Encyclopedia of Food Sciences and Nutrition

Academic Press

EDITOR-IN-CHIEF

Benjamin Caballero
Johns Hopkins University
Center for Human Nutrition
School of Hygiene and Public Health
615 North Wolfe Street
Baltimore, Maryland 21205-2179
USA

EDITORIAL ADVISORY BOARD

FOREWORD

There are no disciplines so all-encompassing in human endeavours as food science and nutrition. Whether it be biological, chemical, clinical, environmental, agricultural, physical – every science has a role and an impact. However, the disciplines of food science and nutrition do not begin or end with science. Politics and ethics, business and trade, humanitarian efforts, law and order, and basic human rights and morality all have something to do with it too.

As disciplines, food science and nutrition answer questions and solve problems. The questions and problems are diverse, and cover the full spectrum of every issue. Life span is one such issue, covered from the nutritional basis for fetal and infant development, to optimal nutrition for the elderly. Another such issue is the time span of the ancient and wild agro-biodiversity that we are working to preserve, to the designer cultivars from biotechnology that we are trying to develop. Still another is the age-old food preparation methods now honoured by the 'eco-gastronomes' of the world, to the high tech food product development advances of recent years.

As with most endeavours, our scientific and technological solutions can and do create new, unforeseen problems. The technologies that gave us an affordable and abundant food supply led to obesity and chronic diseases. The "green revolution" led to loss of some important agro-biodiversity. The technological innovation that gave us stable fats through hydrogenation, flooded the food supply with trans fatty acids. All these problems were identified through a multidisciplinary scientific approach and solutions are known. When technology created the problem and technology has found the solution, implementation is usually more successful. Reducing trans fatty acids in the food supply is case in point. Beyond the technologies, the solutions are more difficult to implement. We know how obesity can be reduced, but the solution is not directly technological. Hence, we show no success in the endeavour.

Of all the problems still confounding us in food science and nutrition, none is so compelling as reducing the number of hungry people in the world. FAO estimates that there are 800 million people who do not have enough to eat. The World Food Summit Plan of Action, the Millennium Development Goals and other international efforts look to food science and nutrition to provide the solution. Yet we only have part of the solution—the science part. The wider world of effort in food science and nutrition needs to be more conscientiously addressed by scientists. This is the world of advocacy and action: advocacy for food and nutrition as basic human rights, coupled with action to get food where it is needed.

But all those efforts would be futile if they are not based on sound scientific information. That is why works such as this Encyclopedia are so important. They provide to a wide readership, scientists and non-scientists alike, the opportunity to quickly gain understanding on specific topics, to clarify questions, and to orient to further reading. It is a pleasure to be involved in such an endeavour, where experts are willing to impart their knowledge and insights on scientific consensus and on exploration of current controversies. All the while, this gives us optimism for a brighter food and nutrition future.

Barbara Burlingame
25 February 2003

INTRODUCTION

There is no factor more vital to human survival than food. The only source of metabolic energy that humans can process is from nutrients and bioactive compounds with putative health benefits, and these come from the food that we eat. While infectious diseases and natural toxins may or may not affect people, everyone is inevitably affected by the type of food they consume.

In evolutionary terms, humans have increased the complexity of their food chain to an astounding level in a relatively short time. From the few staples of some thousand years ago, we have moved to an extraordinarily rich food chain, with many food items that would have been unrecognizable just some hundred years ago.

In this evolution, scientific discovery and technical developments have always gone hand in hand. The identification of vitamins and other essential nutrients last century, and the development of appropriate technologies, led to food fortification, and thus for the first time humans were able to modify foods to better fulfill their specific needs. As a result, nutritional deficiencies have been reduced dramatically or even eradicated in many parts of the world. This evolution is also yielding some undesirable consequences. The abundance of high-density, cheap calorie sources, and the market competition has facilitated overconsumption and promoted obesity, a problem of global proportions.

As the food chain grows in complexity, so does the scientific information related to it. Thus, providing accurate and integral scientific information on all aspects of the food chain, from agriculture and plant physiology to dietetics, clinical nutrition, epidemiology, and policy is obviously a major challenge.

The editors of the first edition of this encyclopedia took that challenge with, we believe, a great deal of success. This second edition builds on that success while updating and expanding in several areas. A large number of entries have been revised, and new entries added, amounting to two additional volumes. These new entries include new developments and technologies in food science, emerging issues in nutrition, and additional coverage of key areas. As always, we have made efforts to present the information in a concise and easy to read format, while maintaining rigorous scientific quality.

We trust that a wide range of scientists and health professionals will find this work useful. From food scientists in search of a methodological detail, to policymakers seeking update on a nutrition issue, we hope that you will find useful material for your work in this book. We also hope that, in however small way, the Encyclopedia will be a valuable resource for our shared efforts to improve food quality, availability, access, and ultimately, the health of populations around the world.

Benjamin Caballero
Luiz Trugo
Paul Finglas

A

Acceptability of Food *See* **Food Acceptability**: Affective Methods; Market Research Methods

ACESULFAME/ACESULPHAME

J F Lawrence, Health and Welfare Canada, Ontario, Canada

Background

Acesulfame K (potassium salt of 6-methyl-1,2,3-oxathiazine-4(3H)-one-2,2-dioxide; Figure 1) is a high-intensity artificial sweetener which is about 200 times as sweet as sucrose (compared to a 3% aqueous sucrose solution). It was accidentally discovered in 1967 by Dr. Karl Clauss, a researcher with Hoechst AG in Frankfurt, FRG, during his experiments on new materials research. The sweetener is not metabolized by the human body and thus contributes no energy to the diet. It is now approved for use in more than 20 countries.

Sweetness

The sweetness properties of acesulfame K are similar to saccharin. It has a clean, sharp, sweet taste with a rapid onset of sweetness and no lingering aftertaste at normal use levels. However, at high concentrations, equivalent to 5% or 6% sucrose solutions, acesulfame K does possess a bitter, chemical aftertaste. The intensity of sweetness of acesulfame K, in common with other artificial sweeteners, varies depending upon its concentration and the type of food application. For example, it is 90 times sweeter than a 6% sucrose solution, 160 times sweeter than a 4% sucrose solution and 250 times sweeter than a 2% sucrose solution. Mixtures of acesulfame K with other intense sweeteners, such as aspartame or cyclamate, result in some synergistic increases in sweetness. Mixtures with saccharin are somewhat less synergistic.

Production and Physical and Chemical Properties

Acesulfame K (Figure 1) is structurally related to saccharin. It also has many of the same physical and chemical properties.

Acesulfame was one of a series of sweet-tasting substances synthesized by Hoechst AG in the late 1960s. All of these had in common the oxathiazinone dioxide ring structure. The synthesis involved reaction of fluorosulfonyl isocyanate with either acetylene derivatives or with active methylene compounds such as α-diketones, α-keto acids, or esters. The latter reaction is used for the commercial production of acesulfame K. A generalized reaction scheme for synthesis of the oxathiazinone dioxide ring structure is shown in Figure 2. Many analoges have been prepared and evaluated for taste properties. The potassium salt of the 6-methyl derivative, acesulfame K, displayed the best sensory and physical properties and thus it has received extensive testing aimed at obtaining approval for its use in diet foods.

Acesulfame K is a white crystalline material which is stable up to 250 °C, at which temperature it decomposes. The free acid form of the sweetener has a distinct melting point of 123.5 °C.

Acesulfame K has a specific density of 1.83. When dissolved in water it produces a nearly neutral solution while the free acid is strongly acidic (pH of a 0.1 mol l^{-1} aqueous solution being 1.15). The sweetener is very soluble in water; a 27% solution can be prepared at 20 °C. The solubility of acesulfame K increases significantly with temperature. At 80 °C, 50% solutions can be prepared; because of this, greater than 99% purity can be obtained by crystallization. It is substantially less soluble in common solvents such as ethanol, methanol, or acetone.

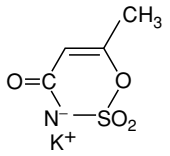

Figure 1 Structure of acesulfame K. Reproduced from Acesulphame/Acesulfame, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

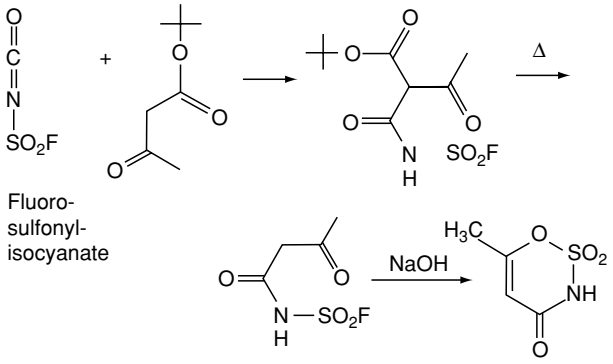

Figure 2 Synthesis of the acesulfame ring structure using fluorosulfonyl isocyanate and *tert*-butylacetoacetate as starting materials. Reproduced from Acesulphame/Acesulfame, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

The stability of acesulfame K in the solid state is very good. It can be stored at ambient temperature for 10 years without decomposition. Aqueous solutions at pH 3 or greater may also be stored for extended periods without detectable decomposition or loss of sweetness. However, below pH 3, significant hydrolysis may occur at elevated temperatures. For example, at pH 2.5 an aqueous buffered solution of acesulfame K would decompose by about 30% after 4 months of storage at 40 °C, whereas no decomposition occurs under the same conditions within the pH range of 3–8. At 20 °C, less than 10% decomposition of acesulfame K occurs after 4 months' storage at pH 2.5, indicating that under normal storage conditions aqueous solutions of the sweetener are very stable.

Acesulfame K is stable under most food-processing conditions, including the elevated temperature treatments encountered in pasteurization and baking.

Food Uses

Because of its stability, acesulfame K has been evaluated in a wide variety of diet food products, including table-top sweeteners, soft drinks, fruit preparations, desserts, breakfast cereals, and chewing gum. Table 1 lists approximate concentration levels of acesulfame K typically used in several types of foods.

Table 1 Typical use levels of acesulfame K in diet foods

Food products	*Concentration (mg kg^{-1})*
Soft drinks	1000
Coffee and tea	267
Jams and marmalades	3000
Ready-to-eat desserts	1000
Chewing gum	2000

Safety and Regulatory Status

Acesulfame K has been subjected to extensive feeding studies in mice, rats, and dogs. The substance is not considered to be carcinogenic, mutagenic, or teratogenic. It is excreted unmetabolized in test animals or humans. The current maximum acceptable daily intake (ADI: the maximum amount that can be consumed daily for a lifetime without appreciable risk) established by the Food and Agriculture Organization/World Health Organization (FAO/WHO) Joint Expert Committee on Food Additives in 1990 is 5 mg per kg body weight. This value is based on the highest amount fed to animals for which there was no effect.

The first regulatory approval for acesulfame K was by the UK in 1983. Since then it has received approval for specific uses in more than 20 countries.

Analysis

Thin-layer chromatography, isotachorphoresis, and high-performance liquid chromatography (HPLC) have been evaluated for the determination of acesulfame K in a variety of matrices, including liquid and solid food products, animal feed, and biological fluids. Of the three, HPLC is perhaps the most useful since the efficiency of the chromatography coupled with selective detection (ultraviolet absorbance) enable quantitative measurements to be made in rather complex food samples. In addition, the sample preparation is minimal, usually involving a water extraction for solid samples or a filtration and dilution of liquid samples before direct HPLC analysis. Acesulfame K has been incorporated into a multisweetener analytical method employing HPLC.

See also: **Carbohydrates**: Sensory Properties; **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **Legislation**: Contaminants and Adulterants; **Saccharin**; **Sweeteners**: Intensive

Further Reading

Franta R and Beck B (1986) Alternatives to cane and beet sugar. *Food Technology* 40: 116–128.

Kretchmer N and Hollenbeck CB (1991) *Sugars and Sweeteners*. Boca Raton: CRC Press.

Lawrence JF and Charbonneau CF (1988) Determination of seven artificial sweeteners in diet food preparations by reverse-phase liquid chromatography with absorbance detection. *Journal of the Association of Official Analytical Chemists* 71: 934–937.

O'Brien-Nabors L and Gelardi RC (1991) *Alternative Sweeteners*. New York: M. Dekker.

ACIDOPHILUS MILK

W Kneifel and C Bonaparte, University of Agricultural Sciences, Vienna, Austria

Background and History

Since the first documentation of the beneficial role of *Lactobacillus acidophilus* in correcting disorders of the human digestive tract in 1922, products containing *L. acidophilus*, especially various types of Acidophilus milk, have become increasingly popular. Today, a multitude of such products are commercially available, many of them being assigned to the category of probiotic foods. Most of these probiotics possess a bacterial microflora of well-documented and scientifically proven bacterial strains with several benefical properties. Besides other categories of foods containing special ingredients, these products have also recently been subclassified under the umbrella of functional foods.

In general, the human body is inhabitated by more than 500 different bacterial species; among them, the lactobacilli play an important ecological role. Besides their important gut-associated function, lactobacilli are also part of various other human-specific microbial ecosystems, e.g., skin, vagina, mouth, nasal, and conjunctival secretions. *L. acidophilus* is the best known of the health-promoting lactobacilli of mammals and a naturally resident species of the human gastrointestinal tract. It colonizes segments of the lower small intestine and parts of the large intestine, together with other lactobacilli species, such as *L. salivarius, L. leichmanii*, and *L. fermentum*. It is interesting to note that these resident *Lactobacillus* species should be distinguished from the spectrum of so-called transient *Lactobacillus* species, which are represented by *L. casei*.

Historically, in 1900, Australian researchers isolated *L. acidophilus* from fecal samples of bottle-fed infants for the first time and named it '*Bacillus acidophilus*.' The actual nomenclature *L. acidophilus* is derived from *acido* (acid) and *philus* (loving) and this designation reflects the acidotolerant potential of this species. In 1959, Rogosa and Sharpe presented a detailed description of this bacterium.

Fundamental Characteristics of *Lactobacillus acidophilus*

Together with 43 other species, *L. acidophilus* is listed as a member of the genus *Lactobacillus* which belongs to the heterogeneous category of lactic acid bacteria. Lactobacilli are Gram-positive, nonmotile, catalase-negative, nonspore-forming rods with varying shapes, ranging from slender, long rods to coccobacillary forms. They are considered as (facultative) anaerobes with microaerophilic properties. *L. acidophilus* usually appears as rods with rounded ends, with a size of 0.6–0.9 × 1.5–6 μm, mainly organized singly or in pairs or short chains (**Figure 1**). The cell wall peptidoglycan is of the Lys-D-Asp type; the mean proportion of guanine and cytosine in the DNA ranges between 34 and 37%. With rare exceptions, this bacterium shows good growth at 45 °C but not below 15 °C, having an optimum growth temperature in the range of 35–38 °C. Substrates with pH values of 5.5–6.0 are preferred. Metabolically, it is a typical obligately homofermentative bacterium and produces racemic lactic acid (both the L+ and the D− enantiomeric forms) from lactose, glucose, maltose, sucrose, and other carbohydrates. Usually, it follows the Embden–Meyerhof–Parnas pathway for glucose metabolism. Important growth factor requirements are acetic or mevalonic acid, riboflavin, pantothenic acid, niacin, folic acid and calcium, but not cobalamin, pyridoxine, and thymidine. Starch and cellobiose are fermented by most strains. Another differential key criterion for the distinction from other lactobacilli (e.g., *L. delbrueckii* subsp. *bulgaricus*) is its capability of cleaving esculin. Further differential criteria are the utilization of trehalose, melibiose, raffinose, ribose, and lactose. While

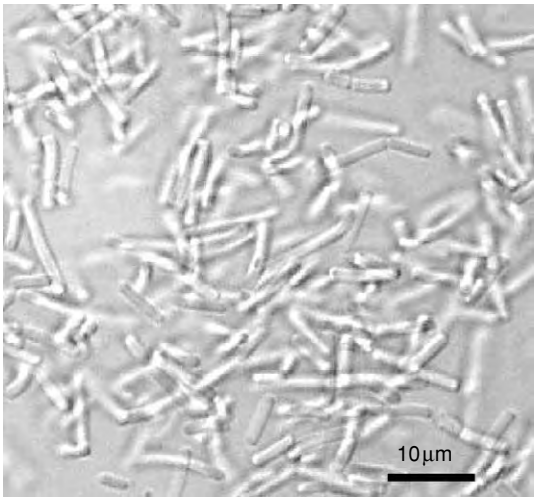

Figure 1 Microphotograph of a *Lactobacillus acidophilus* culture (deep-frozen culture concentrate cultured in MRS broth; for details see **Table 1**).

physiological parameters allow some distinction from other food-relevant lactobacilli, it is not possible to use a phenotypical basis to discriminate sufficiently among *L. acidophilus*, *L. johnsonii*, *L. gasseri*, *L. crispatus*, and *L. amylovorus*. All five species are usually assigned to the *L. acidophilus* cluster. A distinction of these species can be facilitated by applying genotypical techniques and methods based on DNA homology, the molar amounts of guanine plus cytosine in the DNA, or by the analysis of certain cell wall components.

Physiological Actions of *Lactobacillus acidophilus*

Because of the properties described above and its pronounced bile salt resistance, *L. acidophilus* is well adapted to the environmental conditions of the gastrointestinal tract. Proteins in the cell wall may be important in attaching the bacterium to the mucosal cells of the intestine. With strain-dependent variations, *L. acidophilus* contributes to the inhibition of the multiplication of pathogenic and putrefactive bacteria in the intestine due to the production of organic acid and trace amounts of H_2O_2. Furthermore, strain-specific inhibitory substances can be excreted by certain strains. In this context, numerous antagonistic peptides (bacteriocins) have been isolated from certain strains of *L. acidophilus*. For example, some of them were described as lactocidin, acidophilin, acidolin, lactosin B, and lactacin B and possess some 'antibiotic' potential against salmonellae, staphylococci, *Escherichia coli*, and clostridia, and partly also against other species of lactic acid bacteria. Because of their beneficial *L. acidophilus*-related properties, products containing this bacterium have been used in the treatment of gastrointestinal disorders and to reestablish the function of the intestine after treatment with antibiotics. Other features of these products are the provision of β-galactosidase to humans having an enzymatic deficiency for lactose digestion or, particularly when used in conjunction with fructooligosaccharides (oligofructose), the reduction of fecal enzymes (glucuronidase, nitroreductase, azoreductase) which obviously play some role in some stages of precancerogenesis. Since *L. acidophilus* produces equimolar amounts of L(+) and D(−) lactic acid, products fermented with this bacterium offer the advantage of a reduced D(−) lactate content, compared to classical yogurt. However, the acidification potential of this bacterium is often low and varies considerably among strains.

Products with *Lactobacillus acidophilus*

At present, a broad variety of products containing *L. acidophilus* is on the market. This bacterium has been incorporated into fermented as well as nonfermented milks of different levels of dry matter and fat (**Figure 2**). Cows' milk is the main substrate which is processed using the same basal technology as applied for the manufacture of yogurt or other cultured dairy products. Hence, continuous production lines with conventional or aseptic filling systems are used. Some of the products also contain added fruits and flavoring agents.

Fermented dairy products containing *L. acidophilus* as a single bacterial culture are primarily of local importance in Russia, Eastern European countries, and Scandinavia. In contrast, in other European regions, *L. acidophilus* is usually used in combination with other microorganisms (e.g., *Bifidobacterium* spp., *Streptococcus thermophilus, L. delbrueckii* subsp. *bulgaricus*, *L. casei*). Milk cultured with such multicomponent starter cultures (including *L. acidophilus*) is produced in increasing numbers and varieties and consumed frequently by many people. Among these dairy products, a distinction can be made between the so-called 'mild' yogurt products (yogurt-related fermented milks, with or without fruits) which are based on a fermentation with various thermophilic bacteria (many of them are assigned to the area of probiotics), and so-called 'Acidophilus milk' products which are usually fermented by means of mesophilic lactic acid bacteria (e.g., strains of *Lactococcus lactis* or *Leuconostoc cremoris* or combinations of both), in addition to *L. acidophilus*. A general flow diagram for the production of such an Acidophilus milk fermented under mesophilic

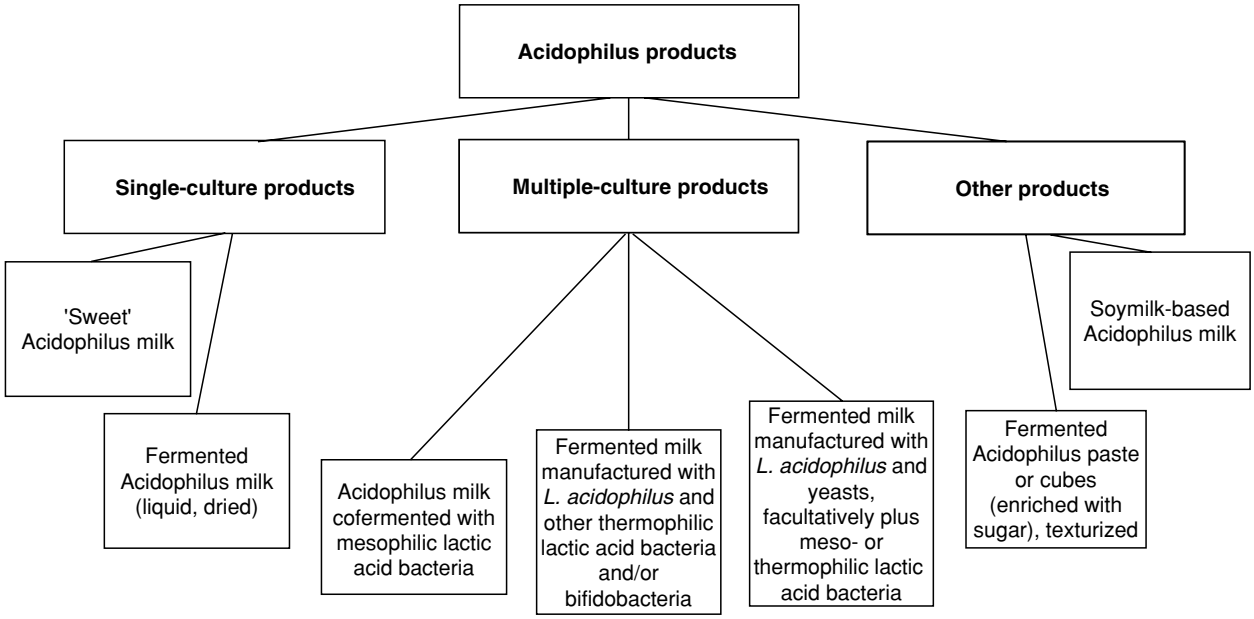

Figure 2 Survey of the diversity of food products containing *Lactobacillus acidophilus*.

conditions is presented in Figure 3. Deep-frozen culture concentrates or freeze-dried bacteria or, very rarely, liquid cultures are inoculated into the milk base. The fermentation is usually performed overnight for 15–20 h. Stirred products, with a liquid character, are usually made, but set-style fermented Acidophilus milk products with increased levels of solid-nonfat are also available.

Other categories of products include specially fermented drinks (e.g., *L. acidophilus* plus yeasts, with or without other lactic acid bacteria, resembling kefir and named acidophilin), texturized products with a reduced water content, which are offered in a pasty form or cut in cubes, or powdered milk which has been fermented with *L. acidophilus* before drying. Many of these product types have a local significance as dietary adjuncts. In Russia, these products even play some role as therapeutic agents and have been well recognized with regard to their medical relevance.

Nonfermented milk containing *L. acidophilus* is also offered by some dairies. Such products are usually produced from standardized milk which is supplemented with a culture concentrate (deep-frozen pellets or lyophylisate) of *L. acidophilus* under cooled conditions, followed by stirring before filling into cartons or beakers. Some of these products are also fortified with fat-soluble vitamins (A, D, E), water-soluble vitamins (thiamin), and trace elements (iron). While a pronounced metabolic activity of the *L. acidophilus* strains is desired for all those products which are produced by fermentation, storage-resistant but not fast-growing cultures (strains) are needed for the manufacture of 'sweet' (nonfermented) Acidophilus milk in order not to alter the sensory properties during storage.

The sensory characteristics of nonfermented 'sweet' Acidophilus milk are comparable with regular milk; those of fermented Acidophilus milk (mesophilic varieties) are similar to those of regular cultured or sour milks which are manufactured using a butter flavor-producing mesophilic culture, since almost no acetaldehyde, which is typical for yogurt, but some diacetyl-based butter aroma is generated during fermentation caused by citrate-fermenting mesophilic lactic acid bacteria. Since *L. acidophilus* possesses alcohol dehydrogenase activity, which is capable of reducing acetaldehyde, only low levels of this compound are found in the corresponding products. Thus, yogurt-related dairy products (thermophilic varieties) containing *L. acidophilus* often exhibit a milder and less acidic taste than classical yogurt, i.e., that manufactured by a cofermention of *S. thermophilus* and *L. delbrueckii* subsp. *bulgaricus*. Sensorically, this classical yogurt is dominated by acetaldehyde which introduces some kind of astringent characteristic and typical sharpness. Moreover, many classical yogurt cultures, in particular owing to the *Lactobacillus* component of the culture, exhibit a continued acidification activity even under cooled conditions on the shelves of retail shops. Besides the sensory changes, this 'overacidification' can also lead to textural problems (syneresis, whey separation).

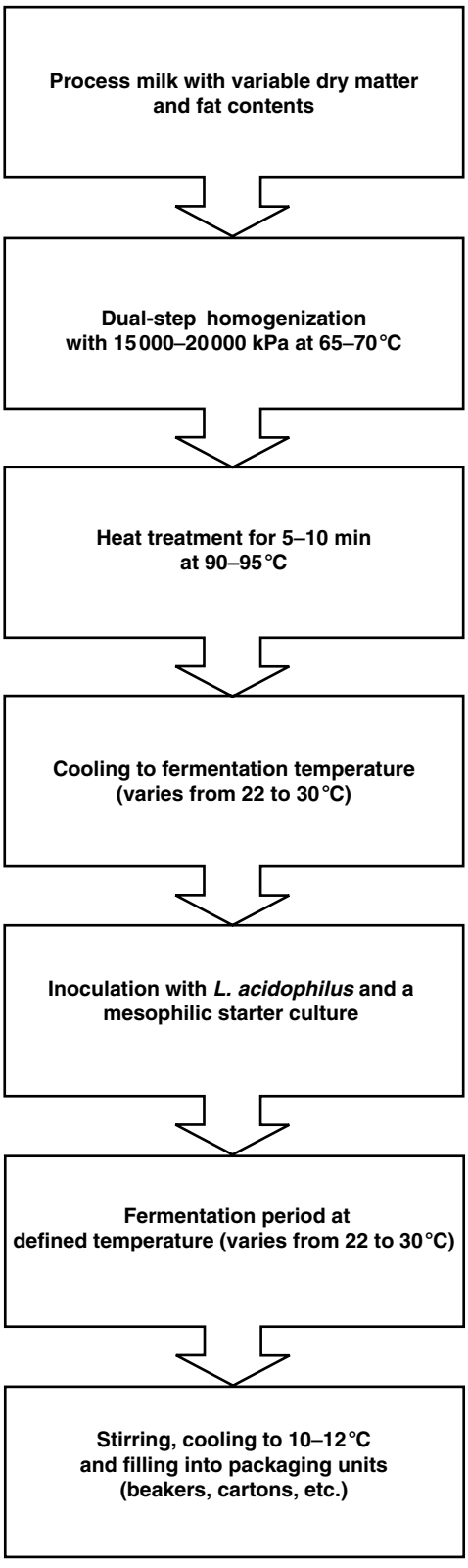

Figure 3 General production steps of the manufacture of fermented Acidophilus milk using a combined fermentation with *Lactobacillus acidophilus* and mesophilic lactic acid starter culture. Data compiled after various manufacturers' recommendations.

Obviously because of these effects, preferences of consumers for the milder yogurts with *L. acidophilus* have been observed in many countries.

A completely different group of Acidophilus products are the pharmaceutical preparations containing *L. acidophilus*. Capsules, suppositories, and soluble powders packaged in sachets are enriched with bacterial lyophylisates and used for therapeutic purposes. They are marketed in pharmacies and health stores.

Bacterial Viable Count and Bacterial Stability of Acidophilus milk

Although milk is a substrate containing almost a universal array of nutrients, it does not fully meet the growth requirements of *L. acidophilus*. For this purpose, additives and growth promoters consisting of a mixture of natural compounds which support and enhance bacterial growth are recommended for supplementation of the fermentation milk by most culture suppliers. Usually, they are added to the milk base in small amounts, before inoculation. In addition, the use of multicomponent cultures offers the advantage of inducing synergistic effects among the bacterial microflora which may also positively influence the propagation rate and the stability of the bacteria.

According to legal aspects and to consumer expectations, products labeled as Acidophilus milk or as containing *L. acidophilus* necessarily have to contain a significant number of these microorganisms. In this context, a group of experts of the International Dairy Federation has recommended that *L. acidophilus* shall be detected in such products at a level of at least 1 million CFU ml^{-1} or g, at their sell-by dates.

Recently, studies performed in several countries have shown that many commercially available products can meet this limit, but with a considerable number of products a decrease in the *L. acidophilus* counts has been observed during a storage period of approximately 3–5 weeks. Due to the fact that the expression of beneficial effects is based on a high number of active bacteria, a high viable count and pronounced bacterial stability have become important goals in product development and optimization.

Viable counts of *L. acidophilus*-containing dairy products are usually enumerated by culture methods based on plate count techniques with media designed for culturing lactic acid bacteria (e.g., MRS, Rogosa agar, TGV agar; for details see Table 1). To enhance the discriminatory power of these media (this is of particular relevance in the examination of products which contain a mixed microflora), media are modified by slight acidification and/or by supplementation

Table 1 Media used for culturing Lactobacillus acidophilus

Media	*References*
MRS agar	Lactobacillus agar according to De Man JD, Rogosa M and Sharpe ME (1960) *Journal of Applied Bacteriology* 23: 130–135.
Rogosa agar	Lactobacillus selective agar according to Rogosa M Mitchell and JA, Wiseman RF (1951) *Journal of Bacteriology* 62: 132–133.
TGV agar	Agar medium according to Galesloot T, Hassing F and Stadhouders J (1961) *Netherlands Milk and Dairy Journal* 15: 127–150.

with antibiotics (e.g., vancomycin at different levels) or with other selective agents (cellobiose, conjugates with chromogenic indicator dyes, esculin, etc.). In many cases, the parallel use of different media selective for each of the bacterial components is necessary to allow the reliable microbiological monitoring of these Acidophilus products. Moreover, microscopical verification of isolates harvested from the different media usually completes their routine assessment. Although a number of media and methodologies have been described in the literature, no official standard method is available yet.

See also: **Fermented Milks**: Types of Fermented Milks; **Functional Foods**; **Lactic Acid Bacteria**; **Probiotics**; **Yogurt**: The Product and its Manufacture; Yogurt-based Products; Dietary Importance

Further Reading

Fondén R, Mogensen G, Tanaka R and Salminen S (2000) Effect of culture-containing dairy products on intestinal microflora, human nutrition and health – current knowledge and future perspectives. In: *IDF Bulletin*, no. 352, pp. 5–30, Brussels: International Dairy Federation.

Hammes WP (1995) The genus *Lactobacillus*. In: Wood JB and Holzapfel WH (eds) *The Genera of Lactic Acid Bacteria*, pp. 19–54. London: Blackie Academic & Professional.

Kanbe M (1992) Uses of intestinal lactic acid bacteria and health. In: Nakazawa Y amd Hosono A, (eds) *Functions of Fermented Milk. Challenges for the Health Sciences*, pp. 289–304. London: Elsevier Applied Science.

Kneifel W and Pacher B (1993) An X-Glu based agar medium for the selective enumeration of *Lactobacillus acidophilus* in yogurt-related milk products. *International Dairy Journal* 3: 277–291.

Lee YK, Nomoto K, Salminen S and Gorbach SL (1999) *Handbook of Probiotics*. New York: John Wiley.

Mital BK and Garg SK (1992) Acidophilus milk products: manufacture and therapeutics. *Food Reviews International* 8: 347–389.

Tamime AY and Robinson RK (1999) *Yoghurt Science and Technology*. Cambridge: CRC, Woodhead Publishing.

ACIDS

Contents

Properties and Determination

J D Dziezak, Dziezak & Associates, Ltd., Hoffman Estates, IL, USA

Background

In very general terms, an acid is a compound that contains or produces hydrogen ions in aqueous solutions, has a sour taste, and turns blue litmus paper red. A more comprehensive definition, given by the US chemist G.N. Lewis, states that acids are substances that can accept an electron pair or pairs, and bases are substances that can donate an electron pair or pairs. This definition, applicable to both nonaqueous and aqueous systems, requires that an acid be either a positive ion or a molecule with one or more electron-deficient sites with respect to a corresponding base.

The definition most widely used to describe acid–base reactions in dilute solution is one that was proposed independently by two scientists in 1923 – the Danish chemist J.N. Brønsted and the US chemist T.M. Lowry. The Brønsted–Lowry theory defines an acid as a proton donor, that is, any substance (charged or uncharged) that can release a hydrogen ion or proton. A base is defined as a proton acceptor or any substance that can accept a hydrogen ion or proton.

Table 1 Structure, ionization constant, pK_a, and key physical and chemical properties of acidulants[a]

Acid	*Structure*	*Ionization constant(s)*	*pK_a*	*Physical form*	*Melting point (°C)*	*Solubility (g per 100 ml of water)*	*Hygroscopicity*	*Taste characteristics*
Acetic acid	CH_3COOH	1.76×10^{-5} at 25 °C	4.76	Clear, colorless liquid	−8.5	Soluble	na	Tart and sour
Adipic acid	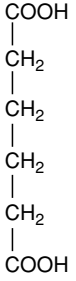	$K_1 = 3.71 \times 10^{-5}$ $K_2 = 3.87 \times 10^{-6}$ at 25 °C	4.43 5.41	Crystalline powder	152	1.9 g at 20 °C 83 g at 90 °C	Low level of hygroscopicity	Smooth lingering tartness; complements grape flavors
Citric acid	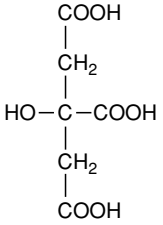	$K_1 = 7.10 \times 10^{-4}$ $K_2 = 1.68 \times 10^{-5}$ $K_3 = 6.4 \times 10^{-7}$ at 25 °C	3.14 4.77 6.39	Crystalline powder			Moderately hygroscopic	Tart; delivers a 'burst' of flavor
Anhydrous Hydrous					153 135–153	181 g at 25 °C 208 g at 25 °C		
Fumaric acid	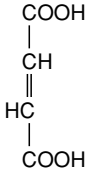	$K_1 = 9.30 \times 10^{-4}$ $K_2 = 3.62 \times 10^{-5}$ at 18 °C	3.03 4.44	White granules or crystalline powder	286	0.5 g at 20 °C 9.8 g at 100 °C	Nonhygroscopic	Tart; has an affinity for grape flavors

Glucono-δ-lactone	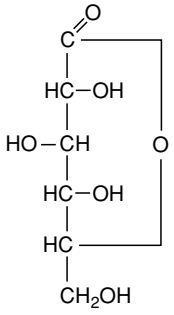	1.99×10^{-4} (for gluconic acid)	3.7	White crystalline powder	153	59 g at 25 °C	Nonhygroscopic	Neutral taste with acidic aftertaste, when hydrolyzed
Lactic acid	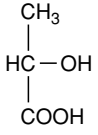	1.37×10^{-4} at 25 °C	3.86	Liquid; also available in dry form	16.8	Very soluble	na	Acrid
Malic acid	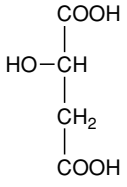	$K_1 = 3.9 \times 10^{-4}$ $K_2 = 7.8 \times 10^{-6}$ at 25 °C	3.40 5.11	Crystalline powder	132	62 g at 25 °C	Nonhygroscopic	Smooth tartness
Phosphoric acid		$K_1 = 7.52 \times 10^{-3}$ $K_2 = 6.23 \times 10^{-8}$ $K_3 = 2.2 \times 10^{-13}$ K_1 and K_2 at 25 °C; K_3 at 18 °C	2.12 7.21 12.67	Liquid		Very soluble in hot water	na	Acrid
Tartaric acid	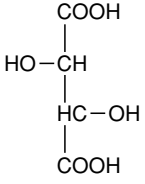	$K_1 = 1.04 \times 10^{-3}$ $K_2 = 4.55 \times 10^{-5}$ at 25 °C	2.98 4.34	Crystalline powder	168–170	147 g at 25 °C	Nonhygroscopic	Extremely tart; augments fruit flavors, especially grape and lime

na, not applicable.

[a]Adapted with permission from *Food Technology*, *Acidulants: Ingredients that do more than meet the acid test*. 44(1): 76–83. Institute of Food Technologists, Chicago, Illinois, USA.

This article discusses the physicochemical properties of acids and describes several methods for their analysis.

Strong Versus Weak Acids

The strength of a Brønsted–Lowry acid depends on how easily it releases a proton or protons. In strong acids, owing to their weaker internal hydrogen bonds, the protons are loosely held. As a result, in aqueous solutions, almost all of the acid reacts with water, leaving only a few unionized acid molecules in the equilibrium mixture. The reaction takes place according to eqn (1):

$$HA + H_2O \rightleftharpoons H_3O^+ + A^- \qquad (1)$$

In this equation, HA represents the undissociated acid, H_3O^+ the hydronium ion formed when a proton combines with one molecule of water, and A^- the conjugate base of HA.

Unlike strong acids, weak acids exist largely in the undissociated state when mixed with water, since only a small percentage of their molecules interact with water and dissociate. Most acids found in foods, including acetic, adipic, citric, fumaric, malic, phosphoric and tartaric acids, and glucono-δ-lactone, are classified as weak or medium strong acids.

Physicochemical Properties

Physicochemical properties, including the ionization constant, pH, the apparent dissociation constant (pK_a) and buffering capacity, are discussed below and are listed in Table 1.

Ionization Constant

The tendency for an acid or acid group to dissociate is defined by its ionization constant, also denoted as pK_a. The ionization constant, given at a specified temperature, is expressed as:

$$K_a = \frac{[H_3O^+][A^-]}{[HA]}, \qquad (2)$$

where the brackets designate the concentration in moles per liter. The ionization constant is a measure of acid strength: the higher the K_a value, the greater the number of hydrogen ions liberated per mole of acid in solution and the stronger the acid.

Acids with more than one transferable hydrogen ion per molecule are termed 'polyprotic' acids. Monoprotic or monobasic acids are those that can liberate one hydrogen ion, such as acetic acid and lactic acid. Those containing two transferable hydrogen ions are called diprotic or dibasic acids and include, for example, adipic acid and fumaric acid. Acids such as citric acid and phosphoric acid, which have three transferable hydrogens, are called triprotic or tribasic acids. Ionization of polyprotic acids occurs in a stepwise manner with the transfer of one hydrogen ion at a time. Each step is characterized by a different ionization constant.

pH

Measurement of acidity is an important aspect of ascertaining the safety and quality of foods. Such measurements are given in terms of pH, which is defined as the negative logarithm of the hydronium ion concentration (strictly, activity):

$$pH = \log_{10}\frac{1}{[H_3O^+]} = -\log_{10}[H_3O^+]. \qquad (3)$$

The lower the pH value, the higher the hydrogen ion concentration associated with it. A pH value of less than 7 indicates a hydrogen ion concentration greater than 10^{-7} M and an acidic solution; a pH value of more than 7 indicates a hydrogen ion concentration of less than 10^{-7} M and a basic solution. When the hydronium and hydroxide ions are equal in concentration, the solution is described as neutral. (*See* **pH – Principles and Measurement**.)

It is also important to note that, because the pH scale is logarithmic, a difference of one pH unit represents a 10-fold difference in hydrogen ion concentration.

pK_a

The term pK_a is defined as the negative logarithm of the dissociation constant:

$$pK_a = \log_{10}\frac{1}{K_a} = -\log_{10} K_a. \qquad (4)$$

The pK_a corresponds to the pH value at the midpoint of a titration curve developed when one equivalent of weak acid is titrated with base, and the pH resulting from each incremental addition of base is plotted against the equivalents of hydroxide ions added.

The pH of a system is at the pK_a when the concentrations of acid (HA) and conjugate base (A^-) are equal. At the pK_a and, to a lesser extent, in the area extending to within one pH unit on either side of the pK_a, the system resists changes in pH resulting from addition of small increments of acid or base. In other words, at the pK_a, acids and their salts function as buffers.

The number of pK_as that an acid has depends on the number of hydrogen ions it can liberate. Monoprotic acids have a single pK_a, whereas di- and triprotic acid have two and three pK_as, respectively.

Strong acids have low pK_a values, and strong bases have high pK_a values.

Buffering Capacity

A solution of a weak acid (or a weak base) and its corresponding salt is called a buffer solution. In these systems, the hydronium ion content is not significantly changed when a small amount of acid or base is added to that solution. The reason that buffer solutions resist appreciable changes in pH can be best illustrated by an example. If a small amount of hydrochloric acid is added to a buffer solution composed of acetic acid and sodium acetate, the protons from the hydrochloric acid would associate with the acetate ions to form unionized molecules of acetic acid. As the newly formed acid molecules ionize, the equilibrium would shift towards forming more hydronium ions (eqn (1)). This would result in only a very slight increase in pH.

Similarly, the addition of a small amount of sodium hydroxide to the same buffer solution would have little effect on pH. Hydroxide ions from the sodium hydroxide would combine with hydronium ions in the equilibrium mixture, forming undissociated molecules of sodium hydroxide. More of the acid molecules would then dissociate to replace the hydronium ions lost; though a new equilibrium system would be created, it would produce only a minimal effect on pH.

The quantity of acid or base that a buffer solution is capable of consuming before a change in pH is realized is termed the 'buffering capacity.' The buffering capacity is defined as the number of moles of strong acid or base required to increase the pH by one unit in 1 l of buffer solution. The buffering capacity of a solution is greatest at its pK_a value where the concentrations of acid and conjugate base are equal.

Analytical Methods

Quantitative determinations of acidity play an important role in ensuring food product quality and stability. Information obtained on acid levels can help in detecting cases of food adulteration, monitoring fermentation processes, and evaluating the organoleptic properties of fermented foods. pH determination, titratable acidity, chromatographic methods, and capillary electrophoresis are procedures commonly employed by the food industry to determine food acids. (*See* **Adulteration of Foods**: Detection.)

pH Determination

pH can be measured by two techniques: colorimetric and potentiometric. The colorimetric method involves adding a suitable indicator to a solution and matching the color of the solution to a standard solution containing the same indicator. This method can estimate pH to the nearest 0.1 pH unit.

A more accurate technique and the one most frequently employed, the potentiometric method, uses a pH meter to determine hydrogen ion concentration. The two electrodes of the meter – a calomel reference electrode and a glass indicator electrode – are immersed in the solution, of known temperature, whose pH is to be measured. The electrode potential of the indicator electrode is linearly related to changes in hydrogen ion concentration and therefore pH.

Titratable Acidity

The total concentration of acid in a solution can be determined by titration. The titration process is performed by placing in a flask a known volume of acid solution whose concentration is unknown. To the flask, a few drops of indicator, e.g., phenolphthalein, which is colorless in acid solutions and pink in basic solutions, is introduced. A base solution of known concentration is then gradually added until the acid is completely neutralized. This point is indicated when the solution permanently changes color. The concentration of acid can then be calculated from the volume of base solution used.

The value obtained, called titratable acidity, is an estimate of the total acid in the solution. It accounts for both the free hydronium ions present in the equilibrium mixture and the hydrogen ions released from undissociated acid molecules. For weak acids, the titratable acidity is different from the actual acidity (hydrogen ion concentration), since these compounds exist largely in the undissociated state in solution. For strong acids, however, titratable acidity and actual acidity are virtually the same, since strong acids and their salts are completely ionized in solution.

Chromatographic Methods

Gas chromatography (GC) and high-performance liquid chromatography (HPLC) have almost entirely replaced paper and thin-layer chromatography as methods for identifying and quantifying food acids.

Gas Chromatography GC has been used to analyze organic acids in fruit and fruit juice. Analysis involves preparing volatile derivatives such as methyl esters of the organic acids, prior to their injection into the gas chromatograph. Derivatives are chromatographed on a nonpolar stationary phase column and detected by a flame ionization detector.

By use of GC, malic acid has been shown to be a major constituent of many fruits, including apples, pears, grapes, peaches, and nectarines, and significant

levels of citric acid have been found in citrus fruits such as orange, lemon, and grapefruit, and in non-citrus fruits, including pears, nectarines, cherries, and strawberries. (*See* **Chromatography**: Gas Chromatography.)

High-performance Liquid Chromatography HPLC is used more extensively than GC to determine organic acids because the technique requires little or no chemical modification to separate these nonvolatile compounds. Separation is usually done on either a reversed-phase C8 or C18 column or a cation-exchange resin column operated in the hydrogen mode. Acids are detected by either refractive index (RI) or ultraviolet (UV) detectors. RI detection requires prior removal of any sugars present that potentially can interfere with quantification; sugar removal is not required for UV detection at 220–230 nm.

Adulteration of a commercial cranberry juice drink was detected using HPLC when the test yielded different results for organic acids, sugars, and anthocyanin pigments than those obtained for a standard juice drink. Atypical citric and/or malic acid contents and presence of a natural colorant, probably grape skin extract, confirmed that the drink was adulterated.

In wine-making, HPLC is used to monitor concentrations of tartaric, malic, succinic, citric, lactic, and acetic acids, which contribute tartness and stability to the finished product. A common approach involves using a column containing a strong cation exchange resin and eluting the sample with dilute sulfuric acid; the eluant is then analyzed for acids by RI detection. This column has the additional advantage of permitting the simultaneous detection and quantification of ethanol and the monitoring of wine for adulteration with methanol. Organic acids in wine can also be separated using ion chromatography with a conductivity detector. (*See* **Chromatography**: High-performance Liquid Chromatography.)

Capillary Electrophoresis

A relatively new technique, capillary electrophoresis, is also useful for separating and quantifying organic acids in food systems. This technique utilizes an electrical field to separate molecules on the basis of their charge and size. Small volumes of sample, usually a few nanoliters, are injected on to a fused silica capillary tube, which is usually less than 1 m in length and 50 μm in internal diameter. The ends of the tube are placed in electrolyte reservoirs containing electrodes. A voltage in the range of 20–30 kV is delivered to the electrodes by a power supply and causes the charged molecules to move. Because organic acids are negatively charged, they migrate away from more neutral or positively charged molecules, such as sugars and phenols, respectively. Acids are detected by a UV detector, and the signal is sent to a data collector. The resulting separation is graphically represented as an electrophoregram.

Enzymatic Analysis

Enzyme assays provide another means of analyzing acids. For example, an enzymatic assay of L-malic acid uses an NAD(P)-linked malic enzyme and involves spectrophotometrically measuring the absorbance of NADPH, a reaction product, at 340 nm.

See also: **Adulteration of Foods**: Detection; **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **pH – Principles and Measurement**

Further Reading

Fennema OR (ed.) (1979) *Food Chemistry. Principles of Food Science, Part 1*. New York: Marcel Dekker.

Lehninger AL (1975) *Biochemistry*, 2nd edn. New York: Worth.

Macrae R (1988) *HPLC in Food Analysis*. London: Academic Press.

Pomeranz Y and Meloan CE (1978) *Food Analysis: Theory and Practice*. Westport: AVI.

Suye S, Yoshihara N and Shusei I (1992) Spectrophotometric determination of L-malic acid with a malic enzyme. *Bioscience, Biotechnology, and Biochemistry* 56(9): 1488–1489.

Natural Acids and Acidulants

J D Dziezak, Dziezak & Associates, Ltd., Hoffman Estates, IL, USA

Background

Acids, or acidulants as they are also called, are commonly used in food processing as flavor intensifiers, preservatives, buffers, meat-curing agents, viscosity modifiers, and leavening agents. This article discusses the functions that acidulants have in food systems and reviews the more commonly used food acidulants.

Functions of Acidulants

The reasons for using acidulants in foods are numerous and depend on what the food processor hopes to accomplish. As outlined above, the principal reasons

for incorporating an acidulant into a food system are flavor modification, microbial inhibition, and chelation.

Flavor Modification

Sourness or tartness is one of the five major taste sensations: sour, salty, sweet, bitter, and umami (the most recently determined). Unlike the sensations of sweetness and bitterness, which can be developed by a variety of molecular structures, sourness is evoked only by the hydronium ion of acidic compounds.

Each acid has a particular set of taste characteristics, which include the time of perceived onset of sourness, the intensity of sourness, and any lingering of aftertaste. Some acids impart a stronger sour note than others at the same pH. As a general rule, weak acids have a stronger sour taste than strong acids at the same pH because they exist primarily in the undissociated state. As the small amount of hydronium ions is neutralized in the mouth, more undissociated acid (HA) molecules ionize to replace the hydronium ions lost from equilibrium (eqn (1)). The newly released hydronium ions are then neutralized until no acid remains. Taste characteristics of the acid are an important factor in the development of flavor systems.

$$HA + H_2O \rightarrow H_3O^+ + A^- \qquad (1)$$

As pH decreases, the acid becomes more undissociated and imparts more of a sour taste. For example, the intense sour notes of lactic acid at pH 3.5 may be explained by the fact that 70% of the acid is undissociated at this pH, compared with 30% for citric acid. In addition to sourness, acids have nonsour characteristics such as bitterness and astringency, though these are less perceptible. At pH values between 3.5 and 4.5, lactic acid is the most astringent. Acids also have the ability to modify or intensify the taste sensations of other flavor compounds, to blend unrelated taste characteristics, and to mask undesirable aftertastes by prolonging a tartness sensation. For example, in fruit drinks formulated with low-caloric sweeteners, acids mask the aftertaste of the sweetener and impart the tartness that is characteristic of the natural juice. In another example, in substitutes for table salt, acids remove the bitterness from potassium chloride and provide the salty taste of sodium chloride. Other acids, such as glutamic and succinic acids, possess flavor-enhancement properties. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics; **Sensory Evaluation**: Taste.)

Because acids are rarely found in nature as a single acid, the combined use of acids simulates a more natural flavor. Two acids that are frequently blended together are lactic and acetic.

Microbial Inhibition

Acidulants act as preservatives by retarding the growth of microorganisms and the germination of microbial spores which lead to food spoilage. The effect is attributed to both the pH and the concentration of the acid in its undissociated state. It is primarily the undissociated form of the acid which carries the antimicrobial activity: as the pH is lowered, this helps shift the equilibrium in favor of the undissociated form of the acid, thereby leading to more effective antimicrobial activity. The nature of the acid is also an important factor in microbial inhibition: weak acids are more effective at the same pH in controlling microbial growth. Acids affect primarily bacteria because many of these organisms do not grow well below about pH 5; yeasts and molds, in comparison, are usually acid-tolerant. (*See* **Spoilage**: Bacterial Spoilage; Molds in Spoilage; Yeasts in Spoilage.)

In fruit- and vegetable-canning operations, the combined use of heat and acidity permits sterilization and spore inactivation to be achieved at lower temperatures; this minimizes the degradation of flavor and structure that generally results from processing. (*See* **Canning**: Principles.)

Acidification also improves the effectiveness of antimicrobial agents such as benzoates, sorbates, and propionates. For example, sodium benzoate – an effective inhibitor of bacteria and yeasts – does not exert its antimicrobial activity until the pH is reduced to about 4.5. (*See* **Preservation of Food**.) Blends of acids act synergistically to inhibit microbial growth. For example, lactic and acetic acids have been found to inhibit the outgrowth of heterofermentative lactobacilli.

Chelation

Oxidative reactions occur naturally in foods. They are responsible for many undesirable effects in the product, including discoloration, rancidity, turbidity, and degradation of flavor and nutrients. As catalysts to these reactions, metal ions such as copper, iron, manganese, nickel, tin, and zinc need to be present in only trace quantities in the product or on the processing machinery. (*See* **Oxidation of Food Components**.)

Many acids chelate the metal ions so as to render them unavailable; the unshared pair of electrons in the molecular structure of acids promotes the complexing action. When used in combination with antioxidants such as butylated hydroxyanisole, butylated hydroxytoluene, or tertiary butylhydroquinone, acids have a synergistic effect on product stability. Citric acid and its salts are the most widely used chelating agents. (*See* **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants.)

Other Functions

One of the most common reasons for adding acids is to control pH. This is usually done as a means to retard enzymatic reactions, to control the gelation of certain hydrocolloids and proteins, and to standardize pH in fermentation processes. In the first example, the lowering of pH inactivates many natural enzymes which promote product discoloration and development of off-flavors. Polyphenol oxidase, for example, oxidizes phenols to quinones, which subsequently polymerize, forming brown melanin pigments that discolor the cut surfaces of fruits and vegetables. The enzyme is active between pH 5 and 7 and is irreversibly inactivated at a pH of 3 or lower. In the second example, acidification to 2.5–3 is required for high-methoxyl pectins to form gels. Because pH influences the gel-setting properties and the gel strength obtained, proper pH control is critical in the production of pectin- and gelatin-based desserts, jams, jellies, preserves, and other products. In the final example, standardization of pH is done routinely in fermentation processes, such as wine-making, to ensure optimum microbial activity and to discourage growth of undesirable microbes. Acids are also added postfermentation to stabilize the finished wine. (*See* **Beers**: Biochemistry of Fermentation; **Colloids and Emulsions**; **Enzymes**: Functions and Characteristics; **Phenolic Compounds**.)

Acid salts function as buffers in various systems. (*See* **Acids**: Properties and Determination.) For example, in confectionery products, acid salts are used to control the inversion of sucrose into its constituents, glucose and fructose, the latter being hygroscopic. The resulting lower concentration of fructose yields a less hygroscopic food system and a longer shelf-life.

Acids are a major component of chemical leavening systems, where they remain nonreactive until the proper temperature and moisture conditions are attained. The gas evolved by reaction of the acid with bicarbonate produces the aerated texture that is characteristic of baked products such as cakes, biscuits, doughnuts, pancakes, and waffles. The onset and the rate of reaction of these compounds are controlled by such factors as the solubility of the acid, the mixing conditions for preparing the batter, and the temperature and moisture of the batter. Many chemical leavening systems are based on salts of phosphoric and tartaric acids. (*See* **Leavening Agents**.)

Acids have also been used for other purposes. For example, they are added to chewing gum to stabilize aspartame and to cheese to impart favorable textural properties and sensory attributes.

Commonly Used Acidulants

Among the most widely used acids are acetic, adipic, citric, fumaric, lactic, malic, phosphoric, and tartaric acids. Glucono-δ-lactone, though not itself an acid, is regarded as an acidulant because it converts to gluconic acid under high temperatures.

Acetic Acid

Acetic acid is the major characterizing component of vinegar. Its concentration determines the strength of the vinegar, a value termed 'grain strength,' which is equal to 10 times the acetic acid concentration. Vinegar containing, for example, 6% acetic acid has a grain strength of 60 and is called 60-grain. Distillation can be used to concentrate vinegar to the desired strength. (*See* **Vinegar**.)

Fermentation conducted under controlled conditions is the commercial method for vinegar production. Bacterial strains of the genera *Acetobacter* and *Acetomonas* produce acetic acid from alcohol which has been obtained from a previous fermentation involving a variety of substrates such as grain and apples. Vinegar functions in pH reduction, control of microbial growth, and enhancement of flavor. It has found use in a variety of products, including condiments such as ketchup, mustard, mayonnaise, and relish, salad dressings, marinades for meat, poultry, and fish, bakery products, soups, and cheeses. Pure (100%) acetic acid is called glacial acetic acid because it freezes to an ice-like solid at 16.6 °C. Though not widely used in food, glacial acetic acid provides acidification and flavoring in sliced, canned fruits and vegetables, sausage, and salad dressings.

Adipic Acid

Adipic acid, a white, crystalline powder, is characterized by low hygroscopicity and a lingering, high tartness that complements grape-flavored products and those with delicate flavors. The acid is slightly more tart than citric acid at any pH. Aqueous solutions of the acid are the least acidic of all food acidulants, and have a strong buffering capacity in the pH range 2.5–3.0.

Adipic acid functions primarily as an acidifier, buffer, gelling aid, and sequestrant. It is used in confectionery, cheese analogs, fats, and flavoring extracts. Because of its low rate of moisture absorption, it is especially useful in dry products such as powdered fruit-flavored beverage mixes, leavening systems of cake mixes, gelatin desserts, evaporated milk, and instant puddings.

Citric Acid

The most widely used organic acid in the food industry, citric acid, accounts for more than 60% of all acidulants consumed. It is the standard for evaluating the effects of other acidulants. Its major advantages include its high solubility in water; appealing effects on flavor, particularly its ability to deliver a 'burst' of tartness; strong metal chelation properties; and the widest buffer range of the food acids (2.5–6.5).

Citric acid is naturally present in animal and plant tissues and is most abundantly found in citrus fruits including the lemon (4–8%), grapefruit (1.2–2.1%), tangerine (0.9–1.2%) and orange (0.6–1.0%). (*See* **Citrus Fruits**: Composition and Characterization.)

The principal method for commercial production of the acid is fermentation of corn. Formerly, the acid had been obtained by extraction from citrus and pineapple juices. Citric acid is available in a liquid form, which solves processing problems related to incorporating the acid into a food system, such as predissolving citric acid crystals and caking or crystallate deposits on processing equipment. Also available are granulated forms which allow the particle size to be customized to meet the particular need.

Citric acid has numerous applications. It is commonly added to nonalcoholic beverages where it complements fruit flavors, contributes tartness, chelates metal ions, acts as a preservative, and controls pH so that the desired sweetness characteristics can be achieved. Sodium citrate subdues the sharp acid notes in highly acidified carbonated beverages; in club soda, it imparts a cool, saline taste and helps retain carbonation. The acid is also used in wine production both prior to and after fermentation for adjustment of pH; in addition, because of its metal-chelating action, the acid prevents haze or turbidity caused by the binding of metals with tannin or phosphate. The calcium salt of citric acid is used as an anticaking agent in fructose-sweetened, powdered soft drinks, where it neutralizes the alkalinity of other ingredients that support browning, such as magnesium oxide and tricalcium phosphate.

Citric acid has also found use in confectionery and desserts. In hard confectionery, buffered citric acid imparts a pleasant tart taste; it is added to the molten mass after cooking, as this prevents sucrose inversion and browning. Citric acid is used in gelatin desserts because it imparts tartness, acts as a buffering agent, and increases the pH for optimum gel strength.

Low levels of the acid, ranging from 0.001 to 0.01%, work with antioxidants to retard oxidative rancidity in dry sausage, fresh pork sausage, and dried meats. Citric acid is also used in the production of frankfurters: 3–5% solutions are sprayed on the casings after stuffing and prior to smoking to aid in their removal from the finished product. Used at 0.2% in livestock blood, sodium citrate and citric acid act as anticoagulants, sequestering the calcium required for clot formation so that the blood may be used as a binder in pet foods.

In seafood processing, citric acid inactivates endogenous enzymes and promotes the action of antioxidants, resulting in an increased shelf-life. Citric acid also chelates copper and iron ions that catalyze the oxidative formation of off-flavors and fishy odors associated with dimethylamine. In processed cheese and cheese foods, citric acid and sodium citrate function in emulsification, buffering, flavor enhancement, and texture development. Sodium citrate is also combined with sodium phosphate as a customized emulsification salt for processed cheese. Cogranulation of citric acid with malic and fumaric acids yields new tart flavor profiles.

Fumaric Acid

The extremely low rate of moisture absorption of this acid makes it an important ingredient for extending the shelf-life of powdered food products such as gelatin desserts and pie fillings. Fumaric acid can be used in smaller quantities than citric, malic, and lactic acids to achieve similar taste effects.

Fermentation of glucose or molasses by certain *Rhizopus* spp. is the method used to produce fumaric acid commercially. The acid is also made by isomerization of maleic acid with heat or a catalyst, and is a byproduct of the production of phthalic and maleic anhydrides. Fumaric acid is also made in particulate form, where the acid makes up about 5–95% of the particulate, with the remainder being other acids such as malic, tartaric, citric, lactic, ascorbic, and related mixtures.

Applications of fumaric acid include rye bread, jellies, jams, juice drinks, candy, water-in-oil emulsifying agents, reconstituted fats, and dough conditioners. In refrigerated biscuit doughs, the acid eliminates crystal formations that may occur in all-purpose leavening systems. In wine, it functions as both an acidulant and a clarifying aid, although it does not chelate copper or iron.

Glucono-δ-lactone (GDL)

A natural constituent of fruits and honey, GDL is an inner ester of D-gluconic acid. Unlike other acidulants, it is neutral and gives a slow rate of acidification. When added to water, it hydrolyzes to form an equilibrium mixture of gluconic acid and its δ- and γ-lactones. The acid formation takes place slowly when cold and accelerates when heated. As

GDL converts to gluconic acid, its taste characteristics change from sweet to neutral with a slight acidic afteraste.

GDL is produced commercially from glucose by a fermentation process that uses enzymes or pure cultures of microorganisms such as *Aspergillus niger* or *Acetobacter suboxydans* to oxidize glucose to gluconic acid. GDL is extracted by crystallization from the fermentation product, an aqueous solution of gluconic acid and GDL.

Because of its gradual acidification, bland taste, and metal-chelating action, GDL has found application in mild-flavored products such as chocolate products, tofu, milk puddings, and creamy salad dressings. In cottage cheese prepared by the direct-set method, GDL ensures development of a finer-textured finished product, void of localized denaturation. It also shortens production time and increases yields. In cured-meat products, GDL reduces cure time, inhibits growth of undesirable microorganisms, promotes color development, and reduces nitrate and nitrite requirements. (*See* **Curing**.)

Lactic Acid

Lactic acid is one of the earliest acids to be used in foods. It was first commercially produced about 60 years ago, and only within the past two decades has it become an important ingredient. The mild taste characteristics of the acid do not mask weaker aromatic flavors. Lactic acid functions in pH reduction, flavor enhancement, and microbial inhibition. Two methods are used commercially to produce the acid: fermentation and chemical synthesis. Most manufacturers using fermentation are in Europe.

Confectionery, bakery products, beer, wine, beverages, dairy products, dried egg whites, and meat products are examples of the types of products in which lactic acid is used. The acid is used in packaged Spanish olives where it inhibits spoilage and further fermentation. In cheese production, it is added to adjust pH and as a flavoring agent.

Malic Acid

This general-purpose acidulant imparts a smooth, tart taste which lingers in the mouth, helping to mask the aftertastes of low- or noncaloric sweeteners. It has taste-blending and flavor-fixative characteristics and a relatively low melting point with respect to other solid acidulants. The low melting point allows it be homogeneously distributed into food systems. Compared with citric acid, malic acid has a much stronger apparent acidic taste. As DL-malic acid is the most hygroscopic of the acids, resulting in lumping and browning in dry mixes, the encapsulated form of this acid is preferred for dry mixes.

Malic acid occurs naturally in many fruits and vegetables, and is the second most predominant acid in citrus fruits, many berries, and figs. Unlike the natural acid, which is levorotatory, the commercial product is a racemic mixture of D- and L-isomers. It is manufactured during catalytic hydration of maleic and fumaric acids, and is recovered from the equilibrium product mixture.

The acid has been used in carbonated beverages, powdered juice drinks, jams, jellies, canned fruits and vegetables, and confectionery. Its lingering profile enhances fruit flavors such as strawberry and cherry. In aspartame-sweetened beverages, malic acid acts synergistically with aspartame so that the combined use of malic and citric acids permits a 10% reduction in the level of aspartame. In frozen pizza, malic acid is used to lower the pH of the tomato paste without chelating the calcium in the cheese, as would citric and fumaric acids. This application improves the texture of the frozen pizza.

Phosphoric Acid

The second most widely used acidulant in food, phosphoric acid, is the only inorganic acid to be used extensively for food purposes. It produces the lowest pH of all food acidulants. Phosphoric acid is produced from elemental phosphorus recovered from phosphate rock.

The primary use of the acid is in cola, root beer, and other similar-flavored carbonated beverages. The acid and its salts are also used during production of natural cheese for adjustment of pH; phosphates chelate the calcium required by bacteriophages, which can destroy bacteria responsible for ripening. As chemical leavening agents, phosphates release gas upon neutralizing alkaline sodium bicarbonate; this creates a porous, cellular structure in baked products. The main reason for incorporating phosphates into cured meats such as hams and corned beef is to increase retention of natural juices; the salts are dissolved in the brine and incorporated into the meat by injection of brine, massaging, or tumbling. When used in jams and jellies, phosphoric acid acts as a buffering agent to ensure a strong gel strength; it also prevents dulling of the gel color by sequestering prooxidative metal ions.

Tartaric Acid

Tartaric acid is the most water-soluble of the solid acidulants. It contributes a strong tart taste which enhances fruit flavors, particularly grape and lime. This dibasic acid is produced from potassium acid tartrate which has been recovered from various byproducts of the wine industry, including press cakes from fermented and partially fermented grape

juice, less (the dried, slimy sediments in wine fermentation vats), and argols (the crystalline crusts formed in vats during the second fermentation step of wine-making). The major European wine-producing countries, Spain, Germany, Italy, and France, use more of the acid than the USA.

Tartaric acid is often used as an acidulant in grape- and lime-flavored beverages, gelatin desserts, jams, jellies, and hard sour confectionery. The acidic monopotassium salt, more commonly known as 'cream of tartar,' is used in baking powders and leavening systems. Because it has limited solubility at lower temperatures, cream of tartar does not react with bicarbonate until the baking temperatures are reached; this ensures maximum development of volume in the finished product.

See also: **Acids**: Properties and Determination; **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; **Canning**: Principles; **Citrus Fruits**: Composition and Characterization; **Colloids and Emulsions**; **Curing**; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Leavening Agents**; **Oxidation of Food Components**; **Phenolic Compounds**; **Preservation of Food**; **Sensory Evaluation**: Taste; **Spoilage**: Bacterial Spoilage; **Vinegar**

Further Reading

Anon. (1995) Spotlight on ingredients for confectionery and ice cream: Pointing and Favex point the way. *Confectionery Production* May: 350–351.

Anon. (1995–1996) Citric acid is no lemon. *Food Review* Dec./Jan.: 51–52.

Arnold MHM (1975) *Acidulants for Foods and Beverages*. London: Food Trade Press.

Bigelis R and Tsai SP (1995) Microorganisms for organic acid production. In: Hui YH and Khachatourians GG (eds) *Food Biotechnology: Microorganisms*, pp. 239–280. New York: Wiley-VCH.

Bouchard EF and Merritt EG (1979) Citric acid. In: Grayson M (ed.) *Kirk–Othmer Encyclopedia of Chemical Technology*, 3rd edn, vol. 6, p. 150. New York: Wiley.

Brennan M, Port GL and Gormley R (2000) Post-harvest treatment with citric acid or hydrogen peroxide to extend the shelf life of fresh sliced mushrooms. *Lebensmittel-Wissenschaft & Technologie* 33: 285–289.

Dziezak JD (1990) Acidulants: ingredients that do more than meet the acid test. *Food Technology* 44(1): 76–83.

Farkye NY, Prasad B, Rossi R and Noyes QR (1995) Sensory and textural properties of Queso Blanco-type cheese influenced by acid type. *Journal of Dairy Science* 78: 1649–1656.

Fowlds R and Walter R (1998) *The Production of a Food Acid Mixture Containing Fumaric Acid*, PCT Patent application WO 98/53705.

Gardner WH (1972) Acidulants in food processing. In: Furia TE (ed.) *CRC Handbook of Food Additives*, 2nd edn, vol. 1, p. 225. Cleveland, OH: CRC Press.

Garrote GL, Abraham AG and DeAntoni GL (2000) Inhibitory power of kefir: the role of organic acids. *Journal of Food Protection* 63(3): 364–369.

Goldberg I, Peleg Y and Rokem IS (1991) Citric, fumaric, and malic acids. In: Goldberg I and Williams R (eds) *Biotechnology and Food Ingredients*, pp. 349–374. New York: Van Nostrand Reinhold.

Hartwig P and McDaniel MR (1995) Flavor characteristics of lactic, malic, citric, and acetic acids at various pH levels. *Journal of Food Science* 60(2): 384–388.

International Commission of Microbiological Specifications for Foods (1980) *Microbial Ecology of Foods*, vol. 1. New York: Academic Press.

Kummel KIF (2000) Acidulants use in sour confections. *The Manufacturing Confectioner* Dec.: 91–93.

Miller Al and Call JE (1994) Inhibitory potential of four-carbon dicarboxylic acids on *Clostridium botulinum* spores in an uncured turkey product. *Journal of Food Protection* 57(8): 679–683.

Oman YJ (1992) *Process for Removing the Bitterness from Potassium Chloride*, US Patent No. 5,173,323.

Phillips CA (1999) The effect of citric acid, lactic acid, sodium citrate and sodium lactate, alone and in combination with nisin, on the growth of *Arcobacter butzleni*. *Letters in Applied Microbiology* 29: 424–428.

Sun Y and Oliver JD (1994) Antimicrobial action of some GRAS compounds against *Vibrio vulnificus*. *Food Additives and Contaminants* 11(5): 549–558.

Suye S, Yoshihana N and Shusei I (1992) Spectrophotometric determination of L-malic acid with a malic enzyme. *Bioscience, Biotechnology, and Biochemistry* 56(9): 1488–1489.

Synosky S, Orfan SP and Foster JW (1992) *Stabilized Chewing Gum Containing Acidified Humectant*. US Patent No. 5,175,009.

Vidal S and Saleeb FZ (1992) *Calcium Citrate Anticaking Agent*. US Patent No. 5,149,552.

ADAPTATION – NUTRITIONAL ASPECTS

P S Shetty, Food and Agriculture Organization, Rome, Italy
J C Waterlow, London School of Hygiene & Tropical Medicine, London, UK

Introduction

The word 'adaptation' is used in many different contexts: biological or Darwinian; physiological or metabolic; behavioral or social. In nutrition, we are concerned with the last two. The difference between 'adaptation' and 'homeostasis' is that the latter represents the maintenance of a set point for some physiological characteristic such as body temperature or pH – this is Claude Bernard's '*fixité du milieu intérieur*.' Adaptation involves a change in the set point, for example, the increase in hemoglobin concentration found in people living at high altitude or the decrease in sodium concentration in the sweat in people exposed to high environmental temperatures. Such adaptations take time; one speaks of people 'becoming adapted,' whereas homeostasis is a rapid and continuous process. For adaptation to be more than just a response, it must represent a new steady state, capable of being maintained, and we think of it as beneficial to the organism, preserving, within limits, normal function. It is here that the real difficulty arises. For most bodily characteristics or functions, there are no clear definitions of a 'normal' range, within which physiological adaptations can operate. Basal metabolic rate (BMR) is an exception, but for most functions that are important for the quality of life, such as work capacity or resistance to infection, there are no such defined limits, so it is difficult to decide whether an adaptation is 'successful.' We shall return to this point later.

In nutrition, it is convenient to look separately at adaptation to inadequate intakes of energy and protein before going on to the more realistic situation of overall deficiency of food and deficiency or excess of micronutrients.

Adaptation to Low Energy Intakes

The human body responds to an inadequate intake of food energy by a whole series of physiological and behavioral responses. Experimental studies of semistarvation in normal adults have helped in understanding the physiological changes that characterize this adaptive response to a lowered energy intake in humans. The metabolic responses that occur during acute energy restriction and the physiological mechanisms that are involved may, however, be different from the changes observed in individuals who are chronically undernourished as a result of long-standing marginal energy intakes.

In previously well nourished adults, a reduction in BMR is a constant finding during experimentally or therapeutically induced energy restriction. This finding has been explained on the basis of a loss of active tissue mass, as a result of the loss of body weight, together with a decrease in the metabolic activity of the active tissues. The latter would indicate a greater efficiency or metabolic adaptation, on the assumption that the same amount of work is being done at lower cost. Recalculating the data from the two separate semistarvation studies, one short term and the other longer term, it has been shown that the early fall in BMR seen during energy restriction is mainly accounted for by enhanced metabolic efficiency (Table 1). This reduction in BMR per kilogram of active body tissues seen in the first 2 weeks of energy restriction remained essentially unchanged over the subsequent period of semi-starvation. The greater contribution to the fall in BMR during prolonged energy restriction, however, was the result of a slow decrease in the total mass of active tissues. It seems reasonable, therefore, to suggest that the reduction in BMR during energy restriction occurs in two different phases. In the initial phase, there is a marked decrease in the BMR, which is not attributable to the changes in body weight or body composition. This decrease in BMR per unit active tissue is a measure of increase in

Table 1 Changes in body weight, active tissue mass (ATM), and basal metabolic rate (BMR) following short- and long-term semistarvation in humans

	Semistarvation			
	Short-term[a]		*Long-term*[b]	
	Baseline	*Day 14*	*Baseline*	*Day 168*
Body weight (kg)	71.6	65.4	67.5	51.7
ATM (kg)	44.9	42.2	38.8	28.7
BMR: MJ day^{-1}	7.3	5.7	6.6	4.2
kcal day^{-1}	1742	1370	1575	1004
BMR decrease:				
kJ day^{-1}		21.4%		36.3%
kJ kg^{-1} ATM		16.3%		13.8%

[a]Data from experiments conducted on 12 human subjects by Grande *et al*. (1958).
[b]Data from experiments conducted on 12 human subjects by Keys *et al*. (1950).

'metabolic efficiency' in well-nourished individuals who are energy-restricted and is often cited as evidence of 'metabolic adaptation.' With continued energy restriction, the lowered level of cellular metabolic rate remains nearly constant, and any further decrease in BMR is accounted for by the loss of body weight. Thus, the longer the duration of energy restriction, the more important the contribution of decreased body tissues becomes to the reduction in BMR. This reduction in lean body tissue with prolonged energy restriction is considered to be a passive process and a consequence of body tissues being used as substrates and metabolic fuel to compensate for the lack of food energy.

The biochemical and physiological mechanisms involved in reducing the cellular metabolic rate are poorly understood. It has recently been estimated that ~90% of BMR is contributed by mitochondrial oxygen consumption, of which only ~20% is uncoupled by mitochondrial proton leak and the rest coupled to ATP synthesis. It is not known how much changes in mitochondrial function contribute to the increasing efficiency of tissue metabolism. Several physiological changes in hormonal and substrate function may operate to influence the changes in metabolic efficiency seen during the early part of energy restriction. Several hormones are now known to be sensitive to changes in the levels of energy intake, dietary composition, and energy balance status of the individual. Changes in sympathetic nervous system (SNS) activity and catecholamines, alterations in thyroid hormone metabolism, and changes in insulin and glucagon play an important role in this response. The reduction in SNS activity and catecholaminergic drive that we observed was counter to traditional views on the control of substrate mobilization during starvation. Traditionally, the increase in lipolysis, maintenance of glucose homeostasis, and increase in glucagon output on fasting have been considered as being the result of an enhanced sympathetic drive during energy restriction. It now appears that the lipolytic activity associated with energy restriction appears to be under the dominant control of declining plasma insulin levels. Insulin is the primary hormonal signal that allows for an orderly transition from the fed to the fasted state without the development of hypoglycemia. While the SNS activity is toned down, signaled by the decrease in energy flux, the energy deficit lowers insulin secretion and initiates changes in peripheral thyroid metabolism. The reduction in the activities of these three thermogenic hormones acts in a concerted manner to lower cellular metabolic rate. Changes in other hormones such as glucagon, growth hormone, and glucocorticoids may also participate and, in association with insulin deficiency, help promote endogenous substrate mobilization leading to an increase in circulating free fatty acids (FFA) and ketone bodies. Contribution may also be made by the reduction in Na^+–K^+ pumping across the cell membrane and futile substrate cycling, although how much they contribute to the reduced energy output is not known. The elevated FFA levels, alterations in substrate recyling, and protein catabolism will also influence the resting energy expenditure. These changes are thus not only aimed at lowering the metabolic activity of the active cell mass but also essential for the orderly mobilization of endogenous substrates and fuels during a period of restricted availability of exogenous calories. These hormonal and metabolic changes aid the survival of the organism and may be considered as being 'adaptive' in nature.

Adaptation to lowered energy intake in chronically undernourished adults on subsistence food intakes in the developing world appears, however, to be different. Ferro-Luzzi summarized the adaptive responses in individuals who were maintaining energy balance in spite of life-long exposure to low energy intakes – the state of so-called 'chronic energy deficiency.' Adaptation was represented as a series of complex integrations of several different processes that occurred during energy deficiency and resulted in a new level of equilibrium being achieved at a lower level of energy intake. People who have gone through the adaptive process may be expected to exhibit more or less permanent sequelae (or costs of adaptation), which include smaller stature and body size, altered body composition and a lower BMR, with the likelihood of enhanced metabolic efficiency of energy handling. However, this has been difficult to prove, largely because marked changes in the body composition (in particular in the fat and lean compartments) make interpretation of changes in the metabolic rate per unit of active tissue mass highly unreliable as indicators of metabolic efficiency. Changes in body composition as well as in body size and dimensions may play a dominant role in adaptation to long-term inadequacy of energy intake from childhood, however undesirable they may be. These physiological adaptations are not beneficial changes, as they influence employability and economic productivity, although they may help in furthering survival of the individual.

Adaptations to a reduction in food energy intake may also be manifest as physiological and behavioral changes in physical activity, aimed at reducing the energy expended by the individual every day to make up for the energy deficit. Reductions in either intensity or duration of physical activity can save much energy and hence may be a crucial response to

energy restriction and an important feature of the adaptive response. Studies on semistarvation of previously well nourished adults showed a marked impairment in both intensity and duration of activity. About 40% of the reduction was attributable to a decrease in actual costs of performing tasks, whereas 60% of the reduction was due to a decrease in tasks undertaken. In previously well-nourished semistarved adults, behavioral reduction in voluntary activity seems to be quantitatively more important. Analysis of the pattern of an individual's physical activity during a voluntary reduction in food intake shows that the behavioral responses were associated with a distinct change in activity pattern.

Physiological changes in the physical work capacity of undernourished young men are also difficult to demonstrate, and the overwhelming evidence seems to support the view that differences, if any, are largely due to changes in body composition and not to adaptive differences in cell function. Spurr summarized the results of several of his studies in Colombia, which demonstrated that maximal oxygen uptake ($Vo_{2\ max}$) was lower in malnourished young adults; the degree of reduction being related to the progressive severity of undernutrition. He was also able to demonstrate that 80% of the reduction in $Vo_{2\ max}$ in moderate and severe categories of undernutrition was accounted for by differences in muscle cell mass. Assessment of endurance at 70–80% of the $Vo_{2\ max}$ in the undernourished also failed to demonstrate any differences in the maximum endurance time. However, assessment of productivity in agricultural environments shows that work productivity is affected indirectly by nutritional status, through its influence on stature, body weight, body composition, and $Vo_{2\ max}$.

Chronically undernourished adults are likely to demonstrate increased ergonomic or 'real life' efficiency. By this is meant a reduction in the effort needed to do any piece of physical work. It is reasonable to suppose that tradition and experience have enabled people living on marginal intakes and hence likely to be chronically undernourished to find the most economical methods of doing the tasks they have to do. This manifestation of increased efficiency might be regarded as a training effect, quite distinct from the behavioral adaptation that accompanies undernutrition, which is mainly related to how individuals allocate time and energy to different productive and leisure activities, with inevitable biological and economic consequences. In undernutrition, more time is given to work activities, while leisure and home production activities are reduced; this is an important form of behavioral adaptation. Marginally undernourished individuals tend to become more sedentary at the expense of decreased social interactions and discretional noneconomic activities. Latham showed that when energy-deficient individuals are forced over a period of time to limit their activities, they forego activities to conserve energy, some of which they do consciously and wilfully, some they do unconsciously. Thus, restricting physical activity or performing it more efficiently is an important coping strategy for undernourished individuals and may form part of the behavioral adaptive response to a lowered intake of food energy.

Adaptation to Low Protein Intakes

Most of our knowledge on this subject has been derived from experimental studies on man. Adaptation to low protein intakes has two proximate functions: to secure nitrogen balance and to maintain lean body mass (LBM). As regarding balance, there is an obligatory loss of nitrogen from the body which has been estimated in male Caucasian adults to amount to about 55–65 mg of nitrogen per kilogram per day and which has to be balanced by the intake. There is little evidence that this loss is lower in people long accustomed to low protein intakes, or to an intake mainly from vegetable sources, so there does not seem to be much opportunity for adaptation at this point. There is, however, evidence, that on lower protein intakes or in children recovering from malnutrition, the efficiency of utilization of food protein may be increased above the usual level of about 70%. This effect may be regarded as a response to depletion, i.e., loss of body nitrogen, but is none the less an adaptive response aimed at conserving body nitrogen.

When a person moves from a normal intake, providing say 1.5 g of protein (250 mg of nitrogen) per kilogram per day to an intake close to the obligatory loss, the nitrogen output falls to a new low level in 7–10 days in the human adult, 1–2 days in the infant and about 30 h in the rat. This is the first stage of adaptation. During this stage, there is a small loss, amounting to 1–2% of body N, which probably has no physiological significance.

The main variable in this adaptation is the urinary excretion of urea. Urea production, which is a measure of amino acid oxidation, is related to nitrogen intake, although at the present time, there is some controversy about the strength of the relationship. Only part of the urea produced is excreted in the urine; the remainder passes into the colon, where it is hydrolyzed by gut bacteria to ammonia. A relatively small part of this ammonia is recycled to urea. The rest of it enters the amino acid pool, and there is increasing evidence that microbes in the gut are

capable of using it to synthesize indispensable as well as dispensable amino acids. In the normal individual, on an adequate nitrogen intake and in a steady state, these reactions are essentially exchanges, and there is no net gain of nitrogen. However, with a deficient intake or an increased demand for growth, amino acids derived from the colonic hydrolysis of urea can make a significant contribution to the body's nitrogen economy. Hence, the term 'urea salvage,' introduced by Jackson is appropriate, salvage representing an important component of adaptation. Since the proportion of urea hydrolyzed to that excreted increases on a low protein intake, it follows that the maintenance of nitrogen balance involves control of the rate of hydrolysis. It is thought that this control may be exerted by a urea transporter, which is sensitive to the protein level of the diet.

A second phase of adaptation comes into play if the protein intake is inadequate to cover the obligatory losses, so that there is a prolonged negative nitrogen balance. This inevitably leads to a loss of body protein. Since the magnitude of the obligatory loss is determined by the body protein mass, as this mass decreases, the loss will decrease until eventually the nitrogen balance is restored. This would represent an adaptation at the expense of a certain loss of lean body mass. Whether that loss is important will be discussed below. An example of such an adaptation is provided by the poor Indian laborers, studied by Shetty's group in Bangalore, whose lean body mass was substantially less (13%) than that of taller controls with the same body mass index (BMI). An important finding was that in these men, the main deficit was of muscle rather than of visceral mass. Presumably, this adaptation has its cost in terms of reduced muscular capacity, but it seems justifiable to regard it as a successful adaptation, since these men could live reasonable lives.

The metabolism of plasma albumin provides an interesting example of adaptation to low protein intake. In children with protein-energy malnutrition, one of the most constant findings is a reduction in plasma albumin concentration. This is accompanied by a fall in the rate of albumin catabolism, as if in an effort to maintain the concentration in plasma. The same effect has been shown in adults on experimental low protein intakes; the relative change in the rate of albumin breakdown was much greater than the change in albumin concentration. Thus, the breakdown rate would provide a much more sensitive measure of the state of protein nutrition than the albumin concentration; unfortunately, it is not a measurement that is practical on a large scale.

In real life, it is in famines, refugee camps, or concentration camps that we are faced with the question: what are the limits of adaptation to a food supply that is inadequate in both energy and protein – in other words, to semistarvation? Nowadays, the response is generally measured by the level of the body mass index (BMI = weight (kg)/height^2 (m)). Factors that affect the response of the BMI are the degree of deficiency, its duration, and the relative deficiencies of energy and protein. In total starvation, of which, as already mentioned, there have been a number of experimental studies, no steady state can be achieved, and no adaptation is possible. In the famous Minnesota semistarvation experiment, subjects were fed half their normal intakes of energy and protein; after 24 weeks, their BMI had fallen to about 16 from an initial level of about 22, and they showed severe functional and psychological impairment. This was in marked contrast to the Indian laborers referred to above who had a similarly low BMI. It seems that by life-long exposure to presumably inadequate food intakes they had adapted to a steady state of what would be currently described as 'chronic energy deficiency,' yet, their vital functions of energy and protein turnover were well maintained.

Some cases of semistarvation present with edema, which is quite commonly seen in famines and in refugee camps. Although the cause of the edema is controversial, it is a reasonable hypothesis that it results from a particular deficiency of protein in relation to energy, although there may be other deficiencies as well. In one study in a refugee camp, subjects with edema had a higher BMI, as might be expected from the accumulation of fluid, than those without edema, but they also had a substantially higher mortality rate. Women adapted better than men; this is apparent in several accounts. It appears, therefore, that when protein is particularly deficient, the capacity for adaptation is reduced.

From a physiological point of view, if the requirement for successful adaptation is the maintenance of LBM within 'normal' limits, it becomes crucial to define those limits. There are many difficulties. The BMI is a crude estimate of LBM, since it does not separate fat from lean tissue. However, the fat content of the body has a bearing on the capacity for adaptation, since it has been shown, not surprisingly, that in starvation, the loss of LBM is inversely related to the size of the initial fat stores. A low BMI with loss of muscle mass would explain the association mentioned above with decreased maximal oxygen consumption and reduced work capacity. However, it does not explain other associations that have been found, such as reduced resistance to infections and low birth weight of infants. Interestingly, there is no effect on breast-milk output, suggesting that this function, basic for the survival of the race, is well

protected. What, then, are the normal limits? Is there a threshold or cutoff point of LBM, as assessed by BMI, above which function is normal and below which it falls off? Some evidence from epidemiological studies suggest that there is no threshold, but a steady fall-off with falling BMI. However, because BMI is influenced by many factors beyond physiological homeostasis, it is difficult to establish with certainty the limits within which adaptation may be regarded as successful.

Adaptation to Variations in Micronutrient (Mineral and Vitamin) Intakes

One of the major processes by which adaptation to changes in nutrient intakes occurs, particularly that of micronutrients, is by changes in gastrointestinal function. The gastrointestinal tract has extensive potential for adaptation. For instance, following intestinal resection, the residual intestine is capable of a considerable increase in size and absorptive capacity. This is achieved by dilatation and an increase in rugosity and by hypertrophy of the villi and microvilli. This increases the available surface area of contact with the nutrients and thus increases the absorptive capacity. The enzyme activities and the turnover of cells are also increased. The ileal part of the intestines adapts better than the jejunum. Changes in the function of the intestines, such as slowing down the transit, also helps the process of adaptation by increasing absorptive capacity. These adaptive changes are maximized by the mucosal exposure to nutrients and by the role played by several key hormones. Intestinal adaptation is, however, limited by inadequate blood supply or poor nutritional status.

Calcium represents the best example of a micronutrient whose absorption by the gastrointestinal tract is modulated to demonstrate adaptation. The physiological need for calcium changes throughout the lifecycle, i.e., growth, puberty, pregnancy, lactation, and menopause. Calcium intakes are also highly variable world-wide, with a more than fourfold difference between the lowest intake and the highest. Hence, the absorption of calcium from the diet must be adaptable and responsive to both dietary and physiological circumstances. This process of adaptation and physiological plasticity is largely orchestrated by vitamin D, which stimulates intestinal calcium absorption by both genomic and nongenomic mechanisms. The renal output of dihydroxy vitamin D_3, which is regulated, reflects the perceived needs of the organism for calcium, which in turn influences the tightly regulated process of intestinal calcium absorption. The latter regulation occurs both by genomic receptor mediated action (i.e., through calbindin) and by nongenomic mechanisms (through transcaltachia). There are other social and behavioral adaptations, too, which influence the individuals' choice of diet and determine what is available for intestinal absorption. It is hence believed that vitamin D-mediated calcium absorption by the intestines satisfies the requirement for it to be considered as an adaptive function.

One would expect that the requirements of most micronutrients are amenable to adaptation when intakes are lowered, although the evidence for such changes is not readily available.

See also: **Calcium**: Properties and Determination; Physiology; **Energy**: Intake and Energy Requirements; Energy Expenditure and Energy Balance; **Famine, Starvation, and Fasting**; **Protein**: Digestion and Absorption of Protein and Nitrogen Balance

Further Reading

Benedict FG, Miles WR, Roth P and Smith HM (1919) *Human Vitality and Efficiency Under Prolonged Restricted Diet*. Publication No. 280. Washington, DC: Carnegie Institute of Washington.

Blaxter KL and Waterlow JC (eds) (1985) *Nutritional Adaptation in Man*. London: John Libbey.

Ferro-luzzi A (1985) Range of variation in energy expenditure and scope of regulation: In: *Proceedings of XIIIth International Congress of Nutrition*, pp. 393–399. London: Libbey.

Jackson AA (1968) Salvage of urea nitrogen in the large bowel: functional significance in metabolic control and adaptation. *Biochemical Society Transactions* 26: 231–236.

James WPT and Ralph A (eds) (1994) Functional significance of low body mass index. *European Journal of Clinical Nutrition* 48 (supplement 3).

James WPT and Shetty PS (1982) Metabolic adaptation and energy requirements in developing countries. *Human Nutrition: Clinical Nutrition* 36: 331–336.

Keys A, Brozeck J, Henschel A, Mickelson O and Taylor HL (1950) In: *The Biology of Human Starvation*. Minneapolis, MN: University of Minneapolis Press.

Latham MC (1989) Nutrition and work performance, energy intakes and human wellbeing in Africa. In: *Proceedings of XIVth International Congress of Nutrition*. London: Libbey.

Norman AW (1990) Intestinal calcium absorption: a vitamin D-hormone-mediated adaptive response. *American Journal of Clinical Nutrition* 51: 290–200.

Shetty PS (1990) Physiological mechanisms in the adaptive response of metabolic rates to energy restriction. *Nutrition Research Reviews* 3: 49–74.

Shetty PS (1993) Chronic undernutrition and metabolic adaptation. *Proceedings of the Nutrition Society* 52: 267–284.

Spurr GB (1993) Nutritional status and physical activity work capacity. *Yearbook of Physical Anthropology* 26: 1–35.

Spurr GB (1987) The effects of chronic energy deficiency on stature, work capacity and productivity. In: Schurch B and Scrimshaw NS (eds) *Chronic Energy Deficiency: Causes and Consequences*, pp. 95–134. Lausanne, Switzerlnd: IDECG.

Waterlow JC (1990) Nutritional adaptation in man: General information and concepts. *American Journal of Clinical Nutrition* 51: 259–263.

ADIPOSE TISSUE

Contents

Structure and Function of White Adipose Tissue

R G Vernon and D J Flint, Hannah Research Institute, Ayr, UK

Distribution and Structure of Adipose Tissue

White adipose tissue is quantitatively the most variable component of the body, ranging from a few percent of body weight to over 50% in obese animals and people. In mammals, adipose tissue is found within the abdominal cavity, under the skin, within the musculature where it is found between muscles (intermuscular) and within muscles (intramuscular) (e.g., marbling of meat) and in a few highly specialized locations such as the eye socket. Within these locations, the tissue occurs in discrete depots (e.g., perirenal, epididymal, omental, popliteal); there are about 16 in most species. Comparative studies have revealed that the distribution of adipose tissue depots evolved early in mammalian evolution and has been retained in most species. In some species (e.g., pigs, whales) subcutaneous depots have become enlarged and have fused to form a continuous layer; this also occurs in obese individuals. Adipose tissue depots are also found in birds, reptiles, and amphibians.

White adipose tissue is a soft tissue, devoid of rigidity, and is well supplied with capillaries and nerve endings from the sympathetic nervous system. In mature animals, adipocytes (fat cells) comprise about 90% of the mass of the tissue but only 25% or less of the total cell population. The 75% or so nonadipocytes are often termed the stromal–vascular fraction and comprise mainly endothelial cells of blood vessels and adipocyte precursor cells. Adipocytes vary enormously in size from several picolitres to about 3 nl in volume, depending on the amount of lipid present. The mature fat cell is essentially a lipid droplet surrounded by a film of cytoplasm (containing mitochondria, endoplasmic reticulum, etc.) and bounded by a plasma membrane; the nucleus is pushed to the periphery and appears as a blip on the surface of the cell. Within a depot, there will be fat cells of various sizes so that it is usual to refer to the 'mean fat cell volume' of a tissue; this varies amongst adipose tissue depots in an individual. The adipocyte mean cell volume also varies with size of the animal, larger animals having larger fat cells; this occurs both within and between species.

Functions of Adipose Tissue

The major function of white adipose tissue is the storage of energy as triacylglycerol (fat, lipid). Fat is a highly efficient form of energy storage, not only because of its high energy content per unit weight, but also because it is hydrophobic. Hence, 1 g of adipose tissue may contain about 800 mg of triacylglycerol and about 100 mg of water. In contrast, glycogen not only has a lower energy content per unit weight than fat, but also is much more hydrated. The development of copious stores of fat was probably very important for the evolution of homeothermy in mammals and birds. Homeotherms have a much higher basal metabolic rate and so need a more substantial energy reserve than poikilotherms (reptiles, fish, and amphibians). The ability to accrue copious amounts of adipose tissue was also essential for exploitation of habitats where food supply is

scarce (e.g., deserts) or seasonal (e.g., arctic). Northern species such as polar bears and reindeer build up substantial depots of fat during the summer to provide reserves of nutrients during the winter. Such species thus have substantial seasonal fluctuations in the amount of adipose tissue in their bodies. Additional reserves of adipose tissue are also accumulated during pregnancy in most species to help support the development of the fetus during the later stages of pregnancy and to facilitate milk production. The use of adipose tissue lipid is very important during early lactation in dairy cows, for example, in which appetite increases more slowly than milk production at the beginning of lactation. It is also important for milk production in some species of bears and seals that fast during lactation.

It is now apparent that adipose tissues are not solely a store of fat. Subcutaneous adipose tissue will act as insulation; adipose depots in the eye socket may have a protective function. More importantly perhaps, adipose tissue produces a number of biologically active substances, e.g., prostaglandins, insulin-like growth factor 1 and binding proteins, adipsin, cytokines (e.g., tumor necrosis factor α), estrogens (primarily estrone), and leptin. Some of these substances are probably important for adipose tissue function and development, but some have other roles. Adipose tissue is the major source of estrogens in postmenopausal women. The mammary gland grows in a bed of adipose tissue and is thought to require factors secreted by adipose tissue for its development. Lymph nodes are located in adipose tissue depots and in some species (e.g., guinea-pigs), at least, there is an interaction between adipocytes and lymphoid cells. Adipose tissue may have another role in defense systems of the body as it secretes adipsin and several other proteins involved in an alternative pathway of complement production. Another important protein produced by adipocytes is the cytokine tumor necrosis factor-α; production of this factor is normally low, but it is markedly increased during obesity, when it appears to play a major role in the development of insulin resistance in the tissue, and hence noninsulin-dependent diabetes.

Perhaps the most important and interesting protein secreted by adipocytes is leptin, which has a key role in appetite control and energy balance (**Figure 1**). Leptin was discovered only recently through studies on the basis of a genetically obese strain of mice (ob/ob mice); these mice produce a nonfunctional form of leptin. Leptin is released into the blood and travels to the brain, where there are leptin receptors in discrete areas involved in appetite control. Low levels of leptin in the blood increase appetite, whereas administration of high doses inhibit appetite. Leptin not only modulates appetite, but also increases energy expenditure, stimulating thermogenesis in brown adipose tissue, suggesting a key role in the control of energy balance in the body. Leptin synthesis is regulated by insulin, glucocorticoids, and catecholamines, but most interestingly, the concentration of leptin in the blood in the fed state is proportional to the amount of fat in the body; this led to the idea that leptin acts as a 'lipostat,' matching appetite to adiposity. However, the leptin concentration in the blood is decreased by fasting, and leptin is involved in the changes in secretion of several pituitary hormones during fasting. Thus, it has been suggested that the major

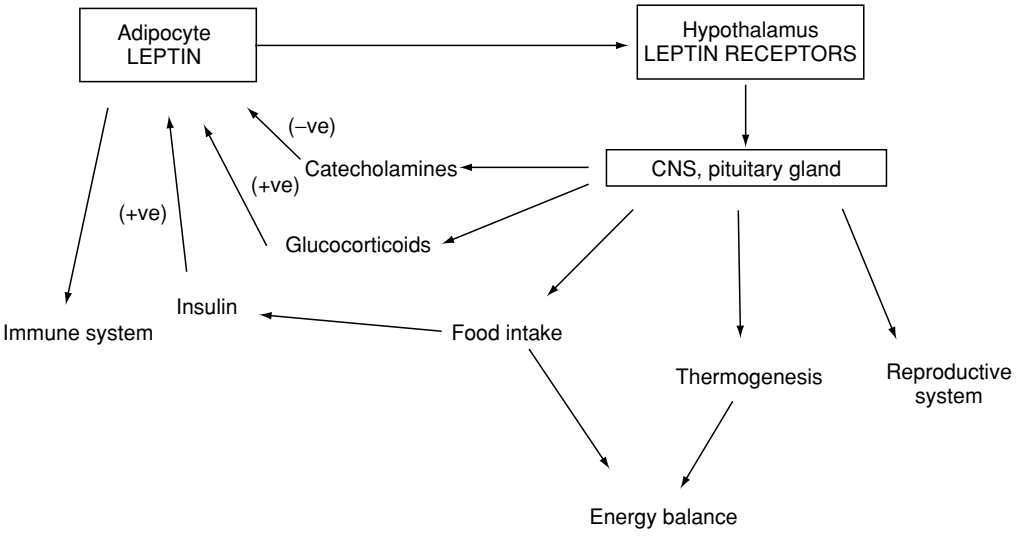

Figure 1 Leptin production and function. CNS, central nervous system.

role of leptin may be in adaptation to fasting and acting as a signal of too little rather than too much adipose tissue. Leptin appears to be required for normal functioning of the immune system and also for reproductive function. Indeed, a lack of leptin may well be the main reason for the failure of the menstrual cycle in anorexics and very lean athletes. This makes good physiological sense as it insures that females do not become pregnant, unless they have adequate reserves of adipose tissue lipid.

Adipose tissue thus has a variety of functions, in addition to being an energy store. While the accumulation of adipose tissue lipid reserves provides a buffer against starvation, and some degree of adiposity is important for the various other functions of the adipose tissue described above, there is a cost in that additional body mass decreases speed and agility and so increases the chance of succumbing to predation. Thus, in most wild animals for which food is generally plentiful, there are usually only small amounts of adipose tissue (predation rather than starvation being the greatest threat to mortality). In such species, it seems likely that the leptin system, and probably other systems, will be acutely tuned to maintain the minimal amounts of adipose tissue needed. In general, it is only species living in environments where the availability of food is erratic or seasonal that accumulate large amounts of adipose tissue since, for these species, starvation is a greater threat than predation. In such species, the leptin system must be modulated to allow the accumulation of adipose tissue lipid. It would also appear that the leptin system can be readily subverted in humans and also domestic pets for excess adiposity is becoming a major problem.

In addition to white adipose tissue, there is also another form, brown adipose tissue, which differs morphologically and biochemically, and has an important role in thermogenesis.

Development of Adipose Tissue

Adipose tissue develops both by accretion of lipid in adipocytes and by increases in the number of adipocytes. Mature adipocytes are thought to be unable to divide; rather, they are produced from a pool of precursor cells within the tissue. The sequence of events in the formation of mature adipocytes (Figure 2) is still partly speculative, and much has been gleaned from studies of certain cell lines (e.g., ob17 and 3T3 L1 cells), which will differentiate and develop into adipocytes in cell culture. Current thinking envisages a pluripotent stem cell that can give rise to muscle and bone cells as well as adipocytes. Once committed to adipocyte formation, this cell is termed an adipoblast.

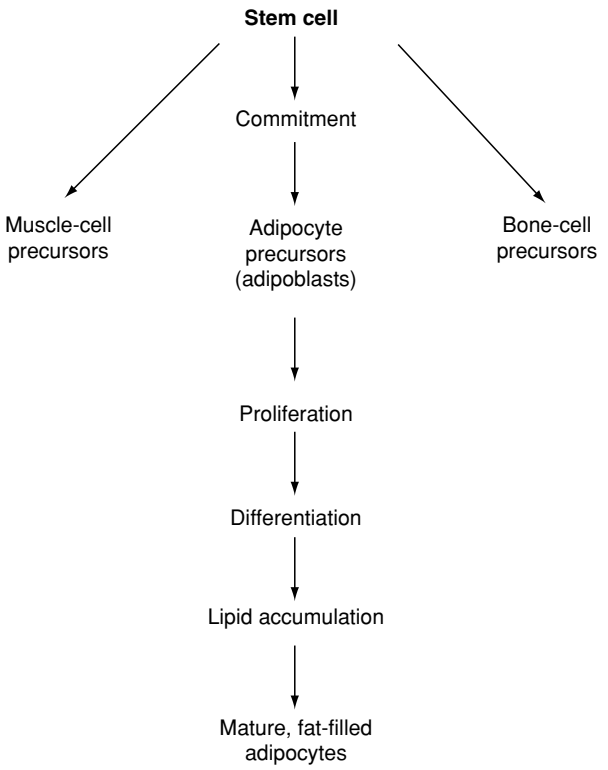

Figure 2 Adipocyte development.

This is envisaged (it has not been isolated) as an undifferentiated cell, devoid of lipid droplets but able to proliferate. At some point, these cells begin to differentiate, acquiring, in stages, the enzymes and other proteins characteristic of adipocytes. Once differentiated, these cells can begin to accumulate lipid, which appears at first as a series of small droplets within the cell. As these become larger, they fuse to form the single lipid droplet characteristic of mature adipocytes. Both differentiating cells and cells with several small lipid droplets (multilocular phase) are often referred to as preadipocytes, the term adipocyte usually being used to describe cells with a single lipid droplet. Multilocular adipocytes are very similar in appearance to mature brown adipocytes, and it was once thought that the brown adipocyte was a stage in the development of the white adipocytes. It is now recognized that this view is incorrect, except possibly for a few special cases (e.g., the perirenal adipose tissue depot of newborn lambs).

Adipocytes begin to appear in the fetus about half way through gestation, developing in small clumps around blood vessels. Within a depot, both the number and size of adipocytes increase in phases (Figure 3). In addition, it is now clear that development is not synchronized in all depots; abdominal depots in general develop earlier than those

associated with the musculature. In most species, the fetal stage is a period of active proliferation but little hypertrophy, so that cells are small at birth (about 10 pl in volume). The suckling period usually results in rapid hypertrophy and hyperplasia; this is followed by a more quiescent period when muscle growth predominates. When the rate of muscle growth begins to slacken, nutrients are diverted into adipose tissue, and the fattening phase begins. This phase is associated with marked hypertrophy, due to lipid deposition, in most depots and further hyperplasia, especially in the carcass depots. During the fattening phase, depot-specific differences in adipocyte size appear. Adipocytes do not increase in size indefinitely; once a maximum is reached (about 1–3 nl, depending on species), this seems to trigger the formation of new adipocytes from the precursor pool. The view prevalent in the 1970s that all hyperplasia occurred in young animals, including humans, is now thought to be invalid.

A great deal of research has gone into identifying the hormones and other factors that promote the proliferation and differentiation of adipocyte precursor cells. At present, the picture is far from clear, in part because of probable species differences and also because much of the work has involved the use of cell lines that do not all appear to have identical hormonal requirements for development. A variety of peptide growth factors (e.g., insulin-like growth factor 1, fibroblast growth factor, platelet-derived growth factor, epidermal growth factor) can stimulate preadipocyte proliferation, whereas insulin, thyroid hormones, and glucocorticoids appear to be important for differentiation of preadipocytes into adipocytes in a variety of species. Glucocorticoid hormones and also testosterone are thought to have important roles in site-specific development of adipose tissue. Derivatives of arachidonic acid (an essential fatty acid) such as 15-deoxy-$\Delta^{12,14}$-prostaglandin J_2 are also thought to have a major role in adipogenesis, acting via the recently discovered (and inappropriately named!) peroxisome proliferator-activated receptor-γ. Growth hormone has a complex role, stimulating insulin-like growth factor 1 production in adipose tissue and hence proliferation of preadipocytes and in addition may be required for the cells to become 'committed' to differentiation. In addition to positive effectors, tumor necrosis factor α and transforming growth factor β can inhibit differentiation. In contrast to hyperplasia, much more is known about the control of hypertrophy, for this is dependent on the metabolic rates of the pathways of lipid synthesis and degradation.

Deposition and Mobilization of Fat

The synthesis of triacylglycerol (esterification) requires a supply of fatty acids and glycerol 3-phosphate (Figure 4). The latter is mostly synthesized from glucose. Fatty acids, however, may be synthesized *de novo* within the cell or obtained from blood triacylglycerols. Fatty acids can be synthesized in adipocytes from a variety of precursors, including glucose, acetate, lactate, and some amino acids. Glucose is quantitatively the most important in man and some laboratory species (e.g., rats, mice), whereas acetate is most important in ruminants. Liver is also an important site of fatty acid synthesis in many mammals and is the major site of fatty acid synthesis in birds (avian adipocytes have essentially no capacity for fatty acid synthesis) and also in humans on a typical Western diet. Some of the fatty acids synthesized in the liver are incorporated into very-low-density lipoprotein

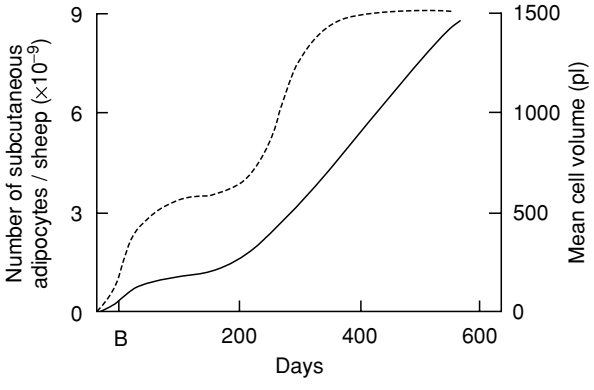

Figure 3 Developmental changes in adipocyte number (broken line) and mean cell volume (solid line) of sheep subcutaneous adipose tissue from 25 days before birth (B) until 600 days after birth.

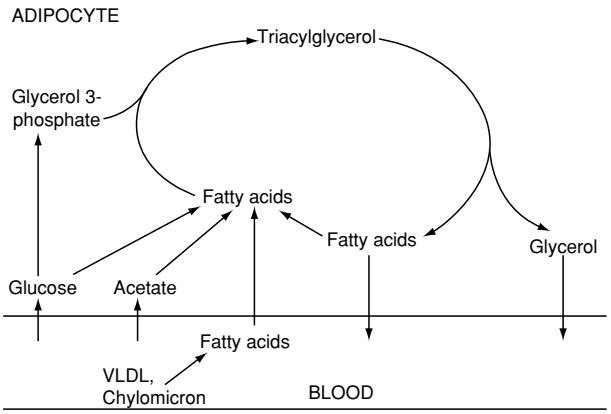

Figure 4 Pathways for synthesis and hydrolysis of triacylglycerol in adipocytes. VLDL, very-low-density lipoprotein.

(VLDL) triaclyglycerols for transport to adipocytes and other tissues. Dietary fatty acids are also incorporated into triacylglycerols in the intestinal cells and secreted as another form of lipoprotein, called chylomicrons. Triacylglycerols are essentially insoluble in water and so cannot be taken up directly by adipocytes from blood lipoproteins; thus, the fatty acids are released by the action of the enzyme lipoprotein lipase. This enzyme is synthesized in adipocytes and then secreted, after which it migrates to the inner surface of the cells lining the blood capillaries. Whereas most of the fatty acids released by the action of lipoprotein lipase are taken up by the adipocytes, some are released into the blood and used by other tissues. The relative importance of *de novo* synthesis and lipoprotein lipase activity as a source of fatty acids for fat synthesis depends on the diet and the species. When animals are fed high-fat diets, chylomicron lipids are the major source. When animals are fed diets rich in carbohydrates, the major source becomes VLDL lipids or *de novo* fatty acid synthesis in adipocytes, depending on whether adipocytes or the liver are the major site of fatty acid synthesis in the species.

Once synthesized within the adipocyte, triacylglycerols are stored in the lipid droplet. Fatty acids are released from them when required by the action of the enzyme hormone-sensitive lipase (distinct from lipoprotein lipase). This enzyme cleaves two molecules of fatty acids to yield a monoacylglycerol that is then hydrolyzed to glycerol and fatty acid by a separate enzyme. Essentially all the glycerol is released from the cell as it cannot be metabolised by adipocytes. Some fatty acids, however, are usually reesterified, and so the ratio of fatty acid to glycerol leaving the cell is normally less than the theoretical 3:1. Released fatty acid is bound to albumin in the blood and transported to the liver and other tissues. Fatty acid esterification and triacylglycerol hydrolysis (lipolysis) occur continuously, i.e., there is a continual turnover of adipocyte triacylglycerol. Net accretion or loss of lipid thus depends on the relative rates of these two processes.

Regulation of Adipose Tissue Metabolism

Both lipid synthesis and hydrolysis are under complex hormonal control. Hormones regulate the amounts of key enzymes and other proteins involved, as well as their activities. In addition, the 'signal transduction' systems (a series of reactions transmitting hormone-induced signals to targets in the cell), through which hormones achieve their effects, are also subject to endocrine control themselves, and changes in the ability of adipocytes to transmit such signals are an important part of the adaptations to some physiological states (e.g., lactation).

Regulation of fatty acid synthesis depends on the precursor. For glucose, control begins at the point of entry into the cell where its transport is dependent on a specific carrier protein (transporter); the major glucose transporter of adipocytes is called 'glut 4.' Insulin stimulates glucose transport both by promoting recruitment of glut 4 into the plasma membrane and by increasing its activity. Within the cell, glucose is initially phosphorylated and then metabolized by a long series of reactions, some in the cytosol, some in the mitochondria, to produce acetyl coenzyme A (CoA) in the cytosol. Several enzymes, in particular phosphofructokinase and pyruvate dehydrogenase, have key roles in controling this flux. Insulin, for example, activates pyruvate dehydrogenase. For acetate, the control is much simpler as its initial reaction results in the production of acetyl CoA. The conversion of acetyl CoA to fatty acid is catalyzed by two enzymes, acetyl CoA carboxylase and fatty acid synthetase. The former is thought to be the most important enzyme controling flux. Both the amount of acetyl CoA carboxylase and its activation status (it is an enzyme that exists in active and inactive forms in the cell) change markedly with physiological, nutritional, and pathological condition. The amount and activity, for example, are decreased by fasting, high-fat diets, diabetes, and lactation. Insulin increases both the amount and activity of the enzyme. These effects of insulin are antagonized by growth hormone. Catecholamines and glucagon also cause inactivation of the enzyme and hence a fall in the rate of fatty acid synthesis.

Insulin increases the synthesis and secretion of lipoprotein lipase; this effect is accentuated by glucocorticoids. Gastric inhibitory polypeptide also increases lipoprotein lipase activity; this effect is likely to be important for promoting fat deposition in animals eating high-fat diets as such diets stimulate secretion of this hormone. Thus, insulin and certain gut hormones increase fat synthesis by increasing the supply of fatty acids for esterification. Insulin also promotes glycerol 3-phosphate formation, in part at least, by increasing glucose uptake by adipocytes. The rate of fatty acid esterification itself may not be stimulated directly by hormones but varies directly with fatty acid availability. Curiously, adipocytes secrete adipsin and two related proteins, which interact in the presence of chylomicrons, to produce acylation-stimulating protein, which then acts on adipocytes to stimulate esterification and glucose uptake.

The enzyme controling lipolysis, hormone-sensitive lipase, exists in active and inactive states in the fat cell. Glucagon and adrenaline (epinephrine), and also

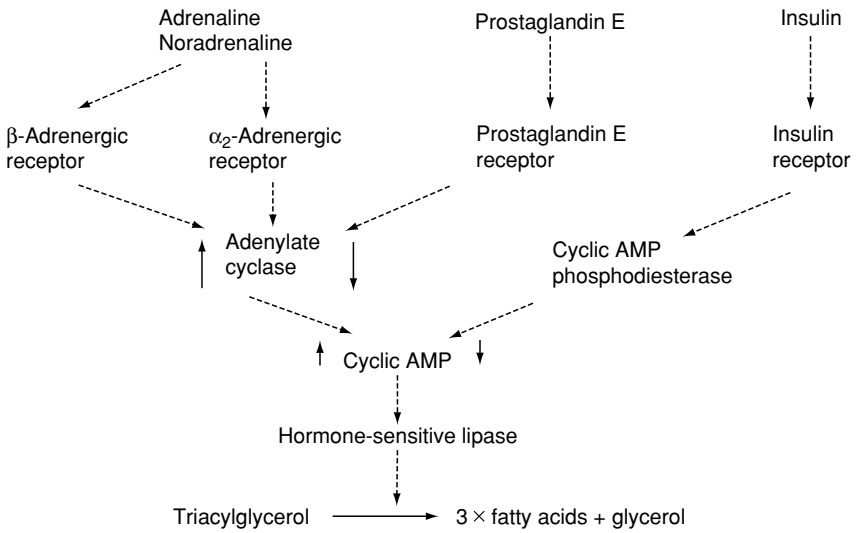

Figure 5 Control of triacylglycerol hydrolysis (lipolysis) by the catecholamines (adrenaline and noradrenaline) and insulin. AMP, adenosine monophosphate; ↑, ↓, activity/concentration increased or decreased by stimulus, respectively.

noradrenaline (norepinephrine) (which is released from nerve endings of the sympathetic nervous system within the tissue itself), interact with specific receptor proteins in the plasma membrane (Figure 5). This causes activation of a key enzyme, adenylate cyclase, which synthesizes cyclic adenosine monophosphate (cAMP). Increased concentrations of cAMP both activate hormone-sensitive lipase and promote its movement from the cytosol to the surface of the lipid droplet, resulting in increased lipolysis. This stimulatory mechanism is attenuated by several inhibitory systems. Adenosine and prostaglandin E_2, which are both produced within adipose tissue, interact with their own receptors, leading to inhibition of adenylate cyclase. Curiously, adrenaline and noradrenaline can both activate and inhibit adenylate cyclase. They activate adenylate cyclase by interacting with β-adrenergic receptors and inhibit by interacting with α_2-adrenergic receptors. The effect of adrenaline and noradrenaline on lipolysis will thus depend in part on the relative number of β- and α_2-adrenergic receptors in the adipocytes. There is considerable site- and gender-specific variation in the ratio of α_2- to β-adrenergic receptor number of adipocytes in some species. For example, in women, intraabdominal adipocytes have a ratio of about 1:1, whereas subcutaneous femoral and gluteal adipocytes have a ratio of about 10:1 α_2-:β-adrenergic receptors. This ratio is thought to be responsible for the very poor lipolytic response to catecholamines of these subcutaneous adipocytes in women and hence the relatively large size of these cells compared with adipocytes elsewhere in the body. In addition to the above, insulin activates the enzyme, cAMP-phosphodiesterase, which catalyzes the degradation of cAMP and so reduces its concentration. The rate of lipolysis then will depend on the concentration of a whole range of hormones, locally produced factors, and neurohumoral transmitters (substances, such as noradrenaline, which are released by nerve endings in tissues). In addition, the ability of the 'signal transduction' system to transmit signals varies with age and with physiological state. For example, during lactation, when fat is often mobilized to support milk production, the system can become more responsive to agents that promote lipolysis. Thyroid hormones, glucocorticoids, sex steroids, and growth hormone all act on one or more components of the signal transduction system, altering its ability to respond to stimulatory and/or inhibitory agents.

Adipose tissue metabolism is thus under complex control. In general, insulin promotes fat synthesis and inhibits lipolysis, whereas catecholamines and glucagon inhibit synthesis and promote lipolysis. In addition, steroid hormones, thyroid hormones, and growth hormone act to modulate the effects of insulin and catecholamines, in part at least, by modifying the ability of the signal transduction systems to transmit signals.

Composition of Stored Fat

Triacylglycerols comprise about 95% of adipose tissue lipid; the remainder includes diacylglycerols, phospholipids, unesterified fatty acids, and cholesterol. The fatty acid composition of the triacylglycerols shows species variation (Table 1), but oleic and

Table 1 Fatty acid composition of adipose tissue triacylglycerols (representative values)

Fatty acids (g per 100 g of total fatty acids)				
Fatty acid	*Humans*	*Pig*	*Sheep*	*Chicken*
Myristic	4	1	3	1
Palmitic	23	26	22	26
Palmitoleic	5	3	4	6
Stearic	6	13	20	7
Oleic	49	42	39	40
Linoleic	9	13	3	19
Linolenic	1	2	2	1
Other	3		7	

palmitic acids are major components in all species. The proportions of polyunsaturated fatty acids (linoleic and linolenic) are usually low in adipose tissue from ruminant animals and higher in chicken and pig adipose tissue. This reflects the dietary supply; as described above, fatty acids are derived both from dietary lipid (via chylomicrons) and from *de novo* synthesis (which produces palmitic acid). There is some capacity for chain elongation of palmitic acid to produce stearic acid, and for desaturation, which converts palmitic to palmitoleic and stearic to oleic acids, but the tissue cannot synthesize linoleic or linolenic acids. In simple-stomached species, such as humans and pigs, varying the fatty acid composition of the diet will alter the fatty acid composition of adipose tissue lipids. For ruminant animals, however, dietary polyunsaturated fatty acids are mostly hydrogenated in the rumen to produce oleic and stearic acids. The small amount of linoleic and linolenic acids escaping this fate is conserved for essential functions (membrane synthesis, prostaglandin production), so that adipose tissue lipids (and milk fat) normally contain little linoleic or linolenic acids. This is ironic, for linolenic acid is the major fatty acid of the ruminant diet. If hydrogenation in the rumen is avoided (e.g., by coating dietary lipid with formaldehyde-treated casein), large quantities of these polyunsaturated fatty acids are absorbed, producing adipose tissue rich in linoleic and linolenic acids.

Minor changes in the fatty acid composition occur during development, and there are minor differences between adipose tissue depots, but these are small compared with the changes that can be elicited by dietary manipulation.

See also: **Fats**: Production of Animal Fats; **Fatty Acids**: Properties; **Hormones**: Adrenal Hormones; Pituitary Hormones; **Obesity**: Etiology and Diagnosis; Fat Distribution

Further Reading

Bjorntorp P (1991) Adipose tissue distribution and function. *International Journal of Obesity* 15: 67–81.

Flier JS (1995) The adipocyte: storage depot or node on the information superhighway? *Cell* 80: 15–18.

Flint DJ and Vernon RG (1993) Hormones and adipose tissue growth. In: Pang PKT, Scanes CG and Schreibman MP (eds) *Vertebrate Endocrinology: Fundamentals and Biomedical Implications*, pp. 469–494. Orlando, FL: Academic Press.

Friedman JM and Halaas JL (1998) Leptin and the regulation of body weight in mammals. *Nature* 395: 763–770.

Gregoire FM, Smas CM and Sul HS (1998) Understanding adipocyte differentiation. *Physiological Reviews* 78: 783–809.

Mohammed-Ali V, Pinkey JH and Coppack SW (1998) Adipose tissue as an endocrine and paracrine organ. *International Journal of Obesity* 22: 1145–1158.

Pond CM (1992) An evolutionary and functional view of mammalian adipose tissue. *Proceedings of the Nutrition Society* 51: 367–377.

Spiegelman BM and Flier JS (1996) Adipogenesis and obesity – rounding out the big picture. *Cell* 87: 377–389.

Vernon RG (1992) Control of lipogenesis and lipolysis. In: Buttery PJ, Boorman KN and Lindsay DB (eds) *The Control of Fat and Lean Deposition*, pp. 59–80. Oxford: Butterworth-Heinemann.

Vernon RG, Barber MC and Travers MT (1999) Present and future studies on lipogenesis in animals and human subjects. *Proceedings of the Nutrition Society* 58: 541–549.

Structure and Function of Brown Adipose Tissue

M J Stock*, St George's Hospital Medical School, Tooting, London, UK

S Cinti, Universita degli Studi di Ancona, Ancona, Italy

Brown Adipose Tissue

Brown adipose tissue (BAT), or brown fat, is a small but highly specialized tissue, the main function of which is to produce heat (thermogenesis). This function requires a good blood supply and a dense population of mitochondria – two features that account for its reddish brown color and distinguish it from white adipose tissue (WAT) (see **Figure 1**). It is found in most mammals, particularly in the neonate, and plays an important role in the control of body temperature during exposure to the cold. There is

*Author deceased.

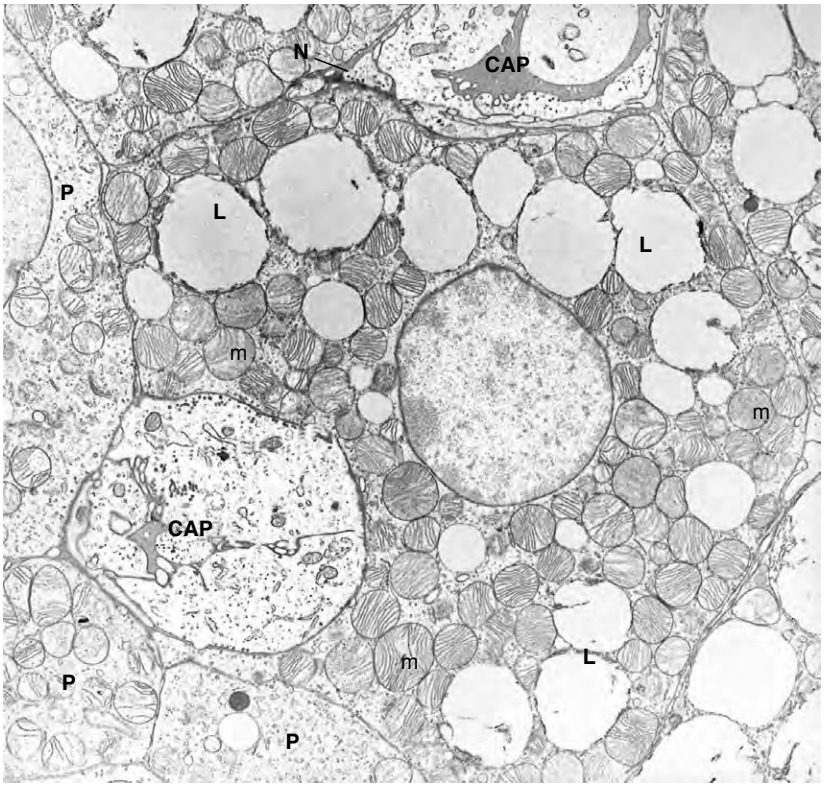

Figure 1 Electron micrograph of brown adipose tissue showing the typical features of a highly thermogenic tissue, i.e., a dense population of well-developed mitochondria, lipid droplets, rich nerve (sympathetic) and blood (capillaries) supply. m, mitochondria; L, lipid droplets; CAP, capillary; N, nerve fiber; P, precursor.

evidence indicating that it is also involved in the regulation of energy balance. The tissue was first described some 300 years ago, but its thermogenic function was not recognized until the early 1960s, and only during the 1980s did its capacity for thermogenesis and its unique metabolism come to be fully appreciated. (*See* **Thermogenesis**.)

Location

BAT is most obvious in small mammals, hibernators, and neonates, and is usually found around the kidneys, heart and aorta, along the intercostal muscles and sternum, in the axilla, in the subcutaneous inter- and subscapular regions, and deep within the neck, around the main arteries and veins. This distribution suggests that the tissues act as a jacket to heat the major organs and warm the blood passing from the periphery into the trunk. The distribution varies considerably between species, and some (e.g., dog, human) have little or no interscapular BAT, whereas in others (e.g., rodents), the interscapular depot may account for 20–30% of the total. BAT rarely exceeds 2–3% of body mass, and is present in such small quantities in large adult mammals that it is often impossible to detect visually. In spite of this, BAT has been identified histologically in human adults up to the age of 80 years or more, and biochemical tests suggest that it might retain its thermogenic activity. BAT depots often contain white adipocytes, and some WAT depots may contain brown adipocytes, but these can be difficult to see.

Histology and Development

Brown adipocytes appear polygonal under the microscope, with a diameter of 10–25 μm, compared with 20–150 μm for white adipocytes. The adipocytes are organized in discrete lobules, surrounded by connective tissue, extensive blood vessels and numerous sympathetic nerves terminating on the adipocytes and blood vessels. Unlike white adipocytes, the nuclei are spherical and located centrally, and the lipid is stored in small, multilocular droplets. Between the droplets and packing the cytoplasm are numerous, well-developed mitochondria that possess distinctive and regular cristae, often traversing the width of the mitochondrion. The endoplasmic reticulum (particularly the rough reticulum) and Golgi apparatus are relatively small, and lysosomes, peroxisomes, and clusters of glycogen granules are often present; adjacent cells are usually connected by gap junctions.

Cytogenic studies indicate that brown adipocytes are derived from stem cells closely associated with vascular structures, and it is now generally agreed that these are distinct from stem cells that give rise to white adipocytes. Mature brown adipocytes cannot undergo mitosis, and the recruitment (hyperplasia) seen during cold adaptation occurs by cytogenesis and mitosis of newly differentiated brown adipocytes. The first appearance of differentiated BAT cells varies between species, and in some neonates (e.g., guinea-pig, rabbit, puppy, lamb), the tissue is well developed and functional at birth. In other species (e.g., rats, mice), the tissue is not fully functional at birth, but becomes thermogenically active within a few days. By contrast, the Syrian hamster is born without BAT, and it takes about 2 weeks for the tissue to develop, during which time, the animal is essentially poikilothermic. Morphology is highly dependent on age, strain, environment, and various physiological and pathological conditions. Brown adipocytes will transform gradually into what look like white adipocytes during prolonged inactivity.

Innervation

The innervation of BAT is another feature that distinguishes it from WAT, since the metabolic activity of the tissue is almost entirely determined by the release of noradrenaline at sympathetic nerve terminals on the brown adipocytes. In some depots (e.g., rodent interscapular BAT), the sympathetic nerves enter as obvious bundles. This makes experimental techniques such as surgical sympathectomy and nerve stimulation and recordings relatively easy to undertake, although there can be problems in distinguishing between effects on adipocytes and those on the vascular supply. The parenchymal sympathetic fibers innervating adipocytes and arterioles release mainly noradrenaline, and this explains why the tissue content and turnover of noradrenaline are high; noradrenaline turnover is a good index of sympathetic activation in response to various environmental and dietary stimuli. Apart from noradrenaline, histamine, adenosine, and various peptides may modulate the sympathetic activation of BAT. Neuropeptide-Y (NPY) is found colocalized with noradrenaline in perivascular sympathetic nerve endings, and the depletion of sensory peptides – CGRP (calcitonin gene-related peptide) and Substance P – by capsaicin suggests that the tissue contains afferent fibers.

Blood Supply

The high oxygen supply required to support thermogenesis is provided by an extensive network of vessels, estimated to be four to six times denser than that in white adipose tissue. The vascular supply can support a blood flow in excess of 20 ml per gram of tissue per minute; during maximal stimulation in cold-adapted rodents, this relatively small mass of tissue can receive over 30% of cardiac output. Blood flow increases result partly from the vasomotor activity of the sympathetic nerves, but also from autoregulatory increases caused by sympathetic activation of metabolism and the release of metabolites. Aerobic heat production can be so intense that the oxygen supplied in arterial blood is almost completely extracted, and the venous blood appears desaturated. The small amounts of oxygen remaining probably represent blood that bypassed the capillary network via arteriovenous anastomoses (i.e., vascular shunts). These vascular shunts, of which there are many, probably act to convect the heat generated away from the tissue, thereby avoiding thermal damage (BAT temperatures can rise to over 44 °C). The thermogenic capacity of BAT can be determined from measurements of blood flow and oxygen extraction, and estimates of up to $500\,W\,kg^{-1}$ can be compared with values of only $60\,W\,kg^{-1}$ for the maximal aerobic power of skeletal muscle. (*See* **Exercise**: Muscle.)

Metabolism

The exceptional heat-producing capacity of BAT is due to its mitochondria, which possess a 32-kDa polypeptide called uncoupling protein (UCP). This is now known as UCP1, since two other, similar mitochondrial proteins (UCP2 and UCP3) have been discovered, but UCP1 is unique to BAT mitochondria and is responsible for the only significant, physiological example of uncoupled oxidative phosphorylation in mammalian metabolism. UCP forms a proton conductance channel in the mitochondrial inner membrane, and dissipates the proton electrochemical gradient generated by oxidation of substrates via the electron transport system. This has the effect of uncoupling oxidation from the phosphorylation of ADP (adenosine diphosphate) to ATP (adenosine triphosphate), thereby dissipating the energy released as heat, as well as increasing the rate of oxidation due to the loss of respiratory control.

The proton conductance pathway is under inhibitory control by purine nucleotides (e.g., ADP, ATP, GDP), which bind to UCP, and is activated following sympathetic activation of the adipocyte β-adrenergic receptors, which also stimulate lipolysis and the release of free fatty acids from the triglyceride droplets. These fatty acids provide the principal fuel for thermogenesis. The rapid activation of the proton conductance pathway following sympathetic

stimulation can be detected by measuring the mitochondrial binding of purine nucleotides – usually GDP (guanosine diphosphate) – *in vitro*, whereas chronic, adaptive changes in thermogenic capacity depend on immunoassay of mitochondrial UCP concentrations.

High rates of oxidation in any tissue require adequate levels of all the enzyme systems of intermediary metabolism, and BAT is particularly well endowed with those required for glycolysis, the tricarboxylic acid cycle, and the mitochondrial electron conductance chain. Since fatty acids are the main fuel for thermogenesis, adenyl cyclase activity and the subsequent cascade that leads to the intracellular release of fatty acids from stored triglyceride are prominent features of BAT metabolism. However, the lipid stored in the multilocular droplets is not sufficient to sustain thermogenesis for long periods, and brown adipocytes then rely on their remarkable capacity for lipogenesis. In cold-adapted rats and mice, the lipogenic capacity of BAT is high enough to account for a major fraction of the amount of dietary carbohydrate that the animal converts to lipid. As well as the fatty acids supplied *de novo* by lipogenesis, the high level of lipoprotein lipase allows BAT to take up fatty acids released by the hydrolysis of circulating triglycerides.

In addition to the normal complement of respiratory enzyme systems, brown fat cells also contain peroxisomes, and these proliferate during chronic stimulation of the tissue. Peroxisomal oxidation of substrates is not linked to phosphorylation, and could therefore make a contribution to cellular thermogenesis. However, the contribution is probably very small, and their function may be more to do with controling levels of free radicals as well as the cytosolic metabolism of fatty acids that are not preferentially metabolized by mitochondria. Another interesting feature of BAT metabolism is the presence of an enzyme, 5′-deiodinase, that converts thyroxine (T_4) to the physiologically active hormone, triiodothyronine (T_3). The enzyme is under sympathetic control, and its activity can increase several hundred-fold in cold-adapted animals. The T_3 produced is more than sufficient to saturate the nuclear receptors, and it is possible that much of the T_3 is exported and exerts effects on other tissues. (*See* **Hormones**: Thyroid Hormones.)

Functions of BAT

Thermoregulation

Shivering is an acute response to cold exposure and not a particularly effective mechanism for protecting the body against hypothermia. As a consequence, many animals resort to a form of heat production called nonshivering thermogenesis (NST), which, unlike shivering, can be sustained without fatigue and disruption of locomotor activity or sleeping behavior. NST appears as an adaptive response to chronic cold exposure in many mammals, but particularly in small animals where heat losses are greater due to the large surface area relative to body mass. The high degree of surface heat loss and immature neuromuscular development also explain why the neonates of most mammalian species (including humans) depend on NST to maintain body temperature until shivering, locomotor activity and other behavioral thermoregulatory responses develop. A third group is the hibernators, who rely on NST for the rapid rewarming that occurs during arousal.

Depending upon the species, NST can raise heat production by 100–300% above that in a warm, thermoneutral environment, and is associated with large increases in the activity of the sympathetic nervous system. Pharmacological blockade (particularly with β-adrenergic antagonists) can inhibit completely the cold-induced rise in heat production, and demonstrates the dominant role of the sympathetic nervous system in mediating NST. The effector tissue is BAT, and a considerable body of evidence now exists to link BAT function to NST. For example, the capacity for NST is inversely proportional to age, bodyweight, and acclimation temperature, and this coincides with histological, physiological, and biochemical indices of BAT activity. Conversely, deacclimation and decreased NST is associated with a parallel decline in BAT activity. Perhaps the most convincing evidence comes from *in vivo* measurements of BAT oxygen consumption, which, in spite of enormous technical difficulties, have shown that the tissue can account for well over 60% of NST. Even this may be an underestimate, since it is not possible to measure the contribution of all the numerous, small and diffuse BAT depots.

Energy-balance Regulation

Evidence linking BAT to energy-balance regulation comes mainly from studies on laboratory rodents that represent examples of two extremes of metabolic efficiency. At one extreme, there are normal, young rats and mice that fail to become obese in spite of an excessive energy intake, and at the other extreme, there are examples of obesity developing in rats and mice (e.g., genetic and hypothalamic obesities), even when energy intake is normal. The explanation for these differences appears to depend on an adaptive form of heat production called diet-induced thermogenesis (DIT), which is absent or defective in obese

animals, but provides a mechanism whereby normal animals can adjust energy expenditure to compensate for energy consumed in excess of requirements. DIT can produce increases in total heat production of 60–70%, and account for up to 90% of the excess energy consumed by hyperphagic rats. In rats feeding normally, the level of DIT is low, but sufficient to control energy balance by compensating for errors in the control of energy intake.

The control and metabolic origins of DIT are identical in almost every respect to NST, although cold is a more potent stimulus and produces more dramatic changes than dietary stimuli. As a consequence, the changes in sympathetic activity, BAT hypertrophy and hyperplasia, mitochondrial proliferation, guanosine diphosphate binding and UCP concentration in rats exhibiting DIT are smaller than those seen in cold-adapted rats. However, these changes in BAT function are sufficient to account for up to 80% of the diet-induced changes in thermogenic capacity seen in hyperphagic rats. By contrast, BAT is usually atrophied and relatively inactive in obese rodents, although it will respond to exogenous noradrenaline, and the animals retain the capacity to adapt to the cold and exhibit NST. This suggests that the defective DIT in these obese rodents is due to a failure of the sympathetic activation of BAT, rather than a defect in BAT itself. This contrasts with what is seen in a transgenic mouse bearing a 'toxigene' that causes a genetic ablation of BAT. These mice fail to exhibit NST and DIT, and become obese – sometimes without eating any more than normal. (*See* **Obesity**: Etiology and Diagnosis.)

Other Functions

In addition to cold- and diet-induced thermogenesis, there are several pathological conditions in which BAT has been implicated as a source of increased heat production. Fever, sepsis, and cancer cachexia are three examples where increased sympathetic activation of BAT is thought to be at least partly responsible for the hypermetabolic response seen in animal models of these conditions, and often involve cytokines such as the interleukins. Patients with pheochromocytoma (adrenomedullary tumor) have very high circulating levels of adrenaline and noradrenaline, and it is thought that the elevated heat production in this condition is due to the stimulatory effect of these catecholamines on BAT; the best examples of active BAT in human adults have been seen in patients with pheochromocytoma.

In spite of increased energy intakes, pregnant rats and mice show little or no change in BAT activity, but during lactation, the tissue atrophies, and its sympathetic activation and thermogenic capacity decline to levels seen after sympathectomy or fasting. Similar reductions can be seen in warm-adapted nonlactating animals, which suggests that BAT thermogenesis declines to compensate for the elevated heat production associated with milk synthesis in the lactating mammary glands. Increased heat production during exercise could also account for the lower BAT activity seen in exercise-trained animals. This is particularly noticeable in cold environments, where exercise can prevent many of the changes in BAT function associated with NST.

Control of BAT

Neural

The control over the sympathetic supply to the various BAT depots originates from the hypothalamus, which receives afferent information on thermal and nutrient status from the periphery, as well as having its own receptor mechanisms and pathways. One of the main thermosensitive and thermoregulatory areas is the preoptic/anterior hypothalamus (POAH), but this is thought to modulate BAT thermogenesis via inhibitory pathways that descend to the lower brainstem. The area that appears to exert a major influence over BAT is one that has been classically associated with the control of energy intake – the ventromedial hypothalamus (VMH), often loosely referred to as the 'satiety center'. Electrical stimulation of the VMH increases BAT thermogenesis, whereas lesions cause the tissue to atrophy, and the latter observation helps explain why VMH-lesioned animals can become obese without overeating. There are connections between the VMH and other hypothalamic areas concerned with feeding behavior (e.g., lateral hypothalamus, paraventricular nucleus), and with the POAH, which provide a neural basis for integrating information on energy intake and body temperature, and modulate the level of NST and DIT accordingly.

Hormonal

Adrenaline stimulates BAT thermogenesis, but it is not as potent as noradrenaline, and in most physiological situations, the circulating levels of adrenaline are probably not sufficient to activate the tissue's β-adrenoceptors. However, views may change on this in the light of recent, more sensitive measurements that show that circulating levels of adrenaline may have been previously underestimated. Although thyroid hormones (T_3 and T_4) are necessary to maintain BAT function, and T_3 is itself produced by the tissue, hyperthyroidism suppresses BAT activity. This is probably due to reduced sympathetic activation

compensating for high levels of heat production in other, thyroid-sensitive tissues.

Glucocorticoids exert little or no direct effects on BAT, even though the tissue has glucocorticoid receptors. However, these adrenocorticoids have central inhibitory actions on the sympathetic outflow to BAT, which are particularly noticeable in genetically obese rodents. Adrenalectomy in these rodents completely restores sympathetic activity, BAT thermogenesis, and DIT to levels seen in lean animals. Only very low replacement doses of glucocorticoids are required to reverse the effects of adrenalectomy, suggesting that these obese animals are hypersensitive to the inhibitory actions of glucocorticoids. The glucocorticoids are thought to inhibit thermogenesis via feedback inhibition of hypothalamic corticotrophin-releasing factor (CRF). In addition to its effects on the pituitary, CRF is a potent stimulus for the sympathetic activation of BAT. Insulin exerts direct, metabolic effects on BAT that are essentially similar to those that it exerts on white adipose tissue, but it also influences the tissue indirectly via its actions on the glucoreceptors of the VMH.

A recent and exciting discovery has been the identification of the lipostatic hormone, leptin. In addition to influencing energy balance (i.e., fat stores) via the hypothalamic control of food intake, leptin also causes sympathetic activation of BAT thermogenesis. This is likely to prove a complex relationship because increased sympathetic activity decreases the expression and release of leptin from both WAT and BAT.

Pharmacological

For many years, the β-adrenergic receptor subtype responsible for activation of BAT lipolysis and thermogenesis was thought to be the β_1-adrenoceptor. However, there is much recent evidence to suggest that the receptor is atypical (neither β_1 nor β_2), and it is now classified as the β_3-adrenoceptor. Much of this evidence came from experiments using novel thermogenic drugs developed for the treatment of obesity, and which were designed to be highly selective agonists of BAT thermogenesis, with little or no effect on β_1-or β_2-mediated functions. Subsequent to the pharmacological identification, the gene sequence for the β_3-adrenoceptor was identified in the human genome. Treatment with novel β_3-adrenoceptor agonists results in significant weight loss (mainly as fat) without affecting food intake in both genetic and dietary obese rodents. However, because the rodent β-adrenoceptor is different, these agonists are not effective in humans, and trials with human β_3-adrenoceptor agonists have yet to be undertaken. BAT also contains α-adrenergic receptors, which are known to be important for activating the conversion of T_4 to T_3 by the 5′-deiodinase, and may also play a minor, facilitatory role in thermogenesis. (*See* **Obesity**: Treatment.)

See also: **Exercise**: Muscle; **Hormones**: Thyroid Hormones; **Obesity**: Etiology and Diagnosis; Treatment; **Thermogenesis**

Further Reading

Cinti S (1999) *The Adipose Organ*. Milan: Editrice Kurtis.

Himms-Hagen J (1990) Brown adipose tissue thermogenesis. In: Schonbaum E and Lomax P (eds) *Thermoregulation and Biochemistry*, pp. 327–414. New York: Pergamon Press.

Rothwell NJ and Stock MJ (1997) Classics in obesity: A role for brown adipose tissue in diet-induced thermogenesis. *Obesity Research* 5: 650–656.

Trayhurn P and Nicholls DG (eds) (1986) *Brown Adipose Tissue*. London: Edward Arnold.

ADOLESCENTS

J J B Anderson, University of North Carolina, Chapel Hill, NC, USA

Introduction

Adequate amounts of energy from carbohydrates, fats, and high-quality protein are considered the critical nutrients for the hormone-driven processes of growth and development of the skeleton and the other organ systems of the body. Developmental sequences vary by gender; Tanner staging is frequently used to assess skeletal development. The needs for nutrients other than those that provide energy and essential amino acids (and N) continue to undergo evaluation as new data are published, and several micronutrients have new recommendations (see below). It has become clear in recent years that adolescents in the USA are not consuming the recommended numbers of servings of foods each day from the basic food groups, which suggests that a very

high percentage are not meeting daily nutrient requirements.

Because of the wide availability of healthful foods but limited physical activities, overweight and obesity have reached epidemic proportions in the US and many technologically advanced nations. The energy imbalance in children as well as in adolescents results from changing societies that value time, speed, and technological advances in their daily lives over physical activities and sports, e.g., physical work, walking, climbing stairs, and other physically demanding chores. It has been estimated in the 1990s that between 30 and 50% of North American adults are overweight (body mass index or BMI >25). Lower percentages of adolescents are overweight, but they represent a much greater prevalence than is desirable because of the tracking of body weight into adulthood and the potential risks for diabetes mellitus type 2, hypertension, and cardiovascular diseases. Obesity in children favors the development of elevated blood pressure and total cholesterol. This trend toward more overweight and obesity in adolescents is facing many nations, especially those in Asia, which have plentiful food available for their youth and declining levels of physical activity.

This review covers the nutritional needs for adolescents, and it also highlights the problem nutrients that are either too little consumed or ingested in excess. Other issues of adolescent nutrition are also addressed.

Nutrient Requirements

Nutrient requirements are divided into macronutrients and micronutrients. Phytomolecules, which are technically not classified as nutrients and which are consumed only in plant foods, may have significant, but under-appreciated, health benefits.

Energy-providing Nutrients

Fats and carbohydrates, when metabolized, generate approximately 85% of the energy used by cells; the other 15% is obtained from the metabolic degradation of the organic backbone of amino acids. Fats also provide the essential fatty acids that exist in plant foods and in fish, but only very minimally in other animal products.

The current adolescent Recommended Dietary Allowances (RDAs) for energy, as kilocalories or kilojoules, are listed in Table 1. As these RDAs are currently under review, it is likely that they will be altered, especially because of the problem of overconsumption of food energy relative to daily energy-expending activities (see above).

Proteins

Proteins supply the essential amino acids, as well as nonessential amino acids; all the amino acids also generate N that is used in the cellular synthesis of many organic molecules, such as nucleic acids.

Table 1 lists the recommended intakes of protein, assuming high-quality protein food sources (protein digestibility and absorption of approximately 75%).

Micronutrients

More than 20 essential micronutrients have been established for humankind, and RDAs exist for most of them. A few micronutrients have become classified as having less scientific basis for establishing RDAs; these micronutrients have therefore been given Adequate Intakes (AIs), even though they may previously have had RDAs. Calcium is a good example. In the 10th edition of the RDAs, calcium RDAs for adolescent girls and boys were 1200 mg per day from 13 to 19 years of age. In the 1997 edition, the calcium AIs have been increased to 1300 mg per day for the same age range.

Other micronutrients with AIs include vitamin D, fluoride, and choline.

Table 2 lists the recommended intakes of the micronutrients.

Water

Often a forgotten nutrient, water carries no specific recommendation. Aside from thirst, a general guide to water consumption is the drinking of 1–1.5 l of water and/or juices, but not counting caffeinated or alcoholic beverages because they have mild diuretic effects. Athletes and those with physically demanding work will need more than the amounts given. The goal is to maintain water balance and, therefore, to avoid dehydration.

Dietary Fiber

No specific recommendation has been made for dietary fiber, but food-labeling guidelines in the US suggest that 25 g of fiber (total of both soluble and insoluble fibers) per day should be consumed with a 2000-kcal diet and 30 g of fiber with a 2500-kcal diet. It is generally assumed that the vast majority of adolescents in the USA consume only 10–15 g of fiber per day.

Phytochemicals

The broad group of phytochemicals, which are not nutrients *per se*, have become increasingly recognized as being important components of our diets because of their numerous roles, such as antioxidants, in

Table 1 Food and Nutrition Board, National Academy of Sciences – National Research Council Recommended Dietary Allowances[a], revised 1989 (abridged), designed for the maintenance of good nutrition of practically all healthy people in the USA

Category	Age (year) or condition	Weight[b]		Height[b]		Protein	Vitamin A	Vitamin E	Vitamin K	Vitamin C	Iron (mg)	Zinc	Iodine	Selenium
		(kg)	(lb)	(cm)	(in)	(g)	(μg RE)[c]	(mg α-TE)[d]	(μg)	(μg)		(mg)	(μg)	(μg)
Infants	0.0–0.5	6	13	60	24	13	375	3	5	30	6	5	40	10
	0.5–1.0	9	20	71	28	14	375	4	10	35	10	5	50	15
Children	1–3	13	29	90	35	16	400	6	15	40	10	10	70	20
	4–6	20	44	112	44	24	500	7	20	45	10	10	90	20
	7–10	28	62	132	52	28	700	7	30	45	10	10	120	30
Males	11–14	45	99	157	62	45	1000	10	45	50	12	15	150	40
	15–18	66	145	176	69	59	1000	10	65	60	12	15	150	50
	19–24	72	160	177	70	58	1000	10	70	60	10	15	150	70
	25–50	79	174	176	70	63	1000	10	80	60	10	15	150	70
	51+	77	170	173	68	63	1000	10	80	60	10	15	150	70
Females	11–14	46	101	157	62	46	800	8	45	50	15	12	150	45
	15–18	55	120	163	64	44	800	8	55	60	15	12	150	50
	19–24	58	128	164	65	46	800	8	60	60	15	12	150	55
	25–50	63	138	163	64	50	800	8	65	60	15	12	150	55
	51+	65	143	160	63	50	800	8	65	60	10	12	150	55
Pregnant						60	800	10	65	70	30	15	175	65
Lactating	1st 6 months					65	1300	12	65	95	15	19	200	75
	2nd 6 months					62	1200	11	65	90	15	16	200	75

[a]The allowances, expressed as average daily intakes over time, are intended to provide for individual variations among most normal persons as they live in the USA under usual environmental stresses. Diets should be based on a variety of common foods in order to provide other nutrients for which human requirements have been less well defined. See text for a detailed discussion of allowances and of nutrients not tabulated.

[b]Weights and heights of Reference Adults are actual medians for the USA population of the designated age, as reported by NHANES II. The median weights and heights of those under 19 years of age were taken from Hamill *et al*. (1979) (see pp. 16–17). The use of these figures does not imply that the height-to-weight ratios are ideal.

[c]Retinol equivalents. 1 retinol equivalent = 1 μg of retinol or 6 μg of β-carotene. See text for calculation of vitamin A activity of diets as retinol equivalents.

[d]α-Tocopherol equivalents. 1 mg of d-α tocopherol = 1 α-TE. See text for variation in allowances and calculation of vitamin E activity of the diet as α-tocopherol equivalents.

This table does not include nutrients for which Dietary Reference Intakes have recently been established [see *Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride*, 1997 and *Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B_6, Folate, Vitamin B_{12}, Pantothenic Acid, Biotin, and Choline*, 1998].

promoting health. Because these nonnutrient molecules come exclusively from plant foods, adolescents who avoid fruits and vegetables will consume little of these molecules and not receive their beneficial effects. Genistein, a soy-derived isoflavone, has been shown to have several possible benefits on health.

Nutrition and Skeletal Growth

Growth of the skeleton (height) is typically completed during adolescence by both girls and boys, and bone mass is practically maximized at this same time. For a large majority of girls in the USA, the bone mass achieved during this adolescent period is considered to be less than optimal because of limited selections of nutrient-dense foods and preferences for low nutrient-dense foods. Figure 1 illustrates the declining consumption of milk and the increasing intake of soft-drink beverages over the last two decades in the USA. Concern has arisen that girls, and possibly boys, are not getting sufficient calcium and other essential nutrients from milk and related dairy products, because so many of these nutrients are essential for the growth of skeletal tissue, and also that adolescent females may be consuming excessive amounts of phosphorus, which alters the calcium:phosphorus ratio in a deleterious way (see Calcium and Phosphorus sections).

Although not the only nutrient required for the extracellular mineral phase of bone, calcium may be the limiting nutrient for mineralization. This divalent cation is provided in the diet by only a few foods that are commonly used and available in most countries. In the USA, dairy foods (exclusive of butter) and calcium-fortified foods provide the bulk of the cation, i.e., approximately 60–70%, in the diets of adolescents. Dark green leafy vegetables, such as broccoli, provide modest, but important, amounts of calcium.

The problem of a shortfall in calcium in the presence of adequate, or possibly, excessive phosphorus consumption is that the low calcium:phosphorus ratio potentially produces a persistent elevation in parathyroid hormone (PTH) that in turn prevents adequate mineralization of the skeleton during the

Table 2 Food and Nutrition Board, Institute of Medicine – National Academy of Sciences Dietary Reference Intakes: recommended levels for individual intake

Life-stage group	*Calcium (mg per day)*	*Phosphorus (mg per day)*	*Magnesium (mg per day)*	*D (mg per day)*[a,b]	*Fluoride (mg per day)*	*Thiamin (mg per day)*	*Riboflavin (mg per day)*	*Niacin (mg per day)*[c]	*B_6 (mg per day)*	*Folate (μg per day)*[d]	*B_{12} (μg per day)*	*Pantothenic acid (mg per day)*	*Biotin (μg per day)*	*Choline*[e] *(mg per day)*
Infants														
0–5 months	210*	110*	30*	5*	0.01*	0.2*	0.3*	2*	0.1*	65*	0.4*	1.7*	5*	125*
6–11 months	270*	275*	75*	5*	0.5*	0.3*	0.4*	3*	0.3*	80*	0.5*	1.8*	6*	150*
Children														
1–3 years	500*	**460**	**80**	5*	0.7*	**0.5**	**0.5**	**6**	**0.5**	**150**	**0.9**	2*	8*	200*
4–8 years	800*	**500**	**130**	5*	1*	**0.6**	**0.6**	**8**	**0.6**	**200**	**1.2**	3*	12*	250*
Males														
9–13 years	1300*	**1250**	**240**	5*	2*	**0.9**	**0.9**	**12**	**1.0**	**300**	**1.8**	4*	20*	375*
14–18 years	1300*	**1250**	**410**	5*	3*	**1.2**	**1.3**	**16**	**1.3**	**400**	**2.4**	5*	25*	550*
19–30 years	1000*	**700**	**400**	5*	4*	**1.2**	**1.3**	**16**	**1.3**	**400**	**2.4**	5*	30*	550*
31–50 years	1000*	**700**	**420**	5*	4*	**1.2**	**1.3**	**16**	**1.3**	**400**	**2.4**	5*	30*	550*
51–70 years	1200*	**700**	**420**	10*	4*	**1.2**	**1.3**	**16**	**1.7**	**400**	**2.4**[f]	5*	30*	550*
>70 years	1200*	**700**	**420**	15*	4*	**1.2**	**1.3**	**16**	**1.7**	**400**	**2.4**[f]	5*	30*	550*
Females														
9–13 years	1300*	**1250**	**240**	5*	2*	**0.9**	**0.9**	**12**	**1.0**	**300**	**1.8**	4*	20*	375*
14–18 years	1300*	**1250**	**360**	5*	3*	**1.0**	**1.0**	**14**	**1.2**	**400**[g]	**2.4**	5*	25*	400*
19–30 years	1000*	**700**	**310**	5*	3*	**1.1**	**1.1**	**14**	**1.3**	**400**[g]	**2.4**	5*	30*	425*
31–50 years	1000*	**700**	**320**	5*	3*	**1.1**	**1.1**	**14**	**1.3**	**400**[g]	**2.4**	5*	30*	425*
51–70 years	1200*	**700**	**320**	10*	3*	**1.1**	**1.1**	**14**	**1.5**	**400**[g]	**2.4**[f]	5*	30*	425*
>70 years	1200*	**700**	**320**	15*	3*	**1.1**	**1.1**	**14**	**1.5**	**400**[g]	**2.4**[f]	5*	30*	425*
Pregnancy														
≤18 years	1300*	**1250**	**400**	5*	3*	**1.4**	**1.4**	**18**	**1.9**	**600**[h]	**2.6**	6*	30*	450*
19–30 years	1000*	**700**	**350**	5*	3*	**1.4**	**1.4**	**18**	**1.9**	**600**[h]	**2.6**	6*	30*	450*
31–50 years	1000*	**700**	**360**	5*	3*	**1.4**	**1.4**	**18**	**1.9**	**600**[h]	**2.6**	6*	30*	450*
Lactation														
≤18 years	1300*	**1250**	**360**	5*	3*	**1.5**	**1.6**	**17**	**2.0**	**500**	**2.8**	7*	35*	550*
19–30 years	1000*	**700**	**310**	5*	3*	**1.5**	**1.6**	**17**	**2.0**	**500**	**2.8**	7*	35*	550*
31–50 years	1000*	**700**	**320**	5*	3*	**1.5**	**1.6**	**17**	**2.0**	**500**	**2.8**	7*	35*	550*

[a]As cholecalciferol. 1 μg of cholecalciferol = 40 IU vitamin D.
[b]In the absence of adequate exposure to sunlight.
[c]As niacin equivalents. 1 mg of niacin = 60 mg of tryptophan.
[d]As dietary folate equivalents (DFE). 1 DFE = 1 μg of food folate = 0.6 μg of folic acid (from fortified food or supplement) consumed with food = 0.5 μg of synthetic (supplemental) folic acid taken on an empty stomach.
[e]Although AIs have been set for choline, there are few data to assess whether a dietary supply of choline is needed at all stages of the life cycle, and it may be that the choline requirement can be met by endogenous synthesis at some of these stages.
[f]Since 10–30% of older people may malabsorb food-bound B_{12}, it is advisable for those older than 50 years to meet their RDA mainly by consuming foods fortified with B_{12} or a B_{12}-containing supplement.
[g]In view of evidence linking folate intake with neural tube defects in the fetus, it is recommended that all women capable of becoming pregnant consume 400 μg of synthetic folic acid from fortified foods and/or supplements in addition to intake of food folate from a varied diet.
[h]It is assumed that women will continue consuming 400 μg of folic acid until their pregnancy is confirmed and they enter prenatal care, which ordinarily occurs after the end of the periconceptional period – the critical time for formation of the neural tube.

This table presents Recommended Dietary Allowances (RDAs) in bold type and Adequate Intakes (AIs) in ordinary type followed by an asterisk (*). RDAs and AIs may both be used as goals for individual intake. RDAs are set to meet the needs of almost all (97–98%) individuals in a group. For healthy breastfed infants, the AI is the mean intake. The AI for other life stage groups is believed to cover their needs, but lack of data or uncertainty in the data prevent clear specification of this coverage.

adolescent growth phase, and also earlier during prepuberty. The net result is suboptimal skeletal development and peak bone mass accrual by the end of adolescence that is less than maximal. So, low dietary calcium intakes, especially in relation to elevated phosphorus intakes, may be disadvantageous for young girls and also for young boys. The problem may be more widespread than previously thought because of the possibility that the fracture rate in adolescent girls may be underestimated.

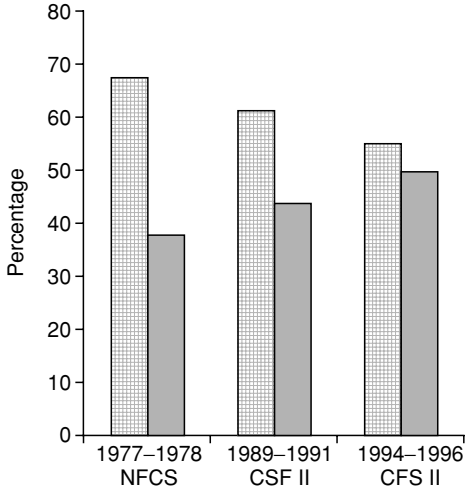

Figure 1 Percentages of US individuals in USDA National Food Consumption Surveys (NFCS and CSF) consuming milk ▦ and soft drinks ■ between 1977–1978 and 1994–1996. Note the decline in milk consumption while soft-drink consumption increased, according to percentage of consumers.

Specific Problem Micronutrients: Deficits and Excesses

Most of the mineral micronutrients reviewed below exist in the body in solutions as cations (positively charged) or anions (negatively charged). The vitamins, however, exist as organic molecules that may carry a slight electrical charge that permits their binding, usually for only brief times, to other molecules.

Calcium

Inadequate consumption of calcium in the diets of adolescents, especially females, in the USA has gained them the label, 'the lost generation.' Calcium intakes have reached an all-time low, and it has been predicted that the present generation of girls, and perhaps boys, are placing themselves at increased risk for skeletal fractures late in life, as longevity is increasing in the USA and most other nations. The concern about increased age-adjusted rates of osteoporosis assumes that tracking, or following the same relative path as age increases, of bone mass and density occur, similar to the tracking of high blood pressure with increasing age. Therefore, an individual who has less bone at peak bone development is, during aging, assumed to follow a path, i.e., track, at the relatively lower level of bone mass to reach critically low bone values sooner, e.g., after the menopause, than an individual who develops a greater peak bone at the end of the skeletal growth phase (Figure 2).

Iron

Low iron consumption by females is one of the most common nutrient problems throughout the world.

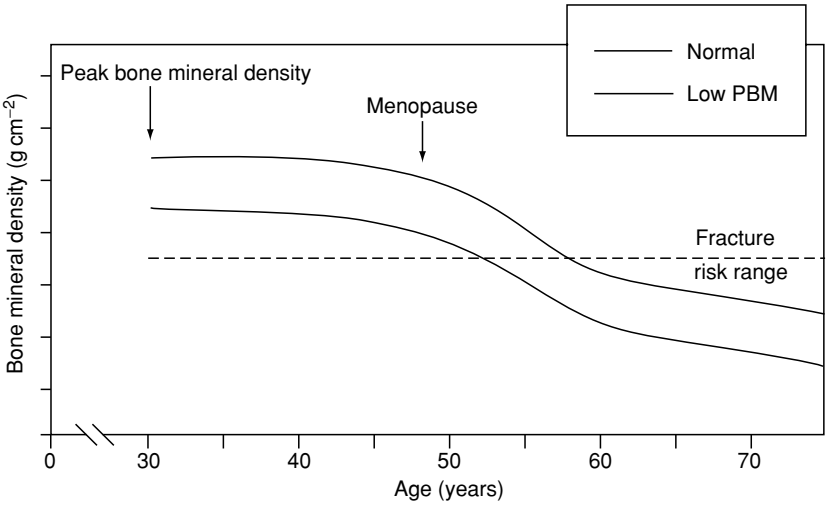

Figure 2 Tracking of bone mass of females from early adulthood to postmenopausal life. Adapted from Anderson JJB (2000) Nutrition in osteoporosis, in Mahan K and Escott-Stump S (eds) *Krause's Food, Nutrition, and Diet Therapy*, 10th edn. Philadelphia, PA: Saunders with permission. PBM = peak bone mass.

Adolescent females are particularly likely to develop iron deficiency, and even the more severe iron-deficiency anemia, if they have little red meat in their diets. In the USA, iron-fortified foods, especially ready-to-eat breakfast cereals, have reduced the severity of the two forms of iron deficits.

Zinc

Too little zinc consumption by adolescents results from poor intakes of plant foods. Red meats have good amounts of zinc, as well as iron, but these would not be consumed by many adolescents who avoid red meats.

Magnesium

Insufficient magnesium typically signifies poor selection of plant foods, especially dark green leafy vegetables. Although magnesium intakes by adolescents are lower than recommended, it has been difficult to pinpoint real physiologic problems, except possibly that optimal heart function may be compromised.

Potassium

Too little potassium consumption reflects a limited selection of fruits and vegetables, the highest potassium-containing sources of this monovalent cation. Both potassium and magnesium have intracellular roles, and in part, adequate intakes of potassium are needed to counter the high intakes of sodium. Therefore, a dietary balance between potassium and sodium is needed to maintain efficient cellular functions that require potassium ions. Renal mechanisms tend to conserve sodium at the expense of potassium and magnesium, which are excreted in increased amounts when sodium intakes are higher than recommended.

Sodium

Excessive amounts of sodium in the diet may contribute to higher blood pressure levels, but still within the normal range, even in adolescents. Sodium consumption by adolescents is a concern, because the taste for this mineral, or perhaps sodium sensitivity, becomes established early in life and continues into adolescence. Higher blood pressure is typically associated with excessive intakes of sodium throughout adolescence tracks with age.

Phosphorus

Excessive consumption of phosphorus from foods and soft drinks that contain phosphoric acid reduces the calcium:phosphorus ratio in the total diet and, thereby, may increase the circulating level of PTH, though the elevated PTH concentration usually remains within the upper range of normality. The net effect is increased calcium losses, i.e., a negative calcium balance. In the USA, food-labeling regulations do not require that information on the phosphorus content of foods be included. This omission means that consumers cannot estimate their daily phosphorus consumption or calculate a calcium:phosphorus ratio of a food or their entire daily intake. In addition, a food may include hidden amounts of phosphorus because of the wide use of phosphate additives in food processing. The magnitude of this potential problem of excessive phosphorus has not been investigated by governmental agencies.

Folic Acid

Recent advances in understanding this nutrient have focused on obtaining sufficient intakes of folic acid to prevent neural tube defects in offspring, but other important roles of this B vitamin exist in cell divisions, e.g., in red blood cell formation, and gut epithelial turnover of cells. In addition, adequate intakes of folate also maintain lower circulating levels of homocysteine, which is now considered a risk factor for coronary heart disease.

Vitamin D

The hormonal form of vitamin enhances intestinal calcium absorption, when calcium intakes are low, and also it increases osteoblastic bone-forming activities. This hormonal metabolite of vitamin D also stimulates osteoblast cells in bone to make new organic matrix (osteoid) and therefore new bone tissue. This metabolite, along with PTH, also contributes to calcium homeostasis, i.e., regulation of blood calcium concentration. Too little intake of vitamin D from the few foods that contain it, and too little sun exposure for skin biosynthesis of this vitamin, puts individuals at risk for reduced functions in the gut and bone tissues. Adolescent intakes of vitamin D are low in the USA because the consumption of milk that is fortified with this vitamin has declined (see above).

Vitamin K

An established requirement for vitamin K exists for both blood clotting and bone health. The daily amount needed for the production of mature osteocalcin may be higher than that needed for the blood-clotting factors, but insufficient information has been generated to be certain. Vitamin K intakes may be insufficient in adolescents because of limited consumption of dark green leafy vegetables and legumes.

Problem Areas

A few problem areas involving meeting nutrient requirements by adolescents need special attention

because of their potentially adverse effects on nutritional status.

Adolescent Athletes

The female athlete triad of disordered eating, amenorrhea or oligomenorrhea, and osteopenia or low bone mass, but generally not osteoporosis, has a marked adverse effect on skeletal mass that has been shown to increase the risk of stress fractures of athletes, especially those participating in cross-country running and other endurance events involving running or repetitive jumping as in volleyball. The risk of fracture, however, is reduced if estrogen is replaced in these athletes.

However, Dutch investigators have suggested that physical activity during adolescence may be even more important for bone development than calcium intake, so long as dietary calcium is in the adequate range, i.e., approximately 1000 mg per day. This conclusion has also been reached for female gymnasts, even amenorrheic gymnasts, because of the positive effects of weight-loading on the skeleton.

Consumption of Alcoholic Beverages

Over-consumption of alcoholic beverages by adolescents in the USA has become a big problem. The adverse effects of excessive alcohol consumption on the nutritional status of adolescents may result from insufficient food consumption, too many empty calories from low-nutrient dense alcoholic beverages, and unwise food choices.

Strict Vegetarians

Vegans or strict vegetarians place themselves at risk for deficits of several micronutrients if supplements are not taken. Fortunately, most vegans know about the pitfalls of this practice of eating, and they do not develop serious deficiencies. However, a few problems, such as iron, zinc, and cobalamin deficiencies, may arise because of the avoidance of all animal foods. Vegans may also not be able to consume sufficient amounts of calcium or of omega-3 polyunsaturated fatty acids. One of the potential disadvantages of a vegetarian diet is a low estrogen status across the life cycle in females. The lower estrogen status may have an adverse effect on the skeleton, although it may be protective against estrogen-dependent cancers and other chronic diseases.

Nutrition Education

Despite the fairly high level of knowledge about nutrition by youth in the US, health practices, including food selection, are too often detrimental to the support of continued skeletal growth. Part of the reason for insufficient calcium consumption by adolescent girls is their limited knowledge about daily calcium allowances and the amounts of calcium in foods. Adolescents, especially females, clearly do not eat the way they should and even, perhaps, know how to eat. These unhealthy eating practices result primarily from bad attitudes toward specific foods, negative reactions to parental influences, and peer pressure.

The serious problem in skeletal development in the more technologically advanced nations relates to the unwise food and drink choices being made by our children, especially after about 10 years of age. In the USA, adolescent females are not only not getting sufficient amounts of calcium, but consuming too little of the nutrient-dense foods in general. The practices of meal-skipping, grazing, snack/convenience food selection, and soft-drink consumption insure suboptimal intakes of several micronutrients in addition to calcium. Surveys of food-consumption patterns of girls demonstrate that inadequate mean consumption exists for the following micronutrients: calcium, magnesium, iron, zinc, and both fat-soluble and water-soluble vitamins, especially E, folate, cobalamin, and other B vitamins. Vitamins D and K typically have not been assessed in these surveys. Mean sodium and phosphorus intakes were higher than recommended.

Avoidance of dairy products by young girls has become widespread, at least in the USA, for fear of weight gain, despite the fact that low-fat milks and no-fat yogurts are widely available. Thus, dairy foods can be consumed with little or no fat. Education of the appropriate fats and amounts of daily fat intake for these females is either lacking or not appropriately highlighted. Some fat is clearly needed for skeletal growth, so that extremely low intakes may be deleterious. Of course, some mono- and polyunsaturated fats found in salad dressings and cooking oils are now being recommended for overall balance within the types of fatty acids and also for optimal health outcomes, but young females are barely aware of these important needs.

Finally, nutrition education messages about the consumption of soft drinks should target both parents and children to limit the number of servings because they displace nutritious beverages, such as milk and fruit juices (see **Figure 1**), especially when two or more 400-ml (12-ounce) soft drinks are consumed in a day (more than 22% of one adolescent population in the USA). Since many of the soft drinks contain phosphoric acid as the acidulant, the calcium:phosphorus ratio becomes even further skewed away from a recommended ratio of 0.70–0.75. The widespread use of soft drinks, in place of milk, has contributed to

the increase in phosphorus consumption and the low calcium:phosphorus dietary ratio of USA adolescents. How serious this trend of increased phosphorus consumption is cannot be resolved without a better understanding of the overall phosphorus intake from processed foods, lowly processed foods, and soft drinks that contain phosphoric acid.

See also: **Alcohol**: Alcohol Consumption; **Calcium**: Properties and Determination; **Dietary Fiber**: Properties and Sources; **Exercise**: Metabolic Requirements; **Folic Acid**: Properties and Determination; **Iron**: Physiology; **Functional Foods**; **Magnesium**; **Phosphorus**: Properties and Determination; **Obesity**: Epidemiology; **Plant Antinutritional Factors**: Characteristics; **Protein**: Requirements; **Sodium**: Properties and Determination; **Vegetarian Diets**; **Vitamin K**: Properties and Determination

Further Reading

Alaimo K, McDowell MA, Briefel R *et al*. (1994) *Dietary Intakes of Vitamins, Minerals, and Fiber of Persons Ages 2 Months and Over in the United States: Third National Health and Examination Survey, Phase 1, 1988–91. Advance Data No. 258 (November 3, 1994)*. Washington, DC: National Center for Health Statistics, CDC, PHS, USHHS.

Anderson JJB (1999) Plant-based diets and bone health: nutritional implications. *American Journal of Clinical Nutrition* 70 (supplement): 539S–542S.

Anderson JJB (2000) Nutrition in osteoporosis. In: Mahan K and Escott-Stump S (eds) *Krause's Food, Nutrition, and Diet Therapy*, 10th edn. Philadelphia, PA: Saunders.

Anderson JJB, Anthony M, Messina M and Garner SC (1999) Effects of phytooestrogens on tissues. *Nutrition Research Reviews* 12: 75–116.

Anderson JJB and Garner SC (eds) (1995) *Calcium and Phosphorus in Health and Disease*. Boca Raton, FL: CRC Press.

Anderson JJB and Miller CP (1998) Lower lifetime estrogen exposure among vegetarians as a possible risk factor for osteoporosis: a hypothesis. *Vegetarian Nutrition* 2: 4–12.

Calvo M and Park YK (1995) Changing phosphorus content of the US diet: Potential for adverse effects on bone. *Journal of Nutrition* 126: 1168S–1180S (1996).

Goulding A, Cannan R, Williams SM *et al*. (1998) Bone mineral density in girls with forearm fractures. *Journal of Bone and Mineral Research* 13: 143–148.

Harel Z, Riggs S, Vaz R, White L and Menzies G (1998) Adolescents and calcium: what they do and do not know and how much they consume. *Journal of Adolescent Health* 22: 225–228.

Harnack L, Stang J and Story M (1999) Soft drink consumption among US children and adolescents: nutritional consequences. *Journal of the American Dietetic Association* 99: 436–441.

Hetzel BS and Berenson GS (1987) *Cardiovascular Risk Factors in Childhood: Epidemiology and Prevention*. Amsterdam: Elsevier Science.

Institute of Medicine, Food and Nutrition Board (1997) *Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride*. Washington, DC: National Academy Press.

Institute of Medicine, Food and Nutrition Board (1998) *Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B_6, Folate, Vitamin B_{12}, Pantothenic Acid, Biotin, and Choline*. Washington, DC: National Academy Press.

McMurray RG, Harrell JS, Levine AA and Gansky SA (1995) Childhood obesity elevates blood pressure and total cholesterol independent of physical activity. *International Journal of Obesity* 19: 881–886.

Munoz KA, Krebs-Smith SM, Bvallards-Barbash R and Cleveland LE (1998) Food intakes of US children and adolescents compared with recommendations. *Pediatrics* 100 (3 Pt 1): 323–329.

National Research Council, Food and Nutrition Board(1989) *Recommended Dietary Allowances*, 10th edn. Washington, DC: National Academy Press.

Simopoulos A (1999) Essential fatty acids in health and disease. *American Journal of Clinical Nutrition* 70 (supplement): 560S–569S.

Tanner JM (1962) *Growth in Adolescence*, 2nd edn. Oxford: Blackwell Scientific.

Thrash L and Anderson JJB (2000) The female athlete triad. *Nutrition Today* 35: 168–174.

Weaver CM, Peacock M and Johnston CC Jr. (1999) Adolescent nutrition in the prevention of postmenopausal osteoporosis. *Journal of Clinical Endocrinology and Metabolism* 84: 1939–1843.

Weaver CM, Proulx WR and Heaney R (1999) Choices for achieving adequate dietary calcium with a vegetarian diet. *American Journal of Clinical Nutrition* 70 (supplement): 543S–548S.

Welten DC, Kemper HCG, Post GB *et al*. (1994) Weight-bearing activity during youth is a more important factor for peak bone mass than calcium intake. *Journal of Bone and Mineral Research* 9: 1089–1095.

Adrenal Hormones *See* **Hormones**: Adrenal Hormones; Thyroid Hormones; Gut Hormones; Pancreatic Hormones; Pituitary Hormones; Steroid Hormones

ADULTERATION OF FOODS

Contents
History and Occurrence
Detection

History and Occurrence

M Tsimidou and D Boskou, Aristotle University of Thessaloniki, Thessaloniki, Greece

Introduction

Food is a basic prerequisite for human survival and also for social and economic welfare and progress. Problems related to food have varied from one period of history to another, from continent to continent, and from country to country. The problem of food adulteration has been a major one and the protection of the consumer has occupied the attention of civil authorities from ancient times.

Food is considered adulterated if it contains poisons or other substances which may render it injurious to the health of the consumer, or if it contains filth or it is decomposed; if it contains a coloring agent or other food additive, that is not approved or contains materials that disguise inferior quality; if any important constituent has been wholly or in part abstracted or any specified ingredient has been substituted by a nonspecified ingredient; if it contains any substance that increases its weight and bulk or changes its strength to improve appearance. A food is misbranded if it is illicitly labeled or it is a food for which standards of identity have been written and it fails to comply with these standards.

History

Food has been liable to adulteration to a greater or lesser extent since very early times. Mosaic and Egyptian laws made provision for preventing contamination of meat, while several centuries before the time of Christ India had regulations prohibiting the adulteration of grain and edible fats. Athens had its public inspector of wines. Adulteration was also common during the Roman period. Evidence for this is given in Apicius' famous cookbook (*De Re Coquinaria*). During the Middle Ages in England, pepper and other costly spices imported from the east were adulterated by mixing with ground nutshells, local seeds, and olive pits.

Between the 13th and 16th centuries, bread, wine, beer, spices, and valuable natural coloring materials were often adulterated. In England in 1319, a meat-market overseer succeeded in putting a butcher in the pillory for selling unsound beef. Wines were 'sophisticated' (adulterated) with burnt sugar, juices, starch and gums, and other substances. Such practices may have reduced the quality of the wine but they were not injurious to health. After the 18th century, however, food adulteration became dangerous. Vinegar was often adulterated with sulfuric acid, wine with preservatives containing lead salts, green vegetables in vinegar with copper (to improve color), essential oils with oil of turpentine, confectionery products with colorings containing lead and arsenic, chocolate with Venetian red (ferric oxide), and red pepper with vermilion (mercury sulfide). 'Black extract', obtained by boiling poisonous berries of *Cocculus indicus* in water and concentrating the fluid, was used in beer. This extract imparted flavor, but also narcotic properties and intoxicating qualities to the beverage.

Bread was not only the basic item of diet for many centuries but also the one most subjected to adulteration. The incorporation of sieved boiled potatoes, chalk, or bone ashes in the flour was a common fraudulent practice, but the most serious example was the addition of alum, which whitened flour of inferior quality.

Recipe books of the 18th century contain instructions, which today would cause alarm. The recipe for preserving green color in pickles is characteristic: 'To render pickles green, boil them with a halfpenny or allow them to stand for 24 hours in copper brass pans.' It is not surprising that records of that period list a number of deaths from copper poisoning.

Frederic Accum, writing from London in 1820, gives a vivid description of the adulteration of bread, alcoholic drinks, tea, coffee, and many other foods. Accum claimed that: 'Indeed, it would be difficult to mention a single article of food which is not to be met

with in an adulterated state; and there are some substances which are scarcely ever to be genuine.' Even if these accusations are somewhat overstated, the seriousness of intentional adulteration of food, which prevailed between 1800 and the early 1900s, is undisputed.

The second decade of the 19th century marks the beginning of the second period in the history of food adulteration. In the first period (ancient times to about 1820) food was procured from small enterprises and individuals who were, to a certain extent, responsible for their own transactions. In the second period (early 1800s to 1900s), the methods of food production changed significantly. Large-scale food production became necessary because of the industrial revolution, which led to a move of population from country to city. This created conditions that were conducive to large-scale adulteration of food. In the second half of the 19th century the first publicly supported analytical laboratories for food inspection were established in Germany and the USA. Intentional adulteration of food remained a serious problem until the beginning of the 20th century. Regulatory pressures and the effectiveness of analytical methods reduced the frequency and magnitude of food adulteration. Further improvements have been achieved up to the present time and, owing to strict legal standards and also to the growth of an increasingly critical public, deliberate adulteration in industrialized countries has become less serious. Of course, fraudulent practices still continue in most countries, especially those which lack adequate means to insure that laws and regulations are enforced.

Food Additives

The excessive use of additives and the use of nonpermitted ones present a serious threat to food safety. It could be argued that a fourth phase of food adulteration began some 40 years ago, when foods containing chemicals became prevalent in the diet. Since then, we have moved gradually from an outrage over fraud, sometimes practiced with dangerous substances, to a growing concern over the more subtle toxic effects of common chemicals added to food to assist manufacture and storage. As a result, there has been a shift in focus from food adulteration to safety of food, and to a different approach which emphasizes not only evaluation of the food in its entirety, but also evaluation of individual components. Current legislation relating to food additives provides statutory lists of all permitted additives which may be used and, where appropriate, the maximum levels for their use. Nevertheless, as the food industry becomes increasingly sophisticated and relies more and more on the use of additives, many consumers believe that governments are unable to control the situation and that preservatives, antioxidants, colorings, flavor or texture modifiers, and other chemicals are and always will remain a potential hazard. (*See* **Food Additives: Safety.**)

Food Legislation

The use of government regulatory powers to control the purity and safety of foods is very old but the first serious efforts to regulate malpractice took place only in the 19th century. The first general food laws were passed in the UK in 1860. Accum's book, *A Treatise on Adulteration of Food and Culinary Poisons* (1820), aroused the indignation of the British public. In 1850, Thomas Wakley, editor of the *Lancet*, established an Analytical Sanitary Commission under the direction of Arthur Hill Hassal, the first man to investigate food adulteration from a scientific point of view. He turned the microscope into a weapon in the detection of adulterants. The Commission's reports, published in the *Lancet* from 1851 to 1854, led to the formation of a parliamentary committee to investigate food adulteration and, in 1860, to the passage of the world's first Food and Drink Act. This law made it a criminal offence to sell adulterated food and drink. It also established the appointment of public analysts to examine foodstuffs produced throughout the country. In 1866, a second country, New Zealand, passed pure food legislation and, in 1874, Canada passed the Food and Drug Law.

In the 20th century, new regulations were developed in industrialized countries, and old regulations were repeatedly updated, taking into account the advice of governmental and international committees. Modern food law establishes and maintains standards for the composition of food, controls the use of additives and extent of contamination, controls packaging materials and unsanitary practices in the production of food, and restrains unwarranted and misleading claims and advertisements. Effective food surveillance systems have also been developed which insure that there is a constant supply of the right kind of food to keep people properly fed and that the food we eat is wholesome as a result of strict adherence to national and international law. The current worldwide access to data and information through the internet may soon reveal that administrative traceability can be successfully used to monitor the sale of fraudulent food products in the global market of the 21st century.

International Standards

The Food and Agriculture Organization (FAO), the World Health Organization (WHO) and the Joint Codex Alimentarius Commission are concerned with the safety of food and food ingredients. The aims of the Codex Alimentarius include protecting the health of the consumer and insuring fair practices in the food trade, coordination of all food standards work, determining priorities and initiating, guiding, and finalizing the preparation of food standards. There are published recommended international standards for individual foods (sugars, edible oils, canned fruits and vegetables, processed meat, fish products, cocoa products, etc.), as well as provisions with respect to food hygiene, food additives, contaminants, and labeling. The key role of the Codex Alimentarius in developing international food standards was recognized when the World Trade Organization (WTO) was set up in 1995. The WTO agreements accept the Codex Alimentarius standards and this enhanced status has raised the profile of the activities of the Codex. However, the impact of WTO on the Codex is not yet clear and fears have been expressed that the body may lose its ability to take decisions based strictly on science.

All of the standards are detailed specifications, most of which require analytical methods for their realization. This task is backed by other standardizing organizations such as the International Union of Pure and Applied Chemistry (IUPAC), the International Standards Organization (ISO), and the Association of Official Analytical Chemists (AOAC), which are involved in the development of methods of sampling and analysis. These organizations consult with the various international bodies representing individual commodity interests, such as the International Dairy Federation (IDF), the International Association for Cereal Chemistry (ICC), the International Olive Oil Council (IOOC), the International Office of Cocoa and Chocolate (IOCC), and others.

European Union Laws

In the European Union (EU), a huge market with a population of 320 million, a programme of food law measures has been designed to create a single European market in foodstuffs. The laws of the community are expressed in the form of regulations and directives. Regulations, made under Article 43 of the Treaty of Rome, are binding on all member states and are directly applicable. The directive has no force of law in a member state until it is incorporated into the legislation of that country. Dates are prescribed in directives to insure compliance within the time stated. Some of the rules and specifications of the EU directives rely on Codex Alimentarius or other international guidelines. The Community is also laying down tougher rules for food analysis which laboratories must adopt. Bringing a multitude of national laws into line with a common objective is a difficult task but food legislation at EU level is expected to deal successfully with questions such as protection of public health, consumer information, and fair-trading. The member states of the EU have speeded up harmonization, since their governments are aware that outdated legislation can no longer act as a basis for competition in this large market. Council directives 89/397/EEC and 93/99/EEC laid down uniform general rules for the official control of foodstuffs and also for laboratories entrusted by the member states to carry out food inspection and analysis. Since 1993 coordinated food inspection programs among the responsible authorities in member states were also set up to establish compliance with food legislation and insure that food is fit for consumption. In July 1996, the so-called Black List Anti-Fraud Regulation came into effect in the EU. This regulation aims to reduce fraud within export refund and intervention measures by intensifying inspection. Those found to have committed a fraud are excluded from EU funds for at least 6 months.

Regulating the Safety of Food in the USA

In the USA there were no food regulation laws prior to the 1900s. At that time food companies could add anything to their products. In 1902, the Director of the Bureau of Chemistry in the US Department of Agriculture (USDA), Harvey Wiley, who was the leader of a campaign against food adulteration, set up a group of male volunteers to evaluate the safety of common food preservatives and ingredients. This group became known as the Poison Squad. At about the same time the Poison Squad was formed, Upton Sinclair wrote a book, *The Jungle*, dealing with the unsanitary conditions of the Chicago meat-packing industry. This crusade resulted in the passage of the Pure Food and Drug Act in 1906: this Act defined food adulteration and made the distribution and sale of adulterated foods and drugs illegal. These early efforts of the Poison Squad marked the beginning of food toxicology and provided impetus for the regulation of substances added to foods.

In 1931, the Food and Drug Administration (FDA) was formed as a separate unit of the USDA to administer the law. This agency regulates food under two general criteria – adulteration and misbranding. The FDA was ineffective until 1938, at which time the

passage of the Federal Food, Drug and Cosmetic Act of 1938 updated and tightened the definition of adulteration. This gave the agency the power to fine violators of the laws.

After 20 years of application, interpretation, and enforcement of this Act, a significant revision took place in 1958. The Food Additives Amendment of 1958 led to a major change in the FDA's approach to its activities. Congress applied the term 'safe' as the criterion for action, supplementing the term 'adulteration.'

Over the years, with advances in science and analytical methods, the FDA has become more confident in its evaluation of food safety and US courts have given general endorsement to the agency's decisions. In the fake food fight, the Agency has sought and won convictions against companies and individuals engaged in selling bogus fruit juices, honey, maple sugar, cream, olive oil, seafood, and many others.

The FDA has major responsibilities for food supply in the USA, but other agencies also exert significant regulatory control over foods and beverages. The major agencies so far involved are the Department of Health and Human Services, the US Department of Agriculture Inspection Services, the US Department of Commerce, the Environmental Protection Agency and Centers for Disease Control and Prevention (CDC). The FDA relies heavily on industry and consumers to help identify instances of economic fraud. The detective work of FDA inspectors in the premises of suspect companies is also valuable. Cooperation with other foreign government bodies prevents adulterated foods banned in one country being shipped to another. Some adulterants recently reported in FDA files are: beet sugar or corn syrup in orange juice; canola oil in olive oil; sugar, water, hydrolyzed inulin syrup in apple juice; corn oil in dairy cream; corn syrup in maple and sorghum syrups or honey; water and sodium tripolyphosphate in scallops; potato starch in horseradish; salt and water in milk and sawdust in ginseng dietary supplement. The Safe Food Act of 1997, by which the food safety and inspection functions of 12 agencies have been consolidated into a single independent agency, is expected to combat even more effectively dishonest or fraudulent practices not only in a federal but also in an international context. (*See* **Food and Drug Administration**.)

Toxicological Implications of Economic Adulteration of Food

In the past, the deliberate adulteration of food with untested chemicals or poisonous substances was widespread. In the 20th century, developments in the analytical field and improvement of legislation have brought many of these illegal and unethical practices under control, but cases of food adulteration with toxicological consequences have not completely disappeared.

One serious case of malpractice in the food industry is the use of denatured rapeseed oil for edible purposes. This caused the outbreak referred to as Spanish toxic oil syndrome. In 1981, a disease broke out in Spain which rapidly took on epidemic proportions. The reported symptoms were respiratory distress, nausea, fever, vomiting, headaches, myalgias, abdominal pain, and skin eruptions. Approximately 20 000 people were affected and more than 400 deaths were recorded. For many years, scientists all over the world tried hard to find an explanation and identify the toxic compounds causing the illness. Today, toxicologists and chemists are convinced that the toxic oil syndrome should be attributed to aniline used for denaturation, but there are different opinions concerning the nature of the toxic compounds. Phenylalamine, originating from the glucosinolates present in rapeseed oil, was for many years believed to be the etiological agent in toxic oil syndrome. The course of events, however, in the two factories processing the aniline containing rapeseed oil and simulating studies conducted in the UK do not confirm this hypothesis.

Another explanation may be the formation of *N*-phenyl-amino propane diol esters of fatty esters under conditions favoring dehydration. This suggests that deodorization sludges containing free fatty acids (and aniline) were catalytically treated with glycerol to obtain reconstruction of triacylglycerols. The problem will remain unresolved unless experts visit Seville again, study in more depth the conditions of the episode, and implement simulation processes that include esterification of fatty acids with glycerol in the presence of aniline.

Cases of Economic Adulteration of Food

Cases of adulteration frequently reported in the literature concern not only the basic foods (oils and fats, milk and dairy products, meat, fish, and cereals) but also fruit juices, alcoholic beverages and wine, honey, chocolate, coffee and spices, essential oils, and flavorings.

Some cases of the last decades with serious economic and other implications are briefly discussed below.

Wines

Some Austrian, German, and Italian wines were found in 1985 to have been illegally supplemented with diethylene glycol (DEG), not a permitted food additive, to improve body and sweetness. Soon after the scandal, a 30% drop in demand for wines was observed in Austria. In 1987 some Japanese wines were also reported to contain the same adulterant. (*See* **Wines**: Types of Table Wine.)

Fruit Juices

One case which involved economic fraud and cover-up efforts is the adulterated apple juice for babies scandal. For many years, Beech-nut Nutrition Corporation purchased and used concentrated apple juice which contained cheap ingredients such as corn syrup, invert sugar, malic acid, caramel, and imitation flavor. Although it was the director of research and development who discovered the use of these adulterants, he was forced to resign. The FDA took action and in 1988 the company was heavily fined.

Citrus juices are frequently adulterated in many ways. The price of grapefruit juice is always lower than that of orange juice, so their admixture can be expected. Misbranding of fruit juices made from imported concentrates is a common problem in Florida.

Vanilla Extracts

The adulteration of vanilla extracts with inexpensive synthetic vanillin represents a loss of millions of dollars to the vanilla extract industry. To obtain equal strength of flavor, either 28 g of synthetic vanillin or 4.5 l of single-fold vanilla extract is required. The ratio of cost is approximately 1:120, and this significant difference is a strong stimulus for adulteration.

Nonvanilla vanillin can be prepared from lignin, eugenol, or guaiacol. Synthetic products have a characteristic vanilla-like note but they are inferior in organoleptic properties because the quality of natural vanilla derives not only from the concentration of vanillin but also from the presence of other valuable volatile flavor compounds.

Olive Oil

Olive oil, a food staple in the warmer regions around the Mediterranean, is now becoming more popular throughout Europe, the USA, Japan, and Australia. Olive oil is often adulterated with seed oils and olive residue oil (olive pomace, olive kernel, or olive husk oil). Esterified oil, which is prepared by re-esterifying low-grade olive oil or olive oil soapstock recovered from alkali refining, has also been used to adulterate olive oil products. A survey in 1982 in the USA demonstrated that many olive oil products (imported or locally produced) contained undeclared olive husk oil, esterified oils, and seed oils. The European legislation for olive oil characteristics issued in 1991 (and amended many times since then) is an important tool against sophisticated frauds. The most recently reported cases of virgin olive oil frauds are those where lower-grade olive oil (desterolized or mildly deodorized), or other oils having a similar composition (hazelnut oil, high oleic-sunflower oil) have been used. (*See* **Olive Oil**.)

Fish, Meat and their Products

Meat and fish commercial products, being the main food protein sources for human consumption, are most frequently adulterated with other proteins of inferior-quality fish species or of nonmeat origin (e.g., undeclared soya protein in pork sausages). Authentication of species is important because all meats are not safe for consumers (e.g., meat from cows which had bovine spongiform encephalopathy (BSE)) or meats which should not be used (endangered or overexploited resources). Identification is more complex, especially when the analysis has to be carried out on cooked products. Fish traded in fillets or as fish fingers is the most frequent case of economic adulteration. Surimi, which was used in the past as a substitute for crab sticks, is no longer a cheap raw material because of overexploitation. Characteristic incidences of substitution of cheaper materials for more expensive ones include the fraudulent replacement of salmon, trout, plaice or lemon sole, and caviar. DNA analysis seems to be the answer to speciation problems related to fish trade. (*See* **Fish**: Fish as Food; **Meat**: Eating Quality.)

Milk and Dairy Products

The addition of water is still the most frequent case of milk adulteration. False labeling of the animal origin of milk, the use of reconstituted powdered milk, and the addition of partially hydrogenated vegetable oils are currently a matter of investigation for milk and dairy products. Roquefort in France, manchego in Spain, pecorino in Italy and feta in Greece are cheeses of large economic importance which are frequently adulterated or misbranded. The manufacture of these products is strictly regulated in these countries and within the EU. Consumption of products containing undeclared milk has been associated with certain food allergies. (*See* **Milk**: Liquid Milk for the Consumer; **Cheeses**: Chemistry and Microbiology of Maturation.)

Wheat and Other Cereals

The main cases of adulteration reported and addressed analytically are the admixture of wheat and barley cultivars, the fraud of durum wheat flour with common wheat (mainly in pasta products), and the misbranding of various commercial types (cargo, milled, broken grains, parboiled) or the origin of Basmati rice. (*See* **Cereals**: Bulk Storage of Grain.)

Genetically Modified Food

The rapid pace of development of genetically modified organisms (GMOs), especially food plants, and the commercialization of novel foods derived from GMOs have caused intense debate and consumer concern. Although there is no evidence of actual or potential harm of GM food, nothing can be certain in a field of rapid scientific and technological progress. The present EU legislation (regulations 258/97/EC and 1139/98) recognizes that a genuine choice of non-GM foods should be available to the public and that food containing identifiable GM material should be properly labeled. Against this background numerous questions have emerged concerning analytical aspects, ranging from sampling procedures to robustness of methods. (*See* **Genetically Modified Foods**.)

See also: **Cereals**: Bulk Storage of Grain; **Cheeses**: Chemistry and Microbiology of Maturation; **Contamination of Food**; **Fish**: Fish as Food; **Food Additives**: Safety; **Food and Drug Administration**; **Genetically Modified Foods**; **Legislation**: International Standards; Additives; Contaminants and Adulterants; Codex; **Meat**: Eating Quality; **Milk**: Liquid Milk for the Consumer; **Olive Oil**; **Wines**: Types of Table Wine

Further Reading

Ashurst PR and Dennis MJ (eds) (1996) *Food Authentication*. London: Blackie Academic and Professional.

Barai BK, Nayak RR, Singhal RS and Kulkarni PR (1992) Approaches to the detection of meat adulteration. *Trends in Food Science and Technology* 3: 69–73.

Farrer KT (1997) Dr AH Hassal and food technology. *Food Science and Technology Today* 11: 81–87.

Fennema OR (ed.) (1996) *Food Chemistry*, 3rd edn, pp. 1–15. New York: Marcel Dekker.

Musgrove S (1998) The enforcement of food authenticity issues in the UK: practical problems in the European context. *Food Science and Technology Today* 12: 84–94.

Robinson C (1997) Genetically modified foods and consumer choice. *Trends in Food Science and Technology* 8: 84–88.

Widmer WW, Cancalon PF and Nagy S (1992) Methods for determining the adulteration of citrus juices. *Trends in Food Science and Technology* 3: 278–286.

Wrolstad RE (1991) Ethical issues concerning food adulteration. *Food Technology* 45: 108–114.

Detection

M Tsimidou and D Boskou, Aristotle University of Thessaloniki, Thessaloniki, Greece

Introduction

Methods of detecting food adulteration are based on physical, chemical, biochemical, and other techniques. All these methods, which have replaced the early organoleptic and other empirical tests, are continuously updated because food adulteration is unceasing, and new problems are always arising. A number of analytical procedures are usually available for most cases of adulteration so that the analyst has the flexibility to choose the appropriate one (number of samples to be analyzed, sensitivity required, etc).

Proximate analysis of a foodstuff indicates the extent of abstraction for the main compositional components – moisture, fat, protein, carbohydrates, fibre, and ash. However, it cannot answer problems of authenticity or speciation. Very often, adulteration can be detected from the presence of minor components that occur in the adulterant and not in the food itself. In such cases, trace chemical analysis is necessary to assess the purity. Data banks of authentic values that account for seasonal, regional or other variation may support authenticity testing in some cases (fruit juices).

This article outlines the main applications of analytical techniques that are routinely used in laboratories authorized to check the purity of foods and drinks, and also presents in brief the state of the art in authenticity testing.

Sensory Methods

Sensory methods have not been incorporated yet into legislation for the detection of food adulteration, although routine analysis of a sample always includes the examination of its organoleptic characteristics. International organizations such as the International Standards Organization (ISO), American Society for Testing Materials (ASTM), and others have developed and recommended sensory methods for foods in general and for particular commodities. Such methods are useful in trade (e.g., grading of goods) and the management of resources. Moreover, the US Food and Drug Administration (FDA) has accepted results of sensory panel tests as *prima facie* evidence of product efficacy, and in some cases, legal

decisions in the USA have relied exclusively upon sensory evaluation data.

The limitations concerning the application of sensory methods for the detection of adulteration are fewer now that multivariate statistical techniques are employed. With these methods, sensory data are better evaluated, and more reliable conclusions can be drawn. For example, the results of Free-choice Profiling tests can be analyzed by the Generalized Procrustes method. The latter adjusts for several types of variation in the way each panelist describes samples and then manipulates the data to combine terms that appear to measure the same characteristics. Such an approach has been meaningfully applied to classify coffee, tea, and alcoholic beverages.

Separation Techniques

Gas Chromatography (GC)

GC has proved to be essential for the analysis of lipids and, therefore, is widely used in the detection of adulteration of edible fats. Adulteration of fats is usually detected by GC analysis of fatty acid methyl esters (FAMEs), triacylglycerols, sterols, and their dehydration products.

Fatty acid composition Separation of FAMEs, based on different molecular weights and degrees of unsaturation, is usually achieved on capillary columns. Positional and geometrical isomers of the FAMEs are also analyzed under the same conditions. Standard methods are available for the analysis of FAMEs. Edible fats have characteristic fatty acid compositions and can be grouped into various categories. The percentage values for each fatty acid are compared with those recommended by national and international organizations. An increase in the levels of certain fatty acids indicates the presence of certain adulterants (e.g., lauric acid, (12:0), content >0.1 indicates the presence of lauric oils to sunflower oil margarine). An increase in the *Trans* isomers, naturally found in low levels in animal fats, is an indication of treatments such as hydrogenation.

Triacylglycerol composition Triacylglycerols are separated according to carbon number. This is important for the detection of cocoa butter equivalents (CBEs) in cocoa products. The addition of 5% noncocoa butter apart from milk fat is permitted in many countries, and it is a common practice in the industry. Fincke, and also Padley and Timms, proposed a methodology based on the well-documented linear relationship between C_{50} and C_{54} species for all cocoa butters. All CBEs contain fewer C_{52} triacylglycerols compared with cocoa butter, and this difference causes a significant deviation from the above relationship. Levels of POP, POS, and SOS (P=palmitic, O=oleic, S=stearic acid) for natural cocoa butters of different origins were used to construct a ternary diagram to discriminate them from synthetic ones.

Analysis of sterols and other minor components Identification and determination of sterols by GC are mainly used to reveal the presence of vegetable fats in animal fats or admixtures of vegetable fats. The principle of recommended procedures involves saponification of the sample, isolation of the sterols from the nonsaponifiable matter on thin-layer plates, and GC analysis of sterols or their derivatives. The separation is usually with SE-30, OV-17 or JXR at elevated temperatures (230–240 °C). Qualitative and quantitative differences among sterol fractions are well established for a great variety of vegetable fats. Sterol dehydration products formed during refining have been successfully used to identify cheaper edible oils in virgin olive oil or vegetable oils in chocolate. Various groups of researchers developed excellent procedures for the determination of steroidal hydrocarbons, which are useful to verify the thermal treatment of oils and to demonstrate the presence of desterolized oils in an admixture. A minor component of olive residue oil, erythrodiol, a triterpenic diol, is used as an indication of adulteration of genuine olive oil with olive residue oil. The detection is based on GC of the combined sterol and triterpenic diol fraction after silanization. The presence of wax esters with 40–60 carbon atoms indicates the presence of solvent extracted or refined oils in virgin olive oils.

Recent applications GC–mass spectroscopy (GC–MS, GC–MS–MS) techniques have been proved unequivocal in identifying the presence of undesirable components and in revealing adulteration. For example, a GC–MS method was used to identify the illegal addition of diethylene glycol in wines or ethylene oxide in spices. Improvements in column technology required new sample application techniques. A variety of injection modes compatible with capillary columns have emerged. Headspace sampling techniques are ideal for trapping very volatile components prior to GC analysis. They can also be employed to explore the volatiles of samples that contain large amounts of nonvolatile components. This is the case in fermented beverages such as wine, beer, and distilled liquors. The profile of the aromatic compounds of fermented beverages chromatographed by headspace techniques is used to check misbranding of

these heavily taxed commodities. The test involves a comparative GC analysis of the suspected sample and a standard for the particular brand.

High-performance Liquid Chromatography (HPLC)

HPLC techniques are increasingly being used in routine analysis and incorporated into the legislation of many countries. Today, well-established procedures for many classes of organic compounds or individual food components are available; many of these procedures can be used to check food adulteration.

Carbohydrate composition Carbohydrate analysis by HPLC is usually performed on amino-bonded reversed-phase columns. A 70–80% aqueous acetonitrile solution is suitable to elute low-molecular-weight sugars, isocratically, within 20 min; for the detection of the eluate, refractive index detectors are necessary. The sample injected (10–100 μl) must be free from particulate material; careful extraction and clean-up are, thus, essential. Simultaneous determination of glucose, fructose, and sucrose is careful to detect the illegal addition of inverted sugar in manufactured products. Nutritional claims may also be checked by HPLC profiling of carbohydrates. For example, lactose content extrapolates the fat level of skim milk powder; raffinose and stachyose determination is an indicator of the additon of texturized soya in meat products, etc. Oligosaccharide analysis is carried out by similar elution systems to those used for low-molecular-weight sugars (40–60% aqueous acetonitrile, ~30 min) or by anion exchange chromatography and pulsed amperometric detection (potassium hydroxide-based mobile phase). Molecular-weight carbohydrate analysis is necessary for the control of the addition of undeclared starch syrups in honey, fruit juices, and other valuable natural products. The authenticity of instant coffee, using measurements of total glucose (to detect maltodextrins or caramel), total xylose, and free mannitol (to detect husks and skins) and free fructose (to detect chicory), is assisted by the use of pulsed amperometric detection, which provides an increased sensitivity over the refractive index detection.

Triacylglycerol composition The HPLC analysis of triacylglycerols is achieved on reversed-phase columns, isocratically or by gradient nonaqueous elution. The separation is based on the concept of equivalent carbon number (ECN): ECN=CN−2DB, where CN is the carbon number of the triacylglycerol molecule, and DB is the double bond number. Triacylglycerols with the same equivalent carbon number are referred to in the literature as 'critical pairs'. The complete resolution of critical pairs is not necessary for detection of adulterated oils and fats or fat speciation, as long as the separation is characteristic and reproducible. The International Union of Pure and Applied Chemistry recommends a simple isocratic procedure using acetone–acetonitrile mixtures and refractive index detection to determine the nature and the relative percentages of triacylglycerols in fat. Based on such simple procedures, the EU has proposed a maximum level of 0.5% trilinolein for olive oil. The distribution of saturated (S) and unsaturated (U) fatty acids in the *1, 2, 3* positions of triacylglycerol molecules is used as a basis for differentiating pork from other fats. The addition of pork to beef would increase the SSU:SUS ratio, thus permitting its detection even at low levels (2%).

Amino acid composition The search for an equally powerful technique for amino acid analysis may explain the fact that the first reported application of HPLC to food and beverage analysis was the determination of free amino acids in cocoa beans in 1980. Precolumn derivatization of amino acids with various reagents (phenyl isothiocyanate, dansyl chloride, 2,4-dinitrofluorobenzene or *o*-phthaldialdehyde), reversed-phase columns, gradient elution with mixtures of acetonitrile (or methanol) and a buffer and a suitable detector (UV-Vis or fluorescence) are the usual chromatographic parameters. Ion-exchange chromatography is followed by postcolumn derivatization with the ninhydrin reaction and detection in the visible region. Among the most important applications are the determination of authenticity and type of juices, and the detection of the origin of honeys. The concentration and relative proportions of amino acids are two of the best criteria for identification of the type of juices and for estimation of the content of juice in a beverage or syrup. This evaluation of results is based on the small variability of relative quantities of L-amino acids in juices, which cannot be easily manipulated. Besides, commercial amino acid mixtures are expensive adulterants. A stepwise regression analysis performed on amino acid data produced by HPLC is very useful in solving complex problems of adulteration such as the presence of more than one adulterant, or the addition of acidulants and coloring agents.

Other applications HPLC is a very promising technique for the determination of minor components and metabolites. For example, illegal addition of coloring matter in juices and wine can be detected by HPLC analysis of pigments; diacetyl determination in butter, flavonoids in apple and pear juice, and naringin and prunin in grapefruit are also used to detect aroma

supplementation or the type of fruit. HPLC analysis of minor components usually involves careful extraction and clean-up of the determinant. The widespread application of solid-phase extraction with commercial cartridges facilitates these steps.

Electrophoresis and Related Techniques

In recent years, electrophoretic techniques have been used to reveal adulteration of many food products rich in protein. Among the most important applications reported are the detection of: the addition of soya protein to meat products (polyacrylamide gel electrophoresis (PAGE)); the substitution of cow's milk for goats milk in goat dairy products (PAGE, isoelectric focusing (IEF)); the origin (season, area, and strain) of wheat flour (PAGE) and the partial replacement of wheat flour by other cereal flours (PAGE). Electrophoretic techniques are used almost exclusively for the species identification of fishery products. The large amount of fish species consumed, the close relations among some of them, and the removal of the external features due to processing are the major difficulties in the identification of fish species. The latter is achieved using, as determinants, sarcoplasmic and myofibrillar proteins under well-established standard procedures. Reference samples are considered necessary for identification. Today, preparations of species-specific proteins are commercially available. Information is also available for the characteristic electrophoretic protein patterns of a great number of fish species. PAGE with an anionic detergent (sodium dodecyl sulfate, SDS) and a reducing agent (mercaptoethanol) used to separate subunits by size is of great value for the identification of salmon, rainbow trout or brown trout eggs. Electrofocusing is suitable because as an equilibrium technique, it is easily standardized. It is the preferred procedure for the identification of raw species. Enzymic staining instead of amido black is used with IEF to differentiate some species having similar protein profiles (e.g., mutton from goat meat using phosphogluconate dehydrogenase). Two-dimensional electrophoresis (IEF and SDS–PAGE) is a powerful technique able to resolve 1000 proteins in a mixture. Conjunction of electrophoretic systems with image analysis equipment is a means of quantitative estimation of the various species present in a mixture. Simplified speeded up PAGE procedures (e.g., the Micrograd system) are applied for discrimination of the most relevant wheat varieties. HPLC is sometimes used to complement the PAGE data and *vice versa*.

Capillary zone electrophoresis The potential of capillary zone electrophoresis will be tested in a routine analysis of various analytes in the near future. At the moment, it is not considered as an alternative method for protein speciation.

Spectrometry–Spectroscopy

Ultraviolet-visible techniques are frequently used for the assessment of adulteration of citrus juices, coffee and tea, virgin olive oils, and colored food products. Minerals in foods are almost exclusively determined using atomic absorption and emission spectroscopy. Infrared radiation is the basis of automated equipment for the quality control of milk, grains, or fats. More sophisticated techniques are employed, where necessary (nuclear magnetic resonance, mass spectrometry).

UV-Vis Spectrophotometry

Citrus products The approximation of orange juice and pulpwash content in citrus products is based on equations derived from absorption data of the alcoholic extract in the visible region (465, 443, and 425 nm; carotenes) and the ultraviolet (325, 280, and 254 nm; polyphenols, flavonoids, and ascorbic acid) regions. The ratio $A_{273\,nm}/A_{326\,nm}$ is found to be constant for lemon juice (A = absorbance). Any deviation from this constant value indicates the addition of grapefruit, orange, or apple juices. Luminescence spectrophotometry is also useful for the characterization of mixtures of citrus fruit juices.

Tea and coffee The caffeine content of these highly priced commodities and their products is a significant quality criterion. The method, which is based on measurement of the absorbance at 276 nm, is sensitive enough to test decaffeinated green or roasted coffee samples.

Oils and fats The state of autoxidation is indicated by the specific extinction values at 232 nm (conjugated dienes) and 268 nm (conjugated trienes). The specific extinction at 268 nm is also suitable to check admixing of virgin olive oil with residue or refined olive oil because conjugated trienes are present in the processed oils.

Colored foods Ultraviolet-visible spectra of permitted and nonpermitted pigments are important for the identification of added coloring matter in food. Careful extraction and isolation of the dyes are necessary prior to analysis. Lists for permitted colors (FDA, EU) and spectra for food colors are available for comparison (German Research Society).

Atomic Absorption and Emission Spectroscopy

Atomic absorption has a high sensitivity but does not permit multicomponent analysis. The latter is feasible when inductively coupled plasma–atomic emission spectrometry (ICP) is coupled with MS. ICP–atomic emission spectrometry coupled with a pattern recognition program can provide information for the presence of pulp wash in juices on the basis of the increased levels of calcium, silicon, and sodium. Barium, rubidium, and copper were the most important elements for the geographical differentiation of orange juices using a similar approach. Sample digestion is a crucial step in this sensitive analytical technique. Wines can be classified according to origin and grape cultivar on the basis of simultaneous determination of a considerable number of minerals and multivariate analysis of data. It is worth noting that for different regions, different elements were found useful for the discrimination. It seems that grape varieties selectively take up certain elements from the soil.

Near-infrared Reflectance or Transmission Spectroscopy (NIR/NIT) (0.7–2.5 μm) and mid-infrared Spectroscopy (2.5–15 μm)

NIR, or NIT, spectroscopy has been used in the analysis of a wide array of raw materials and food products. Many of its industrial applications are designed for specific products and production lines. Quality control of products includes determination of moisture, protein, fat, carbohydrates/starch, and ash content. Milk and dairy products (whole, skim, powdered milk, cheese, caseinates, whey concentrates, etc.), meat and products, grains, vegetables, baked goods, and snacks are the most frequently analyzed materials. The combination of NIR technology with multivariate analysis software enables processing of all recorded absorption data.

Mid-IR spectroscopy was extensively used in the dairy industry in the proximate analysis of milk and soon became a standard method (AOAC, 1980). Infrared Milk Analyzers evolved as the tools for producer payment and efficient, rapid, cheap routine quality control (one milk sample per 20 s). Fourier transform IR spectroscopy and attenuated total reflectance technology are powerful potential methods in the field of proximate analysis of food components.

Isotopic Analysis

The isotopic composition of various constituents has been used since the late 1970s as a means of determining the authenticity of foods of plant origin. ^{14}C activity analysis and stable isotope ratio analysis (SIRA), utilizing an isotope ratio mass spectrometer (IRMS) or a gas chromatography combined with IRMS, are the most frequently used techniques. The ratios are usually expressed as 'δ' units, where

$$\delta_{\text{isotope ratio}} = \left(\frac{R \text{ sample} - R \text{ reference standard}}{R \text{ reference standard}} \right) \times 10^3.$$

^{14}C activity measurement The use of ^{14}C analysis in the food industry for determining the origin of ethanol in wines and other fermented beverages dates back to 1952. Since then, ^{14}C-activity measurement has been extended to other natural products, e.g., spices and flavors. Atmospheric carbon dioxide through photosynthesis provides plants with the unstable isotope ^{14}C. The age of a certain plant or plant material is approximated from the measurement of its ^{14}C content. Hence, inexpensive artificial flavorings from petrochemicals, materials free of ^{14}C activity, can be easily detected. The technique is used in the USA to check the purity of champagne from the addition of artificially carbonated drinks. In a similar way, the vintage of a wine can be effectively determined because the ^{14}C activity of agricultural products has been well documented for the last four decades. However, this approach fails in detecting semisynthetic molecules derived from natural precursors or manipulated levels of ^{14}C to resemble expected levels.

Stable isotope ratio analysis Photosynthesis initiates isotopic fractionation of ^{13}C and ^{12}C. Most plants cultivated for food fix carbon dioxide by Calvin synthesis (or the C_3 photosynthetic mode). Exceptions to this rule are sugar cane and corn syrup, which follow the Hatch–Slack (C_4) pathway. C_4 plant material is more abundant in ^{13}C, and thus ^{13}C:^{12}C ratios are higher than in C_3 plants. Therefore, the ^{13}C:^{12}C ratios are used to detect the low-cost sugar cane and high-fructose corn syrups (HFCS) in valuable natural products such as honey and most juices (except for pineapple, which follows the Crassulacian acid metabolism). No other analysis could identify the above adulterants in these products. The Association of Official Analytical Chemists has adopted a method for honey, which involves complete burning of a sample to carbon dioxide and water, purification of carbon dioxide, and measurement of the ^{13}C:^{12}C ratio in an isotope ratio mass spectrometer.

A very important application of the ^{13}C:^{12}C ratio is the detection of synthetic vanillin in vanilla extract. Vanilla plants follow an intermediate biosynthetic path compared with C_3 and C_4 modes, and the extracts have very characteristic ^{13}C:^{12}C ratio limits. In this way, natural vanillin can be differentiated from its synthetic counterpart. Sophisticate fraud involves enrichment of the plant material with ^{13}C to escape

MS detection. The ^{13}C:^{12}C ratio is also useful for ensuring the authenticity of corn oil and can provide a good indication of the presence of corn oil in other oils. An atypical carbon δ value is also indicative of the addition of cane sugar spirit to brandy. The combination of IRMS and site-specific isotopic fractionation (SNIF) (site-specific natural isotopic fractionation)–NMR appears to be useful to overcome technical limitations in certain cases.

Another very important application is the discrimination between natural citrus juices and reconstituted juices (water addition). Plant water, compared with ground- or rain-water, is richer in ^{18}O due to losses in the lighter ^{16}O isotope during leaf evapotranspiration. ^{18}O:^{16}O values are also used to check wine appellation claims as well as the addition of beet sugar syrups prepared with tap-water.

To establish isotopic methods for the identification of origin, it is necessary to construct the natural variability of the different bioelement's stable isotope ratios within the area of interest. Such data banks have been installed for wine-producing areas within the European Union since the beginning of the 1990s.

The distribution of ^{2}H in water is influenced by similar factors as in the case of oxygen isotopes. Deuterium concentrations are used to detect the addition of groundwater to orange juice (fresh or reconstituted) or the addition of beet invert sugar to orange juice. Analytical difficulties due to the exchangeable oxygen-bound hydrogen atoms are overcome through the formation of nitrate esters (the level of carbon-bound ^{2}H is determined).

Site-specific isotopic fractionation–nuclear magnetic resonance spectroscopy (SNIF–NMR) SNIF–NMR is a sophisticated technique for detecting the presence of added sugars in fruit juices and characterizing the origin of wines. It takes into account the nonrandom distribution of deuterium within organic molecules, in particular ethanol molecules. The methodology involves fermentation of the fruit juice to ethanol and deduction of the botanical origin of the sugars from the $^{1}H/^{2}H$ ratio at the methyl and methylene site of the molecule. The methyl site isotope ratio is much lower in ethanol from beet sugar than that from grape or citrus fruit fermentation. Addition of 1% beet ethanol induces a ^{2}H:^{1}H shift of 0.88 p.p.m. The analytical requirements are great (equipment, sample size, data banks), and the sensitivity of the method is extremely low in the case of commercial orange juice blends.

Microscopic Methods

Standard microscopic methods have been developed for the authentication of herbs, spices, flours, and honey. These examinations are simple in principle, but the evaluation of results requires considerable experience of the microscopist. *Herbs and spices* are ground, and a slide is prepared and examined under different power objectives. For comparison, ISO standards describing histological characteristics of a great number of spices and condiments can be used. The potential of microscopic examination of pollen for the recognition of geographical origin of *honey* has been extensively studied. The principle of the method is the identification of pollen from characteristic and preferably indigenous plant species of a region. The visualization of *proteins* in cereal products, cheese, or chocolate is carried out using fluorescence microscopy with specific stains. Microscopy is used as a screening method in the detection of adulteration or misdescription of meat and meat products. Staining or polarization microscopy enhances the potential of this examination.

Biospecific Methods

Enzymatic Methods

Enzymatic methods are applied to check the purity of fruit juices and starch and to characterize various sugars in foodstuffs and formulae.

Fruit juices Among the various parameters required in the evaluation of fruit juice quality, the determination of sugars and organic acids is extremely important. Enzymatic determination of sucrose and invert sugars is more accurate and specific compared with chemical methods. By applying enzymatic techniques, more reliable sugar/invert sugar ratios can be calculated.

Determination of organic acids is also useful in detecting adulteration or checking the type of processing and the raw materials used. The addition of synthetic D,L-malic acid is revealed by a combination of enzymatic determination of L-malic acid (the natural existing isomer) and HPLC. Enzyme kits are also available to catalyse the oxidation of D-malic acid to pyruvate. The concentration of the formed NADH monitored spectrophotometrically depends on the original concentration of D-malic acid. Deviation from the narrow range of the citric/isocitric acid ratio in orange juice indicates adulteration with sugar solutions and citric acid. Other organic acids such as ascorbic, dehydroascorbic, lactic, and formic acids are also easily determined enzymatically.

Starches There is a need for a reliable and reproducible technique to determine starch purity. The latter, for example, is extremely important for the EU

refunds for food-grade starches. A multienzyme regime consisting of α-amylase, amyloglucosidase, and a debranching enzyme, pullalanase, is utilized to overcome the problem of incomplete starch hydrolysis (EC Regulation 4154/87).

Other applications Enzymatic determination of glucose, sucrose, and starch or glucose, maltose, sucrose, and partially hydrolyzed starch is recommended by various bodies for the characterization of sugars added to meat products, sausages, baby foods, etc. These determinations of complex mixtures are readily carried out with commercial multienzyme test kits.

Very sensitive enzymatic methods are also available for the determination of ethanol. Such methods are particularly useful in the quality control of wine, beer, other alcoholic beverages, alcohol-free products, and alcohol-containing chocolate products.

Immunochemical Methods

Immunochemical procedures were initially developed for clinical analysis, but today, they find many applications in food quality control. Immunodiffusion and newer techniques such as the various forms of enzyme-linked immunosorbent assay (ELISA) and radioimmunoassay (RIA) are the techniques most widely used in food analysis.

Speciation of meat and meat products Identification of meat species is important not only for economic but also for religious reasons (e.g., kosher food). The speciation of fresh, unheated frozen or mildly cooked meat is carried out by immunodiffusion and ELISA tests. Immunodiffusion procedures are used as screening techniques because they are influenced by the antigen content, which varies from sample to sample. ELISA tests are more suitable for quantification.

Identification of thoroughly cooked meat species involves tedious steps for both sample (extraction and protein recovery) and sera preparation. Double-diffusion techniques are capable of detecting 5–10% meat species in cooked beef sausages by employing antigens resistant to thermal denaturation (e.g., adrenal antigens). Recently, multiple species-specific thermostable antigens have been used to detect various species with ELISA systems. In these cases, cross-reactivity has been reported for very closely related systems (beef/lamb and turkey/chicken). Combinations of immunoassays and electrofocusing in order to increase sensitivity have also been used to solve similar problems.

Detection of 'foreign' protein in food products Soya protein in various forms–milk, meal and texturized – is added illegally to dairy and meat products to increase the total protein to a predetermined or a standard level. An array of different immunochemical precipitation techniques have been applied to the detection of soya protein in food products. The main limitation for all the techniques is that the immunoreactivity of soya antigens is influenced by the soya variety and the types of thermal process, which, in practice, are unknown. Such techniques are suitable for screening. ELISA techniques are preferred for quantification, whereas RIA techniques, which offer equally accurate results to the former, are expensive for routine use. Indirect hemagglutination and immunoelectrophoresis, though more sensitive and specific, are laborious and expensive procedures and require trained personnel.

Other applications Gliadin, the ethanol-soluble fraction of gluten, can be routinely determined by ELISA tests, but some precautions have to be taken to eliminate the problem of nonspecific binding. These tests are valuable for monitoring the legal limit of cereal proteins, added to foods for economic or technological reasons, or to ensure the absence of gluten in gluten-free products.

Substitution of barley wort for maize or rice as a cheaper source of starch in beer manufacturing is determined by double diffusion and immunoenzyme assays. This replacement is illegal in some countries.

Recent Challenges

Sophisticated types of adulteration, the development of a new generation of ingredients and foods through biotechnology or using alternative means of cultivation as well as the application of new technologies for processing foods raised new questions and challenges. Moreover, the public is aware about hazards related to diet and demands suitable labeling of the nonconventional products so that free choice is exercised at the time of purchasing, and also potential abuse of the technology is controlled. Therefore, in the last decade, a vast array of new applications and new techniques were developed to reveal any possible illegal practices against food purity and nutritional value.

DNA Techniques

The development of methods using species-specific DNA fragments amplified by polymerase chain reaction (PCR) may be successfully applied to detect some cases of sophisticated types of adulteration, even on heavily processed foods. Nucleic acids are thermostable molecules in comparison with the thermosensitive protein molecules so that tests based on them

are much more sensitive than the conventional ELISA tests. Still, due to specialized training requirements, expensive equipment, and access demands to DNA sequence databases, the techniques are not used routinely unless there is no other method adequate to solve the problem.

Following the DNA changes, treatments that induce DNA fragmentation can be identified.

To determine the cultivar, the restriction-fragment length polymorphism (RFLP) concept is applied. Restriction enzymes are used to cut the strands of DNA at specific points where a specific sequence of nucleotides is recognized. The fragments separated by microelectrophoresis are stained or characterized using specific radioactive gene probes. The presence of common wheat in durum wheat is based on a specific DNA sequence on the D-genome, a set of chromosomes found only in common wheat.

DNA-fingerprinting analysis can be used to identify fish species, but extensive databases on the sequence of DNA of many commercially important species are a prerequisite.

Detection of inserted DNA is based on a comparison with the genetically modified and nongenetically modified material. Several controls are necessary to ensure that the PCR reaction is successful. Products such as coarsely ground cereals, flour, soya lecithin, unrefined fats, chocolate, soy sauce, tofu, cakes and other baked products, cornflakes and popcorn, meat and sausages, tomato, and ketchup are among the products that can be tested for illegal labelling or admixture with genetically modified organisms.

Detection of Irradiated Foods

Methods to identify whether or not a food product has been treated with ionizing radiation without requiring comparison with the respective non-treated counterpart are based on different principles, e.g., thermoluminescence, electron spin resonance (ESR) and gas chromatography of irradiation-induced compounds in fats. In 1996, the European Committee for Standardization (CEN) adopted five standards for methods to detect irradiated foods. The standards are applicable to herbs and spices, shrimps and prawns, frog legs, poultry and other meats, liquid egg, fruits, and vegetables.

Thermoluminescence methods (EN 1788) are based on the emission of light when energy trapped in crystalline lattices during irradiation is released by heat. In food products, such lattices are silicate minerals originated from exposure to wind or soil (herbs and spices) or other sources such as the minerals from the sea sand found in the intestine of shrimps. ESR detects stable radiation-specific radicals produced upon irradiation of food, which are trapped in solid and dry components such as bones or cellulose. ESR does not apply equally effectively to all foods. CEN standard 1786 'for food containing bone' was validated for beef, chicken, and trout bones. The favorable detection limit of the method permits the identification of radiation, even in complex cases (e.g., chicken burgers). Radicals from cellulose are also used as a means for identification in some plant products, but only when the signal is positive (CEN 1787). The sensitivity of the method is dose-dependent. Further validation of the procedure is expected for foods containing crystalline sugars and crustaceae. Gas-chromatographic determination of radiolytic hydrocarbons (EN 1784) or 2-alkylcyclobutanones (EN1785) seems very useful in detecting irradiation of products containing even low levels of lipids.

Detection of 'Organic' Foods

No chemical or other method seems potent enough to distinguish foods produced organically from conventional foods. Therefore, an extensive and formalized system of documentation, inspection, and certification, like that which has been applied to EU countries, seems more realistic for the authenticity control.

See also: **Analysis of Food**; **Carbohydrates**: Determination; **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; Combined Chromatography and Mass Spectrometry; **Consumer Protection Legislation in the UK**; **Fatty Acids**: Analysis; **Food and Drug Administration**; **Food Safety**; **Genetically Modified Foods**; **Irradiation of Foods**: Legal and Consumer Aspects; **Legislation**: History; International Standards; Additives; Contaminants and Adulterants; Codex

Further Reading

AOAC International (1995) *Official Methods of Analysis*, 16th edn. Gaithersburg, MD: AOAC International.

Ashurst PR and Dennis MJ (eds) (1996) *Food Authentication*. London: Blackie Academic and Professional.

Gachet E, Martin GG, Vigneau F and Meyer G (1999) Detection of genetically modified organisms (GMOs) by PCR: a brief review of methodologies available. *Trends in Food Science and Technology* 9: 380–388.

Gordon MH (ed.) (1990) *Principles and Applications of Gas Chromatography.* Chichester, UK: Ellis Horwood.

Guillou C, Remaud G and Martin GJ (1991) Application of deuterium NMR and isotopic analysis to the characterization of foods and beverages. *Trends in Food Science and Technology* 2: 85–89.

Hummel-Liljegren H (ed.) (1990) *Lebensmittelrechts-Handbuch*. Munich: CH. Beck's Verlagsbuchhandlung.

Lindeberg J (1996) Capillary electrophoresis in food analysis. *Food Chemistry* 55: 73–94.

Macrae R (ed.) (1988) *HPLC in Food Analysis*, 2nd edn. London: Academic Press.

Masotti P and Zonta F (1999) Food irradiation: an update of legal and analytical aspects. *Italian Journal of Food Science* 11: 305–316.
Nielsen SS (1998) *Food Analysis*, 2nd edn. Gaithersburg, MD: Aspen.
Piggot JR (ed.) (1986) *Statistical Procedures in Food Research*. London: Academic Press.
Simpkins W and Harrison M (1995) The state of art in authenticity testing. *Trends in Food Science and Technology* 6: 321–328.
Singhal RS, Kulkarni PR and Rege DV (1997) *Handbook of Indices of Food Quality and Authenticity*. Lancaster, PA: Technomic.
Van de Voort FR and Ismail AA (1991) Proximate analysis of foods by mid-FTIR spectroscopy. *Trends in Food Science and Technology* 2: 13–17.
Williams PC and Stevenson SG (1990) Near-infrared reflectance analysis: food industry applications. *Trends in Food Science and Technology* 1: 44–48.

AERATED FOODS

G M Campbell, Satake Centre for Grain Process Engineering, Manchester, UK

Background

Many foods, such as bread, ice cream, and whipped cream, gain their distinctive texture and appearance from the presence of bubbles. This article describes the processes used to create aerated foods, considers methods for characterizing aerated foods, and identifies the factors contributing to the stability of aerated foods during and after manufacture. The creation of aerated foods uses a wide range of aeration processes, including low-viscosity whipping, beating, and shaking, high-viscosity dough and paste mixing, gas injection, and slow and rapid gas generation or expansion through fermentation, leavening agents, extrusion, and puffing. A variety of whisks and mixers are used, both in the kitchen and on the commercial scale, to aerate foods. The air content and bubble size distribution depend on mixer design and operation. Air contents of aerated foods can be determined from density measurements, and cover the range from a few percent for some confectionery products to greater than 95% of the total volume of, for example, popcorn or rice cakes. Aerated food systems are inherently unstable, and the bubbles must be stabilized against drainage of the liquid films, coalescence of bubbles, and disproportionation. Surface-active materials such as emulsifiers, lipids, proteins, fat crystals, and ice crystals stabilize bubbles in foods.

Aerated Foods

Most food processing, either domestic or commercial, is concerned with creating desirable, distinctive, or novel textures, along with an attractive appearance. Many of the most appealing foods deliver their characteristic texture and appearance by exploiting the presence of bubbles. Examples include bread, cakes, ice cream, breakfast cereals, meringues, whipped cream, waffles, soufflés, aerated chocolate bars, beer, champagne, popcorn, and many others (*See* **Beers**: History and Types; **Biscuits, Cookies, and Crackers**: Nature of the Products; **Bread**: Breadmaking Processes; **Butter**: The Product and its Manufacture; **Cakes**: Nature of Cakes; **Cereals**: Breakfast Cereals; **Cheeses**: Types of Cheese; **Cream**: Types of Cream; **Eggs**: Use in the Food Industry; **Ice Cream**: Properties and Analysis; **Snack Foods**: Range on the Market; **Soft Drinks**: Chemical Composition). Bubbles in foods offer no nutritional benefit; they represent pure luxury and proclaim the skill of the chef or industrial counterparts. Aeration transforms food from merely flavorsome fuel into textural novelty. The textures achieved vary from the soft but strong crumb of bread to the crispness of breakfast cereals, the crunch of honeycomb, the smoothness of whipped cream, the 'melting bubbles' of aerated chocolate bars, and the tingle of carbonated beverages.

The results of food aeration are:

1. an increase in volume of the product;
2. a change in the texture and rheology of the product, aiding mastication and giving a different mouth feel and appearance;
3. enhanced ability to hydrate rapidly or to take up sauces, due to increased surface area;
4. possibly a modification of digestibility;
5. the possibility of a reduced shelf-life, as the finely divided air bubbles or increased porosity may enhance oxidation reactions and affect fat and flavor ingredients;
6. a decrease in the intensity of flavors.

Aerated foods offer product differentiation and marketing advantage in the highly competitive, innovative, and dynamic food market. They also inspire delight and praise in the dining room. Their creation requires detailed understanding, empirical and fundamental, of the complex interactions between food chemistry and physics in the kitchen or in the manufacturing environment.

Aeration Processes and Equipment

Aerated foods are produced in one of three ways: (1) liquid is forced around air to form bubbles; (2) gas is sparged into liquid to form bubbles; (3) gas is generated within the food to create bubbles. The first of these can be divided into low- and high-viscosity systems, while the third method can be divided into slow and rapid generation of gas, giving a total of five categories of aeration method. Within these five broad categories a very wide range of specific processes and operations are used, including whipping cream, beating eggs, dough and paste mixing, widgets in beer, fermentation, gas injection, frying, vacuum puffing, and extrusion. Table 1 describes the range of aeration methods used within these five categories, and lists examples of foods created using these processes. Several aeration operations may contribute together or consecutively during the process; for example, in breadmaking, bubbles are incorporated into the high-viscosity dough during mixing (a type 1b process); these bubbles are inflated slowly by carbon dioxide gas generated by yeast fermentation (type 3a), and are further inflated by steam generation and thermal expansion during baking (also type 3a), while also undergoing coalescence and rupture along with setting of the matrix structure.

The major food aeration methods are as described below.

Type 1(a)

- Whipping, beating, or shaking of low-medium viscosity liquids to entrap air.
- Pressure beating (dissolution of air or gas under pressure), for example, in a syrup, fat mixture, or chocolate, for confectionery manufacture.

Type 1(b)

- Mixing of doughs or high-viscosity pastes, in which air bubbles are entrapped as surfaces come together. The high viscosity of the dough or paste prevents rising and disengagement of bubbles. In raised bread, the bubbles incorporated by mixing act as nucleation sites for the CO_2 produced during fermentation. During creaming of butter and sugar, sugar crystals aid aeration; fat crystals in cake batters and ice cream act similarly, with the particle

Table 1 The five categories of processing method used to produce aerated food products

Category	*Processing methods*	*Food products*
1 (a) Low-intermediate-viscosity whipping	Beating, whipping, shaking, pressure beating	Whipped cream, milkshakes, batters (for cakes, pikelets, crumpets, Yorkshire puddings), soufflé, beaten egg whites, meringues, marshmallow, chocolate bars, mousses, dairy desserts, frozen desserts, milkshakes, sherbet, frappé, fruit fool
1 (b) High-viscosity mixing or layering	Dough mixing, paste mixing, entrapment, pulling, layering	Bread doughs, biscuit doughs, meat doughs, soft butter, creamed butter and sugar, ice cream, chocolate bars, puff pastry, choux pastry, croissants, vol au vents, fondant, cream fillings, icing, pulled taffy, peanut butter
2 Gas injection	Sparging, gas injection during pressure beating, widgets, pressurized containers, steam injection	Carbonated soft drinks, widgets in beer, cappuccino, instant whipped cream, ice cream, bubble gum, boiled sweets
3 (a) Slow *in situ* generation or expansion of gas	Fermentation, chemical raising agents (baking powder, sodium bicarbonate), steam generation and thermal expansion during slow cooking processes, vacuum expansion	Proving bread dough, baked goods, unleavened bread, doughnuts, waffles, puff pastry, choux pastry, vol au vents, pizza bases, soda bread, cakes, crumpets, crackers, beer, wine, carbonated soft drinks, ginger beer, Swiss cheese, honeycomb, chocolate bars, boiled sweets
3 (b) Rapid *in situ* generation or expansion of gas	Steam generation and thermal expansion during frying, extrusion, rapid dry heating, puffing, vacuum expansion, microwaving, pressure release	Popcorn, popped sorghum, puffed rice, puffed wheat, cornflakes, extruded breakfast cereals, poppadoms, crispbreads, snacks, potato crisps, micronized wheat, coatings, chocolate bars, beer, wine, soft drinks, bubble and squeak

size of the crystals affecting the size and number of bubbles entrained.
- Entrapment of air between sheeted layers, as in pastries and croissants, or between pulled strands, as in pulled taffy and candies.

Type 2

- Gas injection, e.g., air or nitrogen injection in ice cream and sugar confectionery, carbon dioxide injection in soft drinks, nitrous oxide injection in instant whipped cream, steam frothing of cappuccino.

Type 3(a)

- Fermentation, in which aeration is achieved through carbon dioxide production by yeast in bread, beer, and wine, or by *Propionibacterium* in Swiss cheeses.
- Steam generation during slow to moderate cooking, baking, or frying. Steam generation is often accompanied by thermal expansion of the gases already in the bubbles, and by evaporation of other dissolved components (e.g., CO_2, ethanol).
- The use of chemical raising agents such as baking powders in cakes, or sodium bicarbonate in soda bread, honeycomb, or dulce de leche.
- Vacuum expansion, followed by rapid cooling to set the expanded product, e.g., chocolate bars.

Type 3(b)

- Rapid dry heating or toasting of small or thin products to induce blistering or slight puffing.
- Frying in very hot oil, such that internal steam is formed rapidly, causing the product to puff.
- Expansion extrusion, in which superheated product under pressure emerges suddenly from an extruder, such that internal moisture immediately vaporizes into steam bubbles, to produce crisp snacks, cereals, and sugar confectionery.
- Puffing, in which products such as breakfast cereals containing superheated moisture are subjected to a sudden release of pressure.

In addition, special cases exist, such as popcorn, freeze-dried coffee, and candy floss (cotton candy). Despite the wide variety of processes, aeration equipment comprises primarily mixers of various designs, extruders (*See* **Extrusion Cooking:** Principles and Practice), or specialized puffing or expansion equipment. Most other aeration processes are achieved by heating or by chemical raising agents. (*See* **Leavening Agents**.)

Mixers for food aeration include high-speed whisks and beaters with stainless-steel wire assemblies for egg foams, whipped cream, and cake batters, and high- and low-speed heavy-duty mixers for doughs and pastes. Pressure beating delivers greatly accelerated aeration and produces fine foams, with air consumption of up to $1000\,l\,h^{-1}$. Blades can be mounted horizontally within the mixer bowl, or, more usually, vertically from the top or through the base of the bowl. Industrial-scale high-speed whisks operate at speeds of around 200 rpm with a specific power input of around $50\,W\,kg^{-1}$. Low-viscosity mixers must use high speeds to entrain air and break up the bubbles, while in dough mixing bubbles are entrained unavoidably simply through the action of surface renewal during mixing. In both high- and low-viscosity mixing operations the air content and bubble size distribution depend on the balance between entrainment and disentrainment of air, along with bubble break-up and coalescence.

In modern breadmaking processes the bubble structure created in the dough directly affects the baked loaf structure. Some dough mixers use pressure-vacuum mixing, in which the dough is mixed initially under high pressure to provide additional oxygen (which contributes to development of the gluten network), followed by mixing under a partial vacuum to reduce the air content in the dough. Some batch dough mixers, e.g., the Biplex, operate initially at slow speed to blend and hydrate the ingredients, then at high speed to develop the dough.

Continuous dough mixers lack the versatility of batch mixers to modify the bubble structure in the dough and in the resulting bread. Continuous, tubular, pressurized scraped-surface aerator-freezers with a residence time of about 30 s are used in icecream manufacture to give air contents of up to 50% by volume.

Characterization of Aerated Foods

Aerated foods can be characterized in terms of their rate of aeration, air content, bubble size distribution, the resulting texture, and the stability of the aerated structure. Foamability (the ease with which a foam is formed, in terms of rate and air content) and foam stability are usually applied to transient food foams such as beer and wine foams and beaten egg whites. Foamability may be measured by sparging gas or by rapid whipping to determine the time required to form a prescribed volume of foam, while foam stability is characterized by the half-life of the foam, the rate of foam drainage, or the change of conductivity of the foam. Foamability and foam stability are often inversely related – a greater foam volume is achieved

at the expense of a less stable foam. Stability is conferred through a range of stabilization mechanisms, discussed below.

For more solid aerated foods, texture is of greater importance, and rheometers and textural measurements, including sensory evaluation, are applied. Rheometers may be empirical, imitative, or fundamental. Crisp aerated foods can be characterized by calculating the apparent fractal dimension of the jagged stress–strain curve resulting from crushing. Mechanical properties of solid aerated foods depend on the mechanical properties of the matrix, the amount and distribution of the air, and whether the foam is of open or closed gas cells. Closed gas cells occur in aerated chocolate bars, for example, in which each bubble is discrete, while bread has an open-cell or sponge structure, in which the gas cells are interconnected to form a porous network. Mechanical properties of foams, such as Young's modulus, collapse stresses, crushing strength, and fracture toughness can be related to the density of the foam by a power law model:

$$\frac{\sigma}{\sigma_m} \propto \left[\frac{\rho}{\rho_m}\right]^n \tag{1}$$

where σ is a general mechanical property, σ_m is the mechanical property of the matrix material in the foam, and ρ and ρ_m are the density of the foam and of the matrix material, respectively. The exponent n is usually in the range 2–3. For constant matrix properties, mechanical properties vary with foam density raised to the power of n, such that as air content increases, the foam becomes less strong.

Air contents of aerated foods range from 2% for some confectionery products to greater than 95% for popcorn, rice cakes, and beer foam, with every air content in between. The air content of highly aerated fluid fluids such as icecream and whipped cream is characterized in terms of the overrun, OR, calculated as:

$$OR = \frac{\text{Weight of unwhipped product} - \text{weight of whipped product}}{\text{weight of whipped product}} \times 100\% \tag{2}$$

where weights are measured in a container of constant volume. The overrun represents the additional air added. In aerated foods with lower air contents, the void fraction of air as a fraction of the total volume is more usually calculated as:

$$\alpha = \left(1 - \frac{\rho}{\rho_{gf}}\right) \times 100\% = \left(\frac{OR}{OR + 100}\right) \times 100\% \tag{3}$$

where ρ is the density of the product, and ρ_{gf} is the gas-free dough density. Table 2 gives typical values of the density and air content of a range of aerated foods, from which the other parameters describing aeration can be calculated. Figure 1 shows typical air contents and specific volumes of a selection of aerated food products.

Density measurements indicate the gross air content of a food, but not how that air is distributed. The size distribution of bubbles or gas cells determines the texture, appearance, and mass transfer behavior of an aerated food. The size distribution can be determined by image analysis of an aerated surface or of thin slices of the product. When thin slices of a food material are taken, the holes appearing on the slice are, on average, smaller than the bubbles from which they came. The hole size distribution measured is therefore different from the true bubble size distribution, which must be reconstructed using stereological techniques.

Stabilization of Aerated Foods

The lifetimes of aerated food products range from a few seconds for wine foams, to minutes for beer foams and soufflés, hours for whipped cream, days

Table 2 Typical values of density and air content of aerated foods (dependent in all cases on temperature, composition, and processing factors)

Food	*Density (g cm^{-3})*	*Void fraction of air (%)*
Popcorn	< 0.07	> 95
Rice cakes	0.11–0.13	90–92
Puffed rice	0.13–0.17	88–90
Extruded products	0.10–0.33	75–90
Meringue	0.17–0.18	88–90
Beaten egg whites	0.15–0.20	80–85
Baked loaf	0.20–0.35	72–85
Sponge cake	0.25–0.35	70–80
Risen dough	0.25–0.40	68–80
Marshmallow	0.35–0.45	68–75
Cake	0.35–0.40	68–72
Whipped cream	0.40–0.60	40–60
Ice cream – hard	0.54–0.55	50
Cake batter	0.55–0.80	30–50
Aerated chocolate bar	0.70–0.80	40–45
Nougat	0.80–0.90	30–40
Fruit fool	0.75–0.8	25–30
Ice cream – soft	0.78–0.8	28
Milkshake	0.90–0.95	9–13
Micronized wheat	1.15–1.25	7–11
Bread dough (unrisen)	1.15–1.20	4–8
Wheat grains	1.25–1.35	2–3

Adapted from Campbell GM and Mougeot E (1999) Creation and characterisation of aerated food products. *Trends in Food Science and Technology* 10: 283–296, with permission.

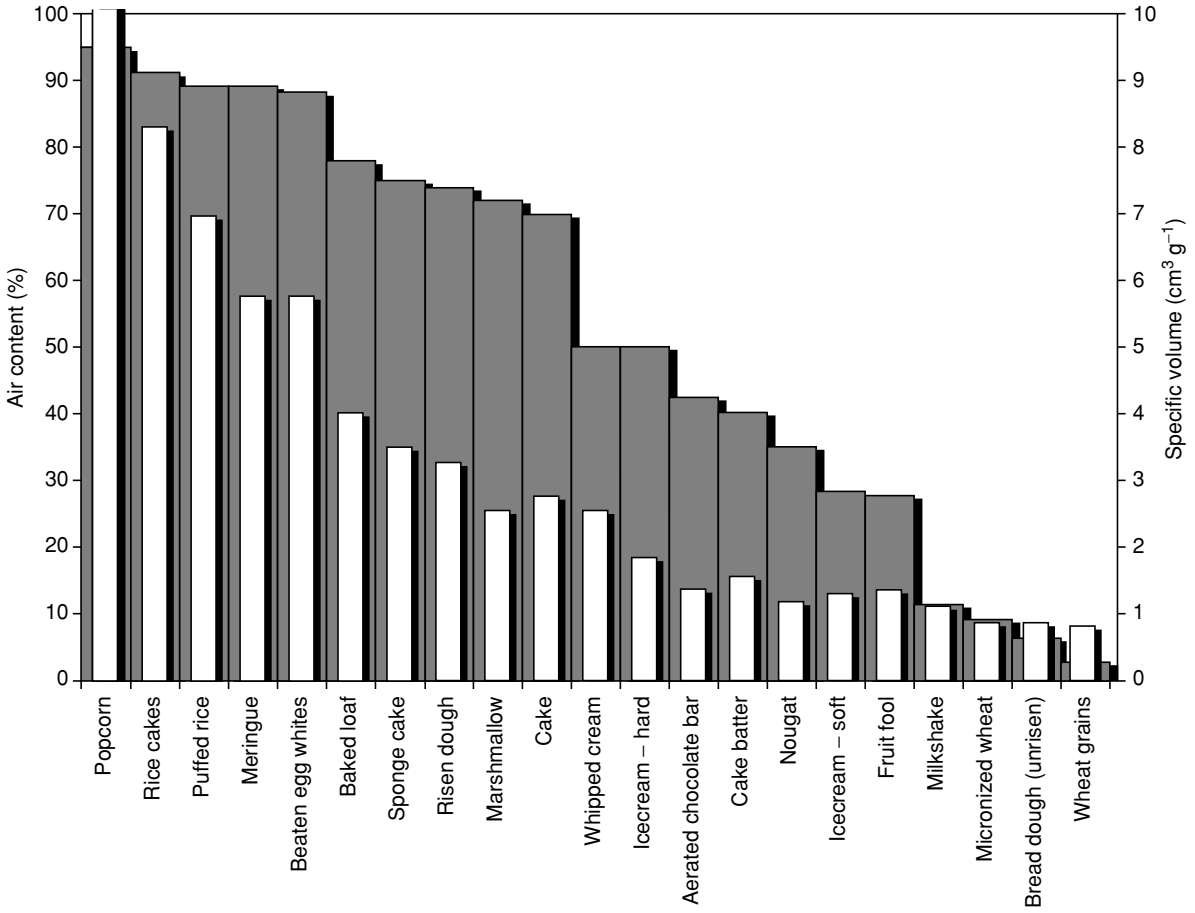

Figure 1 Typical air contents and specific volumes of a selection of aerated foods. Open columns, specific volume; filled columns, air content. Adapted from Campbell GM and Mougeot E (1999) Creation and characterisation of aerated food products. *Trends in Food Science and Technology* 10: 283–296, with permission.

for bread, weeks for cakes, and several months for Swiss cheeses, cereals, chocolate bars, and ice cream. Aerated systems are inherently unstable, and the bubble structure must be stabilized against collapse. In solid aerated foods such as crackers, snacks, and chocolate bars, stability is achieved through the solid matrix. These products are however delicate due to their fine aerated structure, and their high surface area makes them prone to oxidative deterioration or picking up atmospheric moisture; these factors ultimately limit their shelf-life. During the formation of these products, however, and in other more liquid food foams, the bubbles must be stabilized against their thermodynamic tendency to coalesce. Stability is conferred to aerated foods through the action of indigenous or added emulsifiers, proteins (as in beer foams, egg foams, proving bread doughs, cakes), fat or ice crystals or other particles (whipped cream, ice cream, puff pastry), or through a high-viscosity or solid continuous matrix (chocolate bars, meringues, breakfast cereals, Swiss cheese, mousse).

In low-viscosity foams, three distinct mechanisms of destabilization occur: drainage, coalescence, and disproportionation. Drainage occurs when liquid flows from the thin lamellae between bubbles into the Plateau borders that form at the meeting point of three bubbles. Plateau borders form a continuous pathway through the foam, allowing liquid to drain until the bubble lamellae thin sufficiently to allow coalescence. If the foam is stable against coalescence, drainage will continue until a relatively dry polyhedral foam skeleton remains. Higher viscosity in the liquid slows both drainage and coalescence.

Coalescence between bubbles is thermodynamically favorable because of the resulting reduction in surface area. The presence of surface active materials (e.g., emulsifiers and proteins) stabilizes against coalescence; pure liquids, free of surfactants, are unable to produce stable foams. The presence of hydrophilic particles may stabilize a foam, while hydrophobic particles destabilize foams by thinning the film between bubbles, as illustrated in Figure 2.

Disproportionation in foams is the growth of large bubbles at the expense of smaller ones, equivalent to Ostwald ripening in emulsions. The gas pressure in smaller bubbles is greater than that in larger bubbles, due to the contribution of surface tension, γ, as described by the Laplace equation:

$$P_b = P_\infty + \frac{2\gamma}{r} \tag{4}$$

where r is the bubble radius, and P_b and P_∞ are the pressures inside and outside the bubble, respectively. The higher Laplace pressure in smaller bubbles causes a higher gas solubility in the neighboring region. This results in a mass transfer driving force between small and large bubbles that causes diffusion of gas to the latter. The process is self-accelerating, as the loss of gas makes small bubbles even smaller, the Laplace pressure higher, and the gas solubility greater. Ultimately small bubbles implode and the foam coarsens. However, the increasing concentration of surface-active components at the surface of the shrinking bubble slows down and can even arrest the process. Disproportionation is a major factor in foam stability of carbonated beverages such as beer, due to the high solubility of carbon dioxide in water. The presence of a small proportion of nitrogen in the gas phase stabilizes against disproportionation, as the loss of carbon

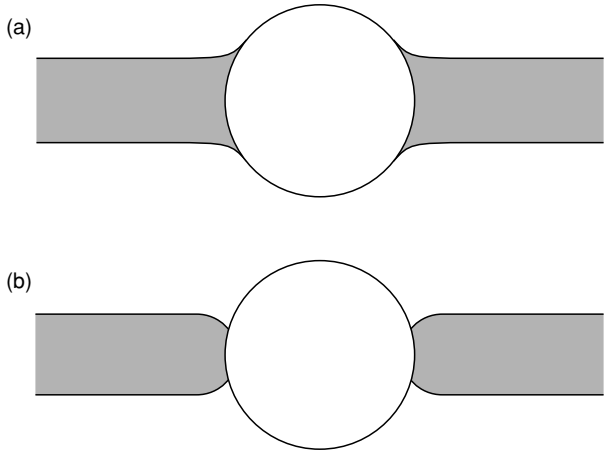

Figure 2 Effect of particles on foam stability: (a) stabilization of bubble lamellae by hydrophilic particles which draw liquid to the particle and oppose film thinning; (b) destabilization by hydrophobic particles.

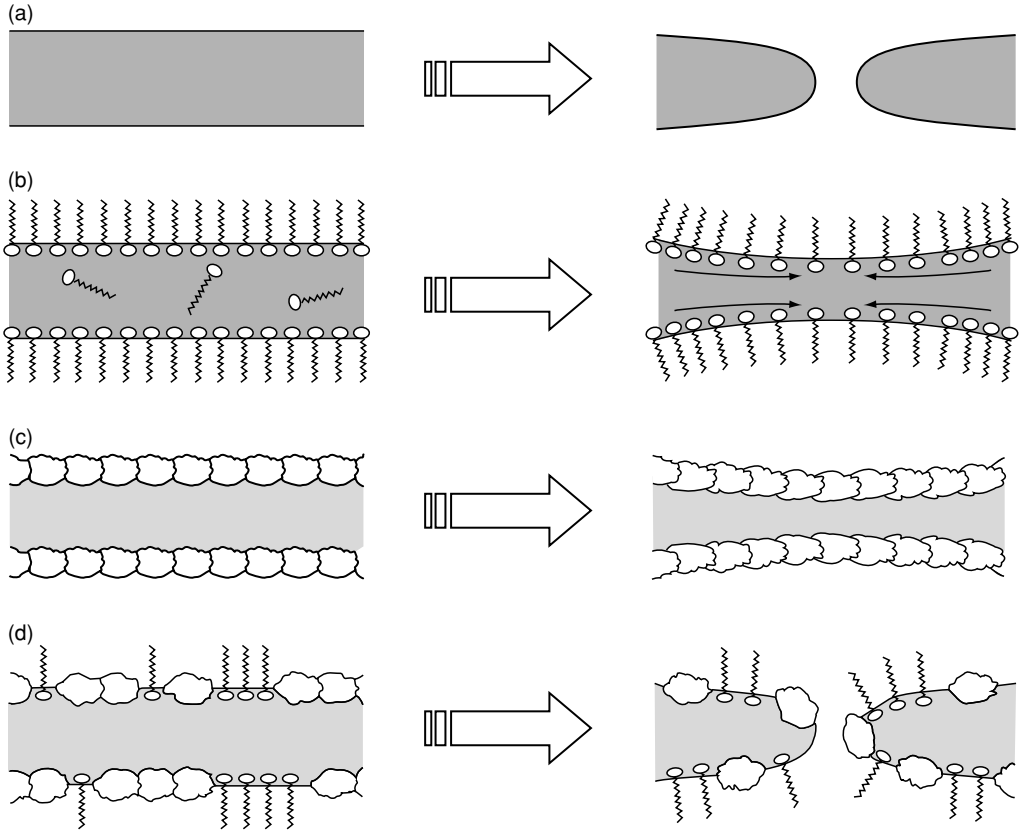

Figure 3 Illustration of film stability between bubbles with (a) no surfactant present; (b) low-molecular-weight surfactants such as lipids; (c) protein-stabilized bubbles; and (d) mixed protein–lipid systems.

dioxide by disproportionation from a small bubble lowers the partial pressure of carbon dioxide remaining in the bubble, lowering its concentration in the surrounding liquid, and thus ultimately removing the driving force for diffusive mass transfer.

Most aerated foods are not low-viscosity foams, and in most aerated foods, bubble coalescence is the most significant destabilization mechanism. Stability against bubble coalescence in both low- and high-viscosity systems is conferred by the presence of surface-active materials such as low-molecular-weight lipids and high-molecular-weight proteins (*See* **Emulsifiers:** Organic Emulsifiers). These two types of surfactant stabilize against coalescence via different mechanisms, as illustrated in Figure 3. Two adjacent bubbles will coalesce when the lamella between them thins and breaks as a result of drainage or mechanical disturbance (Figure 3a). When the lamella thins, it experiences greater stress which rapidly causes further thinning and breakage of the film. If a low-molecular-weight surfactant is present at the surface, its concentration decreases at the point of film thinning. This creates a surface tension gradient which causes surfactant molecules from the adjacent area to diffuse rapidly to the depleted region, sweeping liquid into it and restoring the thickness of the region (the Marangoni effect), and thus safeguarding against coalescence (Figure 3b). The effectiveness of low-molecular-weight surfactants thus depends on their rates of surface lateral diffusion. Large-molecular-weight surface-active molecules such as proteins have low lateral diffusivities; they stabilize against coalescence via a different mechanism, in which they interlink to form a rigid layer at the surface (Figure 3c). If both proteins and low-molecular-weight surfactants such as lipids are present, competing for space at the bubble interface, these two stabilization mechanisms can interfere; the presence of protein prevents rapid surface diffusion of lipid molecules, while the presence of lipids prevents the formation of a rigid interlinked protein network (Figure 3d). Such mixed systems are responsible for the reduction in egg foam volume when a small amount of egg yolk is allowed in with the whites, the reduction in loaf volume when small amounts of polar lipids are added to bread dough formulations, and the disastrous effects of lipids on beer foam stability. In other cases low-molecular-weight lipids may act cooperatively with proteins to improve foam stability.

See also: **Beers**: History and Types; **Biscuits, Cookies, and Crackers**: Nature of the Products; **Bread**: Breadmaking Processes; **Butter**: The Product and its Manufacture; **Cakes**: Nature of Cakes; **Cereals**: Breakfast Cereals; **Cheeses**: Types of Cheese; **Cream**: Types of Cream; **Eggs**: Use in the Food Industry; **Emulsifiers**: Organic Emulsifiers; **Extrusion Cooking**: Principles and Practice; Chemical and Nutritional Changes; **Ice Cream**: Properties and Analysis; **Snack Foods**: Range on the Market; **Soft Drinks**: Chemical Composition

Further Reading

Aguiliera JM and Stanley DW (1999) *Microstructural Principles of Food Processing and Engineering*, 2nd edn. Maryland, USA: Aspen.

Barham P (2001) *The Science of Cooking*. Berlin, Germany: Springer-Verlag.

Blanshard JMV and Mitchell JR (1988) *Food Structure – Its Creation and Evaluation*. London, UK: Butterworths.

Campbell GM and Mougeot E (1999) Creation and characterisation of aerated food products. *Trends in Food Science and Technology* 10: 283–296.

Campbell GM, Webb C, Pandiella SS and Niranjan K (1999) *Bubbles in Food*. St Paul, Minnesota, USA: Eagan Press.

McGee H (1984) *On Food and Cooking: The Science and Lore of the Kitchen*. London, UK: Harper Collins.

Rosenthal AJ (ed.) (1999) *Food Texture*. Maryland, USA: Aspen.

Sadoc JF and Rivier N (1999) *Foams and Emulsions*. The Netherlands: Kluwer Academic.

Stauffer CE (1999) *Emulsifiers*. St Paul, Minnesota, USA: Eagan Press.

Wan PJ, Cavallo JL, Saleeb FZ and McCarthy MJ (1990) *Food Emulsions and Foams: Theory and Practice*. AIChE Symposium Series no. 277, vol. 86. New York, USA: American Institute of Chemical Engineers.

AEROMONAS

S J Walker, Campden & Chorleywood Food Research Association, Gloucestershire, UK

Background

The genus *Aeromonas* belongs to the bacterial family Vibrionaceae and it has been suggested that this genus should be considered to contain some bacterial pathogens for humans, which may cause food poisoning. Other closely related genera within the family are *Vibrio* and *Plesiomonas* species. Both *P. shigelloides* and some *Vibrio* species are known to cause human gastroenteritis. With *Aeromonas* species, however, this pathogenic ability is not yet proven. These bacteria have traditionally been considered to be waterborne, although they may be isolated from a wide range of environments. (*See* **Vibrios**: *Vibrio cholerae*; *Vibrio parahaemolyticus*; *Vibrio vulnificus*.)

Aeromonas Species

Historically, the taxonomy of this genus has been confused, but the genus may be divided into two main groups. The first is the nonmotile species *A. salmonicida*. This microorganism is a pathogen for many species of fish and particularly the salmonids, but is not considered to be of significance in food safety or spoilage. Therefore, this species will not be considered further. The second group is the motile species. A phenotypic classification in 1981 divided this group into the species *A. hydrophila, A. caviae*, and *A. sobria*, which could be distinguished by simple biochemical tests. DNA studies have resulted in the genus being recognized as a separate family, Aeromonadaceae, as part of the Proteobacteria. At least 16 DNA-DNA hybridization groups (equivalent to a genomospecies) are recognized in the genus. The taxonomy of *Aeromonas* continues in this state of flux.

Much of the published literature on *Aeromonas* has not distinguished the motile species or has used out-of-date nomenclature. Consequently, the name *A. hydrophila* will be used as a general name for the motile species in this article. The name is derived from Greek, i.e., *Aeromonas* (a gas-producing unit), *hydrophila* (water-loving).

A. hydrophila as a Pathogen?

The potential for *A. hydrophila* to cause disease has been recognized since 1960. However, the significance of *A. hydrophila* as a direct cause of human gastroenteritis is a matter of debate. Although recognized as an opportunistic pathogen, it is only since the 1980s that there has been increased attention to its clinical significance in healthy individuals.

Some studies have found that the incidence of *A. hydrophila* in diarrheal stools is significantly greater than in those from control groups (e.g., 10.6% and 0.6% of samples positive respectively), and it may often be obtained in high numbers. These observations are most pronounced in infants and young children. The presence of a bacterium in a diarrheal stool, however, does not confirm its role as the causative agent of the syndrome. Other pathogens (including bacteria, viruses, or parasites) may initiate disease and alter the gut environment, and only then may *A. hydrophila* become prominent in the bacterial fecal flora. Much clinical data are of limited use as, in many cases, *A. hydrophila* was not looked for specifically, whilst, in other incidents, the presence of other pathogens could not be ruled out. The confusion in taxonomy and methodology for *A. hydrophila* may also have impaired the abilities of clinical microbiologists to isolate this bacterium from stool samples. A major survey on infectious intestinal disease in the UK found that *A. hydrophila* was as frequently isolated from control cases as from those presenting at general practitioners.

Many strains of *A. hydrophila* possess a wide range of potential virulence factors (i.e., strategies which may aid pathogenicity). These include the production of adhesins, cytotoxic enterotoxins (which lead to fluid secretion in the gut), protease enzymes, and hemolysins (aerolysins), although not all may be present in some isolates. Similar virulence factors are present in many recognized foodborne bacterial pathogens. The virulence factors of *A. hydrophila* produce typical responses in both *in vitro* and *in vivo* model systems for pathogenicity. Human feeding studies with *A. hydrophila*, however, have been inconclusive, with only two of 57 volunteers getting diarrhea after consuming high levels of enterotoxin-producing strains. The validity of this study has been challenged as the strains used may not have possessed all the other virulence factors necessary for gut colonization and so disease.

Consequently, in spite of the abundance of circumstantial evidence for pathogenicity, the case to include *A. hydrophila* as a foodborne pathogen resulting in gastroenteritis has not been conclusively proven.

Symptoms

Human infection with *A. hydrophila* may produce a wide range of localized and systemic illnesses. The symptoms are dependent on the bacterium, host, inoculum size, and infection site. Many of the motile *Aeromonas* species have been implicated in human illness and it is likely that many of them have been mistakenly reported as *A. hydrophila* in the past.

The most common type of infection involves the gastrointestinal tract and this is the illness of most concern to the food industry. Gastroenteritis may occur at all ages and in both sexes, although it is most commonly reported in young children. Symptoms are generally more pronounced in the very young, elderly, and immunocompromised. The most common symptom is an acute profusely watery diarrhea which may last for several days. Other associated symptoms include fever and vomiting. Generally, the disease is self-limiting and symptoms subside within 1 week. In a few cases, the disease may persist for more than 10 days and is classified as chronic. This results in weight loss and may require hospitalization. Chronic disease may last for several months. A second manifestation of gastroenteritis is a 'dysentery-like' illness characterized by diarrhea containing blood and/or mucus and cramping abdominal pain. Hospitalization and antimicrobial therapy are often required. More rarely, a 'cholera-like' illness may occur when the diarrheal stool is turbid and contains white grains ('rice-water stools'). This results in rapid and extensive dehydration, usually requiring hospitalization and aggressive antimicrobial therapy.

Other types of *A. hydrophila* illness include wound infection, septicemia, and localized infections. Wound infections are most common after traumatic injury and in many cases the injury has been associated with water or soil. Infection is most common in the hands, arms, and legs. In most cases with healthy adults, the infection remains localized but, in the immunocompromised, it may become more serious and result in amputation, septicemia, or even death. Septicemia and the infection of a variety of bodily sites occur most commonly, but not exclusively, with immunocompromised individuals. The symptoms of *A. hydrophila* septicemia are similar to those produced by other Gram-negative bacterial pathogens and the mortality level in the immunocompromised is approximately 50%. Both septicemia and localized infections generally require antibiotic therapy and often hospitalization.

Incidents of Foodborne Disease

Although there has not yet been a fully confirmed foodborne outbreak of *A. hydrophila* gastroenteritis, a number of suspected incidents have been reported. Three outbreaks affecting between three and 472 individuals have implicated oysters as the vector of transmission. Levels of the bacterium isolated ranged from 9 per 100 g to 2×10^4 per gram of food, although in the former case, the value obtained was after frozen storage for 1 year.

Cooked prawns have been implicated in three small incidents (1–20 cases) in the UK. Levels obtained in one outbreak were approximately 5×10^6 cells of *A. hydrophila* per gram of food. Several incidents implicating both treated and untreated water have also been reported.

A number of the incidents involving *A. hydrophila* cannot be fully confirmed as the feces of individuals were not specifically tested for this bacterium. Confirmed outbreaks are more likely to be reported with the increased awareness of this bacterium. (*See* **Food Poisoning**: Tracing Origins and Testing.)

Ecology

A. hydrophila may be isolated from a wide range of environmental sources, but has traditionally been considered to be an aquatic bacterium. It has been isolated from fresh waters, estuaries, coastal waters and, occasionally, open ocean waters. The incidence, numbers present, and survival of the bacterium are affected by the pH, temperature, salt content and dissolved oxygen in the water. Several studies have noted that the presence/level of *A. hydrophila* in sources of drinking water does not correlate well with the usual indicators of water quality (e.g., fecal indicator organisms). Furthermore, *A. hydrophila* may be isolated (at low levels) from chlorinated and unchlorinated potable water supplies. A statistical relationship between *A. hydrophila* in water supplies and health risks has been suggested, but there is no firm evidence of serious health effects of humans after exposure to such drinking water. The presence of *A. hydrophila* in potable water supplies may be of concern to food-processing operations, as it may lead to product and/or equipment contamination. Adequate chlorination of the water will insure destruction of the bacterium.

The presence of *A. hydrophila* in water undoubtedly contributes to its common occurrence in fish and seafood. The bacterium is a common component of the intestinal microflora of healthy fish. It is likely that poor processing techniques and poor hygienic practices may result in the contamination of fish flesh and seafood products. (*See* **Fish**: Spoilage of Seafood; **Shellfish**: Contamination and Spoilage of Molluscs and Crustaceans.)

In addition to fish, *A. hydrophila* may be obtained from the feces of a large range of both wild and domestic animals. In these studies, the animals were considered to be healthy. As with fish, poor processing techniques and poor hygienic practices during slaughter, butchery, dressing, and processing of meat and poultry products may contribute to the incidence of this bacterium in raw-flesh foods. (*See* **Poultry**: Chicken; Ducks and Geese; Turkey.)

A. hydrophila is not considered to be a normal inhabitant of the microbial flora of the human gut and its presence is usually transient. The fecal carriage rate for healthy individuals is usually 0–5%. Although asymptomatic carriers could serve as vectors for transmission of this bacterium (e.g., food handlers) no documented evidence of this is available.

As a consequence of this bacterium in human and animal feces, *A. hydrophila* can be commonly isolated from sewage samples and drainage from animal lairages. These may also contribute to the incidence in water samples.

Occurrence in Foods

A. hydrophila may be isolated from a very wide range of foodstuffs. It is not surprising that it is common (up to 100% of samples are positive) in a wide range of seafood, including fish (both wild and farmed), shrimps (raw and cooked), oysters, crabs, and scallops. The bacterium is also common in raw meats and may be isolated from a high proportion (50–100%) of red meat and poultry samples. Cooked meats may occasionally contain *A. hydrophila*, indicating that postprocess contamination has occurred. In fresh (and cooked) meats and poultry this bacterium is generally present in low numbers (less than 100 per gram) but may greatly increase during storage at chill temperatures, particularly when vacuum or modified-atmosphere packaging is used. In the past, *A. hydrophila* has been considered as a spoilage bacterium in these foods. Up to 50% of raw milk samples may contain. *A. hydrophila* and it has also been isolated from pasteurized milk, cream, and some dairy products. The incidence in pasteurized products is much lower than in raw milk and represents postprocess contamination. A wide variety of vegetable products (including parsley, spinach, celery, alfalfa, beansprouts, broccoli, lettuce, and root vegetables) may contain *A. hydrophila* and the incidence is usually between 5 and 40% of positive samples.

In addition to the above foods, *A. hydrophila* has been isolated from drinking-water samples and also from bottled mineral waters, although levels are generally low (less than 10 per ml). Studies have shown that when *A. hydrophila* was inoculated in mineral water, it was able to survive for 100 days.

The high levels of *A. hydrophila* found in some foods have also been associated with spoilage, where it may produce pungent off-odors. In addition the organisms can produce extracellular lipase or protease enzymes which are mildly heat-resistant and still active at temperatures of 0 to $-2\,^{\circ}C$.

Factors Affecting Growth

Of particular concern with the incidence of *A. hydrophila* in foods is the ability of the bacterium to grow at refrigeration temperatures. This bacterium may grow at temperatures as low as $0\,^{\circ}C$. Consequently, refrigerated storage alone cannot be relied upon to inhibit the growth of *A. hydrophila* completely. On occasions, high levels of the bacterium have been isolated from chilled foods which had been stored for prolonged periods. Reducing the storage temperature, however, will reduce the rate of growth of *A. hydrophila* and refrigeration may act in combination with other preservation factors to inhibit growth. The optimum and maximum temperatures for growth are 28–30 and $42\,^{\circ}C$, respectively.

A. hydrophila is not heat-resistant and has a decimal reduction time (*D* value) of between 2.2 and 6.6 min at $48\,^{\circ}C$. Consequently, this bacterium will be readily destroyed by processes such as milk pasteurization and frankfurter processing. It is important to insure that heat processes are carefully controlled to insure microbial destruction.

This bacterium is not a good competitor with many other bacteria present in raw foods, but may reach high levels if the competing bacterial flora has been removed by heating or its growth is retarded. For example, vacuum packaging may prevent the growth of the main spoilage bacteria of raw meats (i.e., *Pseudomonas* species), but *A. hydrophila* may continue to grow and even dominate the bacterial flora. The use of modified atmospheres containing 100% carbon dioxide will inhibit growth of *A. hydrophila*, but a small decrease in this level may permit growth. (*See* **Chilled Storage**: Use of Modified-atmosphere Packaging; Packaging Under Vacuum.)

A. hydrophila is not an acid-tolerant bacterium and at chill temperatures (less than $5\,^{\circ}C$) it will grow only poorly, if at all, at pH values less than 6.0. Therefore, growth would not be expected in even mildly acidic refrigerated products. At lower pH values, the bacterium will slowly die, although this may require in excess of 20 days. When stored at $30\,^{\circ}C$ the minimum pH value for growth has been variously reported to be between 4.0 and 5.0. As with other bacteria, organic acids (e.g., acetic, lactic, and citric acids) which occur

in foods are more effective at inhibiting growth than are mineral acids (e.g., hydrochloric acid).

The maximum level of salt permitting growth depends on the other conditions, but is 6% (w/v) when other conditions are optimal. With refrigerated samples (less than 5 °C), growth is poor when the salt concentration exceeds 2% (w/v).

Irradiation, at the doses proposed for foods (i.e., 10 kGy), would successfully eliminate *A. hydrophila* from foods. The decimal reduction value is reported as 0.14–0.22 kGy. (*See* **Irradiation of Foods**: Basic Principles.) The organism is sensitive to many of the novel technologies being developed, e.g., high-pressure processing.

Little is known about the effect of freezing on *A. hydrophila*, but this process cannot be relied upon to insure the destruction of the bacterium and it has been isolated from a wide variety of frozen foods.

The presence of *A. hydrophila* in chlorinated water suggests that the bacterium is resistant to this biocide. Chlorination can be a very effective means of controlling *A. hydrophila*, but the levels used must be carefully controlled. Other disinfectants, including those commonly used in the food industry, and ultraviolet treatment of water may also be effectively used to control *A. hydrophila*. Overall, *A. hydrophila* is not a hardy bacterium and is readily controlled by many of the processes used in the food industry. Care is needed to insure that postprocess contamination is minimized.

Detection Methods

A. hydrophila is not difficult to grow and produces good growth on many laboratory media. A wide variety of methods are used for this bacterium. Most of the media used have been developed from those used for the detection of Enterobacteriaceae or *Vibrio* species. It is likely that *A. hydrophila* has been misidentified as an enteric bacterium and so its incidence and clinical significance will have been underestimated.

The bacterium may be isolated using quantitative or qualitative procedures. With quantitative methods, the sample is usually inoculated directly on a solid selective agar medium. With qualitative methods, the sample is usually placed in a liquid-selective enrichment medium (to allow multiplication of *A. hydrophila*) prior to inoculation of the agar medium. Incubation temperatures used for both liquid and solid media range between 25 and 37 °C, although 30 °C is being increasingly accepted.

A variety of agar media have been used for food and water samples. Most of these contain selective agents (usually based on ampicillin, bile salts, or ethanol) and differential agents (e.g., starch, dextrin, xylose, amino acids). With liquid media, alkaline peptone water and ampicillin broths are most widely used. Both liquid and solid media containing ampicillin have been shown to be very useful, although a few strains are sensitive to this antibiotic.

At present, there is no general consensus on the best media for *A. hydrophila* and it is likely that the use of multiple solid and/or liquid media will be required for optimal recovery of this bacterium. The individual motile *Aeromonas* species can be separated using conventional biochemical tests. Additional tests to identify biotypes and serotypes are not widely used for food isolates.

Given the complexity of the taxonomy of the genus and that some isolates, even of *A. hydrophila*, are not pathogenic, a variety of molecular techniques have been proposed for the detection and/or identification of this organism in clinical, veterinary, and food situations. These include polymerase chain reaction (PCR) detection of the enterotoxin and hemolysin genes, restriction fragment length polymorphism (RFLP)-PCR to identify the species, and random-amplified polymorphic DNA (RAPD)-PCR probes for detection and identification.

See also: **Chilled Storage**: Use of Modified-atmosphere Packaging; Packaging Under Vacuum; **Dairy Products – Nutritional Contribution**; **Fish**: Spoilage of Seafood; **Gums**: Food Uses; **Irradiation of Foods**: Basic Principles; **Poultry**: Chicken; Ducks and Geese; Turkey; **Shellfish**: Contamination and Spoilage of Molluscs and Crustaceans; **Vibrios**: *Vibrio cholerae*; *Vibrio parahaemolyticus*; *Vibrio vulnificus*

Further Reading

Abeyta C and Wekell MM (1988) Potential sources of *Aeromonas hydrophila*. *Journal of Food Safety* 9: 11–22.

Cahill MM (1990) Virulence factors in motile *Aeromonas* species. *Journal of Applied Bacteriology* 69: 1–16.

Janda JM and Duffey PS (1988) Mesophilic aeromonads in human disease: current taxonomy, laboratory identification, and infectious disease spectrum. *Reviews of Infectious Disease* 10: 980–997.

Kaznowski A (1998) Identification of *Aeromononas* strains of different origin to the genomic species level. *Journal of Applied Microbiology* 84: 423–430.

Palumbo SA and Buchanan RL (1988) Factors affecting growth or survival of *Aeromonas hydrophila* in foods. *Journal of Food Safety* 9: 37–51.

Stelma JR (1989) *Aeromonas hydrophila*. In: Doyle MP (ed.) *Foodborne Bacterial Pathogens*, pp. 1–19. New York: Marcel Dekker.

Tompkins DS, Hudson MJ, Smith HR *et al.* (1999) A study of infectious intestinal disease in England: microbiological findings in cases and controls. *Communicable Disease and Public Health* 2: 108–113.

AFLATOXINS

M O Moss, University of Surrey, Guildford, UK

Background

The aflatoxins are a group of carcinogenic, acutely toxic mold metabolites that may be produced in a range of foods, animal feeds, and the raw materials used in their manufacture. The discovery, physicochemical properties, occurrence, toxicology, control, analysis, and significance will be discussed.

Discovery

A bird food catalog for the year 2000 has a letter to customers advising that adverse climate conditions led to a very poor year for the groundnut crop, resulting in unacceptable levels of aflatoxin contamination, and that there would be limited supplies for bird foods this year. Indeed, this particular supplier withdrew all stocks of peanuts. In 1998, the European Union set a limit of 2 μg kg^{-1} for aflatoxin B_1 in foods for human consumption, a regulation that was included in British legislation in 1999. These two observations make it clear that aflatoxin is an important contaminant, and it is essential to understand its origin, occurrence, and significance.

In 1960, there were reports of large-scale deaths of turkey poults and other poultry in England, which were initially considered to be due to a viral infection and were called turkey X disease. The initial symptoms were a loss of appetite, feeble fluttering, and lethargy, death usually occurring within a week of the appearance of symptoms. Autopsy showed hemorrhages and necroses of the liver. It soon became clear that the birds were poisoned by a contaminant in the pelleted feeds that formed the major part of their food. The toxic component was shown to be associated with groundnut meal contaminated with the green spored mold *Aspergillus flavus*, and it was soon demonstrated that the mold was secreting a toxin that was called aflatoxin to reflect its source. As chromatographic analyses improved, it was shown that aflatoxin is a family of compounds that were labeled according to the color perceived when thin-layer chromatography (TLC) plates were examined under UV light; aflatoxins B_1, B_2 (blue), G_1, and G_2 (green), of which aflatoxin B_1 is the most toxic. Chronic toxicity studies in rats demonstrated that, for this species, aflatoxin was amongst the most carcinogenic compounds known at the time, and the study of aflatoxin became truly international, generating an enormous literature very rapidly.

The chemical structures were illucidated within a few years of the initial outbreak of turkey X disease (Figure 1), and, with an increased understanding of their occurrence and properties, it was appreciated that at least two animal diseases described earlier were due to the presence of aflatoxins. A serious liver disorder of dogs known as canine hepatitis and the presence of liver carcinomas in rainbow trout in fish farms, where protein rich feeds based on slaughter house offal were replaced with plant proteins based on cotton seed meal. Indeed, it is certain that aflatoxins, either alone or in the presence of other mycotoxins, have caused many acute and chronic illnesses in farm animals and several in humans. One of the best documented outbreaks of acute aflatoxicosis in humans occurred in India in 1975. Nearly 1000 people were ill, and almost 100 died following the consumption of contaminated corn.

Physicochemical Properties

The aflatoxins are white crystalline solids that are optically active and have a strong absorbance at about 365 nm with a fluorescence emission of 415–450 nm, depending on the solvent or physical status (Table 1). Fluorescence is especially strong when the aflatoxins are absorbed on silica gel, making this a

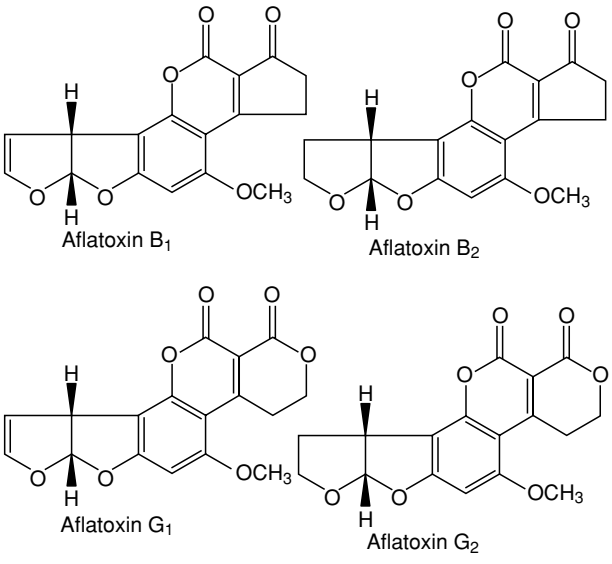

Figure 1 Structures of the four most common naturally occurring aflatoxins.

Table 1 Some physico-chemical properties of the aflatoxins

Aflatoxin	*Formula*	*Molecular weight*	*Melting point (°C)*	*Optical rotation $[\alpha]_D(CHCl_3)$*	*λ_{max} (nm) ε*
B_1	$C_{17}H_{12}O_6$	312.063	268–269	−558°	223 25 600 265 13 400 362 21 800
B_2	$C_{17}H_{14}O_6$	314.0790	287–289	−430°	220 20 500 265 12 700 363 24 000
G_1	$C_{17}H_{12}O_7$	328.0582	247–250	−556°	243 11 500 257 9 900 264 10 000 362 16 100
G_2	$C_{17}H_{14}O_7$	330.0739	230	−454°	217 28 000 245 12 900 265 11 200 365 19 300
M_1	$C_{17}H_{12}O_7$	328.0582	299	−280°	226 23 100 265 11 600 357 19 000

sensitive means of detection on thin-layer chromatograms. The aflatoxins are soluble in a range of organic solvents such as chloroform, ethanol, methanol, and acetone, and insoluble in lipophylic solvents such as hexane, petroleum ether, and diethyl ether.

Occurrence

Because so many compounds have similar chromatographic and fluorescent properties to aflatoxins, there were many early reports of their production by molds from several genera including *Penicillium* and *Rhizopus*. However, the only confirmed sources of aflatoxins are a small group of closely related aspergilli, namely *Aspergillus flavus, A. parasiticus, A. nomius*, and *A. ochraceoroseus*. There are also reports that an isolate of *A. tamarii* produces aflatoxin B_1, although no other isolates of this species have been shown to produce aflatoxins. This isolate has subsequently been referred to a new species, *A. pseudotamarii*. By far the most commonly isolated aflatoxigenic species are *A. flavus*, which produces aflatoxins B_1 and B_2, and *A. parasiticus*, which produces all four aflatoxins shown in Figure 1. On a world-wide basis, perhaps 35% of *A. flavus* isolates are aflatoxigenic, whereas almost all strains of *A. parasiticus* produce aflatoxins.

The most commonly implicated commodities are corn and peanuts, although aflatoxins have been detected in a much wider range of foods. Table 2 provides the results of surveys carried out during the last few years in a number of different countries.

The highest levels of aflatoxin contamination are always associated with postharvest spoilage, when commodities are stored with an inappropriate moisture content and temperature. However, aflatoxin contamination is not simply a problem of poor storage but can occur in the field before the crop is harvested. The spores of these species of *Aspergillus* can land on the stigma of the developing plant, germinate and penetrate to the immature seed tissue just as if they were pollen grains. The mold can establish an endophytic growth within the tissues of the plant without causing any perceptible harm to the plant. Indeed, if the crop continues to develop in a healthy manner and is harvested and correctly stored, there may be no detectable aflatoxins produced, despite the presence of viable mycelium in the plant tissue. However, if the plant is subject to some form of stress, and drought stress is the most common, then low, but significant, levels of aflatoxin

Table 2 Reports of the occurrence of aflatoxins during 1991–1998

Commodity	*Country*	*Year*	*Incidence (%)*	*Range ($\mu g\,kg^{-1}$)*
Almonds	USA	1993	1	tr–372
Brazil nuts	USA	1993	17	tr–619
Chillies	Pakistan	1995	66	1–79.9
Corn (maize)	Argentina	1996	20	5–560
Corn (maize)	India	1997	47	5–666
Dried figs	Austria	1993	13	2–350
Herbs and spices	UK	1996	24	1–51
Nutmeg	Japan	1993	43	0.2–16.6
Peanuts	Brazil	1998	51	43–1099
Peanuts	India	1996	45	5–833
Pistachio nuts	Netherlands	1996	59	2–165
Rice	Equador	1997	9	6.8–40
Soybeans	Argentina	1991	10	1–36
Spices	Sweden	1998	90	0.1–62
Wheat	Uruguay	1996	20	2–20

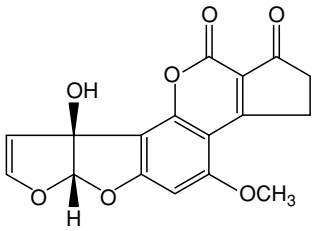

Aflatoxin M_1

Figure 2 Structure of aflatoxin M_1.

may be formed in the seed tissue, even though, on harvest, the crop looks sound and is subsequently correctly stored.

The molds producing aflatoxins are tropical and subtropical, and, although their spores occur in temperate parts of the world, these species do not compete in a cool climate. Thus, the production of aflatoxins is not normally a problem in Britain and Scandinavia. However, in both these parts of the world, a special problem emerged when high-moisture barley was prepared as a winter feed by protecting it with propionic acid, or mixtures of acetic and propionic acids. When properly applied, these antifungal agents protect the material from mold spoilage, but if the dosing is not high enough, *Aspergillus flavus* is able to grow and, under these conditions of partial inhibition, actually produces enhanced levels of aflatoxin.

Aflatoxin B_1 is readily metabolized in the animal body and may be secreted in a number of body fluids. Thus, when cows are fed on animal feed contaminated with aflatoxin B_1, a significant proportion is secreted in the milk as aflatoxin M_1 (Figure 2), which is only slightly less toxic than B_1. Aflatoxin M_1 has also been detected in human milk, and because milk is consumed in relatively large quantities by both the very young and elderly, there have been many surveys to provide some guidance about exposure of the human population in different parts of the world. Table 3 gives some results from such surveys.

Table 3 Selection of surveys for aflatoxin M_1 in milk

Type of milk	*Country*	*Date*	*Incidence (%)*	*Concentration range (ng l^{-1})*
All milks	Thailand	1997	93	50– > 500
Milk powder	UK	1981–1983	13	10–400
Milk powder	Sweden	1985	100	6–57
Pasteurized milk	Greece	1997	89	1–177
Raw milk	India	1995	18	100–3500
Raw milk	India	1997	11	100–1000
Raw milk	Poland	1997	23	3–25
Raw milk	Equador	1997	74	125–6000
Raw milk	UK	1980	31	30–520
Raw milk	UK	1996	16	10–90
UHT milk	Spain	1995	14	10–40

Table 4 Oral LD_{50} values for aflatoxin B_1

Animal species	*LD_{50} (mg per kilogram of body weight)*
Rabbit	0.3
Duckling (1 day old)	0.36
Dog	0.5–1.0
Cat	0.6
Pig	0.6
Baboon	2.0
(Humans	5.0?)
Rat (male)	5.5
Rat (female)	17.9
Macaque monkey	7.8
Mouse	9.0
Hamster	10.2
Chicken embryo	0.025 μg per egg

Table 5 TD_{50} values for carcinogenesis of aflatoxin B_1

Animal species	*TD_{50} (μg per kilogram of body weight per day)*
Fisher rat	1.3 (male); 7.5 (female)
Wistar rat	5.8 (male); 6.9 (female)
Porton rat	3.1 (male); 12.5 (female)
(Human	132?)
Rhesus monkey	156
Cynomolgus monkey	848
Swiss mouse (male)	> 5300

Toxicology

The aflatoxins show both acute and chronic toxicity, and one of the outstanding features in the toxicology of the aflatoxins is the wide variation in response amongst different species of animals and even between the male and female of the same species. Table 4 shows the variation in LD_{50}, which is a measure of acute toxicity, and Table 5 shows the variation in TD_{50} as a measure of carcinogenicity. It can be seen that for some animals, such as the rat, aflatoxins are very carcinogenic, and yet in other species, it is difficult to demonstrate carcinogenicity. This considerable variation in biological response arises from the requirement that the mold metabolite itself has to be metabolized in order that a toxic response occurs, and the metabolites responsible for acute toxicity differ from those reponsible for carcinogenicity. There are several metabolic reactions (Figure 3), such as the demethylation to aflatoxin P_1 and hydration to aflatoxin B_{2a}, which may lead to a decrease in toxicity. A critical reaction associated with the toxic response is

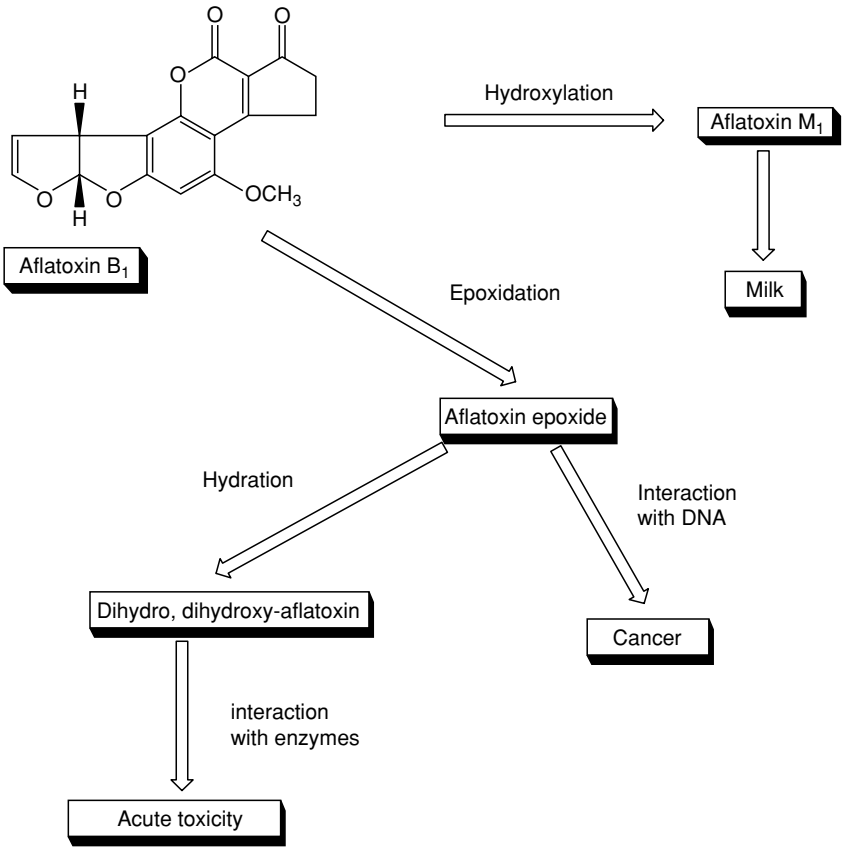

Figure 3 Some metabolic products from aflatoxin B_1.

the formation of the 8,9-epoxide, which may react with guanine residues in DNA, one of these being the third base in codon 249 of the p53 gene, the product of which is an important component in the tumor suppressor system. The epoxide itself can be further metabolized to the 8,9-dihydrodihydroxy derivative of aflatoxin, which can react with the lysine residues in protein molecules and is a good candidate for the acute toxin. An animal that readily forms the epoxide but does not efficiently metabolize this further may be especially sensitive to the carcinogenic activity of aflatoxin. However, an animal that forms the epoxide but effectively metabolizes this further to the dihydroxy derivative may be especially sensitive to the acute toxicity of aflatoxin. The structure of aflatoxin B_1 (which may not itself be toxic) may facilitate its ready transport into the liver cell, where it is metabolized to these far more hazardous molecules.

Where do humans fit into this picture? By extrapolation from both molecular biological studies, and the epidemiology of cases of acute aflatoxin poisoning reported, it would seem that humans are not as sensitive as birds, dogs, or cats but more sensitive than the rat to the acute toxicity of aflatoxin B_1. Similarly, an analysis of exposure versus liver cancer incidence again puts humans in the middle of the range of sensitivity to the carcinogenic activity of aflatoxin B_1. In the early days of the assessment of the role of aflatoxin in the epidemiology of liver cancer in humans, the issue was clouded by the undoubted role of other agents, such as infection with hepatitis B virus, in the initiation of liver cancer. It is probable that these two agents act synergistically, and it is possible to construct a risk assessment model to include both. Such a risk assessment has led to the suggestion that an estimated intake level of between 200 and 400 ng per day would correspond to a lifetime liver cancer excess risk of 1×10^{-5} in the USA. It is clear that it is prudent to assume that aflatoxins are carcinogenic to humans and that efforts should be made to control its occurrence in the food chain.

Control

The highest concentrations of aflatoxins are formed as a result of inappropriate storage conditions post-harvest, and if commodities can be rapidly dried and

stored so that the water activity does not exceed 0.78, growth of aflatoxigenic molds will be inhibited. Aflatoxin biosynthesis itself is inhibited at water activities below 0.83. There is a strong interaction between water activity and temperature in the production of aflatoxins, and if the temperature of storage can be maintained below 15 °C, the minimum water activity for toxin production may be as high as 0.9. Any kind of damage to the commodity can enhance aflatoxin formation.

Lower but significant levels of aflatoxin can be formed before harvest if the crop is subjected to drought stress, and the soil temperature (in the case of peanuts) is between 25 and 32 °C during the 6 or 7 weeks before harvest. Table 6 shows examples of preharvest contamination of corn in North Carolina during the period 1976–1980. It can be seen that during 1980, approximately 65% of the crop was contaminated by at least 20 μg per kilogram of aflatoxins and so could not be used for human consumption leading to an estimated loss of $31 million to the producers. The real cost was even higher if the problems for the animal farming industry were also taken into account.

Because it seemed almost inevitable that contamination will occur in some years, there is interest in the possibility of some form of biological control of preharvest contamination by inoculating the soil with nonaflatoxigenic strains of *Aspergillus flavus*. Greenhouse and field experiments do indicate that such nontoxigenic strains can outcompete or displace resident toxigenic strains leading to reduced preharvest aflatoxin contamination. Alternatively, strains of commodities such as corn and peanuts can be bred, in which aflatoxin biosynthesis is inhibited, even if these aflatoxin producing molds have established an endophytic relationship with the plant.

A number of physical and chemical methods have been tested for the removal of aflatoxins from contaminated materials. Although pure aflatoxins are relatively unstable in the presence of water, these compounds are remarkably stable at elevated temperatures when they are present in the complex matrix of a food. Cooking is thus not effective in removing aflatoxins. In particulate commodities such as peanuts, a significant reduction in contamination levels can be achieved by the identification and removal of damaged kernals. Because contaminated kernals are often (but not always) damaged or discolored, they may be removed by sieving, flotation, and density segregation, or by electronic sorting using a laser beam to identify the discolored kernals. Of several chemical methods tested, the only one that has shown any promise is the use of aqueous or gaseous ammonia at moderate temperatures and elevated pressures. This treatment leads to the production of several products, two of which have been characterized (Figure 4) and shown to be considerably less toxic than aflatoxin B_1.

Many countries have established legislation setting maximum permissible levels of aflatoxins in food and animal feeds, and a selection of these are shown in Table 7. In some countries, there is zero tolerance, which will be based on the sensitivity of analytical

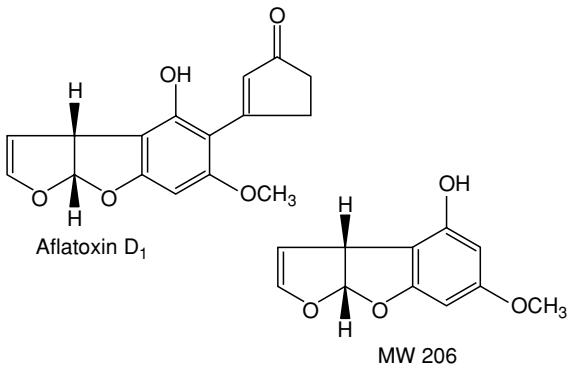

Figure 4 Breakdown products from ammonia treatment of aflatoxin B_1.

Table 6 Preharvest contamination in corn by aflatoxin in North Carolina

Crop year	*Aflatoxin concentration (μg kg^{-1})*		
	0–19	*20–100*	*> 100*
1976	64.2	27.7	8.0
1977	58.1	30.2	11.6
1978	87.0	12.0	1.0
1979	67.3	28.3	4.4
1980	34.3	48.1	17.6

Table 7 Selection of maximum tolerated levels of aflatoxin B_1 in foods

Country	*Maximum level (μg kg^{-1})*	*Products*
Argentina	0	Groundnuts, maize, and products
Brazil	15	All foodstuffs
China	10	Rice and edible oils
Hungary	5	All foods
India	30	All foods
Japan	10	All foods
Nigeria	20	All foods
Poland	0	All foods
South Africa	5	All foods
UK	2	Groundnuts, nuts, dried fruits, and cereals
USA	20	All foods

methods. When the FAO/WHO realized that a valuable plant protein such as groundnut meal, used in the formulation of foods for their food aid programme, could be contaminated with a potential carcinogen soon after the discovery of aflatoxin, they set a maximum concentration of 30 μg kg^{-1}. It was recognized that, in the absence of detailed knowledge of the risk, this was a compromise that recognized that a more stringent level could result in there being no material available for a food aid program. India still accepts this level. The European Union adopted a maximum acceptable level of 2 μg of aflatoxin B_1 per kilogram for groundnuts, nuts, dried fruit, cereals, and their products, and 0.05 μg of aflatoxin M_1 per liter in milk in 1998.

Analysis

The setting of stringent maximum tolerated levels implies the ability to obtain analytical results that are sensitive, specific, quantitative, and reproducible. In the early days of concern over the presence of aflatoxin in foods and animal feeds, there was not sufficient confidence in the chemical analysis of aflatoxins, and biological tests, such as the 1-day-old duckling test and the chicken embryo test, were used because they provided direct evidence of the toxicity of samples. With increasing confidence in the use of confirmatory tests in combination with TLC and high-pressure liquid chromatography (HPLC), these became increasingly acceptable. Most recently, highly specific monoclonal antibodies have become commercially available for the aflatoxins and several other groups of mycotoxins. These can be used in two ways, either directly as enzyme-linked immunosorbent assays (ELISA) or bound to a solid substrate as immunoaffinity columns. ELISA is not generally sufficiently quantitative at very low concentrations but can be very effectively used as a rapid screen for positive samples, which can then be analyzed by physicochemical methods. Immunoaffinity columns provide an excellent way of obtaining a very clean concentrated extract from a sample, which can then be quantified by HPLC.

In many parts of the world, TLC has to be the method of choice because of cost constraints. Using silica-gel plates, the four common aflatoxins can be separated and are sensitively detected because of their intense fluorescence under long-wave ultraviolet. Because so many compounds run with similar R_f values (position on the chromotogram relative to the solvent front both measured from the point of application of the sample) and fluoresce like aflatoxins, it is essential that positive samples are confirmed using a confirmatory test such as derivatization with an acid treatment. Aflatoxins B_1 and G_1 do not fluoresce so intensely in solution, thus reducing the sensitivity of the fluorescence detector in HPLC. However, postcolumn reaction with either bromine or iodine enhances fluorescence considerably and provides a confirmatory test as well.

There have been several collaborative studies in Europe under the Measurements and Testing Programme to validate methods to meet the stringent analytical requirements of European regulations. Such a collaborative study has, for example, evaluated the efficiency of using immunoaffinity columns for the clean-up of samples obtained from a wide range of food matrices for subsequent determination of aflatoxin B_1 and total aflatoxins by liquid chromatography. The European Union legislation for aflatoxin M_1 is especially stringent (0.05 μg l^{-1}), requiring both the validation of methods and the proficiency testing of a network of European Union National Reference laboratories set up to determine aflatoxin M_1 in milk.

Analysis contains three stages, sampling, extraction, and quantitation, and a competent analyst can now achieve the required degree of sensitivity, precision, and specificity required. The analyst needs only 50–100 g of material, and the real problems arise from a consideration of how that sample was obtained from, for example, a shipment of many tonnes of commodity. Except in liquid foods, such as milk, aflatoxins are not uniformly distributed in a commodity, and this was beautifully demonstrated when the UK Ministry of Agriculture, Fisheries and Food (MAFF) purchased a consignment of whole dried figs that had been rejected at the port of entry. The consignment, of just over 10 tonnes made up of 850 boxes each containing 12 kg, had been rejected because an analysis based on a 20-kg sample had indicated a contamination level of 33 μg kg^{-1}. Clearly, with a statutory limit of 4 μg kg^{-1} at that time, the UK had to reject the consignment.

Having purchased the consignment, MAFF analyzed 200 boxes as though each were a single sample. This gave a mean of 15.4 μg kg^{-1}, which still implies rejection but clearly shows that the answer depends on the method of sampling. A study of the distribution amongst these 200 boxes showed that more than 70 had less than 4 μg kg^{-1} and could have been accepted if only they could have been identified! A further 50 boxes had between 4 and 10 μg kg^{-1} and could have been accepted subject to further processing. Thus, about 67% of these 200 boxes were actually acceptable, but a few had contamination levels of > 200 μg kg^{-1}. Indeed, in a further study, individual boxes were divided into 12 × 1-kg samples and each

analyzed separately. In one of these boxes, 11 of the samples had $< 10\,\mu g\,kg^{-1}$, but a single sample had $2063\,\mu g\,kg^{-1}$! There is nothing unexpected in these results, and there are studies on other commodities such as peanuts and pistachio nuts that also demonstrate such a skewed distribution, but they provide quantitative information on which to design a sampling plan. Such a sampling plan has to be acceptable to both the producer and the consumer.

An example of a sampling plan agreed by the European Union for the analysis of figs and other dried fruit is outlined below:

1. Subdivide each lot into units of 25 tonnes (15–30 tonnes).
2. Each unit is sampled separately.
3. The number of incremental samples should be at least 100.
4. Each incremental sample should be about 300 g.
5. The aggregated sample is 30 kg.
6. The aggregated sample is divided into samples of 10 kg, each of which is homogenized and analyzed separately.

Significance

The presence of aflatoxins in foods and animal feeds is quite widespread, and it is possible that, with present agricultural practice, it may not be possible to prevent contamination in some parts of the world. The aflatoxins are acute toxins, and they have undoubtedly caused the death of farm animals and humans on several occasions. They also have chronic toxicity being immunosuppressive and carcinogenic. For these reasons, many countries have set legislative maximum levels in human foods. Some countries also set maximum acceptable levels in animal feeds in order to prevent the contamination of milk by aflatoxin M passing through the food chain. There is as yet no international agreement on these levels, and those countries that are importers of commodities that are susceptible to contamination tend to be more stringent than those that are producers and exporters. There is the potential for difficulties in international trade while this situation remains.

See also: **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; **Mycotoxins**: Toxicology

Further Reading

Anonymous (2002) *Evaluation of Certain Mycotoxins in Food*. 56th report of the Joint FAO/WHO Expert Committee on Food Additives. WHO Technical Report Series 906. Geneva: World Health Organization.

Betina V (1989) *Mycotoxins; Chemical, Biological and Environmental Aspects*. Amsterdam: Elsevier.

Bruce RD (1990) Risk assessment for aflatoxin. II. Implications of human epidemiology data. *Risk Analysis* 10: 561–569.

Chelkowski J (ed.) (1991) *Cereal Grain; Mycotoxins, Fungi and Quality in Drying and Storage*. Amsterdam: Elsevier.

De Vries JW, Trucksess MW and Jackson LS (2002) *Mycotoxins and Food Safety*. New York: Kluwer Academic.

Dragacci S, Grosso F and Gilbert J (2001) Immunoaffinity column cleanup with liquid chromatography for determination of aflatoxin M_1 in liquid milk: Collaborative study. *Journal of AOAC International* 84: 437–443.

Dragacci S, Grosso F, Pfauwathel-Marchond N *et al*. (2001) Proficiency testing for the evaluation of the ability of European Union–National Reference laboratories to determine aflatoxin M_1 in milk at levels corresponding to the new European Union legislation. *Food Additives and Contaminants* 18: 405–415.

Krogh P (ed.) (1987) *Mycotoxins in Food*. London: Academic Press.

Le Bars J and Galtier P (eds) (1998) Mycotox 98; Mycotoxins in Food Chain. *Revue de Médecine Vétérinaire* 149: 469–715.

Moreau C (1979) *Moulds, Toxins and Food*. Chichester, UK: Wiley.

Powell KA, Renwick A and Peberdy JF (eds) (1994) *The Genus* Aspergillus *from Taxonomy and Genetics to Industrial Application*. New York: Plenum Press.

Sinha KS and Bhatnagar D (eds) (1998) *Mycotoxins in Agriculture and Food Safety*. New York: Marcel Dekker.

Smith JE and Henderson RS (eds) (1991) *Mycotoxins and Animal Foods*. Boca Raton, FL: CRC Press.

Stroka J, Anklam E, Jorissen U and Gilbert J (2000) Immunoaffinity column cleanup with liquid chromatography using post-column bromination for determination of aflatoxins in peanut butter, pistachio paste, fig paste, and paprika powder: Collaborative study. *Journal of AOAC International* 83: 320–340.

Van Egmond HP (ed.) (1989) *Mycotoxins in Dairy Products*. London: Elsevier.

AGGLOMERATION

L Coucoulas, Leatherhead Food Research Association, Leatherhead, Surrey, UK

Definitions, History, and Scope

Agglomeration is a powder-size enlargement process, where small particles combine to form large, relatively permanent masses in which the original particles are still identifiable. These agglomerates have a coarse, open structure and a mean particle size ranging from 0.1 mm to a few millimeters. The process uses mechanical agitation in the presence of the required proportion of a liquid phase and perhaps other binding agents and is normally followed by evaporative drying. Agglomeration is sometimes confused with granulation. The latter is, in fact, a term describing all size enlargement processes: compaction, extrusion, prilling, nodulization, sintering, and agglomeration.

Agglomeration as a size enlargement process is not confined in its application to the food industry. The mineral-processing, chemical, surfactant, and pharmaceutical industries use several methods of granulation, including agglomeration, to solve their powder-handling problems. The advantages of agglomeration when compared with other granulation techniques such as compaction are lighter granules with large open pores and superior reconstitution properties.

Agglomeration started to be applied to food powders in the 1960s in the USA. Several processes were developed for dairy products, sugars, flour and cake mixes, chocolate drinks, instant coffee and tea, and other special drinks. Some of these processes are still used today although the majority have been superseded by continuous straight-through fluid bed processes, which are more economical and have less down time or product outside specification.

Agglomeration is employed to improve the handling properties of fine food powder materials when these are used either as intermediate raw materials or as final consumer goods. There are two distinct areas where agglomeration has a beneficial effect for both the consumer and the industrial user: dry handling and reconstitution of fine powders.

Dry Handling

Fine powders are used extensively in the food industry. They are produced from a variety of drying and crystallization processes such as oven drying, spray drying, vacuum or freeze drying, drum drying, etc., which may, when appropriate, be followed by a crushing and milling process. Fine powders invariably exhibit cohesive flow properties, which result in poor or no flowability in silos and hoppers, excessive dust formation, segregation, and loss of control of bulk density. Agglomeration eliminates these problems. (*See* **Drying**: Theory of Air-drying; Spray Drying.)

Agglomerated products exhibit:

1. improved flowability resulting from their having larger particles and a more uniform particle size distribution; for the consumer this means instantly and completely emptying packets with no material clinging to corners;
2. elimination of dust formation, owing to a substantial decrease in the proportion of fines;
3. uniform bulk density, which is usually a function of particle size and porosity, and less segregation in retail packages.

Reconstitution

The reconstitution of fine powders presents problems for the industrial user and the consumer. In industrial use long mixing times must be used with intensive mixing equipment interspersed with screening operations and settling tanks. For the consumer who is only equipped with a spoon and a cup, reconstitution of a soup mix or a chocolate drink needs to be convenient and quick, even in cold water and without the formation of partly hydrated lumps on the surface of the cup that may be mistaken for substandard product. Agglomerated powders are attractive because they have:

1. greatly improved dispersion and rehydration characteristics because of their open porous structure, which allows water to penetrate and disperse the granule with minimum stirring;
2. improved sensory properties, i.e., intense color and an attractive appearance.

Instant Powders

An agglomerated powder with greatly improved dry handling and reconstitution properties is termed an 'instant' powder. The properties and standards for testing the quality of an instant powder differ according to the type of powder. Apart from microbiological requirements that have always had to be fulfilled, the properties for testing the quality of an instant powder can be divided into three groups:

1. appearance, structure, size, and handling properties of the powder;
2. easiness and completeness of reconstitution;
3. appearance and lack of defaults of the reconstituted solution.

Fat-containing powders such as whole milk, spray-dried fats, some chocolate drinks and soup or gravy mixes are difficult to reconstitute in cold water even if they are agglomerated. Such powders are usually covered by a thin layer of surface fat that makes them water-repellent. Coating these powders with small amounts of wetting agents such as lecithin dispersed in fat makes them wettable and cold-water-soluble. This operation has in some instances been termed 'instantization.' In fact, instantization is the production of an instant powder that fulfills certain predetermined standards and may be achieved either by agglomeration or some other form of granulation, lecithination, or both.

The Agglomeration Process: Molecular Bonding and Growth

Knowledge of the types of bonds generated between agglomerating particles and the mechanisms of granule growth is of paramount importance in understanding the agglomeration process. Primarily, the bonds formed between solid particles are due to electrostatic and intermolecular attraction forces. In the presence of a liquid, though, the distance between centers of attraction is reduced and stronger bonds are formed. Their manifestation is termed 'liquid bridges.'

Liquid Bridges

As the name indicates, these are bonds formed between the liquid and the particles. They are strongly dependent on the wetting angle, the adsorption characteristics of the powder, and the amount and viscosity of the binding liquid. When the particles are held together by a low-viscosity liquid, owing to surface tension forces and depending on the liquid saturation, four different states are possible:

1. Pendular state – the liquid is held in discrete lens-like rings at the point of contact between the particles. The air forms a continuous phase throughout the agglomerate.
2. Funicular state – at higher liquid saturations the liquid rings coalesce to give a continuous liquid phase throughout the agglomerate, in which discrete pockets of air are entrapped.
3. Capillary state – the pore space is completely saturated with liquid but the liquid does not cover the external agglomerate surfaces. A negative capillary pressure is built up within the agglomerate, which gives it a degree of tensile strength.
4. In addition to the above, if the liquid covers the external surfaces of the agglomerate, the particles become completely encapsulated within the liquid.

Use of liquids with a high viscosity usually results in agglomerates of higher strength than is achieved using low-viscosity liquids, depending on the fluid cohesion strength and the fluid–solid adhesion force.

Solid Bridges

Liquid bridges hold the particles together during granule growth. Upon drying, these usually result in solid bridges through phase transformation, i.e., recrystallization of the dissolved material, hardening of binder liquid or chemical reaction. In this way the agglomerate comprises particles connected by solid bridges, in which capillary pores are present. The tensile strength of solid bridges is of the same order of magnitude as that of the constituent particles.

Mechanisms of Growth

The overall mechanisms of granule growth determine not only the rate of agglomeration but also the final properties of the agglomerate. Although these mechanisms are very strong functions of the process equipment used, the following steps are generally accepted:

1. nucleation of primary particles by random coalescence;
2. transition stage growth;
3. final 'ball' growth.

In the initial nucleation period, the feed particles grow by rapid random coalescence. During the transition period, particles grow by preferred coalescence between large and small granules (in the case of wide particle size distributions) or by crushing and layering (in the case of a narrow particle size distribution). The growth rate passes through a maximum, which can be perceived as the rate at which the diameter of the granule increases to a size such that any further increase is accompanied by a corresponding increase in attrition crushing.

The principal variable used to control agglomeration is the liquid saturation necessary to achieve the surface tackiness, which leads to a greater propensity of particle coalescence. Research has shown that in many cases this critical moisture content is around 70–90% of that needed to saturate the voidage in the powder being agglomerated.

The choice of equipment and process parameters, such as fluidization velocity and speed of rotation, determine to a large extent the growth

mechanism, which for the same moisture content can be based on coalescence or crushing and layering, or a combination of both. This explains the different densities of agglomerated products obtained from the same material and the same volume ratio of liquid to solids but made on different equipment or of products manufactured on the same equipment but at different agitation motions.

Characteristics of Agglomerated Powders in Contrast with the Starting Materials

It is the physical properties of fine powders that are the source of problems in powder handling. Agglomeration is essential in order to modify these and improve the usefulness of fine powders. There follows a brief description of these properties and how they are modified by agglomeration.

Shape

The shape and surface roughness of particles dictate the number of contact points between adjacent particles and the minimum proximity. The shape has a pronounced effect on the flowability; for example, particles that have a fibrous, highly asymmetric shape can interlock and inhibit free flow. Agglomeration changes the shapes of the particles dramatically and, in general, creates more symmetrical particles with fewer 'cooking points' and therefore improved flowability.

Particle Size

Particle size is a contentious property, because most particulate matter has an irregular shape and its size differs according to the method used to measure it. It is usually defined as the 'dimension of an imaginary spherical particle having the same value in the physical property measured as the irregularly shaped particle.' Not all of the particles of a powder are the same size, and a particle size distribution is usually employed to characterize a powder. The particle size distribution correlates well with many other material properties and bulk characteristics.

Figure 1(a) and (b) shows two particle size distributions. The first is typical of a powder resulting from a milling process. It contains a large proportion of very fine material, causing dust formation in dry handling and wet lump formation during reconstitution. Furthermore, the vast range of particle sizes will lead to acute flowability problems, time consolidation behavior, variable density due to nonuniformity of packing and, finally, segregation or compaction during transportation. Thus, the effects of a wide particle size distribution are spread across the whole range of uses of powdered material. Agglomeration changes the particle size distribution in two ways (Figure 1b). First, it shifts the mean to a larger particle size, thus eliminating all problems arising from the presence of a very fine material. Second, it gives a product with more uniform size, thus narrowing the width of the particle size distribution curve, which eliminates all the other problems outlined above.

Density, Porosity, Surface Area, Pore Size Distribution

In the case of particulate matter, there are four types of density used – true, particle, pouring, and tap density. True density is the density of the material when it does not include any interstitial pores; particle density is the density of the material when it is in

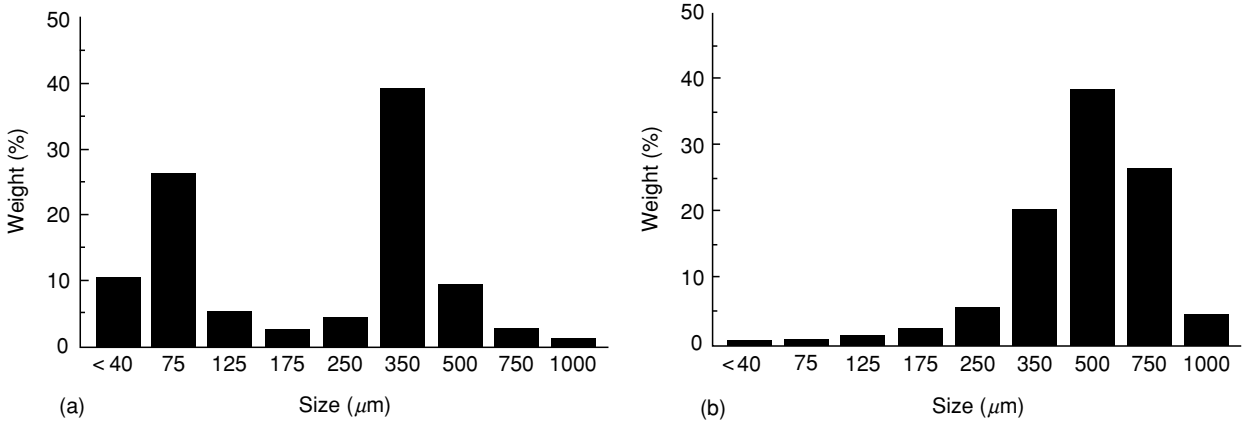

Figure 1 Particle size distributions of (a) a fine powder showing a bimodal particle size range and (b) an agglomerate with a narrow range of particle sizes. Reproduced from Agglomeration, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

particle form and includes interstitial pores; pouring density is the bulk density of the powder when it is poured into a container; and tapped density is the density measured after the container has been shaken and the powder has consolidated. Both of the latter include interparticle porosity. The difference between the poured and tapped densities determines the degree of consolidation and how much empty space will be left on top of a jar filled with product at the factory after it has been shaken through transportation. Density is also important during reconstitution. After wetting, the density of particles determines how fast they sink and whether they have time to dissolve or be stabilized in suspension.

The knowledge of surface area and pore size distribution provides insight into the structure and adsorption potential of particulate matter. This is often calculated from the particle size distribution and by use of the mercury porosimetry method.

In general, agglomerated granules are of larger size with a lower particle or bulk density. Such granules sink slowly into water and therefore can disperse and enter solution before reaching the bottom of the container. Agglomeration also increases the porosity and, particularly, the formation of large numbers of interconnecting micropores, which have potentially large capillary pressures during wetting, i.e., liquid is sucked into them by capillary action. It has been shown that the total wetting time is a decreasing function of porosity. Also, the presence of micropores is instrumental in the spontaneous deagglomeration process, which is essential for instant powders.

It is of interest to contrast the reconstitution of agglomerates with that of aggregates of fine powders. Agglomeration produces an increase in the amount of air between particles. During reconstitution air is replaced by water. In fine powders the surface area is large and the interstitial space small and usually not interconnected. As a result, water penetrates slowly and, after a short time, the space between particles is filled with high-concentration dissolution products, resulting in a sticky jelly with islands of unwetted powder and residual air. Furthermore, lumps form, which are wet and swollen outside, while dry inside. These hamper further reconstitution (even under strong agitation) because their structure is impervious to water. In an agglomerate the large amount of interstitial air results in the particles being dispersed in the liquid before the highly viscous solution is formed.

Cohesiveness

To insure easy and successful handling during processing, powders must be sufficiently 'flowable.' The factors that influence flowability include surface properties, shape and size distribution, and the geometry of the processing equipment. The forces that are involved are gravitational, cohesive (between particles), and adhesive (between particles and walls).

Agglomeration produces granules that show a marked improvement in the flowability properties, which is qualified by a significant decrease in their angle of repose. This occurs because, in general, there is an increase of the mean particle size coupled with a narrowing of the width of the particle size distribution curve.

Hygroscopicity

Hygroscopicity is the quantity that expresses the propensity of the material to pick up moisture, which occurs when the equilibrium relative humidity of the product is below atmospheric humidity. Therefore, when nonagglomerated hygroscopic powders are used in containers open to the atmosphere, the powders tend to pick up moisture and form lumps that are hard on the outside and enclose dry material. These problems of caking and lumping are often encountered in the discharge of powders both in hoppers and in vending machines, particularly since in the latter the atmospheric humidity is high. Anticaking agents are often included in hygroscopic powders and their function is to act as moisture scavengers and thus prevent moisture pick-up by the powder. The equilibrium relative humidity is a function of product moisture; thus, if water is used as an agglomerating medium it will at the same time increase product moisture content and product equilibrium relative humidity, decreasing the likelihood of moisture pick-up and caking.

Wettability

The wettability of a material is a property of the chemical composition of the surface of the material. Changes in the surface structure of a powder can be affected by the adsorption of a surfactant. The term 'instantizing' is often loosely used to describe a process whereby agglomeration is followed by coating with a surfactant, resulting in granules that are easily wettable. Fat-containing powders such as ordinary spray-dried whole-milk powder, chocolate drink powder, or dried soup mix have poor wettability and the object of the instantizing process is to improve it. In the case of powder products, however, the effective contact angle of the material in powder form controls the wetting process. The effective contact angle is a function of particle size, porosity, and moisture content as well as material wettability. The behavior of complex powders cannot therefore be predicted by simple static contact angle measurements. By measuring the height of wetted beds of

powders against time, however, it has been shown that agglomeration decreases the effective contact angle and the time required for bulk wetting provided that the agglomerate size lies within certain limits (0.4–1 mm for some powders).

Sensory Qualities

A final consequence of agglomeration is the change of the color and appearance of the product. These are important qualities from the consumer's point of view and their change might improve or diminish consumer appeal. For example, the color of an agglomerated product may be denser and richer than that of the original fine powder. Alternatively, a composite powder may be agglomerated to give a product of improved speckled appearance, which will also have minimum separation problems and enhanced reconstitution characteristics influenced by the most soluble of its individual components. (*See* **Sensory Evaluation**: Sensory Characteristics of Human Foods.)

Equipment and Processes used for Agglomeration

There are two main types of process used to agglomerate powders – the rewet process and the straight-through process.

The principle of the rewet process is the addition of water, or water containing a binding agent, to a dry powder, and then mixing of the particles in such a way as to cause interparticle collisions to occur and agglomerates to form. The particles of the powder can occur either by addition of the liquid as a spray (droplet agglomeration) or by condensation of water from steam or humid air on to the surface of the powder (surface agglomeration). The liquid may act on the powder material in several different ways:

1. It may dissolve part of it or transform it in such a way as to make the surface of the particles sticky or tacky and therefore make them adhere to each other in clusters.
2. If the fluid contains a binder in solution or in suspension, this binder may be responsible for the bonding function between the particles either on contact with the particles or as the agglomerate is redried and the binder crystallizes or precipitates. This process is commonly used to agglomerate inert materials.
3. It may be a material with a relatively high melting point, used as (1) or (2) above, but which on cooling solidifies to form the interparticle bond.

The straight-through process uses liquid feed as the raw material. In the first stage, this is dried into a powder, usually by spray drying, allowing a moisture content a few percent higher than that anticipated for the finished product. The particles thus formed show an increased tendency to adhere to each other. In the next stages of the straight-through process, the particles are agitated in ways similar to the rewet process.

In the next sections some of the types of plant used in the rewet or straight-through process will be described according to the mechanism used to promote interparticle collisions.

Agitation Tumbling Equipment (*See* Agitation)

This category includes plant that provides agitation by the rotation of a container around its central axis. It can comprise either batch or continuous units and these are mainly variations on the inclined dish agglomerator or the rolling drum.

The essential element of the inclined dish agglomerator is a dish, which is rotated on an inclined axis. The feed or recycled particles enter the dish and are sprayed with water whilst undergoing rolling motion. The relative positions of the input of feed material and sprays influence the size and structure of the agglomerates formed. Other features include a side or bottom scraper. An important aspect of the dish agglomerator is the pronounced size segregation that occurs as a result of the centrifugal forces (**Figure 2**). During continuous operation this allows the largest particles to be discharged first while the smaller particles have more time to grow. Therefore, a product of narrow particle size distribution is obtained and subsequent screening is unnecessary. Drying the product evaporates the moisture, leaving a granular product.

The rolling drum comprises a cylinder rotating around its longitudinal axis, which is slightly inclined to the horizontal so as to assist the continuous circulation of materials through the drum. It normally incorporates dam rings at either end in order to prevent spillback and increase the bed depth in the drum. The geometry of the drum makes it more suitable than the dish for handling dusty materials with minimum loss of fines, although most industrial dishes are covered to prevent fines from escaping in the atmosphere. The solids are wetted near the inlet end of the drum and the wet powder cascades on to the tumbling bed. Build-up of moist material is prevented by incorporating a stationary scraping bar, knockers that rap the walls, and reciprocating scrapers, and by roughening the inner surface of the drum.

An important design parameter affecting the performance of the drum is the critical rotational speed at which the material is carried completely around by centrifugal action. At low speeds the material forms a

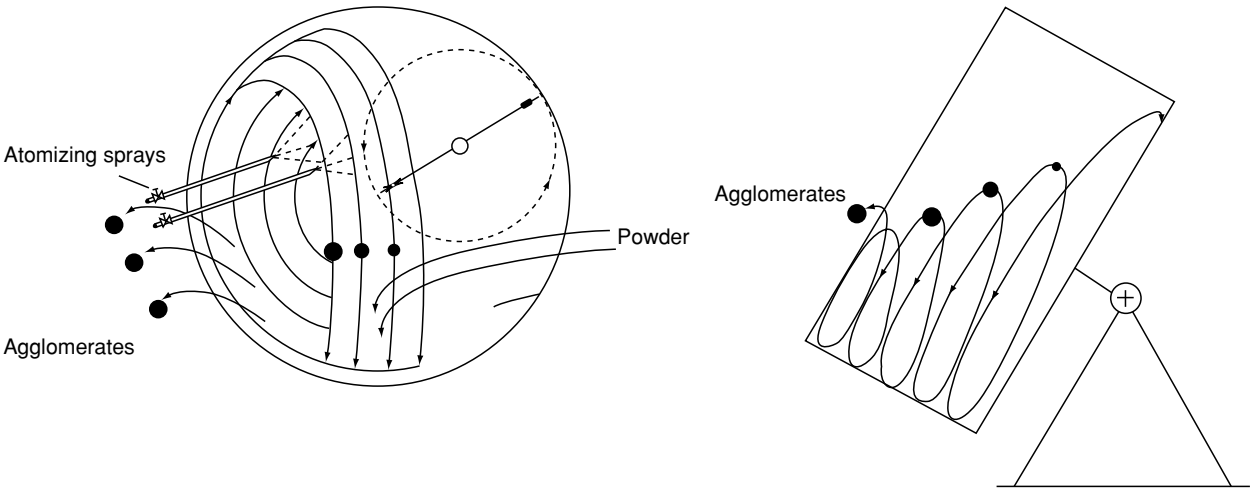

Figure 2 Particle size segregation in the dish agglomerator. Reproduced from Agglomeration, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

static bed and slides along the bottom with very little relative motion between the particles. Cascading takes place at around half the critical speed.

With all agitation and tumbling equipment the growth rate and final size depend on the residence time, binder saturation, and bulk loading. The speed of rotation, the angle of inclination, and the residence time can be varied to affect the density and strength of the granules produced. Sometimes mixers are incorporated inside the dishes to promote further particle collisions and break clusters in the case of very sticky materials. Agitation tumbling equipment has been used extensively in the chemical and mineral-processing industries but its use in the food sector is not so widespread. Maintaining a clean, contamination-free environment can be a problem as most available units are open to the atmosphere. Drying of the granules can be done using rotary drum driers or fluid beds. The large throughput and low cost of these units should make their application financially attractive in the future for food applications provided that the contamination problems are solved.

Mixers/Agglomerators

In this type of equipment water is sprayed on to a powder while the latter is in a turbulent state or in the form of a falling curtain. Shear processing leading to the formation of granules can be carried out in batch equipment including planetary mixers, ribbon blenders, Z-blade units, and high-speed intensive mixers. They rely on time and energy to effect the necessary contact between powder and liquid phases. All have their advantages for specific applications, although final granule properties may be adversely affected by excessive compression during prolonged mixing and the batch-mixed material requires sieving before drying in order to remove lumps.

In the NICA SYSTEM^{TM} turbine mixer/granulator, the powder phase, accurately metered, enters the inner section of a turbine, which deagglomerates the material and throws it outwards as primary particles. A gear pump feeds liquid binder to the underside of the inner turbine, at the periphery of which it is atomized into a mist. Mixing under high shear conditions takes place in the outer turbine; the degree of work and properties of the granule are controlled by varying the area of the exit port. Contact time is thus reduced to a fraction of a second. The resulting agglomerate is fed directly into a close-coupled fluidized bed and drying commences immediately. In this way, time-dependent problems such as solubility of the binder or lumping are eliminated.

The Schugi mixer/agglomerator (Figure 3) is a continuous device for blending small amounts of liquid with solids so that the product remains free-flowing. The mixing chamber consists of a 10-cm diameter hollow vertical cylindrical tube with a shaft running through its center. To the shaft are attached three sets of six knife-edged blades, the relative positions of which can be varied throughout the whole length of the mixing chamber. The shaft belt is driven at 3000 rpm. Liquids are introduced into the mixer through the jets, which are positioned above the top set of blades. The Schugi mixer is mounted vertically and solids are fed into the top. On entering, the solids are caught into contraflow turbulence created by the high peripheral velocity of the mixer blades. Liquids are pumped into the mixer under pressure and enter the mixing area as continuous threads of liquid. In the mixer the liquids are atomized by the combined

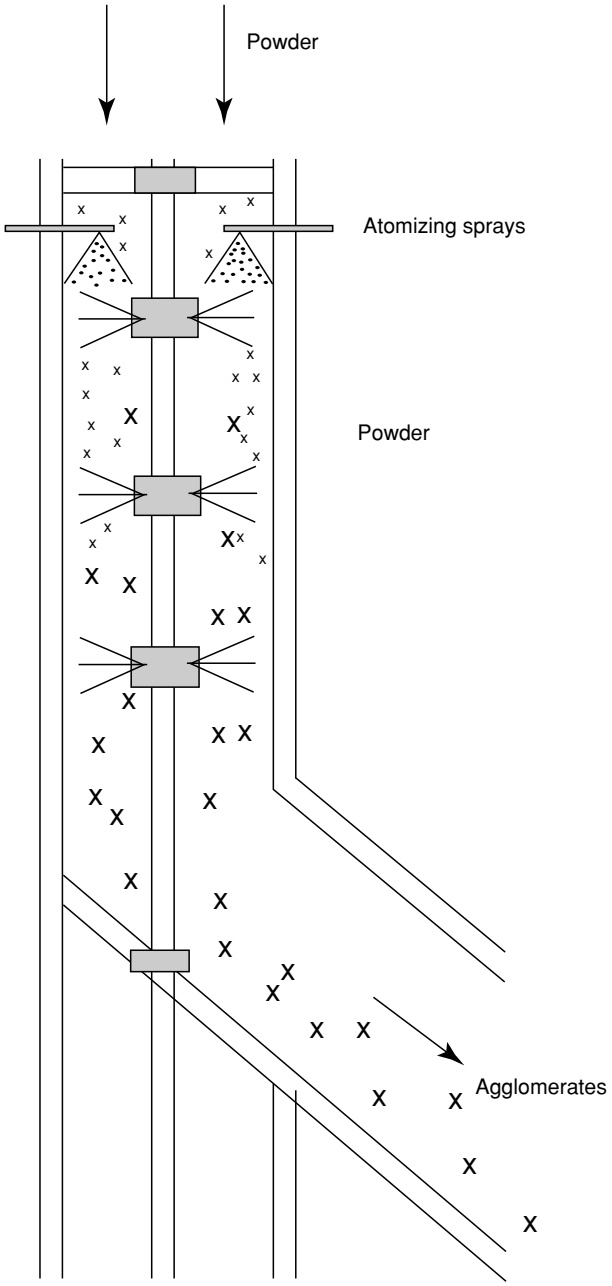

Figure 3 The Schugi mixer. Reproduced from Agglomeration, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

action of the top set of blades, which act in a similar fashion to a disk atomizer, and also by the solids themselves, which because of their high speed have a high energy potential. The atomized liquid is dispersed through the powder and the two adhere together to form agglomerates, which leave by spiralling around the walls of the mixer. The residence time is very short, of the order of 0.5 s, and that limits the temperature increase of the product to 2 °C.

The structure of the product formed depends both on the relative sizes of the liquid droplets and the powder particles and on the feed ratio of the two. Apart from the physical properties of the materials being mixed, the equipment control variables include the rotational speed, the position of the blades, the direction of rotation, and the size of the liquid injection jets. The agglomerated product is dried in a fluid bed.

A third type of mixer agglomerator is referred to as the 'agglomerating tube.' In this device agitation is insured by dropping the powder into a high-velocity stream of humidified air along the length of a tube. The resulting powder clusters are separated in a cyclone and then picked up by a hot air stream and blown through a second tube for drying. A considerable amount of impact between particles takes place inside the agglomerating tube, which makes it suitable for 'difficult' powders that tend to form sticky deposits easily.

In another process the powder is passed through a series of vibrating hoppers or conveyor belts and falls in the form of a thin continuous curtain inside the agglomerating vessel. Steam or atomized water is then injected on to the curtain, which has the force to redirect the traveling plane of the particles. The induced turbulence makes the particles collide and, because they are wet, adhere to each other. The powder clusters formed are carried on to a drying section.

It is evident that some of these units like the 'falling curtain agglomerator' are 'inhouse' types of equipment developed and modified to suit the particular needs of specific products. Recently, they have been superseded by the fluidized bed agglomerators, which together with the straight-through process are the 'state of the art' agglomerators in use today. It is, however, worth considering the use of mixing as the first stage in an agglomerating process, especially in cases where the nonwetting character of the powder requires high shear to effect the initial moistening of the material surface.

Fluid Bed Agglomerators – The Straight-through Process

Fluid bed operations are characterized by:

1. good random mixing;
2. stable, uniform temperature profile leading to effective interphase contact and controlled high rates of heat and mass transfer;
3. high thermal efficiency, which allows use of low exhaust air temperatures and therefore avoids degradation of thermally sensitive food powder components.

These exceptional properties of fluidized beds are the cause of their almost universal use in agglomeration processes – batch, continuous, spouted beds, vibrofluidizers – or in straight-through combinations.

The continuous rewet agglomerator usually consists of three sections that may or may not be separated by weirs. Powder is fed from the feed hopper into the first section of the fluid bed. Here it is rewetted by means of steam, atomized water, or binder solution delivered from jets positioned on top of the fluid-bed layer. The powder particles increase in size either by surface layering or random coalescence. Surface layering is more likely to happen with low-viscosity wetting agents which allow the moisture to distribute evenly around the particles. 'Blackberry'-type agglomerates form when particle surfaces are made just wet enough for collision of particles to lead to capture. Typical moisture contents of the rewetted powder in the first section are 6–10%. Too high wetting will result in deposits in the bed and a blocked distributor plate. Too low wetting will reduce the level of agglomeration and produce less stable agglomerates. Vibration is used to reduce channeling and break powder clusters. The agglomerated material passes into the second section, where it is dried in a hot air stream. After drying it moves to the third section for cooling. Feed particle sizes range between 30 and 300 μm. Agglomerates of 0.5–1.5 mm in size are typically produced.

Straight-through combination systems consist of a spray drier in which liquid feed material is dried to produce a powder at a moisture content slightly higher than would be expected for a final stable product (6–9%). The partially dried powder is then fed to a number of static and/or vibrating fluid beds, where final drying and cooling take place. Fine unagglomerated material is removed pneumatically and taken back to the atomization zone in the spray drier, where it is mixed with the atomized droplets. Agglomeration occurs through contact of fine powder and liquid feed in the atomization zone as well as through interparticle collisions of the damp particles in the fluid beds. The remainder of the operation consists of final drying to 2–5% moisture and cooling and is similar to that in rewet fluid-bed agglomeration. If a lecithination (instantization) stage is required to be included to improve the instant properties of the powder, this is usually carried out by spraying a lecithin-in-oil solution on the agglomerates after the drying stage. Lecithination may be carried out either by direct spraying on the fluidized powder or in a powder trap between two fluid beds. A conditioning fluid-bed stage is then usually added to the process in order to distribute the lecithin uniformly on the surface of the particles.

A number of such straight-through systems differing in geometry and number of stages have been produced by companies such as APV Anhydro (Invensys), NIRO, and others. Powders produced from the straight-through process are nondusty, free-flowing, and with a particle size ranging between 250 and 500 μm. The process has been successfully applied to a wide range of products where a liquid raw material needs to be dried and transformed into an instant powder. Its almost universal application is, however, offset by the high investment cost.

Microbiological Hazards of Agglomeration

Few outbreaks of foodborne diseases have been reported due to dried powders. Microbiological hazards are usually associated with the liquid raw materials and the emphasis of any manufacturer should be to insure that adequate heat treatment is given to any liquid product that is going to be spray-dried and agglomerated. Tests for *Salmonella*, Enterobacteriaceae (*Escherichia coli*, coliforms) and *Staphylococcus aureus* should be carried out routinely. There are, however, occasions where the product may be contaminated after drying through contact or handling. (*See* **Food Poisoning**: Tracing Origins and Testing.)

See also: **Agitation**; **Drying**: Theory of Air-drying; Spray Drying; **Food Poisoning**: Tracing Origins and Testing; **Sensory Evaluation**: Sensory Characteristics of Human Foods

Further Reading

James M, Brault M and Coucoulas L (1991) *Lecithination – Instant Character Dependence and Shelf-life Tests*. LFRA Research Report no. 689. Leatherhead: Leatherhead Food Research Association.

Pisecky J (1980) Instant whole milk powder. *Australian Journal of Dairy Technology* 35(3): 95–99.

Pisecky J (1990) 20 years of instant whole milk powder. *Scandinavian Dairy Information* 2: 74–78.

Retsina T and Coucoulas L (1988) *Agglomeration of Powders*. LFRA Scientific and Technical Survey no. 164. Leatherhead: Leatherhead Food Research Association.

Silliker JH, Baird-Parker AC, Bryan FL *et al.* (eds) (1988) Application of the hazard analysis critical control approach (HACCP) to ensure microbiological safety and quality. In: *Microorganisms in Foods*, vol. 4, Chapter 2, pp. 22–43. Oxford: Blackwell Scientific.

AGING – NUTRITIONAL ASPECTS

N P Hays and S B Roberts, The Jean Mayer USDA Human Nutrition Research Center on Aging at Tufts University, Boston, MA, USA

Aging

Life expectancy at birth in the USA is now 76.5 years, compared with about 47 years at the beginning of this century. Consequently, the proportion of the population that is elderly, as well as the mean age of the population, has increased. A continuation of these trends is anticipated through the beginning of the twenty-first century. One of the consequences of this profound demographic shift is an increased awareness that nutritional influences on health should be optimized for the older population. This manuscript is based on an earlier review by Roberts and Hays, describing the nutritional requirements of older people.

Old age is a time when maintaining a good nutritional status is a critical determinant of health – but at the same time is often more challenging than it is earlier in life. As described below, the body's need for some nutrients is actually increased relative to that in earlier adult life, while the body's subconscious ability to regulate food and nutrient intakes diminishes. For these reasons, the elderly are a group particularly vulnerable to inadequate dietary intakes. The consequences of excessive nutrient intakes, particularly of energy, are also more severe in older adults. Obese elderly persons are much more likely to suffer disabling comorbidities associated with obesity such as heart disease, osteoarthritis, and reduced mobility, which in turn can lead to a downward spiral of disability and frailty. Thus, it is important to know the nutritional requirements of elderly persons, and the food choices that can insure adequate nutrition.

Determining accurate recommended dietary allowances (RDAs) for any age group is challenging. Historically, RDAs have been focused towards preventing nutrient deficiencies, whereas there is now increasing recognition that increased or decreased intakes of some nutrients may also have protective functions against late-life chronic diseases such as coronary heart disease and some cancers. The newest update of the RDAs, now called dietary reference intakes (DRIs), does in fact consider these protective functions as a factor in setting nutrient recommendations. At the time of preparation of this manuscript, 1997 DRIs are available for calcium, magnesium, phosphorus, fluoride, and vitamin D, 1998 DRIs are available for thiamin, riboflavin, niacin, vitamin B_6, folate, vitamin B_{12}, pantothenic acid, biotin, and choline, and 2000 DRIs are available for vitamin C, vitamin E, and selenium. Recommendations for all other nutrients remain based on the 1989 RDAs (new DRIs for the remaining nutrients are expected to be released over the next several years). In some cases, the increased intakes that appear to decrease the risk of late-life chronic disease are unfeasible without resorting to nutrient interventions through fortification or supplementation. Despite increased recommendations for several nutrients in the new DRIs, the question of whether DRIs should recommend levels of nutrients compatible with what can be consumed through foods, or whether in some cases pharmacological supplementation may be appropriate, is an ongoing debate. Of particular relevance to this issue is the fact that, in epidemiological studies linking diet to disease, it is the consumption of foods, rather than nutrients, that is actually measured. Metabolically active 'phytochemicals' in such foods as fruits and vegetables may play important roles in disease prevention and are not at the present time understood well enough to be used as supplements.

The determination of accurate and meaningful DRIs is particularly difficult in the elderly population. Whereas the majority of young adults are healthy, the majority of elderly persons have one or more chronic diseases or disabilities that affect nutrient uptake and utilization and at the same time use multiple medications to treat chronic diseases. If nutritional recommendations are to be made not only for healthy elderly persons, but also for sub-groups with different health limitations, this vastly increases the complexity of formulating accurate recommendations. Moreover, the objective of DRIs may change with adult age, with increasing focus on maintaining current health rather than preventing deficiency or future disease. For example, a recommendation to consume a low saturated fat diet to prevent later coronary heart disease may be inappropriate for 85-year-old men and women, who often find it hard to consume enough food to maintain weight. Finally, the 'elderly' are not one group. Currently, DRIs provide specific recommendations for individuals aged 51–70 and >70 years, which assumes that all individuals over the age of 70 years are metabolically equivalent with regard to their nutrient needs. This is certainly not the case for several nutrients, and individuals in the age

range 70–80 years should be judged differently from those aged 80–90 years or older.

With these considerations in mind, this article focuses on the effects of age on the body's ability to use different nutrients. When an increase in the intake of a particular nutrient is needed to achieve the same circulating level of the nutrient in the body or the same level of adequacy, as judged by a particular enzymatic activity, this is taken as evidence of an increase in requirements. A more advanced treatment of nutritional requirements in the elderly must await the collection of data more sophisticated than is currently available.

Energy Requirements

Current recommendations on dietary energy intake define expected average amounts of dietary energy required for sustaining normal metabolic processes, together with desirable or expected levels of physical activity in healthy individuals. In weight-stable adults, energy requirements are equal to total energy expenditure since body energy stores remain approximately constant. Current dietary guidelines also note that not more than 30% of dietary energy should be supplied by fat, and only 10% or less supplied from saturated fat. This recommendation applies to adults of all ages. A further general consideration concerning energy requirements is that body weight should ideally be maintained within the body mass index (BMI, weight in kilograms divided by the square of the height in meters) range of 20–25 $kg\,m^{-2}$.

Age, body weight, and the level of physical activity are the primary determinants of energy requirements for body weight maintenance in adults of all ages. Concerning the effects of age, the older the person, the lower are the energy requirements. For body mass, the larger the body mass, the greater are the energy needs of the body. Height is an important determinant of body mass since taller persons tend to weigh more than shorter persons. Physical activity is also an important determinant of energy requirements. Table 1 shows the recommended energy requirements of adult men and women of typical weight expressed as 'PAL' (physical activity level) units, which are energy requirements divided by the estimated 'basal metabolic rate' (the minimum expenditure of the body when lying supine at rest at least 12 h after the last meal). PAL units are given for light, moderate, and heavy levels of activity, and the basal metabolic rate can be calculated using the equations at the bottom of Table 1.

As predicted by the US National Academy of Sciences recommendation, elderly adults typically have substantially decreased energy requirements relative to young adults by 1255–2510 kJ per day (300–600 kcal per day) (Table 2). This is thought to be for two main reasons: (1) basal metabolism decreases with age, owing to a reduced amount of lean tissue (the actively metabolizing component of the body) as well as a decrease in the amount of metabolic activity per unit of lean tissue; (2) although not inevitable, the capacity for physical activity is typically decreased in older persons in concert with increased body fat mass.

The estimated energy needs given in Tables 1 and 2 have been challenged as somewhat inaccurate. This is

Table 1 Estimated energy requirements by physical activity level

PAL	*PAL units*[b]	*Energy needs at 65 years (kJ per day)*[a]	
		Man (75 kg)	*Woman (60 kg)*
Light activity	1.50	10 615	8 025
Moderate activity	1.65	11 678	8 828
Heavy activity	2.00	14 154	10 698
Estimated BMR of men (kJ per day)[c]			
Age 18–30	BMR = (64.0 × weight in kg) + 2841		
30–60	BMR = (48.5 × weight in kg) + 3678		
> 60	BMR = (56.5 × weight in kg) + 2038		
Estimated BMR of women (kJ per day)[c]			
Age 18–30	BMR = (61.5 × weight in kg) + 2075		
30–60	BMR = (36.4 × weight in kg) + 3469		
> 60	BMR = (43.9 × weight in kg) + 2494		

[a] 1 kJ ≈ 0.239 kcal.
[b] Physical activity level (PAL) = basal metabolic rate (BMR) multiplied by activity factor.
[c] Based on values obtained from WHO (1985) *Energy and Protein. Requirements*. Report of a joint FAO/WHO/UNU expert consultation. Technical Report Series 724. Geneva: World Health Organization, with permission.

Table 2 Effects of age on estimated energy requirements[a]

Age group (years)	*Body weight (kg)*	*PAL units*[b]	*Estimated energy needs (kJ per day)*[c]
Men			
19–24	72	1.67	12 134
25–50	79	1.60	12 134
51+	77	1.50	9 623
Women			
19–24	58	1.60	9 205
25–50	63	1.55	9 205
51+	65	1.50	7 950

[a] Based on values obtained from National Research Council, Subcommittee on the 10th Edition of the RDAs, Food and Nutrition Board, Commission on Life Sciences (1989) *Recommended Dietary Allowances*, 10th edn. Washington, DC: National Academy Press, with permission.
[b] Physical activity level (PAL) = basal metabolic rate (BMR) multiplied by activity factor.
[c] 1 kJ ≈ 0.239 kcal.

because, based on research using doubly labeled water to measure energy requirements (published after the most recent recommendations were released), they appear to underestimate the usual energy needs of healthy men and women. The underestimation has been suggested to be due to an underestimate of the energy expenditure associated with normal daily living. In a summary of recent energy requirement studies in elderly persons, the measured PAL value of older men and women averaged 1.65, compared with the suggested recommendation of 1.51. This change would mean that the energy requirements of older men and women with typical body weights would average 10 539 kJ (2519 kcal) and 8828 kJ (2110 kcal) per day, respectively. Like the recommended dietary allowances, however, the new research confirms a substantial decrease in caloric requirements with age.

Protein Requirements

The 1985 FAO/WHO Committee on Energy and Protein requirements defines the protein requirements of an adult as 'the lowest level of dietary protein intake that will balance losses of nitrogen from the body in persons maintaining energy balance at modest levels of physical activity'. Nitrogen balance is determined in metabolic studies as the difference between measured nitrogen intake (from dietary protein which is about 16% nitrogen by weight) and nitrogen losses. The major routes of nitrogen loss – in urine and feces – are measured, and miscellaneous nitrogen losses, including those in hair, skin, sweat, and breath ammonia, are assumed to be equal to 8 mg per kilogram of body weight. Based on this nitrogen balance methodology, it is estimated that the mean requirement for good quality protein in adults is 0.6 g per kilogram of body weight, and the recommended safe protein intake (to reflect individual variability in requirements) is 0.75–0.80 g per kilogram of body weight.

No change in protein requirements with age is currently recommended, because of the limited and conflicting information in the literature at the time the most recent recommendations were developed. However, it was recognized in the recommendations that values of 0.75 and 0.80 g per kilogram of body weight provide for higher intakes of protein per unit of lean tissue in older individuals, because of the reduced amount of lean tissue in this group. Thus, even if elderly persons have an increased requirement for protein relative to their amount of lean tissue, this may be accommodated already within the current RDAs.

Recent studies have raised the question of whether the current recommendations may underestimate usual protein needs in elderly persons. In a review and reanalysis of previous studies, protein requirements of elderly persons appeared to be increased to > 1.0 g per kilogram of body weight per day. However, this tentative observation was based on potentially inaccurate assumptions including that miscellaneous nitrogen losses from the body are the same in young and elderly subjects (when they are probably lower in the elderly), and that an equilibration period of only 5 days is sufficient to bring elderly persons into nitrogen balance when protein intake is adequate. Another study found that protein requirements were unchanged in healthy elderly women, compared with younger women, and reduced in elderly men. This controversy over whether the need for dietary protein is increased in the elderly cannot be resolved without further investigation.

Vitamin Requirements

RDAs and DRIs for vitamins in adults are shown in Table 3. Different approaches were taken to determine these estimated vitamin requirements in adults, including depletion–repletion studies conducted in a metabolic ward, measurement of intakes required to maintain constant body stores, and the activity of enzymes requiring vitamins as cofactors. In studies used to determine 1989 RDAs, there was a consistent emphasis on the prevention of clinical or biochemical deficiency as well as consideration of usual dietary intakes, rather than long-term prevention of chronic disease. Although an attempt was made to set the 1997, 1998, and 2000 DRIs at levels designed to reduce the risk of chronic disease, the DRI committees judged that, except for calcium and fluoride, the available scientific evidence was too weak to allow reduction of disease risk to be used as a criteria for setting recommendations. The RDA values listed are derived by adding two standard deviations to the estimated average requirements. The DRI values for vitamin D and biotin, however, are based simply on the observed mean nutrient intakes that appear to sustain a population's normal nutritional status.

The recommendations given in Table 3 are, for the most part, the same for young and older adults – again defined as less than and greater than 51 years of age. The exceptions are vitamin D requirements, which are set at a higher level in older persons based on evidence suggesting that the capacity of human skin to synthesize vitamin D decreases with age, and vitamin B_6 requirements, which are also higher in older persons based on a small number of studies suggesting that the level of vitamin B_6 required to normalize cofactor-dependent enzyme function and excretion levels increases with age.

Table 3 Recommended daily vitamin needs (1989 RDAs except as indicated)

Vitamin	*Younger adults*		*Older adults*		
	Male	*Female*	*Male*	*Female*	*Probable bias in older subjects*
Vitamin A (μg RE)	1000	800	1000	800	Too high
Vitamin D (μg)[a]	5	5	10–15	10–15	Probably OK
Vitamin E (mg-αTE)[b]	15	15	15	15	Probably OK
Vitamin K (μg)	80	65	80	65	?
Vitamin C (mg)[b]	90	75	90	75	Too low
Thiamin (mg)[c]	1.2	1.1	1.2	1.1	Probably OK
Riboflavin (mg)[c]	1.3	1.1	1.3	1.1	Probably OK
Niacin (mg NE)[c,d]	16	14	16	14	?
Vitamin B_6 (mg)[c]	1.3	1.3	1.7	1.5	Too low
Folate (μg DFE)[c,e]	400	400	400	400	Too low
Vitamin B_{12} (μg)[c]	2.4	2.4	2.4	2.4	Too low
Biotin (μg)[c]	30	30	30	30	?
Pantothenic acid (mg)[c]	5	5	5	5	?
Choline (mg)[c]	550	425	550	425	?

[a]1997 DRI range for older adults: the lower value is DRI for 51–70 years; the higher value is DRI for >70 years.
[b]2000 DRI.
[c]1998 DRI.
[d]As niacin equivalents (NE). 1 mg of niacin = 60 mg of tryptophan.
[e]As dietary folate equivalents (DFE). 1 μg of food folate = 0.6 μg of folic acid (from fortified food or supplement) consumed with food = 0.5 μg of synthetic (supplemental) folic acid taken on an empty stomach.

There has been increasing recognition, however, that current RDAs for other vitamins may not be accurate for elderly individuals. In particular, and as summarized in Table 3, there is evidence that DRIs for vitamins C, B_{12}, and folate may be too low and that the RDA for vitamin A may be too high. There is also evidence that the DRI for vitamin B_6 may be inappropriately low as well.

Vitamin C

Vitamin C has three main roles in human metabolism – enyzme cofactor, chemical reductant, and antioxidant. Although the 2000 DRI for vitamin C is increased compared with the previous RDA value, recent research has provided evidence that the currently recommended intake of vitamin C may still be too low for all age groups, including elderly persons. For the same vitamin C intake, elderly persons have lower circulating levels compared with young adults, perhaps because of either impaired absorption of vitamin C from the gut or impaired reabsorption from the kidney. Recent studies examining vitamin C bioavailability, urinary excretion, and steady-state plasma vitamin C concentration as a function of dose suggest that the recommended vitamin C intake should be increased from 90 mg (for men) and 75 mg (for women) to 120 mg per day. Although preliminary studies indicate that even higher doses of vitamin C may prevent low-density lipoprotein oxidation, may improve wound healing, and are inversely correlated with blood pressure in elderly persons, results are inconsistent and further research is needed.

Vitamin B_{12}

Vitamin B_{12} is used in the body as a cofactor for several essential enzymatic reactions, including methionine synthesis and methylmalonyl-CoA mutase. Vitamin B_{12} requirements have been suggested to be increased in elderly persons, on the grounds that 2–3% of individuals over 60 years of age have pernicious anemia (a degenerative autoimmune condition of the stomach) and 10–30% have type B chronic atrophic gastritis (related to nonautoimmune 'environmental' conditions). These conditions result in either impaired intrinsic factor protein or acid and pepsin secretion by the stomach, respectively, which in turn reduces absorption of vitamin B_{12} from the diet. Vitamin B_{12} deficiency may be the one major undetected vitamin deficiency in the elderly population and has been suggested as being responsible for a proportion of the dementia cases among the elderly population. The 1998 DRI committee suggests that older individuals meet their vitamin B_{12} requirement through the use of vitamin supplements or fortified foods, because the absorption of the crystalline form of vitamin B_{12} is not influenced by atrophic gastritis. In older individuals with pernicious anemia, monthly B_{12} injections or daily ingestion of megadose B_{12} supplements ($400 \times$ DRI), allowing absorption by passive diffusion, are other means of normalizing B_{12} status.

Folate

Folate functions as a cofactor for enzymes involved in amino acid metabolism and nucleic acid synthesis. Recently, folate has received attention for its role in homocysteine metabolism and cardiovascular disease (CVD) risk. Excess levels of homocysteine in the blood have been implicated as a CVD risk factor; folate, along with vitamins B_6 and B_{12}, is involved in converting homocysteine to methionine, a harmless amino acid metabolite. In the elderly, studies have shown that folate intakes of at least 500 μg per day, along with a similar amount of vitamin B_{12}, may reduce blood homocysteine levels by as much as one-third. Fortification of grain products and breakfast cereals with folate (140 μg per 100 g) was begun in the USA in 1996 primarily to reduce the risk of prenatal neural tube defects. Recent studies suggest that fortification has also resulted in a substantial improvement in folate status in older adults as well; however, it is unclear whether the current level of fortification will have a significant effect on rates of CVD in the elderly population.

Vitamin B_6

Vitamin B_6 is a cofactor for a large number of enzymes used in the metabolism of amino acids and related compounds. A vitamin B_6 depletion–repletion study showed that vitamin intakes equivalent to a mean of 1.96 mg per day for elderly men and 1.90 mg per day for elderly women were needed to return urinary xanthurenic acid excretion to normal following a tryptophan load (one test for B_6 deficiency), thus indicating that current DRIs may be too low for this age group. This evidence was also consistent with the previous observation that many elderly persons show evidence of marginal biochemical vitamin B_6 deficiency.

Vitamin A

Vitamin A is required for a wide range of metabolic reactions in the body and in particular is used in the differentiation of epithelial cells, support of reproduction, and maintenance of the visual system. Concerning the possibility that the RDA for vitamin A may be too high and should be reduced, studies have indicated that elderly persons have decreased clearance of vitamin A (approximately 50% less than younger adults) by hepatic and other peripheral tissues. This finding helps to explain the observations that elderly persons have normal liver storage of vitamin A despite decreased vitamin A intakes, and that the use of vitamin A supplements by elderly persons is associated with increased circulating indicators of vitamin A overload (high fasting plasma retinyl esters).

Mineral Requirements

RDAs and DRIs for minerals in adults are shown in Table 4. As with the vitamin recommendations, different approaches were taken to determine these estimated mineral requirements in adults, including balance studies and summation of expected mineral needs based on metabolic demands and rates of absorption. The RDA values listed are best estimates of the safe intake for the whole adult population, i.e., the mean requirement plus 2 SD; the DRI values for phosphorus and magnesium are estimated in a similar fashion. The DRIs for calcium and fluoride, however, are based simply on the average nutrient intakes of healthy population groups. The recommendations given in Table 4 are the same for young and older adults, with the exception of iron requirements in older women, which are proposed to be lower on the grounds that iron losses are reduced following menopause, and calcium requirements in older men and

Table 4 Recommended daily mineral needs (1989 RDAs except as indicated)

Mineral	*Younger adults*		*Older adults*		
	Male	*Female*	*Male*	*Female*	*Probable bias in older subjects*
Calcium (mg)[a]	1000	1000	1200	1200	Too low
Phosphorus (mg)[a]	700	700	700	700	?
Magnesium (mg)[a]	420	320	420	320	Probably OK
Iron (mg)	10	15	10	10	Probably OK
Zinc (mg)	15	12	15	12	Probably OK
Iodine (μg)	150	150	150	150	?
Selenium (μg)[b]	55	55	55	55	Probably OK
Copper (mg)	1.5–3.0	1.5–3.0	1.5–3.0	1.5–3.0	Probably OK
Maganese (mg)	2–5	2–5	2–5	2–5	?
Fluoride (mg)[a]	3.8	3.1	3.8	3.1	?
Chromium (μg)	50–200	50–200	50–200	50–200	Too high?
Molybdenum (μg)	75–250	75–250	75–250	75–250	?

[a]1997 DRI.
[b]2000 DRI.

women, which are higher on the grounds that higher intakes ensure a maximal skeletal calcium retention.

Calcium

Based on increased recognition that the 1989 RDA for calcium may be too low, the 1997 DRI was set at 1200 mg per day for men and women 51 years of age and older. This value may still be too low for optimal calcium retention in individuals older than 65 years of age, but there are also issues of nutrient–nutrient interactions when considering an amount greater than 1200 mg per day. Calcium is the primary mineral in bone and teeth and is extremely important for the elderly, because the age-associated bone demineralization that commonly occurs in elderly persons is accelerated when calcium intake is low and leads to increased risks of bone fracture and tooth loss. Aging is associated with a decreased ability to absorb dietary calcium and also a decreased ability to increase the fractional rate of absorption when calcium intake is low. Calcium supplementation (for example using calcium citrate malate, which is better absorbed than calcium carbonate in postmenopausal women) has been shown to prevent age-associated bone loss both in women with initially low calcium intakes and perhaps also in those with intakes in the normal range.

Chromium

Chromium mediates the hormonal effects of insulin and so is needed for a wide range of metabolic functions including maintaining glucose and lipid status. Chromium requirements for older persons are difficult to determine, owing to the scarcity of accurate measures of both chromium status and the chromium content of foods. However, based on limited available intake data, it has been estimated that approximately 12.5 MJ (3000 kcal) of a typical diet would have to be eaten to obtain the minimum estimated safe and adequate daily dietary intake (from 1989 RDAs) of chromium of 50 μg. Such an intake is rare in elderly persons, and at the same time, balance studies have shown that chromium balance can be achieved with dietary intakes less than 50 μg. These findings, coupled with the observation that chromium deficiency is extremely rare in adults of any age, suggest that the current recommended intake may be unnecessarily high for elderly adults.

Fluid Requirements

Water is by far the largest constituent of the body, accounting for between 50 and 80% of the body mass. It is needed for a variety of essential purposes including regulation of cell volume, nutrient transport, waste removal, and temperature regulation. Water is lost from the body in urine, feces, and 'insensibly', i.e., in sweat and transepidermal routes, and so must be replaced. To replace essential water losses, water needs to be consumed at an estimated rate of 0.24–0.36 ml kJ^{-1} (1–1.5 ml $kcal^{-1}$) of energy expended. Thus, an elderly person with an energy requirement of 9200 kJ per day (2200 kcal per day) will need 2.2–3.3 l per day of water. As shown in Figure 1, approximately half of this water intake usually comes from consumed liquids, and the remainder comes from water contained in food and the metabolic water derived from the oxidation of macronutrients in the body.

Water intake is a particular concern in elderly persons. This is not because water requirements are higher in this group (in general, requirements are not higher unless urine concentrating ability is decreased) but because the ability to sense a need for water appears to be blunted in older individuals, with the consequence that they do not feel thirsty when they should and so are at risk of dehydration from inadequate fluid intake. This is particularly true when water requirements are increased – for example, in hot dry weather and when the body temperature is increased during a fever. For this reason, elderly persons should be given the advice that they should not rely on thirst to determine when they drink but instead should have a meal and snack plan that routinely incorporates one or more drinks.

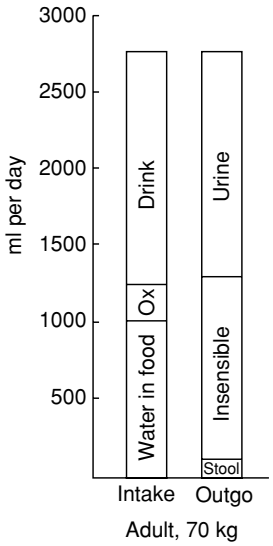

Figure 1 Routes and approximate magnitude of water intake and loss without sweating. Ox, oxidation. Adapted from *Recommended Dietary Allowances* (1989) 10th edn. © National Academy of Sciences. Washington, DC: National Academy Press, with permission.

Appropriate Food Choices for the Elderly

Elderly persons have substantially inadequate intakes of many of the micronutrients, as judged by the adequacy of reported dietary intake in a number of national surveys. This finding may be due, in part, to the inaccuracy of many types of dietary records, especially in elderly persons. However, corroborating evidence from studies of circulating nutrient levels and enzymatic tests of biochemical deficiency suggest that micronutrient inadequacies are widespread in the elderly population. Decreased energy requirements, a reduced ability to accurately regulate food intake, and social factors all contribute to the poor dietary intake of elderly persons. This, coupled with an equivalent or even increased requirement for micronutrients, necessitates a very different dietary pattern in elderly persons with increased emphasis on consumption of fruits, vegetables, legumes, fish, lean meat, low-fat dairy products, and whole grains and decreased intake of highly processed foods with a relative deficiency of micronutrients. Recently, a modified food guide pyramid for healthy individuals aged 70 years and older was developed by Tufts University scientists, placing greater emphasis on the consumption of nutrient-dense foods, high-fiber foods, and water. The modified pyramid also encourages the use of calcium, vitamin D and vitamin B_{12} supplements. While dietary supplements remain a viable alternative to the provision of micronutrients to the elderly population, this should not be a preferred option because supplements do not yet replace the wide range of biologically active food constituents, many of which will probably come to be recognized as playing important roles in the prevention of chronic disease.

See also: **Ascorbic Acid**: Properties and Determination; **Calcium**: Properties and Determination; **Chromium**: Properties and Determination; **Cobalamins**: Properties and Determination; **Energy**: Intake and Energy Requirements; **Folic Acid**: Properties and Determination; **Protein**: Requirements; **Retinol**: Properties and Determination; **Vitamins**: Overview; **Vitamin B_6**: Properties and Determination; **Water**: Structures, Properties, and Determination

Further Reading

Baik HW and Russell RM (1999) Vitamin B_{12} deficiency in the elderly. *Annual Review of Nutrition* 19: 357–377.

Chernoff R (1994) Thirst and fluid requirements. *Nutrition Reviews* 52 (supplement 2): S3–S5.

Garry PJ and Vellas BJ (1996) Aging and nutrition. In: Ziegler EE and Filer LJ (eds) *Present Knowledge in Nutrition*, pp. 414–419. Washington, DC: ILSI Press.

Institute of Medicine, Food and Nutrition Board (1997) *Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride*. Washington, DC: National Academy Press.

Institute of Medicine, Food and Nutrition Board (1998) *Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B-6, Folate, Vitamin B-12, Pantothenic Acid, Biotin, and Choline*. Washington, DC: National Academy Press.

Institute of Medicine, Food and Nutrition Board (2000) *Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids*. Washington, DC: National Academy Press.

Jacques PF, Selhub J, Bostom AG, Wilson PWF and Rosenberg IH (1999) The effect of folic acid fortification on plasma folate and total homocysteine concentrations. *New England Journal of Medicine* 340: 1449–1454.

Levine M, Rumsey SC, Daruwala R, Park JB and Wang Y (1999) Criteria and recommendations for vitamin C intake. *Journal of the American Medical Association* 281: 1415–1423.

National Research Council, Subcommittee on the 10th Edition of the RDAs, Food and Nutrition Board, Commission on Life Sciences (1989) *Recommended Dietary Allowances*. 10th edn. Washington, DC: National Academy Press.

Roberts SB, Fuss P, Heyman MB *et al*. (1994) Control of food intake in older men. *Journal of the American Medical Association* 272: 1601–1606.

Roberts SB and Hays NP (1998) Older people (a) nutritional requirements. In: Sadler M, Caballero B and Strain S (eds) *Encyclopedia of Human Nutrition*, pp. 1466–1473. London: Academic Press.

Russell RM, Rasmussen H and Lichtenstein AH (1999) Modified food guide pyramid for people over seventy years of age. *Journal of Nutrition* 129: 751–753.

Russell RM and Suter PM (1993) Vitamin requirements of elderly people: an update. *American Journal of Clinical Nutrition* 58: 4–14.

WHO (1985) *Energy and Protein Requirements*. Report of a joint FAO/WHO/UNU expert consultation. Technical Report Series 724. Geneva: World Health Organization.

Wood RJ, Suter PM and Russell RM (1995) Mineral requirements of elderly people. *American Journal of Clinical Nutrition* 62: 493–505.

AGITATION

K Niranjan, University of Reading, Reading, UK
A A P de Alwis, University of Moratuwa, Moratuwa, Sri Lanka

Background

Food materials – whether individual or prepared meals – are invariably 'multicomponent' systems. Agitation is both a means of rendering food systems more homogeneous and a means of dispersing further components into existing food systems. Agitation can be used to bring about physical and chemical changes in the system. Under the influence of shear and extensional forces generated by agitation, the rates of heat and mass transfer can be improved during processing. Agitation also enhances the rates of biochemical reactions by insuring that reactants and enzymes are brought into intimate contact. Typical applications of agitation in food processing are (1) blending of soluble solids and liquids, (2) dispersing flavorings, preservatives, colorings, and dressings, (3) disintegrating and comminuting animal and vegetable matter, (4) hydrolyzing and reconstituting dried substances, (5) reducing agglomerates, (6) emulsification, and (7) aeration or de-aeration.

The principles of agitation as applied to food processes, as well as associated problems, are discussed here. In addition, the common types of agitation systems used for different applications are illustrated.

The *characteristic features* of food agitation and associated problems are as follows:

1. Food systems cover a whole spectrum of materials, from dry free-flowing materials through thin and viscous liquids to slurries, pastes and doughs. Agitation of these materials involves many complex interfaces and free surfaces.
2. Flow properties of the components and those of the mixture at any stage during mixing are both very complex and time-dependent. This is further complicated in situations that involve simultaneous chemical reactions.
3. The shear stress generated by an agitator is very important in a food system: processes for making fine dispersions and emulsions need a very high shear, whereas others, such as mixing nuts and chocolates or whole fruit into yogurt, need a low shear.
4. Many food products contain particles of different sizes, some of which are fragile, and these particles may have to be dispersed into viscous liquids such as sauce. Agitation of such systems, to produce uniform product composition, is very difficult.
5. Segregation of particles often occurs when blended products are discharged from mixing vessels. During the production of chilli con carne, for example, beans that have been dispersed uniformly by the agitator aggregate during discharge. This can cause valve blockage, as well as inconsistencies in the packaged product. It is not uncommon to observe significantly higher coefficients of variation between packs than between samples taken from the mixer.
6. Mixing of food materials can involve the addition of high-value components, in very small amounts, into bulk materials. This process is particularly difficult when mixing into dry solids, as well as in the production of dough and batters.
7. It is also difficult to monitor and control the extent of mixing in many food systems, since the end point of agitation occurs some time after agitation has ceased. This is typical of processes involving gas inclusion, crystallization, and texturization.
8. A wide variety of agitating systems are available for use in food materials. Regardless of the size or shape of the mixer, its design features must insure hygiene, as well as provide for in-place cleaning and, possibly, sterile operation. Most equipment is therefore available in crevice-free, polished stainless steel. (*See* **Plant Design**: Designing for Hygienic Operation.)

The principles of agitation can be best illustrated by considering mixing phenomena in a single liquid phase.

Agitation of Liquids

Agitation of liquids can be discussed using stirred-tank agitators as examples, since these devices are not only fairly well understood but also widely used. A stirred-tank system consists of an impeller mounted on a shaft; the impeller rotates, immersed in a liquid contained in the tank. Mixing and dispersion are accomplished by dissipating mechanical energy through the impeller. Figure 1 shows the configuration of a typical stirred-tank system: the dimensions given in the figure are often considered to be 'standard', since most design information is available only for this geometric configuration; however, this configuration is not necessarily the best for any given application. When the height of liquid in a vessel is

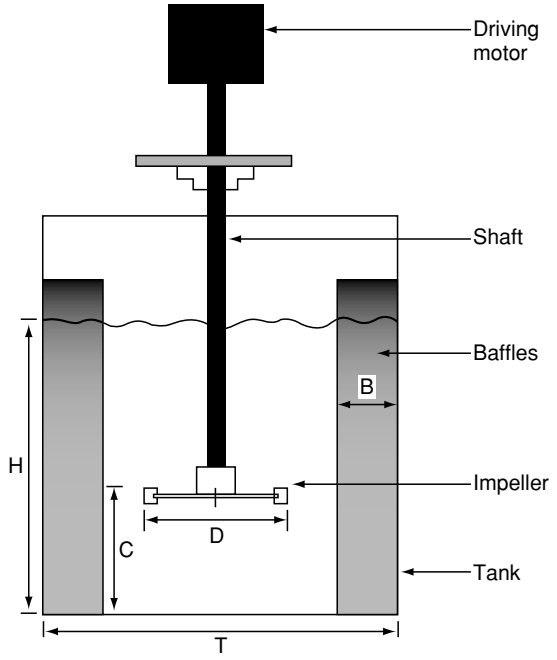

Figure 1 Configuration of a standard agitated, cylindrical stirred tank of diameter *T*, fitted with four baffles, each of width *B*. (Two baffles fitted diametrically opposite each other are shown in the figure; the remaining are fitted along the perpendicular diameter.) Standard ratios: H/T; $D/T=1/3$; $B/T=1/10$; and $C/T=1/3$. Reproduced from Agitation and Agitator Design, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

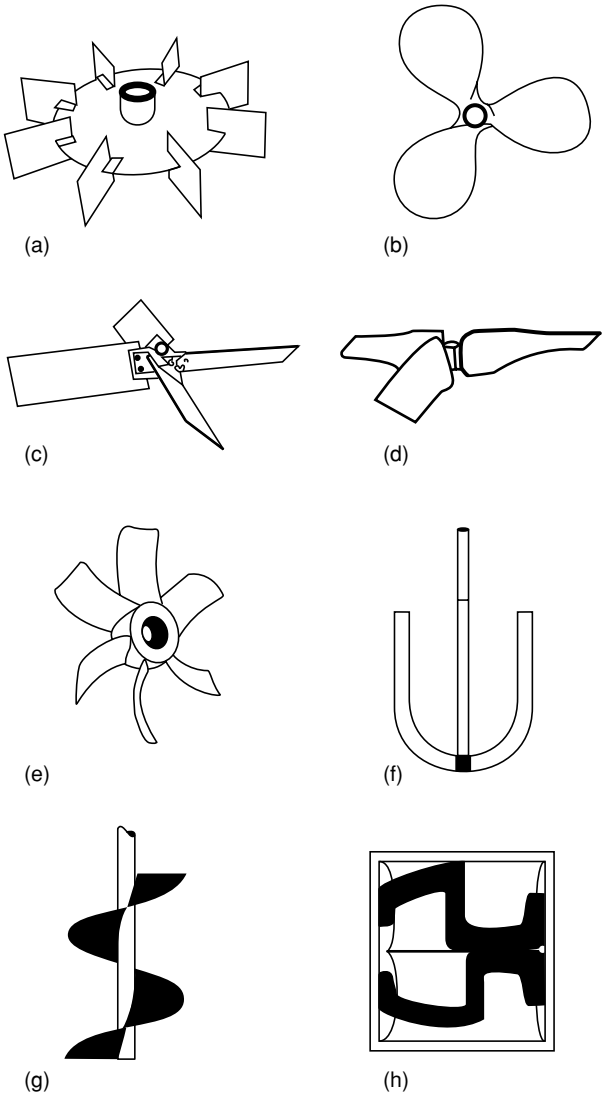

Figure 2 Types of impeller: (a) disk, or Rushton, turbine; (b) propeller; (c) pitched-bladed; (d) hydrofoil; (e) convex-bladed mixed-flow; (f) anchor; (g) helical ribbon; (h) Sigma blade. Reproduced from Agitation and Agitator Design, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

greater than 1.5 times its diameter, it is advisable to use more than one impeller, generally fitted on the same shaft. Most mixing vessels used for low-viscosity liquids (less than 1 Pa s) are fitted with baffles which prevent the formation of a vortex[1]. Baffles also help to establish top-to-bottom circulation of fluid, which moves the fluid to all parts of the vessel. In viscous liquids, unbaffled tanks can be used, since gross vortices are not readily formed.

Mechanically agitated devices are useful for mixing liquids of any viscosity up to 750 Pa s, depending on the size and shape of the impeller. Impellers are often classified according to the direction in which they discharge liquid. The disc turbine, shown in Figure 2a, is a *radial-flow* impeller, since it discharges liquid in the radial direction as it rotates (see Figure 3a). However, a propeller (Figure 2b) is an *axial-flow* impeller (see Figure 3b). Pitched-bladed turbines (Figure 2c) and most of the recently developed impellers, such as the hydrofoil (Figure 2d) and the convex-bladed mixed-flow impeller (Figure 2e), are known as *mixed-flow* impellers, since they develop flows that have components of significant strengths in radial as well as axial directions. Figure 3c shows the flow pattern generated by a mixed-flow impeller.

The above-mentioned impellers are often used in vessels conforming to the standard geometry described in Figure 1. While dealing with viscous Newtonian liquids, or with nonNewtonian systems, which

[1]Vortex is a swirling motion produced when an axially mounted agitator, regardless of type, is used to stir a low-viscosity liquid in an unbaffled vessel. The centrifugal force acting on the rotating liquid raises the level of the liquid at the wall and lowers it at the shaft. Sine the resulting liquid flow is predominantly tangential, mixing is very poor.

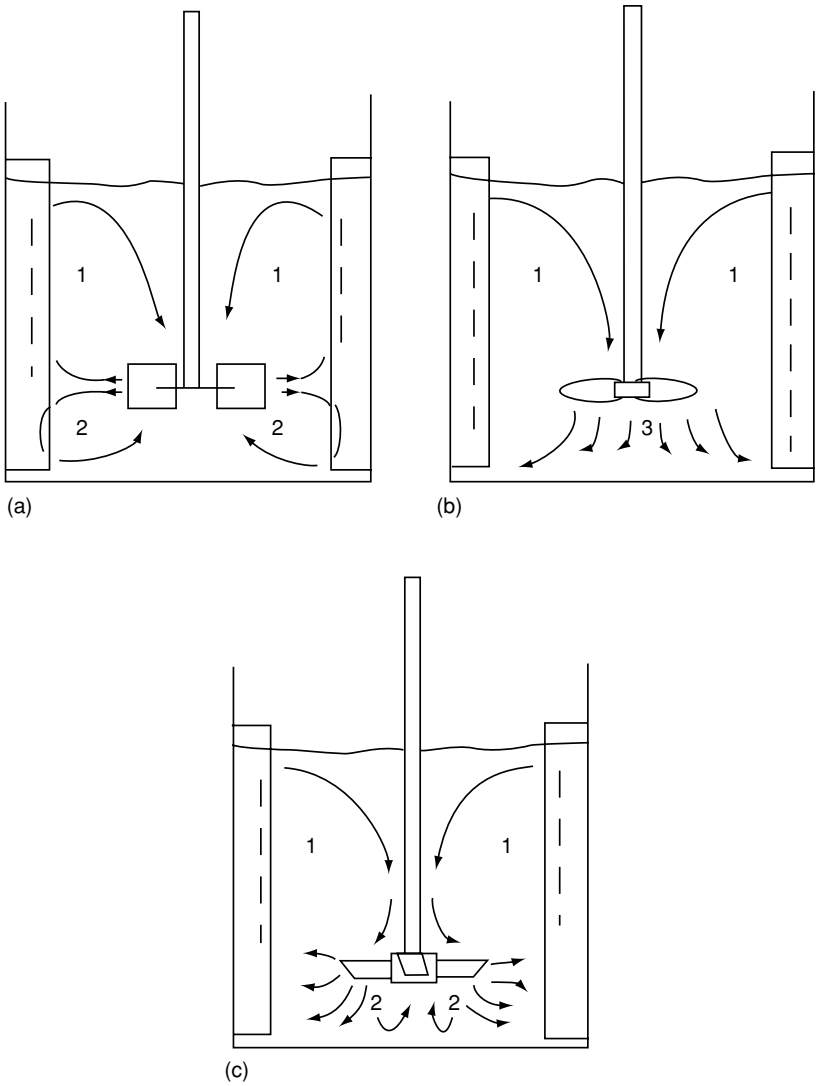

Figure 3 Flow patterns with different impellers: (a) radial-flow; (b) axial-flow; (c) mixed-flow. (1) Upper loop; (2) lower loop; (3) discharged liquid jet. Reproduced from Agitation and Agitator Design, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

are more prevalent in the food industry, the impeller sizes are much larger, and stirrers that have a close clearance between the impeller body and the vessel walls are used. In pseudoplastic systems (fluids exhibiting lower viscosity at higher shear rate), anchor impellers (Figure 2f) are often used. Helical ribbons (Figure 2g) are also used for high-viscosity liquids, as well as for suspensions. These impellers sweep through a larger volume of the mixing vessel and facilitate heat transfer between the liquid and the vessel walls. For mixing applications involving paste-like materials (effective viscosity > 1000 Pa s), Z or Sigma blade agitators (Figure 2h) are used. Figure 4 shows an agitator selection chart based on the effective viscosity of the material to be mixed.

Flow Pattern in Mixing Vessels

The liquid flow pattern generated by a rotating impeller can be either laminar (streamline) or turbulent, depending on the value of the Reynolds number, defined for Newtonian liquids as follows:

$$N_{\mathrm{Re}} = (D^2 N \rho)/\mu. \qquad (1)$$

Here, D is the impeller diameter (m); N is the rotational speed (Hz); ρ is the liquid density (kg m^{-3}) and μ is its viscosity (Pa s). Flow is laminar when N_{Re} is < 10 and it is fully turbulent when N_{Re} is > 10 000. In between these two values, the flow is considered to be in the transition regime. For non-Newtonian fluids, the same definition of N_{Re} is valid, except that the

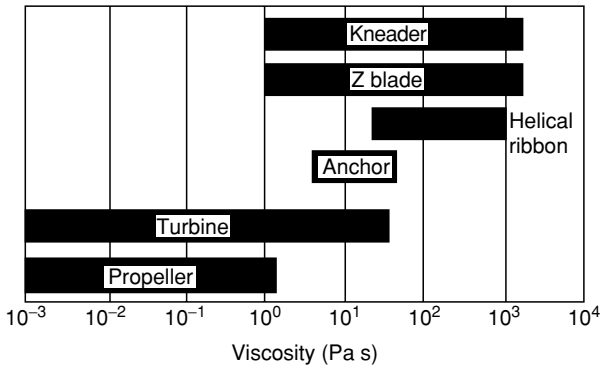

Figure 4 Agitator selection chart based on the viscosity of the material to be stirred. Reproduced from Agitation and Agitator Design, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

denominator is replaced by an 'effective viscosity'. For example, in the case of pseudoplastic liquids following the power law constitutive equation $\tau = k\dot{\gamma}^n$ (where τ is the shear stress, $\dot{\gamma}$ is the shear rate, k is known as the consistency index, and n is known as the power law exponent), the effective viscosity is given by $\mu_{\text{eff}} = k\dot{\gamma}^{n-1}$. It is customary to express $\dot{\gamma}$ in terms of an 'average shear rate' in the agitated vessel, which is known to be proportional to the rotational speed. Thus $\dot{\gamma} = \beta N$, where β is a constant depending on the type of the impeller (β usually takes a value around 10 for most common impellers; see Uhl and Gray 1966). Therefore the effective viscosity of pseudoplastic fluids in an agitated vessel can be expressed as $\mu_{\text{eff}} = k(\beta N)^{n-1}$. With this value of μ_{eff}, the Reynolds number for pseudoplastic fluids can be defined as

$$N_{\text{Re}} = D^2 N\rho / k(\beta N)^{n-1}. \qquad (2)$$

Just as in the case of Newtonian liquids, it is now possible to determine whether the flow pattern in a pseudoplastic fluid is laminar or turbulent or in the transition regime during agitation.

Power Consumption of Impellers

The power consumed by impellers is a very important parameter because it determines the operating cost of running the mixer. For Newtonian liquids, it is related to fluid density and viscosity, rotational speed and impeller diameter. The relationship is graphically expressed by plotting power number (given by $N_{\text{Po}} = P/(\rho N^3 D^5)$, where P is the power drawn by the impeller) against the Reynolds number, N_{Re} (see Figure 5). For any given impeller, the power number decreases with Reynolds number for low values of N_{Re}; for higher values, it is independent of N_{Re}. It may also be noted that, for a given impeller and for low values of N_{Re}, there is no difference in the N_{Po} versus N_{Re} relationship between baffled and unbaffled vessels. However, for high N_{Re} values, baffled vessels consume significantly higher power. Similar relationships between N_{Po} and N_{Re} are also valid for nonNewtonian fluids, provided that the effective viscosity is appropriately defined. Typically, to blend two low-viscosity liquids ($\mu < 0.1$ Pa s), a power dissipation of $0.2\,\text{kW m}^{-3}$ is required; for high-viscosity liquids, the specific power dissipation could be 10 times higher.

Blending of Miscible Substances into Liquids

The *degree of mixing* is a very important indicator of the extent of homogeneity achieved during blending operations. The performance of an agitator is often characterized by the time and energy required to attain a given degree of mixing. In practice, a relatively low, but known, volume of a miscible tracer is added to the contents of the mixer and the time required for the tracer concentration in the mixer to attain 95% of its final estimated concentration is measured; this is known as mixing time. Allowable limits are often placed on the composition variances of blended products. Other statistical methods are also available to describe product uniformity, which can be used for controlling the operation, as well as for testing mixing equipment.

Blending of solute with nonNewtonian liquids (or to result in the formation of nonNewtonian solutions) could pose special problems. *Bingham plastic* fluids (fluids exhibiting shear only when the applied stress exceeds a threshold value known as *yield stress*), when agitated by small impellers, tend to form a cavern around the impeller, where the shear stress is high; there is little or no agitation in the regions remote from the impeller. A different kind of problem can be encountered with certain *viscoelastic* substances, such as those having a consistency similar to a thick sauce or jelly: these substances develop 'normal forces' as a result of elastic properties, in addition to the usual shear forces caused by viscous effects. The normal forces oppose the formation of vortices, and, instead of the liquid level falling around the rotating shaft as described earlier, it rises and climbs up the shaft. The phenomenon is known as the *Weissenberg effect*.

Agitation Equipment for Single-phase Systems

Apart from the impeller-agitated systems described above, various other types of agitating devices are also available. Most of these devices can be operated either batchwise or continuously.

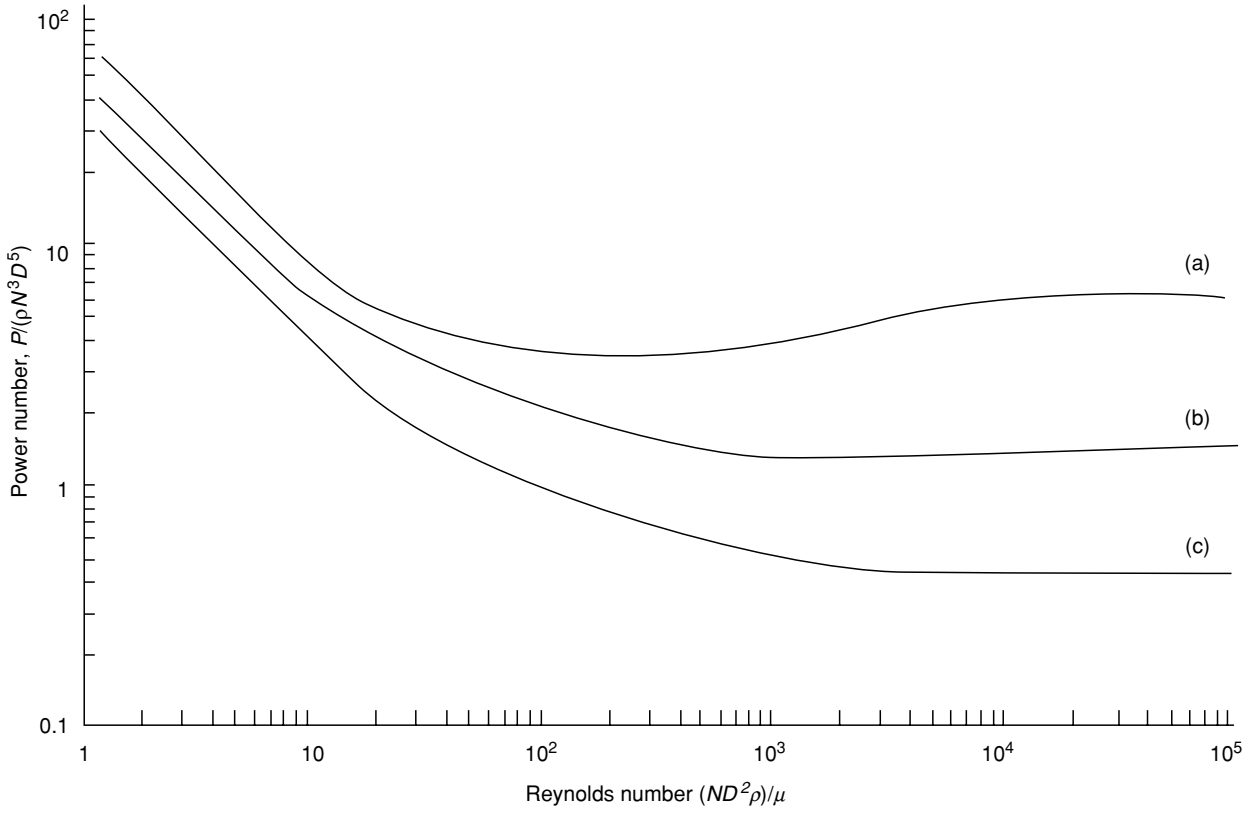

Figure 5 Reynolds number versus power number for some typical impellers – (a) six-bladed turbine, (b) pitched-bladed turbine, (c) propeller – in standard baffle tanks. Reproduced from Agitation and Agitator Design, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

When a blending operation demands or needs only low specific power dissipation levels (e.g., the blending of miscible liquids), a jet mixer can be used. In this device, a continuous recycle of the contents of a vessel is established by an external pump. The stream discharged by the pump re-enters the vessel with sufficient kinetic energy to cause mixing.

Motionless mixers or *static mixers* are a relatively recent development for continuous mixing, especially of viscous liquids. In this type of device, liquid is continuously pumped through conduits fitted with stationary mixing elements. These elements, which are available in various shapes, continuously interchange fluid elements between the walls and the centre of the conduits. The power consumption for agitating action in such devices is simply the power delivered by the pump to the fluid, in order to transport it across the length of the conduit.

Multiphase Agitation

Most practical situations relevant to the food industry involve agitation of several phases, e.g., dispersion of an immiscible liquid into another to form an emulsion, dispersion of air into liquids and pastes, and the mixing of powders (which always involves air and solid phases), pastes and suspensions. The mechanisms of multiphase mixing, as well as the equipment used, are complex, and these areas ought to be the focus of present and future research.

Liquid–Liquid Agitation

Immiscible liquids are agitated for either of the following purposes:

1. To produce a uniform droplet dispersion of one phase (known as the *dispersed* phase) into another (referred to as the *continuous* phase), which would be stable over a period of time, sufficient to allow, for example, mass transfer to occur (as in the case of liquid–liquid extraction); the dispersion is subsequently separated by spontaneous coalescence.
2. Alternatively, and more often, to produce dispersions that are stable over a very long period of time. Such dispersions, known as emulsions, are the most common end products of many agitated liquid–liquid systems. Mayonnaise, salad creams,

can greasing emulsions, and confectionery fillings are all examples of this type of product.

Principles The properties of liquid–liquid dispersions (including emulsions) are related to droplet size distribution, volume fraction of dispersed phase, interfacial tension, and attractive and repulsive forces between droplets. In the production of emulsions, coalescence suppressing agents known as emulsifiers are invariably added; these act by lowering the interfacial tension. Liquid–liquid dispersions are primarily produced by turbulent eddies in the continuous phase, which shear the other phase into droplets. The power dissipated through the agitating device is responsible for the creation of turbulent eddies in the continuous phase. The maximum droplet diameter produced is known to be proportional to the following term:

$$d_{\max} \propto \frac{\sigma^{0.6}}{\rho_c^{0.2}(P/V)^{0.4}},$$

where σ is the interfacial tension, ρ_c is the density of the continuous phase, and P/V is the power dissipated per unit volume of the dispersion.

Practice Mass transfer operations such as liquid–liquid extractions are usually carried out in impeller-agitated vessels; a disc turbine is often the preferred choice for this purpose. Alternatively, spray and packed columns can also be used.

Emulsification is probably the most complicated mixing operation, since the nature of the final product varies greatly in character depending on the method of preparation. The order of addition of the components, as well as the rate at which they are added, can significantly affect emulsion quality. In a typical batch process, an immersion emulsifier is lowered into the aqueous phase (in the case of oil–water emulsions). The various components are drawn into the working head, subjected to very high shear to form the dispersion, and finally the emulsion is expelled. Such devices are described as 'dispersion units.' Different designs of working heads are available, and these can simply be fitted to the end of the frame; care must be taken to select the most appropriate head for any given operation. Emulsification can also be accomplished using in-line static mixers of the type shown in Figure 6. (*See* **Emulsifiers**: Organic Emulsifiers.)

Liquid–Solid Agitation

The agitation of liquid–solid mixtures is carried out in a variety of situations. Dissolving solids into liquids is generally not difficult when the solids are readily soluble, e.g., salt and sugar in water. However, the role of agitation is crucial in dissolving substances such as carboxymethylcellulose, alginates, and gums. In such cases, the rate of dissolution is slow, owing to the low solubility of these substances. The kinetics of this process is further complicated by variations in suspension rheological properties occurring as dissolution progresses. Care must be taken, especially in continuous processes, to provide adequate residence time for dissolution as well as for chemical reactions (e.g., hydrolysis) to occur. In all such situations, it is important to insure that liquids are in a well-agitated state before commencing solid addition. When reconstituting substances such as milk and dried egg powder, the rate of powder addition is important: if the powder is added too quickly, it tends to agglomerate, thus prolonging process times.

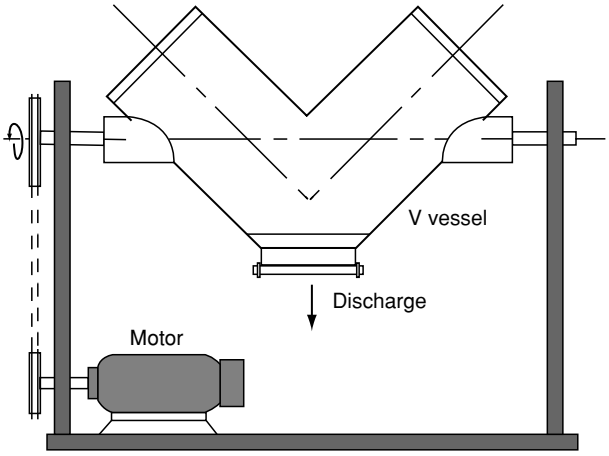

Figure 6 V-shaped tumbler mixer for solid–solid mixing. Reproduced from Agitation and Agitator Design, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

The main difficulty with operations such as dispersing flavorings, spices, coloring, or preservatives is that the quantities to be dispersed are very small in relation to the bulk. In such situations, it is advisable, where possible, to prepare a concentrated dispersion first, and then to mix it with the bulk.

Agitation accompanied by disintegration and comminution of substances such as meat, fruits, and vegetables may be necessary in the manufacture of sauces, ketchups, and infant foods. For each process, a dispersion unit fitted with an appropriate working head, may be used.

In addition to the aforementioned situations, solid–liquid agitation may also be encountered during fermentations (e.g., suspensions of cell cultures and growth media), in holding vessels used as premixers, and in crystallizers. In these situations, the main

objectives of agitation are to keep solids in suspension and to disperse them uniformly so that any draw-off would have an identical solid concentration. Impeller-agitated devices, described earlier, are generally useful for these purposes; propellers, in particular, are known to be efficient for suspending solids in low-viscosity liquids. For a given impeller and solid–liquid system, there is a minimum rotational speed below which the impeller does not produce sufficient 'lift force' to suspend the solids. The threshold rotational speed, in general, depends on the size, shape and density of the solid particles, the solid concentration, the density and viscosity of the liquid, and the geometry of the mixing vessel. In food systems, the density difference between the particles to be suspended and the liquid phase is generally not high. It follows that the minimum rotational speed for solid suspension is not high. Nevertheless, it is quite problematic to ensure homogeneity after the particles have been suspended, especially when the particles are sensitive to collision and high shear. While information on threshold suspension speeds is readily available, the number of studies on the homogeneity of such dispersions in mixing vessels is very limited.

Air–Liquid Agitation

During the agitation of food products, it is generally desirable to exclude air, since entrained air can cause spoilage during storage. However, there are certain situations where aeration is desirable: aerobic fermentations generally demand a continuous supply of air; processes such as cream whipping and ice-cream preparation involve air inclusion. Both aeration and deaeration are therefore important to food processing.

Aeration, whether desirable or not, can occur from the surface of an impeller-agitated vessel. The action of the impeller induces circulation and turbulence in the liquid. When strong eddies are generated at the surface, air is entrained from the head-space to form bubbles, which are then dragged into the bulk by circulation currents. These bubbles are stabilized by surface-active agents that are invariably present in food systems, and the two phases coexist. It has been noted that surface aeration occurs above a minimum impeller speed, and it is possible to estimate its value for a variety of systems. If aeration is detrimental to product storage, then either the impeller speed should be less than the minimum speed for surface aeration, or agitation should be followed by de-aeration.

Aerobic fermenters invariably need a continuous supply of air to sustain microbial growth. The main effect of sparging air into an impeller-agitated tank is to lower its power consumption. The reduction in power is a consequence of the formation of stable air cavities behind the impeller blades; the extent of the reduction depends on the size and shape of the cavities. At a given impeller speed, an increase in air flow rate results in an increase in cavity size. However, the cavity size cannot increase indefinitely: once the cavities have grown to their maximum size at a given flow rate, further increases in air flow rate cause some of the excess air to bypass the cavities and hence the impeller blades. The cavities, in certain cases, may also coalesce. As a consequence, the impeller virtually stops pumping, and the phenomenon is described as 'impeller flooding.' An impeller-agitated fermenter operates well away from the flooding point. In practise, the disc turbine of **Figure 2a** is the preferred impeller used in aerobic fermentation, mainly owing to dispersion ability. This impeller, however, suffers from certain drawbacks: it is a high-shear impeller, and therefore causes shear damage to certain substances, its power consumption is very high, and, more importantly, a large fraction of the supplied power is dissipated in a relatively low volume fraction of the vessel. These disadvantages have been found to be critical to several processes, and there is an attempt to use alternative devices, such as the hydrofoil impeller of **Figure 2d**.

Agitation of Particulate Material

Particle mixing is an extensive food processing operation used for mixing materials that include flour, sugar, dried milk, salt, flavoring materials, cereal flakes, etc. Wide differences among properties such as particle size, shape, density, and surface characteristics (e.g., frictional and electrostatic), make particle mixing a difficult and a complex operation. The process can be further complicated in food systems by a high moisture content, friability, complex flow properties, and agglomeration or segregation. The desired end point of solid-phase mixing is the attainment of a truly random distribution.

The following differences between mixing in particulate systems and fluid systems have been recognized:

1. There is no particulate motion equivalent to molecular diffusive transport in liquids and gases. Thus, when two miscible liquids or gases are in contact with each other, complete mixing eventually occurs. However, blending of particulate materials cannot occur without some input from external energy.
2. Unlike fluids, mixing of particles is reversible, i.e., mixed solids, on storage, tend to segregate, primarily because of size differences; even marginal differences of 15–20% can cause 'unmixing.' Differences in other properties, such as density and

shape, can only accelerate the process of unmixing. It has been recognized that, in general, heavier, smaller, or smoother and rounder particles tend to percolate and sink through lighter, larger or jagged particles.

3. Mixing in liquids and gases is far more intimate than in solids. The ultimate mixing elements of a particulate mixture have a coarser texture, and are of a poorer quality.

Particulate materials may be broadly classified into two groups: (1) free-flowing or cohesionless powders and (2) cohesive powders. Cohesionless powders are mixed by convective transport, surface mixing and/or interparticle percolation; mixing and segregation occur simultaneously. Particles in cohesive powders move in clusters.

Mixers used for solids are normally batch type. Solid mixers can be broadly classified as: (1) tumbler mixers, (2) convective mixers, and (3) hopper mixers. A V-shaped tumbler mixer is shown in Figure 6. The vessel is typically filled to about half its volumetric capacity and is rotated on an axis between bearings, thus causing the particles to tumble and roll over each other. In the case of convective mixers, the mixing vessels are stationary, and the solids are agitated by impeller blades, usually of a ribbon or screw type. Cohesionless powders can also be mixed by discharging a mixture of the components through hoppers. Radial and axial mixing is achieved by returning the discharged material to the top of the hopper and repeating the process until the solids are completely homogeneous.

Mixing of Doughs and Batters

Dough and batter, used in the making of bakery products, are formed by a series of agitation-induced interactions of such diverse components as water, flour, lipids, enzymes, salt, sweeteners, yeast, air, and oxidizing and reducing agents. When classified on the basis of the desired agitating action, most bakery products fall into two broad categories, *extensible doughs*, such as those used to make bread or puff pastry, and *flowable* or *friable mixtures*, used in the preparation of cake batters, icings, etc. (*See* **Bread**: Dough Mixing and Testing Operations.)

Mixing of doughs is more complex and energy-intensive. Although continuous processes are now being developed, mixing of doughs is invariably carried out batchwise in one or many stages. Single-stage mixing is mainly adopted by smaller bakeries, and it aims to complete dough development within its duration. However, in a two-stage process, the initial stage has a shorter duration, and it accomplishes blending and flour hydration (or hydration of the gluten proteins). Dough development and strengthening of the gluten network occurs more gradually during the second stage, promoted by repeated folding and stretching action caused by the mixer. An undesirable effect of the intense shearing action, especially in high-speed mixers, is the rapid increase in dough temperature. Provision for simultaneous cooling therefore exists in most mixers. Careful control of the process is necessary to obtain a dough structure optimum for baking purposes: at the right end point, the dough has maximum consistency or minimum mobility; it appears smooth with a dry surface, and its elastic character is optimized. The end point is reached when the dough starts to pull away from the mixer walls. Overmixed dough, however, is sticky and difficult to handle, and its surface possesses a characteristic sheen.

The agitator in a dough mixer is mounted either on a horizontal shaft or on a vertical shaft. Vertical shafts generally rotate about a fixed axis; in some cases, planetary motion is also superimposed. In either case, there is a close clearance between the agitating blades and the mixing vessel, in order to eliminate stagnant regions and build-up of sticky material on the wall.

See also: **Aerated Foods**; **Bread**: Dough Mixing and Testing Operations; **Emulsifiers**: Organic Emulsifiers; **Plant Design**: Designing for Hygienic Operation

Further Reading

Harnby N, Edwards MF and Nienow AW (eds) (1992) *Mixing in the Process Industries*, 2nd edn. Oxford: Butterworth-Heinemann.

Jones RL (1985) Mixing equipment for powders and pastes. *The Chemical Engineer* 419: 41, 43.

Lindlay JA (1991a) Mixing for agricultural and food materials: 1. Fundamentals of mixing. *Journal of Agricultural and Engineering Research* 48: 153–170.

Lindlay JA (1991b) Mixing for agricultural and food materials: 2. Highly viscous liquids and cohesive materials. *Journal of Agricultural and Engineering Research* 48: 229–247.

Lindlay JA (1991c) Mixing for agricultural and food materials: 3. Powders and particulates. *Journal of Agricultural and Engineering Research* 49: 1–19.

Niranjan K, Smith DLO, Rielly CD, Lindlay JA and Phillips VR (1994) Mixing for agricultural and food materials: 5. Review of mixer types. *Journal of Agricultural and Engineering Research* 59: 145–161.

Niranjan K (1995) An apprisal of the characteristics of food mixing. In: Singh RK (ed.) *Food Process Design and Evaluation*, pp. 47–67. Lancaster, PA: Technomic.

Rielly CD, Smith DLO, Lindlay JA, Niranjan K and Phillips VR (1994) Mixing for agricultural and food materials: 4. Assessment and monitoring of mixing systems. *Journal of Agricultural and Engineering Research* 59: 1–18.

AIDS *See* **HIV Disease and Nutrition**

AIR CLASSIFICATION

Contents

Principles of Operation

P Fedec, POS Pilot Plant Corporation, Saskatoon, Saskatchewan, Canada

Background

In simplistic terms, air classification is a means of using air to effect a dry separation of objects having certain characteristics. These characteristics include physical properties such as size, shape, density, and physicochemical nature. Air classification is an operation that applies the technology to fractionation of nonhomogenous particles, suspended in an air stream, into classes of fairly uniform size based on a common criterion of density or mass. Whereas some segregation of light and heavy particles can be achieved by aspiration where airflow patterns are generated by only a vacuum, true air classification makes use of airflow induced through centrifugal motion with the assistance of a vacuum. Generally, air classifiers will segregate a heterogeneous particulate into two subclasses, one primarily below a targeted particle size (fine fraction) and the remainder above this size (coarse fraction). Air classification becomes a possible alternative when ordinary sieving fails to effect separation below 40 μm. Applications of air classification in the food industry separate particulates in the range of 2–60 μm, seldom exceeding 100 μm.

Factors Affecting Classification

In the ideal case, classification can be best achieved if the particles are uniform and homogeneous, spherical and smooth, and dry and easily dispersible. Practically, very few finely granulated mixtures exhibit these ideal properties. Other features such as particle geometry can affect the tragectory of particles in an air stream. Classification is also affected if some of the particles are porous, allowing fine particles to hide within the pores of the coarser particles.

Dispersibility is a key issue during classification. Materials containing moisture or oil, or exhibiting hygroscopic tendencies, are likely to agglomerate and resist classification. To minimize these effects, pre-classification techniques need to be considered, which may include additional grinding, vacuum drying, extraction by solvent, blending, dispersion, or deagglomeration.

The cell structure of food materials is one of the most important factors for the mechanical separation or dry processing. Application has primarily been in protein and starch separations of legume and cereal flours. Within the cell, the protein forms a matrix in which the starch granules are embedded. During comminution, the breakdown of this matrix into wedge protein or granular protein (and its size or homogeneity) is important and will affect the efficiency of classification (Figure 1). The cell matrix must shatter easily, and in addition, the protein matrix must disintegrate readily. If compound starch granules are present, they must remain intact. A significant number of particles (e.g., starch granules) must be above the selected cut point. Typically, impact milling is the comminution of choice.

Three particle-related factors of primary importance in achieving separation include: particle density, shape, and size. The difference between density of starch (1500 kg m^{-3}) and protein (1300 kg m^{-3}) is too small for density to have a significant role. Shape plays some role; in particular, starch granules flattened by impact milling exhibit increased air drag and unfortunately may concentrate in the fine fraction. Experience has shown that particle size is the predominating influence on the quality of air classification.

The particle size of the starting material is an important factor in that, generally, the particle must

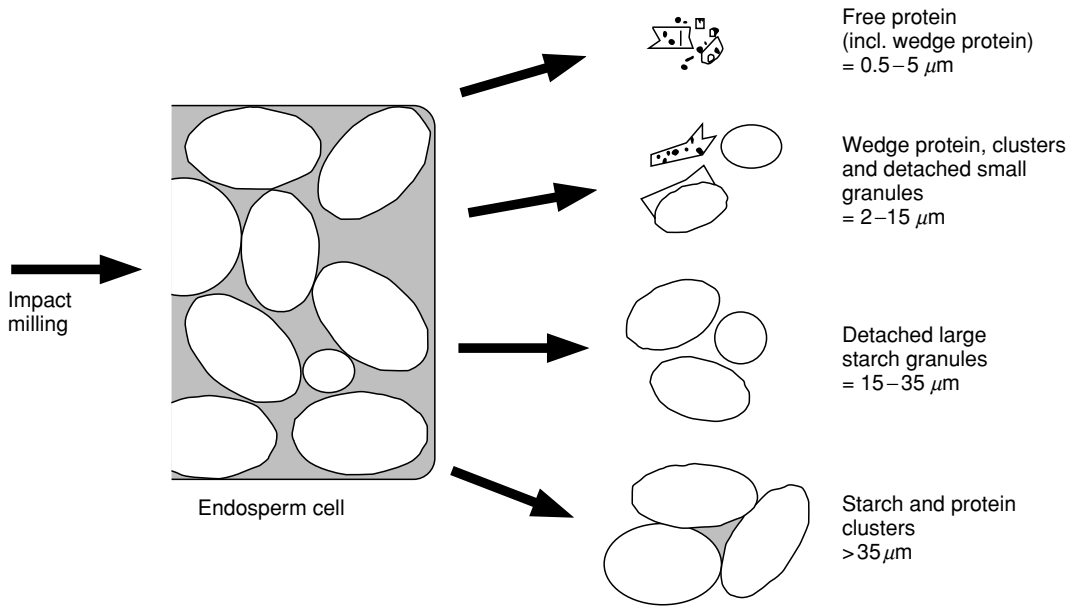

Figure 1 Conceptualized presentation of endosperm cells subjected to milling and subsequent fragmentation.

be less than 100 μm and preferably lower, with a significant number being below the selected cut size. For example, if the granular and wedge protein can be pulverized into aggregates of 2–10 μm in size, while the starch particles remain intact and are in the range of 15–45 μm, separation is possible (Figure 1). Some wedge protein of larger size will spill over into the coarse particle cut, and some small granules will spill into the protein or fine fraction. Products that do not have a significant difference in the sizes of components are not suitable candidates for air classification.

The moisture content of grains and legumes affects the ability to disintegrate the cell contents and dislodge the protein from the starch. Practice suggests that the moisture content for satisfactory comminution preceding air classification should be around 8%. Moisture contents above this value do not lead to an ideal size reduction, but a moisture below 6% can lead to starch granule fragmentation and subsequent spill-over into the fine fraction.

No well-documented research has been carried out on the influence of fat content of the starting material on air classification, but in general, fat contents have to be kept low in order to achieve dispersion in the centrifugal zone and minimize bridging in and around rotors, vanes, and passage-ways. Generally, a 1–2% fat content is the limit, although some results have indicated that flours having up to 7% fat are separable by some classifier models, but prolonged operating times with sustained quality of separation have not been established.

Sharpness of classification is affected by the feed rate, and generally, as rates are elevated above optimal, the sharpness decreases.

Air-classifier Operations

Air-classifier operation is influenced by air-flow dynamics on suspended/dispersed particles of typically less than 100 μm. The term 'cut size' is used in conjunction with discussing the segregation of the two fractions.

If a predetermined particle size were selected (e.g., typically 15 μm in flour applications), a theoretical ideal classification would permit a subdivision of a mixed particle feed stock into two fractions: a fine fraction below the predetermined size and a coarse fraction above it. Figure 2a depicts such a segregation of fine (shaded) and coarse (clear) fractions.

Under practical conditions, such a sharp separation cannot be achieved, and the separation is more realistically observed as shown by the curves in Figure 2b. A single 'cut size' is not achieved; rather, there is a range of particle sizes present in both the fines and coarse fractions, and these overlap in the region of a nominal cut size. The less the overlap, the better is the separation efficiency.

The two major forces affecting air classification are the drag or frictional force exerted on the particle by the air flow and an inertial force exerted by the accelerated motion of the particles. The latter is generally achieved by a centrifugal field or rotational air flow through use of a rotor. A typical example of forces

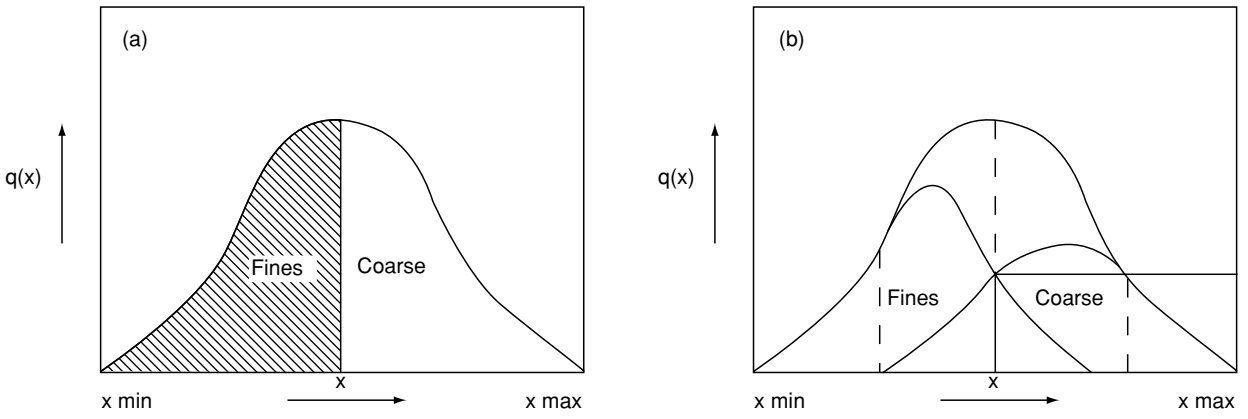

Figure 2 Classification phenomena: (a) an ideal classification; (b) a real classification. Adapted from de Silva SR (1983) Developments in air classifier theory and practice. In: *Institution of Chemical Engineers (Gr. Br.) Symposium Series*, vol. 69, pp. 387–410, with permission.

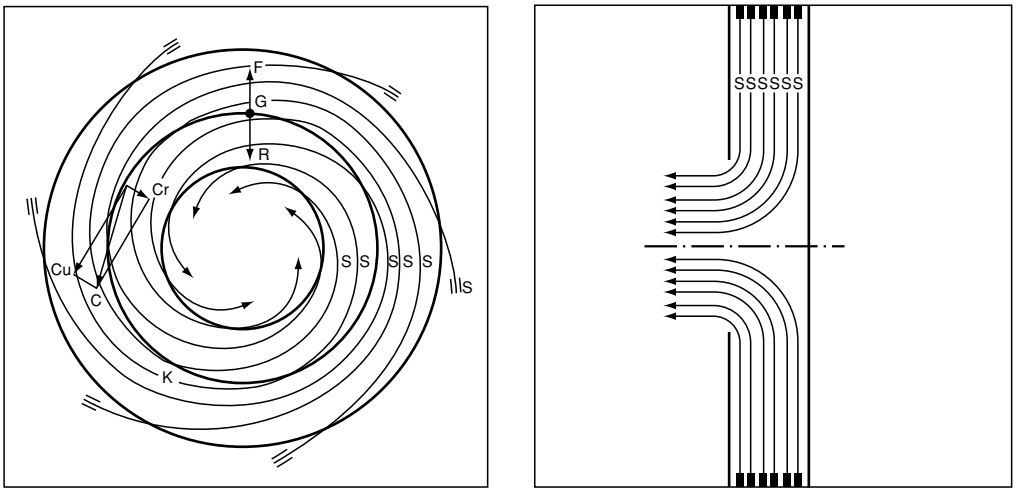

Figure 3 Forces interacting in the spiral flow of an air classifier. From Hosokawa Micron International Inc. Catalogue 31/6e. Summit, NJ with permission.

acting within a classifier is shown in Figure 3. A particle ('G') of a certain diameter and density is exposed to two opposing forces, the centrifugal force (F) produced by a rotating assembly and frictional force (R) produced by an air flow. Additional and more complex mathematical models can be found in articles by Vose and de Silva. As the air flows inwards in a spiral path, particles entrained in the flow are subjected to the opposing forces F and R. Larger particles are dominated by the mass-dependent centrifugal force, whereas small particles are influenced by the frictional force. When the forces are equal, a cut point or particle of a definite size can be established. The cut size can be adjusted by manipulating these forces within the classifier.

Design Styles of Air Classifiers

Many air-classifier variations are on the market. They can be subdivided into types based on their principles of operation. Some of the common types are detailed below.

Gravity Principle

This is the simplest design, has no moving parts, and consists of a zigzag-shaped classifying channel with internal baffling. It is generally used for gross separations, often for dissimilar materials that may be difficult to separate by simple separation techniques such as screening.

Free Vortex Principle

Air classification is achieved by adjusting a set of louvers located at the periphery of the classifying zone, which allows the superimposed radial air flow to enter the zone at an angle. Typical of this group are designs such as Pallmann Galaxy and Alpine Mikroplex. The latter (Figure 4) has been used in a large number of research studies, and significant data have been published. This design incorporates a rotor on a horizontal axis, resulting in a vertical plane of centrifugal force, whereas most others use a horizontal plane that favors greater throughput but with a reduced precision in cut size. The Mikroplex MPS model has been discontinued, but many are still in industrial operation. Feed rates vary from 50 to 5000 $kg\,h^{-1}$, depending on the model size. Classifiers manufactured by Larox operate on the same principle, but the peripheral vanes are not adjustable. The cut size is adjusted by manipulating the air flow produced by a blower.

Forced Vortex Principle

The force is provided by means of a turbine or rotor that disperses the particles into an air stream applied by a suction fan. Variations in equipment design then use air streams to select fine or coarse particles. The Bauer Centri-Sonic was typical of this design and was used commercially for cake flour separation as well as much research, but its precision honeycomb rotor was costly to manufacture, and the equipment has been discontinued. A new breed of classifier, often referred to as the turbine classifier, is now being utilized for fine powder separation (3–150 μm). These may incorporate single or multiple turbine classifier wheels in either vertical or horizontal configuration. The MikroPul Micron Separator illustrates this concept (Figure 5). Other manufacturers using this principle include Hosokawa Micron, ABB Raymond, and Alpine ATP series classifiers, as well as those manufactured by Nisshin, Sturtevant, and Matter and Partner.

Additional but less popular classifier designs used mixed vortex, countercurrent, cocurrent, or crossflow principles. Some are experimental, and others have not been applied to the food industry.

Combined Mill/classifier Units

In the food applications area, no appreciable technological changes to classifier design have been made in

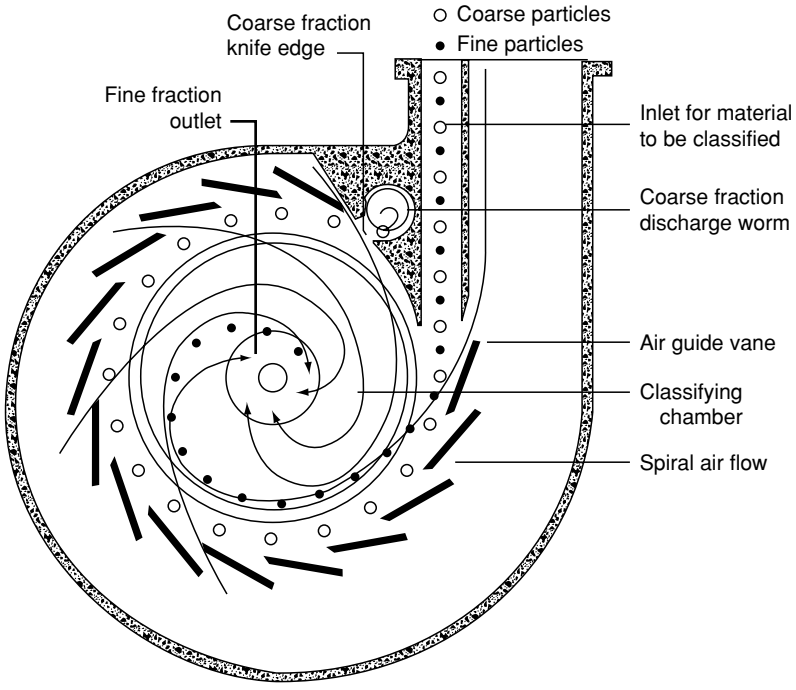

Figure 4 Frontal cross-section view of a Mikroplex MP spiral classifier. ○, coarse particles; ●, fine particles. From Hosokawa Micron International Inc. Catalogue 31/6e. Summit, NJ with permission.

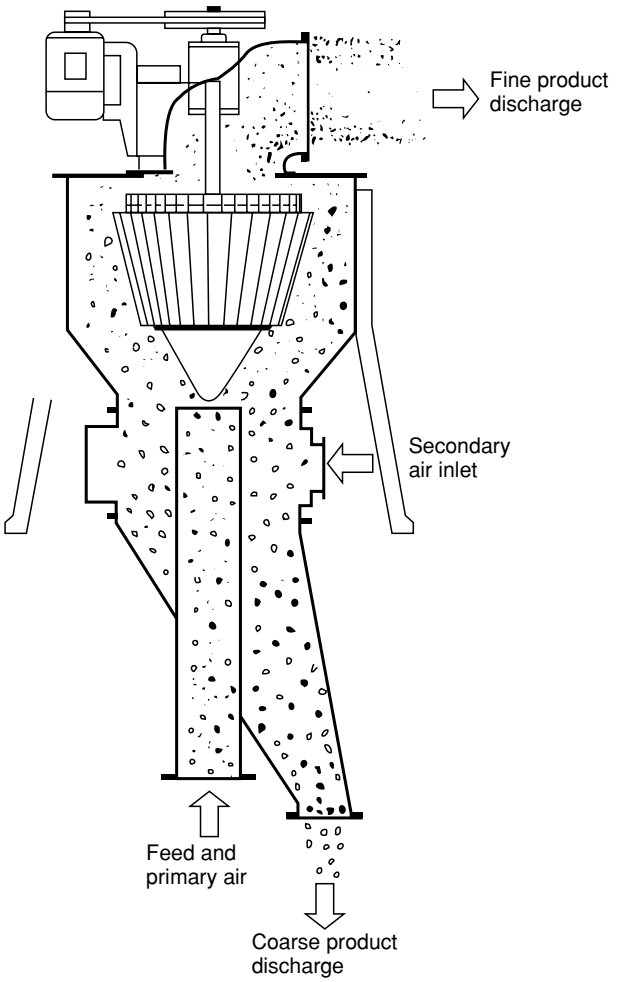

Figure 5 Typical illustration of a turbine-type fine powder classifier (MikroPul Micron Separator MS). From Hosokawa Micron International Inc. (1988) *MikroPul Micron Separators Cat. No. HMS-1 5M-4/88*. Summit, NJ with permission.

recent years, primarily because the profit margins are low in this industry, thus providing little incentive for spending resources on further development and refinement.

To reduce operating costs, interest has focused on equipment that combines the required fine grinding with classification. One such unit is the Mikropul ACM Air Classifier Mill. The basic operation includes a high-speed impact mill through which an air stream is drawn containing the raw material. The product/air mixture enters the milling chamber, after which the product is then accelerated by the rotor towards peripheral guide vanes into the classification zone. Here, the particles are acted upon by centrifugal force in a centripetal airstream, and lower-mass fine particles are separated and conveyed out of the mill to a collector. The coarse particles are directed back to the grinding zone for further size reduction. The rotor speed and air flow are adjusted to maximize mill performance in a given application.

Plant System Designs

The typical installation consists of a centrifugal fan generating a vacuum draw to a baghouse collector. An additional scrubber for fine particulates may be included after the baghouse filter, and some designs use an optional cyclone ahead of the baghouse collector. The fines separated by the classifier are drawn into the collector, whereas the coarse particles are usually rejected by one of several ways, either in an auger system, in an air stream to a cyclone, or within a cyclone built into the classifier. The classifier may be fed from a silo of premilled material or may be directly coupled to a fine grinding mill. This direct coupling method has received criticism in that changes in material composition or particle size, owing to stratification, affect the classifier cut fraction and thereby the final product composition, especially if the latter is being packaged directly from the process line. Ancillary equipment includes rotary air

locks for the collectors, filter assemblies, discharge devices, blow-out relief doors, monitoring and control devices, blowers, silencing devices, and pulse jet filter cleaning apparatus.

Some operating facilities, in particular for legumes, have used a double milling, double classification design sequence, where the initial mill is followed by a classifier, the coarse fraction remilled and reclassified, thereby yielding a final coarse fraction and two fine fractions that are generally pooled. This plan strives to increase the quantity recovered as fines, although their quality may be compromised, and to improve the purity of the coarse fraction.

Safety factors that should be considered in an installation include dust inhalation, noise hazard, vibration, and dangers of explosion from fine dusts, which are ignitable under certain conditions.

Operating Conditions

In addition to the features of the feed material as given by the composition, nature of particles, product uniformity, and premilling influences, the variation of cut point, and, subsequently, the quality/content of the separated fractions can be manipulated by classifier variables such as the rotational speed of the rotor or disintegrator, the amount of air flow through the classifying chamber and the velocity of the air flow (both can be altered by either adjusting air intake or the external booster fan), degree or angle of vane setting, which deflects flow, and finally the rate of feed delivered to the classifier. These options vary in both availability and degree of adjustment, depending on the manufacturer of the equipment. Some classifiers allow for easy adjustment of fineness from an external position while the unit is operating, whereas others can be adjusted only after partial disassembly (i.e., such as alteration or adjustment of numbers or sizes of selection plates). However, with proper manipulation of these parameters, the classifier can produce a quality separation of high efficiency and yield.

Supplemental options are available for special applications whereby water cooling may be applied to the outside cone of certain separators to satisfy the needs of the cooling product (prior to packaging) that has developed a temperature rise during the milling and classification sequence.

When selecting systems for an application, there is no substitute for actually trying out a system on the materials intended for classification. Most manufacturers maintain laboratory and pilot plant test facilities, where various types and sizes of classifiers may be tested for separation efficiencies. Some independent contract research centers are also able to conduct exploration and demonstration testing.

See also: **Air Classification**: Uses in the Food Industry; **Plant Design**: Basic Principles; **Separation and Clarification**

Further Reading

Alpine Aktiengesellshaft (1978) *Alpine Mikroplex Spiral Air Classifiers MP, MPS. Alpine Bulletin 31/6e*. Augsberg, Germany.

de Silva SR (1983) Developments in air classifier theory and practice. In: *Institution of Chemical Engineers (Gr. Br.) Symposium Series*, vol. 69, pp. 387–410.

Galk J (1998) Production of fine powder with the air classifier mill ACM or the newly developed multistage impact mill MSM. *Powder Handling Process* 10(1): 11–13.

Hixon L (1992) Sizing up air classifiers. *Chemical Engineering Progress* 88(7): 59–62.

Hosokawa Micron International Inc. (1988) *MikroPul Micron Separators Cat. No. HMS-1 5M-4/88*. Summit, NJ.

Jones CR (1960) New developments in milling processes. Protein displacement. *Milling* 135: 494–498.

Klumpar IV, Currier FN and Ring TA (1986) Air classifiers. *Chemical Engineering* 93: 77–92.

Lauer O and Prem H (1978) Protein enrichment in vegetable products by air classification. *Zeitschrift für Lebensmittel - Technologie und - Verfahrenstechnik* 29: 212–215.

Schubert H and Heideker H-Th (1986) Protein separation from vegetable sources by selective commination and air classification. In: Le Maguer M and Jelen P (eds) *Food Engineering and Process Applications II, Unit Operations*, pp. 293–303. London: Elsevier.

Thomas ND (1999) The MikroPul air classifier mill. *International Sugar Journal* 101: 480–481.

Vose JR (1978) Separating grain components by air classification. *Separation and Purification Methods* 7: 1–29.

Uses in the Food Industry

P Fedec, POS Pilot Plant Corporation, Saskatoon, Saskatchewan, Canada

Background

Air classification is a technology that is based on physical principles, and therefore, the use of chemicals is precluded, and protein denaturation is avoided, as is the formation of artifacts. This dry treatment also avoids the production of polluting effluents and minimizes or obviates drying costs. Two of its negative aspects are that there may be a retention of antinutritional or undesirable components in one or both fractions, and neither of the

fractions produced is 'pure' when compared to those obtained through conventional wet processing; hence, applications for products produced often need to be developed.

Application to Grain Legumes

Considerable research work on the milling and classification of nonoleaginous, starchy legume seeds has been carried out in Canada at the National Research council Prairie Regional laboratory and the University of Saskatchewan, both in Saskatoon. Although a wide number of species were studied for their potential to commercialization, the focus on field peas led to the adoption of this dry separation process by industry. Similar procedures were adopted in France for the production of fababean protein concentrate. Other centers that have also explored fine grinding and air classification in legumes include the Agricultural and Food Research Council Food Research Center and the University of Reading in the UK, the Technical University of Denmark, the Michigan State University in the USA, and the Korean Food Research Institute. Most researchers have employed pin mills for fine grinding of the seed, and a majority have focused on using the Alpine Mikroplex MPS series classifiers, although Alpine Zig Zag and Bauer Centri-Sonic series were also used. Recently, studies have shown that the use of jet air mills has given flour grinds having a mean particle size of $< 40\,\mu m$, which has improved separation of flour into protein and starch fractions.

A flow diagram for the process developed at the Canadian Prairie Regional Laboratory and applied to legumes is shown in Figure 1. Researchers have found that the milling of the seed is of overriding importance to the air-classification process. Efficient size reduction in the initial milled flour should lead to improved separation of the protein and starch components. Examining Figure 1, whole or dehulled seeds are finely ground in a pin mill and the resulting flour classified in a spiral air stream classifier, set to a predetermined cut point, resulting in the production of a fine (protein) fraction (F) and a coarse (starch) fraction (C). After the first pass, some protein bodies still remain attached to the starch granule surface, and some agglomerates remain, having starch granules embedded in a protein matrix. A second milling of the starch fraction disrupts these complexes further, and a second classification generates additional protein (G), albeit at a reduced yield and protein cont ent, and a final starch fraction (E) with low levels of protein.

In a typical example, the two-run process on field peas generated a combined protein fraction of 34.1% yield and a protein content of 56.6% (Figure 1). For fababeans, a protein fraction of 37.2% yield and a protein content of 68.1% was produced. Starch fractions contained 6–7% residual protein. Repeated pin milling and air classification studies show that the starch fraction can be further purified, but even after four runs, the residual protein is still around 3% for field peas. It follows that little is to be gained from a process based on more than two runs, and, depending on economics, it may be sufficient to use only a single-run procedure.

A comprehensive review of the published data would be beyond this article. However, to give the reader an overview of what may be achieved/expected from various legume flours, a summary table (Table 1) has been prepared. By way of caution, it should be noted that the data are drawn from a range of classifier types and operating parameters on starting flours with varying initial protein levels prepared by employing a wide assortment of grinding mills from seed with varying moisture contents. The data are presented as min–max ranges recorded in the literature. All data shown are based on a single impact milling followed by a single air classification, generally using dehulled seeds.

Low fine fraction yields have a correspondingly higher level of protein content as well as a higher protein level in the coarse (starch) fraction. As the yield of fines increases, the protein content of the fine fraction decreases, and the quantity of the coarse fraction will show a decrease. The protein content in the coarse fraction will decrease, and correspondingly, the starch proportion will increase. In all cases, the recovery of protein into the fine fraction is not as efficient as the recovery of the starch into the coarse fraction.

The initial protein content of the legume flour has an influence on the classification result. At the same yield or split ratio, higher protein levels in the starting flour will result in a higher protein content in the fines fraction. Alternatively, if the protein content in the fines fraction is kept constant, a higher yield of that fraction can be expected. Finely milled, defatted soybean meal has been added to Table 1 for comparative purposes and serves to indicate that, unless there is a high proportion of relatively large and uniform starch granules, a shift of protein cannot be accomplished, despite achieving a separation into two fractions.

Comparing the legumes, good separation may be expected with faba, lima, mung, navy, Great Northern beans, lentils, peas, and cowpeas. Especially high protein concentrates have been achieved from finely ground fababean flour. Chickpeas are a poor candidate for air classification, presumably because of the

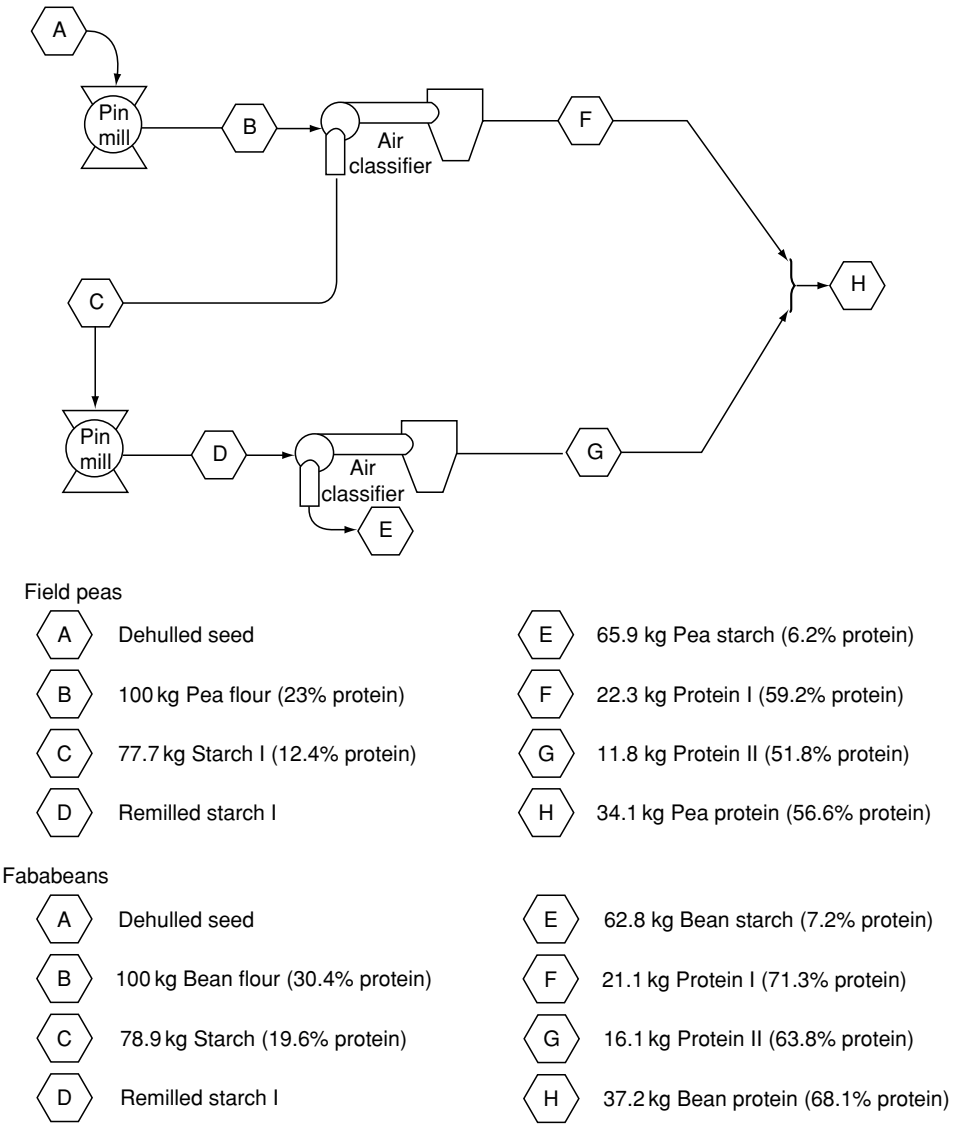

Figure 1 Double-pass milling and air classification sequence for legumes. Adapted from Tyler RT, Youngs CG and Sosulski FW (1981) Air classification of legumes. I. Separation efficiency, yield and composition of the starch and protein fractions. *Cereal Chemistry* 58: 144–148, with permission.

higher oil content in the flour and a high proportion of small starch granules.

The separation efficiency (i.e., the percentage of total protein recovered in the fine fraction or the percentage of total starch recovered in the coarse fraction) is dependent upon moisture at the time the legumes are milled. In peas, an increase in moisture to 14% from 10% reduces the separation efficiency by 20%. As the seed moisture decreases, there is a decrease in the starch yield, in the protein content of the starch fraction and in the protein content of the protein fraction. However, there is an increase in the yield of the protein fraction as well as the protein separation efficiency. It has been suggested that a moisture range of 8–10% is ideal for an impact-milling, air-classification process operation in legumes.

The effect of seed maturity on air classification suggests that the presence of immature seeds, even at high levels, has little effect on subsequent milling and classification. There is concern that green and shrunken seeds may make dehulling difficult and translate to high fiber levels in the starch fraction.

The protein content of the starting flour, and initially of the seed supply prior to milling, has a significant effect on fractionation. In a study of widely varying protein content pea flours, it was noted that the percentage protein in all air-classified fractions

Table 1 Yield, compositional data[a], and separation efficiency of fine and coarse fractions from air classification of milled legume flours[b]

Legume species	*Starting flour (% protein)*	*Fraction yield*		*Protein content*		*Starch content*		*Protein recovery in fines*	*Starch recovery in coarse*
		Fine	*Coarse*	*Fine*	*Coarse*	*Fine*	*Coarse*		
Field pea	20.4–25.3	2.3–53.0	77.7–47.0	62.5–42.6	14.5–6.5	1.4–24.7	60.0–86.2	58.1–88.8	99.8–75.3
Fababean	29.8–33.0	21.1–45.8	78.9–54.2	71.8–60.6	19.6–4.2	0–9.5	57.5–91.1	61.3–92.6	100–95.0
Navy bean	24.5–30.4	19.4–44.6	80.6–55.4	63.8–46.2	21.9–5.7	0.1–6.9	58.0–84.9	49.1–87.1	100–96.7
Lima bean	22.7–22.8	17.7–32.0	82.3–68.0	49.6–39.3	18.7–14.1	0–5.9	56.9–69.6	32.0–55.4	99.1–96.1
Great Northern bean	22.6–24.0	22.5–42.2	77.5–57.8	53.5–45.6	15.6–6.1	0–4.2	51.5–92.3	50.0–87.0	100–96.6
Lentil	22.6–27.1	21.9–44.6	78.1–55.4	64.6–46.5	14.5–3.3	0.2–19.3	65.8–89.7	63.0–91.9	99.2–84.3
Cowpea	23.6–24.4	24.1–33.5	75.9–66.5	51.6–41.8	17.2–14.2	3.2–14.6	67.8–73.7	48.7–60.0	97.0–89.2
Mung bean	25.8–26.5	27.1–29.0	72.9–71.0	61.4–60.4	12.3–11.7	2.0–6.1	67.7–74.0	66.1	96.4–96.3
Chickpea	17.7–19.5	8.5–29.0	91.5–71.0	40.0–28.9	15.5–11.4	3.7–30.3	53.1–62.7	19.2–50.2	94.5–81.4
Soybean	52.5	72.0	28.0	54.2	50.5			74.1	

[a]On a percentage dry-weight basis.
[b]Data are shown as a range of reported values from assorted classifier types and settings.

was positively correlated to the percentage protein in the flour, whereas lipid and cell wall material was negatively correlated. Starch and protein separating efficiencies generally increase with increasing protein content. An unpublished rule of thumb has been applied to evaluating the potential of legume seed for air classification. If the flour is finely milled and a 33% fines fraction yield is sought, the protein content of that fraction is apt to be 2–2.5 times the initial protein value of the flour. To insure uniform products by air classification technology, it is suggested that, in any particular legume being processed, the seed falls within a narrow range of protein content.

Shifting of other components also occurs as a result of air classification. In terms of fiber material present in the flour, the hull fiber tends to segregate into the coarse fraction and the cell wall fiber into the fine fraction. Dehulling may enhance the efficiency of the process somewhat and may be a prerequisite if colored hulls are present. If the coarse (starch) fraction is of lesser importance, dehulling need not be practiced as it only has a small effect on the crude fiber content of the protein fraction.

Legume flours containing α-galactosides, especially raffinose, manninotriose, stachyose, and verbascose, generally show the protein fraction to be 40–90% higher in these components than the initial flour, while the starch fractions become depleted. Legumes with trypsin inhibitors, hemagglutinating factors, saponins, phytic acid, vicine, and convicine will exhibit elevated concentrations of these components in the protein fractions.

The protein fraction from legumes also tends to be enriched in ash and lipid, the latter having undesirable ramifications in storage of the concentrate. The lipids likely occur as a result of the presence of a lipid-rich coating on the protein bodies.

Protein concentrates prepared by air classification may be used as emulsifiers or supplements in meat products and patties, beverages, bread, noodles, and spaghetti. Cowpea flour fractionated into enriched protein and starch concentrates has been applied to produce emulsifiers and starch jellies. There may be a requirement to inactivate some of the antinutritional factors. The starch fractions are increasingly being utilized in the paper industry, especially carbonless paper. Some use of pea starch was made in ore refining, but this has been sporadic. Nevertheless, peas have shown the greatest potential for commercial application of air-classification technology. There are few major nutritional or technological problems except for the 'pea flavor' and, in some cases, a darker green color in some end products. The main obstacles to success in classification stem back to basic agronomics, the guarantee of a year-round supply at a competitive price and of high quality, and establishment of markets for the products.

Application to Cereal Grains

Wheat Flour

The earliest work in application of air classification techniques to food products began in several United States labs (Peoria, Illinois and Manhattan, Kansas) with protein-shifting in wheat flours. Classification has generally been less efficient than in legume flours because of the high proportion of small starch granules that cannot be readily separated from the fine protein particles. Approximately 81% of the granules fall below 7.5 μm, the quantity by weight represented

in those fractions being equal to 4, 3, and 93%, respectively. In some studies, protein fractions having up to 40–50% protein (but yields of only 1.5%) were generated accompanied by a 20–30% protein fraction (< 15 μm) at a yield of 11–25%. Although this fraction could be used to enhance the protein content of foods, commercial use has not occurred. The starch fraction can be further processed by water extraction to remove the protein generating purified starch for food (e.g., puddings and thickeners) and industrial applications.

To obtain a protein-rich flour, a separation cut point of 15 μm is usually sufficient, especially in soft wheat. Two separations may be required, however, to obtain a low-protein starch fraction. As an example, 10.8% protein content soft wheat flour can be fractionated into three cuts: < 15, 15–35, and 35 μm, with yields of 12, 43, and 45% having 19.5, 7.1, and 12.0% protein content, respectively. In a single operation, a 20% yield of 21–22% protein fine fraction and an 80% yield of a coarse fraction having 7.5–8% protein can be achieved from soft wheat flour of around 10.5% initial protein.

Using soft wheat flour, air-classification methods have generated a very fine subfraction that is light, fluffy, and high in protein and total dietary fiber. Applications requiring such high-gluten flour are found in the bagel industry.

The protein-rich fractions generated from wheat can be used as the basis for preparation of protein isolates, or the high-protein fraction can be added to low-protein base flours to improve the dough and baking characteristics. This enhancement is due to the nature of the protein components present and the alteration of starch and ash components during air classification. As in legumes, the soluble sugars are also concentrated into the high protein fraction.

To improve the efficiency of fine grinding and subsequent air classification, it has been suggested that a treatment of 0.1–1.0% sulfurous acid be applied for several hours as an aid in loosening the starch protein complex. Sorghum subjected to such treatment has shown increased yields of protein-rich and starch-rich fractions.

Miscellaneous Cereal Flours

Barley, and especially dehulled malted barley, lends itself to successful separation by air classification. Dehulled malted seed with 11.5% protein can be impact-milled and classified into a high-protein, low-fiber fraction with 26–27% protein content and a starch portion of about 9.5% protein. The split ratio is in the range of 16/84. Regular dehulled barley with 15% protein has up to 40% protein in the fine fraction with a 17.5% yield. The benefits of such application may be in a defibered starchy malt or barley flour or a high-starch malt flour that may be used in rapidly brewed malt drinks. Applications for generating starch have shown that up to three passes have been used in some programs, thereby yielding fractions rich in large granule starch (77–78% starch), which, upon subjecting to wet extraction, produce an almost pure large-granule starch with a higher extraction efficiency than conventional extraction processes.

Similarly, high-fiber byproduct streams from the milling of rye have been impact-milled and reclassified to yield fractions that have enhanced soluble and dietary fiber levels.

β-Glucan-enriched Flours

Nonconventional use of the air-classification technique has been applied to oats for the purpose of generating a β-glucan rich bran fraction. More recently, focus has also turned to barley flour. Instead of classifying finely ground flour, oat groats have been coarsely milled and then subjected to air classification in order to produce a high-bran coarse fraction and a fine flour fraction. The bran fraction can serve as an enriched fiber source for cereal applications or as a starting material for the isolation of β-glucan concentrates. Other researchers have defatted the oat flour prior to air classification and noted that the coarse fraction (> 30 μm) had a β-glucan content of about 17% for a 30% yield. Inactivation of enzymes would be a prerequisite for the preparation of β-glucan fractions by wet extraction. Preparation of barley fractions (2–4% yield) has shown β-glucan values of 15–18%. Numerous researchers have confirmed these results with various waxy and nonwaxy barleys.

Other Minor Uses

Cottonseed proteins are rarely used in edible foods because of pigment glands containing gossypol. Air classification has been shown to produce an edible protein fraction if there is minimal pigment gland damage during the milling of extracted flakes. Yields of 35–40% protein-rich fractions with up to 65% protein and meeting food and drug directorate standards are possible.

Finely ground soybean hulls have been air-classified into fine and coarse fractions with enriched fiber. Application of this fraction in frying batters reduces oil absorption. Substitution of wheat flour by 20% in cakes increases volume and acceptability. The protein fraction from the soybean hulls may yield increased quantities of trypsin inhibitor with anticarcinogenic properties.

Utilization of rapeseed and canola meals could be enhanced if crude fiber levels could be reduced by on-half or less. Small protein shifts of 11–17% have been achieved in roughly 50/50 splits, but in spite of the fiber being shifted into the coarse fraction, the two end products are not sufficiently improved to warrant commercialization.

Although rice protein has a high nutritional value, attempts at recovery of this protein by air classification have failed to demonstrate any significant protein shifting. Air classification of defatted finely ground rice bran has provided a high-protein bran fraction of a 50% yield and 23% protein content from an initial 19.4% protein level in the flour. Unfortunately, the ash content of the protein-enriched fraction is also increased, rendering the product unsuitable for direct food use.

Potato granules produced by spray drying from a wet milling process have been successfully classified. Granules with about 10% protein have been classified into a 15/85 fine:coarse ratio having 38 and 3.1% protein, respectively. Assuming that the economics of spray drying is favorable, the protein fraction can provide a source of nondenatured protein concentrate for food applications.

Other miscellaneous applications of air-classification principles to food products include the production of fine confectioners sugar ($< 5\,\mu$m) and fine cocoa powder ($< 75\,\mu$m), the latter being used to make top-quality, smooth chocolate products. Recent investigations have shown a possible means of generating flax lignan involving the use of hull material removed from flaxseed by pearling then subjected to classification yielding a fraction rich in lignan and water-soluble fiber. The lignan can be extracted sequentially and used in the nutraceutical industry.

The combined air-classifier mills have recently found some applications in the food industry, including milk powder, sugar, alginates, gelatine, wheat gluten and flour, cocoa powder, casein, and starches.

Air-classification techniques applied to finely ground flours have produced a wide range of protein/starch-rich fractions and with variable degrees of success. The process is advantageous in that drying and effluent disposal costs for wet plants are reduced or excluded, a significant factor from an environmental viewpoint. There is also some flexibility for the processor to vary the resulting flours to suit the end-user. Increases in capital and operating costs for the operation must be considered in light of these benefits.

See also: **Air Classification**: Principles of Operation; **Barley**; **Cocoa**: Chemistry of Processing; Production, Products, and Use; **Flour**: Roller Milling Operations; Analysis of Wheat Flours; Dietary Importance; **Rice**

Further Reading

Andersson AAM, Andersson R and Aman P (2000) Air classification of barley flours. *Cereal Chemistry* 77: 463–467.

Gueguen J (1983) Legume seed protein extraction, processing and end product characteristics. *Qualitas Plantarum, Plant Foods and Human Nutrition* 32: 267–303.

Kadan RS, Freeman DW, Ziegler GM Jr. and Spadaro JJ (1979) Air classification of defatted, glanded cottonseed flours to produce edible protein product. *Journal of Food Science* 44: 1522–1524.

King RD and Dietz HM (1987) Air classification of rapeseed meal. *Cereal Chemistry* 64: 411–413.

Kohnhorst AL, Uebersax MA and Zabik ME (1990) Production and functional characteristics of protein concentrates. *Journal of American Oil Chemists' Society* 67: 285–292.

Mok C, Park D-J and Ku K-H (1995) Air classification of barley flour. *Foods and Biotechnology* 4(1): 21–24.

Reichert RD (1982) Air classification of peas (*Pisum sativum*) varying widely in protein content. *Journal of Food Science* 47: 1263–1271.

Sosulski FW, Walker AF, Fedec P and Tyler RT (1987) Comparison of air classifiers for separation of protein and starch in pin-milled legume flours. *Lebensmittel-Wissenschaft & Technologie* 20: 221–225.

Sosulski F and Youngs CG (1979) Yield and functional properties of air-classified protein and starch fractions from eight legume flours. *Journal of American Oil Chemists' Society* 56: 292–295.

Thomas ND (1999) The MikoPul air classifier mill. *International Sugar Journal* 101: 480–481.

Tyler RT, Youngs CG and Sosulski FW (1981) Air classification of legumes. I. Separation efficiency, yield and composition of the starch and protein fractions. *Cereal Chemistry* 58: 144–148.

Vasanthan T and Bhatty RS (1995) Starch purification after pin milling and air classification of waxy, normal and high amylose barleys. *Cereal Chemistry* 72(4): 379–384.

Vose JR, Basterrechea MJ, Gorin PAJ, Finlayson AJ and Youngs CG (1976) Air classification of field peas and horsebean flours: chemical studies of starch and protein fractions. *Cereal Chemistry* 53: 928–936.

Wu YV and Stringfellow AC (1992) Air classification of flours from wheats with varying hardness: protein shifts. *Cereal Chemistry* 69(2): 188–191.

Wu YV and Stringfellow AC (1995) Enriched protein and β-glucan fractions from high-protein oats by air classification. *Cereal Chemistry* 72(1): 132–134.

Air Drying *See* **Drying**: Theory of Air-drying; Drying Using Natural Radiation; Fluidized-bed Drying; Spray Drying; Dielectric and Osmotic Drying; Physical and Structural Changes; Chemical Changes; Hygiene; Equipment Used in Drying Foods

ALCOHOL

Contents

Properties and Determination

A T Bakalinsky and M H Penner, Oregon State University, Corvallis, OR, USA

Introduction

The earliest written records indicate that alcohol (ethanol or ethyl alcohol) has been enjoyed in the human diet for thousands of years. In spite of past and present prohibitions among some groups, an impressive number and variety of alcoholic beverages have been developed, refined, and extolled in an extensive and rich literature. While the discovery of alcohol was undoubtedly accidental, its microbial origin was only established some 120 years ago through the work of Louis Pasteur. Microbial physiologists have since shown that alcohol formed through different fermentative pathways in various microbial species always serves the same purpose: regeneration of an oxidized cofactor, usually the oxidized form of nicotinamide adenine dinucleotide (NAD^+).

Major classes of alcoholic beverages include primary products of alcoholic fermentations (beer and wine), products of mixed alcoholic–lactic acid fermentations, and since the discovery of distillation – known to the Chinese at least 3000 years ago and to the ancient Egyptians even earlier – distilled and fortified beverages. Because alcohol is an excellent extractant, exhibits low toxicity, and possesses antimicrobial activity, it also enjoys wide use as a solvent for food ingredients such as spices and flavors. (*See* **Lactic Acid Bacteria**.)

Physical and Sensory Properties

Ethanol is a clear, colorless, flammable liquid miscible with water and many organic solvents in all ratios. It is hygroscopic and relatively nontoxic, exhibiting an oral LD_{50} in rats of 13.7 g kg^{-1} (grams of ethanol per kilogram of body weight required to kill 50% of the animals). When ethanol is added to water, a rise in temperature due to the heat of solvation occurs with a consequent increase in volume. When the mixture cools to the original temperature, the volume decreases, becoming slightly less than the sum of the initial water and ethanol volumes. Maximum contraction occurs at a molar ratio of eight parts of water to one part of ethanol. At a pressure of 101.3 kPa, a mixture of 95.6% ethanol and 4.4% water on a mass basis forms a constant-boiling-point mixture called an azeotrope. In practice, this means that the concentration of ethanol cannot be increased beyond 95.6%, by simple distillation. Some important physical properties of ethanol are listed in Table 1.

Ethanol has a slightly sweet taste and a characteristic aroma. The aroma threshold determined in one study ranged from 4 to 5 mg of ethanol per 100 ml of

Table 1 Physical properties of ethanol

Formula	CH_3CH_2OH
Molecular weight (Da)	46.07
Boiling point (°C)	78.32
Freezing point (°C)	−114.1
Density, d_4^{20} (g ml^{-1})	0.7893
Refractive index, n_D^{20}	1.361
Viscosity at 20 °C (cP)	1.17
Dielectric constant at 20 °C	25.7
Heat of fusion (J g^{-1})	104.6
Heat of vaporization at 78.32 °C (J g^{-1})	839.31

total aqueous solution. At high concentrations, it causes a burning sensation in the mouth. It moderates the taste of acids, as shown by the greater tartness of dealcoholized wines relative to the same untreated wines. It is said to impart 'body,' which may be due to the fact that at room temperature it is more viscous than water. Addition of sugar to an ethanol–water solution increases the threshold for ethanol, indicating that sugar masks its taste and/or aroma.

Sources of Alcohol

The natural source of alcohol is fermentation which has been defined as the oxidation of organic compounds (generally carbohydrates) in the absence of external electron acceptors. Pasteur called it 'life without air.' Microorganisms ferment in order to obtain the energy necessary for growth and reproduction and alcohol is one of several possible fermentation products. In most countries, use of alcohol in beverages is restricted to that made by fermentation. The vast amount of alcohol used industrially as a solvent and substrate in chemical syntheses is made synthetically from ethylene. The amount derived from the fermentation of agricultural products is largely dependent on economic considerations, e.g., alcohol produced by fermentation is used as a substitute for petrol in Brazil.

Alcoholic Beverages

All alcoholic beverages are derived either directly or indirectly from fermented products. As indicated in Table 2, the major groups are fermented beverages, products of mixed alcoholic–lactic acid fermentations, beverages made by distillation and by fortification. Fruits containing high sugar concentrations at maturity and nutrients at levels sufficient to support growth of fermenting yeasts, primarily *Saccharomyces cerevisiae*, have traditionally been the raw materials from which wines are made. Grapes are unusual in that they contain sufficiently high levels of sugars, nutrients, and acids to produce wines that are microbiologically stable.

Table 2 Major classes of alcoholic beverages

Fermented
Beer
Wine
Sake
Mixed alcoholic–lactic acid fermentation
Kefir
Koumiss
Distilled
Whiskies, vodka
Rum
Brandy
Tequila
Liqueurs
Fortified
Port
Sherry
Vermouth

Beer and sake are produced from barley (other grains are possible sources) and rice, respectively. Unlike wines, these beverages are derived from carbohydrates that are not initially fermentable. Conversion of carbohydrates into fermentable sugars requires the action of amylases produced by barley in the case of beer, and by the fungus *Aspergillus oryzae* during sake production. Kefir and koumiss are examples of beverages produced by the fermentation of cows' and mares' milk, respectively, by a mixture of lactic acid bacteria and lactose-fermenting yeasts. Although almost unknown in some western countries, kefir and koumiss continue to enjoy great popularity in eastern Europe. (*See* **Beers**: History and Types; **Fermented Milks**: Other Relevant Products.)

Distilled beverages are derived from fermented grains and potatoes (whiskies and vodka), sugar cane byproducts (rum), fruits (brandies), and other plants such as mezcal (tequila). Liqueurs are distilled beverages that have been flavored and sweetened. Fortified beverages require the addition of alcohol during production, generally in the form of brandy. Special fruit preserves are fortified to assure preservation. The intact fruits, usually cherries, are steeped in brandy for a period of months prior to consumption. (*See* **Brandy and Cognac**: Armagnac, Brandy, and Cognac and their Manufacture; **Rum**; **Whisky, Whiskey, and Bourbon**: Products and Manufacture.)

Use as a Preservative

Ethanol is not particularly toxic. As a sole agent of preservation in beverages, minimum concentrations ranging from 18 to 21% by volume are required to assure microbiological stability. Table wines which contain significantly less alcohol are stable because of additional factors: their natural high acidity, low pH, high content of phenolic compounds, and lack of sugar. The toxicity of ethanol towards microorganisms is due to a number of effects. At very high concentrations, as in flavor extracts, ethanol acts as a desiccant and protein denaturant. At lower concentrations, 10–20% by volume, the toxicity is believed to result primarily from interactions with cell membranes. In the yeast *S. cerevisiae*, ethanol has been shown to inhibit several solute transport systems. In the presence of sugar, the toxicity of ethanol is enhanced. This explains the stability of a class of sweet

dessert wines which contain several percent sugar and only 7–10% ethanol.

Analysis

Several methods are available for quantitative measurement of ethanol. Some are based on ethanol-specific chemical reactions, such as dichromate and enzyme-catalyzed oxidation. Others are based on selected physical properties of the product or sample which are functions of its ethanol content, such as boiling point or relative density. Often, separation of the ethanol from other compounds by chromatography, distillation, or other means is required prior to quantification. The choice of method depends on the need for accuracy and precision, analysis time, potential for interfering substances, and expense. (*See* **Chromatography**: Principles.)

With respect to interfering substances, it is worth noting that methanol is a minor but potentially toxic constituent of distilled beverages derived from fruits. It is formed from hydrolysis of methoxylated fruit pectin and does not interfere with enzymatic or chromatographic methods of ethanol analysis.

Boiling Point Determinations (Ebulliometry)

Boiling point depression is probably the most commonly used measure of ethanol in liquid systems. The method is based on Raoult's law of partial pressures. Under ideal conditions, the law states that the total pressure above a mixture of two miscible liquids is equal to the sum of their partial pressures, and that the partial pressure of each component is directly proportional to its mole fraction in the mixture. Stated mathematically:

$$P_T = P_A + P_B = P_A^0 X_A + P_B^0 X_B \qquad (1)$$

where P_T is the total pressure above a solution of liquids A and B, P_A is the partial pressure of component A, P_B is the partial pressure of component B, P_A^0 is the vapor pressure of component A above pure liquid A, X_A is the mole fraction of component A in the mixture, P_B^0 is the vapor pressure of component B above pure liquid B, and X_B is the mole fraction of component B in the mixture. Ethanol–water mixtures have boiling points ranging from slightly less than 78.32 °C (100% ethanol) to 100 °C (0% ethanol) at 101.3 kPa pressure. The ethanol–water azeotropic mixture (95.6% ethanol, w/w) boils at 78.2 °C. Deviations from Raoult's law are exhibited by mixtures of nonideal liquids, in which significant molecular interactions occur between components. Ethanol–water mixtures are examples of nonideal systems because ethanol exhibits significant hydrophilic character. Generally, the boiling point of pure water is measured in addition to that of the sample and tabulated ethanol content–boiling point data are adjusted accordingly, to account for variations in atmospheric pressure. To prevent significant evaporative loss of ethanol during boiling and measurement, ebulliometers are fitted with a small condenser, and readings are taken shortly after a constant temperature is attained.

Boiling point measurements require relatively large samples – up to 50 ml – which limits their utility. Their accuracy is reduced in the presence of high concentrations of other dissolved solutes. Since ebulliometers are simple devices, relatively inexpensive, and sufficiently accurate for production purposes, small wineries have used them extensively.

Relative Density Measurements (Hydrometry)

Relative density (specific gravity) is the ratio of the density of a solution to the density of a reference solution (usually water at 4 °C). Relative densities of aqueous solutions are dependent on the concentration of dissolved solutes. Two general approaches are used to obtain relative densities.

The first is to measure the mass of known volumes of sample and standard reference solutions. This is accomplished by use of a pycnometer, a small, relatively lightweight flask of known volume. Since specific volume is temperature-dependent, the pycnometer is immersed in a water bath maintained at a standard temperature. The mass of the sample and reference solutions is then determined on an analytical balance and compared to tabulated standard values. The relative density of aqueous ethanol solutions reflects the concentration of ethanol, assuming it is the only (or major) dissolved solute.

The second general approach is based on Archimedes' principle, which holds that an object immersed in a liquid appears to lose an amount of mass equivalent to the mass of liquid it displaces. The Westphal balance illustrates this principle. Its design is similar to that of an equal-arm, two-pan balance. At one end, a tube of known volume and weight is immersed in a reference solution and the beam is balanced by addition of weights to compensate for the buoyancy of the liquid. The procedure is repeated by immersion in the unknown liquid sample. In this manner, a relative density is determined. A more commonly used device, based on the same principle, is the hydrometer. Hydrometers are glass instruments with narrow, calibrated stems much like a thermometer at the top, but with much expanded bottom halves that are weighted with lead to insure that they float upright. When the hydrometer is placed in a liquid sample, it sinks to a depth that results in displacement of a volume of liquid equal to its weight. The alcohol

content is read from the calibrated stem at the point where it intersects the liquid surface. Hydrometers designed for alcohol measurement are standardized for use at a specific temperature and calibrated either in percent alcohol (v/v), or °proof, where °proof = 2 × % ethanol (v/v). Correction tables are available for measurements taken at nonstandard temperatures. Hydrometers are accurate over a limited range of alcohol concentrations. Consequently, a set of instruments may be required to cover the range of interest.

Since the relative density of a solution depends on the concentration of all dissolved solutes, it is often necessary to separate the ethanol from the other solutes. To accomplish this, the ethanol contained in a specified volume of product is separated by distillation. The distillate is diluted to a known volume with water, and the relative density of the resulting solution is measured. In fermented beverages, the concentration of other volatiles that codistill is significant from a sensory perspective, but is very low compared to that of ethanol and does not significantly affect the relative density. In some beverages, such as bourbon whiskey, the concentration of soluble solids other than ethanol is sufficiently low to permit direct measurement of relative density without need for distillation.

The international Association of Official Analytical Chemists (AOAC) has approved methods based on specific gravity for determination of ethanol in distilled beverages, beers, and wines. The methods require relatively large samples, and are somewhat tedious due to the need for distillation and strict temperature control. However, the equipment is relatively inexpensive and readily available, and precise results can be expected from a skilled technician.

Enzymatic Determination

The enzymatic analysis of ethanol is generally based on the reaction catalyzed by alcohol dehydrogenase. As shown in eqn (2), the oxidation of ethanol by NAD^+ yields acetaldehyde and NADH.

$$\underset{\text{Ethanol}}{CH_3CH_2OH} + NAD^+ \rightarrow \underset{\text{Acetaldehyde}}{CH_3CHO} + NADH + H^+ \qquad (2)$$

The reaction is monitored photometrically at 340 nm since NAD has no significant absorbance at this wavelength in the oxidized form (NAD^+), but has a molar extinction coefficient of $6.23 \times 10^3\,l\,mol^{-1}\,cm^{-1}$ when reduced (NADH). Primary alcohols other than ethanol can interfere, but this is rarely a problem in foods. Determination of ethanol concentration requires that conditions be adjusted such that the equilibrium is shifted toward acetaldehyde formation. One approach is to trap the aldehyde as it forms by reaction with semicarbazide. Another is to carry out the reaction at alkaline pH. A third is to couple the reaction with a second enzyme, aldehyde dehydrogenase, which converts the acetaldehyde to acetic acid. In the latter case, 2 mol of NADH are formed per mole of ethanol. (*See* **Enzymes**: Uses in Analysis.)

Enzyme assays have high sensitivity relative to other methods and good accuracy due to the specificity of the reaction. Often, samples require treatment such as dilution, pH adjustment, or decolorization prior to analysis. Material needs for this type of assay are simple: volumetric glassware, a photometer, and reasonable temperature control. A recurring expense is the enzyme itself, which must be stored and handled appropriately.

Dichromate Oxidation

Chemical oxidation may be used to quantify the amount of ethanol in a sample. In the dichromate procedure, the quantity of dichromate required to oxidize ethanol to acetic acid is measured. In complex beverages, the ethanol is usually separated from potentially interfering compounds by distillation. The ethanol in the ethanol–water distillate is then oxidized with a known excess of dichromate in the presence of sulfuric acid. Residual dichromate is then reduced by ferrous ammonium sulfate in a standard oxidation–reduction titration. A useful end-point indicator is 1,10-*o*-phenanthroline-ferrous sulfate, which turns brownish purple from blue-green. Potassium dichromate is available in sufficient purity to be used as a primary standard. Ferrous ammonium sulfate may be used as an approximate standard, but solutions that are not fresh must be standardized against dichromate.

The dichromate oxidation method for the measurement of alcohol in wines has been approved by the AOAC. The method is appropriate for routine analyses only when the reagents are standardized regularly.

Refractive Index

The refractive index (RI) of a medium is dependent on its chemical composition, since under controlled conditions composition dictates its electrical and magnetic properties. The RI of a sample is defined as the ratio of the speed of light in a vacuum to its speed in the sample medium. Consequently, RI values are always greater than one. The RI of aqueous ethanol solutions can be measured directly to indicate ethanol concentration. When interfering compounds are present, the RI of a distillate can be measured. Since the RI of a liquid is temperature-sensitive, measurements

taken at nonstandard temperatures must be corrected by use of conversion tables.

The AOAC has approved RI-based methods for determining ethanol in beers and wines. For wines, the measurement is made on the distillate. Measured RI values for beer are used in conjunction with specific gravity measurements to calculate ethanol content.

Chromatography

Chromatography is based on the separation of sample components due to their differing affinities within a stationary–mobile-phase system. A successful chromatographic run results in separation of the analyte of interest from all sample components that would otherwise interfere with its analysis. The separated analyte may then be quantified by one of several detectors. The ethanol in most samples can be readily separated by chromatography. Since ethanol is volatile, direct measurement by gas chromatography (GC) is possible. Several commercial vendors have columns specifically designed for determining the ethanol content of beverages. Simple, 2-m packed columns are commonly used. Depending on the sample, it is often beneficial to carry out a simple pretreatment to extend the life of the column by removing nonvolatile components. Ethanol exiting a GC column is commonly quantified with a flame ionization detector or by thermal conductivity. Ethanol separations using high-pressure liquid chromatography (HPLC) are also possible. An advantage of HPLC over GC is the possibility of simultaneous determination of certain nonvolatile sample components. Commercial HPLC columns for ethanol analysis are available. Chromatographic methods, which are expensive, are widely used in analytical laboratories because of their specificity, increasingly simple operation, and potential for automation. (*See* **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography.)

See also: **Beers**: History and Types; **Brandy and Cognac**: Armagnac, Brandy, and Cognac and their Manufacture; **Chromatography**: Principles; High-performance Liquid Chromatography; Gas Chromatography; **Cirrhosis and Disorders of High Alcohol Consumption**; **Enzymes**: Uses in Analysis; **Fermented Milks**: Other Relevant Products; **Lactic Acid Bacteria**; **Rum**; **Whisky, Whiskey, and Bourbon**: Products and Manufacture

Further Reading

Amerine MA and Ough CS (1980) *Methods for Analysis of Musts and Wines*. New York: Wiley.

Amerine MA and Roessler EB (1983) *Wines. Their Sensory Evaluation*. New York: W.H. Freeman.

Association of Official Analytical Chemists (1990) *Official Methods of Analysis*, 15th edn. Arlington, Virginia: Association of Official Analytical Chemists.

Boulton RB, Singleton VL, Bisson LF and Kunkee RE (1996) *Principles and Practice of Winemaking*. New York: Chapman Hall.

Heath HB and Reineccius G (1986) *Flavor Chemistry and Technology*. New York: Van Nostrand Reinhold.

Pederson CS (1979) *Microbiology of Food Fermentations*, 2nd edn. Westport, CT: AVI Publishing.

Sherman PD and Kavasmaneck PR (1980) Ethanol. *Encyclopedia of Chemical Technology*, 3rd edn, vol. 9. New York.

Sommer AE and Bücker R (1983) Ethanol. *Encyclopedia of Chemical Processing and Design*, vol. 19. New York: Marcel Dekker.

Zoecklein BW, Fugelsang KC, Gump BH and Nury FS (1990) *Production Wine Analysis*. New York: Van Nostrand Reinhold.

Metabolism, Beneficial Effects, and Toxicology

C H Halsted, University of California, Davis, CA, USA

Background

Alcohol is both a food that provides 29.7 kJ g^{-1} (7.1 kcal g^{-1}), accounting for 5.6% of energy in the American diet, and a drug with potential for intoxication, addiction, and damage to several organ systems. In the USA, alcohol is consumed to some extent by at least two-thirds of the population, and an estimated 5–10% consume alcohol in excess and are at risk for addiction, liver disease, cardiomyopathy, pancreatic disease, and neurological defects. The overall mortality from alcoholism in the USA is about 100 000 deaths per year, about half due to accidents related to intoxication, and the remainder mainly due to chronic alcoholic liver disease and its consequences.

Metabolism of Alcohol

Unlike most other drugs, alcohol is completely metabolized after ingestion at a rate of 10–15 g h^{-1}, and its energy content is not stored in the body. Alcohol diffuses across all cell membrances, and after equilibration, its concentration is similar in all aqueous compartments. After the human ingestion of 50 g of alcohol, its concentration in the stomach reaches

Table 1 Alcohol metabolizing enzymes

Enzyme	K_m *(mM)*	*Related effects*
Gastric cytosolic alcohol dehydrogenase	500	'First pass'; 30% in men, 10% in women
Liver cytosolic alcohol dehydrogenase (one to two drinks)	2	Increased NADH/NAD ratio promotes fatty acid synthesis leading to fatty liver, preferential oxidation of ethanol over lipid, and reduced gluconeogenesis, lactic acidosis, and hyperuricemia
Liver microsomal CYP2E1 (more than two drinks)	20	Increased NADPH/NADP ratio consumes energy and increases metabolism of many other drugs and vitamin A; increased acetaldehyde inhibits mitochondrial functions and stimulates lipid peroxidation and collagen synthesis

$\sim 1\,\text{mol}\,\text{l}^{-1}$ ($4.2\,\text{g}\,\text{dl}^{-1}$), decreasing by one-half in the upper intestine and equilibrating with the blood alcohol level in the distal intestine. 'First pass' metabolism is achieved in the stomach by gastric alcohol dehydrogenase, an enzyme located in parietal cells with K_m 500 mM that accounts for $\sim 30\%$ of total alcohol metabolism in men and $\sim 10\%$ of alcohol metabolism in women. This gender difference in gastric metabolism partly explains the lower tolerance of women for alcohol. Subsequently, alcohol is metabolized entirely in liver cells by two enzyme systems: cytoplasmic alcohol dehydrogenase and a specific and inducible microsomal cytochrome P450 enzyme, known as CYP2E1 (Table 1). Hepatic alcohol dehydrogenase with K_m 2 mM metabolizes alcohol at blood levels consistent with moderate drinking. During the oxidation of alcohol by this reaction, NAD is reduced to NADH while producing acetaldehyde and subsequently net energy through the production of ATP. However, the increased ratio of NADH to NAD is associated with increased fatty acid and triglyceride synthesis, reduced gluconeogenesis, enhanced lactate production, and reduced urinary excretion of uric acid. The potential clinical consequences of these metabolic side effects include transient fatty liver, hypertriglyceridemia, hypoglycemia, acidosis, and gout. On the other hand, CYP2E1 with a K_m of 20 mM accounts for the metabolism of higher alcohol levels as seen with chronic excessive alcohol use. Alcohol oxidation by CYP2E1 is energy-wasteful by utilizing NADPH at a 2:1 molar ratio while producing greater amounts of acetaldehyde without generation of ATP. Increased intrahepatic acetaldehyde blocks mitochondrial respiration thereby enhancing the accumulation of fat in the liver, increases catecholamine release while promoting ketoacidosis, and, by enhancing lipid peroxidation and collagen synthesis, contributes to the development of alcoholic hepatitis and cirrhosis. The induction of CYP2E1 by frequent alcohol consumption generates a free radical that can trigger oxidative liver injury, accounts for alcohol tolerance in habitual users, and is associated with induction of other microsomal enzymes that accelerate the metabolism of drugs such as acetaminophin, barbiturates, and coumadin, and the catabolism of vitamin A.

Risks Versus Benefits of Alcohol Consumption

The risks or benefits of alcohol consumption are based on quantitative estimates of the numbers of drinks per day, where one drink of 360 ml of beer, 120 ml of wine, or 40 ml of spirits provides ~ 15 g of alcohol. By these standards, 'moderate drinking' is defined as one to two drinks (15–30 g of alcohol) per day or seven to 14 drinks per week on average, and excessive drinking that increases disease risk is defined as more than two to three drinks (30–45 g of alcohol) per day for men and more than one or two drinks (15–30 g of alcohol) per day for women. Excessive drinking may occur through habituation or addiction, and in cultures where wine is consumed daily as a component of the diet (Table 2).

A variety of studies from many developed countries established a J-shaped curve for mortality risk from alcohol consumption, wherein the lowest mortality occurs at consumption of one or two drinks per day compared with abstinence or to the consumption of more than two drinks per day. Thus, the term 'moderate alcohol consumption' is used to define consumption with clinical benefit and no increased mortality risk. Comparing all causes of mortality, the moderate drinking level is somewhat less for women at one to two drinks per day than in men at two to three drinks per day. When all the causes of mortality are taken into account, it appears that the benefit of moderate drinking is confined to the prevention of coronary heart disease and ischemic strokes, both together of which account for the majority of deaths in the USA. Above the level of moderation, i.e., at more than two to three drinks per day, alcohol consumption associates with life-threatening trauma, hemorrhagic strokes,

Table 2 Benefits and risks of alcohol consumption

Benefits (one to two drinks per day)	Protection against ischemic cardiovascular disease
	Improved lipoprotein profile
	Increased tissue plasminogen activator
	Wine flavonoids decrease LDL oxidation, platelet adhesiveness
Risks (more than three drinks per day)	Hypertension
	Cardiomyopathy
	Hemorrhagic stroke
	Pancreatitis
	Anemia
	Neurologic disorders
	Cancer of esophagus, breast, colon
	Alcoholic liver disease
	Malnutrition

hypertension, esophageal, breast and colorectal cancers, acute and chronic pancreatitis, malnutrition, and alcoholic liver disease.

Interest in the cardiovascular protective effects of alcohol stemmed from the French report of the paradoxical improved mortality among wine-drinkers in the Mediterranean provinces compared with residents of northeastern provinces, in spite of the similar content of high-fat foods in the diet and overall alcohol consumption. A subsequent *in vitro* study at the University of California Davis demonstrated that the oxidation of low-density lipoprotein could be prevented by coincubation with phenolic compounds extracted from red wine, suggesting that the improved cardiovascular risk of French red wine-drinkers was related to these nonalcoholic compounds. The principal flavonoid of red wine is catechin, which is also present in nonfermented grape juice. The biological availability of catechin for human subjects is similar in red wine to that in dealcoholized red wine, but a greater amount of catechin may be extracted from fermented grapes during wine-making than from nonfermented grape juice. A long-term prospective Danish study of residents of Copenhagen concluded that the lowest mortality rates occur in wine-drinkers, with the highest mortality in consumers of liquor. A study from Northern California concluded that the improved mortality in wine-drinkers related to their healthy dietary lifestyle, and that red and white wine provided equal protection.

There are several explanations for the cardioprotective effects of moderate alcohol consumption, in particular the effects of red wine. Moderate consumption of all types of alcoholic beverages has a positive effect on the circulating lipoprotein profile, principally by elevating the levels of high-density lipoprotein. Furthermore, moderate alcohol consumption by women results in modest reductions in low-density lipoprotein levels and increases Apo A-1 but decreases Apo B lipoproteins. Whereas serum triglycerides increase in male alcohol consumers, they may decrease in postmenopausal women drinkers. Moderate alcohol consumption also benefits the clotting process by increasing levels of tissue plasminogen activator. In addition to the benefit of wine flavonoids in reducing the oxidation of low-density lipoprotein, they also improve the clotting risk by reducing platelet adhesiveness.

Alcohol Toxicity

With the exception of traffic fatalities that may occur after consumption of an intoxicating dose of alcohol pending timing after the last drink, the deleterious effects of alcohol consumption are limited to those who consume excessive amounts of alcohol (on average, more than two to three drinks per day) on a chronic basis. The clinical risks of alcohol toxicity increase in proportion to increasing average daily amount of alcohol and duration of excessive alcohol consumption. In many instances, the development of these clinical syndromes is modulated by nutrient deficiency, and in the case of pancreatic insufficiency and alcoholic liver disease, the clinical syndromes can contribute to malnutrition.

Clinical Effects of Alcohol Toxicity

Excessive alcohol consumption contributes to increased risks of alcoholic liver disease, cardiomyopathy, chronic pancreatitis and insufficiency, nutritional anemia, certain cancers, abnormal energy metabolism, and malnutrition that is more pronounced after the development of alcoholic liver disease. Abnormal nutrient availability and metabolism play a role in all of these clinical complications of chronic alcoholism (Table 3).

Increased cancer risks Increased risk of cancer of the oropharynx and esophagus is directly related to the amount and duration of alcoholism, is tripled in heavy smokers, and may be related in part to abnormal carcinogen metabolism secondary to zinc and vitamin A deficiency. The incidence of breast cancer is increased in women who drink, in part due to enhanced estrogen secretion, and particularly in women who consume inadequate folate in their diets. Colorectal cancer risk is increased in excessive drinkers of both sexes and is related to both alcohol toxicity and low folate status.

Table 3 Clinical toxic effects of alcohol

Cancers of oropharynx, esophagus, breast and colon (see folate)
Cardiac risk
Arrhythmias: 'holiday heart syndrome'
Cardiomyopathy: low-output failure
Wet beri-beri: high-output failure (see thiamin)
Anemia due to iron, folate, and/or pyridoxine deficiencies
Pancreatitis and pancreatic insufficiency with malabsorption of fat and fat-soluble vitamins
Alcoholic liver disease
Fatty liver: reversible
Alcoholic hepatitis: inflammation, steatonecrosis
Alcoholic cirrhosis: portal hypertension with risk of ascites, renal failure, ruptured esophageal varices, and hepatic encephalopathy
Neurological disorders:
Wernicke–Korsokoff syndrome (see thiamin)
Peripheral neuropathy (see thiamin, pyridoxine)
Intoxication: trauma, automobile accidents
Alcohol withdrawal syndromes

Cardiac risks Alcoholics are at increased risk for three different clinical cardiac complications. The 'holiday heart syndrome' refers to potentially lethal arrhythmias secondary to the excessive consumption of alcohol over a short period of time. Alcoholic cardiomyopathy is caused by the toxic effects of alcohol and/or acetaldehyde or increased circulating fatty acids on mitochondrial respiration in cardiac muscle cells, and results in low-output left-sided heart failure. 'Wet beriberi' represents high-output heart failure secondary to thiamin deficiency in chronic alcoholism and, since thiamin is integral to many carbohydrate and ketoacid reactions, may be triggered or exacerbated by acute fall in thiamin stores following the administration of intravenous glucose in severely malnourished alcoholics.

Pancreatitis and pancreatic insufficiency Chronic alcoholism is the leading cause of acute and recurrent attacks of pancreatitis in developed countries, resulting in anatomic distortion and progressive destruction of the pancreas. Loss of more than 90% of pancreatic function results in pancreatic insufficiency, a condition characterized by glucose intolerance due to destruction of pancreatic beta cells and to decreased production of pancreatic digestive enzymes, resulting in malabsorption of dietary protein, lipid, and lipid-soluble vitamins A, D, E, and K. The net clinical picture is that of late-onset chronic steatorrhea (diarrhea due to excessive unabsorbed fat in the stool), with malnutrition, diabetes, and signs of fat-soluble vitamin deficiencies.

Anemia Anemia is common in chronic alcoholism due to frequent episodes of bleeding from alcohol-associated gastritis or bleeding esophageal varices and to nutritional causes. A study of more than one hundred anemic chronic alcoholics admitted to a large urban hospital found mixed and combined causes, including iron deficiency consistent with episodic blood loss in about one-quarter, megaloblastic bone marrow consistent with folate deficiency in one-third, and sideroblastic bone marrow changes consistent with pyridoxine deficiency in one-quarter of the patients.

Neurological effects of excessive alcohol The neurological effects of alcohol consumption can be categorized as those related directly to alcohol, those that are related to alcoholic liver disease, and those that are secondary to the effects of alcohol consumption on specific micronutrients. The direct toxic effects of excessive alcohol consumption on the brain include intoxication, coma, head injury, and withdrawal syndromes. Intoxication is associated with lowered inhibition, euphoria, poor memory and judgment, and decreased reaction time, which is the principal cause of motor vehicle accidents and other alcohol-related trauma. About half of all motor vehicle accidents involve innocent or intoxicated pedestrians. Legal intoxication in most localities in the USA occurs at blood levels above $0.08\,\mathrm{g\,dl^{-1}}$ or $0.019\,\mathrm{mol\,l^{-1}}$. Severe intoxication resulting in coma and death can occur at levels above $0.4\,\mathrm{g\,dl^{-1}}$ or $0.095\,\mathrm{mol\,l^{-1}}$. Intoxication is often the background for falls that can result in head injury, such as subdural hematoma, that, if unrecognized, can lead to progressive loss of consciousness and death. Alcohol withdrawal syndromes occur after prolonged consumption of excessive amounts of alcohol, usually in binge drinkers who are forced through illness or other circumstances to abruptly stop drinking. The several potential signs of alcohol withdrawal depend upon the length of time after the last drink and are all characterized by hyperexcitability. For example, tremulousness occurs within hours, a general seizure may occur within the first 24 h, and delirium tremens with auditory or visual hallucinations may occur 2–5 days after the last drink. These withdrawal syndromes can be forestalled by reinstitution of alcohol and gradual lowering of intake or by specific anti-anxiety benzodiazepine drugs.

Hepatic encephalopathy represents a disturbance in consciousness that occurs in patients with end-stage alcoholic liver disease. This syndrome is caused by nitrogenous substances, in particular ammonia, that originate through intestinal bacterial digestion of protein, bypass their hepatic site of metabolism due to liver scarring and diversion of blood flow from the portal-splanchnic circulation to the systemic circulation, and then cross the blood–brain barrier

into the brain. Elevations in blood ammonia levels typically occur in patients with severe cirrhosis of the liver after ingestion of excessive animal protein or after acute intestinal bleeding. If untreated, the stages of consciousness proceed from normal to confusion to somnolence to coma to death. Treatment consists of avoidance of excessive dietary animal protein, prevention or prompt treatment of intestinal bleeding, and administration of substances that reduce ammonia levels in the intestine.

Neurologic disorders that are based on micronutrient deficiencies include the Wernicke–Korsokoff syndrome of thiamin deficiency and peripheral neuropathies of thiamin or pyridoxine deficiency, and are described separately below.

Alcoholic liver disease The risk of developing alcoholic liver disease is dependent upon the amount and duration of alcohol consumed. While alcoholic fatty liver is common and reversible with abstinence, alcoholic hepatitis or inflammation of the liver with steatonecrosis of hepatocytes carries a significant mortality and is the precursor of chronic scarring or cirrhosis of the liver. Both alcoholic hepatitis and cirrhosis can result in portal hypertension, which is associated with diversion of portal-splanchnic blood flow through collateral veins to the systemic circulation, and is characterized clinically by increased risk of abdominal fluid accumulation known as ascites, renal failure, acute hemorrhage from esophageal varices, and hepatic encephalopathy.

Retrospective European and American population studies found that the chronic consumption of at least six drinks per day was the threshold for developing alcoholic cirrhosis over time, whereas the statistical risk of developing cirrhosis develops with as little as three drinks per day. A German study of well-to-do corporate executives found a direct relationship between the amount and duration of alcohol consumption and the incidence of cirrhosis by liver biopsy, such that the average daily consumption of 160 g of alcohol (11 drinks, equivalent to 440 ml or a pint of whiskey) over 15 years resulted in a 50% likelihood of cirrhosis. Worldwide, the mortality from alcoholic liver disease increases in direct proportion to the national per capita consumption of alcohol, with the greatest number of cirrhosis-related deaths occurring in Russia and eastern European countries, followed by countries with a high daily cultural wine consumption such as France and Italy. These statistics, together with results from alcohol-fed animal models, support the notion that the pathogenesis of alcoholic liver disease is related mainly to the toxic effects of alcohol and its acetaldehyde metabolite that trigger oxidative and cytokine responses. These responses result in hepatic inflammation, hepatocellular necrosis, collagen production, and cirrhosis. More recent data point to an important role of abnormal hepatic methionine metabolism during excessive alcohol consumption – in particular, decreased hepatic *S*-adenosylmethionine – that is associated with abnormal antioxidative mechanisms and altered hepatocellular DNA metabolism associated with necrosis and apoptosis of hepatocytes.

Nutrient Deficiencies

While alcoholic liver disease is mainly due to the toxicity of alcohol, its presence has profound effects on nutrient metabolism (Table 4). There are essentially three overall etiologies of malnutrition in chronic alcoholism: decreased dietary intake due to anorexia and/or economic deprivation combined with alcohol addiction, decreased intestinal absorption, and abnormal nutrient metabolism with increased nutrient excretion in the bile or urine. With the exception of beer as a source of folate, alcoholic beverages are essentially devoid of micronutrients and of protein, carbohydrate, and lipid sources of energy. Abnormal nutrient absorption of dietary energy from fat and of soluble vitamins is magnified by pancreatic insufficiency due to decreased pancreatic lipase secretion and by alcoholic liver disease due to decreased secretion of bile acids that are required for the micellar solubilization of dietary lipids.

Thiamin Chronic alcoholism is the principal cause of thiamin deficiency in the developed world, and low serum levels of thiamin have been described in 80% of patients with alcoholic liver disease. The etiologies of thiamin deficiency include poor diet and intestinal

Table 4 Micronutrient deficiences associated with chronic alcoholism

Micronutrient	*Deficiency signs*
Thiamin	Wernicke–Korsokoff syndrome
	Peripheral neuropathy
Folate	Macrocytic anemia
	Diarrhea
	Colon and breast cancer
	Hyperhomocysteinemia
Pyridoxine	Peripheral neuropathy
	Sideroblastic anemia
Vitamin A	Night blindness
	Increased cancer risk
Vitamin D	Osteoporosis
Zinc	Night blindness
	Abnormal taste
	Hypogonadism
	Abnormal cellular immunity
Iron	Microcytic anemia

malabsorption secondary to alcohol inhibition of the Na, K-ATPase pump in the basolateral membrane of the intestinal epithelial absorbing enterocytes. The clinical features of the Wernicke–Korsokoff syndrome of thiamin deficiency include progressive cerebral, thalamic, and cerebellar defects that result in cognitive defects, ophthalmoplegia, and a wide-based gait. The Korsokoff syndrome refers to irreversible confabulation, confusion, and profound loss of recent memory. Chronic alcoholics are also prone to develop wet or dry beriberi. Wet beriberi is one of the cardiomyopathies of chronic alcoholism, is characterized by high-output cardiac failure, and can be precipitated by intravenous administration of glucose to severely malnourished patients, with marginal thiamin stores. Dry beriberi represents progressive peripheral neuropathy mainly of the lower extremities and may be associated with painful hypersensitivity of the plantar surfaces of the feet that makes walking difficult.

Folate deficiency and abnormal methionine metabolism Folate deficiency occurs in about 50% of patients with alcoholic liver disease and results from combinations of poor diet, malabsorption secondary to alcohol inhibition of the intestinal reduced folate carrier, abnormal hepatobiliary metabolism, increased urine excretion, and acetaldehyde-triggered oxidation of the folate molecule. While macrocytic anemia is the overt clinical sign of folate deficiency induced by chronic excessive alcohol ingestion, more subtle signs include chronic diarrhea due to abnormal enterocyte function and hyperhomocysteinemia when more than two alcoholic drinks are consumed per day. The combined effects of folate deficiency and chronic alcohol intake may contribute to abnormal hepatic methionine and DNA metabolism. Studies in animal models found that chronic exposure to alcohol impairs the activities of methionine synthase and methyl adenosyl transferase, resulting in hyperhomocysteinemia and *S*-adenosylmethionine deficiency, which in turn impairs the synthesis of the antioxidant glutathione and contributes to the development of alcoholic liver disease. As indicated, impaired DNA metabolism may account for the increased risk for colon and breast cancer found in excessive alcohol consumers with dietary inadequacy of folate.

Vitamin B_{12} The standard Schilling test for vitamin B_{12} absorption is abnormal in chronic alcoholic patients secondary to low uptake of the vitamin by the absorbing ileal enterocytes. However, whereas hepatic levels of vitamin B_{12} are low in patients with alcoholic liver disease, serum vitamin B_{12} levels are typically normal or elevated in chronic alcoholism and correlate with changes in serum enzymes that reflect alcoholic liver injury. This dichotomy may relate to failure of liver uptake and/or increased release of vitamin B_{12} and its analogs from damaged hepatocytes. Owing to the metabolic interaction of vitamin B_{12} and folate in the methionine synthase reaction, the clinical effects of liver vitamin B_{12} deficiency cannot be readily distinguished from those of folate deficiency. Thus, the degree to which liver vitamin B_{12} deficiency contributes to macrocytic anemia and hyperhomocysteinemia in chronic alcoholic patients with liver disease has not been established.

Pyridoxine Pyridoxine, or vitamin B_6, deficiency in chronic alcoholism results from poor diet and the effects of the alcohol metabolite acetaldehyde on hepatic pyridoxal phosphate that releases it from its protein binder and enhances its urinary excretion. The effects of pyridoxine deficiency include sideroblastic anemia, peripheral neuropathy, and a relative decrease in serum alanine transaminase activity that results in decreased ratio of this hepatic enzyme compared with aspartate transaminase during the development of alcoholic liver disease.

Vitamin A While serum levels are unreliable, hepatic vitamin A deficiency, according to measurements in liver biopsies, increases in relationship to the severity of alcoholic liver disease and is essentially universal in alcoholic cirrhosis. The causes of vitamin A deficiency include malabsorption secondary to decreased secretion of pancreatic lipase and of bile acids and increased catabolism due to alcohol induction of microsomal enzymes that convert retinol to polar metabolites that are rapidly excreted in bile, as well as enhanced metabolism of retinol to retinaldehyde by alcohol dehydrogenase. The effects of vitamin A deficiency in chronic alcoholics include decreased dark visual adaptation, which may enhance the risk of automobile accidents, and abnormal epithelial cell turnover, which may enhance the risk of esophageal cancer. Chronic alcoholics have an increased risk of vitamin A toxicity during supplementation of vitamin A at relatively low levels (e.g., 5000 IU per day), resulting in more rapid progression of alcoholic liver injury. Supplementation of chronic alcoholics with β-carotene contributes to a similar deleterious effect on the liver as well as an enhanced risk of lung cancer when smoking is combined with excessive alcohol consumption.

Vitamin D and metabolic bone disease Chronic alcoholics frequently have low circulating levels of 25-hydroxy vitamin D, especially following the development of alcoholic liver disease. The potential

causes of vitamin D deficiency in alcoholic liver disease include poor diet, decreased sun exposure, decreased intestinal absorption secondary to inadequate bile secretion, decreased hepatic 25-hydroxylation, and increased degradation by alcohol induced microsomal enzymes. Vitamin D deficiency is compounded by decreased calcium absorption as a pathway to decreased bone accretion. However, moderate alcohol consumption may be protective against bone disease, since alcohol use also raises estrogen levels in women and is associated with decreased secretion of parathyroid hormone.

Zinc Low serum zinc levels are common in chronic alcoholics and are low in about 50% of liver specimens obtained at all stages of alcoholic liver disease. Zinc deficiency is caused by inadequate dietary ingestion of zinc-rich proteins, decreased production of zinc-rich pancreatic enzymes, abnormal transport through absorbing enterocytes, and enhanced urine excretion due to decreased binding to serum albumin. Among alcoholics, zinc deficiency contributes to night blindness since zinc is a cofactor for retinol dehydrogenase, impaired alcohol metabolism due to decreased activity of gastric alcohol dehydrogenase, altered taste that may contribute to anorexia, decreased testosterone production and hypogonadism that may contribute to osteoporosis, and altered cellular immune function contributing to enhanced infection risk.

Iron Iron deficiency in chronic alcoholic patients is due to frequent episodes of gastrointestinal bleeding, and was found in about one-quarter of consecutive anemic patients admitted to a large urban hospital. The causes of bleeding among alcohol consumers include acute gastritis, esophageal tears due to retching and vomiting (Mallory–Weiss syndrome), and episodic ruptured esophageal varices. Contrary to popular opinion, the incidence of peptic ulcer is no greater in alcoholics than in nonconsumers of alcohol.

Energy Metabolism in Chronic Alcoholism

Alcohol consumers may be severely malnourished, at normal weight, or obese. In moderate drinkers, alcohol-induced thermogenesis accounts for 15% of alcohol metabolism, and preprandial alcohol consumption increases the amount of nonalcohol calories consumed in a meal. However, other studies have shown that women moderate drinkers may become underweight due to the substitution of alcohol calories for carbohydrate. Among excessive alcohol drinkers, the substitution of alcohol for nonalcohol calories decreases body weight due to the energy consuming effect of alcohol metabolism by inducible hepatic microsomal CYP2E1. Conversely, weight gain and obesity may occur when alcohol is added to a typical high-fat diet, potentially due to promotion of fat storage that results from the inhibitory effect of alcohol on oxidation of lipids in the liver (Table 5).

Effects of Alcoholic Liver Disease on Energy and Protein Metabolism

Extensive multicenter clinical studies have established the universality of protein calorie malnutrition in patients with alcoholic hepatitis and cirrhosis (Table 6). Several studies that used anthropometric techniques found that body fat stores were decreased in derelict chronic alcoholics, whereas those who developed alcoholic liver disease had a significant depletion of skeletal muscle as well. Protein calorie malnutrition scores based on anthropometry, urinary creatinine excretion, serum albumin, transferrin, and retinol bound protein, and anergy by skin testing correlated with disease severity and mortality in over 500 male patients with alcoholic liver disease. Retrospective studies showed that malnourished excessive alcohol consumers limit their intakes of nonalcohol calories, whereas prospective studies of hospitalized patients with alcoholic liver disease found significant anorexia with voluntary consumption of less than half the diet offering, which correlated with disease severity and was predictive of mortality risk. On the basis of measured lean body mass or body cell mass, resting energy expenditure by indirect calorimetry is either normal, low, or high in patients with alcoholic cirrhosis. Stable and abstinent cirrhotics tend to have a normal or low energy expenditure if muscle and visceral protein stores are normal or decreased. However, patients with active

Table 5 Energy metabolism during alcohol consumption

	Moderate drinkers	*Excessive drinkers*
Increase in body weight	Appetite stimulation or disinhibition	Increase in fat stores through inhibition of lipid oxidation
Decrease in body weight	Substitution of alcohol for carbohydrate in diet (women)	Substitution of alcohol for nonalcohol calories Energy wasting of CYP2E1 metabolism

Table 6 Effect of alcoholic liver disease on energy and protein metabolism

Protein calorie malnutrition	Universal in advanced liver disease
	Decreased dietary intake due to anorexia
	Increased protein catabolism in active disease
Abnormal energy metabolism	Active disease: hypermetabolic
	Inactive disease or stable cirrhosis
	Normal or hypometabolic
	Prefer fatty acids as fuel source
	Decreased glycogen storage
	Fat malabsorption due to low pancreatic enzyme and bile secretion
Nutritional support	Improves nutritional status
	Not proven to alter the course of disease

alcoholic hepatitis tend to have increased energy expenditure, insulin resistance, and increased urinary excretion of total nitrogen and creatinine, consistent with ongoing inflammatory processes in the liver. Malnourished nonhospitalized patients with stable cirrhosis have normal carbohydrate metabolism but decreased glycogen stores with a greater tendency to utilize fatty acids as a fuel source. Many clinical trials have approached the question of whether improving protein calorie malnutrition by oral, enteral, or parenteral nutrition support will reduce the morbidity and mortality from alcoholic liver disease. Although most trials have found an improvement in parameters of malnutrition, none have shown that correction of malnutrition by nutritional support alters the course of this disease. However, a recent European multicenter clinical trial showed that provision of oral *S*-adenosylmethionine for 2 years decreased the mortality from active alcoholic liver disease by one half.

See also: **Alcohol**: Properties and Determination; Alcohol Consumption; **Anemia (Anaemia)**: Iron-deficiency Anemia; Other Nutritional Causes; **Cancer**: Epidemiology; **Coronary Heart Disease**: Etiology and Risk Factor; **Folic Acid**: Properties and Determination; **Iron**: Properties and Determination; Physiology; **Liver**: Structure and Function; **Thiamin**: Properties and Determination; Physiology; **Vitamin B_6**: Properties and Determination; Physiology; **Zinc**: Deficiency

Further Reading

Bode JC, Hanisch P, Henning H *et al.* (1988) Hepatic zinc content in patients with various stages of alcoholic liver disease and in patients with chronic active and chronic persistent hepatitis. *Hepatology* 8: 1605–1609.

Colditz G, Giovannucci E, Rimm EB *et al.* (1991) Alcohol intake in relation to diet and obesity in women and men. *American Journal of Clinical Nutrition* 54: 49–55.

Doll R (1993) Alcoholic beverages and cancers of the digestive tract and larynx. In: Verschuren PM (ed.) *Health Issues Related to Alcohol Consumption*, pp. 125–166. Washington, DC: ILSI Press.

Giovannucci E, Rimm EB, Ascherio A *et al.* (1995) Alcohol, low-methionine/low-folate diets, and risk of colon cancer in men. *Journal of the National Cancer Institute* 87: 265–273.

Gronbaek M, Deis A, Sorensen TI *et al.* (1995) Mortality associated with moderate intakes of wine, beer, or spirits. *British Medical Journal* 310: 1165–1169.

Halsted CH (1995) Alcohol and folate interactions: clinical implications. In: Bailey LB (ed.) *Folate in Health and Disease*, pp. 313–327. New York: Marcel Dekker.

Hultberg B, Berglund M, Andersson A and Frank A (1993) Elevated plasma homocysteine in alcoholics. *Alcoholism, Clinical and Experimental Research* 17: 687–689.

Jacques PF, Bostom AG, Wilson PW *et al.* (2001) Determinants of plasma total homocysteine concentration in the Framingham Offspring cohort. *American Journal of Clinical Nutrition* 73: 613–212.

Klatsky AL (1994) Epidemiology of coronary heart disease – influence of alcohol. *Alcoholism, Clinical and Experimental Research* 18: 88–96.

Kondrup J and Muller MJ (1997) Energy and protein requirements of patients with chronic liver disease. *Journal of Hepatology* 27: 239–247.

Lelbach WK (1975) Cirrhosis in the alcoholic and its relation to the volume of alcohol abuse. *Annals of the New York Academy of Sciences* 252: 85–105.

Leo MA and Lieber CS (1999) Alcohol, vitamin A, and β-carotene: adverse interactions, including hepatotoxicity and carcinogenicity. *American Journal of Clinical Nutrition* 69: 1071–1936.

Lieber CS (1995) Medical disorders of alcoholism [see comments]. *New England Journal of Medicine* 333: 1058–1065.

Mato JM, Camara J, Fernandez de Paz J *et al.* (1999) *S*-Adenosylmethionine in alcoholic liver cirrhosis: a randomized, placebo-controlled, double-blind, multicenter clinical trial. *Journal of Hepatology* 30: 1081–1089.

Mendenhall C, Roselle GA, Gartside P and Moritz T (1995) Relationship of protein calorie malnutrition to alcoholic liver disease: a reexamination of data from two Veterans Administration Cooperative Studies. *Alcoholism, Clinical and Experimental Research* 19: 635–641.

National Institute on Alcohol Abuse and Alcoholism (2000) *Tenth Special Report to the U.S. Congress on Alcohol and Health*. Rockville, MD: US Department of Health and Human Services.

Renaud S and de Lorgeril M (1992) Wine, alcohol, platelets, and the French paradox for coronary heart disease. *Lancet* 339: 1523–1526.

Savage D and Lindenbaum J (1986) Anemia in alcoholics. *Medicine (Baltimore)* 65: 322–338.

Sellers TA, Kushi LH, Cerhan JR *et al.* (2001) Dietary folate intake, alcohol, and risk of breast cancer in a prospective study of postmenopausal women. *Epidemiology* 12: 420–428.

Alcohol Consumption

A Sierksma and K F A M Hulshof, TNO Nutrition and Food Research, The Netherlands
D E Grobbee, University Medical Center Utrecht, Utrecht, The Netherlands
H F J Hendriks, TNO Nutrition and Food Research, The Netherlands

Background

Alcohol-containing beverages play a prominent role in many societies. For centuries, alcohol (ethanol) formed a part of our diet. Even in the first Bible, wine consumption and drunkenness are mentioned. In the early middle ages, alcoholic beverages (beer) became of great importance as there was a lack of clean drinking water.

Abundant epidemiological and clinical evidence shows that there is a J-shaped relationship between alcohol intake and total mortality, with increases in death from cirrhosis, accidents, cancer, and cerebrovascular disease in heavier drinkers (Figure 1).

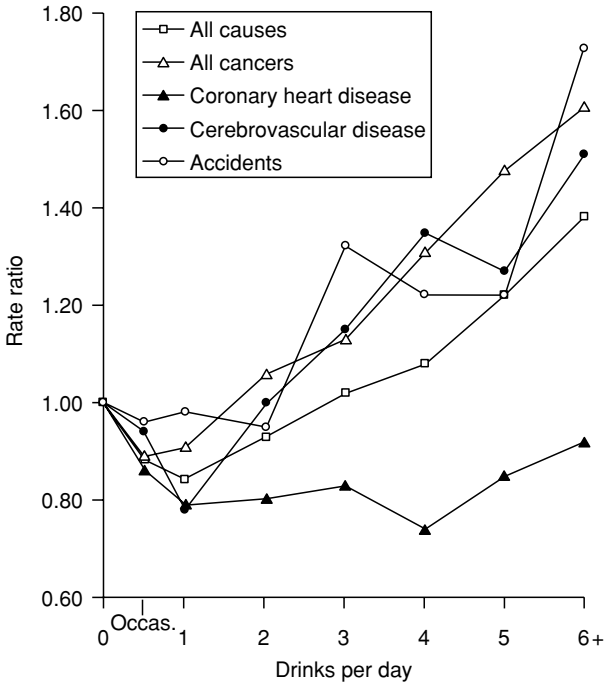

Figure 1 Relationship between alcohol intake and total mortality and death from cirrhosis, accidents, cancer, coronary heart disease, and cerebrovascular disease. From Boffetta P and Garfinkel L (1990) Alcohol drinking and mortality among men enrolled in an American Cancer Society prospective study. *Epidemiology* 1: 342–348, with permission.

Alcohol is consumed according to various patterns. It is becoming increasingly clear that certain drinking patterns may have widely varying effects on health and may lead to social problems. This chapter discusses various aspects of alcohol consumption. First, data on trends in alcohol consumption are given, drinking patterns and their health effects are presented, followed by an overview of the most frequently used methods for the assessment of alcohol intake and their suitability for different purposes. Then, the terms 'moderate and excessive drinking' and 'standard drink' are discussed, and finally, different alcohol control policies and their effectiveness are summarized.

Trends in Alcohol Consumption

Historical Trends

Economic changes, such as those brought about by depression and war, cause fluctuations in the per-capita alcohol consumption. During World War I and II, alcohol consumption was low, and in the mid-1970s after the OPEC oil crisis, per-capita alcohol consumption declined in some countries. In other countries, such a decline occurred in the early 1980s, when unemployment increased, and income stopped rising. Table 1 shows the total alcohol consumption per inhabitant in various developed countries in 1973, 1983, and 1999. Developing countries show increases in alcohol consumption, but these countries still have lower rates of consumption than those in Europe and North America.

Table 1 Total alcohol consumption per liter of pure alcohol per capita in various countries in 1973, 1983, and 1999

Country	*1973*	*1983*	*1999*
Australia	8.6	9.5	7.5
Austria	12.2	10.2	9.3
Canada	7.2	8.2	6.3
Denmark	8.4	10.4	9.5
Finland	5.6	6.4	7.3
France	16.2	14.0	10.7
Germany	11.0	11.0	10.6
Hungary	9.5	11.4	9.7
Italy	13.9	11.4	7.7
Japan	4.9	5.8	6.6
Netherlands	7.6	8.9	8.2
New Zealand	8.8	9.2	7.4
Spain	13.7	12.8	9.9
United Kingdom	6.5	7.2	8.1
USA	7.3	8.1	6.7

Data adapted from *World Drink Trends 2000. International Beverage Consumption and Production Trends*: Henley-on-Thames, UK. NTC Publications, Commodity Board for the Distilled Spirits Industry.

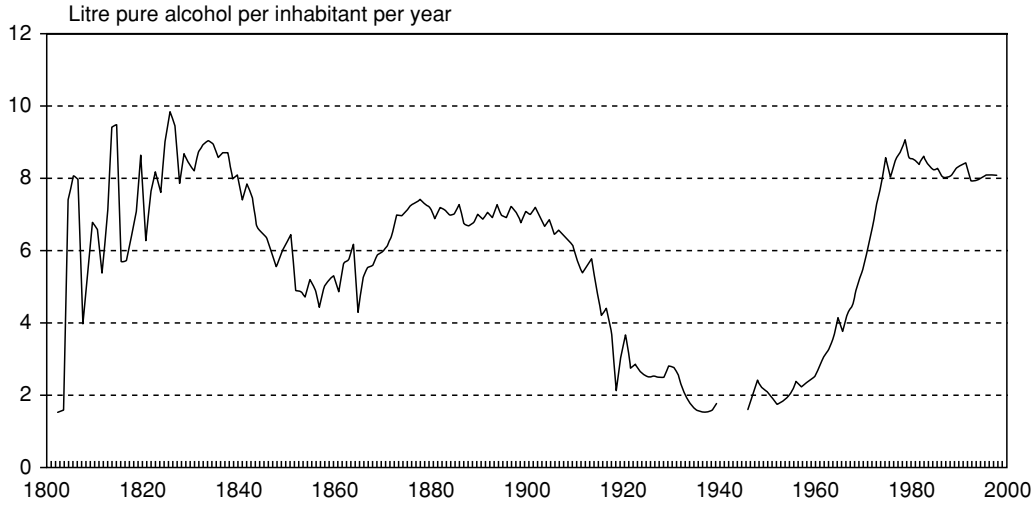

Figure 2 Alcohol consumption in The Netherlands (liters of pure alcohol per inhabitant per year) since 1820. From Statistics Netherlands, with permission.

In The Netherlands, alcohol consumption has been quite stable since 1979, with an average intake of 8 l of pure alcohol per inhabitant per year (Figure 2). The Dutch National Food Consumption Survey over the last 10 years showed that the proportion of alcohol users increased among those aged 13–16 and 16–19 years and that alcohol consumption increased among males aged 16–19 and 19–22 years. Above these age groups, alcohol consumption is somewhat decreased.

Beverage Preference

Wine is the predominant beverage in Mediterranean and Latin American societies. In these regions, wine is commonly consumed with meals. Spirits are the beverages of choice in Nordic and some Eastern European countries, where the drinking pattern is characterized by a low overall consumption and binge-drinking. Beer is the beverage of choice in many parts of Europe, North America, and Australia. In these countries, beer is mainly drunk in leisure time and usually not with a meal. Wine-drinking has become more popular in beer-drinking countries, and beer-drinking in wine-drinking countries. In spirit-drinking countries, both beer and wine consumption have become popular. So differences in beverage preference among preference appear to diminish over time. Table 2 shows the recent consumption figures for alcoholic beverages in various countries.

Table 2 Consumption by beverage type per liter per capita in various countries in 1999

Country	*Spirits*[a]	*Beer*	*Wine*
Australia	3.9	91.2	19.3
Austria	4.0	108.9	30.9
Canada	5.1	68.0	8.9
Denmark	3.2	101.9	29.9
Finland	6.4	80.1	17.5
France	6.9	38.7	57.2
Germany	5.7	127.5	22.9
Hungary	8.6	65.0	30.0
Italy	1.4	27.1	51.5
Japan	7.1	48.0	2.5
Netherlands	4.8	84.2	18.6
New Zealand	4.1	80.7	16.3
Spain	6.9	68.8	33.7
UK	4.3	99.0	14.5
USA	5.3	84.4	7.7

[a]Liters of spirits was calculated, assuming that spirits contain 35 vol% alcohol. Data adapted from *World Drink Trends 2000*. International Beverage Consumption and Production Trends. Henley-on-Thames, UK. NTC Publications, Commodity Board for the Distilled Spirits Industry.

Drinking Patterns

Drinking patterns refer to several aspects of drinking behavior, including the quantity and frequency of consumption, personal characteristics of the drinkers, types of beverage consumed, temporal variations in drinking, settings where drinking takes place, and the activities associated with drinking. These patterns are important determinants of both the positive and negative consequences of drinking.

Quantity and Frequency

The importance of quantity follows from extensive epidemiological research. Light to moderate alcohol consumption is associated with a reduced risk for coronary heart disease, as well as for ischemic stroke

and possibly for type 2 diabetes. Epidemiological and physiological data are in favor of a causal relationship. Proposed protective mechanisms include the stimulation of high-density lipoprotein (HDL)-mediated processes and improvement of hemostatic processes and vascular wall functioning. Excessive alcohol consumption and alcohol abuse may lead to alcoholism or alcohol dependence, with negative effect on almost all parts of the body, including the liver, brain, pancreas, cerebrovascular system, and immune system, and are also associated with premature death.

Both the average amount of alcohol consumed and how drinkers decide to spread the consumption of alcohol over the hours and situations available for drinking are important. Although very little is known about the effects of drinking frequency, intuitively, it seems clear that variations in frequency may have widely varying effects on health and social problems. Binge-drinking, for example, is associated with intoxication, hangover, and injuries.

Individual Characteristics

Drinking is generally associated with particular personal characteristics, like genetics, personality traits, age, gender, and religion.

Genetics Alcohol dependence is a result of an interaction between biological and environmental factors. Twin and adoption studies have shown that there is a genetic factor in the etiology of alcohol dependence, but the exact mechanism of transmission is not known.

Up to about 50% of some Asian groups have inherited deficiencies of the enzyme aldehyde (acetyldehyde) dehydrogenase, inducing relatively high blood levels of acetaldehyde, which causes symptoms such as flushing, nausea, vomiting, palpitations, and hypertension. Deficiency of this enzyme deters the individual from drinking alcohol and reduces the risk of alcohol dependence. About 5–10% of the British and German populations and up to 20% of the Swiss population are reported to have atypical forms of the enzyme alcohol dehydrogenase, which causes more rapid elimination of alcohol.

Personality traits Personality traits such as chronic anxiety, a sense of inferiority, self-indulgent tendencies, antisocial personality disorder, aggressive and impulsive personality, particularly in adolescents, and borderline personality have been linked with alcohol abuse.

Age Young people are more likely to abuse alcohol than older people. They have higher risks of incurring such complications associated with acute intoxication such as accidents, violence, and alcohol poisoning. Physical changes in older people, like increased susceptibility to the depressant effects of alcohol, decreased rates of alcohol catabolism in the liver and decreased percentage of body water, tend to limit excessive alcohol consumption.

Gender Use, abuse, and dependence of alcohol are much more common in males than in females. Social disapproval or prohibition tends to restrict alcohol consumption by females, but it may also promote a different pattern of use, such as drinking at home or in private. The lower percentage of body water and the slower metabolism of alcohol in women may influence the use of alcohol by women and may also be a reason for the increased risk of alcohol problems. Females tend to start abusive drinking much later in life than males and probably also develop alcohol problems later. The inverse association between moderate alcohol consumption and coronary heart disease has been documented in both men and women. Benefits for women are mostly associated with lower quantities of alcohol and appear to be most pronounced in postmenopausal women. This is not surprising since this age group, like men over the age of 50, has a marked increase in cardiovascular risk.

Religion Religion and religiosity are among the most powerful predictors of drinking. The vast majority of persons in Muslim countries abstain from alcohol. Also, among Mormons, Adventists, and various other religious communities, alcohol drinking is not allowed. In countries where drinking is relatively common, attendance at religious services is a powerful predictor of whether and how much a person drinks.

Beverage Type

Some reports have suggested that the type of beverage may be important in coronary heart disease risk. Red wine has received much attention as a potential contributor of antioxidant activity, because of its natural antioxidant compounds, i.e., polyphenols. However, observational and experimental studies on the consumption of beer, wine, and spirits in relation to the risk of coronary heart disease indicate that moderate consumption of all three alcohol-containing beverages is linked with a lower risk. Thus, a substantial portion of the benefit is connected with the alcohol component, rather than with specific non-alcohol components present in the different types of beverages.

Temporal Variations

Drinking is strongly related to leisure time, which is why drinking generally tends to be greatest during nonworking hours and at weekends. Temporal patterns of drinking vary considerably in different cultures and social groups. The term 'alcohol culture' has been used to denote similarities between particular aspects of drinking patterns in different countries. 'Dry' and 'wet' drinking cultures refer to the relatively low or high social acceptance of drinking. In 'wet' cultures, like the Mediterranean area and Latin American societies, drinking with meals and during the day is a common pattern. In 'drier' Northern European cultures, drinking during the day is less accepted and less common. It has been argued that drinking problems are manifested differently in 'dry' cultures: fewer people incur problems, but those who do suffer more severe individual consequences.

Settings and Activities

Another dimension of drinking patterns concerns the settings and activities associated with drinking. Drinking may occur in connection with social or religious rituals or may be related to particular work activities. Drinking can take place in private homes, with or without a meal, or in public places, such as bars, cafés, restaurants, or beer gardens.

Drinking alcoholic beverages leads to a rapid increase in blood alcohol concentration (BAC). After reaching a maximum value, the BAC returns slowly to the base level. The course of the alcohol concentration with time depends on drinking conditions such as the dose and concentration of the alcoholic beverage consumed and whether alcohol is consumed with or without a meal. Studies have shown that BACs after moderate alcohol consumption together with a meal remained half as high as those with moderate alcohol consumption without a meal.

Assessment of Alcohol Consumption

It is very difficult to assess alcohol intake accurately. Although alcohol is part of the normal diet, it is not considered a normal food constituent. It has a highly symbolic value, and its consumption is influenced by cultural differences and social norms.

Methods for determining alcohol intake can be divided into nonintrusive and intrusive estimates. Nonintrusive estimates such as sales figures and collateral reports are less threatening and more objective, because respondents are not aware of them. Intrusive estimates are based on self-reports or biological markers. Unfortunately, a 'gold standard' for alcohol intake assessment is lacking.

Sales Figures

To determine alcohol availability in a population, per-capita sales figures can be used. However, these do not provide information about alcohol consumption in certain groups or individuals. Another drawback is that sales reports do not include untaxed alcohol (duty-free shops), home-brewed alcohol, and export sale.

Self-reports

Methods based on self-reports include face-to-face or telephone interviews, self-administered questionnaires, and diaries. Most of these self-reports can be applied very easily and are relatively cheap. Self-reports are suitable for use in small studies and in large-scale surveys for epidemiological research and can provide beverage-specific information. The quality of the data collected is determined by the order and structure of the questions asked and by the mode of the report and characteristics of the respondent, such as age and culture.

Questions about drinking are threatening and may easily yield socially desirable answers. In population surveys, the average extent of underestimation of alcohol intake can vary between 29 and 83% of the actual intake. When the degree of underestimation is not linked to the actual level of intake, it will not affect the conclusions, and only the quantitative definition of moderate or optimal level of alcohol use will be unrealistic. However, some studies indicate that heavy drinkers underestimate their alcohol intake more than light drinkers do, and in this case, conclusions about dose–response relationships are unreliable. Underestimation may be partly prevented by an anonymous questionnaire. The memory capacity of the respondents may be a limiting factor as well. Common errors in assessing alcohol intake in surveys include selection bias, nonresponse, incomplete time sampling, under- or overestimation of portion size, and recall bias.

Collateral Reports

Collateral reports are provided by a significant 'other person' (e.g., a spouse or friend). The accuracy of these reports is affected by the subject's daily routine making behavior more or less predictable and the frequency of contact between the subject and the collateral. Collateral reports are often used as a reference method.

Biological Markers

The advantage of biological markers is that they are an objective measurement of intake, not affected by socially desirable answers or interviewer effects.

However, they are not suitable for assessing actual level of intake, drinking patterns, or beverage preferences. They can only be used to rank individuals according to their intake. Table 3 gives an overview of the biological markers of alcohol presented in the literature, their purpose and estimated sensitivity (proportion of positives that are correctly identified by the test), and their specificity (proportion of negatives that are correctly identified by the test).

Biological markers of alcohol use are not suitable for assessing habitual low or moderate drinking levels: they can only be used for clinical studies of alcoholics to monitor change in drinking patterns of patients or to check abstinence.

Alcohol Consumption: Moderate and Excessive

Moderate and Excessive Drinking

The literature contains widely different applications of the term 'moderate.' A comparison of average daily amounts defined as light, moderate, and heavy in a sample of recent publications shows that the lower limit of moderate alcohol intakes ranges from 4.5 to 50 g per day, and the upper limit ranges from about 24 to 80 g per day. Table 4 shows some recommendations on drinking levels published by governments or learned bodies.

On a population basis, for the average adult man, the optimal level may be in the range of 10–19 g of alcohol per day, and the noninjurious level may be approximately 30–40 g per day. For a woman, these levels may be < 10 g per day and approximately 10–20 g per day. However, the individual differences in body size, age, and special situations that can increase the degree of risk (pregnancy, driving, diseases, and medications) should also be taken into account.

Light and frequent drinking is suggested to be beneficial, whereas large amounts drunk infrequently are harmful with respect to coronary heart disease.

What is a 'Standard Drink'?

Research outcomes are often confusing when moderate and excessive drinking are expressed in 'standard drinks.' The term 'standard drink' was originally intended to apply to drinks of 'standard' strengths. However, the alcohol content varies among different beers, wines, and distilled spirits, and interpretations differ across countries as to how much alcohol there is in a standard drink (see Table 4). There may also be differences in the standard serving sizes within a given country, depending on the type of beverage served, so the amount of alcohol may be different between the different beverages. Another problem is a lack of uniformity in the definition in which the alcohol content is measured – grams versus ounces of ethanol, American versus British fluid ounces, or measures of alcohol content as a percentage by weight or by volume. The size of drinks poured in serving establishments may not be the same as that defined by the standard drink, and there are significant levels of alcoholic beverage consumption in homes and other private settings where drinks are rarely measured in standard units. These wide discrepancies and variations make any form of international comparison

Table 3 Characteristics of biological markers of alcohol use

Marker	*Purpose*	*Estimated sensitivity (sens.) and specificity (spec.)*
Ethanol in serum, breath, or urine	Recent intake (2–6 h)	Sens. 25% Spec. 100%
Methanol in body fluids or breath	Recent (5–15 h) and chronic intake	Higher sensitivity than ethanol
MCV of erythrocytes	Chronic intake	Sens. 35% Spec. 40–90%
Erythrocyte acetaldehyde in plasma	Recent intake (several days)	Low sensitivity
Transferases (GGT, ASAT, ALAT) in serum	Chronic intake, screening	GGT: sens. 45%, spec. 75% ASAT and ALAT: lower than for GGT
5-Hydroxytryptophol in urine	Recent intake (6–15 h)	Sens. 70% Spec. 95%
HDL in serum	Recent and chronic, moderate and high intake	Sens. 30%
FAEE in serum	Recent intake (24 h)	Higher sensitivity than ethanol
CDT in serum	Chronic excessive intake, screening	Sens. 75% Spec. 90%
AANB in serum	Chronic intake, monitoring	Low sensitivity

AANB, α-amino-*N*-butyric acid; ALAT, alanine aminotransferase; ASAT, aspartate aminotransferase; CDT, carbohydrate-deficient transferrin; FAEE, fatty acid ethyl esters; GGT, γ-glutamyltransferase; HDL, high-density lipoprotein. MCV, mean corpuscular volume.
Data adapted from MacDonald I (ed.) (1999) *Health Issues Related to Alcohol Consumption*. Brussels: ILSI Europe.

Table 4 Alcohol content of a 'standard drink' as defined in various countries, and some daily consumption guidelines recommended by governments or learned bodies

Country	*Size of standard drink (grams of absolute alcohol)*	*Daily consumption guidelines (grams of absolute alcohol)*	
		Men	*Women*
Australia	10	40	20
Austria	10	30	20
Canada	13.5	13.5	13.5
Denmark	12	36	27
Finland	11	41	27
France	12	60	36
Hungary	17		
Iceland	9.5		
Ireland	8		
Italy	10		
Japan	19.75	39.5	
Netherlands	9.9	2–3 units a few times a week	2–3 units a few times a week
New Zealand	10	30	20
Portugal	14		
Spain	10	30	20
UK	8	$\leq$24–32	$\leq$16–24
USA	14	$\leq$28	$\leq$14

Data adapted from ICAP (1998b) *Safe Alcohol Consumption: A Comparison of Nutrition and Your Health: Dietary Guidelines for Americans and Sensible Drinking. ICAP Report No.1*. Washington, DC: International Center for Alcohol Policies; ICAP (1998c) *What is a 'Standard Drink'? ICAP Report No. 5*. Washington, DC: International Center for Alcohol Policies.

of drinks and consumption data difficult, so assessment and comparison of data from many countries would be greatly facilitated by the uniform expression of alcohol intakes and by the adoption of a uniform international definition of a standard drink.

Alcohol Control Policies

In order to reduce the prevalence of alcohol-related problems, many countries apply control measures that affect the availability and/or the price of alcoholic drinks. In addition, governments may use public health campaigns and alcohol advertising as instruments to diminish the prevalence of alcohol-related problems. Alcohol control policies may be directed at reducing the average alcohol consumption among the total population (public health approach) or at reducing the number of heavy users or drinking in high-risk situations (high-risk approach). Alcohol control measures are more effective as public support is higher, and so education of the public and possibly self-regulation by the alcohol industry may be important contributory measures.

Price

Taxation of alcoholic beverages is essentially for the purpose of raising revenue. It is not generally regarded as a primary tool for controlling consumption, although it is used as such in some countries. In most countries, alcoholic drinks are perceived to be luxury items and, as such, candidates for taxation. Alcoholic drinks are price-elastic, which means that the demand is sensitive to price changes. This is the main reason for the argument that taxation should be used to reduce consumption. However, it has been found that elasticity values relating to different regions, different periods, and different types of alcohol vary considerably.

More important is the effect of price increases on different types of consumers. Some claim that heavy drinkers are at least as responsive to price increases than light-to-moderate consumers. Others claim that heavy drinkers reduce their consumption much more than other drinkers after excise duties are increased.

Advertising and Alcohol Availability

A commonly held assumption is that alcohol advertising increases alcohol consumption. The main objective of advertising is to persuade customers to continue to enjoy, or to switch their preference to, the brand being advertised. Several studies have demonstrated that advertising does not increase consumption. Also, restrictions on advertising do not automatically result in a decline in total alcohol consumption.

It has been suggested that there is an association between hours of sale and outlet density, and the consumption of alcohol, but some studies show no

Table 5 Minimum drinking and purchasing age laws in various countries

Country	*Minimum drinking age (MDA)*	*Minimum purchasing age (MPA)*
Australia	18	18
Austria	16	16
		18 (spirits)
Canada	In Alberta and Quebec: 19	All provinces have their own legal MPA
	In all other provinces: 18	
Denmark		18
Finland	18	18 (for all beverages up to 21% alcohol by volume)
		20
France	16	16
Hungary	18	
Iceland	20	
Ireland	18	18
Italy	16	16
Japan	20	20
Netherlands		16 (beer and wine)
		18 (spirits, 16 if accompanied by an adult)
New Zealand	20	20
Portugal	No MDA	No MPA
Spain	16	16
UK	18	18
USA	21	21

Data adapted from ICAP (1998) *Drinking Age Limits. ICAP Report No. 4*. Washington, DC: International Center for Alcohol Policies.

such association. Prohibition of alcohol has been shown to be an ineffective and even counterproductive measure. Although the alcoholic cirrhosis mortality and other alcohol problems may decrease, it might not produce positive results in public health terms as, for example, organized crime increased in the United States after complete alcohol ban.

Drinking Age Limits

Drinking age laws cover a broad spectrum of behaviors concerned with where, when, and under what circumstances alcoholic beverages can be purchased and consumed. Some countries focus their legislation on both a minimum legal drinking age and a minimum age at which beverage alcohol can be purchased, and these ages may be different. Other countries do not address a minimum age for consumption. Table 5 shows some minimum drinking and purchasing age laws. Both consumption and purchasing laws in most countries are set at 18. The national laws generally apply to drinking age limits for venues outside the home, such as taverns, bars, restaurants, nightclubs, and similar establishments. The UK is the only country that legislates a minimum consumption age in the home (from age 5 with parental consent).

Effectiveness of Alcohol Control Policies

Price policy may be an effective means of reducing alcohol consumption and alcohol-related problems as long as increases in price keep up with or outstrip rises in income. However, the effectiveness may depend very much on the population group affected and on socio-economic and demographic factors. Also, stable political and economic conditions as well as appropriate law enforcement and administrative systems are prerequisites for success.

See also: **Alcohol**: Properties and Determination; Metabolism, Beneficial Effects, and Toxicology; **Body Composition**; **Dietary Surveys**: Measurement of Food Intake; Surveys of National Food Intake; Surveys of Food Intakes in Groups and Individuals

Further Reading

Boffetta P and Garfinkel L (1990) Alcohol drinking and mortality among men enrolled in an American Cancer Society prospective study. *Epidemiology* 1: 342–348.

Grant M and Litva J (eds) (1998) *Drinking Patterns & Their Consequences*. London: Taylor & Francis.

ICAP (1998a) *Drinking Age Limits. ICAP Report No. 4*. Washington, DC: International Center for Alcohol Policies.

ICAP (1998b) *Safe Alcohol Consumption: A Comparison of Nutrition and Your Health: Dietary Guidelines for Americans and Sensible Drinking. ICAP Report No. 1*. Washington, DC: International Center for Alcohol Policies.

ICAP (1998c) *What is a 'Standard Drink'? ICAP Report No. 5*. Washington, DC: International Center for Alcohol Policies.

MacDonald I (ed.) (1999) *Health Issues Related to Alcohol Consumption*. Brussels: ILSI Europe.

Pikaar NA, Wedel M and Hermus RJJ (1988) Influence of several factors on blood alcohol concentrations after drinking alcohol. *Alcohol & Alcoholism* 23: 289–297.

Van Tol A and Hendriks HFJ (2001) Moderate alcohol consumption: Effects on lipids and cardiovascular disease risk. *Current Opinion in Lipidology* 12: 19–23.

World Drink Trends 2000 International Beverage Consumption and Production Trends. Henley-on-Thames, UK: NTC Publications, Commodity Board for the Distilled Spirits Industry.

Ales *See* **Beers**: History and Types; Wort Production; Raw Materials; Chemistry of Brewing; Biochemistry of Fermentation; Microbreweries

Algae *See* **Single-cell Protein**: Algae; Yeasts and Bacteria; **Marine Foods**: Production and Uses of Marine Algae; Edible Animals Found in the Sea; Marine Mammals as Meat Sources

ALKALOIDS

Contents

Properties and Determination

M Wink, University of Heidelberg, Heidelberg, Germany

Background

Plants produce a wide variety of secondary metabolites, many of which function as allelochemicals or signal compounds to attract pollinating insects and seed- or fruit-dispersing animals. Their main function is chemical defense against herbivores, microrganisms, or competing plants. Over 13 000 nitrogen-containing secondary metabolites have been described so far in plants. Alkaloids contribute over 12 000 compounds, followed by amines, nonprotein amino acids, cyanogenic glycosides, and glucosinolates.

Definition

The definition of alkaloids has changed throughout the years. Formerly, this class of secondary compounds was restricted to plant bases with a heterocyclic nitrogen atom. Exocyclic nitrogen bases were termed 'pseudoalkaloids.' Other definitions demanded that the skeleton of alkaloids should derive from amino acids or that these bases have explicit pharmacological activities. At present, alkaloids are defined in a more pragmatic way; they include all nitrogen-containing natural products that are not otherwise classified as peptides, nonprotein amino acids, amines, cyanogenic glycosides, glucosinolates, cofactors, phytohormones, or primary metabolites (such as purine and pyrimidine bases). Even a number of antibiotics produced by bacteria or fungi are therefore included in the group of alkaloids.

Occurrence

Alkaloids have been detected in about 15% of plants, bacteria, fungi, and even in animals. Within the plant kingdom, they occur in primitive groups such as *Lycopodium* or *Equisetum*, in gymnosperms and angiosperms. In higher plants (angiosperms) some families contain more alkaloid-containing taxa than

others. Such alkaloid-rich' taxa include Papaveraceae, Berberidaceae, Fabaceae, Boraginaceae, Apocynaceae, Asclepiadaceae, Asteraceae, Liliaceae, Gnetaceae, Ranunculaceae, Rubiaceae, Solanaceae, and Rutaceae. Also, several food plants and food items may contain alkaloids (Table 1).

It has been speculated that alkaloids evolved early in evolution and were already present at the time (*c.* 200 million years ago) when the angiosperms began to radiate. In general, specific alkaloid types are restricted to particular systematic units and are therefore of some importance for the systematics, taxonomy, and phylogeny of plants. For example, benzylisoquinoline alkaloids are typical for Papaveraceae, Berberidaceae, and Ranunculaceae, which seem to be phylogenetically related. Other alkaloids occur in phylogenetically unrelated plant families. Ergot alkaloids occur in fungi (*Claviceps*) but also in members of the Convolvulaceae; quinolizidine alkaloids are typical for some Fabaceae, but have also been detected in Berberidaceae (*Caulophyllum*, *Leontice*). It has been speculated that the genes encoding enzymes leading to the main quinolizidine alkaloid skeleton are distributed much more widely in the plant kingdom, but are normally switched off. Alternatively, convergent evolution and horizontal gene transfer have been suggested.

Isolation and Detection

Alkaloids form free bases in alkaline solutions. The free base is usually lipophilic and mostly insoluble in water but soluble in organic solvents (ethanol, methanol, diethylether, methylene chloride). At higher hydrogen ion concentrations (i.e., pH < 7), alkaloids occur in a protonated state and are usually

Table 1 Alkaloids present in food plants, beverages, stimulants, or hallucinogens

Substance	*Occurrence*	*Activity*
Alkaloids in food		
Solanine and other steroid alkaloids	Solanum (potato, tomato)	Membrane disruption, Neuroreceptor interaction, mutagenicity
Pyrrolizidine alkaloids	*Symphytum* (comfrey); honey (if bees have visited PA plants)	DNA and protein alkylation; mutagenicity, cancer
Cycasin	Cycas, other cycads	Mutagenicity
Ergotalkaloids	*Claviceps purpurea*	Neuroreceptor interaction; vasoconstriction; uterus contraction
Saxitoxin	Algae; ends up in food chain (molluscs; fish)	neurotoxin (Na^+ channel)
Lupanine and other quinolizidine alkaloids	*Lupinus*; other genistoids	Interaction with AChR[a], Na^+, K^+ channels
Pelletierine	*Punica granatum*	Interaction with AChR
Alkaloids in beverages		
Caffeine, theophylline, theobromine	*Coffea, Thea, Paullinia, Ilex paraguarensis*	Stimulant
Quinine	*Cinchona succirubra*	Bitter tonic; neuroreceptor and ion channel interaction; central stimulant
Ephedrine and related compounds	*Ephedra*; *Catha edulis*	
Alkaloids as stimulants and hallucinogens		
Morphine	*Papaver somniferum*, opium	Hallucinogen
N-Methyltryptamine	Fungi, plants	Central stimulant
N, *N*-dimethyltryptamine, 5-methoxy-*N*, *N*-dimethyltryptamine	Mimosaceae: Anadenanthera (syn. Piptadenia), *Mimosa hostilis*; Myristicaceae: *Virola*; Malpighiaceae: *Banisteriopsis*, Poaceae: *Phalaris*	Central stimulant, hallucinogen
Serotonin	Fungi (*Amanita*); stinging hairs of *Urtica*, *Laportea*, *Jatropha urens*, *Mucuna pruriens*; seeds and fruits	
Bufotenine	Fungi (*Amanita*); animal poisons (*Cnidaria*, spider, scorpions, wasps, toads)	
Psilocybin, psilocin	Fungi (*Psilocybe*, *Stropharia*, *Conocybe*, *Panaeolus*, *Gymnophilus*, *Psathyrella*, etc.)	
Mescalin	*Lophophora williamsii*; other cacti	Hallucinogen
Harmalin and other β-carboline alkaloids	*Peganum harmala*, *Banisteriopsis caapi*	Hallucinogen
Cocaine		Hallucinogen
Arecoline	Nuts of Areca palm	Stimulant

[a]AChR, acetylcholine receptor.

soluble in water but insoluble in apolar organic solvents. The different solubilities are useful to isolate and purify alkaloids. In the laboratory, alkaloids are often solubilized from plant material by dissolving them in acidic solutions (e.g., 0.5 M HCl). Treatment of this solution with organic solvents will remove nonalkaloidal substances. The solutions are brought to pH > 12 in the next step and extracted with CH_2Cl_2, ethylacetate, or ethylether to yield free alkaloids.

Plant extracts are often analyzed by thin-layer chromatography (TLC) as a first screening to find out whether alkaloids are present or not. A number of reagents give typical color reactions, such as Dragendorff's, Mayer's reagent, etc. Because alkaloids are typically present in complex mixtures consisting of two to five main and up to 20–50 minor alkaloids, TLC does not have sufficient separation capacities. Better methods are high-performance liquid chromatography (HPLC) and capillary gas–liquid chromatography (GLC). The latter method is extremely useful because it has a strong separation power and is very sensitive and selective if a nitrogen-specific detector is used. Furthermore, GLC can be directly coupled with a mass spectrometer, allowing mass spectra to be obtained, even from very minor components. Since many alkaloids have been analyzed by mass spectrometry, already a large collection of mass spectra is available, making it possible to identify many of the known alkaloids unambiguously. Usually, mass spectra are recorded in the electron impact (EI) mode, which promotes fragmentation. If information on the molecular ions is needed (which can be elusive in EI-MS), other MS techniques, such as chemical ionization (CI), field desorption (FD), or fast-atom bombardment (FAB) are the methods of choice.

HPLC tends to be less sensitive and of lower separation capacity than capillary GLC. However, modern photodiode array detectors are very helpful to identify known metabolites by UV-VIS spectroscopy. Today, also, HPLC and capillary electrophoresis can be coupled to a mass spectrometer, thus widening the use of MS for the analysis of natural products.

In addition, HPLC has a major advantage in that it is possible to isolate a compound in milligram quantities that allow structural elucidation by nuclear magnetic resonance (^1H, ^{13}C). Nuclear magnetic resonance is the method of choice if unknown structures are to be elucidated, whereas mass spectrometry is extremely useful for identifying previously described substances or substances that are slightly different to known compounds.

If small quantities (femtogram or nanogram amounts) of known alkaloids need to be detected routinely, immunological procedures such as radioimmunoassays (RIA) or enzyme immunoassays (EIA, ELISA) should be appropriate. In order to obtain specific antibodies, the alkaloid in question has to be coupled chemically to a large protein, such as albumin, usually by some sort of spacers (e.g., succinic acid).

Biosynthesis

The skeleton of most alkaloids is derived from amino acids (Figure 1, Table 2), although moieties from other pathways, such as terpenoids, are often combined. In addition, in a number of alkaloids (e.g., steroid alkaloids), the nitrogen (deriving from glutamine or other NH_2 sources) is added near the end of a biosynthetic pathway, i.e., the alkaloid skeleton does not stem from amino acids. Biosynthetic pathways have been worked out in detail for a few alkaloids.

Accumulation and Storage

Although the exact site of alkaloid formation in a plant cell has been elucidated for a few species only, it is certainly correct to assume that most compounds are synthesized in the cytoplasm. Membrane-enclosed vesicles have been implied in the biosynthesis of berberine. Not only is the chloroplast the site of photosynthesis and related processes, but it also harbors a number of biosynthetic pathways, such as those for fatty acids and amino acids. In addition, quinolizidine alkaloids (QA) are synthesized in the chloroplast stroma; thus, both the alkaloid and its amino acid precursor, L-lysine, share the same compartment. QA formation is regulated by light and displays a diurnal rhythm. Light regulation seems to be triggered by (1) lysine availability (it is also made during the day), (2) the change in stromal hydrogen ion concentration to pH 8 (enzymes of QA formation have a pH optimum at pH 8), and (3) the reduction of QA enzymes by thioredoxin (reduced thioredoxin is generated under illumination).

Alkaloids are not formed in the extracellular space or in the vacuole. The storage of high concentrations of alkaloids is a prerequisite for their allelochemical roles as defense compounds. Since these concentrations would interfere with the normal metabolisms, the allelochemicals are safely stored in the vacuole (Table 3) of often specialized cells or tissues (such as epidermis). The vacuole of these alkaloid-accumulating cells has been termed the 'toxic' or 'defense' compartment. A number of plants produce latex, which, in addition to its gluing properties (think of insect mandibles!), often contains defense chemicals, such as alkaloids (morphine and related

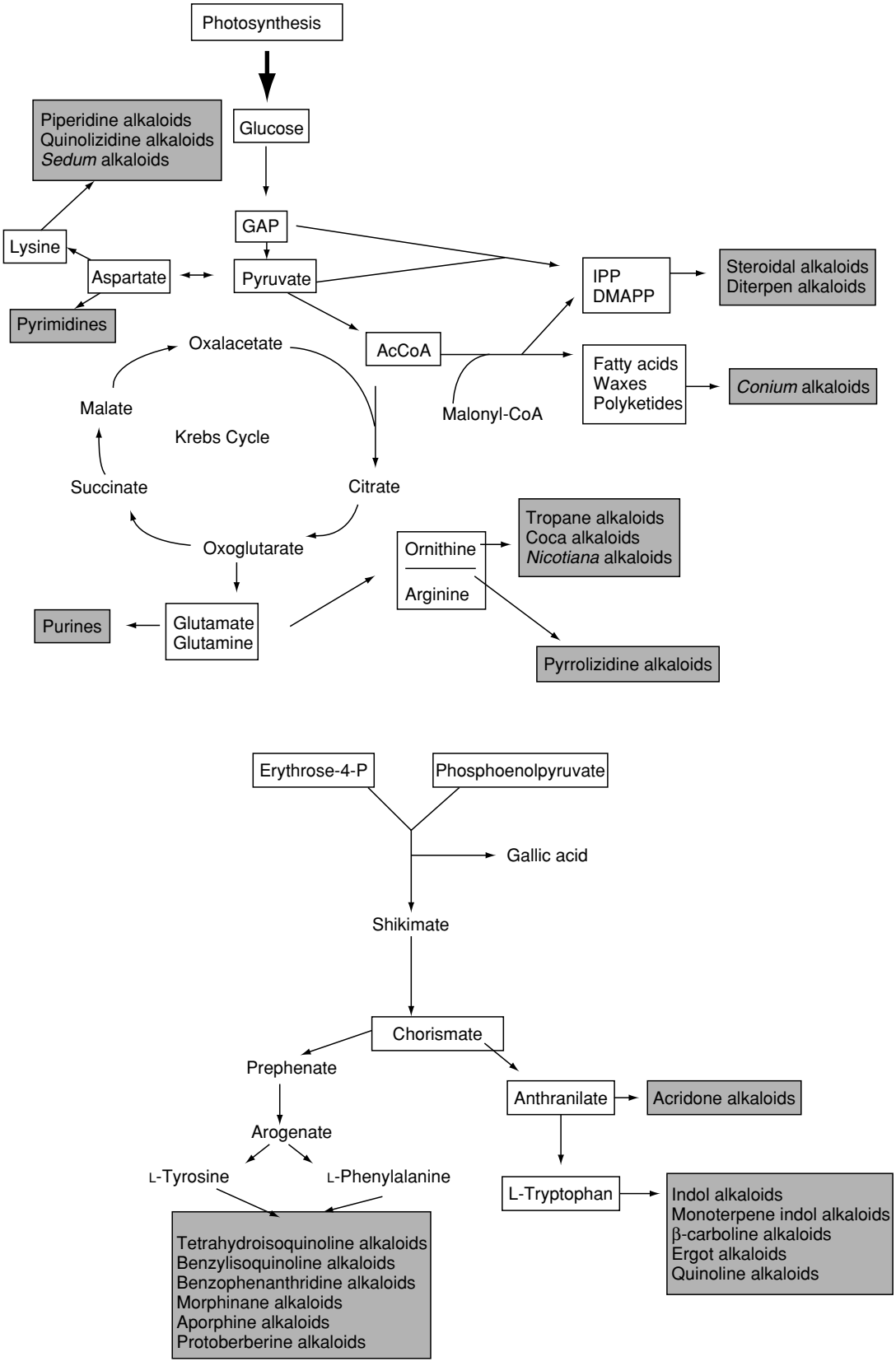

Figure 1 Schematic overview of biosynthetic pathways leading to different groups of alkaloids.

Table 2 Examples for important alkaloids and their biosynthetic precursors

Amino acid	*Alkaloid*	*Main occurrences*	*Example structure*
Lysine	Quinolizidine alkaloids	Fabaceae	Lupanine, sparteine, cytisine
	Lycopodium alkaloids	Lycopodiaceae	Lycopodine
	Piperidine alkaloids	Punicaceae, Crassulaceae	Pelletierine, sedamine
Ornithine	Tropane alkaloids	Solanaceae, Erythroxylaceae	Hyoscyamine, cocaine
	Pyrrolizidine alkaloids	Asteraceae, Boraginaceae, Fabaceae	Senecionine, heliotrine
	Nicotiana alkaloids	Solanaceae	Nicotine, anabasine
Tryptophan	Monoterpene indol alkaloids	Apocynaceae, Aslepiadaceae ...	Strychnine, vincamine, Yohimbine, ajmalicine ...
	Simple indole alkaloids	Fabaceae	Physostigmine
	Quinoline alkaloids	Rubiaceae, Cornaceae	Quinine, cinchonine, camptothecine
	Ergot alkaloids	*Claviceps*, Convolvulaceae	Ergotamine, lysergic acid
	β-carboline alkaloids	Loganiaceae, Zygophyllaceae	Harman, harmaline
Phenylalanine/tyrosine	*Ephedra* alkaloids	Ephedraceae	Ephedrine
	Tetrahydro-isoquinoline alkaloids	Rubiaceae	Emetine
	Benzylisoquinoline alkaloids	Papaveraceae, Berberidaceae	Papaverine
	Benzophenanthridine alkaloids	Papaveraceae	Sanguinarine
	Protoberberine alkaloids	Berberidaceae, Papaveraceae	Berberine
	Morphinane alkaloids	Papaveraceae	Morphine, codeine
	Aporphine alkaloids	Monimiaceae	Boldine
	Phenylethyl-isoquinoline alkaloids	Colchicaceae	Colchicine
	Aristolochia alkaloids	Aristolochiaceae	Aristolochic acid
Anthranilic acid	*Ruta* alkaloids	Rutaceae	Skimmianine, dictamine

Table 3 Storage of alkaloids in the vacuole

Alkaloid	*Species*
Leaf vacuoles	
Lupanine	*Lupinus*
Sparteine	*Cytisus*, *Lupinus*
Hyoscyamine	*Atropa*
Nicotine	*Nicotiana*
S-Coulerine	*Fumaria*
S-Reticuline	*Fumaria*
Ajmalicine	*Catharanthus*
Serpentine	*Catharanthus*
Catharanthine	*Catharanthus*
Betalaines	*Beta*, *Chenopodium*
Senecionine-*N*-oxide	*Senecio*
Capsaicin	*Capsicum*
Latex vesicles	
Sanguinarine	*Chelidonium*
Berberine	*Chelidonium*
Morphine and other morphinane alkaloids	*Papaver*

benzylisoquinoline alkaloids in Papaveraceae; protoberberine and benzophenanthridine alkaloids in *Chelidonium*; lobeline and other piperidine alkaloids in *Lobelia*) and terpenoids (e.g., phorbole esters). The alkaloids are selectively sequestered in small (diameter < 1 μm) latex vesicles and reach local concentrations of up to 1 M.

Storage in the vacuole or in vesicles demands that alkaloids pass the tonoplast and accumulate within the vacuole against a concentration gradient. For the passage across the tonoplast, three mechanisms are plausible:

- simple diffusion (takes place in case of lipophilic alkaloids, e.g., nicotine, ajmalicine, vinblastine, colchicine);
- carrier-mediated transport (in case of polar and charged alkaloids, which is the rule for most alkaloids under physiological conditions, e.g., hyoscyamine, lupanine, reticuline, scoulerine, senecionine);
- membrane fusion (in case of alkaloids that are formed in vesicle-enclosed compartments, e.g., berberine).

Because alkaloids are sequestered against a concentration gradient in the vacuole or in latex vesicles, the driving force of uphill accumulation needs to be determined. In some instances, vesicles or vacuoles contain alkaloid binding or complexing compounds. For example, latex vesicles of *Chelidonium majus* contain between 500 and 1200 mM chelidonic acid, which binds or complexes berberine and benzophenanthridine alkaloids. It has been shown experimentally that chelidonic acid acts as a trapping agent that causes the apparent uphill transport resulting in alkaloid concentrations in vesicles of 500–1200 mM. Protonation, organic acids, and specific peptides may constitute other trapping mechanisms. There is some evidence that amino acids and some ions are transported into the vacuole with aid transporters that are fueled by proton–substrate antiport mechanisms.

Protons are enriched in the vacuole, which generally has hydrogen ion concentrations of 0.001–1 mM (=pH 6–3), by proton-translocating ATPases and pyrophosphatases. In analogy, it has been assumed (supported by experimental evidence) that carrier transport of alkaloids is achieved by a proton–alkaloid antiport mechanism or with the aid of ABC transporters.

Whereas a few alkaloids are formed ubiquitously in all plant organs, organ- or even tissue-specific formation seems to be more common (Table 4). Since most eukaryotic genes are regulated in a cell-, tissue-, and organ-specific manner, genes of alkaloid formation seem to be no exception. The corresponding promotors are presumed to be regulated by respective regulatory proteins. This conclusion is important for the interpretation of results obtained with cell suspension cultures. Whereas alkaloid formation is active in differentiated systems (root or shoot cultures), it is usually reduced, or even absent, in undifferentiated suspension cultured cells. It is likely that the corresponding alkaloid genes are just 'switched off.' The production of berberine and sanguinarine in cell cultures seems to be more the exception than the rule.

Alkaloids are stored predominantly in tissues that are important for survival and reproduction, which include actively growing young tissues, root and stem bark, flowers (especially seeds), seedlings, and photosynthetically active tissues. Alkaloid contents in storing organs can be quite high, reaching up to 10% of dry weight in some instances, which is important if the alkaloids are to function as effective defense compounds. In several herbaceous plants, alkaloids are stored in epidermal and subepidermal tissues (e.g. cocaine, colchicine, aconitine, steroidal alkaloids, nicotine, veratrine, buxine, and coniine), which have to ward off small enemies (insects, microorganisms) in the first place. In lupins and broom, quinolizidine alkaloid concentrations are up to 200 mM in epidermal tissue, whereas mesophyl tissue has values below 5 mM. Some plants possess typical alkaloid-storing cells, called 'idioblasts.' These idioblasts have been detected in *Corydalis* (for corydaline), *Sanguinaria* (for sanguinarine), *Ruta* (for rutacridones), *Catharanthus* (for indole alkaloids), and *Macleaya* (for protopine). Short- and long-distance transport are often required to reach these sites of accumulation.

Alkaloid patterns usually vary between the site of synthesis and the sites of accumulation, since a number of secondary substitutions may take place in the latter tissues, or transport may be selective, distributing differing cocktails. In addition, alkaloid profiles of seeds and seedlings often differ from those of the mature plant. Both patterns and concentrations usually change during the development of plants and the annual cycle. In general, alkaloid levels are markedly reduced in senescing tissues, so that shedded leaves are often nearly alkaloid-free. Alkaloid formation and storage may be influenced by environmental stress, such as wounding or infection.

Transport and Turnover

A number of alkaloids are synthesized and stored in all parts of plants, whereas others are restricted to a particular organ. Alternatively, alkaloids are synthesized but accumulated in many organs, which usually demands long-distance transport. Theoretically, this transport could procede sym- or apoplastically. However, utilization of established transport routes, such as xylem or phloem, seems to be more common. Because it is technically difficult to sample and analyze xylem and phloem sap, only a limited number of appropriate data are available (Table 5). Besides long-distance transport, short-distance and intracellular transport need to be reviewed. In general, alkaloids are synthesized in the cytosol or in membrane-enclosed vesicles (endoplasmic reticulum, mitochondria, chloroplasts), but are accumulated and sequestered in the vacuole.

In general, alkaloids are not end products of metabolism but can be degraded, which seems to be

Table 4 Examples for organ-specific biosynthesis of alkaloids

Alkaloid	*Organ*	*Species*
Tropane alkaloids	Roots	*Atropa*, *Datura*, *Hyoscyamus*, *Mandragora*
Nicotine	Roots	*Nicotiana*
Senecionine and other PA	Roots	*Senecio*
Emetine	Roots	*Cephaelis*
Sanguinarine	Roots	*Sanguinaria*
Betalaines	Roots, shoots	*Beta*
Quinine	Stem bark	*Cinchona*
Berberine	Stem and root bark	*Berberis*, *Mahonia*
Caffeine	Green tissue	*Coffea*
Quinolizidine alkaloids	Leaves and other photosynthetic tissues	*Lupinus, Cytisus, Laburnum*, *Baptisia*
Steroid alkaloids	Roots, tubers, leaves	*Solanum*

Table 5 Long-distance transport of alkaloids by phloem or xylem

Alkaloid	*Xylem*	*Phloem*	*Occurrence*
Lupanine, sparteine		X	*Lupinus*, *Cytisus*
Cytisine		X	*Laburnum*, *Petteria*, *Genista*, *Spartium*
Matrin		X	*Sophora*
Senecionine (*N*-oxide)		X	*Senecio*
Aconitine	?	X	*Aconitum*
Nicotine	X		*Nicotiana*
Hyoscyamine	X		*Atropa*
Scopolamine	X		*Datura*, *Hyosycamus*
Swainsonine		X	*Astragalus*

plausible because nitrogen is a limited nutrient for plants. As discussed previously, alkaloids stored in seeds are partly degraded during germination and seedling development, and their nitrogen is probably used for the synthesis of amino acids. Degradative pathways have not been worked out yet.

In addition to this developmentally specific recycling, there is evidence that a number of alkaloids are turned over all the time, with half-lives of between 2 and 48 h. Examples are nicotine, QAs, and tropane alkaloids. Alkaloid turnover is often quite substantial in cell cultures: *Lupinus* callus cultures are even able to live on the QA sparteine as a sole nitrogen source for more than 6 months.

How can we explain this phenomenon? A number of alkaloids are allelochemicals and affect molecular targets such as receptors of neurotransmitters (tropanes, nicotine, etc.). For this interaction, a correct stereochemical configuration is required. Because alkaloids may oxidize or give rise to racemic mixtures spontaneously, a continuous turnover would make sure that a sufficient concentration of active compounds is always available, similar to the situation of protein turnover.

Functions

Although several alkaloids and other secondary metabolites have been used by mankind for thousands of years as dyes (e.g., indigo, shikonine), flavors (e.g., vanillin, capsaicin, mustard oils), fragrances (e.g., rose oil, lavender oil, and other essential oils), stimulants (e.g., caffeine, nicotine, ephedrine), hallucinogens (e.g., morphine, cocaine, mescaline, hyoscyamine, scopolamine, tetrahydrocannabinol), insecticides (e.g., nicotine, piperine, pyrethrin), vertebrate and human poisons (e.g., coniine, strychnine, aconitine) and even therapeutic agents (e.g., atropine, quinine, codeine, cardenolides, etc.), their putative functions have been discussed controversially.

Alkaloid biology is tightly connected with the basic physiology of plants. Many of the features described before would make no sense if these compounds did not have a vital function for the producer. As a common theme, it has been observed that plants that produce seeds rich in energy supplies (carbohydrates, lipids, proteins) concomitantly accumulate potent chemical defense compounds, often alkaloids, nonprotein amino acids, cyanogenic glycosides, glucosinolates, protease inhibitors, lectins, or other toxalbumins. Their presence in seeds can be mutually exclusive, i.e., legume seeds store either alkaloids (e.g., quinolizidines, pyrrolizidines) or nonprotein amino acids, but not both at the same time. During germination, the breakdown of nutrient reserves is a general procedure and usually includes the nitrogenous defense compounds. They serve a double purpose, i.e., that of N-storage and protection. They are thus degradable and toxic N-storage compounds.

The main function is obviously that of chemical defense against herbivores (insects, other arthropods, and vertebrates), which can be deduced from the fact that many alkaloid have a high affinity for receptors of neurotransmitters that are present only in animals. In some instances, alkaloids play a role (additionally) in the antimicrobial defense (against bacteria, fungi, viruses) and even in the interaction between plants (allelopathy).

Alkaloids are certainly multipurpose compounds that, depending on the situation, may be active in more than one environmental interaction. For example: QAs are certainly the most important defense chemical in Fabaceae against insects and other herbivores, but they also influence bacteria, fungi, viruses, and even the germination of other plants. In addition, they are employed as degradable N-transport and N-storage compounds.

Alkaloids repel or deter the feeding of many animals (many have a bitter or pungent taste to humans and other vertebrates) or are toxic if ingested. In microorganisms and competing plants, a reduction of growth and antibiosis are usually the visible effects of alkaloid intoxication. How are these diverse effects being achieved? Although most compounds have not been studied in full detail, an impressive number of cellular and molecular targets have been identified that are selectively inhibited or modulated by alkaloids. As a consequence of such interactions, organ malfunctions (heart, lung, liver, kidney, CNS disorders) result that may impair reproduction and fertility in animals and other organisms, or simply kill them.

Because many alkaloids have been shaped during evolution by 'molecular modeling,' many of them are

used by humans as medicinal compounds; allelochemicals may have positive effects if used at non-toxic concentrations.

Presence at the Right Concentration at the Right Time and Place

To be effective, alkaloids need to be present at the right time, site, and concentration. Alkaloid metabolism and biochemistry seem to have been optimized and coordinated in most systems to fulfill this prerequisite. An interesting variation can be seen in some plants, where alkaloid formation is enhanced by wounding or microbial attack, i.e., in case of emergency, the production of defense compounds is stimulated.

How effective are alkaloidal defenses? In lupins, which normally produce high amounts of QAs, mutants have been selected (so-called sweet lupins) that accumulate only trace amounts of alkaloids. When both alkaloid-rich and low-alkaloid lupins are planted in the field without any fences and phytoprotectives, a dramatic effect can often be seen (Figure 2). Whereas the low-alkaloid lupins were selectively grazed or infested by herbivores, their alkaloid-rich counterparts remained almost undisturbed. More experimental data are certainly needed, but the defensive role of alkaloids seems to be beyond doubt.

Similar to the situation in sweet lupins, in which plant breeders have eliminated the alkaloid trait, also in other crop plants, secondary metabolites, which had evolved to serve as defense compounds, have been bred away or strongly decreased. As a consequence, many crop plants have lost their original resistance to pests and herbivores. Man-made chemistry (i.e., synthetic pesticides) has to be used if these crop plants have to be cultured.

Exceptions to the Rule – Adaptations of Specialists

No defense is absolute. Whereas chemical defense works against the majority of potential enemies, usually a few specialists have evolved during evolution that have become specialized in the toxin-protected ecological niche. This phenomenon can be clearly seen in many insects, which are often highly host-plant-specific. Some of these insects take up the dietary alkaloids, store, and exploit them for their own chemical defense or that of their offspring.

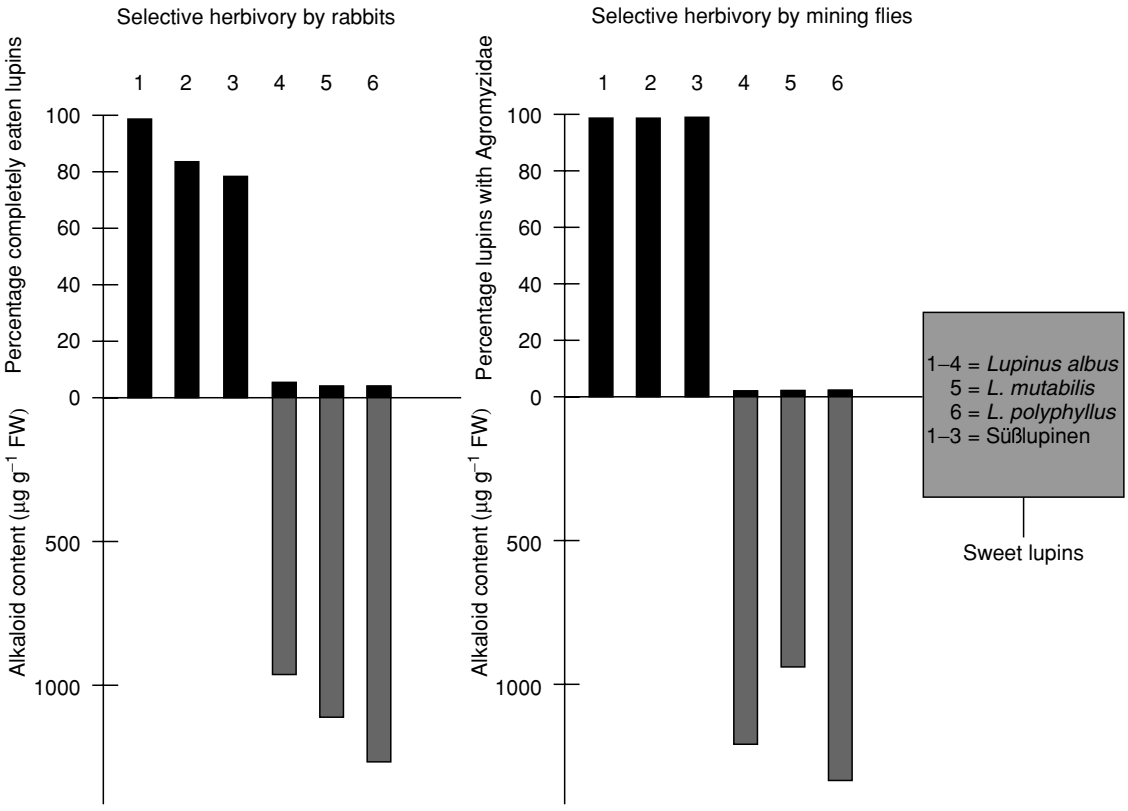

Figure 2 Selective advantage of alkaloids in lupins against herbivores (rabbits, mining flies). Lupins without alkaloids suffer heavily from herbivory, whereas alkaloid-rich lupins are widely protected.

Others additionally transform the alkaloids into pheromones or utilize them as morphogens. Well-studied examples have been published for pyrrolizidine and QAs. Vertebrate herbivores (humans included) have effective liver enzymes that can detoxify xenobiotics. Often, substances become hydroxylated, conjugated, and then excreted via the feces or the kidney and urine.

See also: **Alkaloids**: Toxicology; **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; **Immunoassays**: Radioimmunoassay and Enzyme Immunoassay

Further Reading

Bell EA and Charlwood BV (1980) Secondary plant products. In: *Encyclopedia of Plant Physiology*, vol. 8. Berlin: Springer.

Blum MS (1981) *Chemical Defenses of Arthropods*. New York: Academic Press.

Conn EE (1981) Secondary plant products. In: Stumpf PK and Conn EE (eds) *The Biochemistry of Plants*, vol. 7. New York: Academic Press.

Harborne JB (1988) *Introduction to Ecological Biochemistry*, 3rd edn. London: Academic Press.

Hegnauer R (1962–1990) *Chemotaxonomie der Pflanzen*, vols 1–10. Basel: Birkhäuser.

Luckner M (1990) *Secondary Metabolism in Microorganisms, Plants, and Animals*, 3rd edn. Berlin: Springer.

Mothes K, Schütte HR and Luckner M (1985) *Biochemistry of Alkaloids*. Weinheim: Verlag Chemie.

Roberts MF and Wink M (1998) *Alkaloids: Biochemistry, Ecology and Medicinal Applications*. New York: Plenum.

Robinson T (1981) *The Biochemistry of Alkaloids*, 2nd edn. Heidelberg: Springer.

Rosenthal GA and Berenbaum MR (1991) The chemical participants. In: *Herbivores – Their Interactions with Secondary Plant Metabolites*, vols 1 and 2. London: Academic Press.

Rosenthal GA and Janzen DH (1979) *Herbivores: Their Interactions with Secondary Plant Metabolites*. London: Academic Press.

Rosenthal GA (1982) *Plant Nonprotein Amino Acids and Imino Acids*. London: Academic Press.

Schultes RE and Hofmann A (1980) *The Botany and Chemistry of Hallucinogens*. Springfield, IL: Charles Thomas.

Southon IW and Buckingham J (1989) *Dictionary of Alkaloids*. London: Chapman & Hall.

Wink M (1999) Biochemistry of plant secondary metabolism. In: *Annual Plant Reviews*, vol. 2. Sheffield, UK: Sheffield Academic Press and CRC Press.

Wink M (1999) Function of plant secondary metabolites and their exploitation in biotechnology. In: *Annual Plant Reviews*, vol. 3, Sheffield, UK: Sheffield Academic Press and CRC Press.

Toxicology

M Wink, University of Heidelberg, Heidelberg, Germany

Background

Apparently, most alkaloids play an important role in the ecology of plants or animals. They serve as defense chemicals against herbivores and predators. To a lesser degree they protect against bacteria, fungi, and viruses or provide a means for plant–plant interactions. To be effective defense chemicals, alkaloids must closely interact with specific targets in herbivores, predators, microorganisms, or competing plants, i.e., they must either inhibit or otherwise deregulate important processes that are vital for these organisms. For this purpose the molecular shape of alkaloids has apparently been optimized during a million years of evolution in a process which could be termed 'evolutionary molecular modeling.'

While the structures of more than 12 000 individual alkaloids have been reported, rather limited knowledge is available for most of them in terms of biological activities and functions. In this chapter the modes of action of the better known alkaloids, especially those found in food plants, are summarized and discussed, considering interactions with organs or complete organisms first and then molecular targets. These interactions are the base for understanding the toxic or antinutritional effects that are observed if humans or animals have ingested alkaloids with their diet.

Toxic and Pharmacological Effects at the Organ Level

Many alkaloids are known for their toxic or adverse effects on animals (**Table 1**). In many cases, only the toxicity of an alkaloid has been reported evidencing substantial interactions, but the exact mode of action has not yet been elucidated or is rather complex, involving several molecular targets and organs.

In medicine, alkaloids are employed as local anesthetics, as narcotics, analgesics, as cardiac, uterine and respiratory stimulants, or to raise blood pressure, dilate pupils, and to relax skeletal muscles (**Table 2**). The use of alkaloids as narcotics and hallucinogens causes major social problems.

Ultimately, the toxic and pharmacological effects (**Tables 1** and **2**) observed must be the result of interactions of alkaloids with molecular targets present in or on cells.

Table 1 LD_{50} values of some alkaloids

Alkaloid	*Test system*	*LD_{50} (mg kg^{-1})*
Alkaloids derived from tryptophan		
Brucine	Rat	p.o. 1
Cinchonidine	Rat	i.p. 206
Cinchonine	Rat	i.p. 152
Ellipticine	Mouse	i.v. 1.2
Ergocryptine[a]	Rabbit	i.v. 1.1
Ergometrine[a]	Mouse	i.v. 0.15
Ergotamine[a]	Mouse	i.v. 62
Harman	Mouse	i.p. 50
Harmine	Mouse	i.v. 38
Physostigmine	Mouse	p.o. 4.5
Psilocybin	Mouse	i.v. 285
Quinidine	Rat	i.v. 30; p.o. 263
Quinine[a]	Agelaius	p.o. 100
Reserpine	Agelaius	p.o. 100
Strychnine	Agelaius	p.o. 6
	Rat	i.v. 0.9
Vinblastine	Mouse	i.v. 9.5
Vincamine	Mouse	i.v. 75
Vincristine	Mouse	i.p. 5.2
Alkaloids derived from phenylalanine/tyrosine		
Aristolochic acid	Mouse	i.v. 38–70; p.o. 56–106
Berberine	Mouse	i.p. 23
Bulbocapnine	Mouse	p.o. 413
Canadine	Mouse	p.o. 940
Chelerythrine	Mouse	s.c. 95
Chelidonine	Mouse	i.v. 35
Codeine	Mouse	s.c. 300
Colchicine	Mouse	i.v. 4.1
	Humans	p.o. 0.1–0.3
Emetine	Mouse	s.c. 32
Galanthamine	Mouse	i.v. 8; p.o. 18.7
Morphine	Mouse	i.v. 226–318
Papaverine	Mouse	i.v. 27.5; s.c. 150
Protopine	Mouse	i.p. 36–102
Sanguinarine	Mouse	s.c. 102; i.v. 16
Thebaine	Mouse	i.p. 20
Tubocurarine	Mouse	p.o. 33.2
Steroid alkaloids		
Jervine	Mouse	i.v. 9.3
Protoveratrine	Rabbit	i.p. <0.1
Samandarine	Mouse	i.p. <3.4
Solanine[a]	Mouse	i.p. 42
Veratridine	Mouse	i.p. 1.4
Tropane alkaloids		
Atropine	Rat	p.o. 750
Cocaine	Rat	i.v. 17.5
Pyrrolizidine alkaloids		
Echimidine[a]	Rat	i.p. 200
Heliotrine	Rat	i.p. 300
Jacobine	Rat	i.p. 138
Monocrotaline	Rat	i.p. 175, p.o. 71
Senecionine	Rat	i.p. 85
Seneciphylline	Rat	i.p. 77
Quinolizidine alkaloids		
Cytisine	Mouse	i.v. 1.7
13-Hydroxylupanine[a]	Mouse	i.p. 172
Lupanine[a]	Mouse	i.p. 80
N-Methylcytisine	Mouse	i.v. 21; i.p. 51
Sparteine[a]	Mouse	i.p. 55–67; p.o. 350–510
Miscellaneous alkaloids		
Aconitine	Mouse	i.v. 0.17; p.o. 1
α-Amanitin	Mouse	i.p. 0.1
Arecoline[a]	Mouse	s.c. 100
Caffeine[a]	Mouse	p.o. 127–137
Coniine	Agelaius	p.o. 56
Delphinine	Rabbit	i.p. 1.5–3.0
Maytansine	Rat	s.c. 0.48
Muscimol	Rat	p.o. 45
Nicotine[a]	Agelaius	p.o. 17.8
	Mouse	i.v. 0.3; p.o. 230
Tetrodotoxin[a]	Mouse	i.p. 0.01; s.c. 0.008

[a]Encountered in food plants or food items.
i.p., intraperitoneal; i.v., intravenous; p.o., oral; s.c., subcutaneous.

Central Nervous System and Neuromuscular Junctions

A remarkable number of alkaloids interfere with the metabolism and activity of neurotransmitters in the brain and nerve cells. A disturbance of metabolism or binding of neurotransmitters and related signal pathways impairs learning and memory, sensory faculties (smell, vision, or hearing), and coordination of bodily functions, or produces euphoric or hallucinogenic effects.

Muscle activity (skeletal, heart, etc.) is controlled by acetylcholine (Ach) and norepinephrine (noradrenaline). Any inhibition or overstimulation of neurotransmitter-regulated ion channels will severely influence muscular activity and thus the mobility or organ function (such as heart, lungs, gut). When there is inhibition, muscles will relax; when there is overstimulation, they will be tense or in tetanus, leading to a general paralysis and/or respiratory failure (which is the effect of many of the more toxic alkaloids). Alkaloids which activate (so-called parasympathomimetics) or inhibit (parasympatholytics) neuromuscular action are tabulated in Table 3. These compounds are usually considered to be strong poisons (Table 1).

Inhibition of the Digestive Process

Food uptake can be reduced by pungent or bitter taste in the first instance. The next step can be the induction of vomiting, which is a common reaction to the ingestion of a number of alkaloids; the alkaloid emetine already implies this activity in its name! Causing diarrhea, or the opposite, constipation, would be another activity which negatively influences the digestive system. Many intoxications with alkaloid-containing plants have diarrhea as one of the symptoms. Another way to interfere would be the inhibition of digestive enzymes or of transport proteins for amino acids, sugars, or lipids.

Modulation of Liver and Kidney Function

Nutrients and xenobiotics (such as secondary metabolites) are transported to the liver after resorption in

Table 2 Pharmacological and medicinal properties of alkaloids

Type	*Alkaloid*	*Activity*
Poison	Pyrrolizidine alkaloids (from *Senecio*, *Heliotropium*, *Crotalaria*)	Conversion to DNA and protein alkylating agent in liver Causing liver cirrhosis, mutations, cancer
	Ergot alkaloids (from *Claviceps purpureus*)	Cause vasoconstrictions, hallucinogenic effects, gangrenic limps; disease named ergotism
	Aconitine (*Aconitum*)	Local anesthetic, general paralytic effect
Analgesics	Morphine (*Papaver somniferum*)	Very effective painkiller (used since ancient times); addictive properties
	Codeine (*Papaver*)	Pain and cough depression
	Cocaine (*Erythroxylon coca*)	Local anesthetic
Cardiac stimulants	Quinidine (*Cinchona* spp.)	Antiarrhythmic properties at heart auricle
	Sparteine (*Cytisus scoparius*)	Antiarrhythmic properties
	Ajmaline (*Rauwolfia serpentina*)	Antiarrhythmic properties at ventricle
Respiratory stimulant	Nicotine (*Nicotina*), cytisine (*Laburnum*)	Stimulation of respiration is followed by respiratory depression, asphyxia or even respiratory failure
	Lobeline (*Lobelia* spp.)	Stimulant; used in bronchial asthma
	Coniine (*Conium maculatum*)	Used as a potent poison in antiquity (Socrates)
Constriction of blood vessels	Ergot alkaloids (*Claviceps purpurea*)	Used in obstetrics
	Ephedrine (*Ephedra* spp.)	Employed in the treatment of bronchial asthma, cold, sinusitis
	Scopolamine (*Hyoscyamus, Atropa, Datura*)	Dilatator of vessels
Muscle relaxant	Tubocurarine (*Chondodendron tomentosum*)	Block nAChR*; used in surgery
	Hyoscyamine (atropine, *Hyoscyamus, Atropa, Datura*)	Antispasmodic at smooth muscles (gastrointestine, bladder)
	Papaverine (*Papaver somniferum*)	Smooth-muscle relaxant
Antiparasitic and antimicrobial activity	Berberine (*Berberis, Mahonia*)	Intercalates DNA and inhibits parasites and microorganisms
	Emetine (*Cephaelis acuminata*)	Intestinal amoebiasis, emetic drug
	Boldine (*Peumus boldo*)	Anthelmintic activity
	Quinine (*Cinchona succirubra*)	Antimalarial
Antiinflammatory activity	Colchicine (*Colchicum autumnale*)	Treatment of acute gout, recurrent gout
Eye treatments	Physostigmine (*Physostigma venenosum*)	Reduces intraocular pressure (glaucoma)
	Pilocarpine (*Pilocarpus jaborandi*)	Miotic used in the treatment of open-angle glaucoma
Cytostatic treatment	Taxol (*Taxus brevifolia*)	Treatment of breast and ovary carcinoma; other malignancies
	Vinblastine, vincristine (*Catharanthus roseus*)	Treatment of lymphomas and other tumors

*nAChR, nicotinic acetylcholine receptor.

the intestine. In the liver the metabolism of carbohydrates, amino acids, and lipids and the subsequent synthesis of proteins and glycogen takes place. The liver is also the main site for the detoxification of xenobiotics. Lipophilic compounds, which are easily resorbed from the diet, are often hydroxylated and then conjugated with a polar, hydrophilic molecule, such as glucuronic acid, sulfate, or an amino acid. These conjugates are exported via the blood to the kidney for elimination via the urine. Both organ systems are affected by a variety of secondary metabolites: pyrrolizidine alkaloids are activated during the detoxification process and are converted into potent carcinogens, causing liver cancer. Many other metabolic inhibitors, discussed below, are also liver toxins.

Many alkaloids are known for their diuretic activity. Increased diuresis would also mean an increased elimination of water and essential ions. Since Na^+ ions are already limited in plant food, long-term exposure to diuresis-inducing compounds would reduce the fitness of a herbivore substantially.

Disturbance of Reproduction

Quite a number of allelochemicals are known to influence the reproductive system of animals, which will ultimately reduce their numbers (and fitness as a species). Antihormonal effects could be achieved by mimicking the structure of sexual hormones, such as coumarins which dimerize to dicoumarols, or isoflavones. The next target is the gestation process itself.

Table 3 Examples of alkaloids which bind to neurotransmitter receptors and neurotransmitter-degrading enzymes

Target	*Ligand*	*Alkaloid*	*Occurrence*
Acetylcholine receptors			
Nicotinic receptor	Acetylcholine	Nicotine	*Nicotiana, Duboisia*
		C-toxiferine	*Strychnos*
		Tubocurarine	*Chondodendron*
		Coniine	*Conium*
		Cytisine and other QA	Several legumes
		Lobeline	*Lobelia*
		Anabasine	*Anabasis, Nicotiana*
Muscarinic receptor	Acetylcholine	Hyoscyamine (atropine)	*Atropa, Hyoscyamus, Datura, Mandragora*
		Scopolamine	Several Solanaceae
		Arecoline	*Areca*
		Pilocarpine	*Pilocarpus*
		Muscarine	*Amanita, Inocybe, Clitocybe*, other fungi
		Sparteine and other QA	Several legumes
Adrenergic receptors	Norepinephrine(noradrenaline)/ (adrenaline) epinephrine	Ergot alkaloids	*Claviceps*
		Yohimbine	*Pausinystalia, Aspidosperma*
		Rauwolscine	*Rauwolfia*
		Corynanthine	*Rauwolfia*
		Norlaudanosoline	Papaveraceae
		Ephedrine, norephedrine	*Ephedra*
Serotonin receptor	Serotonin	Ergot alkaloids	*Claviceps*
		Psilocin, psilocybine	*Psilocybe*, other fungi
		N,N-dimethyltryptamine	Several plants and toads
		Bufotenine	*Virola, Anadenanthera*
		β-carboline alkaloids	*Banisteriopsis, Peganum*
		Mescaline	*Lophophora*, other cacti
Dopamine receptor	Dopamine	Ergot alkaloids	*Claviceps*
		Bulbocapnine	*Corydalis*
GABA receptor	GABA	Bicuculline	*Dicentra cucullaria* and other *Corydalis* species
		Muscimol	*Amanita*
		β-carboline alkaloids	*Peganum, Banisteriopsis*
Adenosine receptor	Adenosine	Caffeine	*Coffea, Camellia, Ilex, Paullinia*
		Theophylline, theobromine	*Theobroma*
Glycine receptor	Glycine	Brucine	*Strychnos*
		Strychnine	*Strychnos*
Opioid receptor	Endorphins	Morphine	*Papaver somniferum*
Acetylcholine esterase	Acetylcholine	Physostigmine (eserine)	Physostigma venenosum
		Berberine	Several Papaveraceae
		Coptisine	Several Papaveraceae
		Galanthamine	Several Amaryllidaceae
		Solanine and other steroid alkaloids	*Solanum*
		Huperzine A	*Huperzia serrata*
Monoamine oxidase (MAO)	Norepinephrine, dopamine, serotonin, histamine	Harmaline, harmine	*Peganum*
		Salsolinol	Chenopodiaceae
		Ephedrine	*Ephedra*
Catechol-*O*-methyltransferase	Norepinephrine, epinephrine, dopamine	Tetrahydroisoquinoline	Papaveraceae

QA, quinolizidine analogs; GABA, γ-aminobutyric acid.

As outlined below, a number of alkaloids are mutagenic and lead to malformation of the offspring or directly to the death of the embryo. The last step would be premature abortion of the embryo. This dramatic activity has been reported for a number of allelochemicals, including many mono- and sesquiterpenes and alkaloids. Some alkaloids achieve this by the induction of uterine contraction, as do the ergot and lupin alkaloids.

Molecular Targets of Alkaloids

In the following a number of important cellular molecular targets (Figure 1) have been addressed

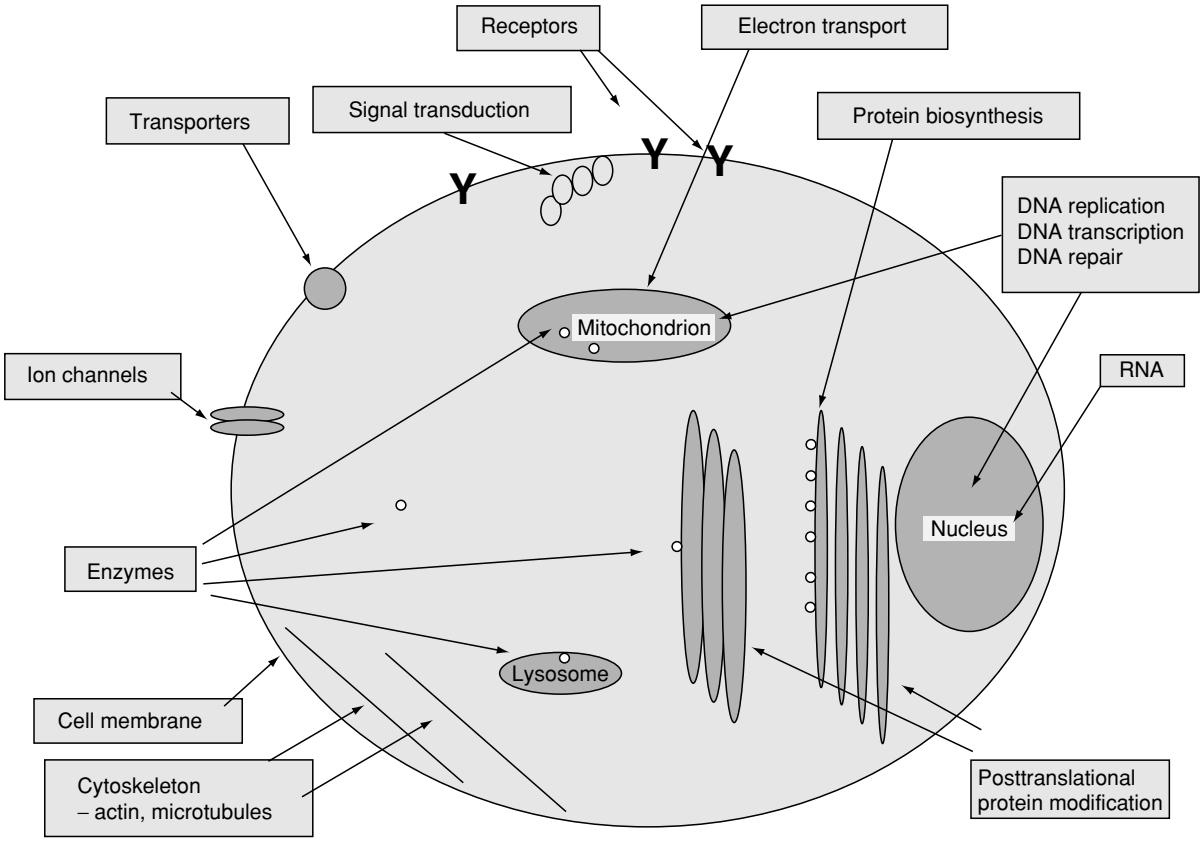

Figure 1 Molecular targets of animal cells that are affected by alkaloids.

which are often affected by alkaloids and other plant toxins.

Biomembranes, membrane transport, and neuronal signal transduction

Cells can only operate effectively if their biomembranes (cytoplasmic membrane, internal membranes) are intact. Biomembranes are almost impermeable for ions and polar molecules. As an exchange of these molecules must take place between cells and organs, specific membrane proteins, which can be ion channels, pores, or carrier proteins, mediate the controlled flux of these compounds across biomembranes. The biomembranes and the complex transport systems are targets of many natural products.

Steroidal alkaloids, such as solanine and tomatine, which are present in many members of the Solanaceae (including potatoes and tomatoes), can form complexes with the cholesterol present in biomembranes. While the steroidal moiety 'dives' into the lipophilic interior of the membrane and interacts with the structurally similar cholesterol, the hydrophilic side chain remains outside and binds to external sugar receptors. Since phospholipids are in a continuous motion, a tension easily builds up which leads to membrane disruption; transient 'holes' occur in the biomembrane, rendering the cell leaky. A similar mechanism has been postulated for saponins, a widely distributed group of natural products, to which the steroidal alkaloids may be assigned. Steroidal alkaloids can also interact with other targets, such as neuroreceptors or even with DNA; malformations have been observed in animal embryos after having been exposed to *Solanum* alkaloids.

Communication between cells is especially important for nerve cells. Signal transduction in the central nervous system and in neuromuscular junctions is mediated by receptor proteins residing in the membrane which are directly or indirectly coupled with ion channels. The neurotransmitters involved include, among others, norepinephrine (noradrenaline), epinephrine (adrenaline), serotonin, dopamine, histamine, glycine, γ-aminobutyric acid (GABA), glutamate, and acetylcholine (ACh).

Neuroreceptors can be ligand-gated channels, i.e., a receptor which is part of an ion-channel complex. When the neurotransmitter binds, a conformational change induces the opening of a Na^+/K^+ channel for

microseconds, allowing Na^+ ions (the external concentration is about 145 mmol l^{-1}) to enter the cell following a concentration gradient (the internal Na^+ concentration is between 5 and 15 mmol l^{-1}). The ligand quickly dissociates from the receptor and, in the case of ACh, is hydrolyzed by ACh esterase (Figure 2). Glutamate (*N*-methyl-D-aspartate, NMDA) and GABA receptors are also ligand-gated ion channels.

More abundant are G-protein-coupled neuroreceptors. A prominent one is the muscarinic ACh receptor; norepinephrine, serotonin, and dopamine receptors also belong to this type. When ACh binds, the receptor changes its conformation, inducing a conformational change in an adjacent G-protein molecule. Its α-subunit dissociates and then activates the enzyme adenylyl cyclase, which in turn produces cyclic adenosine monophosphate (cAMP) from adenosine triphosphate (ATP). The cAMP molecule, a second messenger, activates protein kinases or Ca^{2+} channels directly.

Quite a number of alkaloids are known whose structures are more or less similar to those of endogenous neurotransmitters. They can function therefore as structural analogs. In addition, several plants produce compounds which are identical to animal neurotransmitters, such as ACh and histamine in stinging hairs of *Urtica*, or serotonin and dopamine in several species. Targets can be:

- the receptor itself through inhibition or overstimulation (Table 3)
- the enzymes which deactivate neurotransmitters after they have bound to a receptor (Table 3)
- transport processes, which are important for the uptake of neurotransmitters into the presynapse or their storage in synaptic vesicles (Table 4) or
- enzymes involved in the biosynthesis of a neurotransmitter.

The stimulation of neurotransmitter-activated ion channels leads to a rapid influx of Na^+ ions, which in turn activates voltage-gated Na^+ and K^+ channels, which are essential for further signal transduction. These Na^+ and K^+ channels constitute another important target for alkaloids (Table 5).

Table 4 Alkaloids as inhibitors of neurotransmitter uptake (transport into presynapse or into vesicles)

Transporter	*Alkaloid*	*Occurrence*
Norepinephrine	Reserpine	*Rauwolfia*
(noradrenaline)	Ephedrine	*Ephedra*
Biogenic amines	Tetrahydro-β-carboline	*Peganum*
	Salsolinol	*Salsola*
	Tetrahydroisoquinoline	Papaveraceae
	Tetrahydropalmatine	Berberidaceae
Dopamine	Cocaine	*Erythroxylum*

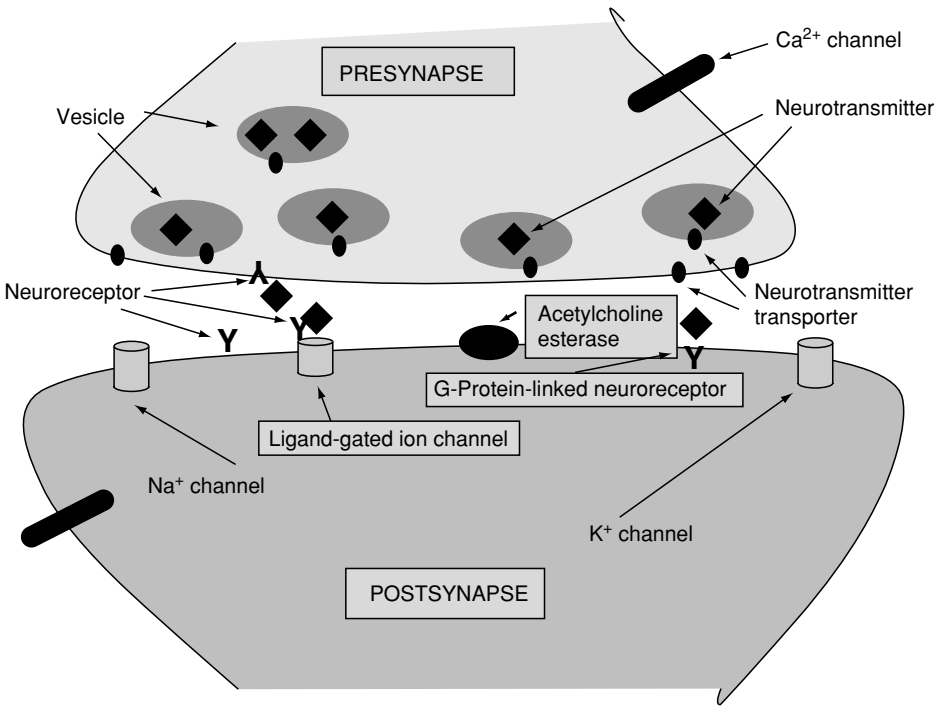

Figure 2 Signal transduction in excitable synapses.

Cells carefully control ion concentrations inside and outside of the cells with the help of specific ion channels (e.g., Na^+, K^+, Ca^{2+}, and Cl^- channels) and of active Na^+, K^+ or Ca^{2+} pumps, such as Na^+, K^+-ATPase and Ca^{2+}-ATPase. Ion gradients and ion fluxes mediated by these channels and pumps are the main elements in the active transport processes, in neuronal and neuromuscular signaling. Cardiac glycosides are potent and well-known inhibitors of Na^+, K^+-ATPase found in plants, some insects, and in the skin of certain toads. A few alkaloids such as harmaline, nitidine, sanguinarine, capsaicine, cassaine, and solenopsine (from ants) inhibit Na^+, K^+-ATPase.

Whereas receptor/ion channel interactions represent the initial part of many signal pathways, key enzymes which produce or inactivate second messengers or amplify the signal can be important targets further down the pathway (Table 6). These enzymes include:

- adenylyl cyclase (making cAMP),
- phosphodiesterase (inactivating cAMP),
- phospholipase (releasing arachidonic acid or inositol phosphates) or
- several protein kinases, such as protein kinase C (which is activated by phorbol esters and the alkaloid chelerythrine) or tyrosine kinase (activating other regulatory proteins or ion channels)

Because these targets are almost exclusively found in animals but absent in plants, the development of active compounds directed to these targets appears to be advantageous for the plants producing them. They can store these compounds without risk of being intoxicated by their own toxins.

Table 5 Alkaloids as modulators of Na^+, K^+, and Ca^{2+} channels

Alkaloid	*Occurence (genera)*	*Action*
Na^+ and K^+ channels		
Aconitine[a]	*Aconitum*	Activation
Ajmaline[a]	*Rauwolfia*	Inhibition
Batrachotoxin[a]	Frogs (Dendrobatidae)	Activation
Harmalin	*Peganum*	Inhibition
Protoveratrine A, B[a]	*Veratrum*	Activation
Quinidine[a]	*Cinchona*	Inhibition
Quinine	*Cinchona*	Inhibition
Saxitoxin[a]	*Protogonyaulax* (algae)	Inhibition
Sparteine[a]	*Cytisus, Lupinus, Genista*	Inhibition
Tetrodotoxin[a]	Algae/fish	Inhibition
Veratridine[a]	*Veratrum*	Activation
Ca^{2+} channels		
Ryanodine	*Ryania speciosa*	Inhibition

[a]Na^+ channel.

DNA/RNA

The genetic information of most organisms is mainly encoded in DNA. Since the integrity of DNA is important for the structure and function of rRNAs, proteins and enzymes which are important for metabolism, structure and development of an organism, DNA is a highly vulnerable target. It is not surprising that a number of secondary metabolites became selected during evolution which interact with DNA or DNA-processing enzymes. Some alkaloids are known to bind or to intercalate with DNA (Table 7). Many of these molecules are planar, hydrophobic molecules which fit between the planar stacks of AT and GC base pairs. Other alkaloids act on the level of DNA- and RNA-polymerases and DNA topoisomerases, thus impairing the process of replication and transcription.

The effects of DNA-binding or intercalating compounds can be mutations, which may result in malformations of newborn animals or in the initiation of cancer. When anabasine, coniine, or anagyrine is administered to pregnant cows or sheep, a large proportion of the offspring develop malformations of the

Table 6 Alkaloids modulating enzymes involved in signal transduction

Enzyme	*Function*	*Alkaloid*	*Occurrence*
Adenylyl cyclase	cAMP formation	Annonaine	Annonaceae
		β-carboline-1-propionic acid	Leguminosae
		Isoboldine	Peumus
		Tetrahydroberberine	Berberidaceae
Phosphodiesterase	cAMP inactivation	Papaverine	Papaver
		Caffeine, theobromine	*Coffea, Camellia, Theobroma*
		Theophylline	*Ilex paraguarensis, Paulinia*
		1-ethyl-β-carboline	*Peganum*
Protein kinases	Protein phosphorylation	Chelerythrine	*Chelidonium majus*
		Lyngbyatoxin A	Marine seaweeds

cAMP, cyclic adenosine monophosphate.

Table 7 Alkaloids interacting with DNA/RNA and related enzymes

Target	*Activity*	*Alkaloid*	*Occurrence*
DNA	Photoaddition	Dictamnine	*Dictamnus*
		Harman	*Peganum*
		Harmine	*Peganum*
	Alkylation	Pyrrolizidine alkaloids	Several Asteraceae, Boraginaceae
		Aristolochic acid	*Aristolochia*
		Cycasin	Cycads
	Intercalation	Ellipticine	*Ochrosia*
		Quinine, quinidine	*Cinchona*
		Skimmianine	*Skimmia*
		Berberine	*Berberis, Mahonia, Thalictrum, Chelidonium*
		Coptisine	Several Papaveraceae
		Fagaronine	Rutaceae
		Sanguinarine	Several Papaveraceae
		Olivacine	*Aspidosperma*
		Ergotamine	*Claviceps purpurea*
		Harmaline, harmin	*Peganum harmala*
		Emetine	*Cephaelis acuminata*
DNA polymerase	Inhibition	Fagaronine	Rutaceae
		Hippeastrine	*Hippeastrum*
		Lycorine	Several Amaryllidaceae
DNA topoisomerase I	Inhibition	Camptothecin	*Camptotheca acuminata*
Reverse transcriptase	Inhibition	Berberine	Several Berberidaceae, Papaveraceae
		Chelidonine	*Chelidonium*
RNA polymerase	Inhibition	Vincristine, vinblastine	*Catharanthus roseus*
Transcription	Inhibition	Colchicine	*Colchicum, Gloriosa*
		amanitin	*Amanita*

legs – so-called 'crooked calf disease.' Some alkaloids of the monocot *Veratrum*, such as jervine and cyclopamine cause the formation of a large central eye, the cyclopean eye, which was probably known to the ancient Greeks and thus led to the mythical figure of the cyclops.

Other alkaloids are known as carcinogens, such as aristolochic acid from *Aristolochia* and pyrrolizidine alkaloids (PA) which are produced by approximately 3% of the higher plants, especially within the families of Asteraceae and Boraginaceae. Aristolochic acid has a nitro group which can be transformed into reactive intermediates in the intestine. If resorbed, these metabolites can alkylate DNA. Pyrrolizidine alkaloids are not carcinogenic in their native form, but become so when they are 'detoxified' in the liver: PA are usually present in the plant as their *N*-oxides, which are polar compounds that cannot pass biomembranes by simple diffusion. In the intestine, PA-*N*-oxides are reduced by gut bacteria. The free base is then readily taken up by the gut cells and transported to the liver. There, the PA are transformed into alkylating compounds, which covalently bind to DNA and proteins. As a result mutations and cancer can be initiated.

Protein Biosynthesis

Protein biosynthesis is essential for all cells and thus provides another important target. Indeed, a number of alkaloids have been detected which inhibit protein biosynthesis *in vitro*. Emetine from *Cephaelis ipecacuanha* (Rubiaceae) is the most potent plant constituent. Other alkaloids with the same ability include harringtonine, homoharringtonine, cryptopleurine, tubulosine, hemanthamine, lycorine, narciclasine, pretazettine, pseudolycorine, tylocrepine, and tylopherine. Several alkaloids which inhibit protein biosynthesis and are also DNA intercalating substances can induce apoptosis in cells.

Electron Chains and Other Enzyme Activities

The respiratory chain and ATP synthesis in mitochondria or photophosphorylation in chloroplasts demand the controlled flux of electrons. These targets seem to be attacked by nicotine, sanguinarine, ellipticine, gramine, alpinigenine, capsaicine, and a few other alkaloids. A multitude of enzymes exist in animal cells and several alkaloids have been reported that interfere with at least one of them.

A recently discovered group of alkaloids are the polyhydroxyalkaloids, such as swainsonine or castanospermine, which inhibit hydrolytic enzymes, such as glucosidase, galactosidase, trehalase (trehalose is a sugar found in some beetle cocoons and fungi which is hydrolyzed by trehalase) and mannosidase selectively.

Cytoskeleton

Microtubules, which are important for cellular movements, vesicle transport in neurons, or the separation of chromosomes during cell division, are composed of tubulin subunits. Movements and some transport processes are mediated through either the rapid assembly or disassembly of microtubules. The assembly of microtubules is inhibited by colchicine, and dimeric indole alkaloids vinblastine and vincristine (important for chemotherapy of certain cancers). These alkaloids thus interrupt cell division. The diterpene alkaloid taxol (used in the treatment of ovarian and breast cancer) affects microtubules in the opposite way; the polymerization of tubulin is enhanced by taxol. As a consequence taxol-induced microtubules are very stable and dividing cells are arrested in the metaphase.

Cell stability, phagocytosis, cell–cell interactions, and cell movements are also controlled by actin filaments, which are rapidly assembled or disassembled from action monomers. Cytochalasin B and latrunculin B bind to the plus end of a growing actin filament, preventing the addition of actin monomers there. Another alkaloid, phalloidin, produced by the fatally poisonous toadstool *Amanita phalloides*, stabilizes actin filaments and inhibits their depolymerization.

Mechanisms of Allelochemical Activities in Antiviral, Antimicrobial, and Phytotoxic Interactions

Circumstantial evidence indicates that some alkaloids protect the producing plant against viruses, bacteria, fungi, and competing plants. A number of antimicrobial alkaloids such as sanguinarine, quinine, or berberine intercalate with viral and microbial DNA or bind to it. These compounds may thus inhibit processes such as DNA replication and RNA transcription which are vital for the microorganisms. Protein biosynthesis in ribosomes is another vulnerable target, attacked by emetine. The stability of biomembranes can be disturbed by steroidal alkaloids and tetrandine. Other targets may be electron chains or just metabolically important enzymes. Phytotoxic properties or germination inhibition, which can be observed in plant–plant interactions, can also proceed via the above-mentioned mechanisms. But interactions with growth hormones and their metabolism must also be considered.

Target specificity of alkaloids

In general, the interactions of a particular alkaloid with a molecular target (as described above) suggest a high degree of specificity. A closer look, however, shows that many alkaloids interfere with more than one target. The phenomenon will be explained for two groups of alkaloids: ergot alkaloids and quinolizide alkaloids (QA).

Ergot Alkaloids

Ergot alkaloids, such as ergotamine, ergometrine, or ergoclavine, are produced by fungi of the genus *Claviceps* which lives in close contact with many grasses (family Poaceae) such as the cereal *Hordeum vulgare*. These alkaloids can modulate several receptors of neurotransmitters, such as dopamine, serotonin, and norepinephrine. As a consequence. the pharmacological action of ergot alkaloids is rather broad, ranging from vasoconstriction and uterus contraction to hallucinations. We can explain these activities through structure similarities between the alkaloid and the different neurotransmitters.

Quinolizidine Alkaloids

QA, such as lupanine, sparteine, or cytisine, are produced by lupins and many members of the Leguminosae. They are bitter for many animals (and plants producing them are therefore avoided as food). If ingested, QA exhibit a broad level of toxicity: they interact with ACh receptors (AChR) as agonists. QA, like many other alkaloids, occur as complex mixtures in plants. Some QA preferentially bind to the nicotinic AChR, whereas others tend more to bind to the muscarinic AChR. Some QA exhibit a prominent cross-reactivity. Additionally, QA such as lupanine and sparteine inhibit Na^+ and K^+ channels, thus blocking the signal transduction in nerve cells at a second critical point. A few particular QA, such as anagyrine, cytisine, and the bipiperidine alkaloid ammodendrine (which cooccurs with QA in many plants), are mutagenic and lead to malformations (see above).

If we accept the hypothesis that alkaloids were developed as chemical defense compounds through a process of 'evolutionary molecular modeling' the 'cross-reactivity' described makes sense: any compound which can interfere with more than one target or with more than one group of adverse organisms is likely to be more effective and thus has a better survival value in general than a more selective allelochemical. In addition, herbivores will try to develop tolerance to or resistance against the dietary toxins. If more than one target is affected by a defense chemical the chances of a herbivore developing specific resistances concomitantly are much smaller than in single-target situations. In conclusion, we can say that Nature has obviously tried 'to catch as many flies with one clap as possible' in the selection of alkaloids during evolution. (*See* **Trypsin**

Inhibitors; Saponins; Antibiotics and Drugs: Uses in Food Production; **Alkaloids:** Properties and Determination.)

See also: **Alkaloids**: Properties and Determination; **Antibiotics and Drugs**: Uses in Food Production; **Cereals**: Dietary Importance; **Coffee**: Analysis of Coffee Products; **Lupin**; **Potatoes and Related Crops**: The Root Crop and its Uses; **Plant Antinutritional Factors**: Characteristics; **Saponins**; **Tea**: Chemistry; **Tomatoes**; **Trypsin Inhibitors**

Further Reading

Alberts B, Bray D, Lewis J, Raff M, Roberts K and Watson JD (1993) *Molecular Biology of the Cell*, 3rd edn. New York: Garland.

Harborne JB (1993) *Introduction to Ecological Biochemistry*, 4th edn. London: Academic Press.

Mann J (1992) *Murder, Magic and Medicine*. London: Oxford University Press.

Roberts MF and Wink M (eds) (1998) *Alkaloids: Biochemistry, Ecology and Medicinal Applications*. New York: Plenum.

Rosenthal GA and Berenbaum MR (1991) *Herbivores: Their Interactions with Secondary Plant Metabolites*, vol. 1. *The Chemical Participants*. San Diego: Academic Press.

Rosenthal GA and Berenbaum MR (1992) *Herbivores: Their Interactions with Secondary Plant Metabolites*, vol. 2. *Ecological and Evolutionary Processes*. San Diego: Academic Press.

Wink M (1998) Modes of action of alkaloids. In: Roberts MF and Wink M (eds) *Alkaloids, Biochemistry, Ecology, and Medical Applications*, pp. 301–326. New York: Plenum.

Wink M (1999) *Biochemistry of Plant Secondary Metabolism. Annual Plant Reviews*, vol. 2. Sheffield: Sheffield Academic Press and CRC Press.

Wink M (1999) *Function of Plant Secondary Metabolites and their Exploitation in Biotechnology. Annual Plant Reviews*, vol. 3. Sheffield: Sheffield Academic Press and CRC Press.

Wink M (1993) Allelochemical properties or the raison d'être of alkaloids. In: Cordell GA (ed.) *The Alkaloids*. vol. 43, pp. 1–118. San Diego: Academic Press.

Wink M (2000) Interference of alkaloids with neuroreceptors and ion channels. In: Atta-Ur-Rahman (ed.) *Bioactive Natural Products*, vol. 11, pp. 3–129. Amsterdam: Elsevier.

Wink M and Schimmer O (1999) Modes of action of defensive secondary metabolites. In: Wink M (ed.) *Function of Plant Secondary Metabolites and their Exploitation in Biotechnology. Annual Plant Reviews*, vol. 3, pp. 17–133. Sheffield: Sheffield Academic Press and CRC Press.

ALLERGENS

E N C Mills, Institute of Food Research, Norwich, UK
A S Tatham, University of Bristol, Bristol, UK

Introduction

This encyclopedia entry describes the nature of food allergens involved in IgE-mediated allergy, their nomenclature, and the properties of proteins thought to predispose them to becoming allergenic. Current knowledge of food allergens of plant and animal origin is summarized, and the role that cross-reacting IgE epitopes play in the relationship between pollen and latex allergy and allergies to certain fruits and vegetables is discussed. The impact of postharvest treatments and food processing on allergen activity is described, together with the problems posed to food allergic individuals by 'hidden' allergenic ingredients and those that result from cross-contamination in the factory and catering outlets.

What is an Allergen?

During the course of normal immune functioning, the body produces a number of different forms, or isotypes, of immunoglobulins such as IgA, IgG, IgM, and IgE, which bind to 'nonself' molecules. These include molecules found in microbial pathogens, parasites, environmental agents such as pollen and dietary proteins. However, in the allergic disease classified as a *Type I hypersensitivity reaction*, this antibody repertoire is altered, and the body synthesizes larger quantities of IgE, an antibody type normally produced only in response to parasitic infections. As yet, we do not understand the mechanisms whereby particular allergens elicit an IgE rather than the normal IgG response in certain individuals. This IgE is directed towards target molecules, which are usually proteinaceous in nature and are known as allergens. IgE can become associated with mast cells, and on binding multivalent allergen, it becomes 'cross-linked' at the mast cell membrane, triggering the

release of immune mediators, such as histamine, which actually cause the symptoms associated with allergic reactions. (*See* **Food Intolerance**: Food Allergies for more information.)

The sites on a protein that are recognized by an antibody are known as epitopes and can be either linear or conformational. In the former, only the primary sequence of a polypeptide is involved in antibody recognition, but in conformational epitopes, the three-dimensional structure of a protein is important. Thus, a number of segments of the polypeptide chain that may be quite distant in the amino acid sequence of a protein are brought together spatially as a consequence of its tertiary and quaternary structure, to form an epitope. Most epitopes are thought to be conformational in nature and are particularly difficult to define in relation to food allergens where processing can have such a disruptive effect on native protein structure (see below).

Allergen Nomenclature

The World Health Organization and the International Union of Immunological Societies produce an official list of allergens, which are designated by the Allergen Nomenclature subcommittee. Allergens included in this listing must induce IgE-mediated (atopic) allergy in humans with a prevalence of IgE reactivity above 5%. An allergen is termed major if it is recognized by IgE from at least 50% of a cohort of allergic individuals but does not carry any connotation of allergenic potency; allergens are otherwise termed 'minor.' The allergen designation is then based on the Latin name of the species from which it originates and is composed of the first three letters of the genus, followed by the first letter of the species finishing with an arabic number, e.g., *Ara h 1* relates to an allergen from *Arachis hypogea* (peanuts).

Types of Food Allergens

There appear to be around eight foods, including those of both plant and animal origin, that are responsible for causing the majority of food allergies, namely peanuts and tree nuts, wheat, soya, cows' milk, egg, shellfish, and fish. The geographical distribution of these allergies reflects patterns of food consumption with, for example, buckwheat allergy being much more common in the Far East, where they are staple foods, than is the case in Western countries. The distribution of allergies also differs across Europe, with reactions to fish and shellfish being higher in Mediterranean countries, where fish consumption is higher.

In addition to foods that sensitize individuals directly (probably via the gastrointestinal tract), there is a group of fruit and vegetable allergies that arise from the presence of cross-reactive epitopes in the allergens. Thus, individuals who become sensitized to pollen are more likely to develop IgE, which recognizes homologous proteins present in fresh fruits and vegetables. Around 70% of individuals who develop birch pollen allergy with IgE towards the pollen protein *Bet v 1*, develop allergies to apples and other fruits of the *Rosacea* because the anti-*Bet v 1* IgE reacts with fruit homologs such as the apple protein *Mal d 1*. Similarly, many of those who develop allergy to latex have IgE, which recognizes the chitin-binding domain of the hevein polypeptides that make up the rubber latex network. An almost identical domain is found in a group of enzymes, the class I chitinases, found in a number of fruits including avocado, chestnut, and banana, which the anti-latex IgE recognizes. As a consequence of this cross-reactivity, individuals with latex allergy frequently suffer from allergic reactions on consuming these fruits (**Figure 1**.)

Protein Properties and Allergenicity

Two factors that are thought to contribute to the allergenic potential of food proteins are their resistance to the effects of thermal processing used in food preparation (discussed below) and breakdown by digestive enzymes in the gastrointestinal tract. The stability of a number of allergenic proteins to pepsin, including those from peanuts, soya, and cows' milk, has been compared with that of nonallergenic proteins. It was found that all the allergens either remained undigested or gave stable fragments that persisted for 8–60 min (depending upon the allergen), whereas the nonallergens were completely digested after less than 15 s. As peptides are required to have a molecular weight of greater than 3000 Da in order to stimulate an immune response, large stable

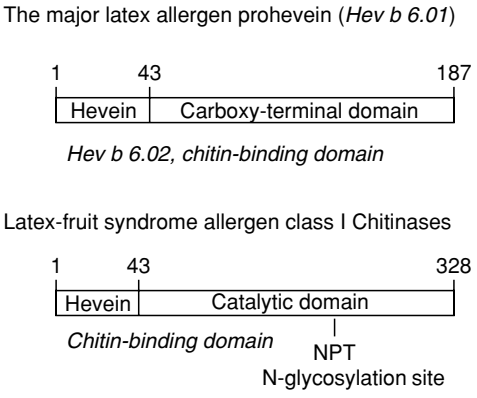

Figure 1 Presence of conserved hevein domains in latex and the class I chitinases of fruits responsible for many latex–fruit cross-reactive allergies.

fragments, as well as intact proteins, have the potential to act as sensitizers. Similarly, such large fragments may be able to cross-link mast cell IgE, leading to histamine release. Whilst the bulk of food proteins are broken down into immunologically inactive fragments, very small proportions (milligram amounts) of material, which are still immunologically important, may escape digestion. Circumstantial evidence that food proteins can resist digestion and cross the gut barrier is offered by the presence of antibodies to a wide range of food proteins in the circulation of normal individuals. The presence of proteins, such as ovalbumin in the blood of individuals after consumption of egg, also supports the premise that intact proteins, albeit in small quantities, can enter the body. However, although the nature of the transported proteins has not been clearly defined, a large proportion of cow's milk allergic patients have specific IgE, which recognizes proteolyzed β-Lg, some of which reacts better with the hydrolyzed protein than the intact form. Thus, the resistance of allergens to the effects of the digestive process plays an important role in determining the allergenic potency of food proteins.

Allergens of Plant Origin

Many plant food allergens are pathogen-related (PR) proteins, which are produced by the plant as a defense against pathogenic microorganisms, insects, and invertebrate pests. They have been classified into a number of families on the basis of sequence similarities or enzymic properties and are receiving attention from plant breeders, as a tool for GM crops, because of their protective activities against agricultural pests. However, they are cross-reactive allergens in fruits, vegetables, and pollen. Allergens, other than PR proteins, are inhibitors of proteases and α-amylases, profilins (actin-binding proteins), seed-storage proteins (albumins, globulins, and prolamins), proteases and lectins (agglutinins). These are summarized in Table 1.

Fruit and Vegetables

Fruit allergy is increasing and is often associated with birch- and grass pollinosis, as well as latex–fruit syndrome. Oral allergy syndrome is a manifestation of food allergy, where certain foods cause itching or hives when they touch the lips and mouth. Fruit and vegetables are often responsible for this reaction, owing to pollen cross-reactivity, and many fruit allergens are PR proteins.

Allergens homologous to the major birch pollen allergen *Bet v 1* (PR group 10), are found in apple *Mal d 1*, cherry *Pru av 1*, apricot *Pru ar 1*, pear *Pyr c 1*, carrot *Dau c 1*, and celery *Api g 1*, but may trigger responses to many other fruits and vegetables. By contrast, in areas free of birch, the major allergens in patients sensitized to Rosaceae fruits are lipid transfer proteins (LTP, PR group 14), 9-kDa proteins that often have antifungal and antibacterial activities. LTP allergens characterized include apple *Mal d 3* and peach *Pru p 3*, but homologous proteins are found in a range of plant-derived foods (vegetables, nuts, and seeds), and as a consequence, LTPs have been described as panallergens.

Class-I chitinases (PR group 3) are a group of carbohydrases active against the exoskeletons of insects and the cell walls of fungi. They are widely distributed in plants and are also termed panallergens, with characterized allergens including those from avocado (*Prs a 1*) and banana (*Mus p 1.2*). Wound-induced chitinases (Win) have been extracted from virus-infected leaves (PR group 4) and have been reported in turnip, tomato, and potato. Another group of carbohydrases, the glucanases (PR group 2), are active against fungal cell walls, have been isolated from latex, and are recognized by IgE from food-allergic patients with hypersensitivity to banana, potato, and tomato. Lastly, thaumatin-like proteins (PR group 5) are another group of allergens that have antifungal and/or antibacterial activity, allergens having been characterized in apple (*Mal d 2*) and cherry (*Pru a 2*).

NonPR protein allergens include the profilins, originally identified from birch pollen, *Bet v2*, which are now recognized as cross-reactive plant allergens in a wide range of fruits, vegetables, nuts, and seeds. Thiol proteases (proteolytic enzymes) have also been identified as allergens in kiwi *Act c 1*, papaya, fig, pineapple, and soybean. Patatin, a storage protein from potato, has also been shown to be an allergen in raw potato and cross-react with the latex allergen, *Hev v 7*.

Tree Nut and Peanuts

Nut allergy is the most common form of food allergy and the most common cause of death by anaphylaxis. The major allergens are 2S albumins and the 7S and 11/12S globulins. In peanut, three of the major allergens are major storage proteins: *Ara h 1* is the 7S globulin, conarachin, a vicilin-like glycoprotein; *Ara h 2* is a 2S albumin; and *Ara h 3*, is a subunit of the legumin-like 11S globulin, arachin. The IgE binding epitopes of *Ara h 1* and *2* have been determined and mutagenized to alter their activity. Brazil nut contains a 2S storage protein allergen, *Ber e 1*; a major allergen is a 2S albumin with a high methionine content. Related 2S albumins are found in walnut, *Jug r 1*, which also contains a 7S globulin allergen, *Jug r 2*. In addition to storage protein allergens, chestnut contains an endochitinase allergen *Cas s 1*, and

Table 1 Major food allergens of plant origin

Food	*Allergen*	*Allergen designation*	*Molecular weight*	*Function and stability*
Fruit				
Apple	Hom: *Bet v 1*	*Mal d 1*	17 700	Possible sterol binding protein (PR-10)
	Hom: thaumatin	*Mal d 2*	31 000	Antifungal activity (PR-5)
	LTP	*Mal d 3*	9 000	Antifungal and antimicrobial activity, heat-stable (PR-14)
Apricot	Hom: *Bet v 1*	*Pru ar 1*	17 700	Possible sterol binding protein (PR-10)
	LTP	*Pru ar 3*	9 000	Antifungal and antimicrobial activity, heat-stable (PR-14)
Avocado	Endochitinase	*Prs a 1*	32 000	Degrade chitin (PR-3)
Banana	Chitinase	*Mus p1.1*	32 000	Degrade chitin (PR-3)
	Chitinase	*Mus p1.2*	34 000	Degrade chitin (PR-3)
Cherry	Hom: *Bet v 1*	*Pru av 1*	17 700	Possible sterol binding protein (PR-10)
	Hom: thaumatin	*Pru av 2*	23 300	Antifungal activity (PR-5)
	Profilin	*Pru av 4*	15 000	Actin (cytoskeleton)-binding protein
Kiwi	Cysteine protease	*Act c 1*	30 000	Proteolytic enzyme
Peach	LTP	*Pru p 3*	9 000	Antifungal and antimicrobial activity, heat-stable (PR-14)
Pear	Hom: *Bet v 1*	*Pyr c 1*	18 000	Possible sterol binding protein (PR-10)
	Profilin	*Pyr c 4*	14 000	Actin (cytoskeleton) binding protein
	Isoflavone reductase	*Pyr c 5*	33 500	Enzyme involved in secondary metabolism
Plum	LTP	*Pru d 3*	9 000	Antifungal and antibacterial activity, heat-stable (PR-14)
Vegetables				
Carrot	Hom: *Bet v 1*	*Dau c 1*	16 000	Possible sterol binding protein (PR-10)
Celery	Hom: *Bet v 1*	*Api g 1*	16 000	Possible sterol binding protein (PR-10)
	Profilin	*Api g 4*	14 300	Actin (cytoskeleton)-binding protein
	Unknown	*Api g 5*	55/58 000	Unknown
Potato	Patatin	*Sola t 1*	43 000	Storage protein
Turnip	Hom: prohevein	*Bra r 2*	25 000	Defense protein
Nuts				
Brazil nut	2S albumin	*Ber e 1*	9 000	Storage protein
Cashew nut	Anacardien	Ana o	50,0000	Storage protein
Chestnut	Chitinase	*Cas s 1*	30 000	Degrade chitin (PR-3)
Hazelnut	Hom: *Bet v 1*	*Cor a 1.0401*	17 000	Possible sterol binding protein (PR-10)
Peanut	Cor Arachin	*Ara h 1*	63 500	Storage protein
	Conglutin	*Ara h 2*	17 000	Storage protein
	Arachin	*Ara h 3*	60 000	Storage protein
	Arachin	*Ara h 4*	37 000	Storage protein
	Profilin	*Ara h 5*	15 000	Actin (cytoskeleton)-binding protein
	2S albumin	*Ara h 6*	15 000	Storage protein
	2S albumin	*Ara h 7*	15 000	Storage protein
Walnut	2S albumin	*Jug r 1*	15 000	Storage protein
	Vicilin	*Jug r 2*	44 000	Storage protein
Seeds				
Barley	Inhibitor	*Hor v 1*	15 000	Glycosylated α-amylase inhibitor
	Inhibitor	CMb	16 000	Glycosylated subunit of barley tetrameric α-amylase inhibitor
Castor bean	2S albumin	*Ric c 1*	10 000	Storage protein
Lentil	Vicilin	Len c 1	45-50,0000	Storage protein
Maize	LTP	*Zea m 14*	9 000	Antifungal and antibacterial activity (PR-14)
Mustard	2S albumin	*Bra j 1*	14 000	Storage protein
	2S albumin	*Sin a 1*	14 000	Storage protein
Rice		*RAP*	14 800	Rice allergenic protein
Rye	Inhibitor	*Sec c 1*	14 000	α-Amylase/trypsin inhibitor
Sesame	2S albumin	*Ses i 1*	10 000	Storage protein
	Vicilin	Ses i 3	60,0000	Storage protein
Soybean	Cys protease (inactive)	*Gly m 1*	7 000	Function
		Gly m 2	8 000	Unknown
	Profilin	*Gly m 3*	14 000	Actin (cytoskeleton)-binding protein
Wheat	Inhibitor	*CM16**	16 000	Glycosylated subunit of wheat tetrameric α-amylase inhibitor
	Prolamin	γ-gliadin	30-50,0000	Storage protein of glutenin

peanuts a profilin, *Ara h 5*. Other nuts, such as cashew, hazelnut, pecan, and pine nuts are known to be allergenic, and contain storage globulin, 2S albumin and LTP allergens.

Seeds

Seed allergens are predominantly nonPR-type proteins, comprising storage proteins, proteases and α-amylase inhibitors. Wheat and related cereals, barley and rye, contain a range of allergens including the prolamins (alcohol soluble storage proteins), which are responsible for food-dependent exercise-induced anaphylaxis and atopic dermatitis; the same proteins are active in celiac disease and dermatitis herpetiformis. Inhibitors of proteases and α-amylases protect the seed from microorganisms and insect pests, characterized allergens including the 12–15-kDa proteins and inhibit α-amylases or trypsin and are both inhalant allergens (e.g., baker's asthma) and food allergens. Those with the highest activity are glycosylated subunits of tetrameric α-amylase inhibitors from wheat (CM16) and its homologs from barley (CMb) and BMAI-1 and rye *Sec c 1*, similar proteins being found in rice. The major allergens in rice are α-amylase and trypsin inhibitors, and in buckwheat, the major allergen is another trypsin inhibitor. Peroxidases in wheat and barley (PR group 9) of 36 kDa have been identified as allergens in baker's asthma.

The major allergens in yellow mustard *Sin a 1* and oriental mustard *Bra j 1*, castor bean, cottonseed, chickpea, oilseed rape *BnIII*, and sunflower SFA8 are 2S albumins (apart from sunflower, which consists of a single polypeptide chain), and they consist of two polypeptide chains linked via disulfide bonds. In soybean, the two major storage globulins, β-conglycinin and glycinin, are major allergens; other allergens are profilin *Gly m 3* and a papain-related thiol protease *Gly m Bd28*.

Allergens of Animal Origin

Egg

The allergens primarily originate from egg white, the major allergens being ovomucoid (*Gal d 1*) and ovalbumin (*Gal d 2*), which constitute 10 and 50% of white proteins, respectively. Both proteins are heavily glycosylated with 25% of the mass of ovomucoid comprising carbohydrate. However, in ovomucoid, the IgE epitopes are clustered in seven regions of the protein sequence and do not encompass glycosylation sites. Whilst IgE epitopes of ovomucoid appear to be conformational in newly acquired sensitivities to egg, linear epitopes are more important in long-standing egg allergy. The IgE epitopes of ovalbumin are also resistant to enzymic digestion and denaturation, and seven IgE-binding regions have been identified, which tend to cluster at the N- and C-terminal regions of the protein. Ovalbumin has also been found in an immunologically active form in the blood of human subjects following consumption of raw egg. Two minor allergens, ovotransferrin (*Gal d 3*) and lysozyme (*Gal d 4*), have also been identified in egg white. In general, cooking has been found to reduce the allergenic activity of egg.

Cows' Milk

Allergens are found in both the whey and casein fractions, the most common allergens being β-lactoglobulin and α-casein, although other IgE-reactive proteins have been identified (Table 2). β-Lactoglobulin is highly resistant to proteolysis and is taken up in an intact form by the gut in experimental systems, with the major IgE epitopes having been in four main regions located on the more mobile surface loops of the protein. The caseins also appear to contain thermostable epitopes, with IgE binding being located in around seven different regions of the protein.

Fish

The major allergen in fish is the muscle protein parvalbumin (*Gad c 1*), which belongs to a group of such proteins unique to fish and amphibians. Whilst it has only been characterized in depth for cod and salmon, parvalbumin is conserved across fish species, a factor responsible for the cross-reactive nature of allergens in cod, salmon, mackerel, herring, and plaice, amongst many other fish species. It acts as a calcium buffer protein in fast muscle, binding the Ca^{2+} ions in a calcium-binding motif known as an EF-hand. *Gad c 1* has been shown to bind IgE in five regions evenly distributed along the length of the protein, one of which encompassed one of the Ca^{2+}-binding sites. Like other calcium-binding proteins, *Gad c 1* is heat-stable, with the holo-form being both more IgE-reactive and more heat-stable than the apo form.

Shellfish and Seafood

Tropomyosin, a heat-stable muscle protein, is the major allergen in shellfish and seafood, with highly homologous proteins being found in the commonly edible crustaceans. These homologies are responsible for the cross-reactive allergies observed between various types of seafood including shrimps, lobsters, crab, squib, and abalone, and inhalant insect allergens, such as those from cockroaches. Two main linear IgE-binding sites have been identified in the shrimp allergen, *Pen i 1*, one in the N-terminus, which

Table 2 Major food allergens of animal origin

Food	*Allergen*	*Allergen designation*	*Molecular weight*	*Function and stability*
Hen's egg	Ovomucoid	*Gal d 1*	28 000	Protein source with protease inhibitor
	Ovalbumin	*Gal d 2*	42 000	Serpin family member
	Ovotransferrin	*Gal d 3*	70 000	Iron-binding protein and bacteriocide
	Lysozyme	*Gal d 4*	14 300	Carbohydrase and bacteriocide
Cows' milk	β-Lactoglobulin	*Bos d 5*	14 000 (monomer)	Dimeric protein belonging to the lipocalin superfamily that binds a number of small lipophiles such as retinol
	α-Lactalbumin	*Bos d 4*	14 200	Calcium-binding protein
	Serum albumin	*Bos d 6*	66 000	Fatty-acid-binding protein
	IgG	*Bos d 7*	150 000	Immunoglobulin
	α-Casein	*Bos d 8*	32 400	Phosphoprotein able to chelate calcium
	β-Casein	*Bos*	26 600	Phosphoprotein able to chelate calcium
Fish	Parvalbumin	*Gad c 1* (cod)	12 500	Calcium-binding muscle protein that is pH- and heat-stable
		Sal s 1 (salmon)		
Shellfish	Tropomyosin	*Pen a 1* (northern brown shrimp)	36 000	Contractile muscle protein that is heat-stable
		Pan i 1 (Indian white shrimp)	38 000	
		Met e 1 (greasy black shrimp)	34 000	
		Pan s 1 (spiny lobster)		
		Hom a 1 (American lobster)	34 000	
		Cha f 1 (crab)	34 000	
		Tod p 1 (squid)	34 000	
		Hal m1 (abalone)	38 000	
			49 000	
Meat	Serum albumin	*Bos d 6* (beef)	68 000	Fatty-acid-binding protein
		Bos d 7 (beef)	150 000	Immunoglobulin

showed no homology with vertebrate tropomyosin, and another in the C-terminal region. The first two residues of the C-terminal epitope appear to be crucial for IgE binding and are not found in vertebrate tropomyosin. As a consequence of the lack of homology in the IgE epitopes, there is no cross-reactivity between IgE from shell-fish allergic individuals and animal muscle tropomyosins. In addition to being found in cooked meat, the allergen also leaches into cooking water.

Meat

Meat allergy is rare, the major allergens having been identified as serum albumins and IgG, both of which are also minor milk allergens. However, unlike shellfish, muscle proteins such as tropomyosin do not appear to be involved. Processing has been found to reduce the allergenic activity of meat allergens, with more severe processes, such as sterilization, causing the greatest reduction.

Factors Altering the IgE reactivity of Food Allergens

Postharvest treatments and processing can potentially alter the IgE reactivity of allergens by altering the levels of allergens present in raw materials and altering IgE-epitope presentation.

Effects of Raw Material and Postharvest Handling

A number of plant protein allergens belong to multigene families, where a number of isoforms differing in only a few amino acid residues may be present. In addition, the levels of different isoforms may vary between cultivars. These changes have been demonstrated to make a difference particularly with regard to the apple allergen *Mal d 1*, where the IgE reactivity of certain cultivars is demonstrably higher than others. Furthermore, one variant, *Mal d 1a*, has a higher IgE-binding capacity than another variant *Mal d 1b*, which differs by only 15 amino acid residues. Allergen levels appear to increase in apples following harvest and during storage at 4 °C over 130 days, probably because of their association with the ripening process, and the fact that they are 'PR' proteins and hence synthesized in response to environmental stress. Storage under modified atmospheres can reduce these increases in allergen levels, as might be anticipated from its ability to delay fruit ripening.

Effects of Processing

There are several food proteins that do not adopt a compact globular three-dimensional structure, but are highly mobile, adopting an ensemble of conformations. These include the γ-gliadin from wheat and caseins from milk. The IgE epitopes on such proteins

are likely to be linear and hence less affected by thermal treatments than those on globular proteins. In contrast, globular proteins undergo a marked denaturation at around 50–60 °C, which is accompanied by a loss of tertiary and secondary structure and is usually followed by aggregation.

Cross-reactive allergens, such as those involved in the pollen–fruit and latex–fruit allergy syndromes, are generally destroyed by cooking, and hence the allergic reactions are confined to raw produce. In addition, heating has been found to reduce the allergenicity of beef and purified bovine allergens. However, many other food allergens are remarkably heat-stable, particularly those originating from plant seeds, which are involved in sensitization via the gastrointestinal tract. This is illustrated by Figure 2, which shows the differential scanning calorimetry (DSC) thermogram for glycinin, indicating that the main thermal transition for this protein is around 90 °C. There is also evidence that, whilst many of highly disulfide bonded plant proteins unfold on heating, they are able to refold on cooling to an almost native structure. Hence, many allergenic proteins may be able to retain both linear and conformational epitopes following processing, allowing an IgE-mediated reaction towards the native and processed protein. However, the unfolding and subsequent aggregation processes that occur in many food proteins offer the possibility of introducing new epitope sites. This applies to many allergens, including those from soya, egg, and shellfish meat, which form aggregates on heating that become incorporated into heat-set gel networks. Thus, allergic IgE binding to roasted nuts has been shown to be 90 times greater than that observed towards raw peanuts, indicating that processing does introduce additional IgE binding sites. The technical difficulties of working with large aggregated protein systems has meant that the characterization of such processing-induced epitopes has been difficult.

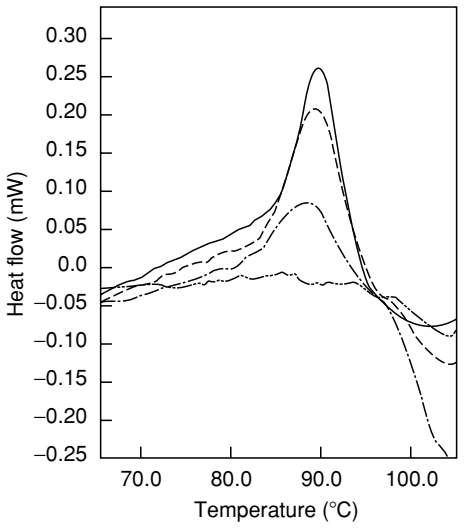

Figure 2 Thermostability of the soya allergen, glycinin, as determined by DSC. Native protein (——); after treatment at 70 °C, 10 min (–); 80 °C, 10 min (-·-·-) and 90 °C, 10 min (-···-···-).

Carbohydrate residues on glycoproteins have also been implicated as thermostable IgE epitopes, although some workers have shown that their removal does not affect IgE reactivity. Furthermore, because of their sparse nature in some allergens, they do not offer the polyvalency necessary IgE cross-linking on mast cells and subsequent histamine release. Whilst they have not been well characterized, there is now evidence that Maillard browning adducts formed following severe heat treatments contribute to the allergenic activity of roasted nuts such as peanuts and pecan nuts. It appears that Maillard modified peanut allergens *Ara h 1* and *Ara h 2* become cross-linked to form high-molecular-weight aggregates that bind IgE more effectively than unmodified allergens and are also more resistant to gastric digestion. The role that other thermally induced modification, such as lactosylation of milk proteins, may play in altering the allergenic activity of products, including severely heat-treated milks, has yet to be characterized in detail.

Problems of 'Hidden' Allergens

Those individuals who suffer from severe forms of food allergy, such as anaphylaxis, can have reactions triggered by very small amounts (as little as 100 μg) of an allergen present in a food. Indeed, traces of nuts found in processed oils and carry-over of allergens on utensils used for serving foods can be enough to cause an allergic reaction. Proteins are widely used as functional ingredients in foods, with soya flours being routinely included in bread and bakery products as an improver, whereas other soya ingredients are used in soups and sauces as well as meat products such as sausages. Similarly, caseinates and whey protein isolates are used in a range of products such as sauces, cakes, and confectionery.

Another difficulty is presented by the uncertainty relating to accidental contamination, where processing lines are shared between products utilizing different ingredients. This is particularly problematic for confectionery, where tree nuts and peanuts are a favored ingredient and has led to labeling of products with the warning 'may contain nuts.' Indeed, surveys have shown that a large proportion of chocolate confectionery is contaminated with hazelnut proteins, probably as a result of cross-contamination in

factories during manufacture. One of the consequences of this for the allergic consumer is that allergenic proteins are hidden as ingredients in manufactured products. Accidental exposure to allergens is a frequent occurrence, and epidemiological studies have shown that around 50% of peanut-allergic individuals have a reaction at least once a year as a consequence of accidental exposure.

In order to address concerns over inadvertent contamination of raw materials by allergenic nuts and seeds, methodologies have been developed for detecting allergens in raw materials and finished products. The favored method of analysis is immunoassay, particularly the form known as enzyme-linked immunosorbent assay (ELISA). In general, the approach adopted has been to develop either a polyclonal or a monoclonal antibody preparation that is specific for a major allergen present in a food stuff, such as *Ara h 1* from peanuts. Whilst it is possible to develop sensitive calibration curves for allergens in buffer, difficulties can be encountered when analyzing foodstuffs, where the presence of fats and polyphenol components found in products such as chocolate can reduce assay sensitivity by 10-fold. Nevertheless, such assays have been developed for a range of allergenic proteins, including milk proteins, hazelnuts, and peanuts, and are available for manufacturers, retailers, and other analysts to detect allergen contamination. Another exciting new development is the application of highly sensitive mass spectrometry methods to allergen analysis, which has the potential to offer a second orthogonal method of analysis to immunoassay, as is desirable for purposes of enforcing labeling regulations and hazard control procedures.

See also: **Food Intolerance**: Types; Food Allergies; Milk Allergy; Lactose Intolerance; Elimination Diets; **Food Labeling (Labelling)**: Applications; **Immunoassays**: Principles; Radioimmunoassay and Enzyme Immunoassay; **Protein**: Chemistry; Food Sources; Determination and Characterization; Interactions and Reactions Involved in Food Processing; Heat Treatment for Food Proteins

Further Reading

Bousquet J, Bjorksten B, Brujijnzeel-Koomen CAFM *et al.* (1998) Scientific criteria and the selection of allergenic foods for product labelling. *Allergy* 53: 3–21.

Breiteneder H and Ebner C (2000) Molecular and biochemical classification of plant-derived food allergens. *Journal of Allergy Clinical Immunology* 106: 27–36.

Hefle SL (1999) Impact of processing on food allergens. *Advances in Experimental Medicine and Biology* 459: 107–119.

Hefle SL, Nordlee JA and Taylor SL (1996) Allergenic Foods. *CRC Critical Reviews in Food Science and Nutrition* 36(S): S69–S89.

Lehrer SB, Horner WE and Reese G (1996) Why are some proteins allergenic? Implications for biotechnology. *CRC Critical Reviews in Food Science and Nutrition* 36(S): S553–S564.

Lessof M (ed.) (1997) *Food Allergy Issues for the Food Industry.* Surrey: British Food Manufacturing Industries Research Association.

Allergy *See* **Food Intolerance**: Types; Food Allergies; Milk Allergy; Lactose Intolerance; Elimination Diets

ALMONDS

D E Kester and A A Kader, University of California Davis, Davis, CA, USA
S Cunningham, Blue Diamond Almond Growers, Sacramento, CA, USA

Introduction

The almond is the major commercial tree nut crop of the world. This importance has been achieved by the very large increase in acreage and production in California during the past 25 years. This article reviews the areas of production, principal cultivars, important uses, and methods of handling and storage.

The Crop and its Importance

Global Distribution

The cultivated sweet almond (*Prunus dulcis* Miller (D. A. Webb) syn. *P. amygdalus* Batsch) originated from within the wild species known originally as

Amygdalus communis L. which grew on the lower slopes of mountains in central Asia. About 30 related almond species have been described occupying specific ecological niches in the arid steppes, mountains, and deserts of central and south-western Asia and southern Europe.

The geographical range of the cultivated almond corresponds to the three stages of cultural evolution: (1) Asiatic (south-west and central Asia); (2) Mediterranean (countries bordering both sides of the Mediterranean sea); and (3) Californian (central valleys of California, parts of Australia, central Chile, and areas of South Africa).

Almonds are adapted to Mediterranean, steppe, and desert climates characterized by mild, rainy winters and hot, dry summers. Although they have traditionally been grown with other arid tree and vine crops, such as olive, pistachio, and grape, almond trees respond so well to supplementary irrigation, fertilizers, good soil, disease, and insect control and other intensive culture methods that yields can be increased five-to 10-fold over the traditional culture practiced for centuries.

Commercial Importance

Since 1950, almonds have become the most important world tree nut crop, with present annual production of approximately $350–400 \times 10^6$ kg ($800–900 \times 10^6$ lb). In any given year, California produces about 70% of the world's supply of almonds (**Figure 1**), with most of the balance coming from Mediterranean countries (Spain, Italy, Greece, Portugal, Morocco, Tunisia, Canary Islands, France). Almonds are also grown, often in primitive culture, as local crops in Iran, Israel, Afghanistan, Pakistan, south-eastern Russia, north-west India, Syria, Turkey, and Iraq.

Cultivars

Older botanical literature described two botanical varieties as *Prunus amygdalus* var. *dulcis*, the sweet almond, and *P. amygdalus* var. *amara*, the bitter almond. In reality, bitter and sweet kerneled trees exist side by side in seedling populations with the same parents; the difference is apparently due to a single dominant gene. Some of the more primitive almond-growing areas (e.g., in Asia and Morocco) still grow almond trees as seedling populations (wild or cultivated).

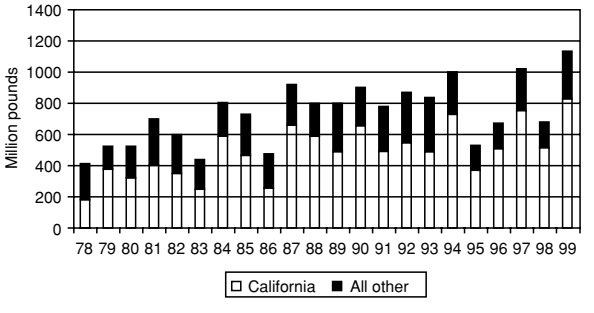

Figure 1 Almonds produced in California in comparison to world production.

Modern orchards grow selected cultivars, which are clones that are vegetatively propagated by budding or grafting upon specific rootstocks. Rootstocks include almond and peach seedlings, as well as certain plums and peach × almond hybrids. Historically, many cultivars were apparently identified when superior individual trees were selected to replace 'bitter' trees within seedling orchards. In recent years, controlled breeding has created cultivars with special qualities. Use of genetic engineering methods will augment genetic improvement.

Although many cultivars exist in the world, a limited number have dominated industrial and local use within specific production areas. Cultivars are selected for performance in the orchard. Flowers on most current cultivars are self-incompatible, so that two or more cultivars that bloom at nearly the same time need to be planted together, preferably in adjoining rows. Hives of honeybees are placed within the orchard to effect cross-pollination. Combinations of cultivars are also chosen to facilitate an economical harvest. Other cultivar characteristics of importance include yield, tree habit, disease and insect resistance, ease of harvest and processing, and other specific characteristics. Nuts of different cultivars vary from very hard-shelled to soft- and paper-shelled. Cultivars are also chosen for their marketing potential, and specific marketing classes have evolved.

California Nonpareil accounts for around 44% of the production. Kernels are uniform in shape, relatively flat, smooth, attractive in appearance, easy to blanch, and are versatile in processing and utilization. The nut has a thin, papery shell which is easy to remove without damaging the kernel. This shell is subject to worm and bird damage. The cultivar is subject to a genetic variant referred to as noninfectious bud failure which is controlled through selection and management of propagation sources.

Mission (Texas) is an old cultivar which produces small, round, plump nuts. Kernels have a relatively strong almond flavor owing to higher levels of amygdalin. The pellicle is difficult to remove during blanching. The shell is hard and protective of worm damage. Padre is a similar cultivar, the kernels of which can be combined with Mission. Peerless is grown to a limited amount as a pollinator of

Nonpareil. The shell is hard and the nuts are sold primarily in the shell.

Ne Plus Ultra has a large, elongated kernel, blooms early, and cross-pollinates Nonpareil. The cultivar is gradually becoming obsolete in California, but orchards still exist and it represents 3% of the total crop.

In the mid-1960s, Merced and Thompson were planted to replace Ne Plus Ultra, Peerless, and Mission. Similarly shaped and blanchable kernels of these and later introduced cultivars have been mixed and marketed under the name California. Other cultivars of this group include Price, Fritz, Monterey, and others.

Carmel became popular in the mid-1970s and is the second most important cultivar in new plantings (18% of almond acreage in 1996). Sonora was introduced in the mid-1970s and is becoming popular as an early-blooming pollinator for Nonpareil. Sonora has large, elongated nuts and is sometimes marketed with Carmel.

Additional cultivars that have recently been planted in California include Butte, Livingston, LeGrand, Mono, and Ruby. Their nuts are grouped either as Mission or California marketing classes.

Spain Marcona and Desmayo are the most important cultivars in Spain, each accounting for about 25% of Spanish production. Marcona has large, heart-shaped, flat, high-quality kernels which blanch easily. Desmayo produces large, elongated, flat kernels, the pellicle of which is thin and can be readily removed after roasting by rubbing between the fingers. Many other cultivars exist in specific production areas of Spain but their kernels are usually mixed together for marketing.

Italy Most of the older areas (Puglia, Sicily, and Sardinia) are reducing production. Kernels of the many cultivars are mixed together for marketing. The main cultivars in newer orchards include Tuono, Felipo Ceo, and Genco, which are self-fertile.

France New, late-blooming, high-yielding, vigorous cultivars derived from breeding programs are being planted not only in France but other countries of the Mediterranean area. These cultivars include Ferragnes and Ferraduel, kernels of which are marketed together with those from other cultivars.

Morphology and Anatomy

Fruit

The almond fruit is known botanically as a drupe, which is composed of three basic parts: (1) exocarp (skin), which in almond is pubescent; (2) mesocarp, which is fleshy and becomes the hull; (3) endocarp (shell). This overall structure is the same as stone fruits, such as peaches, plums, and apricots, but differs in that the almond mesocarp does not enlarge into fruit, but dehisces (splits) at maturity and dries. The hull is an important feed for livestock.

Seed

The almond kernel is the seed which develops from the ovules. Although two ovules are present in the flower, only one normally develops to produce a single kernel. The seed is edible because it lacks amygdalin, the bitter compound in many seeds. The seed consists of the embryo (also referred to as the meat), surrounded by seed coats or testae (also referred to as the pellicle and sometimes the skin). The embryo consists of the hypocotyl root axis with growing points for roots and shoots which appear upon germination, and two massive cotyledons, which are storage organs and contain the high-energy compounds characteristic of almond.

Almond kernels of different cultivars have characteristic sizes, shapes, appearances, thickness of pellicle and, to some extent, flavor. Double kernels result when two kernels develop within the same nutshell. Twin kernels result when two or more embryos develop within the same pellicle.

Harvesting, Handling, and Storage

Harvesting

Almond nuts should be harvested as soon as possible after maturation to avoid quality losses and to minimize problems with fungal attack and insect infestation. Indices used to determine maturity stage and optimum harvest dates for almonds include hull dehiscence (splitting), separation of the hull from the shell, decrease of the fruit removal force, and drying of hulls and kernels. Almond harvesting in California is by machines. Almonds are shaken from the tree to the ground by various kinds of shakers. After about a week of drying, the nuts are raked into rows and picked up by mechanical harvesters. They are then transported in bulk to hulling machines which separate the hull from the in-shell product.

Postharvest Handling

Dehydration of the kernel begins while the nut is still on the tree. Further drying takes place after the nuts are knocked to the ground. Under rainy and/or cool weather conditions, heated-air dehydrators may be used to reduce moisture content to 7% or less. Most almonds sold commercially range from 4% to 5% in

moisture. Dried almonds are transported in bulk to processing plants where they are stored in bins, silos, or other bulk-storage containers for a few weeks to several months before final processing and preparation for market.

In-shell almonds are fumigated with methyl bromide or aluminum or magnesium phosphide, which are lethal to all insect life stages. Residues are monitored to insure that levels in finished products are below legal limits. (*See* **Fumigants**.)

Alternatively, almonds can be kept in a controlled atmosphere of 0.5% oxygen and about 10% carbon dioxide (balance nitrogen) to control insects. The nuts can be frozen for insect control at home, or for small-scale operations. (*See* **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs.)

Storage

Compared to other nuts, almonds and almond products have a long life. This is attributable in part to their low moisture, the presence of relatively low levels of polyunsaturated fatty acids, and high tocopherol levels.

Raw almonds should be stored at 0–5 °C, and 65–70% relative humidity (RH), to minimize deterioration during storage. Prolonged exposure to direct sunlight causes the skins to darken and reduces shelf-life. Since almonds readily absorb odors, they should never be exposed to pungent odors from onions, fresh fruits, fish, cheeses, paint, chemicals, or other products.

Both the maintenance of quality and safety of almond and the extension of storage life depend upon the initial moisture content, the RH and temperature of storage, and the exclusion of oxygen and insect pests. (*See* **Storage Stability**: Parameters Affecting Storage Stability.)

The Food and Drug Administration (FDA) regulations for tree nuts define a safe moisture level (i.e., moisture content which will not support fungal growth) as a water activity level that does not exceed 0.70 at 25 °C. This is equal to a moisture content of about 7%. The optimum range for unroasted almonds is 4–6% moisture. The equilibrium moisture of almonds ranges from 3.8 to 5.0% at 30% RH and from 8.1 to 10% at 75% RH at 5 °C.

The relationship between moisture content and equilibrium relative humidity (ERH) is temperature-dependent. From 20% to 80% ERH (for any given moisture content), ERH rises approximately 3% for every rise in temperature of 10 °C. At a given RH, air will contain more water vapor at a high temperature than at a low temperature. Temperatures between 0 °C and 5 °C are recommended for almonds; the lower temperatures allow longer storage life (up to a year).

Low oxygen (0.5% or lower) atmospheres are beneficial for keeping flavor quality, delaying rancidity, and controlling stored products against insects. Exclusion of oxygen is usually achieved by vacuum packaging or replacement with nitrogen in storage facilities and transport vehicles. The storage life of raw almonds can be extended up to 2 years under low-oxygen atmosphere at 0 °C. (*See* **Chilled Storage**: Packaging Under Vacuum.)

Quality and Safety Factors

Important appearance factors for almonds marketed in the shell include shell integrity, suture opening, and shell color. Kernel defects include damage from insects, mold and mechanical injury, gum, callus growth, shriveling and doubling. Kernels are graded for size. Quality criteria used in the grade standards include freedom from dust particles and other foreign materials, and uniformity of shape and color.

Textural factors, including crispness and firmness, are influenced by moisture content. Thus, roasted almonds are usually more crisp than raw almonds.

Flavor quality depends upon sweetness, oiliness, intensity of almond flavor, and absence of off-flavors resulting from rancidity, staleness, or other causes. A problem called concealed damage, which appears as internal darkening and poor flavor after roasting, is related to wet conditions and high temperatures during temporary storage following harvest.

Safety factors relate primarily to the potential contamination of almonds with mycotoxins, especially aflatoxin produced by the fungus *Aspergillus flavus*. Contamination can occur in the orchard or during postharvest handling if recommended handling and storage procedures are not followed. Elimination of nuts exhibiting any symptoms of fungal contamination is an integral part of market preparation. (*See* **Mycotoxins**: Occurrence and Determination.)

Residues of pesticides used to control the major insect pest, navel orange worm (*Anyelois transitella* Walk), have been reduced by 40% during the last 10 years. This reduction was made possible through such integrated pest management procedures as orchard sanitation, maintaining natural predator balance, and prompt harvest. (*See* **Pesticides and Herbicides**: Residue Determination.)

Propylene oxide, the only sterilant approved for nuts, is generally effective against bacteria, but less effective against yeasts and molds. The nuts are treated in a specially designed vacuum chamber. After treatment, a series of air washes under vacuum removes traces of the remaining gas. Each chamber load then enters a postconditioning staging area until cleared for both residue level and microbial load.

Propylene oxide-treated products are recommended when raw almonds are incorporated into dairy products, such as cheese or yogurt, or in high-moisture foods, or used for other microbially sensitive applications.

Composition and Nutritional Quality

The chemical composition related to nutritive value is summarized in Table 1.

Nuts are a healthy snack if eaten in moderation. Almonds and other nuts are good sources of protein, essential fatty acids, fiber, vitamin E, and minerals. Most of the fat in almonds is monounsaturated. This may be partly responsible for the association observed between frequent nut consumption and reduced risk of coronary heart disease.

Market Forms

Almonds are eaten alone as snacks and are included in many processed foods to enhance their appeal. The wide variety of foods includes sweets, health foods, baked goods, cereals, icecream, dry mixes, garnishes for food entrées, and packaged snacks.

Table 1 Chemical composition and nutritive value of almonds (per 100 g edible part)

Component	*Mean*
Water (%)	5.2
Proteins (%)	21.3
Fats (%)	50.6
Carbohydrates (%)	19.7
Fiber (%)	11.8
Ash (%)	3.1
Sugars (%)	4.8
Food energy (MJ)	2.42
Fatty acids	
Saturated (g)	3.9
Monounsaturated (g)	32.2
Polyunsaturated (g)	12.2
Vitamins	
Vitamin A (IU)	10
Thiamin (mg)	0.2
Riboflavin (mg)	0.8
Niacin (mg)	3.9
Pantothenic acid (mg)	0.4
Folate (μg)	29
Vitamin C (mg)	0
Minerals (mg)	
Calcium	248
Phosphorus	474
Iron	4.2
Sodium	1.0
Potassium	728
Magnesium	275

Source: USDA Nutrient Data Laboratory (http://www.nal.usda.gov/fnic/cgi-bin/list_nut.pl).

Almonds may be used directly as a whole nut, or it may be appropriate for them to be chopped, diced, sliced, split, halved, slivered, or cube-cut. They are available in an array of kernel sizes as segregated by round and slot-holed screens which separate according to kernel width and thickness. There is also a wide range of smaller-sized products, such as flakes, slivers, and meals. The variety of shapes, sizes, and colors available makes almonds appropriate ingredients for many different foods.

The largest usage category is confectionery. Many premium confectionery products are enhanced by the combination of chocolate and almonds. Almonds enhance the flavor and acceptability of confections by lowering the sweetness of the finished piece, by adding crunch, enhancing the nutritional value, and adding sales appeal.

Manufacturing

Roasting

The flavor of most almond varieties is quite mild before they are roasted. The strong flavor and crunchy texture desired in most applications are developed by roasting. Most almonds added to chocolate are roasted almonds. Almonds may be roasted by hot air or in hot oil. Almonds roasted in oil may pick up some of the flavor of the roasting oil. Dry-roasted almonds usually possess the same roasted flavor, a somewhat harder texture, and slightly lower moisture (below 2%).

The amount of oil absorbed by almonds in oil roasting is minimal, usually 3–4%, roughly equal to the amount of moisture driven off by roasting. This small amount of oil has a minimal impact on the fat content of the final product. The degree of roast in a given application is determined by experimentation and/or sensory testing.

Blanching

Blanching is the process of soaking the almonds briefly in hot water, then slipping the skins from the kernels using rollers.

Blanched almonds have a milder flavor and softer texture than unblanched almonds. The only cultivar which is not blanchable is Mission. Blanched almonds are preferred in applications where the brown skin may come loose upon cooking, or a light nut is desired.

Almond Butter

Almond butter is made by grinding (dry) roasted almonds with other ingredients, such as salt, sugar, and stabilizer. The predominant flavor of almond

butter is that of roasted almonds. Almond butter is similar to peanut butter in appearance, with a slightly oilier texture. It is used in a wide variety of food applications.

Almond Paste and Marzipan

Almond paste and marzipan have long been used in baked goods, pastries, and confections, particularly in Europe. Marzipan is traditionally shaped into figures for holidays and in some countries is enclosed in chocolate. Almond paste is among the oldest of confections and is made by grinding raw blanched almonds with sugar.

Extending Shelf-life

The package as well as the moisture, temperature, and RH of the surrounding environment must be considered in predicting shelf-life of products using almonds. With higher moisture and RH in the surrounding environment, the shelf-life is usually shorter.

As a general rule, blanching reduces shelf stability by about 25–50%. Any cutting, such as slicing and dicing, has a similar effect. Roasting accelerates deterioration, and roasted nuts should be packaged to exclude oxygen. When properly packaged in cans, foil, or glass under vacuum or nitrogen-flushed, both dry-roasted and oil-roasted almonds can last a year or longer at room temperature. Dry-roasted almonds tend to have a longer shelf-life than oil-roasted almonds. The quality and stability of oil-roasted almonds depend upon the type and quality of roasting oil. Finely ground products, such as almond paste and almond butter, have a long shelf-life (>1 year) because the particles pack tightly together, excluding oxygen.

Roasted almonds in oxygen-free packaging have a shelf-life of 1–2 years at room temperature. If longer storage is desired, or if harsh storage conditions exist, treatment with antioxidants should be considered. A variety of natural and artificial antioxidants can contribute to a two- to threefold improvement in shelf-life. The ideal package for almonds excludes both moisture and oxygen. Shelf-life in some packages may be improved by the addition of oxygen and/or moisture scavenger packets. (*See* **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants.)

See also: **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; **Chilled Storage**: Packaging Under Vacuum; **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Fumigants**; **Mycotoxins**: Occurrence and Determination; **Nuts**; **Pesticides and Herbicides**: Residue Determination; **Storage Stability**: Parameters Affecting Storage Stability

Further Reading

Fisher MC and Lachance PA (1999) *Nutrition and Health Aspects of Almonds*. Modesto: Almond Board of California (website: www.almondsarein.com).

IPM Manual Group (1985) *Integrated Pest Management for Almonds*. Publication no. 3308. Oakland: University of California Division of Agricultural and Natural Resources.

Kester DE and Gradziel TH (1993) Almond breeding. In: Janick J and Moore JS (eds) *Advances in Fruit Breeding*, 2nd edn, pp. 1–97. Indiana: Purdue University Press.

Kester DE, Gradziel TM and Grasselly C (1991) *Almonds. Germplasm Resources in Fruit and Nut Species*, pp. 70–758. Wageningen, The Netherlands: International Society for Horticultural Science.

Micke WC (1996) *Almond Production Manual Publication No. 3364*. Oakland: University of California Division of Agricultural and Natural Resources.

Rosengarten F Jr (1984) *The Book of Edible Nuts*. New York: Walker.

Ryall AL and Pentzer WT (1982) *Fruits and Tree Nuts*, 2nd edn, vol. 2, ch. 10. Westport, Connecticut: AVI.

Shirra M (1997) Postharvest technology and utilization of almonds. *Horticultural Reviews* 20: 267–311.

Soderstrom EL and Brandl DG (1990) Controlled atmospheres for preservation of tree nuts and dried fruits. In: Calderon M and Barkai-Golan R (eds) *Food Preservation by Modified Atmospheres*, pp. 83–92. Boca Raton, Florida: CRC Press.

Woodruff JG (1979) *Tree Nuts: Production, Processing, Products*. Westport, Connecticut: AVI.

ALUMINUM (ALUMINIUM)

Contents

Properties and Determination

G R Boaventura and J G Dórea, Universidade de Brasília, Brasília, Brazil

Introduction

Aluminum (Al) is a white metal, which has low density, and is ductile and malleable. It is stable in the presence of atmospheric air, and does not react with water. It is easily dissolved in most acidic media liberating hydrogen. The naturally occurring isotope 27 is found with 100% abundance. Three more isotopes – 26, 28, and 29 – can be obtained artificially. Its physical and chemical characteristics are shown in Table 1.

Al is considered one of the most commonly occurring elements in nature. It is found in combination with other elements. After oxygen and silicon it is the most abundant element on the earth crust, comprising 7.5% of its mass. The main ores containing Al include bauxite (AlO(OH)), corundum (Al_2O_3), spinel, and silicates. It is largely used in construction, the metallurgic industry, aviation, maritime equipment, food industry, packaging, and the chemical and pharmaceutical industry.

Al can occur in the metallic particulate form, and also complexed with organic compounds. In foods, including water, it occurs mostly in particulate form, exchangeable Al, nonexchangeable organic, and inorganic Al. In biological samples, Al is normally present in the $\mu g\,g^{-1}$ to $mg\,g^{-1}$ concentration range.

Due to Al abundance on the earth crust (natural), and the many ways it is used (synthetic), people are exposed to Al contamination. Furthermore, as it is frequently used in the food industry as well as in products employed in water treatment, there is increasing interest in its toxic effect on humans. Additionally, industrial activity that causes acid rain increases the availability of Al in the environment.

Because of its low absorption in the human intestine, Al is still considered as nontoxic. However, it is recognized that Al can be toxic to plants, laboratory animals, and humans, depending on the level and chemical form. Absorption, retention, and excretion of Al depend on Al chemical complexes formed with biological ligands. The study and identification of such complexes are treated separately as metal speciation: because of specific techniques of separation and identification, this is beyond the scope of this article.

Table 1 Physicochemical characteristics of aluminum

Symbol	Al
Atomic number	13
Atomic mass	26.98
Melting point	660.45 °C
Boiling point	2520 °C
Density	2.7 g cm^{-3}
Atomic volume	9.99 cm^3
Ionization potential	5.98 eV
Electrode potential (ε^0)	1.66 V
Electronegativity	1.5

Analytical Methods for Al Determination

There are numerous analytical techniques that can be used for Al determination, both traditional and novel. The choice of analytical method is determined by the type of sample and the expected concentration of Al. A typical range of concentrations expected in biological samples is illustrated in Table 2.

Due to the abundance of Al on the earth crust, the first challenge is reducing environmental contamination as this may be an important source of systematic error in sample preparation. Therefore, it is important to consider an Al-free environment, with high-purity water and reagents.

Sample Preparation

Biological samples, especially human, like blood, urine, saliva, and feces, must be handled with extreme caution. In these cases it must be considered that they can be infected with transmissible diseases. Blood samples must always be collected by qualified personnel.

Table 2 Typical concentrations of aluminum in different environments

	Typical concentration (mg kg^{-1})
Marine plants	60
Land plants	500
Soils	71 000
Crustal average	81 300
Granite rocks	74 300
Open sea water	< 0.001
Lake water	0.010–0.5
River water	0.01–40
Groundwater	0.01
Milk	0.02
Cows' liver	0.01–0.05
Beef muscle	0.02–0.06
Human blood	< 1.6–5 mg l^{-1} × 10^{-3}
Orange juice	0.1–0.6
Wine	0.3–2
Potatoes	0.1
Tea leaves	400–1000
Beech leaves	37–163
Beech seeds	6–13

Table 3 Typical concentrations of aluminum in reagents and laboratory glassware

Reagents	*Aluminum (ng ml^{-1})*
Distilled water	
Tin still	0.8
Polypropylene still	0.3
Quartz still	3
Deionized water	0.06
Nitric acid, reagent grade	10
Nitric acid, subboiling	0.9
HCl, high-purity, commercial	100
HCl, subboiling quartz still	0.8

During sample preparation, metal utensils commonly used in handling food are a source of aluminum contamination and should be avoided at all stages. Decomposition of solid samples for total Al is simple, and as with other metals can be done with NaOH fusion, *aqua regia* (HCl:HNO_3) inorganic acids (H_2SO_4, HNO_3, and HCl). For fractionation of Al compounds, exchange resins are employed as required by the nature of the Al complex. In acidic decomposed samples or effluent from exchange resins, Al can be determined by any method suited for the expected concentration. Traditional analytical methods can be used for high aluminum concentrations, using oxine and other specific reagents.

Certified Reference Materials

Reference materials or certified reference materials are essential tools for quality control of analytical results. These materials are accompanied by a certificate of the concentrations of one or more elements by a specified procedure. The commercially available Standard Reference Material (SRM) of the National Institute of Standards and Technology (NIST) of the USA for Al in nonfat bovine liver, corn kernel, bovine muscle, rice flour, egg powders, and others are suitable for most food and nutrition Al determinations.

Purity of Reagent and Quality Control of Sample Acquisition and Storage

In ultratrace Al determination, conventional apparatus and reagents are considered to be far from satisfactory. The most used material sample storage and reagent solution is Teflon (a registered trademark of DuPont de Nemours and Co. Inc. USA) and quartz glass. In these materials the Al contamination levels are very low. Environmental contamination depends on the particulate level of Al in that particular environment. The Al concentration in water, reagents, and other materials generally used in analytical laboratories are presented in Table 3.

Satisfactory Al concentrations in reagents used in Al determination can be obtained by purification. Most water purification equipment available commercially is satisfactory. There is also equipment for acid purification and a simple subboiling apparatus has been proposed that can be assembled in a well-equipped analytical laboratory. Two Teflon bottles joined by a Teflon joint, where one is heated and the other serves as an acid-purified collector, can be used to purify HNO_3, HF, and HCl.

Instruments

The instrumental techniques most commonly used in trace Al determination are spectrofluorometry, molecular spectrometry, atomic absorption spectrometry (AAS) with flame (FAAS) and with graphite furnace (GFAAS), inductively coupled plasma–atomic emission spectrometry (ICP-AES), inductively coupled plasma–mass spectrometry (ICP-MS), X-ray fluorescence spectrometry (XRS), and instrumental neutron activation analysis (INAA).

Molecular Spectrophotometry/Colorimetry

Spectrophotometric (ultraviolet and visible) techniques can be used for sample solutions with concentrations above 0.1 mg l^{-1}. In this technique Al is reacted with a chromogen, usually aurintricarboxilate (aluminon), criocromocianin R, oxine, or alizarin red. Although this technique uses a low-cost instrument, it involves a laborious procedure. There are also interfering substances that can react with chromogen.

Spectrofluorimetry

Compared to other colorimetric methods, fluorimetry offers more sensibility with three orders of magnitude in its detection limit. As with colorimetric methods in general, fluorometry requires the formation of fluorescent complexes: usually Pantochrome dark blue R is used. In chemically complex matrices its selectivity is low and produces low yields. In blood samples fluorescence with 2-hydro-1-naphthaldehyde-*p*-methoxybenzoyl hydrogen has been used. Samples are read at 475 nm with a detection limit of $0.13 \times 10^{-1}\,\mathrm{mg\,l^{-1}}$.

Atomic Absorption Spectrophotometry

The direct determination of Al by AAS is based on the radiation of free atoms with a minimum of interference. The determination is accomplished in the atomized state of the mineral after nebulizing liquid samples. Al is determined at 309.3 nm, using a hollow cathode lamp (HCL) or an electrodeless discharge lamp (EDL) in N_2O/acetylene reducing (rich red) flame. The EDL provides a greater light output and longer life than HCL. In concentrated samples it is possible to use wavelengths of less sensibility, such as 396.2, 237.3, 257.5, and 256.8 nm, where sensitivity can reach $6.0\,\mathrm{mg\,l^{-1}}$. The air/acetylene flame is not recommended because aluminum oxide hinders the acceptable limits of detection for most applications. The partial ionization of Al can be suppressed by adding lanthanum nitrate or potassium chloride in concentrations ranging between 1 and $2\,\mathrm{mg\,l^{-1}}$ in all samples, including blanks and standards. The presence of silicates, nickel, cobalt, chromium, and manganese can decrease sensitivity. Background correction must be used in samples with a high concentration of salts. Sensitivity is in the order of $1\,\mathrm{mg\,l^{-1}}$ with a detection limit of $0.03\,\mathrm{mg\,l^{-1}}$ and good work range within $10–100\,\mathrm{mg\,l^{-1}}$.

In most matrices the use of N_2O/acetylene flame limits the use of AAS for Al determination. However, flameless electrothermal atomization in a graphite furnace (GFAAS) enhances the application of AAS for low Al concentrations. The detection limit in this procedure can reach $0.00003\,\mathrm{mg\,l^{-1}}$ compared to $0.06\,\mathrm{mg\,l^{-1}}$ in N_2O/acetylene flame at 309.27 nm.

Flameless electrothermal atomization can be used in liquid samples that have been predigested or directly in solid samples of biological materials. Although high in sensitivity, the procedure requires analytical skills and longer handling time, which renders it difficult for large numbers of routine analysis. Constant instrument calibration also requires analytical skills.

Compared with other techniques for the same degree of sensitivity, like ICP-MS the GFAAS is considered to be simple, effective, and reliable with a much lower operational cost.

Inductively Coupled Plasma Atomic Emission Spectrometry

In the last 20 years there has been an increase in the use of ICP-AES. Several factors have contributed to its increasing preference over flame AAS. As a source of atomic excitation the plasma has a greater capacity. Advantages include the possibility of multielemental analysis, less interference, mostly of spectral nature, and good sensitivity.

When multielemental analysis is considered there are some operational parameters to be planned. The choices of combination of concentration of the elements in question, its variability and expected range of concentrations can pose a challenge to an experienced analyst.

The use of ICP-AES in food and nutrition is of particular interest, especially when considering multielemental determinations. With advances in microprocessors the complexity of multielement determination has been reduced. The use of simultaneous or sequential spectrometers has greatly improved the routine analysis of matrices containing trace amounts or highly concentrated elements.

Interference in ICP-AES is only found in spectral lines. However, these are fewer and easily solved. In the case of Al there are alternative lines and program designs for instrument operation that take into consideration the concentration of Al. The wavelength at 308.25 nm is the most used and is almost free of interference. For the determination of trace amounts of Al in the wavelengths at 167.08 nm in UV or optic vacuum, spectral interference can be further reduced.

The typical Al detection limit by ICP-AES using a Meinhard nebulizer is $0.08\,\mathrm{mg\,l^{-1}}$ and $0.00033\,\mathrm{mg\,l^{-1}}$ using an ultrasonic nebulizer at 308.215 nm. In biological fluids or tissues where Al concentrations can be low, i.e., less than $0.03\,\mathrm{mg7\,ml^{-1}}$, samples have to be preconcentrated and/or extracted.

Inductively Coupled Plasma Mass Spectrometry (ICP-MS)

ICP-MS combine two established analytical source (ICP and MS) in this relatively new analytical tool. Although one advantage of ICP-MS is the simplicity of the spectra, an overlap of spectra can happen, thus hindering some of its applications. In the determination of Al, both Mg and molecular N cause interference. Background interference is usually related

Table 4 Comparison of technical and operational characteristics of aluminum instrumental determination

Technique	*Fluorimetry* ($mg\,l^{-1}$)	*UV/visible* ($mg\,l^{-1}$)	*AAS* ($mg\,l^{-1}$)	*GFAAS* ($mg\,l^{-1}$)	*XRS* ($mg\,kg^{-1}$)	*ICP-AES* ($mg\,l^{-1}$)	*ICP-MS* ($mg\,l^{-1}$)	*INAA* ($mg\,l^{-1}$)
Limit of detection	0.13×10^{-3}	0.1	0.06	0.00003	50	0.03	0.003	20
Degree of complexity	Simple	Simple	Medium	Complex	Simple	Simple	Medium	Complex
Cost of equipment	Low	Low	Medium	High	Very high	High	Very high	Very high
Cost/analysis	Medium/high	Medium/high	Low	High	Low	Low	Medium	High
Analyst skill	Minimal	Minimal	Minimal	High	High	Intermediate	High	High
Length of time	High	High	Low	High	Low	Low	Medium	High
Food and nutrition application	Rare	Rare	Common	Common	Rare	Common	Novelty	Rare

UV, ultraviolet; AAS, atomic absorption spectrometry; GFAAS, graphite furnace atomic absorption spectrometry; XRS, X-ray fluorescence spectrometry; ICP-AES, inductively coupled plasma–atomic emission spectrometry; ICP-MS, inductively coupled plasma-mass spectrometry; INAA, instrumental neutron activation analysis.

to ArO^+, Cl^+, ClO_2^+, NO_3^+ and isotope analytes which originate from impurities of argon-gas used as a plasma source. Elements with charge can form oxides of the type metal–O^+ and others with double charge metal–OH^+ can cause serious spectral interference.

Limits of detection are reported in the order of $0.003\,mg\,l^{-1}$, and sensitivity is between GFAAS and flame AAS, ICP-AES and INAA. The high cost of equipment, maintenance, and operation is still restricting its use.

INAA is a powerful multielement technique determining trace elements in a variety of sample matrices. The samples are irradiated in a neutron flux inside a nuclear reactor and the energy of the γ-ray peaks forms a spectrum that is used to identify the analyte. However, one of its limitations is the appropriate handling of radioactive materials and chemical interference. ^{28}Al has a half-life of 2.2 min and its decomposition products for complexes with both Si and P. In an ideal matrix, the INAA detection limits for Al are between 1 p.p.b. and 1 p.p.m, which for many biological samples may require sample preconcentration. The INAA determination of Al in food and nutrition samples is rare because of costs and operational difficulties in sample irradiation.

Conclusion

Aluminum determination in biological samples can be accomplished by all available instrumental techniques. Its application in food and nutrition will depend on the availability of instruments. In view of its abundance in the environment, the interest in its determination can range from trace concentrations to relatively large amounts. Although almost any technique is suitable for Al determination in large concentrations there are sample analyses that would require a certain degree of suitability and judgement will depend on the analyst's skill and resources. Although AAS is relatively available in most analytical laboratories, its use in food analysis is suitable for concentrations greater than $mg\,l^{-1}$ because of its limit of detection (Table 4). For trace concentrations of Al, AAS is only possible when coupled with graphite furnace. The versatility of multielement determination offered by ICP-AES is becoming attractive in food and nutrition studies. The limitations of ICP-AES are of the same nature as AAS for trace concentrations of Al but in the case of ICP coupled with MS, very low limits of detection with precision and accuracy are possible. For routine analysis, ICP-AES is superior in speed and operational simplicity to AAS.

See also: **Spectroscopy**: Fluorescence; Atomic Emission and Absorption; Nuclear Magnetic Resonance

Further Reading

Boumans PWJM (1987) *Inductively Coupled Plasma Emission Spectroscopy, Part II: Applications and Fundamentals*. New York: Wiley.

Lide DR (1992) *CRC Handbook of Chemistry and Physics*, 72nd edn. New York: CRC Press.

Mattinson JM (1972) Preparation of hydrofluoric, hydrochloric and nitric acids at ultra-low lead levels. *Analytical Chemistry* 44: 1715–1716.

Montaser A and Golightly DW (1987) *Inductively Coupled Plasmas in Analytical Atomic Spectrometry*. New York: VCH.

Parry SJ (1991) *Activation Spectrometry in Chemical Analysis*. New York: Wiley.

Stoeppler M (1992) *Hazardous Metals in the Enrivonment*. Amsterdam: Elsevier Science.

Thompson M and Walsh JN (1989) *Handbook of Inductively Coupled Plasma Spectrometry*, 2nd edn. New York: Chapaman and Hall.

Varma A (1985) *Handbook of Atomic Absorption Analysis*, vol. II. Boca Raton, FL: CRC Press.

Toxicology

J L Domingo, Rovira i Virgili University School of Medicine, 43201 Reus, Spain

Introduction

The following aspects concerning aluminum toxicology are covered in this article: essentiality and major sources of aluminum exposure in humans, aluminum in drinking water, aluminum metabolism, experimental toxicity including mutagenicity, carcinogenicity, developmental and neurobehavioral toxicity, and human toxicity. The potential role of aluminum in the pathogenesis of Alzheimer's disease is also reviewed. Finally, the prevention and treatment of aluminum accumulation and toxicity are briefly discussed.

Essentiality of Aluminum

Aluminum is the third most abundant element in the earth's crust and it is also ubiquitous in nature. Although biological systems have probably evolved in the presence of high aluminum concentrations, no essential role has been established for this element in living systems. In general terms, the few available data on the essentiality of aluminum are so limited and controversial that it seems that this element is not essential for growth, reproduction, or survival of humans and animals. In spite of the increased number of spontaneous abortions, depressed growth, and decreased life expectancy reported as signs of aluminum deficiency in goats at an apparent deficient concentration in diet of 162 $\mu g\,Al\,kg^{-1}$, nutritional research has been unable to define a deficiency state associated with minimal aluminum intake. Also, until recently aluminum was considered to be an innocuous element.

Sources of Aluminum Exposure

Dietary Exposure

For the general population, most aluminum exposure occurs through the diet. Aluminum in the food supply comes from natural sources including water, food additives, and contamination by aluminum utensils and containers. However, the intake of aluminum from foods has been reported to be less than 1% of that consumed by individuals ingesting aluminum-containing medications such as antacids or phosphorus-binding gels.

While most unprocessed foods, excepting some herbs and tea leaves, contain less than 5 $\mu g\,Al\,g^{-1}$, food additives are probably the major source of dietary aluminum. Sodium aluminum phosphate is the main aluminum compound used in foods. The acidic forms of this compound are commonly added to cake mixes, frozen doughs, pancake mixes, and self-raising flours to react with sodium bicarbonate and leaven these products when they are moistened. Moreover, the alkaline forms of sodium aluminum phosphate are widely used as additives in processed cheeses and cheese foods. In turn, aluminum potassium sulfate is an additive used to clarify sugar as a firming agent and as a vehicle for bleaching agents. Athough this additive is considered safe, in larger quantities it can interfere with the body's retention of phosphorus. Lastly, the aluminum silicates are commonly used as anticaking agents in salt, nondairy creamers, and other dry, powdered products. (*See* **Emulsifiers**: Uses in Processed Foods; **Leavening Agents**.)

One of the most recent surveys on the dietary exposure of aluminum has been carried out in the UK. In that survey, exposure estimates of aluminum for mean (12 $mg\,day^{-1}$) and upper-range (29 $mg\,day^{-1}$) adult consumers were lower than the PTWI (provisional tolerable weekly intake) of 420 $mg\,week^{-1}$ (equivalent to 60 $mg\,day^{-1}$) for a 60-kg adult set by the Joint Expert Committee on Food Additives (JECFA) of the Food and Agriculture Organization of the United Nations and the World Health Organization International Program on Chemical Safety. The higher aluminum concentrations were found in the miscellaneous cereals group, which might reflect the use of aluminum-containing additives in some bakery products. Dietary exposure estimates of aluminum in that survey were similar to those from other countries. In a study carried out in the USA in 1993, dietary exposures of aluminum ranged from 0.7 $mg\,day^{-1}$ for 6–11-month-old infants to 11.5 $mg\,day^{-1}$ for 14–16 year-old males, while average intakes for adult men and women were 8–9 and 7 $mg\,day^{-1}$, respectively. The major contributors to daily intake of aluminum were foods with aluminum-containing food additives, e.g., grain products and processed cheese. Dietary exposures from Italian Total Diet studies were found to range between 2.3 and 6.3 $mg\,day^{-1}$. Finally, in a study conducted in Japan between 1990 and 1995, the median aluminum intake from diet was found to be 2.5 $mg\,day^{-1}$.

Aluminum in Drinking Water

In addition to the naturally occurring aluminum in raw waters, aluminum-based coagulants such as aluminum sulfate (alum) or polyaluminum chloride are often added for water treatment. Although the

treatment of surface water with alum has been in operation for over 100 years all over the world, the use of aluminum-based coagulants for the removal of particulate, colloidal, and dissolved substances from the water often leads to higher aluminum concentrations in the treated water than in the raw water itself. The maximum allowable concentration of aluminum in drinking water established by the former European Economic Community (now European Union) is $0.2\,mg\,l^{-1}$, while the US Environmental Protection Agency promulgated a secondary maximum contaminated level range of 0.05–$0.2\,mg\,l^{-1}$ of aluminum in drinking water. In turn, the World Health Organization has proposed $0.2\,mg\,l^{-1}$ as a guideline value. These values are not based on any assessment of risks to health, but they provide a compromise between the use of aluminum salts in water treatment and discoloration of distributed water.

Other Contributors of Aluminum to Human Diet

In addition to food containing natural aluminum or additives containing aluminum, the intake of aluminum through diet can also be increased from contact of food with containers, cookware, utensils, cans, and foil wrappings. Although the amount of aluminum ingested as a result of preparing food in aluminum cookware (pans, pots, kettles, trays, and foil) does not appear to be of significance in comparison with the amount consumed from other sources, in some cases release of aluminum from aluminum utensils has been found to be remarkable if used frequently. In relation to this, it has been demonstrated that increased concentrations of complexing ions (organic acids, fluoride ion, OH^-, etc.) significantly enhance the release of aluminum from cooking utensils.

One of the possible routes through which aluminum can enter humans can be foods packed in aluminum containers, e.g., aluminum cans. Aluminum cans are now widely used for packaging of different types of food products, e.g., soft drinks, beers, fruit juices, and mineral water. In general, the aluminum content of beverages from aluminum cans has been found to be higher than that from glass containers. However, recent studies have shown that beers or soft drinks packaged in aluminum cans are an insignificant source of dietary aluminum intake – 0.4 and 1.3%, respectively – of the tolerable daily intake according to the JECFA.

Pharmaceutical Products

Aluminum compounds have been widely used in various nonprescription drugs. They include some antacids, buffered aspirins, antidiarrheal products, douches, and hemorrhoidal medications. Aluminum compounds are also used as adjuvant in vaccines. However, due to some doubts in relation to the potential role of aluminum in some neurological disorders, in recent years some companies are removing aluminum from some of their pharmaceutical products. Among the aluminum compounds currently used in nonprescription antacids, aluminum hydroxide has been by far the most common. Although aluminum hydroxide is considered safe its use should be restricted in subjects with chronic renal insufficiency and probably also during pregnancy. On the other hand, various substances administered intravenously, such as albumin, calcium salts, and phosphate salts, have been reported to contain significant quantities of aluminum. Moreover, aluminum is a known contaminant of total parenteral nutrition solutions. In 1990, the US Food and Drug Administration recommended that aluminum concentrations in parenteral feedings not exceed $25\,\mu g\,l^{-1}$.

Infant Formulae

In recent years, several studies have determined the concentrations of aluminum in infant formulae and compared with those found in human milk and cows' milk. The lowest mean aluminum concentrations have been observed in human milk (9.2–$23.4\,\mu g\,l^{-1}$), while the highest values corresponded to infant formulae, which differed markedly among manufacturers (mean, $551\,\mu g\,l^{-1}$; range, 302–$1149\,\mu g\,l^{-1}$). It has been reported that infants consuming milks containing more than $300\,\mu g\,Al\,l^{-1}$ had raised plasma aluminum concentrations. Although some researchers suggested that this might indicate the maximum intake tolerated by these infants, there is not a general agreement about it.

Aluminum in Soils and Plants

Aluminum is one of the most abundant elements in soil. Natural acidification processes result in increasing solubility of aluminum and, as soils become moderately acidic (pH < 5.5), aluminum begins to appear as the exchangeable cation which dominates in the lower mineral horizons. Concentrations of soluble and exchangeable aluminum in acid soils may reach many micrograms per gram of soil and can be toxic to plants. Aluminum toxicity in agricultural plants is a major problem that has been acknowledged at least since 1918. Worldwide, it has been estimated that 40% of arable soils and perhaps about 70% of potential new lands that can be brought under cultivation are acidic enough to have an aluminum toxicity problem. Apart from the Al^{3+} cation, aluminum has also the potential to form various hydroxy-aluminum and polynuclear species in solution. While the

available evidences suggest that the Al^{3+} cation is phytotoxic, it is unclear whether other hydroxy-aluminum species are also toxic. In contrast, anionic aluminum species in solution are considered not to be toxic. On the other hand, a significant correlation between low pH and high aluminum concentration has also been found in acidified fresh water where this element may reach levels of 0.3–1.6 mmol l^{-1} and cause serious metabolic derangement in some water plants.

Absorption, Distribution, and Excretion of Aluminum

Absorption

Among the potential routes of aluminum absorption in humans, only absorption via the gastrointestinal tract has received adequate consideration in scientific research. Despite significant effort, it is still not known how aluminum is absorbed across the gastrointestinal tract. Although some aluminum absorption may occur in the stomach, the majority of aluminum absorption is expected to occur in the intestine. It is believed that intestinal absorption of aluminum includes both paracellular passage routes along enterocytes and through tight junctions by passive processes and transcellular passage routes through the enterocyte, involving passive, facilitated, and active transport processes. On average, about 0.1–0.3% of the aluminum daily intake is absorbed. However, this figure is based in part on the assumption that urinary excretion represents absorption, which does not take into account tissue distribution, as dicussed below. In a recent study in rats using ^{26}Al, aluminum absorption was 0.97% of the dose.

The efficiency of aluminum absorption is also dependent on a variety of other factors. Two important dietary factors affecting absorption of aluminum are organic and inorganic anions. It has been demonstrated that the concurrent ingestion of aluminum compounds and some frequent organic constituents of the diet such as ascorbic, citric, and lactic acid among others (gluconic, malic, oxalic, and tartaric acids) significantly raised the gastrointestinal absorption of aluminum. The reaction of gastric acid with these organic chelators solubilizes aluminum cations, resulting in the equilibrium formation of a soluble complex of aluminum which, by preventing reprecipitation, results in aluminum absorption. Because of the propensity of uremic patients to retain aluminum even if not given oral aluminum compounds, in order to prevent the potential aluminum accumulation and toxicity it is important to carry out a careful surveillance of the diet of patients with chronic renal failure. In contrast to the effects of the dietary organic acids, concurrent ingestion of aluminum compounds with fluoride or silicon significantly reduces aluminum absorption.

Distribution

It has been estimated that the total aluminum load for healthy subjects is 30–50 mg, with about 50% present in skeleton and 25% in lungs. However, studies in rats and mice have shown that aluminum accumulated in tissues as: bone > spleen > kidney $\simeq$ liver > brain. Recently, it was shown that the aluminum burden in liver, kidneys, and spleen of rats was higher in old than in young or adult animals, whereas brain aluminum levels were higher in young than in adult or old rats. The finding of higher aluminum concentrations in brain of young rats can be particularly significant in relation to the controversial role of aluminum in some neurodegenerative diseases.

Excretion

In general, it is accepted that aluminum is mainly excreted in the urine and to a lesser extent in the bile. The susceptibility of patients with renal insufficiency to aluminum toxicity demonstrates the importance of the kidneys as excretory organs following aluminum ingestion. However, it has also been shown that biliary secretion of oral doses of aluminum occurs rapidly after ingestion and probably reflects a response by the liver to increased aluminum loads brought via the portal circulation. Although studies in rats have suggested that the liver is capable of secreting small amounts of absorbed dietary aluminum into bile, the kidneys become the primary excretory organs for aluminum when the liver's secretory capacity is surpassed after ingestion of pharmacological doses of aluminum. However, the occurrence and possible mechanisms of the cellular and tissue excretion of aluminum are almost completely unresolved.

Experimental Toxicity of Aluminum

Acute Toxicity

The intraperitoneal median lethal dose (i.p. LD_{50}) for aluminum nitrate nonahydrate was reported to be between 327 and 901 mg kg^{-1} in rats, and between 320 and 1587 mg kg^{-1} in mice, whereas the reported oral LD_{50} values in rats ranged from 260 to 4280 mg kg^{-1} depending on rat strain, sex, age, and period of observation. The LD_{50} of aluminum sulfate in mice was 1400 mg kg^{-1} injected intraperitoneally and 6200 mg kg^{-1} when given orally. The LD_{50} of aluminum chloride following oral administration

Table 1 Acute toxicity of aluminum compounds given orally and intraperitoneally to rats and mice

	LD_{50} (for Al^{3+})			
	Rats		Mice	
Test compound	Oral	Intraperitoneal	Oral	Intraperitoneal
$Al(NO_3)_3.9H_2O$	261 (223–305)	65 (50–83)	286 (234–346)	133 (104–201)
$AlCl_3.6H_2O$	370 (270–506)	81 (63–106)	222 (161–302)	105 (86–128)
$Al_2(SO_4)_3.18H_2O$	> 730	25 (21–30)	> 730	40 (33–50)
$AlBr_3$	162 (127–205)	82 (66–103)	164 (129–209)	108 (84–138)

Median lethal dose LD_{50} (14 days) in mg kg^{-1}; 95% confidence limits in parentheses.

was 380 mg kg^{-1} in mice, 400 mg kg^{-1} in guinea-pigs, and 400 mg kg^{-1} in rabbits. The LD_{50} values of some aluminum compounds in rats and mice expressed in mg Al^{3+} kg^{-1} are summarized in Table 1. Decreased locomotor activity, piloerection, weight loss, and decreased food and water consumption are the most notable physical signs appearing after acute aluminum intoxication.

Neurotoxicity

It is well established that, regardless of the host, the route of administration, or the speciation, aluminum is a potent neurotoxicant. In susceptible species, aluminum induces cytoskeletal changes in which neurofilaments accumulate in neuronal cell bodies and proximal axonal enlargement. By contrast, neurobehavioral studies of aluminum in rodents have generally not produced robust or consistent results. Thus, while some investigations showed a correlation between elevated brain levels of aluminum with alterations in a variety of behaviors in rats and mice, other studies did not find significant differences in most of the behavioral tests performed in the rat following aluminum exposure. Recently, it has been shown that concurrent oral aluminum exposure and restraint stress enhances aluminum concentrations in the brain and cerebellum of mice, which can be of special interest taking into account the potential role of aluminum in some serious neurological disorders.

Reproductive and Developmental Toxicity

The gonadotoxic effect of aluminum in rats, guinea-pigs, and rabbits was found to be weak following chronic aluminum exposure, while spertamozoa were only affected at the highest aluminum doses tested. When male mice were intraperitoneally exposed to aluminum nonahydrate for 4 weeks before mating with untreated females, the 'no observable adverse effect level' (NOAEL) was 50 mg kg^{-1} day^{-1}. As most oral aluminum is not absorbed from the gastrointestinal tract, and taking into account that in individuals with normal renal function most aluminum ingested is excreted into urine, it was concluded that there is a remarkable safety margin for any adverse reproductive effects in humans due to aluminum ingestion under the intended conditions of use.

On the other hand, it is well established that parenteral exposure to aluminum during pregnancy can cause a developmental syndrome in mammals that includes resorptions and deaths, skeletal, and soft-tissue abnormalities, and growth retardation. However, until recently there was little concern about embryo/fetal consequences of aluminum ingestion because bioavailability was considered low. Nowadays, it is well demonstrated that aluminum may be a developmental toxin depending on the route of exposure and/or the solubility of the aluminum compound administered. Aluminum chloride was found to be embryotoxic and teratogenic when given parenterally at high doses to rats and mice, whereas teratogenic effects of aluminum were also reported in rats given high oral doses of aluminum nitrate that induced concomitant maternal toxicity. In contrast, maternal and embryo/fetal toxicity were not observed when high doses of aluminum hydroxide, the most common aluminum compound given therapeutically, were given by gavage to pregnant mice and rats during organogenesis. However, some signs of maternal and developmental toxicity were found in mice when aluminum hydroxide was given concurrently with citric or lactic acids. In spite of these results, it seems evident that pregnant women are not currently exposed to aluminum doses that can cause adverse effects on health. Moreover, recent studies have shown that maternal stress is not an additional real risk for women consuming current quantities of aluminum during pregnancy.

Carcinogenicity and Mutagenicity

Animal studies have failed to demonstrate carcinogenicity attributable to aluminum powder, aluminum hydroxide, aluminum oxide, or aluminum phosphate administered by various routes to rats, rabbits, mice, and guinea-pigs. Aluminum fibers were not carcinogenic following i.p. injection to rats, whereas

sarcomas were produced at the site of injection of aluminum-dextran in mice. However, since subcutaneous injection of aluminum-dextran is a rather irrelevant model to assess aluminum carcinogenicity, the significance of the positive results appears questionable. (*See* **Carcinogens**: Carcinogenic Substances in Food: Mechanisms.)

Since aluminum and aluminum compounds are not carcinogenic, or even show some antitumor activity in laboratory animals, the negative results in most short-term *in vitro* and *in vivo* mutagenic assays (except some experiments in plants) are not surprising. (*See* **Mutagens**.)

According to experimental data together with certain epidemiological evidence, it can be concluded that aluminum and its salts do not constitute a carcinogenic or mutagenic hazard except, perhaps, in cases of extremely high exposure.

Miscellaneous

No significant aluminum toxicity was observed when aluminum nitrate nonahydrate was given orally to rats at 375, 750 or 1500 mg per kg of body weight per day for 1 month or over a period of 100 days at 360, 720, and 3600 mg per kg body weight per day. However, it has been suggested that oral aluminum exposure may both stimulate and suppress the immune system, depending on the exposure level and the physiological state of the animal and the immune response assessed. Recent studies in rats have shown a slight stimulation of some immune functions in the rats at aluminum plasma concentrations normally found in healthy humans ($< 10\,\text{ng ml}^{-1}$).

Aluminum Toxicity in Humans

Dialysis Encephalopathy Syndrome

Although the possible human toxicity of aluminum has been a matter of controversy for over 100 years, the spectrum of clinical aluminum toxicity has expanded gradually since its initial description as the cause of dialysis dementia. In the 1970s, Alfrey *et al*. suggested that dialysis encephalopathy, previously a uniformly fatal neurological syndrome, resulted from aluminum intoxication. Based on additional biochemical and epidemiological data, this supposition was confirmed, and now it is quite accepted that aluminum is the cause of the dialysis encephalopathy syndrome, the main neurotoxic condition of aluminum. This syndrome includes a mild or severe speech disorder, hallucinations, twitching, myoclonus, seizures, mental changes and typical electroencephalogram abnormalities. Moreover, there is little doubt that aluminum can also cause bone disease and anemia in dialysis patients. Individuals with dialysis osteomalacia typically present with severe symptoms of axial skeletal pain, proximal muscular weakness, and fractures of the ribs, among other symptoms. In turn, aluminum has also been implicated as the cause of a microcytic, hypochromic anemia in patients with chronic renal failure. These disorders can result from introduction of aluminum directly into the blood stream via high aluminum-dialysate or consumption of large oral doses of aluminum-containing phosphate binders. The lack of urine production, which is the major route for aluminum excretion, contributes to the problems in these patients. While the health threat from diffusate fluids was significantly reduced by the recommendation that the diffusate contains less than 10 μg of aluminum per liter, ingesting aluminum-containing phosphate binders is still a matter of concern for uremic patients.

On the other hand, several reports have described aluminum accumulation and toxicity in individuals without chronic renal failure but with iatrogenic exposure to aluminum. These include preterm infants, largely fed intravenously, patients on total parenteral nutrition, subjects with severe burns, individuals receiving alum irrigation in the urinary bladder to prevent bleeding, and patients undergoing cranial bone reconstruction with aluminum containing bone cement.

Occupational Exposure

Occupational exposure to aluminum occurs not only in the primary refining of this metal but also in user industries (e.g., aircraft, automotive, metal products). Aluminum welding is also an important exposure vehicle. Although there are only a few studies on the influence of occupational exposure to aluminum on a worker's cognitive performance, a diverse range of cognitive deficits, including visuomotor or visuospatial problems, attention deficits, impaired verbal or visual memory and learning, and problems with 'concept formation' have been reported. Recent investigations have also corroborated that exposure to aluminum is associated with detrimental effects on certain cognitive functions, with deficits in memory attention, and visuomotor and visuospatial abilities.

For biological monitoring, urinary aluminum appears to be a more sensitive index to current exposure than serum levels. Workers with occupational exposure to aluminum have shown statistically significant increases in urinary aluminum/creatinine ratios in association with borderline changes in serum aluminum values. The mean aluminum concentrations in a healthy nonexposed population who

did not use antacid drugs were 0.06 μmol l^{-1} (range 0.02–0.13 μmol l^{-1}) in serum, and 0.33 μmol l^{-1} (range 0.07–0.82 μmol l^{-1}) in urine, with the upper reference limit for aluminum being 0.1 μmol l^{-1} in serum and 0.6 μmol l^{-1} in urine.

Aluminum and Alzheimer's Disease

Although there are still many unanswered questions, there are strong indications that aluminum may be one of the factors resulting in Alzheimer's disease and perhaps other neurodegenerative disorders, including amyotrophic lateral sclerosis. Aluminum has been detected in both senile plaques and neurofibrillary tangle-bearing neurons in the brains of people with Alzheimer's disease. Several epidemiological studies have shown increased risks for Alzheimer's disease in populations exposed to elevated aluminum concentrations in public drinking water supplies. However, a recent study indicated that any relation at aluminum levels below 0.2 mg l^{-1} is weak. Other independent lines of evidence supporting the hypothesis that aluminum might play a role in the pathogenesis of Alzheimer's disease are: (1) studies on the effects of intracranial aluminum on the cognitive and memory performance of animals; (2) data showing that in laboratory experiments aluminum aggregates Aβ, phosphorylates some sites on tau, the principal subunit of Alzheimer's disease tangles and assembles, *in vitro*, Alzheimer's disease hyperphosphorylated tau into paired helical filaments which are indistinguishable from those of Alzheimer's disease; and (3) the slowing of the progress of the disease with the use of desferrioxamine (or deferoxamine), an aluminum chelating agent that removes aluminum from the body.

Almost certainly Alzheimer's disease can result from many factors, genetic and environmental, or a combination of both. Although the participation of aluminum cannot be discarded, it seems very unlikely that a single environmental factor such as aluminum can constitute a sufficient explanation. A possible interpretation is that aluminum, which is definitely neurotoxic, may have a dementing effect independent of the pathological processes associated with Alzheimer's disease. In contrast to genetic risk factors for which preventive measures cannot yet be instituted in the foreseeable future, human exposure to aluminum, especially in the aging population, could probably be reduced. It would, perhaps, contribute to reducing the incidence of at least some forms of dementia.

Treatment of Aluminum Overload

Reduction of aluminum accumulation and toxicity may benefit a variety of patient populations which have different treatment needs. Prevention is the most important issue to avoid aluminum overload. Because of the propensity of uremic patients to retain aluminum, prevention is of special interest for these subjects. In relation to it, a careful surveillance of the diet of these patients (citrate, ascorbate, lactate, etc.) is recommended. Moreover, sodium bicarbonate and calcium carbonate or calcium acetate are adequate substitutes for Shol's solution and calcium citrate, respectively.

Aluminum mobilization and removal are essential in patients with aluminum toxicity. Currently, the most effective treatment to remove aluminum from the body is parenteral administration of desferrioxamine, an aluminum chelator. In recent years, a number of experimental studies have shown that oral administration of the chelating agent deferiprone (L1) could be an alternative drug to desferrioxamine for aluminum mobilization. However, the potential use of deferiprone for aluminum removal in humans is still controversial.

See also: **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; **Emulsifiers**: Uses in Processed Foods; **Leavening Agents**; **Mutagens**

Further Reading

Alfrey AC, LeGendre GR and Kaenhy WD (1976) The dialysis encephalopathy syndrome. Possible aluminium intoxication. *New England Journal of Medicine* 294: 184–188.

Chadwick DJ and Whelan J (1992) *Aluminium in Biology and Medicine*. Ciba Foundation Symposium. Chichester: John Wiley.

Colomina MT, Sanchez DJ, Sanchez-Turet M and Domingo JL (1999) Behavioral effects of aluminum in mice: influence of restraint stress. *Neuropsychobiology* 40: 142–149.

Domingo JL (1995) Reproductive and developmental toxicity of aluminium: a review. *Neurotoxicology and Teratology* 17: 515–521.

Domingo JL (1996) Adverse effects of aluminium-chelating compounds for clinical use. *Adverse Drug Reactions and Toxicological Reviews* 15: 145–165.

Gomez M, Sanchez DJ, Llobet JM, Corbella J and Domingo JL (1997) The effect of age on aluminium retention in rats. *Toxicology* 116: 1–8.

McLachlan DRC (1995) Aluminium and the risk for Alzheimer's disease. *Environmentrics* 6: 233–275.

Nieboer E, Gibson BL, Oxman AD and Kramer JR (1995) Health effects of aluminium: a critical review with emphasis on aluminium in drinking water. *Environmental Reviews* 3: 29–81.

van Landeghem GF, de Broe ME and D'Haese PC (1998) Al and Si: their speciation, distribution and toxicity. *Clinical Biochemistry* 31: 385–397.

Yokel RA and Golub MS (1997) *Research Issues in Aluminium Toxicity*. Washington: Taylor & Francis.

Zafar TA, Weaver CM, Martin BR, Flarend R and Elmore D (1997) Aluminium (^{26}Al) metabolism in rats. *Proceedings of the Society for Experimental Biology and Medicine* 216: 81–85.

Zatta PF and Alfrey AC (1997) *Aluminium Toxicity in Infant's Health and Disease*. Singapore: World Scientific.

AMARANTH

R Bressani, Instituto de Investígaciones, Universidad del Valle de Guatemala, Guatemala

Background

Amaranth, a legacy of the Aztecs, Mayas, and Incas, continues to be an underexploited plant with a promising economic value due to the variety of uses it can have and the benefits it can provide to producers, processors, and consumers. Present interests have developed because the plant offers leaves of a high nutritional quality when used as a vegetable, because the grains have a high protein content and quality, and because the whole plant offers a highly acceptable forage. The amaranth plant is also attractive since it adapts itself to a large number of environments, grows with vigor, produces large amounts of biomass, and resists drought, heat, and pests. Less attractive features of the plant include leaves that tend to accumulate nitrate and oxalates, and a very small grain size, which necessitates a large labor force when harvesting. With only a relative small research effort, a strong knowledge base is being constructed and is beginning to reveal the economic value and nutritional benefits of this ancient legacy crop.

Origin and Distribution

The Amaranthacea comprise over 60 genera, which include around 800 species of dicotyledonous, herbaceous plants, of either annual or perennial growth. There are three species of the genus *Amaranthus* that produce relatively large inflorescences with often more than 50 000 edible seeds per plant. These are *A. hypochondriacus* from Mexico, *A. cruentus* from Guatemala and *A. caudatus* from Peru and other Andean countries. Vegetable amaranths grow very well in the hot, humid regions of Africa, South-east Asia, Southern China, and India; they are represented by various amaranth species, such as *A. tricolor*, *A. dubius*, *A. cruentus*, *A. edulis*, *A. retroflexus*, *A. viridis*, and *A. hybridus*. Grain amaranth was an important crop for the preHispanic, New World civilizations. Its presence goes back some 4000 years BC in the Tehuacan Valley in Mexico, also the most likely site for the origin of maize. Its use is said to have been highly associated with religious festivities, which were forbidden by the Spanish conquerors and resulted in the elimination of the crop. Its production declined to small and insignificant levels, but it did not disappear. From Mesoamerica and the Andean region, grain amaranth was apparently carried as a weed, ornamental, or grain to other parts of the world. At present, there are three major germ plasm collections: (1) USDA Plant Introduction Center, Ames, Iowa, US; (2) Universidad del Cuzco, Cuzco, Peru; (3) National Bureau of Plant Genetic Resources, Shimla, India.

Classification

Amaranth is a dicotyledonous plant, not a grass like most cereals, which are monocotyledonous; rather, it is a pseudocereal. Amaranth shows an extreme botanical plasticity in adaptation that contributes to difficulties and confusion in its taxonomy. It is classified by means of flower structure, the form and proportion of leaves and inflorescence. These highly variable characters also allow for a high hybridization frequency, and assure the availability of a large germ plasm reserve.

Plant Description

Amaranths are broad-leafed plants that can grow to about 1.5–3.0 m high. The plant has a variable growth pattern in the type and number of branches, all of which end with a small seed head, sometimes maturing at the same time as the main seed head. Leaf shape varies from lanceolate to ovate to elliptic. Leaf number and size show great variability within and between species. The leaves, stems, and flowers of amaranth can be green, gold, reddish purple, or various shades. The seed head, some 50–60 cm high, varies from pendulous to spiked, and seed color is usually black, gold, or cream.

Amaranths, like many cereals and grasses, carry out photosynthesis by the C_4 pathway, which uses carbon dioxide very efficiently by fixing it in the chloroplasts of specialized cells surrounding the leaf vascular bundles. This characteristic is also responsible for lower water losses by transpiration in association with the stomata. C_4 plants also have the capacity to photosynthesize at high rates at high temperatures, and through osmotic adjustment, the amaranth plants can tolerate some degree of lack of water without wilting.

Grain Description

It has been found that the main amaranth inflorescence may produce 49–89 g of grain (50 000–100 000 seeds). The grain is very small, lenticular in shape, 1–1.5 mm in diameter, and weighs 0.6–1.3 mg per seed, with the structure shown in Figure 1. The germ, which is circular in shape, encircles the perisperm in one plane, and represents, together with the seed coat, around 25–26% of the weight of the seed, as obtained by milling techniques. This fraction is relatively rich in fat and protein. The perisperm represents 68–78% of the weight of the grain. In cereal grains, such as maize, the seed coat plus germ represent around 16–18% of the weight of the kernel, with the endosperm representing about 65%.

Production

Grain

Production depends on many variables, such as agronomic practices, plant density per hectare, method of harvest in which seed shattering is important, uniformity of grain maturation, and the type of seed head. Yields in some localities have ranged from 1.1 to 1.5 t ha^{-1}, but in Mexico, some reports indicate yields of 3.0–5.9 t ha^{-1} Annual production has been estimated at 470×10^6 t per year world-wide. High yields can be obtained through high plant population densities (320 000–360 000 plants per hectare), as well as from the appropriate application of organic and inorganic fertilizers. To increase yield, breeding objectives include the development of early maturing varieties, with a reduced plant height and resistance to lodging with insect and disease tolerance and no seed shattering. Likewise, the size of the seed should be increased.

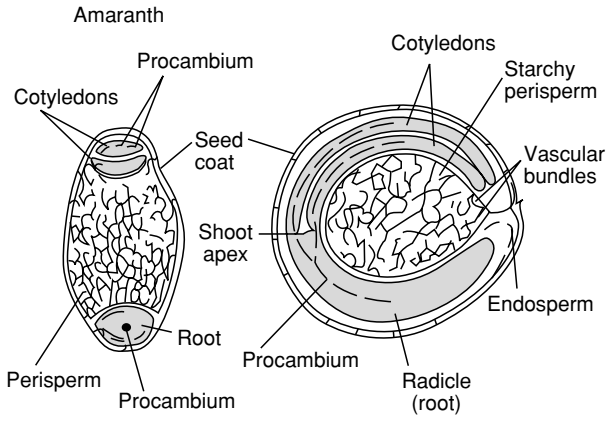

Figure 1 *A. cruentus* seed in (a) cross- and (b) longitudinal sections as viewed in a light microscope, Source: Irving DW, Betschart AA and Saunders RM (1981) Morphological studies on *Amaranthus cruentus*. *Journal of Food Science* 46(4): 1170–1174.

Whole Plant

With respect to vegetable production, yields between 1.4 and 11 t of dry matter per hectare at 4 weeks have been reported. In one study, dry matter yields at 25, 40, and 60 days after emergence were 66.6, 681.8, and 3452.0 kg ha^{-1}, yielding 9.7, 154.3, and 514.7 kg of protein per hectare. Yields of edible leaves are obviously much lower, with values varying from 0.30 to 0.70 t per hectare of dry matter. From a study on 14 varieties, the average dry weight of the whole plant prior to seed harvest, was 208.5 g, of which 40% was represented by the weight of the stems, 26% by the inflorescence, and 34% by the seed.

The amaranth plant may be cultivated as an intercrop with maize to provide a good-quality forage fed fresh or ensilaged to swine and/or to cattle. Alone or combined, it can be dehydrated or pelletized to be used as feed for farm animals.

Vegetable Amaranth

The use of amaranth leaves as a vegetable has been well established in a number of countries. *A. tricolor* is the main vegetable species in Asia, whereas *A. hybridus* and *A. cruentus* predominate in Africa and Latin America. Today, there is probably more amaranth consumed as a vegetable than as grain. The young leaves and the softest sections of the plant are boiled, sometimes in several changes of water, and consumed often mixed with cooked tomatoes and other condiments. Usual forms of consumption include soups, salads, boiled, made into a purée with tomatoes and consumed with cooked root crops or lime-cooked tortillas. The flavor of cooked amaranth leaves and young stems is pleasantly mild and not unlike that of similar greens.

Chemical Composition

The chemical composition of the vegetative part of amaranth is variable, since the content of organic and

Table 1 Chemical composition of amaranth (grams per 100 g)[a]

	Vegetable	*Grain*	*Forage*
Moisture	85.0 ± 4.4	9.9 ± 2.0	87.8 ± 0.76
Dry matter	15.0 ± 4.4	90.1 ± 2.0	12.2 ± 0.76
Protein (N×6.25)	24.1 ± 4.2	15.2 ± 1.7	19.2 ± 5.6
Total lipids	3.8 ± 0.68	7.0 ± 1.6	2.9 ± 1.3
Crude fiber	14.9 ± 3.7	6.2 ± 3.2	16.6 ± 6.2
Dietary fiber		13.6 ± 4.8	
Ash	17.7 ± 1.6	3.3 ± 0.5	19.0 ± 3.9
Carbohydrate	42.9 ± 4.6	62.1 ± 7.6	43.8 ± 8.6
Amylose		6.1 ± 1.2	
Energy (calories per 100 g)	284	366	337
Metabolic energy (kJ^{-1})		12.2	
Nitrate	0.55 ± 0.19		
Oxalate	4.5 ± 1.8		5.86 ± 1.89
Phytate		1.03 ± 1.16	
Tannins		0.18 ± 0.14	
Cell walls			63.5 ± 7.7
Neutral detergent fiber			43.4 ± 10.9
Acid detergent fiber			34.6 ± 15.0
Acid detergent lignin			5.2
Cellulose			23.4
In vitro digestion (%)			60.0 ± 4.1

[a]Data, except moisture are given on a dry-weight basis. Moisture content varies widely depending on processing.

Table 2 Mineral content of amaranth

Mineral	*Vegetable (g per 100 g[a])*	*Grain (mg per 100 g[a])*	*Forage (g per 100 g[a])*
Phosphorus	0.66 ± 0.22	578 ± 38.9	0.70 ± 0.28
Potassium	0.19 ± 0.02	541 ± 80.7	
Calcium	2.57 ± 0.62	212 ± 66.0	2.45 ± 0.76
Magnesium	1.10 ± 0.30	327 ± 42.8	
Sodium		22.5 ± 6.9	
Iron	0.19 ± 0.15	35.6 ± 32.9	
Copper		1.81 ± 1.06	
Manganese		3.42 ± 0.90	
Zinc		3.83 ± 0.38	
Sulfur		150	
Aluminum	0.015	1.0	
Selenium	0.003		

[a]dwb, dry-weight basis.

Table 3 Vitamin content of amaranth (mg per 100 g[a])

Vitamin	*Vegetable*	*Grain*
Thiamine	0.68 ± 0.32	0.136 ± 0.076
Riboflavin	2.24 ± 0.43	0.223 ± 0.056
Nicotinic acid	7.47 ± 1.66	1.153 ± 0.161
Biotin		42.5 ± 1.5
Folic acid		43.8 ± 1.5
Vitamin C	570.7 ± 89.6	4.47
Carotene	33.3 ± 16.68	4.6

[a]dwb, dry-weight basis.

inorganic compounds changes as a result of climate, plant nutrition, agricultural practices, physiological stage, species and cultivar type, and position of leaves taken for analysis (Table 1). The moisture content for all species varied from about 70 to 94%, or from 6 to 30% dry matter. On a dry-weight basis (dwb), protein varies from 18 to 38% and total crude lipids from 1.3 to 10.6%. The lipid fraction contains 53.6% non-polar lipids, 33.8% glycolipids, and 12.6% phospholipids. The main fatty acids include linolenic (30–70%) and palmitic (13–14%) as well as oleic and stearic acid (see Table 5).

The variability in crude fiber is also quite large, from 5.4 to 24.6% (dwb). This variability can be explained on the basis of the age of the plant, with old leaves and shoots showing the highest values. Finally, the ash content ranges from 7.6 to 22.2%.

Minerals and Vitamins

The levels of calcium, potassium, and magnesium are quite high, with calcium present in the largest amount (2–3 g%; Table 2). Large quantities of phosphorus are also found, with values ranging from 0.49 to 0.79 g%. There is a relatively high concentration of iron, varying from 0.08 to 0.50 g% (dwb). Other minerals have been reported in smaller amounts.

The amounts of vitamins B_1 and B_2, nicotinic acid, and ascorbic acid are similar to those found in other greens, but the content of β-carotene is as high as, or higher than, that reported for other green vegetables (Table 3).

Other Compounds

Amaranth, like many other fast-growing plants, has a tendency to accumulate nitrates. Reported values range from 0.27 to 0.74% nitrate (dwb) (0.046–0.104% fresh weight; Table 1). There is a significant variation among cultivars, also due to environmental factors, since the same cultivar grown over different years has different values. The stems contain about twice as much nitrate as the leaves on a fresh-weight basis (fwb). The levels of nitrate in amaranth leaves are similar to those found in spinach and chard. Nitrogen fertilizer increases nitrate accumulation in green vegetables, including amaranth. Nitrates seldom cause problems to consumers because of the relatively low concentration and because, upon cooking, they leach out of the vegetable.

Another compound of greater concern is oxalate, which accumulates in leafy vegetables such as amaranths, particularly during dry weather. Oxalate levels are nutritionally important because they can bind essential divalent minerals, particularly calcium,

Table 4 Essential amino acid content of amaranth (milligrams of amino acid per gram of nitrogen)

Essential amino acid	*Vegetable*	*Grain*
Leucine	421 ± 76.5	356 ± 17.6
Isoleucine	327 ± 38.7	226 ± 15.9
Lysine	300 ± 34.9	351 ± 21.4
Methionine	78 ± 43.6	124 ± 17.7
Cysteine	24 ± 8.1	124 ± 6.4
Phenylalanine	314 ± 78.0	270 ± 40.1
Tyrosine	232 ± 48.7	220 ± 32.6
Threonine	356 ± 75.9	238 ± 21.3
Tryptophan	80	76 ± 12.3
Valine	355 ± 31.4	256 ± 14.1

making it nutritionally unavailable. On a dry-weight basis, the oxalate in amaranth ranges from 1.1 to 7.9% (0.20–1.02% fwb); the stems contain significantly lower amounts than the leaves. The variability is a result of both the genetic make-up of the plant and environmental factors. The oxalate in amaranth is bound to sodium, potassium, calcium, and magnesium, but about 40% is free and therefore capable of binding other minerals. During processing in boiling water, oxalates are removed when water is changed, so that an intake of 100 g of cooked amaranth would provide approximately 0.2–1.0 g of oxalic acid. Assuming that 40% of oxalate is free and lost during cooking, the level drops to 0.12–0.6 g, which is nontoxic. Toxic levels for humans have been indicated to be 2–5 g per day. For populations consuming low levels of calcium, some of this mineral would become unavailable because of its complexing affinity to oxalic acid.

Other Uses

Amaranth leaves have been used in the preparation of leaf protein concentrates. These represent 24–36% of the leaf protein and contain 61–80% protein, which is of a light color, almost odorless, and nonbitter. Their essential amino acid content is attractive and high in lysine, but the sulfur amino acid content is low (Table 4).

Nutritional Value

The chemical composition of amaranth leaves shows that they provide a relatively good nutritional value. The iron and β-carotene contents are highly available. Not much information is available on the quality of the protein in amaranth leaves, but they make a very good supplement to maize and rice, either alone or with common beans at levels of up to 5% (dwb), because of their relatively high content of lysine, an amino acid that is lacking in both cereal grains (Figure 2). At this level of supplementation, amaranth leaves can provide substantial levels of vitamins and minerals to cereal/bean diets.

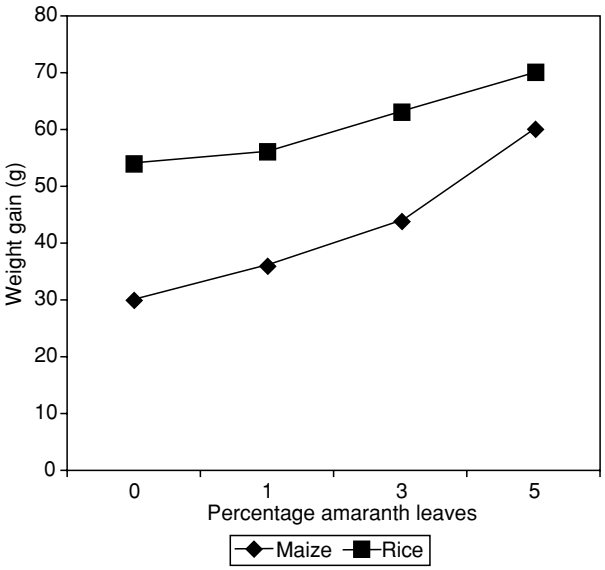

Figure 2 Supplementary effect of amaranth leaves to maize and rice.

Grain Amaranth

Chemical Composition

The gross chemical composition of amaranth grain is shown in Table 1. The crude protein content ranges from 11.8 to 17.6%. The reasons for the large variation amongst the different species are not known. While it may be of a genetic nature, it may also be due to environmental conditions and cultural practices. Nitrogen fertilization of amaranth has not consistently increased the crude protein in the grain.

The variability in total lipid content for all species is 4.8–8.1%. Nutritionally, a high lipid content is of interest for the high energy that grain amaranth can provide in comparison with cereal grains and food legumes.

Dietary fiber data for grain amaranth are lacking. Values for *A. caudatus* have been reported to vary from 7.6 to 16.4%.

The ash concentration is relatively constant among species, and the variability for each element is also small.

Vitamin contents appear to be relatively constant among species. Some of the vitamins, e.g., nicotinic acid, are low in comparison with the nicotinic acid content in cereal grains, which show values about two to three times higher.

Starch is the most abundant chemical component in amaranth grain; the content varies from 48 to 69%. It occurs as very small polygonal or spheric granules (1–3 μm in diameter), and is readily digested by α-amylases. The starch is present mainly as highly branched amylopectin, but some varieties of amaranth grain contain 4.8–7.22% of amylose, a linear starch. Other carbohydrates include sucrose, raffinose, stachyose, and maltose in small but variable amounts. Microgranule starches, like those in amaranth, are being studied for paper coatings, fat substitutes, and biodegradable applications.

Fatty Acids

Amaranth seed oil is rich in the essential fatty acid, linoleic acid, with values ranging from 43.4% in *A. cruentus* to 51.4% in *A. hypochondriacus*. Oleic acid content is second in concentration, with values of 21.3% for *A. hybridus* and 31.9% for *A. cruentus*. Amaranth oil contains 18.6–21.3% palmitic acid. The reported variability suggests differences between species and also differences between varieties of the same species. Other lipid classes include sterols, of which spinasterol is found at around 0.20% of the crude oil. In the unsaponifiable fraction, the isoprenoid, squalene, is found in relatively large amounts, around 6.8%. The total unsaturated fatty acid content is around 77% for amaranth oil (Table 5). (*See* **Fatty Acids**: Dietary Importance.)

Protein

Protein distribution in the grain The endosperm of the grain contains 35.0% of the total grain protein, while the remainder is in the seed coat and germ or embryo. This distribution is quite different from that found in maize, sorghum and rice, wherein the germ contributes between 12.5 and 18.5%, and the endosperm between 81.5 and 87.5% of the total protein of the grain. The physical distribution of the protein in amaranth is responsible for its higher total protein content, as compared with common cereal grains, since the protein concentration in the germ is higher than the concentration in the endospem. (*See* **Protein**: Requirements.)

Table 5 Fatty acid content of amaranth (grams per 100 g)

Fatty acid	*Vegetable*	*Grain*
Myristic acid 14:0	1.1	0.53 ± 0.3
Palmitic acid 16:0	13.3	20.4 ± 2.0
Stearic acid 18:0	4.6	3.5 ± 1.5
Oleic acid 18:1	4.7	26.1 ± 5.6
Linoleic acid 18:2	6.4	48.0 ± 7.3
Linolenic acid 18:3	34.4	1.43 ± 0.8
Arachidonic acid 20:0	0.5	1.6 ± 0.9

Protein fractions in amaranth A few studies have been carried out on amaranth protein fractions. The quantity of albumin was found to vary between 19 and 23%, and that of the globulins between 18 and 21%. The alcohol-soluble prolamines have been found to vary from 2 to 3%, and alkali-soluble glutelin-like proteins from 42 to 46%. Values of around 5–14 g per 100 g of protein have been reported for nonprotein nitrogen. The data available suggest that there are no great differences in the quantity of protein fractions among species or among cultivars. Recently, a purified albumin fraction, 35-kDa monomer, is being used to insert it into potato and rice protein by biotechnological techniques because this albumin fraction has an excellent essential amino acid pattern.

Amino Acids

Table 4 presents average values for all species. Comparison of the essential amino acid content in grain amaranth with the Food and Agriculture Organization/World Health Organization (FAO/WHO) 1973 reference pattern reveals that the most deficient amino acid is leucine. Nevertheless, other amino acids such as valine, isoleucine, and threonine could be limiting. Biological tests, however, have revealed that threonine, rather than leucine, is the most limiting essential amino acid. Amaranth protein is a good source of lysine, tryptophan, and sulfur amino acids. In contrast, cereal grains are deficient in lysine. The essential amino acid balance of grain amaranth protein is significantly better than that of many other proteins of vegetable origin. Since germ proteins are richer sources of essential amino acids (as compared with the storage proteins of the endosperm), the higher proportion of germ proteins in amaranth may explain the higher concentration of lysine in amaranth grain than that in cereal grains. (*See* **Amino Acids**: Properties and Occurrence.)

Antinutritional Factors

As indicated above, amaranth is consumed as a vegetable and as a grain. There are various reports that indicate that both food forms contain antinutritional factors or toxic compounds that could interfere with food uses and limit nutrient utilization. The toxic factors in vegetable and grain amaranth include saponins and phenolic compounds, phytic acid, oxalates and nitrates, protease inhibitors, lectins, alkaloids, and polyphenolics. Appropriate processing, however, has been shown to reduce or eliminate such compounds. (*See* **Plant Antinutritional Factors**: Characteristics.)

Processing

Amaranth grain has been processed by a relatively large number of techniques such as wet and dry cooking, which include atmospheric and pressure cooking, nixtamalization, drum drying, extrusion cooking, and puffing; by biological approaches such as germination, malting, and enzymatic hydrolysis, and by physical means, which include milling, physical fractionation, and starch and protein extraction.

Nutritive Value

Oil

The digestibility of amaranth crude oil varies between 91.7 and 94.1% at the 5% level of addition and between 91.1 and 93.8% at the 10% level. These results are significantly lower than those from refined cottonseed oil at both levels of addition. The lower digestibility is probably caused by the presence of sterols (0.37% total) and unsaponifiable matter (1.4%) in crude amaranth oil; however, no deleterious effects from its consumption have been reported.

Metabolizable Energy of Amaranth Grain

The two main sources of energy in amaranth grain are the carbohydrate fraction and the oil content. The metabolizable energy (ME) of light- and dark-colored *A. cruentus* grain is 11.8 and 11.72 $kJ\,g^{-1}$ (2.81 and 2.79 $kcal\,g^{-1}$), respectively. Heat processing increases the value for both seed colors, with extrusion cooking giving the highest values of 17.72 and 14.11 $kJ\,g^{-1}$ (4.22 and 3.36 $kcal\,g^{-1}$), respectively.

Protein

A number of researchers have uncovered an important conclusion: the protein value (protein efficiency ratio) of the raw amaranth grain does not reflect the amino acid pattern of the protein. Furthermore, amaranth grain processed by wet cooking gives a higher protein quality value than that of the raw grain in all amaranth species. When processed under conditions that do not damage the availability of essential amino acids, its protein quality is very close to that of casein or milk protein. The effect of processing is evident in both the consumption of the diet and the weight gains of test animals. This effect is still unexplained and deserves some research. It has been suggested that the presence of antiphysiological factors that are inactivated by heat may be responsible; well-known antigrowth factors such as trypsin inhibitors, tannins, and lectins are present at low levels in amaranth grain.

Some processes, such as expansion, flaking, and wet cooking, apparently do not affect protein digestibility. However, the product from toasting gives an equal or lower protein digestibility than the raw grain. Table 6 lists the protein quality values of amaranth grain protein and that of other cereal grains. With the exception of quality protein maize, also high in lysine, all other cereal grains have a protein quality below that of amaranth grain.

Table 6 Protein quality of amaranth and various common cereal grains

Grain	*Condition*	*Protein quality (percentage of casein)*
A. cruentus	Raw	74.8 (NPU)
A. cruentus	Popped	74.4 (NPU)
A. caudatus	Drum-dried	89.1
A. cruentus	Drum-dried	86.1
A. hypochondriacus	Drum-dried	85.0
Pearl millet	Raw	55.8
Finger millet	Raw	50.0
Teff	Raw	56.2
Maize	Raw	45.0
Sorghum	Lime-cooked	30.6
QPM	Raw	72.6
QPM	Raw	82.9
Rice	Lime-cooked	60.0
Wheat flour	Raw	26.0
Barley, common	Raw	53.2
Barley, Hiproly	Raw	70.8
QPM	Raw	70.0
Whole wheat	Raw	37.2
Oats	Raw	73.6
Rye	Raw	52.0
Opaque-2 maize	Raw	96.9
Opaque-2 maize	Lime-cooked	92.4

NPU, net protein utilization.

Table 7 Minimum nitrogen intake to obtain nitrogen equilibrium from various food sources

Protein source	*Nitrogen intake*[a]	*Nitrogen absorbed*[a]
Cheese	71	49
Extruded amaranth	80	51
Popped amaranth	84	52
Milk	75–86	48–60
Casein	94	64
Texturized soy protein	97	70
Soybean isolate	87	62
Meat	85	64
Soy/meat	92	68
Maize/beans	97	67
Beans	116	71
Rice/beans	96	
Plantain/beans	112	

[a]Milligrams of nitrogen per kilogram body weight per day.
From Bressani R *et al.* (1993) *Plant Foods for Human Nutrition* 43: 123.

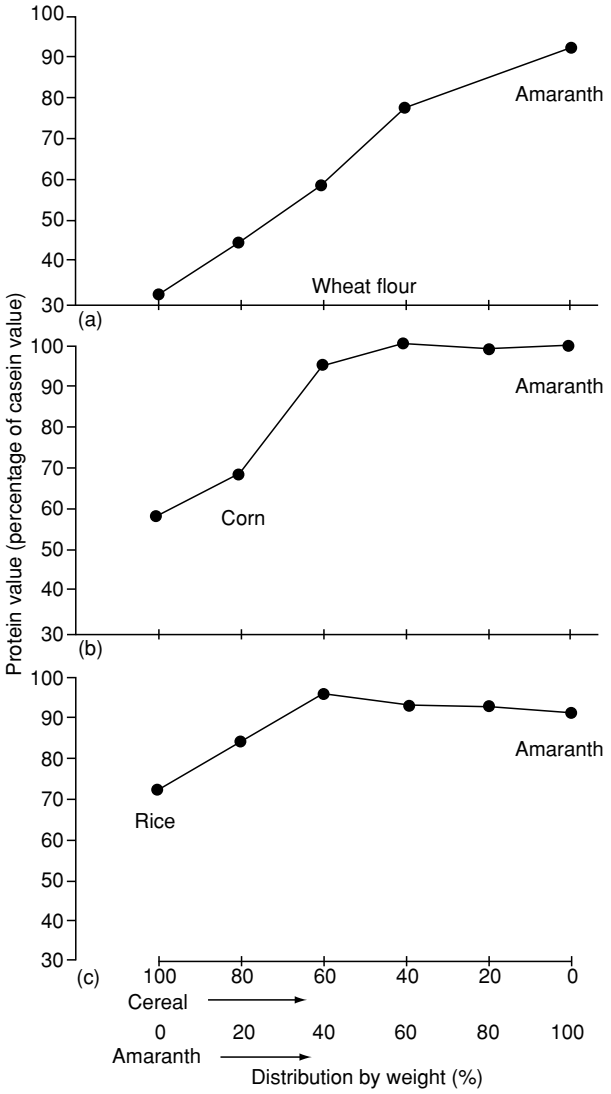

Figure 3 Complementation between amaranth and cereal protein. (a) Wheat flour plus processed amaranth, 10% protein in diet; (b) corn flour plus processed amaranth, 9.0% protein in diet; (c) rice flour plus processed amaranth, 8% protein in diet. Source: Bressani R (1989) The proteins of grain amaranth. *Food Reviews International* 5: 13–38.

Although human studies on grain amaranth protein quality have been conducted, data are not readily available. Some results with adult human subjects indicate that extrusion-cooked grain amaranth has a protein value of 89% relative to the value of cheese, whereas the popped material has a value of 81% of the value of cheese. Digestibility values follow the same order. Table 7 summarizes the protein quality of various protein sources as evaluated in humans. The quality is expressed as the amount of nitrogen ingested from the food protein source needed to obtain nitrogen equilibrium in nitrogen-balance studies. A smaller value indicates a higher protein quality.

In studies with rats, amaranth containing diets reduced cholesterol levels in blood as well as in the liver. In the colon, amaranth diets induced similar changes to those from cellulose feeding, but did not reduce the pH and did not increase cecum surface area. These observations suggest that the soluble and insoluble fiber of grain amaranth behave as that is these fractions of other cereal grains.

Supplementary Value

As a result of its high lysine content, the addition of amaranth grain to cereal flours improves the quality of the mixture (Figure 3). The available data show an increase in protein quality of 62% for wheat flour, 40% for maize, and 25% for rice, from the addition of 30% amaranth flour to 70% of each cereal grain. The high protein quality of amaranth means that it can be used alone or as a fortifier in cereal grain mixtures.

Grain Uses

Popped and parched amaranth are time-honored edible forms of whole amaranth. The methods of preparation have persisted for centuries in Mexico, Peru, and India. In recent research, grain amaranth flour has been tested as a supplement to wheat, maize, and rice flours for use in pastries, baking mixes, pasta products, tortillas, and similar products Likewise, amaranth grain flour has been tested as a component of high-quality protein foods and as a drink in blends with milk and other grain flours. The product possibilities are numerous, but very few have been commercialized. A total of 60 cereal grain foods containing small amounts of amaranth are marketed in the Americas. In Mexico, eight different amaranth grain flours are marketed, with the most popular product being alegria, a mixture of popped amaranth with sugar. Product and market development will be crucial to amaranth's growth in the food chain.

See also: **Amino Acids**: Properties and Occurrence; **Cereals**: Contribution to the Diet; **Fatty Acids**: Properties; **Protein**: Food Sources

Further Reading

Bressani R (1989) The proteins of grain amaranth. *Food Reviews International* 5: 13–38.

Danz RA and Lupton JR (1992) Physiological effects of rats fed amaranth containing diets. *Cereal Foods World* 37(7): 489–494.

Joshi BD and Rana RS (1991) *Grain Amaranths: the Future Food Crop*. Shimla, India: National Bureau of Plant Genetic Resources. Regional Station Phagli.

National Research Council (1984) *Amaranth. Modern Prospects for an Ancient Crop*. Washington, DC: National Academy Press.
Paredes-López O (ed.) (1994) *Amaranth. Biology, Chemistry and Technology*. Boca Raton, FL: CRC Press.
Paredes-López O, Barba de la Rosa AP, Hernández-López D and Carabez-Trejo A. *Amaranto. Características Alimentarias y Aprovechamiento Agroindustrial*. Irapuato, México: Instituto Politécnico Nacional.
Saunders RM and Becker R (1984) Amaranthus: a potential food and feed resource. In: Pomeranz Y (ed.) *Advances in Cereal Science and Technology*, vol. 6, p. 357. St. Paul, MN: American Association of Cereal Chemists.
Singhal RS and Kulkarni PR (1988) Review. Amaranths and underutilized resource. *International Journal of Food Science and Technology* 23: 125.
Teutonico RA and Knorr D (1985) Amaranth: composition, properties and applications of a rediscovered food crop. *Food Technology* 39: 49–60.
Raina A and Dutta A (1992) Molecular cloning of a gene encoding a seed-specific protein with nutritionally balanced amino acid composition from amaranthus. *Proceedings of the National Academy of Science* 89: 11774–11778.

AMINES

M B A Glória, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

Definition, Structure and Synthesis of Amines

It has long been known that certain amines fulfill a number of important metabolic and physiologic functions in living organisms. They are formed during normal metabolic processes and are, therefore, present in foods.

Bioactive or biologically active amines are aliphatic, cyclic, and heterocyclic organic bases of low molecular weight. Most of them have been named after their precursor amino acids, e.g., histamine originates from histidine, tyramine from tyrosine, tryptamine from tryptophan, and so on. However, the names cadaverine and putrescine are associated with decomposition and putrefaction, and spermine and spermidine are related to seminal fluids in which they were first found.

Bioactive amines can be classified on the basis of the number of amine groups, chemical structure, or biosynthesis. According to the number of amine groups, they can be mono-(tyramine, phenylethylamine), di-(histamine, serotonin, tryptamine, putrescine, cadaverine), or polyamines (spermine, spermidine, agmatine). Based on their chemical structure, amines can be aliphatic (putrescine, cadaverine, spermine, spermidine, agmatine), aromatic (tyramine, phenylethylamine) or heterocyclic (histamine, tryptamine, serotonin). They can also be indolamines (serotonin) and imidazolamines (histamine).

According to the biosynthetic pathway, amines are traditionally classified as natural or biogenic. Natural amines – spermine, spermidine, putrescine, and histamine – are formed during *de-novo* biosynthesis (they are produced *in situ* and as required from their precursors). Biogenic amines are formed by bacterial decarboxylation of free amino acids (**Figure 1**). Histamine can be either natural (stored in mast cells or basophils) or biogenic.

Prerequisites for the formation of biogenic amines in foods are the availability of free amino acids, presence of decarboxylase-positive microorganisms, and favorable conditions for bacterial growth and decarboxylase activity. Free amino acids occur as such in foods, but also may be released from proteins as a result of proteolytic activity. Decarboxylase-positive microorganisms may constitute part of the associated population of the food or may be introduced by contamination before, during, and after processing. Several studies have been performed to investigate the production of amines by microorganisms isolated from fish, cheese, meat products, vegetables and alcoholic beverages. Several species of *Escherichia*, *Enterobacter*, *Salmonella*, *Shigella*, and *Proteus* are active biogenic amine formers. Certain species of *Achromobacter*, *Lactobacillus*, *Leuconostoc*, *Pseudomonas*, *Pediococcus*, *Streptococcus*, *Micrococcus*, and *Propionibacterium* are also capable of amine production. In fermented foods, the applied starter cultures also may affect amine production.

Polyamine synthesis is a more complex process, although the first few steps also include decarboxylation reactions. Polyamines are synthesized from their natural precursors – ornithine or arginine and *S*-adenosyl-L-methionine – in different organisms, with putrescine as an obligate intermediate

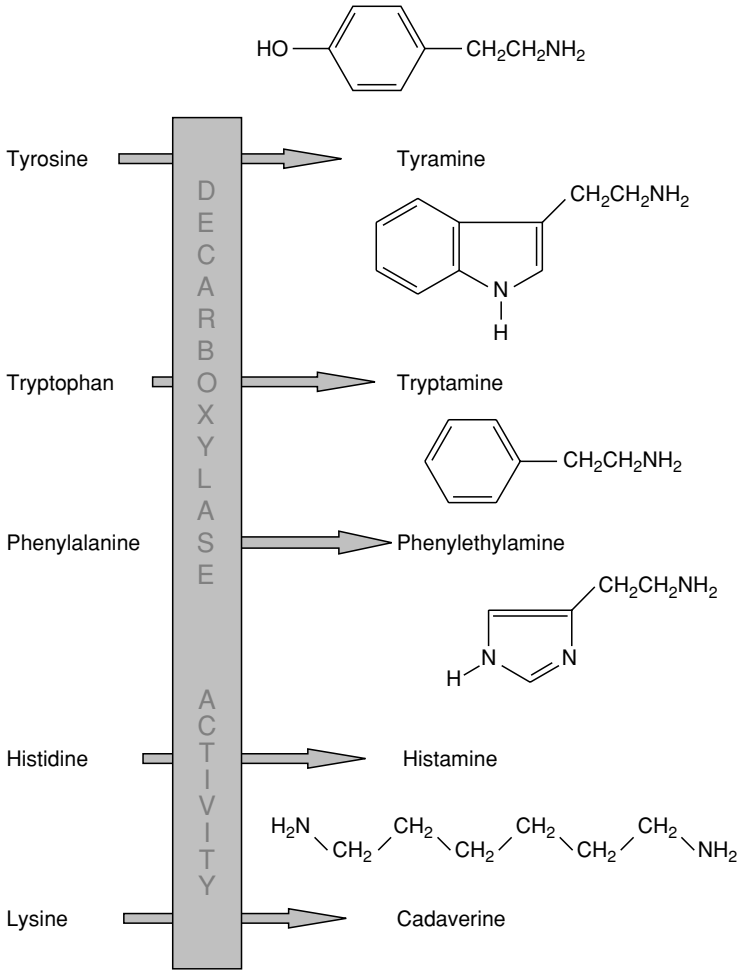

Figure 1 Synthesis of biogenic amines.

(Figure 2). In plants and some microorganisms, an alternative pathway exists to produce putrescine from arginine via agmatine (see dotted lines).

Physiological Importance of Amines

Amines participate in important metabolic and physiological functions in living organisms. Polyamines are essential for cell proliferation, growth, renewal, and metabolism. They are involved in nearly every step of DNA, RNA, and protein synthesis, and regulate the permeability and stability of cellular membranes. Polyamines play indispensable roles in living cells. Therefore, they are important in health and disease. Polyamine requirements are higher in the young during periods of intense growth. In healthy adults, however, these should not be as high, because polyamines are needed only to replace cells and mediate the action of hormones and growth factors. Polyamines are also essential for the maintenance of the high metabolic activity of the normally functioning and healthy gut, repairing the damage caused by deleterious components in food or by bacteria.

Biogenic amines are generally either psychoactive or vasoactive. Psychoactive amines, such as histamine and serotonin, act on the neural transmitters in the central nervous system. Vasoactive amines act directly or indirectly on the vascular system. Pressor amines – tyramine, tryptamine, and phenylethylamine – cause a rise in blood pressure by constricting the vascular system and increasing the heart rate and contraction force. Serotonin is vaso- and broncoconstrictor, reduces the volume and acidity of the gastric juice, has an antidiuretic effect, stimulates smooth muscle, and affects carbohydrate metabolism. Histamine can exert many responses within the body. It can directly stimulate the heart, cause contraction or relaxation of extravascular smooth muscle, stimulate both motor and sensorial neurons, and control gastric secretion. It also mediates primary and immediate symptoms in allergic responses.

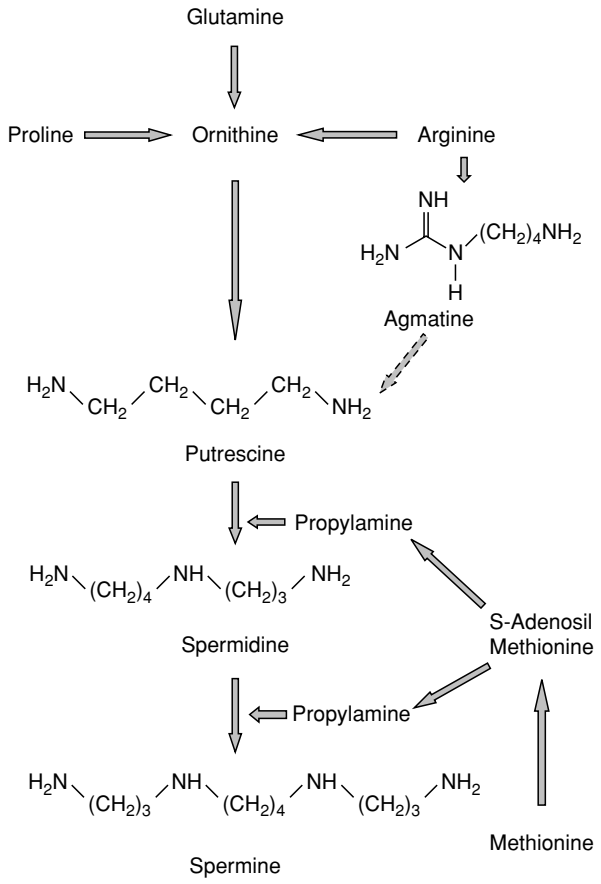

Figure 2 Pathways for the biosynthesis of polyamines.

Toxicology of biologically active amines

Three sources of amines have been identified: biosynthesis *in situ* from amino acids, direct ingestion from the diet, and synthesis and release by the bacterial flora resident in the gastrointestinal tract of the individual. It is now clear that the diet is an important source, supplying a significant part of the polyamines required to sustain normal metabolism.

Because of their biological activity, surplus amines need to be eliminated. Healthy individuals can detoxify amines present in foods by acetylation and oxidation. Biogenic amines are oxidized by monoaminoxidases (MAO; EC 1.4.3.4) and diaminoxidases (DAO; EC 1.4.3.6). Polyamines are usually acetylated first, then oxidized by polyaminoxidases (PAO; EC 1.5.3.11).

Amines in foods do not usually represent any health hazard to individuals, unless excessive amounts are ingested, or the natural mechanism for the catabolism of amines is genetically deficient or impaired by diseases. Individuals with respiratory and coronary problems, hypertension, or vitamin B_{12} deficiency are at risk because they are sensitive to lower amounts of amines. People with gastrointestinal problems (gastritis, irritable bowel syndrome, Crohn's disease, stomach and colonic ulcers) are also at risk because the activity of oxidases in their intestines is usually lower than that in healthy individuals. Patients taking medication that acts as inhibitors of MAO, DAO, and PAO activity also can be affected, as such drugs prevent the catabolism of amines. These MAO and DAO inhibitors are used for the treatment of stress, depression, Alzheimer's and Parkinson's diseases, pulmonary tuberculosis, malaria, panic syndrome, and social phobia. The toxic effect of some amines also can be potentiated by the presence of ethanol and of amines including putrescine, cadaverine, tyramine, tryptamine, 2-phenylethylamine, agmatine, serotonin, spermidine, and spermine. Most of them act by inhibition of amine metabolizing enzymes, whereas spermine causes liberation of endogenous histamine in the intestine.

The toxic effects caused by amines are summarized in Table 1. The most frequent foodborne intoxication caused by amines involves histamine. Histamine intoxication is also referred to as 'scombroid poisoning' because of the association of this illness with the consumption of scombroid fish. The disease manifests several minutes to 3 h after ingestion of the histamine-containing food. At first, a flushing of the face and neck is usually observed, accompanied by a feeling of heat and general discomfort. Often, this is followed by an intense throbbing headache. Other symptoms may be: cardiac palpitations, dizziness, faintness, itching, rapid and weak pulse, and gastrointestinal complaints (abdominal cramps, nausea, diarrhea). In severe cases, bronchospasms, suffocation, and severe respiratory distress are reported. However, the illness usually has a mild character, and symptoms do not last long. In most cases, the illness is not recognized as such. It is also possible that the symptoms are falsely attributed to a food allergy.

In the literature, many cases of food intoxication have been attributed to high histamine levels in foods or to low histamine levels in food containing other amines that can potentiate the toxic effect of histamine. Foods implicated in outbreaks of histamine intoxication include scombroid (tuna, bonito, mackerel, skipjack, herring) and other fish (sardine, anchovy, mahi-mahi), cheese (Gouda, Swiss, Gruyère, Cheddar, Cheshire), sauerkraut, sausage, and red wines.

When foods containing high tyramine levels are ingested, large amounts of nonmetabolized tyramine can reach the blood stream. This causes release of noradrenaline from the sympathetic nervous system,

Table 1 Toxic effects of biologically active amines

Toxic effects	*Amines involved*	*Food associated*	*Symptoms*
Histamine intoxication or scombroid poisoning	Histamine (Toxic effect is potentiated by putrescine, cadaverine, spermine, tryptamine, tyramine, phenylethylamine, ethanol)	Scombroid fish: tuna, bonito, mackerel, skipjack, herring Other fish: sardines, anchovy, mahi-mahi Cheese: Swiss, Gouda, Cheddar, Gruyère, Cheshire Sauerkraut; sausage; wine	Gastrointestinal: nausea, vomiting, diarrhea, cramps Neurological: headache, palpitation, face and neck flushing, burning throat, itching, rapid and weak pulse, dizziness, faintness Hemodynamic: hypotension Cutaneous: rash, urticaria, edema, localized inflammation Severe cases: bronchospasms, suffocation, severe respiratory distress
Tyramine intoxication	Tyramine	Cheese Beer Wine	Headache, fever, increased blood pressure, vomiting, perspiration, pupils and palpebral tissue dilatation, salivation, lacrimation, increased respiration, palpitation, dyspnea
Cheese reaction[a] or hypertensive crisis (associated with patients under MAOI drugs)	Tyramine 2-Phenylethylamine	Cheese: Gouda, Swiss, Gruyère, Cheddar Beer, wine, yeast extract, Pickled herring, dry Sausage, broad beans	Hypertensive crisis, severe headache, cerebral hemorrhage, neuronal sequel, cardiac failure, pulmonary edema, visual alterations, palpitation, nausea, sweat, vomit, muscle contractions, excitation, and mental confusion
Migraine	Tyramine 2-Phenylethylamine Tryptamine Serotonin	Cheese Chocolate Beer Wine	Throbbing headache, migraine attack

[a]MAOI, monoaminoxidase inhibitors.

leading to a variety of physiological reactions. There is an increase in blood pressure by peripheral vasoconstriction and by increasing the cardiac output. Tyramine also can dilate the pupils and the palpebral tissue, cause lacrimation and salivation, and increase respiration.

Ingestion of foods rich in tyramine by individuals under MAO inhibitor drugs treatment results in a dangerous intoxication known as the 'cheese reaction,' so called because most of the cases involved cheese. However, it is not the only type of food incriminated, since cases have also been reported with yeast extracts (marmite), pickled herring, dry sausage, alcoholic beverages, broad beans, etc. The cheese reaction involves a hypertensive crisis, usually accompanied by a severe headache. There can be visual alterations, nausea, vomit, muscle contraction, mental confusion, or excitation. It can lead also to cerebral hemorrhage, neuronal sequel, cardiac failure, and pulmonary edema. Fatal incidents have been reported in the literature.

Several amines, such as tyramine, tryptamine, and phenylethylamine, are considered precipitants of attacks in migraine sufferers.

Even though putrescine and cadaverine have less pharmacological activity than the aromatic amines, after ingestion of very large amounts of these compounds, toxic effects have been observed. Intoxication symptoms reported are hypotension, bradycardia, dyspnea, lockjaw, and paresis of the extremities. However, the most important consequence of these compounds in food is probably the potentiation effect of the toxicity of other amines.

Determination of the exact toxicity threshold of amines is extremely difficult. The toxic dose is dependent on the efficiency of the detoxification mechanism of different individuals. Upper limits of 10 mg of histamine, 10 mg of tyramine, and 3 mg of 2-phenylethylamine in 100 g of foods have been suggested. In the case of alcoholic beverages, the proposed limits are 2–8 mg of histamine and 8 mg of tyramine per liter. Individuals taking MAO inhibitor

drugs should limit tyramine intake to 6 mg per 100 g of foods.

As mentioned before, polyamines play important roles in cellular metabolism and growth. However, the importance of putrescine, spermine, and spermidine in tumor growth is widely recognized, and the inhibition of polyamines biosynthesis in tumor-bearing individuals is a major target of cancer therapy research. A new direction being investigated to inhibit tumor growth is to limit polyamine uptake.

Furthermore, putrescine and cadaverine can be converted to pyrrolidine and piperidine, respectively. These compounds, as well as spermidine, spermine and agmatine, can react with nitrite to form carcinogenic nitrosamines in foods or *in vivo*, in the gastrointestinal tract.

Occurrence of Amines in Foods

Biologically active amines are inherent to living organisms and, therefore, are present in plants, meat, and dairy products. The amount and type of amines in foods depend on the nature and origin of the commodity. However, it can change during processing, fermentation, and storage. It can be affected also by hygienic conditions. Amines are resistant to heat treatment employed in food processing. Based on these findings, amines have been considered good indicators of the freshness, spoilage, and degree of quality of fresh and canned food products.

It is evident that all types of food, whether they originate from plants or animals, contain putrescine and spermidine. In addition to these most common compounds, other amines also may occur naturally. When the food is spoiled by bacteria, the spermine content often decreases, because it can be used as a nitrogen source by some microorganisms. There is also formation and accumulation of different types of biogenic amines.

Much still remains unknown about the concentration of amines in food products. Most data available relate to specific amines, mainly histamine and tyramine in fish, cheese, meat, and alcoholic beverages. A comprehensive study of levels of all amines would be useful from both a toxicological and technological point of view. This is obviously an area deserving further detailed investigation.

Fish

Under normal physiological conditions, fish muscle contains high levels of spermidine and spermine and low levels of histamine and putrescine (Table 2). During storage and deterioration of fish, the levels of spermine and spermidine decrease, levels of putrescine, histamine, and cadaverine increase, and, in some species of fish, the presence of tyramine and tryptamine can be detected also. Based on these findings, a chemical index, calculated by the levels of putrescine, histamine, and cadaverine divided by those of spermine and spermidine, was proposed to evaluate the

Table 2 Mean levels of biologically active amines in different fresh, stored, and deteriorated fish, meat, and meat products

Food	*Mean amine level (mg per 100 g)*[a]					
	Spermine	*Spermidine*	*Putrescine*	*Cadaverine*	*Histamine*	*Tyramine*
Tuna fish						
Fresh	0.95	0.44	0.12	0.15	0.38	–
Decomposed	0.12	0.07	0.25	1.93	25.3	–
Snapper[b]						
Fresh	0.72	0.68	0.99	nd	nd	nd
Spoiled	0.52	1.52	3.32	9.33	0.44	nd
Pork						
Fresh	6.71	0.70	0.78	1.33	0.47	–
15 days/5 °C	3.12	0.31	1.89	4.30	0.99	–
Beef[c]	2.27	0.61	0.51	0.95	0.57	1.07
Chicken						
Fresh	1.79	0.73	0.02	nd	nd	nd
15 days/4 °C	1.12	0.87	2.04	0.43	1.03	1.74
Hot dog	1.08	1.58	0.06	0.04	0.09	0.01
Sausage	1.03	0.66	1.51	0.80	1.35	0.74

[a]nd, not detectable; –, not determined.
[b]Snapper – *Piaractus mesopotamicus*.
[c]Mean of 62 samples.
From Sayem-El-Daher N, Simard RE, Fillion J and Roberge AG (1984) Extraction and determination of biogenic amines in ground beef and their relation to microbial quality. *Lebensm. Wiss. Technol*. 17: 20–23; Halász A, Baráth A, Simon-Sarkadi L and Holzapfel W (1994) Biogenic amines and their production by microorganisms in food. *Trends in Food Science and Technology* 51: 42–49; Silva and Glória (2002) Bioactive amines in chicken breast and thigh after slaughter and during storage at 4 1°C and in chicken-based meat products. *Foods Chemistry* 78: 241–248.

quality of tuna fish. This index was considered adequate to evaluate the quality or freshness of tuna and other types of fish and seafood, among them, salmon, rockfish, snapper, lobster, and shrimp.

Amine formation and build-up in fish normally occur during decomposition or spoilage processes as a result of the liberation of free amino acids through proteolysis together with bacterial production and action of amino acid decarboxylase. Several investigations have demonstrated that storage temperature is a critical factor influencing histamine and other amines formation in fish muscle. The requirement for rapid and uninterrupted refrigeration after catch cannot be overemphasized. The US Food and Drug Administration (FDA) established a guidance level for histamine in fish at the dock or before processing of 5 mg per 100 g. The FDA also states that rapid chilling of fish immediately after catch is the most important strategy to limit histamine formation. The internal temperature of the fish should be brought to 10 °C or below within 6 h of death. Chilling from 10 to 4.4 °C or less should take no longer than 18 h. Any period of time at a temperature above 4 °C significantly reduces the expected safe shelf-life. A maximum average histamine content of 10 mg per 100 g has been established by several countries and communities for acceptance of canned tuna and other fish belonging to the Scombridae and Scomberesocidae families. Because of the potentiating effect of other amines on the toxic effect of histamine, European regulation recommends the use of high-performance liquid chromatography (HPLC) techniques for the determination of the amines.

Scombroid fish, such as tuna, bonito, mackerel, yellowfin, and bluefin, have been involved more frequently in histamine intoxication incidents. Scombroid fish are particularly susceptible to histamine formation, since they contain large amounts of free histidine. However, other fishes have also been implicated, including Scomberesocidae, Pomatomidae, Coryhaenidae, Carangidae, Clupeidae, and Engraulidae. Even though several fish bacteria, either natural or contaminants, are capable of producing histamine, *Morganella morganii*, *Hafnia alvei*, and *Klebsiella pneumoniae* are the only histamine-producing bacteria that have been isolated from fish implicated in outbreaks.

Meat and Meat Products

Putrescine, spermine, and spermidine occur naturally in meat, where they serve as growth factors and are involved in a number of vital processes. Compared with fish, meat contains higher spermine levels. Meat products are also susceptible to proteolysis by enzymes (endogenous or from microbial contaminants), liberation of amino acids, and amine formation by amino acid decarboxylation. Amine levels in different types of fresh and stored meat are listed in Table 2. During storage and putrefaction of pork meat, there is a significant increase in putrescine and cadaverine. A significant correlation has been observed between these amines and total microbial counts.

In poultry, immediately after slaughter, high levels of spermine and spermidine and traces of putrescine have been detected. During storage at 4 °C, there is a decrease in spermine levels, whereas spermidine levels remain constant. The levels of putrescine, cadaverine, histamine, and tyramine increase significantly on the 15th day with higher levels found in breast than in thighs. The amine levels in chicken-based products listed in Table 2 indicate that the ingredients can affect the amine profile. Chicken-based hot dog contains significantly higher spermidine levels than the meat, probably because of the addition of soybean in the formulation. Significantly higher biogenic amine levels have been observed in chicken-based sausages, which could be indicative of the use of low-quality raw material, or of contamination during processing and storage.

Sausages have been implicated in histamine intoxication and hypertensive crisis episodes. Moreover, several reports have shown a wide range and high contents of biogenic amines in sausages. The manufacture of sausages offers favorable conditions for the formation of biogenic amines. The microbial count can be high and the production process lengthy, allowing a certain degree of proteolysis and yielding the amine precursor amino acids. Raw material quality, thawing conditions, ripening temperature and addition of a properly selected starter culture can affect the formation and accumulation of tyramine and histamine and, therefore, are important critical control points in preventing amine formation.

Dairy Products

The polyamine content in milk and dairy products is low, but the biogenic amine levels in cheese can be high. In fact, cheese is one of the foods with the highest amine content. A variety of amines have been found in different types of cheese (Table 3). In general, spermine and spermidine are present at low levels. The levels of other types of amines vary widely. High histamine and tyramine levels have been reported in several types of cheese.

Cheeses represent an ideal environment for amine production. The major factors that affect the formation, accumulation, and type of amines in cheese are the availability of amino acids (and hence proteolysis of cheese) and the presence of microorganisms, either

Table 3 Levels of biologically active amines in different types of cheese

Type of cheese	*Range of amine levels (mg per 100 g)*[a]						
	Spermine	*Spermidine*	*Putrescine*	*Cadaverine*	*Histamine*	*Tyramine*	*Tryptamine*
Blue cheese	–	–	9.6–23.7	42.3–227	nd–409	2.2–166	nd–110
Cheddar	–	–	nd–99.6	nd–40.8	nd–154	nd–153	nd–30
Gorgonzola	–	–	1.2–124	5.8–428	1.7–191	8.9–255	2.4–43
Gouda	nd–1.13	nd–1.35	nd–107	nd–99.5	nd–30.5	nd–67	nd–88
Mozzarella	nd–1.31	nd–1.06	nd–1.37	nd–2.34	nd–11.3	nd–41	nd–10
Parmesan	0.07–0.09	nd–0.15	nd–4.30	nd–9.80	nd–27.2	nd–29	nd–1.70
Provolone	0.07–0.97	nd–2.38	nd–8.70	nd–111	nd–8.2	nd–10.9	nd–1.08
Swiss	–	–	–	–	nd–250	nd–180	nd–1.60

[a]nd, not detectable; –, not determined.

From Stratton JE, Hutkins RW and Taylor SL (1991) Biogenic amines in cheese and other fermented foods: a review. *Journal of Food Protection* 54: 460–470; Halász A, Baráth A, Simon-Sarkadi L and Holzapfel W (1994) Biogenic amines and their production by microorganisms in food. *Trends in Food Science and Technology* 51: 42–49; Vale S and Glória MBA (1998) Biogenic amines in Brazilian cheeses. *Food Chemistry* 63(3): 343–348.

added or as a contaminant, capable of decarboxylating amino acids. Several factors may contribute also, such as pH, salt concentration, water activity, ripening temperature and time, storage temperature, bacterial density, and activity of amine catabolic enzymes by means of mono- and diaminoxidase.

The longer the aging process, the higher the biogenic amine content. Amines concentrate on the rind of cheese. Numerous bacteria, both intentional (starter cultures) and adventitious, have been reported as being capable of amine production. *Lactobacillus* spp. play a major role in histamine, tyramine, and putrescine accumulation. *Enterococcus* spp. are notorious tyramine formers. Representatives of the Enterobacteriaceae, even at low densities, can cause cadaverine and putrescine build-up.

Several outbreaks of histamine intoxication have occurred following the consumption of cheese, particularly Swiss and Cheddar, containing high levels of histamine. Migraine headaches have been observed after ingestion of cheese with high levels of tyramine, tryptamine, or phenylethylamine. Hypertensive crisis has been observed after consumption of Swiss, Cheddar, Gouda, and Gruyère cheeses with high tyramine levels. Therefore, susceptible individuals should be advised to consume cheese with a low biogenic amine content. Furthermore, efforts should be made to elucidate amine formation in cheese in order to optimize the technology and secure low amine levels.

Fermented Beverages

Although tyramine and histamine are the major amines investigated in wine and beer, several different amines have been detected. The type and concentration of amines in wines vary widely (Table 4). In general, red wines contain significantly higher amine levels compared with white. Some amines are normal constituents of grapes, with levels varying with variety and degree of ripening as well as with soil type and composition. Among amines detected in grapes, putrescine and spermidine are usually abundant, whereas agmatine, cadaverine, spermine, and histamine have been found in smaller amounts. During wine-making, several amines can be formed and accumulate. Several factors can affect amine formation, including the presence of precursor free amino acids, must treatment, contact time of must and grape skin, initial microbial population present on the fruit, alcohol content, sulfur dioxide concentration, added nutrients, pH, temperature, quantity and type of finings, and bacterial contamination in wineries. A number of researchers have demonstrated that the amine content increases with microbial growth.

The presence of biogenic amines in wines is also of interest from a toxicological point of view. Histamine has been considered the causative agent of physiological distress experienced by individuals following the consumption of red wines. However, reports have demonstrated that tyramine, tryptamine, 2-phenylethylamine, and serotonin may play an important role as well as the simultaneous presence of potentiating compounds in wines, such as ethanol, aldehydes, and polyamines. Upper limits for histamine in wine have been recommended in Germany ($2\,mg\,l^{-1}$), Belgium ($5–6\,mg\,l^{-1}$), France ($8\,mg\,l^{-1}$) and Switzerland ($10\,mg\,l^{-1}$).

Since beer is generally consumed in greater amounts than wine, it has been suggested that beer might be more of a hazard to the consumer. Mainly, raw materials, brewing techniques, and microbial contamination during brewing affect the types and levels of amines in beers. High levels of putrescine, agmatine, spermine, and spermidine and low levels of histamine, tryptamine, 2-phenylethylamine, and cadaverine are usually found in malt. Tyramine, putrescine, spermine, spermidine, and agmatine have been detected in barley. Relatively high levels of tyramine, 2-phenylethylamine, putrescine, spermine,

Table 4 Types and levels of biologically active amine levels in US wines and Brazilian beers

Food (number of samples)	*Range of amine levels (mg l^{-1})*[a]								
	Agmatine	*Spermidine*	*Spermine*	*Putrescine*	*Cadaverine*	*Histamine*	*Tyramine*	*Tryptamine*	*2-Phenyl-ethylamine*
Wine									
Pinot noir (36)	nd–8.37	nd–2.35	nd–2.38	2.43–203	nd–2.07	nd–23.98	nd–8.31	nd–5.51	nd–0.89
Cabernet Sauvignon (23)	nd–1.61	nd–4.03	nd–1.17	3.15–23.6	nd–1.51	nd–10.10	nd–7.53	nd	nd–0.14
Beer									
Lager (46)	2.10–46.8	nd–6.00	nd–1.41	0.85–9.80	0.15–2.60	nd–0.90	0.30–3.10	nd–0.80	nd–0.70
Stout (10)	2.80–16.8	0.31–1.38	nd–2.05	1.99–5.84	0.30–1.37	nd–0.85	0.48–36.8	nd–10.1	nd–0.69
Ice (5)	3.20–4.00	0.60–0.80	nd–0.30	3.90–4.50	0.10–0.20	nd	0.70–1.40	nd	nd
Bock (23)	4.80–35.1	0.25–2.10	nd–1.73	1.55–6.30	0.15–1.72	nd–1.46	0.81–5.05	nd–3.50	nd–1.72
Nonalcoholic (7)	6.35–8.59	1.35–2.30	nd–1.20	2.30–4.95	nd–0.50	nd–0.62	0.60–3.30	nd–1.41	nd–0.32

[a]nd, not detectable.

From Glória MBA, Watson BT, Simon-Sarkadi L and Daeschel MA (1998) A survey of biogenic amines in Orgeon Pinot noir and Cabernet Sauvignon wines. *American Journal of Enology and Viticulture* 49(3): 279–282; Glória MBA and Izquierdo-Pulido M (1999) Levels and significance of biogenic amines in Brazilian beers. *Journal of Food Composition and Analysis* 12: 129–136.

and spermidine have been detected in hops. However, their contribution to amine levels in beer is not significant, since the amount used is very small. The use of adjunct cereals such as rice has been found to be useful in reducing biogenic amine levels in wort and beer. Beers with a higher acidity usually have higher amine contents.

During beer production, several amines can be formed and accumulate. Tyramine and agmatine can accumulate during mashing and wort boiling, as a result of thermal amino acid decarboxylation and possibly enzymatic activity in the malt. Moreover, amines can be formed during fermentation by contaminating lactic acid bacteria. Therefore, biogenic amines could be associated with hygienic conditions during brewing.

In general, putrescine, agmatine, and spermidine are inherent to beers (Table 4). Large variations in the levels of histamine, tyramine, tryptamine, and cadaverine have been observed. These amines are formed mainly during brewing. Higher total amine levels have been detected in stout and lower in ice beer.

Plant Foods

In plants, putrescine, spermidine, and spermine are implicated in a number of physiological processes, such as cell division, flowering, fruit development, response to stress, and senescence. Polyamines are known to stimulate growth and inhibit ethylene production, and therefore inhibit senescence. Amines are also intermediates in alkaloids synthesis.

The polyamines putrescine, agmatine, spermidine, and spermine are often detected in plants, but spermidine is the prevalent amine, especially in seeds and tissues undergoing growth. In conditions of potassium and magnesium deficiency and high ammonium concentration, putrescine and agmatine can accumulate. Putrescine also builds up in plants subjected to a wide range of stress conditions such as osmotic shock, acidification, high salinity, drought, desiccation, and temperature change.

Phenylethylamine, serotonin, tryptamine, and histamine may be present also in plants, where they have a protective role in deterring predators. Tryptamine and phenylethylamine are precursors of growth substances. These amines have been used as a tool for taxonomic studies. However, microbial contamination can cause increases in the biogenic amine content of fruits and vegetables.

Besides spermidine, putrescine, and agmatine, other types of amines have also been detected in vegetables. Tyramine is present in cabbage, lettuce, chicory, radish, tomato, potato, green onion, spinach, and eggplant. Histamine is found in cabbage, eggplant, tomato, beet, and spinach. Tomato and eggplant also have serotonin and tryptamine. Cadaverine is only found in plants of the Leguminosae family. Sprouts are usually rich in polyamines. High levels of tyramine and histamine have been detected in fermented vegetables such as sauerkraut, table olives, soy sauce, and miso.

Fruits are generally rich in putrescine and spermidine. High amine levels have been found in orange juice (putrescine, tryptamine, tyramine, synephrine), plum (serotonin, tryptamine, tyramine), banana, avocado (serotonin, tyramine), pineapple (serotonin), and raspberry (tyramine).

Other Food Products

Chocolate has been observed to contain high phenylethylamine and serotonin levels. Phenylethylamine is derived from the decarboxylation of phenylalanine during cocoa bean roasting. Mushrooms also contain high levels of phenylethylamine.

There is a clear need for further information on the nature and quantity of amines in food products. There is also a need to understand the effect of production and processing on amine levels, in order to optimize the technology and ensure low amine levels in foods. This will enable food chemists to declare limiting amine concentrations for certain food products. Furthermore, it will allow dietitians to formulate specific diets that are scientifically based, safe, and practical.

See also: **Fish**: Tuna and Tuna-like Fish of Tropical Climates; Spoilage of Seafood; **Food Poisoning**: Classification; **Histamine**; **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage

Further Reading

Bardócz S (1995) Polyamines in food and their consequences for food quality and human health. *Trends in Food Science and Technology* 6: 341–346.

Bover-cid S, Hugas M, Izquierdo-Pulido M and Vidal-Carou MC (2000) Reduction of biogenic amine formation using a negative amino acid decarboxylase starter culture for fermentation of Fuet sausages. *Journal of Food Protection* 63(2): 237–243.

Fuzikawa CS, Hara C, Glória MBA and Rocha FL (1999) Monoamine oxidase inhibitors and diet – Update and practical recommendations for clinical use. *Brazilian Journal of Psychiatry* 48(10): 453–460.

Glória MBA and Izquierdo-Pulido M (1999) Levels and significance of biogenic amines in Brazilian beers. *Journal of Food Composition and Analysis* 12: 129–136.

Glória MBA, Watson BT, Simon-Sarkadi L and Daeschel MA (1998) A survey of biogenic amines in Oregon Pinot noir and Cabernet Sauvignon wines. *American Journal of Enology and Viticulture* 49(3): 279–282.

Halász A, Baráth A, Simon-Sarkadi L and Holzapfel W (1994) Biogenic amines and their production by microorganisms in food. *Trends in Food Science and Technology* 51: 42–49.

Quemener V, Blanchard Y, Chamaillard L *et al.* (1994) Polyamine deprivation: a new tool in cancer treatment. *Anticancer Research* 14: 443–448.

Silva CMG and Glória MBA (2002) Bioactive amines in chicken breast and thigh after slaughter and during storage at 4 1°C and in chicken-based meat products. *Food Chemistry* 78: 241–248.

Smith TA (1981) Amines in food. *Food Chemistry* 6: 169–200.

Stratton JE, Hutkins RW and Taylor SL (1991) Biogenic amines in cheese and other fermented foods: a review. *Journal of Food Protection* 54: 460–470.

Taylor SL (1986) Histamine food poisoning: toxicology and clinical aspects. *CRC Critical Reviews in Toxicology* 17: 91–121.

Vale S and Glória MBA (1998) Biogenic amines in Brazilian cheeses. *Food Chemistry* 63(3): 343–348.

Veciana-Nogués MT, Mariné-Font A and Vidal-Carou MC (1997) Biogenic amines as hygienic quality indicators of tuna. Relationships with microbial counts, ATP-related compounds, volatile amines and organoleptic changes. *Journal of Agricultural and Food Chemistry* 45: 2036–2041.

AMINO ACIDS

Contents

Properties and Occurrence

J Amaya-Farfan, School of Food Engineering, Campinas SP, Brazil
M T Bertoldo Pacheco, Instituto de Tecnologia de Alimentos Av. Brasil, 2880, Campinas SP, Brazil

Introduction

Amino acids can be defined as organic acid molecules that also carry at least one amino group. If one of the amino groups is attached to the α-carbon (relative to the carboxylic group), the amino acid will be an α-amino acid (**Figure 1**). In addition to the above, various distinguishing groups can be attached to the α-carbon, a feature which is responsible for the rather wide variety of amino acids in nature. The distinguishing groups are given the generic name of 'side chains' ('R-groups'). The most common amino acids are those found in proteins, whose side chains are listed in **Table 1**.

Owing to their chemical nature and structure, amino acids have high melting and boiling points,

Table 1 The 20 amino acids found in proteins, showing the polarity of their R-group at pH 7

Polarity of R-group	*Amino acid (abbreviations)*	*R-group or side chain*	*Stick-ball structure*
Nonpolar (hydrophobic)	L-Alaine (Ala or A)	$-CH_3$	
	L-Valine (Val or V)	$-CH(CH_3)_2$	
	L-Leucine (Leu or L)	$-CH_2-CH(CH_3)_2$	
	L-Isoleucine (Ile or I)	$-CH(CH_2-CH_3)(CH_3)$	
	L-Phenylalanine (Phe or F)	$-CH_2-C_6H_5$	
	L-Methionine (Met or M)	$-CH_2-CH_2-S-CH_3$	
	L-Proline (Pro or P)	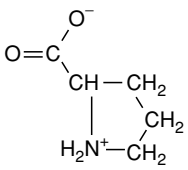	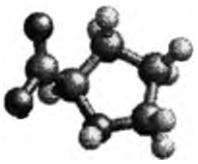
	L-Tryptophan (Try or W)	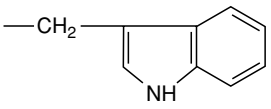	
Polar but uncharged	L-Glycine (Gly or G)	$-H$	
	L-Serine (Ser or S)	$-CH_2-OH$	
	L-Glutamine (Gln or Q)	$-CH_2-CH_2-C(=O)-NH_2$	

Continued

Table 1 Continued

Polarity of R-group	*Amino acid (abbreviations)*	*R-group or side chain*	*Stick-ball structure*
	L-Threonine (Thr or T)	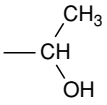	
	L-Cysteine (Cys or C)	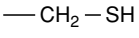	
	L-Asparagine (Asn or N)	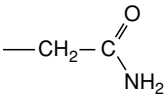	
	L-Tyrosine (Tyr or Y)	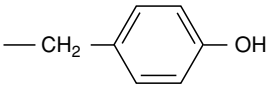	
Polar charged	L-Aspartic acid (Asp or D)	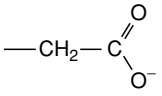	
	L-Glutamic acid (Glu or E)	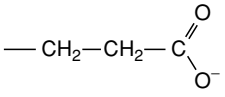	
	L-Lysine (Lys or K)	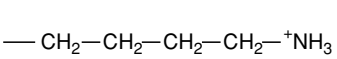	
	L-Arginine (Arg or R)	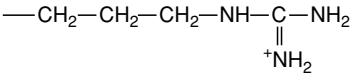	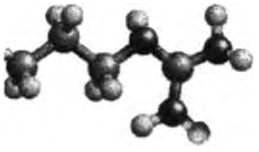
	L-Histidine (His or H)	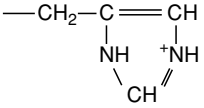	

Figure 1 The structure of an α-amino acid show as a zwitterion.

are water-soluble, and can readily ionize their carboxylic and amino groups in aqueous solutions. Since these molecules can carry both negative and positive charges, they are called zwitterions or ampholytes; that is, they can migrate to either the cathode or the anode in an electric field. By adjusting the pH of the solution, therefore, a point will come when the opposite charges are equal in magnitude and the

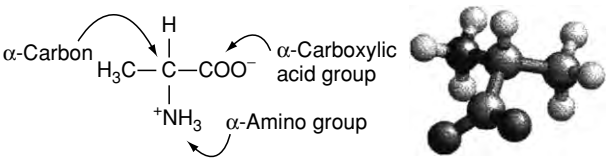

Figure 2 Alanine – a typical α-amino acid.

molecule will not migrate (Figure 2). The amino acid is said to have reached its isoelectric or isoionic point (*p*I). (*See* **pH – Principles and Measurement.**)

Isomerism

With the obvious exception of glycine, all α-amino acids contain at least one asymmetric carbon atom, i.e., their α-carbon atoms are indeed bonded to four different substituent groups. Most amino acids therefore exist in two different spatial arrangements or isomers, whereas isoleucine, threonine, hydroxylysine, and hydroxyproline each have a second asymmetric carbon and thus four stereoisomers.

Standard nomenclature of amino acids is preceded by either a D or an L, indicating whether a structure corresponds to either that of D(+)-glyceraldehyde or L(−)-glyceraldehyde. This designation is the 'absolute configuration' of an amino acid. The relationship between the D and L isomers of glyceraldehyde and alanine can be seen in Figures 3 and 4, respectively.

Since stereoisomers are mirror images of each other and exhibit characteristic angles of rotation for polarized light, nomenclature also includes either a (+) or (−) (formerly *d* or *l*) to indicate that a pure solution of the amino acid would rotate a beam of polarized light clockwise (dextrorotation) or counterclockwise (levorotation), respectively.

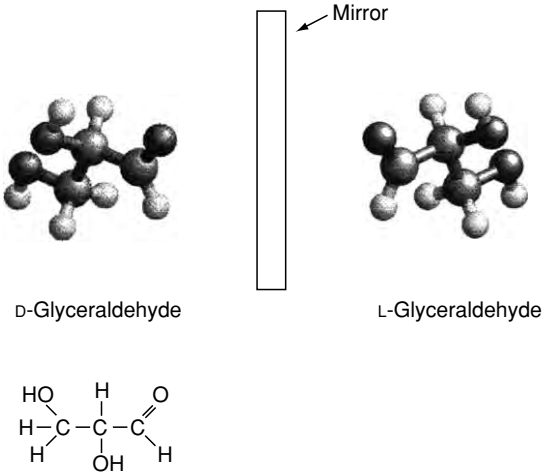

Figure 3 The stereoisomers of glyceraldehyde.

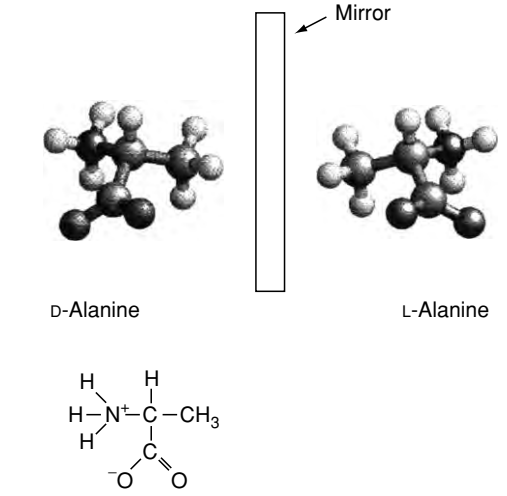

Figure 4 The stereoisomers of alanine.

Amino Acids as Protein Constituents

Over 95% of the amino acids in a food will be encountered in the form of proteins. Proteins are unbranched chains (polypeptides) of L-α-amino (or α-imino) acids linked together by peptide (amide) bonds. Combinations of about 20 amino acids commonly make up the structures of polypeptides. Different sequences made with the same amino acids will spell out different proteins, each sequence corresponding to a different primary structure. The remainder will be found as peptides and free amino acids. (*See* **Protein**: Chemistry.)

Interestingly, only the L isomers of α-amino acids have been found in all proteins. Although D-amino acids are not involved in the metabolism of higher organisms, they can be structural components of cell walls in certain bacteria and part of some peptide antibiotics. L-Amino acids will slowly convert to racemic mixtures of D- and L-amino acids with time or as a result of alkali exposure or considerable heat treatment.

A peptide bond results from the condensation of the α-carboxylic group of one amino acid and the α-amino group of another. Figure 5 shows a dipeptide linkage between two alanine residues.

The peptide bond can be visualized as a plane, in the middle of which lies the amide (C-N) linkage (Figure 6). Two opposite corners of the plane are occupied by the two α-carbons, and each of the remaining two is occupied by an oxygen and a hydrogen atom. The plane is considered a rather rigid structure ($E_{\text{rotation}} = 21\ \text{kcal mol}^{-1}$) due to the tendency of

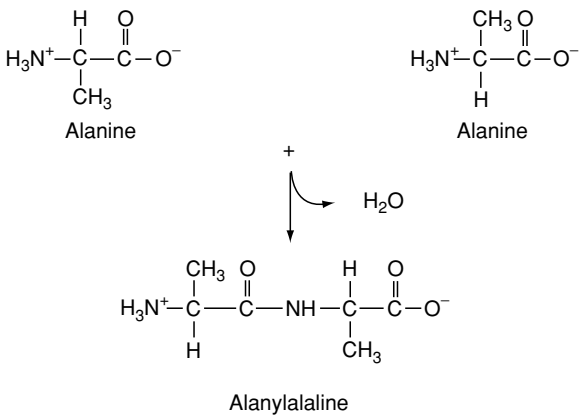

Figure 5 The formation of the peptide bond between two molecules of alanine.

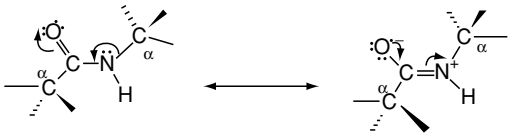

Figure 6 The two structures contributing to the resonance hybrid structure of the peptide bond.

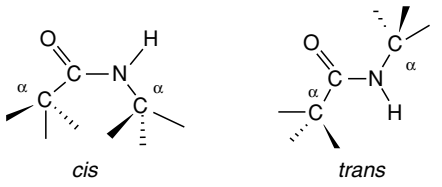

Figure 7 The *cis* and *trans* configurations of the peptide bond.

the electron clouds of the system to remain orbitting, each on a separate side of the plane. Once a free amino acid has been incorporated into a polypeptide or protein, all physical, physicochemical, and chemical properties attributed to the free form are altered.

The peptide bond exists in nature in either the *trans* or *cis* configuration. The *cis* configuration is less common and results from the presence of proline residues (Figure 7).

Imino Acids Found in Proteins

Proteins often contain small amounts of the imino acid L-proline and, more rarely, L-4-hydroxyproline (Hypro). Unlike α-amino acids (with primary amines), the imino acids cannot keep their side chains as far away as possible from those of the neighboring residues, as the *trans* configuration would require. Therefore, the mere presence of a proline (or a hydroxyproline) residue in the sequence introduces a 'kink' or bend in the polypeptide structure (Figure 8).

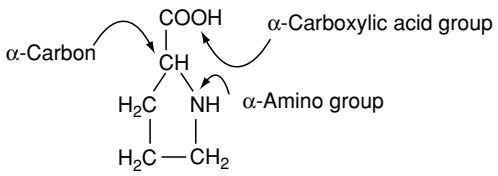

Figure 8 Structure of proline. Note that the α-carbon and 'amino' nitrogen are involved in the side chain, and the amine is thus secondary.

Table 2 The nonessential (dispensible) and essential (indispensible) amino acids for humans

Nonessential	*Essential*
Alanine	Arginine[a]
Asparagine	Histidine[a]
Aspartate	Isoleucine
Cysteine[b]	Leucine
Glutamate	Lysine
Glutamine	Methionine[b]
Glycine	Phenylalanine[c]
Proline	Threonine
Serine	Tryptophan
Tyrosine[c]	Valine

[a]Essential to the young. Not limiting in most foods.
[b, c]Due to one-way interconversions, nutritional evaluation takes into account the pairs.

If the imino acid content of a polypeptide is high (3–10%), then the peptide is likely to fold randomly. If the imino acid residues are repeated regularly along the polypeptide (as is true of collagen, ~23% Pro + Hypro), then the protein is able to fold into a regular structure.

The Classification of Amino Acids

Amino acids can be usefully classified in several ways, depending on their properties. Two common classifications are by their nutritional essentiality to animals and by the polarity of the R-groups at neutral pH (Tables 2 and 3, respectively). As most proteins and amino acids function in aqueous environments at physiological pH, the second criterion is most helpful when considering the function of an amino acid in a protein *in vivo* or in most food systems. Hydrophobic R-groups tend to 'hide' themselves in the interior of folded proteins. Table 1 highlights the structures of the 20 common protein amino acids classified according to their polarity in neutral solutions.

A nutritionally essential amino acid is one that cannot be manufactured, either at all or in sufficient quantities for optimal growth by the body, and

Table 3 Genetic code and some physicochemical properties of the common protein amino acids while in the free form

Amino acid	*Genetic code*	*Solubility ($g\,l^{-1}$)*[a]	*PK_1 (α-COO^-)*	*PK_2 (α-NH_3^+)*	*PK_3*[b]	*PI*
Alanine	GN(N)	167.2	2.34	9.69		6.00
Arginine	AGA/AGG/CG(N)	855.6	2.17	9.04	12.48	10.76
Asparagine	AAU/AAC	28.5	2.02	8.80		5.41
Aspartic acid	GAU/GAC	5.0	1.88	9.60	3.65	2.98
Cysteine	UGU/UGC		1.96	10.28	8.18	5.07
Glutamic acid	GAA/GAG	8.5	2.19	9.67	4.25	3.22
Glutamine	CAA/CAG	7.2[c]	2.17	9.13		5.65
Glycine	GG(N)	249.9	2.34	9.60		5.97
Histidine	CAU/CAC		1.82	9.17	6.00	7.59
Isoleucine	AUU/AUC/AUA	34.5	2.36	9.68		6.02
Leucine	UUA/UUG/CU(N)	21.7	2.30	9.60		5.98
Lysine	AAA/AAG	739.0	2.18	8.95	10.53	9.74
Methionine	AUG	56.2	2.28	9.21		5.74
Phenylalanine	UUU/UUC	27.6	1.83	9.13		5.48
Proline	CC(N)	1620.0	1.94	10.60		6.30
Serine	AGU/AGC	422.0	2.20	9.15		5.68
Threonine	AC(N)	13.2	2.21	9.15		5.68
Tryptophan	AGG	13.6	2.38	9.39		5.89
Tyrosine	AUA/UAC	0.4	2.20	9.11		5.66
Valine	GU(N)	58.1	2.32	9.62		5.96

[a]Solubility in water at 25 °C.
[b]Side chain ionizing groups.
[c]Solubility at 37 °C.

therefore has to be present in the diet. These amino acids are sometimes termed collectively as 'indispensible.' In contrast, nonessential amino acids (indispensible nitrogen) can thus be interconverted in order for the body to meet its needs.

Although the nutritive or biological value of a protein can be measured by its contents of essential amino acids, nonessential amino acids should be regarded as important, as maximal protein efficiency will only be attained if an adequate supply of nonspecific nitrogen is guaranteed in the diet. Besides serving as protein constituents, surplus essential amino acids are used as precursors of other vital metabolites and occasionally as energy. Excess of nonspecific nitrogen, on the other hand, will preferentially be used as energy. (*See* **Amino Acids: Metabolism.**)

Physical, Chemical, and Sensory Properties of Amino Acids Found in Proteins

All amino acids in the pure state are crystalline solids with physical constants much too high for either organic amines or acids of similar structures and molecular weights. They are not volatile, melt at high temperatures, are insoluble in organic solvents, and have low pK values. Their molecular weights range from 75 to 204; average molecular weight is estimated at 110. Therefore, if a protein has a molecular weight of 33 000, one could estimate it contains approximately 300 residues.

Amino Acids Commonly Found in Proteins

Although the following listing refers to those amino acids actually found in proteins, it should be remembered that they also exist in the free form in both vegetable and animal tissues.

Glycine (mol. wt 75.1) The R-group is a single hydrogen atom, making it the smallest amino acid. Because it is the smallest residue, glycine enables a protein to fold into a compact structure. Collagen can fold into a tight triple helix because every third amino acid is glycine. (The uniform sequence of collagen is abbreviated to [-Gly-Pro-X-]$_n$, where X is often hydroxyproline.) The hydrogen atoms attached to the α-carbon atom have too small an influence on the ionization and high degree of polarity of the α-amino/α-carboxylic system. Free glycine is responsible for the sweet taste of shrimp. Its name derives from the same Greek word as glucose (*glycos*).

Alanine (mol. wt 89.1) This is the next simplest amino acid. The side chain of alanine is a methyl group (Figure 2) and performs well in both hydrophobic and hydrophilic environments. Like glycine, this amino acid is found in high concentrations in tightly folded proteins such as keratin and collagen.

Serine (mol. wt 105.1) A hydroxyl group at the end of the side chain makes serine more polar than alanine, yet uncharged at pH 7. Like the other polar,

uncharged amino acids (glycine, threonine, cysteine, tyrosine, asparagine, and glutamine), serine forms hydrogen bonds with water; with sugars it forms glycosidic bonds. Serine is essential for the active sites of enzymes, as in the serine-proteases trypsin and chymotrypsin. (*See* **Enzymes**: Functions and Characteristics.)

Cystine (mol. wt 240.2) and cysteine (mol. wt 121.1) These amino acids have R-groups that contain sulfur either oxidized as a disulfide or reduced as a thiol. Cystine is two molecules of cysteine linked through a disulfide bond (Figure 9). When incorporated into proteins the disulfide bond between cysteine residues effectively 'pins' the protein structure together. The fact that most natural proteins contain disulfide bonds may be a natural mechanism to increase the protein's structure stability under unfavorable conditions and magnify its biochemical function. Intracellular fluids are mildly reducing, favoring cysteine, while the extracellular environment is sufficiently oxidizing to favor cystine. The thiol of cysteine is essential for the catalytic activity of some enzymes, such as papain, and is a prime target for enzyme inactivation by heavy metals and alkylating reagents. Free cystine has low solubility.

Aspartic acid (mol. wt 133.1) This amino acid is normally abbreviated to 'Asp' but if, when it is determined, no distinction is made between aspartic acid and its amide (Asn), it is abbreviated to 'Asx.' Aspartate has an acidic, negatively charged R-group at pH 7. It is electrically neutral at pH 2.98.

Asparagine (mol. wt 132.1) Asparagine is the amide of aspartate. This was the first amino acid to be isolated (1806), from asparagus, the plant from which it derives its name.

Valine (mol. wt 117.1) Valine has a bulky, very hydrophobic side chain. It is an essential amino acid for humans (average adults require 0.7 g day^{-1}) and can be metabolized in the muscle. Valine, leucine, and isoleucine are the branched-chain amino acids.

Leucine (mol. wt 131.2) The side chain of leucine is aliphatic, bulky, and very hydrophobic. It is essential for humans (average adults require 0.98 g day^{-1}) and can be metabolized in the muscle. Leucine, isoleucine, and valine are the branched-chain amino acids.

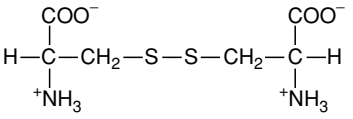

Figure 9 Disulfide bond in a cystine molecule linking two molecules of cysteine.

Isoleucine (mol. wt 131.2) This amino acid is also highly hydrophobic. It has two asymmetric carbon atoms and therefore can exist as any of four isomers. Only one isomer exists in proteins. It is nutritionally essential for humans (average adults require 0.7 g day^{-1}) and can be metabolized by the muscle. Isoleucine, valine, and leucine are the branched-chain amino acids.

Methionine (mol. wt 149.2) Like cysteine, methionine is a sulfur-containing amino acid. The sulfur in methionine can be oxidized, but is unable to form disulfide bonds. The R-group renders methionine quite hydrophobic. In prokaryotic cells a methionine residue is essential for the initiation of protein synthesis. It is nutritionally essential for humans (average adults require Met + Cys 0.91 g day^{-1}) and is limiting in most dietary proteins. It becomes toxic when ingested in excess of 4% of the protein.

Threonine (mol. wt 119.1) This amino acid has a hydroxyl group as part of the side chain, making it polar, yet uncharged at physiological pH. It has two asymmetric carbon atoms and therefore can exist in any of four isomers. Only one form exists in proteins. This was the last of the common amino acids to be identified (1938). It is an essential amino acid for humans (an average person requires 0.49 g day^{-1}).

Glutamine (mol. wt 146.1) The abbreviation for this amino acid is 'Gln' (or 'Q'), but 'Glx' (or 'Z') is used when no distinction can be made between glutamic acid and the corresponding amide glutamine. Glutamine has a polar but uncharged R-group. It is found in large amounts in the gliadin and glutenin fractions of wheat.

Glutamic acid (mol. wt 147.1) This amino acid has a γ-carboxylate group in the side chain. It is therefore found on the surface of proteins. It was discovered in 1908 in edible seaweed, when it was recognized as a flavor potentiator or enhancer. The salt monosodium glutamate (MSG) has a sweet and salty taste at very low concentrations. MSG increases sensitivity to sour and bitter tastes. Other amino acids, such as ibotenic acid and tricholomic acid, have been reported to have similar effects.

Lysine (mol. wt 146.2) A positively charged side chain with a second (ε-)amino group that can form

salt-bridges or react with other residues to form covalent cross-links is found in lysine. These cross-links are important in skin, where they are involved in aging effects. Lysine is known to react with sugars, giving rise to different flavor and color compounds (Maillard reaction). Lysine is an essential amino acid for humans. Cereal proteins are typically low in lysine, while legume proteins are high. The average person requires 0.84 g day^{-1}. (*See* **Browning**: Non-enzymatic.)

Arginine (mol. wt 174.2) This amino acid has a positively charged guanidinium group on the end of the side chain, being a much stronger base than the ε-ammonium group of lysine. Consequently, arginine is 100% protonated with a net charge of +1 at physiological pH. Ammonia intoxication results in children (and cats) if this amino acid is absent from a fasting diet. It is present in high amounts in immature seeds like groundnuts.

Histidine (mol. wt 155.2) A weakly protonated imidazole group is present in the side chain of this amino acid. It is often found in the active sites of enzymes, where it is essential for the catalytic activity. This residue is a primary target for enzyme inactivation by heavy metals and alkylating reagents. Histidine is the only amino acid with significant buffering power near the pH of intracellular fluids and blood. Hemoglobin has a uniquely high content of histidine in order to counteract the effects of excess carbon dioxide. Recommended intake for adults varies from 0.56 to 0.84 g day^{-1}.

Phenylalanine (mol. wt 165.2) This amino acid has an aromatic, nonpolar hydrophobic R-group that makes it ideal for structuring the core of proteins. When present on the surface, it can have a role in substrate recognition or hydrophobic binding of other molecules. Phenylketonuria is a disabling disease that is estimated to occur in 1% of mentally handicapped patients. It is caused by a failure of the liver to metabolize phenylalanine and sufferers must avoid eating foods containing phenylalanine. As natural proteins devoid of phenylalanine are almost nonexistent, feeding phenylketonurics is a technological challenge. Average adults require (Phe + Tyr) 0.98 g day^{-1}.

Tryptophan (mol. wt 204.2) An indole, nonpolar hydrophobic R-group makes this amino acid essential for humans (average adults require 0.25 g day^{-1}). Plant proteins are probably poor in tryptophan due to diversion of the indole group to auxin biosynthesis. Pellagra is now an uncommon disease but may occur in persons subsisting on vegetable proteins that are poor in tryptophan. Symptoms include skin rashes in response to sunlight, diarrhea, severe nervous depression, and partial paralysis. The administration of tryptophan or nicotinic acid reverses these symptoms. (*See* **Niacin**: Physiology.)

Tyrosine (mol. wt 181.2) This amino acid has an aromatic, polar but uncharged side chain (hydroxylated phenylalanine). Tyrosine is an important intermediate in the synthesis of epinephrine, thyroxin, melanin, and biogenic amines. Foods such as chocolate and cheese which is somewhat rich in tyrosine have been suggested to be responsible for some migraine-type headaches. (*See* **Migraine and Diet**.)

Posttranslational Modification of Proteins: Unusual Amino Acids

There exists a number of unusual amino acids, derived from some of the 20 standard protein amino acids. Although a few of these can still be found in the free form (Table 3), the remainder are essential to specific proteins and arise from enzymatic modifications occurring after they have been incorporated into a peptide, as listed below.

Hydroxylation

Hydroxyprolines Proline may be modified to 3-hydroxyproline or 4-hydroxyproline after it has been incorporated into a precursor peptide for collagen. Hydroxylated residues can be further glycosylated.

5-Hydroxylysine 5-Hydroxylysine is another unusual amino acid found in collagen. Hydroxylysine results from the modification of lysine after incorporation into a peptide. Scurvy symptoms occur when hydroxylation does not take place. (*See* **Scurvy**.)

Phosphorylation

The reactivity of tyrosine can be augmented in a peptidic environment. This is illustrated by the fact that tyrosine residues of certain cell proteins are particularly targeted by phosphokinase-phophatase-mediated reactions as an effective way to regulate division, growth, and general cell metabolism in both animals and plants. Epidermal-, platelet- and some insulin like-growth factors (EGF, PDGF, IGF) are examples of proteins that are activated by tyrosine phosphorylation. Mice and human oncogenes can produce tyrosine kinases with uncontrolled phosphorylation activities that lead to leukemia, breast, and prostate cancer.

Carboxylation

γ-Carboxyglutamic acid This amino acid is associated with calcium binding and is found in proteins such as prothrombin, the protein involved in blood clotting, and other proteins in bone tissue.

Methylation

ε-N-Methyllysine This amino acid is abundant in the contractile muscle protein myosin and in cytochrome *c*. (*See* **Exercise**: Muscle.)

Glycosylation

Many proteins have carbohydrate side chains. The sugars (often galactose and glucose) are attached to asparagine, serine, or theonine residues or other hydroxylated amino/imino acid residues. Glycoproteins can form gel structures that influence the water activity and the texture of foods.

Covalent Cross-Linking

Desmosine The structure of desmosine is shown in Figure 10. It is a derivative of lysine and forms a covalent cross-link between polypeptides. It is mainly found in elastin and is relevant to the aging process, as seen in skin. Other cross-links (aldolic) that occur in collagen are associated with meat texture.

Sulfur–Selenium Exchange

Selenocysteine and selenomethionine These are analogs of both cysteine and methionine that result from sulfur–selenium exchange, occurring mostly in plants grown in selenium-rich soils. These amino acids are found in the active site of selenium-dependent enzymes, such as glutathione peroxidase, which are essential for the removal of toxic hydrogen peroxide from cells. (*See* **Selenium**: Physiology.)

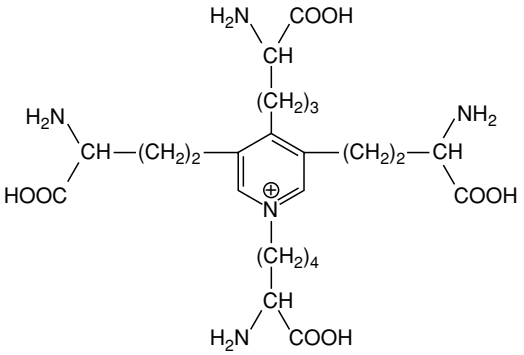

Figure 10 The structure of desmosine.

Occurrence in Foods and Nutritional Implications

All animal foods, such as meat, fish, milk, and eggs, are protein-rich, whereas most plant foods are not. Leguminous seeds, like beans and nuts, where protein and lysine are high, are an exception. These proteins, however, are deficient in the sulfur-containing amino acids, particularly methionine. The proteins found in cereal grains (seeds of maize grain, rice, sorghum), in turn, are deficient in lysine and tryptophan, but not so in methionine. Leafy vegetables are somewhat richer than root vegetables, whereas fruits are very low in protein. The proteins of seeds of pseudocereals like quinoa (*Chenopodium quinoa* Willd.) and amaranthus grain (*Amaranthus* sp.) have amino acid profiles that combine the characteristics of both legumes and cereals. (*See* **Protein**: Food Sources.)

Humans can synthesize only 10 of the 20 amino acids required for protein synthesis. They must obtain the rest from their diet if they are to remain healthy (Table 2). (*See* **Protein**: Synthesis and Turnover.)

Interestingly, the metabolisms of children and adults are slightly different. The growing child has a greater requirement for arginine and histidine than the adult, whereas adults synthesize sufficient arginine for their daily requirement.

Consistent with food composition data, proteins rather than free amino acids are the main source of dietary nitrogen. Some special diets, however, require the use of predigested proteins or free amino acids. About 75% of amino acid metabolism in normal healthy adults is devoted to protein synthesis. This protein is essential for normal growth, repair, and defense, including the production of active enzymes, plasma protein, muscle creatine and, in females, reproduction and synthesis of milk proteins. The remaining 25% of nitrogen metabolism produces intermediates of the tricarboxylic acid (TCA) cycle, hormones, and neurotransmitters.

Amino acids that are not associated with proteins in a given tissue or food can be collectively called 'nonprotein' amino acids or 'nonprotein' nitrogen. In spite of the fact that nonprotein amino acids are present in relatively small amounts, their importance in food science and nutrition can be significant. Table 3 lists the common protein amino acids, as free species, their solubility, genetic code, *p*K and *p*I values.

Less common, nonprotein amino acids, some with very peculiar properties, are listed in Table 4. Worthy of mention are theanine, typical of green tea, which has a relaxing effect in humans (Figure 11) and taurine in human milk. Taurine (β-aminoethanesulfonic acid) is an unusual amino acid, normally found in the

Table 4 Natural nonprotein amino acids found in plant and animal tissues

Name	*Occurrence, function, and importance*
β-Alanine (β-aminopropionic acid)	A component of pantothenic acid, coenzyme A, constituent of carnosine and anserine. Also occurs in the free state
L-α-Aminobutyric acid	Normal human urine, animal and plant tissues. Also as a component of the tripeptide ophthalmic acid
β-Aminoisobutyric acid	Human urine and iris bulb in the free state
γ-Aminobutyric acid	Occurs in the free state in several plants, animal tissues, and certain bacteria. Has been found in the brain of mammals, some amphibia, and birds. This amino acid appears to function in the transmission of nerve impulses
α-Aminoadipic acid	Maize, pea and other plants, human urine
α-Aminopimelic acid	Isolated from *Asplenium septentrionale*
O-acetylhomoserine	This amino acid has been isolated from pea plants
O-Phosphohomoserine	Isolated from extracts of *Lactobacillus casei*
L-Azaserine (*O*-diazetoacetil-L-serine)	Isolated from cultures of a strain of *Streptomyces*. This amino acid inhibits the growth of certain experimental tumors
L-Albizzine (L-2-amino-3-ureidopropionic acid)	It has been found in the seeds of *Albizzia julibrissin* and in other species of Mimosaceae
Alliin (*S*-allyl-L-cysteine)	A constituent of garlic. Alliin is enzymatically converted by alliinase to allicin, which exhibits an odor characteristic of garlic
Betaine	A constituent of plant and animal tissues, an intermediate in the metabolism of lipids
L-Baikiain	Rhodesian teak (*Baikiaea plurijuga*)
Canavanine	Soy and jack bean meals and also in a number of other plants
m-Carboxy-L-tyrosine	Seeds of *Reseda odorata* L.
β-Cyano-L-alanine	Seeds of the vetch (*Vicia sativa*)
L-Citrulline	Fish, juice of watermelon, other plants and animal tissues, important intermediate in urea cycle; precursor of arginine
L-Djenkolic acid	Djenkol bean (*Pithecolobium lobatum*, from Java), *Acacia farnesiana*
Dichrostathinic acid	Seeds of *Dichrostachys glomerata*
β,γ-Dihydroxyglutamic acid	Found in lettuce seeds, rhubarb, and other plants
1,3-Dimethylhistidine	Reported in the Australian seaweed *Gracilaria secundata*
2-(β-D-Glucopyranosyl)-4-alanyl-3-isoxazolin-5-one	Found in peas (*Pisum sativum*)
Glycine	In the free form plays an important role in the sweet flavor of shrimp
L-Homoarginine	Seeds of legumes of the genus *Lathyrus* and *Lotus helleri*
L-Homomethionine	Present in cabbage
Homoserine	Peas and other plants
Homocysteine	Found in adrenal extracts
Hypusine	A basic amino acid reported in bovine brain and other mammalian tissues
3-Hydroxypipecolic acid	Halophyte leaves, flowers, and fruits
5-Hydroxy-L-tryptophan	Found (6–10%, fresh weight) in the West African legume *Griffonia simplicifolia*
Lanthionine	Antibiotics (subtilin, penicillins)
Lathyrine	Seeds of *Lathyrus tingitanus*
Lisopine	Crown-gall tissues of salsify, tobacco, and normal plant tissues
1-Methylhistidine	Anserine, human urine
β-Methylene-L-norvaline	Found in the mushroom *Lactarius helvus*
β-Methylene-L-norleucine	Found in the mushroom *Amanita vaginata*
S-Methylmethionine	Methylsulfonium derivative of methionine found in cabbage and asparagus
N-Methyltryptophan	Methyl ester found in the legume *Aotus subglauca*
Ornithine	Tyrocidine, urea cycle, as ornithuric acid in excreta of birds, some fishes
Octopine	Octopus muscle
Pipecolic acid	Occurs in plant tissues and in mammalian livers as a product of lysine metabolism
L-Saccharopine	Bakers' and brewers' yeasts
Taurine	Sulfur amino acid essential to infants fed cows' milk
Theanine	Green tea. Causes a specific relaxation effect in humans
Tricholomic acid	Responsible for the good taste of mushrooms. Toxic to flies
L-Willardiin	Seeds of several *Acacia* species, including *Acacia willardiana*

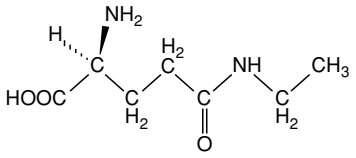

L-Theanine (γ-glutamylethylamide)

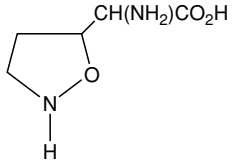

Tricholomic acid

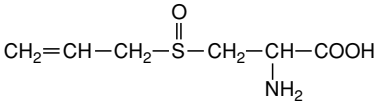

Alliine

Figure 11 The structures of L-theanine, tricholomic acid and alliine.

free form, manufactured by the body from cysteine or methionine and essential for the production of conjugated bile salts (taurocholic acid). Since cows' milk has low concentrations (1–3 μmol 100 ml^{-1}) of taurine in relation to human milk (27–35 μmol 100 ml^{-1}), there is concern that infants fed formulas based on cows' milk may develop a deficiency of taurine.

Amino Acids, Neurotransmitters, and Hormones

The amino acids glutamate, glutamine, asparate, glycine, and γ-aminobutyric acid (GABA) will act as neurotransmitters for specific types of neurones or regions of the brain. Glutamate is one important transmitter in the central nervous system of invertebrates and possibly in humans. GABA is found in relatively high concentrations (around 0.8 mmol l^{-1}), almost exclusively in brain tissue. It acts as an inhibitory transmitter by making it more difficult for a nerve to fire impulses. The structure of GABA is shown in Figure 12.

Thyroxine and triiodothyronine (Figure 13) are hormones derived from the amino acid tyrosine. Both these hormones stimulate metabolism in tissues. (*See* **Hormones**: Thyroid Hormones.)

Chemical and Physicochemical Properties Relevant to Foods

The participation of amino acids in the many changes that occur during ripening, storage, or processing of raw materials and foodstuffs is a direct consequence of their capacity to react chemically or interact physicochemically with other food components. Amino acids, *per se*, for example have little direct relevance on flavor, while proteins have even less. Proteins and peptides however have an important role as emulsifiers, fat and odor binders and thus

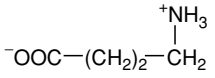

Figure 12 The structure of γ-aminobutyric acid (GABA).

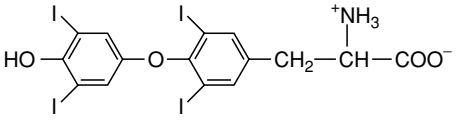

Thyroxine

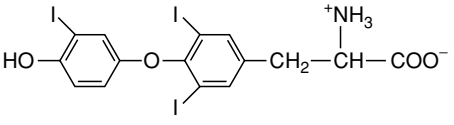

Triiodothyronine

Figure 13 The structures of thyroxine and triiodothyronine.

can affect the flavor and texture characteristics significantly, while some amino acids or their derivatives have direct bearings on food flavor and aroma. Examples are the sweet taste of naturally grown shrimp because of the high concentrations of glycine and the characteristic flavor of mushrooms imparted by tricholomic acid (Table 4, Figure 11).

The shifting of the pK_as of a free amino acid towards lower pH values makes the amine groups more reactive against carbonyl compounds. Side chain reactivity is also enhanced through the effect of the α-amino/α-carboxylic system. Amino acids with aromatic or alifatic side chains are also not as hydrophobic as they would be without the α-amino/α-carboxylic system. The following are reactions responsible for the development of some food attributes.

Deamination

Amino acids undergo deamination during fermentation. This is important in the production of flavor compounds in bread, some cheeses, and beers. If deamination is allowed to proceed extensively, diamines and biogenic amines may form, giving the food a fishy or spoiled odor.

During the ripening of bananas L-leucine can give rise to 13 volatile alcohols, aldehydes and esters, while L-phenylalanine will generate seven and L-valine produce four, some of which contribute to the characteristic banana aroma. Sulfur amino acids can be precursors of such volatile compounds as H_2S, CH_3SH, CH_3CH_2SH, propanethiol, dimethylsulfoxide, methylisothiocyanate, allylthiocyanate,

allylisothiocyanate, thiophene, acetylthiophene, benzothiophene. Garlic, onions chives, leeks, and shallots are rich in alliins (*trans-S*(+)-1-propenyl-L-cysteine-sulfoxide is a major alliin), important sulfur compounds that are released when the bulbs are crushed (Figure 11). (*See* **Flavor (Flavour) Compounds:** Structures and Characteristics.)

Maillard Browning

Under a wide range of moisture and temperature conditions, amino acids and carbonyl compounds can undergo a complex series of condensation, deamination, decarboxylation, and polymerization reactions called 'Maillard browning.' The products of the reaction of amino acids or proteins with aldehydes, ketones, and reducing sugars are numerous. They can be volatile, low-molecular-weight or complex brown-colored, high-molecular-weight, fluorescent compounds referred to collectively as 'melanoidins.' The correct amount of Maillard products is very important in determining the acceptability of baked and roasted foods. Even though an extensive part of the classical nutritive value of the reacted amino acids will be lost and some biogenic amines may form, these products add to the desired aroma, flavor, and color of cakes, pies, powder and condensed milk, roasted groundnuts, chocolate, and coffee.

Processing Aids

L-Cysteine, L-glutamyl-L-cysteine, L-cysteinylglycine and glutathione (γ-L-glutamyl-L-cysteinyl-glycine) are naturally present in wheat flour. In certain dough processes the flour can be enriched with these molecules or with mixtures of these compounds with oxidizers to produce desired properties in the dough or to reduce mixing times. MSG, naturally present in cabbage and other products of either vegetable or animal origin, has a 'meaty' taste and can be added in the crystalline form as a flavor enhancer. (*See* **Flour:** Dietary Importance.)

Amino Acids in Artificial Sweeteners

Aspartame is the dipeptide L-aspartyl-L-phenylalanine methylester. This molecule is about 200 times sweeter than sucrose with almost no aftertaste. It degrades slowly under acid conditions at room temperature and rapidly at higher temperatures, making it unsuitable for baked foods. Phenylketonurics cannot tolerate this sweetener because of the phenylalanine content. (*See* **Sweeteners:** Intensive.)

See also: **Amino Acids**: Metabolism; **Browning**: Nonenzymatic; **Enzymes**: Functions and Characteristics; **Exercise**: Muscle; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Flour**: Dietary Importance; **Hormones**: Thyroid Hormones; **Migraine and Diet**; **Niacin**: Physiology; **pH – Principles and Measurement**; **Protein**: Chemistry; Food Sources; Synthesis and Turnover; **Scurvy**; **Selenium**: Physiology

Further Reading

Bell EA (1973) Amino acids of natural origin. In: Hey DH and John DI (eds) *Amino Acids, Peptides and Related Compounds*, pp. 1–16. London: Butterworth.

Brockway B (1993) Amino acids, properties and occurrence. In: Macrae R, Robinson RK and Sadler MJ (eds) *Encyclopedia of Food Sciences and Nutrition*, pp. 146–153. London: Academic Press.

Damodaran S (1996) Amino acids, peptides and proteins. In: Fennema OR (ed.) *Food Chemistry*, 3rd edn. pp. 321–429. New York, NY: Marcel-Dekker.

Meister A (1965) *The Natural Amino Acids*. Biochemistry of the Amino Acids, 2nd edn. pp. 1–167. London: Academic Press.

Morrey RK, Granner DK, Mayes PA and Rodwell VW (1996) *Harper's Biochemistry*, 24th edn. Stanford, CN: Appleton & Lange.

Vickery HB (1972) The history of the discovery of the amino acids. II. A review of amino acids described since 1931 as components of native proteins. In: Anfinsen CB, Edsall JT and Richards FM (eds) *Advances in Protein Chemistry*, vol. 26, pp. 82–162. New York, NY: Academic Press.

Determination

A P Williams, Leatherhead Food Research Association, Leatherhead, UK

Background

Analysis of foods to determine their amino acid composition has long been of importance nutritionally, but there are an increasing number of other applications, including detection of adulteration and potentially toxic amino acids produced by new food-processing technologies. This article will review current methods of the determination of amino acids.

Isolation of Free and Protein Amino Acids

Although determination of the total amino acid composition of foods, after protein hydrolysis, is the prime requirement, it is sometimes necessary to determine free amino acids in foods, beverages,

physiological fluids, and tissues. To do this, it is necessary to remove any protein before analysis. Methods of deproteinizing include precipitation with acids or alcohols, high-speed centrifugation, ultrafiltration, ion exchange, and equilibrium dialysis. None is perfect, but precipitation with sulfosalicylic acid is the most popular method although not always suitable for all derivatization procedures.

It may be necessary to remove nonprotein substances from samples before hydrolysis, since they can affect the accuracy of the analysis or damage stationary phases by irreversible adsorption. Since the extraction procedures are lengthy and can result in losses of protein, they are normally only used when these substances are present in high concentrations, e.g., lipids in mechanically separated meat. Such samples are homogenized with 15 ml g^{-1} of an acetone/chloroform (3:1) mixture and filtered on a Buchner funnel until air dry. In addition to lipids, it is claimed that this procedure removes nucleic acids and most of the carbohydrates. Nucleic acids may also be removed by heating the lipid-free sample with 10% NaCl at 85 °C, removing the NaCl with water and drying with acetone. After hydrolysis or deproteinization, nonprotein substances can interfere with precolumn derivatization procedures used for reversed-phase chromatography (RPC) and gas–liquid chromatography (GLC). Lipids can be extracted with chloroform, and salts, carbohydrates, and acids can be extracted by cation exchange followed by recovery of the amino acids with 2 M NH_4OH. The strength of the NH_4OH is critical since concentrations > 2 M can lead to degradation of amino acids. (*See* **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography.)

Hydrolysis Procedures

Acids, alkalis, or enzymes may be used for protein hydrolysis. Alkaline hydrolysis is generally only used in the determination of tryptophan, since other amino acids are degraded.

Enzymatic hydrolysis Enzymatic hydrolysis is rarely used except for the determination of glutamine and asparagine, which are converted to aspartic and glutamic acids together with ammonia by acid hydrolysis. Since no single protease will hydrolyze all the peptide bonds in proteins, the procedure is lengthy, and there is evidence of contamination by amino acids derived from the enzymes. A better method is to analyze the sample before and after treatment with bis(1,1-trifluoroacetoxy)iodobenzene, which converts the carboxamide residues to their corresponding amines. (*See* **Enzymes**: Uses in Analysis.)

Acid hydrolysis The most favored procedure involves heating the protein with excess 6 M HCl under reflux or in a sealed tube *in vacuo* or under nitrogen at 110 °C for 24 h. After filtration, the HCl must be removed, usually by rotary evaporation or neutralization with NaOH, before analysis. However, this procedure is a compromise, since no one method can provide satisfactory values for all amino acids. The problem of tryptophan has already been mentioned, but cystine, cysteine, and methionine undergo variable degradation through oxidation during acid hydrolysis, and an alternative procedure must be used. Tyrosine losses can also occur due to oxidation, but this may be reduced by the addition of phenol to the HCl.

There are smaller, but progressive, losses of threonine and serine, which may be compensated for by corrections of 5 and 10%, respectively, or more precisely by hydrolyzing for 24, 48, and 72 h and calculating to zero time. Since isoleucine and valine are difficult to liberate completely in 24 h, for the most accurate values, it is again necessary to hydrolyze for 24, 48, and 72 h and then to calculate to infinite time. This is rarely carried out for foods.

To obtain the ideal amino acid analysis, much time and care must be spent on the hydrolysis stage. Since the analysis time has been reduced from 24 h to around 30 min, there have been many efforts to automate and reduce the time of the hydrolysis. Commercial systems are available for 42 samples to be hydrolyzed in 24 h at 110 °C using standard liquid-phase HCl. The same number can be hydrolyzed in 1 h at 150 °C using vapor-phase hydrolysis. In liquid-phase hydrolysis, HCl is added directly to the sample. In vapor-phase hydrolysis, tubes containing the sample are sealed into a larger vessel containing HCl. As the vessel is heated, the HCl vaporizes so that only the vapor comes into contact with the sample. This has the advantage of preventing contamination from amino acids present in all but the highest-purity HCl. A recent development is microwave irradiation in which samples can be hydrolyzed in liquid phase HCl at 180 ± 5 °C in 5 min in a microwave oven. Special tubes that can resist high temperatures and pressures must be used. In common with most new innovations developed for use with pure proteins, care should be taken in applying them to foods where the presence of carbohydrates often results in losses of amino acids during hydrolysis.

Chromatographic Methods

After hydrolysis or deproteinization, it is necessary to separate the amino acids from each other, and for this, chromatography is the method of choice.

Chromatographic separations may be of three types: column chromatography in which the stationary phase is packed into glass or metal columns; thin-layer chromatography (TLC) in which the stationary phase is coated on to inert plates; and paper chromatography in which the stationary phase is supported by the cellulose fibers of a paper sheet. Neither of the latter two methods is particularly quantitative, and neither is used to any great extent, although TLC is used to monitor inborn errors of metabolism. Column chromatography consists of GLC and high-performance liquid chromatography (HPLC). (*See* **Chromatography**: Principles; Thin-layer Chromatography.)

High-performance Liquid Chromatography

HPLC can be subdivided into methods involving postor precolumn derivatization.

Postcolumn derivatization Postcolumn derivatization involves separating underivatized amino acids on a chromatographic column, mixing a derivatization reagent with the eluent from the column, and passing the mixture through a reaction coil and then through a detection system (spectrophotometer or fluorometer). This type of HPLC is usually carried out using ion-exchange chromatography (IEC) with sulfonated polystyrene cation exchange resins as the stationary phase and aqueous sodium citrate (for hydrolysates) or lithium citrate (for physiological fluids) mobile phases. This, when coupled with derivatization with ninhydrin, is the traditional method and is still considered to be the best. Other derivatizing reagents, *o*-phthalaldehyde (OPA), fluorescamine, dabsyl chloride (DABS-Cl), and 4-fluoro-7-nitro-2,1,3-benzoxadiazole, have been used to increase sensitivity, as the derivatives formed can be detected by fluorescence.

Precolumn derivatization In precolumn derivatization, the mixture of amino acids is treated with a reagent to form derivatives, which are highly fluorescent or ultraviolet-absorbing and can be separated by RPC. RPC, a fairly recent innovation to amino acid analysis, is a partition system in which the mobile phase is more polar than the stationary phase. The most popular stationary phases are octadecyl-bonded silicas with acetate buffer as the mobile phase and a gradient of acetonitrile or methanol.

Many derivatizing reagents have been used, e.g., OPA, DABS-Cl, 1-fluoro-2,4-dinitrobenzene (FDNB), dansyl chloride (DNS-Cl), phenylisothiocyanate (PITC), 9-fluorenylmethyl chloroformate (FMOC), 4-*N,N*-dimethylaminoazobenzene-4′-isothiocyanate, 1-fluoro-2,4-dinitrophenyl-5-L-alanine amide, and 4-chloro-7-nitro-2,1,3-benzoxadiazole (NBD-Cl). Although they give a high sensitivity, there are several problems with all of these reagents (Table 1). Cleanup and derivatization are laborious but can be automated in commercial instruments, where DABS-Cl, PITC, and FMOC are most commonly used.

Gas–Liquid Chromatography

In GLC, the stationary phase is a liquid such as silicone grease supported on an inert granular solid, and the mobile phase is an inert gas (N, He, Ar). Since

Table 1 Comparison of IEC, RPC, and GLC methods of amino acid analysis

	IEC	*RPC*	*GLC*[a]
Analysis time (min)			
Hydrolysates	30	30	10
Physiological fluids	90	50	35
Sensitivity (pmol)	50[b]	10	0.01
Disadvantages	Expensive instruments Complex mobile phases Low sensitivity	Some multiple or unstable derivatives Complex purification and derivatization Interference from salts, lipids, and reagent contaminants Shorter column life Resolution poor for some amino acids	Complex purification and derivatization Interference from carbohydrates and lipids
Recommended applications	Protein hydrolysates Complex mixtures	Individual amino acids Peptide hydrolysates Screening metabolic disorders Enantiomeric analysis	Enantiomeric analysis Identification of amino acids

[a] *N*-HFB isobutyl esters, capillary column, and electron capture detector.
[b] 5 pmol with OPA replacing ninhydrin.

amino acids are not volatile, they must be converted to volatile derivatives before analysis. This is not difficult, but there are even more derivatives than for RPC. However, the *N*-trifluoroacetyl (*N*-TFA) *n*-butyl and *N*-heptafluorobutyl (*N*-HFB) isobutyl esters appear to be the best. Derivatization is laborious, although it has recently been automated. The major difference between these derivatives is their chromatographic separation. All the *N*-HFB isobutyl esters of protein hydrolysates can be readily separated on a single methylsilicone-packed column (e.g., OV 101, or 3% SE30 coated on 100–200 mesh Gaschrom Q) in 35 min or in less than 10 min with a capillary column (DB-1). Capillary columns are recommended for physiological fluids, with analysis taking 1 h. With the popular *N*-TFA *n*-butyl esters, it is impossible to separate all the amino acids on a single column. A column of 1% OV 7 + 0.75% SP2401 on 100–200 mesh Gaschrom Q is used for histidine, arginine, and cystine, and one of 0.65% EGA on 80–100 mesh Chromasorb W for the remainder. A major advantage of GLC is its sensitivity and specificity; since the mobile phase is gaseous, flame ionization, electron capture, and nitrogen-specific detectors can be used. It is possible to detect femtomole levels of amino acids as their *N*-HFB isobutyl esters using capillary columns and electron capture detectors or mass spectrometry. Another advantage of GLC is that it can be linked with a mass spectrometer to provide confirmation of the identity and purity of peaks.

Comparison of IEC, RPC, and GLC

A good agreement between these methods has often been claimed but usually in comparative studies with pure proteins. The choice of method is difficult and often depends on the applications, sensitivity, and urgency required. Table 1 summarizes their main features. It is often suggested that a major advantage of RPC and GLC is that the instruments can be used for other analyses. If the requirement for amino acid analysis is small, and RPC and GLC instruments and expertise are already available, these would be the methods of choice. The traditional, well-tried method is IEC followed by derivatization with ninhydrin. This is ideal for protein hydrolysates and complex mixtures where maximum accuracy and reproducibility are required, and there is no shortage of sample. The sensitivity, analysis times, and cost of the instruments have improved considerably. Sample purification is rarely necessary, since it is almost completely insensitive to sample matrix. The time taken for sample purification and derivatization for RPC and GLC should not be ignored, and recent automation of these stages adds to the overall cost. The higher sensitivities of RPC and GLC are ideal where the amount of sample is limited, but care should be taken to avoid contamination by amino acids from reagents and glassware.

A major problem with RPC is the choice of derivative, with little agreement on which is the best, since none is ideal. Ideally, it would be better if pre- and postcolumn derivatization could be avoided altogether. The separation of underivatized amino acids, at picomole levels, using anion-exchange chromatography followed by pulsed amperometric detection has recently been proposed for protein hydrolysates. This uses complex aqueous mobile phases and would be difficult to apply to more complex mixtures. Whichever method is selected, it should be stressed that the hydrolysis of proteins and the deproteinization of physiological fluids are still major problems.

Methods for Specific Amino Acids

Methionine and Cystine

These amino acids are difficult to estimate because they are present in low concentrations and undergo oxidation to multiple derivatives during acid hydrolysis. To overcome this problem, controlled oxidation of methionine to methionine sulfone and cystine to cysteic acid must be carried out with performic acid prior to acid hydrolysis. This involves oxidizing the sample with performic acid–hydrogen peroxide for 16 h at 0 °C. After removal of excess oxidizing reagents, acid hydrolysis is carried out as normal.

Tryptophan

Tryptophan is also present in low concentrations and extensively degraded during acid hydrolysis. However, there is no measurable end product, and so it is normal to use alkaline hydrolysis specifically for tryptophan analysis. Sodium, barium, or lithium hydroxides may be used at concentrations ranging from 4 to 6 M, with additives such as maltodextrin, starch, or thiodigycol often recommended to reduce tryptophan losses. Hydrolysis may be for 8 h at 145 °C or 20 h at 110 °C using polypropylene vessels. Ideally, the tryptophan should be separated from interfering compounds, e.g., lysinoalanine (LAL) by IEC or RPC. The latter takes only a few minutes, and precolumn derivatization is unnecessary, since tryptophan can be detected by its native fluorescence.

Tryptophan has also been estimated by acid hydrolysis of intact proteins in the presence of ninhydrin with which it reacts before it can be degraded. Corrections must be made for tyrosine.

Lysine and Available Lysine

There has been considerable interest in a specific method for lysine. However, even RPC methods take 15 min and would appear to have little advantage over complete analysis. However, available lysine is usually determined separately. If foods are subjected to heat during processing, lysine can become nutritionally unavailable if its free ε-amino group reacts with, for example, carbohydrates, forming bonds, which are resistant to digestive enzymes. Available lysine can be measured by treating proteins with FDNB or 2,4,6-trinitrobenzene sulfonic acid, which reacts with the free ε-amino groups of lysine to form either ε-dinitrophenyllysine (DNP-lysine) or ε-trinitrophenyllysine (TNP-lysine). The lysine derivatives can be separated by RPC from other DNP or TNP amino acids in 15–20 min.

Lysinoalanine

Alkaline treatment of proteins is used extensively in food processing, e.g., in the preparation of textured proteins. This results in the formation of amino acids such as LAL, ornithinoalanine, lanthionine, and β-aminoalanine. LAL is formed by reaction of the ε-amino group of lysine with the double bond of dehydroalanine, which can result in the loss of available lysine and toxicity problems. LAL can be measured in 16 min by RPC after derivatization with DNS-Cl. Alternatively, LAL and all the common amino acids can be determined by IEC, followed by ninhydrin derivatization in 110 min.

3-Methylhistidine

3-Methylhistidine is an analog of histidine found mainly in skeletal muscle. Methylation of histidine occurs after its incorporation into the peptide chains of actin and myosin. Since, after the catabolism of these proteins, 3-methylhistidine is not recycled but quantitatively excreted in the urine, it has been proposed as an index of muscle protein turnover. It has also been used to determine the meat content of foods, where vegetable or microbial protein has been added for economic or fraudulent reasons. 3-Methylhistidine may be estimated after derivatization with fluorescamine. Histidine and 3-methylhistidine give acid-stable fluorescent derivatives, which can be separated by RPC in 20 min.

Hydroxyproline

If 3-methylhistidine is used as an index for meat protein, hydroxyproline must also be measured to correct for the collagen content. Collagen has a low content of essential amino acids, and excessive amounts in foods reduce their nutritive value. New technologies also make it possible to incorporate collagenous materials into meat products at high levels. There are histochemical, histological, and immunological techniques available, but for routine purposes, there is a British Standards method available. This involves oxidation of the hydroxyproline with chloramine T to pyrrole, followed by photometric determination of the reaction product of the pyrrole with *p*-dimethylaminobenzaldehyde. An RPC method taking 10 min also exists using NBD-Cl derivatization.

D- and L-Amino Acids

During alkali or heat treatment L-amino acids in proteins are racemized to their D-isomers. Since most D-amino acids cannot be utilized by humans, and some are toxic, their determination is of considerable interest. The D- and L-isomers have identical chemical properties and must first be converted to diastereomeric dipeptides by reaction with chiral (optically active) reagents before chromatography or separated by chiral stationary or mobile phases. The leucyl-DL-aspartic acid dipeptides are prepared by coupling aspartic acid with L-leucine *N*-carboxy anhydride (NCA). Basic amino acids are coupled with L-glutamine NCA. *N-t*-Butoxycarbonyl-L-cysteine and OPA are other chiral agents that have been used. Separation of 21 enantiomers in 40 min can be achieved by RPC and fluorescence detection. Precolumn derivatization can be avoided by using a chiral mobile phase, a copper–proline (Cu–Pro) complex, with IEC. Diastereomeric Cu–amino acid complexes are formed on the column and detected by postcolumn derivatization with OPA. Alternatively, an RPC method using a chiral stationary phase, in which the Cu–Pro or Cu–hydroxyproline complex is bound to a silica stationary phase, can be used.

D- and L-amino acids can also be determined by GLC with the introduction of a second, optically pure, asymmetric center into the molecule to make diastereoisomers, which can be separated on conventional packed columns. The use of (+)-butan-2-ol to form (+)-2-butyl esters appears to be the best method. Alternatively, the enantiomers, converted to normal derivatives, e.g., *trans*-fatty acid isopropyl esters, can be separated on capillary columns coated with chiral stationary phases, e.g., *N-trans*-fatty acid-L-valyl-L-valine cyclohexyl ester.

See also: **Chromatography**: Principles; Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; **Enzymes**: Uses in Analysis

Further Reading

Bech-Anderson S, Mason VC and Dhanoa MS (1989) Hydrolysate preparation for amino acid determinations in feed constituents 9. Modifications to oxidation and hydrolysis conditions for streamlined procedures. *Zeitschrift für Tierphysiologie, Tierernährung und Futtermittelkunde* 63: 188–197.

Cohen SA and Strydom DJ (1988) Amino acid analysis utilizing phenylisothiocyanate derivatives. *Analytical Biochemistry* 174: 1–16.

Deyl Z, Hyanek J and Horakova M (1986) Profiling of amino acids in body fluids and tissues by means of liquid chromatography. *Journal of Chromatography* 379: 177–250.

Gehrke CW and Zumwalt RW (1987) Symposium on chromatography of amino acids. *Journal of the Association of Official Analytical Chemists* 70: 146–147.

Williams AP (1986) General problems associated with the analysis of amino acids by automated ion-exchange chromatography. *Journal of Chromatography* 373: 175–190.

Williams AP (1988) Determination of amino acids. In: Macrae R (ed.) *HPLC in Food Analysis*, 2nd edn., pp. 441–470. London: Academic Press.

Metabolism

R M B Ferreira and A R N Teixeira, Universidade Técnica de Lisboa, Lisboa, Portugal

Introduction

Naturally occurring amino acids may be conveniently grouped into three categories: protein amino acids (sometimes known as 'standard,' 'primary,' and 'normal'), uncommon amino acids, and nonprotein amino acids. Protein amino acids are those that are coded for in the genes and incorporated directly into proteins. For some time it seemed well established that all proteins, whatever their origin, were constructed from the same set of 20 amino acids. Recent studies, however, have shaken the foundation of this classical dogma. It now seems that the genetic code may dictate the incorporation of more than 20 amino acids. Thus, for example, selenocysteine and phosphoserine, previously considered to be uncommon amino acids, can be directly incorporated into the polypeptide chain. All protein amino acids are α,L-amino acids. It is not clear why amino acids incorporated by organisms into proteins are of the L form, since L-amino acids have no obvious inherent superiority over their D isomers for biological function. Thus, in this article, unless otherwise stated, an L configuration is assumed.

The 20 classical protein amino acids may be grouped into several classes reflecting important characteristics of their side chains: straight aliphatic amino acids (glycine, alanine), branched-chain amino acids (valine, leucine, isoleucine), hydroxy amino acids (serine, threonine), sulfur-containing amino acids (cysteine, methionine), aromatic amino acids (phenylalanine, tyrosine), heterocyclic amino acids (tryptophan, histidine), basic amino acids (lysine, arginine), acidic amino acids and their amides (aspartate, glutamate, asparagine, glutamine), and imino acid (proline). Amino acids can also be classified on the basis of the polarity of their side chains. (*See* **Protein: Chemistry.**)

Analyses of proteins have revealed that they contain well over 100 different amino acids. The occurrence of uncommon amino acids in proteins is the result of posttranslational, covalent modification of protein amino acids. Cystine, for example, is formed by the posttranslational cross-linking of two cysteine residues. Citrulline, *N*-formylmethionine, *O*-galactosylserine, and *N*-acetylthreonine constitute other examples of amino acids found in proteins.

The amino acids found in proteins are by no means the only ones to occur in living organisms. Thus the term 'nonprotein amino acids' is used to include those naturally occurring amino acids which are present in free or combined forms but not in proteins. Over 200 nonprotein amino acids are known, most of them occurring in plants and frequently limited, in each case, to certain taxonomic groups. Some, such as cystathionine and saccharopine, fulfill important roles in the primary metabolic pathways. However, the great majority of these compounds have obscure functions and are generally regarded as secondary products. Many of the nonprotein amino acids from plants are known to be toxic to animals, plants, and microorganisms. Some accumulate to exceptionally high levels, as in the case of 5-hydroxytryptophan, canavanine, or 3,4-dihydroxy-phenylalanine, which may constitute up to 14% of the seed weight in some Leguminosae species. Storage and protection against predation are probably two of the many possible roles that these amino acids play in plants.

Essential and Nonessential Amino Acids

Organisms differ greatly in their abilities to synthesize the amino acids required for protein synthesis. Many microorganisms and plants are entirely self-sufficient in that they can synthesize the entire basic set

of protein amino acids. However, the bacterium *Leuconostoc mesenteroides* can synthesize only four of the protein amino acids, whereas *Lactobacillus*, which flourishes in milk, must be provided with all amino acids required for protein synthesis. Mammals are intermediate, being able to synthesize about half of the protein amino acids. Amino acids which cannot be synthesized by an organism in adequate amounts are called essential or indispensable because they must be supplied by the diet. Those which can be synthesized by an organism from readily available precursors in sufficient amounts to meet its needs are not required in the diet and are referred to as nonessential or dispensable amino acids. A consensus of current nutritional opinion indicates that the L isomers of 10 amino acids – arginine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine – are considered to be essential for mammals, including humans.

The designations of essential and nonessential amino acids refer to the needs of an organism under a particular set of conditions. Thus essential amino acids are often species-specific, i.e., the set of amino acids that are essential for a particular organism is not necessarily the same for other organisms. The essential amino acid requirements depend on a variety of factors, including age, sex, physiological conditions, and diet. Arginine (see *Urea Cycle*, below) and histidine are synthesized by humans in quantities sufficient to meet the needs of an adult but not those of a growing child. These amino acids have been termed semi- or half-essential. Adults have proportionally lower demands for essential amino acids than infants and children because adults are able to recycle such amino acids efficiently, whereas infants need them for tissue growth. When the ratio of total essential amino acids required to total protein required is considered, it is 0.37 for infants and 0.15 for adults. Tyrosine and cysteine are considered nonessential amino acids for mammals only as long as the diet contains adequate amounts of, respectively, phenylalanine and methionine; this is because, in mammals, tyrosine is formed in one step directly from phenylalanine, and cysteine derives its sulfur uniquely from dietary methionine (see *Synthesis of Amino Acids*, below). Hence the apparent quantitative phenylalanine requirement is actually a requirement for phenylalanine plus tyrosine, whereas that of methionine is for methionine plus cysteine.

The essential amino acids include those with complex structures, which are formed by complex routes, whereas the nonessential are those whose syntheses are the simplest and whose intermediate precursors are always present in all organisms. Indeed, 59 enzymes are required by prokaryotic cells to synthesize the essential amino acids for humans, but only 15 are required for the nonessential. Essential amino acids other than lysine and threonine (the only amino acids that, in mammals, do not participate in transamination reactions; see *Transamination and Deamination*, below) can be replaced by their α-keto analogs in the diet. This indicates that the carbon skeleton of the essential amino acid is the fundamental part of the amino acid molecule.

A deficiency of even one essential amino acid in the diet of an organism results promptly in a negative nitrogen balance, i.e., total nitrogen excretion exceeding total nitrogen intake, indicating that tissue protein is being degraded and used to supply the missing amino acid for those 'high-priority' proteins that need to be continually synthesized. The remaining amino acids then accumulate and are shunted into catabolic pathways – hence the loss of nitrogen. Under these conditions protein synthesis is severely inhibited because the ribosome–messenger ribonucleic acid (mRNA)–nascent polypeptide complex must suspend its operation at the point where the missing amino acid should be incorporated. Thus the degree of negative nitrogen balance is similar whether only one, several, or all of the essential amino acids are missing. This is logical because nearly all body proteins contain all the essential amino acids.

Regardless of the organism or of the essential amino acids considered, the net result of its deficiency inevitably involves a decreased growth rate, increased susceptibility to disease, and biochemical dysfunctions along with ultimate death. However, deficiencies of a specific essential amino acid may also result in disturbances characteristic of that particular amino acid. This is the case for tryptophan in nicotinic acid formation, and lysine in the formation of hydroxylysine in the biosynthesis of collagen. (*See* **Niacin:** Physiology.)

Amino Acid Biosynthesis

Amino acid metabolism involves the dynamic occurrence of anabolic and catabolic pathways. It is sometimes difficult to distinguish between catabolic and anabolic reactions because the catabolism of one amino acid may be involved in the biosynthesis of another. Because of the complexity and multiplicity of these pathways only a simplified version of the major routes will be considered.

Nitrogen Assimilation

Inorganic nitrogen is incorporated into organic nitrogen compounds as ammonium. This process, called

ammonium assimilation, leads to the formation of glutamate, glutamine, and carbamoyl phosphate. Utilization of the nitrogen of carbamoyl phosphate is limited to the biosynthesis of arginine (see *Urea Cycle*, below) and pyrimidine nucleotides. Essentially, all other nitrogen atoms of amino acids and other nitrogenous compounds are derived directly or indirectly from glutamate or glutamine.

The reductive amination of 2-oxoglutarate by ammonium ions (NH_4^+), catalyzed by glutamate dehydrogenase, is the simplest route to the formation of α-amino groups:

$$\text{2-Oxoglutarate} + NH_4^+ + \text{NADPH} + H^+ \longrightarrow \text{glutamate} + H_2O + \text{NADP}^+$$

(NADPH, $NADP^+$ represent the reduced and oxidized forms of the nicotinamide adenine dinucleotides.) This reaction occurs in plants and bacteria only under situations of high NH_4^+ concentration, which is toxic to cells and does not happen frequently under natural conditions, implying that this enzyme does not play a significant role in primary ammonium assimilation. Under natural conditions, the glutamate synthase cycle constitutes the major pathway by which plants and microorganisms assimilate NH_4^+. This cycle involves the sequential action of two enzymes: glutamine synthetase, which catalyzes the adenosine triphosphate (ATP)-dependent amidation of glutamate to produce glutamine, and glutamate synthase, which catalyzes the reductive transfer of the δ-amino group of glutamine to 2-oxoglutarate, to produce two molecules of glutamate. The sum of these reactions is as follows:

$$\text{2-Oxoglutarate} + NH_4^+ + \text{NADPH} + \text{ATP} \longrightarrow \text{Glutamate} + \text{NADP}^+ + \text{ADP} + P_i$$

(NADPH or reduced ferredoxin; $NADP^+$ or oxidized ferredoxin; P_i inorganic phosphate)

Synthesis of Amino Acids

The biosyntheses of protein amino acids arise as branching pathways from a few key intermediates in the central metabolic routes that are common to all cells, namely glycolysis, the pentose phosphate pathway, and the tricarboxylic acid (TCA) cycle. It is convenient to divide the 20 classical protein amino acids into six biosynthetic families according to the central metabolites that serve as starting points for their syntheses. (*See* **Glucose:** Function and Metabolism.)

The glutamate family 2-Oxoglutarate, a TCA cycle intermediate, serves as the starting point in the formation of glutamate and the other members of the glutamate family, glutamine, proline, arginine and, in the fungi and *Euglena*, lysine:

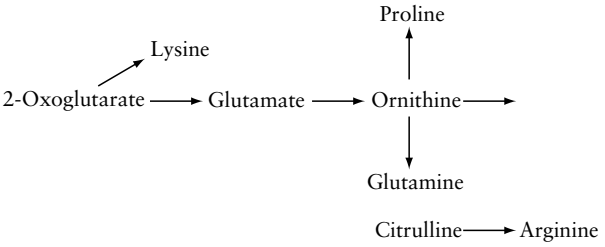

The serine family 3-Phosphoglycerate, an intermediate of the glycolytic pathway, serves as a precursor for the serine family of amino acids, comprising serine and its derivative amino acids, glycine and cysteine:

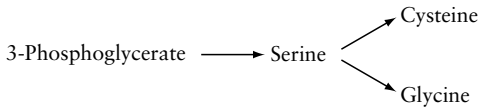

In common with carbon and nitrogen, environmental sulfur is available to organisms in the form of inorganic compounds. Sulfur assimilation is largely confined to plants and microorganisms since higher animals, unable to assimilate inorganic sulfur, must rely on ingested methionine and cysteine. Thus, whilst some microorganisms can reduce sulfate, thiosulfate or elemental sulfur, higher plants use sulfate for amino acid synthesis. Reductive assimilation of sulfate, i.e., incorporation of sulfate sulfur into thiol groups of amino acids and other organic compounds, requires the reduction of sulfate to sulfite and, subsequently, of sulfite to sulfide.

Two major pathways exist for the biosynthesis of cysteine in living organisms. Plants and microorganisms, which utilize H_2S as the source of sulfur, synthesize cysteine by the direct sulfhydrylation pathway. However, in mammals, which synthesize cysteine by the transsulfuration pathway, cysteine derives its carbon skeleton from serine but its sulfur atom is obtained uniquely from methionine.

The aspartate family Oxaloacetate, an intermediate of the TCA cycle, provides the carbon skeleton for the synthesis of six different amino acids: aspartate, asparagine, lysine (in bacteria and plants but not in fungi), methionine, threonine, and isoleucine, which constitute the aspartate family of amino acids:

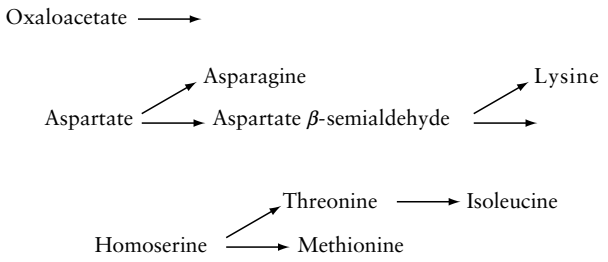

However, isoleucine is frequently included in the pyruvate family since four of its five biosynthetic enzymes are common to the valine pathway. Methionine derives its sulfur atom from cysteine.

The pyruvate family The pyruvate family of amino acids includes alanine, valine, and leucine:

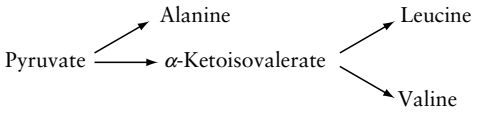

Pyruvate, a glycolytic intermediate, gives rise to the carbon skeletons of alanine and valine, and to four of the six carbons of leucine. In addition, pyruvate also donates two carbon atoms to the synthesis of isoleucine and, on average, 2.5 carbons to the synthesis of lysine in bacteria and plants. As mentioned earlier, isoleucine, a member of the aspartate family, is most conveniently considered along with valine, since the biosynthesis of both involves a common set of enzymes.

The aromatic family Phenylalanine, tyrosine, and tryptophan, which comprise the aromatic family of amino acids, are synthesized from phosphoenolpyruvate and erythrose 4-phosphate, intermediates of glycolysis and the pentose phosphate pathway, respectively:

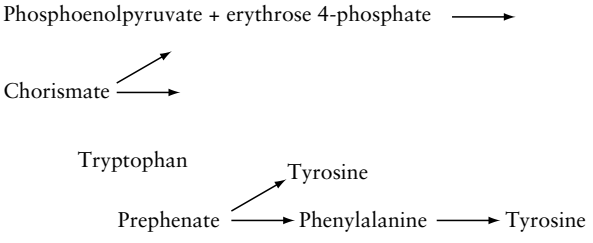

These amino acids are synthesized by a branched pathway in which chorismate is the major branch-point metabolite. Chorismate is synthesized by a seven-step pathway, often referred to as the shikimate or common aromatic pathway, to build the benzene ring.

In some organisms, including humans, tyrosine can be synthesized by hydroxylation of phenylalanine in a reaction catalyzed by phenylalanine hydroxylase. This reaction, the only known reaction of aromatic amino acid biosynthesis in animals, accounts for the nonessentiality of tyrosine in mammals, and is not reversible, which explains why tyrosine cannot replace the nutritional requirement for phenylalanine.

The histidine family Histidine is synthesized from ribose 5-phosphate, a pentose phosphate pathway intermediate, by a pathway unrelated to those of the other amino acids.

Uncommon amino acids are synthesized by post-translational, covalent modification of protein amino acids. These modifications, which may either be enzyme-catalyzed or occur spontaneously, involve a variety of chemical processes including glycosylation, phosphorylation, hydroxylation, methylation, acetylation, and amidation. Hydroxyproline and hydroxylysine, for example, are two uncommon amino acids almost exclusively associated with collagen. The preformed amino acids, as they may occur in ingested food protein, are not incorporated into collagen since there are no transfer RNAs (tRNAs) capable of recognizing and inserting them into a nascent polypeptide chain. Rather, these amino acids are synthesized by hydroxylation of prolyl and lysyl residues, in reactions catalyzed by prolyl hydroxylase and lysyl hydroxylase, respectively. 3-*N*-Methylhistidine constitutes another example of an uncommon amino acid. This amino acid, found in actin and myosin, is synthesized by methylation of a histidyl residue in an enzymatic reaction that utilizes *S*-adenosylmethionine as the methyl group donor. This process is clearly highly specific because only one histidine out of the 35 found in the heavy chain of myosin is methylated. Furthermore, the extent of methylation varies with a number of factors, including age and diet, and is generally not complete in that specific residues are found to be methylated in only a fraction of the myosin molecules.

Relatively little work has been done on the biosynthesis of nonprotein amino acids. There are four different ways by which these amino acids may be formed: (1) as intermediates in protein amino acid synthesis; (2) modification of protein amino acids; (3) modification of pathways to protein amino acids; (4) novel pathways.

Regulation of Amino Acid Biosynthesis

Living cells contain a small pool of free protein amino acids resulting from a precise and coordinated control of the rates at which each amino acid is synthesized and degraded. The mechanisms that control amino acid synthesis vary widely in the various pathways and, for the same pathway, in different organisms. Most studies have been performed with

microorganisms, in particular with *Escherichia coli, Bacillus subtilis*, and *Salmonella typhimurium*. The regulation of amino acid biosynthesis occurs at two levels: regulation of enzyme activity or metabolite flow over a pathway and regulation of enzyme amount. (*See* **Enzymes:** Functions and Characteristics.)

Control of enzyme activity The control over the flow of metabolites into an amino acid biosynthetic pathway can be efficiently achieved by blocking the first, usually irreversible step which is specific for that amino acid. The inhibition of the committed step by the end product, i.e., the amino acid itself, constitutes the simplest kind of feedback inhibition. Some examples include the regulation of the biosynthesis of proline, arginine, histidine, and of the branched-chain amino acids. Alanine, aspartate, glutamate, and glycine are four amino acids for which no form of feedback inhibition is known. However, these amino acids are usually in equilibrium, by means of reversible reactions, with compounds that are key intermediates in the central metabolic routes. Metabolite flow into the biosynthetic pathways of the remaining 16 protein amino acids is controlled by several types of feedback inhibition.

Sequential feedback inhibition regulates the synthesis of aromatic amino acids in *B. subtilis*. The first divergent steps in the synthesis of these amino acids are inhibited by their final products. If all three are present in excess, the branch-point intermediates chorismate and prephenate will accumulate, inhibiting the first common enzyme in the overall pathway, i.e., the first reaction of the shikimate pathway.

Enzyme multiplicity regulates the synthesis of aromatic amino acids in *E. coli, S. typhimurium*, and *Neurospora crassa* and the synthesis of the aspartate family of amino acids in *E. coli*. In the former, those organisms possess three isoenzymes which catalyze the first reaction of the shikimate pathway – one inhibited by phenylalanine, one by tyrosine, and one by tryptophan. In the latter, three forms of the enzyme catalyzing the first reaction of the pathway leading from aspartate to aspartate β-semialdehyde exist – one inhibited by methionine, one by threonine, and one by lysine.

B. polymyxa and *Rhodopseudomonas capsulata* possess a single enzyme catalyzing the first reaction of the pathway leading from aspartate to aspartate β-semialdehyde, and its regulation is achieved by concerted feedback inhibition. Lysine and threonine alone are only weak inhibitors, but when both are present, a strong synergistic inhibition occurs.

The regulation of *E. coli* glutamine synthetase, a key enzyme in the flow of inorganic nitrogen to organic compounds, is an example of cumulative feedback inhibition. Eight inhibitors are either metabolic end products of glutamine (tryptophan, histidine, carbamoyl phosphate, glucosamine 6-phosphate, cytidine triphosphate and adenosine monophosphate, or AMP) or in some other way indicators of the general status of amino acid metabolism (alanine and glycine). Each of the eight compounds alone gives only partial inhibition, but in combination, with each acting independently of the others, the degree of inhibition is increased until the activity is almost completely switched off when all eight compounds are simultaneously present.

Other ways of controlling enzyme activity include the following: (1) activation of enzyme activity by metabolites; (2) modification of enzymes (e.g. adenylation of certain enzymes may render them more susceptible to feedback inhibition); (3) protein–protein interactions (e.g. activity of multienzyme complexes may change with the amounts of its components present).

Control of enzyme amount The amount of an enzyme may be controlled by a number of different mechanisms: (1) end product repression of enzyme synthesis (e.g., the coordinate repression of the synthesis of all the enzymes involved in histidine biosynthesis in *E. coli* by histidine); (2) substrate induction of enzyme synthesis (e.g., the induction of the synthesis of the first enzyme involved in cysteine biosynthesis in *E. coli* by the product of its reaction); (3) metabolite depression of enzyme synthesis (e.g., the synthesis of all amino acid biosynthetic enzymes is strongly reduced when *E. coli* is grown in a rich medium); (4) regulation of enzyme degradation. Very little is known on this last topic. Nevertheless, the protection of a given enzyme against proteolysis is probably an important regulatory process.

Amino Acid Catabolism

All living cells undergo intracellular protein degradation, with the resulting amino acids being recycled into proteins or degraded oxidatively to yield energy. In microorganisms and plants amino acids are not generally present in excessive amounts. In higher animals, however, where amino acid intake may largely exceed the metabolic needs, amino acids present in excess are not stored or excreted as such. Instead, they are used for energy production. It is estimated that amino acids supply about 15% of the total energy required by an average human adult. This value may be increased under conditions of energy insufficiency or nutritional pathologies. Amino acids can also constitute an important energy

source in plants, during the germination of protein-storing seeds, and in microorganisms, when carbohydrates or fatty acids are not available. This is the case in many bacteria that can grow in media containing amino acids as the source of energy, carbon, and nitrogen. These organisms utilize amino acid catabolic pathways analogous to those of higher animals.

The catabolic metabolism of amino acids is mainly concerned with the separation of the amino groups from the carbon skeletons and the subsequent fate of both the amino groups and the carbon chains. (*See* **Energy:** Measurement of Food Energy.)

Transamination and Deamination

In general, one of the first steps in the degradation of amino acids involves the removal of the α-amino group to give the corresponding 2-oxo acid. Two distinct types of reactions are known to accomplish this task: transamination and deamination.

Transamination, the most common mechanism for deamination of amino acids, involves the transfer of an amino group from a donor amino acid to an acceptor 2-oxo acid, with the formation of a new amino acid and a new oxo acid. Transamination reactions are catalyzed by pyridoxal phosphate-dependent enzymes termed transaminases or, more properly, aminotransferases. These enzymes have a twofold specificity in that they are specific for the acceptor 2-oxo acid but nonspecific for the donor amino acid. Most aminotransferases are specific for 2-oxoglutarate as the acceptor 2-oxo acid, although some may use either pyruvate or oxaloacetate. Accordingly, there are three classes of aminotransferases, which form glutamate, alanine, and aspartate, respectively. More than 50 aminotransferases have been identified. With the exception of lysine and threonine, the α-amino groups of all the amino acids found in proteins can be removed by transamination. Moreover, transamination is not restricted to α-amino groups since, for example, the δ-amino group of ornithine is readily transaminated. Transaminases fulfill central catabolic as well as anabolic functions in the metabolism of several amino acids because they catalyze freely reversible reactions, having equilibrium constants close to unity.

Transamination does not result in a net removal of nitrogen from amino acids. It does, however, allow for the collection of amino groups in glutamate. Oxidative deamination of glutamate by glutamate dehydrogenase results in the liberation of ammonium. The 2-oxoglutarate thus produced can either be used as the acceptor 2-oxo acid in further transamination reactions or enter the TCA cycle. Glutamate is the only amino acid for which a specific and highly active dehydrogenase exists. This pathway, i.e., the concerted action of the aminotransferases and glutamate dehydrogenase, is responsible for most of the ammonium produced by the catabolism of amino acids.

Additional minor routes for the deamination of amino acids are provided by amino acid oxidases, capable of oxidizing most naturally occurring amino acids, and by dehydratases, capable of removing non-oxidatively the amino groups of some amino acids.

Urea Cycle

Plants and microorganisms commonly excrete very little nitrogen. Growth of these organisms is often restricted by a limited availability of nitrogen so that nitrogen liberated by catabolic pathways is usually reassimilated. However, because high concentrations of NH_4^+ are extremely toxic to cells, animals must get rid of the excess ammonium produced by the catabolism of amino acids, either by direct excretion or, when removal of NH_4^+ by simple diffusion is difficult, by conversion to less toxic excretory products. Most terrestrial vertebrates, including mammals, excrete ammonia in the form of urea. Urea is highly soluble in water but nontoxic to cells.

Urea is synthesized by the urea cycle, which is carried out almost exclusively in liver cells. This cycle, discovered by Hans Krebs and Kurt Henseleit in 1932, consists of five sequential enzymatic reactions:

1. $NH_4^+ + HCO_3^- + 2ATP \longrightarrow$ carbamoyl phosphate + 2ADP + P_i + H^+
2. Carbamoyl phosphate + ornithine $\longrightarrow$ citrulline + P_i
3. Citrulline + aspartate + ATP $\longrightarrow$ argininosuccinate + AMP + PP_i
4. Argininosuccinate $\longrightarrow$ arginine + fumarate
5. Arginine + H_2O $\longrightarrow$ ornithine + urea

The sum of these reactions is as follows:

$$NH_4^+ + HCO_3^- + 3ATP + H_2O + \text{aspartate} \longrightarrow \text{urea} + 2ADP + AMP + 2P_i + PP_i + H^+ + \text{fumarate}$$

Virtually all organisms synthesize arginine from ornithine by reactions 2–4. However, only ureotelic organisms are capable of catalyzing the hydrolysis of arginine (reaction 5), the reaction responsible for the cyclic nature of the urea cycle. The synthesis of urea is energetically expensive, requiring the hydrolysis of 4 molecules of ATP per turn of the cycle (2 molecules of ATP are needed to convert AMP to ATP). The fumarate produced is hydrated to malate and oxidized to

oxaloacetate by TCA cycle enzymes. Aspartate is then regenerated from oxaloacetate by transamination. Thus both amino groups of urea originate from amino acids: one is derived from ammonium produced by deamination (reaction 1) and the other is provided by aspartate (reaction 3). Bicarbonate (reaction 1) furnishes the carbon atom of urea. In this respect it is interesting to note that not all the urea produced in the human liver is excreted in the urine – a considerable fraction is hydrolyzed in the colon by bacterial ureases. The mucosa of the human colon is relatively permeable to urea. However, the great majority of the urea molecules is rapidly hydrolyzed within the lumen of the colon with a large proportion of the resulting ammonia nitrogen being absorbed into the portal system or metabolized by the intestinal flora. The ammonium absorbed from the colon may be available for transamination into amino acids in the liver or resynthesized to urea also in the liver, with some of this urea distributed back to the gastrointestinal tract for degradation to ammonia and consequent recycling.

Catabolic Pathways

Once the amino groups of amino acids have been removed, the remaining carbon skeletons are funneled into seven major metabolic intermediates, namely pyruvate, acetyl coenzyme A (acetyl-CoA), acetoacetate, 2-oxoglutarate, succinyl-CoA, fumarate, and oxaloacetate (Table 1), which may be either directly oxidized into carbon dioxide and water by the TCA cycle or reincorporated into glucose or fatty acids. Glycogenic amino acids are those possessing carbon skeletons which generate pyruvate or TCA cycle intermediates and can, therefore, be converted to glucose via gluconeogenesis. In contrast, amino acids possessing carbon skeletons which are metabolized to acetyl-CoA or acetoacetate, precursors of fatty acids and ketone bodies, are termed ketogenic. Recall that, with the exception of some species of plants and microorganisms which possess the glyoxylate cycle, all other organisms lack a pathway for the net synthesis of glucose from acetyl-CoA or acetoacetate. A few amino acids are both glycogenic and ketogenic since portions of their carbon skeletons are converted into carbohydrate derivatives whereas other portions are converted into ketone bodies. Note that the classification presented in Table 1 is not universally accepted because several amino acids are glycogenic under some conditions but ketogenic under others.

Regulation of Amino Acid Catabolism

Microorganisms regulate the level of their amino acid degradative enzymes in different ways:

1. The enzymes are subjected to catabolite repression, i.e., repression of the amino acid catabolic pathway by a carbon and energy source, even in the simultaneous presence of that amino acid as the only source of nitrogen. Thus these enzymes are induced only when carbon and energy limit growth (e.g., the induction of tryptophanase – the enzyme which cleaves tryptophan to yield ammonium, pyruvate, and indole – by tryptophan in *E. coli*).

Table 1 Metabolic fates of the carbon skeletons of amino acids

Amino acid	*Product(s) of catabolism*	*Metabolic fate*
Alanine	Pyruvate	Glycogenic
Arginine → glutamate	2-Oxoglutarate	Glycogenic
Asparagine → aspartate	Oxaloacetate	Glycogenic
Aspartate	Oxaloacetate, fumarate[a]	Glycogenic
Cysteine	Pyruvate	Glycogenic
Glutamate	2-Oxoglutarate	Glycogenic
Glutamine → glutamate	2-Oxoglutarate	Glycogenic
Glycine → serine	Pyruvate	Glycogenic
Histidine → glutamate	2-Oxoglutarate	Glycogenic
Methionine	Succinyl-CoA	Glycogenic
Proline → glutamate	2-Oxoglutarate	Glycogenic
Serine	Pyruvate	Glycogenic
Threonine	Pyruvate	Glycogenic
Valine	Succinyl-CoA	Glycogenic
Isoleucine	Succinyl-CoA, acetyl-CoA	Glycogenic and ketogenic
Phenylalanine → tyrosine	Fumarate, acetoacetate	Glycogenic and ketogenic
Tryptophan	Pyruvate, acetyl-CoA, acetoacetate	Glycogenic and ketogenic
Tyrosine	Fumarate, acetoacetate	Glycogenic and ketogenic
Leucine	Acetyl-CoA, acetoacetate	Ketogenic
Lysine	Acetoacetate	Ketogenic

[a]See text (section on urea cycle).

2. The enzymes are induced when nitrogen limits growth (e.g., the induction of proline oxidase – the enzyme which catalyzes the first step of proline degradation – by proline in *E. coli*, even in the presence of ample carbon supply).
3. The enzymes are induced independently of carbon and energy or nitrogen supply (e.g., the induction of threonine dehydrogenase – an enzyme involved in threonine catabolism – in *E. coli* by growth in leucine, even in the presence of other carbon and nitrogen sources). In some microorganisms, catabolite repression can be bypassed by a nitrogen limitation signal that allows the induction of a particular amino acid catabolic pathway. This nitrogen limitation signal is probably related to the complex regulation mechanisms of glutamine synthetase (see *Regulation of Amino Acid Biosynthesis*, above).

In animal cells, amino acid catabolism is also subjected to control mechanisms. Thus, for example, removal of amino groups from amino acids is regulated mainly by control of glutamate dehydrogenase. This enzyme is allosterically inhibited by ATP and guanosine triphosphate (GTP) and stimulated by ADP and guanosine diphosphate (GDP). Hence, when cellular energy charge is low the rate of amino acid oxidation increases. On the other hand, the urea cycle is controlled by *N*-acetylglutamate. This compound is a positive allosteric effector of carbamoyl phosphate synthetase, which catalyzes the first and rate-limiting step in the pathway. *N*-acetylglutamate is also a precursor of arginine and its synthesis is inhibited by arginine. However, the amino acid catabolic enzymes of animal cells are much more often subjected to a hormonal control than are the microbial enzymes. Thus, for example, the synthesis of tryptophan oxygenase, regulated by adrenal activity, is developmentally controlled so that the enzyme is formed only in certain tissues and at certain times during development. A high-protein diet is also a factor that is known to stimulate the formation of a number of amino acid degradative enzymes in liver, namely urea cycle enzymes and tryptophan oxygenase.

Synthesis of Biologically Important Compounds

In addition to their role in protein synthesis, energy production, and gluconeogenesis, many amino acids serve as precursors for the synthesis of other amino acids and other biologically important compounds.

Many oligopeptides containing up to 20 residues, including hormones, antibiotics, and antitumor agents, are synthesized in living organisms by mechanisms different from the usual ribosome-dependent processes of protein synthesis. The dipeptides carnosine (β-alanyl-histidine) and anserine (β-alanyl-1-*N*-methylhistidine) are synthesized enzymatically from β-alanine and histidine, and from carnosine and *S*-adenosylmethionine (SAM), respectively.

Glutathione (γ-glutamylcysteinylglycine) plays a variety of roles in living organisms. This tripeptide is synthesized by a two-step enzymatic pathway:

1. The formation of a peptide linkage between the γ-carboxyl group of glutamate and the amino group of cysteine, to produce γ-glutamylcysteine;
2. The condensation of this dipeptide with glycine, to form glutathione.

Thus the order of the amino acids in glutathione is specifically determined by the enzymes catalyzing the formation of each peptide bond.

At least 90 different peptide antibiotics are produced by strains of *Bacillus subtilis* and *B. brevis*. Gramicidin S, for example, is a cyclic decapeptide composed of two identical pentapeptides (D-phenylalanine-L-proline-L-valine-L-ornithine-L-leucine). Gramicidin S is synthesized by a multienzyme complex, gramicidin synthetase, composed of two enzymes, one of which, serving as a template, specifies the amino acid sequence in the antibiotic.

SAM, the metabolically activated form of methionine, functions as an important source of methyl and propylamino groups for a wide variety of compounds, including alkaloids, choline, creatine, epinephrine (adrenaline), *N*-methylated amino acids, nucleotides, and polyamines, as well as for phospholipids, proteins, polysaccharides, and nucleic acids. It is synthesized from methionine and ATP, in a reaction catalyzed by SAM synthase.

A wide variety of amines occurring in bacteria, plants and animals are derived directly or indirectly from amino acids by decarboxylation; these include ethylamine (from alanine), agmatine (from arginine), γ-aminobutyric acid (from glutamate), methylamine (from glycine), histamine (from histidine), cadaverine (from lysine), putrescine (from ornithine), phenylethylamine (from phenylalanine), ethanolamine (from serine), tryptamine, and 5-hydroxytryptamine (from tryptophan), and tyramine and dopamine (from tyrosine). These amines and their derivatives often play a variety of physiologically important roles. For example, γ-aminobutyric acid, phenylethylamine, tryptamine, 5-hydroxytryptamine, or serotonin, tyramine, dopamine, norepinephrine (noradrenaline), and epinephrine are all neurologically active compounds, whereas histamine, a powerful vasodilator, is involved in allergic reactions. (*See* **Amines**.)

Tyrosine plays several important roles in animal metabolism as a precursor to melanins, thyroid hormones (thyroxine and triiodothyronine), and catecholamines (dopamine, norepinephrine, and epinephrine). In the synthesis of melanins, tyrosinase catalyzes first the hydroxylation of tyrosine to 3, 4-dihydroxyphenylalanine (dopa), followed by the oxidation of dopa to phenylalanine-3,4-quinone (dopaquinone). Dopaquinone undergoes a sequence of reactions, including polymerization, to yield both red and black melanins. Another enzyme forming dopa is tyrosine hydroxylase, which catalyzes the first reaction in the sequential enzymatic pathway leading to the biosynthesis of catecholamines. Dopa is then decarboxylated to yield 3,4-dihydroxyphenylethylamine (dopamine). Dopamine is hydroxylated to norepinephrine, which in turn is methylated by SAM to give epinephrine. (*See* **Hormones**: Adrenal Hormones; Thyroid Hormones.)

Putrescine, or 1,4-diaminebutane, is synthesized by decarboxylation of ornithine, in a reaction catalyzed by ornithine decarboxylase, a highly regulated enzyme. Another route to putrescine formation involves the conversion of arginine to agmatine by arginine decarboxylase, followed by cleavage of agmatine to putrescine and urea by agmatine ureohydrolase. Putrescine is an intermediate in the biosynthesis of two important polyamines, spermidine and spermine. Spermidine is synthesized enzymatically by the SAM-mediated transfer of a propylamino group to putrescine. The enzymatic transfer of an additional propylamino group from SAM to spermidine produces spermine. These polycations play multiple roles in stabilizing negatively charged intracellular components such as nucleic acids and membranes.

Creatine phosphate, which serves as a source of high-energy phosphate in mammalian muscle and brain, is synthesized in three steps from arginine, glycine, and methionine.

There are four classes of tetrapyrrole compounds – hemes, chlorophylls, phicobilins and cobalamins – all of which are synthesized from a common precursor, δ-aminolevulinic acid (ALA). In bacteria and animals, ALA is synthesized by the condensation of glycine and succinyl-CoA, with loss of carbon dioxide, in a reaction catalyzed by ALA synthase. In plants, however, ALA is formed from glutamate by a three-step pathway.

An enormous amount of carbon in the biosphere passes through the pathway leading to lignin biosynthesis, the major constituent of woody tissue. In the first reaction, phenylalanine ammonia lyase catalyses the cleavage of phenylalanine to *trans*-cinnamic acid and NH_4^+. Cinnamic acid is a precursor for the synthesis of a huge number of plant substances, including lignin, tannins, flavonoids, pigments, many of the flavor components of spices, and various alkaloids, such as morphine and colchicine. (*See* **Lignin**.)

In addition, the synthesis of a variety of other important molecules utilizes various amino acids as precursors. Thus β-alanine is a component of CoA, asparagine is a major form of transport of organic nitrogenous compounds in plants, and aspartate is involved in purine and pyrimidine biosynthesis. Glutamate is a precursor of folic acid; glutamine contributes to the synthesis of a variety of substances, including purines, pyrimidines, ATP, cytidine triphosphate, NAD, amino sugars, and glycoproteins; cysteine is a precursor of taurine, isethionic acid, CoA, vasopressin, various types of pigments, including phaeomalins and trichochromes, and other sulfur-containing compounds. Glycine also plays multiple roles, including contributions to the one-carbon pool and as a precursor of purines, glyoxylate, and various conjugates such as hippurate and glycocholate. Histidine is involved in ergothionine and homocarnosine biosynthesis, and methionine, via SAM, is the precursor of the plant hormone ethylene, which influences plant growth and development and induces the ripening of fruits. Serine is involved in the biosynthesis of phospholipids, and tryptophan is the precursor of several important physiological substances, including NAD, NADP, and the plant hormone indole 3-acetic acid. (*See* **Niacin**: Physiology; **Ripening of Fruit**.)

See also: **Amines**; **Energy**: Measurement of Food Energy; **Enzymes**: Functions and Characteristics; **Glucose**: Function and Metabolism; **Hormones**: Adrenal Hormones; Thyroid Hormones; **Lignin**; **Niacin**: Physiology; **Protein**: Chemistry; **Ripening of Fruit**

Further Reading

Devlin TM (1997) *Textbook of Biochemistry with Clinical Correlations*. New York: Wiley-Liss.

Dey PM and Harborne JB (1997) *Plant Biochemistry*. San Diego: Academic Press.

Fowden L, Lea PJ and Bell EA (1979) The non-protein amino acids of plants. *Advances in Enzymology* 50: 117–175.

Garrett RH and Grisham CM (1995) *Biochemistry*. Fort Worth: Saunders College Publishing.

Kulka HE and Schwarz RS (1988) A flexible genetic code, or why does selenocysteine have no unique codon? *Trends in Biochemical Sciences* 13: 419–421.

Miflin BJ and Lea PJ (1977) Amino acid metabolism. *Annual Review of Plant Physiology* 28: 299–329.

Moran LA, Scrimgeour KG, Horton HR, Ochs RS and Rawn JD (1994) *Biochemistry*. Upper Saddle River, NJ: Prentice Hall.

Murray RK, Granner DK, Mayes PA and Rodwell VW (1988) *Harper's Biochemistry*. East Norwalk, Connecticut: Appleton and Lange.
Orten JM and Neuhaus OW (1982) *Human Biochemistry*. St Louis, Missouri: CV Mosby.
Rawn JD (1989) *Biochemistry*. Burlington, North Carolina: Neil Patterson.
Singh BK (1999) *Plant Amino Acids. Biochemistry and Biotechnology*. New York: Marcel Dekker.
Smith EL, Hill RL, Lehman IR *et al*. (1983) *Principles of Biochemistry: General Aspects*. Singapore: McGraw-Hill.
Stenesh J (1998) *Biochemistry*. New York: Plenum Press.
Stryer L (1995) *Biochemistry*. New York: WH Freeman.
Umbarger HE (1978) Amino acid biosynthesis and its regulation. *Annual Review of Biochemistry* 47: 533–606.
Voet D and Voet JG (1995) *Biochemistry*. New York: John Wiley.
Zapsalis C and Beck RA (1985) *Food Chemistry and Nutritional Biochemistry*. New York: John Wiley.
Zubay G (1993) *Biochemistry*. Dubuque, Iowa: Wm C Brown.

ANALYSIS OF FOOD

K Gupta , V Jain, S Jain, K Dhawan and G Talwar, CCS Haryana Agricultural University, Haryana, India

Background

The science of food analysis has developed rapidly in recent years. Analysis of food not only provides information about composition, appearance, texture, flavor, shelf-life, safety, processibility, and microstructure, but also guarantees product quality. Knowledge of the chemical and biochemical composition of foods is important to the health, well-being, and safety of the consumers. Also, it is important for compliance with legal standards, quality assurance, nutritional value determination, and adulteration detection.

Analysis of food involves sample preparation, analysis, and detection of major food components (amino acids, peptides, proteins, enzymes, lipids, phospholipids, carbohydrates, alcohols, fat-soluble vitamins, water-soluble vitamins, organic acids, organic bases, phenolic compounds, bittering substances, pigments, aroma compounds, and dietary fiber) and miscellaneous components (preservatives, antioxidants, colorants, and nonenzymic browning substances). A large number of different analytical techniques have been developed to analyze the properties of food because of the complexity and diversity of food components. These techniques can be divided in to three general segments focusing on physical techniques including spectroscopy, chromatography and electrophoresis, biochemical analysis, and sensory analysis.

Sources of Information

The food analyst has to identify and choose the best analytical method available for analyzing a food sample depending upon its complexity. For this, the food analyst must be familiar with all the recent advances in food analysis methods and must know where to look for the needed information/literature. The main sources of information and general approach to a literature search are detailed below.

Books

The best way to obtain general information about food analysis is to consult textbooks. A number of textbooks are available on food analysis, and these textbooks generally provide information on the background and basic principles of food-analysis methods, sophisticated instruments, and classical procedures that have been in use for many years. The textbooks usually concentrate on the fundamental principles rather than detailed methodology.

For a wide range of information, several encyclopedias exclusively for food chemistry or food technology are good sources of literature, such as *The Encyclopedia of Science and Technology*. Monographs are basically comprehensive surveys of current knowledge on a specific subject. Recently, a number of books (monographs) have become available for specialized areas like the use of spectroscopy and/or HPLC for food analysis. Handbooks are another source of comprehensive information. For example, the *Handbook of Food Analysis*, in its two volumes, deals with the methods for physical characterization as well as for biochemicals, nutrients,

and sensory analysis for a diverse number of food components.

Reviews

Once general information has been obtained, the next step is to consult a critical review of the current knowledge on the particular field. A review may be written about the determination of a specific component, about the specific technique that can be used for different food sample analyses or about the current aspects of a new technique/instrument. Critical reviews pertinent to food analysis are usually published in scientific journals like *Advances in Food Research*, *Annual Review of Nutrition*, *Methods of Biochemical Analysis*, *Recent Advances in Food Science*, etc.

Periodicals

Scientific journals are one of the best sources of information as they deal with the articles/information about current and new analytical methods. Periodicals provide information about the research work or analytical methods developed by other scientists. They cover all scientific aspects of the chemical composition of foods, e.g., *Journal of Food Composition and Analysis, Analytical Chemistry, Food Testing and Analysis*, etc. Related journal articles can now be searched for on the internet on specific sites, such as *www.academicpress.com/jfca* for *Journal of Food Composition and Analysis*.

Official Methods

Several standard analytical methods have been approved by the respective governments/scientific agencies (e.g. Association of Official Analytical Chemists (AOAC), American Association of Cereal Chemists, American Oil Chemists' Society (AOCS), etc.) for use in the food industry. These standard methods are generally compiled in volumes, and they are being improved and updated from time to time.

Symposia

Programs of conferences and symposia are published in several journals and are included on websites. In symposia, prominent scientists present invited review papers in the area of their competence. Abstracts of papers presented at scientific meetings are generally published in the form of a book by the sponsoring scientific society. The Asia Pacific Food Analysis Network and American Chemical Society hold regular meeting/symposia on topics related to food chemistry/analysis, and programs and abstract books are available from the organizing society office bearers.

Trade Publications

The purpose of the companies that trade in chemicals and instruments is to sell products, but advertisements are also sources of valuable information. These publications are the best source on the properties and application of specialized equipment and chemicals. Several manufacturers periodically issue bibliographies and abstracts of technical-scientific articles in a specific area, and some publish periodicals that reproduce specialized articles from regular scientific journals. Even detailed handbooks giving specification, properties, and details of analytical procedures in selected areas are provided by these companies. The advertisements or lists of major commercial companies for chemicals and instrumentation appear in several scientific and trade journals (*Science, Nature, Prepared Foods*, etc.). The instruction manuals supplied with instruments are very useful for installing, using, and servicing equipments.

Modern Information Retrieval Systems

Nowadays, computerized systems have a great capacity for storing and retrieving scientific information. A number of websites are available, on which one can search information on a particular subject. These sites can be accessed from any computer terminal linked to the internet. The proper use of these systems reduces the time, cost, and effort involved in searching the literature. Some of these sites/search engines include http://agrifor.ac.uk, www.altavista.com, www.biacore.com, www.bmn.com, www.cabi.org, www.google.com, www.ncbi.nlm.nih.gov, www.national academics.org, http://highwire.stanford.edu, and www.yahoo.com.

Several more such sites/databases are available for information on food analysis, such as Agricola (USDA, National Agricultural Library, Beltsville, MD), Dissertation abstracts online (University Microfilms International, Ann Arbor, MI), Science Citation Index (Institute for Scientific information, Philadelphia, PA), and Food Science and Technology Abstracts (International Food Information Service, Reading, UK).

Preparation of Samples

Food is a complex inhomogeneous mixture of chemical substances and microorganisms (active or not). Therefore, improper sampling may result in variable analysis results. Ideally, the sample analyzed in a laboratory should have exactly the same properties as the total population it is supposed to represent. There are three steps involved in sample preparation for chemical analysis of foods.

Sampling

An analyst has to decide on the number of samples of a given population to test, the location from which the sample should be selected, and the method used to collect them. The choice of a particular sampling plan depends on the purpose of the analysis, the physical property to be measured, the nature of the total population and of the individual samples, and the type of analytical techniques used to characterize the samples.

Homogenization

The complex structure and composition of food substances necessitate homogenization prior to most chromatographic analysis. Each sample should be homogenized using methods proven to be effective with a particular matrix from experience, from the literature, or by experiment to a degree that meets all the requirements of the assay procedure. This must be confirmed by replicate assays, blanks, and recoveries.

Sample Preparation

Physical and chemical manipulation of the test portion prior to analytical measurements is carried out. This includes storage, weighing, dilution, clean-up, extraction, digestion, purification, separation, derivatization, etc. It is also important to report the results of any analysis in a clear fashion, stating the procedures used, the number of replications performed, and the estimated reliability of the measured value. The analytical method selected must be specific, accurate, precise, and sensitive for the substance to be analyzed.

Selection of a Method for Food Analysis

Once proper sampling and sample preparation have been completed, the food product is ready for analysis. There are numerous analytical procedures available to analyze a particular food product. Selection of the most suitable procedure is often the key to the success of an analysis. Some of the important factors that should be considered are as follows.

Specificity Specificity relates to the absence of interfering substances that give a measurement of the same kind as the substance being determined or the ability to detect and quantify specific components within the food material, even in the presence of similar food components.

Accuracy This is a measure of how close one can approach the true value of parameter being determined.

Precision This is the ability to reproduce a result between the determinations of a substance by the same scientist (or group of scientists) using the same equipment and experimental approach.

Reproducibility This is the ability to reproduce a result by scientists using the same experimental approach in different laboratories using different equipment.

Economy This is the total cost of analysis including reagents, time, and instrumentation.

Simplicity of operation This is the ease with which the analysis may be carried out by other workers.

Sensitivity This is the smallest detectable amount of the substance/sample that can be quantified/detected by the given technique.

Safety This concerns the potential hazards associated with reagents and procedures used in analysis.

Speed This relates to the time required to complete the analysis.

Destructive/nondestructive This indicates whether the sample is destroyed during analysis or remains intact.

Techniques and Methods

The analytical techniques involved in food analysis can be categorized as physical, biochemical, and sensory evaluation techniques.

Physical Techniques

A variety of physical techniques have been developed to analyze the properties of food materials. These have been categorized according to the principles on which they are based.

Spectroscopy Spectroscopic methods involve the quantitative evaluation of the phenomenon resulting from the interaction between a certain form of radiation (electromagnetic, acoustic, electron or neutron beam) and matter (atoms, molecules, ions) to provide information about the properties of a food. Radiation energy interacts with the matter in a variety of ways, including absorption, emission, reflection, refraction, dispersion, and polarization. Spectroscopic methods based on each of these interactions are used in food analysis, and those methods based on absorption and emission are the most common (Table 1).

Table 1 Spectroscopic methods for food analysis

Type of spectroscopy	*Information obtained*	*Material required*	
		Nature	*Quantity*
Atomic absorption and emission spectroscopy	Analysis of elements	Solution	Small
Atomic fluorimetry	Analysis of elements	Solution	Small
UV-visible absorption spectroscopy	Chemical and quantitative analysis	Liquid extracts	Small
Electron paramagnetic	Physical state	Native material	Moderate
X-ray spectroscopy	Structure and composition	Native material	Very small
Nuclear magnetic resonance	Physical state and chemical analysis	Native material or solid or liquid extract	Large
Infra-red spectroscopy	Physical state and chemical structure	Native material depending on the type of IR	Small to large
Mass spectroscopy	Analysis of elements and structure determination	Usually solid extracts	Very small

The quantity very small, sub-microgram level; small, sub-milligram level; moderate, tens of milligrams; large, hundreds of milligrams.

Atomic spectroscopy is based on the transition of outer shell electrons between the different energy levels upon interaction of atoms with electromagnetic radiation in the ultraviolet, visible, and infrared regions. When the electron returns to its normal state, it releases its acquired energy in the form of radiation that is specific to that element, and emitted radiation can then be measured using suitable detectors. Each element has a specific set of energy levels, and so the wavelength of the absorbed or emitted radiation can be used to identify the element, and the extent of absorption or emission can be used to quantify this element. Atomic spectroscopy is used to determine the concentration of mineral elements in food. In atomic absorption spectroscopy, the interaction between the source of energy in the correct frequency region and a sample is measured. An atomic absorption instrument contains a source of radiation with a wavelength that matches those required to excite atoms in the flame, thereby providing a more efficient means of atomic excitation. In atomic emission spectroscopy, a sample is heated to atomize it and to excite electrons to a higher energy state. The emission of radiation by the excited atoms is measured using suitable detectors. The disadvantage of atomic absorption spectroscopy is that it can measure only one element at a time. In comparison, atomic emission spectroscopy (also called flame photometry) can determine several elements simultaneously.

Atomic fluorimetry is based on the absorption of a photon by an atom that causes the passage of a valence electron of one orbital to another orbital with a higher energy. The atom, being unstable, returns to its ground state, emits a photon with a different frequency (usually lower than the absorbed photon) causing fluorescence. This can be used to detect a few mineral elements (Ca, Cd, Fe, Pb, Ni, Cr, Cu, etc.)

UV-visible absorption spectroscopy is one of the most common spectroscopic techniques used to resolve problems of compound identification, quantification, and qualitative analysis of food. Like atomic spectroscopy, it is also based on the transition of outer-shell electrons. A beam of monochromatic radiation is transversed through the sample, and reduction resulting from absorption is measured. A spectrum is obtained by carrying out measurements over a wide range of wavelengths that contain peaks corresponding to the absorbing groups. A compound can be characterized by its spectrum, provided that the absorption coefficients of bands are known, or by comparing it with reference spectra. The quantitative aspects of this technique are based on the Lambert–Beer law, which states that the absorbance of a solution (the log of the ratio of light entering a sample to light leaving the sample) is proportional to its concentration at a particular wavelength under a set of standard conditions. The quantity of a substance can also be determined by measuring the height of one of these absorption peaks (if there is no interference from other molecules). Derivative spectroscopy can be used to eliminate any interference caused by the reaction environment. Photodiode array spectroscopy represents a great advance on absorption spectroscopy allowing simultaneous detection at a wide range of wavelengths in a liquid flow. It provides immediate information of spectra for each eluted compound.

Electron paramagnetic resonance spectroscopy arises from the fact that electrons have a quantum mechanical property, called spin, that results in the generation of a magnetic moment. Only atoms or molecules with unpaired electrons possess a magnetic moment. The transition at this moment when a sample is placed in magnetic field can be detected

and characterized. This technique is employed to detect lipid peroxidation, molecular damage to proteins, and nucleic acids.

Nuclear magnetic resonance (NMR) techniques are concerned with the transition of nuclei between different energy levels when these are placed in a magnetic field. The frequency at which the moment of spin 'wobbles' allows the elementary segments of a molecule to be identified. The nuclei having a net spin angular momentum can be detected by NMR, the most common being ^{1}H and ^{13}C. NMR can be used to determine the moisture, concentration, molecular structures, phase transitions, diffusion, and image, and can be used to check the authenticity of food and drink ingredients whose commercial value is related to their natural origin.

Infrared (IR) spectroscopy techniques are based on the absorption/emission of electromagnetic radiation by a sample resulting from vibration or rotation of molecules. The interaction of IR radiation with matter leads to an IR spectrum containing detailed information on different types of chemical groups of the substance in the form of peaks. The location and magnitude of a peak provide information about the type and concentration of the groups present. IR spectroscopic techniques are available to determine the proteins, lipids, carbohydrates, moisture contents, flavor, and toxic constituents in food.

Mass spectroscopy techniques rely on the production of ions by bombarding organic compounds (in their vapor state) with high-energy electrons. The ions formed are accelerated and separated as a function of their mass and charge under the effect of variable electric and/or magnetic fields, and their relative abundance can be detected. These techniques can be used to identify and determine components in mixtures of organic materials, e.g., flavor components, lipids, amino acids, carbohydrates, vitamins, minerals, toxicants, antinutritional factors, and additives in food, and to identify food microorganisms.

X-ray spectroscopy/spectrometry techniques are based on the emission of X-rays when accelerated electrons/X-rays collide with matter and eject an electron from atom's internal layer. The vacancy is filled by another electron from a higher orbital, and this equilibration leads to the emission of X-rays. The excitation can be produced by accelerated electrons/X-rays, leading to emission spectroscopy or X-ray fluorescence spectroscopy. These techniques provide information about the structure, composition, and purity of the food material.

Optical techniques Analytical techniques utilize the interaction of monochromatic beam of light with matter to provide information about the composition and quality of a food. These techniques include photometry and turbimetry, which concern the dispersion of incident light by suspended particles in liquid and even in gases to measure colloidal particles in water, and clarity of juices, syrups, and wines. Polarimetry relies on the rotation of the plane of polarized light by an optically active solution and is used to determine lactose in condensed milk and sugars in juices. Refractrometry is based on the change in refractive index caused by the displacement of a molecule/substance in solution or by refraction caused by that substance in solution. These techniques can be used in the analysis of lipids, carbohydrates, and protein concentration in liquid and dry matter. In combination with density determination, these techniques can be used to determine unknown substances or structures of known substances. For very precise measurements, differential refractrometry and interferometry can be used.

Chromatography Chromatography designates a variety of techniques that can separate mixture of molecules into individual components on the basis of their differences in size, shape, mass, charge, solubility, and adsorption properties. It is used for purification and constituent preparation in the food industry. A solution of a mixture of molecules (usually in a liquid or gas carrier) is passed through a porous medium. The carrier is called the mobile phase, because it moves through the porous medium, whereas the porous medium is called the stationary phase. In this process, different solute components are separated as a result of the different affinities of the components for a stationary (solid or liquid) phase; the stronger the interaction, the slower they move. The affinity or strength of an interaction depends on the physico-chemical properties of the substance and the stationary phase. Molecules can be separated according to their size, shape, mass, polarity, electric charge, solubility, and molecular interactions. The chromatographic techniques can be divided into different categories depending on their physico-chemical basis of separation (Table 2).

Chromatographic separation can be carried out in one of the three modes, depending on the nature of the support that holds the stationary phase. In paper chromatography, the stationary phase (liquid) is supported by cellulose fibers of a paper sheet. The sample is spotted on to the bottom of a sheet of paper, and the paper is kept in an atmosphere saturated with water vapors to form a thin film of a stationary phase. An appropriate solvent system acting as a mobile phase is then allowed to flow over the sample spot. The various components of the sample are then partitioned between the stationary and mobile phases.

Table 2 Classification of separation techniques of chromatography on physico-chemical basis

Molecular characteristics	*Physical principles*	*Example*
Polarity	Differences in partitioning coefficient between polar and nonpolar solvent	Partition, i.e., liquid–liquid; liquid–solid gas–liquid chromatography and HPLC
Ionic	Differences in electrostatic interaction between charged groups on molecules in the mobile phase and stationary phase	Ion-exchange chromatography
Size	Differences in the size and shape of molecules	Gel-permeation chromatography
Affinity/shape	Specific interactions between molecules in the mobile phase and stationary phase	Affinity chromatography

Components with a higher affinity for the stationary phase move less rapidly than those with a higher affinity for the mobile phase. A more sophisticated separation can be achieved using two-dimensional paper chromatography.

Thin-layer chromatography (TLC) has largely replaced paper chromatography, because it is more convenient, quicker, and more sensitive, and it has a better resolution and reproducibility. This technique is similar to that of paper chromatography, except that paper is replaced by a plastic or glass sheet coated with a thin film of porous material that acts as the stationary phase. In column chromatography, separation is achieved by passing the sample through a vertically fixed tubular glass or polypropylene column using either gravity or a slight vacuum to move the mobile phase through. Steel columns and high pressures are now used to speed up the process. In gas chromatography, the sample is volatilized and then carried along by a gaseous mobile phase through a column coated with the stationary phase. The efficiency of separation can be controlled by selecting the appropriate combination of stationary and mobile phase. In high-performance liquid chromatography (HPLC), the sample dissolved in the liquid mobile phase is passed through a column containing the stationary phase (which may be either solid or liquid). This gives a faster and superior resolution, as high pressure is applied to move the mobile phase.

Electrochemical techniques Analytical techniques involve measuring the electrical properties of food, e.g., potential, current intensity or resistance, to provide information on composition. Potentiometric methods rely on measurements of the potential difference between a specific (an indicator) electrode and a reference electrode, and this is directly proportional to the concentration of the species being measured.

Polarographic methods are based on the potential difference between a mercury electrode and a reference electrode immersed in a solution containing electroactive species. These are used to determine heavy metals polluting water or foods, pesticides, and preservatives. Coulometric methods involve measuring the quantity of electric current required to electrolyze the substance being analyzed; the greater the electric charge, the higher the concentration of the substance. The electrical conductivity is based on the displacement of different components of a substance in the solution, which is equivalent to the current passing between a pair of electrodes. The amount of current is measured and then extrapolated to the composition of a specific component within the food.

Electrophoresis relies on differences in the movement of charged macromolecules in an electrical field applied across a substance. It is used to separate molecules on the basis of a net charge, size, and/or shape. The sample to be analyzed is added to a buffered porous matrix (usually a strip of paper or gel), and a current is applied. Charged molecules move through the matrix in a direction that depends on the sign of their electric charge. The rate of movement depends on the magnitude of the charge and resistance to movement through the matrix. The resistance is determined by the relationship between the effective size of the substance and the size of the pores in the matrix. The larger the size of the pores or the smaller the size of molecules to be separated, the lower the resistance, and therefore, the molecules will move faster through the matrix. Isotachoelectrophoresis (separation of sample components based on differences in net mobility) has a high resolving power, because of its concentrating effect. Components with different isoelectric points are separated better by isoelectric focusing (IEF). Two-dimensional electrophoresis can be carried out to improve the resolution. Substances are separated in one direction, according to their charge by IEF, and in a second dimension (perpendicular to the first), on the basis of their size.

Rheological techniques These techniques are based on the relationship between the stress and strain in a material, with a behavior between that of solids and

liquids. Measurements of these properties of foods can be based on either an analytical or integral approach. In the analytical approach, the properties of a material are related to a simple system such as Newtonian fluids or Hookian solids. In the integral approach, a simple empirical relation between stress, strain, and time is established. The relationship measures the way in which rheological properties vary under a specific system of applied forces. The instruments used provide a measure of baking quality of flours, thickening properties of biopolymers, texture, hardness, flakiness, consistency of food, tenderness of meat, and texture of fruits and vegetables. Rheological techniques are now essential tools, especially as a quality-control tool in food science and technology.

Biochemical techniques There are a number of biochemical techniques available that can be used to determine the identity and/or concentration of specific components of food. Enzymatic analysis means analysis with the aid of enzymes utilizing specific enzyme reactions. The major advantage of enzymatic techniques lie in their ability to react specifically with individual components of a mixture. This avoids lengthy separation of the components and reduces the time required for analysis. These techniques require small amounts of sample and do not require components to be extracted from their native medium.

Immunochemical techniques These rely on specific interactions between an antigen and an antibody. Since macromolecules found in microorganisms, food, and agricultural products are good antigens, antibodies against these molecules can be obtained. The potential applications for immunological techniques are numerous, as much for raw material and final product quality control as for tracking production and conservation effectiveness.

Analytical microbiology involves the use of microorganisms as reagents for the quantitative determination of certain chemical compounds present in food. Since some of the nutritional requirements of microorganisms and experimental animals are similar, it is possible to use analytical microbiological techniques to determine some substances that are essential constituents of living cells. These are used for vitamins, nucleic acids, amino acids, heavy metals, growth factors, and nutritional values of proteins and antibodies. The basic principle is that in the presence of limiting amounts of certain compounds, the amount of microbial growth is a function of the concentration of those compounds. The microorganisms used for assay are primarily bacteria, but yeast, fungi, algae, and protozoa have also been used. These can be assayed after diffusion of the substance through solid inoculated culture media by turbidimetric, dilution, gravimetric, and metabolic response methods.

Sensory evaluation The sensory attributes of the quality of a food are measured to determine consumer preference in order to help in the manufacture of an acceptable product at maximum production economy. Sensory attributes include appearance (color, size, shape, and consistency of liquid and semisolid products), kinesthetic (texture, consistency, and viscosity) and flavor (taste and odor). The evaluation is also done in order to determine the conformity of a food with established government or trade standards and food grades. The technique of sensory evaluation involves using individuals as the data generator, and so, it is influenced by factors such as social, cultural, religious, psychological, and climatic factors, physical status of the individual, availability, and nutritional education. Large, highly trained groups of experts are employed for evaluation to minimize the effects of such factors. Laboratory tests may be conducted to study the human perception of food attributes, to correlate sensory attributes with chemical and physical measurements using principal component analysis, to evaluate raw material selection; to study the processing effects and means of maintaining uniform quality, to establish shelf-life stability, and to reduce costs.

Food Components

Structural changes occur during the processing and storage of foods. Therefore, knowledge of the composition of a food is necessary in order to determine the extent of any change in the concentration of the nutrients. The use of agrochemicals and additives also may affect the quality of the food product. In order to obtain quality assurance, in accordance with legal requirements, nutrition, and food safety, the concentration of major and minor food components must be known. The methods used for the analysis of food components are discussed below, and these methods form part of the techniques already described.

Major Components

Moisture Water is the major component of most foods and is an index of the economic value, stability, and quality of food products. It is important to note that the removal of water either by conventional dehydration or by separation locally in the form of fine ice crystals (freezing) greatly alters the native properties of foods. Therefore, it is very important to use

simple, rapid, and accurate methods for moisture determination in food products. The food sample should be homogenized using a number of electrical/mechanical devices like blenders, mincers, graters, homogenizers, and grinders. Analytical methods of moisture determination can be classified as direct and indirect procedures. Direct procedures include the determination of moisture by drying (oven, vacuum, freezing, chemical desiccation), distillation, chemicals (Karl Fischer titration), and extraction (gas chromatography and refractrometry). The drying method relies on the evaporation of water from the sample and is one of the most commonly used procedures for determining moisture content (conventional ovens, forced air ovens, vacuum oven, microwave oven, and IR oven). The weight of the sample is taken before and after it is dried, and the moisture content is then calculated. In the distillation method, the food sample is mixed with an organic solvent, and water is evaporated and collected in a graduated collection arm, in which its volume can be determined. The most commonly used chemical methods for moisture determination are Karl Fischer titration and the gas-production method. These methods rely on specific chemical reactions between water and other substances that lead to some quantifiable parameters. Indirect methods include rapid and nondestructive methods, and these involve calibration against standard moisture values that have been precisely determined by using one or more of the direct methods. These methods are based on differences in the physical properties of water compared with other components and are widely used to determine moisture content including IR gas chromatography, NMR, and several electrical methods. These instruments are simple to use, provide rapid and reliable measurements, and are, therefore, particularly suitable for routine quality-control applications.

Mineral components Ash is a measure of total mineral components in a food and is regarded as a general measure of quality in certain foods such as tea, flour, and edible gelatin, etc. There are two major procedures of ashing that are used to determine the mineral content of the sample. These are dry ashing and wet ashing. The nature of the ashing procedure is governed by the purpose for which ash is prepared, the particular constituents to be determined, and the method of analysis to be used. In dry ashing, the sample is ignited at 550–600 °C to oxidize all organic material without a flame, whereas in wet ashing, the sample is digested in acid to determine mineral elements. The individual elements within ash can be determined by titrimetric methods, colorimetric methods, or atomic spectroscopy techniques. Atomic spectroscopy techniques can provide a complete profile of different types of elements in a food material, whereas the titrimetric and colorimetric methods are usually designed to determine a particular element. In addition, atomic spectroscopy techniques have a much higher sensitivity and specificity.

Carbohydrates Carbohydrates are present in all grains, vegetables, fruits, and other plant parts and vary in form from simple monosaccharides to oligosaccharides and more complex polysaccharides. These are a major source of energy for the human body, and their varied functional properties are utilized by the food industry to enhance the palatability, acceptability, and shelf-life of foodstuffs. There is thus a continuous need to monitor levels of carbohydrates in foods. The complete analysis of foodstuffs may require the determination of several groups of compounds, e.g., simple sugars, reducing sugars, polysaccharides, and fibers, all of which may play an important role in the quality of the product. For this purpose, a number of physical and chemical methods have become well established and commonly used in many laboratories. Physical methods include refractrometry, polarimetry, and hydrometry, whereas chemical methods include titrimetric and colorimetric, which determine the quantity of a particular sugar present in food. Chromatographic techniques such as paper chromatography, TLC, gas liquid chromatography (GLC), and HPLC provide more rapid analyses with a greater precision and specificity. Nowadays biochemical methods of analysis are often used because these are specific, sensitive, rapid, and reproducible. Flow injection analysis (FIA) gives an excellent performance and has been the basis of a large number of methods for the separation, processing, and analysis of carbohydrates. The future of carbohydrate analysis undoubtedly belongs to three analytical techniques – HPLC, enzymatic assays, and FIA – all having a close relationship to one another.

Proteins The determination of proteins in foods and food products has an important nutritional, functional, and technological significance. The methods for protein estimation depend on the measurement of a specific element or chemical group in the proteins. The elements or groups most commonly used are nitrogen, aromatic amino acids, or peptide linkage. Based on nitrogen content (approx. 16% of protein weight), either protein can be estimated by Kjeldahl's method (AOAC) or ammonia can be directly determined in the digest by color-inducing compounds such as ninhydrin, indophenol, or Nessler's reagent.

A variety of spectrophotometric methods are also available to determine protein content. These include the biuret method (based on the binding of copper (II) to a peptide bond in protein molecules at alkaline pH values), the Lowry method (based on the reduction of Folin–Ciocalteau reagent by oxidation of tyrosine and tryptophan on the polypeptide side chain), and the dye binding method (based on the measurement of excess dye remaining in solution after removal of the precipitated protein–dye complex). Most proteins exhibit an absorption maximum at 280 nm, because of the presence of a side chain of aromatic amino acids, and so UV-spectrophotometry is also used for protein identification and quantification. IR spectrophotometry has become a popular method for the determination of protein in cereal products, because of its speed, simplicity of operation, safety, and low running costs. Specific protein components in food and food products can be quantified by chromatography, electrophoresis, immunology, or a combination of these. Among the methods of chromatography, HPLC is becoming increasing important in the analysis of specific protein components in foods and food products. The separation of proteins in an electric field on the basis of their charge density is a powerful and frequently used technique for the analysis of specific protein components in food. Methods such as immunoelectrodiffusion and enzyme-linked immunosorbent assay are immunological methods (based on the interaction between an antigen and its corresponding antibody) that are highly specific and sensitive means to identify and quantify minor protein components in complex matrices. To improve and optimize the accuracy of the protein quantification in food and food products, further research for new and alternative absolute protein quantification methods is necessary. Amino acids are the building blocks of proteins, and these can be determined after hydrolysis of proteins by colorimetric, enzymatic, and chromatographic (ion-exchange and gas-chromatography (GC)) methods.

Lipids Most of the methods used in food lipid analysis cannot be directly applied to foods themselves because of their complexity. Lipids should first be extracted either for total lipid estimation or for further analysis. Neutral lipids (triacylglycerols) can be easily extracted by nonpolar solvents such as petroleum ether, hexane, or supercritical carbon dioxide. If a sample contains phospho- or glycolipids, polar solvents such as methanol must be used for quantitative determination. Standard methods are used to determine the oil content (AOAC or AOCS) by refluxing the sample in petroleum ether using Soxhlet apparatus. There are a number of procedures for oil estimation in which a sample is not destroyed, such as low-resolution NMR or high-resolution NMR. Methods have been developed to separate different classes of lipids such as neutral lipids, glycolipids and phospholipids. The techniques commonly used are TLC, column chromatography, and HPLC, especially with mass detection.

General information on the constituent fatty acids of fats can be obtained from different chemical analyses giving specific values such as the saponification value, iodine value, hydroxyl value, etc. The presence of free fatty acids in oil, which is an index of rancidity, can be determined by titrating against a standard alkali. Specific fatty acids in fat can be determined by gas–liquid chromatography (GLC), which is still the most informative technique in fatty acid analysis. Mass spectrometry coupled to GC is the most powerful tool for identifying fatty acids separated by GLC. IR spectroscopy and Raman spectroscopy are most often used to detect *trans*-fatty acids and isomers of *cis*-unsaturated fatty acids. UV-spectroscopy at 200–400 nm is generally used to detect the presence of conjugated double bonds. The processing and quality of food oils are often determined by physical characteristics such as their melting point, smoke point, flash point, fire point, solid-fat content, and degree of oxidation. These characteristics can be determined by different analytical techniques.

Other Food Components

A large number of different analytical techniques have been developed to analyze food components such as bitter compounds like terpenoids, flavonoids, phenols; pigments like carotenoids, chlorophyl, anthocyanins; vitamins (water- and fat-soluble); aroma compounds; dietary fibers; residues; alcohols; organic acids; antioxidants; preservatives, etc. The standard methods of the AOAC are generally adopted for the analysis of these components. Apart from this, often techniques that are commonly used for the detection and quantification include colorimetric, enzymatic, chromatography (paper chromatography and TLC), UV-visible absorption spectroscopy, gas chromatography, HPLC (reverse phase, normal phase), capillary electrophoresis, etc. HPLC is the most popular alternative to the standard methods and is commonly used nowadays for the analysis of almost all types of different food components.

See also: **Chromatography**: Principles; Thin-layer Chromatography; **Fatty Acids**: Analysis; **Homogenization**; **Protein**: Chemistry; **Sensory Evaluation**: Sensory Characteristics of Human Foods; **Spectroscopy**: Overview; Atomic Emission and Absorption; Nuclear Magnetic Resonance

Further Reading

AACC (1983) *Approved Methods of Analysis*, 8th edn. St. Paul, MN: American Association of Cereal Chemists.
AOAC (1995) *Official Methods of Analysis*, 16th edn. vols. 1–2. Arlington, Virginia USA: Association of Official Analytical Chemists.
AOCS (1990) *Official Methods of Recommended Practices*. Chicago, IL: American Oil Chemists Society.
Birch GC and Parker KJ (1984) *Control of Food Quality and Analysis*. London: Elsevier Applied Science.
Fennema OR (1996) *Food Chemistry*. New York: Marcel Dekker.
Gruenwedel DW and Whitakar JR (eds) (1984) *Principles and Techniques*, vols. 1–4. New York: Marcel Dekker.
James CS (1994) *Analytical Chemistry of Foods*. London: Blackie Academic & Professional.
Kulp K and Ponte JG Jr. (2000) *Handbook of Cereal Science and Technology*, 2nd edn. New York: Marcel Dekker.
Linden G (ed.) (1996) *Analytical Techniques for Foods and Agricultural Products*. New York: VCH.
Multon JL (ed.) (1997) *Analysis of Food Constituents*. New York: Wiley-VCH.
Nollet LML (ed.) (2000) *Food Analysis by HPLC*, 2nd edn. New York: Marcel Dekker.
Nollet LML (ed.) (1996) *Handbook of Food Analysis* Vols. 1–2. New York: Marcel Dekker.
Poumeranz Y and Meloan CE (1994) *Food Analysis, Theory and Practice*, 3rd edn. New York: Chapman & Hall.
Sorensen (2000) *Chromatography and Capillary Electrophoresis in Food Analysis*. New York: Springer.
Wilson RH (ed.) (1994) *Spectroscopic Techniques for Food Analysis*. New York: Wiley-VCH.

Analytical Techniques *See* **Chromatography**: Principles; Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; Supercritical Fluid Chromatography; Combined Chromatography and Mass Spectrometry; **Immunoassays**: Principles; Radioimmunoassay and Enzyme Immunoassay; **Mass Spectrometry**: Principles and Instrumentation; Applications; **Spectroscopy**: Overview; Infrared and Raman; Near-infrared; Fluorescence; Atomic Emission and Absorption; Nuclear Magnetic Resonance; Visible Spectroscopy and Colorimetry

ANEMIA (ANAEMIA)

Contents

Iron-deficiency Anemia

S R Lynch, Eastern Virginia Medical School, Norfolk, VA, USA

Introduction

A lack of iron is the commonest nutritional deficiency disorder in human beings. Anemia occurs when body iron stores have been exhausted and the rate of delivery of iron to developing red blood cells in the bone marrow is insufficient to support the level of heme synthesis required to maintain a normal circulating red cell mass and hemoglobin concentration. It is estimated that more than 500 million people throughout the world suffer from iron-deficiency anemia. This article will review the causes, methods of detection, physiological consequences, as well as treatment and prevention of iron-deficiency anemia. (*See* **Iron:** Physiology.)

Iron Balance

Iron balance in human beings is uniquely dependent on the body's ability to match the rate of iron absorption from the proximal small intestine to iron requirements. During childhood, iron absorption normally exceeds immediate requirements, ensuring a positive balance and the gradual establishment of an iron store. In the adult, the level of absorption is approximately equal to a relatively fixed rate of loss that is

governed by factors unrelated to iron balance. When requirements are increased, e.g., during periods of rapid growth and in pregnancy, absorption increases to replace the storage iron that is used up. (*See* **Children**: Nutritional Requirements; **Lactation**: Human Milk: Composition and Nutritional Value.)

Once iron stores are exhausted by increased demand, anemia supervenes if adaptive changes in the rate of absorption are insufficient to restore iron balance. Dietary iron intake almost always exceeds the body's requirements by a large margin, but the ability of absorptive mechanisms to extract the iron from food may be limited. Iron-deficiency anemia is therefore most commonly encountered when the relationship between requirement and dietary intake of bioavailable iron is least favorable. Periods of increased vulnerability include infancy, adolescence (particularly among girls after the onset of menstruation), and the child-bearing years for women.

After the first 4–6 months of age, stores are no longer the primary source of iron for growth. By the age of 12 months total body iron should have increased by about 70%. Iron-deficiency anemia is therefore commonly encountered in infancy. The risk tends to decrease after about 18 months of age because the rate of growth slows and the diet becomes richer in bioavailable sources of iron. The adolescent growth spurt and the onset of menstruation in girls constitute the second high-risk period. In adulthood, increased menstrual bleeding or blood loss from the gastrointestinal tract may contribute to an unfavorable balance. For example, the monthly menstrual loss in 10% of healthy women is more than 80 ml (average for most women is 30 ml). This increases the daily iron requirement by about 0.8 mg or 6% of the dietary intake. It would be necessary for such an individual to increase her absorption of dietary iron from about 10% to 16%. Hookworm infestation, which is prevalent in many developing countries, may induce sufficient blood loss to increase iron requirements by 3–4 mg day^{-1}. Iron balance would only be maintained if percentage absorption from the diet were increased by 20–30%. Many of these individuals will become anemic because maximal total absorption from a normal diet is usually only 4–5 mg day^{-1}. Pregnancy presents the greatest physiological challenge to the body's adaptive mechanisms. The daily iron requirement during the last trimester is between 4 and 6 mg. (*See* **Premenstrual Syndrome: Nutritional Aspects**.)

The diets of most western adults contain large quantities of meat, which is rich in highly available heme iron, as well as fresh fruit and other sources of vitamin C, that increase the absorption of inorganic iron. This diet provides sufficient bioavailable iron to prevent anemia even in the presence of higher requirements. For example, a recent study in which dietary iron absorption was measured over a 2-week period indicated that a menstruating woman eating a mixed diet containing meat would absorb about 10% of the iron and maintain an iron store of approximately 350 mg. The composite 10% absorption value is achieved by assimilation of about 30% of the heme iron and 6% of the nonheme iron. A woman with relatively high menstrual losses would also be expected to maintain iron balance by increasing her nonheme iron absorption to 12%, although her iron store would be estimated at only 200 mg.

The iron requirements during the second and third trimesters of pregnancy usually exceed the body's capacity to extract sufficient iron from the diet even if the bioavailability is high. Most women require a small iron supplement to prevent iron imbalance.

As indicated above, relative iron requirements are also high in infants between the ages of 6 and 18 months. Foods given to young children during and immediately after the weaning period usually have low iron bioavailability. However some countries have been very successful in preventing iron deficiency during this vulnerable period by fortifying infant foods with iron.

Iron balance is much less satisfactory in many people living in developing countries. Heme iron is often absent from the diet since meat is not eaten in significant quantities. The bioavailability of the nonheme iron is usually low because the staple foods tend to have high levels of phytate and polyphenols. Iron-deficient individuals may be able to absorb only 5–10% of the nonheme dietary iron. Under these circumstances even modestly increased requirements may not be met. Poor dietary iron availability is often accompanied by high requirements due to factors such as multiple pregnancies and hookworm infestation. As a result, iron-deficiency anemia is widespread.

Prevalence

On the basis of data obtained during the third National Health and Nutrition Examination Survey (NHANES III, 1988–1994) the highest prevalence rates for iron-deficiency anemia in the USA occur in toddlers (3%), and adolescent girls and menstruating women (2–5%). Iron-deficiency anemia is rare in men and postmenopausal women. Among the elderly, inflammatory disease, not iron deficiency, is the commonest cause of anemia. Iron deficiency is much more frequently encountered in developing countries. At least 500 million people are affected globally.

Laboratory Diagnosis

Iron-deficiency anemia may be suspected on the basis of the clinical symptoms and the setting in which it occurs, but precise identification depends on the results of laboratory tests designed to confirm the presence of anemia and to evaluate iron status (Table 1).

Identification of Anemia

Anemia is identified by the finding of a hemoglobin concentration or hematocrit that is below the normal range. This usually presents little difficulty in the clinical situation since symptoms only occur in individuals with moderate or severe anemia. On the other hand, prevalence studies of nutritional anemia depend on the identification of mild anemia in an otherwise healthy population. The World Health Organization (WHO) criteria established in 1975 are the most widely accepted. Hemoglobin concentrations below 13 g dl^{-1} for adult males, 12 g dl^{-1} for menstruating women, and 11 g dl^{-1} in pregnancy are considered indicative of anemia.

In mild iron deficiency, the diagnostic value of a low hemoglobin concentration or hematocrit is limited by the significant overlap in the frequency distribution curves for anemic and normal populations. In one study using a single value criterion, 17% of women responded to oral iron despite the fact that they were classified as normal, while 35% of 'anemic' women failed to show any change in hemoglobin after being treated. The problem is even more complex in children, where developmental changes affect the normal hemoglobin concentration. The accuracy of diagnosis can be improved by using a criterion that employs an abnormal hemoglobin or hematocrit value in combination with one or more laboratory markers of iron deficiency.

Evaluation of Iron Status

Laboratory tests of iron status are used both to distinguish iron deficiency from other causes of anemia and to facilitate the recognition of mild iron-deficiency anemia in nutritional surveys. They supply specific information about the adequacy of iron stores and the supply of iron to the bone marrow and other organs.

Table 1 Laboratory evaluation of iron-deficiency anemia

Purpose of test	*Laboratory test*
Confirm presence of anemia	Hemoglobin
	Hematocrit
Evaluate iron status	
Iron stores	Bone marrow iron
	Serum ferritin
Iron supply	Serum transferrin receptor concentration
	Transferrin saturation
	Free erythrocyte protoporphyrin
	Red blood cell indices

Iron Stores

Bone marrow iron Histological evaluation of storage iron by using the Prussian-blue method to stain macrophage hemosiderin in needle aspirate or biopsy specimens of bone marrow provides a semiquantitative estimate of storage iron. Iron-deficiency anemia is characterized by an absence of stainable bone marrow iron. This procedure remains the gold standard in patients suffering from complicated illnesses, but is rarely necessary under most other circumstances and impractical in nutritional surveys.

Serum ferritin Body iron not immediately required for hemoglobin synthesis or other metabolic processes is stored in cells as ferritin and hemosiderin. Ferritin is also present in very small amounts in the plasma. The development of sensitive immunological methods has made it possible to measure serum ferritin (SF) accurately. This has become the most useful indirect estimate of body iron stores. SF concentration is directly proportional to the size of the storage pool in health and in uncomplicated iron deficiency: 1 μg l^{-1} is equivalent to 8–10 mg storage iron in adults (SF 1 μg l^{-1} is equivalent to 140 μg kg^{-1} storage iron). A serum ferritin below 12 μg l^{-1} is diagnostic of iron deficiency. Unfortunately, higher SF levels do not exclude iron deficiency when disorders such as infection, chronic inflammatory diseases, malignant disorders, and liver disease are present because they elevate SF values independently.

Internal Iron Supply

Serum transferrin receptor concentration Transferrin receptors are expressed on the surfaces of all cells in concert with their need for iron. A truncated form of the extracellular domain of the receptor is produced by proteolytic cleavage and is present in the plasma in direct proportion to the total number of receptors expressed in all body tissues. When functional iron depletion occurs more transferrin receptors appear on cell surfaces and the concentration of the proteolytically cleaved extracellular domain (serum transferrin receptor or STfR) in the plasma increases concomitantly. The magnitude of the increase is proportional to the size of the functional iron deficit. This assay has only recently become available, but may prove to be a very important measure of functional iron deficiency. The transferrin saturation and free erythrocyte protoporphyrin levels

are both abnormal in the anemia of chronic disease even if iron stores are adequate because the iron supply is reduced by impaired release from stores. The serum transferrin receptor concentration may help to distinguish these two common causes of anemia, since serum transferrin receptor levels are normal in the anemia of chronic disease.

Transferrin saturation As body iron stores become depleted, the concentration of the iron transport protein, transferrin, in the plasma increases, leading to a rise in the total iron-binding capacity (TIBC). Once iron stores are exhausted and the tissue demands exceed the rate of supply, the serum iron (SI) concentration falls. Since an impaired iron supply in an iron-deficient individual is characterized by a rise in TIBC and a fall in SI, a reduction in the transferrin saturation ($TS = SI \times 100/TIBC$) from the normal value of approximately 35% to less than 15% is the best single criterion of a suboptimal supply.

Free erythrocyte protoporphyrin An impaired iron supply can also be recognized by measuring the protoporphyrin IX (free erythrocyte protoporphyrin or FEP) in peripheral red blood cells. Insufficient iron limits the final step in heme synthesis, leading to the accumulation of FEP in red cell precursors. The cellular content remains unchanged throughout the life span of the erythrocyte. Peripheral blood FEP values therefore reflect iron supply during red cell production. FEP levels are also raised in lead poisoning. If exposure to lead can be excluded, a raised FEP provides the same information as a reduced TS, but it is more stable. FEP is raised only after several weeks of iron-deficient erythropoiesis and requires a similar period for normalization once an adequate supply of iron is reestablished.

Red blood cell indices As the severity of iron deficiency progresses, morphological abnormalities (microcytosis and hypochromia), resulting from reduced hemoglobin synthesis, appear in the peripheral blood. The sensitivity with which these changes can be recognized has improved dramatically with the introduction of electronic counting and sizing equipment. Both the mean corpuscular volume (MCV) and the mean corpuscular hemoglobin (MCH) are reliable measures of impaired hemoglobin synthesis, although not specific for iron deficiency. The MCV is the more widely used.

The best approach to diagnosing iron-deficiency anemia depends on the setting in which it is being evaluated. The accuracy of prevalence studies of nutritional anemia is improved by the concurrent measurement of hemoglobin or hematocrit and one or more specific tests for iron status, such as SF, STfR, FEP, or TS. The SF concentration is currently the most widely used measure of iron status. A similar approach or a therapeutic trial of oral iron may be satisfactory for clinic or office patients in whom iron deficiency is considered likely. In hospitalized patients, anemia complicated by other illnesses may occasionally require the direct histological evaluation of bone marrow iron stores. (*See* **Nutritional Assessment**: Anthropometry and Clinical Examination; Biochemical Tests for Vitamins and Minerals; Functional Tests.)

Physiological Consequences of Iron-deficiency Anemia

Mild anemia usually causes few symptoms in individuals with sedentary occupations because of compensatory physiological adjustments, including increased cardiac output and a shift in the oxygen dissociation curve with improved oxygen delivery to tissues. Aerobic work capacity in human beings is however reduced when the hemoglobin concentration falls below $12\,g\,dl^{-1}$.

Patients with moderate or severe anemia are pale and have symptoms such as reduced exercise capacity, shortness of breath, and palpitations that are indicative of a decreased cardiopulmonary reserve.

Iron deficiency usually develops gradually. Compensatory adjustments to oxygen transport insure adequate oxygen delivery at rest. However the ability to carry out sustained physical activity is significantly impaired, even in relatively mild iron-deficiency anemia, because of a reduction in available oxygen and the associated tissue iron deficiency that directly affects skeletal muscle function. Manual laborers who suffer from iron-deficiency anemia are less productive and may have a reduced earning capacity.

A study carried out among Indonesian rubber plantation latex tappers, who were paid a daily wage based on the weight of latex they collected, demonstrated a direct correlation between hemoglobin concentration and income. Anemic men were 18.7% less productive than their nonanemic fellow workers; when they were given 100 mg elemental iron per day for 60 days their income increased by 37%. It has been estimated that for every 1% increase in hemoglobin, work output will increase by 1–2%.

Iron-deficiency anemia also has insidious effects in children. Mental and motor development is impaired in infants. Cognitive function and school performance are affected in older children. The observed abnormalities are thought to be due primarily to direct effects of iron deficiency on cellular and neurotransmitter function. They are discussed in the chapter on physiology.

During pregnancy iron-deficiency anemia has significant consequences for the mother and the infant. Maternal anemia is associated with low infant birth weight, premature delivery, and increased perinatal mortality. Furthermore, although fetal requirements tend to be met at the expense of the mother's needs, several studies indicate that infants born to iron-deficient women have lower iron stores at birth and lower SF levels at 3–6 months of age. The risk of maternal death is also increased in women with severe anemia.

Finally iron-deficiency anemia may be associated with several abnormalities that are considered to be specific consequences of a tissue iron lack. They include impaired immunity, mucosal and epithelial changes, the eating of nonfood material (pica), and impaired temperature regulation.

Treatment and Prevention

Treatment

The recognition of iron-deficiency anemia is never a complete diagnosis. Iron administration may produce a rapid and gratifying correction of the anemia, but a satisfactory long-term solution to the problem depends on a clear understanding of the reasons for increased requirements and/or reduced absorption. For example, iron-deficiency anemia in western men is usually a symptom of unsuspected blood loss from the gastrointestinal tract. Identification of the underlying disorder is essential (Table 2).

Oral iron therapy using a soluble ferrous salt, such as sulfate, gluconate, or fumarate, that provides a daily dose of 60–120 mg elemental iron in adults, or 2–3 mg kg^{-1} in children, is usually the most appropriate method both to repair the anemia and restore a normal body iron store in patients with moderate or severe anemia. Once this has been accomplished, continued iron balance can only be assured if the cause of excessive requirements is eliminated or if more available dietary iron is supplied (in situations in which iron-deficiency anemia is the result of an inadequate intake of bioavailable iron).

Prevention

The prevalence of nutritional iron-deficiency anemia in western societies has fallen markedly in recent years. The many factors that have been suggested to contribute to this improvement in adults include dietary fortification, a general improvement in income level and diet, a large increase in the sale of medicinal iron preparations, routine iron supplementation in pregnancy, and reduced menstrual blood loss as a result of the use of oral contraceptives.

Table 2 Causes of iron-deficiency anemia

Underlying disorder	*Site of action*	*Causal factors*
Nutritional deficiency		Low dietary bioavailability
		Unrefined cereal diet
Malabsorption		Gastric surgery
		Clay eating
		Gluten-induced enteropathy
Blood loss	Uterine	Menorrhagia
	Gastrointestinal	Hookworm, schistosomiasis
		Peptic ulcer disease, esophagitis
		Stomach and colorectal cancer
		Angiodysplasia, hereditary telangiectasia
		Aspirin, NSAIDs
		Inflammatory bowel disease
		Diverticulosis, hemorrhoids
	Renal tract	Hemoglobinuria
		Severe hematuria
	Lung	Idiopathic pulmonary hemosiderosis

NSAIDs, nonsteroidal antiinflammatory drugs.

Similarly, increasing awareness of the importance of adequate iron nutrition in early infancy has led to a reduction in the prevalence of iron deficiency in young children. As with adults, improved iron nutrition reflects the effect of several different factors, including the use of iron-fortified infant foods and the provision of supplemental iron when appropriate.

The situation in developing countries is far less satisfactory. Attempts to reduce iron losses or improve the intake of bioavailable iron have had only limited success. Iron supplementation (usually the provision of iron tablets) has been effective in pilot studies, but the inadequacy and cost of distribution systems, as well as poor compliance, have limited the value of this approach as a long-term solution. Nevertheless, supplementation is the preferred method of intervention in severely anemic individuals and during pregnancy.

Increasing the population's intake of bioavailable dietary iron is the only feasible means of improving iron nutrition at the national level. Despite years of research and numerous field trials, this has proved to be extremely difficult to achieve. The traditional diet of people in most developing countries tends to inhibit iron absorption. Dietary customs are not easily modified. In addition, foods that promote iron bioavailability tend to be expensive. Iron

fortification of the diet has also encountered several obstacles. The introduction of iron into the food chain in developing countries is complicated by the difficulty of identifying a suitable vehicle. Most foods are produced at the local level and not processed or distributed through a limited number of centers where the fortificant can be introduced into the food. In addition, the inhibitory effect of vegetable diets on bioavailability tends to reduce the effectiveness of fortification iron. Finally it is often difficult to add iron to dietary items without affecting consumer acceptability. Iron salts that do not alter the appearance or organoleptic and storage properties of foods tend to be poorly absorbed. Despite these formidable obstacles, field trials and targeted national programs have been successful in Chile, South Africa and, more recently, Venezuela. (*See* **Food Fortification.**)

See also: **Children**: Nutritional Requirements; **Folic Acid**: Properties and Determination; Physiology; **Food Fortification**; **Iron**: Physiology; **Lactation**: Human Milk: Composition and Nutritional Value; Physiology; **Nutritional Assessment**: Anthropometry and Clinical Examination; Biochemical Tests for Vitamins and Minerals; Functional Tests; **Premenstrual Syndrome: Nutritional Aspects**

Further Reading

Brock JH, Halliday JW, Pippard MJ and Powell LW (1994). *Iron Metabolism in Health and Disease*. London: WB Saunders.

Cook JD (1982) Clinical evaluation of iron deficiency. *Seminars in Hematology* 19: 6–18.

Cook JD (1990) Adaptation in iron metabolism. *American Journal of Clinical Nutrition* 51: 301–308.

Cook JD, Skikne BS and Baynes RD (1993) Serum transferrin receptor. *Annual Review of Medicine* 44: 63–74.

Looker AC, Dallman PR, Carroll MD, Gunter EW and Johnson CL (1997) Prevalence of iron deficiency in the United States. *Journal of the American Medical Association* 277: 973–976.

Megaloblastic Anemias

N M F Trugo, Universidade Federal do Rio de Janeiro, Instituto de Quimica, Rio de Janeiro, Brazil

Background

Megaloblastic anemia is characterized by morphological abnormalities of hemopoietic cells that include the formation of abnormally large erythrocyte precursors (megaloblasts) and giant metamyelocytes in the bone marrow, and abnormally large (macrocytic) erythrocytes and hypersegmented neutrophils in the blood. Megaloblastic anemia is caused by a reduction in the rate of DNA biosynthesis, resulting in abnormal nuclear maturation and ineffective erythropoiesis. Folate and cobalamin deficiencies are the most common causes of the impaired DNA synthesis that leads to megaloblastic hemopoiesis; other less common factors are the use of drugs that interrupt DNA biosynthesis and inherited conditions presenting defective enzymes of DNA biosynthesis. Megaloblastic anemia caused by folate and/or cobalamin deficiencies present indistinguishable morphological abnormalities, and will be the main focus in this chapter.

Etiology

Consequences of Folate and Cobalamin Deficiencies

Folate and cobalamin deficiencies are primary causes of megaloblastic anemia. Folate is an essential cofactor involved in many single-carbon transfer reactions in intermediary metabolism, including the synthesis of purine bases, thymidylate and methionine. Folate deficiency thus causes impairment in the DNA synthesis and in the methylation reactions that are dependent on *S*-adenosylmethionine.

Both folate and cobalamin are required for the methyl transfer reaction catalyzed by methionine synthase. In this reaction methionine is recycled by remethylation of homocysteine and simultaneous demethylation of 5-methyltetrahydrofolate to tetrahydrofolate, a source of other folate coenzymes. Methionine synthase requires methylcobalamin as a cofactor. Impairment of this reaction by cobalamin deficiency thus causes a decrease in the availability of 10-formyl-tetrahydrofolate for purine synthesis and of 5,10-methylene-tetrahydrofolate for conversion of deoxyuridylate into deoxythymidylate, which might be explained by 'trapping' of methyltetrahydrofolate that cannot be converted back to the other folate forms, resulting in a decrease of folate coenzymes. An alternative hypothesis is that the impairment of the methionine synthase reaction decreases the supply of formate that would be required for generation of 10-formyl-, 5,10-methenyl- and 5,10-methylene tetrahydrofolate, due to the failure of methionine synthesis and resulting low levels of *S*-adenosylmethionine.

Therefore, deficiencies of both or either of these vitamins affect DNA replication and consequently cell division, especially in rapidly proliferating

tissues. In the bone marrow, the impaired DNA replication causes asynchronous maturation of the nucleus and cytoplasm of cells in the erythroid and myeloid series, with cytoplasmic components being synthesized in excessive amounts during the delay between cell divisions. This results in the enlargement of the precursor cells and the presence of larger than normal circulating erythrocytes and leukocytes, which are characteristic features of megaloblastic anemia.

Besides megaloblastic anemia, cobalamin deficiency also causes a neuropathy characterized by an early demyelination of peripheral nerves, as a result of the impaired methylation of the myelin basic protein, a major constituent of myelin. This is currently explained by the inadequate activity of methionine synthase in cobalamin deficiency, which leads to a reduction in the methylation cycle due to the impaired recycling of *S*-adenosylmethionine. Several evidences indicate that the other enzymatic reaction requiring cobalamin as a coenzyme (5′-deoxyadenosylcobalamin), which is the conversion of methylmalonyl-coenzyme A (CoA) to succinyl-CoA by methylmalonyl-CoA mutase, seems not to be implicated in the neuropathy of cobalamin deficiency. Folate deficiency usually does not lead to this neuropathy, unless the deficiency is prolonged and severe, possibly because the neural tissue concentrates more folate than other tissues and may also have alternative pathways to preserve the supply of *S*-adenosylmethionine when there is folate deprivation. (*See* **Cobalamins**: Physiology; **Folic Acid**: Physiology.)

Causes of Cobalamin and Folate Deficiencies

Deficiencies of cobalamin and folate may result from inadequate intake, increased requirements and impaired utilization, due to several underlying factors that are presented in Tables 1 and 2.

Cobalamin deficiency due to insufficient dietary intake is rare, but may occur in strict vegetarians, since cobalamin is supplied primarily by foods of animal origin. Recommended daily intakes of dietary cobalamin are in the range of 1–2 μg for adults. A normal western diet usually provides an excess of the body requirements for cobalamin, and cobalamin body stores represent a further protection, with an efficient conservation through the enterohepatic circulation.

Among the causes of cobalamin deficiency described in Table 1, the most common is intestinal malabsorption, which can be the result of a decreased production or lack of gastric intrinsic factor, a protein required for cobalamin absorption, as occurs in atrophic gastritis, gastrectomy, and pernicious anemia, an autoimmune disease in which there is production of autoantibodies against intrinsic factor and of autoantibodies that cause destruction of gastric parietal cells. Moreover, hypochlorydria or achlorydria and low pepsin secretion in the stomach affect release of cobalamin from its bound state in food. When this happens, but secretion of intrinsic factor is normal, there is protein-bound cobalamin

Table 1 Causes of cobalamin deficiency

Reduced intake
Strict vegetarianism
Breast-feeding by severely cobalamin-deficient mothers
Malabsorption
Pernicious anemia
Gastric dysfunction caused by gastritis due to *Helicobacter pylori* infection or alcohol consumption
Gastrectomy or gastric bypass
Protein-bound cobalamin malabsorption
Enteropathy: Crohn's disease, celiac disease
Ileal resection
Pancreatic insufficiency
Microbial overgrowth in intestinal stasis and parasitic (fish tapeworm) infestation
Use of drugs such as colchicine, neomycine, omeprazole, or *p*-aminosalicylic acid
Congenital absence or dysfunction of intrinsic factor
Lack of ileal receptor for intrinsic factor
Impaired utilization
Congenital transcobalamin II deficiency
Inactivation of cobalamin by prolonged exposure to nitrous oxide
Decreased conversion to active coenzyme forms of cobalamin due to enzyme defect

Table 2 Causes of folate deficiency

Reduced intake
Malnutrition
Alcoholism
Intake of foods poor in folate (raw; processed)
Increased demands
Pregnancy
Rapid growth rate (prematurity; early infancy; adolescence)
Diseases associated with increased cell turnover (hemolytic anemias; leukemia; chronic diseases; exfoliative dermatitis; severe infection)
Excessive losses (hemodialysis; liver disease)
Malabsorption
Enteropathy: Crohn's disease, celiac disease, tropical sprue
Intestinal resection
Chronic giardiasis
Bacterial overgrowth
Use of sulfasalazine and of anticonvulsive drugs (phenytoin, phenobarbital)
Impaired utilization
Congenital abnormalities in enzymes of folate metabolism (5,10-methylene tetrahydrofolate reductase; methionine synthase; di-hydrofolate reductase)
Use of drugs such as methotrexate and oral contraceptives
Alcoholism

malabsorption, because cobalamin from food is poorly absorbed whereas absorption of ingested synthetic cobalamin is not affected.

Differently from cobalamin, dietary inadequacy is a common cause of folate deficiency (Table 2), since folate intake is not usually in great excess of nutritional requirements and the body reserves are meager. Recommended daily intakes of folate are in the range of 0.15–0.20 mg for adults, 0.4 mg for women of reproductive age, and 0.6 mg for pregnant women. Diets poor in folate are characterized by a predominance of starches and grains (white bread, white maize meal, polished rice, etc.) and overcooked vegetables, and insufficient amount or lack of foods of animal origin, dark green vegetables, fresh or lightly cooked, and fruits, as occurs in some parts of Africa and India. In developed countries, nutritional folate deficiency has been found in elderly persons living on a marginal subsistence diet, and in people on very low incomes. Alcoholism may cause folate deficiency, not only because of the reduced folate intake, but also due to altered folate metabolism and interference with the folate enterohepatic cycle.

Intestinal malabsorption can also be a cause of folate deficiency and may be associated with ineffective food folate deconjugation prior to absorption as occurs in subtotal gastrectomy and with the use of sulfasalazine, as well as with ineffective absorption due to extensive intestinal resection, tropical sprue, gluten-sensitive enteropathy, chronic giardiasis, and the use of drugs such as anticonvulsants (phenytoin and phenobarbital).

Due to the high folate requirements, periods of intense cell proliferation such as pregnancy, early infancy, and adolescence predispose to folate deficiency. Increased requirements of folate may also occur in diseases such as hemolytic anemia, malignant metastatic tumors, and leukemias. Folate deficiency may also occur with the use of drugs that are folate antagonists and impair folate utilization, such as those used for cancer chemotherapy (methotrexate) and immunosuppression, as well as antiprotozoal (pyrimethamine) and antibacterial (trimethroprim, pentamidine) drugs.

A genetically carried impairment in the folate utilization, due to one variant of 5,10-methylenetetrahydrofolate reductase that has a lower affinity for folate, and which is apparently common in many populations, increases folate requirements. Another genetic problem related to 5,10-methylenetetrahydrofolate reductase is its deficient production in some patients, leading to hyperhomocysteinemia, homocystinuria, and neurological disturbances, but without any hematological abnormality. (*See* **Cobalamins**: Physiology; **Folic Acid**: Physiology.)

Morphological and Clinical Aspects

The reduction in DNA biosynthesis in the bone marrow delays the development of the erythrocytes, leading to megaloblastosis of erythocyte precursors and macrocytosis of erythrocytes, and reduces their number, leading to anemia. The megaloblastic anemia is morphologically indistinguishable in the deficiency of either vitamin. The changes in the erythroid and myeloid cells result in pallor and weakness.

In the bone marrow, there is an overt megaloblastosis and a reduction in the ratio of myeloid to erythroid cells. Megaloblasts present a characteristic chromatin pattern with a lacy appearance. There are more immature erythroblasts than mature ones, due to selective death of the latter. In the myeloid series, giant and abnormally shaped metamyelocytes (granulocyte precursors) are found. Megakaryocytes (platelet precursors) may also be enlarged, with increased numbers of nuclear lobes.

In blood, erythrocytes are large (macrocytic) and often oval in shape, and some may be fragmented and irregularly shaped. Neutrophilic leukocytes often have abnormal nuclei, with an increase in the number of nuclear segments (hypersegmentation). There is an increased mean corpuscular volume of red cells, and decreased erythrocyte, reticulocyte, leukocyte, and platelet counts in peripheral blood, due to ineffective erythropoiesis that results in cell destruction before maturity. The decrease in erythrocyte counts leads to low hemoglobin values, as the deficiency becomes more severe.

Besides the hemopoietic system, the epithelial cells of gastrointestinal, respiratory, urinary, and female genital tracts, and the gonads are also affected. The decreased replication of enterocytes in the small intestine leads to villus atrophy that, if severe enough, can cause malabsorption. Other symptoms, such as glossitis (sore tongue), cheilitis (sore lips), diarrhea, and weight loss may thus result from impaired proliferation of epithelial cells. Infertility and impotence may also be present.

Cobalamin deficiency may also result in a complex neurologic syndrome, manifested with or without megaloblastic anemia, including peripheral neuropathy, subacute combined degeneration of the spinal cord, and optic atrophy. Common neurological symptoms are paresthesias, ataxia, numbness, impairment of memory, depression, dementia, and psychosis. Folate deficiency can occasionally cause mental changes, such as mental slowing, depression, dementia, and other neurologic syndromes.

There are, however, variations in the hematologic and neurologic expressions of cobalamin and folate deficiencies among patients. In children, for instance,

hereditary and acquired deficiencies result more often in developmental delay and neurological dysfunctions than in megaloblastic anemia. Also, genetic and acquired characteristics of ethnic and racial groups seem to impose variations related to cobalamin deficiency, which could have implications in diagnosis and issues of public health.

During pregnancy, risks from folate deficiency include poor pregnancy outcome, such as prematurity and low birth weight. In addition, there is considerable evidence that folate is a key factor for prevention of neural tube defects (NTD; congenital birth defects such as spina bifida and anencephalus). Although NTD have multiple etiologies, maternal folate supplementation during the periconceptional period can reduce both their occurrence and recurrence. The first 3 weeks after conception represent a crucial period because closure of the neural tube occurs about the 23rd day. Although the underlying mechanism for the folate role in the prevention of NTD is not known, there are indications that in the vast majority of cases the problem is not of maternal dietary deficiency, and may be due to a maternal and/or fetal defect in folate metabolism or transport into the cells, that can be overcome by maternal supplementation. An impairment in homocysteine metabolism seems to be involved in NTD manifestation. Although NTD may affect pregnancies in women who present a wide range of folate erythrocyte levels, from low to normal, the risk is strongly reduced when status is improved, even in those women who do not present low erythrocyte folate levels. Cobalamin status, not necessarily in the deficient range, is also an independent risk factor for NTD. (*See* **Cobalamins**: Physiology; **Folic Acid**: Physiology; **Pregnancy**: Maternal Diet, Vitamins, and Neural Tube Defects.)

Laboratory Diagnosis

Hematology

An elevated blood mean corpuscular volume (MCV) due to macrocytosis is suggestive of megaloblastosis. In cobalamin and folate deficiencies the MCV tends to increase before the decrease in the hemoglobin levels. The increase in MCV may not be present when there is concurrent microcytosis due to iron deficiency and thalassemia. Examination of a bone marrow aspirate may be useful for the diagnosis of megaloblastic anemia, since macrocytosis can be due to other conditions (chronic alcoholism, liver disease, etc.) that do not cause changes in the bone marrow. However, megaloblastosis can also be due to the use of anticancer drugs, mainly methotrexate, which impair folate metabolism. The presence of hypersegmented neutrophils, with five or more nuclear lobes, in the peripheral blood smear is a sensitive indicator of megaloblastic anemia, and is useful for the diagnosis of folate and cobalamin deficiencies when erythrocyte changes are masked by coexistent iron deficiency and when folate and cobalamin deficiencies are still subclinical, because hypersegmentation occurs earlier than the anemia and macrocytosis.

Biochemical Tests

A range of biochemical tests are used to investigate whether megaloblastic anemia is due to folate or cobalamin deficiency. Biochemical indicators of either vitamin status corresponding to different degrees of depletion are used for this purpose, despite the problems in establishing 'cut-off' levels.

Serum (or plasma) cobalamin may be a poor indicator of tissue levels, but levels lower than 200 pg ml^{-1} probably indicate a deficiency state of cobalamin. About one-third of patients with folate deficiency present low serum cobalamin levels, which return to normal with folate therapy. Low serum folate (less than 3 ng ml^{-1}) is an indicator of negative balance that, if persistent, can result in tissue depletion, which is assessed by determination of erythrocyte folate. Values of erythrocyte folate of less than 150 ng ml^{-1} are indicative of an initial folate deficiency (marginal status) and those less than 120 ng ml^{-1} indicate established deficiency. These three indicators are the most widely used to assess cobalamin and folate status. They can be measured by microbiological assays or by competitive protein-binding radioassays, but since values vary between assay procedures, the range of normal values should be established for each procedure.

Once established whether megaloblastic anemia is due to folate or cobalamin deficiency, or to other causes, the underlying causes of these deficiencies should also be investigated. For instance, dietary assessment is useful to detect whether dietary insufficiency of either vitamin is an underlying factor.

Other biochemical and functional tests provide valuable information regarding the nutritional status of folate and cobalamin and the underlying causes of deficiency. Serum homocysteine is a sensitive indicator of intracellular folate and cobalamin, and it is increased when depletion of either vitamin occurs because they are required for conversion of homocysteine to methionine.

The deoxyuridine suppression test in bone marrow cells or in lymphocytes assesses the thymidylate synthesis from deoxyuridylate, which requires 5,10-methylene tetrahydrofolate as a coenzyme and is impaired in folate and cobalamin deficiencies. Exogenous deoxyuridine added to normal cells

suppresses the incorporation of subsequently added radiolabeled thymidine to DNA due to conversion of the added deoxyuridine to thymidylate within the cell. In cells from deficient patients, suppression by deoxyuridine is less efficient due to blockage of this conversion. Folate and cobalamin deficiencies can be distinguished by this test through its correction with the respective vitamins.

Determination of serum holohaptocorrin (cobalamin-carrying haptocorrin), the cobalamin storage glycoprotein, can be used to assess cobalamin body stores. Measurement of holotranscobalamin II, the cobalamin transport protein to tissues, may also be used, because it falls below the normal range before serum cobalamin does. The increase of urinary and serum methylmalonic acid concentrations indicates cobalamin deficiency, because of the impairment in the methylmalonyl-CoA mutase reaction, which requires adenosylcobalamin as a cofactor.

The Schilling test is used to detect cobalamin malabsorption. This test is performed usually in two stages: (1) an oral dose of radiolabeled cobalamin is given and its percentage of excretion in urine during a 24-h collection is determined, after flushing it into urine with a large parenteral dose of unlabeled cobalamin; (2) another test is performed 3–7 days later, with an oral dosing of radiolabeled cobalamin and intrinsic factor. There are several interpretations of the Schilling test regarding the causes of malabsorption in cobalamin-deficient patients, depending on the results in each of the two stages. For instance, when the result is abnormal in the first stage, but normal in the second, pernicious anemia, or any other cause of absence, abnormal or insufficient amount of intrinsic factor, is present. A normal Schilling test in these patients may indicate dietary deficiency or protein-bound cobalamin malabsorption. Pernicious anemia can also be diagnosed by the presence of parietal cell autoantibodies and of antiintrinsic factor autoantibodies in serum.

The biochemical tests described above can all be useful to detect folate and cobalamin deficiencies at earlier stages, the so-called subclinical or marginal deficiencies, well before the manifestation of the clinical and overt signs of deficiency. The recognition of the atypical and subclinical deficiency states is an important step towards preventing not only the progression of deficiency to the clinical manifestations, but also to avoid the consequences that even marginal deficiencies of both vitamins can cause.

Folate subclinical deficiency has been associated with increased risk for dysplasia and various cancers, such as in the uterine cervix, bronchus, and colon, due to uracil misincorporation into DNA, impaired chromosome repair, and DNA strand breaks. Other adverse associations are immunological changes; birth defects, such as the NTD; and abnormalities of the homocysteine metabolism, leading to hyperhomocysteinemia. This last condition is an independent risk factor for cerebral, coronary, and peripheral vascular diseases, and occurs in both folate and cobalamin subclinical deficiencies. (*See* **Cobalamins**: Physiology; **Folic Acid**: Physiology.)

Prevalence

Recent and comprehensive worldwide epidemiological data on megaloblastic anemia and on subclinical folate and cobalamin deficiencies, especially in tropical and developing countries, the ones most likely to be affected by these problems, are scarcely available. Most data concerning the prevalence and distribution of these deficiencies were obtained more than two decades ago.

Those earlier studies showed that nutritional folate deficiency had a worldwide distribution, differing greatly in severity, with the subclinical, nonanemic forms being predominant, and was, in general, a primary cause of megaloblastic anemia. It affected mainly pregnant women in the third trimester of pregnancy, since their increased folate requirements may be difficult to achieve through the habitual diet, especially in low socioeconomic groups. Megaloblastic anemia due to dietary folate deficiency was reported to be highly prevalent in Asian countries such as India, Burma, Singapore, and Malaysia, and in African countries, affecting up to 25% of nonsupplemented pregnancies in certain parts of Asia, Africa, and Latin America, and 2.5–5.0% in developed countries. These values were higher (up to 60% in developing countries) when bone marrow aspirates were used for the diagnosis. Subclinical folate deficiency was estimated to affect up to one-third of the pregnant women on a global scale. The widespread use of iron and folate supplements substantially reduced the prevalence of folate deficiencies during pregnancy. Presently, the patterns of folate deficiency may still be the same, but scattered data indicate a possible decline in the incidence and prevalence of megaloblastic anemia, with a relative increase of subclinical deficiencies. In developing countries subclinical deficiency is high in nonsupplemented pregnancies. Folate deficiency, especially subclinical, is highly prevalent in the elderly.

Cobalamin deficiency due to inadequate intake has been mostly reported in strict Hindus from the Indian subcontinent. Recent studies show that cobalamin deficiency is more prevalent in other countries than formerly believed, but it is mostly due to malabsorption. In developed countries, malabsorption of

cobalamin affects mainly elderly populations. In northern Europe, cobalamin malabsorption is mainly due to pernicious anemia, affecting 1–2.5% of the population over the age of 60. Its frequency in caucasians is higher in women than in men (about 1.5:1), and increases with increasing age, being rare before 30 years of age. There is recent indication that cobalamin malabsorption due to pernicious anemia, and not folate deficiency, may be the main underlying factor of megaloblastic anemia in sub-Saharan Africa, as shown in Zimbabwe, where 86% of patients with megaloblastic anemia had cobalamin deficiency mainly due to pernicious anemia. In rural Mexico, a high prevalence of low levels of cobalamin in plasma of preschool children and in plasma and milk of lactating anemic women has been reported and seems to be caused by malabsorption.

Prevention and Treatment

The prevention and treatment, not only of clinical manifestations of cobalamin and folate deficiencies, but also of the subclinical deficiencies are important to avoid their consequences. The association of subclinical deficiencies of these vitamins with adverse conditions, as mentioned in a previous section, provides support for the case of improving the nutritional status of both vitamins in individuals in the earlier stages of deficiency progression.

The consumption of foods rich in folate and cobalamin is the best prophylaxis for deficiencies of nutritional origin of both vitamins. However, in conditions of increased requirements, such as pregnancy, the adequate intake of food folate may be difficult to attain. Therefore, due to the overall high incidence of folate deficiency in pregnant women, folate supplements ($0.4\,mg\,day^{-1}$) are usually recommended in the latter half of pregnancy during prenatal care. Since folate supplementation in the periconceptional period results in a decreased risk of NTD-affected pregnancies, in developed countries such as the USA, Canada, UK, and Australia it is recommended that all women of child-bearing age consume $0.4\,mg\,day^{-1}$ of folic acid. For women with a prior confirmed or suspected NTD-affected pregnancy, a supplement of 4 mg of folic acid is recommended if they might possibly become pregnant.

Possible public health strategies for supplying women with folate in the periconceptional period are the use of supplements by all women of child-bearing age, changes in food habits, and food fortification with folate. This last alternative, however, has raised concern since a higher folate intake through fortified foods by the general population could be detrimental to individuals with cobalamin deficiency. Although the elderly should benefit from increased folic acid intake added to foods, in those with cobalamin deficiency high intakes of folic acid could contribute to the development of neurologic abnormalities.

In the case of strict vegetarians, supplements or fortified foods providing cobalamin should be used. Folate supplements may also be recommended for people with sickle-cell anemia.

Megaloblastic anemia due to folate deficiency is usually treated with oral syntethic folic acid ($5\,mg\,day^{-1}$) for 2 months, or more depending on the etiology. In the case of malabsorption, parenteral folinic acid (5–10 mg) administered daily may also be used.

Treatment of megaloblastic anemia due to cobalamin deficiency resulting from low intake may be carried out with oral administration of cyanocobalamin or intramuscular injections of $1\,mg\,day^{-1}$ for 1 week, followed by the use of oral supplements or fortified foods containing 2 μg cyanocobalamin if the diet cannot be improved. When megaloblastic anemia is due to malabsorption, intramuscular injections of 1 mg hydroxocobalamin are given daily for 1 month, followed by monthly or bimonthly injections for life in the case of pernicious anemia. In the cases where the causes of cobalamin deficiencies are abnormal transport and utilization, daily doses of 1 or 5 mg are administered indefinitely. For patients with neurological manifestations due to cobalamin deficiency, 1 mg of cyanocobalamin every 2 weeks for 6 months is recommended. In the case of cobalamin deficiency, folic acid supplementation may correct megaloblastic hematologic changes, but allows the development of neurologic abnormalities, because synthetic folic acid, unlike dietary folate, is reduced directly to tetrahydrofolate and not through the methionine synthase reaction.

In any case, treatment of the primary causes of folate and/or cobalamin deficiencies should concomitantly be pursued.

See also: **Cobalamins**: Physiology; **Folic Acid**: Physiology; **Pregnancy**: Maternal Diet, Vitamins, and Neural Tube Defects

Further Reading

Beck WS (1991) Diagnosis of megaloblastic anemia. *Annual Reviews of Medicine* 42: 311–322.

Carmel R (1999) Ethnic and racial factors in cobalamin metabolism. *Seminars in Hematology* 36: 88–100.

Chanarin I (1990) *The Megaloblastic Anaemias*, 3rd edn. Oxford: Blackwell Scientific Publications.

Lindebaum J and Allen RH (1995) Clinical spectrum and diagnosis of folate deficiency. In: Bailey L (ed.) *Folate in Health and Disease*, pp. 43–74. New York: Marcel Dekker.

Scott JM (1999) Folate and vitamin B_{12}. *Proceedings of the Nutrition Society* 58: 441–448.

Snow CF (1999) Laboratory diagnosis of vitamin B_{12} and folate deficiency. *Archives of Internal Medicine* 159: 1289–1298.

Trugo NMF (1999) Nutritional macrocytic (megaloblastic) anemia. In: Strickland TG (ed.) *Hunter's Tropical Medicine and Emerging Infectious Diseases*, 8th edn. Philadelphia: W.B. Saunders.

Wickramasinghe SN (1999) The wide spectrum and unresolved issues of megaloblastic anemia. *Seminars in Hematology* 36: 3–18.

Wintrobe MN, Lukens JN and Lee CR (1993) The approach to the patient with anemia. In: Lee GR, Bithel TC, Foerster J, Athens JW and Lukens JN (eds) *Wintrobe's Clinical Hematology*, 9th edn, vol. 1, pp. 715–790. Philadelphia: Lea & Febiger.

Zittoun J and Zittoun R (1999) Modern clinical strategies in cobalamin and folate deficiency. *Seminars in Hematology* 36: 35–46.

Other Nutritional Causes

F G Huffman, Florida International University, Miami, FL, USA

S Nath and Z C Shah, Florida International University, University Park, Miami, FL, USA

Background

Anemia is a condition where there is a reduction in the hemoglobin content, size, and/or number of erythrocytes. Nutritional anemias may result from various vitamin and mineral, as well as some macronutrient deficiencies, but the most common are megaloblastic anemia, resulting from folic acid or vitamin B_{12} deficiency, and microcytic, hypochromic anemia, resulting from iron (Fe) deficiency. The fat-soluble vitamins A and E, and water-soluble vitamins B_2 (riboflavin), B_6 (pyridoxine), and C (ascorbic acid) have also been implicated as causes of anemia in humans. In addition, deficiencies of the microminerals copper (Cu) and zinc (Zn), essential to erythropoiesis, can cause anemia. Protein–energy malnutrition (PEM), a multifaceted nutritional condition, is also associated with anemia, but this is often not classified as a true nutritional anemia, since the role of PEM is physiologic rather than pathologic. Essential fatty acid deficiency may also play a role in the development of anemia. These nutritional associations with anemia are summarized in **Table 1**.

The two criteria for defining a nutritional anemia are that: (1) deficiency or lack of a specific nutrient must cause the anemia, and (2) replacement of that specific nutrient must correct the anemia. Anemias resulting from the aforementioned deficiencies of vitamins and minerals may not directly conform to these two criteria, because isolated deficiencies are rare, and an anemia may not be directly related to a single nutrient deficiency, but most often is complicated by other nutrient deficiencies.

Fat-soluble Vitamins

Vitamin A

Vitamin A deficiency, or hypovitaminosis A, is a prevalent nutritional problem in both developing and developed countries, especially among children. The biological interactions between vitamin A deficiency, iron metabolism, hematopoiesis, and the development of anemia are clearly established. However, polycythemia, as well as altered water balance leading to dehydration may mask anemia resulting from vitamin A deficiency.

Morphological changes, such as poikilocytosis and anisocytosis in erythrocytes in peripheral blood, may result from vitamin A deficiency. Subjects with vitamin A deficiency often develop microcytic, hypochromic anemia with elevated storage of Fe in the liver and other reticuloendothelial systems. The later finding suggests that the role of vitamin A in erythropoiesis of mobilizing or releasing Fe form the liver and adequate Fe in diet are not sufficient to correct or prevent anemia. Vitamin A deficiency may also decrease transferrin with subsequent reduction in Fe transport. Experimental studies have demonstrated this in subjects with vitamin A deficiency who respond well to vitamin A supplementation, but are refractory to Fe administered to correct anemia. Results from randomized clinical trials document that vitamin A supplementation alone or in combination with Fe supplementation improves anemia. Vitamin A supplementation alone yields a significant increase in serum retinal and Fe, blood hemoglobin (Hb), hematocrit (Hct), and erythrocyte, and transferrin saturation without affecting the total iron-binding capacity and serum ferritin. However, in contrast to Fe-deficiency anemia, serum ferritin levels remain normal, and vitamin A supplementation does not alter the absorption of Fe in the intestines. This implies that vitamin A does not inhibit the absorption of Fe, and that a high liver

Table 1 Selective nutrients and characteristic features of anemias resulting from deficiencies

Nutrient deficiency	*Peripheral blood*	*Bone marrow*
Vitamin A	Microcytic	Not known
Vitamin E	'Distorted and fragmented' RBCs, anisocytosis, poikilocytosis, polychromatophilia, spherocytosis, Heinz bodies	Erythroid hyperplasia
Vitamin B_2	Normocytic, normochromic, sideroblastic	Erythrocytic hopoplasia, increased myeloid erythroid ratio, pronormoblast
Vitamin B_6	Hypochromic, microcytic, sideroblastic	Erythroid hyperplasia, basophilic normoblast
Vitamin C	Normocytic normochromic	Not known
Copper	Normocytic, microcytic or macrocytic	Cytoplasmic vacuoles in myeloid and erythroid precursors, ringed sideroblast
Zinc	Microcytic, normochromic or macrocytic	Ringed sideroblast, cytoplasmic vacuolization in both myeloid and erythroid precursors
Protein	Normochromic, normocytic, variations in size and shape	Normo- or hypocellular, reduced erythroid/myeloid ratio
Essential fatty acid	Aplastic	Not known

concentration of Fe is not the result of hemolysis secondary to the increased erythrocyte fragility associated with vitamin A deficiency. Rather, experts in this field of research speculate that vitamin A may improve the metabolic utilization of Fe.

Several mechanisms of vitamin A action have been proposed in erythropoiesis, including mobilization of Fe from the tissue stores (reticuloendothelial system) through increased receptor synthesis; an influence on the proliferation or differentiation of red blood cells; and facilitation of Fe absorption by forming a complex with nonheme Fe. It has also been noted that vitamin A deficiency increases one's susceptibility to infection, leading to diminished erythropoiesis.

Vitamin A deficiency may also impair the utilization of other nutrients crucial for erythropoiesis. The relationship between vitamin A and Zn has been documented in view of the role of vitamin A in the synthesis of Zn-dependent retinal-binding protein (RBP), and absorption and transport of Zn. In some instances, the association between RBP and hemoglobin synthesis is explained in children and pregnant women.

Vitamin E

Vitamin E, primarily α-tocopherol, acts as a free-radical scavenger and natural antioxidant to protect cell membranes against free-radical reactions or lipid peroxidation. Normal healthy individuals who consume a nutritionally adequate diet rarely develop vitamin E deficiency, and no overt deficiency symptoms have been described. Table 2 describes some circumstances that may result in a low vitamin E level in humans.

In animals, vitamin E deficiency-associated anemia was induced experimentally, with Dinning and Day in 1957 first describing vitamin E deficiency anemia in rhesus monkeys. The monkeys developed a very low reticulocyte count that was reversed by vitamin E supplementation, with subsequent correction of the anemia. Subsequent prolonged vitamin E deficiency studies in similar animals revealed normochromic and normocytic (occasional macrocytic) anemia in peripheral blood, but bone marrow examination showed hyperplasia of erythroid cells and the presence of nuclear chromatin abnormalities with multinucleated forms. A high rate of red blood cell hemolysis has also been reported. It was presumed that all these changes were the result of an abnormal erythropoiesis, but ineffective erythropoiesis was eliminated as the primary cause of anemia, since the stability of reduced glutathione, osmotic fragility, and serum iron were normal.

Table 2 Causes of vitamin E deficiency

Congenital	i. Cystic fibrosis ii. Inherited hemolytic anemia iii. Biliary cirrhosis iv. Abetalipoproteinemia
Acquired	i. Liver disease ii. Gallbladder disease iii. Pancreatitis iv. Sprue v. Inflammatory bowel disease vi. Extensive small bowel resection vii. Malabsorption viii. Steatorrhea
Premature infants	Poor absorption and metabolism of nutrients resulting from underdeveloped systems
Infants and children with protein–energy malnutrition	Poor intake, transport, metabolism, and storage of nutrients.

The relationship between vitamin E deficiency and anemia in humans is less conclusive, since no clinically and hematologically well-defined anemia has been reported in humans. It has been theorized that vitamin E deficiency may cause hemolytic anemia in humans by affecting the hematopoietic system. It is widely accepted that erythrocytes are susceptible to membrane damage in premature infants with vitamin E deficiency because of high lipid peroxidation. Increased erythrocyte fragility, when exposed to hydrogen peroxide (H_2O_2), is considered to be an index of tissue vitamin E deficiency, as the consistent inverse relationship between serum tocopherol concentration and degree of H_2O_2 hemolysis has been established. For example, it has been shown that a serum tocopherol level $< 0.2\,mg\,dl^{-1}$ correlated with a mean H_2O_2 fragility of about 90%. In 1952, increased hemolysis of red blood cells in vitamin E-deficient premature infants was observed when the cells were exposed to H_2O_2. However, *in-vivo* experiments have not documented any significant damage to red blood cells.

In 1967, Oski and Barness first described vitamin E deficiency anemia in a group of premature infants 6–10 weeks of age. Peripheral blood smears showed 'distorted and fragmented' erythrocyte morphology characterized by anisocytosis, poikilocytosis, polychromatophilia, fragmented red cells, and a few mild spherocytosis and Heinz bodies. Examination of the bone marrow of these premature infants revealed erythroid hyperplasia. A low serum concentration of vitamin E ($< 0.5\,mg\,dl^{-1}$), low hemoglobin level ($7.6 \pm 1\,g\,dl^{-1}$), and increased sensitivity of the erythrocytes to hemolysis when exposed to dilute H_2O_2 solution have also been reported. All these hematological abnormalities were corrected by vitamin E supplementation. Other researchers have also reported vitamin E deficiency anemia among premature infants (body weight $< 1500\,g$) with moderately high reticulocyte counts that responded to vitamin E therapy, suggesting an anemia that is hemolytic in nature.

It has been reported that plasma levels of vitamin E are similar in both premature and mature infants, and a linear relationship has been described between body tocopherol content and body weight during gestation. Some studies have found shortened red blood cell survival in vitamin E deficient premature infants and adults, suggesting the presence of hemolytic anemia. Several other studies have observed that, although premature infants and adults with vitamin E-deficient diets did not develop anemia, they experienced decreased plasma tocopherol levels and increased erythrocyte susceptibility to H_2O_2.

Convincing results from studies involving patients with sickle-cell anemia (SCA) also support the association between vitamin E deficiency and anemia. One of the clinical manifestations of SCA is chronic hemolytic anemia. Among sickle-cell patients who usually experience impaired vitamin E absorption, increased susceptibility of sickle erythrocytes to peroxidation has been demonstrated, partly due to abnormal membrane phospholipid organization and low serum and erythrocyte vitamin E deficiencies. It has been documented that preincubation of sickle erythrocytes with vitamin E may significantly reduce susceptibility to oxidation.

It is also plausible that vitamin E plays a role in erythropoiesis. The potential role of vitamin E in erythropoiesis has been highlighted in studies conducted in Uganda, Jordan, and Thailand on protein–energy-malnourished subjects who developed anemia and responded to vitamin E supplementation. In contrast, subjects in India and West Africa with similar deficiencies did not respond to vitamin E supplementation. Chronic vitamin E deficiency is found among patients with cystic fibrosis, biliary atresia, biliary cirrhosis, sprue, pancreatitis, extensive small bowel resection, inflammatory bowel disease, and other malabsorption syndromes without any evidence of related anemia or ineffective erythropoiesis.

In summary, based on currently available information, there are several plausible mechanisms for hemolytic anemia resulting from vitamin E deficiency, such as increased oxidation of erythrocyte cell membranes with increased erythrocyte susceptibility to fragility. There is also the promotion of free radical generation by iron and hemoglobin in vitamin E deficiency, with free radicals acting as potent catalysts for erythrocyte lipid peroxidation.

Water-soluble Vitamins

Vitamin B_2

The two biologically active coenzyme forms of vitamin B_2 (riboflavin) are flavin mononucleotide and flavin adenine dinucleotide. The latter is required for red blood cell glutathione reductase activity, and so a deficiency of riboflavin may lead to anemia. Anemia secondary to riboflavin deficiency was initially observed among animals as hypoplastic anemia with fatty infiltration of the bone marrow and normocytic hypochromic anemia in riboflavin-deficient dogs and calves, respectively. In humans, anemia resulting from riboflavin deficiency has only been reported among patients receiving medications such as galactoflavin, ouabain, theophylline, penicillin, probenecid, and chlorpromazine, which decrease the absorption or metabolism of riboflavin. This anemia

is characterized in peripheral blood smears by red blood cell morphology described as normocytic and normochromic. However, in the bone marrow, there is marked erythrocytic hypoplasia with increased myeloid erythroid ratio, pronormoblasts with cytoplasmic and nuclear vacuoles, and a marked decrease in normoblasts. Reticulocytopenia is also noted. Riboflavin supplementation results in a prompt decrease in the myeloid erythroid ratio and the disappearance of anemia.

New research has identified two paths by which flavoproteins may be linked to sideroblastic anemia. Flavin monooxygenase is an intracellular ferrireductase involved in the reduction of iron (ferric to ferrous) so that it can be incorporated into heme by ferrochetalase. Decreased flavin monooxygenase can result in accumulation of ferric iron in the mitochondria of erythroid cells, the production of ringed sideroblasts, and impaired hemoglobin synthesis. This may lead to sideroblastic anemia and erythropoietic hemochromatosis. Another flavoprotein, pyridoxine phosphate oxidase, is essential in the formation of pyridoxal 5′-phosphate (PLP) from pyridoxine and pyridoxamine phosphate. The riboflavin link to sideroblastic anemia in this case is therefore indirectly through its role in the activation of vitamin B_6.

Vitamin B_6

PLP, the most active coenzyme form of vitamin B_6 is essential for numerous metabolic reactions, including the initial regulatory step of heme biosynthesis involving δ-aminolevulinic acid, the precursor of porphobilinogen (Figure 1).

The typical features of anemia resulting from pyridoxine deficiency have been described as hypochromic, microcytic, and hyperferremic, with bone marrow showing erythroid hyperplasia and maturation arrest with predominantly basophilic normoblasts. Malnourished patients with hypochromic anemia respond to vitamin B_6 supplementation but are unresponsive to iron therapy. Complete recovery from anemia with normal hemoglobin levels can be achieved by adequate pyridoxine therapy, but not by Fe or other vitamins or minerals.

Some researchers have tried to illustrate that the association between pyridoxine and the development of sideroblastic anemia involves the defective conversion of pyridoxine to PLP. However, there was no evidence of low erythrocyte pyridoxine kinase among patients with sideroblastic anemia or there was no deficiency of vitamin B_6. The relationship between pyridoxine deficiency and sideroblastic anemia therefore warrants further investigation.

Vitamin B_6 deficiency has also been reported among patients with sickle cell anemia, who responded well to PLP supplementation. PLP binds specifically to the amino terminus of the β chain of deoxyhemoglobin, and high levels of PLP may be an active inhibitor of erythrocyte 'sickling' *in vitro*. In addition, PLP may increase the affinity of hemoglobin-S to oxygen and, at high concentrations, inhibits the gelation of deoxyhemoglobin-S and reduces the number of sickled erythrocytes. However, *in-vivo* studies have not proven that vitamin B_6 supplementation increases red blood cell counts or improves hemoglobin status.

Although isolated vitamin B_6 deficiency is rare, some populations such as users of some oral contraceptives, amphetamines, isoniazid, chlorpromazine, and reserpine, as well as alcoholics have developed an anemia that can be easily corrected by therapeutic doses of vitamin B_6.

Vitamin C

Although anemia is commonly observed among patients with the vitamin C deficiency disease scurvy, the direct role of vitamin C or ascorbic acid in hemopoiesis is not well documented. Usual hematological findings in this case are normocytic, normochromic anemia with reticulocytosis, which can be corrected, with an accompanying rise in hemoglobin concentration, by vitamin C supplementation.

Vitamin C deficiency may be involved indirectly in the development of anemia through its relation to folic acid and Fe metabolism. It is well established that vitamin C increases the bioavailability of Fe in foods and facilitates intestinal Fe absorption, particularly nonheme Fe. Vitamin C also enhances the utilization of heme Fe and facilitates ferritin synthesis. It has been documented that vitamin C augments the translation of ferritin mRNA by maintaining the Fe-responsive element-binding protein in its active form. In addition, *in-vitro* experiments have shown that vitamin C prevents degradation and enhances the stability of ferritin. A proposed, but indirect, role of vitamin C in erythropoiesis is that vitamin C facilitates the turnover of Fe in the body.

Vitamin C is also required for the maintenance of folic acid reductase, which converts folic acid to its active tetrahydrofolate form. Failure to synthesize tetrahydrofolate leads to megaloblastic anemia. Occasional megaloblastic cells have been observed among scurvy patients.

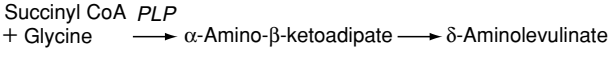

Figure 1 Role of pyridoxal phosphate in the biosynthesis of δ-aminolevulinate.

Minerals Implicated in Anemias

Copper

The essential trace element copper (Cu), is widely distributed in commonly eaten foods, and deficiency of this mineral is extremely rare in people who maintain adequate and recommended caloric intakes. The indispensable role of Cu in humans was first documented during the 1960s among anemic children in Peru, who were refractory to iron (Fe) but responsive to Cu supplementation. Subsequent studies have established that Cu is essential for red and white blood cell maturation and Fe transport.

Cu is a component of several metalloenzymes called the cuproenzymes, with ceruloplasmin (Cp) being the most studied. More than 90% of the body's Cu is found as Cp, and about 60% of the Cp in red blood cells, is bound to superoxide dismutase. Cu-deficiency anemia is thought to result from reduced Cp activity, resulting in impaired iron mobilization. The other Cp-associated enzymes whose impaired function may result in anemia include ferroxidase (which converts Fe^{2+} to Fe^{3+}), ferrochelatase and cytochrome aa_3. Ferroxidase is vital in the incorporation of Fe into circulating transferrin, so that decreased Cp results in poor ferroxidase activity, and Fe, being trapped in the reticuloendothelial system, is unavailable for erythropoiesis. Decreased ferrochelatase and cytochrome aa_3 activity are associated with poor heme synthesis owing to decreased reduction of Fe^{2+} in the mitochondria and poor incorporation of iron into protoporphyrin IX.

Cu deficiency may be acquired easily as a result of an imbalance between body requirements and dietary Cu supply as well as a high zinc intake. In 1956, Sturgeon and Brubaker first described Cu-deficiency anemia in infants. Under normal conditions, liver stores provide adequate amounts of Cu to comply with the infant's need during the first 5 months of life. Acquired Cu deficiency in premature infants is well documented because of a small liver size, low Cu deposits in the liver, and increased need due to the high growth rate. Healthy infants who are fed exclusively cow's milk are prone to Cu deficiency because of the low Cu content and poor bioavailability of Cu from cow's milk. Individuals who receive total parenteral nutrition may also develop Cu deficiency. Table 3 presents several causes of acquired Cu deficiency in humans.

The most common clinical manifestations of acquired Cu deficiency are anemia and pancytopenia observed in both animals and humans. A wide range of morphologic changes in peripheral blood smears and bone marrow biopsies have been reported.

Table 3 Etiology of acquired copper deficiency in humans

Decreased storage in liver	Premature infants
Decreased supply	i. Total parenteral nutrition
	ii. Infants fed cow's milk only
Increased requirements	High growth rate
	i. Celiac disease
	ii. Tropical and nontropical sprue
	iii. Cystic fibrosis
	iv. Short bowel syndrome
	v. Intestinal resection
Decreased absorption	vi. Diarrhea
	vii. Abnormal bile loss
	viii. Intestinal fistula
	ix. High zinc intakes
	x. High iron intakes
	xi. High vitamin C intakes
	xii. Fructose and other refined sugars
Interactions with drugs	i. Penicillamin
	ii. High alkali therapy

The red blood cells can be normo- or macrocytic, dimorphic, and hypochromic. A low reticulocyte count and bone marrow iron content, as well as hypoferremia may exist. Morphological findings in bone marrow cytology include megaloblastic changes with two consistent findings – cytoplasmic vacuoles in the myeloid and erythroid precursors, and the maturation arrest of the myeloid precursors with the appearance of ringed sideroblasts. However, patients with Cu deficiency may not develop megaloblastic changes or there may be the absence of sideroblasts in bone marrow. These changes are easily reversed by Cu supplementation, but refractory to Fe therapy.

Three mechanisms have been proposed for these clinical/morphological manifestations, including reduced Cp activity leading to defective Fe mobilization/transporation. First, even in the presence of normal Fe stores, Fe transfer from macrophages and hepatocytes to plasma is impaired, resulting in hypoferremia. Second, Fe cannot be incorporated into heme and accumulates in the cytoplasm, producing sideroblasts. Third, there is shortened erythrocyte survival as a result of membrane defects secondary to decreased superoxide dismutase activity.

Zinc

The trace element zinc (Zn) functions in several biochemical reactions by way of Zn metalloenzymes. Although several pathological conditions have been related to Zn deficiency in humans, at present, there is no evidence that isolated Zn deficiency causes anemia. In iron depletion, Zn protoporphyrin (ZPP) is formed, and although it is able to bind to heme sites

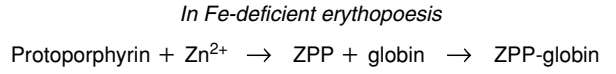

Figure 2 Zinc complexing with protoporphyrin and globin in iron deficiency.

on globin, the resulting molecule is nonfunctional. However, there is evidence that the ZPP–globin complex (Figure 2) accumulates in developing erythrocytes where it inhibits the action of heme oxygenase (involved in heme catabolism), thereby conserving heme.

The involvement of Zn in anemia may be indirectly through its role in Cu metabolism. Several studies have documented that, in the absence of evidence of hemolysis or other nutritional deficiencies, excessive Zn intakes (> 660 mg) may induce Cu deficiency by sequestering Cu in the enterocytes, thus limiting its absorption and increasing fecal losses. Since Cu is essential for mitochondrial Fe transport and heme synthesis, Zn-induced Cu deficiency produces the same anemia as isolated Cu deficiency.

Protein Deficiency and Protein–Energy Malnutrition

Protein is essential for optimal production of hemoglobin and red blood cells. Animal studies have provided evidence of a causal relationship between protein deficiency and the development of anemia. Rats fed a protein-deficient diet show a marked reduction in erythropoiesis and subsequently develop severe anemia, which can be reversed by restoring protein to the diet. In humans, this condition is associated with kwashiorkor. It is thought that protein deficiency causes reduced body mass and a decreased oxygen consumption or requirement, which then leads to reduced erythropoeisis and a fall in red cell mass. Protein deficiency also interferes with the maturation of erythroblasts and decreases the erythropoietin-sensitive pool. The resulting anemia manifests as normocytic, macrocytic, megaloblastic, hypoplastic, normochromic or hypochromic anemia, accompanied by higher-than-normal reticulocyte counts, and a normal to low mean corpuscular hemoglobin concentration. Low serum total protein, albumin fraction, serum iron, and total iron-binding capacity are also observed.

Anemia, as a result of PEM, is poorly understood, and the underlying mechanisms remain obscure. It is thought that this anemia is an adaptation to a lowered oxygen requirement in PEM. Assessing anemia in PEM or kwashiorkor is extremely difficult because of ethical concerns and multiple associations such as infections and deficiencies of iron, folic acid, vitamins B_2, B_{12}, and E, and selenium. In PEM, there is decreased red cell survival, whereas erythropoetin levels are adequate or elevated, and bone marrow iron supplies are normal. Selenium deficiency is thought to significantly influence red cell survival through its relationship with impaired glutathione peroxidase activity. In addition, PEM-related anemia presents with normocytic, normochromic RBCs, low serum Fe, serum iron-binding capacity, serum total protein, albumin, and Fe absorption, but increased γ-globulin. There is evidence that protein supplementation alone in PEM does not correct the related anemia but may exacerbate it. Even when serum vitamin E levels are normal, concurrent protein and vitamin E supplementation appears to significantly increase hemoglobin levels.

Essential Fatty Acid Deficiency

Anemia has also been associated with essential fatty acid (EFA) deficiency. An EFA intake of < 2–4% of total calories is reflected in red blood cells by decreased linoleic acid levels with corresponding increases in eicosatrienoic acid levels. This results in altered triene/tetraene ratios (> 0.4 being considered EFA deficiency). In the plasma fatty acid profile, linoleic acid is generally elongated to arachidonic acid, but in a deficiency, linoleic (tetraene) is replaced by oleic (triene) acid. The RBC fatty acid profile reflects that of the plasma because RBC fatty acids are continually diacylated and reacylated with those in the plasma. Aplastic anemia and thrombocytopenia are the result of altered fatty acid profiles seen in EFA deficiency.

See also: **Antioxidants**: Natural Antioxidants; **Dietary Reference Values**; **Dietary Requirements of Adults**; **Fats**: Requirements; **Malnutrition**: Malnutrition in Developing Countries; **Minerals – Dietary Importance**; **Protein**: Deficiency; **Vitamins**: Overview

Further Reading

Dugdale M (2001) Anemia. *Obstetrics and Gynecology Clinics of North America* 28(2): 363–381.

Fishman SM, Christian P and West KP (2000) The role of vitamins in the prevention and control of anaemia. *Public Health and Nutrition* 3(2): 125–150.

Fussel S and Zeisel SH (2000) *Nutritional Anemias*. Chapel Hill, NC: University of North Carolina.

Gallagher PG and Ehrenkranz RA (1995) Nutritional anemias in infancy. *Clinics in Perinatology* 22(3): 671–692.

Mitrache C, Passweg JR, Libura J *et al.* (2001) Anemia: An indicator for malnutrition in the elderly. *Annals of Hematology* 80(5): 295–298.

Oski FA (1979) Nutritional anemias. *Seminars in Perinatology* 3(4): 381–395.

Oski FA (1995) Anemia due to other nutritional deficiencies. In: Beutler E, Lichtman MA, Coller BS and Kipps TJ (eds) *Williams Hematology*. New York: McGraw-Hill.

Ramakrishnan U (ed) (2001) *The CRC Series in Modern Nutrition: Nutritional Anemias*. Boca Raton, FL: CRC Press.

Rodriguez RM, Corwin HL, Gettinger A *et al*. (2001) Nutritional deficiencies and blunted erythropoetin response as causes of the critical illness. *Journal of Critical Care* 16(1): 36–41.

Silverberg DS, Iania A, Wexler D and Blum M (2001) The pathological consequences of anaemia. *Clinical and Laboratory Haematology* 23(10): 1–6.

Wolfe LC and Lux SE (1979) Nutritional anemias of childhood. *Pediatric Annals* 8(7): 435–443.

World Health Organization (1997) *Malnutrition and the Burden of Disease: The Global Epidemiology of Protein–Energy Malnutrition, Anemias and Vitamin Deficiencies. Global Burden of Disease and Injury Series, Vol. 8*. Cambridge, MA: Harvard University Press.

Animal Fats *See* **Fats**: Production of Animal Fats; Uses in the Food Industry; Digestion, Absorption, and Transport; Requirements; Fat Replacers; Classification; Occurrence

ANIMAL MODELS OF HUMAN NUTRITION

J R Hunt and Z K Roughead, United States Department of Agriculture, Grand Forks, ND, USA

Background

Experimentation with animals has been fundamental to human nutrition knowledge. By using small animals, researchers can economically control experimental diets to determine the biological effects of dietary constituents. Studies with small animals fed controlled diets led to the discovery of most of the vitamins and essential trace minerals. Because of animal research, nutrient-deficiency diseases of humans have been eradicated in many countries. Animal research in the future should help us understand how diet affects disease, reproduction, brain function, and longevity.

Advantages of using animals in nutrition research include: (1) large groups can be studied economically; (2) dietary consumption and living conditions can be carefully controlled; (3) short life-spans improve the feasibility and efficiency of developmental, longevity, or multiple generation studies; and (4) controlled breeding reduces biological variability and can allow the testing of specific peculiarities in metabolism. If dietary deficiencies or imbalances are suspected of causing disease in humans, these hypotheses should be tested in controlled studies with animals. Likewise, new dietary additives should be tested in animals for human-safety purposes. Although animal research is indispensable for our expanding knowledge about diet and health, investigators must weigh the benefits of such knowledge against the discomfort of animals in research. New knowledge for the health of our own and possibly other species should be obtained with an ethical responsibility for the humane treatment of experimental animals.

History

Since the earliest times, humans have recognized that they share with animals a basic need for food and water. Accordingly, the history of nutrition is replete with examples of animal experimentation. Erasistratus (310–250 BC), tested his ideas that *pneuma* in the atmosphere became spirit in the body. He placed fowls in a jar and weighed them and their excreta before and after food. Galen (AD 130–200) studied digestion in hogs and concluded that, in the stomach, foods were reduced to smaller particles that could later be absorbed.

In 1774, Joseph Priestley (1733–1804) exposed mercuric oxide to focused sunlight to produce 'pure dephlogisticated air' (oxygen). He demonstrated that this air caused a candle flame to burn brighter and that mice could live in it.

Guinea-pigs were employed by Crawford, Lavoisier and Laplace, and Despretz in the late eighteenth and early nineteenth centuries to study respiration and body-heat production. Lavoisier and Laplace invented an ice calorimeter and compared the amount of ice melted by a guinea-pig with the amount of carbon dioxide produced. In addition to using guinea-pigs, Lavoisier measured his own respiratory exchanges and those of his friends and assistants. These studies led to the conclusion that respiration was a combustion of carbon and hydrogen which transformed oxygen into carbon dioxide, similar to the burning of a candle or an oil lamp.

Spallanzani (1729–1799) first obtained gastric juice from the stomach of a hawk by using a sponge and string. He demonstrated that it dissolved flesh, bones, and bread without putrefaction.

Boussingault performed the first balance study in 1839 on a cow, by measuring carbon, hydrogen, oxygen, nitrogen, and salts in the feed, urine, feces, and milk. Later, he performed similar experiments with a horse and a turtle dove.

In 1849, Regnault and Reiset used rabbits, dogs, and fowl to determine the effect of various foods on the ratio of carbon dioxide produced to oxygen inspired, later called the 'respiratory quotient.' Dogs were used extensively in the last half of the nineteenth century by German investigators studying protein and energy metabolism. Voit (1831–1908) demonstrated 'nitrogen equilibrium,' and an increase from negative to positive nitrogen balance as dogs were fed greater amounts of meat. He also discovered that muscular work by the dog did not increase protein metabolism. Rubner (1854–1932), a student of Voit, employed dogs to compare energy utilization by direct and indirect calorimetry, and to measure the 'specific dynamic action' (increased heat production) associated with protein consumption. From his measurements of metabolic rates of various species, including humans, Rubner demonstrated that the basal metabolism is proportional to the surface area of the animal.

Animals played an indispensable role in the discovery of the vitamins. In 1897, Eijkman, a Dutch physician in Java, reported that chickens fed polished rice developed a paralysis. This paralysis was similar to that of the disease beriberi in humans. Unmilled rice or a water extract of rice polishings cured the chickens. Hšlst, a Norwegian, repeated Eijkman's experiments, and obtained similar results with pigeons. However, he and Frölich reported in 1907 that guinea-pigs fed a cereal diet developed scurvy, rather than the paralysis of beriberi. This fortunate discovery – of one of the few species in addition to humans that require ascorbic acid – soon led to a biological assay for the antiscorbutic potency of foodstuffs.

In 1906, Hopkins reported from the University of Cambridge, that mice required the amino acid, tryptophan. This initiated the concept of essential amino acids and differences in the biological value of proteins. Hopkins studied both mice and rats and reported in 1912 that animals could not live on purified protein, fat, carbohydrates, and minerals alone; other unknown nutrients were essential.

In 1907, at the University of Wisconsin, McCollum set out to identify what was lacking in purified diets that did not nourish small animals. Frustrated by tedious studies with cows fed various grains, he recognized the advantages of studying animals that have a short lifespan and eat little. McCollum decided to use rats because of their convenient size, omnivorous feeding habits, and lack of economic value. Within 5 years, McCollum and Davis reported a fat-soluble factor (later named vitamin A) in butterfat that was not present in lard or olive oil. Vitamin A deficiency was later found to be a major cause of human blindness.

In the UK, Mellanby used dogs, known to be susceptible to rickets, to show that the disease resulted from a dietary deficiency. Incidental to this research, Mellanby discovered that nitrogen trichloride bleaching of wheat flour caused canine hysteria, a nervous disorder that also occurred in ferrets, cats, and rabbits. Rats that had been used in safety testing were less sensitive to the compound. Commercial use of nitrogen trichloride in flour for human consumption was discontinued after Mellanby's discovery.

McCollum used rats to differentiate between the effects of vitamins A and D. He developed the 'line test,' a measure of the width of a line of new calcification in the bone of a deficient rat given vitamin D.

This became a widespread biological assay for vitamin D in foods.

Rats were the principal animal used to discover most of the vitamins, the essential trace elements, and the essential amino acids. As a result, more is known about the nutritional requirements of the rat than about any other species. Dogs, chickens, pigeons, mice, and guinea-pigs have also made valuable contributions to our knowledge of nutrition.

Animals as Models of Human Nutrition in Research

Factors considered when selecting an animal model include anatomical, physiological, and biochemical similarities to humans, similarities of disease processes, susceptibility to etiological agents, availability, cost, breeding lines available, behavior, size, genetic characteristics, lifespan, resistance to infections, suitability for experimental manipulation, and accumulated pathobiological information. Considerations particularly pertinent to nutrition research are as follows.

Natural Diet in Habitat

The omnivorous feeding patterns of the rat usually make it a better model for human-nutrition questions than a strict herbivore such as the rabbit. Although nonhuman primates may be most genetically related to humans, primates include insectivores, herbivores, and omnivores. Researchers must evaluate each primate species separately. Strict carnivores such as cats have associated differences in nutrient metabolism and requirements. For instance, the cat requires arachidonic acid and preformed vitamin A from animal foods, as it cannot synthesize these from the respective plant food precursors, linoleic acid and β-carotene.

Nutrient Requirements

The requirement for vitamin C is a striking example of qualitative differences between species; few species require it in the diet. Humans share this uncommon requirement with other primates, the guinea-pig, the red-vented bulbul, and some Indian fruit-eating bats.

Dietary fiber may provide a more subtle example of species differences in nutrient requirements. Two herbivores, the guinea-pig and the rabbit, require dietary fiber, according to the US National Research Council. Other laboratory animals are apparently not as vulnerable to a fiber deficit. However, fiber is routinely added to research diets for rodents, and continued research may reveal reduced gastrointestinal diseases associated with fiber in the diets of primates.

Quantitative comparisons of nutrient requirements across species are difficult: limited research on some species has yielded knowledge of adequate intakes without determination of minimum requirements. Vitamin requirements between species are influenced by differences in production by intestinal flora. The mineral requirements of most species are quite similar.

Digestion and Absorption

The relative influence of gastrointestinal microorganisms on metabolism of digesta varies widely between species. Extensive fermentation in the pregastric chambers of ruminants and the hindgut of the horse makes these animals poor models for human nutrition. Dogs and pigs are similar to humans in many aspects of gastrointestinal morphology and physiology, including the relative length of small and large bowels, ingesta transit times, influence of dietary factors on gastric emptying, glucose or xylose absorption, fecal fat excretion, activities of intestinal brush border and pancreatic enzymes at maturity, and colonic volatile fatty acid concentration. However, humans, dogs and pigs differ in the developmental patterns and primary structure of some digestive enzymes. The composition of bile and the structure of bile acids differs substantially between humans and pigs. Colonic volatile fatty acids may be an important energy source for the pig, but not for the dog; the importance of colonic fermentation for humans probably falls between these two species. Numerous strains of miniature pigs have been developed and used in experiments related to digestion and absorption.

Body Composition, Nutrient Metabolism, and Excretion

Species differences in chemical composition may influence the suitability of animals as models for human-nutrition problems. As an example, differences in relative skeletal mass may reduce the usefulness of animal models for bone health and related nutrients. The skeleton constitutes about 16% of the adult human body, as compared to 11% of the rat, 10% of the rabbit, and 7.7% of a 28-week-old pig. This probably accounts for considerably higher amounts of calcium and phosphorus per kilogram of fat-free body tissue in humans than in pigs, cats, rabbits, rats, or mice. Human cortical bones have greater nitrogen and lower calcium:nitrogen ratios than those of pigs, cats, rabbits, rats, and fowl. Differences in nutrient excretion may complicate animal modeling further. For example, the rat excretes less than 1% of dietary calcium in the urine, which is an important route for calcium excretion (and its

hormonal control) in humans. These examples may help explain why there has not been a satisfactory animal model for human osteoporosis.

Coprophagy (the ingestion of feces) by rabbits and rodents can confound dietary intake measurements in nutrition studies. By increasing the total nutrient intake, coprophagy can lower apparent requirements and reduce signs of deficiency for many nutrients. Particularly affected are those nutrients synthesized or made more bioavailable through colonic microbial flora, including the B-complex vitamins and essential fatty acids. To study signs of deficiency for such nutrients, steps must be taken to prevent coprophagy.

Reproduction

Animal studies have demonstrated vital roles for nutrition from estrus to parturition and lactation. For example, animal studies have shown the essentiality of zinc for normal estrus cycling and fertility, development of preimplantation eggs, normal development of organ systems (especially the skeletal and central nervous systems), fetal growth, and normal labor and delivery. Zinc deficiency during specific stages of gestation causes postnatal abnormalities in behavior and immune function.

Special considerations when evaluating animals as models of human reproduction include the number of fetuses, the type of placentation, the tendency to abort or resorb fetuses under teratological conditions, the rate of development, and the maturity of the fetus at delivery. Because members of a litter have a similar genetic inheritance and intrauterine environment, these similarities must be considered to plan experiments with an appropriate number of animals and distribution of experimental treatments. Maturity at birth can determine the suitability of an animal model for studies of late gestational development. For example, epiphyseal calcification of the femur is much greater in newborn piglets than in newborn humans, and pigs walk soon after birth. This would limit the usefulness of the pig in some studies of prenatal bone development. Maturation at birth is also a concern when selecting animals to study early infant development. Another difficulty is that newborn mammals often require maternal lactation, limiting the investigator's control of dietary variables in early life.

Animal Models of Nutrient Evaluation

Historically, animals were useful for food evaluation, such as vitamin assays, because chemical techniques were inadequate for direct analysis of foods. While chemical-analysis difficulties have largely been overcome, difficulties predicting nutrient bioavailability have not.

Bioavailability is the fraction of a nutrient in food that is absorbed and utilized. It is affected by chemical form, interactions with other food components, and, probably, physiological responses to food. For example, the zinc bioavailability of foods fed to rats depends on the amount of food fed, even though the foods are compared in amounts providing similar quantities of zinc. Important *in vivo* variables may be pancreatic secretions in response to different foods, competition between endogenous (pancreatic carboxypeptidase) and dietary sources of zinc for absorption, and coabsorption with ligands such as histidine that have separate facilitated pathways of absorption. *In vitro* estimates of zinc absorption are limited by our incomplete understanding of the dynamic gastrointestinal response to foods.

For investigating the bioavailability of some nutrients, even animal models may not be sufficient. For example, the heme form of iron present in foods of animal origin is much better absorbed by humans than ferrous iron; this is not true for rats.

The nutritional quality of dietary proteins has been evaluated since as early as 1917 by measuring animal growth or nitrogen retention. Despite attempts since the mid-1940s to replace the biological assays with a 'chemical score' of amino acid composition, the biological assays continue to be used extensively. Although chemical scoring accounts for essential amino acid composition and requirements, scoring does not account for differences in amino acid bioavailability, which can be reduced by chemical changes during heating, or by inhibitors of proteolytic digestion.

In vitro testing has progressed substantially and is often preferred by scientists for efficient use of resources. However, the complexity of living organisms remains incompletely understood, and frequently, the nutritional value of foods or diets is more accurately evaluated by direct *in vivo* testing.

Genetically Altered Animal Models

Inheritance affects individual responses to nutritional stress. Genetic factors may determine the absorption and/or utilization of a given nutrient. Also, the amount of the nutrient in the diet may influence gene expression. Genetic engineering technology has been used to study the effects of overexpression of a specific gene on nutrient metabolism. More recently, gene-ablation technology, commonly known as gene knockout, has been used (primarily in mice) to study the effects of inactivating a specific gene. Mice with alterations to specific genes are produced by targeting the specific gene in the embryonic stem cells. Knockout mice provide an opportunity to

analyze the function and physiological effects of specific mammalian genes, to model a range of human inherited disorders (e.g., low-density-lipoprotein receptor knockout mice for the study of familial hypercholesterolemia and vitamin D-receptor knockout mice for the study of dietary calcium and bone health) and to explore potential intervention strategies for disease treatment. Cross-breeding of knockout mice to strains with other mutations can be useful in understanding biochemical controls and interactions. For instance, a mutation in the HFE protein is associated with hemochromatosis, a disorder of excessive iron absorption and accumulation. Similar iron accumulation occurs in HFE knockout mice, and this effect can be exacerbated or reduced by cross-breeding these mice with other knockout mice missing genes for other proteins involved with iron absorption and metabolism. Such mice can be further studied to assess the interaction of the genetic mutation with different forms and concentrations of dietary iron. Newer methods can introduce more subtle gene alterations, including specific point mutations. Animal research is making an important contribution to improved protocols for gene therapy, which should help in the alleviation of many inherited human diseases.

Animal Models of Disease States

With the help of animal models, a major cause of human disease – overt nutrient deficiency – was largely solved as a scientific problem (it unfortunately remains as a political and economic problem). Scientific emphasis has turned to diet and chronic disease. Perhaps because they are as yet unresolved, the problems of chronic disease seem more complex. Dietary variables are likely to be slight deficiencies, excesses, or imbalances in a lifelong interaction with other genetic and environmental factors. Animal models can provide the advantage of controlled environmental and dietary factors over a short lifetime. While genetic differences can limit application to humans, animal models of disease are improving through genetic technology. The use of animals to study nutrition and major chronic diseases is considered briefly below.

Atherosclerotic Cardiovascular Disease

The responsiveness of serum cholesterol and atherosclerosis to dietary cholesterol and saturated fatty acids varies among animal species. Rabbits, guinea-pigs, swine, and rhesus and cynomolgus monkeys are more susceptible than humans. Rats and dogs are more resistant. Baboons and vervet monkeys are in the same moderate range of susceptibility as humans. Genetic differences in lipid metabolism are confirmed by the variation observed within and between species.

Hypertension

Hypertension in response to dietary salt is not observed in all animal models but occurs in specific animal strains without sufficient renal capacity to excrete salt rapidly. The Dahl S and the Kyoto spontaneously hypertensive rat are susceptible to dietary salt, and the addition of potassium to high-salt diets reduces the incidence of stroke in these animals, and in the Sprague–Dawley rat. Research on animal models verifies an interaction between diet and genetic susceptibility to hypertension, and susceptible animals may be useful in identifying dietary variables that affect susceptible humans.

Obesity

Genetically obese strains of rodents – obese (*ob/ob*) and diabetic (*db/db*) mice and obese (*fa/fa*) rats (Figure 1) – confirm that inheritance can contribute to obesity. The obese strains are either resistant to leptin (db/db mice and fa/fa Zucker rats) or deficient in leptin (ob/ob mice), a hormone produced by adipocytes. This in turn results in overfeeding and increased concentrations of Neuropeptide Y in plasma and in the hypothalamus. Obesity in such models has been associated with hyperinsulinemia, abnormal glucose tolerance, and high body-fat composition even with normal diets. Rats without genetic obesity increase their food intake and thermogenesis in response to a 'cafeteria' diet of mixed human food. The relationship between energy intake, physical activity, thermogenesis, and body weight has been extensively studied in rodents. Results of these studies generally support the biological maintenance of a 'set-point' for body weight; that varies only slightly and controls energy balance.

Cancer

Approximately one-third of human cancer mortality may be related to diet. Animal studies, mostly using rats, have indicated that tumor development is enhanced in a variety of tissues by high-fat diets, especially diets rich in ω-6 (but not ω-3) polyunsaturated fatty acids. This effect of high-fat diets has not been completely differentiated from the effect of high energy consumption. Animal models have shown that selenium inhibits both the initiation and proliferative phases of tumorigenesis. The effect of some dietary components may depend on the specific site, type, and stage of carcinogenesis as well as the relative nutrient requirements of host and tumor. For instance, zinc deficiency in animals increases

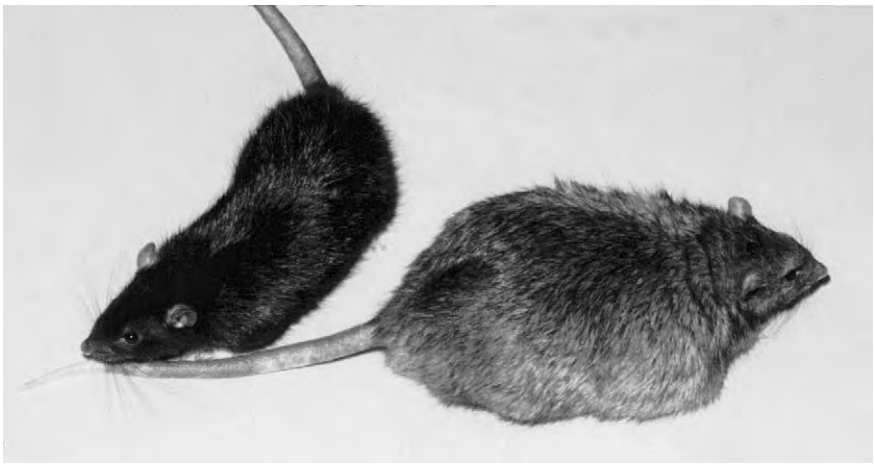

Figure 1 Control and *fa/fa* Zucker rat, a genetic strain used to investigate obesity. From Judith S. Stern, Sc.D., Professor, University of California at Davis, USA with permission.

the incidence of esophageal carcinoma induced by methylbenzylnitrosamine but decreases the incidence of tumors induced by 3-methylcholanthrene and 4-nitroquinoline-*N*-oxide.

Animal models are commonly used to test food additives for carcinogenicity. Careful studies generally involve more than one species and exposure over an extended portion of the lifespan. A US food additive law, known as the Delaney Clause, which prohibits the use of a food additive in any amount if it has been shown to produce cancer in animal studies, has been controversial because it prohibits any amount of a substance that may have been tested only in high amounts. The evaluation of research results must consider the average and peak human exposures to the substance, and the potency of the substance to determine whether dietary characteristics have been experimentally exaggerated, and whether extrapolation to humans is realistic.

Osteoporosis

Diets low in calcium or high in phosphorus increase bone reabsorption and decrease bone mass in rats, mice, cats, dogs, and nonhuman primates. In contrast with experimental animals, calcium balance in humans is relatively insensitive to high dietary phosphorus. There are few animals that experience the spontaneous ovarian failure in middle age that occurs in humans. Researchers have not identified any appropriate animal model of postmenopausal or age-related osteoporosis (see the above section Body Composition, Nutrient Metabolism, and Excretion).

Diabetes Mellitus

Experimental diabetes can be induced in animals by pancreatectomy or injection of β-cell toxins such as streptozotocin or alloxan. There is a genetically obese strain of diabetic (*db/db*) mice. The *Psammomys obesus* (sand rat) strain of rat remains lean and nondiabetic in the wild but becomes obese and diabetic under laboratory feeding conditions. In normal strains of animals, a greatly increased food intake for an extended time leads to adiposity and an increased incidence of insulin resistance, which is reversible with weight reduction. It has not been possible with animal research to show that a high energy consumption causes diabetes.

Interpretation and Extrapolation of Data

Animal studies are most appropriately applied to humans when they complement, and are consistent with, human studies. Relevance of animal research to humans is more likely if the results are confirmed in several animal species.

The highly uniform experimental conditions characteristic of animal models can provide disadvantages as well as advantages. For example, energy restriction reproducibly increases longevity in rodents. However, such studies are usually conducted under laboratory conditions in which the rodents spend their entire lives in a small space, have no social relationships, do not reproduce, have low exposure to hazardous biological or chemical substances, and have little need for physical activity or development of self-preservation skills. Scientists must evaluate whether the food-restricted animals would have an improved *quality*, or even *quantity*, of life if tested under more realistic living conditions.

The control of experimental diets in animal nutrition studies can present similar problems in extrapolation to humans. To enhance experimental control

and comparability between studies, animal diets for nutrition research are often standardized, purified, and fed *ad libitum*. Purified diets composed of a small number of refined ingredients, such as refined proteins, carbohydrates, and fat, with added vitamin and mineral mixtures, reproducibly control specific nutritional variables but also limit the context of the experiment. Unlike many animal research diets, human diets contain thousands of compounds and are usually scheduled in meals. As nutrition research moves from single nutrient deficiencies to more complex dietary interactions that affect health, laboratory diets and other conditions may need to become more like human conditions to facilitate comparability to humans. An example is the 'cafeteria' diet of mixed human food used to attain overeating in rodents.

Limitations of Animal Models in Nutrition Research

Nutritional knowledge has rapidly expanded in the nineteenth and twentieth centuries. During this time, an increasing share of biomedical advances depended on animal research. Two-thirds to three-quarters of major biomedical advances in the 1900s required the use of animals. This trend was accompanied by an antivivisectionist movement in Victorian England in the late 1800s, which may be a predecessor of the current 'animal rights/liberation' movement. High-quality animal care improves experimental reliability and engenders public trust. Scientists have an ethical obligation to ensure animal well-being and minimize pain and suffering. Scientists must be involved in public policy influencing legal regulations that benefit humans through science while treating animals humanely and without inefficient bureaucracy.

Scientists should consider alternative methods, including the use of more socially acceptable species, such as the experimental use of rodents rather than dogs, cats, or monkeys, and the possible use of non-mammalian vertebrates, invertebrates, and micro-organisms. However, depending on the research problem, such alternatives may be unacceptable in answering questions applicable to humans. Nutrition knowledge may also be gained by using cell and tissue cultures, human tissues removed at surgery or at autopsy, and *in vitro* systems and mathematical models. Such methods can be more efficient, controlled, and economical than research with living animals. But again, researchers must be cautious not to extrapolate to human nutrition without adequate validation of the simpler model.

A straightforward alternative to using animals is to study humans directly. Although the technology to study humans safely continues to improve, many nutrition research topics require animal models. Such topics include the effects of nutrition on reproduction, growth and development, longevity, and behavior. Animal experiments can verify observations in humans that often cannot be controlled sufficiently to be conclusive. The use of animals for initial experiments in nutrition research improves efficiency and progress. Animal studies facilitate the investigation of physiological and molecular mechanisms. Animal research has been instrumental in identifying required nutrients and developing the science of nutrition. Animal research continues to answer questions about nutrition and health that can be learned only from living organisms.

See also: **Atherosclerosis**; **Cancer**: Epidemiology; **Coronary Heart Disease**: Etiology and Risk Factor; **Diabetes Mellitus**: Etiology; **Hypertension**: Physiology; **Obesity**: Etiology and Diagnosis; **Osteoporosis**; **Zinc**: Properties and Determination; Physiology

Further Reading

Hui DY (1998) Utility and importance of gene knockout animals for nutritional and metabolic research. *Journal of Nutrition* 128: 2052–2057.

Lusk G (1933) *Nutrition*. New York: Paul B Hoeber.

National Research Council, Committee on Diet and Health (1989) *Diet and Health. Implications for Reducing Chronic Disease Risk*. Washington, DC: National Academy Press.

National Research Council, Committee to Revise the Guide for Care and Use of Laboratory Animals (1996) *Guide for the Care and Use of Laboratory Animals*. Washington, DC: National Academy Press.

Phillips RW (ed.) (1984) *Animal Models for Nutrition Research. Report of the Fifth Ross Conference on Medical Research, Ross Laboratories, Columbus, Ohio*.

Widdowson EM (1986) Animals in the service of human nutrition. *Nutrition Reviews* 44: 221–227.

ANNONACEOUS FRUITS

A P George and R J Nissen, Maroochy Horticultural Research Station, Nambour, Queensland, Australia

Introduction

The main commercial species of *Annona* grown throughout the world are the cherimoya (*A. cherimola*), the atemoya (*A. hybrids*), and the sugar apple (*A. squamosa*). Expansion of production of these three species in many subtropical countries is limited by several factors, including low yields of better-quality cultivars, unattractive external appearance of the fruit, a high susceptibility of the fruit to blemishing, poor and unreliable internal fruit quality, and a very short postharvest shelf- and storage life. In the short term, fruit quality can be improved by field management practices. In the longer term, selection and breeding of new varieties will be necessary.

Classification

The family, Annonaceae, comprises 50 genera. Three genera (*Annona, Rollinia*, and *Asimina*) produce edible fruit, but only two genera are of commercial importance – *Annona*, comprising approximately 100 species, and *Rollinia*, comprising approximately 50 species. The four most commercially important species are the cherimoya, the sugar apple, the atemoya, and the soursop (*A. muricata*). The cherimoya is native to the subtropical highlands of Peru and Ecuador, and is grown commercially in Chile, Spain, California, and New Zealand. The atemoyas, most of which are hybrids between *A. cherimola* and *A. squamosa*, are grown commercially in Florida and Australia. Both the sugar apple and the soursop are widely distributed throughout the tropical regions of South-east Asia and Central America, where they are found growing wild. The rollinia (*Rollinia mucosa*) and biriba (*R. deliciosa*) are native to the jungles of South America. A list of the edible species of *Annona* and *Rollinia* grown throughout the world is presented in Table 1. It should be noted that common names, such as custard apple, have been used with reference to many different species.

Climate

Most *Annona* species are tropical or subtropical in their growth requirements. Both the atemoya and the cherimoya are best grown in frost-free locations as the young trees are killed at −1 °C and mature trees at −3 °C. The cherimoya is more cold-tolerant than the atemoya and is able to stand a longer duration of cold at −3 °C. The sugar apple and the soursop, on the other hand, are very frost-sensitive. Most cherimoya and atemoya cultivars are semideciduous and enter either a true rest or environmentally induced dormancy in the late winter or spring. This dormancy or rest period allows the plant to avoid the effects of late-winter frost and drought.

Different species exhibit varying climatic requirements for fruit development and maturation. Excessively low temperature during the fruit maturation period may retard the process, while excessively high temperatures during this period may cause premature ripening and fermentation of the fruit whilst still on the tree. For example, atemoya cultivars grown in cool subtropical regions of California or New Zealand may fail to mature properly. In these regions the main commercial species grown is the cherimoya. In contrast, cherimoya fruit grown in tropical climates often fails to develop full flavor, and postharvest shelf-life is short.

Besides the influence of temperature on fruit maturation, physiological disorders such as skin russetting are more prevalent under cool night temperatures, particularly when ambient temperatures fall below about 13 °C.

Fruit Morphology

The cherimoya, atemoya, and sugar apple fruits are syncarpiums, made up of many individual carpels which are fused together with the receptacle to form a fleshy mass. Each carpel, if pollinated, contains an oval seed. Depending on species and cultivar the number of seeds per 100 g of flesh can vary from five to 15. The fruit ranges in shape from conical, through globular to ovoid, and the skin type may be smooth, finger-imprinted, or bumpy. The skin is relatively thick and contains numerous stomata. The mesocarp consists mainly of parenchyma, cholenchyma, and sclerenchyma cells, the latter often being highly lignified, similar to those found in pears.

Varietal Selection and Breeding

The atemoya and the cherimoya produce fruit superior in quality to other *Annona* species. The fruit of some species, such as the bullock's hearts (*A.*

Table 1 Important species of the genera *Annona* and *Rollinia*

Species names	*Common names*	*Chromosome number* (2n)	*Gene center*	*Potential uses*
Annona hybrids	Atemoya Cherimorinones Custard apple	16	Various	Fresh fruit Rootstock Processing with dairy products
A. cherimola Mill	Cherimuyu Cherimoya Cherimola Custard apple	16	Andean valleys of Peru and Ecuador	Fresh fruit Rootstock Processing with dairy products
A. squamosa L.	Sugar apple Sweetsop Sitaphal Ata Custard apple	16	West Indies	Fresh fruit Rootstock Processing with dairy products
A. reticulata L.	Bullock's hearts Ramphala Custard apple (often considered the true custard apple)	16	West Indies, Mexico	Fresh fruit Rootstock
A. muricata L.	Soursop Guanabana Corossol Graviola	16	West Indies, Mexico	Fresh fruit Drinks and purées Processing with dairy products
A. glabra L.	Pond apple Alligator apple	28	Florida (USA)	Rootstock
A. senegalensis	Wild Transvaal apple		South Africa	Pollinator species Rootstock
A. diversifolia Safford	Ilama Cherimoyer of the lowlands		Mexico, Guatemala, El Salvador	Fresh fruit
A. montana Macf.	Mountain soursop	16	Cuba, West Indies	Fresh fruit Rootstock
A. purpurea Moc. and Sesse	Soncoya Manirote		Mexico	Fresh fruit Rootstock
A. longiflora	Wild cherimoya		Mexico	Rootstock
A. scleroderma Safford	Posh te		Mexico, Guatemala	Fresh fruit
Rollinia deliciosa Safford	Biriba	48	Brazil	Fresh fruit Rootstock
R. emarginata Schlecht			Brazil, Paraguay	Fresh fruit Rootstock
R. mucosa	Rollinia		Brazil, Paraguay	Fresh fruit Rootstock

reticulata) and ilama (*A. diversifolia*), are barely edible. The fruit of the sugar apple are highly seeded and are not grown commercially in regions where cherimoya or atemoya can be grown. The fruits of rollinia and biriba, although edible, deteriorate rapidly within a few days after harvesting, thus limiting their commercial acceptability for local markets only.

The University of Puerto Rico's Agricultural Experiment Station at one time cataloged 14 different types of soursop. In El Salvador, two types of soursop are grown: guanaba azucarón (sweet), eaten raw and used for drinks, and guanaba acida (very sour), used only for drinks. A yellow, fiberless form of soursop has been selected in Cuba.

The most important commercially grown cultivars are as follows: of cherimoya, Bronceada (Chile), Fino de Jete (Spain), and Bays (California); of atemoya, African pride (Australia) and Gefner (Israel, Florida and Hawaii); of soursop, Cuban fiberless. Wide genetic variability exists between different species of cherimoya and atemoya. Selection of varieties with round, symmetrical shape, smooth skin, high ratios of flesh to seed, and long postharvest storage life is highly desirable. Intensive varietal selection programs for cherimoya are currently being undertaken in Spain, New Zealand, and Chile. A small breeding program, currently in progress in Queensland, Australia, has produced promising new interspecific and intraspecific hybrids.

Maturity and Quality

No reliable quantitative maturity indices for when to harvest *Annona* fruit have been developed. Individual trees are normally harvested up to 10 times, and fruits are judged to be mature when the skin changes from a darker to a lighter green and becomes smoother. In some regions, fruits which develop during the colder winter months may fail to mature properly. Total soluble solids contents at full maturity can vary from 18% to 28%.

Physiological Disorders

Fruit Splitting

Fruit splitting may occur while fruits are still on the tree or after harvest. Splitting appears to be related to sudden changes in fruit moisture content or temperature. Some cultivars appear to be less susceptible than others.

Russetting

Superficial russetting of skin seems to be caused by low night temperatures (less than 13 °C), accompanied by low humidity. Near-mature fruit appear to be particularly susceptible. Cherimoya cultivars are apparently less susceptible than atemoya cultivars.

Crocodile Skin

The extremely wavy and pointed carpels that are symptomatic of this disorder are particularly severe on highly vigorous trees, which may indicate some harmful effect of tree vigor on pollination processes. Based on fruit and leaf analyses, nutrition does not appear to be implicated.

Hard Seed Casing and Brown Lumps

Possible causes are boron deficiency or sudden changes in fruit water content.

Postharvest Handling and Physiology

The skin of most *Annona* fruits, in particular those cultivars which produce fruit with protruding carpels, is easily damaged by rough handling. Fruits are usually size-graded by hand because of their irregular shape, and are packed into single-layer trays before shipment. The cherimoya and atemoya are classified as climacteric fruits which show two peaks in carbon dioxide production. The fruit softens, develops pleasant aroma and flavor, and is considered ripe at the beginning of the second respiratory rise. Compared with other fruits, the cherimoya and atemoya have high rates of ethylene production and respiration. With the exception of a few cultivars, postharvest shelf-life is short (between 7 and 10 days). (*See* **Ripening of Fruit**.)

Storage

The optimum storage temperatures for atemoya and cherimoya fruit are between 13 °C and 16 °C. At these temperatures most atemoya and cherimoya cultivars can be stored for up to a maximum of 2 weeks with minimum loss of quality. One cultivar of cherimoya, Bronceada, has a reported cool storage life of 3 weeks. Storage at temperatures below 13 °C for more than 1 week can cause severe chilling injury. Chilled fruit develops blackened skin, discoloration of the core, and watery patches in the fruit flesh; after removal from storage the fruit fails to ripen properly. However, fruit used for processing can be stored for 1 week at between 5 °C and 10 °C without loss of internal fruit quality. For short-distance transport precooling of fruit at 10 °C for up to 18 h prior to shipment at ambient temperatures has been shown to be beneficial in extending shelf-life by 3–4 days. Because of the presence of numerous stomata on the surface of the fruit, one of the main problems during storage is rapid water loss. Maintenance of relatively high humidity during storage should improve the appearance of fruit after the removal from storage. The main postharvest disease is anthracnose, caused by a range of disease organisms (*Colletotrichum* spp., *Phomopsis* spp., *Rhizopus* spp.). Postharvest fungicidal dips have successfully controlled fruit rots, provided that recommended dipping temperatures and concentrations are adhered to. Only limited studies have been conducted on controlled-atmosphere storage of *Annona* species. Further studies are needed to evaluate the benefits of packaging fruit in polyethylene bags, which may help to prevent water loss and blackening of fruit during storage. (*See* **Controlled-atmosphere Storage: Applications for Bulk Storage of Foodstuffs; Fungicides; Storage Stability: Mechanisms of Degradation.**)

Fresh Food Uses

The flesh of most *Annona* species is most commonly eaten out of hand. Because seeds of the fruit are interspersed throughout the flesh, cultivars containing few seeds are the most desirable. Eating quality of atemoya and sugar apple fruit, which are particularly sweet, may be enhanced by adding a few drops of lime juice to the flesh. Soursops of least acid flavor and least fiber are the only ones suitable for eating fresh. The best eating quality is obtained from fruit that has been ripened at 17–20 °C and then placed in the refrigerator to cool just before eating. In terms of

appearance, color, flavor, and texture, the most acceptable pulp for both eating and processing is to be found 1 day after the first detectable softening. Softened fruit can be stored in the refrigerator (4 °C) for about a week with a minimal loss in flavor, but some skin blackening will result. Segments of the pulp can be added to fruit salads, or used for making sherbets or icecream.

Processed Products

Atemoya, cherimoya, and soursop pulp has potential for mixing with dairy products such as icecream and yogurt, and sweetened soursop juice can be made into a canned beverage. Because of its higher acidity, soursop pulp is the most suitable of all the *Annona* species for processing. Cultivars with a low level of grit or fiber in the flesh have been found to be more suitable for processing than others. Passing the flesh through fine screens (20 000–10 000 μm), and homogenization, as used for guava pulp, should eliminate this problem. The pulp can also be used to make semi-dried fruit 'leathers,' and jams and jellies of fair flavor and quality. In Central America, sweet soursop pulp is sieved and squeezed, diluted with either milk or water, and sweetened, before canning to make a refreshing drink. When mixed with a little cream or icecream, the frozen soursop concentrate makes a delicious dessert.

Some species of *Annona* have been used in the production of soaps, domestic cooking oils, essential oils, herbal medicines, alcohol, fertility drugs, and insecticides. However, the commercial potential of these products at the present time appears limited.

Pulp Extraction

The production of pulp on a commercial scale does not appear to be easy because the skin of *Annona* fruits, at full ripeness, is soft and disintegrates easily if conventional processing techniques are used for skin removal. Other processed fruits, such as mango, passion fruit, and guavas, have leathery skins which are brushed clean of pulp before being discarded. *Annona* fruit disintegrates and is brushed through the screens and mixed in with the pulp. The skins of most *Annona* fruits are high in polyphenols. This causes two problems when the skin is mixed with the flesh – a rapid browning of the pulp and a strong off-flavor. The Food Industry Development Centre of the University of New South Wales, Australia, has evaluated two methods for removing custard apple pulp – a screen pressing process and a reaming process. Although the reaming process offers yield advantages, it requires the use of skilled labor, and this would reduce its potential economic advantages. A prototype screen press extractor has been

Table 2 Chemical and nutritional composition (per 100 g of ripe fruit) of the more important *Annona* spp. fruits

Parameter	*Cherimoya*	*Atemoya*	*Sugar apple*	*Soursop*
Water (g)	74.6–82.8	71.5–78.7	72.5–79.0	77.9–84.0
Fiber (g)	1.5–4.3	0.05–2.5	1.0–1.6	0.8–1.2
Starch (g)	NA	1.1	NA	NA
Sugar (g)	12.0–15.0	18.1	14.6	10.4–12.5
Ash (g)	0.6–1.0	0.4–0.75	0.4–1.4	0.6–0.9
Fat (g)	0.1–0.4	0.4–0.6	0.4–0.6	0.6–1.0
Protein (g)	1.0–2.4	1.1–1.4	1.3–2.4	0.7–1.7
Total acidity (citric acid equivalent)	0.17–0.50	0.2–0.6	NA	0.9–1.3
pH	3.9–4.8	4.4–5.1	3.9–4.8	3.6–4.8
Total energy (kJ)	NA	310–394	368–398	267–297
Ascorbic acid (mg)	4.3–16.8	50	10–51	13–32
Carotene (mg)	0.0–0.02	0.0–0.02	0.01	0.0–0.01
Thiamin (mg)	0.06–0.13	0.05	0.11–0.17	0.05–0.11
Riboflavin (mg)	0.11–0.15	0.07	0.08–0.16	0.03–0.05
Nicotinic acid (mg)	0.73–2.03	0.80	0.7–1.0	0.57–1.28
Calcium (mg)	8.0–32.0	17	19.4–44.7	8–26
Magnesium (mg)	27	32	NA	NA
Phosphorus (mg)	30.2–47.0	NA	23.6–55.3	27–29
Potassium (mg)	298–370	250	NA	179–265
Sodium (mg)	4–6	4.5	NA	9.0–14.0
Zinc (mg)	NA	0.2	NA	NA
Iron (mg)	0.8	0.3	0.28–0.36	0.5–0.8

NA, no data available.

Major sources: Morton JF (1987) *Fruits of Warm Climates*, pp. 65–91. Greenboro, Carolina: Media Incorporated; Duckworth RB (1966) *The Composition of Fruits and Vegetables*, London: Pergamon Press; Nagy S and Shaw PE (1998) *Tropical and Sub-tropical Fruits*. Long Boat Key, Florida, FL: Florida Sc. Source Inc.

developed but this unit has yet to be tested commercially. (*See* **Phenolic Compounds**.)

Pulp Storage

Heating or pasteurization impairs pulp flavor considerably, and often results in the development of a bitter character. However, unblanched frozen pulps treated with either ascorbic acid (1500–2000 mg kg^{-1}) or potassium metabisulfite (500 mg kg^{-1}) can be stored for up to 120 days with minimal loss of flavor. Upon thawing, pulp treated with ascorbic acid may develop a pinkish discoloration, whereas pulp treated with potassium metabisulfite retains a bright, fresh color. (*See* **Ascorbic Acid**: Properties and Determination.)

Nutritional Value

The fruit is rich in starch when firm but increases markedly in sugar as it softens. The main sugars are glucose and fructose (80–90%). There is some phenolic content associated with an increase in the activity of peroxidase, which causes oxidation of the pulp. The volatile fraction consists of alcohol, esters, carbonyls, and hydrocarbons. The nutritional properties of the more important *Annona* species are presented in Table 2. Compared with other fruits, the *Annona* fruits contain significant quantities of vitamin C, thiamin, potassium, magnesium, and dietary fiber content. The calorific value of the flesh is high (300 kJ per 100 g) and is almost double that of peach, orange, and apple. Refer to individual nutrients.

Marketing Characteristics

Both the cherimoya and the atemoya are best eaten as fresh fruits. Both fruits are currently virtually unknown on world markets, and Chile is the only country which exports significant quantities. The major problem facing expansion of cherimoya and atemoya production throughout the world is the high perishability of the product. The transportation of atemoya and cherimoya to distant export markets, given current cultivars and postharvest technology, is difficult. The processing potential of soursop and other *Annona* fruits is yet to be fully exploited. (*See* **Fruits of Tropical Climates:** Commercial and Dietary Importance.)

See also: **Ascorbic Acid**: Properties and Determination; **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Fruits of Tropical Climates**: Commercial and Dietary Importance; **Fungicides**; **Phenolic Compounds**; **Ripening of Fruit**; **Storage Stability**: Mechanisms of Degradation

Further Reading

Bueso CE (1980) Soursop, tamarind and chironja. In: Nagy S and Shaw PE (eds) *Tropical and Subtropical Fruits*, pp. 357–407. AVI Westport, CT: AVI.

Duckworth RB (1966) *The Composition of Fruits and Vegetables*. Fruit and vegetables, appendix A. London: Pergamon Press.

Edwards RA (1989) *The Processing of Custard Apple and the Development of Pulp Extraction Equipment*. University of New South Wales, Australia: Food Industry Development Centre.

George AP and Nissen RJ (1985) The custard apple, part 1. Species, variety and rootstock selection. *Australian Horticulture* 83(10): 100–107.

George AP, Nissen RJ and Brown BI (1987) The custard apple. *Queensland Agricultural Journal* 113(5): 287–296.

Morton JF (1987) *Fruits of Warm Climates*, pp. 65–91. Greensboro, North Carolina: Media Incorporated.

Sanewski GM (1988) *Growing Custard Apple*, p. 86. Brisbane: Queensland Department of Primary Industries.

ANOREXIA NERVOSA

D S Goldbloom and P E Garfinkel, Centre for Addiction and Mental Health, Toronto, Ontario, Canada

Definition

Anorexia nervosa (AN) is an eating disorder that is characterized by self-imposed starvation accompanied by an idealization of thinness and a morbid fear of becoming fat. The resultant emaciation leads to significant medical and psychiatric complications. Contrary to the literal meaning of its name – a nervous loss of appetite – AN does not reflect a true loss of appetite except in the later stages of starvation; rather, it is a determined battle waged against sensations of hunger in order to achieve an illusory sense of control. Since its original modern description by Sir William Gull in 1874, the diagnostic criteria for AN have evolved to reflect the biological, behavioral, and psychological disturbances characteristic of the

disorder. The formal criteria as defined by the American Psychiatric Association in 1994 are listed below:

1. Refusal to maintain body weight at or above a minimally normal weight for age and height (e.g., weight loss leading to maintenance of body weight less than 85% of that expected; or failure to make expected weight gain during period of growth, leading to body weight less than 85% of expected).
2. Intense fear of gaining weight or becoming fat, even though underweight.
3. Disturbance in the way in which one's body weight or shape is experienced, undue influence of body weight or shape on self-evaluation, or denial of the seriousness of the current low body weight.
4. In postmenarcheal females, amenorrhea, i.e., the absence of at least three consecutive menstrual cycles. (A woman is considered to have amenorrhea if her periods occur only following hormone, e.g., estrogen, administration.)

Subtypes

1. Restricting type: during the current episode of AN, the person has not regularly engaged in binge-eating or purging behavior (i.e., self-induced vomiting or the misuse of laxatives, diuretics, or enemas).
2. Binge-eating/purging type: during the current episode of AN, the person has regularly engaged in binge-eating or purging behavior (i.e., self-induced vomiting or the misuse of laxatives, diuretics, or enemas).

Prevalence

AN is, in about 90% of cases, a disorder of females, typically between the ages of 12 and 40. Studies of the age of onset of this disorder reveal bimodal peaks at 14 and 18 years of age; at these ages, the demands of secondary sexual development and autonomy in the context of leaving home may be related to the increased incidence. The current accepted prevalence rate for AN among western adolescent and young adult women is approximately 0.5–1.0%. AN is, relatively speaking, a culture-bound disorder limited to industrialized societies with western values.

Etiology and Groups at Risk

The etiology of this disorder is unknown; most clinicians and researchers endorse a multidimensional model of AN which acknowledges psychological, biological, and social risk factors that exist at the level of the individual, the family, and society. While dieting is almost ubiquitous behavior among women in western society, the significant medical and psychiatric manifestations of AN and its characteristic psychological features argue for the definition of AN as a distinct disorder, and not a simple exaggeration of current western values; indeed, AN was recognized long before our cultural preoccupation with being thin.

At the level of the individual, women who by career choice experience pressures both to be thin and to achieve may be particularly vulnerable. This includes ballerinas, fashion models, and gymnasts, in whom thinness and perfectionism are not only demanded but also equated. Conflicting pressures on women to nurture and to perform professionally may intensify concerns about control that are shifted on to weight regulation as a focus. Further, chronic physical illness, physical or sexual abuse, or other traumas which heighten preoccupation with the body as a reflection of control or self-esteem may increase vulnerability to AN.

At the familial level, there may be a magnification of social idealization of thinness and denigration of obesity. Genetic influences through vulnerability to a variety of psychiatric disorders such as depression, alcoholism, and eating disorders themselves in other family members may also play a role. First-degree relatives of AN subjects also display an elevated prevalence of eating disorders over controls. Twin studies indicate an increased concordance rate for monozygotic (identical) versus dizygotic (nonidentical) twins.

At the societal level, the last 75 years have witnessed shifting models of the ideal female form toward thinness, in contrast to the increasing biologically ideal weights for women in the context of good nutrition. However, it must be remembered that the decription of AN in 1874 long predates our cultural preoccupation with thinness and a broader understanding of this disorder must be sought.

Psychopathology

The focal behavioral point of AN is the individual's relentless pursuit of thinness and the associated conviction that her body is too large. This latter belief may emanate from a variety of sources, from a past history of obesity (with its attendant social humiliation and a sense of failure) to physical or sexual abuse which has led to body loathing. Alternatively, AN may reflect a symbolic focus on the body in a search for personal mastery that in turn reflects a sense of personal ineffectiveness. It may also allow retreat from the maturational demands of adolescence, as expressed by the development of secondary

sexual characteristics. Associated psychopathological features often include a fear of losing control in many spheres, self-esteem that is highly contingent on the opinions of others, an all-or-nothing thinking style that allows no middle ground between being thin and being fat, and a pervasive sense of helplessness.

Clinical Features

The initial manifestations of AN are deceptively benign; the disorder frequently commences with dieting behavior which is itself common in the western female population. However, this particular form of dieting leads to increased criticism of one's appearance and greater social isolation than typical adolescent dieting behavior; further, weight goals continue to drift downward as they are approached. Food aversion becomes more general than specific and, as emaciation progresses, the individual exhibits a marked denial of her illness. A typical but not universal characteristic of AN is a distortion of body image in which the individual overestimates her body size substantially, believing herself to be fat even when thin. This perceptual disturbance is usually accompanied by feelings of loathing of one's appearance.

At the level of behavior, routine dieting evolves into more elaborate food avoidance and other activities to counteract the effect of ingested calories. Individuals restrict their intake dramatically, skip meals, or secretly dispose of food. Elaborate rituals related to eating develop, and decreased consumption of food is paralleled by increasing cognitive preoccupation with it. This translates into thinking, reading, and dreaming about food or working in food-related industries. Weighing oneself becomes not only a frequent ritual but also a regulator of subsequent eating, socialization, and self-esteem. In addition to calorie restriction, individuals may exercise intensively for the express purpose of weight loss, or use purgative techniques such as diuretics, diet pills, laxatives, or self-induced vomiting.

Initially, because the consequences of the disorder are consistent with the individual's desire to be thin and in control, there is seldom the motivation to seek help or to acknowledge the evolving AN as a problem. Rather, the individual with AN may contact a nutritionist, physician, or other health professional, seeking advice on dieting in the absence of obesity or help in dealing with the secondary medical and psychiatric effects of starvation.

Physiological Effects

Virtually no body system is spared the effects of nutritional deprivation in AN. The changes are too extensive to enumerate here; they reflect not only loss but also homeostatic compensatory mechanisms to minimize energy expenditure in the context of decreased energy intake. Loss of menstrual function is the most classically recognized of these sequelae, to the point that it is a diagnostic criterion thought to reflect disturbances in gonadotropin function in the hypothalamus. Other common starvation effects include lowered body temperature and heart rate, dry skin and the development of a fine body hair, abdominal bloating and constipation, and edema.

Psychological Effects

Starvation due to AN also produces distinct psychological effects such as depressed, irritable, or anxious mood, cognitive preoccupation with food, and decreased concentration. Individuals with AN may appear to suffer from a coexistent depressive illness and are vulnerable to the long-term development of such disorders. However, many of the psychological sequelae of the disorder – which are more troubling to the affected individual than the dieting itself – are reversible with adequate nutrition.

Nutritional Status

The food restriction of AN may be severe. Individuals may abstain from all solid foods for days at a time or subsist on as little as 200–500 calories per day. Typically, there is an avoidance of perceived high-calorie, high-fat foods – and foods become imbued with moral as well as nutritional value. Individuals may develop frank cachexia, with vigorous dieting and exercise behaviors persisting at weights as low as 30 kg. Fasting hypoglycemia, abnormal glucose tolerance tests, and increased insulin sensitivity are among the metabolic responses to this form of starvation. Elevations of cholesterol and carotene may occur and plasma zinc levels have variably been reported as low. Low platelet, red blood cell and white blood cell counts, accompanied by a relative paucity of precursors in bone marrow, have been described as a consequence of the disorder.

The nutritional status of AN patients may have implications for the chronicity of the disorder. Brain neurotransmitters such as serotonin and norepinephrine (noradrenaline) are integral in the hypothalamic regulation of appetitive behavior. Serotonin synthesis itself is dependent on the dietary availability of its precursor, the essential amino acid tryptophan. Tryptophan availability may be decreased in AN, and both neuroendocrine provocation tests and cerebrospinal fluid metabolite assays indicate altered serotonin activity in AN. This may be associated with

disturbances in satiety, thought (with increased obsessionality and perfectionism), mood, and sleep found in AN. Delayed gastric emptying may occur in response to the decreased volume of food consumed. This further contributes to altered senses of satiety; an augmented sense of fullness may perpetuate food restriction.

Diagnosis

The diagnostic criteria of the American Psychiatric Association listed above reflect the confluence of biological, behavioral, and psychological factors characteristic of AN. However, clinicians may perceive them as arbitrary, particularly with regard to the degree of weight loss required and the presence of amenorrhea as essential for diagnosis. Epidemiological research requires dichotomous criteria, while clinical practice demands the recognition from a dimensional perspective of disorders in evolution. The diagnosis of AN in a young woman with weight loss requires consideration of other medical and psychiatric illnesses with a superficially similar presentation. These include endocrine disorders, bowel disease, malignancy, depression, schizophrenia, and conversion disorder. All of these differential diagnoses typically lack the central psychological preoccupation of AN – the relentless pursuit of thinness.

Treatment

The goal of treatment of AN includes but is not limited to nutritional rehabilitation. Clearly, if weight gain is the only purpose of therapy for an individual dedicated to the attainment of weight loss, her commitment to the treatment will be poor. At the same time, however, treating personnel cannot ignore the nutritional status of the individual while focusing on psychological issues; when this occurs, it results in collusion with the denial by the individual of her illness and obscures the connection between food deprivation and its sequelae.

The initial phase of treatment is a careful diagnostic assessment, with emphasis on both the evolution of the disorder and its various sequelae. A family assessment is often relevant, particularly in younger patients, both to understand familial influences and to substitute education for guilt. A target weight range is established, allowing the minor fluctuations that are normal (1–2 kg). This range should be able to be maintained without dieting, should allow the return of normal menstrual function, and should reflect consideration of the individual's longitudinal weight history. Generally, a target weight range is above 90% of the average for an individual's weight and height or a body mass index (kg m^{-2}) greater than 20. Despite a seemingly encyclopedic knowledge of nutrition, these individuals usually require directive counseling regarding meal frequency, portion size, and macronutrient selection. A daily diary of eating and associated thoughts, feelings, and behaviors may be helpful. An initial intake of 1500 calories per day is usually sufficient to promote weight gain without inducing the gastric dilatation that can complicate refeeding. A rate gain of 0.5–1.0 kg per week is desirable; more rapid weight gain may induce its own complications, including hypophosphatemia and edema, as well as mistrust in an individual who is reluctantly relinquishing weight control. Caloric intake is usually increased by 200–300 calories per week toward a goal of 2400–3000 calories per day. To date, controlled clinical trials indicate no role for drugs in the promotion of eating or weight gain in AN; food remains the drug of choice; there is modest evidence that the antidepressant fluoxetine may assist weight-recovered AN patients to maintain weight gain, possibly through an antiobsessional effect.

Psychotherapy is usually offered in conjunction with nutritional rehabilitation, and builds on the establishment of a therapeutic relationship. Issues include the recognition of feelings, self-trust, and disconnecting one's sense of self-worth from body weight. A variety of approaches, from psychodynamic to cognitive-behavioral, may be employed. Family therapy may be particularly helpful for the younger AN patient.

Hospitalization is not necessary for the majority of individuals with AN; rather, it is reserved for cases where the weight loss has been either precipitously acute or impinging on basic function, where medical sequelae such as hypokalemia pose an imminent risk, where suicidal tendencies accompany the AN, where AN coexists in a threatening fashion with another illness such as diabetes mellitus, or where other forms of treatment have been ineffective.

Prognosis

AN is a usually gradual, initially covert, and sometimes chronic disorder. Long-term follow-up studies indicate that while two-thirds of patients show eventual improvement or recovery, less than one-third recover within 3 years. More disturbing, some long-term follow-up studies have indicated a mortality rate of 10–20% as a consequence of AN. In addition, these women are vulnerable over the long term to the development of mood disorders, anxiety disorders, and substance abuse, regardless of whether the AN is active or quiescent. They are also susceptible to such diverse medical complications as

osteoporosis with pathological bone fractures and a form of brain tissue atrophy. Beyond reversal of the nutritional deprivation and weight loss, there is no established method of tertiary prevention.

See also: **Adolescents**; **Bulimia Nervosa**; **Famine, Starvation, and Fasting**; **Premenstrual Syndrome: Nutritional Aspects**; **Slimming**: Slimming Diets; Metabolic Conseqences of Slimming Diets and Weight Maintenance

Further Reading

American Psychiatric Association (1994) *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn. Washington, DC: American Psychiatric Association.

American Psychiatric Association (2000) Practice guideline for the treatment of patients with eating disorders.*American Journal of Psychiatry* 157 (suppl): 1–39.

Garfinkel P, Lin E, Goering P *et al.* (1996) Should amenorrhea be required for the diagnosis of anorexia nervosa? Evidence from a Canadian community sample. *British Journal of Psychiatry* 168: 500–506.

Garner DM and Garfinkel PE (eds) (1997) *Handbook of Treatment for Eating Disorders*, 2nd edn. New York: Guilford Press.

Goldbloom D (1997) Evaluating the efficacy of pharmacologic agents in the treatment of eating disorders. *Essential Psychopharmacology* 2: 71–88.

Goldbloom D and Kennedy S (1995) The medical complications of anorexia nervosa. In Brownell KD and Fairburn CG (eds) *Eating Disorders and Obesity: A Comprehensive Handbook*, pp. 266–270. New York: Guilford Press.

Kaplan AS and Garfinkel PE (eds.) (1993) *Medical Issues and the Eating Disorders: The Interface*. New York: Brunner/Mazel.

Katzman D, Lambe E, Ridgley J, Goldbloom D, Mikulis D and Zipursky R (1996) Cerebral gray matter and white matter volume deficits in adolescent females with anorexia nervosa. *Journal of Pediatrics* 129: 794–803.

Kaye WH, Gwirtsman H, Obarzanek E and George DT (1988) Relative importance of calorie intake needed to gain weight and level of physical activity in anorexia nervosa. *American Journal of Clinical Nutrition* 47: 989–994.

Kennedy S and Goldbloom D (1994) Advances in the treatment of anorexia nervosa and bulimia nervosa. *CNS Drugs* 1: 201–212.

Szmukler G, Dare C and Treasure J (eds) (1995) *Handbook of Eating Disorders*. Chichester: John Wiley.

Anthropometry *See* **Nutritional Assessment**: Anthropometry and Clinical Examination

ANTIBIOTIC-RESISTANT BACTERIA

J J Klawe and M Tafil-Klawe, The Ludwik-Rydygier Medical University, Bydgoszcz, Poland

Bacterial Resistance to Antimicrobial Agents

The antimicrobial resistance may be intrinsic or acquired. Intrinsic–innate resistance may be considered a fundamental property of the organism: it was present in bacterial strains isolated before the antibiotic era. Some organisms can produce enzymes that chemically modify a specific drug. Bacteria may also use the system to transport detrimental compounds out of a cell, using efflux pumps.

Acquired resistance is not found in species strains, unless they have been exposed to antimicrobial drugs. Acquired resistance may occur as a result of spontaneous chromosomal mutation or by acquisition of extrachromosomal elements, such as plasmids and transposons.

Chromosomal mutation can occur during treatment or in the absence of the agent, against which resistance develops. Extrachromosomal pieces of DNA, called plasmids (R plasmids), contribute to bacterial virulence or to antimicrobial resistance. Genes of resistance to antimicrobial agents may be transferred from chromosomes to plasmids or from plasmids to chromosomes, as a result of classic recombination. The plasmid and chromosome recombine pieces of DNA in areas of genetic homology. The presence of a *rec* A gene is needed. Many plasmids are composed of two parts: the resistance genes, or R genes, which code for the resistance traits, and the resistance transfer factor, which codes

for the transfer of the plasmid to other bacteria by conjugation.

Transposons, however, can insert into chromosomes in areas where no homology exists, independent of a *rec* A gene. Thus, transposons are less restrictive and play a major role in the transfer of resistance among many different species.

Biochemical Mechanisms of Bacterial Resistance

The biological mechanisms of bacterial resistance may be classified into four groups:

1. modification of the target site of the antimicrobial agent – for example, resistance to rifampin caused by alteration of the targeted RNA polymerase;
2. enzymatic inactivation of the agent – production of β-lactamases, enzymes that hydrolyze the β-lactam ring of the β-lactam antibiotics;
3. decreased membrane permeability and entry of the agent – resistance to tetracycline is due to decreased accumulation of the antimicrobial agent;
4. a combination of two or more mechanisms – for example, resistance of *Neisseria gonorrhoeae*.

Examples of Emerging Antimicrobial Resistance

Enterococci

The Enterococci are a group of bacteria that are a part of the normal intestinal flora. Many enterococci have plasmid-encoded resistance genes. This group of bacteria is intrinsically less susceptible to many common antimicrobials. Enterococci cause nosocomial infections, an old problem that has occurred ever since the first hospitals were established. Nowadays, nosocomial infections make up at least half of all cases of infectious disease treated in the hospitals and may range from very mild to fatal. Enterococci are a common cause of nosocomial urinary-tract infections as well as wound and blood infections.

Staphylococcus aureus

Many people, including healthcare personnel, are carriers of this organism. It is a common cause of nosocomial pneumonia and surgical infections. Hospital strains are resistant to a variety of antibiotics. Over the past 50 years, most strains have acquired resistance to penicillin, due to the acquisition of a gene encoding the enzyme penicillinase. New strains, methicillin-resistant *Staphylococcus aureus*, are resistant to methicillin and to other β-lactam drugs. *Staphylococcus aureus* has also been reported to have become vancomycin-resistant.

Staphylococcus Species other than *S. aureus*

These normal skin flora can colonize the tips of intravenous catheters. The resulting biofilms continuously seed organisms into the bloodstream, thus increasing the risk of systemic infection.

Streptococcus pneumoniae

Some strains, now isolated, are resistant to penicillin. This acquired resistance is due to the modifications in the chromosomal genes coding for different penicillin-binding proteins, which decrease their affinities for the drug. The nucleotide changes are due to the acquisition of chromosomal DNA from other species of *Streptococcus*.

Mycobacterium tuberculosis

Treatment of infections of *Mycobacterium tuberculosis* has always been a complicated process, often requiring a combination of two or three different drugs taken for 6 months or more. *Mycobacterium tuberculosis* is developing resistance to first-line drugs. Strains that are resistant to isoniazid and rifampin are called multiple drug-resistant *Mycobacterium tuberculosis*.

Pseudomonas species

These bacteria grow in a large number of moist, nutrient-poor environments, such as the water in the humidifier of a mechanical ventilator. *Pseudomonas* species are resistant to many antimicrobial drugs. They are a common cause of hospital-acquired pneumonia and infections of the urinary tract and burn wounds.

Slowing down the Emergence and Spread of Antimicrobial Resistance

More effort from healthcare workers to identify the causative agent of infectious diseases and to prescribe suitable antimicrobials is necessary to slow down the spread of antimicrobial resistance. Also, patients need to be vigilant, and a great effort must be made to educate the public about the appropriateness and limitations of antibiotics.

See also: ***Mycobacteria***; ***Staphylococcus***: Properties and Occurrence; Detection; Food Poisoning

Further Reading

Advisory Committee on Microbiological Safety of Foods (UK) (1999) *Report on Microbial Antibiotic Resistance in Relation to Food Safety*. London: The Stationery Office.

Howard BJ (1994) *Clinical and Pathogenic Microbiology*. London: Mosby.

Jacoby GA and Archer GL (1991) New mechanisms of bacterial resistance to antimicrobial agents. *New England Journal of Medicine* 324: 601–612.

McDougal LK and Thornsberry C (1986) Role of beta-lactamase in staphylococcal resistance to penicillinase-resistant penicillins and cephalosporins. *Journal of Clinical Microbiology* 23: 832–839.

Nester EW, Anderson DG, Roberts CE, Jr., Pearsall NN and Nester MT (2001) *Microbiology – A Human Perspective*. London: McGraw-Hill Higher Education.

Patterson JE and Zervos MJ (1990) High-level gentamicin resistance in enterococcus: Microbiology, genetic basis and epidemiology. *Reviews of Infectious Diseases* 12: 644–652.

ANTIBIOTICS AND DRUGS

Contents

Uses in Food Production

K N Woodward, Veterinary Medicines Directorate, Weybridge, UK

Background

Veterinary medicines are used to improve or maintain the health of animal species regardless of whether these are intended for food production. They are designed and manufactured to cover a wide variety of prophylactic and therapeutic purposes, and they are administered to household pets, exotic species, and wild animals in addition to food-producing animals.

Most countries possess a governmental regulatory authority charged with assessing veterinary medicines and granting some form of marketing authorization before a product may be sold, and the UK's system is used here to exemplify this in practice. Veterinary medicines, like their human counterparts, are assessed against three basic criteria – efficacy, quality, and safety – although the terms of these criteria differ from country to country. The efficacy is the ability of the drug to accomplish the task claimed for it by its producer. Quality refers to pharmaceutical quality; the levels of contaminants must meet acceptable criteria, the shelf-life must be as claimed by the manufacturer, and the product must be produced to appropriate standards. Safety means that the product must not harm the animal patient, the environment, the users of veterinary medicines, or the consumer of food of animal origin. This latter aspect is of paramount importance when assessing the safety of veterinary medicines. When a veterinary drug is given to an animal, most of the dose will be excreted in urine, feces, and expired air, but small amounts will remain in body tissues. These are known as residues, and it is the responsibility of regulatory authorities to ensure that these do not pose a threat to the consumer when veterinary medicines are used in food-producing animals.

Uses of Veterinary Medicines

In general, veterinary medicines are available to treat most of the types of animal disease state that are encountered in humans. In practice, diseases such as cancer are not usually treated in food-producing species, largely for economic reasons, although they are frequently dealt with in companion animal medicine. The majority (but not by any means all) of the medicines used in food-producing animals are for the prevention or treatment of diseases caused by infectious agents. These fall into a number of categories, and they are described briefly below.

Antibiotic and Antimicrobial Agents

Antibiotics are antibacterial agents derived from living organisms, but the term includes closely related synthetic analogs. The archetypal compound is penicillin, the structure of which has now formed the basis for a wide range of naturally occurring and semisynthetic analogs. Other major categories of antibiotics

used in veterinary medicine are the tetracyclines, the aminoglycosides, the macrolides, and the polymixins. Sulfonamides constitute the major class of antimicrobial agent, drugs that are unrelated to naturally occurring compounds. A number of sulfonamides have been prepared for fast, medium, or long-acting functions. Sulfadimidine is one of the most widely used sulfonamide drugs in veterinary medicines, particularly in pig production, where it is employed to combat respiratory disease.

Ectoparasiticides

Cattle and sheep are often subject to attack by ectoparasites. For example, in the UK and other parts of Europe (and indeed elsewhere) cattle are susceptible to attack by the warble fly (*Hypoderma* spp.), which, in addition to causing animal suffering, results in considerable economic loss. Sheep scab is another disease that poses serious animal welfare and economic problems, and in several countries, including the USA and the UK, it is a notifiable disease. Both diseases are treated and prevented by the application of organophosphorus compounds. For warble fly, treatment is usually in the form of a relatively viscous formulation that is spread on to the back of the animal (a pour-on formulation); in the case of sheep scab, dipping and showering are the most therapeutically and economically effective methods currently available.

Fish too are susceptible to external parasites, and as fish farming (aquaculture) becomes more important, so the effects of disease become more significant. Salmon farming is a fast-growing industry in many parts of the world, particularly in North America, Chile, Norway, and Scotland. Salmon are vulnerable to infestation by the sea louse, an arthropod that attacks the external surface of the fish and in severe cases causes death.

Anthelmintics

'Anthelmintics' is a term almost universally abused to describe agents used to treat internal parasite infestations and not merely those caused by helminth worms. Many of the diseases caused by internal parasites can be extremely distressing to affected animals, and all can result in substantial economic losses. A number of drugs have been developed to combat these diseases, including the benzimidazoles, levamisole and invermectin.

The first benzimidazole to achieve widespread use in veterinary medicine was thiabendazole, but newer compounds include albendazole, oxfendazole, fenbendazole, and mebendazole. Levamisole is extremely efficacious in the treatment of gastrointestinal nematodes in cattle, sheep, and pigs. Ivermectin is a mixture of two closely related compounds derived from abamectin, a metabolite of *Streptomyces avermitilis*, and it has found wide use as an antiparasitic agent in veterinary medicine (and in some areas of human medicine).

Antifungal Agents

In veterinary medicine, several drugs have found uses as antifungal agents for application to the external surfaces of the body. These include ketoconazole, thiabendazole, aliphatic acids, and benzoic acid. Nystatin and griseofulvin are extensively used as systemic treatments for fungal infections in both food-producing and other species.

Steroid Hormones

It has long been known that testosterone exerts anabolic effects in both humans and animals, and testosterone and chemically related steroid hormones have been widely used for this purpose in beef production. However, the EC has prohibited the growth-promoting uses of steroid hormones in animal production, although certain so-called zootechnical (e.g., synchronization of estrous for mating) and therapeutic (e.g., treatment of recurrent abortion) uses are still permitted using naturally occurring compounds or closely related derivatives. In other countries, notably in the North and South Americas, the growth-promoting uses of steroid hormones are legally permitted.

Regulatory Control of Veterinary Medicines

Regulatory control of the uses of veterinary medicines is almost universal, and, as described earlier, drugs must meet exacting standards of efficacy, quality, and safety. In the UK, for example, human and veterinary medicines are controlled by the Medicines Act 1968 and by integrated EC legislation. Product licences and other forms of marketing authorization are granted by a licensing authority established under the act. The Licensing Authority is defined in the act as the Health and Agriculture Ministers acting together, and in practice, responsibility is given over to civil servants working in the Department of Health for human medicines, and in the Veterinary Medicines Directorate (VMD) of the Ministry of Agriculture, Fisheries and Food (MAFF) for veterinary medicines. Applications for product licences are submitted to the VMD by companies together with the quality, efficacy, and safety data. Staff of the VMD then vigorously assess these data and submit them with their appraisals to

the independent Veterinary Products Committee appointed under the Medicines Act. This committee may then recommend licensing, require further information, or refuse the application. When a product is not recommended for licensing, companies have recourse to an appeals procedure, whereby they can submit their arguments. (*See* **Legislation**: Contaminants and Adulterants.)

Essentially similar systems exist in other countries, including other member states of the EC. For this latter group, national laws are amended from time to time to incorporate EC legislation. Three European directives are worthy of note for the impact that these have had on national legislation within the EEC.

The two Veterinary Medicines Directives (81/851/EEC and 81/852/EEC) provide a legislative framework for the European 'licensing' of conventional veterinary medicines. They establish an expert committee, the Committee for Veterinary Medicines Products, and offer a system of control, including a multistate authorization procedure within the EC. Moreover, they provide comprehensive guidelines on safety, quality and efficacy. The so-called Feed Additives Directive (70/524/EEC) covers both medicinal and nonmedicinal ingredients intended to be added to the animal feeds. Although conventional veterinary medicines and medicinal feed additives are therefore dealt with separately at community level, in the UK, they are subject to the same treatment under the Medicines Act, once they have been accepted by the EC.

On a worldwide basis, the marketing authorization of veterinary medicines and the steps preceding this can be summarized briefly as follows: development of a drug and its formulations; testing to fulfill the requirements of safety, quality and efficacy; submission of an application for marketing authorization with the data from safety, quality, and efficacy tests to the appropriate national authorities; appraisal, assessment, and consideration by those authorities; granting or refusal (or requirement for further studies) of marketing authorization.

Safety Data Requirements

Safety data with respect to consumer safety can be divided into two main categories: toxicity data and residues data. In recent years, there has been an increasing demand for data on microbiological safety with respect to possible effects on the gastrointestinal flora of consumers.

Toxicology Studies

Toxicology data usually take the form of the results from a package of studies using laboratory species and *in-vitro* tests. Pharmacokinetic and pharmacodynamic data are also used where relevant in the interpretation of the results of these endeavors. The precise requirements for toxicology testing differ from country to country, but in general terms, they are exemplified by those of the EC, which include the following:

- single-dose (acute) toxicity;
- repeated-dose toxicity;
- reproductive toxicity: effects on reproduction; effects on the embryo or fetus and teratogenicity (birth defects);
- mutagenicity studies;
- carcinogenicity (induction of tumors) studies;
- pharmacodynamics.
- pharmacokinetics;
- observations in humans.

There are three basic aims in conducting toxicology studies: to identify a no-observed effect level (NOEL), to calculate an acceptable daily intake (ADI), and to determine a maximum residues limit (MRL).

The NOEL is the lowest dose level, usually quoted in milligrams per kilogram bodyweight, below which an effect or range of effects do not occur in the species being investigated (usually mice or rats). It is usually necessary to eliminate those effects thought to be irrelevant – usually because they are known to occur only in rodents or because they represent a phenomenon induced by excessive doses – and then select the study thought to be of most relevance, although in practice, the lowest NOEL from the battery of tests is often used.

The ADI is calculated by dividing the NOEL by an arbitrary safety factor. This is usually 100 for minor toxic effects, but factors of up to 2000 may be used with severe effects or when the battery of tests is considered to be deficient in some relatively minor aspect; if deficient in some major area, e.g., if a carcinogenicity study has not been conducted, an ADI may not be set. The ADI is often expressed in terms of human weight by use of the additional factor of 65 kg, this being widely accepted as representing human adult weight. Hence, the ADI can be expressed (in milligrams per day) as follows:

$$\text{ADI} = (\text{NOEL} \times 65)/\text{safety factor}.$$

The MRL is the maximum permitted level of the drug residue, which regulatory authorities perceive as being acceptable in food of animal origin on the basis of the results of the safety studies described earlier. As such, it is often difficult to describe in purely arithmetical terms, for, although it is based primarily on the ADI, it must take into account other factors, not least of which is the intake of the

food commodity in question. Every consumer has a varied intake of most of the products of food of animal origin, such as muscle, fat, kidney, and liver; some, for example, eat more liver or kidney than others. To try to cover the majority of possibilities, the Food and Agriculture Organization, World Health Organization (FAO/WHO) Joint Expert Committee on Food Additives (JECFA) has recommended standard daily intake values of 300 g of muscle, 100 g of liver, 50 g of fat, 50 g of kidney, 100 g of egg, and 1.5 l of milk to be used in the elaboration of MRLs. The types of food-producing animals are shown in Figure 1. In general terms, the ADI is 'spread' across the different types of food commodity, bearing in mind their relative standard intakes, so that MRLs, which will not result in the ADI being exceeded, can be elaborated.

When considering values for MRLs, a major factor that must be taken into account is the chemical nature of the residues and their resultant biological activity. This is an extremely complex area and one that is not without controversy. Most chemicals, including veterinary medicines, when given to animals are converted to a range of metabolites, usually but not exclusively by the liver. Consequently, the residues present in tissue may be the parent compound or a mixture of metabolites or both. Moreover, either the parent drug or its metabolites may be highly reactive and may therefore combine with cellular constituents to give rise to so-called bound residues. In these circumstances, it is desirable and sometimes essential to know to what extent these may be released to yield potentially harmful substances following consumption of products of animal origin. The most likely place for such a release to occur is in the gastrointestinal tract of the consumer, and a number of animal as well as several chemical models have been developed in an attempt to provide information on this area of hazard assessment. One of these makes use of a technique known as relay toxicology, although in this context, relay pharmacology might be more appropriate. The radiolabeled drug is administered to the food-producing animal, which is later slaughtered, and its tissues, known to contain bound residues, are fed to laboratory animals such as rats. Detection of radioactivity in the plasma (for example) of these animals would indicate that the bound residues are bioavailable and, as a result of digestive processes, have been released for subsequent absorption. Chemical methods using solvent extraction and acid–base treatments investigate the ease with which bound residues can be released. These methods then determine whether or not the contribution from bound residues needs to be taken into account in the MRL setting procedure.

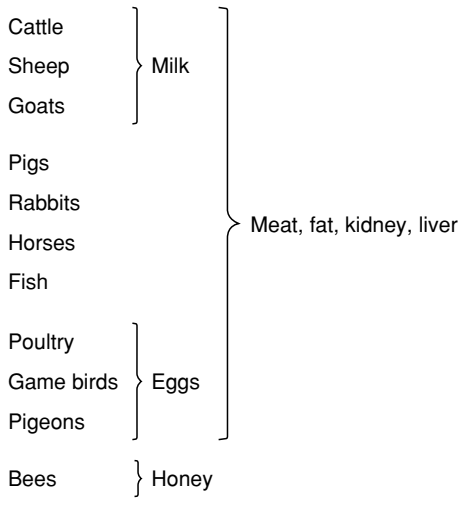

Figure 1 Major food-producing species and their products important in considering maximum residues limits (MRLs) and residues studies. Reproduced from Antibiotics and Drugs: Uses in Food Production, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

As indicated earlier, microbiological safety testing has been proposed for use in the hazard assessment process and indeed has been incorporated into some testing guidelines, including those of the EC. The requirement for such testing is based on the premise that residues of some antibiotic and antimicrobial compounds may have the potential to affect adversely the gastrointestinal flora of the consumer, either by favoring the growth of resistant bacteria or by allowing the growth of organisms not normally found in any numbers in the gastrointestinal tract, by reducing the numbers of other organisms that normally suppress them. This is an extremely controversial area of safety assessments for two main reasons. First, despite the use of antibiotic and antimicrobial drugs in veterinary medicine for many years, there is no evidence to suggest that harmful effects have arisen in the human population as a result of biological changes in the gut flora. Second, the tests proposed to identify these effects have not been validated, so that even if the effects do exist, it is not known whether or not the proposed tests would actually detect them.

There are three tests currently available. Human volunteer studies involve the administration of the drug and subsequent examination of feces and the oral cavity for changes in floral composition. The major alternative involves the use of germ-free rodents in which human gut flora have been seeded into the gastrointestinal tract. These are then given the drug under investigation and changes in the gut

flora monitored. The third approach makes use of *in-vitro* studies of the susceptibilities of various types of bacteria to antibacterial drugs, before and after exposure to low, sublethal levels of the drug. The numerical results in terms of dose can then be employed in the calculation of the ADI.

Residue Studies

Having determined an MRL value (or values) for a drug, it is then essential to ensure that residues in the tissues or other products of food-producing animals do not exceed these. This is achieved by determining the withdrawal (or withholding) period for the drug in question. Small groups of animals are treated with the medicine, usually at the maximum dose recommended for its use in practice. The animals are then slaughtered in a serial manner after various intervals to determine when residue levels fall below the MRL. This time period for depletion of residues to below the MRL is then usually chosen as the withdrawal period; animals must be retained unmedicated during this time as the minimum interval between drug administration and slaughter. Where appropriate, milk, eggs, and honey must be discarded until this minimum time has elapsed.

Enforcement

There is little point in progressing through the various procedures detailed above if the withdrawal period is ignored, and unacceptable levels of residues result. There is often pressure to ensure that animals are sent to slaughter at particular times, and this can lead to abuse.

To ensure that withdrawal periods are observed, most countries operate some form of residues surveillance scheme. In the UK, for example, over 40 000 samples are taken annually from slaughterhouses, and these are tested for residues of various veterinary drugs in compliance with the requirements of the EC Residues Directive (86/469/EEC). Other EC member states carry out similar surveillance in order to comply with the requirements of this directive. The system allows enforcement officers to trace carcases back to farms to determine why the problem arose and, where necessary, to take legal proceedings against offenders.

The Codex Alimentarius Commission

The Codex Alimenatarius Commission is a joint WHO and FAO body, and one of its objectives is to ensure that food of plant and animal origin worldwide is safe for human consumption. It has recently commenced work on establishing MRLs for veterinary drugs, having been involved for many years in setting these for pesticide residues. The Codex is advised on toxicological and residues issues by JECFA, the published scientific reports of which summarize the major conclusions and the proposed MRLs.

Given the international standing of the Codex system and of the JECFA, it seems likely that the MRL values that it promulgates will find their way eventually into various national regulations. In particular, it offers a basis for scientific assessment not currently available in some poorer and less developed nations.

Concluding Considerations

In most countries of the world, veterinary medicines are assessed for their ability to induce toxic effects, and these effects and the doses that produce them are taken into account when calculating the ADI, which is in turn used to determine an MRL value or values. Residues studies are then used to determine when levels of the drug and its metabolites are depleted to below the MRL in food-producing animals. Residues surveillance is essential to ensure that withdrawal periods are observed and, moreover, to ensure that the consumer is not exposed to unacceptable residue levels.

See also: **Hormones**: Steroid Hormones; **Legislation**: Codex; **Parasites**: Occurrence and Detection

Further Reading

Anonymous (1987) *Anabolic, Anthelmintic and Antimicrobial Agents. The 22nd Report of the Steering Group on Food Surveillance. The Working Party on Veterinary Residues in Animal Products*. London: Her Majesty's Stationery Office.

Booth NH and McDonald LE (eds) (1988) *Veterinary Pharmacology and Therapeutics*, 6th edn. Iowa City, IA: Iowa State University Press.

Brander GC, Pugh DM, Bywater RJ and Jenkins WL (1991) *Veterinary Applied Pharmacology and Therapeutics*, 5th edn. London: Ballière Tindall.

Commission of the European Communities (1989) *The Rules Governing Medicinal Products in the European Community*, vol. V. Luxembourg: Office for Official Publications of the European Communities.

European Federation of Animal Health (Seminar) (1989) Rationale view of antimicrobial residues. An assessment of human safety. *Advances in Veterinary Medicine, supplement to the Journal of Veterinary Medicine* No.42.

International Programme on Chemical Safety (1987) *Principles for the Safety Assessment of Food Additives and Contaminants in Food. Environmental Health Criteria 70*. Geneva: World Health Organization.

Rico AG (ed.) (1986) *Drug Residues in Animals*. London: Academic Press.

Woodward KN (1991) The licensing of veterinary medicinal products in the United Kingdom – the work of the Veterinary Medicines Directorate. *Biologist* 38: 105–108.

Residue Determination

A A Bergwerff and J Schloesser, Utrecht University, Utrecht, The Netherlands

Residue Determination

Farming of husbandry animals, as well as aquatic animals, is associated with veterinary application of pharmaceuticals, with the aim of influencing growth, weight gain, fat/protein distribution, milk production, stress, microbial infections, inflammation reactions etc. (*See* **Antibiotics and Drugs: Uses in Food Production**). This administration of pharmaceuticals results in so-called residues in the treated animal, including its edible parts, and products such as muscle (meat), organs, eggs, honey, and milk.

In some cases, residues are persistent, and concentrations may be high enough to affect the health or well-being of the consumer. In the case of direct carcinogenic substances, in theory, a single molecule may cause malignancies in the consumer. The veterinary use of such substances, when xenobiotic, is of course unacceptable. Veterinary medicinal products that have the ability to accumulate in the animal are banned world-wide as well. As well as the risk to public health, the production or quality of food may be affected by residues, for example by inhibition of fermentation processes during cheese production.

This phenomenon of negative effects of lagging drugs is known as the 'residue problem.' Controlling this problem through residue determination in the food chain from 'stable to table' is necessary to guarantee food quality and safety and to secure the consumer's trust.

Test at Maximum Residue Limit

In the case of registered substances, and thus substances allowed for veterinary use, international, national, and supranational legal bodies have agreed on so-called 'maximum residue limits' (MRL) for each edible product, and for each husbandry species. The consumer may be exposed to a nonacceptable health risk when this MRL is offended, i.e., the concentration of the indicator molecule is in excess of this concentration. This makes statutory and nonstatutory control of residues in food of major importance. MRL values may differ from one (inter)national authority to another, as consensus is not always reached on quantitative public health risks.

Authorities have also introduced a definition for residues of veterinary medicinal drugs, which, according to EU council regulation 2377/90, is:

> All pharmacologically active substances whether active principles, excipients or degradation products, and their metabolites which remain in food stuffs obtained from animals to which the veterinary medicinal product in question has been administered.

This description clearly does not include contamination arising from environmental substances, phyto-, phyco-, or mycotoxins, or as a result of food handling, processing, and preparation. The description highlights the environmental influences on the original molecules administered to the animal. The chemical appearance of a residue in muscle (*ante mortem*) or in meat (*post mortem*) is not static. In addition to biochemical processes in the animal, storage, food processing, and food preparation, which may include fermentation and heating, frequently initiate reactions converting the original pharmacon into derivatives. The residue analyst has to account for these processes, since they may mask the presence of residues.

Relatively soon after administration, furazolidone, for example, is metabolized completely and occurs as a so-called bound residue, which is a covalent linkage of the residue molecule with predominantly proteins. Bound residues are, by definition, not extractable. The remaining and nonreacting 3-amino-2-oxazolidone fragment is extractable and is used for furazolidone residue determination. The cephalosporin ceftiofur is converted into a reactive desfuroyl ceftiofur species, which reacts with itself, proteins, and other biomolecules. The residue is released chemically and stabilized prior to residue analysis. Another example is that of the antibiotic sulfadimidine, which may occur in the presence of relatively high concentrations of glucose, as in sausages, as an extractable sulfadiminyl glucose residue. In the living animal, many veterinary drugs are sulfated or conjugated to glucuronic acid in order to detoxify and excrete these substances. Incubation with sulfatases and glucuronidases prior to residue determination may be necessary to obtain reliable results.

Authorities have to decide on the indicator residue molecule for analysis. The MRL value of this indicator molecule should correspond with the total concentration of residue components, including derivatives of *ante-* and *post-mortem* reactions.

Analysis Strategies

Before going into a detailed description of residue analysis techniques, it should be noted that two different situations occur, which dramatically influence the desired degree of reliability of an analytical method.

In the situation of registration of pharmaceuticals, methods are designed and validated in well-defined experiments. A defined number of healthy animals are treated, allowing statistical analysis and averaging of errors. Target tissues may be sampled, and these usually contain easily detectable amounts of residues. In contrast, in an inspection or a monitoring situation, a juridically useful judgment should be made on the basis of a single sample for which the history is generally unknown and the concentrations are usually relatively low. In most cases, meat is sampled, which is mostly not the ideal matrix to trace the (illegal) use of pharmaceuticals. Furthermore, in contrast to the 'registration methods', time pressure is relatively high in the field of 'forensics' especially in the case of inspection of carcasses in the slaughter line. Swine are slaughtered at a rate of up to 12 carcasses per minute, whereas chickens are slaughtered at a rate of up to 270 exemplars each minute. The storage capacity becomes critical when a decision on condemnation of a carcass has to be made. The quality criteria for analytical methods in these different situations may therefore vary considerably.

The situation and analysis strategy therefore determine the characteristics of the assay of choice. The ideal method, which is inexpensive, fast, specific, selective, sensitive and allows a high throughput, does not exist. Compromises have to be made. A low-cost screening method with a high throughput may be followed by a usually more expensive, more selective, more sensitive, and specific confirmatory method. In arbitrage situations, a reconfirmation may be needed through so-called reference methods involving expensive instrumental techniques requiring highly skilled personnel.

Inventory Expected Residue Levels

Of the enormous collection of veterinary drugs, antibiotics are used most frequently and in the largest amounts compared with the other groups listed in Table 1. In the EU, a considerable part of antibiotic use is still for growth promotion. From the point of view of residue analysis strategies, it is important to consider the use of a drug before developing or introducing a new method in the residue laboratory. The different levels at which veterinary drugs are applied will result usually in proportional concentrations in edible tissues. For that reason, legal antibiotic growth promoters do not result in high levels of residues. Tracing the use of illegal antibiotic growth promoters in animal products may be more worthwhile, but will require sensitive tests amenable for detecting relatively low concentrations. Therapeutic use of antibiotics, however, may yield residues at concentrations of milligrams per kilogram, when withdrawal times are not obeyed, or as high as grams per kilogram at the injection site. The withdrawal time is the prescribed interval between administration and allowance to bring products of a treated animal to the market.

Besides antibiotics, the occurrence of substances with hormonal activity is attracting much attention, at least in the EU. Since steroid hormones are effective at very low levels, tracing residues in meat needs ultrasensitive assays for detecting concentrations as low as the subnanogram per kilogram level. These methods also may need to be able to differentiate between endogenous and administered, exogenous hormones, as in the case of nortestosterone of 17β-estradiol, for example. In these situations, it is difficult to answer the questions 'when is a residue a residue, and when is a residue not a residue?'

Monitoring and inspection programs should be aware of the use of 'cocktails' of illegal hormonal compounds. The mixture will exert its biological effect, but the individual components may not be detected as a result of low concentrations per component. The illegal use of β-agonists may co-occur, for example, with corticoids and thyreostats. Residue analysis strategies may therefore focus on the most easily extractable and detectable residue in first screening.

Sampling

In setting up a method, a few steps can be recognized (Figure 1). Each step from sampling, extraction, and concentration to (instrumental) analysis is essential and should be carried out appropriately for successful analysis. The statement: 'collect the wrong sample, or collect the right sample incorrectly, and you will trivialize all that follows, rendering your data worthless' emphasizes the importance of the first step, namely sampling. The selection of the laboratory sample for analysis, its transport, and storage should always be carefully evaluated (cf. Table 2). The storage of a

Table 1 Classes and examples of veterinary drugs[a]

Activity/class	*Subclass*	*Examples*[b]
Antibiotics, including growth-promoters and chemotherapeutics	Nitrofurans	Furazolidine, nitrofurazone, furaltadone
	(Fluoro)quinolones	Enrofloxacin, oxolinic acid, flumequine
	Tetracyclines	Oxytetracycline, doxycycline, chlortetracycline
	Penicillins and cephalosporins	Amoxiciline, cloxacillin, penicillin G, ampicillin, ceftiofur
	Aminoglycosides	Gentamicin, lincomycin, streptomycin
	Sulfonamides	Sulfadimidine, sulphaquinoxaline
	Macrolides	Erythromycin A, spiramycin, tylosine
	Peptides	Virginiamycin, bacitracin
	Miscellaneous	Chloramphenicol, carbadox, salinomycin
Anti-ectoparasitica	Phosphoesters, carbamates, pyrethroids a.o.	Dichlorvos, carbaryl, deltamethrine, lufenurone, cyromazin, lindane
Anthelmintics	Imidazothiazoles	Levamisole, dexamisol
	Benzimidazoles	Mebendazole, fenbendazole, lobendazole
	Tetrahydropyrimides	Pyrantel
	Avermectins	Ivermectin, abamectin, moxidectine
	Salicylanilides	Closantel, rafoxanide, niclosamide
Coccidiostats[c]		amprolium, dimitridazole, monensin,
Antimycotica		nystatine, griseofulvine, enilconazole
Hormones	Glucocorticosteroids	Dexamethasone, cortisone (hydrocortisone)
	Androgenic and estrogenic sex steroids	17β-Estradiol, methyltestosteron, trenbolone
	Progastegens	Megestrol, medroxyprogesteronacetate
	Resorcylic acid lactones	Zeranol, taleranol
	Stilbenes	Diethylstilboestrol, hexoestrol, dienoestrol
	β-agonists	Clenbuterol, salbutamol, bromobuterol
	Thyreostats	Tapazole, thiouracil
	Proteins	Somatropins
Tranquilizers		Azaperone, carazolol, xylazine

[a]This table is a summary and not a complete list of compounds and classes of compounds used in veterinary medicine.
[b]Some of these compounds are registered veterinary drugs. The EU-accepted MRL values in food from bees, cattle, deer, horses, pigs, poultry, salmonidae, sheep, etc. are published on the internet at http://eudrams1.is.eudra.org
[c]Many antibiotics are used as coccidiostats as well.

laboratory sample of homogenated or intact tissue may influence the final determined residue content dramatically. Furthermore, the test sample should reflect the laboratory sample, which usually means thorough homogenization of the latter. The laboratory sample should reflect the tissue, body fluid, or total animal that is sampled. At the same time, the result of an individual sample should not be overvalued. Unfortunately, many mistakes have been made at this level, sometimes with huge political impact.

When tracking the illegal use of prohibited substances, the choice for the most suitable matrix for analysis is obvious, since the MRL in an edible product does not have to be tested. In such cases, urine (anabolics), feces (steroid hormones), bile (antibiotics), retina (β-agonists), or hair (β-agonists), instead of meat, may be most suitable.

Extraction and Clean-up

Milk and diluted honey, but also urine, serum used to trace contamination of the animal, may be analyzed in some cases without extensive sample processing (Figure 1). Solid matrices, including fat, hair, meat, kidney, retina, and skin, however necessitate extraction of residues prior to analysis. Specific solvents are mixed with the homogenized matrix to facilitate solubilization of the residues. As an exception, the collection of meat drip through heating or compressing, or of pre-urine from the kidney cortex may be sufficient for microbiological growth-inhibition assays (see below).

Matrices are complex, and it should be noted that the residue analyst looks for a single particle in a heap of a million to billion other potentially interfering particles. These particles are also for a part solubilized and extracted from the matrix. Clean-up of samples is therefore necessary in most cases to overcome any interference of the analysis method.

A good knowledge of the chemical nature of the analyte in terms of stability, pK_a value(s), polarity, and functional groups will support the design of a selective, effective, and appropriate clean-up procedure. Before the availability of sorbent materials in the mid-1980s, sample clean-up was carried out by

liquid/liquid partition and/or atmospheric pressure chromatography. Sample concentration was obtained through rotary evaporation. A major improvement in this laborious work was obtained by the launch of commercially available, and thus standardized, solid-phase extraction (SPE) cartridges. The cartridges enabled selectivity and concentration in a single step with minimal sample and solvent consumption. The choice of stationary phases and supportive material in SPE is comparable with that used in high-performance liquid chromatography (HPLC), including normal- and reversed-phase, partition and ion-exchange chromatography. Combinations of phases within a single cartridge are available as well. For example, so-called 'restricted access materials', such as alkyl-diol-silica particles, combine the ability of gel filtration and solid-phase extraction within one material. These special SPE materials allow the extraction of aqueous extracts or liquid samples like milk or urine without prior deproteination. With the advent of chelating agents and the availability of antibodies and receptors, it is now possible to develop cartridges facilitating affinity extraction with a high selectivity.

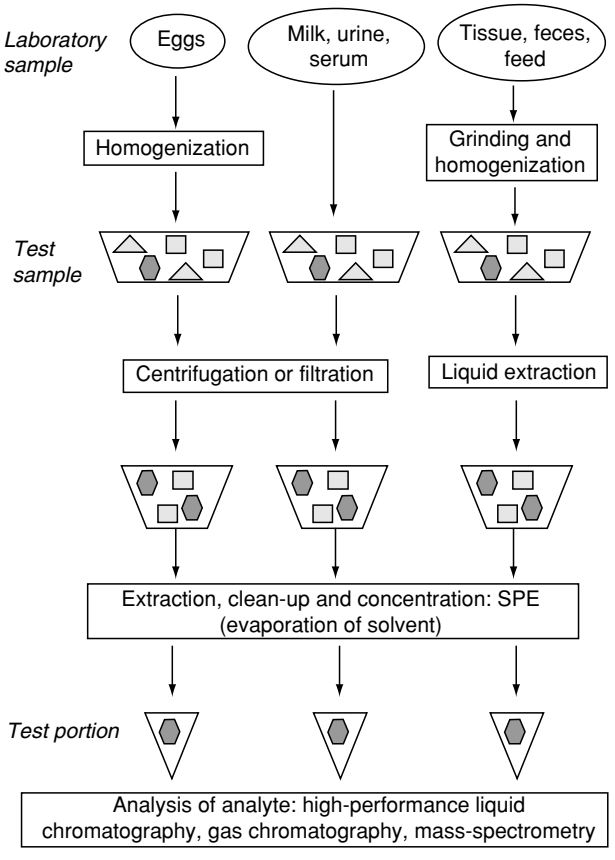

Figure 1 Typical scheme for sample processing involving SPE and determination of residues of veterinary drugs in different matrices.

More recently, disks consisting of membranes of polymers with high, reversed and ion-exchange phase capacity have been introduced. These disks are used as filters, allow fast-flow processing, purify and concentrate analytes like SPE cartridges with small to large sample volumes. Interfaces are available, enabling the coupling of such disks to HPLC columns.

Another sorbent extraction approach is matrix solid-phase dispersion, which allows homogenization and extraction by mixing of sorbent material in a single run omitting liquid extraction of the homogenate as a separate step. Columns prepared with the mixture are eluted in a similar way to SPE cartridges. Alternatives for these extraction procedures are supercritical fluid extraction and accelerated solvent extraction. The approaches are slowly gaining popularity but are still used in niche applications.

After extraction, the analyte-containing solution often has to be concentrated by removing excess solvent through evaporation or sublimation by applying a vaccum or a flow of an inert gas (nitrogen) sometimes in combination with heat. Loss of a part of the amount of the analyte can occur if unsuitable glass- or plastic ware is used. Many analytes, like penicillins and avermectins, tend to become adsorbed by free silanol groups of glass surfaces. Deactivation of free silanol groups through silanization of glassware

Table 2 Stability of veterinary drug residues in different matrices upon cold storage and heating

Pharmaceutical	*Storage at −18 °C*	*Storage at + 4 °C*	*Heated (Percentage activity recovered)*
Penicillin	Stable in muscle for 14 days	Stable in muscle for 14 days	10% in pork at 80 °C for 10 min
	Degraded in kidney	Degraded in kidney	0% in sterilized sausages at 125 °C for 65 min
Furazolidone	Unstable	Unstable	
Oxytetracycline	Stable for 6 months	Stable for 6 months	No recoverable activity in pork at 100 °C for 10 min
Sulfadimidine[a]	Degraded in liver	Degraded in liver	100% in well-done steak at 80 °C for 30 min
Chloramphenicol	Degraded in liver	Degraded in liver and kidney	37% in pork at 80 °C for 10 min
Spectinomycin	Degraded in liver and kidney	Degraded in liver and kidney	

[a]Sulfadimidine reacts with nitrite and glucose in sausages.

using chlorsilane often enhances the recovery in such situations.

Analysis of Analyte

Overviewing all available methods for separation and detection of residues in the nonprocessed test sample or in the obtained extract, an enormous gamma of (immuno)chemical and physicochemical approaches can be disentangled (Table 3). Thin-layer chromatography (TLC) was applied frequently in the early days of residue analyses and was closely followed in time by gas chromatography (GC). Analytes on TLC plates were visualized by generating chromophores or fluorescents by applying (through spraying or dipping) derivatizing molecules. Thermal conductivity, flame ionization, and, in certain applications, electron-capture and nitrogen phosphorus (NPD), detectors were popular in GC analysis. In current residue GC methods, the universality, selectivity and specificity of the mass spectrometer (MS) in combination with electron-impact ionization (EI) is by far preferred.

HPLC, in combination with a range of different detectors, became popular in the 1980s and 1990s. The stationary phases used predominantly are reversed-phase materials, especially phenyl-hexyl, octyl- or dodecyl-modified silica gels, but also, partition and ion-exchange chromatography is found in residue-determination protocols. Highly end-capped silica gels or polymeric stationary phases may be necessary to prevent tailing of amphoteric residues, such as tetracyclines. Polymeric phases also allow elution under alkaline conditions without solving the supporting silica material.

In addition to adequate extraction, clean-up and chromatography, the choice of detection system is an important parameter to achieve selectivity and specificity. Analytes, which are not detected by light absorption, refractive index, or fluorescence at the level they are expected to occur in animal-derived matrices, may require chemical modifications rendering them as fluorescents, chromophores, or UV-light absorbing compounds. For example, spectinomycin is converted postcolumn into a fluorescent derivative following oxidation and reaction of two *o*-phthaldialdehyde molecules with the two primary amines into conjugated *iso*-indoles. In such cases, chemical modification may also introduce a tremendous improvement in the selectivity of the method.

Alternative detection systems compatible with LC that have been used in residue analyses comprise evaporative laser-scattering and electrochemical detection, including (pulsed) amperometric detection.

In electrochromatography, electroosmotic flow is used for mobile phase delivery. Packed and open tubular columns are used in these systems, giving very small band widths, since extremely small particles can be used as the stationary phase. The development of heat is, however, a major drawback of this separation technique. Instead of chromatography, capillary electrophoresis with a high resolving power may be considered as well. However, the maximum sample size, which is very small, limits its use in residue analysis. To obtain namely dissolved extracts, relatively large solvent volumes are required, which reduce residue concentrations.

The identification of analytes in TLC, GC, and LC separation systems is carried out by comparison of the migration or elution positions with those of standard substances under identical chromatographic and detection conditions. However, comigration or coelution is usually not considered as a proof of identity of the analyte, especially if juridical consequences are involved. For this reason, combination of chromatography with spectrometric (MS) or spectroscopic detection techniques (NMR, IR) may provide data with

Table 3 Examples of residue analysis approaches

Analyte	*Matrix type*	*Analytical method*	*Sample pretreatment*
Penicillins (group-specific method)	Meat	GC-NPD or GC-EI-MS	Acetonitrile/buffer extraction, multiple liquid–liquid partition ion-exchange-SPE, elution with sodium chloride, liquid/liquid partition, derivatization
Penicillins	Milk	HPLC, photochemical postcolumn derivatization, UV detection at 300 nm or electrochemical detection	Centrifugation and ultrafiltration, on-line SPE (restricted-access material)
Avermectins	Liver	HPLC, fluorescence detection (λ^{ex}: 365 nm λ^{ex} 470 nm)	Acetone/water extraction, liquid–liquid partition, SPE, concentration, precolumn derivatization
Quinolones	Poultry meat	HPLC-APCI-MS/MS	Phosphate-buffer extraction, SPE
Tetracyclines	Kidney	HPLC-ESI-MS/MS	McIlvain-buffer extraction, on-line SPE
Sulfamethazine, Sulfdaizine	Meat, milk	Biosensor (surface plasmon resonance)	Homogenization with buffer, centrifugation

sufficient evidence. Of the spectroscopic approaches, nuclear magnetic resonance (NMR) and Fourier-transform infrared (FTIR) spectroscopy are most suitable, although certainly not in routine settings. The so-called hyphenated techniques, like in TLC-MS, GC-MS, LC-MS, LC-NMR, and GC-FTIR, have proved very powerful techniques for separation and molecular identification of the residue.

The evolution, and sometimes revolution, of mass selectors, ionization techniques, and their interfacing, has made the mass spectrometer standard equipment in a residue analysis laboratory. Mass selectors comprise sector, quadropole, ion-trap, time-of-flight instruments, and combinations thereof. In particular, mild-ionization techniques, such as atmospheric pressure chemical ionization (APCI) and electrospray ionization (ESI), facilitate MS analysis of small to relatively large and hydrophobic to hydrophilic molecules. APCI and ESI are therefore very suitable for residue determination. Using mass spectrometer-induced molecular fragmentation, it is possible to elucidate the chemical structure of the residue. A major pitfall is matrix-caused interference, which may reduce the response of the analyte to nil.

Besides physicochemical methods, the use of microbiological growth-inhibition assays to test meat and milk for the presence of antibiotics is popular over a long period of time. In terms of the number of analyses performed each year, this type of residue determination is most important. In addition to the low cost, this may be a result of an early implementation of such tests in national legislation. Examples of such methods are the EU Four Plate Test, the New Dutch Kidney Test, and the Delvotest SP, which use very antibiotic-sensitive bacterial reporter strains, such as *Bacillus subtilis* and *Bacillus stearothermophilus* var. *calidolactis*. These bacteria are inoculated under optimal conditions with and without sample. After culturing, results are read from visible inhibition zones or from the color change of the bacterial suspension in agar gels.

The availability of relatively large amounts of immunoglobulins in the 1980s, of which the production could be assured, enabled the development of immunoassays, like radio-immunoassays, enzyme-immunoassays, enzyme-linked immunosorbent assays, and later strip- and dip stick-based immunoassays. Specific enzymes and receptors have been used in a similar way in such analysis approaches, which are referred to as receptor assays, such as the commercially available CHARM Test II and Penzyme III. The major advantage of immunoglobulins and receptors is their high specificity, which reduces the sample clean-up to a minimum. However, their high specificity is a disadvantage in multiresidue methods.

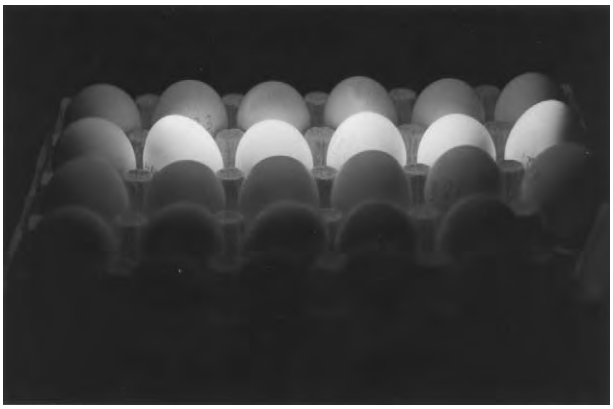

Figure 2 Eggs examined under UV light. The white/yellow fluorescence of the eggs in the middle is a result of tetracycline residues complexed with calcium in the eggshell. Positive eggs were collected from tetracycline treated laying hens. Courtesy of Dr. Michael Petz and Dr. Georg Zurhelle, University of Wuppertal, Germany.

In immuno- and receptor assays, the binding of small-molecular-weight compounds to their corresponding receptors or immunoglobulins is not directly detectable. These assays are therefore mostly competitive assays in which the analyte competes with a radio-isotope- or enzyme-labeled reporter compound for a limited number of binding sites. In contrast to conventional immuno- and receptor assays, advanced biosensor techniques do not require this labeling in most cases. From the point of view of signal transduction, biosensors fall into four categories, namely mass, optical, electrical, and thermal. Of the few applications so far, the 'surface plasmon resonance' optical biosensor has shown benefits in screening residues, including ciprofloxacin, clenbuterol, enrofloxacin, gentamycin, streptamycin, sulfadiazine, and sulfamethazine.

The ability of tetracyclines to fluoresce and to form complexes with alkali metals has revealed its presence in many cases (Figure 2). The article entitled 'The dog with the luminous bone' in the *Daily Telegraph* several years ago is certainly an example of a tetracycline contamination of bone tissue.

Quality Assurance

The validation of a developed method, but also that of an existing method newly introduced in a lab, is of absolute importance to assure reliable outcomes. In fact, international trading requiring certification of foodstuffs dictates that the performance of analytical methods be supported by validation data. The topics that have to be assessed are limit of detection and quantification, specificity, selectivity, accuracy,

precision, reproducibility, repeatability, etc. The robustness of the method should also attract attention. Critical points that may influence the measurement like minor changes in pH, temperature, SPE flow rate, the use of different SPE batches or of different producers, etc. therefore have to be evaluated as well.

In analysis strategies, the predictive value of the method should be considered. The false-negative rate of validated screening methods should be less than 5% (β-error) at the level of interest, whereas up to 20% false positive outcomes are not appreciated but are generally accepted. In contrast, false-positive results are not accepted in confirmatory methods, which usually carry a higher risk with systemic errors and interference. Criteria for the positive identification and quantification of residues are increasingly gaining attention. The suitability of confirmatory and/or reference methods for registered veterinary drugs, as proposed, is given in Table 4. In the case of banned substances, methods have to be based on molecular spectrometry, and the number of so-called 'identification points' (Table 5) should be at least four for positive identification.

Besides validation of methods, quality-assurance systems, such as good laboratory practice (GLP) are also essential for residue laboratories. Such systems not only ensure clarity on the performance of the method, but also increase the traceability of the results. Furthermore, GLP or similar systems assure appropriate laboratory management and properly trained analysts.

Table 4 Suitability of methods for confirmatory analysis of nonbanned veterinary medicinal products

Technique	*Criterion*
LC or GC in combination with MS	Suitable (at least three identification points)[a]
LC-diode array detector	Suitable
Two-dimensional TLC-spectrophotometry	Suitable
GC-electron capture detector	Suitable, only if combined with other methods
LC-immunogram	Suitable, in combination with other techniques
LC-UV/VIS (single wavelength)	Suitable, in combination with other techniques

[a]See **Table 5** for clarification of identification points. Adapted from the pending revised EU commission Decision 93/256/EC.

Table 5 Awarding mass-spectrometric approaches for confirmatory analysis of banned veterinary products with so-called identification points[a]

MS technique	*Identification points awarded per ion*
Low-resolution (LR) MS	1.0
LR-MS^n precursor ion[b]	1.0
LR-MS^n transition products	1.5
High resolution (HR) MS	2.0
HR-MS^n precursor ion	2.0
HR-MS^n transition products	2.5

[a]At least four identification points are necessary for confirmation of banned veterinary drugs. Adapted from the pending revised EU Commission Decision 93/256/EC.
[b]MS^n spectra are generated with tandem MS machines (MS^2) or with ion-trap mass selectors.

Outlook in Residue Analysis

'Multianalyte', i.e., testing various residues in a single run with a single instrument, is a magical word in the field of residue analysis. It requires generic extraction methods and highly resolving and robust analytical techniques. The fulfillment of this ideal still seems to be far away. Computerization and miniaturization and the introduction of SPE cartridges already have revolutionized residue analysis in the past. Advances in computerization and miniaturization will certainly increase sensitivity and selectivity further. This and further technical improvements in instrumental techniques will make multianalyte analyses possible to an increasing degree. Automation and analysis within minutes and even seconds, but also the type of matrices (body fluids and excreta), will transform future residue analysis labs into clinical chemistry labs.

Recent scandals have shown that residue determination at the slaughter-line stage is not effectual in protecting public health. The need for high analysis frequencies to guarantee consumers' trust is obvious. However, it is impossible to test every product for all possible contaminations. Analysis should be performed in the food chain, starting in the stable and at the feed mill. Laboratory samples should be collected not only from animals, but also from the feed and drinking water in order to trace cross- or other accidental contamination. Furthermore, inspection systems will not focus on the individual animal anymore, but on the complete herd or flock. Analysis of a number easily collectable samples will need to reflect the complete herd and predict the level of contamination of the animals, so that its products can be processed and placed on the market accordingly. At the same time, future veterinary control systems are being developed that will monitor residue levels to be in compliance with withdrawal times and MRL values. The residue lab may therefore be (at least in part) situated in the slaughterhouse or located in a vehicle visiting farms, e.g., for certification purposes. In particular, biosensor systems are attracting considerable attention for such targets. (*See* **Antibiotics and Drugs**: Uses in Food Production.)

See also: **Antibiotics and Drugs**: Uses in Food Production; **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; Combined Chromatography and Mass Spectrometry; **Contamination of Food**; **Quality Assurance and Quality Control**; **Spectroscopy**: Overview; Infrared and Raman; Near-infrared; Fluorescence; Atomic Emission and Absorption; Nuclear Magnetic Resonance

Further Reading

Commission of the European Communities, Directorate General SANCO (2002) *Commission Decision of Implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results*. Draft document SANCO/1085/2000 Rev. 7.

Funk W, Dammann V and Donnevert G (1995) *Quality Assurance in Analytical Chemistry*. Weinheim: Wiley-VCH.

Haagsma N and Ruiter A (eds) (1996) *Residues of Veterinary Drugs in Food*. Proceedings of the EuroResidue III Conference, Veldhoven, May 3–6. Utrecht: Utrecht University.

Horwitz W (ed.) (2000) *Official Methods of Analysis of AOAC INTERNATIONAL*, 17th edn. Arlington, VA: AOAC International.

Nollet LML (ed.) (1996) Residues and other food component analysis. In: *Handbook of Food Analysis* (vol. 2) New York: Marcel Dekker.

Oka H, Nakazawa H, Harada K and MacNeil JD (eds) (1995) *Chemical Analysis of Antibiotics Used in Agriculture*, p. 452. Arlington, VA: AOAC International.

O'Keeffe M (ed.) (2000) *Residue Analysis in Food – Principles and Applications*. Singapore: Harwood Academic Publishers.

Stephany RW (1989) Molecular spectroscopy in forensic residue analysis – a general overview of variability, applicability and cost-effectiveness. *Journal of Chromatography* 489: 3–9.

Van Ginkel LA and Ruiter A (eds) (2000) *Residues of Veterinary Drugs in Food*. Proceedings of the Euro-Residue IV Conference, Veldhoven, May 8–10. Bilthoven: National Institute of Public Health and the Environment.

ANTIOXIDANTS

Contents

Natural Antioxidants

M H Gordon, The University of Reading, Reading, UK

Background

Natural antioxidants have been of interest for many years because of their ability to retard the development of off-flavors in foods. However, an upsurge in interest in these components has occurred in recent years because of their importance for the prevention of diseases mediated by free radical reactions *in vivo*. The onset of a variety of major health problems, including cancer, atherosclerosis, rheumatoid arthritis, inflammatory bowel disease, immune system decline, brain dysfunction, cataracts, and malaria may be delayed by natural antioxidants.

Natural antioxidants are primarily plant phenolic compounds which may occur in all parts of the plant. Nonphenolics including carotenoids and phospholipids may also show antioxidant activity under some conditions. Plant phenolics are multifunctional. They can act as radical scavengers, metal chelators, singlet oxygen quenchers, or reducing agents. Though they are considered to be safe at normal levels of consumption they have not been toxicologically tested in many cases and therefore the possibility of mutagenic or carcinogenic activity cannot be discounted for some natural phenolics at unnaturally high levels of consumption. Some antioxidants add color, aftertaste, or off-flavor to a product and this may restrict their use in some types of foods.

Natural antioxidants can be found in a wide range of food raw materials. Tocopherols and related compounds occur at levels of about 50–1000 mg kg^{-1} in vegetable oils. Animal fats contain very low levels (e.g., butter oil 20–50 mg kg^{-1}), but cereal germ oils are very rich sources (1500–5000 mg kg^{-1}).

Tocopherols are added to some foods as antioxidants. They occur with various degrees of methylation of the dihydrochromanol ring as α, β, γ and δ-tocopherol. They have the chemical structure shown in Figure 1.

The activity of the tocopherols depends on their concentration, the food system, and the presence of heavy metals since they can act as prooxidants with iron or copper. Tocopherols are strongly effective as antioxidants in emulsions, but they are less effective in oils. Their antioxidant activity under common test conditions in oils decreases from δ-tocopherol> γ-tocopherol > β-tocopherol > α-tocopherol. They are most effective in the presence of a synergist like ascorbic or citric acid. In biological systems, however, α-tocopherol or vitamin E is the most active antioxidant with a strong antioxidant effect in biological membranes. Tocopherols are commercially produced and they are used as food additives. (*See* **Tocopherols**: Properties and Determination.)

About 50 carotenoids can be found in the human diet as minor components of fruit and vegetables. Carotenoids, including β-carotene (Figure 2), demonstrate antioxidant properties under certain conditions. Due to their multiple conjugated double-bond system, the carotenoids act as singlet oxygen quenchers during photosynthesis. They are also able to scavenge free radicals and act as antioxidants at low partial pressure of oxygen.

There is considerable interest in the nutritional properties of carotenoids. Besides the provitamin A property of some of the carotenoids, the antioxidant properties may also be of nutritional significance.

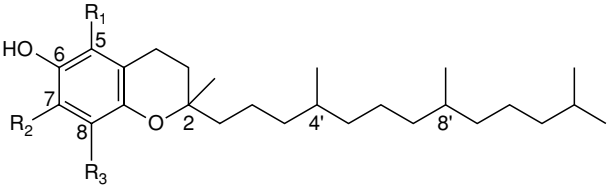

5,7,8-Trimethyltocol (α-tocopherol)	$R_1,R_2,R_3=CH_3$
7,8-Dimethyltocol (β-tocopherol)	$R_1=H$, $R_2,R_3=CH_3$
5,8-Dimethyltocol (γ-tocopherol)	$R_1=CH_3$, $R_2=H$, $R_3=CH_3$
8-Methyltocol (δ-tocopherol)	$R_1,R_2=H$, $R_3=CH_3$

Figure 1 Tocopherol structure.

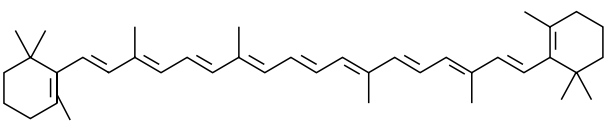

Figure 2 Structure of β-carotene.

Epidemiological studies have found that consumption of fruit and vegetables, which contain carotenoids, is associated with a reduced risk for several chronic diseases, including cardiovascular and photosensitivity diseases, cataracts, and some cancers. However, the results from large controlled trials involving β-carotene supplementation do not always support these beneficial associations, providing in some cases evidence for adverse effects, including an increase in the incidence of lung cancer and overall mortality in smokers.

It has been suggested that other carotenoids, including α-carotene, lycopene, lutein, and zeaxanthin, may show a higher potency than β-carotene in the suppression of the development of various types of cancer, including lung, breast, and stomach cancer. Dietary consumption of tomato products containing lycopene is associated with reduced risk of cancer and coronary heart disease. Although the antioxidant properties of lycopene may be responsible for its beneficial properties, other mechanisms, including effects on intercellular gap junction communication, or on the hormonal or immune system, may be partly responsible. (*See* **Carotenoids**: Occurrence, Properties, and Determination.)

Flavonoids are an important class of natural antioxidants that occur in a wide variety of fruit, vegetables, leaves, and flowers. They are found mainly as glycosides and methylated derivatives. The most important subgroups are the colorless catechins, the red-purple-colored anthocyanidins, the yellow flavonols and flavones, as well as the colorless proanthocyanidins.

The flavonoids share a common C_6-C_3-C_6 skeleton. The flavonoid aglycone normally consists of an aromatic ring (A) condensed with a six-membered oxygen-containing ring (C) which carries a phenyl substituent in the 2-position (B). The six-membered ring condensed with the benzene ring is either a γ-pyrone ring (flavonols and flavones) or its dihydro derivatives (flavanols and flavanones).

Flavonoid Structure (Figure 3)

The position of the attachment of the B-ring to the central C-ring divides the flavonoid class into

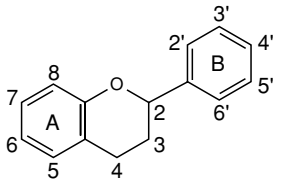

Figure 3 Flavonoid structure.

flavonoids (2-position) and isoflavonoids (3-position). Flavonols including quercetin and myricetin are important natural antioxidants. Quercetin occurs as the glycoside quercetrin in onions, apple skins, and black tea. Anthocyanidins differ from the other flavonoids in having a charged oxygen in the C-ring. The C-ring is open in the chalcones. Flavonoids are often hydroxylated in position C-3,5,7,3′,4′ or 5′.

Flavonol and Flavone Structure (Figure 4)

Attention was focused on the nutritional significance of dietary flavonols and flavones by the Zutphen Elderly Study in The Netherlands. The study started in 1960 with 878 men aged 40–59 years and continued until 1985 with 555 of the original subjects and 250 replacements aged 65–84 years. Nutritional intake was studied by cross-checked dietary history for 5 years. The mean baseline of flavonoid intake was 25.9 mg day^{-1} (as aglycones). The major flavonoids in the food were quercetin, myricetin, kaempferol, apigenin, and luteolin. Their sources of intake were mainly tea (61%), onions (13%), and apples (10%).

After adjustment for age, body mass index, smoking, serum total and high-density lipoprotein cholesterol, blood pressure, physical activity, coffee consumption, and intake of other antioxidants and fiber, the flavonoid intake was significantly inversely associated with mortality from coronary heart disease and showed an inverse relationship with the incidence of myocardial infarction.

In Mediterranean countries, wine is an important source of flavonoids (10–20 mg l^{-1} combined flavonoids). This may partly contribute to the low risk of coronary heart disease of wine drinkers who consume a high level of saturated fats in France, the so-called French paradox.

Tea catechins are very strong antioxidants. For the tea catechins, stronger activity than BHA or α-tocopherol in lard autooxidation systems was shown with activity increasing in the order: (−)-epicatechin < (−)-epicatechin gallate < (−)-epigallocatechin < (−)-epigallocatechin gallate (Figures 5–8).

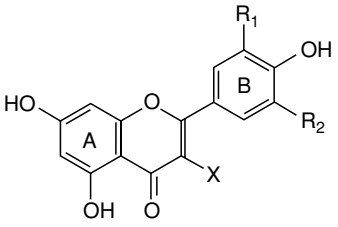

Flavonol aglycone:	X=OH	quercetin R_1=OH, R_2=H myricetin R_1,R_2=OH
Flavonol:	X=glucose	quercitrin R_1=OH, R_2=H
Flavone aglycone:	X=H	luteolin R_1=OH, R_2=H

Figure 4 Flavonol and flavone structure.

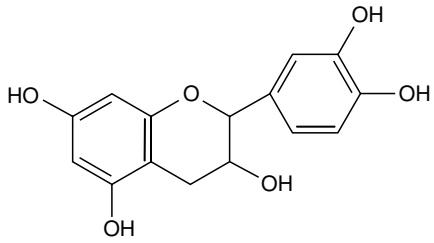

Figure 5 Epicatechin.

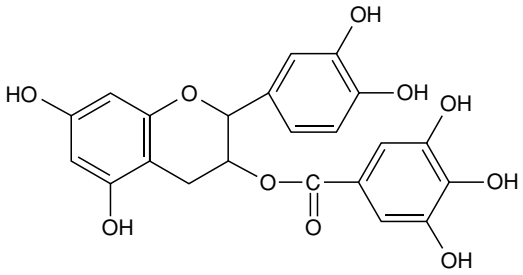

Figure 6 Epicatechin gallate.

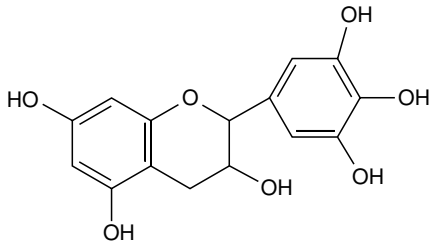

Figure 7 Epigallocatechin.

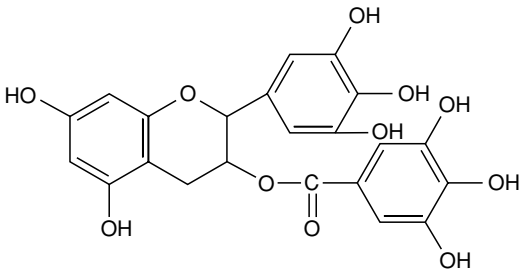

Figure 8 Epigallocatechin gallate.

Isoflavonoids are a subclass of flavonoids which are of interest for their estrogenic activity. The main isoflavones are daidzain and genistein, which occur in soya beans, chickpeas, and lentils. The normal western diet contains much smaller amounts of isoflavones than the Chinese or Japanese diet, due to the low level of consumption of soy products in the west. Epidemiological studies suggest that foods containing phytoestrogens may have a beneficial effect in protecting against several chronic and degenerative diseases, including cancer. Although antioxidant effects may be important, the estrogenic and antiestrogenic effects, induction of cancer cell differentiation, inhibition of tyrosine kinase and DNA topoisomerase activity, and suppression of angiogenesis may be more important.

The unsaponifiable fraction of sesame oil shows good antioxidant activity, and this is mainly attributed to sesamolin which can be hydrolyzed on heating to sesamol which is an active antioxidant (Figure 9).

Olive oil contains a mixture of antioxidants derived from hydroxytyrosol and other polyphenols. From comparisons with a limited number of phenolic constituents, it has been claimed that hydroxytyrosol is the most active antioxidant compound in virgin olive oil, although the concentration of free hydroxytyrosol is low. In the last few years, the antioxidant activity of this compound and its derivatives has been investigated *in vitro*. Oleuropein is an important derivative of hydroxytyrosol, which occurs in large amounts in olive leaves and olive fruits (up to 14% of the dry weight in unripe olives), but it appears to have a weaker antioxidant activity than free hydroxytyrosol in olive oil.

Phospholipids (e.g., commercial lecithin) can also exhibit antioxidant activity in heated oils, where they may act as hydroperoxide decomposers and/or metal chelators. However, they appear to be less active in oils stored at ambient temperature. Phospholipids are present in crude vegetable oils but are removed during oil refining. (*See* **Phospholipids**: Properties and Occurrence.)

Phenolic acids may also play important roles as natural antioxidants.

Hydroxyderivatives of cinnamic and benzoic acids can be found in many plant foods. In oats, esters of caffeic acid (Figure 10, e.g., chlorogenic acid) and ferulic acid seem to be the most important antioxidants. (*See* **Phenolic Compounds**.)

Many spices contain antioxidants, especially oregano, sage, thyme, clove, ginger, and rosemary. The main active compound of rosemary is carnesol (Figure 11).

Organic acids, such as citric and ascorbic acid and its esters, are often used as stabilizers for fats and oils. Citric acid (Figure 12) is a very important antioxidant because it is very effective at chelating traces of metals such as iron and copper that have a strongly detrimental effect on the shelf-life of edible oils. Citric acid is commonly added to oils at the end of the refining process. Ascorbic acid (Figure 13) has a weaker metal-chelating action, but it also removes oxygen and can contribute to the oxidative stability of foods when a limited amount of oxygen is available, e.g., in canned foods.

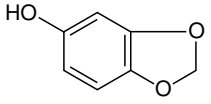

Figure 9 Sesamol.

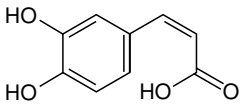

Figure 10 Caffeic acid.

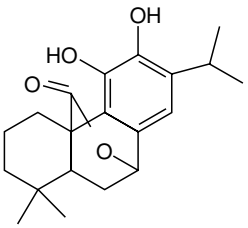

Figure 11 Carnosol.

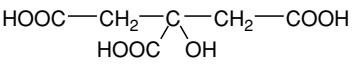

Figure 12 Citric acid.

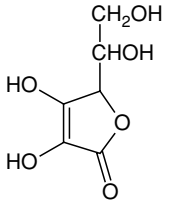

Figure 13 Vitamin C (ascorbic acid).

Amino acids are powerful synergists in combination with phenolic antioxidants, acting as scavengers of heavy metals and promoters of hydroperoxide decomposition. Protein hydrolysates have similar activity. Maillard reaction products (formed by the nonenzymic reaction of amino acids and reducing sugars) have been shown to have some activity, particularly by chelation of heavy metals. Because of their low solubility and color, they can only be added to certain foods. Low-molecular-weight products of carbonyls and amino acids produced by the Maillard reaction, such as dihydropyridines, have strong antioxidant activity. (*See* **Browning**: Toxicology of Nonenzymatic Browning.)

A wide range of other natural plant components have also been shown to have antioxidant activity, e.g., tannins and leaf waxes.

See also: **Amino Acids**: Properties and Occurrence; **Antioxidants**: Role of Antioxidant Nutrients in Defense Systems; **Carcinogens**: Carcinogenicity Tests; **Carotenoids**: Occurrence, Properties, and Determination; **Coronary Heart Disease**: Antioxidant Status; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Oxidation of Food Components**; **Olive Oil**; **Phenolic Compounds**; **Spices and Flavoring (Flavouring) Crops**: Properties and Analysis; **Tocopherols**: Properties and Determination; Physiology

Further Reading

Adlercreutz H and Mazur W (1997) Phyto-oestrogens and western diseases. The Finnish Medical Society. *Annals of Medicine* 29: 95–120.

Adlercreutz H, Goldin BR, Gorbach SL *et al.* (1995) Soybean phytoestrogen intake and cancer risk. *Journal of Nutrition* 125: 757S–770S.

Albanes D (1999) Beta carotene and lung cancer: a case study. *American Journal of Clinical Nutrition* 69: 1345–1350.

Baldioli M, Servili M, Peretti G and Montedoro GF (1996) Antioxidant activity of tocopherols and phenolic compounds of virgin olive oil. *Journal of the American Chemists' Society* 73: 1589–1593.

Burton GW and Ingold KU (1984) Beta-carotene – an unusual type of lipid antioxidant. *Science* 224: 569–573.

Frankel EN (1998) *Lipid Oxidation*. Dundee: Oily Press.

Frankel EN, Kanner J, German JB, Park E and Kiensella JE (1993) Inhibition of oxidation of human low-density lipoprotein by phenolic substances in red wine. *Lancet* 341: 454–457.

Gordon MH (1996) Dietary antioxidants in disease prevention. *Natural Product Reports* 13: 265–274.

Herrmann K (1976) Flavonols and flavones in food plants: a review. *Journal of Food Technology* 11: 433–448.

Hertog MGL, Fresken EJM, Hollman PCH, Katan MB and Kromhout D (1993) Dietary antioxidative flavonoids and risk of coronary heart disease: the Zutphen Elderly Study. *Lancet* 342: 1007–1011.

Hsieh RJ and Kinsella JE (1989) Antioxidant: flavonoids. *Advances in Food and Nutrition Research* 33: 233–341.

Hughes DA (1999) Effects of dietary antioxidants on the immune function of middle aged adults. *Proceedings of the Nutrition Society* 58: 79–84.

Khachik F, Beecher GR and Smith JC (1995) Lutein, lycopene, and their oxidative metabolites in chemoprevention of cancer. *Journal of Cellular Biochemistry* 22: 236–246.

Kritchevsky SB (1999) Beta-carotene, carotenoids and the prevention of coronary heart diseases. *Journal of Nutrition* 129: 5–8.

Loelinger J (1994) Natural antioxidants. In: Allen JC and Hamilton RJ (eds) *Rancidity in Foods*, 3rd edn, pp. 105–124. London: Blackie Academic and Professional.

Lunder TL (1992) Catechins of green tea. In: Huang M-T, Ho C-T and Lee CY (eds) *Phenolic Compounds in Food and their Effects on Health II*, ACS Symposium Series no. 507. Washington, DC: American Chemical Society.

Mayne ST (1996) Beta-carotene, carotenoids and disease prevention in humans. *FASEB Journal* 10: 690–701.

Rao AV and Agarwal S (2000) Role of antioxidant lycopene in cancer and heart disease. *Journal of the American College of Nutrition* 19: 563–569.

Rice-Evans CA, Miller NJ and Paganga G (1996) Structure–antioxidant activity relationships of flavonoids and phenolic acids. *Free Radical Biology and Medicine* 20: 933–956.

Rowe PM (1996) Beta-carotene takes a collective beating. *Lancet* 347: 249.

Synthetic Antioxidants

G Ramis-Ramos, Universidad de Valencia, Burjassot, Spain

Background

Lipid Oxidation

Lipid oxidation is a process that results in rancidity and deterioration of fats and progresses via free-radical propagated chain reactions. Free radicals are produced by the reaction of unsaturated fatty acids with molecular oxygen and traces of other oxidants, as metal ions (Fe^{3+} and Cu^{2+} are particularly effective) and H_2O_2, and by UV light. Primary lipid oxidation yields hydroperoxides that cause a variety of secondary reactions with evolution of aldehydes, ketones, acids, and other low-molecular-weight volatile substances. The ensuing decrease in the flavor and color quality, which renders a food unappetizing in

appearance, and the accumulation of toxic reaction products pose a major problem to the food industry. Natural and synthetic antioxidants and antioxidant synergists are substances of low toxicity that slow down lipid oxidation, and therefore, they are used to increase the stability of fat components of food under a variety of storage and cooking conditions.

Further, there is ample evidence showing that free radical oxidation plays an important role in many pathological situations, including many types of cancer, atherosclerosis, coronary heart disease, rheumatoid arthritis, and other degenerative diseases. Antioxidants are also employed in order to prevent and treat these pathologies. Prevention arises as a consequence of the inhibition of the formation of toxic and mutagenic substances during storing, cooking and in the stomach, and the modification of the basal levels of some enzymes, which results in increased detoxification and reduced rates of carcinogenic activation. Other benefits derived from the intake of antioxidants include the increased resistance to radiation effects, assistance in the treatment of gastroduodenal ulcer and craniocerebral injury (including meningitis effects), and help in slowing down the aging process.

Mode of Action and Functions of Antioxidants and Antioxidant Synergists

Some antioxidants are reducing substances that remove oxygen and other oxidants, thus hindering the initiation of free-radical reactions. Examples are sulfur dioxide, sulfites, and sodium nitrite. The antioxidant activity of tocopherols, alkyl gallates, and other phenol derivatives is mainly due to resonance stabilization of their free radicals. Stabilized radicals are not sufficiently reactive, and the chain reaction is terminated. The antioxidant activity of thiodipropionic acid and its esters relies on a rather different mechanism: these decompose hydroperoxides.

Antioxidant synergists are substances that prevent free-radical initiation and catalysis by complexing metal ions. When bound to a ligand with oxygen as the donor atom, the redox potential of metal ions as Fe^{3+} and Cu^{2+} is largely reduced, and their catalytic activity is seriously limited. Typical examples are the polycarboxylic substances, including citric and tartaric acids, and ethylenediaminetetraacetate salts. Some antioxidants, e.g., ascorbic acid and sodium thiosulfate, exhibit both reducing and metal ion complexing capabilities. Phenol derivatives are free-radical scavengers, but they are also effective in complexing iron, copper, and other metal ions.

When two antioxidants of a different type are jointly used, synergistic antioxidant activity is frequently achieved, although sometimes, antagonistic inhibition can also occur. For instance, polycarboxylic acids are often used in combination with oxygen-reactive antioxidants, such as ascorbates and erythorbates, and with free radical scavengers, as butylated hydroxyanisole (BHA) and butylated hydroxytoluene (BHT), to achieve synergistic food protection.

The substances classified by the Codex Alimentarius Commission (CAC) within the class 'antioxidant' or the subclass 'antioxidant synergist' are listed in Table 1. The order of the INS code (international numbering system for food additives) has been followed. As shown in the table, many antioxidants and antioxidant synergists also belong to other CAC classes and subclasses, whose meaning is indicated in Table 2. The subclasses are not subclassifications but descriptions of technological functions of the additive. The equivalent or similar definitions stated by the FDA regulation 170.3 (reg. of title 21, as all other FDA references within the text) are also given in this table.

It should be remembered that the use of antioxidants is not the only way to protect food against lipid oxidation. For instance, protection from light and packaging under an inert atmosphere, e.g., nitrogen gas (INS 941), using a gas-impermeable packing material, also constitutes a nontoxic and excellent solution.

Natural Versus Synthetic Antioxidants

It is customarily understood that a 'natural' food additive is extracted from a vegetal source (e.g., tartaric acid, which is a byproduct of the wine industry) or produced by microorganisms using substrates of vegetal origin (e.g., this is the case for more than 99% of the citric acid used in food), whereas a 'synthetic' food additive derives from the mineral oil industry or from other mineral resources.

However, this division is at least excessively simple, as can be shown with several examples. Thus, lactic acid is a natural product that is also produced in large amounts by a strongly competitive chemical process. Ascorbic acid, another natural substance, is almost exclusively produced by a process comprising several chemical steps assisted by a single microbiological transformation. The microbiological production of citric acid uses glucose as the substrate, but hydrocarbons can also be used. Finally, bleaching, neutralization, or esterification of the additives using reagents that may be of natural or synthetic origin is frequently performed to improve their technological properties.

The properties of food antioxidants obtained largely or exclusively from mineral resources, or with important concurrence of chemical procedures, are discussed here. Where applicable, emphasis is placed on the similarities and differences between the natural

Table 1 Substances classified by the CAC as food antioxidants or antioxidant synergists, with the exclusion of those of exclusive 'natural' origin

INS substance name (common acronym) (formula); selected synonyms	*CAC functional classes and subclasses*
220 Sulfur dioxide (SO_2); sulfurous oxide; sulfurous anhydride	Preservative, antioxidant
221 Sodium sulfite (Na_2SO_3)	Preservative, antioxidant
222 Sodium hydrogen sulfite ($NaHSO_3$); sodium bisulfite	Preservative, antioxidant
223 Sodium metabisulfite ($Na_2S_2O_5$); sodium disulfite; sodium pyrosulfite	Preservative, bleaching agent, antioxidant
224 Potassium metabisulfite ($K_2S_2O_5$); potassium disulfite; potassium pyrosulfite	Preservative, antioxidant
225 Potassium sulfite (K_2SO_3)	Preservative, antioxidant
226 Calcium sulfite ($CaSO_3$)	Preservative, antioxidant
227 Calcium hydrogen sulfite [$Ca(HSO_3)_2$]; calcium bisulfite	Preservative, antioxidant
228 Potassium hydrogen sulfite ($KHSO_3$); potassium bisulfite	Preservative, antioxidant
250 Sodium nitrite ($NaNO_2$)	Preservative, antioxidant
300 Ascorbic acid ($C_6H_8O_6$); L-(+)-ascorbic acid; vitamin C; antiscorbutic factor	Antioxidant
301 Sodium ascorbate ($NaC_6H_7O_6$); L-(+)-ascorbic acid sodium salt; vitamin C sodium salt	Antioxidant
302 Calcium ascorbate ($Ca\ C_{12}H_{14}O_{12}.2H_2O$); L-(+)-ascorbic acid calcium salt dihydrate	Antioxidant
303 Potassium ascorbate ($KC_6H_7O_6$); L-(+)-ascorbic acid potassium salt	Antioxidant
304 Ascorbyl palmitate ($C_{22}H_{38}O_7$); L-(+)-ascorbic acid 6-palmitate	Antioxidant
305 Ascorbyl stearate ($C_{24}H_{42}O_7$); L-(+)-ascorbic acid 6-stearate	Antioxidant
306 Tocopherol-rich extract (mixed tocopherol concentrate); vitamin E	Antioxidant
307 Synthetic α-tocopherol ($C_{29}H_{50}O_2$); 5,7,8-trimethyltocol	Antioxidant
308 Synthetic γ-tocopherol ($C_{28}H_{48}O_2$); 7,8-dimethyltocol	Antioxidant
309 Synthetic δ-tocopherol ($C_{27}H_{46}O_2$); 8-methyltocol	Antioxidant
310 Propyl gallate (PG) ($C_{10}H_{12}O_5$); *n*-propyl 3,4,5-trihydroxybenzoate; gallic acid propyl ester	Antioxidant
311 Octyl gallate ($C_{15}H_{22}O_5$); *n*-octyl 3,4,5-trihydroxybenzoate; gallic acid octyl ester	Antioxidant
312 Dodecyl gallate ($C_{19}H_{30}O_5$); *n*-dodecyl 3,4,5-trihydroxybenzoate; gallic acid dodecyl ester	Antioxidant
313 Ethyl gallate ($C_9H_{10}O_5$); ethyl 3,4,5-trihydroxybenzoate; gallic acid ethyl ester	Antioxidant
314 Guaiac resin; guaiacum; guaiac gum; gum guaiac	Antioxidant
315 Erythorbic acid ($C_6H_8O_6$); isoascorbic acid; D-araboascorbic acid	Antioxidant
316 Sodium erythorbate ($NaC_6H_7O_6$); sodium isoascorbate	Antioxidant
317 Potassium erythorbate ($KC_6H_7O_6$); potassium isoascorbate	Antioxidant
318 Calcium erythorbate ($CaC_{12}H_{14}O_{12}$); calcium isoascorbate	Antioxidant
319 *tert*-Butylhydroquinone (TBHQ) ($C_{10}H_{14}O_2$); 2-(1,1-dimethylethyl)-1,4-benzenediol	Antioxidant
320 Butylated hydroxyanisole (BHA) ($C_{11}H_{16}O_2$); 3- and 2-*tert*-butyl-4-methoxyphenol	Antioxidant
321 Butylated hydroxytoluene (BHT) ($C_{15}H_{24}O$); butylated hydroxytoluol; 2,6-di-*tert*-butyl-1-hydroxy-4-methylbenzene; 2,[6]-di-*tert*-butyl-*p*-cresol; 2,6-di-*tert*-butyl-4-methylphenol; 3,5-di-*tert*-butyl-4-hydroxytoluene; dibunol	Antioxidant
322 Lecithins; phosphatides; phospholipids	Emulsifier, antioxidant synergist
323 Anoxomer (polymerized mixture of phenolic antioxidants)	Antioxidant
324 Ethoxyquin (EQ, EMQ) ($C_{14}H_{19}NO$); 6-ethoxy-1,2-dihydro-2,2,4-trimethylquinoline	Antioxidant
325 Sodium lactate ($Na[CH_3CH(OH)COO]$); sodium 2-hydroxy-propanoate	Antioxidant synergist, humectant, bulking agent
326 Potassium lactate ($K[CH_3CH(OH)COO]$); potassium 2-hydroxy-propanoate	Antioxidant synergist, acidity regulator
334 Tartaric acid [HOOC-CH(OH)CH(OH)-COOH]; L(+)-2,3-dihydroxybutanedioic acid	Acidity regulator, sequestrant, antioxidant synergist
338 Orthophosphoric acid (H_3PO_4)	Acidity regulator, antioxidant synergist
339 Sodium orthophosphates (NaH_2PO_4, Na_2HPO_4, Na_3PO_4)	Acidity regulator, antioxidant synergist
340 Potassium orthophosphates (KH_2PO_4, K_2HPO_4, K_3PO_4)	Acidity regulator, antioxidant synergist
341 Calcium orthophosphates [$Ca(H_2PO_4)_2$, $CaHPO_4$, $Ca_3(PO_4)_2$]	Acidity regulator, antioxidant synergist
384 Isopropyl citrates; mixture of the citric acid esters of 2-propanol	Antioxidant, preservative, sequestrant
385 Calcium disodium ethylenediaminetetraacetate ($CaNa_2[(OOCCH_2)_2NCH_2CH_2N(CH_2COO)_2]$)	Antioxidant, preservative, sequestrant
386 Disodium ethylenediaminetetraacetate ($Na_2H_2\ [(OOCCH_2)_2NCH_2CH_2N(CH_2COO)_2]$)	Antioxidant, preservative synergist, sequestrant
387 Oxystearin; mixture of glycerides of partially oxidized stearic and other fatty acids	Antioxidant, sequestrant
388 Thiodipropionic acid ($S[(CH_2)_2COOH]_2$)	Antioxidant
389 Dilauryl thiodipropionate ($S[(CH_2)_2COO(CH_2)_{11}CH_3]_2$); didodecyl 3,3′-thiodipropionate; 3,3′-thiodipropionic acid di-*n*-dodecyl ester	Antioxidant
390 Distearyl thiodipropionate ($S[(CH_2)_2COO(CH_2)_{17}CH_3]_2$); dioctadecyl 3,3′-thiodipropionate; 3,3′-thiodipropionic acid di-*n*-octadecyl ester	Antioxidant

Continued

Table 1 Continued

INS substance name (common acronym) (formula); selected synonyms	*CAC functional classes and subclasses*
391 Phytic acid [$(H_2PO_4)_6C_6H_6$]; myo-inositol hexakis(dihydrogen phosphate)	Antioxidant
512 Stannous chloride ($SnCl_2$); tin(II) chloride	Antioxidant, colour-retention agent
539 Sodium thiosulfate ($Na_2S_2O_3$); sodium hyposulfite	Antioxidant, sequestrant
586 4-Hexylresorcinol ($C_6H_3(OH)_2(CH_2)_5CH_3$); 4-hexyl-1, 3-benzenediol; 4-hexyl-1,3-dihydroxybenzene; 4-hexylresorcine; 4-*n*-hexylresorcinol; ascaryl	Antioxidant, colour-retention agent
1102 Glucose oxidase	Antioxidant

Table 2 Meaning of the classes of **Table 1** according to the CAC and equivalent definitions of FDA (170.3)

Additive class	*CAC*	*FDA*
Acidity regulators (CAC) and pH control agents (FDA)	Alter or control the acidity or alkalinity of a food	Change or maintain active acidity or basicity, including buffers, acids, alkalies, and neutralizing agents
Antioxidants	Prolong the shelf-life of foods by protecting against deterioration caused by oxidants, such as fat rancidity and color changes	Preserve food by retarding deterioration, rancidity, or discoloration due to oxidation
Color-retention agents (CAC) and colors and coloring adjuncts (FDA)	Stabilize, retain or intensify the color of a food	Impart, preserve, or enhance the color or shading of a food, including color stabilizers, color fixatives, color-retention agents, etc.
Humectants	Prevent food from drying out by counteracting the effect of a wetting agent atmosphere having a low degree of humidity	Promote retention of moisture. Include moisture-retention agents and antidusting agents
Preservatives (CAC) and antimicrobial agents (FDA)	Prolong the shelf-life of foods by protecting against deterioration caused by microorganisms	Preserve food by preventing growth of microorganisms and subsequent spoilage, including fungistats, mold and rope inhibitors, and the effects listed by the National Academy of Sciences/ National Research Council under 'preservatives'
Sequestrants		Combine with polyvalent metal ions to form a soluble metal complex, to improve the quality and stability of products
Synergists		Act or react with another food ingredient to produce a total effect different or greater than the sum of the effects produced by the individual ingredients

and synthetic products. The reader should resort to section (a) 'Natural Antioxidants' to become familiar with antioxidants of exclusive vegetal or microbiological origin, or with complementary aspects of some of the substances discussed here.

Properties of Synthetic Antioxidants and Antioxidant Synergists

Sulfur Dioxide, Sulfites, Bisulfites and Metabisulfites

Sulfur dioxide is a colorless gas (a liquid below $-10\,^{\circ}C$ or under pressure) with a characteristic, pungent odor, more soluble in cold than in hot water. In moist environments, sulfurous acid, which reacts slowly with oxygen, is formed. This reaction is accelerated at increasing pH values. Sulfur dioxide gas is a corrosive and severe irritant. Some individuals are extremely sensitive to the effects of sulfur dioxide, whereas experienced workers may become adapted to its irritating properties, e.g., cases of acclimatization at 20–30 $\mu g\,g^{-1}$ have been observed. Exposure to high concentrations (e.g., 1000 $\mu g\,g^{-1}$) for several minutes produces respiratory paralysis and pulmonary edema with a risk of death. Liquid sulfur dioxide produces skin burns from the freezing effect of rapid evaporation.

Sodium, potassium and calcium sulfites, hydrogen sulfites (bisulfites), and metabisulfites exhibit almost identical preserving and antioxidant properties to sulfur dioxide. Thus, anhydrous sodium sulfite is relatively stable but reacts rapidly with oxygen in the presence of moisture; all dissolved oxygen is removed in a few seconds. Commercial sodium bisulfite consists mainly of sodium metabisulfite. Sodium and potassium metabisulfites are irritants and possible sensitizers.

At the levels used in food, sulfur dioxide, sulfites, bisulfites, and metabisulfites are generally recognized as safe (GRAS); however, a small segment of the population has been found to develop hives, nausea, diarrhea, shortness of breath, or even fatal shock after consuming these substances. For this reason, its use on fresh fruits and vegetables to be sold or served raw to consumers, or to be presented to consumers as fresh, was banned by the Food and Drug Administration (FDA) (182.3862, 3766, 3637, 3616). Further, these substances decompose vitamin B_1, and so they cannot be used in food recognized as a source of vitamin B_1.

Sulfur dioxide has been used in the manufacture of wine for thousands of years to destroy bacteria, molds, and wild yeasts without harming the yeasts used in fermentation. Sulfur dioxide, sulfites, bisulfites, and metabisulfites are used as antiseptics and bleaching agents in breweries and other fermentation industries, and are used to retard spoilage of meat, egg yolks, grains, dried and fresh fruits, and other products. Sodium sulfite is allowed as a bleaching agent for gelatin, beet sugars, and food starch (FDA 73.85), and sodium and potassium sulfites can be used to assist caramelization (FDA 73.85).

Sodium Nitrite

Sodium nitrite is an odorless, white or slightly yellow, hygroscopic and air-sensitive solid (it slowly oxidizes to nitrate), having both reducing and oxidizing properties. It presents a fire risk when in contact with reducing materials; for example, it is liable to render any organic matter (wood, paper and textiles) dangerously combustible when dry. It is decomposed by acids (even weak acids) with evolution of a brown mixture of nitrogen oxides (NO_x). It is very toxic and dangerous to the environment, and it is an irritant and harmful substance that may damage the cardiovascular and central nervous systems. (*See* **Nitrates and Nitrites.**)

Nitrates and nitrites are used in cured meat products, to control the growth of microorganisms, particularly to avoid the development of *Clostridium botulinum*, and also to serve as color and flavor fixatives. When added to meat, nitrate and nitrite react with the myoglobin and hemoglobin present in the trapped red blood cells, eventually producing nitric oxide myoglobin. After curing, almost all the nitrate and nitrite have reacted with the meat components, and only a very small residual amount of nitrite remains. The FDA allows the use of sodium nitrite in smoked cured fish products (from 10 to 200 $\mu g\,g^{-1}$) and in meat-curing preparations ($< 200\,\mu g\,g^{-1}$, mixed with sodium nitrate, FDA 172.175).

In the stomach, nitrite can produce dangerous carcinogenic nitrosamines; however, these reactions are inhibited in the presence of phenolic antioxidants, ascorbic acid, or other substances that are also added to food or are present in fruit and fresh vegetable juices that may be consumed with the cured meat. Furthermore, nitrite is naturally present in saliva, in concentrations higher than those found in cured meat products, and lettuce, spinach, beets, and many other vegetables contain nitrate, which is reduced in the mouth to nitrite by the action of bacteria. Therefore, the risks associated with the use of nitrate and nitrite in cured fish and meat products seem to be very small, whereas the benefits are large.

Ascorbic and Erythorbic Acids, and their Salts and Esters

Ascorbic acid is a colorless or white crystalline powder with a pleasant, sharp, acidic taste. When pure and dry, it is very stable, but in solution or when moisture is present, the acid and their salts react readily with oxygen. Oxidation is accelerated at both low and high pH values, and is catalyzed by metal ions, particularly for traces of copper ions. The optimum pH for stability in aqueous solution is 5.4.

Ascorbic acid occurs widely in animals and plants, where it constitutes the primary defensive barrier against oxidative stress. It is also required in many metabolic functions. Today, almost all the commercially available ascorbic acid is produced by synthesis. A process in common use (Reichstein–Grüssner second synthesis, 1934) starts with D-glucose and combines chemical and microbiological steps. Natural and synthetic ascorbic acid are exactly the same product, with identical biological activity; however, synthetic ascorbic acid is more economical, and frequently purer (and, therefore, more stable) than the product that can be extracted from lemons and other vegetables.

Ascorbic acid and its salts are added to food as preservatives and antioxidants, to prevent the oxidation of dyes, thus preserving color during storage and cooking. They are used in the preparation of beer, fruit juices, and canned and frozen vegetables and fruits, to prevent browning of cut apples and other

fruits, and in the flour industry to enhance baking quality and appearance of bread. Ascorbic acid is also added together with nitrite to bacon and other cured meat products to avoid the formation of carcinogenic nitrosamines. Ascorbic acid does not protect oils and fats as it is not fat-soluble. For this purpose, the fat-soluble hydrophobic esters ascorbyl 6-palmitate and ascorbyl 6-stearate are used.

Ascorbic acid is also used in nutrition as a dietary supplement in the treatment of vitamin C deficiency. It cures scurvy and acts as an oxidation–reduction catalyst in the cell, thus increasing resistance to infection. It also helps in the prevention of prostate cancer. It may aid in lowering the blood cholesterol level and slows down the aging process. When serum levels of ascorbic acid are high, synovial fluid becomes thinner, allowing freer movements and reducing arthritic pain.

Pure ascorbic acid is moderately toxic and irritating; there have been cases of allergic reaction with eczema, urticaria, and asthma. Ingestion of large amounts (10 g or more) may cause gastrointestinal distress and diarrhea, and increases the excretion of oxalic acid. Urinary and blood uric acid also rise as a result of a high intake of ascorbic acid, and these factors may increase the risk of kidney or bladder calculi. Finally, the mucolytic effect of ascorbic acid might render the cervical mucus less permeable to spermatozoa.

Erythorbic or isoascorbic acid, the epimer of L-ascorbic acid, has the same redox potential but only one-twentieth of the vitamin C activity of L-ascorbic acid. It is prepared by treating methyl 2-keto-D-gluconate with sodium methoxide, although it can also be produced by microorganisms. Sodium erythorbate has the same antioxidant, preservative, and color stabilizer properties as sodium ascorbate. It is used in soft drinks, and poultry and meat curing. When sodium nitrite is added to cure meat, erythorbic acid is also used to inhibit undesirable nitrite reactions. A brief dip in a bath containing 0.25% citric and 0.25% erythorbic acid improves quality retention in frozen fish. This treatment is also applicable to shellfish to sequester iron and copper ions that catalyze bluing and darkening reactions. (*See* **Ascorbic Acid**: Properties and Determination.)

Tocopherols

Vacuum steam distillation of cereal grain, soybean, and other vegetal oils yields a red, almost odorless, viscous oil, unstable on exposure to heat, light, and oxygen, that is a mixture of the four more abundant forms of vitamin E, i.e., the α-, β-, γ- and δ-tocopherols. The proportion of the four forms varies widely from one vegetal source to another. The most biologically active of the four is α-tocopherol, which plays several important roles in metabolism, including cellular respiration, effective use and storage of vitamin A, protection of B-complex and vitamin C from oxidation reactions, and protection of cellular membranes and body tissues from damaging reactions caused by free radicals. The other members of the vitamin E family also have important functions, some of them similar to those of α-tocopherol, whereas others are different and unique. (*See* **Tocopherols**: Properties and Determination.)

As shown in Figure 1, the differences between the natural tocopherols are due to the number and position of the methyl groups on the aromatic ring. Tocopherols other than α-tocopherol can be converted to this form of vitamin E by methylation of the aromatic ring. Tocopherols also have three asymmetric carbon atoms (one on the chromane heterocyclic ring and the other two on the alkyl chain), whose configuration is of prime importance for their biological activity. The three carbon atoms of natural tocopherols have the *R* configuration, e.g., natural α-tocopherol is the 2*R*,4′*R*,8′*R*-α-tocopherol or (*RRR*)-α-tocopherol isomer.

Synthetic α-, β-, γ-, and δ-tocopherols can be produced by condensation of the appropriate tri-, di-, or monomethylhydroquinone with synthetic isophytol (see Figure 2). In each case, a racemic mixture of the $2^3 = 8$ possible diastereoisomers or *all-rac-* mixture is obtained; for instance, 2,3,5-trimethylhydroquinone yields *all-rac*-α-tocopherol. All the isomers have

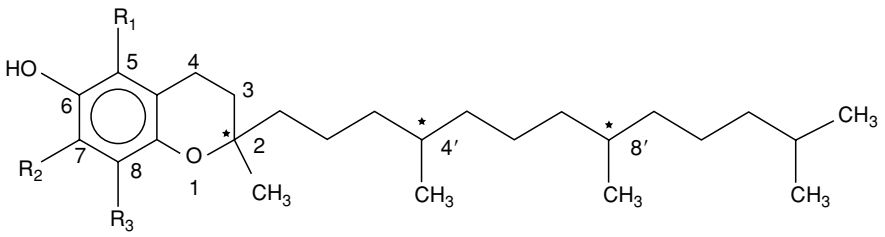

Figure 1 Molecular structure of tocopherols, where R_1═R_2═R_3═CH_3 is α-tocopherol; R_1═R_3═CH_3 and R_2═H is β-tocopherol; R_2═R_3═CH_3 and R_1═H is γ-tocopherol; R_1═R_2═H and R_3═CH_3 is δ-tocopherol.

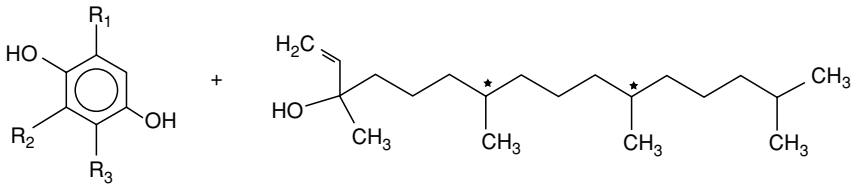

Figure 2 Condensation of 2,3,5-trimethylhydroquinone with synthetic isophytol yields *all-rac*-α-tocopherol, a racemic mixture of eight diastereoisomers.

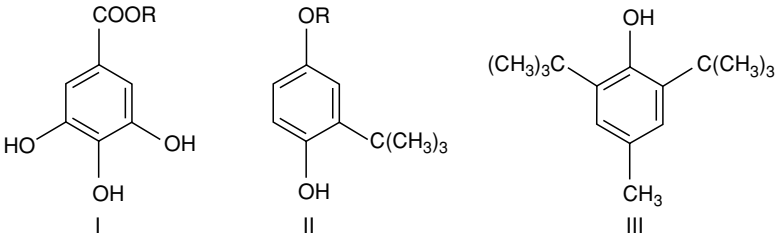

Figure 3 Molecular structures of alkyl gallates (I), *tert*-butylhydroquinone (II with $R = H$), butylated hydroxyanisol (II with $R = CH_3$ is the 2-*tert*-butyl isomer) and butylated hydroxytoluene (III).

similar antioxidant activities, but their biological values differ widely, e.g., (*SRR*)-α-tocopherol (the 2S epimer) has only 21% of the biological activity of the natural (*RRR*)-α-tocopherol. In the body, the 'incorrect' configurations are rapidly excreted, thus, on a weight basis, the synthetic *all-rac*-α-tocopherol is a much less potent vitamin E than the corresponding (*RRR*)-α-tocopherol.

The use of 2,3,5-trimethylhydroquinone and natural phytol (which is a pure *RR* isomer) results in a racemic mixture of only the two (SRR)-α- and (RRR)-α-tocopherol isomers (this latter identical to the natural product) which is called 2-*ambo*-α-tocopherol; 2-*ambo*-mixtures of the β-, γ- and δ-tocopherols can be synthesized in a similar way. Procedures to synthesize the pure (*RRR*)-isomers have also been described.

Both natural and synthetic α-tocopherols are considered as GRAS and can be used as food additives (FDA 184.1890). The recommended dietary intake (RDI) of vitamin E is 30 IU (FDA 101.9) and should be present in infant formulas (FDA 107.100). Tocopherols are used as a dietary supplement, as inhibitors of nitrosamine formation in bacon and other cured meat products, and to prevent rancidity in both chicken and turkey meats during refrigeration or frozen storage. Tocopherols are largely destroyed by cooking and canning. Large amounts of tocopherols are consumed to protect animal feeds, primarily in the poultry industry, and also as supplements and concentrates for cattle and other meat production industries. Finally, vitamin E is used to protect other substances, e.g., vitamin A and unsaturated lipids, during the shelf-life and in the gastrointestinal tract, and to improve the *in vivo* biostability and biocompatibility of medical devices made of polyetherurethane elastomers.

Tocopheryl acetates and succinates are used in pharmaceutical preparations. The esters are less reactive than the corresponding free tocopherols, and thus the shelf-life is much longer. The tocopherols are liberated in the stomach where the ester bonds are readily cleaved off. The international units of vitamin E were standardized on the basis of the *all-rac*-α-tocopheryl acetate (1 IU = 1 mg), because this mixture is the principal commercial form of vitamin E; 1 mg of (*RRR*)-α-tocopheryl acetate equals 1.36 IU.

Alkyl Gallates, BHA, BHT, and Related Compounds

The molecular structures of these phenolic antioxidants are shown in Figure 3. As occurs with other phenols, the antioxidant action of alkyl gallates, *tert*-butylhydroquinone (TBHQ), BHA, and BHT mainly relies on their ability as free radical scavengers. In contrast to tocopherols, ascorbic acid, and other antioxidants, the antioxidant properties of these phenol derivatives are not lost during cooking. According to the FDA, these antioxidants can be added to food products whether singly or in combination of two or more, up to 0.02%, based on the weight of the fat or oil content in the food product.

Propyl gallate (PG) is a white to nearly white odorless crystalline powder that has slightly bitter taste and darkens in the presence of iron salts. It is an

irritant that can cause skin dryness, dermatitis, and sensitization. Inhalation may cause chemical pneumonia. PG is used to stabilize cosmetics, food packaging materials, and foods including edible fats, oils, mayonnaise and other emulsions, shortening, baked goods, candy, dried meat, fresh pork, sausage, and dried milk. The antioxidant activity of PG is synergic with acids, BHA and BHT. It also shows antimicrobial activity, has been reported to be an effective antioxidant-based hepatoprotector, and has been shown to prevent neuronal apoptosis.

PG is markedly hydrophillic, and so its solubility in fats is limited. Substitution of the propyl chain by a shorter (ethyl) or longer (octyl or dodecyl) chain, respectively, decreases or increases fat solubility. Propyl, octyl, and dodecyl gallates (< 0.02%) individually or in combination with BHT, BHA, and ascorbyl palmitate and stearate can be added to margarine (FDA 166.110). Alkyl gallates, although quite effective, can cause discoloration of food substrates, and may also impart off-flavors. Propyl gallate is also used for other industrial purposes, e.g., added to ethers and aldehydes to prevent peroxidation.

TBHQ is sometimes added to food as an antioxidant, and it is also produced in the body as the common metabolite of the two BHA isomers. It is a tan-colored powder with acute irritating properties. It is most effective against oxidation in polyunsaturated vegetal oils, and is often added to soybean oil.

Commercially available food-grade BHA is generally a mixture containing two positional isomers, 3-*tert*-butyl-4-methoxyphenol (> 85%) and 2-*tert*-butyl-4-methoxyphenol (< 15%). It is a white or slightly yellow waxy solid, has an aromatic odor and a slightly bitter burning taste, and degrades following prolonged exposure to sunlight. It is toxic by ingestion, and there is sufficient evidence that it is carcinogenic in animals. It dissolves readily in molten fats, and its antioxidant properties are synergic with acids, BHT, alkyl gallates, hydroquinone, methionine, lecithin, and thiodipropionic acid. At low concentrations, it is used as an antioxidant in fat-containing foods, in edible fats and oils, and as a preservative and antioxidant in cosmetic formulations.

BHT is a white crystalline powder. Contact causes irritation, and inhalation produces lung injury. Depending on the intensity and duration of exposure, effects may vary from mild irritation to severe destruction of tissue. It is more hydrophobic and therefore more soluble in oils and fats than BHA and TBHQ. It is normally used in combination with citric acid in fats and oils, and with phosphate for meat systems. It is also used to protect rubber, plastics, and other industrial products.

BHT combines low costs with heat stability, good antioxidant activity, and high fat solubility, and thus not surprisingly has been the most widely used antioxidant in food in the last quarter of the twentieth century; however, continuing concern over its safety has resulted in decreased usage. Both carcinogenic and anticarcinogenic properties have been reported for BHA and BHT. Thus, metabolites of BHT can promote tumors and produce DNA damage in mice. However, BHT has been found to inhibit the initiation of hepatocarcinogenesis by aflatoxin B1. A study performed in The Netherlands among 120 852 men and women over 6.3 years showed no significant association with stomach cancer risk for an intake of low levels of BHA and BHT. However, when excessively concentrated, phenolic compounds can promote undesired oxidative or degenerative reactions, e.g., natural flavonoids show a marked prooxidant effect that is due to the chain-initiating activity of the radicals formed during inhibition; however, BHA and BHT have no prooxidant activity.

Anoxomer is a complex polymeric substance with a molecular weight of 4500–5000. For its use in food, no more than 5% of the polymer should be greater than 50 000 Da, and no more than 1% of the polymer should be less than 500 Da. It is prepared by condensation polymerization of *m*- and *p*-divinylbenzene with TBHQ, *tert*-butylphenol, hydroxyanisole, *p*-cresol, and 4,4′-isopropylidenediphenol. Anoxomer should be used as an antioxidant at a level no greater than 0.5% of the fat and oil content of foods (FDA 172.105). Gallic acid esters, BHT, BHA, and TBHQ have high vapor pressures at temperatures above 150 °C, and so they are readily lost from thermal processes generating steam. This does not occur with anoxomer. In addition, in contrast to the monomeric phenolic antioxidants, pure anoxomer is not an irritant, thus possibly indicating a greater biological tolerance.

Although not collected in the INS list, there are other phenols with antioxidant properties similar to those exhibited by BHA and BHT, and which have also been used as food additives. Thus, 2,4,5-trihydroxybutyrophenone (THBP) ($C_{10}H_{12}O_4$) or 2,4,5-trihydroxy phenyl propyl ketone has been used as antioxidant for polyolefins, paraffin waxes and food. Nordihydroguaiaretic acid (NDGA) or 4,4′-(2,3-dimethyl-tetramethylene)dipyrocatechol is a natural antioxidant. It can be extracted from an evergreen desert shrub, the creosote bush, *Larrea divaricata*, and can also be prepared by hydrogenation and subsequent demethylation of the guaiaretic acid dimethyl ether, a constituent of gum guaiac. It is added to fats and oils. The molecular structures of THBP and NDGA are shown in Figure 4.

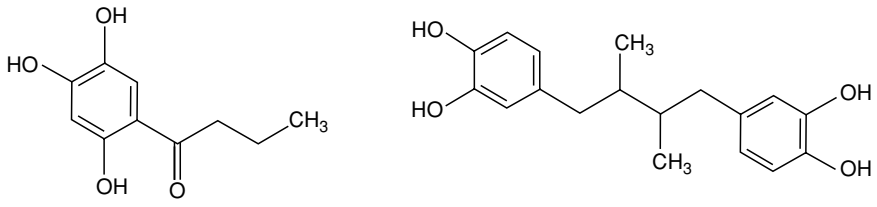

Figure 4 Molecular structures of 2,4,5-trihydroxybutyrophenone and nordihydroguaiaretic acid.

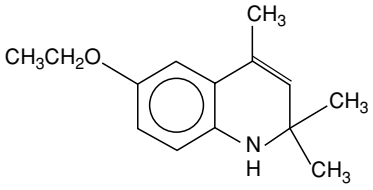

Figure 5 Molecular structure of ethoxyquin.

Ethoxyquin

Ethoxyquin (EQ) (see Figure 5) is a clear, viscous, light yellow to dark brown liquid, with an unpleasant, mercaptan-like smell, that causes irritation of the skin and eyes. It is absorbed through the skin, and is moderately toxic by ingestion. It darkens on exposure to light and air and tends to polymerize, particularly at temperatures above 160 °C (hazardous exothermic reaction). EQ is considered a highly effective antioxidant, is cheap and has a very long shelf-life. It has been the antioxidant of choice in pet food for many years, although it has also been added to other animal feeds and human food. Concern about its toxicity (not yet sufficiently investigated) has promoted the alternative use of tocopherols in pet food, in spite of the higher prices and increasing demand of tocopherols for human consumption. EQ is also used as an insecticide, herbicide, fungicide, postharvest dip to prevent scald on apples and pears, plant growth regulator, and antidegradation agent for rubber.

Sodium and Potassium Lactates

Lactic acid is present in many foods both naturally and as a product of *in situ* fermentation, as in sauerkraut, yogurt, and many other fermented foods. Lactic acid is also a principal metabolic intermediate in most living organisms. Sodium and potassium lactates are produced commercially by neutralization of natural or synthetic lactic acid (FDA 184.1768, 1639). Lactic acid to be used as a food additive can be obtained either by fermentation of carbohydrates or by a chemical procedure involving formation of lactonitrile from acetaldehyde and hydrogen cyanide and subsequent hydrolysis (FDA 184.1061).

The microbiological and chemical procedures to obtain lactic acid are very competitive, with similar production costs. One method of biosynthesis in common use starts with glucose and produces pyruvate, which can be converted to both the L(+) and D(−) isomers using a stereospecific lactate dehydrogenase; however, only the L(+) form is produced commercially. The racemic mixture is always obtained by chemical synthesis. Synthetic lactic acid is free of the contaminants normally found in the product obtained by fermentation, and so it is completely colorless and probably more stable. Lactic acid and its salts are highly hygroscopic, and therefore are usually handled in concentrated solutions (60–80% by weight) rather than in solid form. These solutions are colorless and odorless, and have a mild saline taste.

Lactates complex metal ions, thus improving the stability of food against oxidation. Sodium lactate is used as a pH regulator and antimicrobial, humectant, curing, pickling, and flavor enhancing agent in meat, poultry, seafood, and many other products; potassium lactate is used to replace sodium lactate in low-sodium products. The calcium salt of lactic acid is a good dough conditioner, and the sodium salt is both a conditioner and an emulsifier for yeast-leavened bakery products. The addition of aqueous solutions of lactic acid and its salts during processing of carcasses, particularly those of poultry and fish, increases shelf-life and reduces the growth of anaerobic spoilage organisms such as *Clostridium botulinum*. Sodium lactate is also used as a moisturizer in skin-care products, to supplement the skin's own naturally occurring sodium lactate, as well as in a wide variety of industrial applications. (*See* ***Clostridium***: Botulism.)

Orthophosphoric Acid and Orthophosphates

The antioxidant synergism of orthophosphoric acid and its salts is a result of their ability to complex metal ions. The redox potential of the Fe^{3+}/Fe^{2+} pair is particularly lowered in the presence of phosphates,

even in acid medium. Orthophosphoric acid can be used to assist in caramelization (FDA 73.85), as a neutralizing agent in cacao (FDA 163.110), and as an acidifying agent in cheese manufacture (FDA 133.123, 124). Several sodium and potassium orthophosphates and pyrophosphates can also be used as emulsifying agents to process cheese (FDA 133.169, 173). The acid is also added to soft drinks and has miscellaneous industrial applications.

Ethylenediaminetetraacetates

In an alkaline medium, the dibasic anion of the ethylenediaminetetraacetic acid (EDTA) forms very stable complexes with most divalent and trivalent metal ions. Thus, citric acid and EDTA salts are the most common chelating agents used as antioxidant synergists in food. As a consequence of the concurrence with hydrogen ions, the sequestrant capability of EDTA decreases as the pH is lowered; however, the EDTA complexes of iron, copper, zinc, aluminum, and many other metal ions are still very stable in neutral and slightly acid media. EDTA also lowers the redox potential of the Fe^{3+}/Fe^{2+} pair. The use of calcium disodium EDTA, rather than disodium EDTA, prevents the calcium-depleting action of the ligand in the body. Calcium disodium EDTA can be added to margarine (0.0075%, FDA 166.110), and is also permitted whether alone or in combination with disodium EDTA in a variety of canned vegetables, emulsions, and dressings to preserve color and flavor (FDA 172.120, 169.150).

Thiodipropionic Acid and its Fatty Esters

Thiodipropionic acid and the dilauryl and distearyl thiodipropionates are weak reducing substances, but they are very effective in the destruction of hydroperoxides, which are reduced to alcohols. The sulfur atom is eventually oxidized to sulfate, but the reduced sulfur species originated as reaction intermediates catalyze the decomposition of more hydroperoxide molecules. Up to 20 moles of hydroperoxide can be decomposed by a mole of thiodipropionate. Thiodipropionates display strong synergism with BHT and other phenol derivatives, which also terminate free radical chain reactions through a rather different mechanism.

Thiodipropionic acid and its esters are GRAS for use in food when the total content of antioxidants is not over 0.02% of fat or oil content of the food (FDA 182.3109, 3280). They are used to retain color in dried bananas, beans, peas, canned clams and shrimp, and frozen potatoes. They improve flavor retention in canned carbonated beverages, dressings, mayonnaise, margarine, and sauces. They are also used as antioxidants in soap products and lubricants, and in the manufacture of a variety of polymers (including adhesives). When used in packaging materials, they should not migrate to food in amounts exceeding a concentration of 0.005% in the food (FDA 178.2010, 181.24).

Stannous Chloride

Stannous chloride is a very strong reducing agent that absorbs oxygen from the air. The colorless crystalline solid or the white powder is severely irritating and corrosive, and reacts with alcohols, amines, alkalies, and oxidizers, sometimes with violence; the reaction with hydrogen peroxide is strongly exothermic, even in solution. As a food preservative, it can be added to canned asparagus packed in glass containers in a quantity not exceeding $15\,\mu g\,g^{-1}$ calculated as tin (Sn), but up to $20\,\mu g\,g^{-1}$ can be present if the container lid is lined with an inert material (FDA 155.200). It is also used as a stabilizer for colors and perfumes in soaps, and as a reducing agent in many other industrial applications.

Sodium thiosulfate

The addition of sulfur to sodium sulfite produces sodium thiosulfate, a salt that contains an S–S bond. Its antioxidant properties are due to the presence of a reduced sulfur atom, but it also forms strong complexes with iron, copper, and other metal ions. It can be added to alcoholic beverages (< 0.00005%) and table salt (< 0.1%) (FDA 184.1807). It is also used as a reducing (e.g., to remove residual chlorine from solutions), bleaching, and complexing agent in a variety of industrial applications.

4-Hexylresorcinol

4-Hexylresorcinol is a pale yellow, viscous liquid (becoming solid on standing at room temperature) with a pungent faint fatty odor and a sharp astringent taste. It is an irritant, and its concentrated solutions can cause burns on the skin and mucous membranes. The Canadian Food Inspection Agency allows the addition of up to $1\,\mu g\,g^{-1}$ hexylresorcinol to frozen and fresh crustacean (http://inspection.gc.ca). It is used in medicine as a topical and urinary antiseptic, anthelmintic, cleanser for skin wounds, and anesthetic for throat pain.

See also: **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants, Characterization and Analysis; Role of Antioxidant Nutrients in Defense Systems; **Atherosclerosis**; **Cancer**: Carcinogens in the Food Chain; Diet in Cancer Prevention; Diet in Cancer Treatment; **Coronary Heart Disease**: Antioxidant Status; **Legislation**: Additives; **Mutagens**; **Oxidation of Food**

Components; **Phenolic Compounds**; **Storage Stability**: Parameters Affecting Storage Stability; **Tannins and Polyphenols**; **Tocopherols**: Properties and Determination; **Vitamins**: Overview

Further Reading

Adegoke GO, Kumar MV, Krishna AGG *et al*. (1998) Antioxidants and lipid oxidation in foods. A critical appraisal. *Journal of Food Science and Technology (Mysore)* 35: 283–298.

Bauerova K and Bezek S (1999) Role of reactive oxygen and nitrogen species in etiopathogenesis of rheumatoid arthritis. *General Physiology and Biophysics* 18(special issue) 15–20.

Botterweck AAM, Verhagen H, Goldbohm RA, Kleinjans J and vanden-Brandt PA (2000) Intake of butylated hydroxyanisole and butylated hydroxytoluene and stomach cancer risk: Results from analyses in The Netherlands cohorts study. *Food Chemistry and Toxicology* 38: 599–605.

Budavary S (ed.) (1996) *The Merck Index*, 12th edn. Rahway, NJ: Merck & Co.

Coupland JN and McClements DJ (1996) Lipid oxidation in food emulsions. *Trends in Food Science and Technology* 7: 83–91.

Dangles O, Dufour C and Fargeix G (2000) Inhibition of lipid peroxidation by quercetin and quercetin derivatives: antioxidant and prooxidant effects. *Journal of the Chemistry Society Perkin Transactions* 6: 1215–1222.

Fukumoto LR and Mazza G (2000) Assessing antioxidant and prooxidant activities of phenolic compounds. *Journal of Agricultural and Food Chemistry* 48: 3597–3604.

Gertz C, Klostermann S and Kochhar SP (2000) Testing and comparing oxidative stability of vegetable oils and fats at frying temperature. *European Journal of Lipid Science and Technology* 102: 543–551.

Gower JD (1988) A role for dietary lipids and antioxidants in the activation of carcinogens. *Free Radical Biology and Medicine* 5: 95–111.

Hras AR, Hadolin M, Knez Z and Bauman D (2000) Comparison of antioxidative and synergistic effects of rosemary extract with alphatocopherol, ascorbyl palmitate and citric acids in sunflower oil. *Food Chemistry* 71: 229–233.

Iuliano L, Pedersen JZ, Camastra C *et al*. (1999) Protection of low density lipoprotein oxidation by the antioxidant agent IRFI005, a new synthetic hydrophilic vitamin E analogue. *Free Radical Biology and Medicine* 26: 858–868.

Kirk R and Othmer DF (eds) (1978–1984) *Encyclopedia of Chemical Technology*, 3rd edn. New York: Wiley.

Kirk R and Othmer DF (eds) (1991–1995) *Encyclopedia of Chemical Technology*, 4th edn. New York: Wiley.

Morsel JT and Meusel D (1990) Fortschrittsbericht. Lipidperoxydation. 2. Mitt. Sekundare Reaktionen [Progress report. Lipid peroxidation. Part 2. Secondary reactions]. *Nahrung* 34: 13–27.

Moure A, Cruz JM, Franco D *et al*. (2001) Natural antioxidants from residual sources. *Food Chemistry* 72: 145–171.

Schubert MA, Wiggins MJ, DeFife KM, Hiltner A and Anderson JM (1996) Vitamin E as an antioxidant for poly(etherurethane urea): *in vivo* studies. *Journal of Biomedical Materials Research* 32: 493–504.

Ueda T, Ueda T and Armstrong D (1996) Preventive effect of natural and synthetic antioxidants on lipid peroxidation in the mammalian eye. *Ophthalmic Research* 28: 184–192.

Valenzuela A and Nieto S (1996) Synthetic and natural antioxidants: Food quality protectors. *Grasas y aceites* 47: 186–196.

Weisburger JH (1999) Mechanisms of action of antioxidants as exemplified in vegetables, tomatoes and tea. *Food Chemistry and Toxicology* 37: 943–948.

Code of Federal Regulations, USA http://www.access.gpo.gov/nara/cfr/index.html

FAO/WHO Joint Expert Committee on Food Additives (JECFA). *Monographs and Evaluations* http://www.inchem.org/jecfa.html

International Numbering System (INS) for Food Additives ftp://ftp.fao.org/codex/standard/volumela /en/Cxxot04e.pdf

Miscellaneous databases on chemical products: http://www.chemfinder.com, http://www.nutritionfocus.com, http://www.sigma-aldrich.com

National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, NC http://ntp-server.niehs.nih.gov

School of Food Biosciences, The University of Reading, UK http://www.fst.rdg.ac.uk/foodlaw

US Food and Drug Administration, Center for Food Safety & Applied Nutrition, Office of Premarket Approval, EAFUS: A Food Additive Database http://vm.cfsan.fda.gov/~dms/eafus.html

Synthetic Antioxidants, Characterization and Analysis

G Ramis-Ramos, Universidad de Valencia, Burjassot, Spain

Phenolic Antioxidants

Reversed-phase High-performance Liquid Chromatography

The determination of phenolic antioxidants and ethoxyquin (EQ) in food is most frequently accomplished by reversed-phase high-performance liquid chromatography (RP-HPLC), while both HPLC and thin-layer chromatography (TLC) are normally used

for the screening of antioxidant residues in food products. Usually, sample preparation for RP-HPLC involves dilution or suspension with hexane, extraction with a polar solvent (most frequently acetonitrile mixed with some water), preconcentration by evaporation under reduced pressure, reconstitution (e.g., with isopropanol), and, if necessary, clean-up with unmixable solvents or solid phases. Extraction with supercritical CO_2 has been used for capillary electroseparations. RP-HPLC separations are normally performed on C18 columns using gradient elution with aqueous phases containing methanol or acetonitrile and acetic acid, followed by spectrophotometric detection at 280–285 nm. Pentachlorophenol and 2,4,6-trimethylphenol have been employed as internal standards in the analysis of fats and oils and bakery products.

A modification of the AOAC official method applied to butter oil has yielded recoveries higher than 95% for alkyl gallates, *t*-butylhydroxyquinone (TBHQ), butylated hydroxyanisole (BHA), nordihydroguaiaretic acid (NDGA), 2,4,5-trihydroxybutyrophenone, and 2,6-di-*t*-butyl-4-hydroxymethylphenol, and lower for butylated hydroxytoluene (BHT) (79%). Other reported data indicate recoveries higher than 90% for BHA and BHT in vegetable oil and propyl gallate (PG) in lard, and for dodecyl gallate (DG), BHA, BHT, TBHQ, and NDGA in liver pâté, but the less hydrophobic PG and octyl gallate (OG) gave recoveries around 80% in this sample. Using methanol, recoveries between 85 and 95% have been reported for PG, OG, DG, 2- and 3-BHA, NDGA and TBHQ, but 3,5-di-*t*-butylcresol gave 72%. The intralab relative standard deviations (RSDs) range typically between 3 and 7%, but values lower than 2% have also been obtained. Reported interlab RSDs range from 7% for DG to 26% for TBHQ. Low or rather disperse recoveries can be due to oxidation during sample storage and preparation, which easily occurs with TBHQ.

In one procedure for dry food, the sample was homogenized with hexane, water, and acetonitrile (25:5:75) and the combined acetonitrile extracts concentrated by evaporation, injected, and detected at 280 nm. Separation was performed on a C18 column with gradient elution using water–acetonitrile mobile phases containing 5% acetic acid. The detection limits (DLs) for PG, BHA, BHT, *t*-butylquinol, 2,4,5-trihydroxybutyrophenone, 3,5-di-*t*-butyl-4-hydroxymethylphenol, and NDGA ranged from 0.4 to 2.7 $\mu g\,ml^{-1}$. For a typical 20-µl injection loop, these data roughly match with other DLs reported as absolute values: 25 (OG, DG), 35 (PG) and 100 ng (NDGA). In the determination of nine phenolic antioxidants in corn oil, butter oil, butter, niboshi (dried sardines), and frozen shrimps, the DLs were 1.0 µg per gram of sample (85–102% recoveries).

Lower DLs can be achieved with fluorometric, chemoluminometric, and amperometric detection. For BHA, BHT, and TBHQ, excitation at 280 nm yields maximum emission at *c.* 310 nm. Procedures based on the chemiluminescence of luminol are mainly used to measure antioxidant activity, and these are discussed in detail later in the corresponding section. *t*-Butylquinol, 2,6-di-*t*-butyl-*p*-cresol, and 2- and 3-*t*-butyl-4-methoxyphenol have been determined in oils by HPLC with amperometric detection on a glassy carbon electrode. With a previous modification of the carbon electrode surface by electropolymerization of nickel phthalocyanine, DLs ranging from 0.1 to 0.6 $\mu g\,ml^{-1}$ have been achieved. When used in a flow-injection nonchromatographic system, the modified electrode provides a DL of 2.7 $ng\,ml^{-1}$ with a 1.8% RSD (at 0.5 $\mu g\,ml^{-1}$) for BHA in biscuits. Using 0.3- and 0.7-mm microbore HPLC with electrochemical detection, DLs as low as 0.1 and 0.6 $fmol\,l^{-1}$ have been obtained for BHA, BHT, and PG.

HPLC with Ion-pairing and Micellar Mobile Phases

The selectivity of the HPLC separations can be modified by ion-pairing agents, such as alkyl sulfates and carboxylates, and alkyl ammonium salts. These reagents form micelles when at least one of the alkyl chains has 12 or more carbon atoms; however, micelles are disrupted in the presence of a large organic solvent concentration, thus leading to an ion-pairing rather than a micellar separation mechanism. PG, OG, DG, BHA, and BHT can be determined in spiked olive oil with a mobile phase containing sodium dodecyl sulfate (SDS) and 30% propanol in 18 min.

An aqueous micellar mobile phase, which may also contain a lower percentage of an organic solvent (usually 5–15% of a short-chain alcohol), is used in micellar liquid chromatography (MLC). In MLC, physiological fluids and tissue extracts can be injected without deproteinization, because proteins are solved in the micellar phase and elute with the solvent front. Moreover, hydrophobic samples can be simply diluted with a compatible solvent and injected. Thus, several oils and fats diluted with *n*-pentanol (5–30% solutions) can be injected on a C18 column and eluted with an SDS/*n*-propanol mobile phase; PG, OG, TBHQ, and BHA have been quantified in *c.* 15 min with absolute DLs ranging from 0.05 to 0.3 ng. Avoiding a previous two-phase partitioning step simplifies the procedure and may help in obtaining high recoveries and in reducing random and systematic errors, particularly when easily oxidizable solutes such as TBHQ are quantified.

Further, since retention time in MLC is linearly proportional to solute hydrophobicity, isocratic elution is always suitable to separate solutes with large hydrophobicity differences, whereas gradient elution is required in most cases with conventional hydroorganic phases, owing to the exponential dependence of the retention time with solute hydrophobicity. Also, in comparison to hydroorganic phases, micellar phases have a lower toxicity and produce a milder environmental impact, and owing to the lower vapor pressure of its components, a small volume of phase can be recirculated and reused many times without any perceivable alteration of the chromatograms. To prevent clogging by precipitation of the surfactant, a simple cleaning protocol with water and methanol should be followed before stopping the pumps or storing a used column.

Another advantage of MLC is the excellent reproducibility of the chromatograms and, frequently, the almost perfect Gaussian shape of the peaks, which facilitates modeling and prediction. Thus, the chromatographic retention of seven antioxidants has been described, with errors below 3%, by the equation

$$1/k = c_0 + c_1\mu + c_2\varphi + c_3\mu\varphi$$

where k is the relative retention (capacity factor), μ and φ are the concentrations of SDS and organic solvent (*n*-propanol in this study), and c_0 to c_3 are regression coefficients, which are calculated for each solute from five chromatograms obtained with different μ and φ values. Modeling of retention speeds up the optimization of chromatographic procedures, i.e., the resolution between all the peak pairs of a group of solutes can be maximized, and the interference of matrix peaks can be avoided, by using the model predictions. This procedure can be used to optimize the MLC separation of a mixture of antioxidants in a short elution time.

Capillary Electroseparations

Capillary electroseparations are rapid, versatile, and very efficient, solutions in solvent mixtures of all types with the sole requirement of supporting some conductivity can be used to fill up the capillary, and most hydrophilic and hydrophobic samples can be injected without pretreatment. Further, only the anions or the cations present in a sample can be selectively introduced in the capillary by using electrokinetic injection. Antioxidants can be separated in at least two modes: (1) capillary zone electrophoresis (CZE), which is based on the different electrophoretic mobilities of the charged solutes, and (2) micellar electrokinetic capillary chromatography (MECC), in which differential migration results from the different strength of the association of uncharged solutes with charged discriminant agents such as ionic surfactants (both as individual ions or as micelles), vesicles and derivatized (e.g., sulfated) cyclodextrines. In mixed MECC, charged solutes are separated by the joint action of their own electrophoretic mobilities and the different strength of the solute–discriminant agent interactions.

Using a CZE method with spectrophotometric detection, flavonoids and phenolic antioxidants have been determined in red wine and olive leaves. In one experiment, the wine samples were directly injected, and the leaf extracts were prepared by digestion with boiling water or aqueous methanol, Soxhlet extraction with acetone/CH_2Cl_2, and supercritical fluid extraction with CO_2. Triphenyl acetic and benzoic acids were used as migration time markers, to assist in the identification of the analytes. The absolute DL was 3 pmol.

In a mixed MECC procedure, a buffer containing SDS was used to separate 11 phenolic antioxidants with DLs of 2.8–8.6 $\mu g\,ml^{-1}$. Vesicles of bis(2-ethylhexyl)sodium sulfosuccinate have been used to separate six food antioxidants in 15 min. Sodium cholate and SDS have been used to separate nine phenolic antioxidants and ascorbic acid with DLs in the range of 1–10 $\mu g\,ml^{-1}$ (spectrophotometric detection at 214 nm), and the procedure has been applied to wine and sesame seed oil.

In a previous study, CZE, MECC, and HPLC were compared, using a sodium borate buffer of pH 9.5 for CZE, adding SDS to this buffer for MECC, and performing RP-HPLC separations on a C18 column with isopropanol/water/acetonitrile (2:17:31). The resolution was poor with CZE and excellent with MECC. The antioxidants were separated in 6 and 25 min by MECC and HPLC, with absolute DLs in the picomole and nanomole ranges, respectively. The reader should not be misled by the reported absolute DLs, which are usually extremely low due to the very small injection volumes (in the picoliter range). Owing to the short optical path length (limited to the capillary inner diameter), the relative DLs are in fact higher than in HPLC with spectrophotometric detection. The DLs of capillary electroseparations can be reduced by a factor of five to 10 by using capillaries with specially extended path lengths, or more effectively by implementing fluorimetric or mass-spectrometric detection.

Gas Chromatography

BHA, BHT, TBHQ, α-tocopherol, and α-tocopherol acetate have been determined using a gas chromatograph coupled to a solid-phase extractor device, extracting powdered soup with *n*-hexane in the presence of 2-*t*-butyl-4-methylphenol and 5-α-cholestane

as internal standards, and washing the extract with 0.2 M NaCl to remove any water-soluble compounds. Similarly, oil and margarine samples were spiked with the internal standards and diluted with *n*-hexane. The antioxidants were retained on Amberlite XAD-7 resin and eluted with isopropanol, and portions of the eluate were injected. Using a coated fused-silica column and a flame ionization detector (FID), the DLs ranged from 2 to 8 mg l^{-1}, and the RSDs for 50 mg l^{-1} were 2.7–3.5%. The recoveries of 125 mg of antioxidants per kilogram from spiked sunflower seed oil were 93.6–101.4%. Synthetic antioxidants have also been determined by GC in soya-bean oil and sesame oil.

Thin-layer Chromatography

To screen PG, OG, DG, BHA, and BHT by TLC, oil samples have been diluted with hexane and extracted with 80% ethanol and acetonitrile. The combined extracts were evaporated under reduced pressure, and the concentrate was spotted on the TLC plates. After development, the solutes were revealed by oxidative coupling with *N*,*N*-dimethyl-*p*-phenylenediamine in the presence of hexacyanoferrate (III). The R_F values and the color of the spots obtained with several mobile phases were then tabulated.

Nonchromatographic Techniques

Rapidity at the cost of a lower selectivity can be gained by using nonchromatographic techniques. Selectivity can be improved by using kinetic approaches and other masking techniques, and, as detailed below, if the signals given by several solutes overlap, the required qualitative and quantitative information can be retrieved by using multivariate data treatment techniques, including principal-component regression (PCR) and partial least-squares regression when all solutes exhibit linear signal–concentration relationships, and artificial neural network (ANN) algorithms in case of nonlinearity. Thus, in a differential kinetic determination, the oxidation of several antioxidants with Fe(III) has been monitored with 2,2′-dipyridyl, and the spectrophotometric-time data treated with an ANN algorithm.

In comparison to ordinary spectrophotometry, selectivity can be improved by using derivative spectrophotometry. This technique has been used to determine BHA and BHT in lard, diluting the sample with hexane, extracting with dimethyl sulfoxide, and measuring the second derivatives of the spectrum at 311.5 and 278.5 nm for BHA and BHT, respectively (with recoveries of 99.5–109.3% and 90.6–108.1%).

Sensitization of the luminescence of Tb(III) in a micellar medium has been used to determine PG by a stopped-flow initial-rate kinetic method. Dilution of the oil samples with hexane, followed by extraction in an ammonium acetate aqueous solution, mixing with the Tb(III) reagent and data collection over 1 s has yielded recoveries of 90.1–108.2%, a RSD of $\sim$ 2%, and a DL of 0.02 μg ml^{-1}. Mixtures of PG and BHA have also been determined with a stopped-flow setup, the former being evaluated by the initial rate method, by monitoring the luminescence of Tb(III) at 0.1 s, and the latter being determined from the equilibrium signal of its reaction product with the oxidized form of Nile Blue, measured at 1 s. Both systems are excited at 310 nm, and the emission wavelengths are 545 nm for PG and 665 nm for BHA. Recoveries from food samples have been found to range between 94 and 103%, with RSDs of approximately 2% and DLs of 0.03 and 0.09 μg ml^{-1} for PG and BHA, respectively.

Phenolic antioxidants present well-defined oxidation waves, with peak potentials of 629, 818, 599, and 501 mV, with linear concentration ranges of 1–15, 0.5–15, 0.5–8, and 1–15 mg l^{-1}, for PG, BHA, BHT, and TBHQ, respectively. Square-wave voltammetry on carbon-fiber disk ultramicroelectrodes has been used to determine BHA and BHT in acetonitrile extracts of vegetable oils and lard, yielding DLs of a few micromoles per liter and RSDs of about 6–7%. To determine BHA in chewing gum, samples are shaken with ethyl acetate and the polymer components precipitated by freezing; the supernatant is then diluted with a 1% solution of the polymeric nonionic surfactant Pluronic F-68 and buffered. Differential pulsed voltammetry at a glassy carbon electrode has yielded DLs below 1 μmol l^{-1}.

Problems arising from overlapping oxidation waves in mixtures can be overcome by improving the electrode kinetics through modification of the electrode surface. Thus, PG, BHA, TBHQ, and BHT can be simultaneously determined in mixtures by differential-pulse voltammetry on a carbon-fiber electrode coated with a poly-(3-methylthiophene) film. In micellar solutions and in an oil-in-water emulsion, a DL for PG of 0.8 μmol l^{-1} and a RSD for 50 μM of 1.3% have been found. The method has been applied to dehydrated soups. Cyclic voltammetry with modified Pt electrodes has been used to determine PG in cornflakes and potato flakes, extraction with 1:1 water/methanol leading to quantitative recoveries and a DL of 1.5 μg ml^{-1}.

As indicated above, overlapped signals can be discerned by using multivariate data treatment techniques, and thus PCR and PLS have been applied to voltammetric data. Differential pulse voltammetry with PLS calibration has been used to determine PG, BHA, and BHT in dehydrated soup, with DLs of 0.15,

0.7, and 7 μmol l^{-1}, respectively, and recoveries ranging from 77 to 101%.

Finally, an amperometric enzyme biosensor (BAS VC-2 tyrosinase electrode) has been used to determine BHA in biscuits, extracting the samples with ethyl acetate, concentrating the extracts, and diluting with a medium of dioctyl sulfosuccinate reversed micelles. Under flow conditions, the DL has been found to be 30 μmol l^{-1}.

Characterization and Determination of Anoxomer

Anoxomer is a nonabsorbable polymeric phenol obtained by condensation polymerization of *m*- and *p*-divinylbenzene with TBHQ, *t*-butylphenol, hydroxyanisole, *p*-cresol and 4,4′-isopropylidenediphenol. Two types of analytical problems should be addressed in relation to anoxomer: (1) to verify regulations concerning its composition and purity and (2) to detect and determine anoxomer in food products and wrapping materials. According to the US Code of Federal Regulations (CFR 21, 172.105), the specifications to be met are: (1) the purity should not be lower than 98.0%, (2) the total amount of monomers, dimers, and trimers below a molecular mass of 500 should not be more than 1%, (3) the phenol content should range from 3.2 to 3.8 meq g^{-1} and (4) traces of Pb, As, and Hg should not exceed 10, 3, and 1 μg g^{-1}, respectively. Assays to check specifications (1), (2), and (3) are available from the Center for Food Safety and Applied Nutrition and other US Administration agencies.

Several qualitative and quantitative procedures for anoxomer have been investigated, but no methods for real samples have been published. A linear calibration graph over the range of 5–50 μg ml^{-1} can be constructed by heating the product with nitric acid in a medium containing acetic acid, adjusting the pH to 8 with ammonia, and measuring the absorbance at 420 nm. A linear calibration graph over the range 1–10 μg ml^{-1} has been obtained by measuring the native fluorescence of anoxomer in diethylene oxide. A lower sensitivity has been obtained by measuring the infrared absorption of anoxomer in tetrachloroethylene solutions at 3525 cm^{-1} (-OH stretching; linearity range: 150–580 μg ml^{-1}).

4-Hexylresorcinol

4-Hexylresorcinol has been determined in shrimps by extraction with methanol and RP-HPLC with fluorimetric detection (excitation at 280 nm, emission at 310 nm), with a recovery of ~ 82%, RSD of 2.1% within the range 1.5–2.5 mg kg^{-1}, and a DL of 1.98 ng ml^{-1} in the extract and 80 μg kg^{-1} in shrimp. In another method for shrimps, the ethyl acetate extracts have been cleaned up by solid-phase extraction, followed by gradient elution on a C18 column with trifluoroacetic acid/acetonitrile/methanol/water, and detection at 214 nm (DL = 6 ng, 94% recovery). In a method for crab meat, extraction with acetonitrile, cleaning up by solid-phase extraction, and HPLC with diode array detection has been used (RSD = 7.1%, 86–104% recovery).

Nonphenolic Synthetic Antioxidants

Ethoxyquin

EQ can be determined by extracting the samples with hexane, followed by HPLC with spectrophotometric (absorption bands at 230 and 365 nm), fluorimetric (emission at 435 nm with excitation at 365 nm), or electrochemical (oxidation wave at + 0.45 V) detection, or by GC with a FID or with thermionic detection, or using TLC on fluorescent plates. In a comparative study, significant differences among the RDSs were not found for EQ in fish meals using different techniques, which was attributed to the large variability in the extraction process.

EQ is decomposed by light, which should be taken into account in storing samples. It has been found that EQ is stable for at least 9 days by keeping the apple extracts in deep freeze, and for at least 20 days by storing the pieces of apple in hexane in the dark. Further, its degradation products are found in significant concentrations in the food samples. EQ and up to 12 of its degradation products have been resolved by HPLC. Also, EQ and its two major oxidation products, 1,8′-di-1,2-dihydro-6-ethoxy-2,2,4-trimethylquinoline (DM) and 2,6-dihydro-2,2,4-trimethyl-6-quinolone (QI), have been determined in fish meals by HPLC (DL = 5 mg kg^{-1} for EQ and DM, and 0.5 mg kg^{-1} for QI) and gas chromatography (GC), and the changes in the EQ, DM, and QI concentrations during storage have been studied.

HPLC procedures for EQ in meat meals and extruded pet foods (with fluorimetric detection), apples, skins of apples, pears, citrus fruit, and bananas, chilli powder, and paprika have been developed. For paprika, and using spectrophotometric detection, the DL has been found to be 2 μg ml^{-1}, and the RSDs for 20 and 40 μg ml^{-1}, 2.4 and 1.9%, respectively; however, with fluorimetric detection, the DL has been found to be 0.2 μg ml^{-1}, and the RSD for 0.5 μg ml^{-1}, 3.2%. EQ in apples has been also determined using electrochemical detection (DL = 0.03 ng) [69]. The levels of EQ residues in apple varieties stored under different conditions have been tabulated (with spectrophotometric detection, DL = 0.25 μg ml^{-1} in the extract

and 0.2 mg kg^{-1} in the fruit). EQ has also been determined in apples by GC (DL = 50 μg kg^{-1}, ~ 74% recovery). The use of guaiacol as an internal standard in the GC determination of phenolic antioxidants and EQ has been described.

Thiodipropionic and Thiodiglycolic Acids and their Fatty Esters

3,3′-Thiodipropionic acid, thiodiglycolic acid (or 2,2′-thiodiacetic acid) and their dialkyl esters have been separated and identified by HPLC on a C18 column with a water–acetonitrile mobile phase of pH 7.95 and spectrophotometric detection at 254 nm. The determination of thiodipropionic acid esters of fatty alcohols in polyethylene for food packaging has been performed by Soxhlet extraction with $CHCl_3$, followed by saponification with methanolic KOH and capillary GC of the liberated fatty alcohols. The distearyl thiodipropionate has also been determined by GC analysis of the compounds evolved by pyrolysis of the polymers used for packaging.

Selection of the Analytical Method for Phenols and EQ

For screening purposes, TLC on high-performance microplates has the advantages of simplicity and rapidity, while HPLC with both hydroorganic or micellar mobile phases should be recommended for screening and evaluation. Generally, gradient elution is required with hydroorganic phases, whereas solutes with large hydrophobicity differences can be separated in a short time, with good resolution and excellent reproducibility, by isocratic MLC. In relation to spectrophotometry, the DLs of phenols and EQ can be reduced by several decades using fluorimetric or voltammetric detection.

Rapid and extremely selective separations can be achieved by capillary electroseparation techniques, as CZE and MECC; however, the DLs are larger than in HPLC. To reduce the DLs, laser-induced fluorescence (LIF) can be used. Both the selectivity and sensitivity in HPLC and capillary electroseparation can be largely expanded using MS detection. Moreover, MS is excellent for structural characterization studies at the molecular level, and can be further powered by inserting an ion trap (ITMS) between the mass filter and the ion sink.

GC separations are rapid and very efficient, and the antioxidants are sensibly detected with a FID; however, before injection the hydroxyl and carboxyl groups should be derivatized. Superior selectivity and sensitivity can be achieved with MS detection. Further, many mass spectra within the elution time of a single peak can be obtained with a time-of-flight mass spectrometric detector (TOFMS). The commercial availability of portable gas chromatographs, also including miniaturized ITMS and TOFMS detectors, have increased the interest of GC.

Nonchromatographic kinetic and cyclic voltammetric approaches, although useful to resolve simple mixtures, have a limited selectivity. High selectivity and rapidity, almost without sample preparation, can be achieved with computer-aided near-infrared-spectroscopy (NIR). The poor chemical selectivity is compensated by powerful multivariant data treatment techniques. This approach is also applied to sensor-array- and MS-based electronic noses (EN), but in this case, only the volatile fraction of the sample can be analyzed. An advantage of NIR and EN is the capability of direct evaluation of 'fundamental' variables or key factors, such as specific flavors or off-flavors (including rancidity), while skipping the tedious and sometimes impracticable evaluation of all the chemical species on which a given flavor could rely.

Evaluation of the Prooxidant/Antioxidant Activity

The evaluation of the primary prooxidant/antioxidant activity in food is important to compare the performance of antioxidants in different food products under diverse conditions. In these studies, alkyl gallates, BHT, BHA and α-tocopherol are frequently used as references, to evaluate the antioxidant activity of other compounds, including vegetal extracts, but most of the procedures mentioned below can also be used to determine the antioxidants.

The prooxidant/antioxidant activity can be measured with a wide variety of methods, including the evaluation of sensory quality (color and smell) by a panel of 'rancidity'-trained experts, the spectrophotometric monitoring of the 2,2-diphenyl-1-picrylhydrazyl (DPPH), 2,2′-azinobis(3-ethylbenzothiazoline)-6-sulfonic acid, and *N*,*N*-dimethyl-*p*-phenylenediamine radicals, the bleaching of β-carotene, the inhibition of the chemiluminescence of luminol, the GC evaluation of the C18:2/C16:0 ratio and free fatty acid (particularly, linoleic acid) contents, the iodine, peroxide, and thiobarbituric acid values, and other methods which are discussed below. In spite of their promising capabilities, little work has been still published concerning the use of electronic noses in detecting and monitoring rancidification processes.

In a fast TLC test for the evaluation of the total free-radical scavenger capacity of foodstuffs, silica-gel plates were stained with a methanolic solution of the DPPH radical; this test is sufficiently sensitive

for water- and methanol-soluble products, but not for lipid-soluble compounds. In a method for the screening of complex mixtures for radical scavenging components, postcolumn mixing of the HPLC eluate with the DPPH radical was implemented, the bleaching induced by the presence of antioxidants in the eluate being detected as negative peaks at 517 nm; this simple, rapid, and quantitative method is suitable for both isocratic and gradient HPLC runs, with mobile-phase compositions ranging from 10 to 90% organic solvent, with DLs within the 0.33–94 $\mu g\,ml^{-1}$ range, depending on the compound tested.

The antioxidant activity has been evaluated in edible plant material by the inhibition capacity on the hemoglobin-catalyzed peroxidation of linoleic acid, which can be monitored spectrophotometrically. For eight synthetic antioxidants, the antioxidant activity has been shown to be linear with respect to the logarithm of their concentrations. Hydrophilic and hydrophobic antioxidants can be assayed with this method within 15 min.

Inhibition of luminol chemiluminescence has been used to measure the antioxidant capacity of lipid-soluble antioxidants in the lipid phase of systems as blood plasma and tissue homogenates (measuring range for α-tocopherol: 0.1–3 nmol). In a postcolumn HPLC procedure, radical scavengers were detected in vegetal extracts by mixing the eluate with luminol and hydrogen peroxide in the presence of peroxidase at pH 10; by monitoring the quenching of the luminiscent signal, the antioxidants were detected at the nanogram level.

The antioxidant activity of olive oils and wines has been measured with a flow-injection system with electrochemical detection. Using the 'Rancimat' instrument (Metrohm, Herisau, Switzerland) the sample is heated (under computer control), the volatile organic acids produced by lipid oxidation are collected in water, and the resulting conductivity increase is monitored. The rancidification induction time and stability index of the product are automatically obtained. Reported applications include cereals, cookies and biscuits, lard, vegetal oils, butter and margarine, walnut oil, and olive oil. The antioxidant activities of red grape marc (skin and seeds), BHA, BHT, and PG when used in protecting olive oil have been compared. Similarly, the antioxidant activity in soybean oil of extracts from grape bagasse has been evaluated.

The formation of volatile peroxides during lipid oxidation can be followed by static headspace GC. Using this technique, the ability of BHT, BHA and the natural polyphenol antioxidants from olive oil to inhibit lipid oxidation in tuna can be compared. GC with mass spectrometric detection can also be used. In addition, lipid oxidation can be monitored by headspace GC of hexanal. Capillary GC has been used to determine the oxidized fatty acids.

Finally, the free radicals generated during lipid oxidation can be monitored directly by electron spin resonance (ESR) spectroscopy. The free radicals also induce a strong ESR signal in the exposed proteins and amino acids, and simultaneously the fluorescence of the proteins increases, which has been attributed to cross-linking. When added to the proteins prior to incubation with oxidized lipids, antioxidants, including BHT, BHA, ascorbic acid and α-tocopherol, inhibit both the ESR signal and the increase in fluorescence. The ESR of free radicals has been used to compare the efficiency of OG, BHA, and plant extracts (tea, coffee, rosemary, and others) in protecting dehydrated chicken meat.

See also: **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; **Ascorbic Acid**: Properties and Determination; **Chromatography**: Principles; Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; Supercritical Fluid Chromatography; Combined Chromatography and Mass Spectrometry; **Fats**: Classification; **Food Additives**: Safety; **Legislation**: Contaminants and Adulterants; **Preservation of Food**; **Preservatives**: Analysis; **Tocopherols**: Properties and Determination

Further Reading

Anderson J and Van-Niekerk PJ (1987) High-performance liquid-chromatographic determination of antioxidants in fats and oils. *Journal of Chromatography* 394: 400–402.

Aparicio A, San-Andrés MP and Vera S (2000) Separation and determination of phenolic antioxidants by HPLC with surfactant/n-propanol mobile phases. *Journal of High Resolution Chromatography* 23: 324–328.

Aruoma OI (1996) Assessment of potential prooxidant and antioxidant actions. *Journal of the American Oil Chemists' Society* 73: 1617–1625.

Berthod A, García-Álvarez-Coque MC and Torres-Lapasió JR (1999) *Micellar Liquid Chromatography*, Chromatographic Science Series 91. New York: Marcel Dekker.

Boyce MC and Spickett EE (1999) Separation of food grade antioxidants (synthetic and natural) using mixed micellar electrokinetic capillary chromatography. *Journal of Agricultural and Food Chemistry* 47: 1970–1975.

Dapkevicius A, Van-Beek TA and Niederlander HAG (2001) Evaluation and comparison of two improved techniques for the on-line detection of antioxidants in liquid chromatography eluates. *Journal of Chromatography A* 912: 73–82.

Delgado-Zamarreño MM, Sánchez-Pérez A, González-Maza I and Hernández-Méndez J (2000) Micellar

electrokinetic chromatography with bis(2-ethylhexyl)sodium sulfosuccinate vesicles. Determination of synthetic food antioxidants. *Journal of Chromatography A* 871: 403–414.

Fukumoto LR and Mazza G (2000) Assessing antioxidant and prooxidant activities of phenolic compounds. *Journal of Agricultural and Food Chemistry* 48: 3597–3604.

González M, Ballesteros E, Gallego M and Valcárcel M (1998) Continuous-flow determination of natural and synthetic antioxidants in foods by gas chromatography. *Analytica Chimica Acta* 359: 47–55.

Goupy P, Hugues M, Boivin P and Amiot MJ (1999) Antioxidant composition and activity of barley (*Hordeum vulgare)* and malt extracts, and of isolated phenolic compounds. *Journal of the Science of Food and Agriculture* 79: 1625–1634.

Hall CA, Zhu A and Zeece MG (1994) Comparison between capillary electrophoresis and high-performance liquid chromatography separation of food grade antioxidants. *Journal of Agricultural and Food Chemistry* 42: 919–921.

He P and Ackman RG (2000) HPLC determination of ethoxyquin and its major oxidation products in fresh and stored fish meals and fish feeds. *Journal of the Science of Food and Agriculture* 80: 10–16.

Karovicova J and Simko P (2000) Determination of synthetic phenolic antioxidants in food by high-performance liquid chromatography. *Journal of Chromatography A* 882: 271–281.

Koleva II, Niederlander HAG and Van-Beek TA (2000) An on-line HPLC method for detection of radical scavenging compounds in complex mixtures. *Analytical Chemistry* 72: 2323–2328.

Ni YN, Wang L and Kokot S (2000) Voltammetric determination of butylated hydroxyanisole, butylated hydroxytoluene, propyl gallate and t-butylhydroquinone by use of chemometric approaches. *Analytica Chimica Acta* 412: 185–193.

Ni YN and Liu C (1999) Artificial neural networks and multivariate calibration for spectrophotometric differential kinetic determinations of food antioxidants. *Analytica Chimica Acta* 396: 221–230.

Noguera-Ortí JF, Villanueva-Camañas RM and Ramis-Ramos G (1999) Determination of phenolic antioxidants in vegetable and animal fats without previous extraction by dilution with n-propanol and micellar liquid chromatography. *Analytica Chimica Acta* 402: 81–86.

Page BD (1993) Liquid-chromatographic method for the determination of nine phenolic antioxidants in butter oil: collaborative study. *Journal of the Association of Official Analytical Chemists International* 76: 765–779.

Yankah VV, Ushio H, Ohshima T and Koizumi C (1998) Quantitative determination of butylated hydroxyanisole, butylated hydroxytoluene, and tert-butyl hydroquinone in oils, foods, and biological fluids by high-performance liquid chromatography with fluorometric detection. *Lipids* 33: 1139–1145.

Role of Antioxidant Nutrients in Defense Systems

S B Astley, Norwich Research Park, Norwich, UK

Background

In the body, nutrients such as the carotenoids and vitamins C and E have been suggested to act as antioxidants. Elements, such as selenium, and plant polyphenols, including the flavenoids, have also been shown to be essential in maintaining the body's antioxidant defense systems or act as antioxidants (**Table 1**).

Oxygen Paradox

Aerobic organisms exploit the inherent reactivity of the oxygen molecule to oxidize food constituents efficiently and provide energy for life. Compared with anaerobes, this form of metabolism gives a selective advantage. Reactive oxygen species (ROS), including free radicals, are normal products of an oxidative metabolism. Moderate levels of ROS not only participate in critical physiological pathways, such as signaling cascades and gene expression, but also are responsible for adjusting their rate of reaction. Under normal conditions, the potential toxicity of ROS is kept in check by a highly regulated antioxidant defense system. Aerobes are, however, subject to oxidative stress – when the rate of ROS production outstrips antioxidant defense mechanisms and the oxidant:antioxidant balance is disturbed in favor of the former. Physiological stress, aging, and environmental sources such as polluted air and cigarette smoke can hasten the imbalance.

ROS Damage and Cell Death

The unstable and highly reactive nature of ROS means that severe oxidative activity can wreak havoc within cells and tissues. Such damage may lead to cell death (apoptosis), which is a natural control mechanism that ensures the removal of damaged cells. High levels of oxidative damage that result in cell death are, however, of far less significance to an organism as a whole than oxidative stress, which can cause changes in the structural and functional integrity of a cell. ROS can damage DNA and other cellular components (e.g., lipids, proteins, and carbohydrates). Such damage has been linked to changes that accompany aging (e.g., age-related macular degeneration) and human disease processes (e.g., cancer and cardiovascular disease) (**Table 2**).

Table 1 Compounds with a role in human antioxidant defense systems

Antioxidants	*'Other role' in antioxidant defense*
Water-soluble	*Cell–cell communication, one-carbon metabolism and redox potential*
Vitamin C	Retinoids
Glucosinolates	Folates
Simple and complex phenols including:	B_6
anthocyanins	B_{12}
flavonols, flavones, and flavanones	Riboflavin
chalcones and dihydrochalcones	Thioredoxin
isoflavones and lignans	Coenzyme Q
hydroxybenzoic acids	Tyrosine
proanthocyanidins and tannins	
ellagitannins, oleuropein, and related compounds	
chlorogenic acids and other cinnamates	
Glutathione and its precursors	
Cysteine	
Glucose	
Pyruvate	
Uric acid	
Bilirubin	
Ubiquinol-10	
Quinones	
Lipid-soluble	*Essential metal ions*
Vitamin E (a, b, g, d tocopherol and trienols)	Selenium
Provitamin A carotenoids	Copper
Nonprovitamin A carotenoids	Iron
Lipoic acid	Zinc
Stilbenes	Manganese
Sterols	

Epidemiology

Epidemiological studies have shown that individuals who eat diets rich in fruit and vegetables are at significantly less risk of developing diseases associated with oxidative stress and, in particular, excess free radical production. Fruits and vegetables contain a myriad of putative health-promoting, disease-preventing compounds. For example, folic acid is necessary for accurate DNA replication and decreases plasma homocysteine, which is a risk factor for cardiovascular disease. Fiber increases the bulk of colonic contents, decreases transit time, and alters the type and metabolism of gut bacteria, all of which decrease the risk of colorectal cancers. Other components that demonstrate antioxidant capacity *in vitro* include vitamins C and E, the carotenoids, and plant polyphenolics. While the overwhelming evidence for a protective effect derived from higher amounts of fruits and vegetables in the diet is accepted, the factors responsible for such an effect remain hotly debated. It is likely that all the suggested components elicit a positive, and interactive, dose-related response, although the issue of benefit/risk should not be forgotten, particularly in relation to supplementation and fortification.

If food-derived components are effective in reducing disease risk, they may not act solely via an antioxidant mechanism – scavenging free radicals. However, since oxidative damage to biomolecules is prevalent, it is reasonable to assume that antioxidants or compounds that preserve antioxidant function and prevent or reduce such damage will be beneficial.

Vitamin C

The recommended daily allowance of vitamin C is 40 mg per day according to the UK Dietary Reference Value, or 60 mg per day, according to the US RDA. It is essential for the prevention of scurvy and to aid healing of wounds and fractures. Vitamin C also assists in the absorption of nonheme iron. Deficiency symptoms include bruising, capillary weakness, soft and bleeding gums, slow-healing wounds and fractures, swollen or painful joints, nosebleeds, tooth decay, loss of appetite, muscular weakness, skin hemorrhages, frequent infections, and anemia. Good sources are citrus fruits and juices, strawberries, vegetables.

Figure 1 shows the structure of vitamin C in its different forms. Vitamin C, or ascorbic acid, is a water-soluble crystalline solid that is one of the most labile nutrients in the diet; it is easily destroyed by oxygen, metal ions, increased pH, heat, and light. All

Table 2 Diseases involving reactive oxygen species and/or free radicals[a]

Primary single organ involvement
Cardiovascular system
Atherosclerosis
Alcohol cardiomyopathy
Brain
Stroke/cerebrovascular/traumatic injury
Dementias and amyloid diseases
Demyelinating diseases
Ataxia-telangiectasia syndrome
Aluminum overload
Lung
Effects of cigarette smoking
Inhalation of oxidant pollutants
Emphysema
Hyperoxia
Acute respiratory distress syndrome
Pneumoconiosis
Skin
Solar radiation and photosensitivity
Thermal injury
Porphyria
Contact dermatitis
Eye
Cataractogenesis
Age-related macular degeneration
Degenerative retinal damage
Ocular hemorrhage
Retinopathy
Joint abnormalities
Rheumatoid arthritis
Gastrointestinal tract
Endotoxin liver injury
Carbon tetrachloride liver injury
Free fatty acid-induced pancreatitis
Nonsteroidal antiinflammatory drug-induced lesions, abetalipoproteinemia
Kidney
Nephrotic membrane diseases
Aminoglycoside and heavy metal nephrotoxicity
Renal graft rejection
Erythrocytes
Lead poisoning
Protoporphyrin photooxidation
Malaria
Sickle-cell and other anemias
Multiorgan involvement
Cancer
Aging
Disorders of 'premature aging'
Premature retinopathy
Immune deficiency
Inflammatory-immune injury
Glomerulonephritis (idiopathic, membranous)
Autoimmune diseases
Ischemia-reflow states
Drug and toxin-induced reactions
Iron overload
Idiopathic hemochromatosis
Dietary iron overload
Nutritional deficiencies
Keshan disease (selenium deficiency)
Kwashiorkor
Vitamin E deficiency
Alcohol damage

[a]ROS and free radicals may be the sole cause of only a few but may make some conditions worse or prevent the normal action of the body's healing processes.

animals except humans, other primates, guinea-pigs, fruit bats, and some bird species are able to synthesize ascorbic acid from glucose.

Aqueous solutions of vitamin C are stable unless transition metal ions are present. Copper and iron both catalyze the oxidation of ascorbic acid, utilizing molecular oxygen, to produce hydrogen peroxide and hydroxyl radicals (Fenton chemistry/Haber–Weiss reaction).

Role in the Body

Vitamin C was first isolated in 1928. It is necessary for the synthesis of the body's cementing substances: bone matrix, collagen, dentin, and cartilage, and is needed as a cofactor for several enzymes. These enzymes include proline and lysine hydroylases, which are involved in the biosynthesis of collagen, and the copper-dependent enzyme dopamine-β-hydroylase, which converts dopamine into norepinephrine.

Deficiency of vitamin C results in scurvy, the symptoms of which are largely related to inadequate collagen synthesis and defective formation of intercellular materials due to reduced hydroxylation. Vitamin C was identified as the curative agent for scurvy in 1932. The symptoms of scurvy are usually not seen for 3–6 months in the absence of any dietary vitamin C (< 1 mg per day) because vitamin C is metabolized slowly in humans.

Smokers have an increased turnover of vitamin C, which means that in order to maintain their body pool and circulating levels at a concentration equal to that of a nonsmoker, their intake should be about 80 mg per day, compared with the 40–60 mg required by a healthy nonsmoker.

Vitamin C promotes uptake of nonheme iron by reducing Fe(III) to Fe(II), and inhibits the activity of nitroso compounds by maintaining them in their reduced inactive form. This ability to act as a reducing agent (electron donor) may also be responsible for the antioxidant role of vitamin C.

Antioxidant Role of Vitamin C

Oxygen-based reactive species are reduced by donation of one electron from vitamin C/ascorbic acid. During this process, the semidehydroascorbate

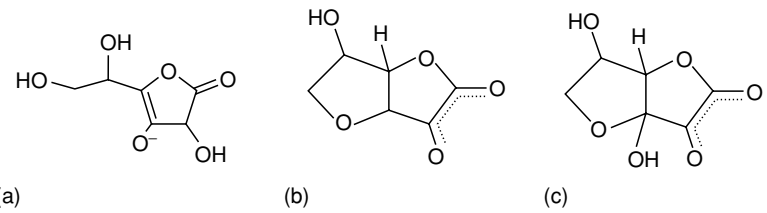

Figure 1 (a) Structure of vitamin C; (b) dehydroascorbate radical; and (c) dehydroascorbic acid.

Table 3 Plasma proteins that bind metals or iron complexes

Protein (plasma concentration)	*Action*
Ceruloplasmin (0.18–0.4 g l^{-1})	Ferroxidase activity
	Inhibits iron and copper-dependent lipid peroxidation
	Reincorporates iron mobilized from the iron storage protein ferritin back to ferritin
	Does not react directly with peroxyl radicals
Transferrin (1.8–3.3 g l^{-1})	Binds iron and stops or slows down its participation in lipid peroxidation and iron-catalyzed Haber–Weiss reactions
	Can take up iron released from ferritin, thereby inhibiting ferritin-dependent lipid peroxidation
Lactoferrin (200×10^{-6} g l^{-1})	Action similar to that of transferrin
Albumin (50 g l^{-1})	Binds copper tightly and iron weakly
Haptoglobin (0.5–3.6 g l^{-1})	Binds free hemoglobin (Hb) and met-Hb, thereby preventing these proteins from catalyzing lipid peroxidation
	Transferrin and hemopexin do not inhibit Hb-mediated lipid peroxidation
Hemopexin (0.6–1.0 g l^{-1})	High affinity for heme ($K_d < 10^{-13}$ mol l^{-1})
	Unlike transferrin and haptoglobin, hemopexin inhibits heme-mediated lipid peroxidation

radical is formed, which, despite being a free radical, is not very reactive (Figure 1). Semidehydroascorbate can undergo disproportionation to reform ascorbic acid and dehydroascorbate. Alternatively, it may be oxidized directly to form dehydroascorbate. Dehydroascorbate is highly unstable and breaks down, via a series of complex reactions, to give oxalic and L-threonic acids. Dehydroascorbate and its breakdown products are, however, toxic and the more usual reaction is the conversion of dehydroascorbate to ascorbic acid. This reaction is catalyzed by the enzyme dehydroascorbate reductase and oxidizes two glutathione molecules in the process. The same enzyme may also recycle the semidehydroascorbate radical. Both vitamin C and dehydroascorbic acid play a pivotal role in protecting lipids against oxidative damage by regenerating vitamin E.

Vitamin C reacts rapidly with superoxide and its derivatives, and even more quickly with hydroxyl radicals. It also scavenges singlet oxygen, and reduces thiyl radicals and hypochlorous acid. The human lens is rich in vitamin C, whereas levels of naturally occurring antioxidant enzymes (e.g., superoxide dismutase) are low. It may, therefore, protect the eye from oxidative damage, in particular the damage caused by UV that is mediated by singlet oxygen.

Prooxidant Potential?

Vitamin C has a potentially duplicitous action. Hydrogen peroxide and hydroxyl radicals produced during oxidation of vitamin C, catalyzed by transition metals, could be responsible for subsequent oxidative damage to the cell and cellular components including DNA.

The sequestration of metal ions *in vivo* by proteins (e.g., transferrin – Table 3) means that the prooxidant role of vitamin C is less likely. It is, however, not impossible because the binding of the metal ions, by the various proteins, does not render them entirely inactive. Furthermore, in the event of physical damage (e.g., bruising) heme iron may be released from erythrocytes, which is available for participation in Fenton chemistry. Oxidative damage caused by the vitamin C in plasma may be a real danger to surrounding tissues.

Risk:Benefit

Megadoses of vitamin C, synthetic or natural, have previously been considered beneficial or, at the very least, harmless. However, there is mounting evidence that doses as small as 500 mg per day may result in increased levels of DNA damage and damage to the lining of major arteries, which subsequently leads to an elevated risk of stroke.

Vitamin E

There are no UK dietary reference values available for vitamin E, but the US RDA is 30 IU. Vitamin E is the major lipid-soluble antioxidant nutrient. Deficiency symptoms are very rare, and usually only occur as a result of fat malabsorption syndromes. Good food sources include cereals, nuts and seeds, wheat germ, vegetable oils, and 'spreads,' which contain vitamin E as an antioxidant.

Chemistry

Vitamin E (Figure 2) is an essential nutrient, which happens to function as an antioxidant. The term 'vitamin E' is generic, and includes all entities that exhibit the biological activity of α-tocopherol.

In nature, eight substances have been found to have vitamin E activity: α-, β-, γ- and δ-tocopherol, and α-, β-, γ- and δ-tocotrienol. The acetate and succinate derivatives of the natural tocopherols also have vitamin E activity, as do the synthetic tocopherols and their acetate and succinate derivatives. Of these, D-α-tocopherol (*RRR*-α-tocopherol) has the highest bioavailability and is the standard against which others are compared.

Natural and synthetic vitamin E are not equivalent in composition, structure, or bioavailability. Natural vitamin E (*RRR*-α-tocopherol or D-α-tocopherol) is a single entity, whereas synthetic vitamin E (all-rac-α-tocopherol or DL-α-tocopherol) is a mixture of eight stereoisomers in equal amounts. The other seven stereoisomers have different molecular configurations and lower biological activities.

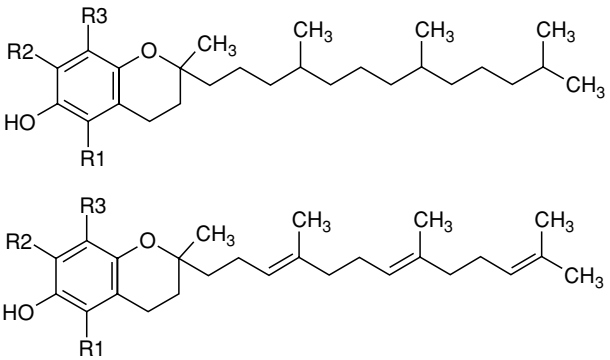

Compound	R1	R2	R3
Alpha (α)-tocopherol	CH_3	CH_3	CH_3
Beta (β)-tocopherol	CH_3	H	CH_3
Gamma (γ)-tocopherol	H	CH_3	CH_3
Delta (δ)-tocopherol	H	H	CH_3
Alpha (α)-tocotrienol	CH_3	CH_3	CH_3
Beta (β)-tocotrienol	CH_3	H	CH_3
Gamma (γ)-tocotrienol	H	CH_3	CH_3
Delta (δ)-tocotrienol	H	H	CH_3

Figure 2 Tocopherols and tocotrienols.

The richest natural sources of vitamin E are vegetable oils, nuts, and whole grains. Wheatgerm oil is the single richest source, whereas foods of animal origin are generally low in vitamin E. In general, the amount of vitamin E (as D-α-tocopherol) supplied by a normal diet ranges from 6 to 13 mg per day (1 IU = 0.67 mg). In most countries, the estimated average intake is below 10 mg per day. During refining, processing, and storage, all foods containing vitamin E lose some of their activity. Vitamin E in foods is destroyed by oxygen, freezing causes degradation, and, like vitamin C, the oxidative process is accelerated by heat and light.

Role in the Body

Those compounds with vitamin E activity have a number of roles in human metabolism that include (1) acting as an antioxidant to protects cells and other body components from free radical attack; (2) stimulation of the immune response; (3) reducing the severity of prostaglandin-mediated disorders such as inflammation, premenstrual syndrome, and circulatory irregularities (nocturnal leg cramps and blood platelet adhesion); (4) inhibition of the conversion of nitrites in smoked, pickled, and cured foods to nitrosamines in the stomach; and (5) regulation of gene expression in particular the proliferation of smooth muscle cells, which are involved in the development of atherosclerosis.

Most vitamins function as cofactors for enzymatic reactions, but vitamin E has no such role. Requirements for, and utilization of, vitamin E vary according to an individual's oxidative status. Furthermore, unlike other vitamins, deficiency of vitamin E does not produce a disease with rapidly developing symptoms. The absence of vitamin E only produces overt deficiency symptoms in cases involving fat malabsorption syndromes, premature infants, and patients on total parenteral nutrition. However, there are noticeable effects of inadequate vitamin E intake, which develop over a long time, typically decades, and have been linked to the development of degenerative diseases.

Antioxidant Role

Vitamin E is present in all lipids in the body, most significantly in lipids of cell membranes and in circulating low-density lipoproteins (LDL – the so-called 'bad' cholesterol). It is recognized to be the major fat-soluble antioxidant in lipid tissue, and the primary defense against lipid peroxidation – oxidation of fats.

When a free radical reacts with lipid in a cell membrane, the resulting chain reactions oxidizing surrounding lipid (lipid peroxidation) damage both the structure and function of cell membranes. Vitamin E is particularly important in tissues that contain relatively high levels of polyunsaturated fatty acids (e.g., brain and central nervous system), those in direct contact with high concentrations of oxygen (e.g., lung) and cell organelles dealing with oxygen or oxygen-free-radical reactions (e.g., mitochondria).

Epidemiology

Epidemiological studies provide data supporting the benefits of higher intakes of foods containing vitamin E as well as other antioxidants. These studies suggest that individuals with low blood concentrations and/or intakes of foods containing vitamin E have increased risk for development of a number of degenerative diseases including certain types of cancer, heart disease, Alzheimer's disease, and cataract. However, as yet, there is no proof that the vital component in the prevention of these diseases is anymore likely to be vitamin E than any of the other compounds contained in such foods.

A limited number of intervention trials have been completed, with mixed results. However, for the most part, such trials have used isolated natural or synthetic vitamin E, and there is increasing evidence that these compounds do not interact with the body in the same way as food-sourced vitamin E (e.g., reduced bioavailability).

Source and Dose

How much vitamin E an individual needs depends on what is required; enough to prevent deficiency disease, prevent oxidation of lipids, prevent free radical damage and promote optimal health, or to produce a therapeutic effect. Deficiency *per se* can be prevented by intakes in line with government recommended dietary intakes. However, if prevention of oxidative damage and promotion of optimal health is the objective, intakes of up to 10 times the current levels have been proposed. If a therapeutic result is the aim, even higher intakes may be necessary.

Research data suggest that considerably more vitamin E needs to be consumed in the average diet to prevent free radical cell damage. Foods are depleted of vitamin E by processing, refining, and storage, and the current trend to reduce fat intake means that a diet-conscious population tends to eliminate those oil-rich foods likely to provide vitamin E. Not only do most people not get enough vitamin E in their diets for optimal health, but a large number are considered to be marginally deficient of the amount necessary to prevent deficiency disease.

Carotenoids

No dietary reference values are available for the recommended daily allowance. Those with a β-ionone ring are precursors for vitamin A.

No deficiency symptoms are known except those associated with vitamin A deficiency. Good food sources are colorful fruit and vegetables (e.g., peas, tomatoes, and carrots) and commercial products such as 'spreads' and milkshake flavoring, which contain β-carotene as an antioxidant.

Links with Vitamin A

In 1913, it was shown that a fat-soluble factor in butter – later characterized as vitamin A – stimulated the growth of rats fed on an incomplete diet. Vitamin A is a fat-soluble vitamin that helps in the formation and maintenance of healthy teeth, skeletal and soft tissue, mucous membranes, and skin. It is also known as retinal because it generates the pigments that are necessary for the functioning of the retina. However, it is not an antioxidant.

Chemistry

Carotenoids are naturally occurring brightly colored plant pigments. More than 600 different carotenoids have been identified, all derived from the same basic C40 isoprenoid skeleton (**Figure 3**).

Mixed carotenoids can be isolated from algae, β-carotene from palm oil, lutein and lutein esters from marigolds, lycopene from tomatoes, and cryptoxanthin from citrus and tropical fruits, in particular orange fruit. β-Carotene has been synthesized for a number of years, and lycopene more recently.

Natural mixed carotenoid supplements contain a variety of carotenoids, including α- and β-carotene. Commercially available synthetic carotenoid supplements contain only β-carotene. Synthetic β-carotene is composed of the all-*trans* isomers, and β-carotene from algae, and fruits, and vegetables is composed of all-*trans*, 9-*cis*, and other *cis*-isomers.

Role in the Body

Humans regularly consume around 40 different carotenoids, but only a small number of these can be detected in appreciable quantities in human blood and tissues. The major carotenoids are α- and β-carotene, lutein, zeaxanthin, cryptoxanthin, and lycopene.

Provitamin A carotenoids, those carotenoids possessing a β-inone ring, have vitamin A activity, and can be converted into vitamin A by the body. β-Carotene has the highest provitamin A activity, but over 60 carotenoids have some provitamin A activity. Vitamin A is toxic when taken in excess, but these

carotenoids are safe sources because they are only converted to vitamin A as and when the body needs it. The other biological roles of carotenoids, including their antioxidant functions, are completely independent of their provitamin A activity. However, the carotenoids are not recognized as essential nutrients, and there are no recommended dietary intakes. The only exception is a recommended dietary intake for vitamin A of 800–1000 retinol equivalents, where one retinol equivalent is 6 mg of β-carotene or 12 mg of another provitamin A-carotenoid.

Role of Antioxidants

Despite the similarities in their structures, the carotenoids have diverse biological functions and potentially different roles within the body. The carotenoids are effective quenchers of singlet oxygen, with lycopene exhibiting the highest singlet oxygen-quenching activity. They are able to function as chain-breaking antioxidants, much like vitamin E, and may protect cells and cellular components from oxidative damage. The carotenoids will protect lipids from oxidation as well but only in the absence of vitamin E, which is far more abundant. Importantly, the carotenoids exhibit the ability to inhibit oxidative damage at different oxygen tensions. At high oxygen tensions, such as those in the lungs, lycopene is more effective then β-carotene, which in turn is more efficient at low oxygen concentrations.

Carotenoids enhance the immune response and protect the skin from redness and damage following exposure to UV radiation. Certain carotenoids (including lycopene and β-carotene) have the ability to enhance cell–cell communication by increasing the exchange of growth regulatory signals. Increased communication between normal cells and damaged cells encourages the latter to undergo apoptosis (commit suicide) and protects the body from proliferation of initiated potentially precancerous cells. Some carotenoids are more effective in tumor suppression in different cultured cell lines, and it has been suggested that others protect against specific cancer types. For example, increased consumption of tomatoes and tomato products (lycopene) is correlated significantly with a decrease in incidence of prostate cancer and foods containing lutein, zeaxanthin, and α- and β-carotene with a reduced risk of lung cancer. Lutein and zeaxanthin are oxycarotenoids (i.e. they contain an oxygen atom) and are the only carotenoids present in the macular region of the retina. These carotenoids are linked to the normal function of the macula, which is responsible for sharp and detailed vision, and are thought to serve as filters for harmful blue light in the macula and as scavengers of singlet oxygen in retinal tissues.

Epidemiology

The majority of data examining carotenoid intakes show that individuals with low carotenoid intakes or low blood carotenoid levels are at a greater risk of developing degenerative diseases associated with aging. High intakes of carotenoid-rich fruits and

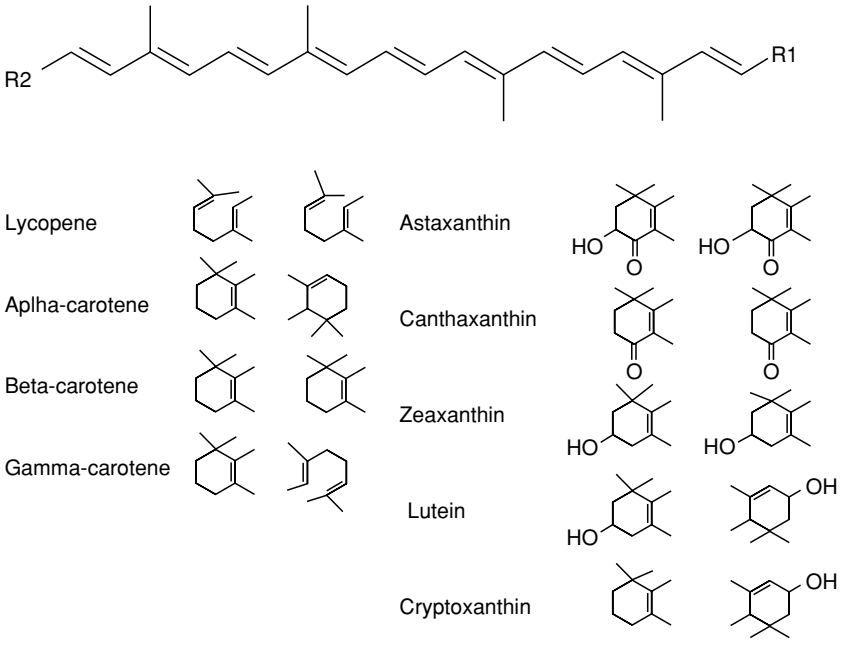

Figure 3 Carotenoids and oxycarotenoids.

vegetables and high carotenoid blood levels are associated with a lower risk of cancer and cardiovascular disease, which may refute the previously held conviction that only vitamin E had the potential to ameliorate risk. In addition, studies have shown an inverse correlation between levels of the oxycarotenoids lutein and zeaxanthin and risk of cataracts and age-related macular degeneration, although there is no evidence that increased intake can reverse existing damage. For the most part, research into the carotenoids has focused on β-carotene, because it is the best known, most abundant, and has provitamin A activity, but recent studies have suggested that lutein, zeaxanthin, lycopene, and cryptoxanthin should also be considered.

Source and Dose

Research suggests that 9-*cis* β-carotene is the most efficient antioxidant and that food sources are better able to increase antioxidant protection than natural isolates. As with vitamin E, the amount of carotenoids required by an individual will depend on the overall aim (i.e., to prevent vitamin A deficiency, prevent oxidation of lipids, prevent free radical damage or promote optimal health). If prevention of vitamin A deficiency is the objective, quantities of provitamin A carotenoids to meet the RDA for vitamin A are adequate for most individuals in developed countries. If prevention of oxidative damage and promotion of optimal health is the objective, higher intakes of carotenoids may be necessary.

The richest dietary sources of carotenoids are brightly colored fruits and vegetables. Currently, the estimated daily intake from a normal diet ranges from 1.5 to 3 mg for β-carotene and from 3.6 to 7.6 mg for total carotenoids. Factors influencing carotenoid intake levels include differences in varieties, variable growth and harvesting conditions, postharvest handling and processing methods, exposure to light and oxygen, and cooking methods. On the whole, mild cooking of food before ingestion improves the bioavailability of carotenoids, because the plant cell wall structure is partially broken down, increasing the ability of the gut to further digest the plant material and release the carotenoids from the cellular structures with which they are intimately associated.

Risk:Benefit

Results from intervention trials using a large dose of synthetic β-carotene supplements over a long period of time, often in combination with vitamin E, have been neutral or negative, suggesting an increase in both cancers and events related to cardiovascular disease. In contrast, studies with other specific carotenoids, mixed carotenoids, or foods with high carotenoid contents have provided neutral or positive results.

The carotenoids are not carcinogenic, mutagenic, teratogenic, or cytotoxic when tested using typical toxicological tests. Doses of 20–180 mg of β-carotene per day, for many years, have been used to treat erythropoietic protoporphyria, with no evidence of toxicity or development of abnormally elevated blood vitamin A levels. Based on current information, therefore, an increased intake of a mixture of natural carotenoids, preferably from food rather than in the form of an isolate or β-carotene alone, may be of greatest benefit in disease prevention.

See also: **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; Synthetic Antioxidants, Characterization and Analysis; **Bioavailability of Nutrients**; **Cancer**: Epidemiology; **Carotenoids**: Occurrence, Properties, and Determination; **Cholecalciferol**: Properties and Determination; **Cobalamins**: Properties and Determination; **Nucleic Acids**: Properties and Determination; Physiology; **Retinol**: Properties and Determination; **Tannins and Polyphenols**; **Tocopherols**: Properties and Determination; **Vitamin K**: Properties and Determination; Physiology

Further Reading

Block G, Patterson B and Subar A (1992) Fruit, vegetables and cancer prevention: A review of the epidemiological evidence. *Nutrition and Cancer* 18: 1–29.

Britton G, Liaaen-Jenson S and Pfander H (1995) *Carotenoids*, vols. 1A, 1B & 2. Boston, MA: Birkhauser Verlag.

Frei B (1994) *Natural Antioxidants in Human Health and Disease*. London: Academic Press.

Fund WCR (1997) *Food, Nutrition and the Prevention of Cancer: A Global Prospective*. Washington, DC: American Institute for Cancer Research.

Gutteridge JMC and Halliwell B (1995) *Antioxidants in Nutrition, Health and Disease*. Oxford: Oxford Academic Press.

Halliwell B and Gutteridge JMC (1999) *Free Radicals in Biology and Medicine*, 3rd edn. Oxford: Clarendon Press.

Institute of Medicine (2000) *Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids: A Report of the Panel on Dietary Antioxidants and Related Compounds*. London: Academic Press.

Sies H (1985) Oxidative stress. In: Sies H (ed.) *Oxidative Stress: Oxidants and Antioxidants*, 1st edn. London: Academic Press.

Appetite *See* **Satiety and Appetite**: The Role of Satiety in Nutrition; Food, Nutrition, and Appetite

APPLES

L J Skog and C L Chu, University of Guelph, Ontario, Canada

Introduction

The apple belongs to the genus *Malus*, and is a member of the rose family, Rosaceae. There are about 25 species of *Malus*; however, most horticulturists refer to *Malus domestica* as the cultivated apple. This article discusses the global distribution, varieties, commercial importance, morphology, nutritional composition, handling, storage, and industrial uses of this crop.

Origin and Global Distribution

The domestic apple of the west appears to have arisen on the northern slopes of the Tien Shan range roughly on the border of what is now Kazakhstan and the Xinjiang province of China. This wild apple is called *M. sieversii*. From this species, it seems, have emerged the cultivated apples of our supermarkets, *M. domestica* and *M. pumila*. Several crabapples and other *Malus* species are considered native to China, India, and North America. Although some species are not hardy, the domestic apple (especially *M. pumila*) is one of the hardiest of temperate-zone fruits. Apple seeds require more than 60 days of cold temperature before germination. In addition, its winter chilling requirement for apple trees to break rest dormancy (1000–1600 h) is high enough that it does poorly in latitudes with warm winters. Apples are grown in most temperate climates and many other countries like Argentina, New Zealand, and South Africa. Apples are also grown at the higher altitudes of some countries with less temperate climates such as Mexico. Production statistics for various areas of the world are presented in Table 1.

Varieties

Variability in the wild species has been used by breeders who use parents with desired traits when crosses are made. When seeds from a cross are planted, the resultant plant is a clone. Thus, breeders select individual plants with desired characteristics rather than selecting a population of plants.

The general objectives of apple-breeding programs are as follows: (1) to enhance yields; (2) to improve fruit quality, which includes factors such as flavor, firmness, or other fruit characteristics; (3) to produce plants adapted to specific environmental or cultural conditions; (4) to develop varieties resistant to diseases and insects; and (5) to develop varieties with specific requirements, such as suitability for processing, or extra firmness for long-distance shipping and improved storage and shelf-life.

New apple varieties are continuously being introduced: worldwide, there are more than 7000 varieties, although most are not currently grown commercially. Some of the more important varieties with their uses are listed in Table 2. Apple characteristics vary greatly with variety, and variety selection is a matter

Table 1 Global distribution of apple production by geographical area, including a listing of countries producing 1.0×10^6 metric tonnes (Mt) or more

Global distribution by geographical area and countries within any one area producing 1.0×10^6 Mt or more	*Production (Mt)*
World total	60 831 349
Africa	1 614 736
North and Central America	5 794 038
USA	4 843 000
South America	3 292 391
Argentina	1 116 500
Brazil	1 160 475
Asia	33 591 517
China	22 887 600
India	1 580 000
Iran	2 200 000
Turkey	2 500 000
Europe	15 710 167
France	2 140 000
Germany	1 800 000
Italy	2 155 600
Poland	1 450 376
Russian Federation	1 200 000
Ukraine	1 325 000
Oceania (includes Australia and New Zealand)	828 500

Data from Food and Agriculture Organization of the United Nations (2000) http://apps.fao.org.

Table 2 Selected apple varieties and their characteristics and uses

Variety	*Flavor*	*Storage life*	*Uses*	*Remarks*
Braeburn	Sweet	Excellent	Cooking, fresh	
Cortland	Slightly tart	Good	Cooking, fresh	White flesh that is slow to darken when cut
Cox's orange	Slightly tart	Good	Cooking, fresh	Orange/red; yellow flesh
Empire	Sweet/slightly tart	Good	Fresh	
Fuji	Sweet	Good	Cooking, fresh	
Gala	Slightly sweet	Short	Cooking, fresh	Orange/red; yellow flesh
Golden delicious	Sweet	Excellent	Cooking, fresh	
Granny Smith	Sweet/tart	Excellent	Cooking, fresh	
Haralson	Tart	Excellent	Cooking, fresh	White, firm flesh
Honeygold	Slightly sweet	Good	Cooking, fresh	
Idared	Slightly sweet	Good	Cooking, fresh	
Jonadel	Slightly sweet	Good	Cooking, fresh	Red fruit
Jonafree	Slightly tart	Good	Cooking, fresh	
Jonagold	Slightly tart	Good	Cooking, fresh	Large yellow fruit with red striping
Jonathan	Slightly tart	Moderate	Cooking, fresh	
Liberty	Slightly tart	Good	Fresh	Greenish yellow with red stripes
McIntosh	Slightly tart	Moderate	Cooking, fresh	Bright red
Mutsu	Slightly tart	Excellent	Cooking, fresh	Yellow-green fruit
Northern spy	Tart	Excellent	Cooking	
Paulared	Slightly tart	Short	Fresh	
Red delicious	Sweet	Excellent	Cooking, fresh	
Rome beauty	Slightly tart	Good	Cooking	
Spartan	Slightly tart	Good	Cooking, fresh	

Primary source: Naeve L (1997). Apple varieties and their uses. *Horticulture and Home Pest News* IC-477(23): 138–139.

of personal preference and intended use. In general, for fresh eating or dessert apples, North Americans prefer mildly sweet, crisp apples, Europeans favor tarter apples with high flavor and firmness, and Asians and people of the Middle East select very sweet apples. Some Japanese growers will leave apples on the tree until the apples become extremely sweet with an almost syrupy center (known as 'water core'). These apples would be considered overripe and unsaleable in most European and North American markets.

In North America, six varieties account for the majority of the crop. In the USA, Red delicious, Golden delicious and Granny Smith account for 65% of the apples grown. In Canada, McIntosh is the most popular, followed by Red delicious and Spartan. In both countries Gala and Fuji are being increasingly planted in response to higher world prices for new apple varieties.

Golden delicious is the number-one variety in Europe, with many clones being selected from this parent, although plantings are decreasing. Red delicious is popular, but some areas in Europe have difficulty growing it due to susceptibility to frost and physiological storage diseases and it too is experiencing decreased plantings. Granny Smith is very popular and increasing slightly in demand. Rome beauty and Cox's orange were quite popular but plantings are now in decline. In Belgium, Germany, and the Netherlands, Jonagold is the most popular variety. Gala is increasing in demand in Europe, as in North America. In 8 years, production of Gala increased by 1200% in the European Union (EU). Braeburn, which is the most popular variety in New Zealand, its country of origin, is also the favorite variety in the UK and Germany. World demand for Fuji has encouraged increased plantings, although there are difficulties finding suitable growing conditions. In China, approximately 40% of the crop consists of the cultivar Fuji.

Commercial Importance

The Food and Agriculture Organization lists apple production statistics for 88 countries. Thirteen countries – USA, Argentina, Brazil, China, India, Iran, Turkey, France, Germany, Italy, Poland, the Russian Federation, and the Ukraine – each produce in excess of 1 million metric tonnes (Mt) annually. China is by far the world's largest producer of apples, accounting for more than 50% of the world's 6.0×10^7 Mt annual production (Table 1). This worldwide production would have a farm-gate value of over $\$3.0 \times 10^{10}$ (US) based on their value in North America. This return would represent $\$0.50\,kg^{-1}$ average price for fresh-market fruit. Return varies greatly with variety depending upon demand and availability. In North America, per capita consumption averages 10 kg for fresh consumption, with an additional 12 kg for processed products (including juice).

In China, much of the best quality Fuji is destined for East Asian markets. In the last 10 years, apple production in China has quadrupled and is expected to double again in the next few years. Approximately 2.5×10^6 ha of land is devoted to apple production, compared to 3×10^5 ha for all of North and Central America. Ten percent of the crop will be processed. Several other Asian countries also produce large quantities; however, distribution is primarily limited to internal markets. Storage facilities are also limited in some Asian countries.

Central and Eastern Europe contains about 83% of Europe's total land mass but produces less than 50% of its apples. Climate and economic conditions vary greatly throughout this area, resulting in a wide variation of quality, production, and return. Newer varieties are not available in most areas. Returns also vary greatly, ranging from 100 to 150% higher than production costs in countries with extensive cultivation, down to 20–40% in less developed countries. Poland, Slovenia, and Hungary produce sufficient crop to permit export. Storage facilities are limited in most countries with only 20–30% of the fruit stored. Storage is often in natural storages as some lack cold and controlled atmosphere (CA) storages.

Apple production in western Europe has been at a surplus for some time. In 1997, almost 50% of apples in storage were unsold as of 1 December. Changing popularity of varieties and increased imports have partially accounted for this surplus. The EU has been directing orchard removals. Orchard removal has been particularly evident in France, which has accounted for 30% of European orchard removals.

Apples are a major fruit crop in North America, although returns are low for some of the traditional varieties and growers are shifting to the more lucrative varieties. Production in North America is concentrated in three main areas: (1) the eastern section, including the north-eastern and Atlantic states and provinces; (2) the central USA (Ohio to Arkansas); and (3) the western section from Colorado to British Colombia.

Morphology

The apple is a pome fruit which develops from the inferior ovary of the flower. Specifically, it develops from the ovary wall and the floral tube (the base parts of the stamens, sepals, and petals). During maturation, the tube fuses to the ovary wall, expands, becomes fleshy and ripens. The fleshy mesocarp is the edible portion of the fruit. Five cavities are also produced which generally contain two seeds each.

Apples differ greatly in their external characteristics. When ripe, they may be green, yellow, or red depending upon variety. Size, shape, sugar content, acidity, and firmness all vary with variety. Apples may be classified into three categories depending upon the concentration of acid (calculated as malic acid) and phenolic compounds or tannins. Bitter-sweet apples contain greater than 0.2% (w/v) tannins and less than 0.45% (w/v) acidity. Sharp apples have less than 0.2% tannins and more than 0.45% acidity. Sweet apples have less than 0.2% tannins and less than 0.45% acid.

The physiological characteristics of apples, such as the waxy skin and low respiration rate (5–10 mg CO_2/kg h^{-1} at 5 °C), make them very suitable for storage and long-distance shipping, thus apples are harvested once a year but available year-round.

Chemical and Nutritional Composition

A list of some of the nutrients in apples is presented in Table 3, with those for strawberries and oranges included for comparison. Apples contain the highest percentage of carbohydrates and food energy of these three fruits. A serving of one 150-g apple provides 8% of the US Department of Agriculture adult daily requirement for ascorbic acid (vitamin C) and 2% of both the vitamin A and iron requirements. Refer to individual nutrients.

Apples contain virtually no fat, cholesterol, or sodium. Apples are rich in pectins (soluble fiber) which has been found to be effective in lowering cholesterol levels. Apples also contain phenolic antioxidants and flavonoids (especially quercetin) which may also provide health benefits. (*See* **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; Synthetic Antioxidants, Characterization and Analysis;

Table 3 Nutrient composition (per 100 g fresh weight) of apples compared to that of strawberries and oranges

Nutrients and units	*Apple*	*Strawberry*	*Orange*
Water (g)	83.93	91.57	86.75
Food energy (kcal)	59	30	47
Food energy (kJ)	247	126	197
Protein (g)	0.19	0.61	0.94
Total lipid (g)	0.36	0.37	0.12
Cholesterol (mg)	0	0	0
Carbohydrate (g)	15.25	7.02	11.75
Fiber (total dietary) (g)	2.7	2.3	2.4
Calcium (mg)	7	14	40
Magnesium (mg)	5	10	10
Phosphorus (mg)	7	19	14
Potassium (mg)	115	166	181
Ascorbic acid (mg)	5.7	56.7	53.2
Folate (μg)	3	18	30
Vitamin A (μg retinol)	5	3	21

Data from United States Department of Agriculture Nutrient Database (2001) http://www.nal.usda.gov/fnic/foodcomp.

Role of Antioxidant Nutrients in Defense Systems; **Pectin**: Properties and Determination; Food Use.)

Handling and Storage

Apples in the northern hemisphere are harvested in the months of August, September, and October. Early varieties usually have a shorter shelf-life than late varieties. Apples are harvested by hand using a lifting and twisting motion to separate the fruit stem from the branch. An apple picker needs to handle the fruit gently to reduce the incidence of bruising on the fruit. Pickers usually carry a harvesting bag to hold the apples while they are walking through the orchard. The bag must be suitable for local conditions and ease of use and not damage the fruit. Often, the bag can be opened at the bottom in order to release apples easily into an orchard bin.

An orchard bin can usually hold approximately 700 l of apples. The size of wood bins is approximately $120 \times 102 \times 74$ cm and that of plastic bins is approximately $122 \times 113 \times 88$ cm. The advantage of using plastic orchard bins is the cleanliness, the faster cooling time, and the ease of fixing damaged bins, which can be simply melted down and reformed. The initial cost of a plastic orchard bin may be higher than the wood orchard bin, but the longer usage life of a plastic bin makes it a more attractive investment for a long-term commercial operation. These orchard bins can be stacked up to eight bins high for the plastic bins and nine bins high for the wood bins when placed inside a cold storage room. Bins need to be stacked closely together in order to force the cold air to pass through the apples.

Most apples can be stored at 0 °C. Some varieties of apples need to be stored at slightly higher temperatures in order to reduce the occurrence of disorders associated with chilling injury. For example, McIntosh apples have to be stored at 2.5–3 °C to avoid brown core disorder. Normally apples can be stored in cold storage for 3–6 months. The length of the storage life of apples depends on apple variety, storage temperature, harvest date, and storage atmosphere. The storage life of apples can be doubled if the apples are stored in CA storage rooms. Oxygen levels in a CA room are maintained at 1–3% and carbon dioxide levels are maintained at 1–5%. A CA room is basically an airtight cold storage room. The pressure in the CA storage room needs to be slightly higher than the atmosphere outside to prevent outside air from seeping into the room. The levels of oxygen can be reduced by burning methane or propane gas or by adding nitrogen gas to the CA room. Combusting oxygen with methane or propane can remove unwanted oxygen and produce the desired carbon dioxide. Nitrogen gas can be provided from a bulk tank of nitrogen gas or by a hollow-fiber membrane separation system. The principle of hollow-fiber membrane air separation is based on varying permeation rates for different gases, such as oxygen, carbon dioxide, and nitrogen, through the membrane.

The hollow-fiber membrane-type gas separator consists of bundles of small hollow polymeric fiber tubes in a vessel. Compressed air is fed into the separator, pressurizing the inside of the individual tubes. As the air flows through the tubes, oxygen and carbon dioxide permeate through the fiber walls faster than the nitrogen. This oxygen-rich stream is vented to the outside. The nonpermeate gas (nitrogen) outflow stream is piped into the storage room. By changing the output flow rate, the purity of the output can be controlled. Nitrogen purity levels of up to 99.9% can be obtained, but for CA applications purity levels of 97–99% are usually used.

The level of carbon dioxide in a CA room can also be achieved by the natural respiration of the fruit. When the level of carbon dioxide reaches 1–5%, excessive amounts of carbon dioxide have to be removed by absorbing carbon dioxide with lime or by adding nitrogen gas to the CA room.

Table 4 shows a summary of recommended storage temperature, levels of oxygen and carbon dioxide, and storage life of apples in CA storage for major apple cultivars in the world. Special care must be exercised when entering a CA storage room because the level of oxygen in a CA room is not enough to support human life. (*See* **Chilled Storage**: Principles; Attainment of Chilled Conditions; Quality and Economic Considerations; Microbiological Considerations; Use of Modified-atmosphere Packaging; **Chill Foods**: Effect of Modified-atmosphere Packaging on Food Quality; **Chilled Storage**: Packaging Under Vacuum; **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; Effects on Fruit and Vegetables; **Ripening of Fruit**.)

Industrial Uses

In most countries, a high percentage of the apple crop is used for processing. In the USA, statistics indicate that, in 1997, 50% of fruit was eaten fresh, 20% was processed into vinegar, cider (fresh or fermented), wine, juice, jelly or apple butter, 17% was canned as apple sauce, baby food, or pie filling, and the remainder was exported to other countries (predominantly for the fresh market). Cider (fermented), wine and brandy production are particularly important uses in Europe. (*See* **Cider (Cyder; Hard Cider)**: The Product

Table 4 Summary of recommended storage temperature, levels of oxygen and carbon dioxide, and storage life of apples in controlled atmosphere storage for major apple cultivars in the world

Cultivar	*Temperature (°C)*	*Oxygen (%)*	*Carbon dioxide (%)*	*Storage life (months)*
Braeburn	0.7	1.8	1.0	6–9
Fuji	0.3	1.4	1.0	7–11
Gala	1.3	1.7	1.6	2–9
Golden	0.5	1.6	2.3	7–11
Granny Smith	0.6	1.4	2.0	7–11
Idared	1.9	2.1	2.5	7–10
Jonagold	0.9	1.4	2.7	5–10
McIntosh	2.5	2.1	2.9	5–10
Red delicious	0	1.6	1.8	6–11
Royal gala	−0.2	1.5	1.8	5–8
Average	0.9	1.7	2.0	

Source: Kupferman E (1997). *Controlled Atmosphere Storage of Apples. CA '97 Proceedings*, vol. 2, pp. 1–30. Postharvest Horticulture series no. 16. University of California, Davis. http://postharvest.ucdavis.edu/produce/storage/CAR apples.pdf, with permission.

and its Manufacture; Chemistry and Microbiology of Cidermaking.) Other minor uses include dried apples, yogurts, cakes, muffins, candy apples, salads, and fried or baked apples.

Apple juice consumption varies widely with region. In North America, citrus juices are more popular and apple juice accounts for only about 8% of total juice consumption. In contrast, in Germany, apple juice accounts for approximately 45% of total juice consumption. In addition to being a popular single-ingredient juice, apple juice is also very popular as an ingredient in blended juice mixes. The USA and Poland are the largest apple juice-producing countries. These countries each account for about 20% of selected country annual output. In the USA, Michigan, California, and New York account for most of the apple juice production with 40–45, 70, and 40%, respectively, of their crops being processed as juice. Most apple juice (95%) produced in Poland is exported, primarily to Germany and other neighboring countries. The profitability of apple juice production has led many Polish growers to switch to varieties which are more suitable for processing. In South America, Argentina and Chile are the primary producers of apple juice, most of which is exported to the USA. Traditionally, apples not suitable for the fresh market are utilized for juice. In recent years, China has become a major apple juice producer.

See also: **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; Synthetic Antioxidants, Characterization and Analysis; Role of Antioxidant Nutrients in Defense Systems; **Chilled Storage**: Principles; Attainment of Chilled Conditions; Quality and Economic Considerations; Microbiological Considerations; Use of Modified-atmosphere Packaging; **Cider (Cyder; Hard Cider)**: The Product and its Manufacture; Chemistry and Microbiology of Cidermaking; **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; Effects on Fruit and Vegetables; **Pectin**: Properties and Determination; Food Use

Further Reading

Committee on Dietary Allowances (1989) *Recommended Dietary Allowances*. Washington, DC: National Academy of Sciences.

De Coster J (1998) European trends in the development of apple varieties. In: *Proceedings of INTERPOMA. The Apple in the World: Growing, Storing and Marketing*. Obstbau Weinbau 35(5).

Food and Agriculture Organization (1998) Production statistics. FAOSTAT. http://apps.fao.org.

Janick J and Moore JN (1996) *Fruit Breeding*. New York: Wiley.

Kupferman E (1997) *Controlled Atmosphere Storage of Apples. CA '97 Proceedings*, vol. 2, pp. 1–30. Postharvest Horticulture Series no. 16. University of California, Davis. http://postharvest.ucdavis.edu/Produce/Storage/CARapples.pdf

Naeve L (1997) Apple varieties and their uses. *Horticulture and Home Pest News* IC-477(23): 138–139.

Nicetto N (1998) Trends in European apple production. In: *Proceedings of INTERPOMA. The Apple in the World: Growing, Storing and Marketing*. Obstbau Weinbau 35(5).

Teskey B and Shoemaker JS (1982) *Tree Fruit Production*. Westport, Connecticut: AVI.

USDA (1982) *Composition of Foods: Fruits and Fruit Juices, Raw Processed, Prepared*. US Department of Agriculture, Human Nutrition Information Services, Agriculture Handbook, pp. 8–9. Washington, DC: US Government Printing Office.

Werth K (1998) Apple growing in Asia. In: *Proceedings of INTERPOMA. The Apple in the World: Growing, Storing and Marketing*. Obstbau Weinbau 35(5).

Westwood MN (1993) *Temperate Zone Pomology*, 3rd edn. Portland, Oregon: Timber Press.

APRICOTS

S M Southwick, University of California, Davis, CA, USA

Introduction

Apricots are one of the most ancient crops produced by humans and the fruit is grown on every continent of the globe. Most apricot production in the world is centered in Europe and Asia with lesser production in the Americas, Africa, and Oceania. In the former USSR, after years of collection and evaluation, there are as many as 1800 cultivars and forms of *Prunus armeniaca* and in China there are more than 2000 cultivars, many of which are purported to be widely adapted, large-fruited, and of excellent quality. The apricot is generally adapted to narrow climatic areas. The chilling requirements required for rest limit adaptability. As a result of narrow adaptation, favorable local varieties are not always produced economically in other regions. Apricots are soft fruits with a hard stone (stone fruits) and are generally sensitive to disease and to high or low temperatures. Apricots can have outstanding organoleptic qualities and the fruit is high in fiber and vitamins, making it ideal in the human diet. Consequently, this versatile fruit is consumed in fresh, canned, dried, and other processed forms. The varied germplasm which exists in the world is being utilized to improve existing apricot cultivars, helping to insure a continued supply of high-quality, healthy apricots for the world population.

Global Distribution and Relative Commercial Importance

Apricot production in the former USSR is located primarily in Ukraine, Russian Federation, Armenia, and in the Central Asian Republics (Uzbekistan, Tajikistan, Turkmenistan), and the Northern Caucasus (Krasnodar region, Daghestan and Karbardino-Balkarsk ASSR). Smaller growing regions exist in Azerbaijan, Georgia, Kirghizia, and Kazakhstan.

A relatively small production of apricots is located in Switzerland and Poland, with increased amounts in Bulgaria, Romania, and Hungary. To a larger extent, apricots are grown in the Mediterranean region, Spain, Portugal, France, Italy, Yugoslavia, Slovakia, Croatia, Macedonia, Albania, Greece, and Israel. In North Africa, Algeria, Tunisia, and Morocco are the leading producers. The central and western Mediterranean countries produce more than one-third of the world's production of apricots. Apricot cultivars of the Near East include cultivars found in Iran, Iraq, Syria, Afghanistan, Pakistan, and Turkey. Existing local varieties and ecotypes were developed by growers and horticulturists, and many small orchards today consist of wild seedlings mixed with cultivated forms. Turkey produces many of the world's apricots and there are an estimated 10×10^6 trees, of which about 50% are budded. Most of the fruit is consumed fresh; the remainder is dried and some of that is exported. In Iran there are an estimated 8996 ha (22 230 acres) of apricots. Most of the apricots are consumed fresh or dried both in Iran and in the previously mentioned countries. The apricots grown in Iraq and Pakistan are mostly generated by seed and consequently there is a wide variation in productivity from year to year.

In Australia and New Zealand, apricots are planted to a limited extent. In 1990, it was reported that approximately 27 300 and 7800 t of apricots were produced in Australia and New Zealand, respectively. Approximately 90% of Australia's apricots are produced in the River Murray irrigation areas in South Australia and Victoria and along the tributaries of the Murray in New South Wales and Victoria. Most of the apricots are dried. In New Zealand the main area of production (85%) is located in the South Island in the Central Otago region. Hawke's Bay is another apricot-producing area in New Zealand. In contrast to Australia, 60% of New Zealand's apricot production is sold fresh and 35% processed. South Africa produced approximately 46 000 t of apricots in 1990 (Table 1). The South African production has increased and by 1993–94 there were reported to be about 5400 hectares.

China appears to possess the most extensive repository of apricot and related germplasm in the world. There are eight species in the *Armeniaca* genus located in China, four of which include the common apricot (*Armeniaca vulgaris* Lam. or *Prunus armeniaca* L.), the Sibirian apricot (*P. sibirica* L.), the Manchurian apricot (*P. mandshurica* Koehne.), and the Tibet apricot (*P. armeniaca* var. *holoserice* Batal.). *P. mume* (mume), a related species, also originated in China. The diversity of apricot germplasm in China, if utilized effectively in genetic improvement research, could substantially broaden the existing commerce of apricots in North America and the world. Japan is a minor producer of apricot and produces more *Prunus mume*. Historically, the apricot and mume were utilized for medicinal purposes, but now the fruit is

Table 1 Apricot production and utilization 1989–98 in major producing/trading countries

Country and year	*Total production (t)*	*Commercial production (t)*	*Domestic fresh consumption (t)*	*Processed (t)*
France				
1989	130 000	114 000	101 200	15 000
1990	107 500	94 300	77 400	15 000
1998	160 000			
Greece				
1989	83 875	83 875	15 603	39 621
1990	113 360	113 360	10 273	41 880
1998	38 000			
Italy				
1989	189 000	181 000	166 725	30 000
1990	181 500	175 000	161 490	30 000
Spain				
1989	155 600	149 900	90 000	45 000
1990	115 200	110 600	63 800	40 000
1998	152 000			
Turkey				
1989	445 000	445 000	152 785	289 250
1990	400 000	400 000	137 000	260 000
1998	538 000			
Yugoslavia				
1989	46 000	37 000	20 879	23 000
1990	40 000	32 000	18 000	20 000
1998	8000			
USA				
1989	120 000	113 000	18 000	95 000
1990	112 000	105 000	18 000	87 000
1998	91 000			
Argentina				
1998	16 550	16 550	7499	9000
1990	15 500	15 500	6950	8500
1998	19 000			
Australia				
1989	27 000	27 000	3300	23 700
1990	27 300	27 300	3500	23 800
1998	27 000			
Chile				
1989	14 000	13 500	6200	6300
1990	14 650	14 150	6500	6550
1998	33 000			
New Zealand				
1989	9000	6800	6500	2282
1990	7800	6000	4700	2500
1998	8000			
South Africa				
1989	43 040	43 040	4652	37 709
1990	46 383	46 383	4650	40 650
1998	65 000			

consumed in processed forms such as syrups, jams, pickles, and liquor. In 1986, mume production approximated 89 100 t and apricot about 1460 t.

Argentina and Chile are the two leading producers of apricots in the southern hemisphere, with production of 15 500 and 15 650 t, respectively, in 1990.

In North America, the USA is the leading producer of apricots, but they are produced to a limited extent in Canada and Mexico. Canadian apricot production is located primarily in the western part with some small production near the Great Lakes. In Mexico, apricots are grown in the high desert east of the Sierra Madre mountain range. In the USA, California is the leading apricot-producing state, followed by Washington and Utah with a smattering of production in Michigan, Colorado, Texas, New Jersey, Idaho, New York, and the Carolinas.

In California, there are approximately 19 000–23 000 acres (7689–9308 ha) planted. Nonbearing acreage accounts for approximately 500–1000 acres (202–405 ha) of that total. Stanislaus county is the leading apricot-producing region in California, with

over 7000 acres (2833 ha). Other important apricot-producing counties in California include Contra Costa, Fresno, Kern, Kings, Madera, Merced, San Benito, San Joaquin, Santa Clara, Tulare, and Yolo-Solano.

Varieties

The most popular apricot varieties in California are described below, and some of their bearing acreages in 1990 are listed in Table 2. The Patterson apricot is being planted more widely and is a multiuse apricot, utilized primarily for canning (processing), but also frozen, dried, and eaten fresh.

Castlebrite

Released as a very-early-ripening, fresh-shipping variety, the fruit is yellow-orange and attractive, with a red blush. It is firm, medium in size, but does not possess finest eating quality. The cultivar suffers from inconsistent production, probably because the chilling requirements are not satisfied.

Derby

Derby or Derby Royal is principally grown in Yolo County, California. It originated from a chance seedling and was first planted in 1895. Derby is suitable for fresh shipping, but not for drying because the stone can cling to the flesh. The fruit may ripen unevenly. At least two distinct strains of the Derby have been recognized around Winters, California area with differing fruit characteristics.

Earlicot and Poppy Cot

Earlicot and Poppy Cot are varieties introduced by Zaiger Genetics of Modesto, California. They are among the earliest ripening apricots grown in California. They are large, firm, and yellow-orange in color with a red blush occasionally found on the skin. They are slightly acidic with a fair flavor overall and are suitable for commercial shipment.

Table 2 California acreage of selected apricot varieties in 1990

Variety	*Acres (ha)*	*Harvest time*
Blenheim/Royal	7303 (2955)	17 June
Castlebrite	986 (399)	7 May
Derby Royal	221 (89)	25 May
Flaming Gold	157 (64)	21 May
Improved Flaming Gold	118 (48)	21 May
Katy	319 (129)	15 May
Modesto	578 (234)	1 June
Patterson	3404 (1378)	12 June
Tilton	5922 (2397)	18 June
Westley	174 (70)	25 June
Other	584 (236)	

Flaming Gold

Originated in Modesto, California, by Floyd Zaiger, Flaming Gold was introduced in 1967 and was a seedling of Perfection. The fruit ripens early and has been used primarily for fresh shipping.

Goldbar and Goldstrike

Goldbar and Goldstrike originated from the breeding program of Dr Tom Toyama at Washington State University in Prosser, WA. Both fresh-market apricots are large-sized, with early maturity (after Castlebrite and before Patterson). They may require cross-pollination and they appear to be cross-compatible. They have smooth, orange skin often with a red blush and orange flesh. The eating quality is good.

Helena

Formerly K210-35, Helena originated from the breeding program of the United States Department of Agriculture (USDA) in Fresno, CA. The fresh-market apricot ripens near the time of Patterson. It is relatively large, firm, and has yellow-orange skin and flesh. It is of fair eating quality.

Katy

Large- to-medium-sized fruit developed by Floyd Zaiger, the Katy is yellow, firm, and an early-ripening fruit used for fresh shipping.

Lorna

Formerly K505-50, Lorna originated from the breeding program of the USDA in Fresno, CA. The fresh-market apricot ripens early in the season ahead of Patterson and is large, with orange skin and flesh. It may often have skin with a blush of red. The eating quality is good.

Modesto

Originated in Le Grand, California, by FW Anderson, the variety was introduced in 1964 and was an F_2, open-pollinated seedling of Perfection. The fruit is firm-fleshed, medium- to-large-sized, often with a red blush, and orange-fleshed. The tree bears regularly, and the fruit ripens about the time of Royal and several days ahead of Patterson. Modesto is resistant to pit burning.

Patterson

Originated in Le Grand, California, by FW Anderson, the Patterson was introduced in 1968 and was an F_2 seedling of Perfection $\times$ unknown. The variety is

firm; it has orange flesh and skin, and is sometimes found with a red blush. The fruit is medium to large, and produces consistently good crops. It appears to be less affected by mild winters than the Tilton variety, and the fruit ripens evenly. Patterson is resistant to pit burning. It is a good variety for processing (canning, freezing), makes a good dried product, and is fair for fresh-eating fruit. The Patterson is the apricot production leader in California.

Robada

Formerly, K106-2, Robada originated from the breeding program of the USDA in Fresno, CA. The fresh-market apricot is early-maturing, ripening ahead of Patterson, and has a very nice orange skin, often with a red blush. This fairly large fruit has orange flesh, which is firm and has good eating quality.

Royal and Blenheim

The Royal and Blenheim varieties are considered together because they have lost their separate identities. The variety was utilized for fresh market and canning, but currently is used for drying and juice. It tends to bear alternately and unless properly thinned may have small fruit. It has excellent flavor when fully matured. It is rather soft and difficult to ship. The flesh is orange, and the skin is yellow-orange; it often has a red blush. The fruit tends to pit burn and is not well adapted for the interior valleys of California. It can develop a condition known as fog spot in the coastal regions of California.

Tilton

The variety originated as a seedling in Kings County, California by JE Tilton in 1885. Tilton was once a leading apricot for processing in California but it has currently been replaced by Patterson. Tilton was used for canning, drying, and shipping, but the fresh flavor is only fair. Tilton is now mostly used for concentrate. The tree tends to bear alternately and is affected by mild winter temperatures. When properly thinned it may bear medium- to large-sized fruit. The fruit is orange to yellowish, ripening late in the season.

Tomcot

The Tomcot originated from the breeding program of Dr Tom Toyama at Washington State University in Prosser, WA. The fresh-market apricot is early-maturing (ahead of Goldbar and Goldstrike, but after Castlebrite). It is not large but adequate and firm. It has very highly colored orange skin and flesh, and the skin often has a red blush. It normally requires hand thinning to insure adequate fruit size and it has good eating quality.

Tri Gems

Developed by Zaiger Genetics, Tri Gems is a firm, medium-sized apricot that yields high-quality canned fruit. It ripens before Patterson and has a yellow-orange skin and flesh. If fruit are removed by hand thinning it can be sold fresh and dried.

Westley

Originated in Le Grand, California, by FW Anderson, the variety was introduced in 1973 and was an F_2, of Perfection × Tilton. The fruit is medium-firm, with good flavor, and is used for some fresh markets and drying. The tree tends to bear alternately.

Rootstocks

Apricots are grown commercially on apricot seedling, peach (*P. persica*), and plum (*P. cerasifera*, myrobalan seedling and 29C; *P. cerasifera* × *P. munsoniana, marianna* 2624) roots. The main plum rootstocks include myrobalan seedling, myrobalan 29C and marianna 2624. Plum rootstocks are characteristically resistant to wet and poorly drained soil conditions. The union between apricot and plum root reportedly is not as sound as apricot on peach or apricot root and occasionally some apricot varieties break off at the union during heavy wind storms. Apricots growing on plum roots may not be as productive as those grown on apricot or peach roots. Peach root for apricot is quite popular. The stocks include those propagated from Lovell, Nemaguard, and Nemared seedling. Apricots grown on peach are productive, but peach cannot tolerate wet and poorly drained soil conditions as well as plum. Peach roots are very susceptible to *Phytophthora* root and crown rots. Apricots grown on apricot seedling roots are not as common as those on peach roots. Apricot roots are not as resistant to wet soil or *Phytophthora* as plum and also do not have the resistance to nematodes as does Nemaguard peach root. However, apricot roots are thought occasionally to impart better fruit quality and productivity to apricots. (*See* **Peaches and Nectarines**.)

A rootstock, developed by Floyd Zaiger and named Citation, is a peach × plum hybrid. It has been shown to increase fruit size and advance fruit maturity of Royal/Blenheim but did not alter the production characteristics of Patterson apricot when compared to marianna 2624. The rootstock may impart some tree size control to selected apricot varieties. However, apricots growing on Citation have shown reduced levels of zinc and nitrogen which suggested reduced uptake of those nutrients by the rootstock and possibly reduced vigor.

Morphology and Anatomy

Botany and Biology

The apricot belongs to the *Rosaceae* family and most cultivated apricots belong to the species *Prunus armeniaca* L. Closely related species include *P. mume* Sidb. and Zucc., the Japanese apricot, *P. dasycarpa* Ehrh., the black apricot, *P. brigantiaca* Vill., the Briancon apricot from the French Alps, *P. ansu* Komar, *P. sibirica* L., and *P. mandshurica* (Maxim.). The apricot is diploid ($2n = 16$, $\times = 8$). Flowers are borne singly or doubly at a node on very short stems (peduncles). Flowers have about 30 stamens with one pistil. The solitary flowers are white or pinkish. Most commercial cultivars in the USA are self-fertile, but Perfection and Riland are examples of self-incompatible cultivars. Trees generally branch well and produce vigorous upright growth, but not as upright as plum. Floral initiation occurs in summer (late May or June) and most of the flowers which set into fruit are produced on spurs. Spurs are productive for several (3–5) years. The highest-quality fruit is borne on younger spurs. Apricots are the first deciduous fruit trees to produce flowers in the spring after almond, and because of that are subject to frost damage. Trees bloom over a period of approximately 1–2 weeks, depending upon weather conditions. Flowers are followed by the appearance of leaves, which are simple, alternate, and serrated, round-ovate to ovate, and sharp-pointed. Apricot cultivars currently available require approximately 450–1200 h of chilling (temperatures at or below 7.2 °C). Vegetative buds require less chilling than reproductive (flower) buds. In those years with insufficient winter (December and January) chilling, flower bud drop can occur leading to reduced cropping, flowering and leaf emergence may occur over an extended period, fruits may develop with few leaves, and small fruits may not persist until harvest. Apricots produce more flowers than are needed to insure the production of an adequate crop and often require the removal of small fruit (thinning) to insure adequate fruit size. The apricot is a stone fruit (drupe). Other stone fruits in the genus *Prunus* include almond, cherry, nectarine, peach, and plum. Other drupe fruit include olive, coconut, mango, etc. Botanically a drupe fruit is a fleshy, one-seeded fruit that does not split open of itself, with the seed enclosed in a stony endocarp, which is called a pit. The fruit of apricot consists of a stony endocarp, a fleshy mesocarp, and an outer exocarp (skin). (*See* individual fruits.)

Approximately 80% of all California apricots are processed (Table 3). California represents nearly 100% of commercially processed fruit in the USA.

Table 3 Volume shares of apricots produced in California, USA in 1998

Product type	*Metric tonnage (t)*	*Percentage of total*
Canned[a]	36 340	40
Dried	8036	9
Fresh	16 518	18
Frozen	9286	10
Concentrate	21 161	23
Total	91 341	

[a]Source: Apricot Producers of California, Modesto, CA. Statistics vary from season to season.

Basic grower production is sold for processing to five different sources: canners, nectar (juice) processors, driers, freezers, and baby food processors. For the purpose of describing these processing methods, only canning, drying, and freezing will be included since nectar and baby food processing are basically similar to the process for producing apricot concentrate of purée by canners.

Processing Apricots in Europe

Canned apricots comprise the major processed product for apricots in Europe, followed by fruit purée and dried fruit last. Fruit for canning must be round, and uniform in shape, size, and maturity. Fruits are pitted by slicing along the suture line of the fruit with rotating knives; stones sticking to flesh are problematic. Often apricots are peeled (the skin is removed) before canning, and this appears to reduce fruit softening. Dipping in hot (boiling) lye (at least 4% concentration) is used for peeling and within 1 min most apricot cultivars are well-peeled. Fruits are packed in cans, glass jars, and plastic bags. Polyvinyl chloride (PVC) bags are covered by thin metallic sheets in order to reduce permeability to oxygen and sterilized with γ-irradiation. Various techniques have been developed to maintain the color, firmness, texture, and flavor of processed apricots.

Apricot fruit used for purée can be harvested with a higher degree of maturity to insure higher organoleptic quality. Stabilization of color and viscosity are essential to maintaining quality of product. Fruit purée is mainly used in baby food, fruit nectars, and jams.

Canned Apricots in California

The majority of apricots for canning are grown in Stanislaus and San Joaquin county. One major canner is located in this area and another operates in the Santa Clara Valley. Most canning apricots consist of

the Patterson and Tilton varieties, with the Patterson predominating.

Canners' prices to the growers are negotiated between the canners and the growers' bargaining association. Current prices are based upon a price per (UK) ton for 'US no. 1s' with a downward sliding scale of five grades below this level. Grading is carried out by a third-party inspection program approved by both parties.

Fruit is picked either by hand or, to a lesser extent, mechanically, and dumped into 272-kg bins. These bins are delivered to a weighing station where they are inspected for grade and weighed. The apricots are then delivered to the canning plant, where they are mechanically dumped into a water bath containing chlorinated water. They travel down an inspection belt for removal of any materials other than apricots (leaves, sticks, stones) or immature, green, overripe or rotted fruit. The apricots are then graded for size; the small fruits, unsuitable for canning, are collected and directed towards eventual use in concentrate and/or nectar-type product items.

The fruit is delivered to mechanical apricot cutters; these machines align the fruit in such a way that the cutting will take place on the apricot suture. Cut fruit is then opened and the loose pits drop through a stainless-steel plate. Cut, pitted fruit is delivered to the sizer/grader to be mechanically sorted, resulting in uniformly sized fruit. Once the cans have been filled, a topping medium such as sugar water or fruit juice is added, then immediately sealed. The product is then ready to be cooked, rendering it commercially sterile. Later, when the product reaches the warehouse it will be labeled and cased as per the customer's request. Canned apricots are packed in heavy syrup, light syrup, juice pack (apricot or pear), and water pack.

Back in the processing plant, fruit that is unsuitable for cutting is delivered to a thermal screw, where it is heated to 98.9 °C and delivered to a pulping unit, which will remove the pit. The heated pulp is pumped through a series of finishers that will remove some of the fibrous material, such as the skin of the apricot. Finished juice is then ready for the evaporation process or delivered to the nectar room where it is mixed with sugar, water, and citric acid. The nectar is filled and sealed, similar to the cut-fruit canning process, and delivered to the warehouse. Fruit juice destined for the evaporator will be condensed, resulting in 32 ° Brix (32%) fruit concentrate. The product is canned in 3.25-l cans or aseptically filled in 250-l drums. The concentrated product is ideally suited for nectar reconstitution or as an ingredient in various sauces.

There are a very few marketers of apricot nectar that purchase off-grade apricots directly from the grower and process their own concentrate or purée according to their own particular specifications. The concentrates are then diluted to an acceptable level as a drinkable apricot juice called nectar.

Dried Apricots

Dried apricots are produced from the plump, ripe, fresh fruit. The fruit is picked and sun-dried in June through early August in California. The Blenheim and Patterson apricots are the principal varieties used for drying. Other varieties are dried but most do not have the unique combination of high flavor density and solids-to-acid ratio that create the luscious, sweet-tart flavor of the California dried apricot.

Sun-Drying of Fruits

The drying of apricots parallels methods used for processing other dried fruits (figs, peaches, pears, etc.). Stage of maturity of the fruit is important in selecting fruits for drying. If the fruit is picked too early, color and flavor are lacking in the final product. When overripe, the final product loses shape and becomes slab-like in appearance. Much of the fine flavor of California dried apricots results from the fact that the dried fruit can be produced from fully tree-ripened fruit.

Preparation of the fresh fruit for sun-drying is simple. The common predrying treatments applied to apricots are as follows: (1) selection and sorting of fresh fruit; (2) washing; (3) cutting into halves and removal of pits; (4) spreading of fruit on drying trays with the cut surface upward; (5) sulfuring with burning sulfur or gaseous sulfur dioxide; (6) placing of trays in full sun in the dry yard. (*See* **Drying**: Drying Using Natural Radiation.)

Sulfuring For many years sulfur dioxide (SO_2) has been used to preserve the color of dried fruits. It is the only chemical added to dried apricots. Sulfur dioxide is recognized as safe for use in dried fruits by the US Food and Drug Administration (FDA). Apricots prepared for drying are usually exposed to gaseous SO_2 before being put in the sun for drying, but treatment with fumes of burning elemental sulfur is also practiced.

In addition to preventing enzymatic browning, SO_2 treatment reduces the destruction of carotene and ascorbic acid which are valuable natural nutrients in apricots; thus, nutritional qualities are preserved.

Sulfured, dried apricots will contain a SO_2 level of about 2500 p.p.m. The amount of SO_2 must also be controlled because regulations on levels of SO_2 permitted vary from one country to the next. Sulfur dioxide in the fruit begins to dissipate as soon as

applied and continues to diminish throughout storage, distribution, and retail shelf-life of the product.

Sun-drying After sulfuring, the trays are placed in the drying yard for sun-drying. Sun-drying is complete when the apricots have a moisture content of 15–20%. Drying time can vary depending on the condition of the fruit, air moisture, and constancy of sun exposure. During drying the cut and sulfured fruit is left in full sun for 5–10 days, followed by further drying away from direct sunlight for a sufficient time to bring moisture to the desired level.

After drying, the apricots are transferred to boxes for curing and for bringing about equilibrium of moisture content. This requires from 2 to 3 weeks or even longer. The dried apricots are then ready for grading and final preparation for packing.

Screening, Inspection, and Washing

Size grading is required to obtain the desired piece size in the end product. Size grading is accomplished by passing the dried product over a shaking perforated metal screen and collecting the fractions separately. Each size fraction passes on to the final inspection operation.

Prior to packing, chopping, or grinding, the apricots are rigorously washed to remove dust, leaf particles, etc. Washing consists of a presoak in water, followed by mechanical scrubbing and rinsing. After washing, the fruit is spread on trays for a second treatment with SO_2 in the finished product to meet shelf-life requirements, or requirements specified by the customer of bulk packs.

Storage

Dried apricots are protected from product decomposition caused by microbial or enzymatic deterioration. The relatively low moisture level, high natural sugar level, high acid content, SO_2 level, and low pH preclude these types of spoilage. (*See* **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability.)

The loss of SO_2 cannot be eliminated entirely, but it can be controlled so that it is of little consequence in normal commercial storage. The single most important factor in determining the storage life in dried fruit is temperature of the storage space. Temperature is so critical that the storage life is cut approximately in half: for every 11 °C increase it is about 4.4 °C.

The following conditions and tips will help to obtain maximum storage life for dried fruits:

1. Store dried fruit at 4.4 °C and 75% relative humidity for excellent keeping for at least 6–9 months (for washed and resulfured apricots).
2. Keep temperatures and humidity constant.
3. Be sure that the product is well wrapped and not exposed to air.
4. Protect dried fruit from strong direct light.

Frozen Apricots

Apricots for freezing include the Patterson and Blenheim varieties, with Patterson predominating.

Frozen apricots are processed for three different product types: sliced halves, slices, and a multiple scored apricot that is not completely sliced. The former two types are sold primarily to bakers, ice-cream makers, and frozen-dessert makers. The latter type are processed for use in jams and jellies and are called 'machine-pitted.'

Receiving

Apricots are received in approximately 363-kg bins. There are normally 48 bins per truckload. Each truckload is weighed and assigned a lot number. Bins from each lot are visually sampled for quality. The apricots are checked for ripeness, insect penetration, rot, flesh damage, and excess foreign materials such as leaves. Fruits are placed in high-temperature cold storage or left in receiving for ripening. As needed, bins of fruit are brought to the processing plant by truck.

Leaf Roller

The apricots are dumped on to a conveyor and pass over a series of parallel rollers. The rollers are separated by small gaps through which leaves and twigs may fall. A worker is stationed there to remove leaves, foreign material, rot, and green fruit. The green fruit are held for rerunning when properly ripened.

From the leaf roller the apricots fall into the shaker-washer which agitates the apricots in water and propels them under fresh-water sprays.

The shaker-washer discharges the apricots on to the first inspection belt. The fruit is examined for rot, leaves, and green color. From the first inspection belt, the fruit drops on to a cleated elevator which carries the fruit to a cross-belt. The cross-belt delivers the apricots to one of two apricot halvers. A shaker-feeder receives the apricots. Rotating scrubbers clean the apricots and single-file them into five lanes. The apricots from the five lanes are fed into pick-up pockets. The pick-up pockets feed the apricots to V trough belts in time with orienting fingers. The orienting fingers cause the apricots to rotate so that the suture is aligned with blades which cut the apricot in half in line with the suture.

A pit remover at the end of the blade is used to separate the pit from the two halves as they pass

through. Approximately 85% of the pits are removed at this point.

A cross-belt carries the apricot halves to a conveyor, which transports them to a cup-up pit shaker. As the apricot halves move across the shaker pan they pass over holes which permit the remaining pits to fall through as they are shaken loose. As the apricot halves are discharged from the cup-up pit shaker they are aligned with the pit cavity, facing up.

The cup-up pit shaker distributes the apricot halves on to the pit inspection belt. On the belt the apricots are inspected for pits and soft fruits. Soft fruits are placed into sort-out buckets and used to make purée. From the pit inspection belt the apricots are transported to a gang blade slicer. The slicer has parallel circular blocks separated by a 1.27-cm gap. The slicer cuts the halves into strips.

The sliced apricots fall on to a final inspection belt. The apricots are inspected for pits, blemishes, and harmless extraneous material. Fruit from the final inspection belt is carried to a fill-weight station. The fruit is placed in prelabeled tins and weighed. A depressor is placed on top of the fruit and 60° Brix syrup with ascorbic acid is added. The container is weighed again. A lid is placed on the can and it is passed through a can washer. The cans are coded, loaded on pallets, and tagged. The completed pallets are promptly moved to cold storage for freezing.

Apricot Usage

At the height of California apricot production in the 1930s and 1940s, the major usage was dried apricots. This was primarily because of the dried product's ability to remain shelf-stable for a relatively long period of time, its versatility as an out-of-hand snack, cooking and baking ingredient, and its sweet-sour flavor.

After World War II, the canning industry began its dynamic growth. Dried apricots gave way to the canned version because of a growing acceptance of canned fruits and vegetables as convenient and economical products, as well as the apricots' greater similarity to the fresh fruit. The dried apricot continued to be the largest usage type through the late 1940s but lost its position to canned apricot products (which included baby food, purée, and nectar) in 1950 and never regained its plurality. Recently, sales of dried apricots have been declining due to foreign competition.

Frozen apricots were introduced to the American public in 1943, diverting about 4046 t out of a total processed crop of 79 248 t, the smallest crop recorded up to that time. Apricots going into frozen production peaked in 1945 at 26 822 t out of a total of 161 544 t and have never regained that level. Today frozen apricots account for less than 10% of the pack.

Apricots sold fresh in retail and food-service outlets have been slow in developing into major usage. Their high point in sales was reached in the disastrous 1943 crop year, with 21% of total production. Most years since 1909 have held at around 10% of the total crop. Since 1984, however, with substantially more marketing effort behind them, fresh apricots have been maintaining a 14–20% share of saleable tonnage.

In 1998, the California Apricot Producers reported 91 341 t by use-type, distributed as shown in Table 3.

Nutrition

Weight for weight, apricots provide more than three times the carotene of peaches, and three times the

Table 4 Nutritional composition (per 100 g fruit) of apricots[a]

		Canned				
	Fresh	*Heavy syrup*	*Light syrup*	*Juice pack*	*Nectar*	*Dried*
Energy (J)	200	350	265	200	235	1000
Protein (g)	1	1	1	1	0	4
Carbohydrate (g)	11	21	16	12	14	62
Fat (g)	0	0	0	0	0	0
Sodium (mg)	1	4	4	4	4	10
Potassium (mg)	295	140	138	165	114	1378
Fiber (g)	2	0	0	1	1	8
Calcium, Ca (mg)	14					
Iron, Fe (mg)	0.54					
Magnesium, Mg (mg)	8.0					
Vitamin C (mg)	10.0					
Vitamin A (IU)	2612					

[a]From the California Apricot Advisory Board (defunct) and Apricot Producers of California, Modesto, CA and The Fruit Pages, http://www.thefruitpages.com/chartapricots.shtml.

vitamin C of pears (based on a 1 cup serving of canned fruit, in juice pack with skin). They are also higher in potassium, calcium, phosphorus, and iron than most other fruits (Table 4). Apricots, especially dried apricots, are an excellent source of dietary fiber. They are low in sodium, calories, and fat, and have no cholesterol. (*See* individual nutrients.)

See also: **Canning**: Principles; Food Handling; Quality Changes During Canning; **Carotenoids**: Occurrence, Properties, and Determination; **Controlled-atmosphere Storage**: Effects on Fruit and Vegetables; **Dietary Fiber**: Properties and Sources; **Drying**: Drying Using Natural Radiation; **Freezing**: Nutritional Value of Frozen Foods; **Fruits of Temperate Climates**: Commercial and Dietary Importance; Factors Affecting Quality; **Iron**: Properties and Determination; **Jams and Preserves**: Methods of Manufacture; **Phosphorus**: Properties and Determination; **Potassium**: Properties and Determination; **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability

Further Reading

Bailey CH and Fredric Hough L (1975) Apricots. In: Janick J and Moore JN (eds) *Advances in Fruit Breeding*, pp. 367–383. West Lafayette, Indiana: Purdue University Press.

CASS (1990) *California Fruit and Nut Acreage 1989*. Sacramento, California: California Agricultural Statistics Service.

Chandler WH, Kimball MH, Philp GL, Tufts WP and Weldon GP (1937) Chilling requirements for opening of buds on deciduous orchard trees and some other plants in California. *University of California Agricultural Experiment Station Bulletin* 611, pp. 35–38.

Chittenden FJ (ed.) (1956) Apricot. *The Royal Horticultural Society Dictionary of Gardening*, vol. 1, pp. 15–156. Oxford: Clarendon Press.

Coe FM (1934) Apricot varieties. *Utah Agricultural Experiment Station Bulletin* 251.

Crosse-Raynaud P and Audergon JM (1987) Apricot rootstocks. In: Rom RC and Carlson RF (eds) *Rootstocks for Fruit Crops*, pp. 295–320. New York: John Wiley.

Day LH (1953) Rootstocks for stone fruits. *California Agricultural Experiment Station Extension Service Bulletin* 736.

FAO Production (1998) *FAO Statistics Series*, vol. 52, no. 148. Rome: Food and Agriculture Organization of the United Nations.

Glucina P, Hosking G and Mills R (1990) Evaluation of apricot cultivars for Hawke's Bay. *The Orchadist of New Zealand* 63: 21–25.

Hedrick UP (1922) *Cyclopedia of Hardy Fruits*. New York: Macmillan.

Hesse CO (1952) Apricot culture in California. *California Agricultural Experiment Station Extension Service Circular* 412.

Manolopoulou H and Mallidis C (1999) Storage and processing of apricots. *Acta Horticulturae* 488: 567–576.

Paunovic SA (ed.) (1986) Second international workshop on apricot culture and decline. *Acta Horticulturae* 209: 13–28.

Southwick SM and Weis KG (1998) Selecting and propagating rootstocks to produce apricots. *HortTechnology* 8: 164–170.

Southwick SM and Yeager JT (1999) Effect of rootstock, cultivar and orchard system on apricot production. *Acta Horticulturae* 488: 483–488.

USDA (1990) *Horticultural Products Review*, circular series, November 1990. Washington, DC: US Department of Agriculture Foreign Agricultural Service.

USDA National Agricultural Statistics Service (1999) *Agricultural Statistics 1999*. Washington, DC: US Government Printing Office.

Westwood MN (1978) *Temperate-Zone Pomology*. New York: WH Freeman.

Aquaculture *See* **Fish**: Introduction; Catching and Handling; Fish as Food; Demersal Species of Temperate Climates; Pelagic Species of Temperate Climates; Tuna and Tuna-like Fish of Tropical Climates; Demersal Species of Tropical Climates; Pelagic Species of Tropical Climates; Important Elasmobranch Species; Processing; Miscellaneous Fish Products; Spoilage of Seafood; Dietary Importance of Fish and Shellfish; **Fish Farming**; **Fish Meal**

Aroids *See* **Vegetables of Tropical Climates**: Commercial and Dietary Importance; Root Crops of Uplands; Root Crops of Lowlands; Edible Aroids

Aroma *See* **Sensory Evaluation**: Sensory Characteristics of Human Foods; Food Acceptability and Sensory Evaluation; Practical Considerations; Sensory Difference Testing; Sensory Rating and Scoring Methods; Descriptive Analysis; Appearance; Texture; Aroma; Taste

Aroma Compounds *See* **Flavor (Flavour) Compounds**: Structures and Characteristics; Production Methods

ARSENIC

Contents

Properties and Determination

N Hata, I Kasahara, S Taguchi and K Goto, Toyama University, Toyama, Japan

Background

Arsenic is located in group 5A of the periodic table and has an atomic number of 33 with an atomic weight of 74.922. The electron configuration is $(1s)^2(2s)^2(2p)^6(3s)^2(3p)^6(3d)^{10}(4s)^2(4p)^3$.

Chemical Properties

The structures of some of several important arsenic compounds and their acid dissociation constants (pK_a) are shown in Figure 1.

Inorganic arsenic (As) occurs in four oxidation states, As(V) as in arsenate (AsO_4^{3-}), As(III) as in arsenite (AsO_3^{3-}), As(0), and As(-III) as in arsine (AsH_3). In aerated water, arsenate is a stable form. Since As(III) is considerably more toxic than As(V), the form in which arsenic exists, whether as arsenate or arsenite, is important. Doses of 70–180 mg of arsenic trioxide (As_2O_3) are fatal. Arsenic acid is a fairly strong acid, while arsenous acid behaves in water practically as a weak, monobasic acid (see Figure 1).

Arsenic forms a variety of organoarsenic compounds. Arsenic compounds of the types, R_3As, $RAs(OH)_2$, R_2AsOH, $RAsO(OH)_2$, $R_2AsO(OH)$, and $R_4As^+X^-$ are called arsines, arsenous acids, arsinous acids, arsonic acids, arsinic acids, and arsonium salts, respectively. It is believed that organoarsenic compounds are much less toxic than the inorganic arsenic compounds. Biochemical conversion of inorganic arsenic compounds to organoarsenic compounds has been known for more than 100 years.

Occurrence

Arsenic is distributed everywhere in the environment. The use of arsenic is decreasing gradually but is still important. Commercial elemental arsenic is mainly employed as an additive for alloys. High-purity arsenic (at least 99.999%) is employed in electronics as gallium arsenic (GaAs) or indium arsenic (InAs). Furthermore, arsenic has been used in wood preservation, medical supplies (Salvarsan), pigment (Scheele's Green: $CuHAsO_4$, Paris's Green: $3Cu(AsO_2)_2{\cdot}Cu(CH_3COO)_2$), agricultural chemicals, food additives (arsanilic acid, $H_2NC_6H_4AsO(OH)_2$), and poison gas (lewisite). British anti-lewisite (BAL), which was developed as an antidote for lewisite, is also one of the antidotes for heavy metal poisoning.

Arsenic occurs naturally in the earth's crust (1.5–2 mg kg^{-1}). Hot-spring and well waters sometimes contain appreciable quantities of arsenic, mostly in

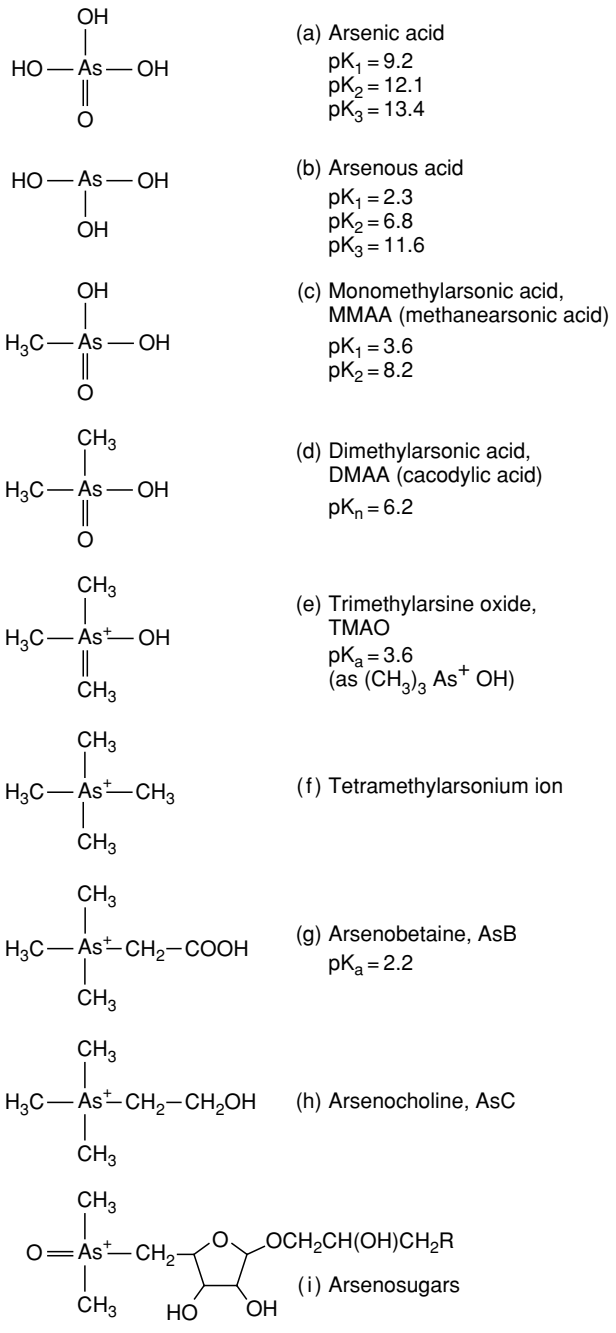

Figure 1 Structure of some important arsenic compounds and their acid dissociation constants. (i1) R = -SO_3H, 2-hydroxy-3-sulfopropyl-5-deoxy-5-(dimethylarsinoyl)-β-ribofranoside; (i2) R = -OH, 2,3-dihydroxypropyl-5-deoxy-5-(dimethylarsinoyl)-β-ribofranoside; (i3) R = -$OPO_3HCH_2CH(OH)CH_2OH$, 3-'glycerophosphoryl'-2-hydroxy-1-[5-deoxy-5-(dimethylarsinoyl)-β-ribofranosyloxy] propane; (i4) R = -OSO_3H, (2S)-3-[-5-deoxy-5-(dimethylarsinoyl)-β-D-ribofranosyloxy]-2-hydroxypropyl hydrogen sulfate.

the form of arsenate. Arsenic is associated with sulfide ores. Waste water discharged from sulfide mines and ore-dressing plants often contains arsenic in appreciable quantities. Arsenic contamination caused by flue gases from lead and copper smelters has been frequently reported. Most authorities adopt a value of $10\,\mu g\,l^{-1}$, measured as total arsenic, as the maximum permissible concentration in drinking water.

The arsenic levels in ocean water are around $2\,\mu g\,l^{-1}$, and although arsenate is the stable form in aerated waters, ocean water sometimes contains appreciable proportions of arsenite. It is believed that some bacteria are able to reduce arsenate to arsenite.

Because arsenic compounds have been used as insecticides, herbicides, and animal-feed additives, some soils, vegetation, swine, and poultry may be contaminated with arsenic. The value of $1\,mg\,kg^{-1}$ as As_2O_3 has been the maximum allowable limit for fruits and vegetables in Japan. Soil containing more than 15 mg of arsenic per kilogram is considered unsuitable for agriculture.

Edible marine organisms, such as lobsters, shrimps, and brown kelps, contain considerable concentrations of arsenic frequently in the $mg\,kg^{-1}$ range, occasionally at concentrations as high as several tens of $mg\,kg^{-1}$ on a wet-weight basis. Fortunately, however, the arsenic is mostly in the form of organoarsenic compounds, which are rapidly excreted and therefore relatively nontoxic to humans. On average, about 80% of the arsenic in marine organisms is found in the water-soluble fraction. Mostly, arsenobetaine (AsB) is the predominant arsenic species in marine animals, with arsenosugars in marine algae and some marine bivalves. Concentrations in freshwater plants and animals, and terrestrials are usually lower than in marine species. Arsenic has been found in mushrooms (*Agaricus bisporus*): As(III), 0.07 $mg\,kg^{-1}$; As(V), 0.14 $mg\,kg^{-1}$; monomethylarsonic acid, 0.05 $mg\,kg^{-1}$; dimethylarsinic acid, 0.28 mg kg^{-1}; and AsB, 0.46 $mg\,kg^{-1}$.

In 1900, there was an outbreak of arsenic poisoning among beer-drinkers in England. The cause of poisoning was traced to the use of invert sugar produced using arsenic-contaminated sulfuric acid. In 1955, there was an outbreak of unusual disease and fatalities in Okayama and Hiroshima Prefectures in Japan among infants drinking arsenic-contaminated milk. The arsenic content of the powdered milk that caused this disaster was estimated to be 15–24 $mg\,kg^{-1}$ as As. The dose of arsenic in the infants was estimated to be 1.3–3.6 $mg\,day^{-1}$ and the source of arsenic was traced to arsenic-contaminated sodium phosphate added as a stabilizer. The sodium phosphate had been produced from waste bauxite processing liquor and renamed by resale, but it contained about 3% arsenic.

Recently, a large number of fatal arsenic intoxications were traced from drinking water contaminated

by inorganic arsenic species. The main sources of arsenic contamination of drinking water are natural, and high concentrations of arsenic in drinking water are found in various parts of the world. In Taiwan, patients with blackfoot disease are found in areas where drinking water contains arsenic. In Bangladesh and West Bengal in India, arsenic poisoning became evident in the 1980s, and tens of millions of people are estimated to be at risk. The survey is not complete because the survey area is so large. The source of arsenic is geological.

Speciation of Arsenic in Food and Water

Various arsenic species show large differences in their toxicity. Their acute toxicity decreases in the following order: As(-III) (AsH_3) >> arsenite {As(III)} > arsenate {As(V)} >> dimethylarsinate > monomethylarsonate >> arsenobetaine and arsenocholine (organoarsenic compounds in seafoods). Therefore, speciation of arsenic is required for reliable toxicological estimation. Arsenic in seafoods has been well studied and is mostly in the form of organoarsenic compounds, including arsenobetaines and arsenosugars.

Extraction and purification are needed, especially for speciation of arsenic. For the speciation of organoarsenic compounds in seafoods, the samples are usually extracted with methanol or a methanol–water mixture. The samples are then purified with a C_{18} cartridge, centrifugation, and/or filtration. The purified sample is subjected to gel-permeation chromatography, ion-pair chromatography, ion-exchange chromatography, and preparative thin-layer chromatography. Buffered ion-exchange chromatography, close to neutrality, is necessary to prevent decomposition of the compounds to dimethylarsinic acid occurring at extremes of pH.

Arsenic compounds extracted with methanol from sea animals (cockle, trough shell, tuna, crab, sea cucumber, squid, etc.) have been separated by liquid chromatography. No single column has been found to be satisfactory for separating all organoarsenic compounds present in the seafoods. Four compounds (**Figure 1f**, **g**, **i2**, **i3**) have been found in animals. Arsenobetaine (**Figure 1g**) was found in all sea animals tested. There is considerable evidence that arsenocholine and arsenobetaine are rapidly excreted and almost nontoxic to humans.

Van Elteren and Šlejekovec investigated the stability of several arsenic compounds in food treatment procedures. Under usual food treatments (microwave treatment of 300 W of power for 120 min and boiling on a hot plate for 100 min), aqueous solutions of monomethylarsonic acid (MMAA), dimethylarsinic acid (DMAA), arsenobetaine (AsB), and tetramethylarsonium ion remained stable. Using γ-irradiation treatment with a dose of 10 kGy, 5% of AsB decomposed to trimethylarsineoxide (TMAO), and 1.5% of MMAA and 1% of DMAA decomposed to inorganic arsenic. After dry heating for 30 min at 160 °C, 10% of AsB decomposed to TMAO, 9% of MMAA to As(III), 6% of DMAA to MMAA, and tetramethylarsonium ions remained stable. Under unusually harsh treatments, however, decomposition of some compounds was found.

Sample Preparation

To determine total arsenic, the sample must be brought into solution. Care must be taken during the destruction of the sample to ensure that no arsenic is lost by vaporization of trivalent arsenic halides. Loss can usually be avoided by oxidizing the arsenic at an earlier stage by boiling with nitric acid under reflux.

In determining arsenic at trace levels in food, the Association of Official Analytical Chemists (AOAC; Article 9.1.01 D) recommends that a sample be digested with HNO_3 (heated in an oven at 150 °C oven for 2 h) in a closed system. In the closed system, a sample is placed in a digestion vessel (or cylinder), usually constructed of a fluorinated polymer such as polytetrafluoroethylene (PTFE) or perfluoroalkoxy (PFA). After the addition of digestion reagents, the cylinder is tightly sealed. The microwave-assisted alkali digestion with tetramethylammonium hydroxide ($(CH_3)_4NOH$, TMAH), which is used as a tissue solubilizer, is useful for the pretreatment of various biological samples prior to analysis. A mixture of nitric, sulfuric, and perchloric acids is also used for wet ashing of food samples. Monomethylarsonate, dimethylarsinate, arsenobetaine, and phenylarsonate are decomposed to inorganic arsenate. Dry ashing at 600 °C with magnesium nitrate for the destruction of meat and poultry is also used. The ash is dissolved in dilute hydrochloric acid.

To examine arsenic contamination of soils in Japan, arsenic is extracted by shaking the soil with 1 M hydrochloric acid (50 ml per 10 g of sample) for 30 min at 30 °C. The mixture is filtered through dry paper and subjected to hydride generation atomic absorption spectrometry.

Low levels of arsenic in water can be enriched by coprecipitation with ferric hydroxide after being oxidized to the pentavalent state with potassium permanganate. Triton X-100 has been used to aid the formation of large precipitates. It is claimed that both arsenate and arsenite are coprecipitated with iron hydroxide.

Methods of Analysis

There are several methods available for the determination of low levels of arsenic, and various analytical methods for determining arsenic in various concentrations are shown in Figure 2.

High-performance Liquid chromatography

High-performance liquid chromatography (HPLC) separation is most frequently employed for the speciation of arsenic species. Gas chromatography (GC) separation can also be used but requires complicated analyte derivatization steps to form volatile derivatives. Ion-pair chromatography in reverse mode, ion-exchange chromatography, and gel-permeation chromatography are often used. In reverse mode, C_{18} columns and polymer resin columns such as Hamilton polymer reversed-phase (PRP) are commonly used. Ion-pair chromatography is based on the distribution of ion-pairs between the aqueous mobile phase and the hydrophobic stationary phase. A quaternary ammonium ion such as tetrabutylammonium ion is often adopted for the separation of the anionic species. An alkyl sulfonate such as hexanesulfonate is often adopted for the separation of cationic species. Detection can be by UV absorption, conductivity, or

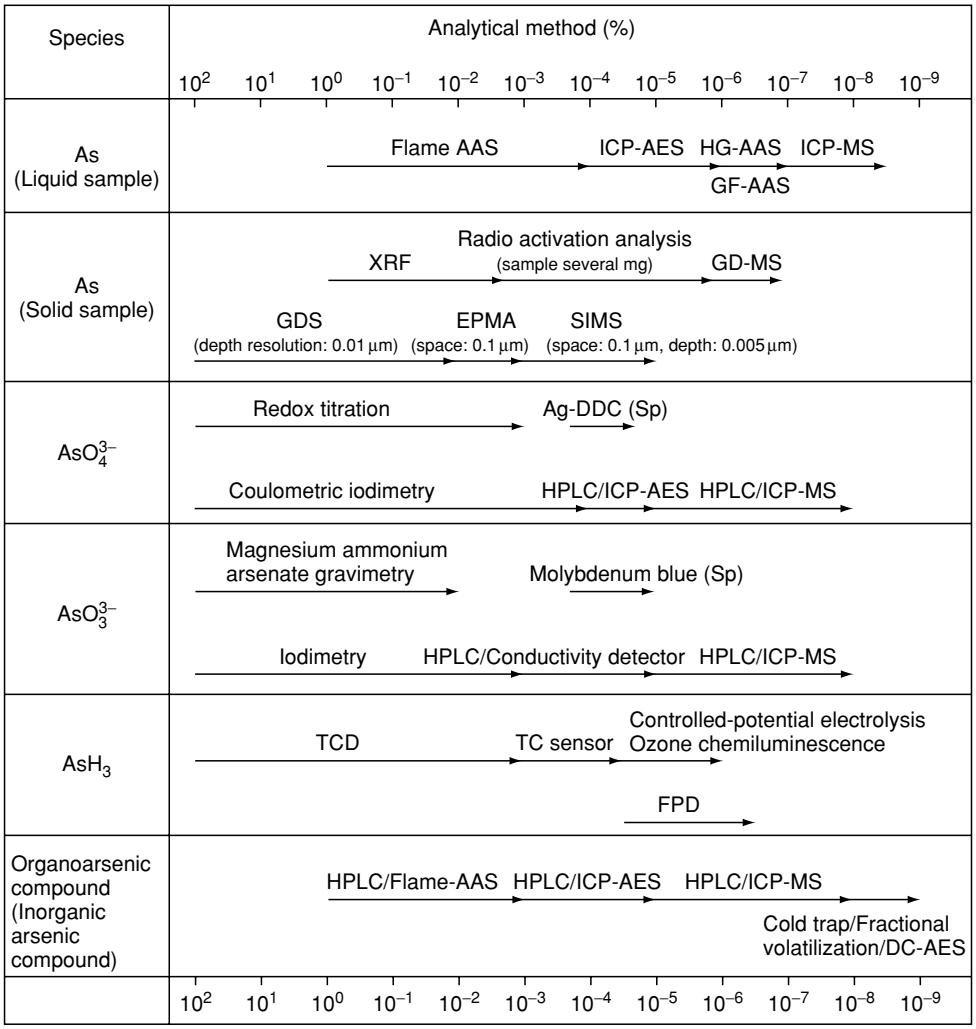

Figure 2 Analytical method for arsenic in various concentrations. From Mochizuki T (1993) Table of determination methods in seventeen elements for various concentrations (in Japanese). *Bunseki*: 262 with permission. AAS, Atomic absorption spectrometry; Ag-DDC, silver diethyldithiocarbamate; DC-AES, direct current 'discharge' atomic emission spectrometry; EPMA, electron probe microanalysis; FPD, flame photometric detector; GD-MS, glow discharge mass spectrometry; GDS, glow discharge emission spectrometry; GF-AAS, graphite furnace atomic absorption spectrometry; HG-AAS, hydride generation atomic absorption spectrometry; HPLC, high-performance liquid chromatography; ICP-AES, inductively coupled plasma atomic emission spectrometry; ICP-MS, inductively coupled plasma mass spectrometry; SIMS, secondary-ion mass spectrometry; Sp, spectrophotometry; TCD, thermal conductivity detector; XRF, X-ray fluorescence.

hyphenated techniques (atomic absorption and mass spectrometry).

Hydride Generation Method

There are two approaches for hydride generation (HG). The first is a selective HG for speciation followed by detector and is limited to the determination of the hydride-forming arsenic species. Differentiation of several arsenic species is possible in the HG technique with sodium borohydride as a reductant. Inorganic arsenic, monomethylarsonic acid, and dimethylarsinic acid are reduced, in acid solution, to arsine, monomethylarsine, and dimethylarsine, respectively, and collected in a cold trap. Differentiation of these arsenic compounds is possible by fractional volatilization of the arsines collected in the cold trap, because there are great differences in boiling point between arsine (AsH_3, $-55\,^{\circ}C$), monomethylarsine (CH_3AsH_2, $2\,^{\circ}C$) and dimethylarsine ($(CH_3)_2AsH$, $36\,^{\circ}C$). The choice of a proper packing material in the cold trap is important for the mutual separation of these arsines. A silanized diatomaceous earth impregnated with phenylmethylsilicon (15%) or hydrofluoric-acid-etched glass beads (0.3 mm) has been used. Arsenite (As(III)) (pH 6) and arsenate (As(V)) (pH 1) can be separated by pH-selective HG. Hydride generation for differentiation of arsenic species is becoming obsolete because very stable organoarsenic compounds such as arsenobetaine cannot form arsine.

The second removes total arsenic from the matrix and is generally combined with HPLC followed by decomposition (microwave digestion or UV decomposition) as a postcolumn prior to detector. The generated arsines are often concentrated by introduction to a cold trap before detection. The HG technique is used in combination with many different detection systems, including atomic absorption spectrometry (flame, quartz furnace, and graphite furnace), atomic emission spectrometry, mass spectrometry, atomic fluorescence spectrometry, spectrophotometry, and colorimetry. The HG technique is incorporated into the Gutzeit method and silver diethyldithiocarbamate (AgDDC) method.

Arsenic is reduced in acid solution by zinc or sodium borohydride to form arsine (AsH_3), which is volatile and is stripped by the hydrogen generated during the reduction; it can be collected by a trap cooled in liquid nitrogen or by a dilute iodine or bromine solution. With zinc as the reductant, prior reduction of arsenic to arsenite is necessary for the conversion of arsenate into arsine. A mixture of potassium iodide and tin(II) chloride also helps the reduction of zinc. In a strongly acid solution, both arsenate and arsenite are converted into arsine by sodium borohydride.

Gutzeit Method (Colorimetry)

The Gutzeit method is becoming obsolete, although it is highly sensitive and has been employed for many years for the detection and semiquantitative determination of traces of arsenic in foods, soils, and fertilizers. The sample is brought into solution by wet ashing or by other means. Arsenic is reduced to arsenite with potassium iodide and tin(II) chloride and further to arsine with zinc in acid solution. The arsine is stripped by the hydrogen formed during the reduction with zinc, and passed through a tube containing a strip of $HgBr_2$-impregnated paper. The paper darkens on reaction with arsine, and the length of the darkened area is proportional to the amount of arsenic.

In the field kit based on this method, the paper turns yellow on reaction with arsine, the concentration of arsenic is determined by comparison on a color chart, and its detection limit is sub-mg kg^{-1} levels.

Spectrophotometry

Spectrophotometric methods used for the determination of arsenic are based on either the formation of heteropolymolybdate, e.g., the molybdenum blue method, or the formation of arsenic diethydithiocarbamate. Only arsenic(V) reacts with molybdic acid to form heteropolymolybdate. For the determination of total inorganic arsenic, prior oxidation of arsenic(III) to arsenic(V) is required.

In the molybdenum (Mo) blue method, arsenic(V) reacts with molybdic acid in the presence of a reducing agent to form heteropolymolybdenum blue. Since this method is interfered with phosphate and silicate, it is used after isolation of arsenic by the hydride generation method. First, arsenic is converted into arsine. The arsine is then passed through an arsine trap solution containing sodium hypobromite. An ammonium molybdate–sulfuric acid solution is added to the arsine trap solution to react as follows:

$$AsO_4^{3-} + 12MoO_4^{2-} + 24H^+ \rightarrow AsMo_{12}O_{40}^{3-} + 12H_2O.$$

After reduction with hydrazinium sulfate, absorbance at 845 nm is measured. The color develops slowly (~ 75 min) at room temperature but is stable.

The silver diethyldithiocarbamate ($(C_2H_5)_2$ NCSSAg, AgDDC) method involves arsine generation and bubbling the arsine formed through a trapping solution in pyridine. Figure 3 shows an example of the apparatus used in this method. The glass wool in the sulfide scrubber is moistened with lead acetate

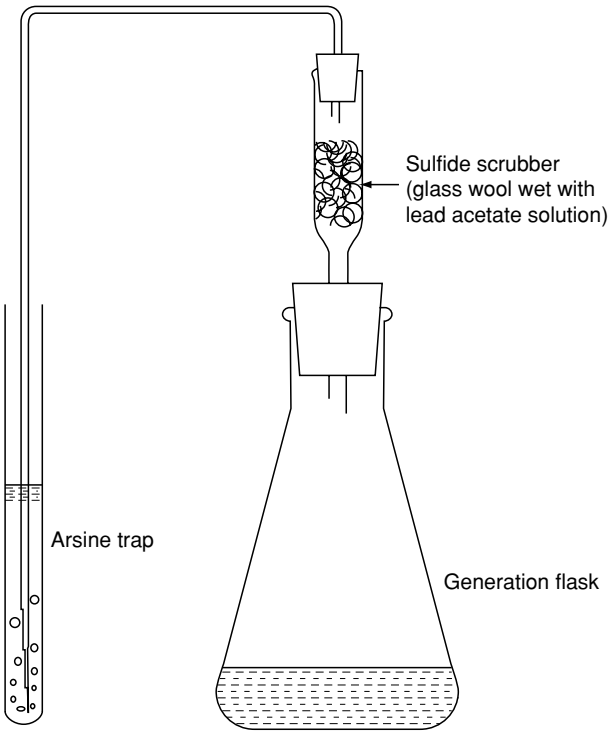

Figure 3 Apparatus for silver diethyldithiocarbamate method.

solution, and AgDDC–pyridine solution is placed in the arsine trap.

An appropriate volume of sample solution is placed into an arsine generator flask, and hydrochloric acid, potassium iodide, and tin(II) chloride are added to the flask. The mixture is allowed to stand for more than 15 min at room temperature. Zinc metal is added, and the generator flask is immediately sealed and the tube connected as illustrated. The solution is then allowed to react for 30 min. The trap is disconnected, and the trapping solution is mixed by gently drawing back and forth five times with an aspirator assembly. The absorbance based on an intense red color is measured at about 522 nm. The wavelength for the absorption maximum changes slightly with the composition of the trapping solution. If AgDDC–brucine–chloroform solution is used as a trapping solution, to prevent the production of foul-smelling pyridine, an appropriate volume of chloroform should be added to the trapping solution at the end of the trapping operation to compensate for the loss of chloroform evaporation.

Atomic Absorption Spectrometry

A hollow cathode lamp or an electrodeless discharge lamp can be employed as a light source. The primary absorption line is 193.7 nm.

Direct nebulization of extracts into a flame atomic absorption spectrometry is impractical at low concentrations because the flame strongly absorbs radiation at the wavelengths of sensitive absorption lines. Better sensitivity is obtained by the hydride generation technique, in which arsenic (as arsine vapor) is introduced into a hydrogen–argon reducing flame or into a heated tube for measurement of atomic absorption.

Graphite furnace atomic absorption spectrometry (GF-AAS) can also be employed. Arsenic sublimes at 613 °C and arsenic trichloride ($AsCl_3$) boils at 130 °C, but nickel arsenide melts at 968 °C. Successful results are obtained when nickel nitrate is added as a modifier to prevent loss of arsenic during the ashing process and to reduce interference. Similarly, magnesium nitrate, palladium and magnesium nitrates, and zirconyl nitrate can also be used as modifiers.

Instead of GF-AAS, quartz furnace atomic absorption spectrometry coupled with HG is also employed. The temperature in the quartz furnace is elevated to 900 °C.

Inductively Coupled Plasma Atomic Emission Spectrometry

Inductively coupled plasma atomic emission spectrometry (ICP-AES) is a simultaneous multielement analysis technique with a dynamic range. In ICP-AES, arsenic can be measured simultaneously in various emission lines (188.979, 180.042, 193.696, 197.192, or 228.812 nm) with different sensitivities. Atomic emission at 193.7 nm is less subject to interference, giving a detection limit of c. 0.02 mg l^{-1}. At 228.8 nm, cadmium interferes. Atomic emission at 188.979 nm, which is recommended in recent years, is giving a detection limit of μg 1^{-1} levels.

At 228.812 nm cadmium interferes. Arsenic (V) is converted into an arsenomolybdate, which is sorbed by the filter as the ion-associate with tetrapentylammonium ion, and the filter and the arsenomolybdate sorbed are dissolved in concentrated sulphuric acid, diluted with water and nebulized into the plasma torch. An emission line at 228.812 nm can be used. Interference from cadmium is eliminated because cadmium is not collected on the filter. Arsenic (III) dose not form arsenomolybdate and is not determined.

The sensitivity of conventional direct nebulization ICP-AES for arsenic is high but not sufficient to determine traces of arsenic. The sensitivities of ultrasonic nebulization ICP-AES for arsenic are about ten times better than that of conventional direct nebulization ICP-AES, but matrix (salt concentration) effect becomes larger. The sensitivity of ICP-AES can be improved by using HG techniques because of a

more efficient sample introduction and matrix removal. A continuous hydride generator is convenient. The detection limits of HG-ICP-AES for arsenic are about two orders in magnitude better than that of conventional direct nebulization. ICP-AES is possible to use a continuous monitoring. Do *et al*. investigated the speciation of arsenic in urine by HPLC-HG-ICP-AES. The limits of quantification were found to be 36 μg l^{-1} for AS(III), 9 μg l^{-1} for DMA, 57 μg l^{-1} for MMA, and 101 μg l^{-1} for As(V), applying a 100 μl sample injected for analysis.

Inductively Coupled Plasma Mass Spectrometry

Inductively coupled plasma mass spectrometry (ICP-MS) uses an inductively coupled plasma as an ion source and is one of the most sensitive methods for elemental determination. Arsenic is a monoisotopic element, that is, 100% of arsenic is ^{75}As.

The determination of arsenic using ICP-MS is interfered with by any molecular ion having a mass/charge ratio (*m/z*) equal to ^{75}As. Thus, the $^{75}As^{+}$ ion is interfered with by $^{40}Ar^{35}Cl^{+}$ formed by the combination of plasma gas and chloride ions, and so separation of the arsenic species from chloride ions is often required. The removal or reduction of argon from the excitation source by the addition of nitrogen gas or helium gas also eliminates, or significantly reduces, the formation of $^{40}Ar^{35}Cl^{+}$. Before ICP-MS, an HG method or a high performance liquid chromatography (HPLC) coupled with ICP-MS is required for the determination of arsenic species. The concentration of organic solvents must be lower because the introduction of organic solvents destabilizes the inductively coupled plasma.

Other Methods

Atomic fluorescence spectrometry (AFS) combined with HG is also available for trace arsenic determination. With LC (anion and cation exchange, liquid chromatography)–UV (decomposition)–HG–AFS system, six arsenic species (As(III), DMAA, MMAA, As(V), AsB, tetramethyarsonium ion) can be separated, with detection limits of about 0.5 μg of As per liter (100 μl injected). The limits of determination for arsenic in AFS, neutron activation without chemical separation, and pulse polarography are 0.8, 0.02, and 0.004 mg l^{-1}, respectively.

Certified Reference Materials

For quality control, the adoption of certified reference materials (CRM) is the best choice. A number of CRM for total arsenic are available, although only a few are available for speciation. Several easily available CRM for arsenic, relating to foods, are as follows: NIES CRM 9 Sargasso (certified value: 115 ± 9 mg kg^{-1}), BCR CRM 422 cod muscle (21.1 ± 0.5 mg kg^{-1}), NIST SRM 1566 oyster tissue (14.0 mg kg^{-1}), BCR CRM 279 *Ulva lactuca* (sea lettuce) (3.09 ± 0.20 mg kg^{-1}), NIST SRM 1548 total diet (0.20 ± 0.020 mg kg^{-1}), NIST SRM 1570a spinach leaves (0.068 mg kg^{-1}), BCR CRM 185 bovine liver (0.024 ± 0.003 mg kg^{-1}), NIST SRM 8433 corn bran (0.002 mg kg^{-1}). CRM are produced by several institutions:National Research Council of Canada, Community Bureau of Reference EC, Brussels, Belgium (Standards, Measurements and Testing Programme), International Atomic Energy Agency, Vienna, Austria, National Institute of Environmental Studies, Tsukuba, Japan, and National Institute of Standards and Technology, Gaithersburg, Maryland, USA.

See also: **Arsenic**: Requirements and Toxicology; **Chromatography**: High-performance Liquid Chromatography; **Mass Spectrometry**: Principles and Instrumentation; Applications; **Spectroscopy**: Atomic Emission and Absorption; Visible Spectroscopy and Colorimetry

Further Reading

Benramdane L, Bressolle F and Vallon JJ (1999) Arsenic speciation in humans and food products: a review. *Journal of Chromatographic Science* 37: 330.

Cullen WR and Reimer KJ (1989) Arsenic speciation in the environment. *Chemical Reviews* 89: 713.

Do B, Alet P, Pradeau D, Poupon J, Gulley-Gaillot M and Guyon F (2000) On-line reversed-phase liquid chromatography hydride generation emission spectrometry: speciation of arsenic in urine of patients intravenously treated with As_2O_3. *Journal of Chromatography B* 740: 179.

Ferguson JF and Gavis J (1972) A review of the arsenic cycle in natural waters. *Water Research* 6: 1259.

The Geological Society of Japan (1998) *Environmental Problems about Arsenic (in Japanese)* Tokyo: Tokai University Press.

Gerhartz W (ed.) (1985) *Ullmann's Encyclopedia of Industrial Chemistry*, 5th edn, vol. A3, pp. 113–141. Weinheim: VCH.

Guerin T, Astruc A and Astruc M (1999) Speciation of arsenic and selenium compounds by HPLC hyphenated to specific detectors: a review of the main separation techniques. *Talanta* 50: 1.

McIver KA (ed.) (1996) *Official Methods of Analysis of the Association of Official Analytical Chemists*, 16th edn. Arlington, VA: Association of Official Analytical Chemists.

Mochizuki T (1993) Table of determination methods in seventeen elements for various concentrations: Arsenic (in Japanese). *Bunseki*: 262.

Morita M and Shibata Y (1987) Speciation of arsenic compounds in marine life by high performance liquid chromatography combined with inductively coupled argon plasma atomic emission spectrometry. *Analytical Science* 3: 575.

Nriagu JO (ed.) (1994) Arsenic in the environment, Part I: Cycling and characterization In: *Advance in Environmental Science and Technology*, vol. 26. New York: John Wiley.

US National Research Council, Committee on Medical and Biologic Effects of Environmental Pollutants (1977) *Arsenic*. Washington, DC: National Academy of Science.

Van Elteren JT and Šlejekovec Z (1997) Ion-exchange separation of eight compounds by high-performance liquid chromatography–UV decomposition–hydride generation–atomic fluorescence spectrometry and stability tests for food treatment procedures. *Journal of Chromatography A* 789: 339.

Van Elteren JT, Šlejekovec Z and Byrne AR (1998) A dual arsenic speciation system combining liquid chromatographic and purge and trap-gas chromatographic separation with atomic fluorescence spectrometric detection. *Analytica Chimica Acta* 358: 51.

Requirements and Toxicology

G I Rehner, Institut für Ernährungswissenschaft, Giessen, Germany

Arsenic in the Environment

Arsenic is widely distributed in the natural environment. Virtually all soils and waters contain a variety of different arsenic compounds. As a byproduct of melting copper, lead, and several other metals, as well as from coal-fired power plants, high quantities of this metalloid are emitted into the atmosphere. Arsenic compounds are also used by the metallurgical and chemical industry. Phosphate fertilizer, as well as many herbicides, fungicides, wood preservatives, insecticides, and rodenticides, contains arsenic. Thus agriculture contributes significantly to arsenic pollution. Legislation in countries with developed conservation systems attempts to minimize the quantity of these pesticides as well as other sources of arsenic pollution.

Both the trivalent (arsenite) and the pentavalent (arsenate) state of arsenic can be detected in biological material. The biochemically important forms of arsenic are however methylated compounds: dimethylarsenic acid, methylarsonic acid, trimethylarsine oxide, trimethylarsonium ion, arsenocholine, arsenobetaine, and unstable arsenyl esters.

Arsenic accumulated in the environment can be subject to various biotransformations – reduction, oxidation, and methylation – performed by microorganisms. Methylation of inorganic arsenic in methanogenic bacteria under anaerobic conditions may be a detoxification mechanism for arsenic. Also marine algae transform arsenate into nonvolatile methylated arsenic compounds (methanoarsenic acid and dimethylarsenic acid) in seawater. These less toxic compounds are concentrated in higher marine organisms consuming the plankton. Several fungi are able to transform aerobically inorganic and organic arsenic compounds into very toxic volatile methylarsines.

Owing to the ubiquity of arsenic in the environment, foodstuffs of plant and animal origin as well as drinking water also contain arsenic. The maximally permissible arsenic concentration in drinking water, vegetables, fruits, and meat is regulated by law in most industrially developed countries. The US drinking water standard, for example, is fixed at $50\,\mu g\,l^{-1}$. However, several springs and spa waters contain 0.5–1.3 mg of arsenic per liter.

The range of arsenic levels in foodstuffs in the absence of significant pollution is as follows: cereals $0.05–0.4\,mg\,kg^{-1}$, fruits $0.03–1.0\,mg\,kg^{-1}$, vegetables $0.05–0.8\,mg\,kg^{-1}$; meat $0.05–1.4\,mg\,kg^{-1}$; dairy products $0.01–0.23\,mg\,kg^{-1}$. Higher concentrations – $1.5–15.3\,mg\,kg^{-1}$ – are registered in seafoods because marine organisms accumulate arsenic in the form of nontoxic organic compounds. The average daily intake of arsenic is estimated in the range of 12–40 μg. When diets contain substantial amounts of seafood, as is usual in Japan, 70–170 μg could be the realistic daily intake. (*See* **Shellfish**: Contamination and Spoilage of Molluscs and Crustaceans.)

Arsenic Metabolism

Intestinal Absorption

Arsenic is generally well absorbed by the gastrointestinal tract, although the solubility of the compounds seems to correlate with the rate of absorption. Intestinal absorption of most inorganic and organic compounds of arsenic is a quick process. However, some unusual compounds, such as Na-P-N-glycolylarsenilate, given orally to humans, are poorly absorbed.

Some investigations in chicks, mice, rats, and rabbits indicate that arsenate as well as arsenocholine readily penetrates both the mucosal and the serosal surfaces of the epithelial membrane and that arsenate and phosphate do not share a common transport pathway. Both arsenic compounds are apparently absorbed by simple diffusion following a concentration

gradient. On the other hand, *in vivo* and *in vitro* experiments in rats revealed a saturable transport process for pentavalent arsenic which was significantly reduced by phosphate. This result might indicate that arsenic and phosphate share the same secondary active carrier-mediated transport system.

Arsenic in Human Tissues

Once absorbed, arsenic is quickly distributed to all organs and tissues, probably as a complex with α-globulins. Apart from inorganic arsenic, blood also contains methylated metabolites formed in the liver.

If the uptake of arsenic is low, no significant accumulation in the soft tissues occurs. However, in unexposed humans relatively high arsenic concentrations are usually found in skin, nails, and hair. Because hair levels correlate with arsenic in the air as well as with renal arsenic excretion, hair concentration has often been used as an index of contamination with toxic amounts of arsenic. The median arsenic content of human hair was found to be $0.51\,\mu g\,g^{-1}$. The hair of workers at plants with arsenic-containing emissions was found to be 10–31 μg of arsenic per g of hair. Values greater than $3\,\mu g\,g^{-1}$ indicate toxic levels of uptake.

Arsenic Excretion

In general, the biological half-life of arsenic is short: the metalloid is excreted mainly in urine with a little in feces. Some compounds, such as arsenobetaine, which is the form found in marine animals, pass through into the urine unchanged. For this reason, despite the high arsenic content, intoxication by seafood is unlikely. Different methylated compounds are the major urinary metabolites.

Excretion via the bile is low after dietary intake of arsenic, but can significantly increase when drugs containing arsenic are administered intravenously. The hepatobiliary transport of arsenic is glutathione-dependent and associated with a profound increase in biliary excretion of glutathione; thus, hepatic glutathione depletion and diminished glutathione conjugation result.

Biotransformation of Arsenic

Inorganic arsenic is methylated in the liver with S-adenosylmethionine as methyl donator. As the first step, arsenate is reduced to arsenite by using glutathione. Arsenite is methylated to dimethylarsenic acid with monomethylarsenic as a precursor. Investigations with germfree animals have shown that methylation is mainly performed by the liver cells, with little contribution from the intestinal microflora.

The rate of methylation of inorganic arsenic varies considerably between species and individuals. As shown by studies in human volunteers exposed to specific doses of inorganic arsenic, the rate of renal excretion of arsenic increases with methylation efficiency. It could be assumed that the variations in interindividual susceptibility to arsenic in humans could be due to differences in the capacity of arsenic methylation.

Some organic compounds, especially arsenobetaine, are excreted unchanged. Arsenocholine can be incorporated into phospholipids, replacing choline, but the major part of this compound is excreted by the kidney.

Some mold species are able to produce volatile neurotoxic compounds by biotransforming arsenical pigments contained especially in wallpapers. Some lethal poisoning cases have been described resulting from such toxic products of aerobic and nonaerobic fungi as well as other primitive organisms.

Biochemical Function and Essentiality of Arsenic

Although a specific function for arsenic on a molecular level is unknown, some evidence has been accumulated indicating a distinct role for this element in several metabolic processes.

This could be the case in CH_3-group metabolism. Almost all organic forms of arsenic produced *in vivo* are methylated compounds. One of the trimethylated ones is arsenocholine, which can be incorporated into phospholipids, replacing choline. Choline is a methyl donator for different methylation reactions in intermediary metabolism. There is some evidence indicating that arsenocholine might have a similar function in labile methyl metabolism. As some experiments have shown that arsenic deprivation in the rat, chick, and hamster affects labile methyl metabolism, it could be speculated that the role of arsenocholine cannot be fulfilled by choline.

Arsenic binds strongly to sulfydryl groups which are known to participate in hundreds of enzymatic reactions. In this context it behaves like heavy metals that impair enzymatic catalysis. It is likely that the biochemical basis is the inhibition of a wide range of SH-enzymes by arsenic. But it cannot be ruled out in advance that such an interaction is not only of toxicological but also of biological significance.

Some investigators have reported that arsenic can provide partial protection from chronic selenosis, with the vitamin E level influencing the incorporation of selenium into the tissues. It has been shown that selenium toxicity for cultured mice fibroblasts can also be counteracted by arsenic. Interactions between

arsenic and cadmium have also been shown. Considering the common mode of action that these elements have at the cellular level, especially through interaction with SH-groups, these facts are not surprising. (*See* **Cadmium**: Toxicology; **Selenium**: Physiology.)

Studies with chicks suggest that arsenic is closely related to the metabolism of zinc. Depressed growth caused by arsenic deprivation was alleviated when dietary zinc was given in excess. It seems, however, not only that interactions between arsenic and zinc exist but also that the arginine status of the chicks plays a distinct role in these complicated correlations between the two trace elements and the amino acid. There is support for the possibility that arsenic participates – in a way that has yet to be clarified – in utilizing amino acids for protein synthesis, or in protein degradation, as well as in uric acid metabolism. (*See* **Zinc**: Physiology.)

Effects of arsenic on carbohydrate metabolism have been observed in experiments with guinea pigs. The most prominent finding after repeated arsenic (As_2O_3) administration was a marked decrease in total carbohydrate content of the liver, mainly owing to depletion of glycogen. (*See* **Carbohydrates**: Digestion, Absorption, and Metabolism.)

Some other biochemical facts in connection with arsenic – used in physiological as well as toxicological amounts – have been registered in the last few years. We are far from being able to classify the frequently conflicting facts and their significance for metabolism.

Since 1975, some conclusive evidence has been published supporting the suggestion of nutritional essentiality of arsenic. An element is considered essential if a dietary deficiency consistently results in a suboptimal biological function that is preventable or reversible by intake of a physiological amount of that element. Signs of arsenic deprivation were studied in four animal species – chick, goat, miniature pig, and rat. Safe and adequate intakes of arsenic for these species are not precisely defined. However, a diet containing less than 50 μg of arsenic per kg is denoted as an arsenic-deficient ration in these animal experiments, whereas 350–500 $\mu g\,kg^{-1}$ is considered a normal arsenic supply and 3.5–5 $mg\,kg^{-1}$ as a therapeutic dose.

The essentiality of arsenic has been systematically investigated in growing, pregnant, and lactating goats, as well as in miniature pigs and their offspring. Deficient rations contained less than 10 μg of arsenic per kg of semisynthetic diet, and control rations 350 $\mu g\,kg^{-1}$.

The following deficiency symptoms were described: feed consumption was reduced and this was correlated with diminished growth rates and milk production; a high abortion rate for goats and increased perinatal mortality for miniature pigs and rats were also registered; intrauterine arsenic depletion in goats, miniature pigs and rats resulted in significantly reduced birth weight of the offspring. The decreased milk-fat concentration could be a consequence of lowered triglyceride concentration in the blood serum of arsenic-deficient goats.

Histological changes, accompanied by ultrastructural alterations of the mitochondria, have been seen, especially in skeletal muscles, the myocardium, and liver tissue. Some modifications in the mineral composition of different organs with a remarkable retention of manganese were also connected to arsenic deficiency.

Many recent findings from studies with sophisticated new techniques have revealed possible sites of essential action of arsenic. It seems to have an important role in the conversion of methionine to different metabolites, e.g., S-adenosylmethionine, S-adenosylhomocysteine, and taurine. Additionally, arsenic affects the formation of the polyamines spermine, spermidine, and putrescine from arginine and the metabolism of labile methyl groups.

As shown in recent studies, arsenic induces the cellular synthesis of certain heat shock or stress proteins. This event is controlled at the transcriptional level and may involve changes in the methylation of core histones. Thus, arsenic seems to regulate gene expression. Arsenic also enhances DNA synthesis in unsensitized and in stimulated human lymphocytes.

Beneficial effects of arsenic preparations have been known – or at least believed – for centuries. Arsanilic acid used to be incorporated into pig and poultry feed in pharmacological doses as a growth-promoting agent. Intake of relatively high amounts of inorganic arsenic salts by humans and horses was reported from several parts of Europe. 'Arsenic eaters' were convinced that these preparations had protective effects against diseases, and increased physical condition and virility. Regular daily intakes of 0.5 g of arsenic have been reported; this kind of dose can only be tolerated once you are accustomed to arsenic for a long time. (*See* **Trace Elements**.)

Nevertheless, there is as yet no convincing evidence that arsenic is essential for humans, but the possibility cannot be ruled out. For definitive judgment of essentiality of arsenic, further critical investigations are necessary.

Only data from animal studies are available for estimating the approximate amount of arsenic required by humans, if it is essential for humans at all. Cautious assessments support a daily requirement of about 12–15 μg for persons consuming 8.4 MJ per

day. This amount is guaranteed by a mixed-food diet. But the evidence is too limited or controversial to recommend an ambiguous apparent beneficial intake for arsenic.

Toxicity of Arsenic

Arsenic has traditionally been associated with the poisoner and it was of interest to forensic medicine as a poison which is frequently used in suicide and homicide. Nowadays, intoxication mostly results from accidental intake of arsenic-contaminated food or beverages, or from environmental pollution. Depending on the dose, the nature of compounds, and the health status of the individual incorporating the poison, arsenic can cause both acute and chronic poisoning.

Acute Toxicity

The smallest lethal dose reported is 130 mg of arsenic, but total recovery can occur even if the ingested dose is several times higher. Signs of intoxication can appear a few minutes after ingesting a solution containing arsenic, but when the arsenic is taken in solid form signs may be delayed for up to 10 h. After ingestion of a lethal dose, death occurs after 12–24 h.

Depending on the severity of exposure to arsine (AsH_3), hemolytic symptoms can occur in a few minutes or after about 24 h. After massive exposure nausea, vomiting, and abdominal cramps begin within 2 h, followed by hematuria and skin color turning yellow within 24 h.

Chronic Toxicity

Although acute intoxication has become rare, arsenic is still a dangerous pollution agent for industrial workers and people living in the vicinity of emission sources.

A maximum daily intake of arsenic that is well tolerated over a longer period of time without impairing health and well-being cannot be fixed precisely. The chemical form of arsenic strongly affects its toxicity and variations in individual susceptibility are large. It was reported that orchard workers ingested for years as much as 6.8 mg of arsenic per day, i.e., about 500 times the estimated daily requirement, without any signs of intoxication. The ratio of the toxic to nutritional dose for rats was found to be about 1250.

The rank order of cytotoxicity of arsenic compounds – based on cell culture experiments – is as follows: arsenite > arsenate > dimethylarsenic acid > trimethylarsine oxide. Arsenobetaine, arsenocholine, and the tetramethylarsonium ion reveal only weak toxicity. Glutathione plays a significant role in protecting the cells from the effects of the toxic arsenic compounds.

The mode of incorporation of arsenic can be by respiration, by oral intake with drinking water and food, and by skin. For incorporation by skin arsenic trichloride is the only compound of importance. Irritations of the respiratory tract are common in areas surrounding coal-burning power stations, especially in eastern Europe where stations use coal containing up to 200 times as much arsenic as average coal. Human exposure to arsenic in the air is likely to increase in the future.

In some areas of the world the concentration of arsenic in drinking water exceeds the fixed standards. However, the correlation of different disorders with drinking water containing arsenic in a range above the drinking water standards of 50 $\mu g\, l^{-1}$ (0.6–6.0 mg l^{-1}) remains questionable. Most of the researchers (California, Oregon, Alaska) reported no adverse health effects; however, investigators form Minnesota, Taiwan, and Chile presented data supporting circulatory and skin disorders from arsenic consumption through drinking water. Reasons for these regional differences may be the additional exposure to other chemicals and/or fungi that potentiated the observed effects. In addition, differences in the nutritional status of the populations investigated remains at issue.

Contamination of foodstuffs with arsenic can result from air pollution, soil pollution by agricultural measures, or by accidental pollution. Use of arsenics as pesticides in vineyards resulted, in some cases, in poisoning by arsenic-contaminated wine. Cirrhosis of the liver in vine-growers was assumed to have at least partially been caused by arsenic.

Until the first decades of the 20th century a preparation of inorganic arsenic known as Fowler's solution was used for the treatment of dermatoses and even as a tonic against anemia. The treatment often lasted several years and caused signs of chronic arsenic poisoning, especially on the skin, with a high incidence of skin cancer.

Clinical symptoms of long-lasting arsenic intoxication develop over weeks and months. The first signs occur in the skin, mucous membranes, gastrointestinal tract, and neural system. The liver and circulatory system are affected later, if at all. The symptom complex resembles that of other progressive diseases, thus arsenic intoxication is difficult to diagnose. For this reason arsenic was one of the most common homicidal poisons for centuries.

Classical initial symptoms of chronic intoxication with arsenic are as follows: loss of appetite and, subsequently, weight; weakness; nausea; vomiting; dry

throat; fatigue; gastrointestinal symptoms; icterus and skin erythema. As the disease progresses, further symptoms can be found: loss of hair, fragile and pigmented nails, eczema, hyperkeratosis, and desquamation of the skin on hands and feet. Later, severe conjunctivitis, bronchitis, neurological symptoms, hearing disturbances and vascular disorder can occur. Some of the symptoms are reversible when exposure is interrupted, but individual variability is great, and neurological symptoms are particularly long-lasting.

Exposure to arsenic has been associated for decades not only with the development of different types of cancer but also with noncarcinogenic injuries, such as diabetes, peripheral neuropathy, and cardiovascular diseases. Some results about the biochemical mechanism of carcinogenesis induced by arsenic were obtained in the last few years. But, there are very few reliable data concerning the specific mechanism(s) of arsenic action on cellular level in context with noncarcinogenic effects. This is astonishing in so far as arsenic is assumed to inhibit more than 200 enzymes. But, direct enzyme inhibition seems not to be the common toxic effect of arsenic.

Therapeutic Measures

The most common antidote to arsenic poisoning is 2,3-dimercaptopropanol (BAL), which is a dithiol binding one arsenic molecule with its two SH-groups. Owing to the fact that this binding is more stable than binding of arsenic to the SH-groups of the amino acid residues, arsenic can be eliminiated via renal excretion.

Successful treatments with BAL have been reported, especially in cases of dermatitis caused by arsenic intoxication. After injection of the drug a powerful stimulation of renal arsenic excretion is registered during the first 3–6 days, accompanied by an improvement of the skin lesions. A further chelating agent recommended in cases of arsenic poisoning is D-penicillinamine. In severe cases of acute intoxication, exchange transfusion or hemodialysis may be necessary.

Teratogenic and Mutagenic Effects

Large doses of arsenic are reported to cause teratogenic alterations in different animal species. Embryotoxic effects of arsenite (1–40 $\mu mol\,l^{-1}$) and arsenate (10–400 $\mu mol\,l^{-1}$) on the development of mouse embryos during early organogenesis have been registered in the form of growth retardation and malformations of the central nervous system and of the extremities. Although no direct evidence exists for embryotoxicity of arsenic in humans, it cannot be conclusively ruled out that this toxic element may be involved in unaccountable early abortions and malformations claimed to be attributable to the toxicity of heavy metals.

Careful studies performed in the last decade point to the mutagenic potency of arsenic. Induction of chromosomal aberrations and sister chromatid exchanges have been described. A role of arsenic as a potential synergist to ionizing radiation, as well as inhibition of DNA repair by this element have been supposed.

The ability of arsenic to induce gene amplification may relate to its carcinogenic effects in humans since amplification of oncogenes is observed in many human tumors. (*See* **Mutagens**.)

Arsenic and Carcinogenesis

A number of epidemiological studies have been performed to elucidate the correlation between chronic arsenic exposure and cancer. The results are controversial. (*See* **Cancer**: Epidemiology.)

Inorganic arsenic causes a variety of benign skin lesions, including hyperpigmentation and hyperkeratosis. Some hyperkeratotic lesions and squamous cell carcinomas-*in-situ* may progress to invasive carcinoma and metastasize. Locally invasive but nonmetastasizing basal cell carcinomas may also occur. A dose–response relationship has been described between medicinally administered arsenic and the frequency of various skin lesions. Chronic intake of arsenical preparations for psoriasis is reported to cause malignant lesions in skin, lung, and liver.

A high incidence of carcinoma of the respiratory tract was registered in several epidemiological studies dealing with populations in areas highly polluted with arsenic. Factory workers exposed simultaneously to arsenic and sulfur dioxide ran a very high risk of developing multiple carcinoma.

In 1970 the occupational Safety and Health Administration of the US government fixed the maximum concentration of inorganic arsenic in the air to 10 $\mu g\,m^{-3}$ for occupational exposure. On the basis of several studies performed in factories causing arsenic pollution, the same institution predicted that the risk of lung carcinoma would rise exponentially with increasing pollution.

During the last decade, several research groups were engaged in clarifying the molecular basis of the carcinogenic effect of arsenic without being able to furnish a homogeneous interpretation.

It is generally recognized that arsenic is a human carcinogen or promoter of carcinogenesis. Arsenite acts as a comutagen by interfering with DNA repair, although a specific DNA-repair enzyme sensitive to arsenic has not been identified. Apparently only a few sensitive enzymes are responsible for arsenic-induced cellular toxicity. Thus, arsenic-induced

comutagenesis and inhibition of DNA-repair are probably not the result of direct enzyme inhibition, but may be an indirect effect caused by arsenic-induced changes in cellular redox levels or alterations in signal transduction pathways and consequent changes in gene expression. Cellular glutathione level was found to be the most important determinant of arsenic sensitivity in cancer cells.

Arsenic was shown to inhibit ubiquitin-dependent proteolysis. This suggests that a tumor suppressor gene product or some other aspect of the ubiquitin system may be a target for arsenic toxicity and that disruption of the ubiquitin system may contribute to the genotoxicity and carcinogenicity of arsenite. Alteration of DNA methylation by arsenic offers a further plausible hypothesis for the carcinogenic mechanism of arsenic effect.

Areas considered in need of future research are: interaction of arsenic with cellular constituents, especially with genetic material; interaction of arsenic with other metals; and development of animal models and/or cell systems to study the effects.

See also: **Cadmium**: Toxicology; **Cancer**: Epidemiology; **Carbohydrates**: Digestion, Absorption, and Metabolism; **Mutagens**; **Selenium**: Physiology; **Shellfish**: Contamination and Spoilage of Molluscs and Crustaceans; **Trace Elements**; **Zinc**: Physiology

Further Reading

Anke M (1986) Arsenic. In: Mertz W (ed.) *Trace Elements in Human and Animal Nutrition*, 5th edn, vol. 2, pp. 347–372. Orlando: Academic Press.

Aposhian HV (1997) Enzymatic methylation of arsenic species and other new approaches to arsenic toxicity. *Annual Review of Pharmacology and Toxicology* 37: 397–419.

Fowler BA (ed.) (1983) *Biological and Environmental Effect of Arsenic*. Amsterdam: Elsevier Science.

Lederer WH and Fensterheim RJ (eds) (1983) *Arsenic: Industrial, Biomedical, Environmental Perspectives*. New York: Van Nostrand Reinhold.

Nielsen FH (1990) Other trace elements. In: Brown ML (ed.) *Present Knowledge in Nutrition*, 6th edn, pp. 294–296. Washington, DC: International Life Science Institute, Nutrition Foundation.

Nielsen FH (1996) Other trace elements. In: Ziegler EE and Filer LJ (eds) *Present Knowledge in Nutrition*, 7th edn, pp. 353–377. Washington, DC: ILSI Press.

Nielsen FH (1996) How should dietary guidance be given for mineral elements with beneficial actions or suspected of being essential? *Journal of Nutrition* 126: 2377S–2385S.

ASCORBIC ACID

Contents

Properties and Determination

M A Kall, Danish Veterinary and Food Administration, Søborg, Denmark

Background

Vitamin C has both enzymatic and chemical functions, and it is a cofactor for several enzymatic reactions involving metabolism of collagen and amino acids, for example. Vitamin C was discovered as an antiscorbutyric factor less than one century ago. The trivial name of the active component L-ascorbic acid was adopted to reflect the antiscorbutyric effect of vitamin C. Ascorbic acid (AA) is a water-soluble vitamin, and under most physiological conditions, it is present in the mono-protonated form. It is considered to be the most important antioxidant in the extracellular compartment. As an antioxidant, AA donates electrons to both intercellular and extracellular reactions. The oxidation product of such reactions, dehydroascorbic acid (DHAA), forms a reversible redox system with AA, and this is an important characteristic of the reactions involving vitamin C. Moreover, AA has scavenger properties and can eliminate free radicals under the formation of semidehydroascorbic acid, a nonreactive radical. The antioxidative properties of AA are frequently utilized in the food industry. Ascorbic acid is widely

Table 1 Physical and chemical properties of ascorbic acid

Formula	$C_6H_8O_6$
Molecular weight (g mol^{-1})	176.13
Melting point (°C)	190–192
pH (5 mg ml^{-1})	3
pH (50 mg ml^{-1})	2
pK_1	4.17
pK_2	11.57
E_o (pH 5) (mV)	+127
Solubility	
Water	1 g per 3 ml
Alcohol	1 g per 30 ml
Absolute alcohol	1 g per 50 ml

Data are taken from *The Merck Index*, 12th edn. (1996) Whitehouse Station, NJ: Merck.

used as a food additive to improve taste and color, and to restore the vitamin C lost during processing and storage.

Both AA and DHAA are very unstable in aqueous solution, and measurable degradation may occur within minutes to hours, because of oxidation caused by a wide range of chemical compounds and enzymes. Vitamin C is present in most foodstuffs of vegetable origin but in varying concentrations. In order to determine vitamin C, it is important to take into consideration the reactivity and lack of stability of AA and DHAA. Over the last three decades, numerous methods have been published describing extraction and analyses of vitamin C in foodstuffs and physiological samples. These methods cover a wide range of analytical principles from direct colorimetric determination or redox titration to modern, advanced high-performance liquid chromatography (HPLC) techniques, and furthermore, they cover a wide range of objectives and different requirements.

This article will focus on the HPLC determination of vitamin C in foodstuffs and physiological samples, and will discuss some important considerations regarding the handling of samples and choice of analytical procedures (see **Table 1**).

Chemical Properties

In aqueous solutions, AA is a strong reductant and is rapidly and reversibly transformed into DHAA in a two-electron oxidation. This reaction may be catalyzed by a variety of oxidative compounds such as molecular oxygen, heavy metal ions, halogens, and enzymes, e.g., ascorbate oxidase. The oxidation of AA to DHAA *in vitro* can be reversed by thiol-containing reductants such as cysteine, homocysteine, glutation, mercaptoethanol, dimercaptopropanol, or dithiothreitol (DTT). Dehydroascorbic acid also has reducing properties, especially in alkaline solutions, and may be irreversibly oxidized to 2,3-diketogulonic acid and further to several other compounds like threonic acid and oxalic acid. The oxidation and ring-opening process of the lactone ring of DHAA causes an irreversible loss of both the vitamin C activity and the antioxidant properties.

Ascorbic acid is a dibasic acid with an enoldiol group built into a five-membered heterocyclic lactone ring. The AA molecule contains two chiral carbon atoms, C4 and C5 (**Figure 1**), and theoretically, three stereoisomers of L-ascorbic acid therefore exist: D-ascorbic acid, D-isoascorbic acid, and L-isoascorbic acid. In the human organism, L-ascorbic acid is the only active form. The synthetic isomer D-isoascorbic (IAA) acid has the same chemical properties as AA, but this enantiomer of L-ascorbic acid has only approximately 5% vitamin C activity compared with AA. D-Isoascorbic acid is often used as an antioxidant additive in processed foods.

The overall reaction for oxidation of AA by dioxygen is given in eqn (1).

$$2AA + O_2 \rightarrow 2DHAA + 2H_2O. \qquad (1)$$

As mentioned previously, AA is easily oxidized, forming an intermediate radical of low reactivity. Ascorbic acid can scavenge very reactive species such as the superoxide anion (O_2^-), hydroxyl radical (HO•), alkoxyl radicals (RO•), and peroxyl radicals (ROO•). The poor reactivity of the semidehydroascorbic acid radical (AA•) (see **Figure 1**) may account for the antioxidant effect of AA. In foods, a free radical (X•) can start a chain reaction and cause damage to the food product (see eqns (2) and (3)).

$$X\bullet + RH \rightarrow R\bullet + XH \qquad (2)$$

$$R\bullet + O_2 \rightarrow ROO\bullet \qquad (3a)$$

$$ROO\bullet + RH \rightarrow ROOH + R\bullet. \qquad (3b)$$

Ascorbic acid can intercept these radicals and stop the chain reaction and, by disproportionation of the formed ascorbic radical, eliminate the unpaired electron (eqn (4)).

$$\begin{aligned} &R\bullet + AA \rightarrow RH + AA\bullet \\ &ROO\bullet + AA \rightarrow ROOH + AA\bullet \\ &2AA\bullet \rightarrow AA + DHAA. \end{aligned} \qquad (4)$$

Vitamin E can also act as a radical scavenger, and AA plays an important role by regeneration of vitamin E after oxidative attack and by elimination of the radical (eqn (5)).

$$\begin{aligned} &\text{Vit-E} + ROO\bullet \rightarrow \text{Vit-E}\bullet + ROOH \\ &\text{Vit-E}\bullet + AA \rightarrow \text{Vit-E} + AA\bullet. \end{aligned} \qquad (5)$$

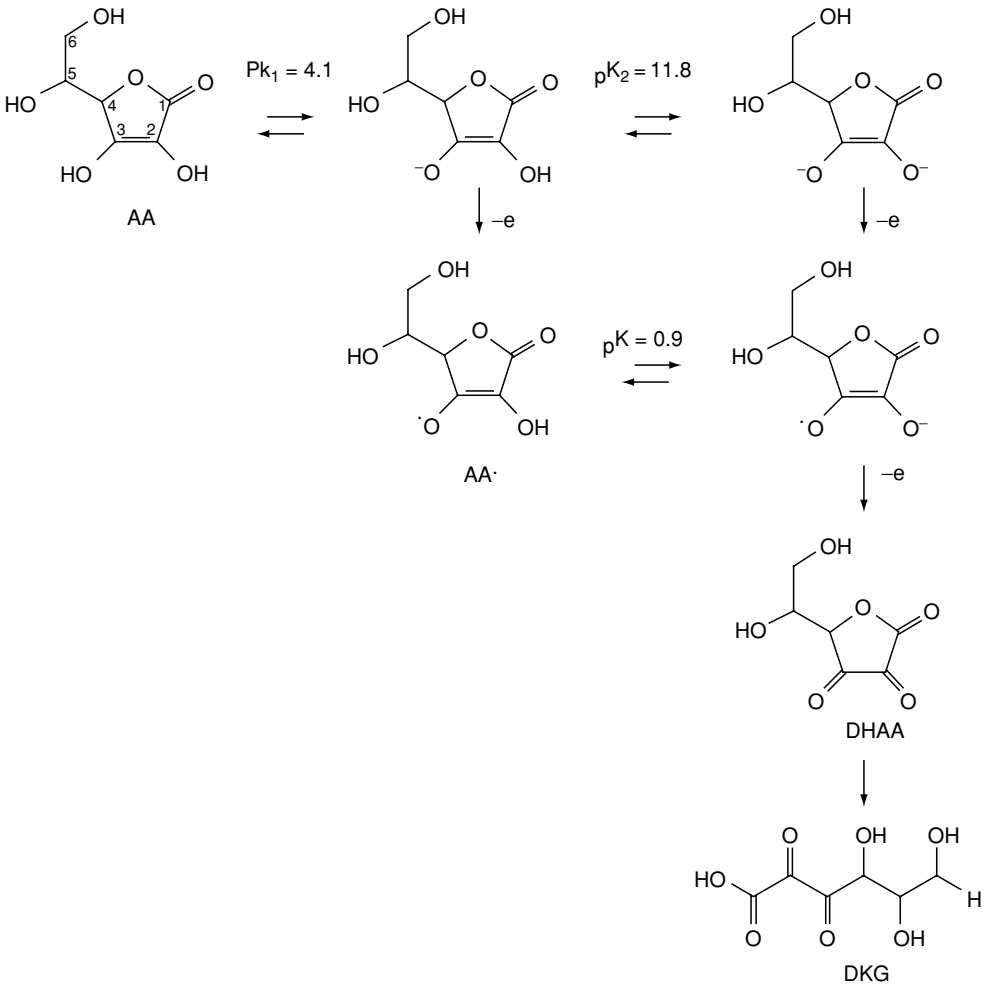

Figure 1 Reaction pathways for oxidation of ascorbic acid. AA, ascorbic acid; AA•, semidehydroascorbic radical; DHAA, dehydroascorbic acid; DKG, 2,3-diketogluconic acid.

Ascorbic acid may also have prooxidative properties. When metal ions are present, AA can stimulate the formation of free radicals. The oxidation of AA by dioxygen (eqn (1)) is very strongly catalyzed by Cu(II) and Fe(III) ions but may also be catalyzed by other metal ions capable of undergoing redox reactions between two valence states. According to eqns (6) to (8), reduction of dioxygen by AA catalyzed by Fe(III) may lead to formation of the hydroxyl radical.

$$\mathrm{Fe(III) + AA + O_2 \rightarrow Fe(III) + DHAA + H_2O_2} \quad (6)$$

$$\mathrm{Fe(III) + AA \rightarrow Fe(II) + AA\bullet} \quad (7)$$

$$\mathrm{Fe(II) + H_2O_2 \rightarrow Fe(III) + OH\bullet + OH^-}. \quad (8)$$

The hydroxyl radical is extremely reactive, and reacts with all organic substances.

Considering analytical aspects of AA, avoidance of oxidation during storage, extraction, and analysis is by far the most challenging aspect of the analysis. Metal ions catalyze the oxidation of the mono-anion of AA only (see Figure 1). Thus, pH has a marker effect on the oxidation rate, especially near the first pK_a value at 4.17. In order to ensure full protonization of AA during extraction of samples and storage of extracts, the pH should not exceed 2. The autooxidation of AA catalyzed by metal ions (eq (6)) is a first-order reaction with regard to AA, metal ions, and oxygen concentrations. Evacuation of dissolved oxygen from samples and solutions is important in preventing oxidation of AA. The catalytic autooxidation can be further eliminated by complex formation of the metal ions with adequate chelators. Complexes tend to stabilize the higher oxidation states of coordinated metal ions, e.g., by lowering the redox potential, and oxidation reactions by complex bound metal ions are therefore generally much slower than reactions with the metal ion aqua complexes. Addition of a chelating agent, such as ethanediaminetetraacetic acid (EDTA) or oxalic acid, to the extraction buffer will in most cases prevent metal-ion-catalyzed

autooxidation or direct oxidation of AA. In several cases, however, it has been reported that the presence of EDTA in sample extract in fact increases oxidation of AA caused by Fe(III) instead of preventing it. One plausible explanation could be the formation of OH• generated by the reaction in (eqn (8)) and that OH• oxidizes AA to AA•. For several ligands examined at neutral pH, it has been found that iron complexes with a high stability were better catalysts for AA oxidation than iron complexes with a low stability. The opposite was found to be true for copper complexes. These trends have been attributed to the change in redox potentials and to the rate of reoxidation of the metal ion. It has been shown that the Fe(II)–EDTA complex reacts with hydrogen peroxide approximately 100-fold faster than the Fe(II) $(H_2O)_6$ complex, i.e., in eqn (8) (see Table 4).

Occurrence in Food

The vitamin C contents of several important food sources are listed in Table 2. The data originate from the Danish Composition of Foods and reveal a large variation in the amount of vitamin C between sources but also notable variations within the sources. This variation within species is known to be true for many vitamins in fruit and vegetables and may be due to different genetic properties as well as different growths conditions, e.g., soil, exposure to sun light, and fertilization. Storage conditions may also affect the content. From Table 2, it is clear that any processing of the food may cause loss of vitamin C. Table 3 shows the amount of vitamin C in some foods distributed as AA and DHAA, and some examples of the occurrence of IAA and DHIAA in processed foodstuffs.

Table 2 Vitamin C content of important food sources

Food source	*Concentration mg per 100 g*	
	Mean	*Range*
Vegetables		
Broccoli, raw	115	92–131
Broccoli, frozen	54	
Broccoli, boiled	37	
Cauliflower, raw	70	50–103
Cauliflower, frozen	49	45–52
Cabbage, raw	45	25–62
Cabbage, boiled	24	
Green beans, frozen	24	12–32
Potato, raw	24	10–40
Potato, boiled	7	
Pepper, red, raw	195	149–240
Pepper, green, raw	86	38–154
Spinach, frozen	16	9–12
Spinach, boiled	12	10–13
Carrots, raw	5	3–7
Carrots, frozen	4	3–5
Carrots, canned	3	2–3
Fruit		
Orange, raw	55	37–79
Orange juice, canned	34	33–35
Lemon, raw	49	35–62
Elderberry, raw	29	12–58
Banana, raw	11	6–16
Grape, raw	11	9–13
Apple, raw	9	2–22
Peach, raw	7	
Apricot, raw	10	
Apricot, dried	2	

Data taken from Saxholt E and MØller A (1996) *The Composition of Food*, 4th edn. Copenhagen: Danish Veterinary and Food Administration, Gylendal.

Table 3 Analytical vitamin C data for some food sources, distributed as AA, DHAA, IAA, and DHIAA

Food source	*Concentration mg per 100 g*			
	AA	*DHAA*	*IAA*	*DHIAA*
Vegetables				
Broccoli, raw	89±2	7.7±0.6		
Broccoli, boiled	37±1	2.6±0.6		
Broccoli, frozen	54±2	2.6		
Cabbage, raw	42±3			
Cabbage, frozen	24±2			
Pepper, red, raw	151±3	4±1		
Pepper, green, raw	129±1	5		
Potato, raw	8	3		
Potato, boiled	7±1	1.3±0.6		
Spinach, raw	52±3			
Spinach, boiled	20±1			
Spinach, frozen	22	3		
Tomato, raw	11±1	3		
Fruit				
Orange	55±2.5	8±1		
Banana	15±.5	3.3±0.6		
Grapefruit	21±0.6	2.3±0.6		
Processed meats				
Canned ham		0.6	23.4	15.6
Canned sausage	0.6	0.8	4.8	3.9
Baby food				
Peaches, summer	31.6	1.9	6.2	0.9
Pears, summer	10.3	0.8	1.5	0.2
Luncheon meat				
Pork		1.3	8.4	10.5
Devil ham spread	4.1	1.2	14	23.4
Pork–chicken–veal spread	4.7	1.1	28.6	15.5
Chicken spread	4.2	0.9	29.5	12.7
Pâté de foie gras	4.5	1.1	16.8	18.7

Data taken from Behrens WA and Madère R (1994) Ascorbic acid, dehydroascorbic acid and dehydroisoascorbic acid in selected food products. *Journal of Food Composition and Analysis* 7: 158–170; Hodiroglou N, Madère R, and Behrens W (1998) Electrochemical determination of ascorbic acid and isoascorbic acid in ground meat and in processed foods by high pressure liquid chromatography. *Journal of Food Composition and Analysis* 11: 89–96; Vanderslice JT, Higgs DJ, Hayes, and Block G (1990) Ascorbic acid and dehydroascorbic acid content of food as eaten. *Journal of Food Composition and Analysis* 3: 105–118.

Occurrence and Stability in Physiological Samples

The concentration of AA and DHAA in blood plasma samples may depend on the nutritional status of the individual. In general, women have higher plasma concentrations of total AA than men, and non-smokers have higher plasma concentrations of total AA than smokers. The plasma concentrations of total AA reported in most studies are in the range from approximately 40 to 85 $\mu mol\,l^{-1}$, corresponding to 7–15 $\mu g\,ml^{-1}$. The stability of AA in plasma is limited to approximately 5 h at 4 °C, but addition of 1:1 of 10% MPA to the plasma can prolong the stability of AA as well as DHAA up to 6 months by storage at − 80 °C. The stability of total AA may be prolonged further to several years by addition of DTT prior to lyophilization of the plasma sample and subsequent storage at − 80 °C.

Measurement

Measurement of AA may be accomplished by several slightly different principles, but in order to choose the optimal analytical conditions, the method by which to determine vitamin C should be considered. Dehydroascorbic acid is not directly measurable by any common HPLC detector principle and has to be transformed prior to measurement. If the sample extract is treated with a reducing agent, DHAA is reduced to AA and can be measured by UV- or electrochemical detection. In the case where determination of the ratio between AA and DHAA is required, samples should be analyzed with and without reduction, and DHAA can be calculated as the difference between the AA concentrations measured in the two analyses. This method is often referred to as the 'subtraction method.' Another possibility is to determine AA and DHAA simultaneously, but this method requires a more complex analytical procedure.

Moreover, one should consider differentiation between AA and IAA, and between DHAA and dehydroisoascorbic acid (DHIAA). Before choosing an appropriate method, it is important to define the objectives for the study and the requirements to the results. For example, in a study of degradation of vitamin C during production or storage of a foodstuff, it may be interesting to determine DHAA as well as AA. However, in order to collect data regarding food composition, to conduct a survey of the occurrence of vitamin C in a customary diet, or to control the declared content of vitamin C in a fortified foodstuff, it may be relevant to determine only the amount of total AA.

Both the capacity and costs of the analysis are important considerations. Different measurement principles often demand different sample preparations, and not all preparation procedures result equally in stabilization of AA. Studies on vitamin C in physiological samples may require determination of DHAA and AA separately, but because of a usually very low DHAA concentration compared with AA, the stability of both AA and DHAA in the extract is important.

Extraction

Regardless of the choice of measurement principle, AA has to be liberated from the sample matrices prior to analysis. The extraction procedure should be able to eliminate oxidative degradation of AA caused by dissolved oxygen, metal ion, and enzymes, and also be able to minimize hydrolysis of DHAA. Owing to the instability and high polarity of AA and DHAA, actual sample clean-up is difficult to perform. Fortunately, suspension or dilution of the sample in diluted acid with subsequent centrifugation and filtration provides a sufficient extraction of AA and DHAA from most matrices. In order to protect AA and DHAA from degradation during the extraction, different extraction solutions have been described in the literature, e.g., metaphosphoric acid (MPA), acetic acid, trichloroacetic acid, and perchloric acid. Optimal stabilization of AA and DHAA is obtained at pH values below 2. The majority of papers dealing with extraction of AA recommend MPA in concentrations ranging from 1 to 10% (w/v), but the addition of a chelating agent may often be necessary in order to complex bind metal ions. Chelators such as EDTA and oxalic acid are often used. As mentioned previously, EDTA may in some cases promote Fe(III)-catalyzed oxidation of AA instead of preventing it (see Table 4).

Oxidation of AA can be reduced further by handling the samples in a N_2 or CO_2 atmosphere, by reducing the temperature, and by light protection.

Table 4 Stability constants for complexes formed between some redox active metal ions and EDTA, oxalic acid and ascorbic acid

	EDTA	*Oxalic acid*	*Ascorbic acid*
Log K			
Fe^{2+}		5.15	1.99
Fe^{3+}	25	18.49	
Cu^{2+}	18.7	10.27	1.57

Data are from Martell AE and Smith RM (1977) *Critical Stability Constants*, vol. 3. New York: Plenum Press; Martell AE and Hancock RD (1996) Metal complexes in aqueous solutions. In: Fackler JP (ed.) *Modern Inorganic Chemistry*. New York: Plenum Press.

Determination

The classic chemistry associated with spectroscopic methods can be divided into two general categories: methods using redox indicators and methods involving formation of chromofor or flourofor derivatives. The principal reactions for some classical methods of direct determination of vitamin C are briefly reviewed, but HPLC methods will be discussed in detail.

Redox Reactions

2,2′-Dipyridyl This method is based on the reduction of Fe(III) to Fe(II) by AA. Fe(II) forms a complex with 2,2′-dipyridyl that can be quantified colorimetrically. Other chromogen formation Fe(II) complexing agents exist, e.g., ferrozine and 2,4,6-tripyridyl-*S*-triazine.

2,6-Dichloroindophenol This method is based on reduction of 2,6-dichloroindophenol by AA. In acidic solution, 2,6-dichloroindophenol has an absorption maximum at 518 nm and, when reduced by AA, this chromophor disappears. The specificity of this method is limited owing to the presence of naturally occurring reductants of 2,6-dichloroindophenol in fruits and vegetables. This method will typically determine 5–10% higher amounts of AA in fruit samples than a HPLC method.

Derivatization Reactions

2,4-Dinitrophyl hydrazine Ascorbic acid has to be oxidized to DHAA prior to the derivatization. The reaction between DHAA and 2,4-dinitrophenyl hydrazine results in the formation of the bis-2,4-dinitrophenylhydrazone derivative, which can be measured at 520 nm. The specificity of the method is limited, owing to unspecific reactions of 2,4-dinitrophenyl hydrazine with 5- and 6-carbon, sugar-like, nonascorbic compounds.

***o*-Phenyldiamine** Ascorbic acid has to be oxidized to DHAA prior to the derivatization. The reaction between DHAA and *o*-phenyldiamine (OPD) results in the formation of an intensive fluorophore: 3(1,2-dihydroxyethyl)furo[3,4-b]quinooxaline-1-one (DFQ) (see Figure 2). The fluorophore can be measured by the emission wavelength at 430 nm when excited at 350 nm. This reaction can be applied to direct measurement or to a chromatographic method. Some interference should be expected from naturally occurring fluorescing compounds if the method is used for direct measurement.

HPLC Determination of Vitamin C

For common foodstuff analysis and control analysis, the amount of vitamin C usually exceeds 1 mg per 100 g of sample, and in this case, a HPLC system coupled with ultraviolet (UV) detection provides adequate sensitivity. In foodstuffs with amounts less than 1 mg per 100 g and in physiological samples, where the concentration often is below 1 μg ml^{-1}, it may be necessary either to use an electrochemical detector or to transform AA into a fluorescent compound in order to increase the sensitivity by fluorometric measurement. In addition, the electrochemical and fluorescent detection are more specific than the UV detection.

In principle, three different strategies for determination of vitamin C by HPLC exist:

1. reduction of DHAA to AA followed by measurement of total AA;
2. oxidation of AA to DHAA followed by measurement of total DHAA; and
3. simultaneous determination of AA and DHAA, by pre- or postcolumn derivatization of DHAA, perhaps with additional postcolumn oxidation of AA.

As mentioned previously, extraction with MPA provides the most efficient protection of AA during extraction and subsequent analysis. However, MPA may cause serious analytical interactions with silica-based column materials. Ion-pair reversed-phase chromatography with C18 column material or pseudonormal-phase chromatography with an aminopropyl column material provide separation of AA and IAA and potential coeluting compounds with high efficiency and selectivity, but the use of MPA in the extraction buffer may affect the robustness of the

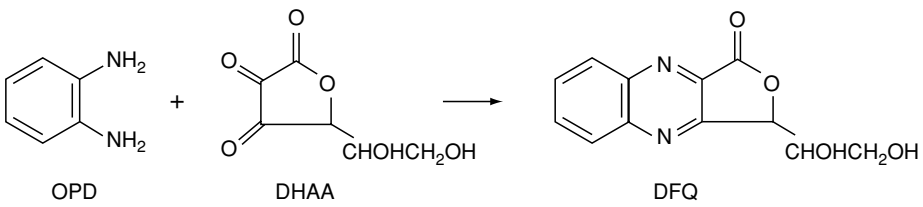

Figure 2 Reaction of DHAA with *o*-phenyldiamin (OPD) under formation of 3(1,2-dihydroxyethyl)furo[3,4-b]quinooxaline-1-one (DFQ).

chromatographic performance. The interactions may result in incomplete separation of the four compounds and cause a variation in retention time of the compound peaks. Furthermore, the use of EDTA in extraction buffers may result in a time-consuming analysis owing to a long retention time of the ligand. The interference problems with the silica-based polymer column materials caused by injection of MPA may be solved by using a polystyrene divinyl benzene (PLRP-S) polymer column material. An effluent based on 0.2 M phosphate buffer applied to the PLRP-S column performs adequate separation of AA and IAA. The PLRP-S column is compatible with injection of MPA in high concentrations in sample extracts and provides a very stable system for separation of the compounds without baseline- and retention drift. However, the efficiency and selectivity of the PLRP-S column are low with regard to AA and coeluting compounds and particularly DHAA and DHIAA, and the column should be used in combination with electrochemical or fluorescence detection only (see Figure 5).

Determination of vitamin C as total ascorbic acid Dehydroascorbic acid is reduced to AA at neutral pH in the presence of thiol-containing reducing agents such as cysteine, homocysteine, 2-mercaptoethanol, and DTT. During incubation of the sample extract with a reductant, DHAA should be transformed quantitatively into AA. Addition of a reducing agent to the sample extract will prolong the stability of vitamin C in the sample extract during the extraction procedure as well as the storage period in the autosampler during HPLC analysis. Ascorbic acid is protected against oxidation, and since DHAA is reduced to AA, DHAA is protected from hydrolyses to 2,3-diketogulonic acid. Opposite AA and DHAA, thiol-containing reductants have both a low stability and a low reducing capacity in acidic environments. Sample preparation is often carried out by MPA extraction at low pH followed by adjustment of the pH to neutral, and incubation with a reductant. Subsequently pH has to be reduced, to ensure the stability of the AA. This procedure is more laborious, and as a result of the degradation of reductants at a low pH, the reducing sample environment, and thus stabilization of the AA in extracts, is maintained for less than 24 h. It should be noted that a similar stability of AA and DHAA without addition of a reductant can be obtained by optimizing the extraction buffer, regarding pH and temperature, and the concentrations of MPA, and chelator.

Recently, tris[2-carboxyethyl]phosphine (TCEP) has been described as a promising reductant of DHAA in physiological samples. The major advantage of this compound is a high stability and a high reductive capacity at a low pH, in contrast to the commonly used reductants. It has been shown that the stabilization of AA in sample extracts can be maintained for at least 96 h and that TCEP may not cause any interference with the chromatographic system (see Figure 3).

Determination of vitamin C as total dehydroascorbic acid This principle is based on oxidation and derivatization prior to HPLC analysis. The quinoxaline derivatives are more hydrophobic than AA and DHAA, and it is possible to obtain a reversed-phase HPLC method with a high efficiency, selectivity, and baseline separation of the two quinoxaline derivatives corresponding to AA/DHAA and IAA/DHIAA. In order to obtain adequate retention times, an organic modifier has to be added to the mobile phase in concentrations that will eliminate the interaction between injected MPA and the column material, as described previously. Preparative oxidation of AA to DHAA during extraction may be accomplished by addition of iodine or heavy metal ions at a low pH or by addition of ascorbate oxidase at neutral pH. Dehydroascorbic acid is much less stable than AA, and therefore the derivatization should be carried out directly after the oxidation. The stability of the quinoxaline derivatives, however, is also limited. A stability of 8–12 h is described for the quinoxaline

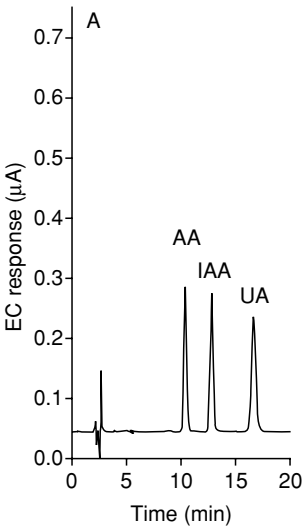

Figure 3 Chromatographic separation of AA and IAA on a C18 column with electrochemical detection. From Lykkesfeldt J (2000) Determination of ascorbic acid and dehydroascorbic acid in biological samples by high-performance liquid chromatography using subtraction method: Reliable reduction with Tris[2-carboxyethyl]phosphine hydrochloride. *Analytical Biochemistry* 282: 89–93 with permission.

derivatives in sample extracts. Further, the preparation of the quinoxaline derivative during extraction requires additional labor compared with the postcolumn derivatization.

In spite of the high selectivity, lack of interaction with MPA, and high sensitivity and specificity, the method is not recommended for purposes involving assay of many samples because of the low stability of the quinoxaline derivatives in the extracts (see Figure 4).

Simultaneous determination of ascorbic acid and dehydroascorbic acid In general, simultaneous determination of AA and DHAA requires more equipment than the other two principles, but it may require less instrumentation. Under the right conditions, oxidation of AA to DHAA is an instantaneous reaction in contrast to reduction of DHAA to AA, which, even under optimal conditions, is a relative slow reaction. In order to determine AA and DHAA simultaneously, it is therefore necessary to perform online, postcolumn oxidation of AA to DHAA and subsequently measure both compounds as DHAA. The oxidation may be performed by addition of a solution containing Cu(II), Hg(II), bromine, or iodine to the effluent after the HPLC column. Alternatively, the effluent can be led through an electrochemical cell or through a solid-state oxidation column packed with active charcoal. In a second postcolumn step, DHAA is mixed with a solution of OPD to form the fluorophore derivative, DFQ (see Figure 2). Another possibility is to measure AA directly and only use the second postcolumn step to measure the original DHAA (see Figures 5 and 6).

In contrast to the subtraction method, the simultaneous determination procedure provides data on AA as well as DHAA from a sample in the same analytical run. The procedure needs no reduction, oxidation, or derivatization prior to the HPLC analysis. However, DHAA spontaneously hydrolyzes to 2,3-diketogulonic acid at pHs as low as 3.2–3.5, and it may be difficult to preserve DHAA in the extracts for 24 h. Furthermore, the concentration of DHAA in most foodstuffs and physiological samples is low compared with AA, and even small degree of oxidation of AA during extraction or in autosampler vials will result in a pronounced increase in the DHAA concentration. Consequently, this analytical principle requires careful handling of samples during extraction and storage.

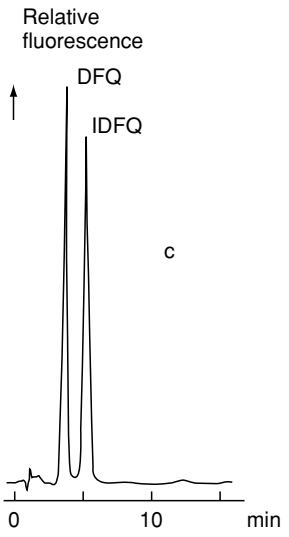

Figure 4 Chromatographic separation of DFQ and IDFQ on a C18 column with fluorometric detection. From Speek AJ, Schrijver J, and Schreurs WHP (1984) Fluorometric determination of total vitamin C and total isovitamin C in foodstuffs and beverages by high-performance liquid chromatography with precolumn deravitization. *Journal of Agricultural and Food Chemistry* 32: 352–355 with permission.

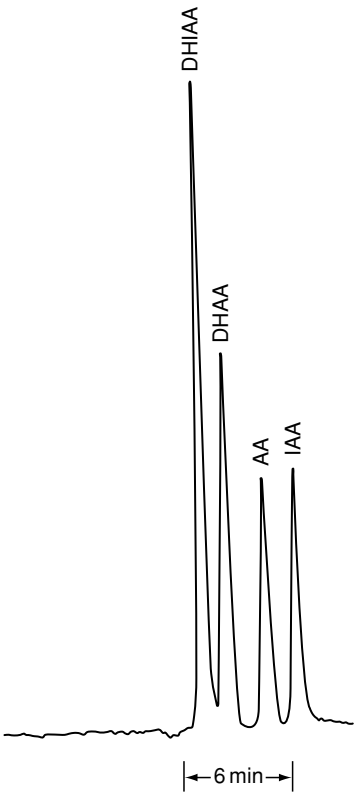

Figure 5 Simultaneous determination of AA, IAA, DHAA, and DHIAA on a PLRP-S column coupled to a postcolumn oxidation and postcolumn derivatization with fluorometric detection. From Vanderslice JT and Higgs DJ (1991) Vitamin C content of foods: sample variability. *American Journal of Clinical Nutrition* 51: 1323S–1327S with permission.

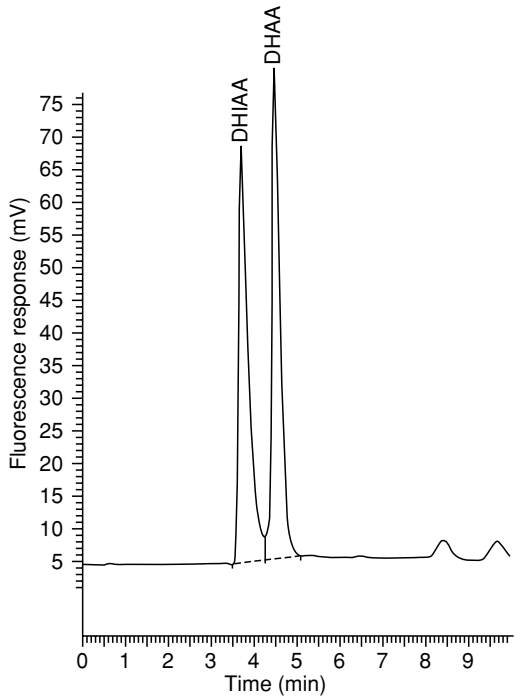

Figure 6 Separation of DHAA and DHIAA on a C18 column followed by postcolumn derivatization and fluorometric detection. From Kall MA and Andersen C (1999) Improved method for simultaneous determination of ascorbic acid and dehydroascorbic acid, isoascorbic acid and dehydroisoascorbic acid in food and biological samples. *Journal of Chromatography* B 730: 101–111, with permission.

Ascorbic acid has a number of chemical properties responsible for the biological reactions that characterize vitamin C. Usually, these characteristics are positively associated with the benefits provided by this vitamin. Paradoxically, these properties provide difficulties in the analysis of ascorbic acid in foods and physiological samples.

See also: **Antioxidants**: Natural Antioxidants; **Browning**: Nonenzymatic; **Cancer**: Diet in Cancer Prevention; **Chromatography**: High-performance Liquid Chromatography; **Controlled-atmosphere Storage**: Effects on Fruit and Vegetables; **Food Fortification**; **Oxalates**; **Oxidation of Food Components**; **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability; **Tocopherols**: Properties and Determination

Further Reading

Behrens WA and Madère R (1994) Ascorbic acid, isoascorbic acid, dehydroascorbic acid and dehydroisoascorbic acid in selected food products. *Journal of Food Composition and Analysis* 7: 158–170.

Budavari S (1996) *The Merck Index*, 12th edn. Whitehouse Station, NJ: Merck.

Hodiroglou N, Madère R and Behrens W (1998) Electrochemical determination of ascorbic acid and isoascorbic acid in ground meat and in processed foods by high pressure liquid chromatography. *Journal of Food Composition and Analysis* 11: 89–96.

Kall MA and Andersen C (1999) Improved method for simultaneous determination of ascorbic acid and dehydroascorbic acid, isoascorbic acid and dehydroisoascorbic acid in food and biological samples. *Journal of Chromatography B* 730: 101–111.

Lykkesfeldt J (2000) Determination of ascorbic acid and dehydroascorbic acid in biological samples by high-performance liquid chromatography using subtraction method: Reliable reduction with Tris[2-carboxyethyl]-phosphine hydrochloride. *Analytical Biochemistry* 282: 89–93.

Martell AE and Hancock RD (1996) Metal complexes in aqueous solutions. In: Fackler JP (ed.) *Modern Inorganic Chemistry*. New York: Plenum Press.

Martell AE and Smith RM (1977) *Critical Stability Constants*, vol. 3. New York: Plenum Press.

Packer L and Fuchs J (eds) (1997) *Vitamin C in Health and Disease*. New York: Marcel Dekker.

Saxholt E and Møller A (1996) *The Composition of Food*, 4th edn. Copenhagen: Danish Veterinary and Food Administration, Gyldendal.

Speek AJ, Schrijver J and Schreurs WHP (1984) Fluorometric determination of total vitamin C and total isovitamin C in foodstuffs and beverages by high-performance liquid chromatography with precolumn derivatization. *Journal of Agricultural and Food Chemistry* 32: 352–355.

Vanderslice JT and Higgs DJ (1991) Vitamin C content of foods: sample variability. *American Journal of Clinical Nutrition* 51: 1323S–1327S.

Vanderslice JT, Higgs DJ, Hayes JM and Block G (1990) Ascorbic acid and dehydroascorbic acid content of food as eaten. *Journal of Food Composition and Analysis* 3: 105–118.

Physiology

G F M Ball, Wembley, UK

Background

Ascorbic acid (vitamin C) has many diverse functions as a reducing agent in the body. It keeps the bound iron or copper ions of several hydroxylating enzymes in the necessary reduced state, and aids the intestinal absorption of iron by forming soluble ferrous–ascorbate complexes. Vitamin C helps in the fight against cancer by reducing harmful free radicals to harmless nonradical species, and inhibiting the formation of

carcinogenic *N*-nitroso compounds in the stomach. There is increasing evidence that vitamin C is involved in the human immune system and in the regulation of prostaglandin synthesis.

Metabolism

Humans and other primates, guinea-pigs, and fruit-eating bats lack the enzyme gulonolactone oxidase, which catalyzes the final step in the biosynthesis of ascorbic acid, and thus rely on their diet to provide the vitamin. Other animal species can synthesize ascorbic acid from glucose and have no need for dietary vitamin C.

Intestinal Absorption

The overall system of intestinal transport and metabolism in humans is shown in Figure 1. Ascorbic acid is ionized within the pH range of the intestinal contents, and therefore, it is the ascorbate anion (specifically L-ascorbate) that is transported. At physiological intakes, ascorbate must move across the brush-border membrane of the intestinal epithelium from a region of lower concentration in the lumen to one of higher concentration in the cytoplasm of the absorptive cell (enterocyte). Such 'uphill' movement requires active transport – a mechanism that depends ultimately upon the expenditure of metabolic energy, i.e., the energy released from the hydrolysis of adenosine triphosphate produced during cellular metabolism. The precise mechanism of ascorbate absorption is secondary active transport in which a transmembrane protein (inappropriately called a carrier) mediates the sodium-coupled transfer of ascorbate across the brush-border membrane. The carrier spans the membrane in a weaving fashion and effects solute transfer through a conformational change in its molecular structure. The immediate energy source for the transport mechanism is the concentration gradient of sodium across the brush-border membrane. The gradient is maintained by the constant extrusion of sodium from the enterocyte by the action of the sodium pump at the basolateral membrane. The sodium pump is driven by metabolic energy and is the primary driving force for ascorbate absorption.

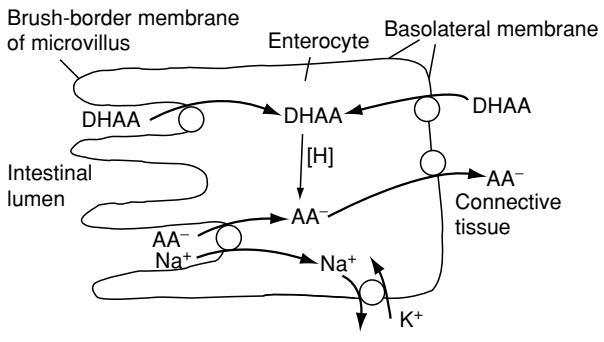

Figure 1 Model of intestinal transport of the L-ascorbate anion (AA^-) and uncharged dehydro-L-ascorbic acid (DHAA) in vitamin C-dependent animals. Thick arrowed lines indicate directional pathways; [H] signifies enzymatic reduction.

An electrical potential difference ranging from −30 to −90 mV (cell interior negative) typically exists across cell membranes. This membrane potential influences the active transport of charged solutes across membranes. Using isolated brush-border membranes from the intestines of guinea-pigs, it has been shown that ascorbate transport is unaffected by changes in the membrane potential. The transport system is therefore electroneutral, indicating a 1:1 cotransport of Na^+ and $ascorbate^-$ by the same carrier. L-Ascorbic acid leaves the enterocyte by sodium-independent facilitated diffusion at the basolateral membrane.

Dehydro-L-ascorbic acid is taken up from the bloodstream across the basolateral membrane of enterocytes as well as being absorbed across the brush-border membrane. The transport mechanism in both cases is sodium-independent facilitated diffusion driven by the steep concentration gradient maintained by the intracellular reduction of dehydroascorbic acid.

The efficiency of vitamin C absorption of foods over a range of usual intakes (20–120 mg per day) is approximately 90%. Absorption by the sodium-coupled secondary active transport mechanism reaches its maximum rate at a relatively low luminal concentration. The higher concentrations resulting from ingestion of vitamin C supplements are absorbed additionally by passive diffusion, which proceeds at a very low rate. Absorption therefore becomes progressively less efficient with increasing dose levels, the efficiency falling from 50% of a single 1.5-g dose to 16% of a single 12-g dose. Absorption efficiency is increased by the ingestion of several spaced doses throughout the day, rather than the ingestion of a single megadose. Ingesting the vitamin in a sustained release form also improves absorption efficiency. Ascorbic acid bioavailability from fruits and vegetables is not significantly different to that from synthetic ascorbic acid, indicating that the bioavailability in these natural food sources is high.

Intestinal absorption of ascorbate is adaptively regulated in a transient and reversible manner by the level of dietary ascorbate. The mechanism of regulation is an increase or decrease in the number of carriers at both brush-border and basolateral membranes in response to low or high concentrations of ascorbate in the blood. The rationale for adaptive

regulation is that carriers are most needed at low dietary ascorbate levels; at excessive levels, the required amount of ascorbate can be absorbed by fewer carriers, aided by passive diffusion. As ascorbate does not provide metabolizable energy, there is nothing to gain from the cost of synthesizing and maintaining carriers when the vitamin supply is in excess.

Postabsorptive Metabolic Events

Vitamin C circulates in the bloodstream mainly as free (nonprotein-bound) ascorbic acid. The kidney actively reabsorbs ascorbate present in the glomerular filtrate, thereby maximizing vitamin C conservation in the body and helping the intestine to maintain the circulating vitamin in its useful, reduced state. Contrary to the intestinal transport system, the renal system is present also in species for which ascorbate is not a vitamin (e.g., rat, rabbit). Renal uptake of the L-ascorbate anion at the brush-border membrane of the absorptive cell of the proximal convoluted tubule is a sodium-coupled, secondary active transport system. Unlike the corresponding intestinal transport system, the renal system is affected by experimental changes in the membrane potential, indicating an electrogenic mechanism with a Na^+/ascorbate$^-$ coupling ratio of 2:1. As the loaded carrier bears a net positive charge, its transport is accelerated by the negative membrane potential. Rapid renal reabsorption of ascorbate is important, as the transit time in the proximal tubule is only about 10 s.

Many cells and tissues can accumulate ascorbate against a concentration gradient. Particularly high concentrations of the vitamin are found in leukocytes (white blood cells), lung tissue, the adrenal and pituitary glands, and compartments of the eye. Skeletal muscle contains much of the body's pool of ascorbate, although the concentration is relatively low. The ultimate metabolic fate of vitamin C is urinary excretion of ascorbic acid or metabolites.

Effect upon Absorption of Inorganic Iron

The iron present in natural foodstuffs exists in two forms, heme iron and inorganic iron. Heme iron is obtained from the hemoglobin and myoglobin present in meat, fish, and poultry. Inorganic iron is found primarily in plant foods; it provides 85–90% of iron in a mixed diet and is the only source of iron in a vegetarian diet. The mechanisms of absorption of heme iron and inorganic iron differ. Unlike heme iron absorption, absorption of inorganic iron is affected by the presence of vitamin C and other constituents of the diet.

Dietary inorganic iron is largely in the ferric (Fe^{3+}) state, which is poorly soluble in the alkaline environment of the small intestine where absorption of iron takes place. Ferrous iron (Fe^{2+}), however, is soluble at alkaline pH. The amount of inorganic iron absorbed by the intestine is influenced by stimulatory and inhibitory factors of either dietary or endogenous origin. Ascorbic acid both reduces and chelates iron, forming ferrous–ascorbate complexes that are readily absorbed. Other weak organic acids (e.g., citric acid), certain sugars, and sulfur-containing amino acids have similar properties, but are less effective. Meat, fish, and poultry also enhance the absorption of inorganic iron, apparently because of the binding of iron by cysteine-containing peptides that result from the digestion of muscle tissues. Phytates that occur in cereals (especially in the bran fraction) bind inorganic iron and retard its absorption. The same is true for tannins (polyphenols) that are found in some vegetables and in tea.

Supplemental ascorbic acid consistently enhances absorption of inorganic iron from single meals, as indicated by incorporation of iron into red blood cells. In one study, 500 mg of ascorbic acid taken with the meal increased iron absorption about sixfold. The most pronounced effects of ascorbic acid are found in meals containing a high content of phytates and/or tannins. This finding has led to the suggestion that ascorbic acid enhances iron absorption by counteracting the influence of inhibitory substances.

Biochemical Functions

Ascorbic acid acts as a cofactor for eight mammalian enzymes (Table 1) by keeping enzyme-bound iron or copper ions in the necessary reduced state. Other reducing agents are far less effective as cofactors

Table 1 Ascorbic acid-dependent mammalian enzymes

Enzyme	*Function*
Prolyl-4-hydroxylase (EC 1.14.11.2)	Collagen synthesis
Prolyl-3-hydroxylase (EC 1.14.11.7)	Collagen synthesis
Lysyl hydroxylase (EC 1.14.11.4)	Collagen synthesis
ε-*N*-Trimethyllysine hydroxylase (EC 1.14.11.8)	Carnitine synthesis
γ-Butyrobetaine hydroxylase (EC 1.14.11.1)	Carnitine synthesis
Dopamine β-hydroxylase (EC 1.14.17.1)	Noradrenaline synthesis
Peptidylglycine α-amidating monooxygenase (EC 1.14.17.3)	Peptide amidation
4-Hydroxyphenylpyruvate hydroxylase (EC 1.13.11.27)	Tyrosine metabolism

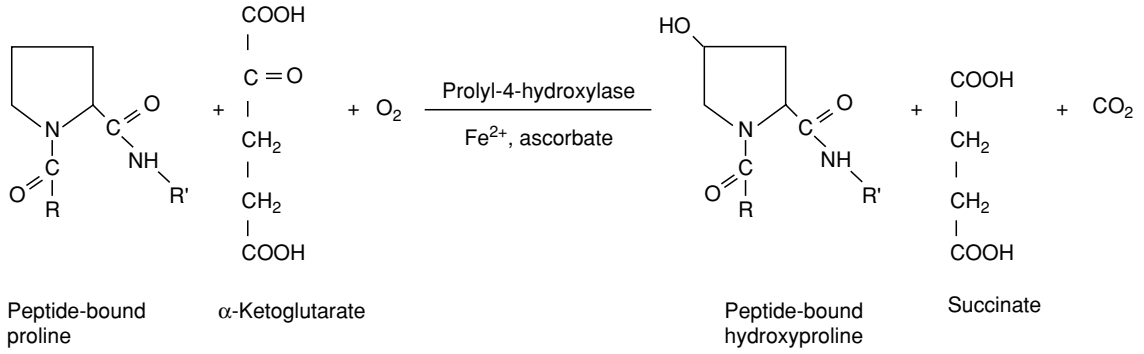

Figure 2 Hydroxylation of proline.

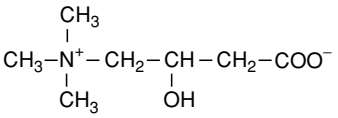

Figure 3 L-Carnitine.

in vitro, and so these enzymes are considered to be ascorbic acid-dependent. The roles of such enzymes in four biochemical processes are discussed below.

Biosynthesis of Collagen

Collagen is the major macromolecule of most connective tissues. It forms a triple helix and is highly cross-linked to give a rigid and inextensible structure. The amino acid composition is unusual among animal proteins, with an abundance of proline and 4-hydroxyproline and a few residues of 3-hydroxyproline and hydroxylysine. The hydroxyproline residues are necessary for proper structural conformation and stability; hydroxylysine residues take part in cross-linking and facilitate subsequent glycosylation and phosphorylation.

Free hydroxyproline and hydroxylysine are not incorporated into procollagen during its synthesis. The presence of these two unusual amino acids in procollagen (the precursor of collagen that contains additional 'propeptides') arises through the post-translational hydroxylation of particular proline and lysine residues in the polypeptide chain. Within the cisternae of the rough endoplasmic reticulum, the newly synthesized procollagen chains encounter three hydroxylating enzymes. Two of these enzymes, prolyl-4-hydroxylase and prolyl-3-hydroxylase, convert proline residues to 4-hydroxyproline or 3-hydroxyproline, respectively, and the third, lysyl hydroxylase, converts lysine residues to hydroxylysine. Each of these enzymes contains an iron ion (maintained in the ferrous state by ascorbate) and requires molecular oxygen and α-ketoglutarate as cosubstrates (Figure 2). In addition to the posttranslational modifications of procollagen, ascorbate stimulates transcription of the type I procollagen gene. The absence of wound healing is one of the features of scurvy that can be attributed to impaired collagen synthesis arising from lack of vitamin C.

Biosynthesis of Carnitine

Carnitine (3-hydroxy-4-*N*-trimethylaminobutyric acid, Figure 3) is essential for the β-oxidation of long-chain fatty acids, producing energy in the mitochondria. Neither free fatty acids nor fatty acid coenzyme As can penetrate the inner membranes of mitochondria, but acylcarnitine can readily do so. The translocation of fatty acids from the cytosol to the β-oxidation site in the matrix of the mitochondrion is therefore dependent upon carnitine.

The synthesis of carnitine commences with the conversion of specific peptide-linked lysine residues to trimethyllysine within certain proteins. Proteins in which trimethyllysine has been found include histones, myosin, calmodulin, and cytochrome c. Free trimethyllysine released by proteolytic cleavage is converted to carnitine in a number of steps, which include two hydroxylations. The enzymes responsible for these hydroxylations are ε-*N*-trimethyllysine hydroxylase and γ-butyrobetaine hydroxylase. Like the hydroxylases involved in collagen biosynthesis, both enzymes contain ferrous ion and require molecular oxygen and α-ketoglutarate as cosubstrates.

Vitamin C deficiency results in a variable decrease in carnitine levels of skeletal muscle, heart muscle, liver, and kidney. Carnitine status depends on dietary carnitine intake, a modest rate of carnitine biosynthesis, and efficient conservation of carnitine by renal reabsorption. The rate of carnitine synthesis in scorbutic guinea-pigs is much higher than the rate

in normal guinea-pigs, and therefore impaired synthesis is not responsible for carnitine depletion. The responsible factor is, in fact, increased urinary excretion resulting from inefficient renal reabsorption.

Biosynthesis of Noradrenaline

As a neurotransmitter, noradrenaline (norepinephrine) is synthesized in the nerve terminals of postganglionic sympathetic neurones, referred to as adrenergic neurones. As a hormone, noradrenaline, along with adrenaline, is synthesized in the chromaffin cells of the adrenal medulla. The final and probably rate-limiting step in the biosynthetic pathway – the conversion of dopamine to noradrenaline (Figure 4) – is catalyzed by dopamine β-hydroxylase in the presence of molecular oxygen. Dopamine β-hydroxylase is a tetramer, each monomer containing two copper ions maintained in the cuprous (Cu^{+}) state by ascorbate. The enzyme is localized within storage vesicles in both adrenergic neurones and adrenomedullary chromaffin cells. Intravesicular ascorbate is maintained in the reduced state by electron transport across the vesicle membrane, cytosolic ascorbate being the most likely electron donor. Cytosolic semidehydroascorbate generated in the extra- to intravesicular electron transfer reaction may be reduced by the outer mitochondrial membrane enzyme, semidehydroascorbate reductase (EC 1.6.5.4), to complete the ascorbate regeneration cycle. In this manner, a stable rate of noradrenaline biosynthesis can be maintained in the absence of exogenous ascorbate.

Amidation of Peptides to Hormones and Hormone-releasing Factors

A large number of peptides act as hormones and hormone-releasing factors. The vast majority of these peptides are initially synthesized as larger, inactive precursor molecules, which are converted to their active forms through a series of posttranslational modifications. One such final step of activation is carboxy-terminal α-amidation, which is essential for biological activity. Examples of α-amidated peptides include α- and γ-melanotropins, calcitonin, pro-ACTH, vasopressin, oxytocin, cholecystokinin, gastrin, gastrin-releasing peptide, and releasing factors for growth hormone, corticotropin, and thyotropin. Direct precursors to α-amidated peptides have glycine as the carboxy terminal amino acid. The enzyme responsible for the amidation is the bifunctional peptidylglycine α-amidating monooxygenase (PAM), which is found in secretory granules in many neuroendocrine tissues. PAM is actually a precursor protein that undergoes endoproteolytic cleavage to generate two enzymes, peptidylglycine α-hydroxylating monooxygenase (PHM) and peptidyl-α-hydroxyglycine α-amidating lyase (PAL). The two-step nature of the amidation is shown in Figure 5. Step 1 requires molecular oxygen, cuprous ion, and ascorbate as the reductant.

Antioxidant Role

Ascorbic acid helps to fight diseases such as cancer by scavenging harmful free radicals. A free radical can be

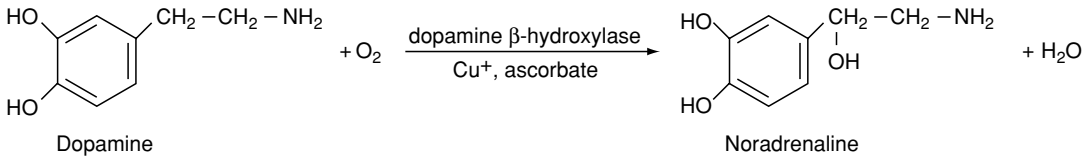

Figure 4 Hydroxylation of dopamine.

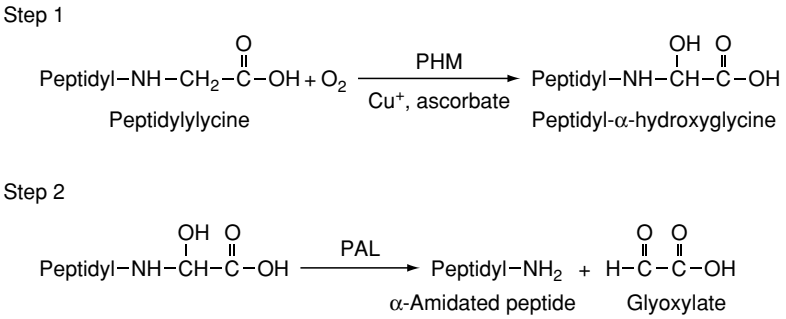

Figure 5 Amidation of peptidylglycine.

defined as an atom or molecule capable of an independent existence that contains an unpaired electron in its outer orbit. A superscript dot is used to denote free radical species; the superoxide anion radical with its two unpaired electrons is usually represented by $O_2^{\cdot-}$. Owing to their unstable electronic configuration, free radicals are much more reactive than nonradicals. They readily extract electrons from other molecules with which they collide, and these molecules in turn become free radicals. Thus, a chain reaction is propagated.

Free radicals can be generated anywhere within the cell by a variety of processes: spontaneous autoxidation reactions (e.g., flavin oxidation), phagocytosis during infection, metabolic processing of foreign compounds (e.g., constituents of tobacco smoke, drugs, pesticides, solvents, and pollutants), ultraviolet irradiation of the skin, and ionizing radiation. Among free radicals of biological importance are the hydroxyl ($HO^{\cdot}$), peroxyl ($ROO^{\cdot}$), superoxide ($O_2^{\cdot-}$), and nitric oxide ($NO^{\cdot}$) radicals.

Many, or perhaps most, free radicals that are generated in the body are metabolized to nonradicals by enzymes such as superoxide dismutase (EC 1.15.1.1), which catalyzes the reduction of the superoxide radical to hydrogen peroxide and water (1).

$$O_2^{\cdot-} + O_2^{\cdot-} + 2H^+ \rightarrow H_2O_2 + O_2. \qquad (1)$$

Although hydrogen peroxide is relatively unreactive, it can traverse biological membranes and cause intracellular damage through the production of hydroxyl radicals catalyzed by ferrous or cuprous ions (2).

$$Fe^{2+}(\text{or } Cu^+) + H_2O_2 \rightarrow Fe^{3+}(\text{or } Cu^{2+}) + HO^{\cdot} + OH^-. \qquad (2)$$

Two enzymes remove hydrogen peroxide in mammalian cells: catalase (EC 1.11.1.6) (eqn (3)) and glutathione peroxidase (EC 1.11.1.9) (eqn (4)), which uses hydrogen peroxide to oxidize reduced glutathione (GSH) to oxidized glutathione (GSSG).

$$2H_2O_2 \rightarrow 2H_2O + O_2 \qquad (3)$$

$$H_2O_2 + 2GSH \rightarrow GSSG + 2H_2O. \qquad (4)$$

Enzymes do not completely prevent the formation of free radicals, and various secondary antioxidants fulfill the role of free-radical scavengers. Ascorbate is one such antioxidant within the aqueous environment of the cytosol and extracellular fluids, and at the water–lipid interface of cellular and subcellular membranes. Figure 6 shows a possible scheme in which ascorbate can be recycled after trapping a free radical. Ascorbate (AH^-) donates an electron to the free radical ($R^{\cdot}$), forming detoxified product (R), and

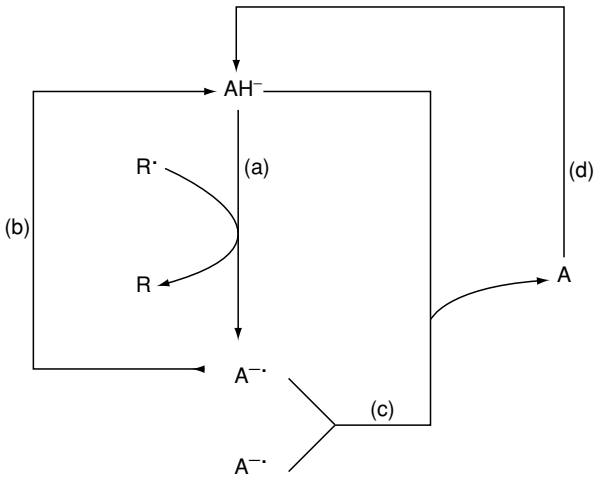

Figure 6 Recycling of ascorbate during the process of free radical scavenging. AH^-, ascorbate; $A^{\cdot-}$, ascorbyl free radical (semidehydroascorbate); A, dehydroascorbic acid; $R^{\cdot}$, free radical species; R, detoxified species. (a) Single-electron transfer; (b) semidehydroascorbate reductase; (c) nonenzymatic dismutation of ascorbyl free radical; (d) dehydroascorbic acid reductase.

is itself oxidized to the ascorbyl free radical ($A^{\cdot-}$), also known as semidehydroascorbate. Pairs of ascorbyl free radicals disproportionate to form one molecule each of dehydroascorbic acid (A) and ascorbate. The high disproportion rate constant of the ascorbyl free radical prevents substantial interaction with other substances. Dehydroascorbic acid can be reduced to ascorbate by either a glutathione-dependent or an NADPH-dependent reductase.

In the lipid environment of cellular and subcellular membranes and plasma lipoproteins, vitamin E (tocopherol, TOH) protects the lipids from peroxidation by reacting with peroxyl radicals, forming the tocopheroxyl radical ($TO^{\cdot}$) and lipid hydroperoxide (ROOH), thereby terminating the chain reaction (eqn (5)). The potentially harmful lipid hydroperoxide can be converted enzymatically to a fatty acid alcohol. Like the ascorbl radical, the tocopheroxyl radical has a low reactivity and can be converted back to tocopherol by enzyme systems.

$$ROO^{\cdot} + TOH \rightarrow ROOH + TO^{\cdot}. \qquad (5)$$

In model *in-vitro* systems, ascorbate regenerates vitamin E by reducing the tocopheroxyl radical back to tocopherol, producing the ascorbyl radical (eqn (6). However, any vitamin E sparing effect exerted by vitamin C *in vivo* is negligible in comparison with other metabolic processes.

$$TO^{\cdot} + AH^- \rightarrow TOH + A^{\cdot-}. \qquad (6)$$

Ascorbate can exert pro-oxidant properties *in vitro* through interaction with iron or copper ions and

hydrogen peroxide to produce hydroxyl radicals. In the healthy human body, these metal ions are largely sequestered in forms unable to catalyze free radical reactions, and so the pro-oxidant properties of ascorbate would not normally be expected to be biologically significant.

Inhibition of *N*-nitroso Compound Formation

Nitrosating agents, for example nitrous anhydride (N_2O_3), are derived from nitrite (eqns (7)–(9))

$$NO_2^- + H^+ \rightleftharpoons HNO_2 \quad (7)$$

$$HNO_2 + H^+ \rightleftharpoons H_2NO_2^+ \quad (8)$$

$$H_2NO_2^+ + NO_2^- \rightleftharpoons N_2O_3 + H_2O. \quad (9)$$

In the acidic contents of the stomach, nitrosation reactions take place between nitrosating agents and secondary amines or *N*-substituted amides to form *N*-nitrosamines (eqn (10)) or *N*-nitrosamides, collectively called *N*-nitroso compounds.

$$\begin{matrix} R_1 \\ \diagdown \\ \quad NH \\ \diagup \\ R_2 \end{matrix} + N_2O_3 \rightleftharpoons \begin{matrix} R_1 \\ \diagdown \\ \quad N-N=O \\ \diagup \\ R_2 \end{matrix} + HNO_2. \quad (10)$$

N-Nitroso compounds are implicated in cancer of the stomach, esophagus, and nasopharynx. Nitrite is present naturally in some foods and is added as a preservative in cured meats. Nitrosatable amines and amides may be ingested as drugs, food additives, or natural constituents of foods. Ascorbic acid is added to the curing brine in the production of bacon and other cured meats in order to prevent the subsequent formation of *N*-nitroso compounds in the stomach. Presumably, the vitamin acts by reducing nitrous anhydride to nitric oxide (eqn (11)), which is not a nitrosating agent.

$$AH^- + N_2O_3 \rightleftharpoons A + 2NO + H_2O. \quad (11)$$

For ascorbic acid to be effective, it must be present in sufficient quantity, since nitric oxide can be oxidized to nitrogen dioxide, and these two species can combine to form nitrous anhydride (eqn (12)).

$$NO + NO_2 \rightleftharpoons N_2O_3. \quad (12)$$

Immune Function

Vitamin C possesses antihistamine activity and is implicated in a variety of immune functions, including neutrophil chemotaxis, T lymphocyte proliferation, natural killer cell activity, and activation of the complement component C1q. Neutrophils and lymphocytes (types of leukocytes) concentrate vitamin C at levels up to 100 times higher than in plasma, suggesting that the vitamin has a physiological role in these immune-system cells.

Antihistamine Activity

In the initial stages of an immune response, histamine amplifies immunoresponsiveness by increasing capillary permeability and smooth muscle contraction, enhancing the flow of immune factors to the site of injury or infection. Subsequently, histamine suppresses the immune response by inhibiting neutrophil lysosomal enzyme release. Excess histamine has a negative effect, and antihistamine therapy is well known for alleviating conditions such as hay fever and inflammatory skin disorders. Ascorbic acid degrades histamine *in vitro*. In humans, chronic oral administration of vitamin C results in a lowering of blood histamine levels, whereas low blood levels of ascorbate are associated with increased histamine levels.

Neutrophil Chemotaxis

Neutrophils (polymorphonuclear leukocytes) destroy pathogenic organisms by phagocytosis (engulfment) and intracellular enzymatic degradation. If foreign materials are too large to be engulfed, enzymes and reactive oxygen species are released extracellularly. Neutrophil chemotaxis refers to the response of neutrophils to chemical signals produced at sites of infection. Histamine lowers the chemotactic responsiveness of neutrophils by elevating intracellular levels of cyclic adenosine 3′,5′-monophosphate (cyclic AMP). A rise in cyclic AMP is associated with depressed neutrophil motility. Data from a study using healthy human subjects suggest that vitamin C at doses of 2 g per day for 2 weeks may indirectly enhance neutrophil chemotaxis through its antihistamine activity.

Proliferation of T Lymphocytes

Vitamin C regulates proliferation of T lymphocytes, which are responsible for destroying viruses that reside and multiply inside living cells of the body. Specialized types of T lymphocyte exert immunoregulatory and effector functions through the secretion of lymphokines such as interferon. Cytotoxic T lymphocytes kill viruses by attaching to the infected cell through recognition of viral antigens and releasing granules whose contents cause perforation of the host cell membrane. The granules also contain substances that induce apoptosis – genetically programmed cell death by fragmentation of nuclear DNA. Since death by apoptosis does not result in the release of the cell's contents, killing by this mechanism may prevent the spread of infectious virus into other cells.

Natural Killer Cell Activity

Natural killer cells are large nonphagocytic leukocytes that kill tumor cells by cell membrane perforation and apoptosis. The cytotoxic activity of killer cells is spontaneous, unlike that of cytotoxic T lymphocytes, which only develops after an immune response has been initiated. An *in-vivo* effect of ascorbic acid on enhancement of human natural killer cell activity has been reported at a dosage of 60 mg per kilogram of body weight.

Regulation of the Complement Component C1q

Complement is a collective term for at least 25 plasma and cell membrane proteins that are important in the host's immune system. Following initial activation by antigen–antibody complexes, the various complement components interact, in a highly regulated enzymatic cascade, to generate various reaction products. Some of these products induce localized vasodilation and attract phagocytic cells chemotactically. Other products coat the foreign particle, enabling it to be recognized by phagocytic cells that bear receptors for these complement products. Finally, the terminal components of the complement system generate the membrane–attack complex, which lyses bacteria and membrane-enveloped viruses.

The first complement component to be activated in the classical cascade pathway is the C1 complex, in which C1q is one of three proteins. When guinea-pigs were fed tissue-saturating amounts of vitamin C, C1q concentrations were significantly higher than in those animals fed only enough ascorbate for adequate growth and for the prevention of scurvy. C1q is a hydroxyproline-containing protein, and so the observed effect is consistent with the known role of vitamin C in hydroxyproline synthesis.

Prostaglandin Synthesis

The prostaglandins (PG) are a group of hormone-like lipids formed in the body from derivatives of essential fatty acids, particularly arachidonic acid (20:4*n*-6). Prostaglandins have a vast range of effects and modulate cardiovascular, pulmonary, immune, and reproductive functions. Vitamin C stimulates the formation of PGE_1 from dihomo-γ-linolenic acid in human platelets. This effect occurs over the physiological range of vitamin C concentrations and is desirable because PGE_1 is an inhibitor of platelet aggregation. PGE_1 is also required for T lymphocyte formation, regulation of collagen and cholesterol metabolism, and regulation of responsiveness to insulin. There are sufficient data to suggest that some of vitamin C's biological actions could be explained by a regulatory effect on prostaglandin synthesis.

Dietary Intake and Supplementation

Recommended Dietary Allowance (RDA)

In the USA, the 1989 RDA for vitamin C is 60 mg per day for adults. At this level, saturation of tissue binding and maximal rates of metabolic and renal tubular absorption seem to be approached. The RDA of 100 mg per day for smokers is higher because the plasma concentration of vitamin C is lowered by the use of cigarettes. Plasma ascorbate levels of women decrease during pregnancy, and the RDA includes a 10 mg per day increment for pregnant women. A daily increment of 35 mg is recommended during the first 6 months of lactation and 30 mg thereafter. Assessments of human vitamin C status are based upon measurement of ascorbate concentrations in plasma and leukocytes.

Deficiency

Scurvy, the classical vitamin C deficiency disease, can occur under circumstances of poor diet, such as arises in chronic alcoholism.

Toxicity

Ascorbic acid is generally regarded as being nontoxic at high intakes. Once the plasma concentration of ascorbate reaches the renal threshold, it is excreted more or less quantitatively with increasing intake. Excessive daily intakes of vitamin C can cause an increased production of oxalic acid in some individuals, leading to an increased risk of kidney stone formation.

Rebound Scurvy

Theoretically, the absorption of ascorbic acid could be impaired on resumption of normal vitamin C inputs following megadosing (> 1 g per day), because of insufficient carriers in the enterocyte cell membranes. Based on experiments with guinea-pigs, it is considered likely that, in humans, renewed synthesis of carriers will take place well before the onset of scurvy. During megadosing, reduced ascorbate absorption is accompanied by increased rates of ascorbate catabolism. In adult guinea-pigs, the accelerated catabolism is not reversible after more than 2 months on subnormal uptake of ascorbate. Guinea-pigs are thus susceptible to a systemic conditioning effect known as rebound scurvy, caused by an induction of ascorbic acid-metabolizing enzymes by high dietary vitamin C. The body stores of vitamin C are depleted more rapidly in juvenile guinea-pigs than in adults,

increasing the likelihood of rebound scurvy in juveniles. In human infants, rebound scurvy could possibly occur, despite an adequate daily intake of vitamin C, if the mother has received large supplements of the vitamin during pregnancy. Reported cases of rebound scurvy in adult humans are rare.

Treatment of the Common Cold

Vitamin C supplementation has been suggested for the prevention of the common cold and the alleviation of its symptoms. Reviews of numerous studies generally conclude that vitamin C megadoses have no consistent effect on reducing the incidence of colds in people in general. However, in four studies with British male schoolchildren and students, a statistically high significant reduction in common cold incidence was found in vitamin C-supplemented groups. This could be interpreted as a correction of a marginal deficiency in the study subjects, rather than an effect of the high dosage. Placebo-controlled studies have consistently shown that high doses of vitamin C alleviate common cold symptoms, although the validity of some of these studies has been questioned.

See also: **Adaptation – Nutritional Aspects**; **Antioxidants**: Natural Antioxidants; **Bioavailability of Nutrients**; **Cancer**: Diet in Cancer Prevention; **Dietary Reference Values**; **Immunology of Food**; **Nitrosamines**; **Prostaglandins and Leukotrienes**; **Scurvy**

Further Reading

Hallberg L, Brune M and Rossander L (1986) Effect of ascorbic acid on iron absorption from different types of meals. *Human Nutrition: Applied Nutrition* 40A: 97–113.

Halliwell B (1996) Vitamin C: antioxidant or pro-oxidant *in vivo? Free Radical Research* 25: 439–454.

Harris JR (ed.) (1996) *Subcellular Biochemistry. Vol. 25. Ascorbic Acid: Biochemistry and Biomedical Cell Biology.* New York: Plenum Press.

Hemilä H (1997) Vitamin C intake and susceptibility to the common cold. *British Journal of Nutrition* 77: 59–72.

Institute of Medicine (2000) *Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids*. Washington, DC: National Academy Press.

Jacob RA (1999) Vitamin C. In: Shils ME, Olson JA, Shike M and Ross AC (eds) *Modern Nutrition in Health and Disease*, 9th edn. Philadelphia, PA: Lippincott Williams & Wilkins.

Johnston CS, Martin LJ and Cai X (1992) Antihistamine effect of supplemental ascorbic acid and neutrophil chemotaxis. *Journal of the American College of Nutrition* 11: 172–176.

Karasov WH, Darken BW and Bottum MC (1991) Dietary regulation of intestinal ascorbate uptake in guinea pigs. *American Journal of Physiology* 260: G108–118.

Packer L and Fuchs J (eds) (1997) *Vitamin C in Health and Disease*. New York: Marcel Dekker.

Siegel BV (1993) Vitamin C and the immune response in health and disease. In: Klurfeld DM (ed.) *Nutrition and Immunology Series. Human Nutrition: A Comprehensive Treatise*. New York: Plenum Press.

ASPARTAME

M B A Glória, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

Background

Aspartame is an intense nutritive sweetener discovered accidentally in 1965 by the chemist James Schlatter from G.D. Searle and Co. It was introduced in 1981 and has been assigned the INS number of 951. At present, it is available under the brand names of Nutrasweet®, Equal®, and Canderel®. Aspartame is a caloric substance because it is a dipeptide that is completely digested after consumption. However, its intense sweetness allows functionality to be achieved at very low levels, providing very few calories.

Aspartame has been approved for food, beverage, pharmaceutical, and tabletop sweetener use in more than 100 countries. Currently, it is the most widely consumed high-intensity sweetener. It is used in approximately 6000 different products and consumed by hundreds of millions of people in countries around the world.

The synthesis of aspartame can be performed by chemical or chemoenzymatic methods. There are over 70 patents on the manufacturing process. It can be made from pure amino acids and is not extracted from any food, and so it is kosher. It is also parve, which means that it contains neither dairy nor meat products. Although kosher and parve, other ingredients in a product containing aspartame must be checked for their kosher status. Aspartame sold for commercial use must meet all requirements of the Food Chemical Codex.

Physical and Chemical Characteristics

Aspartame is a dipeptide composed of two amino acids, L-aspartic acid and L-phenylalanine methyl ester. Its chemical structure is shown in Figure 1. Chemically, aspartame is *N*-L-α-aspartyl-L-phenylalanine methyl ester or 3-amino-*N*-(α-carboxyphenethyl)succinamic acid *N*-methyl ester. It has a molecular formula of $C_{14}H_{18}O_5N_2$ and a molecular weight of 294.30. It is a white, odorless, crystalline powder, with a melting point of 246–247 °C and $[\alpha]_D^{22} = 2.3°$ (1 mol per liter of HCl).

Aspartame is slightly soluble in water (about 1.0% at 25 °C), sparingly soluble in alcohol, and insoluble in fats and oils. The solubility in water is affected by temperature and pH. It increases as the pH is lowered and the temperature is increased.

Sensory Characteristics

Aspartame is described as having a sweet, clean taste and sweetness profile similar to that of sucrose, without any bitter or metallic aftertaste often associated with other high-intensity sweeteners. At concentrations of 0.10–0.89 g l^{-1}, aspartame has no off-tastes. However, it displays a slow onset of sweetness coupled with a lingering sweet taste.

Aspartame has 100–200 times the sweetness of sucrose, depending on the food system. For example, its sweetness potency in flavored yogurt is 175–220 times that of sucrose. The relative sweetness is also affected by pH and the amount of sucrose or other sugars being replaced.

The perception of aspartame sweetness increases linearly with its concentration. However, increasing the ethanol content causes a reduction in the perceived sweetness.

Aspartame extends and intensifies flavors, particularly enhancing fruit flavors, such as orange, lemon, grapefruit, cherry, and strawberry. This flavor-enhancing property, as observed with chewing gum, produces a longer-lasting sweetness and flavor up to four times longer. Such a characteristic is important in many food applications.

Aspartame exhibits synergism, a superior taste profile, and an improved stability when combined with other sweeteners, such as saccharin, cyclamate, acesulfame-K, and stevioside. The flavor-enhancement quality of aspartame masks bitter flavors, even at subsweetening levels. Its use is thus recommended in blends with sweeteners that possess potentially undesirable or more complex taste profile.

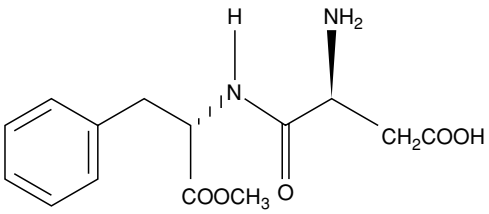

$C_{14}H_{18}N_2O_5$ (FW 294.30)
N-L-aspartyl-L-phenylalanine-1-methyl ester
ASPARTAME

Figure 1 Chemical structure of aspartame.

Food and Beverage Applications

Aspartame provides many opportunities for formulating new products while lowering or limiting calories and sugar consumption. Other benefits of using aspartame are cost reductions in packaging, shipping, and storage. Furthermore, aspartame is generally cheaper than powdered sugar. It has a lower cost per unit sweetness than sucrose.

Aspartame is approved for use as a prepackaged sugar substitute tablet. Its use as a sweetener and flavor enhancer is also approved. It has a potential for a wide range of food applications, including carbonated soft drinks, fruit drinks, dry beverage bases, instant tea, hot chocolate, gelatin, puddings, fillings, yogurt, icecream, frozen novelty, hard candy, confectionery, breath mint, breakfast cereals, chewing gums, dairy products, and toppings.

Aspartame is widely used both as a dry powder and as an aqueous solution. The largest application is in soft drinks followed by dry mix products. The main products in the dry mix market are sweetener tablets and sachets, cold beverage mixes, and instant desserts. In these products, aspartame's sweetness should be released rapidly and evenly throughout the product for maximum performance. In order to insure uniformity of the final product, stability against segregation, and proper reconstitution upon preparation, agglomeration processes are used either as a re-wet or a straight through process (the dry mix is dissolved and spray-dried). Another way to overcome this problem would be by coating food ingredients such as citric, fumaric, and lactic acids, sucrose, dextrose, fructose, oligo-fructose, and maltodextrin with fine-grade aspartame.

Blends or combinations of sweeteners are often used to achieve the desired level of sweetness in food and beverage products that traditionally have been sweetened with single sweeteners. In the USA, almost every major soft-drink manufacturer now uses a combination of aspartame and saccharin in diet drinks. The blending of these two noncarbohydrate sweeteners reduces the cost and helps to prolong the shelf-life of the beverages compared with aspartame alone.

Aspartame hydrolyzes slowly at the low pH values characteristic of soft drinks. Therefore, products containing exclusively aspartame become less sweet on prolonged storage.

When aspartame is substituted for sugar in food products, the sugar bulk must be replaced. Although gums can be used as a thickening agent, modified food starch may be a better choice.

Sucrose can be replaced in nondiet carbonated soft drinks by aspartame–glucose syrup formulations, which mimic sucrose properties in soft drinks. Products containing this blend are liked as much as, or more than, all-sucrose drinks by consumers. They are also significantly cheaper since aspartame has a lower per-unit sweetness than sucrose.

Aspartame may be used as total sugar replacements in whey beverages, resulting in a significant reduction in the amount of carbohydrate present. The amount required to produce a sweetness level equal to that of a beverage containing 10.5 g per 100 ml of invert sugar is 0.25 g l^{-1}. When used together with acesulfame-K, synergism results in a 25% reduction in the total amount of sweetener required.

Stability

In the dry state (e.g., table top sweetener, powdered drinks, dessert mixes), aspartame is quite stable. However, when used in aqueous solutions or intermediate moisture systems, it is susceptible to hydrolysis, other chemical reactions, and microbial degradation, thus limiting its shelf-life.

In aqueous solution, aspartame undergoes decomposition and racemization, forming a variety of products. Degradation is also significantly affected by the pH, temperature, and storage period of the food or beverage. These factors cause a degradation reaction of first-order kinetics.

Aspartame is stable in the pH range of 3–5, which is characteristic of most foods. At 25 °C, the maximum stability is reported at pH 4.3. The good stability observed during refrigerated storage of yogurt is due to its pH, which ranges from 4.0–4.5. A frozen dairy dessert may have a pH ranging from 6.5 to 7.0, but, because of the frozen state, the rate of reaction is dramatically reduced.

The influence of temperature on aspartame depends on the composition and water activity of the food. Maximal stability is observed in the lower temperature range used for refrigerated and frozen storage. The activation energy decreases above and below the isoeletric point, pH 5.4. Aspartame can withstand the heat processing used for dairy products and juices, aseptic processing, and other processes in which high-temperature short-time and ultrahigh temperature conditions are used. However, under conditions of excessive heat, aspartame may hydrolyze or cyclize, limiting its applications. An encapsulated form of aspartame is now available to improve heat stability.

The initial microbial load and subsequent growth of contaminants in yogurt and milk chocolate can affect the rate of aspartame degradation. The stability of aspartame has been observed to be similar for different strains of *Lactobacillus bulgaricus* and *Streptococcus thermophilus*, suggesting that strain differences do not affect stability. To improve stability, aspartame should be added to yogurt after fermentation.

Aspartame can react with aldehydes, which are the principal flavor compounds used in chewing gums and in some soft drinks. However, the reactivity varies with aldehyde type, owing to differences in reaction pathway.

Nonenzymatic browning via the Maillard reaction is another important pathway for aspartame degradation in foods. Owing to its dipeptide structure, aspartame can react with reducing sugars in the presence of water. This reaction is favored at alkaline pH values, because nonprotonated amine groups on the molecule are more available. This interaction may result in off-flavors, loss of sweetening power or undesirable color changes. Another example of carbonyl-amino reactions has been shown for vanillin, resulting in the loss of vanilla flavor. Both temperature and water content during storage can affect reaction rates, which have been observed to follow zero-order kinetics.

Decomposition Products

As indicated in Figure 2, aspartame decomposition [1] results in hydrolysis of the methylester with loss of methanol forming 3,6-dioxo-5-phenylmethyl-2-piperazineacetic acid – diketopiperazine. This compound is then hydrolyzed [2], yielding aspartylphenylalanine, which is eventually hydrolyzed [3] to the free amino acids aspartic acid and phenylalanine. At acid pH values [4], products such as phenylalanine methyl ester and aspartic acid are formed. The first is hydrolyzed [5], forming phenylalanine and methanol. Another degradative pathway involves base-catalyzed hydrolysis with loss of methanol [6] forming α-aspartylphenylalanine, which undergoes further hydrolysis [3] to give phenylalanine and aspartic acid. Racemization of α-aspartame [7] and its degradation products has also been observed.

The major degradation products are diketopiperazine and aspartylphenylalanine. None of the decomposition products have a sweet taste or aftertaste. Therefore, when these compounds are formed in the

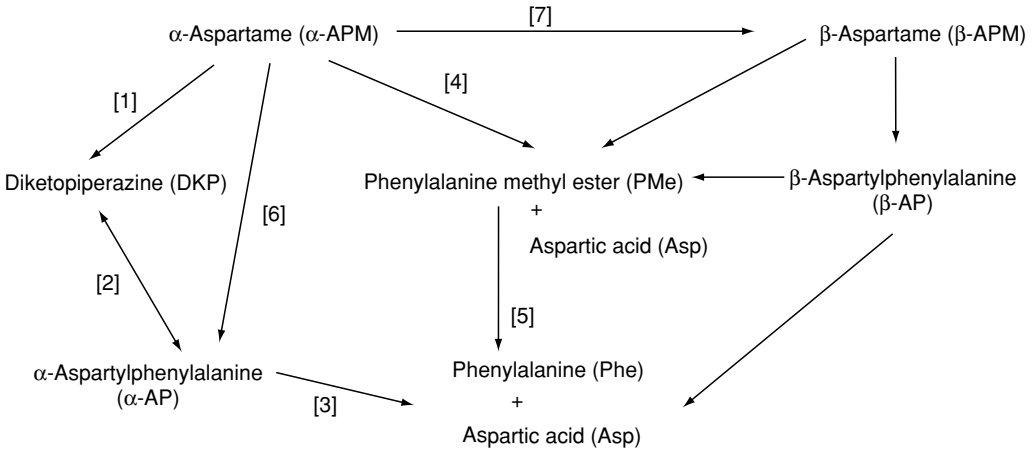

Figure 2 Decomposition products of aspartame.

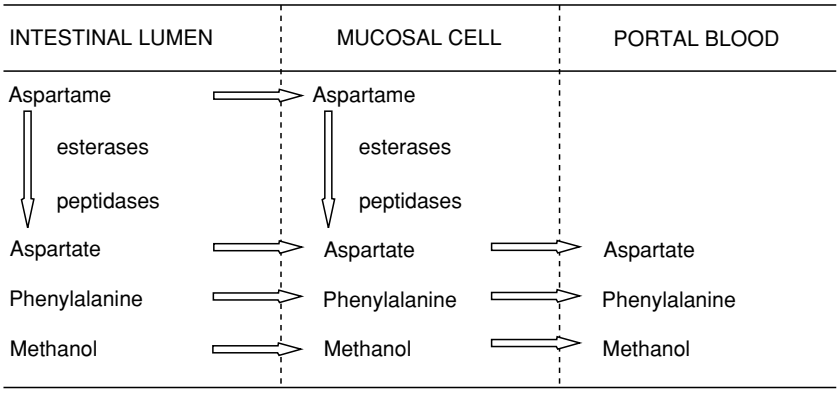

Figure 3 Diagram of aspartame metabolism. From Burini RC and Caballero B (1992) Metabolismo do aspartame em humanos. *Cadernos de Nutrição* 4: 38–47, with permission.

food, a loss of sweetness is perceived. Several investigators have monitored aspartame in different food products in the market and observed that levels varied from 52 to 107% of the value declared on the label.

Metabolism

Metabolism studies in numerous species, including humans, demonstrated that aspartame is rapidly and completely metabolized to its individual constituents phenylalanine, aspartic acid, and methanol. Consequently, it has the same caloric value of proteins – 4 kcal g^{-1}. However, because of its intense sweetness, the amount of aspartame used in foods is small enough for the sweetener to be virtually nonnutritive, providing negligible energy. It is noncariogenic and produces little glycemic response.

Aspartame may be absorbed and metabolized in one of three ways (Figure 3). It may be hydrolyzed in the intestinal lumen by proteolytic and hydrolytic enzymes to form aspartate, phenylalanine, and methanol. These hydrolysis products are absorbed from the intestinal lumen and reach the systemic circulation in a manner similar to that of amino acids, peptides and methanol arising from the digestion of dietary proteins and polysaccharides. Aspartame may be absorbed directly into mucosal cells by one of the peptide transporter systems, with subsequent hydrolysis within the epithelial cells to aspartate, phenylalanine, and methanol. Aspartame may also be de-esterified in the intestinal lumen to yield methanol and aspartylphenylalanine, which is taken up by enterocytes via a peptide transport system and further hydrolyzed to aspartate and phenylalanine. Once absorbed, these components are metabolized, utilized, and/or excreted by the body through similar pathways as when they are derived from any other dietary source.

Safety

Aspartame has been the subject of extensive investigation in animals and humans. Prior to the marketing of aspartame, numerous studies were done to evaluate its metabolism and tolerance in healthy subjects and various subpopulations, including phenylketonuric heterozygotes, children, lactating and pregnant women, and individuals with renal and liver disease. Postmarketing, numerous additional studies were performed, including the evaluation of alleged sensitivity to aspartame. The results of extensive scientific research investigating these allegations did not show any causal relationship between aspartame and adverse effects.

Aspartame has undergone extensive clinical testing prior to its regulatory approval, including acute and long-term studies in the general population and in populations that may be extensive users of a high-intensity sweetener, such as diabetics and obese subjects. In addition, pharmacokinetic and metabolism studies, including radiolabeled studies, were carried out in normal subjects, subjects with a somewhat reduced capacity to metabolize phenylalanine (i.e., heterozygotes for phenylketonuria), and subjects with an almost complete inability to metabolize phenylalanine (i.e., phenylketonurics). Following the approval of aspartame, additional studies were undertaken to further evaluate the safety of aspartame in special populations.

The safety of aspartame has been extensively tested in animal and human studies. Undoubtedly, it is the most thoroughly studied of the high-intensity sweeteners. Its safety has been affirmed by numerous scientific bodies and regulatory agencies, including the Joint Expert Committee on Food Additives (JECFA) of the Codex Alimentarius (Food and Agriculture and World Health Organizations), the Scientific Committee for Food of the Commission of European Communities, the United States Food and Drug Administration, and the regulatory agencies of more than 100 other countries around the world.

On the basis of the results of comprehensive safety studies, a no observed-effect level of more than 2000–4000 mg per kilogram of body weight was established for aspartame. Based on this information, an acceptable daily intake (ADI) of 40 mg per kilogram of body weight per day for aspartame was set by JECFA. On the basis of both animal and human data, the United States Food and Drug Administration set an ADI for aspartame of 50 mg per kilogram of body weight per day. An ADI value of 7.5 mg per kilogram of body weight was given for its metabolite diketopiperazine. The use of aspartame appears to be unanimously authorized in different foods and beverages.

However, products containing aspartame have to be labeled to warn people with phenylketonuria. Methanol released from aspartame does not constitute an excessive metabolic load. Numerous studies have shown that aspartame and diketopiperazine are not associated with any serious adverse health effects.

Estimation of Consumption

Aspartame has been the subject of extensive postmarketing surveillance of consumption. Data from surveys in the USA indicate that approximately 40% of the population is consuming aspartame-containing products. From 1984 to 1992, the 90th percentile 14-day average consumption of aspartame in the general population ranged from 1.6 to 3.0 mg per kilogram per day. This is approximately 5–10% of the US ADI. The consumption by children aged 2–5 years and from all age groups ranged from 2.6 to 5.2 and from 5.2 to 8.5 mg per kilogram per day, respectively. Aspartame consumption by diabetics ranged from 2.1 to 3.4 mg per kilogram per day. Therefore, persons with diabetes who would likely be frequent users of aspartame consume quantities well below the ADI.

In Canada, three different methods to estimate potential aspartame intake have been used. In the first method, it was estimated that aspartame consumption would be 11 mg per kilogram per day if it replaced per-capita disappearance of sugar. In the second method, it was estimated that aspartame consumption would be approximately 4 mg per kilogram per day if it replaced sugar in major types of foods on a per-capita disappearance basis. In the third method, based on a 24-h dietary recall survey of food consumption in over 13 000 individuals, it was estimated that potential aspartame use in these foods would result in an intake ranging from 3.2 to 7.4 mg per kilogram per day. In a survey with over 5000 individuals in 1987, it was observed that 45% of the individuals consumed products with aspartame. Consumption at the 90th percentile for the general population was 5.5 and 5.9 mg per kilogram per day during the winter and summer months, respectively. Intake by diabetics was 5.5 and 11.4 mg per kilogram per day during the winter and summer months, respectively. In addition, the consumption by all age groups of children was below 10 mg per kilogram per day.

In the UK, in a study based on the consumption of saccharin, it was estimated that aspartame consumption by diabetics would be approximately 5.5 mg per kilogram per day at the 90th percentile. In another study, potential aspartame intake in children from soft-drink consumption would be approximately

6.6–6.8 mg per kilogram per day at the 90th percentile.

Nearly 73% of the diabetic children surveyed in Finland consumed aspartame-containing products with a mean intake of 1.15 mg per kilogram per day. Aspartame intake in France from 1991 to 1992 was 0.6 mg per kilogram per day and in Germany, consumption in 1988–1989 was 2.75 mg per kilogram per day at the 90th percentile. The average intake of aspartame among Italian teenagers users of diet products was 0.03 mg per kilogram per day, with a maximum of 0.39 mg per kilogram per day.

In Brazil, data on potential intake of aspartame have been based on questionnaires to users of sweetener in the cities of Campinas, state of São Paulo, and Curitiba, state of Paraná, during the winter of 1990 and summer of 1991. Data showed that 40% of the population consumed aspartame alleging reasons of weight-control diet (36%), diabetes (35%), and weight loss (23%). Table-top sweeteners were the major source, followed by soft drinks. The median daily intake by users represented 2.9% of the ADI. Diabetics had a much higher intake within the studied population.

Although survey methodologies differed among these evaluations, aspartame intake is remarkably consistent across studies and well below the IDA.

Because aspartame is metabolized to common dietary components – aspartate, phenylalanine, and methanol – it is useful to evaluate the impact of its consumption on the normal dietary intake of these components. At the current 90th percentile 14-day average levels of consumption, aspartame contributes approximately 1.5 and 3% of an adult's daily dietary intake of aspartate (80 mg per kilogram per day) and phenylalanine (52 mg per kilogram per day), respectively.

Aspartame yields approximately 10% methanol by weight. This amount is well below the normal dietary exposure to methanol from many fruits, vegetables, and juices. Therefore, normal diets provide several times more methanol than that from foods sweetened with aspartame. For example, tomato juice provides approximately six times more methanol than an equal volume of aspartame-sweetened beverage. Whereas methanol exposure at the 90th percentile of chronic aspartame consumption is 0.3 mg per kilogram per day, the FDA has established acceptable levels of exposure to methanol at 7.1–8.4 mg per kilogram per day for 60-kg adults.

Methods of Analysis

Many analytical procedures have been proposed to check food labeling or to measure the degree of aspartame decomposition during manufacturing and storage of food products. Several types of methods have been described for the analysis of aspartame, including spectrophotometric, enzymatic, titrimetric, capillary isotachophoretic, thin-layer chromatography, and liquid and gas chromatographic methods.

HPLC is by far the most frequently used technique to monitor aspartame, intermediates formed during its synthesis, and decomposition products formed during manufacture and storage of food products. It is also used to check for amounts declared on food labels.

Sample Preparation

Sample preparation generally depends on the type of food matrix. Samples can be directly injected into the HPLC or after minimal pretreatment. However extraction, clean-up, and/or purification might be necessary, depending on the complexity of the matrix, in order to eliminate any interference.

Clarification with Carrez solutions can be used to eliminate any interfering compounds. Purification, isolation, or concentration of the extract can also be performed by solid-phase extraction with C8 or C18. Sep-Pak C18 cartridges can be used for the separation of aspartame from its degradation products. The use of ion-exchange cartridges, such as Bond Elut SCX, Bond Elut C18, and Amberlite CG-120, has also been described.

Separation of Aspartame from Other Sweeteners

During HPLC analysis of aspartame, the most commonly used stationary phase is reverse-phase C18. Two main types of mobile phase have been used: an alcohol (methanol or isopropanol) associated with acetate or phosphate buffer or acetonitrile associated with phosphate buffer. In both cases, the pH is adjusted to 3.0–4.5. Ion-pair chromatography has also been used to separate aspartame from other sweeteners.

Separation of Aspartame from Synthesis Intermediates, Stereoisomers and Degradation Products

The most commonly used stationary phase for the separation of aspartame from synthesis intermediates, stereoisomers and degradation products, is the reverse-phase C18 column. The main type of mobile phase used is a phosphate buffer at a pH ranging from 2.5 to 5.0, associated with acetonitrile. Reverse-phase HPLC with gradient elution of acetonitrile in phosphate buffer has also been used.

Ion-pair chromatography has also been used to separate aspartame from synthesis intermediates and

degradation products. The method using sodium hexanesulfonate and gradient elution to separate aspartame from two decomposition products and seven related synthesis byproducts has been used routinely for the control of aspartame synthesis and as a check for the purity of both finished bulk and mother liquor. Chiral chromatography can also be used to obtain the resolution of stereoisomers from aspartame, its precursors, and its degradation products.

Detection Systems

Aspartame has been quantified by UV detection (254, 200–217 nm). However, because aspartame has a relatively low extinction at 254 nm, quantification at lower wavelengths provides an increased response. Detection can also be performed with increased specificity by fluorescence after postcolumn derivatization with *o*-phthaldehyde. In order to improve the detection sensitivity, aspartame can also be separated by HPLC as a fluorescent derivative. Fluorescamine derivatives are separated on LiChrosorb RP-8 using acetonitrile in 50 mM acetate buffer, pH 6, 22:78, v/v or on Spherisorb S5 ODS-2 RP using 0.2 M phosphate buffer (pH 9):acetonitrile:methanol, 2:1:1, v/v. Detection is performed at 397 nm excitation and 482 nm emission.

See also: **Acesulfame/Acesulphame**; **Browning**: Nonenzymatic; **Chromatography**: High-performance Liquid Chromatography; **Saccharin**; **Sweeteners**: Intensive; Others

Further Reading

Burini RC and Caballero B (1992) Metabolismo do aspartame em humanos. *Cadernos de Nutrição* 4: 38–47.

Butchko HH and Stargel WW (2001) Aspartame: scientific evaluation in the postmarketing period. *Regulatory Toxicology and Pharmacology* 34(3): 221–233.

Duffy VB and Anderson GH (1998) Position of the American Dietetic Association: use of nutritive and nonnutritive sweeteners. *Journal of the American Dietetic Association* 98(5): 580–587.

FDA (1981) Aspartame: commissioner's final decision. *Federal Drug Administration Register* 46: 38285–38308.

FDA (1984) Food additives permitted for direct addition to food for human consumption; aspartame. *Federal Drug Administration Register* 49: 6672–6682.

Glória MBA (2000) Intense sweeteners and synthetic colorants. In: Nollet LML (ed.) *Food Analysis by HPLC*, pp. 523–573. New York: Marcel Dekker.

Grenby TH (1991) Intense sweeteners for the food industry: an overview. *Trends in Food Science and Technology* 2: 2–6.

JECFA (1980) Aspartame. Evaluation of certain food additives. *FAO/WHO Technical Report Series* 653: 20–21.

Sherlock JC (1986) The estimation of intake. In: *International Aspartame Workshop Proceedings. Marbella, Spain*. Washington, DC: International Life Sciences Institute – Nutrition Foundation.

Tschanz C, Butchko HH, Stargel WW and Kotsonis FN (1996) *The Clinical Evaluation of a Food Additive – Assessment of Aspartame*. Boca Raton, FL: CRC Press.

Assays *See* **Immunoassays**: Principles; Radioimmunoassay and Enzyme Immunoassay; **Analysis of Food**

ATHEROSCLEROSIS

A H Lichtenstein, Jean Mayer USDA Human Nutrition Research Center on Aging at Tufts University, Boston, MA, USA

Introduction

Atherosclerosis is the leading cause of death and disability in developed countries and its rates are rising rapidly in developing countries. The disorder is characterized by the thickening of artery walls. There are three major types of atherosclerotic lesions; fatty streaks, fibrous plaques, and complicated lesions. The latter two are raised and because they are on the surface of a blood vessel they result in decreased lumen size and impeded blood flow (ischemia). The blood vessel involved, location of the lesion(s), and extent of ischemia determine the clinical manifestations of the disorder. Primary risk factors for atherosclerosis include elevated low-dens-

ity lipoprotein (LDL) cholesterol levels, male gender, increased age, positive family history, current cigarette smoking, hypertension, diabetes mellitus, and low high-density lipoprotein (HDL) cholesterol levels. This article emphasizes dietary modification as a nonpharmacological approach to treating the major risk factor for atherosclerosis, elevated LDL cholesterol levels. The topics discussed include saturated fat, omega-3 fatty acids, *trans* fatty acids, dietary cholesterol, and soluble fiber.

The term atherosclerosis describes a subgroup of arteriosclerotic disorders characterized by the presence of fatty plaque (atheromas) within the intima and media of medium and large arteries. The word atherosclerosis is derived from a combination of the Greek words *athere*, gruel, and *skleros*, hardening. The World Health Organization defines atherosclerosis as a 'variable combination of changes in the intima of the arteries involving focal accumulations of lipids and complex carbohydrates with blood and its constituents accompanied by fibrous tissue formation, calcification and associated changes in the media.' The American Heart Association defines atherosclerosis as 'a process that leads to a group of diseases characterized by a thickening of artery walls.'

Incidence

In the late 1990s, 15 million deaths were attributed to cardiovascular disease worldwide. This accounted for about 30% of all the deaths reported. In industrialized countries, the rate of cardiovascular disease has been declining for a number of years. However, it still accounts for about 50% of all deaths. In developing countries and eastern Europe, the rate of cardiovascular disease has been increasing. It accounts for about 25% and 60% of all deaths, respectively. As compiled by the American Heart Association from World Health Organization data, the mortality rates for coronary heart disease, stroke, and total cardiovascular disease, relative to total deaths, are shown in Table 1.

Pathology

The development of atherosclerotic lesions is relatively slow and the etiology has yet to be fully elucidated. Evidence suggests that there are multiple factors which independently impinge on lesion progression. Atherosclerotic lesions are frequently subdivided into three categories: fatty streaks, fibrous plaque, and complicated lesion. The first to appear, fatty streaks, are flat or slightly raised accumulations of lipid-rich macrophages and smooth-muscle cells. Fatty streaks are ubiquitous in humans and appear early in life. They rarely impede blood flow and their presence goes largely unnoticed. Although some fatty streaks can progress to fibrous plaque, their distribution suggests this is not always the case.

Fibrous (atherosclerotic) plaque is made up of macrophages which have accumulated lipid and are transformed into what are commonly referred to as foam cells. These lesions are elevated and composed of a core of cholesteryl ester, cellular debris, and cholesterol crystals, and a fibrous cap made up of smooth muscle and foam cells, collagen, and lipid. Due to their raised nature, fibrous plaques frequently obstruct the arterial lumen and impede blood flow. Rupturing of the cap may lead to clot formation and occlusion of the artery.

Fibrous plaque frequently progresses to complicated lesions. Complicated lesions severely or even completely obstruct the arterial lumen, with consequent ischemia or infarction. They are frequently associated with mural thrombosis (frequently leading to complete obstruction), ulceration, hemorrhage into the lesion, and calcification.

There are a number of theories on how and why both fibrous plaques form and progress to complicated lesions. The lipid hypothesis suggests that persistent hypercholesterolemia leads to LDL infiltration of the epithelial cells, smooth-muscle cells, and macrophages. This hypercholesterolemia not only leads to cholesterol accumulation but also activates protein growth factors which stimulate smooth-muscle cell proliferation into the media and subsequent lipid uptake. The response to injury hypothesis suggests that injury results from mechanical factors, chronic hypercholesterolemia, toxins, viruses, or immune reactions and leads to monocyte aggregation and adherence at the site of injury. Injury causes the release of growth factors which may arise from endothelial cells, monocytes, macrophages, platelet, smooth-muscle cells, and T cells. Monocytes and smooth-muscle cells carry 'scavenger' receptors which bind oxidized but not native LDL in a nonsaturable fashion. Uptake of oxidized LDL converts macrophages and smooth-muscle cells into foam cells. The immunological hypothesis suggests that an autoimmune reaction, potentially to oxidized LDL, precipitates a chain of events which results in lipid accumulation and the development of fibrous plaque.

Clinical Events

Atherosclerosis is a disorder that is slow to be detected, primarily because it is silent throughout most of its natural history. Its stealthy presentation is explained by the fact that under normal circumstances

Table 1 Death rates for coronary heart disease, stroke, total cardiovascular disease, and all causes in selected countries

State	Total cardiovascular disease[a]				Coronary heart disease[b]				Stroke[c]			
	2000		1940		2000		1940		2000		1940	
	Rank[d]	Death rate	Rank[d]	Death rate	Rank[d]	Death rate	Rank[d]	Death rate	Rank[d]	Death rate	Rank[d]	Death rate
Alabama	43	414.3	47	202.2	19	157.5	27	80.7	37	68.4	45	31.3
Alaska	7	316.0	4	137.7	4	124.9	5	62.0	41	69.3	25	25.5
Arizona	10	322.9	13	147.4	17	157.1	15	72.2	9	56.7	10	22.8
Arkansas	47	420.8	42	198.1	47	208.1	46	102.6	51	86.1	51	36.7
California	19	347.5	15	153.7	28	176.3	20	76.2	26	63.2	29	26.3
Colorado	4	305.7	2	131.5	6	130.6	4	60.3	1.1	57.1	5	22.1
Connecticut	17	345.4	16	154.3	16	156.8	11	69.8	6	52.5	3	21.4
Delaware	30	375.6	31	174.9	22	162.5	23	78.2	4	52.5	12	23.4
District of Columbia	33	383.9	50	206.4	7	131.3	9	67.2	29	64.2	47	33.5
Florida	15	339.5	17	155.2	35	184.8	30	82.7	5	52.5	8	22.6
Georgia	45	417.6	46	200.6	25	168.0	31	84.1	47	75.8	48	34.0
Hawaii	1	295.7	6	139.2	1	110.7	2	53.1	14	60.0	30	26.3
Idaho	9	322.7	7	141.5	18	157.3	19	75.4	34	65.6	23	25.1
Illinois	34	389.9	34	180.1	42	199.0	38	90.8	33	64.9	35	27.4
Indiana	40	401.6	37	183.7	43	204.3	44	98.2	44	71.1	39	28.7
Iowa	24	356.3	24	157.9	32	180.8	29	82.5	21	62.3	16	24.2
Kansas	21	352.6	20	156.6	24	166.1	21	77.8	25	63.0	19	24.9
Kentucky	46	420.0	43	198.9	45	205.7	48	104.2	39	68.7	38	28.5
Louisiana	44	415.5	45	200.1	30	177.8	39	90.9	40	69.2	44	30.8
Maine	23	354.8	27	160.8	29	177.5	32	85.3	10	56.9	9	22.6
Maryland	27	360.5	29	167.5	15	154.8	12	71.4	20	61.8	28	26.1
Massachusetts	12	324.7	10	144.7	20	159.3	17	73.4	2	50.9	2	20.0
Michigan	39	399.1	36	182.7	40	197.6	37	89.5	31	64.7	34	26.9
Minnesota	5	305.9	3	134.9	10	140.4	8	67.0	32	64.8	20	24.9
Mississippi	52	474.8	52	235.5	34	184.3	42	97.3	46	71.8	49	34.2
Missouri	42	408.9	41	188.5	51	220.4	47	104.1	35	65.7	36	27.4
Montana	8	319.2	9	144.0	5	129.8	7	64.7	22	62.8	21	25.0
Nebraska	25	359.3	25	158.0	14	154.2	15	72.4	16	60.5	17	24.3
Nevada	35	390.9	39	183.8	11	141.9	10	67.6	30	64.3	33	26.8
New Hampshire	28	360.9	19	156.4	33	181.1	26	80.2	24	63.0	13	23.7
New Jersey	29	364.5	28	163.1	41	197.6	34	86.7	3	51.8	7	22.5
New Mexico	6	311.6	5	138.0	2	112.4	1	50.4	8	55.5	6	22.2
New York	41	402.2	35	180.2	52	253.1	52	108.4	1	44.2	1	19.5
North Carolina	37	394.3	40	187.4	37	188.5	41	95.6	49	79.6	46	33.4
North Dakota	16	340.7	14	153.1	26	168.2	25	80.0	27	63.8	24	25.4
Ohio	38	399.1	38	183.7	46	206.0	43	97.6	19	61.5	26	25.7
Oklahoma	49	427.8	44	199.6	50	219.7	51	107.2	43	70.2	41	29.4
Oregon	14	334.3	11	145.7	13	152.0	13	72.0	48	77.5	43	29.7
Pennsylvania	36	391.5	33	179.2	38	190.7	36	88.9	15	60.2	22	25.0
Puerto Rico[e]	3	304.7	22	157.2	9	139.0	18	73.9	12	59.7	40	29.2
Rhode Island	20	351.8	23	157.4	44	205.1	40	91.5	7	53.3	4	21.4
South Carolina	48	423.7	51	208.1	39	192.2	45	99.6	52	88.5	52	39.3
South Dakota	18	345.4	18	155.9	31	178.0	33	86.3	18	61.4	15	23.9
Tennessee	51	433.4	49	204.9	49	215.5	49	104.4	50	81.9	50	34.8
Texas	32	380.4	30	174.4	36	187.1	35	87.9	38	68.5	37	28.2
Utah	2	299.9	1	128.2	3	124.7	3	56.6	23	63.0	14	23.8
Vermont	26	360.5	26	158.0	23	164.2	22	77.8	13	59.9	11	23.1
Virginia	31	379.3	32	175.3	21	160.5	24	79.2	45	71.4	42	29.5
Washington	11	324.5	8	141.6	8	138.5	6	63.5	42	69.5	31	26.5
West Virginia	50	430.6	48	202.5	48	214.8	50	105.1	17	61.0	27	25.8
Wisconsin	22	354.7	21	157.1	27	174.7	28	81.4	36	67.7	32	26.5
Wyoming	13	328.1	12	146.0	12	151.2	16	73.1	28	64.1	18	24.4
Total USA		375.1		171.5		186.5		86.3		63.0		26.3

[a]Total cardiovascular disease is defined here as ICD/9 390–459.
[b]Coronary heart disease is defined here as ICD/9 410–414.
[c]Stroke is defined here as ICD/9 430–438.
[d]Rank is lowest to highest.
[e]Cardiovascular disease data for Puerto Rico are for major cardiovascular disease (390–448) and for 1996–97 only.
Source: NCHS compressed mortality file for the years 1994 to 1997. Age adjustments are based on the 2000 and 1940 standards. Reproduced from American Heart Association website http://www.american heart.org with permission.

unobstructed arteries have an excess capacity for blood flow, even under conditions of high demand, for example during prolonged physical exertion. Therefore, the gradual progression of the disease can go undetected until fibrous or complicated lesions form and blood flow is impeded to a significant extent. Because lesion development is silent, a surrogate measure of lesion progression, elevated LDL cholesterol levels, is used as a guide to determine whether individuals are at increased risk of developing atherosclerosis and whether the elevated levels, frequently in combination with other established risk factors, warrant treatment. Treatment of asymptomatic individuals with established hyperlipoproteinemia has been demonstrated to delay or prevent the onset of atherosclerotic events.

There are a number of disorders that are classified under the generic rubric of atherosclerosis. Angina pectoris (chest pain) is attributed to myocardial ischemia, at times precipitated by increased physical activity, accompanied by increased oxygen demand of the heart. A myocardial infarction describes a complete or near complete obstruction of a coronary vessel by thrombosis, hemorrhage, or ulceration of an atherosclerotic plaque or spasm of an artery, and usually involves vessels that are already partially obstructed. The result can be damage to the myocardium or sudden death. Cerebral ischemia refers to compromised blood flow to the brain due to atherosclerosis and can involve the carotid, subclavian, or vertebral arteries, as well as intracranial vessels. Clinical manifestations can include episodic cerebral ischemia with motor function, monocular blindness, or loss of consciousness. A complete occlusion results in ischemia due to thrombosis, embolism, and hemorrhage (either intracerebral or subarachnoid) of a cerebral artery and can result in neurologic deficiency. Atherosclerosis of the renal artery can lead to the development of hypertension. Occlusive atherosclerosis of the femoral, popliteal, and tibial arteries is associated with intermittent claudication and eventually gangrene. The presence of these lesions tends to correlate with atherosclerosis in other parts of the vascular tree.

Etiology

Both observational and interventional data indicate that the major contributing factors to the etiology of atherosclerosis are elevated LDL cholesterol levels, male gender, increased age, positive family history, current cigarette smoking, hypertension, diabetes mellitus, and low HDL cholesterol levels (Table 2). It is difficult to rank the relative contribution to each to the natural progression of the disease because the relative importance of each depends on severity and likely varies from individual to individual. Additional factors, albeit not always independent factors, such as elevated levels of homocysteine, lipoprotein (a) (Lp(a)), triglyceride, physical inactivity, and obesity also likely determine the course of development and rate of progress of atherosclerotic lesion formation. For this reason, the importance of devising individualized treatment plans when counseling patients cannot be overstated.

Primary Risk Factors

Nonmodifiable Risk Factors

Age Major prospective epidemiological investigations have provided strong evidence that the risk of developing atherosclerosis is increased with advancing age. These observations are consistent with what is known about the etiology of atherosclerotic lesions and the latency period between the appearance of fatty lesions and development of complicated plaque. In developed countries atherosclerosis

Table 2 Risk factors

Positive risk factors
Age
Male $\geq$ 45 years
Female $\geq$ 55 years or premature menopause without estrogen replacement therapy
Family history of premature coronary heart disease
Myocardial infarction or sudden death before 55 years or 65 years of age in parent or first-degree relative, male or female, respectively
Current cigarette smoking
Hypertension
Blood pressure $\geq$ 140/90 mmHg (should be confirmed) or taking antihypertension medication
Low high-density lipoprotein (HDL) cholesterol ($<$ 40 mg dl^{-1} (1.0 m mol l^{-1}))
Diabetes mellitus
Negative risk factor (subtract one positive risk factor)
High HDL cholesterol ($\geq$ 60 mg dl^{-1} (1.6 $mmol^{-1}$))

becomes the leading cause of death in males during the fourth decade of life and in females during the sixth decade of life. Males above the age of 45 years and females above the age of 55 years (or premature menopause without estrogen replacement therapy) are considered to be a positive risk factor.

Family history of premature coronary heart disease (genetics) Early studies have identified a trend towards familial aggregation of atherosclerosis among the relatives of individuals with established disease compared to those without disease. Similarly, studies of monozygotic and dizygotic twins provide supportive evidence for the heritability of atherosclerosis through comparisons of concordance rates of clinical events. Offspring of patients with a premature myocardial infarction have been reported to have higher circulating cholesterol levels than age-matched offspring of patients without a premature myocardial infarction. Familial combined hyperlipidemia is one of the major genetic causes of atherosclerosis within families. Numerous genetic mutations associated with different risks of developing atherosclerosis have been identified. As the methodology for screening advances, a greater understanding of how individual mutations alter the clinical progression of this multifactorial disease will emerge.

Modifiable Risk Factors

Elevated LDL cholesterol levels Elevated LDL cholesterol levels are currently considered to be the primary risk factor for the development of atherosclerosis. LDL levels below 2.57 mmol l^{-1} are classified as desirable, 2.58–3.34 mmol l^{-1} are classified in the near or above optimal, 2.35–4.41 mmol l^{-1} are classified in the borderline high, 4.12–4.89 mmol l^{-1} are classified in the high and above 4.91 mmol l^{-1} are classified as very high (Table 3). There are three categories of risk that determine LDL cholesterol lowing goals (Table 4). Risk category is determined on the basis on the presence of existing coronary heart disease (CHD) and CHD risk equivalents (other clinical forms of atherosclerotic disease such as peripheral arterial disease, abdominal aortic aneurysm, and symptomatic carotid artery disease; diabetes; or multiple risk factors that confer a 10-year risk for CHD > 20%) or the number risk factors (Table 2). Ten-year risk for CHD can be estimated using the criteria presented in Table 5. The higher the estimated risk for CHD the lower the LDL cholesterol goals. Cross-cultural comparisons have clearly identified a positive relationship between LDL cholesterol levels and the incidence of atherosclerosis. Similar observations have been made within populations. Migration studies which have resulted in elevations in LDL cholesterol levels, presumably due to environmental factors (diet and lifestyle), have supported these observations. Primary prevention trials resulting in decreased LDL cholesterol levels have led to reduced rates of atherosclerosis. More recent interest has centered on the potential for regression of atherosclerotic lesions by nonpharmacological means. Efforts in this area have been hampered by the lack of a safe, noninvasive, inexpensive, and accurate measure of atherosclerotic plaque. The primary nonpharmacological approach to reducing LDL cholesterol levels is to limit saturated fat, *trans* fatty acids, and cholesterol, increased physical activity, and maintain a healthy body weight. The goals for lowering LDL cholesterol levels as established by the National Cholesterol Education Program are shown in Table 5.

Table 3 Classification on the basis of total and LDL cholesterol levels

Total cholesterol;		
≤200 mg/dl	(≤ 5.2 mmol/L)	Desirable
201–239 mg/dl	(5.2–6.2 mmol/L)	Borderline–high
≥240 mg/dl	(≥6.2 mmol/L)	High
LDL cholesterol;		
< 100 mg/dl	(< 2.57 mmol/L)	Optimal
100–129 mg/dl	(2.58–3.34 mmol/L)	Near or above optimal
130–159 mg/dl	(2.35–4.11 mmol/L)	Borderline high
160–189 mg/dl	(4.12–4.89 mmol/L)	High
≥190 mg/dl	(4.91 mmol/L)	Very high

Current cigarette smoking A dose-dependent relationship has been observed between cigarette smoking and atherosclerotic events. The cessation of cigarette smoking is associated with a precipitous drop in the risk of a clinical event. This observation suggests that cigarette smoking, *per se*, may act as a triggering agent that results in a series of reactions culminating in an atherosclerotic event. Physiological responses associated with cigarette smoking include undesirable effects on platelet adhesiveness and clotting factors, increased heart rate, catecholamine levels, and myocardial oxygen demand, and decreased oxygen-carrying capacity of the blood. Similar effects of pipe and cigar smoking are not normally observed.

Table 4 Categories of risk and LDL cholesterol goal

Category	*LDL Goal*
CHD and CHD risk equivalents	< 2.59 mmol/L
Multiple (2 or more) risk factors	< 3.36 mmol/L
0–1 risk factors	< 4.14 mmol/L

Table 5A Estimate of 10-Year Risk for Men (Framingham Point Scores)

Age, y	*Points*
20–34	−9
35–39	−4
40–44	0
45–49	3
50–54	6
55–59	8
60–64	10
65–69	11
70–74	12
75–79	13

Total Cholesterol, mg/dL	*Points*				
	Age 20–39 y	*Age 40–49 y*	*Age 50–59 y*	*Age 60–69 y*	*Age 70–79 y*
< 160	0	0	0	0	0
160–199	4	3	2	1	0
200–239	7	5	3	1	0
240–279	9	6	4	2	1
≥ 280	11	8	5	3	1

	Points				
	Age 20–39 y	*Age 40–49 y*	*Age 50–59 y*	*Age 60–69 y*	*Age 70–79 y*
Nonsmoker	0	0	0	0	0
Smoker	8	5	3	1	1

HDL, mg/dL	*Points*
≥ 60	−1
50–59	0
40–49	1
< 40	2

Systolic BP, mm Hg	*If Untreated*	*If Treated*
< 120	0	0
120–129	0	1
130–139	1	2
140–159	1	2
≥160	2	3

Point Total	*10-Year Risk, %*
< 0	< 1
0	1
1	1
2	1
3	1
4	1
5	2
6	2
7	3
8	4
9	5
10	6
11	8
12	10
13	12
14	16
15	20
16	25
≥17	≥30

Table 5B Estimate of 10-Year Risk for Women (Framingham Point Scores)

Age, y	*Points*
20–34	−7
35–39	−3
40–44	0
45–49	3
50–54	6
55–59	8
60–64	10
65–69	12
70–74	14
75–79	16

Total Cholesterol, mg/dL	*Points*				
	Age 20–39 y	*Age 40–49 y*	*Age 50–59 y*	*Age 60–69 y*	*Age 70–79 y*
< 160	0	0	0	0	0
160–199	4	3	2	1	1
200–239	8	6	4	2	1
240–279	11	8	5	3	2
≥ 280	13	10	7	4	2

	Points				
	Age 20–39 y	*Age 40–49 y*	*Age 50–59 y*	*Age 60–69 y*	*Age 70–79 y*
Nonsmoker	0	0	0	0	0
Smoker	9	7	4	2	1

HDL, mg/dL	*Points*
≥ 60	−1
50–59	0
40–49	1
< 40	2

Systolic BP, mm Hg	*If Untreated*	*If Treated*
< 120	0	0
120–129	1	3
130–139	2	4
140–159	3	5
≥160	4	6

Point Total	*10-Year Risk, %*
< 9	< 1
9	1
10	1
11	1
12	1
13	2
14	2
15	3
16	4
17	5
18	6
19	8
20	11
21	14
22	17
23	22
24	27
≥25	≥30

Hypertension Elevated levels of both systolic and diastolic blood pressure (blood pressure $\geq 140/90$ mmHg) are associated with increased risk of developing atherosclerosis. The relationship appears to be linear. Recent evidence has confirmed that sodium restriction is an efficacious dietary approach to reducing blood pressure. In addition, compared to more conventional dietary patterns of developed countries, diets high in fruits, vegetables, and low-fat dairy products have also recently been demonstrated to result in lower blood pressures in both normotensive and hypertensive subjects.

Low HDL Cholesterol Levels HDL cholesterol levels are inversely correlated with the risk of developing atherosclerosis. HDL particles are involved in the process of reverse cholesterol transport, the only means by which cholesterol in peripheral tissues can be transported to the liver for metabolism and excretion from the body. Low HDL cholesterol levels ($< 1.03\,\text{mmol}\,\text{l}^{-1}$) are a risk factor for the development of atherosclerosis. High HDL cholesterol levels ($> 1.6\,\text{mmol}\,\text{l}^{-1}$) have been designated a negative risk factor (to be subtracted from the total number of risk factors when ascertaining an individual's risk profile) (Table 2). It is more difficult to raise HDL cholesterol levels than lower LDL cholesterol levels with nonpharmacological intervention. Common approaches are increased exercise, cessation of smoking, weight loss, and moderate alcohol intake. The magnitude of each intervention varies between individuals.

Diabetes mellitus Atherosclerosis is the most common cause of death in diabetic patients. Individuals with diabetes frequently have what is commonly referred to as an atherogenic lipoprotein profile: low HDL cholesterol levels and elevated triglyceride levels. In addition, insulin resistance, hyperinsulinemia, and glucose intolerance have been found to be atherogenic. Some evidence suggests that the increased risk of developing atherosclerosis observed in diabetic patients is contributed to by both the adverse lipoprotein profile and hyperinsulinemia itself. Adult-onset diabetic patients (type II) frequently present with a combination of obesity, adverse lipoprotein profile, and elevated blood pressure. This combination is frequently termed the 'insulin resistance syndrome' and is strongly associated with increased risk of developing atherosclerosis. Weight loss, in addition to the other dietary and lifestyle approaches to decreasing atherosclerosis risk, is among the most efficacious modalities of treating these patients.

Secondary Risk Factors

Certain risk factors are informally referred to as secondary risk factors, either because they are thought to be a consequence of a primary risk factor or due to insufficient evidence to classify them as a primary risk factor at the current time. The following gives details of some of these. As new data are accumulated, some may in the future shift from one category to another.

Obesity

Increased body weight for height is associated with various primary risk factors for atherosclerosis. These include hypertension, elevated LDL cholesterol levels, and low HDL cholesterol levels. Additionally, central or abdominal obesity is more strongly related to increased risk of developing atherosclerosis than overweight alone. When treating overweight patients cessation of weight gain should be the first goal and then a slow gradual approach to weight loss that combines energy restriction with increased energy output should be the second goal. Emphasis should be on permanent lifelong changes rather than drastic approaches.

Physical Inactivity

Physical inactivity has been associated with increased risk of developing atherosclerosis. It is difficult to determine whether physical fitness itself is the determining factor or whether individuals who habitually have low levels of physical activity also have a greater number of risk factors for atherosclerosis (i.e., obesity, hypertension, insulin resistance) that contribute to the clinical outcome. Regardless of the relationship, increased levels of physical activity should be encouraged. Currently, a number of health organizations are emphasizing the importance of maintaining a physically active lifestyle throughout the life cycle as a means of decreasing the risk of developing atherosclerosis and other degenerative diseases.

Elevated Homocysteine Levels

Homocysteine is a metabolite of methionine, an essential amino acid. Elevated homocysteine levels have been associated with an increased incidence of atherosclerosis. High levels of homocysteine can result from deficiencies in folic acid and vitamins B_6 and B_{12}. Since 1998 the USA has mandated that folic acid be added to all grain products. As a result there has been a secular decrease in homocysteine levels. There are ongoing studies to determine whether the decline in homocysteine levels due to folic acid supplementation or fortification will result in decreased rates of vascular disease.

Lp(a)

Lp(a) is a lipoprotein particle formed by the addition of a carbohydrate-rich protein (apolipoprotein; apo(a)) via a disulfide bond to apo B-100 on an LDL particle. High levels of Lp(a) are associated with an increased risk of developing atherosclerosis. There is a high degree of homology between Lp(a) and plasminogen. Plasminogen is converted to plasmin in plasma. Plasmin functions by dissolving blood clots. It has been hypothesized that high levels of Lp(a) compete with plasminogen, counteracting the fibrinolytic system, and increasing the risk of thrombosis. Whether Lp(a) levels are an independent risk factor for atherosclerosis have yet to be determined. Lp(a) levels are, for the most part, determined by genetics. Although dietary modification, hormone replacement therapy, and some lipids-lowering medications have been reported to alter Lp(a) levels, the changes tend to be small and, although statistically significant, may not be of a great enough magnitude to have a physiologically significant effect. More work in this area is likely to emerge in the near future.

Inflammation

Recent interest has centered on markers of inflammatory response, such as C-reactive protein and the risk of developing atherosclerosis. Epidemiological studies have noted an association between C-reactive protein and risk of an atherosclerotic event in apparently healthy males and those with a history of a prior event. Again, at this point it is premature to designate C-reactive protein as a primary risk factor for atherosclerosis. However, the results of future work will likely cause the issue to be revisited.

Dietary Modification to Decrease Risk of Developing Atherosclerosis

Diet

Current recommendations (individuals over the age of 2 years) to decrease the risk of developing atherosclerosis are to limit fat intake to 25–35 of energy, saturated fat to < 7% of energy, and cholesterol to less than 200 mg day^{-1}. This diet is termed a therapeutic lifestyle change (TLC) diet. Additional lifestyle recommendations include maintenance of body weight and regular physical activity. Table 6 includes the commonly occurring fatty acids, in both the body and food.

Saturated fatty acids Early evidence has clearly demonstrated that the consumption of foods relatively high in saturated fatty acids increases serum total cholesterol levels and that not all saturated fatty acids behave in the same way. Relatively short-chain fatty acids (C6:0 to C10:0) and stearic acid (C18:0) produce little or no change in serum cholesterol levels. Saturated fatty acids with chain lengths of C14:0, C16:0, and possibly C12:0 appear to be the most potent in increasing serum lipid levels. It has been postulated that stearic acid (18:0) is rapidly converted to oleic acid (18:1) and for this reason has a relatively neutral effect on plasma lipid levels. Subsequent work has further refined what is known about the hypercholesterolemia effect of saturated fatty acids. Dietary saturated fatty acids result in an increase in both LDL cholesterol and HDL cholesterol levels with little or no effect on triglyceride levels.

Table 6 Common fatty acids

Code	*Common name*	*Formula*
Saturated		
12:0	Lauric acid	$CH_3(CH_2)_{10}COOH$
14:0	Myristic acid	$CH_3(CH_2)_{12}COOH$
16:0	Palmitic acid	$CH_3(CH_2)_{14}COOH$
18:0	Stearic acid	$CH_3(CH_2)_{16}COOH$
Monounsaturated		
16:1*n*-7 *cis*	Palmitoleic acid	$CH_3(CH_2)_5CH{=}(c)CH(CH_2)_7COOH$
18:1*n*-9 *cis*	Oleic acid	$CH_3(CH_2)_7CH{=}(c)CH(CH_2)_7COOH$
18:1*n*-9 *trans*	Elaidic acid	$CH_3(CH_2)_7CH{=}(t)CH(CH_2)_7COOH$
Polyunsaturated		
18:2*n*-6,9 all *cis*	Linoleic acid	$CH_3(CH_2)_4CH{=}(c)CHCH_2CH{=}(c)CH(CH_2)_7COOH$
18:3*n*-3,6,9 all *cis*	α-linolenic acid	$CH_3CH_2CH{=}(c)CHCH_2CH{=}(c)CHCH_2CH{=}(c)CH(CH_2)_7COOH$
18:3*n*-6,9,12 all *cis*	γ-linolenic acid	$CH_3(CH_2)_4CH{=}(c)CHCH_2CH{=}(c)CHCH_2CH{=}(c)CH(CH_2)_4COOH$
20:4*n*-6,9,12,15 all *cis*	Arachidonic acid	$CH_3(CH_2)_4CH{=}(c)CHCH_2CH{=}(c)CHCH_2CH{=}(c)CHCH_2CH{=}(c)CH(CH_2)_3COOH$
20:5*n*-3,6,9,12,15 all *cis*	Eicosapentaenoic acid	$CH_3(CH_2CH{=}(c)CH)_5(CH_2)_3COOH$
22:6*n*-3,6,9,12,15,18 all *cis*	Docosahexaenoic acid	$CH_3(CH_2CH{=}(c)CH)_6(CH_2)_2COOH$

c, *cis*; t, *trans*.

Foods relatively high in saturated fatty acids include meat and full-fat dairy products. Saturated fatty acids tend to be solid at room temperature. The notable exceptions are the tropical oils; palm, palm kernel, and coconut. Efforts to reduce saturated fatty acid intake should include use of lean meat, trimming of excess fat or the skin of poultry, and the substitution of fat-free or low-fat dairy products for their full-fat counterparts. Judicious use of ingredient listings and nutrient labels on processed foods will help individuals achieve the goal of reducing the saturated fatty acid content of their diet.

Omega-3 (n-3) polyunsaturated fatty acids Reports of low rates of atherosclerosis in Greenland Eskimos and the inverse relationship between fish consumption and mortality from atherosclerosis raised interest in oils derived from fish and marine mammals. These oils, high in polyunsaturated fatty acids, are distinguished from other oils high in polyunsaturated fatty acids, such as vegetable oils, by the presence of *n*-3 fatty acids. As a class, *n*-3 fatty acids are defined as having a double bond at the third carbon from the methyl end of the acyl chain. In contrast, the more common *n*-6 fatty acids are defined as having a double bond at the sixth carbon from the methyl end of the acyl chain. Potential mechanisms of *n*-3 fatty acids in decreasing risk of developing atherosclerosis include decreased postprandial triglyceride levels, reduced risk of arrhythmic events, decreased blood pressure, altered endothelial function, and release of cytokines and other inflammatory mediators. The relative importance of each of these factors to the epidemiological observation of an inverse relationship of atherosclerosis incidence and fish or *n*-3 fatty acid intake has yet to be determined. The major dietary *n*-3 fatty acids from plant sources is alpha-linolenic acid (18:3*n*-3) and from marine sources are eicosapentaenoic acid (20:5*n*-3), and docosahexaenoic acid (22:6*n*-3) (Table 6). The different physiological effects of each is under active investigation at this time.

***Trans* fatty acids** *Trans* fatty acids are fatty acids that contain at least one double bond in the *trans* configuration. Dietary *trans* fatty acids occur naturally in meat and dairy products as a result of anaerobic bacterial fermentation in ruminant animals. *Trans* fatty acids are also created when liquid vegetable or fish oil is hydrogenated; this is frequently done to increase their plasticity and chemical stability for subsequent food processing. *Trans* fatty acids raise LDL cholesterol levels, similarly to saturated fatty acids. However, *trans* fatty acids do not raise HDL cholesterol levels, as do saturated fatty acids. For this reason there is concern that dietary *trans* fatty acids are worse than saturated fatty acids. Additionally, some literature suggests that *trans* fatty acids may increase Lp(a) levels. Over the past decade it appears that as a result of changes made by industry the intake of *trans* fatty acids is decreasing. Due to the considerably higher intake of saturated than *trans* fatty acids it still seems prudent to maintain emphasis on reducing saturated fat intake.

Dietary cholesterol In humans there is a positive correlation between dietary cholesterol intake, and both serum cholesterol levels and risk of developing atherosclerosis. Interpretation of the dietary cholesterol literature is complicated by the large degree of variation among studies and the tendency for responsiveness to vary over a wider range than other dietary variables. Nonetheless, the majority of evidence supports the premise that increased dietary cholesterol intake results in increased LDL cholesterol levels. The major sources of dietary cholesterol are eggs, meat, and dairy products. Since dietary cholesterol is frequently found in foods which are also relatively high in saturated fatty acids, reducing the saturated fatty acid intake should also result in a reduction in the cholesterol content of the diet.

Soluble fiber Dietary soluble fiber has been reported to have a modest hypocholesterolemic effect. A major contributor to this outcome is the β-glucan fraction of the soluble fiber. A variety of mechanisms have been proposed for the cholesterol-lowering effect, most likely relating to decreased absorption of cholesterol. There is no evidence that insoluble fiber has an effect on plasma lipid levels.

Dietary pattern It can be difficult for individuals, however well intended, to modify their diet to meet certain dietary goals, for example, to consume < 7% of energy as saturated fat. However, with the exception of high-risk individuals who need more individualized instruction provided by a qualified professional, the general recommendation to consume a diet high in fruits, vegetables, and whole grain which also includes nonfat or low-fat dairy products, fish, poultry, and lean meat is consistent with the more technical recommendations and should be consistently encouraged on a population-wide basis. Additionally, it is important to couple such dietary recommendations with a prescription that further encourages a pattern of regular activity, through the life cycle, that matches levels of physical activity to energy intake with the expressed intent of avoiding overweight or obesity.

Summary

Atherosclerosis is a disorder of blood vessels characterized by a thickening of the arterial wall. This process frequently leads to partial or total occlusion of the vessel to blood flow. The location of the impeded or blocked blood flow determines the clinical consequence of the disorder. A major contributing factor to the development of atherosclerotic lesions is elevated LDL cholesterol levels, followed by other risk factors. The primary risk factors include advanced age, family history of premature atherosclerosis, current cigarette smoking, hypertension, lower HDL cholesterol levels (high HDL cholesterol levels are classified as a negative risk factor), and diabetes mellitus. Sometimes referred to as secondary risk factors, either because they are thought to be a consequence of a primary risk factor or due to lack of sufficient evidence to classify them as a primary risk factor at the time, are obesity, physical inactivity, elevated homocysteine levels, elevated Lp(a) levels, and elevated inflammation markers. With the intent of lowering LDL cholesterol levels, the major dietary interventions are decreased saturated fat, cholesterol, and *trans* fatty acid intakes, and increased *n*-3 fatty acid intakes and soluble fiber. These changes from the current diet of developed countries can best be achieved by encouraging the consumption of a diet high in fruits, vegetables, and whole grain which also includes nonfat or low-fat dairy products, fish, poultry, and lean meat. Additionally, these recommendations should be coupled with a pattern of regular activity that is matched to energy intake to avoid overweight or obesity.

See also: **Amino Acids**: Properties and Occurrence; **Cholesterol**: Role of Cholesterol in Heart Disease; **Coronary Heart Disease**: Etiology and Risk Factor; **Diabetes Mellitus**: Etiology; **Exercise**: Metabolic Requirements; **Hypertension**: Hypertension and Diet; **Obesity**: Etiology and Diagnosis; **Smoking, Diet, and Health**; **Fatty Acids**: *Trans*-fatty Acids: Health Effects

Further Reading

American Heart Association website. http://www.americanheart.org.

Benowitz NL (1997) The role of nicotine in smoking-related cardiovascular disease. *Preventive Medicine* 26(4): 412–417.

Brown L, Rosner B, Willett WW and Sacks FM (1999) Cholesterol-lowering effects of dietary fiber: a meta-analysis. *American Journal of Clinical Nutrition* 69(1): 30–42.

de Valk B and Marx JJ (1999) Iron, atherosclerosis, and ischemic heart disease. *Archives of Internal Medicine* 159(14): 1542–1548.

Hankey GJ and Eikelboom JW (1999) Homocysteine and vascular disease. *Lancet* 354(9176): 407–413.

Hobbs HH and White AL (1999) Lipoprotein(a): intrigues and insights. *Current Opinion in Lipidology* 10(3): 225–236.

Journal of the American Medical Association (2001) Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *Journal of the American Medical Association* 285(19): 2486–2497.

Krauss RM, Eckel RH, Howard B *et al.* (2000) AHA Dietary guidelines. Revision 2000: a statement for healthcare professionals from the Nutrition Committee of the American Heart Association. *Circulation* 102: 2296–2311.

Kritchevsky D (1999) Diet and atherosclerosis. *American Heart Journal* 138(5): S426–S430.

Ordovas JM and Schaefer EJ (1999) Genes, variation of cholesterol and fat intake and serum lipids. *Current Opinion in Lipidology* 10(1): 15–22.

Sacks FM, Svetkey LP, Vollmer WM *et al.* Effects on blood pressure of reduced dietary sodium and the dietary approaches to stop hypertension (DASH) diet. *New England Journal of Medicine* 344: 3–10.

World Health Organization website. http://www.who.int.

Authenticity *See* **Adulteration of Foods**: History and Occurrence; Detection

AVOCADOS

A K Cowan, Stockholm, Sweden
B N Wolstenholme, University of Natal, Scottsville, South Africa

Background

The fruit of the avocado is regarded as one of the most nutrient-dense fruits available. Although its high oil content has been mistakenly thought of as a disadvantage, closer scrutiny of the nutritional facts reveals that avocado eaten in moderation does not contribute to weight gain, contains large quantities of the desirable unsaturated and monounsaturated fatty acids, and is rich in fiber, protein, vitamins, antioxidants, and minerals. This nutrient-rich fruit is therefore without rival and, in many respects, a veritable fruit of paradise.

In this chapter, brief descriptions of avocado production, emphasizing the 'subtropical' ecotype, and processing are provided. These are followed by an account of the nutritional status of avocado as a food source. Data on nutrient, fat, vitamin, antioxidant, mineral, and sugar content and composition are supplied and are synthesized based on minimum and maximum values published in the literature. The nutritional benefit of a diet containing avocado is also discussed. Reference is also made to the major allergen in avocado that is relevant in the 'latex-fruit syndrome' that affects latex-allergic individuals.

Avocado Fruit Production

History, Cultivation, and Fruit Growth

Avocado originated in central America, and the fruit has been consumed as part of the diet by the indigenous peoples of that region for more than 5000 years. The word 'avocado' derives from a corruption of the Spanish words 'ahuacate' or 'aguacate,' which are adaptations of the Aztec 'ahuacatl.' The fruit of the first vegetatively propogated cultivar 'Fuerte' is pear-shaped, and this characteristic probably resulted in the inappropriate colloquialism, 'avocado pear.' However, the horticulturally and agriculturally correct term for the fruit is simply avocado.

The botanical name for avocado is *Persea americana* Mill., and three ecological races (today regarded as botanical varieties or subspecies) are recognized; *Persea americana* var. *drymifolia* (Mexican ecotype), *Persea americana* var. *guatemalensis* (Guatemalan ecotype), and *Persea americana* var. *americana* (West Indian ecotype). Fruits of the West Indian ecotype are usually large with a thick peel and have a low oil content (< 8%) and higher water and sugar content. Trees of the Mexican ecotype yield fruit with a high oil content (up to 30%). Those of the Guatemalan ecotype are recognized as having the most horticulturally desirable traits, have an intermediate oil content, and are characterized by a 'nutty' flavor. Amongst the most important commercial cultivars of the 'subtropical' avocado are 'Hass,' 'Fuerte,' 'Ettinger,' and 'Pinkerton,' which, are all chance seedlings with superior fruit quality. Several countries have extensive selection and breeding programs which are starting to produce fruit of improved quality. 'Hass' is the most sought after subtropical cultivar, having almost eclipsed 'Fuerte' in California, and forms the basis of the subtropical avocado export industry to Europe.

The Mexican and Guatemalanm ecotypes, whose progeny today form the basis of the technologically advanced 'subtropical' avocado industries, are indigenous to subtropical and tropical highland climates. So-called 'West Indian' avocados are lowland tropical in origin. Mexico (by far the largest producer), California, Israel, Spain, Chile, South Africa, and Australasia all have industries based on Guatemalan and Guatemalan × Mexican hybrid cultivars. However, virtually all subtropical countries grow avocados to some extent. Large crops of 'West Indian' and 'West Indian' × Guatemalan hybrid avocados are also produced in tropical countries, from less technologically advanced industries. Tropical avocado industries, with the conspicuous exception of Florida, USA, are usually based on seedling (not grafted) trees. The fruits, although making a substantial contribution to the diets of mostly poor, tropical lowland peoples, are horticulturally and nutritionally inferior to the selected subtropical cultivars.

The avocado tree is evergreen and flowers in early spring, and fruitlets are usually visible by late spring/early summer. Typically, careful and correct orchard management practise is required to ensure production of quality fruit. Amongst the most important ecophysiological factors that affect fruit production are irradiance, temperature, water stress, and salinity. Botanically, the fruit of the avocado is described as a berry. Fruit growth continues for anything from 20 to 60 weeks, depending on the cultivar, environment, and cultural practice, and the fruit is harvested when horticulturally mature as it does not ripen on the tree. Harvested fruit are washed, graded, often

waxed, and packed in cartons typically holding 4–6 kg of uniform fruit. European markets prefer the individual fruit weight to be in the range of 200–350 g, for which higher prices are paid. It is now known that the final fruit size of avocado is the result of the number of cell divisions during flower set and fruit growth and that cell division is dependent on isoprenoid biosynthesis, carbohydrate metabolism, and maintenance of plant hormone homeostasis. Zinc, boron, magnesium, and calcium are considered important elements in the production of quality fruit. Nitrogen is a 'manipulator element' used to control tree vigor, and must be appropriately managed as excessive tree vigor is counterproductive to good fruit quality.

Maturity

Oil content and/or dry matter of the fleshy mesocarp, which surrounds the single large seed, is used by most producers to determine harvestable maturity and consumer acceptability. For the cultivar 'Fuerte,' a minimum oil content of 8% (dry mass) was applied in California for many years, but today, this has been replaced by a minimum dry matter standard of *c*. 21%, depending on the cultivar and region. This is based on the percentage moisture (determined using a microwave oven), which is an indirect measure of oil content based on the cultivar-specific constant of the sum of percentage moisture and percentage oil. By 1983, percentage dry matter (or its reciprocal, percentage moisture) had become standard for the determination of avocado fruit maturity and is now used worldwide.

Ripening

Avocado fruit are classified as climacteric (i.e., fruit that can be harvested unripe and then undergo normal ripening). Although detachment of the fruit is not a prerequisite for ripening of most climacteric fruit, avocado fruit do not ripen while attached to the tree. Postharvest, the fruit ripen within 5–10 days at temperatures between 15 and 24 °C. Both ripening and fruit softening are delayed by precooling the fruit immediately after harvest to 5–6 °C, and this strategy is used by most producers to maintain fruit quality during transport/export of fruit. A return to ambient temperature after a period of cold storage sees the acceleration of the ripening process. Ripening in avocado can also be stimulated by exposure to the plant hormone, ethylene.

During ripening, the fruit undergoes marked changes in a variety of biochemical processes. These include an increase in respiration (i.e., consumption of oxygen and release of carbon dioxide), increased ethylene production, changes in texture (i.e., fruit softening), and the production of volatiles such as β-caryophyllene (28% of total volatiles), α-copaene (11%), α/β-cubebene (8%), α-farnesene (6%), decanal (6%), and heptanal (3%), which impart flavor. The substrate for respiration in avocado appears to be carbohydrate and not oil. However, some degradation of oils does take place. An increase in ethylene biosynthesis is typical of climacteric fruits such as avocado, and the maximum ethylene production usually coincides with the maximum respiration rate (amount of carbon dioxide released). Cold storage reduces this climacteric rise in respiration.

Fruit softening is a consequence of changes in cell-wall metabolism resulting from alterations in cellulase (EC 3.2.1.4), polygalacturonase (EC 3.2.1.15) and pectinesterase (EC 3.1.1.11) activities. In some varieties (e.g., 'Hass'), the color of the peel changes from green to purple/black owing to the accumulation of anthocyanin pigments. Other changes include a slight increase in the concentration of glucose and fructose in the fruit flesh.

Processing of Avocado

Avocado is usually consumed fresh in salads, as a savory dish, as a sandwich filling, as guacamole, or sweetened as a dessert. It was the extraction of oils from avocado and the production of guacamole that really started avocado processing. Apparently, the production of guacamole on a commercial scale was started as long ago as 1964 in California, and by 1981, 6000 tonnes of fresh avocado was being processed annually.

For processing, the preferred cultivar is 'Hass' because of its superior flavor, mesocarp color, 'keeping' quality, and year-round availability. Fruit with a 25% dry matter content (or 13% oil) are considered ideal for processing. Avocado flesh is also marketed frozen, processed into a sauce, dehydrated to a powder, and extracted for its oil.

Guacamole, Avocado Sauce, and Oil

Before processing can take place, fruit are uniformly ripened at 22 °C in the presence of ethylene gas for 2–5 days. When the fruit are soft, they are cooled to 5 °C and surface-sterilized in 200 mg of hypochlorite per liter. After removal of the stems (pedicels) and seed, the pulp (mesocarp) is mechanically separated from the peel. The pulp is then mixed with the ingredients (usually includes lime or lemon juice, salt, and salsa or Picante) to yield a uniform textured guacamole. The guacamole is then placed in containers and stored frozen.

The process for the production of avocado sauce is similar to that used for the preparation of guacamole up to pulp extraction. For sauce, the pulp is sheared in a high-speed blender with water, emulsifiers, and spices. Sauces are usually packaged in polyethylene containers and then stored frozen.

Dehydration of avocado is either by spray- or drum-drying in much the same way as for the preparation of milk powder, but the product is less attractive than that of sugar-storing fruits.

Oil is extracted from avocado with organic solvents, hydraulically (pressing), or by centrifugation. The last method is more desirable as the product is free from solvent residue. Nevertheless, solvent extraction enables the extraction of various fractions, some of which are used in the pharmaceutical industry. For example, avocado unsaponifiables (including hydrocarbons, tocopherols, triterpenes, sterols, and other unidentified compounds), which seem to be beneficial in the treatment of periodontal and osteoarticular diseases, function by stimulating deposition of repair material in affected areas by enhancing transforming growth factor β (TGF β) in articular cartilage. Avocado oil is also used in cosmetic preparations such as skin moisturizers and body lotions.

Nutritional Status

Avocado is considered a nutritious fruit and contains higher quantities of soluble and insoluble fiber and protein than many other fleshy fruits. Avocado is also a rich source of potassium and the vitamins E and C, and β-carotene (provitamin A). Although low, a vitamin A value of 150 μg RE (retinol equivalents) per kilogram has been reported. Furthermore, the monounsaturated fatty acids in avocado effectively reduce blood levels of the undesirable low-density lipoprotein (LDL, cholesterol) while increasing levels of the beneficial, high-density lipoprotein (HDL). (*See* **Cholesterol**: Factors Determining Blood Cholesterol Levels; **Vitamins**: Overview.)

Nutrient Composition/density

Unlike foods that are described as providing 'empty calories,' the avocado is a nutrition-rich 'protective' fruit. The energy content per 100-g serving has been estimated at 800 kJ, depending on the cultivar and growing conditions. A detailed analysis of the composition of avocado mesocarp (the edible portion of the fruit) of a typical subtropical cultivar has revealed a nutritional status, which is summarized in Figure 1.

Avocado contains approximately 2.3% protein on a fresh-weight basis, which is between two and 10 times that of most other fleshy fruits and vegetables analyzed. Although avocado contains all of the essential amino acids (i.e., it is 'complete'), it is not generally regarded as nutrient-dense for protien, certainly not in comparison to meat or eggs, for example. A mixture of soluble and insoluble fiber is ideal in the diet, and avocado contains appreciable quantities of both (2.1 and 2.7%, respectively).

In avocado, the sum of percentage water plus percentage oil is constant in the range 88–91%, and for each gram of water lost, there is a 1 g increase in oil, in fruit containing between 8 and 22% oil. Changes in oil content are also associated with changes in fiber and protein. Therefore, the energy value of avocado is almost entirely due to its oil content, and any variation is due solely to changes in the percentage oil. It is probably this observation that led to the mistaken idea that consuming avocado increases body mass. On the contrary, addition of avocado to the diet has been shown to cause a small but significant average weight loss. One possible explanation is that avocado speeds up the basal metabolic rate in humans. Another hypothesis relates to the type of oil, and avocado is rich in the beneficial monounsaturated fatty acids.

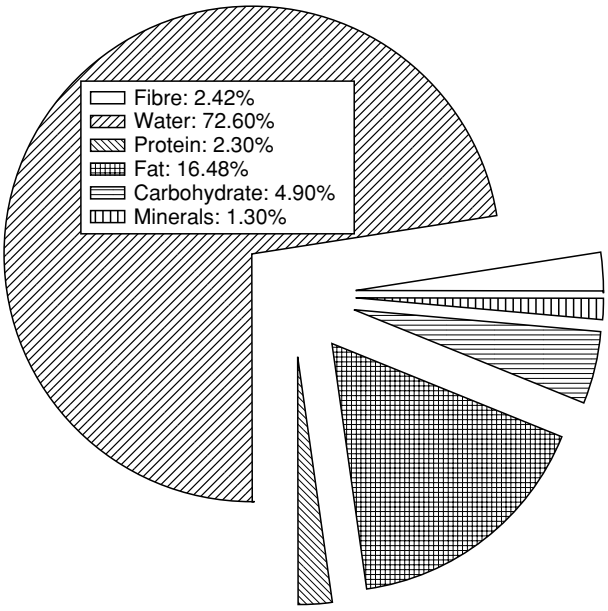

Figure 1 Nutrient content and composition of subtropical fruit flesh. Data are the average between the minimum and maximum values reported by Ahmed EM and Barmore CD (1980) Avocado. In: Nagy S and Shaw PE (eds) *Tropical and Subtropical Fruits: Composition, Properties and Uses*, pp. 121–156. Westport, CT: AVI; Seymour GB and Tucker GA (1993) Avocado. In: Seymour GB, Taylor JE and Tucker GA (eds), *Biochemistry of Fruit Ripening*, pp.53–81. London: Chapman & Hall; Slater GG, Shankman S, Shepherd JS and Alfin-Slater RB (1975) Seasonal variation in the composition of Californian avocado. *Journal of Agricultural and Food Chemistry* 23: 468–474, and references cited therein.

The fatty acid content and composition of avocado have been iterated in most texts on the physiology and biochemistry of fruit growth and ripening. Avocado lipids can be divided into the following fractions: (1) neutral lipids (tri-, di-, and monoacylglycerols), (2) phospholipids, (3) glycolipids, and (4) free fatty acids. The neutral lipid fraction constitutes 96% of the total lipid content of avocado, and the majority are triglycerols. The major triglycerols identified in avocado are dioleyl palmitin, triolein, dioleypalmitolein, linoleyl oleyl palmitin, and linoleyl diolein. Within each of these triglycerols, 18:1, 18:2, 16:0, and 16:1 are the major fatty acids present, and the relative concentration (percentage of total lipid) of each is in the range 59–81% (18:1), 7–14% (18:2), 7–22% (16:0), and 3–11% (16:1). These values are comparable with those for olive oil, making avocado which is easily digested, an ideal substitute for olive oil in cooking, and in the preparation of salad dressings, although it is currently more expensive.

Unsaturated fats are either mono- or polyunsaturated, and the ratio of polyunsaturated fatty acids to saturated fatty acids (P/S), which is used as an indicator of nutritional value by nutritionists, is about 0.74 for avocado. A diet high in polyunsaturated fatty acids reduces the level of the undesirable LDL, whereas a high dietary level of monounsaturated fatty acids also maintains levels of the desirable HDL. Thus, it should not be surprising that results from trials in which the monounsaturate-rich avocado was used as a dietary component revealed a significant decline in total cholesterol with preservation of the HDL level. Likewise, substituting avocado for butter, margarine, and cheese significantly reduced blood cholesterol levels, and increased HDL up to 16%. Accordingly, avocados have gained acceptance by the Heart Foundations of several countries, and can even be prescribed for convalescent heart patients.

Fat- and Water-soluble Vitamins, and Antioxidants

The vitamin content and composition of avocado are shown in Table 1. Although vitamin D has been identified in avocado, no values have been published. It has been reported, however, that the vitamin D content of avocado is higher than that of butter and eggs. Vitamins C and E and provitamin A (β-carotene) are believed to function as antioxidants and protect against damage from oxygen 'free radicals.' Although oxygen is essential to life processes, it is also damaging if converted to reactive oxygen species (e.g., the superoxide anion or the hydroxyl radical), which can cause cell mutation and contribute to aging, cancers, arthritis, and heart disease. The three antioxidant vitamins are effective at disarming these reactive oxygen species, and it has been stated that for each of the three, and for any given daily kJ proportion, the avocado provides about twice the proportion of this nutrient.

Table 1 Vitamin content and composition of avocado fruit flesh

Component (per 100 g fresh weight)	*Concentration range*[a]
β-Carotene (provitamin A) (IU)	370–750
α-Tocopherol (vitamin E) (IU)	1.6–2.4
Ascorbic acid (vitamin C) (mg)	1.6–30.0
Biotin (μg)	3.2–10.0
Choline (mg)	17–22
Folacin (μg)	30–62
Niacin (mg)	1.4–3.5
Pantothenic acid (mg)	0.25–1.14
Pyridoxine (vitamin B_6) (mg)	0.22–0.62
Riboflavin (vitamin B_2) (μg)	95–230
Thiamine–HCl (vitamin B_1) (μg)	60–240
Phytyl menaquinone (vitamin K) (μg)	0–8
Calciferols (vitamin D)	Unknown

[a]Concentration is dependent on the cultivar and stage of ripening. Values were derived from data published by Ahmed EM and Barmore CD (1980) Avocado. In: Nagy S and Shaw PE (eds) *Tropical and Subtropical Fruits: Composition, Properties and Uses*, pp. 121–156. Westport, CT: AVI; Seymour GB and Tucker GA (1993) Avocado. In: Seymour GB, Taylor JE and Tucker GA (eds), *Biochemistry of Fruit Ripening*, pp. 53–81. London: Chapman & Hall; Slater GG, Shankman S, Shepherd JS and Alfin-Slater RB (1975) Seasonal variation in the composition of Californian avocado. *Journal of Agricultural and Food Chemistry* 23: 468–474, and references cited therein.

In addition to the antioxidant vitamins, 2(R)-(12Z, 15Z)-2-hydroxy-4-oxoheneicosa-12,15-dien-1-yl acetate and persenone A and persenone B have been isolated from avocado mesocarp as inhibitors of superoxide and nitric oxide generation. These compounds do not scavenge reactive oxygen species (like the vitamins) but rather suppress free-radical generation and may therefore be effective chemopreventive agents in inflammation associated carcinogenesis. Thus, the consumption of avocado may be one way of ingesting more of the antioxidants that help to protect against cancers, heart disease, and aging.

Mineral Composition

The mineral content and composition of avocado are presented in Table 2. As shown, avocado is a rich source of potassium, which is purported to protect against the risk of strokes in humans and reduces the incidence of strokes by up to 40%. One of the confounding factors in the cause of strokes is increased blood pressure, which has been associated with a high intake of sodium. The avocado is as low in sodium as it is high in potassium. (*See* **Minerals – Dietary Importance**.)

Phosphorus, calcium, and magnesium are abundant minerals in avocado. Most other minerals, including iron, occur in amounts of less than 1 mg per gram fresh weight of edible fruit.

Table 2 Mineral content and composition of avocado fruit flesh

Mineral (mg per 100 g fresh weight)	*Concentration range*[a]
Potassium	340–723
Magnesium	40–60
Phosphorus	20–80
Calcium	10–15
Sodium	5–15
Iron	0.5–2
Boron	1–3

[a]Concentration is dependent on the soil type, cultivar, and cultural practice. Values were derived from data published by Ahmed EM and Barmore CD (1980) Avocado. In: Nagy S and Shaw PE (eds) *Tropical and Subtropical Fruits: Composition, Properties and Uses*, pp. 121–156. Westport, CT: AVI; Slater GG, Shankman S, Shepherd JS and Alfin-Slater RB (1975) Seasonal variation in the composition of Californian avocado. *Journal of Agricultural and Food Chemistry* 23: 468–474, and references cited therein.

Table 3 Sugar content and composition of ripe avocado fruit flesh

Sugar (mg per 100 g fresh weight)	*Concentration range*[a]
Glucose	0.14–1.28
Fructose	0.08–0.65
D-*Manno*-heptulose	0–3.82
Perseitol	0–2.06

[a]Concentration is dependent on the cultivar and stage of ripening. Values were derived from data published by Liu X, Robinson PW, Madore MA, Witney GA and Arpaia ML (1999) 'Hass' avocado carbohydrate fluctuations. II. Fruit growth and ripening. *Journal of the American Society of Horticultural Science* 124: 676–681; Seymour GB and Tucker GA (1993) Avocado. In: Seymour GB, Taylor JE and Tucker GA (eds), *Biochemistry of Fruit Ripening*, pp. 53–81. London: Chapman & Hall, and references cited therein.

Carbohydrates

During avocado fruit ripening, there is an increase in the concentration of glucose and fructose. In addition to these monosaccharides, avocado also contains several unusual sugars including the seven-carbon sugar alcohol, perseitol, and its reduced form, D-*manno*-heptulose. The sugar content and composition of avocado are given in Table 3. (*See* **Carbohydrates**: Classification and Properties.)

Studies in the 1940s indicated that ingestion of avocado could lead to the presence of sugar in urine. It was later discovered that *manno*-heptulose, the major reducing sugar in avocado, could induce hyperglycemia when administered to humans and other animals orally. Even so, this author knows of no unequivocal information linking the consumption of avocado with diabetic symptomology. In fact, the *manno*-heptulose content of the edible portion of avocado declines by up to 80% as the fruit ripens. Thus, any amounts of *manno*-heptulose ingested are likely to be very low, and some cultivars appear to have none in harvested fruit.

Interestingly, *manno*-heptulose is regarded as an inhibitor of hexokinase activity in plants. This enzyme phosphorylates glucose during glucose metabolism to give glucose-6-phosphate, which is then metabolized in the glycolytic pathway to provide energy and organic substrates. Whether this biochemical pathway is affected by a diet rich in avocado (and therefore some *manno*-heptulose) in humans is unknown.

Nutritional Benefits and Costs

Nutrient Density

The extraordinary nutrient density, at least of the fruits of subtropical avocado cultivars which form the basis of world trade and of consumption in the affluent first world, has been noted, and the subtropical avocado has been described as very nutritious of all fleshy fruits. Thus, comparatively small amounts of avocado are extremely helpful in upgrading the quality of the diet in an increasingly health-conscious world. Furthermore, the relatively small amounts consumed per serving (80–100 g) are no cause for concern where affluence has contributed to overweight or even obesity. The problem in poor, mostly tropical third world countries is very different, and here the energy and nutrient density of the fruit, even of less nutritious cultivars, can make a substantial contribution to guarding against malnutrition.

Protection from Coronary Heart Disease

Coronary artery disease results from the build-up of cholesterol and other lipids in coronary arteries. Cholesterol is carried by lipoproteins, and LDL causes the most damage. HDL, however, protects vessel walls from atherosclerosis. Monounsaturated fatty acids increase the blood level of HDL. Avocado is high in monounsaturated fats and contains no cholesterol. A diet rich in avocado therefore has a favorable effect on blood fats, decreases cholesterol, and preserves HDL. Trials have revealed that a diet rich in avocado causes a decline in total serum cholesterol of 16% in healthy individuals. In hypercholesterolemic subjects, a decrease in serum cholesterol of 17%, LDL-cholesterol of 22%, and triglycerols of 22% was observed, while HDL-cholesterol increased by 11%. There has, therefore, been a welcome change in attitude towards avocado consumption in people at risk of coronary heart disease. (*See* **Coronary Heart Disease**: Prevention.)

Hypersensitivity

Hypersensitivity occurs in latex-allergic individuals who react to avocado (and other fruits including

banana and kiwifruit) within 60 min of ingestion and typically display one or more of the following symptoms: mouth irritation, angiodema, urticaria, asthma, nausea, vomiting, diarrhea, rhinitis, or anaphylaxis. Allergy to avocado is increasing, especially in the USA and Mexico, and has been estimated at around 1% of the general population. Avocado allergy is of particular importance in the 'latex-fruit syndrome' observed in at least 40% of latex-allergic individuals. As mentioned above, the allergens elicit diverse immunoglobulin E-mediated reactions in sensitized individuals. Many of the known plant food allergens are in fact proteins produced in plants either following pathogen attack or after induction of the plant stress syndrome. These are called pathogenesis-related proteins (PR), and 14 families are now recognized. The major allergen in avocado has been identified: a 32-kDa protein known as '*Prs a 1*,' which has endochitinase activity and is recognized by 15 out of 20 avocado and/or latex-allergic individuals. This allergen belongs to the PR-3 family of pathogenesis-related proteins and is induced in plants by ethylene treatment, but allergic activity is lost in heating. This probably explains why plant foods containing this allergen that are consumed after cooking (e.g., green beans) are not usually associated with the latex-fruit syndrome.

See also: **Allergens**; **Atherosclerosis**; **Coronary Heart Disease**: Etiology and Risk Factor; Prevention; **Diabetes Mellitus**: Etiology; **Fatty Acids**: Dietary Importance; **Food Intolerance**: Food Allergies; **Minerals – Dietary Importance**; **Salad Crops**: Dietary Importance; **Slimming**: Slimming Diets; **Vegetable Oils**: Oil Production and Processing; **Vitamins**: Overview

Further Reading

Ahmed EM and Barmore CD (1980) Avocado. In: Nagy S and Shaw PE (eds) *Tropical and Subtropical Fruits: Composition, Properties and Uses*, pp. 121–156.Westport, CT: AVI Publishing.

Bergh B (1992) Nutritious value of avocado. *California Avocado Society Yearbook* 76: 123–135.

Bower JP and Cutting JG (1988) Avocado fruit development and ripening physiology. *Horticultural Reviews* 10: 229–271.

Breiteneder H and Ebner C (2000) Molecular and biochemical classification of plant-derived food allergens. *Journal of Allergy and Clinical Immunology* 106: 27–36.

Cowan AK, Cripps RF, Richings EW and Taylor NJ (2000) Fruit size: Towards an understanding of the metabolic control of fruit growth using avocado as a model system. *Physiologia Plantarum* 111: 127–136.

Kurlaender A (1996) Avocados. In: Somogyi IP, Barrett DM and Hui YH (eds) *Processing Fruits: Science and Technology*, vol. 2, pp. 445–457. Lancaster, PA: Technomic.

Lui X, Robinson PW, Madore MA, Witney GA and Arpaia ML (1999) 'Hass' avocado carbohydrate fluctuations. II. Fruit growth and ripening. *Journal of the American Society of Horticultural Science* 124: 676–681.

Seymour GB and Tucker GA (1993) Avocado. In: Seymour GB, Taylor JE and Tucker GA (eds) *Biochemistry of Fruit Ripening*, pp. 53–81. London: Chapman & Hall.

Slater GG, Shankman S, Shepherd JS and Alfin-Slater RB (1975) Seasonal variation in the composition of Californian avocado. *Journal of Agricultural and Food Chemistry* 23: 468–474.

Sowka S, Hsieh L-S, Krebitz M *et al.* (1998) Identification and cloning of Prs a 1, a 32-kDa endochitinase and major allergen of avocado, and its expression in the yeast. *Pichia pastoris*. *Journal of Biological Chemistry* 273: 28091–28097.

Whiley AW and Schaffer B (1994) Avocado. In: Schaffer B and Anderson PC (eds) *Handbook of Environmental Physiology of Fruit Crops, Vol. II. Sub-Tropical and Tropical Crops*, pp. 3–35.Boca Raton, FL: CRC Press.

B

BACILLUS

Contents

Occurrence

R Leuschner, Central Science Laboratory, York, UK

Systematics and Ecology of *Bacillus* Species

Aerobic endospore-forming bacteria are currently assigned to four genera in the family Bacillaceae. Within this family, the genus *Bacillus* was established to include the rod-shaped bacteria that grew in the presence of air, thus distinguishing them from the strictly anaerobic *Clostridium* spp. It is possible to allocate many *Bacillus* species to one of six taxa that have distinguishable physiologies and this is generally consistent with the devision of the genus based on spore morphologies. The six groups are: *B. polymyxa* group (I), *B. subtilis* group (II), *B. brevis* group (III), *B. sphaericus* group (IV), and thermophiles (V and VI). In group I all species are facultative anaerobes and grow strongly in the absence of oxygen. Acid is produced from a variety of sugars. Endospores are ellipsoidal. The group II species (*B. subtilis* group) are phylogenetically and phenetically consistent. *B. subtilis* is an appropriate representative of the taxon that includes many common names. All these bacteria produce acids from a range of sugars and some, *B. cereus* and *B. licheniformis*, are facultative anaerobes. *B. licheniformis* grows poorly anaerobically and can use glucose only under anaerobic conditions. Although *B. subtilis* is generally regarded as an aerobe, it can grow and sporulate slowly under strict anaerobic conditions. Given glucose, with nitrite as a terminal electron acceptor, it grows strongly anaerobically. These bacteria are therefore an intermediate stage between the true facultative anaerobes of the group I strains and the strict aerobes in groups III and IV. This is reflected in their production of acid from several sugars. All these bacteria produce oval endospores that do not swell the mother cell and are generally located centrally or subterminally. Group III represents strict aerobes that generally do not produce acid from sugars, with the exception of two species. They produce ellipsoidal spores that swell the mother cell. In group IV all species produce spherical spores that may swell the mother cell and contain L-lysine or ornithine in the cell wall. All species are strictly aerobic, but some have limited ability to produce acid from sugars. Group V represents thermophilic species that grow optimally at $> 50\,^{\circ}C$. Physiologically and morphologically, they are heterogeneous, but most produce oval spores that swell the mother cell. In group VI are thermophilic, acidophilic species with membraneous ω-alicyclic fatty acid.

Occurrence in the Environment and Food

Members of the genus *Bacillus* have a ubiquitous environmental distribution. The endospore is important for the dispersion of *Bacillus* spp. Spores are readily blown about in dust and air currents and are prevalent in animal feces. Bacteria related to *B. subtilis* are commonly encountered and easily identified. The soil is the reservoir of these bacteria. From the soil, they are transferred to various associated environments, including plants and plant materials, foods, animals, and marine and fresh-water habitats. *B. cereus*, *B. licheniformis*, *B. pumilus*, and *B. subtilis* are prevalent in soils, particularly low-nutrient soils. They are also common on straw and cereals, including rice and pulses, which they presumably colonize

from wind-blown soil particles and dust. Endospores of bacilli represent a metabolically inactive survival form. They are characterized by high resistance against heat, dryness, irradiation, and other unfavorable environmental conditions. The success of the endospore for the survival of the species depends on an effective mechanism to resuscitate and enter the vegetative cell cycle to multiply. It is usually the case that spores resuscitate under good environmental conditions that allow germination and cell growth. Food of various origin and compositions can offer these conditions and this can lead to spoilage or even food poisoning due to growth of bacilli in food.

Spores and vegetative cells of *Bacillus cereus* and other food-poisoning *Bacillus* spp. occur widely in soils (10^5–10^6 CFU g^{-1}) and may be found in raw materials and processed foods which were not sterilized by heat or irradiation. They constitute a major portion of the microbial flora of raw milk and easily contaminate various dairy products, causing spoilage with their proteolytic, lipolytic, and saccharolytic activities. *B. cereus* was detected and enumerated in milk, vegetable, and meat-based products (Table 1). *B. cereus* is a problem to the dairy industry because it contaminates the udders of cows while they graze in the fields and can be introduced into milk. *B. licheniformis*, *B. pumilus* and *B. subtilis* were the most commonly isolated *Bacillus* species in bakeries and milk (Tables 2 and 3). The incidence of *B. cereus* and *B. subtilis* in various food products was investigated in the Netherlands. The total number of products was 229 samples and *B. cereus* and *B. subtilis* were isolated at contamination levels between 10^2 and 10^6 CFU g^{-1} or ml in a wide range of food products such as milk, yeast, flour, pasta products, cocoa, chocolate, bakery products, meat products, and herbs and spices. *B. cereus* was present in 48% and *B. subtilis*

Table 1 Prevalence of *Bacillus cereus* in raw and processed food products

Food product	*No. of samples examined*	*No. positive for B. cereus (%)*	*B. cereus counts g^{-1} or ml^{-1}*
Rice and oriental foods			
Raw rice	13	46	10^2–10^3
Boiled rice	32	38	10^3–10^7
	14	93	10^1–10^3
	252	10	10^2–10^5
Fried rice	14	86	10^1–10^3
	204	24	10^2–10^5
Egyptian rice dishes	172	40	10^1–10^4
Japanese noodles	200	8	10^1–10^4
Soybean curd (tofu)	257	56	10^1–10^4
	467	31	10^1–10^6
Sashimi (raw fish)	228	14	10^2–10^3
Milk and dairy products			
Raw milk	100	9	10^1–10^2
Pasteurized milk	100	35	10^1–10^3
Milk powder	120	27	10^1–10^3
Icecream	100	48	10^1–10^3
Spices	110	53	10^1–10^4
Meat and meat products			
Raw meats	133	2	10^1–10^2
	452	6	10^1–10^2

Modified from Kramer JM and Gilbert RJ (1989) *Bacillus cereus* and other *Bacillus* species. In: Doyle MP (ed.) *Foodborne Bacterial Pathogens*, pp. 21–70. New York: Marcel Dekker.

Table 2 Distribution of 170 bakery *Bacillus* isolates identified using API 50CHB

Isolate	*Bread ingredients and uncooked dough (%)*	*Swabs of processing line (%)*	*Cooked loaves and bread crumbs (%)*
B. subtilis	22.9	4.1	5.3
B. amyloliquefaciens	1.2	1.2	0
B. licheniformis	18.2	5.9	4.1
B. pumilus	11.2	2.4	2.4
B. circulans	2.9	0.6	0.6
B. megaterium	1.8	1.2	0.6
B. polymyxa	2.9	0	0.6
B. macerans	2.4	0	0.6
B. cereus (including *B. mycoides*)	0.6	1.8	0
B. stearothermophilus	0	0	0.6
Unacceptable profile	0.6	2.4	1.2

Reproduced from Thompson JM, Dodd CER and Waites WM (1993) Spoilage of bread by *Bacillus*. *International Biodeterioration and Biodegradation* 32: 55–66, with permission.

Table 3 Mesophilic *Bacillus* species isolated from milk samples

Bacillus *species*	*No. of samples where a species was isolated (total number of samples: n = 250)*
B. licheniformis	131
B. pumilus	156
B. subtilis	127
B. lentus	52
B. pantothenticus	33
B. amyloliquefaciens	25
B. stearothermophilus	19
B. mycoides	22
B. cereus	16
B. circulans	15
B. firmus	10
B. megaterium	27
B. shaericus	6
B. macerans	3
B. laterosporus	2
B. polymyxa	5
Unknown	120

Reproduced from Sutherland AD and Murdoch R (1994) Seasonal occurrence of psychrotrophic *Bacillus* species in raw-milk, and studies on the interactions with mesophilic *Bacillus* sp.

in 25% of all samples examined. In spoiled bread that developed symptoms of ropiness after 2 days' storage at ambient summer temperatures, identified species were *B. subtilis* (70%), *B. licheniformis* (24%), *B. pumilus* (2%), and *B. cereus* (2%).

Food Spoilage and Poisoning

The occurrence of *Bacillus* species in raw materials used for food processing is generally below the infectious dosis required to affect negatively the well-being of consumers: this is normally above 10^5 cells g^{-1} or ml^{-1} food (Table 4). This indicates that *Bacillus* cells or spores must have the opportunity to multiply in the food environment to enable them to cause food spoilage and poisoning.

B. cereus food poisoning is principally associated with the storage of cooked foods at temperatures and times that allow growth and result in cell numbers above 10^6 CFU g^{-1} or ml^{-1}. Rapid cooling of cooked food below 10 °C is generally an effective control measure. Low-pH foods (< pH 5.0) and dry foods will not support the growth of *B. cereus*, although many dried foods will be contaminated with spores of this microorganism. Spores are more hydrophobic than spores of other *Bacillus* spp., which enables them to adhere particularly well to many types of surfaces. This makes them difficult to remove during cleaning and a difficult target for disinfection.

B. cereus can cause infections and intoxications. In addition to foodborne diseases, *B. cereus* causes septicemia, meningitis, and ocular infections. It can cause two types of food poisoning: the emetic disease, characterized by a short incubation period (1–5 h) displaying symptoms such as nausea, vomiting, and stomach cramps, and the diarrheal disease, characterized by an incubation period of 8–16 h with symptoms including abdominal pain, watery diarrhea, and rectal tenesmus. Usually both types of food poisoning are relatively mild and last for less than 24 h.

B. subtilis was reported to be the cause of illness in meat dishes with elements of vegetables, seafood and rice, bread and pastry products, sandwiches, and pizza. The infectious dose was > 10^5 CFU g^{-1} with an incubation period of 10 min to 14 h and a duration of illness of 2–8 h. Symptoms were vomiting (80%), diarrhea (49%), abdominal pain/cramps (27%), and nausea, headaches, flushing, and sweating. Bacilli are capable of causing food spoilage in bread, and this is known as ropiness. *B. subtilis* has been described to be the main cause of the bread spoilage 'ropiness.' Ropiness is the most important spoilage of bread after moldiness and occurs particularly during the summer when the climatic conditions favor growth of the bacteria. This involves mainly *B. subtilis*. *Bacillus* counts in white and wholemeal wheat loaves produced without preservatives or sourdough were consistently 10^6 CFU g^{-1} after 2 days of storage at ambient summer temperatures.

B. licheniformis was identified in meat dishes with elements of vegetables, bread, and pastry products, and chicken as a cause of food poisoning. The infectious dose was > 10^6 CFU g^{-1}, the incubation period 2–14 h and the duration of illness 6–24 h. Symptoms were vomiting (54%), diarrhea (92%), and abdominal pain/cramps (46%). *B. pumilus* in meat products, sandwiches, and canned tomato juice was found to be responsible for food poisoning. The infectious dose was > 10^6 CFU g^{-1} and the incubation period 15 min to 11 h. Symptoms were vomiting and diarrhea.

B. stearothermophilus and *B. coagulans* were reported to be the cause of flat sour spoilage in evaporated milk.

Growth Characteristics and Resistance of Bacilli

Knowledge about the resistance and growth characteristics of *Bacillus* species is important to enable food manufacturers and consumers to avoid handling practice that enable spores to germinate and vegetative cells to multiply in a food.

Psychrotrophic strains of *B. cereus* were able to grow and produce diarrheal toxin at temperatures down to 4 °C. Most mesophilic strains are able to grow in low-acidic foods at temperatures between 15 and 55 °C, with an optimum range of 30–40 °C. Growth can generally occur between pH 5.0 and 8.8,

Table 4 Infective or intoxication dosis food poisoning caused by *Bacillus* species

Bacillus *species*	*Intoxication dosis*	*Infective dosis*
B. cereus (emetic toxin)	12–32 μg kg^{-1} toxin (*Suncus murinus*: a small monkey)	10^5–10^8 cells g^{-1} food
B. cereus (diarrheal disease)		Usual 10^5 g^{-1} or ml^{-1}
B. subtilis		> 10^5
B. licheniformis		> 10^6
B. pumilus		> 10^6

Modified from Granum PE and Baird-Parker TC (2000) *Bacillus* species. In: Lund BM, Baird-Parker TC and Gould GW (eds) The *Microbiological Safety and Quality of Food*, vol. 2, pp. 1029–1039. Maryland: Aspen.

Table 5 Decimal reduction (D) values (min) at 95 °C of *Bacillus cereus* spores for various pH and organic acid types

pH	*Citric acid*	*Lactic acid*	*Acetic acid*	*Malic acid*
6.5	1.03	1.09	1.21	0.96
6	0.95	0.85	0.99	0.96
5.5	0.81	0.82	0.82	0.77
5	0.74	0.66	0.67	0.69
4.5	0.64	0.54	0.55	0.62
4	0.62	0.40	ND	0.59

ND, not determined.
Modified from Leguerinel I and Mafart P (2001) Modelling the influence of pH organic acid types on thermal inactivation of *Bacillus cereus* spores. *International Journal of Food Microbiology* 63: 29–34.

with an optimum between 6.0 and 7.0. Spores are variable in their heat resistance, which is normally moderate, but can be increased by fat components. The emetic toxin is very heat-resistant and can survive heating at 126 °C for 90 min. The diarrheal toxin is heat-sensitive and is inactivated at 56 °C after 5 min. Growth of *B. cereus* was found at water activities of 0.93 depending on the acidulant and humectant. Other food-poisoning *Bacillus* spp. have similar growth and resistance characteristics to *B. cereus*. *B. licheniformis* and *B. subtilis* were not found to grow at pH 4–4.2; however, growth in tomato juice at a pH of 4.4 in the presence of oxygen was reported. Spores of thermophilic bacilli such as *B. stearothermophilus* are far more heat-resistant and are used to test sterilization effectiveness. Heat resistance of bacilli spores is dependent on the temperature during spore formation and on the pH of the food. Further dependence was observed on various acidulants (Table 5).

Sterilization will inactivate all *Bacillus* spores. This is however also imposing a harsh treatment towards other valuable food ingredients. The development of novel and milder food preservation technologies, such as high-pressure and electromagnetic fields, have proven to be effective in *Bacillus* spore inactivation. Pressure treatments of 400 MPa for 25 min at 30 °C resulted in a 0.45 log inactivation of *B. cereus* spores. Pressure treatments at lower temperatures were less effective. Spores of *B. subtilis* proved to be more resistant to ultraviolet B range (280–330 nm) than spores of *B. cereus*. An exposure of 30 min reduced the viability of spores of *B. cereus* by 50% and that of *B. subtilis* spores by 10%. Levels of resistance to ultraviolet B of spores from different *Bacillus* species appear to be related to the quantity and quality of small acid-soluble proteins and to activities of DNA repair systems. The novel, milder nonthermal preservation technologies need further evaluation with particular emphasis on spore inactivation mechanisms and synergistic effects by combination of various techniques in food systems.

See also: ***Bacillus***: Detection; Food Poisoning; **Food Poisoning**: Classification; **Spoilage**: Bacterial Spoilage

Further Reading

Claus D and Berkeley RCW (1986) Genus *Bacillus*. In: Sneath PHA (ed.) *The Bergey's Manual of Systematic Bacteriology*, vol. 2, pp. 1105–1139. Baltimore: Williams & Wilkins.

Granum PE and Baird-Parker TC (2000) *Bacillus* species. In: Lund BM, Baird-Parker TC and Gould GW (eds) *The Microbiological Safety and Quality of Food*, vol. 2, pp. 1029–1039. Maryland: Aspen.

International Commission on Microbiological Specifications for Foods (1996) *Bacillus cereus*. In: Roberts TA, Baird-Parker AC and Tompkin RB (eds) *Microbiological Characteristics of Food Pathogens*, pp. 335–344. London: Blackie Academic and Professional.

Kramer JM and Gilbert RJ (1989) *Bacillus cereus* and other *Bacillus* species. In: Doyle MP (ed.) *Foodborne Bacterial Pathogens*, pp. 21–70. New York: Marcel Dekker.

Priest FG (1993) Systematics and ecology of *Bacillus*. In: Sonenshein AL, Hoch JA and Losick R (eds) Bacillus Subtilis *and Other Gram Positive Bacteria*, pp. 3–16. Washington, DC: American Society of Microbiology.

Detection

L A Shelef, Wayne State University, Detroit, MI, USA

Background

The genus *Bacillus* is large, comprising more than 60 species that are mostly saprophytes, widely distributed in nature, spreading from soil to water, plants, and animals. The genus shows a great diversity of strains and species. The organisms are Gram-positive or Gram-variable spore-forming bacilli, mostly catalase-positive, that may be motile by peritrichous flagella. Most are mesophiles, but some are psychrothrophs and thermophiles. *Bacillus* contains strict aerobes (e.g., *B. megaterium*), as well as facultative anaerobes (e.g., *B. cereus, B. licheniformis*). The vegetative cells are rods ranging from $0.5 \times 1.2\,\mu m$ to $2.5 \times 10\,\mu m$, and the endospores are in the central or paracentral, subterminal or terminal position. Survival of the organism results from the resistance of the spores to adverse conditions.

Of the identification schemes proposed, that of Gordon and co-workers divides the genus *Bacillus* into three groups, according to cellular morphology and physiological properties. Group 1 consists of

Gram-positive species that form ellipsoidal or cylindrical spores that do not appreciably distend the sporangia. This group is further divided according to the vegetative cell dimensions and presence of lipid globules in the protoplasm of the species. The 'large-celled' group 1 *Bacillus* species have a cell width greater than 0.9 μm and include *B. megaterium, B. cereus, B. cereus* var. *mycoides, B. thuringiensis*, and *B. anthracis*. The 'small-celled' species have a cell diameter < 0.9 μm, and lipid globules are not formed. They include *B. subtilis, B. pumilus, B. licheniformis*, and *B. coagulans*. Organisms of groups 2 and 3 are characterized by swollen sporangia and either ellipsoidal (group 2) or spherical (group 3) spores. Studies of DNA composition and other studies suggest that the genus *Bacillus* consists of some bacterial families that will necessitate further classification of the genus.

Most *Bacillus* species are harmless saprophytes characterized by their proteolytic and saccharolytic properties and are rarely associated with human or animal disease. However, several species are proven or suspect pathogens. Of these, *B. cereus* is a recognized food pathogen, and other species, particularly those of the *B. subtilis* group (*B. subtilis, B. licheniformis*, and *B. pumilus*), have been implicated in foodborne illnesses. *B. brevis*, also identified as a causative agent of foodborne disease, has been split and transferred to the new genus *Brevibacillus*. *B. anthracis* is the agent of anthrax. Five *Bacillus* species, *B. thuringiensis, B. popilliae, B. sphaericus, B. larvae*, and *B. lentimorbus*, are insect pathogens, the first three being used in commercial insecticides for the control of crop insects.

Of the *Bacillus* species associated with foods, detection and identification methods have focused on *B. cereus* as a recognized food pathogen. Spores and vegetative forms of *B. cereus* are frequent inhabitants of soils, sediments, dust, natural water sources, vegetation, and many foods such as grains, dairy products, dried foods, spices, meats, and vegetables. Examination of various raw and processed foods shows contamination levels ranging from 10^1 to 10^6–10^7 cells per gram or milliliter. It is estimated that *B. cereus* accounts for ~2–3% of reported foodborne outbreaks in the USA. In the UK, a single outbreak was reported during the period of October–December, 2000. In this outbreak, 30 people became ill after consuming contaminated soup with cream in a canteen, and a statistical association was established between consumption of the food and the cause.

Detection of *B. cereus* and its toxins and differentiation between the species within the *B. cereus* group are presented in this chapter. Following a brief review of the properties of the organism and its toxins, traditional and novel detection of the organism and its toxins are discussed.

Properties of *B. cereus*

Cells of *B. cereus* are large, Gram-positive rods that are motile by means of peritrichous flagella. They are typically 1.0–1.2 μm in diameter and 3.0–5.0 μm in length. Endospores are in either the central or paracentral position without distention of the sporangium. The organism sporulates freely on many media under aerated conditions. The optimum growth temperature is 28–35 °C, with a minimum of 5 °C and a maximum of 48 °C, and the pH for growth is 4.9–9.3. *B. cereus* strains are able to utilize glucose, fructose, and trehalose. Sucrose, salicin, maltose, mannose, glycerol *m*-inositol, and lactose are utilized by some, but not all, members of the species. Pentoses, mannitol, and many of the sugar alcohols are not metabolized. Urease is produced by few strains. Most of the strains produce acetylmethylcarbinol (Voges–Proskauer-positive), reduce nitrate to nitrite, hydrolyze soluble starch, tyrosine, gelatin and casein, and produce a neutral metalloprotease. *B. cereus* and *B. thuringiensis* produce a broad spectrum of β-lactamase and are, therefore, resistant to penicillin, ampicillin, and cephalosporins. Their ability to grow on agar containing penicillin aids in differentiating them from *B. anthracis*. Certain strains have been reported to synthesize a dehydropeptide reductase that catalyzes cleavage of the bacteriocins nisin and subtilin. In contrast to *B. megaterium, B. cereus* strains are not sensitive to lysozyme.

The organism elaborates a number of toxins with distinct diarrheal and emetic syndromes. The diarrheal syndrome is caused by bacterial growth and toxin formation in the small intestine, whereas the emetic syndrome is caused by preformed toxin resulting from growth of the pathogen in the food. Of the enterotoxins responsible for the diarrheal syndrome, hemolysin, designated BL, comprises three heat-labile peptides, B (37.8 kDa), L_1 (38.5 kDa), and L_2 (43.2 kDa), and the genes that encode for these components have been cloned. The L_1 and L_2 peptides appear to be responsible for binding to erythrocytes and the B component for lysing them. Cereolysin (56 kDa), a cytolytic protein responsible for hemolysis, consists of phospholipases that hydrolyze phosphatidylinositol, some of which are metalloenzymes requiring divalent cations for activity, and sphingomyelinase, also a metalloenzyme that has hemolysin-like activity, but shows no phospholipase activity. The nonhemolytic enterotoxin is heat-labile, comprising three proteins (39, 45, and 105 kDa). Cereulide is a putative emetic toxin. It has been

isolated and identified as a small (1.2 kDa), heat-stable dodecadepsipeptide, structurally related to the ionophore valinomycin, producing a vacuole response in HEp-2 cells. Psychrophilic strains of *B. cereus* capable of producing toxins at refrigerator temperatures have been reported.

Isolation and Identification of *B. cereus*

Standard Agar Media

The characteristics of *B. cereus* and of the culturally similar *B. mycoides*, *B. thuringiensis*, and *B. anthracis* are summarized in Table 1. Most procedures for the isolation, identification, and enumeration of *B. cereus* involve direct plating techniques. Media developed for the identification of the organism provide conditions for the organism to exhibit fermentation properties, hemolysin production, lecithinase activity, and morphological characteristics. Selective agents to inhibit competitive microorganisms, such as antibiotics (most often 10 μg of polymyxin per milliliter), are normally incorporated into the media.

MYP agar Mannitol–egg yolk-polymyxin agar (MYP) was the first selective medium, introduced in 1967 by Mossel and coworkers. It relies on lecithin hydrolysis (the egg-yolk reaction), negative mannitol reaction for differentiation of *B. cereus* from other commonly occurring *Bacillus* species, and polymyxin B. The agar plate is surface-inoculated, and typical colonies of *B. cereus* after 20–24 h of incubation at 30–32 °C have a dry, flat, 'ground-glass' appearance, translucent to creamy white with a violet–red background surrounded by a readily visible zone of egg-yolk precipitate (Figure 1). Less commonly, colonies may be large (5 to over 10 mm in diameter), amorphous, spreading, and highly irregular. MYP remains to date a popular choice for the isolation of *B. cereus* and is available commercially. Plates can be stored for up to 7 days at 4 °C.

Table 1 Characteristics of *Bacillus cereus* and other members of the *B. cereus* group

Reaction	B. cereus	B. mycoides	B. thuringiensis	B. anthracis
Egg-yolk reaction	+[a]	+	+	(+)[b]
Acid from mannitol	−	−	−	−
Catalase	+	+	+	+
Gram reaction	+	+	+	+
Motility	±[c]	−[d]	±	−
Hemolysis (sheep RBC)	+	(+)	(+)	−
Rhizoid growth	−	+	−	−
Production of toxin crystals	−	−	+	−
Anaerobic utilization of glucose	+	+	+	+
Nitrate reduction	+	+	+	+
VP reaction	+	+	+	+
Tyrosine utilization	+	(+)	+	(+)
Resistance to lysozyme	+	+	+	+

[a] +, 90–100% of strains are positive.
[b] (+), usually weakly positive.
[c] ±, 50–90% of strains are positive.
[d] −, < 10% of strains are positive.

KG agar In 1971, Kim and Goepfert developed the Kim–Goepfert (KG) medium with a similar sensitivity and selectivity to MYP. It contains no carbohydrates and low levels of peptone to promote free spore formation within 24 h of incubation at 37 °C. Surface colonies of *B. cereus* after overnight incubation at 37 °C are flat, dry, and of a rough appearance, similar to the appearance of the colonies on MYP agar. The composition of KG agar allows direct confirmation of lecithinase-producing organisms by microscopic examination, differentiation of *B. cereus* from *B. thuringiensis* based on production of parasporal crystals of δ-endotoxin by the latter species during sporogenesis, and the use of fluorescent-labeled antibodies to spore-coat antigens as a rapid and specific serological test for confirmation of identity. *Bacillus* spp. that may produce lecithinase, such as *B. polymyxa* and *B. laterosporus*, are unable to do so under the nutritionally poor KG medium.

PEMBA medium Polymyxin–pyruvate–egg yolk–mannitol–bromothymol blue agar (PEMBA), developed by Holbrook and Anderson in the early 1980s, also relies on the mannitol-negative, lecithinase-positive properties of the organism. The medium contains pyruvate to reduce the tendency of *B. cereus* to form rhizoid colonies, improve the egg yolk reaction, and enhance sporulation. After 18–24 h of incubation at 37 °C, *B. cereus* strains form flat, crenate to rhizoid, turquoise–peacock blue colonies, 2–5 mm in diameter, with a 'ground-glass' surface appearance. After a further 24 h of incubation at ambient temperature, all colonies turn peacock blue. *B. cereus* is differentiated microscopically from other mannitol-negative species by microscopic examination of the stained cells for the presence of lipid granules and spores in the cytoplasm.

PEMPA medium Polymyxin–Pyruvate–egg yolk–mannitol–bromocresol purple agar (PEMPA), a modified agar, gives comparable results to PEMBA after shorter incubation times (18–24 h).

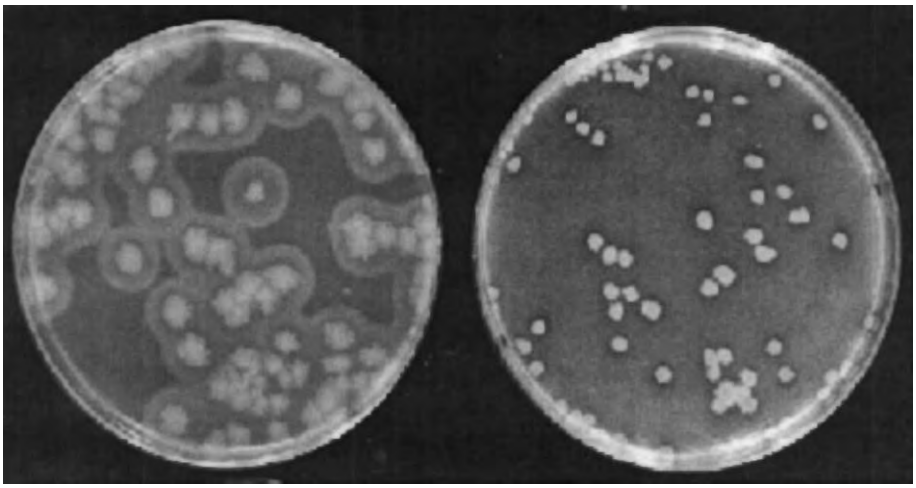

Figure 1 Typical colonies of *B. cereus* on plates of MYP agar (left) and BCM® agar (right).

Blood agar On 5% horse blood agar, after 24 h at 35–37 °C, *B. cereus* colonies are 3–8 mm in diameter, flat to low convex, gray–white to gray–green, with a matt 'ground glass' surface. Growth is also associated with a strong 'mouse-like' odor. Hemolytic reactions on the medium vary from partial lysis of erythrocytes to complete β-hemolysis. Small colonies, 2–3 mm in diameter, form during anaerobic incubation. The irregular, β-hemolytic colonies resemble those of *Clostridium perfringens* and require further identification.

Of the plating media, MYP agar is the official medium in the USA (Association of Official Analytical Chemists), Canada, France, and several other countries. PEMBA is widely used in the UK. These plating media do not differentiate between *B. cereus* and *B. thuringiensis* organisms (Table 1). In general, the numbers of *B. cereus* recovered from foods are sufficiently high for direct plating, and preenrichment is unnecessary. However, preenrichment may be required when low viable counts or damaged cells are present. The most probable number (MPN) technique, using trypticase soy-polymyxin B broth and incubation for 48 h at 30 °C, is recommended for estimation of *B. cereus* numbers in foods expected to contain less than 10 cells per gram. Tubes showing growth are streaked on to MYP agar. On egg-yolk-containing agar plates, zones of egg-yolk precipitation from individual colonies tend to coalesce if high levels of the organisms are present (Figure 1), and enumeration of colonies may become difficult. Further dilutions of samples to three to 30 colonies are therefore recommended in quantitative analyses.

Chromogenic Agar Media

Chromogenic substrates that utilize specific enzymatic activities are increasingly used for rapid detection and identification of microorganisms. Their incorporation into selective agar media facilitates identification of colonies, improves the accuracy of the test and reduces the need for isolation of pure cultures and confirmation. The enzyme inositol-specific phospholipase C (PI-PLC) was first reported in culture filtrates of *B. cereus* in 1965, and a similar enzyme was later identified in *B. thuringiensis*. On a chromogenic, selective, and differential plating medium that detects the enzyme phosphatidylinositol-specific phospholipase C (PI-PLC) in *B. cereus* and *B. thuringiensis* (BCM® *Bacillus cereus* and *Bacillus thuringiensis*, Biosynth, Staad, Switzerland), these organisms form large, 2–7-mm, turquoise, flat colonies with or without turquoise halos on the medium. Other *Bacillus* spp. form white, 1–2-mm colonies or are inhibited; some *Listeria* strains form colonies less than 1 mm in diameter, white or turquoise on the medium. Comparison with MYP agar has shown a significantly higher inhibition of Gram-positive non-*B. cereus*/*B. thuringiensis* organisms that interfere with enumeration of small numbers of *B. cereus* in heavily contaminated samples. Typical colonies of *B. cereus* on the chromogenic agar plate are shown in Figure 1. The shelf stability of the chromogenic agar plates (3 months at 2–5 °C) is superior to that of MYP plates. As with other procedures, the chromogenic medium does not differentiate between *B. cereus* and *B. thuringiensis* organisms.

Confirmatory Procedures

The specific confirmation tests that distinguish *B. cereus* from the other members of the *B. cereus* group include motility, toxin crystal formation, rhizoid growth, and hemolysin activity (Table 1).

Motility is determined by stabbing the center of a tube of semisolid medium and incubating for 18 h at 30 °C. The motile organisms (most *B. cereus* and *B. thuringiensis*) diffuse out from the stab, forming an opaque growth pattern. The only established difference between *B. cereus* and *B. thuringiensis* strains is the presence of genes encoding for the insecticidal toxins, usually present on plasmids. The ability of *B. thuringiensis* to form toxin crystals is determined by inoculating nutrient agar slants with the overnight culture and incubation for 2–3 days at room temperature. Toxin crystals appear on a smear of a microscope slide stained with 0.5% basic fuchsin as dark-staining diamond-shaped objects after their release from the sporangium upon lysis. The test is inconclusive if the release is not observed. Rhizoid growth, characteristic of *B. mycoides*, is observed by the production of hair- or root-like structures projecting from an agar plate inoculated with the organism and incubated for 48–72 h. Rhizoid growth may be lost by the organisms. All three species show strong hemolytic activity on trypticase soy–sheep blood agar plates inoculated with an overnight culture and incubated at 35 °C for 24 h.

A rapid confirmatory staining technique for the identification of *B. cereus* developed by Holbrook and Anderson uses colonies grown on PEMBA medium and a combined spore and intracellular lipid stain. Only the *B. cereus* group bacilli produce the intracellular lipid globules.

The US Food and Drug Administration confirms presumptive *B. cereus* colonies from MYP agar plates by anaerobic production of acid from glucose, reduction of nitrate to nitrite, production of acetylmethylcarbinol, decomposition of L-tyrosine, and growth in the presence of 0.001% (wt/vol) lysozyme. Additional tests for rhizoid growth, hemolysis of sheep erythrocytes, motility, and production of parasporal δ-endotoxin crystals enable differentiation of *B. mycoides, B. thuringiensis*, and *B. anthracis*. The Central Public Health Laboratory, London, UK confirms presumptive *B. cereus* by testing motility, hemolysis, rhizoid growth, susceptibility to γ-phage (*B. cereus*-negative; *B. anthracis*-positive), and fermentation of ammonium salt-based glucose, but not of mannitol, arabinose, or xylose, after 5 days of incubation at 36 °C.

Serological Tests

The serological relationships within the genus *Bacillus* are not as well defined as for other genera, and the development of serological procedures for use as diagnostic aids has met with limited success. Studies with spore antigens have shown cross-reactions with other *Bacillus* species. The multiplicity of spore surface antigens in a single strain of *B. cereus* has also been demonstrated. Some of the problems in the use of spore antigens stem from contamination of their preparations with vegetative cell debris and formation of antivegetative cell immunoglobulins, germination of spores during serum production, which also results in the formation of antivegetative cell immunoglobulins, and *in-vitro* spore germination during immunoassay.

Forty-two flagellar (H) and 13 somatic (O) antigens have been identified. A serological typing scheme for *B. cereus* strains based on flagellar antigens was developed by Gilbert and coworkers and is used routinely by the Food Hygiene Laboratory, Central Public Health Laboratory, Colindale, London, UK for investigations of food-poisoning outbreaks and for serotyping strains from nongastrointestinal infections. It is composed of 42 H antigens raised against prototype strains of diverse origin. Twenty-three of the 42 H serotypes are linked to illness, and serotype H1 is dominant and contains both emetic and diarrheal strains. A system similar to that used in the UK was also developed by the Tokyo Metropolitan Research laboratory of Public Health in Japan.

B. thuringiensis strains have been classified on the basis of their flagellar antigens. A serological typing scheme for this species has been developed, and 14 serotypes have been recognized. There is evidence that certain strains of *B. cereus* and *B. thuringiensis* share common antigens.

Rapid Identification Methods

The term *rapid methods* refers to a variety of tests, which can shorten the analysis time and reduce labor (i.e., by automation). Some eliminate the need for confirmation, whereas others require enrichment prior to testing, which limits the assay speed but provides more reliable identification of the target organism. All require pure culture isolates of the bacteria. Rapid methods applicable to *B. cereus* are discussed below.

Miniaturized biochemical kits to identify *Bacillus* species have been developed to produce standardized test media that improve reproducibility, have a long shelf-life, are rapid and convenient to use, economize the use of media, and are inexpensive. The API 20E and 50 CHB strips (bioMérieux, Marcy l'Etoile, France; Hazelwood, MO, USA), are ready-to-use microtube systems that contain dehydrated media for standard biochemical tests. The API 50 CHB for the identification of the genus *Bacillus* tests the fermentation of 49 carbohydrates, and the phenotypic profile obtained from these tests is used to identify the

strain. The API 20E strips are used for the identification in the Vitek automated identification system, which eliminates the need for manual identification. Incubation for 24–48 h at 30 or 37 °C (mesophiles) is required before reading the results. Other manual and automated systems for identification of *Bacillus* based on biochemical properties are available. MicroLog™ microbial carbon-source utilization (Biolog Inc., Hayward, CA) generates an identity profile based on the ability of the organism to oxidize different carbon sources based on changes in the redox potential following respiration of the viable bacterial cells. Oxidation is determined colorimetrically with a tetrazolium-based compound, and the color change data measured automatically by a spectrophotometer are fed into a computer. Bacteria are identified by comparison with a reference data base. The BBL Crystal™ Gram-positive Identification System (Beckton Dickinson, Sparks, MD) is another example of miniaturized identification method for Gram-positive aerobes employing modified conventional, fluorogenic, and chromogenic substrates.

Another approach for the identification of the organism is based on its fatty acid profile. Analysis of the cellular fatty acid composition is carried out by gas chromatography following extraction of pure cultures and injection into a gas chromatograph (Microbial Identification System, MIS, Microbial-ID, Newark, DE). The cellular fatty acid profile is analyzed by computer and compared with database values to provide the best-match identification of the organism.

Other diagnostic schemes include polyacrylamide gel electrophoresis analysis, pyrolysis mass spectrometry, and Fourier-transform infrared spectroscopy. Classification and identification based on differences in rRNA are available commercially for food pathogens including *B. cereus*. Ribotyping (RiboPrinter® Microbial Characterization System, DuPont Qualicon, Wilmington, DE) generates precise fingerprints that can be used to classify isolates at the species level and trace them to the contaminated source. Steps of the identification, from cell lysis to data capture and database comparisons, are automated. All of the above identification methods require large updated databases. DNA sequence data for the design of specific primers for *B. cereus* are being investigated. The design of primers within the coupled sequences for phospholipase C and sphingomyelinase of *B. cereus* has been reported, and PCR tests with these primers are positive only with isolates of the *B. cereus* group. Moreover, positive results have also been observed with an egg-yolk-negative *B. cereus* isolate, suggesting that the primers could be specific for atypical *B. cereus*.

Phage typing of *B. cereus* strains is another epidemiological tool in outbreaks of *B. cereus* food poisoning. Phages are used in routine test dilution, and drops are deposited on bacterial lawns on nutrient agar plates that are incubated overnight at 32 °C and examined for lysis. A typing scheme using a set of 12 phages selected on the basis of host range, stability, and reproducibility of reaction showed that most isolates of the pathogen were typeable. A good correlation (80–100%) was observed between phage types of strains isolated from specimens of suspected foods and from symptomatic patients. Most *B. thuringiensis* strains were also typeable by the same set of phages.

Detection of *B. cereus* Toxins

In-vivo and *in-vitro* methods for the detection of *B. cereus* diarrheal and emetic toxins are summarized in Table 2. *In-vivo* methods for the detection of *B. cereus* diarrheal enterotoxins include the rabbit or guinea-pig ileal loop test, dermonecrotic test in guinea-pig, mouse lethality test, vascular permeability test, and Rhesus monkey feeding trials. In the ileal loop assay, the lower portion of the intestine of the animal is surgically exposed to allow ligation, usually of the ileal regions, thereby testing different samples in a single animal. The sample is injected into a ligated section, and fluid accumulation is monitored with time and compared with control sections. Animal feeding trials (Rhesus monkey and *Suncus marinus*) and lethality tests (mouse and *Suncus marinus*) are used to detect the presence of the emetic toxin. Animals (a set of six) are observed for about 5 h after samples have been introduced

Table 2 Detection methods of *B. cereus* toxins[a]

In-vivo *methods*	In-vitro *methods*
Diarrheal syndrome	
Rabbit or guinea-pig ligated ileal loop	Gel-diffusion assays
Guinea-pig dermonecrotic test	Cell cytotoxicity (hemolysin BL)
Mouse lethality test	BCET-RPLA kit (L_2 of BL)
Vascular permeability test	BDE-VIA kit (nonhemolytic toxin)
Rhesus monkey feeding trials	
Emetic syndrome	
Feeding trials (Rhesus monkey and *Suncus marinus*)	Cytotoxicity (proliferation of HEp-2 cell line)
Lethality tests (mouse and *Suncus marinus*)	

[a]See text for details.

using a stomach tube. An emetic response in two of the six animals is considered positive for the toxin.

The disadvantages of bioassays are the time to complete the assay, the need for special animal facilities, the cost, as well as the concern about the use of experimental animals in research. Alternatives to the *in-vivo* tests are *in-vitro* assay methods that include diffusion techniques, enzyme-linked immunosorbent assays (ELISA), and cell cytotoxicity tests. The gel diffusion assay detects hemolysin BL, producing a discontinuous hemolysis pattern on blood agar plates. The microslide immunodiffusion assay and the fluorescent immunodot assay detect enterotoxin by line of identity and fluorescence, respectively.

Two immunoassay kits are available commercially for the detection of the *B. cereus* diarrheal toxins. These include the *B. cereus* enterotoxin-reversed passive latex agglutination (BCET-RPLA) assay (Oxoid, Unipath, Basingstoke, UK), and the *Bacillus* diarrheal enterotoxin visual immunoassay (BDE-VIA) (TECRA Diagnostics, Batley, UK). The RPLA procedure uses latex particles to amplify the antibody: antigen reaction. The antibody detects the L_2 component of hemolysin BL, and can provide a semiquantitative measure of the enterotoxin in foods. In the TECRA immunoassay, a sandwich ELISA is used, in which the antibody is absorbed on the solid phase, and the enterotoxin is added. The colorimetric reaction detects the 45-kDa and 105-kDa proteins of the nonhemolytic enterotoxin and other proteins. The presence of preformed toxin is detected in 4 h, and production of the toxin in samples containing enterotoxigenic *Bacillus* spp. is detected after overnight enrichment. Since these two kits detect components in the different enterotoxins, isolates should be tested by both methods. Many *B. cereus* strains have been shown to react with both the Oxoid and the TECRA detection kits, suggesting that they are able to produce both enterotoxin complexes.

B. thuringiensis, B. subtilis, B. licheniformis, and *B. pumilus* strains have been involved in foodborne illnesses. Although the nature of their toxins is not well defined, reaction with antibodies to *B. cereus* enterotoxins in both the Oxoid and the TECRA detection kits has been demonstrated for some of them.

Cell cytotoxicity techniques have been used for both the diarrheal toxin and the emetic toxin. Cell lines used include HeLa, HEp-2, Vero, and others, and cellular responses in the presence of the toxin range from morphological changes (subjective), to more specific, e.g., metabolic status of the cells and detection of lactate dehydrogenase release from damaged cells.

Taxonomic Relationship of *B. cereus*, *B. thuringiensis*, and *B. anthracis*

The lower end of the GC range of the genus *Bacillus* (32–38%) is occupied by *B. cereus* and the closely related species *B. anthracis*, *B. mycoides*, and *B. thuringiensis*. It was suggested in 1952 that there are no consistent phenotypic properties that differentiate these species and that these three species be designated varieties of *B. cereus*. However, this has not been accepted for *B. anthracis* and *B. thuringiensis* because of their pathogenic qualities. The high homology between DNA from *B. anthracis*, *B. thuringiensis*, and *B. cereus* suggests that organisms in these taxa should have the same name. Recently, researchers used multilocus enzyme electrophoresis and sequence analysis of nine chromosomal genes to provide further evidence that *B. anthracis* should be considered a lineage of *B. cereus*. Evidence of the close taxonomic relationship of *B. cereus*, *B. thuringiensis*, and *B. anthracis* was indirectly obtained from serological studies that showed extensive cross-agglutination between the spores of the three species.

The only established difference between *B. cereus* and *B. thuringiensis* strains is the presence of genes encoding for the insecticidal toxins, usually present on plasmids. None of the detection methods discussed above distinguishes between the two species.

See also: ***Bacillus***: Occurrence; Food Poisoning

Further Reading

Ahmed R, Sankar-Mistry P, Jackson S, Ackerman H-W and Kasatiya SS (1995) *Bacillus cereus* phage typing as an epidemiological tool in outbreaks of food poisoning. *Journal of Clinical Microbiology* 33: 636–640.

Berkeley RCW and Goodfellow M (1981) *The Aerobic Endospore-Forming Bacteria: Classification and Identification*. London: Academic Press, Society for General Microbiology.

Claus D and Berkeley RCW (1986) Genus *Bacillus* Cohn 1872, 174. In: Sneath PHA, Mair NS, Sharpe ME and Holt JG (eds) *Bergey's Manual of Systemic Bacteriology*, vol. 2, pp. 1105–1139. Baltimore, MD: Williams & Wilkins.

Granum PE and Lund T (1997) Mini review: *Bacillus cereus* and its food poisoning toxins. *FEMS Microbiology Letters* 157: 223–228.

International Commission on Microbiological Specifications for Foods (1996) *Bacillus cereus*. In: *Microorganisms in Foods 5. Characteristics of Microbial pathogens*, pp. 20–35. London: Blackie Academic & Professional.

Kramer JM and Gilbert RJ (1989) *Bacillus cereus* and other *Bacillus* species. In: Doyle MP (ed.) *Foodborne Bacterial Pathogens*, pp. 21–70. New York: Marcel Dekker.

Kramer JM and Gilbert RJ (1992) *Bacillus cereus* gastroenteritis. In: *Food Poisoning Handbook of Natural Toxins*, vol. 7, pp. 119–153. New York: Marcel Dekker.

Logan NA and Turnbull PCB (1999) *Bacillus* and recently derived genera. In: *Manual of Clinical Microbiology*, 7th edn. pp. 357–369. Washington, DC: American Society for Microbiology.

Rhodehamel EJ and Harmon SM (1995) *Bacillus cereus*. In: *Bacteriological Analytical Manual*, 8th edn. Gaithersburg, MD: AOAC International.

Schultz FJ and Smith JL (1994) *Bacillus*: Recent advances in *Bacillus cereus* food poisoning research. In: Hui YH, Gorham JR, Murrell KD and Cliver DO (eds) *Foodborne Disease Handbook, Diseases Caused by Bacteria*, vol. 1. New York: Marcel Dekker.

Tortorello ML and Gendel SM (1997) *Food Microbiological Analysis: New Technologies. IFT Basic Symposium Series*. New York: Marcel Dekker.

van Netten P and Kramer JM (1992) Media for the detection and enumeration of *Bacillus cereus* in foods: a review. *International Journal of Food Microbiology* 17: 85–99.

Table 1 Different *Bacillus* species involved in food poisoning[a]

Species	
B. cereus	Frequently involved in food poisoning
B. anthracis	Has caused food poisoning (see text)
B. thuringiensis	Involved in food poisoning
B. mycoides	Might have been involved in food poisoning
B. weihenstephanensis	Might have been involved in food poisoning
B. pseudomycoides	Might have been involved in food poisoning
B. brevis	Involved in food poisoning
B. circulans	Probably not involved, but has been shown to produce *B. cereus*-type enterotoxin
B. licheniformis	Involved in food poisoning
B. pumilus	Involved in food poisoning
B. sphaericus	Involved in food poisoning
B. subtilis	Involved in food poisoning

[a]The first six species belong to the *B. cereus* group and are very closely related.

Food Poisoning

P E Granum, The Norwegian School of Veterinary Science, Oslo, Norway

Background

There are several *Bacillus* species that have been involved in food poisoning (**Table 1**) although the only species that frequently causes problems is *Bacillus cereus*. There are six species that belong to the *B. cereus* group including *B. anthracis*. All of these species can cause food poisoning and in most cases are not distinguished in routine food laboratories, apart for *B. anthracis*, which is usually not hemolytic and is sensitive to penicillin. It has also been suggested that these species are so closely related that they should be considered as one species. In this article, the *B. cereus* group (apart from *B. anthracis*) is dealt with mostly as one species, although *B. thuringiensis* and *B. weihenstephanensis* are dealt with separately. *B. cereus* is a food-poisoning bacterium of growing concern, although it does not cause the types of illness that makes newspaper headlines. However, outbreaks of both the emetic- and the diarrheal type have caused some deaths in recent years. In some countries, where relatively few outbreaks of campylobacteriosis and salmonellosis are recorded, a steady increase in *B. cereus* food poisoning cases has been observed. As the Norwegian reference laboratory for the spore-forming food-poisoning bacteria, we can follow this trend closely. The same increase in *B. cereus* outbreaks is probably hidden behind the usually more serious outbreaks caused by other bacteria in Europe and the USA. And indeed it has been reported to be the most important cause of foodborne disease in The Netherlands, together with *Salmonella* spp., in the years between 1985 and 1991.

Bacillus cereus is a Gram-positive, spore-forming, motile, aerobic rod that also grows well anaerobically. It is a common soil saprophyte and is easily spread to many types of foods, especially of plant origin, but is also frequently isolated from meat, eggs, and dairy products. *B. cereus* and other members of the *B. cereus* group can cause two different types of food poisoning: the diarrheal type and the emetic type. The diarrheal type of food poisoning is caused by complex enterotoxins produced during vegetative growth of *B. cereus* in the small intestine, whereas the emetic toxin is preformed during the growth of cells in the food. For both types of food poisoning, the food involved has usually been heat-treated, and the surviving spores are the source of the food poisoning. *B. cereus* is not a competitive microorganism but grows well after cooking and cooling ($< 42–50\,^{\circ}C$). The heat treatment will cause spore germination, and in the absence of competing flora, *B. cereus* grows well. Invasion by psychrotolerant strains from the *B. cereus* group in the dairy industry has led to increasing surveillance of *B. cereus* in recent years.

B. cereus food poisoning is a nonreportable disease in all of Europe and the USA. Because of this, and because it is usually a relatively mild and short lasting

disease (< 24 h), it is highly underreported in official statistics. However, occasional reports have described more severe forms of the diarrheal type of *B. cereus* food poisoning, including a necrotic enteritis type causing three deaths. A Swiss boy also died after eating spaghetti containing large amounts of emetic toxin a few years ago.

The closely related *B. thuringiensis* is reported to produce enterotoxins and has been shown to cause food-poisoning symptoms when given to human volunteers. It has also been reported to cause food poisoning in regular outbreaks. The extensive use of this organism as a protective agent against insect attacks on crops may be part of the increasing problems with organisms of the *B. cereus* group observed in the food industry. Normal procedures for confirmation of *B. cereus* would not differentiate between the two species, if at all possible from heat-treated food products, since *B. thuringiensis* frequently will throw out insecticidal plasmids when grown above 30 °C. This makes it difficult to investigate the real numbers of food-poisoning cases caused by commercially used *B. thuringiensis*. To assure safe spraying with *B. thuringiensis*, the organism in use should be unable to produce food-poisoning toxins. The Health & Consumer Protection Directorate-General (European Commission) has already accepted that only non-toxin-producing *Bacillus* spp. should be allowed to be used in animal nutrition. *B. weihenstephanensis* is the psychrotolerant species within the *B. cereus* group, and most of the strains of this species are non- or low toxin producers, although there are exceptions. There are also *B. cereus* strains that are able to grow temperatures as low as 4 °C.

The other *Bacillus* species that might cause food poisoning are all isolated from soil and foods. However, in contrast to the members of the *B. cereus* group, only a small number of isolates of these species have the ability to produce toxins (harboring toxin genes) that can result in food poisoning.

Taxonomy of the *B. cereus* Group

The aerobic endosperm forming bacteria have traditionally been placed in the genus *Bacillus*. Over the past three decades, this genus has expanded to accommodate more than 100 species. Analysis of 16S ribosomal RNA sequences from numerous *Bacillus* species has indicated that the genus *Bacillus* should be divided into at least five genera or rRNA groups. The species treated in this text (Table 1) do all still belong to the genus *Bacillus*.

Bacillus anthracis, B. cereus, B. mycoides, B. thuringiensis, and, more recently, *B. pseudomycoides* and *B. weihenstephanensis* comprise the *B. cereus* group. These bacteria have highly similar 16S and 23S rRNA sequences, indicating that they have diverged from a common evolutionary line relatively recently. Extensive genomic studies of *B. cereus* and *B. thuringiensis* have shown that there is no taxonomic basis for separate species status. Nevertheless, the name *B. thuringiensis* is retained for those strains that synthesize a crystalline inclusion (Cry protein) or δ-endotoxin that may be highly toxic to insects. The *cry* genes are usually located on plasmids, and loss of the relevant plasmid(s) makes the bacterium indistinguishable from *B. cereus*. It is now clear that most strains in the *B. cereus* group, including *B. thuringiensis*, carry enterotoxin genes.

Foodborne Outbreaks Caused by the *B. cereus* Group

B. cereus is now well recognized as a food-poisoning organism. Outbreaks can be divided into two types according to their symptoms. The diarrheal type is far more frequent in Europe and the USA, whereas the emetic type appears more prevalent in Japan. Typical foods implicated are stews, puddings, sauces, and flour and rice dishes. When expressed as a proportion of all reported food poisonings, outbreaks ascribed to *B. cereus* seem to be concentrated in Scandinavia and Canada (10–47% of the total) and less frequent in Central Europe, UK, USA, and the Far East (1–5% of the total). Although these differences might partly be due to different consumer habits, they are also not comparable, because of dissimilar reporting practices. Thus, in the Netherlands, in 1991, *B. cereus* was responsible in 27% of outbreaks in which the causative agent was identified. However, the incidence was only 2.8% of the total, since the majority of cases of food poisoning were of unknown etiology. In addition, when the number of food poisoning cases ascribed to *B. cereus* are expressed on a per capita basis, many of the large regional differences in incidence disappear. Examples of foods involved in different outbreaks are listed in Table 2.

Characteristics of the *B. cereus* Disease

The emetic toxin results in vomiting, and the second type, caused by enterotoxins, leads to diarrhea. In a small number of cases, both types of symptoms are recorded, probably due to ingestion of preformed emetic toxin together with living *B. cereus* cells that may produce enterotoxins in the small intestine. There has been some debate about whether or not the enterotoxin(s) can be preformed in foods and cause intoxication. From a review of the literature, it is clear that the incubation time is slightly too long

Table 2 Examples of the variety of foods involved in *Bacillus cereus* food poisoning

Type of food	*Country*	*Number of people involved*	*Type of syndrome*
Barbecued chicken	Many countries		E, D
Cooked noodles	Spain	13	D
Cream cake	Norway	5	D
Fish soup	Norway	20	D
Lobster pâté	UK		D
Meat loaf	USA		D
Meat with rice	Denmark	> 200	D
Milk	Many countries		E, D
Pea soup	The Netherlands		D
Sausages	Ireland, China		D
School lunch	Japan	1877	E
Scrambled egg	Norway	12	D
Several rice dishes	Many countries		E, D
Stew	Norway	152	D
Turkey	UK, USA		D
Vanilla sauce	Norway	> 200	D
Wheat flour dessert	Bulgaria		D

E, emetic syndrome; D, diarrheal syndrome.

for that (> 6 h; average 12 h), and in model experiments, it has been shown that the enterotoxins are degraded on the way to the ileum. Although the enterotoxin(s) can be preformed, the number of *B. cereus* cells in the food would need to be at least two orders of magnitude higher than that necessary to cause food poisoning, and such products would no longer be acceptable to the consumer. The characteristics of *B. cereus* food poisoning are listed in Table 3.

In recent years, food poisoning resulting from ingestion of *B. anthracis*-infected meat has been reported in three continents (Asia, Africa, and America). Patients who do not develop anthrax symptoms instead develop diarrhea, sometimes bloody, with fever and rashes. It is not yet known whether the enterotoxins known from *B. cereus* are at least partly involved in such symptoms. However, it is known that enterotoxin genes (*nhe*) are present in *B. anthracis* (from the total genome sequence), although the positive regulator for enterotoxin production (see Table 4) is nonfunctional, but in other strains, this might not be the case.

Infectious Dose of *B. cereus* Group

Counts ranging from 10^4 to 10^9 per gram (or milliliter) *B. cereus* have been reported in affected foods after food poisoning, giving total infective doses of about 5×10^4–10^{11}. The variation in the infective dose is partly due to the differences in the amount of enterotoxin produced by different strains (at least two orders of magnitude) and partly due to the difference in infective dose between vegetative cells and spores, since all the spores will survive the stomach-acid barrier. Thus, any food containing more than 10^3 *B. cereus* per gram cannot be considered completely safe for consumption.

Virulence Factors/Mechanisms of Pathogenicity

Very different types of toxins cause the two types of *B. cereus* food poisoning. The emetic toxin, causing vomiting, is a ring-formed small peptide, whereas the diarrheal disease is caused by several different enterotoxins (Table 4).

Emetic toxin

The emetic toxin causes emesis (vomiting) only, and for many years, its structure was a mystery, as the only detection system involved living primates. The discovery that the toxin could be detected (vacuolation activity) by the use of HEp-2 cells led to its isolation and structural determination. The emetic toxin has been named cereulide, and consists of a ring structure

Table 3 Characteristics of the two types of disease caused by *Bacillus cereus*

	Diarrheal syndrome	*Emetic syndrome*
Infective dose	10^5–10^7 (total)	10^5–10^8 (cells per gram food to produce enough emetic toxin)
Toxin produced	In the small intestine of the host	Preformed in foods
Type of toxin	Protein(s)	Cyclic peptide
Incubation period	8–16 h (occasionally > 24 h)	0.5–5 h
Duration of illness	12–24 h (occasionally several days)	6–24 h
Symptoms	Abdominal pain, watery diarrhea (occasionally bloody diarrhea) sometimes with nausea	Nausea, vomiting, and malaise (sometimes followed by diarrhea, due to additional enterotoxin production?)
Foods most frequently implicated	Meat products, soups, vegetables, puddings/sauces and milk/milk products	Fried and cooked rice, pasta, pastry, and noodles

Table 4 Toxins involved in food poisoning produced by *Bacillus cereus*

Toxin	*Type/size*	*Food poisoning*
Hemolysin BL (Hbl)	Protein, three components transcribed from one operon: HblC (L_2-component), HblD (L_1-component) and HblA (B-component); the proteins have a molecular mass of 37–46 kDa	Probably
Nonhemolytic enterotoxin (Nhe)	Protein, three components transcribed from one operon: NheA, NheB, and NheC; the proteins have a molecular mass of 36–41 kDa	Yes
Cytotoxin K (CytK)	Protein, one component, 34 kDa	Yes, three deaths
Emetic toxin (cereulide)	Cyclic peptide, 1.2 kDa	Yes, one death

of three repeats of four amino-and/or oxy-acids: [D-*O*-Leu-D-Ala-L-*O*-Val-L-Val]$_3$. This ring structure (dodecadepsipeptide) has a molecular mass of 1.2 kDa and is chemically closely related to the potassium ionophore valinomycin. The emetic toxin is resistant to heat, pH, and proteolysis, but is not antigenic. The biosynthetic pathway and mechanism of action of the emetic toxin still have to be elucidated, although it has been shown recently that it stimulates the vagus afferent through binding to the 5-HT$_3$ receptor. It has just been shown that cereulide is synthesized non-ribosomally by a peptide synthetase.

Cereulide was responsible for the death (fulminant liver failure) of a 17-year-old Swiss boy a few years ago. A large amount of *B. cereus* emetic toxin was found in the residue from the pan used to reheat the food (pasta) and in the boy's liver and bile. In a recent experiment, mice were injected i.p. with synthetic cereulide, and the development of histopathological changes was examined. At high cereulide doses, massive degeneration of hepatocytes occurred. The serum values of hepatic enzymes were highest on days 2–3 after the inoculation of cereulide, and rapidly decreased thereafter. General recovery from the pathological changes and regeneration of hepatocytes was observed after 4 weeks.

Enterotoxins

As shown in Table 4, three different enterotoxins, believed to be involved in food poisoning, have been characterized to date. The three-component hemolysin (Hbl; consisting of three proteins: B, L_1, and L_2) with enterotoxin activity was the first to be fully characterized. This toxin also has dermonecrotic and vascular permeability activities, and causes fluid accumulation in ligated rabbit ileal loops. Hbl has been suggested to be a primary virulence factor in *B. cereus* diarrhea. Convincing evidence has shown that all three components are necessary for maximal enterotoxin activity. It has been suggested, from studies of interactions with erythrocytes, that the B protein (HblA) is the component that binds Hbl to the target cells, and that L_1 (HblD) and L_2 (HblC) have lytic functions. More recently, another model for the action of Hbl has been proposed, suggesting that the components of Hbl bind to target cells independently and then constitute a membrane attacking complex, resulting in a colloid osmotic lysis mechanism. Substantial heterogeneity has been observed in the components of Hbl, and individual strains have been shown to produce various combinations of single or multiple bands of each component.

More recently, a nonhemolytic three-component enterotoxin (Nhe) has been characterized. The three components of this toxin are different from the components of Hbl, although there are similarities. The three components of Nhe enterotoxin were first purified from a *B. cereus* strain isolated after a large food-poisoning outbreak in Norway in 1995. Binary combination of the components of this enterotoxin shows some biological activity, but not nearly as high as when all the components are present.

Almost all tested *B. cereus*/*B. thuringiensis* strains produce Nhe, and about 60% produce Hbl. At present, we do not know how important each of them is in relation to food poisoning. There are several food-poisoning strains that do not produce Hbl, but none that do not produce Nhe.

The newly discovered cytotoxin K (CytK) is similar to the β-toxin of *Clostridium perfringens* (and other related toxins) and was the cause of the symptoms in a severe outbreak of *B. cereus* food poisoning in France in 1998. In this outbreak, several people developed bloody diarrhea, and three died. This could be considered an outbreak of *B. cereus* necrotic enteritis, although it is not nearly as severe as the *C. perfringens* type C food poisoning.

There is significant sequence identity between the three proteins of Nhe and between the Nhe and Hbl proteins. The identity is highest in the N-terminal third of the proteins. The most pronounced gene

sequence similarities are found between *nheA* and *hblC, nheB* and *hblD*, and *nheC* and *hblA*. This is not only in direct comparison of the sequences, but also in predicted transmembrane helices for the six proteins. NheA and HblC have no predicted transmembrane helices, whereas NheB and HblD have two each. Finally, NheC and HblA have one predicted transmembrane helix each, in the same position in the two proteins. Although there are some similarities among the components of Hbl and Nhe, they cannot be substituted for each other to give biological active complexes.

Other Possible Virulence Factors

The *B. cereus* spore is more hydrophobic than any of the other *Bacillus* spp. spores, which makes it adhesive to several types of surfaces. This makes it difficult to remove during cleaning and a difficult target for disinfection. The *B. cereus* spores also contain appendages and/or phili that are, at least partly, involved in adhesion (Figure 1). These properties of the *B. cereus* spore not only allow them to survive sanitation, and thus become available for contamination of different foods, but also aid adherence to epithelial cells. Experiments have shown that spores, at least from one strain isolated after one outbreak, can indeed adhere to Caco-2 cells in culture and that these properties are linked to hydrophobicity and possibly to the appendages. A longer incubation period is observed in this case, as expected, as the spore would first have to germinate.

Commercial Methods for Detection of the *Bacillus cereus* Toxins

Neither of the two available commercial immunoassays can quantify the toxicity of the enterotoxins from *B. cereus*. The assay from Oxoid measures the presence of the HblC (L_2) component, whereas the Tecra kit mainly detects the NheA protein. However, if one or both of the commercial kits react positively with proteins from *B. cereus* supernatants, it is likely that the strain is enterotoxin-positive, specifically if the strains are cytotoxic on epithelial cells. If the supernatants are shown to be cytotoxic, the strains can be regarded as enterotoxin-positive. At present, there is no commercial method available to detect CytK, and, unfortunately, no commercial method to detect the emetic toxin either. However, a specific, sensitive, semiautomated, and quantitative Hep-2 cell culture-based 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide assay for *B. cereus* emetic toxin has been developed.

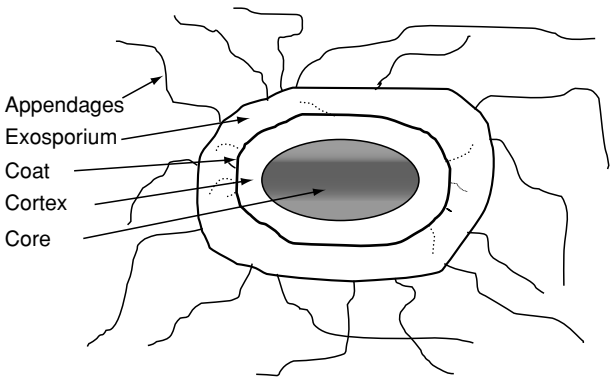

Figure 1 *Bacillus cereus* spore with the different layers and appendages.

Other Bacillus Species

Of the other *Bacillus* species involved in food poisoning, only three have been involved frequently enough to give a relatively clear picture of the diverse food-poisoning symptoms (Table 5). For *B. sphaericus* and *B. brevis*, only a few cases have been recorded. The infective dose variation for these species is not clear, although they are probably quite high ($> 10^6$). From the data in Table 5, it is clear that many types of toxins are involved. For incubation periods shorter than 6 h, preformed toxins are expected. Enterotoxins similar to those of *B. cereus* have been detected from several of the other *Bacillus* species (Table 1) using immunological methods and polymerase chain reaction-based methods. However, until the toxins have been isolated and characterized, there structure and function are mostly speculative.

One of the toxins has been well characterized. From a strain of *B. licheniformis* isolated for milk powder (which caused the death of a young child), a toxin similar to the *B. cereus* emetic toxin was isolated. This toxin was also found in 13 out of 210 isolates (6%) of *B. licheniformis* from different sources (mainly from foods). The toxins isolated from three strains of different origins contained the same component each of which had the same amino acid residues L-Gln, L-Leu, D-Leu, L-Val, L-Asp, D-Leu, and L-Ile, in that order. Toxins were identified as lichenysin A, a cyclic lactonic heptalipopeptide in which the main 3-hydroxy fatty acids are 13–15 carbons in length. The molecular mass of this toxin is about 1 kDa.

Prevention and Control of *Bacillus* Food Poisoning

Prevention and control of *Bacillus* spp. is relatively easy, apart from in the dairy industry, where *B. cereus*

Table 5 Characteristics of the illness caused by the three *Bacillus* species other than *B. cereus* causing food poisoning

	Bacillus subtilis	*Bacillus licheniformis*	*Bacillus pumilus*
Foods involved	Meat dishes with elements of vegetables Seafood with rice Bread and pastry products Sandwiches Pizzas	Meat dishes with elements of vegetables Bread and pastry products Chicken	Meat products Sandwiches Canned tomato juice
Infective dose	$> 10^5$	$> 10^6$	$> 10^6$
Incubation period	10 min to 14 h	2–14 h	15 min to 11 h
Duration	2–8 h	6–24 h	
Symptoms:			
Vomiting	80%	54%	Yes
Diarrhea	49%	92%	Yes (not all)
Abdominal pain/cramps	27%	46%	?
Other symptoms	Nausea, headaches, and flushing/sweating	Not reported	Not reported

Table 6 Factors determining the growth of the four most common food-poisoning *Bacillus* species

Species	*pH range*	*Temperature range*	*Other factors*
Bacillus cereus	4.3–9.3	4–50 °C	Growth at a_w to 0.92 Inhibited by 0.2% sorbic acid (pH < 5.0) Growth in 7% NaCl
Bacillus subtilis	5.0–8.5	10–50 °C	Growth in 7% NaCl Growth with only glucose and ammonium
Bacillus licheniformis	5–?	> 15–55 °C	Growth in 7% NaCl
Bacillus pumilus	5–?	10–45 °C	Growth in 7% NaCl

causes major problems. There are also a growing number of precooked long-life products on the market that are difficult to produce completely free from *Bacillus* spp. spores. The growth characteristics of the four most common food-poisoning *Bacillus* spp. are listed in Table 6. If foods are not kept at 6–60 °C for too long, no spore germination and growth of these species will occur. Rapid cooling and proper reheating of cooked food are essential if the food is not consumed immediately. Long-term storage must be at temperatures below 8 °C (or preferably 4–6 °C to prevent growth of *B. cereus*). Low-pH foods (pH $\leq$ 4.3) can be considered safe from growth of the food-poisoning *Bacillus* spp. *Bacillus* spores are commonly isolated from spices, cereals, and dried foods. For example, during 1960–8, *B. cereus* was the third most common cause of food poisoning in Hungary, and meat dishes were frequently involved. The reason for this was the preference for well-spiced meat dishes in Hungary. Several cases of *B. cereus* food poisoning, (involving many people) resulting from consumption of meat-containing dishes have also been reported recently in Norway.

The emetic syndrome of *B. cereus* food poisoning is often connected with the consumption of rice in Chinese restaurants. The predominance of cases in Chinese restaurants is linked with the common practice of saving portions of boiled rice from bulk cooking. The boiled rice is then stored, usually at room temperature, overnight and *B. cereus* is then able to multiply. The same problem may occur when foods such as pasta and pizza are stored for long periods of time at room temperature.

At present, the main problem with *B. cereus* seems to be in the dairy industry, where the keeping quality of milk is determined by the number of *B. cereus* cells/spores in the product. The bacterium may cause aggregation of the creamy layer of pasteurized milk, known as bitty cream, which is explained by the lecithinase activity of *B. cereus*. Further, *B. cereus* is responsible for sweet curdling (without pH reduction) in both homogenized and nonhomogenized low-pasteurized milk. It seems impossible to completely avoid the presence of *B. cereus* in milk where it already infects raw milk at the farm. Soiling of the udders of cows is the principal source of contamination of milk with *B. cereus*. Soil has been shown to contain 10^5–10^6 spores per gram. It is very important, therefore, that the udder and the teats are cleaned to reduce the contamination of raw milk. Transport and further storage in the dairy may result in further

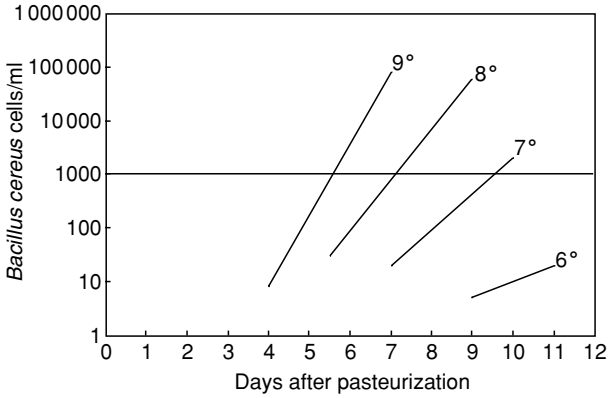

Figure 2 Growth rate of *Bacillus cereus* cells in milk stored at 9, 8, 7, and 6 °C.

contamination of the raw milk from *B. cereus* spores already present (adherent) in the tanks or pipelines. The vegetative bacteria are killed in the pasteurization process, but the spores survive. Pasteurization might activate at least some of the spores (heat activation), which might start germinating.

To control *B. cereus* in milk and milk products, it is very important to trace the presence of spores from farmer to package. The storage temperature is the most important factor in keeping *B. cereus* numbers to a minimum. Figure 2 shows the number of days required before milk contains more than 10^3 *B. cereus* per milliliter at different temperatures. An increase of just 2 °C during storage, from 6 to 8 °C, increase the growth rate substantially. At the dairy, the milk is generally kept at 4 °C, thus ensuring a good keeping quality. However, during distribution, where the energy costs are often valued higher than food quality, temperatures up to 8 °C and above are common. Further, the consumer often exposes milk to higher temperatures for longer periods, for example at the breakfast table. The majority of the strains from the *B. cereus* group that grow at low temperatures (including *B. weihenstephanensis*) are usually low in enterotoxin production, although there are exceptions.

As mentioned before, the spores of *B. cereus* are very adhesive to different hydrophobic surfaces (such as glass and stainless steel), and for this reason, *B. cereus* is often found in food plants and kitchen environments. The strong adhesion of *B. cereus* spores is mainly due to three characteristics: the high relatively hydrophobicity, the low spore surface charge, and the spore morphology (appendages). At present, the only way to overcome this problem, when first introduced, is through the use of hypochlorite (0.2% at pH 7–8) or UVC light, since neither low- nor high-pH cleaning is sufficient to control the problem.

See also: ***Bacillus***: Occurrence; Detection; **Food Poisoning**: Classification; Statistics

Further Reading

Agata N, Ohta M, Mori M and Isobe M (1995) A novel dodecadepsipeptide, cereulide, is an emetic toxin of *Bacillus cereus*. *FEMS Microbiology Letters* 129: 17–20.

Andersson A, Granum PE and Rönner U (1998) The adhesion of *Bacillus cereus* spores to epithelial cells might be an additional virulence mechanism. *International Journal of Food Microbiology* 39: 93–99.

Beecher DJ, Schoeni JL and Wong ACL (1995) Enterotoxin activity of hemolysin BL from *Bacillus cereus*. *Infection and Immunity* 63: 4423–4428.

Drobniewski FA (1993) *Bacillus cereus* and related species. *Clinical Microbiological Reviews* 6: 324–338.

Granum PE (2001) *Bacillus cereus*. In: Doyle M, Beuchat L and Montville T (eds) *Food Microbiology. Fundamentals and Frontiers*, pp. 373–381. Washington, DC: ASM Press.

Granum PE and Baird-Parker TC (2000) *Bacillus* spp. In: Lund B, Baird-Parker T and Gould G (eds) *The Microbiological Safety and Quality of Food*, pp. 1029–1039. Guithersburg, MD: Aspen.

Granum PE and Brynestad S (1999) Bacterial toxins as food poisons. In: Alouf JE and Freer JH (eds) *The Comprehensive Sourcebook of Bacterial Protein Toxins*, pp. 669–681. London: Academic Press.

Kramer JM and Gilbert RJ (1989) *Bacillus cereus* and other *Bacillus* species. In: Doyle MP (ed.) *Foodborne Bacterial Pathogens*, pp. 21–70. New York: Marcel Dekker.

McKillip JL (2000) Prevalence and expression of enterotoxins in *Bacillus cereus* and other *Bacillus* spp. *Antonie Van Leeuwenhoek* 77: 393–399.

Salkinoja-Salonen MS, Vuorio R, Andersson MA *et al.*(1999) Toxigenic strains of *Bacillus licheniformis* related to food poisoning. *Applied and Environmental Microbiology* 65: 4637–4645.

Bacteria *See* **Microbiology**: Classification of Microorganisms; Detection of Foodborne Pathogens and their Toxins; **Antibiotic-resistant Bacteria**

Baking *See* **Biscuits, Cookies, and Crackers**: Nature of the Products; Methods of Manufacture; Chemistry of Biscuit Making; Wafers; **Bread**: Dough Mixing and Testing Operations; Breadmaking Processes; Chemistry of Baking; Sourdough Bread; Dietary Importance; Dough Fermentation; **Cakes**: Nature of Cakes; Methods of Manufacture; Chemistry of Baking

Baking Powder *See* **Leavening Agents**

BANANAS AND PLANTAINS

J W Daniells, Queensland Department of Primary Industries, Queensland, Australia

Introduction

Bananas and plantains (*Musa* spp.) are grown extensively throughout the tropical and subtropical regions of the world. Together they represent the number-one fruit crop in the world, in terms of both production and trade, exceeding oranges by 37×10^6 t per year for production and by 10×10^6 t per year for trade. Most of the exported fruit, which makes up only 15% of the total, are Cavendish-type dessert bananas. However, bananas and plantains that are cooked account for almost half of the world production. This article describes their global distribution and importance, the morphological and nutritional characteristics of the fruit, handling and storage, and other uses of fresh and processed fruit. (*See* **Fruits of Temperate Climates:** Commercial and Dietary Importance.)

The terms 'bananas' and 'plantains' require clarification. 'Bananas' refers to all the members of the genus *Musa*. In the narrow sense, plantains are a defined group within this genus which have the AAB genome and are characterized by the orange-yellow color of both the compound tepal of the flower and the fruit pulp at ripeness. When ripe, a relatively high proportion of starch (10–25% of fresh weight) is present in the pulp. The fruits are slender, angular to pointed, and are generally palatable only after cooking. Plantains, in the broad sense, include other members of the genus *Musa* that are starchy at ripeness but which lack the other characteristics. In this article, plantains are referred to in the narrow sense, except when global production figures are discussed, as these refer to plantains in the broad sense.

Global Distribution

Before international trade began a little over 100 years ago, bananas in temperate regions were merely a curiosity. Bananas are grown in commercial plantations, but more often in household gardens because of their almost universal appeal as a food, ease of growing, quick production, and attractive plant appearance.

Bananas and plantains are suited to the warm, high-rainfall regions of the lowland tropics. They originated in South-east Asia (including Papua New Guinea) and have been spread to other regions of the world in the past 2000 years. They are now grown from the equator to about 35 °N and 30 °S, and within these latitudes they may be grown in semiarid locations with irrigation or in large plastic greenhouses.

Bananas for export are produced mainly in Central and South America as well as the Philippines (**Tables 1–4**). Most exported fruit goes to North America, Europe, and Japan, and in 2000 was valued at $\$4.306 \times 10^9$. Exports of plantains have increased in importance in recent years but still represents less than 2% of the combined total exports of bananas and plantains. Other major areas of banana production are South-east Asia and East Africa (according to Food and Agriculture Organization (FAO) statistics*) but little of this is exported. Plantain production is most important in East and West Africa as well as South America.

Average world consumption of bananas and plantains in 2001 was about 16 kg per person per year. However, in much of Africa and Latin America it is 5–10 times this amount, and it is as much as 450 kg

*FAO production statistics need careful interpretation because the meaning of the words 'bananas' and 'plantains' differs among countries, particularly in East Africa. For instance information from Burundi on bananas largely refers to plantains used for cooking purposes.

Table 1 Proportion of world production of 69 million tonnes of bananas among regions and countries 2001

Region	*Country*	*Percentage of total world production (69×10^6 t)*
Asia	India	23.3
	China	7.9
	Philippines	7.4
	Indonesia	5.2
	Thailand	2.5
	Vietnam	1.6
	Other	2.8
South America	Brazil	8.4
	Ecuador	11.0
	Colombia	2.0
	Other	3.0
Central America	Costa Rica	3.3
	Mexico	2.9
	Honduras	0.7
	Guatemala	1.1
	Panama	0.7
	Other	0.4
Africa	Burundi	2.3
	Cameroon	1.2
	Uganda	1.4
	Other	6.3
Other		4.6

Source: Food and Agriculture Organization 20 August 2002: www.fao.org.

Table 2 Proportion of world production of 29 million tonnes of plantains among regions and countries 2001

Region	*Country*	*Percentage of total world production (29×10^6 t)*
Asia		3.9
South America	Colombia	9.7
	Peru	5.0
	Venezuela	2.4
	Ecuador	1.6
	Other	0.8
Central America	Dominican Republic	1.2
	Cuba	1.3
	Other	3.5
Africa	Uganda	32.7
	Rwanda	5.4
	Congo Democratic Republic	1.8
	Ghana	6.6
	Nigeria	6.5
	Ivory Coast	5.2
	Cameroon	4.8
	Tanzania	2.2
	Other	5.4

Source: Food and Agriculture Organization 20 August 2002: www.fao.org.

Table 3 Proportion of world exports of 14 million tonnes of bananas among regions and countries 2000

Region	*Country*	*Percentage of total world exports (14×10^6 t)*
Asia	Philippines	11.2
	Other	1.2
South America	Ecuador	28.1
	Colombia	12.0
	Other	1.0
Central America	Costa Rica	14.7
	Guatemala	5.6
	Honduras	1.3
	Panama	3.4
	Other	1.3
Africa		2.9
Other		17.3

Source: Food and Agriculture Organization 20 August 2002: www.fao.org.

per person per year in some parts of East Africa. The major export production areas are located close to the equator, where an even supply of fruit throughout the year is possible, and where they are less damaged by cyclones or typhoons.

The optimum temperature range for bananas is about 22–31 °C. Chilling injury of fruit occurs when the latex coagulates at temperatures below 13 °C. Bananas are susceptible to frost and in subtropical regions are often planted on hillsides to avoid it. The growth cycle is greatly prolonged outside the tropics and productivity is generally reduced. The industries that have developed in the subtropics have largely done so because of proximity to markets.

Varieties

All the edible bananas belong to the Eumusa (sometimes referred to as Musa) section of the genus *Musa*, except for the Fe'i bananas of the Pacific region which belong to the Australimusa section (Figure 1). The Fe'i bananas are characterized by erect bunches and pink-red sap, and an orange, slimy fruit pulp that generally requires cooking. There are numerous wild seeded species in each of the *Musa* sections.

Edible bananas and plantains belonging to the Eumusa section are believed to contain genomes from two wild species, *M. acuminata* (A) and *M. balbisiana* (B). Most cultivated bananas are triploid and are classified according to characteristics estimating the contribution of the two parent species. Because the binomial Latin nomenclature for edible varieties, e.g., *Musa cavendishii* cultivar (cv.) Williams, proved unsatisfactory, they are referred to as, for example, *Musa* spp. (AAA Group, Cavendish Subgroup) cv. Williams. There are about

Table 4 Proportion of world imports of 14 million tonnes of bananas among regions and countries 2000

Region	*Country*	*Percentage of total world imports (14×10^6 t)*
Asia	Japan	7.6
	China	4.2
	Other	5.8
North America	USA	28.2
	Canada	2.8
Europe	Germany	7.8
	Belgium	7.2
	UK	5.2
	Italy	4.2
	Russian Federation	3.5
	France	2.4
	Poland	2.0
	Other	12.5
Other		6.6

Source: Food and Agriculture Organization 20 August 2002: www.fao.org.

500 varieties of bananas and plantains in existence, but only 150 or so are primary clones and the remainder are somatic mutants.

Bananas for export now come almost entirely from varieties of the Cavendish subgroup. From the 1940s to 1960s, these replaced cv. Gros Michel, which was devastated by a soil-borne fungus called Panama disease (fusarium wilt) in one of the largest disease epiphytotics that has ever occurred in crop plants. The greatest diversity of varieties occurs in Southeast Asia where particular varieties are especially favored for various culinary uses (Figure 2).

The leaf disease, black Sigatoka, and new races of Panama disease currently pose major threats to world banana and plantain production. As resistant varieties are developed by conventional breeding programs, somacloning and genetic engineering, further changes in the varieties cultivated can be expected.

Fruit Morphology and Anatomy

The banana plant is a large, tree-like, determinate perennial herb with a basal rhizome, a pseudostem composed of leaf sheaths, and a terminal crown of large leaves (Figure 3). The terminal inflorescence is initiated near ground level and is then thrust up the center of the pseudostem by elongation of the true stem. The basal flower clusters (hands) are female and form the fruit bunch. Distal flower clusters are male, do not produce fruit, and are commonly deciduous. The banana has the largest inflorescence of any plant grown as a crop. The world record for a mature fruit bunch is 144 kg, obtained in the Canary Islands.

Banana bunches have from one to 20 hands and take 2–6 months to reach maturity. Bunches are pendant or subhorizontal and usually weigh between 10

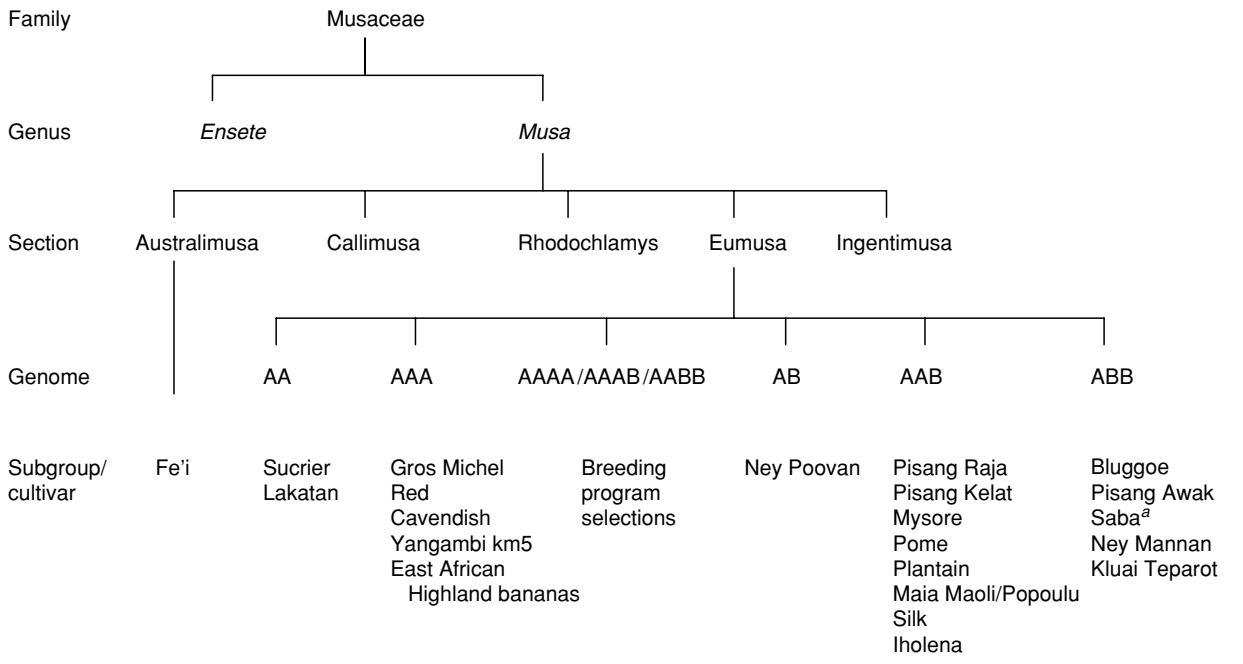

Figure 1 Systematic position of banana varieties. [a]Considered as BBB in the Philippines. Modified from Stover and Simmonds (1987).

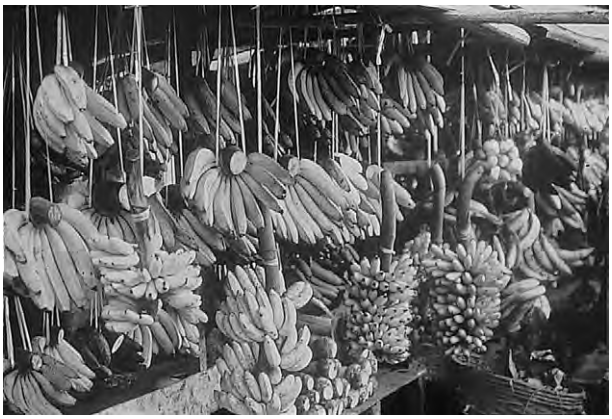

Figure 2 (see color plate 1) A wide range of banana varieties are available in South-east Asian markets. Reproduced from Bananas and Plantains. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

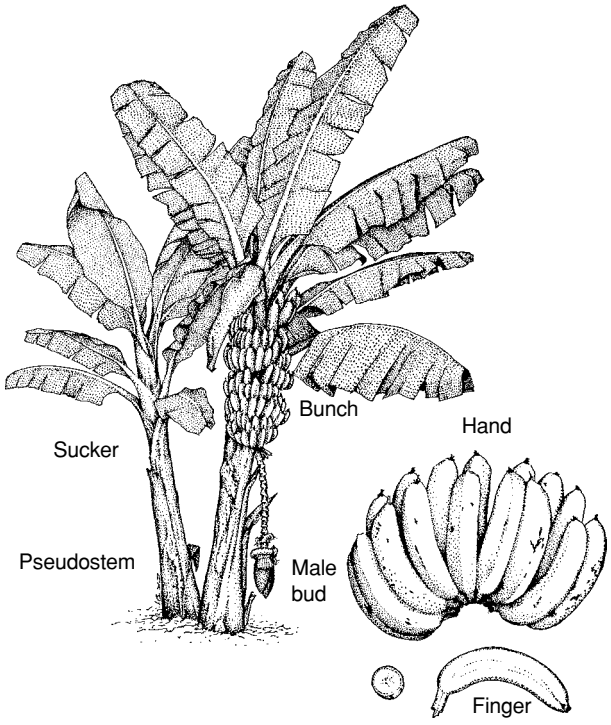

Figure 3 Banana plant, bunch, and fruit morphology. Reproduced from Bananas and Plantains. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

and 60 kg. Individual fruits (fingers) can number up to 300 on a bunch. Fingers of different varieties can be anywhere from 6 to 60 cm in length and from 50 to 1000 g in weight, but are more usually 15–30 cm long and weigh 50–200 g. The curved shape of the fruit is caused by a negative geotropic growth response. Immature bananas are usually green, and when mature they ripen to a yellow color. Young fruits are somewhat angular, but this disappears with maturity in the AAA varieties. AAB and ABB varieties may still be angular at maturity.

The fruit is a berry that develops from the inferior ovary of the female flower. It is parthenocarpic, i.e., it develops without the stimulus of pollination. The ovules shrivel early but can still be recognized as brown specks in the center of mature fruit. Most varieties are sterile or have very low fertility. If a pollen source, such as a wild species, is nearby, some varieties will set an occasional dark hard seed about 3–5 mm in diameter. These are a hazard to the teeth of consumers so it is fortunate that seeds are an extremely rare event.

In the growing fruit, the pulp-to-peel ratio of fresh weight varies from about 1:1 to 4:1, depending upon variety and maturity at harvest. When the mature fruit ripens, the pulp-to-peel ratio increases, and this is thought to be in part as a result of water movement from the peel to the pulp associated with the hydrolysis of starch to osmotically active sugars.

Nutritional Composition

Ripe dessert bananas are considered by many to be a complete food if taken in association with a protein source such as milk. They are favored as food for young babies and elderly people because they are easily digested and are very nutritious. They are excellent for people with stomach complaints, particularly ulcers, and are ideal for diets with low levels of cholesterol, fats, and sodium. The potassium concentration is about 350 mg per 100 g of fresh pulp with trace amounts of sodium (Table 5). Bananas are also recommended in the treatment of infant diarrhea, celiac disease, and colitis. They are a good source of vitamins C and B_6 (Tables 5 and 6). A major feature is the high sugar-to-acid ratio (ranging from 100 to 180) compared with 7–10 for citrus. This is confirmed by the carbohydrate concentrations shown in Tables 5 and 7. If we assume 20% of the fresh banana pulp is sugar and 76% water, then 83% of the dried solids are sugar (Tables 5 and 7). This is why bananas have recently received much attention from people in sport because of the high carbohydrate content, capable of rapidly releasing energy necessary for vigorous sporting events. The sugars are almost entirely glucose, fructose, and sucrose, in a ratio of 20:15:65. The high sugar concentration of ripe bananas is exceeded in fresh fruit only by that of dates, jujube, tamarind, and carob. (*See* **Celiac (Coeliac) Disease**; **Colon**: Diseases and Disorders; **Exercise**: Metabolic

Table 5 Compositional data per 100 g edible portion of plantains (unripe) and bananas (ripe)

	Proximate analysis (g)			*Mineral content (mg)*			*Vitamin content*	
Components	*Plantain*	*Banana*	*Minerals*	*Plantain*	*Banana*	*Vitamins*	*Plantain*	*Banana*
Water	67.5	75.1	Sodium	4	1	Retinol (μg)	0	0
Sugars	5.7	20.9	Potassium	500	400	Carotene (μg)	360	21
Starch	23.7	2.3	Calcium	9	6	Vitamin D (μg)	0	0
Dietary fiber	2.3	3.1	Magnesium	37	34	Thiamin (mg)	0.10	0.04
Total nitrogen	0.18	0.19	Phosphorus	36	28	Riboflavin (mg)	0.05	0.06
Protein	1.1	1.2	Iron	0.5	0.3	Nicotinic acid (mg)	0.7	0.7
Fat	0.3	0.3	Copper	0.08	0.1	Ascorbic acid (mg)	15.0	11
			Zinc	0.1	0.2	Vitamin E (mg)	0.20	0.27
			Chlorine	80	79	Vitamin B_6 (mg)	0.30	0.29
						Vitamin B_{12} (mg)	0	0
						Folate (μg)	22	14
						Pantothenate (mg)	0.26	0.36

Source: Holland B, Welch AA, Unwin ID, Buss DH, Paul AA and Southgate DAT (1991) *McCance and Widdowson's The Composition of Foods*, Fifth revised and extended edition. London: The Royal Society of Chemistry and Ministry of Agriculture, Fisheries and Food.

Table 6 Nutrient content of the fruit pulp of banana varieties

	Percentage of US RDA[a] *per 100 g*		
Nutrient	*Gros Michel*[b]	*Cavendish*[c]	*Horn Plantain*[c]
Vitamin A	3.8	5.1	61.6
Ascorbic acid	13.3	20.0	26.7
Vitamin B_6	25.0[c]	NA[d]	NA
Thiamin	3.3	2.6	2.9
Riboflavin	3.8	5.3	5.9
Nicotinic acid	4.7	4.8	4.0

[a]US recommended dietary allowance.
[b]Source: USDA (1963) *Composition of Foods*. Agriculture Handbook 8. Washington, DC: US Government Printing Office.
[c]Source: Anonymous (1959) *Bananas: Versatile in Health or Illness*. Boston: United Fruit.
[d]NA, not available.

Requirements; **Infants**: Nutritional Requirements.) *See also* individual nutrients.

The characteristic aroma of bananas has received considerable attention, with more than 350 volatile compounds having been identified. The major constituents appear to be amyl and isoamyl esters of acetic, propionic, and butyric acids.

The major differences between bananas and plantains are: (1) the lower moisture percentage of the pulp in green plantains compared with ripe bananas (Tables 5 and 7); (2) the lower sugar concentration in ripe plantains compared with ripe bananas; and (3) plantains are a much richer source of vitamin A than bananas (Tables 5 and 6). These differences are valid for the edible product, but are less important if fruits are compared at the same stage of ripeness. For example, the moisture percentage can increase from about 60% in preclimacteric banana fruit to 70% after ripening and 75–80% at the stage of senescence. In addition, there are major inconsistencies in the literature regarding the conversion of starch to sugar in plantains, with some references indicating only traces of starch in the ripe fruit. Some of the differences are probably due to different ripening conditions and the stage of ripeness selected for measurement. (*See* **Ripening of Fruit**.)

Fresh Fruit Handling and Storage

Handling procedures vary substantially in the different banana-producing countries and depend upon whether fruit is to be exported. This discussion is based largely on procedures adopted in commercial plantings producing fruit for export because it provides an understanding of the factors involved.

Bananas travel best and receive a minimum of mechanical damage while they are in a hard, green condition. It is necessary to transport the fruits to the marketplace in this hard, green state so that they can be uniformly ripened with ethylene gas (1000 mg l^{-1}) in humidified rooms at 15–18 °C. This greatly facilitates marketing. Fruits can take from a few days to 3 weeks to reach the marketplace, and so must have sufficient greenlife (the period after harvest for which the fruit stays in a hard green condition) to survive this journey. No postharvest treatment can improve upon the inherent greenlife, but treatments can reduce its rate of decline. In general, the earlier the harvest in the life of the fruit, the greater is the fruit greenlife, but any gain in greenlife must be balanced against the loss in bunch weight (5–10% per week). A bunch that would ripen in the field 5 months after it emerged may be harvested and ripened satisfactorily as early as about 10 weeks (pulp-to-peel ratio of 1:1), when the fingers are still quite thin. A key to profitable banana growing is to maximize yield without premature ripening occurring. Other considerations

Table 7 Proximate analysis (g per 100 g) of banana and plantain

	Banana		*Plantain*	
	Edible portion	*Oven-dried solids*	*Edible portion*	*Oven-dried solids*
Water	75.7		66.4	
Carbohydrate	22.2	91.4	31.2	92.8
Protein	1.1	4.5	1.1	3.3
Fat	0.2	0.8	0.4	1.2
Ash	0.8	3.3	0.9	2.7

Source: USDA (1963) *Composition of Foods*. Agriculture Handbook 8. Washington, DC: US Government Printing Office.

Figure 4 (see color plate 2) Bananas for export are usually transported to the packing shed by cableways to minimize skin blemishes due to mechanical damage. Reproduced from Bananas and Plantains. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

are that greater fruit maturity leads to a higher proportion of edible pulp and possibly an increase in fruit flavor. In practice, maturity standards are varied, so that more mature bunches are harvested for nearer markets and less mature ones for more distant markets.

Maximum greenlife for a particular finger diameter (grade) at harvest is achieved by manipulating the plant and environment to obtain the maximum possible rate of fruit filling. Finger diameter, bunch age, and degree of fruit fullness (loss of angularity) are used as criteria for harvest. The combination of age-grade control for determining time of harvest insures the production of high-quality fruit with sufficient greenlife at harvest. To minimize the risk of premature ripening after harvest, it is important to keep the fruit as cool as possible (but above 13 °C to prevent chilling injury) and not to expose it unnecessarily to light.

Banana bunches are harvested by hand and usually carried on the shoulder to nearby cableways (Figure 4) or tractor-drawn, padded trailers. The plastic sleeve, known as a bunch cover, that is applied to the young bunch to improve fruit quality is retained to minimize mechanical damage during handling. In the packing shed, the bunch cover is removed, the bunches are dehanded, usually into tanks of water, and then fruit is sorted, graded, divided into clusters of 4–10 fingers, weighed, labeled, and packed into fiberboard cartons containing plastic liners. Fungicide may be applied in a separate operation before the fruit is packed. Cartons hold from 12 to 18 kg of fruit. The packed fruit is sent by refrigerated transport to the marketplace and usually involves transport by rail, road, and sea. The transport system is highly integrated and regular (at least weekly) so that fruit of uniform maturity is marketed. This has the effect of avoiding the need for storage beyond the shipment period. (*See* **Fungicides**.)

Bananas are usually consumed within 3–4 weeks from the day of harvest, with no long-term commercial storage possible, as for citrus and pome fruits. It is possible to delay the onset of ripening by a few weeks by the use of modified atmosphere storage with high carbon dioxide (5%) and low oxygen (2%) and with ethylene scrubbers, such as potassium permanganate, but because banana fruit is available year round there is usually no advantage in doing this. However, in places where the expensive technology of refrigeration is not available, this technique may be valuable for marketing fruit in distant markets.

Once the fruit ripens, the period over which it may be eaten, its shelf-life, is relatively short. It is usually of the order of 2–10 days depending upon variety and ambient temperature. The extremely perishable nature of banana is associated with its high rate of metabolism – the rate of respiration during the climacteric is 100–180 ml of oxygen per kg of fruit per h and 40–60 ml kg^{-1} h^{-1} whilst green. This is higher than apples and pears, which have a respiration rate of 6–40 ml of oxygen per hour at similar temperatures. Lower temperatures will reduce the metabolic rate. A reduction of 10 °C usually halves the rate. Chilling injury occurs when the fruit is kept for a long period below 13 °C.

Fruit Utilization

Almost half of the bananas and plantains produced are eaten raw as a dessert fruit; the other half is cooked, usually by frying, boiling, roasting, or baking. Virtually all varieties of bananas and plantains may be either eaten raw when ripe, or cooked when either green or ripe. Cultural preferences govern the choices made.

Bananas can be processed in various ways so that they may be stored and utilized for other purposes. Fruits that are unmarketable because of small size or some peel blemishes are suitable for processing.

Banana purée is the most important processed product made from the pulp of ripe fruit. The purée is canned and used as an ingredient in dairy desserts, bakery items, drinks, processed foods, and sauces, and as a part of special diets in hospitals and nursing homes. Ripe bananas are also sliced and canned in an acidified syrup and are used in desserts, fruit salads, cocktail drinks, and bakery items.

Chips are made by deep-frying thin slices of unripe fruit, with the optional addition of various flavorings, and sold as a snackfood like potato crisps. Ripe banana may be dried (known as banana figs in some regions) and is said to store satisfactorily for over 10 years without the addition of preservatives. This would presumably be due to the high sugar content, which is in excess of 50%. The ripe fruit is ideal for ice-blocks (water ices, ice lollies), when peeled and frozen, with the optional addition of toppings such as chocolate and chopped nuts. It is an excellent base for icecreams because of its creamy consistency.

The other major processed products are flour made from dried unripe fruit, and essence which is extracted from the pulp of ripe fruit. When ripe fruit is fermented it makes a low-alcohol beer. Beer manufacture is confined to East Africa, with consumption being as much as 1.2 l per person per day in Rwanda. Less important processed products include a clarified juice, powder, jams, flakes, freeze-dried slices, ketchup 'filler', vinegar, and wine. Unripe bananas have been used to make starch, but none is currently produced.

Green banana fruit, pseudostems, and foliage are suitable as animal feed. They provide a source of energy and require supplementation with a protein source. Bananas are only economical as a source of animal feed when the livestock are nearby, because of the high cost of transport. The corms, shoots, and male buds find widespread use as an animal food in Asia and Africa.

Conclusion

Bananas have become the major fresh fruit consumed around the world, even in the temperate zone where they are not grown, despite the fact that they cannot be stored for more than a few weeks. This has been possible because they can be produced all year round, are competitively priced, come packed in an hygienic, easily opened peel, are extremely convenient to eat in virtually all situations and, as William Forsyth put it, 'The suave melting texture of a fully ripe banana combined with its distinctive mellow flavour makes a delicious combination.'

See also: **Celiac (Coeliac) Disease**; **Colon**: Diseases and Disorders; **Exercise**: Metabolic Requirements; **Fruits of Tropical Climates**: Commercial and Dietary Importance; **Fungicides**; **Infants**: Nutritional Requirements; **Ripening of Fruit**

Further Reading

Champion J (1963) *Le Bananier*. Paris: Maisonneuve et Larose.

Forsyth WGC (1980) Banana and plantain. In: Nagy S and Shaw PE (eds) *Tropical and Subtropical Fruit*, pp. 258–278. Westport, Connecticut: AVI.

Gowen S (1995) *Bananas and Plantains*. London: Chapman & Hall.

Hassan A and Pantastico EB (eds) (1990) *Banana – Fruit Development, Postharvest Physiology, Handling and Marketing in ASEAN*. Jakarta: ASEAN-COFAF.

Israeli Y and Blumenfeld A (1985) Musa. In: Halevy A (ed) *CRC Handbook of Flowering*, pp. 390–409. Boca Raton, Florida: CRC Press.

Jones DR (ed) (1999) *Diseases of Banana, Abaca and Enset*. Wallingford: CAB International.

Marriott J (1980) Bananas – physiology and biochemistry of storage and ripening for optimum quality. *CRC Critical Reviews in Food Science and Nutrition* 13: 41–88.

Soto M (1985) *Bananos: Cultivo y Comercialización*. Costa Rica: Litogratia e Imprenta LIL.

Stover RH and Simmonds NW (1987) *Bananas*, 3rd edn. London: Longman.

Turner DW (1997) Bananas and plantains. In: Mitra SK (ed) *Postharvest Physiology and Storage of Tropical and Subtropical Fruits*, pp. 47–83. Wallingford: CAB International.

Von Loesecke HW (1950) *Bananas*. New York: Interscience.

BARLEY

A W MacGregor, Canadian Grain Commission, Winnipeg, Manitoba, Canada

Introduction

Barley is one of the major cereal crops in the world. This article reviews its origin, botany, distribution, chemical composition, and uses as food and feed. Utilization of barley in the malting and fermentation industries (brewing, distilling, alcohol production) is covered in other articles.

Botany

Barley is a member of the grass family Gramineae, the subfamily Festucoideae, the tribe Hordeae and the genus *Hordeum*. Cultivated barleys belong to the subspecies *vulgare*, whereas the wild forms of barley belong to the subspecies *spontaneum*. The basic chromosome number of the genus *Hordeum* is 7, and all cultivated barleys are self-fertilizing, diploid annuals ($2n = 14$).

The spike or head of barley consists of a series of spikelets that are attached – in sets of three – at nodes to alternating sides of the rachis. Each spikelet contains a floret. In six-rowed barley (Figure 1a), each floret is fertile and develops into a kernel, leading to the formation of six rows of kernels. Only the central kernel in each triplet is symmetrical; the lateral kernels are twisted to varying degrees. Therefore, two-thirds of the kernels in a sample of six-rowed barley will be nonsymmetrical. In two-rowed barley, the lateral florets are not fertile, so that only the central floret in each triplet develops into a kernel (Figure 1b), leading to the formation of two rows of symmetrical kernels in the mature spike. All cultivated barleys are either two- or six-rowed, but some six-rowed cultivars appear to have only four rows of kernels; thus, reference is sometimes made to four-rowed barleys, although these are really six-rowed barleys.

The barley kernel consists of many different tissues; some of these are shown in Figure 2. Outer layers of the kernel and the area between the embryo and the starchy endosperm are shown in detail in Figure 2a and **b**. In botanical terms, the barley seed is that part of the kernel enclosed by the testa. Hence, the testa is often referred to as the seed coat. Surrounding the testa and fused tightly to it is the pericarp, the outer tissue of the grain or caryopsis. Since a caryopsis is a fruit, barley is a fruit containing a seed. The hull or husk completely surrounds the grain and adheres tightly to the pericarp in most barley cultivars. In some cultivars, however, the so-called naked or hull-less types, the hull is loosely attached to the grain and is removed during threshing. (*See* **Wheat**: Grain Structure of Wheat and Wheat-based Products.)

Origin, Adaptation, and Production

Neither the place of origin nor the progenitor of cultivated barley is known with certainty. There is archeological evidence that from about 7000 BC barley was cultivated in the fertile crescent of the Middle East. This area includes present-day Israel, Jordan, Syria, and Iran. There is evidence also that barley may have been cultivated even earlier than that in the Nile Valley of Egypt, but considerably later in India and China. It is probable, but by no means certain, that two-rowed *Hordeum spontaneum* is the ancestor of cultivated barley, but this subspecies

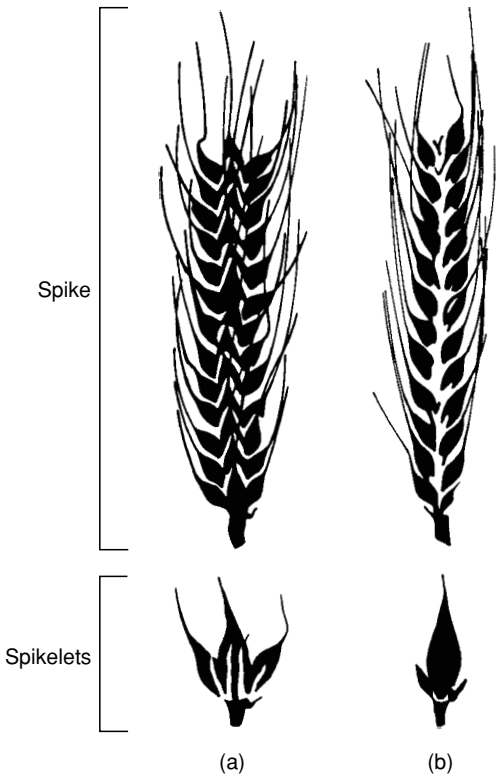

Figure 1 Barley spikes and spikelets. (a) Six-rowed barley. (b) Two-rowed barley. Reproduced from Barley, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

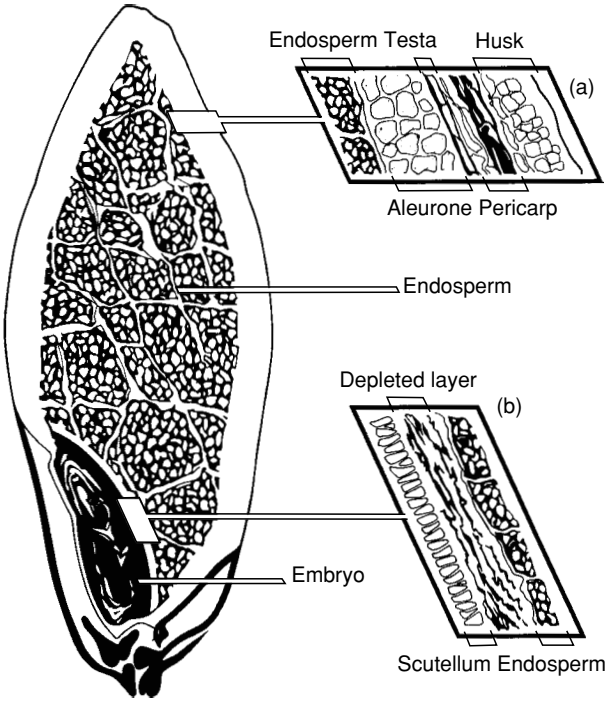

Figure 2 Diagram of a longitudinal section of a barley kernel. (a) Outer layers of kernel. (b) Junction between the embryo and endosperm. Reproduced from Barley, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 1 World production of barley

Country	*Average production (1991–2000) (thousand metric tonnes)*
Former USSR	39 935
Canada	12 870
Germany	12 397
France	9 551
Spain	8 537
USA	8 131
Turkey	7 150
UK	6 918
Australia	5 568
China	3 867
Denmark	3 859
Kazakstan	3 681
Poland	3 344
Other	25 551
Total	151 349

From *Canadian Grains Council Statistical Handbook* (2001). Winnipeg: Canada Grains Council.

may be an intermediate between the actual barley progenitor and cultivated barley. Six-rowed forms of barley are thought to have developed from early two-rowed forms through mutation and hybridization.

Barley is a widely adapted crop and grows under a wider range of environmental conditions than any other cereal. It appears to require fewer heat units than other cereals to reach physiological maturity and can therefore be grown successfully at higher latitudes and altitudes than other cereal crops. Barley is relatively tolerant to drought and to alkaline and salt conditions but grows best in temperate regions of the world where growing seasons are long, cool, and moderately dry.

Winter barley is less cold-tolerant than either winter wheat or winter rye. Nevertheless, in some countries with temperate climates, such as the UK, most of the barley is grown during winter. In countries such as Australia and Spain, which experience hot summers, a significant proportion of the barley is grown during the winter. Conversely, the harsh winters of North America preclude the growing of winter barley, except for some areas in the southern USA.

Over the past 10 years, world-wide production of barley has averaged 51.0×10^6 t (Table 1). This places barley fourth after wheat, maize, and rice in terms of total annual production. In 2000, barley accounted for 8.4% of the world's total cereal production. The former USSR is the largest producer of barley, but it is all consumed internally. Canada, the USA, and Australia each produce much smaller quantities but they, as well as members of the EU, are major barley exporters. (*See* **Cereals**: Contribution to the Diet.)

Composition

The major constituent of barley kernels is starch, which is present in the endosperm in the form of discrete granules (Figure 2) and represents, on average, 60–64% of the weight of the kernel (Table 2). Starch consists of two high-molecular-weight polymers of glucose: amylose and amylopectin. Amylose is composed of long chains of glucose residues linked by α-(1,→4) bonds with a few side-chains joined to the main chains through α-(1,→6) bonds. Amylopectin contains much shorter unit chains of α(1,→4)-linked glucose residues that are inter-linked through α-(1,→6) bonds. In other words, amylose is lightly branched, and amylopectin is heavily branched. In general, barley starch contains about 75% amylopectin and 25% amylose. However, starch from 'waxy' barley cultivars contains 95–100% amylopectin, and starch from high-amylose barley cultivars may contain more than 40% amylose. These starches have different functional properties, but they have not yet been utilized widely in the food industry.

The major component of endosperm cell walls is the polysaccharide, (1,→3), (1,→4)-β-D-glucan, which

Table 2 Barley composition

Component	*Content (%, dry weight)*
Starch	60–64
Arabinoxylans	4.4–7.8
β-Glucans	3.6–6.1
Cellulose	1.4–5.0
Simple carbohydrates (glucose, fructose, sucrose, maltose)	0.41–2.9
Oligosaccharides (raffinose, fructosans)	0.16–1.8
Proteins	8–15
Lipids	2–3
Minerals	2–3

Barley also contains small quantities of the B-complex vitamins, including thiamin (B_1), riboflavin (B_2), nicotinic acid, pyridoxine (B_6) and pantothenic acid, biotin, folic acid, and vitamin E.

forms highly viscous, aqueous solutions. These high viscosities may cause filtration problems during brewing and digestive problems when untreated barley is fed to chickens. Conversely, there is substantial evidence that ingestion of β-glucan can lead to lowered levels of serum cholesterol and of glucose and insulin responses. These are all beneficial responses. Arabinoxylans, which are polymers of xylose and arabinose, are present, in both soluble and insoluble forms, mainly in the cell walls of endosperm and aleurone tissues. They absorb many times their own weight of water and also form viscous solutions in water. Small amounts of cellulose are present in the outer tissues of barley kernels.

Only small amounts of the simple sugars, glucose and fructose, are present in the kernel; the major sugars are sucrose and raffinose, which are present mainly in the embryo. Varying amounts of fructosans, which are fructosyl polymers of sucrose, are also present in the mature kernel.

Barley contains a large number of proteins, which can be separated into four major groups on the basis of solubility. Albumins and globulins (15–30% of barley protein) are soluble in water and salt solutions, respectively; hordeins (35–50% of barley protein), the major storage proteins in barley, are soluble in aqueous alcohol solutions; glutelins (15–20% of barley protein), a mixture of storage and structural proteins, are soluble in alkali. Although aleurone cells are rich in protein, mainly globulins, most of the barley protein is in the endosperm, especially in the subaleurone region.

About 70% of barley lipid is present in the endosperm, 20% is in the embryo, and the remainder is in the outer layers of the kernel.

Minerals such as magnesium, sulfur, sodium, potassium, zinc, and calcium are concentrated in the outer layers of the kernel. Phosphorus is present, mainly, as phytic acid in the aleurone. Refer to Individual nutrients (*See* **Cereals**: Dietary Importance.)

Uses

Animal Feed

The largest use of barley, world-wide, is as animal feed, especially for cattle and pigs. Barley provides a good balance of high energy, because of its starch content, and reasonable protein quality and content. Hulled barley is less desirable than other cereal grains, such as maize, for chicken feed formulations because of its relatively high fiber content. There is increasing interest in developing hull-less barleys for feeding to pigs.

Malting, Brewing, and Distilling

The proportion of the total barley production used for malting varies widely among countries, but world-wide utilization is about 15% of total production. This aspect of barley utilization is discussed more fully in another entry. (*See* **Beers**: Wort Production; **Malt**: Malt Types and Products; Chemistry of Malting.)

Human Food

Less than 5% of total barley production is used for human food in most developed countries, but in some countries in the Far East, Middle East, and North Africa, barley products form an important part of the diet. The most common products are blocked, pot and pearl barley, barley flakes, and barley flour. Blocked barley is prepared by lightly scarifying the grain between abrasive disks to remove the husk. More severe treatment (pearling) results in removal of the outer kernel layers and rounding of the grain to produce pot barley, which amounts to about 65% of the starting material. Extensive pearling leads to removal of the outer endosperm layers and the embryo to leave pearled barley, which is only about 35% of the original barley. These products are used in soups, breakfast cereals, and stews, as tea or coffee substitutes after roasting, and as rice extenders. They can also be milled into flour and used in baby foods, porridge, and some baked goods when blended with wheat flour. Barley may be puffed by heating it under pressure to gelatinize the starch and then releasing the pressure suddenly. The loss in pressure allows rapid expansion of the water vapor and significant puffing of the grain. Barley flakes are made by cooking the barley, rolling the hot, moist product between heavy rollers, and drying the flakes to about 10.5% moisture content. (*See* **Cereals**: Breakfast Cereal.)

Nutrients are not distributed uniformly throughout the barley kernel. Vitamins, minerals and lipids are concentrated in the outer layers and embryo of the grain, and peripheral regions of the endosperm have significantly more protein than central regions. During pearling, therefore, the nutritive value of the product decreases as the outer layers of the kernel are removed. Pearl barley, for example, is nutritionally inferior to pot barley because it does not contain the embryo or any of the nutritionally rich outer layers of the kernel. Hull-less barley may offer some advantages over hulled barley in the food industry.

Sound, bright, uniformly plump kernels with a high test weight ($> 60\,kg\,hl^{-1}$) are preferred for all barley uses. In general, barley cultivars possessing yellow aleurones are also preferred. There is a small, decreasing demand for malts prepared from blue-aleurone barley.

Storage

The end-use quality of grain stored under high moisture conditions deteriorates through loss of viability and attack by insects and microorganisms. The temperature recommended for safe storage depends on the moisture content of the grain. Dry grain (12–13% moisture) can be stored safely at 17 °C, but grain at higher moisture contents deteriorates and should be dried with cool, dry air before storage. Barley intended for malting should not be dried at temperatures over 45 °C because the germination potential of the grain may be damaged. For food and feed uses, however, temperatures up to 60 °C may be used without impairing quality of the grain. Ideally, barley should be stored at about 12% moisture and at 15 °C or less. Grain may be treated with insecticides and pesticides, but some of these may leave unacceptable residues on the grain. It is possible to store grain under anaerobic or low oxygen conditions (e.g., under carbon dioxide or nitrogen), thus slowing down or inhibiting the development of microorganisms and insects. Such treatment does not impair the germination potential of the grain, nor does it leave residues on the grain. (*See* **Cereals**: Handling of Grain for Storage; Bulk Storage of Grain.)

See also: **Beers**: Wort Production; **Cereals**: Contribution to the Diet; Bulk Storage of Grain; Handling of Grain for Storage; Breakfast Cereals; Dietary Importance; **Malt**: Malt Types and Products; Chemistry of Malting; **Protein**: Food Sources; **Starch**: Structure, Properties, and Determination; **Trace Elements**; **Vitamins**: Overview; **Wheat**: Grain Structure of Wheat and Wheat-based Products

Further Reading

Bhatty RS (1999) The potential of hull-less barley. *Cereal Chemistry* 76: 589–599.

Briggs DE (1978) *Barley*. London: Chapman & Hall.

Canada Grains Council Statistical Handbook (2001) Winnipeg, Manitoba, Canada: The Canada Grains Council.

Henry RJ (1988) The carbohydrates of barley grains – a review. *Journal of the Institute of Brewing* 94: 71–78.

Hockett EA (1991) Barley. In: Lorenz KJ and Kulp K (eds) *Handbook of Cereal Science and Technology*, pp. 133–198. New York: Marcel Dekker.

Kent NL (1983) *Technology of Cereals*, 3rd edn. Oxford: Pergamon Press.

MacGregor AW and Bhatty RS (eds) (1993) *Barley: Chemistry and Technology*. St. Paul, MN: American Association of Cereal Chemists.

Rasmusson DC (ed.) (1985) *Barley*. Madison, WI: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America.

Shewry PR (ed.) (1992) *Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology*. Wallingford, UK: C.A.B. International.

USDA (1968) *Barley: Origins, Botany, Culture, Winterhardiness, Genetics, Utilization, Pests. USDA Agriculture Handbook 338*. Washington, DC: US Department of Agriculture.

BARRELS

Contents

Beer Making

E R Partington, Nickel Development Institute, London, UK

History

The art of brewing, which became highly developed in most English monasteries, can be traced back to Roman times. For 2000 years beers have been produced and stored in wooden vessels which have been lined with a variety of materials such as pitch to help seal them against leakage. Originally, beer was brewed to meet the needs of small communities and was consumed at the production site but, as demand for it grew, and beer had to be taken to more distant points of sale, transportable casks were required and these, too, were made from wood. The most common size of cask held one 'barrel,' a brewing unit which, in Medieval times, was a volume of 32 imperial gallons (145.5 l) but which is now standardized at 36 imperial gallons (164 l). This volume naturally gave its name to the cask of that capacity but was eventually adopted colloquially for all sizes of cask, despite their having their own names; the 4.5 gallon (20.5-l) 'pin,' the 9-gallon (41-l) 'firkin,' the 18-gallon (82-l) 'kilderkin,' and the 54-gallon (246-l) 'hogshead.' Wooden casks were made from vertical strips of oak, or staves, held tightly together by horizontal steel hoops. For this arrangement to be watertight, the staves were not only tapered so that together they created a circular cross-section but also bowed so that steel hoops could be forced down from the circular end to squeeze them together. This gave rise to the bellied shape of casks, which offered the practical advantages that even the hogshead, which weighed nearly one-third of a ton when full, could be easily rolled and steered along the ground with a stick or by gentle kicking. Then, when it needed to be lifted up to, and laid horizontally on, the rack (or 'stillage') in the customer's cellar, the bellied shape allowed the container to be rocked backwards and forwards longitudinally until it could be lifted smoothly on to its end and then swung completely over and on to the stillage (see Figure 1). There, it was stored until the natural conditioning processes were complete and the beer was ready for drinking. The belly also retained the yeast sediment, which settled during conditioning such that, even as the level in the cask fell, the beer was constantly drawn off from above the sediment, keeping it clear and 'bright.'

Eventually, with the advent of brewing on an industrial scale, metallic materials were introduced for the production vessels, but these were selected essentially for their strength. From the middle of the nineteenth century (with little recognition of the effects of corrosion and the resultant product contamination) process tanks were made out of soldered copper sheet and massive open fermenting chambers were lined with lead. Vessels were sometimes named after the materials that the industry had begun to use and the ingredients for beer are even today boiled in 'coppers.' However, until the mid-1900s, wood remained the only material commonly used for casks for storage, distribution and dispense, with the advantage that staves damaged by impact during delivery could be replaced individually by the cooper.

Although, from 1934, Flowers' Indian Pale Ale was, for a short time, exported in experimental steel casks, it was not for another quarter-century that metals became extensively used for the bulk packaging of beers. Stainless steel was introduced in the late 1950s for the smaller sizes of cask, the dimensions of which mimicked as far as possible those of the traditional oak ones so that they could operate side by side with them, and the weights of which were less than those of wooden casks of the same strength. In the early 1960s, aluminum alloys were introduced because they offered the advantage that, for an equivalent size and strength of container, they were approximately 30% lighter even than stainless steel.

Terminology

'*Beer*' is the generic name encompassing ales, lagers and stouts. '*Container*' is the generic term including all sizes of both casks and kegs (definitions below) (although all are often colloquially referred to simply as 'barrels').

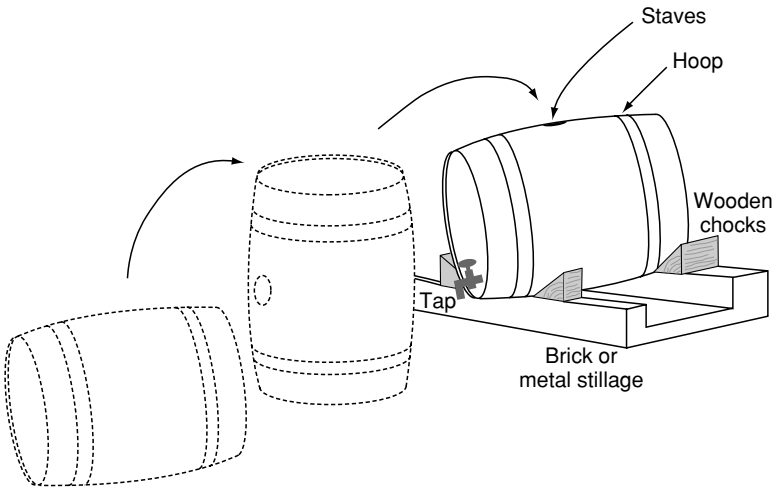

Figure 1 Cask of traditional beer, stillaged and tapped.

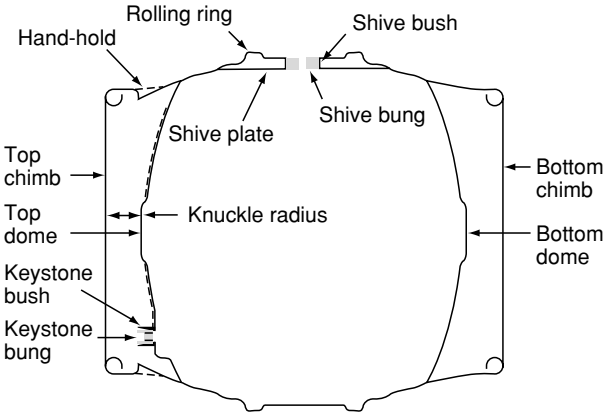

Figure 2 Typical cask for traditional beers.

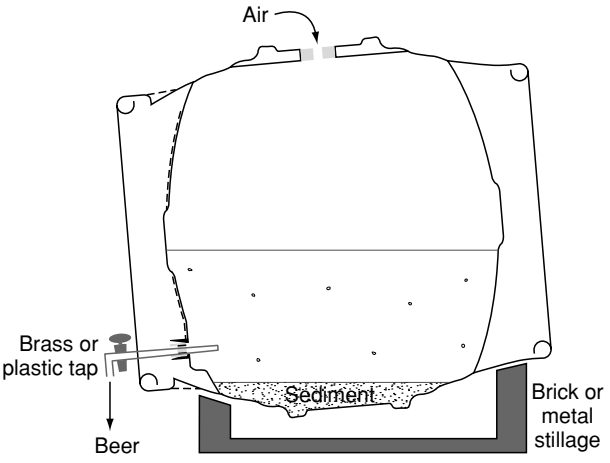

Figure 3 Dispense of beer from a traditional cask.

Traditional ales complete their conditioning in the containers in which they have been transported to the sales outlet. Typically, they have a low carbon dioxide content and can therefore be packaged into dual-aperture containers which, since at least as far back as 1727, have been called *casks*. Originally these were wood, but they are now almost all metallic (see Figure 2) and they are sealed only by tapered wooden (or, latterly, plastic) bungs driven into *shive* (inlet) and *keystone* (outlet) bushes. These bungs are removed when the empty cask returns for washing and are renewed before reuse. Figure 3 shows a cross-section through a cask during dispense.

Pasteurized beers (both ales and lagers) are conditioned in the brewery and have a relatively high gas content (which may be carbon dioxide or a mixture of this gas with nitrogen). In the packaging of these beers, the superior ability of metallic containers to contain a gas pressure becomes of paramount importance, as it is essential to maximizing the shelf-life of the product. Such beers are therefore packaged into single-aperture metal *kegs* (see Figure 4), first developed in the UK in the early 1960s, which incorporate a semipermanent '*extractor*' (or 'spear,' 'closure,' or 'valve'). This is commonly screwed into a '*Barnes Neck*' welded to the keg body (and named after its inventor, Australian Roy Barnes, who was employed by one of the major UK container manufacturers). The extractor remains in the keg whilst it is being cleaned, filled and subsequently emptied. It sits on a synthetic sealing-gasket in the neck and features two concentric, spring-loaded valves (see Figure 5), through the outer of which a gas pressure can be applied at dispense to force the beer up the downtube and through the inner valve to the dispense point on

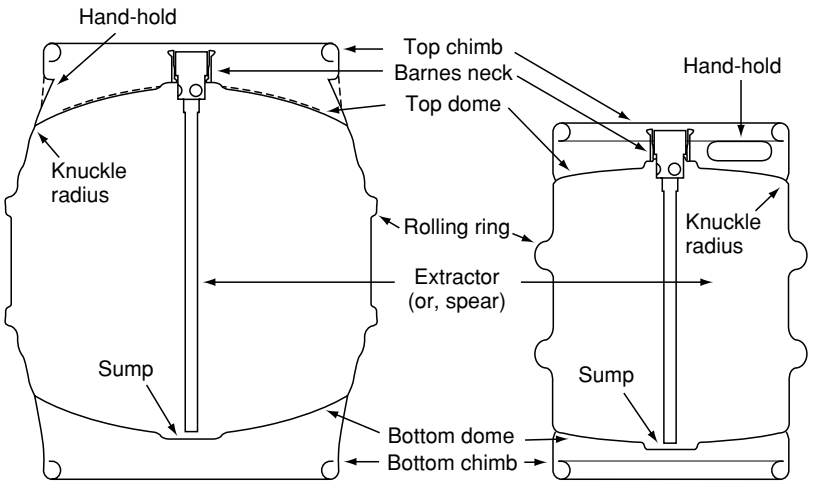

Figure 4 Typical kegs for pressurized beers.

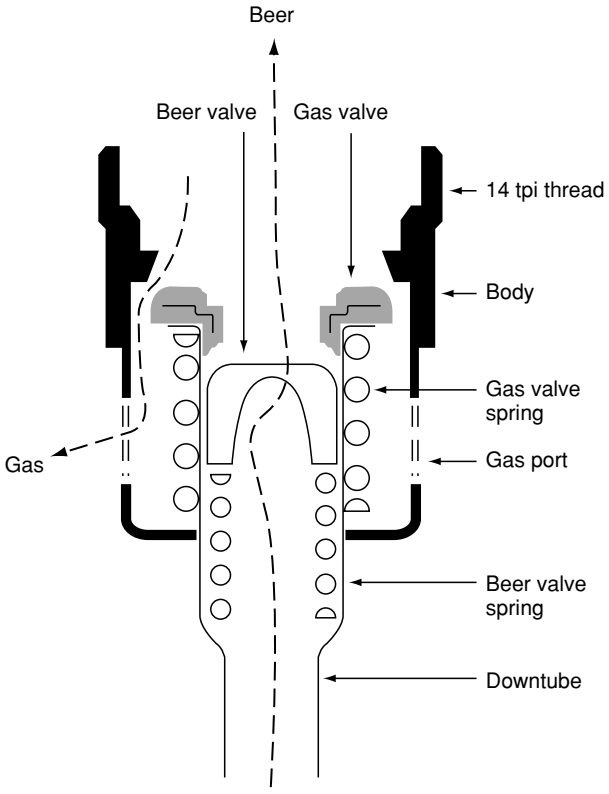

Figure 5 Typical extractor.

the bar (see Figure 6). It is because of this extractor that cleaning and filling can be mechanized and the costs of packaging significantly reduced.

Similar kegs and extractors (with slight design and material modifications to accommodate the more aggressive environment) are also used to package ciders.

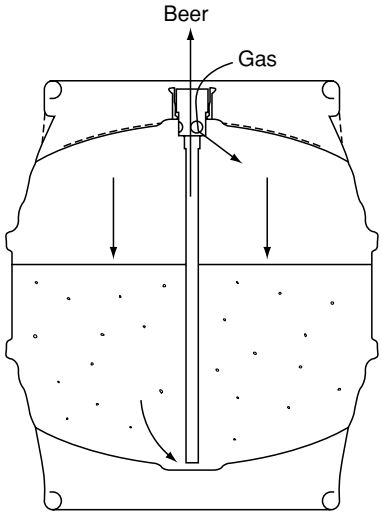

Figure 6 Dispense of beer from a pressurized keg.

Regulations and Codes of Practice

The then-Ministry of Agriculture, Fisheries and Food stipulated that materials must not 'react with, or alter the organoleptic properties of, the foods with which they come into contact.' This legislation has necessitated much research into materials, including the epoxy resins used to line aluminum containers, higher grades of stainless steels, and the synthetics used for the gaskets and valves of the extractors.

The UK Brewers' and Licensed Retailers' Association (now the British Beer and Pub Association) issues the instruction that all pressure kegs 'shall be tested at the manufacturer's works to at least 1.5 times their Safe Working Pressure,' this SWP being 'the maximum gauge pressure to which equipment

should be subjected and which must not be exceeded by any planned method of working.' It further stipulates that the 'maximum test pressure should not subject the material to stresses in excess of 90% of the minimum specified yield for the material [and] shall be maintained for a sufficient length of time to permit a thorough examination to be made of all seams and joints.' In practice, the industry voluntarily applies these same procedures to the manufacture of traditional casks and, in view of this self-regulation, beer kegs and casks are currently exempt from all EU legislation applicable to the design, manufacture and testing of pressure vessels.

Design Considerations

Existing Standards

Few design standards currently exist for beer containers. In the absence of any liaison between brewing companies or container manufacturers in the early days of container production, as many slightly different designs were created as there were customers. Even with the introduction by a significant number of UK brewers of the European cylindrical 11-gallon (50-l) stainless steel keg in the late 1980s, there was little industry-wide standardization of dimensions as each brewer's keg had to meet slightly different operational constraints, particularly those of compatibility with their preexisting packaging, handling, storage and transportation systems. In 1984, however, the major UK brewers came together to form the InterBrewer Technical Liaison Group (INTEL), and this body recommended procedures for materials selection and standards for the performance testing of kegs. In conjunction with INTEL, the UK Brewers' and Licensed Retailers' Association issued in the early 1990s detailed specifications for the two most common designs of Barnes Neck.

Compatibility with Existing Machinery and Equipment

In view of the extent to which the operation of both kegs and casks are now mechanized, it is essential that their designs are compatible with the plant on which they will be cleaned and filled (the 'washer/racker'), the machinery that will palletize them, the cradles, boards or pallets on which they will be stored, the road vehicles on which they will be transported and the dispense environment, including stillages for casks and extractors in kegs.

Strength

Almost all damage to kegs happens in the distribution cycle. Drop tests can simulate a container falling the 4 ft 6 in (1.4 m) from the bed of a delivery vehicle on to a concrete pavement, and evaluate the strength of the rolling rings and the chimbs (the end-rings on which containers stand when vertical and by which they are usually picked up), particularly when a container falls at 45° onto its handholds.

Testing the strength of domes can simulate static loads, such as 'topping.' This is the practice of stacking a small container horizontally on top of a larger, vertical container. Whilst this can save space on a delivery vehicle bed, it can also damage the Barnes Neck or the keystone bush of the lower container.

When a keg is to be cleaned and refilled in the brewery, the washer/racker effects a seal between the washing/filling head and the rim of the Barnes Neck by means of a pneumatically operated clamp. However, a 3-in (75-mm) air-ram at 50 psig (3 bar) exerts a force of nearly 500 lbf (0.25 tonne) axially on to the keg neck – when applied gently. If, however, the head impacts the neck suddenly, the effect can be equivalent to a far higher static force.

If, during the cleaning process, steam injected to purge the detergent is followed by rinse-water, the steam in the keg will condense rapidly to water, creating a vacuum. A hard vacuum applies a force of 700 lbf (300 kgf) to the end-domes of an 11-gallon (50-l) keg, which may cause them to collapse inwards.

Design features such as impressed stars (or, 'cruciforms') spanning most of the diameter of end-domes can significantly increase their resistance to deflection without any increase in weight (see **Figure 7**).

Weight

The strength of a container can be improved by increasing the thicknesses of its materials, but this will increase both its tare weight and its cost, and

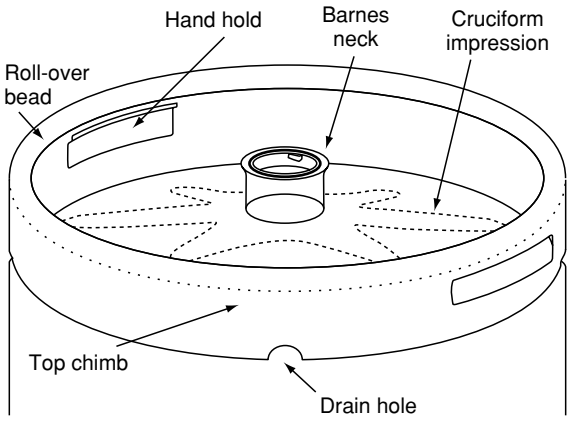

Figure 7 Top dome of a typical cylindrical keg.

will reduce road-vehicle payloads. The use of superior materials, such as half-hard stainless steels (some chimbs being rolled from sheet of over 1000 N mm^{-2}) and the addition of cruciforms will both permit higher specific strengths. However, it is advantageous for there to be as little variation as possible between different manufacturers' tare-weights because of the weighing scales at the end of the filling line in the brewery which automatically check for kegs not sufficiently filled.

In addition to consumer demand for a wider choice of beers at the point of sale, Manual Handling legislation is resulting in a trend towards a greater number of smaller (and, therefore, lighter) containers. Some aluminum hogsheads and barrels remain in service, but most are being phased out in favor of containers of 22 gallons (100 l) or less, which remain of an acceptable weight even in stainless steel. Table 1 shows the weights of typical metal casks. A keg of the same capacity will weigh about 1 lb (0.5 kg) more than a cask because of its extractor.

Volume

Even today, casks are filled manually through the shive bush. It is therefore possible to brim-fill them, and so the content of each cask is the same as its capacity.

Until 20 years ago, kegs were brim-filled in the upright position with the Barnes Neck uppermost, and so, again, the contents were always the same as the capacity. It is now, however, almost universal practice that the machine fills kegs in the inverted position, because the beer can be injected faster (and less turbulently) through the gas ports of the extractor than through the narrow downtube (see Figure 8). However, an inverted keg cannot be filled completely because of the gap between the tip of the downtube and the dome of the keg. There is always a mushroom-shaped gas space left, and this can account for as much as 0.5 pint (284 ml) in a 72-pint (41-l) firkin, and 0.75 pint (426 ml) in a 288-pint (164-l) barrel, depending on the shape of the dome and the distance between it and the tip of the downtube.

However, UK Trading Standards stipulate that, when a brewery fills a batch of containers all of a nominal capacity of (for instance) 18 gallons (82 l) and to be sold as '18-gallon casks' or '18-gallon kegs,' the actual contents of all of the containers in that batch must average 18 gallons (82 l) and each individual container must hold at least nominal-less-3% (17.46 gallons, 79.44 l) in the case of casks or nominal-less-2% (17.64 gallons, 80.26 l) in the case of kegs. Similarly, each individual 11-gallon (50-l) keg in a production batch must hold at least nominal-less-2% (i.e. 49 l). This 'Declaration of Contents' regulation has to be allowed for when specifying the capacities of new containers as containers can change in capacity over long periods of service. Aluminum containers with rolling-rings swaged out of the body tend to grow in length and capacity with time because the rings flatten out, but stainless steel containers tend to shrink by about 0.05% of their original capacity for every year in service because of all the small dents they accumulate. The design specification for the lower limit (i.e., nominal less manufacturing tolerance) of the volume of a container should therefore

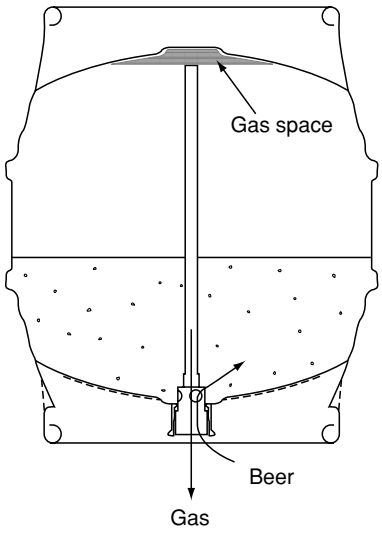

Figure 8 Filling a keg through the gas ports of the extractor.

Table 1 Weights and nominal capacities of typical metal beer casks

Cask			*Material*	*Weight*			
	Nominal capacity			*Empty*		*Full*	
	Imperial gallons	*(liters)*		*lb wt*	*(kg wt)*	*lb wt*	*(kg wt)*
Firkin	9	(41)	Stainless steel	25	(11)	119	(54)
Kilderkin	18	(82)	Stainless steel	51	(23)	236	(107)
Barrel	36	(164)	Aluminum	63	(29)	435	(197)
Hogshead	54	(246)	Aluminum	85	(39)	641	(291)

reflect both filling practice and in-service changes in shape usually by including an over-measure of approximately 1% of the nominal capacity.

Practicality

The profile of the bottom dome of a container can significantly affect both the volume of beer left in the keg after normal dispense and the effectiveness of the on-line deterging procedures. Too shallow a dome radius may result in too much beer being unextractable and, for this reason, many kegs incorporate in their bottom dome a small sump (or 'dimple') of approximately 3 in (75 mm) diameter and 0.25 in (6 mm) depth (equivalent to about one fluid ounce, or 25 ml) into which the tip of the extractor just reaches. However, if the shape of this sump is not carefully chosen, the detergents that are lanced into the inverted keg through the extractor downtube during the washing cycle may not spray evenly over all internal surfaces and clean them effectively.

Tight radii can create areas inside containers that are difficult to clean. Top chimbs usually incorporate small drain points just above the butt-weld, which attaches them to the top dome such that extraneous water drains away easily (see Figure 7).

Material Selection

Operating Environment

The conditions under which casks and kegs must operate influences the choice of the materials used for the bodies of the containers themselves, any protective interior linings they may have, the synthetics of extractor components and any plastics used for shive or keystone bungs.

Beer has a pH of about 4 when fresh, but this can drop to 3.5 or below if the beer is exposed to oxygen, such that it sours, as is inevitable in a traditional cask after dispense. Fresh ciders may have a pH as low as 3.3 and, when oxidized, even below 3. Stainless steel is generally impervious to these levels of acidity, but the oxide layer with which aluminum alloys protect themselves from corrosion is attacked by any pH less than about 4 or over 9. Aluminum alloy containers are therefore internally lined at manufacture by a sequence of steam-sealing, anodizing, and epoxy lacquering. However, if that lacquer lining is broken down (as may occur by impact to the keg during handling), not only can flakes of lacquer get into and jam the extractor valves, but also the keg itself can be corrosively attacked. This is most prevalent at exposed welds and can threaten the structural integrity of the container.

Typically, beer contains chlorides at concentrations up to 350 mg l^{-1} (p.p.m.) and sulfates at up to 300 mg l^{-1}. This environment might only be corrosive to stainless steels if combined with an abnormally high temperature (over about 55°C), which would itself be deleterious to the beer, and might affect areas such as rolling-rings, which retain high stresses from the manufacturing processess. However, for this reason, stainless steel containers should not be exposed to hot, salty conditions such as exist at the seaside in summer, even when empty.

Even during filling and dispense using mixtures of carbon dioxide and nitrogen, the pressures in kegs should rarely exceed 50 psig (3 bar). All containers made in Europe (whether kegs or casks) are designed for a working pressure of 60 psig (4 bar), and every one is tested at manufacture and after repair to 90 psig (6 bar). In practice, aluminum containers rarely fail at less than 300 psig (20 bar), and stainless steel containers commonly withstand pressures of 1000 psig (70 bar).

During the washing and refilling processes, steam at temperatures of up to 145 °C is used to sanitize kegs, and this has generally proved to be too hot for synthetics to be used as a material of construction for beer kegs. If steaming is immediately followed by a charge of inert gas to remove all oxygen before the new beer is added, the material of the container can suffer thermal shock from approximately 120 down to 0 °C. If steaming follows a cold-water rinse, the thermal shock can be from 20 to 140 °C. Such sudden changes in temperature can crack the epoxy linings of aluminum kegs, exposing the substrate to subsequent corrosion by the beer.

Commonly, hot 1% phosphoric acid is used to remove process soils from the interiors of metallic kegs and warm 4% phosphoric solution to remove normal dirt from their exteriors. Alternatively, a hot 2% caustic soda solution with ethylenedianine tetraacetic acid (EDTA) may be used to clean both the interior and exterior of stainless steel containers (but not aluminum containers as caustic soda is highly corrosive).

Some lubricants used on conveyors may embrittle synthetics.

Storage temperatures normally range between 0 and 25 °C. However, temperatures may fall to −20 °C if the container is left outside during winter, under which circumstances the 9% expansion of the water content of the beer as it turns to ice can create internal pressures in excess of 400 psig (27 bar); high enough to distend outwards the end domes of stainless kegs or burst most aluminum containers (especially at sites of corrosion) and all wooden casks. Such pressures in casks may be relieved by the shive

or keystone bungs being blown out, but this may not always happen, particularly if an ice-plug has formed beneath them first. The temperature of a container may rise to 60 °C if it is left exposed to strong sunshine for extended periods in the summer, and this can cause synthetics to soften.

In the event of a fire, the materials of containers should neither ignite easily nor support a flame.

When a container is damaged irreparably, or its design becomes obsolete, the material from which it is made must satisfy the Packaging and Packaging Waste Directive requirements for recyclability.

Materials selected

During the 1950s, attempts were made to develop containers that were stronger, cheaper to maintain and more hygienic than timber alone. A number of material combinations were tried, some more successfully than others:

- A thin stainless steel liner was encased in a wooden cask. This presented to the beer a more hygienic surface than timber alone, but the wood still broke during handling and the stainless steel liner could be dented easily.
- A stainless steel vessel was complete encased in a mild steel jacket. This was known as a 'Brown Bomber' and was strong but very heavy. External rusting of the jacket also presented a poor image of the contents.
- A thin stainless steel body had a pair of mild steel chimbs longitudinally bolted together to entrap the vessel. This was quite strong but, again, unacceptably heavy.
- A cylindrical stainless steel body had interference-fit galvanized mild steel chimbs with integral rolling-rings pressed on to each end. This 'Sunbrite' design, developed by GKN-Sankey, is still in service today.

In the 1980s and 1990s, synthetic materials were introduced:

- Polyurethane was used to jacket thin-walled stainless kegs, with the benefits of both having stainless steel in contact with the beer – but at a thinner, cheaper gauge than would be strong enough to withstand everyday handling on its own – and a synthetic exterior material that could not only be decorative and promotional but also support the stainless lining whilst making the keg considerably quieter to roll around than an all-stainless keg. However, these kegs were not only as expensive overall as all-stainless steel ones kegs, but they were also very difficult to repair, particularly if an impact dented them and delaminated the thin stainless skin from the plastic.
- Under development are all-synthetic kegs made from plastics that can withstand the high stresses imposed during manual handling and the high steam temperatures required to sterilize beer kegs before they are refilled, but which do not taint the flavor of the beer.

The two most prevalent materials now, however, are aluminum alloy and stainless steel. Aluminum alloys, borrowed from the aircraft industry for their lightness and strength, were introduced in the early 1960s. Initially, container bodies were cast, but these were found over time to deteriorate to the point where they could suffer fast fracture at normal working pressures, and so the Brewers' and Licensed Retailers' Association Code of Practice now prohibits the use of cast aluminum for pressure kegs. It may be used for the chimbs welded to the ends of the beer-containing section, but the bodies themselves are manufactured from sheet HE30, a 1% silicon, 0.7% magnesium, 0.6% manganese heat-treatable aluminum alloy. Any welded-on rolling-rings are manufactured from an extruded version of this alloy. Aluminum alloy containers have the advantage of a high specific strength but the disadvantages of high purchase and operating costs. The material requires heat treatment before and after both manufacture and repair and, to minimize its corrosion by the beer, a series of expensive processes including internal steam-sealing, anodizing, epoxy-lacquer spraying and stoving to provide a barrier between the alloy and the beer. However, this protection is still very susceptible to crazing as a result of impact damage during handling and delivery of the container. Furthermore, in the case of pressure kegs (see **Figure 4**), a special design of Barnes Neck is required to insulate the aluminum body of the keg from the stainless steel of the extractor components. It comprises an outer, aluminum bolster welded to the keg body and an interference-fit stainless steel insert (which carries in its bore a female thread for the extractor), the two being separated by an electrically insulating nylon sleeve. Failure of this sleeve, as is common in service, results in the creation of a galvanic cell, the beer acting as an electrolyte between the stainless steel and the aluminum, and this can increase the rate of general corrosion of the aluminum alloy some 30 times. There is also a significant trade in stolen aluminum beer containers, as the material is easily smelted. For these reasons, aluminum containers are almost exclusive to the UK and, indeed, up to the 22-gallon (100-l) size are being superseded by stainless steel containers.

Stainless steels are readily formed and welded, robust, totally inert in normal usage, impervious to the most effective internal and external detergents, simple and economic to repair and very safe. Deep-drawing of the bodies offers a higher strength for the same weight and at about half the cost per unit volume that was possible 10 years ago. The most common sheet material used is EN: X5 CrNi 18/10 (1.4301; AISI 304 – 17/19.5% chromium, 8/10.5% nickel), and there are now approximately 10 million stainless steel beer kegs and casks in service in the UK made from this grade. With its higher resistance to general pitting corrosion, EN: X5 CrNiMo 17/10/2 (1.4436; AISI 316 – 16.5/18.5% chromium, 10.5/13.0% nickel, 2.5/3.0% molybdenum) has been used where the environment was particularly aggressive, such as for ciders to which metabisulfite preservatives had to be added. The 1.4301 series is also used to make the Barnes Necks, but here it is important to ensure its compatibility with the materials of the extractors as two such stainless steel components, connected by fine (14 tpi) threads, may gall (or cold-weld) themselves together unless their individual chemical compositions and their surface topographies and finishes are specified correctly.

Manufacture and Repair

Processes and practices

Aluminum alloys must be solution-treated at over 600 °C before manipulation. Both stainless steel and aluminum alloy containers were originally manufactured with five major components (see Figure 9), to which the 'fittings' (Barnes Necks, shive bushes, and keystone bushes as appropriate) were attached. The center section of the body was formed from rolled sheet, longitudinally welded into a cylinder and then profiled. Two pressed domes were then circumferentially welded on and, to these, the two chimbs were also circumferentially welded. In the 1980s deep-drawing was introduced for both materials, discs being drawn into half-shells to form the center section, to which the two chimbs were attached, all three welds being circumferential (see Figure 10). For both materials, deep-drawing reduced the amount of welding necessary, but in the case of stainless steel it had the added advantage that the material was significantly work-hardened between the two knuckle radii, enhancing the strength of the cylindrical body.

After assembly, aluminum alloy containers are internally steam-sealed, anodized, and sprayed with an

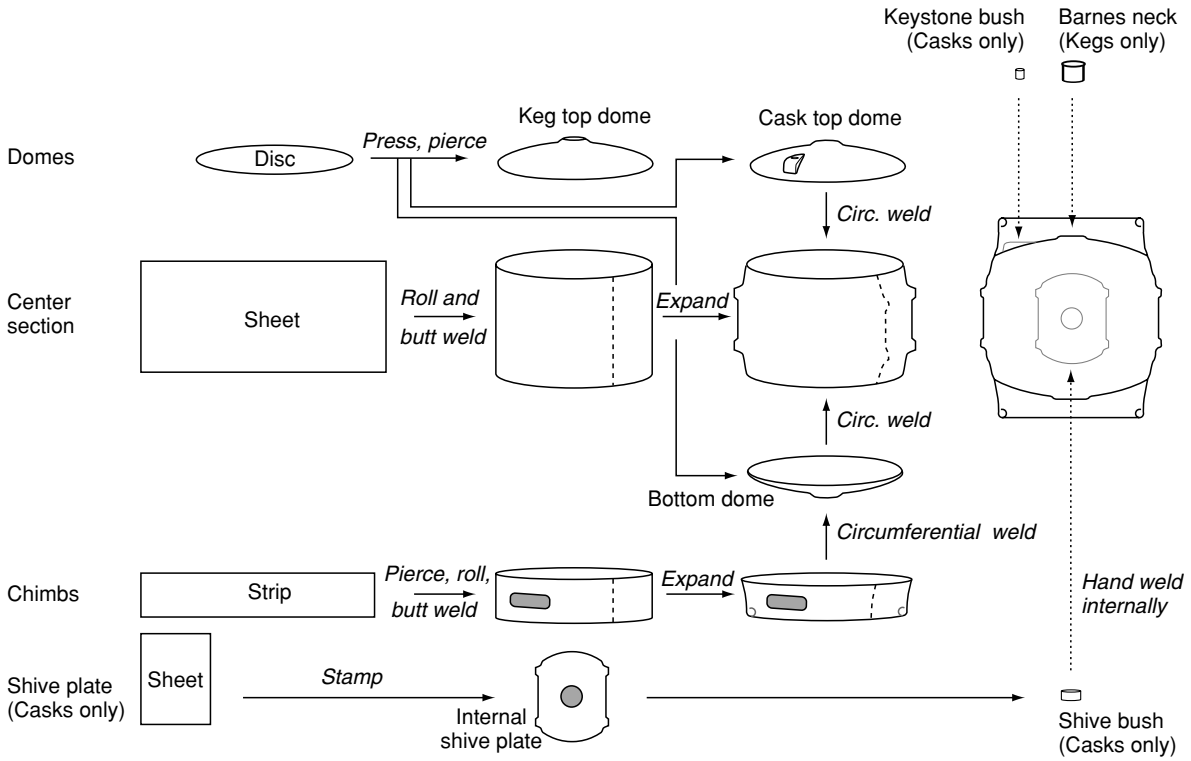

Figure 9 Manufacture of five-piece metal beer containers.

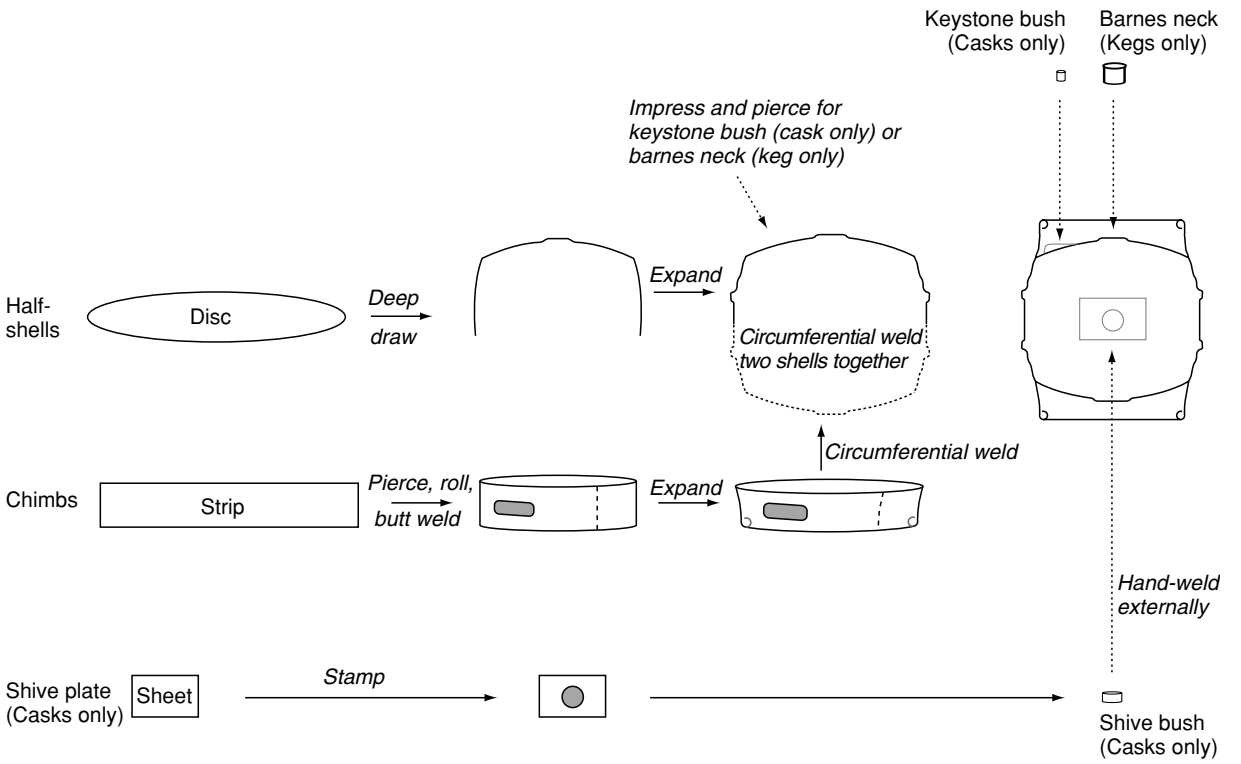

Figure 10 Manufacture of deep-drawn metal beer containers.

epoxy-resin lining. They are then heat-treated at approximately 190 °C both to cure that lining and to precipitation-harden the alloy. Stainless steel containers do not require lining, but must be pickled in a 3% hydrofluoric/10% nitric acid to descale the welds.

After manufacture, all containers are immersed in water and leak-tested to 40 psig (3 bar) with air, and then pressure-tested to 90 psig (6 bar) hydraulically.

Welding Standards

All welding of aluminum alloy containers is by Gas Metal Arc Welding (sometimes known as 'Metal Inert Gas' welding) which involves a filler wire. All welding of stainless steel containers is by Gas Tungsten Arc Welding (sometimes known as 'Tungsten Inert Gas' welding) which does not. Welding standards should control:

- the alignment of components (particularly at butt welds) and the maximum projection of the weld-bead into the beer (both to maximize cleanability);
- weld penetration (to preclude crevices that could be hygiene hazards);
- porosity (to maximize strength); and
- straightness of circumferential weld runs.

Performance in Service

Damage Repair

Although stainless steel does suffer more bending and indentation in service than aluminum alloy, it can be straightened again relatively easily and repeatedly. A cooperage (often sited beside the brewery washer/racker) can hydraulically pull a depressed keg top-dome back up and reround badly damaged chimbs. However, repairs involving capacity correction, pressure testing, welding or fitting of new parts are generally carried out by specialist companies external to the brewery.

All repairs to aluminum containers have to be carried out by specialist manufacturers/repairers, as the alloy requires re-solution-treatment to soften it before reshaping it and then re-precipitation-hardening afterwards.

Corrosion

Once the protective internal lining of an aluminum alloy container is flawed (as can result from impact damage during handling, thermal shock, or physical damage when fragments of the wooden shive or keystone bungs are spiked from inside casks), the material will suffer corrosion by the beer. Attack is

particularly prevalent at exposed welds (these effectively comprising cast material and therefore having a lower corrosion resistance than the parent plate), and ultimately this can threaten the structural integrity of the keg. In normal service, stainless steel kegs exhibit no corrosion.

The Future

To meet the continuing demand for traditional ales, most UK brewing companies still maintain large stocks of casks, these generally being of 4.5, 9, or 18 gallons (20.5, 41, or 82 l, respectively) capacity. However, existing 36-gallon (164-l) and 54-gallon (245-l) casks are generally being replaced by smaller casks, both because of their weight and because of the shorter shelf-life of traditional beers once stillaged and tapped in the cellar, which necessitates high sales volumes.

Existing 9-gallon (41-l) and 18-gallon (82-l) stainless steel kegs remain in common use, but are being supplemented by deep-drawn 30-litre, 11-gallon (50-l) and 22-gallon (100-l) stainless kegs.

Glossary

Barnes Neck A bush welded to the top dome of a metal keg to accept the extractor, which commonly screws into this neck with a 7 tpi or 14 tpi, 2 in. diameter thread (see Figure 4). Between the extractor and the neck is a synthetic sealing gasket to hold the gas in the keg.

Barrel A volume of 36 imperial gallons (163.7 l) which gives its name to a wooden or metal container holding that nominal volume of beer. However, the name 'barrel' is often applied colloquially to other sizes of beer or cider container.

Board An abbreviation of 'Locator Board'.

Cask A re-usable, dual-aperture storage and transportation container for traditional beers which are not ready to drink when they leave the brewery but must complete their conditioning both in these containers and at the sales outlet (see Figure 1). Traditional beers (sometimes referred to as 'real ales') typically have a low gas-content and can therefore be packaged into such containers which are sealed only by tapered bungs driven into their shive and keystone bushes.

Chimb An end-ring formed in a wooden cask by the tips of the staves and in a metallic container usually by a separate component welded to the body (see Figures 2 and 4). Chimbs provide stability when the container is stood upright and also protect the Barnes Necks of kegs.

Closure See 'Extractor'.

Containers The generic term for casks and kegs.

Cooper A craftsman who manufactures and repairs wooden beer containers essentially using hand tools.

Cruciform A star-shape impressed into each dome of a metal keg to enhance its strength; particularly its resistance to end-loads such as those caused by 'topping' or those applied by the 'washer/racker' (see Figure 7).

Dimple An indentation in the bottom dome of a keg to minimize the volume of beer or cider which cannot be extracted. May also be referred to as a 'sump'.

Extractor A bi-directional valve unit which is fitted into the Barnes Neck of a keg and which remains in place whilst the keg is being cleaned, filled and subsequently emptied (see Figure 4). It features two concentric, spring-loaded valves (see Figure 5), through the outer of which a gas pressure can be applied at dispense to force the beer up the downtube and through the inner valve to the dispense point on the bar (see Figure 6). May also be referred to as a 'spear' or 'closure' or 'valve'.

Firkin A beer or cider container with a nominal capacity of 9 imperial gallons (40.9 l).

Handhold An aperture pierced through the chimb of a container to facilitate its being manhandled (see Figures 2 and 4).

Hogshead A beer container with a nominal capacity of 54 imperial gallons (245.5 l).

Hoop A steel band forced down from the end of a wooden cask towards its belly in order to contain and compress the staves and ensure their water-tightness (see Figure 1).

Keg A metallic single-aperture storage and transportation container for brewery-conditioned beers or ciders which are ready to drink immediately upon arrival at the sales outlet (see Figure 4). Because such beers and ciders typically have a higher gas-content than traditional ales, they depend upon the pressure-integrity of a metal container to maintain their condition.

Keystone A wooden bung driven into the outlet bush of a cask (see Figure 2). It is replaced each time the cask returns to the brewery for washing and re-filling.

Kilderkin A beer container with a nominal capacity of 18 imperial gallons (81.8 l).

Knuckle The radius between the dome and the wall of a metallic container.

Locator Board A pallet-like component made of timber or plastic on which four or six containers are stacked on end in a rectangular array. Another board is placed on top of the containers and another array of containers on top of that, and so on. Two or three layers of containers at a time (according to their size) may be moved by a fork-lift truck which is fitted with special tines which clamp the sides of the containers

with a horizontal pincer-action. In the warehouse, stacks of containers will comprise at least six layers.

psig Pounds force per square inch (gauge) – i.e.: above atmospheric pressure.

Pin A beer container with a nominal capacity of 4½ imperial gallons (20.5 l).

Rolling rings Rings around the belly of a container transverse to its longitudinal axis to provide integral 'wheels' to facilitate its being rolled in a straight line. They are formed either by swaging them out from the material of the wall of the container or by welding on separate components (see Figures 2 and 4).

Roll-over bead The curling-over of the extremity of the chimb of a metallic container to form a bead which provides a comfortable and safe area to grasp the container, add strength to the chimb and offer a firm ring on which the container can be stood (see Figure 7).

Shive A wooden (or, latterly, plastic) bung driven into inlet bush of a cask (see Figure 2). It is replaced each time the cask returns to the brewery for washing and re-filling.

Spear See 'Extractor'.

Staves Shaped strips of oak, longitudinal to the axis of a cask and held tightly together by horizontal steel hoops (see Figure 1).

Stillage A wooden, brick or concrete cradle in the cellar of a public house on which casks are laid essentially horizontally while the conditioning process is completed and the sediment of their traditional ale settles into the belly of the cask, leaving the beer clear and ready to drink (see Figures 1 and 3).

tpi Threads per inch.

Topping The practice of stacking a small container which is horizontal on top of a larger container which is vertical (i.e.: standing on one end) in order to save space on a delivery vehicle bed.

Washer/racker The item of plant in a brewery which automatically cleans and re-fills metallic beer kegs. It manipulates the kegs and locates emptying- and filling-heads onto their Barnes Necks by means of pneumatic rams.

Valve See 'Extractor'.

See also: **Beers**: History and Types; Quality in Process Control

Further Reading

BLRA (1990) *Code of Practice for the Dispense of Beer by Pressure Systems in Licensed Premises*.

BLRA Technical Circular 249.

INTEL Procedure IP1: Selection and Certification of Materials for Beer and Cider Containers and their Extractors.

INTEL Procedure IP2–2: Test Procedure for the Evaluation of Proposed Designs of Beer or Cider Container or of the Effects of Design Modifications.

MAFF Statutory Instrument 1523. Materials and Articles in Contact with Food.

Peuch JL and Moutounet M (1993) Barrels. In: Macrae R, Robinson RK and Sadler MJ (eds) *Encyclopaedia of Food Science, Food Technology and Nutrition*, pp. 312–317. London: Academic Press.

Wines, Spirits, and Other Beverages

J R Mosedale and J -L Puech, UMR Sciences Pour L'oenologie, Montpellier, France

Introduction

The use of oak wood for making barrels is primarily explained by its mechanical properties (strength, hardness, and flexibility) and low permeability to liquids. Nevertheless, contact with oak wood also strongly influences the flavor of barrel-stored wines and spirits. This article principally covers oak wood properties, methods of barrel construction, the organoleptic effects of oak wood, and its use for aging wines and spirits.

Oak Wood

Types of Oak

Oaks belong to the genus *Quercus*, which is divided into two subgenera, *Cyclobalanopsis* and *Euquercus*. The subgenus *Euquercus* is divided into six sections. The section *Lepidobalanus* is widely distributed, being found in Europe, Asia, North Africa, and North America, and includes the important species used in cooperage.

In Europe two species provide wood for the construction of barrels: *Quercus robur* (also known as *Q. pedunculata* or pedunculate oak) and *Q. petraea* (*Q. sessiliflora* or sessile oak). In North America, 10 species of white oak are reported to be used for oak cooperage, the most important of which is *Q. alba*. While additional oak species may be suitable for cooperage it is unlikely that their use will ever become anything other than a novelty.

Different ecological preferences characterize the two European species. *Q. robur* is a faster-growing,

pioneer species and grows best in fertile soils with good water supply. *Q. petraea*, however, is more drought-resistant and better adapted to dry, acid soils. The geographical range of *Q. robur* extends further east than *Q. petraea*, as far as the Ural mountains. However, the distribution of these species in much of Europe no longer reflects their ecological preferences due to considerable transportation and indiscriminate planting of seeds. Within most of Europe, both pure and mixed forests of the two species are observed and there is strong evidence indicating that some degree of interbreeding between the species may occur.

American oak cooperage comes from central and eastern USA, mainly from Indiana, Tennessee, and Missouri. While the dominant species *Q. alba* is found across all states, certain oak species have very localized distributions.

Wood Structure

Two different wood zones are clearly distinguishable within a cut oak tree. Beneath the bark and cambium is a layer of lightly colored sapwood generally representing 10–20 years of growth. The rest of the tree is composed of older, darker-colored heartwood and it is only from this part that wood is cut for cooperage. The transformation of sapwood into heartwood is characterized by the laying down of tyloses – membraneous growths that block large vessels and reduce wood permeability. Some oak species, notably red oak (*Q. rubra*), lack tyloses and are therefore not suitable for cooperage. Heartwood formation is also characterized by cell death, the removal of starch, and the laying down of diverse extractible compounds in the dead cells of the heartwood. Despite the lower permeability and greater durability of heartwood, the woods are structurally very similar.

Oak wood is known as a 'ring porous' wood, with each year of growth defined by clearly visible annual rings. Each of these rings consists of two types of wood: the early (or spring) wood and late (or summer) wood (Figure 1). The early wood is predominantly made up of large vessels and is laid down early in the growing season. The width of early wood tends to vary little between growing seasons. In contrast, the width of late wood, made up of mostly wood fibers and small vessels, varies according to the growth rate. Thus, the relative proportion of early and late wood is highly dependent upon the growth rate of the tree, with fast-growing trees containing a higher proportion of late wood.

Heartwood Chemistry

Wood has three main structural components: lignin, cellulose, and hemicellulose. These complex polymers are mostly insoluble in alcohol–water solutions. The remaining 5–12% of wood mass is composed of extractible compounds. This fraction consists mainly of the ellagitannins with a variety of other compounds belonging to diverse chemical classes (Figure 2).

The concentration of particular heartwood compounds varies among the different woody tissues. Furthermore, certain compounds undergo gradual transformation over the lifetime of the tree. The ellagitannins become increasingly insoluble due to

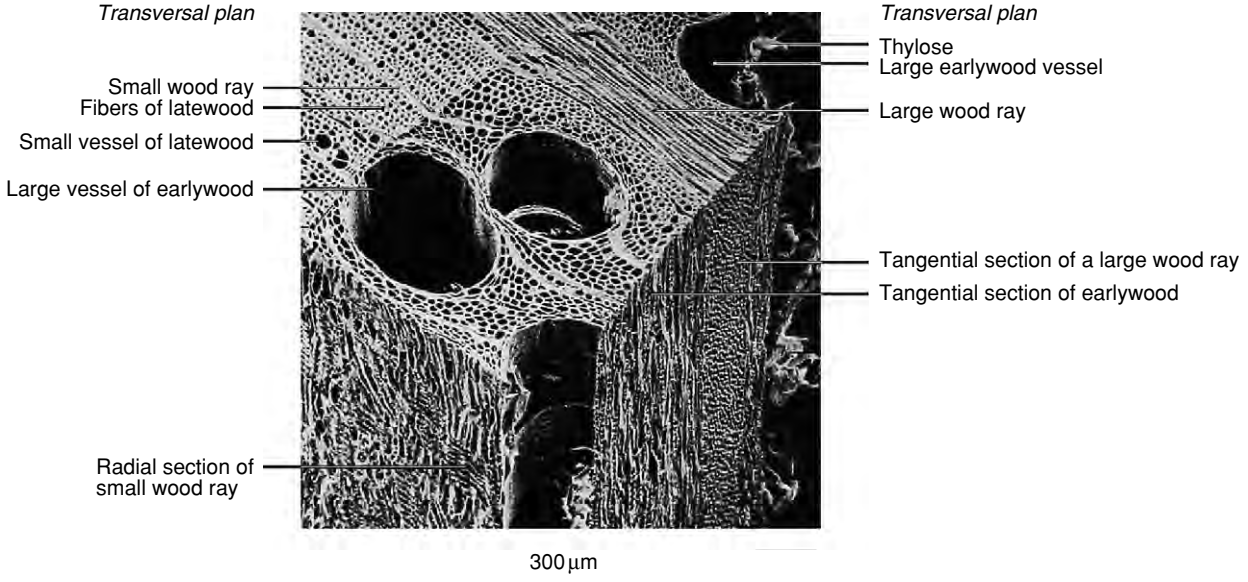

Figure 1 Oak wood anatomy.

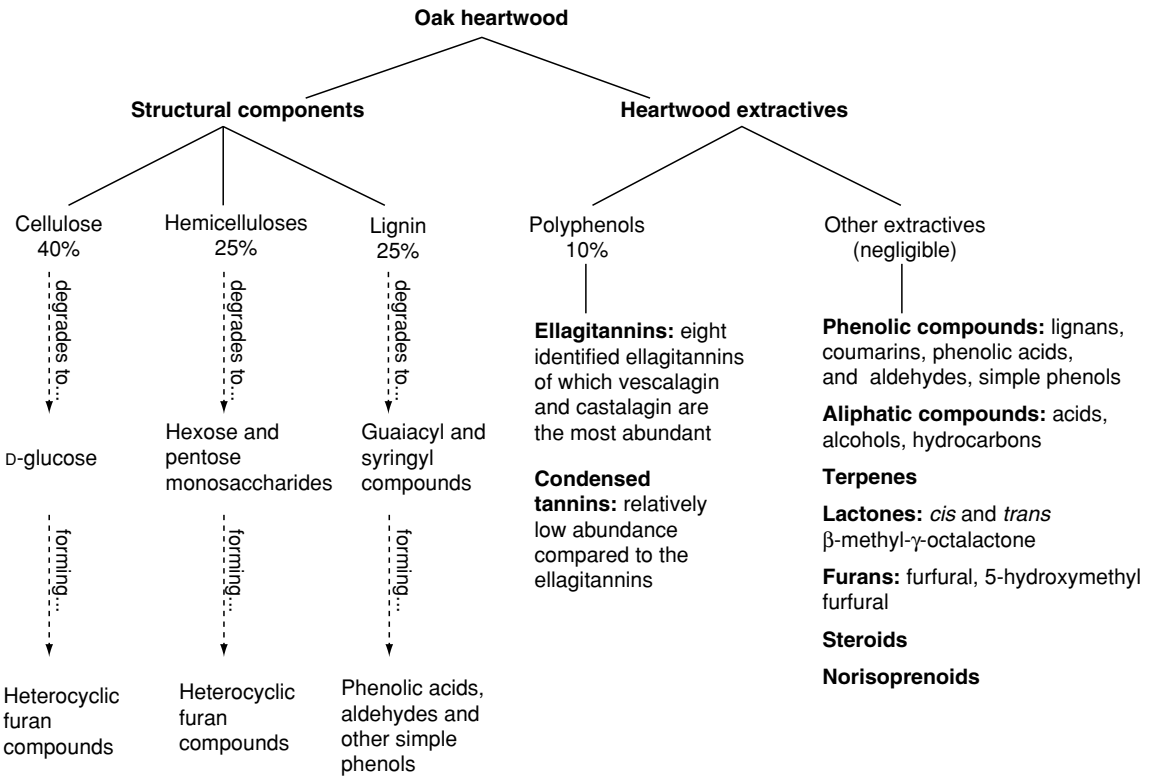

Figure 2 The major chemical constituents of oak heartwood. Percentages indicate the percentage of wood mass made up by the constituent.

polymerization and therefore levels of soluble polyphenols decline in the central, older heartwood.

Method of Manufacture

Barrels are made in several stages: preparation of the wood, making the staves, 'raising' the barrel, bending staves by heating, beveling, grooving and placing heads, hooping, and checking for defaults (Figure 3).

The Cooperage Industry

The two main sources of oak cooperage have traditionally been France and the USA, although recent years have seen a growing supply of oak wood and casks from eastern Europe.

In France, the cooperage industry is one of the most important purchasers of top-quality oak wood, with 250 000 m^3 of oak timber used in 1997, representing 400 000–450 000 barrels of an average capacity of 220 l. Over half of this production is destined for export. The cognac and armagnac market accounts for about 10% of barrel production, with the rest destined for wine aging.

In contrast, the much larger American cooperage market is dominated by the spirits industry. The much larger scale of production has also led to greater automation in cooperage methods. Nevertheless, smaller cooperages, often affiliated to French *tonnelleries*, are also found and cater in particular to wine makers.

Selection and Cutting of Stave Wood

High-quality wood needed for barrel-making is cut from trees with a girth of at least 35 cm and generally 100 years old or more, depending on growth rate. Perfect straight-grained logs are used, with the timber selected for its mechanical properties, its suitability for shaping, and its porosity.

European oak is traditionally split along the grain into bolts, before quarter-sawing staves (Figures 4 and 5). The finished staves are about 3 cm thick. This method of cutting is considered to produce staves possessing the best mechanical properties and to improve the impermeability of barrels. However, 5 m^3 of logs is required to produce 1 m^3 of stave wood.

American logs are rarely initially split into bolts, but rather staves are directly sawn from cut logs. While this method is both quicker and less wasteful of wood, it is not considered suitable for European wood. The greater abundance of tyloses observed in most American oak may render it less permeable than

Barrel manufacture

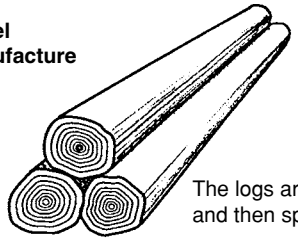

The logs are cut to the desired length and then split into bolts

Splitting

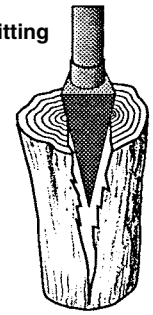

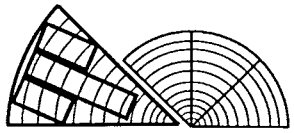

The bolts are cut into stave wood

Drying

The wood is dried outdoors exposed to the weather for 3 years.
It is sometimes kiln-dried after air drying...

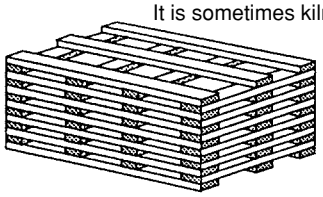

...the staves are shaped

Planing and hollowing **Shaping and jointing**

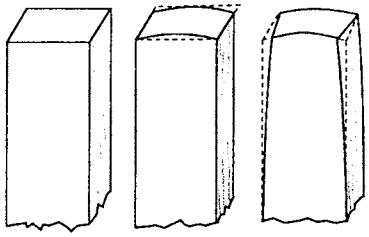

Raising the barrel and bending

This is carried out with a windlass or a bending machine...

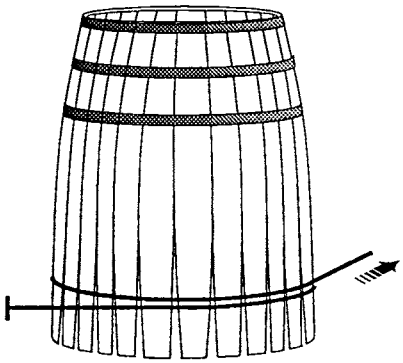

...the wood is dampened outside and heated inside

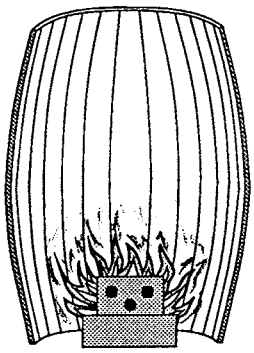

The European technique
Bending with a wood-fired brazier for about 20 min followed by further heating:

5–10 min: light heating
10–15 min: medium heating
15–20 min: heavy heating

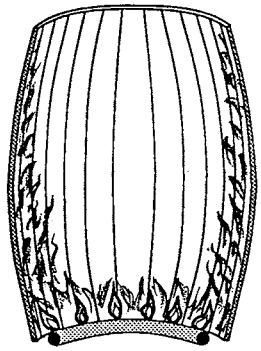

The American technique
Steam bending followed by charring with a gas burner:

15 s: light char
30 s: medium char
45 s: heavy char

Making the heading pieces

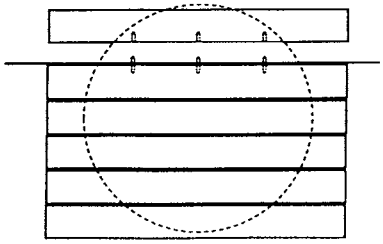

The heads are made up of 7–9 boards assembled with dowels. Strips of reed make them liquid-tight

The final hoops are fitted and the barrel is tested with hot water

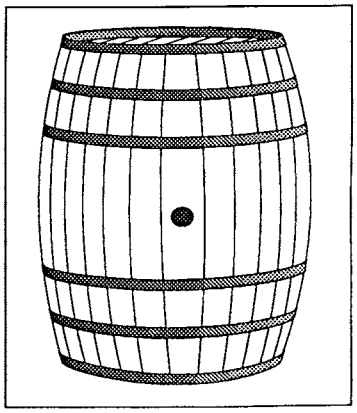

Figure 3 Barrel manufacture.

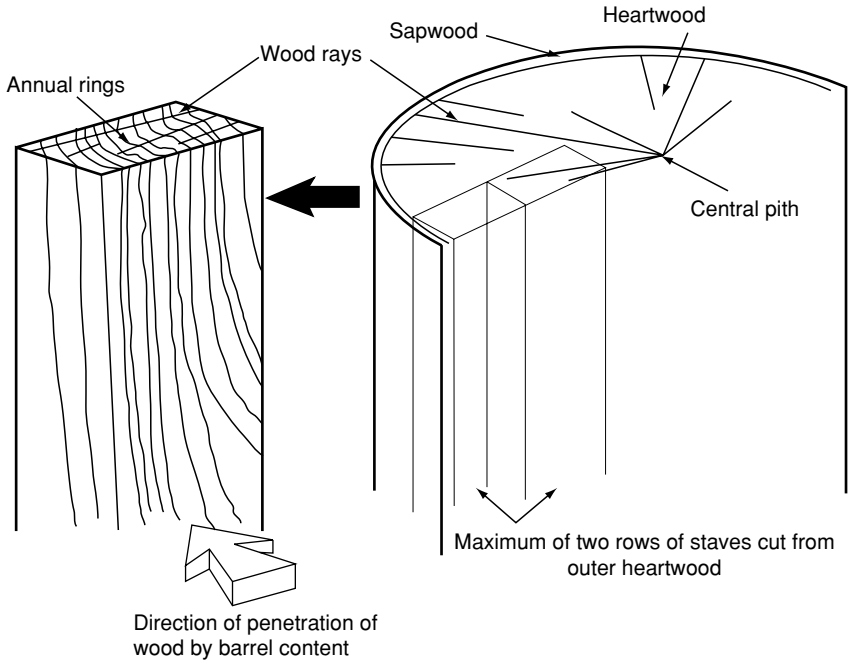

Figure 4 Quarter-sawn stave: orientation and location in the bole of the tree.

European oak wood and therefore this reduces the importance of careful splitting and cutting of barrel staves.

Drying

Wood that is dried below what is termed the fiber saturation point (about 30% water content for oak wood at 20 °C) undergoes shrinkage. The degree of shrinkage varies along the different dimensions of the wood, with radial and tangential shrinkage being much greater than longitudinal shrinkage. Oak used for barrel construction, therefore, is dried to 14–18% water content to insure that any further moisture loss will cause little dimensional change.

In France, oak wood is traditionally dried (or 'seasoned') using drying stacks (Figure 6) in the open air for about 3 years; it is considered that drying requires 1 year for each centimeter of stave thickness. During this period, changes may occur in the physical, mechanical, and biochemical properties of the wood. These modifications are influenced by the initial properties of the wood (e.g., density) and the stacking technique and climate. During air-drying, a diverse fungal flora also develops on the surface of staves, down to a depth of 3–4 mm.

An alternative drying method is by forced-convection kiln-drying. As oak is one of the most difficult types of wood to dry, the conditions of kiln-drying are very gentle (40–65 °C) to avoid any physical degradations (deformations, splitting, wood collapse). Despite the difficulties of kiln-drying oak, the greater flexibility and speed are clearly an advantage compared to the long process of air-drying. It is common for wood to be initially air-dried for 12 months or so and then kiln-dried according to need. An alternative kiln-drying method, used primarily in the USA, consists of rapidly kiln-drying the wood to a reduced moisture content of between 25 and 40%. The drying is then completed by traditional kiln or air-drying.

Barrel Assembly

An average of 28–32 staves and 12–16 head pieces are used for the construction of a barrel. The first phase of construction involves the preparation of barrel staves. After being cut to the same length, they are planed, hollowed, shaped, and jointed. Planing renders the outer surface of each stave convex. The inside of the stave is hollowed to give the barrel a regular, concave shape. The staves are shaped to make them narrower at the ends than in the middle. A typical stave is 27 mm thick at the top and 24 mm thick in the middle (the bilge). The next step of construction is 'raising' the barrel, when the staves are assembled in a 'truss hoop' before tightening.

Bending

This is a very important part of the barrel-making process as it governs, to a great extent, the quality and longevity of the barrel; poor heating can lead to defective products. In Europe, bending is traditionally

Figure 5 Splitting into bolts.

Figure 6 (see color plate 3) Stave wood drying in the open air.

Figure 7 (see color plate 4) A barrel placed over a brazier for bending.

Figure 8 (see color plate 5) Preparation for bending with a brazier.

performed over a wood fire, whereas steam is most frequently used in the USA.

In the traditional wood fire technique, the cooper burns oak chips and off-cuts in a brazier to make a fire inside the assembled barrel (Figure 7 and 8). The barrel may be turned to insure even heating of all the staves. As the barrel begins to grow hot, the cooper windlasses the base of the barrel while continuing to feed the fire. He takes care to heat the central part of the staves so that they bend without breaking and he may cover the barrel shell to concentrate the heat inside.

The wood is heated to temperatures of up to 200 °C during the operation. Such temperatures modify the thermoplasticity of certain polymeric constituents of wood (polyoses and lignin), allowing it to bend without splitting or breaking. The inside and outside of the barrel are dampened during bending to prevent warping and difficulties when the heads are inserted.

The barrel is then turned over and new hoops fitted. They are hammered on strongly while the barrel is still hot to tighten the staves and make the

joints perfectly liquid-tight. Once the barrel has been formed, coopers may continue heating the barrels (Figure 9).

Toasting and Charring of Barrels

The degree of additional heating (or 'toasting') influences the organoleptic qualities of the wood. Both the duration and intensity of heating may vary. Light and medium heating differ mainly in the duration of heating (5–10 min), with temperatures generally varying from 150 to 200 °C. Intense heating (heavy toast) occurs at higher temperatures and is stopped when exothermic reactions begin, that is to say at about 280 °C, as the barrel may catch fire at higher temperatures. Traditionally, only the barrel staves are heated. The heads, forming nearly 25% of the barrel surface in contact with the liquid contents, do not generally receive heat treatment.

In the USA, most barrel staves are more intensely treated to produce a layer of char on the inner surface of the barrel. After being assembled by steaming, barrels are then put over firing pots to dry the staves and remove the surface moisture acquired in the steam box. Next, they are sent to the charring fires where the inside of each barrel is treated for 15–45 s according to the depth of char desired. Barrel heads are also charred on a heading char machine.

Final Assembly

The heads are assembled with wooden dowels; a length of reed is fitted in each joint to give a good seal. They are cut to the required length and fitted to each end of the barrel. The barrels are crozed – a groove is cut in the ends of the staves to hold the head.

The truss hoops used during assembly are removed and replaced by the final hoops. The bung hole is drilled and boiling water or steam is injected to check for any leakage.

Figure 9 (see color plate 6) Heating barrels.

Organoleptic Properties of Oak Barrels

The cooperage industry used to manufacture barrels in chestnut or other woods such as acacia for transporting beverages or dry goods. Today, oak casks are used exclusively for the aging of wines and spirits and a few other food products (e.g., certain vinegars and sauces). The barrel plays an important role in the final flavor of the wines and spirits that it contains. (*See* **Wines**: Types of Table Wine; **Whisky, Whiskey and Bourbon**: Composition and Analysis of Whisky.)

The Effect on Flavor of Wood Constituents

The barrel directly influences flavor through the following mechanisms:

- direct extraction of soluble oak wood compounds
- partial disintegration of structural macromolecules and extraction of the resulting compounds
- transformation of extracted compounds and interactions with other wine or distillate compounds

Only a small fraction (4–5%) of lignin slowly degrades through ethanolysis and is extracted from the wood. In contrast, 55% of phenolic compounds in the wood are extracted. The most important polyphenol compounds extracted are the ellagitannins, yet these are not found in mature spirits and only very low quantities are found in oak-aged wines. Recent research has revealed that these molecules react with ethanol and form hemiketal or ketal derivatives. These derivatives have astringent properties and oak polyphenols are also responsible for the color of oak-aged spirits. Nevertheless, the exact role of polyphenolic compounds in the aging of wines and spirits remains poorly understood.

The principal volatile compound extracted from oak wood is *cis*- and *trans*-β-methyl-γ-octalactone (also known as oak or whisky lactone), and the quality of both wines and spirits has been found to be correlated with levels of this compound. Other compounds which are extracted from the wood and which may affect the characteristics of wines and spirits include volatile phenols, furans, and pyrazines. The levels of these compounds, however, are generally below individual flavor-threshold values and many are only formed in abundance as a result of the heat treatment of barrels.

The oak barrel also indirectly influences the flavor of aging beverages. The barrel defines the physical conditions of storage which will influence, among other factors, the evaporative losses of water, alcohol, and other volatile compounds. The layer of char in barrels used for aging spirits serves as an adsorbent for volatile substances formed during distillation. Storage in barrels also results in special redox

conditions, with the wood allowing continuous, if limited, interaction between the wine and the outside atmosphere.

Oak constituents of the wood may also indirectly influence flavor by chemical or physical interactions with wine and spirit constituents, influencing conditions such as pH, oxygen availability, the solubility of other volatile compounds, and the catalysis of various reactions. During the maturation of spirits, the gradual evolution of distillate compounds over the years is as important for the flavor of the final product as any direct effect of the wood.

The Influence of the Type of Wood and Barrel

Oak wood is a heterogeneous structure, the properties of which are highly variable. Furthermore, the wood used for barrels derives from different countries, species, and forests. In addition to this 'natural' variation, the method of construction also influences the properties of barrels. Different methods of wood drying and particularly barrel heating have important effects on barrel properties. It is therefore not surprising that the flavor effect of barrels has been found to be highly variable.

Natural variation of wood properties Different species of oak are characterized by different levels of heartwood extracts. The dominant American species (*Q. alba*) is characterized by low levels of oak tannins but high amounts of certain volatile compounds, notably β-methyl-γ-octalactone. American oak wood is also generally denser and has more tyloses. Of the two European species, *Q. petraea* has been found to possess levels of extracts closer to those of American oak, namely high oak lactone and low tannin concentrations, than the other European species, *Q. robur*, even when growing in the same forest (Table 1). Recent studies suggest that the heartwood properties of different American species may also vary.

Not all natural variation is explained by species differences. There is also high heterogeneity in heartwood properties among individual trees of the same species.

A recent study comparing barrels constructed from the wood of individual trees by the same cooper found that the resulting wines were very different. The heartwood concentrations of *cis*-β-methyl-γ-octalactone in the oak wood was the factor that most influenced wine flavor.

Barrel variation Differences in the method of constructing barrels may either augment or reduce natural differences in the properties of oak wood. The most important factor influencing barrel properties is the heating of the wood during construction. The various heating and charring operations during barrel construction modify the chemical composition and macromolecular structure of the wood. These changes depend on the intensity and duration of treatment.

The most susceptible polymers to heating are the oak polyphenols, followed by the hemicelluloses and finally the lignin–cellulose matrix. Thermal degradation leads to the appearance of volatile compounds which are present in only negligible quantities in unheated wood. Thermo-sensitive hemicelluloses and celluloses degrade to hexose and pentose monosaccharides which are in turn susceptible to degradation, forming furan products such as furfural and 5-hydroxymethyl furfural (Figure 2). The thermal degradation of lignins produces aromatic aldehydes (cinnamic and syringyl aldehydes). The intensity of heating influences the composition of degradation products. At higher temperatures, aromatic aldehydes give way to their respective acids and dimethoxy

Table 1 Concentrations of heartwood extractives measured in three oak species by studies of (I) barrel staves and (II) wood cores of individual oak trees. Maximum, minimum, and mean values for the sum of eight identified oak ellagitannins (mg g^{-1}) and the sum of *cis*- and *trans*-β-methyl-γ-octalactone (μg g^{-1})

				β-Methyl-γ-octalactone			Ellagitannins		
Study	*Oak species*	*Forest origin*	*Sample number*	*Mean*	*Min.*	*Max.*	*Mean*	*Min.*	*Max.*
I[a]	*Quercus robur*	Limousin	6	1	1	1.3	27	13	42
	Quercus petraea	Tronais	6	29	1	78	20	6	30
	Quercus alba	Missouri and Virginia, USA	6	39	24	77	14	3	32
II[b]	*Quercus robur*	Cîteaux	22	2	1	8	38	23	52
	Quercus petraea	Cîteaux	28	5	1	28	29	8	40

[a]Results from (i) Masson *et al*. (1995) Ellagitannin content of oak wood as a function of species and of sampling position. *American Journal of Enology and Viticulture* 46: 262–268. (ii) Masson *et al*. (1995) Stereoisomers of β-methyl-γ-octalactone. II Contents in the wood of French and American oaks. *American Journal of Enology and Viticulture* 46: 424–428.

[b]Results from Mosedale *et al*. (1998) Variability of wood extractives among *Quercus robur* and *Quercus petraea* trees from mixed stands and their relation to wood anatomy and leaf morphology. *Canadian Journal of Forest Research* 28: 1–13.

(syringyl) thermolysis products become increasingly prevalent compared to methoxy (guaiacyl) compounds. (*See* **Browning**: Nonenzymatic; **Phenolic Compounds**.)

Structural modification of the wood due to the thermolysis of parietal structures also leads to increased accessibility of solvents to extractible compounds.

Wood-drying conditions also affect the chemical composition of oak wood. Air-drying leads to a greater loss of oak wood ellagitannins than does kiln-drying. The effect of kiln-drying varies according to the temperatures used, with the loss of many volatile compounds (e.g., oak lactone, eugenol) increasing at higher drying temperatures. At drying temperatures of 65 °C or more, a slight degradation in the xylan and glucomannan hemicelluloses is observed.

Barrel Use

Barrels are selected and used according to the properties of the spirit or wine being aged. The correct selection and use of barrels is essential in achieving the desired final product.

Barrel-aging Wines

The costs of barrel-aging restrict their use to only the more prestigious and expensive wines. Traditionally, red wines are aged in barrels directly after vinification, for a period of 6–18 months, before bottling. White wines are sometimes barrel-fermented, the wine being transferred to the barrel before alcoholic fermentation is completed. Both red and white wine may also undergo malolactic fermentation during barrel-aging.

Barrel selection and aging practices depend on the type of wine; the cask should contribute subtle fineness and aromatic complexity without overwhelming the inherent character of the wine. In general, the barrels used for wines are less intensely heated than those for spirits. The type of bung, the frequency of topping-up barrels after evaporation loss, the proportion of new to old barrels, and the cellar conditions will also all affect the aging process. The maintenance of good barrel hygiene is more important in wine than spirit aging due to the greater risk of bacterial contamination. Barrels may be a refuge for microbial contamination and the build-up of tartrate deposits can prevent easy cleaning.

Some wines, and especially sherry, are aged in oak butts using the 'solera' system, which is in effect a process of blending across vintages. After use, these barrels are much sought-after for aging Scotch whisky.

Barrel-aging Spirits

Both the type of cask and the duration of maturation vary widely and are often defined by legislation according to the product name. Thus, American corn whisky must be matured in new, charred barrels, while in Canada and Scotland, maturation must be for a minimum of 3 years. Therefore, the bourbon industry is the main user of new, charred American barrels, which are subsequently reused for aging Scotch whisky and rum. These spirits may remain in barrels for over 10 years. Cognac and armagnac are matured in European oak (**Figure 10**). Coarse-grain (wide annual rings) wood is preferred, as it is considered more porous and releases more tannins than fine-grain oak. Coarse-grain wood is predominantly of the species *Q. robur* and fine-grain is mostly *Q. petraea*. This difference of species is likely to explain the difference in the levels of tannins extracted. The spirits are first lodged in new barrels for several months and then transferred to old barrels. Aging periods may range from 3 years to several decades, with a minimum of $2\frac{1}{2}$ years' aging required for armagnac and cognac under European Union law. While it is in the wood, the spirit acquires color and its bouquet becomes richer and more complex. (*See* **Tannins and Polyphenols**.)

Exhaustion and Rejuvenation of Barrels

The amount of wood constituents extracted and the effect on flavor decline with repeated (or extended) barrel use. A number of methods are used to restore some of the effects lost. The most common techniques include the scraping out of the inner surface of barrels to expose new, unextracted wood, and/or reheating or recharring the inside of barrels. These techniques rarely fully restore the barrel; recharring will not restore products other than those deriving from the heat degradation of wood macromolecules.

The change from using ex-sherry casks to ex-bourbon casks for Scotch whisky maturation led some producers to adopt various cask treatments such as allowing them to absorb white wine or a very sweet, dark sherry under pressure, before using them for whisky. The beverage previously aged in the barrel is considered to influence the flavor of the whisky.

Barrel Alternatives

The use of cheaper alternatives to barrel-aging have become increasingly popular, particularly in the new-world wine and brandy sectors. The adoption of such alternatives is, however, often limited by legislation preventing their use. This is notably the case for the whisky and European wine and brandy sectors.

Figure 10 (see color plate 7) Aging spirits in oak barrels. Reproduced from Barrels *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

The simplest method of 'oak-aging' is by the use of oak chips. Many different types of chips are available according to their size, wood origin, and treatment (degree of toasting). Other barrel-alternatives include various systems of removable staves which can be simply replaced once exhausted. These systems are used in conjunction with either inert tanks or with old oak barrels. Such methods are widely used for wine-aging outside Europe.

Another barrel-alternative is oak wood extract, whose use tends to be limited to brandy production. Wood extracts (*boisés*) allowed in the production of armagnac and cognac must be obtained by extracting oak chips using boiling water and are therefore normally devoid of volatile components, consisting mostly of oak polyphenols and their degradation products. More sophisticated extracts are available as either liquids or lyophilized powders, deriving from an infusion or extraction of oak wood that may have been previously subjected to physical or chemical treatments to promote degradation.

Future Developments

Most attention has recently focused on the variability of oak barrels. There is greater interest in defining the natural properties of stave wood influencing the maturation of wines and spirits. The dominant factor appears to be the biological species, but further studies are needed to estimate the relative importance of natural variation between oak trees, forests, and species.

Coopers also continue to optimize barrel production and quality. In particular, toasting and charring processes are more carefully supervised and controlled.

One of the main unknowns is the future development of barrel-alternatives. Current systems are likely to be refined and perhaps new methods proposed that more closely simulate barrel-aging. However, despite the widespread use of barrel-alternatives for new-world wines, the demand for barrels has showed no sign of significantly decreasing. If oak chips and other systems allow an economical alternative to oak barrels, this has not prevented the continued use of barrels for the most prestigious of oak-aged wines.

See also: **Browning**: Nonenzymatic; **Phenolic Compounds**; **Tannins and Polyphenols**; **Whisky, Whiskey, and Bourbon**: Composition and Analysis of Whisky; **Wines**: Types of Table Wine

Further Reading

Maga J (1989) The contribution of wood to the flavor of alcoholic beverages. *Food Reviews International* 5: 39–99.

Mosedale JR (1995) Effects of oak wood on the maturation of alcoholic beverages with particular reference to whisky. *Forestry* 68: 3, 203–230.

Mosedale JR and Puech J-L (1998) Wood maturation of distilled beverages. *Trends in Food Science and Technology* 9: 3, 95–101.

Mosedale JR, Puech J-L and Feuillat F (1999) The influence on wine of the oak species and natural variation of heartwood components. *American Journal of Enology and Viticulture* 50: 4.

Nishimura K and Matsuyama R (1989) Maturation and maturation chemistry. In: Piggott JR, Sharp R Duncan REB (eds) *The Science and Technology of Whiskies*, pp. 253–263. Harlow: Longman Scientific and Technical.

Philp JM (1989) Cask quality and warehouse conditions. In: Piggott JR, Sharp R and Duncan REB (eds) *The Science and Technology of Whiskies*, pp. 264–294. Harlow: Longman Scientific and Technical.

Puech J-L, Feuillat M, Boulet JC *et al*. (1998) Elevage des vins. In: Flanzy C (ed.) *Œnologie: Fondements Scientifiques et Technologiques*, pp. 1002–1052. Paris: Lavoisier.

Puech J-L, Leaute R, Mosedale JR and Morgues J (1998) Barrique et vieillissement des eaux-de-vie. In: Flanzy C (ed.) *Œnologie: Fondements Scientifiques et Technologiques*, pp. 1109–1142. Paris: Lavoisier.

Puech J-L, Feuillat F and Mosedale JR (1999) The tannins of oak heartwood: structure, properties and their influence on wine flavor. *American Journal of Enology and Viticulture* 50: 4.

Sefton MA, Francis IL and Williams PJ (1989) Volatile flavour components of oak wood. In: Williams PJ, Davidson DM and Lee TH (eds), *Proceedings of the 7th Australian Wine Industry Technical Conference*, pp. 107–111. Adelaide: Winetitles.

Singleton VL (1974) Some aspects of the wooden container as a factor in wine maturation. In Webb AD (ed) *Chemistry of Wine Making*, pp. 254–277. Washington, DC: American Chemistry Society.

Vivas N (1998) *Manuel du Tonnellerie*. Bourdeaux: Editions Fêret-Bordeaux.

BEANS

S K Sathe, Florida State University, Tallahassee, FL, USA
S S Deshpande, Sunnyvale, CA, USA

Background

Throughout human history, more than 3000 species of plants have been used as foods. On a global basis, plants provide 65% of food protein and over 80% of food energy, and account for 85% of gross tonnage. Excluding the large number of fruit and vegetable species, only about 50 crop species make a significant contribution to human diet. Of these, cereals are the largest group, followed by legumes, in terms of global production. However, because legumes contain almost two to three times more protein than cereals, their dietary importance as protein source is well appreciated. Of more than 1300 species of legumes, only about 20 are most commonly consumed by humans. Among these, the common dry bean, *Phaseolus vulgaris*, is consumed in the largest quantity on a world-wide basis. Dry beans are low in fat (excluding oilseeds), low in sodium, and contain no cholesterol. They are a rich source of proteins, complex carbohydrates, fiber, vitamins, and certain minerals. On a caloric basis, dry beans are more nutrient dense than cereals. Dry beans are less expensive than animal food products and, when stored properly, have a considerably longer shelf-life than several animal, fruit, and vegetable products. Since legumes have the ability to fix atmospheric nitrogen and therefore add nitrogen to the crop–soil ecosystem, they are important in soil conservation and maintenance of soil quality.

Global Distribution, Varieties, and Commercial Importance

The word 'legume' is derived from the Latin word '*legumen*,' which means seeds harvested in pods. The term 'pulse' (from the Latin word '*puls*,' meaning pottage) is used for legume seeds that contain small amounts of fat, whereas for those containing large amounts of fat (such as soybeans and peanuts), the term 'leguminous oilseed' is used. According to the Food and Agriculture Organization (FAO), the word 'legume' is used for all leguminous plants. The legumes most commonly used as human food are listed in **Table 1**.

Although legumes have been cultivated for several thousand years, the chronology and origins of domestication of food legumes are almost impossible to reconstruct. Some legumes (lentils for example) have been dated back to 7000–6000 BC. Leguminosae (or Fabaceae) is the third largest family of flowering plants (after Compositae and Orchidaceae) in size and economic importance, and is second only to the grasses (Gramineae). Current estimates indicate that Leguminosae has about 16 000–19 000 species in about 750 genera. The subclassification is somewhat controversial. The generally accepted subclassification is shown in **Figure 1**. Almost all of the domesticated legumes used as food are members of Papilionoideae. All of the common beans belong to the tribe Phaseolaeae.

In terms of global production, legumes (including oilseeds) rank fifth in annual world grain production. Dry beans account for approximately 30% of the total world legume (pulses) production (**Table 2**). On a world-wide basis, the common beans (*Phaseolus* spp.) are the number one crop among dry beans

Table 1 Grain legume species commonly used for food purposes[a]

Botanical name	*Common name*
Aracahis hypogaea	Groundnut, peanut, monkey nut, goober pea, nguba
Cajunus cajan	Pigeon pea, arhar, red gram, tur, *toovar* Angola pea, gandal, ambre vade, alverja
C. indicus	Pigeon pea, congo pea, yellow dhal
Canavalia ensiformis	Jack bean, horsebean, gotani bean, haba de burro, chickasaw, lima
C. gladiata	Sword bean, maxima
Cicer arietinum	Chick-pea, bengal gram, chana, deshi chana, kabuli, chiche
C. minotinum	Chana, garbanzo
Cyamopsis tetragonoloba	Cluster bean, guar, aconite, cyamopse
Dolichos biflorus	Horse gram
D. lablab	Hyacinth bean, bonavist, field bean, caballeros, Indian butter bean Egyptian kidney bean,
Ervum vulgaris	Lentils, masur dhal
Faba vulgaris	Windsor bean
Glycine max	Soybean, soja
G. hispida	
G. soja	
Lablab niger	Lablab bean
L. purpureus	Kidney bean, hyacinth bean, Indian bean, lubia bean
Lathyrus sativus	Grasspea, kesari dhal, vetch, chickling vetch, chicaro
Lens esculenta	Lentils, masur dhal, red dhal, lentille, split pea, lentija
L. culinairs	
Lupinus spp.	Lupins, tarwi, tarin, pearl lupin, wolf bean, tremoco
Macrotyloma uniflorum	Horse gram, Madras gram, Kallu, Kulthi bean
Mucuna pruriens	Velvet bean, cowage, Mauritius bean, stizolobia
Phaseolus aconitifolius	Moth bean
P. acutifolius	Tepary bean, pavi, Yorimuni, dinawa
P. angularis	Adzuki bean, *feijao*
P. aureus	Mung bean, green gram, golden gram, chiroko, chicka sano pea
P. calcaratus	Rice bean, frijol arroz
P. lunatus	Lima bean, sieva bean, Madagascar bean, sugar bean, Burmabean, towe bean, pole bean, caraota, panguita
P. mungo	Mung bean, mungo bean, urd dhal, black gram, *urad*, woolly pyrol, kambulu
P. radiatus	Mung bean, golden graham, green gram
P. vulgaris	Dry bean, haricot, common bean, kidney bean, navy bean, pinto or snap bean, feijao, opoca, *rajma*, French bean, chumbinho
Pisum sativum	Dry pea, green pea, garden pea, field pea
P. angularis	
P. arvense	
Psophocarpus tetragonolobus	Winged bean (humid tropics), goa bean, asparagus bean, Colombo, four-angled bean, princess bean
Sphenostylis stenocarpa	Yam bean
Stizolobium spp.	Velvet bean
Tetragonolobus purpureus	Winged bean (Europe)
Trigonella foenumgraecum	Methi, fenugreek
Tylosema esculentum	Marama bean
Vicia faba	Broad bean, horsebean, faba bean, field bean, Windsor bean
V. sativa	Vetch
Vigna aconitifolia	Moth bean, matki, mouth bean, mat, math
V. aureus	Mung bean
V. radiata	
V. mungo	Black gram, *urd*, *urad*, kambulu, pyrol
V. sinensis	Dry cowpea
V. umbellata	Rice bean, red bean, mambi bean
V. unguiculata	Black-eyed cowpea, black-eye pea, cowpea, kaffir bean, Hindu pea, asparagus pea
Voandzeia subterranea	Bambara groundnut, Madagascar groundnut, earthpea, Congo goober, kaffir pea, jugo bean, haricot pistache

[a]Compiled from Deshpande SS and Srinivasan D (1990) Food legumes: Chemistry and technology. *Advances in Cereal Science and Technology* 10: 147–241 and Doughty J and Walker A (1982) *Legumes in Human Nutrition, Food and Nutrition Paper 20*. Rome: Food and Agriculture Organization of the United Nations.

(excluding oilseeds) in both production and consumption and are therefore economically an important crop. In 1997, the global dry bean supply in kilograms per capita per year was 2.4. Asia produces the largest quantity (48.80% of total world production) of dry beans, followed by Latin America

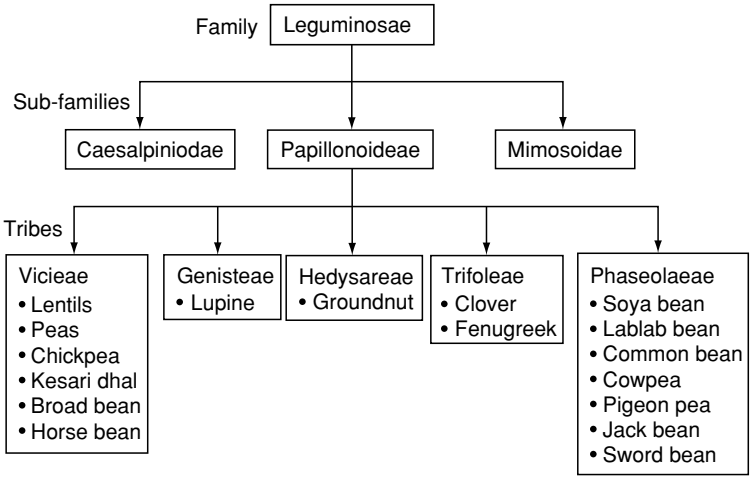

Figure 1 Botanical classification of food legumes.

and the Caribbean (26.98%), North Central America (19.07%), South America (17.19%), and Africa (11.00%). In 1998, India (17.81%), Brazil (13.06%), China (10.15%), US (8.19%), Mexico (7.48%), and Myanmar (6.40%) accounted for 63.09% of the global dry bean production. Latin America and Caribbean, Europe, and Asia, lead the world imports of dry beans in terms of dollar value (Table 2). Brazil, Mexico, Japan, Italy, France, and Venezuela, respectively, accounted for 13.18, 11.17, 8.02, 5.06, 4.55, and 4.08% of the total dollar value of world-wide imports ($1 164 946 000) of dry beans. Asia and North and Central America lead the world-wide export of pulses and dry beans (Table 2). In 1998, the leading exporting countries of dry beans were US, Argentina, China, Myanmar, and Canada, respectively, accounting for 24.37, 18.32, 15.95, 12.27, and 7.36% of the total dollar value (78.27%) of world-wide dry bean exports.

Morphology of the Pods/Seeds

Regardless of the fat content, most legume seeds have a similar structure. Mature legume seeds have three major components: the seed coat, the cotyledons, and the embryo axis. In most dry seeds, they account for 8–20, 80–90, and 1–2% of the seed weight, respectively. The majority of the nutrients are present in cotyledons. Typical seed structure and various anatomical parts of legume seeds are shown in Figure 2. The outermost layer of the seed is the seedcoat or testa. The external seed structure includes the hilum, micropyle, and raphe. The hilum is a scar-like structure (usually oval) near the middle edge where the seed breaks away from the stalk. The micropyle is the small opening in the seedcoat where, originally, the pollen tube enters the valve. The raphe is the ridge at the side of hilum opposite the micropyle and represents the base of the stalk that fuses with the seed coat upon seed maturation. In most legumes, the endosperm is short-lived and shrinks to a thin layer surrounding the cotyledons (or embryo). On soaking the seeds, the endosperm can be easily removed, along with the seed coat. The remainder (embryo) of the seed consists of shoot (which contains two cotyledons), and a short axis above and below the cotyledons and terminates in the shoot tip. The plumule or embryonic stem is well developed in the resting seed and lies between two cotyledons.

The outermost layer of seed coat is the cuticle, which has papillae or papillae-like growth in some legumes (e.g., green gram), but in most legumes, it is a smooth structure. The thickness of the seed coat is quite variable, depending upon the type of bean. Generally, seeds containing thick seed coats tend to have a high fat content. Both hilum and micropyle are important in water imbibition by testa. Palisade cells derived from the outer epidermis of the outer integument, next to the cuticle, are either loosely packed or densely packed depending upon the seed maturity and may affect seed hydration. Next to palisade cells are the hourglass cells. The shape of the hourglass cells can vary from bottle-shaped (guar) to dumbbell-shaped (broad beans) to hourglass-shaped (soybeans). Only a few legume species (such as *Dolichos*, *Cajanus*, and *Vigna*) have more than one layer of hourglass cells. The remainder of the testa contains primarily mesophyl cells.

Legume cotyledons are primarily composed of parenchyma cells, which vary in size (70–100 μm) and act as storage sites for most nutrients. Each cell of the cotyledon is bound by the cell wall and the middle lamella. Vascular bundles in cotyledons generally are devoid of any filling material. Vascular bundles

Table 2 Acreage, production, yield, and import/export data for pulses and dry beans[a]

	Pulses							*Dry beans*						
	Area harvested ($\times 10^6$ ha)	*Yield* (kg ha^{-1})	*Production* ($\times 10^3$ t)	*Import*		*Export*		*Area harvested* ($\times 10^6$ ha)	*Yield* (kg ha^{-1})	*Production* ($\times 10^3$ t)	*Import*		*Export*	
				($\times 10^3$ t)	($\times 10^3$ $)	($\times 10^3$ t)	($\times 10^3$ $)				($\times 10^3$ t)	($\times 10^3$ $)	($\times 10^3$ t)	($\times 10^3$ $)
World	67.651	826	55 933					24.26	694	16 848				
Africa	15.898	519	8 259	130 107	69 299	13 018	4 231	2.97	622	1 852	160 148	83 770	80 022	38 628
Asia	35.258	744	26 244	477 634	201 938	132 532	62 215	13.064	629	8 221	527 742	259 426	1 034 829	416 628
Europe	4.549	2 034	9 256	54 994	14 790	52 096	10 567	0.488	124	607	508 948	355 337	106 562	89 832
Latin America and Caribbean	7.350	699	5 143	104 779	51 198	881	684	6.807	667	4 545	590 809	408 994	370 350	274 625
North and Central America	5.540	1 220	6 764	119 016	57 587	18 533	6 902	3.645	881	3 213	349 826	230 445	677 224	400 858
Oceania	2.056	1 088	2 237	6 141	2 356	8 400	2 291	0.56	100	56	10 086	7 059	32 690	14 877
South America	4.348	729	3 171	1 346	991	517	482	4.031	718	2 896	315 485	228 909	354 176	261 751

[a]Data from FAO (2000).

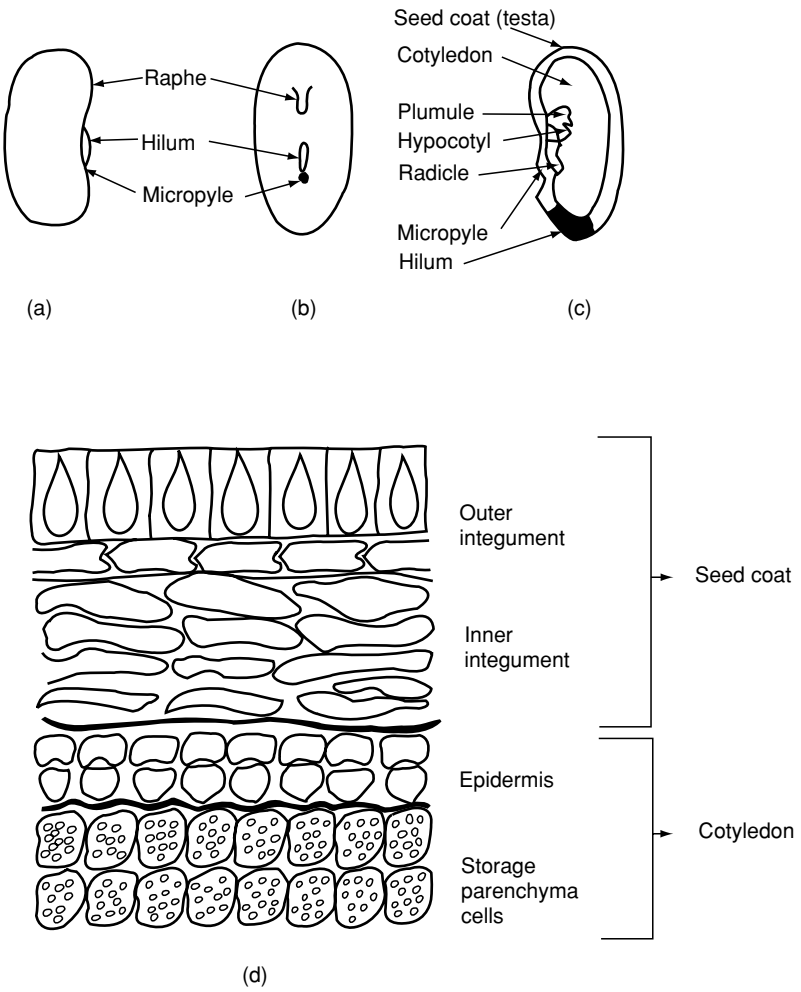

Figure 2 Dry bean (*Phaseolus vulgaris*) seed: (a) external side view, (b) external face or edge view (viewing at hilum side), (c) cross-section (one cotyledon removed), and (d) detailed cross-section across the seed coat and cotyledon.

are often used as key structures in the identification of different plant types.

Chemical and Nutritional Composition

The majority of nutrients in dry beans are primarily located in the cotyledons and account for up to 90% of the total nutritive value. Typically, dry beans provide 300–350 kcal per 100 g of dry seeds. The majority of constituents of cotyledons are proteins and carbohydrates and respectively account for 15–25% and 50–75% of the total seed weight. The remainder consists of fat, minerals, fiber, and vitamins. With the exception of oilseeds, dry beans generally contain low amounts (1–3% of seed weight) of fat. Although most minerals are present in cotyledons, some (such as calcium and iron) may be present in seed coat in significant proportion. The typical nutrient composition of several *Phaseolus* beans is shown in Table 3. Dry beans contain not only significant amounts of nutrients but also several undesirable components and attributes: inhibitors of enzymes such as trypsin, chymotrypsin, subtilisin, amylases, and elastase; lectins; phenolic compounds including tannins; phytates; toxic amino acids mimosine and djenkolic acid; cyanogenic glycosides, which produce HCN; flatulence-causing oligosaccharides raffinose, stachyose, and verbascose; lipoxygenases, which catalyze the development of rancidity, and off-odors often described as beany, grassy, painty, cardboard-like, and chalky.

Proteins

Phaseolus beans are not only important for their caloric contribution to human diet but are especially valued for their protein content, since they are a major protein contributor to the human diet on a global basis. In certain parts of the world, they are

Table 3 Proximate composition of *Phaseolus* beans[a]

Bean	Moisture (%)	Protein (%)	Carbohydrates (%)	Fat (%)	Ash (%)	Crude fiber (%)
Adzuki	11.0	20.2	49.8	1.9	4.39	4.9
Black beauty	10.41	22.87	70.79	4.48	1.86	
Black gram	10.2–10.9	19.7–24.0	56.6–63.4	1.3–1.6	3.2–3.4	4.4–6.4
California small white	9.65	25.90	58.00	0.25		
Cranberry	12.71	23.43	71.26	1.09	4.22	
Great Northern	8.5–13.3	21.0–24.37	61.2–71.07	1.0–3.48	3.5–4.86	6.7
Lima						
Baby	13.30–20.40	20.40	62.10	0.80	3.40	6.0
Large	8.90	22.30	63.80	0.80	4.20	7.4
Mung						
Green	17.92	27.12	62.85	1.53	4.01	4.5
Black	13.64	25.68	64.20	0.45	4.32	5.3
Navy	9.4–18.2	23.13–24.65	61.2–66.19	1.5–4.3	2.90–4.27	3.4–6.6
Pinto	9.05–14.70	18.8–24.97	61.8–69.47	1.2–3.6	3.07–4.10	3.9–6.3
Red kidney						
Light	10.52	20.89	73.20	1.52	4.39	
Dark	13.22	20.32	73.68	1.58	4.42	
Rice bean		18.0–25.0	60.0–77.0	1.0–1.6	3.8–4.3	3.3–4.8
Roshina G2	9.89–11.11	25.77–26.30	63.33–64.02	1.85–2.00	3.19–3.79	4.6–5.1
Roshina pink	4.90	19.40	68.80	3.50	3.40	4.6
Sanilac	11.61	18.98	75.09	1.65	4.28	
Tepary		21.0–25.0	70.0–73.0	0.8–0.9	4.1–4.8	
Small red	13.12	22.45	71.97	1.43	4.15	
Small white	13.03	19.73	74.34	1.99	3.94	
Viva pink	12.69	21.30	73.23	1.06	4.41	

[a]Data compiled from Sathe *et al*. (1984) and Salunkhe and Kadam (1989). Data expressed on a dry-weight basis.

the most important source of dietary protein. Dry bean proteins can be classified as storage and metabolic proteins. The original protein classification proposed by T. B. Osborne was based on the solubility of proteins in a series of solvents. In this scheme, the water and dilute salt soluble proteins were termed as albumins and globulins, respectively. Dry beans contain 40–60% globulins and 20–40% albumins, based on Osborne's protein classification. Globulins are exclusively storage proteins, whereas the albumin fraction contains both storage and metabolic proteins. The protein content of dry beans is usually calculated by multiplying the Kjeldahl nitrogen content by a factor of 6.25. Because dry beans contain 10–15% of total nitrogen as nonprotein nitrogen, most dry bean protein values are typically overestimated by 1–2%.

The major storage proteins in dry beans have been identified by several names, and therefore, their nomenclature is somewhat confusing. Based on the nomenclature using the ultracentrifugation sedimentation coefficient (S), dry beans contain both 7S (vicilin-like) and 11S (legumin-like) storage proteins. Depending on the bean variety, the relative proportion of these two types of proteins varies considerably. The 11S type proteins typically are nonglycosylated proteins with an estimated molecular weight (MW) in the range of 300 000–400 000. They are usually composed of six subunits (MW 60 000), each consisting of one acidic (MW 40 000) and one basic (MW 20 000) polypeptide linked by disulfide bond(s). Usually, 11S proteins are present in minor amounts in *Phaseolus* beans. The 7S globulin in *Phaseolus* beans has also been referred to as glycoprotein II, globulin 1, euphaseolin, globulin, and phaseolin. Depending on the type of bean, the 7S globulin type and quantity vary considerably. In *Phaseolus* beans, the 7S globulins are, however, the major storage proteins and account for 40–60% of the total proteins. The three major types of 7S proteins that have been identified, biochemically purified, and characterized are: (1) phaseolin, (2) lectin (also called glycoprotein II, phytoagglutinins or phytohemagglutinins, and protein II), and (3) arcelin in wild bean accessions from Mexico (named after the town Arcelia in Mexico, where some of the accessions were collected). All the 7S globulins are glycosylated and contain D-mannose and D-glucosamine as the major sugar constituents.

Phaseolin Phaseolin is the major globulin in domesticated *Phaseolus* beans. It is a trimeric, vicilin-like 7S globulin known to exhibit polymorphism. The polypeptide polymorphism is believed to be due not only to the differential glycosylation but also to phaseolin polypeptides being encoded by a small multigene

family. Phaseolin is soluble in 0.5 M NaCl at all pH values. It undergoes reversible pH-dependent dissociation–association with sedimentation coefficients of 3.0S (pH 12.0), 7.1S (pH 7.0), and 18.2S (pH 3.6) known as peptides (MW 44 000), protomers (MW 163 000), and tetramers of protomers (MW 653 000). Phaseolin consists of a group of subunit polypeptides with MWs 43 000–54 000 and isoelectric points from pH 5.6 to 5.8, depending on the phaseolin type. Among *Phaseolus* beans, three distinct types of phaseolins, named after the cultivars tendergreen (T), Sanilac (S), and Contender (C), have been identified. Screening of 107 cultivars has revealed that S-, T-, and C-type phaseolins accounted for 69, 25, and 6%, respectively, of the total cultivars. These types can be easily distinguished by one-dimensional or two-dimensional gel electrophoresis using sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) for one-dimensional or isolectric focusing in the first dimension followed by SDS–PAGE in the second dimension for two-dimensional gel electrophoresis. The C-type phaseolin is believed to have originated from the T and S types. Regardless of the type of phaseolin, phaseolin contains 3–5% carbohydrates, and its amino acid composition is dominated by acidic amino acids (30–40% of the total). Typically, the N-glycosylation sites in phaseolin polypeptides occur at amino acid residue numbers 252 and 341. Sulfur-containing amino acids (notably methionine) are the limiting amino acids in phaseolin as well as other major storage proteins in *Phaseolus* beans. The secondary structure of phaseolin in 0.5 M NaCl typically has a low α-helix content (10%) and β-turns (9.0%) and a large amount of β-sheet (48.0%) and random coils (33.0%). Native phaseolin is quite resistant to digestive proteases such as pepsin, trypsin, and chymotrypsin, and is degraded to polypeptides with MWs 24 000–28 000. Heat-denatured phaseolin is easily digested by these proteases, however. (*See* **Protein**: Chemistry.)

Lectin Bean lectins agglutinate erythrocytes owing to their ability to bind with cell-surface glycoproteins and glycolipids. Although the precise function of lectins in beans is not known, they are thought to offer protection to the plant. Since lectins are toxic, they are of nutritional concern. Lectins occur in both albumin and globulin fractions. Certain bean cultivars lack lectins. When present, lectins represent 6–12% of total protein. In addition to agglutinating activity, many bean lectins have mitogenic activity. Most lectins of *Phaseolus* beans have subunits of MWs 29 000–36 500 with an isoelectric pH in the range of 4.9–7.9 (most are in the 5–6 pH range). The majority of native lectins have tetrameric nature (MWs 100 000–150 000) although some (lima beans, for example) have a dimeric nature. The majority of *Phaseolus* lectins have 4–6% carbohydrate, low sulfur-containing amino acids, and sugar specificity towards D-acetylgalactosamine. Lectins are very resistant to common digestive proteases and are slowly hydrolyzed *in vitro*, even after extensive heat denaturation. Proper moist heat denaturation can completely inactivate the biological activity of lectins and therefore can render them nontoxic. (*See* **Hemagglutinins (Haemagglutinins)**.)

Arcelin Arcelin was first discovered in wild accessions of Mexican *Phaseolus* beans. It also occurs in lines that contain phaseolin as well as lectin. Because it is present in equal or greater levels than phaseolin in certain lines, it is one of the major storage proteins in *Phaseolus* beans. The MWs of arcelin subunit polypeptides range from 35 000 to 42 000, depending on the variant, and are more basic than both lectin and phaseolin. The native arcelin has a MW of 89 000 and is therefore a dimeric protein. Arcelin has many similarities with lectin (including agglutinating activity) with respect to chemical composition.

Other proteins *Phaseolus* beans contain trypsin inhibitors (many of them also inhibit chymotrypsin), amylase inhibitors, lipoxygenases, and several other minor protein components. Most of these proteins are a part of albumins. Trypsin and chymotrypsin inhibitors in *Phaseolus* beans typically account for up to 10% of the total proteins and are generally rich in sulfur amino acids. The MWs of these inhibitors range from 2000–23 000. Most *Phaseolus* beans lack Kunitz-type (inhibitors with 170–200 amino acids with MW of approximately 20 000) trypsin inhibitors. Amylase inhibitors in dry beans have been characterized from only a few cultivars and therefore not yet extensively studied. The MW of kidney bean amylase inhibitor (a glycoprotein) has been shown to be 50 000. Appropriate moist heat treatment (such as cooking or autoclaving) can inactivate both the protease and amylase inhibitors.

Carbohydrates

Total carbohydrates in *Phaseolus* beans contribute 50–70% of the seed weight and include mono-, di-, and oligosaccharides, starch, and other polysaccharides. Starch is the most abundant nutrient in *Phaseolus* beans, accounting for up to 70–80% of total carbohydrates. Among the simple sugars, oligosaccharides (raffinose, stachyose, verbascose, and ajugose) are the major constituents (up to 10% of seed

weight) and are at least partially responsible for flatulence. Crude fiber is primarily composed of cellulose, hemicellulose, lignins (not a carbohydrate), and other nonstarchy polysaccharides such as arabinogalactans, arabinoxylans, glucomannans, galactomannans, and pectins. The hypocholesteremic effect of dry beans is partially attributed to the presence of nonstarchy polysaccharides. *Phaseolus* bean starch granules are quite variable in shape (round, oval, oblong, elliptical, spherical, kidney shape, and irregular) and size (5–60 μm) and typically contain 10–45% (of total starch) amylose. The average degree of polymerization for amyloses and chain lengths for amylopectins of *Phaseolus* bean starches range from 1600 to 1900 and 22 to 26, respectively. Based on X-ray diffraction spectra *Phaseolus* dry bean starches are mostly of the C type [a mixture of A (typical of cereal starches) and B (typical of root and tuber and high amylose cereal starches) types]. These starches have restricted swelling, a gelatinization temperature range of 60–89 °C, high solution viscosities, and good thermal stabilities. Upon gelatinization, they produce opaque gels. Dry bean starches (especially if cooked) are well digested (their digestibility is comparable with those of many cereal and tuber starches) by humans. Because dry bean starches are digested slowly, however, they are hypoglycemic and therefore useful in the diets of diabetics. (*See* **Carbohydrates**: Classification and Properties.)

Vitamins and Minerals

Phaseolus beans are a good source of B-vitamins, especially thiamin, riboflavin, niacin, and folacin. Typically, the thiamin, riboflavin, niacin, and folacin contents (on a dry-weight basis) of *Phaseolus* beans are, respectively, 0.5–1.14, 0.1–0.25, 0.4–3.14, and 0.037–0.676 mg per 100 g. The vitamin E content ranges from 0.72 to 1.97 mg per 100 g, and the B_6 content ranges from 0.2 to 0.659 mg per 100 g. *Phaseolus* beans are poor sources of vitamin A and C.

Phaseolus beans are excellent sources of several minerals, including Ca, Fe, Cu, Zn, P, K, and Mg. Typically, raw beans contain 70–260, 0.5–1.40, 3.34–13.5, 160–320, 1.0–2.1, 380–570, 1320–1780, 4.0–21.0, and 1.9–6.5 (all expressed as milligrams per 100 g, dry-weight basis) of Ca, Cu, Fe, Mg, Mn, P, K, Na, and Zn, respectively. The low sodium and high potassium content of raw beans makes them a desirable constituent of human food (especially for people with hypertension).

Antinutritional Factors

In addition to protease inhibitors and amylase inhibitors, *Phaseolus* beans contain several other antinutritional and antiphysiologic factors such as phytic acid, tannins, cyanogenic glycosides, saponins, and allergens. Because these components are usually present in small quantities (less than 5% of the total seed weight), they do not pose a serious health hazard under normal conditions (that is, when beans are a part of total diet and are properly processed prior to consumption). Of these antinutritional factors, phytates and tannins are of particular concern, because both are heat stable and cannot be easily removed from beans during normal home processing. Phytate is a general term used for mono to dodeca anions of phytate along with esters lower than hexaphosphate. Ca–Mg salts of phytic acid [myoinositol 1, 2, 3, 4, 5, 6-hexakis (dihydrogen phosphate)] are referred to as phytin. In most dry beans, phytate phosphorus accounts for up to 80% of the total phosphorus. The amount of phytate in *Phaseolus* beans ranges from 0.6 to 2.1% (by weight) of the total seed weight. Because phytates are chelating agents, they may interfere in mineral utilization. Germination, fermentation, and soaking followed by cooking (if both soak- and cook-water are discarded) are effective methods of removing phytates (50–80% reduction). (*See* **Plant Antinutritional Factors**: Characteristics.)

Tannins (especially condensed tannins) are heat-stable compounds and are present in *Phaseolus* seeds (especially in colored varieties) up to 2% of total seed weight. Because of their ionic character, they may interact with other constituents (notably proteins) and adversely affect the nutritional bioavailability of that constituent. Tannins are thought to offer protection to the plant from insects and pests. (*See* **Tannins and Polyphenols**.)

Lipids

Phaseolus beans contain 1–3% lipids (by weight), depending on the species. Neutral lipids (30–50% of total) and phospholipids (25–35% of total) are the major constituents and glycolipids may account for up to 10% of the total lipids. Regardless of the variety, *Phaseolus* bean lipids primarily contain palmitic, oleic, linoleic, and linolenic acids. Polyunsaturated fatty acids and saturated fatty acids typically account for 55–87% and 12–28% of total lipids. (*See* **Fats**: Classification.)

Beneficial Bioactive Compounds

Continuing research on raffinose oligosaccharides, phytates, tannins, simple phenolic compounds, sterols, protease inhibitors, and several other major and minor constituents of dry beans suggests that some of these compounds, under appropriate conditions, may offer beneficial effects for human health.

Long-term detailed and fundamental research is needed to fully realize the potential health benefits of dry bean consumption.

Grading, Handling, and Storage

Dry beans are usually harvested at maturity. The seeds are removed from the pods either manually or mechanically and cleaned to remove dirt, stalks, leaves, blemished, and wrinkled seeds, and packaged prior to storage. The grading of seeds is usually based on external characteristics such as color, gloss, seed size, seed soundness, seed firmness, and presence of contaminating substances. The seeds are stored at farmer, trader, or government levels. Typically, farmers hold up to 8% of harvested seeds until the next season so that they can be used as planting seeds if crop failures occur. In developed and developing nations, a majority of the seeds are stored by traders and/or governments to protect against subsequent crop failures (or low yields), price fluctuations, and fluctuations in supply (change in demand, shortages, and famines). Losses in seeds occur both pre- and postharvest. The normal preharvest losses are mainly due to birds and mammals feeding on bean plant seeds. Drought/floods, insects, rodents can also contribute to preharvest losses. In developed nations, preharvest losses are usually small (as low as 1% of the crop). A majority of the losses occur during postharvest handling and storage, and can range from 8% to as high as 50% of total crop. It is estimated that as much as 48% of food produced in the world is lost (due to pre- and postharvest losses). Factors that influence postharvest losses of legumes include moisture, temperature, respiration rate, insect damage, microbial spoilage, and damage caused by mites and rodents. Properly packaged dry beans should be stored at low relative humidity and temperature conditions. A high relative humidity and temperature favor the 'hard-to-cook' beans. These conditions also favor the growth of molds and insects. Three major insect genera that cause much of the damage to stored legumes are *Bruchus, Acanthoscelides*, and *Callosobruchus*. Dehusked, split stored pulses are also damaged by *Rhizopertha, Trogoderma*, and *Tribolium* species. Usually, pests seem to have a preference for the type of bean they infest, although the basis for such a preference (or the lack of it) has not been elucidated yet. The major microbial problem during bean storage is contamination by aflatoxin-producing molds (*Aspergillus flavus* and *Aspergillus parasiticus*). Mites can consume food up to their own weight (6–8 μg) and, because of their large numbers, can cause serious losses. Rodents cause twofold damage to stored legumes by not only consuming but also contaminating plants (up to 20 times the amount they would eat). Because rodents are carriers of many communicable diseases, they pose serious damage to stored beans. The species that most commonly cause damage include *Rattus rattus, R. norvegicus, Bandicota indica, B. bengalensis*, and *Mus musculus*.

Processing and Food Uses

Phaseolus beans are processed and used in a variety of ways. The processing of beans is mostly at the household level in developing and underdeveloped countries, whereas in most developed countries, the majority of the processing is done at the industrial level. Home processing methods include milling, soaking, cooking, frying, germination, fermentation (either alone or in combination with cereals), roasting, puffing, parching, extrusion and frying, and toasting. The method(s) used for home processing depend on the regional preference for bean variety and the desired end product. For example, mung beans in sprouted form are popular on a global scale, and therefore, germination is one of the preferred household processing methods used. Black gram, however, is extensively used for preparation of *idli*, a breakfast food popular in India and Sri Lanka, after fermenting it with rice. Industrial processing includes freezing for such beans as green French beans, snap beans, etc., milling (production of flours and high-protein flours), baking (baked beans), cooking and frying (refried beans), and canning (alone in salt water or tomato juice or in combination with meats such as beef and pork). In developing countries, dehusking and splitting to produce *dhal* are also performed on an industrial scale.

In developing and underdeveloped countries, *Phaseolus* beans are used in numerous ways, depending on the type of bean and regional preference. They may be eaten as raw, immature seeds; cooked as green vegetables (such as French beans); consumed as part of salads; used in making curry; used as a soup ingredient; cooked, mashed, mixed with condiments and spices and used as gruels and porridges; prepared as pastes to be extruded to prepare fried snack products; sprouted; puffed or roasted and eaten as snack foods; and fermented to prepare numerous fermented products.

In developed countries, *Phaseolus* beans are consumed as a salad and soup ingredient, sprouts, canned, frozen, and refried beans. They are also extensively used in the preparation of Mexican-style preparations, such as burrito, chimichanga, taco, bean dips, tamale, etc., and often canned with meats such as beef and pork. In many South American

countries, cooked black beans are a preferred part of breakfast.

Although there is a good potential for the preparation of protein concentrates and isolates and the development of food starches, *Phaseolus* beans have not been used on a large scale for such purposes. In many countries, especially the developed countries, *Phaseolus* beans have been extensively used as animal feed. In developing and underdeveloped countries, the green foliage, deseeded pods, and roots and shoots of bean plants are used as natural fertilizers, especially after composting. Because the legume roots fix nitrogen, they help conserve soil quality. For this reason, in many developing countries, they are extensively used for soil-quality conservation.

See also: **Carbohydrates**: Classification and Properties; **Legumes**: Legumes in the Diet; Dietary Importance; **Fats**: Classification; **Phytic Acid**: Properties and Determination; Nutritional Impact; **Tannins and Polyphenols**; **Vitamins**: Overview

Further Reading

Derbyshire E, Wright DJ and Boulter D (1976) Legumin and vicilin, storage proteins of legume seeds. *Phytochemistry* 15: 3–24.

Deshpande SS (1992) Food legumes in human nutrition. A personal perspective. CRC *Critical Reviews in Food Science and Nutrition* 32: 333–363.

Deshpande SS and Srinivasan D (1990) Food legumes: Chemistry and technology. *Advances in Cereal Science and Technology* 10: 147–241.

Doughty J and Walker A (1982) *Legumes in Human Nutrition, Food and Nutrition Paper 20*. Rome: Food and Agriculture Organization of the United Nations.

FAO (2000) Statistical Databases, Food and Agriculture Organization of the United Nations, Rome, Italy http://apps1.fao.org/serlet.

Kami JA, Velasquez VB, Debouck DG and Gepts P (1995) Identification of presumed ancestral DNA sequences of phaseolin in *Phaseolus vulgaris*. *Proceedings of the National Academy of Sciences of the USA* 92: 1101–1104.

Muntz K (1998) Deposition of storage proteins. *Plant Molecular Biology* 38: 77–99.

Osborn TC (1988) Genetic control of bean seed proteins. *CRC Critical Reviews in Plant Sciences* 7: 93–116.

Patwardhan VN (1962) Pulses and beans in human nutrition. *American Journal of Clinical Nutrition* 11: 12–30.

Reddy NR, Pierson MD and Salunkhe DK (1986) *Legume-based Fermented Foods*. Boca Raton, FL: CRC Press.

Reddy NR, Pierson MD, Sathe SK and Salunkhe DK (1984) Chemical, nutritional and physiological aspects of dry bean carbohydrates – a review. *Food Chemistry* 13: 25–68.

Reddy NR, Pierson MD, Sathe SK and Salunkhe DK (1989) *Phytates in Cereals and Legumes*. Boca Raton: CRC Press.

Salunkhe DK and Kadam SS (1989) *Handbook of World Food Food Legumes: Nutritional Chemistry, Processing Technology, and Utilization*, vols. I–III. Boca Raton, FL: CRC Press.

Sathe SK, Deshpande SS and Salunkhe DK (1984) Dry beans of *Phaseolus* A review. Part 1. Chemical composition Proteins. *CRC Critical Reviews in Food Science and Nutrition* 20: 1–46.

Takahashi CK, Liang YW and Pueyo MT (2000) Phaseolin RFLP in bean (*Phaseolus vulgaris*) cultivars. *Food Biochemistry* 24: 129–145.

BEEF

R L Joseph, The National Food Centre, Dublin, Ireland

Introduction

Beef is 'the flesh of ox, bull or cow' (*Oxford English Dictionary*). *Boeuf* (Anglo-French) is derived from the Latin *bos, bovis* (ox), and until the eighteenth century beef meant both animal and meat. The etymology suggests that the meat was for the Norman ruling class, and was not commonly eaten by Anglo-Saxon folk even in the late thirteenth century as Middle English was forming.

Beef has enjoyed a special status since *Bos* spp. were domesticated in Neolithic times in South-west Asia. The bull was a totem for warrior nomads; eating his flesh transferred his power. The docility and strength of draft oxen, the versatile milk, the valuable hides and attractive meat have made the *Bos* genus economically vital to humans.

Today beef cattle are raised all over the world (**Table 1**). The natural environment of *Bos taurus* is temperate grasslands. *B. indicus* prefers savannah.

Table 1 World production of beef and veal[a]

Region	Live ($\times 10^3$)	Slaughtered ($\times 10^3$)	Carcass weight (kg, mean)	Production (kt)
1979–81				
World	1 218 277	223 143	198	44 090
Africa	172 203	20 683	142	2932
North/Central America	173 510	48 911	254	12 402
USA	112 152	37 292	271	10 092
South America	239 246	34 023	201	6853
Asia	350 404	21 382	114	2435
China	52 567	2350	75	177
India	186 500	1710	80	137
Europe	133 377	48 037	220	10 551
France	23 825	7841	234	1832
Germany[b]	20 672	7353	259	1890
UK	13 321	4097	260	1063
Oceania	34 790	12 715	173	2198
USSR	114 748	37 391	180	6720
1989				
World	1 281 472	236 577	209	49 436
Africa	185 794	23 277	148	3455
North/Central America	166 999	52 631	267	14 028
USA	99 180	36 376	293	10 655
South America	261 096	34 006	209	7100
Asia	391 556	29 599	119	3532
China	661 141	6658F	99	662F
India	195 500	2925	80	234F
Europe	125 569	43 388	242	10 500
France	21 780	6900F	249	1716F
Germany[b]	20 369	7814	258	2005
UK	11 902	3374	284	958
Oceania	30 858	10 375	195	2021
Former USSR	119 600	43 300F	203	8800F

'F' is Food and Agriculture Organization estimate.
All data from FAO (1989) *FAO Yearbook of Production*, vol. 43. Rome: Food and Agriculture Organization.
[a]Average per annum.
[b]Summed over FRG and GDR figures.

Crossbreeds such as Santa Gertrudis grow well in hot dry areas, such as the south-west USA.

Different breeds have been developed over the centuries to maximize returns under local conditions. The compact, early-fattening Aberdeen Angus and Hereford are productive in harsher nutritional and climatic environments. The great draft cattle of Europe, like Charolais and Chianina, put on muscle quickly in larger quantity and are heavier at slaughter. They are leaner, more productive, and are gradually replacing the earlier-maturing breeds.

Many European countries, and New Zealand, produce beef from the dairy herd, whose cows are selected mainly for milk production. Beef bulls on surplus heifers produce beef calves. Pure dairy calves may be raised for beef, in the UK, Ireland, and New Zealand, or veal, in France and The Netherlands, or pet food, in the USA.

Modern economic pressures intensify and concentrate animal production. Many once strong local rare breeds are facing extinction; in the British Isles and USA alone about 30 are listed. Their rich gene pool must be maintained so a wide genetic base will be available to accommodate future natural or market constraints on production.

Comparison of systems converting feed into saleable meat is complex. Prices of land, stock, and feed change yearly. Support schemes alter markets. Health and diet are important to consumers, so a demand for 'natural' feeding has reinforced economic pressures towards increased use of grass and less use of animal-derived feeds.

Typical Production Systems

Two-Year-Old Spring-Born Steer (Friesian/Friesian Cross)

This tightly managed semiintensive system is suitable for north-west Europe. In February, male calves are bought and finished cattle are sold. In early April yearlings start grazing 45% of the farm; 55% is closed for silage. By mid-May the calves are grazing ahead of the yearlings. In late May silage is cut, 45% of the farm is closed for a second cut, taken in mid-July, then slurry is spread. The calves are castrated in late September. Finishing cattle are housed in mid-October, weanlings in early November.

Fifty cattle may be produced off 22 ha. The main inputs per animal sold per year are the 0.45 ha of productive grassland, to provide grazing and 10 t of silage; the fertilizers are 110 kg of nitrogen equivalent, 41 kg of potassium, and 11 kg of phosphorus. Each animal receives 25 kg of reconstituted milk powder, roughage, and concentrates *ad lib*. In the first winter the weanlings are fed silage and concentrates (150 kg). Yearlings receive 600 kg of concentrates during a 150-day finishing period.

Live-weight gains are from 0.6 to 0.9 kg day^{-1}. Slaughter weight (Friesian) is about 600 kg, to give a carcass of 320 kg, with about 220 kg of saleable lean beef, some 55 kg of bone and 45 kg of fat trim. Charolais $\times$ Friesian reach 680 kg, dressed 380 kg; giving 270 kg of lean meat, 60 kg of bone, and 50 kg of fat trim.

Suckler Beef

This is extensive and more common on marginal land. Calves suckle dams; a 'cow unit' consists of cow, calf, and yearling. Cows are cross-bred Hereford or Limousin on Friesian to give adequate milk with

low maintenance requirements. Mating is to a third breed: Hereford, Aberdeen Angus, or Limousin for the first calving and large breeds, Charolais or Simmental for later pregnancies.

Calving takes place in February to April. Calves grow rapidly on the milk, reaching about 300 kg at weaning in November. Each cow unit grazes 0.32 ha in April–June, with 0.45 ha reserved for silage. Then, until August, 0.48 ha is grazed; 0.29 ha is for silage. Thereafter, 0.77 ha is grazed until housing in November. Silage is cut, in late May and late July, totaling 37 t ha^{-1}. About 175–225 kg ha^{-1} of nitrogenous fertilizer is applied yearly.

Housed yearlings have silage *ad lib* with 1 kg per head per day of concentrates. After the second summer on grass, the heifers are slaughtered, at 20 months, in November–December, finishing on silage and 3 kg day^{-1} of concentrates. Slaughter at 500–600 kg gives carcasses of 260–330 kg. Steers are fed silage and 4 kg day^{-1} of concentrates to finish in early spring at 24 months, at a live weight of 620–680 kg and a carcass weight of 340–380 kg.

Bull Beef

In Europe, except the British Isles, most male beef cattle are bulls. From about 4 months they grow up to 8% faster than steers, giving perhaps 40 kg more carcass. Bulls kill out at about 57% (dressed weight as percentage of carcass), 2% more than comparable steers, with carcasses up to 15% heavier. About 74% of a bull's carcass is lean meat, compared to 66% for steers. Bulls have a heavier forequarter, but a higher yield of high-price cuts.

Most bull beef is from the dairy herd, being Holstein/Friesian dual-purpose and dairy beef crosses. Calves bought at 1–3 weeks of age are reared to 100–150 kg on 25–50 kg of milk replacer, concentrates, and roughage. Slaughter at 12 months requires a high-cereal or -fodder beet diet, giving a daily gain of 1.1–1.4 kg and a carcass weight of 180–220 kg. The meat is very pale. Slaughter at 18–21 months requires maize silage with 15–20% concentrates or beet pulp. Friesians reach 550 kg; dual-purpose and crosses may attain 600 kg.

In France, Spain, and Italy the suckler herd is an important source of bull beef. Veau de Lyons is Limousin, and at 12 months the slaughter weight is 450 kg; the meat is pale and not marbled. Slaughter at 18–21 months requires a diet like the dairy bull, but with more concentrates (25–30%). Carcass weights are 300–430 kg, depending on the region and breed.

Nearly all bulls are housed, usually loose in pens of 2–5 m^2 per bull with up to 25 per pen. Farmers raising bulls must insure safety, with secure housing and fencing, warning notices, and careful work practice.

Bovine Spongiform Encephalopathy

Bovine spongiform encephalopathy (BSE) is a progressive, fatal neurological disease. Since 1986 some 26 000 cases have been recorded in the UK, with sporadic cases in other countries. BSE presents, in adult dairy cows, as abnormalities of behavior and gait. The causal agent was in meat and bone meal in concentrates fed to calves. In 1988, feeding ruminant protein to ruminants was banned, and notification and compulsory slaughter were introduced. (*See* **Bovine Spongiform Encephalopathy (BSE)**.)

Meat Trade – Transport and Slaughter

The trade is closely controlled both nationally and internationally, e.g., by the EC. Feeds and therapeutics must not endanger consumers' health. Cattle handling must be humane, so animals must be led, not driven, and moved in production groups with minimal goading and noise. Journeys should be short and packing must allow fallen cattle to rise. A 12-m lorry will take 22 650-kg cattle.

Humane treatment improves meat quality and reduces cross-contamination; nervous overfrequent defecation increases *Salmonella* and other infections. Lairage at the meat plant must have drinking water and shelter. Groups should not be mixed as this causes fighting. Stock must not be fed for 24 h before slaughter.

Western slaughter is preceded by stunning with a blow or captive bolt. The unconscious animal is hoisted by a hind leg, over a 'bleeding trough' and stuck in the arteries of the throat. Delayed sticking allows blood pressure to rise; arteries in muscles burst so meat is blood-splashed and unsightly at sale. (*See* **Meat**: Slaughter.)

Jewish or Muslim ritual slaughter, kosher or halal, despatches the conscious animal with one cut across the throat; slaughtermen must be licensed by the religious authorities. Bleeding-out takes about 10 min. Blood supports microbial growth and its removal is important for increasing meat shelf-life.

The carcass, suspended by both hind legs, is dressed on the line. A moving chain carries it past successive work stations. Hooves and head are removed. The esophagus and rectum are tied and freed before the ventral abdominal wall is opened. The green offal – the alimentary tract – and the red offal – the heart, lungs, and liver – are removed on to conveyors. All parts are inspected by veterinarians. The carcass is flayed and split with a power saw into two sides. Trimming is to specifications agreed with the farmers' associations.

Sides are weighed 'hot' and graded for fatness and conformation (shape) by public inspectors. Payment

is generally on 'hot weight minus 2%' to accommodate evaporative weight loss in the chills. Sometimes grade will affect price.

Sides are washed and pass into the chill room about 1 h postmortem. Chilling is mandatory. Chilling rates are determined by carcass weight and fatness and by air speed and temperature. Faster cooling facilities throughput and reduces bacterial growth and evaporative weight loss, but is more expensive in equipment and energy. Cold toughening may result if muscle cools below 10 °C sooner than 10 h postmortem.

Cutting

EC regulations stipulate that, before cutting, sides must be chilled to below 7 °C. A 140-kg side of average fatness will achieve 7 °C in the center hip, in about 32 h under air of 0 °C at 1 m s^{-1}, but this rate could cause cold toughening in the striploin (longissimus dorsi, LD), so a slower chilling may be required.

Chilled beef sides are quartered according to local usage, often near the end of the rib cage, and go for shipping, to the boning hall or retail shops. The primal cuts of the side differ slightly by country, but are based on convenient anatomical divisions, eating quality and costing.

Figure 1 shows London and Home Counties cuts. Vacuum-packed, usually boneless, primals now predominate over quarters for distribution.

Beef differs in eating quality depending on the cut, the age of the animal, and pre- and postmortem handling. The relative price of sides and cuts depends chiefly on expected tenderness. Prime 2-year-old steer is more expensive than 7-year-old culled cow. Fillet (psoas major) is always tender, and is the most expensive; striploin and rump (gluteus medius) are slightly tougher, more unpredictable, and are cheaper in cost. Shin beef is cheapest of all. *B. indicus* meat is tougher than that from *B. taurus*, but breed differences are not large; reports of tougher bull meat are now attributed to faster cooling in the leaner sides.

Butchers, and customers, may identify different cuts, but once the beef is off the bone and trimmed of fat it is difficult to distinguish lean beef from, for example, adults of different ages or from fast- and slow-cooled sides. This has made beef marketing less transparent than that of other meats.

The trade has developed empirical techniques for improving quality. Prime beef is from 18–36 -months-old cattle. Aging, holding in chill for some days, effects some improvement in tenderness, but will not make cold-toughened beef as tender as slow-cooled beef.

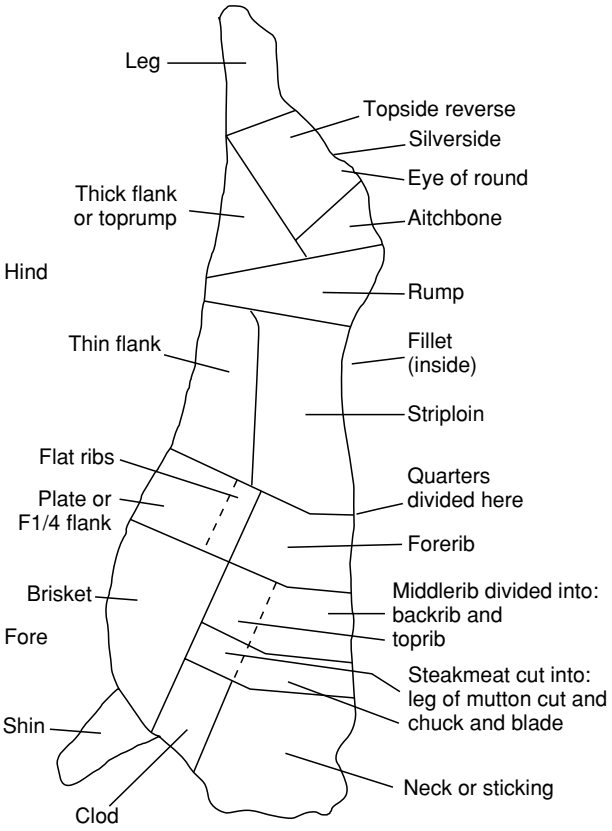

Figure 1 London and Home Counties primal cuts. Adapted from Gerrard F, Mallion FJ and Quin (eds) (1977) *The Complete Book of Meat*. London: Virtue.

Muscle biochemistry now provides explanations for much meat quality variation. In cattle, prolonged stress premortem depletes muscle glycogen reserves and postmortem production of lactic acid from glycogen is inadequate to bring the physiological pH of muscle from 7 to 5.4–5.6, the normal level 48 h after slaughter. Stressed cattle give meat of above pH 6.0 that is described as 'dark, firm, and dry' (DFD). Consumers react against the darker appearance compared to the customary bright cherry-red due to oxymyoglobin at pH 5.4–5.6; the higher pH facilitates microbial spoilage. (*See* **Meat**: Structure.)

Young bulls, and most cattle in the autumn, are more prone to stress and thus to produce DFD meat. Biopsies have shown that depleted muscle glycogen is not replaced in life for several days after the stress insult. Stressed cattle must be rested before slaughter. Humane handling will prevent most DFD meat.

Tenderness variation among muscles is partly due to differences in connective tissue collagen content. More collagen implies lower tenderness (Table 2(a)). Decreasing tenderness with age is due to increasing collagen molecular cross-linking. Solubility is an inverse measure of cross-linking and is less in older

Table 2 Collagen and tenderness

(a) Differences between muscles

Muscle	*Cut*	*Collagen (%)*	*Tenderness*
Psoas major	Fillet	2.24	6.7
Longissimus dorsi	Striploin	2.76	5.2
Gluteus medius	Rump	3.64	4.5
Semimembranosus	Topside	4.09	3.3
Semitendinosus	Eye of round	4.75	3.6
Biceps femoris	Silverside	5.07	2.8

Collagen is expressed as percent fat-free dry matter (Dransfield E (1977) Intramuscular composition and texture of beef muscles. *Journal of the Science of Food & Agriculture* 28: 833–842). Tenderness is on a linear scale: 0, extremely tough; 8, extremely tender (Joseph RL and Connolly J (1977) The effects of suspension method, chilling rates and post mortem ageing period on beef quality. *Journal of Food Technology* 12: 231–247).

(b) Collagen cross-linking and age

Extraction temperature of medium (°C)	*Animal age groups*			
	I (40–49 days)	*II (403–495 days)*	*III (5 years)*	*IV (10 years)*
60	544 ± 29	168 ± 10	82 ± 1	38 ± 1
70	863 ± 34	409 ± 26	171 ± 4	82 ± 1

Figures are average ± SE μg collagen extracted from samples into medium (Goll DE, Hoekstra WG and Bray W (1964) Age associated changes in bovine muscle connective tissue. II. Exposure to increasing temperature. *Journal of Food Science* 29: 615–628).

cattle. Table 2(b) shows that extraction, that is, solubility, of collagen falls significantly with increasing animal age.

Cold toughening seems to invoke two mechanisms. Early exposure to a temperature of about 10 °C causes the release of Ca^{2+} ions, which stimulate contraction in slack muscles like LD. As the muscle contracts into rigor, extra bonds are formed between the thick and thin filaments of the contractile myofibrils, and the strength and toughness of the meat are increased. This is cold shortening, and is confirmed by micrometry of the sarcomeres, units of the myofibrils.

Although ultrafast cooling may enhance tenderness, normal cold inhibition of protein-splitting enzymes during the very early period, 2–4 h postmortem, probably causes more tough beef in commerce than cold shortening. A near-physiological temperature should be maintained up to 4 h postmortem to maximize tenderness. Lower temperatures reduce tenderness *pro rata*. Subsequent aging at low temperature, and low pH, may not facilitate the splitting of particular protein links in the myofibrils which would be critical in providing reentrant cracks for the breaking of connective tissue during chewing. (*See* **Meat**: Preservation.)

Fillet is always consistently tender because it is low in collagen, it is stretched giving minimal thick–thin filament overlap and, being inside the rib cage under kidney fat, cools slowly.

A coherent explanation of beef tenderness is awaited; the synergy of temperature, pH, and Ca^{2+} concentration in controlling enzyme activity and shortening is complex, as is the interaction of connective tissue, myofibrils, and cooking.

Electrical stimulation (ES) of prerigor carcasses was patented 40 years ago to improve tenderness. ES causes muscular contractions, depleting the glycogen and other energy reserves in muscles, so preventing the cold shortening of slack muscles like LD. It may toughen muscles if the concomitant fast pH fall inactivates certain proteolytic enzymes. Its effect may be compounded by the tearing of opposing simultaneously contracting muscles. The New Zealand lamb industry and the North American beef industry use ES, but it is not widely used elsewhere.

Nutritional Value

Beef is central in most societies for celebrations and family meals, but has lost a little ground to white meats recently. Beef is an excellent food, supplying most nutritional requirements save fiber (Table 3); its vitamin B_{12} content promotes vitamin absorption from other sources. The three essential unsaturated fatty acids account for 3–6% of total fatty acid in lean beef. Trace elements (iodine, managanese, zinc, cobalt, selenium, nickel, chromium, molybdenum, fluorine, vanadium, silicon, arsenic) occur in beef so a normal diet will ensure an adequate supply. (*See*

Table 3 Contribution of 100 g of beef to recommended daily allowances (RDAs)

Nutrient	*US RDA*	*From beef*	*%RDA*
Energy	11.3 MJ	2.4 MJ	21.0
Protein	56 g	31.1 g	55.5
Calcium	800 mg	22.6 mg	2.8
Phosphorus	800 mg	312.8 mg	39.1
Iron	10 mg	4.7 mg	47.0
Magnesium	350 mg	31.7 mg	9.0
Vitamin A[a]	1000 μgRE	455 μg RE	45.5
Vitamin C	60 mg	1.16 mg	1.9
Thiamin	1.4 mg	0.50 mg	35.7
Riboflavin	1.6 mg	0.43 mg	26.9
Niacin	18 mg	7.95 mg	44.2
Vitamin B_6	2.2 mg	0.59 mg	26.8
Vitamin B_{12}	3.0 μg	5.23 μg	174.3
Folacin	400 μg	17.5 μg	4.4
Zinc	15 mg	7.1 mg	47.3
Copper	2.0–3.0 mg	0.21 mg	10.5

[a]1 retinol equivalent (RE) = 3.3 international units (IU) of vitamin A activity.

Meat: Nutritional Value.) Refer to individual nutrients.

The culinary uses of beef are numberless and it does not pall. Stock from beef and bones is the 'fond de cuisine.' The beefburger has been universally accepted by the world's youth. Charqui (biltong), sun-dried salted beef, provides protein and salt throughout the tropics. Salamis were developed to preserve beef, and other meats with salt and spices; now thousands of varieties are available. The 'roast beef of old England' still forms the centerpiece of state banquets. Beef is indeed the universal meat.

See also: **Bovine Spongiform Encephalopathy (BSE)**; **Meat**: Structure; Slaughter; Preservation; Nutritional Value

Further Reading

de Boer H and Martin J (eds) (1978) *Patterns of Growth and Development in Cattle*. The Hague: Martinus Nijhoff.

Dransfield E (1977) Intramuscular composition and texture of beef muscles. *Journal of the Science of Food and Agriculture* 28: 833–842.

FAO (1989) *FAO Yearbook of Production*, vol. 43. Rome: Food and Agriculture Organization.

Gerrard F, Mallion FJ and Quin (eds) (1977) *The Complete Book of Meat*. London: Virtue.

Goll DE, Hoekstra WG and Bray W (1964) Age associated changes in bovine muscle connective tissue. II. Exposure to increasing temperature. *Journal of Food Science* 29: 615–628.

James SJ and Bailey C (1990) Chilling of beef carcasses. In: Gormley TR and Zeuthen P (eds) *Chilled Foods – the Ongoing Debate*. Amsterdam: Elsevier.

Joseph RL and Connolly J (1977) The effects of suspension method, chilling rates and post mortem ageing period on beef quality. *Journal of Food Technology* 12: 231–247.

Lawrie RA (1985) *Meat Science*, 4th edn. Oxford: Pergamon.

Pearson AM (1981) Meat and health. In: Lawrie RA (ed.) *Developments in Meat Science*, vol. 2, ch. 8. London: Applied Science Publishers.

Tarrant PV (1989) Animal behaviour and environment in the dark-cutting condition in beef – a review. *Irish Journal of Food Science and Technology* 13: 1–21.

Wallis D (1986) *The Rare Breeds Handbook*. Poole: Blandford.

Beekeeping *See* **Honey**

BEERS

Contents

History and Types

C W Bamforth, University of California, Davis, CA, USA

The Origins and History of Beer

We must look back to Babylon some 8000 years ago for the origins of beer. It is assumed that barley, which was beginning to be cultivated, was found to be both more nutritious and tasteful once it had been wetted and subsequently dried after sprouting, Furthermore the ensuing product, perhaps baked in the form of a bread, once more wetted and spontaneously fermented by adventitious yeasts, will have been particularly pleasurable. It was soon realized that the dough could be thinned by the plentiful addition of water and then strained to remove most of the solids prior to fermentation of the liquid component in earthen vats. Furthermore, the ancient brewers will have realized the merits of adding a proportion of the previous brew to 'kick-start' the fermentation. Addition of certain other plant materials, including mandrake, afforded flavors to complement the elements of the brew.

Beer soon became the staple drink of the Egyptian diet, as well as playing a key role in matters from currency to religious offering. It is claimed that the court of the Pharaoh Rameses included a brewery producing 10 000 hectoliters of beer each year.

Although brewing technology passed to the Greeks and Romans, beer was only consumed by their lower classes. In the warmer southern countries wine was the preferred beverage. The procession of beer-brewing technology took a more northerly route, with the Gauls, Germanic tribes, and Celts. In part this reflected their abilities to work with wood for the fashioning of vessels for brewing and storage. By the first century AD there is mention of Britons and Hiberni (Irish) making an alcoholic drink from barley.

Ale was a major beverage for the Danes and Anglo-Saxons and was regarded as the ideal drink for heroes. The Vikings drank heartily before setting sail next day, hence the origins of the phrase 'three sheets to the wind.' The word beer originates from the Anglo-Saxon *bjor*.

Right the way through and beyond the middle ages, much beer was brewed in the home, by the female of the household (the 'brewster'). However whenever history has seen the development of larger communities, it has been accompanied by the establishment of more substantial, commercial brewing operations. The development of new technologies also allows the scaling-up of brewing concerns, and so it was in the early AD centuries, with individuals learning to smelt and shape metal into vessels. Cleanable, flame-proof, heat-conducting copper brewhouse equipment emerged, and metal could also be used to hold together ever larger wooden fermenters, as well as barrels.

Through the ages some of the most substantial breweries have been located in ecclesiastical settings. Remnants today are the formidable Trappist beers of Belgium. The Domesday Book (1086) records that the brewery of St Paul's Cathedral was producing the equivalent of some 3150 hectoliters of ale. In every walk of life, ale was the principal beverage for all members of the family, young and old.

The roots of a truly major commercial brewing industry in the west were sown first in Bohemia and Bavaria. The importance of good malt to good beer was recognized, causing the development of specialist malthouses. Many of the European beers were flavored and preserved with proprietary mixtures of herbs and spices, including bog myrtle, coriander, cumin, and even strychnine. Such mixtures were known as gruit. Hops were first used in beer in Bavaria in the late eighth century, yet did not gain the ascendancy as the flavorant for several centuries in many countries, notably England. The term 'ale' was originally retained in that country for description of

products that did not contain hops, as opposed to beers, which did.

Until the sixteenth century ale (in later years incorporating hops) was the major beer style universally, such beers being produced in open vessels by yeasts that rose to the surface during fermentation (top fermentation). Bottom fermentation probably started in Bavarian monasteries in the early sixteenth century. A major factor was an edict of Prince Maximilian I in 1533 that largely precluded brewing in the summer months. The ale-style products from top fermentation had been brewed in warm conditions, but now bottom fermentation was performed in the winter, with the beer being stored ('lager' = 'store') until the subsequent autumn, when brewing could start again.

In Bavaria in 1516, Dukes Wilhelm IV and Ludwig X introduced the Reinheitsgebot to preclude the use of undesirable materials in brewing, a practice that had crept in due to shortages of barley and hops. The law has survived to this day and now extends throughout Germany for domestic production, restricting the raw materials of brewing to malt, hops, yeast, and water.

James I of England, whose exchequer took four pence for every quarter of malt, levied the first tax specifically on brewers in 1614.

Just 6 years later the first ales were landed in the New World with the Pilgrim Fathers, who actually put into Plymouth Rock because: 'We could no longer take time for further search or consideration, our victuals being much spent, especially our beer.'

In the late eighteenth century, with the Industrial Revolution, there was a decline in the proportion of beer brewed at home and a growth in brewing companies producing beer for consumption for a working population. The development of transportation systems enabled the increasingly growing brewing companies to distribute their products ever more widely. The railways and roads were fast becoming main arteries for beer distribution, but beer was also being shipped internationally, with products such as the India Pale Ale from England having a suitable long shelf-life, its high hopping acting as a preservative. Use of James Watt's steam engines (1765) within the brewery facilitated increased efficiencies and outputs. The development of artificial refrigeration by von Linde (1876) divorced the production of lager-style products from an association with natural ice such as was to be found in the winter months adjacent to the great lakes in North America, as well as allowing the shipping of beer ever greater distances in cold railcars.

The ability to regulate the process was, by now, manifest. In 1762 Michael Combrune had employed a thermometer to enable the checking and regulation of temperatures in the mash tun and fermenter. Prior practice was to dip the thumb into the brew to ensure that the wort wasn't too hot for the yeast – the 'rule of thumb.' Another Englishman, John Richardson, further facilitated enhanced process control with the construction in 1784 of the first brewer's saccharometer. Other significant technological developments included Pasteur's understanding (1860) of the natural spoilage of foodstuffs by microorganisms and the role of heat in countering this, as well as the development of bottle-closing technology, the crown cork. This was invented in Baltimore by William Painter, a man of British descent who accordingly took out his first patent on closures in London in 1885 and who, by 1891, had arrived at the familiar crimped metal device. Accordingly, beer for domestic consumption, as opposed to draught beer for retail in pubs and bars, could now reach a vast hinterland. The appearance of beer had been revolutionized by Enzinger's (1879) introduction of filtration while the consistency of beer was dramatically enhanced by Hansen's realization in 1883 of the importance of using isolated, pure yeast strains for fermentation of individual beers. The advent of cylindroconical vessels through the patents of Nathan filed between 1899 and 1905 allowed the large-scale fermentation of beer. The first cans suitable for filling with beer were produced by the American Can Company in 1934, and the initial company to go ahead with a canned product was Krueger's of Newark with its cream ale, in January 1935.

Other key developments in the evolution of a refined brewing process were the establishment of breeding programs for the key raw materials barley and hops. The first named barley variety came in 1823 in the UK through the skill of Chevalier, and it was duly named for him. The first established hop-breeding program was also in the UK, with Howard at Wye, commencing in 1904. By the 1950s for barley and 1970s for hops the most significant advances in how to process them had been put in place. Thus the importance of steeping barley intermittently with air rests between water soaks had led directly to a halving of total malting time and huge enhancement of output and efficiency. For hops, the establishment of pellet technology and, later, extraction of the key principles using liquid carbon dioxide allowed enhanced utilization of resins and oils and flexibility of product design.

Styles of Beer

One of the most fundamental approaches to classifying beers is based on whether they are produced by

'top fermentation' or 'bottom fermentation,' which basically refers to whether the yeast congregates at the top of the vessel or sinks to the base. With modern cylindroconical fermenters and their high hydrostatic pressures there tends to be less of a distinction.

Top fermentation, with its vastly older pedigree, tends to be performed at relatively warm temperatures (15–25 °C), under which conditions yeast tends to produce higher levels of flavor volatiles such as esters, affording fruity characteristics. Bottom-fermentation beers are produced at much lower temperatures (say 6–15 °C) and frequently possess significant sulfury notes.

Top-Fermentation Beers

Ales Perhaps the major product in the genre is the English pale ale, which is referred to as bitter when dispensed on tap (draught). Alcohol content will be in the range 3–7.5% by volume (ABV). Classically it is produced from well-modified malt, kilned to a relatively high-temperature regime to impart a copper color. The products are relatively malty, especially in the case of Scotch ales, which may be sweeter and darker than their English counterparts. After fermentation ales are traditionally primed with extra sugars, 'dry-hopped' with a handful of cones and dosed with isinglass finings prior to racking in casks. Residual yeast uses the sugars to introduce natural carbonation (albeit at a low level of perhaps 2.5 g l^{-1} CO_2), prior to being sedimented along with other insolubles by the finings. The product is not pasteurized and is distributed for retail in pubs, whether by gravity dispense or hand-pumping via a beer engine. Because of the skill needed to deal with cask ales (if infections and clarity problems are to be avoided), it has become increasingly common for such ales to be conditioned in the brewery and pasteurized prior to packaging in keg (for draught dispense) or bottle or can. Such products have a carbon dioxide similar to that of other beers sold from kegs or small packs, between 4.5 g l^{-1} for large packs and up to 6 g l^{-1} for small packs. Lower carbonation products in both keg and small packs have become increasingly prevalent over the past decade. They incorporate nitrogen gas, which enhances foam stability and also affords a smooth texture to the product. Because of the greatly reduced tendency for foam to form at the lower carbonation levels, it is common to use nucleation devices called 'widgets' in cans and even bottles containing beer of this type.

Mild ale, a style which is in decline and largely perceived as old-fashioned, is a sweeter, darker product, the color being either due to caramel or in part to a low proportion of heavily kilned malt, though not so much as to impart burnt flavors. It tends to have a lower alcohol content (less than 3.5% ABV) and when bottled may be referred to as brown ale. Traditional terms in Scotland have been heavy, for their stronger bitters, and light, for the milds.

Barley wines are fermented at very high gravities (the higher the specific gravity, the greater the content of sugars available for fermentation), which is an additional factor impacting on the very high ester/fruity character and, of course, accounting for a very high alcohol content of up to 10% ABV. They are usually sold in smaller volumes, in bottles called nips.

Porter In early eighteenth-century London there were three main categories of beer: ale, which was strong, the weaker beer, and two-penny, a better quality of beer. There were brown, pale, and amber versions of each. People tended to ask for 'half and half' – equal measures of ale and beer – or 'two-thirds' – (three threads): ale, beer and 'tuppenny' in equal measure. One far-sighted brewer in 1722 conceived of a product based on the second of these but with the three styles premixed in the brewery, and as a consequence saving the landlord's time. Because most of the customers wanting their beer in a hurry were porters in the local markets, it became known as porter. There are few, if any, genuine porters to be had these days. They are traditionally very dark, due to the use of roasted barley, and not overwhelmingly strong (about 5% ABV).

Stout A close relative of porter, stout originated in Ireland, of course, with intense color and burnt, smoky flavors due to the use of roasted barley adjuncts and high bitterness. These robust flavor characters have been tempered for many years now by the use of nitrogen gas, which 'smoothes' the palate as well as affording the rich, white, and creamy foam. Alcohol content may be between 4 and 7%, with up to 10% in imperial stouts. Sweet stouts are a British variant, of lower alcohol content (up to 4% ABV), with less roast character (often due to the use of caramel and less roast barley as colorant).

Belgium top-fermentation beers There is a greater diversity of beer styles in Belgium than any other country. Trappist beers are relatively dark, intensely bitter, acidic products of up to 12.5% ABV, with clear fruity notes. The Lambic and Gueuze products possess a diversity of complex flavor characteristics, imparted through the agency of the metabolic activities of a more complex microflora than brewing yeast alone. They tend to be quite sour (low pH) and are frequently not clear. Various flavorants may be added

to certain Belgian beers, including cherries (Kriek) or raspberries (Framboise).

German ales Alt from the German means 'old,' is brewed 'in the old way' from a grist of barley and (to a lesser extent) wheat malts. It is a dark-copper color, relatively bitter, and approximately 5% ABV. Kolsch, from Cologne, is a much lighter beer, of similar alcohol content. The darker malts used in the production of alt are replaced by a less intensely kilned malt, called Vienna malt.

German wheat beers Weizenbier is made from a grist comprising at least 50% wheat malt. The products are relatively highly carbonated, affording a refreshing character alongside the fruity and phenolic (clove-like) characters. Kristallweizen denotes that the beers have been filtered to remove yeast, whereas Hefeweizens are cloudy due to yeast, which is traditionally employed to carbonate the bottled product through 'natural conditioning.' The products are relatively lightly colored (straw-like) and have alcohol contents of 5–6% ABV. Weissbier ('white beer') is much weaker (e.g., 2.8% ABV), and made from a grist of less than 50% wheat malt, with the addition of lactic acid bacteria to generate a low pH of 3.2–3.4. Accordingly such beers are quite sour, and may be taken with raspberry or sweet woodruff syrups.

Bottom-fermentation beers

The classic style originated in mid nineteenth-century Bohemia, in the Pilsen city brewhouse. It is quite a malty brew (the malt traditionally being relatively undermodified), typically with 4.8–5.1% ABV and a pale gold color. Particularly prized is the late hop character, afforded by retaining a proportion of the hops for addition late in the kettle boil to ensure that the desired fraction of essential oils presents.

All too often the term *lager* is used synonymously with pilsner (pils). Lager as a term is really an umbrella description for all relatively pale brews, fermented and dispensed at low temperatures.

The greatest breadth of styles of this type of product is to be found in Germany. To many the distinction between some of the styles is subtle and, for those seeking to replicate them internationally, may own as much to marketing as to major differences in brewing practice.

Export is a little stronger than Pilsner (5.5% ABV) and darker, though somewhat less bitter.

Bock is a stronger beer (6–8% ABV) with sulfury and malty flavors and colors ranging from straw to dark brown. The related Doppelbock may contain up to 12% ABV.

Marzen ('March beer') was historically brewed to a high strength (up to 6.5% ABV) with pale and dark versions. It has 'middle-of-the-road' characters, with some toasted notes.

Helles (4.5–5.5% ABV) is pale amber, very malty, with a low bitterness and hop character.

Dunkel has a similar alcohol content but is copper-brown, whilst Schwarzbier is black, dry, and between 3.8 and 5% ABV.

Malt liquor is a term used to describe alcoholic products (6–7.5% ABV) which are very pale, very lightly hopped and quite malty and sweet, and originating in the USA.

The other lager-style product indigenous to the USA is steam beer. The origins of this product can be traced to the California gold rush and a demand for light and refreshing drinks despite the unavailability of ice for cold storage and conditioning. Bottom yeasts were used at warmer fermentation temperatures in shallow vessels into which the 'steaming' wort was introduced to cool.

Other classifications of beer

There is considerable blurring of the boundaries that divide the classic beer styles. For example, products both light and dark have been developed with all manner of additional flavors, such as citrus character, cf. the traditional mixers of ale with lemonade (shandy) or lager with a splash of lime or lemon (lager top). Flavors as diverse as oyster, chocolate, coriander, and chilli have been used. Black lagers have been developed, as indeed have colorless ones. There is probably no limit to the ingenuity of new product development teams, although most of these products tend to be quite short-lived as brands.

Ice beers The original provenance of this style stems from the trick (mostly in Canada) of concentrating beer by placing it on the window sill in the heart of winter, under which conditions water freezes out first, leaving a more alcoholic and warming solution behind. In fact, the freeze concentration technique used in the manufacture of such beers in the brewery produces ice in only sparing amounts, but there is a loss of some of the tannin materials, which may make these products slightly less astringent than the norm.

Dry beers These are lagers with a relatively low proportion of residual sugars making for a low sweetness. They are related to the light beers, the most rapidly growing genre. Standard beers contain a proportion of carbohydrate surviving the fermentation process, whereas a light beer has most or all of this sugar converted. These beers therefore have fewer

calories, provided that the extra alcohol is diluted to the level found in 'normal' beers. They are produced using a more convoluted mashing procedure to allow as much starch conversion as possible, followed by the use of malt extracts (or microbial enzymes if the brewer is content to use additives) in the fermenter.

Draft beers The term 'draft' ('draught') is nowadays used not only to describe large pack dispense (keg, cask) but also nonpasteurized beers in small pack.

Non- and low-alcohol beers There are many definitions worldwide about what constitutes low-alcohol products. Perhaps the most stringent is in the UK, where non- and low-alcohol beers contain less than 0.05% or 1.2% ABV respectively. They are produced either by removing the alcohol from a full-strength brew, by techniques such as vacuum distillation or reverse osmosis, or by restricting the ability of yeast to ferment wort, either by making a wort containing very low levels of fermentable sugars or by ensuring that the contact between yeast and wort is at a very low temperature and for a relatively brief time.

See also: **Beers**: Biochemistry of Fermentation

Further Reading

Bamforth C (1998) *Beer: Tap into the Art and Science of Brewing*. New York: Insight.

Corran HS (1975) *A History of Brewing*. Newton Abbot: David and Charles.

Raw Materials

P Hughes, Heineken Technical Services, Zoeterwoude, The Netherlands

Background

The production of beer has a long history, dating back to the Sumerians and the Egyptians. A unique beverage, the quality of beer is substantially dependent on the identity and quality of its raw materials. The major raw materials of beer are cereals and hops. Both have a number of roles in beer production and final beer quality, but the cereals principally provide sugars and amino acids for fermentation. Hops are used to provide a unique flavor and to stabilize the beer foam. Both can be processed for optimal and more flexible usage for beer production. Here, aspects of hops, and cereals as they pertain to the brewing industry, are described. For clarity, brewing specific terms are defined in **Table 1**.

Hops

The cultivated hop (*Humulus lupulus* L.) was originally added to beers during their production for flavor and to extend the beer shelf-life. Historically, herbs have also been used, but today, essentially only the hop is used commercially. Hops are cultivated solely for beer production and grow almost exclusively in the temperate regions of the world, between the latitudes of 35 and 55°, both north and south of the equator. Owing to continuing improvements in the utilization of hops and the gradual reduction of beer bitterness levels, the global production of hops is currently decreasing (**Table 2**), whereas beer production is increasing on a world-wide basis.

The hop plant is a hardy, climbing, herbaceous perennial plant. The rootstock stays in the ground from year to year. The vines that grow from the rootstock are not self-supporting, and the provision of a support structure or wirework is a significant capital investment when setting up a hop garden. Traditionally, hop plants attained heights in the region of 5 m, but the recent development of commercial dwarf varieties (*c*. 2.5 m in height) allows more cost-effective supporting structures such as hedges to be used.

Hops for Beer Production

The brewing value of the hop is embodied in its cones. These are structures that bear small, physically robust yellow lupulin glands. It is these glands that contain the hop essential oil and the hop acids that are responsible for hoppy and bitter flavors, respectively. Traditionally, cones were added to the wort boil, during which time extraction and chemical conversion of the appropriate brewing value occurs. However, neither the extraction nor chemical conversion is particularly efficient during a typical beer production process, and an industry has grown up, whereby hop cones are processed to enhance efficiency, or utilization, of this raw material. Today, around 85% of all hop products used are either extracts (45%) or pelleted hops (40%). The remainder is made up of reduced extracts and the unprocessed hop cones themselves (**Table 3**). These various products are used at a number of stages during beer production (**Figure 1**).

Hop Pests and Diseases

These are numerous and can have widespread effects on the resulting crop. The sensitivity of hops to various pests and diseases is in some cases dependent on

Table 1 Definitions of common malting and brewing terms

Term	*Definition*
Boiling	Operation that follows lautering. The sweet wort is boiled to ensure its sterilization. α-Acids added with hops or hop products are partially isomerized here, and there is also substantial development of solids that are later removed. Yeast is added to the cooled product to begin fermentation.
Fermentation	Yeast added to the cooled boiled wort uses the sugars, free amino acids and residual lipid present to grow and generate ethanol.
Lautering	A common process in many breweries. The mash is allowed to run through a coarse slotted plate. Insoluble husk material settles on to this plate and creates an effective filter bed.
Malting	Process of partial germination to convert barley (or, less commonly, wheat) into malt.
Modification	In terms of malt, this indicates the degree of biopolymer degradation in the grain.
Pasteurization	Commonly applied operation that neutralizes any residual bacteria in the final product by controlled heating. Some companies prefer to sterile filter their product.
Run-off	Refers to the gravitational separation of malt extract (wort) from residual solids, usually during the lautering operation. This can be seriously hampered by a number of wort constituents, such as β-glucan gums.
Trub	Solids formed during wort boiling and subsequent cooling. Mainly composed of protein–polyphenol complexes.
Utilization	Traditionally refers to the ratio of the moles of bitter iso-α-acids recovered in the final beer to moles of α-acids added to the wort boil. Now often modified to include efficiency of isomerized hop product dosing.
Wort (sweet)	The liquid that recovered, for instance after lautering. The term 'sweet' refers to the high levels of fermentable sugars that are released during the mashing operation. The addition of hops at wort boiling gives a bitter wort.
Whirlpool	A common device in breweries for effective solids removal from hot boiled wort. Wort is passed into the whirlpool vessel tangentially, and the solids present accumulate at the center and deposit as a cone at the base of the vessel.

Table 2 Global production of hops

Country	*Global production (percentage of global production)*	
	By area (hectares)	*By mass (tonnes)*
Germany	18 600 (31.5)	29 290 (30.3)
USA	14 630 (24.8)	30 650 (31.7)
Czech Republic	6 110 (10.4)	4 890 (5.1)
China	4 930 (8.4)	13 000 (13.5)
Poland	2 250 (3.8)	3 060 (3.2)
England	1 980 (3.4)	2 800 (2.9)
Slovenia	1 620 (2.7)	1 760 (1.8)
Russia	1 520 (2.6)	820 (0.8)
Total	59 020 (100)	96 610 (100)

Source: The Barth Report (2000/2001).

Table 3 Distribution of global hop sales

Hop product class	*Proportion of total sales (%)*	*Comments*
Cones	7	Traditional. Use declining
Bittering pellets	19	Enhanced control of bitterness
Aroma pellets	22	Enhanced control of hoppy aroma
Extracts	44	Highly concentrated aroma/ bitter products
Reduced extracts	8	Chemically modified by selective reduction

variety, so that breeding programs, such as those at the Hop Research Unit at Wye College, specifically test for resistance to a range of challenges as part of its selection criteria. For instance, resistance to verticillium wilt, endemic in some parts of the UK, is essential here for acceptance to the grower.

The Hop Market

Hop producers rely almost solely on brewers to purchase their product. In previous years, the susceptibility of the hop crop to a range of pests and diseases has affected their market price. This resulted in the practice of the forward contract, whereby a grower or merchant would undertake to supply certain quantities of hops or hop products to a buyer for a given period. Hops that were not put aside for forward contracts were sold on the spot market. At the end of the 1990s, the hop industry suffered from overproduction, so hops and hop products were available for unprecedented low prices on the spot market, and forward contracts became unpopular. Many growers, particularly in Europe, ceased growing what became an uneconomical crop. The cost of the infrastructure to grow hops and the need for two to three seasons before a high-quality crop can be expected mean that there will be an inertia for farmers to turn back to hops. At the beginning of the new millennium, the reductions in acreage worldwide have begun to redress the supply and demand balance.

Impact on Beer Production and Quality

The use of hops in brewing is famous for the provision of bitterness in the final beer. However, they have a number of other effects on the final beer quality and on the production process itself.

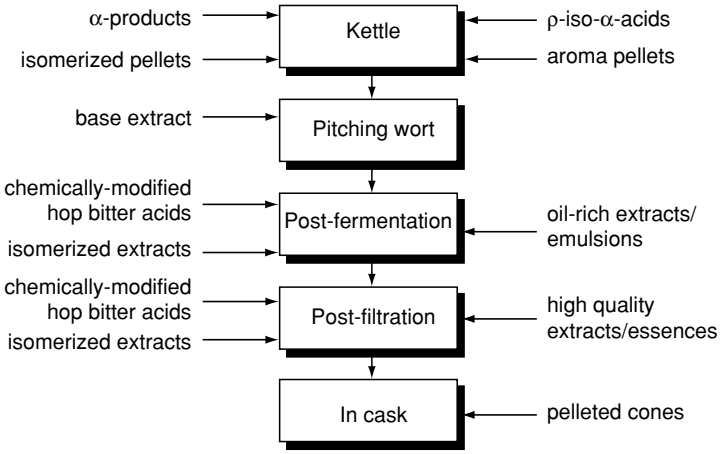

Figure 1 Points-of-addition of hops and hop products during beer production. Generally, the boiling stage is mandatory if isomerization of the α-acids is necessary.

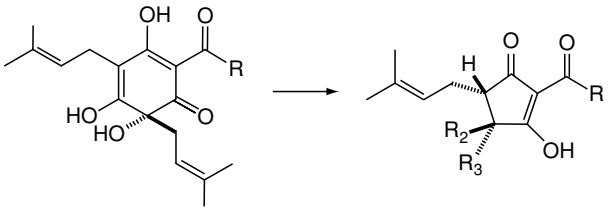

Figure 2 Isomerization of α-acids to the bitter iso-α-acids. The mechanism of the reaction is unclear, but the result is a ring contraction and the production of *cis*- and *trans*-isomers. R represents isopropyl, isobutyl, or *sec*-butyl substituents. R_2 and R_3 are either hydroxyl or 4-methylpen-3-enoyl, depending on whether a *cis*- or *trans* isomer is formed.

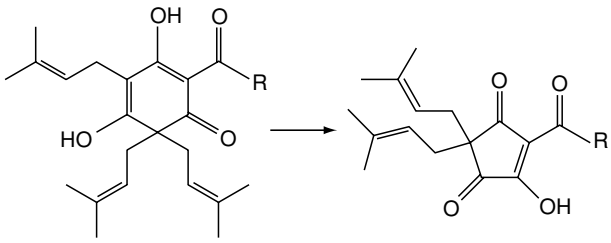

Figure 3 Degradation of hop β-acids to bitter hulupones. This reaction is far from quantitative, although it is believed by some to impart a good-quality bitterness to the final product. See Figure 2 for the definition of R.

Bitterness Most, if not all, of the bitterness of beer is due to the presence of the iso-α-acids. These are a mixture of six major components, which are three stereoisomeric pairs of compounds derived from each of the three hop α-acids (Figure 2). These α-acids are isomerized to the iso-α-acids during the wort boiling stage, although in practice, the final utilization is around 30–40%. In contrast, isomerization by hop processing gives yields in excess of 90%. There are probably minor differences in the bitterness of the individual compounds. Nevertheless, whilst hops come in a wide range of varieties, current thinking is that variety has little effect on bitterness intensity or quality. High-α-acid hop varieties have become a lucrative breeding target, because of the generally favorable yields of α-acids per hectare. Today, α-acids are considered a commodity on the hop market. There are some who consider the cohumulone content of the α-acids to have a poorer-quality bitterness, but to date, this has not been conclusively demonstrated.

Hops also contain the β-acids. Although these do not undergo isomerization in the sense of α-acids, they do form bitter degradation products – the hulupones – when present during wort boiling (Figure 3). In practice, this contribution to bitterness is usually minor.

Hoppy flavor Hop cones have a characteristic aroma, which is particularly evident when hops are rubbed between the hands. This is due to the presence of over 300 volatile compounds in the essential oil of hops. In addition to volatility, some hop oil compounds are chemically reactive, so that after wort boiling, very little of the original hop aroma compounds remain, with some being lost completely. Nevertheless, what is left, subject to the impact of fermentation, stays in the beer to give a unique flavor. Beers vary substantially in the intensity and quality of their hoppy flavor. The intensity is dependent on the brewing practice and dosing level, whereas the quality is mostly dependent on the varieties used

Table 4 Common hop varieties widely available to brewers, in decreasing order of tonnage (these varieties account for > 60% of all hops produced in 2000)

Hop variety	*Growing areas*[a]
Qingdao Flower	China
Nugget	USA, Germany, Spain
Perle	Germany, (USA)
Hallertauer Magnum	Germany, (Spain)
Galena	USA
Willamette	USA
Zatec/Saaz	Czech Republic, Slovenia
Spalt Select	Germany
Hersbruck	Germany
Hallertauer Tradition	Germany
Zeus	USA
Northern Brewer	Germany

[a]Countries in brackets are very minor growers of variety indicated.
Source: The Barth Report (2000/2001).

(Table 4). Hops with perceived high-quality aromas can command a premium price in the marketplace. This is in contrast to hops sold for bitterness, where trading is often carried out on a kilogram α-acid basis.

Foam The chemical structure of the iso-α-acids is conducive to beer foam stabilization. Beer foam is predominantly made up of malt-derived polypeptides, and the iso-α-acids can cross-link these polypeptides to provide a more stable foam structure. Furthermore, the foam that adheres to the inside of a beer glass during consumption – the lacing or cling – is reliant on the presence of iso-α-acids in beer.

Antibacterial potency The initial reason for adding hops to beer – to stabilize it against flavor deterioration – is well founded. At levels found in beers and at the pH of beer (*c*. 3.9–4.5), the iso-α-acids effectively kill many Gram-positive bacteria. This bacteriostatic effect is aided by the presence of ethanol. Given the global trends for increased beer pH and decreased bitterness of beer, there is a school of thought that considers that this mode of beer protection could be compromised. In any case, other measures, such as good brewery hygiene, sterile filtration, and pasteurization also help to insure that beer is not affected by microbial action.

Hops and Beer Production

The traditional use of hop cones themselves meant that a significant quantity of leaf material needed to be removed from the wort after the boiling stage. This was done with a hop back, a specifically designed sieve to clarify the wort. The widespread use of pellets and extracts means that, today, the hop back is absent from many modern breweries. Alternative clarification procedures were therefore required, particularly for the remains of pellet powder. Based on an initial discovery by Einstein, the whirlpool was developed. This is a vessel whereby the wort, after boiling, is pumped in tangentially. The circular motion of the liquid results in the deposition of solids (trub) in the center of the vessel, where it sinks to the bottom. This development signified the beginning of hop-pellet penetration into the brewing industry.

The presence of polyphenols in hops also assists in the precipitation of excess polypeptides extracted from the malt. If this precipitation is limited, there is an increased risk of the final beer producing a haze during its shelf-life. These polyphenols are also attracting interest because of their antioxidant properties, although, as yet, the complexity of these components in hops means that few specific details on their properties are known.

Quality-control Aspects of Hops

Hops and hop products purchased principally for bittering the final beer are assayed for α-acids. Today, this is commonly achieved by high-performance liquid chromatography (HPLC), although traditionally, this was carried out using spectrophotometry or the measurement of the so-called lead conductance value. For aroma attributes, judgements are primarily on the basis of the name of the variety: there is no specific test available for predicting hop aroma performance. The time-honored approach of rating hops according to their visual characteristics and the aroma released on rubbing hop cones in the hands is still very much in use.

Hops and hop products should be stored cold (but above 0 °C), and, where possible, sealed packages should not be breached until the contents are required. Both aroma and bittering components of hops are sensitive to heat and oxidation, although there is some varietal dependence of the rate of hop degradation during storage. (*See* **Chromatography**: High-performance Liquid Chromatography.)

Cereals

Cereals form the basis of beer. Although quantites vary, for a typical beer, some 200 g of cereals per liter of beer are used during production. Cereals provide the carbohydrates and amino acids necessary for the yeast to grow, generate ethanol and carbon dioxide, and produce key flavors in the final beer. Malted wheat and rye are used in some German beers, and malted sorghum is used in some African beers. Raw cereals, in the form of flour or grits, and cooked cereals, generally as flakes or torrified (syn. torrefied)

are still widely used as adjuncts to form part of the grist. Today, most brewing malt (with the exception of some specialist beers made from wheat malt) is made from barley, since it has been recognized that in certain essential properties, barley is superior to many other cereals. In any case, with the exception of the Japanese happoshu beer brands, which may use as little as 25% malt in the grist, barley malt makes up at least 40% of the total grist. Nevertheless, perhaps with the exception of the Japanese brewing industry, the proportion of extract derived from malt remained essentially constant over much of the twentieth century.

Malted barley, as well as adjuncts and syrups, are usually assessed for extract yield. This is the quantity of sugars released per unit mass of malt, adjunct, or syrup. This extract yield relates directly to the quantities required and therefore the impact on cost. Extract yield, though, is measured by density and reflects the total sugar content. However, a proportion of the sugar content is nonfermentable, and this is assessed by a further measurement – that of fermentability. This is determined by carrying out a test fermentation and measuring the maximum amount of extract consumed. Again, this is done by density measurements. The remaining sugars are termed dextrins. Whilst they may not contribute directly to the yield of ethanol, dextrins may have a subtle role to play in the sensory qualities of the final beer. Two other common measurements are the total nitrogen content, a measure of protein content, and, in malt, *S*-methylmethionine. The significance of the latter is that it is readily degraded to yield dimethyl sulfide, a distinctive sulfur compound, with an aroma reminiscent of sweetcorn. At the correct levels, it is a key flavor impact component of many beers, particularly lager-style beers. For pale malts, the levels of enzyme activity – especially α- and β-amylase, are important parameters for satisfactory performance of the malt in the brewery. Typical analytical data for various cereals and their products are given in Table 5. (*See* **Cereals**: Contribution to the Diet.)

Barley

The most popular cereal is barley that has been malted. Its use has several advantages: it has a suitable level of protein (8–12% dry weight) and low levels of lipids, it develops high levels of amylolytic enzymes during germination, and it has a husk that assists in the separation of insoluble material from the required extract.

Barley is essentially a temperate plant that is nevertheless distributed widely in both tropical and subtropical zones. The main commercial growing areas of barley are in the grassland belt, across the northern hemisphere from the northern USA and Canada, across northern Europe into Asia and North Africa, where it probably originated. Again, in the southern hemisphere, barley is grown commercially in South America, on the fertile uplands of South Africa and in Australia and New Zealand. (*See* **Barley**.)

Barley malting The malting of barley has several purposes. Firstly, the development of amylolytic enzymes is essential for the subsequent brewing operation, as these are required for the formation of fermentable sugars from the malt starch. This diastatic power is more than enough for conventional brewing, so this can be diluted to some extent with nondiastatic materials, such as syrups and speciality malts. Secondly, on a practical level, it is desirable to reduce or eliminate the β-glucans present in barley. These form gels that substantially slow down or even stop the separation of the sweet wort from the residual solid material. Thirdly, there needs to be enough free amino acids released, either during malting or in the mashing stage of beer production, to support the yeast growth required during the early stages of fermentation. Finally, in terms of flavor, 'grainy' or 'green' flavors should be eliminated, to be replaced by desirable 'malty' flavors. Other aspects are assuming importance, namely the reduction in levels of agrochemical residues during malting, which in turn means lower levels finding their way into the final

Table 5 Typical malt specifications

Parameter	*Level*	*Applicability*
Malt		
Moisture (%, w/w)	5	To ensure storage stability
Hot water extract (l °kg^{-1})	310	Indicates the amount of soluble extract available
Color (°EBC)	3	Has a large impact on final beer color
S-Methylmethionine (mg kg^{-1})	4.5	Indicative of likely dimethyl sulfide levels in final beer
Total nitrogen content (%, w/w)	< 1.65	Ratio with soluble nitrogen gives an indication of the degree of malting (modification)
Total soluble nitrogen (%, w/w)	0.70	Indicates the amount of soluble nitrogen (protein)
Free amino nitrogen (mg l^{-1})	140	Indicates the degree of proteolysis during malting

beer. This also applies to the presence of barley microflora, which can be accompanied by a range of mycotoxins.

The malting process involves several stages. The malt, once produced, is considered to be relatively stable if the grain moisture levels are below about 12% (w/w), and can be stored under appropriate conditions for many months. Low atmospheric moisture and attention to possible fungal growth are essential during prolonged storage. The final malt color will depend on the kilning regime used to dry the malt so that it is stable during storage. Thus, so-called lager malts are dried at temperatures well below 100 °C, whereas an ale malt is dried at around 100 °C. These higher temperatures stimulate nonenzymic browning reactions between sugars and free amino compounds, thus resulting in a darker color and, generally, a higher degree of flavor. Crucially, though, both still retain enzyme activity essential for the mashing operation. The degree of enzyme development is termed 'modification' and has an impact on how the malt is used in the brewery.

Brewing with malt The early stages of beer production have been tailored to the properties of malt. Thus, to effect recovery of as much brewing value as possible from the malt, it must be ground, and the enzymes generated during malting need to be activated (Table 6). There are some variations on the milling and mashing operations, but essentially, milled malt is blended with water in a specific ratio, typically 1:3 (w/w).

Specialty malts Some beers are brewed with relatively low concentrations of specialty malts. Generally, these retain little enzymic activity, but are used because of the specific flavor or color attributes that they confer on the final beer (Table 7). Whilst not a malt, roasted barley may also be considered, together with specialty malts, as it is used for similar reasons. Many darker beers contain small percentages of specialty malts as part of the grist.

Adjuncts

This is a term reserved for nonbarley malt derived cereal substitutes in beer production. Whilst barley malt is best tailored to the requirements of brewing, there are good reasons for partial substitution of barley malt for other cereals. These include flavor modification, increasing brewery capacity, potentially favorable costs (malting barley often commands a premium price), enhancing the stability of the final product with respect to haze formation, and as a point of beer brand differentiation (e.g., oatmeal stouts, wheat beers). The lack of diastatic power in adjuncts is not usually a serious issue as good-quality malt is characterized by having amylolytic activity surplus to requirements. As a rule of thumb, there is little need to be concerned about the final diastatic power if the level of diastatic malt in the grist is at 70% or higher.

Adjuncts come in a variety of forms. Some of these forms generally require a specific brewery plant, the installation of which implies a future commitment to the chosen adjuncts. A brewer therefore needs to be

Table 7 Typical color and flavor descriptor attributes of specialty malts

Specialty malt	*Color*	*Typical flavor characteristics*
Pale malt	4.5–4.8	Biscuit-like
Caramalt	25–35	Sweet, nuitty, cereal, toffee
Crystal	100–300	Malty, toffee, caramel
Amber	40–60	Nutty, caramel, fruity
Chocolate	900–1200	Mocha, treacle, chocolate
Black	1250–1500	Smoky, coffee
Roasted barley[a]	1000–1550	Burnt, smoky

[a]Not strictly a malt, but included here for completeness.

Table 6 Some enzyme activities present in milled malted barley during mashing

Type	*Enzyme*	*Activity*
Amylases	α-Amylase	An endoenzyme, mainly responsible for the breakdown of high-molecular-weight starch
	β-Amylase	An exoenzyme that releases maltose
	Limit dextrinase	Breaks down branch points of starch
	α-Glucosidase	An exoenzyme that releases glucose
Proteases	Endopeptidases	A sulfydryl protease of general activity
	Exopeptidases	Important for free amino nitrogen generation
	Aminopeptidases	Exopeptidase that attacks the amino end of protein
	Carboxypeptidases	Most important proteases, attacking the carboxyl end of protein
Other carbohydrases	β-Glucan solubilase	Releases β-glucan from cell walls
	Endo-β-glucanases	Break down potentially troublesome β-glucans
	Pentosanases (xylanase)	Degrade pentosans
	Xylosidase	Breaks down xylan polymers

sure that the source of required adjunct is and will be available at the required quality for the foreseeable future.

Raw cereal grains In principle, suitably milled barley, wheat, rye, and *Triticale* grains can be used during mashing without prior treatment. Of these, most interest has focused on the use of whole barley. Careful milling of the cleaned barley, preferably retaining the husk as intact as possible, provides a suitable grist addition. Hammer milling of raw cereal grains is generally avoided, as problems with β-glucan extraction and subsequent run-off difficulties can occur. If the proportion of raw barley in the mash is high, enzyme supplementation is highly desirable to degrade the proteins and carbohydrates present. As a total replacement for malt, raw barley is financially attractive, although a balanced mixture of enzymes is essential to give the required extract for subsequent fermentation.

Cooked intact cereal grains Roasted barley, torrefied or micronized barley, and wheat grains contain no active enzymes. However, their preheating results in partly precooked starch and a disrupted grain structure. Such products are readily milled and easily handled, giving extracts generally higher than those for raw cereals. The heat also partly degrades β-glucans, so that the resulting worts from torrefied grains are less viscous than those from raw materials.

Grits Grits may be defined as uncooked, nearly pure fragments of starchy endosperm from cereal grains. If they are used directly in the brewery, they must be cooked or flaked (see below). The removal of other parts of the grain reduces the levels of lipid, ash, and fiber. Rice grits, or broken rice, are often a byproduct of debranning and milling rice for human consumption. Nowadays, milling efficiency has improved such that less of this byproduct is produced. Rice grits are generally low in lipid and fiber but must be cooked well if adequate extract is to be recovered. There are substantial differences in the suitability of various rice cultivars for beer production, but rice is considered to be flavor-neutral and is therefore a popular adjunct.

Maize grits are the most common adjunct used in the USA. They can be prepared in a number of ways, but in any case, the aim is to remove the outer layers and the lipid-rich germ. The resulting grits are rich in starch and contain much less lipid than their parent grain. They are popular, as, after cooking, they yield a good extract during mashing.

For use in the brewery, all grits must be cooked to disrupt them and gelatinize the starch. They are then usually mixed with conventional barley malt mash to complete the starch conversion. The grits are mixed with a small proportion of highly diastatic malt or with microbial amylase and then heated. Temperatures may be raised to boiling or cooking can be carried out under pressure to effect gelatinization. The swollen and gelatinized starch is partly degraded or liquefied before the heat inactivates the amylase activity. Consequently, viscosity neither increases to unmanageable levels nor sets to a gel on cooling. This is important from a practical point of view, as it can be pumped and does not burn on to the heating surfaces. Such burning can lead to problems such as flavor pick-up, color generation, and deteriorating heat-transfer characteristics.

Flaked cereals These are produced either by cooking whole grains with steam or by micronization. In both cases, the heated grain is passed through feed rolls to flaking rolls, typically held at 85 °C. For the more traditional steam process, the flakes are dried prior to packaging. The flakes may also be lightly crushed, as the flakes themselves have a low bulk density. Because the starch is already gelatinized, the flakes do not need cooking and can be added directly to the mash. Flakes also produce relatively little soluble nitrogen, so they can be used as a wort nitrogen diluent.

A range of cereals are flaked. Whilst flaked rice is popular because of its good extract yield and neutral flavor and color, it is not used for economic reasons. Flaked maize grits and flaked whole wheat grains and barley grains are currently in use.

Flours In principle, both wheat and barley can be hammer-milled in the brewery to produce coarse flours. These are suitable for addition directly into the mash. Nevertheless, flour milling is a rather specialist process, and in practice, the brewer buys flour from certain suppliers.

Syrups Sucrose, and its hydrolysis product, inverted sugar, are used by some brewers during beer production. Produced either from sugar cane or sugar beet, the range of products available is large, in terms of color, mineral content, and residual syrup present. They can be supplied for the special flavors that they can confer. These can include descriptors such as burnt, raw sugar, and bland.

Syrups and sugars from the partial or essentially complete hydrolysis of refined maize starch, unlike beet or cane syrups, contain no significant quantites of nitrogenous components or other yeast nutrients. Thus, like sucrose syrups, they are wort extenders rather than wort replacers. Their production requires

two hydrolytic stages. The first is the solubilization of starch slurry to form dextrins. Here, the starch is completely solubilized, so much so that it does not form gels or retrograde (precipitate) on cooling. In the second stage, the dextrins of the first stage are converted into the desired spectrum of sugars. Dilute mineral acid and enzymes can be used in both stages to accurately control the desired result. (*See* **Syrups**.)

Malt extracts and wort-replacement syrups Malt extracts are essentially sweet wort (i.e., unhopped), from which the water has been removed under reduced pressure. Although the extracts may be either diastatic or nondiastatic, the former are rarely, if ever, used. Malt extracts are an expensive source of extract, because the malt itself is relatively expensive, as are the operations for water removal.

Summary

The hop- and beer-production industries are, for the foreseeable future, inextricably linked, as there is no viable alternative to hops that brewers can employ and few, if any, other outlets for hops other than brewing. Like hops, cereals are essential to beer, and indeed, the two have been intimately linked ever since beer was first produced. Specific barley varieties are bred and tested for their performance in all aspects of agronomic, malting, and brewing characteristics.

See also: **Beers**: History and Types; Wort Production; Chemistry of Brewing; **Caramel**: Properties and Analysis; **Cereals**: Contribution to the Diet; **Chromatography**: High-performance Liquid Chromatography; **Malt**: Malt Types and Products; Chemistry of Malting

Further Reading

Bamforth CW (1998) *Beer – Tap into the Art and Science of Brewing*. London: Plenum Press.

Barth HJ, Klinke C and Schmidt C (1994) *The Hop Atlas*. Nuremburg: Joh. Barth & Sohn.

Briggs DE, Hough JS, Stevens R and Young TW (1994) *Malting and Brewing Science, Volume I, Malt and Sweet Wort*, 2nd edn. London: Chapman & Hall.

Corran HS (1975) *A History of Brewing*. Newton Abbot, UK: David & Charles.

Hough JS, Briggs DE, Stevens R and Young TW (1993) *Malting and Brewing Science, Volume II, Hopped Wort and Beer*, 2nd edn. London: Chapman & Hall.

Hughes PS and Baxter ED (2001) *Beer: Quality, Safety and Nutritional Aspects*. Chemistry: Royal Society of Chemistry.

Kunze W (1999) *Technology – Brewing and Malting, International Edition*. Berlin: VLB Berlin.

Neve RA (1991) *Hops*. London: Chapman & Hall.

Verzele M and de Keukeleire D (1991) *Chemistry and Analysis of Hop and Beer Bitter Acids*. London: Elsevier.

Wort Production

M Ortega-Heras and M L González-Sanjosé, University of Burgos, Burgos, Spain

Background

One of the most common and simple definitions of beer is: 'beer is the drink resulting from the fermentation by means of selected yeast of the wort, which is obtained from malted barley, alone or mixed with other starch sources, and it is flavored by boiling with hop flowers.'

Brewing process can be divided into three phases: malting, wort production and fermentation, and maturation. Starting with the malting process, this includes different phases, such as steeping, germination, drying, and toasting. In this way, barley is transformed into malt.

During these phases, barley undergoes the following changes:

1. Steeping: The barley grain is humidified in large tanks of water into which air is bubbled to prevent anaerobic conditions developing. The temperature is also controlled. When this phase ends, the barley grain has a humidity of between 42 and 46%.
2. Germination of the grain: This involves partial development of the embrion, activating several endogenous enzymatic systems, such as the α- and β-amylases and the proteases, which are capable of hydrolyzing the starch and proteins, respectively. This germinated barley is known as 'green malt.'
3. Drying and toasting: In this phase, the grain is dried, the embrion is destroyed, and the enzymatic activity is stopped to avoid using up the reverse substances. In this way, the malt is obtained.

The aroma and taste of the toasted grain are formed simultaneously by Maillard reactions. The beer color is lighter or darker depending on the toasting of the malt.

The second phase includes different processes that convert malt into wort. Wort is a liquid extract containing fermentable sugars (sweet wort), amino acids, vitamins, and other compounds, and it is the principal substrate in which beer is obtained by fermentation.

Before fermentation, the wort is boiled with hop flowers to flavor it (sweet hopped wort).

The third phase is associated with wort fermentation and maturation, a process in which yeast transforms the wort into beer, by consuming sugars and releasing ethanol and CO_2. This phase ends with the filtration process, to produce a clear beer, and then bottling.

Nowadays, only some breweries have their own malt factory, and most of them have to buy the malt, as well as several other raw materials. In such cases, the first stage of the brewing process is wort production.

Wort Production

During this phase, malt and other starch sources (adjuncts) undergo different processes, in which the main aim is to obtain a fermentable extract, the wort.

Adjuncts are used in order to reduce the quantity of malt used, since malt is more expensive than these products, and so the use of adjuncts in relatively large amounts (20–30%) significantly reduces the cost of the brewing process. The most common adjuncts are rice, corn, wheat, or even sugars, syrups, and liquid starches, etc.

It is important to note that some legislation, especially European legislation, does not permit the addition of exogenous enzymes during the brewing process. Therefore, the use of malt is essential in order to obtain a fermentable extract, because when the exogenous enzymes cannot be used, malt is the only source of enzymatic activities, as will be shown later.

However, some special beers, such as wheat beers or sorghum beers (very common in African countries) are produced without the use of malt. In these cases, cereals are previously malted in a process similar to that for barley, or exogenous enzymes are usually added during mashing. The most recent adjuncts described are 'chufa nuts' or sedge tubers, which, after commercial enzymatic treatment, produce a wort comparable with sorghum wort, which may even have a better filterability, probably because of its higher fiber content.

In most of the cases adjuncts arrive at the brewery in a ready-to-use state, usually in the form of flour or other similar product, or may even be pretreated by different processes such as micronization, extrusion, etc. In this way, they can be used without reducing their particle size, and pretreatment gelatinizes the starch, so that their transformation into a liquid fermentable extract is easier.

Recently, a study on the use of nonmalted barley has been published, in which barley was gelatinized by cooking and thus was able to replace up to 10% of the malt, and extruded barley substituted up to 50% of the malt. However, barley treated previously with commercial enzymes did not yield satisfactory results.

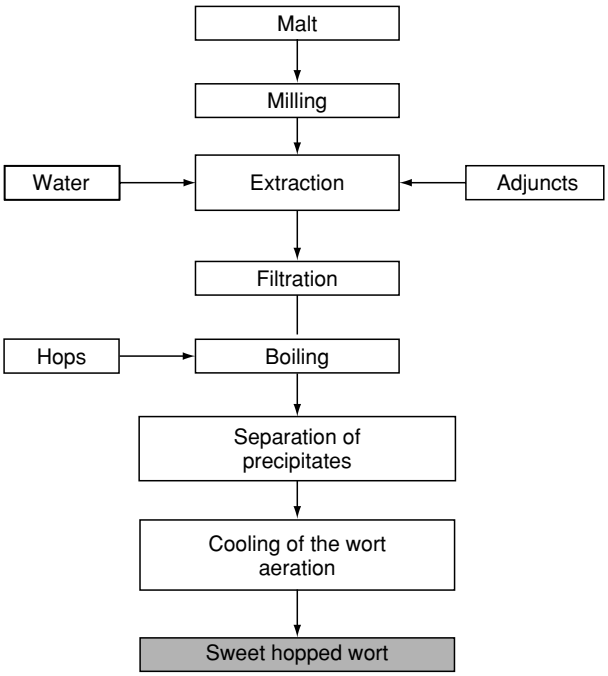

Figure 1 Process of wort production.

The process to convert malt into wort comprises several stages as can be seen in Figure 1.

Malt Milling

Malt or other malted cereals (e.g. sorghum) arrive at the brewery as a complete kernel, so the first process is malt milling, which is carried out in order to reduce the size of the particles, thus facilitating extraction and transformation of the desired compounds.

This process involves separating the endosperm from the husk and reducing the particle size to yield thick seeds, fine seeds, and mainly flour. However, it is important that the husk remains as intact as possible, so as to act as a natural filter during the filtration process.

The most commonly used milling system is the roller mill, generally consisting of three pairs of rollers. Placed between the rollers these are a series of sieves that separate the grist fraction by size. The final product of the milling process can be classified as follows:

1. Husks: These are the largest particles and consist of the outer case of the barley grain. They contribute very little to the wort.

2. Thick particles: These contain the least modified endosperm. They account for 20–25% by weight of the milled malt and yield 10–12% of the soluble extract.
3. Fine particles or flour: These contain the most modified endosperm. They account for 50–60% by weight of the milled malt and yield 89–90% of the final extract.

The milling system must be adjusted in order to improve the production of fine particles, without fractionating the husk. In this sense, malt is sometimes steam-treated to increase the moisture of the husk, which remain then more elastic and is more difficult to fragment.

The next process is mashing, in which the liquid extract is obtained.

Mashing

Once the malt has been milled, it is mixed with water to form the 'mash.' The mash is heated under controlled conditions, and then important transformations of the malt components are carried out by the action of endogenous malt enzymes.

After the hydration process, the enzyme activity that was latent in the dry malt is renewed. Enzymes work best under the following conditions: abundant water (high water activity), high-surface-area particles (allowing better access to their substrates, i.e. starch, proteins, lipids, etc.), and optimum temperatures, which are reached in subsequent stages.

A recent study has shown that during mashing, it is very important to control the oxygen availability, since some of the enzymatic systems that could be activated during this step are oxygen-dependent. Furthermore, the presence of oxygen determines the development of nonenzymatic oxidation, which can be very important in the formation of undesirable products such as aldehydes formed from unsaturated fatty acids or the formation of brown pigments.

The mashing phase is usually carried out in a single vessel, lasts between 4 and 6 h, and is often carried out in accordance with a particular temperature profile, since each type of enzyme has its optimum working temperature:

1. 35–50 °C: Dominated by the action of proteases and phosphatases. Proteins are hydrolyzed, and amino acids and other nitrogen compounds are released.
2. 50–65 °C: Dominated by the action of the β-amylases, which convert starch into maltose and limiting dextrines.
3. 65–75 °C: α-Amylase activity is dominant. Starch hydrolysis continues, and its saccharification is completed.
4. > 75 °C: All enzymes in the mixture are rendered inactive.

The exogenous enzymes can be added sequentially at each stage or at the beginning as a complex mixture that usually contents proteases and amylases.

Different breweries apply different extraction methods. The three most commonly used methods are described below.

Infusion Mashing

This is the traditional method employed in the manufacture of English 'ale.' The method uses a single vessel, and mashing is carried out at a single temperature, known as the conversion temperature, usually within 62–67 °C. This extraction method requires well-modified or broken-down malts. It is not suitable for use with additional starch sources that require prior gelatinization.

Decoction Mashing

This is the traditional method for making 'lager' beer. The equipment used for this type of process include a mash tun, a mash cooker, and a filter vat.

In this process, part of the mash is pumped into a second vessel, heated to boiling temperature, and re-added to the main mash. In mixing the two parts, the whole mash reached the different expected temperatures. This method is especially suitable for use with malts that have not been broken down to any large degree or are only slightly modified.

The traditional system uses three decoctions (the Pilsen method), in which the temperature is raised in three steps (34–40 °C, 50–65 °C, 75 °C). This system has been replaced by a double and finally a single decoction, since malts used nowadays are more modified.

This decoction procedure can also be applied to adjuncts in which the starch usually has a gelatinization temperature higher than that of the malt starch.

Double Mashing

This is the traditional method used in the USA. The adjuncts, which represent 30–50% of the total extract required and 10% of the malt or an external source of α-amylase, are heated in a mash cooker to 70 °C and maintained at this temperature for a specified period of time.

The temperature is then raised to the boiling point and maintained for 20 min under these conditions. After a while, the part that is removed and boiled off is pumped into the mash tun, which has been

kept at 35–40 °C. The temperature of the mixture is controlled by the speed of pumping.

Recent studies have focused on the use of different exogenous enzymatic systems in order to improve the quality of the wort during mashing, such as studies in the use of a new purine-nucleosidase to reduce the levels of purine nucleosides of worts and the use of proteinases to stabilize the foam.

Moreover, several studies on the thermostability of endogenous enzymes of different varieties of barley, an important factor for the malting and mashing process, are being carried out, in order to determine the best varieties for the brewing process. It is interesting to note that some breweries usually add some selected *Lactobacillus* spp. to increase the acidity of the mash and wort.

After mashing, all desired compounds are extracted and transformed into their respective byproducts, thus yielding a fermentable extract. However, this extract is cloudy, with a large amount of solids in suspension, so filtration is required to separate the liquid from the solid particles.

Filtration

In this phase, the insoluble part of the mixture, the spent grain, is separated from the liquid part, the 'sweet wort'.

Filtration is usually carried out using a lauter tun or by a mash filter (filter press). The lauter tun is a vessel with a flat, perforated bottom. First, the husk rapidly forms a filter cake on the bottom, which contains a natural filter through which the wort can be filtered. This filter yields wort of an excellent quality, but the filtration time is long, and it presents several problems in terms of removal of the spent grain. There is a special type of lauter tun, the strainmaster, which has a different bottom with slotted triangular pipes instead of a perforated bottom. In this way, the filtration area is larger, and the run-off is quicker. The quality of the wort is also good, but the filter cake retains a large amount of liquid, and discharge is not easy.

Several hollow frames and plates, separated by filter cloths, form the mash filter or filter press. In a similar way to the lauter tun, when wort crosses the filter, husk is retained on the frames forming the filter cake. The wort obtained with this type of filtration system is not as clear as that obtained using the lauter tun, and the wort also has higher levels of lipids. However, it is a more rapid process, and it is easier to automate. Usually, mash filtration is carried out to a high temperature in order to improve the filtration process and reduce the duration. This step is very important in the brewery; it could be considered one of the limiting steps. Brewers and related commercial industries have spent much time researching how to develop fast filtration systems that do not result in any losses of the wort quality. Some of the first studied systems were high-pressure mash systems, but the filter in these types of system had several technical shortcomings and did not meet the requirements. Later, a membrane mash filter was developed. This system yields wort of a good quality, and enables good separation of the solid part, even when a fine grist is used. Membranes can be inflated with compressed air to remove more of the liquid remaining in the filter cake, thus improving the extraction yield. This system can be easily automated. In the early 1990s, different breweries installed this type of filter. Another system developed at the beginning of the 1990s was the 2001 filter. This is an automatic system equipped with polypropylene plates and a frame filter, and comprising the following stages: filling, filtration, precompression, sparging, compression, and spent-grain removal. The advantages of this filter include: production of a clear wort with a low fatty acid content, reduced volume of sparging liquor, production of a dryer spent grain, reduced O_2 uptake, zero heat radiation and waste water, high efficiency (12 brews per day), simplified milling technology, flexible load (70–110% normal capacity), and flexibility with regard to raw material and mashing procedures.

The final product of this step is more or less a 'clean' liquid extract called a 'sweet wort.'

Wort Boiling and Addition of Hop

Wort boiling is carried out for different reasons, which can be summarized in terms of the main effects and secondary effects that are very important for the final characteristic of the final beer, such as stabilization, improve flavor, modify color, etc.

The main effects are: to render enzymes inactive, once starch, proteins, and other constituents have been transformed (hydrolyzed), and to sterilize the wort, converting it into a sterile medium for the subsequent fermentation.

Some of the most important secondary effects are:

- Coagulation of the remaining proteins (hot break), which is very important in preventing cloudiness in the beer;
- Precipitation of calcium phosphate, which causes a fall in pH.
- Extraction and transformation of important compounds as boiling is carried out in the presence of hop. Therefore, during this process, humulones and lupulones are extracted, and then, after heat

treatment, they are isomerized to give the typical bitter taste of beer. In addition, the isoacids formed help to stabilize the foam.
- Removal of volatile products (e.g., undesirable flavor compounds such as sulfated compounds) and some of the essential oils extracted from the hop.
- Partial caramelization of the wort sugars and the formation of Maillard reactions products, improving the flavor, and slightly changing the color. Some of these products (e.g., melanoidins) induce changes in the surface tension and also help stabilize the foam.
- Partial oxidation of phenols and their condensation with proteins. Therefore, they can be removed more easily from the medium, the beer is more stable, and cloudiness is reduced.
- Evaporation of water and wort concentration.

Boiling is traditionally carried out in boiling vessels, where hop is added as hop flowers, pellets, or even hop extracts. In the manufacture of 'ale,' the hops are usually added in the initial stages of boiling, as formation of foam can be controlled during the process. In 'lager' production, the hop is usually added manually and at different times during the boiling process, thus imparting a hop aroma. Different types of vessel are used in the wort boiling process. The most common types are as follows.

Traditional Boiling Vessels

Open-top vessels, in which the process takes places at atmospheric pressure. Traditionally, boiling vessels were made of copper and had an external heat source. Modern vessels have an internal heat source and are made of stainless steel, which is a low heat conductor, thus minimizing heat loss. However, some authors claim that the traditional copper vessel seems to impart a special flavor characteristic to the wort.

In traditional systems, the boiling process lasts between 1 and 2 h. In order to reduce the energy needs for this process, several heat systems have been developed. The use of an external heat exchanger, or calandria, requires a shorter boiling time, because, in this system, the wort circulates through the heat exchanger and reaches higher temperatures (106–110 °C).

Other systems have focused on recycling the vapor formed. Mechanical vapor-compression systems have been used as external boiling systems.

Continuous High-temperature Systems

Focusing on energy reduction, these systems consist of several consecutive heat exchangers, in which the wort temperature is increased to approximately 140 °C and held for 3–5 min. The pressure is then reduced using two expansion vessels, and the vapors are finally recycled to heat the heat exchanger. In this way, the brewing process time is reduced, and less energy is consumed, so the final costs can be reduced considerably. Several different systems, with diverse designs, have been patented in recent years. Some authors have indicated that there are remarkable differences between the sensorial characteristics of the worts produced by both methods.

The final product from the boiling step is called 'sweet hopped wort,' which must be filtered before fermentation in order to eliminate all the precipitates formed during boiling (hot precipitates).

Separation of Hot Precipitates (Filtration)

The first point to note is that after the boiling process, all operations must be carried out under sterile conditions in order to avoid any recontamination of the wort.

Hot precipitates, together with hop flowers if they have been used, must be removed from the medium. In these cases, the hop flowers form a filter bed or cake through which the wort is filtered. When pellets or extracts are used, wort is usually clarified by the addition of a clarifying agent, such as bentonite or, even better, silica gel, before it is pumped into the filtration systems. The types of equipment employed in this filtration process include the so-called windmill, Whirlpool, centrifuge, or hot settling tank.

Different modifications of the filter systems have been developed in recent years, with the main aim to improve the filtration process by reducing the operation time and increasing the cleanness of the sweet hopped wort.

After hot filtration, the clean wort must be cooled in order to reduce the temperature and to allow the addition of yeast, and must be aired to promote the growth of the yeast, so that the wort will be in the best condition for the next step: fermentation.

Cooling and Aeration of the Sweet Hopped Wort

There are several systems of cooling and aeration used in beer manufacture. These include:

1. *Flat recipients*: In this traditional system, the wort was loaded on to the recipients to a height of 25 cm. Cooling was achieved by evaporation of the water through ventilation holes through which air flowing over the wort could enter. These containers also allowed aeration of the wort and decanting of hot and cold precipitates. The air used would be

passed through aseptic filters that were free of undesirable microorganisms.

2. *Refrigeration by means of plate or tubular heat exchangers*: The cooling agent used is usually cold water, or refrigerated glycolate, alcohol or brine. Air is introduced, in the form of sterile air or oxygen, usually between the two sections of the plate heat exchanger, when the wort enters the final stage of cooling. The fall in temperature leads to the formation of new precipitates, which are eliminated by sedimentation, although some breweries do this by filtration or centrifugation of the wort, or even by flotation separation, which is more effective than sedimentation when the particle size is very small. Furthermore, studies carried out in 2000, showed that flotation techniques minimize wort losses. Cross-flow, micro- and ultrafiltration have also been described as interesting techniques by which to eliminate the wort cold break precipitates.

The aeration is usually controlled by specific mechanisms based on concurrent flow of the liquid and gas, and on the residence time of the liquid in the treatment unit.

The final product is a clean, cold, and aired sweet hopped wort, ready to be transformed into beer after its fermentation by the action of selected yeast.

Special Wort

Several patents have been developed in recent years for the production of 'special wort,' such as 'concentrate wort,' which can be used as a raw material in small breweries or in home installations common in some countries such as the UK.

In addition, special wort with lower levels of fermentable sugars is used in the brewing process for low-alcohol beers. Limiting the action of amylases, especially β-amylase, reduces the production of fermentable sugars. This can be achieved with a high-temperature mashing process, in which β-amylase is thermal destroyed by heat. This wort has a high level of dextrin. Low-gravity wort is also employed in low-alcohol beer manufacture. This type of wort is obtained by low-gravity brewing or, more usually, by diluting normal wort.

See also: **Beers**: History and Types; Raw Materials; Chemistry of Brewing; Biochemistry of Fermentation; **Yeasts**

Further Reading

Freeman GJ and McKenie MT (1995) Filtration and Stabilization. In: Lea AGH and Piggot JR (eds) *Fermented Beverage Production*, pp. 325–358. London: Blackie Academic & Professional.

Hornsey IS (1999) *Brewing*. Cambridge: Royal Society of Chemistry.

Hough JS (1991) *The Biotechnology of Malting and Brewing*. Cambridge: Cambridge University Press.

Iserentant D (1995) Beers: recent technological innovations in brewing. In: Lea AGH and Piggot JR (eds) *Fermented Beverage Production*, pp. 45–61. London: Blackie Academic & Professional.

Kender RMJ (1990) Hop products and their usage – a review. *Ferment* 3: 38–41.

Lewis MJ and Young TW (1995) *Brewing*. London: Chapman & Hall.

McElroy D and Jacobsen J (1995) What's brewing in barley biotechnology? *Bio/Technology* 13: 245–249.

Biochemistry of Fermentation

G G Stewart, Heriot-Watt University, Riccarton, Edinburgh, UK

Introduction

The characteristic flavor and aroma of any beer are, in large part, determined by the yeast strain and the fermentation conditions. Thus, proprietary strains belonging to individual brewing companies are usually (but not always) jealously guarded and conserved. In Germany, most of the beer is produced with only four lager strains, and approximately 65% of the beer is produced with one strain.

The Fermentation Process

The genus *Saccharomyces* has many species that are generally regarded as safe and produce the two important primary metabolites – ethanol and carbon dioxide (CO_2). Lager and ale, the two main types of beer, are fermented with strains of *Saccharomyces uvarum (carlsbergensis)* and *Saccharomyces cerevisiae*, respectively. The scientific literature increasingly refers to them as *Saccharomyces cerevisiae* (ale type) and *Saccharomyces cerevisiae* (lager type).

Traditionally, lager is produced by 'bottom-fermenting yeasts' at 7–15 °C, which, at the end of primary fermentation, flocculate and collect on the bottom of the fermenter. 'Top-fermenting yeasts,' used for the production of ale, ferment at temperatures between 18 and 22 °C. At the end of fermentation, the culture forms into loose clumps of cells that are adsorbed on to CO_2 bubbles and are carried to the surface of the wort. Consequently, top yeasts are

collected (skimmed) for reuse from the surface of the fermenting wort, whereas bottom yeasts are collected (cropped) from the bottom of the fermenter. The difference between lager and ales on the basis of bottom and top cropping has become less distinct with the advent of cylindroconical fermenters, where the yeast sediments to the base of the vessel, and centrifuges, where the yeast remains in suspension throughout the fermentation.

There is a plethora of literature describing the genetics and biochemistry of *Saccharomyces cerevisiae* laboratory strains, but there is a lack of knowledge regarding the genetics and biochemistry of industrial *Saccharomyces* strains. The haploid strain that the molecular biologist employs in the university laboratory as the organism of choice is usually totally unsuitable for use in breweries. Brewing yeasts and many other industrial yeasts have been selected over time for those characteristics that render them unamenable to easy genetic manipulation in the laboratory. They are usually polyploid or aneuploid, lack a mating-type characteristic, and sporulate poorly, if at all, and the spores that do form are usually not in fours, exhibiting a poor spore viability and rendering tetrad analysis difficult.

The objectives of wort fermentation are to consistently metabolise wort constituents into ethanol and other fermentation products in order to produce beer with a satisfactory quality and stability. Another objective is to produce yeast crops that can be confidently repitched into subsequent brews. During the brewing process, overall yeast performance is controlled by a plethora of factors. These factors include:

- the yeast strains employed and their condition at pitching and throughout fermentation;
- the concentration and category of assimilable nitrogen;
- the concentration of ions;
- the fermentation temperature;
- the pitching (inoculation) rate;
- the tolerance of yeast cells to stress factors such as osmotic pressure and ethanol;
- the wort gravity;
- the oxygen level at pitching;
- the wort sugar spectrum; and
- yeast flocculation characteristics.

These factors influence yeast performance either individually or in combination with others and also together permit the definition of the requirements of an acceptable brewer's yeast strain.

> In order to achieve a beer of high quality, it is axiomatic that not only must the yeast be effective in removing the required nutrients from the growth/fermentation medium (wort), able to tolerate the prevailing environmental conditions (for example, ethanol tolerance) and impart the desired flavour to the beer, but the microorganisms themselves must be effectively removed from the wort by flocculation, centrifugation and/or filtration after they have fulfilled their metabolic role.

It is worth noting that brewing is the only major alcoholic beverage process that recycles its yeast. It is therefore important to jealously protect the quality of the cropped yeast because it will be used to pitch a later fermentation and therefore will have a profound effect on the quality of the beer resulting from it.

Over the years, considerable effort has been devoted in many research laboratories to the study of the biochemistry and genetics of brewer's yeast (and industrial yeast strains in general). The objectives of these studies have been twofold:

- to learn more about the biochemical and genetic makeup of brewing yeast strains; and
- to improve the overall performance of such strains, with particular emphasis being placed on broader substrate utilization capabilities, increased ethanol production, and improved tolerance to environmental conditions such as temperature, high osmotic pressure and ethanol, and finally, to understand the mechanism(s) of flocculation.

When yeast is pitched into wort, it is introduced into an extremely complex environment due to the fact that wort is a medium consisting of simple sugars, dextrins, amino acids, peptides, proteins, vitamins, ions, nucleic acids, and other constituents too numerous to mention. One of the major advances in brewing science during the past 30 years or so has been the elucidation of the mechanisms by which the yeast cell, under normal circumstances, utilizes in a very orderly manner the plethora of wort nutrients.

Pathway and Regulation of Fermentation

Wort contains the sugars sucrose, fructose, glucose, maltose, and maltotriose, together with dextrin material. In the normal situation, brewing yeast strains (ale and lager strains) are capable of utilizing sucrose, glucose, fructose, maltose, and maltotriose in this approximate sequence (or priority), although some degree of overlap does occur. The majority of brewing strains leave the maltotetraose and other dextrins unfermented, but *Saccharomyces diastaticus* is able to utilize dextrin material as a result of the secretion of the extracellular enzyme glucoamylase. The initial step in the utilization of any sugar by yeast is usually either its passage intact across the cell membrane or its hydrolysis outside the cell membrane, followed by

entry into the cell by some or all of the hydrolysis products. Maltose and maltotriose are examples of sugars that pass intact across the cell membrane, whereas sucrose (and dextrin with *Saccharomyces diastaticus*) is hydrolyzed by an extracellular enzyme, and the hydrolysis products are taken up into the cell. Maltose and maltotriose are the major sugars in brewer's wort, and as a consequence, a brewer's yeast's ability to use these two sugars is vital and depends upon the correct genetic complement. Brewers' yeast may possess independent uptake mechanisms (maltose and maltotriose permease) to transport the two sugars across the cell membrane into the cell. Once inside the cell, both sugars are hydrolyzed to glucose units by the α-glucosidase system. It is important to reemphasize that the transport, hydrolysis, and fermentation of maltose are particularly important in brewing, since maltose usually accounts for 50–60% of the fermentable sugar in wort. Maltose fermentation in *Saccharomyces* yeasts requires at least one of five unlinked (each independent) MAL (maltose) loci each consisting of three genes encoding the structural gene for α-glucosidase (maltase) (MAL S), maltose permease (MAL T) and an activator (MAL R) whose product coordinately regulates the expression of the α-glucosidase and permease genes. The expression of MAL S and MAL T is regulated by maltose induction and glucose repression. When glucose concentrations are high (greater than 1% (w/v)), the MAL genes are repressed, and only when 40–50% of the glucose has been taken up by yeast from the wort will the uptake of maltose and maltotriose commence. Thus, the presence of glucose in the fermenting wort exerts a major repressing influence on wort fermentation rate. Using the glucose analog 2-deoxy-glucose (2-DOG), which is not metabolized by *Saccharomyces* strains, spontaneous variants of brewing strains have been selected in which the maltose uptake is not repressed by glucose, and as a consequence, these variants (called derepressed) have increased wort fermentation rates. Once the sugars are inside the cell, they are converted via the glycolytic pathway into pyruvate. To date, only derepressed ale strains have been isolated. Repressed lager strains have not been found.

Active yeast growth involves the uptake of nitrogen, mainly in the form of amino acids, for the synthesis of proteins and other nitrogenous compounds of the cell. Later in the fermentation, as yeast multiplication stops, nitrogen uptake slows down or ceases. In wort, the main nitrogen source for synthesis of proteins, nucleic acids and other nitrogenous cell components is the variety of amino acids formed from the proteolysis of barley proteins. Brewer's wort contains 19 amino acids, and, as with wort sugars, the assimilation of amino acids is ordered. Four groups of amino acids have been identified on the basis of assimilation patterns (Table 1). Those in group A are utilized immediately following yeast pitching, whereas those in group B are assimilated more slowly. Utilization of group C amino acids commences when group A types are fully assimilated. Proline, the most plentiful amino acid in wort and the sole group D amino acid, is utilized poorly or not at all. Proline is usually still present in beer at 200–300 $mg\,l^{-1}$. However, under aerobic conditions, proline is assimilated after exhaustion of the other amino acids, since its uptake requires the presence of mitrochondrial oxidase.

The regulation of amino acid uptake by brewers' and related yeast strains is complex, involving carriers specific to certain amino acids and a general amino acid permease of broad substrate specificity. The utilization pattern of wort nitrogen is due to a combination of the range of permeases present, their specificity, and feedback inhibition effects resulting from the composition of the yeast intracellular amino acids.

The metabolism of assimilated amino nitrogen is dependent on the phase of the fermentation and on the total quantity provided in the wort. The majority of amino nitrogen is ultimately utilized in protein synthesis and, as such, is vital for yeast growth. It would appear that amino acids are not usually incorporated directly into proteins but are involved in transamination reactions, a significant proportion of the amino acid skeletons of yeast protein being derived via the catabolism of wort sugars. This explains why the total amino content of wort is important in determining the extent of yeast growth, the amino acid spectrum being somewhat secondary. However, the amino acid spectrum of wort does influence beer flavor.

The yeast assimilates the wort amino acids, a transaminase system removes the amino group, and the

Table 1 Classification of amino acids according to their spread of absorption from wort by a brewing yeast strain

(A) Fast absorption	*(B) Intermediate absorption*	*(C) Slow absorption*	*(D) Little or no absorption*
Glutamic acid	Valine	Glycine	Proline
Aspartic acid	Methionine	Phenylalanine	
Asparagine	Leucine	Tyrosine	
Glutamine	Isoleucine	Tryptophan	
Serine	Histidine	Alanine	
Threonine		Ammonia	
Lysine			
Arginine			

carbon skeleton is anabolized, creating an intracellular oxo-acid pool. The oxo-acid pool generated by the transaminases and anabolic reactions is a precursor of aldehydes and higher alcohols that contribute to beer flavor. Thus, the formation of higher alcohols (i.e., higher in number of carbon atoms than ethanol) is tied in with nitrogen metabolism.

The main nitrogen composition of wort has far-reaching effects on fermentation performance and on beer flavor. Where malt is used as the principal source of extract, the quantity and composition of amino acids are such that these problems are not encountered. However, care must be exercised when using adjuncts, many of which are relatively deficient in amino nitrogen.

Wort fermentation in beer production is largely anaerobic, but when the yeast is first pitched into wort, some oxygen must be made available to the yeast. Indeed, it is now evident that this is the only point in the brewing process where oxygen is beneficial. Oxygen must be excluded as far as possible from all other parts of the process, because it will have a negative effect on beer quality. Specifically, it will promote beer flavor instability. The widespread adoption of high-gravity brewing procedures has increased our awareness of the importance of oxygen during wort fermentation and has stimulated basic and applied research on the mechanisms of oxygen interactions during cell growth and the application of this knowledge in the process.

Oxygen has a profound influence on the activity of yeasts and particularly on yeast growth. Certain yeast enzymes only react with oxygen, and this cannot be replaced by other hydrogen acceptors. This applies to the oxygenases involved in the synthesis of unsaturated fatty acids and sterols, which are vital components of cell membranes. Quantitative studies on the effect of aeration on yeast growth and fermentation have been given little serious consideration until the last 25 years. The traditional concept of beer fermentation was that growth occurred prior to the fermentation of most wort sugars and that fermentation was carried out by nongrowing, stationary phase cells. It is now known that yeast growth, sugar utilization, and ethanol production are coupled phenomena. For example, the rate of fermentation by growing, exponential phase cells of an ale yeast strain is 33-fold higher than that of nongrowing cells.

For a brewery fermentation to proceed rapidly, there must be sufficient amounts of yeast synthesized. Inadequate growth of a brewer's yeast culture will result in poor attenuation, altered beer flavor, inconsistent fermentation times, and recovered pitching yeasts, which are undesirable for subsequent fermentations. Trace amounts of oxygen have profound stimulatory effects on yeast fermentation and particularly on yeast growth. Pasteur demonstrated that oxygen was necessary for normal yeast reproduction, although excessive wort aeration caused undesirable flavor effects on the finished beer. Oxygen requirements were confirmed by such early notable brewing researchers as Adrian Brown, Horace Brown, and Frans Windisch. Windisch concluded that overvigorous aeration of fermenting worts led to yeast 'weakness,' illustrated by increasingly sluggish fermentations characterized by longer lag phases, a slower specific rate of fermentation, and/or residual sugar remaining in the final beer. The critical importance of oxygen was confirmed when, in 1954, it was shown that, under anaerobic conditions, *Saccharomyces* yeast strains require both preformed sterols and unsaturated fatty acids as growth factors. These two lipids are both found in membranes and are critical for membrane function and integrity. Both of these lipid classes require molecular oxygen for their biosynthesis.

Lipids in beer quantitatively form an almost negligible component, but can influence a beer's organoleptic and physiochemical properties. Malt is the main source of unsaturated fatty acids in wort. Wort concentrations of these acids are suboptimal and can be growth-limiting. During fermentation, yeast can take up free fatty acids from wort, most of which are incorporated as structural lipids.

Yeast cultures synthesize fatty acids throughout fermentation, but the ratio of the acids varies with time. Unsaturated fatty acid (for example, palmitoleic (C16:1) and oleic (C18:1) acids) synthesis only occurs in the presence of dissolved oxygen. Oxygen is present in aerated/oxygenated pitched wort for a relatively short period (3–9 h), and during this period, there is a large increase in the percentage of unsaturated fatty acids. When oxygen is depleted, there is an increase in the production of short-chain fatty acids (C6–C12).

The sterol component of brewing yeast ranges from 0.05 to 0.45% of the cellular dry weight (depending on the prevailing environmental conditions) and accounts for less than 10% of the total cell lipid. Ergosterol is the major sterol in brewing-yeast strains and can account for over 90% of the total sterol. The biosynthetic pathway for sterol formation is complex. The important fact for this chapter is that the precursor sequences can be synthesized anaerobically, but the final reaction that produces ergosterol requires molecular oxygen. The major function of sterols in yeast is to contribute to the structure and dynamic state of the membranes. The primary role is to modulate membrane fluidity under fluctuating environmental conditions. For example, ergosterol confers

increased resistance to ethanol and multiple freeze–thaw effects. A decrease in the ergosterol level of membranes has been directly related to a reduction in cell viability in the presence of ethanol.

Pitching yeasts are propagated under weakly aerated conditions or recovered from previous fermentations. In both cases, the cells are lipid-depleted, and to promote normal growth and attenuation, either preformed lipids must be added to the wort or oxygen must be made available for their synthesis. In commercial brewing, only the second alternative is feasible. Wort is cooled and aerated/oxygenated to 8–16 $mg\,l^{-1}$ dissolved oxygen. Within a few hours of pitching, most of this oxygen is removed from the wort. During this time, there is intensive synthesis of lipid (sterol and fatty acid) and a decrease in cellular glycogen. In practice, sterol synthesis by brewing yeasts in the presence of oxygen appears to be of greater significance than unsaturated fatty acid synthesis. This may be due to the contribution of wort to the fatty acid pool. Wort does not contribute exogenous sterol to the fermentation.

Although ethanol is the major excretion product synthesized by yeast during wort fermentation, this primary alcohol has little impact on the flavor of the final beer. It is the type and concentration of the many other yeast excretion products formed during wort fermentation that primarily determine the flavor of the beer. The formation of these excretion products depends on the overall metabolic balance of the yeast culture, and there are many factors that can alter this balance and, consequently, beer flavor. Yeast strain, fermentation temperature, adjunct type and level, fermenter design, wort pH, buffering capacity, wort gravity, etc., are all influencing factors.

Flavor Formation

Some volatiles are of great importance and contribute significantly to beer flavor, whereas others are important in building background flavor. The following groups of substances are found in beer: organic and fatty acids, alcohols, esters, carbonyls, sulfur compounds, amines, phenols, and a number of miscellaneous compounds. In flavor terms, the higher alcohols (also called fusel oils) that occur in beer and many spirits are: *n*-propanol, isobutanol, 2-methyl-1-butanol, and 3-methyl-1-butanol. However, more than 40 other alcohols have been identified. Regulation of the biosynthesis of higher alcohols is complex, since they may be produced as byproducts of amino acid catabolism or via pyruvate derived from carbohydrate metabolism.

Esters are important flavor components that impart flowery and fruit-like flavors and aromas to beers, wines, and spirits. Their presence is desirable at appropriate organoleptic concentrations, but failure to properly control fermentation can result in unacceptable beer ester levels. Organolepiticaly important esters include ethyl acetate, isoamyl acetate, isobutyl acetate, ethyl caproate, and 2-phenylethyl acetate. In total, over 90 distinct esters have been detected in beer.

Some 200 carbonyl compounds are reported to contribute to the flavor of beer and other alcoholic beverages. Those influencing beer flavor, produced as a result of yeast metabolism during fermentation, are various aldehydes and vicinal diketones, notably diacetyl. Also, carbonyl compounds exert a significant influence on the flavor stability of beer. Excessive concentrations of carbonyl compounds are known to cause stale flavor in beer. The effects of aldehydes on flavor stability are reported as grassy notes (propanol, 2-methyl butanol pentanol) and a papery taste (*trans*-2-nonenal, furfural).

Quantitatively, acetaldehyde is the most important aldehyde. This is produced via the decarboxylation of pyruvate and is an intermediate in the formation of ethanol. It may be present in beer at concentrations above its flavor threshold (approx. 10 $mg\,l^{-1}$), at which it imparts an undesirable 'grassy' or 'green apple' character. Acetaldehyde accumulates during the period of active growth. Levels usually decline in the stationary phases of growth late in fermentation. As with higher alcohols and esters, the extent of acetaldehyde accumulation is determined by the yeast strain and the fermentation conditions. Although the yeast strain is of primary importance, elevated wort oxygen concentration, pitching rate, and temperature all favor acetaldehyde accumulation. In addition, the premature separation of yeast from fermented wort does not allow the reutilization of excreted acetaldehyde associated with the later stages of fermentation.

Other important flavor-active carbonyls, whose presence in beer is determined in the fermentation stage, are the vicinal diketones, diacetyl (2,3-butanedione) and 2,3-pentanedione. Both compounds impart a 'butterscotch' flavor and aroma to beer. Quantitatively, diacetyl is the most important, since its flavor threshold is approx. 0.1 $mg\,l^{-1}$ and is 10-fold lower than that of 2,3-pentanedione. The organoleptic properties of vicinal diketones contribute to the overall palate and aroma of some ales, but in most lagers, they impart an undesirable character. A critical aspect of the management of larger fermentations and subsequent maturation is to ensure that the mature beer contains concentrations of vicinal diketones lower than their flavor threshold.

Diacetyl and 2,3-pentanediones arise in beer as byproducts of the pathways leading to the formation of valine and isoleucine. The α-acetohydroxy acids, which are intermediates in these biosyntheses, are in part excreted into the fermenting wort. Here, they undergo spontaneous oxidative decarboxylation, giving rise to vicinal diketones. Further metabolism is dependent on yeast dehydrogenases. Diacetyl is reduced to acetoin and, ultimately, 2,3-butanediol, and 2,3-pentanedione to its corresponding diol. The flavor threshold concentrations of these diols are relatively high, and therefore the final reductive stages of vicinal diketone metabolism are critical in order to obtain a beer with acceptable organoleptic properties.

The diacetyl concentration peak occurs towards the end of the period of active growth. The reduction of diacetyl takes place in the later stages of fermentation when active growth has ceased. In terms of practical fermentation management, the need to achieve a desired diacetyl specification may be the factor that determines when the beer may be moved to the conditioning phase, filtered or centrifuged (depending on the processing procedures). Thus, diacetyl metabolism is an important determinant of overall vessel residence time, which clearly affects the efficiency of plant utilization.

Sulfur compounds make a significant contribution to the flavor of beer. Although small amounts of sulfur compounds can be acceptable or even desirable in beer, in excess, they give rise to unpleasant off-flavors, and special measures such as purging with CO_2 or prolonged maturation times are necessary to remove them. Many of the sulfur compounds present in beer are not directly associated with fermentation but are derived from the raw materials employed. However, the concentrations of hydrogen sulfide (rotten egg aroma) and sulfur dioxide (burnt match aroma) are dependent on yeast activity. Failure to manage fermentation properly can result in unacceptably high levels of these compounds occurring in the finished beer.

The concentration of hydrogen sulfide and sulfur dioxide formed during fermentation are primarily determined by the yeast strain used, although the wort composition and the fermentation conditions are major factors, particularly where levels are abnormally high. Both compounds arise as byproducts of the synthesis of the sulfur-containing amino acids cysteine and methionine from sulfate. Their synthesis is influenced by wort composition in that the yeast will preferentially assimilate sulfur-containing amino acids. It is only when wort is depleted in such amino acids that the biosynthetic route comes into operation.

Flocculation

The flocculation property or, conversely, lack of flocculation of a particular yeast culture is one of the major factors when considering important characteristics during brewing and other ethanol fermentations. Unfortunately, some degree of confusion has arisen by the use of the term flocculation in the scientific literature to describe different phenomena in yeast cell behavior. Specifically, flocculation, as it applies to brewer's yeast is ' the phenomenon wherein yeast cells adhere in clumps and either sediment from the medium in which they are suspended or rise to the medium's surface.' This definition excludes other forms of aggregation, particularly that of 'clumpy-growth' and 'chain formation.' This nonsegregation of daughter and mother cells during growth has sometimes erroneously been referred to as flocculation. The term 'nonflocculation' therefore applies to the lack of cell aggregation and, consequently, a much slower separation of (dispersed) yeast cells from the liquid medium. Flocculation usually occurs in the absence of cell division, but not always, during late logarithmic and stationery growth phase and only under rather circumscribed environmental conditions involving specific yeast cell surface components (proteins and carbohydrate components) and an interaction of calcium ions. Although yeast separation often occurs by sedimentation, it may also be by flotation because of cell aggregates entrapping bubbles of CO_2 as in the case of 'top-cropping' ale brewing yeast strains.

Individual strains of brewer's yeast differ considerably in flocculating power. At one extreme, there are highly nonflocculent, often referred to as powdery, strains. At the other extreme, there are flocculent strains. The latter tend to separate early from suspension in fermenting wort, giving an underattenuated, sweeter, and less fully fermented beer. Beers of this nature, because of the presence of fermentable sugars, are liable to biological instability. By contrast, poorly flocculent (nonflocculent or powdery) yeasts produce a dry, fully fermented, more biologically stable beer in which clarification is slow, leading to filtration difficulties and the possible acquisition of yeasty off-flavors. The disadvantages presented by the two types of yeast strain are especially relevant to more traditional fermentation systems where the fermentation process is dependent upon the sedimentation characteristics of the yeast.

Contemporary brewing technology has largely reversed this situation where yeast sedimentation characteristics are now fitted into the fermenter design. The efficiency, economy, and speed of batch fermentations have been improved by the use of

cylindroconical fermentation vessels and centrifuges (which are often, but not always, employed in tandem). There is no doubt that differences in the flocculation characteristics of various yeast cultures are primarily a manifestation of the culture's cell wall structure. Several mechanisms for flocculation have been proposed. One hypothesis is that anionic groups of cell wall components are linked by Ca^{2+} ions. In all likelihood, these anionic groups are proteins. Another hypothesis implicates mannoproteins specific to flocculent cultures acting in a lectin-like manner to cross-link cells; here, Ca^{2+} ions act as ligands to promote flocculence by conformational changes. Most people working in the field agree that the latter hypothesis is the most credible. In addition to flocculation, there is the phenomenon of coflocculation. Coflocculation is defined as the phenomenon where two strains are nonflocculent alone, but flocculent when mixed together. To date, coflocculation has only been observed with ale strains, and there are no reports of coflocculation between two lager strains of yeast. There is a third flocculation reaction that has been described where the yeast strain has the ability to aggregate and cosediment with contaminating bacteria in the culture. Again, this phenomenon appears to be confined to ale yeast strains, and cosedimentation of lager yeast with bacteria has not been observed.

As described above, flocculation requires the presence of surface protein and mannan receptors. If these are not available or are masked, blocked, inhibited, or denatured, flocculation cannot occur. The onset of flocculation is an aspect of the subject where there is great commercial interest but about which relatively little is known. As previously discussed, the ideal brewing strain remains in suspension as fermenting single cells until the end of fermentation when the sugars in the wort are depleted, and only then does it rapidly flocculate out of suspension. What signals the onset of activation or relief from inhibition? This is still an unanswered question that is currently being studied by a number of research laboratories.

Yeast flocculation is genetically controlled, and research on this aspect of the phenomenon dates from the early 1950s. However, because of the polyploid/aneuploid nature of brewing yeast strains, most, but not all, of the research on flocculation genetics has been conducted on haploid/diploid genetically defined laboratory strains. Numerous genes have been reported to directly influence the flocculent phenotype in *Saccharomyces* spp. Four dominant flocculation genes have been identified: *FLO1* (whose alleles are *FLO2*, *FLO4*, *FLO8*), *FLO5*, *FLO9*, and *FLO10*, as well as semidominant gene, *flo3*, and two recessive genes, *flo6* and *flo7*. In addition, mutations in several genes, including the regulatory genes, *TUP1* and *SSN6*, have been found to cause flocculation or 'flaky' growth in nonflocculent strains. In total, at least 33 genes have been reported to be involved in flocculation or cell aggregation. Although the role of many of these genes is far from understood, *FLO1* and other FLO genes have been successfully cloned into brewing strains and the flocculation phenotype expressed.

See also: **Alcohol**: Properties and Determination; **Beers**: History and Types; Raw Materials; Chemistry of Brewing; **Yeasts**

Further Reading

Reed G and Nagodawithana TW (1991) *Yeast Technology.*New York: Van Nostrand Reinhold.

Smart K and Boulton C (2000) *Brewing Yeast Fermentation Performance*. New York: Blackwell Science Inc.

Stewart G G and Russell L (1986) One hundred years of yeast research and development in the brewing industry. *Journal of the Institute of Brewing* 92: 537–558.

Chemistry of Brewing

C W Bamforth, University of California, Davis, CA, USA

Introduction

The word 'brewing' is used in at least two contexts in beer making. Strictly speaking, the most accurate usage of the term is to describe the process by which is produced the feedstock, wort, that will be fed to the yeast. Thus it is customary for those within the industry to talk of 'brewing and fermentation.' Frequently, however, the word brewing is used to describe the entirety of the operation by which malted barley (and other sources of sugar) and hops are converted to beer. For the purposes of a description of the chemistry of brewing I am taking the latter definition, but we must not ignore the prior process of malting (the controlled germination of barley) because the conversion of barley into wort involves this stage also. It is very difficult to divorce a discussion of the mashing stage of brewing from malting, for in reality they are but successive stages in the enzymic conversion of barley into wort. **Figure 1** summarizes the structure of barley and the distribution of the key polymers within the starchy endosperm, whilst **Table 1** summarizes the key process stages of malting and brewing. I will approach this somewhat complex

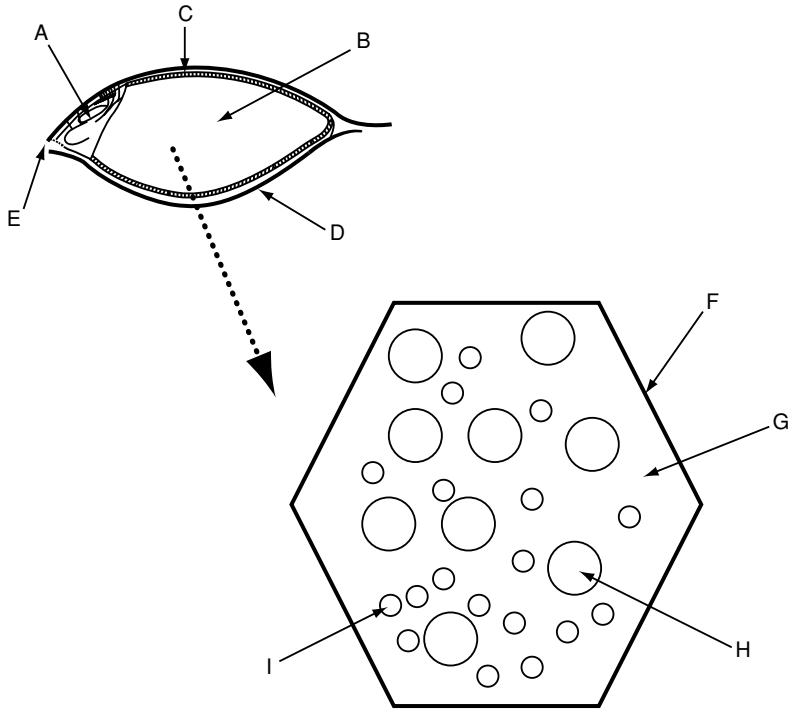

Figure 1 The structure of barley and of one of the cells of the starchy endosperm. A, embryo; B, starchy endosperm; C, aleurone; D, hull; E, micropyle; F, cell wall (75% β-glucan; 20% arabinoxylan; 5% protein; traces ferulic and acetic acids); G, protein; H, large starch granule; I, small starch granule.

issue by focusing on the chemistry of the finished product and highlighting how the balance of components in beer is determined by the various stages occurring in the maltings and brewery.

The Chemistry of Beer

The properties of beer result from the presence of a diversity of chemical compounds, some large (macromolecules) and some small. For the majority of beers these materials are derived in their entirety from the raw materials or are generated through the metabolism of brewing yeast, *Saccharomyces cerevisiae*. Most brewers are reluctant to introduce additives at any stage of the process and will also place stringent specifications on the water, grist, and hop materials in respect of what may or may not be used in the processes of their suppliers. Some brewers will use propylene glycol alginate to protect foam on the product from the deleterious impact of lipophilic materials entering the beer at point of sale, for example, fats from food or detergent in inadequately rinsed glasses. Sulfur dioxide or ascorbic acid is sometimes (but increasingly sparingly) employed to protect beer from staling. (*See* **Yeasts**.)

Ethanol

The equation at the heart of the production of all alcoholic beverages is:

$$C_6H_{12}O_6 \xrightarrow{\text{Yeast}} 2C_2H_5OH + 2CO_2$$

In the case of the majority of beers the sugar is derived during mashing of malted barley. In fact, most of the sugar derived in conventional mashing operations is maltose, with lesser quantities of glucose, maltotriose, sucrose, fructose, and dextrins.

The majority of beers worldwide contain between 3.5 and 5.5% alcohol by volume (ABV).

Ethanol impacts on the quality of beer in several ways. It contributes directly to flavor, with a recognizable warming characteristic absent from nonalcoholic or low-alcohol products. It also affects the contribution to aroma of other volatiles by influencing their partitioning between the beer and its head space.

Ethanol lowers surface tension, thereby promoting bubble formation. Conversely, it competes with other surface-active molecules for sites in the bubble wall, leading to a lessening of foam stability. (*See* **Alcohol**: Properties and Determination.)

Table 1 The key events in malting and brewing

Process stage	*Treatments*	*Events*
Steeping of barley	Water added, separated by air rests, to raise water content of embryo and endosperm (target in whole grain is usually 44–46%); takes up to 48 h at 14–18 °C	Water enters through micropyle. Distribution through embryo, aleurone, and starchy endosperm critical to enable modification. Synthesis of hormones (including gibberellins) by embryo, hydration of substrate (starchy endosperm)
Germination of barley	Controlled sprouting (modification) of grain – typically 4–5 days at 16–20 °C	Synthesis of enzymes by aleurone (triggered by hormones) and migration into starchy endosperm; sequential degradation of cell walls, some protein, small starch granules, and pitting of large granules
Kilning	Heating of grain via increasing temperature regime (50–220 °C) for desired properties: enzyme survival, removal of moisture for stabilization, removal of raw flavors, development of malty flavors and color	Enzyme survival greater with low-temperature start to kilning and lower final curing temperature. Increased heating of malts of increased modification (i.e., higher sugar and amino acid levels) gives increasingly complex flavors and colors via Maillard reactions
Malt storage	3–4 weeks' ambient storage, otherwise wort separation problems occur later	Unknown
Mashing	Extraction of milled malt at temperatures between 40 and 75 °C; typically < 1–2 h. Separation of wort from spent grains by lautering or mash filtration with sparging to recover extract fully. Spent grains go to cattle feed	Enzymolysis continued, especially of starch after gelatinization at > 62 °C
Use of adjuncts	Solid adjuncts used in brewhouse, taking advantage of malt enzymes; rice and corn need 'cooking' at up to 100 °C, < 1 h (liquid sugars are products of acid and enzyme action in sugar factory and added at boiling stage)	Cereals with higher starch gelatinization temperatures than for barley need precooking before combining with main mash
Boiling	1–2 h at 100 °C, before clarification (< 1 h) and cooling	To sterilize, extract hops, concentrate, and kill all residual enzymes. Precipitation of insoluble materials as trub, enhancing final product stability
Fermentation	Wort pitched with yeast and fermented for 3–14 days at 6–25 °C	Fermentation of glucose, maltose, sucrose, maltotriose to alcohol; enzymic production of various flavorsome compounds (alcohols, esters, fatty acids, sulfur-containing compounds, etc.) Synthesis and removal of diacetyl as an offshoot of amino acid production
Cold conditioning and filtration	− 1 °C for ≥ 3 days; possibly stabilization (silica hydrogels, tannic acid or papain to remove protein; PVPP to bind polyphenol); then filtration	Precipitation, sedimentation, and removal of solids

PVPP, polyvinylpolypyrrolidone.

Carbon Dioxide

A typical brewery fermentation will produce some 150 g CO_2 per hectoliter for every degree Plato (1 degree Plato is basically equivalent to a 1% sugar solution). Most of this CO_2 sweeps out and is lost during fermentation. Many packaged beers will contain 500–550 g $CO_2\,hl^{-1}$ and extra gas needs to be introduced to the product prior to packaging. At 1 atm pressure (as CO_2) and 0 °C, a beer will dissolve no more than 200 g hl^{-1} CO_2. Achievement of the very high levels of CO_2 demands the pressurizing of beer. None the less, when a beer container is opened, the gas usually stays in solution: the beer is 'supersaturated' with CO_2. When fobbing occurs spontaneously when a can or bottle is dispensed, it is called 'gushing.' In the absence of agitation as a cause, the most likely reason for the phenomenon is the presence of low-molecular-weight, highly hydrophobic peptides contributed by infection of barley, notably from *Fusarium*.

CO_2 affords the 'sparkle' to beer, through its reaction with the pain receptor mechanism of the trigeminal nerve. Apart from this influence on mouth feel, CO_2 establishes the extent of foam formation during

dispense and of course impacts heavily the delivery of volatiles into the head space of beers and therefore the aroma.

Other Gases in Beer

Oxygen has major negative effects on beer via its oxidation of various components, leading to cardboard and other stale notes and to the formation of haze. Brewers therefore strive to minimize the oxygen level in beer (preferably to $0.1\,mg\,l^{-1}$ or ideally even less) by avoiding ingress into the product downstream of the fermenter (yeast is a powerful oxygen scavenger and many brewers feel that it is only after the removal of yeast that oxygen is a concern).

The stale papery notes are due to a series of unsaturated carbonyl compounds, which are usually held to include *trans*-2-nonenal, though this has been no means fully substantiated. Several components of beer can degrade to such carbonyl substances during oxidation, including the 'higher' alcohols, the iso-α-acids, and the unsaturated fatty acids.

Nitrogen has long been introduced to beer to promote foam stability. The foams produced on beers containing N_2 tend to have populations of much smaller bubbles. These foams are more resistant to collapse and the beer displays a much smoother mouth feel. Typical levels of N_2 are $20–50\,mg\,l^{-1}$ which are three orders of magnitude lower than those of CO_2.

Water

The vast majority of beers contain between 90 and 95% water: apart from ethanol and carbon dioxide, the remaining constituents usually amount to less than 1% of the product. The water used for brewing must have no taints or hazardous components and water is often treated coming into the brewery to adsorb any contaminants by charcoal filtration and ultrafiltration. The ionic composition of water is also significant (although the grist also contributes salts). High levels of calcium, e.g., $200–350\,mg\,l^{-1}$, are claimed to be desirable in the production of ales. Much lower levels of calcium are traditionally associated with the brewing of lagers, for example, the water in pilsen (the home of the classic lager-style) contains less than $10\,mg\,l^{-1}$ calcium. A scientific justification for this is not entirely proven, but may relate to the role of calcium in promoting the surface behavior of the top-fermenting ale yeasts (calcium promotes the flocculation of yeast). Brewers may add salts or remove them (e.g., by reverse osmosis) to adjust the ionic composition of the product.

Two key impacts of calcium are on pH and the levels of oxalic acid surviving into beer. Calcium reacts with phosphate to lower the pH to the appropriate level for mashing:

$$3Ca^{2+} + 2HPO_4^{2-} \rightarrow Ca_3(PO_4)_2 + 2H^+$$

By precipitating out malt-derived oxalic acid in the brewhouse, calcium eliminates a material that, if surviving into beer, causes problems such as haze, gushing, and the blocking of dispense pipes.

Carbohydrates

Although most of the sugars found in wort are converted to ethanol by yeast, those containing four or more glucoses (i.e., maltotetraose and bigger, and known as dextrins) are not fermented. In order to sweeten the finished beer, some brewers may contrive to leave a proportion of fermentable sugar unconverted, or may add 'priming' sugar.

Most of the starch survives malting, because it is relatively resistant to enzymatic hydrolysis when in the form of large granules. When starch is gelatinized by heating, however, its constituents (amylose and amylopectin) become accessible. Mashing, therefore, incorporates a stage, typically at around 65 °C, to allow for the gelatinization of malt starch. The starch in some other cereals has higher gelatinization temperatures, e.g., rice and corn starches gelatinize over the range 70–80 °C and so are cooked separately from the main mash and then mixed with the malt mash to be degraded by the amylases from malt. (*See* **Carbohydrates**: Classification and Properties.)

Starch hydrolysis starts with α-amylase, a highly heat-resistant enzyme abundantly present in malt. It is endo-acting, catalyzing the hydrolysis of α-1→4 bonds within the starch, releasing dextrins. From amylose the enzyme release linear dextrins, and from amylopectin it produces dextrins with side chains. (*See* **Starch**: Structure, Properties, and Determination.)

β-Amylase is an exoenzyme, which removes maltose units from the nonreducing ends of starch and dextrin molecules. This enzyme has sufficient stability at 65 °C to achieve most of the conversion of starch and dextrins of which it is capable but if mashing is carried out at somewhat higher temperatures, e.g., 72 °C, then β-amylase is rapidly inactivated and the ensuing wort contains a high dextrin level and low content of fermentable sugars. Such high-temperature mashing can be used in the production of low-fermentability worts which will give low-alcohol beers.

β-Amylase cannot pass the α1→6 branch points in the dextrins formed from amylopectin. Limit dextrinase carries out this function but is generally present in fairly low concentrations because it is developed

very late during barley germination and because it is inactivated by association with another protein. Furthermore, limit dextrinase is relatively heat-labile. It is because of the limited action of limit dextrinase that 20–25% of the starch remains as dextrins in most conventionally mashed beers.

In so-called 'light' or 'lite' beers all of the starch is converted to ethanol. One way of achieving this is to add heat-stable glucoamylase or pullulanase from microbial sources to the mash or to the fermenter, although those brewers seeking to avoid any additions may modify their brewhouse and fermenter operations so as to take maximum opportunity of endogenous enzymes.

The major component of the barley cell walls is a linear β-glucan comprising approximately 67% β1→4 links with 33% β1→3 bonds that are for the most part every third or fourth linkage. The key enzyme hydrolyzing this molecule is an extremely heat-sensitive endo-β-glucanase, which is destroyed within 5 min of mashing at 65 °C. Accordingly, one of the main purposes of malting is to remove the cell walls, in large part through the action of this enzyme. In practice some cell wall material (at the distal end of the barley) will survive and if it is not properly degraded it can afford increased viscosity to the wort and cause sluggish wort separation in the brewhouse and retarded beer filtration. Some brewers mash-in at low temperatures such as 40–50 °C to allow β-glucanase to act, before raising the temperature to that needed for starch gelatinization. Alternatively, heat-stable β-glucanases from bacteria (e.g., *Bacillus subtilis*) or fungi (e.g., *Trichoderma, Penicillium*) may be used.

Proteins, Polypeptides, and Amino Acids

Amphipathic polypeptides in beer stabilize the bubbles in foam. Their hydrophobic regions 'drive' them into the surfaces produced during foaming and are largely responsible for their interaction with other hydrophobic molecules, notably the iso-α-acids. The matrix formed counters the force of surface tension that seeks to minimize increased surface area generated in foaming. Some of the protein in beer can react with polyphenols to form hazes.

There is no advantage in having significant levels of residual amino acids in beer; rather they are a risk in so far as they represent assimilable nitrogen sources for spoilage microorganisms. However, it is important that wort contains the correct balance of amino acids to support yeast growth and fermentation. Sufficient proteolysis must occur during malting and mashing to yield these amino acids and to remove haze-potentiating proteins, whilst leaving ample foam-positive polypeptide. (*See* **Protein**: Chemistry.)

The native proteins of barley undergo considerable degradation and denaturation in malting and brewing. During germination of barley, endoproteinases develop and attack the heart of substrate proteins to release polypeptides and peptides. An exoacting enzyme, carboxypeptidase, acts during malting and mashing by splitting off amino acids successively from the carboxyl-terminus of the peptides produced by the endoproteinases.

Lipids

Barley comprises some 3% by weight lipid, most of it in the embryo and aleurone. Very little survives into beer, because it is removed by adsorption on insoluble matrices (e.g., spent grain, trub) during the process. Lipids are severely detrimental to beer foam, disrupting the network of proteins and iso-α-acids in the bubble wall. As observed earlier, the unsaturated fatty acid component of lipids is suspected to be at least one precursor of the stale flavors (e.g., 'cardboard') that develop in beer. (*See* **Fats**: Classification.)

Flavors from Hops

Hops afford the bitterness (from the hop resins) and aroma (from the essential oils) to beer.

The most important resins are the α-acids (Figure 2), accounting for 2–15% of the dry weight of the hop,

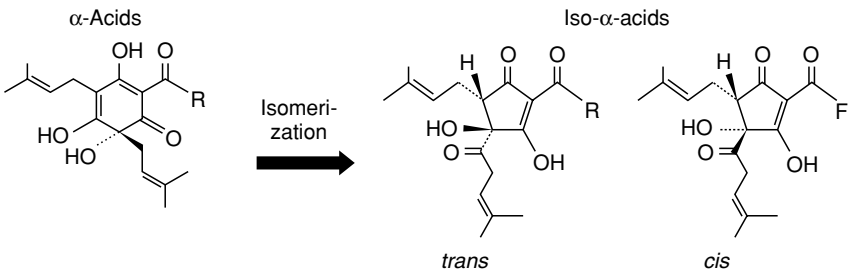

Figure 2 The isomerization of α-acids. R = $-CH_2CH(CH_3)_2$ in humulone and isohumulone. R = $-CH(CH_3)_2$ in cohumulone and isocohumulone. R = $-CH(CH_3)CH_2CH_3$ in adhumulone and isoadhumulone.

depending on variety and environment. The more α-acids, the greater the bitterness potential.

There are three different α-acids in hops, differing in side-chain structure. The α-acids are isomerized to form the more soluble and bitter iso-α-acids during wort boiling (Figure 2). Each iso-α-acid has two isomers, *cis* and *trans*, with differing orientation of the side chains. The six iso-α-acids have a range of bitterness intensities and it is generally held that the hops should have a relatively low level of cohumulone.

Apart from their impact on bitterness and foam, the iso-α-acids have strong antimicrobial properties and inhibit many Gram-positive bacteria.

Hops comprise 0.03–3% w/w essential oil, comprising a complex mixture of more than 300 compounds (Figure 3). The components are very volatile and tend to be lost during wort boiling. To ensure hoppiness in beer it is necessary to add a proportion of the hops late in the boil (late hopping, associated with lager-style products) or to the finished beer (dry hopping, associated with some ales).

Hops are either used in a traditional way as whole cones, though increasingly after hammer milling and extrusion into pellets, in which form they are more stable and more efficiently utilized, or after extraction by liquid carbon dioxide and isomerization by weak alkalis. The latter products can be added to beer downstream, greatly increasing the efficiency of bitterness utilization and flexibility of product formulation. The oils in such extracts may be separated from the resins and also added to beer to provide hoppy aroma.

In some preparations the iso-α-acids are reduced, using hydrogen gas in the presence of a palladium catalyst. Such reduction prevents the formation of the skunk-flavored 3-methyl-2-butene-1-thiol from the side chains when the bitter compounds are exposed to light. The reduced preparations can be used in the production of beers for packaging into green and clear glass bottles, which are particularly prone to light damage.

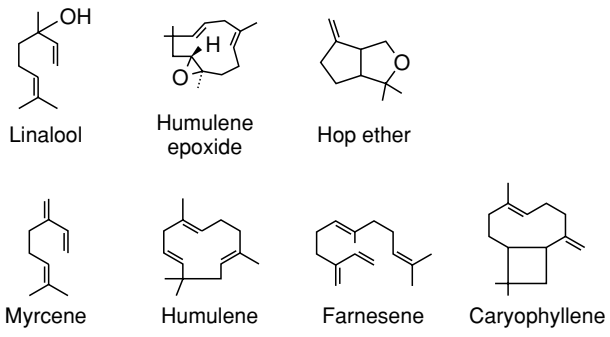

Figure 3 Some components of the essential oil fraction in hops.

Phenolic Materials

The simplest phenolic acid in beer is ferulic acid, a molecule found associated with the arabinoxylan component of the starchy endosperm cell walls. It may have antioxidant properties both for the beer and for the drinker's body. When decarboxylated, by an enzyme present in the yeast strains used to make wheat-based beers though not barley-based ones, 4-vinylguaiacol is produced, which has a distinct clove-like character.

Certain flavanoids, including catechin and quercetin, derive from the outer layers of malt and from hops. They, too, may have antioxidant properties, but when oxidized they polymerize, to produce tannoids that cross-link through the proline groups in certain beer polypeptides to form the insoluble complexes responsible for haze. Haze formation is lessened by reducing oxygen ingress and by reducing the levels of haze-forming polypeptides and phenolics. Silica hydrogels, tannic acid, and the proteinase papain have been used to attend to the former, while polyphenols can be removed using polyvinylpolypyrollidone (PVPP). Vigorous wort boiling and chilling of beer to as low a temperature as possible without freezing (−1 °C) are both important stages in the colloidal stabilization of beer.

The level of tannic materials in beer is probably too low to have any significant impact on astringency. (*See* **Tannins and Polyphenols**.)

Other Flavor Components of Beer

The four main flavor features detected by the tongue are bitterness, sweetness, sourness, and saltiness. The nose detects the other characteristic aromas of beer and this is a balance between positive and negative notes, each of which may be due to more than a single compound from different chemical classes. Although some of these substances originate in the malt and hops, many are products of yeast metabolism.

Esters Esters afford a fruity character to beer: two of the most important components are ethyl acetate and iso-amyl acetate. Esters are formed from their equivalent alcohols when the acetate group is available by not being needed for the synthesis of key components (lipids) of the yeast membranes. Therefore, factors that promote cell production lower ester production, and vice versa. Ester levels in beer are impacted, *inter alia*, by the ratio of carbon to nitrogen in the wort and by the amount of oxygen available to the yeast. The yeast strain is very

important; some strains inherently produce much higher ester levels.

Alcohols The alcohols are the immediate precursors of the esters. As the esters are substantially more flavor-active, it is important to regulate the levels of the higher alcohols if ester levels are also to be controlled.

The higher alcohols (i.e., those larger than ethanol) are produced by transformation of amino acids, so the levels of amino acids in wort have a major impact: the more amino acids available to the yeast, the greater the production of higher alcohols. However, *Saccharomyces* will also produce higher alcohols as an offshoot of the metabolic pathways responsible for amino acid elaboration, pathways that are particularly significant when the amount of assailable nitrogen in the wort is low. Hence higher alcohol production is increased when both too much and too little amino nitrogen is available to the yeast. Ale strains produce more higher alcohols than do lager strains. Conditions favoring increased yeast growth (e.g., excessive availability of oxygen) promote higher alcohol formation.

Sourness This is due to the organic acids, such as acetic, lactic, and succinic, produced by yeast during fermentation. It is the H^+ ion produced by their dissociation that causes sourness. Higher levels of the acids are produced in vigorous fermentations.

Vicinal diketones The vicinal diketones (VDKs), diacetyl and pentanedione, afford highly undesirable butterscotch and honey characters respectively to beer. These substances are offshoots of the pathways by which yeast produces certain amino acids. Precursors leak from the yeast and decompose spontaneously to form VDKs. Yeast, however, can reassimilate the VDK, provided the cells are healthy and remain in contact with the beer. Brewers may allow a temperature rise of 2–3 °C at the end of fermentation to speed up the removal of VDK. Additionally, a proportion of freshly fermenting wort is introduced to the beer as an inoculum of healthy yeast (a practice called krausening). Alternatively a bacterial enzyme (acetolactate decarboxylase) can be added to a fermentation; this enzyme converts acetolactate to acetoin, thereby avoiding the much more flavor-active diacetyl.

Persistently high VDK levels may be a symptom of infection by *Pediococcus* or *Lactobacillus* bacteria.

Sulfur compounds In the right levels and proportions, the sulfur compounds play a substantial role in determining the character of diverse beers. Some ales display a distinct hydrogen sulfide character when first dispensed, albeit one which subsides to reveal the dry hop character. Lagers often have a more complex sulfury character, which for many includes dimethyl sulfide (DMS), with its cooked corn/parsnip note. DMS is the best understood of all the sulfur compounds in terms of its production route.

All of the DMS ultimately originates from a precursor, *S*-methylmethionine (SMM), which develops in the barley embryo during germination. SMM is heatlabile, breaking down rapidly at temperatures above about 80 °C in malting and brewing. Accordingly, SMM is at a lower level in the more highly kilned ale malts and, therefore, there tends to be less DMS in ales. SMM is extracted into wort during mashing and is broken down during boiling and in the whirlpool. In a vigorous boil most of the SMM is converted to DMS and lost by volatilization. In the whirlpool the temperature is still hot enough to degrade SMM, but conditions are nonturbulent and the DMS tends to linger. Brewers aiming for a finite level of DMS in their beer specify a target level of SMM in the malt and will adjust the boil and whirlpool stages so as to deliver a certain level of DMS in the pitching wort. In fermentation a great deal of DMS is swept away with the CO_2, hence the level of DMS targeted in the wort is higher than that specified for the beer. Some of the DMS produced during malt kilning is oxidized to dimethyl sulfoxide (DMSO). This is not volatile but is water-soluble and enters wort, to be reduced by yeast to DMS. Hence the level of DMS in the finished beer is a function of how much DMS is present in pitching wort, how much is volatilized, and how much is replenished via the reduction of DMSO.

There is no simple relationship between the level of DMS in beer and the perception of its flavor. This is because phenylethanol and phenylethylacetate interfere with the perception of DMS. There must be many other antagonisms of this type contributing to the complexity of beer flavor, but they have not been studied.

Hydrogen sulfide (H_2S) is produced by yeast via the breakdown of cysteine or glutathione, or by the reduction of sulfate and sulfite. A vigorous fermentation purges H_2S and factors that hinder fermentation (such as a lack of zinc or vitamins) will increase H_2S levels in beer.

Malty notes As well as being the source of the DMS character in beer, malt contributes in other ways to flavor. Malty character is in part due to isovaleraldehyde, produced by the reaction of leucine with reductones in the malt. The toffee and caramel

Table 2 Order-of-magnitude composition of a typical pilsner beer

Component	*Typical level (mg l^{-1})*
Original wort gravity	120×10^3
Ethanol	40×10^3
Total carbohydrates	30×10^3
Glucose	150
Fructose	30
Sucrose	5
Maltose	1500
Maltotriose	2000
Dextrins derived from starch	24×10^3
β-Glucan	350
Pentosan	50
Protein	5000
Amino acids	1100
Thiamin	0.3
Riboflavin	0.4
Vitamin B_6	0.6
Pantothenic acid	1.5
Niacin	8.0
Biotin	0.01
Vitamin B_{12}	0.0001
Folic acid	0.2
Potassium	500
Sodium	30
Sulfate	200
Chloride	200
Organic acids (acetic, pyruvic, citric, succinic, malic, lactic, etc.)	700
Polyphenols	150
Iso-α-acids	30
Sulfur dioxide	4
Carbon dioxide	5×10^3
Nucleotides and nucleosides	300
Glycerol	1500
Higher alcohols	100
Ethyl acetate	15
Iso-amyl acetate	1
Acetaldehyde	5
Essential oils from hops	< 1
Dimethyl sulfide	0.06
Total organic sulfur compounds	< 1

character from crystal malts and the roasted, coffee-like notes in darker malts are due to complex components derived from amino acids and sugars when they cross-react during kilning. Of equal importance during kilning is the disappearance of grassy and beany notes, due *inter alia* to *cis*-3-hexen-1-ol, *trans*-2-hexenal, *trans*-2-*cis*-6-nonadienal, and 1-hexanol.

The cross-linking of sugars and amino acids induced by heating in kilning and wort boiling leads to the formation of melanoidins via the Maillard reaction. The melanoidins are responsible for imparting color to beer: darker beers are produced from grists incorporating malts and other adjuncts that have been kilned to more intense regimes. Polyphenol oxidation, occurring during wort production, can make a significant contribution to color in some of the paler beers.

Miscellaneous Acetaldehyde, which is converted to ethanol in actively fermenting yeast, imparts an undesirable 'green apples' character to beer. High levels of acetaldehyde are due to premature separation of yeast before fermentation is complete, poor yeast quality, or infection by the bacterium *Zymomonas*.

The short-chain fatty acids, with their rancid notes, are offshoots in the synthesis of membrane lipids by yeast. When yeast needs fewer lipids (when it needs to grow less), these compounds accumulate.

Table 2 offers a summary of the approximate chemical composition of a pilsner-type beer.

See also: **Alcohol**: Metabolism, Beneficial Effects, and Toxicology; **Antioxidants**: Natural Antioxidants; **Barrels**: Beer Making; **Barley**; **Beers**: Raw Materials; Wort Production; Biochemistry of Fermentation; Microbreweries; **Carbohydrates**: Classification and Properties; **Malt**: Malt Types and Products; Chemistry of Malting; **Packaging**: Packaging of Liquids; **Phenolic Compounds**; **Protein**: Chemistry; **Starch**: Structure, Properties, and Determination; **Tannins and Polyphenols**

Further Reading

Bamforth C (1998) *Beer: Tap into the Art and Science of Brewing*. New York: Insight.

Boulton C and Quain D (2001) *Brewing Yeast and Fermentation*. Oxford: Blackwell.

Briggs DE, Hough JS, Stevens R and Young TW (1981) *Malting and Brewing Science*. Volume 1: *Malt and Sweet Wort*. London: Chapman and Hall.

Hornsey IS (1999) *Brewing*. London: Royal Society of Chemistry.

Hough JS, Briggs DE, Stevens R and Young TW (1981) *Malting and Brewing Science*. Volume 2: Hopped Wort and Beer. London: Chapman and Hall.

Hughes PS and Baxter ED (2001) *Beer: Quality, Safety and Nutritional Aspects*. London: Royal Society of Chemistry.

Kunze W (1996) *Technology: Malting and Brewing*. Berlin: VLB.

Lewis MJ and Young TW (1995) *Brewing*. London: Chapman and Hall.

MacDonald J, Reeve PTV, Ruddlesden JD and White FH (1984) Current approaches to brewery fermentations. In: Bushell ME (ed.) *Progress in Industrial Microbiology*, Vol. 19: *Modern Applications of Traditional Biotechnologies*. Amsterdam: Elsevier.

Moll M (1991) *Beers and Coolers*. Andover: Intercept.

Microbreweries

A Nothaft, BrewTech Serviços Ltd, Jacarepaguá-Rio de Janeiro, Brazil

Background

Although the term 'microbrewery' strongly suggests a small-scale beer production plant, which is usually true, the microbrewery activity is actually what makes the difference in comparison with a large-scale beer factory. A microbrewery is a driving force in different marketplaces to reintroduce or, even more often, to develop many different types of beer with varying flavors. Tradition and diversity are key words when describing specialty beers, which once was defined by Charlie Papazian with the following statement.

> Beer is an expression of the human endeavor of living. We use science as a tool to create it, but its essence is and always will be a form of art expressing the variety of the world's lifestyles. The diversity of the world's food, beverages (other than beer), art, religion, music, clothes offer a great deal of varietal choice and contribute to the quality of life. Beer belongs to this list. The variety of specialty beers currently produced by small breweries and brewpubs create the opportunity for the beer industry to join global trends in offering diversity while also enhancing the image of beer.

History

After World War II, the scenario in the brewing industry world-wide began to change with the formation of large national breweries that, through the acquisition of regional breweries and the consolidation of the production sites as well as the product lines, strongly influenced the marketplace. In England, this consolidation happened during the 1960s and 1970s when six national operating breweries were formed as a result of the concentration of the production that changed the industry and the way in which beer was marketed.

In 1971, four men, Jim Makin, Bill Mellor, Michael Hardmann, and Graham Lees, unhappy with the changes in the beer industry, especially with the fact that they could no longer drink beers with traditional flavor, decided to create a movement aimed at changing this tendency. Initially, an association, named 'The Campaign for the Revitalization of Ale,' shortly after renamed 'Campaign for Real Ale' (CAMRA) set the starting conditions for the microbreweries on the way we know them today.

In the late 1970s, interest in cask ales was regained, and national breweries started to produce different brands, encouraging regional brewers to maintain efforts in producing the traditional beers. These changes and the response from the consumers were the catalyst for a wave of microbreweries in the early 1980s.

The success of CAMRA in England was a starting point for similar international activities. Soon after, CAMRA Canada was created, and the English CAMRA become one of the members of the European Beer Consumer's Union acting in the European Parliament.

In Canada, where beer production was extremely consolidated and the industry had to deal with strong legal restrictions for production and commercialization, the major activity of the organization was to legalize home brewing and the brewpubs.

In the USA, the driving organization was the American Home Brewer's Association, active since the early 1980s, providing information and promoting beer culture. This organization evolved to the Association of Brewers, which is an educational and trade association, still loyal to the initial motivation of promoting beer culture, through their publications, technical discussions on forums, conferences, and expositions, as well as beer festivals such as The Great American Beer Festival and, more recently, the World Beer Cup.

Table 1 Major beer types and styles

Type	*Style*
Ale	Belgian Witbier; German Weissbier; Dunkel-weizen; Weizenbock; Berliner Weisse; lambic; Gueuze; Faro; Kriek; Framboise
	Sweet Stout; Oatmeal Stout; Dry Stout; Imperial Stout Porter
	Pale Mild; Dark Mild; Bitter; Best Bitter; Strong Bitter; Brown Ale; Old Ale; Irish Red Ale; Scotch Ale; Belgian Brown; Pale Ale; India Pale Ale; Belgian Ales; Saisons; Trappisten; Altbier; Kölsch; American Ale; Cream Ale; Barley Wine
Lager	Pilsener; Dortmunder; Export; Strong Lagers; American Malt Liquor
	Vienna; Märzen; Oktoberfest; Amber; Red Lager
	Munich Pale; Munich Dark; Dark Bock; Pale Bock; Double Bock
Special	Fruit Beers
	Chilli Beers
	Honey Beers
	Spiced Ales
	Smoke Beers
	Stone Beers
	California Common

Beer Styles

Diversity is the driving force for microbreweries, and many different types of beer have been offered to the public. Some would certainly have disappeared if microbrewery activity did not exist, being found only in history books. Additionally, the segment is constantly introducing new styles of beers reflecting the evolution of the market and technology.

Usually, beers are divided into ales and lager, distinguished by the fact that ales have a fruity aroma and flavor, and lagers are smoother and crisper. Also, some beers are made using specific techniques and ingredients, rendering their classification as 'specialty' beers, and they can be either ales or lagers.

Table 1 lists the major types of beer; note that they may have additional differentiations depending on local conditions in different marketplaces around the world.

Microbrewery Operation

The size of the microbrewery plant size varies according to the type of beer being produced, as detailed in Table 2. Differences in plant size also entail different methods of processing and quality control. In larger plants, the same standards as those in major breweries apply. Measurements to insure that potential processing deviations can be identified and corrected before they can affect the quality of the final product are mandatory.

In microbreweries, where usually just a small number of tanks are in use at a given time, and the products are in different stages of process, a completely different approach for quality control is necessary. The brewer has to focus on all precautions necessary during the process to guarantee that the product meets the desired standards. This is a relevant difference between microbreweries and large-scale production.

Looking further at the personal skills needed to manage the different types of breweries, people involved with microbreweries are usually individuals

Table 2 Characteristics of microbreweries

Brewpub	Most of the production is dedicated for in-house sales
Microbrewery	Produces for external sales in a reduced geographical area
Regional brewery	Produces for external sales in a larger geographical area
Contract brewing	The company develops the product formula, merchandising and sales, and production is hired from a third party company

Figure 1 (see color plate 8) View of the brewhouse in a microbrewery.

Table 3 Recommended tasks for quality control in a microbrewery

Process step	*Minimal level of control*
Raw materials	
Water	Sensorial evaluation before every use and periodical external evaluation
Malt	Sensorial evaluation, malt analysis from the supplier, and periodical external evaluation
Adjuncts	Sensorial evaluation (hot water extract) and supplier analysis
Wort production	
Raw material addition	Weigh control
Malt grist	Visual evaluation and periodical external evaluation
Water addition	Volume control
Saccharification	Iodine test
Evaporation rate	Calculation
Production yield	Calculation
pH	Determination
Break formation	Visual evaluation
Bitterness	Sensory evaluation and periodical external evaluation
Color	Visual evaluation and periodical external evaluation
Wort clarity	Visual evaluation
Wort cooling/aeration	
Cooling temperature	Temperature control during the whole process
Process time	Record time needed to cool
Oxygen intake	Visual evaluation of the wort after the oxygen intake
Terminal load on wort	Total time the wort was kept over 85 °C, measured from the filling of the kettle to the final of the wort cooling
Yeast dosing	
Yeast control	Check flavor and appearance of the yeast, frequent renewal, and periodical microbiological control
Yeast storage	Control how long yeast was stored between brews
Dosing volume	Weight of yeast added to fermentation
Wort clarity	Visual evaluation
Fermentation/aging	
Extract reduction	Record the extract reduction against the fermentation time
pH	Check the pH on the end of fermentation and aging
Time/temperature/pressure	Record the process parameters
Diacetyl	Sensorial evaluation at the end of fermentation and aging
Cells in suspension	Visual evaluation at the end of fermentation and aging
Oxygen intake	Preventive operational procedures
Filtration/packaging	
Appearance	Visual evaluation of the brightness of the beer
Carbonation	Sensory evaluation and periodical external evaluation
Foaming	Visual evaluation and periodical external evaluation
Alcohol/extract	Control with a saccharometer; although the analysis is not correct, it gives indications about fluctuations. Periodical external evaluation
pH	Control of the final product. Deviations are strong indicators of microbiological contamination
Bitterness	Sensory evaluation and periodical external evaluation
Diacetyl	Sensory evaluation and periodical external evaluation
Oxidation	Sensory evaluation and periodical external evaluation

who constantly like to face new challenges. They are involved in the operation in all different aspects, often being involved at different stages, from providing the resources for brewing to sales or technical support in the marketplace, and are often keen to play an active part in developing a new beer formulation.

This does not mean that they are better brewers than those running large breweries. Large brewery professionals gain their motivation by managing such a complex structure involved in the beer production.

A plant showing the brewing house of a microbrewery is presented in Figure 1.

Quality Control

The construction of a laboratory with sufficient capacity to run all necessary analyses to control all phases of the process can easily double the cost of the investment made in a microbrewery, and, as discussed previously, this will not insure the quality of the final product. This means that strict control of the process, by recording and evaluating processing data and also by sensory evaluation of all process steps, will contribute effectively to a successful quality control.

Table 3 lists the recommended control tasks for the operation of a small-scale microbrewery.

Perspectives

Microbrewing has become a reality in most countries, mainly due to the diversity of products offered to consumers in general and to beer lovers in particular. It has become a well-established segment of the brewing industry in Europe, North America, Asia, and Oceania. However, in South America and Africa, this tendency has not yet achieved the same degree of development. Clearly, there is potential for growth, and it is anticipated that the formation of many new microbreweries around the world will introduce a variety of products and flavors to beer consumers.

See also: **Beers**: History and Types; Raw Materials; Chemistry of Brewing; Biochemistry of Fermentation

Further Reading

Beer Seller's Guide (1997) Institute for Brewing Studies.(ibs@aob.org)

Brewery Planner (1996) Brewers Publications.

Daniels R (1996) *Designing Great Beers*. Brewers Publications.

Eschenbach R (1993) *Gasthausbrauerein*. Verlag Hans Carl Getränke-Fachverlag.

Evaluating Beer (1993) Brewers Publications.

Jackson M (1988) *The New World Guide to Beer*. Courage Books.

Papazian C (1998) *Zymurgy, The Best Articles and Advice from America's # 1 Home Brewing magazine*. Avon Books.

Protz R (1995) *The Ale Trail*. Eric Dobby.

BEHAVIORAL (BEHAVIOURAL) EFFECTS OF DIET

D A Booth, University of Birmingham, Edgbaston, Birmingham, UK

Nutritional Effects on Behavior

Behavior other than eating and drinking can be influenced by physiological effects of the constituents of foods and beverages. Indeed, consumers' attributions of effects on their subjective feelings or objective performance appear to be among the major determinants of the acceptance or refusal of certain items of the diet. This aspect of the many relationships among food, nutrition, and behavior is the topic of this article.

Scientific Study of Effects of Diet on Behavior

The possible influences of diet on behavior have attracted a considerable amount of public and professional attention in recent years, and yet reliable scientific information on such effects remains very limited.

This is partly because studies claiming positive findings are often open to serious criticism of their design and also of the theory behind them. Even arguably supportable findings have frequently failed to be substantiated after some years, at least on the scale initially claimed. A field that acquires a dubious reputation, or at best an air of scientific intractability, will naturally have difficulty attracting much good research. Nevertheless, definite scientific progress could be made by close collaboration among social cognitive psychologists, applied nutritionists, and food technologists.

The measurement of behavioral performance or experiences is not the real difficulty in studying dietary effects on behavior, despite a widespread misconception to the contrary. Even the answers to questions about personal viewpoints, which seem such 'soft' data to biologists and chemists, can be shown, by the well-established criteria of multivariate psychometrics, to reflect underlying determinants. Such adequately scaled instruments provide valid and reliable expressions of differences in emotional state and in judgments of one's own ability. Sound measurements of intellectual performance, such as ability to concentrate or remember, or indeed physical performance, such as athletic endurance, can be constructed by cognitive or exercise scientists at least as easily as self-ascription scales such as mood or ability.

The greatest difficulties arise in the design and interpretation of investigations of real-life psychological phenomena and, indeed, in measuring the composition of the everyday diet as well. As in any other area of science, effective research requires a good theoretical grasp of the mechanisms involved and hence of the methods for obtaining evidence of their operation. There are no standardized tests for

effects of diet on behavior and there never will be, any more than there can be standardized tests for dietary effects on children's growth or for consumer perception of food quality.

Hence what follows is an impressionistic sketch of the theoretical issues of how diet might affect behavior, with brief comments on the current state of evidence for the mechanisms involved. Detailed reviews referring to original studies are included in the Further Reading section.

Long-Term Effects of Diet on Behavior

There are broadly speaking two ways in which constituents of the diet could have long-term psychological effects. They may affect the general physical growth or deterioration of the nervous system. Alternatively, diet might permanently affect the functioning of the brain, via short-term effects on changes within specific neural connections that have long-term psychological consequences.

Brain Development

The nerve cells in the human brain have virtually all been formed prior to birth. Considerable elaboration of interneuronal connections occurs postnatally, however. Maternal nutrition might therefore affect the fetus or the breast-fed infant. Nevertheless, the brain is very effective at extracting essential nutrients from the blood supply and is not susceptible to overdevelopment. Hence there is reason to think that only metabolic disorder or extreme deficiency in the diet is liable to prejudice development of the brain.

In the rare inherited metabolic disorders, the accumulation of unusual catabolites damages the brain unless prevented by dietary or genetic engineering. As well as neurological disorders, more or less severe learning difficulties (mental handicaps) can result. (*See* **Inborn Errors of Metabolism**: Overview.)

Possibilities under study include psychological sequelae of neurological defects, prematurity or very low birth weight, to which maternal deficiency (such as in folic acid) might contribute. There is concern that some fatty acids in breast milk may be needed in infant milk formulae, at least in premature babies, in order for the brain to develop to full intellectual potential.

Energy–protein deficiency early in childhood is associated with slow intellectual development. It is difficult to disentangle effects of malnutrition and disease on brain growth from functional handicaps arising from effects of economic and social disadvantage. Nutritional supplementation and psychosocial stimulation exert separate effects. In addition, school failure in low-income areas is associated with missing breakfast. However, regular provision of supplements or breakfasts can be an important environmental change, aside from its physiological effects. (*See* **Breakfast and Performance**; **Protein**: Deficiency.)

Brain Damage

Nutritional factors have repeatedly been proposed for schizophrenic and other psychiatrically diagnosed disorders. None has yet been substantiated, however, when subjected to controlled investigation. Thiamin deficiency produces neurological symptoms, and lack of this and other B vitamins is generally considered to contribute to some encephalopathies that involve several memory defects. Therapeutic effects of supplementation remain to be established. (*See* **Thiamin**: Physiology.)

Toxic contaminants of the diet may damage the brain, with psychological sequelae. Nevertheless, no long-term psychological effect of the present dietary levels of residues, heavy metals, etc., has been established to date. Ingestion of flakes of paint or contaminated dust can contribute to an accumulation of lead in the brain but this generally results from hands or objects being brought to the mouth, not from dietary consumption. (*See* **Heavy Metal Toxicology**.)

Short-Term Mediation of Long-Term Effects

Long-term psychological effects of dietary constituents on children, such as those on intelligence quotient (IQ), personality, or behavior disorders, are likely to be mediated by acute psychological effects of the daily diet. For example, if school performance were permanently improved by regular provision of breakfast, it could be because each meal facilitated learning that morning.

Additives and hyperactivity Feingold's initial hypothesis that intolerances to food additives such as colorings cause widespread behavioral problems in children was narrowed at an early stage to tartrazine and salicylates. Much subsequent investigation has established the very low incidences of food intolerances and the infrequency and transience of proven allergenic reactions to food proteins in young children. Moreover, recent work has shown no systematic behavioral effect when a toxicological sensitivity has been provoked. (*See* **Food Additives**: Safety; **Food Intolerance**: Types; Food Allergies; Lactose Intolerance.)

Vitamins and IQ Several recent reports have provided data interpreted as supporting the hypothesis that supplementation with one or more unspecified vitamins and/or minerals prevents a supposed slowing of the normal rise in nonverbal intelligence test

scores (IQ) in schoolchildren whose diets are alleged to be deficient in those micronutrients.

As yet, however, there is no satisfactory evidence that those whose IQ scores appear to respond to supplements are deficient in micronutrients, because the only relevant dietary data are the children's dietary records and these are likely to be confounded by factors that influence scores in IQ tests. Furthermore, children who are estimated to be consuming less than recommended levels of a micronutrient generally appear also to have inadequate energy intakes. The accuracy of these records is therefore suspect. Alternatively, these children may be chronically hungry or reacting to other sorts of deprivation. This might account at least in part for the fidgetiness and apparent lack of attention that have been reported in children having the poorer dietary records, which in turn could make for poor learning and hence slowed development of performance in some IQ tests. (*See* **Energy**: Measurement of Food Energy.)

Another (not incompatible) hypothesis is that the high sugar intake of some children might have acute sedative effects. This could result in difficulty in thinking clearly, so that the child becomes restless when faced with an intellectual challenge. (*See* **Sucrose**: Dietary Importance.)

No biochemically sound mechanism has yet been proposed by which moderate vitamin or mineral deficiency could affect nonverbal learning. Theory is hardly feasible in any case because, although some speculations have been offered, the neural bases of individual differences in human intelligence are as yet unknown.

Short-Term Effects of Diet on Behavior

The acute effects on behavior of a constituent of a drink or even of the composition of a meal are both theoretically and methodologically much more accessible than chronic dietary effects. Even so, they pose formidable multidisciplinary challenges to investigators, in design and analysis of characteristics of the diets, the consumers and the situations to be observed.

Design of Investigations

The foods When the issue is physiological effects in the brain, the postingestional effects of the diet must be distinguished from its sensory effects. This is crucial in principle because a flavor or texture may suggest effects to the consumer (e.g., filling, nutritious, junk food, luxury) that could confuse the search for physiologically mediated effects. The distinction has become demonstrably necessary in some cases, such as reduction of distress – just the taste of sugar can quieten a baby and raise the pain threshold. Hence differences in dose of the hypothesized active constituent must not be detectable to the subject. If the agent cannot be swallowed in a capsule, demonstration of effective sensory matching or masking is necessary. (*See* **Sensory Evaluation**: Sensory Characteristics of Human Foods.)

For major constituents of the diet, the technological requirements to design unambiguous research protocols may be so unusual and difficult as to be impossible at reasonable expense. Experimental cognitive psychology can provide a way round by measurements that pick out the effects of sensorily triggered expectations, but this demands sophisticated design and analysis (and a modicum of good luck).

The eaters Individuals may vary from one another physiologically in susceptibility to a dietary factor. They will certainly differ in prior experiences with foods and drinks containing the component(s) of interest.

Where the psychological effect being studied is known to the public, and especially if it is commonly sought or avoided, the effectiveness of a study is likely to depend on designing and analyzing it around the participants' uses of items and the occasions when they are consumed. This will not only help to accommodate differences in both physiology and experience; it can also enable differences to be exploited. For example, disguised variations of a constituent, around the level at which an individual normally consumes it with the expectation of a relevant psychological effect, are likely to be a more sensitive design than the same levels imposed on everyone tested, without regard to their normal expectations.

The test tasks and situations A psychological effect may be wanted only in certain situations by a given consumer. Indeed, the effect may only occur in specific circumstances. That can be because the effect is part of conventional behavior and experience in those circumstances or because the user personally discovered something, when the conditions were right, that the item could do for mood or performance. It may well be crucial, therefore, that the setting investigated and the state of mind of the research participant are suited to the mood or performance under study and that the behavioral tests measure those particular benefits that each participant obtains during normal use of the dietary item.

Alcohol

Alcoholic beverages have been an important part of the diet in many cultures from time immemorial. The

alcohol content is widely regarded as a source of good cheer on social occasions and soothing or even emotionally anesthetic at times of distress. However, incapacitating effects of alcohol are also well known, as is the risk of problems from its continuous heavy use. (*See* **Alcohol**: Metabolism, Beneficial Effects, and Toxicology.)

Neurophysiological evidence suggests that ethanol has psychoactive effects by acting on the γ-aminobutyric acid (GABA) receptor system at inhibitory synapses throughout the nervous system. Such neural inhibition is critical to the precision of information processing. Thus the postingestional psychological effects of ethanol are likely to be protean. It may be that all forms of fine control, including precise physical movements (walking down a straight line), vigilance against subtle dangers (crossing in front of approaching traffic), and self-critical social performance (ethical inhibitions and fears for self-esteem), are rendered less competent.

However, ethanol has been regarded in animal pharmacology as specifically anxiety-reducing at subsedative doses. The benzodiazepines (such as Valium), that also act on the GABA system, are used as antianxiety agents, as well as muscle relaxants and sedatives. Thus the neural actions of ethanol may be more effective at reducing tensions than at incapacitating generally.

Many of the effects of common levels of consumption of alcoholic beverages on mood and social behavior appear to have at least as much to do with the drinking situation as with neural actions of ethanol. Merriment and perhaps sexual predation are what is expected at parties; personal aggressiveness and vandalism became a norm for soccer fans, and gloom is natural for the lone(ly) drinker. All these effects have been seen in experimental studies, but there tend to be large 'placebo' or expectancy effects too. It seems that ethanol contributes some disinhibition or incapacitation but a participative spirit achieves the rest.

The behavioral effects of ethanol in the diet are therefore complex to investigate. The sensory qualities of ethanol and the aftereffects of its ingestion on bodily sensations and mental and physical abilities are well known to experienced drinkers. This weakens the interpretations of sophisticated experiments on the behavioral effects of ethanol (as also for other familiar substances).

One example is the so-called 'balanced placebo' design. This has four conditions, two drinks with ethanol and two without, where one of each pair is stated by the investigator to contain alcohol and the other is said not to. The presence and absence of alcohol are supposed to be masked but, when the sensory disguise is checked, it is often found to have been ineffective. Furthermore, characteristic effects of ethanol are liable to be noticed some minutes after ingestion of the ethanol-spiked drink that was alleged to be an alcohol-free drink. This is likely to provoke an emotional reaction and to change the strategy in the task set by the experimenter. An experienced user not feeling the usual effects when the drink was falsely said to contain alcohol is also likely to react to that disparity but in a different way, perhaps more disappointed than angry. Thus the effects on behavior of stated and actual alcohol contents cannot be separated out by analysis of this two-by-two design on an additive model: it is not balanced and placebo control is impracticable (as generally for familiar psychoactive substances). Detailed evidence is needed on the cognitive processes after drinking more or less alcohol with the normal approximate knowledge of amount.

Traditional dose–response studies of behavioral effects of alcohol are subject to similar problems, even when the variations in alcohol content are not detected during consumption and aftereffects are hard to distinguish, e.g., within the lower range of doses. Sensitive tests of psychomotor performance can show deficits at low doses that are proportionate to those at higher doses, thus justifying an argument for a zero blood alcohol limit on drivers. However, the consumer of a known amount of alcohol before driving or working may pay closer attention to the task and put extra effort into control and decision. In some circumstances, such an effort can overcompensate for the detrimental effect of ethanol, associating a low dose with objectively improved performance. Of course, such an effect should not be confused with the personal belief that a little alcohol has improved one's performance, since that is liable to be an illusion fostered by ethanol's disruption of self-critical abilities. Nevertheless, this phenomenon does illustrate how actively people use the effects of food and drink; they are not just affected passively or automatically.

Caffeine

Caffeine is thought to be able to act as a mild alerting agent by blocking synaptic receptors for endogenous adenosine, which is sedative. However, the experimental literature on human behavior has been confused by the use of large doses relative to those obtained through normal coffee, tea, and cola drinking, by differences between people in responsiveness to caffeine, and in the benefits to performance or mood habitually obtained from such drinks, and by unrealistic tests for such benefits. (*See* **Caffeine**.)

In a study using normal doses, quite strong and consistent effects of caffeine (at a dose as low as

32 mg) were shown on tests of mental concentration (attentive and integrative thinking), effects which the participants themselves noticed in ratings of their state. Furthermore, in another study the participants rated themselves also to be more cheerful and less anxious, perhaps because they felt that they were doing better at the cognitive tests. Such conscious benefits may mediate some of the attractions of caffeinated drinks, over and above cultural norms and advertisers' implications.

A remaining weakness in even these studies is a lack of dose–response relationships. As recognized by physiologists, engineers, and others, if an effect does not become stronger with increasing strength of an influence, that is a sign we are not looking at the actual mechanism. Grouped designs may sample a wide range of personal dose optima, depending on the benefits habitually gained from the use of caffeine in particular contexts by different individuals. Hence studies of behavioral effects of caffeine (and of other dietary constituents) should investigate individuals' habitual uses at doses ranging around that which is usual for each person in that situation.

Blood Glucose and Behavior

Hypoglycemia has been blamed for aggressiveness in members of societies subsisting on low energy intakes and for restlessness and poor attention in children consuming large amounts of sucrose. However, such claims are not well supported by measurements of blood glucose levels. Hypoglycemia is in fact quite rare, in contrast to the prevalences claimed for such problematic behavior. (*See* **Hypoglycemia (Hypoglycaemia)**.)

The consumption of sucrose with relatively little complex carbohydrate and other nutrients might indeed lead to reactive hypoglycemia, arising from overstimulation of insulin secretion. However, the behavioral aftereffect of consuming a large amount of sugar is if anything drowsiness, not agitation. Sucrose challenges specifically to children diagnosed as suffering from attention-deficit hyperactivity disorder have mostly shown no effect on physical activity. However, the sedative effects could acutely impair attention and, in theory, a child's awareness of this might exacerbate problem behavior in attention-demanding situations. It must be noted, on the other hand, that a parent's or institution staff's concern about the sugar intake of a problem child may be no more than a desperate hope for some remedy for the unmanageable behavior. These relations between diet and behavior also need careful sociopsychological analysis before biomedical investment.

Somewhat paradoxically in the light of the above, it has been recently suggested that administration of glucose might improve memory in the elderly, perhaps via norepinephrine (noradrenaline) systems in the brain. However, animal experiments have involved administering concentrated glucose solutions; these are stressful and may improve memory simply by alerting the rat.

Dietary Effects on Monoamine Neurotransmitters

A meal that is high in carbohydrate and low in protein content stimulates insulin secretion in the rat. The insulin facilitates uptake by muscle of circulating branched-chain amino acids (BCAAs: leucine, isoleucine, and valine). These amino acids compete with other large neutral amino acids (LNAAs) for transport from the blood into the brain. The LNAAs include tryptophan, the precursor of the neurotransmitter 5-hydroxytryptamine (5-HT, or serotonin), and phenylalanine and tyrosine, precursors of the catecholamine transmitters dopamine and norepinephrine (noradrenaline). The supply of precursor can limit the rate of synthesis of the transmitter, especially in the case of 5-HT. Thus reduced competition by BCAAs for brain tryptophan uptake is liable to increase the activity of serotonergic (5-HT-transmitted) synapses. (*See* **Amino Acids**: Properties and Occurrence.)

Serotonergic neurons are important in the control of sleep. Oral administration of a substantial dose of tryptophan is sedative. This tryptophan supply effect on brain 5-HT probably explains why a high-carbohydrate meal promotes postprandial sleep in the rat. (*See* **Carbohydrates**: Requirements and Dietary Importance.)

The LNAAs are abundant in protein mixtures of high biological quality. Thus, although a high-protein, low-carbohydrate meal also provokes insulin secretion in the rat, plasma levels of BCAAs are kept high by absorption and are not reduced enough to have a substantial effect on competition with tryptophan for transport into the brain. Hence the high-protein meal does not increase 5-HT activity and induce sedation by that mechanism. (*See* **Protein**: Requirements.)

Relatively modest dietary levels of protein, e.g., 10–15% in the rat, keep the ratio of tryptophan to other LNAAs in blood plasma low enough to have no effect on brain 5-HT levels. However, as little as 4% protein keeps the plasma ratio low in human subjects. Few eating occasions provide that little protein. Even chocolate and sugar confectionery may contain a milk protein and/or grain protein. Hence it is unlikely that carbohydrate-rich foods induce sedation or other mood changes in people via the 5-HT mechanism.

This is a difficulty for the suggestion that drugs and psychiatric disorders affecting serotonergic activity

induce a craving for carbohydrate via the action of carbohydrate-rich foods on tryptophan uptake, 5-HT, and mood. Another difficulty is that many of these foods are sweet, an oral sensation that by itself apparently dampens distress via opioid mechanisms. In addition, the high-carbohydrate foods reportedly craved are high-fat foods, and the creaminess or crispiness, with or without sweetness, makes them highly palatable. They can therefore be pleasurable and cheering to eat, independently of postingestional factors.

Finally, in a further illustration of the need for psychosocial analyses of diet and behavior, these craved foods are generally convenience products that are recognized as nutritionally less desirable (so-called junk foods). Hence they may be avoided and as a result become tempting and craved for. Their consumption could then have the powerful impact on mood of any guiltridden sensual indulgence ('naughty but nice'), by purely cognitive processes with no particular neurotransmitter mediation.

Meals

A modest amount of food is widely regarded as mentally and physically energizing or refreshing. However, a heavy meal is expected to make one drowsy (while not necessarily promoting a good night's sleep). Recent behavioral research has provided some support for these conventional beliefs but the physiological mechanisms involved remain obscure.

There is a semicircadian rhythm of arousal, including a period of reduced performance in midafternoon as well as a more profound reduction in the small hours after midnight. A series of experiments has shown that a substantial lunch tends to depress objective and subjective alertness further for a few hours. Going without lunch can therefore improve cognitive performance in midafternoon, although other consequences may not be desirable. Caffeine helps to counter the 'postlunch dip' and alcohol makes it worse. Different aspects of attention are affected by different protein:carbohydrate ratios in the meal. There is as yet no clear basis for this theory, in terms of either the cognitive processes or the physiological actions of food involved in these effects.

Breakfast is reputed to improve performance at work, although the evidence has been largely correlational with accident rates or school reports. Such effects are likely to depend on the size and composition of the breakfast, the physiology, personality, and attributions of the consumer and the activities and tasks that follow the meal. (*See* **Breakfast and Performance**.)

See also: **Alcohol**: Metabolism, Beneficial Effects, and Toxicology; **Breakfast and Performance**; **Caffeine**; **Carbohydrates**: Requirements and Dietary Importance; **Energy**: Measurement of Food Energy; **Food Additives**: Safety; **Food Intolerance**: Types; Food Allergies; Lactose Intolerance; **Heavy Metal Toxicology**; **Inborn Errors of Metabolism**: Overview; **Protein**: Requirements; Deficiency; **Sensory Evaluation**: Sensory Characteristics of Human Foods; **Sucrose**: Dietary Importance

Further Reading

Bendich A and Butterworth CE (eds) (1991) *Micronutrients in Health and in Disease Prevention*. New York: Marcel Dekker.

Blane HT and Leonard KE (eds) (1987) *Psychological Theories of Drinking and Alcoholism*. New York: Guilford Press.

Dews PB (ed.) (1984) *Caffeine*. Berlin: Springer-Verlag.

Heuther G (ed.) (1988) *Amino Acid Availability and Brain Function in Health and Disease*. Berlin: Springer-Verlag.

Lieberman HR, Wurtman RJ, Garfield GS, Emde GG and Coviella ILG (1987) The effects of low doses of caffeine on human performance and mood. *Psychopharmacology* 92: 308–312.

Shepherd R (ed.) (1989) *Handbook of the Psychophysiology of Human Eating*. Chichester: John Wiley.

Smith AP, Leekam S, Ralph A and McNeill G (1988) The influence of meal composition on post-lunch changes in performance efficiency and mood. *Appetite* 10: 195–203.

Thayer RE (1989) *The Biopsychology of Mood and Arousal*. New York: Oxford University Press.

BERIBERI

K J Carpenter, University of California, Berkeley, CA, USA

Occurrence in Asia

The first Western physicians allowed to work in Japan in the 1870s were surprised to discover the existence of a serious disease previously unknown to them and 'second only to smallpox in its ravages.' In Japan, it was known as 'kakké,' but was soon recognized as being identical to the disease known in South-east Asia as 'beriberi,' a native name now universally adopted, which may originally have meant 'great weakness.' Characteristically, it began with a feeling of weakness in the legs and a loss of feeling in the feet. Then, in many but not all cases, the legs and then the trunk would swell with retained water. Finally, the heart would be affected so that the subject gasped for breath, and would die from heart failure.

Older records from both Japan and China showed that it had been known for some centuries, although it had been the opinion of two eighteenth century Japanese physicians that the disease had become worse after about 1750. The early records also indicated that it was largely a disease of the wet summer months and could attack even the well off.

Infection or Malnutrition?

Those most at risk were men in the newly modernized Japanese army and navy, and also prisoners. As these were all people living together in large groups, and with the excitement in this period at other diseases being traced to the transmission of pathogenic bacteria, this seemed a likely cause for beriberi also. Yet, it was difficult, on this basis, to explain the frequent observation that a naval ship would leave its base with all its crew in good health, yet, after a month or more in isolation at sea, the disease would sweep through the crew.

Kanehiro Takaki, a surgeon on the naval staff, was directed in 1878 to work on the problem. He knew that the ships had been built in Britain and that they followed the general practices of the British navy where there was no beriberi. The only difference that caught his attention was in the rations issued to the men: the Japanese issues contained less protein and did not meet the high standard in force at that time in Europe. He therefore persuaded his superiors to permit a trial of modified rations with a proportion of the rice being replaced by meat, condensed milk, vegetables, and barley. The change was a complete success, and it was found that even just the use of barley in place of one half of the rice staple was enough to prevent the disease, which Takaki now believed to have resulted from a deficiency of protein in the earlier rations.

The Japanese army, perhaps as the result of inter-service rivalry, did not follow the navy and in the short Russo-Japanese war of 1904–1905, some 100 000 of their soldiers had to be invalided home from Manchuria suffering from beriberi.

A Disease in Chickens

Meanwhile, the disease had become an equally serious problem in the Dutch East Indies (now Indonesia) (Figure 1). After a punitive military expedition had to be withdrawn because of a beriberi epidemic, the Dutch government dispatched a small commission to try to identify the bacteria responsible for the disease. After a few months, it was thought that the microorganism had been found, and Christiaan Eijkman, a young Army physician, remained behind to confirm its activity in animal models.

Some of the chickens that Eijkman had injected with blood from beriberi patients developed signs of leg weakness, but so did some of his uninjected controls, suggesting that the condition was so infectious that it could 'jump' from cage to cage. Autopsies of the affected birds showed degenerated peripheral nerves. But in the following months, none of the next batch of birds developed the condition. Eijkman discovered that when the leg weakness had appeared, the man in charge of the birds had been feeding them on cooked white rice left over from feeding the beriberi victims in the adjoining hospital, instead of buying rough, feed-grade rice.

A long series of feeding trials confirmed that birds fed on white rice would become sick with leg weakness, whereas those given supplements of rice polishings (still present in feed-grade rice) remained healthy. This was only an animal disease, but a survey by the medical inspector of prisons in Java showed that prisoners who had been receiving white rice as their staple issue were susceptible to beriberi, whereas those receiving brown rice were not.

The Concept of a Vitamin

Eijkman, who believed that the disease was a kind of starch poisoning, now had to be invalided home with malaria. His successor, Gerrit Grijns, found that birds

Figure 1 Two prisoners in Java with beriberi and needing assistance to walk. From Vorderman (1897) *Onderzoek naar het gevangenissen op Java en Madoera en het voorkomer van beri-beri order de geintemeerden* Batavia: Jav. Boekh & Drukkerij.

became sick even when fed on meat that had been autoclaved. After further work, his statement in 1901 was perhaps the progenitor of the 'vitamin era' in nutritional research: 'there occur in various natural foods substances which cannot be absent without serious injury... they are easily disintegrated... and cannot be replaced by simple chemical compounds.'

The work of Eijkman and Grijns was confirmed by British investigators in Malaysia and by Americans in the Philippines, and attempts began to extract the active material from rice polishings and to concentrate it. There are moving accounts of scientists in Manila being implored by local doctors to bring a few spoonfuls of extracted syrup to save the lives of infants with beriberi, and of the babies' spectacular recoveries. Women themselves seemed less susceptible to beriberi than men, but when mothers were receiving a diet of low thiamin content, their breast-fed babies were at a high risk of dying with acute infantile beriberi.

Isolation of Thiamin

Isolation of the active factor proved very difficult. Each stage of extraction, and then further partitioning by reprecipitation involved biological assays with birds. Eventually, the next generation of Dutch workers in Indonesia was successful, obtaining a few milligrams of active crystalline material after starting with one-third of a ton of rice polishings and going through at least 16 separation stages in which much of the vitamin was lost. It was found that adding just 2 p.p.m. of the crystals to white rice was enough to keep birds healthy. In 1931, the crystals were found to contain sulfur as well as carbon, hydrogen, nitrogen, and oxygen, and the chloride salt was shown to have the empirical formula $C_{12}H_{18}N_4SO_2Cl_2$.

There were, of course, almost innumerable ways in which these atoms could be combined. By good fortune, Robert R. Williams in the USA found that adding sodium sulfite to a solution of the vitamin led to its division into two roughly equal halves. Further work at a number of centers showed that one of these compounds contained a pyrimidine and the other a sulfathiazole ring. By 1937, a synthesis of the active molecule was achieved. It was named 'thiamin or thiamine' (i.e., the sulfur-containing vitamin) and soon began to be produced and marketed as a pharmaceutical.

The Analysis of Foods

Thiamin can be oxidized to a highly fluorescent derivative, 'thiochrome.' This property is used to measure the thiamin contents of different foods, even at levels of less than 1 p.p.m. The procedure is specific, and no other naturally occurring compounds have been found that give the thiochrome reaction. However, thiamin can react with polyphenols and a compound present in garlic to give derivatives that are still biologically active (as will be referred to again) but do not give the thiochrome reaction. The analytical procedure may therefore underestimate the efficacy of a product. It is essential, therefore, to have a confirmatory bioassay before knowing for certain that any kind of processing has caused significant loss of thiamin.

Rice and Other Staples

Figure 2 illustrates the thiamin levels in the world's major staple foods, both when fully milled and when minimally processed. In the case of grains, the latter means removal of the husk (or hull) but no more. It is clear that, for each grain, the full milling that removes both the bran and germ results in the loss of a major portion of the thiamin originally present.

White rice is not that much lower than white wheat flour in its content of the vitamin, but after the grains have been prepared for consumption, the difference is increased. White rice is normally washed several times, and this alone can remove half the thiamin present, and boiling in excess water can again halve the level of remaining vitamin. In contrast, white wheat flour is most commonly baked into bread with yeast as the raising agent, and this causes little loss of thiamin.

There is no evidence that cooked white rice has any positively harmful qualities, but if it is the major item in a diet that contains only small amounts of foods that are richer in thiamin, so that the diet as a whole provides no more than about 0.25 mg per 1000 kcal, it is not surprising that beriberi should gradually develop.

The data in Figure 2 also explain the Japanese experience that serious problems with beriberi in their navy in the late 1800s disappeared when one-half of their rice ration was replaced by barley.

The same figure also shows the low thiamin content of tapioca prepared from cassava roots. This explains the existence of beriberi in Brazil at the same period among even well-off people whose favorite foods were tapioca and molasses. Their preferred protein supplement was dried, salted cod, which had to be soaked for several days to leach out most of the salt, which also removed most of the vitamin.

The very first reports of beriberi to reach Europe came from Portuguese priests working in the Molucca

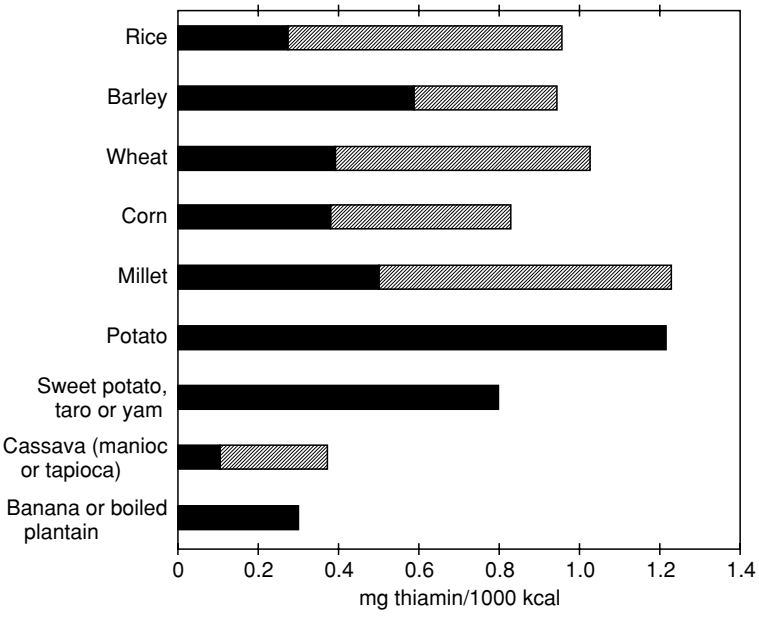

Figure 2 Representative analytical values for the thiamin content of different staple foods: (a) after husking only (■ ▨) and (b) after full processing (■), as explained in the text, and also for some staple root crops, etc. Sago meal is not shown, as it contains only an insignificant level of thiamin. From Carpenter KJ (2000) *Beriberi, White Rice and Vitamin B*. Berkeley, CA: University of California Press, with permission.

Islands (at the Eastern end of Indonesia) in the 1500s. Their staple was the locally produced sago meal, now realized to be almost pure starch, and the priests correctly attributed their weakness to a lack of 'something' in this food and asked to be provided by their superiors with wheat flour.

Beriberi was also a serious problem in early spring in isolated communities in Newfoundland in the early years of the twentieth century. Their families, who would be cut off for the winter, had to buy 6 months of supplies, with white flour as their staple. They were also apparently not familiar with using yeast to leaven bread, but cooked their flour with baking soda, and it is known that much of the thiamin present would be destroyed under the alkaline conditions during this procedure.

The Improvement of Rice

Once the association of beriberi with white rice in Asia had been established, attempts were made to replace it in some way with other foods. As already mentioned, barley was an economic and well-accepted alternative in the diet of the Japanese armed forces.

In the Philippines, a proposal was made to enforce the use of brown rice by the local military. This, of course, is rice from which the husk has been removed but not the entire bran layer and germ (embryo and scutellum) (Figure 3). This was the traditional staple of villagers in South-east Asia who had no access to mechanical rice mills. They would pound their paddy (i.e., rice still in the husk) in some kind of bowl and then winnow the product so that the lighter husks blew away, and the grains fell in a pile.

This procedure was time-consuming but created no problems when only enough was pounded for immediate use in the next 24 h. However, it was repeatedly found that in the tropics, brown rice on storage would become infested with insects of different kinds, and the oil in the bruised germ would become rancid. Since large organizations, or an army on the move, needed large-scale supplies ready for cooking, brown rice did not provide a practicable staple.

The early workers who discovered the association of beriberi with white rice had assumed that the important micronutrient was concentrated in the bran of the grain. However, it was later realized that more was present in the germ area (Table 1). Japanese millers then attempted to modify their machinery so as to remove the bran without removing the germ from the grain. The so-called 'germ rice' that they were able to produce proved to be both palatable and an improved source of thiamin. However, millers were only able to produce it with certain varieties of rice, and not with the bulk of the rice favored in Japan.

A traditional method of processing rice common in Bengal is called parboiling. It had been found that if rice in the husk were to be steeped for a period in hot water and then allowed to dry in the sun, the husks

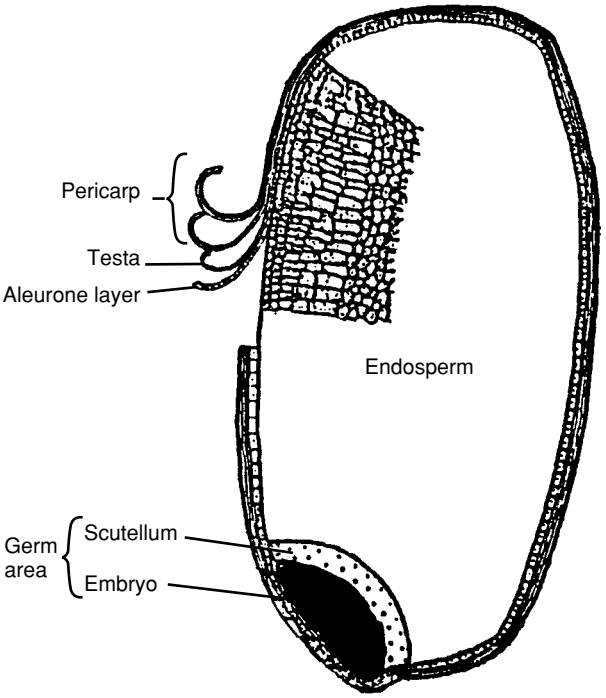

Figure 3 Dissection of a dehusked rice grain. From Carpenter KJ (2000) *Beriberi, White Rice and Vitamin B*. Berkeley, CA: University of California Press, with permission.

Table 1 Thiamin contributed by the different parts of a sample of dehusked rice grains

Dissected parts of rice grain		*Proportion by weight (%)*	*Thiamin*		
			Concentration in fraction (mg per 100 g)	*Contribution to 100 g grain (mg)*	
Pericarp + testa + aleurone layers		6	3.1	0.186	
Germ area	Embryo	1	5.9	0.059	0.248
	Scutellum	1	18.9	0.189	
Endosperm		92	0.05	0.046	
Total		100		0.480	

From Carpenter KJ (2000) *Beriberi, White Rice and Vitamin B*. Berkeley, CA: University of California Press, with permission.

cracked off more easily on pounding, and there was less breakage of the grains. Broken grains had a lower commercial value.

In Malaysia, where many immigrant groups were employed as laborers but living on their habitual diets, it was realized in about 1910 that Bengalis were remarkably free from beriberi, and studies in a mental hospital showed that replacing ordinary white rice with polished rice prepared from parboiled grains relieved the inmates from the disease. After thiamin had been identified, analyses showed that the initial soaking of the grain caused thiamin to diffuse into the endosperm and to remain there as the grains dried out.

Again, this does not appear to be something that could be applied more widely. The traditional soaking and sun-drying leaves the rice with a characteristic musty flavor and slight discoloration, which is acceptable only to people who have grown up with it. The process can be modernized, with controlled autoclaving and vacuum-drying for so-called 'conversion' of the rice, but this makes it too expensive for mass consumption in developing countries.

The final procedure for the production of 'enriched' rice is to fortify it with the synthetic vitamin. For a flour, such as white wheat flour, this is relatively simple, requiring only very careful mixing of the traces of vitamin with a small quantity of flour, then the blending of the premix to larger batches. This is familiar, and indeed compulsory practice, in the USA and UK, together with other vitamins and trace minerals. As a powder, thiamin cannot be blended with grains of white rice. However, methods have been developed of preparing vitamin-rich granules with the size and appearance of rice grains, and blending these in at the ratio of one granule to 200 grains, so that the mix has at least the thiamin content of brown rice. To reduce loss of the vitamin during washing and cooking, the granules are coated with a nontoxic resin at the final stage of their production. This method of enrichment has been tested in an area of the Philippines where beriberi was endemic and has proved successful in greatly reducing its incidence. Unfortunately, where widely separated villages each had a small electric mill, there were practical problems in persuading millers to pay for the premix when the product appeared unchanged, but the price had to be a little higher.

Supplementing Foods

Unfortunately, there is no convenient food that is extremely rich in thiamin. Dried brewers yeast contains 15 mg per 100 g, but many people cannot tolerate it in more than extremely small regular doses. Of the meats, pork is richest with lean pork containing 1 mg per 100 g. Beef has only about one-tenth as much. Dry peas and beans contain about 0.5 mg per 100 g. Potatoes are another useful supplement – on a dry matter basis, they have slightly more than brown rice.

In practice, reaching a desirable intake of thiamin comes usually from eating a wide variety of foods. Beriberi, being restricted almost entirely to the summer months among nineteenth century Japanese, may be explained by their, at that time, regarding most foods other than rice as being 'heating' and therefore to be restricted in that season.

Is there an Antithiamin Problem?

There are many references in the literature to at least a suspicion that certain foods and beverages may be responsible for beriberi appearing in people whose intake of thiamin would otherwise be adequate.

Thiaminases

Many species of fish contain enzymes in their viscera that split thiamin molecules at the junction between its two ring structures. This was discovered when foxes, being reared for their fur and fed on a mix containing a large proportion of whole raw fish, developed a form of paralysis that responded to injections with thiamin. A similar condition was seen later in cats that had been fed on a canned food containing a large proportion of whole fish. It was believed that the thiamin in the mix had been largely destroyed after the mix had been prepared and was waiting to be autoclaved.

These experiences led to investigations as to whether humans could be similarly at risk, but it appears not. The enzymes are not present in fish muscles (i.e., fillets), and even where small fish are eaten whole, they are not ground up with other items of diet before being cooked. Lastly, it was found in animal studies that a subsequent meal with different constituents was not affected by thiaminases being present in an earlier meal.

Another source of thiaminases was found to be bracken, and their presence explained the condition known as 'staggers' that occurs in horses that have been feeding on bracken. Cooked bracken, in which the thiaminase was inactivated, proved harmless to horses. The only known case of thiaminase poisoning in humans occurred in a group exploring the interior of Australia in the 1860s. Running out of provisions on their return journey, they lived on the sporocarps in the fronds of a particular fern that is now known to contain a high level of a particularly heat-resistant thiaminase. All four men developed leg weakness

and lassitude; three died, and the survivor remained lame even after his safe return.

Heat-stable Antithiamins?

It has been found that when the thiamin in a food comes into contact with polyphenols such as caffeic acid, it no longer gives a fluorescent product in the usual thiochrome procedure for the estimation of thiamin. This led some workers to suppose that drinking large quantities of tea or coffee, or chewing betel nut – all sources of polyphenols – might induce a condition of thiamin deficiency. However, biological assays have indicated that the vitamin is still fully available.

When thiamin is incubated with garlic extracts, it undergoes a reaction with the allicin present in which the thiazole ring opens, and the sulfur atom in the ring links to the alkyl sulfide to form a disulfide compound. This is not measured in the thiochrome reaction, but in the body, it is reduced to re-form the active vitamin. In fact, compounds of this type can be absorbed more efficiently by alcohol-damaged intestinal walls than ordinary thiamin. Thiamin tetrahydrofurfuryl disulfide in particular is approved for this purpose in some countries.

Although there is no confirmed evidence of naturally occurring heat-stable compounds that inactivate thiamin, chemists have synthesized such materials. One, named 'oxythiamin,' has the amino group attached to the pyrimidine group in thiamin replaced by a hydroxy group. Giving it to animals results in the more rapid production of some of the signs of thiamin deficiency, although it differs from thiamin in being unable to pass the blood–brain barrier.

Acute Deficiency in the West

With the discovery that autoclaving yeast powder would destroy the thiamin, whereas the other B-vitamins were retained, it was possible to place volunteers on an artificial diet essentially free of thiamin. To the surprise of investigators, some subjects had lost appetite within 2 weeks and became nauseated and dizzy, with other mental symptoms, by 6 weeks, but with no sign of peripheral nerve damage or cardiac abnormality, which are characteristic of classic beriberi.

Trials using pigeons and rats with very deficient diets also produced appetite loss and death before any sign of leg weakness had developed. It appeared that in both humans and animals, acute deficiency of thiamin resulted in damage to the central nervous system. With slightly higher intakes, the CNS had priority, whereas peripheral nerves gradually degenerated.

Deficiencies in Total Parenteral Nutrition

There are many reports of people recovering from surgery of the gastrointestinal tract who have developed acute thiamin deficiency. They had been fed intravenously with a solution providing energy, amino acids, and minerals but no vitamins. This is adequate for a short period, but thiamin is the first vitamin to become depleted.

In a number of cases where this type of parenteral feeding has continued for some weeks, a condition called 'Wernicke's encephalopathy' has developed. Patients are confused and have characteristic involuntary eye movements. Where patients have died, autopsies have shown brain lesions analogous to those found in acutely deficient animals. The same outcome has been seen in subjects voluntarily fasting for long periods or being unable to take food because of persistent vomiting in pregnancy.

Alcoholism

One material that can be responsible for the production of thiamin deficiency is ethanol (i.e., 'alcohol' in everyday speech). In developed countries where nearly everyone can afford a well-balanced diet, most of those diagnosed as being thiamin-deficient are 'alcoholics.' The continued ingestion of high levels of alcoholic beverages has many undesirable effects. In the present context, two are relevant. First, the alcoholic typically no longer bothers to eat a normal range of foods, partly because the beverages provide a large portion of their calorie needs and partly from nothing but their next drink being of immediate interest. Second, the high level of alcohol ingestion damages the intestinal wall so that thiamin is absorbed less efficiently, and the requirement for the vitamin increases.

Unfortunately, a small proportion of such victims develop Wernicke's encephalopathy, which may lead in turn to Korsakoff's psychosis. Such people, sometimes referred to as suffering from the Wernicke–Korsakoff syndrome, are at present incurable and have to be maintained in a mental hospital for the rest of their life. The cost of this to the state is such that some specialists have argued that it would actually be cheaper to have all beer and wine fortified with thiamin as a preventive.

There may be a genetic factor making some Western people susceptible to the cerebral form of beriberi and the Wernicke–Korsakoff syndrome, since it was seen even in Western prisoners of the Japanese in World War II who had white rice but no alcohol, but apparently has not been seen in Asian subjects.

See also: **Alcohol**: Alcohol Consumption; **Cassava**: The Nature of the Tuber; Uses as a Raw Material; **Food Fortification**; **Rice**; **Thiamin**: Properties and Determination; Physiology

Further Reading

Carpenter KJ (2000) *Beriberi, White Rice and Vitamin B*. Berkeley, CA: University of California Press.

Cook CC, Hallwood PM and Thomson AD (1998) B vitamin deficiency and neuropsychiatric syndromes in alcohol misuse. *Alcohol and Alcoholism* 33: 317–336.

Eijkman C (1929) Antineuritic vitamin and beriberi. In: *Nobel Lectures: Physiology or Medicine*, pp. 1922–1941. Amsterdam: Elsevier.

Hinton JJC (1948) The distribution of vitamin B_1 in the rice grain. *British Journal of Nutrition* 2: 237–241.

Jansen BCP and Donath WF (1926) On the isolation of the anti-beriberi vitamin. *Koninklijke Akademie von Wetenschappen, Amsterdam, Proceedings* 29: 1390–1400.

Sauberlich HE, Herman YF, Stevens CO and Herman RH (1979) The thiamin requirement of the adult human. *American Journal of Clinical Nutrition* 32: 2237–2248.

Shimazono N and Katsura E (1965) *Review of Japanese Literature on Beriberi and Thiamine*. Kyoto: Vitamin B Research Committee of Japan.

Takaki K (1906) The preservation of health amongst the personnel of the Japanese army and navy. *Lancet* i: 1369–1374.

Truswell AS and Apeagyei F (1982) Alcohol and cerebral thiamin deficiency. In: Jelliffe EF and Jelliffe DB (eds) *Adverse Effects of Foods*, pp. 253–258. New York: Plenum Press.

Williams RR (1961) *Toward the Conquest of Beriberi*. Cambridge, MA: Harvard University Press.

Williams RR and Cline JK (1936) Synthesis of vitamin B_1. *Journal of the American Chemical Society* 58: 1504–1505.

Wuest HM (1962) The history of thiamine. *Annals of the New York Academy of Sciences* 98: 385–400.

BIFIDOBACTERIA IN FOODS

A L McCartney, University of Reading, Reading, UK

Historical Beginnings

Historically, bifidobacteria have been incorporated into foods for preservation purposes or to combat spoilage organisms. With modern technologies, including sterilization/pasteurization techniques, cold-storage facilities, preservatives, and transportation, more traditional microbiological methods are less necessary. In spite of this, the last decade has seen a minor resurgence in biopreservative applications of bifidobacteria, with extended shelf-life demonstrated for meat and fish products treated with organic acid salts and bifidobacteria. In general, however, modern foods containing bifidobacteria are produced and marketed for prophylactic and/or therapeutic applications, so-called 'well-being' foods or functional foods. The aim of this review is to discuss the relevant issues and current status of foods containing bifidobacteria. As such, we will cover the history of functional foods, the bifidobacterial microflora of humans, potential health benefits of bifidobacteria in humans, considerations for incorporating bifidobacteria in food products, and the regulations/legislation surrounding such products.

The concept that diet impacts the health and well-being of the host animal has long been established. Historically, the ingestion of foods containing lactic acid bacteria (including bifidobacteria) has been commonplace. However, the traditional consumption of fermented milk products largely resulted from the need to reduce putrefaction of foods and preserve them for leaner times. The first scientific contribution to the phenomenon that is now known as 'Functional Foods' was in the early 1900s by Elle Metchnikoff. In his book, 'The Prolongation of Life,' Metchnikoff promoted the ingestion of fermented milks for health and longevity of life. At the same time, work by Tissier demonstrated that there were distinctive bacterial populations harbored by breast- and formula-fed infants. Bifidobacteria were the predominant organisms in fecal samples from breast-fed infants, whereas their formula-fed counterparts harbored a more diverse fecal microbiota in which bifidobacteria were less dominant. Taken together with the general acceptance that breast-fed infants are healthier and better able to resist enteric infections than those fed formulae, the interest in bifidobacteria and methods of enhancing their dominance in the human gastrointestinal (GI) tract was born.

Functional foods are defined as foods that provide benefits to the consumer beyond that of simple nutrition. Much interest has been seen in the potential of

functional foods to modify the gut microflora, both of humans and animals, and in the health benefits of such modulation. Initial work in this area concentrated on *probiotics* (live microbial feed supplements that beneficially affect the host animal by improving its intestinal microbial balance). Lactic acid bacteria (LAB), especially lactobacillus and bifidobacteria, have accordingly received enormous attention from both the scientific community and the food industry. More recent developments in the area of functional foods have targeted *prebiotics* (nondigestible food ingredients that beneficially affect the host by selectively stimulating the growth and/or activity of one or a limited number of bacteria in the colon). Such food components are considered to be preferable to probiotics, since they lack stability problems, both within the food product and during transit through the upper GI tract, and they are aimed at enhancing the indigenous LAB. However, this is beyond the scope of this review, and we will concentrate on bifidobacteria in foods rather than bifidogenic foods (foods containing bifidobacteria and/or compounds that enhance bifidobacteria). Additionally, there are some instances where probiotics are advantageous, owing to species- and/or strain-specific properties and/or limited indigenous LAB.

Bifidobacteria in the Human Colon

The normal colonic microflora of humans is extremely large and complex. The composition of this bacterial population is affected by host-mediated factors, microbial factors, microbial interactions, and environmental factors. Examples of each of these factors are listed in Table 1. Bifidobacterial levels fluctuate in humans throughout their lifespan (although they are relatively stable during adult life), with the largest numbers usually seen in exclusively milk-fed infants and reduced levels seen in the elderly. Compromised individuals can also harbor low bifidobacterial populations. Indeed, extensive antibiotic treatment and chemotherapy have been demonstrated to markedly reduce the bifidobacteria levels of humans. Low bifidobacterial levels often correspond to greater susceptibility to enteric infections and outbreaks of diarrhea.

Feeding bifidobacteria to humans (either in food products or as freeze-dried preparations) has most often led to recovery of the administered strain in intestinal or fecal samples. The ingested probiotic is usually lost from the intestinal flora relatively soon after administration ceases (often within 4–7 days). When bifidobacteria are administered to healthy humans, whose microflora is undisturbed, the effect on the bacterial population levels is usually low. However, metabolic changes have been observed in some investigations of bifidobacterial feeding in healthy humans.

Table 1 Examples of factors affecting the composition of the normal microflora of humans

Type	*Examples*
Host-mediated	Gastric secretions
	Gastrointestinal motility
	Host physiology
Microbial characteristics	Ability to adhere
	Nutritional flexibility
	Generation time
Microbial interations	Synergistic
	Antagonistic
Environmental	Diet
	Stress

Potential Health Claims of Bifidobacteria

A number of health benefits have been attributed to LAB, including: colonization resistance, stimulation of host immune function, anticarcinogenic activity, lowering cholesterol levels, alleviation of lactose intolerance, vitamin synthesis, and re-establishing a balanced intestinal flora. These are discussed in more detail below. However, the scientific evidence for some of these claims is weak and/or contradictory at best. Often, inappropriate animal models or poorly designed *in vitro* and *in vivo* studies have been employed. In the following pages, we will endeavour to summarize the current knowledge on the benefits of bifidobacteria to human health. Table 2 lists a number of human studies investigating various effects of oral administration of bifidobacteria.

Colonization Resistance

One of the most common claims associated with dairy foods containing bifidobacteria, and other LAB, is the 'maintenance or re-establishment of healthy intestinal microflora.' The normal colonic flora provides an important barrier function against pathogenesis, often termed 'colonization resistance.' Multiple mechanisms may be involved in the exclusion of undesirable organisms by bifidobacteria, including competition for receptor sites and/or nutrients and production of inhibitory factors and/or conditions [e.g., organic acids or antimicrobials, as well as physiological factors (lowering of pH and/or stimulating the immune system)]. Perhaps the most compelling evidence of the therapeutic effect of bifidobacterial food products is seen during dietary intervention studies of individuals with a disturbed colonic flora. The best examples of such data have

Table 2 Compilation of experimental results showing the effect of feeding bifidobacteria in humans

Organism	*Control*	*Test*	*Effect*
Bifidobacterium bifidum, Lactobacillus acidophilus	15 humans receiving no fermented milk	15 humans Fermented milk	Increased total IgA levels
B. bifidum, L. acidophilus	Baseline[a]	14 humans	Increased phagocytic activity
B. bifidum, Streptococcus thermophilus	Placebo	55 hospitalized infants	Reduced incidence of diarrhea (70% of observed diarrhea cases caused by rotatvirus)
Bifidobacterium spp.	Baseline	Humans Fermented milk	Lowered β-glucuronidase levels
B. bifidum, L. acidophilus, S. lactis, S. cremoris	Baseline	Humans Fermented milk	Lowered nitroreductase levels
B. longum	Baseline	10 humans Fermented milk	Lower fecal pH
B. longum, L. acidophilus	Placebo	30 humans Fermented milk	Lowered fecal volatile fatty acids levels and reduced GI discomfort
B. longum	Baseline	6 humans Fermented milk	Increased bifidobacterial dominance
B. longum	Placebo	12 humans Fermented milk	Increased fecal bifidobacteria levels
B. bifidum, L. acidophilus	Placebo	Humans Lyophilized bacteria	Delayed reduction of bifidobacteria due to antibiotic treatment and induced earlier re-establishment of bifidobacteria, postantibiotic therapy
B. bifidum	Baseline	12 elderly dyspepsia patients Lyophilized bacteria	Increased fecal bifidobacteria levels and reduced symptoms
B. bifidum, L. acidophilus	Placebo	20 humans Lyophilized bacteria	No significant differences during amplicillin administration
B. bifidum	Placebo	15 humans Lyophilized bacteria	Decreased chronic inflammation of sigmoid colon and increased humoral immunity
B. longum BB536	*B. longum* ATCC15707	48 Humans Fermented milk	Reduced clostridia, bacteroides, coliforms population levels
Bifidobacterium spp.	Pasteurized bifidus-fermented milk	60 Humans Fermented milk	Reduced colonic transit time
Bifidobacterium spp.	Baseline	Patients with hepatitis or liver cirrhosis	Lowered blood ammonia and phenol levels, reduced urinary indican levels, increased fecal bifidobacteria levels and improved appetite and weight gain
B. lactis	13 control subjects given placebo milk	12 healthy elderly humans Supplemented milk	Increased interferon-α levels and polymorphonuclear cell phagocytic capacity
B. longum	Baseline	5 humans Lyophilized bacteria	Reduced lecithinase negative clostridia numbers, lowered NH_3 and β-glucuronidase levels in feces and reduced fecal pH
Bifidobacterium spp.	Baseline	Bedridden elderly humans Fermented milk	Increased stool frequency
B. bifidum, L. acidophilus	10 elderly humans	15 elderly humans	Reduced inflammation of sigmoid and descending colon
Bifidobacterium spp.	Baseline	15 humans	Reduced stool frequency in intractable diarrhea
B. longum	Yogurt without bifidobacteria	10 Humans Fermented milk	Reduced erythromycin induced gastrointestinal effects
Bifidobacterium spp.	Baseline	Humans Fermented milk	Lowered nitroreductase and glucuronidase levels

[a]Same subjects prior to feeding.

come from investigations of the antidiarrheal effect of bifidobacteria. Indeed, the biotherapeutic potential of foods containing bifidobacteria (and other LAB) has been well established for antibiotic-associated diarrhea outbreaks and rotavirus infections. In a double-blind, placebo-controlled trial in 55 hospitalized infants, oral administration of *Bifidobacterium bifidum* and *Streptococcus thermophilus* reduced the incidence and severity of rotavirus diarrhea. There is also some evidence that fermented milk products containing bifidobacteria may protect against traveler's diarrhea. However, work to date suggests that the

probiotic strain/mixture as well as the destination are paramount.

Physiological Effects

The production of organic acids (acetic and lactic acid) by bifidobacteria inhibits the growth of pathogenic organisms (directly and indirectly) and stimulates intestinal peristalsis. Acetic acid is a stronger antagonist against Gram-negative bacteria than lactic acid. As such, the potential applications of bifidobacteria against microbial perturbation may surpass those of lactobacilli. Additionally, the organic acids produced by bifidobacteria have been shown to inhibit the growth of many nitrate-reducing bacteria. Human studies using bifidobacteria as a dietary adjunct have shown the relief of symptoms in constipated elderly persons (Table 2). Several investigations have also reported a reduction of fecal pH in individuals consuming fermented milk products containing bifidobacteria. Lowering the pH of the intestinal tract produces an environment less favorable to some pathogenic bacteria and thus helps prevent their overgrowth.

Anticarcinogenic Activity

Increasing evidence, from *in vitro* experiments and animal studies, indicates the potential protective influence of probiotic bacteria (including bifidobacteria) against cancer. To date, three plausible mechanisms have been identified: (1) inhibition of putrefactive organisms that produce carcinogens (such as *N*-nitroso compounds, phenolic products of tyrosine and tryptophan, and metabolites of biliary steroids); (2) binding and/or inactivation of carcinogens; and (3) inhibition of tumor cell formation. Bacterial enzymes (including β-glucuronidase, β-glucosidase, nitroreductase, and azoreductase) are responsible for converting some procarcinogens into carcinogens. As such, the levels/activities of these enzymes are considered a useful biomarker for cancer risk in humans, enabling noninvasive estimation of carcinogen levels. Several human studies have employed this method to determine the effects of different diets on the risk of colon cancer. A number of investigations have shown a reduction in fecal enzyme activity during ingestion of bifidobacteria-containing milk products, compared with initial levels before bifidobacterial consumption (i.e., baseline levels). Although these results are encouraging, they are inconclusive regarding the effects on cancer risk. Much more research is required to establish clear links between enzyme activity and cancer risk. Animal studies have enabled more definitive investigations of the anticancer effects of bifidobacterial feeding. Several such reports from murine models have identified an antitumor effect of bifidobacterial feeding (with or without prebiotics). Cases affording protection against both aberrant crypts (precancerous lesions) and tumors have been published. However, animal studies are speculative at best, regarding the protective nature of these strains against the development of cancer in humans.

Stimulation of Immune Functions

Several probiotics, including some bifidobacterial strains, are claimed to enhance the immune system in a nonspecific manner, thereby stimulating immunity to a number of antigens. A number of studies have also shown the ability of certain LAB strains to alter cytokine production and/or increase secretory IgA levels. However, most work in this area has either involved inappropriate animal models or concentrated on intermediary endpoints (such as cytokine production) rather than disease symptoms. Poor study designs have also led to difficulty in confirming the immunomodulatory agent. Preliminary data demonstrate the potential of probiotics in modulating certain immune responses and indicate their potential role in allergy, autoimmunity, and gastrointestinal disease.

Cholesterol-lowering Ability

Mann and Spoerry identified correlations between daily consumption of fermented milk and low serum cholesterol levels in East African Masai warriors. Subsequent *in vitro* work has demonstrated the ability of bifidobacteria to both assimilate cholesterol and coprecipitate it with deconjugated bile acids. Such observations have led to great interest in the cholesterol-lowering capacity of a diet containing fermented milks. However, much contradictory data exist regarding the effects of consuming foods containing bifidobacteria on serum cholesterol levels. Confounding the issue has been the use of different strains, dosages, and/or food vehicles in the various studies carried out so far. Additional criticisms of the current data have included lack of stabilization of baseline cholesterol levels, inadequate size and/or duration of studies, and difficulty in controlling the diet and physical activity of subjects.

Lactose Intolerance

Lactose malabsorption affects large portions of the population (estimated by some to affect over half the world's population), with a higher prevalence in those of Oriental or African ancestry. Symptoms normally include abdominal discomfort, flatulence, and/or diarrhea. However, lactose-intolerant individuals

can consume cultured milk products (containing bifidobacteria and/or lactobacilli) without any deleterious effects. Two mechanisms have been proposed for the improved digestibility of lactose in such products: (1) the β-galactosidase activity of the probiotic strains; and (2) stimulation of host mucosal β-galactosidase activity by the ingested strains.

Nutritional Value

Bifidobacteria are known to produce thiamine, riboflavin, vitamin B_6, and vitamin K. There have also been reports of their ability to synthesize folic acid, niacin, and pyridoxine. These vitamin B complexes are slowly absorbed in the human body. However, the impact on human nutrition of such vitamin synthesis by bifidobacteria in the colon is unknown. Available information on the nutritional properties of fermented milks containing bifidobacteria indicates that they have lower residual lactose and higher levels of free amino acids and vitamins than nonfermented milks. Additionally, they preferentially contain L(+)-lactic acid [produced by bifidobacteria in addition to acetic acid, whereas lactobacilli produce D/L(−)-lactic acid], which is more easily metabolized by humans. This is particularly important for infants less than 1 year old, in whom metabolic acidosis can be a problem. Consuming bifidobacterial food products may also improve the bioavailability of certain minerals, including calcium, zinc, and iron, by lowering the gastric pH (facilitating ionization of minerals, which is necessary for their uptake).

Product Development

Technically, bifidobacteria may be incorporated into foods either to produce certain product characteristics, such as organoleptic and/or nutritional properties, or as a probiotic. In the latter case, the viability of the strain is essential during both manufacture and storage of the food product. Indeed, the generally accepted recommendation is that probiotic strains must be present at levels of $> 10^6$ viable cells per milliliter at the time of consumption. This is based on the minimum therapeutic daily dose of 10^8–10^9 cells.

Bifidobacterial cultures incorporated into foods for therapeutic purposes may be added either during fermentation (along with or as part of the starter culture) or to the finished fermented or fresh product prior to shipment. Storage of food products under the appropriate refrigeration conditions after manufacturing restricts subsequent fermentation. As such, products to which bifidobacteria are added immediately prior to distribution are organoleptically indistinguishable from the unfortified version. Products fermented with bifidobacteria, however, often have a mild acidic flavor. The selection of culture strains and ratios is as critical in determining the final product characteristics as the fermentation parameters (temperature, duration, chemical composition and level of inoculum).

Selection Criterion

The selection of bacterial strains is vital since different strains (even within the same species) have distinctive metabolic and probiotic characteristics, as well as different growth rates. The main factors to consider include rate of acid production, type of polysaccharide fermentation, ability to synthesize vitamins, proteolytic characteristics (which may result in bitter compounds) and capacity to produce flavor compounds. An additional consideration is the growth characteristics of the bacterial strains. This is particularly important regarding mixed inocula and determining both the optimal mixture and the ratio of strains, as well as the appropriate time in processing to add the organisms. Recent studies have shown that bifidobacteria inoculated into sour cream and buttermilk either at the time of setting or at the time of breaking allowed appropriate numbers of bacteria to survive during normal storage of the products. Additional work suggests that the best procedure for incorporating bifidobacteria in cottage cheese is in the creaming mixture, whereas addition immediately after cooking (prior to hooping) is suitable for Edam cheeses. In the production of yogurts containing bifdiobacteria, it is common practice to use premixed cultures of *Streptococcus thermophilus, Lactobacillus delbrueckii* ssp. *bulgaricus, L. acidophilus*, and *Bifidobacterium* species. However, bacterial 'loading' is also practiced, where the lactobacilli and bifidobacteria are grown separately prior to addition to the product. A number of studies have also demonstrated successful manufature of probiotic cheeses, especially hard cheeses (e.g. cheddar, Canestrato Pug liese and Gouda-type), by either incorporating bifidobacterial strains in the starter culture or using bifidobacterial fermented cream.

As bifidobacteria are anaerobes, it is useful to select species or strains that are less oxygen-sensitive (such as *B. infantis* and *B. longum*), perform gas flushing with nitrogen (a process often used for preserving yogurts), or include an oxygen scavenger in the fermentation inocula (such as *S. thermophilus*). The growth and survival characteristics depend on the pH (and its buffering capacity), the fermentation and storage conditions (especially temperature and duration), and the presence of competing microorganisms or microbial inhibitors (such as H_2O_2 and NaCl).

Regarding biotherapeutic considerations, the strain(s) must be selected, based on their ability to elicit the desired beneficial effect at the target. This necessitates their survivability within the product and during transit through the upper GI tract. As such, it is pertinent to use bifidobacteria of human intestinal origin as they are better suited to the physiological needs of the host than strains from other sources. Finally, it is essential to demonstrate the safety of probiotic strains. They must be confirmed as non-pathogenic, have a well-established safety record, or have 'generally regarded as safe' (GRAS) status. Many publications have listed recommendations of selection criteria for probiotic strains. These fall into the four general categories of (1) appropriateness, (2) technological suitability, (3) competitiveness, and (4) performance and functionality.

Technical Considerations

Besides strain selection, careful monitoring throughout the manufacturing process is required to insure efficient control of the final product (pH, metabolic products, cell density, and flavor). Incorporation of LAB above the necessary starter organism to produce specific product characteristics demands examination of the final product to confirm the desired property. For example, *in vitro* demonstration of a strain's ability to synthesize vitamins is inadequate on its own, as the fermentation characteristics and final product composition (including the levels of the desired vitamins) need to be determined.

A number of methods have been reported to increase the survival of LAB whilst maintaining the desirable attributes of yogurt. These include enrichment with whey protein concentrate; lowering incubation temperature (to 37 °C); terminating fermentation at a higher pH (e.g., > 5); hydrostatic pressure treatment or heat-shock of yogurt (to prevent postmanufacture acidification); and storing below 3–4 °C.

Market for Foods Containing Bifidobacteria

To date, food products containing bifidobacteria have largely been of dairy origin and include yogurts, fermented milks, fresh milks, cheeses, buttermilk, sour cream, cream cheeses, cottage cheeses, frozen dairy desserts (including icecream and frozen yogurts), and dips. Research has further demonstrated the feasibility to incorporate bifidobacteria in vegetable spreads, fermented vegetable products (such as kimchi, a traditional Korean food), mayonnaise, cookies, and some beverages. A list of some of the commercially available fermented milk products containing bifidobacteria and some of their claims is provided in Table 3.

The increased awareness of consumers towards general health and well-being, together with the accepted impact of diet on health, has led to a greater demand for healthy and/or functional food products. As such, the upward trend currently seen in bifidobacterial foods is projected to continue. As much as 70% of milk products on the market in some European countries, such as Sweden, contain bifidobacteria. The Japanese and European food industry and markets are well ahead of the USA and Canada, regarding functional foods. However, the influx of foods containing bifidobacteria on these markets has begun. A recent article estimated double-digit annual growth rates of the world-wide market for such products.

Regulations and Legislation

The generally accepted minimum level of probiotics in foods at the time of consumption is $> 10^6$ cells per milliliter. However, the standard developed by the Fermented Milks and Lactic Acid Bacteria Beverages Association of Japan is $> 10^7$ cells per gram at the time of manufacture. In the USA, 10^8 bacteria per gram at the time of manufacture is the criterion of the National Yogurt Association (NYA). As well as the concern of viability of probiotics within the consumed product, there is increasing interest (from the scientific community, food industry, and national authorities) in establishing codes of practice regarding the legislation and regulations of health claims for food products. To this end, definitive evidence demonstrating the efficacy of products is essential. As such, statically significant, well-designed (double-blind, placebo-controlled) clinical trials are demanded for health-promoting food products. The need to prove safety is equally important to substantiating efficacy. Key areas for demonstrating nutritional safety include intake, extent of use, implications/impact on gut microflora and metabolic pathways, compositional analysis (including the presence of antinutritional or toxic factors), and potential effects in compromised individuals or specific target populations.

Currently, there is no world-wide regulatory or legislative control on the labeling and health claims of food products. However, most legislative bodies prohibit the use of 'medicinal' claims for food products (including functional foods). That is, claims suggesting the prevention of, treatment, and/or cure for human disease are not permissible for food products. However, health claims are deemed acceptable and usually involve the promotion of health benefits,

Table 3 Commercially available fermented milk products containing bifidobacteria

Product	*Bacterial strains*	*Claims*	*Country*
A-38	*L. acidophilus, B. bifidum, Leuconostoc mesenteroides* ssp. *cremoris*, mesophilic lactococci		Denmark
A-B Yogurt	*B. bifidum, L. acidophilus*		France
Aktifit	*S. thermophilus, L. casei, L. acidophilus, B. bifidum*	Defense system for bowel	Switzerland
Akult	*B. bifidum, L. acidophilus, B. breve, L. casei* ssp. *casei*	Normalize the balance of human intestinal flora	Japan
B-Active	*L. acidophilus, B. bifidum, S. thermophilus, L. delbrueckii* ssp. *bulgaricus*		France
Bifidus milk	*B. bifidum, B. longum*		Several
Bifiel	*B. breve* strain Yakult, *S. thermophilus, Lactococcus lactis*	Normalize the balance of human intestinal flora	Japan
Bifighurt	*B. bifidum, S. thermophilus*		Germany
Bifilus	*B. longum, L. bulgaricus, S. thermophilus*		Sweden
Biokys	*B. bifidum, L. acidophilus, Pediococcus acidilactici*		Czech Republic
Biomilk	*Bifidobacterium* spp., *L. acidophilus*		Several
Bio Pot	*L. acidophilus, B. bifidum*	Improve intestinal flora	Germany
Bio-Garde	*L. acidophilus, B. bifidum, S. thermophilus*		Australia
Biola	*B. lactis, L. acidophilus, L*. GG	Balance microflora	Norway
Bulla AB Live	*L. acidophilus, Bifidus*		Singapore, Australia, Indonesia
GEFILUS	*S. thermophilus, L. bulgaricus, L. acidophilus, B. bifidum, L. casei* GG		Finland
LA7	*L. acidophilus* LA7, *Bifidus*	Improve intestinal flora	Germany
Leisure Live	*L. acidophilus, B. bifidum*		UK
Little Swallow	*L. acidophilus, B. bifidum*		UK, Germany
Mil-Mil	*B. breve* strain Yakult, *B. bifidum* strain Yakult, *L. acidophilus*	Normalize the balance of human intestinal flora	Japan
Mil-Mil E	*B. breve* strain Yakult, *L. acidophilus, S. thermophilus*	Normalize the balance of human intestinal flora	Japan
Morinaga Bifidus	*B. longum, L. acidophilus*	Normalize the balance of human intestinal flora	Japan
Morinaga Caldus	*B. longum, L. acidophilus*	Normalize the balance of human intestinal flora	Japan
Nu-trish A/B	*B. bifidum, L. acidophilus*		USA
Procult 3	*B. longum*		Germany
Progurt	*B. bifidum, L. acidophilus*, mesophilic lactococci		Several
Ski	*L. acidophilus, Bifidus*	Promote health	Australia, Singapore
Sym-Balance	*L. reuteri, L. acidophilus, L. casei*, bifidobacteria	Defense system for bowel	Switzerland
VAALIA	*L. acidophilus, B. lactis, L*. GG	Promote health, balance intestinal microflora	Australia
Vifit	*S. thermophilus, L. bulgaricus, L. acidophilus, B. bifidum, L. casei* GG	Keep the intestinal flora in good shape, enhance immune response	Netherlands, Belgium, UK
Vifit	*S. thermophilus, L. bulgaricus, L. acidophilus, B. bifidum, L. casei* GG	Keep the intestinal flora in good shape, enhance the immune response	Germany
Yoplait Yoplus	*L. acidophilus, Bifidus*		Singapore, Australia
Yukijirushi Nachure	*L. acidophilus, B. longum*	Promote the maintenance of a good intestinal environment	Japan

reduction of disease risk, or general well-being of the host. Health claims may not include specific claims such as 'improves resistance against pathogenic infection' but could state 'improves normal intestinal function.' As one can see, the line between medical and health claims is very gray, and the presentation or wording of the claims is usually important in defining under which heading they fall. Indeed, one published explanation of medical foods (medicine) and functional foods gave the following definitions. Medicines include 'a food product which has a direct preventative effect on a disease,' whereas health claims may

include maintenance of 'good health as part of a life style' and/or potential reduction of 'the risk of disease occurring.'

In Japan, a voluntary system is in place where industry may apply for 'Foods for Specified Health Use' (FOSHU) approval. Information on the safety and efficacy in humans (including definitive scientific evidence) and nutritional analyses are required for such approval. FOSHU labeling, therefore, effectively demonstrates that the product is approved as a functional food, according to the advertised claim.

In the USA, Federal Drug Agency (FDA) approval is necessary regarding any labeling of health claims on food products. Furthermore, the claims must be founded on an 'authoritative statement' from an appropriate Government scientific agency responsible for 'public health protection or research.' However, 'structure/function' statements are permitted on dietary supplements relating the food ingredient to 'normal, healthy functioning of the human body.' Such claims are not normally used on conventional foods, but rather for foods intended to supplement the diet. That said, foods containing bifidobacteria are generally marketed as dietary supplements in the USA, and therefore fall within a loophole. Currently, however, most manufacturers of such products simply name the added bacteria (at least to genus level) for labeling purposes and do not state any functions or claims.

In Europe, there is no collective regulation or legislation regarding functional foods. In general, the primary focus is on the safety of the products. Food products or food ingredients that were not available on the European market for human consumption prior to April 1997, and can be classified as novel (owing to raw materials, composition, or manufacturing process), must undergo regulatory approval. The second regulatory focus in Europe is the labeling and health claims of the food product. The general rule of thumb is that health claims may not attribute the property of preventing, treating, or curing a human disease, or any such implications. Currently, seven European Union member states are developing their own Codes of Practice regarding legislation and regulation of health claims and labeling of functional foods. The level of scientific substantiation needed for such health claims differs between some codes. However, the Codes of Practice for the UK, Belgium, and The Netherlands all include principles for human (clinical) trials.

Summary

Results to date clearly indicate the potential benefits of consuming a diet incorporating foods containing bifidobacteria. Several food vehicles have been identified within the dairy industry, and appropriate manufacturing techniques are available to insure that the desired product characteristics and viability of probiotic strains are attained. Essential to the future of functional foods are adequate studies confirming the safety, efficacy, and viability of such products (especially well-designed human clinical trials). Current developments within the scientific community, food industry, and regulatory bodies are all pursuing this end.

See also: **Dairy Products – Nutritional Contribution**; **Fermented Milks**: Types of Fermented Milks; Products from Northern Europe; Other Relevant Products; Dietary Importance; **Lactic Acid Bacteria**; **Microbiology**: Detection of Foodborne Pathogens and their Toxins; **Microflora of the Intestine**: Role and Effects; Probiotics; **Prebiotics**; **Probiotics**; **Yogurt**: The Product and its Manufacture; Yogurt-based Products; Dietary Importance

Further Reading

American Society for Nutritional Sciences (various authors) (2000) Symposium: Probiotic bacteria: Implications for human health. *Journal of Nutrition* 130: 382S–416S.

Arunachalam KD (1999) Role of bifidobacteria in nutrition, medicine and technology. *Nutrition Research* 19: 1559–1597.

de Roos NM and Katan MB (2000) Effects of probiotic bacteria on diarrhea, lipid metabolism, and carcinogenesis: a review of papers published between 1988 and 1998. *American Journal of Clinical Nutrition* 71: 405–411.

Gibson GR and Williams CA (2000) *Functional Foods: Concept to Product*. Cambridge: Woodhead Publishing.

Heller KJ (2001) Probiotic bacteria in fermented foods: product characteristics and starter organisms. *American Journal of Clinical Nutrition* 73(Suppl): 374S–379S.

Kailasapathy K and Rybka S (1997) *L. acidophilus* and *Bifidobacterium* spp. – their therapeutic potential and survival in yogurt. *The Australian Journal of Dairy Technology* 52: 28–34.

Lee Yuan-Kun Nomoto K, Salminen S and Gorbach SL (1999) *Handbook of Probiotics. Wiley-interscience Publication*. New York: Wiley.

Rasic JL and Kurmann JA (1983) *Bifidobacteria and Their Role. Experientia Supplementum*, vol. 39. Basel: Birkhauser.

Sanders ME and in't Veld JH (1999) Bringing a probiotic-containing functional food to the market: microbiological, product, regulatory and labeling issues. *Antonie van Leeuwehoek* 76: 293–315.

Tannock GW (1999) *Probiotics: A Critical Review*. Wymondham, Norfolk, UK: Horizon Scientific Press.

BILE

A Lanzini, University of Brescia, Brescia, Italy

Background

Bile is a complex aqueous solution of organic and inorganic compounds secreted by the liver. It contains bile acids, a class of detergent-like molecules that exert their biological functions as lipid solubilizers within the biliary tree and the gut lumen. The bile acids are restricted to the enterohepatic circulation, so as to reutilize many times these biologically valuable molecules. This chapter reviews the composition and physicochemical properties of bile, the mechanism involved in bile formation, and the function of bile acids as solubilizers of biliary and dietary lipids.

Composition of Bile

Inorganic compounds in human hepatic bile comprise electrolytes in concentrations similar to those in plasma, with the noticeable exception of HCO_3^- concentration which is higher in bile. The osmolarity of human hepatic bile is also similar to plasma osmolarity, at about 300 mOsm l^{-1}.

Organic compounds in bile comprise protein and bile pigments in addition to the three biliary lipids, bile acid, cholesterol, and phospholipid. Proteins account for 4.5% of organic compounds in bile, and their concentration ranges between 0.3 and 3.0 g ml^{-1} in typical human bile. The most abundant biliary protein is albumin derived from the plasma pool. In general, the biliary concentration of other plasma proteins is inversely related to their molecular weight. Bile also contains immunoglobulin A, lysosomal enzymes, and plasma membrane enzymes. The secretion of these latter proteins is influenced by the bile acid secretion rate.

Bile pigments constitute less than 0.3% of organic compounds in bile, and their concentrations range between 0.8 and 3.2 mmol l^{-1} in human hepatic and gallbladder bile, respectively. Conjugation with glucuronic acid is essential for biliary bilirubin excretion, and bilirubin diglucuronide is the major pigment in human bile. The maximal bilirubin excretion capacity depends on bile flow and is increased during enhanced bile acid secretion rates. About 30% of bilirubin excretion is thought to be bile acid-independent, suggesting that incorporation of bilirubin into mixed micelles is not essential for its biliary excretion.

Many other endogenous substances are present in human bile in addition to those listed above. These include vitamins (mainly D_2, B_{12}, folic acid), steroid (estrogens), and thyroid hormones. Exogenous compounds such as contrast media, some antibiotics (ampicillin, metronidazole), cardiac glycosides, and opiates may also be excreted in bile and undergo some degree of enterohepatic circulation.

Biliary Lipid Chemistry and Physicochemical Properties

Bile Acids

Bile acids comprise a class of molecules derived from the hepatic catabolism of cholesterol. More than 200 bile acids have been isolated in human bile, and 92–99% of total bile acids are constituted by mono-, di-, and tri-hydroxy derivatives of a 24-carbon atom steroid, cholanoic acid (Figure 1). The brief and systematic names and the relative proportions of individual bile acids in human bile are shown in Table 1.

Virtually all biliary bile acid are amidated in man almost exclusively with glycine or taurine at a ratio of about 2:1. Sulfation on the steroid nucleus of bile acids occurs to a significant extent in man only for lithocholic acid, and glucuronidation is a trace metabolic pathway in man.

From a physicochemical point of view, bile acids are planar amphiphiles in that they exhibit both hydrophilicity with one part of the molecule and hydrophobicity with the remainder. In commonly occurring bile acid molecules, the hydrophobic face is constituted by the β side of the steroid nucleus, and the hydrophilic face by the α side of the nucleus and by the side chain. The relative potency of the hydrophilic and hydrophobic functional groups in affecting the physicochemical properties of bile acids is referred to as the hydrophilic–hydrophobic balance. The order of increasing hydrophilicity, as assessed by high-pressure liquid chromatography, follows the order taurine conjugates > glycine conjugates > unconjugated bile acid; and trihydroxylated > dihydroxylated>monohydroxylated bile acids.

Bile acid solubility is strongly dependent on amidation. Unconjugated bile acids (pK_a 5.0) are insoluble at a pH below 6–7, and glycine-(pK_a 3.8) and taurine-conjugates (pK_a < 1.0) are soluble at pH 4–5 and below 2, respectively. Fully ionized common bile acids are present in physiological solutions as their sodium salts. They are extremely water-soluble, with a monomeric solubility of 1–3 mmol l^{-1} (Figure 2).

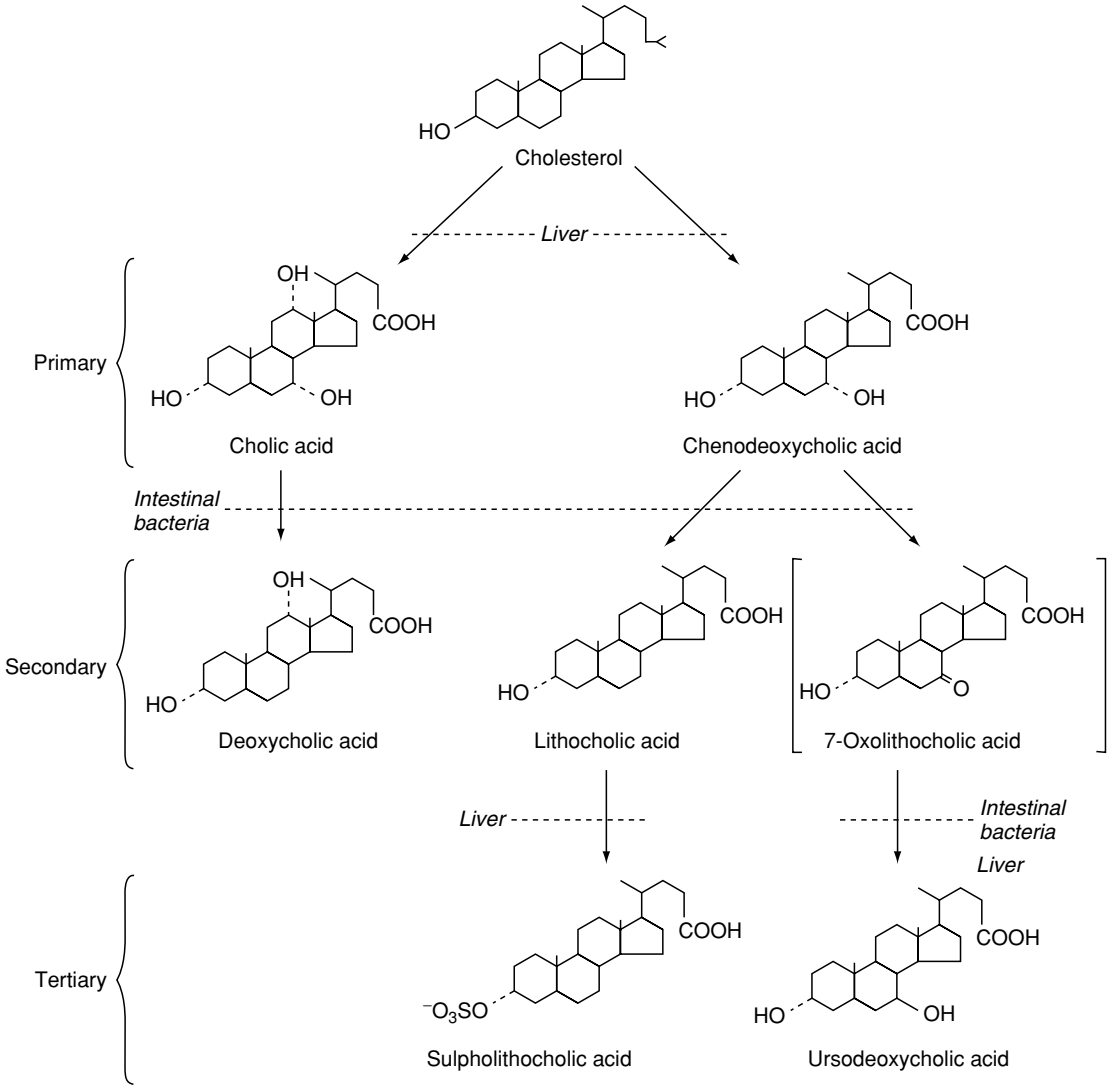

Figure 1 Structural formulae and sites of synthesis and metabolism of common bile acids in man. From Carey MC and Duane WC (1994) Enterohepatic circulation. In: Arias IM, Boyer JL, Fausto N *et al*. (eds.) *The Liver: Biology and Pathobiology*, 3rd edn, pp. 719–767. New York: Raven Press, with permission.

They are present in diluted water solutions as monomers in the bulk water phase, and as an unstable film on the water surface. Upon concentration, a critical monomeric concentration is reached, called the critical micellar concentration (CMC), at which bile acid monomers self associate to form multimolecular aggregates called micelles. Micelle formation involves back-to-back agglomeration of the hydrophobic side of the bile acid molecules, with the hydrophilic side facing water. At concentrations above the CMC, bile-salt micelles have the capacity of incorporating and solubilizing otherwise insoluble compounds, thus acting as detergents. In general terms, hydrophobic bile acids form micelles at lower concentration than hydrophilic bile acids, and the CMC ranges between 0.5 and 11 mmol l^{-1} for common bile acid (Table 1).

Phospholipids

The most abundant biliary phospholipid species are phosphatidylcholines (lecithins), which account for 80–95% of phospholipids in human bile. Phosphatidylcholines are insoluble swelling amphiphiles, with a monomeric solubility in water of about 1 pmol l^{-1} (Figure 2). Upon hydration (> 45% water), phospholipids swell to form liquid crystalline phases consisting of choline bilayers with interposed water layers. These liquid crystalline phases may fold and aggregate to form vesicular structures.

Cholesterol

From a physicochemical point of view, cholesterol is an insoluble nonswelling amphiphile, with a monomeric solubility in water of about 1 nmol l^{-1} at 37 °C (Figure 2). Upon hydration, crystals of cholesterol monohydrate are in equilibrium with water-containing monomers of cholesterol monohydrate.

Table 1 Name and composition of biliary bile acids of typical human bile

Brief name of bile acid	*Systematic name*	*CMC (mmol l^{-1})*	*Percentage in bile*[a]
Cholic	3α,7α,12α-Trihydroxy-5-cholanoic acid	11	35
Chenodeoxycholic	3α,7α-Dihydroxy-5-cholanoic acid	4	35
Deoxycolic	3α,12α-Dihydroxy-5-cholanoic acid	3	24
Ursodeoxycholic	3α,7β-Dihydroxy-5-cholanoic acid	7	tr.–4
Lithocolic	3α-Monohydroxy-5-cholanoic acid	0.5	tr.–3

[a]Individual bile acids as a percentage of total bile acids. CMC, critical micellar concentration; tr, trace.

Bile Formation

Bile formation involves a number of secretory processes in the hepatocyte and in the ductular cells generating an osmotic gradient between the cells and the biliary canaliculus and ductulus, with passive water movement along this gradient until osmotic equilibration is achieved. Canalicular and ductular bile flow contribute 470 and 150 ml, respectively to the typical total daily bile flow of 620 ml in man. Bile flow varies from 1.5 to 15.4 μl min^{-1} kg^{-1}, and is linearly related to bile acid secretion rate (Figure 3). The contribution of canalicular to total bile flow is defined by the linear relationship between bile acid secretion and biliary clearance of sugars (erythritol or mannitol) acting as bile-flow markers. This relationship identifies the so-called bile acid-dependent canalicular bile flow (about 60% of total canalicular bile flow) generated by active bile acid transport into the canaliculus. Extrapolation of the linear regression function relating bile acid secretion rate and erythritol clearance to a value of zero bile acid secretion rate yields a positive zero intercept. This linear extrapolation identifies the so called bile acid-independent bile flow (about 40% of total canalicular bile flow).

Bile acid-independent bile flow is generated by active transport of glutathione and organic anions

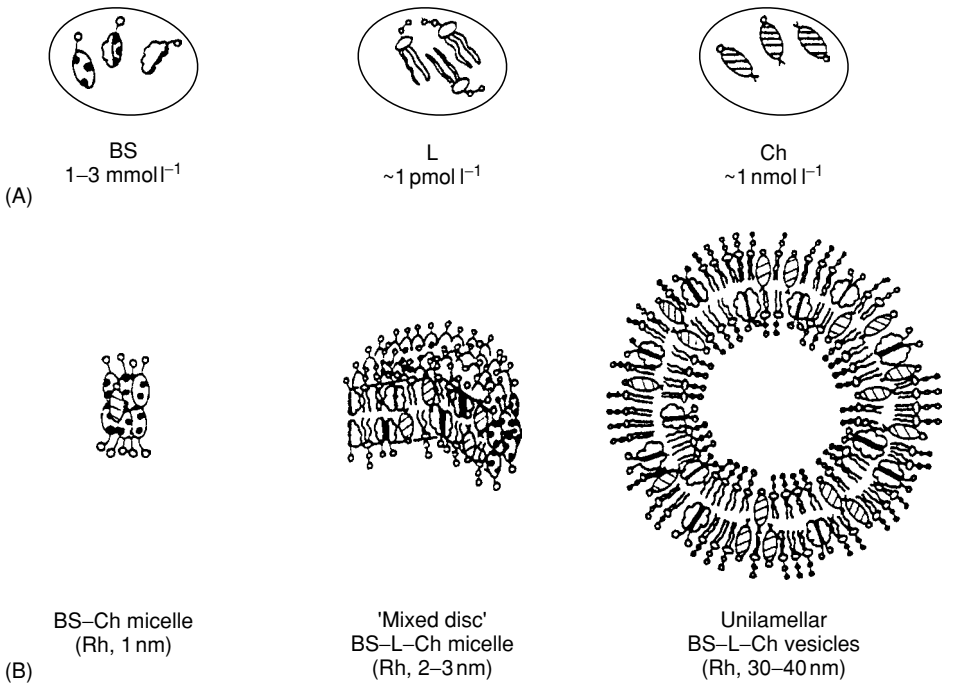

Figure 2 Molecular models of (A) monomeric and (B) aggregated biliary lipids. Monomeric solubility in water and mean hydrodynamic radius (Rh) of aggregated lipids are also shown. BS, bile acids; L, lecithin; Ch, cholesterol). From Carey MC and Duane WC (1994) Enterohepatic circulation. In: Arias IM, Boyer JL, Fausto N *et al*. (eds.) *The Liver: Biology and Pathobiology*, 3rd edn, pp. 719–767. New York: Raven Press, with permission.

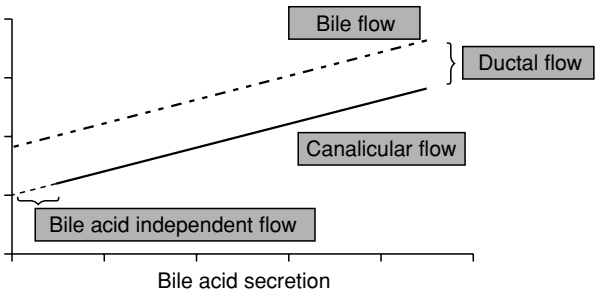

Figure 3 Relationship between bile acid secretion into bile and bile flow.

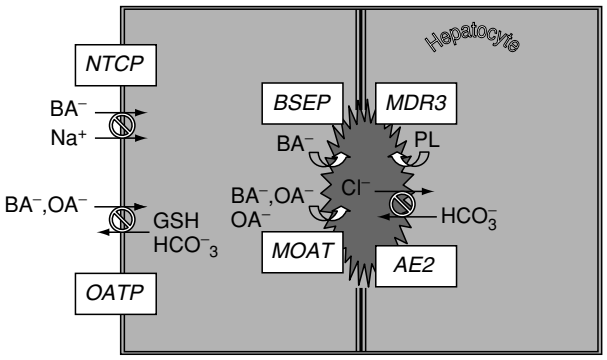

Figure 4 Major membrane transporters involved in bile formation. NTCP, Na^+ taurocholate cotransporter polypetide; OATP, organic anion transporting polypetide; BSEP, bile salt export pump; MOAT, multispecific organic anion transporter; MDR3, multi drug resistance 3; AE2, anion exchanger 2.

(bilirubin glucuronide, thyroid and steroid hormones, drugs) into the canaliculus. Glutathione is synthesized in the hepatocyte, secreted into bile, where it is hydrolyzed into the constituent amino acids. These amino acids exert the osmotic gradient responsible for bile acid-independent paracellular entry of water into the biliary canaliculus. Na^+ and HCO_3^- also contribute to generating the osmotic gradient responsible for the bile acid-independent bile flow.

Bile acid-dependent bile flow is a complex phenomenon involving captation, intracellular transport and biliary secretion of bile acids (Figure 4). Bile acid captation occurs on the basolateral membrane of the hepatocyte and involves two transport systems. The first is a Na^+-taurocholate cotransporter polypeptide dependent on a sodium gradient energized by the Na^+–K^+–ATPase system. The second transport mechanism is independent of sodium gradient and consists of an organic anion transporting polypeptide. This system is not specific for bile acids, and transports other organic anions such as bilirubin, bromosulfophthalein, and estrogens.

The process of intracellular bile acid transport is poorly understood, and it is thought to occur by diffusion of bile acid bound to cytosolic protein from the sinusoidal to the canalicular pole of the hepatocyte. Vesicular transport may also play a role in regulating intracellular bile acid traffic, as suggested by the finding that vesicle poisoning results in cholestasis.

Bile acid secretion into the biliary canaliculus involves two ATP-dependent transport systems and an electrogenic system driven by membrane potential. Monovalent bile acids are transported by an ATP-dependent canalicular bile salt export pump, and sulfated or glucuronated divalent bile acids are transported by a canalicular multispecific organic anion transporter. This latter transport system is also involved in biliary excretion of bilirubin diglucuronide, glutathione, bromosulfophthalein, and other anionic dyes.

Phospholipid secretion into bile is also an active ATP-dependent process mediated by *p*-glycoprotein encoded by two genes in humans (multidrug-resistant; MDR1 and MDR3) that make cells resistant to a variety of amphipatic drugs (colchicine, vincristine, doxorubicin, and others) by stimulating their extrusion from the cell. MDR3 protein is expressed exclusively on the bile canalicular membrane and is specialized in biliary export of phospholipid, as indicated by the finding that homozygous disruption of the murine mdr2 gene (the equivalent of human MDR3) leads to complete absence of phospholipid in bile.

Cholesterol secretion into bile is tightly coupled with that of phospholipid, but it is not clear whether cholesterol secretion is carrier-mediated, as demonstrated for phospholipids. It is likely that cholesterol translocation to the outer leaflet of the bile canalicular membrane is mediated by a spontaneous flip-flop mechanism.

The current hypothesis to explain the phenomenon of lipid secretion into bile is summarized in Figure 5. Bile acids are continuously pumped out of the bile canalicular membrane into bile. When the outer leaflet of the bile canalicular membrane is sufficiently enriched with phosphatidylcholine, it 'buds out,' forming a phospholipid–cholesterol bilayer that adsorbs to bile acids and dissolves into mixed bile acid–cholesterol–phospholipid micelles (see below).

Bile flow and composition changes along the biliary tree, due to water and inorganic electrolyte movements across the duct system. The net result of these movements is an increase in bile flow due to secretion of bicarbonate and water. This process is under control of the enteric hormone secretin that stimulates secretion of Cl^- by a 'cystic fibrosis

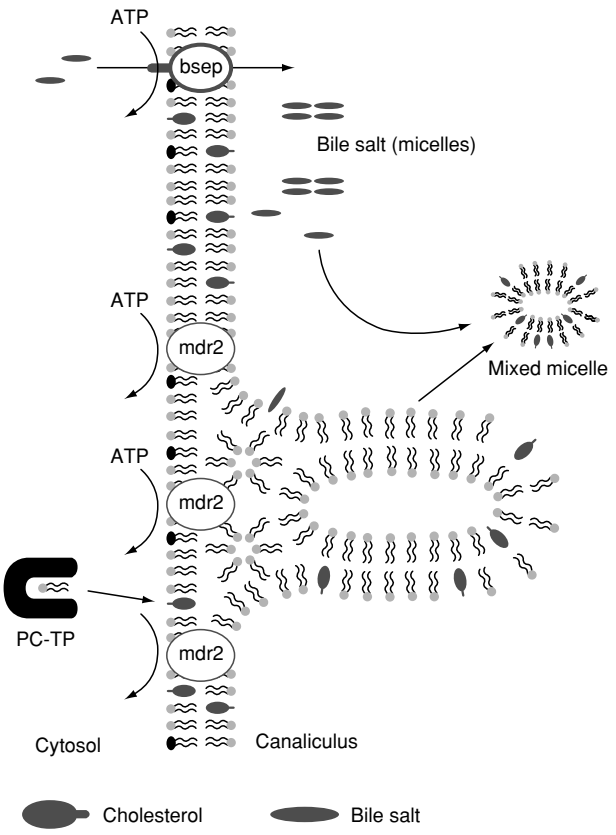

Figure 5 Biliary canalicular events involved in lipid secretion. PC-TP = phosphatidylcholine transporter; bsep, bile salt export pump; mdr 2, multidrug resistance 2. From Hofmann (1999) The continuing importance of bile acids in liver and intestinal disease. *Archives of Internal Medicine* 159: 2647–2658, with permission.

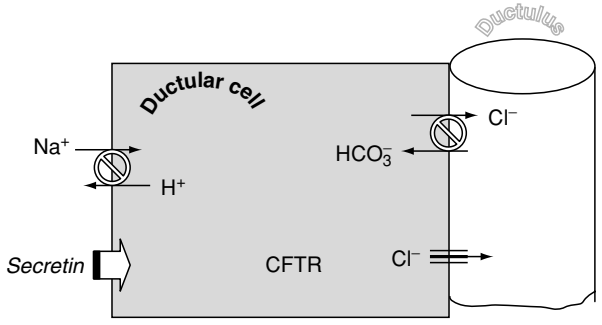

Figure 6 Mechanism for HCO_3^--dependent choleresis in the biliary epithelium. CFTR, cystic fibrosis conductance regulator.

transmembrane conductance regulator' acting as a chloride channel on the apical membrane of the biliary epithelial cells. Cl^- secretion stimulates a Cl^- bicarbonate exchanger, resulting in bicarbonate secretion and bicarbonate driven choleresis (Figure 6).

Bile Acid Synthesis and Enterohepatic Circulation

Bile Acid Synthesis

Biliary bile acids are derived from two sources, *de-novo* hepatic synthesis and hepatic recycling via the enterohepatic circulation. The former mechanism compensates for a daily fecal loss of about 200–600 mg. The liver is the unique site for bile acid synthesis, and cholesterol is the obligatory precursor. Bile acid synthesis involves both nuclear and side-chain transformation of cholesterol molecules, and about 25 intermediate compounds have been isolated along this pathway. A 'neutral' and an 'acidic' bile acid biosynthesis pathway have been identified, and the end products of the 'neutral' synthetic pathway in man are two bile acids, cholic and chenodeoxycholic acid, conventionally called 'primary' bile acids (Figure 1). 7-α-hydroxylation of cholesterol, a process catalyzed by microsomal 7-α-hydroxylase, is the rate-limiting step in 'neutral' bile acid synthesis. The first step in the 'acidic' pathway is catalyzed by sterol 27-hydroxylase, a mitochondrial P_{450} enzyme, and chenodeoxycholic acid is the predominant end product of this pathway. The contribution of the 'acidic' pathway to the total synthesis is unclear, but it can be as large as 50% of total synthesis. Bile acids resulting from subsequent bacterial modification of the two primary bile acids in the enterohepatic circulation are termed 'secondary' bile acids (see below). Bile acids resulting from hepatic sulfation of lithocholic acid and hydrogenation of 7-oxo-lithocholic acid are termed 'tertiary' bile acids (Figure 1).

It is generally agreed that bile acids returning to the liver via the enterohepatic circulation regulate their own synthesis by a negative-feedback mechanism. Bile acid synthesis may increase as much as 20-fold during interruption of the enterohepatic circulation in the bile fistula rat, and an opposite effect is achieved by increasing bile acid return to the liver. Inhibition of both cholic and chenodeoxycholic acid synthesis accompanies intraduodenal or intravenous infusion of either bile acid in the bile fistula rat. Individual bile acids appear to exhibit a specific inhibitory capacity for synthesis, which is less marked for the more hydrophilic bile acid. This inhibitory effect of hydrophobic bile acids involves transcriptional downregulation of cholesterol 7-α-hydroxylase. By contrast, increased supply of cholesterol to the cell turns on the 7-α-hydroxylase gene, thus enhancing cholesterol elimination via bile acid synthesis.

Enterohepatic Circulation and Gallbladder Storage of Bile

The movement of bile acid molecules is mainly restricted to the biliary tree, intestinal lumen, portal blood and liver, topographically referred to as the enterohepatic circulation. The driving force for the enterohepatic circulation is provided by two chemical pumps and two mechanical pumps. The chemical pumps are provided by the active transport processes underlying ileal absorption and hepatic uptake, and the mechanical pumps are gallbladder-emptying and small intestinal transit. (*See* **Liver**: Enterohepatic Circulation.) Gallbladder motor function is the major factor affecting delivery of bile into the intestine, thus influencing bile acid recycling frequency in the enterohepatic circulation. As a result of gallbladder storage and emptying functions, duodenal bile acid output typically varies from 5 μmol kg^{-1} h^{-1} in the interdigestive period to 25 μmol kg^{-1} h^{-1} postprandially. Low duodenal bile acid output during the interdigestive period is accompanied by diversion of about 90% of hepatic bile into the gallbladder. Gallbladder filling alternates with gallbladder emptying occurring synchronously with the periodic aboral progression of the intestinal interdigestive migrating motor complexes. This alternation of filling and emptying also occurs postprandially, and this 'bellows' effect may play a role in mixing of gallbladder content during the whole 24 h. The net effect of these alternating episodes of filling and emptying is that the gallbladder stores only about 50% of the hepatic bile that enters it during overnight fasting. Given a capacity of 20–40 ml, the gallbladder would be filled very soon during fasting beyond its storage capacity if it were not for its marked capacity for active NaCl and $NaHCO_3$, and passive water reabsorption.

Functions of Bile Acids

Bile

Despite the very low solubility of both phospholipid and cholesterol in water, bile is an optically clear solution that does not contain a macroscopically detectable solid phase in health. At physiological bile salt–lecithin–cholesterol molar ratios (approximately 75%–15%–5%), cholesterol monomers are solubilized in large mixed micelles having a mixed disk structure (Figure 2). This structure envisages bile acid covering the external surface of the micelles, but also dispersed inside the micellar core together with phospholipid.

The amount of cholesterol in a given bile sample may exceed the solubilization capacity of the micellar mechanism, as determined from *in-vitro* studies on model bile solutions (supersaturated bile). Hepatic bile is normally supersaturated with cholesterol at low bile acid secretion rates (< 10–15 μmol kg^{-1} h^{-1}), and gallbladder bile is often supersaturated in healthy subjects. Cholesterol monohydrate crystals are not found in these supersaturated biles when analyzed immediately following collection. However, if these supersaturated biles are stored under appropriate laboratory conditions, phase separation and cholesterol crystal formation occur within days or weeks (nucleation time). These observations indicate that bile behaves as a metastable system, and a nonmicellar mechanism of cholesterol transport must be involved to account for this prolonged metastability. This mechanism involves the formation of spherical unilamellar vesicles composed mainly of cholesterol and lecithin (ratio ranging between 1:1 and 5:2, Figure 2). The proportion of cholesterol carried in vesicles relative to that carried in mixed micelles is about 50% in hepatic bile and 40% in gallbladder bile. This proportion is not constant, and the vesicular system prevails at low bile acid secretion rates. The degree of metastability of supersaturated human bile, as assessed by *ex-vivo* nucleation time, is greater than that predictable from the nucleation time of model bile systems. A protein component of bile, probably a lipid-containing bile protein, may act as a 'stabilizer' of supersaturated normal biles.

Phase transition occurs during passage of bile down the biliary tree and during storage in the gallbladder (Figure 7). Cholesterol/phospholipid vesicles are probably excreted by the biliary canaliculus as such by a mechanism independent of that involved in bile acid monomer secretion (Figure 5). As soon as the bile acid concentration reaches the CMC within the biliary canaliculus, vesicular lipids are dissolved in mixed micelles. Supersaturated mixed micelles in hepatic bile rapidly form stable cholesterol/lecithin vesicles, probably unsaturated with cholesterol. On entry into the gallbladder, this lipid fraction may be slowly redissolved into mixed micelles unsaturated with cholesterol. This latter effect occurs as a result of the increased total lipid concentration in concentrated gallbladder bile, and this is known to enhance the micellar capacity for cholesterol. In supersaturated gallbladder bile, cholesterol-rich vesicles may remain stable as a result of protein–vesicle interaction.

Small Intestine

According to the classical theory of Hofmann and Borgstrom, normal fat digestion can be divided into two closely related events: a chemical event – lipolysis;

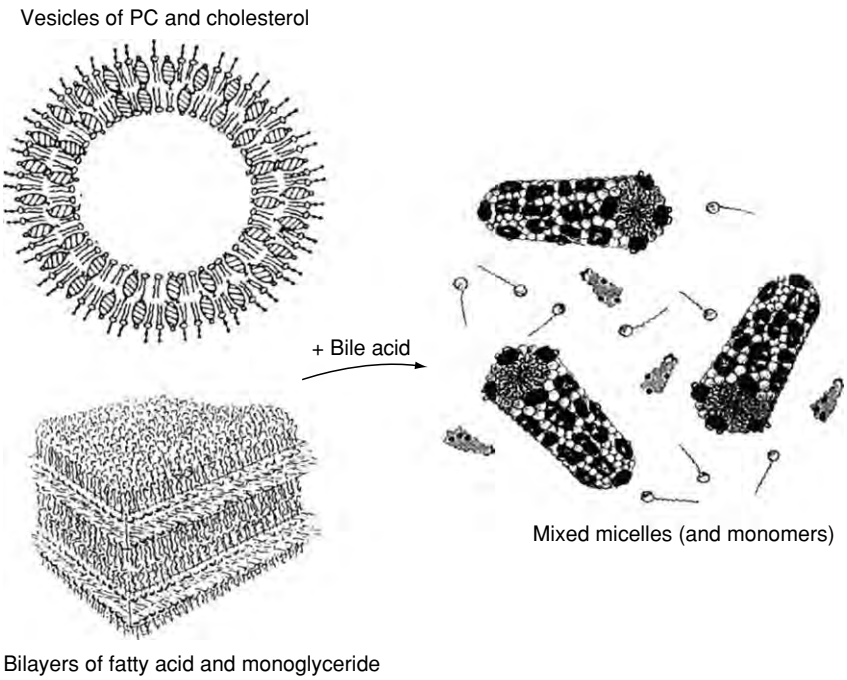

Figure 7 Conversion of phosphatidylcholine (PC)–cholesterol bilayer to mixed micelles in bile, and of fatty acid–monoacyglycerol lamellae to mixed micelles during fat digestion in the intestinal lumen. From Hofmann (1999) The continuing importance of bile acids in liver and intestinal disease. *Archives of Internal Medicine* 159: 2647–2658, with permission.

and a physical event – acqueous solubilization. Pancreatic lipolysis of dietary fat is a prerequisite for its micellar solubilization, and micellar solubilization is a prerequisite for its absorption. Bile acids play a key role in both these events.

From a physicochemical point of view, fat droplets consist of long-chain triacylglycerol stabilized mainly by phospholipids in natural food, and by other emulsifiers (arabic gum, mono- or diacylglycerols) in processed food. Dietary triacylglycerol is hydrolyzed by pancreatic lipase (lipolysis) to monoacylglycerol and fatty acid. In the absence of bile acids, access of lipase to lipid droplets is prevented by the emulsifiers surrounding the oil droplets. Bile acids have the ability to wash these off the oil droplets, thus giving access to lipase. Bile acids would also wash off the lipase, a phenomenon prevented by colipase. This coenzyme binds to lipase and bile acid, thus fixing the lipase molecule in contact with the underlying triacylglycerol.

Following digestion, the products of lipolysis are kept in solution by the natural detergents, bile acids, mainly in form of mixed bile acid–phospholipid–fatty acid–monoacylglycerol micelles. Unilamellar vesicles consisting of the products of lipolysis have been observed under the microscope to 'bud off' the lipid emulsion surface following lipase digestion. These vesicles are rapidly dissolved by biliary micelles freshly delivered by bile, so that during intestinal fat digestion, saturated mixed-disk micelles coexist with vesicles having the same lipid composition (Figure 7). Both micelles and vesicles contribute to transport of solubilized lipids across the intestinal unstirred water layer lining the gut lumen. It is not known whether lipid absorption involves monomeric absorption or vesicle fusion with the plasma membrane of the enterocyte.

Colon

Two hundred to 500 mg of bile acids escape the enterohepatic circulation and enter the colon daily. In the human colon, bile acids are exposed to obligate anaerobes bacteria capable of producing hundreds of metabolic products of bile acid. The two main biotransformations of bile acids in the colon are bacterial deconjugation and 7-α-dehydroxylation. As a consequence of the former biotransformation, conjugated bile acids are totally absent from the colon. As a result of 7-α-dehydroxylation of the primary bile acids, cholic and chenodeoxycholic acid, the majority of fecal bile acids consist of deoxycholic acid (about 50%) and lithocholic acid (33%) (Figure 1). Since these two bile acids are reabsorbed and resecreted in bile along with the two primary bile acids, they are termed 'secondary' bile acids.

Most of the bile acids precipitate in feces, since the slightly acidic pH of feces is lower than the pK_a value for unconjugated bile acids. The bile acid concentration in fecal water is about 0.5 mmol l^{-1}, and a small proportion of these bile acids are reabsorbed from the colonic mucosa by passive nonionic diffusion and reenter the enterohepatic circulation. If the value of 1.5 mmol l^{-1} for bile acid concentration in fecal water is exceeded, bile acids exert a cathartic effect by inhibiting water absorption and, at higher concentrations, by promoting colonic electrolytes and water secretion. The cathartic effect of bile acids depends on the bile acid structure, and it is at a maximum for α-dihydroxylated bile acids.

See also: **Cholesterol**: Properties and Determination; Absorption, Function, and Metabolism; Factors Determining Blood Cholesterol Levels; Role of Cholesterol in Heart Disease; **Colon**: Structure and Function; **Gallbladder**; **Liver**: Enterohepatic Circulation; **Phospholipids**: Properties and Occurrence; Determination; Physiology

Further Reading

Bahar RJ and Stolz A (1999) Bile acid transport. In: Cooper AD (guest ed.) Bile salts: metabolic, pathologic, and therapeutic considerations. *Gastroenterology Clinics of North America* 28(1): 27–58.

Borgstrom B, Barrowman JA and Lindstrom M (1985) Roles of bile acids in intestinal lipid digestion and absorption. In: Daniellsson H and Sjovall J (eds) *Sterols and Bile Acids*, pp. 405–425. Amsterdam: Elsevier Science.

Carey MC and Duane WC (1994) Enterohepatic circulation. In: Arias IM, Boyer JL, Fausto N *et al.* (eds) *The Liver: Biology and Pathobiology*, 3rd edn, pp. 769–786.New York: Raven Press.

Elferink RPJ and Groen AK (1999) The mechanism of biliary lipid secretion and its defects. In: Cooper AD (guest ed.) Bile salts: metabolic, pathologic, and therapeutic considerations. *Gastroenterology Clinics of North America* 28(1): 59–74.

Erlinger S (1994) Bile flow. In: Arias IM, Boyer JL, Fausto N *et al.* (eds) *The Liver: Biology and Pathobiology*, 3rd edn., pp. 769–786. New York: Raven Press.

Hofmann AF (1994) Bile acids. In: Arias IM, Boyer JL, Fausto N *et al.* (eds) *The Liver: Biology and Pathobiology*, 3rd edn, pp. 677–718. New York: Raven Press.

Hofmann AF (1999) The continuing importance of bile acids in liver and intestinal disease. *Archives of Internal Medicine* 159: 2647–2658.

Lanzini A and Northfield TC (1991) Biliary lipid secretion in man. Review. *European Journal of Clinical Investigation* 21: 259–272.

Muller M and Jansen PLM (1997) Molecular aspects of hepatobiliary transport. *American Journal of Physiology* 272: G1285–G1303.

Vlahcevic ZR, Pandak WM and Stravitz RT (1999) Regulation of bile acid biosynthesis. In: Cooper AD (guest ed.) Bile salts: metabolic, pathologic, and therapeutic considerations. *Gastroenterology Clinics of North America* 28(1): 1–26.

BIOAVAILABILITY OF NUTRIENTS

S J Fairweather-Tait and S Southon, Institute of Food Research, Norwich, UK

Concept and Definition of Bioavailability

Bioavailability (biological availability) is a term used to describe the proportion of a nutrient in food that is utilized for normal body functions. Although the overall concept is simple, it is very difficult to describe the bioavailability of most nutrients in quantitative terms. Therefore, to facilitate its measurement and interpretation of the data, the bioavailability of nutrients can be subdivided into its three constituent phases: (1) availability in the intestinal lumen for absorption; (2) absorption and/or retention in the body; and (3) utilization. The reasons for studying bioavailability are to evaluate the nutritional quality of foods and diets and to provide data for establishing dietary requirements for nutrients.

The amount of any nutrient that is available for absorption depends upon its release from the food macro- and microstructure during digestion, chemical and physical interactions with other food components during digestion, the form in which the nutrient is presented to the absorptive cells of the intestine, and regulatory mechanisms in the intestinal mucosa that reflect the individual's physiological need for the nutrient. These factors contribute to the (often) highly variable proportion of absorbable nutrient which is taken up and used by the body and may explain the generally poor association between total amounts of micronutrients consumed and indices of micronutrient status.

Low absorbability of some minerals and vitamins may be the underlying cause of certain nutritional inadequacies. If intakes are well in excess of

physiological requirements but bioavailability is a limiting factor, nutritional deficiency disorders may occur, as for example with iron.

Determinants of Bioavailability

There are many factors, both dietary and physiological, that influence nutrient bioavailability. Examples include: (1) the physical form of the nutrient within the food structure and the ease with which the nutrient can be released from that structure; (2) the chemical form of the nutrient in a foodstuff and its solubility in the lumen; (3) the presence of proteolytic enzyme inhibitors (commonly associated with legumes such as soybeans) which reduce the body's ability to digest protein; and (4) the presence of enzymes such as thiaminase which partially hydrolyzes thiamin and makes it less biologically active.

Diet-related factors include:

- Food structure
- Physicochemical form of the nutrient
- Enhancers of absorption, e.g., ascorbate (for iron), some organic acids, sugars, amino acids, bulk lipid (for fat-soluble vitamins), and specific fatty acids
- Inhibitors (primarily of inorganic micronutrient absorption), e.g., phosphates (especially phytate), polyphenols (including tannins), and oxalate
- Competition for transport proteins or absorption sites, e.g., between metals.

Physiological factors include:

- Gastric acidity
- Intestinal secretions
- Gut motility
- Luminal redox state
- Body status (e.g., tissue levels, nutrient stores)
- Short-term homeostatic mechanisms mediated through the mucosal absorptive cells
- Anabolic demands (e.g., growth in infancy and childhood, pregnancy, and lactation)
- Endocrine effects
- Infection and stress
- Genetic polymorphisms and inborn errors of metabolism
- Gut microflora.

Certain food constituents have the ability to bind nutrients, thereby rendering them more or less absorbable. For nutrients that are transported into the mucosal cell by means of a carrier-mediated pathway, the degree of binding is an important determinant of bioavailability. If the chelating compound has a higher affinity for the nutrient than the specific carrier molecule, then the net effect is a reduction in bioavailability. Conversely, a weak chelator may act as an absorption promoter by holding the nutrient in a suitably soluble form ready to be taken up into the intestinal mucosal cells. Binding compounds that impair vitamin bioavailability include the protein avidin in egg white, which binds biotin, making it biologically unavailable.

Competitive inhibitors of nutrient metabolism make up another category of dietary factors affecting bioavailability. It has been suggested that minerals with similar chemical properties may compete for common binding sites or carriers. Transition metals such as iron, zinc, and copper are typical examples of competitive inhibitors. This will only take place at high levels of intake when the sum of ionic species present at the site of absorption exceed the critical threshold relating to the absorption kinetics of the minerals in question. (*See* **Copper**: Physiology; **Iron**: Physiology; **Zinc**: Physiology.)

The bioavailability of lipid-soluble nutrients (including carotenoids, tocopherols, and other fat-soluble vitamins) is markedly influenced by their physicochemical availability for incorporation into mixed micelles during digestion. The release of lipid-soluble components is thought to occur primarily upon ingestion and initial digestion within the stomach. The factors which may influence their release are: localization within the food matrix; physical break-up of the food; chemical/enzymatic breakdown of the food during ingestion and initial digestion; and the presence of a suitable lipid phase in the form of emulsion droplets or free lipid.

The absorbability of a nutrient also depends upon certain physiological factors, such as the composition and volume of gastric and intestinal secretions. The absorption and utilization clearly depend upon a number of host-related variables. Most of these are key participants in the body's homeostatic regulatory mechanisms such as nutritional status, developmental state, gastric and intestinal secretions, mucosal cell regulation, and gut microflora.

Methods of Determining Bioavailability

Various techniques have been employed to assess the bioavailability of nutrients. The best estimate is probably obtained by combining the results from several different methods. By and large, the methods chosen depend on the resources available since detailed human studies on nutrient bioavailability use specialized techniques, which can be very expensive.

In Vitro Techniques

These range from simple measures of chemical solubility and fractional dialysability following simulated

digestion to measures of uptake and transport by mucosal tissue from experimental animals. The substance of unknown bioavailability is incubated with a variety of intestinal preparations, including intestinal rings, everted loops (sacs), isolated mucosal cells from cell culture, and vesicles. Uptake of the nutrient into the tissue is determined and equated with absorbability. Perfused loops, with an intact blood supply, can be used to measure the transport of the nutrient into the body. More recently, Caco-2 cell systems have been used for iron transport studies and these have generated good predictive data for iron bioavailability. This cell system may be useful for predicting the bioavailability of other micronutrients. Measurements *in vitro* of the kinetics and extent of partitioning of the lipid-soluble nutrients into oil and micellar phases also provide a means of determining the amount of these nutrients available for absorption.

Generally speaking, *in vitro* techniques are reproducible and can be used for studying dietary factors that affect nutrient bioavailability. They are relatively simple and inexpensive, but their relevance to utilization in humans requires validation. Results obtained from *in vitro* studies are only an indication of the proportion of a nutrient that is available for absorption, and cannot be used to predict utilization *in vivo* as this is dependent on many host-related factors.

Balance Studies

Fecal Metabolic Balance Approach

This relies on measurement of the difference between input and excretion of the nutrient and is used to attempt to quantify amount absorbed. Analysis of intake and fecal content of the nutrient over the balance period, or after an acute dose, must be extremely accurate because relatively small differences need to be detected. The diet must be carefully controlled during and sometimes for periods before and after the dose. Fecal markers need to be employed to determine the start and end of the balance period and to insure complete collection, and it must be assumed that feces are the only significant excretory mechanism, and that none of the unabsorbed nutrient has undergone biotransformation (due to digestive processes or microbial action), or otherwise been lost. It is also essential to consider enterohepatic circulation because the nutrient may have been absorbed, re-excreted in the bile, and then lost to feces. Similarly, enteric recycling can occur within the gut, involving absorption, excretion, and reabsorption of nutrient over several cycles. In both cases, the nutrient enters the body pool where it will have an impact on the metabolic kinetics and would therefore be considered to be bioavailable.

It is likely that for many nutrients these assumptions cannot be made because of their susceptibility to microbial and oxidative degradation in the gut, so that the unabsorbed portion will not be quantitatively recovered from the feces. Nevertheless, much of our information on the extent of absorption of certain nutrients is based on either acute or chronic fecal mass balance methods which, not surprisingly, show great variability in results.

The Gastrointestinal Lavage Technique

This approach has been used in an attempt to overcome some of the problems inherent in the 'simple' fecal mass balance approach. In the gastrointestinal lavage technique the entire gastrointestinal tract is washed out by consuming a large volume of Colyte containing polyethylene glycol and electrolyte salts. Washout is complete with the production of clear rectal effluent and the volunteers then consume the test meal and are permitted only water or 'diet' soft drinks for the next 24 h. All the effluent is collected and pooled with the effluent collected on the following day when another dose of Colyte is given to wash out the remainder of the test meal. The compound recovered in the stool is subtracted from that fed to obtain an absorption figure. The advantage of this approach is that the residence time of the compound in the gastrointestinal tract (particularly the large bowel, where fermentation occurs) is reduced and is standardized. The disadvantages are that the method is time-consuming, may give an underestimation of absorption if Colyte causes excessive transit rates and, again, the method depends upon there being no degradation or loss of unabsorbed compound.

The Ileostomy Mass Balance

In individuals who have undergone ileostomy the colon has been surgically removed and the terminal ileum brought to a stoma on the abdominal wall. Ingested food passes through the stomach and ileum in around 6 h as it would in the intact individual. The digesta (ileal effluent) can be collected at regular intervals (over about 12 h) from a test meal given in the morning after an overnight fast. During this time the volunteers may consume meals and beverages which are free of the nutrient being studied. Using this approach, the unabsorbed compound can be recovered from the ileal effluent in real time without the delay of the colon and rectum, or the confounding influence of the colonic microflora.

Bioassays

The response of experimental animals or humans to graded levels of the nutrient in the diet can be used to assess bioavailability. The response criteria employed should specifically reflect the utilization of the nutrient of interest, either by direct measurement of the nutrient or its metabolites in tissues, blood, or urine, or by employing functional measurements, such as enzyme levels.

Plasma Response: Acute Doses

Measurement of changes in plasma, serum, or whole blood concentration following acute or chronic dosing is currently the most well-used approach for estimating absorption and clearance. However, changes in blood concentration, particularly following a single acute dose of dietary amounts of a nutrient, can be difficult to detect and the results from such studies are those most frequently misunderstood and misinterpreted. The plasma response approach depends upon frequent blood sampling after an acute dose of an isolated compound, or after a single meal. Changes in concentration of the compound of interest are measured and plotted against time to produce a response curve. The area under the curve (AUC) is then calculated and used to determine the extent of absorption.

At its simplest, the AUC is dependent upon the rate at which the newly absorbed nutrient enters the blood and the rate of disposal to other body compartments, both of which are occurring at the same time. Changes in the characteristics of the AUC within a study may therefore be due to differences in absorption, or differences in kinetics of disposal, or both. There may also be a problem of reexportation of the compound between the plasma and other compartments, so that the AUC now consists of three components: newly absorbed compound entering the plasma; disposal of the compound to other tissues and possibly urinary excretion; and reexportation of the compound from a tissue (or tissues) into the plasma. In isolation, and without a mechanistic understanding of the processes of absorption and disposal, the AUC cannot usually provide quantitative data on the extent of absorption. However, if sufficient time points are obtained, and if metabolic modeling techniques are employed, it is possible to calculate both the rate and extent of absorption. This approach is particularly useful for hydrophilic nutrients present in the aqueous plasma phase. For lipophilic nutrients quantifying amounts absorbed is more complex because the whole plasma response is a multicomponent response involving the transfer of the nutrient (and/or its metabolites) to, and between, carriers in the blood which each have their own absorption and clearance kinetics. A better approach for lipophilic nutrients is to examine that blood fraction which contains the newly absorbed nutrient, e.g. the triglyceride-rich lipoprotein (primarily chylomicron) fraction.

Plasma Response: Chronic Doses

Chronic dosing with foods or isolated nutrients needs to be carried out until the plasma concentration reaches a plateau, being careful not to 'flood' the system, e.g., not to exceed the renal threshold. Absolute absorption cannot be determined by chronic dosing (only relative absorption) and differences between individuals in disposal rate to the tissues can be a confounder when interpreting changes in plasma concentration. As with acute plasma response AUC, chronic response cannot be used to compare different nutrients whose absorption and clearance kinetics are not known.

Isotopes

Many of the problems of lack of measurement specificity can be overcome by using radio- or stable isotopes. Foods can be labeled with isotopes, and the fate of the labeled nutrient following ingestion monitored in the body. The method used to label the food is an important consideration. Extrinsic labeling, as used for heme and nonheme iron, is the simpler technique. The isotope is mixed into the food just before consumption and is assumed to exchange fully with the nutrient in the food. However, intrinsic labeling, in which the isotope is incorporated into the foodstuff whilst it is growing, is necessary for some minerals (e.g., selenium) and most vitamins.

Isotopes can be used to measure the absorption, and sometimes the utilization, of nutrients in the body from different foods and under different dietary and physiological conditions. A number of isotope techniques can be used and alternative approaches or refinements of existing methods are currently under development, particularly for stable isotopes. These include whole-body counting (for gamma-emitting isotopes), fecal and urinary monitoring, double isotope techniques (urinary and plasma measurements), and plasma appearance/disappearance kinetics (similar to methods used for measuring the bioavailability of drugs). Selection of technique depends on the nutrient under study, the questions to be answered, and the experimental conditions.

The use of radioisotopes for human studies is restricted by ethical considerations. The hazards associated with exposure to ionizing radiation preclude the use of radioisotopes in infants, children, and

pregnant women. Since these groups are most vulnerable to nutritional deficiency, they are the subject groups for whom nutrient bioavailability is of particular interest. Thus alternative methods to study bioavailability have been sought using stable isotopes, which are safe and ethically acceptable. The main disadvantages of stable as compared to radioisotopes stem from the larger quantities that have to be employed, usually far exceeding the tracer amounts required for radioisotope studies, and the difficulties associated with their measurement. (*See* **Immunoassays**: Radioimmunoassay and Enzyme Immunoassay.)

Minerals

For many nutrients the bioavailable fraction is high (i.e., 90% or more), but for a number of inorganic elements it is low and variable. Examples of ionic species that have a low (< 25%), medium (25–75%), and high (> 75%) absorption are:

- Low: Fe, Mn, Cr, Ni, V
- Medium: Ca, Mg, Zn, Cu, Se, MO, CO, PO_4
- High: Na, K, Cl, I, F.

Iron

Iron deficiency is a worldwide problem, yet iron is one of the most abundant elements in the earth's crust. This paradox can be explained in terms of bioavailability. The main factors that influence the bioavailability of iron from the diet are the amounts of heme and nonheme iron, the presence and amounts of dietary factors influencing iron absorbability, and the iron status of the individual. (*See* **Iron**: Physiology.)

Heme and nonheme iron are absorbed by two separate pathways: the heme complex is absorbed intact into the mucosal cell, whereas nonheme iron is bound to a transport protein. The absorption of nonheme iron is very variable and generally much less than heme iron. Absorption of heme iron is relatively constant (approximately 20–30%), and only marginally affected by dietary factors that enhance or inhibit nonheme iron absorption. These include meat, ascorbic, and certain other organic acids, all of which increase iron bioavailability; tannins, phytate, and calcium decrease bioavailability. (*See* **Ascorbic Acid**: Physiology; **Phytic Acid**: Nutritional Impact; **Tannins and Polyphenols**.)

Physiological variables have a profound influence on iron bioavailability, notably body iron status. Previous dietary iron intake will also affect the bioavailability of subsequent iron, probably mediated via mucosal cell regulation. Iron is one nutrient for which it is possible to estimate true bioavailability because absorbed iron is utilized for hemoglobin production. Therefore the incorporation of isotopically labeled food iron into red blood cell hemoglobin can be measured to assess dietary iron bioavailability. Another method involves measuring the rate of hemoglobin regeneration in deficient subjects in response to different dietary sources of iron compared with a reference substance (e.g., ferrous sulfate).

Zinc

A number of dietary factors affect zinc bioavailability. Phytate is probably the most important zinc antagonist, especially in the presence of calcium, as it forms a chelate with zinc which is unavailable for absorption. Other cations with similar physicochemical properties (copper and cadmium) or mutual affinity for carrier protein (iron) also reduce zinc bioavailability. Some proteins have been shown to improve zinc bioavailability, but the mechanisms for the effect are not yet clear. (*See* **Cadmium**: Toxicology; **Copper**: Physiology; **Zinc**: Physiology.)

Zinc homeostasis in the body is controlled by alterations in efficiency of absorption coupled with changes in the quantities of endogenous zinc secreted into the intestine. Thus, body zinc status plays an important role in determining dietary zinc bioavailability.

Calcium

Calcium absorption is vitamin D-dependent; therefore bioavailability depends upon vitamin D intake and status. The efficiency of absorption is related to physiological requirements for calcium and is dose-dependent. Dietary inhibitors of calcium absorption include substances that form complexes in the intestine, such as phytate. Protein and sodium may also modify calcium bioavailability in that high levels increase urinary calcium excretion. Although this is accompanied by an increase in intestinal absorption, the net result may be a reduction in the proportion of dietary calcium utilized by the body, i.e., lower bioavailability. Lactose, on the other hand, promotes calcium absorption. Other potential enhancers of calcium absorption are currently under investigation with regard to the development of functional foods for the prevention of osteoporosis, such as caseinophosphopeptides and nondigestible oligosaccharides. (*See* **Calcium**: Physiology; **Cholecalciferol**: Physiology; **Protein**: Interactions and Reactions Involved in Food Processing.)

Vitamins

A number of vitamins are not fully bioavailable to the human body. The more important ones are folate, vitamin B_6, niacin, vitamin A, carotenoids, and vitamin E. (*See* **Vitamins**: Overview.)

Folate

Folate exists in nature as a range of vitamers based upon the parent compound folic acid (pteroyl monoglutamic acid). Vitamers differ in the extent of the reduction of the pteroyl group, the presence of one-carbon substituents, and the number of glutamyl residues. The polyglutamyl form of folate makes up approximately 80% of dietary folate and must be cleaved by folate conjugase to the monoglutamate form for absorption.

There appears to be little or no difference in the extent of absorption of the various monoglutamyl forms, although stability in the gastrointestinal tract and *in vivo* retention may differ.

Numerous dietary and physiological factors influence the deconjugation and absorption of folate, and its subsequent utilization in the body. Dietary factors that may reduce folate bioavailability include conjugase inhibitors in foods, as found for example in pulses and yeast extracts, and dietary fiber such as wheat bran. Physiological factors that may reduce folate bioavailability include intraluminal pH (maximal absorption occurs at pH 6.3), and certain nutrient deficiencies, including zinc, vitamin B_6, and vitamin B_{12} deficiency. Although food processing and storage can cause losses of folates, it is conceivable that the action of endogenous conjugases during food preparation may increase the bioavailability of naturally occurring polyglutamyl forms. (*See* **Cobalamins**: Physiology; **Folic Acid**: Physiology.)

Vitamin B_6

Vitamin B_6 (3-hydroxyl-5-hydroxymethyl-2-methylpyridine) exists in foods as either the free or phosphorylated form of pyridoxine, pyridoxamine, and pyridoxal. Plant foods also contain pyridoxine glycosides (forming 5–50% of the total vitamin B_6 content), and it has been suggested that the observed lower bioavailability in plant than animal foods is associated with this form of vitamin B_6. (*See* **Vitamin B_6**: Properties and Determination.)

Niacin

Niacin deficiency is associated with diets containing high levels of maize because the niacin is chemically bound to carbohydrate by ester linkages and unavailable for absorption. However, alkali treatment, as used by certain cultural groups when processing foods in the traditional way, renders the niacin more bioavailable. Other grains, such as sorghum, wheat, barley, and rice, contain niacin in chemically bound forms. (*See* **Niacin**: Physiology.)

Vitamin A, Carotenoids, Vitamin E

These vitamins and provitamins need to be dissolved and carried in lipid and lipid plus bile salt systems (micelles) in order to be absorbed at the enterocyte brush border. This mass transfer from bulk aqueous food to lipid and micellar phases is a complex process that is hindered by the presence of food structure. Digestion products of dietary proteins (peptides) may play a role in stabilizing micelles because of their ability to act as surface active agents. Efficient absorption of retinol is dependent on the presence of normal mature enterocytes and is compromised in inflammatory bowel disease, protein-calorie malnutrition, zinc deficiency, and alcohol abuse. (*See* **Retinol**: Physiology.)

In addition to the level and type of fat, a number of other factors influence the bioavailability of carotenoids from foods. The major controlling factors are food structure and the physical form of the carotenoids within the food matrix (e.g., crystalline or lipoprotein complexed). The absorption of carotenoids from uncooked foods can be very low but cooking can enhance absorption by increasing the ease with which carotenoids are released from the food structure. Chopping and other types of food preparation leading to particle size reduction also improve absorbability. *In vitro* results indicate that gastric pH is an important physiological determinant of carotenoid availability. (*See* **Carotenoids**: Physiology.)

There are at least eight compounds of plant origin that have vitamin E activity, but α-tocopherol accounts for almost all of the vitamin E activity in foods of animal origin. This isomer has by far the highest biological activity of all the natural isomers. Under normal dietary conditions about 20–80% of ingested vitamin E is absorbed, depending on dose and lipid content of the meal. Efficiency of absorption decreases when large amounts of tocopherols are consumed, or when individuals suffer from disorders associated with fat malabsorption. It is reported that high intakes of pectin, wheat bran, alcohol, and polyunsaturated fatty acids also reduce vitamin E absorption. Dietary constituents such as vitamin A, iron, selenium, and zinc may affect vitamin E utilization. (*See* **Tocopherols**: Physiology.)

Nutritional Implications of Bioavailability

The ultimate goal in nutritional science is to understand the interactions between diet and health. This requires a detailed knowledge of nutrient bioavailability, without which it is impossible to relate dietary intakes to indices of physiological function used to assess health. Estimates of dietary requirements cannot be made without the appropriate bioavailability factors. Where diets are marginal or inadequate in one or more micronutrient it is often possible to improve the quality of the diet by increasing the bioavailability of the nutrients in question, with consequent improvements in the nutritional status of people consuming it. Future research should utilize recently developed techniques to quantify the different phases of bioavailability, differentiating where possible between dietary and physiological factors. The ultimate goal is to combine all the information to form a coherent picture of bioavailability of nutrients for different foods and diets.

See also: **Ascorbic Acid**: Physiology; **Carotenoids**: Physiology; **Cholecalciferol**: Physiology; **Cobalamins**: Physiology; **Folic Acid**: Physiology; **Immunoassays**: Principles; Radioimmunoassay and Enzyme Immunoassay; **Niacin**: Physiology; **Phytic Acid**: Nutritional Impact; **Protein**: Interactions and Reactions Involved in Food Processing; **Retinol**: Physiology; **Tannins and Polyphenols**; **Tocopherols**: Physiology; **Vitamins**: Overview; **Vitamin B_6**: Physiology

Further Reading

Fairweather-Tait SJ (1998) Trace element bioavailability. In: Sandstrom B and Walter P (eds) *Role of Trace Elements for Health Promotion and Disease Prevention*, no. 54, pp. 29–39. Basel: Karger.

Fairweather-Tait S and Hurrell RF (eds) (1996) Bioavailability of minerals and trace elements. *Nutrition Research Reviews* 9: 295–324.

Faulks RM, Hart DJ, Wilson PDG, Scott KJ and Southon S (1997) Absorption of all *trans* and 9-*cis* β-carotene in human ileostomy volunteers. *Clinical Science* 93: 585–591.

Glahn RP, Lee OA, Yeung A, Goldman MI and Miller DD (1998) Caco-2 cell ferritin formation predicts nonradiolabeled food iron availability in an *in vitro* digestion/Caco-2 cell culture model. *Journal of Nutrition* 128: 1555–1561.

Gregory JF (1997) Bioavailability of folate. *European Journal of Clinical Nutrition* 51: S54–S59.

Hurrell R (ed.) (1999) *The Mineral Fortification of Foods*. Surrey: Leatherhead International.

Jackson MJ, Fairweather-Tait SJ, van den Berg H and Cohn W (eds) (1997) Assessment of the bioavailability of micronutrients. Proceedings of an ILSI Europe workshop. *European Journal of Clinical Nutrition* 51 (suppl 1): S1–S90.

Rich G, Fillery-Travis A and Parker M (1998) Influence of pH on carotene transfer from carrot juice to olive oil. *Lipids* 33: 985–992.

Southgate DAT, Johnson I and Fenwick GR (eds) (1989) *Nutrient Availability: Chemical and Biological Aspects*. Cambridge: Royal Society of Chemistry.

Biochemical Pathways *See* **Tricarboxylic Acid Cycle**; **Oxidative Phosphorylation**

BIOFILMS

G Wirtanen, E Storgårds and T Mattila-Sandholm, VTT Biotechnology, Espoo, Finland

Formation of Biofilm in the Food Industry

When hygiene is poor, foodstuffs can be contaminated by microbes, e.g., bacteria or fungi from various sources, including raw materials, process equipment, packaging materials, or air. Common contamination sources in the process include transfer gears, packaging machines, flow plates, unhygienic valves, pumps, and pipe fittings. The number of microbes in packages is dependent on the materials from which they are manufactured. In packages made from recycled paper materials, fungal counts may rise to >100 colony-forming units (CFU) 100 cm^2, while foils containing plastics usually show significantly lower levels. The distribution of microorganisms in the air is highly dependent on local air flow, humidity, temperature, air pressure, settling properties of the

microbes, and their resistance to dehydration and ultraviolet (UV) radiation from the sun. High relative humidity leads to a higher number of airborne microbes. Airborne mold and yeast particles in general have a medium equivalent diameter of approximately 2–20 μm. Therefore, gravitational and inertial forces play a major role in their distribution, resulting in higher fungal counts close to surfaces than at higher levels in the air.

The formation of biofilm in air-conditioning systems does not occur without a water reservoir. Normally, there is no water in air-conditioning systems, but it can accumulate through condensation. A study carried out in various industrial ventilation ducts has shown that bacterial and fungal biofilms can accumulate rapidly within the ducts, especially where oil has been left on the duct surfaces. The microbial population in the air channels is dependent on the environment, filtration membranes, and positioning of air. The membranes in the air-conditioning system and the walls in the air-conditioning channels are places where biofilms start to grow.

Biofilm formation causes problems in many areas, such as industrial water systems, medicine, and in the food-processing industry. According to the literature, biofilm problems in food processes have been found in air-handling systems, cooling systems, milk transfer lines, on conveyors, in packaging machines, on vegetable-processing surfaces, heat exchanger surfaces, ultrafiltration and reverse osmosis membranes, in blancher extractors, on mixers, slicers, gaskets, floors, and in drains. In addition to causing problems in cleaning and hygiene, biofilm can cause energy losses and blockages in condenser tubes, cooling fill materials, water circuits, and heat exchangers. Biofilms may enter a food-processing system by causing reduced effectiveness in ion-exchange and membrane processes. In food-processing water supply systems, biofilms cause problems in granular activated-carbon columns, reverse osmosis membranes, ion exchange systems, degasifiers, water storage tanks, and microporous membrane filters. It can be seen from this list that problems generating from biofilm can occur anywhere in the food process if design and maintenance are not carried out properly. Biofilm formation in these systems is a symptom of disturbance in the process. Therefore, equipment design plays the most important role in combating biofilm formations. Dead ends, corners, cracks, crevices, gaskets, valves, and joints are vulnerable points for biofilm accumulation. Poorly designed sampling valves can destroy an entire process or give rise to incorrect information due to biofilm effects at measuring points.

Biofilm can generally be produced by any microbes under suitable conditions, although some microbes naturally have a higher tendency to produce biofilm than others. Common contaminants on food contact surfaces include bacilli, enterobacteria, lactic acid bacteria, micrococci, staphylococci, thermophilic streptococci, and pseudomonads. The bacterial slime of one *Bacillus* strain enhanced the heat resistance of the bacterium, extending the autoclaving time required for successful sterilization to several hours. It is somewhat alarming that pathogens such as *Listeria monocytogenes, Salmonella typhimurium*, and *Yersinia enterocolitica* can readily produce biofilms, causing severe disinfection and cleaning problems on surfaces in the food industry. The most common microbes in cooling waters are the slime-forming *Gallionella* and *Pseudomonas*, as well as common algae. From a hygienic and health point of view, infection and disease problems with *Legionella pneumophila* biofilms can occur in hot-water systems. The slime-forming microbial flora involved in biofilm formation in manufacturing of packaging material can be divided into primary slime-formers, e.g., bacteria of the *Bacillus*, *Pseudomonas*, and *Enterobacter* strains and fungi of the *Aspergillus*, *Mucor*, and *Penicillium* strains, and secondary slime-formers, e.g., bacteria of the *Alcaligenes*, *Flavobacterium*, *Klebsiella*, *Micrococcus*, and *Staphylococcus* strains and fungi of the *Paecilomyces*, *Tricoderma*, and *Trichosporum* strains.

The lubricants used in conveyors are a problem, especially in dairies and breweries. *Acinetobacter* spp., *Algaligenes* spp., *Pseudomonas* spp. and sulfate-reducing bacteria have been isolated from lubricants. The biofilm microbes in the lubricant can indirectly promote corrosion. *Listeria monocytogenes*, an opportunistic pathogen, has been isolated from lubricants in dairies. Yeasts belonging to *Saccharomyces*, *Candida*, and *Rhodotorula* have been isolated from biofilms on conveyor tracks.

There are very few published studies concerning yeast biofilms in food processing. As mentioned above, there are many publications dealing with food-processing biofilms; the biofilms dealt with are mainly bacterial in nature. Yeasts belonging to *Candida*, *Rhodotorula*, and *Saccharomyces* have also been isolated from biofilms on can and bottle warmers in packaging departments of the beverage industry. Additionally, these biofilms contained bacteria and molds. Yeasts have also been observed in biofilms in draught beer dispensing lines at pubs, where they often grow mixed with bacteria. Contaminants introduced into the dispensing system are attracted to the pipe surface by electrostatic interactions but cannot actually adhere due to close-range repulsion. Yeasts overcome this charge barrier by extending surface fimbriae, which anchor them to the conditioning film (**Figure 1**).

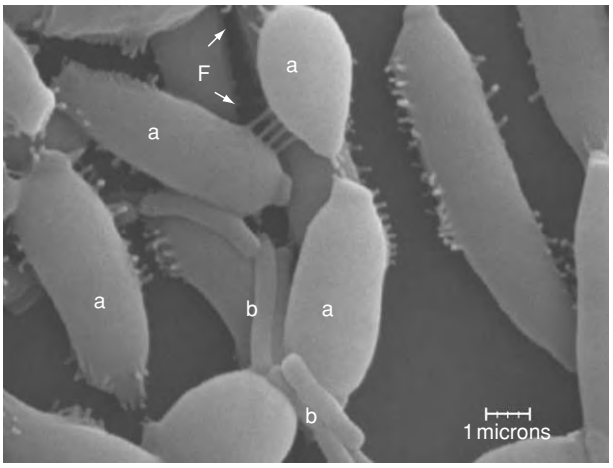

Figure 1 Scanning electron micrograph of a mixed biofilm on a polyvinyl chloride (PVC) surface. a, yeast cells; b, bacteria, F, protruding fimbriae. Reproduced from Storgårds E (2000) *Process Hygiene Controls in Beer Production and Dispensing*. VTT publications 410. Espoo: Otamedia Oy, with permission.

The biofilm formation tendency of *Dekkera anomala*, *Issatchenkia orientalis* (anam. *Candida krusei* var. *krusei*), *Pichia anomala*, *P. membranaefaciens*, *Rhodotorula mucilaginosa*, and *Saccharomyces cerevisiae* isolated from brewery samples has been studied. The biofilm production was found to be strain-dependent rather than species-dependent and additionally dependent on medium and incubation time. The area covered by biofilm varied from 2 to 4% in 10 days for slower-growing strains to approximately 100% in 2 days for fast biofilm producers. Reliable identified strains of the same yeast species differed markedly in their tendency to form biofilm under constant conditions. Sugars included in the media increased biofilm formation. Later, mixed biofilms containing lactic acid bacteria, enterobacteria, and yeasts have been used in cleanability.

The choice of materials and their surface treatments, e.g., grinding and polishing, are important factors in inhibiting the formation of biofilm and in promoting the cleanability of surfaces. The surface structure of stainless steel is very important in avoiding biofilm formation; it has been reported that, although the grain boundaries of AISI 316L stainless steel constitute 3–20% of the total surface area, over 90% of the adherent bacteria were found attached to the grain boundaries. Hoses, tubes, and filters containing polyvinylchloride increase the risk of contamination. The number of free planktonic cells in the water does not necessarily correlate with the amount of biofilm on the pipe surfaces. Costerton *et al*. found that the numbers of microbial cells adhered to surfaces were 500–50 000 times the numbers of planktonic cells in water. Inadequate cleaning and sanitation of surfaces coated with biofilm cause contamination, because the biofilm protects the microbes against sanitizers and disinfectants.

Elimination of Biofilms

Physical, chemical, and microbiological cleanliness are essential in food plants. Cleaning agents are applied to remove soil, microbes, and biofilms from surfaces. Factors governing the selection of detergents and disinfectants in the food industry are that the agent should be efficient, safe, not damage or corrode equipment, be easily rinsable, and not affect the sensory values of the product produced. Physical cleanliness means that there is no visible waste, foreign matter, or slime on the equipment surfaces. Chemically clean surfaces are surfaces from which undesirable chemical residues have been removed, whereas microbiologically clean surfaces imply freedom from spoilage microbes and pathogens. Removal of biofilm is important for the maintenance of the equipment, since the debris left on surfaces can act as nutrients for the build-up of new biofilm. Attached bacteria or bacteria in biofilms can be a problem in food processing, because they adhere to the surfaces, and if cleaning is insufficient the remaining bacteria begin to grow and multiply after cleaning and contaminate the product. To minimize biofouling, it is best to clean the equipment at frequent and regular intervals using suitable, efficient cleaning procedures for the process before the biofilm has the opportunity to develop.

Metal chelators have been used in the break-up of biofilm layers. Disinfection is required in food plant operations where wet surfaces provide favorable conditions for the growth of microbes. Disinfectants approved for use in the food industry contain chlorine-based compounds, iodine compounds, alcohols, quaternary ammonium compounds, oxidants, or surfactants. The use of disinfectants in food plants is dependent on the material used and the adhering microbes. Disinfective agents are not very effective at penetrating the polysaccharide and glycoprotein matrix left on the surfaces after an ineffective cleaning procedure, and thus do not destroy all the microbes within biofilms.

Microbes have also been found in disinfectant solutions. This means that microbial contaminants can be spread on the surface to be cleaned instead of being cleaned. It was reported that chlorhexidine mixtures were contaminated with *Pseudomonas* spp. *Pseudomonas* spp. have also been found in concentrated iodine solutions. Viable *Serratia marcescens* cells were found in a disinfectant containing 2% chlorhexidine after 27 months. Microbial contamination has

also been found in solutions of aldehydes, quaternary compounds, and amfotensides. Other microbes that have been isolated from disinfectants include *Alcaligenes faecalis*, *Enterobacter cloacae*, *Escherichia coli*, *Flavobacterium meningosepticum*, and *Pantoea agglomerans*. *Pseudomonas* spp. are relatively resistant to chlorine treatments and can even multiply when chlorine has been used. Even increased amounts of chlorine ($2.0\,mg\,l^{-1}$) did not kill *Escherichia coli* grown in biofilm. Capsular *Klebsiella pneumoniae* grown on glass surfaces has been shown to have a 150-fold resistance to chlorine compared with that grown in suspensions. On metal and carbon surfaces the corresponding resistance ratios were about 2400 and 3000 times the value in suspension, respectively.

Detection of Biofilm

Many reviews have been published on the formation and detection of biofilms in the laboratory. Methods of studying biofilm formation include microbiological, physical, chemical, and microscopic methods. The biofilm consists of about 85–96% water, which means that only 2–5% of the total biofilm volume is detectable on dry surfaces. If the actual monitoring practice involves only sampling of the liquid phase, it does not reflect the location or extent of microbes growing on surfaces. Advances in molecular biology are producing methods by which detection and enumeration of specific organisms on surfaces can be performed.

Hygiene testing in the food industry is currently based on conventional cultivation using swabbing or contact plates. These classical evaluation methods suffer from several serious deficiencies. Conventional cultivation measures only the number of living cells able to grow on the chosen agar. The quantification of cells in the biofilm is difficult to perform because they adhere strongly to the surfaces (Figure 2). In the cultivation of biofilm microbes it is important that the sample is detached and mixed properly. It has been reported that ultrasonics provide a suitable tool for sampling of biofouled surfaces. Cultivation showed that the use of ultrasonics detached about 10 times the number of cells from the surface compared with swabbing.

Despite extensive efforts to develop reliable control methods for practical hygiene assessment in the food industry, there are still no such methods available. The conventional swabbing procedure of equipment surfaces should be modified, e.g., by chemical loosening of the remaining biofilm cells from the surface at the moment of assessment. A study performed on industrial premises showed that a a mild detergent solution improved microbial sampling. The surface-wetting protocol was developed in laboratory studies using stainless-steel surfaces covered with biofilm. The evaluated test kit comprises a composition designed for the pretreatment of process and equipment surfaces before sampling.

The chemical methods used in the assessment of biofilm formation are indirect methods based on the utilization or production of specific compounds, e.g., organic carbon, oxygen, polysaccharides, and proteins, or on microbial activity (adenosine 5′-triphosphate or ATP) content. ATP measurement is a luminescence method based on the luciferine-luciferase reaction. The ATP content of the biofilm

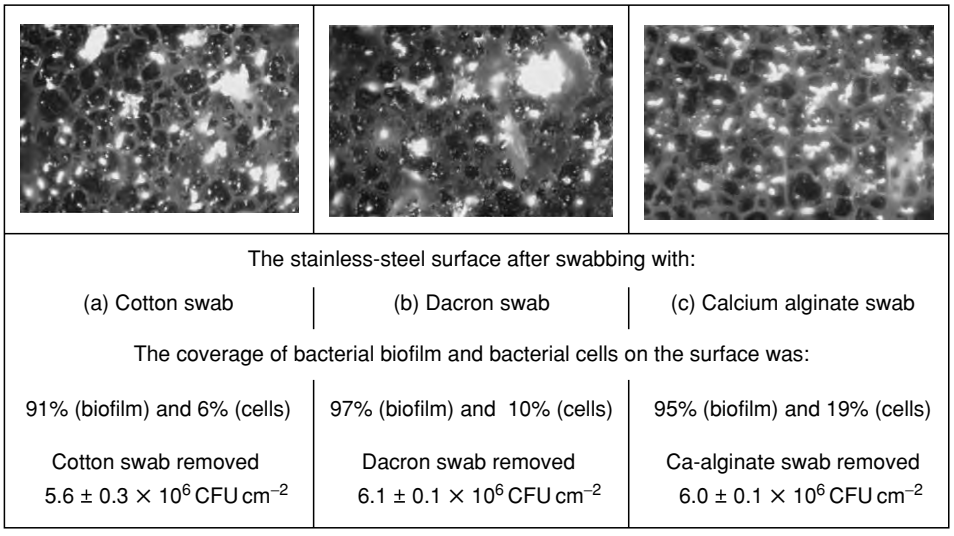

Figure 2 (see color plate 10) Epifluorescence images of 5-day-old *Pseudomonas fragi* biofilm on stainless-steel surfaces (AISI 304, 2B) after sampling with cotton, Dacron, and calcium alginate swabs.

is proportional to the number of living cells in the biofilm and provides information on their metabolic activity. Kinetic data obtained for freely suspended cells should not be used to assess immobilized biomass growth, e.g., biofilm. The ATP method is insensitive and therefore not suitable for hygiene measurements in equipment where absolute sterility is needed. Impedance measurement is used in the food industry to control product quality and to assess the effect of cleaning agents and disinfectants.

Microscopic techniques such as epifluorescence microscopy and, especially, scanning (SEM) and transmission electron microscopy (TEM) are very informative tools in biofilm research and hygiene studies. Electron microscopy has been used to identify biofilm structures. SEM provides accurate information on surface materials and the position of biofilm cells, but it does not provide quantitative data for statistical analysis.

One advantage of epifluorescence image analysis is that it measures the adherent cells on the surface, rather than cells that have been detached by some method from the surface. Epifluorescence microscopy of multilayered biofilms can only be counted two-dimensionally. This technique has been used to determine the effects of antimicrobial agents and biofilm formation by monitoring attachment rates, detachment rates, growth, motility, viability, morphology, and cell area. More accurate information is obtained on the chemical and biological relationships between microbes and their microenvironment *in situ* in real time using confocal microscopy. Laser microscopy permits optical sectioning of biological materials without optical interference from other focal planes. Confocal microscopy has been used to study both growth and metabolism of living cells in biofilms.

Conclusions

Equipment surfaces in the food industry provide the microbes growing in biofilms with the interface needed for growth and with easily obtainable liquids and nutrients needed for growth. Biofilm formation on food-processing surfaces should be combated, using efficient cleaning procedures to improve process hygiene. Cleaning in the food industry should be based on systematic planning. The knowledge that microbes grow differently on surfaces and in suspensions is the first step towards developing advanced regimes in process hygiene. It is very important to bear in mind that the biofilm contains 85–96% water. Biofilms are less likely to accumulate in well-designed systems that can be cleaned effectively. Detection of deposits built up on equipment surfaces at an early stage enables effective countermeasures and thus results in improvement of process hygiene. Methods used for monitoring process hygiene are often based on conventional cultivation, using various types of agar plates or ATP. Conventional cultivation requires several days before the result can be obtained, and it measures those cells that are able to form colonies on the given agar. ATP is used for measuring total hygiene. Many valuable methodological tools used in biofilm research can be applied in industrial monitoring. Online monitoring techniques transferred from various industrial fields, e.g., glass-fiberoptics, infrared techniques, ion mobility techniques, bioluminescence, microelectrodes, and heat transfer could be used for control of microbial and chemical contaminants on process surfaces.

See also: ***Bacillus***: Occurrence; **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant; Overall Approach; ***Escherichia coli***: Occurrence; ***Listeria***: Properties and Occurrence; **Microscopy**: Scanning Electron Microscopy; Transmission Electron Microscopy; ***Salmonella***: Properties and Occurrence; **Yeasts**; ***Yersinia enterocolitica***: Properties and Occurrence

Further Reading

Austin JW and Bergeron G (1995) Development of bacterial biofilms in dairy processing lines. *Journal of Dairy Research* 62: 509–519.

Block JC (1992) Biofilms in drinking water distribution systems. In: Melo LF, Bott TR, Fletcher M and Capdeville B (eds) *Biofilms – Science and Technology*, pp. 469–485. Dordrecht: Kluwer Academic.

Carpentier B and Cerf O (1993) A review: biofilms and their consequences, with particular reference to hygiene in the food industry. *Journal of Applied Bacteriology* 75: 499–511.

Characklis WG (1981) Fouling biofilm development: a process analysis. *Biotechnology and Bioengineering* 23: 1923–1960.

Chisti Y and Moo-Young M (1994) Cleaning-in-place systems for industrial bioreactors: design, validation and operation. *Journal of Industrial Microbiology* 13: 201–207.

Costerton JW, Cheng K-J, Geesey GG *et al*. (1987) Bacterial biofilms in nature and disease. *Annual Review of Microbiology* 41: 435–464.

European Hygienic Equipment Design Group (EHEDG) (1993) Hygienic equipment design criteria. *Trends in Food Science Technology* 4: 225–229.

Gibson H, Taylor JH, Hall KE and Holah JT (1995) *Biofilms and their Detection in the Food Industry*. Chipping Campden: CCFRA.

Gilbert P, Brading M, Walker J and Wimpenny J (eds) (1999) *Biofilms: The Good, the Bad and the Ugly*. Cardiff: BioLine.

Hugenholz P, Cunningham MA, Hendrikz JK and Fuerst JA (1995) Desiccation resistance of bacteria isolated from

an air-handling system biofilm determined using a simple quantitative membrane filter method. *Letters in Applied Microbiology* 21: 41–46.

Lappalainen J, Loikkanen S, Havana M *et al.* (2000) Microbial testing methods for detection of residual cleaning agents and disinfectants – prevention of ATP bioluminescence measurement errors in the food industry. *Journal of Food Protein* 63: 210–215.

Mattila-Sandholm T and Wirtanen G (1992) Biofilm formation in the industry, a review. *Food Review International* 8: 573–603.

Mettler E and Carpentier B (1998) Varations over time of microbial load and physicochemical properties of floor materials after cleaning in food industry premises. *Journal of Food Protein* 61: 57–65.

Salo S, Laine A, Alanko T, Sjöberg A-M and Wirtanen G (2000) Validation of the microbiological methods Hygicult dipslide, contact plate and swabbing in surface hygiene control: a Nordic collaborative study. *Journal of AOAC International* 83: 1357–1365.

Scholte RPM (1996) Spoilage fungi in the industrial processing of food. In: Samson RA, Hoekstra ES, Frisvad JC and Filtenborg O (eds) *Introduction to Foodborne Fungi*, 5th edn, pp. 275–288. Baarn: Centraalbureau voor Schimmelcultures.

Storgårds E (ed.) (2000) *Process Hygiene Control in Beer Production and Dispensing*, VTT publication 410. Espoo: Otamedia Oy.

Tarvainen K, Wirtanen G and Mattila-Sandholm T (1994) Preventing airborne microbial risks in rooms with special hygiene requirements. In: *Healthy Buildings '94*, pp. 217–222. Budapest: Technical University of Budapest.

Wirtanen G (1995) Biofilm formation and its elimination from food processing equipment. VTT publication 251. Espoo: Technical Research Centre of Finland.

Wirtanen G, Salo S and Mikkola A (eds) (1999) *30th R^3-Nordic Contamination Control Symposium*, pp. 1–198. VTT Symposium 193, Espoo: Libella Painopalvelu Oy.

Wirtanen G, Saarela M and Mattila-Sandholm T (2000) Biofilms – impact on hygiene in food industries. In: Bryers J (ed.) *Biofilms II: Process Analysis and Applications*, pp. 327–372. New York: John Wiley-Liss.

Biogenic Amines *See* **Amines**

BIOSENSORS

I E Tothill and A P F Turner, Cranfield University, Silsoe, UK

Introduction

Consumers and legislators are placing farmers, food manufacturers, and retailers under increasing pressure to provide food that is high in quality, free from polluting compounds and chemical residues, and consistent in nutritional value. The demand for fresh natural foods, ready-prepared, and cook-chill foods containing less preservatives and additives, more nutritional value and free from pathogenic microorganisms and genetically modified (GM) material has fueled demands for rapid sensing methods. Food analysis is usually carried out by 'off-site' methods, in which the samples are sent to a laboratory for testing. These methods suffer from the disadvantage of being expensive and time-consuming, and require the use of highly trained personnel. However, central analytical facilities do generally provide the highest accuracy in quantification and the lowest detection limits. Many assays are available on the market today that promise results within 24 h or less, such as immunoassays, polymerase chain reaction assays (PCR), and adenosine triphosphate (ATP) detection methods. These assays are used to measure the concentration of a variety of hormones, allergens, viruses, bacteria, and chemical pollutants. However, for today's environment and lifestyle, there is an increasing need for more easy-to-use, robust, rapid, and relatively inexpensive technologies. The availability of these assays would facilitate 'on-site' sample analysis, and the number of samples requiring further analysis using laboratory-based instruments could then be reduced. Hence, there is a clear need for simpler, alternative analysis systems and preferably methods that allow real-time monitoring in the field such as biosensors.

Clark and Lyons pioneered the underlying concept of the biosensor in a paper published in 1962. They reported an 'enzyme electrode,' comprising an immobilized enzyme at an electrochemical detector, and

hence produced a sensor that was responsive to the substrate of the enzyme. Since then, the field of biosensors has greatly expanded due to advances in signal transduction technologies and also to huge advances in the receptor technology used as the sensing element. There is now a wide array of biological and synthetic sensing receptors and signal transducers available to construct a particular biosensor. The approach generally employed is to identify the analytical target and then to choose the sensing molecule that gives the required affinity, selectivity, and dynamic range for the analyte or analytes of interest. The assay format signal transduction strategy can then be selected based on the analytical requirements of the final device (Table 1).

The study of biosensors has been strongly driven by the needs of medical diagnostics. Instant analysis of medical samples can provide better patient care and disease management. There has also, however, been a wide appreciation of biosensor technology in other areas such as veterinary care, animal husbandry, food and process control, and environmental analysis. While work initially concentrated on biosensors for the aqueous phase, more challenging environments such as the gas phase and organic solvents have been explored recently. The diagnostics industry is very diverse, with numerous markets requiring different products. The main controlling influences are price of both the instrumentation and the test, the accuracy, sensitivity and number of parameters required, speed, and portability. The types of instruments required for the diagnostics market can be divided into large multianalyte analyzers, bench-top portable instruments and one-shot disposable sensors. Many of the instrumentations developed to date have been designed for the medical diagnostics market. However, a number of these have been adapted for the agrofood market. In addition, the market for diagnostics continues to grow. To date, an astounding diversity of research on biosensors has been reported in the literature, and these papers offer an almost infinite number of suggestions for biosensor construction.

Table 1 Components used for biosensor construction

Receptors (recognition elements)	*Transducers*
Tissues	Optical
Microorganisms	Electrochemical
Organelles	Thermometric
Cell receptors	Piezoelectric
Enzymes	Magnetic
Antibodies	
Nucleic acids	
Synthetic receptors	

Definition

Biosensors are analytical devices incorporating a biological material, a biologically derived material or a biomimic as the recognition molecules, which is either intimately associated with or integrated within a physicochemical transducer or transducing microsystems. The usual aim is to produce a digital electronic signal that is proportional to the concentration of a specific analyte or group of analytes (Figure 1). This signal can result from a change of proton concentration, uptake or release of gases such as oxygen and ammonia, light emission, reflectance or absorption, heat emission, or other mechanisms, brought about by the action of the sensing molecules. This signal can then be converted by the transducer into a measurable response. Both the biological and the electrical signal can be manipulated further by amplification or processing. While the signal may in principle be continuous, devices can be configured to yield single measurements to meet specific market requirements.

The analytical capabilities of biosensors have been further increased by miniaturization and the improved processing power of modern microelectronics. Different biosensor formats have been developed for single target analytes and for broad-spectrum monitoring. From this definition, it can be seen that a number of different biosensor devices either already exist or are theoretically possible.

Sensing Elements Used in Biosensors (Receptors)

Natural Receptors

There is a wide range of naturally produced molecules from plants, animals, and microorganisms that can be used as the receptor in biosensors. These molecules may be broadly classified into two groups: the biocatalytic group (e.g., enzymes, whole cells, mitochondria, and tissue slices); and the affinity reaction group (e.g., nucleic acids, antibodies, and cell receptors). In each case, the biosensor exploits the selective binding capabilities inherent in biomolecules. Table 2 list the most frequently used sensing molecules and their respective analytes. In catalytic sensors, the change in the concentration of a component resulting from the catalyzed reaction is detected to give the sensor signal. In the case of an affinity sensor, the binding event itself (between the receptor and the target analyte) is monitored.

Because of their specificity and catalytic properties, enzymes have found widespread use as sensing elements in biosensors (enzyme biosensors). Since

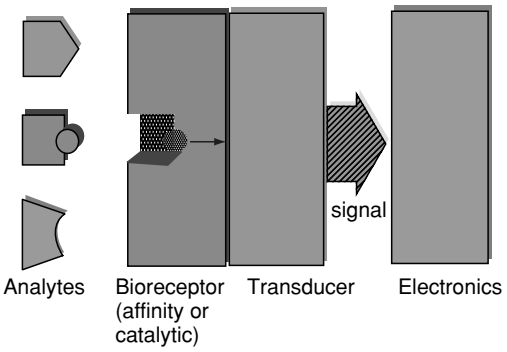

Figure 1 The biosensor.

Table 2 Biological systems most frequently used in biosensor construction for sensing their respective binding substance

Sensing molecule	*Analyte to be detected*
Enzyme	Substrate analog, inhibitor, cofactor, metal chelaters, dyes
Antibody	Antigen, virus, cell
Nucleic acid	Complementary base sequence, histone, nucleic acid polymerase, binding protein
Cell	Cell-surface-specific protein, lectin
Receptor, carrier protein	Hormone, vitamin

the development of the first glucose sensor based on the immobilization of glucose oxidase on an oxygen-sensing electrode, there has been a proliferation of applications involving a wide variety of substrates. A variety of enzymes belonging to classes of oxidoreductases, hydrolases, and lyases have been integrated with different transducers to construct biosensors for applications in health care, veterinary medicines, the food industry, environmental monitoring, and defence. However, the most important group has been the oxidoreductases, catalyzing oxidation or reduction events using either oxygen or cofactors. Enzymes are either used alone, as in catalytic biosensors, or used in conjunction with other components, such as antibodies as the signal marker.

Whole cells of living organisms, such as bacteria, yeast, fungi, plant and animal cells, or even tissue slices, have been used as the recognition component in biosensors (whole-cell sensors). The sensing principle is usually based on interrogating the general metabolic status of the cells by detecting oxygen or substrate consumption, the production of carbon dioxide or metabolites, bacterial luminescence, or direct electrochemical sampling of the electron transport chain. A range of whole-cell or microbial sensors have been constructed to detect toxic and pollutant compounds. Interest in these types of devices for environmental and toxicity monitoring is increasing because of their ability to monitor directly biological effects, as opposed to simply providing a total chemical analysis, which may not be related to the bioavailability of the contaminant.

Antibody–antigen binding reactions are used mainly in affinity sensors. However, other biological components such as cell receptors, single-stranded DNA (to bind and detect its complementary sequence), and lectins (plant proteins used to bind carbohydrates) are also being used. Combinations of the above biological components have been utilized to provide new or improved analytical capabilities. The developments of polyclonal antibodies (Pabs) and the huge advances in monoclonal antibodies (Mabs) and recombinant antibody production have created a major force in the diagnostics market. Antibodies are widely applied in a range of devices from rapid wet chemistry assays such as immunoassay kits, to dipsticks and biosensors. The use of antibody fragments and molecularly engineered antibodies is a future area of growth for immunosensors. A new research approach is to recover the genes of useful antibodies, express them in bacteria or plants, develop three-dimensional structural models, and then derive improved variants by directed and combinatorial mutagenesis. The affinity and selectivity of the recombinant antibody can thus be adapted to make it more suitable as an analytical tool.

The main disadvantage of using biological molecules in sensor devices is their poor stability, which is often the stumbling block preventing wider commercialization. Research in improving the stability of these compounds is expanding, and methods include the use of soluble, positively charged polymers, such as diethyl amino ethyl, dextran, lacitol, and sugar derivatives. A new approach to overcome the stability problem of biological molecules is to replace them with artificial receptors or biomimics.

Artificial Receptors (Biomimics)

New advances in receptor discovery such as the development of artificial receptors using the combined approach of computer (molecular) modeling and molecularly imprinted polymers or combinatorial synthesis has increased the range of receptors that can be used for the construction of suitable sensing layers for biosensors.

Molecular modeling is a powerful tool to study molecular recognition, the specific interaction between substrate and receptor. The affinity interactions between synthetic receptors and target analytes comprise hydrogen bond interactions, π-stacking interactions, Van der Waals interactions and electrostatic interactions, which constitute molecular forces

involved in molecular recognition processes. To design the artificial receptors, computer modeling is used to provide structure information for the target analyte, which will then be used to guide the design of combinatorial libraries and rational design of the artificial receptors. Molecular modeling allows the prediction of ligands that are expected to bind strongly to key regions of biologically important molecules (e.g., enzymes, macromolecular receptors) of known three-dimensional structure, so as to inhibit or alter their activity.

Molecular-imprinting polymerization is currently the focus of intense research interest and is being used in a wide range of application areas, e.g., in the preparation of selective separation materials, artificial antibodies, and synthetic enzymes and in molecular recognition. Molecular imprinting relies upon the presence of complementary interactions (noncovalent or reversible covalent) between sites in the template molecules (the analyte) and the functional monomer(s) used in the polymerization process (Figure 2). The functional monomer(s) is a key component in the imprinting process, though generally constituting a relatively small percentage of the resultant MIP. The complexation of a template by these building blocks prior to, and during, polymerization to yield the complementary positioning of functionalities, is the central dogma of molecular imprinting.

The combinatorial library technique (combinatorial chemistry) is a fast developing area of receptor discovery and has been widely used by the drug industry to both discover and optimize lead compounds. An example is the optimization of the antidepressant drug, benzodiazepine. Combinatorial libraries consist of a large array of diverse molecular entities, generated by the systematic and repetitive covalent connection of a set of different 'building blocks' of varying structures. The number of possible compounds (n) generated in a single library can be found using the relationship:

$$n = b^x,$$

where b represents the number of building blocks for each step and x, the number of synthetic steps.

Enormous effort has been devoted to the development of new strategies for peptide and nonpeptide libraries. The principle is illustrated in Figure 3. A large number of compounds can be generated in a very short time using combinatorial library techniques, which would be near impossible using classical organic synthesis methodologies. Molecular modeling is also used as a combined approach with combinatorial chemistry to facilitate a more efficient receptor discovery process. Combinatorial chemistry has been used to generate affinity ligands, which can be used as synthetic receptors in biosensor devices.

Research activity in the area of artificial receptor discovery has increased in the last few years, in particular the production of synthetic receptors for medical and environmental diagnostics. The development of artificial receptors for various purposes remains an important challenge. Combinatorial synthesis and molecularly imprinted polymers are two of the most exciting and rapidly growing areas in ligand discovery, which can overcome the stability problems inherent in natural ligands. However, the challenge is to produce receptors that compete with the natural molecules with respect to sensitivity and stability.

Transducers

The transduction element of a biosensor must be capable of converting a specific biological response

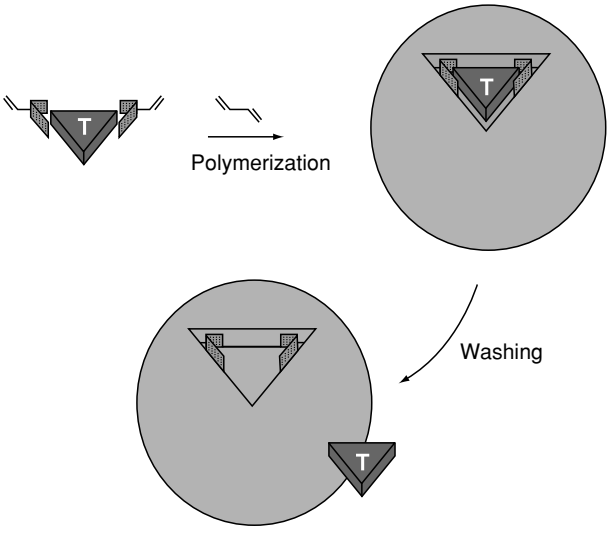

Figure 2 Molecularly imprinted polymers (MIPs).

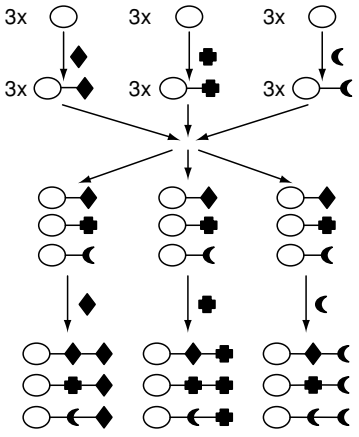

Figure 3 Combinatorial chemistry.

into a quantifying signal. The transducers must also be suitable for receptor immobilization at, or close to, its surface. A list of transducers that have been used in the construction of biosensors is given in Table 3. Considerable work is being carried out to improve transducer designs, with the blending of a variety of different technologies. The operation principle of commonly used transducers is described briefly below.

Electrochemical Transducers

Electrochemical devices usually monitor the current at a fixed voltage (amperometry) or the voltage at zero current (potentiometry), or measure conductivity or impedance changes. Amperometric biosensors were the first type to be developed and have been in use as glucose biosensors for over 35 years. These devices have continued to be the most popular, largely due to their simplicity, ease of production, and the low cost of the devices and instruments. The signal in amprometric devices is dependent on the rate of mass transfer to the electrode surface. The most commonly used devices monitor the analyte concentration by measuring the fall in O_2 tension, the production of H_2O_2 or determination of the redox reaction of a mediator molecule (electron shuttle). A large number of mediators are available, but the most widely used include potassium ferricyanide, tetrathiafulvalene (TTF), tetracyanoquinodimethane (TCNQ), and the ferrocenes.

A potentiometric sensor operates under conditions of near-zero current flow and measures the difference in potential between the working electrode and a reference electrode. Ion-selective electrodes (ISEs), of which the pH electrode is a well-known example, are the most important of this class of transducer.

Table 3 Transducers used in biosensors

Transducer	*Output*
Electrochemical (electrodes)	
Amperometry	Applied current
Potentiometry	Voltage
Conductimetry/impedimetry	Conductivity/impedance
Optical (optrodes)	
Colorimetric	Color
Luminescence	Light intensity
Fluorescence	Light intensity
Interferometry	
Calorimetric	
Thermistor (heat-sensitive sensor)	Temperature
Mass	
Piezoelectric (surface acoustic wave devices)	Acoustic wave
Magnetic	

Specificity is conferred by selective membranes, which may be formed from metal salts or polymer membranes containing ion-exchangers or neutral carriers. To date, there is a range of commercially available ISEs that can detect specific ions (i.e., calcium, nitrate, potassium, copper, barium, chloride, etc.). Devices can also include the gas-sensing electrodes (GSEs), which are ISEs modified by the use of a gas-permeable membrane, and the chemically sensitive field-effect transistors (FETs), which are semiconductor devices that respond to the surface electric gradient or charge at the gate electrode. More recently, light-addressable potentiometric sensors (LAPSs) have been developed, based on silicon technology. A chemical reaction at the surface of the silicon chip will shift the surface potential and hence affects the current, which allows the rate of the reaction to be monitored.

Conductimetric biosensors are based on measuring the time dependence of the change in conductivity as a result of the receptor recognition of its complementary analyte. Impedance is the total electrical resistance to the flow of an alternating current being passed through a given medium. Typically, during measurement, impedance decreases, while conductivity and capacitance increase.

Optical Transducers

Optical transducers use a number of principles, such as the effect of the biological event on light absorption, fluorescence, refractive index, or other optical parameters. These types of transducers have become increasingly popular over the last few years with many devices now commercially available. The early-developed optoelectronic devices utilize a color change induced in pH-sensitive dyes by the action of an enzyme that generates or consume protons. More recent approaches to measure pH changes related to the analyte use optical sensor devices based on dynamic quenching by molecular oxygen of a fluorescence of an indicator.

Most of the recent work on optical biosensors has concentrated on the use of evanescent wave technology. When light is totally internally reflected during, for example, transmission through an optical waveguide, a small amount of energy propagates beyond the optical interface. In an unclad waveguide, this can be coupled to fluorescent labels constrained close to the optical interface as a result, for example, of antibody–antigen interactions. Alternatively, a direct immunosensor can be constructed, which does not require the use of labels, by measuring the change of refractive index due to the sensing receptor interaction with the target analyte. Surface plasmon resonance (SPR) is currently receiving the most interest

in this area and forms the basis of a successful commercial instrument. The technology is based on the excitation of the electron plasma (surface plasmon) of a thin metal layer covering the surface of the waveguide. The change in angle of the incident light or the reflectance minima due to the change in the refractive index in the vicinity of the metalized surface can be monitored.

Calorimetric Transducers

Thermometric devices operate by measuring enthalpy changes during the biological reaction using one of a range of thermometers, thermopiles (array of thermocouples), or thermistors.

The detection of heat during the enzyme catalysis can be used in the construction of calorimetric enzyme biosensors. Since heat is either produced or consumed in most reactions, these can be converted to analytical signal to monitor analyte concentrations.

Piezoelectric Transducers

Sensors based on piezoelectric principles use the change in the resonant frequency of wave propagation through a piezoelectric material. These principles can be used to measure mass, viscosity, or density changes at the sensor surface. These devices are able to generate and transmit acoustic waves in a frequency-dependent manner. Bulk wave devices operate by transmitting a wave from one side of the crystal to the other, while surface acoustic wave (SAW) transmit waves along a single crystal face.

Magnetic transducers

Very recently, a few papers have been published describing biosensors based on magnetic transduction. Hall-effect transducers, for example, have been used to detect magnetic labels in immunosensors. Practical application of this type of devices has been limited, but they are included here for completeness.

Fabrication of Biosensors

The commercial production of devices such as the ExacTech home blood-glucose biosensor (MediSense, Inc., Cambridge, MA) has verified the importance of automated manufacturing technologies where large numbers of inexpensive, reproducible electrochemical devices are required. The capability of printing materials at high precision and speed is very desirable for the mass production of analytical devices such as biosensors. Techniques such as screen-printing, ink-jet printing, air-brushing, and Cavro deposition have been developed under microprocessor control and adapted for biosensor fabrication, especially

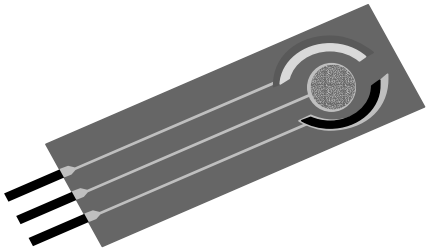

Figure 4 Screen-printed biosensor.

thick-film biosensors (Figure 4). These deposition techniques are rapid and flexible with regard to the ink characteristics employed. Thin-film deposition has also been used, which involves the application of material through a mask, under vacuum, via evaporation due to heating, or by placing the substrate to be coated between two electrodes (sputtering). This process is able to create patterns less than 1 μm thick, making it suitable for the construction of microsensors. Thin-film sensors are generally produced by a variety of vapor-deposition techniques, electrochemical methods, and, more recently, by the use of Langmuir–Blodgett technology. Thin metallic films, usually deposited by vapor deposition, have been used extensively for the production of sensors utilizing surface plasmon resonance. These sensors have been primarily directed towards affinity sensors.

Silicon fabrication technology has been increasingly used as a means of mass production of biosensors. The use of silicon microfabrication for both electrochemical and optical sensors is expanding and the capability of on-chip electronic signal amplification and data processing is very attractive. This is an area of intense research activity mainly by diagnostics companies. Several companies are using microelectronics technology based on silicon to produce optical chips for biosensor application.

Biosensors Based on Biocatalitic Interactions (Biocatalytic Sensors)

The biosensor market has been dominated for many years by electrochemical biocatalytic sensors. The majority of biocatalytic sensors are based on enzyme or whole-cell sensors. Amperometric enzyme biosensors form the majority of commercial biosensors available today. In the enzyme electrode, for example, an enzyme specific for the substrate of interest is immobilized on the sensor surface. The chemical conversion of the analyte is catalyzed as it enters the enzyme layer, and either the cosubstrate or the product of the reaction is transduced to an electrochemical signal, which is proportional to the concentration of

the substrate. The simplest biosensor approach is to measure amperometrically the oxygen depletion or hydrogen peroxide production associated with the reactions involving oxidase enzymes. The combination of simplicity, ease of manufacture, high sensitivity, availability, and low cost of instrumentation is highly attractive in this type of devices. A variety of enzymes belonging to classes of oxidoreductases, hydrolases, and lyases have been applied in enzyme-based sensors. These enzymes can act on fatty acids, sugars, amino acids, organic acids, and phenols. Biosensors based on the use of redox mediators to shuttle the electrons, resulting from the catalytic oxidation or reduction of the substrate of interest, to the electrode has been an area of great interest to many researchers. Much of the work is centered on the use of redox mediator compounds such as ferrocene, TTF, TCNQ, hexacyanoferrate (III), methylene blue, quinones, and their derivatives. Careful selection of the mediator can also eliminate interference problems from other substances in the sample.

To date, there are several successful biosensor devices on the market based on the different amperometric transduction principles outlined above. These include the MediSense mediated blood glucose analyzer for medical application and the Yellow Springs hydrogen peroxide-based systems (YSI, Yellow Springs, OH) developed for both medical and food analysis and for bioprocess control. The YSI 2700 SELECT™ Biochemistry Analyser (**Figure 5**) can be used for on-line fermentation monitoring or off-line sample analysis for food and drink fermentation. Nutrient and bioproducts analysis can be carried out using the biosensor system including glucose, L-lactate, L-glutamine, L-glutamate, ethanol, lactose, sucrose, galactose, and methanol. Other commercially available biosensors for food analysis include; Sensomat B10, for alcohol analysis (Gonotec, Germany); Microzym-L for lactate analysis (Kalger GmbH, Germany) and Glu-11 for glucose analysis (TOA Electronics, Japan). A range of single and multianalyte biosensors based on the same principle are being developed for a diverse range of applications in the food industry.

Figure 5 YSI 2700 SELECT™ Biochemical Analyzer. Photo courtesy of YSI incorporated, Yellow Springs, USA.

The majority of biosensors developed for food analysis are water-based systems. Organic-phase enzyme electrodes, however, have been developed to work in organic solvents and have been used for the analysis of compounds such as cholesterol, alcohols, organic peroxides, and phenols. This technique has the advantages of ease of sample preparation from materials, such as fats and oils, and increased enzyme stability. In principle, analytes soluble in fatty substances can also be measured using biosensors. Enzyme biosensors have also been widely applied to the detection of phenolic compounds. Contamination by these aromatic compounds results from the production of drugs, dyes, and antioxidants and also from the paper and pulp industry. Enzymes used in biosensor development for this type of compounds are phenol oxidases (tyrosinases), laccases, and peroxidases.

Biocatalytic sensors based on enzyme inhibition are the most commonly reported biosensors for the detection of toxic compounds and heavy metal ions. These sensors are based on the selective inhibition of specific enzymes by classes of compounds, or by more general inhibition of enzyme activity. The measurement of the enzyme activity can be used as a screening test to evaluate the contamination by compounds such as organophosphorous and carbamate insecticides and the triazine herbicides and other compounds that display similar toxicological behavior. Enzymes used include: choline esterase, horse-radish peroxidase, polyphenol oxidase, urease, and aldehyde dehydrogenase.

Biosensors based on the use of whole cells as the sensing layer generally respond to more than one analyte, but can offer increased stability when compared with enzyme-based sensors. Whole-cell biosensors have been utilized commercially for environmental monitoring and capitalize on the broad sensitivity of microorganisms to a wide range of toxins. Whole-cell biosensors may contain living bacteria, yeast, fungi, plants, or animal cells. Biosensors able to monitor indicators of pollution, such as biochemical oxygen demand (BOD), have been developed. Sensors such as the ARAS BOD Biosensor System (Dr Bruno Lange GMbH, Düsseldorf,

Germany), the BOD module biosensor produced by (Prufgerate-Werk Medigen, Germany), and Nissin's (Japan) BOD Sensor are commercially available. A range of biosensors based on this principle have been developed for polluting compounds such as pesticides, in particular for water monitoring.

Electrochemical sensor approaches to hygiene monitoring based on the metabolism of living cells have also been commercialized. The Bactometer (Bactomatic Inc., Princeton, NJ) and the Malthus 2000 (Malthus, Stoke-on-Trent, UK) use impedance and conductance, respectively, to plot bacterial growth. A more rapid approach is the use of amperometry. The method comprises the use of mediators to detect the metabolic activity of the cells, and the resulting current can be related to the numbers of organism present in the sample.

Catalytic optical biosensors are usually based on optical fibers or planar wave-guide films. Various fiber-optic biosensors based on pH opt(r)odes have been constructed for glucose, urea, penicillin, and creatinine analysis. Commercial examples of fiber optic sensors for monitoring blood pH, dissolved oxygen and carbon dioxide concentrations. The transfer of these technologies to the agrofood and environmental sector can be anticipated in the near future.

Affinity Sensors

Antibodies have been used as recognition elements in tests and assays since the 1950s and in biosensor devices since 1970s. In the past decade, there has been phenomenal growth in affinity-based sensor research and development. Affinity-based sensors can be define as analytical devices that use an antibody, receptor protein, or other molecules with affinity-recognition powers interfaced to a signal transducer to measure a binding event. Various research groups have pioneered different types of affinity-based sensor configurations. By utilizing the immunorecognition properties of antibodies in immunosensors and affinity sensors, the range of analytes that can be measured has been broadened considerably.

Immunosensors offer a wide range of potential applications to the food industry, water companies, and regulatory authorities. In immunosensors or immunoprobes, the antibody (Ab) or the antigen (Ag) constitutes the biospecific component in the sensor structure. A range of immunosensors are being developed, and with minor adaptations, these can replace immunoassays. At present, immunodiagnostic tests are used in pollution detection in food and water samples and also on-farm monitoring of livestock reproduction (milk progesterone), and quality control of foodstuffs originating from livestock production (authenticity and adulteration testing). These tests can be developed to a new generation of rapid immunosensors for real-time and on-site applications. A range of research projects have been undertaken to develop immunosensors for residues in food (antibiotics, toxins, and pesticides), the presence of additives, and hormones. The availability of the required antibody can limit the potential for diverse analyte detection by immunosensors. Increased research in the development of specifically tailored antibodies such as plantibodies produced in plants, recombinant antibodies, catalytic antibodies or abzymes, artificial receptors, and molecularly imprinted polymers should overcome some of these problems.

Immunosensors can be divided into classes depending on the transducer technology employed. Electrochemical immunosensors generally rely on the use of electroactive labels, usually based on enzyme labeling and amplification techniques. The sensors can be inexpensive and may achieve low detection limits (< 1 p.p.b.). Different types of electrochemical immunosensors and affinity sensors have been developed for environmental analysis using various transducers. The scientific literature increasingly contains reports of the development of one-shot disposable or on-line immunosensors for a diverse range of analytes. However, the technology required to manufacture these types of sensors has not yet reached the level of sophistication achieved for enzyme electrodes, and hence, these devices are at an earlier stage of commercialization.

Optical devices have a clear advantage over electrochemical methods due to their ability to monitor binding reactions directly. The major drawback to the application of optics remains the high cost of many optical components, but these costs are constantly falling. Optical phenomena such as SPR and evanescent wave (EW) technology have shown promise in providing direct measurement of Ag–Ab interactions occurring at the surface–solution interface, the surface usually consisting of a glass prism on a gold or silver metal layer. SPR is a phenomenon that occurs when a beam of light is directed on to a glass–metal interface, which results in changes in the resonance angle. The BIAcore™ biosensor system based on SPR technology was developed by a Pharmacia (Uppsala, Sweden) spin-off company and is now commercially available from Biacore AB (Uppsala, Sweden). This instrument represents a significant breakthrough in optical immunosensor technology. BIA technology enables detection of biomolecules in real time without the use of labels, but the instruments are expensive and mainly laboratory-based, and require trained

personnel to interpret the results. The Biacore system has been reported to be able to detect atrazine at $0.05\,\mu g\,l^{-1}$ and *E. coli* O157 in food samples. Recently, Biacore AB have launched the BIAcoreQuant™, for the automated analysis of vitamins in foods (Figure 6). The assay is designed as an inhibition assay, and similar principles are applied for the determination of drug residues in meat and milk products using BIACORE 2000. Amersham International plc has been active in research into SPR-based immunosensors for infectious diseases and has developed immunotechnology that employs antibodies labeled with latex particles (beads). The beads amplify the change in refractive index at the sensor surface–solution interface, which occurs when the antibodies bind to the immobilized antigen layer and thus cause a large change in the resonance angle. An alternative direct optical immunosensor to compete with SPR technology has been developed by Affinity Sensors (Cambridge, UK). The resonant mirror technology is marketed as the IAsys (IAsys and IAsys *Auto*+) and also enables analysis of biomolecular interactions in real-time. Reusable cuvettes offer a choice of derivatized sensor surfaces and chemistries. The ligand is simply and efficiently immobilized on to the sensor surface, and binding of the analyte can be detected immediately.

The alternative approach of using fluorecent labels detected via evanescent wave interactions with an optical fiber has also been applied successfully. One example is the Raptor developed by the US Navy Research Laboratory for the detection of biological

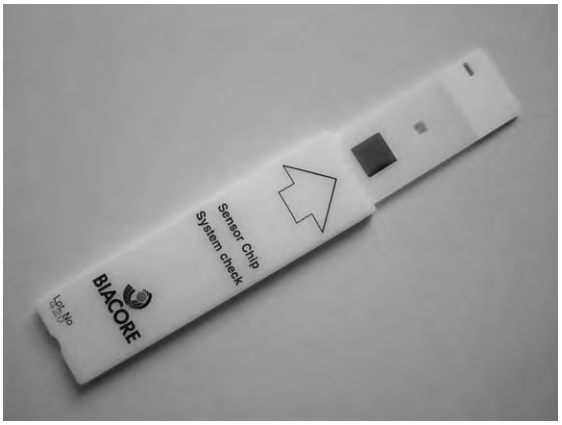

Sensor chip

(a)

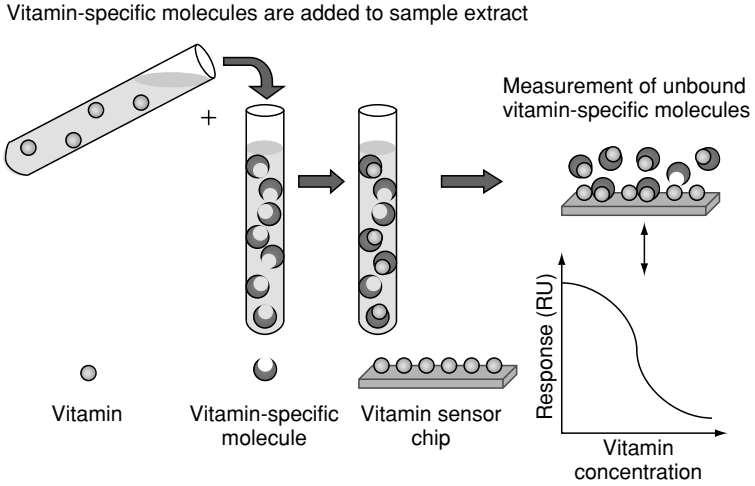

(b)

Figure 6 The BiacoreQuant™ surface plasmon resonance system for the automated analysis of foods, (a) the sensor chip, (b) the procedure. Photo courtesy of Biacore AB.

warfare agents. Unilever (Bedfordshire, UK) has developed a fluorescence-based EW immunosensor that incorporates a novel capillary-fill design. The system consists of two glass plates separated by a narrow capillary gap of $\sim 100\,\mu m$. The lower plate acts as an optical waveguide and contains an immobilized layer of antibodies on its surface. These sensors benefit from the capillary fill system, which draws a fixed volume of sample into the space between the plates, regardless of bulk sample volume.

Piezoelectric transduction approaches and, in particular, SAW devices have also been widely used to detect antibody binding to an immobilized antigen. The most widely used acoustic transducer is the quartz crystal microbalance, which measures small changes in surface properties, such as bound surface mass and viscosity, resulting from the binding of molecules to the sensor surface. Again, the attraction of piezoelectric sensors is their ability to monitor directly the binding of Ab–Ag reaction encountered in affinity sensing, but they suffer from the disadvantage of having a low sensitivity for small molecules.

Potential areas of application for affinity-based sensors are similar to those for biocatalytic sensors. However, affinity sensors can offer enhanced sensitivity and specificity when compared to biocatalytic sensors. Commercial immunosensors may not have been developed specifically for use in the agrofood and environmental sectors yet, but their long-term potential is evident. As the commercial success of immunoassays becomes more evident in health care, food, and environmental monitoring, the demand for faster techniques will be sufficient for continued affinity sensors development. The development of disposable electrochemical affinity sensors for the rapid detection of residues of anabolically active illegal androgens, antibiotics, and other residual compounds in food samples would allow regulatory bodies to improve their screening program. A potential market for immunosensors is genetically modified (GM) foods testing for protein analysis. Immunoassay tests as a plate test or dipstick devices for the detection of GM foods in plant tissues, raw agricultural commodities, and also food ingredients are marketed today by SDI Europe Ltd. (Hampshire, UK). These types of tests may well be developed into immunosensors devices in the near future.

DNA Biosensors

DNA is a double-stranded helical molecule that can be split into its two complementary strands by treatment with alkali. By using the correct environmental conditions, the two strands can come together again (reassociate) in a process that forms the basis of all DNA hybridization techniques. This technique involves a small quantity of a pure DNA probe that is labelled and made single-stranded so that it can seek out and hybridize to its complementary sequence in a large amount of DNA. The probes can be labelled with an enzyme marker. Nucleic acids, the building blocks of DNA and RNA, are universally present in all living cells and are utilized as a general method of microbial detection and identification. PCR is a widely used method for amplifying trace amounts of DNA for analysis. To date, it has been used mainly for qualitative analysis, but the need for quantitative information has resulted in further development of PCR protocols. Quantitative PCR methods are shown to be extremely sensitive, and fully automated detection systems that should be rapid, simple to use, and inexpensive are being developed. DNA biosensors are currently used in the detection of infectious diseases and genetic abnormalities, where the main approach has been by the use of immunobased sensors. This technology is well established, and the performance of these tests is good. PCR is widely used to amplify the signal in DNA probes, but it is time-consuming. There is intense current interest in microsystems for DNA analysis. An electrochemical DNA sensor based on a gold electrode modified with DNA probes and an electroactive hybridization indicator has been reported in the literature. A range of detection systems have also been applied in DNA probes such as the use of enzyme label, an SPR-based biosensor, an EW biosensor and an acoustic-based sensor. Affymetrix (Santa Clara, CA) has developed a platform for acquiring, analyzing, and managing complex genetic information. The system is based on disposable DNA probe arrays (GeneChip) containing selected gene sequences on a chip, and an instrument to process each chip, and analyze the information. This 'lab-on-a-chip,' offers traditional, laboratory-based biological assays in convenient on-site or on-line instruments. DNA chips are used for microorganism identification or detection of GM foods. The company is developing new chips for microbial contamination diagnosis in environmental samples. GeneScan Europe (Hanse Analytik GmbH, Freiburg, Germany) markets test kits for the detection of genetically modified components in human and animal foods by DNA analysis using PCR techniques. These tests can also be developed into DNA biosensor devices.

Electronic Noses

The application of the electronic nose concept or artificial olfaction has been of tremendous emerging importance in recent years. Electronic noses are based

on the use of artificial receptors in the sensor array, and their application in food analysis justifies their inclusion in this article. In recent years, a great deal of research in the development of the electronic nose has taken place with a diverse range of applications. This type of sensor instrument mimics the olfactory system in the nose. The instrument consists of an array of gas sensors with different selectivity patterns, a signal collecting unit, and data analysis software, which analyses the signal by pattern-recognition methods, such as principal component analysis, discriminant function analysis, cluster analysis, and artificial neural networks. The results are comparative, rather than quantitative, and are presented as a 'fingerprint.' A variety of sensors are used, ranging from metal oxides to conducting polymers, which can be highly sensitive but not specific and can respond to volatile compounds with molecular weight ranging from 30 to 300. Molecules such as alcohols, ketones, fatty acids, and esters give a strong response, whereas fully oxidized species, such as CO_2, NO_2, and H_2O have a lower response. The sensor array can also recognize molecules containing sulfur and amine groups. Furthermore, they are not in direct contact with the measuring media, since they are used for gas-phase measurements.

There has always been considerable waste during the storage and handling of food. Moreover, food often constitutes an ideal growth medium for microorganisms. The composition of volatile compounds evolved from the food can reflect the activity and type of microorganisms present and also can be related to food quality. Thus, electronic noses appear to be a useful means of detecting these kinds of changes. Electronic noses have been used in medical, environmental, and food diagnoses. The majority of the current research concentrates on quality control in the foods and drinks industry, such as the detection of microbial contamination (bacteria, fungi, and yeast), freshness of meat and fish, and authenticity (beverages, coffee, and meat).

Conclusions

The food and drink industry deals with a wide range of materials and products with different characteristics and diverse processing procedures. The shared theme, however, is that the raw materials are organic, subject to environmental pollution, degradations and microbial contamination. The food diagnostics market is expanding rapidly and covers a wide range of disciplines. The establishment of appropriate technologies to apply biosensors to practical agriculture and horticulture is expected to produce a significant effect on quality improvement and cost reduction. Biosensor applications in the medical diagnostics market have been highly successful, but their potential success in the food, agriculture, veterinary diagnosis, and environmental market still remains to be established. Biosensor systems, which are relatively small portable instruments, have an on-site application, and their relatively low cost is clearly desirable in agrofood analysis. Research is being undertaken in diagnostic companies and research institutions to develop biosensor technologies for the food and environmental sector. Nevertheless, moving the technology to the market place involves many challenges. Biosensors will have to offer perceived advantages over existing and competing technologies.

See also: **Enzymes**: Functions and Characteristics; **Nucleic Acids**: Properties and Determination; **Genetically Modified Foods**.

Further Reading

Cunningham AJ (ed.) (1998) *Introduction to Bioanalytical Sensors*. New York: Wiley-Interscience.

Fellows PJ (ed.) (1996) *Food Processing Technology, Principles and Practice, Series in Food Science and Technology*. Cambridge: Woodhead.

Kress-Rogers E (ed.) *Handbook of Biosensors & Electronic Noses*. Boca Raton, FL: CRC Press.

Mathewson PR and Finley JW (eds) (1992) *Biosensors Design and Application, ACS Symposium Series 511*. Washington, DC: American Chemical Society.

Mulchandani A and Rogers KM (eds) (1998) *Methods in Biotechnology; Enzyme and Microbial Biosensors. Techniques and Protocols*. Totowa, NJ: Humana Press.

Rogers KR and Mulchandani A (eds) (1998) *Methods in Biotechnology 7, Affinity Biosensors, Techniques and Protocols*. Totowa, NJ: Humana Press.

Scott AO (ed.) (1998) *Biosensors for Food Analysis*. Cambridge: The Royal Society of Chemistry.

Turner APF (ed.) *Advances in Biosensors*. volumes 1; 2; Suppl 1; 3; 4 (1991; 1992; 1993; 1995; 1998). London/Amsterdam: JAI Press/Elsevier.

Turner APF (2000) *Biosensors McGraw Hill Yearbook of Science and Technology 2000*. New York: McGraw Hill.

Turner APF, Karube I and Wilson GS (eds) (1989) *Biosensors: Fundamentals and Applications*. Oxford: Oxford University Press.

BIOTECHNOLOGY IN FOOD PRODUCTION

S Pérez-Magariño and M L González-Sanjosé,
University of Burgos, Burgos, Spain

Introduction

Biotechnology's origins lie in the ancient crafts of brewing, baking, and the production of fermented foods such as yogurt and cheese, although it was not until 1859 that microorganisms were identified as the cause of both desirable and undesirable changes in food. During the 1940s, microorganisms, as submerged cultures in large vessels, were developed. Nowadays, similar fermentation technology is applied to obtain a wide range of bioproducts.

In the 1950s, the primary structure of proteins and DNA was first determined. These discoveries led to the development of molecular biology in the 1960s and genetic engineering techniques in the 1970s. Techniques developed using microorganisms have been applied to plants and animals, and these transgenic plants and animals have been produced to make many useful products.

For centuries, humans have been selecting, sowing, and harvesting seeds to produce better food more efficiently. In recent decades, global food demand has increased enormously, and biotechnology has offered the technology required to produce more nutritious and better-tasting foods, higher crop yields, and plants protected against disease and insects.

Definition of Biotechnology

The term 'biotechnology' has come to mean different things to different people. In general, biotechnology is the integration of natural sciences and engineering in order to provide new products and services by the application of organisms, cells, parts thereof, and molecular analogs.

In a broad sense, biotechnology covers many of the tools and techniques that are commonplace in agriculture and food production. However, in a narrow sense, biotechnology considers only the new DNA techniques, molecular biology, and reproductive technological applications. Thus, 'biotechnology' for many people refers specifically to the genetic engineering techniques that have been developed in the last 30 years.

Food Biotechnology

Biotechnology provides powerful tools for the sustainable development of agriculture as well as the food industry, improving the production of foods and making them more available and nutritious. Food biotechnology is a rapidly expanding field of science with many different applications, which is clear from the large number of scientific papers published in the last 10 years. The Food Science and Technology Abstracts (FSTA) database reveals at least 4150 articles on Food Biotechnology published between 1990 and 2000, which represents almost 4% of the number of total published papers on food.

According to FSTA information and the current situation of the food industry, some of the main areas of application of food biotechnology are: the production of new kinds of foods and drinks, both by modern developments of conventional techniques and by genetically modifying the products themselves, or producing them using genetically modified (GM) organisms or their products.

The majority of articles deal with fermentation processes, the use of enzymes and microorganisms (Figure 1). These fields have been the most important and traditional applications of biotechnology in food production. However, in the last few years, investigations into genetic engineering, molecular biology, and cell cultures have been increasing significantly.

Traditional Food Biotechnology

Traditional biotechnology refers to the conventional techniques that have been used for many centuries to produce beer, wine, cheese, bread, and other foods, all of them foods obtained by fermentation processes.

The fermentation of food is the oldest biotechnological process, and is the oldest form of food preservation. In brief, fermentation is anaerobic conversion of certain carbohydrates (such as reducing sugars, mainly glucose) to other products such as:

- alcohol, glycerol and carbon dioxide, the principal products of the fermentation of yeast sugars (alcoholic fermentation);
- butyl alcohol, acetone, lactic acid, and acetic acid, the principal products of bacterial action (such as acetic, lactic, and malolactic fermentation).

Of the many possible types of fermentation processes, lactic acid fermentation and alcohol fermentation are the two most common. These processes are carried out by specific microorganisms (several yeasts and

Figure 1 Percentage of published papers on different fields of biotechnology in food production.

bacteria), and they are regulated by the action of specific enzymes.

Traditional biotechnology is most closely related to food production both in the selective breeding of food plants and animals and in food processing using microbial enzymes. Traditional selection techniques have been employed to develop a great variety of plants, animals, and microorganisms for the production of a wide range of food products and ingredients for processed foods.

In addition, enzymes have been applied to practical uses for thousands of years, in the form of crude animal or plant preparations or as a consequence of permitted microbial development. Improvements in modern techniques, which allow purification and purification of specific fungal and microbial enzymes and their production at an industrial level, has made it possible to use enzymes in order to facilitate and improve some food manufacturing processes.

Current technology for dealing with complex organisms advances modern biology and permits us to understand many traditional processes in which biological agents are involved, allowing the application of new fermentation and enzyme technological processes to food production. This new situation has opened the doors to an exciting dimension of food biotechnology, which has been called 'modern food biotechnology.'

Modern Food Biotechnology

Modern biotechnology embraces all methods of genetic modification by recombinant DNA and cell-fusion techniques together with the modern developments of 'traditional' food biotechnological processes. Genetic modification techniques (GMT) are now being used in the production of new foods and drinks. Genetic modification involves the insertion of one or a small number of scientifically well-characterized genes into the food plant, animal, or microorganism. These techniques will play an increasingly useful and economic role, and they are strongly related to fields such as genetic engineering, molecular biology, protein engineering, biochemical engineering, and processes involving monoclonal antibodies, which are beginning to have a considerable impact on food processing.

Modern biotechnology will pursue not only the production of new foods but also the preservation of raw materials, increased shelf-life of final products, and the planned alteration of their nutritional and functional properties. It also affects the development of processing aids and direct additives that can improve the overall utilization of materials. GMT are being used to achieve many of the same aims as traditional breeding and selection methods but have at least two main advantages. First, they provide the means of controlling the introduction of genes with much greater prediction and precision than can be achieved by traditional methods. Second, they make it possible to introduce copies of genetic material into unrelated species hitherto impossible to achieve by traditional techniques.

Some applications of genetic engineering techniques are:

- increasing the resistance of plants to diseases and pests, herbicides or pesticides;
- developing plants to withstand more extreme environmental conditions (drought, frost, and soil salinity, amongst others);
- developing foods with specific properties and improved quality (sensorial and nutritional characteristics);
- adapting microorganisms for more efficient food production and to produce natural food ingredients (amino acids, organic acids, volatile fatty acids, vitamins, etc.);
- obtaining enzymes, antibodies and microorganisms to monitor food production and processing

systems for quality control. Microbial probes and biosensors are being used experimentally as indicators of bacterial and other types of contamination; and
- improving animal farming by genetic modification of crop plants for animal feed, of microorganisms or enzymes to improve the nutritional value of animal food stuffs and in the animal health sector for pharmaceuticals.

Benefits and Risks of Food Biotechnology

For centuries, farmers have relied on the newest technology to produce and enhance foods that possess specific beneficial traits. The use of biotechnology benefits not only the grower or producer but also the consumer.

The current benefits of biotechnology include disease resistance, reduced pesticide use, herbicide tolerance, more rapid growth of crops, producing higher yields. Hence, biotechnology benefits the environment, including water and soil conservation, and is safer for workers and the ecosystem. Also, it can enhance the nutritional characteristics of foods and improve food taste and quality.

Biotechnology has been used in a number of crops for several years. The volume of biotechnological crops in developmental stages continues to grow. In the early 1990s, many people were skeptical that transgenic crops could provide improved products. During the last five years, the planting of transgenic crops has increased, thus demonstrating that the early promises of transgenic crops are meeting the expectations of farmers of both large and small farms planting these crops in both industrial and developing countries. However, two-thirds of total transgenic crops are in the developing countries, where yields are lower, and the need for improved production of food is the greatest.

More genetically enhanced products are expected to be on the market in the years to come, many of which are similar in nature to products already on the market. Benefits that can be expected in the near future include:

- reduced levels of natural toxins in plants;
- simpler and faster methods to locate pathogens, toxins, and contaminants; and
- extending the time before spoilage.

The world population is expected to double by the year 2050. Few other new technologies will be able to approach biotechnology's potential to help avoid starvation in the next century. Another economic and environmental benefit is in the area of fertilizer use. Much of the fertilizer used is wasted and can end up in water sources, damaging the environment. Some plants, such as corn, might also be modified to draw nitrogen from the soil, thereby reducing the need for fertilizer. Benefits that can be expected further down the road include:

- producing safer foods through reduction of allergenic proteins;
- drought and flood tolerance;
- salt and metal tolerance; and
- heat and cold tolerance.

In general, according to most scientific papers published, the use of biotechnology poses no specific risk, and products should not be discriminated against on the grounds of their method of production. The scientific consensus is that the risks associated with food biotechnology are fundamentally the same as for other foods. Current science shows that foods made using biotechnology are safe to consume and safe for the environment. Although there is no such thing as 'zero risk' for any food, consumers can be confident that foods produced using biotechnology meet the government's most stringent food-safety standards. Years of research and no evidence of promoting health harm indicate that the benefits of agricultural biotechnology far outweigh any risks.

However, there are people, amongst them ecologists and activist groups, who do not agree with this. Some of their arguments against biotechnology are that the use of GM products could reduce genetic diversity indirectly if, for example, farmers adopt genetically uniform varieties of plants and other organisms. The inclusion of novel genes for herbicide resistance in plants may increase the occurrence of weeds with resistance to certain agrochemicals. Others are concerned about the possible appearance of allergens or toxins and possible causes of different human diseases.

Thus, despite the benefits to be gained from modern biotechnology, the products obtained need to have appropriate legislation and safeguards to protect human health and the environment.

Current Concerns, Regulations, and Labeling

According to the concerns expressed about the possible existence of health hazards associated with the consumption of food containing modified genetic material, it is clear that rational policies are needed, regarding the regulatory status of bioengineering products.

Studies about public perceptions, consumer attitudes, and ethical implications of biotechnology in food and drink production have shown that certain

applications are more acceptable than others. For example, according to the European Federation of Biotechnology, the use of biotechnology to change plants was considered more acceptable than using it to change animals. In addition, around 85% of interviewees said that food obtained with genetic techniques should be clearly labeled. It will be necessary, therefore, to develop programs to educate the public about the benefits and problems of biotechnology in the production and processing of food. Governments and scientific institutions must be encouraged to share knowledge on transgenic foods with society, such that people are well informed about the impact of biotechnology on the environment, food safety, sustainability, and global food safety. In addition, as with any technology, there are limits that will and must be subjected to law. Countries must be helped to develop appropriate legislation and to set up proper regulatory bodies for all aspects of biosafety. Appropriate regulation is a prerequisite from the point of view of both the general public and the food industry.

In the USA, there are three organizations responsible for controlling GM organisms and products. The Environmental Protection Agency (EPA) and the United States Department of Agriculture (USDA) oversee the effects of GM organisms on the environment and on plant production. The Food and Drug Administration (FDA) is responsible for overseeing the safety of food, which is determined by the characteristics of the consumed or processed product rather than the method by which it was produced. The FDA policy statement also addresses the labeling of foods obtained from new plant varieties. It states that labeling is only required when a food obtained from a GM variety differs from that obtained from a conventional counterpart.

Nowadays, in Europe, there is no equivalent overall agency for food regulation, although a European Institute will be in operation in the near future. Until this time, the European Commission has therefore proposed harmonizing regulation between member states. The Council of the European Community (EC) adopted the Directive 90/220/EC on the deliberate release into the environment of GM organisms. This directive had two main objectives: to protect human health and the environment and to provide a harmonized regulatory framework of GM organisms with the EC.

In the EC, the safety of GM foods is controlled by Regulation (EC) No. 258/97 concerning novel foods and novel food ingredients, and the labeling of GM foods is regulated by two regulations, with important differences with American Regulations. First, Commission Regulation (EC) No. 49/2000 establishes that the foodstuffs shall not be subject to the additional specific labeling requirements when the material derived from the GM organisms is present in a proportion no higher than 1% of the food ingredients individually considered or food comprising a single ingredient, or this presence is adventitious. Second, Commission Regulation (EC) No. 50/2000 establishes the labeling of foodstuffs and food ingredients containing additives and flavorings that have been genetically modified or have been produced from GM organisms. In addition, novel foods and GM foods also have to fulfill the general labeling requirements set out in the Directive 2000/13/EC, which deals with an approximation of the laws of the Member States relating to the labeling, presentation, and advertising of foodstuffs.

With so many foods traded internationally, there is a clear need to reach a harmonized approach to labeling GM foods if potential trade barriers are to be eliminated. Therefore, it will be necessary to develop international safety guidelines.

In 1990, Food and Agriculture Organization (FAO) and the World Health Organization (WHO) took the first steps towards international harmonization of the food safety assessment of GM foods. It was recommended that the safety assessment of foods produced by biotechnology should take into account the molecular, biological, and chemical characteristics of these foods. The 1996 FAO/WHO report defined substantial equivalence as 'established by a demonstration that if the characteristics assessed for the GM organisms, or the specific food product derived therefrom, are equivalent to the same characteristics of the conventional food, then it can be assumed that the new food is as safe as the conventional equivalent.' In 1999, the FAO/WHO of Codex Alimentarius established the Codex *Ad Hoc* Intergovernmental Task Force on Foods derived from Biotechnology to elaborate norms and guidelines for the foods derived from biotechnology. The Task Force held its first session in March 2000, and approved the elaboration of major texts with a set of broad general principles for risk analysis and specific guidance on the risk assessment of foods derived from biotechnology. It will also have to prepare a list of available analytical methods including those for the detection or identification of these kinds of foods.

The Task Force agreed that the labeling was covered by the Codex Committee on Food Labeling, and the environmental risk was addressed by other bodies such as the Cartagena Biosafety Protocol under the Convention on Biological Diversity, the International Plant Protection Convention, and the Commission on Genetic Resources for Food and Agriculture.

Biotechnology in the Food Industry

As we said earlier, biotechnology applied to food production has a long history, beginning in cottage industries such as bread-making, cheese-making, and brewing. Many biotechnology advances have been made in the food industry over the last century. A major advance has been in food preservation, which is basically the prevention of undesirable microbial and enzyme activity in foodstuffs. This allows the storage of food over much longer periods and permits transport over great distances. The development of genetic techniques also represents a great advance, providing opportunities for much innovation.

According to the data obtained from the FSTA data base (Figure 2), food biotechnology processes are applied more to manufacturing foods of vegetable origin (38%) than to foods of animal origin (26%). In addition, the fields of study covered by the two largest groups of papers were vegetables and fruits first (19%) with dairy products in second place (15%). Papers on beverages, cereal products, meat products and fish products have been published more or less in the same quantity. During the last few years, the most frequent applications of biotechnology in food industries have been as follows.

Biotechnology in the Vegetable and Fruit Industries

Enzyme applications are very common in this industry. The presence of acid phenols and tannins in fruits has been studied, owing to their effect on bacterial growth during lactic fermentation.

Pectinolytic enzymes and cellulolytic enzymes are the most commonly used. These enzymes have applications in plant tissue maceration and in fruit juice processing, improving the clarification of juices and manufacturing processes of fruit such as peeling. Owing to the importance of these enzymes, many works have studied their activities during the ripening of fruits. These enzymes have even been extracted from fruits and purified later to be characterized and applied in different food industries.

Vegetables and fruits are used as substrates in some reactions, owing to their enzymatic systems, to produce other chemical compounds, such as volatile flavor compounds, lipids, vitamins, enzymes, pigments, etc.

The genetic variation of some vegetable and fruit crops, and the effect of genotype and environment on yield and quality, have been investigated, using random amplified polymorphic DNA. This DNA-based technology and others have resulted in a substantial increase in the number of genes identified and characterized, which will make it possible to produce transgenic crops of several vegetables and fruits.

Biotechnology in the Beverage Industries

In general, fermented beverages are obtained by the presence of determinate microorganisms, so the majority of published papers deal with this topic.

Many applications focus on the selection of yeasts capable of performing alcoholic fermentation under extreme conditions (high sugar concentration, high amounts of SO_2 or ethanol). GM amylolytic yeasts, which can hydrolyze starch, have been used to improve fermentation processes.

Some studies on the evaluation of the use of non-*Saccharomyces* yeasts, separating those that could positively affect the taste and flavor of alcoholic beverages from others that produce large amounts of negative byproducts, have also been developed. Yeasts selected have been applied to obtain fruit drinks with a slightly acid and less sugary taste. The use of immobilized cells in fermented beverages for continuous production has also been studied.

Another field is the use of different methods such as the combination of pressure, heat, and bacteriocins, and a flow-through photoreactor, to render inactive foodborne pathogens in some juices and some microorganisms in water, respectively.

Other studies have investigated the molecular and catalytic properties of different types of enzymes (proteinases, glucosidases), which have particular

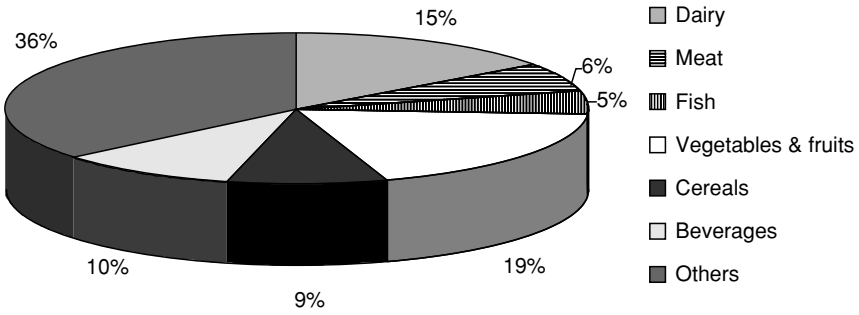

Figure 2 Percentage of food biotechnology published papers in different sorts of food industries.

importance in the production of fermented beverages. Commercial purified pectinolytic enzymes have been applied to improve clarification processes and aroma and color compound extraction, in beverage industries. The beverage industry has used genetic engineering techniques mainly for the selection of yeasts.

Biotechnology in Cereal Industries

The most important field has been fermentation. Isolation and characterization of both desirable and undesirable microorganisms (pathogens) have been carried out. The role of cereal as a substrate to obtain antimutagenic or anticarcinogenic fungic compounds has been evaluated.

The cereal breeding industry is also investigating the application of genetic techniques to improve major grain parameters (e.g., resistance to herbicides, viruses, fungal disease, temperature variations).

Biotechnology in Dairy Industries

A great deal of research has been carried out on selection, industrial production conditions, immobilizing processes, and applications of different dairy starters. The most studied microorganisms group has been the lactic acid bacteria (LAB), which have been applied to improve and increase industrial productions of traditional food, without any loss of popular acceptance; to increase the nutritional and health value of dairy foods; to increase selected flavor generation; and to improve the texture of dairy products.

Potential health benefits of bifidobacteria have been studied, and their application in yogurt and cheese production has been described. Some LAB have been genetically modified to utilize the lactose, and have been applied to obtain new yogurts with a sweet alanine flavor, longer shelf-life, or improved clinical probiotic value.

Other important studies have been concerned with the isolation of microbial enzymes, developments in GM enzymes, and their application in dairy industries. Clotting enzymes (chymosin), proteases, and lipases are the most studied and applied. Immobilized or free lipases have been used in many different aspects, such as for the production of food additives (e.g., antioxidants by esterification of phenols) and desirable flavors (e.g., in cheese, buttermilk), for synthesis of flavor thioesters, or as emulsifiers or food spoilage agents. Immobilized esterases have been employed in butter oil to obtain dairy flavor products. Thermally stable microbial β-glycosidase has been used to avoid problems of human lactose intolerance.

Some microorganisms are selected to improve the safety of dairy products, which are able to biosynthesize biopreservative compounds (e.g., reutenin). Different biotechnological practices (selected microorganisms or their metabolites) can be applied to control and predict bacteriophage contamination in dairy products, and bacteriocins isolated from different cultures have been applied to render inactive milk pathogens.

Many papers on probiotic dairy food have been published in relation to mechanisms, underlying their effects on human health, and some microorganisms have been applied to reduce the level of cholesterol and increase the availability of provitamin D_3.

Recent works have studied the use of genetic engineering to modify the composition of milk indirectly by using transgenic animals. These studies have been carried out to obtain low-lactose milk or to improve the quality of milk (increasing the fat level) to make special cheeses such as mozzarella.

Different papers have shown biotechnological applications on dairy effluents such as cheese whey. Microbial fermentation of whey for the production of biosurfactant sophorolipids or some enzymes with important industrial and environmental applications has been described.

Biotechnology in the Meat-product Industries

Fermentation processes in order to increase microbial activities and the production of bacteriocins have been extensively studied with satisfactory results, and new starter cultures (LAB) have been developed.

In addition, different kinds of proteinases alone or accompanied by lipases and esterases have been studied and applied usefully for the tenderization and ripening of meats and their products. Proteinases have been also used to hydrolyze peptides isolated from different animal meats to obtain new flavors.

In meat production, amylases, cellulases, and pectinases have been mixed and applied to obtain new animal feed, which increases the body weight and the percentage of protein and ash of the meat.

The main genetic technique used is the polymerase chain reaction, which uses a DNA-based identification system for testing the species origin of meat samples and is suitable for authenticating processed meat products.

Biotechnology in the Fish-product Industries

The development of new fermented products (fish and sauces) with special attention on desirable sensory properties has been one of the most important fields of study in these industries.

The second field has been the production of ω-3 polyunsaturated fatty acids, using immobilized lipase, or looking for marine organisms (bacteria, microalga,

fungi, and zooplankton) that can biosynthesize these acids.

Other applications have been the use of different enzymes for the tenderization and ripening of fish and to obtain new flavors. Fish enzymes and proteins have been isolated and applied in the food industry. LAB and their bacteriocins isolated from fish intestines have been genetically characterized, and their potential use in food preservation evaluated.

See also: **European Union**: European Food Law Harmonization; **Food and Drug Administration**; **Food and Agriculture Organization of the United Nations**; **Genetically Modified Foods**; **Legislation**: International Standards; Codex; **World Health Organization**

Further Reading

Altman A (1998) *Agricultural Biotechnology*. New York: Marcel Dekker.

European Federation of Biotechnology (EFB) Task Group on Public Perceptions of Biotechnology (1997). *Biotechnology in Foods and Drinks Inform*. London, UK: EFB

IFT (2000) Genetically modified organisms. *Food Technology* 54: 42–45.

Kaufman PB, Cseke LJ, Warber S, Duke JA and Brielmann HL (1999) *Natural Products from Plants*. Boca Raton, FL: CRC Press.

Lappe M and Bailey B (1999) *Against the Grain. The Genetic Transformation of Global Agriculture*. London: Earthscan Publications.

Lee BH (1996) *Fundamentals of Food Biotechnology*. New York: VCH Publishers.

Menrad K, Agrafiotis D, Enzing CM, Lemkow L and Terragni F (1999) *Future Impacts of Biotechnology on Agriculture, Food Production and Food Processing*. London: Springer.

Nelson GC (2001) *Genetically Modified Organisms in Agriculture, Economics and Politics*. London: Academic Press.

Pandey A, Benjamin S, Soccol CR *et al*. (1999) The realm of microbial lipases in biotechnology. *Biotechnology and Applied Biochemistry* 29: 119–131.

Ratledge C and Kristiansen B (2001) *Basic Biotechnology. Part I and II*. New York: Cambridge University Press.

Roller S and Harlander S (1998) *Genetic Modification in the Food Industry: A Strategy for Food Quality Improvement*. London: Blackie Academic & Professional.

Ryan-Harshman M (1998) Food biotechnology: food industry, nutrition and public health. *Proceedings of the Nutrition Society* 56: 845–854.

Shahidi F, Jones Y and Kitts DD (1997) *Seafood Safety, Processing and Biotechnology*. Basle: Technomic Publishing.

Thompson PB (1997) *Food Biotechnology in Ethical Perspective*. London: Blackie Academic & Professional.

Tzotzos GT and Skryabin KG (2000) *Biotechnology in the Developing World and Countries in Economic Transition*. Wallingford, UK: CABI Publishing.

BIOTIN

Contents

Properties and Determination

R Bitsch, Friedrich-Schiller-University, Jena, Germany

Biotin

The molecular structure of biotin is characterized by three chiralic C atoms leading theoretically to eight stereoisomers. Only the D(+)-isomer is biologically active. The most commonly used microbiological assays (bioassays) for quantitation are more recently replaced by protein binding assays, with avidin showing higher precision and enhanced detection limits. Besides these methods, biotin can also be analyzed mainly in pharmaceuticals using photometric and fluorometric procedures or chromatographic techniques (TLC and HPLC).

Properties

The molecular structure of biotin (CAS No. 58–85–5) is a bicyclic ring system comprising an imidazolidone ring fused with a tetrahydrothiophene ring that is substituted with a valeric acid side chain. The stereochemistry reveals that both ring systems are fused in a

boat-like configuration at an angle of 62° (Figure 1). The chemical nomenclature of this compound is like a *cis*-hexahydro-2-oxo-1H-thieno[3,4-D] imidazole-4-pentanoic acid as was elucidated by Kögl 1937 and Du Vigneaud in 1941. The molecule contains three chiralic centers (*), and so eight stereoisomers are possible. Only the D-(+) isomer with a *cis*-configuration at the chiralic centers possesses vitamin activity and is found in natural products. The empiric formula is $C_{10}H_{16}N_2O_3S$ (molecular weight 244.3 Da). The molecular structure was confirmed in 1943 by Harris *et al.*, who were the first to achieve total biotin synthesis. In the meantime, a stereospecific synthesis for commercial production was developed by Goldberg and Sternbach in 1949.

Physicochemical Characteristics

D-Biotin as a free acid with a pK_a value of 4.51 crystallizes in colorless needles with a melting point of 230–232 °C (decomposition). The substance is slightly soluble in water (0.82 mM = 0.2 g l^{-1} at 25 °C) and ethanol 95% (3.27 mM = 0.8 g l^{-1}), readily soluble in dilute alkali, and almost insoluble in most organic solvents. The specific optical rotation $[\alpha]_D^{22}$ in 0.1 N NaOH is + 92 °C ($c = 0.3$) (Table 1).

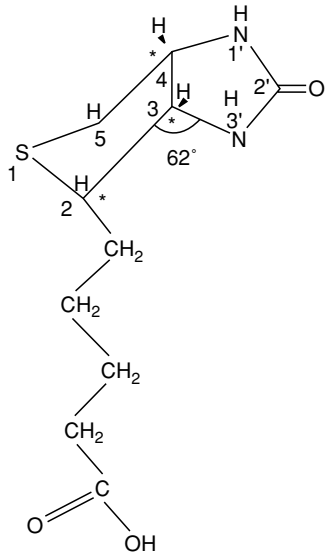

Figure 1 D(+)-Biotin. *, chiralic centers.

Crystalline D-biotin is fairly stable on exposure to air, daylight, and heat. Gradual destruction is observed on exposure to ultraviolet radiation. In weakly acid or alkaline solutions between pH 4.0 and 9.0, it is relatively stable, too. The molecule is also resistant to autoclaving in 2 N or 6 N sulfuric acid for 2 h (the commonly used extraction procedure).

Biotin is, however, susceptible to oxidation at the sulfur atom of the thiophene ring with the sequential formation of biotin sulfoxide and biotin sulfone. This oxidation, which is negligible in concentrated solutions, provides a primary step for vitamin activity losses in processed foods producing a mixture of D- and L-sulfoxides. While D-biotin sulfoxide exhibits a similar growth promoting activity for some microorganisms to that of biotin (e.g. *Lactobacillus* and *Saccharomyces* sp.), subsequent oxidation to biotin sulfone leads to complete loss of the biological activity. Biotin sulfoxide also seems to be a product of the biotin metabolism and can be detected as a minor excretion component in urine. In addition, microbial growth can drastically lower the concentration of nonsterile dilute solutions of this vitamin.

Besides the sulfoxide and sulfone, further metabolites and analogs have been identified as intermediate products of the biosynthesis (e.g., dethiobiotin) or of the metabolism (e.g., bis- and tetranorbiotin) in bacteria, yeasts, and fungi or as byproducts of the chemical synthesis (e.g., oxybiotin) (Figure 2). The growth response of microorganisms to these analogs is variable and thus may lead to conflicting results in terms of vitamin activity in biological samples when measured by means of microbial assays (see Table 2). Bisnorbiotin, tetranorbiotin, and biotin sulfone, as well as other derivatives, are potent antagonists of biotin and do not support microbial growth. Apart from the sulfoxide and sulfone, these analogs are of little practical importance in foods. Nevertheless, more recent results indicate that in human urine samples, besides unchanged biotin, about 11% D-biotin-sulfoxide, 29% bisnorbiotin, and 4% biotin sulfone plus bisnorbiotinmethyl-ketone and tetranorbiotin-L-sulfoxide could be identified, but no

Table 1 Physicochemical properties of biotin

Brutto formula	*Molecular weight*	*Melting point*	*Specific optical rotation*
$C_{10}H_{16}N_2O_3S$	244.3 Da	230–232 °C (decomposition)	$[\alpha]_D^{22} = +92°$ (0.1 N NaOH, $c = 3.0$)
Solubility			
Water	*Ethanol* (95%)	*Lipophilic solvent*	*Dilute alkali*
0.82 mM = 0.2 g l^{-1} (25 °C)	3.27 mM = 0.8 g l^{-1}	Insoluble	Readily soluble

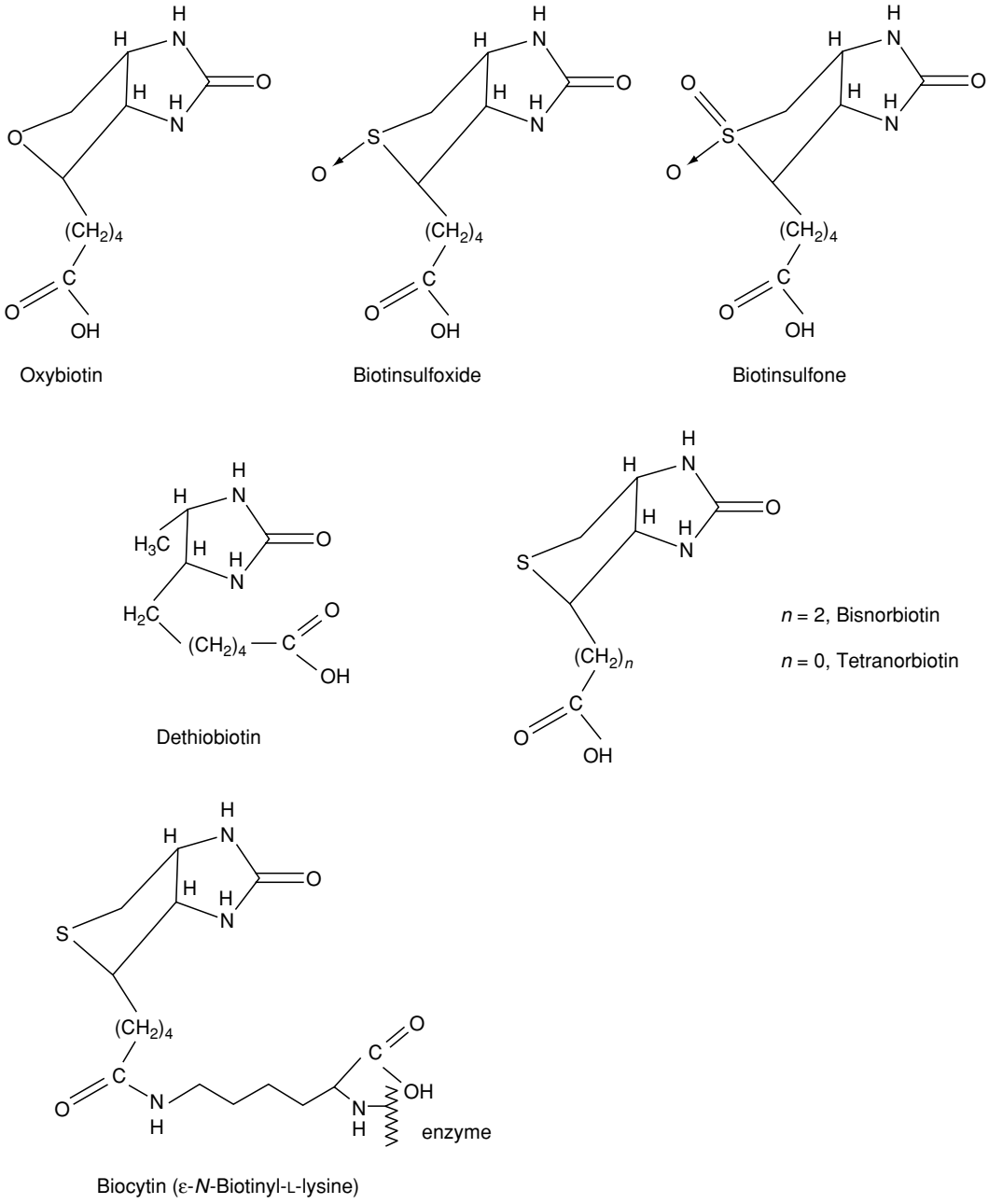

Figure 2 Biotin derivatives and analogs.

degradation products of the cyclic urea ring were found to be excreted.

In natural products, biotin exists either in free form or protein-bound. Generally, the free form predominates in plant food, whereas a greater part in animal food is protein-bound. Therefore, less drastic hydrolysis with 2 N sulfuric acid is required to liberate biotin from plant material than from products of animal origin, which need 6N sulfuric acid. The biotin content in selected foods is shown in Table 3.

If protein-bound, the vitamin is linked through an amide bond between the carboxyl group of the valeric acid side chain and an ε-amino group of terminating lysine. This ε-*N*-biotinyl-L-lysine (biocytin) is naturally occurring, too, and is partly formed by proteolysis of biotin enzymes. Occasionally, biocytin is indicated as the real prosthetic group of biotin-dependent enzymes.

The most prominent antagonist of biotin is avidin. This is a glycoprotein from egg white with a MW of 68 000, which is formed within the oviduct of avian

Table 2 Growth response of biotin derivatives to some microorganisms

Microorganism	*Biotin-D-sulfoxide*	*Oxybiotin*	*D-Dethiobiotin*	*Biocytin*
Lactobacillus plantarum (ATCC 8014)	++	+	+	−
Lactobacillus casei (ATCC 7469)	−	+	−	++
Ochromonas danica	−	−	−	++
Saccharomyces cerevisiae (ATCC 7754)	++	(+)	+ +	++
Neurospora crassa	−	−	+ +	++
Kloeckeria brevis (ATCC 9774)	++		−	++

++, identical activity to equimolar biotin; +, about half of the activity of equimolar biotin; −, no growth response.

Table 3 Biotin content in selected foods[a]

Food	*Biotin content (μg per 100 g)*	*Nutrient density (μg biotin per 1000 kcal or 4.2 MJ)*
Fruits		
Apples, apricots, bananas, grapefruits, pears, strawberries	0.4–5.5	8–125
Vegetables		
Asparagus, cauliflower, peas, potatoes, spinach, tomatoes	1–4	51–314
Eggs (whole)	25	162
Nuts		
Almonds, peanuts, walnuts	10–34	18–61
Milk and cheese		
Whole milk, brie, gouda	1.5–6	4–54
Meat		
Pork and beef, lean	4–5	27–28
Liver		
Pork and beef	23–92	194–536
Fish		
Cod, herring, redfish, trout	2–10	22–80
Cereals		
Wheat flour, oats, rice polished, rice (wholegrain), rye (wholemeal)	1.5–17	4.5–34

[a]Modified, according to the databases of Bundeslebensmittelschlüssel II.2, 1996, of the Bundesinstitut für gesundheitlichen Verbraucherschutz und Veterinärmedizin, Berlin, with permission.

species and reptiles. It has a tetrameric structure with four identical subunits (MW 17 000), each containing a binding site for biotin, so 4 moles of biotin are bound by 1 mole of avidin.

The interaction of avidin with biotin is the strongest noncovalent binding ever found in living organisms with a dissociation constant of the binding sites, $K_d = 10^{-15}$ M. Because avidin is a basic protein with an isoelectric point of pH 10.5, by oral intake, not only the dietary but also the enterally synthesized and enzymatically bound biotin can be irreversibly bound. The avidin–biotin complex is not cleaved within the gastrointestinal tract.

Raw egg white contains 2 mg of avidin per 100 g, so about 15 μg of biotin are fixed by one egg (c.50 g) and made unavailable for man and animal. Not only free or enzymatically bound biotin but the structural isomers with an intact ureido ring in the *cis* configuration compete for complexing, too, as is shown in Table 4. The very stable avidin–biotin complex is only split by autoclaving with strong acid. Avidin, however, is only in raw egg white active as an inhibitor of the biotin availability and is denaturated by heat, thereafter losing its binding capacity.

The exact physiological role of avidin is not well understood, but it is generally thought to be effective as a bacteriostat. With respect to the stoichiometric binding, avidin has served as a useful diagnostic tool in the identification and study of biotin-dependent enzymes.

Besides avidin, streptavidin (MW 60 000) has been isolated from *Streptomyces avidinii*, which is also composed of four subunits with a biotin-binding affinity similar to that of avidin. On the basis of the stoichiometric binding of biotin by avidin and streptavidin in a stable complex, current analytical methods to quantitate biotin and its metabolites in biological materials have been developed.

Determination

Microbiological Assays

As with other B vitamins, biotin has been exclusively assayed for longer times by microbiological methods. Biotin is a growth factor for a variety of bacteria, fungi, yeasts, and flagellatae. For assay purposes, in foods and other biological materials such as blood, plasma/serum, urine, and various tissues (liver, brain), *Lactobacillus plantarum*, *L. casei*, *Saccharomyces cerevisiae*, *Neurospora crassa*, and *Ochromonas danica* have been used. Besides the variable response to biotin analogs, as shown in Table 2, the lack of specifity is also due to certain amino acids, fatty acids and their metabolites, which are able to stimulate the growth and replace biotin, thus affecting the validity of the assay. Results achieved with different microbial assays are often therefore hardly comparable, particularly if the relationship of biotin to these interfering factors is poor.

Table 4 Quantitation (%) of biotin analogs and derivatives by various binding proteins and microbial assay (comparing equimolar amounts), according to Baur, Helbich-Endermann, and Salz (unpublished results)

Derivative	*Avidin*	*Streptavidin*	*Neutravidin*	*Lactobacillus plantarum*
Biotin	99	88	100	98
Biocytin	103	98	83	0.1
Biotin-D-sulfoxide	97	87	35	83
Biotin-L-sulfoxide	96	36	24.5	14
Biotinsulfone	96	54	28	0.2
Bisnorbiotin	49	0.4	5.5	0.2
Dethiobiotin	10	–	–	0

Minor growth activity for *L. plantarum* exhibits oleic acid, elaidinic acid, linoleic and linolenic acid, as well as a few other lipids, which have to be removed either by filtration or by ether extraction. Aspartate and several other amino acid metabolites, pimelic acid, and oleic acid interfere with the *Saccharomyces* assay. Less frequently used microorganisms are *N. crassa* and *O. danica*, although *O. danica* is rather specific and responds, besides biotin, only to biocytin and protein-bound vitamin. The *O. danica* assay was applied for biotin assessments in blood, serum, urine, liver, brain, and muscle. The validity of the method established by recovery experiments was demonstrated to 96–103% biotin recovery.

In order to extract the vitamin, in most assays, the biological material is subjected to acid (sulfuric acid) or enzymatic hydrolysis. Protein-bound biotin is also cleaved by papain, a proteolytic enzyme hydrolyzing specifically lysine-bound substances.

Principle of the *Lactobacillus plantarum* (ATCC 8014, formerly *L. arabinosus*) assay To liberate protein-bound biotin, foods, feeds and other biological samples are extracted with sulfuric acid for 2 h at 121 °C. Plant products as well as whole blood, plasma, and urine samples are autoclaved with 2 N sulfuric acid, and animal samples otherwise with 6 N sulfuric acid. The ratio of sample to acid should be 1:20, whole blood, plasma or urine at a ratio of 3 ml of sample to 8 ml of acid. For pharmaceuticals and premixes, dilution with 0.1 N sodium hydroxide at a concentration of about 0.2 ng ml^{-1} and, if needed, centrifugation or filtration to clarify is the preferred extraction mode. Stronger acid concentrations for sample extraction should be avoided, because a high sodium sulfate concentration in the assay is inhibitory to the test organism. Furthermore, using strong sulfuric acid, followed by autoclaving, biotin may be partly oxidized to inactive biotin sulfone.

Hydrochloric acid should not be used in place of sulfuric acid, as the hydrochloric acid may inactivate biotin probably because of the presence of chlorine. High-fat samples must be extracted, prior to acid cleavage, with hexane or ethyl ether to avoid fatty acid stimulation.

After cooling, the samples are adjusted to pH 6.75 with concentrated NaOH, and biotin in the extract is determined by means of its growth response to an inoculum suspension prepared from a stock culture of *L. plantarum* and diluted with culture medium. After incubation at 37 °C, the absorbance (nephelometric) is measured at 660 nm, and the biotin content is calculated by comparing with a standard curve using D(+)-biotin concentrations of up to 0.80 ng per 5 ml. The detection limit of this microbiological assay is 122 pg biotin ml^{-1} (0.5 pmol ml^{-1}). The coefficient of variation (CV) for whole blood analysis amounts to about 19%.

A modification of the standard biotin assay with *L. plantarum* has been developed by a radiometric-microbiological assay (RMA) based on the measurement of $^{14}CO_2$ from the metabolism of L-1-[^{14}C]-methionine by *Kloeckeria brevis* (ATCC 9774). The amount of $^{14}CO_2$ is measured with the aid of a gas-flow ionization chamber system and is proportional to the present biotin amount. The assay sensitivity has been found to be 10 pg of biotin (41 fmol)/vial (= 5.2 ml). This test has been employed for plasma, as well as food, analysis. Biocytin and biotin-D-sulfoxide have the same biological activity as biotin in equimolar concentrations, but dethiobiotin and other analogs as well as fatty acids and aspartate did not interfere with the biotin measurement (see Table 2). The RMA allows some simplification of the sample preparation and assay procedure (no filtration of the sample extract and no turbidimetric measurement). Its application for biotin analyses, however, is limited by the need for special laboratory devices.

The application and detection limits of the most commonly used microbial assays are summarized in Table 5.

Isotope Dilution Assays (Radioligand Binding Assays)

The strong and specific affinity of the glycoprotein avidin to biotin was taken very early on as the basis

Table 5 Microbiological assays

Microorganism	*Principle of assay*	*Biotin release*	*Application*	*Detection limit*
Lactobacillus plantarum ATCC 8014	Growth-promoting acticity	2–6 N H_2SO_4 (121 °C, 2 h)	Biological samples, food analysis	122 pg ml^{-1} (0.5 pmol ml^{-1})
Ochromonas danica	Growth-promoting activity	Papain proteolysis	Biological samples	3 pg ml^{-1} (12 fmol ml^{-1})
Kloeckeria brevis ATCC 9774	$^{14}CO_2$ detection from [^{14}C]-methionine	2 N H_2SO_4 (foods) papain proteolysis (plasma)	Plasma, food analysis	10 pg per vial (5.2 ml) (41 fmol)

for methods to quantitate biotin and certain metabolites in biological materials. The competition between radioactive and nonradioactive (cold) biotin for avidin binding is the principle of isotope dilution assays. The radioactivity of the avidin–biotin complex depends on the dilution of the added radioactively labeled biotin by the unknown (nonradioactive) vitamin content in the sample. ^{14}C-Biotin, ^{125}I-biotin, or ^{3}H-biotin can be used as a competing tracer. Techniques to separate protein bound from unbound biotin differ in adsorption and separation of either the avidin–biotin complex (bentonite or nitrocellulose) or the exceeding free biotin (charcoal) in the sample.

Radiochemical assays previously using ^{14}C-labeled material (biotin) have often lacked the sensitivity necessary to measure the physiological concentrations of biotin because of the low specific activity of the available [^{14}C]-carboxybiotin (45 mCi per mmol of biotin). The detection limit of 1 ng of biotin per assay (4.0 pmol) or 5 μg of biotin kg^{-1} (20.5 mmol) was only capable of detecting the biotin content in tissue samples and animal feed.

By using a ^{125}I-labeled conjugated of biotin, a biotin detection limit in the femtomolar range has been made possible, thus being sensitive enough to determine the biotin content in the small volumes of blood available in pediatric studies. The drawback of this assay, however, is the extensive and cumbersome synthesis of the ^{125}I-biotin beforehand.

The lack of sensitivity has been overcome when using ^{3}H-labeled biotin as a tracer, which is commercially available with a considerably enhanced specific radioactivity (30–46 Ci per mmol of biotin). Assays based on ^{3}H-biotin are sensitive enough to quantify biotin in the picomolar, or even femtomolar, range, as required for biological material (tissue, blood, and urine).

Non-bound biotin, as well as free ^{3}H-biotin, is separated by means of adsorption of the avidin–biotin complex on bentonite or filtration through nitrocellulose filters. Alternatively, nonbound biotin can be adsorbed on dextran-coated charcoal. In principle, biotin analogs or metabolites with an intact ureido ring as well as protein-bound biotin are also complexed by avidin and, if present in the sample, should be assayed by the radioisotope dilution methods (see Table 4).

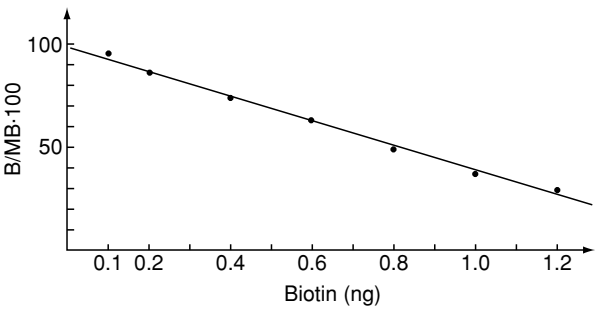

Figure 3 Calibration curve. B, bound ^{3}H-biotin; MB, maximal binding capacity. B = bound ^{3}H-biotin, MB = maximal binding capacity. From *Deutsche Lebensmittel-Rundschau* 82 (1986) 80–83, with permission.

A problem may arise in terms of the extraction procedure of bound biotin from the sample material. With acid hydrolysis, there is the problem of achieving maximal liberation of bound forms without any detectable vitamin losses or oxidative damage. Furthermore, large amounts of salt are produced on neutralization of the acid extracts, which may interfere with the following adsorption of biotin on bentonite or charcoal. Dilution is necessary, therefore, to lower the high salt concentration, but this may create problems in bringing the biotin content to within an appropriate sensitivity range. Such difficulties can be avoided by means of enzymatic cleavage by papain digestion, instead of acid hydrolysis. A typical calibration curve using ^{3}H-biotin as a tracer and avidin as a binder is shown in Figure 3.

This modification of the radioligand test has been applied to biotin determinations in plasma, urine, and animal and plant food samples. Food samples are pretreated with liquid nitrogen followed by homogenization and papain digestion. Comparing analyses by use of the microbiological assay with *L. plantarum* correlated well with this radioligand test. The detection limit is 0.4 pmol of biotin per tube (= 0.1 ng of biotin).

Another modification makes use of a sequential solid-phase assay based on ^{125}I-labeled avidin. A fixed amount of ^{125}I-labeled avidin is incubated with either standards of biotin or sample biotin in serum or urine. An aliquot of this incubate is then transferred to microtiter plates previously coated with biotin linked to bovine serum albumin. Depending on the amount of biotin bound to ^{125}I-biotin in the first incubation step, the counts bound in the second incubation step vary and can be compared with the counts for the standards. The assay sensitivity is 10 fmol per assay well (100 fmol ml^{-1}).

Table 6 summarizes details of the aforementioned isotope dilution assays.

Assessment by Competitive Enzyme Protein Binding Assay (EPBA) or Sequential Solid-Phase Assay

Very recently, highly improved and sensitive protein binding assays have been developed by means of either enzyme-linked avidin or streptavidin, or biotinylated chromophores or fluorophores, respectively.

The principle of the EPBA is similar to isotope dilution assays, to prevent enzyme-coupled avidin from binding to a biotinylated protein adsorbed to plastic by the unknown sample's vitamin content. These tests have been improved by the introduction of microtiter plates in solid-phase sequential assays. In most cases, the detector molecule is avidin linked to horse-radish peroxidase (HRP), and the microtiter plates are coated with bovine serum albumin (BSA) covalently linked to biotin.

The typical sequence of a sequential solid phase assay is as follows:

- A constant amount of avidin linked to horse-radish peroxidase (HRP-avidin) is incubated in a microtiter plate with known amounts of biotin and with dilutions of the sample to produce the standard curve and the unknown response (incubation 1).
- An aliquot of the first incubation is then transferred to corresponding microtiter plates previously coated with biotin covalently linked to bovine serum albumin (biotinyl-BSA) and incubated, too (incubation 2). During this incubation step, HRP-avidin molecules in surplus with unoccupied biotin-binding sites will bind to the biotinyl-BSA. Unbound HRP-avidin molecules with no available biotin-binding sites are washed away thereafter.
- The amount of HRP-avidin bound to the biotinyl-BSA is quantitated by measuring the rate of

Table 6 Isotope dilution techniques of biotin determination

Tracer	*Binder*	*Application*	*Separation of free and bound tracer*	*Detection limit*	*Author*
^{125}I-Biotin conjugate	Avidin	Physiological fluids	Adsorption of unbound biotin on dextran-coated charcoal	< 1.25 pmol ml^{-1}	T Horsburgh/D Gompertz: *Clinica Chimica Acta* 82 (1978) 215
^{14}C-Biotin	Avidin	Biological tissue, animal feed	Precipitation of avidin–biotin	20 pmol ml^{-1}	R L Hood: *Methods in Enzymology* 62 (1979) 279
^{3}H-Biotin	Avidin	Blood serum	Precipitation of avidin–biotin	1.0 pmol ml^{-1}	
^{3}H-Biotin	Avidin	Tissues, serum	Adsorption of avidin–biotin on bentonite	0.5 pmol ml^{-1}	K Dakshinamurti/L Allan: *Methods in Enzymology* 62 (1979) 284
^{3}H-Biotin	Avidin	Biotin analogs	Adsorption of avidin–biotin on bentonite	4 fmol ml^{-1}	A L Landman: *International Journal for Vitamin and Nutrition Research* 46 (1976) 310
^{3}H-Biotin	Avidin	Plasma, urine	Adsorption of avidin–biotin on nitrocellulose	1.0 pmol ml^{-1}	R S Sanghvi/R M Lemons/H Baker/J G Thoene: *Clinica Chimica Acta* 124 (1982) 85
^{3}H-Biotin	Avidin	Plasma, urine, animal and plant food	Adsorption of unbound biotin on dextran-coated charcoal	0.2 pmol ml^{-1}	R Bitsch/I Salz/D Hoetzel: *International Journal for Vitamin and Nutrition Research* 59 (1989) 59
^{125}I-Avidin	Biotin–bovine serum albumin	Serum, urine	Sequential solid-phase assay (molecular sieve chromatography on Sephadex)	0.1 pmol ml^{-1}	D M Mock: *Methods in Enzymology* 184 (1990) 224

peroxidation of *o*-phenylenediamine (OPD) used as an indicator dye. Peroxidation is stopped by the addition of acid (e.g., H_2SO_4), which also intensifies the absorbance of the peroxidized dye (OPD). The absorbance maximum at 490 nm is finally quantitated in a plate-reader spectrophotometer.

Analogous to RIA, a biotin increase in the standard additions or unknown samples decreases the number of free biotin-binding sites on HRP-avidin during the initial incubation. Thus, fewer molecules of free HRP-avidin are bound to the plates coated with biotinyl-BSA, and the final color development is progressively less.

This sequential solid-phase test was applied for biotin measurement in serum, urine, and milk. The detection limit is reported to about 5 fmol of biotin and about 10 parts per trillion (p.p.t.). Separate quantitation of biotin and analogs such as biotinsulfoxide and bisnorbiotin is necessary, because the biotin-binding affinity of HRP-avidin varies among biotin and its metabolites. These derivatives thus have to be separated beforehand by HPLC on RP18 columns, and the binding affinity of each metabolite has to be determined separately. The measured total avidin-binding substances therefore do not equal the stoichiometric sum of the vitamin and its metabolites.

Variations of this sequential solid phase test are described using streptavidin as the binder and biotinylated enzyme, e.g., alkaline phosphatase, and free sample vitamin as the competitors for binding. The activity of alkaline phosphatase as an indicator is determined by measuring the absorbance of *p*-nitrophenylphosphate as a chromogenic enzyme substrate at 405 nm. The biotin-binding affinity of streptavidin is slightly lower than that of avidin, but biotin metabolites are bound much more weakly by streptavidin (see Table 4).

Assays utilizing enzyme-linked biotin as an indicator are applicable to quantify either biotin or avidin/streptavidin. The lower limits of those assays are 0.2–20 pg of biotin per well (0.8 fmol to 0.08 pmol).

A superior replacement for avidin or streptavidin appears to be the introduction of neutravidin, a modified avidin with no carbohydrate components. The deglycosylation procedure is carried out under mild conditions, so as not to affect the stability or the biotin-binding affinity. The suitability for biotin assessments seems to be more advantageous in so far as the cross-reactivity for structure analogs is weaker compared with the other binding proteins (Table 4). An ELISA test has been developed, based on neutravidin-coated microtiter plates and biotinylated HRP competing with sample biotin. The absorbance of the chromogen tetramethylbenzidine after oxidation by the enzyme conjugate and urea peroxide is detected at 450 nm in a plate-reader photometer. The detection limit of this commercial test is 1.1 pg of biotin (4.5 fmol) or 11 ng of biotin per liter of sample (45 pmol). Using this assay, the biotin content of biological materials (plasma, urine) as well as in several plant and animal food samples was analyzed after release of the protein bound vitamin by papain digestion. The biotin content correlated well with microbiologically determined vitamin using *L. plantarum* ($r = 0.973$). A typical biotin calibration curve obtained when using the RIDASCREEN® assay with neutravidin as a solid-phase binder is shown in Figure 4. The principles and detection limits of typical competitive EPBAs are summarized in Table 7.

Enzymatic Determination

An older enzymatic assay is based on the formation of holopyruvate carboxylase from apopyruvate carboxylase and biotin, followed by determination of the pyruvate carboxylase activity in the presence of holoenzyme synthetase, measuring the incorporation of $H\text{-}^{14}CO_3^-$ into oxaloacetate, which is reduced to acid-stable malate. The apoenzyme and the holoenzyme synthetase have been obtained from the biotin-free cultivated yeast, *S. cerevisiae*.

The reaction of the test is as follows:

$$\text{Apopyruvate carboxylase} + \text{biotin} \xrightarrow[\text{Synthetase}]{\text{Holoenzyme}} \text{Pyruvate carboxylase}$$

$$\text{Pyruvate} + H\text{-}^{14}CO_3^- \xrightarrow[\text{Pyruvate carboxylase}]{\text{ATP, } Mg^{2+}, K^+} {}^{14}\text{C-oxaloacetate.}$$

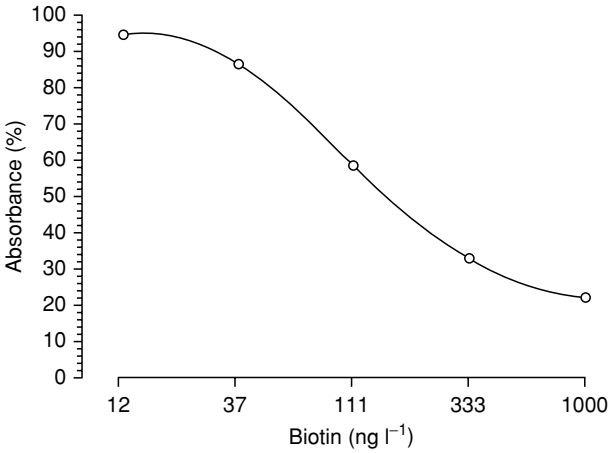

Figure 4 Calibration curve of RIDASCREEN® with biotin.

Table 7 Enzyme immuno(sorbent) assays

Enzyme conjugate	*Indicator dye*	*Solid phase (microtiter plates)*	*Application*	*Detection limit*	*Author*
HRP-Avidin	OPD	Biotinyl-BSA	Serum, milk, urine	5 fmol per well (10 p.p.t.)	D Mock: *Methods in Enzymology* 279 (1997), 265
HRP-Streptavidin	OPD	Biotinyl-BSA	Aqueous solution	0.1 pmol per well	E Z Huang: *Methods in Enzymology* 279 (1997) 304
HRP-Streptavidin	[2,2-azinodi(3-ethyl-benzthiazoline) sulfonic acid	Biotinyl-goat antirabbit IgG	Aqueous solution, clinical samples	1 pg ml^{-1} (4 fmol)	D Shiuan: *Methods in Enzymology* 279 (1997) 321
Biotinyl-HRP	Tetramethylbenzidine + urea peroxide	Neutravidin	Plasma, urine, food samples	4.5 fmol per well	M Helbich-Endermann: Doctoral thesis (1999) Jena/ Germany

The enzyme activity is proportional to the amount of available biotin. The method is somewhat difficult, and the absolute minimum amount of biotin detectable is 5 pg. Of the biotin analogs and precursors, only biocytin has been found to interfere in this test.

Photometric and Fluorimetric Determinations

Based on the high affinity of biotin, earlier methods took advantage of the spectral shift, the fluorescence change, or the chemiluminescence induction after displacement of an avidin-bound dye by biotin. After addition of the azodye 2-(4′-hydroxyazobenzene) benzoic acid (HABA) to avidin, a complex is formed with an absorbance maximum at 500 nm. Titration with biotin quantitatively displaces the dyefrom avidin, leading to a spectral shift, which can be monitored spectrophotometrically. This test can be employed for the quantitation of biotin as well as avidin. The sensitivity is no higher than 2 μg of biotin per ml (8 nmol ml^{-1}). Biotin analogs with an intact ureido ring react in a similar way. Depending on their binding constants, these analogs may reduce the sharpness of the endpoint but do not affect the titer.

Another spectrophotometric assay for biotin and avidin has been proposed, which is based on the inactivation of biotinylated glucose-6-phosphate dehydrogenase when complexed with avidin. The degree of inactivation of the biotinylated enzyme in relation to the avidin concentration is detected in a coupled reaction with NADH at 340 nm. The detection limits have been reported to be 100 ng of biotin per ml (0.4 nmol ml^{-1}) and 2.5 ng of avidin per ml.

Fluorescence and chemiluminescence determinations The competition between biotin and a biotin luminol derivative for avidin was utilized for a chemolumimetric detection. Biotin is covalently linked to the amine function of isoluminol to yield the conjugate. This conjugate emits light when oxidized by either a hydrogenperoxidase–lactoperoxidase or a superoxide. The peak light intensity increases about 10-fold when the conjugate is incubated with avidin prior to oxidation. When the isoluminol-biotin conjugate and free biotin compete for the avidin binding sites, the peak light intensity produced in subsequent chemiluminescent reactions decreases with increasing biotin levels. In this way, biotin is measured quantitatively at levels as low as 50 nM.

A greater sensitivity is achieved by measuring the quenching of the chemiluminescence of aminobutylethylisoluminol (ABEI)-labeled biotin when bound to fluorescein-labeled avidin. The chemiluminescence decrease in the range of 460–525 nm compared with unbound ABEI-biotin is capable of detecting biotin in human sera in the range of 1.2–4.3 nM. However, the disadvantage with this is the laborious preparation of the ABEI-biotin and its purification by TLC and HPLC.

Another modification of fluorescence detection in a competitive binding assay includes HPLC separation of biotin and biocytin coupled with a postcolumn reaction detection system. In one analytical approach, biotin or biocytin displacement of 2-anilinonaphthalene-6-sulfonic acid from its binding site to avidin yields a decrease in the fluorescence intensity at 438 nm. The other analytical approach utilizes the increase in fluorescence of (strept) avidin labeled with fluorescein isothiocyanate (FITC) at 518 nm after binding with the analyte (biotin, biocytin) in the postcolumn eluate. However, the detection limits of both methods are relatively low in the range of 5×10^{-7} M up to 1.8×10^{-8} M biotin or biocytin,

so these methods are only applicable for analyses of the biotin content of vitamin preparations, liquid infant formula, etc.

A maximal intensification of the chemolumimetric biotin detection can be achieved by using the bioluminescent protein aequorin as a label in a competitive binding assay. Aequorin is a photoprotein from the jellyfish *Aequorea victoria*. Aequorin associates noncovalently with its chromophore coelenterazine, which is oxidized into a metastable excited state, resulting in the emission of photons at 469 nm.

In a homogenous assay, the aequorin–biotin conjugate competes in a one-step procedure with free biotin for binding to avidin in solution. Binding of the conjugate to avidin results in partial quenching of the bioluminescence signal, and so the decrease in luminescence can be related to the concentration of biotin in a sample. The detection limit is reported to be around 10^{-14} M biotin (= 10 fM).

The competition assay can also be performed in a heterogenous system by using immobilized avidin (agarose, polyacrylamide). Both methods need special equipment (luminometer operating in a photon-counting mode). The amounts of avidin and biotinylated aequorin must be optimized beforehand, as the relative amounts of avidin and biotinylated aequorin influence the response characteristics of the assay (e.g., detection limits and sensitivity).

Chromatographic Analysis

Earlier methods attempted to separate and quantify biotin from vitamin mixtures or feed supplements. After dissolution of the lyophilized preparation and addition of the internal standard (2-imidazolidone), the sample was applied on a TLC plate and eluted with chloroform/methanol/formic acid. Biotin was visualized by spraying with *p*-dimethylaminocinnamaldehyde and determined *in situ* by reflectance measurements. Spraying with paraffin after the coloring procedure increased the sensitivity to a detection limit of 10 ng of biotin per spot.

Quantitation by HPLC requires derivatization of the molecule, because biotin does not absorb in the visible or UV region. For the derivatization 9-anthryldiazomethane was used to produce fluorescent biotin-9-anthrylmethylester followed by separation on C 18 bonded columns with acetonitrile/water as the mobile phase and fluorimetric detection (excitation = 365 nm, emission = 425 nm). Alternatively, *p*-bromophenacyl bromide esters and 4-bromomethylmethoxycoumarin derivatives of biotin were produced for UV and fluorescence detection, respectively, after separation on RP 18 columns with methanol/water or tetrahydrofuran/water as eluents.

Standards of biotin and 13 metabolites and analogs can be separated by HPLC on a C 18 RP phase using gradient elution with trifluoroacetic acid/acetonitril and UV detection at 220 nm. An HPLC separation of six vitamins including biotin from almonds has been performed on an LC-8-DB column with hexanesulfonic acid sodium salt/methanol as the eluent and detection at 200 nm. The limit of biotin detection has been found to be as low as 0.9 mg per 100 g.

When biotin is derivatized with panacyl bromide in the presence of crown ether, a fluorescence compound results with fluorescence maxima at 380 nm (excitation) and 470 nm (emission). After extraction from biological tissue with trichloroacetic acid and purification by solid-phase extraction combined with ion-exchange chromatography on DOWEX and TLC, biotin is derivatized with panacyl bromide in the presence of crown ether. The resulting panacylester can be separated on normal-phase HPLC using isocratic elution with methanol/dichloromethane as well as on RP phase HPLC using gradient elution (water/methanol). Dethiobiotin is taken as an internal standard.

With this method, the biotin content in rat small intestine has been assessed. The detection limits are reported to be 10 pmol of biotin (normal phase HPLC) or 100 pmol of biotin (RP-HPLC), which is less sensitive than with protein-binding assays. Criticism results from the facts, that the clean-up required for biological sample extracts prior to derivatization is very extensive, the derivatization should not only include biotin, and the cross-reactivity to biotin analogs has to be examined.

See also: **Analysis of Food**; **Chromatography**: High-performance Liquid Chromatography; **Enzymes**: Uses in Analysis; **Food Composition Tables**; **Immunoassays**: Principles; Radioimmunoassay and Enzyme Immunoassay; **Lactic Acid Bacteria**; **Microbiology**: Classification of Microorganisms; **Spectroscopy**: Fluorescence; **Vitamins**: Determination

Further Reading

Baker H, Frank O, Matovitch VB *et al.* (1962) A new assay method for biotin in blood, serum, urine and tissue. *Analytical Biochemistry* 3: 31–39.

Bitsch R, Salz I and Hoetzel D (1986) Biotin assessment in foods and body fluids by a protein binding assay. *International Journal for Vitamin and Nutrition Research* 59: 59–64.

Bonjour JP (1991) Biotin. In: Machlin LJ (ed.) *Handbook of Vitamins*, pp. 393–427. New York: Marcel Dekker.

Eitenmiller RR and Landen WO, Jr. (1999) *Vitamin Analysis for the Health and Food Sciences*. Boca Raton, FL: CRC Press.

Fidanza F (1991) *Nutritional Status Assessment*. London: Chapman & Hall.

Green NM (1970) Spectrophotometric determination of avidin and biotin. *Methods in Enzymology* 18: 418–424.

Guilarte T (1985) Measurement of biotin levels in human plasma using a radiometric-microbiological assay. *Nutrition Reports International* 31: 1155–1163.

Hentz NG and Bachas LG (1997) Fluorophore-linked assays for high-performance liquid chromatography postcolumn reaction detection of biotin and biocytin. *Methods in Enzymology* 279: 275–286.

Lizano S, Ramanathan S, Feltus A, Witkowski A and Dannert S (1997) Bioluminescence competitive binding assays for biotin based on photoprotein aequorin. *Methods in Enzymology* 279: 296–303.

Schroeder HR, Vogelhut PO, Carrilco RJ, Boguslaski RC and Buckler RT (1976) Competitive protein binding assay for biotin monitored by chemoluminescence. *Analytical Chemistry* 48: 1933–1937.

Physiology

J Zempleni, University of Nebraska-Lincoln, Lincoln, NE, USA

D M Mock, University of Arkansas for Medical Sciences, Little Rock, AR, USA

Introduction

Biotin is a water-soluble vitamin that serves as essential coenzyme for four mammalian carboxylases. Biotin may also play a role in the replication and transcription of DNA, possibly by an effect on biotinylation of histones. Frank biotin deficiency causes clinical findings such as hair loss and skin rash. Marginal biotin deficiency is frequently seen in pregnant women and theoretically could have teratogenic effects. Biotin status may be assessed using urinary excretion of biotin and its metabolites, and certain organic acids that are byproducts of deficiency of biotin-dependent carboxylases. Direct measurement of biotin-dependent carboxylases and biotin in lymphocytes and fibroblasts also show promise as indicators of biotin status. Inborn errors of biotin metabolism (e.g., biotinidase deficiency and holocarboxylase synthetase deficiency to a variable degree) can produce symptoms similar to biotin deficiency. For adults, the safe and adequate daily dietary intake of biotin is 30 μg. Available data suggest that pure biotin administered orally is absorbed completely. Uptake of biotin into cells of the intestine and peripheral tissues is mediated by a sodium-dependent transporter that requires metabolic energy. Mammals catabolize biotin by β-oxidation of the valeric acid side chain and by sulfur oxidation in the thiophane ring. Biotin, bisnorbiotin, and biotin, *d,l*-sulfoxide are the most abundant biotinyl compounds in urine and plasma.

Biochemical Function

In mammals, holocarboxylase synthetase (EC 6.3.4.10) catalyzes the covalent binding of biotin to the ϵ-amino group of lysine in four different apocarboxylases to form the active holocarboxylases (**Figure 1**). For acetyl-CoA carboxylase (EC 6.4.1.2), a cytosolic (acetyl-CoA carboxylase α) and a mitochondrial form (acetyl-CoA carboxylase β) have been identified. Both the α and β forms catalyze the binding of bicarbonate to acetyl-CoA to form malonyl-CoA; the latter is a substrate for fatty acid synthesis. Although the α and β forms of acetyl-CoA carboxylase catalyze the same reaction, they appear to have different roles in intermediary metabolism. Acetyl-CoA carboxylase α controls fatty acid synthesis in the cytosol by providing the substrate malonyl-CoA; acetyl-CoA carboxylase β controls fatty acid oxidation inside the mitochondria via malonyl-CoA inhibition of fatty acid transport into mitochondria. (*See* **Fatty Acids**: Metabolism.)

The three other mammalian biotin-dependent carboxylases are located exclusively in mitochondria: pyruvate carboxylase (EC 6.4.1.1), propionyl-CoA carboxylase (EC 6.4.1.3), and β-methylcrotonyl-CoA carboxylase (EC 6.4.1.4). By synthesizing oxaloacetate, pyruvate carboxylase provides a tricarboxylic acid cycle intermediate and catalyzes a step in gluconeogenesis. Propionyl-CoA carboxylase catalyzes an essential step in the metabolism of isoleucine, valine, the cholesterol side chain, and products of dietary carbohydrate breakdown by intestinal microorganisms. β-Methylcrotonyl-CoA carboxylase catalyzes an essential step in leucine metabolism.

Biotin deficiency causes reduced carboxylase activities; substrates are shunted to alternative pathways (**Figure 1**). For example, reduced activity of propionyl-CoA carboxylase results in increased formation of 3-hydroxypropionic acid and 2-methylcitric acid; reduced activity of β-methylcrotonyl-CoA carboxylase results in increased formation of 3-hydroxyisovaleric acid and 3-methylcrotonyl glycine. Increased urinary excretion of these organic acids has been used to diagnose biotin deficiency, as discussed below.

Biotin may also play a role in the regulation of DNA transcription and replication on the basis of the following observations:

1. Biotinidase (EC 3.5.1.12) specifically biotinylates histones. Histones are DNA-binding proteins that

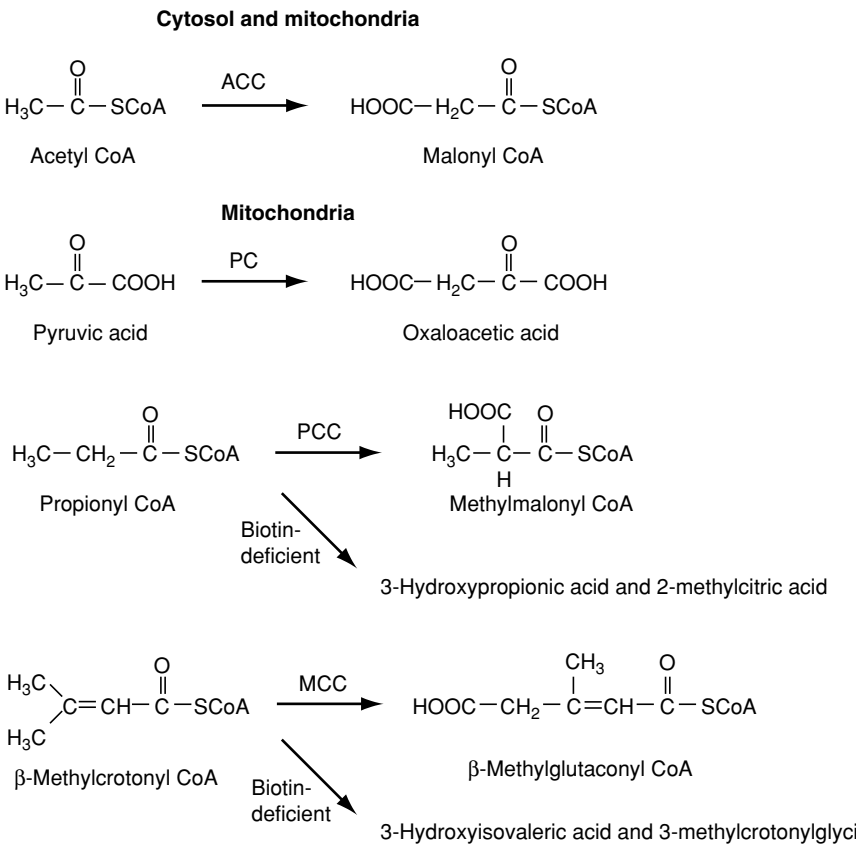

Figure 1 Biotin-dependent carboxylases. ACC, acetyl-CoA carboxylase; PC, pyruvate carboxylase; PCC, propionyl-CoA carboxylase; MCC, β-methylcrotonyl-CoA carboxylase.

regulate transcription and replication of DNA. Biotinylation of histones might play some role in DNA packaging in analogy to other covalent modifications such as acetylation, methylation, phosphorylation, or adenosine diphosphate (ADP)-ribosylation of histones. Biotinidase is ubiquitous in mammalian cells, and 25% of the cellular biotinidase activity is located in the nucleus. Indeed, the following observations are also consistent with a role for biotin in modifying histones and in turn affecting the packaging of DNA: first, histones dissociate from the DNA in biotin-deficient rats, and second, biotin deficiency in rats results in decreased phosphorylation and methylation of histones as well as increased acetylation of histones.

2. Rates of biotin uptake are greater in proliferating than in nonproliferating lymphocytes. This increase is mediated by an increased number of biotin transporters on the cell surface. Cell proliferation requires a substantial increase of both replication and transcription. Theoretically, the observed increased uptake of biotin by proliferating cells may be in response to increased demand for biotin to biotinylate histones. This is a potential mechanism by which biotin exerts an effect on cell replication.
3. Biotin status affects gene expression. Biotin deficiency causes a 40–45% reduction of rat liver glucokinase activity; the activity is restored by biotin administration. Injection of biotin into rats causes a 4–19-fold increase of mRNA encoding glucokinase.

In addition to this effect on mRNA synthesis, biotin may affect the synthesis of some proteins at the posttranscriptional level. When human-derived liver (HepG2) cells were cultured in biotin-deficient medium, expression of the asialoglycoprotein receptor was reduced despite normal protein synthesis, total cellular protein content, and mRNA coding for asialoglycoprotein receptor; addition of biotin or biocytin restored receptor expression. These observations suggest that a biotin-dependent posttranscriptional event can affect the ultimate expression of asialoglycoprotein receptor.

Storage, Metabolism, and Excretion

Depletion and repletion experiments of biotin-dependent carboxylases in rat liver provided evidence that mitochondrial acetyl-CoA carboxylase may serve as a reservoir for biotin. Neither cytosolic acetyl-CoA carboxylase nor the mitochondrial pyruvate carboxylase, propionyl-CoA carboxylase, or β-methylcrotonyl-CoA carboxylase seem to serve as biotin reservoirs.

The liver accumulates a significant percentage of ingested biotin. For example, rat liver accumulates approximately 4% of [^{14}C]biotin within 1 h of intravenous administration. Once steady-state conditions are attained, biotin is located mainly in liver mitochondria (30% of total [^{14}C]biotin) and cytosol (59%); smaller amounts can be found in microsomes and nuclei. Greater than 80% of the [^{14}C]biotin present in the cytosolic fraction is acid-precipitable, suggesting that biotin has become covalently bound.

During the normal intracellular turnover of proteins, holocarboxylases are degraded to biotin linked to lysine (biocytin) or biotin linked to an oligopeptide containing at most a few amino acid residues. Biotinidase catalyzes the hydrolysis of the amide bond between biotin and lysine. Biotinidase likely serves in both absorption of biotin (cleavage of protein-bound dietary biotin) and recycling of biotin (cleavage of biocytin). Biotinidase also catalyzes a specific biotinylation of histones, as noted above. (*See* **Protein**: Synthesis and Turnover.)

A significant proportion of biotin undergoes catabolism before excretion (Figure 2). Two principal pathways of biotin catabolism have been identified in mammals. In the first pathway, the valeric acid side chain of biotin is degraded by β-oxidation. β-Oxidation of biotin leads to the formation of bisnorbiotin, tetranorbiotin, and related intermediates that are known to result from β-oxidation of fatty acids. The cellular site of this β-oxidation of biotin is uncertain. Spontaneous (nonenzymatic) decarboxylation of the unstable β-keto acids (β-keto-biotin and β-keto-bisnorbiotin) leads to formation of bisnorbiotin methyl ketone and tetranorbiotin methyl ketone; these catabolites appear in urine.

In the second pathway, the sulfur in the thiophane ring of biotin is oxidized, leading to the formation of biotin *l*-sulfoxide, biotin *d*-sulfoxide, and biotin sulfone. Sulfur oxidation may be catalyzed by a NADPH-dependent process in the smooth endoplasmic reticulum. Combined oxidation of the ring sulfur and β-oxidation of the side chain lead to metabolites such as bisnorbiotin sulfone. In mammals, degradation of the biotin ring to release carbon dioxide and urea is quantitatively minor.

On a molar basis, biotin accounts for approximately half of the total avidin-binding substances in human serum and urine (Table 1). Bisnorbiotin, bisnorbiotin methyl ketone, biotin *d,l*-sulfoxide, and biotin sulfone account for most of the balance. Using thin-layer chromatography and staining with *p*-dimethylaminocinnamaldehyde, tetranorbiotin *l*-sulfoxide was also identified in human urine. However, avidin binding of this metabolite was too weak to allow quantitation.

The biliary route of biotin excretion is quantitatively minor. For example, in rats, approximately 2% of intravenously administered [^{14}C]biotin was excreted in bile but approximately 61% was excreted in urine.

The relationship of metabolite profile to biotin nutritional status has not been fully elucidated. In human and rat urine, the percent excretion of biotin increases when the biotin intake is increased from physiologic to pharmacologic amounts. This may reflect saturation of renal reabsorption, metabolic pathways, or both.

Deficiency

Clinical Symptoms of Frank Biotin Deficiency

The fact that humans have a requirement for biotin has been clearly documented in two situations: first, parenteral nutrition without biotin supplementation in patients with short-gut syndrome and other causes of malabsorption; and second, prolonged consumption of raw egg white. The critical event in the egg white-induced biotin deficiency is a highly specific and very tight binding ($k_b = 10^{15}\,\mathrm{mol\,l^{-1}}$) of biotin by avidin, a glycoprotein found in egg white. Avidin is resistant to intestinal proteolysis in both the free and the biotin-bound form. Thus dietary avidin binds and prevents the absorption of biotin. Cooking denatures avidin, rendering it susceptible to digestion and hence unable to interfere with absorption of biotin. (*See* **Eggs**: Dietary Importance.)

Biotin deficiency has also been reported or inferred in several other circumstances, including pregnancy, individuals undergoing dialysis, individuals suffering from chronic gastrointestinal disease, Leiner's disease, and sudden infant death syndrome.

The clinical findings of frank biotin deficiency in adults and older children are similar regardless of whether deficiency is caused by egg-white feeding or omission of biotin from parenteral nutrition. Typically, the findings begin to appear gradually after an interval of 6 months to 3 years of parenteral nutrition or after 6 weeks to several years of egg-white feeding. Thinning of hair, often with loss of hair color, was

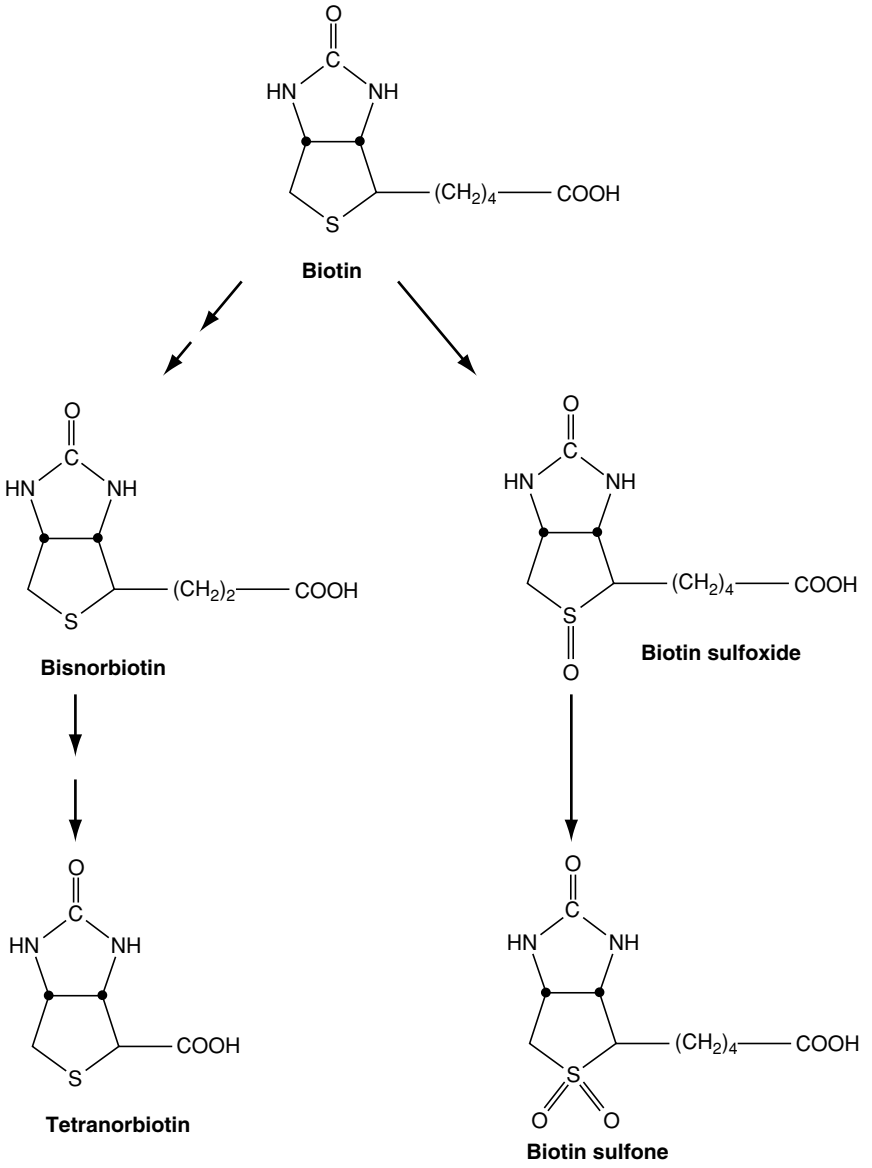

Figure 2 Pathways of biotin catabolism.

reported in most patients. A skin rash described as scaly (seborrhoeic) and red (eczematous) was present in the majority; in several, the rash was distributed around the eyes, nose, and mouth. Depression, lethargy, hallucinations, and paresthesias of the extremities were prominent neurological symptoms in the majority of adults.

In infants who developed biotin deficiency, the signs of deficiency began to appear within 3–6 months of initiation of total parenteral nutrition. The rash initially appeared around the eyes, nose, and mouth; ultimately, the ears and perineal orifices were involved. The appearance of the rash was similar to that of cutaneous candidiasis (i.e., an erythematous base and crusting exudates); typically, *Candida* could be cultured from the lesions. The character and distribution of the biotin deficiency rash are quite similar to the rash of zinc deficiency. In some infants, hair loss can progress to total baldness, including loss of eyebrows and lashes. The most striking neurological findings in biotin-deficient infants was hypotonia, lethargy, and developmental delay. In these infants, a peculiar withdrawn behavior was often noted and may have reflected the same central nervous system dysfunction diagnosed as depression in adult patients.

Biochemical Markers of Biotin Deficiency

Our studies provide evidence that marginal biotin deficiency is much more common than frank biotin

Table 1 Biotin and metabolites in human serum and urine

Compound	*Serum (pmol l^{-1})*	*Urine (nmol 24 h^{-1})*
Biotin	244 ± 61	35 ± 14
Bisnorbiotin	189 ± 135	68 ± 48
Biotin *d,l*-sulfoxide	15 ± 33	5 ± 6
Bisnorbiotin methyl ketone	ND[a]	9 ± 9
Biotin sulfone	ND[a]	5 ± 5
Total biotinyl compounds	464 ± 178[b]	122 ± 66

Means ± SD are reported ($n = 15$ for serum; $n = 6$ for urine).
[a]ND, not determined. Bisnorbiotin methyl ketone and biotin sulfone had not been identified at the time when this study of serum was conducted and hence these unknowns were not quantitated against authentic standards.
[b]Including three unidentified biotin metabolities.
From Zempleni J and Mock DM (1999) Biotin biochemistry and human requirements. *Journal of Nutrition and Biochemistry* 10: 128–138, with permission.

deficiency. Diagnosis of marginal biotin deficiency requires sensitive metabolic indicators. The best validated indicators of marginal biotin deficiency are urinary excretion of biotin and metabolites and urinary excretion of organic acids such as 3-hydroxyisovaleric acid (see above: Figure 1). These organic acids are synthesized and excreted in increased quantities in biotin-deficient individuals because the normal catabolism of their precursors by biotin-dependent carboxylases is reduced. Activities of biotin-dependent carboxylases in lymphocytes and fibroblasts and plasma concentrations of biotin and metabolites have been reported to be abnormally low in individuals who developed frank biotin deficiency after 1 month of biotin-free intravenous feeding.

Teratogenic Effects of Biotin Deficiency

Biotin deficiency is teratogenic in several animal species at degrees of deficiency that produce no obvious findings in the pregnant animal. Hens with marginal biotin deficiency produce eggs with higher embryonic mortality, reduced hatchability, chrondystrophy ('parrot beak' deformity), perosis (an abnormality of bone tendon formation that results in a deformity analogous to 'club foot'), micromelia, and syndactyly. Effects on hatchability and viability have been reported in turkey poults. In some strains of mice, asymptomatic biotin deficiency during pregnancy causes substantial increases in fetal malformations and mortality. Ninety-four percent of pups from biotin-deficient dams had cleft palate, 85% had micrognathia, and 41% had micromelia; multiple malformations were common.

Biotin status may be reduced during human pregnancy. In the majority of pregnant women, excretion of 3-hydroxyisovaleric acid is significantly greater than the upper limit of normal. The serum concentration of biotin commonly decreases significantly from early to late pregnancy and reaches values that are below the lower limit of normal in some individuals. Likewise, the urinary excretion rates of biotin and bisnorbiotin are often significantly less in late pregnancy than in early pregnancy.

Inborn Errors of Metabolism

Two important classes of inborn errors of biotin metabolism have been identified:

1. Biotinidase deficiency is an inborn error of metabolism that is characterized by partial (70–90%) or profound (>90%) reduction of biotinidase activity. The combined incidence of partial plus profound biotinidase deficiency is 1: 60 000 newborns. Clinical findings of biotinidase deficiency largely mimic biotin deficiency and include developmental delay, seizures, skin infection, hepatomegaly, and splenomegaly. These symptoms are thought to be a consequence of one or more impairments leading to reduced biotin status: (i) deficient salvage of covalently bound biotin during holocarboxylase turnover may result from incomplete release of biotin from biocytin; (ii) substantial amounts of biocytin are lost in urine; (iii) release of covalently bound dietary biotin may be impaired, leading to malabsorption. Biotinidase-deficient patients respond well to physiologic or pharmacologic doses of biotin.
2. Multiple carboxylase deficiency is an inborn error of metabolism in humans that is caused by deficiency of holocarboxylase synthetase, leading to less efficient biotinylation of apocarboxylases. Isolated carboxylase deficiencies are caused by mutations of genes encoding the mammalian apocarboxylases.

(*See* **Inborn Errors of Metabolism**: Overview.)

Intestinal Absorption

Studies in rats provided evidence that intestinal uptake of biotin occurs by two distinct mechanisms. The first is a saturable, structurally specific transporter located in the brush border membrane. This transporter is dependent on sodium and energy, and is electroneutral. As judged by the maximal transport rate (V_{max}) of the transport process, the transporter is appropriately up- and downregulated by biotin deficiency and biotin excess, respectively. In the rat, biotin transport is most active in the upper small bowel; activity of biotin transport declines proceeding from the jejunum to the ileum to the colon. However, activity in the colon is still 3% of jejunal values, leaving open the possibility that biotin synthesized by

enteric flora might contribute to absorbed biotin. In the rat, activity of the biotin transporter increases with age from suckling to weanling to young adult (aged 3 months). As with the other changes observed in the biotin transporter, the mechanism leading to increased transport appears to be an increase in the number or activity of the carriers, rather than changes in the affinity of the carrier.

The second mechanism of intestinal biotin absorption is passive diffusion. This process predominates at large luminal biotin concentrations; at that point, the biotin transporter is likely saturated.

The biotin transporter in human intestine is similar to that identified in rats. The human transporter is saturable, electroneutral, sodium-dependent, and capable of accumulating biotin against a concentration gradient. The anatomical distribution along the human gastrointestinal tract is similar to that observed in the rat. Changes in transport activity appear to be mediated primarily by an increase in the number or activity of transporters, rather than by changes in the affinity of the transporter. The exit of biotin from the enterocyte (i.e., transport across the basolateral membrane) is also carrier-mediated; however, biotin export is independent of sodium, is electrogenic, and cannot accumulate biotin against a concentration gradient.

Recently, a biotin transporter in human tissues has been cloned, sequenced, and expressed in mammalian cells. This multivitamin transporter binds and transports not only biotin but also pantothenic acid and lipoic acid. Binding affinities are similar for these nutrients.

Clinical studies have also provided evidence that biotin is absorbed from the human colon. When biotin is instilled directly into the lumen of the colon, the plasma concentration of biotin increases; however, when the same dose of biotin is given orally, the increase in plasma concentration is greater.

Bioavailability

Recent reports provide evidence that biotin is absorbed almost completely even if large doses of biotin are ingested. However in these studies, biotin was administered in aqueous solution rather than being endogenous to the diet. A large percentage of biotin in food may be bound to protein and therefore is likely to require cleavage from the protein in order to be transported into enterocytes by the biotin transporter. Biotinidase is thought to be responsible for release of biotin from protein. The release of biotin might occur in or near the intestine mucosa via the action of mucosal biotinidase; biotinidase in pancreatic juice might release biotin during the luminal phase of proteolysis. Theoretically, biotinyl oligopeptides might also be absorbed directly, either by a specific biotin transporter or by a peptide transporter. However, the mechanism remains to be elucidated. Biotinidase is present in pancreatic juice and in intestinal mucosa, but biotinidase activity is not enriched in intestinal brush border membranes. Currently, only limited data are available regarding the bioavailability of protein-bound biotin in food. (*See* **Bioavailability of Nutrients**.)

Transport of Biotin from the Intestine to Peripheral Tissues

Biotinidase has been proposed to serve as a biotin-binding protein in plasma and as a carrier protein for the transport of biotin into the cell. Two biotin-binding sites on biotinidase have been identified. One had an equilibrium dissociation constant (k_d) of $3\,\text{nmol}\,\text{l}^{-1}$ and a maximum binding capacity ($B_{\max}$) of 0.065 mol of biotin per mol of biotinidase; the other had a k_d of $59\,\text{nmol}\,\text{l}^{-1}$ and a $B_{\max}$ of 0.79 mol of biotin per mol of biotinidase. However, empirical studies with [^{3}H]biotin provide evidence that biotin binds primarily to albumin. Of the total biotin in human plasma, 81% is free, 12% is covalently bound, and 7% is reversibly bound.

Transport of Biotin in Liver and Peripheral Tissues

Biotin Uptake into Liver

Rat liver cells accumulate biotin by a transporter-mediated process that depends on metabolic energy; uptake is sodium-dependent and Na-K-ATPase-dependent. The ureido portion of the biotin molecule plays a role in binding to the transporter. Organic anions such as bilirubin and cholic acid compete with biotin uptake by hepatocytes.

Biotin uptake into HepG2 cells is also transporter-mediated, energy-dependent, sodium-dependent, and Na-K-ATPase-dependent; a free carboxyl group on the valeric acid side chain of biotin is essential for recognition by this transporter. Apparent Michaelis constant (K_m) and $V_{\max}$ in HepG2 cells are $19.2\,\mu\text{mol}\,\text{l}^{-1}$ and $6.8\,\text{nmol mg protein}^{-1}\,\text{min}^{-1}$, respectively.

Biotin Uptake into Lymphocytes

In analogy to liver cells, human lymphocytes accumulate biotin by a transporter that requires metabolic energy and is dependent on Na-K-ATPase. Apparent K_m and $V_{\max}$ values of this biotin transporter are

2.6 nmol l^{-1} and 2.9 fmol 10^6 cells^{-1} 30 min^{-1}, respectively. In lymphocytes, the thiophane portion of the biotin molecule is more important than the ureido portion for binding to the transporter. The biotin transporter in lymphocytes is very specific for biotin; inhibition of biotin transport by hexanoic acid, lipoic acid, bilirubin, and pantothenic acid is minor. The same transporter that mediates biotin uptake into lymphocytes also accounts for biotin efflux from these cells as judged by countertransport experiments.

Biotin Uptake into the Central Nervous System

Biotin enters the central nervous system by a saturable transport system. A free carboxylic acid group is important for transport across cerebral capillaries; pantothenic acid and nonanoic acid compete for uptake, suggesting that uptake is mediated by an organic anion carrier. In the brain, biotin is covalently bound to proteins, presumably biotin-dependent carboxylases.

Placental Transport of Biotin

A biotin transporter is present in the placental brush border membrane. In mammals, a single protein termed the multivitamin transporter can transport biotin, pantothenic acid, and lipoic acid across the placenta; this protein has been cloned and functionally expressed. A recent study of fetal and maternal plasma concentrations of biotin at 18–24 weeks' gestation of normal human pregnancies reported a fetal-to-maternal biotin ratio ranging from 3 to 17:1, consistent with active transport across the placenta.

Requirements and Adequate Intakes

Adequate Intakes

For biotin, scientific knowledge is currently insufficient to estimate average requirements and formulate recommended dietary allowances. Notwithstanding, the Food and Nutrition Board of the National Research Council has published recommendations for adequate intake of biotin (Table 2). These data are based on empirical biotin intakes in a group of healthy people.

Table 2 Adequate intake of biotin

Life-stage group	*Adequate intake (μg day^{-1})*
Infants	
0–6 months	5
7–12 months	6
Children	
1–3 years	8
4–8 years	12
Males and females	
9–13 years	20
14–18 years	25
≥19 years	30
Pregnancy	30
Lactation	35

From Yates AA, Schlicker SA and Suitor CW (1998) Dietary reference intakes: the new basis for recommendations for calcium and related nutrients, B vitamins, and choline. *Journal of the American Dietary Association* 98: 699–706, with permission.

Factors that Affect Biotin Requirements

Pregnancy (see above) and lactation appear to produce an increased demand for biotin. Biotin secretion into breast milk accounts for the higher recommendations for adequate intake of biotin in lactating compared to normal women (Table 2). At 8 days postpartum, biotin in human milk is approximately 8 nmol l^{-1} and accounts for 44% of biotin plus measured catabolites; bisnorbiotin and biotin *d,l*-sulfoxide account for 48% and 8%, respectively. By 6 weeks postpartum, the biotin concentration increases to approximately 30 nmol l^{-1} and accounts for about 70% of biotin plus catabolites; bisnorbiotin and biotin *d,l*-sulfoxides account for about 20% and less than 10%, respectively.

Accumulating data provide evidence that long-term anticonvulsant therapy in adults and children can lead to biotin depletion severe enough to interfere with amino acid metabolism. This conclusion was originally based on decreased plasma concentrations of biotin as determined by the *Lactobacillus plantarum* bioassay; the incidence of biotin concentrations below the normal range was approximately 75% in a cumulative study group of 274 adults undergoing long-term therapy with a variety of anticonvulsant drugs. The diagnosis of biotin deficiency was strengthened by the demonstration of increased urinary excretion of the characteristic organic acids (e.g., 3-hydroxyisovaleric acid) in adults and children receiving long-term anticonvulsant therapy. (*See* **Drug–Nutrient Interactions**.)

The mechanism of biotin depletion during anticonvulsant therapy is not known. The anticonvulsant drugs implicated include phenobarbital, phenytoin, carbamazepine, and primidone. These drugs each have a carbamide (-NH-CO-) moiety in their structure, as does biotin; in some cases they incorporate a full ureido group (-NH-CO-NH-). Physiological concentrations of primidone and carbamazepine specifically and directly inhibit biotin uptake by brush border membrane vesicles from human intestine. This finding suggests that impaired intestinal absorption of biotin may contribute to biotin deficiency. In addition, phenobarbital, phenytoin, and carbamazepine displace biotin from biotinidase and thus

conceivably could have an effect on plasma transport of biotin, renal handling of biotin, or cellular uptake of biotin. Recently, increased urinary excretion rates of biotin *d,l*-sulfoxides and bisnorbiotin have been demonstrated in adults and children receiving long-term anticonvulsant therapy, raising the possibility that accelerated catabolism of biotin contributes to the biotin deficiency associated with long-term anticonvulsant therapy.

Intake

Biotin is widely distributed in foodstuffs, but the content of even the richest sources is low when compared to the content of most other water-soluble vitamins. Foods relatively rich in biotin include egg yolk, liver, and some vegetables. The average dietary biotin intake has been estimated to be 70 μg per day for the Swiss population. These data are in reasonable agreement with the estimated dietary intake of biotin in a composite Canadian diet (62 μg per day). Calculated intake of biotin for the British population was 35 μg per day.

Infants who ingest 800 ml of mature breast milk per day ingest approximately 6 μg (24 nmol) of biotin. It remains unclear whether biotin synthesis by gut microorganisms contributes importantly to the total biotin absorbed. However, an infant developed biotin deficiency while consuming a biotin-free elemental formula.

See also: **Bioavailability of Nutrients**; **Drug–Nutrient Interactions**; **Eggs**: Dietary Importance; **Fatty Acids**: Metabolism; **Inborn Errors of Metabolism**: Overview; **Protein**: Synthesis and Turnover

Further Reading

Hymes J and Wolf B (1999) Human biotinidase isn't just for recycling biotin. *Journal of Nutrition* 129: 485S–489S.

Kim K-H (1997) Regulation of mammalian acetyl-coenzyme A carboxylase. In: McCormick DB, Bier DM and Goodridge AG (eds) *Annual Review of Nutrition*, pp. 77–99. Palo Alto, CA: Annual Reviews.

Knowles JR (1989) The mechanism of biotin-dependent enzymes. *Annual Review of Biochemistry* 58: 195.

Mock DM (1996) Biotin. In: Ziegler EE and Filer LJ Jr (eds) *Present Knowledge in Nutrition*, 7th edn, pp. 220–235. Washington, DC: International Life Sciences Institutes, Nutrition Foundation.

Wang H, Huang W, Fei Y-J *et al.* (1999) Human placental Na^+-dependent multivitamin transporter. *Journal of Biological Chemistry* 274: 14875–14883.

Wolf B (1991) Worldwide survey of neonatal screening for biotinidase deficiency. *Journal of Inherited Metabolic Disease* 14: 923–927.

Wolf B and Heard GS (1991) Biotinidase deficiency. In: Barness L and Oski F (eds) *Advances in Pediatrics*, pp. 1–21. Chicago: Book Medical Publishers.

Wood HG and Barden RE (1977) Biotin enzymes. *Annual Review of Biochemistry* 46: 385–413.

Yates AA, Schlicker SA and Suitor CW (1998) Dietary reference intakes: the new basis for recommendations for calcium and related nutrients, B vitamins, and choline. *Journal of the American Dietary Association* 98: 699–706.

Zempleni J and Mock DM (1999) Biotin biochemistry and human requirements. *Journal of Nutrition and Biochemistry* 10: 128–138.

Zempleni J and Mock DM (2000) Marginal biotin deficiency is teratogenic. *Proceedings of the Society for Experimental and Biological Medicine* 223: 14–21.

BISCUITS, COOKIES, AND CRACKERS

Contents

Nature of the Products

S Zydenbos and V Humphrey-Taylor,
New Zealand Institute for Crop & Food Research Limited, Christchurch, New Zealand

Background

Biscuits, cookies, and crackers are enjoyed the world over by consumers who select their favorites based on flavor and appearance, but to bakers, distinguishing between these baked products can be a complex task. The basic characteristic that separates a biscuit, cookie, or cracker from other baked products, such as bread or cake, is a moisture content below 5%. Biscuits, cookies, and crackers have a cereal base of at least 60%. This is usually wheat but is sometimes oat, barley, rye, or rice. Other major ingredients are fat or shortening, and sugar. While water is also added to a biscuit, cookie, or cracker dough, this is not a major component of the final product since it evaporates during baking.

In this article, the history and usage of biscuits, cookies, and crackers are examined. This is followed by a discussion of the classification of biscuits, cookies, and crackers, the definitions of which overlap and are geographically dependent. Finally, some features of the different types of biscuits, cookies, and crackers are considered.

History and Usage

The origin of the name biscuit comes from Latin, where *bis coctus* means twice-baked. There is also an Old French word, *bescoit*, that has a similar meaning. It is thought that these products have been baked for thousands of years. The original process consisted of baking the biscuits in a hot oven and subsequently drying them in a cool oven. It is very rare to find this double baking technique in modern biscuit production. However, a special type of microwave oven is often connected after a gas-fired oven in factories. Cookie is derived from a Dutch word, *koekje*, which means little cake, while the sound of a cracker being eaten probably led to use of that name.

Biscuit, as defined above, is a term used in the UK (and in New Zealand, Australia, and South Africa). The word biscuit is also used for a baked product in the USA. This is a leavened bread-like product that is similar to the UK scone and is not discussed in this article. In the USA, the word cookie describes the same type of products that are called biscuits in the UK. Cracker is a generic term used throughout the world and refers to products with very low sugar and fat contents.

The low moisture content of biscuits, cookies, and crackers means that these products have a long shelf-life and a relatively low risk of spoilage by micro-organisms. This has led to a history of usage in epic journeys, such as the flight of the Israelites from Egyptian slavery and the sea voyages of the fifteenth-century explorers. British and European traditions usually served biscuits in a semiformal situation with tea or coffee between main meals, especially in the afternoon. Small biscuits were usually made so that a range of appearances and flavors could be eaten without a large intake of food. Modern biscuits, cookies, and crackers are often used as a casual snack food, especially in the USA. Many of these are much larger than traditional biscuits, as they are sometimes eaten to replace meals by consumers who have busy lifestyles.

Classification

Biscuit classification varies according to geographic location. Different and quite specific systems are used in the UK and USA, while in other parts of the world, the distinction between biscuits or cookies and cakes is not as clear; these products are often classified as small confectionery products. For example, biscuits in France are often classified under the title *petits fours sec*. These words literally translate as small oven dry, giving the implication that these are small, dry baked goods. However, petits fours secs may also encompass other products that are outside the biscuit

definition given above, such as cake-, meringue-, and pastry-based products. Biscuits that are too large to be put in the mouth at one time, would be called *les petits gâteaux secs*, meaning small dry cakes. A similar distinction exists in Germany where biscuits are called *kekse* or *konfekt*, but only if they are small.

In the UK, biscuits are classified on their formulation, the two major categories being hard and short doughs. Hard doughs are similar to bread doughs but have a much stiffer consistency. With hard doughs, the three-dimensional gluten network is well developed during the mixing process, so the dough is both elastic and extensible. The formula for hard doughs is very lean, as the content of fat and sugar relative to flour is very low. Crackers are made from hard doughs, as are semisweet, unsweetened, and savory biscuits. Short doughs are more closely related to cake doughs than bread doughs, although they have less water than a standard cake formula. Short doughs are named for the higher content of shortening or fat than hard doughs. They also have a higher sugar content. The increased levels of fat reduce dough extensibility, causing this type of biscuit or cookie to break easily. The flour in short doughs is given very little mixing, so the development of a gluten network is minimized. The consistency of short doughs has been compared with wet sand; they hold together under pressure but crumble easily. Short doughs have a subgroup called soft doughs that have higher levels of fat and sugar, resulting in a dough with an even softer consistency.

Biscuit classification in the USA is based on the method of processing and, in particular, the way the biscuits are shaped. The four main categories are sheeting or cutting (sometimes called cutting machine doughs), rotary molding, wire cutting, and depositing. Sheeting is used for doughs that would be classified as hard doughs in the UK and is always used for making crackers. The dough is passed through a series of rollers until it reaches the required thickness. The biscuit or cracker shapes are then cut out of the dough sheets using a plastic or metal die. It is important that the dough is strong and elastic so that the biscuits and cracker shapes do not deform when excess scrap dough is removed from around the shapes. Rotary molding is used for short doughs, and requires a dough with a relatively stiff consistency that is not too sticky. The dough is compressed into dies mounted on the surface of a roller, and any excess dough is scraped off. The molded dough piece retains its shape as it is pushed out of the die on to the baking sheet or band. Short doughs may also be shaped by wire cutting, which involves extruding the dough through a die and slicing it with a tight wire at appropriate intervals. The tension placed on the dough during extrusion and the thickness of the wire will vary for different types of doughs. Because soft doughs are similar to a batter, having a semifluid consistency and lacking cohesiveness, they are shaped by depositing. The dough is extruded through a nozzle and dropped on to the baking sheet or band. To achieve uniformity in size and shape of the biscuits or cookies, the flow of dough is cut off at regular intervals.

Features of Biscuits

To further explore the nature of biscuits, cookies, and crackers, this section examines the main types of these products in more detail. Products are categorized by the hard and short dough classification with a discussion of the distinct features necessary for the different forming processes. The wide range of shapes, sizes, textures, flavors, and colors of biscuits is illustrated in Figure 1.

Crackers made from Hard Doughs

Crackers are characterized by very low fat and sugar levels. They are often used as a base for a savory topping, such as cheese or tomato, but the modern snack cracker, with its tasty coatings and flavors, may be eaten without additions. Crackers may be further subdivided into those that are fermented, including soda crackers, saltines, and cream crackers, and those that are chemically leavened like the popular snack cracker.

Soda crackers have been popular in the USA for over 150 years. They are typically 4 mm thick and 50×50 mm square. The shortening content is 8–10%. Saltines are a smaller, more dainty type of soda cracker with an increased amount of shortening. Traditional soda cracker manufacture involves a long fermentation using a sponge starter. Soda crackers have a significant amount of sodium bicarbonate added to the dough (1%), which increases the alkalinity and is the reason for the name of the cracker. Once the dough is mature, it is sheeted to about 4 mm and then laminated six to eight times. The cracker is cut by making lines of perforations and baked as a whole sheet, a process that minimizes the amount of waste dough. A feature of soda crackers is the nine hole docking pattern on each cracker, set out in a 3×3 grid pattern. After baking, the sheet of crackers is split along the perforation lines, usually to make a block of four crackers for packaging. The final product is quite flaky but crisp. The spring between the docking holes on the top of the cracker should be even, and the bottom surface should be almost flat with numerous small blisters. Because soda crackers are quite dry and bland, they are not usually eaten

Figure 1 (see color plate 9) Biscuits come in a wide range of shapes, sizes, textures, flavors, and colors.

alone and are often used as an accompaniment to soup.

In the UK, soda crackers are rare, and cream crackers fill a similar market niche. Despite their name, cream crackers contain no cream! They have a slightly higher fat content (12–18%) than soda crackers. Cream crackers are generally a 65×75 mm rectangular shape and are slightly thicker than soda crackers with around 6.5 mm between biscuits stacked in a column. Unlike soda crackers, they are usually produced as individual units. While, traditionally, the long sponge and dough fermentation process was used for manufacturing cream crackers, modern techniques involve a single-stage mixing and fermentation process, which takes from 4 to 16 h. As with soda crackers, cream crackers are sheeted and laminated. A feature of cream crackers is the laminating 'dust', consisting of flour, shortening, and salt, that is applied between the layers of dough. This causes the laminations to lift apart during baking, giving an extra flaky structure. The movement of the laminations and subsequent surface blistering are quite irregular, giving the cream cracker its characteristic uneven surface. Blisters are present on both the top and bottom surfaces. Cream crackers have a final moisture content of 3–4%, which is quite high for a cracker, and along with the increased fat content, the cracker is relatively soft, will not crumble and should 'melt in the mouth'. Without chemical leavening, cream crackers are bland but have a slightly nutty flavor. They are mostly eaten with a savory topping and are often buttered.

Snack or savory crackers have a more recent history than soda and cream crackers. They have two distinguishing characteristics. The first is that the cracker is sprayed with hot oil as it leaves the oven, and the second is that a topping is applied to the crackers to add flavor. Snack crackers usually contain some sugar (4–10%), which also adds flavor and texture. They are usually chemically leavened, but some snack crackers are made from fermented doughs. Because snack crackers do not have the long fermentation to mature the dough, proteolytic enzymes or sulfites are used to relax the doughs so that the crackers do not deform during sheeting and cutting. Snack crackers come in a wide range of shapes and sizes, but are often round and have docking holes to allow an even lift during baking. Only snack crackers produced by fermentation are laminated. Toppings are generally applied before baking, and include herb, cheese, salt, chicken, and smoke flavors. Sometimes, the crackers are decorated with small seeds, such as poppy, sesame, or celery. Snack crackers have a dense texture and are quite soft. The hot oil spray improves the mouth feel and gives an attractive appearance to the finished product.

Sweet and Semisweet Biscuits made from Hard Doughs

Sweet and semisweet biscuits made from hard dough are generally more popular in the UK than the USA. While the gluten network is relatively well developed by mixing, the higher amounts of sugar (around 20% of flour) and fat (16–20%) than in cracker doughs

make the gluten less elastic and more extensible. Chemicals, such as sodium metabisulfite or other chemical derivatives of sulfur dioxide, can also be used to condition and relax the dough to facilitate processing. Unlike cracker doughs, most hard sweet and semisweet biscuit doughs are chemically leavened. After mixing, the dough is sheeted and formed into shapes. The individual biscuits are cut out of the dough sheet, leaving a web of scrap dough to be removed and incorporated back into the main batch of dough. These biscuits are generally docked and marked with a name or pattern before baking. Sometimes, a milk or egg/milk wash is applied to enhance their appearance after baking, and occasionally, a garnish of sugar or other granular material is applied.

Baked biscuits should have a smooth, even surface with a pale color. The texture of the biscuit is open and even, giving a delicate bite, although this is somewhat dependent on the formula – lower levels of sugar result in a harder bite. The ingredients in sweet and semisweet biscuits are quite plain, so the flavor is usually a mild vanilla, caramel, or buttery flavor. The biscuits are generally served without accompanying food and are eaten with tea or coffee. There are some interesting variations of sweet and semisweet biscuits, such as those that are processed after baking to incorporate a cream sandwich or chocolate coating. The garibaldi fruit sandwich biscuits have a layer of currant or small sultana filling between two layers of hard dough.

Biscuits or Cookies made from Short Doughs

The vast majority of biscuits and cookies consumed world-wide are made from short doughs, and consequently the range of shapes, sizes, flavors, and ingredients is huge. Formulae are correspondingly variable, but there are some consistent requirements. Flour is usually weak, with less than 9.5% protein. While there are no rules for the proportions of fat and sugar, which can range up to 100 and 200% of flour weight, respectively, the quality of these ingredients is important since they make up such a large proportion of the dough. Short doughs are usually mixed in a two-stage process with an initial creaming of the fat and sugar, although modern techniques tend to use the 'all-in' mixing method. The doughs are cohesive and plastic but lack extensibility and elasticity. The consistency of the dough will vary according to the requirements of the machinery used to form and shape the biscuits.

As described above, the two main processes for forming short dough biscuits and cookies are rotary molding and wire-cutting. Rotary molding originated from the simple wooden molds often used in monastery kitchens to produce biscuits and cookies containing inscriptions. With a rotary mold, dough is continually fed from a hopper and forced into a metal die on a rotating roller. The formed dough shape is pressed out of the mold on to the baking band as the roller rotates. Unlike hard doughs, which tend to shrink during baking, short and soft doughs generally spread because of the high sugar and fat content. This is a particular disadvantage of rotary molding as the inscription can become blurred. Tight control of the formula is necessary to reduce spread. Rotary cookies should be thin and smooth with no surface cracks or irregularities. Because of their regular size and shape, rotary cookies may be used to make cream sandwiches in the same way as hard sweet biscuits.

Wire-cutting is a form of extrusion; the dough passes through a die and is sliced at intervals by a tight wire so that the formed dough shape drops on to the baking band. Doughs intended for wire-cutting are usually softer than rotary molding doughs, and more chunky ingredients, such as chocolate chips, nuts, or raisins, may be incorporated. A variation on the wire-cutting process is a rout press that extrudes dough continuously. The dies on a rout press are designed to produce strips, which are cut into short lengths before baking. Both of these techniques can be adapted to coextrude by having two different doughs or a dough and a filling coming from separate hoppers into a single die that forms a dual layer cookie (e.g., chocolate/vanilla) or a dough tube often filled with a fruit paste of similar consistency.

Biscuits or Cookies made from Soft Doughs

Soft doughs have a pourable consistency, are typically rich in fat (65–76% of flour weight), and may be based on whipped egg whites (15–25%). Sugar is around 35–40% of the flour weight. Weak flour is used, and mixing is in a two-stage process. The flour and other dry ingredients are added last, and only minimal mixing takes place to prevent the dough from becoming tough. Often, rich, expensive ingredients, such as ground almonds, coconut flour, or cocoa, are used in soft doughs. However, coarse particles are avoided, because they may block the nozzles during depositing. Dough temperatures are important to achieve the correct consistency for forming the specific type of biscuit or cookie required. Temperatures generally range between 10 and 17 °C. The dough flows from the hopper through a nozzle on to the baking band. The nozzles may be of different shapes and sizes to alter the appearance of the cookie, but because the dough flows a little after depositing, these designs may be rather irregular. Some depositor heads can be rotated to make swirls and circular shapes, while two or more depositors may be

synchronized to combine doughs of different flavors or colors within the same biscuit or cookie. Soft dough biscuits and cookies have a soft, delicate texture and a 'melt-in-the-mouth' feel. However, these properties make them fragile and subject to breakage, and packaging can also be difficult because of the irregular shapes.

Finishing and Decorating Biscuits

While the basic biscuit, cookie, or cracker is distinguished by the processes described so far, the nature of the final product can be significantly altered by secondary processing after baking. A wide range of processes are used to finish and decorate biscuits, cookies, or crackers, some of which have already been referred to. Two of the most popular are cream sandwiches and chocolate coating or enrobing. In a cream sandwich, the cream usually makes up about 30% of the final biscuit weight and consists of sugar, fat, and flavorings, such as fruit acids, cocoa powder, and skimmed milk powder. The cream is either poured into a stencil positioned over the top of the biscuit base or deposited directly on to the base. Working with chocolate is technically difficult because of the necessity to control the temperature accurately. The properties of chocolate are quite variable and depend on the ingredients that were used to make the chocolate. Various methods of fully enrobing or partially dipping the biscuit or cookie into a chocolate bath are used, depending on product specifications. Icings, made from icing sugar, water, and sometimes fat or a gelling agent such as gelatine, are another popular finishing on biscuits. Jellies, jams, marshmallow, caramel, fruit, seeds, and nuts may also be used to enhance the taste and texture of biscuits, cookies or crackers.

See also: **Biscuits, Cookies, and Crackers**: Methods of Manufacture; Chemistry of Biscuit Making; Wafers

Further Reading

Anon (1979) *The Cookie and Cracker Manufacturing Correspondence Course. Unit IV Crackers*. Washington, DC: The Biscuit and Cracker Manufacturers' Association.

Anon (1999a) *Biscuit Production. Baking Distance Learning Course*. Christchurch, New Zealand: New Zealand Institute for Crop & Food Research.

Anon (1999b) *The Science of Baking Correspondence Course. Saltines/Snack Crackers*. Manhattan, KS: The American Institute of Baking.

France WJ and Small M (1986) *The New International Confectioner*, 5th edn. London: Virtue & Company.

Kulp K (1994) *Cookie Chemistry and Technology*. Manhattan, KS: The American Institute of Baking.

Manley D (2000) *Technology of Biscuits, Crackers and Cookies*, 3rd edn. Cambridge, UK: Woodhead.

Matz SA (1992) *Cookie and Cracker Technology*, 3rd edn. McAllen TX: Pan-Tech International.

Smith WH (1972) *Biscuits, Crackers and Cookies*, vols. 1 and 2. London: Applied Science.

Wade P (1988) *The Principles of the Craft. Biscuits, Cookies and Crackers*, vol. 1. London: Elsevier.

Methods of Manufacture

L Doescher Miller, Formerly of The Pillsbury Company, Minneapolis, MN, USA

Background

The manufacturing process used to produce all biscuits, cookies, and crackers consists of a mixing step, a shaping or forming step, and a baking step. The mixing and baking steps are common to the manufacture of all types of these products. What is distinct for the products are the shaping or forming steps. In this article, the processes used to produce these products will be described.

Mixing

Mixing is commonly defined as a process designed to blend separate materials into a uniform, homogeneous mixture. In the context of cookie and cracker doughs, the term takes on a broader meaning in that it also applies to the development of gluten from hydrated flour proteins, the aeration of a mass to give a lower density, and the dispersion of solids in liquids. One or more of the functions is required for the formation of cookie and cracker doughs. These processes are accomplished with three principal types of mixers: vertical spindle mixers, horizontal drum mixers, and continuous mixers. (*See* **Bread**: Dough Mixing and Testing Operations.)

Vertical Spindle Mixers

Vertical mixers were used extensively before the development of the high-speed mixer. Although usage has decreased overall, this type of mixer is still used for crackers and semisweet doughs, for which the vertical spindle design is particularly well suited. The feature that all vertical mixers have in common is the portable dough trough, which serves as the mixing vessel. The trough is usually a wheel-mounted, heavy tub with vertical sides, round ends, and a flat bottom. The trough is designed so that it may be

wheeled into position below the mixer head and locked into place. Mounted on an overhead frame are the drive mechanism and spindles with horizontal paddles or arms. Most mixers of this type are equipped with two or three spindles. The spindles are lowered into the trough and move in either a planetary or a stationary, circular motion when activated. The shape of the arms forces the dough upwards and downwards in the trough, generating a mixing action. The mixer blades are designed to provide a cutting action rather than kneading or stretching the dough.

The advantage of vertical spindle mixers is that they can be used for almost any product. They are very well suited for products such as soda cracker doughs. These doughs require a two-stage mixing sequence separated by a fermentation period. The sponges and doughs do not have to be transferred in and out of the trough at each mixing stage. Instead, they remain in the same trough for the entire mixing/fermentation sequence. Because there is almost no heat generated during the mixing process, this mixer is ideal for doughs that must remain cool. The mixing action is very gentle. Doughs containing ingredients or particulates that are easily damaged may be safely mixed with this mixer type.

The primary disadvantage of vertical mixers is their slow operating speed. The spindles rotate over a limited range at two or three speeds, up to a maximum of approximately 20 rpm. For doughs requiring the development of a gluten network, the mixing time required may be as long as 90 min. Because this is slow relative to the speed of the remainder of the processing line, several mixers may be needed to maintain an uninterrupted dough supply. Other disadvantages include lack of uniformity of the mix, and the labor-intensive nature of the system's design. The latter disadvantage relates to the fact that the dough troughs are very heavy, requiring special equipment to move, lift, and tilt the tubs at each stage of the process.

Horizontal Mixers

The feature that all mixers of this type have in common is a horizontal bowl mounted on a rigid frame that encloses the drive motor. The bowl may be stationary with a vertical front wall that slides down so that the dough may be ejected. More commonly, the mixer bowl is designed to tilt, and the dough is ejected from the tilted position. In both designs, the top cover is fixed so that mixing takes place in an enclosed space.

The blades of horizontal mixers are mounted on a horizontal shaft and may have any of several different shapes, depending on whether a cutting, scraping, or kneading action is needed. Mixing times for cookie and cracker doughs range from 5 min for soft cookie doughs to 30 min for the hard doughs, which require gluten development. Most mixers of this type are fitted with a jacket surrounding the mixing chamber. This allows a coolant to be circulated, so that dough temperatures can be controlled. Without such control, horizontal mixers can increase the dough temperatures to the point where dough handling and finished product properties are adversely affected.

The advantages of horizontal mixers are their high speed, ability to supply dough to a processing line continuously, uniformity of the mixes they produce, and their potential for complete automation. Horizontal mixers may be operated at speeds of 15–80 rpm. Commonly, their power train is equipped for two-speed rotation. Unlike most vertical spindle mixers, ingredients may be added with the blades in motion, and the discharge of the dough from the mixer is simple. Accurate control of dough temperature is possible, because there is a continuously circulating refrigerant.

Horizontal mixers require that all ingredient charging spouts be located overhead or that ingredients be added manually. Charging the mixer with ingredients is typically a significant portion of the total mixing cycle time. During mixing, the blades may throw material to the top of the chamber, so that it is never fully incorporated. With the shaft and mixing blades located in the center of the bowl, dough discharge is not always rapid and may occur in several large fragments. Only one operation (charging, mixing or discharging) may be performed at any one time, so that fast cycling of the dough batches may not be possible.

Horizontal mixers usually have capacities of up to 550 kg. The weight and the vibrations generated by their relatively high operating speeds place special design demands on the production facility. In addition to a reinforced floor, special mounts are necessary to secure the mixer in place.

Continuous Mixers

Continuous mixers are best described as a rotor or screw operating within a barrel jacketed for temperature control. The ingredients are fed continuously, either from one end of the barrel or in successive ports at intervals along its length. The mixing action may be altered from gentle blending and dispersing to vigorous or high-intensity kneading by varying the arrangement of different mixing arms along the length of the barrel. The amount of work put into the dough during its transit may be controlled additionally by restricting the dimension of the outlet orifice, creating a back pressure inside the barrel.

Continuous mixers are favored in some plants because they are capable of providing a constant supply of dough to production line, all of uniform age. Continuous mixers are small in size relative to horizontal mixers and are suitable for complete automation. In spite of this advantage, continuous mixers are not common in cookie or cracker plants, where a single line may be required to produce a variety of products. They are usually used only on high-output lines having a single purpose or similar product types.

The primary disadvantages of this mixer type are its high cost and the additional cost of the associated automated, continuous ingredient feeding systems. Beyond this, continuous mix systems are not particularly flexible; different products usually require completely different types of machines. In addition, the initial process set-up can be difficult, requiring considerable experimentation to determine optimum mixing conditions and sequences. Finally, starting and stopping the process are difficult in the event of any problem along the rest of the production line.

The Forming Process

While the same mixing and baking process may be used for many types of cookies and crackers, the forming step is specific to each product type. There are three processes used to form cookie and cracker doughs: (1) cutting or stamping from a continuous sheet of dough, (2) rotary molding by shaping dough in die cavities cut into the surface of a metal cylinder, and (3) extruding dough through a shaped die.

For each of these methods, the rheology of the dough is different and is designed to be compatible with the process. In general, doughs that are to be sheeted possess a significant gluten network as a result of mixing, and are both elastic and extensible. Those destined for rotary molding lack gluten development and are best described as cohesive. Doughs intended for extrusion are soft, frequently high in shortening, and spread while baking.

Sheeting and Cutting

The most common and versatile method to form cookie and cracker doughs is sheeting and cutting. This method involves the production of a thick sheet of dough, evenly reducing the thickness of the sheet, cutting out the desired shapes, and returning the scrap dough to be reincorporated either in the mixer or early in the sheeting process. This method is used for the production of crackers, semisweet biscuits, and selected soft doughs.

After mixing, the dough is fed into a hopper, below which lie the sheeting rollers. There are typically three rollers below the hopper arranged in a triangular fashion (Figure 1). At least one of the top two rollers, known as forcing rollers (labeled A in the figure), is grooved so that a positive feed is provided to the gauge or gauging roller. The gauging roller (B), which is always smooth, serves to deliver the dough to the conveyor belt (C). The purpose of the sheeting unit is to compact the mass from dough hopper uniformly and provide a sheet of even thickness having the width of the processing line.

The relatively thick dough slab from the sheeter then passes through a series of reduction or gauge rollers (D). These are smooth steel rollers used to reduce the dough sheet to the thickness that is desired before cutting of the finished dough piece. The gauge rollers occur in pairs mounted vertically. For products having sticky or adherent doughs, it may be necessary to mount a scraper blade against one or both of the rollers to release the sheet of dough. On most process lines, there are two or three pairs of rollers. This ensures that the thickness is reduced no more than 50% at any one rolling operation.

Some doughs, such as those of saltines and cream crackers, are laminated before cutting. The lamination occurs by lapping the dough back upon itself in the process direction. At the lapper, the take-away conveyor lies at a 90° angle relative to the line delivering the dough. The number of layers is controlled by

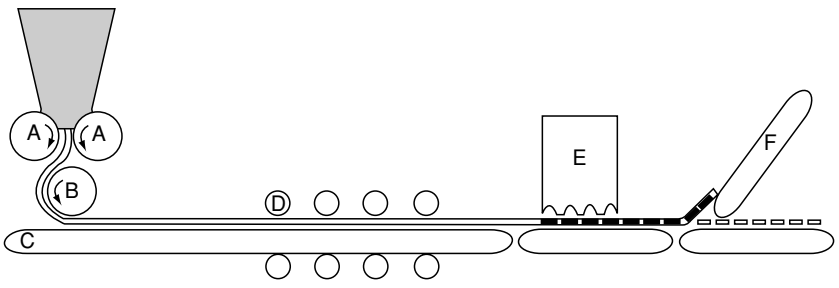

Figure 1 Sheeting and cutting process. See text for explanation. Reproduced from Biscuits, Cookies and Crackers: Methods of Manufacture, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

the relative rate of the lapper and take-away conveyor. The lapped dough then passes through several more sets of gauging rollers to bring the dough sheet to the desired thickness prior to cutting.

The repeated working of the dough in one direction results in the accumulation of stress. If the dough were cut at this point, the resulting pieces would shrink to relieve the stress, and misshapen or distorted products would result. Therefore, it is normal to relax the dough after reduction and before cutting. The relaxation is accomplished by transferring the dough to a conveyor, still moving in the same direction, but at a slower speed.

Once the dough has been relaxed, it passes on to the cutting operation. Two different types of cutting methods exist: reciprocating cutters and rotary cutters. The reciprocating cutters are heavy block cutters that stamp out one or more pieces at a time. The cutter head (E) may have a dual action, whereby the cutter drops first, followed by a docking head or an embossing plate. The equipment operates via a swinging mechanism so that the dough sheet moves at a constant speed, the cutter drops and moves with the dough, and then rises and swings back to the original position. The second type of cutter, the rotary cutter, consists of a rotating metal cylinder. On the face of the roll are formed the desired shapes with a sharp metal edge. As the cutter rotates with the dough conveyor, the metal edges cut into the dough sheet to form the product. The product pieces are then conveyed into the oven.

As a result of either cutting process, 20–60% of the dough sheet remains as scrap. The scrap dough (F) is lifted way from the cut dough pieces and returned either to the mixer or to the sheeter. Return to the mixer permits uniform incorporation of the scrap into the dough mass. However, most systems route the scrap back into the sheeter either along the full length of the hopper or at the back side of the hopper. If dough is incorporated behind the new dough, imperfections will be on the bottom side of the dough sheet and will not be visible on the finished product.

Rotary Molding

The principle of rotary molding is illustrated in Figure 2. Three rollers are placed in a triangular arrangement below a dough hopper. A roller, called the forcing or feed roller (A), has deep grooves designed to pull dough down from the hopper. The dough is forced into the cavities of the engraved roller (B) by the forcing roller. A scraper blade (C) is mounted against the engraved roller to remove any excess dough and return it to the hopper via the forcing roller. Beneath the engraved roller is a rubber-covered extraction roller (D) that serves to drive the take-away belt (E). The extraction roller applies pressure to the engraved roller via the belt, causing the dough to adhere preferentially to the conveyor belt. Dough pieces are dropped from the take-away belt into pans or directly on to the baking belt (F).

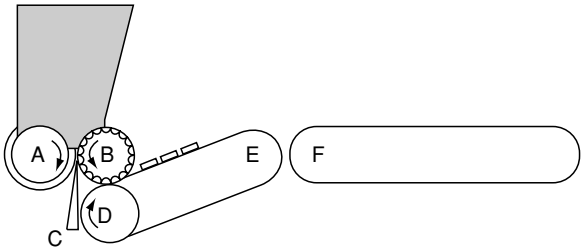

Figure 2 Rotary molding process. See text for explanation. Reproduced from Biscuits, Cookies and Crackers: Methods of Manufacture, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

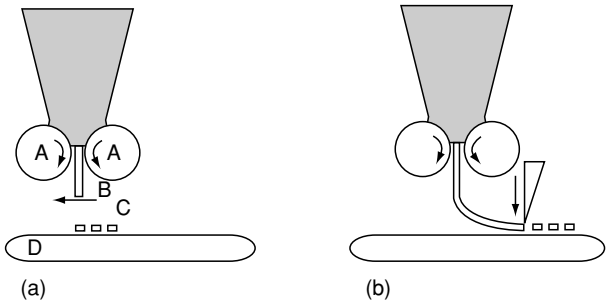

Figure 3 Extrusion process: (a) wire-cut machine; (b) bar/rout-press. See text for explanation. Reproduced from Biscuits, Cookies and Crackers: Methods of Manufacture, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

The rotary molding process is suitable only for dry, crumbly doughs. This process offers advantages over sheeting and cutting in that there is no scrap to recycle, and there are very low labor requirements to run the process.

Extrusion

There are two types of devices used in the production of extruded cookies: wire-cut machines and bar/rout-presses. Both systems are very similar in design (Figure 3). A hopper is placed over a system of two or three rollers (A) that force dough into a pressure chamber (B). The rollers may run continuously or intermittently to force dough out of the pressure chamber at the die. For wire-cut cookies, the dough is extruded through a row of dies, and a wire or blade (C) mounted on a frame moves through the dough just below the die nozzle outlet. The cut dough pieces

then drop into a conveyor band (D) for transport to the oven. The wire usually moves only in one direction through the dough, opposite that of the conveyor. The wire-cut machines operate at rates of up to 100 strokes per minute. Difficulties encountered with this type of production are distortion of the extruded dough piece during cutting, and inconsistent placement or drop of the cut piece on to the conveyor. (*See* **Extrusion Cooking**: Principles and Practice.)

The design of the bar- or rout-press is very similar to that of the wire-cut machine. The hopper rollers and pressure chamber are essentially identical to their wire-cut counterparts. Unlike the wire-cut machine, the base of the pressure chamber has a die plate that is inclined in the direction of the extrusion. A continuous ribbon of dough is extruded from a nozzle, which is shaped to impart the desired finished product design. The dough ribbon can be cut into individual pieces by a vertically operating guillotine before the oven or after baking. If the product can be baked as a continuous ribbon the dough is extruded directly on to the oven band; otherwise, it is extruded and cut on to a conveyor belt.

Baking

The cookie and cracker industry relies almost exclusively on band or traveling ovens to bake its products. The band oven is essentially an insulated, heated tunnel equipped with a continuous conveyor. The ovens vary both in length (from 30 to 150 m) and in band width (from 1.0 to 1.5 m). More modern ovens frequently consist of a series of modular units or zones. Each of the zones is equipped with its own set of controls so that the temperature and air flow may be controlled within that zone. The oven band is typically continuous, passing on to a drive drum at the end of the oven and returning underneath the baking chamber to a tension drum at the feed or input end of the oven. The chamber through which the oven belt returns may or may not be enclosed. Frequently, the oven band serves as the baking surface for the product. Depending on the product type, the oven band may be solid or any of a variety of open wire-mesh types. The choice of mesh is a critical factor in the process, as it affects the heat transfer at the bottom of the product. This, in turn, can have a marked effect on the quality of the finished product.

There are three basic types of ovens: direct fired, indirect fired, and fully indirect fired. Ovens are usually heated by the combustion of gas, although there are a few manufacturers who use oil or electricity for economic reasons. The most common type is the direct-fired oven, in which gas is burned inside the baking chamber itself. In these ovens, the burners are placed across the width of the oven at regular intervals, both above and below the oven band. In other oven types, termed 'indirect ovens,' the gas or oil is burned outside the baking chamber, and the heated combustion gases are circulated into and throughout the baking chamber. Indirect-fired ovens typically have a single burner for each section. The hot gases from the burner pass along pipes parallel to the length of the oven, both above and below the oven band. The products of combustion are circulated throughout the baking chamber by large fans. Fully indirect ovens are those in which the heat source is independent from the baking chamber and heat transfer occurs via a heat exchanger. None of the products of combustion circulate inside the baking chamber. This type of oven is not common, except when oil is used as a combustible material. If circulated, the products of this type of combustion would impart an unacceptable flavor to the products.

Cooling

Products hot from the oven must be cooled prior to packaging for several reasons: the products may not be firm enough to withstand the packaging process while warm, the packaging material may shrink around a warm product, or the quality of the products would deteriorate if palletized while warm because the cooling rate across the pallet would be quite slow.

The normal method of cooling products is to place them on an open conveyor and transfer them along a distance 1.5–2 times the length of the oven. The products cool naturally in the ambient factory atmosphere. In a few cases, it is necessary to provide forced air to aid the cooling process.

See also: **Bread**: Dough Mixing and Testing Operations; **Extrusion Cooking**: Principles and Practice

Further Reading

Almond N (1989) *The Biscuit Making Process. Biscuits, Cookies and Crackers*, vol. 2. London: Elsevier.

Manley D (1991) *Technology of Biscuits, Crackers and Cookies*, 2nd edn. Chichester, UK: Ellis Horwood.

Matz SA (1978) *Cookie and Cracker Technology*, 2nd edn. Westport, CI: AVI.

Matz SA (1988) *Equipment for Bakers*. McAllen, TX: Pan-Tech International.

Pyler EJ (1988) *Baking Science and Technology*, 3rd edn, vol. II. Merriam, KS: Sosland.

Smith WJ (1972) *Technology, Production and Management. Biscuits, Crackers and Cookies*, vol. I. New York: Magazines for Industry.

Wade P (1988) *The Principles of the Craft. Biscuits, Cookies and Crackers*, vol. I. London: Elsevier.

Chemistry of Biscuit Making

J L Hazelton, J L DesRochers and C E Walker,
Kansas State University, Manhattan, KS, USA

Introduction

What is a biscuit? Depending on the part of the world in which you live, the answer will vary. In the USA it would be described as a chemically leavened bread-type item similar to a scone in the UK. However, in the UK a biscuit is equivalent to a cookie or cracker in the USA. The term 'biscuit' will be used here to include biscuits, cookies, and crackers.

Most biscuits differ from other baked cereal products by having a very low moisture content. Finished breads and cakes typically have moisture contents of 35–40% and 15–30% respectively, whereas biscuits usually contain only 1–5% moisture. Wheat flour, sugar, and fat are the basic ingredients, but thereafter the variety is almost endless. Biscuits can be grouped in many ways, based on their texture and hardness, their change in outline during shaping and baking, the extensibility or other characteristics of the dough, or the ways that the doughs are handled prior to biscuit formation. The groups often overlap, so it is important to see how several types of biscuits are related, based on their proportions of fat, sugar, and water, and then to compare their typical characteristics and processing means.

These differences in formulation, processing, and finished product attributes are all a function of the dough consistency or dough rheology. Certain key ingredients, such as flour, fat, and sugar largely determine the dough rheology and thus what type of forming equipment can be used to produce the biscuit. Based on these fundamental distinctions, five basic categories can be devised: (1) wire-cut; (2) rotary-molded; (3) bar-shaped; (4) deposited; and (5) cut or stamped. Regardless of the category, there are certain rheological requirements for all biscuits, namely the dough must be adequately cohesive for molding/forming, without excessive stickiness, and the dough must have a short, cuttable texture.

Likely the single greatest influence on (biscuit) dough rheology is the degree of gluten development within a dough. Biscuit doughs generally fall into two broad categories – hard and soft doughs. The viscoelastic properties of hard doughs represent the presence of a protein (gluten) matrix which is developed during mixing and sheeting. Hard doughs are stiff, tight doughs that require extensive mixing (work) with a resulting increase in dough temperature. They are similar to bread doughs, except that the sugar and fat contents modify their viscoelastic properties. Hard doughs are usually laminated and sheeted before cutting or stamping. The formed pieces will generally shrink because of the elastic quality of the gluten. During baking, the biscuits may continue to shrink in outline, but become thicker. This type of dough formulation may also be suited for rotary-molded biscuits, due to its firm consistency. (*See* **Bread**: Chemistry of Baking.)

Soft doughs do not have a formed gluten structure, because of their high levels of shortening and sugar, and are generally mealy or sandy in texture. They are usually formed by compressing into dies (rotary-molded) or by extruding and cutting, but some types can be sheeted, then cut. Dough pieces formed from soft doughs tend to retain their shape until baking, but then they spread or flow, becoming thinner.

Deposit biscuits are the machine-made counterpart of the hand-bagged or 'drop' version. Such formulae have been successfully adapted to automated production. Deposit biscuits contain about 25–40% sugar, 65–75% shortening, and 15–25% eggs and possess a spread factor of 79–80 (percentages based on 100 parts flour). Formulations should be such that excessive spread does not occur and the top design of the biscuit is preserved during baking. Adequate adhesive characteristics of the dough are also needed, so that it will adhere to the band and separate from the main tube of dough when deposited (**Figure 1**). This type of formula may also apply to wire-cut biscuits, where the dough is forced through a die and cut into disks by a reciprocating wire.

Although there are many similarities between biscuit and cracker formulations, a few key differences should be pointed out. One is the obvious difference in sugar content. Unlike biscuits, crackers are usually made from laminated dough, where thin sheets of dough are alternately layered with fat. Whereas biscuits are chemically leavened, some crackers, i.e. saltines, are yeast-fermented over long time periods to develop the characteristic flavors. The flour used in crackers is commonly stronger than biscuit flour and often dough relaxers, e.g., proteases, are required to increase the extensibility of the cracker dough. Salt further strengthens dough, increasing its resistance to extension. A series of biscuit and cracker formulae is shown in **Tables 1** and **2**.

Dough Composition

Wheat Flour

Wheat flour is unique among the cereal grain flours in that, when mixed with water, its protein components form an elastic network capable of holding gas and developing a firm spongy structure during baking. The protein substances contributing these properties (gliadin and glutenin), when combined with water and mixed, are known collectively as gluten. The suitability of a flour for biscuit making is generally determined by its gluten. Gluten characteristics are determined by genetics, the wheat's growing conditions, and the milling process. (*See* **Flour**: Dietary Importance.)

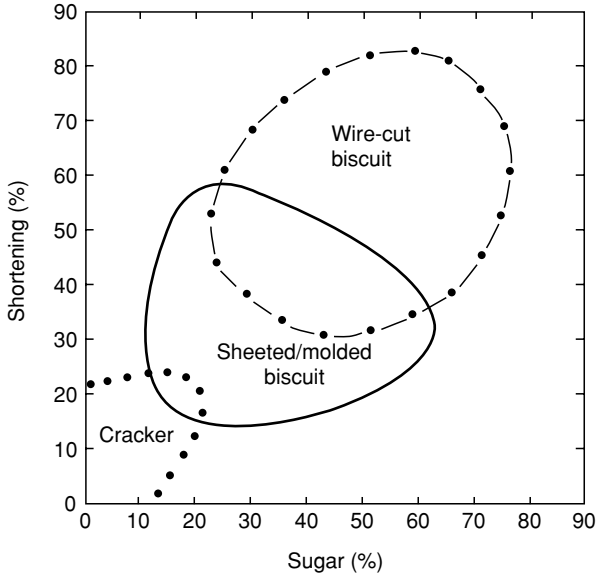

Figure 1 Biscuit composition in relation to sugar and shortening, based on 100 parts flour. (Each dough is processed according to its consistency or water content.)

Wheats are described by millers as hard, medium, or soft, based on the grain's physical characteristics. Hard types tend to have higher protein quantity and quality, possessing a vitreous endosperm, with starch granules tightly packed in a protein matrix. During milling, some starch grains are damaged, resulting in an increased surface area that leads to higher water absorption. The softer wheats have a less compact starch–protein complex which results in less starch damage and lower water absorption. The protein level of soft wheats is usually lower, producing a less resistant, more extensible dough. These 'weaker' flours are traditionally deemed more suitable for biscuit making than the harder flours used for breads. (*See* **Flour**: Roller Milling Operations; Analysis of Wheat Flours.)

Flour protein levels needed for biscuit making typically range from 7 to 10% and are often selected by functionality for end use, price, and availability. Flour should contain no more than about 14% moisture. Biscuit flour is typically left unbleached and unchlorinated.

Sweeteners

All biscuit formulae contain sweeteners, which constitute the bulk of dissolved materials in most doughs. Sugars impart sweetness, act as vehicles for other

Table 1 Representative formulae for various sweet biscuits

Ingredient	*Animal cracker (molded) %*	*Sugar cookie (cut/stamped) %*	*Chocolate chip (wire-cut) %*
Soft flour	55.72	45.00	22.00
Shortening	11.36	11.20	19.40
Granular sugar	11.00	27.10	29.70
Water	7.70	9.80	NA
Sucrose syrup	7.33	NA	NA
Honey	2.35	NA	NA
Arrowroot flour	1.83	NA	NA
Molasses	0.88	2.70	NA
Salt	0.88	0.40	0.70
Monopotassium tartarate	0.37	NA	NA
Sodium bicarbonate	0.37	0.40	0.10
Lecithin	0.22	NA	NA
NFDM	NA	1.30	NA
MCP	NA	0.30	NA
Liquid whole eggs	NA	1.80	9.10
Vanilla flavor	NA	NA	0.20
Chocolate chips	NA	NA	18.70

NFDM, non-fat dry milk; MCP, monocalcium phosphate; NA, not applicable.

Table 2 Representative formulae for various cracker types

Ingredient	*Snack cracker %*	*Soda (saltine) %*	*Graham %*
Strong flour	66.37	69.60	25.40
Soft flour	NA	NA	25.40
Graham flour	NA	NA	12.60
Water	18.10	21.58	16.40
Granular sugar	3.85	NA	7.62
Shortening	3.85	6.61	5.08
Ammonia bicarbonate	2.05	NA	0.38
Corn syrup	1.90	NA	NA
Malt syrup	1.28	0.64	0.44
Brown sugar or honey	NA	NA	4.64
Meal	0.96	NA	NA
Sodium bicarbonate	0.71	0.44	0.63
Salt	0.60	0.97	0.89
Monocalcium phosphate	0.30	NA	0.25
Protease	0.03	NA	NA
Lecithin	NA	NA	0.25
Optional seasonings	As desired	NA	0.02
Yeast	NA	0.16	NA

NA, not applicable.

flavors, and create an attractive finish. Sweeteners also increase tenderness, crust color, volume, and moisture retention, while maintaining the proper balance between liquids and solids responsible for product contour/shape. Machining properties and baking characteristics are also closely related to sugar. Sweeteners tenderize the finished product by interfering with gluten hydration and starch gelatinization. (*See* **Sweeteners**: Intensive.)

High-fructose corn (maize) syrups are becoming increasingly important, but sucrose from sugar cane or beet is still the major sweetener. Commercial sugars are often categorized as granulated or powdered. Granulated sugars range from coating sugar (extremely fine grain) to coarse. Powdered sugars are made by grinding granulated sugars and screening through fine bolting cloths. Generally, as the size of the sugar crystal increases, the size and symmetry of the biscuit decrease, while the thickness and color increase. Often anticaking agents, e.g., corn starch, are added to insure proper flow of the sugar during raw material handling. Where pumpability may offer a processing benefit, granular sucrose can be combined with water to form 'liquid sugar,' usually at 67% sugar solids or 67°Brix.

Sucrose may be hydrolyzed (inverted) to glucose (dextrose) and fructose (laevulose) by heating it in the presence of a dilute weak acid or mixing it with invertase enzyme. Sucrose is the standard for sweetness, with an arbitrary rating of 100. Fructose and glusose are rated at 170 and 74, respectively. Their combined sweetness in a completely inverted sugar is 127, sweeter than the starting sucrose. Inverted sugars are used primarily for their hygroscopicity and browning reactions, which contribute moisture retention and color development in the biscuits. (*See* **Carbohydrates**: Sensory Properties.)

Shortenings and Emulsifiers

Fats are the third major component used in biscuit making, but are considerably more expensive than flour or sugar. Besides being used in the doughs, fats or oils are used as surface sprays, in cream fillings and coatings (such as chocolate), and as release agents. In dough, they tenderize (impart shortness to) the crumb by being dispersed in films and globules during mixing, which interferes with gluten development. Shortening also aids dough aeration during the creaming step. The overall effect improves palatability, extends shelf-life, improves flavor and, of course, adds caloric energy. (*See* **Fats**: Uses in the Food Industry.)

Animal fats, primarily lard, were originally used by bakers. Compound (part animal and part vegetable source) shortenings and all-vegetable shortenings were then developed. Soya bean, cottonseed, palm, coconut, and peanut oils are the primary vegetable sources used in shortening production. Continued advancements in purification and hydrogenation developed vegetable oils that could replace animal fats with equal or better flavors, melting points, consistency, and availability. Because of current health concerns, most bakeries have switched to fats of plant origin. (*See* **Vegetable Oils**: Types and Properties.)

Surfactants (surface-active agents) are given many names by bakers: crumb softeners, emulsifiers, antistaling agents, or dough conditioners. Examples include lecithin, mono- and diglycerides, diacetyl

tartaric acid esters of fatty acids, polysorbate 60 and sodium stearloyl 2-lactylate. Surfactants at low concentrations act to modify the surface behaviors of liquids. They are believed to complex with the protein–starch structure, thereby strengthening the film, and to delay dough setting during baking. The behavior of surfactants is due to their amphoteric (possessing both hydrophilic and hydrophobic molecular regions) properties. Their behavior varies according to the charges on the molecules, their solubility, the hydrophilic–lipophilic balance, and the type of functional groups involved. (*See* **Emulsifiers**: Uses in Processed Foods.)

Surfactants modify dough consistency and reduce stickiness by reacting with the gluten. The greasiness of biscuits with high fat content is also reduced by surfactants. Crumb softeners also complex with the starch molecules to delay retrogradation and texture staling. The grain pattern and volume of the finished product are often improved, as surfactants increase dough gas-retaining properties.

Antioxidants retard the development of oxidative rancidity during product storage. All fats are subject to oxidative or hydrolytic rancidity which causes objectionable odors and flavors, but antioxidants delay these reactions from occurring within the biscuits' shelf-life. They are usually added to bulk shortenings and are important for preserving low-moisture products, which are expected to remain edible for several months. (*See* **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants.)

Is Water an Ingredient?

Water is often thought of as a processing aid or catalyst, rather than as an ingredient. It is incorporated at the dough stage but driven off during baking. Water functions in several ways, including hydrating flour proteins and starch, dissolving sugars, salts, and various leavening chemicals, aiding in ingredient distribution and helping control dough temperature.

A dough's consistency is directly related to its water content, or absorption. Many factors affect dough absorption. Approximately 46% of flour's total absorption is associated with the starch, 31% with protein, and 23% with the pentosans. Acceptable consistency can be obtained only after sufficient water is present to hydrate the flour. This is regarded as bound water and controls the dough's consistency. As bound water layers are 'stacked up,' some of the water is held less and less strongly, resulting in water that can escape (evaporation) and/or migrate as free water. Water activity (a_w) is an important way to measure and monitor water's mobility in baked products.

Making Biscuits Lighter

Leavening agents aerate the dough or batter to make it light and porous. The leavening action is responsible for good volume, improved eating quality, and a uniform cell structure. Leavening can be achieved by various methods, including yeast fermentation, the mechanical incorporation of air by mixing and creaming, formation of water vapor during baking, and the creation of carbon dioxide and/or ammonia by chemical leaveners. However, creation of the initial air bubbles during the mixing phase is critical before any of the other leavening agents can take effect. (*See* **Leavening Agents**.)

Small products like biscuits that bake quickly need a fast-acting leavener that will release the gas before the structure sets. The most widely used source of carbon dioxide in chemically leavened systems is the reaction of sodium bicarbonate or baking soda ($NaHCO_3$) with an acid, usually the acidic salt of a weak mineral acid. The leavening acid promotes a controlled and nearly complete evolution of carbon dioxide from sodium bicarbonate in an aqueous solution. Some examples include monocalcium phosphate monohydrate ($CaH_4PO_4)_2.H_2O$), sodium acid pyrophosphate ($Na_2H_2P_2O_7$) and potassium acid tartrate ($KHC_4H_4O_6$). When these agents combine with water, they react to form controlled amounts of carbon dioxide. Sodium bicarbonate also raises dough pH.

Ammonium bicarbonate (NH_4HCO_3) generates carbon dioxide, ammonia, and steam when heated. It increases spread and gives a larger, more desirable surface 'crack' in some types of hard, high-sugar biscuits. However, it can be used only with low-moisture biscuits that are baked sufficiently to drive off all residual ammonia.

Other Ingredients

Milk products, eggs, and salt are added for variety. Milk and eggs are viewed as wholesome ingredients by consumers; however, they are among the most expensive.

Milk and whey are good sources of protein and lactose, which aid in shape retention and browning reactions. They add flavor and nutrients, improve texture, crust color, moisture retention, and control spread. They are usually added in dried form. (*See* **Whey and Whey Powders**: Production and Uses.)

Eggs contribute color, structure, nutritional value, and some flavor. They affect texture as a result of their emulsifying, tenderizing, leavening, and binding actions. The form of eggs used can be fortified or whole components (yolks and whites) in the liquid, frozen or dried state, or combined with sugar.

Salt performs two principal functions in biscuit doughs. The first is flavor. It accentuates or potentiates the flavor of other ingredients (e.g., the sweetness of sugar is emphasized), and it removes the flatness or lack of flavor in other foods. Moreover, salt has a slight effect on the consistency of hard doughs, because it has a strengthening effect on gluten. Salt also controls fermentation and aids in suppressing undesirable bacteria.

Minor ingredients include malt, proteases, mold inhibitors, spices and flavorings. Though used in relatively small amounts, these ingredients have quite important effects on the sensory and physical qualities of biscuits.

Malt is prepared from barley by sprouting it, then drying it at controlled temperatures. Diastatic malt provides enzymes that break starch into simple sugars and add flavor and color. If the malt is heated sufficiently to inactivate the (amylase) enzymes, for use as a flavorant, it is called 'nondiastatic.' (*See* **Malt**: Chemistry of Malting.)

Proteases are important in crackers (low sugar). They are often added to modify the gluten framework. The effect of a protease is to make the dough less elastic, so that shrinkage does not occur during sheeting and cutting. Proteases may occasionally be used in cookie production. A dough containing gluten that is too strong will decrease biscuit spread, so proteases can improve the spread ratio. (*See* **Enzymes**: Uses in Food Processing.)

Mold inhibitors are not generally used in low-moisture cookies and crackers, but some types which are higher in moisture may benefit from their inclusion. Other food products used in biscuits (i.e., fillings, toppings, and creams) may require an inhibitor. Some examples include sodium diacetate ($Na_2C_2H_3O_2.HC_2H_3O_2.\ ½H_2O$), calcium propionate ($Ca(CH_3CH_2COO)_2$) and sodium propionate (CH_3CH_2COONa).

Added Flavors

Our choice of food is largely influenced by taste and flavor. Smell, taste, touch, and sight are all influenced by the chemical and physical properties of the food. Sweet bakery foods are often selected for this reason. Therefore, selecting the correct flavor is extremely critical. Selection of the best flavor is most often achieved through experience and trial and error. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Biscuits may be flavored in one or more of three ways: adding the flavoring to the dough before baking; dusting or spraying the flavor on after baking; or by flavoring a nonbaked portion such as cream filling, icing, or jam that is applied after baking.

Spices are aromatic vegetable products (tree bark, seeds, fruits, and roots), and are usually finely ground. They improve quality through smells and tastes. The most commonly used spices are cinnamon, mace, nutmeg, caraway, anise, allspice, poppy seed, coriander, ginger, cloves, and fennel.

Flavorings are alcohol extracts from fruits or beans. Vanilla is the most common because it blends with and enhances other flavors. Flavorings, such as vanilla or almond, can be either natural or synthetic. Because flavorings are volatile, much of them may be lost during baking.

Added Colors

Color additives are used in biscuits and in fillings, icings, and coatings to create a perception of quality and richness. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments; Properties and Determinants of Synthetic Pigments.)

Chemical Changes During Baking

Biscuits are usually baked in a tunnel oven. The dough pieces are placed on either a flexible metal or wire mesh band that travels continuously through the oven's length, which may reach 100 m. Baking is controlled by varying the temperatures in the individual zones within the oven. It is not uncommon for the drier to consist of 5–7 zones. Baking time is regulated by the band speed. Crackers and some biscuits are docked by pressing blunt pins into the dough sheet. Docking dough pieces before baking creates air passages through the crust and seals the top to the bottom, reducing big blisters.

Dough undergoes several changes during baking (Figure 2). Changes in dimension and texture, loss of moisture, and color and flavor development are the most important. Baking is divided into three phases. The first involves dough expansion and the start of moisture loss. Dough expansion and water

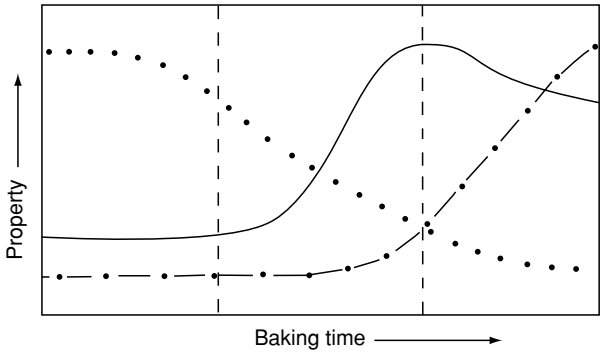

Figure 2 Physical changes in biscuits during baking. Key: —●—, color; —, thickness; ●●●, weight.

loss reach maximum rates and colour development starts during the second phase. The third phase concludes baking with a lower rate of moisture loss, thinning of the biscuit, and increasing surface color.

Dimension and Texture Changes

As dough pieces enter the oven, the biscuit's internal temperature rises, and the sugars and fats melt. The complex matrix contains liquified fat and dissolved sugars. Water evaporation causes the solution to become more concentrated, and the dough temperature continues to increase. Leaveners and steam cause the dough volume to expand. With sufficient heat, limited protein denaturation and starch swelling occur, helping to create some structure. However, the high sugar and fat content inhibit full starch gelatinization.

Doughs rich in fat and sugar, but containing little water, often have unhydrated proteins and ungelatinized starch when baked. With this formulation, a rigid structure cannot be achieved, and is therefore replaced by a soft, sugary matrix. During baking, the dough expands greatly via steam and leavening, but does not set properly, resulting in collapse of the biscuit upon removal from the oven. This expansion–collapse effect is responsible for the characteristic cracked surface of some biscuits.

Moisture Loss

Dough moistures average 11–30% before baking and 1–5% after. Moisture can be lost only from the biscuit surface. It migrates from the center to the surface of the biscuit by capillary action and diffusion. The free water is evaporated very readily, but part of the bound water will also be liberated, in association with color development. Moisture content and water activity are critical parameters, as both directly influence storage life.

Color Changes

One of the most important color reactions is the Maillard reaction. It involves the interaction of reducing sugars with amino groups in the proteins, mainly from lysine, and produces an attractive reddish-brown hue. It is also associated with the dextrinization of starch and the caramelization of sugars. These reactions require very high temperatures, which are reached only at the biscuit surface. (*See* **Browning**: Nonenzymatic.)

Storage and Staling

Most freshly baked biscuits are cooled before packaging or secondary processing, e.g., icing or sandwiching with a cream filling. Snack crackers are sprayed with oil and salted/seasoned prior to cooling. The cooling period is usually 1·5–2 times longer than the baking period. During cooling and storage, the biscuit continues to undergo texture changes. Hard dough biscuits are rigid and crisp immediately after baking. Soft dough products are still flexible at the end of baking, but become firm and crisp after cooling, as a result of sugar recrystallization and glass transition temperature.

Changes in Moisture Distribution

The residual moisture is not uniformly distributed as a biscuit leaves the oven. Most of the moisture lies in a lamella near the center, leaving the surface and the outer periphery almost dry.

'Checking' is a change associated with uneven moisture distribution. Dimensional changes within the biscuit after baking cause it to crack. The center shrinks as it loses moisture, but the rim expands as it absorbs moisture. Checking can happen in almost any type of biscuit or cracker, but is most commonly found in semisweet products.

See also: **Bread**: Chemistry of Baking; **Browning**: Nonenzymatic; **Carbohydrates**: Sensory Properties; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; Properties and Determinants of Synthetic Pigments; **Emulsifiers**: Uses in Processed Foods; **Enzymes**: Uses in Food Processing; **Fats**: Uses in the Food Industry; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Flour**: Roller Milling Operations; Analysis of Wheat Flours; Dietary Importance; **Malt**: Chemistry of Malting; **Vegetable Oils**: Types and Properties; **Whey and Whey Powders**: Production and Uses

Further Reading

Blanshard JMV, Frazier PF and Gailliard T (1986) *Chemistry and Physics of Baking*. London: Royal Society of Chemistry.

Fennema OR (1976) *Principles of Food Science*. New York: Marcel Dekker.

Manley DJR (1983) *Technology of Biscuits, Crackers and Cookies*. Chichester: Ellis Horwood.

Matz SA (1991) *The Chemistry and Technology of Cereals as Food and Feed*, 2nd edn. New York: Van Nostrand Reinhold/AVI.

Matz SA (1992) *Cookie and Cracker Technology*, 3rd edn. McAllen, TX: Pan-Tech International

Pomeranz Y (1978) *Wheat Chemistry and Technology*, 2nd edn. St Paul, MN: American Association of Cereal Chemists.

Pyler EJ (1988) *Baking Science and Technology*, 3rd edn. Kansas City, MO: Sosland.

Smith WH (1972) *Biscuits, Crackers, and Cookies: Technology, Production, and Management*. Barking: Applied Science.

Stauffer CE (1990) *Functional Additives for Bakery Foods*. New York: Van Nostrand Reinhold/AVI.
Sultan WJ (1986) *Practical Baking*, 4th edn. Westport: AVI.
Wade P (1988) *Biscuits, Cookies and Crackers*, New York: Elsevier.

Wafers

K Tiefenbacher, Vienna, Austria

Background

Wafers – thin, crisp and precisely shaped products from cereals – are a very special part of the biscuits, cookies, and crackers area. The name covers several types of wafers, quite different in recipes, manufacturing equipment, and end use. The fundamental steps of making flat wafers, hollow wafers, molded cones, rolled wafer cones, and wafer sticks are outlined in some detail.

Concepts like water activity and glass transition are of critical importance here. Moreover, the function of wafers as a part of combined food products is discussed. Finally, some important trends in both products and manufacturing equipment should be noted.

Main Wafer Features

Wafers are baked as sheets, cones, and sticks or with different fancy shapes. The characteristic features with respect to other bakery products are:

1. Wafers are very thin biscuits; the overall thickness is usually between less than 1 and 5 mm. They often have a typical 'wafer pattern' on one surface or on both. The surfaces are smooth and precisely formed, with the dimensions and all the details – engravings, logos, etc. – of the baking molds. See **Figure 1** for some examples.
2. Wafers are cereal-based low fat products made of wheat flour, sometimes with addition of other flours or starches. The product density is in the range of 0.10–0.25 g cm^{-3}. In cross-section, the wafer matrix is highly aerated and primarily of gelatinized starch.
3. Wafers, by their typical delicate and crisp texture, combine well with different fillings (cream, ice-cream, foam) and coatings.

There is sometimes confusion in terminology between the crisp 'wafers' as described here and 'waffles,' which are of a soft, cake-like texture but show some kind of wafer pattern, too.

Basic Wafer Types

There are two basic types of wafers:

1. No- or low-sugar wafers. After baking, these contain from zero to a very low percentage of sucrose or other sugars. Typical products are flat and hollow wafer sheets, molded cones, cups, and fancy shapes.
2. Higher-sugar wafers. Well over 10% of sucrose or other sugars are responsible for the plasticity of the hot, freshly baked sheets. These are formed into different shapes before sugar recrystallization occurs. Typical products are rolled sugar cones, rolled wafer sticks or tubes, and deep-formed fancy shapes.

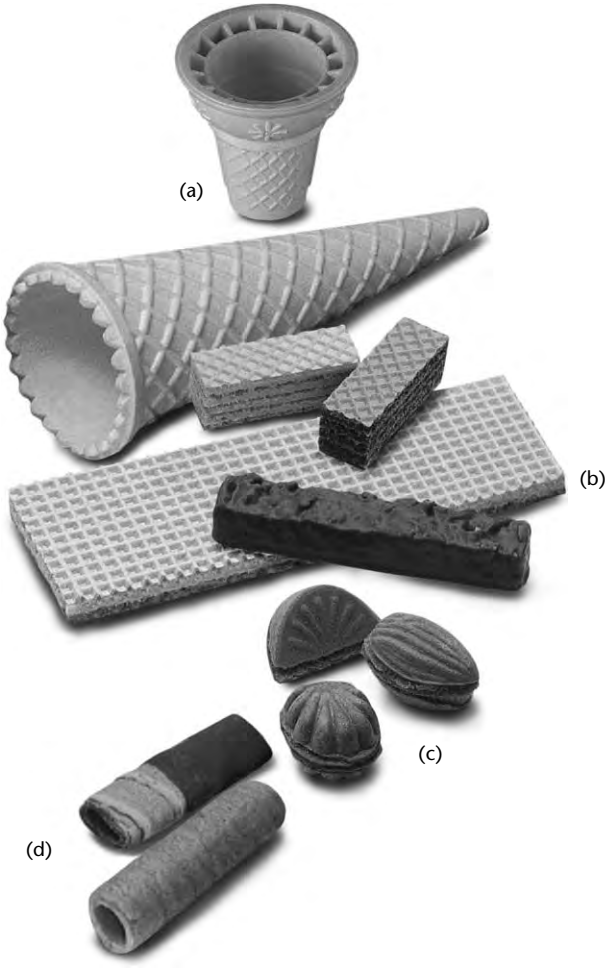

Figure 1 Wafer products. *From top to bottom*: (a), molded cones; (b), flat wafers, creamed/enrobed; (c), hollow wafer pieces; (d), wafer sticks.

As in both wafer types, the main ingredient is flour, wafers fit very well into current dietary recommendations to consume more cereals. They are high-carbohydrate, low fat products. (*See* **Cereals**: Dietary Importance.)

The baking of 'wafers' between hot metal plates has been known since medieval times, but these first wafers were more similar to our waffles or pancakes in their high fat and egg contents and their texture.

Modern wafers are low-fat cereal products, very similar to the altar breads for Christian churches, and are basically made out of flour and water. The first wafer ovens were used after World War I, but more automatic manufacturing lines have been available since the mid-1950s.

Wafer Recipes

Before creating a wafer recipe, two main questions have to be answered:

1. What is the end use of the wafer? If it is part of a cream-filled, chocolate-covered biscuit, where contributing a crisp texture element is far more important than the taste of the wafer itself, recipes with few components are recommended. If the wafers are consumed directly as wafer bread or wafer sticks, more sophisticated recipes are chosen.
2. What kind and quality of raw materials are available? Low- to medium-protein soft wheat flours with a low water absorption work best, especially for no-sugar wafers. Problems with suboptimal flours must be balanced by variations in minor ingredients in the recipe. The use of wholemeal flour is possible, and in some regions, other cereals such as rice or corn are used for wafer production.

Table 1 lists the common ingredient ranges for both types of wafer.

In-line Manufacturing of Wafer Biscuits

Wafer biscuits are the most important products by volume. To explain the different steps of manufacturing, we follow Figure 2.

Batter Preparation (Figure 2, process 1)

Wafers are made from a fluid batter with a typical viscosity in the range of 300–2000 mPs. First, the water-soluble components are dissolved. By adding the farinaceous ingredients, a homogenous suspension, the 'wafer batter,' is obtained within a few minutes of mixing.

Batter Transport and Depositing (Figure 2, process 2)

From an intermediate tank via a ring main, the batter is pumped to the oven and spread on to the baking molds by a depositor head.

Wafer Sheet Baking (Figure 2, process 3)

The baking of wafer sheets is performed in 'tongs,' i.e., pairs of cast-iron metal plates with a hinge and

Table 1 Wafer batter ingredient ranges (weight parts, flour = 100)

	No (low)-sugar wafer	*Higher-sugar wafer*
Wheat flour	100	100
Water	120–160	100–140
Starch	0–12	0–5
Sucrose	0–4	25–75
Oil/fat	0.5–2	1–6
Milk powder	0–2	0–2
Soya lecithin	0.2–1	0.2–1,5
Salt	0–0.6	0–0.6
Sodium bicarbonate	0.1–0.5	0–0.3

Optional minor ingredients: other cereal flours, soya flour, other sugars and syrups, egg-based ingredients; whey powder, yeast, caramel color, cocoa powder, colors, ammonium bicarbonate, enzymes.

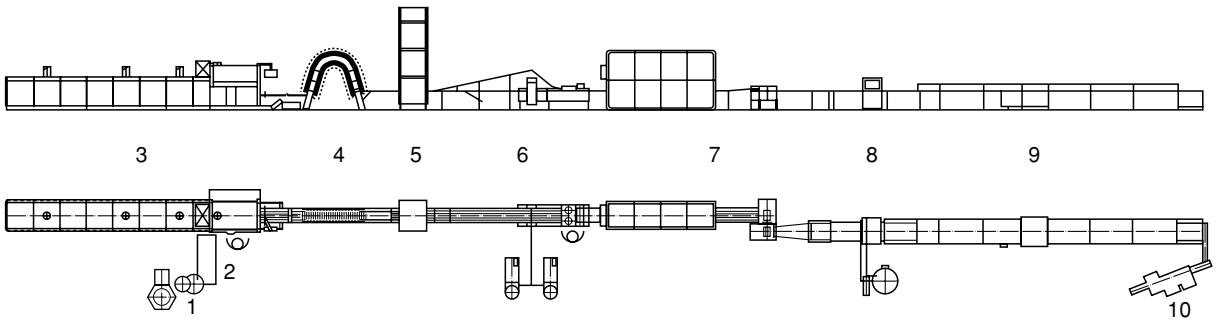

Figure 2 Manufacturing line for creamed and enrobed wafer biscuits from flat wafers. (1), batter preparation; (2), batter transport and depositing; (3), wafer sheet baking; (4), wafer sheet cooling; (5), conditioning of sheets; (6), creaming and wafer book building; (7), cooling and cutting of books; (8), enrobing; (9), cooling of enrobed pieces; (10), packaging.

latch on opposite sides. Baking plate sizes up to 350 × 730 mm are available. The precisely machined baking plates carry reedings or other engravings. We call the resulting wafers 'flat' wafer sheets, with an overall thickness of no more than 2–5 mm.

But such baking plates can also carry special figures (nuts, sticks, hemispheres, fancy shapes) up to a depth of approximately 20 mm, thus yielding the so-called 'hollow' wafer sheets.

Modern wafer-baking plates often are surface-plated, e.g., with chromium for easier release and reduction of cleaning stops. The plates are edged with metal strips to give a closed baking mold, except for small venting channels for steam release. Wafer-baking ovens frequently have 32–104 pairs of plates, continuously circulating on a chain. They are mostly gas, sometimes electrically heated and operate at mold temperatures between 160 and 190 °C.

The baking process Within a few seconds after batter deposition, the baking molds close and are locked. At first, the batter is distributed mechanically, but then the mold is filled completely by the steam that evolves. A small quantity of batter is extruded as baking waste 'bobbles' through the venting channels. As gelatinization of starch starts immediately, the pressurization of the mold by steam occurs at the right moment, resulting in a well-aerated starch foam.

When most of the water has been driven off, the glass temperature of the wafer matrix rises, and the stable structure is formed. The temperature of the wafer increases to 160–190 °C, the temperature of the baking mold. Then, from Maillard reactions, the typical wafer color and flavor develop. (*See* **Browning**: Nonenzymatic.)

The overall baking times are between 1.5 and 2.5 min, depending on the wafer thickness and baking temperature.

During the wafer manufacturing process, there is no substantial degradation of starch molecules compared with other bakery products such as extruded cereals. Therefore, wafers have two unique textural properties:

1. Extreme crispness on biting and initial chewing.
2. Good mouth feel during prolonged chewing and swallowing owing to the absence of sticky, glutinous stimuli.

Wafer Release and Cooldown (Figure 2, process 4)

At one end of the oven, the plates open to release the baked sheets and to spread fresh batter, and then reclose very quickly. The sheets are cooled to room temperature while passing over an arch-type sheet cooler.

Wafer Conditioning (Figure 2, process 5)

Next, the wafers optionally pass a conditioning unit, where the moisture content of the sheets is carefully increased so as to achieve some stability in both the texture and size of the wafer.

Moisture sorption and wafer texture After baking, the residual moisture is 0.8–1.5%. Looking to the sorption isotherm, this corresponds to a water activity of around 0.1 or even lower. As both the water activity of the air in the production area and the water activity of filling creams or coatings are well above that, wafers pick up moisture very easily. In line with this sorption, the dimensions of the sheet increase by 0.2–0.3% for every 1% of additional moisture. This can result in cracking of the coating in enrobed wafer biscuits.

To compensate for the low water activity, humidity conditioning up to approximately 4.5% wafer moisture is possible. This is recommended, especially if enrobed or chocolate molded wafer products are made, in order to anticipate this first dimension increase and to avoid any cracking of the coating during its shelf-life.

Moreover, with increasing water activity, the wafer texture changes from a soft to a harder crispness, accompanied by a higher mechanical stability, which is good both for handling and for the final product texture. Up to 5–6% moisture content, the wafer sheets keep their typical crisp texture, but higher moisture levels will result in most cases in inadequate, tough, or even soft and soggy textures (see Table 2 for details).

Table 2 Water activity, water content, and texture of wafers

Wafer condition	*Wafer texture*	*Water activity, approx.*	*Moisture (%), approx.*
Freshly baked	Very tender, crisp	< 0.1	< 2
After conditioning	Crisp, harder	0.3	4.5
Limit of crispness	Crisp to tough	0.5/0.55	≥6
Wafers, foam-filled	Soft to flexible	0.7	≥12
Collapse of structure	Very soft, shrinks	> 0.85	> 20

Creaming and Book Building (Figure 2, process 6)

The sheets then pass the creaming station, where a layer of a sugar- and fat-based cream is applied to one side. The type of cream flavor liked most varies regionally, but chocolate, vanilla, hazelnut, milk, strawberry, and lemon are the most common. Creaming is done at temperatures of 30–40 °C either by contact spreading or by depositing a preformed cream film. For flat wafers, several creamed sheets together with a noncreamed top sheet form a so-called 'wafer book.'

For hollow wafers, the cream is added to the hollow parts either by spreading or by controlled single depositing. Again, either two hollow wafer sheets or a hollow and a flat sheet are combined to form a 'book.'

Cooling and Cutting (Figure 2, process 7)

The wafer books pass a cooling tunnel to set the cream, after which they are wire or saw-cut into small biscuits.

Enrobing or Molding in Chocolate, Cooling (Figure 2, processes 8 and 9)

The cut biscuits may be enrobed with chocolate-type coatings, sometimes after the application of chopped nuts or crispies to the top wafer. Molding in chocolate is another possibility. After a final cooling step, the biscuits are ready for packaging.

Packaging (Figure 2, process 10)

The biscuits have to be packed tightly to protect against humidity, but also against oxygen and light to prevent oxidative deterioration and to insure a shelf-life of 6–9 months. Inadequate packaging-film moisture barriers and bad sealing are the most frequent reasons for later complaints by customers. Laminated or specially coated films are used for the typical wafer product packaging in flow packs, boxes, or bags.

Manufacturing of Molded Wafer Cones

Another type of 'hollow' wafers, cones, cups, and fancy shapes with up to 185 mm in length are produced in cast-iron molds. Holes for four to six items are provided in each of these molds, and 12–72 molds are circulated in one oven. The lower part of the mold is made of two symmetric halves that open to release the baked pieces. After their reclosure, they take up a fresh batter deposit, and finally the 'core,' the upper part of the mold, closes the mold for a new baking cycle. The baked cones are cooled down and stacked for packaging.

Recipewise, there are two groups of molded cones:

1. No- or low-sugar cones, generally known as 'cake cones,' whose recipes are similar to those for sheets (see Table 1).
2. The so-called 'molded sugar cones' have an intermediate sugar content, usually below 20 parts of sucrose for 100 parts of flour.

Recent Developments in Molded Wafer Cones (see Figure 2)

Whereas traditionally molded cones have flat tops and regular, symmetric reedings now for a few years molds for more sophisticated products can be manufactured. We now see cones and cups in the market showing curvilinear tops and artfully designed outsides. Figure 3 shows the first of these, which comes from a Japanese cone manufacturer.

Manufacturing of Rolled Wafer Cones (see Figure 4)

'Rolled sugar cones' need a concentration of more than 20% of sucrose or other sugars in the finished product. The first three steps of rolled sugar cone manufacturing are rather similar to no/low-sugar wafer sheet baking with the exception of reduced steam pressure as there are no baking ledges to build up pressure and to extrude 'bobbles.' The deposit is just to form an oval or circular sheet.

1. Batter preparation (Figure 4, process 1)
2. Batter transport and depositing (Figure 4, process 2)

Figure 3 Molded wafer cone: example for new, nonflat top design.

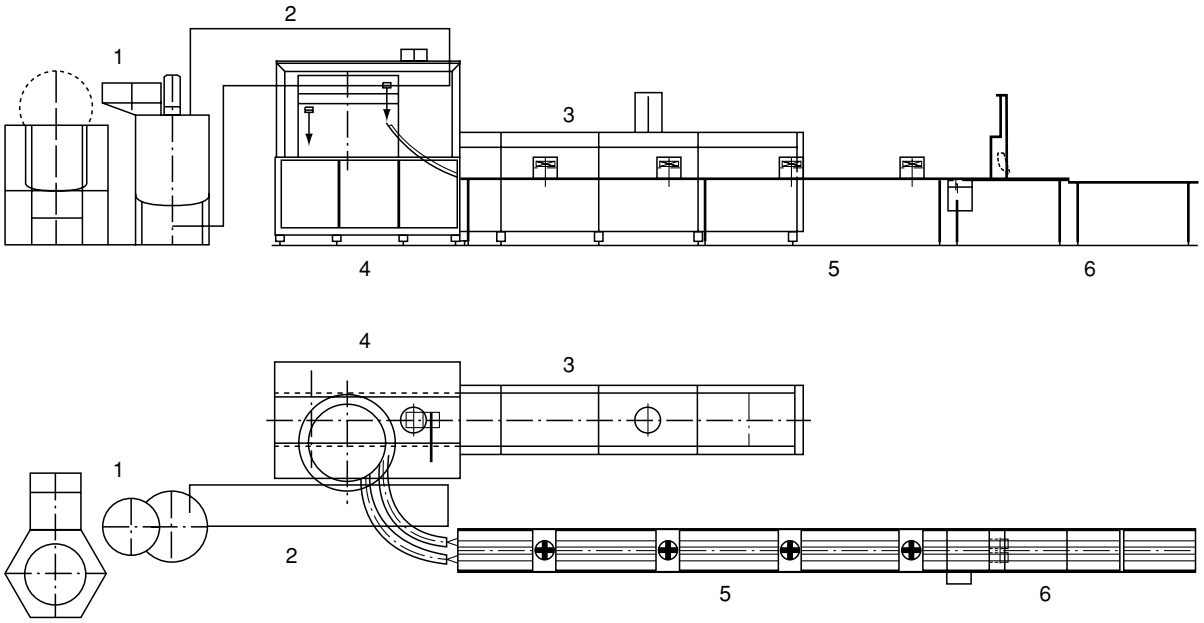

Figure 4 Manufacturing line for rolled sugar cones. 1, batter preparation; 2, batter transport and depositing; 3, wafer sheet baking; 4, take-off, rolling and cone release; 5, cooling of cones; 6, cone sleeving, stacking.

3. Wafer sheet baking (Figure 4, process 3): Nowadays, equipment is available to manufacture up to 13 000 cones per hour.
4. Wafer take-off, rolling, and cone release (Figure 4, process 4): When, after baking, the mold reopens, the sheet is automatically stripped off the plate and rolled immediately on tapered mandrels to form the finished cone. A series of rolling devices mounted on a round table operates continuously: Sheet removal, rolling, release, etc.

Glass Transition and Formability

During the rolling procedure, the wafer temperature is still above the glass transition point. The molten sugar acts as a plasticizer, and so any rolling or deep-forming operation can be performed. Besides the rolled cones, three more types of products are manufactured by the same principle:

1. Wafer 'rolls': Wafer sheets of rectangular shape are baked, stripped off the baking plate, and rolled into a sugar wafer 'rod' without a center hole (different from wafer 'sticks' with a center hole discussed later). The rod is later cut into smaller cylindrical pieces. Typically, these pieces are finally partially enrobed with chocolate-type coatings.
2. 'Deep-formed' wafers (cups, shells), where the hot, flat wafer piece is introduced into a forming tool, sometimes with an additional embossing or stamping device, which defines the final shape.
3. Rolled wafer sticks. These are discussed in a separate chapter below.

Cone Cooling (Figure 4, process 5)

The fresh cones pass through a cooling section, where ambient or cooled air is applied. Here, sugar recrystallization finishes to give the final strong and brittle texture.

Cone Sleeving and Stacking (Figure 4, process 6)

Now, paper sleeves are applied automatically, and these form the packaging material for the typical end product, industrially manufactured icecream cones. Next, is automatic cone stacking and packaging.

Later in the icecream plant, the wafer cones are sprayed inside with chocolate and filled, and finally, the paper cone is closed.

Manufacturing of Rolled Wafer Sticks

Rolled wafer sticks are hollow tubes with walls consisting of very thin multiple layers, similar in texture to crunchy cereals. These layers do not carry a wafer pattern and are approximately 0.5 mm in thickness. But from the recipe, the sticks are typical higher-sugar wafers (see Table 1).

In manufacturing, a stripe of batter is applicated to a heated drum and baked into a continuous wafer band. The band is rolled immediately while hot into a continuous still formable tube with an internal

diameter of about 6–36 mm. From here, there are several options to create a large family of finished products:

1. The tube is automatically cut into wafer stick pieces and cooled. These are consumed as sweet snacks or with icecream, for example. Other end uses include intermediate products for confectioners to finish by additional filling or decorating operations.
2. During the rolling operation, the tube is coated inside with compound chocolate or filled with cream, either partially or fully. Then, cutting and cooldown, followed optionally by a coating and/or decorating process, give the finished product.
3. The filled tube is transformed into small pillow-like bits by a combined squeezing–cutting operation.
4. The wafer tube pieces, nonfilled, coated inside, or partially cream-filled immediately after cutting while still hot, pass a pressing station to form oval or even flat pieces. Such flat pieces may form the center of premium-type confectionery items with a chocolate coating, including chopped nuts or cereal crispies, for example.
5. Larger-diameter tubes are pressed flat, and by an embossing and stamping operation, so-called 'fan wafers' or other elaborate shapes can be generated.

The diameter and length of the sticks as well as the number of very thin, 'glassy' sheets forming the wall of the stick can be adjusted individually. Rolled wafer sticks have a unique brittle wall texture without being too hard in biting and chewing.

Savory Wafer Products

Whereas many end uses of wafers in combination with sweet fillings, coatings, etc. are well known, savory or neutral-tasting wafer products are becoming increasingly popular in some areas. Some examples:

1. Wafers as crispbread, delivering a crunchy but softer texture than traditional hard crisp breads. Besides wheat flour there are several options for other cereal and noncereal flours as raw materials. The products are either neutral in taste to be eaten with sweet or nonsweet toppings or have spices, cheese, and other flavorsome ingredients.
2. Wafers – flat or hollow wafers – with savory fillings, e.g., peanut butter or cheese creams.
3. Wafer sticks, a recent development, made from a nonsweet but still rollable material. These are typically combined with savory cream fillings.

Trends – From the Very Wafer to a More Sophisticated End Product

After discussing the different wafer types and their traditional end uses (for an overview, see Table 3), a sharply increasing number of new confectionery brand products during the last decade are worth mentioning. Here, wafers are only a smaller, but still important, part of the deal. These are small confectionery pieces, mostly in bite-size or sectioned bars, often chocolate-covered, where the wafer has three main functions:

1. To add its typical texture and crispness, e.g., to a soft filling cream with a piece of nut and a chocolate cover, thus imparting a multitextured impression.
2. To make the overall product 'lighter' in calories and nutritionally, as the wafer part is of an extremely low density and cereal-based.
3. To define the precise structure of the piece, e.g., when holding a soft filling, or, for noncoated pieces, to keep the consumer's hands clean if the filling melts due to warmer conditions.

Looking to some of the current trends, we see increasing interest in combining lower fat fillings such as toffee (caramel). Another trend may be newer fat free fruit-type fillings with water activities that are so low that the wafers stay crisp. Even some new examples of fillings with a higher moisture, resulting in a 'soft' wafer part, can be rolled out.

Table 3 Which product from which wafer type?

End product	*No (low)-sugar wafer*	*Higher-sugar wafer*
Wafer crisp bread	Flat wafer sheet	
Sugar wafer biscuits, creamed	Flat wafer sheet	
Wafer biscuits, creamed and enrobed or molded in chocolate	Flat wafer sheet	
Containers, e.g., for icecream	Molded cake cone	Molded sugar cone, rolled sugar cone
Confectionery, filled, enrobed, decorated	Hollow wafer *or* hollow + flat wafer	Wafer stick *or* sugar cone, small
Fan wafers		Wafer stick, pressed, embossed
Wafer rolls, partially enrobed		Rolled sheet
Wafer bowls		Deep-formed sheet

Figure 5 New, inductively heated, low-emission oven for rolled wafer stick manufacturing.

Developments in Wafer-manufacturing Equipment

After a long period of just 'routine' technical optimizations, recent years have seen a few new technical concepts in wafer-manufacturing equipment:

1. The 'stack oven,' with the baking plates vertically stacked. Here, the weight of the stack itself eliminates the need for hinges and for closing mechanisms as in the traditional chain oven. Moreover, the floor area needed is greatly reduced. The coming years will show whether that development will find its way into industrial practice.
2. A new 'low-emission' heating concept together with a 60% reduction in energy consumption for rolled wafer stick ovens. The key is the replacement of the gas-heated drum by a ring, heated by induction. This results in a new, more consistent and more controllable product quality, and results in the elimination of flue gases and a huge reduction in energy consumption for baking, as the first industrial oven prototypes (Figure 5) showed.

See also: **Browning**: Nonenzymatic; **Cereals**: Dietary Importance

Further Reading

Barron LF (1977) The expansion of wafer and its relation to the cracking of chocolate and 'baker's chocolate' coatings. *Journal of Food Technology* 12: 73–84.

Blanshard JMV, Frazier PJ and Galliard T (eds) (1986) *Chemistry and Physics of Baking*. London: The Royal Society of Chemistry.

Haas F, Haas J and Sachsenhofer J (1998) Bakers' oven for producing thin-walled shaped products with baking forms each consisting of two form halves which open and close. *PCT Patent Application* WO 98/51156.

Manley DJR (2000) *Technology of Biscuits, Crackers and Cookies*, 3rd edn, pp. 290–306. Cambridge, UK: Woodhead Publishing.

Negri G (1984) Waffeln und Verfahren zu deren Herstellung. *European Patent* 0,022.901.

Oliver G and Sahi SS (1995) Wafer batters: a rheological study. *Journal of the Science of Food and Agriculture* 67: 221–227.

Pritchard PE and Stevens DJ (1974) The influence of ingredients on the properties of wafer sheets. *Food Trade Review* 42(8): 9–15.

Roos YH (1995) *Phase Transitions in Foods*. London: Academic Press.

Schwartzberg HG and Hartel RW (eds) (1992) *Physical Chemistry of Foods*. New York: Marcel Dekker.

Seibel W, Menger A, Ludewig H-G, Seiler K and Bretschneider F (1978) Standardisierung eines Backversuches für Flachwaffeln. *Getreide, Mehl und Brot* 32: 188–193.

Stevens DJ (1976) The role of starch in baked goods. Part 1 The structure of wafer biscuit sheets and its relation to composition. *Starch* 28: 25–29.

Tiefenbacher K and Dobrovics M (1999) Mehle für die Waffelherstellung. *Getreide, Mehl und Brot* 53: 349–353.

Wade P (1988) *Biscuits, Cookies and Crackers*, vol. 1, pp. 139–150. Barking, UK: Elsevier.

BLACKBERRIES AND RELATED FRUITS

D L Jennings, Otham, Maidstone, UK

Introduction

Although in the same genus as raspberries, blackberries are often regarded by the food industry as the Cinderella of the soft fruits. Nevertheless, they have considerable potential, which is likely to be realized to an increasing extent as new and improved cultivars become available. This article first describes the main types of blackberries and their botanical classification. Their anatomy, fruit pigments, and some of the characteristics that determine their fruit quality are then discussed, particularly color and flavor, followed by a discussion on the problems of postharvest storage and the various market outlets.

Types of Blackberries

The botanical classification of blackberries is complex, but from the food viewpoint, it is useful to consider just three types: the traditional blackberries of Europe and eastern North America typified by Evergreen Thornless, Loch Ness, and Triple Crown, the indigenous blackberries of western North America typified by Marion, Kotata, and Helen and blackberry–raspberry hybrids typified by loganberry, tayberry, and boysenberry. World production is concentrated in a few areas, but there are low levels of production in most of the areas where raspberries are grown. Two major production areas are Washington and Oregon on the west coast of America, where cultivars of each type are grown but with a predominance of European forms, another is in New Zealand, where the boysenberry, a blackberry–raspberry hybrid, predominates, and a third is in Arkansas, USA, where locally bred cultivars are grown.

The flavor usually associated with blackberries is that of the European type. The flavor of the western North American type is more intense and pleasantly acidic, and has strong aromatic overtones. The loganberry, tayberry, boysenberry, and youngberry are all hybrids derived from western North American blackberries and red raspberries. They have inherited much of the characteristic flavor of this type of blackberry. Of these hybrids, the loganberry and tayberry contain one-third of raspberry germplasm and have purple fruits; the others contain a lower proportion and have black fruits. The Andean blackberry of South America, *Rubus glaucus*, is the only blackberry–raspberry hybrid to be derived from a black raspberry, which has undoubtedly contributed to its excellent flavor.

Until recently the available cultivars of the western North America type have been low yielding and not widely grown outside their area of origin, but new cultivars such as Kotata, Siskiyou and Helen are much improved in productivity and quality, and are providing new opportunities for blackberry production and usage both in North America and in Europe. Their excellent flavor and early cropping are so different from those of the European type that the two forms complement each other to provide a choice of flavor and a longer season of production.

Fruit Structure

Blackberry fruits are aggregates of drupelets. A drupe fruit is one that develops entirely from a single ovary, and the aggregate of drupelets that makes a blackberry fruit is formed by the ripening together of some 50–100 ovaries from the same flower and adhering to a common receptacle: in a sense, each drupelet is a complete fruit in itself and a miniature homolog of such drupe fruits as the cherry or plum.

The cohesion of these drupelets depends on the entanglement of epidermal hairs, which are unicellular linear trichomes arising from surface cells. The hairs are abundant at the bases and sides of the drupelet, and shorter and less profuse over its dome; they are so tightly enmeshed that the drupelets cannot normally be separated without tearing the skin. A poor set of drupelets or poor cohesion between them causes 'crumbly fruit' and a serious loss of quality.

Internally, the soft part of the drupelet consists of thin-walled parenchymatous cells, radiating from the pyrene in the center to a region of larger, oval-shaped cells that underlie the epidermis, and a slightly collenchymatous hypodermis of one to three cells. Fruits are succulent because the walls of the parenchyma are thin and the cells turgid: variations between cultivars in fruit firmness are influenced by variations in the frequency of cells of small diameter, tissue compactness, and overall cell size, which influences the amount of supporting cell wall. The strength of the epidermis and cuticle influences the 'skin' strength of the berry.

Pyrenes

In the center of each drupelet is a pyrene (often erroneously referred to as the seed) with a hard endocarp

enclosing one or two true seeds. The endocarp consists of two layers of elongated, parallel scleroid cells at right angles to each other and each several cells thick, the outer one forming the characteristic ridged pattern of the pyrene.

The large size of most blackberry pyrenes is one of the main reasons why relatively few blackberries are used for jam-making. Large pyrenes are also undesirable in fresh fruit. Blackberry cultivars show a wide range in this characteristic: surveys showed a range in pyrene weight, which is indicative of size, from 2.10 to 4.83 mg for American cultivars and from 2.10 to 6.10 mg for breeding material at the Scottish Crop Research Institute. The American cultivars Waldo and Kotata have individual pyrene weights of 2.18 and 2.41 mg, respectively, and are examples of cultivars that have small pyrenes and are suitable for jam-making. However, small pyrenes are only advantageous if their small size is associated with good development of the juicy soft tissues that surround them. Another potentially important quality of the pyrenes is their tendency to become 'blind' in jam. Nothing is known about this characteristic in blackberries, because few of them are used in jam-making. Blindness occurs during the storage of raspberry jam when a slow displacement of air from within the pyrenes allows the surrounding syrup to infiltrate and causes them to lose their opacity. They then appear to have merged with the surrounding jam, which acquires a dull appearance and gives the illusion of a low fruit content. The problem could appear if small-seeded blackberry cultivars become popular for jam-making. The rate at which the process proceeds would be expected to vary with cultivar, method of fruit storage, and manufacturing procedure.

Fruit Ripening and Abscission

All the drupelets of a blackberry usually ripen together, but ripening at the top of the fruit is sometimes slightly later to give a red tip to an otherwise black fruit, especially in large, elongated fruits. Ethylene production begins at a late stage of color development in Maryland, USA, where the blackberries studied can therefore be described as 'climacteric' fruits. No ethylene production was detected in a study in Scotland, however. The difference between these two results may be attributable to the temperatures prevailing at the ripening time, because blackberries ripen in mid-summer in Maryland and much later in Scotland, or to the different cultivars studied. It is interesting that in the Scottish study, the behavior of two blackberry–raspberry hybrids, tayberry and tummelberry, was intermediate between that of raspberries and blackberries, and was therefore closer to that described for blackberries in Maryland. For blackberries grown in the USA, the ethylene-releasing chemical 'Ethrel' has been used to hasten ripening for machine harvesting. An abscission layer forms in the tissues as the fruit ripens. In blackberries, it is a single layer at the base of the receptacle, which consequently separates from the plant with the fruit, in contrast to raspberries, in which a large number of abscission layers form, one at the point of attachment of each drupelet to the receptacle, which is therefore retained on the plant after fruit abscission. This is the most fundamental difference between blackberries and raspberries, and it leads to an interesting situation in blackberry–raspberry hybrids, which are genetically heterozygous for both types of abscission. Dominance is not always complete, and the fruits of some hybrids do not abcise exactly like those of either parent, although most of them have the system of their blackberry parent. The tayberry, for example, has a poorly developed abscission layer, but the fruits separate like a blackberry except that the point of separation is frequently proximal to the calyx, which is then retained on the fruit when it is picked.

This difference between blackberries and raspberries in their method of fruit abscission explains why the core of most blackberries is soft and palatable, while that of some blackberry–raspberry hybrids such as tayberry is less so: blackberry fruits have been selected for the edibility of their receptacles, which become the cores of the ripe fruits, whereas the receptacles of raspberries are always left on the plant and are never eaten. Fruits of the hybrids are intermediate and often have a central core that is harder than that of the true blackberries.

Fruit Color

The color of blackberries results from the presence of anthocyanin pigments. The anthocyanin molecule consists of an aglycone with a varying number of sugar residues attached to the hydroxyl group, usually in the 3-position. Only two anthocyanins have been found in blackberries, which have a predominance of the monosaccharide form, cyanidin-3-glucoside, and a trace of the diglycoside cyanidin-3-rutinoside, in which an additional rhamnose sugar is present. This is a very small number of pigments for a *Rubus* fruit, and the presence of other anthocyanins indicates that a fruit has a hybrid origin. Thus, the diglycoside cyanidin-3-sophoroside, which is a characteristic pigment of the red raspberry, occurs in the loganberry, boysenberry, youngberry, and mertonberry, and is consistent with these fruits being related to the red

raspberry. Similarly, the occurrence in the Andean blackberry of the diglycoside cyanidin-3-sambubiose, which is a characteristic pigment of black raspberries, supports the hypothesis that this blackberry is related to the black raspberry.

The identity of the sugar residues attached to the aglycone part of the anthocyanin molecule is thus useful to complement taxonomic observations and establish the hybrid origin of some cultivars. However, there is no evidence to suggest that any particular glycoside should be preferred from the viewpoint of the quality of fruit color.

The concentration of anthocyanins in blackberries is intermediate between those of red raspberries and black raspberries. For example, fruits of the cultivar Ashton Cross, a typical nonhybrid blackberry, were found to contain 65.45 mg of cyanidin per 100 g of fruit (62.56 mg of cyanidin-3-glycoside and 2.89 mg of cyanidin-3-rutinoside), compared with 22.91 mg for a typical red raspberry and 253.77 mg for a typical black raspberry. Hence, black raspberries, and not blackberries, are used as a source of this natural pigment.

Expression of Fruit Color

The color of anthocyanin pigments is influenced by any chemicals that influence the pH of the solutions that contain them. The pigments are chemical indicators that are red in acid solutions, violet, or purple in neutral solutions and blue in alkaline solutions. This may explain why the color of blackberries frequently changes from black to red when the fruit is frozen: freezing causes widespread cellular disruption, which possibly allows mixing of the cells' plasma and vacuolar contents and places the anthocyanins in a solution of lower pH than where they occur in nonfrozen fruits. However, thoroughly ripe blackberries are less prone to turn red when frozen than less ripe fruits, and a more likely explanation of the color change is that many fruits are not fully ripe when picked. Anthocyanins develop first in cells near the surface of the fruit, where they are exposed to sunlight, so that slightly immature fruits cannot be recognized immediately. The immature fruits have a lower concentration of anthocyanins, and the cell disruption following freezing allows the anthocyanins to spread within the tissues and become diluted by cytoplasmic sap: for these fruits, this pigment dilution is probably sufficient in itself to cause the color change. The problem is greater when fast freezing is used, and cultivars with purple–black fruits, such as Bedford Giant and Silvan, are more prone than cultivars with intensely black fruits.

Fruit Chemistry

Blackberries contain about 14% solids, which are approximately equally divided between soluble and insoluble forms. The size of the pyrene and the relative development of the surrounding soft tissues influence the proportion of soluble to insoluble solids. Pectins are an important constituent of the soluble fraction and average 0.8% (w/w, expressed as calcium pectate) with a range of 0.35–1.19%. Loganberries have a lower content. Protopectins form the intercellular cement and contribute towards the firmness of fruit texture, but they decrease with ripening, owing to hydrolysis. Flavor is determined by the content of sugars, acids, and volatiles, all of which vary with cultivar and growing conditions. Flavor is not perceived as a number of separate characteristics but rather as an overall impression. It is therefore not strongly correlated with any of these factors. Fruits of the western North American cultivars have a more aromatic flavor than those of the European types, and fruit grown in areas that have warm, dry summers have more sugars and are more aromatic than fruit grown in wetter and milder regions. The ratio of sugars to acids plays a major part in determining flavor.

The main sugars are the reducing sugars, glucose and fructose, and there is a smaller amount of sucrose; these form the major soluble component of the juice. The main acids are malic acid and isocitric acid with its lactone; there is only a trace of citric acid (Table 1). In boysenberries, the proportion of the different acids was found to change as ripening proceeded: the proportion of malic acid decreased, whereas the proportion of citric and isocitric acids increased. A number of trace acids also occur, especially in blackberry–raspberry hybrids.

Table 1 Typical sugar and acid contents (% w/w) of European blackberries and two blackberry–raspberry hybrids

	Sugars			*Acids*				
	Reducing	*Sucrose*	*Total*	*pH*	*Typical ripe fruit*	*Range*	*Citric*	*Malic*
Blackberry			4.30	3.00	1.50	0.68–1.84	Trace	0.82
Boysenberry	4.20	1.14	5.34	2.91	1.51		1.24	0.21
Loganberry			3.40	2.90	2.63	1.02–3.12	2.10	0.53

Green A (1971) Soft fruits. In: Hume AC (ed.) *The Biochemistry of Fruits and their Products*. London: Academic Press.

The acids have a high buffering capacity, which maintains a stable pH close to 3.0. The best measurement of the amount of acid present is therefore titratable acidity. As fruit development proceeds, this quantity increases at first and then decreases as ripening starts. It is lower at high temperatures. The relationship between titratable acidity and ripeness is so close that it is the best quantitative measure of fruit ripeness and has been used to assess the ripeness of fruits harvested by machine.

Although a large number of volatile compounds that probably contribute to the flavour of *Rubus* fruits have been isolated and identified, blackberry fruits have received relatively little attention, and no conclusions on the identity of the important components are available.

Blackberries are generally poor sources of vitamins, but provide useful amounts of ascorbic acid and vitamin E. They also provide a useful source of fiber. Mineral content is also low, with a predominance of potassium and calcium, and there are low contents of proteins and polypeptides and traces or larger amounts of a number of amino acids (Table 2).

Post-harvest Management

The commercial life of blackberries is short, partly because of the physiological activity of the fruit after harvest, and partly because of the common occurrence of infection by fruit-rotting pathogens, especially the ubiquitous *Botrytis cinerea*. This pathogen infects before fruit harvest, remains quiescent within the fruit until after harvest, and then becomes aggressive and causes gray mold disease.

The fungus can infect blackberries in various ways. It can probably infect at the flowering stage, as in raspberries, it can enter minute wounds in the fruit, and it can infect by the contact of diseased fruit. Its incidence can be reduced by sprays applied from flowering time until a short interval before harvest, and after harvest, the development of the disease can be inhibited by rapid cooling within 3 h of picking. This also prolongs the fruits' shelf life by inhibiting physiological activity. It is done by placing the containers of fruit in a cold but humid airstream in a refrigerated store designed to lower the fruits' temperature from the field level of 20–2 °C in under 4 h.

Respiration is the physiological activity mostly responsible for reducing shelf life. Its rate is closely related to temperature within the range 0–16 °C and is very high at 21 °C. These high rates of respiration are associated with an increasingly high heat of respiration. It is therefore recommended that blackberries for fresh consumption be stored at 0–0.5 °C with a relative humidity of 90–95%, to avoid loss of moisture from dehydration.

Table 2 Typical vitamin, protein, and mineral contents of European blackberries

	Content (mg per 100 g)
Vitamins	
Carotene (total)	0.10–0.59
Thiamin	0.03
Riboflavin	0.034–0.038
Nicotinic acid	0.4
Ascorbic acid	20
Total nitrogen	181
Proteins	0.56
Amino acids	2.25
Minerals	
Ash	0.5
Phosphorus	23.8
Potassium	208.0
Sodium	3.7
Calcium	63.3
Magnesium	29.5
Iron	0.85
Copper	0.18
Sulfur	9.2
Chloride	22.1

Green A (1971) Soft fruits. In: Hume AC (ed.) *The Biochemistry of Fruits and their Products*. London: Academic Press.

Uses

In Europe, most blackberries are marketed as fresh fruit. Although consumers associate them with the autumn, it is interesting that the introduction of varieties of the western North American group that ripen in midsummer has also proved popular: clearly, blackberries can compete with other soft fruits if they are of a high quality. A proportion of the crop is also marketed as individual quick-frozen fruit, mostly for use in the processing and catering industries for pie fillings and flavorings, etc., though this usage is largely limited to North America, where the highly flavored varieties indigenous to the west coast are particularly popular for this purpose. Many of the boysenberries produced in New Zealand are exported in this way. A few blackberries are used for canning, but the large size of their pyrene ('seed') limits the fruit's use for jam manufacture, although significant amounts of seed-free blackberry conserve are produced. The introduction of very highly flavored cultivars with small seeds may change this.

Blackberries and tayberries are used to a small extent in Europe for the production of liquors and wines, for which the tayberry is well suited because of its rich purple color and excellent flavor. The Andean blackberry of Colombia has such a rich flavour that it is used to produce a popular breakfast fruit juice, but little blackberry juice is consumed elsewhere.

See also: **Ascorbic Acid**: Properties and Determination; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Jams and Preserves**: Chemistry of Manufacture

Further Reading

Green A (1971) Soft fruits. In: Hume AC (ed.) *The Biochemistry of Fruits and their Products*. London: Academic Press.

Jennings DL (1988) *Raspberries and Blackberries: their Breeding, Diseases and Growth*. London: Academic Press.

USDA (1984) *The Commercial Storage of Fruits, Vegetables, Flowers and Nursery Stock. US Department of Agriculture, Handbook 66*. Washington, DC: US Government Printing Office.

Blueberries *See* **Fruits of Temperate Climates**: Commercial and Dietary Importance; Fruits of the Ericacae; Factors Affecting Quality; Improvement and Maintenance of Fruit Germplasm

BODY COMPOSITION

J R Lustig and B J G Strauss, Monash Medical Centre, Clayton, Melbourne, Australia

Introduction

The assessment of nutritional status involves an understanding of the energy stores of the body. In developing countries, lack of food supply leads to wasting, whilst in developed countries obesity and cardiovascular disease are also important causes of illness. The prevention and management of these disorders require techniques for measuring the extent of these changes in individuals and in populations.

Investigation into body composition began in the nineteenth century. Chemical analyses of fetal and adult cadavers were followed by the development of concepts of different functioning body compartments. In 1906, Adolf Magnus-Levy coined the term fat-free mass. This was further refined in 1937, by Hastings and Eichelberger, who recognized that neutral fat does not bind water, nitrogen, or electrolytes.

Current concepts consider the composition of the human body at a number of different levels. These levels are (1) atomic; (2) molecular; (3) cellular, (4) tissue/organ; and (5) whole body. In the steady state, there are clear relationships within and between each of these levels. These categories also allow human body composition to be related to other areas of biology and clinical medicine. For example, the molecular level relates to endocrinology and biochemistry, the cellular level relates to physiology, the tissue/organ level relates to diagnostic imaging and pathology, and the whole-body level relates to bedside assessment and epidemiology.

Models of Composition of the Human Body

Atomic

The human body is composed of many different elements. The International Commission on Radiation Protection compiled a list of the multitude of elements found in the body of the adult male. These data were based on the chemical analysis of human organs by several investigators, and represent the best estimate of total body content based on a wide review of the literature. Thirty-six elements have been listed as being present in the human body. Approximately 18–20 of these elements have been shown to have a physiological function. The most abundant elements are carbon, hydrogen, oxygen, nitrogen, calcium, and phosphorus. Many factors will determine the amounts of a particular element in the body. Not only the health of the individual at the time of death, but also environmental exposure to an element will contribute to individual differences, for example, exposure to compounds containing lead or mercury, or exposure to radiation fall-out leading to accumulation of caesium and strontium. Gender, age, and size are also important determinants of the amount of differential elements in the body.

Molecular

Water is the most abundant molecule in the human body in health, comprising between 55 and 60% of body weight. It is obtained from food and fluid as well as being a metabolic byproduct of chemical reactions within living cells.

The next most abundant molecule is neutral fat or triglyceride which is stored within adipose cells but is also an essential component of all cell membranes. The fat compartment varies considerably in healthy individuals, but it has a relatively constant density of $0.9\,g\,cm^{-3}$ at 37 °C, and a very small water and potassium content.

Protein, either as a structural molecule or as a component of cellular cytoplasm, is produced under genetic control assuming an adequate food supply. Total-body protein is best assessed by extrapolation from the mass of elemental nitrogen in the body since human protein consists of about 16% nitrogen by weight. This is an example of a fixed relationship between the atomic level and the molecular level of measurement.

Glycogen represents a short-term carbohydrate storage molecule but is difficult to measure by any of the currently available body composition techniques.

There are many other different kinds of molecules in the body, for example DNA and RNA, but their whole-body mass is relatively small.

Cellular

The cellular level of body composition consists of body cells (body cell mass) and their surrounding extracellular water, plus the skeleton and connective tissue. Although there is some lipid in the form of cell membranes, this compartment is largely fat-free and these components are sometimes termed the fat-free mass (FFM) or in older terminology the lean body mass (LBM). The body cell mass is responsible for almost all of the basal energy expenditure of the body, since that is where cellular metabolic and respiration processes take place. Together with the adipose tissue compartment (which consists mostly of fat), this level is often referred to as a two-compartment model, i.e., FFM and fat mass (FM). In the healthy individual, the FFM has a relatively constant composition, with a water content of 72–74%, an average density of $1.1\,g\,cm^{-3}$ at 37 °C, a potassium content of $60–70\,mmol\,kg^{-1}$ in men and $50–60\,mmol\,kg^{-1}$ in women, and a protein content of 20%.

Tissue/Organ Level

More than 80% of the mass of the body is accounted for by skeletal muscle, adipose tissue, the skeleton, and blood volume. Of course, other organs such as the liver, kidneys, heart, and brain are of major importance. Most of these tissues can be imaged and their volume estimated. Magnetic resonance imaging (MRI) or computed tomography (CT) can be used to measure skeletal muscle mass, total fat mass, and its visceral fat subcompartment, and renal and liver volumes. The quality of muscle with respect to its fat content is also able to be estimated using these techniques. Echocardiography can estimate cardiac left ventricular mass, and dual-energy X-ray absorptiometry (DEXA) can measure bone mass and bone density. Dilution techniques can measure blood volume and red cell mass.

Whole-Body-Level

Clinical examination allows the measurement of anthropometric indices, such as height, weight, limb spans and circumferences, and truncal girths, as well as skinfold thicknesses. Using underwater weighing, another whole-body index, total-body density can be assessed.

Gender and Age Differences in Body Composition

Fetal growth rate reaches a maximum during the latter stages of pregnancy. Much of this growth is due to an increase in fat-free weight, consisting of water and protein as components of cell growth. Skeletal muscle contains about the same proportion of total-body nitrogen, potassium, and water in the newborn.

Growth velocity falls during infancy and childhood, reaching a nadir at about 4 years of age. A slow increase follows, culminating in the adolescent growth spurt. During this time there are differential rates of growth of different organs.

There is a small sex difference in FFM throughout infancy and childhood, but in the adolescent period the rate of increase in growth to adult levels of the FFM in males is approximately twice as much as in females. The male adolescent increase in FFM is also of longer duration. This is a function of hormonal action which continues to be an important determinant of body composition throughout life, together with genetic, nutritional, and activity influences.

Adult men have a greater FFM and less body fat then adult women. Women have approximately two-thirds the FFM of their male counterparts, whilst also having a greater relative fat mass. In young adults, the gender difference in FFM is greater then the difference in stature and in body weight.

There is much less variability between FFM values in individuals matched for age and height, compared with the variability in body fat.

The skeletal mass continues to grow until the late third decade. In women, there is an estrogen withdrawal-related decline in bone density over the period of the menopause. (*See* **Osteoporosis**.)

In old age, there is a decline in FFM, with a shift in body water towards the extracellular phase. Bone mass declines further in both sexes. Body fat content shows more variability than FFM, but generally increases with age, and accounts for most of the variance in weight. It also becomes more centrally or truncally distributed.

Fat cell numbers appear to be synthesized and turned over throughout life.

Methods of Measuring Body Composition (Table 1)

Anthropometry

Anthropometric techniques are readily portable and inexpensive. The equipment required includes a tape measure, height stick, scales, and skinfold callipers. Various formulae have been developed which allow the rapid calculation of different aspects of body composition, including percentage fat derived from triceps skinfold (Table 2), and the following anthropometric equations:

1. Body mass index (BMI) or Quetelet's index can be readily calculated from height and weight data: BMI = (weight in kilograms)/(height in meters)^2
2. arm muscle circumference = mid upper-arm circumference − ($n \times$ triceps skinfold)
3. abdominal circumference-to-gluteal circumference ratio
4. abdominal circumference

In populations, an increased BMI has a good correlation with body fatness, as measured by other techniques, and correlates well with morbidity and mortality in the obese individual. In underweight populations, a reduced BMI correlated well with loss of FFM. In individuals, the BMI should be treated as an index of nutrition with more caution, since other factors determining it, such as fitness or the presence of edema, are of nutritional importance.

Trunk circumferences define fat distribution. An abdominal circumference to-gluteal circumference ratio greater than 0.95 (males) and greater than 0.85 (females) is consistent with abdominal obesity, as is an abdominal circumference greater than 103 cm in males and 93 cm in females. The abdominal circumference is measured at the midpoint between the costal margin and the iliac crest in the midaxillary line. The gluteal circumference is measured at the maximal gluteal circumference.

Skinfold thicknesses have been used to measure body fat. This method assumes that subcutaneous fat measurements represent total body fat. Various sites can be assessed and equations applied to derive body density and hence subcutaneous fat mass. Durnin and Womersley developed the regression equation using four skinfolds (biceps, triceps, subscapular, and suprailliac), gender, and age. Equations have been developed using multiple or single skinfold sites.

Precision of skinfold measurement depends on the skill of the operator as well as the character of the subcutaneous fat. In general, the error is 5%, although this can be higher in the very obese individual.

Measurement of the midarm muscle circumference has been used to estimate total body protein stores. The triceps skinfold thickness by itself has been used as a measure of total body subcutaneous fat (Table 3).

Table 1 Summary of body composition measurement techniques

Level	*Method of measurement*	*Component*
Atomic level	Whole-body gamma counting	Total-body potassium
	Neutron activation analysis	Total-body nitrogen
Molecular	Mass spectrometry	Total-body water
	Bioelectrical impedance	
	Derived from total-body nitrogen	Total-body protein
Cellular	Derived from total-body potassium	Body cell mass
	Bromide dilution measured by HPLC	Extracellular water
	Dual-energy X-ray absorptiometry	Bone mass/density
Organ	Magnetic resonance imaging	Skeletal muscle mass
		Visceral fat mass
Whole-body	Anthropometry	Body mass index
	Underwater densitometry	Total body fat

HPLC, high-performance liquid chromatography.

Table 2 Percentage fat derived from triceps skinfold

Triceps skinfold (mm)	*Relative fat mass (%)*									
	Males (years)					*Females (years)*				
	17–19	*20–29*	*30–39*	*40–49*	*50 +*	*17–19*	*20–29*	*30–39*	*40–49*	*50 +*
5	8.0	10.0	18.0	16.5	18.5	12.5	9.5	13.0	15.5	16.0
7	11.5	13.5	20.0	20.0	23.0	16.5	14.0	17.0	20.0	20.5
9	14.5	16.0	22.0	23.0	26.0	19.5	18.0	20.5	23.0	24.5
11	17.0	18.0	23.5	25.5	29.0	22.0	21.0	23.0	26.0	27.5
13	19.0	19.5	24.5	27.7	31.0	24.5	23.5	25.5	28.0	30.0
15	20.5	21.0	25.5	29.5	33.0	26.0	25.5	27.5	30.0	32.5
17	22.0	22.5	26.5	31.0	35.0	28.0	27.5	29.0	32.0	34.0
19	23.5	23.5	27.0	32.5	36.0	29.0	29.5	30.5	33.5	36.0
21	25.0	24.5	28.0	33.5	37.5	30.5	31.0	32.0	35.0	37.5
23	26.0	25.5	28.5	35.0	39.0	31.5	32.5	33.5	36.0	39.0
25	27.0	26.5	29.0	36.0	40.0	33.0	33.5	34.5	37.5	40.5
27	28.0	27.0	30.0	37.0	41.0	34.0	35.0	35.5	38.5	41.5
29	29.0	28.0	30.5	38.0	42.0	35.0	36.0	36.5	39.5	43.0
31	30.0	29.0	31.0	38.5	43.0	35.5	37.0	37.5	40.5	44.0
33	30.5	29.5	31.0	39.5	44.0	36.5	38.0	38.5	41.5	45.0
35	31.0	30.0	32.0	40.0	45.0	37.5	39.0	39.5	42.5	46.0
37	32.0	30.5	32.0	41.0	45.5	38.0	40.0	40.0	43.0	47.0
39	32.5	31.0	32.5	41.5	46.5	38.5	41.0	41.0	44.0	47.0
40	33.0	31.5	33.0	42.0	47.0	39.0	41.5	41.0	44.5	48.0

Adapted from Durnin JVGA and Wormersley J (1974) Body fat assessed from body density and its estimation from skin fold thickness: Measurements on 481 men and women aged from 16 to 72 years. *British Journal of Nutrition* 32: 77–97.

Table 3 Error of the methods

Method	*Coefficient of variation (%)*
Anthropometry	1
Weight	1
Height	1
Circumferences	1
Skinfold thicknesses	3–6
Bioelectrical impedance	3–4
Dual-energy X-ray absorptiometry	
Bone density	1–2
Fat mass, fat-free mass	2–3
Deuterium dilution	1–2
In vivo *neutron activation*	
Total body protein	3–4
Total body potassium	3–4

Sources: unpublished data from the Body Composition Laboratory, Monash Medical Centre, Melbourne, Australia; Lohman (1981); Lukaski *et al*. (1985); Mazess *et al*. (1989); Mazess *et al*. (1990): see Further Reading for details.

The arm muscle circumference (or arm muscle and bone circumference) can be derived from the arm circumference and the triceps skinfold, and it gives a good indication of protein stores. This correlation holds true particularly in underdeveloped countries, where populations tend to have little subcutaneous fat, and it can be a useful tool in diagnosis and monitoring of progress in the management of protein-energy malnutrition.

In the nonambulant elderly, knee height can be used to predict stature.

Bioelectrical Impedance

Water and electrolyte distribution determine electrical conductance in the living organism. Virtually all the water and electrolytes in the body is found within the FFM which represents a low-resistance pathway. Fat and bone are poor conductors. Reactance is the opposition to the flow of electrical current caused by capacitance. The cell membrane, by maintaining an osmotic gradient between extra- and intracellular compartments, serves as a capacitor. Reactance is a measure of the quantity of cell membrane capacitance and may give an indication of the quantity of the intracellular cell mass. Whereas fat and water offer resistance to an electric current, only cell membranes have reactance.

Application of a constant, low-level alternating current to the body can be used principally to determine total-body water (TBW) and, by regression analysis from other techniques, to determine FM and FFM. Both resistance and reactance are measured.

The measurement of total-body impedance (resistance and reactance) is a vector sum of resistance and reactance in the limbs and torso, with the limbs making the major contribution.

Use of this method has been validated in healthy populations as a determinant of TBW and deduced lean and fat mass. In disease states such as renal failure and dehydration, metabolic function may alter the compartmentation of TBW and bioelectrical

impedance may be less useful and less reproducible, with the literature reporting conflicting results.

The advantage of bioelectrical impedance is that it is cheap, portable, and simple to use.

Heavy-Water (D_2O) Dilution

Water is not present in stored triglyceride and occupies a relatively fixed fraction of the FFM (73%). Estimation of total-body water (TBW) can therefore be used as an index of body composition. Several isotopes have been used, but deuterium oxide (D_2O), the naturally occurring nonradioactive isotope of water, containing hydrogen protons with two neutrons – heavy water – and present as 0.01% of naturally occurring water, is now seen as the gold standard for measurement of TBW.

The technique involves the administration of a known quantity of D_2O, an equilibrium period, and a sampling period. It assumes that the D_2O has the same distribution volume as water and is exchanged by the body in a manner similar to water.

Sampling can be from either serum or saliva and, whilst the analytical equipment is only suited to a laboratory, it enables collection of specimens in the field for later analysis.

A variety of analytical techniques have been used to measure D_2O, including mass spectrometry and Fourrier transform infrared spectroscopy.

Whole-Body Densitometry

Whole-body densitometry is the gold standard for the measurement of body fat. The technique assumes that the body is composed of two distinct compartments (fat and fat-free, each of a known or assumed density) and that the relative amount of each can be determined by measurement of the whole-body density. Underwater weighing is the most widely used technique, based on Archimedes' principle which states that the volume of an object submerged in water is equal to the volume of water that the object displaced. The mass in air of the object and the mass in water of the object is then converted to a total-body density. Body fat mass can then be calculated using one of the empirical equations describing the relationship between fat content and body density.

Valuable body composition data can be obtained using underwater weighing, although there are several inherent disadvantages. Subjects must be accustomed to swimming and submersion in water and must be medically fit enough to endure such procedures. These caveats to its use exclude use of this technique in many hospitalized patients. The apparatus required is substantial in size, and thus only suited for use in large institutions. In addition, variation in bone density owing to ethnicity, gender, or aging is not taken into account in the constant used for nonfat density.

A recent development is whole-body air plethysmography which allows estimation of body volume without underwater immersion and may allow the technique of whole-body densitometry to become available on a wider scale.

In Vivo Neutron Activation Analysis

In vivo neutron activation analysis (IVNAA) provides the only *in vivo* technique currently available to determine simultaneous multielemental composition. Calcium, phosphorus, sodium, and chloride content have been measured, but in current clinical practice only nitrogen is measured (from which total-body protein is calculated).

Neutron activation involves delivery of a beam of neutrons to the subject. These neutrons are captured by the target atoms in the body, creating unstable isotopes; in the case of protein, the isotope formed is ^{15}nitrogen. The isotope reverts to a stable state by the emission of γ-rays of a characteristic energy, which can then be detected by the use of standard γ-spectrographic analysis. This method, targeting nitrogen, allows: (1) the determination of total-body nitrogen and, therefore, total-body protein, which is the principal nitrogen-containing component of the body; and (2) in combination with total-body potassium measurements, the indirect determination of skeletal muscle mass.

Whilst IVNAA is a very useful tool for measurements of body composition, its development in only a few centers has limited the wider application of this methodology. In addition, concern has been expressed over the large expense of this technique.

Total-Body Potassium

The naturally occurring radioactive isotope of potassium, ^{40}K, is present in a known, constant, very low percentage of total potassium. Since body potassium is essentially intracellular, and not present in stored fat, measurement of ^{40}K not only provides an estimate of total-body potassium, but also allows estimation of body cell mass. If total body water is known (from deuterium dilution), extracellular water can be calculated.

The technique requires a highly shielded environment in which to detect the ^{40}K in the body, as ^{40}K also occurs in most environmental structures. In addition, the requirement for appropriate γ-ray detectors and corrections for factors such as body geometry makes the technique expensive.

Dual-Energy X-Ray Absorptiometry

DEXA is a recent addition to the body composition analysis field. It was originally developed to measure regional and total-body bone mineral content, but it is also capable of measuring fat mass. It is more sensitive than dual-photon absorptiometry, which it has now replaced, and exposes the subject to substantially lower radiation doses than total-body calcium measured by neutron activation analysis.

The DEXA technique exposes the subject to low energy irradiation at two different energies. As there is differential absorption by tissues of different densities (bone, lean and fat tissues), values for bone mineral in the hip, spine, whole body, or specialized regions, as well as values for fat or soft tissues, can be derived.

It is a sensitive technique for determining bone mineral content and densities, and has become the gold standard for clinical and research work in osteoporosis.

Because the fat-free nonbone component of the limbs effectively measures limb skeletal muscle mass, of which the proportion to whole-body skeletal muscle is fairly constant, DEXA can be used to measure skeletal muscle mass.

Other Techniques

Other techniques for measuring body composition include CT scanning and MRI. These techniques have been used for accurate measurement of various body compartments. These compartments include visceral adipose tissue mass, increased amounts of which are related to risk for diabetes and vascular disease, and skeletal muscle volume. Whilst their clinical and research use is limited by expense, availability, and, in the case of CT scanning, patient exposure to ionizing radiation, much helpful information has been obtained from the research groups who have utilized such technology.

Error of the Methods

Each of the methods for measuring body composition has intrinsic and biological errors. In general, these are very comparable to the errors of biochemical methods commonly performed in hospital laboratories (Table 3).

Effect of Disease Processes on Human Body Composition

Obesity

The fat compartment – expressed as mass or percentage fat – has importance as an expression of adiposity. The association between increased adiposity and morbidity is well documented. The incidence of diabetes, hypertension, ischemic heart disease, and gallbladder disease is increased at higher levels of adiposity.

Fat distribution is being increasingly recognized as a clinically relevant risk factor. Abdominal obesity, expressed as the abdominal circumference-to-gluteal-circumference ratio, has been shown, independent of other factors, to represent an increased risk for morbidity and mortality. First described in the 1950s, this association has been verified and strengthened by intense research interest in the last decade. More recently, the abdominal circumference alone has also been shown to predict morbidity and mortality. The correlation is not just an epidemiological one: many metabolic abnormalities, such as insulin resistance, are associated with this condition.

With increasing adiposity, it is usual to find an increase in lean mass. This occurs because of the increased skeletal muscle mass required to carry the increased fat mass.

In obese people who have undergone repeated near-starvation dieting, there may be marked wasting of skeletal muscle, even in the presence of adiposity. Similarly, muscle wasting, or even marasmus, can occur in the obese person who has concomitant severe illness. For example, a chronic alcoholic with a poor quality of food intake and liver impairment may be obese (excess fat stores) and also have skeletal muscle wasting.

Undernutrition

In global terms, undernutrition remains one of the greatest determinants of health status. The most common expressions of undernutrition are commonly known as kwashiorkor and marasmus. Intermediate forms are commonly seen. In marasmus, there is a wasting of total-body protein without expansion of the TBW compartment. In kwashiorkor, there is an expansion of the extracellular component of TBW, giving rise to peripheral edema and ascites. Exactly why undernutrition results in these two different forms remains unclear.

Both forms are associated with a reduction in the capacity of the cell-mediated immune system of the body, giving rise to an increased risk of infection. Where food intake is reduced, it is likely that other nutrient deficiencies, such as iron, folate, or vitamin A deficiency, will be present and may mask the extent of the underlying body composition changes.

In developing countries, these forms of undernutrition usually arise from the combination of inadequate food resources and chronic infection. In malnutrition

caused by inadequate food intake, many studies have shown an increase in the ratio of extracellular fluid volume to TBW. Electrolyte abnormalities are common. Reduced levels of sodium, potassium, and magnesium may be found. Protein stores, as measured by IVNAA and expressed as nitrogen index, are reduced.

In developed countries, these conditions are frequently seen in conjunction with malignancy, psychiatric conditions, organ failure, and conditions in which self-feeding is difficult, e.g., stroke or arthritis. (*See* **Malnutrition**: Malnutrition in Developed Countries.)

The wasting seen in many malignancies, even when nutritional intake and absorption is apparently adequate, may be caused by cytokines such as cachectin and tumor necrosis factor.

Visceral mass may be better preserved in cancer patients, suggesting that loss of muscle accounts for the major proportion of weight loss. This may be the result of a difference in metabolic rate between malignancy and anorexia; the rate is often increased in patients with malignancy, but reduced in anorexia nervosa. (*See* **Anorexia Nervosa**.)

In anorexia nervosa, many of the body composition changes seen are similar to those found in starved subjects. Total-body nitrogen, total-body potassium, and blood volume are all reduced. Unlike primary starvation, extracellular water may be reduced by relatively greater amounts than the reduction in TBW owing to induced vomiting or purging. Interestingly, many patients with anorexia nervosa maintain normal levels of hemoglobin and serum albumin and rarely exhibit vitamin deficiencies or develop edema.

Conclusion

An understanding of body composition and its measurement provides the clinician with further scientific data on which to base a nutritional assessment. Advances in this area are now available to provide accurate measurements which previously could only be estimated. Some of these techniques are easily applicable to office general practice or field studies, whilst others require more sophisticated equipment which is only available in specialized centres. Body composition changes in health and disease help to explain the pathogenesis of illnesses which involve alteration in food intake or absorption and metabolism.

See also: **Anorexia Nervosa**; **Malnutrition**: Malnutrition in Developed Countries; **Osteoporosis**

Further Reading

Alleyne GEO (1975) Mineral metabolism in protein-calorie malnutrition. In: Olsen RE (ed.) *Protein-Calorie Malnutrition*, pp. 201–208. New York: Academic Press.

Brozek J, Grande F, Anderson IT and Keys A (1963) Densitometric analysis of body composition: revision of some quantitative assumptions. *Annals of the New York Academy of Sciences* 110: 113–140.

Durnin JVGA and Womersley J (1974) Body fat assessed from body density and its estimation from skin fold thickness: measurements on 481 men and women aged from 16 to 72 years. *British Journal of Nutrition* 32: 77–97.

Forbes GB (1987) *Human Body Composition*, pp. 2–171. New York: Springer-Verlag.

Forbes GB (1990) Body composition. In: Brown M (ed.) *Present Knowledge in Nutrition*, 6th edn. Washington, DC: International Life Sciences Institute.

Forbes GB, Kriepe RE, Lipinski BA and Hodgman CH (1984) Body composition changes during recovery from anorexia nervosa: comparison of two dietary regimes. *American Journal of Clinical Nutrition* 40: 1137–1145.

Fowler PA, Fuller MF, Glasby CA *et al*. (1991) Total and subcutaneous adipose tissue in women: the measurement of distribution and accurate prediction of quality by using magnetic resonance imaging. *American Journal of Clinical Nutrition* 54: 18–25.

Hastings AB and Eichelberger L (1937) The exchange of salt and water between muscle and blood. *Journal of Biological Chemistry* 117: 73–93.

International Commission on Radiological Protection(1975) *Report of the Task Group on Reference Man*, no. 23. Oxford: Pergamon Press.

Jelliffe EPP and Jelliffe DB (1969) The arm circumference as a public health index of protein-calorie malnutrition of early childhood. *Journal of Tropical Paediatrics* 15: 179–192.

Kvist H, Choudury B, Grangard U, Tyler U and Sjostrom L (1988) Total and visceral adipose tissue volumes derived from measurements with computed tomography in adult men and women: predictive equations. *American Journal of Clinical Nutrition* 48: 1351–1361.

Lohman TG (1981) Skin-folds and body density and their relation to body fatness: a review. *Human Review* 53: 181–225.

Lukaski HC (1987) Methods for the assessment of human body assess human body composition. *Journal of Applied Physiology* 60: 1327–1332.

Lukaski HL, Johnson PE, Bolonchuk WW *et al*. (1985) Assessment of fat-free mass using bioelectrical impedance measurements of the human body. *American Journal of Clinical Nutrition* 41(4): 810–817.

Mazess RB, Collick B, Trempe J, Barden HS and Hanson JA (1989) Performance evaluation of a dual-energy x-ray bone densitometer. *Calcified Tissue International* 44: 228–232.

Mazess RB, Barden HS, Bisek JP and Hanson J (1990) Dual energy X-ray absorptiometry for total body and regional

bone mineral and soft tissue composition. *American Journal of Clinical Nutrition* 51: 1106–1112.
Mazess RB, Hanson JA, Trempe J and Bonnick SL (1991) Ultrasound Measurement of the Os Calcis. Proceedings of the 8th International Workshop on Bone Densitometry, 28 April, Bad Reichenhall, FRG.
Ross R, Goodpaster B, Kelley D and Fernando B (2000) Magnetic resonance imaging in human body composition research: from quantitative to qualitative tissue measurement. *Annals of the New York Academy of Science* 904: 12–17.
Stroud DB, Borovnicar DJ, Lambert JR *et al.* (1990) Clinical studies of total body nitrogen in an Australian hospital. In: Yasurnura S (ed) *In vivo Body Composition Studies, Basic Life Sciences*, vol. 55, pp. 177–182.New York: Plenum Press.
Vague J (1956) The degree of masculine differentiation of obesities: a fact for determining predisposition to diabetes, atherosclerosis, gout, uric calculus disease. *American Journal of Clinical Nutrition* 4: 20–34.
Wang ZM, Pierson RN and Heymsfield SB (1992) The five level model: a new approach to body composition research. *American Journal of Clinical Nutrition* 56: 19–28.
Widdowson EM and Dickerson JWT (1964) Chemical composition of the body. In: Comari CL and Bronner F (eds) *Mineral Metabolism*, vol. 2, part A, pp. 2–247.London: Academic Press.

BONE

K D Cashman, University College, Cork, Ireland
F Ginty, MRC Human Nutrition Research, Cambridge, UK

Introduction

Bone is a specialized connective tissue that makes up, together with cartilage, the skeletal system. The skeleton serves three important functions: (1) mechanical; support and site of muscle attachment for locomotion; (2) protective; for vital organs and bone marrow; and (3) metabolic, as a reserve for ions, especially calcium and phosphate; for the maintenance of serum homeostasis, which is critical to life.

In bone, as in all connective tissues, the fundamental constituents are the cells and the extracellular matrix (the osteoid). Osteoid is particularly abundant in bone and is composed of collagen fibers and noncollagenous proteins. Bone is distinguished from other forms of connective tissue by the fact that it becomes extremely hard, owing to the deposition within a relatively soft organic matrix of a complex mineral substance, largely composed of calcium, phosphate, and carbonate. Bone is built up during youth to a peak reached probably in the 20s. It then diminishes again, first slowly, then more rapidly, especially in women after the menopause, and in both sexes in the elderly. The regulation of this skeletal growth and development is complex and involves genetic, mechanical, hormonal, and nutritional influences. At a cellular level, bone metabolism or remodeling is coordinated by a multiplicity of interacting local hormones, cytokines, and growth factors. Such bone-associated factors may play a role in the pathophysiology of bone disorders.

Anatomical Features of Bone

Anatomically, two types of bone can be distinguished in the skeleton: flat bones (i.e., the skull bones, scapula, mandible, and ilium) and long bones (tibia, femur, humerus, etc.). External examination of a long bone (**Figure 1**) shows the two wider extremities (the epiphyses), a more or less cylindrical tube in the middle (the diaphysis or midshaft), and a development zone between them (the metaphysis). In a growing bone, the epiphysis and the metaphysis, which originate from two independent ossification centers, are separated by a layer of cartilage, the epiphyseal cartilage or growth plate, which has an important role in the longitudinal growth of bones. The external part of bone is formed by a thick and dense layer of calcified tissue, the cortex or compact bone, which in the diaphysis encloses the medullary cavity where the hematopoietic bone marrow is housed. Cortical bone makes up about 80% of the total skeleton. Toward the metaphysis and the epiphysis, the cortex becomes progressively thinner, and the inner space is filled with a network of thin, calcified trabeculae: this is cancellous bone, also named spongy or trabecular bone. The spaces enclosed by these thin trabeculae also are filled with hematopoietic bone marrow and are in continuity with the medullary cavity of the diaphysis. Cancellous bone is also found in the vertebrae and the majority of flat bones. It makes up about 20% of the total skeleton. The relative proportions of cortical and cancellous bone differ at different sites in

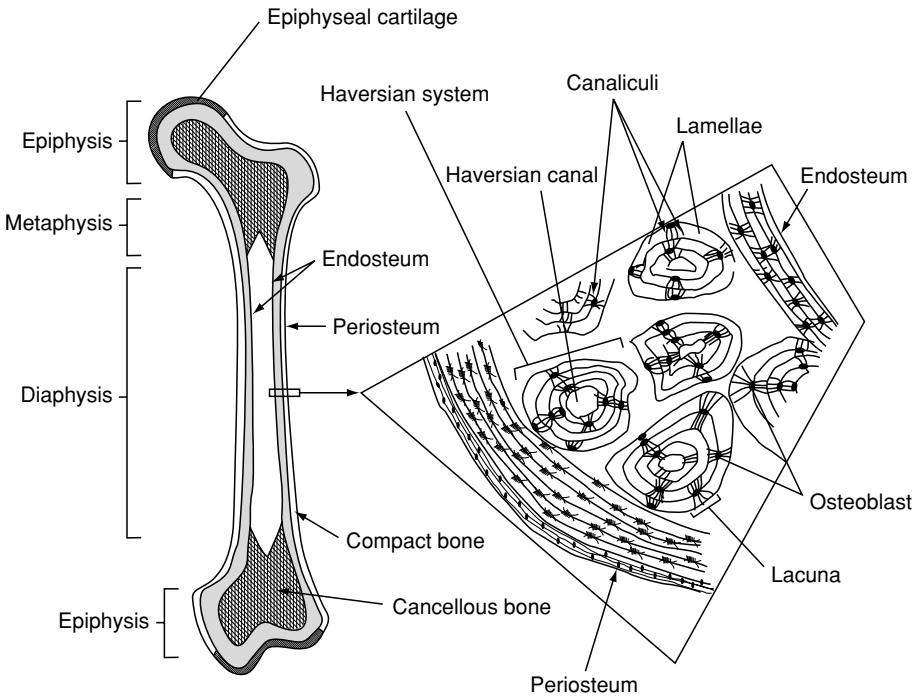

Figure 1 Schematic view of a longitudinal section through a long bone (insert; transverse section through the diaphysis).

the skeleton (e.g., cancellous bone comprises >66% of the total bone in the lumbar spine, but only 25% of the total bone in the neck of the femur). The bone surfaces at the epiphyses that take part in the joint are covered with a layer of articular cartilage that does not calcify.

There are two bone surfaces at which the bone is in contact with the soft tissues (Figure 1): an external surface (the periosteal surface) and an internal surface (the endosteal surface). The surfaces are lined with osteogenic cells organized in layers, the periosteum and the endosteum. Cortical and trabecular bone are constituted of the same cells and the same matrix elements, but there are structural and functional differences. The primary structural difference is quantitative: 80–90% of the volume of compact bone is calcified, whereas only 15–25% of the trabecular bone is calcified (the remainder is occupied by bone marrow, blood vessels, and connective tissue). The result is that 70–85% of the interface with soft tissues is at the endosteal bone surface, which leads to the functional difference: the cortical bone fulfills mainly a mechanical and protective function, while the trabecular bone has a metabolic function. Therefore, at a macroscopic level, bone can be regarded as an outer cortical sheath and an inner three-dimensional trabecular network, which together allow optimal mechanical and biochemical function with minimal weight.

Bone Matrix and Mineral

Chemically, bone is made up of one-third protein matrix and two-thirds mineral. Bone matrix is formed by type I collagen fibers (which make up 90% of the total protein of bone) and numerous noncollagenous proteins (which make up the remaining 10% of total protein of bone; Table 1). Type I collagen is a highly cross-linked fibrillar protein which, through its tridimensional structure similar to rope, gives bone its tensile strength. The functions of the noncollagenous proteins are diverse and some are still poorly understood (Table 1). The urinary excretion and the plasma or serum concentrations of some of these components of the matrix are used chemically to assess bone metabolism (Table 1). Spindle- or plate-shaped crystals of hydroxyapatite $Ca_{10}(PO_4)_6(OH)_2$ are found on and within the collagen fibers. They tend to be oriented in the same direction as the collagen fibers.

The preferential orientation of the collagen fibers alternates in adult bone from layer to layer, giving to this bone a typical lamellar structure. The lamellae can be parallel to each other if deposited along a flat surface (trabecular bone and periosteum), or concentric if deposited on a surface surrounding a channel centered on a blood vessel. In cortical bone, the tissue is mainly organized as Haversian systems which represent its basic structural building blocks (Figure 1).

Table 1 Principal noncollagenous proteins in the bone matrix

Protein type/name	*Potential function(s)*
Serum proteins	
Albumin	Transports proteins; inhibits hydroxyapatite crystal growth
α-2HS glycoprotein	Promotes endocytosis; chemoattractant for monocytic cells; mineralization inhibitor
Glycoproteins	
Alkaline phosphatase[a]	A phosphotransferase; potential Ca^{2+} carrier; hydrolyzes inhibitors of mineral deposition
Osteonectin	May mediate deposition of hydroxyapatite; binds to growth factors; may influence cell-cycle antiadhesive protein
Tetranectin	Binds to plasminogen; may regulate matrix mineralization
RGD-containing glycoproteins	
Thrombospondins (I,II,III,IV)	Cell attachment
Fibronectin	Binds to cells
Vitronectin	Cell-attachment protein
Osteopontin	Binds to cells; inhibits mineralization; may regulate proliferation; inhibits nitric oxide synthase; may regulate tissue repair
Bone sialoprotein[a]	Binds to cells; binds Ca^{2+} with high affinity; may initiate mineralization
BAG-75	Binds to Ca^{2+}; may act as a cell-attachment protein; may regulate bone resorption
γ-Carboxy glutamic acid-containing proteins	
Matrix Gla protein	May function in cartilage metabolism, may inhibit mineralization
Osteocalcin[a]	May regulate activity of osteoclasts and their precursors; may mark turning point between bone formation and resorption; regulates mineral maturation
Protein S	Protein S-deficiency may result in osteopenia
Glycosaminoglycan-containing	
Veriscan	May 'capture' space that is destined to become bone,
Decorin	Binds to collagen and may regulate fibril diameter; binds to TGF-β; inhibits cell attachment to fibronectin
Biglycan	May bind to TGF-β
Fibromodulin	Binds to collagen; may regulate fibril formation; binds to TGF-β
Osteoadherin	May mediate cell attachment
Hyaluroan	May work with versican-like molecule to capture space destined to become bone

[a]Their plasma or serum concentrations are used chemically to assess bone metabolism.
RGD, Arg-Gly-Asn; TGF-β, transforming growth factor-β.
Adapted from Lian JB, Stein GS, Canalis E, Robey PG and Boskey AL (1999) Bone formation: osteoblast lineage cells, growth factors, matrix proteins, and the mineralization process. In: Murray FJ (ed.) *Primer on the Metabolic Bone Diseases of Mineral Metabolism*, 4th edn, pp. 14–30. Philadelphia, USA: Lippincott/Williams & Wilkins.

Each Haversian system is made up of a central canal, the Haversian canal, which contains an artery, a vein, a lymphatic vessel, a nerve, and some bone cells. This is surrounded by successive rings of lamellae. Lying between the lamellae are a number of small cavities, the lacunae, which contain bone cells, or osteocytes. Fine protoplasmic processes from the osteocytes branch profusely throughout the matrix in very small canal-like structures, the canaliculi (Figure 1). These serve to interconnect the lacunae. Lamellar bone is the form present in adult cortical and cancellous bone. However, when bone is being formed very rapidly (during embryonic development and fracture healing), there is no preferential organization of the collagen fibers. They are not so tightly packed and found in somewhat randomly oriented bundles. This type of bone is called woven bone. This immature bone is usually replaced by lamellar bone, so that it is practically absent from the adult skeleton, except for some that may persist near tendon insertions and ligament attachments.

Bone Cells

Bone is a dynamic connective tissue, comprising an exquisite assembly of functionally distinct cell populations that are required to support its structural, biochemical, and mechanical integrity and its central role in mineral homeostasis. Bone tissue (both cortical and cancellous bone) is constantly turned over by two cell-mediated processes, bone modeling and bone remodeling. In the former, new bone is formed at a location different from the one where it was destroyed (resorbed). This allows bone to elongate and widen, altering the shape of the skeleton. This change takes place principally in the child and adolescent, allowing the development of a normal architecture during growth. It is also the process by which vertebrae increase in size during life. During adulthood, bone remodeling is the primary process of bone turnover, however, modeling does continue to take place at certain skeletal sites such as the periosteum. The physiological repair of fractured bone is also

part-achieved by modeling. Bone remodeling involves resorption and formation of bone and takes place at the same site (known as the bone-remodeling unit), leading to no change in bone shape, but allowing bone rejuvenation, adaptation to stress and strains, the repair of microfractures and therefore the maintenance of mechanical integrity. In addition, bone remodeling has an important role in the maintenance of calcium homeostasis. This bone cell-mediated process replaces 2–10% of the adult skeleton per annum.

The principal cells that mediate the bone-forming processes of the mammalian skeleton are the osteoblasts that synthesize the bone matrix on bone-forming surfaces; osteocytes, organized throughout the mineralized bone matrix, that supports bone structure; and the protective bone-surface lining cells. The principal cells that mediate the bone-resorbing process are the multinucleated giant cells, known as osteoclasts.

Osteoblasts

Osteoblasts are the cells responsible for the synthesis and mineralization of type I collagen. They originate from pluripotent mesenchymal stem cells which, by following a developmental pathway, they initially differentiate into preosteosblasts, and then to mature osteoblasts. A mature osteoblast which is actively synthesizing bone matrix has many distinctive ultrastructural characteristics, including a large nucleus, enlarged Golgi apparatus, and extensive endoplasmic reticulum. It also has unique biochemical characteristics, including high levels of type I collagen, alkaline phosphatase, and osteocalcin (Table 1). These cells form an epithelial-like structure at the surface of bone where they unidirectionally secrete the osteoid. In a second step, this matrix than calcifies extracellularly; the role of the osteoblast in this process is still unclear.

Osteocytes

The osteocytes are the most abundant cells in the bone matrix, with about 25 000 cells per mm^3 bone. However, despite their abundance, their function is still poorly understood. They originate from mature osteoblasts which become embedded in the bone matrix after it becomes mineralized. Osteocytes are physically different from osteoblasts and have long protoplasmic extensions which permeate the bone matrix and allow them to interconnect and communicate with other osteocytes and osteoblasts. Recent research has shown that osteocytes may act as sensors for stresses and strains within the bone matrix and thus may be important in the adaptation of bone to weight-bearing exercise.

Bone-Lining Cells

Bone-lining cells are osteoblasts which are not actively synthesizing bone and thus are known as resting osteoblasts or surface osteocytes. They occupy about 80% of the skeletal surface where they act as a kind of blood–bone barrier. It is thought that they are essential for the maintenance of blood calcium levels, perhaps by actively pumping calcium ions from the bone fluid compartment to the extracellular fluid compartment.

Osteoclasts

Osteoclasts originate from a lineage different from that of other bone cells, namely the hemopoietic system, more specifically from the granulocyte–macrophage colony-forming unit for granulocytes and macrophages (CFU-GM). Before developing into mature resorbing osteoclasts the CFU-GM cells undergo proliferation and further differentiation, after which they migrate to the bone surface where they fuse to become large multinucleated cells. These large, motile, multinucleated osteoclasts, situated either on the surface of cortical or trabecular bone, or within the cortical bone, where they are located at the tip of the remodeling units, are responsible for the resorption of bone at these sites.

Osteoclast differentiation and activation appear to be primarily regulated by the osteoblast and its precursors. Active osteoclasts resorb bone in a closed, sealed-off microenvironment. This is formed by attachment of the osteoclast to the mineralized surface by a marginal rim of contractile proteins, called clear zone or sealing zone. This attachment involves cell membrane receptors, known as integrins, which recognize specific peptide sequences in the matrix. The osteoclast surrounds a chamber lined with membrane folds (the ruffled border) and secretes self-generated protons (via a proton adenosine triphosphatase) which then dissolve bone mineral. Proteolytic enzymes, especially cathepsins (cathepsin K) and collagenases, are also synthesized by the osteoclast and are secreted through the ruffled border into the extracellular bone-resorbing compartment where they digest collagen. Recent evidence suggests that the osteoclast undergoes apoptosis (programmed cell death) after a cycle of resorption.

Integrated Activity of Bone Cells (Bone-Remodeling Unit)

The bone-remodeling process represents the coordinated actions of osteoclasts and osteoblasts on trabecular bone surface and in Haversian systems. Bone remodeling occurs in the bone-remodeling units,

which are focal and discrete packets of bone throughout the skeleton. The remodeling of each packet takes a finite period (taking up to 6 months but differing in cortical and cancellous bone, and probably longer in cancellous bone). The sequence of cellular activity – the bone-remodeling cycle – within every bone-remodeling unit is always identical: (1) activation of osteoclast precursors (possibly by a member of the tumor necrosis factor (TNF) ligand family, known as RANK-L (receptor activator for NF-κB ligand) released by the osteoblast); and then (2) osteoclastic bone resorption (over a period of 20 days), followed by: (3) osteoblastic bone formation to repair the defect (Figure 2). However, the remodeling that occurs in each packet is geographically and chronologically separated from other packets of remodeling. It is believed that this can occur because the activation of the sequence of cellular events responsible for remodeling is locally controlled, possibly by local mechanisms in the bone microenvironment. In a healthy adult, it has been estimated that 12 new bone-remodeling units are started every minute in trabecular bone and 3 per minute for cortical bone.

Coupling of Bone Formation and Bone Resorption

In most circumstances, the coupling of bone formation to previous bone resorption occurs faithfully. Packets of bone removed during resorption are replaced during formation. The cellular and humoral mechanisms responsible for mediating this coupling process (or disrupting it, as in some bone diseases) are still not clear but may involve an osteoblast-stimulating factor (such as insulin-like growth factor (IGF)-I, IGF-II, or transforming growth factor-β).

Remodeling Process: Aging and Disease

With aging there is a bone-remodeling imbalance, with the rate of resorption exceeding that of formation. Bone formation rates have been shown to be dramatically decreased with age and the percentage of bone surface with no activity increases and osteoblasts disappear. In cortical bone, the 'physiologic' imbalance between the two processes leads to an increased porosity which, in turn, is followed by mobilization of bone mineral and a decrease in bone mass. In trabecular bone there are normal or deeper resorption cavities that are incompletely filled with new bone by the osteoblasts, leading to a thinning of the trabeculae, and, in some cases, a perforation of the trabeculae. This will remove the structural basis for the following formative period and the final result will be a hole in the trabecular network, a loss of bone and bone mineral, and a reduced strength of the remaining bone. Therefore, an imbalance of the remodeling process reduces bone mass and causes detrimental architectural changes and increases fracture risk.

Therefore, the balance between bone formation and bone resorption, as effected through cellular

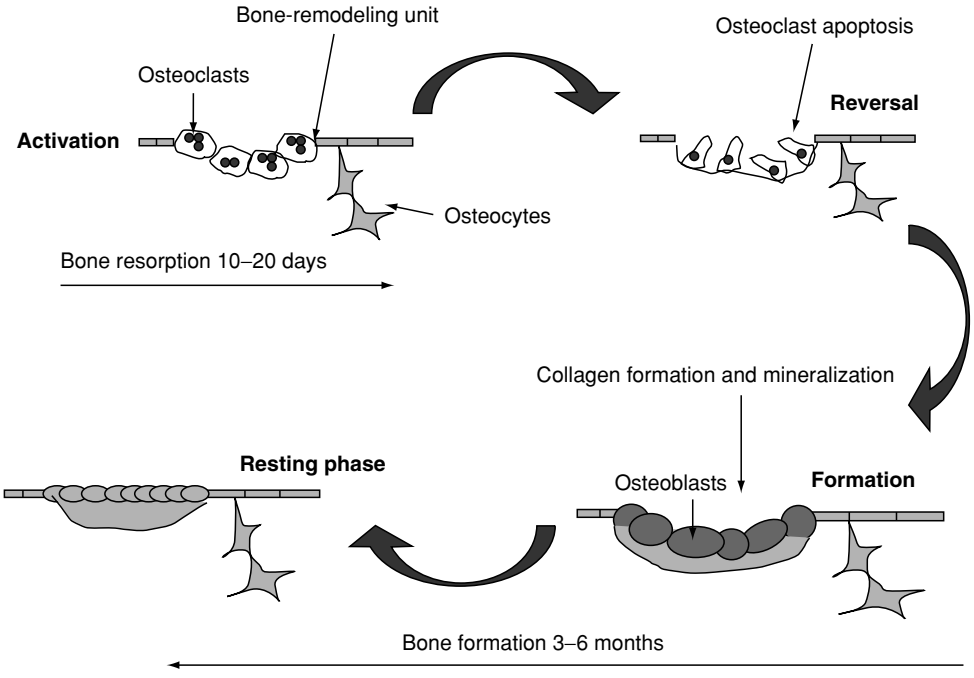

Figure 2 Bone-remodeling cycle.

events, is an important determinant of bone mass of an individual at any stage of the life cycle. In addition, all of the diseases of bone are superimposed on this normal cellular-remodeling sequence. In diseases such as primary hyperparathyroidism, hyperthyroidism, and Paget's disease, in which osteoclasts are activated, there is a compensatory and relatively balanced increase in the formation of new bone. In elderly patients with osteoporosis, there is a decrease in mean wall thickness, presumably reflecting the inability of osteoblasts to repair adequately the resorptive defects made during normal osteoclastic resorption.

Regulation of Bone Cell Activity

Bone remodeling is regulated by a variety of systemic hormones and by local regulatory factors (Table 2) that affect cells of the osteoclast and osteoblast lineage and exert their effects on: (1) the replication of undifferentiated cell; (2) the recruitment of cells; and (3) the differentiated function of cells.

Table 2 Systemic hormonal and local regulation of bone remodeling

Type of regulatory factor/name	*Effect(s)*
Systemic hormones	
Parathyroid hormone (PTH)	Stimulates differentiation of committed progenitors to form mature osteoclasts; activates preformed osteoclasts to resorb bone
Parathyroid hormone-related protein (PTHrP)	Effects identical to those of PTH on osteoclasts
1,25 $(OH)_2D_3$	Potent stimulator of osteoclastic bone resorption
Calcitonin (CT)	Potent inhibitor of osteoclastic bone resorption (transient effect); causes cytoplasmic contraction of the osteoclast cell membrane; causes dissolution of mature osteoclasts into mononuclear cells; it can also inhibit osteoclast formation, inhibiting both proliferation of the progenitors and differentiation of the committed precursors
Insulin	Stimulates bone matrix synthesis and cartilage formation; necessary for normal bone mineralization
Growth hormone (GH)	May regulate bone formation
Glucocorticoids	May regulate bone resorption
Sex steroids	Estrogen deficiency is associated with increased osteoclastic bone resorption (either via a direct or indirect mechanism)
Thyroid hormones	May stimulate osteoclastic bone resorption
Local factors	
Interleukin-1 (α, β)	Potent stimulators of osteoclasts; stimulates the proliferation of the progenitors and differentiation of committed precursors into mature cells; activation of mature multinucleated osteoclasts
Tumor necrosis factor (TNF)	Stimulates the proliferation of osteoclast progenitors, causes fusion of committed precursors to form multinucleated cells, and activates multinucleated cells to resorb bone.
Colony-stimulating factor 1	Mediates osteoclast formation
Interleukin-18	Inhibitor of osteoclast formation
Osteoclastogenesis-inducing differentiation factor (ODIF)	Mediator of osteoclastic bone resorption
Interleukin-6	Weak stimulator of osteoclast formation
Interferon-γ	Inhibitor of osteoclastic bone resorption; inhibits differentiation of committed precursors to mature cells
Transforming growth factor-β	Inhibits osteoclast formation by inhibiting both the proliferation and differentiation of osteoclast precursors; directly inhibits the activity of mature osteoclasts; stimulates osteoblast proliferation and synthesis of differentiated proteins; increased mineralized bone formation
Insulin-like growth factors (I, II)	May enhance bone collagen and matrix synthesis and stimulate the replication of cells of the osteoblast lineage
Fibroblast growth factors (1, 2)	Increases bone cell population capable of synthesizing bone collagen
Platelet-derived growth factors	Stimulates bone collagen synthesis
Retinoids	Stimulatory effect on osteoclasts
Prostaglandins	Complex and multiple effects on osteoclasts; prostaglandins of the E series may stimulate osteoclastic bone resorption; may act as mediators of pro-bone-resorptive growth factors
Leukotrienes	Linked to osteoclastic bone resorption

Adapted from Lian JB, Stein GS, Canalis E, Robey PG, and Boskey AL (1999) Bone formation: osteoblast lineage cells, growth factors, matrix proteins, and the mineralization process. In: Murray FJ (ed.) *Primer on the Metabolic Bone Diseases of Mineral Metabolism*, 4th edn, pp. 14–30. Philadelphia, USA: Lippincott/Williams & Wilkins.

Systemic regulation of bone turnover is controlled by numerous hormones, including the mineral homeostasis hormones (parathyroid hormone, calcitonin and 1,25-dihydroxvitamin D; (*See* **Calcium**: Physiology)), estrogen, growth hormone, thyroid hormone, and others (Table 2). The mineral homeostasis hormones ensure that the skeleton fulfills its metabolic function as a source of ions, by regulating osteoblast and osteoclast maturation and activity and thus ensuring that ions are released or retained by the skeleton as necessary. Circulating hormones may act on skeletal cells either directly or indirectly, modulating the synthesis, activation, receptor-binding, and binding proteins of a local growth factor, which in turn stimulates or inhibits bone formation and bone resorption. It is likely that hormones are important in the targeting of growth factors to tissue expressing specific hormonal receptors.

Growth factors may play a critical role in the coupling of bone formation to bone resorption, and possibly in the pathophysiology of bone disorders. The local factors are synthesized by skeletal cells and include growth factors, cytokines, and prostaglandins (Table 2). Growth factors are polypeptides that regulate the replication and differentiated function of cells. Growth factors have effects on cells of the same class (autocrine factors) or on cells of another class within the tissue (paracrine factors).

Bone Development and Growth

Bone is often described as developing by two different methods: intramembranous (in membrane) and endochrondral (in cartilage) ossification. The fundamental process is, however, similar. The osteoid is laid down by osteoblasts, and this then becomes calcified. The flat bones (such as the bones of the calvarium of the skull) are formed by intramembranous ossification, whereas the basal bones of the skull and the majority of bones of the skeleton, including the long bones, are formed by endochrondral ossification. The main difference between the processes is the presence of a cartilaginous phase in the latter.

Endochondral Ossification (Development of Long Bones)

Formation of a cartilage model Long bones begin as cartilaginous regions in the early embryo. Mesenchymal cells proliferate and differentiate into prochondroblasts and then into chrondroblasts. These cells secrete the cartilaginous matrix. Like the osteoblasts, the chrondoblasts become progressively embedded within their own matrix, where they lie within the lacunae; they are then called chrondocytes. Unlike the osteocytes, they continue to proliferate for some time, this being allowed in part by the gel-like consistency of cartilage. At the periphery of this cartilage (the perichrondium), the mesenchymal cells continue to proliferate and differentiate. They lay down a layer of osteoid, which immediately calcifies, so becoming a collar of periosteal bone directly in contact with the cartilaginous model. This is called appositional growth. Later on, the chrondocytes enlarge progressively, become hypertrophic, and undergo apoptosis.

Longitudinal growth, growth in diameter, and shape modification The embryonic cartilage is avascular. During its early development, a ring of woven bone is formed by intramembranous ossification in the future midshaft area under the perichondrium (which is then the periosteum). Just after the calcification of this woven bone, blood vessels (preceded by osteoclasts) penetrate the cartilage, bringing the blood supply that will form the hematopoietic bone marrow.

The growth plate in a growing long bone has a proliferative zone at the top, where chondroblasts divide actively, while their more mature descendants synthesize the matrix and enlarge, thereby producing longitudinal growth. Ultimately, the chondrocytes hypertrophy and calcify their matrix. Osteoclasts present in the marrow cavity invade this calcified cartilage, destroying the horizontal septa separating the chondrocytes. After osteoclastic resorption, osteoblasts differentiate and form a layer of woven bone on top of the cartilaginous remnants of the septa. This is the first remodeling sequence, and cartilage is replaced by woven bone. Still lower in the growth plate, this woven bone is subjected to further remodeling, in which the woven bone and the cartilaginous remnants are replaced with lamellar bone, representative of mature trabecular bone. Complete calcification of the growth plate at the end of puberty marks the end of longitudinal growth.

Growth in diameter of the shaft is the result of a deposition of new intramembranous bone beneath the periosteum that will continue throughout life. The midshaft is narrower than the metaphysis, and the growth of a long bone progressively destroys the lower part of the metaphysis and transforms it into a diaphysis, accomplished by continuous resorption by osteoclasts beneath the periosteum.

Intramembranous Ossification (Development of Flat Bones)

In intramembranous ossification, a group of mesenchymal cells within a highly vascularized area of the embryonic connective tissue proliferates and differentiates directly into preosteoblasts and then into osteoblasts. These cells synthesize and secrete osteoid which is calcified to become woven bone.

Blood vessels incorporated between the woven bone trabeculae will form the future hematopoietic bone marrow. Later, the woven bone is remodeled and is progressively replaced by mature lamellar bone. In early human fetal life, resorption and apposition begin to take place so that the cancellous bone occupies the center of the mass while a layer of cortical bone is formed on each surface by the continuous addition of new sheets of bone by active osteoblasts. Osteoclasts resorb bone from the inner surface to maintain proportional thickness and shape.

Age-Related Changes in Bone Mass

Bone mass in later life depends on the peak bone mass achieved during growth and the rate of subsequent age-related bone loss. Therefore, development of maximal bone mass during growth and reduction of loss of bone later in life are the two main strategies of preventing osteoporosis. Several factors are thought to influence bone mass. These can be broadly grouped into factors that cannot be modified, such as gender, age, body (frame) size, genetics, and ethnicity, and those factors that can be modified, such as hormonal status (especially sex and calciotropic hormone status), lifestyle factors, including physical activity levels, smoking and alcohol consumption patterns, and diet. The interaction of these genetic, hormonal, environmental, and nutritional factors influences both the development of bone to peak bone mass at maturity and its subsequent loss.

After the initial formation and development of bone, and during childhood and adolescence, there is a rapid linear and appositional skeletal growth, the former reaching a maximum between ages 15 and 20 years. Bone mass then continues to increase by appositional growth and the peak bone mass is probably attained during the third decade of life (Figure 3), although the exact timing is not certain and may vary between different regions of the skeleton. Peak bone mass and the subsequent rate of loss of bone are important determinants of risk of osteoporosis in later life. The regulation of peak bone mass is not fully understood but a number of factors have been identified. Of these, the most important are genetic influences; other determinants, which are potentially modifiable, include physical activity, nutritional factors, and hormonal status:

- Genetic factors – studies in twins show that 60–80% of peak bone mass is genetically determined; there is also some evidence that some aspects of bone architecture and geometry relevant to bone strength are inherited. The heritability of peak bone mass is believed to be polygenic (i.e., influenced by numerous genes) and there is currently a lot of interest in identifying these genes.
- Nutritional factors adequate intakes and status of a number of dietary components and nutrients are thought to be important in achieving optimal peak bone mass. These include calcium, phosphorus, magnesium, while high salt intakes may be detrimental.
- Physical activity – physical activity has important effects on bone growth and architecture during childhood and adolescence.
- Hormonal status – primary hypogonadism in either sex is associated with low bone mass, while secondary amenorrhea in women (due to anorexia nervosa, excessive exercise, or disease) results in low peak bone mass.

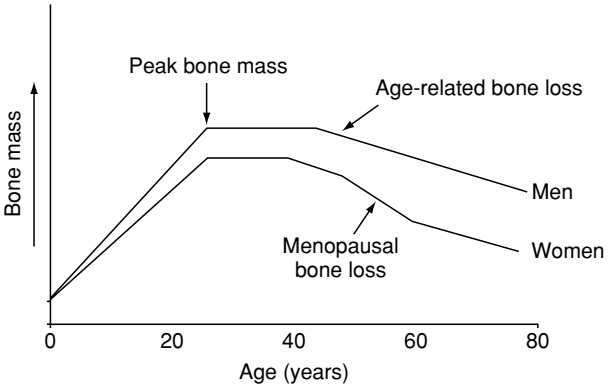

Figure 3 Age-related changes in bone mass in men and women.

Age-Related Bone Loss

After peak bone mass has been attained, there is a period of consolidation in which the transverse diameter of the long bones and vertebrae continues to increase by subperiosteal appositional growth. The age at which bone loss commences is uncertain but is believed to be around the age of 40 years in both men and women. Bone loss then continues throughout life, affecting both cortical and cancellous bone throughout the skeleton. In men, bone loss averages between 0.5 and 1.0% per annum. In women, there is an acceleration in the rate of bone loss around the time of the menopause to about 2%, although rates of bone loss vary widely, from less than 1% to 6% per annum. In the early postmenopausal years, bone loss from the spine exceeds that at other sites and overall it is estimated that, in women, approximately 35% and 50% of cortical and cancellous bone respectively are lost from the skeleton over the course of a lifetime.

The factors that contribute to age-related bone loss are incompletely understood. Potential contributory factors include:

- Hormonal factors – estrogen deficiency is an important determinant of menopausal bone loss. Both osteoclasts and osteoblasts possess estrogen receptors and thus may be influenced by estrogens. Declining production of sex hormones may also contribute to age-related bone loss in men, although this is less well documented than in women. Changes in the growth hormone–insulin-like growth factor I axis may contribute to age-related bone loss. The age-associated increase in parathyroid hormone is also thought to be a contributory factor.
- Nutritional factors have also been implicated; calcium deficiency due to an inadequate intake, reduced intestinal absorption, or increased renal deficiency may contribute to age-related bone loss. Vitamin D deficiency is common in many elderly populations and results in secondary hyperparathyroidism and increased bone turnover. Malnutrition, also common in some older individuals, may lead to acceleration of bone resorption, possibly as a result of reduced protein intake. Vitamin K-deficiency may contribute to an increased risk of osteoporotic fractures, although the pathogenesis of this condition has not been precisely delineated.
- Heritable factors – age-related bone loss can be very dramatic in certain individuals, and this decline cannot be attributed solely to hormonal or nutritional factors. This has led some researchers to hypothesize that genetic programming, when triggered, may lead to bone loss. However, the identity of genes involved in this process remains elusive.
- Environmental factors – decreasing physical activity with age is another likely contributor to age-related osteopenia, in both men and women. Similarly, smoking, alcohol, and some medications (such as glucocorticoids and anticonvulsants) may contribute to an excessive rate of bone loss in some elderly individuals.

Age-Related Fracture and Osteoporosis

Advancing age is a relatively strong risk factor for an osteoporotic fracture. Osteoporosis is defined as a systemic skeletal disease characterized by low bone mass and microarchitectural deterioration of bone tissue, with a consequent increase in bone fragility and susceptibility to fracture. Fragility fractures are the hallmark of osteoporosis and are particularly common in the spine, hip, and distal forearm (Colles fracture), although they can occur throughout the skeleton. These fractures constitute a major public health problem. Osteoporosis can be classified into type I (or postmenopausal osteoporosis) and type II (or senile osteoporosis). A more indepth discussion on the etiology, treatment, and prevention of osteoporosis as well as its consequences for quality of life are discussed elsewhere in this encyclopedia. (*See* **Osteoporosis**.)

Other bone disease states result from nutritional deficiency, e.g., rickets/osteomalacia (*See* **Cholecalciferol**: Properties and Determination; Physiology), or from genetic defects, e.g., osteopetrosis (characterized by defective osteoclast function and a consequent failure of bone remodeling) or osteogenesis imperfecta (characterized by an osteopenia associated with recurrent fracture and skeletal deformity stemming from a qualitative or quantitative abnormality of type I collagen).

See also: **Body Composition**; **Calcium**: Physiology; **Cholecalciferol**: Properties and Determination; Physiology; **Osteoporosis**

Further Reading

Baron R (1999) Anatomy and ultrastructure of bone. In: Murray FJ (ed.) *Primer on the Metabolic Bone Diseases of Mineral Metabolism*, 4th edn, pp. 3–10. Philadelphia, USA: Lippincott Williams & Wilkins.

Canalis E (1996) Regulation of bone remodeling. In: Murray FJ (ed.) *Primer on the Metabolic Bone Diseases of Mineral Metabolism*, 3rd edn, pp. 29–34. Philadelphia, USA: Lippincott-Raven.

European Commission (1998) *Report on Osteoporosis in the European Community – Action for Prevention*. Luxembourg: Office for Official Publications of the European Communities.

Lian JB, Stein GS, Canalis E, Robey PG and Boskey AL (1999) Bone formation: osteoblast lineage cells, growth factors, matrix proteins, and the mineralization process. In: Murray FJ (ed.) *Primer on the Metabolic Bone Diseases of Mineral Metabolism*, 4th edn, pp. 14–30. Philadelphia, USA: Lippincott Williams & Wilkins.

Mundy GR (1999) Bone remodeling. In: Murray FJ (ed.) *Primer on the Metabolic Bone Diseases of Mineral Metabolism*, 4th edn, pp. 30–38. Philadelphia, USA: Lippincott Williams & Wilkins.

Rodan GA (1998) Control of bone formation and resorption: biological and clinical perspective. *Journal of Cellular Biochemistry* 30/31: 55–61.

Thomson BM (1998) Bone: composition, metabolism and bone growth. In: Sadler MJ, Strain JJ, and Caballero B (eds) *Encyclopedia of Human Nutrition*, vol. 1, pp. 185–190. San Diego: Academic Press.

Vaughan J (1981) Bone as a tissue. In: Vaughan J (ed.) *The Physiology of Bone*, 3rd edn, pp. 1–26. Oxford: Clarendon Press.

BORON

C D Hunt, Grand Forks Human Nutrition Research Center, Grand Forks, ND, USA

Background

Boron is a light, nonmetallic element known to be essential for all vascular plants and some lower plants and prokaryotes. The discovery of two classes of naturally occurring organoboron compounds has provided some insight into the biomolecular role of boron. Recent research findings indicate that boron is important for the embryological development of lower vertebrates. In human and subhuman animals, boron modifies, and possibly regulates, energy substrate and mineral metabolism, and immune function.

Boron Chemistry

Boron Structure and Bonding

Boron is the fifth element in the periodic table and the only nonmetal in Group III. Although boron has five electrons and a ground-state electronic configuration of $1s^2\ 2s^2\ 2p^1$, it is always trivalent; simple electron loss to form a cation does not occur in boron chemistry. Boron has a small atomic size; the first ionization potential of boron ($800.5\,kJ\,mol^{-1}$) is rather high, and the next two are much higher (2426.5 and $3658.7\,kJ\,mol^{-1}$). It requires less energy to promote one 2s electron and mix the remaining 2s electron with the resultant $2p_x$ and $2p_y$ electrons to form a hybridized valence state of the sp^2 type. The three resulting B–X σ-bonding electrons can be regarded as occupying three equivalent sp^2 hybrid orbitals. Thus, the characteristics of the boron atom dictate covalent rather than ionic bonding. BX_3 compounds have a trigonal–near planar configuration.

In geological systems, boron neither occurs free nor binds directly to any element other than oxygen except for trivial exceptions, e.g., $NaBF_4$ (ferrucite) and $(K,Cs)BF_4$ (avogadrite). Only organoboron compounds are apparently important in biological systems during normal physiological conditions, and they are defined for this discussion as those organic compounds that contain B–O bonds, i.e., the orthoborates $(B(OR)_3$, $(RO)B(OR')_2$ and $(RO)B(OR')(OR'')$, and orthoborates of polyhydric alcohols. Organoboron compounds include B–N compounds, because B–N is isoelectronic with C–C.

Boron Speciation

Boric acid is an exclusively monobasic acid and is not a proton donor, but rather accepts a hydroxyl ion (a Lewis acid) to form the tetrahedral anion $B(OH)_4^-$ (eqn (1)):

$$B(OH)_3 + 2H_2O \leftrightarrow H_3O^+ + B(OH)_4^- \qquad pK_a = 9.25(25°C). \qquad (1)$$

At typical physiological boron concentrations (6.0×10^{-7} – $\sim 9.0 \times 10^{-3}\,mol\,l^{-1}$) in plants, animals (Table 1), or humans, inorganic boron is essentially present only as the mononuclear species boric acid $B(OH)_3$ and as borate $B(OH)_4^-$. Within the normal pH range of the gut and kidney, $B(OH)_3$ would prevail as the dominant species (pH 1: ~100% $B(OH)_3$; pH 9.3: 50%; pH 11: ~0%). Undissociated (and uncharged) boric acid is very soluble in water ($B[OH]_3$-saturate solution at $20\,°C = 0.75\,mol\,l^{-1}$) and has a high lipid solubility of the same order as urea. Polyborate species can form near neutral physiological conditions (pH ~7.4) when borate concentrations exceed $\sim 0.025\,mol\,l^{-1}$, an unusually high boron concentration in biological systems, but still lower than that found in the snap bean leaf ($0.1\,mol\,l^{-1}$).

The speciation of boron in foods has not been determined, but if plant and animal boron absorption mechanisms are analogous, the organic forms of boron *per se* are probably unavailable. However, the strong association between boron and polyhydroxyl ligands (described below) is easily and rapidly reversed by dialysis, change in pH, heat, or the excess addition of another low-molecular-weight polyhydroxyl ligand. Thus, within the intestinal tract, most ingested boron is probably converted to $B(OH)_3$, the normal end product of hydrolysis of most boron compounds.

Boron Isotopes

The study of boron kinetics in biological systems is hampered because the radioisotopes ^8B, ^{12}B, and ^{13}B all have half-lives of less than 1 s. However, two stable boron isotopes, ^{10}B and ^{11}B, are distributed unequally in nature (19.8 and 80.2%, respectively). This phenomenon has been successfully exploited to determine boron translocation in plant species and boron kinetics in animal models of human physiology. The ^{10}boron has a very large cross-section for thermal-neutron capture, and this property has been

Table 1 Concentration of boron in selected foods (wet-weight)[a]

Food	*Micrograms per 100 g*
Animal products	
Milk, low fat, 2% milkfat	27.0 ± 3.0[b]
Chicken, fryer, light meat, cooked with meat and skin, flour, with added fat	12.0 ± 2.0
Beef, ground, cooked, pan-fried	9.0 ± 2.0
Butter, with salt	4.0 ± 0.0
Cereal grain products	
Cornbread, homemade-type	56.0 ± 4.0
Cereal, ready-to-eat, wheat, shredded	34.0 ± 2.0
Bread, white, enriched	17.0 ± 1.0
Rice, white, long-grain, instant, enriched	9.0 ± 0.0
Macaroni, cooked, enriched, with salt	8.0 ± 1.0
Sweeteners/catsup	
Jelly, grape	145.0 ± 8.0
Catsup	100.0 ± 14.0
Sugar, white, granulated	10.0 ± 1.0
Infant foods	
Fruit, prunes with tapioca	359.0 ± 8.0
Carrots and beef, strained	132.0 ± 9.0
Vegetable, corn, creamed, strained	19.0 ± 2.0
Infant formula, Gerber, with iron	12.0 ± 0.0
Fruits and fruit beverages	
Avocado, raw	14 300.0 ± 42.0
Grape juice, unsweetened	342.0 ± 2.0
Cherries, sweet, raw	228.0 ± 3.0
Pears, raw	227.0 ± 11.0
Apple, juice, bottled, unsweetened	190.0 ± 3.0
Peach, raw	171.0 ± 7.0
Orange juice, unsweetened, reconstituted from concentrate	81.0 ± 1.0
Pineapple juice, canned	78.0 ± 1.0
Vegetables	
Snacks, potato chips, plain, salted	228.0 ± 6.0
Squash, winter, cooked, baked	161.0 ± 10.0
Carrots, raw	130.0 ± 7.0
Radishes, raw	66.0 ± 11.0
Potato, boiled, cooked without skin	48.0 ± 2.0
Beverages	
Alcoholic beverage, wine, table, white	364.0 ± 3.0
Coffee, instant, regular, prepared with water	24.0 ± 0.0
Carbonated beverage, cola, contains caffeine	10.0 ± 0.0
Tea, leaf, brewed in a porous bag	6.0 ± 0.0
Water, spring, bottled	2.0 ± 0.0

[a]Data from Meacham and Hunt (1998).
[b]Mean ± standard deviation ($n = 3$) of values from a single package of a single lot of a single brand.

used in neutron capture therapy (NCT) for inoperable brain tumors. During NCT, a boronated compound, preferentially bound to the tumor cells, is irradiated with thermal neutrons. Following neutron capture, the energy is released as a γ ray, α particle and lithium nucleus, the latter two traveling 9 and 5 μm, respectively. Thus, heavily ionizing radiation is deposited in single targeted cells with minimal collateral damage.

Boron Biochemistry

Boron Esters

Boroester formation Boroesters result from reactions between boron oxo compounds and certain mono or polyhydroxy compounds to form specific organoboron complexes. Boroesters probably represent the most biologically relevant boron species because of the vast number of biomolecules that contain one or more hydroxy groups with suitable positions. Boric acid reacts with suitable dihydroxy compounds to form the corresponding boric acid monoesters ('partial' esterification) (e.g., Figure 1) that retain the trigonal–planar configuration and no charge.

A borate monoester ('partial' esterification; monocyclic) (Figure 2) with a tetrahedral configuration and a negative charge is created when borate forms a complex with a suitable dihydroxy compound. A compound of similar configuration and charge is also formed when a boric acid monoester forms a complex with an available hydroxyl group. These two types of boromonoesters can react with another dihydroxy compound to give a corresponding spirocyclic borodiester ('complete' esterification) that is a chelate complex with a tetrahedral configuration and negative charge (Figure 3). Boric acid and boric acid-like structures, instead of borate, are most likely the reactive species with biological ligands, because it is probably easier for a diol to substitute for a relatively loosely bound water molecule associated with boric acid or a boric acid-like structure than it is for the diol to substitute for a hydroxyl ion in borate or a borate-like structure.

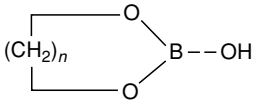

Figure 1 Structure I.

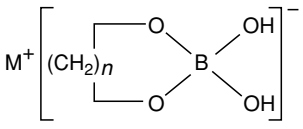

Figure 2 Structure II.

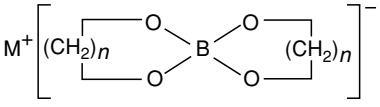

Figure 3 Structure III.

Ligand structure Ligands that contain adjacent and *cis* hydroxyl groups are those most likely to react with boron oxo compounds to give a boroester, and the reactivity of boric acid with the ligand generally increases as the number of these *cisoid* groups increases. The relevant *cis*-diol conformations for boron complexation are present in several biologically important sugars, their derivatives (sugar alcohols, -onic, and -uronic acids), and some polymers. Common examples include mannitol, ribose, erythrose, and glycerol. Sugars often form intramolecular hemiacetals in water where those with five-membered rings are called furanoses, and those with six-membered rings are called pyranoses. Compounds in a configuration where there are *cis*-diols on a furanoid ring (e.g., ribose, apiose, and erythritan) are rare in nature and form stronger complexes with boron than do compounds configured to have *cis*-diols predominately on a pyranoid ring (e.g., the pyranoid form of α-D-glucose). Glucose is virtually all in the pyranose form. The pyridine (e.g., NAD^+ or NADP) and flavin (e.g., FAD) nucleotides have received special attention, because they contain a ribose moiety with a *cisoid* diol conformation.

***p*K$_a$** Because borate complexes often have a much lower pK_a (e.g., boron-mannitol, 5.2) than free boric acid (9.25), the pK_a of many boron complexes is considerably below the physiological pH. This phenomenon has functional consequences and becomes the mechanism of action for some enzymatic activities. For example, boric acid is a competitive inhibitor of *Streptomyces griseus* Protease 3, and there is decreased affinity of the protein for boric acid below pH 7.0. This is directly attributable to the stability of the enzyme inhibitor complex, which is dependent on an ionizing group in the protein with an apparent pK_a of 6.6. This finding indicates that boric acid is an inhibitor of the enzyme at physiological pH by shifting the pK_a.

Boron–Nitrogen Compounds

The known ability of boron to form covalent bonds with the nitrogen atom of amine groups and the observation that boron binds near the coordinating iron site of hemerythrin (the nonheme iron-containing, oxygen transport protein of the sipunculid worm, *Golfingia gouldii*) suggest that a large array of biocompounds other than polyols can react with boron to form complexes. For example, there is experimental evidence to suggest that the mechanism for inhibition of a sub-subclass of enzymes (the serine proteases) by boron involves the formation of a covalent bond between boron and a specific nitrogen at the active site of these enzymes (Figure 4).

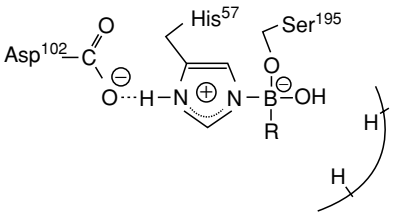

Figure 4 Structure IV.

Boron Occurrence in the Environment

Anthropogenic Sources

Each year, coal combustion is estimated to release between 28 000 and 330 000 tonnes of boron to the atmosphere and between 24 000 and 287 000 tonnes of boron go to landfill/soil. Also, approximately 300 000 tonnes of boron are recovered from extracted and refined deposits destined for five main product groups: insulation fiberglass/textile fiberglass/borosilicate glass (39%), ceramics (13%), detergents (19%), fertilizers (5%), cellulose insulation (2%), and other products (23%) including nuclear shielding, leather tanning, and catalysis. Boron, in the form of borax, has been used since Egyptian and Roman times for the preparation of hard (borosilicate) glasses. Boron leaches readily from most glassware, especially under conditions of alkalinity.

Soil

The largest single natural deposit of boron is located at Boron, California, an area of former volcanic activity and associated with the waters of former hot springs. Within geological boron deposits, boron is usually present as a borate mineral such as borax, $Na_2B_4O_5(OH)_4{\cdot}8H_2O$. A 1956 survey of US soils indicated an average total boron content of 30 μg g^{-1}. Less than 5% of the total soil boron is available for crop uptake and depends on release from soil minerals, mineralization from decomposed organic matter, artificial addition to soils by irrigation or fertilization, and subsequently distribution between the solid and liquid phases of the soil. Boron is easily leached out of soils by rainwater. Thus, plant boron deficiency is of concern in humid regions with light-textured, acid soils.

Water

Weathering of clay-rich sedimentary rock is the major source of total boron mobilized into the aquatic environment, although sewage effluent can be an important source of boron, especially in areas where boron compounds are major constituents of detergent products. Undissociated boric acid is the

predominant species of boron in most natural freshwater systems. The concentration of boron in surface water may vary substantially between geographic locations and water sources (values in $\mu g\,ml^{-1}$; in parentheses, $\mu mol\,ml^{-1}$): UK, 0.046–0.822 (0.004–0.077); Italy, 0.400–1.00 (0.037–0.093); the USA, 0.001–5.0 (0.00009–0.467). The concentration of boron in seawater is $4.5\,\mu g\,ml^{-1}$ ($0.42\,\mu mol\,ml^{-1}$), whereas the median boron concentration in USA drinking water is $0.031\,\mu g\,ml^{-1}$. Thus, the usual boron intake through drinking water is usually not high. For example, an adult consuming 2 l of water with the median concentration and ingesting 1.5 mg of boron in the diet would be exposed to < 1.6 mg or < 10% of the acceptable reference dose (18 mg of boron per day).

Foods

Boron is unequally distributed within Angiospermae, the class of plants most often utilized in human and animal diets. Most species within the subclass Dicotyledoneae, which includes fruits, vegetables, tubers and legumes, have much higher concentrations of boron than do species in the subclass Monocotyledoneae, especially gramineaceous species (the grasses), i.e., rice, corn, rice, and wheat (Table 1). Foods cooked in boiling water are quickly leached of boron content. Also, $B(OH)_3$ is volatile in steam. Boron loss could conceivably be further enhanced during pressure cooking with chlorinated or fluorinated water; boron–halogen compounds are extremely volatile at high temperatures.

Estimated daily intakes of boron vary with age and sex. Infants, toddlers, adolescent females and males, adult females and males, and senior females and males in the USA consume 0.545, 0.534, 0.590, 0.848, 0.698, 0.907, 0.731, and 0.863 mg of boron per day, respectively. The concentration of elemental boron in animal muscle and milk is relatively low. Because whole milk (0.29 µg of boron per gram) is consumed in such large quantities, it is the typical primary source of boron for infants, toddlers, and adolescents. Coffee is the primary source of boron for adult females and males, and senior females and males in the USA. There is a close relation between boron and vitamin intakes; a diet low in boron is typically low in vitamins, especially ascorbic acid.

Boron Essentiality and Function

The high affinity of boric acid for adjacent and *cis*-hydroxyl groups present on biomolecules may have exerted evolutionary pressure to select for carbohydrate energy sources with low percentages of the furanose forms described above. Conversely, there may have been selection for two natural sugars, apiose and ribose, because their physiological derivatives have a strongly borate-complexing furanose configuration and they serve as components of important structural or enzyme-related biomolecules, as described below.

Prokaryotes

Boron essentiality for the heterocystous Cyanobacteria, predominant organisms during the Middle Pre-Cambrian Period, indicates that boron was an essential element during the early evolution of life. Boromycin, an antibiotic from *Streptomyces antibioticus*, was the first well-defined boron-containing biological organic compound. Antibiotics with a similar structure are excreted by a marine isolate belonging to *Streptomyces griseus* (aplasmomycin) and from the myxobacterium *Sorangium cellulosum* (tartrolon B; Figure 5). These antibiotics are ionophoric macrolides and have four inward directed oxygen atoms (two *cis*-hydroxyl groups) that provide an ideal geometry for accommodation of the boron atom and thus formation of a stable borodiester.

Plants

Whether certain lower plants lack a boron requirement is being revisited. Findings in 1999 that boron stimulated growth in the fungus *Saccharomyces cerevisiae* (a yeast) supplanted findings in 1968 to the contrary. There is a demonstrated boron requirement for *Penicillium italicum* but apparently not for the fungi *Neurospora crassa* or *Aspergillus niger*. Boron stimulated penicillin production, but not growth, in *Penicillium chrysogenum*, and boron enhanced growth of *Chlorella vulgaris*, but no apparent boron requirement has been demonstrated among the many species of chlorophyta tested. At least 16 species of

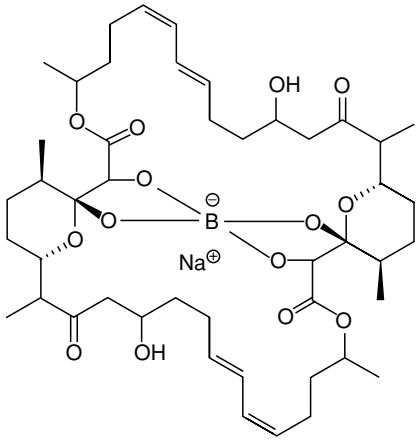

Figure 5 Structure V.

marine and eight species of freshwater diatoms and some species of marine algal flagellates have an absolute boron requirement.

All vascular plants have acquired a unique requirement for boron. However, the primary role and mode of action of boron remain controversial. Boron seems to be involved in such diverse cellular processes as sugar transport, cell-wall synthesis, lignification, cell-wall structure, carbohydrate metabolism, RNA metabolism, respiration, indole acetic acid metabolism, phenol metabolism, and membranes. Rhamnogalacturonan-II (RG-II) is a small, structurally complex polysaccharide of the pectic fraction of primary plant cell walls. It is present in at least 24 plant species and represents an extreme example of the evolutionary conservation of wall polysaccharide structure. RG-IIs contain the unusual monosaccharide apiose described above and an atom of boron that cross-links two RG-II dimers at the site of the apiose residues to form a borodiester (Figure 6). Knowledge that borate ester cross-linking of polysaccharides *in vitro* is pH-dependent has led to the suggestion that the boron cross-links in RG-II are the 'load-bearing,' acid-labile linkages that are hydrolyzed by a decrease in wall pH during auxin-induced cell expansion. However, graminaceous plants, with an absolute but low requirement for boron, have very low amounts of boron, pectin, and complexable polysaccharides in their cell walls, a finding that suggests unknown additional roles for the element.

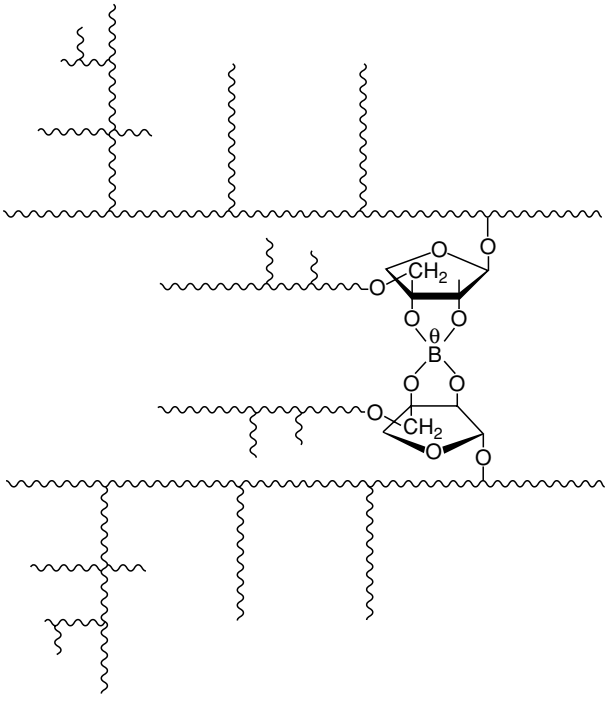

Figure 6 Structure VI.

Humans and Animal Models

Boron homeostatic control Gastrointestinal absorption of inorganic boron and subsequent urinary excretion are near 100%. In humans, a lack of boron accumulation and relatively small changes in blood boron values during a substantial increase in dietary boron support the concept of boron homeostasis. Female rats given water high in boron (9.25 mmol l^{-1}) for 21 days exhibited increased plasma boron concentrations, although some type of homeostatic mechanism concurrently eliminated any excess of boron from two examined organs, liver and brain, against their own concentration gradients. In cows, the percentage of filtered boron reabsorbed by the kidneys decreased significantly with increased boron intake, a finding that suggests physiologic regulation of the element.

Boron and reproduction Boron may be essential for embryological development in at least two separate vertebrate phylogenetic classes. In the South African clawed frog, *Xenopus laevis*, dietary boron deprivation (4.16 μmol kg^{-1}) causes a high proportion of necrotic eggs, fertilized eggs with a high frequency of abnormal gastrulation, and abnormal development of the gut, craniofacial region, and eye; visceral edema; and kinking of the tail musculature and notochord. In mated zebrafish (*Danio rerio*) fed boron-depleted brine shrimp and maintained in low-boron water (0.1 μmol l^{-1}), the boron content of the blastulas was only 5% of that in blastulas from boron-supplemented (45.0 μmol l^{-1}) parents. The early cleavage stage of development was the most sensitive to boron deficiency, although repletion of low-boron embryos during the first hour after fertilization rescued them from death.

Boron and cartilage and bone structure Interest in mammalian boron nutrition was renewed after a 36-year hiatus with the discovery that boron supplementation of a boron-low diet reduced gross bone abnormalities in the vitamin D-deficient chick model system. Other subsequent findings at the microscopic level indicated that physiologic amounts of boron (321 μmol kg^{-1}) function to modify mineral metabolism in vitamin D deficiency (3.13 μg kg^{-1}) by suppressing initiation of growth cartilage mineralization in magnesium deficiency and enhancing cartilage calcification in magnesium adequacy (see Figure 7). The effects of boron in this model on cartilage calcification apparently are beneficial in both magnesium deficiency and adequacy because the vitamin D deficiency-induced mortality was substantially reduced by dietary boron. Furthermore, supplemental boron

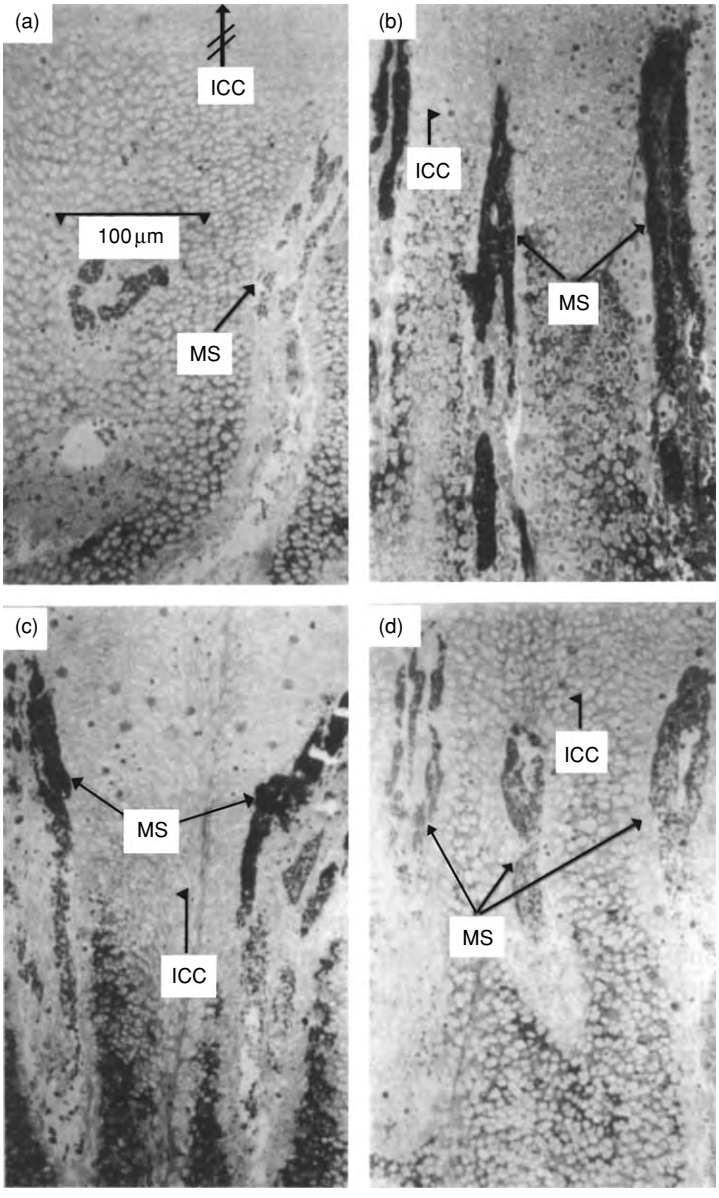

Figure 7 Microphotographs (all at same magnification) depicting marrow sprout (MS) orientation and initiation of cartilage calcification (ICC) in the chick proximal tibial epiphysial plate. ICC beyond a frame is denoted by slashed arrows. Boron supplementation corrected the cholecalciferol-deficiency-induced disorientation of the MS (compare a and b, and c and d). ICC is inhibited by boron in magnesium inadequacy (compare a and b), but enhanced in magnesium adequacy (compare c and d). Dietary supplements (milligrams of supplement per kilogram of diet): (a) B, 0; Mg, 300; (b) B, 3; Mg, 300; (c) B, 0; Mg, 500; (d) B, 3; Mg, 500. Data from Hunt (1989).

alleviated distortion of the marrow sprouts, a distortion characteristic of vitamin D deficiency.

Boron also has an apparent undefined role in metabolic events not directly related to extracellular matrix calcification. *In ovo* injections of boron (90 μmol) or 1-25-dihydroxycholecalciferol reduced the abnormal height of the growth plate of 1-day-old chicks hatched from vitamin D-deficient eggs. In addition, physiological supplements of boron (130 μmol kg^{-1}) to a boron-low diet (16 μmol kg^{-1}) increased chondrocyte density in the proliferative zone of the growth plate in vitamin D-deficient chicks.

Boron and mineral metabolism In postmenopausal volunteers fed marginal amounts of magnesium (4.5 mmol per day), boron supplementation (280 μmol per day) of the low-boron diet (33 μmol per day) significantly decreased the percentage of dietary calcium lost in the urine (1.4%) but significantly increased that percentage (2.5%) in volunteers fed

ample amounts of magnesium (14.0 mmol per day), a relation that may be important in understanding metabolic mineral disorders that perturb calcium balance. A similar phenomenon occurred in either free-living sedentary or athletic premenopausal women consuming self-selected typical Western diets not different in analyzed calcium content; boron supplementation (280 μmol per day) increased urinary calcium loss. In a different study of older volunteers fed a low-magnesium (4.73 μmmol 8330 kJ^{-1}), marginal copper (25.3 μmol 8330 kJ^{-1}) diet (men, and women on or not on estrogen therapy), boron repletion (300 μmol 8330 kJ^{-1}) after boron depletion (21.3 μmol 8330 kJ^{-1}) significantly decreased calcitonin and significantly increased ionized calcium, but not total calcium concentrations in serum in a manner similar to that caused by estrogen therapy. In the vitamin D-deficient rat model fed a low-boron diet, supplemental dietary boron enhanced the apparent absorption and retention of calcium and phosphorus and increased femur magnesium concentrations.

Boron and steroid metabolism There is clear evidence that dietary boron affects steroid metabolism. In particular, circulating concentrations of vitamin D metabolites are sensitive to boron nutriture. The number of vitamin D-deficient chicks with detectable serum concentrations of 25-hydroxycholecalciferol and the overall serum concentrations of 25-hydroxycholecalciferol increased when boron-deprived chicks (15 μmol B kg^{-1}) were supplemented with physiological amounts of boron (280 μmol kg^{-1}). In the same study, physiological supplements of boron improved growth, feed efficiency, leg conditions, and bone calcium, and serum ionized calcium in broiler chicks fed low amounts of vitamin D, but not in vitamin D-deprived or adequate chicks. These findings indicate that dietary boron enhances the efficacy of vitamin D but cannot substitute for the vitamin. Further research is needed to determine whether boron enhances vitamin D absorption or increases cholecalciferol hydroxylase activity. In volunteers (men, and women on or not on estrogen therapy), boron supplementation after consumption of a low boron diet increased serum 25-hydroxycholecalciferol (62.4 ± 7.5 vs. 44.9 ± 2.5 mmol l^{-1}, mean $\pm$ SEM), an effect that may be especially important during the winter months when those concentrations normally range between 35 and 105 mmol l^{-1}.

The circulating concentrations of 17β-estradiol also respond to boron nutriture. Perimenopausal women who excreted < 93 μmol of boron per day during the placebo period exhibited increased serum concentrations of estradiol after boron supplementation (231 μmol of boron per day) of self-selected diets. In a separate study, postmenopausal women on estrogen therapy, but neither men nor postmenopausal women not ingesting estrogen, also exhibited increased serum concentrations of estradiol after boron supplementation (280 μmol of boron per day) of a low boron diet (23 μmol of boron 8330 kJ^{-1}). However, plasma estradiol, but not testosterone, concentrations increased in young male volunteers when their self-selected diets were supplemented with ample amounts of boron (10 mg per day).

Boron and enzyme regulation Boron influences the activities of at least 26 enzymes examined (various oxidoreductases, transferases, hydrolases, and isomerases), most often in an inhibitory manner, by binding to cofactors (e.g., NAD) or substrates, or by unknown mechanisms. Reversible enzymatic inhibition as an essential role for an element is unusual. However, there is irrefutable evidence that boron serves to inhibit or dampen several metabolic pathways in higher plants. A serious outcome of boron deficiency in plants is starch accumulation, a condition thought to be caused by increased activity of an enzyme in that pathway, starch phosphorylase. Oxidoreductase enzymes that require pyridine (e.g., NAD^+ or NADP) or flavin (e.g., FAD) nucleotides (EC 1.1.1, 1.1.3, 1.2.1, 1.3.5, 1.6.2) are competitively inhibited by borate or its derivatives as boron competes for the NAD or flavin cofactor. The serine proteases (EC 3.4.21) represent one critical sub-subclass of hydrolases that have many essential regulatory roles including the blood coagulation system (e.g., coagulation Factor Xa). Boron reversibly inhibits the activity of these enzymes by serving as a transition-state analog. It is possible that boron acts as an unobtrusive metabolic regulator by quenching the activity of some enzymes and/or stabilizing reactive compounds to limit the hyperactivity of several physiological systems including the normal inflammatory response.

Boron and membrane function There are numerous indications that boron is important for membrane function. The esterification reaction that produces boromonoesters (e.g., **Figure 1** or **2**) is easily reversible, because these esters are completely hydrolyzed when transferred into water. Therefore, it is reasonable that boromonoesters in the hydrophobic environment of the lipid portion of the plasmalemma should have a prolonged functionality, because the absence of water in these environments shifts the equilibrium to the right (eqn (2)).

$$B(OH)_3 + 3ROH \leftrightarrow B(OR)_3 + 3H_2O. \qquad (2)$$

Heterocystous cyanobacteria may require boron interactions with polyhydroxy components of the

cellular membrane to maintain efficient conformation of cell membranes and thus preservation of the anaerobic environment of the heterocyst. Very early research on plant boron nutrition indicated that boron deficiency caused a reduction in ion uptake in *Impatiens balsamina*, and considerable evidence has accumulated since then that boron is involved in plasmalemma-bound transport processes for higher plants and diatomaceous species. Low-boron zebrafish zygotes began to die during the one-cell stage where blebbing of the cell and yolk-sac membranes marked the early stages of death followed by extrusion of cytoplasm and yolk-sac contents. In boron-deprived rats, increases in dietary potassium from marginally deficient to adequate to luxuriant induced increases in the transport of calcium into thrombin-activated platelets. However, physiological amounts of boron muted the effect of potassium on calcium transport. Boron may have reacted directly with thrombin to somehow modify platelet activation because thrombin is a serine protease inhibited by boron. However, these findings are similar to those expected for plasmalemma ion transport involving regulation of the voltage-dependent Ca^{2+} channel by boron through inhibition of NADPH or cAMP concentrations.

Boron and immune function There is evidence that dietary boron helps control the normal inflammatory process. The antigen-induced arthritis rat model exhibited reduced paw swelling and circulating neutrophil concentrations when fed with physiological amounts of boron ($190\,\mu mol\,kg^{-1}$). Commercial rat chow contains considerable higher amounts of boron ($\sim 1100\,\mu mol\,kg^{-1}$). Perimenopausal women excreted 110 and 304 μmol of boron per day during the placebo and boron supplementation periods, respectively, and exhibited an increased percentage of polymorphonuclear leukocytes during the boron supplementation period. Boron-containing rhamnogalacturonan-II from *Panax ginseng* leaves enhanced expression of the Fc receptor (that internalizes antigen–antibody complexes) and interleukin-6 (IL-6) production activity of mouse macrophages. Dietary boron may serve as a signal suppressor that downregulates specific enzymatic activities typically elevated during inflammation at the inflammation site. Suppression, but not elimination, of these enzyme activities by boron is hypothesized to reduce the incidence and severity of inflammatory disease.

Boron Toxicity

As with all other elements, boron produces toxicity in all tested biological organisms when excessive amounts are absorbed. The associated toxicity of boric acid when used as an antiseptic in lieu of antibiotics on abraded epithelium (i.e., surgical wounds and diaper rash) was overlooked for many years, even though signs of poisoning were reported soon after its introduction into clinical use. Boron is more of a bacteriostat than a bactericide and as such is not sufficient for most modern usage. The minimum lethal dose of boron for humans has not been established, although single doses of 18–20 g in adults have been fatal. Signs of acute boron toxicity, regardless of route of administration, include nausea, vomiting, headache, diarrhea, erythema, hypothermia, restlessness, weariness, desquamation, renal injury, and death from circulatory collapse and shock. Autopsy may reveal congestion and edema of brain, myocardium, lungs, and other organs, with fatty infiltration of the liver. Chronic boron toxicity symptoms include poor appetite, nausea, weight loss, decreased sexual activity, seminal volume, sperm count and motility and increased seminal fructose. At present, death from boron poisoning is exceptionally rare probably because of the emphasis placed on maintaining the electrolytic balance and supporting kidney function during the worst part of the illness. Depending upon boron blood levels, treatment ranges from observation to gastric lavage to dialysis.

See also: **Bone**; **Enzymes**: Functions and Characteristics; **Hormones**: Steroid Hormones; **Immunology of Food**

Further Reading

Culver BD, Sullivan FM, Murray FJ, Coughlin JR and Strong PL (eds) (1998) Proceedings of the Second International Symposium on the Health Effects of Boron and its Compounds. *Biological Trace Element Research* 66: 1–473.

Fort DJ, Stover EL, Strong PL, Murray FJ and Keen CL (1999) Chronic feeding of a low boron diet adversely affects reproduction and development in *Xenopus laevis*. *Journal of Nutrition* 129: 2055–2060.

Hunt CD and Idso JP (1999) Dietary boron as a physiological regulator of the normal inflammatory response: a review and current research progress. *The Journal of Trace Elements in Experimental Medicine* 12: 221–233.

Hunt CD, Herbel JL and Nielsen FH (1997) Metabolic responses of postmenopausal women to supplemental dietary boron and aluminum during usual and low magnesium intake: boron, calcium, and magnesium absorption and retention and blood mineral concentrations. *American Journal of Clinical Nutrition* 65: 803–813.

Johnson SL and Smith KW (1976) The interaction of borate and sulfite with pyridine nucleotides. *Biochemistry* 15: 553–559.

Loomis WD and Durst RW (1992) Chemistry and biology of boron. *BioFactors* 3: 229–239.

Meacham SL and Hunt CD (1998) Dietary boron intakes of selected populations in the United States. *Biological Trace Element Research* 66: 65–78.

Nielsen FH and Penland JG (1999) Boron supplementation of peri-menopausal women affects boron metabolism and indices associated with macromineral metabolism, hormonal status and immune function. *The Journal of Trace Elements in Experimental Medicine* 12: 251–256.

Rowe CD and Eckhert CD (1999) Boron is required for zebrafish embryogenesis. *Journal of Experimental Biology* 202: 1649–1654.

Schummer D, Irschik H, Reichenbach H and Höfle G (1994) Antibiotics from gliding bacteria, LVII. Tartrolons: new boron-containing macrodiolides from *Sorangium cellulosum*. *Liebigs Annalen der Chemie* 1994: 283–289.

Botanicals *See* **Herbs**: Herbs and Their Uses; Herbs of the Compositae; Herbs of the Labiatae; Herbs of the Umbelliferae; Herbs Used in Alcoholic Drinks

Botulism *See* ***Clostridium***: Occurrence of *Clostridium perfringens*; Detection of *Clostridium perfringens*; Food Poisoning by *Clostridium perfringens*; Occurrence of *Clostridium botulinum*; Botulism

Bourbon *See* **Whisky, Whiskey, and Bourbon**: Products and Manufacture; Composition and Analysis of Whisky

BOVINE SPONGIFORM ENCEPHALOPATHY (BSE)

R Bradley, Private Consultant, UK

Introduction

Bovine spongiform encephalopathy (BSE) is a feed-related infectious disease of cattle, *par excellence*. It is a new disease first identified as a subacute spongiform encephalopathy (SE) in November 1986, though the earliest clinical case probably occurred in the UK in April 1985. BSE was later determined to be a transmissible SE (TSE) and a prion disease. Such diseases had previously been confined to humans (Creutzfeldt–Jakob disease (CJD), Gerstmann–Sträussler–Scheinker (GSS) disease, and kuru), sheep and goats (scrapie), farmed mink (transmissible mink encephalopathy (TME)) and certain species of deer in North America (chronic wasting disease (CWD)). Of these, only TME was certainly contracted via feed. All these diseases, with the possible exception of scrapie and now BSE, are rare. For example, the annual incidence of CJD is between one and two cases per million per annum worldwide.

In the years following the discovery of BSE, a number of other species of the families Bovidae and Felidae became affected with similar diseases, almost certainly following feed exposure to the BSE agent. Finally, in March 1996 10 cases of a new variant form of CJD (vCJD) in the UK were announced. In the absence of any other plausible explanation, exposure to the BSE agent before 1989 was presumed to be the cause. In 1989 a specified bovine offals (SBO) ban had been put in place in the UK to protect public health to supplement the 1988 measure of removing all clinically suspect animals from all food and feed chains. SBO are the offals from cattle that were, on

scrapie evidence, likely to contain BSE infectivity. In fact, subsequent transmission studies from natural cases of BSE have only detected infectivity in the central nervous system. Other tissues, including blood, milk, and muscle meat, show no detectable infectivity. Subsequently it was shown that the agent causing BSE in cattle was indistinguishable biologically from the vCJD agent, i.e., the strain type was the same. Furthermore, the strain-type characteristics were different from the agents causing conventional forms of sporadic CJD and scrapie. Molecular analysis of the prion protein (PrP) from some of these species has also shown distinct similarities between them. In particular, that from vCJD cases has a unique molecular signature that distinguishes it from conventional sporadic and other forms of CJD.

The following is a brief account of the BSE epidemic in cattle, where it exists, how it is being eliminated, and how both public health and animal health are being protected.

The Cause of BSE

BSE is caused by the BSE agent. TSE agents can be biologically strain-typed in inbred strains of mice, a process that usually takes years. The BSE agent exists as a single major strain and is different from the agents that cause scrapie, TME, CWD and all forms of human TSE except vCJD.

The Nature of TSE Agents and Diagnostic Methods

The nature of TSE agents is disputed. They have been variously described as prions, virinos, and unconventional viruses. They are unlike bacteria and conventional viruses as they provoke no conventional immune response in infected hosts. Therefore there is no immunological test that can be used to identify an infected individual before the onset of clinical signs. TSE agents are the nearest things to indestructible forms of life. They can resist boiling, conventional autoclaving, dry heat to over 300 °C (and perhaps more) for an hour, ionizing and ultraviolet radiation, burial for up to 3 years, formaldehyde, and a range of chemicals that are effective against conventional bacteria and viruses. Nevertheless, several of these treatments do reduce the titer of infectivity, often substantially. Specified autoclave conditions and incineration, sodium hypochlorite, and sodium hydroxide are either effective or useful in practice.

The Prion Hypothesis and PrP Detection

The prion hypothesis, in its purest form, proposes that the agent is an infectious protein, an altered form of a host protein known as PrP. This protein is coded for by the *PrP* gene that exists in all cells of an individual. The gene is highly preserved in mammals and some other orders of animals. PrP is found in several tissues of mammals but is most abundant in nervous tissue and particularly nerve cells, where it is regarded as a neuronal membrane protein. The function of this protein is unknown though it may control programmed cell death. Mice appear to live successfully if the *PrP* gene is deleted (*PrP* knockout mice), though some models have shown altered circadian rhythm and others deficits in synaptic function.

The protein found in normal individuals is called PrP^{C}. In disease there is a post-translational conformational change that results in the formation of PrP^{Sc}. This involves a change in shape from mainly α helix to mainly β sheet. PrP^{C} is denatured by proteases such as proteinase K. PrP^{Sc}, on the other hand, is insoluble and partially protease-resistant. It is then known as PrP-res. These features have been exploited to develop diagnostic methods to identify the protein in infected tissues. Such methods include immunoblotting (Western blotting), dot blotting, and immunohistochemistry (the detection of PrP in tissue sections). Disease-associated PrP tends to aggregate in tissues. Diagnostic use is made of this phenomenon by detection of disease-specific fibrils (scrapie-associated fibrils or SAF) by electron microscopy. This is done on detergent extracts of brain or spinal cord, treated with proteinase K, and followed by electron-dense staining. More recently Rapid diagnostic tests have been developed for use in cattle and sheep to aid surveillance for the disease (active surveillance). All Rapid tests are based on the detection of PrP^{Sc} in central nervous tissue collected after death.

The Virino Hypothesis

There is no doubt that infectivity and PrP are closely associated and difficult to separate. However, doubt has been cast on the validity of the prion hypothesis. This is because some features of the transmissibility and dynamism of TSE agents, such as mutation and agent strain differences, for which there is abundant evidence, are not well explained by the basic 'protein-only' hypothesis. These observations have been made following transmission experiments between species. This has meant that various modifications to the prion hypothesis have been proposed over the years to explain the observations. One of these modifications is the proposition that 'protein X', an unidentified protein, is required for infectivity. This may act as a molecular chaperone. A second is that in some prion disease neurodegeneration occurs in the absence of PrP^{Sc}. Instead, it may be caused by a transmembrane form of PrP, called ^{Ctm}PrP.

Although the biological features of agents isolated from cattle with BSE and several other species with recently discovered, natural, or experimental TSE, including three human patients with vCJD, are indistinguishable (the biological strain type is the same), the primary PrP amino acid sequence from each is different. This feature, mutation, and strain variation in general can be more readily explained by conventional molecular theory, namely that the agent has a genome that is, perhaps, a small nucleic acid. The agent therefore is hypothesized to be a hybrid mixture of PrP and the agent genome. However, the weakness in the hypothesis is that this genome has remained undiscovered.

Unconventional Virus Hypothesis

There are generally fewer supporters of this hypothesis but most of them claim it has not been disproved. An infection-specific nucleic acid has not been identified.

Whatever is the structure of the causal agent it is clear that the PrP plays a central role and this must be accommodated in any hypothesis. Claims that BSE is an autoimmune disease or alternatively that organophosphorus compounds are a prime cause are not viable theories.

Epidemiology of BSE

Origin of BSE

The UK epidemic of BSE had an extended common source. The features were consistent either with an origin from sheep affected with a scrapie-like agent or from a cattle-adapted scrapie-like agent. They were not consistent with exposure to a mutant form of scrapie. The source of infection has been identified as concentrate feed and the vehicle is ruminant-derived protein in the form of meat-and-bone-meal (MBM). Bovine blood, fetal calf blood, components of blood, and bovine milk have shown no detectable infectivity following bioassay. Positive epidemiological evidence for transmission of BSE via milk is also absent. All authorities, including the World Health Organization (WHO), Office International des Epizooties (OIE), European Commission (EC), Food and Drug Administration (FDA), and Spongiform Encephalopathy Advisory Committee (SEAC), having reviewed the evidence conclude that bovine milk is a safe product in regard to TSE, for all purposes. This means that suckled cattle, not the offspring of a BSE case, reared entirely on milk and subsequently on grassland have a negligible risk of developing BSE. It also follows that unadulterated bovine blood and milk products can safely be used in pharmaceuticals, biologicals, and cosmetics. However, in the case of bovine blood an additional safeguard against contamination is to source the material from countries that are classified in the highest category of safety.

The genetic component in BSE is small. Thus, the disease can be considered as an infectious disease transmitted mainly via feed.

There is a possibility that maternal transmission may play a minor role in the transmission of BSE in the UK but, even if it does, it will not itself prevent the elimination of the disease.

It has not been possible to transmit BSE via embryos derived from clinical cases of BSE either to the offspring so derived, or to recipient dams after a 7-year period. Male or female reproductive tissue, including embryos, fluids used for flushing embryos from the uterus, placenta, and milk, have not revealed detectable infectivity following bioassay. Nevertheless a cohort study, confounded by accidentally feeding MBM to both cases and controls, provided some evidence that there was a risk of maternal transmission at a level around 10% in the study. The risk was highest for calves born whilst the dam was clinically affected but there was also a significant risk for calves born in the 6 months prior to onset of clinical signs in the dam. Subsequent experience of BSE in the UK as the epidemic declines and following detailed epidemiological study of a few BSE cases born after 1 August 1996, when the second feed ban was considered complete, has not supported such a high incidence of maternal transmission. Furthermore, no BSE has been reported in any offspring of dams with BSE in any other country, even though over 3000 cases or positive animals have occurred. In summary, maternal transmission cannot be entirely ruled out but evidence for it is meager and no plausible method has been proposed for its occurrence.

A recent review reports that, during the period 1970–88 (a ruminant feed ban was introduced in July 1988 in GB) feed manufacturers introduced MBM into rations for calves from the first or second week of age. Accordingly, in the UK, a high proportion of dairy cows and cross-bred suckler cows derived from dairy cows will have been fed MBM as calves. This was an important risk factor for BSE and accounted for most cases of BSE being in dairy cows or cows derived from the dairy herd.

Commercially and safety-stimulated changes in the processing of animal waste around 1980 in the UK apparently permitted an increase of BSE agent infectivity to enter the MBM, sufficient to cause detectable disease by 1985–86. Disease resulted after the mean 60-month incubation period was complete. Recycling BSE infection from cattle via MBM exacerbated the

epidemic, whether its origin was from sheep or cattle. Recycling commenced from infected tissues of incubating cattle, before BSE was even discovered, and resulted in a marked increase in incidence from mid-1989. MBM was probably also the source of infection for a small number of captive wild Bovidae that developed TSE during the late 1980s and early 1990s in the UK. The origin of TSE in captive wild Felidae was consumption of uncooked carcass material from fallen stock (animals that die on the farm or in transit) that contained central nervous system (CNS) tissue. It is these tissues that contain the highest levels of infectivity in clinically affected animals. The precise origin of feline spongiform encephalopathy (FSE) in domestic cats is less certain but was via feed.

The Scientific Steering Committee (SSC) of the EC has frequently been confronted with unresolved issues relating to the origin of BSE and with alternative hypotheses for the transmission of this disease other than via animal proteins and maternal transmission. They reviewed the evidence in 2001 provided in reports by the TSE/BSE ad hoc group and gave an opinion. They concluded that:

- Epidemiological studies, rendering studies, and the effect of feed bans in all countries with BSE very clearly support the hypothesis of infected mammalian protein in the form of MBM being the major vehicle for BSE transmission in cattle. It can enter the feed deliberately or accidentally by cross-contamination that can occur readily during feed preparation in feed mills, during transportation, or on the farm.
- Maternal transmission is theoretically a possible route of transmission, although it has not yet been demonstrated.
- Other infected materials, such as gelatin and/or fat, incorporated in feed represent another possibility of transmission.
- The PrP theory remains central to BSE and other TSE.
- It is wise at present to maintain an open mind on the nature of the BSE agent(s).
- A number of other hypotheses have been proposed on the nature of the responsible agent(s) of BSE (e.g., a toxin, a bacterium, alkaloidal glycosidase inhibitors, autoantibodies, and a single-stranded DNA), but for none of them is sufficient scientific evidence available and some are clearly implausible or very difficult to investigate.
- The origin of the BSE prion is not known. Many hypotheses have been suggested, including, for example, an origin from mammalian species other than cattle (a mutant form of scrapie agent, a natural TSE in Bovidae or Felidae or other wild animals whose carcasses were rendered into MBM, the existence of a form of sporadic TSE-like sporadic CJD of humans, a spontaneous mutation of normal bovine PrP into an infectious and protease-resistant TSE prion, etc.). For none of these hypotheses are there enough data either to substantiate or to reject it.
- In regard to routes of transmission other than those mentioned above (the so-called third route), not one of the many possibilities considered has been substantiated by scientific evidence so far.
- Factors that have been suggested as having a potential for affecting susceptibility of cattle to BSE include some metal ions (e.g., copper and manganese), organophosphorus compounds, and antioxidants. However, supporting data for all these hypotheses are very limited and do not allow any conclusions to be drawn.

Geographical Distribution of BSE and Related Diseases

Imported cases

In cattle BSE has occurred in a small number of cattle exported from countries with BSE (mostly the UK) to other countries. These countries are as far afield as Canada, the Falkland Islands, the Sultanate of Oman, and some European countries. In these countries, so long as cases occur only in imported animals, the surveillance for disease is sound, affected cattle are completely destroyed, infected cattle and their offspring are quarantined, and no infected organs are permitted to enter any food or feed chain, any residual risk is likely to be extremely small.

In other species FSE has been reported in a small number of cheetahs exported from the UK and imported into captivity in the Republic of Ireland, France, and Australia.

Cases in native-born animals

In cattle Up until June 2002, countries with BSE in native-born cattle, mostly dairy cattle, are (cumulative numbers of confirmed cases including cases found by active surveillance using 'rapid' tests): Austria (1), Belgium (81), Czech Republic (2), Denmark (8), Finland (1), France (567), Germany (163), Greece (1), Italy (54), Israel (1), Japan (3) Liechtenstein (2), Luxembourg (1), Netherlands (35), Northern Ireland (1977), Poland (1), Portugal (631), Republic of Ireland (953), Slovakia (5), Slovenia (1), Spain (137), Switzerland (416), UK (179 488) and other British Isles (1287). (Data courtesy of DEFRA and DARDNI for the UK and OIE/EC for other countries).

In other animal species A total of 37 cases of SE in captive wild Bovidae and Felidae and 90 cases of FSE in domestic cats have been confirmed in the UK to June 2002. One case each of FSE in domestic cats has been reported in Norway, Liechtenstein, and Switzerland. (Data courtesy of DEFRA, UK).

In humans Up until June 2002, there has been a total of 113 definite and probable cases of vCJD in the UK, four cases in France, and one each in the Republic of Ireland and Italy. Each of these countries has had BSE in native-born cattle. By contrast, in 2001 the total number of all forms of definite and probable CJD in the UK was 77 (sporadic 50, iatrogenic 3, familial 2, GSS 2 (all these completely separate from vCJD), and vCJD 20: data courtesy of Department of Health). Two or three further cases of vCJD have been confirmed in countries without BSE, but these are included in the UK total because the epidemiological evidence points to them having been exposed in the UK during residence there.

The SSC of the EC has classified a range of countries wishing to trade in bovine products with member states of the European Union (EU) and including all member states into one of four categories on the basis of an analysis of submitted dossiers containing relevant data. This analysis is called the Geographical BSE Risk (GBR) risk assessment. Although the list is dynamic, at the present time the UK and Portugal are in the highest risk category and countries like Argentina, Australia, and New Zealand that have reported no BSE and have good surveillance for the disease are in the lowest risk category. The OIE has defined five categories dependent upon the occurrence and incidence of BSE but has categorized no individual country. The two apparently different approaches are in fact very closely aligned and it is anticipated that eventually they will coincide. The EC is currently re-evaluating the GBR for over 80 countries to reflect the current situation more accurately and since BSE has been reported in widely separated regions of the world.

Processing of Waste Animal Tissues – Rendering

Epidemiological studies In the UK, MBM had been fed to cattle for decades before the first case of BSE was discovered in 1986, without any knowledge of a problem and despite the fact that scrapie-infected sheep tissues were in the starting material. At this time and until it was stopped many years later by EC decision, the starting material had included certain bovine TSE-risk byproducts from abattoirs. Also included were condemned materials and fallen stock. During the 1970s and early 1980s commercially motivated changes were occurring in the UK rendering industry.

First, there was a change from batch processing to continuous processing. But this did not result in an equivalent change in the processing parameters (such as processing at lower temperatures or for shorter times) that might reasonably have been associated with a reduced inactivation of a scrapie-like agent.

Second, there was a phasing-out of the hydrocarbon solvent extraction of tallow, particularly in the period of the putative first exposures to the BSE agent. This was believed to be critical. This change appears to have permitted an increased amount of TSE agent infectivity to escape inactivation, and this was sufficient to produce disease at a detectable incidence. However, subsequent laboratory inactivation studies using hydrocarbon solvents previously used to extract tallow, and the additional heat process used to recover the solvent, showed that these had little effect on TSE agent inactivation, though possibly just sufficient to restrict an effective oral dose being delivered. Nevertheless, some doubt was cast on the 'rendering' hypothesis, but no plausible alternative explanation has been proposed for the timing of the first effective exposures.

Experimental studies Experimental studies on animal waste processing, as conducted throughout the EU, showed that viable scrapie and BSE agents could survive some processes that were used in the countries where BSE occurred, thus supporting the notion that MBM was the vehicle of transmission. As a result of all these studies, the procedures adopted in the EU for processing mammalian waste now require the following parameters: a maximum particle size of 50 mm in the raw material, followed by processing (pressure cooking) at a minimum temperature of 133 °C at 3 bar pressure, for 20 min. However, laboratory studies suggest that no rendering process, including pressure-cooking, can guarantee complete inactivation of TSE agents.

A new European regulation to control animal waste is likely to become law by 2003 and covers all member states of the EU. Amongst other things, this determines that waste is divided into three categories of risk, each of which has a predetermined disposal method applied to it. The highest risk material would be incinerated and the lowest could only come from animals passed fit for human consumption and then be treated in approved ways before being permitted into animal feed or for other purposes. A major step forward was the earlier introduction, consolidated into the new regulation, of the rule whereby animal feed derived from animal byproducts can only use starting materials passed fit for human consumption

and processed by approved methods. Furthermore, within-species recycling of animal protein (other than milk and a few other exceptions) will not be permitted. Of course ruminants would not be permitted to receive protein from any animal species. In addition, there is a temporary ban on feeding animal protein to any food animal species, mainly to combat the problem of cross-contamination of ruminant feed and the occurrence of cases of BSE born after the feed ban was put in place.

Gelatin, Skull, and Vertebral Column

Gelatin is a product with a wide range of uses, including in food, confectionery, and medicines. The raw materials are animal bones or skin from pigs and cattle. The only starting material with which there is concern is bovine bone, not because of any inherent infectivity in it in clinically healthy animals, but rather because of the possible contamination with specified risk materials (SRM), particularly brain, spinal cord, and associated ganglia. Thus the bones most likely to be a risk are ruminant skull bones (that are SRM throughout the EU) and vertebral column, excluding the tail vertebrae. These, respectively, could easily retain residual brain and spinal cord and the vertebral column would certainly retain dorsal root ganglia in which TSE infectivity has been found in experimental BSE. Depending on the GBR category to which a country has been allocated by the EC, the risk bones are designated SRM and cannot be used for gelatin manufacture. In addition, the processing parameters must follow prescribed rules that have been shown experimentally to reduce the titer of spiked BSE material by substantial amounts. If this is done the residual TSE risk in gelatin is negligible.

Because of the same TSE concerns with the starting material (bovine skull and vertebral column) it is now prohibited to use ruminant bones in the EU to manufacture mechanically recovered meat that has been widely used in human food in the past and was thought by some to be a possible source of vCJD infection from cattle. All ruminant bones are now proscribed because it would be impossible otherwise to police the ban.

The Source of BSE Infection and Effect of the Removal of the Source

The coordinated epidemiological studies on BSE commenced in mid-1987 when there were fewer than 10 confirmed cases of the disease. By the end of the year, there had been almost 200 cases, sufficient to enable a working hypothesis for the origin to be drafted. Thus, the MBM hypothesis was established. The results from this study led to the introduction, in July 1988, of a major measure to control the disease, the ban on feeding ruminant protein to ruminant animals, with certain exceptions like milk protein.

Withdrawal of mammalian MBM from the rations of cattle and other Bovidae in 1988 effected a marked reduction in incidence of BSE and feed/BSE-related ruminant TSE in the UK, and strongly supports the hypothesis of an origin from TSE-infected raw material that somehow survived processing. The decline in incidence became clear from 1992–93 onwards; the delay was due to the mean 5 years' incubation period for BSE.

Descriptive Epidemiology

All breeds of cattle and both sexes appear to be equally susceptible; however, the disease affects mostly breeding dairy cows due to the methods of feeding the calves of these animals. Unlike beef breed calves that are suckled, dairy calves are removed from the dam after sucking colostrum. They are then fed artificial milk and later concentrates that, until 1988 in the UK, often contained MBM. Amongst dairy cattle in the UK, the incidence of disease between breeds is similar, but the number of confirmed cases by breed is directly proportional to the numerical size of the breed. Thus, most cases occur in Holstein-Friesian cattle and crosses with this breed, because they are the most abundant. Because of the widespread use of artificial insemination, bulls are much less numerous than cows and this largely explains why so few bulls (just under 500 in GB) have been affected by BSE compared with cows.

The risk of BSE occurrence in a herd increases with increase in herd size. This is attributed to the greater risk attached to purchasing an infected ration because in large herds more batches of feed would be bought. Of herds affected by BSE in the UK, nearly 35% have had only one case and over 80% have had seven or less. The mean within-herd incidence has remained more or less constant at less than 3% in any 6-month period, though it is now falling as the epidemic declines. Collectively these data and the results of a case-control study do not support contagion being important in BSE. Instead, it seems that infection and disease result from a random low-dose exposure to a common source of infection in MBM that is in 'packets' rather than being evenly distributed within feed, though this hypothesis is disputed by some workers.

Control and Elimination Measures

Animal Health Control Measures

The objective of member states of the EU and Switzerland is elimination of BSE from the territory.

A new regulation of the European Parliament and of the Council (EC no. 999/2001 of 22 May 2001) details the rules for the prevention, control, and eradication of TSE. Eradication (elimination of BSE from all countries in the world) is the eventual objective, but with several countries reporting BSE for the first time in the period 1999–2002, this task will not be achieved in the immediate future.

The main measures to protect ruminants from BSE are a ban on the inclusion of ruminant protein (subsequently mammalian MBM) in ruminant diets and the authorization only of the most effective rendering processes (pressure-cooking or methods giving equal or better inactivation of BSE and scrapie agents) where this is permitted for processing mammalian waste. Legislation to achieve the former has been adopted in the major cattle-rearing countries of Europe, North and South America, Australasia, and some other countries. Rendering process controls are in place throughout the EU.

In the UK a ban on the use of certain specified bovine offals (SBO) thought most likely to contain BSE infectivity in cattle incubating BSE, originally introduced to protect public health in 1989, was extended in 1990 to protect all species of animal and bird. More recently, the SRM ban was introduced in the UK to protect public and animal health. SRM include some tissues from sheep and goats. An EU-wide SRM ban was introduced in June 2000 and has subsequently been modified to improve its scope. SRM must be destroyed completely by approved processes.

The feed ban, introduced in its initial form in 1988 in the UK, has a sound scientific basis but was not completely effective in reducing exposure from feed sources, so that cases of BSE have occurred in animals born after the ban was introduced. Indeed, over 43 500 such 'born after the ban' or BAB cases have been reported in GB and most current cases of BSE in Europe have occurred in animals born after the date of feed bans applied in individual countries. In the UK, new exposures from feed in cattle born since 1 August 1996 should in theory be entirely prevented as a result of banning the use of mammalian MBM in the feed for any food animal species, including horses and fish. An even more extensive ban now applies throughout the EU and essentially prohibits the feeding of any animal protein, except milk protein, to food animals. The major stimulus to introduce these extended and strict bans is because cross-contamination of ruminant diets with MBM or diets for pigs and poultry containing MBM (nonruminant diets legitimately could, prior to the regulation, include MBM as a constituent) could not be prevented in feed mills, during transport, or on-farm. Research has shown that very small amounts (< 1 g) of brain material can carry a lethal infective oral dose for a cow.

The epidemics of BSE in the UK and Portugal are now declining towards elimination. It is hoped that Switzerland and all EU member states will follow as the extended bans become completely effective.

The OIE (World Organization for Animal Health) has produced an *International Animal Health Code* chapter on BSE that determines the requirements to facilitate safe trading in cattle and cattle products in regard to BSE throughout the world.

Public Health Control Measures

In addition to the SRM ban mentioned above, all cattle suspected of having BSE are compulsorily slaughtered and totally destroyed (since 1988 in the UK and 1990 in the EU) so they can enter no food or feed chain. This is now practiced in all countries and is endorsed by WHO that has also recommended global surveillance and diagnosis of all forms of CJD. The WHO collaborates with the OIE in recommending surveillance and monitoring for animal TSE, including BSE in all countries.

The UK BSE Situation Contrasted with that in other European Countries

It is clear that the total number of cases of BSE and the incidence of BSE in GB (> 179 000 to date) exceeds, by far, the number of cases and the incidence in other countries (collectively > 3000 to date). However, the annual number of suspect cases reported and confirmed in the UK has declined since 1992 and is still declining toward extinction and in line with predictions. As a result of passive surveillance, at the peak of the epidemic in GB in 1992 > 36 000 cases of BSE were confirmed. In 2001 only 781 were confirmed. There is greater than a 40% reduction in cases by year, though the epidemic is expected to have a long 'tail'. Furthermore, the residual risk to humans and animals is probably lower in the UK than anywhere else, because of the extensive interlocking measures applied and enforced to protect animal and human health. A significant measure applied after March 1996 in the UK is the scheme that prohibits the use for any purpose of products from cattle over 30 months (OTM) of age. A conditional exclusion is permitted for hides. This OTM scheme exists in no other country. However, cattle from herds in the UK beef assurance scheme may be slaughtered for human consumption up to 42 months of age. Because cattle in these closed herds are mainly grass-fed and have not received MBM in their diet, they present a low risk from BSE.

Some British and continental veterinarians have formally expressed in 1997 concern that there could be more cases of BSE occurring outside the UK than were being officially reported. Indeed, there have been studies into the expected incidence of BSE in cattle exported from the UK to other member states. It was concluded that it was likely that BSE cases had occurred in cattle in member states that had been imported from the UK. The (average) number of potential cases in animals imported between 1985 and 1989 was estimated to be 1642, a much higher number than had been officially reported to the EC or OIE. BSE cases in imported cattle could also have been the source of secondary cases of BSE. Such cases could have resulted from the recycling of infected cattle tissues through the rendering system (depending on the efficacy of the processing in the country in question) and then back to cattle via contaminated feed. These predictions have come true.

Following the announcement of 10 cases of vCJD in March 1996, several countries reported an increase in the number of BSE cases. For example, all three Benelux countries reported their first BSE cases in 1997 and Denmark, Germany, and Spain their first native-born case in 2000, and several other European countries followed suit during 2001 and 2002. BSE has now been reported in Japan and Israel. The number of confirmed cases currently being reported in these countries is small and less than the number currently being reported in the UK. However, the epidemics in most countries, as judged by the numbers of cases reported in successive years and identified by both passive and active surveillance, are increasing, whereas those in the UK and Portugal are declining. The occurrence of clinical cases refers back to exposures 5 years previously, so it is quite possible that new infections in most European countries are currently very low and will be reflected as a declining incidence in the coming years.

Feed Bans in other European Countries

It is only since 1994 that a uniform MBM feed ban (for ruminants) and rendering control has been determined by EC legislation. Switzerland (who alone adopted an SBO ban like the UK in 1990), Denmark, France, Ireland, the Netherlands and some other countries imposed a national ruminant feed ban from 1990. In at least one country, the ban was applied to cattle feed only, and not to other ruminant feed, thus potentially exposing sheep and other ruminants to infected MBM via feed until the law was changed some years later. All other member states with BSE in native-born cattle have had, like the UK, cases in animals born after the respective dates of the feed ban in their own country. Most agree that these probably result from cross-contamination of ruminant feed with infected MBM intended for monogastric species (mainly pigs and poultry). Mammalian MBM, including ruminant MBM, was permitted until 1 January 2001 for feeding to non-ruminant farm animals in all member states except the UK and Portugal.

Offal Bans in other European Countries

In 2000, uniform SRM controls in the EU were introduced, though several countries with BSE in native-born animals had some form of ban prior to this. However, some member states with no BSE in native-born animals until recently had no SRM ban at all, despite the known risks. This meant that it was possible that consumers of permitted but dangerous offals (SRM) in those countries had risked exposure to BSE if the offals were derived from incubating animals. Nowhere were the various offals/materials bans so extensive as in the UK. However, before 1996 in the UK, compliance with and enforcement of the ban were incomplete. There was leakage of some infected material, largely into waste destined for rendering, which was recycled to pigs and poultry via MBM. This contributed to an extension of the cattle epidemic because of accidental cross-contamination of ruminant diets with MBM intended only for non-ruminants. There is no clear indication that such bans, where they existed, were as rigorously enforced as they are now in the UK. The EU-wide SRM ban in place since 2000 will only be effective if it is completely and continuously enforced. However, Rapid tests applied to cattle over 24 or 30 months old in EU abattoirs and destruction of the whole carcasse of those that fail the test are added consumer safeguards.

Passive Surveillance

For the most part, the surveillance for BSE in all member states is passive, relying on the clinical identification of suspect animals by farmers and veterinarians. This is followed by compulsory slaughter and by microscopic examination of brains to confirm or deny a diagnosis of SE, according to standard criteria set down by the EC and the OIE. Such surveillance relies for its success on the vigilance of the observers and awareness of the clinical signs of the illness. Such surveillance is likely to underestimate the real number of cases and cannot detect preclinical cases (or, if they exist, subclinical cases) of BSE.

Active Surveillance

As a result of research into methods of detecting PrP in the brain or spinal cord of BSE-affected cattle, a number of confirmatory tests have been developed

and evaluated by the EC. Using one of these tests, in January 1999, the Swiss State Veterinary Service initiated a targeted, active surveillance system for three categories of cattle. These are fallen stock, cattle slaughtered in emergency, and a proportion of healthy cattle slaughtered for human consumption. Twelve out of 4868 tested fallen stock were positive and four out of 2274 tested emergency slaughter animals were positive. Two out of 4847 healthy slaughter animals were positive, probably because they were in the last few months of incubation. Thus, 18 additional cases of BSE that otherwise would not have been detected have been identified by this method before the end of 1999. This was about the same low number of cases detected by conventional passive surveillance in that period. The underreporting of BSE in 1999 in Switzerland was thus estimated to be about 50%.

Future Outlook

A successful EC-funded comparative evaluation of four Rapid tests for PrP using brain or spinal cord tissue for the postmortem confirmation of BSE in clinically affected cattle has been reported. The approved tests are now being used throughout the EU on fallen stock and cattle sent for emergency slaughter over 24 months of age and on all cattle for slaughter for human consumption over 30 months of age. Some countries test all animals for human consumption over 24 months of age. Any animals that fail the test must be completely destroyed. Application of several million tests has already identified several hundred additional animals with BSE infection that would otherwise have escaped detection. Results of tests must be regularly reported to the EC and are published. The true incidence of disease and the extent of underreporting are thus being revealed.

Though not currently validated for use in clinically healthy cattle, studies are in progress to determine how effective some of these tests are for detecting preclinical cases of BSE. This will be done by using CNS tissue derived from a study of the pathogenesis of experimental BSE in cattle. In contrast to the situation in scrapie-infected sheep, the distribution of infectivity in cattle with naturally occurring BSE is restricted to inaccessible tissues even in the clinical phase of disease, namely the brain spinal cord and the retina. In the experimental disease, infectivity has been detected in the distal ileum in the preclinical phase of disease, and in central ganglia and possibly in bone marrow during the clinical phase. Because none of these tissues is easily accessible, the prospect of a reliable and practical test being soon available to detect preclinical BSE in live cattle is remote.

Conclusions

The epidemic of BSE in the UK is clearly under control and heading for elimination, as predicted. It is probably similarly heading for elimination in Portugal. All products permitted for sale from cattle slaughtered in the UK in 2002 are as safe, if not safer in regard to the BSE risk, as those are from any other country with BSE. The situation in other countries is less certain but, provided all those countries at risk adopt the lessons learned in the UK, Portugal, and Switzerland, the future prospects for the elimination of BSE are also good. The critical issue is to insure that ruminant feed is not contaminated with infected MBM. This is a difficult task and interlocking and enforced measures are essential to success. Also essential is to conduct surveillance of targeted populations of cattle using the most sensitive and specific tests available.

Surveillance for all forms of CJD in humans is also vital. Of current concern is the future size and duration of the vCJD epidemics and of the possible occurrence of scrapie in sheep caused by the BSE agent. There is no evidence for the latter. In regard to the former, predictions are not possible due to several uncertainties, including the length of the incubation period, the number of humans exposed, and the precise source of infection. Concerns have been expressed about the possible transmission of vCJD from human to human, particularly via blood transfusion or surgical instruments. Currently, there is no evidence for transmission of vCJD by any means between humans, but nevertheless, strict risk management procedures have been put in place to reduce any possible risk.

See also: **Beef**; **Offal**: Types of Offal

Further Reading

Anderson RM, Donnelly CA, Ferguson NM *et al*. (1996) Transmission dynamics and epidemiology of BSE in British cattle. *Nature*, 382: 779–788.

Bradley R (1997) Animal prion diseases. In: *Prion diseases*. Collinge J and Palmer MS (eds) pp. 89–129. New York: Oxford University Press.

Bradley R (2001) Bovine spongiform encephalopathy and its relationship to the new variant form of Creutzfeldt-Jakob disease. An account of BSE, its cause, the clinical signs and epidemiology including transmissibility of prion diseases with special reference to the relationship between BSE and vCJD. In: *Prions: A challenge for science, medicine and public health system. Contributions to Microbiology*, vol. 7. Karger Basel, Rabenau HF, Cinatl J and Doerr HW (eds) pp. 105–144.

Brown P (2001) Bovine spongiform encephalopathy and variant Creutzfeldt-Jakob disease. *British Medical Journal*, 322: 841–844.

Bruce ME, Will RG, Ironside JW *et al*. (1997) Transmissions to mice indicate that 'new variant' CJD is caused by the BSE agent. *Nature* 389: 498–501.

Clarke AR, Jackson GS and Collinge J (2001) The molecular biology of prion propagation. *Philosphical Transactions of the Royal Society of London B*. 356: 185–195.

Collee JG and Bradley R (1997) BSE a decade on – Part 1. *Lancet* 349: 636–641.

Collee JG and Bradley R (1997) BSE a decade on – Part 2. *Lancet* 349: 715–721.

Collinge J, Sidle KCL, Meads J, Ironside J and Hill AF (1996) Molecular analysis of prion strain variation and the aetiology of 'new variant' CJD. *Nature* 383: 685–690.

Defra (2001) Bovine spongiform encephalopathy in Great Britain. *A Progress Report, December 2001*. London: DEFRA.

Doherr MG, Oesch B, Moser M, Vandevelde M and Heim D (1999) Targeted surveillance for bovine spongiform encephalopathy. *Veterinary Record* 145: 672.

European Commission. Transmissible spongiform encephalopathies (1994) Protocols for the laboratory diagnosis and confirmation of bovine spongiform encephalopathy and scrapie. A report of the Scientific Veterinary Committee. EC, Brussels.

European Commission (2001) Regulation (EC) No. 999/2001 of the European Parliament and of the Council of 22 May 2001 laying down rules for the prevention, control and eradication of certain transmissible spongiform encephalopathies. EC, Brussels.

European Commission: Opinion of the Scientific Steering Committee (2001) Opinion on: Hypotheses on the origin and transmission of BSE. Adopted by the Scientific Steering Committee at its meeting of 29–30 November 2001. EC, Brussels.

European Commission: Opinion of the Scientific Steering Committee. Updated Opinion on: The safety with regard to TSE risks of gelatine derived from ruminant bones or hides from cattle, sheep or goats. Adopted by the Scientific Steering Committee at its meeting of 28–29 June 2001. EC, 2001, Brussels.

Hegde RS, Mastrianni JA, Scott MR *et al*. (1998) A transmembrane form of the prion protein in neurodegenerative disease. *Science* 279: 827–834.

Hill AF, Desbruslais SJ, Joiner S *et al*. (1997) The same prion strain causes vCJD and BSE. *Nature*, 389: 448–450.

Hope J, Reekie LJD, Hunter N *et al*. (1998) Fibrils from brains of cows with new cattle disease contain scrapie-associated protein. *Nature* 336: 390–392.

Horn G Review of the origin of BSE. 5 July 2001, pp 66. London: DEFRA.

Hunter N, Goldmann W, Smith G and Hope J (1994) Frequencies of PrP gene variants in healthy cattle and cattle with BSE in Scotland. *Veterinary Record* 135: 400–403.

Kimberlin RH and Wilesmith JW (1994) Bovine spongiform encephalopathy epidemiology, low dose exposure and risks. *Annals of the New York Academy of Sciences* 724: 210–220.

Moynagh J and Schimmel H (1999) Tests for BSE evaluated. *Nature* 400, 105.

Oie (2000) *Manual of standards for diagnostic tests and vaccines*. pp 954. Paris: OIE.

Oie (1992) *Transmissible spongiform encephalopathies of animals*. Bradley R and Matthews D (eds). Scientific and Technical Review, 11. pp 634. Paris: OIE.

Oie Bovine spongiform encephalopathy. *OIE International Animal Health Code, Chapter 2.3.13*., pp 11. Paris: OIE.

Rabenau HF, Cinatl J and Doerr HW (2001) Prions: A challenge for science, medicine and public health system. *Contributions to Microbiology*, 7. pp 164. Karger, Basel.

Safar J, Wille H, Itri V *et al*. (1998) Eight prion strains have PrP^{Sc} molecules with different conformations. *Nature Medicine* 4: 1157–1165.

Schreuder BEC, Geertsma RE, van Keulen LJM *et al*. (1998) Studies on the efficacy of hyperbaric rendering procedures in inactivating bovine spongiform encephalopathy (BSE) and scrapie agents. *Veterinary Record* 142: 474–480.

Schreuder BEC, Wilesmith JW, Ryan JBM and Straub OC (1997) Risk from BSE from the import of cattle from the United Kingdom into countries of the European Union. *Veterinary Record* 141: 187–190.

Stevenson MA, Wilesmith JW, Ryan JBM, Morris RS, Lockhart JW, Lin D and Jackson R (2000) Temporal aspects of bovine spongiform encephalopathy in Great Britain: individual animal-associated risk factors of the disease. *Veterinary Record* 147: 349–354.

Taylor DM, Woodgate SL and Atkinson MJ (1995) Inactivation of the bovine spongiform encephalopathy agent by rendering procedures. *Veterinary Record* 137: 605–610.

Taylor DM, Woodgate SL, Fleetwood AJ and Cawthorne RJG (1997) Effect of rendering procedures on the scrapie agent. *Veterinary Record* 141: 643–649.

Wells GAH, Hawkins SAC, Green RB *et al* (1998) Preliminary observations on the pathogenesis of experimental bovine spongiform encephalopathy (BSE): an update. *Veterinary Record*, 142: 103–106.

Wells GAH, Scott AC, Johnson CT *et al* (1987) A novel progressive spongiform encephalopathy in cattle. *Veterinary Record* 121: 419–420.

Wells GAH and Wilesmith JW (1995) The neuropathology and epidemiology of bovine spongiform encephalopathy. *Brain Pathology* 5: 91–103.

Wilesmith JW, Ryan JBM and Atkinson MJ (1991) Bovine spongiform encephalopathy: epidemiological studies on the origin. *Veterinary Record* 128: 199–203.

Wilesmith JW, Ryan JBM, Stevenson MA *et al* (2000) Temporal aspects of bovine spongiform encephalopathy in Great Britain: holding-associated risk factors for the disease. *Veterinary Record* 147: 319–325.

Wilesmith JW, Wells GAH, Cranwell MP and Ryan JBM (1988) Bovine spongiform encephalopathy: epidemiological studies. *Veterinary Record* 123: 638–644.

Wilesmith JW, Wells GAH, Ryan JBM, Gavier-Widen D and Simmons MM (1997) A cohort study to examine maternally-associated risk factors for bovine spongiform encephalopathy. *Veterinary Record* 141: 239–243.

Wrathall AE, Brown KFD, Sayers AR *et al*. (2002) Studies of embryo transfer from cattle clinically affected by bovine spongiform encephalopathy (BSE). *Veterinary Record* 150: 365–378.

BRANDY AND COGNAC

Contents

Armagnac, Brandy, and Cognac and their Manufacture

A Bertrand, Université de Bordeaux II, Talence Cedex, France

Brandy

Regulations

European legislation (EC regulation no. 1576/89) distinguishes between wine spirits and brandy.

Wine spirits According to regulations, wine spirits is a spirit drink, obtained exclusively by distilling wine or 'winey' wine (vin viné) to 86% vol, or by redistilling wine distillates to less than 86% vol (wine distillates are halfway between spirits and ethyl alcohol; they must have retained the flavor and aroma of wine).

Wine spirits have a volatile substance content equal or superior to 125 g hl^{-1} of 100% vol alcohol (volatile substance content includes higher alcohols, esters, aldehydes, volatile acids, and furfural). Maximum methanol content is 200 g hl^{-1} of 100% vol alcohol. When aged, this beverage can continue to be marketed with the designation 'wine spirits' if the period of aging is equal or superior to the period provided for brandy.

According to this rule, general designations (20 different kinds of spirits) can be completed by geographical indications. The most famous French geographical designations are Cognac and Armagnac.

Brandy Brandy is a spirit drink which is obtained from wine spirits either blended or not with wine distillates distilled to less than 94.8% vol, on condition that such distillates do not exceed maximum 50% proof in the finished product.

Brandy is aged in oak containers for at least 1 year, or for a minimum period of 6 months if the capacity of the oak casks is less than 1000 l.

Brandy has a volatile substance content equal or superior to 125 g hl^{-1} of 100% vol alcohol resulting exclusively from the distillation or redistillation of the raw materials brought into play, and with a maximum methanol content of 200 g hl^{-1} of 100% vol alcohol.

Distillation

As a rule, brandies are distilled in columns containing several dozen plates. Old stills were made entirely of copper, but because the sulfur dioxide in the wine corrodes the stills and particularly the 'phlegm' collector plates, copper has gradually given way to stainless steel; today, copper is only used for the topmost parts of the rectifying columns (to draw the spirits).

Wine rectifiers For a long time, only phlegms were rectified, that is, the distillery carried out two successive operations: (1) wine distillation and recuperation of phlegms; (2) rectification of phlegms.

With advances in the conception and building of distilling equipment, continuous rectifiers now exist which process the fermented wine, making it possible to obtain 96.5 ° rectified alcohol in a single operation.

Wine rectifiers can be divided into three categories: indirect, direct, and semidirect appliances.

Indirect rectifiers (**Figure 1**). The term indirect means that rectification involves a traditional distillation process, except that phlegms are not cooled and

sent into trays but go directly into the continuous rectifier on the end of the column.

A is the distillation column producing the phlegm; while hot, phlegm is then extracted from the distillation column and sent directly to the purifying column (B) of the continuous phlegm rectifier.

The purifying column extracts the head foreshots, after which the purified phlegm is sent to the rectifying column (C) from which the drinkable pasteurized alcohol is drawn off; the unpasteurized alcohol, which is downgraded, and the low- and high-fusel oils are also extracted.

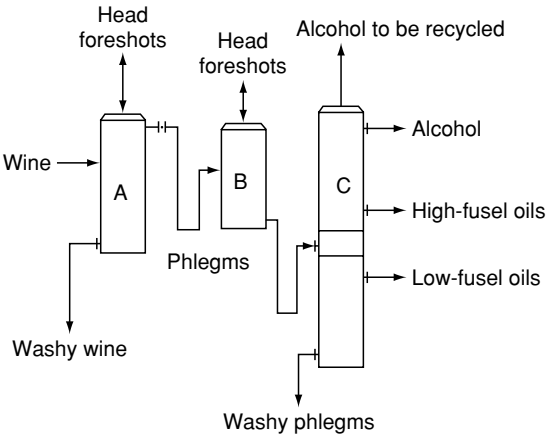

Figure 1 Indirect continuous wine rectifier. A, distillation column; B, purifying column; C, rectification column. Reproduced from Mariller C (1925) *Distillation et Rectification des Liquides Industriels*. Paris: Dunod, with permission.

Special devices have been added to the columns to recycle the most volatile substances which are of organoleptic interest (Figure 2).

This specific draw-off determines the final amount of tailings in the distillate.

For the subsequent production of wine spirits, only the higher alcohols are partially eliminated (high- and low-fusel oils) in order to recompose the distillate as desired. In this case, the resulting alcohol can only be called distillate and not 'spirits', because of the high percentage of alcohol – higher than 86% vol. – required for sorting heavy impurities.

Methanol is extracted on a special multiple-plate column (about 50 plates). This column has the disadvantage of eliminating not only the methanol but also volatile compounds of organoleptic interest; a small column can be added which separates the methanol and the volatile substances, and the latter are reintroduced into the distillate.

Batch rectification for the production of wine spirits or distillate This kind of still has a boiler similar to that of the Charente still or of cylindrical boiler stills. Primary spirits can be slightly oxidized before redistillation. This batch rectier comprises a steam-producing device, a 30-plate column, and cooler satellites and circuits. Head and tail products can be recycled in the boiler (Figure 3).

Composition of Brandies

Because of the nature of the raw material brought into play (usually red wine) and how they are made, brandies contain far fewer volatile substances than

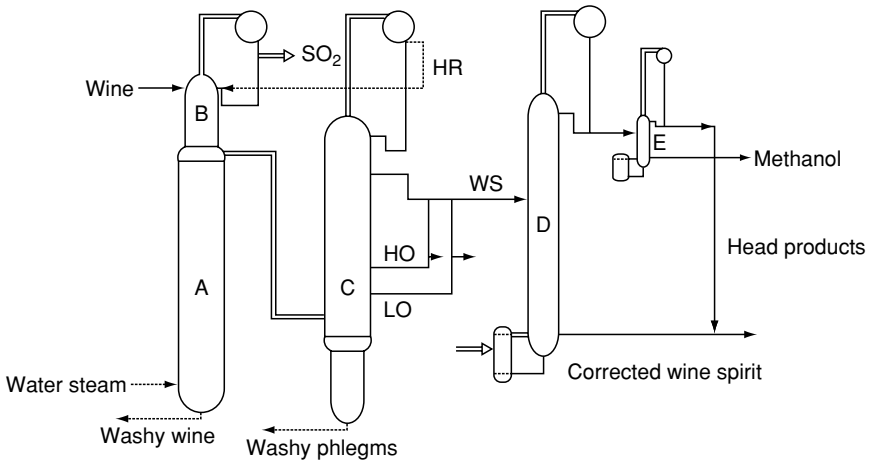

Figure 2 Continuous wine rectifier with head foreshot recuperation. A, distillation column; B, desulfitation column; C, concentration column; D, demethanolation column; E, column for the separation of esters from methanol (head products); HR, head product recycling; LO, low-fusel oils; HO, high-fusel oils; WS, wine spirit. Reproduced from Guigon and Cogat P (1991) Elimination des défauts dans les eaux-de-vie de vin traditionelles. In: *Les Eaux-de-vie Traditionelles d"Origine Viticole*, pp. 110–113. Paris: Lavoisier-TEC & DOC, with permission.

Appellation d'Origine Contrôlée (AOC) spirits (Table 1); higher alcohols are rectified by specific fusel oil separation processes; head products lose most of their ethyl acetate and esters during the ethanal and sulfur dioxide elimination process.

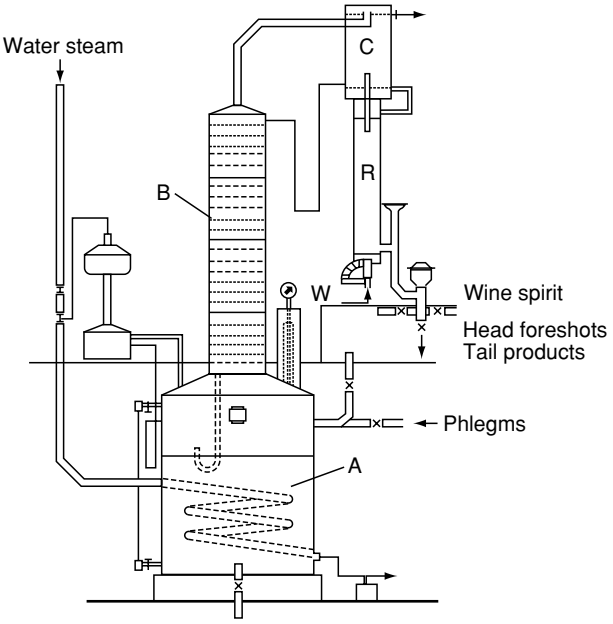

Figure 3 Batch rectifying still for the production of wine spirit or distillates. A, boiler; B, distillation column (30 plates); C, condenser; R, refrigerator (cooler); W, water. Reproduced from Guigon D and Cogat P (1991) Elimination des défauts dans les caux-de-vie de vin traditionelles. In: *Les Eaux-de-Vie Traditionelles d'Origine Viticole*, pp. 110–113. Paris: Lavoisier-TEC & DOC, with permission.

Making brandy is usually a way of salvaging defective wines or production surpluses. However, there are some good-quality brandies made from wines which are specially grown for the purpose and vinified with limited quantities of SO_2.

Gas chromatography is used to analyze volatile compounds. As an example, the chromatogram in Figure 4 shows some of the volatile alcohols, esters, acids, and various aromatic components contained in an ether-hexane extract of a brandy. Other analyses can be carried out by injecting the spirits directly into the chromatograph. The high levels of diethyl succinate indicate that the wine distilled was older than the wine used for Cognac and Armagnac.

Aging and Merchandising Preparations

As specified in the law, brandies are aged in wood containers (sous bois); different processing methods can be used: cold processing, ion exchange resins to eliminate sulfur dioxide, calcium, and copper cations; sugar, caramel, and boisé (infusions from oak shavings) may be added. Brandies are reduced with deionized water and must have a minimum 37.5% vol alcohometric titer.

The entire production of French brandy is exported.

Summary

Wine spirits of viticultural origin belong to French cultural heritage; they also constitute a major part of our economy.

Furthering our knowledge of their composition reveals that traditional wine spirits have exactly the

Table 1 General composition of different wine spirits. Results are given as g hl^{-1} pure alcohol (100% vol) of an average 15 samples

	Armagnac		*Cognac*		*Brandy*[a]	
	Average	*SD*	*Average*	*SD*	*Average*	*SD*
Alcohol strength (% vol)	41.4	1.6	40.04	0.75	45.46	11.1
Dry extract (g l^{-1})	4.5	3.5	6.7	3	8.43	1.58
Total acidity (as acetic acid)	153.9	57.6	103.6	28.2	31.46	8.23
Volatile acidity (as acetic acid)	106.5	37.5	59.3	19.3	19.06	3.53
Aldehydes (as ethanal)	23.3	6.4	19.3	8.25	25.33	7.32
Esters (as ethyl acetate)	109.6	34.7	72.9	7.2	54.8	6.72
Among esters, ethyl acetate	78.3	24	45	5.9	38.5	4.2
Furfural	1.2	0.8	2.45	0.93	0.35	0.14
Higher alcohol	441.4	42.3	444.4	127.5	258.4	23.28
Butan-2-ol	0.5	0.85	0.7	1.6	3.39	1.64
Propan-1-ol	49.4	13.5	43	7.8	25.06	2.22
2-Methylpropan-1-ol	104.5	19	121.7	18.8	55.43	5.09
Butan-1-ol	0.2	0.5	0.1	0.3	1.34	0.67
2-Methyl + 3-methylbutan-1-ol	286.6	33.2	312.3	29.6	172.73	20.46
Sum of volatile substances[b]	682.1	83.6	632	42.2	357.5	31.36
Methanol	47	10.9	49.7	11.4	69.2	16.5

[a]Most brandies are not yet ready for consumption.
[b]Ethyl acetate is included with esters.

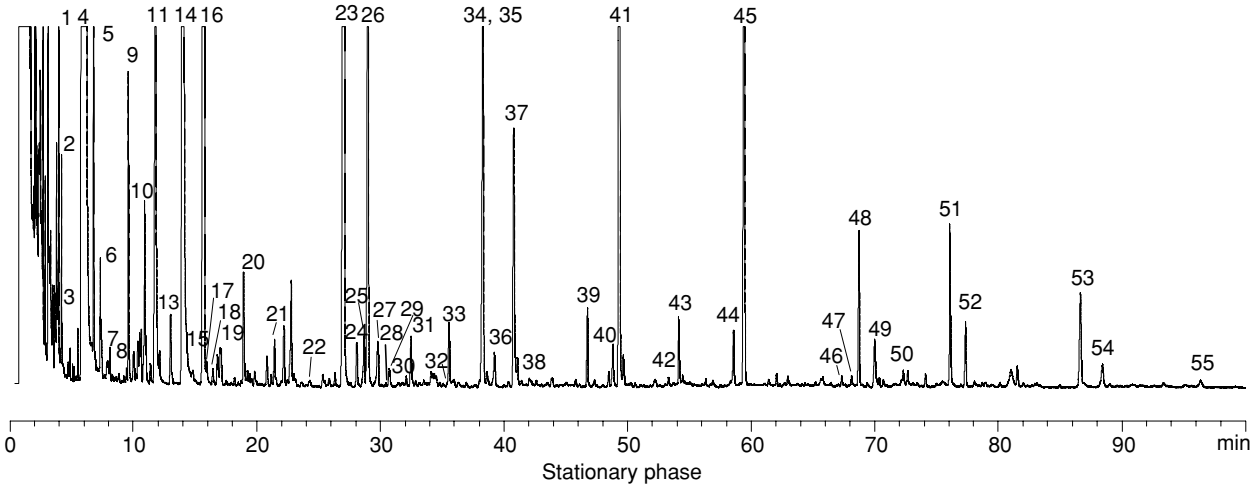

Figure 4 Chromatogram of the extract of a brandy by ether hexane (concentration about three times); column FFAP 50 m × 0.22 mm; injection by splitless mode; temperature programming from 40 to 200 °C. Identification of the peaks: 1, Ethyl butyrate; 2, 2-methylpropan-1-ol; 3, isoamyl acetate; 4, isoamyl alcohols; 5, ethyl hexanoate; 6, hexyl acetate; 7, styrene; 8, acetoin; 9, ethyl heptanoate; 10, ethyl lactate; 11, hexan-1-ol; 12, trans-hex-3-en-1-ol; 13, *cis*-hex-3-en-1-ol; 14, octan-3-ol (internal standard 1); 15, *trans*-hex-2-en-1-ol; 16, ethyl octanoate; 17, *trans*-linalol oxide (furane); 18, *cis*-linalol oxide (furane), 19, acetic acid; 20, benzaldehyde; 21, linalol; 22, 2-methylpropionic acid; 23, ethyl decanoate; 24, butyric acid; 25, 3-methylbutyric acid; 26, diethyl succinate; 27, α-terpineol; 28, 1,1,6-trimethyl-1,2-dihydronaphthalene (TDN); 29, methionol; 30, unknown; 31, unknown; 32, damascenone; 33, phenylethyl acetate; 34, ethyl dodecanoate; 35, hexanoic acid; 36, benzyl alcohol; 37, 2-phenylethanol; 38, heptanoic acid; 39, 4-ethylgaiacol; 40, ethyl myristate; 41, octanoic acid; 42, 4-allylgaiacol (eugenol); 43, 4-ethylphenol; 44, ethyl palmitate; 45, decanoic acid; 46, ethyl stearate; 47, ethyl oleate; 48, dodecanoic acid; 49, ethyl linoleate; 50, ethyl linolenate; 51, unknown; 52, myristic acid; 53, palmitic acid; 54, palmitoleic; 55, linoleic acid.

same composition as wine; only the heaviest and most polarized products are rectified (acetic acid, phenyl ethanol, polyols).

Major defects can only be eliminated with a distillation column, but at the cost of eliminating components which contribute to quality.

Good-quality wine spirits can only be made with good-quality wines; expensive aging in oak casks should be reserved for noble products.

Armagnac

Historical Background

According to scholars, Armagnac is the oldest known wine spirit; it has been produced without interruption since the early fifteenth century.

The name is the result of successive transformations of the word 'arminiacum,' meaning owned by or estate of Arminius. Arminius may be the latinized form of Hermann, a Saxon warrior and a companion of Clovis, the Frank king, who crossed the Rhine in 406 to wage war in the south-west of France. A Count of Armagnac is recorded in about 1032.

As in many other regions of France, the vineyards developed considerably under the Romans, until their expansion was temporarily interrupted by a decree of Emperor Domicien in 92 (in fact, half of the vines were pulled up and not planted again until 267, when Emperor Probus published an edict authorizing wine growing in Gaul). Vines continued to be planted all through the barbarian invasions and developed essentially around the monasteries. A tithe levied by the monks from the Abbey of Saint Mont (Gers) testifies that wine was being produced in the tenth century.

In 1254, Bayonne was actively trading with merchants from northern Europe; wines from Chalosse and probably from the west of Armagnac were shipped down the Adour. Wines from Haut Armagnac passed in transit on their way to Bordeaux, but, in 1241, the Bordeaux authorities decided to regulate, to their advantage, the Haut-Pays wines arriving in Bordeaux.

For a region so far removed from any natural outlet to the sea, distillation proved very early to be a reasonable solution, with smaller volumes to transport and without the quality of the wine deteriorating. This explains why the Comté of Armagnac became the first region of France to develop distillation. At the time, most of Aquitaine was English and wine growing was certainly favored by the flourishing trade between England and Northern Europe, as well as by the pilgrimage to Santiago de Compostela.

In 1411, the archives of the Haute-Garonne show that a distiller, M Antoine (aygua ardentarius), was producing a wine spirit, 'aygue de Bito,' in Toulouse; 28 years later, in 1439, the same source mentions other distillers: Jean Nouvel and his wife 'facientes aquam ardentum.' One document, dated 1461, mentions the levy of taxes, which testifies to the beginning of an Armagnac trade in Saint-Sever (Landes). In 1489, a document in the Gers archives mentions the presence of a still in Solomiac (Gers). By 1550, both Bordeaux and Bayonne were selling Armagnac to northern Europe.

Until the eighteenth century, wines were distilled in pot stills; these were simple devices similar to those described by Savonarole in 1440 in his treaty on distillation, 'Confidencia aquae vitae,' in which he describes the first copper still with a coil plunged in cold water. It was only toward the mid eighteenth century that distillation methods began to be improved. In 1761, with the help and advice of the chemist Chaptal, Menier invented a new process known as continuous distillation, and this process was patented by Adam in 1801.

This new distillation technique rapidly spread to the whole Armagnac area and is still the method used today, with few alterations. However, the type of still used in Charentes had to gain formal approval again in 1972.

The French Revolution boosted production: 50 000 hl of pure alcohol were produced in 1810, and over 100 000 hl in 1873, corresponding to 107 000 ha of vines. At that time, the Armagnac vineyard was one of the largest in France but in 1879 it was hit by phylloxera; 10 years later, alcohol production had been reduced by two-thirds. The vineyard was never to be entirely replanted and present-day production wavers between 20 000 and 50 000 hl per year as pure alcohol.

Appellation Areas, Soils, Climate, Vine stocks

The region of production covers about 12 000 ha, divided into three areas: Bas-Armagnac, Ténarèze, and Haut Armagnac (Figure 5).

Bas-Armagnac Bas-Armagnac is the most productive area – about 60% of the vineyard – with acid, predominantly sandy soils. Conversely, the soil in Haut-Armagnac is mainly composed of calcareous clay and has virtually ceased producing wine spirits. A blend of wine spirits from several minor appellations can only be called plain Armagnac.

Bas-Armagnac is a flat area, on the edge of the Landes forest, and the slightly acid, siliceous clay soil often contains iron oxides which endow it with a dark color – hence the name 'fawn sands.'

The annual mean temperature is 13 °C, with 7.5 °C in winter and 20 °C during the three summer months. Annual rainfall is 892 mm. The rain is well spread out over the year, with 1 mm minimum in July and 86 mm maximum in May.

Ténarèze The hilly landscape of the Ténarèze area consists of a variety of soils: sand, outcrops of calcareous clay, and boulbènes (a mixture of clay and fine sand) in the valleys. Being further away from the ocean, and especially from the Landes forest, the climate is a little more 'continental,' entailing atmospheric disturbances and more rainfall.

Mean annual rainfall is 804 mm, the main difference from Bas-Armagnac being the frequent thunderstorms in October.

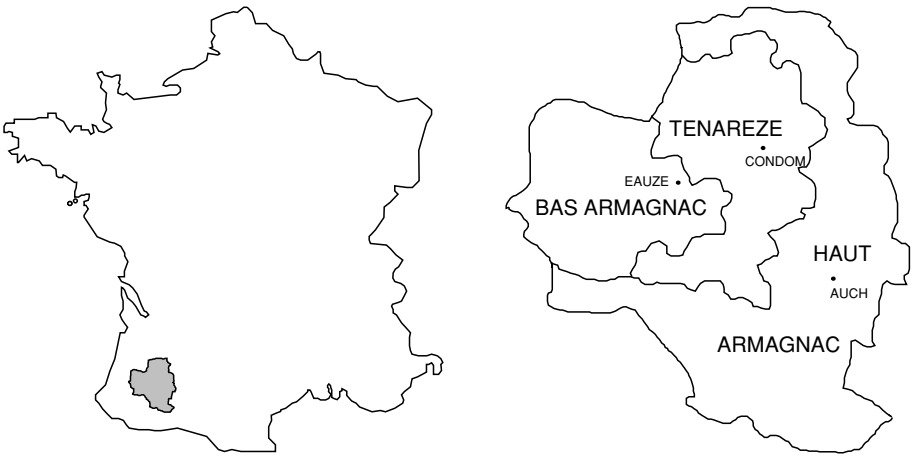

Figure 5 Armagnac in France and the three appellation areas.

Likewise, temperatures range between 7.2 °C in winter and 26 °C in summer. This difference is due to wind conditions; the area is still exposed to the east wind (vent d'Autan), which brings the end of summer droughts and storms.

Haut-Armagnac In Haut-Armagnac, soils are more calcareous and the climatic differences even more pronounced. The average period of sunshine in Auch is 1934 h, surprisingly, this is 5% less than in Bordeaux, perhaps because of a longer period of cloud cover.

The main different vine stocks allowed are Ugni blanc, Baco 22 A, Folle Blanche and Colombard.

Ugni blanc, an Italian vine stock related to Trebbiano, is rapidly gaining ground, as it offers the advantage of coming to maturity in the 'troisième époque' (45 days after Chasselas maturation), thus eliminating the problem of rot, and it produces low-alcohol wines suitable for distillation. Baco 22 A, which ranks second for making Armagnac, is a hybrid of Noah (*Vitis labrusca* × *V. riparia*) and Folle Blanche; it matures 30 days or more after Chasselas ('maturité de deuxième époque'), is very disease-hardy and does not require grafting thanks to the sandy soil of the Landes. This variety stands out as an exception among the vine stocks used for French AOCs. The Institut National des Appellations d'Origine des Vins et Eaux-de-Vie (INAO: French National Institute for Wine and Spirits Appellations of Origin) is planning to eliminate it from the Armagnac vineyard by 2010. The rather 'foxy' flavor of Baco wines, due to the presence of furaneol and methylic anthranilate ('*labrusca*' character), disappears during distillation and neither substance is detectable in the spirits.

The vineyards are of average size; pruning is managed according to the Guyot method in one or two stages; the vines are trellised and there are 4000–5000 vines per hectare.

Vinification

Most vinegrowers use a grape harvester to harvest the grapes and a continuous wine press to extract the juice; such methods, which are usually too brutal for immature grapes, have no particular drawbacks in this case as long as the dregs are not strained too much.

The temperature of alcoholic fermentation is rarely checked. Wines are not subjected to any enological process, and sulfur dioxide is strictly forbidden (Decree of 6 August 1936, modified 24 May 1956). The alcohol strength of the wine varies considerably, ranging between 8 and 11.5% vol or more; acidity is average (4–6.5 g l^{-1}, expressed as H_2SO_4). Malolactic fermentation usually occurs spontaneously just after alcoholic fermentation.

According to the Bureau Interprofessionnel de l'Armagnac (BNIA: French Interprofessional Office for Armagnac), 60 922 hl of pure alcohol were distilled in 1990, from 567 000 hl of wine. Production this year has been particularly good. The average annual production is about 20 000 hl 100% vol alcohol.

By law, wines have to be analyzed in an appointed laboratory to prove they are free of sulfur dioxide and of sufficient quality to be used to make Armagnac.

Distillation and Regulations

Wines have to be distilled in the appellation area. Two sorts of stills are used; the BNIA has listed the following ones which are in use:

1. 112 continuous stills: Armagnacais
2. 12 pot stills, producing less than 10% of the total volume of Armagnac.

The maximum distillation alcohometric titer allowed is 72% vol, the same as other French AOC wine spirits (Decree of 6 August 1936). The minimum alcohometric titer for white spirits (when they come out of the still) is set at 52% vol.

Wines are required to be distilled between the end of the harvest and 31 March of the following year (Decree of 15 March 1988, modified).

The continuous Armagnac still (Figure 6) The continuous still used in Armagnac is made entirely of annealed electrical-grade copper; it works very much like steam-powered distillation and plays an essential role in the specificity of armagnac.

The boilers, distillation columns, wine-heater, and cooler constitute the still's main parts.

The volume of the boiler ranges from 5 hl to a maximum of 35 hl; it is divided into two or three sections by separation plates. The total capacity of the boilers must be at least equal to that of the cooling unit, including the wine-heater and the cooler.

The column has five to 15 plates (the most recent stills have about 12 plates). These plates are fitted with different-shape bubble-through devices: bubble caps, bell-shaped tunnels or oil grooves. The plates above the wine arrival pipe are called dry plates and help reduce tailings while increasing the alcohometric titer.

The wine-heater is used to preheat the wine; the coil inside raises the temperature to 70 or 85 °C, which condenses the alcoholic vapors issuing from the column. The capacity of the wine-heater varies considerably (5–15 hl).

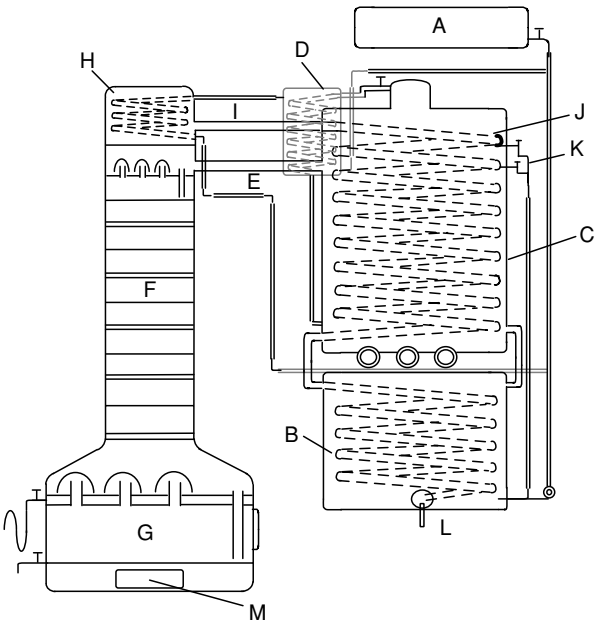

A: Head of wine, B: cooler, C: wine heater, D: head condenser,
E: wine arrival, F: column, G: boilers, H: head of column,
I: swan neck, J: coil, K: drawing and recycling of tailings
L: alcohometer holder, M: furnace.

Figure 6 Armagnac still.

The cooler, which is generally smaller (3–10 hl) than the wine-heater, is placed under the wine-heater. After going through the wine-heater, the coil goes into the cooler to achieve complete condensation and cooling of alcoholic vapors.

Sometimes, a head foreshot condenser system can be fitted to the still, above the wine-heater; more frequently, a tail product condenser is placed level with the alcoholic vapor pipe, between the column and the wine-heater. Tailings can also be collected in the first turns of the coil. The condensed fractions can be returned to the wine and recycled. A coil can be added to the head of the column, to circulate pre-heated wine in the head or tailings condensers; this also helps to condense the least volatile products and to increase the percentage of alcohol.

The wine from the loading vat goes into the still, at the bottom of the cooler, by force of gravity. The flow is regulated with a gate-valve equipped with a flowmeter. The spirits come out of the still through the alcohometer holder where the temperature and alcohometric titer of the distillate are measured. Washy wines are evacuated continuously through a syphon connected to the boiler.

The daily quantity produced cannot exceed one and a half times the capacity of the whole cooling system.

The wine is always heated over an open fire, usually propane gas, although wood is still frequently used to heat small stills.

Because of the way it works, Armagnac distillation is not only far more economical than two-stage distillation, but also three times as fast.

When it has been running for about 2 weeks, the still is turned off and cleaned. Sediments accumulated on the plates and residues prevent the copper from fixing the volatile acids and sulfurous compounds. Inadequate cleaning soon results in the appearance of an unpleasant flavor and a greasy rancid smell, like seconds.

To start the still, one proceeds as follows: first, the boiler and the column are filled with water; once the wine-heater and the cooler are full of wine, the fire is lit; when the water begins to distill, the wine inlet pipe is opened. As soon as the desired alcohometric titer is reached (60% vol, for instance), the spirits are recuperated.

With the Armagnac distillation method, the volatile substances are either entirely distilled (higher alcohols), or are more or less rectified (2-phenyl ethanol, ethyl lactate, 2,3-butanediol), according to their polarity. The fatty acid ethyl esters and fatty acids with a high molar weight are released by heating the yeasts, which means that the quantity of these acids depends on the the wine's yeast content; generally speaking, there are about four times fewer fatty acid ethyl esters with 8, 10, and 12 atoms of carbon in Armagnac than in Cognac. This leads connaisseurs to claim that the quality of Armagnac comes from the specific nature of its soils and the aromas of vines harvested at maturity and not only from their yeast ester content (Table 2).

Table 2 Distillation of volatile substances (average from 50 wines and their corresponding spirits)

	Wine ($mg\,l^{-1}$)	*Spirit* ($mg\,l^{-1}$)	*% Recovery*
Ethanol, (% vol)	11.1	59.7	
Higher alcohols	373	2043	102
Methanol	41.4	194	87
2-Phenyl ethanol	53	32	10
Higher alcohol acetates	2.61	12.7	90
Volatile acid ethyl esters	1.46	14	177
Ethyl acetate	41	207	94
Diacetyl	0.65	2.91	83
Volatile acids (C3 – iC5)	4.12	10.79	38
Volatile acids (C6 – C12)	11.2	34.2	56
Ethyl lactate	340	248	14
Acetic acid	400	118	5.5
Butane-2,3-diol	549	14.9	0.5

To modify the composition of the spirits, the distiller mainly controls two parameters: wine flow and heating.

How the still is adjusted plays an essential role in the composition of the spirits: lowering the heating or increasing the wine flow brings down the temperature at the head of the column and results in a higher alcohometric titer; in this case higher alcohol and ester concentrations correspond exactly to the percentage of alcohol, in other words, the quantity of these substances expressed in g hl^{-1} remains constant (Figure 7).

Conversely, the amount of substances called tailings, of which there is usually a surplus in Armagnac, decreases exponentially when the percentage of alcohol increases. For prolonged aging, a large quantity of tailings is an advantage because of the 'winey' character of their molecules; but if the Armagnac is to be marketed soon, it is preferable to make a high-proof distillate to limit the amount of such substances.

Two-stage pot stills These stills are used in the same way as for Cognac; however, in Armagnac, sediments are never reintroduced in suspension into the wine before distillation. Two-stage distillation has its advantages: the Armagnac ages more quickly and can thus be marketed much earlier, although it loses some of its specificity in the process.

Analysis

Two sorts of analysis are carried out on wine-spirits: (1) traditional analyses; and (2) analyses to improve knowledge of the products and establish correlations between their chemistry and organoleptic observations.

Traditional analyses Their purpose is to determine real and raw volume-percentages of alcohol (titration), dry extract, total acidity, and the ratio of non-alcoholic elements, i.e., volatile acidity, aldehydes, esters, furfural, and higher alcohols; methanol values are determined separately.

Table 1 shows mean values for Armagnac and Cognac, compared to brandy. The sum of volatile substances of brandy is noticeably lower than the others; this is due to the fact that it is elaborated in columns.

Conversely, brandy contains more methanol than Armagnac or Cognac, because it is made essentially with red wines.

Gas chromatography Gas chromatography is used to analyze volatile compounds. As an example, the chromatogram in Figure 8 shows some of the volatile alcohols, esters, acids, and various aromatic components contained in an ether-hexane extract of Very Special Old Pale (VSOP) Armagnac. Other analyses can be carried out by injecting the spirits directly into the chromatograph. The high levels of butanediol found in Armagnac show that it is rich in tail products.

A set of average values can be established from these analyses (Table 1).

Carbonyl compounds are currently being analyzed; long-chain aldehydes seem to be specific to Armagnac; methyl ketones, glyoxal, and methyl glyoxal levels increase with age. These carbonyl compounds contribute to the rancid character of old wine-spirits.

Different high-performance liquid chromatography (HPLC) analyses can also be carried out by absorption spectroscopy to measure the principal phenolic compounds extracted from wood which contribute to the vanilla flavor that develops with age.

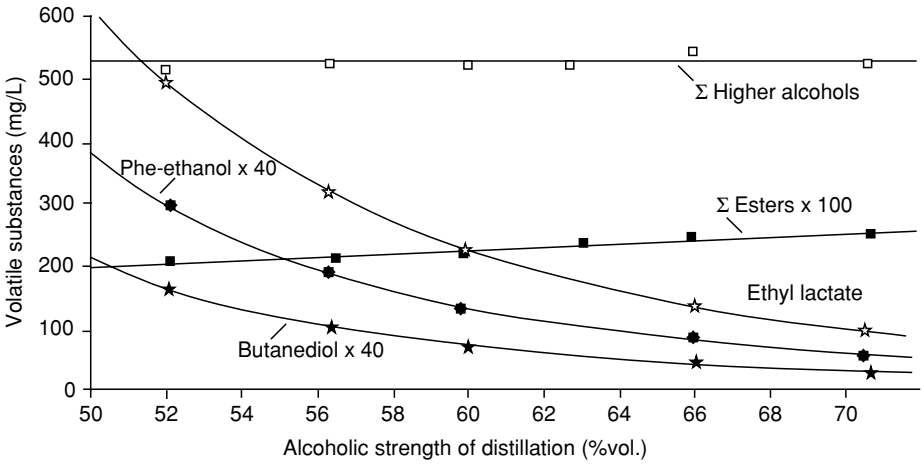

Figure 7 Relationship between alcoholic strength of distillation and content of volatile substances.

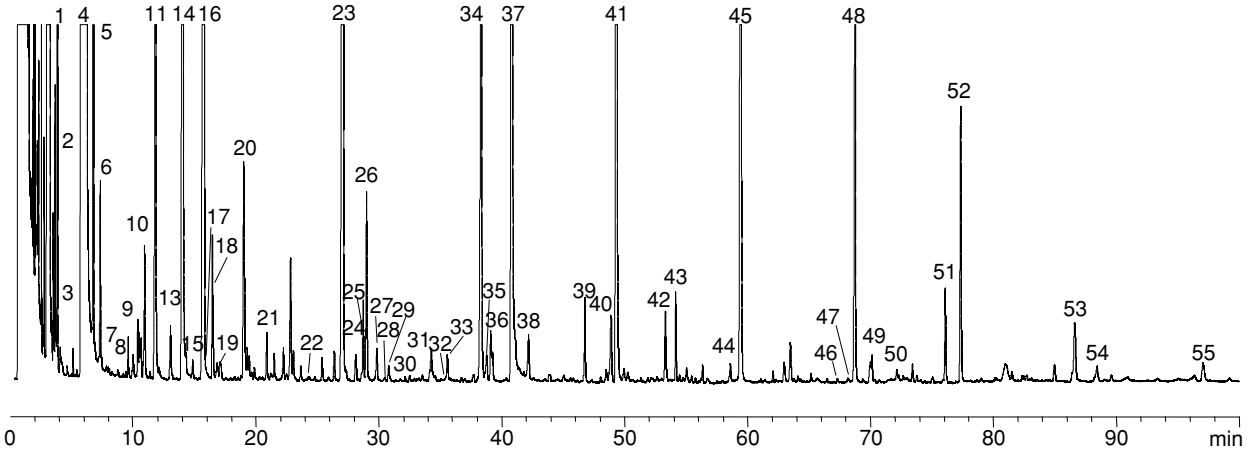

Figure 8 Chromatogram of the extract of an Armagnac by ether hexane (concentration about 3 times); column FFAP 50 m × 0.22 mm; injection according to splitless mode; temperature programmation from 40 to 200°C. Identification of the peaks: 1, Ethylbutyrate; 2, 2-methylpropan-1-ol; 3, isoamyl acetate; 4, isoamyl alcohols; 5, ethyl hexanoate; 6, hexyl acetate; 7, styrene; 8, acetoin; 9, ethyl heptanoate; 10, ethyl lactate; 11, hexan-1-ol; 12, trans-hex-3-en-1-ol; 13, cis-hex-3-en-1-ol; 14, octan-3-ol (internal standard 1); 15, trans-hex-2-en-1-ol; 16, ethyl octanoate; 17, trans-linalol oxide (furane); 18, cis-linalol oxide (furane), 19, acetic acid; 20, benzaldehyde; 21, linalol; 22, 2-methylpropionic acid; 23, ethyl decanoate; 24, butyric acid; 25, 3-methylbutyric acid; 26, diethyl succinate; 27, α-terpineol; 28, 1,1,6-trimethyl-l,2-dihydronaphtalene (TDN); 29, methionol; 30, inconnu; 31, inconnu; 32, damascenone; 33, phenylethyl acetate; 34, ethyl dodecanoate + hexanoic acid; 35, benzyl alcohol; 36, trans-whisky lactone; 37, 2-phenylethanol; 38; cis-whisky lactone; 39, 4-ethylgaiacol; 40, ethyl myristate; 41, octanoic acid; 42, 4-allylgaiacol (eugenol); 43, 4-ethylphenol; 44, ethyl palmitate; 45, decanoic acid; 46, ethyl stearate; 47, ethyl oleate; 48, dodecanoic acid; 49, ethyl linoleate; 50, ethyl linolenate; 51, inconnu; 52, myristic acid; 53, palmitic acid; 54, palmitoleic; 55, linoleic acid.

Sensory analyses Sensory analyses can also be carried out by experienced wine-tasters. Tasters fill in a multiple-choice question tasting-card, and tick off the boxes corresponding to their perceptions; there are also questions related to 'hedonistic' sensations. Since Armagnac comprises several hundreds of substances, it is extremely difficult to describe it accurately in just 20 questions (which is a maximum), but no other solution is forthcoming. Results are often computer-processed using multidimensional analysis methods; multiple correspondence factorial analysis is the best suited.

The main features of old Armagnac are its aroma of prunes, its 'rancio' taste, its complexity; it is vigorous and even rough, with a long-lasting palate.

Aging and Merchandising Preparations

Wine spirits are usually aged in oak casks. Coarse-grained wood is preferred (Gascony or Limousin) to fine-grained wood, as it is slightly more permeable to oxygen and yields more tannin.

At this stage, oxidation is of prime importance, not only for the development of the substances originating from the wood, but also for the distillate itself. Alcohol is oxidized into acetic acid, the quantity of which increases threefold in 20 years; pH, which is 5 in a young wine spirit, drops to 3.5; acids in turn partially become esters; and, in the end, only the higher alcohols remain relatively unchanged in relation to ethanol.

Oak consists of 40–45% cellulose, 20–25% hemicellulose, 25–30% lignin, and 8–15% tannins. The optimal alcohometric titer to extract these components with wine spirits is around 55% vol. Armagnac just out of the still lends itself perfectly to harmonious aging.

The amount of substances extracted from the wood depends on whether the cask is new or old; over a 12-year period, a new cask can produce three times as much as an old one.

Spirits with too much tannin (castalagin and vescalagin) can be harsh and astringent.

With time, lignin is transformed into aromatic aldehydes and phenolic acids.

Armagnac contains vanillin, syringaldehyde, coniferaldehyde, and sinapaldehyde, but only the vanillin is detectable at tasting.

Although there are a variety of aging methods, spirits are usually kept in new casks for 6 months to 1 year before being transferred to old casks.

Prior to being marketed, several wine spirits are blended and the alcohometric titer of the blend (the cut) is reduced to a minimum of 40% vol with distilled water. The naturally golden-yellow color can be enhanced with caramel. Sometimes, infusions or decoctions made from oak shavings are added to

Table 3 Major commercial designations of Appellation d'Origine Contrôlée (AOC) wine spirits

Category	*Minimum age (compte d'† ge)*	*Average age*
Three Stars	2	About 2 years
VO, VSOP	4	5 years
XO, Extra, Napoleon Vieille Réserve, Hors d'Age	5	6 years

VO, Very Old; VSOP, Very Special Old Pale; XO, extra old.

make the Armagnac more astringent, to give it more body; however, these preparations must be at least the same age as the youngest spirit used for the commercial designation of the final product. Sugar solutions are sometimes added to attenuate the 'burn' of the alcohol (about $6\,g\,l^{-1}$).

Finally, before being bottled, the spirits are cold-processed (usually 1 week at $-5\,^{\circ}C$) and passed through a cellulose filter to eliminate any possible cloudiness due to an excess of calcium or fatty acids.

Vintage spirits from a single harvest of a particular year are sometimes sold with no prior reduction of their natural alcohometric titer. Vintage Armagnacs constitute an exception among wine spirits; some of them are extremely valued and fetch high prices.

The various commercial designations are based on the youngest spirit in the blend (cut) the BNIA keeps updated registers listing the age (compte d'âge) and volume of all the different Armagnacs stored in any given storehouse (Table 3).

Cognac

The first distillation of wine for the production of Cognac took place in Charente in 1622. In fact it was in the province of Aunis that this distillation took place, but historically the development of the vine in this area of the south-west of France is related to the economic development of the old maritime province of Saintonge. This was much vaster and extended from the Sèvre Niortaise to Angoulème, the north of the Dordogne, and the Gironde river. But, it only really developed from 276 AD with the edict of the Roman emperor Probus who granted the privilege of cultivating the vine which had hitherto been granted only to Romans in 92 AD.

During the thirteenth century trade flourished, in particular with the shippers of the Nordic countries who came to buy not only the salt of the Saintonge salt-pan created by the Romans but also the wine of Ritsel (La Rochelle) and Saint Jean (d'Angély). Since demand perpetually increased, the wine of the 'haut pays' began to be traded as the small coastal rivers were navigable (Boutonne river which goes to Saint Jean d'Angély and especially the Charente river, navigable up to Saintes, Cognac, Jarnac, and Angouleme). Dutch sales representatives settled in Tonnay-Charente, then in Cognac, which was a 'tax stage' city between the provinces of Saintonge and Angoumois and which in the fifteenth century had an administrative infrastructure appreciated by the traders.

Another important event was the marriage of Eleanor of Aquitaine with Henri Plantagenet in 1152. This opened up an era of great prosperity for this province which could practise the trade of wine since it was now attached to England. The province then began to plant vines massively. At the end of the sixteenth century, when the Dutch became the masters of maritime trade, brandy gradually replaced wine and it was also used as a currency of exchange for spices, metals, and slaves. The Dutch stimulated the production of aquavits and brandevins which insured a perfect conservation of the distilled product and made it possible to reduce the volume to be transported. However, because of the proximity of the sea, the Dutch could distil the wine on their premises because the trip was not too long. This explains the relatively late appearance of distillation in Saintonge (two centuries later than in Armagnac).

At the beginning the stills were imported from Holland. They were manufactured with Swedish copper but the production of stills in Bordeaux and then in the Cognac area began about 1724. At that time there occurred a kind of wine revolution which led certain areas to increase their production, in order to have the greatest volume for the Dutch merchants. On the other hand, other areas preferred the production of high quality, thereby implying limited yields. Such was the case in particular for the wines of Bordeaux, Burgundy, Champagne, for the liqueur wines of Sauternes, Bergerac, and Borderies (area to the north-west of Cognac), and finally for Cognac brandies.

In fact, it seems that tax reasons encouraged the middle-class rich of the city of London to search for top-of-the-range products. Since taxation of wines and alcohols was very high from 1628, it was necessary to pay high prices, so the product had to be of excellent quality. By storing brandy in wood barrels, the English to some extent invented the aging that is now known to transform brandies into delicious, much more agreeable products for consumption. Moreover, the term 'Old Cognac Brandies' appeared in London in 1700. With the Augier family in 1943, the British, for example, Jean Martell (1715) and Richard Hennessy (1765), were partly responsible for setting up the true commercial structures which still remain today.

At that time, the trade in wines and brandies was of equal importance but thereafter, brandy production really came to the fore.

The production of Cognac began on the chalky slopes located to the south-east of Cognac, an area called 'Champagne' because of the nature of the grounds. The vineyards extended gradually to the surrounding wooded regions. Starting from 15 000 ha in 1600, the vineyard reached 100 000 ha in 1800 and 255 000 ha in 1878, just before the phylloxera epidemic which destroyed the Charentais vineyard.

Nevertheless, the multiple obstacles to trade of all types, in particular the wars, resulted in huge volumes of spirit being stored. In 1873 this was evaluated at 7 277 000 hl of Cognac. This made it possible to overcome the period of reorganization of the vineyard without losing important customers, in particular the English.

The Viala mission (1887) found an American vine in Texas, *Vitis berlandieri*, which was resistant to phylloxera and adaptable to the charentais soils which are very rich in limestone. This made it possible to create rootstocks which were not susceptible to chlorosis (when the leaves cannot synthesize chlorophyl owing to iron deficiency). Under the control of Alexis Millardet, the father of the bouillie bordelaise (copper sulfate + lime) used for treatment against mildew, research led to the 41 B rootstock, a hybrid of chasselas and *V. berlandieri*. This rootstock was used to reconstitute the vineyard, and gradually the old traditional type of vines like folle blanche or Colombard were replaced by the better-adapted white Ugni blanc cultivar. The average quality of brandies from Ugni blanc is higher than that of folle blanche brandies from the last century because the latter was subject to gray rot.

The Vineyard

Double Guyot pruning of the vine is practiced. Nowadays, the plants are 1 m 30 on the rows and 3 m between the rows. Pruning is done with two long branches arched downwards. The number of buds is 50 000–60 000 per hectare, i.e., 20–22 buds per foot. Normally the vine is trellised with five or six wires.

Blossoming (débourrement) of the buds take place between the end of March and beginning of April, flowering is about June 20–25, ripening (veraison) between August 25 and September 10, and harvesting generally begins about October 10.

The Climate

The climate is moderate and wet, rather similar to that of Bordeaux, with average temperatures + 6.5°C in winter and 21.7 °C in summer. Rainfall (800 mm of rain per year) is distributed over 130–150 days.

The Appellation Areas

There are six appellation areas (crus) (**Figure 9**) corresponding to particular soils, which may also be influenced by climate. For a long time tasters classified brandies according to the sectors. The decree fixing these areas was promulgated in 1909. The limits were modified in 1938.

1. The Grande Champagne comprises 27 communes of the Cognac district. The soils are clay-calcareous surmounting a friable calcareous layer of cretaceous soil which absorbs excess water and which, in dry periods, constitutes a reservoir locally, called the volant of quality. From a climatic point of view, this is both the hottest and driest area, with the nearby sea having relatively little influence.
2. The Petite Champagne is concentric in relation to Grande Champagne. Its characteristics are similar but slightly different brandies have produced. A mixture of grande and petite champagne is harmonious, and is called fine champagne.
3. Borderies, on the right of the Charente river, to the north-west of Cognac, has soil composed of sand and clay (instead of calcareous and clay in the other areas) on a calcareous base, and gives particular brandies which age with a typical bouquet.
4. The Fins Bois were formerly wooded areas (before the nineteenth century). The soils composed of hard stones are from Jurassic, and with a cretaceous calcareous base. They give brandies which are not as fine as the preceding ones but which age more quickly.
5. The bons bois soils contain less limestone, so may also be more influenced by the climate.
6. The bois ordinaires are primarily the north of maritime Charente and the islands. Here, the sea influence is more marked, and the wines and spirits may sometimes reflect this in taste.

Production

Between 1956 and 1973, the possible yield per hectare rose from 3 hl of pure alcohol (PA) to 11.5 hl PA. During the years 1967–1973 the average yield was approximately 7 hl. AP ha^{-1}. Nowadays it is about 9 hl ha^{-1}. For the 1999 vintage approximately 3 850 000 hl of white wine was distilled, giving 325 000 hl PA of Cognac (approximately the equivalent of 100 000 000 bottles of Cognac). The same quantity has been sold in 2000 with more than 90% for the export market.

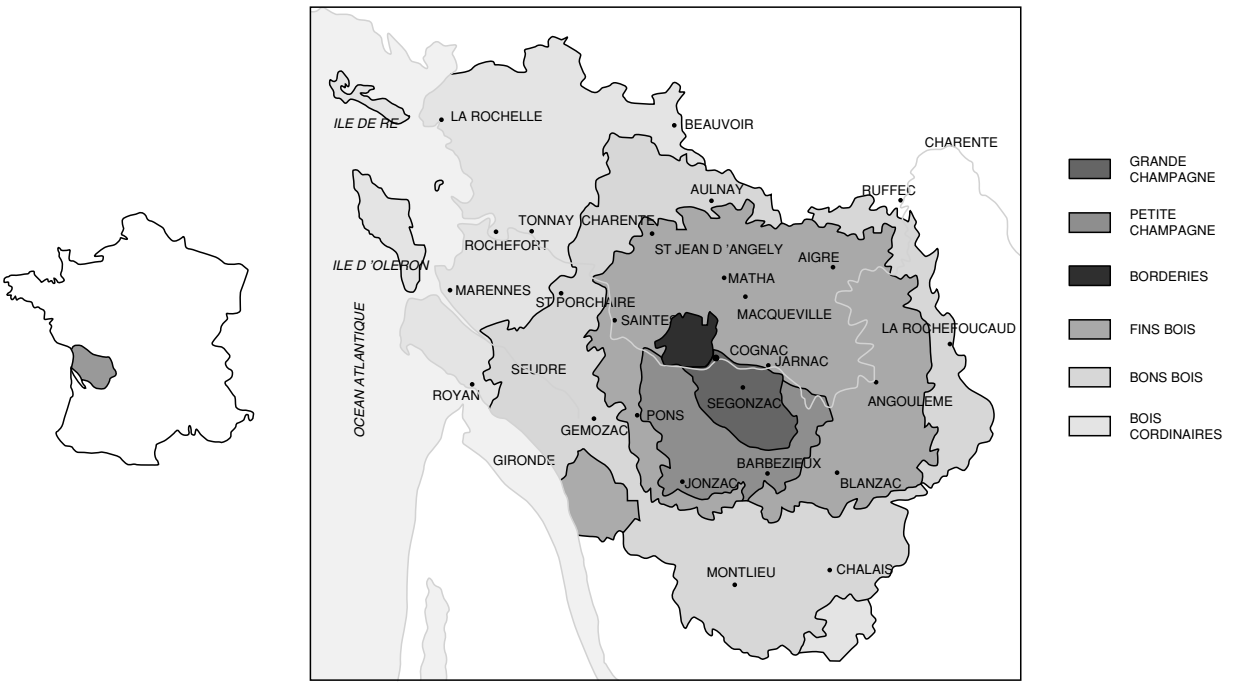

Figure 9 (see color plate 11) Cognac in France and the different appellation areas. See text for further details.

Wine Making

This is very rudimentary. The grape harvest takes place in the second fortnight of October. About 90% of the production is mechanically harvested. Then the grapes are pressed in horizontal presses. The continuous presses (called the Archimedes screw) are prohibited because they may induce herbaceous tastes. Sulfur dioxide is not used because it binds ethanal which is then released during distillation and tastes pungent. Nor is there any racking of the lees before fermentation (débourbage), and addition of sugar to the must to increase the potential alcoholic strength (chaptalization) is strictly prohibited.

Sometimes, alcoholic fermentation starts badly in cold and rainy years, but generally it is very fast (8 days for the first vat), then 3–5 days, when the wine material has been used and allows seeding of musts.

The wines are kept on the lees, and malolactic fermentation is conducted in the month following alcoholic fermentation, even if the pH is low, because absolutely no sulfur dioxide is used.

Distillation can immediately take place after the end of the fermentations, and the best brandies are those which are distilled first. Ethyl butyrate levels increase with time of conservation and especially when the temperature increases. This is explained by the total absence of disinfectants to protect the wine against oxidation and bacterial accidents. As long as the wine ferments, the medium is reduced (low oxidoreduction potential) and there is no risk. However, at the end of fermentation, the wine must be stored in full, hermetically closed vats. The most frequent accident is the browning of wine (casse brune) caused by an enzymatic oxidation of the phenols in quinones in years of rot. It seems that this has no effect on the quality of the spirits since the polyphenol does not distil.

The bacterial risks are the bacterial degradation of tartaric acid call tourne (which is far from frequent), and the problem of bitterness due to acrolein which arises from the decomposition of glycerol in propenal. This is not very frequent but is very harmful for the brandies, which then become undrinkable. Several cases occurred in the past when the tanks were situated under the ground; the wine cooled too slowly and lactic bacteria degraded the glycerol and butanediol. Since the use of disinfectants is prohibited, it is obviously necessary to work with perfectly clean wine-making equipment because of possible contamination. The screw of the press must be lubricated with paraffin oil because mineral oil could give a tainted taste to the Cognac.

Wines

The alcoholic strength is generally relatively low (8–10% vol). The wines are too acid for direct consumption at pH 3 or even less, and total acidity ranges

from 7 to 10 g l^{-1} (expressed as tartaric acid), even if malolactic fermentation has taken place. The acidity of the must reaches 15–20 g l^{-1}. This acidity makes it possible to some extent to compensate for the absence of sulfur dioxide. However, according to regulations, the wines must be distilled before the end of March because after that time heat would lead to inevitable deterioration caused by the growth of lactic bacteria. These wines contain only very little ethanal since sulfur dioxide is not used.

The flavor must be very neutral. When the specific fruitiness of the type of grape is too marked, it is not possible to obtain fine brandies. For example, muscadelle or sauvignon cultivars would not be suitable to obtain good Cognac. In contrast, with the very neutral Ugni blanc, the flavor comes from fermentation byproducts. This flavor is probably mainly due to higher alcohols which are present in high quantities, since fermentation occurs in the presence of the less of the must (see Table 1). The esters of fatty acids, ethyl caproate, caprylate, caprate, and laurate are also important because of a relatively low fermentation temperature (late grape harvest, low sugar content).

The herbaceous tastes in Cognac are due to carbonyl compounds with six carbons. They may occur when the harvested grapes contain too many leaves or when they are crushed too hard by the machines. Oily tastes are due to 1,1,6-trimethyl-1,2-dihydronaphtalene (TDN) for the same reasons. This chemical substance seems to be more or less specific to Ugni blanc, especially when it is not ripe enough.

Volatile acidity is generally very low (0.2–0.4 g l^{-1} expressed in acetic acid). Ethyl acetate levels are also generally low (30–60 mg l^{-1}). The dry extract varies between 15 and 24 g l^{-1}. The quantity of unfermented sugar is very low, and often nil (less than 50 mg l^{-1}).

Distillation

The charentais still (Figure 10) The charentais still is made of copper of electrolytic quality (no hydrogen bubbles in metal). Copper is essential since it is a good conductor, is not attacked by wine acids, and fixes the traces of H_2S arising from the late treatment of the vine with sulfur to combat oidium infection. Copper combines a part of the caprylic, caproic, and lauric fatty acids, whose odor resembles that of cheese, and long-chain fatty acids which could give insoluble soaps (eliminated by filtration). Distillation in a stainless-steel boiler gives poor-quality brandies. The process can be improved by adding turnings of copper or copper sulfate. The copper is rolled and hammered to increase its hardness and to make its surface smoother.

There are two shapes of stills according to the shape of the boiler, which is either straight or onion-shaped. The volume of the boiler is 30 hl maximum for the second distillation called the bonne chauffe; exceptionally, the volume of the boiler may be 150 hl but only for the distillation of the wine (first distillation).

The upper part is called the chapiteau (olive or onion form). Its volume represents 10% of the volume of the boiler and it is topped by the swan neck. A wine heater generally makes it possible to recover the calories provided by the alcoholic vapors. The temperature is about 50 °C before the boiler is filled. The serpentine, which is also made of copper, is about 30 m long. It is placed in the cooler, called

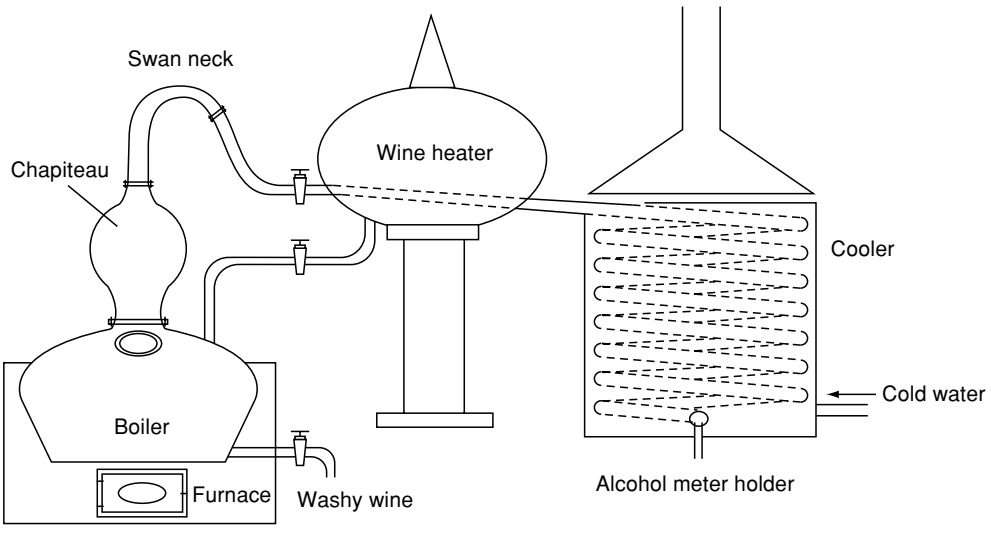

Figure 10 Charente still for the production of Cognac spirit.

la pipe, which is filled with water; its capacity is approximately twice that of the boiler.

The boiler is assembled on a solid mass of masonry. The boiler is heated by direct flame (no steam or electric heating). At the lower part is the furnace. On the sides of the boiler are what is known as the 'turns with fire' in which the gases from the hearth are burned before being evacuated towards the chimney. From 600 °C in the hearth, the evacuated gases drop to no more than 250 °C.

The liquid in the boiler is at boiling point (between 90 and 102 °C) and the temperature in the chapiteau is lower than 5–8 °C. Part of the vapor (3–5%) flows back, which makes it possible to obtain a slight rectification of the distillate. At 1 mm above the bottom of the boiler, there occurs an overheating of about 5 °C before boiling begins. Excessive overheating induces a certain bitterness of the spirit.

Formerly wood was used as fuel, then coal, and nowadays butane and generally propane. It is obviously necessary to use burners which work efficiently without overheating. The adjustments are similar to those of a central heating system.

Distillation method Traditionally the wines are distilled with their lees. However, in recent years only the finest lees (yeast) have been used, so it is possible to speak of wines with fine lees or even wines free of lees. This has repercussions on the ester content (Table 4).

Table 4 Effect of the lees of the wine on the composition of the distillate

	Few lees in the wine	*With lees in the wine*
Ethyl caproate	6.76	8.3
Ethyl caprylate	8.95	23.6
Ethyl caprate	13.8	63
Ethyl laurate	12.45	36.2
Ethyl myristate	5.4	9.8
Ethyl palmitate	9.77	13.2
Ethyl palmitoleate	1.44	1.8
Ethyl stearate	0.59	0.61
Ethyl oleate	1.19	1.22
Ethyl linoleate	7.69	9.2
Ethyl linolenate	1.86	2.58
Isoamyl caprylate	0.42	2.48
Isoamyl caprate	1.67	5.76
Isoamyl laurate	0.78	1.83
2-phenylethyl caprylate	Traces	1.2
2-phenylethyl caprate	0.25	1.55
Summ of aromatic esters	73.02	186.65 (+ 150%)

Results as $mg\,l^{-1}$ of the distillate at 70% vol.
Reproduced from Cantagrel R, Lurton L, Vidal JP and Galy B (1992) La distillation caharentaise pour L'obtention des eaux-de-vie de Cognac. In: Bertrand A (ed.) *Les Eaux-de-vie Traditionnelles d'Origine Viticole*, pp. 60–69. Paris: TEC & DOC, with permission.

First distillation or distillation of the wine, known as chauffe du brouillis The boiler is filled to 95% of its volume. The wine boils after 1.5 h of heating, and the distillate which runs out first contains 55–60% vol of alcohol. Distillation takes place until all alcohol is distilled and is stopped when the distillate contains 2% vol. The liquid thus collected is called the brouillis, its alcoholic strength ranges from 26 to 31% vol. It is obtained in approximately 12 h. The temperature of flowing distillate must be about 15 °C.

Second distillation, called the bonne chauffe This is carried out under the same conditions. After 1.5 h of heating, the distillate starts to run out. It contains 75–80% vol of alcohol: 1% of total volume collected constitutes the heads which are separated and mixed with wine or brouillis. The distillate is collected up to 60% vol of alcohol; this is the heart, whose average alcoholic strength is nearly 70% vol. The distillate which then runs out is called the seconds (which taste of fat); these seconds are mixed with wine or brouillis to be redistilled.

The temperature of the distillate which runs out of the cooler must be about 18 °C to obtain a good brandy. The rate of flow is $1\,l\,min^{-1}$.

The seconds can either be mixed with the wine and in this case the alcoholic strength is not of major importance (first principle, Figure 11), or they may be added to the brouillis, in which case the strength is 28% vol. Therefore, it may be necessary to reduce the distillation of the brouillis according to the alcoholic strength of the initial wine (second principle, Figure 12). In this case the seconds are redistilled only once, which increases the strength of the brouillis.

The second distillation is performed at a lower temperature, thus obtaining better rectification of the spirit. Here, the tails of the brouillis need to be cut.

There are also variants of these two methods. Varying the intensity of heating is important according to the strength of brouillis required. Slow distillation gives good rectification. An odorous fine brandy is obtained but with dryness that may be detected on tasting due to the lack of certain products of tail distillation (e.g., ethyl lactate, diethyl succinate). In contrast, fast heating involves the formation of a marrowy brandy with little bouquet. Excessive heating results in a heavy taste.

Aging

The cognac ages in barrels of between 200 and 600 l. The barrels must be in oak of special quality. Traditionally these oaks came from the forests of Tronçais, Allier, Limousin, and the Vosges in France. The stave woods, i.e., pieces of wood used to make the barrels,

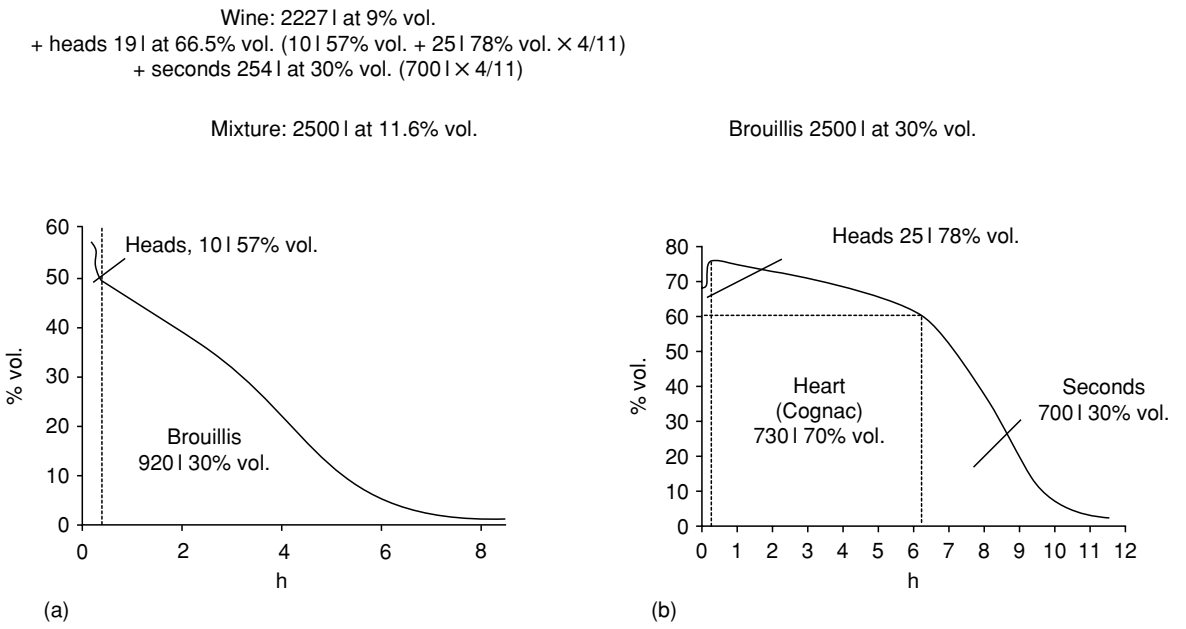

Figure 11 Distillation of the cognac, first principle. (a) First distillation of the wine, to obtain the broiullis; (b) second distillation, bonne chauffe. Eleven brouillis must be distilled to give four bonnes chauffes.

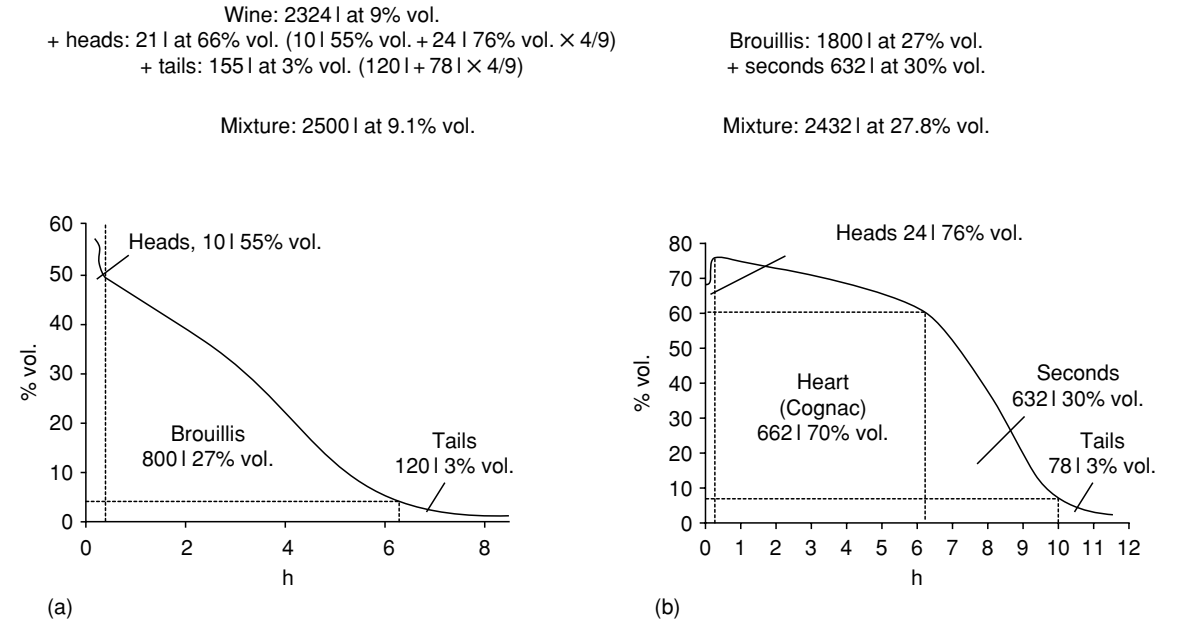

Figure 12 Distillation of the cognac, second principle. (a) First distillation of the wine to obtain the brouillis; (b) second distillation, bonne chauffe. Nine brouillis must be distilled to give four bonnes chauffes, and all the tails are mixed with the wine.

must be stored for 3 years outside so that the bitter substances that they contain are transformed by enzymatic reaction. The meta-digallic acid is hydrolyzed into the less aggressive gallic acid. The same occurs with aesculine and scopoline, glucosides that are transformed into aesculetine and scopoletine, which are less bitter. The transformations can take place only with time; fast drying obtained by stoving (heating) is not sufficient and the brandies placed in such barrels are bitter.

During aging the brandy is oxidized slowly. Its acidity grows by oxidation of alcohol into volatile

acids and by dissolution of the acid substances in the wood. Moreover, acetals are formed, and their odors are softer than those of aldehydes. The odor characteristic of young brandy becomes blurred, and eventually disappears, to be replaced by a vanilla odor induced by vanillic aldehyde, together with other phenolic aldehydes and acids arising from the alcoholysis of lignin in oak wood. The color becomes brown by dissolution of tannin. Moreover, the taste softens with the appearance of sugars arising from the hydrolysis of wood hemicelluloses.

There is also the appearance of a rancid taste by oxidation of fatty acids. Several stages in aging may be distinguished:

- From 1.5 to 5 years the main process is dissolution of substances in the wood.
- By 5–10 years astringency decreases and the brandy becomes rounder.
- From 10 to 35 years a rancio taste appears.
- After 40 years one should no longer keep brandies out of the barrel.

A barrel can yield substances to the cognac for about 40 years.

During aging, there is a loss of volume known as 'the angels' share.' This represents 3% on average per year, with a 1% vol reduction in the alcoholic strength.

During conservation there is little evolution of the volatile substances. The alcohols concentrate, the esters are slightly hydrolyzed, and the unsaturated fatty acids oxidize, which gives the rancio taste. The Cognac ages by slow oxidation in barrels. In bottles there is no further evolution.

Analysis

Traditional analyses Their purpose is to determine real and raw volume percentages of alcohol (titration), dry extract, total acidity, and the ratio of non-alcoholic elements, i.e., volatile acidity, aldehydes, esters, furfural, and higher alcohols; methanol values are determined separately.

Table 1 shows values for Cognac compared to those of brandy and Armagnac. The sum of volatile substances of Cognac is very high because the distillation process does not eliminate higher alcohols; however some tail products such as volatile acidity are noticeably lower than in Armagnac.

Gas chromatography Gas chromatography is used to analyze volatile compounds. As an example, the chromatogram in Figure 13 shows some of the volatile alcohols, esters, acids, and various aromatic components contained in an ether-hexane extract of VSOP cognac. Other analyses can be carried out by injecting the spirits directly into the chromatograph.

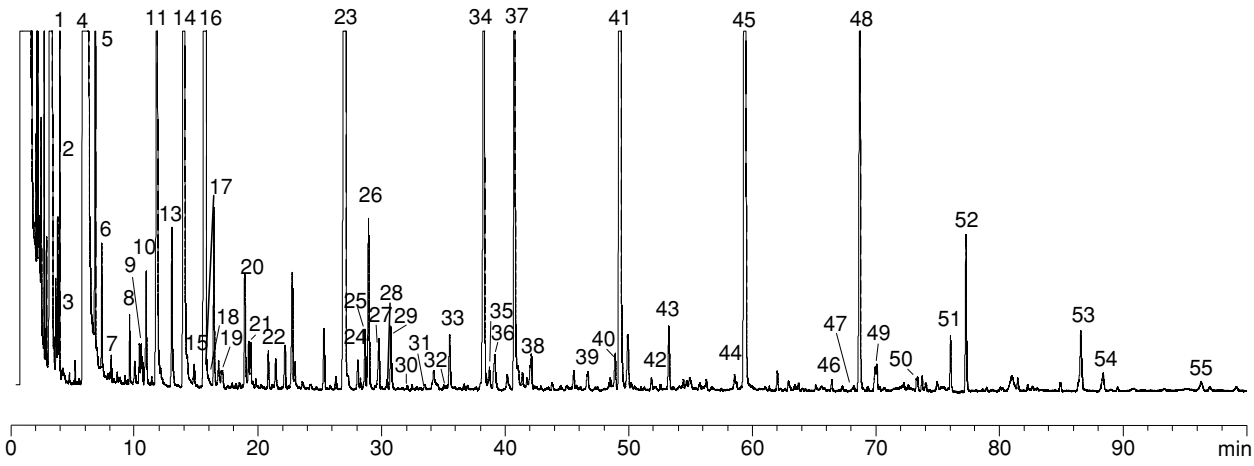

Figure 13 Chromatogram of the extract of a Cognac by ether hexane (concentration about three times); column FFAP 50 m × 0.22 mm; injection according to splitless mode; temperature programming from 40 to 200 °C. Identification of the peaks: 1, Ethyl butyrate; 2, 2-methylpropan-1-ol; 3, isoamyl acetate; 4, isoamyl alcohols; 5, ethyl hexanoate; 6, hexyl acetate; 7, styrene; 8, acetoin; 9, ethyl heptanoate; 10, ethyl lactate; 11, hexan-1-ol; 12, *trans*-hex-3-en-1-ol; 13, *cis*-hex-3-en-1-ol; 14, octan-3-ol (internal standard 1); 15, *trans*-hex-2-en-1-ol; 16, ethyl octanoate; 17, *trans*-linalol oxide (furane); 18, furfural, 19, acetic acid; 20, benzaldehyde; 21, vitispirane; 22, linalol; 23, ethyl decanoate; 24, butyric acid; 25, 3-methylbutyric acid; 26, diethyl succinate; 27, α-terpineol; 28, 1,1,6-trimethyl-1,2-dihydronaphthalene (TDN); 29, methionol; 30, unknown; 31, unknown; 32, damascenone; 33, phenylethyl acetate; 34, ethyl dodecanoate + hexanoic acid; 35, benzyl alcohol 36, *trans*-whisky lactone; 37, 2-phenylethanol; 38; *cis*-whisky lactone; 39, 4-ethylgaiacol; 40, nerolidol + ethyl myristate; 41, octanoic acid; 42, 4-allylgaiacol (eugenol); 43, 4-ethylphenol; 44, ethyl palmitate; 45, decanoic acid; 46, ethyl stearate; 47, ethyl oleate; 48, dodecanoic acid; 49, ethyl linoleate; 50, ethyl linolenate; 51, unknown; 52, myristic acid; 53, palmitic acid; 54, palmitoleic; 55, linoleic acid.

The high levels of esters and fatty acids indicate that wine is distilled with lees (yeast).

Disturbances and Deposits

These can be:

- of organic nature: insolubilization of certain substances when the temperature decreases; to avoid this accident the brandies are filtered cold after cooling at $-5\,^{\circ}C$.
- of mineral origin due to calcium arising from the barrels (1–1.5 mg in 15 years' storage) or from the filter plates, thus requiring the use of plates without $CaCO_3$. Beyond $3\,mg\,l^{-1}$ the brandy deposits at cold temperature. Certain treatments by ion exchange resins could be used to eliminate the excess of calcium and copper.

Preparation for Commercialization

During aging, once a year, the cognac is racked and all the barrels of the same production are mixed together. The alcoholic strength is gradually diminished by adding demineralized water to finally obtain an alcoholic strength of 40.0% vol in commercialized bottles (never less). During aging it is common to mix spirits of different origins, quality and age in order always to obtain the same odor and taste for a given brand name. For instance, some Cognac firms use up to 60 different brandies to prepare their blend but the brand name (i.e., ***, VSOP) must consider only the youngest spirit of the blend.

The color can be adjusted with the addition of caramel. The taste can be adjusted with the addition of woody water extract obtained from small pieces of oak wood to give more body to the spirit, more astringency, and a little bitterness. On the other hand, excessive hardness can be diminished by the addition of sucrose syrup, generally less than $8\,g\,l^{-1}$. Generally, all these procedures should be carried out at least 2 or 3 months before bottling.

Tasting

Just as for wines, the length of a brandy can be measured by taking it into the mouth, spitting it out, and counting the number of seconds during which an intense aromatic persistence may be perceived, just before the appearance of a certain dryness. The word 'caudalies' is used to indicate the number of seconds elapsed. For a Grande Champagne Cognac the sensation lasts 7–12 s; for Bois ordinaire Cognac it is only 2 s.

The Grande Champagne In addition to the length of aromatic persistence, this brandy is characterized by a bouquet of great smoothness and distinction, odors of vine flower, dried lime, flowers and dry vine shoots.

The Petite Champagne The same characters are found but are less accentuated.

Borderies The persistence is the same as in Petite Champagne. These brandies have a developed odor which is more suave than that of the champagnes. At the time of the folle blanche before 1880, brandies had an odor of violet. Unlike in Armagnac, this is not considered a quality in Cognac spirits. They are less light than the Champagnes.

Fins bois The fins Bois brandies are rounder than the Champagnes. They are more marrowy but have a heavier odor more reminiscent of the fruit than the flower. From the point of view of character, the length decreases. They age faster, so together with the bons bois producting *** and VSOP Cognac, they constitute the most readily available of the Cognacs.

Acknowledgments

The author is grateful to the Bureau Interprofessionnel de l'Armagnac for its collaboration and financial support in the form of research contracts, and to his students at the Institute of Oenology: P. Wildbolz, Ph. Jadeau, Miss M.C. Segur, Miss S. Biau, M.K. Bertsch, Miss R. Vanderlinde, and E. Herve.

See also: **Alcohol**: Properties and Determination; Metabolism, Beneficial Effects, and Toxicology; Alcohol Consumption; **Brandy and Cognac**: Chemical Composition and Analysis of Cognac; **Grapes**; **Sensory Evaluation**: Aroma; Taste

Further Reading

Bertrand A (1988) Role of the continuous distillation process on the quality of Armagnac. In: Piggott JR and Paterson A (eds) *Distilled Beverage Flavour. Recent Developments*, pp. 97–115. Edinburgh: Ellis Horwood.

Bertrand A (1989) L'analyse chimique et sensorielle d'eaux-de-vie d'Armagnac. Conséquence de la technologie sur la qualité. In: *La Règlementation Communautaire des Eaux-de-vie et des Liqueurs*, vol. 9, pp. 101–111. Verona: Bollettino del Cideao OIV (ed.).

Bertrand A and Segur M-C (1991) L'alambic armagnacais. In: Bertrand A (ed.) *Les Eaux-de-vie Traditionnelles d'Origine Viticole*, pp. 70–74. Paris: Lavoisier-TEC & DOC.

Bertsch K (1992) *Les Eaux-de-vie d'Armagnac, Critères Analytiques de Qualité, étude sur le Carbamate d'Èthyle*. Thesis: l' Université de Bordeaux II.

Cantagrel R, Lurton L, Vidal JP and Galy B (1992) La distillation caharentaise pour l'obtention des eaux-de-vie de Cognac. In: Bertrand A (ed.) *Les Eaux-de-vie Traditionnelles d'Origine Viticole*, pp. 60–69. Paris: TEC & DOC.

Casamayor P and Cousteaux F (1985) *Le Guide de l'Amateur d'Armagnac*. Toulouse: Laffont/Briand.

Caumeil M (1983) Le Cognac. *Pour la Science* December: 48–57.

Dufor H (1982) *Armagnac – eaux-de-vie et terroir*. Toulouse: Privat.

Guigon D and Cogat P (1991) Elimination des défauts dans les eaux-de-vie de vin traditionnelles. In: Bertrand A (ed.) *Les Eaux-de-vie Traditionnelles d'Origine Viticole* pp. 110–113. Paris: Lavoisier-TEC & DOC.

Jadeau P (1987) *Incidence du Débourbage des Moûts et de la Fermentation Malolactique des Vins sur la Composition des Eaux-de-vie d'Armagnac*. Thesis: Université de Bordeaux II.

Lafon J, Couillad P and Gay-Bellyle F (1973) *Le Cognac, sa Distillation*. Paris: JB Bailliere.

Marche M and Joseph E (1975) Etude théorique sur le cognac. *Revue Francaise d'Oenologie* 57: 1–22.

Mariller C (1925) *Distillation et Rectification des Liquides Industriels*. Paris: Dunod.

Puech J-L (1986) Le vieillissement de eaux-de-vie. In: *Les Arômes des Vins*. Montpellier: Journée de Rencontres Œnologiques.

Roget J and Garreau C (1991) Recherches sur l'origine de la distillation et la préparation de l'alcool. In: Bertrand A (ed.) *Les Eaux-de-vie Traditionnelles d'Origine Viticole*, pp. 54–59. Paris: Lavoisier-TEC & DOC.

Segur M-C (1988) *Approche de la Dégustation Analytique des Eaux-de-vie d'Armagnac. Interprétation Statistique*. Thesis: Université de Bordeaux II.

Sempe A (1988) *La Grande Messe de l'Armagnac*. Paris: Laffont.

Chemical Composition and Analysis of Cognac

R Cantagrel, Bureau National Interprofessionnel du Cognac (BNIC), Cognac, France

Introduction

Cognac is one of the most prestigious products in the world. It has received an 'Appellation d'Origine Contrôlée' or registered designation of origin, and the term 'Cognac' must only be used for spirits from Cognac or the Charente region. It is subject to strict production regulations to guarantee origin and quality.

Cognac is characterized by a rich aroma and flavor, which distinguishes it from other spirits distilled from wine, the production of which is not governed by such strict regulations.

Chemical Composition and Sources of Variations

The aroma of Cognac is the result of the characteristic aromas of many of its constituents, enhanced or modified by synergistic or masking effects. More than 500 substances have been detected in spirits; they belong to a large number of chemical classes: alcohols (ethanol, methanol, higher alcohols, etc.), aldehydes, esters, volatile acids (acetic acid, fatty acids, etc.), ketones, acetals, nitrogen-, oxygen- and sulfur-containing heterocyclic compounds, phenolic acids, and aldehydes. These different aromatic constituents are contributed by:

- the grape (primary aromas);
- wine-making (fermentation aromas);
- distillation (specific aromas produced by the heating process);
- aging (aromas imparted by the oak wood).

The main aim in the production of a spirit is to form pleasant, distinctive flavors and fragrances to be conveyed by the alcohol. (*See* **Wines**: Production of Table Wines.)

The quality of a given Cognac depends on a number of factors:

1. In the production process, variations can be due to:
   - the weather conditions, which may vary from year to year, causing differences in the chemical composition of the resulting wines and spirits;
   - the diversity of grape producers;
   - the different 'crus' or growth regions within the area of designated origin;
   - the several hundered distillers;
   - the several types of stills used with different distillation methods.
2. In the aging process, variations can be due to:
   - the origin of the oak wood used for the barrels or casks;
   - cask-making techniques;
   - the layout and geographical location of the cellar (chai) where the spirits are aged;
   - the length of the aging period.

Variability of the Raw Material

The wines used in producing Cognac spirits are made essentially from the Ugni-blanc grape variety, which

produces wine with a relatively low alcohol content (generally 8–10% v/v) and a fairly high acidity (6–12 g of sulfuric acid per liter). This high degree of acidity ensures good conservation before distillation. Most of the constituents that make up a Cognac are present in the wine and lees used in the distillation.

Weather conditions are never the same from one year to the next. They determine the size of the grape harvest and the different qualities of wine produced by the vineyard. For example, the average alcoholic strengths (v/v) of wines over the last 12 years in Cognac were:

- 1987, 7.5%;
- 1988, 9.6%;
- 1989, 10.9%;
- 1990, 9.8%;
- 1991, 9.6%;
- 1992, 8.0%;
- 1993, 7.8%;
- 1994, 8.6%;
- 1995, 9.7%;
- 1996, 9.8%;
- 1997, 10.3%;
- 1998, 9.2%.

As a result, the chemical composition of the resulting spirits varies from year to year.

The figures below illustrate the diversity of the raw material obtained:

- the total area of the registered designation region is 80 000 ha, comprising six growth zones ('crus') within the designation region;
- the number of grape growers producing the white wine sold for Cognac is 8000;
- the number of producers with vineyards larger than 3 ha is 5000;
- the number of industrial-scale distillers is 150;
- the number of growers distilling on their own premises is 2500.

Cognac is an internationally recognized product, highly specific, and yet obtained from a number of different sources.

During the distillation period (about 5 months), the composition of the wines varies (see Table 1). The higher alcohol and polyol content does not change, but there is a significant decrease in the content of esters with aromatic properties of interest (isoamyl, hexyl and phenylethyl acetates, ethyl caproate, caprylate, caprate and laurate) and an increase in ethyl acetate, ethyl lactate, diethyl succinate, acetaldehyde (ethanal) and acetic acid. For this reason, the distillation is not performed in the same way in November, when the fermentation is completed, as in March.

Table 1 Variations in the concentration of wine constituents during storage for 5 months

Constituents		*Variation (%)*
Ethyl acetate		+24.1
Isoamyl acetate	1[a]	−55.7
Hexyl acetate	2[a]	−60.7
Phenylethyl acetate	3[a]	−62.1
Ethyl caproate	4[a]	−22.0
Ethyl caprylate	5[a]	−14.2
Ethyl caprate	6[a]	−18.9
Ethyl laurate	7[a]	−52.0
Total esters	1–7[a]	−25.8
Ethyl lactate		+49.1
Diethyl succinate		+226
Acetaldehyde (ethanal)		+117
Acetic acid		+18.8

[a]Organoleptically aromatic esters present in wine in low quantities. Ethyl acetate and ethyl lactate concentrations are much higher.

Charente Distillation Method

The distillation period in Charente begins as soon as the alcohol fermentation ends (generally the beginning of November) and continues until March 31.

The Charente method of distillation is performed in two stages. From the first distillation stage, the brouillis (27–30% (v/v) alcohol) is obtained; the second distillation stage yields the bonne chauffe (70% (v/v) alcohol). The purpose of this is threefold:

1. To extract the volatile compounds contained in the wines used. Besides water and alcohol (ethanol), other compounds contribute more through their intense characteristics rather than their quantities; they represent 0.3–1% (v/v) of the alcohol.
2. To select (rectification) from among the volatile substances present with the alcohol. Certain highly volatile products are undesirable at high doses (ethanal, ethyl acetate, acetal, etc.).
3. To perform certain chemical transformations favorable to the quality of the spirit. The heating time has a decisive influence on the combination and decomposition of compounds.

This reactive function of the Charente pot still is the source of chemical reactions such as:

Esterification/hydrolysis:

$$\text{Acid} + \text{Alcohol} \rightleftharpoons \text{Ester} + \text{Water}.$$

Acetal formation:

$$\text{Aldehyde} + \text{Alcohol} \rightleftharpoons \text{Acetal} + \text{Water}.$$

Maillard reaction:

$$\text{Sugar} + \text{Amino acid} \rightarrow \text{Pyrazine, Furans (cocoa flavors)}.$$

Table 2 Effect of lees on ester contents (mg l^{-1} spirit at 70% vol.)

	Distillation	
Constituents	*With few lees*	*With lees*
Ethyl caproate	6.76	8.3
Ethyl caprylate	8.95	23.6
Ethyl caprate	13.8	63.0
Ethyl laurate	12.45	36.2
Ethyl myristate	5.4	9.8
Ethyl palmitate	9.77	13.2
Ethyl palmitoleate	1.44	1.8
Ethyl stearate	0.59	0.61
Ethyl oleate	1.19	1.22
Ethyl linoleate	7.69	9.52
Ethyl linolenate	1.86	2.58
Isoamyl caprylate	0.42	2.48
Isoamyl caprate	1.67	5.76
Isoamyl laurate	0.78	1.83
2-Phenylethyl caprylate	Trace	1.20
2-Phenylethyl caprate	0.25	1.55
Total aromatic esters	73.02	182.65 (= + 150%)

Strecker degradation:

$$\alpha\text{-Amino acid} \rightarrow \text{Aldehydes} \rightarrow \text{Acetals}.$$

In order to prevent a coarse character appearing in the spirits, the wines are generally racked in order to leave only part of the lees. In keeping with tradition and local custom, the natural wine is distilled on fine lees. But of course, each firm gives its distillers specific instructions in order to obtain products in keeping with the distinctive types they each market. These instructions concern in particular the proportion of fine lees that must be kept in the wines used. Obviously, as a result, there are considerable analytical differences among the various finished products. The presence of lees results in a greater quantity of aromatic esters, which have a high sensory impact, and produces spirits higher in floral aromas (Table 2). Care must be taken that the lees are of good quality; if not, they are a major source of defects.

Given the very high number of chemical classes present in wines, all wine aromas that are volatile or carried over by water–alcohol vapors do not pass over with the same speed during distillation.

Obtaining a good-quality spirit necessarily implies a harmonious balance of all the constituents: the distiller's entire art is required to achieve this balance.

Aging Cognac

The Contribution to Product Diversity

Aging is a basic factor in the quality of Cognacs; it is also the most costly. Cognac can be kept for many years in oak barrels; the longer the aging, the more the intrinsic quality is enhanced. In order to maintain this quality, the producer is very careful in producing and aging the product.

Aging is an active process, and there are several parameters that must be controlled:

- the cellar or chai;
- the humidity;
- the temperature;
- the type of container used for aging.

The cellar may have an earthen floor, which provides good humidity, or it may be made of cement, in which case, the humidity is lower. The degree of humidity influences the evaporation rate. In a dry cellar, evaporation will consist mainly of water loss. In a very humid cellar, evaporation will affect mainly the alcoholic strength as alcohol is lost; Cognacs will then be smoother and mellower, and have a greater 'rancio' taste.

Difference in Cognac temperature from 7 °C on average in winter to 22 °C in summer are reasonable. This is one of the factors in the ability of a cellar to age spirits. Chemical reactions occurring during aging depend on the temperature (esterification, for example).

There are several types of containers. There are 1000- and 50 000-l barrels (for delivery of Cognac and blends) and casks of 100, 270, 350, 400, and 550 l (aging). Currently, most casks are of the 350-l type manufactured by coopers (cask makers). The cask performs several functions during aging. (*See* **Barrels**: Wines, Spirits, and Other Beverages.)

Quality Factors

The oak imparts aromatic elements to the Cognac in greater or lesser amounts according to:

- the choice of oak – coarse grain oak imparts tannins more easily than fine grain;
- the amount of time the staves are allowed to dry and their thickness;
- the method of heating when bending the staves to make the casks;
- the alcoholic strength of the spirits during aging;
- the age of the cask or barrel;
- the grape production zone ('crus') and the distillation method.

Oak wood consists of cellulose (wood fiber), hemicelluloses, lignin, and tannins. Tannins are highly soluble in spirits.

The first notable modification during aging is the change in color. This phenomenon is directly linked to the extraction and oxidation of tannins. As a direct result, a young Cognac in a new cask will color

quickly, even excessively, if kept there for a long time, whereas a young Cognac in an old cask will change color more slowly, remaining clear even after several years. Excess tannin (bitter taste) is often associated with a very high degree of extracted colour. (*See* **Tannins and Polyphenols**.)

The aromatic aldehydes, mainly vanillin, appear when the lignin breaks down. They are produced by aging and are very important for the bouquet of Cognac. Different types of casks result in substantial variations to their contents of these substances. These aldehydes, predominantly vanillin, contribute a vanilla note, which is not too heavy and is greatly appreciated.

The stave acts as a selective membrane enabling exchanges to take place between the Cognac within the cask and the air in the cellar by way of the evaporation of the most volatile and smallest molecules, a phenomenon that contributes to oxidation.

The quality of a Cognac's aging is measured by its 'rancio taste' (hydrolyses of fatty acid esters together with oxidation and transformation into ketones). The formation of the rancio taste during aging is accompanied by the continual extraction of tannins, slow oxidation, and various chemical reactions. (*See* **Fatty Acids**: Properties.)

The conditions necessary to make a good cask container are:

- the use of good-quality wood;
- the use of wood dried in the open air, exposed to inclement weather for at least 3 years;
- the careful preparation of a new cask and the judicious racking into older casks after proper storage of the recently distilled Cognac in new casks.

Important research work on the aging of Cognac spirits has been a major contribution in identifying the molecules extracted from the oak by the spirits and in understanding the reaction mechanisms underlying the aging process.

The development of gas–liquid chromatography (GLC) and high-performance liquid chromatography (HPLC) techniques has enabled the analytical changes noted during aging to be quantified with greater detail. Several aspects of variations observed in the aging of Cognacs are shown in Tables 3 and 4 by way of example. The extraction of gallic acid is much greater in a cask made of coarse grain oak, $35.7\,mg\,l^{-1}$ compared with $14.5\,mg\,l^{-1}$ in a cask made of fine grain wood. In an old cask, the extration of phenolic compounds is less (Table 3), but the oxidation and exchange with the atmosphere continue to occur. (*See* **Pesticides and Herbicides**: Types, Uses, and Determination of Herbicides; **Phenolic Compounds**.)

Table 3 Effect of cask on a spirit aged 13 years

	Concentrations ($mg\,l^{-1}$)	
Constituent	*New cask*	*Old cask*
Gallic acid	15.3	5.4
Vanillic acid	2.8	1.0
Syringic acid	7.0	2.2
5-Hydroxymethylfurfural	6.3	1.8
Furfural	21.3	10.8
5-Methylfurfural	1.6	0.2
Vanillin	8.8	2.9
Syringaldehyde	17.6	4.6
Coniferaldehyde	6.7	0.6
Sinapaldehyde	17.0	1.7

Table 4 Effect of length of aging period

	Concentrations ($mg\,l^{-1}$)		
Constituent	*0.7 years*	*5 years*	*13 years*
Gallic acid	4.6	9.0	15.3
Vanillic acid	0.3	1.4	2.8
Syringic acid	0.6	2.6	7.0
5-Hydroxymethylfurfural	4.2	4.2	6.3
Furfural	26.8	24.7	21.3
5-Methylfurfural	1.5	1.4	1.6
Vanillin	0.9	4.4	8.8
Syringaldehyde	2.25	8.9	17.6
Coniferaldehyde	3.65	5.9	6.7
Sinapaldehyde	9.45	17.8	17.0

Cognac is sold throughout the world mainly under commercial designations; among those currently in use are ☆☆☆, VSOP, Napoleon, and X.O. These commercial qualities correspond with products of different ages. The price increases with age. These commercial products are obtained through blends (also called 'coupes') of different cognacs. Each dealer has their own blending technique and style. The blend of different Cognacs is above all a subjective choice. It enables a representative product of constant and reproducible quality to be obtained. Blending is therefore an important step in the production of Cognac, since it is the key link between production and marketing.

Sensory Analysis

Cognac spirits are obtained from different 'crus' or growth zones within the area of designated origin: Grande Champagne, Petite Champagne, Borderies, Fins Bois, Bons Bois, and Bois Ordinaires. They show major sensory differences (see Table 5). The blending of several crus influences the quality of the product. (*See* **Sensory Evaluation**: Sensory Characteristics of Human Foods.)

Table 5 Main organoleptic characteristics of spirits from different Cognac crus

Crus	*Sensory characteristics*
Grande Champagne (GC)	Subtle fragrance, finish floral, fruity, persistent aromas, slightly viney
Petite Champagne (PC)	Elegant, added distinction with aging, slightly heavier than GC, stronger alcoholic aroma, supple
Borderies	Discreet violet aroma, ages more rapidly than the GC and PC 'crus', floral aromas + vinosity, fine, elegant
Fins Bois	Powerful, rich and well balanced, vinous, fruity grapey aromas, heavier alcohol aromas
Bons Bois	Less elegant than Fins Bois, heavier, rough, less persistent aromas
Bois Ordinaries	Alcoholic aroma, heaviness, vinosity, rough

During aging, the differences are increased with regard to both the quality of the flavors and the length of the aging. A fins bois spirit is aged over a maximum period of 30 years, whereas a Grande Champagne spirit can be aged in casks over 60 years. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Initial blending of new spirits provides more uniform batches.

Product Authenticity

The production of cognac spirits is subjected to very strict regulations with regard to both production and marketing. Regulations deriving from the Decrees of 15 May 1936 and 13 January 1938 are supplemented by genuine and consistent application of local custom for the area, based on several parameters:

- the area of registered origin;
- the choice of grape varieties;
- the materials used in the wine-making;
- the distillation method;
- the containers for the spirits;
- the minimum period of aging for each of the commercial designations.

Various governmental authorities are responsible for ensuring that these regulations are followed:

- the Institut National des Appellations d'Origines (INAO);
- the French internal revenue service (Direction Générale des Impôts, or DGI);
- the Directorate-General for Consumer Affairs, Competition and Fraud (Direction Générale de la consommation, de la Concurrence et de la Répression des Fraudes, or DGCCRF);
- Customs Administration.

The Bureau National Interprofessionnel du Cognac (BNIC) carries out inspections for age and issues certificates for age, designation of origin, and analysis, certifying the authenticity of the products placed on the market.

See also: **Barrels**: Wines, Spirits, and Other Beverages; **Fatty Acids**: Properties; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Pesticides and Herbicides**: Types, Uses, and Determination of Herbicides; **Phenolic Compounds**; **Sensory Evaluation**: Sensory Characteristics of Human Foods; **Tannins and Polyphenols**; **Wines**: Production of Table Wines

Further Reading

Cantagrel R, Lurton L, Vidal JP and Galy B (1990) La distillation Charentaise pour l'obtention des eaux-de-vie de Cognac. *1er Symposium International, les Eaux-de-vie Traditionnelles d'Origine Viticole*, Bordeaux, 26–30 June.

Cantagrel R, Maignial L, Ferrari G and Snakkers G (1998) Elaboration du Cognac. La qualité se protège, gage de sécurité pour le consommateur. *Bulletin de l'OIV*. 71: 1029–1057.

Cantagrel R, Mazerolles G, Vidal JP, Lablanquie O and Boulesteix JM (1990) L'assemblage: une étape importante dans le processus d'élaboration des Cognacs. *1er Symposium International, les Eaux-de-vie Traditionnelles d'Origine Viticole*, Bordeaux, 26–30 June.

Goldberg DM, Hoffmann B, Yang J and Soleas GJ (1999) Phenolic constituents, furans and total antioxidants status of distilled spirits. *Journal of Agricultural and Food Chemistry* 47: 3978–4002.

ter Heide R, de Valois PJ, Visser J, Jaegers PP and Timmer R (1978) Concentration and identification of trace constituents in alcoholic beverages. In: Charalambous G (ed.) *Analysis of Foods and Beverages. Headspace Techniques*, pp. 249–281. London: Academic Press.

Joseph E and Marche M (1972) Contribution à l'étude du vieillissement du Cognac. Identification de la scopolétine, de l'aesculine, de l'ombelliférone, de la β-méthyl ombelliférone, de l'aesculine et de la scopoline, hétérosides provenant du bois. *Connaissance de la Vigne et du Vin* 6(3): 1–58.

Lafon J (1971) Problèmes actuels de technologie et de vieillissement du Cognac et des eaux-de-vie de vin. *Bulletin de l'OIV*, No. 482, April, pp. 339–355.

Lafon J, Couillaud P and Gay-Bellile F (1973) *Le Cognac, sa Distillation*. Paris: Editions J. B. Baillière.

Leaute R (1990) Distillation in alambic. *American Journal of Enology and Viticulture* 41(1): 90–103.

Poiron T (1997) *La distillation charentaise dans l'élaboration d'une eau-de-vie de Cognac*. Thèse, Doctorat en

Pharmacie, Faculté de Médecine et de Pharmacie, Université de Poitiers, France.

Puech JL (1986) Le vieillissement des eaux-de-vie. *Revue Française d'Oenologie* 102: 11–16.

Puech JL, Leaute R, Clot G, Nomdedeu and Mondies H (1982) Etude de la lignine et de ses produits de dégradation dans les eaux-de-vie de Cognac. *Bulletin de liaison No. 11 du Groupe Polyphénols, Compte-rendu des Journées Internationales d'Etude et de l'Assemblé Générale Toulouse*, 29–30 September, 1 October, pp. 605–611.

Puech JL, Leaute R, Clot G, Nomdedeu L and Mondies H (1984) Evolution de divers constituants volatils et phénoliques des eaux-de-vie de Cognac au cours du vieillissement. *Sciences des Aliments* 4(1): 65–80.

Viriot C (1995) *Polyphénols et polyosides du bois de chêne. Contribution au vieillissement des Cognacs en fût*. Thèse, Institut National Agronomique Paris-Grignon, France.

Vlassov VN and Maruzhenkov DS (1999) Application of GC/MS method for the identification of brandies and Cognacs. *Analysis* 27: 663–667.

BRASSICAS

J W Fahey, Johns Hopkins University, Baltimore, MD, USA

Background

The brassicas comprise a large and diverse group of widely consumed vegetables. Brassica is the Latin name of a genus that is taxonomically placed within the Brassicaceae (Cruciferae), which is one of the 10 most economically important plant families in the world. The genus *Brassica* includes, but is not limited to, the following vegetables: bok choy, broccoli (calabrese, sprouting broccoli), Brussels sprouts, cabbage, cauliflower, Chinese cabbage, kale, kohlrabi, mizuna, swede (rutabaga), and turnips. Rapeseed or canola, one of the world's most widely grown oilseed crops, is also a nonvegetable *Brassica* species and is discussed elsewhere. Other, closely related vegetables within the Brassicaceae family (also known as brassica or cruciferous vegetables, crucifers, cole crops, or mustards) include radish (*Raphanus sativus*), watercress (*Nasturtium officinale*), arugula (*Eruca sativa*), horseradish (*Armoracia rusticana*), maca (*Lepidium meyenni*), mashua (*Tropaeolum tuberosum*), wasabi (*Wasabia japonica*), and cress (*Lepidium sativum*). One of the common characteristics of brassica vegetables is that they all contain glucosinolates, sulfur-containing compounds that are hydrolyzed to produce so-called mustard oils, which impart characteristic tastes and odors to these vegetables.

Vegetables are considered to be an essential part of a balanced diet. The brassica vegetables in particular, add considerable visual or esthetic appeal to a meal, since many of them are leafy and green. They are rich sources of dietary fiber, any many are good sources of calcium, provitamin A, vitamin C, and certain beneficial phytochemicals. They also have distinctive flavors and textures, which enhance palatability in the eyes of many consumers, though they evoke quite violent negative emotions in others, prodding them to wax poetic:

> Lurking in basement kitchens, consorting with boiled beef, tripe, watery-fleshed German cooks, and roaches, the kohlrabi, with darkened eyes, has designs upon the very young. Bilious whitey-green, pretending that its thickened basal part is a vegetable, it awaits the pink mouths of children, corrupting thick-witted kitchen help in their seduction.
>
> Oh, it is the whole puddle and sink of all sins against God and man, the kohlrabi.
>
> Think of it, hiding in vegetable bins like Nazis in Argentina.
>
> Noxious white root, it is the very nose and vegetable essence of Andy Warhol, silk-screened on pale velvet and hawked along the highways of America.
>
> The Kohlrabi, by William A. Fahey

Historical Origins of Modern Brassica Vegetables

Most scholars place the wild cabbage at the head of the family tree that has led to the development of the 400 or so varieties of *Brassica oleracea*. Although there is still considerable disagreement among scientists as to precisely how all of the modern brassica vegetables arose, there are widespread references to these vegetables in writings dating back many centuries. Their taxonomy, too, is a subject of considerable debate and should be regarded as being in a state of flux. Genus, species, variety, subvariety, subspecies, botanical group, and cultivar designations are frequently interchanged in the descriptive literature. Commonly recognized taxonomic designations are used herein, wherever possible.

It is widely agreed that there are six major vegetable *Brassica* species. Three are monogenomic, and three are amphidiploid (being diploid for two genomes, each originally contributed by a different species). These major categories are thus commonly designated as having A, B, or C genomic compositions, and their relationships are shown in Table 1.

The ancestral wild cabbage was almost certainly a seaside plant of northern European or Mediterranean origin. All of the wild brassicas today occur in cliffs and rocky islets in fairly isolated places. Wild *Brassica oleracea* varieties still grow as perennials along the coasts of northern Spain, western France, and southern and southwestern Britain. Over time, some 400 varieties have been created, which include cabbages, kohlrabi, oilseed rape, Brussels sprouts, cauliflower, and broccoli. Cabbages were not eaten by the Hebrews or the Egyptians. None of the crucifers are mentioned in the Old Testament, and the only crucifer mentioned in the New Testament (Mark 4:30–32) is mustardseed. The ancient Romans and Greeks, however, were quite familiar with cabbages and cauliflower, and believed that eating cabbage during a banquet would prevent one from becoming drunk. Dietary cabbage intake was discussed in the writings of Pythagoras, Diogenes, and Cato, to name a few. Cato, for example, recommended cabbage in the diet to prevent disease and prolong life, and claimed to owe the existence of his 28 sons to cabbage. In the Middle Ages, cabbage plasters were used for medicinal purposes, and cabbage was used for cough syrup and wound dressings. In the 1700s, they were used aboard ships for their vitamin C content and were used as dressings to combat gangrene.

Broccoli (from the Italian 'brocco,' arm or branch) is widely presumed to have developed from the wild cabbage that was native to coastal Europe and spread through the Near East to the Orient between 2000 and 2500 years ago. Some authorities consider that sprouting broccoli (something resembling modern-day broccoli) was first domesticated and cultivated in Italy during ancient Roman times. A vegetable that was probably broccoli was described by the Roman botanist Pliny (first century CE). There is no consensus, however, on the translation of the early name 'cyma,' and there is concern that the writings of early botanists may have confused 'broccoli' and 'cauliflower' as we know them today. Others maintain that early selection and domestication of broccoli may have been made in Asia Minor, with a cultivated form being brought to Italy by early traders in the 1500s. What is certain, however, is that broccoli (also known at that time as Italian asparagus or sprouting broccoli), was introduced to England around the 1720s. The following passage from Stephen Switzer's *The Practical Kitchen Gardiner* (1727) illustrates the point:

> As for the broccoli, there are three kinds of it, one of which yields sprouts button'd at their points, or headed like small collyflowers; another sort with curl'd leaves, which produce sprouts button'd on the points like asparagus; and a third with curl'd leaves of a pale green colour, which yield sprouts like the red kind; the two are to be had at several places about London; but the first is very rare to be had, but from some few gentlemen that have them yearly from Italy…

By the late 1700s, broccoli was introduced to the American colonies, where it was grown by Italian immigrants on the east coast of the USA but was otherwise little known until the 1920s. In 1912, the Stokes seed company brought broccoli seed into the USA and started selling to growers in 1918. In 1923, the D'Arrigo Brothers Company initiated field trials in California and, by 1925, was shipping ice-pak freight car loads of broccoli back to the east coast.

Commonly Cultivated Brassica Vegetables

Broccoli, Including Broccoli Sprouts (*Brassica oleracea* var. *italica*)

This is also known as calabrese or sprouting broccoli. The most valued portions of broccoli plants are the heads, which are inflorescences consisting of immature fully differentiated flower buds and tender upper stems. Both primary and secondary inflorescences are eaten, and lower stems are eaten, too but are not as prized due to their tough outer 'rind.' Broccoli is available commercially as fresh or frozen florets and is used raw in salads or as vegetable crudités. It is also frequently cooked and served by itself as well as being a component of many cooked and stir-fried dishes. There are over 100 commercial hybrid cultivars of broccoli, derived from a limited number of landraces or open pollinated cultivars that include purple sprouting, purple cape, purple Sicilian, white sprouting, and calabrese or green sprouting broccoli. Cut broccoli shoots or florets are very perishable and must thus be cooled (e.g., vacuum cooling) very soon after picking. Crushed ice or an ice slurry is typically blown into cartons of broccoli within a few hours of their being picked. California and Mexico in North America, and Italy, France, and Spain in Europe are the major production areas. Consumption of broccoli in the USA has been steadily climbing since about 1970. In 1970, the total per-capita consumption was about 0.7 kg and is presently about 3.6 kg. The development of hybrids in the late 1970s and their subsequent marketing by the vegetable seed companies

Table 1 Nutrient composition per 100 g serving of selected brassica vegetables[a]

Nutrients:	*Broccoli*	*Brussels sprouts*	*Cabbage*	*Red cabbage*	*Savoy cabbage*	*Pak choi*	*Pei tsai*	*Cauliflower*	*Collards*	*Cress*	*Kale*	*Kohlrhabi*	*Mustard Greens*	*Tendergreens*	*Swede*
Gross composition															
Water (g)	90.69	86.00	92.15	91.55	91.00	95.32	94.39	91.91	90.55	89.4	84.46	91.00	90.80	92.20	89.66
Energy (kcal)	28	43	25	27	27	13	16	25	30	32	50	27	26	22	36
Energy (kJ)	117	180	105	113	113	54	67	105	126	134	209	113	109	92	151
Protein (N × 5.95) (g)	2.98	3.38	1.44	1.39	2.00	1.50	1.20	1.98	2.45	2.60	3.30	1.70	2.70	2.20	1.20
Total lipid (g)	0.35	0.30	0.27	0.26	0.10	0.20	0.20	0.21	0.42	0.70	0.70	0.10	0.20	0.30	0.20
Carbohydrate (g)	5.24	8.96	5.43	6.12	6.10	2.18	3.23	5.20	5.69	5.50	10.01	6.20	4.90	3.90	8.13
Fiber, total dietary (g)	3.0	3.8	2.3	2.0	3.1	1.0	3.1	2.5	3.6	1.1	2.0	3.6	3.3	2.8	2.5
Ash (g)	0.92	1.37	0.71	0.68	0.80	0.80	0.98	0.71	0.89	1.80	1.53	1.00	1.40	1.40	0.81
Sugars, total (g)															
Minerals															
Calcium (mg)	48	42	47	51	35	105	77	22	145	81	135	24	103	210	47
Iron (mg)	0.88	0.50	0.59	0.49	0.40	0.80	0.31	0.44	0.19	1.30	1.70	0.40	1.46	1.50	0.52
Magnesium (mg)	25	23	15	15	28	19	13	15	9	38	34	19	32	11	23
Phosphorus (mg)	66	69	23	42	42	37	29	44	10	76	56	46	43	28	58
Potassium (mg)	325	389	246	206	230	252	238	303	169	606	447	350	354	449	337
Sodium (mg)	27	25	18	11	28	65	9	30	20	14	43	20	25	21	20
Zinc (mg)	0.40	0.42	0.18	0.21	0.27	0.19	0.23	0.28	0.13	0.23	0.44	0.03	0.20	0.17	0.34
Copper (mg)	0.045	0.070	0.23	0.097	0.062	0.021	0.036	0.042	0.039	0.170	0.290	0.129	0.147	0.075	0.040
Manganese (mg)	0.229	0.337	0.159	0.180	0.180	0.159	0.196	0.156	0.276	0.553	0.774	0.139	0.480	0.407	0.170
Selenium (μg)	3.0	1.6	0.9	0.9	0.9	0.5	0.6	0.6	1.3	0.9	0.9	0.7	0.9	0.8	0.7

Vitamins															
Ascorbic acid (mg)	93.2	85.0	32.2	57.0	31.0	45	27	46.4	35.3	69	120	62	70.0	130	25
Thiamin (mg)	0.065	0.139	0.050	0.050	0.070	0.040	0.040	0.057	0.054	0.080	0.110	0.050	0.080	0.068	0.090
Riboflavin (mg)	0.119	0.090	0.040	0.030	0.030	0.070	0.050	0.063	0.130	0.260	0.130	0.020	0.110	0.093	0.040
Niacin (mg)	0.638	0.745	0.300	0.300	0.300	0.500	0.400	0.526	0.742	1.000	1.00	0.400	0.800	0.678	0.700
Pantothenic acid (mg)	0.535	0.309	0.140	0.324	0.187	0.088	0.105	0.652	0.267	0.242	0.091	0.165	0.210	0.178	0.160
Vit B-6 (mg)	0.159	0.219	0.096	0.210	0.190	0.194	0.232	0.222	0.165	0.247	0.271	0.150	0.180	0.153	0.100
Folate (μg)	71	61	43	21	80	66	79	57	166	80	29	16	187	159	21
Vit B-12 (μg)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Vit A (IU)	1542	883	133	40	1000	3000	1200	19	3824	9300	8900	36	5300	9900	580
Vit A, RE (μg)	154	88	13	4	100	300	120	2	382	930	890	4	530	990	58
Vit D (IU)															
Vit E, α-TE (mg)	1.660	0.88	0.105	0.105	0.105	0.120	0.120	0.040	2.260	0.700	0.800	0.480	2.010	1.704	0.300
Lipids															
Saturated, total (g)	0.054	0.062	0.033	0.034	0.013	0.026	0.043	0.032	0.055	0.023	0.091	0.013	0.010	0.015	0.027
Monounsaturated (g)	0.024	0.023	0.019	0.019	0.007	0.015	0.017	0.014	0.030	0.239	0.052	0.007	0.092	0.138	0.025
Polyunsaturated (g)	0.167	0.153	0.122	0.125	0.048	0.096	0.072	0.099	0.201	0.228	0.338	0.048	0.038	0.057	0.088
Cholesterol (mg)	00	0	0	0	0	0	0	0	0	0	0	0	0	0	
Phytosterols (mg)	24	11						18							
Pigments (carotenoids)[b]															
β-carotene (μg)	779	450	65	3323							9226				
lutein/zeaxanthin (μg)	2445	1590	310								39550				
lycopene (μg)	0	0	0								0				
Refuse[c] (% of total)	39	10	20	20	20	12	7	61	43	29	39	54	7	7	15

[a]Data from USDA Nutrient Database for Standard Reference, Food Group 11, Release 13 (1999).

[b]Data from USDA-NCC Carotenoid Database for U. S. Foods – 1998; (not all vegetables have been analyzed for all carotenoids).

[c]Refuse is some combination of the total biomass, which is typically not used in food preparation, e.g. stems; crowns; spoiled, damaged croute leaves; leaf stalks; cores; trimmings; or root bases.

were responsible for much of the increased consumption in the 1970s and 1980s. Impetus for the dramatic increase in consumption over the past 10 or so years has come from the health aura which broccoli has recently enjoyed. It is regularly identified as the vegetable eaten most often for health reasons, including cancer prevention, high fiber, vitamin C, folate, and calcium content.

Broccoli raab (*Brassica rapa* var. *rapa*)

Also known as broccoli rabe, broccoli babe, rapini, broccoletti di rape, brocoletto, turnip broccoli, cima di rapa, Italian turnip, sparachetti, taitcat. Broccoli raab has also been classified botanically as *B. campestris* or *B. ruvo*. It is a bitter tasting vegetable similar in appearance to Chinese kale, but nonheading and with thinner stems and smaller flowers than broccoli. It is long been considered a delicacy in Italy, typically lightly sautéd with garlic. It has recently gained interest in the USA and is considered a 'new' vegetable by some.

Brussels sprouts (*Brassica oleracea* var. *gemmifera*)

Brussels sprouts are a relatively young member of the Brassica family. They originated in Belgium in the 1500s and by the 1700s were appearing on tables around the world. Brussels sprouts grow as a tall (*c.*1 m) single stem biennial from which the axillary buds, resembling miniature cabbage heads, are harvested. They are generally eaten after cooking (steaming or boiling) and are available commercially either fresh or frozen. Brussels sprouts require a long growing season, and vegetable quality is adversely affected by warm weather.

Cabbage (*Brassica oleracea* var. *capitata*)

Cabbage has been cultivated, and even revered, as a vegetable by the ancient Greeks as far back as 2600 years ago, and it has also had a long history of medicinal use. There are scores of references to its use for such diverse purposes as the prevention of drunkenness, headache, stomach ailments, and even cancer. Cabbage leaves have long been used as poultices for application to tumors, and even in modern times, they have been under investigation as a means for preventing or treating breast engorgement in nursing mothers.

All cabbages have heads formed from tightly packed leaves. There are many distinct head types, the most common being: Wakefield (with small, early, white, pointed heads; for fresh market), Red (leaf surfaces are pigmented, medium very firm, round heads; for fresh market and storage); Ballhead or Danish (round, very firm heads, light green leaves; for fresh market and storage); Savoy (some authorities designate this *Brassica oleracea* var. *sabauda*) (round, loose heads with crinkled or blistered leaves; for fresh market and storage). Cabbage is typically eaten fresh, processed into a salad-like dish called coleslaw, boiled, or as a fermented and pickled product called sauerkraut. Cabbages are important as a fresh market crop as well as a processing crop in most parts of the world and rank in the top 10 vegetables in both sales and volume in North America and much of Europe.

Cauliflower (*Brassica oleracea* var. *botrytis*)

The edible portion or head of cauliflower is composed of curds. In contrast to the head of broccoli, cauliflower heads consist of tightly clumped, undifferentiated shoot apices on top of hypertrophied, highly branched fleshy stem tissue. Referred to by Mark Twain as 'nothing but a cabbage with a college education,' cauliflower curds are actually degenerate shoot tips which are most frequently white in color (lacking chlorophyl), although purple, green, and orange cultivars now exist. Cauliflower is not as cold-tolerant as many other brassica vegetables. It is grown commercially in France, Italy, the UK, and North America, and it is eaten in the same manner as broccoli as well as being pickled. It is available commercially as fresh or frozen curds.

Charlock (*Brassica kaber*, also *Brassica arvensis*, *Sinapis arvensis*)

Charlock is also known as kaber. The seeds of this plant, likely of Mediterranean origin, are used as a condiment, and the leaves are eaten as a potherb. The seeds are not as pungent as those of other mustards.

Chinese Cabbage (*Brassica rapa* var. *pekinensis*)

Also known as napa, napa cabbage, pe-tsai, wongbok, chihli. This is a vegetable of major importance in China (over 300 000 ha grown), Korea, Taiwan, and Japan. Grown as an annual crop, most cultivars are biennial and produce tight, compact, cylindrical heads. This vegetable has been cultivated in China for over 1600 years and accounts for a major fraction of the total vegetable consumption in certain (northern) areas of the country. In Korea, it is is fermented to produce (preserved) Kimchi, which is thus a yearlong, ubiquitous commodity in that country.

Chinese Kale (*Brassica oleracea* var. *alboglabra*)

This is also known as gai lan, Chinese broccoli, gai lon, gai larn, kai laan, white-flowered broccoli, fatshan. Compared to broccoli, Chinese kale has more,

slender, dark green leaves, longer stems, and very few florets, which are similar to those of broccoli. Flower buds, flower stalks, and young leaves are consumed, primarily in salads and stir fries. Relatively new to Japan, Western Europe and US cuisines, but extensively grown in Taiwan, South-east Asia and China, Chinese kale is relatively fast growing and heat-tolerant.

Chinese Mustards (*Brassica rapa* ssp. *chinensis*)

These include bok choy, pak choy, choi sum, Shantung cabbage and are also known as Chinese white cabbage and celery mustard. Pak choi and bok choy are sometimes errantly referred to as Chinese cabbages. *Brassica rapa* ssp. *parachinesis*, also known as mock pak choy, choy sum, cai tai, saishin, has been cultivated since the fifth century CE in Asia and continue to be very important vegetables, especially in China. Pak choy is the more leafy cultivar, and bok choy is notable for its massive leaf midribs, which are white and fleshy. This subspecies is a biennial, which is grown as an annual for its edible leaves. Plants can reach 0.6 m tall and can weigh over 2 kg. Leaves are usually consumed fresh but are also dried after blanching, for use through periods when fresh vegetables are not plentiful.

Collards (*Brassica oleracea* var. *sabellica*)

Available commercially as fresh, canned, or frozen leaves, collard greens are popular in the southern USA, where they are grown along the eastern portion of the country. As such, it is a much more heat tolerant crop than its close relative, kale. The plants are nonheading and up to 1.25 m tall, and broad, flat, or slightly furrowed leaves form as a rosette on a minimal stem.

Colza (*Brassica napus* var. *napus*)

Colza is also known as vegetable rape, xi yang you cal, and chou navet. Colza oil (expressed from seeds) is commonly consumed in India and China. Foliage and sprouted seeds are eaten in salads or as a potherb; inflorescences are prepared like broccoli.

Kale (*Brassica oleracea* var. *acephala*)

There are many 'kales,' some of which are classified taxonomically into other *Brassica* species or varieties. Kale includes kitchen kale, green kale, dwarf Siberian kale, marrow stem kale, tronchuda kale, curly leaf kale, Scotch kale, tree kale, and borecole. Although highly variable, kale is characterized by a nonheading rosette-like whorl of foliage, a short, erect stem, and large, upright, curly leaves. They are used mainly for their edible foliage and are generally eaten cooked, but are sold fresh, canned, and frozen, or used as garnish. Among the commonly grown kales are:

- Branching bush kales (sometimes classified as *Brassica oleracea* var. *fruticosa*): also known as cow-kale, borecole (sometimes classified as *B. oleracea* var. *selenesia*), these were often cultivated in the past for their edible foliage and have been used extensively for animal fodder.
- Thousand-headed kale (*Brassica oleracea* var. *ramosa* or *B. oleracea* var. *millecapitata*).
- Inflorescence kales (a term used by some to describe cauliflower, broccoli, and related brassica vegetables).
- Galega kale (also known as couve galega): a traditional and widely grown Portugese nonheading kale with long petioles, large midribbed leaves, and an elongated stem that can reach 3 m; leaves are picked one by one and used for traditional soups or for animal feed.
- Marrow stem kale (*Brassica oleracea* var. *medullosa*): a particularly prolific type of kale used exclusively for animal feed.
- Siberian kale (also classified as *Brassica napus* var. *pabularia*): also known as Hanover salad; a leafy vegetable, similar to collards, which is used fresh in salads and cooked as a potherb.

Kohlrabi (*Brassica oleracea* var. *gongylodes*)

This is also known as knol khol or turnip cabbage. The edible portion of kohlrabi is the fleshy, swollen, tuber-like enlargement of the short, unbranched stem, which may be white, green, or purple and develops just above the surface of the soil. This vegetable developed in northern Europe about five centuries ago. It resembles turnip and rutabaga in flavor and texture but becomes highly fibrous if not harvested at peak maturity.

Mizuna (*Brassica rapa* ssp. *japonica*)

Also known as mibuna, curled mustard, or Japanese greens, mizuna is a cool-tolerant relative of the leafy turnips that has recently been introduced to the West.

Mustard (Various Latin Binomials)

Plants from this diverse group have been used worldwide for centuries as a condiment (mustardseed; also known as black mustard, brown mustard, *Brassica nigra*) and as a vegetable (mustard greens; *Brassica juncea*). Most well-known among the mustards are:

- white or yellow mustard (*Brassica hirta*, *Brassica alba*, *Sinapis alba*);
- chinese mustard (*Brassica japonica*);
- black mustard (*Brassica nigra*, *Sinapis nigra*);

- field mustard (*Brassica campestris*);
- Ethiopian mustard; Abyssinian mustard (*Brassica carinata*).

Many of the so-called brown mustards (*Brassica juncea*) have been assigned unique variety names. Common English and Chinese names are as follows:

- *B. juncea* var. *capitata* (capitata mustard, jie qiu jie);
- *B. juncea* var. *crassicaulis* (bamboo shoot mustard, sun zi jie);
- *B. juncea* var. *crispifolla* (cut leaf mustard, curled mustard, mi tuo jie);
- *B. juncea* var. *foliosa* (small leaf mustard, xiao ye jie);
- *B. juncea* var. *gemmifera* (gemmiferous mustard, bao zi jie);
- *B. juncea* var. *involuta* (involute mustard, juan xin jie);
- *B. juncea* var. *latipa* (wide petiole mustard, kuan bing jie);
- *B. juncea* var. *leucanthus* (white flowered mustard, bao hua jie);
- *B. juncea* var. *linearifolia* (line mustard, feng wei jie);
- *B. juncea* var. *longepetiolata* (long petiole mustard, chang bing jie);
- *B. juncea* var. *megarrhiza* (tuberous-rooted mustard, dal tou jie);
- *B. juncea* var. *multiceps* (tillered mustard, fen nie jie);
- *B. juncea* var. *multisecta* (flower-like leaf mustard, hua ye jie);
- *B. juncea* var. *rugosa* (large leaf, brown mustard, Indian mustard, mustard greens, rai, dai ye jie);
- *B. juncea* var. *strumata* (strumous mustard, tsatsai, zha cai);
- *B. juncea* var. *tumida* (swollen stem mustard, jing liu jie);
- *B. juncea* var. *utilis* (peduncled mustard, tai jie);
- *B. rapa* var. *narinosa* (broad beaked mustard, wu ta cai, taasai).

Swede (*Brassica napus* var. *napobrassica*)

Also known as rutabaga in the USA, swede is considered a root crop, though, technically, this is not accurate. It is an annual crop grown as animal fodder and consumed by human beings after cooking or pickling. It has been grown for about three centuries, originating in Sweden and spreading throughout Europe. The flesh is white or orange and similar in flavor and texture to turnips with equivalent, excellent storage characteristics but low commercial value. Swede is much hardier than the turnip but also takes much longer to mature. The leaves are used as a potherb.

Tendergreen (*Brassica rapa* ssp. *perviridis*)

Also known as spinach mustard, mustard spinach or komatsu-na, this leafy relative of the turnip is reasonably cold-tolerant, surviving temperatures as low as $-15\,^{\circ}C$. It has large, dark green, mild-flavored foliage, which is eaten fresh and pickled, primarily in Korea, Taiwan, and Japan.

Texsel greens (*Brassica carinata*)

Of Ethiopian origin, the early growth of this plant is valued for its high protein and vitamin C content and is eaten raw in salads or lightly boiled as a spinach substitute. It has a milder flavor than collards or mustard greens.

Tronchuda cabbage (*Brassica oleracea* var. *costata*)

Also known as Portugese cabbage, Couve Tronchuda, Galician cabbage, braganza, or sea-kale cabbage, these are loose-headed cabbages that have large leaves with succulent midribs. It is believed that the many landraces of this vegetable arose from an initial hybridization of cabbage and kale.

Turnip (*Brassica rapa* var. *rapifera*)

Turnip is very similar to rutabaga in that it is a root crop (technically incorrect) that produces high amounts of biomass per hectare, is high in starch content, and has very favorable storage characteristics. It appears to have been around for about 4000 years, originating in eastern Europe and Siberia and gradually spreading across Europe. As with swedes, turnips are generally eaten after cooking and can also be processed for use in pickled or mixed vegetables. Turnip greens are eaten in-season as a fresh leafy green vegetable.

Turnip rape (*Brassica campestris* var. *oleifera*)

The seeds of this plant produce an oil that is sometimes used in cooking, and it has relatively high levels of unsaturated lipids. It is becoming more popular as an oilseed but is distinct from the very widely grown oilseed rape (*B. napus* var. *oleifera*). The foliage of the plant is used as a potherb and garnish.

'New' Brassica Vegetables

A number of 'new' brassica vegetables have been produced under trade names, primarily by cross-hybridization between existing taxa. These include:

- broccolini: a cross between Chinese kale and broccoli trademarked by Mann Packing Co. (California, USA);
- asparation: a cross between Chinese kale and broccoli trademarked by Sakata Seed Inc. (California, USA);
- broccoflower (*Brassica oleracea* var. *botrytis*): a bright green cauliflower originating in Holland, and trademarked by T & A (Tanimura & Antle, California, USA) about a decade ago.

Regional Preferences

Brassica vegetables are a large group of primarily herbaceous plants that includes a number of the world's most commonly cultivated vegetables. Though the progenitor species likely originated in the Mediterranean region, the cultivated brassica vegetables are of cosmopolitan distribution. Cultivars have been adapted for worldwide production, from the tropics to the Arctic Circle. The largest cabbages in the world have, in fact, been grown near Fairbanks, Alaska, USA.

Brassica vegetables include a large number of taxonomically closely related, but morphologically and organoleptically diverse, plants. These species have been cultivated for many centuries and have been extensively crossed and hybridized. Many cultivars have been developed in microenvironments or very small geographical regions where they have remained essentially isolated for decades, even centuries. For example, certain small villages or regions in Italy have their own very distinctive broccoli cultivars. Since these vegetable gene pools have remained isolated for hundreds of generations of selection, there has been considerable development of phenotypes that diverge from a common ancestor. The relationships of some of the more common brassicas are detailed in Figure 1.

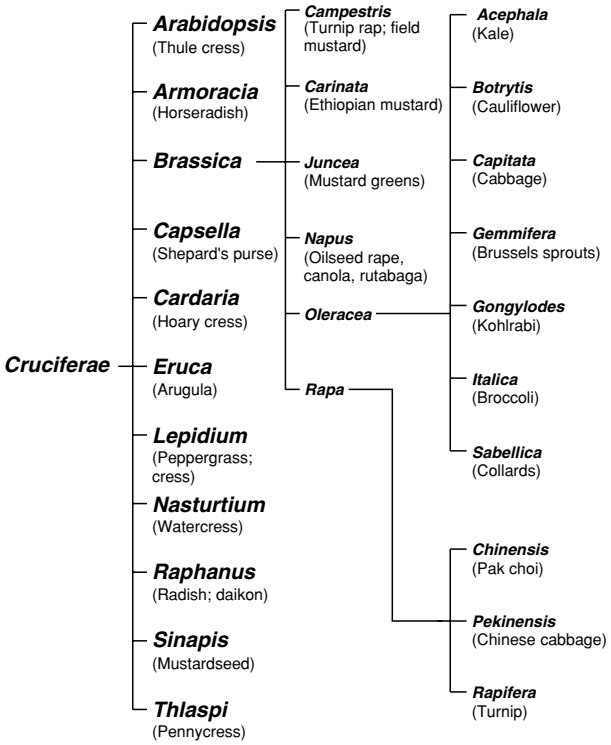

Figure 1

Dietary and Commercial Importance

From a nutritional standpoint, these vegetables are perhaps best recognized as excellent sources of vitamin C, fiber, calcium, and certain phytochemicals such as carotenoids (provitamin A) and glucosinolates (which have been the source of considerable recent scientific research as cancer protective agents). The contents of vitamins, minerals, and phytochemicals are, however, highly dependent on both genetic and environmental variables. Plant cultivar or variety, the environment in which the plant is grown (e.g., amount of sunlight, drought stress, temperature), the conditions under which it is harvested, stored, and transported to market, and the way in which it is prepared for the table and consumed all play key roles in the ultimate nutritional value of that vegetable. Water-soluble components such as vitamin C and glucosinolates (see below) are easily leached out in the pot liquor during cooking. These and other beneficial chemicals can exhibit a tremendous gradient from one portion of the plant to another, and it is often quite difficult to assess these differences without performing sophisticated chemical analyses. Nonetheless, the brassica vegetables remain among the best sources of the dietary components mentioned above and should be consumed regularly as part of a diet rich in fruit and vegetables.

In the West, broccoli, Brussels sprouts, cauliflower, cabbage, and kale are the most significant brassica components of the diet. In the East, the so-called oriental brassicas (Chinese cabbage, tendergreen (spinach mustard), bok choy, pak choy, mizuna, celery mustard, and Chinese mustard) as well as spinach mustard and mizuna greens, Chinese kale (Chinese broccoli), and daikon (Japanese radish – a cruciferous root crop) are the main brassica vegetables of commercial and dietary importance.

Cultivation and Postharvest

The brassica vegetables are hardy, cool-season vegetables that grow best in temperatures in the range of

15–20 °C and have similar cultural requirements. Cabbage plants that have been hardened off can tolerate temperatures as low as −4 °C for short bursts, and broccoli and cauliflower plants thrive in light frosts. Almost all cole crops decline in quality when temperatures are in excess of 26–27 °C. Under irrigation (guaranteed water supply), broccoli crops can be harvested in 10–15 weeks from direct seeding, cabbage crops in 13–17 weeks, and cauliflower crops in about 13–16 weeks, depending on the temperature and climate. Brassica vegetables can be grown on a wide range of soil types, and in addition to direct seeding, many crops are now grown from glasshouse transplants. All brassica vegetables are more or less susceptible to the same diseases and insect pests. More specific agronomic advice is beyond the scope of this article.

Although some brassica crops do not need cooling after harvest, broccoli and cauliflower require immediate prechilling to 4 °C. These crops are typically hydrocooled or immersed in an ice-slurry. They must be refrigerated for transport and storage and have a storage life of about 2 weeks (broccoli), to 4 weeks (cauliflower), to as much as 6 months for some cabbage varieties grown in cooler climates.

Nutritional Value and Chemical Composition

Table 2 summarizes the nutritional value and chemical composition of broccoli, Brussels sprouts, cabbage, red cabbage, savoy cabbage, Chinese cabbage, cauliflower, collards, cress, kale, kohlrhabi, mustard greens, tendergreen, and swede.

Vitamin A, Vitamin C, Selenium, Calcium, and Fiber

The brassica vegetables are excellent sources of calcium, fiber, vitamin A (in the form of β-carotene, or provitamin A, and vitamin C (especially broccoli, kale, and tendergreen). The element selenium can be incorporated into the tissues of brassica vegetables (in particular, broccoli), where it can reach rather high levels, and the vegetable may be of therapeutic value as a source of this antioxidant element due to its special availability from such tissues.

Table 2 Haploid chromosome number and genomic compositions of the six major *Brassica* species

Species	*Haploid chromosome number*	*Genome*
B. rapa	10	A
B. nigra	8	B
B. oleracea	9	C
B. juncea	18	AB
B. napus	19	AC
B. carinata	17	BC

Phytochemical Attributes (e.g., Glucosinolates/isothiocyanates, Carotenoids, and Flavonoids)

All of the brassica vegetables contain glucosinolates, at concentrations of up to 3% by weight in the seeds of some plants. These sulfur-containing compounds and their breakdown products have long been known for their fungicidal, bactericidal, nematocidal, and allelopathic properties, as well as for the goitrogenic or antinutritional glucosinolates in the protein-rich, defatted meal from widely grown oilseed crops (e.g., rapeseed) and in some domesticated brassica vegetables (e.g., Brussels sprouts). When used as animal feed, rapeseed meal can have pronounced deleterious health consequences on livestock due to ingestion of excessive quantities of 'progoitrin'; (e.g., when livestock are fed a meal produced from defatted rapeseed containing the progoitrin, which may interfere with thyroxine production, drastically reducing iodine supply to the thyroid gland, and resulting in the development of goiter and other associated problems. Recently, however, other members of this large group of compounds (e.g., glucoraphanin, sulforaphane, phenethyl isothiocyanate, benzyl isothiocyanate, crambene, and indole-3-carbinol) have attracted intense research interest because of their cancer chemoprotective and antioxidant attributes. Certain compounds (e.g., sulforaphane) have been shown to be potent inducers of mammalian detoxication enzymes, which facilitate the deactivation and excretion of many carcinogens from the body. The use of a number of these compounds in a dietary strategy for cancer prevention is now being investigated in clinical trials worldwide.

The brassica vegetables are generally very rich sources of the antioxidant and provitamin A carotenoids (e.g., lutein, zeaxanthin, and β-carotene). These compounds are long-chain, fat-soluble substituted hydrocarbons which are the light-gathering accessory pigments typically found in the leaves, stems, and inflorescences of most plants. Since these are typically the plant organs which are eaten, they can be regarded as typical 'dark green leafy vegetables' and good sources of such compounds (Table 2). Certain brassica vegetables, such as kale and broccoli, are particularly rich in these compounds, and the edible head of cauliflower is particularly devoid of them. A mutant, orange cauliflower plant was found growing in a Canadian field about 30 years ago, however, which is currently being investigated as a very potent source of β-carotene and a very useful system for scientists to unravel some of the biochemical and molecular mysteries still surrounding carotenoid production in

plants. Though an orange cauliflower is now being marketed, the application of this research may result ultimately in the introduction of higher-carotenoid varieties of some of the world's staple crops and a reduced incidence of vitamin A deficiency (which can be reduced by ingestion of β-carotene-rich foods).

The brassica vegetables, though not unique in this respect, are also good sources of various flavonoids and their conjugates to which significant antioxidant activity has been ascribed. For example, the prevention of lipid peroxidation for which these compounds appear to be reasonably well suited may be an important mechanism for reducing the severity of age-related degenerative diseases such as arthritis, cardiovascular disease, and cancer.

See also: **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; Synthetic Antioxidants, Characterization and Analysis; Role of Antioxidant Nutrients in Defense Systems; **Cancer**: Diet in Cancer Prevention; **Functional Foods**; **Glucosinolates**; **Goitrogens and Antithyroid Compounds**; **Salad Crops**: Dietary Importance; Leaftypes; **Vegetables of Temperate Climates**: Cabbage and Related Vegetables; Oriental Brassicas; **Vegetarian Diets**; **Vitamin K**: Properties and Determination; Physiology

Further Reading

Facciola S (1990) *Cornucopia: A Source Book of Edible Plants*. Vista, CA: Kampong.

Fahey JW and Stephenson KK (1999) Cancer chemoprotective effects of cruciferous vegetables. *HortScience* 34(7): 1159–1163.

Fahey JW, Zhang Y and Talalay P (1997) Broccoli sprouts: an exceptionally rich source of inducers of enzymes that protect against chemical carcinogens. *Proceedings of the National Academy of Sciences of the United States of America* 94: 10367–10372.

Gómez-Campo C (ed.) (1999) *Biology of Brassica Coenospecies*. Amsterdam: Elsevier.

Gray AR (1982) Taxonomy and evolution of broccoli. *Economic Botany* 36: 397–410.

Kays SJ and Dias JCS (1996) *Cultivated Vegetables of the World*. Athens, GA: Exon Press.

Lucier G (1999) Broccoli: Super food for all seasons. *Agricultural Outlook* April: 8–12. Washington, DC: Economic Research Service, USDA.

Rupp R (1987) *Blue Corn and Square Tomatoes*. Pownal, UT Storey Communications/Garden Way Publishing.

Seelig RA (1971) *Fruit & Vegetable Facts and Pointers*. Alexandria, UA: United Fresh Fruit and Vegetable Association.

Switzer S (1727) *The Practical Kitchen Gardiner*. London: Thomas Woodward.

Toussaint-Samat M (1987) *A History of Food* (trans., 1992). Cambridge, MA: Blackwell.

Tsunoda S, Hinata K and Gomez-Campo C (1980) *Brassica Crops and Wild Allies: Biology and Breeding*. Tokyo: Japan Scientific Societies Press.

Vaughan JG, MacLeod AJ and Jones BMG (eds.) (1976) *The Biology and Chemistry of the Cruciferae*. London: Academic Press.

Zhang Y, Talalay P, Cho C-G and Posner GH (1992) A major inducer of articarcinogenic protective enzymes from broccoli: Isolation and elucidation of structure. *Proceedings of the National Academy of Sciences of the United States of America* 89: 2399–2403.

BRAZIL NUTS

G T Prance, University of Reading, Reading, UK

Background

Brazil nuts are produced in the crown of one of the tallest trees of the Amazonian rainforest. An important dietary item of the indigenous Indian and local populations, the Brazil nut later entered into commerce and is a major export of the Amazonian countries, Bolivia, Brazil, and Peru. The nuts (**Figure 1**) are harvested almost exclusively from wild trees in the forest, rather than from plantations. They are high in protein and oils and are also a useful source of thiamin (vitamin B_1). (*See* **Thiamin**: Properties and Determination.)

Description and Taxonomy

The Brazil nut comes from the species *Bertholletia excelsa* Humb. & Bonpl., the only species in this genus of the Brazil nut family, Lecythidaceae. Several other species of the family have edible seeds, the best known being the sapucaia nut or monkey pot (*Lecythis pisonis* Cambess.). However, the sapucaia fruits open at maturity on the trees, and the seeds are removed by bats, rendering large-scale harvesting almost impossible. The Brazil nuts are produced in large, woody, round fruits that are slightly larger than

Figure 1 Brazil nuts. Reproduced from Brazil Nuts, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds.), 1993, Academic Press.

a baseball. These fruits, or pyxidia as they are termed, drop to the ground intact when they are mature, 14 months after flowering has taken place. Between 10 and 25 seeds, or nuts, as they are incorrectly called according to the botanical definition of a nut, are arranged rather like the segments of an orange inside the pyxidium. This explains their characteristic flattened triangular shape. In nature, a large terrestrial rodent, the agouti (*Dasyprocta*), or squirrels open the pyxidium and bury the nuts in caches some distance from the parent tree; this is the means of seed dispersal for this species. Brazil nut gatherers go to the trees in January to March, when the fruits are falling and before the rodents have removed all the nuts. The fruits are generally gathered in the morning, because there is less danger of injury from fruit fall. The fruits weigh 0.5–2.5 kg and fall from a height of 40–50 m, and injury and deaths have been reported of collectors struck by falling fruits. The gatherers open the pyxidia skillfully with machetes (see **Figure 2**) and pour out the nuts into their baskets. They trade their

Figure 2 Brazil nut gatherer opening nuts. Reproduced from Brazil Nuts, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds.), 1993, Academic Press.

harvest for goods, and the nuts are shipped to factories for processing. Before exportation, they are pasteurized with steam to kill toxic fungi, and in many cases, they are also shelled and canned.

Distribution and Ecology

Brazil nuts grow in nonflooded rainforests of the Guianas and Amazonian Colombia, Venezuela, Peru, Bolivia, and Brazil. They tend to grow in natural clusters or stands of 50–100 trees. These stands are known as *manchales* in Peru and *castanhais* in Brazil. Usually, there are one or less trees per hectare, but stands of 15–20 trees per hectare have been reported in Madre de Dios, Peru. Some of the clustering may be the result of planting by Indians who formerly occupied the area. Brazil nuts grow in forests with an annual rainfall of 1400–2800 mm and do better where there is at least a short, dry season. The trees are light-demanding, i.e., they need gaps in the forest before they can grow to adult size.

Flowering and Pollination

Flowering occurs in the dry season, which begins in September and can extend until February. The peak of flowering is September and October throughout most of central and eastern Amazonia, and slightly earlier to the west. After flowering and pollination, the fruits take approximately 14 months to mature and fall off in January and February.

The flowers have a complicated androecium structure, which forms a closed hood over the top of the flower. Nectar is produced inside the hood. Large bees are the only insects that have the strength to force open the flower to reach the nectar. As the bees force open the springed hood, their backs are forced against the pollen-bearing stamens and the stigmatic surface where pollen is deposited. The bees therefore carry the pollen from flower to flower. Bees that have been found visiting Brazil nut flowers are from the genera *Bombus*, *Centris*, *Epicharis*, and *Eulaema*. The Brazil nut is largely self-incompatible, i.e., the tree requires pollen from another tree in order for seed set to take place. Low seed set has been found in trees that are isolated in forest clearings and in areas of disturbance. The maintenance of the pollination system is essential for the production of Brazil nuts. Therefore, plantations that are surrounded by forest where the pollinators thrive are much more likely to be productive than plantations that cover larger open areas away from the forest.

Chemistry

The Brazil nut is rich in oil and protein (see Table 1). The oil extracted from the seed is bright yellow, almost odorless, and with a pleasant nutty taste. It has a specific gravity of 0.9165 at 15 °C, solidifies at −4 °C and does not become rancid easily. The first extraction of oil from the seeds yields an excellent cooking oil, and the second extraction produces an oil suitable for making soap and burning in lamps. Hence, it furnished many of the needs of the Indian population of Amazonia. The residue left after pressing the seeds can be used as an animal feed. The average protein content of defatted Brazil nut flour is about 46% and contains all the essential amino acids, making both the nut and the flour highly nutritious. The principal amino acid content of the nuts is given in Table 2, which shows them to be rich in leucine. The nuts also contain vitamin A, vitamin B_1 (thiamin), vitamin B_2 (riboflavin), and niacin, being a particularly good source of thiamin and niacin.

Brazil nuts contain selenium, a powerful antioxidant. The amount in a Brazil nut (16–30 $\mu g\,g^{-1}$) exceeds the US recommended daily dose. Some people consume a Brazil nut a day to achieve the recommended dose. Brazil nuts can cause hair loss if consumed in large quantities owing to their selenium content. Brazil nuts accumulate selenium because of the similarity of this element to sulfur. There is often a sulfur deficiency in Amazonian soil, especially after long-term harvesting of the nuts. If the soil is selenium-rich, this is used by the plant as a substitute for sulfur. When sulfur fertilization is used, the selenium content of the nuts is reduced. Brazil nuts are rich in the sulfur amino acids methionine and cysteine. (*See* **Amino Acids**: Properties and Occurrence; **Niacin**: Properties and Determination; **Retinol**: Properties and Determination; **Riboflavin**: Properties and Determination; **Thiamin**: Physiology.)

Table 2 Principal amino acids found in the Brazil nut[a]

	Amino acid content (mg per kilogram fresh weight)
Lysine	4313
Threonine	3263
Valine	4888
Methionine	4088
Leucine	9250
Isoleucine	3450
Phenylalanine	4385
Tryptophan	1400

[a]From SUDAM (Superintendência do Desenvolvimento da Amazonia) (1976) *Estudos e pesquisas sobre a castanha-do-Par²*. Belém, Brazil: Ministerio do Interior, Departamento de Recursos, Naturais.

Table 1 Chemical properties of Brazil nuts (as a percentage of the total weight)

	Seeds from Peru[a]	*Seeds from Brazil*[a]	*Seeds from Brazil*[b]	*Seeds from Brazil*[c]
Oil	65.0	67.0	70.0	67.45
Protein	17.0	17.0	13.9	15.48
Ash	3.0	4.0		3.89
Crude fiber	0.9			3.21
Water	4.0	5.0	2.0	5.49
Carbohydrates	10.1	7.0		3.83
Ash/fiber/carbohydrates			14.1	
	100.0	100.0	100.0	99.80

[a]From Sánchez JS (1973) Explotación y comercialización de la Castaña en Madre de Dios. In: *Informe No. 30*. Lima: Ministerio de Agricultura, Dirección General de Forestall y Caza.
[b]From SUDAM (Superintendência do Desenvolvimento da Amazonia) (1976) *Estudos e pesquisas sobre a castanha-do-Par²*. Belém, Brazil: Ministerio do Interior, Departamento de Recursos, Naturais.
[c]From Knuth (1939) Lecythidaceae. In: Engler A (ed.) *Pflanzenreich IV*, vol. 219a, pp. 1–146.

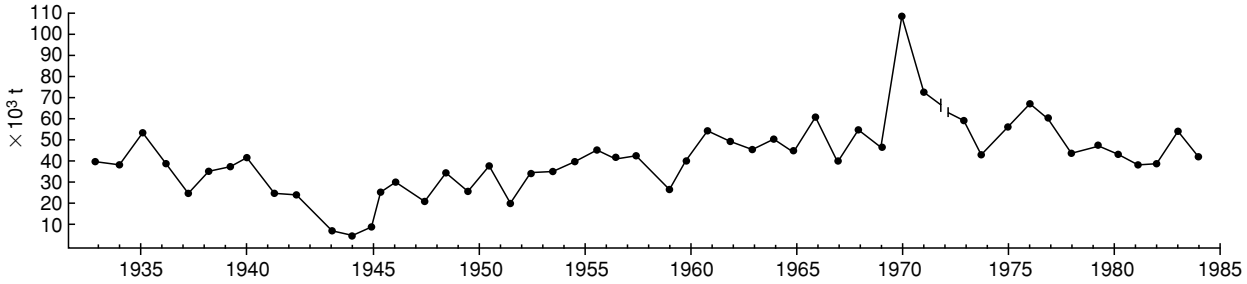

Figure 3 Brazil nut seed production from 1933 to 1984 (data from the *AnuÃrio EstatÚstica do Brasil*). Reproduced from Brazil Nuts, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds.), 1993, Academic Press.

Economic Botany

Until recently, Brazil nut seeds were second only to rubber as an export crop from Amazonian Brazil. The production from 1933 to 1986 has ranged from a low of 3557 tonnes in 1944 to 104 487 tonnes in 1970 (Figure 3). The production in 1986 was 36 136 tonnes. The low of 1944 was caused by World War II. In the Department of Madre de Dios in Peru, 20% of the forests are rich in Brazil nuts, and it is estimated that two-thirds of the population are engaged in Brazil nut extraction and processing. The annual production of nuts from the state of Pando in Bolivia is estimated to be between 10 and 12 000 tonnes. Most of the seeds are exported to France, the UK, the USA, and Germany. Very little on the crop is consumed domestically in either Brazil or Peru. Recent price fluctuations of Brazil nuts on the commodity market have made it harder for the producers. The nuts are exported intact or with the seed coat removed. In the latter case, the seeds are soaked for 8–10 h in a tank of water, after which they are submerged in boiling water for 1–2 min. The seeds are dried to make the embryo shrink away from the seed coat; this facilitates the removal of the seed without breakage. Broken seeds have a lower export value and are often used locally for oil extraction and in the soap industry. The dried seeds are placed in a manual press, and pressure is applied at both ends to crack open the seed coat (see Figure 4). The seeds are canned for export. Brazil nuts can be eaten raw or roasted, and are used in confectionery.

Brazil nut oil is often used in soaps, shampoo, and hair and skin conditioning products. It is a good moisturizer for the skin and provides antioxidant benefits. Brazil nuts require careful handling and steam sterilization, because they are highly susceptible to bruising and molding, which can cause the accumulation of aflatoxins. Sometimes, whole batches of nuts exported in their shells have to be rejected because of the danger of aflatoxins. This is rare in properly prepared and packaged nuts.

Figure 4 Cracking Brazil nuts in a processing factory, Óbidos, Brazil. Reproduced from Brazil Nuts, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

In addition to producing the nut, Brazil nut trees yield excellent timber. Although it is against the law to fell Brazil nut trees in Brazil, there is still a clandestine market for its timber. The timber is also used in Peru. The fibrous bark of the tree is used for caulking boats, and in folk medicine, a bark tea is used as a cure for liver ailments. The empty fruit cases are used for fuel and for making a vast array of tourist souvenirs. The world-wide Brazil nut market in 1999 was estimated to be $35 million.

See also: **Amino Acids**: Properties and Occurrence; **Niacin**: Properties and Determination; **Retinol**: Properties and Determination; **Riboflavin**: Properties and Determination; **Thiamin**: Properties and Determination; Physiology

Further Reading

Chang JC *et al.* (1995) Selenium content of Brazil nuts from two geographic locations in Brazil. *Chemosphere*.

Knuth R (1939) Lecythidaceae. In: Engler A (ed.) *Pflanzenreich IV* 219a: 1–146.
Mori SA and Prance GT (1990a) Taxonomy, ecology, and economic botany of the Brazil nut (*Bertholletia excelsa* Humb & Bonpl.: Lecythidaceae). *Advances in Economic Botany* 8: 130–150.
Mori SA and Prance GT (1990b) Lecythidaceae – Part II. *Flora Neotropica Monograph* 21: 1–375.
Moritz A (1984) Estudos biológicos da castanha-do-Brasil (*Bertholletia excelsa* H.B.K.). EMBRAPA, Centro de Pesquisa Agropecuário do Trópico Umido *Documentos* 29: 1–82.
Müller CH (1981) Castanha-do-Brasil; estudos agronómicos. EMBRAPA, Centro de Pesquisa Agropecuário do Trópico Umido. *Miscelânea* 2: 1–25.
Nelson BW, Absy ML, Barbosa EM and Prance GT (1985) Observations on flower visitors to *Bertholletia excelsa* H.B.K. and *Couratari tenuicarpa* A.C.Sm. (Lecythidaceae). *Acta Amazônica* 15(supplement 2): 225–234.
Sánchez JS (1973) Explotación y comercialización de la castaña en Madre de Dios. In: *Informe No. 30*. Lima: Ministerio de Agricultura, Dirección General de Forestall y Caza.
Souza AH (1963) Castanha-do-Pará; estudo botánico, químico e tecnológico. Edições S.I.A. *Estudos Técnicos* 23: 1–69.
SUDAM (Superintendência do Desenvolvimento da Amazonia) (1976) *Estudos e pesquisas sobre a castanha-do-Pará*. Belém, Brazil: Ministerio do Interior, Departamento de Recursos, Naturais.
Sun SS *et al*. (1987) Properties, biosynthesis and processing of a sulfur-rich protein in Brazil nut (*Bertholletia excelsa* H.B.K.). *European Journal of Biochemistry* 162: 477–483.
Vaz Pereira IC and Lima Costa SL (1981) *Bibliographia de Castanha-do-Pará* (Bertholletia excelsa H.B.K.). Belém, Brazil: EMBRAPA, Centro de Pesquisa Agropecuário do Trópico Umido.

BREAD

Contents

Dough Mixing and Testing Operations

S D Deshpande, Central Institute of Agricultural Engineering, Bhopal, India

Dough Mixing

One of the most important operations in a bakery is the accurate and thorough mixing of ingredients. Mixing reduces nonuniformities of gredients in composition, properties, or temperature of bulk materials or products in and out of a reactor, and was defined by Quillen as the 'intermingling of two or more dissimilar portions of a material, resulting in the attainment of a desired level of uniformity, either physical or chemical, in the final product.' The term 'mixing' includes more than just the mechanical actions that blend ingredients to a homogeneous mass. There may also be an incorporation of air and the formation of minute bubbles in a fat/water phase, a development of hydrated flour proteins into an elastic mass known as gluten, and an increase of temperature by a build up of mechanical heat. There is no fundamental basis for classifying the gluten proteins on the basis of aqueous alcohol extractability, although such a classification does have technological significance. Gliadin and glutenin are known to impart entirely different physical properties to the gluten network in a wheat-flour dough. Gliadin behaves mainly as a viscous liquid, when hydrated, and confers extensibility, allowing the dough to rise during fermentation, whereas glutenin provides elasticity and strength, preventing the dough from being overextended and collapsing either during fermentation or in baking. This article discusses the purpose of kneading of dough, types of mixers for dough, dough mixing and dough development, and empirical methods in physical dough testing.

Purpose of Dough Mixing

The dough is kneaded before it is allowed to ferment, and again before it is molded into the desired shapes. The purpose of kneading is to remove some of the excess carbon dioxide in order to prevent overstretching of the gluten strands, and to distribute the yeast cells throughout the dough. Kneading also tends to keep the dough at a uniform temperature. The dough is kneaded sufficiently when it has a smooth, 'satiny' surface, and small bubbles appear under the surface.

Mixing of Dough – Dimensional Analysis

Dough mixing is the most common unit operation in the food industry, finding widespread application in the bakery, snack food, and breakfast cereal industries. Dimensional analysis of this operation reveals how potential scale-up problems can be diagnosed before any experimentation is initiated.

Consider the schematic representation of a dough mixer that is shown in Figure 1. Cereal chemists generally consider the specific energy input as being indicative of the degree of 'development' of the dough. Specific energy is defined as the total energy dissipation per kilogram of dough. Assuming that the dough is Newtonian, one can perform a dimensional analysis for the power consumed by a dough mixer. The key variables used are given in Table 1.

By inspection, the dimensionless relationship that is sought is

$$\frac{P}{\rho N^3 D^4 L} = f\left(\frac{ND^2\rho}{\mu}, \frac{d}{D}, \frac{F}{D^3}, \frac{h}{D}, \frac{t}{D}\right). \qquad (1)$$

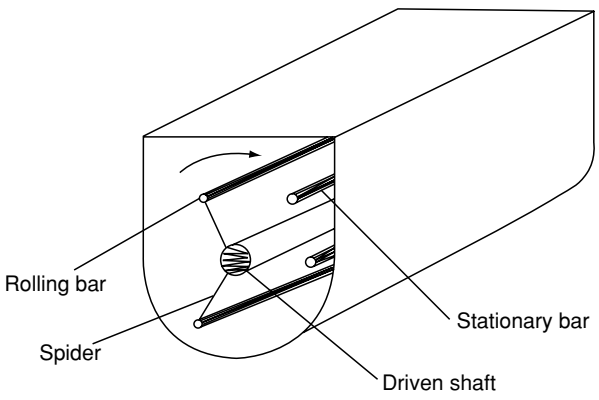

Figure 1 Schematic representation of a dough mixer. From Valentas KJ, Leuine L and Clark JP (1991). Dimensional analysis – mixing and sheeting of doughs. In: *Food Processing Operations and Scale-up*, pp. 359–361. New York: Marcel Dekker, with permission.

Table 1 Key variables used in the text

Variable	*Description*	*Units*
P	Power	ML/T^3
ρ	Density	M/L^3
μ	Viscosity	M / LT
N	Rotational speed	1 / T
F	Volumetric charge	L^3
d	Roller diameter	L
b	Breaker bar diameter	L
D	Mixer diameter	L
h	Clearance between rollers and wall	L
t	Clearance between breaker and rollers	L
L	Length of mixer	L

Solving for power and making use of the general rule that in laminar flow, the power is proportional to the velocity squared, it can be shown that

$$P\alpha\mu N^2 D^2 L g\left(\frac{d}{D}, \frac{F}{D^3}, \frac{h}{D}, \frac{t}{D}\right). \qquad (2)$$

The specific energy is simply the power dissipation multiplied by the mixing time divided by the mass of dough in the mixer. Then, the required time of mixing is given by

$$t = \frac{\omega}{P} \qquad (3)$$

where ω = specific energy input. If the scaled-up mixer is geometrically similar to the pilot mixer, this simplifies to

$$t\alpha\frac{1}{N^2}. \qquad (4)$$

A conclusion can be drawn from eqn (4). If the plant mixer operates at the same speed as the pilot plant mixer, the times required for developing the dough are identical. This provides the designer with a clear-cut direction as to how to design the scaled-up mixer. Unfortunately, for most actual applications, geometric similarity is not maintained when scaling up bakery mixers.

How does one handle a real dough with non-Newtonian properties? Complete analysis is impossible since the dough exhibits very complicated, time-dependent behavior. However, one can draw some useful inferences from the non-Newtonian analyses that follow.

If the dough were a simple power law fluid, the viscosity in the list of variables would be replaced by the flow consistency and the flow index. The result of the dimensional analysis would be

$$\frac{P}{\rho N^3 D^4 L} = f\left(\frac{N^{2-n}D^2\rho}{m}, \ldots, n\right). \qquad (5)$$

The time required to mix the dough is then

$$t\alpha\frac{1}{N^{1+n}}. \tag{6}$$

As before, if geometric similarity is maintained, the time required to mix the dough is held constant on scale-up if the speed of the mixer is held constant.

For a viscoelastic dough, which can be approximated as a Maxwell body, the variable list would have the relaxation time added. The new dimensionless group that appears to incorporate this new term is the Weissenberg (Deborah) number. The result is

$$\frac{P}{\rho N^3 D^4 L} = h\left(\frac{ND^2\rho}{\mu}, \theta_R N\right), \tag{7}$$

where θ_R is the relaxation time of the dough. The conclusions are unchanged. No matter how complicated a model, even a model containing an infinite number of relaxation times, the conclusion is the same.

Types of Mixers for Dough

Mixers and kneaders are necessary for the proper blending or mixing of the various ingredients. Liquids can be easily blended using propeller-type agitators in a tank. For blending of premixes or preparation of adjuncts, ribbon blenders, twin-cone, and V-shell blenders are usually employed when dry ingredients are to be mixed.

Dough mixing can be carried out using horizontal, vertical, reciprocating, or continuous mixers as required. The horizontal type may have a fixed or tilting bowl, while the vertical mixer may be a planetary or fixed spindle type. The continuous mixers may be of the agitator-in-tube type, or the rotor and stator head type.

Horizontal Dough Mixers

These mixers can be used to mix various mixtures ranging in consistency from thin batters to tough doughs. They must be used when gluten development is desirable, e.g., in breadmaking. These mixers have a horizontal U-shaped bowl mounted on a rigid frame with a drive motor. Their capacities range from 10 kg upwards. Bowls are usually made of stainless steel or stainless steel-clad mild steel. Bowls are sometimes provided with jackets for the circulation of cold or hot water.

High-speed mixers for gluten development usually contain a single axle on which two or more arms parallel to the front of the mixer are mounted. The limited clearance between the jacket wall and the agitator bars causes the dough to be repeatedly stretched and kneaded, resulting in quick gluten development. These mixers can usually be operated in two speed ranges, namely 30–60 and 40–80 r.p.m. In the slow-speed mixers, agitators of various forms, e.g., eight-arm, Z, etc., are employed on one or two axles.

The agitator configuration and its speed affect the action. Slow-speed Z or eight-arm mixers have a speed range of 14–60 r.p.m. (25 r.p.m. on an average).

The high-speed horizontal mixers are also called kneaders, and may contain bars projecting from the axle at right angles, while the bowl has radical bars intermeshing with axle bars. This construction gives very efficient dough kneading, and is particularly suitable for tough doughs.

The dough is discharged from these mixers by tilting the bowl so that the top is brought to a forward-facing position. In some models, the front of the bowl consists of a tightly fitting door, which can be raised and lowered independently of the immobile section of the bowl, and the dough is discharged through this door. The bowl-tilting mechanism may be hand-operated or it may be operated by a separate motor.

Vertical Mixers

Vertical mixers consist of a movable bowl or trough with one or two vertical shafts that may be stationary or may have planetary movement. The agitators themselves may be of various sizes and configurations. The most important design in the vertical mixer is the planetary mixer, which is particularly suitable for mixing batters or adjuncts such as icings. The agitator is called 'planetary' as it, apart from revolving around its own vertical axis at a relatively high speed, also moves in circles as it rotates around the bowl. This motion ensures 'beater' action throughout the bowl contents. The agitator may be provided with a variable-speed mechanism, and the bowl can be raised and lowered by a separate drive. The agitator may be of any of the various types available, e.g., simple, curved single-arm, wire meshing, or paddle gate.

The hook-type agitator gives a good kneading action to the dough, resulting in development of the gluten with a minimum of tearing. The wire-mesh type agitator gives a good 'beater' action with maximum air incorporation. The paddle-type agitator ensures good mixing coupled with scraping of the bowl sides. Bowls are usually provided with covers to prevent splashing or dusting. Capacities range from 10 kg upwards.

Spindle-type mixers are used mainly for biscuit doughs. Their special advantage is with saltine doughs, in which case they are adapted for mixing

in special mobile troughs used for fermentation. Thus, sponges and doughs need not be transferred in and out of the mixer between the various stages. The spindle speed is usually low, and spindles can be raised to allow the trough to be rolled into place under them. The mixer blades are so designed as to give a cutting action rather than a kneading or stretching action to the mix, thus a toughening effect on the cracker or biscuit doughs is avoided. (*See* **Biscuits, Cookies, and Crackers**: Methods of Manufacture.)

In reciprocating agitator or revolving bowl mixers, a pair of agitator arms travels through intersecting elliptical paths in a shallow, slowly revolving bowl. These mixers are useful for mixing temperature-sensitive doughs, and also where nuts and raisins in the mix are to be kept unbroken. Their output is usually low compared with horizontal mixers, and they are particularly suitable for pie doughs and puff pastry.

Dough Mixing and Dough Development

Dough Properties and Mixing Behavior

It is generally agreed that the uniqueness of wheat is due to its breadmaking quality. Among plant crops, only wheat flour and, to a limited extent, rye flour have the ability to form a dough that retains gases and produces a baked product, particularly leavened bread, with the desired eating qualities. Wheat-flour doughs exhibit a wide range of properties when different flour samples are compared. Dough properties influence both the efficiency of throughput in the manufacturing plant and the quality of the final baked product. In discussing the state of dough prior to baking, we need to differentiate between the contributions from the individual flour and those from the various treatments, which include added ingredients, mixing, and intermediate punching, and molding steps. Dough is a complex material from a rheological point of view. Knowledge about its structure comes from fundamental rheological studies, from standard physical dough testing, including mixing, and from microscopy.

Because of the trend towards greater use of baking processes involving intensive mixing and short fermentation times, much more attention has been given to basic studies of dough mixing in recent years. These studies have highlighted the sensitivity of dough structure to the conditions to which the dough has been subjected during its development. It is clear that meaningful interpretations of rheological measurements can be made only if the state of dough development is well understood.

Factors Affecting Dough Development

The first step in converting flour into bread is the mixing of flour, water, and other additives. Specific additives have been incorporated as improvers to wheat flour. The effects of addition of sodium alginate at 0.1–0.5% levels to wheat flour have been studied, and it has been observed that sodium alginate can improve the gas formation/retention power of dough, thus making it suitable for producing better bread. As mixing proceeds, the initially incoherent dough mass develops viscoelastic properties and finally acquires a sheen, this stage being termed 'clearing.' Prolonged mixing past this point causes the dough to lose its strength and become plastic and very sticky to the touch. If the torque on the pins, blades, or arm of the mixer is monitored against time, the trace rises, reaches a maximum, and then falls steadily. This is illustrated in Figure 2 for doughs mixed in a mixograph. Optimum bread-baking performance is usually achieved in a region at, or slightly past, the peak development point.

Critical mixing intensity and critical work input Two parameters are critical for optimum development and hence best performance. These are that

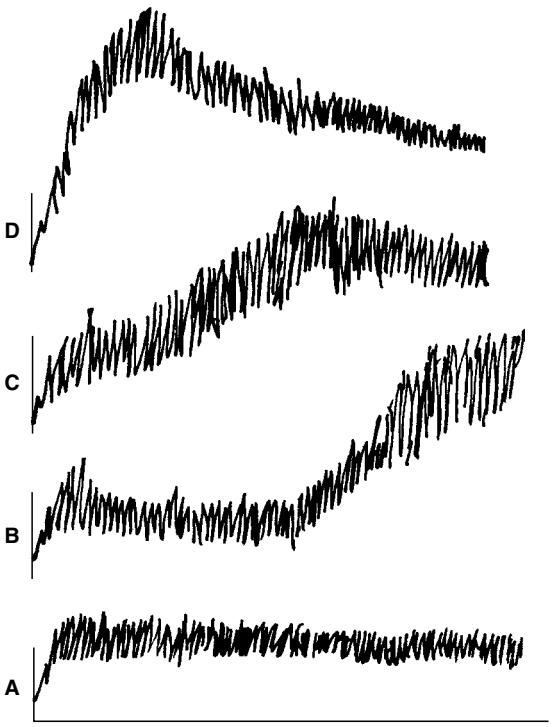

Figure 2 Mixogram traces. The sequence from A through D may be used to represent either doughs of increasing mixing requirements mixed at a fixed intensity, one dough mixed at increasing intensities or one dough mixed at a fixed intensity but containing increasing additions of cysteine.

(1) the mixing intensity must be above a minimum critical value, and (2) the total work imparted to the dough must be above a certain minimum value. Both parameters vary according to the flour used. Before discussing these requirements more fully, it is useful to consider the mechanical and mixing efficiencies of different dough mixers. The distinction between these two quantities is illustrated in Figure 3. A percentage of the total energy consumed by the mixer motor is available for performing mechanical work on the dough. This is called the mechanical efficiency. However, not all of this energy is utilized for mechanical development of the dough but that portion which is utilized is denoted the 'mixing efficiency' of the mixer.

Mixing requirements of flours vary widely, and Figure 2 illustrates the mixing behavior of four flour types. The mixing curves are for doughs that have been mixed at a constant and moderate speed. Flour A has high mixing requirements, and the rate of mixing used is insufficient to develop the dough; such a dough would be a failure in a baking test. Flour B is a moderately strong flour that does develop at this mixing speed, although there is an initial induction period before the dough begins to exert any appreciable torque on the mixer. Flour C is of medium strength, and flour D is a comparatively weak flour that develops quickly. We may also make use of Figure 2 to illustrate the effect of varying the mixing intensity (i.e., mixing speed) on the mixing curve of a given dough. At a low speed, the dough does not develop (trace A). As the mixing speed is increased, the dough develops after a lag phase (trace B) and thereafter develops more easily (traces C and D). With increasing speed, and providing the amount of work applied to the dough is above a certain level (corresponding to a point at, or slightly past, the peak in the trace), the baked loaf volume increases up to a critical mixing speed, above which it plateaus.

For every dough, there exists a critical mixing speed (which may vary for each mixer) below which dough handling properties and baked loaf characteristics are unsatisfactory. In addition, it is necessary to have imparted a minimum level of work input to the dough in order to bring it to an optimal developed state. If the mixing speed or work input level is below the critical values, unsatisfactory baking results are obtained. Baking results are less affected by exceeding the critical values than when these values are not reached. However, dough-handling properties deteriorate, and doughs become sticky; weaker flours are more sensitive to overmixing.

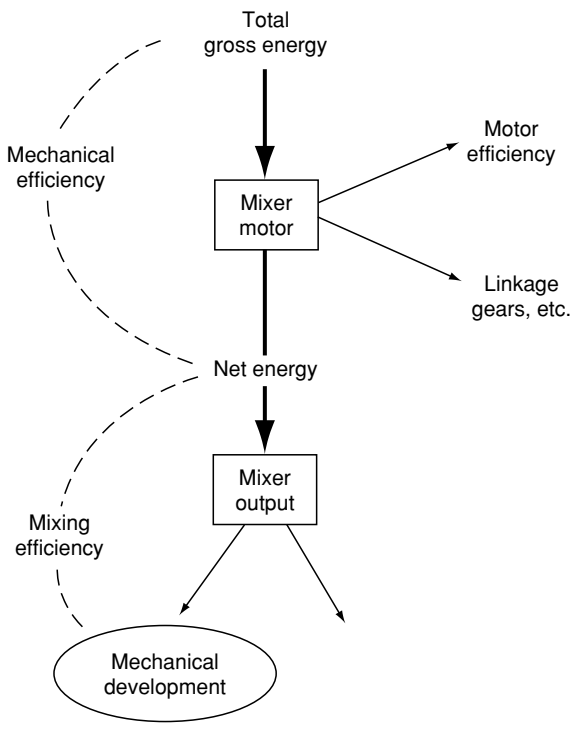

Figure 3 Diagram illustrating the distinction between mechanical and mixing efficiencies. From Kilborn RH and Tipples KH (1972) Factors affecting mechanical dough development. I. Effect of mixing intensity and work input. *Cereal Chemistry* 49: 34–47, with permission.

Activated dough development Because some flours have very long mixing requirements or have critical mixing speeds that exceed the capacity of many dough mixers, methods of overcoming these limitations have been devised. One method involves the addition of reducing agents, particularly sodium metabisulfite (SMS) and L-cysteine hydrochloride (cysteine), to doughs. Cysteine reduces the energy level required to achieve peak dough development, and also the critical mixing speed necessary to produce bread of satisfactory loaf volume. We may again make use of Figure 2 to illustrate the effect of increasing additions of cysteine in altering the mixing characteristics of a strong dough. As the amount of added cysteine is increased, the nature of the mixing curve changes from A through D. Cysteine also slightly increases the rate of energy input at a given mixing speed and increases the tolerance to undermixing, i.e., enables satisfactory bread to be produced with energy levels less than those required to achieve peak dough development. There is an optimum level of cysteine addition, and too high a level causes deterioration of loaf characteristics.

Intelligent Mixer for Bread Dough

To detect 'dough development' through digital signal processings of the mixer's torque and control mixing without an ordinary mathematical model, intelligent

mixer (speed variable farinograph, robots for flour–water addition) was developed.

A smoothed farinogram was obtained at constant mixing intensity, which has a linear part and a maximum point. Dough at the maximum point was judged to be fully developed by experienced mixing operators and was confirmed by the baking test. An empirical equation about torque mixing and fuzzy rules to correct nonlinear effects by three parameters (water content, temperature, and mixing speed) were obtained. Applying the equation gave a good statistical relation between 'torque increasing rate' and maximum torque value. This statistical relation led to regulation of the maximum torque value (a good indicator of dough properties) through flour (or water) addition by robots. Addition of flour or water in the early mixing stage caused no apparent influence on developed dough and baked products. The wave form of oscillating components, though there seemed a lot of background noise, had a good correlation with the degree of dough development. Combining wave-form analysis with step-wise change of mixing intensity, it was easy to find the optimum mixing intensity at each mixing stage. Applying these results, the intelligent mixer produced a dough with the desired maximum torque (a good indicator of dough properties) in a short time.

Empirical Methods in Physical Dough Testing

A great variety of commercial testing instruments are used in routine physical tests on dough, either in quality control or in research work. The deformation to which dough is subjected in these instruments is so complicated that it does not allow the evaluation of the properties of the material in simple physical terms. Though the results are evaluated more on an empirical basis than through theoretical analysis, they provide valuable information on the baking characteristics of the tested material and are a useful tool.

Over the years, a physical testing system has been developed on the basis of the 'three-phase concept of breadbaking.' The system has been generally accepted and is used in many commercial milling and baking operations. It applies three principles of testing: dough mixing, dough stretching (load–extension), and viscosity measurements on buffered flour suspensions at elevated temperatures.

Mixing Tests

The two most common instruments used for testing wheat flour and wheat-flour dough during the mixing operation are the Brabender farinograph and National Recording Dough Mixer (mixograph).

The Brabender farinograph is essentially a torque-measuring dough mixer that measures the plasticity and mobility of dough upon a relatively gentle mixing at constant temperature. The resistance of the dough to Z-shaped mixing blades is transmitted to a dynamometer connected to a mechanical recording system, which records a curve (farinogram) on a kymograph chart. Farinograms provide information on optimum mixing time and dough stability on prolonged mixing (Figure 4). Doughs are tested at a standard water content known as the 'farinograph water absorption' value. This value has to be determined by 'titration' of flour with water, and is the amount of water needed for a standard optimum consistency of the dough of 500 Brabender units. 'Forinograph water absorption,' dough development time, and dough stability are useful parameters for the evaluation of the strength of a flour. In general, the higher the value of these parameters, the stronger the flour. To express the strength of a flour on the basis of farinograph data as a single score, the 'valorimetric value' may be determined by the dough development time and the decreasing slope of the curve (degree of softening) – the higher the value, the stronger the flour.

The National Recording Dough Mixer is another widely used instrument for flour testing. Like the farinograph method, the mixograph method has become a standard physical dough test. The resistance offered by the dough to four vertical pins revolving around three stationary pins in a mixing bowl creates a force that deviates the mixing bowl from its original position. The torque is proportional to the shear strength and elasticity of the dough, and may be used as an index of the dough strength. Both Brabender farinographs and National Recording Dough Mixers may be modified by replacing the mechanical recording device by an electronic strain-gauge system. Electronic strain-gauge recording offers several advantages: ease of calibration in terms of physical units, greater accuracy because of the elimination of friction from the torque measuring system, a wide range of sensitivity, and a record on rectangular coordinates.

For testing the mixing characteristics of dough under conditions more closely related to those during mechanical development, a variant of the farinograph, known as the Brabender Do-Corder, was developed. This has a nearly closed mixing bowl with heavier mixing blades. The dough can be mixed at higher rates than in the standard farinograph. A more recent development is the Brabender resistograph. The instrument was developed to meet the needs of modern bread-baking technology in which high-speed mixers subject the dough to both kneading and stretching. The characteristics of doughs mixed

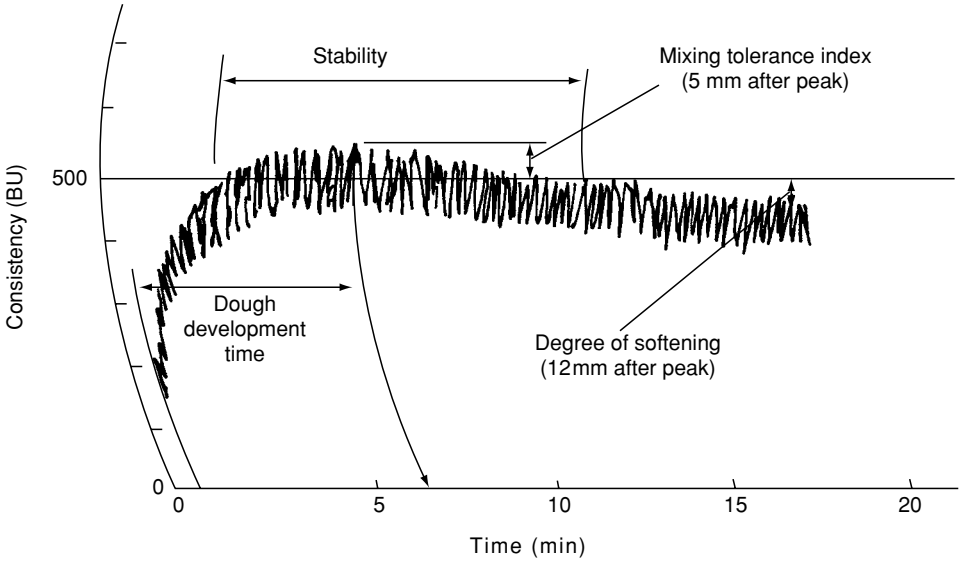

Figure 4 Representative farinogram showing the most commonly measured parameters. Brabender units (Bu) are a measure of optimum consistency of the dough. From Bloksma AH (1971) Rheology and chemistry of dough. In: *Wheat Chemistry and Technology*. St. Paul, MN: American Association of Cereal Chemists, with permission.

under these conditions differ from those produced by conventional mixers, where mixing is accomplished more by pressure than stretching. The resistograph has a mixing head that combines blending with stretching, pressing, and kneading. It imparts a high work input to the dough and can be used for the evaluation of dough response to high-speed mixing. Strong flours show a sharp increase in mixing resistance, and broadening and narrowing of the band. The resistograms of medium strength and weak flours are characterized by two pronounced maxima. The first is related to waterbonding and dough development and indicates the dough development time. The second measures the stickiness and extensibility at breakdown of the dough; the time to reach this point is most important in testing medium and weak flours. Both optimum and breaking points are reached when the blades of the mixer are completely covered with dough.

Load–Extension Tests

Several commercial instruments are available for the routine testing of dough at large elongations. From the recorded curves, various characteristics, such as resistance to deformation, extensibility, and energy needed to rupture the dough, can be computed. Among these instruments, the Brabender extensigraph is the most common. It was developed in about 1936 as a supplement to the farinograph, and has become particularly useful in the study of the effect of various chemical improvers on the rheological properties of dough. The extensigraph data (Figure 5) reported for control purposes are usually the curve length in centimeters (extensibility), the curve height in extensigraph units either at the maximum or 5 cm from the start of the curve (resistance), the area under the curve in square centimeters (strength value), and the quotient of height over length – the greater this quotient, the stronger the dough. Several attempts have been made to use the extensigraph for more fundamental studies, and to transform extensigraph data into rheological terms. The main problem in these transformations is the calculation of the actual cross-sectional area based on the 'effective mass' of the dough, i.e., the mass between the edges of the cradle supporting the dough. The effective mass progressively increases during stretching because some dough is pulled out from the cradle due to its resistance to extension. The effective mass and the actual cross-sectional area of the dough can be calculated from multiple regression equations. The coefficient in these equations, however, change with the type of dough and have to be determined experimentally.

A higher sensitivity and wider applicability of the Brabender extensigraph may be achieved by replacing the mechanical recording device by an electronic strain gauge system.

Another load–extension meter used in cereal laboratories is the research extensometer (Halton extensigraph) designed by Halton and associates. While the actual extensometer works on a similar principle to the Brabender extensigraph, the required water content for dough preparation is measured on

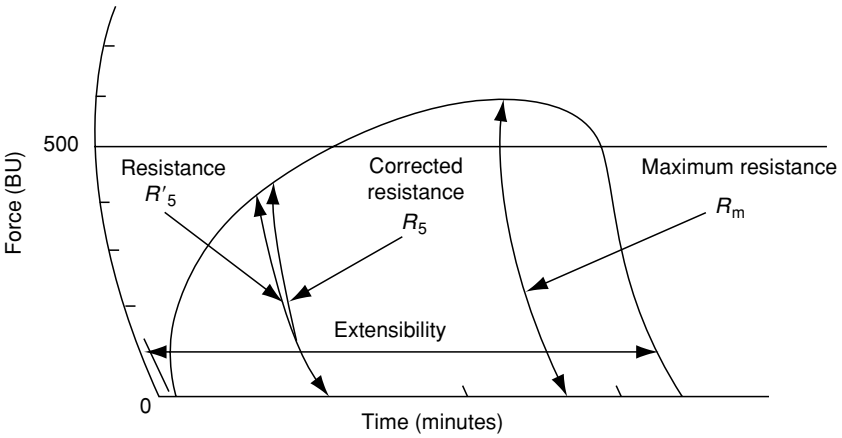

Figure 5 Representative extensigram showing the most commonly measured parameters. From Bloksma AH (1971) Rheology and chemistry of dough. In: *Wheat Chemistry and Technology*. St. Paul, MN: American Association of Cereal Chemists, with permission.

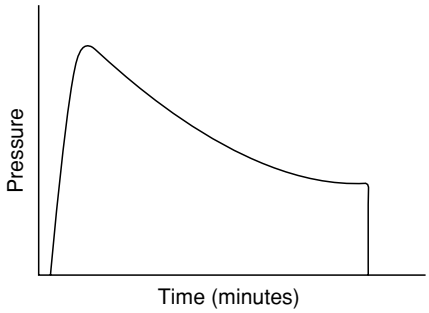

Figure 6 Representative alveogram. From Bloksma AH (1971) Rheology and Chemistry of dough. In: *Wheat Chemistry and Technology*. St. Paul, MN: American Association of Cereal Chemists, with permission.

the principle of extrusion by means of a special water-absorption meter, which, together with a mixer/shaper, belongs to the three-unit extensometer. The interpretation of the curves is the same as with Brabender extensigrams. A great advantage of this instrument lies in its applicability to fermenting doughs.

The Chopin alveograph is more popular in Europe than in North America. The instrument consists of three parts: mixer, bubble blower, and a recording manometer. The procedure involves using air pressure to blow a bubble from a disk of sheeted dough until it reaches its breaking point. A recording manometer records the presure of the air in the bubble as a function of time. The manometer records a curve (Figure 6) from which three basic measurements are taken: the height and the length of the curve, and the area under the curve. The area under the curve is proportional to the work of deformation.

Viscosity Measurements

The third step in the process of the physical testing of flour or dough, based on the 'three-phase concept of bread baking,' is the measurement of changes in flour-paste viscosities at temperatures above the gelatinization temperature of starch present in flour. The data obtained by these measurements are not only related to the gelatinization characteristics of the crumb in the baking oven, but also a good indicator of the diastatic activity of the flour. These measurements do not fall into the category of dough rheology, because all tests are done on flour slurries only. However, they help to evaluate factors that considerably affect the consistency of dough as well as the texture of bread crumb, its color, general appearance, and shelf-life. The most common instrument used for these tests is the Brabender amylograph. It is a torsion viscometer that provides a continuous record of changes in viscosity of the flour slurry at a uniform rate of temperature increase (decrease) of 1.5 °C min^{-1} under constant stirring. Although the instrument was originally developed for testing rye-flour pastes, the amylographic method has become a standard method for controlling α-amylase activity in wheat flour. The higher the activity, the lower the hot paste viscosity due to the liquefying effect of the enzyme. The effect of other amylolytic enzymes present is also reflected in the amylogram.

See also: **Biscuits, Cookies, and Crackers**: Methods of Manufacture

Further Reading

AACC (1969) *Approved Methods*, 8th edn. St. Paul, MN: American Association of Cereal Chemists.

Bloksma AH (1971) Rheology and chemistry of dough. In: *Wheat Chemistry and Technology*. St. Paul, MN: American Association of Cereal Chemists.

Chichester CO, Mrak EM and Schweigert BS (1984) Dough properties and mixing behaviour. In: *Advances in Food*

Research, vol. 29, pp. 230–234. London: Academic Press.
De Man JM, Voisey PW, Rasper VF and Stanley DW (1975) Empirical methods in physical dough testing. In: *Rheology and Texture in Food Quality*, pp. 310–323. Westport, CT: AVI.
Khatkar BS, Bell AE and Schofield JD (1995) The dynamic rheological properties of glutens and gluten subfractions from wheats of good and poor breadmaking quality. *Journal of Cereal Science* 22: 29–44.
Khatkar BS and Schofield JD (1997) Molecular and physico-chemical basis of breadmaking – properties of wheat gluten proteins: A critical appraisal. *Journal of Food Science and Technology* 34(2): 85–102.
Kilborn RH and Tipples KH (1972) Factors affecting mechanical dough development. I. Effect of mixing intensity and work input. *Cereal Chemistry* 49: 34–47.
Kilborn RHI and Tipples KH (1973) Factors affecting mechanical dough development. IV. Effect of cysteine. *Cereal Chemistry* 50: 70–86.
MacRitchie F (1992) Physico-chemical properties of wheat proteins in relation to functionality. *Advances in Food Nutrition Research* 36: 1–87.
Quillen CS (1954) Mixing – the universal operation. *Chemical Engineering* 61: 177–224.
Ravi R and Haridas Rao P (1995) Factors influencing the response of improvers to commercial Indian wheat flours. *Journal of Food Science and Technology* 32: 36–41.
Shewry PR, Halford NG and Tatham AS (1994) Analysis of wheat proteins that determine breadmaking quality. *Food Science and Technology* 8: 31–36.
Sidhu JS (1999) Effect of improvers and hydrocolloids on rheological and baking properties of wheat flour. Ph.D. thesis, GND University, Amritsar, India.
Sidhu JS and Bawa AS (1998) Effect of gum acacia on rheological properties of wheat flour. *Journal of Food Science and Technology* 35: 157–159.
Sidhu JS, Singh J and Bawa AS (1999) Effect of carboxy methyl cellulose on the rheological gas formation and gas retention properties of wheat flour. *Journal of Food Science and Technology* 36: 139–141.
Sidhu JS, Singh J and Bawa AS (2000) Effect of incorporation of sodium alginate on rheological, gas formation/retention and baking properties of wheat flour. *Journal of Food Science and Technology* 37: 79–82.
Stauffer CE (1990) Oxidants. In: Stauffer CE (ed.) *Functional Additives of Bakery Foods*, pp. 1–41. New York: Von Nostrand Reinhold.
Torikata Y and Ban N (1994) Intelligent mixer for bread dough. In: *Developments in Food Engineering. Part I*, pp. 266–267. Glasgow: Blackie Academic and Professional.
Valentas KJ, Levine L and Clark JP (1991) Dimensional analysis – mixing and sheeting of doughs. In: *Food Processing Operations and Scale-up*, pp. 359–361. New York: Marcel Dekker.
Voisey PW and Kilborn RH (1974) An electronic recording. Grain Research Laboratory mixer. *Cereal Chemistry* 51: 841–848.

Breadmaking Processes

M Collado-Fernández, University of Burgos, Burgos, Spain

Background

Bread is one of the most consumed food products known to humans, and for some people, it is the principal source of nutrition. Bread is an inexpensive source of energy: it contains carbohydrates, lipids, and proteins, and it is important as a source of essential vitamins of the B complex and of vitamin E, minerals and trace elements.

The history of bread can be traced back about six millennia. Breadmaking is an ancient art that is closely connected with the development of the human race and civilization, but the development of the baking oven, the industrial production of baker's yeast in the nineteenth century, was decisive for the technology of breadmaking. The twentieth century led to technical advances and the rationalization of bread production. Some of these advances in breadmaking include knowledge about physical–chemical changes in dough, the rheology of flour and dough, and the development of different instrumentations of rheology. Research into breadmaking is currently concerned with staling, the influence of the different additives on the breadmaking process, and the rheological properties of flour, dough, and bread, in order to improve its quality.

There are numerous variations of the breadmaking process. The choice of one or another is related to: tradition, cost, the kind of energy available, the kind of flour, the kind of bread required, and the time between baking and consumption of bread.

Bread

The basic formulation of bread is wheat flour, yeast, salt and potable water. Optional ingredients are nonwheat flours, shortening, nutritive carbohydrate sweeteners, skim-milk powder, enzyme active preparations, dough conditioners, and so on. The Food and Drug Administration legally defined the specified amounts of these optional ingredients in the production of various standardized bread products. (*See* **Legislation**: Additives.)

Bread is the food produced by baking a dough obtained by mixing wheat flour, salt, and potable water, leavened to specific microorganisms of bread fermentation, as *Saccharomyces cerevisiae*. Each country has many different kinds of bread, and these differ from one country to another (bread, rolls, pan

bread, pita bread, French bread, toast bread, baguette, etc.).

The breadmaking process changes the rheological and organoleptic properties of bread; this also changes the quality. In each country, the definition of quality may also change according to specific consumer preferences. However, the term 'bread' describes a range of products of different shapes, sizes, textures, crust, colors, softness, eating qualities, and flavors. For this reason, we will consider only the influence of the breadmaking process on the organoleptic properties of bread, and not if this bread is of a good or bad quality. The principal attributes of bread are as follows: loaf volume, crumb softness, grain uniformity, silkiness of texture, crust color, flavor and aroma, softness retention, and nutritive value. Although the breadmaking process influences the end product, the bread, the principal influence is the quality of different ingredients and especially the kind of flour used.

Principal ingredients

Flour

The breadmaking quality of flour depends on the variety of the grain, all agricultural and climatic conditions, including the harvest, and the milling process. Flour is the major ingredient in most formulations. The most important characteristics of flour are as follows: the protein content, especially the quantity and quality of gluten, the water absorption capacity, and the diastatic activity. The kneading of the flour and water gives the dough a cohesive, viscoelastic mass that retains the gas formed during fermentation. Thus, flour is responsible for the structure of bread. (*See* **Flour**: Analysis of Wheat Flours.)

Yeast

The principal function of yeast is the conversion of the simple sugars into carbon dioxide and ethanol. The production of gas causes the expansion of dough. Yeast also has an important role in the rheological properties of dough. (*See* **Bread**: Dough Fermentation.)

Salt

Salt is used in all baked goods to provide flavor, and because of its effect on the breadmaking process, salt has an influence on the rheological properties of dough. Salt inhibits the hydration of gluten; the gluten shrinks, the dough does not collapse, and gas retention is improved. Higher concentrations of salt inhibit enzymatic reactions and also inhibit the fermentation activity of yeast. In general, the proportion of salt used is 1–2% (based on flour weight).

Water

Only potable water may be used for the production of bread, and its quality is very important in breadmaking. Mineral constituents of the dough water (mainly carbonates and sulfates) give a firmer, more resistant gluten; the doughs do not collapse during fermentation, the gas retention is improved, and with a normal volume, the grain is finer and more elastic.

Optional Ingredients

Fat

The use of fats is essential to be able to keep the bread in storage. The functions of fats are as follows: fat increases the shelf-life, produces a finer grain, and, if used in small concentrations, yields a greater volume of baked foods (10%). The crust is more elastic and softer. The shortening effect is due to the formation of a film between the starch and protein layers of the flour. Surface-active materials, such as mono and diglycerides or lecithin, promote the formation of this film of fat and have a fat-sparing effect. The shortening effect is greater for fat with a lower melting point than for harder fats. The best fats are hydrogenated vegetable fats with a solidifying point between 30 and 40 °C. Fat produces a dough with more plasticity, so the use of fat requires less water in the formulation.

Sugar

Sugar promotes fermentation, browning of the crust, and a sweeter taste. In addition, it makes the dough more stable, more elastic, and shorter, and the baked goods more tender. For increasing additions of sugar and fat, the amount of added liquids must be reduced for a given dough consistency.

Milk and Dairy Products

These include milk (usually skim-milk powder) and whey containing lactose, which promotes browning, a softer crust, and a longer shelf-life.

Oxidants

The function of oxidants is the oxidation of -SH groups of protein to -SS-groups, resulting in an improvement in rheological properties of dough and of gas retention. The disulfide bonds thus established within and between protein chains lead to a firmer gluten structure. The time of dough maturation is shorter, the oven spring is greater, the volume is large, and the quality of the grain is better. The oxidant commonly used is ascorbic acid (E300). (*See* **Legislation**: Additives.)

Enzyme Active Preparations

Flours with a low enzyme activity produce breads that do not brown well, have a crumbly crumb, and stale rapidly. Such flours must have fermentable sugars or enzyme active preparations added (malted flour, malt extract, and bacterial or fungal α-amylases). The addition of malted flour increases bread volume slightly; the texture is better and decreases the keyholing (especially of pan breads). The addition of α-amylases is most important to start the hydrolysis of starch and to form the substrate for β-amylase action. The addition of fungal or bacterial α-amylases makes a difference to the thermal stability. The heat stability of bacterial α-amylases is higher. These amylases often remain active after baking and may cause some problems (moist and slim crumb) in the final product. (*See* **Enzymes**: Functions and Characteristics.)

Emulsifying Agents

Emulsifying agents or shortening or tensoactive agents are used to make bread softer during storage, especially pan bread. The effect of emulsifying agents on dough and baked products is based on their reaction with the starch–protein–fat–water system. The dough improves the strength of the gluten structure and improves the handling characteristics of dough as well as gas retention, and staling is delayed. Some of the emulsifiers in use are monoglycerides (E471), esters from monoglycerides and diacetyltartaric acid (DATA esters; E472e), sodium or calcium stearoyl-2-lactylate (SSL, E481 or E482), lecithin (E322), and other ingredients that perform a similar function. (*See* **Fats**: Uses in the Food Industry; **Legislation**: Additives.)

Preservatives

Preservatives are used in bread with a higher moisture content (40%) and packed hermetically because these breads are more sensitive to mold growth. The most important preservatives are calcium propionate (E282), sorbic acid (E200), and vinegar to prevent mold growth or rope. Calcium propionate also affects the fermentation process. More yeast can be added to compensate for this effect. (*See* **Legislation**: Additives.)

Breadmaking Process

The elaboration of bread may be divided into three principal stages:

1. Kneading of Dough (Dough formation): This stage marks the difference between all the breadmaking processes. In general, the kneading of dough involves the thorough mixing of ingredients. Some ingredients such as gluten and starch are hydrated; others are dissolved in water. The mechanical work produced by mixers contributes energy to the development of a gluten, incorporates air bubbles within the dough, and produces a dough with the rheological properties required, such as viscosity, elasticity, extensibility, gas retention, etc.
2. Fermentation, Proofing, or Proving: Fermentation is the action of the yeast on fermentable sugars to produce carbon dioxide (CO_2) and ethanol (which has a neutral taste). The gas is retained by dough and increases its volume. The production of gas and gas retention depends on the quality of the flour. The modifications of this stage are controlled fermentation and ultrafrozen dough.
3. Baking: This is the last stage, where, by the action of heat, the dough is converted into the final baked products by firming (stabilization of the structure) and by the formation of the characteristic aromatic substances. A modification of this stage is prebaked bread.

The different steps in breadmaking are as follows (Figure 1):

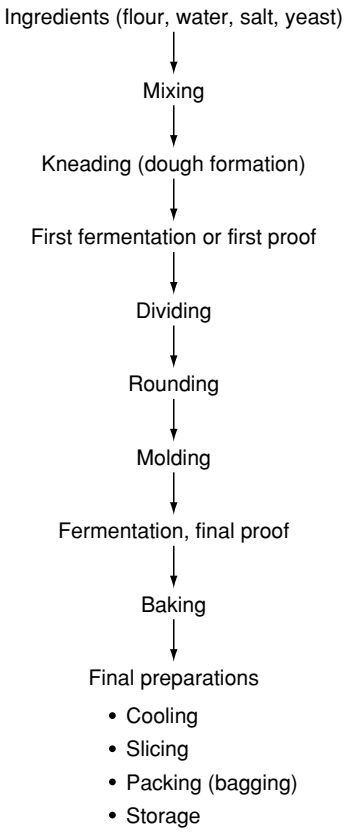

Figure 1 Different steps in the breadmaking process.

Table 1 Comparison of breadmaking processes

Breadmaking process	*Mixing*	*Kneading*	*Fermentation*	*Mixer*	*Dividing*	*Rounding*	*First proof*	*Molding*	*Second proof*	*Oven*
Straight Dough Method	Yes	20 min	Bulk 2–4 h	No	Yes	Yes	10–20 min floortime	Yes	70–80 min	Yes
Sponge and Dough Process	Sponge or predough	No	Sponge 4 h	Yes, and dough stage	Yes	Yes	20 min	Yes	60 min	Yes
Poolish	Preferment liquid	No	Preferment 1–2 h	Yes	Yes	Yes	20 min	Yes	60 min	Yes
System Do-Maker	Preferment without flour	No	Preferment 2.5 h	Mixer developer	No	No	No	Panning	60 min	Yes
Amflow System	Preferment with flour	No	Preferment 2 h	Mixer developer	No	No	No	Panning	60 min	Yes
Chorleywood Process	High-speed mixer	2 min 30 s	No	No	Yes	Yes	15 min	Yes	50 min	Yes
ADD	Yes	20–25 min	No	No	Yes	Yes	15 min	Yes	50 min	Yes

Source: Kulp K and Ponte JG (2000) *Handbook of Cereal Science and Technology*, 2nd edn. New York: Marcel Dekker; Quaglia G (1992) *Scienza e tecnologia della panificazione*. Pinerolo: Chiriotti Editori;
Rehm HJ and Reed G (1983) *Baked goods. In: Biotechnology: a Comprehensive Treatise, Vol. 5: Food and Feed Production with Microorganisms*, pp. 1–83. Weinheim: Verlag Chimie.

- *Ingredients*: Depending on the kind of bread required (flour, water, salt, yeast). (*See* **Flour**: Analysis of Wheat Flours.)
  - *Kneading*: This stage makes the greatest difference between all the breadmaking processes. (*See* **Bread**: Dough Mixing and Testing Operations.)
  - *First Fermentation or First Proof*: Dough has a resting period (floortime) in bulk after mixing and before dividing. In general, this stage tends to be substituted by intensive kneading or by other methods. Its function is to adapt the fermentation agent to the medium, production of dioxide carbon and other components, and the physical transformation of dough. The duration of this stage depends on the kind of process and equipment. The temperature conditions are 27 °C, relative humidity 75% (to avoid dried surfaces), and draught-free.
- *Dividing*: Dough may be divided by weighting or volumetrically (which is more common). At this stage, dough must be fluid, and have plegability and extensibility. After dividing, the dough is degassed and has an irregular form, the surface of dough has a higher viscosity, the structure of gluten has been damaged and the surface does not retain gas, and the dough has less plegability. For this reason, dough needs to undergo the next operation, rounding.
- *Rounding*: Dough with a rotary motion produces a ball-shaped piece with smooth skin. The alveolar structure is redistributed and improves the retention of gas. Dough is less viscous. After rounding, dough needs a floortime (between 2 and 20 min) in order to increase the plegability. The temperature conditions are 27–29 °C. If the temperature increases, dough becomes stale, decreases the retention of gas, and becomes more sticky; if the temperature decreases, dough becomes colder, and the fermentation time is greater. The relative humidity is 75%; if the relative humidity increases, this also increases the stickiness of dough.
- *Molding*: Two successive steps are responsible for the final shape of dough: laminated and curled. The molded dough is placed either in tins or on a baking tray and kept in a proofing cabinet to continue fermentation (final proof). This step is necessary to work the fermented dough, divide the alveolus, and provide a uniform redistribution.
- *Fermentation, Final Proof, or Proving*: During fermentation, starch is converted into sugars by enzyme action. The sugars feed the yeast, and the breakdown products are carbon dioxide and ethanol. As carbon dioxide is produced, the dough expands and retains it, and it is important that the skin remains flexible. There is a relationship between produced and retained gas that depends on the quality of the gluten structure. The more retained gas, the more bread volume. The temperature used depends on the kind of bread and the breadmaking process (28–30 °C): the lower the temperature, the longer fermentation time (2–4 h or 1–1.30 h). The relative humidity is between 60 and 90%, depending on the formulations and fermentation temperature. There is a variation of *P/L* during fermentation; dough has less tenacity (*P*) and more energy (*w*). (*See* **Bread**: Chemistry of

Baking; Dough Mixing and Testing Operations; Dough Fermentation; **Sensory Evaluation**: Aroma.)

- *Baking*: Baking is the final operation. There are physical, chemical, and biological transformations in dough that allow a stable product to be obtained, with excellent organoleptic and nutritive characteristics, the bread. (*See* **Bread**: Dough Mixing and Testing Operations; **Sensory Evaluation**: Aroma.)
- *Final Preparations*: Different steps for retention of quality: cooling, slicing, and packing.

- *Cooling*: During cooling, the moisture of the crumb goes to crust. It is necessary to reduce the temperature gradually to maintain the bread quality, and to take into account the relative humidity and temperature. If the relative humidity is low, the bread will lose too much moisture and may have no uniform crust. If the relative humidity is high, the crust of the bread will not be crunchy. This step is very important to maintain the bread quality when bread needs to be sliced and packed.
- *Slicing*: The bread must be cold for slicing; if not, it will crumble.
- *Packing*: Packing maintains the product quality by reducing crumb drying, minimizes the risk of contamination, and gives an appealing and informative package to the consumer.

Description of Breadmaking Processes

The different breadmaking systems are related to the diverse methods for making the dough. Although there are numerous variations, the major processing methodologies or breadmaking process are as follows:

Straight Dough Method

Flour, water, yeast, salt, and all other ingredients are added at the same time by single-stage mixing into a dough of optimum physical properties (i.e., the softness and appearance required, as well as the elasticity). The amount of water required depends on the water adsorption capacity of the flour determined by the farinograph. Usually, the amount of water is about 55–60% of the flour weight. (*See* **Bread**: Dough Mixing and Testing Operations.)

The kneading time depends on the speed and type of mixers used as well as the flour characteristics. Strong flours need more kneading time than soft flours. The speed differs, and energy imparted to the dough has repercussions on the efficiency of mixing of ingredients. (*See* **Bread**: Dough Mixing and Testing Operations.)

The dough is removed from the mixer, which may have a temperature between 26 and 28 °C, and is kept in the proof cabinet to ferment for 2–4 h, depending on the concentration of the yeast used. Afterwards, the dough is divided, rounded, molded, fermented (final proof), and baked.

- Advantage: It requires less processing time, labor, power, and equipment. Fermentation losses are smaller than with other processes.
- Disadvantage: Relatively small variations in processing may lead to noticeable variations in the final quality of breads. It is less flexible than the sponge dough, requires limited fermentation time, and does not permit correction of overfermentation.

Straight dough breads have a blander flavor than sponge dough breads. They do not have a soft texture, and the bread volume is lower, but this depends on the quantity of additives used.

Sponge and Dough Process

This method is the most common manufacturing commercial process used in the USA. In this indirect process, the flour, water, and a portion of the yeast are first mixed into a sponge or predough (temperature 25 °C). When the predough has fully fermented, it is mixed with the remaining flour, water, and other ingredients to make the final dough (temperature 26–28 °C), which is then given a short fermentation. Afterwards, the dough is divided, rounded, molded, fermented (final proof), and baked. (*See* **Bread**: Sourdough Bread.)

This method permits the use of strong and soft (weak) flour, using the strong flour to make the sponge and the soft in the final dough. Usually, the amount of strong flour is about 50%.

For rich dough high in fat or sugar, the fermentation time is often only 20–30 min. In this case, all the yeast is used with one-fourth to one-third of the flour and water to form a warm, soft predough. This causes better swelling and acclimatizes the yeast to the conditions of the dough.

- Advantage: It permits greater variations in the operations of the process and improves the volume, texture, and shelf-life of the bread. In addition, the longer fermentation time improves the aroma.
- Disadvantage: It requires more fermentation time and labor than the straight dough method, and it is more difficult to control. Sometimes, the dough has more tenacity, and it is difficult to divide, laminate, and mold.

Liquid Fermentation Process (Poolish)

The principle of this process is essentially the same as the sponge and dough method, except that it uses a

liquid instead of solid sponge. Its origin is Poland, hence the name Poolish, or liquid sponge. This method has been used since the 1920s with some variation in Vienna (for the production of Viennese bread); and in France (for the production of French bread). It has been used in American bakeries since the 1950s. (*See* **Bread**: Sourdough Bread.)

These preferments contain flour, water, and yeast. Usually, the amount of water is 1 l per kilogram of flour, and the amount of yeast depends on the fermentation time (12–15 g of yeast to 3 h; 7–8 g of yeast to 6 h; and 5 g of yeast to 8 h). The temperature of fermentation is 23–31 °C. The fermented liquid is stirred slowly to prevent foaming. After the desired fermentation time (the liquid triples its volume), the liquid is cooled by a heat exchanger to 8–10 °C and kept in a refrigerated storage tank until used. It may be stored for a 24-h period without any loss of activity.

The second step is the addition to this preferment liquid of the remainder of the flour, water, and salt (1%) and the other ingredients in the formulation. Afterwards, the final dough is fermented for 1 h, and then the remaining operations in the breadmaking process continue: dividing, rounding, molding, fermentation (1 h), and baking.

- Advantage: The same as the sponge and dough process, and also the consistency of the liquid sponge permits its transfer by pumping and better control of the production process. It permits work with weak flours, increasing its energy and elasticity, a cause of better starch gelatinization, and it has a good machinability. It requires less pressed yeast, and a single preferment may be used for the production of a variety of breads and baked goods.
- Disadvantage: It requires more fermentation time and labor than the straight dough method.

Continuous Bread Process

The development of continuous mix processes started in 1926 with the observation that the enzymatic and oxidative changes in dough structure during fermentation may be replaced by intensive mixing of dough at a high rpm. This process, introduced in the USA in the 1950s, was represented by two systems: the 'Do-Maker' process, based on patents granted to John C. Baker, and the 'Amflow' process, introduced by the American Machine and Foundry Co. It was observed that high-speed mixing could not replace the development of CO_2 and the aroma substances during fermentation. High-speed mixing was used with a liquid preferment to insure a desirable aroma in the final bread.

Do-Maker system (preferment liquid without flour) All operations are automatic. The first phase consists of preparations of the preferment liquid, which is a solution with all ingredients except the flour and shortenings (sugar, yeast, and nutrients to yeast and water). The development of the dough is due to the high degree of mechanical work, and not to floortime. It is also necessary to add higher levels of oxidants. The fermentation requires 1–4 h (usually 2.5 h) at 30–32 °C, with constant slow stirring in a vertical tank. The liquid preferment is then cooled (4 °C) and can be stored for a long time. During the fermentation, the pH decreases, so liquid preferment buffer substances like soya flour and powdered milk are added.

The second phase involves mixing the liquid preferment with the flour, shortening, and oxidants for 30–60 s and then pumping it to the developer. Usually, the rate of addition of the ingredients is 18–19 kg min^{-1} for flour and 14 kg min^{-1} for the preferment. The dwell time in the developer is only about 60 s. It is then extruded into pans.

- Advantages: The bread has a very uniform and fine grain, with a very soft crumb and a thin crust. It requires less time than the other systems with bulk fermentation.
- Disadvantages: The taste of the bread is different from that of the conventional sponge and dough bread.

Amflow system (liquid preferment with flour) This system is similar to the Do-Maker, except that the preferment is formed with flour, and the tank is horizontal.

Obtaining the preferment is a double step. In the first step, water, yeast, yeast nutrient, and a portion of sugar are mixed and kept in a holding tank to ferment for 1 h, and the aroma is developed. In the second step, salt, skim-milk powder, additional sugar, and 12% of flour are added. This viscous liquid is pumped to another tank for a second fermentation for 1 h with slow stirring at 30–32 °C. Finally, the preferment (sponge liquid) is pumped into a horizontal tank and cooled. The sponge liquid is pumped into the premixer, where the remaining ingredients (flour, water, sugar, fat, and oxidant) are mixed. This mixture is then developed in the mixer-developer. It is necessary to decrease the pressure created in the dough in order to avoid excess air, and directly extrude into baking pan. The bread has a very fine grain.

No-time Doughs (Short-time Doughs)

Dough fermentation and dough maturing may be accelerated by the addition of reducing and oxidant

substances. This has led to the development of 'no time dough' in which dough maturation is obtained by mechanically and chemically decreasing the fermentation stage.

Chorleywood process This system is also known as 'mechanical dough development' and 'intensive kneading.' This process originated in the UK and is used extensively throughout the world. The characteristic of this process is mixing of the dough intensively in a batch, high-speed mixer (Tweedy kneader) for 3–5 min with a controlled energy input. The optimum work input is 11 Watt-hours per kilogram of dough ($40\,J\,g^{-1}$). However, the energy requirements differ for flours of varying quality (soft flour needs less energy than strong flour) and the type of bread. This energy is five to eight times the energy used in conventional kneading. The mixing is generally done under vacuum, so the dough must contain ascorbic acid (75 p.p.m.). In this process, it is necessary to add high-melting-point fat (0.7% of flour), more water (3.5%) to soften the dough, since it has to support excessive mechanical work, and more yeast (50–100%).

- Advantages: This process saves time and space. The bread may be ready in 1.30 h. The production of bread is increased through the addition of extra water, and wastage is reduced by absence of bulk fermentation. Less, but more skilled, labor is necessary. More conservation and less staling of bread and soft flour may be used.
- Disadvantages: There is a reduction in the crumb of bread, flavor, and aroma because of the shorter processing times. For this reason, preferments may be added.

Activated dough development (ADD) This system was developed after research carried out on the Chorleywood system. It was found that good bread could be made with classic kneading, with a low-speed kneader, and a similar time to kneading but with the addition of chemical substances to eliminate the first fermentation or bulk fermentation. In this system, it is necessary to use fat and greater quantities of yeast and water.

Chemicals must be added before kneading, and are as follows: L-cysteine hydrochloride (35 p.p.m. based on flour) with ascorbic acid (35 p.p.m.) and potassium bromate (25 p.p.m.). Some legislation does not permit these oxidants. (*See* **Legislation**: Additives.)

- Advantages: This system is similar to the Chorleywood system, with the additional advantage of not requiring a high-speed mixer. The cost of chemicals is similar to the cost of the supplementary energy of the high-speed mixer.
- Disadvantages: The bread ADD crumb flavor and aroma are worse than that of traditional bread (straight dough) or Chorleywood bread.

In relation to fermentation, recent work has sought to control fermentation, which permits the regulation of bread production, and ultrafrozen dough. Controlled fermentation is one of the new techniques used to apply cold to stop or to decrease the extent of fermentation. (*See* **Bread**: Dough Fermentation.)

In ultrafrozen dough, the fermentation is interrupted for some time to renew it afterwards. The prefermentation time must not be more than 15 min, so as to avoid any type of fermentation. After molding, the dough is ultrafrozen (−35 to −40 °C), packed to avoid dry surface during conservation, and stored at −20 or −18 °C, until it is defrosted to go on to fermentation and baking.

In relation to baking, recent research has investigated prebaked bread. Prebaked bread is bread for which baking has been interrupted and may be continued afterwards. The baking is divided into two steps. The dough is prepared by the traditional method and is placed into an oven for 10 or 15 min until the starch has coagulated and the bread has gained structure, and then the bread is taken out of

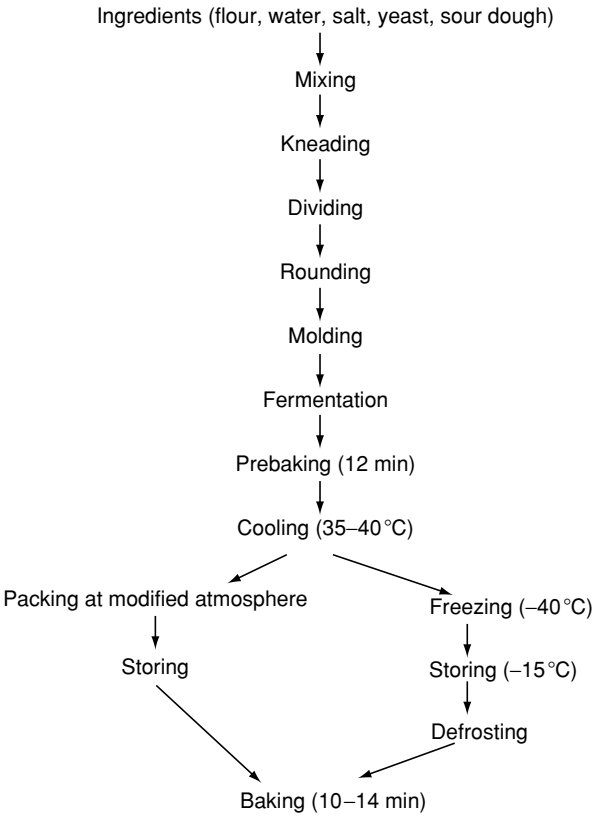

Figure 2 Different steps in the prebaked bread process.

the oven. After prebaking, the bread must be cooled down for an hour to a temperature of 35–40 °C, without any draughts so as to avoid cracking the crust, and packed in a modified atmosphere or frozen. This bread has a white color, contains more moisture, and is denser. The second phase is to finish the baking, and then the bread looks exactly like traditional bread.

- Advantages: Fresh bread can be obtained in a few minutes without having to ferment previously.
- Disadvantages: The bread has less volume, the crust is rougher and the crumb denser, and it stales more rapidly compared with the traditional system (straight dough).

See also: **Bread**: Dough Mixing and Testing Operations; Chemistry of Baking; Sourdough Bread; Dietary Importance; Dough Fermentation; **Enzymes**: Functions and Characteristics; **Fats**: Uses in the Food Industry; **Flour**: Analysis of Wheat Flours; **Legislation**: Additives; **Sensory Evaluation**: Aroma; **Yeasts**

Further Reading

Cauvain SP and Young LS (1998) *Technology of Breadmaking*, 1st edn. London: Blackie Academic & Professional.

Code of Federal Regulations (2000) *Requirements for Specific Standardized Bakery Products. Food and Drugs Title 21 Part 136.110*, pp. 347–349. Washington, DC: Office of Federal Register, National Archives and Records Administration. US Government Printing Office.

Council of European Communities 89/107/CE (1989) Relativa a la aproximación de las legislaciones de los estados miembros sobre aditivos alimentarios autorizados en los productos alimenticios destinados al consumo humano. *Official Journal of the European Communities* Doce-n ° 049/1989 serie L.

Hoseney RC (1994) *Principles of Cereal Science and Technology*, 2nd edn. St. Paul, MN: American Association of Cereal Chemists.

Kent N and Evers AD (1994) *Technology of Cereals: an Introduction for Students of Food Science and Agriculture*, 4th edn. Oxford: Pergamon.

Kulp K and Ponte JG (2000) *Handbook of Cereal Science and Technology*, 2nd edn. New York: Marcel Dekker.

Lorenz KJ and Kulp K (1991) *Handbook of Cereal Science and Technology*. New York: Marcel Dekker.

Pomeranz Y (1987) *Modern Cereal Science and Technology*. New York: VCH.

Pomeranz Y (1988) *Wheat: Chemistry and Technology*, 3rd edn. St. Paul, MN: American Association of Cereal Chemists.

Quaglia G (1992) *Scienza e tecnologia della panificazione*. Pinerolo: Chiriotti Editori.

Real Decreto 145/1997 (1997) Lista Positiva de Aditivos distintos de Colorantes y Edulcorantes para su uso en la elaboración de productos alimenticios, así como sus condiciones de utilización. In: *Boletin Oficial del Estado n° 70, 22 de marzo*, pp. 9378–9418. Madrid: Ministerio de la Presidencia.

Rehm HJ and Reed G (1983) *Baked goods*. In: *Biotechnology: a Comprehensive Treatise, Vol. 5: Food and Feed Production with Microorganisms*, pp. 1–83. Weinheim: Verlag Chimie.

Chemistry of Baking

M Collado-Fernández, University of Burgos, Burgos, Spain

Background

Baking is the last stage of the breadmaking process, in which there are several physical–chemical and biological changes as the result of the action of heat. These changes transform the dough into an edible final product with excellent organoleptic and nutritive characteristics – bread.

At this stage of baking, it is very important to control the baking temperature and time, which depend on the type of oven used, as well as the size of pieces and the kind of bread desired (the formulation used). All these possible variations may have repercussions on the different quality of bread depending on the bakery. The transfer of heat to dough may be by conduction, convection, and radiation. The different temperatures reached inside and outside the dough cause the formation of the crust and crumb of bread. The different phases during baking are as follows: oven spring (enzyme active zone), gelatinization of starch, and browning and aroma formation.

Baking

Working Conditions

During baking, the following highly correlated factors need to be controlled: the temperature of baking, the type of baking oven, the relative humidity, and the duration of baking.

Temperature of Baking

Usually, the baking temperature of 200–275 °C is a function of the type of baking oven, the duration of baking, and the size and type of bread desired (formulation used). Generally, the small items require a higher temperature and a shorter time, whereas large items need a lower temperature and a longer time.

This is also limited by the consistency of the dough; soft dough (low consistency) requires a higher temperature than hard dough (high consistency). In hard dough, where there is less water to evaporate, high temperatures produce cracks because dehydration is too rapid. If the temperature is too high, an overdone thick crust and, unbaked crumb result.

Depending on the type of bread, the temperature during baking will be uniform or will be higher at the beginning of the baking process with a slowly decreasing temperature afterwards.

Duration of Baking

The duration of baking depends on the temperature used. During baking, the following changes take place: gelatinization of the starch, denaturation of the protein, and the formation of aromatic substances. Thus, the duration of baking is limited, on the one hand, to a minimum in which all these changes are produced and, on the other hand, to a maximum limited by the loss of vitamins, excess evaporation of water, and also economic considerations.

Relative Humidity

The injection of water vapor into the oven favors the increase in relative humidity. The dough is covered with a fine surface layer of water, which keeps it moist and elastic for some time. This presents the following advantages: it avoids the immediate formation of crust, increases the volume (with a satisfactory oven spring), improves the color and shine, and produces a fine crust. Any excess water leads to the formation of drops of water on the surface of dough that give rise to bubbles on the bread.

Type of Oven

Heat is transferred by conduction, convection, and radiation. Coal, oil, or electricity may be used to heat the oven. There are several different types of oven, the most common being the hot air oven, in which the air is heated in a closed system and introduced into the oven by forced convection. Other types include the rotary hearth oven, the reel oven, the pan rack oven, and a continuous belt or tunnel oven.

Baking Process

The dough is placed into the oven when it reaches its optimum fermentation. If the fermentation is too short, the dough has not matured enough (underoxidized or green dough), and when it is placed into the oven, it does not support the increase (oven spring), resulting in a flat bread. If there is overfermentation (overoxidized dough), the crumb is very rough. To make bread with a crunchy crust, a cut is made on the dough before placing it into the oven. The cut should be oblique so that some of the wet zones are protected, thus achieving a greater volume of bread, because it can support the pressure better; if the cut is vertical, the bread opens and dries.

Transfer of heat Once placed into the oven, the dough increases its volume, reaching a maximum volume after some 5–10 min, depending on the weight, size, and quality of the dough. This increase in dough volume is due to the heat in the oven. Convection and radiation from the walls of the oven heat the dough, which comes into contact with the hot air. Dough on the baking pan is heated by conduction. The heat is transferred by conduction from the outside to the core of the dough, so the dough reaches different temperatures, leading to the formation of a crust on the surface ($> 100\,^{\circ}C$) and the crumb inside ($< 100\,^{\circ}C$) the bread.

Evaporation of water When the dough reaches $100\,^{\circ}C$, water inside evaporates. This process occurs superficially and water moves from the inside to the outside. When the migration of water is slow, the crust begins to form, the thickness of which depends on the duration of baking. At the beginning of baking, steam is added to help development and increase the bread volume. Later, the bread itself produces the steam, which condenses on the surface of the dough. The dough thus remains moist for longer, acquiring a satisfactory oven spring, and the surface is prevented from drying too rapidly. The resulting bread has a good external quality, color and shine, with a thick, crunchy crust.

Vaporization of aromatic substances During baking, together with the evaporation of water, there is also vaporization of all compounds that have a boiling point of less than $100\,^{\circ}C$, particularly ethanol, carbon dioxide, and some aromatic compounds formed during fermentation and baking (e.g., aldehydes, esters, and volatile acids). The vaporization depends not only on the concentration of these substances in dough but also on the capacity of gas retention in the dough and the permeability and elasticity of the network.

Cooling It is very important to cool the bread correctly after baking, otherwise the bread will crumble upon slicing. Cooling is also very important if the bread needs to be packed. The moisture of the crumb becomes crust during cooling. Depending on the relative humidity of the environment, the bread will lose or take up excessive moisture, and this will affect the quality of the bread.

Physical–Chemical Changes During Baking

Baking converts the dough into the final baked product by firming (stabilization of the structure) and by the formation of characteristic aroma substances. There are three phases depending on the temperature the dough reaches:

1. oven spring (enzyme active zone) (from 30 to 60 or 70 °C);
2. gelatinization of starch (55–60 °C) to no higher than 90 °C;
3. browning and aroma formation above 100 °C

Oven Spring

In this stage, the dough temperature reaches about 50 °C. There is very intensive fermentation because yeast activity continues until the dough reaches about 50 °C. Enzymatic activity is also very high because of the high temperatures, with a maximum at about 60 °C, and causes the conversion of starch to sugars. There is a liquefaction of dough and decrease in viscosity. The enzymatic activity continues until the dough reaches about 80 °C. There is an increase in the production of carbon dioxide from the conversion of sugar by the action of yeast. The solubility of carbon dioxide decreases with increasing temperature, and the production of carbon dioxide ceases when the yeast dies.

With the action of heat, the gas pressure increases. There is an inverse relationship between the size of the gas bubbles and gas pressure: the smaller the gas bubble, the higher the gas pressure. If the flour is of a poor quality, the gas bubbles cannot withstand the pressure and widen; then, the pressure decreases, and the resulting bread is flat (the bread deflates).

Enzymatic activity has an influence on the structure, crumb quality, and volume of the bread. If there is little enzymatic activity, the bread is small because the structure of the starch becomes rigid too soon, whereas too much activity leads to a structure that is too fluid, and the bread collapses completely.

Gelatinization of Starch

Dough undergoes increases in temperature from 50–60 °C to 90 °C. This promotes the gelatinization of starch and the denaturation of protein with the coagulation of gluten. The denaturation of protein begins at about 70 °C and is important for a stable bread structure. During kneading, the proteins are hydrated by the absorption of water, and during baking, water is transferred from gluten to starch, which swells. This dehydration of gluten continues until the increase in temperature leads to gluten coagulation. The denaturation of protein is more intensive on the crust than in the crumb of the bread because of the high temperature on the crust.

The capacity of starch to absorb water depends on the quantity of starch damaged during milling. First, starch absorbs the free water in the dough. It swells, and in the oven, the starch begins to gelatinize due to the high temperature. During baking, the gluten gives up water to starch, and the gelatinization is complete. The starch undergoes several changes such as loss of birefringence, increase in viscosity, and exudation of amylose (very important during staling). Dough undergoes an increase in viscosity by gelatinization of starch, and the dough becomes more consistent. The plasticity decreases, and the rigidity of the dough increases. When the gluten collapses, starch supports the structure of dough. During cooling, starch crystallizes, contributing to the firmness of bread. These changes in starch brought about by the action of heat may be measured using an amylograph, which gives an idea of the behavior of the flour in breadmaking and the retrogradation of starch, associated with staling. (*See* **Starch**: Functional Properties.)

Browning and Aroma Formation

The reaction of starch is different inside and outside the dough due to the different temperatures. The temperature outside the dough is more than 100 °C, there is a high degree of evaporation of water on the surface, and crust formation and coloration begin. The temperature inside the dough is less than 100 °C, which leads to crumb formation. The crumb in the center of the dough and the crumb near crust are different.

Color develops on the crust, as a result of the temperatures reached in the oven. Thus, the temperature and the time in the oven affect the color of the crust and the aroma of the bread. There are two different reactions: browning (or the Maillard reaction) and sugar caramelization. The formation of different aromatic compounds influences the color, aroma, and flavor of bread. Different dough ingredients, especially salt and sugar, also influence the flavor. The Maillard reaction is the reaction between reducing sugars (glucose, fructose) and the amino group of proteins to form dark brown melanoidins and other intermediate compounds such as maltol, isomaltol, and α-dicarboniles. Sugar caramelization is brought about by dehydration of sugar, as a result of the high temperatures, to form very reactive aldehydes (furfural aldehydes) and other dark compounds, which may form the polymers or react with the amino compounds, furan derivates, and isomaltol.

It is principally the quantity of the aromatic compounds formed during fermentation that determines the aroma of bread, but the compounds formed during baking as a result of browning reactions and formation of melanoidins and caramel polymers also have some influence. Some of the aroma components detected include 1-propanol, acetaldehyde, propanal, butanal, furfural, acetic acid, and ethyl acetate. The more volatile compounds are evaporated during baking, as mentioned before. Depending on the quality of the network, gas-retention capacity, and permeability, the aroma will be more or less retained by the bread. The different types of bread also have different aromas and intensities. (*See* **Browning**: Nonenzymatic; **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Development of Bread

The factors that influence the development of bread are as follows: the quantity and quality of gluten, the quantity of sugars (owing to their capacity to produce gas), and the quantity of yeast, salt, and other ingredients such as fat. Other technological factors include: the breadmaking process and duration of fermentation, temperature of baking, duration of baking, and characteristics of dough pieces.

During baking, there is an increase in dough volume owing to the following: (1) water in the aqueous phase in dough is vaporized by the action of heat; (2) carbon dioxide retained in the dough is dilated by the heat; and (3) the network becomes more elastic, owing to the change in thiol/disulfide bonds between adjacent protein chains, and favors a satisfactory oven spring.

Thus, the dough development is related to the following factors: the concentration of gas, the elasticity and resistance of the network, and the gas-retention capacity. If the dough has a high elasticity and a great capacity to retain gas, the bread will rise greatly, have a less specific weight, and have a uniform soft crumb, whereas, if the dough has little elasticity or little gas-retention capacity because of a short and rigid network, the bread will be small, with a nonhomogenous alveolus and agglomerate crumb.

Loss During Baking

During baking, dough loses weight mainly from evaporation of water (about 95%) and minimally from vaporization of ethanol, CO_2, volatile acids, and esters. These losses occur principally on the crust, which withstands higher temperatures than the crumb, which contains more moisture. A short baking time reduces the weight loss, but the resulting bread is of a poor quality because of insufficient browning and lower crumb elasticity. In addition, the amount of thermolabile vitamins, such as riboflavin (B_2) and thiamine (B_1), is smaller.

Defect of Baking

It is very important to control the temperature of the oven and the injection of steam. The main defects during baking are due to insufficient or excessive heat, and insufficient or excess steam. Insufficient heat affects the quality of the bread, giving an excess of volume, a rough crumb, thick, uncolored crust, and an excessive loss of weight. Excessive heat produces bread with little volume, a dark crust with bubbles, nonuniform crumb and an underbaked appearance. Insufficient steam leads to cavities between the crumb and crust. Excessive steam produces a crust that is too soft and has white bubbles.

In conclusion, the effects of baking are as follows: an increase in the volume of bread, the formation of crumb and crust, the decrease of thermolabile vitamins, and the development of flavor and color.

See also: **Bread**: Dough Mixing and Testing Operations; Breadmaking Processes; Dough Fermentation; **Browning**: Nonenzymatic; **Carbohydrates**: Metabolism of Sugars; **Enzymes**: Uses in Food Processing; **Flour**: Analysis of Wheat Flours; **Protein**: Chemistry; Interactions and Reactions Involved in Food Processing; **Starch**: Structure, Properties, and Determination; Functional Properties; **Wheat**: Grain Structure of Wheat and Wheat-based Products

Further Reading

Cauvain SP and Young LS (1998) *Technology of Breadmaking*, 1st edn. London: Blackie Academic & Professional.

Hoseney RC (1994) *Principles of Cereal Science and Technology*, 2nd edn. St. Paul, MN: American Association of Cereal Chemists.

Kent N and Evers AD (1994) *Technology of Cereals: An Introduction for Students of Food Science and Agriculture*, 4th edn. Oxford: Pergamon.

Kulp K and Ponte JG (2000) *Handbook of Cereal Science and Technology*, 2nd edn, revised and expanded. New York: Marcel Dekker.

Lorenz KJ and Kulp K (1991) *Handbook of Cereal Science and Technology*. New York: Marcel Dekker.

Oura E, Suomalainen H and Viskari R (1982) Breadmaking. In: Rose AH (ed.) *Economic Microbiology. Vol. 7: Fermented Foods*, pp. 87–146. London: Academic Press.

Pomeranz Y (1987) *Modern Cereal Science and Technology*. New York: VCH.

Pomeranz Y (1988) *Wheat: Chemistry and Technology*, 3rd edn. St. Paul, MN: American Association of Cereal Chemists.

Quaglia G (1992) *Scienza e tecnologia della panificazione*. Pinerolo: Chiriotti Editori.

Rehm HJ and Reed G (1983) Baked goods. In: *Biotechnology: A Comprehensive Treatise, Vol. 5: Food and Feed Production with Microorganisms*, pp. 1–83. Weinheim: Verlag Chimie.

Sourdough Bread

B J B Wood, University of Strathclyde, Glasgow, UK

Background

Sourdough breads, although their consumption is lower than that of yeasted breads in developed countries, probably predate the latter products. This review examines the history, microbiology, and nutritional significance of sourdough breads. Outlines of some typical production processes are also provided, but the reader is referred to the sources cited in the Further Reading section for detailed accounts of sourdough bread manufacture.

Sourdoughs are so called because the bread is leavened with a mixture of yeast(s) and lactic acid bacteria. It should be appreciated that a conventional yeast leaven will always contain some lactic acid bacteria, and that there will be a contribution to the final flavor that is attributable to them, but this will be minor and certainly will not result in a significant acid taste in the final bread. In a true sourdough, however, the contribution of the lactic acid bacteria will always be detectable, at least to the experienced palate, and quite often, the acidity is sufficient to be noticeable to the normal consumer. The sourdough way of making bread is of great antiquity, and many experts would agree that it is almost certainly the oldest way of leavening doughs, as a sour leaven will arise naturally from organisms already present in the flour if a mixture of flour and water is left in a warm place for a few hours; if this is used to inoculate a fresh flour/water mixture (dough), which is then allowed to ferment and a portion of it is used to inoculate a third portion of dough, and so forth, there will be a gradual evolution of a compatible mixture of yeast and *Lactobacillus* species. This type of continuing re-inoculation has much in common with maintaining liquid inocula by the process sometimes referred to as 'backslopping.' From the first mixing of flour and water, the developed sour ferment can be used to leaven a batch of bread dough, but as the sour is developed through succeeding generations, its quality and the consistency with which it leavens bread dough will improve, and the resulting bread will be more uniform and predictable in its quality, other things, such as flour quality, being equal. In some traditions, the sour dough starter is thus maintained for a very long time, possibly centuries, whereas in other practices, the sour starter is begun again every so often.

Historical Background

Many people, especially in the USA, think first, or only, of the San Francisco area when sourdough breads are mentioned. These bakers trace their roots back to the days of the pioneers, when sourdough bread, pancakes, etc. were an essential part of life in the covered wagon trains. The same heritage is found in Alaska, where the pioneers and especially the gold-miners were called 'sourdoughs,' as any reader of Jack London's stories and Robert Service's poems will know. Indeed, the latter called one of his poem collections *Songs of a Sourdough*. Dried yeast was not available at that time, so sourdough was the only way to have bread and other biologically leavened foods on the trail or in remote encampments.

This way of making bread, pancakes, and other foods would not have seemed particularly strange in those days as many of the settlers would know, or have kinship with, Europeans with traditions of making this type of bread.

This type of baking is still very widespread in Germany, Scandinavia, Eastern Europe, etc., but, perhaps surprisingly, French breads also quite often use this type of leaven. Even in Scotland, there was a traditional way of making bread using a liquid leaven called a 'Parisian barm,' which was still practised in some areas until quite well into the twentieth century, and one commerical bakery was still using a combination of Parisian barm and a final leavening with pressed yeast until a few years ago.

Allied Products

Many foods with a relationship to sourdough breads are produced in various parts of the world. Surplus barm is used in making soups, e.g., Kvass is a sharp, refreshing acidic drink found in Russia and much of Eastern Europe and is traditionally made by toasting or rusking sourdough rye bread, crumbling it into water, then allowing a fermentation to take place, after which the golden-brown, sharp, slightly sparkling, rather nutty-tasting liquid is poured off and drunk. The brew is kept going by replenishing the mixture with fresh toast and water at intervals, but if a brew is to be started *de novo*, fermentation is assured by inoculating with a portion of sourdough.

The Italian festive cake called 'panettone' is properly made with a sourdough starter especially maintained for this purpose by regular re-inoculation into a basic panettone dough, and there are reputed to be a number of similar products in various parts of the world. There is also a link with products such as the Indian food called 'Idli,' although it is a little tenuous, as Idli is made from a mixture of black gram (a variety of mung bean) and rice, and the leavening action is due to gas produced by the heterofermentative lactic acid bacterium *Leuconostoc*. Another variation is found in the use of presoaking to prepare certain ingredients for use in specialist breads. In a particular case, a Scottish bakery was using whole and kibbled (i.e., coarsely ground, crushed into large pieces or bruised) grains in breads such as the Vogel formula loaf and Turkestan-style bread. The whole and kibbled grains and coarse bran were mixed with water then left in the bakery overnight to soak up all the water that they could. The next morning, this mixture was incorporated with the other ingredients to make the final dough, which was leavened with pressed yeast in the usual manner. The bakery had not appreciated that the overnight soaking in the warm environment of the bakery was having any particular effect. They assumed that the only purpose of it was that which they had originally intended, namely to insure that the materials in question had imbibed all the water that they could. However, the bakery workers had noticed that the vessels containing the materials had a sour aroma and a marked frothiness when uncovered in the morning. In fact, there was a vigorous growth of lactic acid bacteria and some wild yeasts in the mixture overnight, and the acidity carried through to the finished bread, where it made an appreciable contribution to the flavor of the product.

Effects of Sour Fermentation on the Bread

Several effects are observed in addition to the obvious effect of an acid note to the flavor. The dough proteins are particularly affected, with changes in elasticity and extensibility that may result in the dough becoming in effect stronger, thus improving the baking performance of weak flours, and it is thus clear why sourdough formulations are so often associated with rye breads, as rye flour is much weaker than the typical wheat flour. Rye flour also differs from wheat flour in that it possesses higher levels of α-amylase, so that if a rye-based dough is fermented at neutral pH, excessive breakdown of the starch is observed. At the lower pH associated with the sour dough process, the amylolytic activity is reduced to an acceptable level.

Sourdough fermentations result in an increased hydrolysis of phytate from the grain; two factors are probably at work here, the fact that these are rather slow fermentations, so allowing more time for the seeds' phytases to effect hydrolysis, plus the lower pH, which is optimal for phytase activity. There is evidence for both phytase and phosphatase activity in lactic acid bacteria isolated from such ferments (especially those from the 'pre-soaks' described in the previous section). However, the significance of the results or the contribution that such activities (if confirmed by more detailed work) may make to the finished bread has not yet been established. The dietary significance of phytase activity is that phytic acid is a powerful chelating agent for a number of nutritionally important metal ions, including calcium, copper, iron, and zinc. Therefore, reducing the level of phytic acid will aid the absorption of these nutrients by the bread's consumer, which may be particularly important to those on vegetarian and 'macrobiotic' diets, although some authorities suggest that even the supposedly over-fed US consumer may not be consuming enough of these nutrients and may even be nutritionally compromised for them. The six phosphate residues that are present in each molecule of phytic acid and the inositol molecule are also made available to the consumer of the bread only upon hydrolysis of the phytic acid.

The acids formed in the fermentation are principally lactic and acetic (ethanoic) acids, and it is believed that their presence may contribute to the storage life of the bread by acting as mild antimicrobials, particularly effective against molds, which are often the first agents of biodegradation of bread. There is also evidence that the acids are damaging to the spores of *Bacillus* species, bacteria often present in the ingredients used for breads, particularly the wholegrain and so-called 'organic' types of materials favored by some makers of sourdough breads, because of their association with particular lifestyles. There is also good evidence that the acidity and possibly other factors associated with sourdough fermentations increase the time between baking and the onset of evident staling of the loaf.

Sourdough Bread Processes

From the foregoing, it is clear that there are many variations on the theme of sourdough bread production, but in order to keep this discussion within reasonable bounds, it is convenient to present two basic types, bearing in mind that other approaches are not necessarily just variants on the processes described here, as many of them have a quite clear identity, e.g., the Scottish 'Parisian barm' process mentioned

above. However, the majority of the sourdough bread produced commercially can be legitimately regarded as variants on one or other of the two main types.

Rye Breads (Particularly in Germany)

The work of the Detmold (Germany) group of Spicher and colleagues over many years has provided a deep understanding of the sourdough breads of Germany. They report that several microbial species participate in the fermentations. From some sources, it seems to be the practice to restart the German sour ferments fairly frequently, and this probably helps to bring about the diversity of species. However, Spicher notes that 'The development of commercial sourdough starter cultures over several decades has led to a certain natural selection of organisms that predominate in the fermentation.' This observation is in agreement with the work of other investigators who have studied sourdoughs from various parts of the world. Spicher also reports that attempts to replace these natural associations of organisms with pure cultures, as has been done with success in the production of commercial pressed yeast for the baking industry, has not been possible. An opposite claim will be discussed later.

Among the processes in German practice according to Spicher are multistage, two-stage, and single-stage sourdoughs. The last require the addition of yeast for good leavening, whereas the two-stage and multistage processes develop a yeast flora sufficient to leaven the dough fully. The single-stage processes are classified into the Detmold, Berlin, and Manheim (salt-sour) methods. In the Detmold process, the amount of starter and the temperature used for the fermentation are such that the process takes about 15–24 h. The temperature is in the range of 20–28 °C, with the amount of starter (as a proportion of the sourdough) ranging from 2% at 27–28 °C to 30% at 20–23 °C. The sourdough is made from a mixture of 55.6% rye flour and 44.4% water. When acidification is complete, this sourdough is blended with a mixture of wheat and rye flours and water plus the necessary yeast. Spicher quotes a typical mixture as 49.5 kg of sourdough, 42.5 kg of rye flour, 30.0 kg of wheat flour, and 46.0 l of water, giving a final proportion of 70% rye flour to 30% wheat flour. In the Berlin process, the combination of a high temperature and a high addition of seed sour, plus a softer mixture, gives adequate acidification in only 3–4 h. Spicher quotes a typical mix for the souring stage of the Berlin process as 8.0 kg of seed sour, 40.0 kg of rye flour, and 36.0 l of water, giving a proportion of 52.6% flour to 47.4% water. The sourdough is mixed with 30.0 kg of rye flour, 30.0 kg of wheat flour and 32.0 l of water, again giving a final ratio of 7:3 for rye to wheat flour. In the salt-sour process, the fermentation to produce the sourdough is longer than in the preceding processes, about 48 h being typical, although the presence of the salt will permit times up to 80 h to be used without loss of quality. The mixture for the souring stage contains 5.0 kg of seed sour, 25.0 kg of rye flour, and 25.0 l of water, plus 500 g salt, and is fermented at an initial temperature of 30–35 °C, dropping to 15–20 °C during 48 h of fermentation. This dough is mixed with 45.0 kg of rye flour, 30.0 kg of wheat flour, and 43.0 l of water to give the final 7:3 ratio of rye to wheat flour. Because of the inhibitory effects of the salt, it is necessary to add 2.5–3.5% w/w of compressed yeast to the final dough mixture, as compared with the 1.0–1.5% used with the preceding processes.

In some other processes described by Spicher, and referred to above as multistage and two-stage, the sour is increased in volume by successive inoculation of larger mixes of rye flour and water, each stage used to inoculate the next. The processes seem to be rather more favorable to the growth of desirable types of wild yeast alongside the lactic acid bacteria (LAB), with the result that either less compressed yeast is added than in the single-stage processes or no added yeast at all is needed in some of the multistage processes.

In these multispecies fermentations, the major LAB, according to a number of authors, are *Lactobacillus plantarum* (homofermentative), *L. brevis*, and *L. fermentum* (both heterofermentative). From the hexose sugars available for fermentation, homofermentative bacteria produce lactic acid only, or to such an overwhelming extent that other fermentation products are negligible in practice. Heterofermenters produce a mixture of products, including lactic and acetic acids, ethanol, and carbon dioxide, the ratios of these products being influenced by various environmental factors. Other products exist of course, and some of these play a part in the flavors that are characteristic of lactic-fermented products. In addition, LAB can utilize various pentose sugars, with the production of equimolar amounts of lactic and acetic acids. However, it is the homo- or heterofermentative aspect of the fermentation that is important here. According to previous reports, the best-quality bread results when the three species of bacteria are all present. The heterofermenters are essential for development of the 'typical flavor of sour rye bread.' However, the use of the heterofermenters alone gives a bread whose crumb lacks elasticity, but inclusion of the homofermentative *L. plantarum* restores this desirable attribute. The latter organism on its own gives the desired crumb characteristics, but the bread lacks aroma. Thus, both types

of organisms are essential for production of the finest quality of bread.

Spicher and coworkers have reported that the yeasts found in sourdoughs used in the production of rye breads fall into four types: *Candida krusei, Saccharomyces cerevisiae, S. exiguus*, and *Pichia saitoi*. Again, the environmental conditions influence the distribution of yeast species in doughs. Interactions between the yeasts and LAB are complex, and they can stimulate, inhibit, or have no obvious influence, one upon another.

A previous study of a sourdough started spontaneously on Swedish rye meal obtained a rather different result, with some similarities to the flora of the wheat sourdoughs (to be discussed below) in that the fermentation came to be dominated by only one species of yeast, although in association with three principal LAB isolates. The yeast was identified as *Saccharomyces delbrueckii*. The identities of the LAB were less certain, except that of *Pediococcus pentosaceus*; the other two dominant isolates were assigned as *Lactobacillus* sp. I and sp. II (probably *L. alimentarius*).

Wheat Sourdough Breads

The most important study of these products is that done on the San Francisco process by a group of workers from the US Department of Agriculture's Western Regional Research Laboratories, including Kline, Sugihara, Miller, and McCready. They reported that the fermentation is dominated by just two microorganisms, *Lactobacillus sanfrancisco* and the yeast *Saccharomyces exiguus* (sometimes referred to as *Torulopsis holmii*), now reclassified as the new species *Candida milleri*. Thus, both of the dominant organisms are types first isolated from this fermentation, although both have since been reported from other bread fermentations and also from other types of fermentation.

The most striking aspect of these studies is that the workers have been able to offer an explanation for the remarkable stability of the fermentation, characterized by a constant ratio of yeast to LAB cells of 1:100. The yeast, unlike the usual compressed baker's yeast (*S. cerevisiae*), cannot utilize maltose, whereas the LAB can utilize only this sugar, so they live in harmony because they do not compete for food. Indeed, they appear to have a symbiosis, as the LAB utilizes maltose by the phosphorolytic route, thus discarding one glucose molecule for every maltose used; the yeast assimilates the glucose, and there is evidence to suggest that the yeast liberates compounds that are stimulatory, or even essential, for the growth of the LAB, which is known to be exceptionally fastidious in its nutritional requirements. Growth of the LAB has been shown to be strongly stimulated by a small peptide, which is present in freshly prepared yeast extract, and also to be somewhat stimulated by various minerals and vitamins present in the extract. The yeast also has the unusual property that it is exceptionally resistant to the antibiotic cycloheximide, which is used in media for enumerating LAB because of its ability to completely inhibit the growth of most yeasts. In collaborative studies of Polish bakeries with scientists at the Politechnika Lodzka and in an examination of a Polish-Jewish bakery in Glasgow, the presence of yeasts and LAB with characteristics similar to those described above has been demonstrated, and Magdalena Wlodarczyk has independently discovered the biochemical basis of the important symbiosis between the yeast and the LAB.

In the San Francisco process, the starter is routinely rebuilt every 8 h, although it can be stored in the refrigerator for several days without any evident deterioration. According to Sugihara, the rebuilding is done by mixing 100 parts of developed sponge with 100 parts of high-gluten wheat flour and 46–52 parts of water (all by weight). During incubation, the pH of this mixture, initially 4.4–4.5, decreases to a final value of 3.8–3.9. To make bread, 20 parts of this starter sponge are mixed with 100 parts of what in the USA is called regular patent wheat flour, 60 parts of water and two parts salt. From an initial pH of 5.2–5.3, the value drops during 7 h of proofing to a final value of 3.9, when it is ready for baking.

Pure Cultures in Sourdough Bread-making

Sugihara reports a French study in which pure cultures of a yeast (*Candida tropicalis*) and a LAB (*L. plantarum*) were inoculated into a mixture of one part of wheat flour and 10 parts of water plus nutrients, with equal numbers of each organism, then incubated at 30 °C for 17 h, when the count of each organism had increased 100-fold. This was then used as an inoculum for breadmaking. Polish work has demonstrated that separately cultured *C. milleri* and *L. sanfrancisco* can be mixed in the bakery and used for the production of a starter, which is reported to give results superior in flavor, general quality, and consistent production characteristics to those obtained with their conventional starters; this technology is now being applied, together with improved designs of equipment for developing the sour, in bakeries in Lodz. These results are in marked contrast with the claims reported earlier in this review that pure cultures were not satisfactory for production of sourdough breads; this difference may relate to the type of bread that is being produced.

The Danish company, Chr. Hansen's Laboratorium A/S, is now offering a selection of freeze-dried starter cultures under the brand name 'Florapan.' Florapan L-22 is *Lactobacillus delbrueckii*, a homofermentative strain reported to give a 'mild and pleasant flavor' to the bread in which it is used. L-73 is also homofermentative (*L. plantarum*) but is reported to give the bread a 'piquant, spicy flavour.' Finally, L-62, the heterofermentative *L. brevis*, gives a 'penetrating but rounded taste and a long shelf-life of the bread.' The bacteria are particularly recommended for whole-grain flours because of their requirements for vitamins and minerals. The company advises that a freshly inoculated dough be allowed to sour overnight before use, but that where a fully developed dough is added to a fresh dough at a ratio of up to 1:20, 8 h of development can be sufficient. The developed dough is then used at about 30–40% of the final mixture, which includes baker's yeast. Their literature makes no comments on the crumb texture of the breads produced with the different bacteria, although it should be noted that the German workers attached significance to this aspect, concluding that for the best combination of texture and flavor, it is necessary to use a mixture of homo- and heterofermentative bacteria. As far as can be determined, no biochemical or other reason has been advanced for the claimed texture differences, so there would seem to be some scope for some useful research here.

A combination of technical advances in our understanding of, and capacity to control, sourdough fermentations has made these processes more reliable on a commercial production scale. At the same time, an increased public interest in the more traditional and 'ethnic' types of foods has created an increasing demand for these breads. The development of small bakeries of the type often referred to as 'craft' or 'artisan' bakers, has contributed to this process. Further improvements in processes, products, and marketing strategies are possible and can further increase the market penetration of these bread types.

See also: ***Bacillus***: Occurrence; **Lactic Acid Bacteria**; **Leavening Agents**; **Macrobiotic Diets**; **Organically Farmed Food**; **Phytic Acid**: Properties and Determination; **Rye**; **Starter Cultures**; **Vegetarian Diets**; **Yeasts**

Further Reading

Hammes WP and Gänzle MG (1998) Sourdough breads and related products. In: Wood BJB (ed.) *Microbiology of Fermented Foods*, 2nd edn. vol. 1, pp. 199–216. London: Blackie Academic & Professional.

Jenson I (1998) Bread and baker's yeast. In: Wood BJB (ed.) *Microbiology of Fermented Foods*, 2nd edn. vol. 1, pp. 172–198. London: Blackie Academic & Professional.

Kent NL (1983) *Technology of Cereals*, 3rd edn. Oxford: Pergamon Press.

Lonner C, Welander T, Molin N, Dostalek M and Blickstad E (1986) The microflora in a sourdough started spontaneously on typical Swedish rye meal. *Food Microbiology* 3: 3–12.

Oura E, Suomalainen H and Viskari R (1982) Breadmaking. In: Rose AH (ed.) *Economic Microbiology, Vol. 7. Fermented Foods*, pp. 88–147. London: Academic Press.

Rocha JM and Malcata FX (1999) On the microbiological profile of traditional Portuguese sourdough. *Journal of Food Protection* 62: 1416–1429.

Service R (1957) *Songs of a Sourdough*. London: Ernest Benn.

Spicher G and Pomeranz V (1985) Bread and other baked products. In: *Ullmann's Encyclopedia of Industrial Chemistry*, 5th edn, vol. A4, pp. 331–389. Weinheim: VCH Verlagsgesellshaft.

Steinkraus KH (1983) *Handbook of Indigenous Fermented Foods*, pp. 1131–1188. New York: Marcel Dekker.

Sugihara TF (1985) Microbiology of breadmaking. In: Wood BJB (ed.) *Microbiology of Fermented Foods*, vol. 1, pp. 249–262. London: Elsevier.

Wlodarczyk M (1985) Associated cultures of lactic acid bacteria and yeasts in the industrial production of bread. *Acta Alimentaria Polonica* 11: 345–359.

Wlodarczyk M (1986) Characteristic of yeasts from spontaneously fermenting bread starters. *Acta Alimentaria Polonica* 12: 103–108.

Wood BJB, Cardenas OS, Yong FM and McNulty DW (1975) Lactobacilli in the production of soy sauce, sour dough and Parisian barm. In: Carr JG, Cutting CV and Whiting GC (eds.) *Lactic Acid Bacteria in Beverages and Food*, pp. 325–335. London: Academic Press.

Dietary Importance

D A T Southgate, Norwich, Norfolk, UK

The History of Bread

Bread has played such a key role in the human diet that 'bread' is often taken as being synonymous with food as a whole, and in some dialects, it is used as a slang term for money, the means of obtaining food. Bread indeed has a very long history of use in the human diet.

The cultivation of cereals, especially wheat, appears to have started in the Middle East, and alongside the evidence for the cultivation of wheat, one finds evidence of loaves apparently produced with

yeast, and baked in ovens. Wheat is the primary cereal for producing bread, although in Northern Europe at the northern limit of cultivation for wheat, rye is widely used in breads, and wheat/rye mixtures are widely consumed. Earlier archeological evidence provides artefacts in the form of querns used for grinding cereal grains mainly barley, which was usually eaten as solid cakes.

Wheat as a cereal has the special compositional property that when ground grains are mixed with water and yeast, the fermentation of the starch produces a very attractive type of food. There are many references to leavened and unleavened breads in the Old Testament, and the Greeks and Romans both produced breads analogous to those consumed today. The cultivation of wheat provided an abundant source of food energy, together with a protein level and a protein with a balanced amino acid composition that made it a very nutritious food. In addition, the wheat grains could be stored from one harvest to the next and, if necessary, for longer periods to cover periods when crops failed. This led to wheaten bread becoming an important staple food. There is clear evidence of the importance of bread in the diet of both the Greeks and the Romans. The milling of wheat grains in querns and wind- or water-powered millstones exerts a shearing force on the grain separating a flake of the outer branny layers while progressively breaking the endosperm into smaller fragments and eventually flour. The bran and flour can be separated by sieving to produce a white flour, which, from earliest times, was used to produce a more prestigious food. Both the Greeks and the Romans argued about the merits of white versus wholemeal breads, a debate that continues to the present time.

In medieval times, wheaten bread was extremely important in the diet of the poorer classes, and a system of rigorous price control was developed to ensure that the peasants were kept supplied with ample bread. Severe penalties were also implemented to prevent the bread being adulterated with inferior foods such as pulses. In the UK, the price of bread has continued to be seen as extremely important by successive governments, as a means of keeping the people well fed and contented. To a certain extent, bread has become a 'token food' with regulations as to the size of loaves, the price of loaves, and permitted ingredients and additives. During both world wars, the extraction rate of bread flour was controlled at a high level to reduce shipping requirements, and in World War II, fortification with calcium and, later, iron and some vitamins was obligatory to maintain the nutritional value of the diet, with bread being used as a vehicle. Recently, proposals have been made to fortify bread with folic acid, to prevent neural-tube defects, and vitamin D, to ensure that Asian immigrants to the UK obtain this vitamin from their wholemeal chapatis.

Thus, since it first became part of the human diet, wheat and the breads made from it have provided an important staple food, and, despite the growth of affluence and the wider variety of foods available to most developed populations, it remains an important and nutritious food. Most current dietary advice regarding healthy eating advocates the consumption of a substantial amount of cereal foods in the diet, and this of course includes breads. The recognition of the role of dietary fiber in the diet has emphasized the importance of wholemeal and higher extraction breads, but most advice returns to the theme that all breads are good foods.

Nutritional Composition of Bread

Although much of the focus of this article will be on wheaten bread, some information on rye breads has also been included because these are also widely consumed, and in the UK, mixed grain breads have achieved some popularity. The focus on breads has been limited to those breads prepared by fermentation of the dough with yeast, although many communities consider many nonfermented rye products as breads.

Bread Production

Wheat and, to a lesser extent, rye and barley proteins contain gluten, which, as the dough is kneaded (or developed mechanically) forms a network within the dough that traps the carbon dioxide produced by fermentation of the carbohydrates in the flour by the yeast. This causes the dough to rise and leads to the development of the crumb texture when the dough is baked. The increased volume of the bread is an important factor in the sensory quality of bread and is possibly why breads have become established in the diet. Baking produces a crust that contains some starch granules that are not fully gelatinized and, in well-baked bread, can produce starch degradation products and even some carbon. These modified carbohydrates in the crust are important for the sensory characteristics of bread, but do not materially affect the nutritional value. Some salt is usually added to the dough to assist in the activation of the yeast, and other 'flour improvers' are added to facilitate the development of cross-linking in the gluten and so strengthen the network. Vitamin C is the only permitted improver in wholemeal bread in the UK, but the baking process effectively destroys any residual vitamin.

Composition of Wheaten Breads

The composition of some typical wheaten breads is given in Table 1 with a range taken from British, Italian, and Danish Food Composition tables to illustrate values from southern and northern Europe. It should be stressed that there is some variation in composition; the protein content, for example, depends very much on the selection of wheat mixture chosen by the bakers. The majority of home-grown wheats in the UK are soft, lower-protein wheats and need to be mixed with harder, high-protein wheats, typically from North America. Table 1 illustrates the difference that the inclusion of the bran makes, with the wholemeal breads typically containing two to three times as much dietary fiber (the values cited are for nonstarch polysaccharides). In most bread mixes, a small quantity of fat, usually as oil, is incorporated into the mix. In Denmark, there is a preference for softer crumbs, and the fat levels are higher. In the USA, milk powder may be included to produce similar softer crumbs. The levels of fat in Italian white bread are very low and probably represent an anomaly due to the choice of method for fat. Most of the carbohydrate is starch, with small amounts of free sugars and dextrins in the crust.

The water contents vary with the extent of staling, but the Italians seem to prefer a drier white bread. Table 2 lists the corresponding values for several inorganic constituents. The sodium contents are substantial, and bread is an important source of sodium in the diet. Recently, there have been moves within the bread producers to lower the levels as a public health measure. Salt plays an important role in bread flavor, in addition to activating the yeast, so reductions will have to be phased in.

The wholemeal breads have higher potassium, phosphorus, iron, and zinc levels, representing the higher concentrations of these elements in the outer layers of the grain. Incorporation of the germ in the bread also increases the levels over the values for white bread. The UK data for calcium and iron are higher than in other countries because of fortification of all flours except wholemeal.

Table 3 lists values for a number of vitamins. The concentration of most B vitamins is higher in the outer layers of the grain, so the values for the wholemeal breads are higher, except for folate in Danish breads. The thiamin and niacin values in the British breads represent the effects of fortification.

Rye Breads

Table 4 lists the compositions of several rye breads taken from the British and Danish food composition tables. This shows that rye breads fermented with yeast are very similar in overall composition to wheaten breads. The Danish appear to prefer a

Table 1 Major constituents of wheaten breads (per 100 g)

Bread type	*Water (g)*	*Protein (g)*	*Fat (g)*	*Carbohydrate (g)*	*Dietary fiber (g nonstarch polysaccharides)*	*Energy (kJ)*	*Country of origin*
Wholemeal	38.3	9.2	2.5	41.6	5.8	914	UK
Brown	39.5	8.5	2.0	44.3	3.5	927	UK
Wheatgerm	40.3	9.5	2.0	41.5	3.3	899	UK
White	37.3	8.4	1.9	49.3	1.5	1002	UK
Wholemeal	38.4	8.2	2.0	49.3	5.3	1054	Denmark
White	37.9	7.2	3.2	49.9	3.9	1092	Denmark
Wholemeal[a]	36.6	7.5	1.3	53.8	5.7	1017	Italy
White[b]	29.0	8.1	0.2	64.7	2.8	1167	Italy

[a]Pane integrale.
[b]Pane commune.

Table 2 Major inorganic constituents of wheaten bread (per 100 g)

Bread type	*Sodium (mg)*	*Potassium (mg)*	*Calcium (mg)*	*Magnesium (mg)*	*Phosphorus (mg)*	*Iron (mg)*	*Zinc (mg)*	*Country of origin*
Wholemeal	550	230	54	76	200	2.7	1.8	UK
Brown	540	170	100	53	150	2.2	1.1	UK
Wheatgerm	600	200	120	56	190	3.7	2.1	UK
White	520	110	110	24	91	1.6	0.6	UK
Wholemeal	511	198	49	39	268	1.65	1.25	Denmark
White	453	136	44	20	128	1.19	0.65	Denmark
Wholemeal	550	210	25		180	2.5	1.6	Italy
White	665	161	64		100	1.7	0.8	Italy

Table 3 Vitamins in wheaten breads (per 100 g)

Bread type	*Thiamin (vitamin B_1) (mg)*	*Riboflavin (vitamin B_2) (mg)*	*Niacin (mg)*	*Vitamin B_6 (mg)*	*Folate (μg)*	*Country of origin*
Wholemeal	0.34	0.09	4.1	0.12	39	UK
Brown	0.27	0.09	2.5	0.13	40	UK
Wheatgerm	0.80	0.09	4.2	0.11	39	UK
White	0.21	0.06	1.7	0.07	29	UK
Wholemeal	0.22	0.12	2.5	0.16	49	Denmark
White	0.18	0.10	1.2	0.07	70	Denmark
Wholemeal	0.10	0.12	2.9	0.12	39	Italy
White	0.02	0.03	1.4	0.07	29	Italy

Table 4 Composition of some rye breads (per 100 g)

Type of bread	*Water (g)*	*Protein (g)*	*Fat (g)*	*Carbo-hydrate (g)*	*Dietary fiber (g)*	*Na (mg)*	*K (mg)*	*P (mg)*	*Fe (mg)*	*Thiamin (mg)*	*Niacin (mg)*	*Folate (μg)*	*Country of origin*
Wholemeal	37.4	8.3	1.7	45.8	5.8	580	190	160	1.0	0.29	2.3	24	UK
Dark wholemeal	41.1	6.2	1.7	48.7	8.6	573	294	266	2.28	0.19	1.1	43	Denmark
Dark	41.5	6.2	1.7	48.1	9.2	558	296	285	2.3	0.21	1.1	43	Denmark
Light	42.8	6.2	1.3	47.6	7.5	562	266	208	2.1	0.20	0.7	43	Denmark

slightly higher moisture content, which reflects the higher water-binding activity of the carbohydrates in rye. The rye breads tend to contain higher amounts of dietary fiber, and this has a higher proportion of soluble nonstarch polysaccharides than wheat. Rye grains have a different structure to wheat, and the bran is less easily separated from the endosperm, so the light rye bread is of a higher extraction than white wheaten breads.

The wider range of rye breads cited in the Danish tables is typical of northern Europe, where rye breads and rye/wheaten mixtures are widely consumed. Rye proteins contain gluten levels that are less effective in increasing the volume of the bread, and the inclusion of a proportion of wheat is widely practiced in Germany and Scandinavia.

Nutritional Significance of Bread in the Diet

This derives both from the quantitative nutritional contribution of breads in the diet and from the qualitative characteristics of the components of both wheat and rye grains.

Qualitative Aspects

Protein Breads are often seen as important in the provision of carbohydrates, and consequentially as a source of energy, and over the past 30 years as a source of dietary fiber. This limited view of breads as carbohydrate foods is misleading because their proteins are important nutritionally. White bread contains 8.4 g per 1002 kJ (235 kcal), so the protein: energy ratio is well above that considered nutritionally essential. Protein–energy malnutrition is virtually unknown amongst populations where wheat is the stable cereal. The amino acid compositions of both wheat and rye proteins are reasonably well balanced with the sulfur-amino acids being limiting, so a combination of bread and vegetables provides a very good combination. Baking results in a small reduction in the available lysine content but still retains the overall protein quality.

Carbohydrates The major carbohydrate in bread is starch, with relatively minor amounts of free sugars. Some degradation of the starch occurs in the crust during baking, producing soluble dextrins and some Maillard reaction products. The production of bread leads to the formation of some retrograded starch, which is physiologically resistant to digestion in the small intestine. The resistant starch is formed from the amylose portion of the starch that combines to form insoluble crystalline products. The proportion of 'resistant starch' increases with staling of the bread. The bulk of the starch, however, in breads, especially white bread, is rapidly digestible and is responsible for the high glycemic index of white bread.

The dietary fiber in wheat comes principally from two sources: the outer branny layers contain cellulose and noncellulosic polysaccharides. Wheat and rye bran noncellulosic polysaccharides are both rich in arabino-xylans, which are important in the water-binding properties of the flour. The bran layers are lignified and resistant to bacterial degradation in

the large intestine. The dietary fiber in wheat bran is one of the more effective sources of dietary fiber for increasing fecal bulk.

The endosperm contains smaller quantities of thin-walled cells, the walls of which are rich in cellulose and β-glucans.

Minerals and vitamins Wholewheat breads contain significant amounts of potassium, magnesium, phosphorus, and iron, together with important amounts of thiamin, niacin, and small amounts of folates. A substantial proportion of the phosphorus is present as phytic acid, and this has been shown to reduce the bioavailability of calcium and iron. At the start of World War II, there was nutritional concern that the availability of dairy products would be reduced and that any increase of the extraction rate to conserve the imported grain would result in reduced amounts of calcium being available in the diet. This led to the fortification of wheat flours with calcium carbonate. When the extraction rate of bread was increased to make a higher proportion of the imported wheat grain to be used for human consumption, this also increased the amounts of thiamin, niacin, and iron in the bread. When, in the early 1950s, the production of white bread was again permitted, it was decide that the nutritional composition of all flours, except wholemeal, should be maintained by fortification with iron, thamin, and niacin to maintain the levels in the higher extraction breads available during the war.

Consumption and Nutritional Contributions

Since classical times, bread, especially wheaten bread, has provided the staple source of energy and the other nutrients it contains. This was particularly true for the mass of the population in ancient Egypt and classical Greece and Rome. This reliance of the lower socio-economic groups on bread as the staple source of energy and nutrients continued through medieval times to the early years of the 20th century in Europe. In the UK, bread was seen as a special food, and successive governments saw controlling the price of bread as the key to keeping the 'masses' contented, just as it was in Roman times. Even amongst the poorer members of the community, however, there was a preference for white 'low-extraction' flours that were seen as purer and conferred some measure of status on the consumer. This continues, despite repeated advice from the medical and other health professionals that higher-extraction breads are better for health. Probably, such breads were less easily adulterated in earlier times with cheaper beans or barley flour.

The key to assessing the dietary importance lies in considering the amounts of bread consumed together with the composition of those breads. The UK National Food Survey provides a continuous record of food consumption over the last 47 years. In its earlier years, it measured only foods purchased to be consumed within the home, but more recently, because the consumption of meals outside the home has increased, the Survey has included estimates of them.

During World War II, bread was not rationed and thus provided a major source of energy. As mentioned previously, the extraction rate of bread-making flours was controlled to around 80–85%, resulting in a bread that was gray/brown for most of the 1940s. Consumption was then of the order of 1721 g per person per week (Table 5), but when the war ended, rationing was introduced, and consumption fell to around 1420 g.

Control on flour was abolished in 1956, and consumption switched very rapidly from the 'national loaf' at about 80–85% extraction to white at about 1200 g, illustrating very clearly the popular preference for white bread. Since that time, the overall consumption of breads has shown a steady decline, reflecting the improved economic status of the population and the increased availability of other foods. The National Food Survey data for the earlier years are, however, for foods purchased for the home, and with improving economic conditions, the consumption of meals outside the home increased in importance. In recent years, the growth in 'sandwich shops' providing bread-based meals for consumption at midday means that the consumption of bread *per se* may not have declined so extensively.

Table 6 lists some data for 1997 on the patterns of bread purchases within different income groups,

Table 5 Changes in total bread consumption in the UK (grams per person per week)

Year	*1942*	*1950*	*1960*	*1970*	*1990*	*1990*	*1997*
Consumption	1721	1640	1291	1082	884	798	745
Comment	War time	Rationing	After controls were relaxed				

From MAFF (1991) *Household Food Consumption and Expenditure 1990*. London: HMSO; MAFF (1998) *National Food Survey 1997*. London: The Stationery Office.

Table 6 Consumption of different breads by income groups[a] for UK in 1997 (grams per person per week)

Income group	*A*	*B*	*C*	*D*	*State pensioners*
Bread type					
White	312	393	461	409	458
Brown	73	68	77	95	273
Wholemeal	87	83	79	63	139
Other breads	171	156	147	104	136
Total bread	635	700	765	772	867

[a]Income groups are based on the total household income: Group A highest 10%; Group B next 40%; Group C next 40%; Group D lowest 10% of families with earner in family.

Table 7 Nutrient contributions of breads to UK intake in 1997 combined contributions of white, brown, and wholemeal) (per person per day)

Nutrient	*Contribution*	*Percentage of total intake*
Energy (kcal)	142	7.9
Protein (g)	11.0	17.0
Fat (g)	1.2	1.5
Sugars (g)	1.6	1.3
Starch (g)	27.6	21.1
Dietary fiber (nonstarch polysaccharides) (g)	1.8	14.5
Calcium (mg)	60	7.3
Iron (mg)	1.3	13.9
Sodium (mg)	345	13.4

showing the greater consumption of bread by the lower income groups and especially amongst those receiving the state old age pension. This group includes those with additional pensions from employment, and the group as a whole may have a net income above those in the lowest income group, D. It is interesting to note the dominant position of white bread in the lower income groups. Bread consumption amongst the large families in the lower income groups tends also to be higher, emphasizing that bread still provides an important stable food.

Importance as a Source of Nutrients

Table 7 shows the contribution of bread to the intakes of some nutrients (per person per day) and the percentage contribution to the total intakes. These show that bread makes a greater contribution to protein intakes than energy. The contribution to fat intakes is very minor, but it must be recognized that bread acts as a vehicle for introducing fat into the diet via spreads and sandwich fillings. The dietary fiber intake comes almost equally from white and wholemeal breads because of the relative proportions consumed. Despite fortification, bread is a minor source of calcium in the diet, although iron fortification means that bread is a significant iron source. The contribution of bread to sodium intake has been a cause of some nutritional concern and is currently being addressed by the bread industry.

Conclusions

Bread, especially wheaten bread, has played an important role in the human diet ever since wheat began to be cultivated in the Middle East. It remains an important stable source of nutrients and energy, and as a source of complex carbohydrates, a food whose consumption, especially the higher-extraction types, should be encouraged as part of a healthy diet.

See also: **Protein**: Food Sources; **Rye**; **Starch**: Structure, Properties, and Determination

Further Reading

Holland B, Welch AA, Unwin ID, Buss DH, Paul AA and Southgate DAT (1991) *McCance and Widdowson's The Composition of Foods*, 5th edn. Cambridge: Royal Society of Chemistry and Ministry of Agriculture Fisheries and Food.

McCance RA and Widdowson EM (1956) *Breads White and Brown: Their Place in Thought and Social History*. London: Pitman Medical Publishing.

MAFF (1991) *Household Food Consumption and Expenditure 1990*. London: HMSO.

MAFF (1998) *National Food Survey 1997*. London: The Stationery Office.

Salvini S, Parpinel M, Gnagnarella P, Maisonneuve P and Turrini A (1998) *Banca Dati di Composizione per Studi Epidemiologici in Italia*. Milan: Tipografia F lli Verderio.

Dough Fermentation

M Collado-Fernández, University of Burgos, Burgos, Spain

Background

Fermentation involves a complex system of reactions brought about by microorganisms that may be present simultaneously. There are different types of fermentation resulting from the action of yeast and other microorganisms such as lactic acid bacteria. The principal type of fermentation is the action of the yeast on fermentable sugars to produce carbon dioxide (CO_2), ethanol, and some aromatic compounds.

The organoleptic characteristics of bread, namely, the texture, flavor, and aroma, depend on the conditions of the dough fermentation.

Fermentation begins when the yeast is in contact with the dough and finishes some minutes after the dough has been placed into the oven. There are two phases in the fermentation: the bulk fermentation or first fermentation or first proof and the fermentation or final proof or proving. The bulk fermentation tends to be substituted by intensive kneading or by other methods.

During fermentation, there must be a balance between gas production and gas retention and the characteristic viscoelastics of dough. The dough undergoes several changes like weight loss, changes in carbohydrates, and phytates.

Yeast

Yeast is a living unicellular organism, has a length of 6–8 μm, and is round to oval in shape. One gram of yeast contains around 15×10^9 cells. The yeast cell has an external membrane called the cell wall. This has a selective permeability and has an important role in controlling (by osmotic processes) the movement of nutrients and other substances into the cell and the release of other substances such as CO_2 and ethanol. Bakers' yeast is *Saccharomyces cerevisiae*. Commercial production of yeast is based on growth, on an industrial scale, of a pure selected strain. The cell culture obtained is separated from the culture medium and prepared for processing. The yeast can be compressed (most common), granular, dried in the form of a pellet, instantaneous, encapsulated, frozen, or in the form of a 'cream.' These variations are related to the physical form of the yeast and, mainly, the yeast moisture content. In terms of the speed of yeast fermentation, there are two types, standard and rapid, related to the maltose adaptation. Rapid yeast is more active than standard yeast, and is adapted for use in accelerated breadmaking systems.

Yeast is a versatile organism that can act under aerobic and anaerobic conditions. Under aerobic conditions, yeast grows quickly, owing to the oxygenation of sugar, which produces new yeast, water, and the energy needed for its growth. Under anaerobic conditions, yeast transforms the sugar into carbon dioxide and ethanol by fermentation.

The speed with which yeast consumes sugars and produces CO_2 depends on factors such as: the yeast strain used, resistance to osmotic pressure, maltose adaptation, yeast concentration, flour fermentative capacity, temperature and pH of the dough, and presence of dough inhibitors.

Strain

The selection of yeast strain is a function of the fermentative properties of the dough, the breadmaking process, the high stability of its characteristics, and a good industrial yield.

Resistance to Osmotic Pressure

Some strains have a high tolerance to osmotic pressure, but others do not and consequently die. The increase in osmotic pressure in the cell wall is due to factors such as salt and sugar concentration, the invertase action of the yeast itself. The decrease in water in the formulation of the doughs that include high levels of fat and sugar, and especially problems presented by frozen dough sugar; during freezing, the free water decreases, and the osmotic pressure increases.

Maltose Adaptation

In order to maintain a good production of gas, yeast requires a plentiful and continuous supply of glucose as an energy source. The action of α- and β-amylases of flour in damaged starch gives maltose (starch damage results from modifications in the structure of starch during the milling process). This disaccharide is transported into the cell, in which it is broken down into two molecules of glucose. The metabolic changes to this new source of glucose from starch are only active when yeast detects the presence of maltose in the medium, and its functioning takes some time (lag phase). Consequently, the rate of gas production decreases until this mechanism is working fully, and then, gas production increases. The decrease in gas production is lower as soon as the maltose adaptation begins, which is different for rapid and standard yeast strains. The choice of one over another depends on the breadmaking process to be used.

Yeast concentration

When sugar is not added to dough, the limiting factor of gas production is the quantity of maltose available. The increase in yeast concentration at one temperature leads to an increase in the rate of gas production and greater sugar depletion.

Fermentative Capacity of Flour

The quantity of amylase in flour determines the degree of glucose production from starch, so long as there is enough damaged starch.

Temperature

Temperature has an important influence on yeast activity: at 4 °C, there is no activity, and fermentation

does not occur, at 10–15 °C, the yeast activity and fermentation are slow, at 20–40 °C, fermentation is very active, at 45 °C yeast activity slows down, and at 55 °C, the yeast dies. Thus, the best storage temperature for yeast is 4 °C (0–10 °C). Depending on the strain used, an increase of 1 °C in the dough gives an 8–12% increase in fermentation speed. Although, at 38 °C, there is more gas production, there is also lactic and butyric fermentation (which is undesirable).

pH

Yeast is very tolerant to fluctuations in environmental pH (2–8), but the optimum pH for yeast activity is between 4 and 6. Usually, the pH of dough is 5, although the pH may be lower when sourdough is used.

The undissociated forms of organic acids produced from parallel fermentation to alcoholic fermentation, that is, produced by lactic acid bacteria of flour and by contamination of yeast, are inhibitors of fermentative activity. They have greater effects at lower pHs, the critical zone being between 4 and 5, and these values correspond to the final fermentation.

Presence of Dough Inhibitors

The copper and chlorine present in water used in breadmaking processes can act as fermentation inhibitors. Organic acids, such as those used as preservatives (acetic and propionic acid) and others formed during fermentation, can also act as fermentation inhibitors.

Sources of Sugar used by Yeast

Yeast uses four different sources of sugar: (1) trehalose endogenous, (2) free sugar from flour, (3) sugars obtained by the action of enzymes in oligosaccharides and polysaccharides in the flour, or added as additives, and (4) added sugars such as sucrose. Fermentation begins with trehalose, then later with the free sugars and sugars added, and finally with maltose. The hydrolysis of sucrose to glucose and fructose is faster than the fermentation of glucose and fructose. The consumption of glucose is greater than that of fructose, and initially, the consumption of maltose is low, but when most of the other sugars have been consumed, the yeast ferments the maltose rapidly. This also depends on the capacity of yeast to adapt to maltose.

Carbohydrate Metabolism

Yeast ferments monosaccharides – glucose (mainly), fructose, and mannose, disaccharides – sucrose, maltose, and trehalose – and trisaccharides (rafinose).

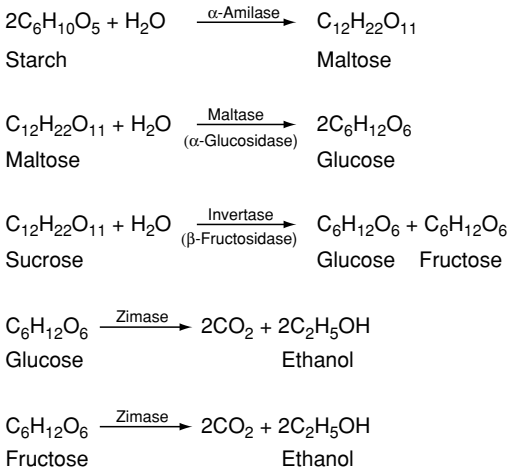

Figure 1 Carbohydrate transformation during dough fermentation.

Yeast does not consume lactose. The transformation of carbohydrates during dough fermentation is shown in Figure 1.

Yeast can transform carbohydrates in different ways: via the Embden–Meyerhof–Parnas pathway (EMP, glycolysis) and from hexose monophosphate (HMP). Under anaerobic conditions, the carbon dioxide formed is by the EMP pathway, and under aerobic conditions, the carbon dioxide may be formed by HMP from pyruvate (via the tricarboxylic acid (TCA) cycle or Krebs cycle). Under aerobic conditions, yeast grows rapidly and accumulates a small amount of carbon dioxide and ethanol, whereas under anaerobic conditions, yeast grows slowly and produces a large amount of carbon dioxide and ethanol.

The pyruvate formed in these ways is very important as a precursor to acids, alcohol, and other volatile compounds, which have an influence on the flavor and aroma of bread (Figure 2).

Production and Retention of Gas

The quality of gluten is related to gas retention, but also, the degree of gas retention depends on the rate of diffusion. During kneading, air is incorporated into the dough, and air bubbles are formed. These bubbles contain mainly nitrogen, because it has a low solubility in water, and oxygen is quickly used up by the yeast cells within the dough. Yeast cannot create new bubbles, as the pressure (P) in a bubble is related to the radius (r) of the bubble and the interfacial tension (γ) by the law of Laplace: $P = 2\gamma/r$. The creation of new bubbles is only possible by punching or remixing the dough, or by the use of surfactants that change the interfacial tension. The more bubbles are formed, the finer the grain is.

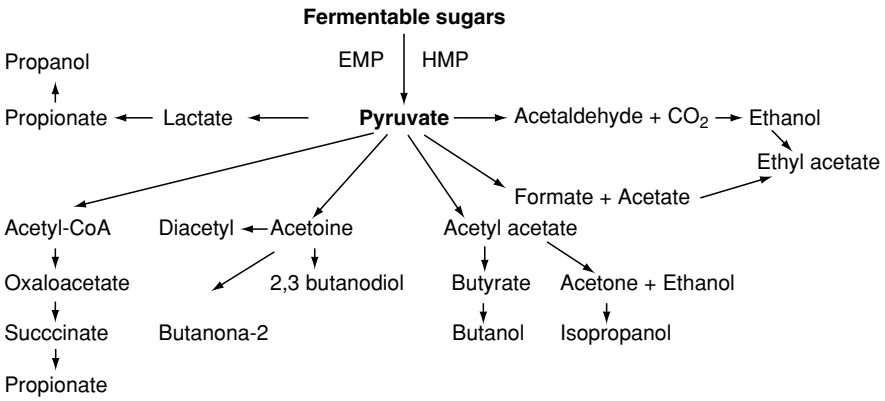

Figure 2 Some of the breakdown products from sugar metabolism.

During kneading, yeast is mixed into the dough, and fermentation starts. The major breakdown products are carbon dioxide and ethanol. Initially, carbon dioxide is produced in the aqueous phase, the pH decreases, and this phase becomes saturated with carbon dioxide. As the fermentation proceeds, carbon dioxide is retained in the air bubble, and nitrogen is displaced. Then, the dough is leavened, because the aqueous phase is saturated by CO_2 and the newly formed CO_2 is retained in preexisting bubbles. At this point, the quality of gluten is also very important to retain the gas, and thus, carbon dioxide is retained in the dough in two phases: contained within the gas cells and dissolved in the aqueous phase. At the pH of dough, most of the carbon dioxide is present as CO_2 and a small amount of this as CO_3^{2-}, HCO_3^-, or H_2CO_3.

Only 45% of the total gas produced is present at the end of fermentation. The loss of carbon dioxide is due to the different steps of breadmaking process, namely initial fermentation, punching, rounding, molding, and final fermentation.

The gas-retention capacity of the gluten decreases with time and is critical when the dough is placed into the oven. The high temperatures expand the retained gas and increase the volume of the dough, and so if the gluten has a good gas-retention capacity, it will yield bread with a good volume, otherwise the bread will be flat. The vaporization of ethanol and water present in the dough also contribute slightly to dough expansion.

Fermentation

There are two phases in fermentation: bulk fermentation, first fermentation, or first proof and the main fermentation, final proof, or proving. Bulk fermentation takes place from final kneading to the dividing of the dough. In this phase, the dough undergoes several physical modifications that allow it to be cut and molded. This phase tends to be substituted by intensive kneading or other methods. The final proof takes place from molding to baking, so this phase really is the fermentation.

The relative lengths of time of these phases depend on the physical properties of the dough, which depend on the kind of wheat used: when using soft flours that produce dough that is short and not very elastic, the fermentation time should be reduced.

During fermentation, the temperature and relative humidity need to be taken into account. The temperature used depends on the kind of bread and the breadmaking process (usually 28–30 °C), with lower temperatures requiring longer fermentation times (1–1.30 h or 2–4 h), but also depends on the yeast concentration used. If the temperature is lower, the dough will be very cold, the fermentation will be longer, and gas production will decrease. As the fermentation is longer, the concentration of sugar increases, and the texture of the bread will be of a lower quality and rough grained, and will have a colorful crust with shorter shelf-life. If the temperature is higher, the crumb will be rougher, and nonalcoholic fermentation may occur. If the relative humidity is < 75%, the skin of the dough will be very dry and will lose its elasticity.

During fermentation, the pH of dough changes, principally owing to the formation of lactic acid. Initially, dough has a pH of about 6.2, and during fermentation, the values are about 5.76 or 5.67. This acidic environment improves the formation of gluten, which is more extensible, and in bread, this slows down the growth of mold.

The pH of dough has a great influence on fermentation, because yeast needs an acid environment (pH 5.8–6.2), and also influences the properties of protein and the activity of enzymes.

During fermentation, dough increases the force (*W*, work (related to the curve of the alveogram)) and decreases the tenacity (*P*, pressure (related to the height of the curve of the alveogram)) so the relation *P*/*L* decreases (*L* being extensibility). The capacity of dough to withstand excessive mechanical work is called dough tolerance. Generally, doughs that ferment slowly are more tolerant.

Fermentation Types

Not all sugar is transformed by yeast. Lactic acid bacteria (usually present in dough, especially in sourdough) carry out fermentation, consuming sugar and producing CO_2 (Figure 3). The types of fermentation are: alcoholic fermentation, lactic fermentation, butyric fermentation, and acetic fermentation.

Alcoholic Fermentation

This is the principal type of fermentation. Yeast transforms fermentable sugars, mainly glucose, into carbon dioxide and ethanol. The optimum pH for alcoholic fermentation of dough is 5. If the pH is higher (> 6), the breakdown products of fermentation are glycerine and acetic acid as well as carbon dioxide and ethanol.

Lactic Fermentation

Lactic fermentation is achieved by the hydrolysis of lactose or glucose by lactic acid bacteria, giving lactic acid. Usually, yeast is contaminated with lactic acid bacteria. The optimum temperature for lactic fermentation is 35 °C, so this fermentation is slow at the temperature of yeast fermentation. The lactic acid acts in the dough, giving a more elastic gluten.

Butyric Fermentation

Different bacteria, such as *Clostridium butyricum*, may use the lactic acid to produce butyric acid. The temperature of this reaction is about 40 °C, so during dough fermentation, these bacteria cannot act. However, if the dough has a long fermentation time, it increases its temperature (more than 32 °C), and butyric fermentation may occur, affecting the aroma of bread. The pH of this reaction is 3.2–3.8.

Acetic Fermentation

Acetic fermentation is carried out by *Mycoderma aceti*, which transform the ethanol into acetic acid. The optimum reaction for *Mycoderma aceti* occurs under aerobic conditions. The contribution of acetic acid alone to the total acidity of the dough is 5%. The excess of acetic acid content gives short and rigid

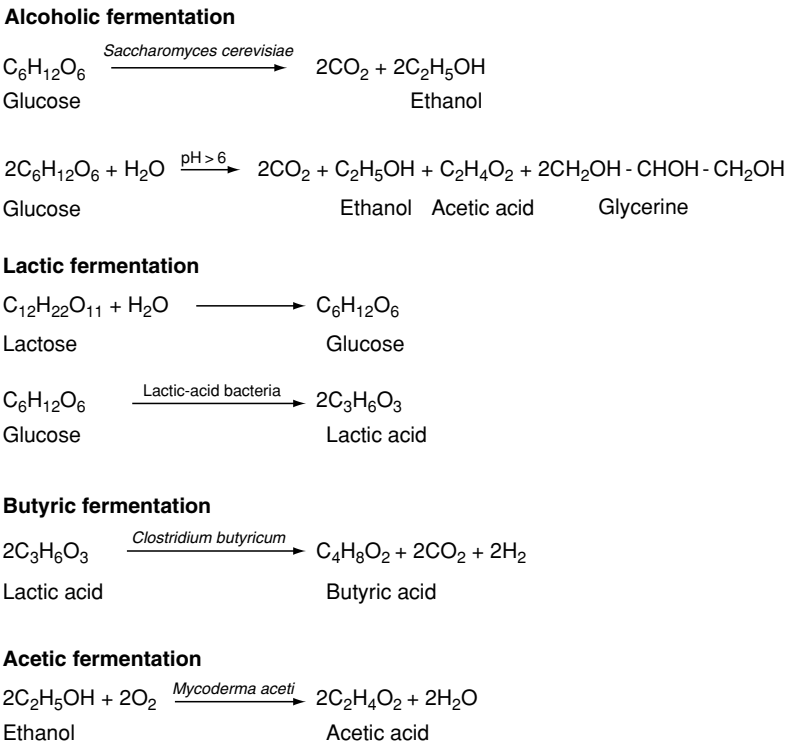

Figure 3 Types of fermentation during the breadmaking process.

gluten. Therefore, it is necessary to insure that the ratio of lactic acid to acetic acid formed is 3/1.

Controlled Fermentation

Controlled fermentation is one of the new techniques to apply cold, to stop, or to decrease the fermentation. The possibility of controlling and programming fermentation has the advantage in that it is possible to regulate bread production, avoid working at night, produce fresh bread any time, and utilize the advantages of long fermentation, i.e., a better shelf-life aroma, and flavor.

Raw Materials

The flour must be of good quality ($W = 160$–180; $P/L = 0.4$–0.6, falling number 300–350 s), because the dough must not ferment before the block phase, and flour needs less enzymatic activity than traditional fermentation. Standard yeast must be used, because the fermentation must not begin before the block phase. The yeast concentration is usually 2–3% (based on flour weight). Fresh sourdough is also added in the bread mixture.

Manufacturing

The kneading time required for dough development is 20% longer than the traditional breadmaking process, because the flour is stronger. Two to 3% less water in the formulation is recommended, to insure that the dough is not too soft, and the final temperature of the dough should be 22 °C to avoid premature fermentation. The dough is divided, rounded, and molded as soon as possible, and then pieces of the dough are placed into a controlled fermentation cabinet.

Proof Cabinets

The design of the proof cabinet has evolved with time. Traditional bakeries used wood cupboards with large boxes to ferment the dough, and some still use these. These proof cabinets present some problems, because they are affected by the temperature of the workroom, and the fermentation is not the same in all the pieces of the dough. Nowadays the proof cabinets are isolated and have automatic control of temperature and relative humidity. The ideal relative humidity should be the result of the addition of dough hydration and moisture from the flour (for example, if water added to the mixture is 55%, and the flour moisture is 14%, the ideal relative humidity will be 69%), so the dough will be neither rough nor sticky.

Types of Proof Cabinet

There are several different types of proof cabinet: the traditional fermentation cabinet, the controlled fermentation cabinet, and the block fermentation cabinet.

Traditional fermentation cabinet In this cabinet, only heat and humidity are used. The usual conditions are: temperature of 28–32 °C and relative humidity of 70–85%. If a high temperature (> 30 °C) is used, a high relative humidity (> 75%) is also required.

When the temperature is > 28 °C, during dough fermentation, lactic and butyric acid are formed, increasing in numbers with increasing temperature. Also, the enzymatic reactions in dough are more active at these temperatures, so the dough is softer, the bread volume increases considerably, and the bread is insipid and has a shorter shelf-life.

If the temperature is < 26 °C, there is less production of lactic and butyric acid during the fermentation, the fermentation time is longer, the enzymes are less active, and the bread volume is lower, but the bread has a good flavor.

The reduction in the amount of yeast and the increase in fermentation time lead to a bread with a firm crumb and thick crust.

Controlled fermentation cabinet This cabinet controls dough fermentation temperature allowing progressive heating. Previously, these cabinets did not canalize the air, and the dough was rough. Heating was rapid, and there was condensation on the dough surface, which gave a red and translucent crust. The result was bread with great variations in quality, color and volume. Nowadays, these cabinets have a regulating system, depending on the size and volume of the pieces, but there are four phases: block, refrigeration, heating, and fermentation.

Block In the first phase, the dough is cooled, and fermentation is stopped when the temperature in the dough reaches about 2 °C. Usually, temperatures of about −2 °C (0 to −14 °C) are used.

Refrigeration In this phase, the temperature of the dough is usually at 0–2 °C until fermentation begins.

Heating In this phase, there is a gradual increase in temperature to prevent the dough surface becoming rough. The dough temperature increases gradually until the fermentation temperature is reached.

Fermentation It is very important to regulate the temperature and relative humidity to ensure uniform fermentation of dough.

Block fermentation cabinet This cabinet has only two phases: block and refrigeration. The functions of these phases are the same as that of the controlled fermentation cabinet, but the dough temperature needs to be increased before fermentation. The dough is kept in this cabinet for some time before fermentation, and another cabinet is needed for fermentation.

Physical and Chemical Changes in Dough During Fermentation

The transformation of the dough into bread is very complex, because dough contains very reactive chemical compounds, enzymes, microbial flora, different ingredients, and additives (depending on the formulation). In addition, the formation of dough may have different processing conditions (depending on the breadmaking systems used). Therefore, bread shows great variations in organoleptic characteristics. Most of the chemical, biochemical, and microbiological modifications that take place, during fermentation depending on processing conditions, thus it may be possible to lead fermentation in some way by controlling the conditions.

Yeast's functions during fermentation are:

- to produce enough CO_2 and in time increase the dough and make the dough softer;
- to modify the physical properties of gluten, for the maturation of the dough;
- to decrease the pH of the dough, thus allowing lactic bacteria to grow;
- to produce certain chemical compounds that give bread its characteristic flavor; and
- to modify the bread volume and texture.

The functions of lactic acid bacteria during fermentation are:

- to produce acids that inhibit the development of undesirable microorganisms, producing the conditions for enzymatic processes.
- to produce catabolites that influence the dough rheology and the sensorial characteristics of bread.

During fermentation, the dough undergoes several changes, such as weight loss and changes in carbohydrates, phytates, lipids, nitrogen compounds, and flavor and aroma compounds.

Weight Loss

The decrease in dry matter content (fermentation loss) during fermentation observing can be estimated by the volume of carbon dioxide evolved. The major proportion of carbon dioxide produced during fermentation results from alcoholic fermentation, and other fermentations may contribute about 5%. During alcoholic fermentation, the consumption of 1 mol of glucose produces 2 mol of carbon dioxide, according to Gay Lussac's law, and 1 ml of carbon dioxide released corresponds to a decrease in dry matter content of 3.65 mg. Thus, fermentation loss is influenced by the breadmaking processes used, and the straight dough method has less fermentation loss than the sponge and dough process (3.2% versus 5.5%, based on flour weight). In the Chorleywood process, the yield of bread from flour is about 4% higher than in bread made by traditional methods. This difference is due to a substantial decrease in fermentation losses.

Changes in Carbohydrates, Phytates, and Lipids

The carbohydrate concentration decreases during fermentation. The most important change during the fermentation is the transformation of fermentable sugar into carbon dioxide and ethanol by yeast action.

Sugar metabolism by yeast has already been mentioned. Sugar metabolism may be changed by temperature conditions (fermentation being controlled). Increasing the sugar concentration increases gas production up to the limit of osmotic pressure at which yeast cannot survive. Decreasing the pH increases gas production.

Another path for sugar metabolism is by lactic acid bacteria. Heterofermentative lactic acid bacteria produce carbon dioxide, but this gas production is less than that from yeast (Figure 4). The principal function of lactic acid bacteria is to decrease the pH, and increase the aroma and flavor of bread.

The lactic acid/acetic acid relationship is very important, because it influences the aroma and flavor of bread. For rye bread, this relationship is closer than that of wheat bread, and the intensity of bread flavor is higher. The production of acids may be controlled

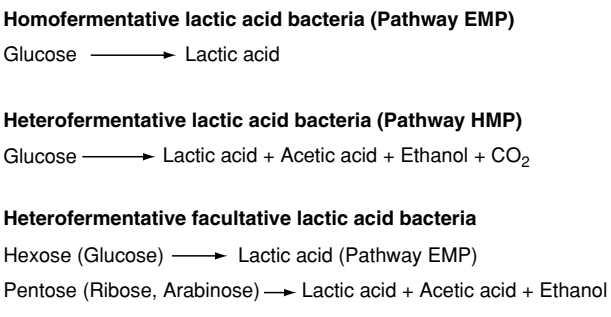

Figure 4 Metabolism of sugar by homofermentative and heterofermentative lactic acid bacteria.

by adding acid-producing microorganisms, such as heterofermentative facultative lactic acid bacteria, which produce acetic acid. The hydration of dough, temperature, and presence of yeast also have an influence on production of acids.

In terms of chemical changes in starch, as mentioned earlier, starch is a source of sugar for fermentation. α-Amylases hydrolyze damaged starch forming the substract for β-amylases action yielding the breakdown products dextrin and maltose.

In terms of nondigestible carbohydrates, the concentration of fiber increases during fermentation, and in long periods of fermentation, there is an increase in the amount of hemicelluloses.

Phytates are associated with the fiber content. Phytic acid in wheat and flour is present as phytine associated with calcium and magnesium. Phytine has a heterogeneous distribution in the kernel of wheat. There is a greater proportion of phytine in the germ and bran than in the endosperm. During fermentation, there is dephosphorylation of phytic acid by phytase, and the concentration of phytate decreases, decreasing the capacity to form a complex with minerals and protein.

Concerning lipids, polar lipids improve the structure of bread and increase the soft crumb. Nonpolar lipids deteriorate the structure of bread and decrease the bread volume.

Changes in Nitrogenous Compounds

The effects of fermentation on nitrogenous compounds are as follows: (1) there is a modification in the structure of gluten, thus making the dough more extensible and improving the gas retention; (2) there are changes in amino acids, peptides, and other soluble compounds, which are precursors of the flavor and aroma of bread, and a source of nutrients.

The protein network in dough is able to retain carbon dioxide in the air bubbles and is sufficiently elastic to expand without disrupting during the fermentation and when the gas volume increases during baking. This is possible because of the chemical changes in the bonds of gluten and other wheat proteins. The mechanical work during the dough mixing process breaks some of the hydrogen and disulfide bonds, linking adjacent protein chains, and the new bonds formed are in different positions. Expansion of the gas formed during fermentation produces new bond changes, thus improving the dough extensibility. When the dough matures, there is a balance between the consistency of the three-dimensional protein network, its extensibility, and its permeability.

Amino acids and peptides may be formed by proteolysis of the dough. Lactic acid bacteria have proteinases and peptidases, and the free amino acids and peptides may be metabolized by the yeast. During fermentation, the amino acids may lose an atom of carbon giving an aldehyde or two or more atoms of carbon, yielding inferior organic acids or hydrocarbons. The breakdown products of oxidation and reduction of aldehyde yield acid and alcohol, respectively. Alcohols and acids react to yield esters, and thus, there are a many compounds related to aroma and flavor (Figure 2).

The amino acids are precursors of aroma and flavor from the Maillard reaction, which takes place during baking. The overfermentation of dough depletes the amino acids and the color and flavor of bread decreases.

In conclusion, depending on the breadmaking process used, and if sourdough is added, the organoleptic characteristics of bread will be different, due to their influence on changes that take place in dough during fermentation.

See also: **Bread**: Dough Mixing and Testing Operations; Breadmaking Processes; Sourdough Bread; **Carbohydrates**: Metabolism of Sugars; **Dietary Fiber**: Effects of Fiber on Absorption; **Flour**: Analysis of Wheat Flours; **Phytic Acid**: Nutritional Impact; **Protein**: Requirements; Functional Properties; Quality; **Starch**: Structure, Properties, and Determination; Functional Properties; **Wheat**: Grain Structure of Wheat and Wheat-based Products; **Yeasts**

Further Reading

Cauvain SP and Young LS (1998) *Technology of Breadmaking*, 1st edn. London: Blackie Academic & Professional.

Hoseney RC (1994) *Principles of Cereal Science and Technology*, 2nd edn. St. Paul, MN: American Association of Cereal Chemists.

Jenson I (1998) Bread and baker's yeast. In: Wood BJB (ed.) *Microbiology of Fermented Foods*, 2nd edn. vol. 1, pp. 172–198. London: Blackie Academic & Professional.

Oura E, Suomalainen H and Viskari R (1982) Breadmaking. In: Rose AH (ed.) *Economic Microbiology. Vol. 7: Fermented Foods*, pp. 87–146. London: Academic Press.

Kent N and Evers AD (1994) *Technology of Cereals: An Introduction for Students of Food Science and Agriculture*, 4th edn. Oxford: Pergamon Press.

Kulp K and Ponte JG (2000) *Handbook of Cereal Science and Technology*, 2nd edn. revised and expanded. New York: Marcel Dekker.

Lorenz KJ and Kulp K (1991) *Handbook of Cereal Science and Technology*. New York: Marcel Dekker.

Pomeranz Y (1987) *Modern Cereal Science and Technology*. New York: VCH.

Pomeranz Y (1988) *Wheat: Chemistry and Technology*, 3rd edn. St. Paul, MN: American Association of Cereal Chemists.

Quaglia G (1992) *Scienza e tecnologia della panificazione*. Pinerolo: Chiriotti Editori.

Rehm HJ and Reed G (1983) Baked goods. In: *Biotechnology: A Comprehensive Treatise, Vol. 5: Food and Feed Production with Microorganisms*, pp. 1–83. Weinheim: Verlag Chimie.

Ward OP (1989) *Fermentation Biotechnology: Principles, Processes, and Products* Milton Keynes, UK: Open University Press.

Zimmermann FK and Entian KD (1997) *Yeast Sugar Metabolism: Biochemistry, Genetics, Biotechnology, and Applications*. Lancaster, PA: Technomic.

BREADFRUIT

D Ragone, National Tropical Botanical Garden, Kalaheo, HI, USA

Introduction

Breadfruit has long been a staple food in the Pacific islands and is now widely distributed and used throughout the tropics, as is a related species known as breadnut. The discovery of breadfruit by westerners and subsequent distribution of breadfruit and breadnut in the 18th to 20th centuries are discussed. Two species of breadfruit (*Artocarpus altilis* and *A. mariannensis*) and breadnut (*A. camansi*) are described. The importance and use of breadfruit in various Pacific island groups and methods of preparation, including traditional techniques to preserve fruits, are summarized. Production and use of breadfruit and breadnut in the Caribbean includes information on local and export markets. Harvest and yields of both crops are discussed. Use and nutritional composition of breadfruit and breadnut for fresh consumption are provided, as is an overview of commercial processing and postharvest handling of breadfruit.

History and Distribution of Breadfruit and Breadnut

Breadfruit derives its name from the fact that the fruits, when baked or roasted, have a starchy, dense consistency similar to bread or root crops such as potatoes, yams, or sweet potatoes. Native to New Guinea and possibly also the Moluccas, breadfruit has been an important staple crop for more than 3000 years. Islanders spread it throughout the vast Pacific during their voyages of exploration and discovery. By the late 1500s, when western ships first ventured into the region, breadfruit had been established in every island group settled by indigenous islanders. The only exceptions were New Zealand, which was too cool for this subtropical plant, and certain small coral atolls in Micronesia and Polynesia, where it was too dry to grow the tree.

Breadfruit was a remarkable and delicious food for western sailors arriving on those distant shores. Voyages to the Pacific from Europe were arduous, months-long passages, and the quality of food served to the crew became more and more unpalatable as the voyage progressed. The main provision was generally weevil-infested hardtack (flour and water baked into a biscuit). The sailors were delighted with this new island food. Fresh-roasted in a fire, breadfruit has a satisfying, aromatic smell, much like bread as it comes out of the oven. Its moist, dense, doughy texture and subtle flavor round out its similarity to bread. The scientific name, *Artocarpus altilis*, is derived from the Greek, with *artos* meaning bread, *carpus* meaning fruit, and *altilis* originating from the word *utile*, meaning useful, and has memorialized it as the tree which produces fruits like bread.

As those voyagers returned home they spread the word about this wondrous new plant and its potential as a food crop for other tropical countries. The growing interest in breadfruit set the stage for one of the grandest sailing adventures of the 18th century – Captain William Bligh's ill-fated mission on *HMS Bounty* to introduce Tahitian breadfruit to the Caribbean – and the extraordinary tale of courage and sailing skill that saved the survivors of the mutiny. The hundreds of breadfruit plants aboard ship were less lucky; the mutineers tossed them all overboard and sailed back to Tahiti and then on to Pitcairn Island.

Lesser known is that in 1793 Captain Bligh successfully completed his mission by introducing more than 600 plants of several seedless Tahitian cultivars to the islands of St Vincent and Jamaica. Most of the breadfruit in the Caribbean today originated from those plants. During this period the British and French also distributed breadfruit to Mauritius, the Maldives, Indonesia, and Sri Lanka. In the 19th and early 20th century breadfruit was widely distributed throughout the tropical countries of the world: Central and South America, India, South-east Asia, northern Australia, Africa, and Madagascar. However, its popularity and use vary greatly by locale. Breadnut, *A. camansi*, a closely related species from

the Philippines – grown for its nutritious, high-protein seeds – was also introduced to other tropical areas where it is now widespread, especially in the Caribbean, parts of Central and South America, and coastal West Africa.

Description of Breadfruit and Breadnut

A member of the Moraceae or fig family, breadfruit is a stately and attractive evergreen tree reaching heights of 15–20 m or more. The tall, straight trunks may be as large as 2 m in diameter at the base. It has large glossy, dark-green, leathery leaves ranging from almost entire with only slight lobing to deeply pinnately lobed (Figure 1). The long-lived, multipurpose trees begin bearing in 3–5 years (Figure 2) and are productive for many decades. They provide nutritious food, timber, and medicine, as well as feed for animals. The trees require little attention and input of labor or materials and can be grown under a range of ecological conditions.

Fruits are usually globose (Figure 3) to oblong, ranging from 8 to 20 cm wide and up to 30 cm long and weigh from less than 0.5 kg to more than 5 kg, depending on the cultivar. Each fruit is composed of the fused perianths of up to 2000 flowers attached to the fruit axis or core. As the fruit develops this area grows vigorously and becomes fleshy at maturity, forming the edible portion of the fruit. The rind is attractively patterned with five- to seven-sided disks, each representing the surface of an individual flower. In the center of each disk is a distinctive scar formed by the withered remains of the stigmatic lobes. The fruit rind is light green, yellowish-green, or yellow when mature and often stained with dried latex exudations. The surface varies from smooth to slightly bumpy or spiny. The flesh is creamy white or pale yellow in color and may be seedless except for numerous minute, abortive seeds surrounding the fruit core. Seeded breadfruit contain from one to many normal seeds (Figure 4). Breadfruit is considered to be parthenocarpic and does not require pollination to develop normally.

A third species of breadfruit, *A. mariannensis*, grows wild in the islands of Palau, Guam, and the Mariana Islands and is cultivated throughout the islands of Micronesia. The leaves are generally smaller and glossier than those of *A. altilis* and

Figure 1 Typical fruit and leaf of seedles breadfruit.

Figure 2 Productive 4-year-old breadfruit tree.

Figure 3 Smooth-textured globose breadfruit.

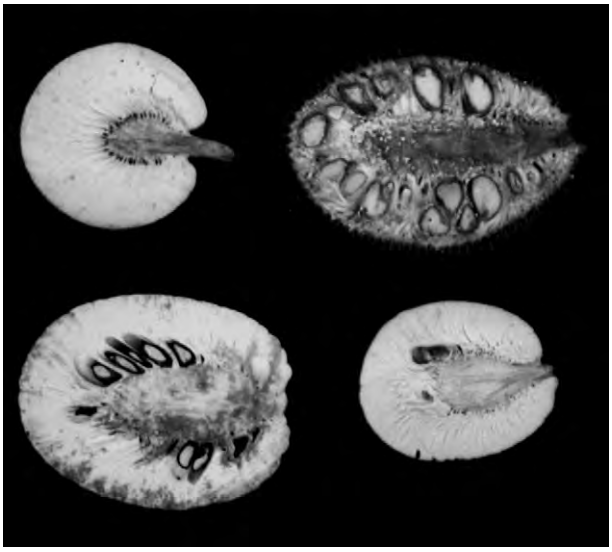

Figure 4 Variation in seed number of breadfruit and breadnut.

A. camansi. They are usually entire or have a few lobes, typically in the upper third of the leaf. The dark green cylindrical or asymmetrical fruits are generally smaller than breadfruit, with deep yellow-colored flesh when mature. Each fruit contains several to many seeds. Hybrids between *A. altilis* and *A. mariannensis* are common in Micronesia.

Production and Use

In the Pacific Islands

Breadfruit has long been an important staple crop in the Pacific Islands, and the islanders have selected and cultivated hundreds of cultivars. In Melanesia, seeded breadfruit grows wild in lowland areas of New Guinea, and the seeds are a valued food and an important component of the subsistence economy. Breadfruit is cultivated throughout the Solomon Islands and is a major subsistence food only in the easternmost islands. Breadfruit is equally important in Vanuatu, and it is still widely grown and used in certain areas of Fiji. Throughout Polynesia, breadfruit is an important staple food crop, especially in Samoa. Breadfruit remains the main staple food on atoll islands in Micronesia, and is the most important crop in season on the high volcanic islands where it is a primary component of traditional agroforestry systems.

While most breadfruit is produced for subsistence purposes, small quantities of fresh breadfruit are available for sale in village and town markets, especially in population centers where urban households must purchase traditional foods and rely heavily on imported foods. In Hawaii, fresh breadfruit is occasionally available in ethnic grocery stores and from local farmers' markets, but demand far exceeds supply. There is interest in establishing commercial breadfruit plantings to provide fresh fruits and chips for the local market and export. Although little breadfruit is currently exported from the Pacific islands, a large potential market for fresh fruits exists in the large communities of Pacific islanders living in Auckland, Honolulu, Los Angeles, and other urban areas of the USA and Canada.

Everywhere in the Pacific the starchy fruit is basically prepared and consumed in a similar fashion. It is typically cooked by roasting halved or whole in an open fire or earth oven, boiling, or occasionally is fried as chips. Because breadfruit, like all starch crops, is relatively bland, islanders have developed simple methods to enhance its flavor. Freshly grated and squeezed coconut cream is often added. A popular preparation is pudding made from mature or ripe cooked breadfruit that is grated or pounded, mixed with coconut cream, wrapped in leaves and baked. In many islands, these starchy puddings are the main form in which breadfruit is consumed. Puddings make a portable, tasty travel food that keeps for several days. Steamed or boiled breadfruit is also pounded until it becomes paste-like or doughy. It is preferred when at least 1 day old and can be kept for several days, when it begins to ferment and sour.

Traditional methods of preserving fruits Since breadfruit is a seasonal crop that produces much more than can be consumed fresh, Pacific islanders have developed many techniques to utilize large harvests and extend its availability. Preserved breadfruit also adds diversity to the daily diet. Drying is the simplest method for preserving breadfruit. Cooked breadfruit is halved or sliced into bite-sized pieces and dried over hot stones heated in a fire. Occasionally, thin slices are dried in the sun. Dried breadfruit is usually eaten without additional preparation, but it can be made into soup or ground into meal and mixed with water or coconut cream to make a porridge.

The most common method of preservation is the preparation of pit-preserved breadfruit. Pit storage is a semianaerobic fermentation process involving intense acidification which reduces fruit to a sour paste. Analysis of breadfruit fermented for 7 weeks showed that the starch was broken down first to maltose, then glucose, and eventually to lactic acid and carbon dioxide. Protein and carbohydrate levels remained relatively unchanged, while fat content and iron and calcium levels increased slightly (Table 1).

Fermented breadfruit is still made every season throughout Micronesia and to a limited extent elsewhere in the Pacific. Mature fruits are peeled, cored, and in the atolls, soaked in sea water for 12–48 h. They are then placed in a leaf-lined pit (Figure 5) and covered with leaves and a thick layer of soil or rocks. Fermented breadfruit can last for a year or more, and it is removed and eaten at various stages of fermentation depending upon need and taste preference. Fermented breadfruit is usually washed, then pounded or kneaded, and cooked before being eaten. Traditional storage practices are being modernized to fit contemporary lifestyles. Fermented breadfruit can be made in plastic tubs, coolers, or other containers that can be made relatively airtight. There are several advantages: it yields a cleaner, more uniform product, it is less labor-intensive, and smaller quantities can be made.

Table 1 Proximate composition of breadfruit and breadnut seeds per 100 g (dry weight) edible portion

	Breadnut seeds[a,c,d]	*Breadfruit seeds*[b]
Water (%)	56.0–66.2	47.7–61.9
Protein (g)	13.3–19.9	7.9–8.1
Carbohydrate (g)	76.2	26.6–38.2
Fat (g)	6.2–29.0	2.5–4.9
Calcium (mg)	66–70	46.6–48.3
Potassium (mg)	380–1620	
Phosphorus (mg)	320–360	186–189
Iron (mg)	8.7	2.3
Magnesium (mg)	10.0	
Niacin (mg)	8.3	1.8–2.1
Sodium (mg)	1.6	
Thiamin		0.13–0.33
Riboflavin		0.08–0.10
Ascorbic acid		1.9–22.6

Data from [a]McIntoch C and Manchew P (1993) The breadfruit in nutrition and health. *Tropical Fruits Newsletter* 6: 5–6; [b]Murai M, Pen P and Miller CD (1958) *Some Tropical Pacific Foods*. Honolulu: University of Hawaii; [c]Negron de Bravo E, Graham HD and Padovani M (1983) Composition of the breadnut (seeded breadfruit). *Caribbean Journal of Science* 19: 27–32; [d]Quijano J and Arango GJ (1979) The breadfruit from Colombia – a detailed chemical analysis. *Economic Botany 33: 199–202.*

In the Caribbean Islands

In the Caribbean breadnut and breadfruit are both widely grown, with seedless breadfruit more commonly cultivated and used. Since its introduction in the 1790s it gradually became an accepted food and an important component of the daily diet. In most cases, breadfruit is used to supplement other staple starchy foods in season and is the main source only in certain rural areas. It is a major source of food in Jamaica where more than 2 million trees were estimated to be growing in the 1970s. In the Windward Islands breadfruit is very popular for backyard planting in urban areas, and scattered trees are found island-wide. Breadfruit is prepared boiled, steamed, or roasted, and has lent itself to the creation of regional dishes such as 'oil down' – made with salt-cured meats, breadfruit, coconut milk, and dasheen leaves – which is popular in Trinidad, Tobago, and Grenada. In season, fresh breadfruit can be found in any local market, and a small quantity of processed breadfruit products such as frozen, dehydrated, and canned slices, flour, chips, and candied male flowers are also locally available. Breadnut is popular mainly in Trinidad, Tobago, and Guyana where immature and mature seeds are boiled and eaten as a snack.

The Caribbean is the major supplier of breadfruit to Europe, the USA, and Canada, providing more than 90% of the fruits for the UK market, with the rest coming from Mauritius. Jamaica is one of the largest exporters in the region. Most of the fruits are

Figure 5 Leaf-lined breadfruit fermentation pit in Micronesia.

harvested from backyard trees and small plantings, but there are efforts underway to establish orchards to support and expand the export market in the Caribbean.

In Other Tropical Areas

Breadfruit is more a subsistence crop than a commercial crop in most areas of the world, with the Pacific and Caribbean islands being the major production areas. In season, breadfruit and/or breadnuts supplement other staple foods for home consumption in Indonesia, the Philippines, and parts of West Africa, Central and South America, South-east Asia, India, and Sri Lanka. Some fruits may be available at local markets in these areas, especially Indonesia. Breadfruit has never been adopted as a major foodstuff in Africa, possibly because of failure to introduce ways of making the fruit into a food acceptable to local tastes. Recently, farmers in Queensland, Australia, have become interested in establishing breadfruit orchards for the export market.

Harvest and Yields

Breadfruits are harvested as needed and generally picked when mature, but not yet ripe. Breadnut seeds are harvested from ripe fruits and the seeds are separated from the soft pulp. Breadfruits are usually harvested with a sharp scythe or curved knife attached to the end of a long, sturdy pole (Figure 6) and are allowed to drop to the ground in most areas. Fruits that fall to the ground are damaged and soften sooner than those that are hand-picked. Most yield estimates are very general and a figure often cited is 700 fruits per tree per year, each averaging 1–4 kg. An average of 200 fruits per tree, each weighing from 1 to 2 kg, has been estimated for the Caribbean. Yields for the South Pacific are generally given as 50–150 fruits per tree per year, although a study in Pohnpei recorded from 212 to 615 fruits per tree depending on the cultivar. Hectare yield estimates range from 16 to 50 tons based on 100 trees ha^{-1}. Mature breadnut trees in the Philippines have been reported to produce 600–800 fruits per season. The average number of seeds per fruit is variable, ranging from 32 to 94 per fruit, each seed weighing an average of 7.7–10 g. Based on 100 trees ha^{-1}, each average 200 fruits per tree, an average yield of 11 MT ha^{-1} of fresh seeds has been estimated.

Figure 6 Using long-handled poles to harvest breadfruit from a 20-year-old tree.

Composition and Use

Seeds

The seeds of breadfruit and breadnut have a thin, dark-brown outer skin about 0.5 mm thick and an inner, fragile paper-like membrane which surrounds the fleshy white edible portion of the seed. Seeds are firm, close-textured, and have a sweet pleasant taste that is most often compared to chestnuts. The small, immature fruits of breadnut are sliced and cooked as vegetables, seeds and all. Seeds are harvested from ripe fruits and boiled or roasted with salt. Breadfruit seeds are usually cooked with the raw breadfruit or are boiled or roasted. They generally are smaller than breadnut seeds and sometimes have a slightly bitter flavor.

The nutritional composition of breadnut and breadfruit seeds is shown in Table 1. The seeds, especially breadnuts, are a good source of protein and are low in fat, compared to nuts such as almond, brazil nut, and macadamia nut. The fat extracted from the seed is a light yellow, viscous liquid at room temperature with a characteristic odor similar to that of peanuts. It has a chemical number and physical properties similar to those of olive oil. Seeds are a good source of minerals and contain more niacin than most other nuts. Four amino acids (methionine 3.17 g, leucine 2.6 g, isoleucine 2.41 g,

and serine 2.08 g) comprised 50% of 14 amino acids analyzed.

Fresh Fruits

Breadfruit is a versatile food that can be cooked and eaten at all stages of maturity, although it is most commonly consumed when mature but still firm, and is used as a starchy staple. The relatively bland fruit can form the basis for an array of dishes, and it takes on the flavor of other ingredients in the dish. Very small fruits, 2–6 cm or larger in diameter, can be boiled and have a flavor similar to that of artichoke hearts. These can be pickled or marinated. Mature and almost mature breadfruit can be boiled and substituted for potatoes in many recipes. Ripe fruits are very sweet and can be eaten raw or used to make pies, cakes, and other desserts. In the Philippines, candied breadfruit is made by boiling slices in coconut and sugar.

Breadfruit's carbohydrate content is as good as or better than that of other widely used major carbohydrate foods. Compared to other staple starch crops, it is a better source of protein than cassava and is comparable to sweet potato and banana. It is a relatively good source of iron, calcium, potassium, riboflavin, and niacin. A comparison of nutrient composition of fresh and fermented mature fruits, and breadfruit flour is shown in Table 2. The nutritional composition of breadfruit varies among cultivars and should aid in the selection of cultivars for different uses for fresh consumption and processed products.

Fruit quality and attributes Fruit texture is an important attribute that affects cooking and processing. Seedless and few-seeded breadfruit both exhibit a wide range of textures at the mature stage. Preferred fruits are generally those that are dense, smooth, and creamy when cooked. There are cultivars with mealy flesh, as well as ones with fibrous, stringy flesh, and spongy ones which are full of what appear to be fine threads of latex. The quality of cooked fruit also depends on the method of preparation: different cultivars provide different results when boiled, roasted, or baked. Some cultivars are suitable for roasting but become mushy and fall apart when boiled. The potential for wide-scale processing by freezing, canning, or production of flour will be enhanced by selection of suitable cultivars. The presence or absence of seeds will affect how fruits are handled and processed. Fruits with seeds are probably inappropriate for large-scale canning or chip-making operations but are excellent for home use because they are a good source of protein and make breadfruit a more complete food.

Table 2 Proximate composition of breadfruit per 100 g edible portion

	Fresh[a,b,d]	*Flour*[c,e]	*Fermented*[a]
Water (%)	63.8–74.3	2.5–19.0	67.3–71.2
Protein (g)	0.7–3.8	2.9–5.0	0.7
Carbohydrate (g)	22.8–77.3	61.5–84.2	27.9
Fat (g)	0.26–2.36	1.93	1.13
Calcium (mg)	15.2–31.1	50	42.0
Potassium (mg)	352	1630	20–399
Phosphorus (mg)	34.4–79.0	90	
Iron (mg)	0.29–1.4	1.9	0.73–1.18
Sodium (mg)	7.1	2.8	
Thiamin (mg)	0.07–0.12		
Riboflavin (mg)	0.03–0.1	0.2	
Niacin (mg)	0.81–1.96	2.4	
Ascorbic acid (mg)	19.0–34.4	22.7	4–20
β-carotene	0.01		0.04–0.29

Data from [a]Aalbersberg WGL, Lovelace CEA, Madhoji K and Parkinson SV (1988) Davuke, the traditional Fijian method of pit preservation of staple carbohydrate foods. *Ecology of Food and Nutrition* 21: 173–180; [b]Dalessandri KM and Boor K (1994) World nutrition – the great breadfruit source. *Ecology of Food and Nutrition* 33: 131–134; [c]Graham HD and Negron de Bravo E (1981) Composition of the breadfruit. *Journal of Food Science* 46: 535–539; [d]Murai M, Pen P and Miller CD (1958) *Some Tropical Pacific Foods*. Honolulu: University of Hawaii; [e]Wootten M and Tumalii F (1984) Breadfruit production, utilisation and composition. A review. *Food Technology in Australia* 36: 464–65.

Postharvest Handling

A major limitation on the greater utilization of breadfruit is the seasonal nature of the crop – the trees typically bear fruits for just several months of the year – and the highly perishable nature of the fruit. Small-scale commercial processing of breadfruit involves canning slices in brine in Jamaica for local and export markets. Breadfruit flour and chips have been made on a limited basis, and the flour has been evaluated as a substitute for enriched wheat flour and as a base for instant baby food. The starch has been extracted and may find use in industrial applications such as textile manufacture.

The keeping quality of breadfruit is limited by a rapid postharvest rate of respiration, with the fruits typically ripening and softening in just 1–3 days after harvest. Breadfruit's perishability restricts local marketing and limits its export potential since fruits ripen before they reach their destination. Shelf-life can be extended by several days by careful harvesting and precooling fruits with chipped ice in the field and during transport. Refrigeration markedly increases shelf-life for both untreated fruits and ones stored in sealed polyethylene bags. However, fruits show symptoms of chilling injury at temperatures below 12 °C. Satisfactory fruit quality is best maintained at 12–16 °C with a shelf-life of 10 days for unwrapped fruits and 14 days for packaged fruits. Controlled-atmosphere storage has good possibilities. Fruits kept

at 16 °C in atmospheric containers of 5% carbon dioxide and 5% oxygen remained firm for 25 days with some browning of the skin.

Cooked breadfruit can be frozen, and this storage method deserves greater attention as it may provide a simple, effective means to utilize this crop better. Small sections of fruit which have been boiled for 2–5 min compare most favorably in flavor, color, and texture of fresh cooked breadfruit and can be kept at −15 °C for 10 weeks or more. Sections of fruit that are frozen without first being boiled discolor on cooking after storage and have poor flavor.

Since breadfruit is generally preferred while mature and still firm, nutritional studies, development of commercial products, and research to extend shelf-life have focused on this stage. Ripe fruits generally go to waste or are used as animal food, and there has been little attention given to expanding the use of ripe fruits. A greater proportion of the breadfruit crop could be utilized and marketed if food products incorporating ripe breadfruit, such as baby foods, baked goods, and desserts, are developed. Breadfruit is a reliable, easy-to-grow, nutritious staple food which has the potential to be more widely grown and used in the world's tropical regions. This can best be accomplished by selecting cultivars that extend the fruiting season and developing better methods of postharvest handling and simple, economical means of processing the fruits.

See also: **Fermented Foods**: Origins and Applications; **Figs**; **Fruits of Tropical Climates**: Commercial and Dietary Importance; **Preservation of Food**

Further Reading

Aalbersberg WGL, Lovelace CEA, Madhoji K and Parkinson SV (1988) Davuke, the traditional Fijian method of pit preservation of staple carbohydrate foods. *Ecology of Food and Nutrition* 21: 173–180.

Coronel RE (1983) Rimas and Kamansi. In: Coronel RE(ed.) *Promising Fruits of the Philippines*, pp. 379–398. Los Banos: College of Agriculture, University of the Philippines at Los Banos.

Dalessandri KM and Boor K (1994) World nutrition – the great breadfruit source. *Ecology of Food and Nutrition* 33: 131–134.

Graham HD and Negron de Bravo E (1981) Composition of the breadfruit. *Journal of Food Science* 46: 535–539.

Leakey CLA (1977) *Breadfruit Reconnaissance Study in the Caribbean Region*. CIAT/InterAmerican Development Bank.

McIntoch C and Manchew P (1993) The breadfruit in nutrition and health. *Tropical Fruits Newsletter* 6: 5–6.

Murai M, Pen P and Miller CD (1958) *Some Tropical Pacific Foods*. Honolulu: University of Hawaii.

Narasimhan P (1990) Breadfruit and jackfruit. pp. 193–259. In: Nagy S, Shaw PE and Wardowski WF (eds) *Fruits of Tropical and Subtropical Origin. Composition, Properties and Uses*, pp. 193–259. Lake Alfred, Florida: Florida Science Source.

Negron de Bravo E, Graham HD and Padovani M (1983) Composition of the breadnut (seeded breadfruit). *Caribbean Journal of Science* 19: 27–32.

Quijano J and Arango GJ (1979) The breadfruit from Colombia – a detailed chemical analysis. *Economic Botany* 33: 199–202.

Ragone D (1997) *Breadfruit*. Artocarpus altilis *(Parkinson) Fosberg. Promoting the Conservation and Use of Underutilized and Neglected Crops*. 10. Institute of Genetics and Crop Research, Gatersleben/International Plant Genetic Resources Institute, Rome, Italy.

Roberts-Nkrumah LB (1993) Breadfruit in the Caribbean: a bicentennial review. *Extension Newsletter*. Dept. of Agriculture, Univ. of the West Indies (Trinidad and Tobago). *Tropical Fruits Newsletter*. 24: 1–3.

Wootten M and Tumalii F (1984) Breadfruit production, utilisation and composition. A review. *Food Technology in Australia* 36: 464–465.

BREAKFAST AND PERFORMANCE

C Ani and S Grantham-McGregor, University College London, London, UK

Introduction

Breakfast is the first main meal of the day. As the name suggests, it is a meal taken after a long period of overnight 'fasting.' Starting the day without breakfast has been likened to embarking on a long-distance drive with an empty fuel tank. In this article, we will discuss what constitutes an adequate breakfast, and review the evidence for the effect of breakfast on children's cognitive performance and school achievement. We will also explore some of the possible mechanisms underlying these effects. Although there is no precise standard of what an adequate breakfast should contain, it is generally recommended that it should provide between

20 and 25% of the recommended dietary allowance of energy and other nutrients. In western countries, the widespread fortification of breakfast products (e.g., cereals) with vitamins and minerals means that eating an adequate breakfast usually increases the chances of meeting daily micronutrient requirements. Many school-aged children miss breakfast or have inadequate breakfast. Children from poor or disorganized families are at high risk of missing breakfast, as are children of working mothers. In developing countries, children may have to walk long distances to school as well as do household or farming work before going to school and thus use a considerable amount of energy before arriving at school. Therefore eating breakfast is likely to be particularly important.

Studies from several western countries show that skipping breakfast is common among adolescent girls. For example, 30% of 14–15-year-old girls in a Swedish study skipped breakfast. This trend has been blamed on efforts by young girls to fit 'desirable' body images portrayed by the media. Although the obvious goal for skipping breakfast is to limit calorie intake, paradoxically, skipping breakfast may actually increase overall daily calorie intake by increasing the likelihood of eating more at lunch, or relieving hunger with higher-fat snacks.

An important question is, does missing breakfast affect children's ability to achieve in school? Several studies have shown that children who usually miss breakfast have poorer cognition and school achievement than children who eat breakfast. However, missing breakfast is often associated with poverty and disorganized families, which may independently affect children's function. Therefore the best evidence of the effect of eating breakfast comes from experimental studies in which children are given breakfast or asked to miss it. We will first review evidence for breakfast affecting children's cognition, including attention, memory, and problem-solving.

Breakfast and Cognitive Performance

The short-term effects of having or missing breakfast have been examined in several studies conducted in schools. In these, children who had breakfast were compared with those who did not have breakfast on tests of cognitive performance. The results were inconsistent. A possible explanation is that the studies had no control over the children's prior diet and activity and there were no baseline measures before treatment to compare with after-treatment measures. The most rigorous experiments have been done in a laboratory setting, where the effect of missing or receiving breakfast was examined in the same children in a crossover design. Children were given cognitive tests on the mornings when they received breakfast and when they did not, and their performance was compared. The studies achieved tight control over the children's prior diet and activity by admitting them overnight to the laboratory before the experiments. The rigor of the studies was improved by random assignment of children to receive breakfast or placebo on the first occasion. Two such studies in the USA found that the children's performance on cognitive tests improved when they received breakfast. In two other studies in Jamaica and Peru, undernourished children's cognition deteriorated when they missed breakfast whereas well-nourished children were not affected.

It appears that the quality of breakfast is also important. One study from Sweden using a crossover design found that children who received adequate breakfast (e.g., cereal, yogurt, sandwiches, and milk) performed better in tests of creativity and addition as well as physical exercise compared with those who ate inadequate breakfast (e.g., sweet rolls and sweet drink).

In a Jamaican study conducted in schools, children were given breakfast at school regardless of whether they had it at home. The study had a crossover design and all children were tested after 1–2 weeks of receiving breakfast and after not receiving breakfast. In spite of loss of control over previous diet and activity the findings were similar to the previous Jamaican study. The undernourished children showed benefits in their performance on cognitive tests on the mornings when they received breakfast.

In summary, the evidence from the best-controlled studies indicates that missing breakfast or having an inadequate breakfast affects children's cognitive performance especially if they are already undernourished.

How Can Breakfast Affect Cognitive Performance?

The cognitive effects of breakfast are likely to be mediated by several factors. The relief of hunger may play a role. Children who miss breakfast may suffer short-term hunger during midmorning school work. Hungry children may be easily distracted and undermotivated and thus not attend to learning tasks. Eating breakfast can relieve these adverse effects of hunger on children, thereby improving the children's ability to learn.

Second, the nutritional components of breakfast may have direct metabolic benefits on cognition. In particular, glucose or its precursors is thought to be important. Several studies in elderly subjects, young adults, and children have found that glucose drinks

improve memory and concentration. Reaction times and driving ability have also been improved with glucose drinks. In one study, memory in people who ate breakfast correlated with their levels of blood glucose. Another study showed that a glucose drink improved the memory of those who missed breakfast to the same level as those who had eaten breakfast. However, a glucose drink did not improve memory any further in those who had eaten breakfast. This finding suggests that beyond a certain threshold of blood glucose level, higher levels of intake confer no extra cognitive advantage. Interestingly, glucose did not improve all cognitive functions in subjects who missed breakfast and recall of a story was not improved with glucose. The findings suggest that, while glucose may explain some of the cognitive benefits of breakfast, it does not account for all of them.

In order for glucose to influence mental performance, there should be a biochemical mechanism linking it with mental function. Such a link is thought to exist through the role of glucose in the biosynthesis of acetylcholine, a neurotransmitter associated with memory. Acetylcholine is synthesized from choline and acetyl-coenzyme A (acetyl-CoA), and glucose is the major substrate for the synthesis of acetyl-CoA. It has been shown that increased glucose levels can enhance the synthesis of acetylcholine and that giving glucose can ameliorate falling levels of the neurotransmitter. Animal studies have also shown that giving glucose ameliorates the loss of memory produced by drugs that inhibit acetylcholine. While glucose plays a role in some of the short-term cognitive benefits of eating breakfast, other nutrients (e.g., protein, energy, iron, iodine, etc.) may contribute to mental performance in the medium to long term.

Millions of children in developing countries are underweight and short for their age (stunted). These children generally have poorer cognitive function than adequately nourished children. Several studies of children under 3 years have shown that nutritional supplementation improves concurrent mental development and in some cases long-term development. We are unaware of long-term supplementation studies beginning at an older age. Micronutrient deficiencies, including iron, iodine, and zinc, have also been implicated in causing poor cognition in school-aged children. So providing an adequate breakfast containing protein, energy, and micronutrients over a long period of time should improve undernourished children's nutritional status and thereby improve their cognitive development.

Breakfast and School Performance

School achievement depends on three main factors, including the children's innate cognitive ability and biological state, school characteristics such as the quality of teaching and availability of reading materials, and home factors such as parental educational level and their attitude to school. The amount of time children spend engaged on a task is a critical determinant of whether they learn the task. Time on task depends not only on the amount of time actual teaching occurs but also on the children's attendance and, when present, if they are actively engaged in a task or distracted. As shown in Figure 1, there are several ways breakfast (especially when provided in schools) could improve school performance. First, as discussed above, eating breakfast is likely to prevent midmorning hunger and improve children's cognitive ability. In addition, it may improve classroom behavior,

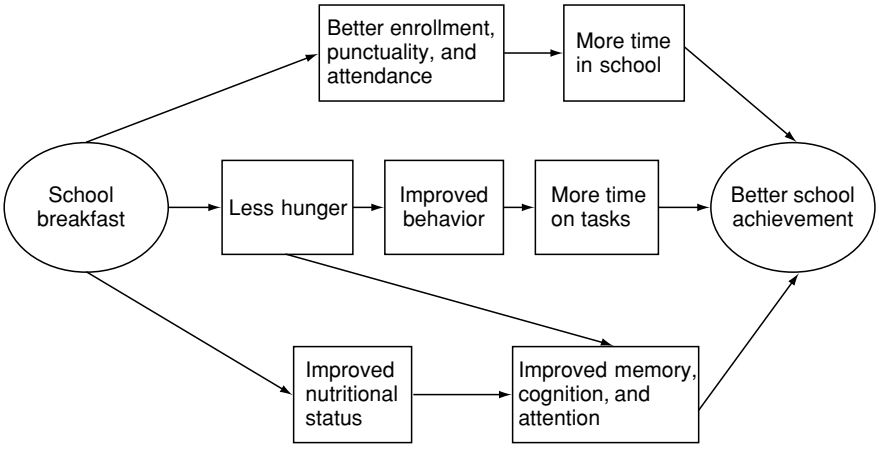

Figure 1 Mechanisms for the effects of breakfast on school performance. Adapted with permission from Ani C and Grantham-McGregor SM (1999) The effects of breakfast on educational performance, attendance and classroom behavior. In: Donovan N and Street C (eds) *Fit for School. How Breakfast Clubs Meet Health, Education and Childcare Needs*, pp. 14–22. London: New Policy Institute.

including reducing disruptive behavior and improving concentration. The opportunity to eat breakfast at school may also encourage children to attend school more frequently, particularly in situations where food is short at home. The children may also be more punctual if they want to eat breakfast before school begins. All these factors may increase the time the children spend actively engaged on learning tasks.

In addition, where undernutrition is prevalent, children who eat adequate breakfasts regularly are more likely to become well nourished and to perform better in school. We will now review studies examining the relationship between breakfast, classroom behavior, and school achievement.

Classroom Behavior

Although early studies reported improvements in children's classroom behavior when they were given breakfast, they were generally poorly designed, making interpretation difficult. The measurements were often subjective and depended on reports from teachers, who were aware of whether the children had been given breakfast. There are a few more recent and better-designed studies. These studies have shown when children received breakfast they were off-task and out of their seats less often, and participated more actively in teaching activities and showed better peer interaction. However, the situation is complicated and, in a Jamaican study, children's behavior in response to receiving a school breakfast varied according to the atmosphere and organization of the classroom. In well-organized classrooms their behavior improved with breakfast but in poorly organized, overcrowded classrooms their behavior actually deteriorated and they were out of their seats more often and less on-task.

One study showed that children who were provided with a glucose drink were more attentive and showed fewer signs of frustration when given a difficult task.

School Attendance

In the USA, it has been shown that students participating in the Department of Agriculture School Breakfast Program were more likely to attend and less likely to be late to school compared with nonparticipants. Other studies in poorer countries like Peru and Jamaica have also consistently found improved attendance among participants in school breakfast programs.

School Achievement

There are few well-designed evaluations of school breakfast programs; most have failed to have control groups or pretreatment measures. Low-income students who participated in a school breakfast program in the USA were found to perform better in tests of mathematics, reading, and language compared with nonparticipating students from similar backgrounds. In a randomized controlled trial in Jamaica, students given breakfast over two school terms performed better in arithmetic compared with those who received a low-calorie drink. Another study in Peru showed that the vocabulary of nutritionally at-risk students improved when given school breakfast for as little as a month, compared with similarly disadvantaged but nonparticipating children. However, the effects of breakfast on school achievement have not been studied more than one school year, so it is unknown whether the improvements continue.

Overall Conclusions

There is good evidence that missing breakfast detrimentally affects children's cognitive function, especially if they are undernourished. Glucose benefits cognition in children who have missed breakfast and there is some suggestion that the short-term effects of breakfast on cognition may be partly but not wholly explained by the metabolic effect of glucose. In undernourished populations, long-term benefits could be mediated by improvements in nutritional status.

There is also good evidence that if breakfast is provided at school in high-risk populations, the children's attendance and punctuality improve. There is some evidence that classroom behavior and school achievement also improve. However, there is a need for more rigorous studies in both developing and developed countries and the long-term benefits need to be evaluated. In view of the demonstrable benefits of breakfast to children's performance, policy-makers should consider implementing school breakfast programs in high-risk populations.

See also: **Carbohydrates**: Requirements and Dietary Importance; **Eating Habits**; **Glucose**: Function and Metabolism; **Hypoglycemia (Hypoglycaemia)**; **Malnutrition**: The Problem of Malnutrition

Further Reading

Ani C and Grantham-McGregor S (1999) The effects of breakfast on educational performance, attendance and classroom behavior. In: Donovan N and Street C (eds) *Fit for School: How Breakfast Clubs meet Health, Education and Childcare Needs*, pp. 14–22. London: New Policy Institute.

Benton D and Parker PY (1998) Breakfast, blood glucose, and cognition. *American Journal of Clinical Nutrition* 67: 772S–778S.

Kanarek R (1997) Psychological effects of snacks and altered meal frequency. *British Journal of Nutrition* 77: S105–S120.

Kennedy E and Davis C (1998) US Department of Agriculture school breakfast program. *American Journal of Clinical Nutrition* 67: 798S–803S.

Myers A, Sampson A, Weitzman M, Rogers B and Kayne H (1989) School breakfast program and school performance. *American Journal of Diseases of Childhood* 143: 1234–1239.

Pollitt E (1990) *Malnutrition and Infection in the Classroom*. Paris: UNESCO.

Pollitt E, Cueto S and Jacoby ER (1998) Fasting and cognition in well and undernourished school children: a review of three experimental studies. *American Journal of Clinical Nutrition* 67: 779S–784S.

Powell C, Walker S, Chang S and Grantham-McGregor S (1998) Nutrition and education: a randomized trial of the effects of breakfast in rural primary school children. *American Journal of Clinical Nutrition* 68: 873–879.

Richter L, Rose C and Griesel R (1997) Cognitive and behavioral effects of a school breakfast. *South African Medical Journal* 87: 93–100.

Wyon D, Abrahamsson L, Jartelius M and Fletcher R (1997) An experimental study of the effects of energy intake at breakfast on the test performance of 10-year-old children in school. *International Journal of Food Sciences and Nutrition* 48: 5–12.

Breast-feeding *See* **Infants**: Nutritional Requirements; Breast- and Bottle-feeding; Weaning; Feeding Problems; **Lactation**: Human Milk: Composition and Nutritional Value; Physiology

BROWNING

Contents

Nonenzymatic

Jennifer M Ames, University of Reading, Reading, UK

Introduction

The term 'nonenzymatic browning' refers to the chemical reactions that result in the formation of brown color when food is heated. In contrast to enzymatic browning, no enzymes are involved in nonenzymatic browning reactions. The most important nonenzymatic browning reaction is the Maillard reaction, which encompasses the cascade of reactions that occur when reducing sugars are heated with compounds possessing a free amino group (e.g., amino acids, amines, and proteins) and which result in thousands of reaction intermediates and products. The Maillard reaction is named after the French scientist, Louis Camille Maillard, who first investigated reducing sugar–amino acid interactions in 1912. Other nonenzymatic browning reactions include Maillard-type reactions between amino compounds and other compounds possessing a free carbonyl group, e.g., ascorbic acid and lipid oxidation products. A further type of nonenzymatic browning is caramelization, which occurs when sugars discolor in the absence of amino compounds. In all types of nonenzymatic browning mentioned, the chemistry involved is very similar. In this article (unless stated otherwise), the terms 'nonenzymatic browning' and the 'Maillard reaction' are used interchangeably and refer to reactions between reducing sugars and amino compounds.

As well as giving brown colors, the Maillard reaction results in various other outcomes in foods. These include development of flavor, increase in antioxidative capacity, loss of nutritional and functional properties, and formation of compounds with potential adverse effects on human health. Since the Maillard

reaction has so many important consequences for food quality, it is without doubt one of the most significant reactions that occur in food. Various factors affect the course of the Maillard reaction. These include the nature of the reactants, temperature, and time of heating, the pH and water activity (a_w) of the food and the presence of reaction inhibitors, e.g., sulfite. Due to the importance of the reaction, its control is important for food quality and kinetic approaches to modeling and control are showing considerable promise.

Chemistry

A summary of the chemical pathways comprising the Maillard reaction is given in Figure 1. The early stage of the reaction is well understood. It involves the condensation of the reducing sugar with the amino compound to give an *N*-substituted glycosylamine which rearranges to the 1-amino-1-deoxy-2-ketose, also known as the Amadori rearrangement product (ARP). The relative importance of the subsequent pathways largely depends on the pH of the food. At a pH below 5, the route involving 3-deoxyosone as the intermediate prevails but, if the pH is above 7, the pathways involving the 1- and 4-deoxyosones and the 1-amino-1,4-dideoxyosone are more important. Since most foods are in the pH range 4–6, the 3-deoxyosone route usually predominates. The 3-deoxyosone reacts with amino acids to give melanoidins (the final products of the Maillard reaction which are macromolecular and colored), Strecker aldehydes via the Strecker degradation (Figure 2) and pyrroles and furfurals and their derivatives. Enaminols, which are products of the Strecker degradation, subsequently condense to give pyrazines (Figure 2). The other deoxyosones are much less stable than the 3-deoxyosone. They are converted into reductones which may react with other intermediates to give melanoidins. In addition, these deoxyosones (as well as other Maillard intermediates, e.g., the *N*-substituted glycosylamine) readily fragment via retro-aldolization. The low-molecular-weight products that result are highly reactive and may participate

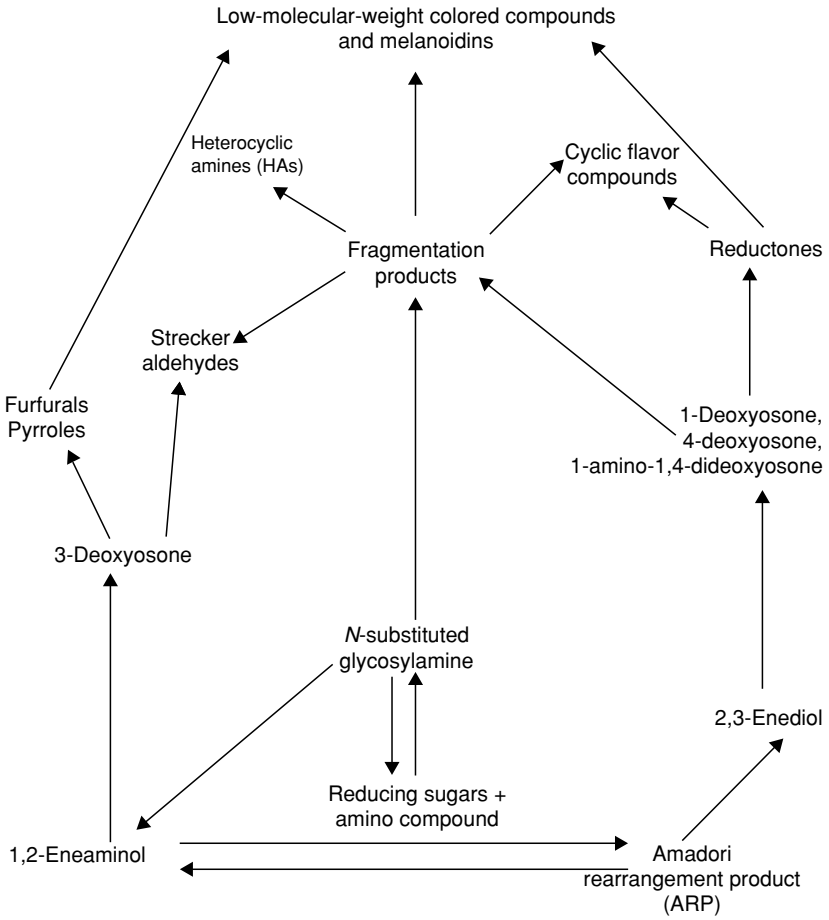

Figure 1 Outline of the chemical pathways of the Maillard reaction.

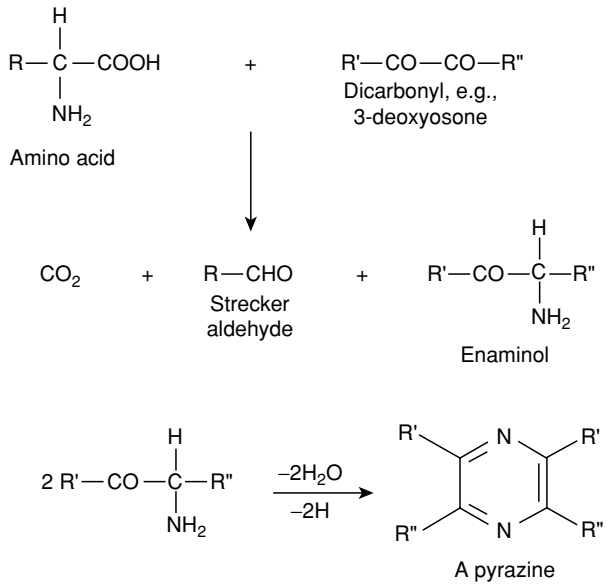

Figure 2 Strecker degradation and pyrazine formation.

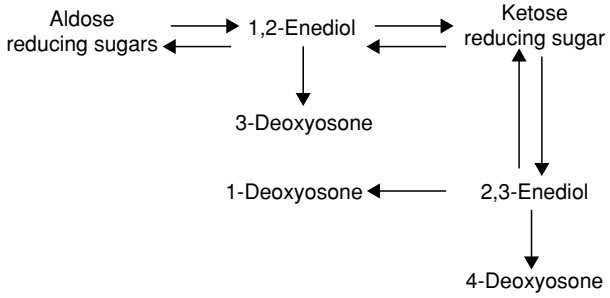

Figure 3 Degradation of reducing sugars at temperatures above 130 °C, or at pH values below 3 or above 8, and in the absence of amino compounds.

in the Strecker degradation, condensation reactions resulting in cyclic products, and reactions leading to melanoidins.

At temperatures above 130 °C or at pH values of below 3 or above 8, sugars degrade in the absence of amino acids to give key intermediates that also form in the Maillard reaction (Figure 1), as shown in Figure 3.

Flavor

Hundreds of flavor compounds formed by the Maillard reaction have been identified. The compounds concerned are largely those that contribute to aroma (i.e., they are volatile and are thus perceived by the nose), although a few components that possess specific tastes (and are therefore perceived by the tongue) have also been reported. Roasted coffee, cooked meat, bread, cookies, breakfast cereals, roasted nuts, and baked potatoes are among the foods that owe their particular flavor to the Maillard reaction. Also, the Maillard reaction is employed by the flavor industry for the production of process or reaction flavors. In addition to the development of desirable flavors, the Maillard reaction is responsible for off-flavors in certain foods, particularly in commodities in which the original fresh flavor of the product is desired. Thus, it can lead to cooked or stale flavors in processed orange juice or dried milk powder. (*See* **Flavor (Flavour) Compounds: Structures and Characteristics.**)

Table 1 Volatile compounds derived from the sugar, the amino acid, and sugar/amino acid interactions

Compounds derived from the sugar	*Compounds derived from sugar–amino acid interactions*	*Compounds derived from the amino acid*
Furans	Pyrroles	Sulfur compounds
Pyrones	Pyridines	Aldehydes
Cyclopentenes	Pyrazines	Thiazoles
Carbonyls	Oxazoles	Cyclic polysulfur compounds
Acids	Thiophenes	

The volatile Maillard flavor compounds may be divided into three groups: those formed from the sugar, those formed from the amino acid, and those requiring both the sugar and the amino acid for their formation (Table 1). Strictly speaking, it is only the compounds in the latter group that are formed by the Maillard reaction, i.e., reducing sugar–amino compound interactions. Members of a further category of aroma compounds are formed by reactions involving intermediates of both the Maillard reaction and lipid degradation. These components often possess fried odors and contribute fried aromas to foods such as potato chips. Much less is known about the products of the Maillard reaction that may contribute to taste. However, it has been established that some pyrrole derivatives contribute to the bitter taste of roasted coffee and malt. The structures of some specific compounds, together with their odor (or taste) threshold values and contributions to food flavor, are given in Table 2.

Color

The compounds formed early on in the Maillard reaction are colorless and absorb in the ultraviolet (UV) region of the electromagnetic spectrum. As the reaction progresses, absorption in the UV region intensifies and begins to tail into the visible region, resulting in the appearance of yellow, orange, and brown

Table 2 Properties of selected flavor compounds

Compound	*Odor threshold value ($\mu g\,kg^{-1}$)*	*Odor*
Butanedione (diacetyl) $CH_3COCOCH_3$	7	Buttery
3-(Methylthio)propanal (methional) $CH_3SCH_2CH_2CHO$	200	Cooked potatoes
2,5-Dimethyl-4-hydroxy-3(2*H*)furanone (furaneol) 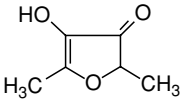	0.04	Caramel, burnt pineapple, strawberries
Methylpyrazine 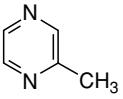	6×10^4	Nutty
2-Acetyl-1-pyrroline 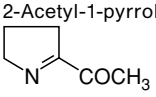	0.1	Freshly baked wheat bread crust
2-Acetyl-1,4,5,6-tetrahydropyridine 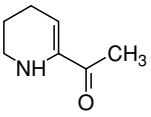	1.6	Crackers
Trimethyloxazole H_3C 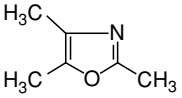	5	Green, nutty, sweet
2-Acetylthiophene 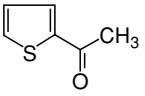	0.1	Onion-link, mustard-like
bis(2-Methyl-3-furyl) disulfide 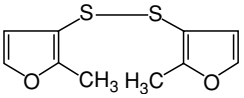	2×10^{-5}	Cooked meat
2-Actylthiazole 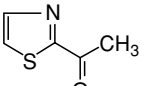	10	Nutty, popcorn
3-Methyl-5-pentyl-1,2,4-trithiolane 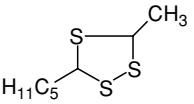		Fried chicken
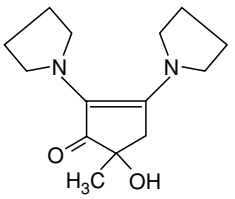	5–10 $\times$ 10[3][a]	Bitter

[a]Taste threshold.

colors. The Maillard reaction is responsible for the desirable yellow or brown colors in foods, including bread crust, potato chips, popcorn, cookies, and other flour confectionery, but it can also lead to discoloration, e.g., in heated fruit juices and milk powder.

The colored Maillard reaction products may be divided into two groups, i.e., the low-molecular-weight compounds that possess up to about five linked rings, and the macromolecular melanoidins. Most work on colored Maillard reaction products has concerned the low-molecular-weight compounds isolated from model food systems. A selection of compounds isolated from aqueous sugar–amino acid systems, heated with or without additional furfural, is shown in Figure 4. Progress in the field of low-molecular-weight colored Maillard products has advanced rapidly over the last 5 years. Nevertheless, much remains to be done.

No homogeneous melanoidin has been isolated from a food or model food (sugar–amino acid) system to date and no structures for melanoidins isolated from food are available. Progress concerning the structures of high-molecular-weight melanoidin-like structures has been made by investigating the polymers produced on polymerization of first, 2-hydroxymethyl-*N*-methylpyrrole in trichloromethane and, second, *N*-methylpyrrole with 2-formyl-*N*-methylpyrrole (or furfural) in methanol. Figure 5 gives an example of the structures identified by $^{1}H/^{13}C$ nuclear magnetic resonance spectroscopy, fast atom bombardment–mass spectrometry and matrix-assisted desorption ionization–time of flight–mass spectrometry. Although the formation of such regular homopolymers in food systems is unlikely, it is reasonable to expect that the type of structure depicted in Figure 5 represents a domain of a food melanoidin.

It has been demonstrated that heating food proteins (casein) in the presence of glucose results in protein cross-linking and a parallel increase in color

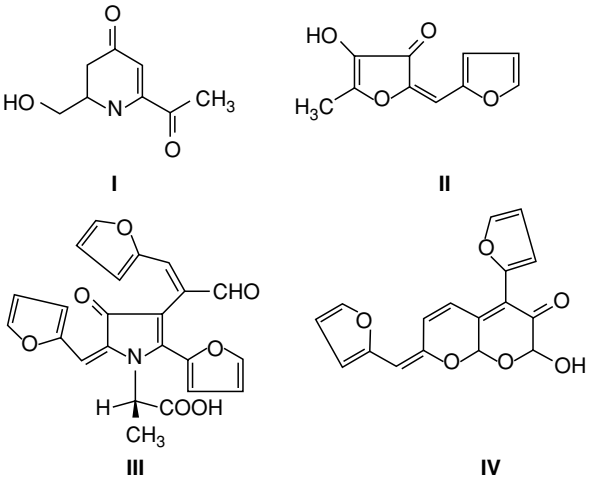

Figure 4 Selected low-molecular-weight colored compounds. **I**, from xylose and glycine; **II**, from xylose and glycine or xylose and lysine; **III** and **IV**, from xylose, alanine and furfural. (**I**: 2-Acetyl-6-(hydroxymethyl)-5,6-dihydro-4*H*-pyridinone; **II**: 2-furfurylidene-4-hydroxy-5-methyl-3(2*H*)-furanone; **III**: 4-[1-formyl-2-(2-furyl)ethenyl)]-5-(2-furyl)-2-[(2-furyl)methylidene]-2,3-dihydro-α-methyl-3-oxo-1*H*-pyrrole-1-acetic acid; **IV**: 4-(2-furyl)-7-[(2-furyl)-methylidene]-2-hydroxy-2*H*,7*H*,8a*H*-pyrano[2,3b]pyran-3-one.)

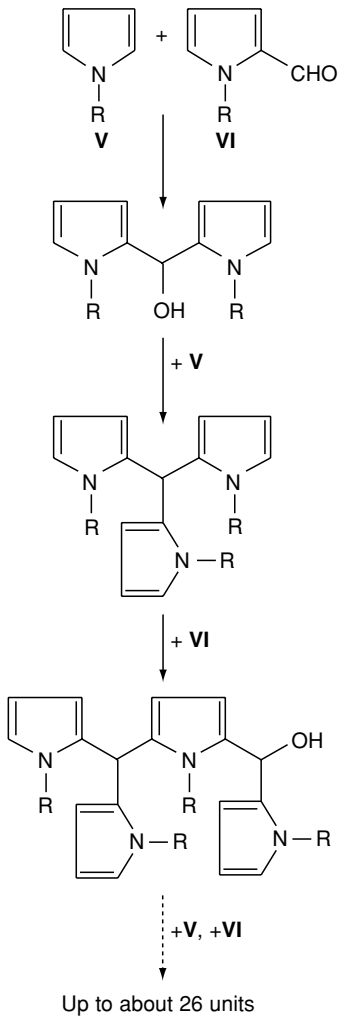

Figure 5 Structure of the polymer prepared from *N*-methylpyrrole and 2-formyl-*N*- methylpyrrole.

intensity of the reaction products. This color development is due to the incorporation of colored Maillard reaction products into the cross-linked protein. Therefore it is likely that, in foods, a portion of the brown color formed on heating is due to low-molecular-weight colored Maillard products binding to colorless protein backbones giving protein cross-linking and browning.

Antioxidant Activity

The Maillard reaction may influence the oxidative stability of foods because some Maillard reaction products are able to retard lipid oxidation. An increase in stability may be achieved by modifying the heat process applied, e.g., to milk, adding Maillard reaction precursors, e.g., to cookie dough, prior to heat treatment, or addition of Maillard reaction products to susceptible foods, e.g., margarine.

An increase in pH and the amino compound–sugar molar ratio enhances the formation of nondialyzable melanoidins and the inclusion of nitrogen in the Maillard reaction products. There is a positive correlation between all these factors and antioxidant activity. Nondialyzable melanoidins can scavenge hydroxyl radicals and other active oxygen species. Antioxidant activity increases with formation of melanoidins up to a maximum and then decreases with further heating and browning. Reducing activity and metal chelating ability are factors that may account for antioxidant activity. A strong candidate for such activity is the reductone group. However, enediol-like reductones and amino reductones (examples of the latter being those formed by the Amadori rearrangement) may act as antioxidants or prooxidants, depending on the reaction conditions, and the structures responsible for antioxidant activity remain uncertain. (*See* **Antioxidants**: Natural Antioxidants; Role of Antioxidant Nutrients in Defense Systems.)

Nutritional Effects

The Maillard reaction can result in serious loss of availability of essential amino acids, both in the free form and in peptide chains. Loss of the available amino acid occurs as soon as the ARP is formed. In most foods, lysine destruction is most significant, due to its ε-amino group. Ascorbic acid and also vitamins possessing a free amino group, e.g., thiamine, can participate in Maillard-type reactions leading to loss in activity of the vitamin concerned. Various Maillard reaction products may complex nutritionally important metals, e.g., copper, zinc, iron, thus making them unavailable. ARPs or more advanced reaction products fed orally to humans increase zinc excretion via urine. The effect is more marked in patients fed intravenously.

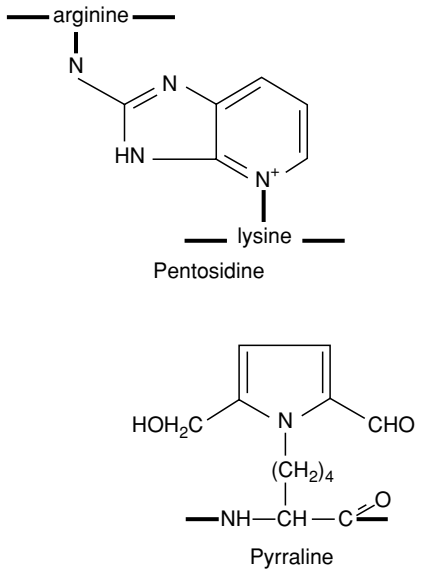

Figure 6 Structures of pyrraline and pentosidine. The peptide backbone is represented by a solid line.

The posttranslational modification of proteins and the development of protein cross-links as a consequence of the Maillard reaction reduce protein digestibility (as well as functionality). Pyrraline (identified in heated milks) and pentosidine (identified in roasted coffee) are examples of such modifications and their structures are shown in Figure 6. Reduction in protein digestibility may be due to decreased access by digestive enzymes to the modified protein or by the direct inhibition of the enzymes. The main effect is to block normal tryptic hydrolysis on the carboxyl side of lysine.

Mutagen Formation

More than 20 heterocyclic amines (HAs) that are products of the Maillard reaction have been identified at the p.p.b. level in cooked beef, pork, lamb, chicken, and fish muscle. Amino acids, sugars, and biochemicals present in muscle tissue, e.g., creatinine and creatine, are all required for their formation. The structures of a selection of HAs are shown in Figure 7. The amounts produced increase with temperature and time of cooking; temperatures encountered during, for example, grilling and frying favor their formation. Most HAs cause mutations in bacteria, mammalian cells and cancer in animals. They are the most potent mutagens examined by the Ames test and are able to produce frameshift mutations in

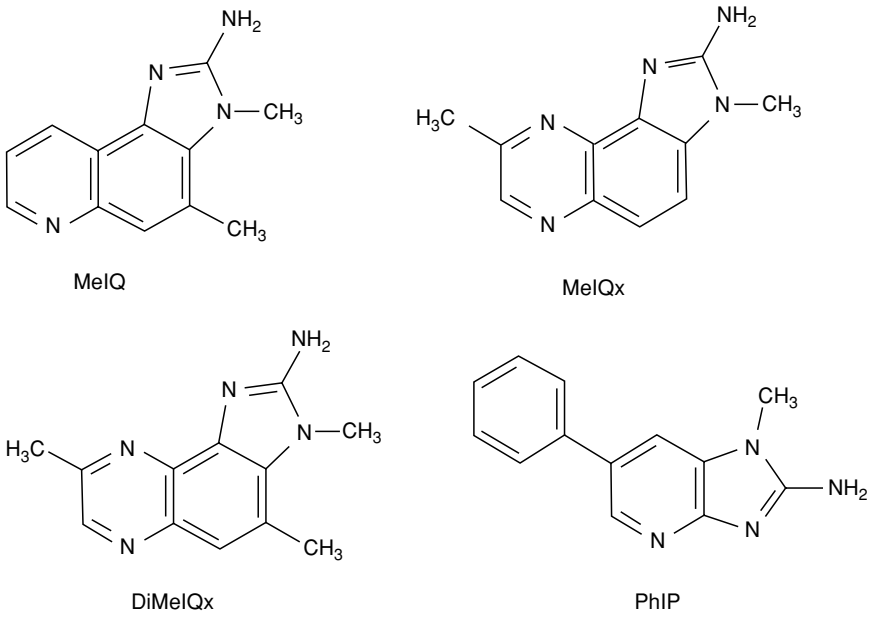

Figure 7 Structures of selected heterocyclic amines. MeIQ, 2-amino-3,4-dimethylimidazo[4,5-*f*]-quinoline; MeIQx, 2-amino-3, 8-dimethylimidazo[4,5-*f*]quinoxaline; DiMeIQx, 2-amino-3,4,8-trimethylimidazo[4,5-*f*]quinoxaline; PhIP, 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine.

Salmonella and base pair substitutions in hamster ovary cells. The risk to humans of consuming foods containing measurable amounts of HAs is difficult to estimate and requires epidemiological studies. Estimates based on animal cancer data and surveys of human food intake suggest that the average exposure to cooked food containing HAs gives a 10^{-4} risk of lifetime cancer. This average value is subject to substantial variation due to differences in diet, cooking methods, and genetic factors.

Although HAs are the mutagens formed during the Maillard reaction that have received most attention, other Maillard products are also reported to be mutagenic, including methylglyoxal and 5-hydroxymethylfurfural (HMF). (*See* **Mutagens**.)

Factors Affecting the Maillard Reaction

The extent and course of the Maillard reaction are affected by the profile of components in the food capable of acting as reaction precursors, temperature and time of heating, pH, and a_w. The reaction is also affected by metals, oxygen, and the presence of inhibitory agents, e.g., sulfite. When attempting to control the Maillard reaction, the aim may be to stop it or to steer it towards a particular outcome in terms of flavor or antioxidant activity. In any case, the reaction is difficult to stop once it has started.

The Maillard reaction may be minimized by removing one of the required reaction partners, i.e., the reducing sugars for protein-rich foods or the amino compounds for carbohydrate-rich foods. In jam-making, limiting the cooking time after addition of sucrose reduces its inversion to glucose and fructose in the acidic medium. An alternative to removing the reducing sugars or amino compounds is to process separately the carbohydrate and protein components of a food, e.g., cream-style soups.

Pentose sugars, e.g., ribose, react faster than hexose sugars, e.g., glucose and fructose, which in turn react faster than reducing sugars that are disaccharides, e.g., maltose and lactose. Amino acids also react at different rates. Of the free amino acids, when color formation is measured, lysine and glycine are among the fastest while cysteine is the slowest. In peptide chains, the amino acid residues with reactive side chains, e.g., lysine and arginine, react the fastest. The amino acid has a large effect on the flavor formed. Cysteine gives meaty notes, methionine results in a potato aroma, and cracker and bready odors are produced by proline.

The Maillard reaction increases with temperature. Q_{10} values range from 2 to 8 according to the system and parameter being measured. Therefore, if the aim is to stop the reaction, the temperature should be reduced to as low a level as possible. The profile of reaction products formed varies according to the temperature and time of heating and it is not possible to produce the profile obtained at a high temperature by heating at a lower temperature for a longer time. This

can be illustrated by comparing the aroma of meat resulting from boiling at 100 °C and from roasting at 200 °C. The two aromas are easily distinguished because different reaction pathways, and thus different aroma compounds, are favored by the different cooking methods.

The pH of the system plays a key role in determining the relative importance of the reaction pathways followed (see section on Chemistry, above). At pH values below about 5, the 3-deoxyosone route is favored while routes involving the other deoxyosones are favored above pH 7. In fact, there is no sharp cut-off point. All the pathways operate at all pH values of relevance to food, the ratio of products being influenced by the pH. Thus, pH is a parameter that can be manipulated to achieve the required profile of reaction products.

Since water is a product of the Maillard reaction, the reaction occurs less readily in high-a_w foods. In low-a_w foods, the reactants are less mobile and this also impedes the reaction. In practice, the Maillard reaction is favored by intermediate a_w values (0.5–0.8). When the aim is to prevent the Maillard reaction, a_w is of great concern during the drying or concentration of food. In such cases, the food must be taken through the intermediate a_w stage as quickly as possible, especially if heat is used. The formation of individual Maillard reaction products is favored by different a_w values and therefore a_w represents another means of optimizing the profile of reaction products formed.

Since the Maillard reaction is so difficult to stop once it has begun, the food industry may resort to the use of sulfite (normally a solution of sodium bisulfite or sodium metabisulfite) to control it. The sulfite combines with reducing sugars and Maillard intermediates possessing carbonyl and dicarbonyl groups to inhibit the reaction.

Kinetics and Modeling

The Maillard reaction is much more complex than most reactions that have been modeled by physical chemists. However, since the reaction is of such interest to the food industry, attempts have been made to describe it using kinetics and mathematical equations, in order to make valid predictions concerning color, flavor, and antioxidant activity for various food compositions processed and stored under different conditions. As a result, it is possible to predict accurately the color development (absorbance at a single wavelength of a model system) at a particular time. Rate constants have also been published for a limited selection of Maillard aroma compounds. However, since the Maillard reaction is qualitatively and quantitatively affected by many factors, the direct application of models may be of very limited use. There is a great need for further application of kinetics principles to the Maillard reaction so that various aspects of the reaction may be predicted and, ultimately, controlled in foods.

See also: **Antioxidants**: Role of Antioxidant Nutrients in Defense Systems; **Browning**: Toxicology of Nonenzymatic Browning; **Casein and Caseinates**: Uses in the Food Industry; **Flavor (Flavour) Compounds**: Structures and Characteristics; **pH – Principles and Measurement**

Further Reading

Ames JM (1992) The Maillard reaction. In: Hudson BJF (ed.) *Progress in Food Proteins – Biochemistry*, pp. 99–153. London: Elsevier Applied Science.

Ames JM and Hofmann T (eds) (2000) *Chemistry and Physiology of Food Colorants*. American Chemical Society Symposium Series 775. Washington, DC: ACS.

Eriksson C (ed.) (1981) *Maillard Reactions in Food*. Oxford: Pergamon Press.

Finot PA, Aeschbacher HU, Hurrell RF and Liardon R (eds) (1990) *The Maillard Reaction in Food Processing, Human Nutrition and Physiology*. Basel: Birkhäuser.

Friedman M (1996) Food browning and its prevention: an overview. *Journal of Agricultural and Food Chemistry* 44: 631–653.

Fujimaki M, Kato H and Namiki M (eds) (1986) *Amino-Carbonyl Reactions in Food and Biological Systems*. Amsterdam: Elsevier.

Hodge JE (1953) Dehydrated foods: chemistry of browning reactions in model systems. *Journal of Agricultural and Food Chemistry* 1: 928–943.

Ikan R (ed.) (1996) *The Maillard Reaction. Consequences for the Chemical and Life Sciences*. Chichester: John Wiley.

Ledl F and Schleicher E (1990) New aspects of the Maillard reaction in foods and in the human body. *Angewandte Chemie International Edition in English* 29: 565–706.

Leong LP and Wedzicha BL (2000) A critical appraisal of the kinetic model for the Maillard browning of glucose and glycine. *Food Chemistry* 68: 21–28.

Namiki M (1988) Chemistry of Maillard reactions: recent studies on the browning reaction mechanism and the development of antioxidants and mutagens. *Advances in Food Research* 32: 115–184.

O'Brien J and Morrisey PA (1989) Nutritional and toxicological aspects of the Maillard browning reaction in foods. *Critical Reviews in Food Science and Nutrition* 28: 211–248.

O'Brien J Nursten HE, Crabbe MJC and Ames JM (eds) (1998) *Maillard Reaction in Foods and Medicine*. Cambridge: Royal Society of Chemistry.

Waller GR and Feather MS (eds) (1983) *The Maillard Reaction in Foods and Nutrition*. ACS Symposium Series 215. Washington, DC: ACS.

BROWNING

Toxicology of Nonenzymatic Browning

P A Finot, Nestec, Nestlé Research Centre, Lausanne, Switzerland

Introduction

Reactions between reducing sugars and free amino groups in foods, without any catalytic involvement of enzymes, lead to nonenzymatic browning (Maillard reaction) causing a reduction in nutritive value and certain physiological and/or toxicological effects. The browning reaction develops during both home cooking and industrial processing of foods. Whilst contributing to an improvement of the organoleptic properties of foods, through aroma development, browning is often an undesirable side-effect of obligatory heat treatments applied for microbiological (sterilization and drying) or nutritional (cooking) reasons and for convenience (storage). Since the Maillard reaction occurs so frequently and Maillard reaction products are present in practically all meals, an understanding of its biological implications is of importance.

Because of the multiplicity of the food systems, the complexity of the chemical reactions, and the large variety of heat treatments involved, any generalization on the biological outcome of nonenzymatic browning of foods is not easy. Nevertheless, a number of studies of model amino acid/protein and sugar reaction systems permits a classification of the biological outcome as: (1) nutritional; (2) physiological; and (3) toxic (including genotoxic) effects.

Nutritional Effects

The nutritional effects of the Maillard reaction in foods are due both to the chemical modification of essential nutrients which thereby become unavailable (direct effects) and to the presence of Maillard products which reduce the bioavailability and disturb the metabolism of other nutrients (indirect effect).

Since it proceeds through many chemical routes and produces a large number of chemical species, the Maillard reaction has been divided for clarity and convenience into two distinct stages, early and advanced, which are associated with different nutritional and physiological consequences (Figure 1).

The Early Maillard Reaction

This first stage involves the reaction between a free amino group and a reducing sugar to form, through the Amadori rearrangement, a stable deoxyketose addition compound, also called the Amadori compound. This is the obligatory step for the continuation of the reaction to the advanced stage.

At this Amadori stage, nutritional damage has already been done. No enzyme in animal tissues can split these complexes to regenerate the amino compounds, which are nutritionally unavailable. In the case of amino acids, rat growth and metabolic studies have shown that the Amadori compounds are biologically unavailable. Cecal and large intestinal microorganisms in experimental animals can liberate the complexed amino acids, but this occurs too late in the digestive/absorptive process for a significant *in vivo* utilization. The nutritional loss includes those free amino acids and peptides that have reactive α-amino groups, protein-bound lysine which has a reactive ε-amino group, and vitamins (thiamin, pyridoxine, and folic acid).

In a model system representative of the early stage, comprising milk powder stored at 60 °C or lower for several weeks, the most quantitatively important nutrient which is damaged is lysine. This is due to its high level in milk protein compared to the other amino compounds and of the high reactivity of its ε-amino group (Table 1). Milk is one of the most sensitive foods to this reaction because of its high content of the reducing sugar lactose in addition to lysine. The most important negative consequence of the early Maillard reaction is therefore the 'blockage' of lysine in milk-based products (infant formulae and weaning foods). Lysine is essential for growth and its requirement is high ($103\,mg\,kg^{-1}\,day^{-1}$ for babies as compared to $12\,mg\,kg^{-1}\,day^{-1}$ for adults). The high, recommended lysine level in baby formulae (minimum of 6.7 g per 100 g of protein, equal to the level in mother's milk) is reached using cows' milk which contains an excess of lysine of at least 20% compared to mother's milk.

Industrially treated milks contain a certain amount of such blocked lysine as its Amadori compound, the amount varying between 0 and 15% depending on the treatments applied (Table 2). Higher values are reached with the roller drying process, though this process is no longer used in the industrial-scale production of milk formulae.

Lysine bioavailability is also affected by the Maillard reaction in other heat-treated foods like bread, biscuits, and pastas, but the negative nutritional consequences

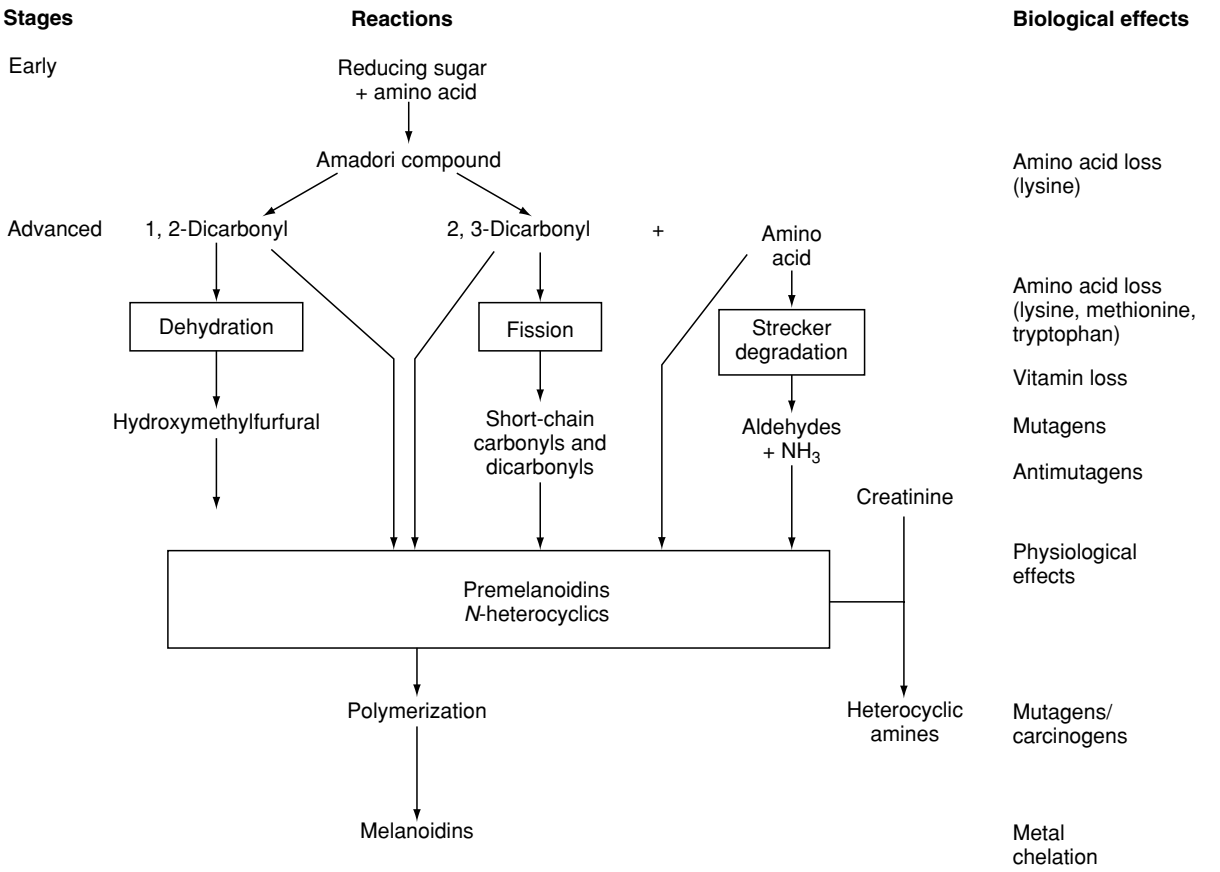

Figure 1 Browning reaction. Chemical pathways and biological effects. Reproduced from Browning: Toxicology of Nonenzymatic Browning, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 1 Nutritional effects of early and advanced Maillard reactions. Model system: dried milk powder

Maillard status treatments	*Percentage of original values of untreated sample*							
	Reactive lysine	*Lysine as Amadori compound*	*Lysine in advanced Maillard products*	*Vitamin B_1*	*Pantothenic acid*	*Vitamin B_6*	*Folic acid*	*Vitamin B_{12}*
Early								
60 °C, 4 weeks	74	26	0	81	90	84	29	100
Advanced								
70 °C, 4 weeks	17	14	69	5	7	18	3	31

			Tryptophan		
	Reactive lysine	*Nitrogen digestibility*	*Chemical analysis*	*Bioavailability*	*Methionine*
Early					
50 °C, 9 weeks	79	98	100	100	100
Advanced					
60 °C, 4 weeks	20	75	100	92	92

are much lower than for milk-based infant formulae as these cereal foods are poor sources of lysine.

Quantification of Lysine Damage

Many biological and chemical methods have been developed to quantify available and reactive lysine respectively. Available lysine (= reactive lysine × digestibility) may be evaluated after enzymic hydrolysis, by microbiological tests and, in animals, by growth tests. Reactive lysine may be measured as ε-*N*-dinitrophenyl-lysine after derivatization with fluorodinitrobenzene, as homoarginine after derivatization with *O*-methylisourea, and as lysine after reduction of the Amadori compound with sodium

Table 2 Lysine blockage in processed milk

Heat process	*Percentage lysine as Amadori compound*[a]
Freeze-drying	0
Pasteurization	0
UHT sterilization	0–2
Spray-drying	0–2
Spray-drying infant formula	5–10
HTST sterilization	5–10
Conventional sterilization	10–15
Roller drying	20–50

[a]According to the furosine method.
UHT, ultra heat-treated; HTST, high-temperature, short-time.

borohydride. Reactive lysine can also be determined after conventional acid hydrolysis, by the furosine method. This method is based on the property of fructosyl-lysine to produce upon acid hydrolysis a new amino acid, i.e., furosine. The measurement of the furosine level allows quantification of the Amadori compound of lysine and calculation of the amount of reactive lysine.

The Advanced Maillard Reaction

This stage of the reaction starts with the degradation of the sugar moiety of the Amadori compound involving dehydration, scission, oxidation, and Strecker degradation, leading to the generation of new molecules (premelanoidins) contributing to aromas and flavors that vary in their characteristics according to the conditions employed. Some of these molecules are chemically more reactive than the initial sugar, e.g., α-dicarbonyls, reductones, and aldehydes. They react not only with free amino groups still present in the food, but also with a larger number of radical groups such as hydroxyl, amines, and other nitrogen-containing molecules, vitamins, and the side chains of amino acids.

Polymerization of premelanoidins leads to the formation of the high-molecular-weight melanoidins, the reactions involved being responsible for the developments of pigments.

During the advanced stage of the Maillard reaction, loss of lysine and of B-vitamins in foods is accelerated. Other amino acids become less bioavailable because they are either chemically modified or a decrease in the total food nitrogen digestibility occurs. Lysine lost as new unavailable addition product(s) can be calculated by difference as 'lysine in advanced Maillard compound(s)' (Table 2). These addition products do not regenerate lysine on acid hydrolysis and their chemical structures are still unknown.

Indirect Effects

Associated with the presence of Maillard reaction products, a good example of an indirect negative nutritional effect of nonenzymatic browning is the reduction in protein digestibility. This is due to the inability of proteases and peptidases to hydrolyze the peptide bonds now containing modified amino acids. Feeding such a protein results in an increase in the amount of fecal nitrogen and a decrease in the bioavailability of the chemically unmodified amino acids. Another indirect effect is a modification of the metabolism (intestinal absorption, utilization, and urinary excretion) of some minerals and trace elements (calcium, zinc, iron, and copper). The increase in the urinary excretion of zinc, iron, and copper in patients fed intravenously a sterilized aqueous solution of free amino acids and glucose, and the hyperzincuria observed in rats fed heated casein–glucose and casein–lactose mixtures, would be due to the chelating effect of Maillard reaction products. However, this increase in urinary zinc being a relatively minor route of excretion compared to fecal excretion, zinc retention, and zinc status is not reduced in rats.

Physiological Effects

Many physiological changes in tissues of animals have been attributed to feeding Maillard reaction products. It has not always been easy, however, to distinguish between the changes that are directly due to the reduced nutritive value of the reacted protein and those that are probably induced by the Maillard components of the test protein in the diet. Furthermore, in many of these studies which corresponded to model systems, the reaction products tested were incorporated in the diets at such high levels that interpretation of observation has remained difficult.

Reduced growth rate and increased relative weight of liver, kidney, and cecum have been noted in rats fed a severely reacted (brown) compared to an unreacted mixture of egg protein and glucose. Parallel increases in values for blood glucose and urea nitrogen, and serum transaminases and alkaline phosphatase were also observed with the reacted mixture. The lower growth rate, attributable to the reduced nutritive value of the reacted egg protein, may partly explain the above differences. Long-term rat feeding trials, designed to eliminate the influence of the nutritional factors through increased dietary protein level, improved protein quality, and a lower, more realistic level of incorporation of the reacted egg protein under test showed less dramatic effects. Nevertheless, an apparent effect of the reacted brown Maillard compound was evident in the depressed growth rate, enlarged liver, spleen, and cecum, lowered serum triglycerides, increased total iron, and histopathologically observed fatty changes in the liver.

Activities of intestinal mucosal dissacharidases (lactase, sucrase, and maltase) were also reduced in rats fed the reacted, brown egg protein, compared to those fed the unreacted material. Some improvement occurred on fortification of the diet with amino acids (apparently) destroyed during reaction, although a persistent decrease in the activity of these enzymes demonstrated a specific inhibitory effect of the Maillard compounds.

Diarrhea is often observed in animals fed high levels of severely browned Maillard products. This is likely to be a direct consequence of the increased amount of indigestible protein available for fermentation in the cecum, which is often found to be enlarged.

During the progress of the nonenzymatic browning reaction, additional glycosyl residues become fixed to specific milk proteins leading to an increase in their allergenicity. Using a skin reactivity test, a good correlation has been shown between the degree of browning and allergenic response of a modified bovine β-lactoglobulin. Whilst this may explain one of the mechanisms responsible for the enhancement of the allergenicity of a protein, it does not explain the 'how and why' of allergy to cows' milk protein.

Maillard reaction products of low molecular weight which are partially absorbed can affect the activity of detoxifying enzymes. Rats fed browned egg albumin exhibited increased hepatic benzo[*a*]-pyrene hydroxylase activity and decreased colonic aminopyrine *N*-demethylase activity compared to animals fed nonbrowned egg albumin. This suggests that Maillard reaction products may modify the metabolism of endogenous substrates, exogenous drugs, and other xenobiotics.

Dietary melanoidins appear to decrease plasma cholesterol in rats and to modify the composition of fecal, neutral steroids, suggesting that melanoidins influence the intestinal metabolism of cholesterol.

Maillard reaction products also interact with the microorganisms of the digestive tract. Thus they strongly inhibit the activity of glycosyl transferase of *Streptococcus mutans* which has an active role in the development of dental caries. This inhibition would hence reduce the adherence of this microorganism to the tooth surface. Also, in the gut, the microflora is able to regenerate the amino acids from their Amadori compound and oxidize the advanced Maillard reaction products.

Toxicity

Studies of the toxicity of Maillard reaction products, both *in vitro* and *in vivo*, are a recent development. This interest comes from the observation that Maillard reactions involving creatinine (a typical molecule of animal tissues) or protein pyrolysis represent a carcinogenic potential due to the formation of heterocyclic amines. In addition to these amines, there exist mutagenic products of the premelanoidin and melanoidin types resulting from heat-processed carbohydrate-rich foods such as bakery products, caramel, coffee, and other milk and meat products.

Mutagenic/Carcinogenic Heterocyclic Amines

Using the short-term Ames microbiological test, it has been possible to detect mutagenic activity in broiled fish and meat. A series of heterocyclic amines has been isolated from cooked foods and from protein and amino acid pyrolysates. Some of them, 2-amino-3-methylimidazo [4,5-*f*]quinoline (IQ), 2-amino-3,4-dimethylimidazo [4,5-*f*]quinoline (MeIQ), 2-amino-3-methylimidazo [4,5-*f*]quinoxaline (IQx), 2-amino-3,8-dimethyl-imidazo [4,5-*f*]quinoxaline (MeIQx), 2-amino-3,4,8-trimethylimidazo[4,5-*f*]quinoxaline (4,8-DiMeIQx), and 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine (PhIP), possess an imidazole ring which is formed from creatinine. The remaining parts of the molecules (pyridines and pyrazines) are derived from Maillard reaction products formed by amino acids and sugars. Since creatinine is present only in animal protein sources, this class of substances is found only in these foods, albeit at low levels in the charred portions of broiled fish and the surface of grilled meat (PhIP 1 p.p.b; MeIQx 1 p.p.b.) and in beef extracts (IQ 70 p.p.b; MeIQx 8–90 p.p.b.).

The other heterocyclic amines detected in foods are pyrolysis products formed from tryptophan: 3-amino-1,4-dimethyl-5*H*-pyrido[4,5-*b*]indole (Trp-P-1), 3-amino-1,4-dimethyl-5*H*-pyrido[4,5-*b*]indole (Trp-P-2), from glutamic acid: 2-amino-6-methyldipyrido-[1,2-*a*: 3′,2′-*d*]imidazole (Glu-P-1), 2-aminodipyrido-[1,2-*a*: 3′,2′-*d*]imidazole (Glu-P-2), from phenylalanine: 2-amino-5-phenylpyridine (Phe-P-1), and from soya bean globulin: 2-amino-9*H*-pyrido[2,3-*b*]indole (AαC), 2-amino-3-methyl-9*H*-pyrido[2,3-*b*]indole (MeAαC). The heterocyclic amines which are produced in very small amounts in foods have been found to be carcinogenic in rats and mice (Table 3). The major target organs are the small and large intestines, liver, lung, and blood vessels.

Mutagenicity of Melanoidin-Type Products

Heated mixtures of free amino acid or food proteins and sugars possess a weak mutagenic activity in the Ames test. This activity varies with the composition of the mixture and the conditions of the reaction. Generally, the mutagenic activity observed *in vitro* correlates quite well with the formation of brown

Table 3 Mutagenic and carcinogenic effects of heterocyclic amines

Compound	*Source*	*Mutagenicity*	*Carcinogenicity*			
		Revertants (μg) TA 98	*Species*	*Concentration (%)*	*Experimental period (weeks)*	*Target organs*
Products formed from creatinine						
IQ	Broiled sardines	433 000	Rats	0.03	72	Liver, intestine, skin
			Mice	0.03	96	Liver, lung
MeIQ	Broiled sardines	661 000	Rats	0.03	40	Large intestine, mammary gland
			Mice	0.04	91	Liver, forestomach
MeIQx	Fried beef	145 000	Rats	0.04	61	Liver, skin
			Mice	0.06	84	Liver, lung
Pyrolysis products formed from tryptophan						
Trp-P-1	Tryptophan pyrolysate	39 000	Rats	0.02, 0.015	52	Liver
			Mice	0.02	89	Liver
Trp-P-2	Tryptophan pyrolysate	104 200	Mice	0.02	89	Liver
Pyrolysis products formed from glutamic acid						
Glu-P-1	Glutamic acid pyrolysate	49 000	Rats	0.05	67	Liver, small and large intestines
			Mice	0.05	68	Liver, blood vessels
Glu-P-2	Glutamic acid pyrolysate	1 900	Rats	0.05	104	Liver, small and large, intestines
			Mice	0.05	84	Liver, blood vessels

IQ, 2-amino-3-methylimidazo [4,5-*f*]quinoline; MeIQ, 2-amino-3,4-dimethylimidazo-[4,5-*f*]quindine; MeIQx, 2-amino-3,8-dimethyl-imidazo [4,5-*f*]quinoxaline; Trp-p-1, 3-amino-1,4, dimethyl-5*H*-pyrido [4,5-*b*] indole; Trp-P-2, 3-amino-1-methyl-5*H*-pyrido[4,5-*b*]indole; Glu-P-1, 2-amino-6-methyldipyrido-[1,2-*a*: 3′,2′-*d*]imidazole; Glu-P-2, 2-aminodipyrido-[1,2-*a*: 3′ 2′-*d*]imidazole.

pigments and with the presence of molecules with high antioxidant capacity.

The demonstration that carbonyl compounds can produce free radicals lends confirmation to the observation that the mutagenicity of the antioxidant fraction of the Maillard products is deactivated by catalases. This apparent contradiction suggests that, when applied at high concentration to *in vitro* tests, Maillard products can induce mutations due to an 'overpowering' of their prooxidant activity, a phenomenon already observed with other antioxidants, such as the antioxidant vitamins and phenolic compounds.

When considering model systems reflecting normal cooking conditions a relatively weak mutagenic activity was observed compared to that of heterocyclic amines. The typical mutagenic effects of this class of compounds reflect base pair, in contrast to frameshift mutations, as caused by heterocyclic amines. Furthermore, their mutagenic effect has been found to be effectively inactivated by the liver microsomal enzymes (S9 liver fraction), in contrast to heterocyclic amines which require S9 activation to mutagenic/carcinogenic species. (*See* **Mutagens**.)

Examples of heated foods with such a weak mutagenic activity that is metabolically deactivated *in vitro* are some commercial caramel preparations (ammonia and sulfite/ammonia processes), hydrolyzed plant proteins and roasted coffee. However, it has been confirmed in animal feeding studies that melanoidin-containing foods such as roasted coffee and roasted cocoa are neither mutagenic nor carcinogenic. Moreover, tested *in vivo* under well-controlled physiological conditions, Maillard reaction products with antioxidant activity are found to act as antimutagens and/or anticarcinogens.

Antioxidant and Antimutagenic Activity of Melanoidin-Type Products

Many Maillard reaction products exhibit antioxidant properties which are utilized with success in different types of food. Such reaction products include premelanoidins which are rich in carbonyl compounds as well as melanoidins. The observed antioxidant property depends on the conditions of the reaction and on the amino acids. Thus, reaction mixtures containing histidine have been found to have the highest antioxidant activity. The mechanism itself of the antioxidant

activity in food systems involves both the reducing capacity of the carbonyl compounds in the premelanoidins and melanoidins, and the property of the melanoidin brown pigments to chelate prooxidant metals, copper, and iron.

That Maillard reaction products possess antimutagenic and/or anticarcinogenic properties has been evidenced. Thus, such reaction products inhibit mutagenic activity of several foodborne heterocyclic amines and of aflatoxin B_1. Antimutagenic and possible anticarcinogenic activities have been attributed to melanoidins and α-dicarbonyl compounds generated either during the aminocarbonyl or caramelization reactions. The scavenging of active oxygen by melanoidins, demonstrated by means of electron spin resonance, can explain their antimutagenic effects. The involvement of antioxidant reactions in the antimutagenicity of heat-reaction products suggests that these compounds might have anticarcinogenic properties which include inhibition of nitrosamine formation, scavenging of active oxygen involved in cancer initiation and promotion, and also change in the chemical structure of carcinogens during the browning process.

Nitrosamines of Amadori Compounds

Amadori compounds have been shown to be nitrosated to mutagenic *N*-nitroso derivatives. This has led to the postulation that such compounds, if present in human food or if formed in the stomach, will constitute a health hazard. However, heated foods rich in Maillard reaction products with antioxidant properties can both inhibit nitrosamine formation and reduce nitrosamine-induced carcinogenicity. (*See* **Nitrosamines**.)

See also: **Amines**; **Browning**: Nonenzymatic; **Nitrosamines**

Further Reading

Eriksson C (ed.) (1981) *The Maillard Reaction in Food. Progress in Food and Nutritional Science 5*. Oxford: Pergamon Press.

Finot PA, Aeschbacher HU, Hurrell RF and Liardon R (eds) (1990) *The Maillard Reaction in Food Processing, Human Nutrition and Physiology. Advances in Life Sciences*. Basel: Birkhauser.

Friedman M (ed.) (1991) *Nutritional and Toxicological Consequences of Food Processing. Advances in Experimental Medicine and Biology 289*. New York: Plenum Press.

Lee TC, Pintauro SJ and Chichester CO (1982) Nutritional and toxicological effects of non-enzymic Maillard browning. *Diabetes* 31 (suppl. 3): 37–46.

Enzymatic – Biochemical Aspects

J Nicolas and C Billaud, Conservatoire National des Arts et Métiers, Paris, France
J Philippon (Retired), Centre National du Machinisme Agricole du Génie Rural des Eaux et des Forêts, Antony, France
M A Rouet-Mayer, Centre National de la Recherche Scientifique, Gif sur Yvette, France

Introduction

The discoloration which occurs in plant material after cell disruption and results in the formation of brown or sometimes yellow, black, or pink pigments is mainly due to the process of enzymatic browning. Besides plant materials which are primarily concerned, enzymatic browning also affects some animal materials, e.g., melanosis (black spot) of shrimp.

Cell disruption can arise either from mechanical injury or temperature changes which lead to physiological disorders or even cell death (e.g., for deep-frozen products). This loss of cell integrity results in the decompartmentation of phenolic substrates and enzymes and then, in the presence of molecular oxygen, the oxidative production of colored pigments.

Brown pigmentation following this enzymatic reaction, and subsequent nonenzymatic reactions, is generally considered to be detrimental to food quality from both the organoleptic and nutritional points of view. The prevention of enzymatic browning has always been a challenge to food scientists owing to the loss of quality that it causes in many food products, e.g., fruits and vegetables during either storage or processing. It is only in a few exceptional cases that the enzymatic browning is desirable (prunes, dates, tea, tobacco, etc.).

With a better knowledge of the mechanism of browning reactions, it will be possible to propose processes which avoid or at least minimize this discoloration and which can be adapted to each particular product. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments).

Nomenclature

Two kinds of enzymes are able to act upon diphenols in the presence of molecular oxygen according to the reaction scheme shown in **Figure 1**. Both have the trivial name polyphenol oxidases, but they are somewhat different in nature. The first class of enzymes, catechol oxidases (1,2-benzenediol: oxygen

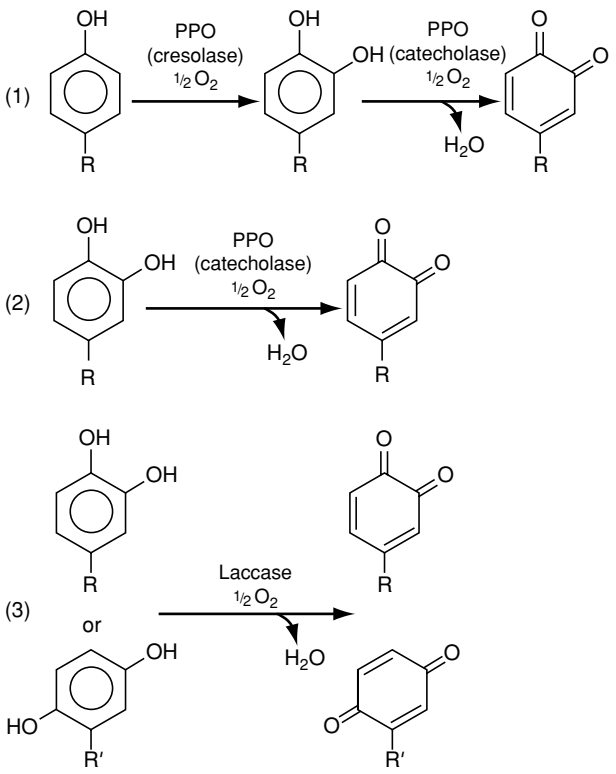

Figure 1 Reactions catalyzed by polyphenol oxidases. Reproduced from Browning: Enzymatic-Biochemical Aspects. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

oxidoreductase, EC1.10.3.1), catalyze two distinct reactions (reactions (1) and (2) of Figure 1), namely hydroxylation of monophenols into *o*-diphenols (cresolase activity) and oxidation of *o*-diphenols into *o*-quinones (catecholase activity). Both reactions consume oxygen and the overall stoichiometry is 1 mol of oxygen for 1 mol of monophenol, giving 1 mol of *o*-quinone and 1 mol of water. The second class, laccases (benzenediol: oxygen oxidoreductase, EC1.10.3.2), oxidize *o*-diphenols as well as *p*-diphenols, forming their corresponding quinones (reaction (3) of Figure 1). The stoichiometry is one atom of oxygen for 1 mol of diphenol giving 1 mol of quinone and 1 mol of water. The unique ability to oxidize *p*-diphenols can be used to distinguish the laccase activity from that of the first class of polyphenol oxidases. In all cases, the quinones formed are very reactive and, depending on the nature and the concentration of the reactive species in the medium, they can enter in numerous secondary enzymatic and nonenzymatic reactions which will be described later.

The nomenclature of these enzymes is somewhat confusing since, besides the two designated as EC1.10.3.1 and EC1.10.3.2, a third one, EC1.14.18.1 (monophenol, L-DOPA: oxygen oxidoreductase), exists. It is referred to as monophenol monooxygenase (tyrosinase) and corresponds to the same enzymes as EC1.10.3.1 but which always catalyze the hydroxylation of monophenols. For the sake of simplicity, we will use the general terms of polyphenol oxidase (PPO) for the first class (EC1.10.3.1) and laccase for the second class (EC1.10.3.2). (*See* **Enzymes:** Functions and Characteristics.)

Enzymatic Browning Factors

The different factors of enzymatic browning are examined, including some minor ones which may occasionally play an important role, e.g., peroxidases.

Enzymes

Polyphenol oxidases (EC1.10.3.1) PPOs are copper oxidoreductases exhibiting both cresolase and catecholase activities. However, many enzymatic preparations have a very low cresolase activity or none at all. It is difficult to insure that the latter activity was originally absent in the plant or has not either been extracted or has been destroyed during extraction owing to its lability. Cresolase activity is often lost during purification. Thus, in a preparation, the ratio of catecholase to cresolase activity (when the latter is present) can vary from 1 to 40.

PPO is present in a wide variety of plants. In a particular plant, PPO activity varies from one organ to another and varies inside an organ, depending on the tissue considered. PPOs have been found in different cell fractions, in organelles (chloroplasts and, more precisely, in thylakoids, mitochondria, peroxisomes) where the enzymes are tightly bound to membranes and in the soluble fraction of the cell. The degree of binding to membranes varies with the tissue and its ontogenic state. Thus, the overall PPO activity is higher and mostly present in bound forms in young green fruits, whereas it generally decreases and the proportion of soluble forms increases in ripe fruits.

Extraction of the PPO activity from plant sources is complicated by the presence of endogenous phenolic substrates which are oxidized and then interact with proteins. Besides destroying activity, this may induce 'new' enzymatic forms corresponding to artifacts. As a consequence, this oxidation has to be prevented by adding a reducing agent (e.g., ascorbic acid or a thiol) and/or a phenol-complexing compound (e.g., polyamide, polyvinylpolypyrrolidone (PVPP), or polyethylene glycol) to the extracting solution. Solubilization after preparation of an acetone powder is a frequently used procedure. Detergents (Triton

X100 or X114) are also used, but probably induce modifications of the enzyme structure and properties. Most of the purification procedures are based on fractional precipitation by ammonium sulfate followed by one or several chromatographic steps. Until the mid-1980s, the molecular mass of PPOs was believed to be around 40–45 kDa. However, PPOs from a number of plant species have now been characterized with molecular masses ranging from 20 to 180 kDa. This disparate estimation of molecular mass, traducing the heterogeneity and multiplicity of forms of the enzyme, can partly be attributed to artifacts of protein isolation and electrophoresis, as exposed above. Moreover, they can correspond to small multigene families for PPO. In fact, PPO genes characterized from different leaf and fruit plant species (including faba bean, apple, tomato, potato, spinach, Virginian pokeweed, and grape berry) are present in the plant genome as gene families. To date, cDNAs for PPOs cloned from these various plants have been found to encode mature peptides of 56–67 kDa with 8–12 kDa putative transit peptides. The nuclear encoded-plastid protein is thus translocated as a precursor with this transit peptide that can direct its transport into the thylakoid membrane or lumen. The copper content of various plant PPOs has not yet been well established, but their deduced amino acid sequences from the cloned PPOs genes have revealed two highly conserved copper-binding domains, showing significant homology with that of closely related bacterial, fungal, or mammalian tyrosinases.

Most of the PPOs studied show an optimum activity between pH 4 and 7. However, there are often discrepancies in the published values for PPOs from the same source. In addition, several reports indicate differences in pH optimum depending on cultivars and maturity. These differences have been attributed to variations in the proportions of isoenzymes which have distinct pH optima. Moreover, pH optima were found to vary with the phenolic substrate used for several enzymatic preparations. The effect of temperature on PPO activity has been less investigated than that of pH. The optimum temperature ranges from 15 to 40 °C, depending on the same factors as pH.

Laccases (EC1.10.3.2) The laccases catalyze the oxidation of *o*- and *p*-diphenols into their corresponding quinones according to reaction (3) in **Figure 1**. They occur much less frequently in the plant kingdom than PPOs. Their presence has been mainly reported in many fungi and in species of the genus *Rhus* (e.g., the Japanese lacquer tree). They are almost absent from fruits and vegetables, with the exception of certain peach cultivars and apricots. Nevertheless, the distinction between PPO and laccase activities is not always clear-cut since the presence of endogenous phenols can induce the coupled oxidation of *p*-diphenols and therefore leads to a false conclusion on the presence of a laccase type of activity. Selective inhibitors can be used to ascertain the type of activity. Thus, cinnamic acid, salicylhydroxamic acid, 4-hexylresorcinol, phenylhydrazine, and carbon monoxide inhibit PPO activity more specifically, whereas cetyl trimethyl ammonium bromide, a cationic detergent, is more specific for laccases.

Most fungi laccases are glycoproteins with a basic subunit consisting of a single polypeptide chain of 50–70 kDa containing a considerable amount of carbohydrate (10–45%) and four atoms of copper. The effect of pH on laccase activity is similar to that observed for PPO activity, i.e., an optimum pH which ranges between 4 and 7.5 and depends on the substrate being used.

Peroxidases (EC1.11.1.7) The peroxidases are enzymes whose primary function is to oxidize hydrogen donors at the expense of peroxides. They are highly specific for hydrogen peroxide, but they accept a wide range of hydrogen donors, including polyphenols. The overall reactions catalyzed are shown in eqns (1) and (2).

$$AH_2 + H_2O_2 \rightarrow A + 2H_2O \qquad (1)$$

$$2AH + H_2O_2 \rightarrow AA + 2H_2O \qquad (2)$$

Peroxidases are glycoproteins with a hematin compound as cofactor. Their molecular weights range between 30 and 55 kDa. Depending on the enzyme source, the isoenzyme considered, and the hydrogen donor substrate, the optimum activity is between pH 4 and 7. In many, but not all cases, the peroxidase isoenzymes are activated by calcium ions. The higher thermostability of some isoenzymes is well known, and the residual peroxidase activity after blanching is often used as an index of thermal treatment.

Although peroxidases are widely distributed, especially in plants, they generally appear to be little involved in enzymatic browning of fruits and vegetables following a mechanical stress. The explanation could be that the peroxidase activity is limited by the internal level of hydrogen peroxide. However, their involvement in slow processes such as internal browning during cold storage of fruits is possible.

Nevertheless, the direct involvement of peroxidase in browning still needs confirmation, just as does that of laccase which may not be present in some fruits and vegetables. Therefore, the following sections will be mainly devoted to PPO activity.

Substrates

Susceptibility to browning varies widely from one plant to another. This variation is linked to both quantitative and qualitative aspects of their phenolic content. Thus, browning after bruising of fruit cultivars with similar amounts of total phenolics can be either more or less intense. Among the wide variety of phenolic compounds found in fruits and vegetables, only a small number serves as direct substrates for PPO. Caffeic acid derivatives and monomeric flavan-3-ols (mainly (+)-(gallo)catechin and (−)-(gallo)epicatechin) often appear to be the best substrates. Other quantitatively important classes of phenols, namely anthocyanins, flavonols, and condensed forms of flavan-3-ols (tannins), are weakly if not directly oxidized by PPO. The same holds for other less important classes of phenols (flavones, flavanones, flavononols, chalcones, and dihydrochalcones). This restricted activity is probably related to the presence of a sugar moiety in many of these molecules which could cause steric hindrance since the aglycone forms are often good substrates of PPO. Nevertheless, phenolic compounds which are not direct substrates can actively participate in browning through coupled oxidation reactions. Thus, in model systems, it has been shown that the degradation of anthocyanins, procyanidins, and flavonols by PPO is greatly accelerated in the presence of caffeic acid derivatives or catechins. The *o*-quinones enzymatically formed from either of the latter compounds are able to promote cooxidation reactions leading to both the degradation of the former compounds and the regeneration of good substrates for the enzymatic reaction. This degradation and the phenolic copolymerization resulting from nonenzymatic coupled oxidations (reactions (2) and (3) of Figure 2) lead to products which may be intensely brown.

Many studies have been devoted to the specificity of PPO towards phenolic substrates. Usually, the apparent Michaelis constant (K_m) is higher than 1 mmol l^{-1}, indicating a relatively low affinity. However, for a particular substrate, the K_m values can vary widely depending on the PPO source, and the same holds for the relative rate of oxidation among different

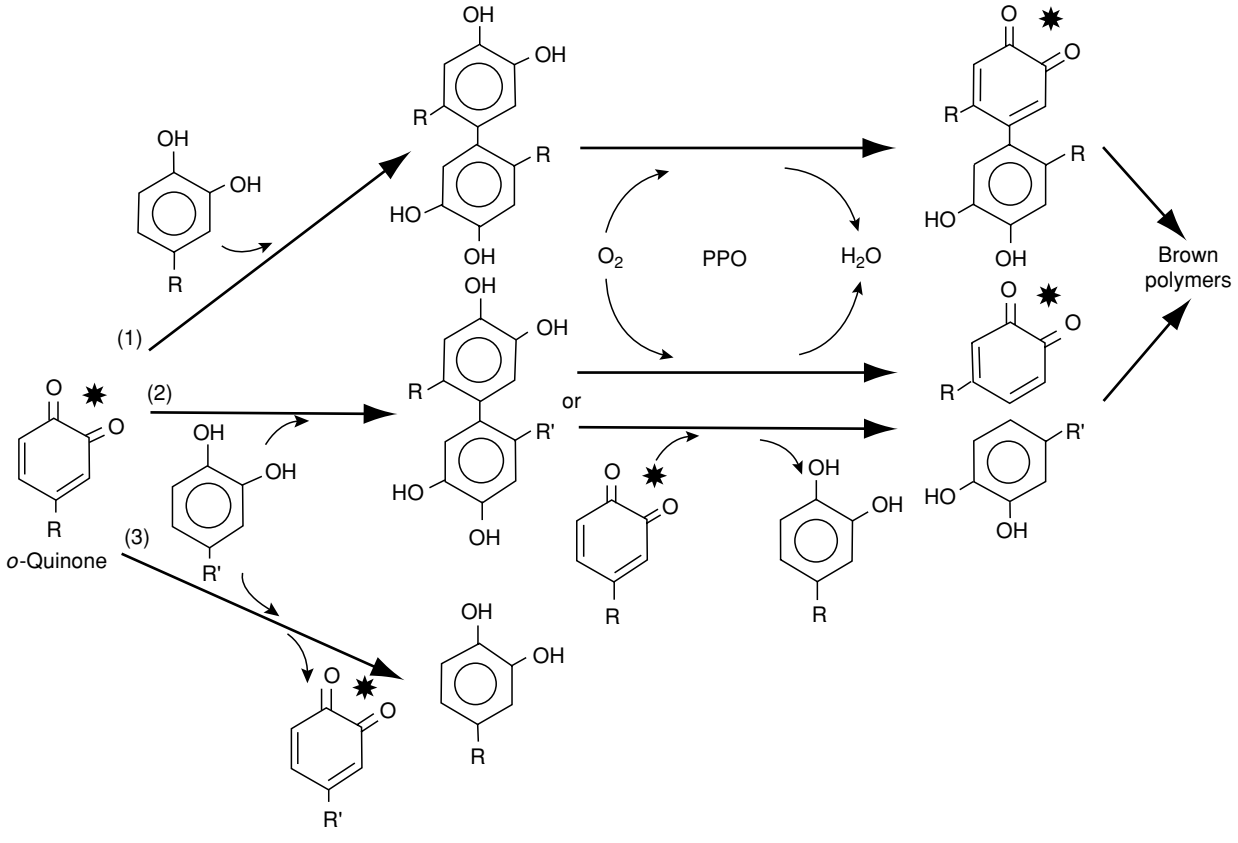

Figure 2 Reactions of *o*-quinones with phenolic compounds. (All reactions are nonenzymatic except those with polyphenol oxidases (PPOs). Reactions (2) and (3) are able to regenerate the original phenol.) Products with differing color intensities are indicated by asterisks. Reproduced from Browning: Enzymatic-Biochemical Aspects. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 1 Michaelis constant (K_m: mmol l^{-1}) and V_m (maximal rate) values (expressed as a percentage of the V_m of chlorogenic acid) of polyphenol oxidases from different sources for three common natural substrates

	Apple		*Grape*		*Pear*		*Peach*	*Apricot*	*Potato*
	V_m[a]	K_m[a]	V_m[b]	K_m[c]	V_m[b]	K_m[c]	V_m[b]	K_m[c]	K_m[c]
Chlorogenic acid	100	4.2	100	2.5	100	16.1	100	1.2	1.4
(+)-Catechin	58	6.2	64	1	60	2.1	373	0.74	
Caffeic acid	8.1	0.14	69	5.5	43		61	0.5	2.4–2.9

Adapted from:
[a]Janovitz-Klapp A, Richard F, Goupy P, Nicolas J (1990) Kinetic studies on apple polyphenol oxidase. *Journal of Agricultural and Food Chemistry* 38: 1437–1441.
[b]Macheix JJ, Fleuriet A, Billot J (1990) *Fruit Phenolics*. Boca Raton: CRC Press.
[c]Vamos-Vigyazo L (1981) Polyphenol oxidase and peroxidase in fruits and vegetables. *CRC Critical Reviews in Food Science and Nutrition* 15: 49–127.

phenolics (Table 1). It is often suggested that the preferred substrate is the most abundant phenolic, although this is not always true.

By contrast, few studies have concerned oxygen, the other substrate. Steady-state kinetics carried out on PPO indicate that it probably follows an ordered Bi-Bi mechanism in which the oxygen binds first. The reported values for equilibrium constant (K_{mapp} O_2) are in the range 0.1–0.5 mmol l^{-1}, corresponding to a rather low affinity for the oxygen (compared to, for example, cytochrome oxidase: 0.5–1 μmol l^{-1}). (*See* **Oxidation of Food Components; Phenolic Compounds; Tannins and Polyphenols.**)

Reaction Products

The primary products of enzymatic oxidation are *o*-quinones. These molecules have different spectral properties and their color depends on the pH and the phenol from which they originate. Some of the molar extinction coefficients at maximum wavelengths are given in Table 2, and show a wide range of variation.

The colors are different from their precursors, since after oxidation catechin is bright yellow, chlorogenic acid is dull orange-yellow, whereas DOPA is pink. Moreover, the *o*-quinones are reactive compounds, as illustrated in Figures 2 and 3. Consider Figure 2 for the reactions with phenolic compounds. *o*-Quinones can react with another phenol molecule, resulting in a dimer of the original phenol (reaction (1) of Figure 2). This dimer, having an *o*-diphenolic structure, can be subject to reoxidation either enzymatically or by another *o*-quinone and gives larger oligomers with different color intensities. The *o*-quinones can also react with a different phenol molecule, either leading to a copolymer (reaction (2) of Figure 2) or regenerating the original phenol and giving a different *o*-quinone (reaction (3): coupled oxidation of Figure 2). In Figure 3, corresponding to reactions with nonphenolic compounds, another coupled oxidation reaction is observed with ascorbic acid (reaction (1) of Figure 3), since the phenol is regenerated with the simultaneous formation of dehydroascorbic acid.

Table 2 Molar extinction coefficients of quinones from different *o*-diphenolic substrates of polyphenol oxidases

Substrate	*Wavelength (nm)*	*Extinction coefficient*
Pyrocatechol[a]	390	1417
3,4-Dihydroxyphenylacetic acid[a]	390	1311
4-t-Butylcatechol[a]	400	1150
4-Methylcatechol[a]	400	1400
L-DOPA[a]	480	3388
Hydrocaffeic acid[a]	412	1124
Chlorogenic acid[b]	420	2000
(+)-Catechin[b]	380	1200

Adapted from:
[a]Waite JH (1976) Calculating extinction coefficients for enzymically produced *o*-quinones. *Analytical Biochemistry* 75: 211–218.
[b]Rouet-Mayer MA, Ralambosoa J, Philippon J (1990) Roles of *o*-quinones and their polymers in the enzymic browning of apples. *Phytochemistry* 29: 435–440.

With sulfites colorless addition compounds are formed, together with regenerated diphenol (reaction (2) of Figure 3). The *o*-quinones can form additional compounds with thiol groups by nucleophilic substitution (reactions (3) and (4) of Figure 3). Cysteine, either free or bound in small peptides (e.g., glutathione) or in large proteins, generates colorless compounds. However, due to their *o*-diphenolic structures, these compounds can be either oxidized by laccase or react with excess *o*-quinones (by a coupled oxidation mechanism) and form intensely colored products. The same kind of addition reactions occur with amino groups (primary or secondary amines), although a little less readily ((reactions (5–7) of Figure 3). Furthermore, substitutions with other thiol or amino groups of proteins may occur, leading to intra- and intermolecular cross-links. Lastly, water slowly adds to the *o*-quinones to form triphenols which are readily oxidized by excess *o*-quinone (by a coupled reaction mechanism), leading to the *p*-quinones (reaction (8) of Figure 3). (*See* **Amines; Ascorbic Acid**: Properties and Determination.)

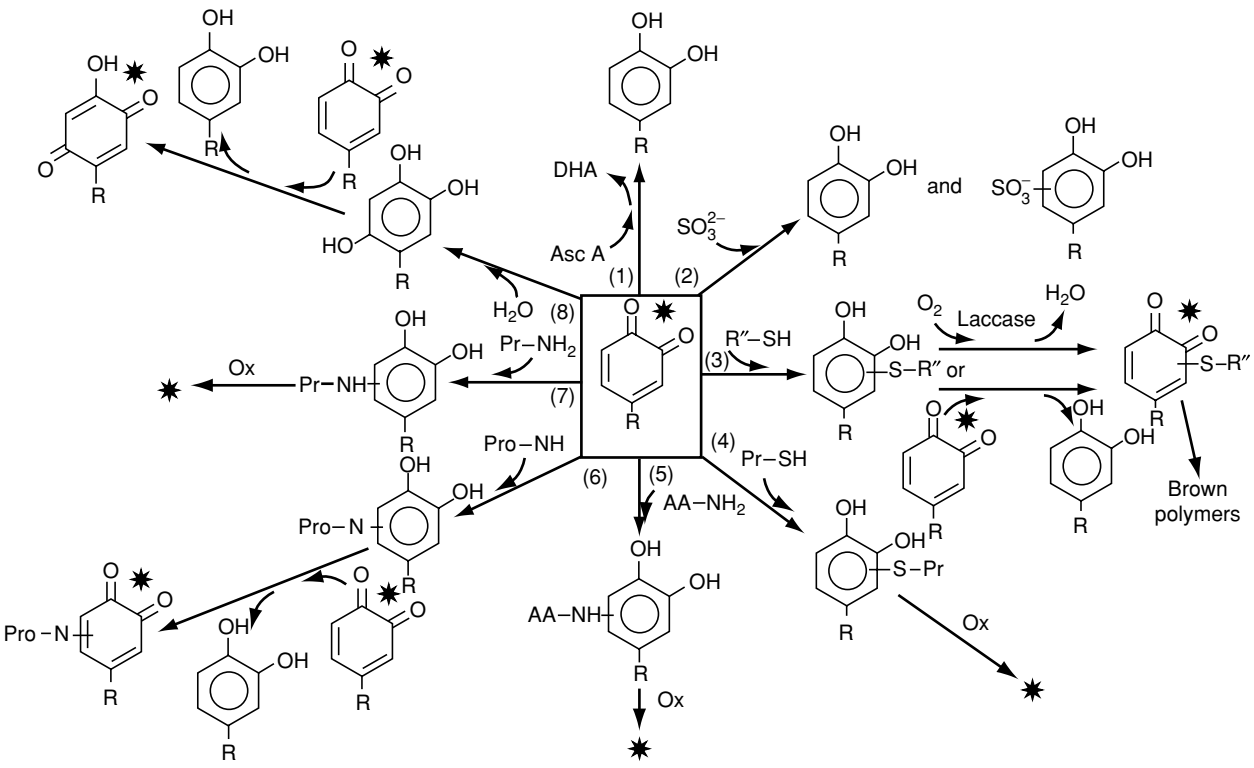

Figure 3 Reactions of *o*-quinones with nonphenolic compounds. (All reactions are nonenzymatic except those with laccase. Reactions (1), (2), (3), (6), and (8) are able to regenerate the original phenol.) Products with differing color intensities are indicated by asterisks. Ox, further oxidation reactions by oxygen or *o*-quinone; Pr-SH and Pr-NH_2, proteins; Pro-NH, proline; AA-NH_2, amino acids; Asc A, ascorbic acid; DHA, dehydroascorbic acid; R″SH, small thiol compounds (e.g., cysteine and glutathione). Reproduced from Browning: Enzymatic-Biochemical Aspects. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

The reactivity (or in other words, stability) of the *o*-quinones in these different cases is very variable. It depends strongly on the substituted nature of the parent phenol, and on the medium (composition, pH, temperature, etc.). Thus, in the same conditions, the *o*-quinones derived from 4-methylcatechol are more stable than those from chlorogenic acid which, in turn, are more stable than those from catechins. Obviously, the presence of reactive molecules, with amino or thiol groups, in the medium can greatly affect the stability of the *o*-quinones. Moreover, the reactivity with reducing compounds, namely those entering in coupled oxidation reactions, is under the control of the redox potentials of the systems involved. Thus, the *o*-quinones of chlorogenic acid are able to cooxidize catechins into *o*-quinones of catechin and regenerate chlorogenic acid, whereas the reverse is not true.

Inhibition of Enzymatic Browning

Enzymatic browning is often an undesirable reaction and its prevention is a major concern for food scientists. Therefore, a large part of the many studies devoted to enzymatic browning concern its control. This section gives only the physicochemical basis for the control of the reaction. The more applied view of the problem in food technology will be described in the next article. The different ways of browning control can be divided into three classes, depending on whether they affect enzymes, substrates, or reaction products. However, some inhibitors are able to act simultaneously on more than one of these factors.

Action on Enzymes

First, heat treatment is often employed, and usually short exposure to temperatures between 70 and 90 °C is sufficient to inactivate PPOs. High-pressure treatments have also been proposed for the denaturation of PPO. Very high pressures (400–500 MPa) are needed to inactivate the enzyme at 25 °C. However, PPO appears to be more vulnerable to high-pressure CO_2 treatment (close to 6 MPa) or to a combined treatment of heat and ultrasonic waves under moderate pressure (manothermosonication). Since the optimum pH of most PPOs lies between pH 4 and 7 (see earlier) on

the one hand, and the phenols are more readily oxidized when the pH increases on the other, acidification below pH 4 is often proposed, although PPOs are still active in some cases (grape, apple, etc.).

Besides these general treatments, more specific chemical inhibitors are known. PPOs having copper as a prosthetic group are inhibited by many metal-chelating agents such as cyanide, azide, diethyldithiocarbamate, and ethylxanthate, although inhibition depends on the source of the PPO. Inhibition by halide ions observed in numerous cases is pH-dependent, increasing when the pH decreases; it may be caused by the formation of a complex between the halide ion and copper which is enhanced by low pH values. Due to their structural resemblance to phenolic substrates, aromatic carboxylic compounds are usually found to be competitive inhibitors. As has been shown for apple PPOs, the inhibitory properties are highly dependent on structure (Table 3). Thus, for a particular substituent, inhibition decreases in the order cinnamic, benzoic, phenylpropionic, and phenylacetic acid. In each series, inhibition is slightly enhanced by *p*-hydroxy substitution and greatly reduced by *o*-methoxy substitution. Moreover, the presence of a benzene nucleus is not an absolute requirement for the inhibitory effect since sorbic acid, an aliphatic carboxylic acid with two conjugated double bonds, is almost as effective as benzoic acid. In all cases, inhibition increases as pH decreases and it was shown that the un-ionized form of the carboxylic group is mainly responsible for inhibition.

More recently, substituted resorcinols have been recognized as PPO inhibitors. Although resorcinol is a poor PPO inhibitor, the 4-substituted resorcinols exhibit the highest inhibition properties. Among them, the 4-hexylresorcinol has been proposed as a valuable sulfite alternative for use in the food industry in order to prevent enzymatic browning. It has been shown that 4-hexylresorcinol inhibits PPO in a non-classical mechanism involving the rapid formation of an enzyme–inhibitor complex which undergoes a slow reversible reaction (slow-binding competitive inhibitor).

Action on Substrates

Inhibition of enzymatic browning can also be obtained by removal of one of the two substrates (oxygen and phenolic compounds) from the reaction medium. Complete oxygen removal is the most satisfactory way to control the phenolic oxidation catalyzed by PPOs. However, although this method may be applied to dead tissues, either by creating a physical barrier to oxygen diffusion or by vacuum infiltration (a more detailed description on this subject will be given in the next chapter), it is inapplicable to living tissues due to the risk of metabolism deviations caused by anaerobic conditions. Concerning the phenolic substrates, two methods are available. The first one is the physical elimination by ultrafiltration or by specific adsorbents. In the latter case, the most widely used, polyvinylpyrrolidone (PVP) and its insoluble form (PVPP), are very effective in controlling enzymatic browning. Thus, PVP has also been shown to be a competitive inhibitor of PPOs. Other complexing agents of phenols can be used, such as polyethylene glycol or polyamide. In the same way, borate acts by complexing the *o*-dihydroxy groups of phenolic substrates. Lastly, the inhibitory properties of β-cyclodextrins, observed in apple juice browning, are related to their abilities to form inclusion complexes with PPO substrates. A similar mechanism of browning inhibition could be ascribed to some polysaccharides such as carragenan, amylose sulfate, xylan, and chitosan.

The second way of removing phenolic compounds is by their modification. This can be performed by two kinds of enzymes. The first modification is methylation of the *o*-diphenolic substrates of PPOs by an *O*-methyltransferase (e.g., caffeic acid is converted to ferulic acid). Unfortunately, this elegant

Table 3 Inhibition constants of sodium halides and some carboxylic acids for apple polyphenol oxidases at pH 4.5. All are competitive inhibitors, with the exception of sodium chloride, which is noncompetitive

Inhibitor	K_i *(inhibition constant)* $(mmol\,l^{-1})$	*Inhibitor*	K_i $(mmol\,l^{-1})$
Sodium iodide	117	Sodium bromide	106
Sodium chloride	20	Sodium fluoride	0.07
Benzoic acid	0.64	*p*-Hydroxybenzoic acid	0.57
Vanillic acid	10	Syringic acid	34.5
Cinnamic acid	0.092	*p*-Coumaric acid	0.04
Ferulic acid	0.29	Sinapic acid	15
Phenylacetic acid	13	Phenylpropionic acid	1.4
p-Hydroxyphenylpropionic acid	1.1	Sorbic acid	0.51

Adapted from Janovitz-Klapp A, Richard F, Goupy P, Nicolas J (1990) Inhibition studies on apple polyphenol oxidase. *Journal of Agricultural and Food Chemistry* 38: 926–931.

process is impeded by the high cost of the enzyme (*S*-adenosylmethionine: catechol *O*-methyltransferase) and one of its substrate (*S*-adenosylmethionine). The second modification is oxidative ring opening by protocatechuate 3,4-dioxygenase. However, in addition to its cost, this enzyme has poor catalytic activities on caffeic acid and its derivatives.

Action on Reaction Products

Acting on reaction products is the third way of controlling enzymatic browning. *o*-Quinones are very reactive primary products (Figures 2 and 3) and, using chemical means, they can either be reduced back to *o*-diphenols or trapped as colorless addition compounds. However, secondary products resulting mainly from the oxidative polymerization of *o*-quinones often give highly colored compounds which become less reactive as the browning reaction proceeds. Therefore, almost all the compounds listed in the section on reaction products which act on *o*-quinones, and more especially ascorbic acid and its derivatives, thiol compounds and sulfite derivatives, can be used for the control of enzymatic browning. However, their effectiveness in preventing browning is greatly reduced if their use is delayed until after the reaction has started.

Besides chemical means, brown pigments can be removed more or less completely by physical treatments such as ultrafiltration or use of resins in order to trap polymers of *o*-quinones. This can be an efficient method for the clarification of liquids such as fruit juices and wine.

Lastly, several studies have been devoted to the isolation of 'natural' inhibitors of enzymatic browning. Thus, low-molecular-weight compounds exhibiting inhibitory properties have been detected in cultures of *Penicillium expansum* and *Dactylium dendroides*. Inhibitors have also been found in anise oil, cumin, and several medicinal plants. Similarly, it has been shown that Maillard reaction products and honey extracts contain substances which are able to inhibit the PPO activity. Nevertheless, in most cases, these inhibitory compounds have been only partially characterized and their exact mechanism of inhibition is not fully understood. (*See* **Antioxidants:** Natural Antioxidants; Synthetic Antioxidants.)

In recent years, considerable progress has been made in understanding PPOs and polyphenols of fruits and vegetables. However, many studies were restricted to model systems with an enzyme acting on one phenolic substrate only. There is a great need for studies on enzymatic action on mixtures of phenols on the one hand, and on nonenzymatic reactions that follow *o*-quinone formation on the other. This research would give valuable information on how phenols affect both the enzymatic activity and the kind of brown pigments formed, and it would help to give a better understanding of the relation between the extent of browning and the phenolic composition of fruits and vegetables.

See also: **Amines**; **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; **Ascorbic Acid**: Properties and Determination; **Browning**: Toxicology of Nonenzymatic Browning; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Enzymes**: Functions and Characteristics; **Oxidation of Food Components**; **Phenolic Compounds**; **Tannins and Polyphenols**

Further Reading

Amiot MJ, Fleuriet A, Cheynier V and Nicolas J (1997) Phenolic compounds and oxidative mechanisms in fruits and vegetables. In: Tomas-Barberan FA, Robins RJ (eds) *Phytochemistry of Fruits and Vegetables*, pp. 52–85. Oxford: Clarendon Press.

Billaud C, Regaudie E, Fayad N, Richard-Forget F and Nicolas J (1995) Effect of cyclodextrins on polyphenol oxidation catalyzed by apple polyphenol oxidase. In: Lee CY, and Whitaker JR (eds) *Enzymatic Browning and its Prevention*, pp. 295–312. Washington: American Chemical Society.

Cheynier V, Fulcrand H, Guyot S, Oszmianski J and Moutounet M (1995) Reactions of enzymically generated quinones in relation to browning in grape musts and wines. In: Lee CY and Whitaker JR (eds) *Enzymatic Browning and its Prevention*, pp. 130–143. Washington: American Chemical Society.

Flurkey WH, Ratcliff B, Lopez L, Kuglin J and Dawley RM (1995) Differentiation of fungal tyrosinases and laccases using selective inhibitors. In: Lee CY and Whitaker JR (eds) *Enzymatic Browning and its Prevention*, pp. 81–89. Washington: American Chemical Society.

Lerch K (1995) Tyrosinase: molecular and active-site structure. In: Lee CY and Whitaker JR (eds) *Enzymatic Browning and its Prevention*, pp. 64–80. Washington: American Chemical Society.

Macheix JJ, Fleuriet A and Billot J (1990) *Fruit Phenolics*. Boca Raton: CRC Press.

Marques L, Fleuriet A and Macheix JJ (1995) Fruit polyphenol oxidases. New data on an old problem. In: Lee CY and Whitaker JR (eds) *Enzymatic Browning and its Prevention*, pp. 90–102. Washington: American Chemical Society.

Mayer AM (1987) Polyphenol oxidases in plants. Recent progress. *Phytochemistry* 26: 11–20.

McEvily AJ and Iyengar R (1992) Inhibition of enzymatic browning in foods and beverages. *Critical Reviews in Food Science and Nutrition* 32: 253–273.

Nicolas J, Cheynier V, Fleuriet A and Rouet-Mayer MA (1993) Polyphenols and enzymatic browning. In: Scalbert A (ed.) *Polyphenolic Phenomena*, pp. 165–175. Paris: INRA Editions.

Nicolas JJ, Richard-Forget FC, Goupy PM, Amiot MJ and Aubert SY (1994) Enzymatic browning reactions in apple and apple products. *Critical Reviews in Food Science and Nutrition* 34: 109–157.

Vamos-Vigyazo L (1995) Prevention of enzymatic browning in fruits and vegetables. A review of principles and practice. In: Lee CY and Whitaker JR (eds) *Enzymatic Browning and its Prevention*, pp. 49–62. Washington: American Chemical Society.

Walker JRL (1995) Enzymatic browning in fruits. Its biochemistry and control. In: Lee CY and Whitaker JR (eds) *Enzymatic Browning and its Prevention*, pp. 8–22. Washington: American Chemical Society.

Zawistowski J, Biliaderis CG and Eskin NAM (1991) Polyphenol oxidases. In: Robinson DS and Eskin NAM (eds) *Oxidative Enzymes in Foods*, pp. 217–273. London: Elsevier Applied Sciences.

Enzymatic – Technical Aspects and Assays

J Nicolas and C Billaud, Conservatoire National des Arts et Métiers, Paris, France
M-A Rouet-Mayer, Centre National de la Recherche Scientifique, Gif sur Yvette, France
J Philippon (retired), Centre National du Machinisme Agricole du Génie Rural, des Eaux et des Forêts, France

Background

Enzymatic browning affects numerous plant organs and seafoods which are rich in oxidizable phenols such as fruits, vegetables, aromatic plants, fungi, and shrimps. It occurs during various procedures used in the food industry: deep-freezing, dehydrofreezing, freeze-drying, hot-air drying, etc. In practice, there are three approaches to prevention of this browning: varietal selection, physical methods, and chemical inhibitors.

The various methods currently used or likely to be used in the future for prevention of browning are reviewed in this article. Deep-frozen fruits and vegetables are considered in detail, since enzymatic browning is most difficult to resolve in these products. Also, the preventive methods used during industrial drying and freeze-drying, and even canning and bottling, are in fact very similar.

A brief review is then given of the methods of measurement of browning and of the questions raised by objective assessment.

Practical Methods of Preventing Browning

Choice of Raw Material

Some plant tissues, such as fruits or vegetables, are less sensitive to browning than others. Most berries – raspberry, blackcurrant, redcurrant, bilberry – are virtually unaffected. Other species, however, quickly brown on freezing and thawing unless previously subjected to antioxidant treatment. They are generally rich in total phenols and poor in ascorbic acid in cases of fruits with pips (apple, pear, quince), and stone fruits (apricots, peaches, plums).

The phenol content of fruits and vegetables varies naturally according to the degree of ripeness, the soil and climate, but also depends substantially on intraspecific genetic characteristics. The Sunbeam peach variety, for instance, which does not brown, contains 20 times less total phenols than the Elberta variety, which is sensitive to browning. This plainly illustrates the value of the development by genetic selection of new, low-phenol varieties. (*See* **Ascorbic Acid**: Properties and Determination; **Phenolic Compounds**.)

Physical Methods of Prevention

Mechanical Factors

Browning is enhanced by wounding of plant tissue, notably during industrial processing. Upkeep of machine blades is important to insure that fruits and vegetables are cut cleanly.

Effect of Temperature

Cold A marked slowing in enzymatic browning is induced by a reduction in temperature. None the less, color change is still rapid at 0 °C. This means that the most sensitive products are frozen by methods allowing the crystallization temperature of water to be reached as rapidly as possible. This is the case with mushrooms, for which cryogenic techniques are obligatory when they are frozen without prior antioxidant treatment. The same is true for peeled slices of white peach which brown during freezing if the cooling exceeds 10 min.

Once frozen, color change is practically blocked at the temperature of commercial storage (−18 °C). When the temperature rises, browning starts again, and will be more marked if the cellular structures of the plant organ have been severely damaged by freezing, chemical peeling, and slicing.

Heat Heat treatment, or blanching, without doubt constitutes the simplest and most direct method

of enzyme inactivation. It consists of brief immersion (from 1 to 6 min depending on size) of the product in water, boiling syrup, or steam close to 100 °C.

Catechol oxidases are inhibited above approximately 70 °C. Their heat stability is closely pH-dependent; it is maximal at about pH 6 and decreases very rapidly below and above this value.

In the deep-freezing of foodstuffs, blanching is principally applied to vegetables that are eaten cooked (potatoes, salsify, asparagus, etc.), but it is also selected to stabilize frozen fruit purées, particularly of apricots. Blanching is little used with whole or sliced frozen fruits that are eaten raw, since it results in loss of firmness and in flavor changes which are unacceptable to the consumer. This is why blanching is only used for fruit with tissue sufficiently impermeable to oxygen, to escape deep browning. Examples are apricot and peach, for which the brief surface treatment scarcely alters the organoleptic qualities. When blanching of pears and, particularly, apples (20–25% lacunae) is incomplete, marked deep browning occurs on thawing.

Other Physical Methods for Prevention

Electrodialysis, ultrafiltration Electrodialysis, based on dissociation of the water molecule using a bipolar membrane, is applicable for the control of enzymatic browning in food liquids. This technique, applied to cloudy juice extracted from MacIntosh apples, results in the reduction of juice pH to 2 with a complete inhibition of polyphenol oxidase (PPO). After readjustment at pH 3.5 (the initial juice pH), the electrodialysis treatment leads to an improvement in the juice color during storage, a large decrease in mineral concentration, and a slight loss in malic acid without noticeable change in the composition of sugars. The latter modifications result in a less acidic and somewhat milder apple juice.

Ultrafiltration can be used as a clarification process in fruit juice processing. Use of polyethersulfone/polyvinylpyrrolidone membranes permits reduction of the amounts of polyphenols in apple juice, leading to a large improvement of the stability of color and clarity during storage.

High-pressure, ultrasonic waves High-pressure processing appeared in the last decade as an alternative to heat treatments for inactivation of microorganisms and food-quality enzymes (including PPO). However, at room temperature, complete inactivation of PPO requires pressures higher than 400–500 MPa which are not industrially relevant for the moment. Therefore, several processes using high pressures in combination with other treatments, such as acidification, moderate heat treatment, CO_2, and ultrasonic waves, have been proposed to inhibit enzymatic browning in fruits, vegetables, and crustaceans.

Protection from Oxygen

When plant organs cannot be blanched, protection from air constitutes another way of preventing browning. Prevention fails, however, as soon as the anoxic conditions no longer hold, and color changes are even accelerated by the destruction of cell structures caused by freezing and thawing.

In the food industry, products are protected from oxygen using a variety of methods, either singly or combined.

Airtight packages These are indispensable for preventing the browning of sensitive fruits with active oxidase systems. In hermetically sealed metal cans, pitted cherries in syrup do not brown, even at −7 °C, whereas spoilage occurs very rapidly if the container is not airtight. The differences between airtight and nonairtight packages increase as the temperature rises. Browning is also a function of the degree of can filling, and hence the quantity of residual air in the headspace.

Partial vacuum A partial vacuum of 380 mmHg is enough to block the browning of frozen peaches in syrup stored at −7 °C. However, stronger vacuums should be avoided since, on thawing, raw fruit will have an unpleasant spongy consistency, and fruit packaged in syrup will tend to acquire an unwanted translucent appearance. (*See* **Chilled Storage**: Packaging Under Vacuum.)

Oxygen-poor atmospheres Browning can also be avoided by keeping the foodstuff in oxygen-poor atmospheres. This can be achieved chemically, or by use of an inert gas. The chemical approach involves the enzymatic oxidation of the sugar present in the medium (enzymatic deoxygenation), thus leading to the consumption of oxygen trapped within an airtight package. The sugar combines with oxygen under the effect of added glucose oxidase, giving gluconic acid. This process has been shown to be effective in laboratory tests with peaches, but to our knowledge has not undergone industrial development.

The use of carbon dioxide or pure nitrogen atmospheres also effectively protects frozen fruits against enzymatic browning. Nitrogen, furthermore, affords better protection of the original flavor and aroma.

Minimally processed products (i.e., ready-to-use, cut fruits and vegetables) are very susceptible to enzymatic browning since they are wounded, living

tissues. In this case, complete removal of oxygen is inapplicable due to the risk of metabolic deviations caused by anaerobic conditions. According to the type of plant material, the packaging film has to be chosen by taking into consideration both the film-selective permeability towards oxygen, carbon dioxide and water on the one hand, and the plant respiratory metabolism on the other, in order to obtain the optimum atmosphere composition allowing an increased shelf-life upon storage at refrigerated temperatures.

Sugaring of fruits In this process, fruit is immersed in a covering liquid, either of a chosen concentration of sugar syrup, or of the fruit's own juice drawn out by contact with dry sugar.

Sugar (generally sucrose) has several effects. It increases the osmotic pressure of the surrounding liquid, which contributes to the reduction in size of ice crystals on freezing, thereby protecting cellular structures. It also increases the viscosity of the medium, thus slowing the rate of enzymatic spoilage reactions. It may also have a slight inhibitory effect on the enzymes of browning when the syrup concentration exceeds 20% (w/w). (*See* **Sucrose**: Properties and Determination.)

Sugar is also a well-known and valued enhancer of taste and aroma. However, it does result in some dilution of aroma when used in the form of syrup. This defect is attenuated when sugar is added to fruit in the dry form, but severe changes in consistency result, thus harming the fruit's appearance. In addition, the exuded juice collects in the bottom of the package and only effectively protects part of the package contents.

In commercial practice, sugaring is only used with frozen fruits intended for the manufacture of juice, syrup, liqueur, or purée; the concentrations used range from 1:9 to, more commonly, 1:4 (dry sugar: fruit).

Frozen fruits intended for direct consumption are frequently packaged under a covering syrup. The syrup concentrations are about 30–40°Brix for cherries, pears, and plums, 40–50°Brix for bananas, apples, and nectarines, and 40–60°Brix for strawberries, apricots, and peaches. None the less, syrup alone inadequately protects the most sensitive fruits, and an antioxidant is therefore incorporated. The most used antioxidant is ascorbic acid, and its derivatives, although other substances such as citric acid are employed, notably in the case of peeled, frozen bananas. The covering liquid comprises either a syrup of pure sucrose, or a mixture of sucrose and maize glucose. (*See* **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants.)

Chemical Methods of Prevention

Numerous chemical compounds inhibit enzymatic browning. Of these, few combine effectiveness with recognized innocuousness, unaltered product flavor, and reasonable cost. The use of most of these compounds is banned by legislation.

In the case of plant foodstuffs, the legislation of most countries only permits ascorbic acid and its derivatives, sodium chloride and, within stricter limits, citric acid and sulfur dioxide (as well as its derivatives, sodium sulfite and sodium metabisulfite). Recently, the GRAS (generally recognized as safe) status has been attributed to the 4-hexylresorcinol for use in the prevention of shrimp melanosis (black spot). (*See* **Legislation**: Additives.)

Ascorbic Acid

Ascorbic acid, which, with its first oxidation product dehydroascorbic acid, constitutes vitamin C, reduces *o*-quinones progressively as they are formed. Catechol oxidase, which indirectly controls the oxidation of ascorbic acid, is virtually unaffected during the reaction, so that browning begins as soon as the reductant has been consumed. (*See* **Ascorbic Acid**: Properties and Determination.)

In practice, concentrations of 0.1–0.3% prove adequate when, as is often the case, the external conditions are unfavorable to browning, i.e., low temperature, acidic and poorly oxygenated medium.

Ascorbic acid is effective above all when thoroughly mixed in liquid or paste products (apple juice, purée). It is also used at concentrations of 0.2–0.25% in syrups covering frozen fruit to enhance the protective effect against enzymatic browning.

One of the drawbacks of ascorbic acid (apart from its relatively high cost) is that it is readily oxidized and rapidly disappears from the medium. To counter this rapid disappearance various derivatives have been proposed as substitutes, such as ascorbic acid-2-phosphate (AAP), or ascorbic acid-triphosphate (AATP). These compounds are not reducing agents, and therefore remain stable in the presence of oxygen, but they do hydrolyze progressively, releasing ascorbic acid under the action of the phosphatases present in the plant tissue. Such compounds are not, however, effective if the medium is too acid (e.g., apple juice, pH 3.7), since the activity of phosphatases is then too low. They are effective, though, in solutions of soaked, sliced apples (0.8%) in cold storage.

Ascorbyl palmitate is a fat-soluble substance which over a long period is more effective than ascorbic acid in fruit juice when used at a concentration of 200 p.p.m. Its poor aqueous solubility, however, makes it ineffective in soaking solutions.

Sodium Chloride

Salt is frequently used in the food industry to slow microbial growth (bacteria, fungi), but is inadequately protective at low concentrations and imparts an unwanted salty flavor at high concentrations. At a pH below 5.5, however, salt is a good inhibitor of enzymatic browning, hence the value of using it in conjunction with citric acid for neutral plant products, such as cultivated mushrooms. (*See* **Preservation of Food**.)

Sodium chloride is sometimes used at low concentrations (1–2%) to afford temporary protection from surface browning of sliced, peeled apples intended for freezing, or to be sold as ready-to-use fresh foods. The soaking time is short, usually less than 1 min.

Used in combination with ascorbic acid, sodium chloride prolongs the duration of action of the acid.

Citric Acid

Oxidative enzyme activity is very pH-dependent (*See* previous article **Browning**: Toxicology of Nonenzymatic Browning). Highly acidic (pH < 2.5) or basic (pH > 10) media which result in protein denaturation destroy its activity. Although alkalization of the medium in which fruits are immersed is undesirable (spontaneous browning following nonenzymatic oxidation of *o*-diphenols), acidification can prove advantageous in certain circumstances. Apart from acetic acid used in vinegar production, citric acid is virtually the only substance authorized for this use. It is sometimes used in deep-freezing and in fruit conserves in syrup; thus, it is used to neutralize the rinsing water of halved, chemically peeled peaches, and at the same time to protect the fruit temporarily against enzymatic browning. The dose used is about 1% of the volume of water. (*See* **Acids**: Natural Acids and Acidulants; **Enzymes**: Functions and Characteristics.)

Sulfur Dioxide and its Derivatives

Sulfur dioxide, which is an antiseptic widely used in foodstuffs, is by far the most effective of currently authorized chemical inhibitors of browning. It even acts at very low concentrations and is inexpensive. It can also be used with certain fruits to eliminate unwanted red pigmentation, as in the case of Bigarreau Napoléon cherries, used in confectionery, which must first be decolorized using sulfite to insure that the pigment does not turn brown during heat treatment.

As an inhibitor of browning, sulfur dioxide has three actions: formation of colorless addition products with (colored) *o*-quinones; reduction of *o*-quinones to their original *o*-diphenols (colorless); denaturation of catechol oxidase (*See* **Browning**: Toxicology of Nonenzymatic Browning). Sulfur dioxide also reacts with many other chemical constituents of plant tissues (notably ketones and aldehydes). The strength of the chemical bonds formed varies with the pH and the type of molecule involved. It would appear that such products are the cause of the abnormal flavors which sometimes develop when sulfur dioxide is used at too high a concentration and is not subsequently eliminated.

Sulfur dioxide can be used in the form of a gas or in aqueous solution as sodium sulfite or disulfite. As a disulfite, it acts less rapidly, but is more easily controlled and results in a few flavor changes.

Of deep-frozen products, apples and apricots are most suited to sulfur dioxide treatment. Fruit slices or halved fruits are soaked for 3–4 min in solutions containing 0.4–0.5% sulfur dioxide.

High doses are sometimes authorized and used to stabilize fruit pulp intended for the making of jams of standard quality (antiseptic action) and to clear fruits before drying (notably apricots). In this case, however, these semiprocessed products will lose most of the treatment compound during subsequent operations. Nevertheless, the tendency of national and international regulations is progressively to restrict the use of sulfur dioxide, or even to ban it altogether, since it has been shown on occasion to be implicated in severe disorders in asthmatics.

4-Hexylresorcinol

4-Hexylresorcinol is probably the most promising among the recently identified browning inhibitors. This compound has obtained the GRAS status for use in the prevention of shrimp melanosis. In this case, a low concentration of 4-hexylresorcinol, up to 25-fold lower than sulfites, is required to obtain a comparable browning inhibition. Moreover, several studies have shown that this compound used either alone or in combination with ascorbic acid could adequately replace sulfites in different plant foodstuffs such as apple slices, potatoes, avocadoes, or apple and grape juices.

Assays for Evaluation of Browning

Accurate methods are required for the measurement of browning in tissue slices and extracts. This need is obvious when different cultivars are to be compared for susceptibility to browning or for evaluation of experimental treatments designed to control enzymatic browning.

Basically, two kinds of techniques are available. The first uses absorbance measurements, usually in the 400-nm region, on solutions after extraction and purification of the brown pigments. The second uses

tristimulus reflectance colorimetry and can be applied directly to cut surfaces or fruit purée. Although both methods are easy and rapid, they do have serious disadvantages. (*See* **Spectroscopy**: Visible Spectroscopy and Colorimetry.)

Absorbance measurements evaluate only the soluble pigments. It is well known that, as the reaction proceeds, polymerization occurs (*See* preceding article **Browning**: Toxicology of Nonenzymatic Browning) and the solubility of a large part of the brown pigments decreases. The insoluble entities are eliminated during the filtration and centrifugation steps in the purification process. Moreover, according to the kind of pigments, which in turn depend on the original phenols and on their relative proportions, the wavelength of maximum absorbance ranges between 360 and 500 nm in the visible region. Therefore, absorbance measurements at a single wavelength are poorly correlated with the visual evaluation of browning.

Most of the data obtained in tristimulus reflectance colorimetry are now given as L (lightness), a (greenness or redness), and b (blueness or yellowness) values, or a combination of these factors. However, the values obtained are highly dependent on the method of measurement and on the state of the surface of the examined object. Thus, during browning, variations in tristimulus data are the result of both chemical and physical changes, the relative importance of the two processes being difficult to assess. Most authors use the decrease in lightness δL (i.e., the difference in L values before and after browning) to evaluate the extent of browning. Some authors have proposed a more sophisticated parameter, the color difference (before and after browning) δE, which can be calculated as:

$$\delta E = \sqrt{[(\delta L)^2 + (\delta a)^2 + (\delta b)^2]}$$

This value gives the distance in color space between two colors but not the direction in which they differ. Moreover, neither δL nor δE gives an idea of the resulting perceived color.

In intervarietal comparisons, numerous authors have tried to correlate the extent of browning, measured by one of the above methods with the phenolic content and/or the PPO activity (Table 1). Clearly, there is no clear-cut relationship between browning and any of these chemical parameters for all fruits. Thus, depending on the fruit species, a correlation can be obtained with both the parameters, one of them, or neither. There are two potential explanations for these difficulties. First, methods for evaluation of browning are only approximate and improved specificity is required. Second, the chemical parameters, and especially, the total phenolic content, are probably not sufficient to explain the extent of browning. In the preceding article **Browning**: Toxicology of Nonenzymatic Browning, it has been stressed that PPO has a wide specificity towards phenolic compounds, which are enzyme substrates to varying degrees. Moreover, the pigments resulting from these phenolics vary widely in color intensity and hue. Therefore, the relative balance among the different classes of phenols as well as the total phenol content are probably fundamental in explaining the degree of browning. Studies carried out both on apple cultivars and on model solutions of phenolic compounds oxidized under standardized conditions by purified apple PPO illustrate these remarks. Thus, the browning susceptibility of apple cultivars can be adequately determined by the simultaneous measurements of soluble (absorbance at 400 nm) and insoluble (lightness) brown pigments. Soluble pigments are correlated to the amount of hydroxycinnamic (mainly chlorogenic acid) esters degraded, whereas insoluble pigments are correlated to the flavan-3-ols ((−)-epicatechin and procyanidin B_2) degraded. Similarly, in model solutions containing both chlorogenic acid and (−)epicatechin oxidized by a constant amount of purified apple PPO, the tristimulus color values and the absorbance values at 400 nm are strongly correlated to the initial and degraded amounts of chlorogenic acid and (−)-epicatechin

Table 1 Relations between the extent of browning, phenolic content, and polyphenol oxidase (PPO) activity in different fruit species[a]

		Correlation with[b]		
Species	*Number of cultivars*	*Phenolics*	*PPO*	*Method used*[c]
Apple	7	0	0	S (A = 440 nm)
	3	+	0	S (A = 440 nm)
	4	0	+	S (A = 440 nm)
	6–8	+		R (tristimulus)
Avocado	3	0	+	Visual
	6	0	+	Visual
Banana	3–5	0	+	Visual
Eggplant	3		+	Visual
Grape	9 (red)	0	+	S (A = 430 nm)
	19 (white)	0	0	S (A = 430 nm)
Olive	6	+	+	R (545 nm)
	5		+	S (A = 410 nm)
	5	+	+	S (A = 410 nm)
	9	+		S (A = 400 nm)
Peach	6	+		S (A = 395 nm)
Pear	6	+		R (540 nm)

[a]Adapted from Macheix JJ, Fleuriet A and Billot J (1990) *Fruit Phenolics*, Boca Raton: CRC Press.
[b]0, no correlation; +, positive correlation.
[c]S, Absorbance data; R, reflectance data (either at a single wavelength or tristimulus).

Table 2 Equations relating tristimulus color parameters *L* (lightness), *a* (redness), and *b* (yellowness) to initial (in) and oxidized (ox) chlorogenic acid (CA) and (−)-epicatechin (EP) in binary mixtures obtained by multilinear regression analysis

Tristimulus color parameters	*Initial amount*		*Oxidized amount*	
L	99–21.0 CA_{in}–25.2 EP_{in}	$r = 0.96$	98.5–19.6 CA_{ox}–30.4 EP_{ox}	$r = 0.97$
a	−4.21 + 27.9 CA_{in} + 0.48 EP_{in}	$r = 0.95$	−4.57 + 31.8 CA_{ox}–2.0 EP_{ox}	$r = 0.95$
b	2.58 + 51.5 CA_{in} + 98.2 EP_{in}	$r = 0.98$	4.62 + 41.3 CA_{ox} + 118.3 EP_{ox}	$r = 0.99$

Adapted from Goupy P, Amiot MJ, Richard-Forget F, Duprat F, Aubert S and Nicolas J (1995) Enzymatic browning of model solutions and apple phenolic extracts by apple polyphenoxidase. *Journal of Food Science* 60: 497–501, 505.

Table 3 Equations relating tristimulus color parameters *L* (lightness), *a* (redness), and *b* (yellowness) to initial (in) and oxidized (ox) amount in each phenolic class in apples obtained by multilinear regression analysis

Tristimulus color parameters	*Initial amount*		*Oxidized amount*	
L	97.6–5.61 HD_{in}–3.23 FO_{in} +2.1 FA_{in}–2.39 DC_{in}	$r = 0.87$	98.1–9.29 HD_{ox}–0.04 FO_{ox} + 0.53 FA_{ox} + 0.02 DC_{ox}	$r = 0.99$
a	−0.02 + 2.75 HD_{in} + 0.49 FO_{in} –0.74 FA_{in} –0.9 DC_{in}	$r = 0.92$	−0.47 + 2.99 HD_{ox} + 0.72 FO_{ox} −0.93 FA_{ox} – 1.14 DC_{ox}	$r = 0.92$
b	5.74 + 18.3 HD_{in} + 16.2 FO_{in} –5.46 FA_{in} + 14.9 DC_{in}	$r = 0.84$	3.92 + 38 HD_{ox}– 1.42 FO_{ox} + 2.77 FA_{ox} + 4.88 DC_{ox}	$r = 0.99$

HD, hydroxycinnamic acid derivatives; FO, flavonols; FA, flavanols; DC, dihydrochalcones.
Adapted from Goupy P, Amiot MJ, Richard-Forget F, Duprat F, Aubert S and Nicolas J (1995) Enzymatic browning of model solutions and apple phenolic extracts by apple polyphenoxidase. *Journal of Food Science* 60: 497–501, 505.

(Table 2). Multilinear regression analysis indicates that both phenolics play a role in the darkening (decrease in *L*) and yellowness (increase in *b*). Chlorogenic acid contributes to redness (increase in *a*) whereas (−)-epicatechin is almost without effect. The same kind of experiments carried out on phenolic compounds extracted from nine different apple cultivars shows that the browning extent of the phenolic solutions after oxidation by purified apple PPO is strongly correlated to the initial and degraded amounts of hydroxycinnamic derivatives and flavan-3-ols (Table 3). However, besides these two classes, the flavonols and the dihydrochalcones appear to have a significant influence on the color development (mainly on the *b* value). Lastly, additional factors (other than phenols) may also be influential: some are chemical, such as the acidity and the concentration of reductants (e.g., as ascorbic acid), while others are physical, such as texture. The relative importance of each of these factors probably varies greatly from one species to another.

In conclusion, enzymatic browning can be controlled by physical treatments, which are not applicable in all circumstances, and by the use of chemical compounds. Very few of the latter are presently available to prevent enzymatic browning in foodstuffs, and none is ideal.

Ascorbic acid is a good inhibitor of browning, but cannot always be used and is, moreover, relatively expensive. Citric acid and sodium chloride have only a limited usefulness and are valuable principally as a complement to another treatment. Only sulfur dioxide is effective and inexpensive, but it has undesirable side-effects and research is therefore necessary to find a suitable substitute. For certain food products such as shrimp, 4-hexylresorcinol could be this substitute.

Existing techniques for the measurement of browning provide useful information on differences in color and color intensity. However, these techniques give a very imperfect assessment of the degree of browning as noted by an observer. The main reason is that human ocular acuity in perception of a difference in hue or color decreases sharply as the color of the object darkens or tends towards saturation. Much more work will be necessary to establish for each product a strict correlation between instrumental measurements and consumer perception.

See also: **Acids**: Natural Acids and Acidulants; **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; **Ascorbic Acid**: Properties and Determination; **Browning**: Toxicology of Nonenzymatic Browning; **Chilled Storage**: Packaging Under Vacuum; **Enzymes**: Functions and Characteristics; **Legislation**: Additives; **Phenolic Compounds**; **Preservation of Food**; **Spectroscopy**: Visible Spectroscopy and Colorimetry; **Sucrose**: Properties and Determination

Further Reading

Amiot MJ, Tacchini M, Aubert S and Nicolas J (1992) Phenolic composition and browning susceptibility of various apple cultivars at maturity. *Journal of Food Science* 57: 958–962.

Amiot MJ, Fleuriet A, Cheynier V and Nicolas J (1997) Phenolic compounds and oxidative mechanisms in fruits and vegetables. In: Tomas-Barberan FA, Robins RJ (eds) *Phytochemistry of Fruits and Vegetables*, pp. 52–85.Oxford: Clarendon Press.

Clydesdale FM (1984) Color measurement. In: Gruenwedel DW and Whitaker JR (eds) *Food Analysis: Principles and Techniques I: Physical Characterization*, pp. 95–150. New York: Marcel Dekker.

Goupy P, Amiot MJ, Richard-Forget F, Duprat F, Aubert S and Nicolas J (1995) Enzymatic browning of model solutions and apple phenolic extracts by apple polyphenoxidase. *Journal of Food Science* 60: 497–501, 505.

Macheix JJ, Fleuriet A and Billot J (1990) *Fruit Phenolics*. Boca Raton: CRC Press.

McEvily AJ and Iyengar R (1992) Inhibition of enzymatic browning in foods and beverages. *Critical Reviews in Food Science and Nutrition* 32: 253–273.

Nicolas J, Cheynier V, Fleuriet A and Rouet-Mayer MA (1993) Polyphenols and enzymatic browning. In: Scalbert A (ed.) *Polyphenolic Phenomena*, pp. 165–175.Paris: INRA Editions.

Nicolas JJ, Richard-Forget FC, Goupy PM, Amiot MJ and Aubert SY (1994) Enzymatic browning reactions in apple and apple products. *Critical Reviews in Food Science and Nutrition* 34: 109–157.

Philippon J, Rouet-Mayer MA, Gallet D and Herson A (1982) Précongélation des pêches entières à l'usage des conserveries de fruits au sirop. Conséquences des différentes étapes de la nouvelle filière technologique sur la qualité des fruits décongelés et sur celle du produit élaboré. *Refrigeration Science and Technology* 4: 302–310.:

Sapers GM, Hicks KB, Phillips JG *et al*. (1989) Control of enzymatic browning in apples with ascorbic acid derivatives, polyphenol oxidase inhibitors and complexing agents. *Journal of Food Science* 54: 997–1002, 1012.

Vamos-Vigyazo L (1995) Prevention of enzymatic browning in fruits and vegetables. A review of principles and practice. In: Lee CY and Whitaker JR (eds) *Enzymatic Browning and its Prevention*, pp. 49–62. Washington: American Chemical Society.

BUCKWHEAT

G Mazza and B D Oomah, Pacific Agri-Food Research Centre, Summerland, Canada

Buckwheat

Buckwheat is a crop commonly grown for its black or gray triangular seeds. It can also be grown as a green manure crop, a companion crop, a cover crop, a source of buckwheat honey, and as a pharmaceutical plant yielding rutin. Buckwheat belongs to the Polygonaceae family and is not a true cereal. Like cereals such as wheat, corn, and rice, which belong to the grass family, however, the grain of buckwheat is a dry fruit.

Buckwheat is believed to have originated in central and northeastern Asia and was probably cultivated in China during the fifth and sixth centuries. It was introduced into Europe through Turkey and Russia during the fourteenth and fifteenth centuries and into North America in the seventeenth century. World production of buckwheat peaked in the early nineteenth century, and has declined since then. During the past 5 years (1995–1999) it has averaged about 2.7 million tonnes, with China and the former Soviet Union accounting for 58 and 34% of the global production, respectively. Other major producing countries are Brazil, Poland, the USA, Japan, Canada, France, Bhutan, Republic of Korea, and South Africa.

Types and Cultivars

Although the genus *Fagopyrum* contains at least 15 species of buckwheat, only two are utilized as food or feed, and wild buckwheat (*F. cymosum*), mainly found as tetraploid, is used on a sporadic basis as a green vegetable or as cattle forage. There are three major species of buckwheat: common buckwheat, *F. esculentum*, tartary buckwheat, *F. tataricum*, and perennial buckwheat, *F. cymosum*. Common buckwheat is also known as *F. saggitatum*, and a form of tartary buckwheat may be called *F. kashmirianum*. It is generally believed, however, that perennial buckwheat, particularly the diploid type, is the ancestral form of both tartary and common buckwheat.

Tartary buckwheat is cultivated in the Himalaya region of India and China, in eastern Canada, and occasionally in mountain areas of the eastern United States. It has some frost tolerance and is therefore

grown at higher altitudes where risk of frost damage is greater. The seeds and products made from them are greenish in color and somewhat bitter in taste. It is used primarily as feed or as a mixture of wheat and buckwheat flour, and often used as a source of rutin.

Common buckwheat is by far the most economically important species of buckwheat, accounting for over 90% of the world production. Many types, strains, and cultivars of common buckwheat exist. There are late and early maturing types, Japanese and European types, summer and fall types. Within a given type, there may be strains or varieties with tall or short plants, gray or black seeds, white or pink flowers. In general, however, common buckwheat varieties from different parts of the world may be divided into two major groups. The first group includes tall, vigorous, late maturing, photoperiod sensitive varieties found in Japan, Korea, southern China, Nepal, and India. The second group is rather insensitive to photoperiod, small and early maturing. All of the varieties in Europe and northern China belong to this type.

Prior to 1950, producers of buckwheat planted unnamed strains of buckwheat harvested from their own fields or obtained from their neighbors or local stores. Named varieties developed through plant breeding were first released in the 1950s. Tokyo, the oldest of the named cultivars introduced into North America, was licensed by the Agriculture Canada Research Station, Ottawa, in 1955. Other cultivars licensed for production in Canada are: Tempest, Mancan, and Manor, all developed at the Agriculture Canada Research Station, Morden, Manitoba since 1965. Mancan, which has large, dark brown seeds, thick stems and large leaves, is the Canadian cultivar preferred in the Japanese market because of its large seed, desirable flavor and color, and high yield of groats in the milling process. Recently, these older cultivars have been supplemented by newer cultivars such as AC Manisoba, Koban, and Koto. These new cultivars produce larger seeds and have an increased seed yield.

Cultivars licensed in the USA are Pennquad, released by the Pennsylvania Agricultural Experimental Station and the USDA in 1968, and Giant American, a Japanese-type cultivar apparently selected by a Minnesota farmer.

Cultivars developed in the former USSR since the 1950s include: Victoria, Galleya, Eneida, Podolyaka, Diadema, Aelita, and Aestoria. Representative cultivars from other areas of the world include: Pulawska, Emka, and Hruszowska from Poland: Bednja 4n from Yugoslavia; and Botan-Soba, Shinano No. 1, Kyushu-Akisoba Shinshu, and Miyazaki Oosoba from Japan.

Plant and Seed Morphology

Buckwheat plant is a broad-leaved, erect, herbaceous plant that grows to a height of 0.70–1.5 m. It has a main stem and several branches, and can reach full maturity in less than 90 days. The stem is usually grooved, succulent, and hollow except for the nodes. Before maturity, the stems and branches are green to red in color, and after maturity, they become brown. The plant has a shallow taproot from which branched, lateral roots arise. Its root system is less extensive than that of the cereals and constitutes only 3–4% of the dry weight of the total plant.

Buckwheat has an indeterminate flowering habit. The flowers of common buckwheat are perfect but incomplete. They have no petals, but the calyx is composed of five petal-like sepals that are usually white, but may also be pink or red. The flowers are arranged in dense clusters at the ends of branches or on short pedicels arising from the axils of the leaves. Common buckwheat has plants bearing one of two types of flowers. The pin-type flower has long styles, or female parts, and short stamens, or male parts, and the thrum-type has long styles and short pistils. The pistil consists of a one-celled superior ovary and a three-part style with knoblike stigmas and is surrounded by eight stamens. Nectar-secreting glands are located at the base of the ovary. The plants of common buckwheat are generally self-infertile, as self-fertilization is prevented by self-incompatibility. Seed production is usually dependent on cross-pollination between the pin and thrum flowers. Honey bees and leaf-cutter bees are effective pollinators. They increase seed set and seed yield.

The buckwheat kernel is a triangular, dry fruit (achene), 4–9 mm in length, consisting of a hull or pericarp, spermoderm, endosperm, and embryo. Large seeds tend to be concave-sided, and small seeds are usually convex-sided. The hull may be glossy, gray, brown, or black, and it may be solid or mottled. It may be either smooth or rough with lateral furrows. The hulls represent 17–26% (in tartary buckwheat 30–35%) of the kernel weight. Diploid varieties usually have less hull than tetraploids.

Structure of Kernel

Scanning electron microscopy of the buckwheat kernel has revealed that the hull, spermoderm, endosperm, and embryo are each composed of several layers. For the hull, these are in order from the outside toward the inside: epicarp, fiber layers, parenchyma cells, and endocarp. The spermoderm is composed of the outer epiderm, the spongy parenchyma, and the

inner epiderm. The endosperm is composed of an aleurone layer, 10–15 μm thick, and subaleurone endosperm containing starch granules surrounded by proteinaceous matrix. The embryo, with its two cotyledons, extends through the starchy endosperm. The terminal parts of cotyledons are often parallel under the kernel surface.

Composition

Proximate composition and selected mineral profiles of Mancan, Manor, and Tokyo buckwheat groats are shown in Table 1.

Carbohydrates

Starch is quantitatively the major component of buckwheat seed, and concentration varies with the method of extraction and between cultivars. In whole grain of common buckwheat, the starch content ranges from 59 to 70% of the dry matter. The chemical composition of starch isolated from buckwheat grains differs from the composition of cereal starches. The amylose content in buckwheat granules varies from 15 to 52%, and its degree of polymerization varies from 12 to 45 glucose units. Buckwheat starch granules are irregular, 4–11 μm in size, with noticeable flat areas due to compact packing in the endosperm. The starch has a water binding capacity of 79–104%, a blue value of 0.35, a swelling power and solubility at 60 °C of 2.35 and 0.5%, respectively, an amylograph viscosity at 92 °C of 640 Brabender Units (BU), an enzyme susceptibility (percentage of solubilized starch) of 2.63%, and an initial and final gelatinization temperature of 61.5 and 76.0 °C, respectively.

Buckwheat grains also contain 0.65–0.76% reducing sugars, 0.79–1.16% oligosaccharides, and 0.1–0.2% nonstarchy polysaccharides. Among the low-molecular-weight sugars, the major component is sucrose. There is a small amount of arabinose, xylose, glucose, and probably the disaccharide melibiose. (*See* **Sucrose**: Properties and Determination.)

Proteins

The protein content in buckwheat varies from 7 to 21%, depending on the variety and environmental factors during growth. Most currently grown cultivars yield seeds with 11–15% protein. The major protein fractions are globulins, which represent almost half of all proteins. The globulins consist of a major 13S legumin-like and a minor 8S vicilin-like fractions. The 13S fraction consists of eight subunits with molecular weights between 43 000 and 68 000. The 8S fraction consists mainly of 57 000 polypeptides with a few subunits in the 26 000–36 000 range. The albumin represents about 25% of the total seed proteins and consists mainly of a low-molecular-mass single-chain polypeptide of 8–16 kDa. Glutelin represents only 4% of the total seed protein. Prolamin has been fractioned into at least two peaks by gel filtration of Sephacryl S-200 and into three major and several minor components by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE). (*See* **Protein**: Chemistry.)

Buckwheat proteins are particularly rich in lysine. They contain less glutamic acid and proline, and more

Table 1 Proximate composition, on a dry-weight basis, and selected mineral profile of three dehulled buckwheat cultivars

Assay	*Cultivar*		
	Mancan (± SD)	*Manor (± SD)*	*Tokyo (± SD)*
Moisture (g per 100 g)	16.2 ± 0.9	10.1 ± 0.2	10.9 ± 0.1
Protein[a] (g per 100 g)	14.2 ± 0.6	14.6 ± 0.3	11.9 ± 0.4
Crude fiber (g per 100 g)	1.57 ± 0.30	1.21 ± 0.03	1.57 ± 0.10
Ash (g per 100 g)	1.85 ± 0.01	1.66 ± 0.01	1.39 ± 0.01
Lipids[b] (g per 100 g)	2.6 ± 0.3	2.2 ± 0.3	2.1 ± 0.2
Carbohydrates[c] (g per 100 g)	79.8 ± 1.6	80.3 ± 0.8	83.0 ± 1.1
K (mg per 100 g)	440 ± 0.005	419 ± 0.009	407 ± 0.005
P (mg per 100 g)	359 ± 0.018	347 ± 0.003	262 ± 0.016
Mg (mg per 100 g)	214 ± 0.002	201 ± 0.012	195 ± 0.010
Ca (mg per 100 g)	18 ± 0.74	180.5 ± 10.6	220.5 ± 6.5
Fe (mg per 100 g)	2.5 ± 0.18	21.4 ± 0.3	21.2 ± 0.9
Zn (mg per 100 g)	2.3 ± 0.04	22.0 ± 1.6	22.8 ± 1.1
Mn (mg per 100 g)	1.0 ± 0.02	10.0 ± 0.5	10.2 ± 0.4
Cu (mg per 100 g)	0.03	3.74 ± 0.1	4.3 ± 0.2

[a] N × 6.25.
[b] Soxhlet, petroleum ether for 8 h.
[c] By difference.
Data from Mazza G (1988) Lipid content and fatty acid composition of buckwheat seed. *Cereal Chemistry* 65: 122–126, with permission.

arginine, aspartic acid, and tryptophan, than cereal proteins. Due to the high lysine content, buckwheat proteins have a higher biological value (BV) than the cereal proteins such as those of wheat, barley, rye, and corn. The digestibility of buckwheat protein, however, is rather low, and this is probably due to the high fiber content (17.8%) in buckwheat, which may be desirable in some parts of the world. Buckwheat fiber is free of phytic acid and is partially soluble. (*See* **Dietary Fiber**: Properties and Sources.)

Lipids

Whole buckwheat seeds contain 1.5–3.7% total lipids. The highest concentration is in the embryo and the lowest in the hull, each containing 7–14% and 0.4–0.9%, respectively. Groats or dehulled seeds of Mancan, Tokyo, and Manor buckwheat contain 2.1–2.6% total lipids (Table 1), of which 81–85% are neutral lipids, 8–11% phospholipids, and 3–5% glycolipids. The major fatty acids of buckwheat lipids are palmitic (16:0), oleic (18:1), linoleic (18:2), stearic (18:0), linolenic (18:3), arachidic (20:0), behenic (22:0), and lignoceric (24:0). Of these, the 16 and 18 carbon acids are commonly found in all cereals. The long chain acids, arachidic, behenic, and lignoceric, which represent, on average, 8% of the total acids in buckwheat, are only minor components or are not present in cereals. (*See* **Phospholipids**: Properties and Occurrence.)

Phenolic Compounds

The content of phenolics in hulls and groats of common buckwheat is 0.73 and 0.79%, and that of tartary buckwheat is 1.87 and 1.52%, respectively. Common buckwheat grown in Canada contains 1.2–1.6% of total phenolics with 70–79% of the phenolic acids being etherified. The total flavonoid concentration in common buckwheat seed and hulls varies with cultivar and location, and ranges from 19 to 387 and from 74 to 1314 mg per 100 g, respectively. Six flavonoids, rutin, orientin, vitexin, quercetin, isovitexin, and isoorientin, have been isolated and identified in buckwheat. Rutin and isovitexin are the only flavonoid components of buckwheat seed, while the hulls contain all six compounds. Rutin (quercetin 3-rutinoside), a well-known flavonol diglucoside used as a drug for treatment of vascular disorders caused by abnormally fragile or permeable capillaries, occurs in leaves, stems, flowers, and fruit of buckwheat. On average, about 300, 1000, and 46 000 p.p.m. of rutin are found in leaves, stems, and flowers, respectively, of common buckwheats. Other reported flavonols are quercitin (quercetin 3-rhamnoside) and hyperin (quercetin 3-galactoside). At least three red pigments have been found in hypocotyls of buckwheat seedlings. One of these is cyanidin; the other two are presumed to be glycosides of cyanidin. The phenolic acids of buckwheat seed are the hydroxybenzoic acids, syringic, *p*-hydroxylbenzoic, vanillic, and *p*-coumaric acids. Also in seed are soluble oligomeric condensed tannins, which, along with the phenolic acids, provide astringency and affect the color and nutritive value of buckwheat products. (*See* **Phenolic Compounds**.)

Biological Effects

Common buckwheat and its components are regarded as functional foods in Japan because of its rutin content. Rutin is known to increase the elasticity of blood vessels and therefore prevent hardening of the arteries. According to Chinese medicine, the leaf and stem of tartary buckwheat have a therapeutic function in treating ulcer, hemostasis, and hypertension, may improve vision and hearing, and acts as an adaptogen. In Nepal, consumption of tartary buckwheat is reported to aid in stomach disorders. Consumption of buckwheat has been associated with prevention and treatment of hypercholesterolemia. This effect of common buckwheat is due to its protein, which has a stronger hypocholesterolemic activity than that of soy protein isolate and behaves similar to dietary fiber. Intake of buckwheat protein extract retards mammary carcinogenesis in rats by lowering serum estradiol and reduces body fat because of its low digestibility. Bioactive peptides from buckwheat pollen have an immunostimulatory effect on lymphocytes and may therefore act as an immunomodulator in increasing immunity and curing cancer.

Buckwheat flour has been proposed to have physiological activity in the treatment of hyperlipemia, hyperglycemia, diabetes, and cardiovascular disease. A decrease in blood sugar level has been observed clinically in 75 diabetic patients treated with tartary buckwheat biscuits. Buckwheat bran is known to contain fagopyritols, a group of phytochemicals acting as an insulin mediator, that may have an important pharmacological use in the treatment of noninsulin-dependent diabetes mellitus (NIDDM). Rutin from buckwheat prevents the oxidation of ascorbic acid and protects the endothelium when given with vitamin C. Hence, the combination of fagopyritols and rutin can have a significant effect on the prevention of diabetes.

Buckwheat herb tea standardized to a total flavonoid content of 5% has been reported to prevent edema in patients with chronic venous insufficiency (CVI). This positive pharmacological effect is thought to be due to the capillary action of rutin resulting in

venous microcirculation. Leaves and flowers of buckwheat have been dehydrated and packaged with other essential oils for protection of the vascular system as a result of the rutin content. Tartary buckwheat has been reported to treat periodontitis and gum bleeding because of the high contents of quercetin and rutin. Green flour obtained by milling dried flowering buckwheat plants is already used in Japan as a natural food colorant for pasta, ice cream, and other food products as a source of dietary rutin.

Grading, Handling, and Storage

Grading

In most countries, buckwheat grain is priced according to its physical condition in terms of size, soundness, and general appearance. In Canada, buckwheat is marketed according to grades established under the Canada Grain Act. The grades are No. 1, No. 2, and No. 3 Canada, and Sample. The grade determinants are minimum test weight, degree of soundness, and maximum limits of foreign material (Table 2). Grades No. 1 and 2 Canada must be free from objectionable odors; No. 3 Canada may have a ground or grassy odor but should not be musty or sour. Test weight, seed size and foreign material are determined on a dockage-free sample. Seed size is determined with a No. 8 slotted sieve (3.18×19.05 mm) and is added to, and becomes part of, the grade name, e.g., buckwheat, No. 1 Canada large. Foreign material refers to cereal grains (wheat, rye, barley, oats, and triticale), weed seeds, and other grains that are not readily removable by mechanical cleaners and may include peas, beans, corn, and other domestic or wild weeds. Samples containing in excess of 5% are graded buckwheat, sample Canada (size) account admixture. Damaged seeds include frosted, moldy, distinctly green, or otherwise unsound and dehulled seeds.

In the USA, buckwheat is not marketed under official grades established by the USDA. However, some states (e.g., Minnesota) have Official Grain Standards that specify the use of Grades 1, 2, 3, and Sample. The grade determinants are similar to those of the Canadian grading system. The Russian standards for food buckwheat are: moisture ≤ 14.5, extraneous matter $\leq 3\%$ (including 0.2% mineral admixture of which 0.1% may be stones), pernicious admixtures $\leq 0.2\%$, spoiled grain $\leq 0.5\%$. The Japanese Buckwheat Millers Association prefers buckwheat, which has large and uniform seeds with black hulls and greenish groats.

Handling

The method of handling varies among production areas. In most cases, however, losses and quality changes of the grain occur at various postharvest stages. During harvest, shattering and losses due to germination, animals such as birds and rodents, and infection by molds occur in all countries. Threshing is done with combines or by beating the dried plants against stones or wooden bars, or by trampling the plants under bullock feet, carts, or tractor wheels. Transportation of grain from the field to market results in losses and quality deterioration. Losses during transportation are mainly due to spillage. However, exposure of the grain to rain or frost during transit leads to subsequent spoilage due to infection by microorganisms. (*See* **Cereals**: Handling of Grain for Storage.)

Storage

Like other grain crops, buckwheat is stored to ensure an even supply through time, to preserve the surplus grain for sale to deficit areas, and to be used as seed in the next planting season. Storage of the seeds may be at the farm, trader, market, government, retail, or consumer levels. Storage containers range from sacks to straw huts to bulk storage bins. In developing countries, traditional storage structures include: granaries of gunny, cotton, or jute bags, or structures manufactured from reed, bamboo, or wood and plastered with mud and cowdung. In North America, storage structures include metal, concrete, or wooden bins at the farm level, elevators and annex at

Table 2 Primary-grade determinants of buckwheat (Canada)

Grade	*Minimum seed density (kg hl^{-1})*	*Degree of soundness*	*Maximum limits of foreign material (%)*				
			Stones[a]	*Ergot*	*Sclerotinia*	*Cereal grains*	*Total foreign material*
No. 1	58	Well matured, cool, and sweet	3.0	0.0	0.0	1.0	1.0
No. 2	55	Reasonably well matured, cool, and sweet	3.0	0.05	0.05	2.5	3.0
No. 3	No minimum	May have a ground or grassy odor, but should not be musty or sour	3.0	0.25	0.25	5.0	5.0

[a]Number of kernel size stones in 500 g.

centralized receiving, storage and shipping points, and concrete silos at grain terminals. Bagged buckwheat is highly susceptible to attack by insects and rodents. Bulk storage in bins, elevators, and silos is recommended. (*See* **Cereals**: Bulk Storage of Grain.)

A moisture content of 16% of less is required for safe storage of buckwheat. If the seed requires drying, the maximum temperature of the drying air should not exceed 43 °C (Table 3). This temperature limit applies to seed for both seeding and processing. During storage at ambient temperature and relative humidity, the color of the aleurone layer changes from a desirable light green to the undesirable reddish brown. This undesirable quality change can be reduced by storing the seed at a lower temperature and at a relative humidity below 45%. Seeds can also be stored under controlled atmospheres (97% N_2, 1.5% O_2, and 1.5% CO_2). However, concentration of the major volatiles decreases, while the content of odor-active volatiles such as hexyl acetate and 1-hexanol increases in seeds stored under controlled atmosphere for 60 weeks. (*See* **Storage Stability**: Parameters Affecting Storage Stability.)

Primary Processing

Primary processing of buckwheat includes cleaning, dehulling, and milling. The aim of seed cleaning is to remove other plant parts, soil, stones, weed seeds, chaff, dust, seeds of other crops, metallic particles, and small and immature buckwheat seeds. The extent and sophistication of the cleaning equipment depend largely on the size of the operation and the requirements for the finished product(s). Milling of buckwheat seed can be carried out virtually by any equipment capable of milling cereal grains. Hammer mills, stone mills, pin mills, disk mills, and roller mills have been used to mill buckwheat. Of these, stone mills and roller mills are probably the most extensively used today.

The milling process may be of two types. In the first and most common type, the whole seeds are first dehulled and then milled. In the second type, the seeds are milled and then screened to remove the hulls. When dehulling and milling are separate operations, the seeds are segregated according to size and may be steamed and dried prior to dehulling. Dehulling is carried out by impact or abrasion against emery stones or steel, followed by air or screen separation of groats and hulls. A widely used buckwheat dehuller is built on the principle of the stone-milling with emery stones set to crack the hull without breaking the groat. The effectiveness of this type of dehuller depends on the clearance between the seed cracking surfaces, and for any seed size, there is an optimum setting. The ease of dehulling and percentage recovery of undamaged groats depends on variety and moisture content. From the dehuller, the groats are separated over sieves of different mesh for sizing into whole groats and two or more sizes of broken groats. Flour is produced by passing the groats through stone and/or roller grinders. (*See* **Flour**: Roller Milling Operations.)

When buckwheat seed is to be processed only into flour, and production of groats is not a requirement,

Table 3 Amino acid composition of buckwheat protein and its fractions

Amino acid	*Whole buckwheat*[a]	*Buckwheat groats*[a]	*Light flour*[a]	*Salt-soluble proteins*[b]	*Albumins*[b]
Lysine (Lys)	6.0	5.9	5.7	5.7	5.6
Histidine	2.6	2.6	2.7	2.4	1.6
Arginine	9.2	10.0	7.9	9.0	5.1
Aspartic acid	11.4	11.4	10.5	9.0	6.4
Serine	4.9	4.6	4.7	4.7	4.6
Glutamic acid	18.5	19.3	17.6	16.6	11.2
Proline	3.8	3.8	5.1	3.7	2.2
Half cystine	1.6	1.8	1.8	NB	0.7
Glycine	6.6	6.2	6.2	5.6	3.0
Alanine	4.3	4.4	4.5	4.4	2.3
Valine	5.3	4.9	5.4	4.4	2.3
Methionine (Met)	2.3	2.8	2.8	2.0	9.2
Isoleucine	4.0	3.7	4.2	3.5	2.1
Leucine	6.7	6.2	7.0	6.4	5.6
Tyrosine	2.0	2.1	2.9	2.8	5.7
Phenylalanine	4.8	4.8	4.8	4.4	2.0

[a]Data from Marshall HG and Pomeranz Y (1982) Buckwheat: Description, breeding, production, and utilization. In: Pomeranz Y (ed.), *Advances in Cereal Science and Technology*, vol. 5. St. Paul, MN: American Association of Cereal Chemists, with permission.

[b]Data from Radovic RV, Maksimovic RV, Brkljacic MJ *et al*. (1999) 2S albumin from buckwheat (*Fagopyrum esculentum* Moench) seeds. *Journal of Agriculture and Food Science* 47: 1467–1470, with permission.

the seeds are ground on break rolls or stone mills and then screened to separate the coarse flour from the hulls. The coarse flour is further reduced by a series of size reduction rolls, each grinding operation followed by sifting to fractionate the mixture of particles according to size. The flour yield ranges from 50 to 75%, depending on the size, shape, and condition of the seeds and efficiency of the dehulling and milling operations.

End Products

Buckwheat has multiple uses. It is used to make alcoholic drinks, the liquor prepared from tartary buckwheat being ascribed medicinal qualities. In China, it is used in the production of vinegar. Common buckwheat is also used as a source of nectar for honey production in many countries. Buckwheat is useful as a green manure crop for renovation of low-productivity land. The plant has been used as a smother crop as it germinates rapidly and produces a dense canopy that controls broad-leaved weeds. It also acts as a feed and cover crop for wildlife including deer, wild turkeys, and other birds.

Buckwheat flour is generally dark in color due to the presence of hull fragments. In North America, it is used primarily for making buckwheat pancakes, and is commonly marketed in the form of prepared mixes. These mixes generally contain buckwheat flour mixed with wheat, corn, rice, oat or soybean flours, and a leavening agent. Buckwheat is also used with vegetables and spices in kasha and soup mixes, and with wheat, corn, or rice in ready-to-eat breakfast products, porridge, bread, and pasta products.

In Japan, buckwheat flour is used primarily for making *soba* or *sobakiri* (buckwheat noodles) and *Teuchi Soba* (hand-made buckwheat noodles). These products are prepared at *soba* shops or at home from a mixture of buckwheat and wheat flours. The wheat flour is used because of its binding properties and availability. *Soba* is made by hand or mechanically. In both methods, buckwheat and wheat flours are mixed with each other and then with water to form a stiff dough that is kneaded and rolled into a thin sheet (1.4 mm) with a rolling pin or passed between sheeting rolls and cut into long strips. The product may be cooked immediately, sold fresh, or dried. (*See* **Wheat**: The Crop.)

In Europe, most buckwheat is milled into groats, which are used in porridge, cabbage rolls, or meat products (especially hamburger), or consumed with fresh or sour milk. Buckwheat groats with cottage cheese, sugar, peppermint, and eggs is used as stuffing in a variety of dumplings. Buckwheat flour is used with wheat or rye flour and yeast to make fried specialty products such as bread, biscuits, and other confectioneries. An extruded, ready-to-eat corn–buckwheat breakfast product of high nutritional value is being produced and marketed in western Europe. This product contains over 14% protein and 8% soluble fiber. Similar products have also been developed in Poland and the former USSR.

In most countries, the quality of buckwheat end products is controlled by law. According to Canadian Government Specifications, buckwheat flour must have $\leq 1.5\%$ ash and $\geq 1.1\%$ protein nitrogen on a 14% moisture basis, and contain $\leq 12\%$ moisture when delivered. Class B Pancake Mix should contain more than 40% and less than 50% buckwheat flour with admixtures of 50% wheat, corn, rice, or soybean flour.

See also: **Cereals**: Bulk Storage of Grain; Handling of Grain for Storage; **Dietary Fiber**: Properties and Sources; **Fats**: Classification; **Phenolic Compounds**; **Phospholipids**: Properties and Occurrence; **Protein**: Chemistry; **Storage Stability**: Parameters Affecting Storage Stability; **Sucrose**: Properties and Determination; **Wheat**: The Crop; Grain Structure of Wheat and Wheat-based Products

Further Reading

Campbell CG (1997) *Buckwheat: Fagopyrum esculentum Moench*. Rome: International Plant Genetic Resources Institute (IPGRI).

Campbell CG and Gubbels GH (1986) *Growing Buckwheat*. Agriculture Canada Research Branch. Technical Bulletin 1986–7E.

DeJong H (1972) Buckwheat. *Field Crops Abstracts* 25: 389–396.

Ihme N, Kiesewetter H, Jung F *et al*. (1996) Leg oedema protection from a buckwheat herb tea in patients with chronic venous insufficiency: a single-centre, randomised, double-blind, placebo-controlled clinical trial. *European Journal of Clinical Pharmacology* 50: 443–447.

Institute of Soil Science and Plant Cultivation (eds) (1986) *Buckwheat Research 1986*, Parts I, II and III. Pulawy, Poland: Laboratory of Science Publisher.

Kreft I, Javornik B and Dolisek B (eds) (1980) *Buckwheat Genetics, Plant Breeding and Utilization*. Ljubljana, Yugoslavia: VTOZD za agronomijo Biotech. facultate.

Marshall HG and Pomeranz Y (1982) Buckwheat: description, breeding, production and utilization. In: Pomeranz Y. (eds.) *Cereals '78: Better Nutrition for the World's Millions*, pp. 201–217. St. Paul: MN: American Association of Cereal Chemists.

Mazza G (1988) Lipid content and fatty acid composition of buckwheat seed. *Cereal Chemistry* 65: 122–126.

Nagatoma T and Adachi T (ed.) (1983) *Buckwheat Research 1983*. Miyazaki, Japan: Kuroda-toshado Printing Co.

Oomah BD and Mazza G (1996) Flavonoids and antioxidative activities in buckwheat. *Journal of Agricultural and Food Chemistry* 44: 1746–1750.

Wang J, Liu Z, Fu X and Run M (1992) A clinical observation on the hypoglycemic effect of Xinjiang buckwheat. In: Lin-Rufa, Zhou Ming-De, Tao Yongru, Li Jianying and Zhang Zongwen (eds) *Proceedings of the 5th International Symposium on Buckwheat, 20–26 August 1992, Taiyuan, China*. Beijing: Agricultural Publishing House.

BUFFALO

Contents
Meat
Milk

Meat

C D Khedkar, College of Dairy Technology, Warud (Pusad), Maharashtra, India
G D Khedkar, Dr Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India
G Gyananath, Swami Ramanand Teerth Marathwada University, Nanded, Maharashtra, India
S D Kalyankar, Marathwada Agricultural University, Parbhani, Maharashtra, India

Introduction

The world population of buffaloes has been estimated to be over 140 million heads. It can be seen from **Table 1** that about 97% of these are found in Asia and the Pacific regions. India has 75 million, China 21 million, Pakistan 14 million, and Thailand 6 million, which makes up most of the animals. About 98% of these animals are raised by small farmers owning not more than 5 acres of land and not more than five buffaloes.

Buffaloes (*Bubalius bubalis*) originated from the Indo-Gangetic plains, thrived in Asia, and have been a symbol of life, religion, and endurance. Buffaloes are classified into two classes: swamp buffaloes and water buffaloes. Buffalo raising mainly supports crop farming, especially rice cultivation. It is suggested that water buffaloes were in the service of humans as early as 2500–2100 BC.

The awareness of the potential of water buffaloes for meat has increased in recent years throughout the world due to its high content of lean meat. Buffalo meat is 25% higher in protein than beef and 50% lower in cholesterol. Most buffalo meat is derived from old animals slaughtered at the end of their productive life. As a result, much of the buffalo meat sold is of poor quality, but when buffaloes are properly reared and fed, their meat is tender and palatable.

Average cattle carcass weight in developing countries was much lower than that in developed countries and is not increasing rapidly. Therefore, despite a large increase in the number of animals slaughtered between 1974–76, 1980–82 and 1995–96, the contribution of developing countries to world beef supplies

Table 1 World distribution of buffaloes with slaughter rate and meat production in some Asian countries

Country	*Population (1000 heads)*	*Slaughtered (1000 heads)*	*Slaughter rate as a % of the population*	*Annual meat production (1000 tonnes)*
World	129 283	7461	5.8	1006
Asia	125 413	6435	5.1	857
India	64 500	980	1.5	135
China	19 547	1125	5.8	125
Pakistan	13 070	2450	18.7	245
Thailand	6250	288	4.6	73
Nepal	4500	175	3.9	23
Philippines	2980	289	9.7	46
Vietnam	2800	350	12.5	75
Indonesia	2424	221	9.1	35
Burma	2100	118	5.6	20
Bangladesh	1800	21	1.2	2
Laos	1200	125	10.4	27
Sri Lanka	900			
Kampuchea	685	53	7.7	8
Malaysia	260	37	14.2	7
Iran	230	66	28.7	10
Iraq	145	20	13.8	3

Reproduced from: FAO (1985) *Production Year Book of the Food and Agriculture Organization of the United Nations*. Rome: Food and Agriculture Organization, with permission.

has not increased proportionally. In view of the immense cattle resources of developing countries – some 65.2% of the world's total – a priority development objective in the tropics, where most developing countries are located, must be to increase productivity per head of cattle. Developing countries contributed some 27.8% of world beef and veal production in 1974–76, rising to 33.3% in 1982 and 39% in 1996.

Buffalo Nutrition

Buffaloes are mostly located in countries where land, cultivated forage crops, and pastures are limited. They are mainly raised on crop residues, sometimes supplemented with green fodder or byproducts, available from the processing of cereal grains, oil seeds, fruit, and vegetables. Buffaloes are known to be good converters of poor-quality roughage into milk and meat. They are reported to have a 5% higher digestibility of crude fiber than high-yielding cows and a 4–5% higher efficiency of utilization of metabolic energy for milk production. Buffaloes can gain as much as 1 kg in weight per day in good-quality roughage and concentrates. They digest feed more efficiently than cattle, particularly when feeds are fibrous and high in lignin and cellulose.

Carcass Dressing

The best and most hygienic way to dress a carcass is with the carcass hanging from some kind of hoist. Whether or not the carcass is hung, the nature of the dressing technique and the way that the carcass is cut will depend upon local custom and, of course, upon variations in marketing techniques which may differ even within a community. One may find different categories of butchers, working with different facilities and often situated adjacent to each other, preparing different cuts of meat to suit a particular market.

In major meat-producing countries the whole side of meat is dressed and cut, but in many tropical countries this is not practical. The hot carcass is cut into several recognizable pieces that facilitate transport which may be on a man's head, on a cart, or even in the boot of a taxi. Although such methods expose the carcass cuts to contamination, they do at least insure rapid distribution of fresh meat. Most consumers in the tropics still demand fresh and not chilled or frozen meat.

Hide should be carefully removed, using proper knives to avoid damage to this valuable byproduct. The carcass should be washed after dressing. The stomach and intestine should be cleaned out and washed, and the contents and washings disposed of separately, as is the blood. Stomach and gut casings may be processed in the tripery.

General Examination of the Carcass

This should be done in a methodical manner, whatever the status of the slaughter house or abattoir facilities. The criteria laid down, for example in the UK, regarding methods and stipulations of meat inspection can also apply to meat inspection in a tropical environment. Every carcass should be examined for the following.

State of Nutrition

Physiological leanness resulting from lack of feed and water must be differentiated, as far as possible, from pathological leanness due to disease. In trade cattle in the tropics this may be difficult since, in many cases, cattle on long treks are not only deprived of food, but exposed to tick-borne diseases. In a temperate climate the percentage of water within the meat of normal buffalo is not above 76.5% and the percentage of protein is about 22% or a ratio of less than 4:1. In extremely emaciated animals the ratio may be 80% water and 19% muscle, or greater than 4:1. Emaciated meat has no place in the export market.

Bruising, Hemorrhage, and Discoloration

This condition is undesirable in any carcass, whether it is due to some localized condition, or more generalized conditions, such as helminthiasis, trypanosomiasis, and tick-borne diseases. Edema reduces carcass quality and storage stability of the meat.

Efficiency of Bleeding

This is pertinent to any community since it affects both keeping quality and flavor of meat. Most consumers in the tropics do not eat meat that is improperly bled and because of this local butchers are often highly suspicious of stunning and new slaughter methods. They insist on casting the animal and cutting the jugular vessels or cutting off the head, pithing, or some other slaughter technique that insures good bleeding-out.

Swelling and Deformities

This receives much less attention under slaughter conditions in the rural tropics, where meat is more of a luxury for special occasions, than in the larger abattoirs adjacent to urban communities.

Age and Sex

These factors are also not considered in most tropical countries unless quality meat is required for elite households or special occasions.

Abnormal Odors

These do not arouse particular attention in the cattle and buffalo meat market unless, for example, there is some specific odor encountered, as in putrefying meat or in the sweet-smelling meat of rinderpest-infected animals.

Buffalo Meat Quality

Buffalo carcass has rounder ribs, a higher proportion of muscle, and a lower proportion of bone and fat than beef. Buffalo meat is higher in protein content than the meat of other species. The amino acid composition, especially of critical amino acids of skeletal muscles, is higher than that of the ox. Buffalo beef has more isoleucine, leucine, lysine, phenylalanine, tyrosine, threonine, valine, and histidine but less tryptophan, cystine, and methionine than beef. Except for cystine, the daily allowance recommended by the Food and Agriculture Organization can easily be met by carabeef as the principal source of protein. About 100 g of carabeef lean meat per head per day is enough to meet all the amino acids (except cystine) required for the body. Buffalo meat has about 44% more lysine than egg protein. Buffalo meat and beef are in many ways quite similar, e.g., muscle pH (5.4), shrinkage on chilling (2%), moisture (76.6%), protein (19%), and ash (1%). Buffalo fat, however, is always white and buffalo meat is darker in color than beef because of more pigmentation and less intramuscular fat (2–3%), which leads to 'marbling,' compared with beef (3–4% intramuscular fat). Some of the data regarding carcass components of buffaloes slaughtered at different body weights is given in Table 2.

Detailed Inspection of the Carcass

After a general examination the carcass is subjected to a more detailed examination. The following is a summary of procedures applicable in a tropical environment.

Table 2 Effect of body weight on carcass components of buffaloes

Slaughter body wt (kg)	*Carcass wt (kg)*	*Dressing (%)*	*Meat (%)*	*Fat (%)*	*Bone (%)*
480	277	59.7	66	17	14
446	239	53.6	69	13.1	17.9
356	193	62.8	66.7	12.9	19.7
300	151	51	70	6	19
260	141	55.5	67	12	18
258	139	53.9	67.5	12.7	19.8
161	83	51.4	66.8	9.7	23.5

Reproduced from Mudgal VD and Sharma DD (1993) Buffalo meat. In: *Encyclopaedia of Food Science, Technology and Nutrition*, vol. 1, pp. 521–529. London: Academic Press, with permission.

The Head

1. The tongue should be loosened but not detached and the surface and deep structures inspected.
2. The gums and root of the mouth should be inspected.
3. The lymph glands around the pharynx and large salivary gland behind the angle of the jaw are examined. Lymph glands act as filters of disease agents and, if swollen, hemorrhagic, or abscessed, indicate disease in the tissues that they drain.
4. The cheek muscles should be free from oozing serum and should not be of yellow or greenish color.
5. The eyes should be examined for microbial defects.

Abdominal Cavity

Stomach, intestines, and spleen The outer, and when necessary, the inner surfaces of the stomach and intestines, the surface and substance of the spleen, and the surfaces of the omentum or loose covering over the stomach should be examined. The lymph glands related to these organs are also examined.

Liver The surface and substance of the liver should be examined and the bile ducts incised to check for the presence of the liver fluke. The thick end of buffalo livers should be incised and examined in detail.

Kidneys The renal, lymph glands and adrenal glands should be examined and, when necessary, the kidneys should be exposed and incised.

Uterus and Ovaries The substance of the uterus, its inner and outer surface should be examined.

Chest Cavity

The pluck should be examined before the organs are separated from each other. Lungs should be examined by observation and palpitation and, unless obviously diseased, they should be incised at the base. The heart covering should be opened and the heart examined and, if necessary, incised.

Udder

This should be incised and examined by observation and palpation and the associated supermammary lymph glands should be examined in detail.

Testicles and Penis

The outer surface and substance of the testicles and penis should be examined. The superficial inguinal lymph glands or those situated under the skin in the groin area should be examined in detail.

Feet

The feet should be examined, especially for lesions suggesting foot-and-mouth disease.

Buffalo Meat Products

Buffalo meat is used as a food item in a similar way to other meats such as beef (e.g., grilled, fried, etc.). In South Asia meat curry is also prepared. Buffalo meat has very good emulsification and binding properties and is ideal for sausage making. Buffalo meat has all the properties needed to prepare direct meat products such as pastarma and gabrovi in Bulgaria, and biltong in Nepal and Brazil. The common Asian uses of buffalo meat are kebabs, koftas and tikkis, bheja, curries, corned meat, and sausages.

Preservation of Buffalo Meat

Although fresh meat has the highest nutritional value and is preferred by most consumers, it is frequently preserved and stored for future consumption. The four most widely used preservation methods are drying (including smoking), curing, canning, and refrigeration. In many cases, combinations of these, for example, drying and salting, are used together to produce the desired results.

Drying

Air drying of meat dates back to early human history and, although commercial techniques for the preparation of dehydrated meat are available, these are generally very costly to operate. Air drying is a comparatively simple process, provided it is done in hot, dry weather. Physiological emaciation of cattle, commonly found in starved animals, does not interfere with the final dried product, although the yield from such carcasses may be low. One of the best known dried meat products is biltong. Although simple and cheap to prepare, preslaughter care of animals is essential for the production of biltong that has good keeping qualities. The procedures for biltong production are as follows:

1. The carcass meat is cut into long strips of equal thickness, cutting the fibers lengthwise. The best dried meat is obtained by physically tearing rather than cutting the muscles so that groups of fibers can be dried as a unit.
2. Strips of meat are then salted either by rubbing the salt by hand just before hanging or strips are left covered with salt overnight. Spices and peppers may be added, according to local taste, also saltpeter, which is a mild preservative which also imparts a good red color to the meat.
3. The strips of meat are then hung in the sun on galvanized wire in such a fashion that they should not touch one another, to allow free circulation of air.
4. After about 12 h drying the strips should be removed, straightened, and turned round for the other side to be properly dried. The time taken for good drying will depend upon the weather.
5. Where dried meat is to be kept for periods longer than 4–6 weeks, it may be necessary to smoke it or to add surface insecticide in low concentrations, if this is allowed by local ordinances. Such dried meat is best stored in polythene bags.

The advantages of dried meat are due to its high content of protein, its good keeping qualities if protected from moisture and attacks of insects such as beetles, and its retention of minerals and vitamins after processing. The disadvantages are that exposure to air and light can bring about rancidity of the fat. Exposure to moisture will also lead to decomposition and/or moldiness. Dried meat can be eaten dry, or raw, or it can be treated as fresh meat, after soaking in water.

Salting and Drying

Salted meat, known as charqui, is first cured in salt and then dried, and differs from biltong in that it contains more moisture. When processed, it resembles dried salted fish in appearance. Pieces of fresh meat for initial curing should be 1.5–2.5 kg in weight and about 2.25 cm thick. These are first hung for a period of about 1 h to cool, after which they are then allowed to drain off on slats. Final curing consists of placing the pieces of meat in a series of layers with about 2.25 cm salt between them, to make up a series of meat/salt layers some 1.25 m high. The pile of layers is then covered with planks and heavy stones to press the pile. The next day the layers in the pile are reversed in position, and this reversal of layer position continues for 5 days, after which the meat is ready for drying. This is done by pressing the meat between planks, or by passing it through a sort of wooden mangle. In hot, dry climates, sun drying on bamboo slats or on chicken wire in the early morning and evening is sufficient. The meat should not be allowed to overheat, otherwise the fat will melt. This type of dried, salted meat has good keeping quality. Before eating it should be soaked in water, which is then poured away in order to get rid of excess salt.

Dehydration

Dehydrated meat is prepared using special equipment to control temperature and humidity, as well as being carried out under a vacuum. The process is generally not suited to the developing stages of beef and buffalo meat industries. The chief advantages of dehydrated meat are its light weight, ease of storage and packing, and the reliability and excellence of the water-reconstituted product. Two major processes of preparation are the vacuum contact plate process and the Platt or Zimmerman process. In the Zimmerman process fresh meat is immersed in heated oil under vacuum; the oil acts as a medium for the transfer of heat to the meat. There are certain technical difficulties in removing all the oil from the meat by centrifugation, but this method is cheaper than the vacuum contact plate process.

Heat Sterilization (Canning)

The principle of this process is that the produce is sealed in a container after applying heat to kill microorganisms which may cause spoilage of the product. The process is not dissimilar to that of home bottling of fruit. The basic principles involve heating the food product sufficiently to kill the microorganisms, followed by sealing in a vacuum. The length of heating time will depend on several factors, including the type of microorganisms to be destroyed, the moisture content and acidity of the product, the amount of protection offered by fats or oils, and the quantity of other substances, such as salt, that may be present. Meat for this process must be properly prepared before canning and, if necessary, precooked. Oxygen and air are driven off to provide the vacuum, either by closing the can under vacuum, or by passing the can through an exhaust box, where the air and oxygen are driven off by heat and the can is sealed before cooking.

Smoking

This is one of the oldest methods of preserving meat, the simplest method being to hang slices above an open fire. It imparts good keeping qualities and, provided the right woods are used, a pleasant flavor. Smoked meat does not easily go bad and it is also resistant to flies and beetles. Its longer keeping properties can be improved by a combination of curing and smoking. Smoked meat is often prepared for export in developed countries. For these and to dampen down the smoke, an alternative method is to use a large oil or petrol drum, the bottom of which is removed and the top made into a lid. A fire pit is dug near the drum and the smoke from the pit is directed through a covered trench into the base of the drum, in which the meat is hung. The smoke escapes through the top, where the lid can be adjusted to control its rate of flow. Meat for smoking must not come from fevered or emaciated animals. Such meat remains wet and flabby and does not smoke well.

Meat may be smoked raw or after being cured or previously cooked. It should be cut along the muscle fibers into flat slices to allow good smoke penetration. Large pieces of meat need higher temperatures and a longer period of smoking. Where salt-cured meat is used for smoking, the brine should be washed off the surface of the meat with clean water, otherwise the salt will form a crust on the surface of the meat during smoking.

Refrigeration

Preservation of meat by this method was first used on a commercial basis in northern Europe, using natural ice shipped from the Arctic. Mechanical refrigerators came into use during the latter part of the nineteenth century in order to ship meat from the Argentine, Australia, and New Zealand to Europe. There are now rapid developments in refrigerator transport of meat carcasses from production areas to internal consumer markets and for export in some tropical countries. There are three major temperature ranges at which buffalo meat is processed by cooling. These are: the initial cooling range (0 to $-5\,^{\circ}C$), the chilling range (-3–$10\,^{\circ}C$), and the freezing range (-10 to $-18\,^{\circ}C$).

Curing and Pickling

There are several methods and substances used for preserving by curing. Some of these are used for curing pork and other meats, but only those relevant to beef and buffalo meat will be briefly mentioned here. Curing is not the same as preserving meat by the use of antiseptics. Curing agents such as salt, sugar, or vinegar are harmless for the consumer, and while they do not kill microorganisms they prevent their multiplication. At the same time, they may also impart special flavors or even improve the flavor of some products. Antiseptics and certain other chemical preservatives may kill microorganisms, but if improperly used may be harmful to the consumer, so there is strict legal control over their use. The following ingredients are used for curing and preservation: salt and sugar used in high concentration and also jointly with other curing ingredients, Bengal saltpeter or potassium nitrate, Chile salt peter or sodium nitrate, vinegar, and various spices. Although spices are mainly used to impart flavor, they also have some inhibitory action in high concentrations on microorganisms. Cures may be used dry, when the dry ingredient is rubbed

into the surface of the meat, or in solution as in pickling and brining, or as a mixture.

Utilization of Waste from Buffalo Meat Industry

The waste (excluding meat, bone, fat, and hide) amounts to 25% of the live weight. The list of various wastes obtained from the buffalo meat industry and their uses is given in Table 3.

Microbiological Quality of Meat

Contamination of buffalo meat may occur through the water used in the abattoir and meat-processing plant, equipment used for processing, and personnel handling the product. The microflora consists of bacilli, micrococci, streptococci, and staphylococci, *Escherichia coli* and other coliforms, pseudomonads, salmonellae, clostridia, yeasts, and fungi. The type of microflora depends upon the initial source of contamination of the carcass and additional contamination during subsequent handling of the meat. The standard plate counts of meat processed in plants are reported to be around 10^5 cfu g^{-1}. The lower number of organisms in buffalo meat processed in modern slaughterhouses could be attributed to the better hygienic measures employed during processing.

Table 3 Important biochemicals and pharmaceuticals obtained from the buffalo meat industry

Byproducts	*Biochemicals /pharmaceuticals*
Blood	Plasma, serum, albumin, hematinics, blood-meal
Fibrin	Peptone, fibrin foam, fibrin powder, bioplasm
Pancreas	Insulin, glucagons, pancreatin, trypsin, and chymotrypsin
Lungs and intestinal mucosa	Heparin
Liver	Liver extract
Thyroid	Thyroxine
Pituitary	
Anterior	Follicle-stimulating hormone (FSH), adrenocorticotropic hormone (ACTH), prolactin
Posterior	Oxytocin, vasopressin
Adrenals	Epinephrine (adrenaline)
Testis	Hyaluronidase
Gallbladder	Bile, bile salts
Heart	Cardiolipins, cytochrome C
Spinal cord, brain	Cholesterol, lecithin
Bones	Gelatin, bone morphogenetic protein (BMP)
Hooves	Neats foot oil

Undesirable microbial contamination is possible at any stage if proper procedures are not followed. Certain strains of *E. coli* can cause gastrointestinal disease in humans, either by enterotoxin production or by penetrating the intestinal epithelium. *Staphylococcus aureus* has also been reported in dressed meat. *S. aureus* has been reported to be widely distributed in market meats.

It has been confirmed that the hygienic processing of meat improves the bacteriological quality of meat and reduces the incidence of food-poisoning organisms. The concept and basis of meat hygiene and inspection relate to the fundamental principle that the meat reaching the market is sufficiently well examined before and after slaughter, as well as during the process of cutting, dispatching and retailing, to insure, as far as possible, freedom from diseases which can be transmitted to humans. In developed countries optimum precautions can be taken at all stages, during ante- and postmortem inspection and transportation and processing stages of the carcass and byproducts. In developing countries the rigid criteria of inspection accepted in the developed world cannot be applied to local marketing, although much can be and is being done by local authorities to protect the consumer and abattoir workers from exposure to infection. The following describes some of the diseases carried in buffalo meat.

Bovine Tuberculosis

There is increasing evidence that bovine tuberculosis, which is also transmissible to humans, is becoming more prevalent in the tropics. This may be due to an increase in the incidence of the disease, or may be an indication of improved detection. Since infection is by ingestion of infected meat or milk, adequate inspection of the carcass, with particular reference to the lymph glands, is essential. This is a bacterial disease of both animals and humans and causes gastrointestinal infections as well as lung infections. Meat-borne infections can occur from contamination of meat by tuberculosis-infected human carriers such as butchers and handlers and by contact with meat derived from infected animals during the abattoir routine. Careful antemortem and postmortem inspection combined with good slaughter hygiene are essential to reduce the danger of widespread contamination.

Salmonellosis

Salmonella bacteria affect a wide range of domestic and wild animals, as well as humans. They cause mainly enteric infections and in the tropics human gastrointestinal infections caused by salmonellae are common. Meat-borne infections can frequently occur, not only by contact with meat derived from

infected animals, but also from meat which has been contaminated by human carriers of salmonellae during slaughter procedures. Contamination of meat from infected animals during slaughter is a considerable problem and careful ante- and postmortem inspection combined with a satisfactory standard of slaughter hygiene are essential to reduce the danger of widespread contamination and subsequent dissemination of the bacteria.

Anthrax

This is a highly infectious bacterial disease of animals, quite commonly encountered in tropical countries. Humans can acquire the disease from infected carcasses or byproducts by handling, inhaling contaminated air, or eating infected meat. Anthrax can often be detected by proper antemortem inspection, but the disease can be overlooked. Even an experienced inspector may have difficulty under certain conditions in differentiating anthrax from other septicemic diseases. For these occasions minimal laboratory facilities for early on-the-spot provisional diagnosis are essential.

Once an infected anthrax carcass has been removed unnoticed from the abattoir the disease can spread rapidly. The anthrax bacillus forms resistant bacterial spores which can remain viable in the soil and elsewhere for very long periods. Care should be taken to remove a suspected carcass from the abattoir floor. To avoid further contamination the premises should be disinfected with a 2% caustic soda solution. Carcasses infected with anthrax should be destroyed completely, if possible by burning. If buried, the pit should be sufficiently deep to prevent the carcass being dug up by dogs or other predators. Lime sprinkled on the carcass before burial also helps reduce the danger of the disease spreading.

Cysticercosis

This condition is prevalent in cattle in many tropical areas and is a particular health hazard to humans because infected cattle and buffalo muscles contain the cysts, which are an intermediate or developmental stage of *Taenia saginata*, one of the large human tapeworms. Cattle and buffalo grazing on pastures contaminated with human feces containing these tapeworm eggs can become infected. The muscular tissue of the cattle and buffalo then become infected with the tapeworm cysts and if humans eat improperly cooked beef they become infected and the cycle of development can recur.

Meat inspectors usually cut into the cheek muscles and heart muscles to check for the presence of cysts. In Zebu-type cattle the hump is also a further useful site for muscle incisions.

Hydatidosis

The importance of this disease lies in both its human health and economic aspects. The dog is the most important link in the biological cycle of this tapeworm and for this reason control should be directed at proper treatment of dogs with anthelmintics, keeping them out of slaughter premises, and the proper disposal of the viscera of all slaughter animals. The hydatid cysts are intermediate developmental stages of dog tapeworms. These cysts can develop not only in large and small ruminants and swine, but also in humans, where they may even locate themselves in the brain. From the human infection standpoint, the hydatid cysts in the viscera of sheep and swine are the most dangerous as far as potential infection of scavenging dogs is concerned, because cysts in these animals are much more fertile than those found in cattle and buffalo.

See also: **Canning**: Principles; **Chilled Storage**: Principles; **Curing**; **Dehydration**; **Parasites**: Occurrence and Detection; **Pickling**; **Sterilization of Foods**

Further Reading

FAO (1985) *Production Year Book of the Food and Agriculture Organization of the United Nations*. Rome: Food and Agriculture Organization.

FAO (1988) FAO statistics of livestock population of the countries in Asia Pacific Region. *Asian Livestock* 13: 94.

Hill D (1988) Slaughter methods, meat hygiene and carcass quality. In: *Cattle and Buffalo Meat Production in the Tropics*, pp. 142–149. Longman Scientific and Technical.

Mudgal VD (1987) Prospectus of buffalo as dairy animal in India. *Asian Livestock* 12: 28–32.

Mudgal VD and Sharma DD (1993) Buffalo meat. In: *Encyclopaedia of Food Science, Technology and Nutrition*, vol. 1, pp. 521–529. London: Academic Press.

Milk

C D Khedkar, College of Dairy Technology, Warud (Pusad), Maharashtra, India
G D Khedkar, Dr Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India
M R Patil, College of Dairy Technology, Warud (Pusad), Maharashtra, India
S D Kalyankar, Marathwada Agricultural University, Parbhani, Maharashtra, India

Introduction

Buffalo plays a pivotal role in the agricultural economy of many Asian, European, African, and Latin

American countries. Among farm livestock buffalo is being increasingly recognized as an animal for milk, meat, and draft. Despite its significant contribution as a multipurpose animal, this species has received little of the care and attention which it rightly deserves. A major bottleneck in the enhancement in production potential of buffaloes has been inadequate research and paucity of information. It is often falsely presumed that the scientific information generated on cattle can be extrapolated to buffaloes.

Milk Production

In terms of the percentage of world milk production from buffaloes, India ranks first (65%), followed by Pakistan (22%), China (5%), Egypt (4%), and Nepal (2%). Various compositional and functional properties render buffalo milk (BM) eminently suitable for the manufacture of dairy products such as ultra heat-treated (UHT) cream, dried icecream mixes, dairy whiteners, edible casein, and caseinates. However, from a technological point of view, BM is often considered not an ideal fluid for the manufacture of several types of cheeses, milk powders, evaporated and condensed milk, and infant formulae. Due to several biochemical differences between BM and cows' milk (CM), conventional processing technologies are often unsuitable and cannot be applied directly for processing BM.

Emerging research and development trends in BM processing suggest that there is ample scope for tailoring the technology, particularly in the developing world where buffaloes enjoy a preeminent position in milk production. Several region-specific traditional milk products owe their unique characteristics to BM. The suitability of buffalo as a milk producer is now distinctly gaining importance throughout the world.

Comparative Appraisal of Gross Composition of Buffalo and Cows' Milk

BM is reputed for its richness and creaminess. The compositions of BM and CM are compared in Table 1; BM has a higher content of fat, lactose, casein, whey proteins, and minerals than CM. All of the casein in BM is present in the micellar form while in CM only 90–95% of the casein is in the micellar state and the rest is present in serum phase. The proportion of bovine casein which is micellar depends on the temperature range and the gravitational force used to sediment casein micelles. The size of the casein micelles is bigger (110–170 nm) in BM compared to CM (range 50–500 nm, mean 120–150 nm). The voluminosity of the casein micelles in BM is 2.68–3.72 ml g^{-1} compared to 4.18 ml g^{-1} for CM casein. Similarly, the solvation (hydration of casein micelles) of BM is lower (2.60–2.90 g water g^{-1} casein compared to 3.48 g water g^{-1} of casein from CM). The opacity of BM casein micelles is three times higher than CM casein. BM casein micelles contain higher levels of calcium and magnesium but lower levels of sialic acid and hexose. The lactoferrin content of BM is 32 mg 10 ml^{-1} compared to 15 mg 100 ml^{-1} of CM. No genetic polymorphism is exhibited either by the caseins or whey proteins in BM.

Table 1 Concentration of major constituents in buffalo and cows' milk

Constituents	*Concentration (%)*	
	Buffalo milk	*Cows' milk*
Water	83.18	85.5
Total solids (TS)	16.82	13.5
Solids not fat (SNF)	10.01	9.11
Fat	6.71	4.39
Proteins	4.52	3.3
Lactose	4.45	4.44
Total ash	0.80	0.73
Calcium	0.18	0.12
Magnesium	0.02	0.01
Sodium	0.05	0.05
Potassium	0.11	0.15
Phosphorus	0.10	0.10
Citrate	0.18	0.18
Chloride	0.07	0.1
Ca/P ratio	1.8	1.2

Adapted from Sindhu J (1999) Physico-chemical profiles of cow and buffalo milk in relation to milk processing. In: *Advances in Processing and Preservation of Milk: A Compendium of Short Term Course Notes*, pp. 97–103. Karnal, India: National Dairy Research Institute.

Milk Lipids

BM fat contains higher proportions of high melting triglycerides (9–12%) than CM (5–6%) and as a result it is more solid. Similarly, the proportion of butyric acid-containing triglycerides is higher (50%) in BM than in CM (37%). As a consequence of the higher proportion of butyric acid-containing triglycerides, the emulsifying capacity of BM fat is superior to CM fat. Fat globules are bigger (4.15–4.6 μm) in BM than in CM (3.36–4.15 μm). These are rendered chargeless at much higher pH (4.5–4.6) in BM compared to CM (pH 4.3). BM fat contains less (0.22%) free fatty acids than CM (0.33%). The concentration of unsaponifiable matter (392–398 mg 100 ml^{-1}) is also lower in BM than in CM (414–450 mg 100 ml^{-1}). Similarly, the phospholipid content of BM is also less (21 mg 100 ml^{-1} compared to 37.37 mg 100 ml^{-1}. The total and free cholesterol content is 275 mg and 210 mg, respectively, per

100 g of BM ghee, which is much less than corresponding values of 330 mg, and 280 mg $100\,g^{-1}$ in CM ghee. On the other hand, the esterified cholesterol of BM is much higher (64 mg $100\,g^{-1}$) CM than (48 mg $100\,g^{-1}$).

Milk Salts

The concentration of calcium and magnesium is about 1.5 times higher in BM than in CM. On the other hand the concentration of sodium, potassium, and chloride is lower than in CM. The content of colloidal calcium and magnesium (160 and 9 mg, respectively $100\,ml^{-1}$) in BM is much higher than in CM (80 and 3 mg, respectively, $100\,ml^{-1}$). Only about 20% of calcium and 55% of magnesium in BM is present in a dissolved state compared to 33% and 75% respectively in CM. The Ca/P ratio is much higher in BM (1.8) than in CM (1.2).

Vitamins

BM is a rich source of most water-soluble and fat-soluble vitamins. The average concentrations of vitamins in milk of buffalo, Indian, and western cow are summarized in Table 2; vitamin A content is higher in BM ($340\,IU\,kg^{-1}$) than in CM ($230\,IU\,kg^{-1}$). However, due to the absence of carotenoids and the high fat content in BM, its total vitamin A potency per unit weight of fat is less than CM. Similarly, the tocopherol (vitamin E) content of BM is slightly higher ($334\,ng\,kg^{-1}$ milk) than $312\,ng\,kg^{-1}$ CM. However, due to the higher fat content in BM, its fat is poor in tocopherol ($26\,ng\,g^{-1}$ fat compared to $35\,ng\,g^{-1}$ fat in CM).

Pigments in Buffalo Milk

Biliverdin IX alpha, a latent blue-green pigment, occurs in fresh BM. This pigment is absent from CM and is considered to be an important characteristic of BM. The average concentration of biliverdin in skim milk of Murrah and Surati buffaloes is 51.8 and 65 μg $100\,ml^{-1}$, respectively. The concentration of this pigment in BM varies significantly at different stages of lactation and lactation number. Biliverdin was primarily associated with α-, β-, and κ-caseins and the proteose-peptone fraction of BM. Biliverdin is converted to bilirubin on storage and souring of BM. This pigment binds to BM lipids and imparts the characteristic greenish-yellow appearance to BM fat and butter prepared by the traditional fermentation process.

Other Constituents

BM is rich in taurine ($6\,mol\,l^{-1}$), compared to 4.0 mol l^{-1} in CM. In contrast, the concentration of urea in BM, 17–22 mg $100\,ml^{-1}$ is much lower than the level in CM (37–40 mg $100\,ml^{-1}$). The levels of lipase and alkaline phosphatase are lower in BM than in CM. However, free amino acids are present at a higher concentration (0.44%) in BM than in CM (0.15%).

Differences in Physicochemical Properties of Cow and Buffalo Milk

Precise and indepth information on the physicochemical and functional attributes of milk is an essential prerequisite for automated industrial processing. Most of the prevalent processing technologies apparently originated in the western world, where CM and milk products predominate. Therefore, processing strategies were essentially based on knowledge of the chemistry and functionality of CM. Species-related differences in milk composition have become the subject of topical interest, as processing of milk from species other than cow is being adopted by the dairy industry in several countries. With the emergance of buffaloes as a predominant milk species, particularly in countries of Asia, Africa, and Latin America, coupled with rapid dairy development,

Table 2 Average concentration of some vitamins in milk of Buffalo, Indian, and western cow

Vitamins	*Concentration in milk*		
	Buffalo (Bubalus bubalis)	*Zebu* (Bos indicus)	*Western cow* (Bos taurus)
Vitamin A ($IU\,ml^{-1}$)	340	230	136–157
Thiamin ($\mu g\,ml^{-1}$)	0.2–0.5	0.2	0.2–0.8
Riboflavin ($\mu g\,ml^{-1}$)	1.59	2.33	1.7
Pyridoxine ($\mu g\,ml^{-1}$)	3.25	2.6–3.0	
Ascorbic acid (mg $100\,g^{-1}$)	6.72	1.94	1.65–2.75
Tocopherol ($\mu g\,g^{-1}$)	334.2	312.2	

Adapted from Sahais (1996) Compositional profile of buffalo milk. In: *Buffalo Milk: Chemistry and Processing Technology*, pp. 20–57. Karnal, India: SI Publishers.

Table 3 Physicochemical properties of buffalo and cows' milk

Properties	*Buffalo milk*	*Cows' milk*
Specific gravity (20 °C)	1.03	1.03
Viscosity (in cP at 20 °C)	2.25	1.45
Curd tension (in g)	32.85	28.54
Surface tension (dyn cm^{-1} at 20 °C)	45.5	42.5
Freezing point (°C)	−0.55	−0.53
pH (at 20 °C)	6.7	6.6
Acidity as % lactic acid	0.16	0.15
Redox potential (volts)	0.31	0.26
Electrical conductance (m ohm)	43.22–6.77	4.0–5.50
Thermal conductance (kcal h^{-1} m °C at 37 °C)	0.57	0.46
Heat capacity (cal g^{-1} °C at 30 °C)	0.83	0.93–0.95
Heat-stability (in min at 130 °C)	32.28	32.0

Adapted from Sindhu JS (1999) Physico-chemical properties of cow and buffalo milk in relation to milk processing. In: *Advances in Processing and Preservation of Milk: A Compendium of Short Term Course Notes*, pp. 97–103. Karnal, India: National Dairy Research Institute.

efforts have been made to develop and adopt appropriate technologies for BM processing. It becomes essential to understand the unique physicochemical and functional properties of BM to overcome the various challenges encountered during its processing. BM has higher specific gravity, viscosity, curd tension, pH, oxidation reduction potential (E_h), thermal conductivity, and thermal expansion than CM but its heat capacity and rennet stability are lower than CM. In the fluid state, BM is as stable as CM to various physicochemical factors. In contrast, the heat-stability of concentrated milk product is significantly lower than BM. Owing to differences in physical properties, milk from the two species behave differently when processed for manufacturing of the products. A comparison of the physicochemical properties of BM and CM is given in Table 3.

Comparison Between the Quality of Dairy Products from Cows' and Buffalo Milk

Edible Casein and Caseinates

Due to the higher content of casein in the form of larger micelles and the presence of all of the casein in the micellar state, it is easier to manufacture edible casein and caseinates from BM. The yield of these products is also higher from BM due to lower losses in the whey because of greater size and low hydration of the micelles.

Coffee and Tea Whiteners

Due to higher protein, fat, and calcium content in BM, the yield of whitener from BM is higher. The product is superior due to higher whitening capacity when made from BM. The larger size and greater opacity of casein micelles from BM may be responsible for a better-quality product. The higher emulsifying capacity of BM fat may also be responsible for the better dispersion of whitener used in coffee or tea.

Icecream

Buffalo milk is considered as a better source of fat for icecream due to higher emulsifying capacity. Further, BM ingredients produce better body and texture in icecream.

Infant and Health Foods

Better absorption of fat due to higher emulsifying capacity, better absorption of calcium due to higher concentration of calcium, magnesium, lactoferrin, free amino acids, esterified cholesterol and taurine and lower concentration of sodium, potassium, chloride, urea, and free and total cholesterol in BM compared to CM is beneficial for human nutrition. These attributes make BM superior to CM as an ingredient for infant and health foods provided its curd tension is reduced to improve digestibility.

Certain Varieties of Cheese

Certain varieties of cheese, like Mozzarella and white pickled Domiati cheese, are superior in quality when made from BM. The manufacture of Domiati cheese is easier from BM as it is easier to handle the curd to as it is firmer and the yield is higher.

Yogurt

Buffalo milk is better suited for the manufacture of yogurt as its manufacture is easier and there is no need for prior concentration or the addition of dried milk due to higher total solids in fat.

Problems Associated with Buffalo Milk Cheese

Cheddar cheese is manufactured mainly from CM in major cheese-producing countries. However, in India the major share of milk production is from buffaloes. The adaptation of well-known technology for the production of various products from BM posed many problems, primarily because of its qualitative and quantitative differences. CM, in general, is considered to be the most suitable raw material for cheese. The main problems encountered in the manufacture of hard-type cheese from BM have been faster renneting, lower retention of moisture, slower lipolysis and proteolysis, and poor flavor, body, and texture development. BM cheese is criticized for its higher fat content and hard, rubbery, and dry body and texture.

The defects in BM Cheddar cheese are mainly attributed to the physicochemical and compositional characteristics of milk. The high buffering capacity of BM due to its higher calcium phosphate and casein content is the cause of a slower development of acidity. Faster renneting time may be attributed to its higher colloidal content (about 160 mg $100\,ml^{-1}$ compared to only 8 mg $100\,ml^{-1}$ in CM). The lower retention of moisture in the curd may be the result of low solvation (hydration) of its calcium compared to CM casein. Hard, rubbery, and dry body may be due to the high curd tension which, in turn, is the result of higher content of casein with bigger size of the micelle, high content of calcium and magnesium, more so in the colloidal state, large proportion of solid fat with bigger size of globules, and low voluminosity and solvation of casein micelles compared to the same in CM. The slower rate of proteolysis and lipolysis is the cause of the higher curd tension of BM.

Technology for Manufacture of Buffalo Milk Cheese

In view of the above-mentioned problems associated with the manufacture of BM cheese, research was initiated in the area of cheese in the early 1960s at the National Dairy Research Institute, Karnal, India. Since a major share of milk production is from buffaloes and BM is known to be unsuitable for cheese production, attention was given to this raw material. Cheddar cheese, the most common variety made in India, does not develop proper flavor, body, and texture when it is made from BM. The main problem is the faster rate of syneresis which results in lower moisture content in finished cheese. This in turn adversely affects the three most important biochemical reactions, i.e., glycolysis, proteolysis, and lipolysis, which constitute the major activities in cheese flavor development. In order to overcome this problem, attempts should be made to develop a manufacturing technique which would insure greater retention of moisture and accelerated rate of glycolysis, proteolysis, and lipolysis. A presalting method has been developed which envisages the addition of 1% salt to the cheese milk. This method resulted in the best product. However, addition of salt to the milk makes the whey unsuitable for use in food products.

See also: **Casein and Caseinates**: Methods of Manufacture; **Cheeses**: Types of Cheese; **Ice Cream**: Properties and Analysis; **Milk**: Liquid Milk for the Consumer; Analysis; Dietary Importance; **Yogurt**: The Product and its Manufacture

Further Reading

Kanawjia SK (1998) Modified practices for Cheddar cheese making from buffalo milk. In: *Advances in Cheese and Fermented Milk Products: A Compendium of Short Term Course Notes*, pp. 54–59. Karnal, India: National Dairy Research Institute.

Sahai D (1996) Compositional profile of buffalo milk. In: *Buffalo Milk: Chemistry and Processing Technology*, pp. 20–57. Karnal, India: SI Publishers.

Sindhu JS (1999) Physico-chemical properties of cow and buffalo milk in relation to milk processing. In: *Advances in Processing and Preservation of Milk: A Compendium of Short Term Course Notes*, pp. 97–103. Karnal, India: National Dairy Research Institute.

BULIMIA NERVOSA

D S Goldbloom and P E Garfinkel, Centre for Addiction and Mental Health, Toronto, Ontario, Canada

Definition

The symptom of bulimia refers specifically to binge eating, i.e., the consumption of a large amount of food in a discrete period of time with a sense of loss of control. As an isolated symptom, bulimia may occur in a number of normative and pathological states. Isolated episodes of pleasurable binge eating among adolescents and young adults are relatively common. Pathological binge eating may occur in medical conditions such as Prader–Willi syndrome or idiopathic Parkinson's disease.

The term bulimia nervosa (BN), however, refers to a distinct psychiatric disorder of eating behavior which includes but is not limited to episodes of binge eating. It is also characterized by distinct psychological preoccupations regarding fears of becoming fat and self-esteem yoked to body weight that parallel concerns in its companion disorder anorexia nervosa (AN). Accompanying the episodes of binge eating, and fueled by the fears of obesity, are feelings of loss of personal control and efforts to counteract the effects of ingested calories through a variety of purgative techniques.

Unlike the long-recognized AN, BN was formally characterized as an autonomous disorder as recently as 1979; the formal diagnostic criteria for BN as described by the American Psychiatric Association in 1994 are:

A. recurrent episodes of binge eating. An episode of binge eating is characterized by both of the following:
   1. eating, in a discrete period of time (e.g., within any 2-h period), an amount of food that is definitely larger than most people would eat in a similar time and under similar circumstances
   2. a sense of lack of control over eating during the episode (e.g., a feeling that one cannot stop eating or control what or how much one is eating)

B. recurrent inappropriate compensatory behavior in order to prevent weight gain, such as self-induced vomiting; misuse of laxatives, diuretics, enemas, or other medications; fasting; or excessive exercise

C. the binge eating and inappropriate compensatory behaviors occur, on average, at least twice a week for 3 months

D. self-evaluation is unduly influenced by body shape and weight

E. the disturbance does not occur exclusively during episodes of anorexia nervosa.

There are two subtypes:

1. Purging type: during the current episode of BN, the person has regularly engaged in episodes of self-induced vomiting or the misuse of laxatives, diuretics, or enemas.
2. Nonpurging type: during the current episode of BN, the person has used other inappropriate compensatory behaviors, such as fasting or excessive exercise, but has not regularly engaged in self-induced vomiting or the misuse of laxatives, diuretics, or enemas.

Prevalence

BN is largely but not exclusively a disorder of females, typically between the ages of 16 and 40. The existence of a frequency criterion for binge eating has on the one hand reduced estimates of the prevalence of this disorder since its initial description; on the other hand, evidence exists that this frequency criterion is highly arbitrary and does not capture the full spectrum of the disorder. Approximately 1% of western adolescent and young adult females meet diagnostic criteria for BN.

Etiology and Groups at Risk

The etiology of BN is unknown; physiological, psychological, and cultural cues have been invoked in its pathogenesis and risk factors in these three domains have been identified. The commonest acute precipitant of binge eating *per se* is food restriction, whether experimental or self-imposed, as in dieting or AN. This may result from adaptive physiological signals in the hypothalamus to ingest calories. At the same time, the cognitive consequences of having eaten when trying to abstain from food may provoke a transient relaxation of rules about food that results in a binge-eating episode. Finally, the abundant and frequently instant availability of food in western society may facilitate impulses to binge eat.

At the level of the individual, risk factors include dieting behavior and weight preoccupation,

particularly in the context of AN; up to 50% of AN patients also exhibit bulimic behaviors. Premorbid obesity is a risk factor in the sense that these individuals are likely to be dieting and experiencing the prominent negative social attitudes to obesity. Impulsivity – as manifested by substance abuse, recurrent self-harm behavior, intense and unstable interpersonal relationships, and an inability to tolerate mood states such as depression and boredom – may be an independent risk factor. This cluster of symptoms is often labeled borderline personality disorder.

At the familial level, psychiatric disorders such as depression, substance abuse, eating disorders, and antisocial behavior are overrepresented compared to control populations. It has been argued that BN may represent the cultural shaping of an underlying familial vulnerability to psychiatric disturbance. As with AN, families of individuals with BN may place undue emphasis on the importance of thinness.

At the level of society, the emergence of BN as an autonomous disorder in the latter third of the 20th century parallels the increased emphasis on thinness, dieting, and dieting disguised as thinness, as well as the mushrooming of technology that allows more widespread promulgation of social values. Many women with BN have 'failed' to achieve the weight loss requisite for a diagnosis of AN but share a set of beliefs and values commonly seen in AN. For others, BN represents only one expression of impulsive dyscontrol.

Psychopathology

A morbid fear of fatness and a self-appraisal based largely on body weight and shape are the overriding psychological preoccupations characteristic of BN. Psychometric assessment of eating and related attitudes parallels the findings obtained in AN. It is important to recognize that in BN the act of binge eating seldom occurs in response to normative hunger. The extent to which it represents a physiological response to food deprivation usually reflects intensive dietary restriction for the purpose of weight loss. However, the subjective perception of binge eating falls typically into two categories: counterregulation and distraction. Counterregulation refers to the phenomenon whereby individuals who pursue dietary restraint find they have violated their self-imposed limits; their response is one of resignation to loss of control which then triggers binge eating. In the case of BN, this permission to binge eat is facilitated and perpetuated by the availability of purging behaviors. Distraction refers to the role of binge eating in escaping from or quelling unpleasant or intolerable psychological states, such as depression, anger, boredom, or conflict; indeed, some individuals describe their eating binges as automatic states where they dissociate, losing touch with all feelings. Street drug and alcohol misuse as well as nonlethal self-injurious behavior may serve a similar function.

Clinical Features

As with AN, the commonest precipitant of BN is dieting behavior. However, BN may occur without the emaciation characteristic of AN. This may be especially true where women have been premorbidly overweight. Women with BN who have never had AN still experience up to a 30% fluctuation in adult body weight, compared to the 10% fluctuation among women without an eating disorder. Food that is usually avoided or assigned a negative moral value inevitably becomes food consumed during eating binges where dietary restraint is abandoned.

BN is typically a secretive behavior. Eating binges often occur in the evening after a day of caloric restriction and/or psychological stressors. Individuals may consume between 3000 and 6000 calories in an hour or less, and they describe their eating as rapid, often without savoring the taste, and with a sense of loss of control and inability to stop. Macronutrient analysis of binge food indicates a high quantity of carbohydrate, although perceived high-carbohydrate foods often have a significant fat content as well. A loss of the normal sense of satiety is evident from both subjective ratings and objective measures of the quantity of food consumed. Ironically, the aftermath of a binge-eating episode is often heightened dysphoria, weight concern and self-loathing, which then precipitate purging behavior.

The commonest form of purgation is self-induced vomiting, with individuals using a finger, toothbrush, or pen to stimulate an oropharyngeal vomiting reflex. The commonly available emetic, ipecac, has also been used for this purpose with catastrophic results: ipecac contains emetine, which is directly toxic to cardiac muscle. Other methods of counteracting the effects of ingested calories include laxative, diuretic, and diet pill abuse. These lead to a variety of physiological and psychological complications, including low serum and intracellular potassium levels from losses through vomiting and diuresis; this may precipitate cardiac arrhythmias. An associated state of metabolic alkalosis results from chloride losses through vomiting. Alternating diarrhea and constipation may result from laxative abuse. Diet pills are typically amfetamine-like psychostimulants that not only suppress hunger but also produce anxiety, insomnia, irritability, and dependence. Less obvious forms of purgation include

intensive exercise for the primary purpose of weight loss and severe dietary restriction after binges; this latter feature serves to perpetuate the disorder.

There are few physical characteristics of BN that facilitate its diagnosis. However, two potential signs of the disorder are hypertrophy of the parotid salivary glands (likely secondary to overstimulation) and calluses or erosions on the dorsum of the hand (caused by friction against the incisor teeth while inducing vomiting). The acid content of the vomitus has an erosive effect on dental enamel and may promote gum recession, leading at times to an initial diagnosis of BN by dentists. Feelings of depression and anxiety are common in BN; indeed, these women and their families are at higher than normal lifetime risk for the development of autonomous mood and anxiety disorders. However, when these symptoms coexist with BN, it is often difficult to separate them from other sequelae of the eating disorder.

Because BN is a disorder that generates secrecy, isolation, and shame, affected individuals may wait months or even years prior to seeking treatment. This contributes to the chronicity of the disorder.

Diagnosis

The diagnostic criteria of the American Psychiatric Association listed above reflect the confluence of behavioral and psychological features characteristic of BN. Earlier definitions of this disorder were less restrictive and included a wider range of eating disturbance without regard to binge-eating frequency or weight and shape concerns. This led to epidemiological overdiagnosis without regard to clinical severity. However, problems remain with respect to the definition of a binge-eating episode. For some individuals, the sense of loss of control associated with eating 'forbidden' foods such as two biscuits or a piece of cake may constitute a qualitative or subjective binge, whereas there is consensus that eating more than is normal under the circumstances constitutes a quantitative or objective binge. The accuracy of self-reported binge-eating quantities hampers research in BN.

The diagnosis of BN must be considered in the absence of disclosure by the individual where sequelae such as unexplained low potassium are evident. BN may also contribute to disruption of normal menstrual function among women of a statistically normal body weight.

Treatment

The treatment of BN includes but is not limited to the reestablishment of normal eating behavior, whether through education, psychotherapy, or drug therapy. Unlike AN, individuals with BN are often eager to overcome their eating disturbance, although this is accompanied by a morbid fear of weight gain.

Treatment begins with a careful history which includes longitudinal weight history, including weight prior to the onset of dieting and desired weight; the discrepancy may reflect a conflict between culture and biology. The acute precipitants of binge-eating states are identified and the nonbinge meals are quantified. Methods of purgation and their potentially lethal complications are elicited. The associations between psychological states, eating behaviors, and weight and shape concerns may be facilitated by self-report diaries where individuals with BN discover these connections. Unlike AN, weight gain may not be a goal but weight stabilization may occur as a result of cessation of dieting and bulimic behaviors; typically, these individuals experience significant weight fluctuations.

An educational approach may include readings on cultural bias and body weight, the effect of dietary chaos and deprivation on physiological and psychological states as well as eating behavior, and the biological regulation of body weight. Indeed, research demonstrates that guided reading may be an effective first step in the treatment of individuals with BN. Individual psychotherapy may explore the origins and extent of dysfunctional attitudes related to eating, weight, and shape, as well as more global underlying disturbances related to self-esteem and self-appraisal; this approach, entitled cognitive behavioral therapy, is the most validated treatment for BN and is usually administered over 8–20 sessions. Often a specific meal plan may be prescribed, not to promote weight gain but to avoid long intervals of deprivation that may precipitate binge eating. Learning to tolerate feelings and impulses through understanding, or through learning more adaptive coping strategies, may also be useful. Interestingly, research evidence also supports the efficacy of another short-term intervention called interpersonal therapy, which focuses on current relationships and does not include direct examination of eating behavior.

The use of a wide range of antidepressant drugs has been studied since 1983. The early evidence confirmed that tricyclic antidepressants, monoamine oxidase inhibitor antidepressants, and selective serotonin reuptake inhibitor antidepressants exert an antibulimic effect which is independent of their effect on mood. The most extensively studied and validated such antidepressant is fluoxetine, which is licenced for the treatment of BN in many countries. Their mechanism of action in BN is unclear but may relate to effects on neurotransmitters involved in

hypothalamic regulation of appetitive behavior. Extensive abnormalities of serotonin and norepinephrine (noradrenaline) have been documented in BN. More recent studies which have compared and combined antidepressants and individual cognitive behavioral therapy have suggested that while both are effective alone, the combination confers no particular advantage and there is greater compliance and duration of effect with psychotherapy. To date, no predictor variables exist which will help determine which of these disparate treatments will be most effective for an affected individual.

Hospitalization is rarely necessary for individuals with BN. However, when the disorder is accompanied by suicidal tendencies, serious electrolyte compromise, or coexistent diabetes mellitus, hospitalization may assist in crisis resolution or metabolic stabilization.

Prognosis

The long-term prognosis of BN is relatively less well known than for AN because of the recency of both description of the disorder and the development of effective interventions. It is known that individuals with BN are vulnerable to mood, anxiety, and substance misuse disorders, as well as a general impairment in social adjustment. Some follow-up studies of 5 years indicate that two-thirds of the treated sample recovered from BN; low self-esteem at the onset of treatment and persistent vomiting at the end of treatment have been identified as poor prognostic factors. In contrast to AN, the mortality rate of BN is very low (0.3%) in outcome research.

See also: **Adolescents**; **Anorexia Nervosa**; **Famine, Starvation, and Fasting**; **Premenstrual Syndrome: Nutritional Aspects**; **Slimming**: Slimming Diets; Metabolic Conseqences of Slimming Diets and Weight Maintenance

Further Reading

American Psychiatric Association (1994) *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn. Washington, DC: American Psychiatric Association.

American Psychiatric Association (2000) Practice guideline for the treatment of patients with eating disorders. *American Journal of Psychiatry* 157(suppl.): 1–39.

Brownell KD and Fairburn CG (eds) (1995) *Eating Disorders and Obesity: A Comprehensive Textbook*. New York: Guilford Press.

Fairburn CG and Wilson GT (eds) (1993) *Binge Eating: Nature, Assessment and Treatment*. New York: Guilford Press.

Garfinkel P and Goldbloom D (1993) Bulimia nervosa: a review of therapy research. *Journal of Psychotherapy Practice and Research* 2: 38–50.

Garfinkel P, Lin E, Goering P *et al.* (1995) Bulimia nervosa in a Canadian sample: prevalence and a comparison of subgroups. *American Journal of Psychiatry* 152: 1052–1058.

Garner DM and Garfinkel PE (eds) (1997) *Handbook of Treatment for Eating Disorders*, 2nd edn. New York: Guilford Press.

Goldbloom D (1997) Evaluating the efficacy of pharmacologic agents in the treatment of eating disorders. *Essential Psychopharmacology* 2: 71–88.

Goldbloom D, Olmsted M, Davis R *et al.* (1997) A randomized controlled trial of fluoxetine and cognitive behavioral therapy for bulimia nervosa: short-term outcome. *Behaviour Research and Therapy* 35: 803–811.

Kaplan AS and Garfinkel PE (eds) (1993) *Medical Issues and the Eating Disorders: The Interface*. New York: Brunner/Mazel.

Treasure J, Schmidt U, Troop N, Tiller J, Todd G and Turnbull S (1995) Sequential treatment for bulimia nervosa incorporating a self-care manual. *British Journal of Psychiatry* 167: 1–5.

Bulk Storage *See* **Cereals**: Contribution to the Diet; Dietary Importance; **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; Effects on Fruit and Vegetables

BURNS PATIENTS – NUTRITIONAL MANAGEMENT

J A Marlett, University of Wisconsin-Madison, Madison, WI, USA

Background

Successful nutritional therapy for the person with a thermal injury is among the most challenging and diverse tasks for the nutritionist. Appropriate therapy is determined by the extent and body location of the burn, the presence of related conditions such as inhalation or electrical injuries, and the age and previous health status of the patient. Some epithelial elements remain after second or partial thickness burns, making regeneration of the skin possible in most cases. No epithelial components remain after third-degree or full-thickness burns, and skin grafting or other coverings are needed for wound closure. While partial thickness burns can cause nutrition-related problems, nutritional therapy is an essential component of the treatment of full thickness burns. The focus of this chapter is the nutritional care of patients with 20% or more full thickness burns or more than 25% partial thickness burns who require special nutritional attention. Pediatric and geriatric patients may tolerate less extensive burn, whereas a healthy young adult may tolerate more without special nutrition intervention.

With a severe thermal injury comes hypermetabolism and a catabolic response. Levels of immune and inflammatory mediators and hormones are elevated. Provision of adequate nutrition is critical for adequate treatment, but determining adequate intake is difficult, because the hypermetabolic response includes negative nitrogen balance and inappropriate insulin responses. Fluid status and pathological state are dynamic in the burn patient, which causes continuous fluctuation in nutrient needs. Further, surgery to excise and close wounds, and dressing changes temporarily change energy needs.

The severity of a thermal injury is usually defined by the extent of body surface area (BSA) burned. Although specialized burn units have diagrams that allow accurate calculation of BSA, the 'rule of nines' is a convenient method of estimating the total surface affected. Each upper extremity and the head is considered to be 9% of body surface; each lower extremity, the anterior trunk and the posterior trunk is 18% of body surface. The remaining 1% is assigned to the perineum.

Energy Needs and Sources

Traditionally, the energy needs of the burn patient are estimated using a formula that includes the preinjury weight, basal energy expenditure, and the percentage BSA burned (Table 1). These formulas may overestimate energy needs, and it is now recommended that the individual energy needs of both children and adults be estimated using indirect calorimetry. Respiratory gases are used in an indirect calorimetry formula to determine daily energy needs as kcal per day ($\times 4.18$ for kJ per day) $= 1.05-1.2 \times [3.78$ (oxygen consumed, $l\,min^{-1}) + 1.16$(carbon dioxide produced, $l\,min^{-1}) - 2.98$(urea nitrogen excretion, g per h)]. Complete urine collection is needed for 12–24 h to determine urea nitrogen.

Most clinicians now recommend that protein intake comprise 20% of energy intake; this level also is suitable for the pediatric patient, provided that indices of nitrogen status (e.g., blood urea nitrogen, nitrogen balance, fluid status) are monitored. The most common proportion of fat used for the burn patient is 25–30% of energy, although a lower fat intake has been shown to reduce infectious morbidity, including pneumonia, and duration of hospitalization. If a low-fat diet is used, linoleic acid should provide 4% or more of the kilocalories to prevent essential fatty acid deficiency. Excess fat intake can lead to hypderlipidemia and an immunosuppressive response. The remainder of kilocalories, 50–65% of the total, is carbohydrate.

Micronutrient Needs

Electrolyte requirements are addressed as part of fluid resuscitation and maintenance. Needs for other minerals by the burn patient have not been well defined, and, with the exception of zinc, it is generally thought that body stores, dietary intakes, and the amounts included in a standard multivitamin are sufficient to meet needs during the recovery phase. A daily zinc supplement as 220 mg per day of zinc sulfate (100 mg per day for patients 3 years or younger) is recommended, on the assumption that it will promote healing of burn wounds, as has been shown for other types of wounds. Vitamin intakes provided by the diet are usually supplemented with a standard daily multivitamin. Ascorbic acid is further supplemented at 500 mg (250 mg) twice daily, as is vitamin A at 10 000 IU per day (5000).

Table 1 Equations for prediction of energy needs of burn patients (kcal/24 h)[a]

For adults:

Curreri (Curreri PW, Richmond P, Marvin J, and Baxter CR (1974) *Journal of the American Dietetic Association* 65: 415)

(25 kcal × preburn weight, kg) + (40 kcal × %BSA)

Long (Long C (1979) *Journal of Trauma* 19: 904–906)

Male: 66.5 + (13.7 × weight, kg) + (5 × height, cm) − (6.8 × age, years) × activity factor × injury factor

Female: 665 + (9.6 × weight, kg) + (1.8 × height, cm) − (4.7 × age, years) × activity factor × injury factor

where the activity factor is 1.2 confined to bed and 1.3 out of bed, and the thermal injury factor is 2.1

Toronto (Allard JP, Pichard C, and Hoskino E (1990) *Journal of Parenteral and Enteral Nutrition* 14: 115–118)

−4343 + (10.5 × %BSA) + (0.23 × previous day's kilocalorie intake) + (0.84 × basal energy expenditure) + (114 × previous day's average body temperature) − (4.5 × days postburn)

where basal energy expenditure is determined using the Long equations without the activity and injury factors

Ireton-Jones (Ireton-Jones C, Turner WW, Liepa GU, and Baxter CR (1992) *Journal of Burn Care and Rehabilitation* 13: 330–333)

629 − (11 × age, years) + (25 × weight, kg) − (609 × obesity)

where the presence of obesity is 1, and the absence is 0.

For children:

Mayes (Mayes T, Gottschlich MM, Khoury J, and Warden GD (1996) *Journal of the American Dietetic Association* 96: 24–29)

0–3 years: 108 + (68 × weight, kg) + (3.9 × %BSA)

5–10 years: 818 + (37.4 × weight, kg) + (9.3 × %BSA)

[a]%BSA is the percentage of body surface area burned. Multiply the result of a formula by 4.18 to convert to kilojoules.

Meeting Nutrient Needs

Enteral formulas, which are nutritionally complete liquid diets administered through a tube placed in the gastrointestinal tract, are employed in most instances, as the burn patient cannot usually voluntarily ingest sufficient food to maintain an adequate energy intake. Burn patients have been successfully treated with the end of the tube placed either in the stomach or in the upper part of the small intestine. It is difficult to provide adequate energy with a low-fat enteral diet, because the volumes required are so great. Nutrition by a parenteral or intravenous route may be indicated in special circumstances, for example if the gastrointestinal tract cannot be used. Occasionally, parenteral nutrition may be used in combination with enteral, when use of the latter is not providing adequate nutrition. Parenteral nutrition, however, is not the standard more of nutrition for the burn patient, as it is more costly than, and has not been shown to be superior to, other routes.

Oral intake is the first choice, but it should be used only when the patient can consume sufficient food at meals and in between meal snacks to achieve an adequate energy intake. Nutrient-dense snacks and liquid supplements help to achieve an adequate intake. Occasionally, *ad libitum* oral intake is combined with enteral formula administered overnight to provide adequate nutrition.

Since early introduction of nutrition improves patient response to treatment, nutrient intake is initiated within the first 24–72 h after injury, as soon as fluid and hemodynamic status are stabilized. Early feeding is also important, as it has been shown to help maintain the integrity and motility of the gastrointestinal tract.

Table 2 Pathophysiological complications of the burn patient

Related to thermal injury
Electrical burn
Inhalation damage
Gastrointestinal damage
Infection
Preexisting conditions
Cardiovascular disease
Hypertension
Obesity
Diabetes mellitus

Nutritionally Relevant Complications

Complications that affect nutrition therapy of the burn patient can be divided into three classes: (1) those related to the thermal injury; (2) preexisting metabolic conditions; and (3) those related to the method of feeding, e.g., enteral or parenteral nutrition (Table 2). Electric shock can have substantial neuromuscular effects; the ability to ingest food orally or gastrointestinal motility may be hampered (Table 2). Inhalation injury usually means that the patient will be placed on a ventilator. A very high carbohydrate intake, such as would occur if a low-fat diet were used, can exacerbate ventilator dependency, and the overall macronutrient distribution of energy needs to be evaluated in this context. Oral and esophageal damage also can occur with prolonged exposure to smoke and toxic fumes and would counterindicate oral ingestion of food. A stress ulcer in the stomach, known as Curling's ulcer, which frequently occurs in the burn patient, is now prevented by prophylatic antiacid therapy. The potential for a gastric ulcer and the poor gastric emptying that

also may occur in the patient with thermal injury are reasons why enteral (vs. oral) nutrition is used; the stomach may be bypassed, and the feeding tube is placed in the upper part of the small intestine.

The skin is the primary barrier for the body against the external environment. When it is lost, as in the severely burned patient, body tissues become exposed to the environment, which contains a variety of bacteria and viruses. Current therapy includes a variety of techniques and medications to prevent and treat infection in the burn patient. None the less, a burn patient frequently has cutaneous infections that may pass into the blood, causing septicemia. The fever that accompanies an infection increases energy needs, and septicemia also produces a hypermetabolic state. Metabolic rate is estimated to increase about 10% for each 1 °C increase in body temperature. Nutrient requirements need to be reevaluated in the burn patient when an infection ensues.

A thermally injured patient may have one or more preexisting diseases that affect the overall diet pattern (Table 2). Consideration should be given to lowering dietary fat if the patient has significant cardiovascular disease, and sodium if chronic severe hypertension is present. Obesity complicates the determination of adequate energy intakes, as well as wound closure and healing in the burn patient. During the hypermetabolic phase, protein becomes a primary substrate for energy; there is little value in providing a hypocaloric diet during this phase, as the obese patient will metabolize body protein if adequate amounts are not provided. When normal metabolic pathways resume, consideration can be given to a modest (10–25%) reduction in estimated energy needs when the burn patient is obese. Careful and continual coordination of insulin availability, blood glucose concentrations, and food intake is needed in the burn patient who also has type 1 diabetes mellitus, the form of the disease in which the body does not produce insulin. Similar monitoring is also needed for the person with type 2 diabetes, in whom insulin response to food intake is not normal, but monitoring is less frequent. Extra care also is needed if the patient with diabetes is given a very high (more than 55–60% of kilocalories) carbohydrate diet.

Many of the problems associated with tube feedings can be avoided by the experienced clinician. If the infusion rate is too fast and administration of the feeding too frequent, there can be potential problems, particularly for burn patients, for the simple reason that they have to be given more in order to meet their nutrient needs. Gastrointestinal stress that usually occurs in the burn patient is manifest as a poor emptying of the stomach; thus, 'residuals,' or the amount of feeding remaining 2 h after its administration, need to be determined frequently if the feeding tube is placed in the stomach. It is also important to keep the head of the patient elevated (at least a 30° angle) to prevent aspiration, which is the movement of gastric contents toward the mouth and subsequent entrance into the trachea and ultimately the lungs.

Since the feeding tube may be in place for a long period of time, smaller-diameter tubes are used, because they are less irritating. These smaller-bore tubes, however, cause mechanical problems, because they become more easily clogged. Clogged tubes are prevented by a rigorous maintenance protocol and by not administering ground, solid medications through the tube.

Gastrointestinal complications of tube feedings include nausea and vomiting, diarrhea, constipation, and abdominal cramping or pain, all of which also can be caused by the injury and by medications unrelated to nutrient intake. Many of the gastrointestinal problems can be prevented by maintaining constant administration of the feeding with a pump. Nausea, vomiting, and abdominal cramping may be a result of retention of food in the stomach; if this does not resolve spontaneously within a couple of days, the tube can be repositioned so that food is delivered to the gastrointestinal tract beyond the stomach. Diarrhea can be caused by gastrointestinal dysfunction, too rapid a rate of administration, and bacterial infection in the gut.

Metabolic Modulators

Several new therapies have been proposed to counteract the catabolism and hypermetabolism that accompany a burn injury. However, many of the studies to evaluate these new therapies have been very limited in scope or conducted using animal models or critically ill patients who were not burned. This is an active area of research, but data available at this time do not support the use of most of these therapies. Growth hormone administration does not appear to benefit the burn patient. Very high doses of insulin appear to reduce protein losses, but such doses have not been well accepted by burn centers, as they require frequent blood glucose monitoring. Arginine and glutamine are viewed as conditionally essential amino acids, that is they are indispensible in situations of severe metabolic stress when body synthesis may be inadequate. No convincing evidence exists that they are of benefit specifically in the burn patient, although they are included in some protocols, on the basis that they may help and are unlikely to hamper recovery.

The use of a particular group of enteral feedings, called immune-enhancing therapies, for the burn

patient is clearly controversial, with the results of studies indicating a beneficial effect, no effect, or possibly a detrimental effect on the burn patient. These formulas contain various amounts of dietary nucleotides, arginine, glutamine, carnitine, taurine, and/or ω-3 fatty acids. These formulas are three two four times more costly than standard enteral formulas, and few of these ingredients have been evaluated for their independent effects.

Metabolic complications of enteral feedings include hyperglycemia, hypertonic dehydration, hypernatremia, hyperkalemia, hypokalemia, and hypophosphatemia. Hyperglycemia can be caused by sepsis or severe infection and the hypermetabolism that accompanies a thermal injury. Hyperkalemia may signal compromised cardiac output or renal function. Generally, these metabolic complications are very rarely caused by the enteral feeding in the burn patient, because intake is so closely monitored.

See also: **Children**: Nutritional Requirements; **Coronary Heart Disease**: Intervention Studies; **Dietary Requirements of Adults**; **Electrolytes**: Analysis; Acid–Base Balance; **Energy**: Measurement of Food Energy; Intake and Energy Requirements; Measurement of Energy Expenditure; **Enteral Nutrition**; **Hypertension**: Hypertension and Diet; **Infection, Fever, and Nutrition**; **Parenteral Nutrition**; **Protein**: Digestion and Absorption of Protein and Nitrogen Balance

Further Reading

Berger MM and Shenkin A (1998) Trace elements in trauma and burns. *Current Opinion in Clinical Nutrition and Metabolic Care* 1: 513–517.

Deitch EA (1995) Nutritional support of the burn patient. *Critical Care Clinics* 11: 735–750.

Garrel DR, Razi M, Larivière F *et al.* (1995) Improved clinical status and length of care with low-fat nutrition support in burn patients. *Journal of Parenteral and Enteral Nutrition* 19: 482–491.

Gottschlich MM and Warden GD (1990) Vitamin supplementation in the patient with burns. *Journal of Burn Care and Rehabilitation* 11: 275–278.

Gueugniaud P-Y, Bertin-Maghit M, Petit P and Cardin H(2000) Current advances in the initial management of major thermal burns. *Intensive Care Medicine* 26: 848–856.

Hansbrough JF (1998) Enteral nutritional support in burn patients. *Gastrointestinal Endoscopy Clinics of North America* 8: 645–667.

Heyland DK and Novak F (2001) Immunonutrition in the critically ill patient: more harm than good? *Journal of Parenteral and Enteral Nutrition* 25: S51–S55.

Khorram-Sefat R, Behrendt W, Heiden A and Hettich R (1999) Long-term measurements of energy expenditure in severe burn injury. *World Journal of Surgery* 23: 115–122.

Mayes T (1997) Enteral nutrition for the burn patient. *Nutrition in Clinical Practice* 12: S43–S45.

Raff T, Germann G and Hartman B (1997) The value of early enteral nutrition in the prophylaxis of stress ulceration in the severely burned patient. *Burns* 23: 313–318.

Rodriguez DJ (1995) Nutrition in patients with severe burns: state of the art. *Journal of Burn Care and Rehabilitation* 17: 62–70.

Scaife CL, Saffle JR and Morris SE (1999) Intestinal obstruction secondary to enteral feedings in burn trauma patients. *Journal of Trauma: Injury, Infection and Critical Care* 47: 859–863.

Schloerb PR (2001) Immune-enhancing diets: products, components, and their rationales. *Journal of Parenteral and Enteral Nutrition* 25: S3–S7.

Vandevoort M (1999) Nutritional protocol after acute thermal injury. *Acta Chirurgiae Belgium* 99: 9–16.

BUTTER

Contents

The Product and its Manufacture

E Richards, Bridgend, UK
A M Fearon, The Queen's University of Belfast, Belfast, UK

Background

Butter has long been recognized as a valuable food. Its method of manufacture has changed little over the years except in respect of the equipment used. The traditional craft methods have given way to large-scale sophisticated continuous production, but both can still be found.

In the process of buttermaking, cream is subjected to severe agitation or 'churning.' This causes physical damage to the milk fat globules and results in phase inversion from the oil-in-water emulsion of cream to the water-in-oil emulsion of butter. Butter is composed of a continuous fat phase and a dispersed phase comprising water and globular fat. (*See* **Colloids and Emulsions**.)

Butter is permitted to contain certain other ingredients. These include salt (sodium chloride), coloring agents such as carotenes and annatto, acidity regulators, and starter cultures for the manufacture of lactic or ripened butter.

Production of premium quality butter requires attention to the quality of the raw material, process hygiene, and efficiency. The equipment used in creameries today makes a product of more uniform flavor and texture than the older craft methods and from a better-quality raw material.

Economic forces are emphasizing the need to maximize efficiency and flexibility in processing and production methods. Butter manufacturers have had to reexamine their product properties and investigate technological and other means to enhance these in the eyes of consumers. (*See* **Butter**: Properties and Analysis.)

Types of Butter

Sweet cream:

- salted with a salt content normally about 2%, but can vary from 1.5 to 3%;
- unsalted.

Lactic:

- slightly salted with a salt content of approximately 1%;
- unsalted.

Whey:
Sometimes called 'farmhouse butter.' This is found in localized areas and is manufactured from whey cream – a by-product of cheesemaking. The salt content is about 2%. (*See* **Whey and Whey Powders**: Production and Uses.)

Raw Materials

The quality and handling of the raw material are of paramount importance in achieving a premium-quality end product.

Cream

The essential elements are:

- clean milk;
- efficient separation to a specific fat content;
- efficient pasteurization (heat treatment) and cooling;
- good temperature control during storage;
- care in the physical handling of the cream.

(*See* **Cream**: Types of Cream.)

Raw milk for butter manufacture must contain less than 100 000 microorganisms per milliliter. In fact, milk with a much lower microbiological content (less than 20 000 organisms per milliliter) is now commonly obtained from UK herds. The quality schemes in operation within the UK insure good milk quality, and this is reflected in dairy product standards. (*See* **Milk**: Processing of Liquid Milk.)

Milk is separated into cream and skimmed milk by means of a centrifugal separator. At all stages in the handling of the milk and in the preparation of the cream, care is taken to avoid damage to the fat globule membrane (FGM) by excessive pumping or turbulence. The FGM is composed primarily of protein

and phospholipid and stabilizes the fat phase in milk while also protecting the fat from oxidation and enzyme attack. Separators are designed to cause as little damage as possible, while achieving a high degree of efficiency.

The optimum fat percentage of the cream, to obtain the maximum churning efficiency, needs to be adjusted according to the method and equipment used to make the butter. Traditional methods advise a fat content of 30–35%, whereas modern continuous buttermakers normally operate at 40–44% butterfat for sweet cream and 38–40% butterfat for an acid or cultured cream.

Cold separation (<10 °C) has the advantage that milk may be processed on arrival at the factory. Although the separation is less efficient, free fat levels are low, and the cream has a higher phospholipid content. This improves its whipping properties.

Heat treatment of the cream can be achieved on a small scale by heating in a vat to a temperature of not less than 63 °C and holding the cream at that temperature for at least 30 min. This 'holder method' is slow and inefficient, as heating and cooling are carried out in the same vessel. It is only suitable for small quantities of cream.

For larger-scale production, cream is treated in a continuous high-temperature, short-time (HTST) plate heat exchanger to a minimum of 72 °C for at least 15 s. At higher processing temperatures, there is a danger of oxidative rancidity promoted by migration of copper from the serum to the fat globules. For this reason, it is recommended that cream used for buttermaking be processed at a maximum of 77 °C for 15 s. However, in practice, flash heating to 85 °C is often used to produce a desirable slightly nutty or caramelized flavour. (*See* **Pasteurization**: Principles.)

Feed and weed taints concentrate in the fat phase of milk, and these may be removed by vacreation, a form of heat treatment under vacuum. The heated cream is held in contact with steam under reduced pressure, then separated to allow volatile taints to be drawn off. This is a combined heating and flavor-stripping treatment. Evaporative cooling to approximately 60 °C may also be used to adjust the fat content of the cream. The final cooling of the cream takes place in a plate heat exchanger. A milder, flavor-stripping method using vacuum treatment after an HTST process may also be used.

The cream used for buttermaking must be cooled and stored at a low temperature to encourage crystallization of the fat. This period, usually a minimum of 8 h, is often called 'ageing' the cream. The latent heat released will increase the temperature of the cream to 7–8 °C from the initial cooled temperature of 5 °C. Aging is essential to achieve a low fat loss into the buttermilk and produce butter of the desired texture. (*See* **Butter**: Properties and Analysis.)

Both the rate of cooling and the temperature of holding are important in determining the size of the fat crystals and the proportion of solid-to-liquid fat achieved within the final butter. Large silo tanks for storage are preferable to a series of small cream tanks. There is improved homogeneity, allowing the buttermaking equipment to operate with a more consistent product for a longer period.

Cream temperature treatments ('Alnarping') may be applied to improve butter spreadability and produce a consistent butter despite seasonal variation in milk fat composition. However, these have had limited success. A more effective means has been through altering the cow's diet to increase the proportion of unsaturated fatty acids in milk fat. This reduces butter firmness at low temperatures. Incorporating vegetable oils into cream and applying margarine process technology have been very successful developments, but the resultant products may not be identified as 'butter'. (*See* **Butter**: Properties and Analysis.)

Water

Water, wherever used, must be of the highest possible microbiological quality.

Lactic Cultures

A soured or ripened cream is traditionally required to produce lactic butter. Although it is possible to allow milk or cream to sour naturally, it is neither advisable nor practical. A culture of lactic microorganisms – *Lactococcus lactis* subsp. *cremoris* (formerly *Streptococcus cremoris*), *Lactococcus lactis* ssp. *lactis* (formerly *Streptococcus lactis*), *Lactococcus lactis* biovar. *diacetylactis* (formerly *Streptococcus diacetylactis*) – may be added to the cream to produce the desired acidity, flavor, and aroma. The primary aroma producers are *L. lactis* biovar. *diacetylactis* and *Leuconostoc mesenteroides* ssp. *cremoris*. (*See* **Lactic Acid Bacteria**.)

Salt

Salt adds flavor and also acts as a preservative in sweet cream butter. For short-term storage, bulk butter is stored at temperatures of −18 °C whether it has added salt or not. Intervention (EC-subsidized) butter – for long-term storage – is unsalted and is stored at −25 °C.

Only pure, finely milled, vacuum-dried salt (of at least BSI 998 (1969) or equivalent standard) should be used for butter.

Manufacturing Processes

Churn Method

Batch process A batch-type butter churn may vary in capacity from a few liters up to maximum of about 45 000 l. These churns were originally made of wood (Figure 1) but latterly of stainless steel (Figure 2).

After cleaning and disinfection, the churn must be specially prepared to prevent the butter from sticking to the surface. With wood, this is achieved by scalding with boiling water and immediately cooling with chilled water. This treatment leaves a film of water on the surface of the wood and prevents the butter from adhering to it. All wooden equipment must be kept wet until used.

The surface of stainless steel equipment also requires special physical preparation. The detergents used in the cleaning of such equipment must contain silicates to maintain the special 'nonstick' surface. (*See* **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant.)

Batch butter churns may be barrel- or cone-shaped with fixed or rotating internal 'workers.' As the churn is rotated, the combined actions of rotating and beating cause the cream emulsion to break, forming the butter grains (fat phase) and buttermilk (aqueous phase).

During the first few turns, gases, e.g., carbon dioxide from heterofermentative fermentation, may be liberated from the cream. In order to maintain an even pressure within the churn, it is necessary to release these gases. This is done by depressing a small valve in the lid of the churn.

Each churn has an indicator glass – a small window through which it is possible to see what is happening inside the churn. When hand churning, the cream feels heavier as it begins to thicken. This takes about 15–20 min from the beginning of churning. The cream emulsion breaks and small grains of butter form. These are clearly seen on the indicator glass. The actual size of the butter grains varies according to the type and size of the churn. It is essential not to allow them to grow and form lumps, which will cause an uneven distribution of buttermilk.

For hand churning, the grains should be kept small, approximately 3 mm in diameter – traditionally stated as the size of wheat grains.

Chilled water at approximately 5 °C is added to harden and control the size of these grains, as well as removing the traces of buttermilk. Washing reduces the yield, and is not necessary if the cream is of good quality and all the necessary hygienic precautions have been observed in the preparation of the butter churn.

Salt may be added dry or in the form of brine as a final wash. The addition of brine (10% solution) to butter grains has been used to reduce the need for chilled water. This can be important during warm weather when there is a lack of chilled water. It will

Figure 1 Wooden butter churn (nineteenth century). From National Museums and Galleries of Northern Ireland, Ulster Folk & Transport Museum, with permission.

Figure 2 (see color plate 12) Stainless steel batch butter churn (twentieth century). From APV Unit Systems, Denmark, with permission.

also prevent streakiness due to uneven mixing of the salt. For dry salting, the calculated quantity is sprinkled on to the butter grains to give approximately 2% in the final product.

The butter grains are 'worked' to expel excess moisture, create an even, fine distribution of water droplets and produce a close textured, evenly colored product. This may be carried out using the workers inside the churn or externally on a small scale, by using Scotch hands. These are made of grooved wood (or plastic) and are also used to shape and print attractive designs on the finished packs (Figure 3).

During the period of working, drainage, and addition of dry salt, samples are tested to determine the salt and moisture contents. The operator determines the 'end point' of working when the moisture content is between 15.5 and 16% and by visual assessment of the butter. At this stage, the butter is removed from the churn in readiness for packing.

The moisture content of butter must not exceed the legal maximum of 16%. Manufacturers attempt to be as near to that limit as possible to ensure the maximum yield.

Cultured butters Traditionally, cream is inoculated with specific cultures of bacteria. The reduction in pH and development of flavors produce a lactic-flavored end product. This is regarded by some as a more desirable product, i.e., a fuller-flavored butter. (*See* **Starter Cultures**.)

This method of making butter requires starter culture facilities with the necessary laboratory controls, and additional equipment such as cream-ripening tanks as well as cooling and aging facilities. There must also be a well-controlled program of temperature controls and pH monitoring.

Figure 3 Using Scotch hands to work butter. From National Museums and Galleries of Northern Ireland, Ulster Folk & Transport Museum (L882/10), with permission.

The cream is inoculated with approximately 1% of culture and incubated at 20–27 °C to achieve a final pH of 5.3–4.7, depending on the preferred lactic flavor. The cream is then cooled to stop the fermentation and to attain the desired fat crystallization.

The method of manufacture on a batch basis is no different from that employed for sweet cream. The disadvantage of this traditional system is that the buttermilk has a lactic content that causes problems in its disposal. The behavior of the starter culture is not always consistent, and this can result in end-product variations.

Because of the problems and expense of culturing cream and the disposal of 'lactic' buttermilk, several methods have been developed to make lactic or cultured butter from sweet cream.

The NIZO method developed in the mid-1970s consisted of churning sweet cream and adding a special mix containing cultured whey concentrate and bacterial culture while working the butter grains. This gave the major advantage of being able to manufacture from sweet cream, thus producing sweet cream buttermilk, which has a far greater commercial value than cultured cream buttermilk.

The indirect biological culturing system, as described, for example, by Pasilac – Danish Turnkey Dairies Ltd, involves the addition of two types of prepared starter culture to the churn at the working stage. The combination of aroma-producing bacteria and the acidity of the culture mix result in the final pH and flavor of the butter conforming to that of traditional cultured butter.

The addition of a starter distillate provides an alternative method of flavoring butter without the need for culturing equipment, while acidity is increased by addition of lactic acid.

Continuous Buttermaking

Continuous buttermaking equipment began to be widely used in the 1960s. The success of this equipment was such that, within a decade, most of the batch churns used in commercial butter manufacture had been superseded.

Continuous buttermaking equipment gives an advantage over the batch process in terms of hygiene, uniform product quality, and process efficiency.

Cream processing is an integral part of the whole buttermaking system. The preparation of the cream is similar to that for traditional manufacture. From the storage tanks, it is pumped into the first stage of the buttermaker at a constant speed and temperature.

The capacity of continuous buttermakers varies from small units of $12\,kg\,h^{-1}$ to more than $10\,000\,kg\,h^{-1}$ (Figure 4).

Although design features vary, the basic principles remain the same. A continuous buttermaker consists of: (1) the churning section; (2) the separating section and (3) the working section(s) (Figure 5).

Churning section A multibladed beater operating at high speeds (approximately 1000 rpm) within a cylindrical chamber introduces air into the cream and damages the fat globules. The distance between the cylinder wall and the beater blades is only 2–3 mm. In just a few seconds, the cream emulsion is broken, and the initial agglomeration of fat globules takes place. A mixture of small butter grains and buttermilk is then transported from this beating chamber into the next unit, the separating chamber.

Separating chamber This consists of a rotating drum where the final churning takes place and in which there is a perforated filter – a separation drum – to separate the buttermilk from the butter granules. The drum rotates slowly so that as the buttermilk is drained away, the butter grains are gradually clumped together.

Cooling may be achieved by circulating chilled water in the walls of either churning chamber. In some machines, the first buttermilk is cooled and recirculated. It is in this particular section that the grains of butter are allowed to grow to the required size.

The speed of the beaters, the temperature of churning, and the butterfat content of the cream will all vary slightly. An experienced butter manufacturer will adjust these parameters according to the season, equipment, texture, and consistency of the resultant butter. A firmer milk fat and, therefore, firmer butter are obtained in winter than in summer. In addition, the temperature of the cream has to be maintained at a lower temperature, e.g., 5–7 °C, in summer, whereas it could be at 10 °C in winter.

There is an observation window for the operators – similar to the sight glass of traditional equipment. The control panels may also have display screens, allowing operators to observe the processes inside the machine.

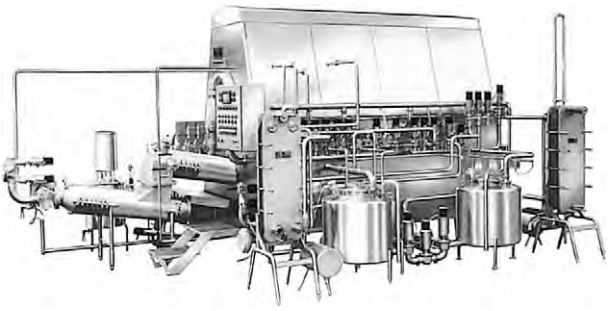

Figure 4 Continuous buttermaking machine. From APV Unit Systems, Denmark, with permission.

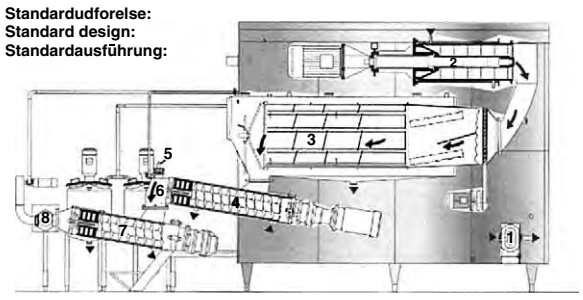

1. Flodepumpe	1. Cream pump	1. Rahmpumpe
2. Kæmeafdeling	2. Chuming section	2. Butterungsabteilung
3. Separeringsafdeling	3. Separating section	3. Trennabteilung
4. Æteafdeing I	4. Working section I	4. Knetabteilung I
5. Reguleringsspjæld	5. Regulating gate	5. Reglerplatte
6. Vakuumkammer	6. Vacuum chamber	6. Vakuumkammer
7. Æteafdeing II	7. Working section II	7. Knetabteilung II
8. Smorpumpe	8. Butter pump	8. Butterpumpe

Figure 5 Schematic outline of a continuous buttermaking machine. From APV Unit Systems, Denmark, with permission.

Working sections Augers transport the butter along the working sections and through aperture plates. The process kneads or works the butter, expelling more buttermilk and influencing the final body and texture of the finished product. The moisture droplets must be fine and evenly dispersed.

During this process of working, salt (if required) is added in the form of a 50% saturated slurry. Water may be added to adjust the final moisture content, and in the case of lactic butters, the mix of flavor distillates or concentrated bacterial cultures is added at this stage. The second part of the working section operates at a much higher speed to ensure that the culture or salt is correctly distributed.

The working sections are cooled with chilled water. The link between the first and second working sections operates under vacuum. This provides controlled deaeration of the butter, thus giving the end product a very close texture and improving the shelf-life. The body and texture of butter worked under vacuum are quite different from the open structure of traditionally worked butter.

To allow for the stoppages that must occur during normal production, a 'balance tank' is maintained between the buttermaker and packing equipment to buffer flow variation (Figure 6). This butter 'trolley,' as it is known, is constructed of stainless steel and performs the essential task of maintaining the flow of

Figure 6 Butter buffer tank. From APV Unit Systems, Denmark, with permission.

butter both from the buttermaker and to the packing equipment.

Packing

Wholesale The butter is packed in bulk or retail packs directly from the churn or from the trolley of a continuous buttermaker. Bulk butter is normally packed in 25-kg quantities in cardboard cartons. These can be lined with parchment paper, but colored polythene, enabling the inner wrapper to be readily seen, is the preferred lining material.

The simplest form of bulk butter packer is the 'Vane'-type packer – this is simply a hopper into which the butter is fed, either manually or pumped from the buttermaker trolley. The butter is extruded by a screw-conveyor through a suitably sized nozzle into the lined carton. When the carton is full, the flow of butter is stopped, and the butter is 'cut' with a heated wire. The full carton is then removed and checked for weight, which is adjusted manually. The liner is closed, ensuring there is no exposed product, and the carton is sealed, coded, and palletized. The normal quantity is 50 cases of 25-kg butter per pallet.

For large-scale production, automatic packers are normally an integral part of continuous buttermaking installations (Figure 7). The cardboard cartons are loaded flat, and the inner wrap is polythene from a reel. The machine forms the carton, lines it, and presents it for filling. Once full, the weight is checked,

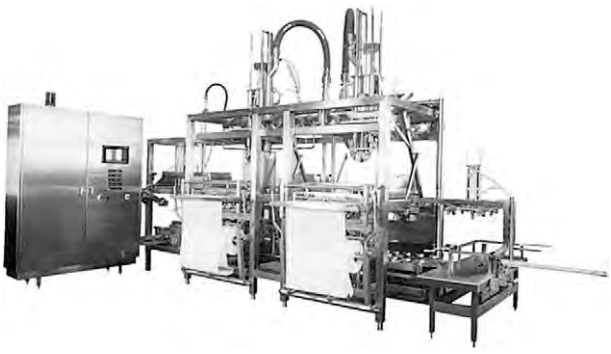

Figure 7 Butter packing line. From APV Unit Systems, Denmark, with permission.

and if necessary, additions are made via an 'injection unit,' which adds the required quantity of butter below the original surface.

On achieving the correct weight (25 kg), the liner is folded over the top surface, and the carton is closed, sealed, and coded. This minimal exposure to possible contamination has contributed greatly to the extension of the keeping quality of butter. Automatic palletization is also a feature of some installations.

Retail Butter is required to be sold in metric quantities. Most retail packs are in 250- or 500-g weights. The shape of the packs vary from a brick of differing dimensions to rolls, and packs may be foil- or parchment-wrapped. Attractively designed plastic tubs holding 250 or 500 g are also available.

For retail packing, the butter is formed in an appropriately shaped chamber on a rotating drum. This formed portion (roll or brick) is pushed out of the chamber into the waiting coded wrapper (parchment or foil), which is then folded. The portions are checkweighed before passing onwards to an automatic case packer and palletizer. The butter at this stage is still very soft in consistency, and any mishandling is likely to deform the portion.

Catering butter portions are packed in foil or plastic tubs with foil lids. The machines developed to pack this size of butter portion must operate at maximum efficiency. The quality of the butter, in terms of body and texture, is of secondary importance. The butter is extruded into the partly folded foil held within a shaped chamber, and the folding is then completed. Alternatively, the butter is discharged into a plastic tub before sealing.

Larger catering packs, e.g., 2-kg plastic tubs, are filled by extrusion before the lid is applied.

Reworked butter For marketing or distribution reasons, it may be necessary to rework bulk butter.

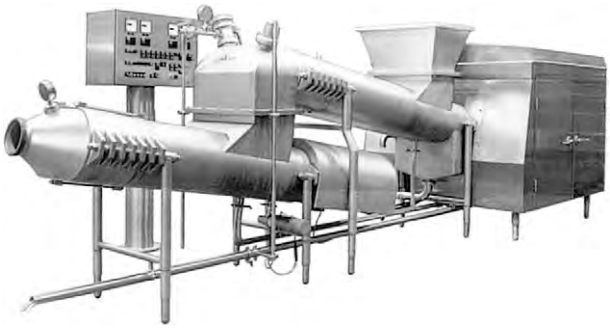

Figure 8 (see color plate 13) Butter reworking system. From APV Unit Systems, Denmark, with permission.

This butter, in 25-kg blocks that have been stored at either −18 or −25 °C, can be either attemperated, i.e., brought to a temperature suitable for packing, before working, or handled frozen.

Attemperation traditionally involves placing bulk butter in a store at 5–8 °C for a period to attain that temperature. The use of microwave tunnel heaters has reduced the space and time necessary, and is a more efficient way of attemperation. Blocks of butter may be brought up to blending temperature within hours, then blended and standardized on a batch or continuous basis to the required salt and moisture contents for packing.

An alternative process employs a butter reworking system that first chops the frozen butter blocks into strips. The spaghetti-like strips of butter leave the chopping unit at a temperature of 0–2 °C. They are then transferred into the blender where they are worked several times and vacuum-treated before repackaging in retail units (Figure 8).

Product Evaluation

Notwithstanding the aesthetic requirements of packaging and any legal requirements, the most important quality parameters of butter are its taste and keeping quality.

The modern continuous machines and their associated equipment are normally cleaned in place (CIP). Close collaboration between the design engineers, operators, and quality controllers is necessary to ensure effective cleaning. (*See* **Plant Design:** Basic Principles; Designing for Hygienic Operation; Process Control and Automation.)

Attention to detail, monitoring the critical aspects of the process, whether it be batch or continuous, is essential. It is, however, normal to sample the finished product to confirm that the desired microbiological, chemical, and organoleptic qualities have been achieved.

Microbiological

Microbiological guidelines for the acceptance of bulk butter into intervention storage are described by the Intervention Board for Agricultural Produce. For butter to be acceptable, it must have:

Total viable count: < 1000 organisms per gram (maximum 5000);
Coliforms: absent in 1 g;
Yeasts and molds: less than $10\,g^{-1}$.

Chemical

Butter composition and protection of use of the designated term 'butter' are described by the EC Council Regulations No. 1898/87 and No. 2991/94, respectively. Butter must contain not less than 80% milk fat but not exceeding 90%; a maximum water content of 16% and a maximum dry nonfat milk material (milk solids not fat) of 2% are permitted. Substances necessary for the manufacture of butter may be added, but such substances may not be used to replace in whole or part any milk constituent. Regulations for butter composition in other major butter-producing countries, New Zealand (Regulation 111. Butter, Food Regulations 1984) and USA (US Code of Federal Regulations Title 7, Volume 3, 7CFR58.345 and Title 21, Section 321a), are similar to the EC regulations. However, the New Zealand regulations also permit the fat phase used to prepare butter to be standardized for consistency by the removal or addition of physically fractionated fractions, provided that the fatty acid profile of the butter falls within the typical range. Such a product must then be identified as 'standardized butter'. The Australian New Zealand Food Standards Code may be found at www.anzfa.gov.au/foodstandardscode/.

Constant monitoring of the moisture and salt levels – where appropriate – is carried out during processing. On-line devices are available that link the machine controls to the moisture and salt determinations. These are necessary to match the manufacturing speeds now possible where traditional chemical analysis is too slow. However, these automated techniques still require calibration against standard laboratory methods.

Organoleptic Grading

The grading of butter involves much more than the tasting of the product. Grading normally takes place no less than 48 h after manufacture – a period necessary to allow the butter to 'cool' and 'settle.' Scoring standards and identification of defects are described for butter for intervention storage in the Official Journal of the European Community (1995) No.

L46/27. (*See* **Sensory Evaluation**: Sensory Characteristics of Human Foods; Practical Considerations.)

The butter must first be stored or 'tempered' at 10 °C for 24 h prior to examination. A sample (plug) of bulk or tub butter is taken using a borer. For other retail portions, the brick or roll is cut half way down, and the lower half is broken apart.

The qualities measured are:

1. Flavor and aroma. These are judged by the smell and taste. A portion of the butter should be tasted but not swallowed. (*See* **Sensory Evaluation**: Aroma; Taste.)
2. Consistency. A very good butter should have a close body and waxy texture and should be well spreadable. The appearance of the butter will provide the experienced butter grader with much information. A portion of the plug is broken off to enable the texture to be examined, and a section is cut to observe the cut surface. Scoring reflects the desirability of the compact 'plastic' nature of butter produced by modern technology. The more open and grainier consistency of traditional butter may be viewed as a defect or alternatively considered in the 'specialties' category. (*See* **Sensory Evaluation**: Texture.)
3. Appearance and finish. Evenness of color and absence of free moisture, giving a 'clean, dry' butter, are the requirements for good-quality butter. Traditionally made butter had a very open texture, and free moisture was a frequent fault. With the advent of the vacuum section in continuous buttermakers, this particular characteristic or fault is now rarely seen. (*See* **Sensory Evaluation**: Appearance.)

Each of these attributes is given a score, dependent on the perceived relative importance of the characteristic. Scoring for acceptance for intervention storage is shown in Table 1. The United States Department of Agriculture (USDA) defines detailed criteria for butter grading based on flavor, body, color, and salt characteristics (USDA Standards for Butter, 58.2621–27, 1989).

Defects

Defects that arise in the finished product can normally be attributed to problems arising from two main areas: (1) the quality of the original milk or cream and its handling; (2) manufacturing defects; or a combination of both.

Taints and poor microbiological quality will give rise to off-flavors. A reduced shelf-life and physical defects can be caused by poor hygiene, temperature abuse, the use of inappropriate pumps, and over-agitation.

Utilizing cream received from other sources can give rise to numerous problems unless the receiving plant can be absolutely sure of the conditions under which it was produced and handled.

Operating faults, such as an imbalance between the speed of the first section of the buttermaker and too slow a cream flow, will cause the grains to be too large. Consequently, the buttermilk will not drain away cleanly, and the result will be a streaky, weak-bodied butter with free moisture.

Underchurning, with too high a cream flow and too slow a churn speed, results in small grains and incomplete separation of the fat and aqueous phases. This gives a butter with a very high moisture content and a pale color.

Overworking due to, for example, too much product in the working section, excessive conveyor speed, or apertures being too restricted, will give a weak-bodied, lifeless, and sticky butter, difficult to handle and likely to lose points at grading.

An open-textured butter with uneven salt and moisture distribution may be caused by too slow an auger speed, inadequate vacuum or too little product in the working section, or insufficient restriction at the adjustable apertures.

A mottled butter with excess moisture or excess salt can be the result of an inaccurate salt:water ratio in the slurry, or an improperly mixed slurry.

Packed butters held under chilled conditions have a finite life. Defects can develop from exposure to light, and taints are easily absorbed if butter is stored near strong flavors or smells. Foil-wrapped or sealed retail packs stored under the recommended chilled conditions should retain their premium quality for 2 months, whereas parchment-wrapped butters will develop surface faults after some 4–6 weeks.

Table 1 Scoring for acceptance for intervention storage

	Maximum	*Required*
Appearance	5	4
Consistency	5	4
Flavor	5	4

See also: **Butter**: Properties and Analysis; **Colloids and Emulsions**; **Cream**: Types of Cream; **Lactic Acid Bacteria**; **Milk**: Processing of Liquid Milk; **Plant Design**: Basic Principles; Designing for Hygienic Operation; Process Control and Automation; **Sensory Evaluation**: Sensory Characteristics of Human Foods; Texture; Aroma; Taste; **Starter Cultures**; **Whey and Whey Powders**: Production and Uses

Further Reading

Murphy MF (1990) In: Robinson RK (ed.) *Dairy Microbiology*, 2nd edn, vol. II. London: Elsevier.

Rajah KK and Burgess KJ (eds) (1991) *Milk Fat, Production, Technology and Utilization*. Huntingdon, UK: Society of Dairy Technology.

Romney AJD (ed.) (1990) *CIP: Cleaning in Place*, 2nd edn. Huntingdon, UK: Society of Dairy Technology.

Rothwell J (ed.) (1988) *Cream Processing Manual*, 2nd edn. Huntingdon, UK: Society of Dairy Technology.

Wilbey RA (1991) In: Russell JB and Pritchard L (eds) *Analysis of Oilseeds Fats and Fatty Foods*. London: Elsevier.

Wilbey RA (1994) In: Robinson RK (ed.) *Advances in Milk Processing. Modern Dairy Technology*, 2nd edn. vol. 1. London: Chapman & Hall.

Properties and Analysis

A M Fearon, The Queen's University of Belfast, Belfast, UK

Introduction

In the manufacture of traditional creamery butter, cream is churned to bring about a phase inversion and then worked to give a product having a continuous fat phase and a disperse phase containing water and globular fat. The properties of butter are determined both by its chemical composition and by the physical treatments applied throughout its manufacture. This article reviews the physical properties and chemical composition of butter, what factors affect these properties, and how they may be determined.

Physical Changes During Churning

The first stage of the physical transformation of cream into butter by churning occurs during the aging or tempering of the cream prior to churning. During this period, milk fat crystallizes and the fat globule membrane is weakened. The churning process is based on the incorporation of air into cream to create a foam and is achieved by vigorous agitation. Fat globule membranes, already weakened by cold aging, are further damaged by the mechanical stresses applied, and liquid fat escapes. This brings about collapse of the foam and clumping of the fat globules to produce butter granules which are then worked together to form butter. Some globular fat is present in the final product, but the amount will depend on the prechurning tempering treatment and the severity of the working conditions before and after processing.

Butter Microstructure

The microstructure of traditional creamery butter and the influence of processing factors thereon have been studied using freeze-fracture techniques in association with electron microscopy. These have shown how cream tempering conditions can influence the crystallization pattern of milk fat within the globules and hence their process stability and, ultimately, product firmness. The cold–warm–cold cream tempering procedure, which was developed to improve the spreadability of winter butter, results in globules with a thick surface shell of solid fat and crystal aggregates of varying shape and size in the liquid fat in the interior. This type of cream globule can withstand mechanical stress during churning and consequently gives a softer butter with a higher proportion of globular fat than that obtained from a low-temperature cream treatment. Electron microscopy studies have also shown how intensive mechanical working of butter during processing destroys fat globules, resulting in a highly homogeneous structure with a more crystalline interglobular phase and consequently a firmer texture. A recent development in butter manufacture has been the application of scraped surface technology, developed by the margarine industry, to produce a more spreadable product.

Chemical Changes During Churning

Butter is basically a concentration of milk fat along with some water and milk solids-not-fat (MSNF). The composition of the fat in butter reflects that of the original milk fat, although there is some loss of phospholipids, sterols, and free fatty acids, particularly volatile fatty acids, into the buttermilk during separation and churning. Greater change occurs in the physical state of milk fat during churning than in the chemical nature of its constituents. However, the combination of agitation of the milk during milking with extended holding of the milk on the farm and at the factory before pasteurization lead also to an increase in concentration of free fatty acids and consequently lipolyzed flavor in the product. The increase in lipolysis is probably due to increased accessibility of the fat to lipolytic enzymes because of damage to or loss of the protective milk fat globule membrane. Lipolyzed flavor may increase in stored butter due to the preferential release of short-chain fatty acids by enzymes with a high specificity

for these acids (secreted by psychrotrophic bacteria). In sweet cream butter (as opposed to lactic or ripened butter), in particular unsalted butter, the primary cause of flavor impairment is lipolytic rancidity. Thus, good manufacturing practice, at both the farm and the factory, is necessary to prevent a high level of free fatty acids in the final product.

Chemical and Physical Properties of Butter

Chemical Composition and Analysis

The composition of butter is described in the food regulations of most countries. Typically, butter contains 80–84% milk fat, 15.3–15.9% water, about 1% milk solids other than fat (casein, lactose, and minerals) and 0.03–1.8% salt. Milk fat is the only fat permitted in butter, although in some countries lower-melting milk fat fractions may be added during manufacture to produce a softer product which complies with the definition of 'butter.' A maximum water content (16%) is usually stated as it is important for the good keeping properties of the product. Butter may be salted or unsalted, but may not contain any added antioxidants. In some countries, natural coloring agents, such as annatto, turmeric, carotene, or curcumin are permitted. Neutralizing salts and lactic acid cultures are allowed for the manufacture of ripened or lactic butters.

Analytical methods for butter are included in the inventory prepared by the International Dairy Federation (IDF), the Association of Official Analytical Chemists (AOAC) and the International Standards Organization (ISO) on methods of analysis for milk and milk products. Apart from standard methods of analysis of butter components (moisture, fat, solids-not-fat, and salt), other analytical methods are cataloged, including methods to characterize milk fat, to determine the presence of a foreign fat, and to detect rancidity in butter.

Not surprisingly, many of the analyses for butter are concerned with the composition and properties of its fat. Some classical chemical constants for milk fat are summarized in Table 1.

Milk fat contains a comparatively high proportion of water-soluble volatile fatty acids, in particular butyric acid. The presence of these is the basis of both traditional (Reichert–Meissl, Kirschner) and modern (chromatography) tests to detect the presence of adulterating fat. Iodine value (a measure of unsaturation) is a useful identifier of summer or winter butter but it has commercial value too. Butter-makers in mainland Europe and Ireland apply the appropriate cream-tempering treatment, the Alnarp process, to produce harder or softer butter on the basis of cream iodine value.

Table 1 Chemical constants for milk fat

Chemical constant	*Value*
Saponification value (SV) SV = mg KOH needed to saponify 1 g fat	220–240
Iodine value (IV) IV = g iodine reacted with 100 g fat	26–42
Reichert – Meissl number (RMN) RMN = cm^3 of 0.1 mol l^{-1} alkali needed to neutralize the water-soluble volatile fatty acids distilled from 5 g saponified fat	20–35
Polenske number (PN) PN = cm^3 of 0.1 mol l^{-1} alkali needed to neutralize the water-insoluble volatile fatty acids distilled from 5 g saponified fat	1.0–3.3
Kirschner number (KN) KN = cm^3 of 0.1 mol l^{-1} alkali needed to neutralize the water-soluble volatile fatty acids distilled from 5 g saponified fat and which form soluble silver salts	18–30

Data from Walstra P, Jenness R (1984) *Dairy Chemistry and Physics*, pp. 377–457. Chichester: Wiley.

Table 2 Major fatty acids found in milk fat triacylglycerols

Fatty acid	*g per 100 g fatty acid*	*Class*
4:0	3.3	Saturated short-chain
6:0	1.6	
	4.9	
8:0	1.3	Saturated medium-chain
10:0	3.0	
12:0	3.1	
	7.4	
14:0	9.5	Saturated long-chain
16:0	26.3	
18:0	14.6	
	50.4	
16:1	2.3	Monounsaturated
18:1	29.8	
	32.1	
18:2	2.4	Polyunsaturated
18:3	0.8	
	3.2	

Data from Gurr MI (1989) In: Cambie RC (ed.) *Fats for the Future*, pp. 41–61. Chichester: Ellis Horwood.

Although more than 450 fatty acids have been identified in milk fat, only about 12 of these (Table 2) play a significant role in determining its physical, chemical, and nutritional properties. For detailed information on the composition of milk fat it is necessary to use chromatographic techniques. The fatty acid profile of milk fat is obtained by first preparing the more volatile fatty acid methyl esters and then separating these by gas–liquid chromatography (GLC) using a capillary column and a flame ionization detector. A triacylglycerol fingerprint of the fat may also be obtained using capillary column GLC, with resolution on the basis of carbon number and degree of unsaturation. Rapidly developing high-performance liquid chromatography (HPLC) techniques have found application for the resolution

of milk triacylglycerols whilst a combination of techniques, e.g., silver ion chromatography and reversed-phase HPLC, and new methods combining gas chromatography–mass spectrometry (GC-MS) can provide more detailed information on the molecular structure of milk fat.

Factors influencing chemical composition The fatty acids in milk fat originate from two main sources: those acids synthesized *de novo* in the mammary gland, C_4–C_{14} and a proportion of C_{16} acids, and those arising directly from the diet and taken up by the mammary gland from the circulating blood, i.e., the remainder of C_{16} and the longer-chain C_{18} acids. The principal factors which influence the relative contributions of fatty acids from these two sources are stage of lactation (shorter-chain acids) and composition of diet (longer-chain acids). Summer grazing of fresh pasture leads to a softer milk fat with a decrease in $C_{16:0}$ and an increase in $C_{18:0}$ and $C_{18:1}$ acids, whilst the reverse occurs during winter feeding of concentrates and silage.

Supplementation of the cows' diet with fats or oils to increase energy input, especially in early lactation, can also affect fatty acid composition. Generally, such a diet will tend to increase the yields of fatty acids $C_{18:0}$ and $C_{18:1}$ while decreasing the levels of the shorter- and medium-chain acids, C_6–C_{16}. If the supplemented fat is presented in a protected form, it can pass through the rumen without being subjected to ruminal hydrolysis or biohydrogenation and the composition of the resulting milk fat will reflect that of the supplement. Protected fats also allow rumen metabolic activity to proceed unimpaired by the adverse effect of large amounts of fat. The technology to produce protected fats has developed considerably since the early 1970s and is currently in use in Australia to produce butter with higher proportions of $C_{18:1}$, $C_{18:2}$ and $C_{18:3}$ fatty acids and lower proportions of the saturated C_{14} and C_{16} fatty acids. This butter has greatly improved spreadability whilst changes in the proportions of saturated-to-unsaturated fatty acids and the presence of *n*-3 fatty acids enhance the contribution of milk fat to human nutrition. Previous problems of rapid oxidation of milk fat arising from the high content of polyunsaturated fatty acids have been avoided by including additional vitamin E in the cows' diet. This additional natural antioxidant is transferred into the milk.

An alternative approach to feeding protected fats has been based on exploiting the $C_{18:0}$ to $C_{18:1}$ conversion which occurs during milk fat biosynthesis by the action of desaturase enzymes in the cows' intestine and mammary gland. The cows' diet must supply a high proportion of C_{18} fatty acids (e.g., from oilseed cereals) in order to optimize the desaturating activity of the bovine tissue. The resulting milk fat has an increased level of $C_{18:1}$ (oleic) acid and a reduced level of $C_{16:0}$ (palmitic) acid. Monounsaturated acids are less susceptible to oxidative reactions than polyunsaturated acids and the increased content of oleic acid in butter greatly enhances its spreadability at low temperatures.

Physical Properties

Physical constants Values for physical constants of milk fat are presented in Table 3. The refractive index for milk fat at 40 °C was once a valuable indication of its purity, but nowadays many fats used in the margarine industry will give a similar figure. The specific gravity of milk fat may be measured at different temperatures, although the difference in specific gravity between milk fat and other fats is greatest at around 40 °C.

Milk fat melts and solidifies over a temperature range so, rather than having a melting or solidification point, it has a melting and solidification interval (Table 3). Ideally, these ranges should coincide, but the dependence of the solidification interval on the rate of cooling and the influence of previous thermal history on the melting interval, as well as the dissolution rather than melting of fat crystals during heating, means that they rarely do so. Further information on the melting behavior of milk fat may be obtained using differential scanning calorimetry. This analysis is based on the thermal transitions that occur in a substance during heating and cooling.

The solid fat content in milk fat over a range of temperatures may be measured by nuclear magnetic resonance (NMR). This technique operates on the principle that protons placed in a strong magnetic field can, under certain circumstances, absorb energy from electromagnetic waves. This absorption, termed NMR, depends on the physical state of the protons and allows determination of the solid fat content (% SFC). The method was previously applied to butter only for research purposes but is now a regular quality control test for the dairy spreads, the physical

Table 3 Physical constants of milk fat

Physical constant	*Value*
Refractive index (40 °C)	1.4524–1.4561
Specific gravity (37.8 °C)	0.910–0.913
Melting interval	28–33 °C
Solidification interval	24–19 °C

Data from Kirk RS and Sawyer R (eds) (1991) *Pearson's Chemical Analysis of Foods*, 9th edn, pp. 574–608. Essex: Longman.

properties of which at refrigerator temperatures are an important selling point.

Rheology Butter may be described as a plastic fat and depicted for rheological purposes by the Bingham model which describes ideal plastic behavior. In rheological terms, a plastic material flows when a stress greater than a limiting value (yield value) acts on it. However, aspects of butter behavior, such as work softening and thixotropic hardening, can only be explained by viscoelastic theories. Various instrumental techniques to measure butter firmness were evaluated by the IDF, which recommended the cone penetrometer as easy, quick, and cheap to use and with acceptable reproducibility. The IDF method relates penetration depth (p) to apparent yield stress (AYS) by the equation $AYS = gW/\pi p^2 \tan^2(\frac{1}{2}\alpha)$, where W is the cone mass, g the acceleration due to gravity and α the cone angle. For more detailed information on the textual characteristics of butter, a two-bite compression test using the Instron Universal Testing Equipment can be carried out. This provides a texture profile from which such properties as fracturability, hardness, cohesiveness, and springiness can be measured. German butter manufacturers favor the sectility test which measures the force at a set rate needed to cut through a portion of tempered butter with a wire of certain diameter and length.

Factors that influence butter consistency A number of factors, not all within the control of the manufacturer, influence butter consistency. The proportion of solid fat in butter is highly correlated with product firmness and is strongly influenced by the cows' diet. Feeding is responsible for the difference in firmness between summer and winter butters and can be directly related to the change in fatty acid composition of the milk fat, which is due largely to the changeover from summer grazing to winter silage and concentrates.

The number and size of fat crystals also affect consistency and are usually determined by crystallization temperature and rate during aging of the cream. Slow or stepwise cooling promotes the formation of fewer, larger crystals and a lower SFC which favor a softer fat. Cream aging is not required when butter is produced by scraped surface technology. The manufacturing process creates very many small (β') crystals in a firm, homogeneous product. Firmness is greatly reduced by mechanical working and, as fat crystallization is completed during production, postmanufacture hardening, common in traditional creamery butter due to crystal growth, is avoided.

Plastic fats like butter have a three-dimensional crystal network held together by reversible weak van der Waals attractive bonds and irreversible stronger bonds formed where crystals have grown together postmanufacture. During mechanical working of butter, such as the microfixing of bulk butter prior to retail packaging, the hardness of butter is reduced (work softening); and although the butter becomes harder again over several weeks it will not reach its original value. Work softening may be explained by the breaking of the bonds in the crystal network, while the reformation of reversible bonds into a new network structure is responsible for the gradual increase in hardness. This property of work softening has been exploited by butter manufacturers to produce an easier-spreading butter. Recovery of hardness during storage is slower at low temperatures but is accelerated by temperature fluctuations, hence such a product requires careful control during marketing.

Nutritional properties of butter Butter, along with other fat-containing dairy products, contributes just under 25% of dietary lipid in the UK and concern has been focused on the relationship of dietary lipid to coronary heart disease (CHD) and atherosclerosis. Two inversely related cholesterol-containing blood lipid fractions, low-density lipoprotein (LDL) and high-density lipoprotein (HDL), have an important influence on these diseases. High concentrations of plasma LDL generally indicate an increased risk of CHD with dietary saturated fats tending to increase LDL concentrations and polyunsaturated fats tending to reduce them. Milk fat, although containing about 30% oleic acid ($C_{18:1}$), is defined as a saturated fat, probably because of the low levels of polyunsaturated fatty acids present. Butter itself has been shown to cause elevated LDL levels and this has been linked not only to its content of saturated fatty acids but also to their esterified configuration. However, there is now evidence that *cis* monounsaturated fatty acids, far from occupying a neutral position in the diet and CHD, actually lower LDL concentrations whilst maintaining beneficial HDL levels. Consequently, the new 'monobutters' containing elevated levels of monounsaturated fatty acids (mostly oleic acid) will be attractive to consumers not only for improved spreadability but also for health reasons.

Interest has been shown recently in a branched-chain unsaturated fatty acid which is normally found in relatively small quantities in milk fat, conjugated linoleic acid (CLA). CLA, or its major isomer found in milk, *cis*-9, *trans*-11-octadecadienoic acid, has shown significant anticarcinogenic activity, antiatherogenic activity and other beneficial biological activity in studies with human cells and animals. CLA can arise directly in milk fat from biohydrogenation of dietary linoleic (18:2) acid in the cows' rumen.

Its major route into milk fat however, is via the desaturation of *trans* vaccenic acid, *trans*-11 18:1, (TVA), which is formed during the biohydrogenation of both linoleic and linolenic (18:3) acid. On reaching the cows' mammary gland, TVA is desaturated to CLA by the Δ^{-9} stearoyl Co-A desaturase enzyme. There are a number of research programs directed at increasing this and other functional fatty acids in milk fat by modifying the cows' diet.

Government recommendations for a healthy diet have included a reduction in total fat intake, particularly saturated fat. However, the balance of nutrients, fat types, and ratio of *n*-6:*n*-3 fatty acids must also be considered. Milk fat is an important source of fat-soluble vitamins, especially vitamin A, and provides small but useful amounts of vitamin D to diets of children, pregnant and lactating women, whose requirements are particularly high.

Dietary lipids, of the correct balance, are essential for our health but, in general, food will be eaten only if it is palatable. Milk fat contributes not only a pleasant flavor and aroma, but also an attractive texture and mouth feel.

Additives and contaminants Certain additives such as salt, natural coloring agents, lactic cultures, and neutralizing salts may be permitted in butter but these vary from country to country. However, butter may also contain very low concentrations of other components, called contaminants. Although contaminants are primarily synthetic chemicals, they also include microbial toxins, endogenous plant toxicants, heavy metals, and radionuclides. A comprehensive review of milk contaminants in Europe has been published by the IDF.

Chemicals such as the organocholorine-type pesticides and polychlorinated biphenyls have been scrutinized as they are lipophilic and tend to accumulate in the fat. However, levels of these chemicals detected in butter are below limits set in most countries and there is legislation in place governing their use. Organophosphates and carbamates are also widely used as pesticides but, as they are not lipophilic and are readily broken down by the cow, they give less cause for concern.

The risk of contamination of milk by such substances as detergents, disinfectants, and plasticizers (from packing materials and pipelines) may be minimized by good production practice both on the farm and at the factory. Antibiotics in milk can have a detrimental effect on the manufacture of cultured products and their use on farms is governed by legislation. Strict control of feed manufacture should avoid mold growth and the possibility of mycotoxin production and contamination of the milk.

There is little heavy metal contamination of butter through the feed, the cow acting as an effective filter. However, some contamination with copper and iron from dairy equipment can occur during and after milking; these metals act as prooxidants and the manufacturing treatments which are applied to butter make its component fat phase particularly susceptible.

There is little cause for concern over radionuclide levels in milk. Nevertheless, following the Chernobyl reactor incident in May 1986 and the associated increase in radioactive fallout over Europe, especially Scandinavia, monitoring of agricultural produce still-takes place. The isotopes concerned were iodine-131, cesium-134, and cesium-137. Although iodine-131 has a short half-life of 8 days, iodine accumulates in the thyroid gland and high concentrations can be reached. Milk was found to contain all three radioactive isotopes of concern but butter had very low levels, with the isotopes being distributed in the serum rather than the fat phase.

Dairy Spreads

Traditional table spread products such as butter must compete with the plethora of blended dairy spreads on the market. These blends have improved spreadability over butter, are more competitively priced, and appeal to the health lobby owing to their increased levels of unsaturated fatty acids and, in some instances, lower fat contents.

The simplest type of dairy spread may be obtained by blending cream or butter with a liquid vegetable oil such as soya bean oil. The mixture of cream and vegetable oil may be churned in a batch or continuous butter-maker, but if oil is added to butter itself, high shear rates are required to insure good mixing. Increasing the level of oil to improve spreadability at low temperatures results in oiling-off and loss of body at higher temperatures. This can be avoided by mimicking margarine manufacturers and including a proportion of saturated (hydrogenated) fat to maintain body and aid emulsion stability. Typically, such a product would contain a vegetable oil such as soya bean oil, a partially hydrogenated oil and cream, and may be manufactured in a continuous butter-maker or, more commonly, by using scraped-surface technology. The fat content of these two types of dairy spreads is usually in the region of 70–80%.

The third type of dairy spread available is the low-fat product. In low-fat spreads the aqueous phase makes up about 52–75% of the product compared with a maximum of 16% in butter. The fat phase is composed of vegetable oils, hydrogenated vegetable oils and possibly milk fat, with sodium caseinate or a

buttermilk protein concentrate added for both flavor and water-binding/emulsification purposes. The early low-fat products contained milk fat as the principal fat but both financial and consistency considerations have led to the development of products having no or very low levels of milk fat and depending on the milk protein for flavor. One particular problem arises with these products due to the use of milk protein. As the fat content is reduced, the water-in-oil emulsion becomes less stable and milk protein, when added to such products with fat levels around 40%, tends to promote a phase inversion to an oil-in-water emulsion. This problem may be overcome by increasing the milk protein level (and increasing the cost), modifying their properties by heat treatment and by carefully selecting the emulsifier and stabilizer levels needed to maintain a stable emulsion. This type of product, which most closely resembles margarine, is manufactured using margarine technology in a scraped-surface cooler to bring about crystallization, mainly in the β' form. During manufacture, it is critical for the keeping properties of the product that good moisture distribution with large numbers of discrete moisture droplets and an absence of channeling is achieved. All three types of dairy spreads, as with the new spreadable butters, require packaging in tubs, as the traditional foil laminate or parchment wrapping provides inadequate support. Because of increased levels of unsaturated fats and an increased aqueous phase (which results in a larger size of moisture droplet), these products should be stored at a low temperature to preserve chemical and microbiological quality.

See also: **Butter**: The Product and its Manufacture; **Contamination of Food**; **Dairy Products – Nutritional Contribution**; **Fatty Acids**: Dietary Importance; **Milk**: Physical and Chemical Properties

Further Reading

Baer RJ (1996) Production and utilization of dairy cows' milk and products with increased unsaturated fatty acids. In: Phillips CJC (ed.) *Progress in Dairy Science*, pp. 247–262. Oxon: CAB International.

Christie WW (1997) *Advances in Lipid Methodology – Four*. Dundee: Oily Press.

De Man JM (1999) *Principles of Food Chemistry*, pp. 311–353. Gaitersburg, Maryland: Aspen.

Gurr MI (1989) Dairy fats: nutritional nasties or dietary delights? In: Cambie RC (ed.) *Fats for the Future*, pp. 41–61. Chichester: Ellis Horwood.

IDF (1981) *Evaluation of the Firmness of Butter*. International Dairy Federation Bulletin Document no. 135. Brussels: International Dairy Federation.

IDF (1991) *Residues and Contaminants*. Special Issue 9101. Brussels: International Dairy Federation.

IDF (1996) *Inventory of IDF/ISO/AOAC Adopted Methods of Analysis and Sampling of Milk and Milk Products*, 5th edn. International Dairy Federation Bulletin no. 312. Brussels: International Dairy Federation.

Juriaanse AC and Heertje I (1988) Microstructure of shortenings, margarine and butter – a review. *Food Microstructure* 7: 181–188.

Keogh MK, Quigley T, Connolly JF and Phelan JA (1988) Anhydrous milk fat. 4. Low fat spreads. *Irish Journal of Food Science and Technology* 12: 53–75.

Kirk RS and Sawyer R (eds) (1991) *Pearson's Chemical Analysis of Foods*, 9th edn, pp. 574–608. Essex: Longman.

Pariza MW (1997) Conjugated linoleic acid, a newly recognised nutrient. *Chemistry and Industry* No. 12. June: 464–466. West Sussex: Society of Chemical Industry, John Wiley & Sons Ltd.

Walstra P and Jenness R (1984) *Dairy Chemistry and Physics*, pp. 377–457. Chichester: Wiley.

Wilbey RA (1994) Production of butter and dairy based spreads. In: Robinson RK (ed.) *Advances in Milk Processing. Modern Dairy Technology*, vol. 1, 2nd edn, pp. 107–158. London: Chapman & Hall.

C

Cabbage *See* **Vegetables of Temperate Climates**: Commercial and Dietary Importance; Cabbage and Related Vegetables; Leaf Vegetables; Oriental Brassicas; Carrot, Parsnip, and Beetroot; Swede, Turnip, and Radish; Miscellaneous Root Crops; Stem and Other Vegetables

CADMIUM

Contents

Properties and Determination

G Zurera Cosano and M A Amaro López, University of Córdoba, Córdoba, Spain

Background

Cadmium is a natural element in the earth's crust (atomic number = 48, atomic weight = 112.411, electronic configuration (K_r) $4d^{10}$ $5s^2$, and oxidation states 2). Cadmium metal is generally described as bluish or silver white in color, and as being relatively soft, ductile, and malleable, but this form is not common in the environment. It is usually found as a mineral combined with other elements such as oxygen (cadmium oxide), chlorine (cadmium chlorine), or sulfur (cadmium sulfate, cadmium sulfite). These compounds are all stable solids that do not evaporate, although cadmium oxide is often found as part of small particles present in air.

Most cadmium used is obtained as a byproduct from the smelting of zinc, lead, or copper ores, probably owing to its chemical similarity. Cadmium has a number of important industrial applications, but it is used mostly in metal plating (electroplating, alloys, and solders), pigments, plastics (stabilizers for polyvinyl chloride), batteries, photovoltaic cells, etc. Most uses of cadmium are 'dissip–ative,' the products using it are widely distributed in the environmental and hard to recycle.

Cadmium is fairly widely distributed and is found in shales and igneous rocks, coals, sandstone, limestone, lake and oceanic sediments, and soil. Cadmium has been listed as a pollutant of concern due to its persistence in the environment, potential to bioaccumulate, and toxicity to humans and the environment. Though cadmium has come into widespread use only fairly recently, it is quite likely that for many centuries this highly toxic metal has caused incidents of food poisoning. After mercury and lead, cadmium can probably be considered next in importance as an environmental pollutant.

Sources of Cadmium Exposure

Sources of exposure are many and include:

- Breathing contaminated work-place air (battery manufacturing, metal soldering or welding, smelting and plating operations, lithography, and engraving).
- Exposition of dust and fumes from the mining and refining of metals and contaminated air from the burning of fossil fuels (such as coal or oil) or the incineration of municipal waste materials.
- Eating foods containing cadmium; there are low levels in all foods (highest in shellfish, liver, and kidney), and drinking contaminated water.

- Inhalation of cadmium in cigarette smoke (most people who smoke have about twice as much cadmium in their bodies as do nonsmokers).
- Application of phosphate fertilizers or sewage sludge which increases cadmium levels in crops.

The primary source of cadmium exposure is food, since food materials tend to take up and retain cadmium (for example, plants from the soil, fish from water).

Occurrence in Foods

In order to understand the biological effects of cadmium in foods, it is necessary to have some idea of the distribution and occurrence of cadmium in air, water, and soil. Cadmium enters the air from mining, industry, and burning coal and household wastes and cadmium particles in air can travel long distances before falling to the ground or water. It enters water and soil from waste disposal and spills or leaks at hazardous waste sites; it binds strongly to soil particles, and some cadmium dissolves in water.

Furthermore, it is essential to try to determine how cadmium normally moves from one sector to another and to explore the way in which human activities may influence these processes. The importance of atmospheric cadmium in relation to its potential effects on humans lies not so much in its direct inhalation as in its possible contribution to soil, water, and vegetation through either dry or wet deposition and, ultimately, to food. In this respect, the solubility, pH, and chemical nature of the deposited fraction may be essential factors for controlling concentrations in food.

Cadmium in Air

Cadmium is universally present in the atmosphere in amounts varying from $1\,ng\,m^{-3}$ or less in rural or uncontaminated areas to 10–$50\,ng\,m^{-3}$ in congested urban areas. There is no definite relation of cadmium levels to the size of cities; some small but highly industrialized cities have levels which are greater than those of some very large metropolitan areas. The principal sources of cadmium in the atmosphere are thought to have resulted from emissions from smelters, reprocessing of various cadmium-containing metals, burning of plastics, or power generation from coal.

In areas around cadmium-emitting factories, cadmium concentrations in air are several hundred times greater than those in noncontaminated areas. The compounds of cadmium which exist in the atmosphere are not known with any certainty but it can be assumed that they are nonvolatile and that they are present in the particulate phase. Some of the probable forms are the oxide, sulfide, sulfate, chlorides, and other more complex ions or compounds.

Cadmium in Water

The cadmium content of water is extremely low. The occurrence in drinking water has of course been of special interest since this directly affects human intake. Cadmium levels in water, in the absence of contamination, are seldom above $1\,\mu g\,l^{-1}$. Contamination may occur as a result of the use of galvanized pipes and cisterns. Cadmium-containing solders in water heaters and other fittings can be another cause. Pollution of water due to the use of cadmium-rich sewage sludge for agricultural purposes has been reported on a number of occasions. The mean concentration of cadmium in rainwater is $0.001\,\mu g\,ml^{-1}$ in clean areas and $0.0037\,\mu g\,ml^{-1}$ in contaminated areas. The US Environmental Protection Agency estimates that consumption of the reference dose (RfD) of $0.0005\,mg\,kg^{-1}\,day^{-1}$ for cadmium in drinking water would likely not result in the occurrence of chronic noncancer effects.

Cadmium in Soils

Agricultural soils receive inputs of cadmium from soil supplements, such as phosphate fertilizers and sewage sludge, and from atmospheric deposition. Although the significance of sewage sludge as a source of cadmium contamination of soil is well established, European soil inventory studies indicate that, on a national basis, phosphate fertilizer and atmospheric inputs are the greatest sources of cadmium addition to soils. Levels in soils vary widely from about $0.06\,mg\,kg^{-1}$ in virgin soil to 10–$20\,mg\,kg^{-1}$ in the vicinity of some smelters.

Cadmium in Plants

Some species of plants have an ability to concentrate cadmium at levels that are sometimes much greater than those in the immediate environment. The ratio of cadmium concentration in plants to that in the corresponding soil has been reported to be as high as 10:1. The average cadmium content of terrestrial plants is about $0.6\,mg\,kg^{-1}$, although there is obviously a very wide variation even in uncontaminated areas. The range of values in species growing in clean areas is of the order of 0.01–$10\,mg\,kg^{-1}$ or more, but most plants contain below $1\,mg\,kg^{-1}$. However, some root crops such as carrots and parsnip, and some leafy crops such as spinach and lettuce, tend to contain more cadmium than other plant foods. This is also true of cereals and certain mushrooms, and this indicates that plants tend to take up cadmium from the soil, unlike lead. Variation in species is very great even

in plants growing in the same area, and there is also variation in different parts of the same plant. There seems to be good evidence that plants generally show increased tissue cadmium levels in response to an increase in the environment from normal sources or from soil, water, or air contamination.

Plants primarily absorb cadmium from the soil through their roots. This absorption is influenced by, among other factors, particularly the pH of the soil and the organic matter content. The cadmium concentration in plants is inversely related to the amount of organic matter in the soil. A large amount of organic matter in the soil limits the availability of cadmium to plants. Organic soils form the strongest metal complexes and hence retain the metal more firmly. An increased accumulation of cadmium by plants fertilized with superphosphate which contains cadmium as an impurity has been reported. With aerial application, it has been shown that the leaves of plants are an important entry point for cadmium.

Contamination of Food During Processing

The processing of vegetables prior to packaging can modify the content of cadmium. Washing and blanching decrease cadmium to varying degrees. These decreases are probably caused by cadmium being leached during washing and blanching and by thermal processing where cadmium is extracted into the water of brine.

Cadmium contamination occurs as a result of its use in plate-manufacturing equipment. Cadmium is readily soluble in weak acid solutions, so its use in food-processing plants should be confined to parts which do not come directly into contact with food. Another source of cadmium contamination is the utilization of ceramic and enameled utensils. Some glazed pottery, especially of the craft and homemade kind, is capable of releasing toxic amounts of cadmium and other metals into food.

Enameled kitchenware can also be a source of cadmium contamination of food, especially those which have bright colors, indicating the possible use of cadmium pigments. Another domestic source of cadmium contamination is the decoration and printing applied to glass tumblers and other containers. Furthermore, the spices, condiments, coloring, and preservatives which are added to foodstuffs may contain cadmium and be a vehicle of contamination. (*See* **Food Poisoning**: Classification.)

Cadmium Level in Foods

As cadmium is ubiquitous in the environment, all food is exposed to and contains cadmium. Except where there has been pollution, cadmium is normally found at a fairly low concentration in foodstuffs that generally contain less than $50\,ng\,g^{-1}$.

The Codex Committee of Food Additives and Contaminants (Joint Food and Agriculture Organization (FAO)/World Health Organization (WHO) Food Standards Programme) indicates that the dietary intake of cadmium fluctuates between 10 and 50 μg $person^{-1}\,day^{-1}$ for noncontaminated areas. The major sources of cadmium dietary intake are cereals, green leafy vegetables, potatoes, liver, and kidney (80% of total cadmium intake). Most fish contains little cadmium, with the average being less than $200\,ng\,g^{-1}$. Shellfish from unpolluted waters rarely have an average cadmium concentration greater than $1000\,ng\,g^{-1}$. A wide range of concentrations has been reported by different workers (Table 1).

Analysis of Foods for Cadmium

For cadmium analysis, an important aspect is the avoidance of airborne particulate contamination. The low level of cadmium in foods and the ubiquitous presence of significant levels of this element in some reagents, containers, and air make it necessary for the analyst to evaluate and eliminate all potential sources of contamination at each stage of the analysis.

Sample Preparation

Prior to the destruction of the organic matter, the sample should be subjected to drying and grinding in order to remove water from the product without reducing its dry-matter content. Normally, the moisture is removed and the analytical result expressed on a dry-weight basis. However, for nutritional purposes, it is common practice to express the concentration on a fresh-weight basis (i.e., wet-weight basis).

The drying procedure recommended by the Association of Official Analytical Chemists (AOAC) is vacuum oven drying at 100 mmHg pressure and at a temperature of 70 °C. The primary objective of grinding a sample is to homogenize it and, at the same time, reduce it to a suitable physical consistency so that it can be easily measured or weighed. A number of grinding procedures can be employed, but in any case reducing the sample to a finer particle size is desirable if the subsample chosen for analysis is to be less than 1.0 g. With all grinding procedures, contamination and segregation are potential sources of error.

Organic Matter Destruction

The traditional methods to solubilize biological and food samples are wet ashing with suitable oxidizing agents and dry ashing. Wet ash digestions are characterized by short ashing times (normally 3–4 h) and are

Table 1 Cadmium levels (ng g^{-1}) in individual food categories

Food category	*Cd level*	*Food category*	*Cd level*
Milk and dairy products			
Butter	0.26–1.33[a]	Dairy products	1–3[f]
Cheese	0.42–0.84[a] 0.18–0.55[b] 3–40[c] 3.15–5.63[d]	Icecream, mixed	1.85–5.55[a]
Cheese, processed	0.35–57.8[a]	Pasteurized milk	0.69–0.89[g]
Milk, whole	0.13–0.80[a] 0.044[e] 1–2[f]	Yogurt, mixed	0.38–1.60[a] 0.56[b]
Meat and poultry			
Meat	2–36[f]	Luncheon meat, canned	1.13–5.09[a]
Beef steak, raw	0.56–4.73[a]	Pork sausage	20–36[i, j]
Beef meat	1–40[h]	Meat products	3–75[f]
Pork, raw	0.46–15.8[a]	Veal, raw	0.64–9.09[a]
Pork, cured	2.90–14.7[a]	Lamb, raw	0.53–1.73[a]
Ham, boiled and cold	37–51[i, j]	Poultry, raw	0.76–3.14[a]
Meat organs	85.1–167[a]	Eggs	0.42–5.02[a] 1–2[f]
Fish			
Fresh-water fish, raw	0.54–7.72[a]	Fish canned	8.36–13.8[a]
Marine fish, raw	1.21–1.87[a]	Fish and shellfish	20.4–300[l]
Fish	7–97[f]	Shellfish	8.10–38.6[a]
Crayfish	0.032[k]		
Soups			
Soups, meat, canned	1.71–5.13[a]	Tomato soups, canned	7.64–15.1[a]
Bakery goods and cereals			
Apple pie	5.21–10.2[a]	Oatmeal cereal	1.68–15.5[a]
Bread	3–27[f]	Pasta, plain, cooked	15.6–39.3[a]
Cake, chocolate	8.10–15.8[a]	Pizza	13.6–27.5[a]
Cake, white and yellow	2.97–8.45[a]	Rice cereal, cooked	5.23–12.7[a]
Cereal	3–15[f]	Waffles and pancakes	6.25–14.7[a]
Cookies, all	13.3–20[a]	White bread, all	17.2–35.9[a]
Corn cereal	1.64–5.18[a]	Wheat and bran cereals	45.3–70.7[a]
Crackers	17.5–26.8[a]	Wheat flour	23.6–33.4[a]
Danish and donuts	9.88–15.5[a]		
Vegetables			
Beans, raw	1.38–3.76[a]	Onions, raw	7.07–44.2[a]
Broccoli, raw and cooked	6.08–22.3[a]	Peas, raw	1.92–4.04[a]
Carrots, raw	17.8–38.9[a]	Peppers, green and red	11.5–22.8[a]
Cauliflower, raw	4.59–28.2[a]	Potatoes, raw	14.0–61.1[a] 5–61[f]
Celery	3.74–46.2[a]	Potato chips	63.6–129.5[a]
Corn, raw	1.86–23.6[a]	Tomatoes, raw	5.51–13.4[a]
Cucumber, raw	0.77–3.10[a]	Tomatoes, canned, ketchup, sauce	25.8–42.2[a]
Lettuce	5.63–71.1[a]	Vegetables	4–17[f]
Mushrooms, raw	5.05–18.2[a]		
Fruits and fruits juices			
Apples	0.17–0.46[a]	Fruits	1–2[f]
Bananas	0.16–0.66[a]	Melons	2.69–8.41[a]
Citrus fruit, raw	0.14–1.91[a]	Peaches, raw	1.38–4.88[a]
Citrus juice	0.02–0.33[a]	Pears, raw	1.57–6.05[a]
Cherries, raw and canned	0.70–27.4[a]	Plums, raw	0.11–10.9[a]
Grapes	0.22–1.05[a]	Strawberries	11.0–22.7[a]
Fats and oils			
Fats and oils	0.3–1[f]	Peanut butter and peanuts	31.1–62.2[a]
Margarine	0.35–2.38[a]		
Sugar and candies			
Sugar	0.20–0.48[a] 2–20[f]	Honey	0.12–2.61[a]
Syrup	0.13–0.25[a]	Candy, chocolate	8.60–19.5[a]
Jams	1.70–9.78[a]	Candy, other	0.18–3.38[a]
Beverages			
Alcoholic beverages	1[f]	Soft drinks	0.07–0.32[a]
Coffee	0.08–1.73[a]	Wines	0.20–0.88[a]
Tea	0.03–0.80[a]	Beer, cans	0.08–1.61[a]

[a]Dabeka RW and MacKenzie AD (1995) *J. Assoc. Off. Anal. Chem*. 78: 897–909.
[b]Larsern EH and Rasmussen L (1991) *Lebensm. Unters. Forsch*. 192: 136–141.
[c]Gabrielli Favretto L (1990) *Food Add. Cont*. 7: 425–432.
[d]Zurera G, Sánchez PJ, Amaro MA and Moreno R (1997) *Food Add. Cont*. 14: 475–481.
[e]Jeng SL, Lee SJ and Lin SY (1994) *J. Dairy Sci*., 77: 945–949.
[f]Urieta I, Jalón M and Eguileor I (1996) *Food Add. Cont*. 13: 29–52.
[g]Moreno R, Sánchez PJ and Amaro MA. y Zurera (1999) *Milchwissenschaft* 54: 210–212.
[h]Kramer HL, Steiner JW and Valley PJ (1983) *Environ. Cont. Toxicol*. 30: 595–599.
[i]Ybañez N, Montoro R, Catala R and Flores J (1982) *Rev. Agroqui. Tecno. Alimentos* 22: 419–425.
[j]Ybañez N, Montoro R and Bueso A (1983) *Rev. Agroqui. Tecno. Alimentos* 23: 510–520.
[k]Rincón F, Zurera G and Pozo R (1988) *Archv. Environ. Cont. Toxicol*. 17: 251–256.
[l]Harrison N (1993) In: Watson DH (ed.) *Safety of Chemicals in Food. Chemical Contaminants*. England: Ellis Horwood.

performed by adding acid or an acid mixture to the test portion and boiling until digestion is complete.

Numerous wet oxidation procedures have been proposed but they all fit basically into three categories of digestion mixtures: (1) HNO_3 and H_2SO_2, (2) HCl, H_2SO_2, and H_2O_2 and (3) mixtures containing $HClO_4$. Wet digestion is frequently criticized because elemental loss by volatilization can occur if the temperature of the digestion vessel is greater than 250 °C. A variety of procedures have been developed using different types of digestion apparatus to retain volatile elements. 'Hot-block' digesters have been used for improved temperature control and their reflux capabilities and 'PTFE (polytetrafluoroethylene) bomb' digesters have been used because digestion can take place in a sealed vessel, thereby retaining volatile elements.

The latest papers tend to suggest that the method of wet digestion in a microwave oven is perhaps the most interesting development in sample preparation

techniques. Current systems incorporate many features to improve digestion time, element recoveries, contamination, and safety. Improved recoveries can be achieved with systems operating under power-regulated control of pressure because these systems allow digestion to be completed without venting volatile elements. Sealed digestion vessels also reduce acid consumption and, in combination with Teflon construction, reduce contamination problems.

The advantages of microwave digestion with respect to the dry-ashing sample preparation method (at 550 °C) are rapid dissolution, complete digestion, sample integrity, minimal reagent use, lower reagent blank, simultaneous multiple sample digestion, possible automation, and contamination-free digested samples.

For the determination of cadmium in brine shrimp, four different digestion procedures were compared (dry ashing, wet ashing on a hot-block, high-pressure bomb digestion and microwave-oven digestion) and the conclusion was that microwave digestion (in polyethylene autosampler cups) was an excellent technique for the rapid digestion of submiligram amounts of biological materials.

Dry ash digestions use a long, slow ashing step, usually performed overnight in a muffle furnace, accomplished by high-temperature oxidation. The critical requirements are: (1) the nature of the ashing vessel; (2) position in the muffle furnace; (3) ashing temperature; and (4) time. Silica is probably one of the best materials for the ashing vessel, although glass beakers and well-glazed porcelain crucibles are also suitable. A critical initial requirement is to keep the temperature sufficiently low to prevent flaming or burning.

The ashing process is completed by the addition of a small quantity of inorganic acid to the residue and evaporation on a hot plate. The residue is redissolved in acid and the solution is brought to volume with distilled or deionized water. Element loss during the ashing process is greater for dry ash procedures than for wet ash procedures and 'ashing aids' are often employed to prevent loss of the more volatile elements. The addition of 1:1 dilute sulfuric acid to food materials prior to dry ashing eliminates losses of cadmium, although losses due to volatilization are possible as well as lack of recovery due to the formation of insoluble silicates.

Both techniques, wet oxidation and dry ashing, give reasonably comparable results and, in may instances, it is the analyst who makes the difference whichever technique is employed. However, wet digestion is recommended, using sulfuric acid and hydrogen peroxide: dry ashing can result in low recoveries since cadmium is volatile at temperatures over 500 °C, temperatures of below 450 °C being more appropriate. Furthermore, the digestion procedure selected needs to be tested for appropriate quality assurance and quality control guidelines that include the use of contamination control, digestion blanks, spiked test portions, replicate analyses, an appropriate reference material, and recovery calculations.

Analytical Techniques

Analysts nowadays have a number of techniques to choose from when making an elementary assay. The choice to be made depends on a number of factors, such as accuracy, precision, detection limit, sensitivity, the user's experience, performance with standard samples, and safety considerations. However, it may be the more practical considerations of instrument availability, cost, and sample form and quantity that become the governing factors rather than the choice being based on the above criteria. The analyst's skill, past experience, or lack of experience may dictate the selection of the analytical technique chosen. Personal preference may equally be a significant factor in this decision. (*See* **Analysis of Food; Heavy Metal Toxicology.**)

Spectrometric techniques are most commonly used and include flame atomic absorption spectrometry (FAAS), graphite furnace, or electrothermal atomization atomic absorption spectrometry (ETAAS) and inductively coupled plasma-atomic emission spectrometry (ICP-AES). (*See* **Spectroscopy**: Atomic Emission and Absorption.)

The general trend for solid sample introduction is slurry atomization. The slurries are prepared by grinding the sample with 3-mm diameter ZrO_2 beads in a laboratory shaker or mixing by ultrasonic agitation with dilute HNO_3. For determination of cadmium in protein foodstuffs and vegetables by ETAAS, the samples were stabilized with a thixotropic agent (Viscalex HV-30); to aid ashing O_2 was added during the ashing phase of the furnace program and matrix modifiers ($NH_4H_2PO_4$ and Pt) were included. The size of the food sample particles determined the stability of the slurry and also influenced the extent of matrix and chemical interferences. With respect to the direct insertion of solid samples into a graphite furnace, the use of a novel graphite tube described as a 'ring chamber tube' is reported; this obtains good accuracy for National Bureau Standards (NBS) and Standards Reference Materials (SRMs).

The main area of research in liquid sample introduction is based on flow injection systems to improve the analytical sensitivity. Preconcentration procedures are described, such as the use of microcolumn packed with a chelating resin. The compounds used most commonly as ligands are dithiocarbamates such as ammonium pyrrolidine dithiocarbamate (APDC),

diphenyl thiocarbazone (dithiozone), and others. They vary in their selectivity and are usually very dependent on the pH at which extraction is carried out. The organic solvent used most is methyl isobutyl ketone (MIBK).

The gaseous sample introduction is of a particular interest in the food industry, for technological and toxicological reasons, to determinate Sn, Se, As, and Hg in food samples by hydride generation AAS. For analysis of cadmium in food, the hydride generation is not used extensively.

Developments in methodology for atomic absorption spectrometry Atomic absorption spectrometry has been the most popular method for metal determination in general and the most widely used technique for analyses of trace elements in foods. The popularity of this method arises from its analytical specificity, good detection limits, excellent precision, and relatively low cost. The atomic absorption spectrophotometer consists of a light source (wavelength for Cd = 228.8 nm), an atomization source, and a dispersion/detection device. Three types of atomizers are commonly used: flame, graphite-furnace and chemical vaporization.

With regard to FAAS, the air–acetylene gas mixture is the most commonly used, producing a flame of about 2400 °C, but it is not totally interference-free. There are various ways to improve FAAS sensitivity for determination of cadmium in foods, such as the use of slotted quartz tubes, atom-trapping techniques (water-cooled and slotted-tube atom traps), where atoms are trapped (condensed) on to a cool silica tube which is situated in an air-C_2H_2 flame and flow injection systems with incorporation of minipreconcentration columns.

Graphite furnace or ETAAS has been used for the measurement of selected elements, including cadmium, when FAAS detection limits were insufficient. Perhaps the main development has been the introduction of graphite probe atomization to ETAAS.

The main problems with FAAS and ETAAS analysis are matrix and spectral interferences. Deuterium arc and Zeeman-effect background corrections have been applied to reduce spectral interference in cadmium analysis. While deuterium arc background corrections are generally inadequate, Zeeman-effect background correction appears to be satisfactory for routine analysis. Various 'matrix modifiers' (organic and inorganic reagents) have been used to reduce matrix effects, such as molybdenum, lanthanum, and ammonium dihydrogen phosphate. Solvent extraction of cadmium complexes is another approach taken to minimize matrix effects.

Developments in Multielement Analysis

Inductively coupled plasma (ICP) excitation has revolutionized the field of emission spectroscopy, providing the analyst with a powerful multielement technique with good sensitivity for cadmium at 214.440 or 226.502 nm. Food and food products can be prepared for ICP analysis by either wet oxidation or dry ashing and samples with a wide range of elemental concentrations can usually be handled without the need to dilute or concentrate, since the ICP has a fairly large dynamic range. The current qualities of ICP – high speed, low detection limits (less than 0.01 mg kg^{-1} for cadmium), and excellent instrument stability – make this technology attractive for a wide range of analytical applications.

ICP-AES is used preferentially due to the possibility of handling a wide range of samples with minimum interference and of quantifying many of the metals, reducing analysis time, and providing a more sensitive determination of refractory elements. The potential of inductively coupled plasma mass spectrometry (ICP-MS) is affected by polyatomic interferences for some mineral elements that can be overcome by different sample introduction methods (hydride generation, flow injection, slurry nebulization) or separating procedures with chelating resins.

X-Ray Fluorescence (XRF) and Neutron Activation Analysis (NAA)

X-ray fluorescence (XRF) and neutron activation analysis (NAA) are the two main methods for noninvasive *in vivo* determination of heavy-metal concentrations in humans. Various XRF techniques have been developed for the measurement of cadmium, mercury, and lead, primarily in occupationally exposed persons to evaluate whether these cadmium concentrations could reveal previous occupational cadmium exposure in humans. Today, the technique can also be used for measuring kidney cadmium levels in the general population.

Both techniques are being used in foods for the analysis of trace elements. In XRF, a powdered sample is bombarded with X-rays and the Kα line of the element is measured using one of several different detectors depending on the wavelength of the emitted radiation. This technique is somewhat limited in scope as to detected elements and their concentration range. By XRF spectrometry is able to determine Cd in food with a detection limit of 2.5 $\mu g\ g^{-1}$.

In NAA, the sample is bombarded with elementary particles to light atomic nuclei to generate radioactive nuclides. The properties and intensity of the resulting radioactivity permit the identification and quantitative determination of the elements present in the

bombarded sample. The main advantages of neutron activation are the excellent accuracy at microgram and picogram ranges, matrix effects are minimal, and the sample can be either liquid or solid in form. For determination of Cd, the characteristics pertaining to neutron activation are nuclear reaction with ^{114}Cd (n, γ) ^{45}Cd, the half-life of the product is 54 h, and sensitivity (g) 10^{-8}.

Electrochemical Methods

The electrochemical methods are polarographic, both direct and alternating current, and anodic or cathodic stripping voltametry and polarography. Both techniques rely on the fact that different metals require the application of different electrical potentials before they are deposited from solution on to the cathode. The polarographic methods are suitable for dilute solutions with sensitivities of 10^{-6} to 10^{-9} mol l^{-1}.

Anodic or cathodic stripping voltametry is even more sensitive than usual polarography. After the cadmium ions have been preconcentrated from the solution and amalgamated into a stationary hanging mercury drop electrode, the process is reversed by using a more negative potential than the reduction potential of cadmium. Cadmium is oxidized and stripped anodically, using a slowly increasing positive potential. The measured current recorded during the stripping step is a direct linear function of the bulk concentration of cadmium. The detection limit for cadmium determined by anodic stripping voltammetry was 1×10^{-9} mol l^{-1}.

AOAC International indicates a multielementary method including the determination of cadmium in foods (AOAC Official Method 986, 15, 1996) adopted as a Codex Alternative Approved Method (type III) for anodic stripping voltammetry of cadmium, lead, and zinc in all foods, and another anodic stripping voltammetry method for the determination of Cd and Pb in foods (AOAC Official Method 982, 23, 1996), which is not applicable to fats and oils.

Speciation Studies

The role of atomic spectrometric techniques in chemical speciation is well documented using separation techniques prior to cadmium detection. The type of fractionation technique and its coupling to atomic spectrometers can only produce meaningful results when particular attention is paid to the methods used for sample collection, pretreatment, and storage to avoid the changes that can occur. The principal applications in this field have been the use of ICP-MS or ICP-AES as a multielement detector for high-performance liquid chromatography (HPLC) for the analysis of cadmium bound to metallothionein. Other applications are ETAAS coupled with HPLC for the determination of inorganic species. (*See* **Chromatography**: High-performance Liquid Chromatography.)

All methods must be validated by the laboratory prior to routine sample analysis by quality assurance/quality recommended procedures to avoid variability in reporting elemental contents in foods and improve the quality and reliability of published analytical results and data quality indicators. (*See* **Quality Assurance and Quality Control**.)

See also: **Heavy Metal Toxicology**; **Quality Assurance and Quality Control**

Further Reading

Benton Jones J Jr (1984) Developments in the measurement of trace metal constituents in foods. In: Gilbert J (ed.) *Analysis of Food Contaminants*, pp. 157–202. Barking: Elsevier.

Codex Committee of Food Additives and Contaminants (1999) *Discussion Paper on Cadmium*. CX/FAC 99/21.

Food Research Institute (1993) *Food Safety*, pp. 319–322. New York: Marcel Dekker.

Harrinson N (1993) Metals. In: Watson DH (ed.) *Safety of Chemicals in Food. Chemical Contaminants*, pp. 113–115. London: Ellis Horwood.

Reilly C (ed.) (1991) *Metal Contamination of Food*, 2nd edn. London: Elsevier.

Shibamoto T and Bjeldanes LF (1993) *Introduction to Food Toxicology*, pp. 136–139. London: Academic Press.

Toxicology

M Nordberg, Karolinska Institutet, Stockholm, Sweden

Introduction

Cadmium toxicology has been extensively reviewed during the last decades. Focus on food sciences and nutrition has been directed to dietary food intakes of cadmium in the population. Toxicity of metals depends on the bioavailability, species, source, and route of exposure. Nutritional factors such as protein, calcium, zinc intake and iron status have an impact on the toxicity of cadmium. Acidification of lakes and soils influences the mobility of cadmium. In the population, different groups and subjects, e.g. with a low iron status (mainly women) have been identified as being more susceptible to cadmium exposure. It is thus important to identify vulnerable groups when setting recommendations for daily intake of cadmium.

Importance

Cadmium is considered an important world-wide pollutant. In the environment, humans are exposed to cadmium via food and drinking water. Cadmium bromide and iodide is used in photography and photoengraving, and cadmium sulfide (cadmium yellow) is used in high-quality paints, glazes, inks, and artists' pigments. Cadmium is present in the air as a result of incineration of household wastes and emission from industry, and mining. Cadmium particles can be transported in the air over long distances, and thus the ground and water can be contaminated far away from the emission source. Cadmium remains at levels depending on pH, bound to other compounds, in the soil and water. The body burden of cadmium is higher in smokers than in nonsmokers, because tobacco often contains cadmium. The mechanism for toxicity of cadmium involves metallothionein (MT), which also can be used as a biomarker in environmental exposure to cadmium and in biological monitoring.

Excessive exposure to cadmium gives rise to several human diseases, with itai-itai disease being the most severe. Cadmium has been shown to cause damage to the reproductive organs and to induce carcinogenesis in laboratory animals in organs such as the prostate and testes. Long-term human exposures have given rise to lung cancer in industrial workers. Cadmium was classified in 1993 as a human carcinogen (Group 1) by the International Agency for Research on Cancer (IARC), based on experience in experimental animals and in welding environments, where inhaled cadmium could cause lung cancer. Some uncertainty exists in this assessment, based on recent data. Because of the toxicological properties of cadmium, legislation implying considerable restrictions in its use has been passed in some countries. Industrial exposures as well as exposures in the general environment have to be kept below recommended exposure limits in order to avoid adverse health effects.

Discovered in 1917, cadmium is a soft, silver–white, tasteless, and odorless metal. It is a group IIB element in the periodic table with an atomic number of 48 and an atomic mass of 112.411. It appears together with zinc, but is softer, and is used, to some extent, in a similar way to zinc. Cadmium is found as an impurity of zinc carbonate, which, upon heating, changes color due to cadmium impurities. Several natural isotopes of cadmium exist. The melting and boiling temperatures are 320.9 and 765 °C, respectively.

Cadmium occurs in the earth's crust with an average distribution of 0.1 mg kg^{-1} and occurs naturally, associated with zinc, in the geosystem. It is found in sulfide ores, and many of the inorganic compounds, e.g., chloride and sulfate, are soluble in water. Cadmium oxides and sulfides are regarded as almost insoluble species of cadmium. Knowledge of cadmium's solubility in biological media is limited.

Occurrence

The frequent use of fertilizer explains the increased cadmium concentrations in the most heavily polluted areas. Cadmium occurs naturally, is widely dispersed in the environment, and is produced as a byproduct in the production of other metals. Human exposure occurs as a result of inhalation of cadmium-containing dust in industry and consumption of cadmium-containing food in the general population. Owing to its natural occurrence in the geo-environment, some farming products, including tobacco, may be high in cadmium. The concentration of cadmium can be 1–2 μg per cigarette.

Cadmium is present in food as a natural component. Foodstuff contains cadmium, with the highest concentrations being found in liver, kidney, cadmium-contaminated rice, and shellfish.

The Swedish National Board of Food Safety has calculated the contribution of cadmium from foodstuff to give a daily intake of 12 μg per day. This assumption is based on the concentrations of cadmium in foodstuff: meat, fish, and fruit 1–5; cereals, potatoes, root fruits 10–50; bran 150 and kidney and liver 100–400 μg kg^{-1}.

Metabolism and Kinetics

Absorption

Uptake of cadmium via inhalation is 5–35%, depending on the aerodynamic properties, solubility, and particle size. Higher uptake values are related to aerosols with a small particle size, e.g., cadmium in tobacco smoke. Uptake via the gastrointestinal tract in humans is estimated to be 4.6–7%. For humans with low iron stores, absorption values of up to 20% can be seen. Nutritional factors and dose influence the proportion that is absorbed. Conditions increasing the uptake of cadmium via the gastrointestinal tract include low intake of calcium, iron, zinc, and copper. A high intake of fiber can result in a lower level of intestinal absorption of cadmium.

Excretion/elimination

Excretion of cadmium via urine or feces largely depends on the route of exposure. Urinary excretion of cadmium has been demonstrated in a number of experimental studies in laboratory animals to represent about 0.01–0.02% of the total body burden

upon long-term exposure. In many mammalian species, urinary excretion increases slowly upon exposure to cadmium, and after renal damage has occurred, excretion of cadmium increases. For humans, it has been estimated that approximately 0.01% of the body burden is excreted in urine. Urinary excretion, like the body burden of cadmium, is age-dependent. If tubular proteinuria occurs, cadmium excretion increases. A high level of cadmium excretion without proteinuria may occur in short-term, high-level exposure. Cadmium is excreted in urine bound to MT.

Since it is not possible to distinguish net gastrointestinal excretion from unabsorbed cadmium in feces, it is very difficult to study net fecal excretion of cadmium. In oral cadmium exposure, approximately 95% of fecal cadmium represents unabsorbed cadmium. Fecal cadmium can be used as an indicator of oral intake. Studies of injected cadmium in experimental animals show that, initially, fecal excretion is greater than urinary excretion, calculated on a percentage basis. This is probably due to contribution from the bile. Data on fecal excretion of cadmium in humans are almost nonexistent.

Biological Half-life

The biological half-life of cadmium in the kidneys is estimated to be around 20 years in humans. Such a long biological half-life explains why cadmium accumulates constantly up to approximately 50 years of age in humans. Cadmium accumulating in the kidneys is probably largely bound to MT that is synthesized *de novo*. This process may be responsible for the long biological half-life of cadmium in the kidneys and its accumulation in long-term exposure.

Toxicokinetics

Cadmium accumulates in the body with age and has an extremely long biological half-life. Because of its long biological half-life of around 20 years, long-term toxicity is of particular interest. Also, short-term toxicity can be of interest. The kinetics of cadmium is related to chemical species of cadmium. Cadmium in the form of inorganic cadmium ions has a different metabolic pathway in the body compared with cadmium as metallothionein. For all types of damage caused upon cadmium exposure, MT is involved.

Toxicokinetic aspects including the mechanism for transport of cadmium to the kidneys by MT have been extensively studied in mammals and have provided evidence indicating that MT is important for transport of cadmium to the kidneys and that intracellular MT protects renal tubule cells from toxic insult by cadmium. The ratio of MT bound cadmium and nonMT bound cadmium is highly important.

After a single exposure, cadmium in blood plasma is mainly bound to albumin and other larger proteins immediately after uptake from the gastrointestinal tract or the lungs. There is time dependence, with a larger proportion of plasma cadmium being bound to low-molecular-weight plasma proteins (probably MT) at longer time intervals after a single administration. Cadmium bound to albumin is taken up in liver. In liver cells, cadmium–MT complex is dissociated, cadmium not bound to MT induces *de novo* synthesis of MT in liver cells, and the proportion of liver-cadmium bound to MT increases. In long-term exposure, there is a slow release of cadmium-MT from the liver to the blood. As mentioned before, after a single administration of cadmium, plasma-cadmium is mainly bound to albumin, and this form of cadmium is not taken up by the kidneys to any great extent. However, for cadmium injected into plasma as MT, the opposite is found, which means that a major part is taken up in the kidneys. In the tubular cells of the kidneys, the cadmium–MT complex is transported into the lysosomes, where MT is catabolized. The rate of influx of cadmium–MT into the renal tubular cell and the rate of *de novo* synthesis of MT in these cells regulates the pool of intracellular 'free' cadmium ions that can interact with cellular membrane targets in the tubules. When cadmium–MT influx into the lysosomes is limited, MT synthesis may be sufficient to produce enough MT to bind the limited amounts of cadmium. The 'free' cadmium pool is small, and no membrane damage occurs. Calcium transport in the cell is normal. When cadmium–MT influx into the lysosomal compartment is high, and *de novo* synthesis of MT is deficient, the 'free' cadmium pool is sufficiently large to interact with membrane targets to block calcium transport routes. This results in impaired uptake and transport of calcium through the cell with subsequent cell damage and increased excretion of calcium and proteins in urine. Cadmium–MT injection has frequently been used as a model to study several features of nephrotoxicity seen in long-term exposure to cadmium.

There has been some controversy over the role of cadmium–MT as a sequestering agent in the course of cadmium-induced renal tubular toxicity, and also the sensitive site in the cell has not been identified.

A role of metallothionein in transport of cadmium and nephrotoxicity, in modulating cadmium hepatotoxicity was advanced in the 1970s and has gained further support over the years. Cadmium bound to albumin is taken up in the liver. By transplanting cadmium-containing livers to noncadmium-exposed animals, i.e., animals with almost no tissue concentration of cadmium, a gradual uptake of cadmium in

the kidneys was demonstrated. These results support the transport model of cadmium in the body.

Iron deficiency increased levels of MT-1 in bone marrow of rats with hemolytic anemia, whereas the unchanged concentrations in liver and with a reduced renal concentration of MT indicate that MT-1 in blood reflects erythropoietic activity. This might explain how iron deficiency increases the absorption of cadmium. Cadmium is also recognized as a potent inducer of MT synthesis. Heavy exposure to cadmium gives rise to hemolytic anemia with increased urinary excretion, supporting the involvement of MT.

Health Problems Upon Cadmium Exposure

Cadmium has health effects upon both acute and long-term exposure.

Acute Toxicity

Cases of acute toxicity have been described in the past from ingestion of cadmium-contaminated food or drink. The symptoms of acute toxicity are nausea, vomiting, and abdominal pain. Acute toxicity after ingestion of drinks with more than 15 mg per liter of cadmium occurred in children after exposure to cadmium via a soft-drink machine. High cadmium concentrations can occur if acid food comes into contact with cadmium-plated utensils. A fatal case has been described after oral intake of a high dose of cadmium–iodine.

Chronic Toxicity

Human disease has occurred in some special situations, such as itai-itai disease in Japan, and renal dysfunction and increased occurrence of bone effects in Belgium and in China as a result of long-term effects related to cadmium pollution of the general environment.

Long-term exposure to cadmium via air, food, or water increases the cadmium concentration in the kidneys and gives rise to kidney disease. Other adverse effects that may be caused by cadmium exposure are lung damage, bone effects, liver dysfunction, and reproductive toxicity.

Chronic cadmium poisoning occurs as a result of long-term intake of food with increased cadmium concentrations or long-term inhalation of airborne cadmium in industry. Inhalation of cadmium affects the lungs. The most prominent adverse health effects upon long-term exposure to cadmium, regardless of exposure route, are renal tubular dysfunction and glomerular disturbance.

One of the most extreme forms of disease related to cadmium exposure by the oral route is itai-itai disease, described in Japan in the 1960s. Itai-itai disease is characterized by osteomalacia, osteoporosis, renal tubular dysfunction, malabsorption, and anemia. The bone effects are prominent features in humans suffering from itai-itai disease. The characteristics are multiple fractures and deformities of the spine and the long bones, and so far, these have been described only in Japan. Patients would suffer a considerable reduction in body height due to deformities of the spine, resulting from multiple compression fractures. In the long bones, characteristic pathologic fractures with osteoid formation (Milkmans pseudofractures) occurred, which is a typical finding in osteomalacia. Other characteristics included increased excretion of cadmium in urine and kidney damage with proteinuria. Itai-itai disease was a form of renal osteomalacia. The cadmium contamination of rice was due to cadmium contamination from a smelter that discharged material into the Jinzu River, which was used for irrigation of rice fields. Postmenopausal women were affected almost exclusively. The low calcium content of Japanese food may have been a contributing factor, like the tradition of dress to screen away from the sunshine, which gave the women only a low contribution of vitamin D, synthesized by the action of UV-light on the skin.

Epidemiological studies starting in the 1980s from Belgium, e.g., Cadmibel and PheeCad, concerning adverse health effects in humans in Belgium in the general environment reported an increased incidence of health effects occurring at much lower tissue concentrations of cadmium than were previously considered to be related to adverse effects by cadmium. It is interesting to investigate whether bone mineral disorders due to oral intake of cadmium can be related to the estrogen receptor, since it is known that itai-itai disease occurred almost exclusively in postmenopausal women. It has been reported that polymorphism for the estrogen receptor alpha is related to a reduction of bone mineral density in postmenopausal women in Japan. The genetic distribution of the receptor was, however, the same in itai-itai disease subjects. Nothing has been reported for the estrogen beta-receptor.

Epidemiological evidence indicates that diabetics might have an increased susceptibility to develop renal dysfunction upon cadmium exposure.

Kidney

Cadmium is a toxic metal with an extremely long biological half time, 15–20 years in humans. Upon exposure, adverse health effects may occur primarily as renal tubular dysfunction. Thus, the kidney is regarded as a critical organ independent of the exposure route. Cadmium is classified as a human

carcinogen based on the occurrence of lung cancer after inhalation.

After uptake cadmium in the blood is initially taken up by the liver and subsequently slowly redistributed to the kidneys because of its binding to the low-molecular-weight protein, MT. Cadmium bound to this protein in blood accumulates selectively in the kidneys. In long-time exposures, cadmium gradually accumulates, particularly in the kidneys but also in other tissues. When the concentration in the kidney cortex reaches 50 mg kg^{-1} or higher, damage to renal tubular cells may occur in the most sensitive persons, and the kidneys have a decreased ability to reabsorb proteins with increased concentrations of low-molecular-weight proteins and calcium in urine. As a result of long-term calcium losses and other metabolic changes related to the renal damage, osteoporosis, and, in severe cases, osteomalacia, occur (itai-itai disease; see Chronic Toxicity).

Cadmium-induced renal dysfunction influences cellular calcium metabolism. Cadmium accumulates in the kidneys, probably largely bound to MT that is synthesized *de novo*. This process may be responsible for the long biological half-life of cadmium in the kidneys and its accumulation after long-term exposure. The biological half-life of cadmium in the kidney is estimated to be around 20 years in humans. This explains why cadmium accumulates constantly up to approximately 50 years of age in humans. Renal toxicity develops when the total tissue concentration of cadmium in the renal cortex reaches levels between 50 and 300 $\mu g\, g^{-1}$. Renal tubular dysfunction occurs in various individuals at different total concentrations of cadmium in the kidney, depending on the variation among individuals in the ability to synthesize MT and other protective components.

Liver

Liver toxicity caused by cadmium can be modified by MT. Cadmium bound to albumin is taken up in the liver. Upon catabolism of MT, unbound cadmium is free to cause toxicity to liver cells. This happens after high cadmium exposure when MT levels are not high enough to handle all nonbound cadmium ions, e.g., the 'free' cadmium ions, Genetic polymorphism can influence the toxicity of cadmium, which is supported by studies in laboratory knock-out animals. The role of MT in cadmium-induced hepatotoxicity and nephrotoxicity, as well as in zinc-induced protection showed that MT-null mice were more sensitive to i.p. $CdCl_2$ hepatotoxicity. Zinc pretreatment by subcutaneous injection increased hepatic MT 80-fold in control mice, but not in MT-null mice, and prevented $CdCl_2$ hepatotoxicity in control mice only.

Endocrine Effects

Cadmium does not pass the blood–brain barrier, except when exposure is via nasal mucosa. Low-level cadmium exposure of lactating rats caused alterations in brain serotonin levels in the offspring.

Epidemiological evidence indicates an increased susceptibility in diabetics to developing cadmium-induced renal dysfunction, which might be related to changes in MT expression in diabetics.

Experimental studies on cadmium–MT nephrotoxicity show an increase in genetically diabetic as compared with normal Chinese hamsters. The Chinese hamster develops a subnormal pancreatic insulin release, insulin resistance in the liver, and sometimes hyperinsulinemia. These animals develop diabetic nephropathy with glomerular and increased effects of the glomerulus. The exact mechanisms behind cadmium-induced renal damage in diabetics need to be elucidated by further studies.

Physiological Conditions and Pregnancy

Pregnant rats have a faster decrease in hepatic cadmium levels and a faster increase in kidney cadmium levels compared with nonpregnant rats. Pregnant rats also have higher plasma cadmium levels and MT levels compared with nonpregnant rats. It is known from that that placental necrosis, fetal death and a condition in the dam similar to 'toxemia of pregnancy' develops after injection of cadmium in the late stages of pregnancy, including congenital malformations. Changes in brain development and related changes in brain enzymes have been reported in pups from rats treated orally with cadmium during pregnancy and lactation. Cadmium does not pass the placental barrier. This may have several explanations, including a sequestering role of MT. MT increases in the placenta of smoking pregnant women. The fetus is likely to be protected from cadmium exposure because the MT concentration increases in placental tissue. Human fetal liver contains MT rich in copper. There is a tendency in non-/ex-smokers towards increasing cadmium-blood levels during pregnancy, reaching a maximum at delivery, whereas an opposite tendency has been shown for smokers. The increase in nonsmokers may be related to the mobilization of cadmium from the liver to the blood during pregnancy.

Biological Monitoring

Biological monitoring of cadmium can be performed by measurements of cadmium in blood and urine. Cadmium in blood indicates ongoing or recent exposure to cadmium, and cadmium in urine reflects body burden. However, after external exposure to cadmium

has stopped, blood cadmium also reflects the body burden as the cadmium stored in the liver is released into the bloodstream. A metabolic model for cadmium can be used to estimate closeness to renal failure. Concentrations in urine and blood related to the risk of adverse effects are presented in the section on Risk Assessment below. Placental samples, if taken correctly, constitute a useful material for monitoring cadmium exposure in the human population. However, it is important to be aware that differences in cadmium within parts of the human placenta exist. Placental cadmium concentrations were four and six times higher than maternal and umbilical cord blood cadmium, respectively, Calcium, protein, and magnesia are excreted in urine after exposure to cadmium.

Cadmium analyses in biological samples should be handled carefully. The risk of contamination during sampling is largely due to contamination from smokers and if the persons handling the samples are smokers. Lack of knowledge of the smoking habits of the subject such as whether they are a present smoker, a previous smoker, or a person who has never smoked, is an another pitfall in interpreting concentration results in the sample. The long biological half-life of cadmium can influence the results up to 10 years after cessation of smoking. It is important to take into account smoking habits when drawing conclusions about the status of cadmium exposure and risk of development of adverse health effects due to cadmium exposure present or past. MT can be used as a bioindicator in monitoring both exposure and effects related to metal exposure. Humans with a low iron status and diabetics have been found to be more sensitive to cadmium exposure.

Risk Assessment

Risk assessment serves as a basis of preventive action for avoidance of adverse health effects in populations exposed to chemicals. Whether an agent causes reversible or irreversible effects is of importance in determining whether an effect is to be classified as a critical effect. It is important to define which characteristics and which cut-off level are chosen. The critical effect of cadmium exposure both in the general and work environment is renal tubular dysfunction. Renal damage occurs at a much lower cadmium concentration than previously estimated. Evaluations of recent studies in the general population concluded that an average level of cadmium in urine of 2.5 μg per gram of creatinine is related to a few percent excess prevalence of renal tubular dysfunction. A kidney cadmium concentration of 50 μg per gram of cortex wet weight, is reached after decades of a daily cadmium intake of 50 μg and results in excretion of cadmium in urine of 2.5 μg per gram of creatinine. This constitutes a higher risk of tubular damage than indicated by earlier estimates. These risk estimates are made for a population group in total. Estimates of risks for the development of renal tubular dysfunction have to be made for certain identified groups with increased risk, such as subjects with low iron stores and smokers. It has been suggested that a small percentage (1–5%) of such groups may develop renal tubular dysfunction with a lifelong daily intake of as little as 15 μg.

Oral intake of cadmium in high single dose via food or drink gives rise to vomiting, abdominal pain, and diarrhea in humans. Drinks of 16 mg of cadmium per liter and higher, corresponding to a dose of 3 mg and higher can result in the aforementioned symptoms. Long-term exposure via food in the form of rice with concentrations of around 1 mg kg^{-1}, which corresponds to an intake of 600 μg of cadmium per day, causes signs of malabsorption and some less pronounced symptoms. A weekly intake of 7 μg per kilogram of body weight corresponds to a daily intake of 70 μg or 1 μg per kilogram of body weight, which is equal to the provisional tolerable weekly intake (PTWI) for cadmium set at 500 μg. Lower lifelong exposures via food or drinks cause renal dysfunction in humans. This will likely reduce the present PTWI considerably.

After intake of 43 μg per kilogram of body weight as a single oral dose, early symptoms from the gastrointestinal tract in humans can be seen. This dose is set as an estimated lowest-observed-adverse-effect level (LOAEL) for a single dose of cadmium. Calculations based on LOAEL and a safety factor of 3–10 give a tolerable single dose between 0.3 and 1 mg (4–14 μg per kilogram of body weight). Long-term exposure, e.g. from months to a few years of daily intakes of 200 μg of cadmium (3 μg per kilogram of body weight), might be tolerated with the lack of symptoms from the gastrointestinal tract.

These doses for humans have not, at present, been connected with convincing evidence to cause reproductive or developmental effects. Because of this, a daily intake above 1 μg per kilogram of body weight should not be acceptable because of sensitive subpopulations such as children and pregnant or lactating women and humans low in iron status.

The PTWI is set to allow a certain variation of intake during a week provided that the weekly intake is not exceeded. Exposures above PTWI have to be compensated for by intakes below the PTWI for some time in order to avoid effects of the cumulative dose by long-term effects, e.g., renal dysfunction. Dietary intake of cadmium has been estimated to be 12–50 μg per day on average in various countries. The FAO/WHO has set a

PTWI of cadmium to 400–500 μg for adult persons, which corresponds to approximately 1 μg per kilogram of body weight for each day of the week.

A drinking-water guideline value of $5\,\mu g\,l^{-1}$ has been set by the WHO, which the US Environmental Protection Agency allows in drinking water. The Environmental Protection Agency regulates the cadmium concentration in lakes, rivers, waste sites, and cropland. The US Food and Drug Administration has a cadmium limit in food colors of $15\,mg\,kg^{-1}$. Sweden lacks threshold limit values for cadmium in food and is actively encouraging international organizations to set recommendations for the concentration of cadmium in food. A national recommendation restricts the intake of mammalian liver and kidney to one to two per week, or no consumption at all, depending on the age of the animal.

See also: **Copper**: Properties and Determination; Physiology; **Zinc**: Properties and Determination; Physiology; **Risk Assessment**

Further Reading

IARC (1993) *Beryllium, Cadmium, Mercury and Exposures in the Glass Manufacturing Industry Monograph*, vol. 58, pp. 119–237. Lyon, France: IARC.

Jarup L, Berglund M, Elinder CG, Nordberg G and Vahter M (1998) Health effects of cadmium exposure – a review of the literature and a risk estimate. *Scandinavian Journal of Work and Environmental Health* 24 (supplement 1): 1–51.

Nordberg GF, Kjellström T and Nordberg M (1985) Kinetics and metabolism. In: Friberg L, Elinder CG, Kjellström T and Nordberg GF (eds) *Cadmium and Health*, vol. 1, pp. 103–178. Boca Raton, FL: CRC Press.

Nordberg GF, Alessio L and Herber RFM (eds) (1992) *Cadmium in the Human Environment: Toxicity and Carcinogenicity*. IARC Scientific Publications No. 118. Lyon, France: IARC.

Nordberg M (1984) General aspects of cadmium: Transport, uptake and metabolism by the kidney. *Environmental Health Perspectives* 54: 13–20.

Nordberg M and Kojima Y (1979) Metallothionein and other low molecular weight metal-binding proteins. In: Kägi JHR and Nordberg M (eds) *Metallothionein*, pp.41–116. Basel: Birkhauser.

Nordberg M and Nordberg GF (2000) Toxicological aspects of metallothionein. *Cellular and Molecular Biology* 46: 451–463.

Nordberg M and Nordberg GF (2002) Cadmium. In: Sarkar B (ed.) *Handbook of Heavy Metals in the Environment*, pp. 231–269. New York: Marcel Dekker.

WHO/IPCS (1992) *WHO/IPCS Environmental Health Criteria Document 134 Cadmium*. Geneva: WHO.

CAFFEINE

J Watson, Southampton General Hospital, Southampton, UK

Background

Caffeine is a purine-based chemical that is widely consumed throughout the world. In this article, consumption of caffeine and its subsequent metabolism are considered in detail. This is followed by a detailed description of the important physiological and psychological effects of this stimulant. The phenomenon of caffeine withdrawal is described because it is so commonly experienced and has important implications on caffeine consumption on a worldwide level. Finally, the therapeutic uses of this substance are considered.

Caffeine is a chemical substance that has been isolated from approximately 60 plants. In its pure form, it is a bitter-tasting, white, odorless, crystalline powder. Its botanical function is thought to be both an antifungal and insecticidal agent.

Caffeine was first isolated in 1820, by Runge and van Giese. It was included in the medical vocabulary in 1823, but it was not until 1875 that its structural formula was described by Medicus. Caffeine has a purine base, and its chemical name is 1,3,7-trimethylxanthine (**Figure 1**).

Caffeine is one of the most widely consumed substances in the world and, in moderate doses, is generally considered to have the effect of a mild stimulant, which is helpful in relieving minor fatigue and boredom, with a small risk of harmful effects. In view of

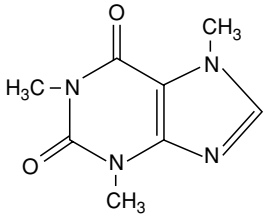

Figure 1 Chemical structure of caffeine: $C_8H_{10}N_4O_2$. Molecular weight: 194.19.

its ubiquitous use and popularity, an enormous investment has been made to examine health issues associated with its regular use.

Caffeine Consumption

There are various vehicles for caffeine consumption: some are internationally recognized, such as tea and coffee, but others are only locally consumed, e.g., maté in South America and miang in Thailand. Worldwide, the average caffeine consumption is estimated at 50 mg per day.

The properties of caffeine have been known since preliterate times, when its use was confined to specific areas – those where coffee plants grew wild (e.g., Ethiopia), until the time of European colonization, in the seventeenth century. At this time, caffeine-containing products were cultivated and subsequently became widely distributed. The combined physiological and psychological impact of caffeine depends on the pattern and degree of exposure to the drug. Thus, there has been extensive investigation into the caffeine content of widely consumed beverages and their consumption.

Caffeine Content of Beverages and Food

There is a marked variation in the estimates of caffeine content in different beverages. This arises from differences in:

- methods of preparation, e.g., methods of brewing tea and coffee;
- product source, e.g., black or green tea;
- reference volumes; and
- analytical methods.

Table 1 lists standard values that are widely quoted with respect to caffeine content.

Caffeine Consumption Studies

It is estimated that at least 80% of the adult population in the majority of countries worldwide consume caffeine on a regular (daily) basis. Variation exists in the nature of the caffeine vehicle. In the UK, 81% of adults drink tea, whereas in France, 81% drink coffee. In the USA, it is estimated that only 52.1% drink coffee, but more soft-drink beverages are consumed than in other countries.

Data for estimating a population's caffeine intake have been available from prospective and retrospective studies. In general, the former has gained more credence, and it has been shown repeatedly that there is a good correlation between reported caffeine intake and subsequent appropriate measurement of caffeine or caffeine metabolite levels. Table 2 shows the estimated mean intake of caffeine in several countries.

Table 1 Caffeine content of beverages and foodstuffs

Drink	*Caffeine content*
Coffee	
Ground roast	85 mg per 150 ml
Instant	60 mg per 150 ml
Decaffeinated	3 mg per 150 ml
Tea	
Leaf or bag	30 mg per 150 ml
Instant	20 mg per 150 ml
Cola	18 mg per 180 ml
Caffeine-free cola	0
Cocoa (hot chocolate)	4 mg per 150 ml
Chocolate milk	4 mg per 150 ml
Chocolate candy	5–20 mg per 100 g

Table 2 Estimated average daily consumption in certain countries

Country	*Mean intake (mg kg^{-1})*
USA	4
UK	3
Denmark	7

Pharmacokinetics of Caffeine

Absorption

Caffeine absorption from the gastrointestinal tract is rapid. About 90% of the caffeine in a cup of coffee is cleared from the stomach in 20 min, although the stomach is also able to absorb this substance. A wide variation in the time to reach peak plasma concentrations has been reported in humans (i.e., 15–120 min). However, absorption may be slowed by the presence of food in the intestine or when large amounts of the drug are ingested. On an empty stomach, the time taken is around 40 min, with 99% absorbed after 45 min.

Distribution

Once absorbed, caffeine, because of its hydrophobic properties, can pass through all biological membranes, including the blood–brain barrier. Salivary caffeine concentrations have been shown to correlate closely with plasma concentrations, reaching up to 85% of plasma levels.

After oral ingestion of 5–8 mg kg^{-1} in humans, the peak plasma caffeine concentration equals 8–10 mg l^{-1}. From this, it can be estimated that ingestion of a single cup of coffee, which provides a dose of 0.4–2.5 mg kg^{-1}, would give a peak concentration of 0.25–2.0 mg l^{-1}. In normal adults, the plasma half-life is between 3 and 7 h. Whereas there is little or no difference with increasing adult age, premature and full-term neonates show a markedly increased plasma

half-life, owing to the lower activity of cytochrome P450 and the relative immaturity of some elimination pathways. This state continues for up to 12 months but is then essentially stable throughout life.

Although the clearance rate is overall similar for men and women, it is related to the menstrual cycle, being slower in the luteal phase (6.85 h luteal vs. 5.54 h follicular). In the third trimester of pregnancy there is a significant prolongation of the plasma half-life (up to 15 h). In adult males who smoke, the caffeine half-life is decreased by up to 50% compared with nonsmokers.

Caffeine Metabolism

Caffeine is metabolized by the cytochrome P450 enzyme system. The major process is demethylation at the C3 position to produce paraxanthine, which accounts for up to 80% of caffeine metabolism. The other 24 excretory products include theophylline (1,3-dimethylxanthine), theobromine (3,7-dimethylxanthine), and both uric acid and uracil derivatives. Many of the metabolic steps may be saturable in humans, as the elimination half-time for caffeine and some of its metabolites is dose-dependent. Less than 2% of the ingested dose of caffeine is recoverable from the urine unchanged.

Mechanism of Caffeine Action

Caffeine exerts a variety of pharmacological actions at diverse sites, both central and peripheral. Whilst several processes have been investigated as being responsible for these effects, the mechanism must be activated by caffeine consumption within normal human amounts. This effectively rules out some hypotheses (Table 3).

Thus, the general consensus now is that the only known mechanism that is significantly affected by the relevant doses of caffeine (i.e., able to achieve an extracellular concentration of 10–50 μM) is binding to adenosine receptors and competitive antagonism of adenosine. Although the effects of caffeine on behavior have been well documented, understanding of its cellular mechanisms of action remains incomplete.

Adenosine receptors are widely distributed throughout the body, which befits the mechanism of caffeine action because of the substance's wide range of effects. For example, adenosine causes bronchoconstriction and negative inotropic effects on the heart, and inhibits gastric secretions, lipolysis, and renin release, all of which are the opposite of the effects of caffeine. Centrally mediated effects of adenosine are also the reverse of caffeine.

Four distinct adenosine receptors have been identified and cloned: A_1, A_{2A}, A_{2B}, and A_3. In humans, A_1 and A_{2A} have been shown to be important, as the other two are only stimulated with toxic concentrations of caffeine. Whereas caffeine is equally active at adenosine A_1 and A_{2A} receptors, it has been shown recently that the stimulatory central action of caffeine is mainly mediated through an inhibition of transmission of adenosine A_{2A} receptors.

Physiological Effects of Caffeine – General

The effects of caffeine are legion. Table 4 summarizes the effects of caffeine on the body, with the following sections describing the physiological effects on the functions of the cardiovascular, central nervous, and neuroendocrine (including adrenomedullary) systems. Psychological effects are considered separately.

Physiological Effects of Caffeine on the Cardiovascular System

The effects on caffeine on the cardiovascular system have been extensively researched in both experimental and epidemiological studies. It has generated much interest because of the debate regarding the

Table 3 Mechanisms of caffeine action

Potential mechanism of caffeine action	*Reference*	*Relative caffeine concentration required for effective activation (1 = a cup of coffee)*
Adenosine receptor antagonism	Fredholm (1980)	× 1
Inhibition of phosphodiesterase		× 10
γ-Aminobutyric acid receptor blockade	Marangos *et al*. (1979)	× 20
Mobilization of intracellular Calcium	McPherson *et al*. (1991)	× 500

Fredholm BB and Hedqvist P (1980) Modulation of neurotransmission by purine nucleotides and nucleosides. *Biochemical Pharmacology* 29(12):1635–1643. Marangos PJ, Paul SM, Parma AM, Goodwin FK, Syapin P and Skolnick P (1979) Purinergic inhibition of diazepam binding to rat brain (in vitro). *Life Science* 24(9): 851–857. McPherson PS, Kim YK, Valdivia H *et al*. (1991). The brain ryanodine receptor: a caffeine-sensitive calcium release channel. *Neuron* 7(1):17–25.

Table 4 Summary of the physiological effects of caffeine on the body

	Effect of caffeine
Gastrointestinal system	Increases gastric acid secretion in susceptible individuals, not related to gastric ulcers
	Hepatic dysfunction increases caffeine half-life
	Digestive absorption of alcohol is increased when caffeine is consumed beforehand
Respiratory system	Relaxes bronchial smooth muscle
	Increases CNS sensitivity to carbon dioxide
Renal system	Stimulates renin release
	Diuresis resulting from increased glomerular filtration rate and decreased tubular reabsorption
Adipose tissue	Stimulates lipolysis
Bone tissue	In large amounts, coffee can contribute to a negative calcium balance in those with a low calcium intake

involvement of caffeine in cardiovascular disease. Experimental studies have investigated the effects of controlled caffeine exposure, whereas epidemiological studies have addressed the relationship of caffeine to cardiovascular disease.

Caffeine and Blood Pressure

The major cardiovascular effect of caffeine is peripheral vasoconstriction, resulting in elevated blood pressure, which is not accompanied by an increase in cardiac output. Acute tolerance to the pressor response does occur in regular consumers but is reinstated by a brief period of abstinence, as demonstrated by the referenced studies. This may be as short as 3 h. Other evidence has demonstrated that a second cup of coffee induces a pressor effect, albeit diminished. Finally, ambulatory blood pressure measurements on days of caffeine intake compared with placebo days are higher in usual caffeine consumers. Both systolic and diastolic blood pressures have been reported as being more significantly affected by an acute caffeine challenge.

In addition, caffeine increases cardiovascular responses to stress. This includes the pressor response to mental or physical stress, as well as other responses, e.g., cardiac output and forearm blood flow. As caffeine has pressor effects, it has been recommended as a treatment for orthostatic and postprandial hypotension.

Caffeine and Heart Rate

The effect of caffeine on heart rate is less well defined. Consumption within normal limits has been reported as being associated with a decrease, an increase or no change in heart rate. Whilst caution is advised with regard to tachyarrhythmias induced by caffeine, this is associated with excessive doses or underlying cardiac disease. Interestingly, it is widely believed that caffeine is associated with palpitations, tachycardia, and arrhythmias, by physicians.

Physiological Effects of Caffeine on the Central Nervous System

The physiological effects on the central nervous system are manifold. These include the effects of caffeine on cerebral blood flow, cerebral metabolism, and electrophysiological activity. The psychological effects will be considered separately.

Cerebral Blood Flow

Caffeine causes cerebro-vasoconstriction 30 min after ingestion, via antagonism of the vasodilatory properties of endogenous adenosine. This effect was first demonstrated over 60 years ago. A dose of 250 mg of caffeine causes a decrease in cerebral blood flow of 15–18% as measured by the Xenon inhalation technique. The effects of caffeine on cerebral blood flow have been demonstrated to last for 90 min and are independent of mood and arterial partial pressure of carbon dioxide.

Cerebral Metabolism

Following caffeine administration, the stimulant effects on the central nervous system (CNS) are associated with changes in local rates of cerebral energy metabolism. In rat studies, a caffeine dose of $10\,mg\,kg^{-1}$ causes increases in the motor and limbic systems, as well as the thalamus (specific nuclei only), substantia nigra, ventral tegmental areas, locus ceruleus, and raphe nuclei. These changes in glucose utilization correlate with changes in locomotor activity in rats and other behavioral modifications. The multiple areas of the brain that are affected by caffeine mean that, although the primary action of caffeine is to block adenosine (a neuromodulator), other neurotransmitters such as dopamine, norepinephrine, serotonin, acetyl choline, and glutamate are involved.

Neuroelectrophysiology

Caffeine is a cerebral stimulant, and thus its effects can be measured on an electroencephalograph. The

most consistent results from a number of studies is that moderate doses of caffeine effect the decrease in spectral power in the lower alpha and theta ranges. In addition, P300 amplitude increases with confidence, attention, arousal level, and reward. Such changes have been demonstrated with caffeine consumption.

Physiological Effects of Caffeine on the Neuroendocrine Function

The effects of caffeine in humans on the endocrine system appear to be confined to the endocrine response to stress, especially production of catecholamines and cortisol, although other hormones may be important too, e.g., growth hormone. Plasma concentrations and urinary excretion of epinephrine and norepinephrine are increased by caffeine consumption. Although the methods of augmentation have not been described in detail, it has been suggested that caffeine-stimulated release of adrenomedullary catecholamines is mediated by peripheral sympathetic nerves. Thus caffeine antagonism of adensoine receptors may result in increased sympathetic stimulation of the adrenal gland, leading to increased thyroxine hydroxylase activity and thus catecholamine synthesis. Like increased adrenaline production, norepinephrine synthesis is controlled centrally by sympathetic outflow.

Cortisol production is similarly augmented by caffeine and stress. This is thought to be mediated by an increase in adrenocorticotrophin. This action could also be partly due to caffeine's action on prostaglandins. The effects of caffeine on growth hormone and prolactin are less well defined.

Psychological Effects of Caffeine Consumption

Having previously discussed the effects of caffeine at a receptor and neuronal level, this section discusses the effects of caffeine on psychological responses. A large number of functions are affected, which would be beyond the range of this review, and so three specific functions are considered.

Behavior

Caffeine is erratically self-administered in animal models of drug self-administration. In humans, both self-administration and choice procedures have shown that caffeine acts as a reinforcer. In choice studies, subjects typically first sample caffeine or placebo and later have the opportunity to choose to self-administer one. Caffeine is chosen in preference to placebo in subjects with or without a prior history of caffeine consumption and is especially marked in those with a history of drug abuse. To date, the mechanism of action of caffeine on behavior in humans has not been extensively investigated. It is likely that an interaction between adenosine and dopamine receptors is important.

Performance

Owing to the enormous consumption of caffeine worldwide, there has been a vast amount of work on the performance effects of caffeine. It is far from clear how the benefits and drawbacks of its consumption balance out. In this section, the evidence for the psychostimulant effects of caffeine on performance is considered, as these are widely held to be the main benefits of caffeine use.

Numerous placebo-controlled studies have been published. Although the results are varied, with some studies showing no effect of caffeine, taken as a whole, they confirm a psychostimulant action of caffeine.

Whilst a wide spectrum of cognitive tests have been investigated in the above studies, usually, prior abstinence of caffeine was required at least overnight. This limits the relevance to caffeine consumption in everyday life. The problem with caffeine deprivation prior to testing, especially for a short period, means that symptoms of caffeine withdrawal can be invoked. Thus, it is not possible to describe the net effects of caffeine with these studies, especially as caffeine withdrawal has also been shown to decrease performance in both humans and animals. However, evidence in this area is inconclusive. A caffeine challenge of only 70 mg with 1.5 h of caffeine deprivation has been shown to improve simple reaction time.

Another standpoint in this discussion is founded on the demonstration of improved psychomotor performance in noncaffeine-deprived individuals. A large (9003 subjects) correlational study showed that there was a relationship between habitual caffeine consumption and performance (reaction time – simple and choice, verbal memory, and visuospatial reasoning), even when all variables were controlled for.

Mood

The effects of caffeine on mood have been studied extensively, and the results of these studies have been more consistent than in other areas of caffeine research. Caffeine in lower doses (up to 250 mg) has positive effects on mood in both caffeine-deprived and caffeine-tolerant studies. From these and many other studies, the benefits of acute caffeine ingestion on mood include an increase in energetic, confident, and alert feelings, as well as motivation to work, concentrate, or socialize.

Caffeine Withdrawal

The phenomenon of the caffeine withdrawal syndrome is well substantiated and characterized in detail. Table 5 lists the symptoms identified to occur in the caffeine-withdrawn state.

It is noteworthy that the prevalence of headache in the general population is 27%, yet only 6% coffee drinkers and 6% tea drinkers associate headache with omission of caffeine intake.

The time course for these symptoms to develop is short. Overnight abstinence is part of the usual pattern of caffeine consumption, and it is usual for withdrawal symptoms to develop. The symptoms increase for 24–48 h and then begin to improve. Usually, they have subsided by day 7. A high daily intake of caffeine can lead to a prolongation of the symptoms for up to 10 days or more. This may be explained by the fact that the rate of caffeine metabolism may vary by 13-fold throughout the population. It has been shown that caffeine consumption is directly related to the symptoms of caffeine withdrawal. A dose of 100 mg of caffeine per day is sufficient for caffeine withdrawal symptoms to occur. Three days of caffeine consumption is required before withdrawal symptoms occur.

Much work has been published on the subjective effects of caffeine withdrawal, but less work has considered performance. Although psychomotor impairment is not associated with up to 18 h of caffeine deprivation, impairments have been reported in a variety of tasks after 24 h.

The above summarizes the parameters whereby caffeine withdrawal occurs in an otherwise healthy population. It can be so disruptive to life that in 1994, the American Psychiatric Society included it in the DSM IV, although only as a 'proposed,' rather than an official, category.

Table 5 Symptoms of caffeine withdrawal

Increased	*Decreased*
Irritability	Well-being
Blurred vision	Desire to talk/socialize
Drowsiness/sleepiness	Urge to do work-related tasks
Yawning	Energy/active
Lethargy/fatigue/tiredness/ sluggishness	Contentment
Thick head	Flu-like feelings
Headache	
Self-confidence	
Heavy feeling in arms and legs	
Hot or cold spells	

Therapeutic Uses

A number of clinical uses have been suggested for caffeine. However, the only widespread therapeutic use at present is stimulating apneic premature infants. Initially, theophylline was identified to increase the respiratory rate, but caffeine had succeeded this because of its wider therapeutic index. Following a loading dose of caffeine, plasma caffeine levels are maintained carefully between 10 and 20 mg l^{-1}.

Other applications for caffeine include treatment of asthma and chronic obstructive pulmonary disease, because of its bronchodilatatory effects. In addition, it has been used to extend the duration of seizure, induced by electroconvulsive therapy. Uses in hypotension and allergic rhinitis have also been described. Caffeine may also aid the recognition of hypoglycemia induced by treatment of diabetes mellitus. Finally, caffeine is available in over-the-counter medicines for simple analgesia, cold remedies, and stimulatory preparations.

See also: **Behavioral (Behavioural) Effects of Diet**; **Cocoa**: Chemistry of Processing; Production, Products, and Use; **Coffee**: Green Coffee; Roast and Ground; Instant; Analysis of Coffee Products; Decaffeination; Physiological Effects; **Hormones**: Adrenal Hormones; Thyroid Hormones; Steroid Hormones; **Hypertension**: Hypertension and Diet; **Tea**: Types, Production, and Trade; Chemistry; Processing; Analysis and Tasting

Further Reading

Barone J and Roberts H (1996) Caffeine consumption. *Food and Chemical Toxicology* 34: 119–129.

Debry G (ed.) (1994) *Coffee and Health*. Paris: John Libbey Eurotext.

Nehlig A (1998) Are we dependent upon coffee and caffeine? A review on human and animal data. *Neuroscience and Biobehavioral Reviews* 23: 563–576.

Silverman K, Evans SM, Strain EC and Griffiths RR (1992) Withdrawal syndrome after the double blind cessation of caffeine consumption. *New England Journal of Medicine* 327: 1109–1114.

CAKES

Contents

Nature of Cakes

S P Cauvain, Campden and Chorleywood Food Research Association, Chipping Campden, UK

Definition

The definition of cake will vary slightly in different parts of the world but essentially the term refers to products which are characterized by formulations based on (wheat) flour, sugar and (whole) eggs and other liquids (milk), to which fat or oil may be added. The level of added liquids is such that a low-viscosity batter is formed rather than a dough. A key difference between cake and bread is that the former is most commonly baked in a mold, pan, or tin in order to contain the batter and to form a product of a given shape. Minimal processing of the batter is required before baking.

There is no significant gluten formation in cake batters and, indeed, cake-making technology exploits positive steps to prevent gluten formation. Modern cakes are produced by forming a complex emulsion and foam – the batter – which is processed by being heat-set. As a broad generalization, modern cake batters can be considered as fat, or oil, in water emulsion systems. The aqueous phase contains the dissolved sugar and suspended flour particles. In many batter systems air bubbles are incorporated by entrainment in the solid fat and they are held there rather than the aqueous phase. However, as the batter warms during baking, the air bubbles transfer from the fat to the aqueous-phase foam and expand. Later in the baking process the foam sets to yield the cake structure.

The key structure-forming component of cakes is the starch contributed by wheat flour and the modification of its gelatinization characteristics through the addition of sugars and high levels of liquids (egg, milk, water).

Cakes are characterized by being relatively high-density products, typically 0.4–0.7 ml g^{-3}, with a tender-eating, friable crumb, and sweet taste. Their final moisture contents typically lie in the region of 18–28%, lower than that of bread but higher than that of pastries, biscuits, and cookies. The organoleptic and mold-free shelf-life of cake products varies according to their formulations.

A Brief History of Cake Making

Representations of products described as 'cakes' are known from the twelfth century BC in Egypt. To what extent the products depicted equate to our modern view of cake is uncertain. Baking molds appear to have been used, since many of the products take the form of animals of the time. For example, a fried product resembling a snake is described. The first likely use of sweet substances in cooking/baking appears to belong to the ancient Greeks who knew of the special properties of the 'sackharon' or 'sweet reed' (sugarcane) from the Indian subcontinent. The advance of cane sugar into the Middle East was complete by about 700 BC. Sugar consumption continued to rise and spread, first with the expansion of the Roman Empire and later throughout medieval Europe.

Early cake recipes were probably based on the use of honey and in Roman times such confections were made for special occasions such as the Saturnalia and the New Year's fair. Recipe books of the time record the addition of honey to dough to make such cakes. In the medieval period cake recipes and their production are recorded in many European contexts. Access to sugarcane gradually improved with the voyages of discovery to the east and the Americas and the production of sweetened cakes increased and their forms proliferated through the sixteenth, seventeenth, and eighteenth centuries AD.

Types of Cakes

The term 'cake' is taken to include the forms listed in Table 1 and variants thereof. Cakes may be classified into six arbitrary categories:

- Low-fat sponges, containing less than 5% fat.
- High-fat sponges, containing more than 5% fat.

Table 1 Examples of cake formulations

	Sponge cake	*High-ratio cake*	*Chocolate cake*	*Fruited cake*
Flour	100	100	100	100
Sugar (sucrose)	105	115	130	120
Fat/oil	5	35	55	50
Baking powder	3	5	5	3
Whole eggs	50	30	30	45
Milk solids	5	6	8	5
Water	65	100	120	65
Cocoa powder			25	
Dried vine fruits				140

Recipes are based on 100 parts (e.g., 100 g, 100 kg, 100 lb) of flour.

- Low-ratio cakes, in which the levels of sugar and liquids in the formulation are individually less than the flour weight.
- High-ratio cakes, in which the levels of sugar and liquids in the formulation are individually greater than the flour weight.
- Fruited cakes, which have a proportion of dried vine fruits, fresh fruit, nuts, or other particulate materials in the formulation.
- Chocolate cakes, containing a proportion of cocoa solids.

Some examples of typical sponge and cake formulations are given in Table 1.

The Nature and Roles of the Ingredients Used

Flour

Cakes can be manufactured with most types of wheat flour. The most important contribution of the wheat flour is the starch, the gelatinization of which forms the major structural component of cakes. The hydrated wheat flour proteins form a significantly weaker gluten structure in cake batters than in bread dough because of the disruptive effects of the recipe sugars and fats in the cake formulation. In addition, the high levels of recipe liquids produce a low-viscosity batter which reduces the ability of the mixing action to impart the necessary energy for gluten formation. A protein continuum may form in the batter and this does make a small contribution to cake strength and gives a firmer eating quality.

Most commercial cake flours are characterized by the following key features:

- Particle size – improved cake quality is obtained when the particle size is reduced below that normally obtained from a standard flour-milling process. This is usually accomplished by regrinding the flour or through the application of air-classification techniques. The aim of both methods of particle size reduction is to free the starch granules from the enveloping protein matrix. Ideally, the maximum particle size of cake flours should be $< 90\,\mu m$.
- Postmilling treatment – the most common forms of treatment are exposure to chlorine gas and dry-heat treatment. Although the two treatment methods are quite different, the resulting change in flour properties is quite similar–notably, they cause changes in the hydrophobicity of the starch granules. The treatment makes the flours more suited to the production of the high-ratio type of cake. (Note that treatment of cake flours with chlorine gas is not now permitted in many countries.)

Sugars and Sweeteners

Sugar provides sweetness in cakes but also plays a significant role in batter aeration and structure formation. The influence on structural formation comes from the effect of the sugar solution formed in the batter on the gelatinization temperature of the starch. The higher the sugar concentration, the higher the starch gelatinization temperature and therefore the longer the time that the batter has to expand before it sets. Figure 1 illustrates the effects of increasing sugar level.

The most common sugar used in cake making is sucrose, though glucose and other sugars may be used. If reducing sugars are used, there is a tendency for discoloration of the crumb to occur at higher rates of addition.

Eggs

Whole egg is a common component of cakes, providing color, flavor, aeration, and structure-forming properties. The aeration and structure-forming properties come from the proteins which are present in the egg white. The addition of fat will interfere with the aeration power of egg proteins.

Some cake types, e.g., angel cake, may be made without the use of egg yolks. Using only the egg white confers a white color to the final product.

Milk

Milk products are commonly used to provide flavor and a source of water to dissolve the added sugar. Dried-milk products may also be used, provided another source of water is added in the recipe. Skimmed milk powder can be used for color and flavor. The powder should have been heat-treated before use, otherwise it may adversely affect product volume. The heat treatment is required to denature the globulins which are present, otherwise they can interfere with the stability of the gas bubbles in the batter and cause loss of product volume.

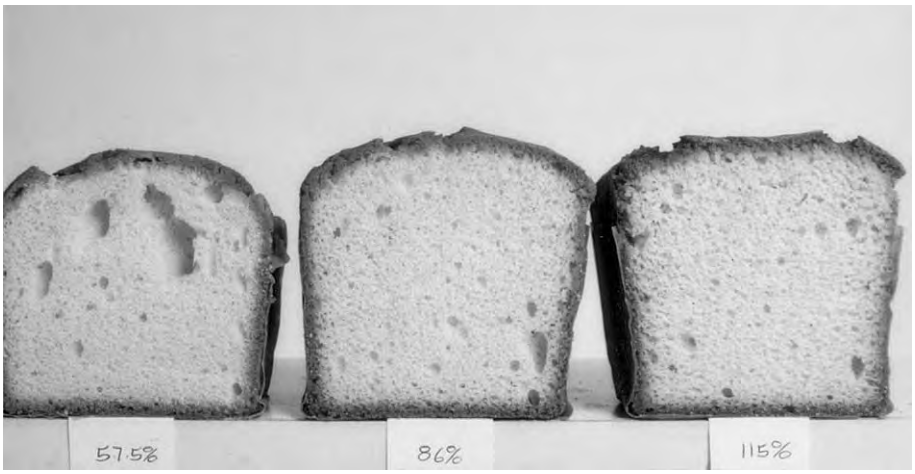

Figure 1 (see color plate 14) Cakes with (left) 57.5, (middle) 86 and (right) 115% sucrose based on flour weight.

Water

Water may be added to dissolve the recipe sugars, especially if liquid milk products are not used.

Fats and Oils

The addition of solid fat to a cake formulation is used to provide tender eating qualities and restrict the potential for gluten formation. The solid fat also aids the incorporation and stabilization of air bubbles in the batter. Polarization microscopy shows that, when subjected to agitation, the fat crystals align themselves around the air bubbles, thereby entrapping and stabilizing them. This is possible because the oil fraction in the fat allows the fat crystals to move and at the same time adhere to one another like links in a chain. It is apparent that in such a mechanism the ratio of liquid to crystalline fat and the size of the individual crystals are very important. If oil is used in the batter, the bubble stabilization effects will be absent and the cake will lack volume.

The crystalline form of the fat is important in promoting batter aeration. The three fat polymorphs of concern are the alpha, beta-prime and beta. Air incorporation in cake batters is greatest with fats in the beta-prime form, less with the alpha form, and least with the beta form.

Emulsifiers

Emulsifiers are used to aid the incorporation and stabilization of air bubbles in the batter, especially in the presence of fats or oils. The most commonly used emulsifiers for this purpose are glycerol monostearate (GMS) and polyglycerol esters, with the former being the more effective of the two on a weight-for-weight basis. Both emulsifiers are commonly used in a paste form, i.e., dispersed in water with other ingredients which promote gel stability. Emulsifiers like GMS may exist in a number of forms when dispersed in water and it is important it is in the active alpha-gel form when used for cake making.

Without GMS the egg protein will largely stabilize the air bubbles and the sponge will have a reasonable volume, but often with an area of coarse open-cell structure in the crumb. The addition of a small level of GMS somewhat unexpectedly produces a worse cake, with a dip in the top surface. It is apparent from this result that the GMS and the egg proteins do not combine their stabilizing power. At low levels the GMS is able to stabilize the bubbles at low temperatures, but it cannot continue to do so as the temperature rises and the bubbles expand during baking. Further increases in the level of added GMS eventually provide a stable batter and improved cake quality.

Baking Powders

A significant proportion of the final baked cake is air. This is a mixture of gases, partly air incorporated during mixing, and partly carbon dioxide generated from the action of sodium bicarbonate with a suitable food acid. The production of carbon dioxide gas and its release when the batter first enters the oven is an important part of the cake expansion mechanism. Without this extra release of gas the batter can only expand from the temperature effect on the air trapped in the batter and the cake will remain small in volume. Providing an extra release of gas produces a larger cake specific volume and gives it a more tender eating quality. The relative importance of mechanical and chemical aeration is shown in Figures 2 and 3.

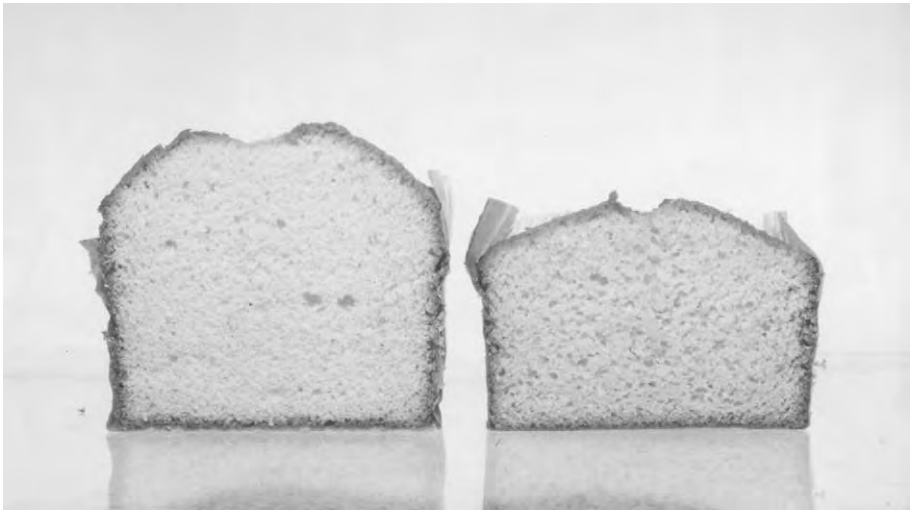

Figure 2 (see color plate 15) Cake (left) with and (right) without chemical aeration.

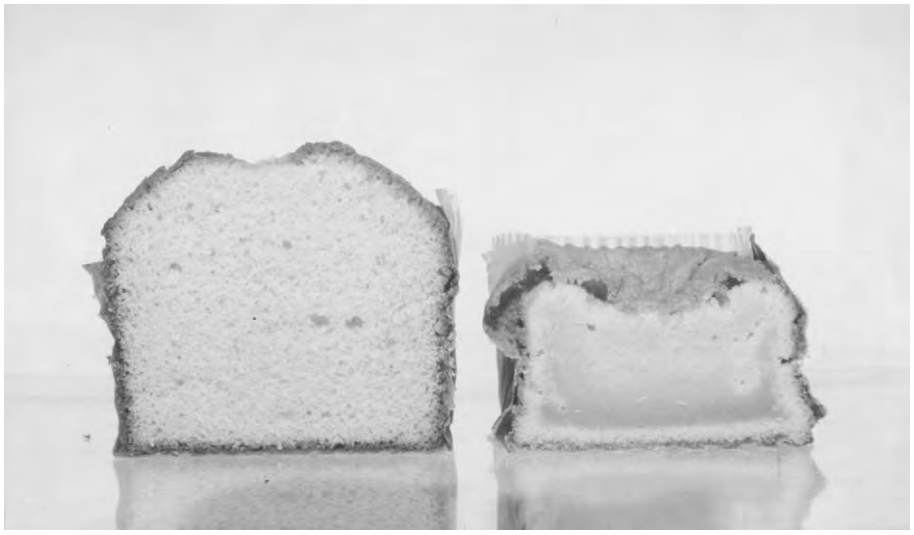

Figure 3 (see color plate 16) Cake (left) with and (right) without mechanical aeration.

The action of the baking acid on the sodium bicarbonate normally proceeds to completion, i.e., to release all of the carbon dioxide gas from the mixture. The flavor of the cake is influenced by the residual salt from the acid–base reaction. The rate at which carbon dioxide is released from the baking powder reaction is controlled by using acids of different types. The most commonly used baking acids and their rates of addition to achieve neutrality with one part of sodium bicarbonate are:

- Tartaric acid, 0.89, fast reaction
- Potassium hydrogen tartrate (cream of tartar), 2.25, fast reaction
- Acid calcium phosphate (monocalcium phosphate), 1.25, moderately fast reaction
- Glucono-delta-lactone, 2.10, moderately fast reaction
- Sodium acid pyrophosphate, 1.33, slow reaction
- Sodium aluminum phosphate, 1.00, slow reaction

Other Dry Ingredients

- Whey solids may be used for color and flavor. The powder should have been heat-treated before use otherwise it may adversely affect product volume.
- Cocoa powder is used in the production of chocolate cakes. It can come in several different forms.
- Salt may be used for flavor and improving product mold-free shelf-life.
- Humectants such as glycerol and sorbitol may be used to extend product mold-free shelf-life.

- Dried vine fruits are used in the production of fruited cake varieties.
- Candied orange and lemon peel and glacé cherries may be used in the production of fruited and speciality cakes.
- Dried nuts may be added to fruited and speciality cakes.

The Importance of Recipe Balance

The final quality of cakes is strongly influenced by the proportions of the ingredients that are used in the recipe. Since the wheat flour provides the main structure-forming component – the starch – it is common practice to construct the recipe based on the flour weight and express the mass of other ingredients as 'percentage flour weight'. This method of determining ingredient levels has the advantage that it is possible to deduce a series of 'rules' by which any cake recipe can be constructed and the causes of quality losses rectified.

Typical of the rules which are applied to cake making include the following:

- The level of sugar will lie between 95 and 135% flour weight for high-ratio cakes and 65 and 95% for low-ratio cakes.
- The level of water should be equal to and not exceed 1.25 times the sugar weight.
- The level of fat should not exceed the level of total egg used (i.e., solids and water combined).
- The level of fat should not exceed the level of sugar.

Imbalances in ingredient ratios will manifest as quality losses in the baked product. Many more specialized rules exist whereby the level of other functional ingredients, e.g., baking powder, may be determined. The range of acceptable ingredients ratios varies according to the type, size, and shape of cake being made.

Cake Shelf-Life

Water Activity

Cakes are commonly referred to as intermediate-moisture food products because the moisture content of typical cake products (approximately 18–28%) is high enough to confer soft and tender eating properties. At such moisture contents cakes are susceptible to microbial spoilage within a few days or weeks. However, it is the water activity of a cake product that controls the products microbial shelf-life rather than the moisture content alone. Examples of water activity in cakes are given in Table 2.

The formulation of the cake recipe is the main factor which affects its water activity and thus its microbial shelf-life. Particularly important in controlling water activity is the ratio of recipe water to the soluble materials. The latter will include salt and sugars. In addition, some other ingredients which do not go into solution but which have the ability to hold water and reduce its availability for microbial activity may be present. An example of the latter is flour which hydrates during mixing because of the presence of proteins and damaged starch. On baking, the flour proteins coagulate and the starch gelatinizes.

When soluble materials like salt and sugar dissolve in water they both reduce the vapor pressure of the solution. The potential effect of a particular ingredient depends on its individual chemistry. For example, salt is ionic and has a great affinity for water. Sucrose and other sugars (e.g., glucose) do not possess that same affinity of water and so for a given weight is less effective than salt at locking up water. The sucrose equivalent of an ingredient is often used to identify its likely affect on cake water activity. Some typical sucrose equivalents for cake ingredients are given in Table 3. The sucrose equivalent is based on a method which assigns a value of 1 to sucrose and compares all ingredient effects on an equal-weight basis. Thus, in Table 3 the value for salt is 11, which indicates that, weight for weight, salt is 11 times more effective at lowering product water activity than sucrose. Other materials worthy of note are glycerol four times more effective, sorbitol twice, and dextrose monhydrate 1.4. Many ingredients are much less effective than sucrose and therefore have values of less than one.

Table 2 Examples of water activity in cakes

Cake product	*Water activity range*
High-moisture cakes, such as carrot cake	0.9–0.95
Yellow, white, chocolate, and layer cakes	0.8–0.89
Fruited cakes	0.7–0.79
Celebration cakes	0.6–0.69

Table 3 Sucrose equivalent of some bakery ingredients

Ingredient	*Sucrose equivalent*
Salt	11.0
Glycerol	4.0
Sorbitol solids	2.0
Dextrose monohydrate	1.4
Sucrose	1.0
Flour	0.2

Spoilage Inhibitors

Where the microbial shelf-life of cake products is insufficient it is common practice to add a suitable mold inhibitor to add to the effect of product water

Table 4 Approximate increases in mold-free shelf-life with 1340 p.p.m. potassium sorbate in cake (pH 6.5)

Water activity	*Extra days at 21 °C*
0.92	2.5
0.86	8.0
0.80	40.0

activity. Commonly used mold inhibitors are sorbic acid and its salts (e.g., potassium sorbate). Levels of addition will vary according to the shelf-life requirements, with upper limits being controlled by legislation. In practice, high levels of addition will lead to changes in product flavor which may make the product unacceptable to consumers.

The effect of mold inhibitors in cake making is enhanced by the effect of water activity. An example of this relationship is given for potassium sorbate in Table 4.

See also: **Cakes**: Methods of Manufacture; Chemistry of Baking; **Eggs**: Use in the Food Industry; **Emulsifiers**: Uses in Processed Foods; **Flour**: Dietary Importance; **Leavening Agents**; **Milk**: Dietary Importance; **Powdered Milk**: Characteristics of Milk Powders; **Sweeteners**: Intensive; Others

Further Reading

Bent AJ (ed.) (1997) *Technology of Cakemaking*, 6th edn. London: Blackie Academic & Professional.

Cauvain SP and Cyster J (1996) Sponge cake technology. *CCFRA Review No. 2*.

Cauvain SP and Young LS (2000) *Bakery Food Manufacture and Quality: Water Control and Effects*. Oxford, UK: Blackwell Science.

Cauvain SP and Young LS (2001) *Baking Problems Solved*. Cambridge, UK: Woodhead Publishing.

Cauvain SP, Hodge DG, Muir DD and Dodds NJ (1976) *Improvements in and Relating to Treatment of Grain*. British patent no. 1 444 173. London, UK: HSMO.

Cauvain SP, Gough BM and Whitehouse ME (1976) The role of starch in baked goods: part 2. The influence of purification procedure on the surface properties of the granule. *Die Starke* 41: 461–467.

Doe CAF and Russo JVB (1968) *Flour Treatment Process*. British patent no. 1 110 711. London, UK, HSMO.

Gough BM, Whitehouse ME and Greenwood CT (1978) The role and function of chlorine in the preparation of high-ratio cake flour. *CRC Critical Review in Food Science and Nutrition* 10: 91–113.

Grover DW (1947) The keeping properties of confectionery as influenced by its vapour pressure, *Journal of the Society of Chemistry and Industry* 66: 201–205.

Kent NL and Evers AD (1994) *Kent's Technology of Cereals*, 4th edn. Oxford, UK: Elsevier Science.

Sahi SS (1999) Influence of aeration and emulsifiers on cake batter rheology and textural properties of cakes. In: Campbell GM, Webb C, Pandiella SS and Niranjan K (eds) *Bubbles in Food*, St Paul, USA: AACC. pp. 263–271.

Spies RD and Hoseney RC (1982) Effects of sugar on starch gelatinisation. *Cereal Chemistry* 59: 128–131.

Street CA (1991) *Flour Confectionery Manufacture*. London, UK: Blackie Academic & Professional.

Young LS, Davies PR and Cauvain SP (1998) Cakes – getting the balance right. In: Milne R, Macintosh A and Bramer M (eds) *Proceedings of ES98, the Eighteenth Annual International Conference of the British Computer Society Specialist Group on Expert Systems*, pp. 42–55. London, UK: Springer.

Methods of Manufacture

S P Cauvain, Campden and Chorleywood Food Research Association, Chipping Campden, UK

Aims of Mixing

As discussed above, gluten formation is undesirable in cakes and is limited, in part, by the formulation, most notably high levels of sugar, liquid, and fat. Gluten formation may also be limited by the methods used to mix cake batters and the avoidance of any processing steps that may induce gluten formation. As a broad generalization, cake batters can be considered as fat, or oil, in water emulsion systems. The aqueous phase contains the dissolved sugar and suspended flour particles. In cake-batter systems, air bubbles are incorporated and held in the fat phase rather than the aqueous phase. However, as the batter warms during baking, the air bubbles transfer from the fat to the aqueous phase foam and expand. Later in the baking process, the foam sets to yield the familiar cake structure.

The aims of cake-batter mixing can be summarized as follows:

- to uniformly disperse the ingredients;
- to incorporate and stabilize the necessary air bubble nuclei; and
- to minimize gluten formation.

The main manufacturing methods for cakes are mostly based on the mixing process that is used. There are two main types:

- *Multistage*, in which various ingredients are mixed together as subcomponents before blending into the final batter. The aims of multistage mixing methods are to minimize gluten formation and to aid with the incorporation and stabilization of air

bubbles in batter. Such methods are traditional and probably derive from days when ingredient qualities were less reliable than today.

- *All-in*, in which all of the ingredients are placed in the mixing bowl at the same time and beaten together. All-in or single-stage mixing relies on standardized ingredients being available.

Mixing Methods

Sugar–Batter Processes

The essential features of this process are:

1. Creaming of the fat and sugar together. The sugar crystals aid the incorporation of air bubbles into the mixture, and the density falls as the mixing continues. This will be seen as a progressive 'whitening' of the mixture. The ideal creaming time depends on the type of fat being used, with higher-melting-point fats requiring longer creaming times. A medium speed, about 150 rpm, is typically used.
2. The addition of egg to the fat–sugar mixture in four to six individual portions with some blending between the addition of each portion. The aim should be the uniform dispersion of the egg without de-aeration of the mixture. It is important that the egg should not be at a lower temperature than the mixture, otherwise separation of the fat from the mixture may occur – this is known in the baking industry as 'curdling.' Additions of egg are usually made at low speed, typically around 50 rpm, with subsequent blending on a medium speed for short lengths of time.
3. The addition of the flour and other 'dry' ingredients with subsequent blending. The mixing time at this stage will be short and only sufficient to blend in the flour. A low speed is used.
4. The addition of final liquids, e.g., milk and water, blended in at low speed.
5. The addition of dried fruit, if required.

Flour–Batter Process

1. The fat and a portion of the flour (typically three-quarters of the flour mass) are creamed together at low speed for about 10 min, during which time, air is incorporated into the mixture, and the density falls.
2. The egg and sugar are whisked separately at high speed.
3. The fat–flour and egg–sugar are blended at low speed.
4. Any water or other liquids are blended into the mixture at low speed.
5. The remaining flour and other dry ingredients (e.g., skimmed milk powder, salt, and baking powder) are blended into the mixture at low speed.

Egg-separated Process

1. The fat, a portion of the flour (typically three-quarters of the flour mass), salt, and other dry ingredients are creamed together at low speed.
2. The egg is separated into its white and yolk components.
3. The egg white and sugar are whisked together at high speed to form an aerated foam.
4. The egg yolk and additional water or milk are blended into the egg white foam before being blended with the fat cream.
5. The remaining flour is blended into the mixture at low speed.

Boiled Process

1. The egg and sugar are whisked at high speed to form a foam.
2. The fat is heated.
3. The flour and other dry ingredients are blended into the hot fat.
4. The egg foam and the fat-flour blend are mixed together.

The final batter temperature is around 35–40 °C. The method is not usually considered suitable for recipes that contain baking powder because the high temperature causes premature release of carbon dioxide.

All-in Process

Using this method, all of the ingredients are placed into the mixing bowl and blended together at low and medium speed. The all-in method is particularly suitable for use with recipes that contain an emulsifier.

Sponge Cakes

Sponge cakes may be made with a simple formulation of flour, whole eggs, and sugar. The eggs and sugar are whisked together at high speed, typically 200–300 rpm. During this whisking process, large numbers of minute air bubbles are incorporated into the batter. Without some form of stablization, the air bubbles occluded during mixing would rapidly coalesce, rise to the surface of the batter, and be lost. In this sponge system, the surface-active proteins of the egg white and the lipoproteins migrate to the interface with the air occluded during beating, and form a protective film around the nascent gas bubbles and prevent them coalescing. When a stable foam has

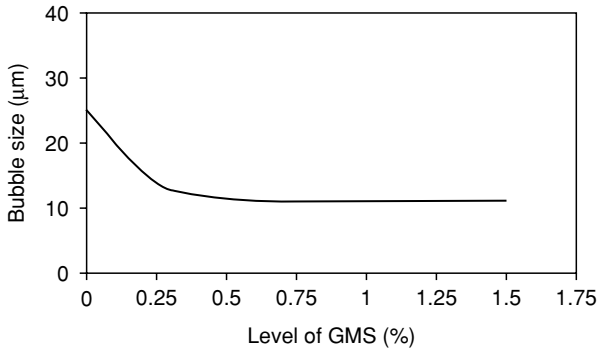

Figure 1 Effect of glycerol monostearate on gas-bubble size in sponge cake batter.

been formed, the flour is added carefully, trying not to de-aerate the foam. After depositing, the bubbles expand as the batter temperature rises in the oven, and eventually, just as the mass of the batter begins to set, the bubbles burst into one another to form the porous cake structure.

The addition of a small level of fat or oil to the recipe disrupts the interfacial films and prevents the egg proteins from stablizing the air bubbles. The result is a less well-aerated batter and restricted cake volume. To overcome this potential loss of volume, the fat can be melted and carefully added at the end of the mixing process once a stable foam had been formed. More commonly fat-enriched sponge cakes are made with the addition of glycerol monostearate (GMS), which displaces the egg proteins at the gas-bubble interface and provides the stabilizing role.

GMS controls not only the size of the air bubbles (see Figure 1) but also the quantity of air mixed into the batter. High levels of GMS allow large quantities of air to be incorporated into the batter in the cold, but unless there is a sufficient amount to stabilize the bubbles during baking, the sponge inevitably will collapse. Measurement of bubble sizes in cake batters at ambient temperatures shows that this property is governed, in part at least, by the amount of GMS present. However, an excess of GMS above an optimum restricts bubble coalescence during baking so that the foam remains intact and tends to shrink back on cooling. This gives the cakes an unacceptable wrinkled surface.

Delayed Soda Process

In some more traditional sponge making processes using fast-acting baking acids, the addition of the sodium bicarbonate in the recipe may be delayed to the end of the mixing cycle. This delayed addition of sodium bicarbonate insures that losses of carbon dioxide before baking are minimized.

Mixing Equipment

Planetary Mixers

The most common form of cake batter mixer is the vertical form that comprises a removable bowl in which the mixing tool describes a planetary motion. Typically, the mixing tools are removable, with the most common forms being a flat beater or a wire whisk. The former is commonly used with batters containing higher fat levels in the recipe, whereas the latter is mostly used with sponge batters. Planetary mixers can operate over a range of mixing speeds (usually three).

Horizontal Mixers

The mixing chamber of this type of mixer is fixed within the mixer frame but can be rotated through a limited arc to allow for the discharge of the cake batter. The mixing tools revolve around a horizontal axis within the mixing chamber. The chamber comprises two U-shaped sections with a single mixing blade operating in each section. Two mixing speeds are usually possible with this type of mixer. Ingredients are fed in from the top of the mixer.

Pressure Whisks

In this type of machine, the mixing is based on a rotating whisk, vertically mounted, and moving within a sealed mixing chamber. Since the chamber is sealed, it is possible to vary the pressure of the air within the mixer headspace during the mixing cycle. This increases the quantity of air that can be incorporated into the batter and so reduces batter density. Pressure whisks are most commonly used in the production of low-density sponge batters. The pressure whisk may be operated over a range of mixing speeds. Discharge of the batter from the mixing chamber may be under pressure direct to a depositor.

Continuous Mixers

Continuous mixers are most commonly used in larger-scale cake-batter production. The mixing chamber is small compared with other cake mixing machines, because the batter residence time for mixing is in the order of 10–20 s. The mixing tool comprises a form of round-sectioned, double-sided rotor covered in small, rectangular pins able to rotate in a round-shaped mixing chamber with similar-shaped pins projecting from the inner walls. The effect is to create a series of very narrow gaps through which the batter can pass, and the action provides an intimate mix. At the same time as being thoroughly mixed, the batter is aerated by the injection of air under pressure. The high shear of the mixing action

aids the dispersion of the air and the formation of smaller and more uniformly sized gas bubbles in the batter.

Mixing and batter aeration in continuous mixers are far more controlled than with any other type of cake mixer. All of the ingredients must be blended to form a slurry on a separate mixer before being fed to the holding tank of the continuous mixer. Aeration of the slurry is avoided. The residence time of the batter in the mixing head is a function of the throughput rate. The speed of the rotor may be varied. Levels of aeration can be varied by adjusting the air flow rate into the mixing chamber.

Continuous mixers may be used with most forms of cake batter, except fruited varieties, and also for the aeration of creams, meringue, and some forms of sugar confectionery.

High-speed, Vertical Mixers

High-speed, vertical mixers are most commonly used in the production of bread and other fermented doughs. However, they may be used in the production of pastes and cake batters. Mixing times are very short, typically less than 3 min, and the levels of aeration are lower than those achieved with most other mixer types.

Batter Handling and Depositing

The low viscosity and fluidity of most cake batters make them relatively easy to handle. However, batter handling should be kept to a minimum, because, while in motion, there is the potential for de-aeration of the batter, especially the loss of carbon dioxide gas from the baking powder reaction.

Individual portions of many cake batters are deposited into some form of container for baking. Typically, the baking container will be lined with greaseproof or silicone paper to provide easy release of the product after baking. Product shapes can vary, though round or rectangular are the most common. Some cake batter types, e.g., Swiss roll and sponge drops, may be deposited directly on to the oven band or sheet for baking.

Cake Baking

Cake batters may be baked in most oven types, e.g., deck, rack, and continuous. The baking temperatures are lower than used for bread baking and vary according to the type and size of cake. In many ovens, heat input will be greater at the bottom of the product than at the top. High heat inputs directed on to the top of the baking product commonly lead to problems with product shape and surface color.

Secondary Processing

While many cake types may be eaten in their baked form, a large number of cake products are combined with other food products to present consumers with a wide variety of products. Examples of other food products that may be used as a filling or topping with cakes include, dairy and nondairy creams, chocolate, icing and fondants, jams, and jellies.

Packaging

Some form of packing is required to hold cakes, protect them from microbial contamination, and prevent moisture losses that would have an adverse effect on the cake-eating qualities. Product packs take many forms, but most often, the material used will be impermeable to moisture to prevent drying out of the product. Wrapping composite cake products in moisture impermeable film may lead to problems of moisture migration between components, unless steps are taken to adjust the water activities of individual components.

Storage

The shelf-life of cake products is directly related to the storage conditions: the warmer the storage temperature, the shorter the microbial shelf-life and the greater the problem with moisture migration between the cake and other components (e.g., cream). Low-temperature storage is particularly important for dairy cream-filled cakes because of their susceptibility to bacterial growth. Refrigerated storage does not accelerate cake staling (unlike the situation with bread). Cake staling proceeds fastest with cake at temperatures around 25 °C.

See also: **Cakes**: Chemistry of Baking; **Celiac (Coeliac) Disease**; **Emulsifiers**: Uses in Processed Foods

Further Reading

Bent AJ (ed.) (1997) *Technology of Cakemaking*, 6th edn. London: Blackie Academic & Professional.

Cauvain SP and Cyster J (1996) Sponge cake technology. In: *CCFRA Review No. 2*, Chipping Campden, UK: CCFRA.

Cauvain SP and Young LS (2000) *Bakery Food Manufacture & Quality: Water Control and Effects*. Oxford: Blackwell Science.

Street CA (1991) *Flour Confectionery Manufacture*. London: Blackie Academic & Professional.

Chemistry of Baking

J L DesRochers, K D Seitz and C E Walker, Kansas State University, Manhattan, KS, USA

Principal Ingredients in Cake Batter

Flour is the most important ingredient in cakes, functioning primarily to establish crumb structure. Cake flour is milled from soft white or red wheat with low protein and ash (mineral) levels and has a fine particle size. In the USA, cake flour is commonly treated with chlorine gas, causing hydrolytic depolymerization of the starch molecules, which increases the water-holding capacity of the flour. During mixing, proper gluten development is critical, to insure a fine foam structure without excessive toughening. Protein level and type and mixing procedure are key factors for producing cakes with desirable crumb texture. The optimal specifications for a typical cake flour are: protein content $8.5 \pm 0.5\%$, ash content $0.36 \pm 0.04\%$, pH 4.7 ± 0.2, and average particle size $10 \pm 0.5\,\mu m$. (*See* **Flour**: Roller Milling Operations; Analysis of Wheat Flours; Dietary Importance.)

Shortening performs three basic roles in cakes. First, it aids in aeration or leavening of the batter and finished cake by entrapping air during the creaming process. These minute air cells provide the nucleus for bubble expansion via steam and carbon dioxide during baking. Second, shortening coats the protein and starch particles, preventing hydration and formation of a continuous gluten–starch network. Third, it emulsifies liquids in the batter, which increases the moisture of the crumb. The last two functions contribute to a soft, tender crumb texture. (*See* **Fats**: Uses in the Food Industry.)

Eggs perform a variety of functions in cake production, providing structure, volume, tenderness, and nutritional quality to the product. They act as a binding agent due to their high protein content and ability to form a complex network with gluten. They contribute to cake volume and structure by their ability to be whipped into a relatively stable foam. Upon heating, the proteins are denatured, thus setting and stabilizing the crumb structure. The yolk portion of the egg imparts an emulsifying and tenderizing effect because of the high lipid and lecithin content. Eggs help stabilize the emulsion, retain gas generated by the leaveners, and prevent air cell coalescence in the batter, resulting in a uniform crumb grain and desirable texture. Eggs also add a mild but distinctive flavor and color (from the yolk) and enrich cake's nutritive value. (*See* **Eggs**: Dietary Importance.)

Chemical leavening agents are added to aerate the batter and produce a light, tender, porous product. The porosity of the batter directly translates into good volume, uniform cell structure, bright crumb color, tender texture, and overall eating quality of the finished cake. In lieu of yeast, chemically leavened cakes utilize sodium bicarbonate (baking soda) plus an acidic agent to generate carbon dioxide (CO_2) gas when combined in the presence of water. (*See* **Leavening Agents**.)

Baking powders are a combination of sodium bicarbonate and the salt of a weak acid (see Table 1 for examples). Baking powder should yield no less than 12% carbon dioxide based on the weight of the product. Starch or flour may be added as a diluent to standardize the baking powder's strength.

Baking powders are classified by their reaction rates as fast, slow, or double-acting. The fast-acting types release most of their available carbon dioxide within the first few moments of contact with liquid. If these batters or doughs are not processed very quickly, the gas will be lost before the structure is set and the volume will decrease significantly.

Slow-acting baking powders do not react at low temperatures and therefore require oven heat to release gas. Double-acting types react partially at low temperature, but need higher temperatures to complete the reaction. The double-reacting powders are most commonly used in commercial cake production.

Table 1 Acidic agents used in leavening

Acid	*Formula*	*Neutralization value*
Monocalcium phosphate (MCP)	$Ca(H_2PO_4)_2 \cdot H_2O$	80
Anhydrous monocalcium phosphate (AMCP)	$Ca(H_2PO_4)_2$	83.5
Sodium acid pyrophosphate (SAPP)	$Na_2H_2P_2O_7$	72
Sodium aluminum phosphate (SALP)	$NaH_{14}Al_3(PO_4)_8 \cdot 4H_2O$	100
Monopotassium tartrate (cream of tartar)	$KHC_4H_4O_6$	45
Sodium aluminum sulfate (SAS)	$Al_2(SO_4)_3 \cdot Na_2SO_4$	100
Dicalcium phosphate dihydrate (DCP)	$CaHPO_4 \cdot 2H_2O$	33
Glucono-δ-lactone (GDL)	$C_5H_{10}O_6$	50

The neutralization value (NV) is used to give the proper balance of the alkaline and acidic agents. The NV indicates the amount of sodium bicarbonate required to release all the available carbon dioxide from 100 units of the acid leavener (Table 1). NVs enable the formulator to achieve the desired batter pH, which is important to crumb structure and grain, color, and flavor.

Sugar acts primarily as a sweetener in cakes and aids in air incorporation during creaming. In the mid-1900s, emulsified shortenings enabled bakers to produce richer cakes with higher sugar and liquid levels. These high-ratio cakes generally contain 120% sugar based on flour weight, and have an extended shelf-life and tender texture. The type and form of sugar used are also important; liquid sugar or syrup acts as a moistener whereas crystalline or granular sugar functions as a drier. Granulation affects how quickly the sugar dissolves and generally increases cake volume as granulation becomes finer. Invert or reducing sugars, e.g., high-fructose corn syrup, honey, or molasses, can affect the crust and crumb color, and texture.

Milk, either in fluid or dry powdered form, is a source of protein and lactose, which aid in crust development and browning reactions. Milk also stabilizes the foam and contributes to cake structure.

Finally, salt is added as a flavor enhancer.

Formula balance of the major cake ingredients – flour, shortening, sugar and eggs – is important to produce a cake with good volume, grain, texture, and eating properties. Some general guidelines for high-ratio cakes are as follows:

- The weight of the sugar should be greater than the flour weight
- The weight of the eggs should exceed the weight of the shortening
- The weight of the total liquids (milk, eggs, and/or water) should be slightly greater than the sugar weight

Some exceptions to these rules include foam-type cakes, i.e., angel food or sponge which contains little or no shortening. In these cakes, whipped egg whites, sugar, and flour are the primary ingredients, along with some minor ingredients. In a typical angel food cake formulation (Table 2), the weight of the sugar is usually equal to the weight of the egg whites, and the flour is approximately one-third the weight of the sugar. Figure 1 shows the relationship between basic ingredient composition and cake type.

Sponge cakes use whole egg instead of only egg whites. To minimize the toughening effect of the eggs, an equal or slightly higher amount of sugar is added. The liquid and flour levels are balanced around the eggs and sugar. In general, the total liquids should be 25% greater than the weight of the sugar.

Table 2 Angel food cake formula

Ingredient	*Baker's %*[a]
Flour	100
Sugar	280
Egg white	280
Salt	4
Cream of tartar	4
Vanilla	5

[a]Baker's % based on 100 parts flour.

Flour level should be less than the weight of either the sugar or eggs. Combined, the weight of eggs plus flour should exceed the weight of sugar plus nonegg liquids (milk or water). A typical sponge cake formula is listed in Table 3.

Pound cake is one of the oldest types of cake, deriving its name from the original recipe which called for 1-lb (0.45-kg) increments each of flour, butter, eggs, and sugar. The expense of butter and eggs prompted commercial bakers to modify the formula, producing a lighter cake with improved volume and eating quality (Table 4).

The modified pound cake formula led to the basic yellow or white layer cake (Table 5). Unlike yellow layer cake, no yolks are used for white layer cake. These layer cakes adapt readily to many formula variations by the addition of fruits, nuts, spices, cocoa, etc. In chocolate layer cakes, the sugar level appears higher, which is due to a reduction in the flour level when cocoa is added. Functionally, low-fat cocoa acts much like flour. High-ratio cakes evolved from layer cakes through the use of emulsified shortenings and may contain as much as 140% sugar (based on flour weight).

Chemical and Physical Changes during Mixing

Mixing plays an important role in the production of quality of batter-type cakes. The objectives of mixing are to disperse the various ingredients uniformly, and to incorporate air into the batter while minimizing gluten development. There are four different basic methods of mixing: creaming, blending, single-stage, and foam-type methods.

In the creaming, or conventional method, fat and sugar are mixed at low to medium speeds until thoroughly blended and aerated. Large volumes of air are incorporated into the fat phase in the form of small air cells and the sugar crystals are encased in the shortening. Next the eggs are added with continuous beating until the mixture is fluffy and well aerated. Last, the flour and milk are added. The main advantages of this method are: (1) the large number of small

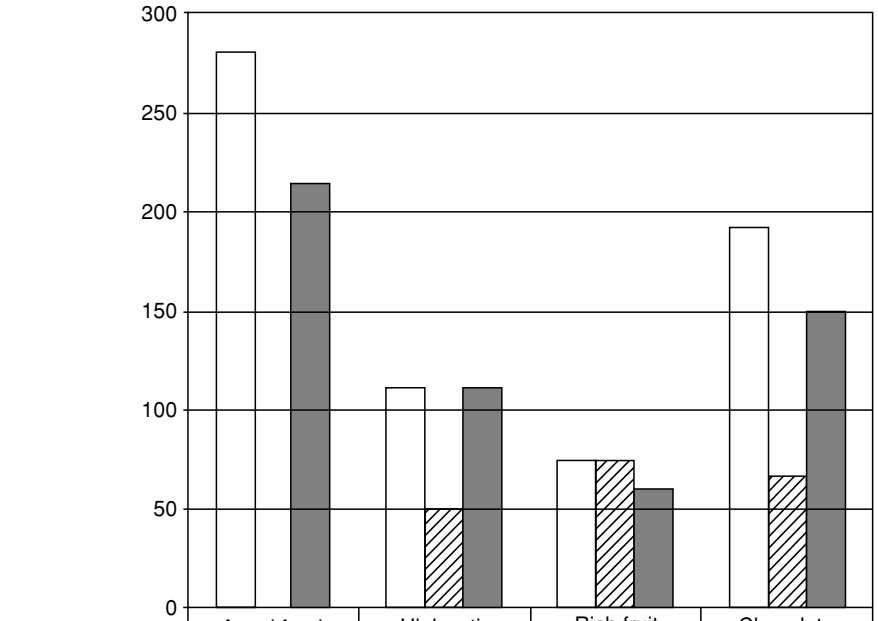

	Angel food	High-ratio	Rich fruit	Chocolate
% Sugar	280	110	73	190
% Shortening	0	50	73	65
% Liquid	214	110	59	150

Figure 1 Relationship between cake type and basic composition based on 100 parts flour.

Table 3 Sponge cake formula

Ingredient	*Baker's %*[a]
Flour	100
Sugar	95
Corn syrup	12
Eggs	105
Water	12
Vanilla	3
Baking powder	1.5
Salt	0.75

[a]Baker's % based on 100 parts flour.

Table 4 Commercial pound cake formula

Ingredient	*Baker's %*[a]
Flour	100
Sugar	100
Shortening	50
Whole eggs	50
Milk	50
Vanilla	2
Salt	1.5

[a]Baker's % based on 100 parts flour.

Table 5 Ordinary yellow cake formula

Ingredient	*Baker's %*[a]
Flour	100
Sugar	85[b]
Shortening	45
Whole eggs	50
Milk	50
Baking powder	2.5
Salt	2
Flavor	1.5

[a]Baker's % based on 100 parts flour.
[b]In the USA, sugar would run to 120 or more.

air cells in the batter; and (2) gluten development is limited due to delayed hydration and solubilization of the fat-coated flour and sugar particles.

The blending method consists of two separate mixing steps. The shortening and flour are beaten until fluffy in one bowl and the eggs and sugar are whipped separately in a second bowl. These two mixtures are blended together followed by slow addition of the milk. This method produces a cake with a very fine grain and uniform texture. Compared with the creaming method, aeration is lessened, thus reducing cake volume, but higher sugar and liquid levels are possible.

The single-stage mixing method was devised to reduce the number of steps and shorten the mixing time requirements. All the major ingredients are mixed together at once. Although the procedure is simple, the texture and stability are sacrificed somewhat. This method is generally used for premixed box cakes, but not by the wholesale trade.

Unlike shortening-based cakes, foam cake mixing depends on air incorporation in an egg–protein matrix for volume and structure development. Egg whites plus sugar are whipped into a stiff foam, into which the flour and other dry ingredients are gently folded. In angel food cake batters, the egg whites must be fat-free to achieve maximum volume; however, in sponge cakes where whole eggs are used, a lower-volume, emulsified foam is formed.

The batter specific gravity is a measure of how much air has been incorporated into the batter during mixing. It is defined as the ratio of the weight of a known volume of batter to the weight of the same volume of water, at a given temperature. Batter specific gravity is directly related to the volume, texture, and grain of the finished cake. In general, the lower the specific gravity, the higher the finished cake volume, but optimal ranges for different types of cake batters have been established.

Chemical and Physical Changes during Baking

Air bubbles, creamed in the fat, are released into the aqueous phase as the fat is melted at approximately 40 °C. Carbon dioxide is generated by the baking powder and collects in these air bubbles. As the batter heats, the batter begins to flow due to natural convection currents, as the batter temperature next to the sides and bottom of the pan increases first and that in the center heats last.

The batter viscosity decreases initially upon heating as the fat melts and before the starch gelatinizes. Between 60 and 70 °C, the starch granules absorb several times their weight in water, increasing the batter viscosity considerably. Major swelling of the gelatinized starch granules insures that the cake structure will not collapse. The amount of sugar and some emulsifiers in the formula control the temperature at which the starch gelatinizes. The cake normally sets into a solid system well below the boiling point of water (100 °C).

As the batter temperature reaches about 80 °C, the air bubbles enlarge rapidly, causing the cake to rise. The liberation of carbon dioxide, the expansion of air cells, and the formation of steam all contribute to the leavening effect. Heat causes the gas to expand and the pressure inside the gas cells to increase. Resistance to expansion results from protein coagulation and starch gelatinization. Timing is critical for the protein film to enlarge with the expanding gases, prior to protein denaturation and starch gelatinization, which set the structure. Emulsifiers, whether occurring naturally in eggs or added, improve the elasticity of the protein film surrounding the gas bubbles, enabling them to increase without rupturing.

Moisture evaporates from the cake surface during baking, keeping it cool. However, as most of the water is removed, the surface gets hot enough to brown. Lower baking temperatures should be used for richer (high sugar and fat content) cakes, as the sugars can cause excessive browning of the crust before the interior is set. The baking time for all cakes should be kept as short as possible to avoid too much color development and formation of a thick crust.

A summary of common cake faults and possible ingredient, mixing, or baking-related causes is given in Table 6.

Role of Additives

One of the main additives in cake production is added to the flour itself. In the USA, cake flour is made by

Table 6 Troubleshooting guide for cakes

Possible causes	*Common cake faults*						
	Low volume	*Toughness*	*Lacking resilience*	*Poor crust appearance*	*Irregular grain*	*Peaked center*	*Undesirable color*
Improper chemical leavening	×				×		×
Low batter viscosity	×				×	×	
Excessive oven temperature	×	×		×	×	×	×
Insufficient oven temperature					×		×
Egg/milk protein level incorrect		×	×				
Incorrect sugar level or type		×		×			×
Overmixed batter		×		×	×		
Undermixed batter			×		×		

adding chlorine gas to soft wheat flour at a rate of 0.5–2.5 oz per 100 lb (0.3–1.5 g kg^{-1}) of flour. This lowers the pH and improves overall baking performance by increasing the volume, and improving grain, texture, and symmetry. The optimal pH range is between 4.5 and 4.8. The mechanism of chlorine on flour is not completely understood: various researchers have shown that it affects the gluten, starch, and/or lipid components of wheat flour.

Overchlorinated cake flour will cause the batter to set too quickly around the sides of the pan before full expansion has been reached. The center continues to rise, and the result is a cake with a strong peak. If the flour is underchlorinated, the structure sets too slowly, allowing the leavening gases to escape, and the center of the cake collapses upon cooling.

Emulsifiers promote air incorporation in the form of fine bubbles and disperse the shortening into small-sized particles. Emulsifiers' unique behavior is due to their ability to bridge the inseparable oil and water phases at the interface. When their concentration exceeds the solubility limit, they form an interfacial membrane whose hydrophilic portions extend into the aqueous phase. The membrane surrounds the dispersed oil and prevents the emulsion from breaking.

Hydrogenated shortenings typically contain 3% emulsifiers, with glycerol mono- and distearate being the most common, although many others, including blends, are also used. (*See* **Emulsifiers**: Uses in Processed Foods.)

Antioxidants are sometimes added to cake mixes to retard the development of oxidative rancidity during storage. All fats are subject to oxidative and hydrolytic rancidity, which causes objectionable odors and flavors, but antioxidants delay these reactions from occurring within the products' shelf-life. Four compounds commonly used as antioxidants are butylated hydroxyanisole, butylated hydroxytoluene, t-butyl hydroquinone, and propyl gallate. Citric and phosphoric acids have a synergistic effect when combined with the antioxidants. The levels are limited by law, economics, and functionality, and vary with the additive and product application, but generally fall between 0.005 and 0.1% of the product weight. Natural antioxidants, e.g. tocopherals, offer a desirable alternative to synthetic varieties, but have other usage issues such as heat lability. (*See* **Antioxidants**: Synthetic Antioxidants.)

Color additives are used in many baked products, including cakes and their icings. Added color can help the product fulfill consumer perceptions and expectations regarding quality, richness, and overall visual appeal. There are two types of color additives – certified and uncertified. The certified colors are synthetic and strictly regulated, whereas the uncertified usually come from natural sources and usage level varies greatly. The certified colors used in the USA include FD&C blue no. 1, FD&C red no. 3, FD&C yellow no. 5 and FD&C red no. 40. Uncertified color additives include annatto extract, β-carotene, beet powder, β-apo-8-carotenal, xanthins, caramel, carmine, carrot oil, cochineal extract, toasted partially defatted cottonseed flour, fruit and vegetable juices or concentrates, paprika and paprika oleoresin, riboflavin, saffron, titanium dioxide, tumeric, and tumeric oleoresin. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments; Properties and Determinants of Synthetic Pigments.)

Many flavoring agents are used in cake batters, icings, and/or fillings. Spices are processed from different parts of aromatic plants, including fruits, barks or seeds. Some spices commonly used in cakes include allspice, anise, caraway seed, cardamom, cinnamon, cloves, coriander, ginger, mace, nutmeg, poppy seed, saffron, and sesame seed. Some of these act as both flavoring and coloring agents.

Alcohol extracts can also be used to enhance the flavor of cakes. The sapid and odorous volatile components are extracted from aromatic plants or parts of the plant and solubilized in ethanol or propylene glycol. For example, vanilla extract is a ubiquitous flavoring in cakes, derived from the vanilla bean.

Chocolate and cocoa from the cacao tree bean are also popular flavoring agents. However, defatted cocoa powder also adds bulk to the cake, often replacing up to 10% of the flour weight. Often in chocolate cakes, the sugar level and leavening system must be adjusted to compensate for the cocoa. (*See* **Cocoa**: Production, Products, and Use.)

See also: **Antioxidants**: Synthetic Antioxidants; **Cocoa**: Production, Products, and Use; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; Properties and Determinants of Synthetic Pigments; **Eggs**: Dietary Importance; **Emulsifiers**: Uses in Processed Foods; **Flour**: Roller Milling Operations; Analysis of Wheat Flours; Dietary Importance; **Leavening Agents**

Further Reading

Bennion M (1980) *The Science of Food*. San Fransisco: Harper and Row.

Bennion EB and Bamford GST (1997) *The Technology of Cake Making*, 6th edn. In: AJ Bent (ed.) London: Blackie Academic.

Blanshard JMV, Frazier PJ and Galliard T (1986) *Chemistry and Physics of Baking*. Special publication no. 56.London: Royal Society of Chemistry.

Charley H (1982) *Food Science*, 2nd edn. New York: Macmillian.

Hoseney RC (1986) *Principles of Cereal Science and Technology*. St Paul, MN: American Association of Cereal Chemists.
Matz SA (1992) *Bakery Technology and Engineering*, 3rd edn. New York: Van Nostrand Reinhold.
Pyler EJ (1988) *Baking Science and Technology*, 3rd edn. Kansas City, MO: Sosland.
Stauffer CE (1990) *Functional Additives for Bakery*, 5th edn. New York: Van Nostrand Reinhold.

CALCIUM

Contents

Properties and Determination

K A Cockell, Sir F.G. Banting Research Centre, Ottawa, Ontario, Canada

Calcium

Calcium is an alkaline earth element and a mineral nutrient essential for plant and animal life. This article covers the basics of calcium chemistry, its occurrence in foods, influences of food processing on calcium content and bioavailability, and methods for analysis of calcium in foods and biological samples.

Chemistry of Calcium

The chemical element calcium, symbol Ca, atomic number 20, is a group IIA metal (alkaline earth metal), first isolated by British chemist Sir Humphry Davy in 1808. The name is derived from '*calx*,' Latin for lime (calcium oxide), which was used by the Romans in mortar for construction. Calcium has an atomic weight of 40.08 and six stable isotopes, with ^{40}Ca predominating at over 96% relative abundance (**Table 1**). There are several calcium radionuclides available, including ^{41}Ca, ^{45}Ca, and ^{47}Ca. For practical reasons, ^{45}Ca is the most commonly used calcium radionuclide for biological investigations.

The electron configuration of calcium is $1s^2 2s^2 2p^6 3s^2 3p^6 4s^2$. Calcium thus has two valence electrons and commonly occurs as a divalent cation (Ca^{2+}). Calcium does not occur as a free metal in nature, being found rather in several common mineral compounds including carbonates (e.g., limestone, marble, chalk), sulfates (e.g., gypsum, alabaster), fluorides (e.g., fluorite), phosphates, silicates, or as ionic calcium in seawater ($\sim 400\,mg\,l^{-1}$). Calcium is the fifth most abundant element, and third most abundant metal, in the earth's crust (comprising almost 4%), and ranks seventh in terms of content in seawater.

In living organisms, calcium occurs in the form of solid mineral salts as well as dissolved in solution, in contrast to the other major alkali and alkaline earth metals (sodium, potassium, and magnesium) where solution chemistry predominates. Calcium in solution has an affinity for oxygen-containing ligands such as carboxy and phosphate groups.

Calcium as an Essential Nutrient

Calcium is an essential nutrient for all plants and animals including man. It is the most abundant

Table 1 Stable isotopes and selected radioactive nuclides of calcium

Stable isotopes		*Radioactive nuclides (partial list)*				
Isotope	*Relative abundance (%)*	*Nuclide*	*Decay*	*Half-life*	*Energy (MeV)*	*Intensity (%)*
^{40}Ca	96.941	^{41}Ca	Electron capture	1.02×10^5 years	0.42	100
^{42}Ca	0.647	^{45}Ca	β^- (no γ emitted)	162.7 days	0.26	100
^{43}Ca	0.135					
^{44}Ca	2.086	^{47}Ca	β^-	4.536 days	1.98	16
^{46}Ca	0.004				0.68	84
^{48}Ca	0.187		γ		1.30	75

metal and fifth most abundant element (after oxygen, carbon, hydrogen, and nitrogen) in the human body. In both plants and animals, calcium plays key structural and metabolic roles.

In plants, calcium forms salt bridges between pectin molecules in the middle lamella, providing the 'glue' that holds the walls of adjacent cells together. Calcium in solution regulates the activity of many enzymes in plant cells, and protoplasmic calcium concentrations are regulated by sequestration of calcium in vacuoles where it is precipitated in an inert form such as calcium oxalate.

In animals, calcium in mineral form is plentiful as the carbonate ($CaCO_3$) in shell and coral and as hydroxyapatite ($Ca_5(PO_4)_3(OH)$) in bone and tooth enamel. Calcium in solution acts an important intracellular signal mechanism in animals, essential for muscle contraction, blood clotting and many other functions. (*See* **Calcium**: Physiology.)

Overview of Occurrence in Foods

Roughly 70% of the calcium present in the food supply in the USA is in the form of milk and dairy products, with about 10% in fruits and vegetables, 5% in grain products, and the remainder in all other foods. Similar proportions have been reported from other Westernized countries. The relative importance of these food sources to people's calcium intakes depends upon the relative amounts consumed. Even so, the main food sources of calcium for most people consuming omnivorous diets in Western countries are milk and dairy products, typically accounting for over 50% of total calcium intake. Although the concentration of calcium in cereal products and vegetables is typically much lower than in dairy products, the amount of these foods eaten means that foods of plant origin do make a significant contribution to total calcium intakes for most people (20–25% from cereals in the USA and the UK, with about 14% from bread due to calcium fortification of white flour). It is commonly estimated that the total contribution of nondairy foods eaten over the course of a day provides 200–300 mg of calcium to those consuming a mixed diet, or about the equivalent of one glass of milk. Other foods that contain considerable amounts of calcium include tinned fish such as sardines or salmon (largely due to the calcium in the bones and skin), tofu (especially if it has been processed with calcium) and certain seeds and nuts, but typical eating habits indicate that these do not contribute much to total calcium intake for most people. Vegans, or others who do not consume milk and dairy products, may need to take particular care to ensure adequate calcium intake from their diet or may require a supplemental source of calcium, particularly during periods of rapid growth. Lactovegetarians, who do not avoid milk and dairy products, generally have little difficulty in attaining adequate calcium intakes.

Factors Influencing Calcium Bioavailability

Knowledge of the calcium content of a food must be tempered to some extent by understanding the dietary and physiological factors that can influence the bioavailability of calcium. Calcium bioavailability can be operationally defined as the amount of calcium that is available to be absorbed and used for normal metabolic functions in the body. Physiological factors influencing calcium bioavailability include rate of growth, calcium status and age (*See* **Calcium**: Physiology). A number of dietary factors can influence the bioavailability of calcium, including the level of calcium intake and its chemical form, whether calcium is taken with food or as a supplement on an empty stomach, and the presence of inhibitory or enhancing compounds coingested with the calcium. Calcium is more bioavailable when taken with food than on an empty stomach, due at least in part to the longer passage time through the GI tract. Oxalate and phytate are well-known inhibitors of calcium uptake through the formation of insoluble complexes within the gut. With the exception of vitamin D, which is involved in the active transport uptake of calcium from the gut (*See* **Calcium**: Physiology; **Cholecalciferol**: Physiology), potential enhancers of dietary calcium absorption are generally less well defined than the inhibitors, and may have only limited influence on the intestinal absorption of calcium ingested with a meal. Absorption from the gastrointestinal tract is not the only consideration in bioavailability; dietary sodium and protein (particularly proteins high in sulfur amino acid content due to the physiological acidification effect) influence the amount of calcium excreted in urine and consequently not available for essential functions in the body.

Sources of Calcium Intake

As alluded to previously, calcium is quite unevenly distributed in the diet of humans. Some foods are rich sources of calcium, whereas others contain relatively little. With developments in food processing and fortification policies and practices, the increasing availability of calcium-fortified foods and dietary supplements containing calcium salts is leading to a wider range of rich dietary sources of calcium.

Foods of Animal Origin

Milk and many dairy products are good sources of bioavailable calcium (Table 2). Removal of the milk fat fraction from fluid milk leads to slight increases in the calcium concentration as one proceeds from whole milk to skim milk. Fluid milk with added modified milk ingredients is a relatively new entry into selected marketplaces, containing some 20–25% more calcium than regular milk due to the addition of whey calcium concentrates. Naturally, foods containing milk or milk products, whether 'hidden' in the recipe as in cakes or other baked goods or more obviously added as with cream or cheese sauces, will also provide a fair amount of calcium.

Soft cheeses and cream cheese generally contain less calcium on a wet-weight basis due to their higher water content. The calcium content of cheeses also depends on the process used in preparation and whether the calcium precipitates with the milk solids or remains in the whey. Cheeses precipitated with lactic acid, such as cottage cheese or cream cheese, contain lower calcium levels in the curd, as most of the calcium remains soluble in the acid whey. Rennin coagulation (e.g., in production of cheddar or mozzarella cheese) yields cheeses with a higher calcium content as the curd is formed before significant acidification takes place.

Meat and fish are not typically rich sources of calcium (Table 2). Mechanically deboned meats can contain much higher levels of calcium due to the abrasion of bone during the deboning process. Tinned salmon provides higher levels of calcium than fresh salmon fillet (Table 2), because the bones and skin tend to be consumed along with the flesh. Similar reasoning applies to small fish consumed whole, such as sardines and anchovies. Calcium is generally somewhat higher in crustaceans than in fin fish.

Foods of Plant Origin

Foods of plant origin are not generally very rich sources of calcium, and some contain significant levels of calcium absorption inhibitors such as phytate or oxalate. However, due to the large amounts consumed, this class of foods typically makes a fairly significant contribution to total calcium intake. Whole grains and seeds generally provide more calcium than most fruits or vegetables. Fortification of refined flours and breakfast cereals can significantly increase the contribution of these foods to total calcium intake. Similarly, the high calcium content of baking powder and the variety of calcium-containing food additives used (e.g., as dough conditioners or yeast nutrients – see below) adds to the amount of calcium in baked goods.

While spinach appears to contain a reasonable amount of calcium (129 mg per 95-g serving, see Table 3), the high oxalate content of spinach renders much of the calcium insoluble and much less

Table 2 Calcium content of selected animal-source foods[a]

Food name	*Serving size*	*Weight (g)*	*Calcium (mg)*	*Concentration (mg g^{-1} as is)*
Milk and dairy products				
Milk, whole, 3.3% M.F.[b]	250 ml	258	308	1.19
Milk, partly skimmed, 2% M.F.	250 ml	258	314	1.22
Milk, skim	250 ml	259	319	1.23
Milk, skim, with added milk solids		100	143	1.43
Milk, condensed, sweetened, canned	300 ml	388	1100	2.84
Milk, evaporated, partly skimmed, canned, undiluted, 2% M.F.	250 ml	268	739	2.76
Yogurt, fruit bottom, 1–2% M.F.	175 g	175	214	1.22
Cheese, cheddar	4 slices	52	378	7.27
Cheese, mozzarella, shredded, 22.5% M.F.	125 ml	60	321	5.35
Cheese, cottage, 2% M.F.	125 ml	119	82	0.69
Meat, fish, and egg				
Egg, poached	1 large	50	24	0.48
Beef, ground, lean, medium broiled	1 patty	80	9	0.11
Beef, mechanically deboned, raw		100	485	4.85
Pork chop, centre cut, lean, pan-fried	1 chop	69	16	0.23
Chicken, broiler, breast, roasted	½ breast	98	5	0.05
Salmon, Sockeye, baked or broiled	½ fillet	155	11	0.07
Salmon, Sockeye, canned, solids+bone+liquid	125 ml	79	181	2.29
Sardine, Atlantic, canned in oil, drained, with bone (7.5 cm long)	4 sardines	48	183	3.81

[a]Data from Health Canada (1999) *Nutrient Value of Some Common Foods*. Ottawa: Health Protection Branch, Health Canada, Canadian Government Publishing, except where no serving size is specified, which data are from the Canadian Nutrient File, Health Canada, 1997 revision.
[b]Milk fat.

Table 3 Calcium content of selected plant-based foods[a]

Food name	*Serving size*	*Weight (g)*	*Calcium (mg)*	*Concentration (mg g^{-1} as is)*
Vegetables and fruit				
Beans, snap, frozen, boiled, drained	125 ml	71	35	0.49
Carrots, raw	1 medium	80	22	0.28
Peas, green, frozen, boiled, drained	125 ml	85	20	0.24
Spinach, boiled, drained	125 ml	95	129	1.36
Apple, raw, with skin (7 cm diameter)	1 apple	138	10	0.07
Banana, raw	1 medium	115	7	0.06
Oranges, raw	1 fruit	131	52	0.40
Grain and cereal products, nuts, and seeds				
Wheat, flour, all purpose	250 ml	132	20	0.15
Bread, white, commercial	1 slice	25	27	1.08
Corn meal	125 ml	73	4	0.05
Flax seeds	125 ml	73	196	2.68
Soy flour, defatted	250 ml	106	255	2.41
Tofu, firm, prepared with magnesium chloride	1 piece	80	164	2.05
Tofu, firm, prepared with calcium sulfate		100	683	6.83
Tofu, regular, prepared with calcium sulfate		100	350	3.50
Almonds, dry-roasted, salt added	125 ml	73	206	2.82
Peanuts, dry-roasted, without shell	125 ml	78	42	0.54
Hazelnuts, chopped, dried	125 ml	61	119	1.95
Sesame crisp	4 pieces	35	247	7.06

[a]Data from Health Canada (1999) *Nutrient Value of Some Common Foods*. Ottawa: Health Protection Branch, Health Canada, Canadian Government Publishing, except where no serving size is specified, which data are from the Canadian Nutrient File, Health Canada, 1997 revision.

Table 4 Bioavailability of calcium from various food sources[a]

Food	*Serving size (g)*	*Ca content (mg)*	*Fractional absorption (%)*	*Servings equivalent to 240 ml milk*
Milk	240	300	32.1	1
Soy milk (unfortified)	120	5	31.0	60.4
Almonds, dry-roasted	28	80	21.2	5.7
Sesame seeds, no hulls	28	37	20.8	12.2
Broccoli	71	35	52.6	5.0
Brussel sprouts	78	19	63.8	8.0
Spinach	90	122	5.1	15.5
Turnip greens	72	99	51.6	1.9
Tofu, calcium set	126	258	31.0	1.2

[a]Selected data from Weaver CM and Plawecki KL (1994) Dietary calcium: adequacy of a vegetarian diet. *American Journal of Clinical Nutrition* 59 (supplement): 1238S–1241S.

bioavailable, so that it has been estimated that over 15 servings of spinach would be required to obtain the same level of absorbable calcium as one glass of milk (Table 4). Turnip greens contain a similar level of calcium to that found in spinach, but without the oxalate, so that about 2 servings of turnip greens provide a similar level of absorbable calcium to one glass of milk (Table 4). However, these types of food calcium absorption values are often determined under laboratory conditions while feeding only the single food. Calcium bioavailability from a given food source can be substantially modified by other foods eaten at the same time. For example, milk calcium absorption has been shown to decline from 33 to 27% when coingested with spinach, whereas spinach calcium absorption showed an increase from 3% when fed alone to 11% when coingested with milk. In a Western mixed diet, the overall calcium absorption by adults typically averages about 25–30%.

Processing Effects on Food Calcium

There are numerous food additive uses of calcium salts that may add appreciable amounts of calcium to some foods. The Food Chemicals Codex, fourth edition (1996) lists over 30 calcium compounds used as food additives or processing aids. Some of the common functions of calcium-containing food additives (and examples of the calcium compounds used) include: dough conditioners such as calcium carbonate, calcium iodate, calcium lactate, calcium oxide, calcium phosphate or calcium sulfate; pH adjusters such as calcium hydroxide or calcium phosphate; antioxidants such as calcium ascorbate; preservatives such as calcium propionate, calcium disodium EDTA

or calcium sorbate; anticaking agents such as calcium stearate, calcium phosphate or calcium silicate; and thickeners such as calcium gluconate or calcium alginate. The percentage composition of calcium in these compounds ranges considerably, e.g., calcium gluconate is 9% calcium, calcium carbonate is 40% calcium, and calcium oxide is 71% calcium by weight. Solubility also differs considerably between compounds, with some such as calcium carbonate being relatively insoluble at neutral pH, whereas others like calcium acetate or calcium chloride are highly soluble.

Tofu set with calcium sulphate contains considerably more calcium than tofu set with magnesium chloride (Table 3). Firm tofu contains a higher concentration of calcium than does regular tofu due to the lower water content (and consequently higher solids). Fermentation (including leavening of bread) or germination of selected plant foodstuffs can increase the availability of calcium due to breakdown of complexing compounds such as phytic acid. The calcium content of meat can be increased by processing or cooking in acidic solutions due to the dissolution of calcium from bone. For example, pork spare ribs or chicken cooked with vinegar contain more calcium than the uncooked flesh.

Tap water typically contains calcium and many other elements, and may make some contribution to total calcium intake particularly in hard water areas. Water hardness is measured as milligrams of calcium carbonate equivalents per liter, and hard water may be over 300 mg of $CaCO_3$ per liter. Calcium carbonate is 40% elemental calcium by weight. Thus, a 250-ml glass of water may provide 30 mg of elemental calcium in some areas. The hardness of water added in processing may also influence the calcium content of foods to some extent. Conversely, losses through soaking and/or boiling may amount to 5–25% of the calcium content of a variety of vegetables that are commonly prepared in this way.

Foods may have nutrients added for a variety of reasons including restoration of processing losses, or to provide similar nutrient levels in substitute foods (e.g., soy 'milk' and other plant based beverages, which may be fortified with calcium in some countries, as they are used as a substitute for dairy products in the diet of their consumers), for fortification (which, in this context, refers to addition of calcium with the intent of raising the calcium content of the food for the consumer, e.g., enriched flour or infant cereals), or for special purpose foods such as meal replacements or formulated liquid diets (*See* **Legislation**: Codex). Regulations and practices concerning food fortification may vary considerably between countries.

Supplements

Nutritional supplements and antacid medications can make a significant contribution to calcium intakes for some people. Calcium supplements are available in a number of forms, including carbonate, phosphate, lactate, or citrate salts of calcium, among others. Calcium carbonate is found in many over-the-counter antacid preparations. Multivitamin and multimineral supplements may also contain calcium, though frequently in smaller amounts. Natural source supplements include oyster shell and dolomite, though it should be noted that concerns have been raised with the potential for lead contamination of these forms. It has been estimated that, among supplement users in the USA (who represent approximately 20% of the population), an average of 300 mg of Ca per day is consumed from supplements. In addition, calcium-containing antacids, supplying 200–400 mg of Ca per tablet, are consumed by about 18% of USA adults.

Analysis of Calcium in Foods and Biological Samples

There is no satisfactory routine biochemical method for assessing calcium nutritional status (*See* **Nutritional Assessment**: Biochemical Tests for Vitamins and Minerals). Serum calcium concentration varies very little, even across a wide range of dietary calcium intakes, because of adaptive alterations by which the endocrine system regulates the level of this mineral in blood (*See* **Calcium**: Physiology). Deviations of calcium concentration from this narrow range are medically significant, and in this setting, the measurement of total plasma or serum calcium or specific measurement of the ionized calcium fraction is important. Total calcium in plasma or urine can be measured by atomic absorption spectrophotometry, either directly on the diluted sample or following sample mineralization ('ashing' – see below), or by colorimetric assay (e.g., using arsenazo III or orthocresolphthalein complexone) on a clinical autoanalyzer. Ionized calcium in plasma (the fraction that is not bound to proteins or low-molecular-weight ligands, and consequently is most physiologically active) represents about 45% of plasma calcium under normal conditions and can be assessed using an appropriate ion-sensitive electrode. Assessment of intracellular calcium levels in the research laboratory can be addressed using fluorescent probes such as FURA 2 or QUIN 2, among others. Calcium nutritional status is more often assessed indirectly through balance studies or by measurement of bone mineral concentration or bone mineral density, reflecting the major structural role played by this mineral nutrient.

Sample Preparation

Most foods and biological tissue samples can be prepared for analysis of their calcium content either by dry ashing (in a muffle furnace) or wet ashing (acid digestion) techniques or some combination of the two. Dry ashing techniques typically involve combustion at elevated temperature (e.g., 500–550 °C) until organic matter is fully destroyed, followed by dissolution of the ash in a suitable acid for subsequent analysis. Wet ashing techniques typically involve destruction of organic matter by heating in concentrated nitric acid until a pale straw color is attained, and may feature additional oxidation steps with perchloric acid or hydrogen peroxide to clarify the sample solution. In wet ashing procedures, particularly for samples with a high calcium content such as bone meal, sulfuric acid is generally to be avoided due to the ready precipitation of calcium from solution by sulfate, forming plaster of Paris (calcium sulfate hemihydrate) or gypsum (calcium sulfate dihydrate).

Determination of Calcium Content of Foods and Tissues

Determination of calcium content of foods and tissues is commonly done by atomic absorption spectrophotometry, though a number of alternative methods exist including titrimetric methods using EDTA or $KMnO_4$, neutron activation analysis, or inductively coupled plasma emission spectrometry. Details of sample preparation and analytical procedures for a variety of food sample types are published by the Association of Official Analytical Chemists.

Atomic Absorption Spectrophotometry

For atomic absorption spectrometric analysis of calcium, 0.1–1.0% (w/v) lanthanum is included in the analytical working solution as a matrix modifier to reduce anion interferences due to phosphate or sulfate, which otherwise can form refractory complexes and depress the absorption of light by atomic calcium. Absorbance is measured at the 422.7-nm calcium spectral line, following atomization in a reducing air–acetylene flame, and compared with certified analytical standard calibration solutions. A reducing flame gives a higher sensitivity, though an oxidizing flame may give a higher precision where this is critical. Typical analytical working ranges are obtained up to $5\,mg\,l^{-1}$ in the analytical working solution when using a standard nebulizer assembly, and may be approximately doubled with the use of a high sensitivity nebulizer. (*See* **Spectroscopy**: Atomic Emission and Absorption.)

$KMnO_4$ Titrimetric Method for Calcium in Wheat Flour (Method from the Association of Official Analytical Chemists)

After dry ashing and dilution of the ash to a suitable volume with demineralized water, bromocresol green indicator is added along with enough 20% sodium acetate solution to change the pH to 4.8–5.0 (blue). The sample solution is covered and heated to boiling. The calcium is precipitated by slow addition (1 drop every 3–5 s) of 3% oxalic acid solution (w/v) until pH is 4.4–4.6, as indicated by a distinct green shade. The solution is then boiled for 1–2 min and allowed to settle overnight. The supernate is filtered through quantitative paper, Gooch, or fine fritted glass filter, and the beaker and precipitate are washed with small portions of ammonium hydroxide (1 + 50). A mixture of 125 ml of water and 5 ml of sulfuric acid is added to the precipitate with heating to 80–90 °C. The solution is finally titrated at 70–90 °C with 0.01 M $KMnO_4$ to a slight pink endpoint, 1 ml of permanganate solution equating to 1 mg of calcium.

See also: **Calcium**: Physiology; **Cholecalciferol**: Physiology; **Nutritional Assessment**: Biochemical Tests for Vitamins and Minerals; **Spectroscopy**: Atomic Emission and Absorption

Further Reading

Bender AE (1978) *Food Processing and Nutrition*. London: Academic Press.

Flynn A and Cashman K (1999) Calcium. In: Hurrell R (ed.) *The Mineral Fortification of Foods*, pp. 18–53. Leatherhead, UK: Leatherhead International.

Fraser D, Jones G, Kooh SW and Radde IC (1986) Calcium and phosphate metabolism. In: Tietz NW (ed.) *Textbook of Clinical Chemistry*, pp. 1317–1372. Philadelphia, PA: W.B. Saunders Company.

Guéguen L and Pointillart A (2000) The bioavailability of dietary calcium. *Journal of the American College of Nutrition* 19(2): 119S–136S.

Health Canada (1999) *Nutrient Values of Some Common Foods*. Ottawa: Health Protection Branch in cooperation with Health Promotion and Programs Branch, Health Canada. Canadian Government Publishing.

Hibbins SG (1992) Calcium and calcium alloys. In: *Kirk–Othmer Encyclopedia of Chemical Technology*, 4th edn, vol. 4, pp. 777–786. New York: John Wiley.

Horowitz W (ed.) AOAC (2000) *Official Methods of Analysis of AOAC International*, 17th edn. Gaithersburg, MD: The Association of Official Analytical Chemists.

Institute of Medicine (1996) *Food Chemicals Codex*, 4th edn. Washington, DC: Committee on Food Chemicals Codex, Food and Nutrition Board, Institute of Medicine, US National Academy of Sciences. National Academy Press.

Institute of Medicine (1997) *Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride*. Washington, DC: Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine, US National Academy of Sciences. National Academy Press.

Karmas E and Harris RS (eds) (1988) *Nutritional Evaluation of Food Processing*, 3rd edn. New York: Van Nostrand Reinhold Company.

Petersen RL and Freilich MB (1992) Calcium compounds (survey). In: *Kirk–Othmer Encyclopedia of Chemical Technology*, 4th edn, vol. 4, pp. 787–796. New York: John Wiley.

Weaver CM and Plawecki KL (1994) Dietary calcium: adequacy of a vegetarian diet. *American Journal of Clinical Nutrition* 59(supplement): 1238S–1241S.

Physiology

M R L'Abbé, Nutrition Research Division, Ottawa, Ontario, Canada

Background

Calcium (Ca, element number 20, atomic weight 40.08, oxidation state + 2) is the most abundant cation in the human body and is a nutritionally essential mineral. More than 99% of body calcium is in the skeleton, accounting for 1.5–2.0% of body weight (approximately 1.5 kg). Calcium has two main types of functions: structural and metabolic. It is responsible for the mechanical and structural functions of bone and teeth as well as many metabolic functions involving numerous regulatory enzymes that require calcium. The physiological functions of calcium are so vital that total serum calcium is maintained in a very narrow range through a highly integrated system of hormones (parathyroid hormone, calcitonin, and vitamin D). If dietary sources of calcium are inadequate, bone reserves are mobilized; inadequate dietary calcium has long been associated with osteoporosis – a disease characterized by low bone mineral density.

Calcium Metabolism

Absorption

Intestinal absorption of calcium proceeds by two independent routes, a saturable, active absorption pathway regulated almost exclusively by vitamin D and a nonsaturable passive pathway that depends on the concentration of calcium in the intestinal lumen. Most calcium absorption occurs in the duodenum and jejunum, but minor amounts can also be absorbed in the colon. The active absorption pathway is located primarily in the duodenum and upper jejunum. Absorption by this pathway varies with age, calcium and vitamin D status, and total calcium intake. It constitutes the mechanism whereby the body increases relative absorption during periods of low dietary calcium intakes or low calcium status. Most calcium is absorbed by the nonsaturable passive route in adults with a mature skeleton, and active transport is of relatively minor importance to the total calcium absorbed. Thus, the major factor that determines the amount of calcium absorbed is the quantity ingested. Although the efficiency of calcium absorption declines with increasing calcium dose, there is evidence that the total amount absorbed may continue to rise.

The absorption of calcium from food can vary from 5 to 60%. It is influenced by calcium status, vitamin D metabolites, pH of the gut lumen, and dietary factors such as protein, phosphorus, sodium, phytate, and oxalate (see sections on Bioavailability Dietary Factors Affecting Calcium Balance, for more details).

Bioavailability

Foods vary in both calcium content and bioavailability. Together, these two factors determine how much calcium a particular food source will provide per serving. Information on the content and bioavailability of calcium in various food products is given in **Table 1**. Milk and milk products have a high calcium content and bioavailability, and thus are major food sources of calcium in Western diets. Many vegetables have a higher fractional absorption (bioavailability) of calcium but a lower content of calcium than is found in milk and milk products. In addition, some vegetables and plant products contain inhibitors of calcium absorption, such as oxalate (e.g., spinach), or phytate (e.g., wheat bran, legumes, nuts, oats, soy beans, maize), which reduce the bioavailability of calcium.

Balance

Calcium balance is controlled through an integrated response to calcium-regulating hormones that affect calcium transport at three primary sites of regulation: intestine, bone, and kidney. The most important calcium-regulating hormones are parathyroid hormone (PTH), calcitonin and vitamin D, and the effects of these are briefly described below. Glucocorticoids, thyroid hormones, growth hormone, insulin, estrogen, and testosterone also affect bone turnover and calcium metabolism.

Table 1 Content and bioavailability of calcium in various food products[a]

	Serving size (g)	*Calcium[b] content (mg)*	*Fractional absorption (%)*	*Estimated absorbable Ca/serving (mg)*	*Servings needed to = 1 cup of milk*
Milk (1 cup)	240	300	32.1	96.3	1.0
Almonds, dry roasted (1 oz)	28	80	21.2	17.0	5.7
Beans, pinto	86	4.7	17.0	7.6	12.7
Beans, red	172	40.5	17.0	6.9	14.0
Beans, white	110	113	17.0	19.2	5.0
Broccoli	71	35	52.6	18.4	5.2
Brussel sprouts	78	19	63.8	12.1	8.0
Cabbage, Chinese (bok choi)	85	79	53.8	42.5	2.3
Cabbage, green	75	25	64.9	16.2	5.9
Cauliflower	62	17	68.6	11.7	8.2
Kale	65	47	58.8	27.6	3.5
Kohlrabi	82	20	67.0	13.4	7.2
Mustard green	72	64	57.8	37	2.6
Radish	50	14	74.4	10.4	9.2
Rutabaga	85	36	61.4	22.1	4.4
Sesame seeds no hulls (1 oz)	28	37	20.8	7.7	12.2
Soy beverage, not fortified	120	5	31.0	1.6	60.4
Spinach	90	122	5.1	6.2	15.5
Tofu, calcium set	126	258	31.0	80.0	1.2
Turnip greens	72	99	51.6	51.1	1.9
Watercress	17	20	67.0	13.4	7.2

[a]Selected data from Weaver CM and Plawecki KL (1994) Dietary calcium: adequacy of a vegetarian diet. *American Journal of Clinical Nutrition* 59(5): 1238S–1241S.
[b]Based on a half-cup serving size unless otherwise noted.

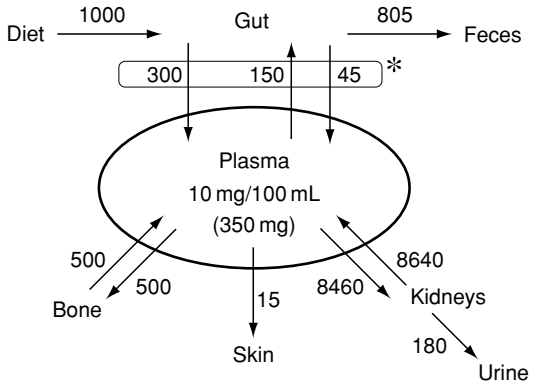

Figure 1 Outline of calcium balance (mg per day) under equilibrium conditions (input = output) in adult humans. *Assuming equilibrium conditions and 30% absorption, 300 mg of the daily dietary intake of 1000 mg are absorbed from the gut. Endogenous calcium secretion returns 150 mg per day to the gut lumen. Using the same assumption of 30% absorption, 45 mg per day of the endogenously secreted calcium are reabsorbed. Net calcium uptake is therefore 195 mg per day under these conditions, which offsets the net loss of 195 mg per day through the skin and in urine.

At equilibrium, calcium balance is maintained through regulation of plasma calcium levels, with the calcium influx/efflux occurring at several sites, as illustrated in Figure 1. If dietary calcium intakes are insufficient, or in the case of malabsorption syndromes, calcium is mobilized from bone to maintain blood concentration.

Hormonal Regulation

Vitamin D is an important nutrient for regulating plasma calcium concentration and bone metabolism. The two most physiologically active forms are vitamin D_2 (ergocalciferol), found mainly in yeast and plant foods, and vitamin D_3 (cholecalciferol), synthesized by skin and the most common form added to foods. Vitamin D is hydroxylated first in the liver at the 25-carbon position to form the 25-hydroxyvitamin D (25(OH)D), which enters the circulation. Circulating (25OH)D is a good measure of vitamin D status. In the kidney, the 25(OH)D is further hydroxylated to form the dihydroxy compound (1,25 $(OH)_2D$), which is the most active metabolite. The production of $1,25(OH)_2D$ in the kidney is tightly regulated via PTH (see below). The $1,25(OH)_2D$ increases calcium absorption in the intestine and is absolutely critical during growth as well as during times of low calcium intake. In calcium deficiency, $1,25(OH)_2D$ production is increased, causing enhanced intestinal absorption and renal reabsorption of calcium, and increased bone resorption.

PTH is released in response to a decrease in extracellular fluid calcium concentration. PTH directly stimulates bone and renal tubular calcium resorption and indirectly increases intestinal absorption through

enhanced formation of $1,25(OH)_2D$ in the kidney from the monohydroxy form, 25(OH)D.

Calcitonin is synthesized by the C cells of the thyroid in response to a rise in serum calcium; it inhibits bone resorption and agents that have resorptive effects on bone (PTH, vitamin D, prostaglandins, and vitamin A).

Excretion

The kidneys filter 8–10 g of Ca per day, with almost all of it reabsorbed; only about 100–200 mg per day is excreted in urine. Tubular reabsorption therefore plays an important part in the conservation of calcium, and a reduction in the number of functional nephrons or in their capacity to reabsorb normally can result in hypocalcemia.

Under ideal conditions (i.e., in calcium balance), excretion via the urine (and, to a lesser degree, via sweat and feces) is matched by the influx into the extracellular fluid compartment, usually through intestinal absorption.

Dietary Factors Affecting Calcium Balance

Vitamin D

As mentioned above (see section Hormonal Regulation), adequate vitamin D is critical to the absorption of calcium D; yet, there is little or no natural vitamin D in most foods that humans normally eat, except some fish. Vitamin D is present in the flesh of fatty fish such as sardines and herring, and in the liver oils of other fish such as cod and tuna. Meat such as poultry, pork, and beef contain small amounts of vitamin D. In Canada and in many other countries, vitamin D fortification of milk and/or margarine provide the major sources of dietary vitamin D.

The other major source of vitamin D for humans is synthesis by skin exposed to ultraviolet or sunlight irradiation, and most people could obtain sufficient vitamin D from this source. However, this vital photosynthetic process can be affected by anything that limits the amount of solar ultraviolet B photons to reach the skin's surface and penetrate into the viable epidermis. In northern latitudes, the photosynthetic production of vitamin D is negligible during the winter months (e.g., from November to March, or virtually half the year in much of Canada). The elderly also have a reduced capacity for vitamin D synthesis, and deficiency is most prevalent among the elderly, particularly among institutionalized or house-bound individuals. Factors such as pollution in the air and melanin of the skin can affect the cutaneous synthesis of vitamin D. In addition, sunscreens block the absorption of the sunlight spectrum that helps the cutaneous synthesis of vitamin D, and there has been a marked increase in the use of sunblocks in recent years in many countries. Thus, providing a reliable, safe, and adequate source of vitamin D in the diet is essential for optimum calcium absorption.

Phosphorus

There is some controversy concerning the effects of high phosphorus on calcium balance. In the past, considerable emphasis was placed on the Ca:P ratio of the diet, particularly in infant nutrition. In the most recent review of calcium and phosphorus requirements done by the National Academy of Sciences, this concept was considered to have no demonstrable relevance in adults, although the report stated: 'The concept still has some utility under conditions of rapid growth (in which a large share of the ingested nutrients is converted into tissue mass).' The ratio of calcium to phosphorus by itself does not take into account the differing bioavailabilities of the two nutrients, but estimates of optimal Ca:P ratios have frequently been based on the calcium and phosphorus needs of bone building. It was estimated that an absorbed Ca:P ratio sufficient to support the sum of bony and soft tissue growth would be approximately 1.3:1, and correcting for the relative net absorptions, this would be met by Ca:P intakes in the range of about 2:1.

Currently, in the USA, the intake of phosphorus is high relative to calcium. Furthermore, an individual's actual phosphorus intake may be higher than estimated due to inaccurate nutrient composition tables for phosphorus values, coupled with underreporting of processed foods. In recent years, the usage of phosphate salts as food additives has increased by nearly 20%. Currently, there are more than 45 phosphorus-containing compounds approved for use in food processing as nutrients, dietary supplements, or for functional purposes such as to preserve moisture or color, and as emulsifiers. Cola beverages and root beer beverages can be a significant source of phosphorus, containing about 70 mg of phosphorus per serving. One study has suggested that high intakes of phosphoric acid-containing cola beverages were associated with slight reductions in serum calcium in Mexican children, but it is unclear to what extent the effect was due to the acid load of the cola beverages, low calcium intakes, or the phosphorus itself.

Because phosphorus absorption is not strictly regulated, a high-phosphorus diet will directly increase the level of serum phosphate, which in turn lowers

serum calcium, a signal for PTH release. Current research describing diets in terms of the Ca:P ratio has suggested that diets with low ratios (i.e., high phosphate intakes where Ca:P equals 0.25:1) cause serum calcium to drop, thereby increasing PTH secretion, and inducing secondary hyperparathyroidism. This increase in PTH is often thought to be harmful to bone. Other researchers, however, have shown that calcium and phosphorus have no effect on each other in the healthy adult population. For example, balance studies in human adults with Ca:P ratios ranging from 0.08:1 to 2.40:1 had no effect on either calcium balance or calcium absorption.

Protein

North Americans consume considerably more calcium than people in developing countries; yet, the incidence of osteoporosis is much higher in North America. Recommended intakes for calcium in developed countries is much higher than for countries that consume a plant-based diet. One possible reason for this is the high amount of processed foods and animal protein found in Western diets that are not present in diets from primarily cereal-based food economies. With regard to processed foods, phosphorus has been discussed above, and salt and caffeine will be discussed below. Animal proteins contain a high proportion of sulfur amino acids, and the oxidation of these sulfur amino acids to inorganic sulfate generates endogenous acid, which in turn increases urinary excretion of acid and calcium. Increasing urinary calcium excretion will result in bone loss when there is insufficient capacity to replace the loss by absorbing more dietary calcium. High-protein-induced hypercalciuria is a misnomer, in that simply comparing high protein with low protein does not show hypercalciuria unless phosphorus is controlled or the amount of protein added is exceedingly high, or there is a change in the type of protein from one low in sulfur amino acids to a protein that is high in sulfur amino acids. In general, for every gram of total protein consumed, urinary calcium increases by approximately 1 mg.

Salt

High intakes of sodium chloride (NaCl) result in increased renal calcium loss, which promotes negative calcium balance. Urine calcium level increases by approximately 26 mg for every gram of salt ingested. Habitual excessive sodium chloride intakes have been suggested as a factor in promoting bone loss. However, given the small amount of calcium loss for every gram of salt consumed, the overall impact from dietary salt on calcium metabolism is likely minimal unless excessive amounts are consumed.

Caffeine

Oral doses of caffeine increase the urinary excretion of calcium for at least 3 h after consumption. Calcium balance studies have been performed in which caffeine was shown to have no effect when dietary calcium levels were adequate. The amount of caffeine in a small cup of coffee could raise urinary calcium excretion by approximately 4 mg, an effect that is less than that for salt. However, for persons with a low calcium intake, caffeine ingestion can be a significant predictor of low bone mineral density. Caffeine is found in only a small number of food products (e.g., colas, coffee, tea); its consumption would be unlikely to have a negative impact on calcium status, as urinary losses would be more than compensated for by a teaspoon of milk.

Methods for Assessment of Calcium Status

Serum concentration of calcium varies very little, even across a wide range of dietary calcium intakes, because of the adaptive alterations by which the endocrine system regulates the level of this mineral in blood. Therefore, there is no satisfactory routine biochemical method for assessing nutritional calcium status. If serum calcium levels are found to be outside the normal range, it is usually due to underlying disease conditions rather than nutritional problems. Calcium is found in several forms in serum: ionized, protein bound, and other minor blood components, with the ionized form of calcium being the functionally regulated form. Numerous biological markers or techniques have been used to assess calcium status. These can be classified into several groups: serum calcium or ionic calcium, urinary excretion of calcium, bone mineral density measurements or other techniques to measure the degree of bone loss, quantification of markers of bone resorption or bone remodelling, and finally calcium balance or absorption studies. Information on each of these measures is given in the following section.

Serum Calcium

Serum calcium concentration cannot be used as an index of calcium status, as it is homoeostatically controlled. Calcium concentration is serum varies very little, despite large changes in dietary calcium, because of the adaptive alterations by which the endocrine system regulates this mineral (see section Hormonal Regulation). Calcium in serum is distributed in the following forms: ionized 47.5%, protein bound 46%, Ca citrate 1.7%, $CaHPO_4$ 1.6%, and unidentified complexes 3.2%.

Serum-ionized Ca Concentration

This test measures 'physiologically active' calcium and thus better reflects calcium metabolism than total calcium values. It is useful in determining calcium changes in patients with altered proteins (e.g., chronic renal failure, nephrotic syndrome, hyperparathyroidism, malabsorption, multiple myeloma) but is not well correlated with dietary intakes of calcium.

Urinary Calcium Excretion

Urinary excretion of calcium can be elevated following high intakes of dietary calcium, although urinary excretion of calcium accounts for only 0.2% of the calcium filtered by the kidneys, because 99.8% of filtered calcium is reabsorbed. Many conditions can be associated with hypercalciuria, including metabolic disturbances such as hyperparathyroidism, bone metastases, osteoporosis (especially after immobilization), Cushings syndrome, and vitamin D intoxication. High dietary sodium or high dietary protein can also promote increased excretion of calcium. Hypercalciuria usually occurs prior to hypercalcemia.

Bone Mineral Density

Bone mineral density (BMD) is used as an index of body calcium stores, and techniques such as radiography, single or dual energy X-ray absorptiometry, and computerized tomography have been developed to measure BMD. The standard technique for diagnosing the presence or absence and severity of osteoporosis is based on BMD and a comparison with reference ranges established for normal BMD in young adults (see section Osteoporosis).

Markers of Bone Resorption

Blood biochemical markers of bone resorption rates tend to be less sensitive and specific than those for bone remodeling and include: urinary hydroxyproline, a measure of total collagen turnover, and pyridinoline and deoxypyridinoline, which are 3-OH-pyridinium derivative cross-links more specific to bone collagen. These measures indicate the degree by which bone reserves of calcium are being depleted.

Markers of Bone Formation

Markers of bone formation include serum osteocalcin (bone gla-protein), bone alkaline phosphatase, and urinary nephrogenous cyclic AMP.

Calcium Balance

Calcium status has been assessed indirectly by assessing calcium balance data, as increased retention is usually seen when status is poor. Techniques used include stable isotope or radiolabeled absorption and calcium retention measures to estimate calcium balance.

Function

Calcium has two main types of functions: structural and metabolic. It is responsible for the mechanical and structural functions of bone and teeth as well as many metabolic functions involving numerous soft tissue regulatory enzymes that require calcium. These include such processes as neuromuscular transmission of chemical and electrical stimuli, enzyme activation, membrane transport of inorganic ions, muscle contraction, hormone and cellular secretions, signal transduction, blood clotting, and reproductive functions such as sperm motility and fertilization of the ovum. Calcium regulates release and storage of neurotransmitters and hormones, the uptake and binding of amino acids, absorption of vitamin B_{12}, and gastrin secretion. The activities of many enzymes are affected by calcium, either acting directly (e.g., glyceraldehyde phosphate dehydrogenase, pyruvate dehydrogenase, α-ketoglutarate dehydrogenase) or indirectly through its activation of calmodulin (e.g., adenylate cyclase, cyclic nucleotide phosphodiesterase, Ca, Mg-ATPase, myosin light chain kinase, phosphorylase b kinase, NAD kinase, ornithine decarboxylase). Calcium therefore plays a role in normal cardiac function, renal function, respiration, blood coagulation, cell and membrane capillary permeability, cyclic nucleotide metabolism, glycogen metabolism, microtubule and microfilament function, and cell division.

The physiological functions of calcium are so vital to survival that total serum calcium is maintained in a very narrow range (2.2–2.5 mmol l^{-1}). Calcium in bone is in constant exchange with calcium in plasma. When calcium absorbed from the diet is insufficient to balance obligatory fecal and urinary losses, it is drawn from bone to maintain plasma levels of ionic calcium within this tightly controlled range.

Calcium Requirements

Recommended intakes for calcium for North America were recently released by the Standing Committee on the Scientific Evaluation of Dietary Reference Intakes (DRI), US National Academy of Sciences, Food and Nutrition Board, Institute of Medicine, under a program jointly commissioned by federal government departments in the USA and Canada. These revised standards encompass both recommended dietary allowances (RDA) and tolerable upper intake levels for nutrients. It should be noted

Table 2 Adequate intakes (AI) and tolerable upper intake levels (UL) for calcium by life stage group for Canada and the USA[a,b]

Age	*AI (mg per day) (1997)*	*Criteria on which AI was based*	*UL (mg per day)*
Children (male and female)			
0–6 months	210	Human milk content	nd
7–12 months	270	Human milk content + solid food	nd
1–3 years	500	Extrapolation of desirable calcium retention from 4 through 8 years	2500
4–8 years	800	Calcium accretion/ΔBMC/calcium balance	2500
9–13 years	1300	Desirable calcium retention/factorial/ΔBMC	2500
14–18 years	1300	Desirable calcium retention/factorial/ΔBMC	2500
Adult males			
19–30 years	1000	Desirable calcium retention/factorial	2500
31–50 years	1000	Calcium balance	2500
51–70 years	1200	Desirable calcium retention/factorial/ΔBMD	2500
> 70 years	1200	Extrapolation of desirable calcium retention from 51–70-year group/ΔBMD/fracture rate	2500
Adult females[c]			
19–30 years	1000	Desirable calcium retention/factorial	2500
31–50 years	1000	Calcium balance	2500
51–70 years	1200	Desirable calcium retention/factorial/ΔBMD	2500
> 70 years	1200	Extrapolation of desirable calcium retention from 51–70-year group/ΔBMD/fracture rate	2500

[a]As set in National Academy of Sciences, Institute of Medicine, Food and Nutrition Board, Standing Committee on the Scientific Evaluation of Dietary Reference Intakes (1977) *Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D and Fluoride*. Washington, DC: National Academy Press.
[b]The AI is a recommended average daily nutrient intake level based on observed or experimentally determined approximations of estimates of nutrient intake by a group (or groups) of apparently healthy people that are assumed to be adequate – used when an RDA cannot be determined. The UL is the highest average daily nutrient intake level likely to pose no risk of adverse health effects to almost all individuals in the general population. As intake increases above the UL, the potential risk of adverse effects increases.
[c]For pregnancy and lactation: no additional requirements for pregnancy and lactation; requirements are the same as for the corresponding age group.
nd, not determined due to a lack of data of adverse effects in this age group and concern with regard to lack of ability to handle excess amounts. The source of intake should be from food only to prevent high levels of intake.

that although the DRI Panel revised calcium recommendation levels upwards, the panel set adequate intake levels (AIs) rather than RDAs 'because reducing the risk of chronic disease was the intended endpoint [but] there were many uncertainties about the epidemiologic and experimental data. Other than for breast fed infants, the AI is believed to cover needs of all individuals in the group, but lack of data or uncertainty in the data prevent being able to specify with confidence the percentage of individuals covered by this intake.' The AIs for calcium were determined mainly from the intakes necessary to achieve maximal calcium retention. Table 2 list the AIs for calcium by life stage group for the USA and Canada.

Current Intakes of Calcium

Dietary Intakes

No national data on the calcium intakes of Canadians are available, although several provincial surveys have reported relatively recent data for calcium intakes (data were obtained from surveys conducted in the 1990s). Mean calcium intake data as well as percentile distributions for Canada and the USA (from the 1994 CSFII survey) are shown in Table 3. It should be noted that these data do not reflect consumption of the recently introduced orange juices with added calcium, milks with added milk solids (approx 30% more calcium), fortified soy beverages or the increased level of calcium in enriched flour. These additional sources of calcium reflect recent changes in fortification policy and industry practices in Canada and the USA since these data were collected.

Supplement Use

In the Quebec survey (1990), 8–13% of men and 15–29% of women reported taking nutritional supplements containing calcium. In the National Health Interview Survey conducted in 1986, 25% of women in the USA took supplements containing calcium compared with 14% of men and 7.5% of children.

Calcium and Osteoporosis

Definition of Osteoporosis

Osteoporosis is a disease characterized by low bone mass and microarchitectural deterioration of bone tissue, leading to enhanced bone fragility and a

Table 3 Intake of calcium from foods by age and gender for the USA and Canada (mg per day)[a]

Age	USA				Canada (data from province of Quebec)[b,c]			
	Mean	25th	50th	75th	Mean	25th	50th	75th
Infants (males and females)								
0–6 months	461	343	457	569	na			
7–12 months	725	544	703	883	na			
1–3 years	793	599	766	957	na			
4–8 years	838	649	808	993	na			
Males								
9–13 years	1025	756	980	1245	na			
14–18 years	1169	834	1094	1422	na			
19–30 years	1013	729	954	1232	1114	744	984	1227
31–50 years	912	651	857	1112	922	613	769	961
51–70 years	748	545	708	908	736	575	701	845
$\geq$ 70 years	748	548	702	880	771	476	645	904
Females								
9–13 years	919	705	889	1100	na			
14–18 years	753	541	713	922	na			
19–30 years	647	464	612	792	788	530	688	831
31–50 years	637	461	606	779	658	407	555	770
51–70 years	599	441	571	727	622	409	539	730
$\geq$ 70 years	536	407	517	644	574	394	524	660

[a]From 1994 *USDA Continuing Survey of Food Intakes by Individuals*, as reported by the Institute of Medicine, *Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D and Fluoride*, 1997.
[b]From Santé Québec, *Rapport de l'enquete quÕbecoise sur la nutrition*, 1995 (infants, children and adolescents not assessed).
[c]Age groups do not exactly match CSFII groupings.
na, not applicable.

consequent increase in fracture risk. The World Health Organization expert committee on osteoporosis has defined four different diagnostic categories: *normal* if BMD or bone mineral content (BMC) is within one standard deviation (SD) of young adult values; *osteopenia* (low bone mass) if the BMD or BMC is between 1 and 2.5 SD below young adult values; *osteoporosis* if the BMD or BMC is 2.5 SD or more below young adult values; and *severe osteoporosis* if the individual is osteoporotic and has suffered at least one fragility fracture.

Bone Structure

Bone tissue is a matrix of collagen and other proteins into which minerals are deposited, calcium and phosphorus being the most abundant. Calcium in bone is primarily (>90%) in the form of hydroxyapatite [$Ca_{10}(PO_4)_6(OH)_2$], and amorphous calcium phosphate, which lacks a coherent structure, found in areas of active bone formation. Bone is composed of two different types of structural tissue: cortical (the outer, denser envelope which plays a major structural support function) and trabecular or cancellous tissue (the metabolically more active form which is structurally like a fine sponge with numerous small voids). Cortical bone is found primarily in the appendicular skeleton (bones of the limbs), whereas trabecular bone is found mainly in the axial skeleton (skull and vertebral bones). The total skeleton is composed of approximately 80% cortical bone and 20% trabecular bone.

Bone contains two types of cells, osteoblasts and osteoclasts. The osteoblasts secrete an organic matrix composed largely of collagen, which is then hardened by hydroxyapatite. Osteoclasts are active at numerous sites within bone, secreting enzymes that dissolve bone before it is reformed. Bone remodeling, or turnover, is the cycle of bone breakdown by osteoclasts and bone rebuilding by osteoblasts and continues throughout life. In the normal young state, total osteoblast activity exceeds osteoclast activity, resulting in bone accretion.

Risk Factors for Osteoporosis

Osteoporotic fractures are multifactorial and may result from a variety of causes, either singly or in combination. For this reason, no single risk factor can identify all potential fracture cases. Peak bone mass is the major factor determining the risk of developing osteoporosis. By about age 20, the human skeleton reaches 90–95% of its peak bone mass. Then, over the next 10 years, the final 5–10% of bone mineral is added. People who have achieved a greater peak bone mass are less susceptible to osteoporosis. Numerous factors have been identified as the major risks for osteoporosis, including low calcium intakes and other dietary factors, BMD, estrogen status, as well as use of alcohol and some drugs (see Table 4).

Table 4 Major risk factors for osteoporosis

Low bone mineral density; BMD is the most useful risk factor for stratifying people by level of fracture risk
Estrogen deficiency due to menopause (natural or surgical), ovulatory disturbances among premenopausal women
Inadequate dietary calcium and vitamin D
Excessive intakes of caffeine
Excess dietary sodium intake
Alcohol abuse
Extended periods of immobility
Long-term use of corticosteroids
Other medications (heparin, anticonvulsants, excess thyroid hormone, and others)
Hereditary factors

Adapted from Ross PD (1996) Osteoporosis. Frequency, consequences, and risk factors. *Archives of Internal Medicine* 156: 1399–1411.

Physiologic Relationship Between Dietary Calcium and Osteoporosis

While it seems reasonable to predict that calcium intake and bone loss would be linked, the evidence showing this relationship has been less than obvious. Studies comparing calcium intakes and osteoporosis incidence or fracture rates among different countries with varying cultures tend to show no relationship or a negative association between calcium intakes and osteoporosis. However, within countries or population groups, a large body of evidence has accumulated that shows that higher calcium intakes at various times in the life cycle are linked with increased BMD or BMC. Controlled clinical trials of calcium supplementation have been conducted for almost every age group (children and older) and, for the most part, support the benefits of calcium for increasing BMD or BMC, or in some groups reducing losses. Fewer studies have investigated the impact of calcium on osteoporosis or fracture rates.

In children, clinical trials have shown a modest but positive effect of calcium supplementation, particularly in those children with intakes < 1000 mg per day, on bone mineral accretion. In general, supplementation resulted in 1–5% greater gains in bone mineral density or bone mineral content compared with controls. However, the long-term benefits of such an increase and whether the increase is sustained remain unclear at the present time.

Data concerning the role of calcium in bone health of young adults are particularly lacking compared with other age groups. Physical activity appears to be a significant determinant of bone health in this age group.

Studies show several consistent effects regarding the role of calcium in bone loss in postmenopausal women. Early postmenopausal women are less responsive to calcium supplementation than late postmenopausal women; where effects were seen, they tend to be in cortical bone, with spine being less responsive to calcium. Late menopausal women with low calcium intakes tend to gain more BMD from calcium supplementation than do women with higher usual intakes of calcium. Observational studies in postmenopausal women and one study that included men generally indicated a positive effect of calcium on bone density. Several studies also suggested that higher intakes of calcium earlier in life were related to reduced incidence or fractures or increased BMD in postmenopausal women.

Many studies conducted in the elderly have shown a benefit of supplements or higher intakes of calcium on the clinically important outcome – fracture rate. Almost half of the trials in the elderly found decreased fracture incidence, in addition to changes in BMD. More studies in this age group than in any other have included male subjects, and benefits appeared to apply equally to men. In the elderly, most studies have provided a supplement of vitamin D along with calcium. Given that vitamin D deficiency is most prevalent among the elderly, particularly among institutionalized or house bound individuals, it is important that the elderly have adequate vitamin D in order to utilize calcium or to benefit from additional calcium.

See also: **Bioavailability of Nutrients**; **Bone**; **Calcium**: Properties and Determination; **Dairy Products – Nutritional Contribution**; **Dietary Requirements of Adults**; **Hormones**: Thyroid Hormones; Pituitary Hormones; **Milk**: Dietary Importance; **Minerals – Dietary Importance**; **Osteoporosis**; **Phosphorus**: Physiology

Further Reading

Brazier M, Kamel S, Maamer M *et al*. (1995) Markers of bone remodeling in the elderly subject: effects of vitamin D insufficiency and its correction. *Journal of Bone Mineral Research* 10: 1753–1761.

Gibson RS (1990) *Principles of Nutritional Assessment*. New York: Oxford University Press.

Lobaugh B (1996) Chapter 3: Blood calcium and phosphorus regulation. In: Anderson JJB and Garner SC (eds) *Calcium and Phosphorus in Health and Disease*, pp. 27–43. New York: CRC Press.

National Academy of Sciences, Institute of Medicine, Food and Nutrition Board, Standing Committee on the Scientific Evaluation of Dietary Reference Intakes (1977) *Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D and Fluoride*. Washington, DC: National Academy Press.

Packard PT and Heaney RP (1997) Medical nutrition therapy for patients with osteoporosis. *Journal of the American Dietetic Association* 97: 414–417.

Ross PD (1996) Osteoporosis. Frequency, consequences, and risk factors. *Archives of Internal Medicine* 156: 1399–1411.

Wardlaw GM (1996) Putting body weight and osteoporosis into perspective. *American Journal of Clinical Nutrition* 63: 433S–436S.

Wood RJ (2000) Chapter 28: Calcium and phosphorus. In: Stipanuk M (ed.) *Biochemical and Physiological Aspects of Human Nutrition*, pp. 643–670. Philadelphia, PA: W.B. Saunders.

World Health Organization (1994) Assessment of fracture risk and its application to screening for postmenopausal osteoporosis. Report of a WHO Study Group. *World Health Organization Technical Report Series* 843: 1–129.

Calorific Values *See* **Energy**: Measurement of Food Energy; Intake and Energy Requirements; Measurement of Energy Expenditure; Energy Expenditure and Energy Balance

CAMPYLOBACTER

Contents

Properties and Occurrence

M B Skirrow, Gloucestershire Royal Hospital, Gloucester, UK

Introduction

Campylobacters are the leading cause of infective diarrhea in most industrialized countries. Yet, they have hitherto enjoyed a low public profile that belies their importance as a cause of foodborne disease. There are several reasons for this. First, their role as enteric pathogens was not discovered until the late 1970s; second, infection is rarely fatal; and third, they seldom cause spectacular outbreaks that catch the attention of the media. Yet, they cause much morbidity and constitute a huge economic burden to society.

After a brief description of campylobacters and their properties, this article describes their wide distribution in wild and domestic animals, notably birds, and shows how they are introduced into the food chain from these sources. It attempts to define the foods of greatest importance as a cause of human infection, and the circumstances that allow infection to arise. Although there must be many sources and modes of transmission that remain hidden, there can be no doubt that poultry are the most important single source. It has been estimated that poultry account for about half of all recorded cases.

Classification

Campylobacters are small, spirally curved, flagellated, microaerophilic Gram-negative bacteria that were formerly classified in the genus *Vibrio*. Together with *Arcobacter*, *Helicobacter*, and *Wolinella*, they form a phylogenetic group distantly related to other eubacteria. Several species are associated with human disease, but only *Campylobacter jejuni* ssp. *jejuni* and *C. coli* are of major importance as a cause of campylobacter enteritis. *C. upsaliensis* and *C. jejuni* ssp. *doylei* are associated with diarrheal disease in children in developing countries, but to only a minor extent in industrialized countries. *C. lari* is present in abundance in birds (notably gulls) and natural water, but most strains appear to be nonpathogenic, as they seldom cause human disease.

C. fetus, the type species of the genus, is an uncommon cause of systemic (bloodstream) infection in people with impaired immunity. Although *C. fetus* infection has been caused by eating raw lambs' liver, its low virulence for healthy people makes it of minor concern to the food microbiologist.

Arcobacters were initially regarded as aerotolerant *Campylobacter* species (unlike campylobacters, they grow freely in air), but in 1991, the new genus *Arcobacter* was created to accommodate them. Although they cause abortion in pigs and cattle and are commonly found in poultry (24–96% of fresh carcasses) and meat (80% of ground pork samples in one study), arcobacters (usually *A. butzleri*) have only occasionally caused intestinal infection in humans. Until more is known about their role in human disease, arcobacters remain a research topic. Our main concern is with *C. jejuni* and *C. coli*.

Helicobacter pylori, which colonizes the stomach of roughly half of the world population and causes gastric and peptic ulceration, can probably be transmitted via food and water. Evidence is currently indirect and inconclusive, but progress needs to be monitored.

Growth Factors and Properties

Campylobacters are strictly microaerophilic bacteria that require oxygen for growth yet are poisoned by concentrations much above 10%. Unlike most other campylobacters, *C. jejuni*, *C. coli*, and *C. lari* have an optimum growth temperature of 42–43 °C, a feature that has earned them the informal title of 'thermophilic campylobacters.' They do not grow below a temperature of 28 °C.

Campylobacters are more sensitive to physical and chemical agents than enterobacteria such as *Escherichia coli*. Dense saline suspensions (10^{10} cfu ml^{-1}) dried on hard surfaces survive for 2–10 h at 37 °C, but suspensions in broth or skimmed milk dried and held at 4 °C survive for several weeks. Campylobacters have a pH growth range of 6.0–8.0, but they are inactivated below pH 5.5 or above pH 9.0. Hypochlorites, phenols, iodophors, and quaternary ammonium compounds kill campylobacters within 1 min at standard working dilutions.

These factors combine to prevent their multiplication in food. Unlike salmonellas and other enterobacteria, they tend to die in food stored at ambient temperatures, owing to exposure to air and overgrowth by other bacteria. Their survival is prolonged by refrigeration.

Occurrence in the Environment

C. jejuni and *C. coli* live mainly as commensals in the intestinal tract of a wide variety of warm-blooded animals, notably birds (Table 1). Carriage rates in wild mammals are generally lower than in birds, but a rate of 24% has been found in urban rats (*Rattus norvegicus*). Domestic and food-producing animals are commonly colonized and constitute the main source of human infection. Poultry are a particularly important source (see below). *C. jejuni* is generally the predominant species, but in pigs, virtually all the thermophilic campylobacters are *C. coli*. Horses are seldom colonized.

Table 1 Occurrence of campylobacters (*C. jejuni* and *C. coli*) in wild birds from five continents[a]

Ducks and geese (Anatidae)	37 (7)
Gulls (*Larus* spp.)	34 (8)
Pigeons (mainly *Columba livia*)	25 (6)
Sparrows (*Passer* spp.)	18 (5)
Starling (*Sturnus* spp.)	47 (2)
Crows (Corvidae)	55 (4)

Also isolated from cormorant (*Phalacrocorax olivaceus*), caracara (*Milvago chimango*), sandhill crane (*Grus canadensis*), waders (Haematopodidae, Charadriidae, Scolopacidae), puffin (*Fratula arctica*), owls (Strigidae).

[a]The data are given as the mean percent positive, with the number of surveys from which the data were extracted in parentheses.

The fecal shedding of campylobacters by domestic and wild animals results in widespread environmental contamination, notably surface water, even in remote areas. Campylobacters survive well at temperatures below 15 °C in fresh water, sea water, and sewage. In these conditions, they can adapt to aerobic metabolism, which is presumably a defense strategy. Surface water is probably the chief vehicle of infection for farm animals, and drinking untreated water carries a high risk of infection for human beings.

Occurrence in Foods

Although carriage rates in food-producing animals vary widely, both seasonally and between herds or flocks, average figures compiled from numerous surveys give an idea of the high prevalence of these bacteria in poultry and animals produced for red meat.

Table 2 shows the occurrence of campylobacters in eviscerated poultry, in poultry processing plants and

Table 2 Occurrence of campylobacters (*C. jejuni* and *C. coli*) in poultry[a]

		Retailed product	
	Birds at processing plant	*Fresh*	*Frozen*
Broiler chickens	66 (22)	62 (32)	47 (9)
Chicken livers	76 (3)	48 (3)	9 (2)
Turkeys	48 (7)	22 (3)	36 (2)
Ducks	58 (5)		
Geese		38 (1)	
Game birds		14 (2)	

[a]The data are given as the mean percent positive with the number of studies from which the data were extracted in parentheses.

Table 3 Occurrence of campylobacters (*C. jejuni* and *C. coli*) in red meats and their source animals[a]

Product	Source animal	Carcass before chilling	Retailed product	
			Fresh	Frozen
Cattle/beef	25 (14)	12 (3)	2.9 (11)	0 (1)
Sheep/lamb	30 (9)	11 (4)	5.8 (5)	2.0 (1)
Pigs/pork	82 (14)	27 (7)	3.4 (9)	0 (1)
Offal[b]			22 (8)	1.8 (3)
Miscellaneous red meats			1.3 (1)	
Cooked meats			2.3 (1)	

[a]The data are given as the mean percent positive with the number of studies from which the data were extracted in parentheses.
[b]Mostly liver, kidney, and heart from cattle, sheep, and pigs.

Table 4 Occurrence of campylobacters (*C. jejuni* and *C. coli*) in milk and miscellaneous foods[a]

Raw cows' milk (mainly bulked)	4.8 (9)
Seafood[b]	9.7 (6)
Mushrooms	1.5 (1)
Salads (unsealed)	0 (1)
Salads (MAP[c])	22 (1)
Fresh vegetables[d]	2.3 (1)

[a]The data are given as the mean percent positive with the number of studies from which the data were extracted in parentheses.
[b]Includes oysters, mussels, scallops, cockles, and shrimps.
[c]Modified-atmosphere-packaged.
[d]From farmers' markets; samples from supermarkets were negative.

at the point of retail sale. The initial high contamination rates persist in the retailed product, with only a modest fall brought about by freezing and thawing. This high persistence is not surprising, considering that cecal contents of chickens may contain as many as 10^8 campylobacters per gram. Cuts of chicken with skin are significantly more contaminated than those without skin.

Table 3 shows the occurrence of campylobacters in cattle, sheep, and pigs together with the meat products derived from them. Butchered carcasses in the abattoir reflect the high intestinal colonization rates in the live animals, but in contrast to poultry, there is a sharp reduction of campylobacters between the abattoir and the retailed product. This is largely due to the forced air chilling process used on these large carcasses (see below). Thus, contamination rates on red meat are generally low, as are the campylobacter counts on the positive samples. Note that offal, which is not subjected to the same treatment, remains more contaminated.

Table 4 shows the occurrence of campylobacters in milk and other foods. Raw milk is clearly a major potential source of infection, and there are many instances of campylobacter enteritis resulting from its consumption. Pasteurization or other conventional heat treatment renders it safe from this threat.

Food Poisoning from Campylobacters

The proportion of campylobacter infections attributable to the consumption of food is not precisely known, but it is believed to be the great majority. A few infections are acquired through direct contact with infected animals or their products. These are mostly occupational, for example, in workers in poultry-processing plants, but some arise through contact with domestic pets, notably puppies and kittens that are themselves suffering from campylobacter diarrhea. Animal contact is more important in developing countries, where transmission rates are far higher than in industrialized countries, and the disease is virtually confined to young children.

Most campylobacter infections are sporadic. Unlike salmonella infection, outbreaks of campylobacter enteritis are uncommon. In the UK, for every campylobacter outbreak reported, there are 34 salmonella outbreaks. Large campylobacter outbreaks are mostly milkborne (see below) or waterborne. Five waterborne outbreaks, affecting a total of nearly 7000 people, have arisen from the distribution of contaminated supplies, including unchlorinated water from oligotrophic lakes in Scandinavia.

Incidence

As diagnosis depends on the identification of campylobacters in patients' feces, measurement of incidence depends on laboratory reports. This in turn is dependent on the availability and use of medical and laboratory services, which are highly variable. Laboratory reports give incidences of 100/100 000 per year in the UK, 25/100 000 in the USA, and 318/100 000 in New Zealand. However, extrapolations from community surveys and outbreaks give figures of 1100/100 000 and 1000/100 000 for the UK and USA, respectively. The costs in terms of health care and lost productivity are immense. World Health statistics for foodborne

campylobacter enteritis in the USA cite estimates of US$ 1.2–6.6 billion annually.

Foods Implicated in Campylobacter Enteritis

Having identified the foods that are regularly contaminated with campylobacters (Tables 2–4), we need to define which of them are actually implicated as sources of infection. This is not easy. As already mentioned, outbreaks of campylobacter enteritis, which provide good opportunities to pinpoint sources, are uncommon. There are several approaches to the problem:

1. Comparison of the campylobacter strains commonly found in foods with strains found in patients. For example, strain identification tells us that pigs are a relatively minor source of campylobacter enteritis, as almost all strains in pigs are *C. coli*, yet in most regions, *C. coli* accounts for only about 5% of human infection. The common serotypes of *C. jejuni* found in man are well represented in poultry, cattle, and sheep. In the UK, about 70% of the strains most commonly isolated from patients are also found in chickens. (*See* **Essential Oils**: Isolation and Production; **Pork**.)
2. Identification of foods that have been implicated in outbreaks of campylobacter enteritis. These foods are listed in Table 5, from which it can be seen that poultry feature prominently. The locations of these outbreaks are shown in Table 6.
3. Identification of foods that are statistically associated with infection in case-control studies. Such studies provide two valuable indices: the relative risk of consuming or handling a food (usually expressed as the odds ratio); and the etiological fraction, i.e., the proportion of infections attributable to that food. Table 7 shows the foods or activities that have been associated with a risk of acquiring campylobacter infection. Again the consumption of chickens, especially if they are undercooked, stands out as a prominent risk factor.
4. Observation of the effects of eliminating or reducing campylobacters from a food on the incidence of infection in the community that consumes the food. In a UK study, the control of *C. jejuni* serotype HS4:HL1 in a heavily colonized chicken farm was followed by a fall in the proportion of human infections with this serotype from over 30% to 5% in the community supplied with the chickens. Likewise, in Tasmania, the escalation of a salmonella control program in chickens, which incidentally would have controlled campylobacters, was followed by a 50% reduction in human campylobacter infections.

The role that these various foods play in causing campylobacter infection will now be considered in more detail.

Poultry Poultry feature prominently in all of the above categories, especially if undercooked. It is notable that any meat cooked by barbecue, char-grill, stir-fry, or fondue methods feature in these lists. These hot rapid methods of cooking are liable to leave the deeper parts of meat uncooked, even when the outside is well browned. Handling and preparing chicken were found to carry a much higher risk of infection than eating chicken, in a case-control study in the USA.

Broiler chickens are consumed in vast numbers, so even though the relative risk (odds ratio) of eating chicken is only around 3 (Table 7), the total number of cases resulting from their handling and consumption is large. Extrapolating case-control data to estimate the proportion of sporadic infections attributable to poultry must be viewed with caution, but two studies, one in the USA and the other in the UK, came up with similar figures, namely 48 and 50%, respectively.

Table 5 Foods other than milk implicated or suspected as the source in 83 outbreaks of campylobacter enteritis (average attack rate 41%)[a]

						Miscellaneous		
Poultry		*Red meat*		*Seafood*		*Salads*	*foods*	*Specific meal*
Chicken	8	Raw beef	2	Raw clams	1	12[b]	7[c]	36
Undercooked chicken	7[d]	Vinegared pork	1	Prawn, salmon vol-au-vent	1			
Chicken liver	2[e]	Red meat products	3					
Turkey	2[f]							

[a]The figures shown are the numbers of outbreaks reported in publications from 11 countries.
[b]Evidence of cross-contamination from raw chicken in 7.
[c]Includes pâté vol-au-vent (1), tuna/egg vol-au-vent (1), mayonnaise (1), frozen egg (1), undercooked egg (1), cake icing (1).
[d]Barbecued (2), char-grilled (1), stir-fried (1).
[e]Fondued (1), pâté (1).
[f]Processed and sliced (1), boned, stuffed, and rolled (1).

Table 6 Places where 67 outbreaks of foodborne campylobacter enteritis arose (excluding milkborne outbreaks)

Place	*Number of outbreaks*
Restaurant, hotel	34
Institution (e.g., colleges, military barracks)	13[a]
School (residential and nonresidential)	8
Private home or function	4
Home for aged	3
Outdoor camps	3
General community	2

[a]Includes one hospital.

Table 7 Foods identified as carrying a risk for acquiring sporadic campylobacter enteritis, as determined by case-control studies

Food	*Number of studies*	*Odds ratio, mean (range)*
Chicken (any)	8	2.6 (0.44[a]–5.2)
Chicken (raw/undercooked)	5	6.0 (2.7–9.0)
Chicken eaten in restaurants	1	3.8
Poultry liver	1	5.7
Game hens	1	3.3
Barbecued sausages	1	7.6
Raw/rare fish	1	4.0
Raw/rare shellfish	1	1.5
Mushrooms	1	1.5
Raw (unpasteurized) milk	5	5.8 (3.9–9.3)
Raw (unpasteurized) cream	1	12.0
Milk from bird-pecked bottles	1	42.1

[a]This odds ratio of 0.44 ostensibly shows a protective effect from eating chicken. A possible explanation for this exceptional result is that many of these chicken eaters habitually consumed chicken and had thereby developed immunity to infection.

Milk and Dairy Products Raw milk commonly contains campylobacters (Table 4). The distribution of inadequately pasteurized milk was responsible for the largest outbreak of campylobacter enteritis on record. In 1979, in England, an estimated 3500 people, mostly primary-school children, were affected over a period of 3 weeks. Since then, at least 58 milkborne outbreaks have been reported in the UK and 26 in the USA, affecting over 8000 people. The average attack rate in exposed persons in 14 outbreaks in which data were available was 54% (range 24–79%).

In the USA, milkborne outbreaks accounted for 50% of all reported foodborne campylobacter outbreaks. Roughly one-half the milkborne outbreaks arose in the general community, 20% in schools or other institutions, 17% on school visits to farms, and 8% in holiday or rural camps. In all but two instances, the implicated milk was drunk raw; in the two exceptions, the milk was distributed as pasteurized, but there had been unsuspected processing failures. It is notable that four of 10 people who drank their raw milk in coffee became infected.

Most milkborne infections are associated with cows' milk, but raw goats' milk has occasionally been implicated. In a survey in the UK, only one out of 2493 (0.04%) goat's milk samples in the UK was found to contain *C. jejuni*.

Cheese Most cheeses are too acidic for the survival of campylobacters, but a soft cheese made from unpasteurized sheep's milk was incriminated in an outbreak of campylobacter enteritis in Czechoslovakia, and there is a single instance of infection associated with home-made ricotta cheese prepared from unpasteurized goats' milk.

In several case-control studies, the consumption of dairy products was associated with a reduced rate of infection.

Eggs The various egg-containing foods listed as suspect sources of infection in Table 5 may have been cross-contaminated from other foods. However, shell eggs may become externally contaminated with campylobacters from chicken excreta and may thus contaminate egg contents when broken. Campylobacters do not penetrate intact shell or membranes into the egg contents. Vertical transmission to eggs from laying hens has not been proven, but there is indirect evidence to support it.

Seafood Raw seafood may harbor campylobacters (Table 4) as well as other pathogens. In temperate zones, contamination may be highly seasonal – high in winter and low in summer. A study in Australia showed that Sydney Rock Oysters (*Crassostrea commercialis*) allowed to feed in waters containing about 10^4 cfu campylobacters per milliliter concentrated them to 10^2–10^3 cfu g^{-1} in their tissues within 1 h, but they were effectively cleared by depuration for 48 h. Yet, depuration does not always eliminate campylobacters. A high proportion of mussels have been found to harbor campylobacters, but conventional steaming completely inactivates them. Seafood kept in sea water may be contaminated from the water rather than containing campylobacters itself.

Campylobacters survive for several months in frozen oysters and for 8–14 days in refrigerated oysters. (*See* **Shellfish**: Contamination and Spoilage of Molluscs and Crustaceans.)

Introduction of Campylobacters into Food

Contamination of Meat at Slaughter

The contamination of animal carcasses with intestinal contents at slaughter and the persistence of the bacteria in raw meat are the most important sources

of campylobacters in food. Because these are inextricably bound with processing methods, they are considered below in the section on the effects of food processing on campylobacters. (*See* **Meat**: Slaughter.)

Contamination of Milk from Milk-producing Animals

At any one time, an average of about 25% of milking cows excrete campylobacters in their feces. Carriage in individual animals is variable, and several strains may circulate in a herd over a period. Some degree of fecal contamination is unavoidable, even in well-run milking parlors, and several outbreaks have been caused by drinking top-grade certified milk in the USA. Table 4 shows that, on average, 4.8% of bulked milk contains campylobacters. Positive samples usually have high counts of *E. coli*, but occasionally, campylobacters have been found without *E. coli*, probably through excretion from the udder. High counts of *C. jejuni* have been found in milk from cows with campylobacter mastitis, an uncommon naturally occurring infection. In experimental mastitis, many campylobacters may be excreted from an infected quarter before the milk appears obviously granular.

Contamination of Milk by Wild Birds

In the UK, where it is common practice for fresh milk to be delivered to the doorstep in metal-foil-topped bottles, magpies (*Pica pica*) and jackdaws (*Corvus monedula*) have developed the habit of pecking through the foil tops and contaminating the contents. This is a seasonal activity when the birds are raising fledglings (April to June), which is also a time when they frequently probe cow dung for invertebrates. The habit is geographically variable; in areas of highest prevalence, case-control studies have shown it to account for a high proportion of campylobacter infections at that time of year.

Cross-contamination in the Kitchen

Items of food, such as raw poultry that are heavily contaminated with campylobacters, provide a reservoir of bacteria in the kitchen from which other foods can become contaminated. This may come about through direct contact (e.g., from the liquor of a thawing chicken), or via common utensils, counter tops, chopping boards, and the hands of kitchen staff. Campylobacters have been isolated from all of these sites in working kitchens and from 76% of the hands of staff who had held contaminated chicken.

A frequent finding in campylobacter outbreak investigations is a failure to separate raw meats from other foods in food preparation areas – and often a disturbing ignorance on the part of the staff of the need for this separation. Although it is seldom possible to prove cross-contamination, it is probably the most frequent event leading to infection.

Insects Theoretically insects could be passive vectors of infection via food. Campylobacters have been isolated from houseflies (2.4% of flies caught in houses and gardens in the UK; 5.5% in Pakistan) and cockroaches, but substantial carriage has been found only in the close vicinity of chicken sheds and piggeries. Campylobacters failed to survive for more than 2 days in flies artificially fed with campylobacter broth cultures. (*See* **Insect Pests**: Problems Caused by Insects and Mites.)

Fate of Campylobacters During Food Processing

Animal Slaughter and Meat Processing

Contamination of animal carcasses with intestinal contents during slaughter and preparation is universal and, to some degree, unavoidable. Fortunately, in the case of large animals (cattle, sheep, pigs), conventional forced air evaporative chilling with frequent mechanical spraying greatly reduces the numbers of live campylobacters remaining on the surfaces of carcasses, mainly through the effect of drying. Cooling without mechanical assistance causes only about a 50% reduction in campylobacter numbers.

Broiler chickens and other poultry pose a greater problem owing to the high degree of mechanization required for processing birds at rates that satisfy public demand. Two stages of processing are particularly problematic:

1. Mechanical defeathering, which is preceded by immersion in hot water ('scalding') to facilitate plucking. 'Hard scalding' at temperatures above 58 °C causes a reduction in bacterial counts, but 'soft scalding' at 50–52 °C, used for birds that are to be sold fresh, has little effect; thus, leakage of intestinal contents, which is common, results in gross cross-contamination.
2. Mechanical evisceration, which often results in the rupture of the gut of birds that are not of consistent size. Again, widespread cross-contamination is the inevitable consequence; birds that are initially clean become contaminated during processing. Contamination can be reduced, but not eliminated, by chilling in well-designed reverse-flow systems containing chlorinated water (10 p.p.m. chlorine). Uneviscerated ('New York dressed') birds are the most heavily contaminated; campylobacter

counts of up to $2.4 \times 7\,10^7$ per chicken have been recorded. Frozen birds are least contaminated, owing to the damaging effect on the bacteria of freezing and thawing.

Effects of Storage

The most critical factors in the survival of campylobacters are temperature and oxygen concentration: cold and reduction of oxygen tension are protective. Campylobacter counts have been shown to decline on stored meat by 1–2 $\log_{10}$ at 4 °C and 2–5 $\log_{10}$ at 23 °C over 2 weeks, but much depends on the particular campylobacter strain and the type of meat – minced beef, for example, is unusually protective, probably on account of its reducing properties. Reduced conditions, such as are present in vacuum or modified-atmosphere-packaged (MAP) foods, are certainly protective. In one study, campylobacters were detected in 22% of MAP salads from supermarkets (Table 4), although counts were mostly below 80 cfu g^{-1}. Freezing and thawing cause a 1–2 $\log_{10}$ reduction in campylobacter counts, but once frozen, survival is measured in months or even years. (*See* **Storage Stability**: Mechanisms of Degradation.)

Salting

C. jejuni and *C. coli* do not grow in the presence of salt at concentrations greater than 1.5%. They can survive for 3–5 days at room temperature in 4.5% NaCl and for 3 weeks at 4 °C in 6.5% NaCl, albeit with a several $\log_{10}$ fall in viable counts. However, the addition of 2.5% NaCl to aerobically packaged turkey hams reduced *C. jejuni* counts from about 6.0 $\log_{10}$ cfu g^{-1} to below detectable limits within 6 days at 4 °C. Campylobacters can survive on washed pig intestines after overnight salting.

Acidification

Campylobacters cannot grow at pH values of 5.5 or less and are inactivated at pH 5.0 or below.

Irradiation

Campylobacters are considerably more sensitive than *E. coli* to gamma (D_{10} values 0.12–0.25 kGy) and ultraviolet irradiation, so they are destroyed by conventional irradiation regimens. A gamma irradiation dose of 1.5 kGy is sufficient to cause a 7 $\log_{10}$ reduction in *C. jejuni*. (*See* **Irradiation of Foods**: Basic Principles.)

Survival in Milk, Dairy Products, Spices, and Fruit

Milk Most campylobacter strains are recoverable from raw milk held at 4 °C for 7 days if the initial concentration is heavy, e.g., 10^7 cells ml^{-1}. Inactivation of strains parallels a rise in the aerobic plate count and fall in pH. Thus, survival in raw milk is about half that in sterilized milk. The lactoperoxidase system also reduces survival in milk. Survival becomes progressively shorter as the temperature is increased, so that at 20 °C, it is only 1–2 days. Campylobacters are readily destroyed by conventional pasteurization and give *D*-values of 1 min or less at 60 °C in raw, pasteurized, or skimmed cows' milk.

Cheese and Yogurt In one study, campylobacters were not recovered from cheddar cheese curd after 30 days of curing, or from whey or curd of cottage cheese after cooking for 30 min at 55 °C. In another study, no strain survived for more than 25 min in yogurt. (*See* **Yogurt**: The Product and its Manufacture.)

Egg Campylobacters can grow in egg yolk and yolk–albumen melanges, but albumen alone is toxic. The major factor in the sensitivity of egg white is apparently conalbumen.

Spices and Fruit The spices oregano, sage, and ground cloves are mildly inhibitory to *C. jejuni*, but not sufficiently to reduce numbers in refrigerated food. A study showed that *C. jejuni* survived for 6 h on cubes of water melon and papaya at 25–29 °C (8–62% survivors), but much less readily when lemon juice was added to the surface of the cubes (0–14% survivors).

See also: ***Campylobacter***: Detection; Campylobacteriosis; **Contamination of Food**; **Eggs**: Microbiology; **Fish**: Spoilage of Seafood; **Food Poisoning**: Classification; **Insect Pests**: Problems Caused by Insects and Mites; **Irradiation of Foods**: Basic Principles; **Meat**: Slaughter; **Milk**: Processing of Liquid Milk; Analysis; ***Salmonella***: Salmonellosis; **Storage Stability**: Mechanisms of Degradation; **Yogurt**: The Product and its Manufacture

Further Reading

Advisory Committee on the Microbiological Safety of Food (1993) *Interim Report on Campylobacter*. London: HMSO.

Altekruse SF, Stern NJ, Fields PI and Swerdlow DL (1999) *Campylobacter jejuni* – an emerging foodborne pathogen. *Emerging Infectious Diseases* 5: 28–35.

Busby JC and Roberts T (1997) Economic costs and trade impacts of microbial foodborne illness. *World Health Statistics Quarterly* 50: 57–66.

Friedman CR, Neimann J, Wegener HC and Tauxe RV (2000) Epidemiology of *Campylobacter jejuni* infections in the United States and other industrialized nations. In: Nachamkin I and Blaser MJ (eds) *Campylobacter*, 2nd edn. pp. 121–138. Washington, DC: American Society for Microbiology.

Jacobs-Reitsma W (2000) *Campylobacter* in the food supply. In: Nachamkin I and Blaser MJ (eds) *Campylobacter*, 2nd edn. pp. 467–481. Washington, DC: American Society for Microbiology.

Nachamkin I and Skirrow MB (1998) *Campylobacter, Arcobacter*, and *Helicobacter*. In: Collier L, Balows A and Sussman M (eds) *Topley and Wilson's Microbiology and Microbial Infections*, 9th edn. vol. 2, pp. 1237–1256. London: Edward Arnold.

Skirrow MB (1998) Infection with *Campylobacter* and *Arcobacter*. In: Collier L, Balows A and Sussman M (eds) *Topley and Wilson's Microbiology and Microbial Infections*, 9th edn. vol. 3, pp. 567–580. London: Edward Arnold.

Solomon EB and Hoover DG (1999) *Campylobacter jejuni*: a bacterial paradox. *Journal of Food Safety* 19: 121–136.

Todd ECD (1997) Epidemiology of foodborne diseases: a worldwide review. *World Health Statistics Quarterly* 50: 30–50.

Wesley IV (1996) *Helicobacter* and *Arcobacter* species: risks for foods and beverages. *Journal of Food Protection* 59: 1127–1132.

Detection

M B Skirrow, Gloucestershire Royal Hospital, Gloucester, UK

Introduction

The conventional means of detecting campylobacters is by isolation in culture. In this article their microaerobic growth requirements and special needs for isolation from foods are described, including selective enrichment and plating on selective agar. Direct detection of campylobacters by immunological and DNA probe techniques are becoming more sensitive and specific by the application of polymerase chain reaction (PCR) methods, but they are as yet too complex to be adopted routinely. Campylobacters can be provisionally identified by phenotypic methods in most laboratories, but definitive identification requires molecular methods that are available only in reference laboratories. Strain typing of campylobacters (serotyping, phage typing, molecular typing) is also best performed in reference laboratories. Advances in molecular typing in particular are paving the way for better source tracing and understanding of the epidemiology of campylobacter infection.

Principles of Campylobacter Cultivation

The failure of traditional cultural methods to provide the conditions necessary for the growth of campylobacters was the main reason why their role in human disease remained hidden for so long. The main peculiarity of campylobacters is that they are strictly microaerophilic and require specially prepared atmospheres for satisfactory growth. Specific selection, either by means of selective media or filtration, is essential for their isolation from feces or other specimens containing mixed organisms. Moreover, they only grow well on supplemented media, and *Campylobacter jejuni, C. coli*, and *C. lari* – the so-called thermophilic group that causes campylobacter enteritis – have an optimum growth temperature of 42–43 °C.

Oxygen Sensitivity

Oxygen sensitivity is the most critical factor in the cultivation of campylobacters. Campylobacters have a respiratory type of metabolism based on the krebs cycle, so oxygen is necessary for growth; yet oxygen at atmospheric pressures is toxic. This toxicity is due to superoxides and free radicals that form in unsupplemented media exposed to air, particularly in the light. Campylobacters are especially vulnerable to these compounds despite being able to produce superoxide dismutase and catalase. They are most vulnerable in the resting phase; once they have started to grow they become more oxygen-tolerant, probably owing to the production of nicotinamide adenine dinucleotide (NAD) or nicotinamide adenine dinucleotide phosphate (NADP), which forms a physiological barrier that protects oxygen-sensitive enzymes. The practical points that arise from this are twofold:

1. Heavy inocula are required to initiate growth on unsupplemented media. Large populations usually have a few bacteria that are able to start growing and thereby create conditions suitable for other cells to grow. The result of seeding campylobacters on unsupplemented media tends to be either confluent growth or none at all.
2. The key supplements for growing campylobacters are agents that neutralize superoxides and free radicals; any nutrient value they have is incidental. Blood, hemin, hematin, activated charcoal, dihydroxyphenyl, and iron compounds are examples of such agents. A simple and widely used supplement is the triple combination of ferrous sulfate, sodium metabisulfite, and sodium pyruvate, each at a final concentration of 0.025% (FBP).

Detection in Food by Culture

The isolation of campylobacters from food is more difficult than from the feces of a patient with campylobacter enteritis, as there are usually far fewer organisms present. For fecal specimens direct plate culture is normally adequate – indeed, organisms may be so plentiful that a presumptive diagnosis can be made by direct microscopy – but for food, selective enrichment cultures are essential. Possible exceptions are raw poultry products, which may be so heavily contaminated that enrichment is not critical; indeed, for some chicken samples enrichment has proved to be less sensitive than direct plating.

The principle of the enrichment method is to incubate the food (after stomaching or other treatment to break it down) in a selective broth for 24–48 h and then subculture on a selective agar for the isolation and identification of individual colonies. However, for campylobacters that have been subjected to hot or cold stress, which causes sublethal injury, other techniques may be necessary. As the treatment the foods have received is not always known, there is much to be said for including the more sensitive techniques routinely.

When properly carried out, cultural methods are able to detect 1 campylobacter colony-forming unit (CFU) in 2 g food that contains at least $10^6\,g^{-1}$ of other bacteria.

Specimen Handling and Transport

The correct handling and preservation of specimens are crucial for food specimens; incorrect handling negates the value of sensitive isolation procedures. The three most important factors are refrigeration (as near to 4 °C as possible), exclusion of oxygen, and prevention of drying. Oxygen may be excluded by gassing with 100% N_2, but this is not always practicable. Provided specimens are not unduly large, they can be placed in a semisolid transport medium, such as Cary-Blair medium (0.16% agar), or directly into a campylobacter-selective broth, each containing either FBP or sodium bisulfite 0.01%. The choice of medium may be critical, and it is often better to use a highly selective formulation, as campylobacters do not compete well with other microorganisms. The choice will be influenced by the nature of the sample, the physical state of the target organisms, and the time likely to elapse before cultures are set up.

Milk and liquid foods should be held refrigerated after adding sodium bisulfite 0.01% and either sodium thioglycollate 0.15% or 100% N_2. Under these conditions there is little loss of viability over several days, but the sooner specimens are cultured the better.

Enrichment Techniques

Campylobacters that have been frozen, chilled (4 °C), or subjected to mild heating become sublethally injured. This is manifested by increased susceptibility to antibiotics (notably rifampicin and polymyxin), deoxycholate, oxygen radicals such as peroxides, and high incubation temperatures. This susceptibility is thought to be due to damage to the outer membrane of the bacterial cell and it can significantly reduce the efficacy of detection by culture.

Thus the first stage of an isolation procedure is designed to allow the organisms time to recover. Ideally this is done in a nonselective broth incubated at 37 °C. This is feasible with samples such as water containing few competing flora, and in which campylobacters are often severely injured. However, for samples such as chicken, preincubation in nonselective media often results in a reduced isolation rate because the high numbers of competing flora overwhelm *Campylobacter*. Given the unpredictability of bacterial contamination levels, it is better to use selective media for the routine examination of food, water, and environmental samples. The broth should be at about room temperature when inoculated, and if the volume is large, incubation is best done in a water bath.

In choosing a liquid-selective medium it is important to take one that is able to suppress the growth of competing flora adequately, particularly *Pseudomonas* and *Proteus* spp. The choice will depend partly on the type of sample being examined, as some of the many available formulae perform better with some materials than others. A good all-round highly selective medium is Exeter broth (Table 1).

Various protocols have been laid down for incubation procedures. In general, maximum sensitivity can

Table 1 Formula for Exeter campylobacter-selective enrichÙment broth (and agar)

Nutrient broth[a]	1 l
Sodium metabisulfite	0.25 g
Sodium pyruvate	0.25 g[b]
Ferrous sulfate ($7.H_2O$)	0.25 g
Polymyxin	2500 IU
Rifampicin	5 mg
Trimethoprim	10 mg[c]
Cefoperazone	15 mg
Amphotericin	2 mg
Lysed horse blood	50 ml
(New Zealand agar[d]	15 g)

[a]Nutrient broth no. 2 (CM67, Oxoid Ltd, Basingstoke, UK) in original formula.
[b]Commercially available as MAST SV61 Campylobacter growth supplement.
[c]Commercially available as MAST SV59 Campylobacter-selective supplement.
[d]Only for solid plating medium.

be attained by initial incubation at 37 °C for 18 h followed by incubation at 42–43 °C for 30 h. This latter period can be extended for a further 24 h, and with some lightly contaminated samples the use of this extension has been shown to improve isolation rates. With highly selective media like Exeter broth it is possible to incubate samples at 37 °C throughout. In general, as large a volume of broth as is practicable should be used in order to accommodate as much of the sample as possible; a minimum of 25 g in 225 ml broth is recommended.

For liquid samples such as milk, it may be convenient to add the sample to an equal volume of double-strength medium. For maximum sensitivity, however, it is necessary to examine large volumes (around 200 ml) and this can present problems. Alternative methods of testing large volumes are to culture gauze or cotton wool pads through which the milk has been filtered, or to culture centrifuged deposits.

It is generally acceptable to incubate fluid cultures in closed vessels containing air. Slight increases in sensitivity can be attained by the use of Erlenmeyer flasks with stoppered side-arms, so that air can be partially extracted and replaced with a mixture of 10% carbon dioxide, 10% hydrogen, and 80% nitrogen. Best results are obtained if there is a constant flow of gas mixture with agitation of the broth.

Subculture to Agar Plates

Mature enrichment cultures are subcultured on a campylobacter-selective agar. There are numerous formulae, but either charcoal, cefoperazone, deoxycholate (CCDA) or Exeter agars are recommended. Hydrogen peroxide and other potentially toxic oxygen metabolites can accumulate in agar even if the plates are stored in the dark under refrigeration. It has been shown that these compounds act synergically with antibiotics, such as rifampicin, to inhibit campylobacter growth severely. Thus for maximum sensitivity it is important to use fresh plates and include FBP supplement in the agar, even if blood is also included.

Plates are incubated at 42–43 °C in a microaerobic atmosphere containing 5–10% oxygen and preferably some hydrogen. If cylinder gases are available, the easiest method is to extract two-thirds of the air from a closed container, such as an anaerobic jar (without catalyst), and replace it with a mixture of 10% carbon dioxide, 10% hydrogen, and 80% nitrogen. If there is no cylinder gas, an anaerobic-type gas-generating envelope can be used, and the jar partially evacuated after closure. Alternatively, special gas-generating envelopes for campylobacters are commercially available for use in jars with a catalyst.

Growth is usually visible after 24 h, but occasional strains do not appear until incubation has been maintained for 48 h.

The foregoing protocols are designed to detect the main causes of campylobacter enteritis, namely *C. jejuni* and *C. coli* (they will also detect *C. lari*). Other species of campylobacter and related bacteria, such as *C. uppsaliensis*, *C. hyointestinalis*, and *Arcobacter butzleri*, which have been implicated to a lesser degree, are unlikely to be detected. Those wishing to isolate these bacteria, for research or whatever reason, must use special selective media incubated at 37 °C rather than 42 °C. Alteratively, nonselective media may be used in conjunction with membrane filters placed on the agar.

Fecal Indicator Bacteria and Campylobacters

It will be evident from the foregoing account that the culture of campylobacters from food is time-consuming and costly, especially if regular monitoring is undertaken. Campylobacters are rarely found in the absence of fecal bacteria, so for processed foods it is reasonable to screen for the latter rather than look for campylobacters. Most raw meats have fecal coliforms on them, and it must be assumed they also have campylobacters. In short, it pays to think carefully about the value of screening for campylobacters before undertaking a monitoring program; it could be unproductive and uneconomic.

Identification

In general, most *Campylobacter* colonies are typically flat, effuse, and wet-looking, although other morphotypes can occur. Identity is confirmed by a positive oxidase reaction and the presence of characteristic motile, spiral, or S-shaped Gram-negative organisms in wet preparations and Gram-stained smears. Additional basic identification tests are shown in Table 2. Sets of these tests are available from commercial sources and one includes simple resistogram typing. Latex agglutination tests for identification are also available commercially, but they need careful assessment. Molecular methods that depend on the analysis of DNA fragments can perform well in identification tests, but they tend to be too complex for general use. In choosing a test appropriate to one's needs it is important to know whether it discriminates between groups of species, single species, or strains within a species.

Meat samples, especially chicken meats, are likely to carry *Arcobacter* spp. These bacteria closely resemble *Campylobacter* spp. on agar plates. *Arcobacter* spp. can be distinguished by their ability to grow in air at 25 °C (Table 2). Not all *Arcobacter* spp. grow at 42 °C.

Table 2 Basic differential characters of campylobacters likely to be found in food

Species	Growth at 25 °C	Hippurate hydrolysis	Indoxyl acetate hydrolysis	Sensitivities	
				Nalidixic acid	Cephalothin
C. jejuni	−	+	+	S	R
C. coli	−	−	+	S	R
C. lari	−	−	−	R	R
C. fetus	+	−	−	R	S
Arcobacter spp.	+[a]	−	+	S	R

−, negative; +, positive; R, resistant (no zone of inhibition to 30 μg disk); S, sensitive (any zone of inhibition to 30 μg disk).
[a]Growth in air.

Typing

Serotyping on the basis of heat-stable lipopolysaccharide antigens (Penner system) and heat-labile flagellar and outer-membrane protein antigens (Lior system) is the main typing method in current use. There are several hundred designated serotypes, but only a limited number are found frequently. The discriminatory power of serotyping can be enhanced by the parallel use of biotyping and/or phage typing. Typing facilities are only available at reference laboratories.

In recent years a plethora of molecular-typing methods have appeared based on the analysis of fragments of DNA split by specific enzymes. Their discriminatory powers and complexity vary widely. Some are simple enough for wide application, but all suffer from a lack of standardization. Until this problem is resolved, serotyping, supplemented with phage typing, remains the basic typing method.

Detection in Food by Noncultural Methods

Immunological methods that make use of high-affinity monoclonal antibodies were the first noncultural methods to be used for the direct detection of campylobacters. The antibody/bacterial antigen reaction was detected by latex agglutination or enzyme-linked immunosorbent assay (ELISA). Work on these techniques was largely foresaken in favor of molecular methods, in which fragments of DNA specific to the bacterium are recognized by specific probes and then amplified by PCR. A sensitivity of 3–15 *C. jejuni* cells in 100 ml water and $3\,g^{-1}$ has been claimed. These methods work better with samples that do not contain an excess of other bacteria or biological material, but they have recently been made to work with feces. The main problem is that they need specialized skills and reagents. False-positive results are likely unless scrupulous care is taken to avoid contamination with DNA. These methods are likely to come into their own as they become refined and easier to use.

Detection in Suspected Food Poisoning

Outbreaks of campylobacter enteritis usually come to light through the recognition of a sudden excess of campylobacter isolations in a clinical laboratory, or, less often, by the recognition of an excess of diarrhea in a community. Either way the laboratory plays a key role, as the infection cannot be diagnosed clinically. Unfortunately, outbreak investigations are hampered by the long time that elapses between food consumption and awareness of an outbreak: the average incubation period is 3 days; people tend not to seek medical attention until they have been ill for several days; sampling, testing, and reporting may take 2–3 days; and several more days may pass before someone realizes that cases are epidemiologically connected. Thus, the investigation may only start 2 weeks or more after ingestion of the food, long after suspect items have been discarded. Even if foods are available, the vulnerability of campylobacters mitigates against their survival in detectable numbers. (*See* **Food Poisoning**: Tracing Origins and Testing.)

For the above reasons reliance must often be placed on statistical associations of infection with foods eaten. Such evidence can be powerful and prove beyond all reasonable doubt that a particular food was the vehicle of infection. Protocols for investigating such outbreaks can be complex and require close cooperation between community doctors, public health workers, microbiologists, veterinarians, and food producers. The prime essential is to have all efforts directed and coordinated by one person.

See also: ***Campylobacter***: Properties and Occurrence; Campylobacteriosis; **Food Poisoning**: Tracing Origins and Testing

Further Reading

Advisory Committee on the Microbiological Safety of Food (1999) *Report on Microbial Antibiotic Resistance in Relation to Food Safety*. London: Stationery Office.

Corry JEL, Post DE, Colin P and Laisney MJ (1995) Culture media for the isolation of campylobacters. *International Journal of Food Microbiology* 26: 43–76.

Hurtado A and Owen RJ (1997) A molecular scheme based on 23S rRNA gene polymorphisms for rapid identification of *Campylobacter* and *Arcobacter* species. *Journal of Clinical Microbiology* 35: 2401–2404.

Johnson LG and Murano EA (1999) Comparison of three protocols for the isolation of *Arcobacter* from poultry. *Journal of Food Protection* 62: 610–614.

Mason MJ, Humphrey TJ and Martin KW (1999) Isolation of sublethally injured campylobacters from poultry and water sources. *British Journal of Biomedical Science* 56: 2–5.

Nachamkin I and Skirrow MB (1998) *Campylobacter, Arcobacter*, and *Helicobacter*. In: Collier L, Balows A and Sussman M (eds) *Topley and Wilson's Microbiology and Microbial Infections*, 9th edn, vol. 2, pp. 1237–1256. London: Edward Arnold.

Nachamkin I, Engberg J and Aerestrup FM (2000) Diagnosis and antimicrobial susceptibility of *Campylobacter* species. In: Nachamkin I and Blaser MJ (eds) *Campylobacter*, 2nd edn, pp. 45–66. Washington DC, ASM Press.

Owen RJ (1995) Update on epidemiological typing of *Campylobacter*. *PHLS Microbiology Digest* 12: 2–6.

Owen RJ, Lorenz E and Gibson J (1997) Application of the Mast resistotyping scheme to *Campylobacter jejuni* and *C. coli*. *Journal of Medical Microbiology* 46: 34–38.

Roberts D, Hooper W and Greenwood M (eds) (1995) *Practical Food Microbiology*, 2nd edn. London: Public Health Laboratory Service.

Waage AS, Vardund T, Lund V and Kapperud G (1999) Detection of small numbers of *Campylobacter jejuni* and *Campylobacter coli* cells in environmental water, sewage, and food samples by a seminested PCR assay. *Applied and Environmental Microbiology* 65: 1636–1643.

Campylobacteriosis

M B Skirrow, Gloucestershire Royal Hospital, Gloucester, UK

Introduction

Campylobacteriosis, or *Campylobacter* enteritis, is an infection of the intestines manifest by acute diarrhea and abdominal pain. It is the most frequently reported form of infective diarrhea in the human population in most industrialized nations. Although the illness seldom lasts more than a few days, occasional patients develop complications, such as reactive arthritis or Guillain–Barré syndrome (GBS), a serious neurological disorder that can result in permanent disability or even death. The burden of campylobacteriosis on society in terms of health care, economic costs, and suffering is immense. This article describes the clinical manifestations, pathology, and treatment of the disease, and ends by considering options for its prevention and control. There are many approaches to control, but ultimately the only one that will have lasting impact is the control of infection in food-producing animals, especially broiler chickens.

Clinical Features

Campylobacter enteritis is caused mainly by *Campylobacter jejuni* and *C. coli* (*See* ***Campylobacter***: Properties and Occurrence), but before describing the disease due to these organisms, brief mention is made of an uncommon low-grade septicemic form of infection known as systemic campylobacteriosis, usually caused by *C. fetus*. Although blood stream invasion by *C. fetus* may arise from intestinal colonization, diarrhea is not a common feature. This form of infection is limited to patients with disease of the immune system, cancer, or some other chronic disease compromising their immunity. It is of minor importance relative to *Campylobacter* enteritis, but it earns a place here because *C. fetus* infection has been acquired by eating raw calves' liver. *C. fetus* is a common inhabitant of the intestinal tract of sheep and cattle, in which it may cause septic abortion.

Campylobacter enteritis is an acute self-limiting diarrheal disease clinically indistinguishable from *Salmonella* enteritis, although minor differences become apparent when groups of patients are compared.

Infection can be established with as few as 500 organisms. The average incubation period is 3 days (range 1–7 days). In most patients the illness starts with abdominal pain and diarrhea, but in about one-third of patients there is an influenza-like prodrome of malaise, headache, and fever, sometimes with rigors, for a period ranging from a few hours to a day or so before the onset of gastrointestinal symptoms. Profuse diarrhea lasts for 2–3 days and is almost invariably accompanied by abdominal pain, which is sometimes severe enough to simulate acute appendicitis and cause some patients to undergo emergency surgery. Occasionally there is genuine appendicitis. Children sometimes have abdominal pain without diarrhea. Surveys of patients affected in community outbreaks show that, on average, about 50% have fever, 40% myalgia, 8% rigors, 14% frank blood in their stools, but only 15% vomit, even though nausea is common.

Some patients present with acute colitis that is difficult to distinguish from an acute attack of idiopathic ulcerative colitis (inflammatory bowel disease). Acute

complications are rare but include intestinal haemorrhage, toxic megacolon, hemolytic–uremic syndrome, pancreatitis, and septic abortion.

After a few days the diarrhea eases and about 80% of patients recover within a week, irrespective of any treatment they may receive. Minor relapses have been reported in about 20% of patients after a day or two without diarrhea.

Young and middle-aged adults seem to be most severely affected by the infection. Unlike salmonellosis, *Campylobacter* enteritis in infants is often mild, and although infection in old people tends to be more invasive, it is seldom dangerous. Death is rare and virtually limited to frail patients debilitated from other causes.

Late Complications

There are two late complications that typically arise 1–2 weeks after the onset of illness: reactive arthritis and GBS.

Reactive Arthritis

Reactive (aseptic) arthritis affects about 1% of patients, particularly those possessing the human leukocyte antigen (HLA) B27. The ankles, knees, wrists, and the small joints of the hands are most frequently affected, often in migratory fashion. It can cause painful debility lasting for several months, but it is ultimately benign. Clinically it is the same as the reactive arthritis that sometimes follows other forms of acute bacterial enteritis.

Guillain–Barré Syndrome

Another name for GBS is postinfective polyneuropathy. It is a serious, though fortunately rare complication causing paralysis that does not always resolve completely. Severely affected patients develop respiratory paralysis and have to be nursed on a ventilator. Surveys indicate that *Campylobacter* infection is the antecedent event in 14–41% of patients with GBS – it is the most commonly identified trigger for the disease.

The mechanism of this distressing condition, in which the myelin in peripheral nerves becomes lost or damaged, is thought to be immunological. Certain *C. jejuni* strains are particularly associated with GBS, and some have been found to possess cell wall epitopes that cross-react with proteins in peripheral nerve myelin. It is believed that this triggers immunological attack and destruction of the myelin.

Convalescent Excretion

Excretion of *Campylobacter* in feces falls off exponentially after illness. About 50% of patients are culture-negative after 3 weeks, and virtually all are negative after 3 months. Long-term carriage is unknown except in patients with severe immune deficiency, such as hypogammaglobulinemia or acquired immune deficiency syndrome (AIDS). Spread of infection from a healthy excreter is a remote risk. In the case of food handlers, provided they have formed stools there is no need to exclude them from work.

The basic features of *Campylobacter* enteritis are summarized in Table 1.

Table 1 Basic features of *Campylobacter* enteritis

Feature	*Duration and details*
Incubation period	3 days (range 1–7 days)
Main symptoms	Diarrhea, abdominal pain, fever
Mean duration of illness[a]	
Incapacity	3.8 days
Symptoms	14.6 days
Mean duration of fecal *Campylobacter* excretion[a]	38 days (maximum 69 days)
Exclusion from food handling	Until stools formed

[a]Figures taken from a Norwegian study of 135 sporadic cases.

Pathology

The essential lesion in *Campylobacter* enteritis is an acute inflammation of the jejunum and small intestine, which progresses to the cecum, colon, and rectum. The histopathology of the intestinal mucosa is one of acute inflammatory cellular infiltration indistinguishable from that seen in *Salmonella, Shigella*, or other acute bacterial infections of the gut. The terminal ileum and cecum are often particularly affected, and it is this feature that causes the symptoms that mimic acute appendicitis.

Campylobacter are invasive and may be present in the blood stream at the onset of illness, as indicated by the high fever and rigors experienced by some patients. Unlike *C. fetus, C. jejuni* and *C. coli* are sensitive to the natural killing power of normal serum, so any bacteremia is likely to be transient and hence seldom detected in blood cultures (bacteremia reported in only 0.15% of infections). Regional lymph nodes are often enlarged and inflamed. A cholera-like enterotoxin and at least one cytotoxin are produced in modest quantities by many strains of *C. jejuni* and *C. coli*, but the role played by these in pathogenesis is unclear.

Immunity

Circulating (humoral) antibodies are produced within a few days of illness and for at least several months patients are refractory to reinfection with the same strain. People regularly exposed to infection develop

solid immunity. Children in developing countries, where transmission rates are high, become immune within the first 2 years or so of life, and occupationally exposed groups in industrialized countries also become immune. For example, in a milkborne outbreak of *Campylobacter* enteritis in the USA, 76% of people exposed for the first time were ill, whereas no person who habitually drank raw milk became ill. In the UK significant amounts of *Campylobacter*-specific antibody were found in 18% of veterinary assistants, 36% of cattle abattoir workers, and 27–68% of workers in poultry-processing plants, compared with 2–5% in nonexposed adults. Infection with one or two strains may be sufficient to give broad immunity, although there are patients who have had two attacks of *Campylobacter* enteritis several months apart; in each case the second infection has been with a different strain. It is not known how long immunity lasts, but regular exposure is probably necessary for its maintenance.

Treatment

As *Campylobacter* enteritis almost always resolves spontaneously, treatment is primarily symptomatic and supportive. The most important element, as with all forms of acute diarrhea, is the correction of dehydration and electrolyte loss, particularly in infants and old people, who are least able to do this of their own accord. In many cases simply getting patients to increase their fluid intake with assorted drinks is sufficient, but giving a balanced electrolyte preparation such as Dioralyte is ideal.

Antimicrobial therapy is effective if given within a day or two of the onset of illness, but a bacteriological diagnosis has seldom been made at this stage. If a patient is still suffering from acute symptoms when a positive report is received, antimicrobial therapy is justified. Erythromycin is the antibiotic of choice for both children and adults. In patients with severe diarrheal illness, who might be put at risk if treatment were delayed until a bacteriological report was available, there is a case for treating empirically with an antimicrobial agent effective against the common enteropathogenic bacteria. Ciprofloxacin, or an equivalent quinolone drug, was once the ideal agent here, and it still may be in some places. Unfortunately, in recent years resistance rates in *Campylobacter* (as well as *Salmonella*) have soared to around 50% in some European countries, so that its use has been drastically compromised. This is almost certainly a result of the widespread use of quinolones in poultry.

Fortunately erythromycin resistance is less of a problem. In industrialized countries resistance has remained at about 1–5% of strains, but higher rates have been found in some developing countries, notably 65% in Thailand. Erythromycin resistance is much more common in *C. coli* than *C. jejuni*, probably due to the use of tylosin, a similar macrolide antibiotic, in pig rearing.

Prevention

Basic measures for the prevention of any communicable intestinal infection apply equally to *Campylobacter* enteritis. These include the treatment and safe disposal of sewage, the provision of potable water, the pasteurization or equivalent heat treatment of milk, and the general hygienic processing and handling of food. An important aspect of the prevention of *Campylobacter* infection is the interruption of transmission from food-producing animals, particularly poultry, to the final food product. Each of these is now considered.

Sewage Disposal

Campylobacter are plentiful in raw sewage. The discharge of untreated sewage is undesirable as it helps to maintain the *Campylobacter* population in the environment. Conventional sewage treatment is effective in eliminating *Campylobacter*. In one study over 78% were removed after primary sedimentation, and less than 0.1% remained in the final effluent.

Water Purification

Campylobacter are ubiquitous in natural surface waters, including sea water. They are capable of surviving for several weeks in cold water (below 15 °C). Survival is increased in waters that have an autochthonous microflora. *Campylobacter* also appear to be capable of forming viable but nonculturable forms, which are alleged to survive for several months or longer. However, the potential role of such dormant cells in the epidemiology of *Campylobacter* has yet to be established. Conventional treatments are fully adequate for removing *Campylobacter*, which are generally more sensitive to chlorine and other disinfecting agents than *Escherichia coli*. Contact with monochloramine 1.0 mg l^{-1} for 15 min or free chlorine 0.1 mg l^{-1} for 5 min causes more than 99% inactivation. *Campylobacter* are also sensitive to ultraviolet light and are destroyed by protocols designed to kill *E. coli*.

Major community outbreaks of *Campylobacter* enteritis have been caused by the distribution of unchlorinated or inadequately chlorinated water (*See* ***Campylobacter***: Properties and Occurrence), which serves to emphasize the importance of proper treatment. This applies equally to remote oligotrophic lakes and surface waters, which become contaminated from wild

birds and other animals. A number of outbreaks have arisen as a result of the failure of outdated, overloaded, or inadequately maintained water works or distribution systems. (*See* **Water Supplies**: Water Treatment.)

Heat Treatment of Milk

Pasteurization and other orthodox treatment processes are fully effective in destroying *Campylobacter*. Despite this, milkborne outbreaks of infection regularly occur, even where it is illegal to sell raw milk to the public. Some of the largest outbreaks of *Campylobacter* enteritis on record have been milkborne, mostly from the distribution of raw milk, but a few of them from ostensibly pasteurized milk. Thus, although compulsory pasteurization is a highly desirable goal, it is unlikely to eliminate all risks of milkborne infection. Bad design, malfunctioning of machinery, or those running the machinery, the difficulty of designing satisfactory small pasteurizing units at reasonable cost are all factors that need attention in order to minimize risks. Electricity supply failures putting pasteurizing machinery out of action and bad weather preventing the delivery of milk to central pasteurizing plants have also been precipitating factors in outbreaks of milkborne infection.

Food Hygiene

Measures taken to maintain good hygiene and food-handling practice in restaurants and other catering establishments are fundamental to the prevention of *Campylobacter* enteritis. Much infection is believed to be transmitted through the cross-contamination of cooked or ready-to-eat foods, such as bread and salads, from raw poultry or other raw meats. The separation of raw from cooked foods and the use of dedicated equipment for each, such as chopping boards, mixers, knives, and other utensils, is the single most important principle to be observed. If raw meats cannot be stored separately they should always be placed on a bottom shelf so they cannot drip juices on to other foods.

Raw poultry should be handled as little as possible; handling and even washing carcasses only spreads organisms around work areas. The discipline of washing hands after touching raw meats of any sort should be rigorously observed.

It has been shown that *Campylobacter* can be isolated from the flesh of chicken carcasses even when the surface has been disinfected by immersion in boiling water. Chicken muscle affords the organism some protection from the effects of heat and this is the probable explanation for the implication of barbecued chicken in outbreaks of *Campylobacter* enteritis. It is therefore particularly important that chicken and other poultry meat is properly cooked.

These rules must be taught to all professional food handlers and it is the responsibility of health authorities to see that correct instruction is carried out and maintained to satisfactory standards. Analysis of outbreaks shows that many arose through ignorance of these basic rules. (*See* **Poultry**: Chicken.)

Control of *Campylobacter* in Poultry

Elimination of *Campylobacter* from food-producing animals is clearly impracticable owing to the wide natural distribution of the organisms. Yet there are ways of reducing colonization in farm animals, especially in poultry.

Campylobacter colonization in broiler chicken flocks does not usually appear until 2 weeks or more after hatching. This is taken to indicate horizontal transmission of *Campylobacter* from the environment either within or outside the broiler house. Thus there is the possibility of preventing colonization, or at least extending the *Campylobacter*-free period, by improving hygienic practice among broiler-house attendants and proofing houses against intrusion by birds and rodents. Such interventions have met with partial success, notably in Scandinavia and The Netherlands.

In some cases water has been shown to be a crucial factor in promoting colonization. Many broiler houses are served by unchlorinated private borehole supplies that are suspect. Even in mains-supplied houses *Campylobacter* can become established in distribution systems beyond header tanks. Experimental interventions, in which header tanks and distribution pipework have been thoroughly cleaned and disinfected, have been followed by dramatic falls in colonization rates in the broilers. Unfortunately such measures are difficult to sustain under agricultural conditions and further research is needed.

It is generally believed that, unlike *Salmonella* infection, vertical transmission of *Campylobacter* from parent flocks does not occur. This view has been challenged by the results of a 5-year study in the UK, in which the distribution of *C. jejuni* serotypes in broiler flocks supplied with day-old chicks from two hatcheries pointed to vertical rather than horizontal transmission. *C. jejuni* readily survives in embryonated eggs. If vertical transmission is confirmed, there is the possibility of preventing *Campylobacter* colonization of broilers by ensuring that breeder flocks are kept free from infection.

The principle of competitive exclusion, in which normal intestinal flora is artificially established in newly hatched chicks in order to exclude pathogens such as *Salmonella*, is not applicable to *Campylobacter* control, as the intestinal flora has already become established by the time *Campylobacter*

make their appearance. An experimental oral vaccine has been used with some success in broiler chicken flocks. This is a sound approach that could have major benefit.

The total elimination of *Campylobacter* from broiler flocks is probably not practicable, but reducing the extent of colonization may be sufficient to curtail transmission to the human population. The logistics, precise methods, and costs of such measures have yet to be worked out, but they are urgently needed if there is to be any hope of reducing the present unacceptable burden of the disease. These measures need to be studied in relation to the problems of cross-contamination during the mechanical processing of poultry, which are discussed elsewhere (*See* ***Campylobacter***: Properties and Occurrence.)

Terminal irradiation of dressed poultry is an option that removes all pathogens, including *Campylobacter*, but there are problems with this, not least public acceptability. (*See* **Irradiation of Foods**: Basic Principles.)

Waste disposal from poultry farms An added advantage of controlling *Campylobacter* colonization of poultry is that fewer bacteria would be returned to the environment and to surface waters. Heavily colonized flocks have a prodigious output of *Campylobacter* that are likely to be recycled in farm stock via surface water and wild or feral animals that are closely associated with farms, such as magpies, jackdaws, starlings, pigeons, sparrows, rats, and mice. Cattle with access to surface water are commonly colonized, whereas those with access only to mains water have been found to be free of *Campylobacter*.

Public Education

It has been said that if everyone handled their foods properly, foodborne *Campylobacter* enteritis would virtually disappear. This is probably true, but it is an unrealistic expectation. Yet much could be done to improve matters. Surveys to assess public knowledge of the rudiments of food hygiene have revealed abysmal ignorance. It should not be difficult to remedy this. Even to impart the simplest facts about handling poultry could have an impact that far outweighs the effort – and it would help control salmonellosis as well as *Campylobacter* enteritis. Television is the most influential medium and a substantial impact could be made by broadcasting high-quality entertaining yet informative spot 'advertisements.' The understandable fears of the poultry and retail industries could be allayed by inviting their participation and ensuring that the content of such spots was sensibly balanced. With appropriate professional skills, it could even be turned to their advantage. Public moneys used to fund such a venture should be amply offset by the savings from infections prevented.

See also: ***Campylobacter***: Properties and Occurrence; **Irradiation of Foods**: Basic Principles; **Poultry**: Chicken; **Water Supplies**: Water Treatment

Further Reading

Advisory Committee on the Microbiological Safety of Food (1999) *Report on Microbial Antibiotic Resistance in Relation to Food Safety*. London: Stationery Office.

Humphrey TJ (1989) Salmonella, campylobacter and poultry: possible control measures. *Abstracts on Hygiene and communicable diseases* 64: R1–R8.

Lastovica AL and Skirrow MB (2000) Clinical significance of *Campylobacter* and related species other than *Campylobacter jejuni*. In: Nachamkin I and Blaser MJ (eds) *Campylobacter*, 2nd edn, pp. 89–120. Washington, DC:American Society for Microbiology.

Nachamkin I, Mishu Allos B and Ho TW (2000) *Campylobacter jejuni* infection and the association with Guillain–Barré syndrome. In: Nachamkin I and Blaser MJ (eds) *Campylobacter*, 2nd edn, pp. 155–175. Washington, DC: American Society for Microbiology.

Newell DG and Wagenaar JA (2000) Poultry infections and their control at farm level. In: Nachamkin I and Blaser MJ (eds) *Campylobacter*, 2nd edn, pp. 497–509. Washington, DC: American Society for Microbiology.

Piddock LJV (1995) Quinolone resistance and *Campylobacter* spp. *Journal of Antimicrobial Chemotherapy* 36: 891–898.

Ransom GM, Kaplan B, McNamara AM and Wachsmuth IK *Campylobacter* prevention and control: the USDA-Food Safety and Inspection Service role and new food safety approaches. In: Nachamkin I and Blaser MJ (eds) *Campylobacter*, 2nd edn, pp. 511–528. Washington, DC: American Society for Microbiology.

Rice BE, Rollins DM, Mallinson ET, Carr L and Joseph FW (2000) *Campylobacter jejuni* in broiler chickens: colonization and humoral immunity following oral vaccination and experimental infection. *Vaccine* 15: 1922–1932.

Skirrow MB and Blaser MJ (2000) Clinical aspects of *Campylobacter* infection. In: Nachamkin I and Blaser MJ (eds) *Campylobacter*, 2nd edn, pp. 69–88. Washington, DC: American Society for Microbiology.

Smith KE, Bender JB and Osterholm MT (2000) Antimicrobial resistance in animals and relevance to human infections. In: Nachamkin I and Blaser MJ (eds) *Campylobacter*, 2nd edn, pp. 483–495. Washington, DC:American Society for Microbiology.

Van de Giessen AW, Tilburg JJHC, Ritmeester WS and van der Plas J (1998) Reduction of campylobacter infections in broiler flocks by application of hygiene measures. *Epidemiology and Infection* 121: 57–66.

Wassenaar TM and Blaser MJ (1999) Pathophysiology of *Campylobacter jejuni* infections in humans. *Microbes and Infection* 1: 1023–1033.

Yang S, Leff MG, McTague D *et al.* (1998) Multistate surveillance for food-handling, preparation, and consumption behaviors associated with foodborne diseases: 1995 and 1996 BRFSS food-safety questions. *Morbidity and Mortality Weekly Report* 47: 33–56.

CANCER

Contents

Epidemiology

A Trichopoulou and P Lagiou, University of Athens Medical School, Goudi, Athens, Greece
D Trichopoulos, Harvard School of Public Health, Boston, MA, USA

Nutritional Epidemiology of Cancer

Methodological Complexities

Epidemiological studies exploring the role of diet in the etiology of cancer have several methodological complexities, which may be classified into partially overlapping categories as follows:

- Instruments for ascertainment of past diet are inherently weak; therefore, random misclassification of dietary factors of causal importance may be extensive, leading to substantial underestimation of genuine effects and to false-negative results.
- Random misclassification may be so extensive as to substantially reduce the power of a study, to the extent that an apparently positive result may be as likely to reflect a statistical artefact (alpha error) as a biologically genuine association.
- Mutual confounding among nutritional factors is extensive and complex, because there are many nutritional factors, and many of them have a common dietary origin (**Figure 1**).
- Random misclassification of nutritional factors may substantially reduce the ability to control their confounding influences.
- Most nutritional factors implicated in human carcinogenesis are considered to be late-stage agents (promoters or growth enhancers) rather than early stage agents (initiators). All of them, however, have their own specific latent periods. This poses theoretical and practical difficulties in the definition of a common latency when the effect of more than one nutritional factors are considered in relation to a specific cancer in any particular study.
- The exposure–response curve describing cancer risk in relation to a particular nutritional variable may be sigmoid, implying the existence of thresholds below or above which no association could be documented between the two variables.
- When a particular cancer has several independently operating nutritional causes, the relative risk associated with any particular nutritional cause decreases with the prevalence or the abundance of the other nutritional causes.
- On the contrary, when two factors are jointly necessary for the causation of a particular cancer, for example, an unknown initiator and a particular

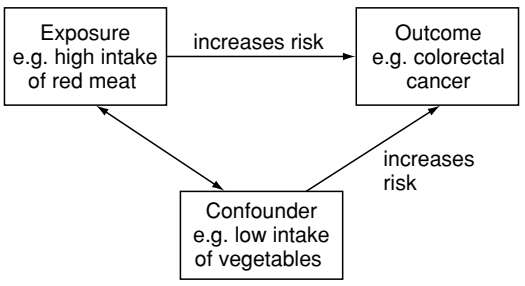

Figure 1 Confounding.

food item or nutrient as promoter, the relative risk associated with the nutritional factor increases with the prevalence or the abundance of the unknown initiator.

- When many nutritional factors are examined in the context of a particular study, multiple comparisons can generate both false-positive and false-negative results.
- Food patterns, for instance the Mediterranean diet, may have effects that are considerably different to those of the constituent foods or nutrients because of complex interactions.
- Interindividual variability of certain food items or nutrients may be too small in comparison with the corresponding intraindividual variability, thus considerably reducing the power of analytical epidemiological studies. Indeed, in many populations, the diet may be too uniform and etiologically important food items may be too common or too rare to allow statistical substantiation of risk differentials with standard analytic epidemiological studies.
- Energy intake may be an important cause, an equally important confounder or an indicator of bias, particularly in case-control studies. Control of energy intake is thus necessary whenever particular foods or nutrients are evaluated.

Epidemiological Evidence on the Nutritional Etiology of Cancer

The effect of diet on cancer occurrence is poorly understood in molecular or mechanistic terms, although it is generally assumed that dietary factors act at a late stage in the long process of carcinogenesis. Several patterns have emerged, however, from the many epidemiological studies that have been undertaken during the last 30 years on the relation between diet and cancer, and these are summarized in Tables 1–4. The associations are distinguished into convincing, probable, and possible, and can be either positive (increasing intake of the dietary factor increases cancer risk) or inverse (increasing intake of the dietary factor reduces cancer risk).

Table 1 refers to the cancer risk implications in relation to the intake of major food groups. Vegetables and, to a slightly lesser extent, fruits are inversely associated with risk for several forms of cancer. There is also evidence that intake of red meat increases the risk for some forms of cancer, particularly cancer of the large bowel.

The collective evidence for macronutrients, as summarized in Table 2, is considerably weaker for obvious reasons. In the absence of long-term randomized intervention studies, the evidence concerning macronutrients is derived from the evidence concerning food groups with additional uncertainties introduced by variability in food composition tables and analytical methods, as well as by uncontrollable confounding due to the possible operation of unidentifiable or immeasurable factors that tend to coexist with some of the macronutrients.

Table 3 summarizes the risk implications for major forms of cancer by intake of selected micronutrients. Although the study of micronutrients shares the same constrains as that of macronutrients,

Table 1 Risk implications for major forms of cancer by consumption of major food groups

Cancer site	*Cereals*	*Pulses*	*Vegetables*	*Fruits*	*Meat (red)*	*Fish*	*Dairies*	*Eggs*	*Sugars*
Mouth/pharynx			↓↓↓	↓↓↓					
Nasopharynx									
Esophagus			↓↓↓	↓↓↓					
Stomach			↓↓↓	↓↓↓					
Large bowel			↓↓	↓	↑↑			↑	↑
Liver			↓						
Gall-bladder									
Pancreas			↓	↓	↑				
Larynx			↓↓	↓↓					
Lung			↓↓↓	↓↓					
Breast			↓↓	↓↓					
Endometrium			↓	↓					
Cervix uteri			↓	↓					
Ovary			↓	↓					
Prostate			↓		↑		↑		
Urinary bladder			↓↓	↓↓					
Kidney			↓		↑				

↓↓↓, convincing inverse association; ↑↑↑, convincing positive association; ↓↓, probable inverse association; ↑↑, probable positive association; ↓, possible inverse association; ↑, possible positive association.

Table 2 Risk implications for major forms of cancer by consumption of macronutrients

Cancer site	*Protein*	*Starch*	*Fiber*	*Saturated lipids*	*Monounsaturated lipids*	*Polyunsaturated lipids*
Mouth/pharynx						
Nasopharynx						
Esophagus						
Stomach		↑				
Large bowel	↑		↓	↑		
Liver						
Gall-bladder						
Pancreas	↑	↑	↓			
Larynx						
Lung				↑		
Breast					↓	
Endometrium			↓	↑		
Cervix uteri						
Ovary						
Prostate				↑		
Urinary bladder						
Kidney						

↓↓↓, convincing inverse association; ↑↑↑, convincing positive association; ↓↓, probable inverse association; ↑↑, probable positive association; ↓, possible inverse association; ↑, possible positive association.

Table 3 Risk implications for major forms of cancer by consumption of selected micronutrients

Cancer site	*Vitamin A and carotenoids*	*Folate*	*Vitamin C*	*Vitamin D*	*Vitamin E*	*Calcium*	*Selenium*
Mouth/pharynx			↓				
Nasopharynx							
Esophagus	↓		↓				
Stomach	↓		↓↓				
Large bowel	↓	↓		↓	↓	↓	
Liver							
Gall-bladder							
Pancreas			↓				
Larynx							
Lung	↓		↓		↓		↓
Breast	↓						
Endometrium							
Cervix uteri	↓		↓		↓		
Ovary							
Prostate	↓↓				↓		↓
Urinary bladder							
Kidney							

↓↓↓, convincing inverse association; ↑↑↑, convincing positive association; ↓↓, probable inverse association; ↑↑, probable positive association; ↓, possible inverse association; ↑, possible positive association.

this table appears more informative than Table 2 for three reasons, two of which are purely artificial. First, of the many micronutrients, only those for which some evidence exists are included. Second, several carotenoids are included in the single column dedicated to these compounds. Thus, the evidence for lung cancer mostly refers to β-carotene, whereas the evidence for prostate cancer refers almost exclusively to lycopene. The third reason, which is more substantive, is that the physiologic and pathophysiologic effects of certain micronutrients have been extensively studied, so that the criteria of biologic plausibility can be more successfully relied upon for micronutrients rather than for macronutrients. For instance, several carotenoids, vitamins C and E, and selenium have been intensively studied on account of their experimentally documented antioxidant potential.

Table 4 refers to nonnutrients and nutritional correlates like height, obesity, and physical activity. Entries in the same column do not imply identical products or processes. For instance, alcohol causes liver cancer through cirrhosis and cancer of the esophagus and larynx, mostly through interaction with tobacco smoking, and cancer of the breast possibly by increasing levels of estrogens. Moreover,

Table 4 Risk implications for major forms of cancer by exposure to selected nonnutrients and nutrition-related indicators

Cancer site	*Alcohol*	*Salt*	*Coffee*	*Hot drinks*	*Height*	*Obesity*	*Physical activity*
Mouth/pharynx	↑↑↑		↑ (mate)				
Nasopharynx		↑↑					
Esophagus	↑↑↑		↑↑ (mate)	↑↑			
Stomach	↑	↑↑↑					
Large bowel	↑				↑	↑	↓↓↓
Liver	↑↑↑						
Gall-bladder						↑↑	
Pancreas						↑	
Larynx	↑↑↑						
Lung							
Breast	↑↑				↑↑↑	↓↓/↑↑↑	↓
Endometrium						↑↑↑	
Cervix uteri							
Ovary							
Prostate					↑	↑	↓
Urinary bladder			↑				
Kidney						↑↑	

↓↓↓, convincing inverse association; ↑↑↑, convincing positive association; ↓↓, probable inverse association; ↑↑, probable positive association; ↓, possible inverse association; ↑, possible positive association.

coffee and mate are mixtures that have both similarities and differences. Lastly, the effect of obesity may refer to stages of carcinogenesis that are not necessarily identical in the various forms of cancer. At the extreme, obesity among premenopausal women reduces the risk of breast cancer, whereas it considerably increases this risk among postmenopausal women.

Additives and Contaminants

Some readers may be surprised by the absence of a table referring to additives and contaminants that are frequently reported as carcinogens in the lay press. There are several reasons for this. First, the intake of additives and contaminants cannot be documented in most epidemiological research. Second, additives and contaminants are extensively scrutinized in laboratory assays, and whenever evidence incriminates them as carcinogens, these are removed from the diet. Third, the alleged breast carcinogenicity of organochlorine compounds, which can be measured in the blood or adipose tissue, has not been documented in large and sophisticated epidemiological investigations; nor has there been any direct epidemiological support for a role of nitrosamines in human carcinogenesis. At this stage, the only additive that is likely to play a major role in human carcinogenesis is salt. Among dietary contaminants, only aflatoxin has been conclusively linked to hepatocellular carcinoma and may be causing a nonnegligible fraction of liver cancer in developing countries. The suspicion that chlorination byproducts in drinking water may be involved in urinary bladder carcinogenesis has not been documented but is currently under investigation.

Prevention of Diet-related Cancers

On the basis of studies in migrants and ecologic investigations, it has been estimated that diet may be responsible for up to 30% of cancer mortality in developed countries. Although we know little about the mechanisms of diet-related carcinogenesis, we have a very good idea of what a prudent diet that reduces cancer in humans should be. This diet should be high in vegetables, fruits, pulses, and cereals, low in red meat and salt, and low in saturated fat of animal origin. Added lipids should be of plant origin, and, among them, olive oil has a safety record of several thousand years. Obesity should be avoided, preferably by increasing physical activity, which, in itself, can reduce the incidence of colorectal cancer and, perhaps, other forms of cancer as well. Regular physical activity during childhood and adolescence may also slow down excessive growth, as reflected in attained height, and may have beneficial consequences on several cancer types. Alcohol, which is known to protect against cardiovascular diseases, should be consumed in moderation, particularly by women, because it contributes to the causation of cancer at several sites, including the breast.

See also: **Aflatoxins**; **Alcohol**: Metabolism, Beneficial Effects, and Toxicology; **Antioxidants**: Natural Antioxidants; **Cancer**: Carcinogens in the Food Chain; **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; **Colon**: Cancer of the Colon; **Contamination of Food**

Further Reading

Lagiou P, Trichopoulou A and Trichopoulos D (2002) Nutritional epidemiology of cancer: accomplishments and prospects. *Proceedings of the Nutrition Society*. 61: 217–222.

MacMahon B and Trichopoulos D (1996) *Epidemiology: Principles and Methods*, Second Edition. Boston: Little Brown.

Willett W (1998) *Nutritional Epidemiology*, 2nd edn. New York: Oxford University Press.

Willett WC and Trichopoulos D (eds) (1996) Nutrition and cancer. *Cancer Causes and Control* 7: 3–180.

World Cancer Research Fund and American Institute for Cancer Prevention (1997) *Food, Nutrition and the Prevention of Cancer: A Global Perspective*. Washington, DC: World Cancer Research Fund and American Institute for Cancer Prevention.

Carcinogens in the Food Chain

M J Prival, United States Food and Drug Administration, Washington, DC, USA

Background

It has become increasingly clear in recent decades that food can have a substantial effect on cancer risk. This conclusion is based on the results of both human population studies and animal studies. Presently available evidence indicates that dietary fat may be a risk factor for cancer and that fresh fruits and vegetables may have a protective effect. In addition, the human diet contains specific carcinogenic substances, most, but not all of which are present at such low levels that their effects on human cancer rates, if any, are likely to be small. This article summarizes current information on carcinogens in the diet that are (1) naturally occurring components of food, (2) produced by the processing of food, or (3) added intentionally to food. Carcinogenic contaminants of food, including mycotoxins and pesticides, are discussed elsewhere. (*See* **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; Carcinogenicity Tests; **Fatty Acids**: Dietary Importance; **Contamination of Food**; **Mycotoxins**: Occurrence and Determination.)

Naturally Occurring Components of Foods

Bracken Fern

The toxic effects of bracken fern on livestock have been observed for a century or more. Bracken fern is not only toxic, but also carcinogenic when fed to rats and other experimental animals at high levels. Both the toxicity and the carcinogenicity are attributed to the same chemical constituent, ptaquiloside (**Figure 1**).

Young bracken fronds in the fiddlehead or crosier stage are used as a salad green in some countries, including Japan. In Japan, bracken fern is boiled or pickled before eating to remove the astringent taste; this reduces, but does not eliminate, its carcinogenicity. Milk from cattle that were fed bracken fern has been found to be carcinogenic to rats. This raises the possibility that there may be a risk associated with the drinking of milk in areas where dried bracken fern is used for bedding in winter or where range-fed cattle may consume bracken fern; this occurs, for example, in Turkey and Bulgaria.

Pyrrolizidine Alkaloids

Pyrrolizidine alkaloids occur in a variety of plants used as food and herbal medicines in many parts of the world, particularly in Africa and Asia. The first natural substances shown to cause liver tumors were pyrrolizidine alkaloids from plants. Among the pyrrolizidine-alkaloid-containing plants used by humans are *Petasites japonicus* Maxim (a kind of coltsfoot), which is used as a food and herbal remedy in Japan, and *Symphytum officinale* L. (comfrey), which is widely used as a medicinal herb, salad ingredient, or 'tea.' Both of these plants have been shown to be carcinogenic when fed to rats. Senkirkine (**Figure 1**), which is found in *P. japonicus* M., induced liver tumors when injected intraperitoneally into male rats, as did symphytine (**Figure 1**), extracted from roots of comfrey (*S. officinale*). Both alkaloids can also induce mutations in the bacterium *Salmonella typhimurium*; this is an indication of their ability to react with deoxyribonucleic acid (DNA) and may be relevant to the mechanism of their carcinogenicity. (*See* **Alkaloids**: Toxicology.)

Mate

Mate (pronounced 'mah-taý') is a beverage that is widely consumed in southern Brazil, Paraguay, Uruguay, and northern Argentina. It is made by brewing the toasted, dried, and aged leaves of a particular type of tree, *Ilex paraguariensis*, which is indigenous to this area. Several studies have compared mate consumption among patients with cancers of the esophagus, larynx, oral cavity, or pharynx with that of comparable individuals who did not have these cancers. Taken as a whole, these studies indicate that consumption of hot mate is probably carcinogenic to the upper gastrointestinal tract. There is currently no way of knowing to what extent the

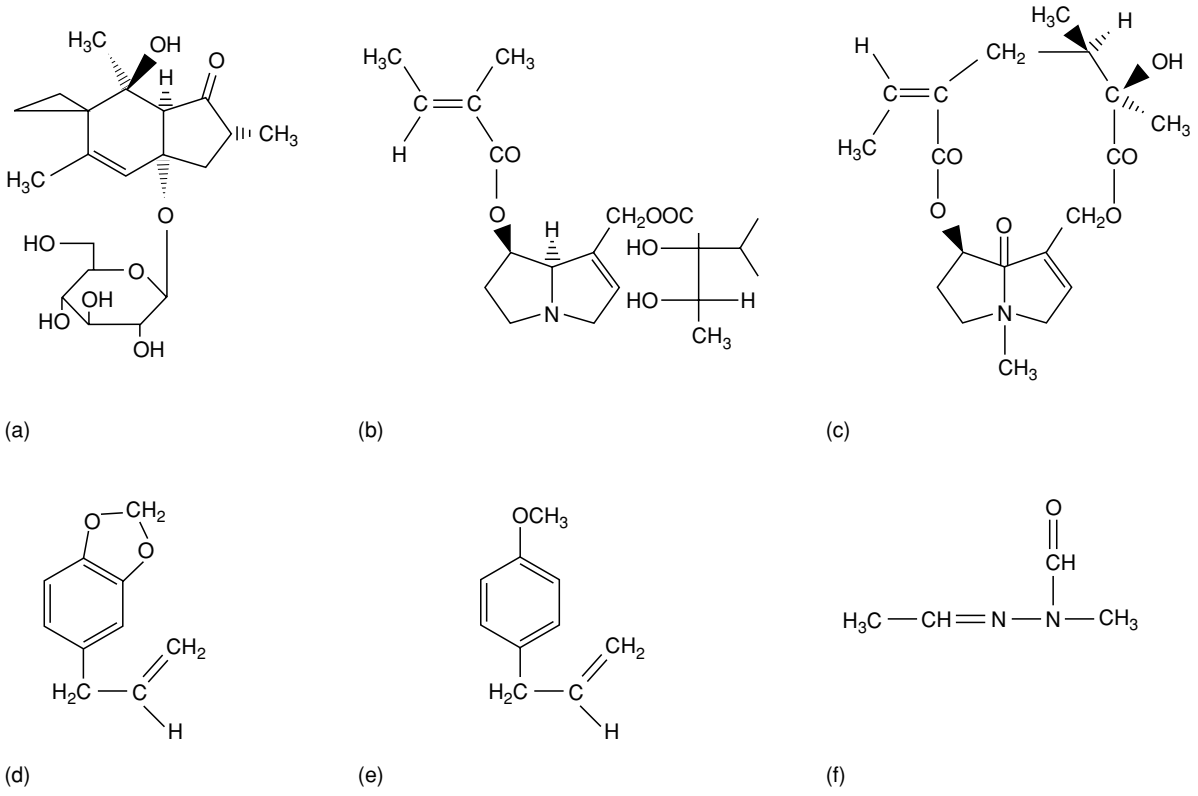

Figure 1 Chemical structures of some naturally occurring dietary carcinogens: (a) ptaquiloside, the carcinogenic constituent of bracken fern; (b) symphytine and (c) senkirkine, two pyrrolizidine alkaloids; (d) safrole and (e) estragole, two alkenylbenzene carcinogens found in foods; and (f) gyromitrin, a toxic and carcinogenic compound found in the false morel mushroom, *Gyromitra esculenta*. Reproduced from Cancer: Carcinogens in the Diet, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

chemicals present in the mate contribute to its apparent carcinogenicity and to what extent the high temperature of the consumed beverage is the problem. Experiments on the carcinogenicity of mate on animals might help to resolve this question, but unfortunately, no such experiments have been reported.

Coffee

The many studies on the possible association between coffee consumption and cancer in people have given conflicting results. Overall, the evidence suggests that coffee consumption may increase the risk of bladder cancer in humans, but this conclusion is by no means definitive because of the possible failure to account properly for other factors that can affect cancer rates. Some studies, but not others, have linked high levels of coffee consumption to cancer of the pancreas, but again, the positive findings may result from methodological problems. Unfortunately, currently available studies on the carcinogenicity of coffee to experimental animals do little to shed light on the issue owing to the inadequacy of these studies. Caffeine, which is found in tea, cola drinks, and cocoa as well as coffee, has not been found to be carcinogenic when tested in animals. (*See* **Caffeine; Coffee**: Physiological Effects.)

Alkenylbenzenes and Benzene

Several alkenylbenzene compounds have been shown to induce liver tumors in experimental animals. One of these compounds, safrole (Figure 1), constitutes 80–90% of the volatile oil of sassafras and is also present at much lower concentrations in certain spices and flavoring ingredients such as sweet basil, cananga oil, nutmeg, pepper, tamarind, and ylang ylang oil. As a result of the carcinogenicity of safrole, the US Food and Drug Administration (FDA) prohibited the use of safrole or sassafras oil in food, which resulted in the cessation of its use as the principal flavoring ingredient in root beer. Furthermore, sassafras extracts or leaves can only be sold in the USA if the safrole has been removed from them.

Another carcinogenic alkenylbenzene related in chemical structure to safrole is estragole (Figure 1; also called methylchavicol), which is found in

tarragon, sweet basil, anise, West Indian bay, fennel, chervil, and marjoram. The level of estragole from these sources in a typical diet is very low, and the resulting risk is, at most, vanishingly small.

Benzene is a well-established human and animal carcinogen, inducing leukemia in occupationally exposed workers and several types of cancer in benzene-treated rodents. Traces of benzene have been detected among the volatile components of a variety of foods, including oat groats, processed pork and ham, cooked meats and baked potatoes, and in the aromas of coffee and cocoa.

Mushrooms and Hydrazine Derivatives

Gyromitra esculenta is one of the mushrooms known as false morels. It is harvested and eaten by many people in northern Europe and also in the USA, although it is poisonous, even fatal, unless properly dried and boiled. The principal poisonous ingredient is gyromitrin (Figure 1), which is carcinogenic to animals and can give rise to the carcinogen methylhydrazine under acidic conditions, such as exist in the stomach. Both methyl-hydrazine and another carcinogen, *N*-methyl-*N*-formyl-hydrazine, appear to form during the preparation of false morel for consumption. In fact, cooking must be performed in an open vessel to allow methylhydrazine to escape; otherwise, poisoning can occur.

The widely available edible mushroom cultivated in countries with a temperate climate is *Agaricus bisporus*. The principal hydrazine derivative in *A. bisporus*, agaritine, has not been shown to be carcinogenic, but there are reports of carcinogenicity for some hydrazine derivatives related to compounds that have been identified in this mushroom. When uncooked *A. bisporus* was fed to mice as their only food for 3 days per week throughout their lifetimes, tumors developed in several organs. No adequate study has yet been performed on the carcinogenicity of cooked mushrooms, and cooking may inactivate carcinogens that may be present. Additional studies are needed to determine the significance of the results with uncooked *A. bisporus* and the hydrazine derivatives.

Carcinogens Produced by the Processing of Food

Alcoholic Beverages

Alcoholic beverages probably constitute the most widely consumed class of substances for which human carcinogenicity is definitively established. The most clearly affected sites are the mouth, pharynx, larynx, esophagus, and liver. Smoking of tobacco and consumption of alcohol, each of which is carcinogenic independently, appear to have a far greater additive effect in inducing human cancer at these sites, excluding the liver. There are also studies implying that alcoholic beverages increase the risk of breast cancer, but the data in this regard are not nearly as definitive as those for the other sites mentioned.

It is not clear which components of the alcoholic beverages are responsible for the human cancers. Animal experiments have failed to show that ethanol (alcohol) itself is carcinogenic, although the major metabolite of ethanol, i.e., acetaldehyde, induces respiratory-tract tumors when inhaled by experimental animals. Alcoholic beverages contain numerous chemicals that may contribute to the carcinogenicity of the beverages, including traces of acetaldehyde, *N*-nitroso compounds (including *N*-nitrosodimethylamine), and urethan (ethyl carbamate). Urethan has been found in the highest concentrations in 'stone fruit' (i.e., cherry, plum, apricot, etc.), brandies, sake, and rice wine; at lower levels in various other types of distilled spirits and wines; and at even lower levels in beer. Urethan is also found in other fermented food products such as soya sauce, bread, yogurt, and some cheeses. The carcinogenic risks from urethan appear to be extremely low and are probably negligible, but may be somewhat increased among those who frequently consume alcoholic beverages with the highest levels of this carcinogen. (*See* **Alcohol**: Metabolism, Beneficial Effects, and Toxicology; Alcohol Consumption; **Smoking, Diet, and Health**.)

Carcinogens Produced by Cooking

Under most circumstances, the cooking of meat, fish, or eggs results in the formation, at very low concentrations, of a group of heterocyclic amines that are very potent mutagens when tested on bacteria. A few of these mutagens, all of which are carcinogenic when fed to experimental animals, are shown in Figure 2 with their commonly used abbreviated names.

The formation of mutagenic activity has been observed in beef, pork, ham, bacon, lamb, chicken, fish, and eggs after broiling (grilling), frying, and barbecuing. Other high-protein foods, such as tofu, beans, and cheese, gave little or no mutagenic activity when cooked under similar conditions. Beef extract is also mutagenic. (*See* **Browning**: Nonenzymatic.)

While the cooking-induced mutagens that have been isolated are extraordinarily potent mutagens when tested on bacteria, these chemicals tend to be only moderately potent as carcinogens. The fact that the mutagens are found in foods in the parts per billion range, and the fact that their carcinogenic

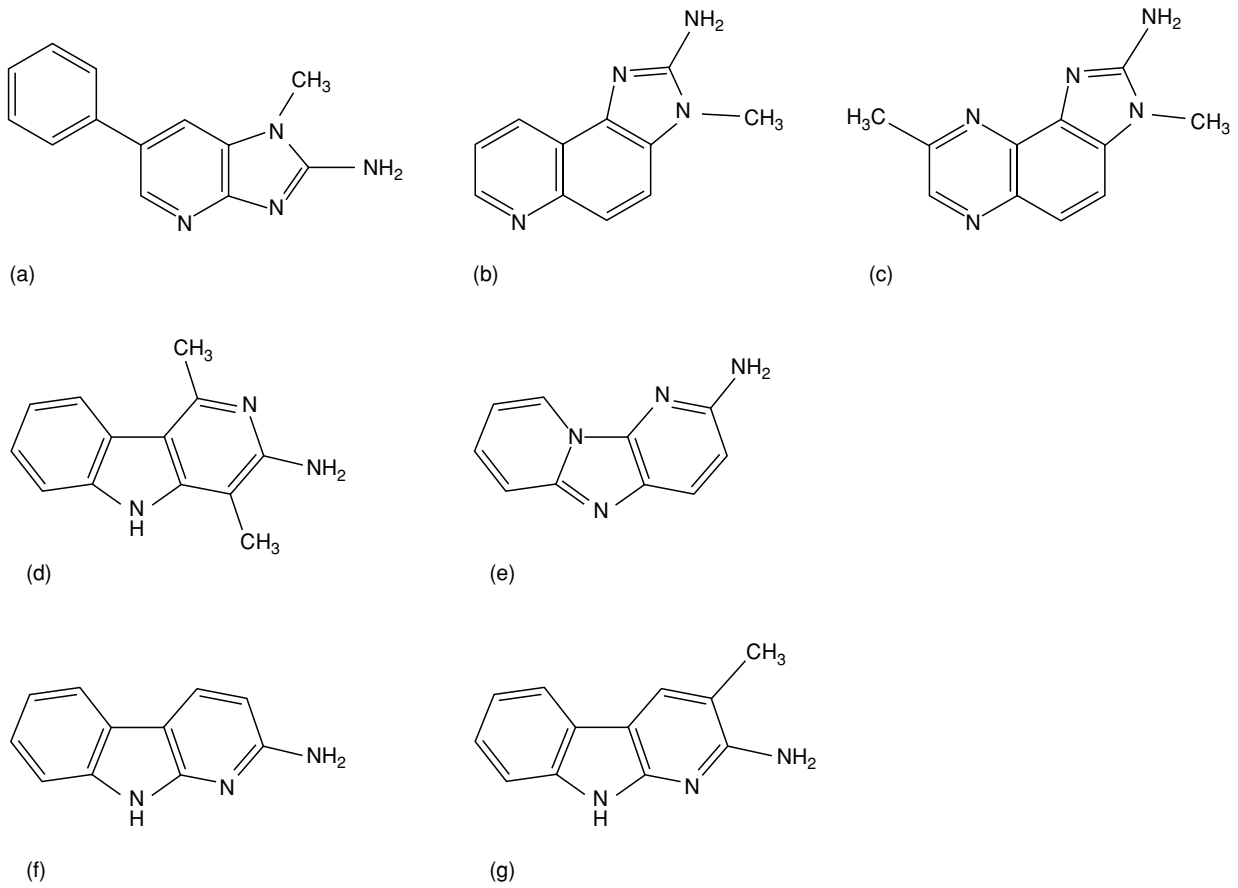

Figure 2 Chemical structures of some cooking-induced heterocyclic amine carcinogens. Their commonly used short names are (a) PhIP, (b) IQ, (c) 8-MeIQx, (d) Trp-P-1, (e) Glu-P-2, (f) A-α-C, and (g) MeA-α-C. Reproduced from Cancer: Carcinogens in the Diet, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK, and Sadler MJ (eds), 1993, Academic Press.

potency, in animals at least, does not seem to be particularly great, are factors that minimize the anticipated risk from these substances. However, exposure to them occurs each time a person eats cooked meat or fish, and such exposure often occurs at a high frequency over the lifetime of the individual. While it is possible to reduce the level of carcinogenic heterocyclic amines by reducing the temperature or time of cooking, the reduction in carcinogenic risk thus effected must be weighed against the very real possibility of increased risk from bacterial or parasitic infection from undercooked meat or fish. Recent reports concerning the isolation of an anticarcinogen from fried ground beef also imply that caution is necessary before recommending any reduction in the degree of cooking. Microwave cooking without surface browning apparently does not result in the formation of heterocyclic amine mutagens or carcinogens. (*See* **Mutagens**.)

Another class of chemicals containing a number of carcinogens, the polynuclear aromatic hydrocarbons, are also introduced into food when meat or fish are grilled, but not when cooking is performed by the lower-temperature processes of frying, roasting, or microwave cooking. The pyrolysis of fat dripping on to the heat source appears to be a major source of this class of carcinogens. Benzo[a]pyrene and other polynuclear aromatic hydrocarbon carcinogens appear to be relatively weak carcinogens when administered orally to experimental animals, but much more potent when applied to the skin. Whether this is also true for humans is unknown, and the risks associated with these carcinogens are difficult to evaluate. (*See* **Polycyclic Aromatic Hydrocarbons**.)

Salting of fish

The salting of fish in southeastern China is carried out for several days using sea salt, followed by drying in direct sunlight. Sometimes, the fish is permitted to soften by decomposition. Several studies have found that the consumption of such salted fish contributes to the high incidence of cancer of the nasal passages and pharynx (nasopharyngeal carcinoma) among Chinese people in southeastern China, Hong Kong,

and Malaysia. Nasopharyngeal carcinoma is rare in most populations but accounts for about 15% of all cancer deaths among males in Guangdong Province, China. Feeding of salted fish to children and to babies during weaning seems to be a particularly strong risk factor.

The conclusions of these population studies have been confirmed by animal studies in which feeding of Cantonese-style salted fish to rats resulted in the induction of nasopharyngeal carcinomas. *N*-Nitroso compounds formed as a result of the salting process may be responsible for the carcinogenicity of the salted fish. Nitrate, which contaminates the salt used to prepare the fish, can be reduced to nitrite during the processing, and the nitrite can react with amines in the fish to form *N*-nitroso compounds. Since salted fish consumption appears to account for most of the nasopharyngeal carcinoma cases among Chinese populations, it must be considered to be a very high risk factor when consumed regularly, particularly by children and infants.

Chemicals Added to Foods

Saccharin

Saccharin has been used as a nonnutritive sweetening agent since 1907, and its safety has been a subject of debate by scientists and public health officials from that time. Experiments conducted in the 1970s and 1980s have demonstrated that high levels of sodium saccharin in the diets of rats resulted in the appearance of tumors of the urinary bladder. In spite of this, the controversy concerning the safety of saccharin continues. (*See* **Saccharin**.)

Studies on human populations have failed to show any association between bladder cancer incidence and the consumption of saccharin or other artificial sweeteners, but such studies have a limited ability to detect such associations, even if they exist. The fact that saccharin is generally not mutagenic, or is very marginally mutagenic, in a variety of test systems may imply that saccharin does not interact with DNA and, therefore, that it may not pose a significant carcinogenic risk at the usually consumed human doses, which are far below the doses fed to the rats in the positive carcinogenicity studies. Saccharin continues to be widely used in food products around the world.

Cyclamates

Other widely used nonnutritive sweeteners include sodium cyclamate and calcium cyclamate. In a study performed in the 1970s, rats that were fed high doses of a mixture of sodium cyclamate and saccharin had an increased incidence of bladder tumors. Some additional studies reported that rats receiving sodium cyclamate alone developed bladder tumors, but the results were not statistically significant. Based on these findings, cyclamates were banned in some countries, including the UK and the USA. However, many consider the experimental reports of carcinogenicity to be of questionable value, and cyclamates continue to be approved for use in many countries. As mentioned above, there are no data from human populations indicating that consumption of nonnutritive sweeteners are associated with increased cancer risk, although such studies are, by their very nature, insensitive. (*See* **Cyclamates**.)

Nitrate and Nitrite

Sodium nitrate and sodium nitrite are used as preservative agents in cured meats, such as bacon, sausage, and ham, and in some cheeses. While neither salt has been shown to be carcinogenic, there is some concern because nitrite can react with (nitrosate) amines present in food to form *N*-nitroso compounds, many of which are carcinogens; this reaction can occur under the acid conditions in the stomach. Since nitrate can be converted to nitrite either in a food product or in the mouth or gastrointestinal tract, both nitrate and nitrite must be considered together in assessing any possible hazard. (*See* **Curing**; **Nitrosamines**.)

There are several dietary sources of nitrate and nitrite other than cured meats, including vegetables and some drinking waters. Over the past few decades, the levels of nitrate and nitrite in cured meats have been reduced considerably, and the potential for formation of *N*-nitroso compounds has been further mitigated by the addition to cured meats of antioxidants such as ascorbic acid, which inhibit the nitrosation of amines.

The addition of nitrate and nitrite to food is only one of many sources of *N*-nitroso compounds to which humans are exposed; salt-preserved fish and smoked foods are among the other sources of *N*-nitroso compounds in the human diet. While there are some data indicating that nasopharyngeal, esophageal, and stomach cancer are more common in populations that consume high levels of nitrate or certain types of preserved meat or fish, the specific role of *N*-nitroso compounds or of added nitrate or nitrite in human cancer remains to be clarified. It is possible that the presence of inhibitors of nitrosation in fresh fruits and vegetables is a factor in explaining why consumption of these foods appears to have a protective effect against cancer in human populations.

Butylated Hydroxyanisole

The antioxidant butylated hydroxyanisole (BHA) has been used since 1947 to prevent rancidity in edible

fats and oils and in fat-containing foods. Experiments performed since 1982 have shown that BHA in the diet can cause tumors of the forestomach in rats and hamsters. However, the relevance of this finding has been questioned because humans do not possess a forestomach, and their stomachs do not contain the type of tissue (squamous epithelium) found in the rodent forestomach. Although humans have squamous epithelium at other sites along the digestive tract, no tumors have been noted in such sites in experimental animals. The possibility that BHA may be a carcinogenic hazard at normal human dietary levels, which are much lower than the experimental tumor-inducing levels, has also been questioned on the grounds that it is not mutagenic in a variety of test systems. Because of the questions that have been raised concerning the relevance of the experimental forestomach tumors induced by BHA to human exposure at low dietary levels, and the effectiveness of BHA as an antioxidant in fatty foods, it continues to be approved for use in many countries around the world. (*See* **Dietary Fiber:** Properties and Sources; **Food Additives:** Safety.)

See also: **Alcohol**: Metabolism, Beneficial Effects, and Toxicology; Alcohol Consumption; **Alkaloids**: Toxicology; **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; Carcinogenicity Tests; **Contamination of Food**; **Curing**; **Cyclamates**; **Dietary Fiber**: Properties and Sources; **Food Additives**: Safety; **Mutagens**; **Mycotoxins**: Occurrence and Determination; **Polycyclic Aromatic Hydrocarbons**; **Saccharin**; **Smoking, Diet, and Health**

Further Reading

Aeschbacher H-U (ed.) (1991) Potential carcinogens in the diet. *Mutation Research* special issue, vol. 259, nos 3 and 4.

Hayatsu H (ed.) (1991) *Mutagens in Food: Detection and Prevention*. Boca Raton, FL: CRC Press.

Hirono I (ed.) (1987) *Bioactive Molecules, vol. 2. Naturally Occurring Carcinogens of Plant Origin*. Amsterdam: Elsevier Science.

IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: vol. 22, *Some Non-Nutritive Sweetening Agents* (1980); Vol. 31, *Some Food Additives, Feed Additives, and Naturally Occurring Substances* (1983); vol. 40, *Some Naturally Occurring and Synthetic Food Components, Furocoumarins, and Ultraviolet Radiation* (1986); vol. 44, *Alcohol Drinking* (1988); vol. 51, *Coffee, Tea, Mate, Methylxanthines, and Methylglyoxal* (1991) Lyon: International Agency for Research on Cancer.

National Research Council (1989) *Diet and Health: Implications for Reducing Chronic Disease Risk*. Washington, DC: National Academy Press.

Pariza MW, Aeschbacher H-U, Felton JS and Sato S (eds)(1990) *Progress in Clinical and Biological Research*, Vol. 347. *Mutagens and Carcinogens in the Diet*. New York: Wiley–Liss.

Weisburger EK (1987) Carcinogenic natural products in the environment. In: Mehlman M (ed.) *Advances in Modern Environmental Toxicology*, vol. X, pp. 243–266.

Diet in Cancer Prevention

I T Johnson, Institute of Food Research, Norwich, UK

Introduction

Cancer is primarily a disease of old age. In societies where infant mortality and infections have ceased to be major causes of death, a large proportion of the population will survive to die eventually of cancer. Nevertheless, the age-specific incidence of many common types of cancer varies by a factor of 10-fold or more between populations, the age-standardized incidences of particular cancers change within populations over time, and the risks alter consistently in populations migrating between countries. Since only a small proportion of human cancers are directly attributable to inherited susceptibility, most must be caused by environmental factors that are, in principle, avoidable. In their 1981 review of the causes of cancer, Sir Richard Doll and Richard Peto showed that, apart from tobacco, which at that time was the direct cause of 30% of deaths from cancer in the USA, the largest proportion of avoidable cancer (approximately 35%) could be attributed to diet. However the uncertainty attached to this estimate was very high (10–70%). Surprisingly little has changed in the intervening years. The American epidemiologist WC Willet has recently suggested that the true proportion of cancers avoidable by dietary change is probably between 20 and 42%, but the means by which such reductions might be achieved remain stubbornly obscure. There are two major sources of difficulty. First, there are many different types of cancer, and the various mechanisms by which they arise have only recently begun to be understood. Second, human diets are immensely complex and difficult to measure. The biology of cancer, and the epidemiological evidence for effects of diet, will be briefly reviewed here, and the current state of knowledge regarding the role of dietary factors in the prevention of cancer will be discussed.

The Nature of Cancer

Cell division is fundamental to all life. The human body begins life with the fertilization of a single cell, yet by the time maturity is reached the total number has risen to about 10^{13}. Throughout the earliest stages of development all the embryonic cells divide constantly and undergo differentiation to form the specialized tissues and organs. Cell division is minimal in many adult tissues but those of the blood-forming organs, the reproductive tissues, and the epithelia lining the gut continue to divide at a high rate throughout life. A tumor can be defined as a localized increase in the number of cells in an organ, beyond what is necessary for tissue growth and repair, or for the reproductive needs of the host. Benign tumors grow relatively slowly, and retain an approximately normal anatomical location, whereas malignant tumors acquire the ability to invade surrounding tissues and form secondary tumors in remote organs such as the liver or brain. Until relatively recently it was assumed that the crucial abnormality of cancerous cells was their loss of proliferative control, but their ability to evade apoptosis, the internal pathway which obliges damaged cells to undergo suicide, is at least as important.

Carcinogenesis occurs in stages that can be broadly characterized as initiation, promotion, and progression. The earliest event is thought to be the acquisition of an unrepaired mutation by a single cell. DNA damage and repair occur constantly, but a mutation which remains unrepaired, fails to trigger apoptosis, confers some growth advantage on the cell, and can be passed on through successive cellular divisions, is potentially carcinogenic. During the promotion stage, the normal constraints on proliferation and spatial location are disrupted by the acquisition of further mutations within genes essential for the normal growth and differentiation. At the progression stage the lesion has made the transition to malignancy and can give rise to secondary tumors at remote sites. This multistage model of carcinogenesis is exemplified by colorectal cancer, which in most cases develops through the adenoma–carcinoma sequence. Around 40% of all males over 50 in the UK and other western industrialized countries have one or more colorectal polyps. Only a few percent of these will ever progress to carcinomas, but regular screening and removal of such polyps are an effective, though costly means of preventing cancer.

Cancer and Populations

Some examples of the geographical variations that occur in the incidence of important cancers of the alimentary tract are given in Table 1. These data are of comparable reliability, and have been selected to illustrate some important features of cancer incidence across the globe. For example, colorectal carcinoma is about 12 times more common in New Zealand than in India. This cancer is consistently more common in richer, westernized countries than in the developing nations of Asia and Africa that have primarily agrarian economies. Cancer of the stomach is much less consistently related to economic development. The highest incidence of gastric cancer in the world occurs in Japan, but rates are also high in other much poorer countries of eastern Asia, and in Central and South America. Within countries, gastric cancer tends to decline with increasing socioeconomic status. The third example, esophageal cancer, shows an extraordinary degree of geographical variation. In general it tends to be more common in poorer societies, but there is a recognized "esophageal cancer belt" which runs from China through central Asia to Iran with incidence rates hundreds of times higher than the least affected countries of the world. Such extremes cannot be attributed to general economic conditions. They suggest that the esophagus is exquisitely sensitive to some combination of environmental conditions amongst which diet must play a major role.

Table 1 Geographical variations in the incidence of major cancers of the alimentary tract

Country	*Incidence of carcinoma for both sexes (cases/100 000)*		
	Esophageal	*Gastric*	*Colorectal*
Brazil	13.6	31.3	27.3
Denmark	8.3	12.6	69.3
Germany	7.0	25.3	77.0
India	12.7	8.5	8.0
Israel	3.1	22.1	78.1
Japan	11.6	97.8	68.5
New Zealand	7.9	16.2	98.7
South Africa	22.6	12.6	23.0
Uganda	24.5	12.5	14.4
UK	13.2	17.9	60.7
USA	6.3	11.2	71.2

Data are predicted age-specific incidence rates for the year 2000. Source: Globocan 2000 database of the World Health Organisation, Geneva (http://www.who.int).

Cancer and Diet

Three types of interaction between diet and the carcinogenic process can be envisaged.

1. The diet may contain specific carcinogens or promoters that initiate or accelerate carcinogenesis.
2. The provision of known nutrients and /or energy may be a key determinant of the body's susceptibility to cancer.

3. Biologically active nonnutrients may modulate the initiation or progression of carcinogenesis at various stages.

Carcinogens and Promoters

The Swiss toxicologists Lutz and Schlatter analyzed the contribution of various dietary factors to the risk of human cancer. They compared human exposure to known carcinogens with the activity of these substances in animal models, and concluded that only a small proportion of the human cancer cases attributed to diet by Doll and Peto were likely to be caused by foodborne carcinogens. Certainly there is very little evidence that food additives and pesticide residues impose any carcinogenic risks on the general public in industrialized societies. Food additives are rigorously screened to eliminate any risk of carcinogenic effects, and synthetic pesticide residues are no more likely to be mutagenic than natural plant constituents. In contrast, food supplies in developing countries are often at risk of contamination with carcinogenic toxins derived from fungi. Aflatoxin, a product of the fungus *Aspergillus flavus*, frequently contaminates groundnuts stored in hot climates. Dietary exposure to aflatoxin is associated with increased risk of hepatocellular carcinoma, a form of liver cancer. Food supplies in western countries are screened for the presence of aflatoxins, but socioeconomic conditions make this difficult to achieve in developing countries where locally produced groundnuts are a dietary staple. Another potent class of carcinogens known to be present in foods are the nitrosamines. The best established example of a probable link between these compounds and disease is the high risk of nasopharyngeal cancers in southern Chinese populations consuming high levels of salted fish. In Japan, a high intake of various pickled and highly salted vegetable and meat products is believed to be one cause of the very high incidence of gastric cancer.

Epidemiological evidence derived from both case-control and cohort studies suggests that regular consumption of more than about 80 g of red meat (beef, pork, or lamb) per day leads to an approximately 2.5-fold increase in the risk of colorectal cancer. Around half of these studies show a statistically significant effect. The strongest evidence of an adverse effect comes from North America. However at the time of writing, the European Prospective Investigation of Cancer and Nutrition (EPIC) has begun to publish its first findings, and preliminary reports suggest a moderate adverse effect of red meat consumption on colorectal cancer risk, and a somewhat stronger effect of processed meat products. Cooked meat contains carcinogens, particularly heterocyclic amines, but at rather low levels. A second hypothesis is that heavy meat consumption increases the formation of nitrosamines by colonic bacteria. It is not entirely clear why processed meat should entail a higher risk, but one possibility is that the iron-containing protein heme is the risk factor, and that processed meats contain a higher concentration. In any case meat consumption is only one among many nutritional factors that influence colorectal carcinogenesis.

Alcohol is not a genotoxic carcinogen but there is good evidence that even moderate consumption increases the risk of breast cancer. Heavy consumption is an important cause of oropharyngeal, esophageal, laryngeal, and hepatic cancers. The risk is dose-related but also strongly linked to smoking.

Nutritional Factors

A large proportion of human diet-related cancers appear to be due to the effects of overconsumption of energy. Evidence that restricting energy intake would inhibit tumor growth in experimental animals first began to appear in the early part of the twentieth century. More recent studies have confirmed the relationship, and shown that increased physical exercise, leading to greater energy expenditure, has a similar effect. There are several plausible mechanisms of action, including a suppression of endogenous DNA damage caused by reactive oxygen species, enhanced DNA repair, and suppression of tumor growth, mediated by insulin and other endocrine factors. Direct evidence that similar mechanisms are at work in humans is difficult to achieve, but restricted energy intakes may well contribute to the reduced risk of colorectal cancer observed in agrarian communities of Asia and Africa. It is difficult to see how deliberate restriction of energy intake could be made to work as a cancer prevention strategy in the west, but regular exercise and avoidance of obesity are likely to be beneficial. An understanding of the mechanisms involved may eventually lead to other biological strategies for suppression of carcinogenesis.

It has long been thought that dietary fat may have an effect on the risk of cancer, independently of any effect of energy intake. International comparisons suggest positive links between fat intake and breast and colon cancer, and there is evidence that total fat intake is positively related to induction of tumors by chemical carcinogens in animals. However, the evidence is much less convincing when comparisons are made within populations, and there is no clear scientific consensus that total fat intake is an important risk factor. There is stronger experimental evidence to show that the fatty acid composition of the diet is

important. Polyunsaturated fatty acids (PUFA) of the *n*-6 series tend to promote mammary carcinogenesis in experimental animals, whereas *n*-3 PUFA such as eicosapentaenoic acid, which is found in fish oils, tend to suppress mammary and colorectal carcinogenesis. Epidemiological evidence has begun to emerge for protective effects of fish consumption against cancer of the colon, breast, and prostate, but these findings require validation.

Micronutrient status is another aspect of general nutrition that has received a great deal of attention in relation to several types of cancer. Small quantities of many minerals and organic micronutrients are essential to normal growth and function. Substances for which a specific deficiency condition has been identified are classified as nutrients, but many have biological effects beyond their classical function. Vitamin C (ascorbic acid) is one example. Ascorbic acid is essential for the prevention of scurvy in humans but it is also a strong antioxidant and numerous studies have shown that a relatively low level of ascorbate in the serum is a risk factor for several forms of cancer. However the evidence that supplementation with vitamin C reduces the risk of cancer is much less convincing. It is possible that plasma vitamin C is a biomarker for a high intake of fruit and vegetables. Similar arguments apply to a number of other antioxidant nutrients, including β-carotene, but strong protective effects against several types of cancer have been demonstrated in a human intervention study with selenium.

Low intakes of folate, one of the B-group vitamins, have been linked to increased risk of cancers of the colon, pancreas, lung, and breast. Individuals with one or more common inherited abnormalities of methylenetetrahydrofolate reductase, one of the principal enzymes involved in folate metabolism, may be at higher risk. The relationship is biologically plausible because folate is essential for normal synthesis of DNA, and for other aspects of cell proliferation. Folate deficiency causes misincorporation of nucleotides and DNA breaks similar to those caused by ionizing radiation.

Of all human cancers, squamous cell carcinoma of the esophagus seems the most likely to be caused, at least partially, by one or more deficiencies of known micronutrients. Populations characterized by a very high incidence of the disease often consume restricted diets, with low intakes of protein and many micronutrients, including riboflavin, niacin, vitamins C and E, β-carotene, and retinal. It seems probable that in high-risk communities, specific nutrient deficiencies act synergistically with other factors, such as specific foodborne carcinogens or the irritating effects of very hot or abrasive foods and drink.

Dietary fiber

The dietary fiber hypothesis has complex origins, but its best-known advocate was the colonial medical officer Dennis Burkitt, who sought to explain the rarity of colorectal carcinoma and other bowel disorders amongst rural Africans. Burkitt proposed that the various nondigestible polysaccharides of plant cell walls, collectively termed dietary fiber, are protective against cancer of the large bowel. His original hypothesis was based largely on the concept of fecal bulk. Working from clinical observations, Burkitt deduced that populations consuming traditional rural diets, rich in vegetables and cereal foods, produced frequent bulky stools. The increased risk of colorectal cancer in western populations eating low-fiber diets was attributed to a low volume of feces and infrequent bowel movements, causing prolonged exposure of the colonic epithelial cells to mutagens. The mildly laxative effects of dietary fiber are now well recognized, and numerous human intervention trials have confirmed that dietary supplements containing wheat bran or other types of dietary fiber can increase fecal volume and reduce transit times. Reports from such bodies as the World Cancer Research Fund, the European Cancer Prevention Organization, and the UK Department of Health have agreed that the evidence for a protective effect of dietary fiber against colorectal cancer is at least moderately convincing. The fecal-bulking hypothesis provides the principal rationale for the current dietary reference values for nonstarch polysaccharides in the UK, which recommend that adults should consume an average of 18 g of fiber per day.

Apart from the simple dilution effect of increased fecal bulk, the nonfermentable particulate components of plant cell walls may sequester bile acids, thus reducing their concentration in the aqueous phase of the fecal stream. This hypothesis is consistent with the results of human intervention studies with high-fiber low-fat diets, which have been shown to reduce fecal bile acids and slow the appearance of large adenomas.

In the three decades or so that have elapsed since the original work of Burkitt, great progress has been made in our understanding of the gradual transition from a normal colonic epithelium to a malignant tumor. Apoptosis is thought to provide a defense mechanism against the appearance and survival of cells carrying unrepaired mutations, and it is probably also important at a later stage in the adenoma–carcinoma sequence because the speed at which tumors grow depends upon their relative rates of cell production and death. The colonic microflora is rich in bacteria that can break down carbohydrates to yield the short-chain fatty acids acetate, propionate, and butyrate, and these are absorbed and metabolized by human tissues. Butyrate

is utilized preferentially by the colonic mucosa, and besides supplying metabolic energy, it suppresses proliferation, increases differentiation, and induces apoptosis of tumor cells *in vitro*. Thus the principal role of fermentable carbohydrate may be to regulate mucosal cell turnover and death by providing a constant endogenous supply of butyrate.

The dietary fiber hypothesis has never been disproved. Some large North American cohort studies have failed to confirm an inverse relationship between fiber intake and colorectal cancer, but it is doubtful whether the highest levels of fiber consumption in such studies were great enough to provide the protective effects seen in international comparisons. Preliminary reports from the EPIC study do provide some evidence of beneficial effects of dietary fiber in Europe, where the range of fiber intakes may be greater.

The colon is perhaps the most obvious site for fiber to exert anticarcinogenic effects, but there is evidence that it protects against other cancers. One analysis of a series of case-control studies on breast cancer and diet has suggested that a 20-g increase in daily fiber intake was associated with a statistically significant reduction in risk of breast cancer of about 15%. Diets containing this rather high level of fiber may be associated with low levels of body fat, reduced plasma estrogen levels, and improved insulin sensitivity, all of which might inhibit the growth of hormone-dependent tumors. Furthermore, fiber-rich plant foods contain a variety of other biologically active plant cell constituents, including antioxidants and phytoestrogens, which probably exert a variety of anticarcinogenic effects in their own right.

Phytochemicals

The strongest association between dietary factors and cancer in industrialized societies is a protective effect of high intakes of fruit and vegetables. The phenomenon extends to most carcinomas (cancers of epithelial tissues) but is not consistently observed for leukemia, nor for cancers of lymphoid or neural tissues. Fruits and vegetables are rich in a variety of important micronutrients, including antioxidants and folate, and they probably contribute to the protective effects of these foods (Table 2). However, studies with experimental animals have shown that plant tissues also contain a variety of nonnutrient phytochemicals that can act as blocking agents, which interfere with the mutagenic effects of carcinogens at target cells, or suppressing agents, which slow or reverse the development of precancerous lesions or tumors, after the initial mutagenic effect of the carcinogen has occurred. Some compounds can act in both ways.

Although much of the epidemiological data is relatively nonspecific in that it simply indicates that a variety of vegetables and fruits are protective against an equally diverse range of cancers, there is also a growing body of evidence relating specific plant foods, and hence specific types of chemical compound, to particular disease processes. Evidence of this type forms the starting point for focused mechanistic studies. Some examples of potentially anticarcinogenic phytochemicals and their principal sources are given in Table 1. A thorough discussion of these compounds cannot be given here but one group, the glucosinolates, will be discussed in some detail.

Table 2 Some examples of potentially anticarcinogenic phytochemicals

Class of compound	*Dietary source*
Isothiocyanates	Brassica vegetables and condiments
Flavonoids	Apples, onions, broccoli, tea infusions, chocolate
Isoflavonoids	Soy beans, soy products
Carotenoids	Tomatoes, carrots and other colored vegetables
Protease inhibitors	Legumes

Many cohort and case-control studies indicate that a high consumption of brassica vegetables is associated with a decreased risk of carcinomas of the lung, stomach, colon, and rectum. Probably the most important common characteristic of cabbages, broccoli, cauliflower, and Brussels sprouts is that they all contain glucosinolates, a large family of sulfur compounds linked to glucose, and bearing a side chain derived from various amino acids. Following tissue damage caused by pests, harvesting, food processing, or chewing, glucosinolates are broken down by the endogenous enzyme myrosinase. Hydrolysis of the glucosidic bond releases glucose and an unstable intermediate, which spontaneously degrades to form a variety of breakdown products, the most important of which are the isothiocyanates. These hot and bitter compounds, commonly termed "mustard oils," provide much of the flavor of brassica vegetables.

Isothiocyanates such as phenethyl isothiocyanate, benzyl isothiocyanate, and sulforaphane modify the balance of phase I and II xenobiotic-metabolizing enzymes that are expressed in liver, and in epithelial cells including those of the colon. Phase I enzymes such as the cytochrome p450 family metabolize procarcinogens, often converting them to highly carcinogenic products in the process. Phase II enzymes such as the glutathione transferase family (GST) convert these products to inactive, water-soluble conjugates, readily excreted in urine. In one animal study in which a carcinogenic nitrosamine and phenethyl isothiocyanate (a constituent of watercress) were administered simultaneously, lung tumors were induced in

70% of the control rats given only the carcinogen, but only 5% of those cotreated with isothiocyanate.

One problem with such animal models is that the concentrations of both carcinogens and anticarcinogens used are often unrealistically high. However there is also evidence from epidemiological studies to show that this work has real implications for human health. In a recent study by London *et al*. the excretion of isothiocyanates in urine was used as a marker of brassica vegetable consumption in a large cohort of Chinese men. The presence of inherited defects in two phase II enzymes, *GSTM1* and *GSTT1*, was recorded at the same time as the urinary analysis, and the cohort was followed up for 10 years. Individuals with detectable levels of isothiocyanate metabolites in urine were found to be at a reduced risk of cancer, but the effect was only observable in subjects who lacked *GSTM1* or *GSTT1*, and strongest in those with deletion of both enzymes. The GST enzymes play a major role in the detoxification of carcinogens, but they also break down the isothiocyanates themselves. These observations provide evidence of a very complex interplay between environmental factors, genotype, and risk of cancer. The protective mechanisms may include suppression of the carcinogenic process by induction of apoptosis. Like butyrate, glucosinolate breakdown products block the cell cycle *in vitro*, and induce apoptosis in the rat colon after treatment with a chemical carcinogen. It is interesting to note that a protective effect of broccoli consumption against precancerous adenomatous polyps of the colon has been observed in individuals with the *GSTM1* null genotype.

Conclusion

Clearly there is still a huge disparity between the proportion of human cancers that are theoretically avoidable, and the small number of well-established dietary strategies for actually reducing the risk of cancer in practice. By far the most important cause of human cancer is cigarette smoke. Its avoidance is the surest way for individuals to minimize their risk of cancer. Nevertheless, there are subtle interactions between dietary components, genetic profile, and exposure to carcinogens that shape the individual's risk of disease to an important degree. Understanding such mechanisms in greater depth is the key to harnessing the protective effects of diet more effectively in the future. The current rapid developments in genetics and molecular biology will do much to facilitate this process. In the meantime, the best advice available is to eat large quantities of fruit and vegetables. The current recommendation in many countries is to eat five portions (80 g) of leafy vegetables and fruits per day from a variety of sources. This is already well in excess of the average intake in the UK, but even higher quantities may be of additional benefit. The bulk of such a diet, coupled with quantities of complex carbohydrates sufficient to provide around 18 g of dietary fiber per day, will ensure a relatively low energy density. If combined with regular exercise, this should help to maintain body weight within the desirable range. The most recent evidence suggests that moderating the consumption of red and processed meat products in favor of oily fish and poultry may also help to reduce the risk of bowel cancer. Alcohol consumption should be maintained within current Department of Health guidelines. There is no convincing evidence that very large doses of any particular vitamin are protective against cancer, but in view of the evidence linking suboptimal micronutrient status to some forms of carcinoma, consumption of a daily multivitamin supplement may be prudent.

See also: **Aflatoxins**; **Alcohol**: Properties and Determination; **Ascorbic Acid**: Properties and Determination; **Cancer**: Carcinogens in the Food Chain; **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; Carcinogenicity Tests; **Folic Acid**: Properties and Determination; **Nitrosamines**; **Functional Foods**; **Fatty Acids**: *Trans*-fatty Acids: Health Effects

Further Reading

Ames BN and Gold LS (1998) The prevention of cancer. *Drug Metabolism Reviews* 30: 201–223.

Doll R (1996) Nature and nurture: possibilities for cancer control. *Carcinogenesis* 17: 177–184.

Doll R and Peto R (1981) The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *Journal of the National Cancer Institute* 66: 1191–1308.

Johnson IT and Fenwick GR (2000) *Dietary Anticarcinogens and Antimutagens: Chemical and Biological Aspects*. Cambridge: Royal Society of Chemistry.

Johnson IT, Williamson GM and Musk SRR (1994) Anticarcinogenic factors in plant foods: a new class of nutrients? *Nutrition Research Reviews* 7: 175–204.

London S, Yuan JM, Chung FL *et al*. (2000) Isothiocyanates, glutathione S-transferase M1 and T1 polymorphisms, and lung-cancer risk: a prospective study of men in Shanghai, China. *Lancet* 356: 724–729.

Lutz WK (1999) Carcinogens in the diet vs. overnutrition. Individual dietary habits, malnutrition and genetic susceptibility modify carcinogenic potency and cancer risk. *Mutation Research Reviews* 443: 251–258.

Lutz WK and Schlatter J (1992) Chemical carcinogens and overnutrition in diet-related cancer. *Carcinogenesis* 13: 2211–2216.

Willett WC (2001) Diet and cancer: one view at the start of the millennium. *Cancer Epidemiology, Biomarkers and Prevention* 10: 3–8.

Diet in Cancer Treatment

J W T Dickerson, University of Surrey, Guildford, Surrey, UK

Introduction

The nutritional requirements of persons with cancer, and hence the diet required to meet them, may be affected by the body's response to the disease itself, a condition known as 'cancer cachexia,' and by the methods used to treat the disease. Assessment of the factors involved and their interrelationships can help towards the adequate nutrition of cancer patients and the maintenance of their maximum possible quality of life. Furthermore, it is now becoming evident that some food components may be able to modify the cachectic process, whilst others may have a direct effect on some kinds of cancer cells. This latter effect raises the possibility that diets may be used in the future to treat, or at least to control the growth of some kinds of cancer.

Cancer Cachexia

Weight loss occurs in many but not all patients with cancer. Thus, whilst patients with cancer of the gastrointestinal tract may become progressively wasted, those with breast cancer may maintain or even increase their body weight. Variability in weight change both within the same patient and between different patients has been reported in patients with lung cancer. In some patients, weight loss may be the first indication of the presence of a tumor. This is particularly so in patients with gastrointestinal cancer in whom protein-energy malnutrition (PEM) may develop somewhat insidiously owing to a decrease in food intake, though this may be denied by the patient.

In healthy individuals the body adapts to starvation by conserving protein and reducing energy expenditure. In many patients with cancer these adaptations do not adequately occur and a wide spectrum of changes in nutrient metabolism occur, similar to those found in the so-called 'metabolic response to trauma.' Some of these changes are mediated by cytokines released by macrophages. At least four of these have been described: tumor necrosis factor-α, interleukins-1 and -6 (IL-1 and IL-6 respectively) and interferon-γ. In addition, tumors themselves elaborate catabolic factors which cause direct breakdown of host lipid and skeletal muscle protein stores with the production of acute-phase proteins, such as C-reactive protein. The cachexia syndrome is characterized by anorexia, early satiety, changes in taste perception, weight loss, weakness, anemia, and edema.

Anorexia

A failure of appetite is the most important cause of a low food intake. The etiology of anorexia is, however, often complex with a number of interacting etiological factors, rather than one single factor.

The factors responsible for anorexia in the cancer patient may be psychological, physical, and metabolic. Moreover, the anorexia may be one of three types:

1. Transient anorexia is caused by emotional distress such as during a diagnostic work-up, or when recurrence or metastases are diagnosed, or when pain or discouragement is experienced.
2. Iatrogenic anorexia is related to treatment by surgery, the result of radiation sickness, or a side-effect of chemotherapy.
3. Pathological anorexia is related to the disease, particularly gastrointestinal or bronchial cancers, or part of the anorexia–cachexia syndrome of advanced disease.

The importance of psychological factors may decrease with the progress of the disease as physical factors become more prominent; nevertheless, they may make a substantial contribution to appetite failure. Feelings of worthlessness, loss of self-esteem, pessimism, guilt, and suicidal ideas may be experienced. Anorexia and cachexia may be the physical result of the patient's own beliefs and attitudes towards the disease and its treatment. Patients may become quite unable to overcome a sense of fear and hopelessness, leading to withdrawal.

We know little about the pathology of disturbed appetite control in cancer patients. It is part of the cachexia syndrome and it seems certain that the release of cytokines and associated acute-phase proteins is involved but the mechanism of their action is not clear. In hospitalized cancer patients a low food intake has been found to be significantly associated with a lower quality of life, but whether by cause or effect is not known. Consuming insufficient food to meet energy requirements leads to weight loss and a low body mass index (BMI). In the community, cancer patients with a BMI below $20\,kg\,m^{-2}$ have been found to have higher rates of consultation with their doctor, higher rates of prescription, and higher death rates during the follow-up period compared with those with a BMI of $20–<25\,kg\,m^{-2}$.

Taste Changes

There is strong evidence that taste changes and associated alterations in food preferences occur during the development of neoplastic disease. These can occur irrespective of the location of the primary tumor or

the therapeutic management of the patient. Measurements of taste detection thresholds (TDTs) have shown the threshold for bitterness to be reduced, that is, that patients could taste bitterness at lower concentrations than their noncancerous counterparts. Conversely, patients with cancer had elevated thresholds for sucrose and could only taste sucrose (sweet) at higher concentrations than noncancerous subjects. The changes in TDT were more common in patients with weight loss and in early reports there was a suggestion that they decreased as the energy intake increased. The metabolic cause of the taste changes is not at present clear but studies in end-stage cancer patients with low thresholds to bitter have been found to have an ongoing metabolic response with higher levels of the acute-phase reactant protein, C-reactive protein, when compared with an age-matched group. Moreover, patients with the lowest bitter thresholds had the highest plasma levels of the cytokine, tumor necrosis factor-α. Heightened recognition of household odors has also been reported in patients with end-stage disease.

Attempts to correlate changes in TDT with plasma vitamin A (retinol) and plasma zinc in patients with cancer are complicated by the interpretation of such levels. Plasma retinol levels depend on the availability of retinol-binding-protein (RBP), the levels of which are susceptible to nutritional status, and plasma zinc concentrations are an uncertain measure of tissue zinc status. Magnesium status has been found to correlate with salt detection thresholds.

Taste changes can have important consequences for the feeding of cancer patients. Reduced threshold for bitter taste (as determined with urea) results in beef and pork becoming less acceptable, poultry and fish rather intermediate, and cheese and eggs pleasurable. Taste abnormalities vary between patients and this is one aspect of the condition which needs to be individually assessed.

Taste changes may result in aversions, but aversions may also develop in patients who do not have disturbances in taste. There is some evidence that food aversions may be associated, or learned, by the association of certain foods with side-effects of drug treatment. Whatever the mechanisms by which they develop, aversions to foods may be persistent, lasting for several months, and ignoring them will have a definite effect on food intake.

Energy Expenditure

The resting energy expenditure (REE) of cancer patients has been reported to be increased, decreased, or the same as that of control subjects. Some of this variation is due to the subjects' different kinds of cancer, the cross-sectional nature of the studies, and to the different body composition of the subjects. The resting metabolic rate (RMR) of a group of lung cancer patients has been found to be 108% of predicted values at diagnosis, with a range of 96–151%. In those who experienced a partial or complete remission, maintenance of body weight was associated with a significant decrease in RMR. In contrast, patients who did not respond to treatment and who had lost weight had an unchanged RMR. It seemed that the elevated RMR in those with active disease was a response mediated by cytokines and acute-phase proteins rather than to the metabolic activity of the small tumors themselves.

However, RMR is only part of the total energy expenditure (TEE) and changes in physical activity can exert a greater influence on TEE. It is a matter of common observation that sick people reduce their physical activity and this could counterbalance the increase in RMR. This, indeed, seemed to be so in patients with lung cancer, and this meant that the energy requirement of lung cancer patients is unlikely to be higher than their preillness levels.

Glucose Metabolism

Altered glucose metabolism in the patient who is losing weight is the result of glucose intolerance, changed β-cell function, and decreased sensitivity and responsiveness to insulin. The causes of these changes are not clear but it seems likely that decreased food intake, protein loss, and hormone changes contribute, as do specific characteristics of the tumor. Increased glucose turnover is a composite of increased Cori cycle activity (80%), glucose oxidation (15%) and lipid synthesis, and ketone formation (5%). (*See* **Glucose**: Function and Metabolism.)

Protein Metabolism

Muscle wasting is clear evidence of altered protein metabolism. Most studies on this subject have been performed using experimental animals but insofar as similar studies have been carried out in patients, they have confirmed those in animals. Cancer increases whole-body protein turnover, with an increased rate of protein synthesis in the liver, but a decreased rate of synthesis in skeletal muscle with a simultaneous increase in skeletal muscle degradation. It is likely that a number of factors affect protein metabolism in the cancer patient, e.g., tumor specificity, the staging of the tumor, age, sex, and previous nutritional condition of the patient. (*See* **Protein**: Digestion and Absorption of Protein and Nitrogen Balance; Synthesis and Turnover.)

The biggest challenge in the nutritional management of a patient with cancer is to reverse the changes in protein metabolism which result in the severe muscle wasting so characteristic of cachexia.

Hypoalbuminemia is common in cancer patients and is often interpreted as evidence of malnutrition. However, this is only so if the other causes have been eliminated. These other causes are decreased albumin synthesis, increased plasma volume, and passage of protein-rich fluid into the gut, i.e., protein-losing enteropathy.

Fat Metabolism

Loss of subcutaneous fat is an obvious feature of cancer cachexia. Patients with progressive disease metabolize more fat than those with nonprogressive disease. If a rise in plasma free fatty acids (FFA) can be taken as an index of the mobilization of fat stores, then a significant increase has been found to correlate with the clinical activity of a tumor. However, these findings have not been confirmed. There does seem to be a failure of normal homeostatic mechanisms, possibly owing to insulin resistance, because glucose does not depress FFA oxidation to the same degree in cancer patients as it does in control subjects.

Whatever their cause, loss of body fat, inappropriate energy consumption, and loss of lean body mass can only contribute to a negative energy balance and continuing weight loss. (*See* **Fats**: Digestion, Absorption, and Transport; **Fatty Acids**: Metabolism.)

Vitamins and Minerals

Overt vitamin-deficiency diseases are rare in western societies but their absence is no guarantee of satisfactory status with respect to any vitamin. Patients with malignant disease are, indeed, likely to suffer some degree of deficiency with respect to a number of vitamins which may affect their continuing resistance to the disease and determine their response to treatment. Poor food intake and abnormal losses through the alimentary and renal tracts, possibly coupled with increased requirements which may be drug-induced, may contribute to vitamin depletion. If such deficiencies are suspected during recovery from active treatment, or during palliative care, patients should be given appropriate supplements. However, patients receiving anticancer treatment are advised not to take vitamin supplements, unless under medical supervision, as they may interfere with the action of the drugs.

Some vitamins, particularly those that are antioxidants (vitamin C, β-carotene, vitamin A and α-tocopherol) have been suggested, on the basis of epidemiological studies in humans and experimental studies in animals, to give protection against the development of some kinds of tumors, particularly lung cancer. However, large-scale trials with supplements of these substances have so far yielded adverse and contradictory results and it is considered at present that there is no place for supplements of these vitamins in the prevention or treatment of cancer.

Of the essential minerals investigated in relation to cancer, selenium has attracted most attention, largely because of its antioxidant properties. Though in rodents dietary supplementation has been found to inhibit colon, mammary gland, and stomach cancer, there is as yet no evidence for such an effect in humans. Iron transferrin and ferritin have been linked to cancer risk and cancer cell growth, possibly because many cancer cells require iron for growth. The relevance of this to dietary management seems unclear.

Diet in the Chemoprevention of Cancer

Epidemiological studies have provided evidence that diets low in fat and high in vegetables, whole-grain cereals and fruits are associated with lower risk of colon, breast, prostate, and endometrial cancer. Although such diets are rich in antioxidant vitamins, it seems that their role in cancer prevention is due to the presence in these foods of other substances with anticancer properties. Foremost amongst those at present being investigated are the phytoestrogens (coumestans, isoflavonoids, flavonoids, and lignans). A number of these compounds have been identified in fruits, vegetables, and whole grains commonly consumed by humans and research is under way to determine the quantities present. Studies in humans, animals, and cell culture systems have suggested a number of possible mechanisms for their anticancer action, including estrogenic and antiestrogenic effects, induction of cancer cell differentiation, and suppression of angiogenesis (i.e., extension of existing blood vessels).

Good sources of phytoestrogens are legumes (particularly soybeans), dark green leafy vegetables, nuts (including peanuts), seeds (especially linseed), whole grains and fresh fruits (including apples and bananas) and dried fruits (especially dates, prunes, and raisins). Antibreast-cancer diets including rich sources of phytoestrogens have been proposed. Genistein, a prominent isoflavonoid in soy products, may be responsible for lowering breast cancer risk in Asian women. Other plant constituents which are under investigation for their anticancer properties include indole-3-carbinol and glucosinolates. Indole-3-carbinol is found in vegetables of the *Brassica* genus and

has been found to induce cell cycle arrest of breast cancer cells. Glucosinolates are often responsible for the bitter taste of condiments (e.g., in mustard and horseradish) and are found in turnips, cabbage, and Brussels sprouts. Some glucosinolates are effective inhibitors of cancer induction in rodents by modifying the metabolism of carcinogens. The use of these compounds in humans as anticancer agents may be limited by the fact that some of them are goitrogenic.

Effects of Antineoplastic Treatment on Nutritional Status

Surgery

Not only are cytokines produced as a result of the presence of cancer in the body, they are also produced as a result of surgery, and a decrease in muscle mass characteristic of the cachectic process also occurs as part of the neuroendocrine response to surgery. The initial effect of the cytokines produced by macrophages is on the hypothalamus and this can cause anorexia, even in a previously noncachectic patient. The hypothalamus produces corticotropin-releasing factor which stimulates the pituitary to secrete adrenocorticotropic hormone which causes the adrenals to produce glucocorticoid hormones, and it is these hormones that stimulate the breakdown of muscle proteins. Some of the amino acids produced by muscle proteolysis, principally alanine, are converted to glucose by gluconeogenesis, whilst others are used to synthesize acute-phase proteins, one of which, C-reactive protein, is also produced during the cachectic process caused by cancer. The glucose produced by gluconeogenesis following surgery causes a rise in blood glucose levels which has been described as the 'diabetes of trauma.'

Largely experimental studies have shown that the production and suppression of cytokines can be manipulated by nutritional means, particularly fats, and this offers scope for their potential use in the treatment of severely ill patients.

The extent of these changes, their duration, and the period of convalescence are directly proportional to the severity of the surgery. This is important for the patient with cancer in whom surgery may need to be extensive in order to remove all cancerous tissue. Moreover, surgery is the treatment of choice for cancers of the alimentary tract and may have direct nutritional consequences (Table 1) which will be superimposed on the general 'metabolic response to injury' experienced by all surgical patients. Nutritional management of these patients requires careful attention to specific problems, possibly over an extended period of time.

Table 1 Possible nutritional effects of gastrointestinal surgery

Surgical procedure	*Possible effects*
Resection of the oropharyngeal region	Chewing or swallowing difficulties Prolonged dependence on tube feeding
Esophagectomy	Gastric stasis
Esophageal reconstruction	Reduced acid secretion (both these effects are due to cutting the vagus nerve) Fat malabsorption Diarrhea
Gastric resection (partial or total)	'Dumping syndrome' Malabsorption Reduced acid secretion Reduced absorption of vitamin B_{12} Low blood glucose level
Small intestine: duodenum	Reduced secretion of pancreatic juice and bile
Jejunum	Reduced efficiency of absorption
Ileum	Reduced absorption of bile salts Reduced absorption of vitamin B_{12} Increased urinary excretion of oxalate
Massive resection	Severe malabsorption Malnutrition Metabolic acidosis
Colon	Fluid and electrolyte imbalance
Removal of pancreas	Reduced secretion of pancreatic enzymes Reduced secretion of insulin – diabetes mellitus

Radiation Therapy

The purpose of radiation therapy is to kill cells and the effects may be dramatic. As the cells are destroyed, toxic waste products are liberated, causing an inflammatory response with consequent edema. Cells may be lost from tissue surfaces which then become ulcerated and clinically the patient shows symptoms such as mucositis and stomatitis, nausea, vomiting, cystitis, or diarrhea. Irradiation may have both immediate and late effects on tissues and hence on nutritional status. Normal and tumor cells are affected in similar ways and cell death may be immediate, delayed, or 'natural':

1. Immediate cell death is the result of irreversible damage to DNA; this mechanism is responsible for the immediate side-effects.
2. Delayed cell death occurs after mutation of DNA; limited functions usually continue until the M phase in the multiplication process, when the cells are unable to divide; this mechanism is responsible for late effects.
3. 'Natural' cell death involves the formation of giant (sterile) cells which function but cannot divide.

Some of the nutritional consequences of radiation therapy may be severely disabling, even life-threatening,

Table 2 Effects of radiation and drug treatment of cancer and their palliative treatment

Treatment	*Undesirable effects*	*Possible palliative treatment*
Radiation		
Mouth and throat	Nausea	Small frequent meals; reduced fat
	Vomiting	Total parenteral nutrition (TPN)
	Anorexia	Small meals, served attractively
	Severe anorexia	Tube feeding
	Loss of taste	Emphasis on aroma of food
	Dental deterioration	Reduced sugar intake
	Sore mouth	Bland soft foods; tube feeding
	Lack of saliva	Avoidance of dry foods
	Swallowing and chewing difficulties	Soft, semisolid small meals at correct temperature; between-meal supplements
Upper abdomen	Nausea and vomiting	As above
	Sense of fullness	Small frequent meals; reduced fat
	Reduced secretion of digestive enzymes	Soft, moist, easily digested foods
Lower abdomen	Intestinal cramping	Soft, bland, low-residue diet
	Diarrhea – dehydration	Low-residue foods, high fluid intake
	Malabsorption	Low-fat (?), lactose-free (?), elemental diet; TPN
Drugs		
General effects	Nausea, vomiting, anorexia, diarrhea	As above
Particular effects		
Nitrogen mustard	Metallic taste in mouth	Elimination of any food accentuating this taste; individual tolerance
Cyterabine	Abdominal pain	Bland low-residue diet
Hydroxyurea	Intermittent constipation	Prune juice, large fluid intake (at least 3 l day^{-1})
Vinca alkaloids	Intermittent constipation	As above
	Abdominal pain	As above

Adapted from Dickerson JWT and Williams CM (1988) Nutrition and cancer. In: Dickerson JWT and Lee HA (eds) *Nutrition in the Clinical Management of Disease*, 2nd edn, pp. 350–373. London: Edward Arnold.

and occur in patients whose nutritional status is already compromised. Food intake is substantially reduced and malabsorption resulting from treatment of tumors in or near the gastrointestinal tract may be very difficult to control. The severity of these problems means that aggressive, symptom-related nutritional support (Table 2) is an essential adjunct to radiation therapy in order to prevent deterioration of nutritional status and to allow compensatory hyperplasia of undamaged bowel.

Chemotherapy

Anticancer drugs have profound cytotoxic effects and since they cannot discriminate between 'normal' and 'cancer' cells they may seriously affect 'normal' tissues. Contributions to host malnutrition may be by both direct and indirect mechanisms that cause many symptoms. Current medical practice is to administer drugs cyclically in multiple combinations, sometimes called 'cocktails.' This increases the possibility of 'killing' cancer cells at different stages of cell multiplication, but it can increase both the range and severity of side-effects. Different anticancer drugs cause a slightly different pattern of side-effects (Table 2), but the general effect is to reduce food intake and since this may occur over prolonged periods it will lead to weight loss and debility.

Indirect effects of chemotherapy that may contribute to a reduced food intake include moniliasis or candidiasis. These are particularly common in patients with leukemia and lymphoma and affect the oral cavity, pharynx, or esophagus, causing a sore, painful mouth, odynophagia, and dysphagia.

Possible palliative treatment for some of these problems is shown in Table 2.

Nutritional Management of Cancer Patients

The feeding of cancer patients should be seen as part of their total care. Provision of food for these patients requires the following:

1. Understanding. Carers and providers should appreciate the clinical problems and difficulties experienced by individual patients so that the 'diet,' whether oral, enteral, or parenteral, is suited to the patients' needs, including the effects of the anticancer treatment.
2. Flexibility. Dietary provision must be flexible, with menus to cater for a wide range of needs with respect both to the nature of the food and the timing of its availability.
3. Psychological importance. The provider should be sensitive to the psychological condition of patients and the importance of food to them. This involves

the way in which food is provided, as well as its nature. Food can be a 'messenger' between the carer and the patient.

Oral food is always the first choice and considerable skill may be needed to obtain an appropriate intake. Detailed suggestions for dealing with specific problems are given in some of the works quoted in the Further Reading section. Practical suggestions for increasing energy and protein intakes are as follows:

1. Plan a definite eating schedule and then adhere to it. Do not omit meals or between-meal snacks.
2. Add cereal with banana or other fruit, sugar, and cream to the breakfast menu.
3. Butter toast or bread when it is hot because more butter can be used.
4. Use jam, marmalade, or cheese with toast or bread.
5. Add cream to milk beverages.
6. Add skim milk powder to milk, soups, pudding, mashed potatoes, etc.
7. Add icecream or whipped cream to desserts (cake and icecream, apple pie with whipped cream, etc.).
8. Use mayonnaise, salad dressing, butter, or margarine with sandwiches, salads, and vegetables.
9. Serve thickened gravy with meat and potatoes.
10. Have bacon, ham, or sausages with eggs at breakfast.
11. Add eggs to beverages or recipes to increase the protein.
12. Use milk instead of water to prepared canned condensed cream soups.

Oral food can be supplemented with 'sip' feeds. These may be homemade, milk- and egg-based, or commercially available complete feeds such as Ensure, Clinifeed, or Fortisip. If elemental feeds such as Flexical or Vivonex are needed due to malabsorption, they should never be given orally because of their taste. Patients with inadequate oral intake should be fed through a fine-bore tube (1–2 mm in diameter) inserted nasogastrically, or directly into the stomach or intestine. Feeds used as sip feeds should be introduced from a reservoir either by gravity or with a pump. If the gastrointestinal tract cannot be used, or if the patient cannot tolerate a tube, complete nutrition can be given parenterally through a central vein. Parenteral feeds (for TPN) contain energy (usually fat and glucose), amino acids, vitamins, and minerals. Given under expert supervision, their use in cancer patients before surgery has been found to reduce complications and in some reports to reduce the length of hospital stay. Usually, the greatest benefit has been in the most malnourished individuals. TPN may also widen the scope for radiation and drug treatment. Indeed, the survival of some cancer patients may depend on this kind of treatment. A potential problem with TPN is that it results in a decrease in the barrier function of the gut, but this can be prevented by giving some food enterally.

Conclusions

The feeding of cancer patients is part of their total care at all stages of the disease and its treatment. It is at different times 'supportive,' 'adjunctive,' and 'definitive.' It is supportive as part of the preparation for definitive treatment. It is adjunctive as a part of the overall therapy. In some individuals it becomes definitive as the therapy that permits survival. In the future, it is envisaged that diets containing foods with anticancer properties may be used to treat specific tumors.

See also: **Fats**: Digestion, Absorption, and Transport; **Fatty Acids**: Metabolism; **Glucose**: Function and Metabolism; **Protein**: Digestion and Absorption of Protein and Nitrogen Balance; Synthesis and Turnover

Further Reading

Armstrong F and Mathers JC (2000) Kill and cure: dietary augmentation of immune defences against colon cancer. *Proceedings of the Nutrition Society* 59: 215–220.

Davidson HIM, Pattison RM and Richardson RA (1998) Clinical undernutrition states and their influence on taste. *Proceedings of the Nutrition Society* 57: 633–638.

Dickerson JWT and Williams CM (1988) Nutrition and cancer. In: Dickerson JWT and Lee HA (eds) *Nutrition in the Clinical Management of Disease*, pp. 350–373. 2nd edn, London: Edward Arnold.

Edington J, Winter PD, Coles SJ, Gale CR and Martin CN (1999) Outcomes of undernutrition in patients in the community with cancer or cardiovascular disease. *Proceedings of the Nutrition Society* 58: 655–661.

Gescher AJ, Pastorino U, Plummer SM and Manson MM (1998) Suppression of tumour development by substances derived from the diet – mechanisms and clinical implications. *British Journal of Clinical Pharmacology* 45: 1–12.

Grimble RF (1998) Dietary lipids and the inflammatory response. *Proceedings of the Nutrition Society* 57: 535–542.

Heys SD, Walker LG and Eremin O (1997) The value of peri-operative nutrition in the sick patient. *Proceedings of the Nutrition Society* 56: 443–457.

Jebb SA (1997) Energy metabolism in cancer and human immunodeficiency virus infection. *Proceedings of the Nutrition Society* 56: 763–775.

Kurzer MS and Xu X (1997) Dietary phytoestrogens. *Annual Review of Nutrition* 17: 353–381.

Wheatley C (1998) Vitamin trials and cancer: what went wrong? *Journal of Nutritional and Environmental Medicine* 8: 277–288.

Candies *See* **Sweets and Candies**: Sugar Confectionery

CANNING

Contents

Principles

M R Berry, Cincinnatti, OH, USA
I J Pflug, Minneapolis, MN, USA

Introduction

Canning is the general term applied to the process of packaging a food in a container and subjecting it to a thermal process for the purpose of extending its useful life. An optimal thermal process will destroy pathogenic (disease-causing) bacteria, kill or control spoilage organisms present, and have minimal impact on the nutritional and physical qualities of the food. Although we think of canning in terms of steel or possibly aluminum cans, the principles apply equally well to a variety of food containers such as glass jars, plastic and foil-laminated pouches, semirigid plastic trays or bowls, as well as metal cans of any one of several shapes, including cylindrical, oval, oblong, or rectangular. The concept of aseptic packaging (sterilizing the food and the container prior to filling and sealing) also follows the same principles.

Basic Concepts

In 1810, Nicolas Appert reported the first methods on the thermal treatment of food. His method of preservation was primarily aimed at the elimination of the use of large quantities of sugar, salt, and vinegar as preserving agents because they detracted from the natural flavor and quality of the food. His methods of food preservation developed over the years into procedures that not only prevented the large economic loss associated with microbial spoilage, but also destroyed the food-borne microorganisms that are capable of causing illness, or even death, in humans.

Heat processing coupled with hermetic packaging is used to preserve a wide variety of products. Microbial control processes at temperatures in the 65–95 °C range are often called pasteurization, those from 100 to 150 °C, sterilization. Pasteurization processes are designed to kill pathogenic microorganisms and extend product life under refrigerated storage; sterilization processes make possible indefinite product life at ambient temperatures. Whereas the principles of thermal process design are the same for all conditions, the concepts for process establishment that will follow are those for the sterilization of foods known as low-acid canned foods (LACFs) packaged in hermetically sealed containers. Low-acid foods have a pH greater than 4.6 and a water activity (a_w) greater than 0.85 – a combination capable of supporting the growth of *Clostridium botulinum*, a spore-forming bacterium that produces an exotoxin which is one of the most deadly neuroparalytic toxins known. *C. botulinum* is ubiquitous; it occurs in both cultivated and forest soils, sediments of streams, lakes, and coastal waters, the intestinal tracts of fish and mammals and gills and viscera of crabs and other shellfish. For many years, canning industry laboratories have devoted much attention to *C. botulinum*. Early in the 20th century, thermal-processing technologists divided high-moisture foods into acid (pH less than 4.6) and low-acid (pH greater than 4.6). The basis for this decision was that, at a pH of less than 4.6, *C. botulinum* will not grow and produce toxin. At a pH above 4.6, in a favorable medium *C. botulinum* will multiply and produce toxin. Examples of

foods with a pH greater than 4.6 are vegetables, fresh meats, and seafood. Tomatoes normally have a pH that is less than 4.6 and require a less severe heat treatment (pasteurization) to achieve preservation.

a_w is a measure of the amount of available water in the food. The a_w of fresh fruits, vegetables, and meats is normally greater than 0.85. Dried fruits, honey, and salami have insufficient water content to support the growth of most hazardous microorganisms and thus do not require a sterilization process to produce a shelf-stable product.

Establishment of the Thermal Process

The establishment of the thermal process for sterilization of canned foods results from a successful marriage of microbiological science and physical science, specifically thermobacteriology and heat penetration testing, their validation and iteration, as shown diagrammatically in Figure 1.

Thermobacteriology

Thermobacteriology is the science that studies the potential microbiological contaminants in foods, the relationship between temperature and time levels required to destroy them, and the influence of the food itself on the destruction rates.

There are three microbiological parameters which are involved in all process establishment work, namely D_T, z, and F. These variables define the thermal resistance of bacteria and indicate how much of an effect a particular thermal process is likely to have. The D_T value, which is defined graphically in Figure 2a, is the time in minutes at constant temperature (T) to inactivate 90% (one log reduction) of the target organisms present in a food. The D_T value is also known as the 'death rate constant' or 'decimal reduction time.'

Thermal resistance, or thermal destruction tests (TDTs), that measure D_T are conducted using small food samples inoculated with known levels of microorganisms. The samples, contained in specially designed, low-profile TDT cans or glass tubes, are heated in chambers capable of rapidly heating the sample to a precise temperature, holding for a precise time period, and rapidly cooling to sublethal temperatures. Common heating devices are the TDT retort and the thermoresistometer.

A plot of the thermal resistance (or survival) data must approximate a straight line on semilogarithmic graph paper (as in Figure 2a) for the D_T value to be meaningful. Each TDT curve is unique for the microorganism, food medium, and exposure temperature. The D_T value describes the time effect of heat on a population of microorganisms exposed at constant temperature for a precise time period, without influence of a heating (come-up) or cooling period effect.

The $D_{121.1^\circ c}$ value for *C. botulinum* is normally taken as 0.2 min. This is based on thermal resistance studies conducted in the early 1920s on spores harvested from the most heat-resistant strains known. These studies demonstrated that, by extrapolation from the semilogarithmic survival curve, it was necessary to heat a spore suspension in phosphate buffer for 2.78 min at 121.1°C to reduce the survival population from about 10^{11} spores per unit to less than one spore per unit (12-log reduction). Later, correcting the data for come-up time resulted in a reduction of the heating time to 2.45 min to achieve the same lethal effect, hence, a $D_{121.1^\circ c}$ value of 0.2 min.

The time–temperature data in Figure 3 (see Thermal Process Calculations, below) are typical of the way in which cans of food heat, and illustrates that food in containers does not heat (or cool) instantly. To be efficient in the thermal process design, we must take advantage of the microbial kill at each step along the thermal process path. The thermal resistance curve shown in Figure 2b is the vehicle that makes this possible. A series of TDT tests are conducted to determine the effect of different temperatures (D_T values) on the thermal resistance of an organism. By plotting the measured D_T values on a logarithmic scale against temperature on a linear scale (Figure 2b), a thermal resistance curve is constructed. The thermal resistance curve relates time for a one log kill with the kill temperature. From this plot, the z value can be obtained; it is the inverse slope of the curve and represents the number of degrees of temperature required for the curve to traverse one log cycle. In other words, the z value denotes the number of degrees of temperature required to effect a 10-fold change in the time to achieve the same lethal effect. A higher z value means that a greater change in process temperature is required for the same change in the destruction rate of an organism. The z value makes it possible to quantify the microbial kill at the product temperature that exists at all times during a thermal process.

A range of z values from 7 °C to 12 °C have been measured over the years for *C. botulinum*. These differences are attributed to the spore type (strain), heating system, test substrate, and method of calculation. Much effort has been expended on determining the appropriate z value for LACF process establishment. Consensus led to the conclusion that the use of a single z value of 10 °C – which has been in general use for 80+ years – is still the best recommendation for calculating LACF sterilization processes that are to be safe from a public health standpoint. It is

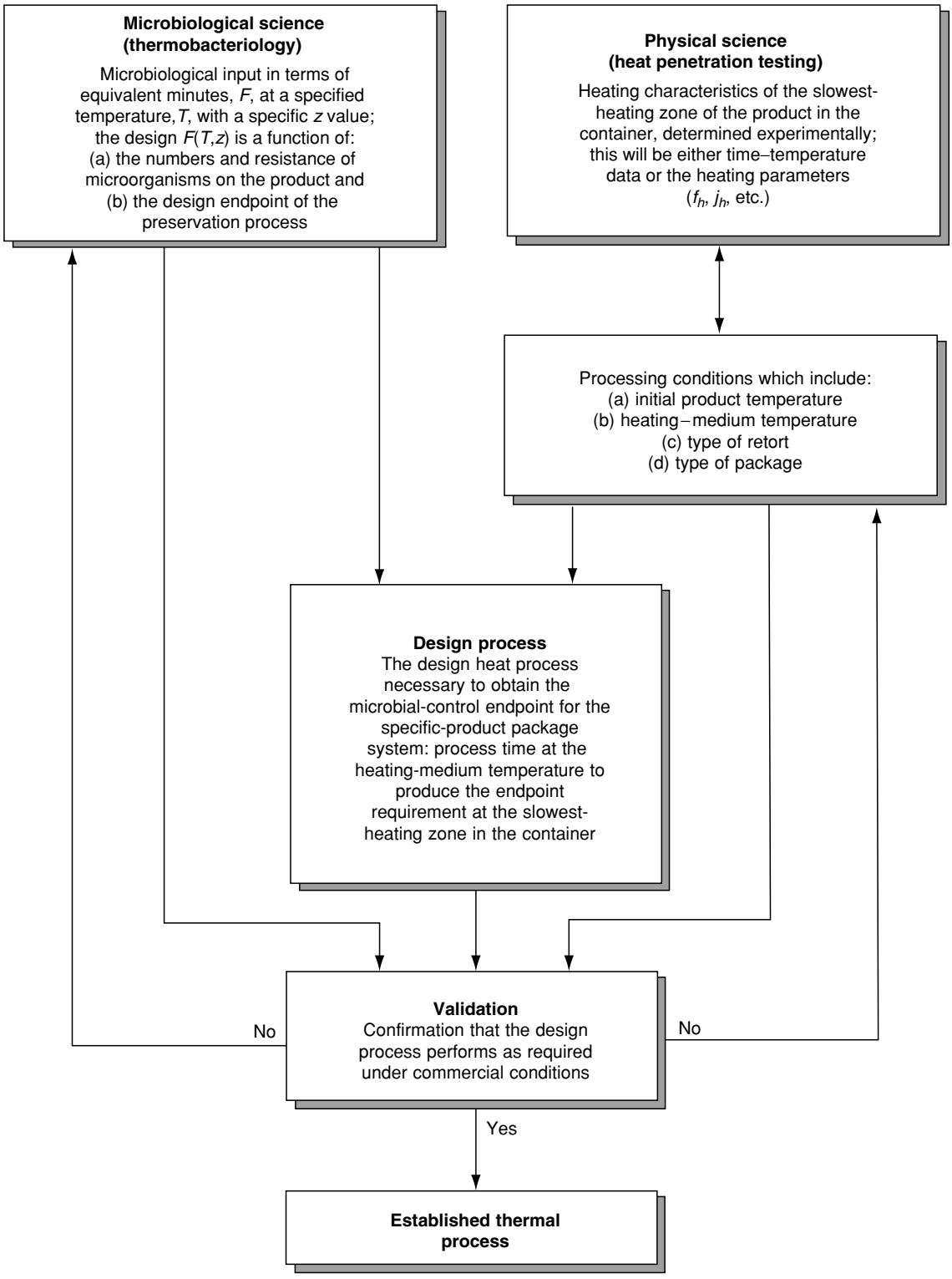

Figure 1 Establishment of the thermal process.

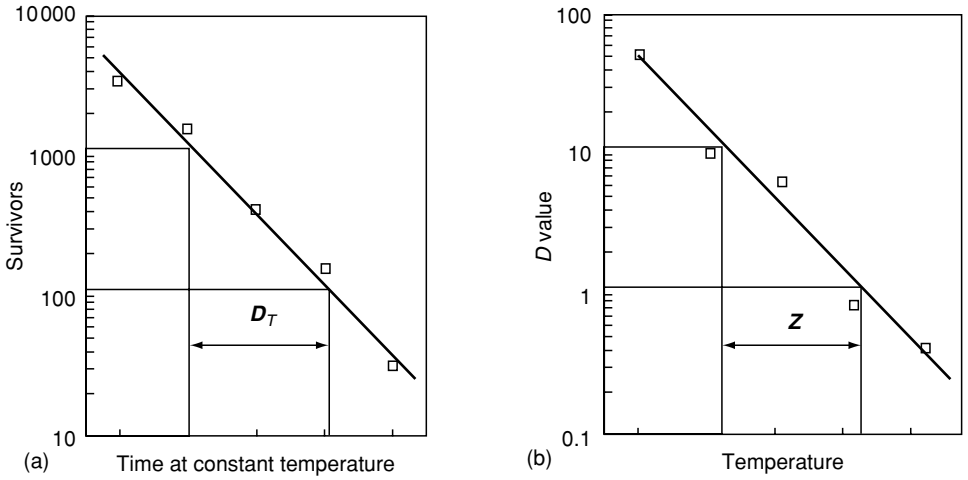

Figure 2 Graphical representations of *D* and *z* values. (a) *D* value; time required at temperature to reduce survivors by 90%; (b) *z* value; temperature change required for a 10-fold change in destruction rate.

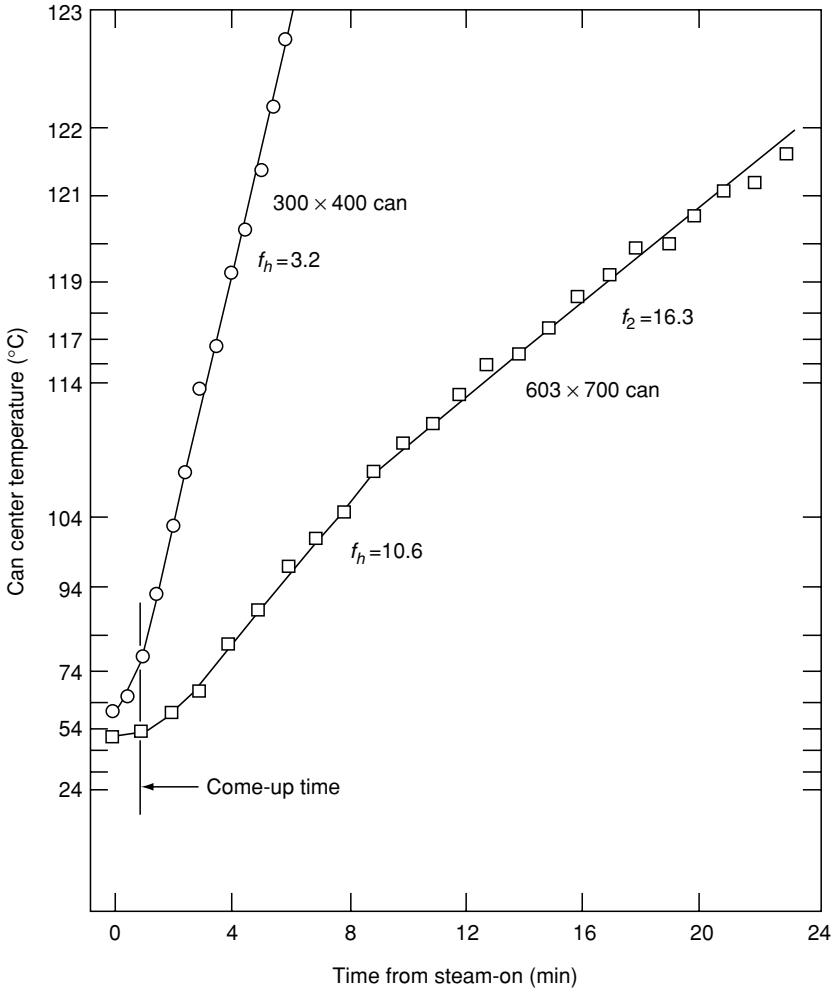

Figure 3 Typical simple and broken heat penetration curves. Tests were for two can sizes of mushrooms in brine, heated simulating the FMC Sterilmatic continuous agitating retort. Can sizes are American designations: 603 × 700 implies an outside diameter of 6 3/16 in (approx. 16 cm) and a height of 7 in (approx. 18 cm). From Berry MR, Bradshaw JG (1982) Heat penetration for sliced mushrooms in brine processed in still and agitating retorts with comparisons to spore count reduction. *Journal of Food Science* 47: 1699, with permission.

possible that this choice was fortuitous; on the other hand, it may have been the purposeful result of research by the canning industry pioneers of the 1920s and 1930s.

Lethality

Containers of food do not heat instantaneously, and since all temperatures (above a minimum value) have a lethal effect and contribute to the destruction of microorganisms, a mechanism to determine the relative effect of a changing temperature while the food is heated and cooled during thermal processing is necessary. The z value is the parameter that allows us to calculate the lethal effect of various temperatures on the destruction of microorganisms. The lethal rate (L) describes, through use of the z value, the relative effect of temperature on microbial destruction with respect to the effect of a certain reference temperature (T_{REF}). In descriptive terms, L is the equivalent minutes at the reference temperature per minute at any temperature T:

$$L = 10^{\frac{T-T_{\mathrm{REF}}}{z}} \qquad (1)$$

Table 1 shows lethal rates at five temperatures for *C. botulinum* assuming a reference temperature of 121.1 °C and a z value of 10 °C, and times required at each temperature for a 12-log spore reduction. If the initial *C. botulinum* population per container (N_0) is 10^3 and we desire a final probability (N_{F}) of 10^{-9}, then a 12-log spore reduction is required. The difference in each temperature in Table 1 is one z value (10 °C), which illustrates that changing the exposure or processing temperature by one z value will require a 10-fold change in processing time.

Sterilization Value

The parameter that accumulates the lethal effect as a function of time (t) during the thermal process is the sterilization value, defined as

Table 1 Lethal rates and times required for a 12-log reduction, applicable to the destruction of *Clostridium botulinum* spores (reference temperature 121.1 °C; z 10 °C)

Temperature, T (°C)	*Lethal rate (min at 121.1 °C/min at T)*	*Time (F_T)[a] required for a 12-log spore reduction*
101.1	0.01	4 h
111.1	0.1	24 min
121.1	1	2.4 min
131.1	10	15 s
141.1	100	1.5 s

[a] $F_T = D_T\, Y_n$ where $D_{121.1} = 0.2$ and Y_n is the spore log reduction ($\log N_0 - \log N_F$) where N_0 and N_F are 10^3 and 10^{-9}, respectively.

$$F^z_{T_{\mathrm{REF}}} = \int_0^t 10^{\frac{T-T_{\mathrm{REF}}}{z}} dt \qquad (2)$$

Or, in terms of lethal rate, L (eqn 1),

$$F^z_{T_{\mathrm{REF}}} = \sum L\Delta t \qquad (3)$$

When temperature (T) characterizes the slowest-heating zone in the container of food and when the reference temperature and z value are 121.1 °C and 10 °C, respectively, then the sterilization value is known as the F_0 value for the thermal process. The F_0 value is specific for the food, container, processing conditions, processing system, and thermal process (processing time, temperature, and other physical factors affecting the process). The F_0 value is the equivalent value of the process in terms of minutes at 121.1 °C, as if no time is involved in heating to 121.1 °C and cooling to sublethal temperatures. An F_0 value of 3.0 min ($z = 10$ °C) is generally accepted as a realistic, minimum botulinum thermal process that will produce LACFs that are safe from a public health standpoint.

Commercial Sterility

Commercial sterility of food means the condition achieved by application of heat which renders such food free of viable forms of microorganisms having public health significance, as well as any microorganisms of nonhealth significance capable of reproducing in the food under normal nonrefrigerated conditions of storage and distribution.

Several additional considerations go into the decision as to the design F_0 for commercial sterility which may be as much as 20 F_0 units higher than the minimum *C. botulinum* public health thermal process. By convention in North America, the F_0 used in calculating a thermal process includes the processing equipment heat transfer safety factor along with the requirement for killing the microbial contaminations. These considerations include: initial bacterial level of the food product, physical parameters of the food itself (style, consistency, particle size, liquid-to-solid ratio, etc.); food container; processing system (still, hydrostatic, continuous agitating retorts, etc.); conditions of storage and distribution; natural or added ingredients that prevent spoilage; economics and the general experience of the food processor. As examples, foods that will be distributed to a high-temperature geographical area may require an F_0 of 15–20 min to afford the same protection from economic loss due to spoilage as an F_0 of 5–7 would afford for a moderate temperature area. An F_0 of 8–12 min is recommended for products heated with induced agitation.

The initial bacterial level of the food has a direct effect on the outcome of a specific process; the same thermal process (F_0) does not guarantee the same process endpoint. The F_0 value is a measure of processing conditions necessary to affect a specific process condition, for example, the level of *C. botulinum* spores by a certain number of log reductions, such as 12 $D_{121.1^{\circ}c}$ values. The higher the initial spore concentration, the higher the spore concentration after processing for a process delivering the same F_0 value.

There is a potential hazard in expressing process requirements as '12*D*,' in that only the spore-log reduction is specified. A spore-log reduction of 12 will yield a probability of a spore surviving of 10^{-9} (one spore per 10^9 containers) only when the initial spore contamination level is 10^3. To provide all consumers of canned food with equal protection, regardless of the initial numbers of *C. botulinum* spores, the heat process F_0 value should always satisfy a constant, agreed endpoint value of probability of a surviving spore.

Heat Penetration Testing

The purpose of the heat penetration (HP) test is to determine, accurately, the temperature of the slowest-heating zone in the food container during thermal processing. The results of the HP test are experimentally determined time–temperature relationships describing the heating and cooling of the product. These data are derived from tests which duplicate commercial processes with a high degree of reliability. The HP data are normally collected in the laboratory because of difficulty in making accurate measurements and having close temperature control in plant equipment. This is especially true for products that heat by natural or forced convection by conduction (i.e., no product movement). The HP test provides the temperature history of the product during the process which, when combined with the thermal resistance information for the organism of concern (the required F_0 value), allows us to calculate the length (process time) of the thermal process at the specified processing temperature.

The factors which affect the HP results are numerous and tend to be more complex as the food product, package, and processing systems (retorts or autoclaves) become more complex. The following factors must be considered by the HP technician during the conduct of a HP test because they may influence the resulting heating and cooling temperature profiles:

- Processing (retort) temperature
- Processing time
- Processing medium (steam, water, etc.)
- Initial temperature and temperature distribution within the container
- Size and shape of container
- Container orientation and distribution within retort
- Agitation of containers during processing*
- Container fill and head space*
- Product formulation and preparation procedures*
- Proportion of solids to liquid*
- Size, shape, arrangement, and composition of food particles
- Consistency of product*
- Product drained weight after processing
- Style of container (plastic or metal; rigid, semirigid, or flexible)
- Vacuum or air remaining in container
- Temperature distribution (uniformity) in processing vessel
- Operating conditions during processing (come-up time, sequence of events, controller function, rotational speed, etc.)*
- Location and type of temperature sensor in container
- Ability of test retort to duplicate commercial conditions*

Items marked with an asterisk are particularly significant when processing with agitation.

Every thermal process will have factors that are critical for delivery of the design F_0. For example, the critical factors of retorting systems that are designed to agitate the contents of the container during processing, to increase the rate of heat penetration, will differ from those of a still retorting system for the same product. It is the responsibility of the person establishing a thermal process to understand all the factors that may influence the way the product heats and cools. It has been repeatedly observed that the HP testing program must continue until all parameters are fully understood. Only accurate and applicable HP factors are meaningful for thermal process establishment.

The instrument of choice for measuring food temperature during HP testing has historically been the thermocouple (TC) with recording potentiometer. Normally nonprojecting style TC receptacles are attached to the container and hard-wired to the potentiometer. The TCs are placed to measure the temperature at the slowest-heating zone within the container; this is determined *a priori* by ancillary testing. Since the objective of the HP test is accurate time–temperature data, care must be exercised in the selection and use of the TC. For products that have considerable natural or induced convection, such as whole-kernel corn in brine, a TC of small diameter is

used in order to avoid interfering with product movement. For conduction-heating foods that remain motionless during processing, such as a viscous stew, the TC support material is selected to have thermal properties similar to the food to minimize the conducting of heat to or from the TC junction. If the TC and/or container are not adequately grounded, especially in water processes, stray voltages may cause large temperature errors.

Temperature measuring systems in use today usually use personal computers as the output device and include resistance temperature devices (RTDs) and miniature telemetry or data logger systems. These systems have allowed HP testing in systems that were previously not possible since they have removed the requirement of direct wiring to the container.

The accuracy of the measuring instrument is extremely important. A 0.5 °C difference in the temperature of the thermal process results in more than a 10% difference in F_0. For minimally processed foods, this could result in severe underprocessing and survival of numerous pathogenic or spoilage organisms.

Thermal Process Calculations

The methods for calculating the sterilization value (F_0) from the HP and TDT data can be classified as either a general or formula method. The two methods use similar principles but the procedures are distinctly different.

The general method is essentially a graphical or numerical integration of eqn 2, using the time–temperature data obtained during the HP test. It is the most accurate and simplest method of determining the F_0 delivered by the thermal process. The disadvantage is that the method affords little or no ability first, to change the process time, heating parameters, or initial product temperature and predict their influence on F_0, or second, to use F_0 as an input to predict required process time. An example of calculating F_0 using the general method is given in Table 2. In this example, eqn 3 is numerically integrated using typical time–temperature data at 2-min intervals from a HP test. The resulting F_0 for the 30-min combined heating and cooling phases (Table 2) is 9.8 min at 121.1 °C. In the general method, improved accuracy may be achieved by reducing the HP data measurement time interval.

The various formula methods are mostly iterations and improvements of the formula method first proposed by C.O. Ball in 1923. The HP data are first plotted on semilogarithmic paper as either simple- or broken-heating curves, as shown in Figure 3. The shapes of the respective heating curves are defined in terms of parameters, commonly known as HP factors: a heating lag factor (j), a temperature response parameter that is a function of the slope of the heating curve (f_h), and the second slope and time of the break point (f_2 and X_{bh}) when the heating curve has a change of slope and can be better represented by two straight-line segments.

The simple (single, straight-line) heating curve normally occurs for food product heating by conduction, or by forced convection induced by mechanical agitation of the container. Broken-heating curves normally occur for product heating by natural convection in still retorts, and for products that undergo a change in their thermophysical properties during processing (such as a rapid increase in viscosity as temperature increases).

In the formula methods, the temperature of the food during the process is described by equations that utilize the HP factors for the heating, cooling, and transitional phases of the processing cycle. When these expressions are substituted into eqn 2, the equivalent minutes at heating medium temperature of the process can be calculated. F_0 values are calculated by either numerical procedures that use high-speed computing equipment or by classical 'cookbook' procedures using supplemental tabulated data. The versatility of the formula method makes it possible to vary the heating time, process temperature, design F_0, and even can size, using the same HP data, and to determine the influence of each of these factors on the thermal process delivered to the product.

Sterilization values calculated using the original Ball formula method are reported to be conservative. Numerous investigators have offered modifications which have improved the method, resulting in F_0 values nearer those calculated by the general method.

Process Validation

The delivered F_0 in the process establishment procedure is a value calculated using experimental data that can be related to the reduction in microorganisms that occurs when the process is used for the commercial production of canned foods.

The final step in the process establishment process is the validation or confirmation of the design process to provide assurance that the design F_0 will be delivered to the product under the commercial processing conditions (Figure 1). It is not possible to measure the design level of sterilization processes (microbial survival probability of about 10^{-9} spores for *C. botulinum* or 10^{-6} for nonpathogenic organisms) using the organism for which the process is intended to destroy. Process validation is normally performed

Table 2 Example heat penetration data and calculation of the sterilization value by the general method (reference temperature 121.1 °C; *z* 10 °C)

Time[a] *t (min)*	*Temperature*[a] *T (°C)*	*Lethal rate L (eqn 1)*	*Lethality* $(L \times \Delta t)$	*Cumulative lethality* F_0 *(eqn 3)*
0[b]	58.0	0.00	0.00	0.00
2	81.0	0.00	0.00	0.00
4	96.0	0.00	0.01	0.01
6	104.0	0.02	0.04	0.05
8	109.0	0.06	0.12	0.17
10	114.0	0.19	0.39	0.56
12	116.0	0.31	0.62	1.18
14	118.5	0.55	1.10	2.28
16	119.8	0.74	1.48	3.76
18	120.7	0.91	1.82	5.58
20	121.6	1.12	2.24	7.83
22[c]	120.1	0.79	1.59	9.42
24	114.0	0.19	0.39	9.81
26	100.0	0.01	0.02	9.82
28	79.0	0.00	0.00	9.82
30	60.0	0.00	0.00	**9.8**

[a]During heat penetration test.
[b]Begin heating.
[c]Begin cooling.

using microbial techniques involving the inoculation of calibrated bacterial spores into the cans before the containers are sealed and processed. After the cans receive the thermal process, they are incubated. At the end of 2 to 4 weeks incubation, all cans are examined, the number of cans that show evidence of microbial growth is determined.

The bacteria used in biovalidation have a heat resistance higher than *C. botulinum* and are typically spore-forming, putrefactive mesophiles or thermophiles. A commonly used organism is PA3679 which is nontoxic and, therefore, safe for use in food plants and not hazardous to microbiologists conducting the validation tests.

Experience in commercial processing indicates that the microbial kill measured by biological methods does not always agree with the measurements of physical parameters (HP and TDT). This is why each process must be validated biologically. If the bacterial spores have been adequately calibrated, they give an indication of the actual killing power of the thermal process as delivered by the commercial processing equipment. A common biological validation is to carry out an inoculated pack where 10 000 resistant PA3679 spores ($D_{121.1°c}$ of between 1.0 and 1.5 in phosphate buffer) are added to each container of product before processing. After processing inoculated containers are incubated. An acceptable process should produce a greater than 5-log reduction of the PA3679. The test must be carried out with good technique and appropriate controls. If the results of the validation tests do not agree with the physical process design, it is an indication that the critical processing parameters were not adequately understood and the differences should be resolved, typically using the iterative process depicted in Figure 1.

See also: **Canning**: Quality Changes During Canning; **Consumer Protection Legislation in the UK**; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; **Microbiology**: Classification of Microorganisms; Detection of Foodborne Pathogens and their Toxins; **Packaging**: Packaging of Liquids; Packaging of Solids; **Pasteurization**: Principles; **Preservation of Food**; **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage; **Storage Stability**: Mechanisms of Degradation; **Water Activity**: Principles and Measurement; Effect on Food Stability

Further Reading

Ball CO and Olson FCW (1957) *Sterilization in Food Technology*. New York, NY: McGraw-Hill.

Downing DL (1996) *A Complete Course in Canning*, 13th edn. Baltimore, MD: CTI.

Goldblith SA, Joslyn MA and Nickerson JTR (1961) *An Introduction to the Thermal Processing of Foods*. Westport, CT: Avi.

Joslyn MA and Heid JL (1967) *Fundamentals of Food Processing Operations*. Westport, CT: Avi.

National Food Processors Association, formerly National Canners Association (1968) *Laboratory Manual for Food Canners and Processors*. Westport, CT: Avi.

Pflug IJ (ed) (1988) *Selected Papers on the Microbiology and Engineering of Sterilization Processes*, 5th edn.

Minneapolis, MN: University of Minnesota Environmental Sterilization Laboratory.
Pflug IJ (1999) *Microbiology and Engineering of Sterilization Processes*, 10th edn. Minneapolis, MN: University of Minnesota Environmental Sterilization Laboratory.
Pflug IJ and Gould GW (1999) Heat treatment. In Lund BM, Baird-Parker TC and Gould GW (eds) *Microbiological Safety and Quality of Food*, vol. 1, pp. 36–63. Gaithersburg, MD: Aspen.
Stumbo CR (1973) *Thermobacteriology in Food Processing*, 2nd edn. New York, NY: Academic Press.

Cans and their Manufacture

M de F Filipe Poças, Escola Superior de Biotecnologia, Universidade Católica Portuguesa, Portugal

Cans

Cans for food and drinks may be manufactured in tin plate, tin-free steel, or aluminum. Depending on the metal to be used and on the type of can, different production methods may also be used to manufacture cans. The characteristics of the metals and the operations to produce and close the cans are described below. Additionally, information is given on the polymeric coatings used to avoid undesirable interactions between the product and the metal surface.

Types of Cans for Foods and Drinks

Metal cans for foods and drinks are usually classified in three-piece cans and two-piece cans. The can components and the terms commonly used to designate different parts of a can are shown in **Figure 1**. Cans in the first group are composed of a welded body and two seamed ends and are usually made in tinplate. The two-piece cans have the body and the bottom end in a single piece and a seamed top end. They are made in tinplate, aluminum, or tin-free steel, and are produced by the Draw–Redraw (DRD) process or by the Draw-Wall-Ironing (DWI) process. The DRD process is used to produce shallow cans, with a low height/diameter ratio, whereas the DWI process is typically used for drink cans, commonly with a high height/diameter ratio. These cans have a very thin wall, thus lacking mechanical resistance. They are used for carbonated drinks where the high pressure from the product (very often around 4 atm) imparts the required resistance. In still drinks, the application of liquid nitrogen in the headspace yields a high internal pressure.

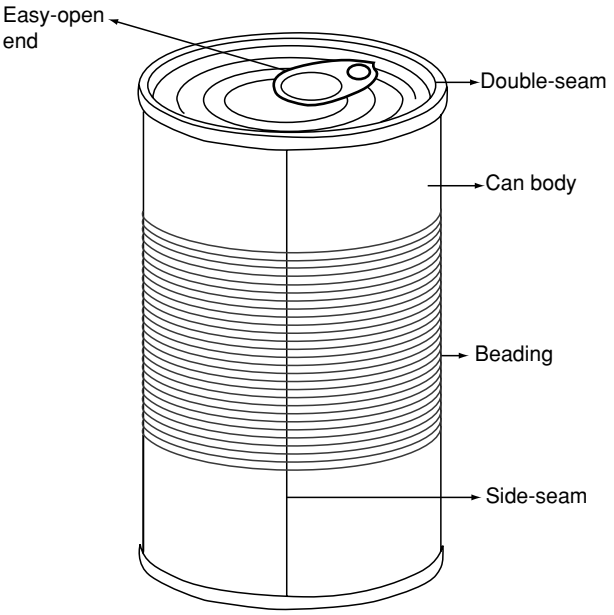

Figure 1 Can terminology.

A wide variety of can shapes are available: round, rectangular, oval, trapezoidal, etc. The circular can is the most popular shape because it is the easiest shape to seam and uses the least metal sheet area for a given volume content. Rectangular shapes are common for processed fish because this format benefits the product presentation when the consumer opens the can.

For a given can capacity, the surface area of metal required for round cans is minimal when the can diameter equals its height. The dimensions of cans are designed taking into consideration this diameter/height ratio so as to maximize the efficiency of metal usage. However, it is easier and less expensive to reset a production line to make a can with a different height than to change its diameter, and thus a standard can diameter system was developed (**Table 1** shows the standard diameters of round cans in both imperial and metric units). Therefore certain can dimensions have been commonly used for certain capacities of cans. The nominal size of round cans is given as diameter × height. The dimensions of rectangular cans are given as three sets of numbers: the first two sets are base dimensions, and the third is the height dimension. The units conventionally used are millimeters for metric units; imperial dimensions are given in three digits: the first digit is in whole inches, and the second two digits indicate 16ths of an inch. For example, a can designated as 307 × 403, is 3–7/16 inches in diameter and 4–3/16 inches in height. **Table 2** shows some of the more widely used cans for foods.

Metallic Materials

Tinplate is the material used more widely in cans for foods. The term tinplate refers to a low-carbon mild steel sheet with a coating of tin on each surface of the material (Figure 2a). The steel varies from around 0.12 to 0.5 mm in thickness. The thickness and hardness of the tinplate are selected as a function of the can size and format, as well as taking into consideration the mechanical solicitations during the thermal processing and handling. In general, the smaller the diameter of the can, the thinner the body wall and ends may be to withstand the imposed loads. Larger cans with thinner walls may be used if wall beading is provided. Double-reduced steel (steel that has undergone a second cold reduction before tinplating) also enables downgauging with no loss of performance. This steel has additional strength, but is less ductile compared with the single reduced plate, which may impose some limitations to the can-manufacturing process.

The tin coating is applied in weights from 1.0 to 11.2 g m^{-2}, with the same or different amounts of tin on each surface. Differentially coated tinplate is identified by marking a set of parallel lines, the line pattern being related to the differential weight combination. Normally, the pattern is applied on the more heavily coated surface, inside the can, where greater protection is needed. Figure 3 shows the tin coating weights used and, in the case of differentially coated steel, the pattern of marking lines.

The steel strip usually has a passivation treatment to render its surface more stable and resistant to the atmosphere, as well as to improve lacquer adhesion. Passivation treatment results in the formation of a top layer of chromium and chromium oxides and tin oxides. After the passivation, the plate is given a light oiling to help preserve it from attack and to assist the passage of sheets through container-forming machines without damaging the soft tin layer.

Tin-free steel (TFS) or electrolytic chromium-coated steel (ECCS) is low-carbon steel coated with metallic chromium and chromium oxides with a weight of around 80 and 20 mg m^{-2}, respectively (Figure 2b). The surface of TFS has a better adhesion to protective lacquer coatings or printing inks and varnishes than tinplate. It requires shorter times in curing enamels (since higher temperatures may be used, due to the lack of a low-melting-point tin

Table 1 Nominal diameters of round cans

Imperial units	*Metric units*
202	52
211	65
300	73
307	83
401	99
404	105
502	127
603	153
610	168
700	176

Table 2 Examples of common sizes of food cans

Type	*Capacity (ml)*	*Dimensions (mm)*
Round three-piece cans	142	55 × 67
	212	65 × 71
	212	73 × 58
	236	65 × 78
	335	73 × 88
	340	99 × 52
	350	83.7 × 69
	425	73 × 109
	850	99 × 118
	945	99 × 123
Rectangular two-piece cans	75	104 × 59 × 19
	125	104 × 59 × 28
	150	154 × 55 × 23
	250	105 × 76 × 38

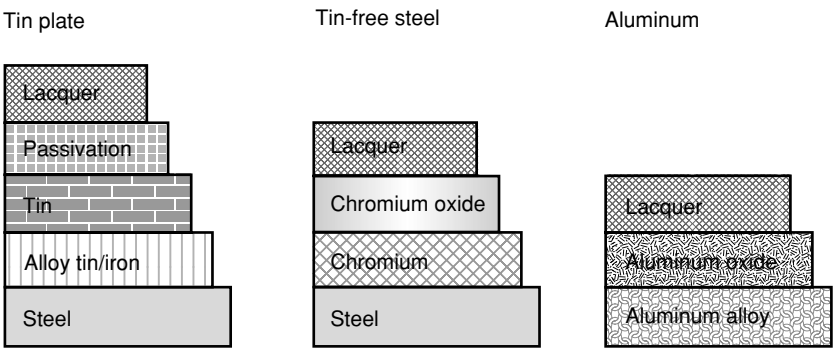

Figure 2 Metallic materials for can-making.

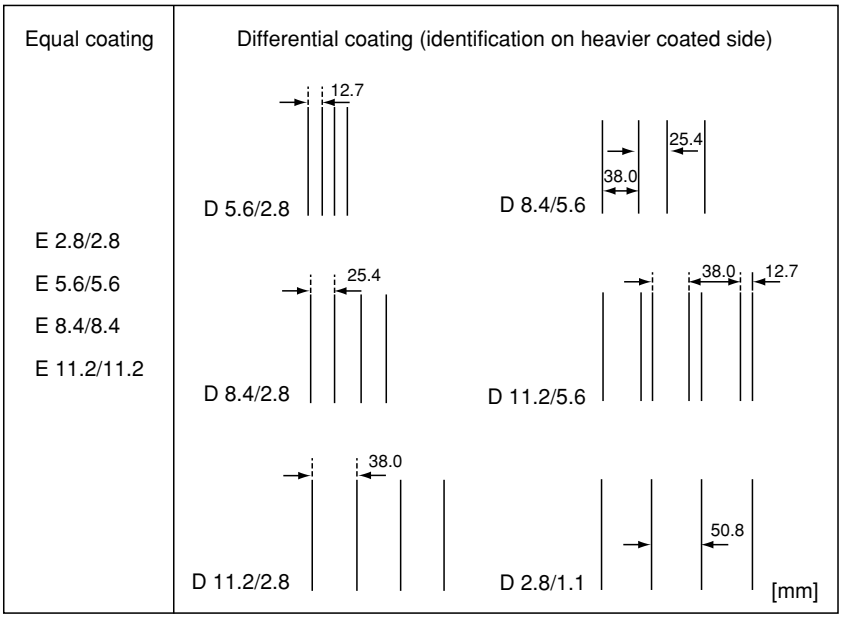

Figure 3 Tin coating weights.

layer) but is less resistant to corrosion than tinplate. TFS is not used in three-piece cans, because it cannot be soldered by traditional techniques, nor in DWI two-piece cans because its hard and rough surface makes drawing operations in uncoated metal impracticable. Therefore, TFS is frequently used for can ends and DRD two-piece cans.

Aluminum is a light metal, easy to work throughout most of the conventional can-forming processes. Aluminum is used for two-piece containers only, DWI and singly or multiply drawn cans, but it is also formed into easy-open ends. Aluminum cans may also be formed by impact extrusion, this type of can normally being used for monobloc aerosols.

Aluminum body stock is available in a range of thicknesses from 0.25 to 0.30 mm, whereas end stock covers the range from 0.30 to 0.38 mm. Most commercial applications of aluminum require special properties that the pure metal cannot provide. Therefore, alloying agents are added to impart strength, improve formability, and influence corrosion characteristics. A wide range of alloys are commercially available for packaging applications, depending on the container design and fabrication method being used. The alloys are identified by four-digit numbers where the value of the first digit indicates the alloy type and the principal alloying ingredient. For example: 2xxx for copper, 3xxx for manganese, 4xxx for silicon, and 5xxx for magnesium. Series 1xxx represents aluminum with 99.00% or higher purity. Commercially pure aluminum (types 1100 and 1050) is used for the manufacture of foil and extruded containers since it is the least susceptible to work hardening. For rigid packaging, the alloy elements used are manganese and magnesium (alloy types 3004 and 5184). When the aluminum surface is exposed to the atmosphere, a layer of aluminum oxides is naturally formed, which protects the metal from further oxidation. This self-protecting characteristic gives the aluminum its high resistance to corrosion, unless it is exposed to some substance that destroys this passivation layer. Aluminum will corrode on contact with products with a pH lower than 4 or higher than 9, and when used for manufacturing cans for food and beverage products, it is always coated with a organic lacquer in the internal surface.

Manufacturing Processes

Steel and aluminum are initially in coil form, so sheet cutting is often the first manufacturing process. This operation can be performed by the can-maker if coil cut-up lines are available in the plant. Otherwise, material must be acquired already in sheet form. Sheets to be used for the production of can ends or round DRD cans are scroll-cut (Figure 4). This is a method used to cut coil into sheets in a pattern, which reduces metal wastage. It enables circular blanks to be packaged more densely: skeleton scrap is reduced by a third compared with straight-cut sheet.

Three-piece Cans

Starting with metal sheets, rectangular in format and with dimensions suiting the can body size (Figure 5a),

the sheets are coated on the side that will become the internal surfaces of the finished cans (Figure 5b). This special lacquer protects the can itself from corrosion and limits the interaction between the contents and the metal. However, welding margins should be left uncoated since the polymeric nature of the lacquer would influence the electric current passage through the metal (see details below). The lacquer is cured as sheets pass in an oven (Figure 5c). The sheet may also be printed for decoration on the other side, unless a paper label is to be used.

The sheets are fed into a slitter, where they are cut into body can blanks (Figure 5d). The blank is then fed to the forming rolls, which generate a cylinder with the edges of the side seam (Figure 5e). The cylinder edges are welded by squeezing them together whilst passing an electric current through them (Figure 5f). This heats up the metal sufficiently for a sound join to be made. To protect the weld seam, a coating is applied immediately after the can body exits the welder, in both sides of the side seam (Figure 5g). Powder coating is usually applied electrostatically to the interior of the side seam.

The cans are passed through a flanger, where the top and bottom of the can are flanged outwards to accept the ends (Figure 5h). Ends, produced elsewhere, are seamed to the can bodies to close one end (called the 'maker's end') of every can (Figure 5i).

Depending on the can design, cans are passed through a beader, where the walls of the cans have circumferential beads formed (Figure 5j). Beading of the can body wall, although reducing the top-load resistance of the can, allows for a reduction in wall thickness without sacrificing paneling performance.

Testing is performed with high-pressure air to detect pinholes and poor seams on every can, and the finished can bodies are then transferred to the warehouse to be automatically palletized before delivery to the filling plant.

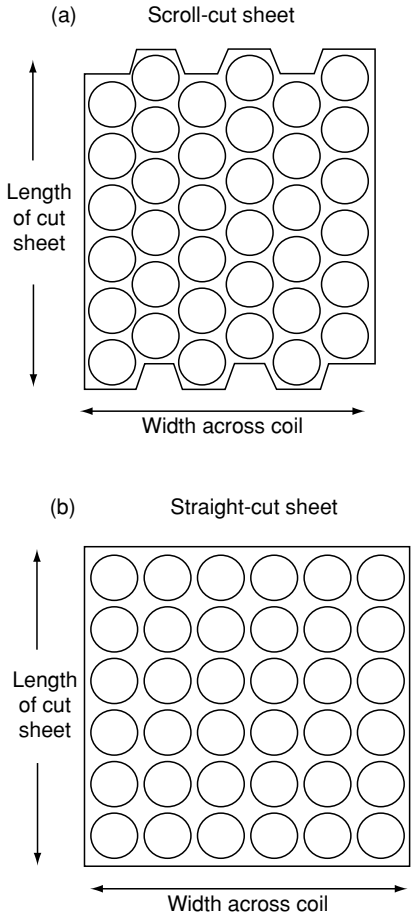

Figure 4 Comparison between scroll and straight cut sheet.

DWI Two-piece Cans

Aluminum or steel coil (Figure 6a) is lubricated with a thin film of oil and then fed continuously through a cupping press (Figure 6b), which stamps and draws disks into cups at a rate of more than 1500 cans per

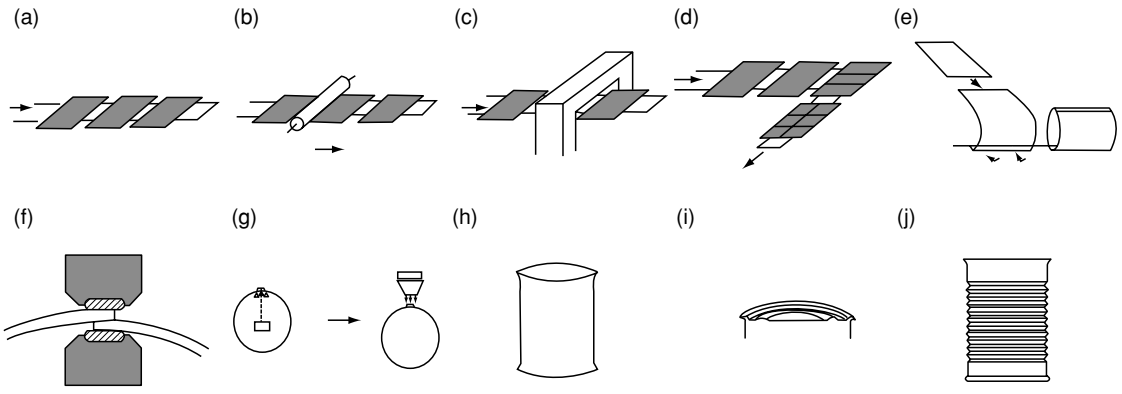

Figure 5 Three-piece can-manufacturing process.

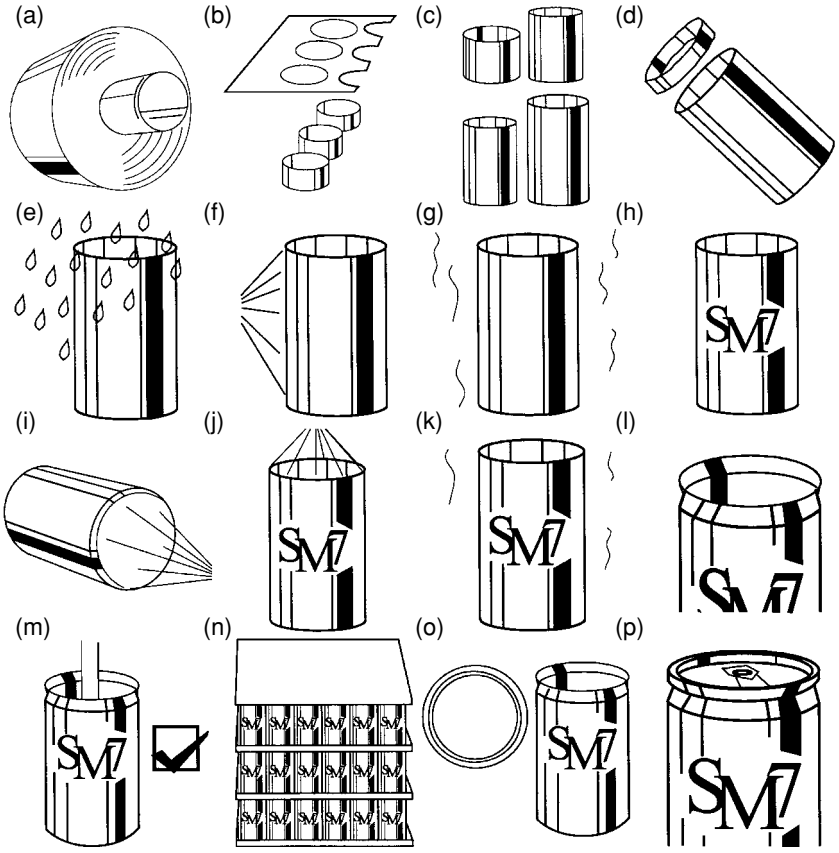

Figure 6 Two-piece DWI can-manufacturing process.

minute. Each cup is rammed through a series of tungsten carbide rings for drawing and ironing. This process redraws the cup to smaller diameter and thins the walls, whilst increasing the height leading to untrimmed cans (Figure 6c).

Trimmers remove the surplus irregular edge and cut each can to a precise specified height (Figure 6d). The trimmed can bodies are passed through washers (Figure 6e) for removal of all traces of lubricant and then dried in preparation for internal and external coatings. The clean cans are coated externally (Figure 6f) with a clear or pigmented base coat, which forms a good surface for the printing inks. The cans pass through a hot air oven to dry the lacquer (Figure 6g).

The next step is the application of the print design in up to six colors, and a varnish, by a highly sophisticated printer/decorator (Figure 6h). A coat of varnish is also applied to the base of each can by the rim coater (Figure 6i). The cans pass through a second oven, which dries the ink and the varnish. Spray machines apply an internal coating of protective lacquer (Figure 6j) to maintain product quality, and the cans are then dried in the final oven (Figure 6k).

The cans are passed through a necker/flanger (Figure 6l), where the diameter of the body is reduced (necked in), and the tops of the cans are flanged outwards to accept the ends after the cans have been filled.

Every can is tested at each stage of manufacture. At the final stage, they pass through a light tester, which automatically rejects any can with pinholes or fractures (Figure 6m). Cans are then moved to the warehouse for palletizing and delivery to the filling plant together with the ends.

DRD Two-piece Cans

This process is similar to the DWI process, except that the final height and diameter of the container are produced by sequentially drawing cups to a smaller diameter, i.e., causing metal to flow from the base to the wall of the container rather than ironing the container wall. The wall and base thickness of DRD cans, as well as the surface area, are identical to the original blank, as opposed to the DWI cans, where the wall thickness is much lower than the base thickness (Figure 7).

The process starts with blanking and cupping, in which cups are produced from prescrolled sheets, lacquered, and/or decorated. The cups then go

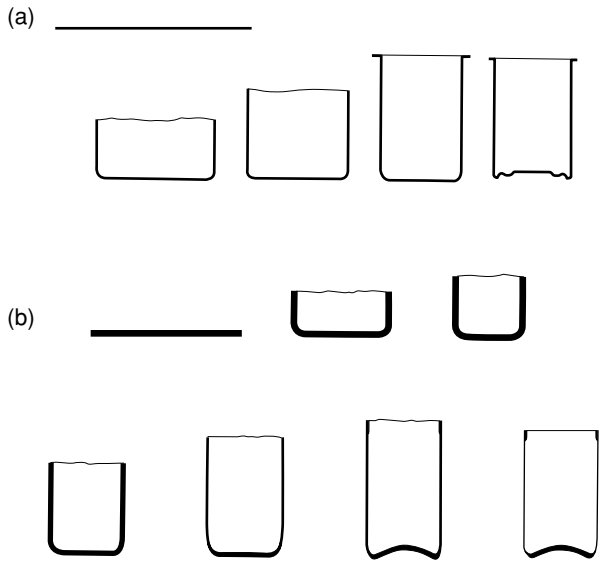

Figure 7 Comparison between (a) DRD and (b) DWI processes. From Bakker M and Eckroth D (1986) *The Wiley Encyclopedia of Packaging Technology*. New York: John Wiley with permission.

through the first drawing operation, where they are drawn by a punch through a die. The cups produced at this stage may undergo a similar redraw process for further diameter reduction as the height is progressively increased (Figure 7b).

There is a linear relationship between the maximum height to which a container can be drawn in a single operation and its diameter. The taller the can, the more drawing operations are required. Also, single-reduced steels can be drawn more than double-reduced, which can be used in lower gauges for the same strength.

The bottom end is the thickest region and governs the material gauge, often resulting in an excessive side wall thickness. Typically, a prelacquered tinplate and TFS with a thickness of 0.2 mm, is used for the DRD process. DRD cans are currently used in the packaging for food rather than beverages, where a greater wall thickness is required to withstand pressure reversals. The body is beaded, and TFS is used more than tinplate because a better enamel adhesion is achieved with the former.

Ends

The can end is designed for optimum deformation behavior during the food-heating process, thereby preventing permanent distortion: it should resist the high relative internal pressure achieved in the heating process that tends to deform the can ends outwards, and it must also be sufficiently flexible to return to its original profile as the internal pressure becomes slightly negative inside the processed can. This behavior depends on

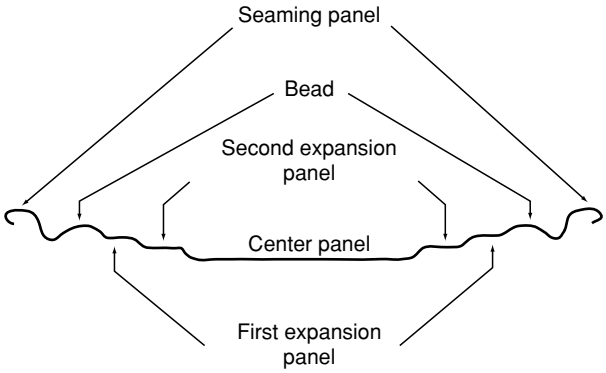

Figure 8 Can end profile.

the metal thickness, the profile of the expansion rings, and the countersink depth (Figure 8).

The scroll-cut sheet, previously lacquered, is fed through a press that stamps the ends. After stamping, the edges are curled and passed through a lining machine, which applies a precise amount of a sealant compound around the inside of the curl. This sealant assists the formation of a hermetic seal by providing a gasket between adjacent layers of metal in the double seam.

Easy-open ends are produced, from a previously stamped shell (Figure 9a), in a press comprising four basic operations: a bubble-like structure is drawn (Figure 9b), a cylindrical button is drawn from the bubble (Figure 9c), the score line is performed (Figure 9d), and the tab is attached at the rivet station by compression of the button, the tab being formed on a separate operation (Figure 9e). In this way, the tab is affixed to the end without any loss of end integrity. Figure 9 illustrates the basic operations.

Double Seaming

The end is joined to the can body by a double-seaming operation that is carried out in two stages (Figure 10): in the first operation, the end curl and the body flange are brought together and rolled inwards so that the end hook is well tucked up underneath the body hook. The shape of the seaming roll determines the contour of the seam at the end of this stage; in the second operation, the seam is tightened by a second set of seaming rolls. The final quality of the double seam is defined by its length, thickness, and the extent of the overlap of the end hook with the body hook (Figure 11). Rigid standards are laid down for an acceptable degree of overlap and seam tightness.

Welding Side Seams of Three-piece Cans

The side seam is made by a resistance-welding process, using the lost-wire-electrode principle. After the can body blank has been formed into a cylinder over

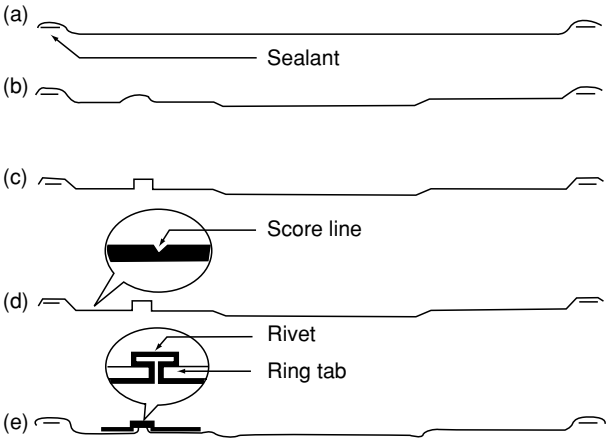

Figure 9 Easy-open end production.

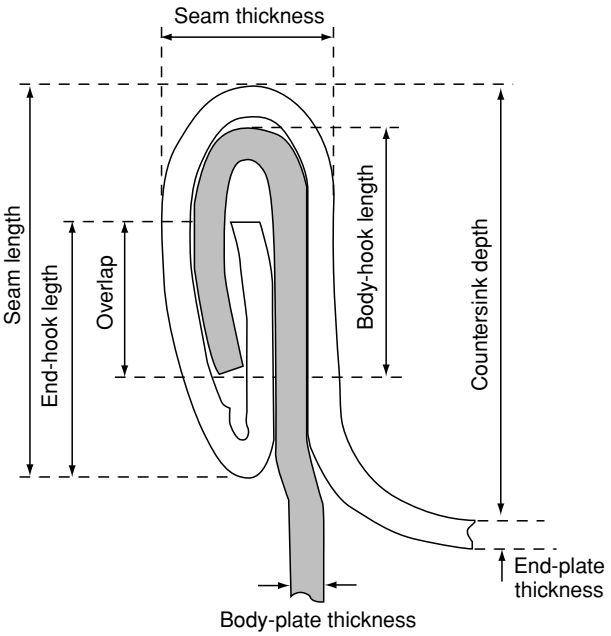

Figure 11 Main components of a double seam.

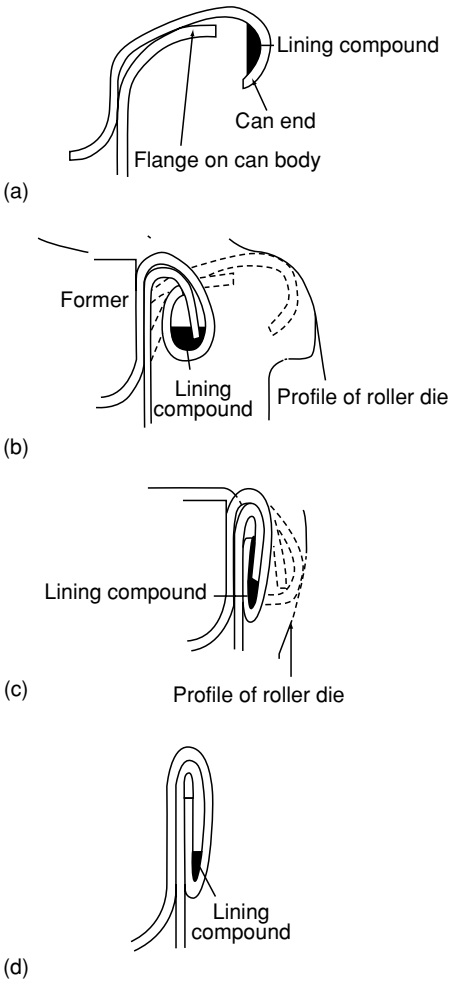

Figure 10 Double-seaming: (a) and (b) first seaming operation; (c) and (d) second seaming operation.

the welding arm, the overlapping seam (0.4 mm for most cans) passes between two copper weld rolls where the weld is formed (Figure 12).

The two layers in the seam are squeezed together between the upper weld roll located outside the can body and a smaller-diameter inner weld roll. An alternating current is passed to the upper roll, and high electrical resistance causes the interface temperature to rise rapidly to at least 900 °C, resulting in solid-phase bonding. As the can body moves continuously between the two weld rolls, a spot weld is created at each peak of the alternating current. By adjusting the frequency of the current to the linear speed of the can body, the correct number of spot welds per unit of can length can be achieved. In a good weld, the individual spots should merge into a continuous weld. Although the welded seam is free from the danger of lead pick-up, the weld has to be effectively coated to prevent any traces of iron being picked up by some types of beverages and acidic foods. The external side of the seam is also coated for protection.

Protective Organic Coatings

The primary function of interior can coatings, enamels or lacquers, is to prevent any interaction between the can and its contents, although some enamels have special properties, and others are used merely to improve the appearance of the pack. Exterior can coatings may be used to provide protection against the environment, as well as for decoration and product labeling.

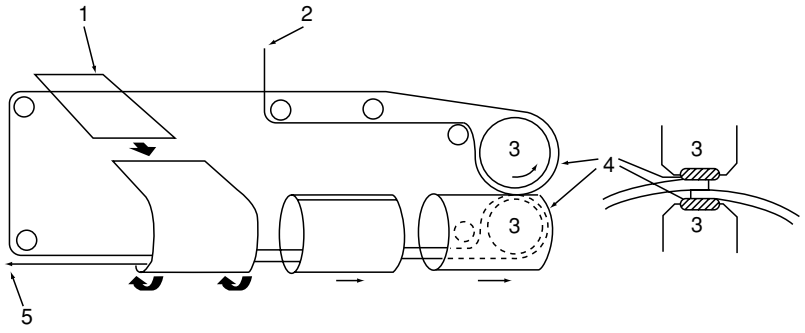

Figure 12 Welding side seam of a three-piece can: 1, blank feed; 2, wire fed; 3, welding rolls (electrodes); 4, copper wire (welding contact); 5, used wire. From Bakker M and Eckroth D (1986) *The Wiley Encyclopedia of Packaging Technology*. New York: John Wiley, with permission.

Internal lacquers must be inert, must provide a good barrier, and must also have a good mechanical resistance, as well as thermal resistance if the product is to be thermally processed. The lacquer must comply with the regulations for materials intended for food contact, i.e., only approved substances that had sustained successful migration or extraction tests and that do not impart any flavor to the contents, may be used.

The barrier provided by the lacquer depends on its chemical composition and on its porosity (which must be as low as possible) and which in turn depends on the thickness of the lacquer. Control of the amount of lacquer applied, usually measured in terms of the mass of dry film per area of metal sheet, is very important since a thin layer will not cover the surface completely, and a layer that is too thick will lead to brittleness, lack of flexibility, and poor adhesion, as well as being uneconomical.

The lacquer must be resistant to mechanical solicitations during can and ends manufacturing operations, as in the case of three-piece cans and most DRD two-piece cans that are produced from a precoated metal sheet. Flexibility and adhesion are also important characteristics of lacquers, as during retorting, the can tends to deform, due to a pressure imbalance.

The lacquer is usually applied by roller coating to the metal in the flat form sheet or coil before fabrication. Spraying is used for coating the internal surface of DWI two-piece cans that cannot be produced from prelacquered sheets, owing to the considerable amount of metal deformation and surface disruption occurring during ironing operations. Powder coating, where the resin is applied dry in the form of a fine powder, usually under the direction of an electrostatic field, is used for the protection of welded side seams (internal side), where heavy coatings are required. A postfabrication repair lacquering is applied to cans used to pack products with very demanding protection requirements, to compensate for natural application imperfections and layer damage during can fabrication.

After application, the lacquer must be dried by solvent removal, oxidation, and/or heat polymerization. This process is usually performed in a forced convection oven using hot air up to 210 °C for up to 15 min. Recent developments include resin formulations, which need lower temperatures and shorter curing times, often through the use of ultraviolet radiation (these ultraviolet-curing resins are virtually solvent-free).

The more common chemical families of protective lacquers are the phenols, epoxy, polyesters, and vinyl. The resins are rarely used pure. They are modified and/or combined with other resins and additives that confer specific characteristics. Table 3 summarizes the major lacquers' characteristics and applications.

Phenolic resins are produced by the action of formaldehyde on phenol and are characterized by having a high degree of reticulation enhancing barrier, hardness and chemical inertia, but limited flexibility due to the presence of the bulge phenolic ring.

Epoxy resins are produced by polymerization of bisphenol A with glycerol epichlorohydrine. Epoxy is highly reactive, allowing for combination with other resins, and it has a good chemical resistance. It is very flexible because of its linear structure, without reticulation, and it is often combined with phenolics giving the epoxy-phenolic family, one of the most widely used families of resins for lacquering. A three-dimensional structure is formed during curing, either by straight blending of a solid epoxy resin with a solid phenolic resin or by condensation of a mixture of two resins in appropriate solvents. Epoxy-phenolic lacquer combines the good adhesion properties of the epoxy resin with the high chemical resistance properties of the phenolic resin. The balanced properties of

Table 3 Application and characteristics of lacquers

Family	*Application*	*Flexibility*	*Adhesion*	*Resistance to heat-processing*
Phenolic	Fruits, vegetables, meat: very good barrier	Bad	Bad	Very good
Epoxy-phenolic	Wide use, can be pigmented with Al, ZnO; characteristics depending on formulation	Good	Good	Good
Vinylics	Beverages	Excellent	Good	Bad
Organosols	Large use in two-piece cans	Very good	Very good	Good
Acrylics	Pigmented with TiO_2	Good	Very good	Medium
Epoxy-urea	Beverages	Good	Good	Medium
Polyesters	Pigmented with TiO_2	Medium	Good	Good

epoxy-phenolic coatings have made their use almost universal in food-can applications. Epoxy may also be combined with amines to produce epoxy-urea and with fatty acids to produce epoxy-esters.

Vinyl coatings are based on copolymers of vinyl chloride and vinyl acetate. They can be used with or without pigmentation, and they can be blended with alkyd, epoxy, and phenolic resins to enhance their performance. Their main disadvantage is the high sensitivity to heat and retorting processes, restricting their application to cans that are hot-filled and to beer and beverage products.

Vinyl organosol coatings are dispersions of PVC (homopolymer), sometimes reinforced with soluble thermosetting resins, such as epoxy or phenolic, to enhance thermal resistance and adhesion. Plasticizers are also added to aid film formation. These coatings are typically white due to the addition of titanium dioxide.

Polyesters have an excellent resistance to high temperatures and are often used for can external coating. When used for heat-processed foods, they are modified with phenolic resins and often white pigmented with titanium dioxide.

The choice of lacquer depends on the nature of the food and the can to be used. Products that contain sulfur-containing amino acids require sulfur-resistant lacquers to prevent staining of tinplate surfaces. Sulfur compounds may break down during heat processing to release sulfides, which react with tin to form brown–violet tin sulfide (SnS) that affects the global can surface or with iron to form iron sulfide (FeS) in spots where, due to lacquering and tin-coating imperfections, iron may be in contact with the food. These compounds do not affect the product safety or quality but are suspicious to the consumer. To overcome this problem, lacquers offering a good physical barrier, such as phenolic or epoxy-phenolic, are used. In addition, these are pigmented with zinc oxide, which reacts preferentially with the sulfur compounds, thus acting as a chemical barrier. Lacquers may also be pigmented with aluminum powder or white pigments, to obscure any tin sulfide that might be formed.

Acidic foods require a very good protection to prevent can corrosion. A vinyl organosol with a high thickness, pigmented with aluminum powder or titanium dioxide, or a double coating of epoxy-phenolic plus vinyl organosol may be used for this purpose.

Beer and other drinks susceptible to flavor contamination by metallic traces require very good protection as well. DWI cans are coated after fabrication, usually with two coats of lacquer: epoxy-phenolic plus vinyl, epoxy-urea plus vinyl or a double layer of epoxy-urea.

See also: **Canning**: Principles; Food Handling; Quality Changes During Canning; **Packaging**: Packaging of Liquids; Packaging of Solids

Further Reading

Bakker M and Eckroth D (1986) *The Wiley Encyclopedia of Packaging Technology*. New York: John Wiley.

Robertson GL (1993) *Food Packaging Principles and Practice*. New York: Marcel Dekker.

Soroka W (1995) *Fundamentals of Packaging Technology*. Herndon, VA: Institute of Packaging Professionals.

Recent Developments in Can Design

M de F Filipe Poças, Escola Superior de Biotecnologia, Universidade Católica Portuguesa, Portugal

Short History of the Can

Metal cans have been on the marketplace since the early nineteenth century, and this early-born packaging system keeps its relevant position in the food packaging area owing to constant developments and

innovation to meet the demands from the food and beverage industry, distribution, and the consumer. Recent developments in can design include lightweighting, shaping cans, double seams, the use of polymer coated metals, and new can opening features.

The start of the history of the can is attributed to Nicholas Appert as a result of the method he developed for storing and sterilizing food using glass bottles. This was followed by the registration of the first patent, by Peter Duran, for food processing using iron and tin containers. Since then, the modern can has evolved from a simple cylinder with two flat ends into a sophisticated package, with a design that incorporates features to improve performance, convenience, and appearance, as well as to reduce costs. Environmental protection has also played a role as a driving force in can design, particularly in issues related to the reduction of metal and energy usage, and air emissions and liquid effluents on can-manufacturing operations. Table 1 summarizes the major landmarks on food and drink can history, and Figure 1 shows, as an example, changes in Coca-Cola beverage can design over decades. These include changes in the easy-open end, can body neck-in, and innovations in can shape.

Table 1 Brief history of can making

Period	*Innovation*
1809	Nicolas Appert wins Napoleon's prize for his method of processing food
1810	Peter Durand patents the idea of using tinplate for making a container
1853	The canned condensed milk is introduced
1900s	The first can with soldered side-seamed and double-seamed ends is produced
1930s	Cans for beer and soft drinks appear
1942	Electroplating for applying tin coating is introduced
1940s	The draw–redraw process is developed
1947	Draw-wall-ironing aluminum cans are introduced
1959	The Soudronic Company is founded (side-seam welding machines)
1962	Aluminum easy-open ends for beverage cans are produced
1965	Tin-free steel is available
1975	The Wima (wire mash) welding system is developed by Soudronic

Lightweighting

Lightweighting means the reduction in the amount of material used in the can-making process. This may be achieved at the expense of a reduction in the metal thickness and/or improvements in can construction, leading to savings on the metal sheet surface required for each can. Statistics show that lightweighting is a fact. For example, over the last 20 years, a typical 33 cm^3 steel beverage can has become on average 30% lighter (Figure 2). In general, the steel thickness of the can body has decreased by around 4.5 μm per year, and the steel thickness of can ends has decreased by 2 μm per year, i.e., a 40 and 18% reduction, respectively, in thickness over the last 20 years.

Downgauging has been possible through the use of double-reduced steels and through can wall beading. Double-reduced steel has suffered a second cold-reduction process, between annealing and tinplating, where the thickness is reduced by about 35%, whereas the elastic limit of the steel has increased to more than 500 $N\,mm^{-2}$. Therefore, double-reduced steels have additional strength although at the expense of their ductibility and formability properties. This process allows for a 3/100 mm reduction in steel thickness without compromising mechanical performance. Double-reduced steels have been used for can bodies and more recently for can ends.

Beading has the effect of reinforcing the side wall, thus allowing for a reduction in metal thickness without compromising paneling performance. The

Figure 1 Coca-Cola beverage can design over decades. From The CanMaker November 98, Copyright 1998 Sayers Publishing Group Ltd. Reprinted with permission.

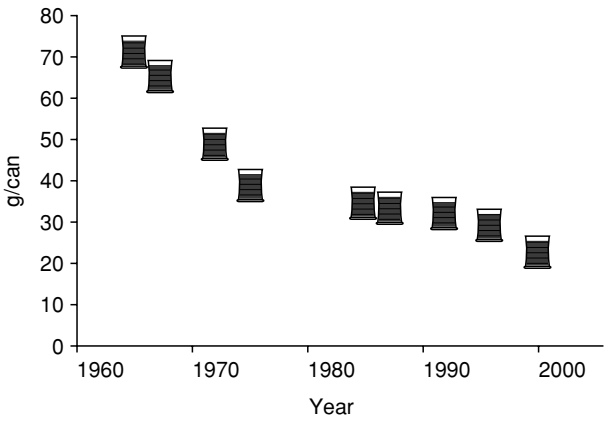

Figure 2 Weight of a 33 cm^3 steel beverage can. Source: Apeal.

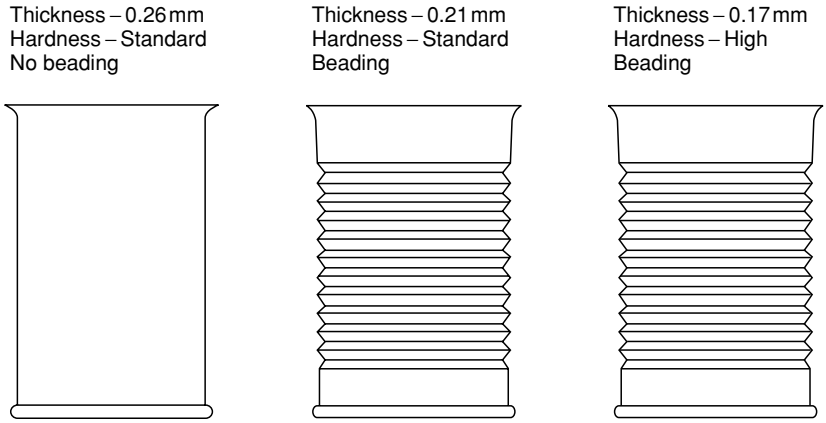

Figure 3 Relationship between beading and metal thickness and hardness in can design.

retorting process of thermal processed food induces positive and negative pressures within the can at different stages. The can must withstand high internal pressures that occur when the product is heated, this resistance being mostly conferred by the expansion rings of the can ends. It must also withstand external pressure to paneling owing to the pressure difference during the early stages of retorting, while there still is a slight vacuum inside the can, and the steam pressure acts outside the can. Circumferential beads on the walls greatly increase can resistance to implosion. However, beading also reduces the resistance to vertical load, which is important during filling, storage, and handling, and so a compromise between beading profile, pitch, and depth must be achieved. Beading, thickness, and hardness of the metal sheet are three variables that the can designer may use to optimize the can performance. Figure 3 shows how these three factors can be combined in the design of a specific can.

Beading was first introduced in 1970 and was commonly used in 1975–80. More recently, a new beading design was developed by Hoogovens and Impress Metal – The Varybead (Figure 4). This design has a constant bead depth of 0.4 mm and a varying bead pitch from 3.67 mm at the extreme ends of the 20 beads to 2.68 mm in the middle. The end beads control the mode of axial collapse, and the middle smaller beads maximize the paneling resistance.

Aluminum drink cans have also been reduced in weight. The number of cans produced from a kilogram of aluminum increased by 35% between 1975 and 1995. Figure 5 shows the net weight of aluminum required to make 1000 cans from 1990 to 2000. Lightweighting has been possible by decreasing the starting gauges of the aluminum coil (Figure 6), by improvements in the draw-wall-iron process, and by using smaller diameter can ends, which is related to the can body neck-in. Downgauging of the aluminum coil requires refinement in dome design as the starting gauge of the metal determines the can bottom dome thickness, critical to the performance of the can in terms of its resistance to internal pressure.

Figure 4 New beading design – the Varybead. From The Can-Maker December 1998, Copyright 1998 Sayers Publishing Group Ltd. Reprinted with permission.

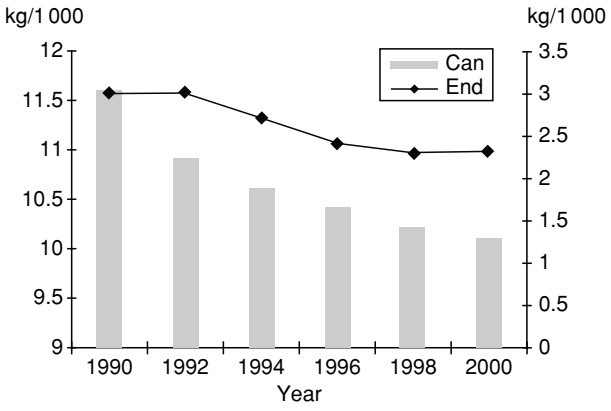

Figure 5 Number of cans and ends produced per kilogram of aluminum. Source: Alcoa.

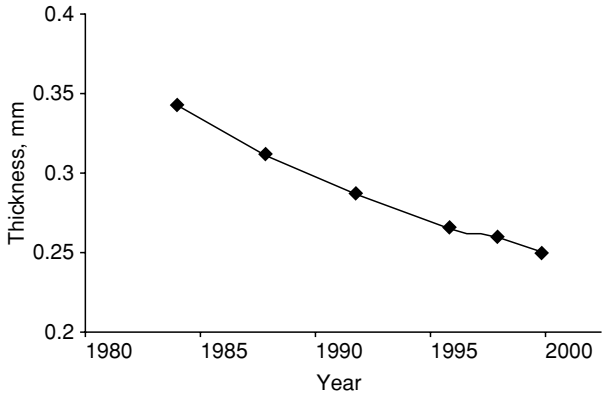

Figure 6 Starting gauges of aluminum coil for beverage cans. Source: Redicon Corporation.

Double Seam

A double-seam fixes the end to the can body and provides a hermetic seal. It is one of the most critical elements in the safety of canned shelf-stable foods, and great effort is devoted to controlling the double seam dimensions, in compliance with rigorous standards and tight tolerances.

Developments to reduce costs have led to smaller seam profiles. The Microseam was developed by Kramer in Brazil, and has a structure similar to the conventional double seam. However, the dimensions of the seam height, body, and cover hooks are 50% of a conventional seam. The design is only possible when using double-reduced steel (0.16 mm or less) for can body and ends. When applied to a 73-mm-diameter three-piece can, it allows for the end diameter before seaming to be reduced from 88.5 to 81.5 mm, and the blank body height to be reduced from 98.7 to 93.5 mm (Figure 7). The main advantages claimed

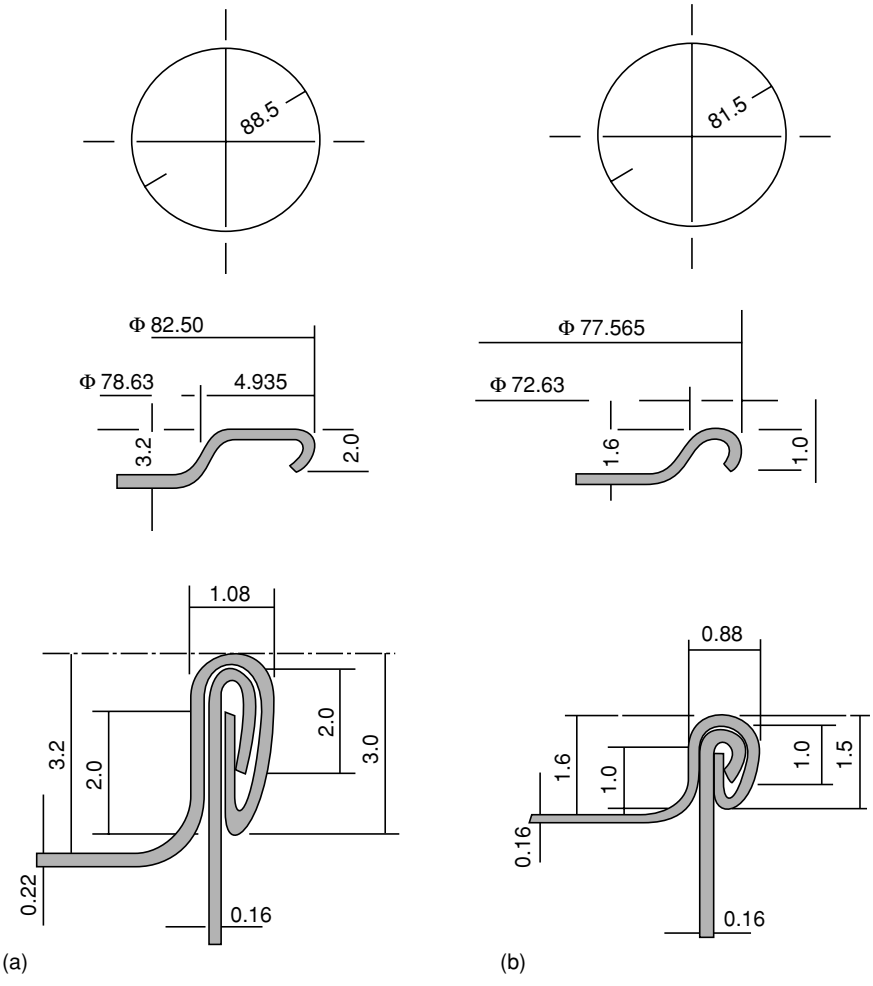

Figure 7 Comparative dimensions of (a) conventional and (b) microseam three-piece can double-seam. From The CanMaker September 1992, Copyright 1992 Mayo Sayers Ltd. Reprinted with permission.

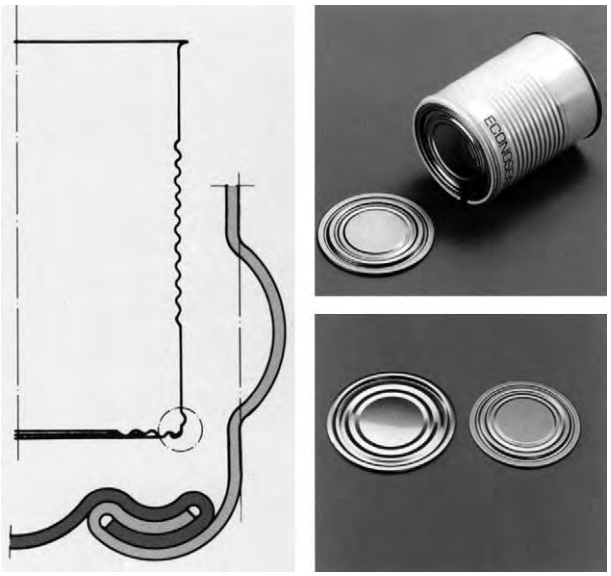

Figure 8 Novel seam design – Econoseam.

Figure 9 Hexacan with honeycomb beading. From The Can-Maker January 1997, Copyright 1997 Sayers Publishing Group Ltd. Reprinted with permission.

are a stronger marketing appeal, metal savings of around 15%, depending on the can diameter, and a significant cutback of end sealing compound. Tests have shown no significant differences in the performance between cans closed by the conventional double-seam and those closed by the Microseam. More recently, Alcan has also developed smaller seam designs, using specifications similar to those developed for Microseam, but for beverage aluminum can ends: Mini and Reduced seams.

The Econoseam is another novel seam design developed by Krupp (Figure 8). This is a method of seaming in which the folds are made in the horizontal rather than in the vertical plane. The material consumption for can ends to be fixed through Econoseam is around 28% lower for 73-mm-diameter cans, as compared with conventional seams. Besides metal savings, the elimination of the base rim may be an advantage in preventing transit damage resulting from rim-riding impact. However, this design did not achieve widespread acceptance in commercial terms, because of competition with two-piece cans.

Shaped Cans

Recently, new styles of cans appeared in the market. Odd-shaped cans were developed because of marketing factors, as their novelty leads to product differentiation. However, other issues such as material saving and performance improvements are also behind innovations regarding shaping cans.

As mentioned previously, beading provides extra strength to the can wall. Although circumferential beads are commonly used, other beading designs have been developed. The honeycomb beading (Figure 9) is claimed to offer at least 15% material savings, because it supports thinner gauge steel at all levels of performance. These cans are marketed as Hexacan and are commercially available in diameters of 73, 83, 99, 127 and 153 mm.

Another enhanced can design is the fluted can. The food steel three-piece can developed by CMB Foodcan in the early 1990s and designated as the Quantum can is an example of a fluted can. It has vertical body flutes that act as a support to provide a strong, high-quality, thin-walled can, with enough elasticity to withstand the expansion and contraction stress experienced during the product thermal treatment (Figure 10a). More recently, Anheuser-Busch developed a similar design for aluminum beer cans. This development was marketing-driven, since the fluted can assists in product differentiation (Figure 10b).

Shaped aluminum draw-wall-ironed cans often require thicker starting metal gauges, as in the case of the 500-ml beer-glass-shaped can of Heineken beer. The can is manufactured by CMB UK, using the high-pressure gas expansion technique. The can is produced in a conventional manufacturing line and then expanded in the radial direction at the top and bottom to form a shape common to glasses used for drinking beer. The shape, together with the printed decoration showing a typical gold yellow beer color with white foam on the top, yields an attractive and different image (Figure 11).

Figure 10 Two examples of fluted cans. From The CanMaker November 1992 and January 1998, Copyright 1998 Sayers Publishing Group Ltd. Reprinted with permission.

Figure 11 Beer-glass-shaped can. From The CanMaker June 1997, Copyright 1997 Sayers Publishing Group Ltd. Reprinted with permission.

Another marketing-driven innovation on can design is the Coca-Cola contour can, a bottle-shaped two-piece aluminum can (shown in Figure 1). The can's external dimensions are similar to the industry standards, thus requiring no significant changes in filling lines. The can is higher to compensate for the loss in volume resulting from the ribbed body. The can requires 15% more aluminum to impart the strength required during the filling and seaming operations.

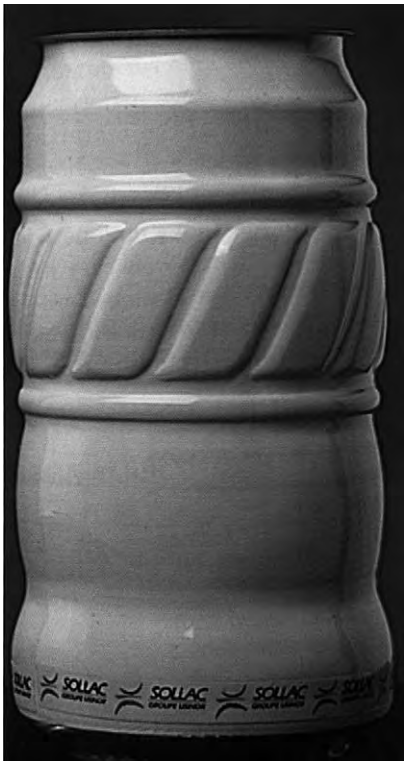

Figure 12 New shaped steel can-making process from Sollac. From The CanMaker September 1998, Copyright 1998 Sayers Publishing Group Ltd. Reprinted with permission.

Sollac has recently launched a new shaped steel can-making process – the full body necking process. Unlike other can-shaping systems, this process uses standard gauge steel and centers on a necking process that first reduces the diameter of a standard can, which is then expanded into the desired form (Figure 12). The finished can shape is formed by high-pressure water jets forcing the reduced can diameter out into a shaped mold. The major advantages of this process are claimed to be the possibility of fitting into the existing draw-wall-ironing production lines whilst maintaining identical productivity, no increase in the quantity of steel, and costs similar to those of straight-walled aluminum cans.

Can performance may also be optimized by using a design that allows for a greater amount of product to be packed. The Caronde can from Carnaud Metalbox (Figure 13) is one such example. This can is manufactured from a cylindrical can body wall, which is shaped into a squared body, affording around 18% more volume of product.

The stackability of three-piece cans is a feature introduced only recently and is becoming increasingly important, because of the risk of products tipping over in retailers' displays. The design of the base of most two-piece cans and the typical neck-in in

Figure 13 Caronde can from CMB Packaging (France). Reprinted with permission.

Figure 14 Stackable three-piece can from Ferembal (France). Reprinted with permission.

two-piece drink cans allows these cans to be stacked. This does not apply to regular three-piece cans, in which the dimensions of the top and bottom of the can are the same, and so the bottom of the upper can does not fit in the top of the underlying can. The formation of a neck-in in just one end of the can reduces the diameter and allows that end to fit in the opposite end of the next can, thus forming a stable pile (Figure 14). Furthermore, the use of one end smaller in diameter contributes to a reduced can weight and hence reduced costs. Stackability and weight reduction are the major drivers for the increased use of neck-in in three-piece cans.

Polymer-coated Metal Cans

Changes in can coatings have been influenced by environmental legislation, food-contact legislation, and trends in can making. Laminates and extruded coatings are currently under development, some of them having already been introduced on to the market. These new types of can coatings imply new can-manufacturing technologies and are related to can design and decoration.

In laminates, the metal sheet is coated by lamination with prefabricated thermoplastic polymeric films. The metal substrate, whether tinplate, tin-free steel, or aluminum, is coated on both surfaces with one layer of polyethylene terephthalate (PET) or polypropylene (PP). The advantages of laminates over traditional wet coatings include the good properties for food contact and the negligible emission of volatile materials. Also, laminates confer an excellent appearance and abrasion resistance to the external layer, combined with very good barrier properties between the food product and the metal of the internal layer. Cost is the major drawback, and some laminates require the use of a tie adhesion layer (epoxy, phenoxy, or phenolic) between the PET and the metal substrate, which may raise problems regarding the suitability of the can for food contact.

Ferrolite is a laminate produced by Corus Packaging Plus under license from Carnauld Metalbox. The metal substrate (tinplate or tin-free steel) is heated before application of PET to the external surface of the steel and PP to the internal surface. PET film imparts a high resistance to abrasion and a glossy appearance, whereas PP provides a protective barrier between contents and metal. The system allows for different types of film, colors, and thickness to be applied to each opposite surface. The resultant composite sheet passes through a second induction heater, which melts the polymeric films. After holding for a few seconds at high temperature (260 °C), the

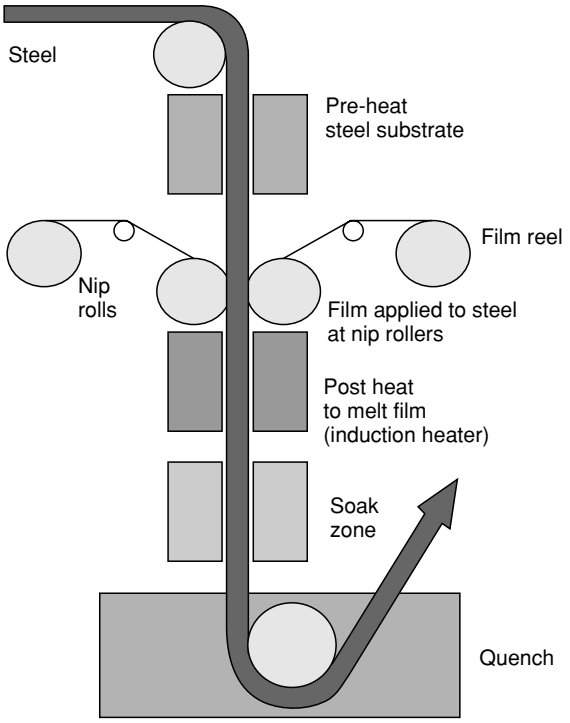

Figure 15 Ferrolite-producing process.

composite sheet is cooled down by quenching in controlled conditions in a water bath (Figure 15).

The other major laminate, Andrafol, is produced by Toyo Kohan (Japan). This is a tin-free steel laminated with PET and/or PP. The polymeric film may also be combined with lacquer on the opposite metal surface. The production system has three alternative methods for application of the polymeric film. The first method is a solvent-free process, where the PET film is thermally laminated (as described earlier for Ferrolite). The second method uses a PET film to which a tie layer was previously applied, promoting adhesion between the PET and the metal surface. Alternatively, the adhesive may be applied to the metal surface prior to laminating the polymeric film. When lacquer is used in one surface, this is applied by a roller system. The lacquer must be dried and cured in an oven.

Polymer-coated steel is used for steel aerosol components, such as tops, bottoms, and valve cups, but also for food and beverage applications. In this segment, it is used for easy-open ends and for containers such as food trays and drink cans. Food trays in polymer-coated steel are microwaveable, strong, and remain rigid when hot. Fabrication of two-piece drink cans from laminated steel requires specific technologies, but developments in this area have led to a considerable growth on the interest for this new type of cans.

Figure 16 RBS cans and TULC can. From The CanMaker November 1998 and June 1997, Copyright 1998 Sayers Publishing Group Ltd. Reprinted with permission.

The TULC can (Toyo Ultimate Can) and the RBS can (Redicon British Steel) are examples of this type of cans. The TULC can is a development of Toyo Seikan Kaisha, and it has been on the Japanese market since 1992, currently representing around 14% of the Japanese can manufacturing market. The RBS can was launched by Corus Packaging Plus and Redicon Corporation in 1997 (Figure 16). These are two-piece cans made from PET laminated tin-free steel, in the TULC case, and from Ferrolite, in the RBS case. The cans are produced by a draw–stretch–redraw technique, an improvement on the draw–redraw process, in which body-stretching operation is included. The RBS production line includes a mechanical body-shaping machine. Recent developments in the TULC can manufacturing process have led to the use of wall ironing, which implies considerable can lightweighting. Another recent innovation in the TULC can is the new wall design in a triangulated pattern (the Diacut pattern), which using slightly more material thus providing a much better resistance to denting and allowing for a global weight reduction. The polymer-coated steel cans weigh more than a regular draw-wall-ironing steel can, but further weight reductions are expected by producers of these cans.

Polymer laminated steel is also used for three-piece cans, as the Crystal can from the Japanese producer Hokkai. In this case, tin-free steel is coated on both surfaces with PET. The lamination requires an adhesive layer and therefore a curing operation. The sheets are cut into can body blanks and welded. The external layer of PET is preprinted with a high-definition decoration, providing a very appealing image.

An alternative to laminated metal, using a polymer coating, is extrusion coating. This technology is commonly used in other areas of the packaging industry, such as polyethylene extrusion coating of paper and paperboard. In extrusion coating, the polymer is

extruded and applied, in the melted state as it leaves the extruder die, on to the metal substrate. Advocates of this technology claim that, compared with the laminates, extrusion coating has the following advantages: thinner coatings, a wider range of materials available, better adhesion and flexibility, and lower costs. This technology is, however, still under development and, besides some unsolved issues such as can decoration, still lacks consolidation and proven results at the industrial level.

Alcoa has been developing its Flex-Coat extrusion coating system, devoted to aluminum coating for beverage draw-wall-ironed cans. Hoogovens developed the Protact, a multilayer coated steel intended for aerosol components, full aperture easy-open ends, and twist-off caps, although applications in shallow cans and trays are also expected. The composite has three layers of polymer on each side of chromium-coated steel applied by coextrusion. The top layer provides wear resistance, gloss, and ink adhesion, the center layer gives strength and coloring, and the third layer provides adhesion to the metal substrate.

Can Opening

One of the most remarkable features introduced in metal cans is the easy-open end. Easy opening is considered by consumers to be a major advantage and marketing argument for the metal can, and its rate of use is increasing. Easy-open ends are of two broad types: those that provide a pouring aperture for dispensing liquids, used in beverage cans, and those that give a full aperture to remove solid products, used in food cans.

The ring-pull easy-open end for drink cans was first introduced in 1962. The tab was detachable, causing litter problems. To overcome this, stay-on tabs were incorporated later on. Easy-open ends for food cans, full aperture ends, were introduced in 1965. The first ends were made of aluminum (since this is an easily workable material), and only in the middle of the 1980s were steel easy-open ends being widely used.

The importance of this feature in selling cans has led producers to invest great effort and economic resources in optimizing the concept. Easy-open ends must accommodate a compromise between easy opening and resistance during the product's thermal processing and can handling during the distribution operations. Balancing these opposite requirements is a technical challenge, and the design of an easy-open end includes several aspects that should be precisely detailed:

- the levering tab (produced separately and attached to the end) should not be too hard to lift for breaking the can end (initial pop); the tab ring should also allow easy access to the consumer's finger at the peeling back stage;
- the rivet formed in the end panel through which the tab is attached without loss of the end integrity; if the rivet is not strong enough, when peeling the end, the consumer might remove the ring without opening the can;
- the panel profile that is responsible for the end behavior both during the thermal processing (providing the required reversible deformation) and when peeling off the end; and
- the score line that allows for the end to be torn off; the depth of the score line has very tight tolerances, as it controls the force needed to tear the end off but also constitutes a weak region of the can.

Current developments and improvements focus on making ends safer and easier to open (that is avoiding sharp edges that might hurt the consumer and requiring less effort to open the can) while maintaining the required mechanical strength and reducing cost by reducing thickness and diameter.

The Eole II from Carnauld Metalbox Food, the Top Can from Impress and the Meteor from Usinor Packaging are some examples of recent developments in steel easy-open designs for food cans. The first two are manufactured from single-reduced steel, and the last is produced with double-reduced steel.

Eole II is an update of the Eole first version, and the producer claims that improvements in the panel design and in the score line profile have led to an end that is 30–40% easier to open, with equal or better performance in mechanical and seaming load tests. The initial pop force required is lower, and the size of the finger well is larger for easier access. The Eole family of ends is supplied with a printed diagram on the panel, to inform the consumer of the best way to open the can.

The Top Can follows a different philosophy: instead of reducing the initial pop force, the end is designed to reduce the amount of energy globally expended in the opening process (pop plus tear). The panel profile insures that the end does not bend during the peeling part of the opening process, which is claimed to reduce the tearing force. The total energy expended is reduced, even with an increased pop force.

The Meteor design uses double-reduced steel combined with a new method of producing the rivet. Producers claim that this allows for an increased ease of opening, and for the possibility of reducing the thickness of the end without compromising strength.

In the beverage area, Crown Cork & Seal introduced very recently the SuperEnd. This new design is

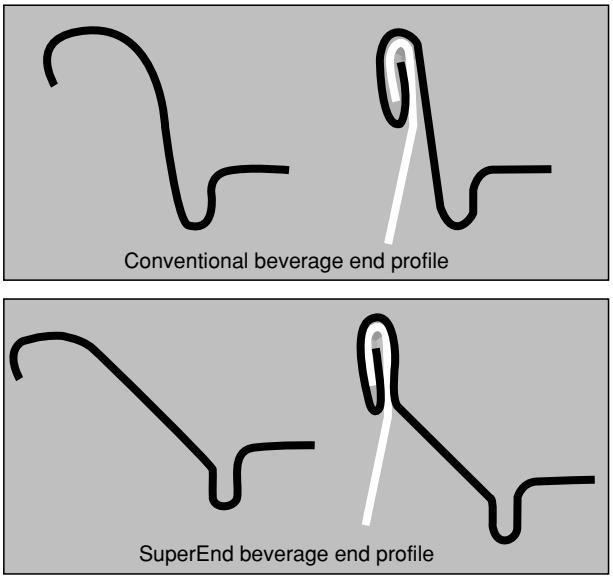

Figure 17 Recent developments in easy-opening. From The CanMaker November 1998, Copyright 1998 Sayers Publishing Group Ltd. Reprinted with permission.

claimed to provide a better seam integrity, to be stronger, to be easier to open and to drink from, and to have better pouring characteristics. Figure 17 shows the profiles of the current industry standard and that of the new end design, before and after seaming.

Other developments in cans opening and closing systems are resealable drink cans (screw-top cans) that allow closure between uses, and the peelable end, where a foil end is heat-sealed to a metal ring, which is then double-seamed on to the can body.

See also: **Canning**: Cans and their Manufacture

Further Reading

Baker M and Eckroth D (eds) (1986) *The Wiley Encyclopedia of Packaging Technology*. New York: Wiley.

Goddard R (1992) *Metals in Packaging*. Leatherhead, UK: Pira International.

Page B (1998) Design for the future. *The CanMaker* November: 47–48.

Robertson GL (1993) *Food Packaging Principles and Practice*. New York: Marcel Dekker.

Soroka W (1995) *Fundamentals of Packaging Technology*. Herndon, VA: Institute of Packaging Professionals.

Food Handling

P Rutledge, Food Processing Consultant, NSW, Australia

Introduction

There is a large variety of affordable, canned foods available on supermarket shelves. These foods comprise the full range of the food groups which include meat, fruit, vegetables, cereals, and dairy products. Preparation methods and canning technology vary depending upon the product. Vegetables and fruits are canned directly after harvest and require washing, blanching, and sometimes peeling before canning. Meat, once it is butchered, is cut into chunks or minced, while cereals are often used as the ground flour.

Farm produce is generally seasonal and is only available for a matter of weeks each year. Canneries require enough produce during this short time to maintain stock in the supermarket from season to season. This means the scale of operation is in tonnes per hour for sometimes 24 h day^{-1} to process enough product while it is available. Mechanical handling and processing systems are essential under these conditions.

Preparation of Vegetables

Vegetables are normally mechanically harvested. Mechanical harvesters do not distinguish between the wanted vegetable material and other extraneous plant material and soil. Some of this is removed by fans and belts on the harvester before the crop is transported to the cannery. Cannery cleaning operations must be very thorough to removed the unwanted material and wash the vegetable.

Cleaning

The preparation of vegetables harvested from under the ground, such as carrots and potatoes, requires the removal of adhering soil and stones. Several systems are available to remove the soil, and they start with dry brushing or soaking to remove the bulk of the dirt, followed by washing. Washing involves wet scrubbing with rotary brushes or rubber fingers, followed by rinsing in a rod washer. In the rod washer, the vegetables are tumbled in a cylinder made from steel rods while being washed by water sprays from inside the cylinder. These vegetables are usually peeled in a subsequent operation, which insures a clean product.

Legumes, such as green peas and beans, are mechanically harvested and transported in bulk ready for

canning to the processing plant. Specialized equipment is used, for example, to carry out operations such as husking corn and breaking up clusters of green beans. Dry cleaning with air blowers to remove extraneous leaf material from this type of vegetable is followed by washing in tanks which removes mud balls and stones. Some systems use froth flotation cleaners to remove small pieces of vegetable material. A final rinse with water follows the cleaning treatment to remove the last of the soil.

Leaf vegetables, such as spinach, are difficult to clean as extraneous material lodges between the leaves. Leaves are cleaned by floating them in tanks of water where the water is agitated with air or water injection. This separates the leaves and removes the soil.

Inspection of the washed vegetable by instruments, such as color sorters or electronic sorters, at this stage will remove any remaining unwanted material. Manual inspection is still carried out where the removal of specific material cannot be achieved mechanically.

Peeling

Vegetable peeling may be by mechanical cutting or abrasion, by the use of high-pressure steam, or by chemical treatment. Abrasion peeling uses carborundum-covered rollers or disks which come in contact with the skin of the tumbled vegetable. The abraded skin is removed with water sprays. To remove material from depressions in the vegetable, some of the flesh of the vegetable is also removed with abrasion and, as such, gives the lowest yield of the three methods.

Vegetables are steam-peeled by holding them for a short time in high-pressure steam. This superheats a layer of tissue under the skin. When the pressure is suddenly released, the tissue explosively boils, releasing the skin from intimate contact with the vegetable. Typical conditions for steam peeling are 17 atm steam pressure for about 30 s. The skin is washed from the vegetable. This system is the most efficient in removing the skin and in yield of peeled vegetable.

A hot lye (sodium hydroxide) solution is used for chemically peeling vegetables and some fruits. It is important that the vegetable is thoroughly washed after peeling to remove all traces of the lye. The treatment varies depending upon the skin to be removed, but a boiling solution of 10% lye will remove most skins in less than 1 min. The skin is removed by rotating brushes or rubber fingers and water sprays. Caustic waste, both liquid and solid, has to be neutralized with acid before disposal, which increases the cost of peeling.

Slicing and Dicing

Cutting operations are used to supply the required size of vegetable for canning. Asparagus spears are sawn to the correct length for the can, and carrot and potato are sliced or diced to give an attractive piece of canned vegetable. Canned food in which the particles are of uniform size generally has a superior appearance to that with large variations in particle size. Some dicers are designed to give an irregular-shaped dice so the product looks home-made. The size and shape of the canned vegetable depend upon the style of pack that the market requires.

Preparation of Fruit

Washing

Generally fruit is more easily damaged than vegetables. Therefore, fruit is washed by submersion in tanks of water, which is agitated, followed by water sprays on the elevators as the fruit is removed from the tank. Rod washers may be used with fruit, such as citrus, which are not easily damaged.

Peeling and Pitting

Pome and some stone fruit require peeling and pitting before canning. Pitting is a mechanical operation, and each type of fruit has specialized equipment designed for pit removal. Mechanical apple and pear peelers remove the core section and skin before halving the fruit. Stone fruit that is peeled for canning is usually chemically peeled with lye followed by washing. Other fruit with stems, such as cherries, are rolled over sets of small rotating rollers which pick up the stems and remove them.

Blanching

Blanching is a heat treatment in a near-boiling water or steam followed by rapid cooling given to vegetables and some fruit. Blanching removes gases from within the tissue and softens the product. Blanching makes the product easier to fill into the can and to obtain the correct fill weight. The removal of the gas also reduces the oxidation of the product, maintains the vacuum in the can, and prevents excessive can corrosion.

Blanching gives the product another washing treatment and inactivates enzymes which may cause deterioration of the food. Enzyme inactivation is not as important for canned foods as it is for frozen foods, as canned foods receive a far greater heat treatment during thermal processing of the can. It can be important if there is a long delay between filling the can and retorting. Typical blanch times in near-boiling

water are 60–90 s for small objects, such as green peas and diced carrot, and up to 3 min for larger pieces.

Preparation of Juices

Juices are the liquid which may be squeezed from fruits and vegetables. The major methods of extracting juices are to apply a force to the whole or pulped material followed by screening out pulp from the resultant liquid. This can be carried out continuously in screw presses and belt presses, and there are many different types of batch presses. Citrus fruit is reamed on mechanical reamers or crushed in such a way as to remove the edible portion from the skin. Unwanted material is removed from the juice in paddle or brush finishers, or in some cases small screw presses. These machines push the juice through a screen while separating and removing the pulp which is too large to pass through the screen.

Citrus juice is pasteurized, which is a heat treatment of 95 °C, immediately after extraction to inactivate pectinase, which will cause the cloud in the juice to 'fall.' The cloud in the juice is held by naturally occurring pectin which, when attacked by pectinase, allows the juice to separate into a clear serum and a solid deposit. Conversely, pectinase may be added to other juices, such as apple juice, to manufacture a clear juice. If the juice is not hot-filled it will receive another heat treatment during the canning process.

Meat Preparation

Meat preparation, after slaughter and deboning, mainly consists of removing unwanted tissue such as fat, skin, and visible arteries. Meat shrinks by about 30% when it is cooked, so for formulated products using meat it is usually precooked before filling into the cans. Some meat products are corned, which means they are cooked with a cure containing salt and nitrite. Nitrite causes the meat to turn a characteristic pink color during heating and, because of its antimicrobial action, permits a less severe heat treatment during retorting. (*See* **Curing**; **Meat**: Slaughter.)

Fish such as tuna are cleaned and then steamed to allow for the easy removal of skin and bones. The steamed fillets are filled into a machine which shapes and cuts them to the can size before filling into the can. Other fish are cut to size and filled raw. Fish are canned with brine or oil or, in some cases, a formulated sauce. (*See* **Fish**: Processing.)

Preparation of Formulated Products

There is an infinite variety of formulated products from meat stews to dairy desserts, and beverages such as beer, that are canned. Most of these products are cooked or blended or brewed prior to filling. Those that are cooked in the preparation stage are filled into the can hot.

Carbonated products, such as beer and sparkling fruit juices, are filled at temperatures just above freezing to maintain the carbonation. Soft drinks, although they are often packed in metal cans, are not 'canned products' as preservatives are used to maintain their microbiological stability. (*See* **Preservation of Food**.)

Cans

Cans are delivered to the factory on pallets. The pallets are automatically unloaded into the can race system. This delivers the correct-size can to the can filler. Canning lines operate between 200 and 2000 cans per minute, so this system must work effectively. Cans may be contaminated so they are thoroughly washed prior to filling.

Filling

There are many types of machines used to fill cans, and the type employed depends, of course, on the product to be filled. Volumetric piston fillers can be used with liquid products and liquid products with entrained solids, such as mushrooms in butter sauce. A turntable is used containing several filling 'heads' so that several cans are filled sequentially as the turntable rotates; filler speed depends upon the number of heads. A slow filler might have 12 heads, while high-speed fillers have up to 72 heads.

Tumble fillers are used to fill solid materials, such as vegetable pieces, into cans. The washed cans move through a large rotating drum containing the pieces of product. The product falls into the can and excess is removed by tilting and shaking the can at the exit of the filler. Other volumetric fillers wipe the solid products into pockets on a turntable and the products are then dumped by gravity into the can.

Hand filling is used for products that are difficult to fill by machine. The filling of asparagus into cans can be carried out by machine but in some cases the tips are damaged, so some processors prefer hand filling.

Some products are canned with syrup or brine. This is a separate operation to filling the solid food. It may take place before or after the filling of the solids. Some solid products may have pockets of air held between the pieces and, in these cases, the liquid is added prior to filling the solids so the liquid will fill these spaces. Sometimes the cans are topped up after the solids are filled.

A head space must be left in the top of the can after filling. This small space is evacuated on closing, but is important to the integrity of the can. An overfilled can prevents the expansion of the product during the thermal process, which may result in permanent damage to the can end. Also the thermal characteristics of the can change, which may invalidate the calculated thermal process.

Exhausting

Oxygen remaining in the head space of cans accelerates corrosion of the tin plate in the head space area. To prevent this occurring, the volume of gas between the product and the lid of the can (known as the head space) must contain a partial vacuum.

The conventional system of exhausting the cans is to clinch the lid on the can. Clinching is a partial, first operation seaming roll which holds the lid loosely on the can. Exhausting is carried out by passing the filled cans with clinched lids through a steam-filled compartment for several minutes to heat the can contents and displace the air in the can with steam. This is immediately followed by the completion of the seaming operation.

Another method of exhausting is by filling the product hot into the can followed by hot brining or syruping which has the same effect as steam exhausting. This is followed by 'steam flow closing.' Steam is injected between the can and the lid as they come together in the can closer. This displaces the air in the head-space region. When the steam condenses, a vacuum is formed in the head space of the can.

Alternatively, mechanical vacuum pumps can be used to reduce the air in the head space of the can. This can be done during filling, as fruit may be vacuum-syruped by pulling a vacuum on the can and air is replaced by the incoming syrup. Some closers have a vacuum chamber in which the can is seamed. This is common in meat and fish canning, where the canning lines are slowed by the relatively slow operation of the vacuum closer.

Acid products, such as fruit juices, jams, pickles, and chutneys, may be filled into the can at near-boiling temperatures. The cans are seamed, inverted to sterilize the lid, and then cooled. This is called the hot-fill process. The vacuum results from the shrinkage of the product on cooling and the displacement of air by the steam from the hot product.

Can Closing or Seaming

The modern can closer also has a turntable similar to the filler where the cans are fed into a closing 'station.' The lid is aligned over the can and the baseplate of the station raises the can and lid to engage the top chuck. The seaming rollers then roll around the seam to form the seal. The rollers retract, the baseplate is lowered, and the closed can exits from the closer. Closers with four to six stations are common.

The seaming operation is carried out by two rolling operations. The first operation roller bends the two flanges together, and the second operation roller flattens them to form a seal. The seal is insured by a thin layer of mastic-like material deposited in the flange of the can end called 'compound.'

Rates of can seaming depend more on the speed of the filler than that of the can closer. Beverages which are easy to fill can be closed at speeds of 2000 cans per minute. New fillers for semisolid foods, such as pet food, allow closer speeds of about 1300 cans per minute and vegetables are often closed at 500 cans per minute.

Thermal Processing

There are two important factors in canning which make the food safe for long-term storage. Food is sealed to prevent recontamination, and it is heated to inactivate microbes which spoil the food. The seaming operation on the can hermetically seals the food, and the thermal process provides the sterilization step.

Products which have a pH of less than 4.5 are called 'acid foods' and can be thermally processed at temperatures less than 100 °C, which is called a pasteurization process. Those with a pH above 4.5 are called 'low-acid foods' and must be thermally processed at temperatures between 110 and 125 °C in a retort or pressure vessel.

Postprocessing Operation

Cans exiting from water-cooling operations are wet with chlorinated water and must be dried before they can be handled safely. Some canners label the cans directly after processing and other palletize the cans for storage before labeling. Lithographed cans do not require further labeling.

Can stores should be maintained at a temperature designed to prevent water vapor condensation which could rust the outside of the cans. The temperature must, however, not be too high as this can, on very rare occasions, promote the growth of heat-resistant thermophilic bacterial spores which may have survived the thermal process.

Cans are held by the canner in store until an incubation period has passed. This provides added protection for the customers as it insures that only safe wholesome food is placed in the marketplace. All

cans are packed in cardboard outers to prevent damage to the cans during transport and handling, and also as a convenient package for the supermarket.

Storage of Canned Food in the Home

Canned food should be stored in a dry cupboard. Most canned food is safe for at least 2 years, but care should be taken to use the oldest cans first. Stock rotation is important at home as well as in the supermarket.

See also: **Canning**: Principles; **Curing**; **Fish**: Processing; **Meat**: Slaughter; **Preservation of Food**

Further Reading

Arthey D and Ashurst PR (eds) (2000) *Fruit Processing*, 2nd edn. London: Blackie Academic and Professional.

Arthey D and Dennis C (1991) *Vegetable Processing*. Glasgow: Blackie.

Downing DL (1996) *A Complete Course in Canning and Related Processes, Book III, Processing Procedures for Canned Food Products*, 13th edn. Timonium, Maryland, USA: CTI Publications Inc.

Hersom AC and Hulland ED (1980) *Canned Foods – Thermal Processing and Microbiology*, 7th edn. Edinburgh: Churchill Livingstone.

Lopez A (1987) *A Complete Course in Canning*, 12th edn. Baltimore: Canning Trade.

Quality Changes During Canning

R J Pither, Campden & Chorleywood Food Research Association, Gloucestershire, UK

Introduction

Canned foods are a significant component of the diet of most individuals in developed countries, offering food in a convenient form with year-round availability. The canning process relies on heat treatment for the destruction of microorganisms and preservation of the food, which is then generally considered to have an indefinite microbiological shelf-life providing that pack integrity is maintained. The extent of the thermal processing, in terms of both temperature and duration of the treatment, is dependent upon the chemical and physical composition of the product. Both physical and chemical changes occur during processing and, to a lesser extent, during storage, and it is these that determine the product quality in terms of its sensory properties and nutrient content. These changes, which can be either desirable or undesirable, are influenced by the time and temperature of the process, the composition and properties of the food, the canning medium, and the conditions of storage.

This article will consider the changes that can occur during canning and their effect on the quality of the product.

Changes in the Sensory Properties of Foods

The sensory properties of a food, its flavor, color, and texture can all be affected by thermal processing. Changes in these properties may be in the form of direct effects of heat on food constituents (e.g., starch gelatinization, protein denaturation, and cell separation; see **Table 1**) or through heat-induced reactions such as the Maillard reaction. Significant changes in all three sensory properties can also be brought about by oxidation reactions that can occur, not only during processing, but also on subsequent storage of the canned product.

Oxidation reactions in fruits and vegetables occur mainly during preparation prior to processing, due to the effects of oxidative enzymes. Heat treatments, such as blanching prior to canning and the sterilization process itself, cause inactivation of these enzymes. Along with the low oxygen tensions in canned products, this limits oxidation in all but the most oxygen-sensitive constituents.

Table 1 The effect of heat processing on sensory quality

Chemical/physical reactions or changes occurring	*Impact on sensory attribute*
Flavor	
Volatile loss (scalping oxidation)	Loss of flavor
Volatile formation	
Maillard	Roasted flavor, bitterness
Oxidation	Rancidity
Pyracaines	Roasted flavor
Texture	
Cell membrane damage	Loss of crispness
Cell separation	Loss of firmness
Protein denaturation	Gelling, firming
Starch gelatinization	Gelling
Color	
Natural pigment breakdown	Bleaching Loss of color
Maillard reactions	Browning
Others, e.g., metals and polyphenolic compounds	Discoloration

Flavor Changes

The flavor of food can be retained, modified, or, occasionally, significantly changed during heat processing, with most of the changes seen occurring in the volatile flavor components.

Lipid Oxidation

Lipids occur in almost all foodstuffs, particularly meat and fish products, and lipid oxidation can therefore occur during canning of most foods. Saturated fatty acids are relatively stable at the temperatures used in standard canning operations. However, unsaturated fats are degraded, under the conditions of oxygen and heat, to a large number of volatile compounds, which give rise to both desirable and undesirable flavors. (*See* **Fatty Acids: Properties**; **Oxidation of Food Components**.)

The first stage of the oxidation reaction involves uptake of oxygen in the presence of catalysts, such as transition metals and hemoproteins, and is initiated by heat or light. Highly reactive hydroperoxides are formed, which undergo secondary reactions giving rise to a complex mixture of low molecular compounds, including aldehydes, ketones, alcohols, acids, alkanes, alkenes, and alkynes. A certain level of volatile compounds is generally considered necessary to give characteristic odor and flavor properties to many foods, but as many of the volatile compounds give rise to typical rancid or stale off-flavors, an ideal balance needs to be achieved in the food.

Maillard Reaction

In foodstuffs, reactions that occur between reducing sugars and amino acids or proteins are known as Maillard reactions. A very important aspect of the Maillard reaction in food preparation is the production of flavors and aromas. The rate of Maillard reaction increases with temperature, although pH and water content also influence the reaction. The Maillard reaction can be split into three main stages. The first is a condensation reaction between the carbonyl group of a reducing carbohydrate and the free amino group of an amino acid or protein, followed by rearrangement of the glycosylamines. These early Maillard reactions can lead to losses in protein quality but do not give rise to flavors in the food.

The second stage, the advanced Maillard reactions, involve many complex reactions and pathways. The rearranged glycosylamines can degrade via a multiplicity of routes, including dehydration, elimination, cyclization, fission, and fragmentation, which results in a pool of flavor intermediates and flavor compounds being formed. The flavors produced by the Maillard reaction can be classified into four main groups, nitrogen heterocyclics and cyclic enolenes, which give characteristic flavors to heated foods, and monocarbonyls and polycarbonyls, which include more volatile supplementary flavors not necessarily associated with product characteristics.

The final stage of the Maillard reaction is the polymerization of the highly reactive compounds formed in the second stage, which results in the formation of brown melanoidin pigments.

Most Maillard reaction products contribute highly desirable flavors to heated foods such as bread, toast, cereal products, and meat. These flavors are often described as baked, nutty, roasted, caramel, and burnt, but even these can be considered to be off-flavors in certain other foods, e.g., the burnt caramel taste in heat processed milk.

Taints

Other types of off-flavor in canned foods can be caused by contamination of the product leading to an undesirable taint. One particularly unpleasant taint that has been found in a range of foods is the 'catty taint.' This is caused by a heat-dependent reaction between natural sulfur-containing compounds of the food and unsaturated ketones, such as mesityl oxide, which are widespread in many solvent-based products.

Examples of causes of catty taint in processed meat products include: meat that has been stored in a cold store painted with a material containing mesityl oxide as a solvent contaminant, ox tongues that have been hung on hooks coated with a protective oil, and pork packed in cans where the side seam lacquer has been dissolved in an impure solvent. Catty taint has also been a problem in canned rice pudding, where the dye used for printing on the rice sacks contained traces of mesityl oxide, which was picked up by the rice and reacted with the trace amounts of hydrogen sulfide in the milk during processing.

Texture

Canning can bring about both desirable and undesirable changes in texture of foods, primarily through starch gelatinization, protein denaturation, and pectin changes.

Starch Gelatinization

Starch gelatinization commences at a range of temperatures dependent on the type of starch, i.e., proportions of amylose and amylopectin present, and the availability of water. The two components of starch

behave differently on canning, with amylose giving an opaque solution that sets to a firm gel on cooling and with amylopectin forming a translucent paste that remains fluid when cooled. The swelling of the starch granules during canning or other heat processes causes cellular disruption, which, together with the starch gelatinization, gives a softening in texture and increased palatability to the product. During canning of vegetables, starch can often be leached into the brine making it viscous or turbid, e.g., in mature canned peas. (*See* **Starch**: Structure, Properties, and Determination.)

Pectin Changes

Canning of plant materials can lead to loss of semipermeability of cell membranes, and solubilization and breakdown of pectic substances in the cell walls and middle lamellae. The resultant cell separation leads to a loss of crispness and a softening of the product. This is generally a desirable effect, improving the palatability of foods, but overprocessing can lead to excessive softening of fruits and vegetables. Low-temperature blanching ($\sim 70\,^{\circ}C$) of fruits and vegetables can cause firming, due to enzymic demethylation of pectins followed by calcium bridges forming across the liberated carboxylate groups. (*See* **Pectin**: Properties and Determination.)

Protein Denaturation

Application of heat in meat canning causes protein denaturation, which leads to changes in texture. The hydrogen bonds maintaining the secondary and higher structure of the protein rupture and form into a predominantly random coil configuration, which affects the solubility, elasticity, and flexibility of the proteins. The sarcoplasmic and myofibrillar proteins in meat coagulate during heat processing, resulting in a firming of the texture, while the collagen proteins become more soluble, softening as they take up water. The resultant texture of canned meat is similar to that of conventionally cooked products.

Struvite

Canned seafood products such as tuna, salmon, and shrimps can occasionally contain small crystalline fragments of magnesium ammonium phosphate, commonly known as struvite. Struvite is formed during the canning process by the union of mineral elements that occur naturally in the flesh of seafood.

The formation of struvite is a rare event that appears to occur sporadically. Although struvite has no odor or taste and is harmless, it does resemble broken glass and, therefore, is undesirable from the consumer's point of view.

Color

Color changes during canning can be brought about by the breakdown of natural pigments, by the production of colors due to oxidation reactions, by the Maillard reaction, and by interactions between product constituents, e.g., metals and polyphenolic compounds.

Chlorophyl

Chlorophyl pigments are present in all green vegetables, but canning leads to breakdown with the associated color change from bright green to olive green or brown. The chlorophyl is converted to pheophytin by the loss of magnesium ions (Mg^{2+}), with heat and low pH during processing greatly accelerating this change. The addition of alkaline salts in the canning liquor to maintain a pH of 6.2–7.0 and high-temperature, short-time (HTST) processing have both been used in canning to reduce chlorophyl degradation. (*See* **Chlorophyl**.)

Heme Pigments

The red coloration in meat is due to two heme pigments, hemoglobin in the blood and myoglobin in the muscle tissue; as most of the blood is removed after slaughter, the main pigment is myoglobin. Canning causes oxidation of myoglobin to metmyoglobin, which gives the typical brown color of cooked meats. It is this reaction that is also the main cause of color change on canning of dark fleshed fish, e.g., tuna and mackerel. Green discolorations can also occur as a result of a reaction of myoglobin with sulfur compounds.

Carotenoids

Carotenoids are fat-soluble, highly unsaturated red, orange, or yellow pigments that are susceptible to oxidation and isomerization under the conditions of heat and low pH, such as those used during the canning process.

Carotenoids are generally found complexed with either proteins or fatty acids, which protect them from oxidation. It is the breakdown of these complexes during processing that leads to degradation of the carotenoids resulting in bleaching or discoloration of the product.

In crustaceans, the denaturation of carotenoprotein on heating releases the carotenoid astaxanthin, which causes a change in color from natural blue–grey to pinky red. Two types of isomerization, *cis–trans* and epoxide, can also occur, giving rise to a slight lightening in color in canned fruits, e.g., pineapple. (*See* **Carotenoids**: Occurrence, Properties, and Determination.)

Anthocyanins

Anthocyanins are water-soluble, red–violet pigments that can take part in a wide range of reactions during canning. A combination of heat and oxygen can lead to the hydrolysis of the glycosidic bonds, resulting in loss of color and formation of yellow or brown precipitates, but a low pH gives rise to greater color stability. Aldehydes, produced by the breakdown of sugars during canning, and ascorbic acid both accelerate the breakdown of the anthocyanins. These color losses are a particular problem in canned red fruits, e.g., strawberries.

Anthocyanins can also be produced, as a result of excessive thermal processing, from leucoanthocyanidins, giving rise to defects such as red gooseberries and pink pears. (*See* **Phenolic Compounds**.)

Tin and Iron

Both desirable and undesirable color effects can occur in canned goods due to reactions involving metal ions.

Tin coating of the internal can surface produces a chemically reducing environment, which minimizes oxidation and helps prevent color and flavor degradation in certain products, e.g., tomato-based products and asparagus. Asparagus canned in a lacquered can may develop a dark discoloration due to the formation of a complex between rutin (a flavanol glycoside) and iron, whereas even partial exposure to tin, such as one plain can end, will retain product color.

Certain anthocyanin pigments can form metal complexes with tin and iron produced through internal can corrosion, causing pink discoloration, especially in pears and peaches, and bluing of red fruits.

Current UK legislation limits the level of tin in canned foods to 200 mg kg^{-1}. Tin can cause short-term gastric irritation and nausea, but it is not absorbed by the gut in significant quantities, and there is no evidence of any long-term or cumulative toxic effects.

Maillard Reaction

The Maillard reaction, the principles of which are discussed earlier in this article, can cause off-colors, particularly browning on canning of a wide variety of products.

Color compounds can be grouped into two general classes, low-molecular-weight compounds and the melanoidins, which have much higher molecular masses.

The browning of navy beans in tomato sauce is primarily due to the formation of melanoidins through the Maillard reaction. Melanoidins are also partly responsible for the browning effect seen in canned apricots. Browning during canning of dark flesh fish such as mackerel and tuna is of little consequence. However, in white fish, such discoloration is a major problem, and for this reason, white fish is not routinely canned. Canning of milk can also lead to a brown tint, although cream is less affected.

Betalains

Betalains are water-soluble and split into two main groups, red betacyanins and yellow betaxanthins. The most important pigment in this group is betanin, the red pigment in beetroot which is often used as a natural colorant. Betanin is susceptible to oxidation during canning, which leads to a loss of color, although this is often not noticeable due to the large amount present.

The occurrence of a black ring discoloration in canned beetroot is due to enzymatic oxidation of polyphenols that occur during steam peeling, and not to betalain reactions.

Changes in Nutritional Properties of Foods

Both physical and chemical changes can occur during the canning of food, which may affect its nutritional status (Table 2). As well as damage to heat-labile

Table 2 The effect of heat processing on the major nutritional components

Nutrient	*Effect*
Water	Loss of total solids into canning liquor Dilution Dehydration
Proteins	Enzymic inactivation Loss of certain essential amino acids Loss of digestibility Improved digestibility
Carbohydrates	Starch gelatinization and increased digestibility No apparent change in content of carbohydrate
Dietary fiber	Generally no loss of physiological value
Lipids	Conversion of *cis* fatty acids to *trans* fatty acids Loss of essential fatty acid activity
Water-soluble vitamins	Large losses of vitamins C and B_1 due to leaching and heat degradation Increased bioavailability of biotin and nicotinic acid due to enzyme inactivation
Fat-soluble vitamins	Mainly heat-stable Losses due to oxidation of lipids
Minerals	Losses due to leaching Possible increase in sodium and calcium levels by uptake from canning liquor

nutrients and physical loss of nutrients due to leaching, there are many reactions that occur on canning that affect the availability of nutrients within the foodstuff and therefore their usefulness to the body.

If comparing the nutritive value of canned foods with that of fresh foods, it is also important to consider any changes that occur during conventional preparation and cooking techniques.

Moisture

The movement of water and solids during canning can cause major changes in the nutritional status of the foodstuff. If the complete can contents are consumed, these changes can be largely disregarded, but in products where the canning liquor is discarded, the effects of dilution, dehydration, and loss of total solids must be taken into consideration. Dilution or dehydration will affect the relative proportions of other constituents in the food, while soluble nutrients can be leached into the liquor.

Proteins

Heating of proteins, as in canning, causes denaturation, i.e., rupturing of the hydrogen bonds and other noncovalent bonds, leading to changes in the conformation of the protein. The degree of protein denaturation depends on the level of heat treatment applied, but it can also be caused by oxidation and reaction with other food constituents, e.g., reducing sugars and lipid oxidation products. The total level of crude protein is generally unaffected by canning, but both desirable and undesirable changes can occur in its nutritive quality and availability. Mild heating of proteins leads only to changes in tertiary structure, which have little nutritional effect, although there is usually a loss in solubility. More severe heating, as in canning of vegetables, results in the Maillard reaction and the consequent loss in protein quality. These reactions occur mainly between lysine and sugars and cause a loss in availability of lysine, through cross-linking, with a loss of up to 40% being seen on canning of potatoes. Canning of meat also leads to the reduction in availability of lysine and other essential sulfur containing amino acids and can lead to a reduction in the digestibility of the meat. (*See* **Protein**: Chemistry; Functional Properties.)

The losses in protein availability that occur under normal canning conditions are, however, quite small and not nutritionally significant for most people in developed countries, as lysine is rarely the limiting amino acid in the diet. Canning can, in fact, lead to improved protein availability and digestibility by denaturing antidigestive factors and by denaturing proteins. Some examples of advantageous effects of heat processing are mentioned below.

Heating of milk results in the proteins being precipitated by stomach acids as finely dispersed particles, making attack by digestive enzymes more effective than in raw milk. This can also enhance the formation of disulfide bonds, e.g., between β-lactoglobulin and K-casein, which leads to greater stability of the normally unstable β-lactoglobulin. The canning of legumes improves their digestibility by unfolding the major seed globulins, as well as increasing nutritional availability by inactivation of trypsin inhibitors.

Lipids

The nutritive value of the fat content of foods is not generally significantly altered during normal heat processing. Hydrolysis reactions can occur, resulting in separation of the fatty acids from the glycerol unit, but this does not adversely affect the nutritional value of the fat as the resulting free fatty acids are available for digestion.

Saturated lipids are relatively stable, but unsaturated lipids are prone to oxidation when heated in the presence of oxygen or air. The exclusion of oxygen or use of antioxidants prevents the oxidation of lipids during canning, so that losses in the nutritional value of fats are unlikely to be significant. It is worth considering the effects of lipid oxidation, however, as any contact with oxygen during the history of the food can be sufficient for oxidation reactions to occur.

The major effect of lipid oxidation is related to the flavor of food, but it can also result in the conversion of *cis* fatty acids to *trans* fatty acids. The nutritional value in terms of energy is similar for the two fatty acid types, but the *trans* fatty acids do not generally possess essential fatty acid activity. The availability of the fat-soluble vitamins A, D, and E, as well as vitamin C and folate, can also be reduced during lipid oxidation.

Carbohydrates

Carbohydrates have numerous characteristic properties, and the effects of canning are therefore varied. The levels of total and available carbohydrates have been found to be largely unaffected during canning of fruit and vegetables. In general, the effects of canning on carbohydrates are related not directly to their nutritional value but to their interaction with other food constituents and to the overall eating quality of the foodstuff.

Losses on canning can be caused by reducing sugars reacting with protein through the Maillard reaction, which also causes a loss in availability of certain amino acids.

Gelatinization of the starch granules improves the texture and, thereby, palatability of the food; it also

aids the digestibility of some foods (e.g., potatoes, rice) that are largely indigestible in the raw state.

The other major constituent of carbohydrate in food is the indigestible dietary fiber, which consists mainly of cellulose. Cellulose and other polysaccharides, i.e., hemicelluloses and pectins, are largely responsible for the texture and structure of plant foods.

Canning appears to have little effect on total dietary fiber levels, but the exact effects of canning on the various constituents of dietary fiber and the effect of heat-induced fiber breakdown are not fully known. (*See* **Carbohydrates**: Classification and Properties; **Dietary Fiber**: Properties and Sources.)

Minerals

Total mineral levels are not generally adversely affected by the canning process as they are relatively stable under conditions of heat, acid, or alkali (Table 3). However, minerals are susceptible to changes in bioavailability due to interactions with other food components. The bioavailability of iron can be enhanced during canning in the presence of vitamin C, or reducing sugars with which it forms available complexes. Both oxalates (which occur naturally in many acidic plants) and phytates (from cereals) can inhibit calcium bioavailability.

The major changes that can occur in mineral levels on canning are caused by movement between the foodstuff and the canning liquor. Certain minerals, especially sodium and calcium, can be taken up by the food from the canning liquor; this can be seen especially during the canning of vegetables in brine. Minerals can also be leached out from the foodstuff into the canning liquor. Potassium is particularly prone to leaching with losses of between 15 and 50% seen on canning of vegetables. Zinc, manganese, and cobalt are also susceptible to leaching.

No further substantial changes in sodium or calcium levels are seen on storage of canned vegetables, but slight further leaching of potassium and zinc does occur.

Table 3 Mineral content in freshly cooked and canned cooked peas on a wet-weight basis

	Mineral (mg per 100 g)				
Sample	*Ca*	*Na*	*K*	*Zn*	*Fe*
Fresh	48	65	179	0.82	1.4
Zero time canned	47	320	152	1.0	1.4
Canned stored					
3 months	40	315	79	0.72	1.3
6 months	31		82	0.44	0.9
9 months	28	295	84	0.53	1.5
12 months		280	108	0.55	1.2

Vitamins

Most vitamins are unstable under conditions of heat and are susceptible to loss during the canning process (Table 4). The fat-soluble vitamins are generally more stable than the water soluble vitamins, but losses can occur during canning due to oxidation. Carotenoids are particularly prone to oxidation during heat processing, but this can be greatly reduced by the addition of antioxidants, e.g., vitamin C.

Losses of water soluble vitamins during canning can be quite considerable, with vitamin C being the most labile. Vitamin C can be lost through the following processes: (1) oxidation, which can occur in the early stages of heat treatment before the ascorbic oxidase is inactivated; (2) chemical degradation, e.g., losses due to nonenzymatic browning reactions in fruit products with high vitamin C levels; or (3) through leaching into the canning liquor, which is generally the major cause of vitamin C loss in canned fruits and vegetables. The levels of vitamin C remaining in canned vegetables have been shown to be as low as 20% of the vitamin C found in the fresh raw produce; however, vitamin C is lost during all stages of fresh storage, preparation, cooking, or processing, and some canned vegetables have levels of vitamin C similar to the fresh product, which has been stored prior to cooking (Table 5). Thiamin (vitamin B_1) is the most heat-sensitive of the B vitamins, especially under alkaline conditions and in the presence of bisulfite ions. Thermal degradation of thiamin involves the cleavage of its methylene bridge, which gives rise to many volatile products, producing a meat-like aroma in cooked foods. In the presence of reducing sugars, thiamine may take part in nonenzymatic browning reactions, and it also reacts with aldehydes in the presence of ascorbic acid. Thiamin can also be lost through leaching, but it is less soluble than vitamin C, and retention levels of between 60 and 90% are usual on canning.

Folic acid is lost on canning through heat degradation and oxidation, although it is stabilized in the presence of ascorbic acid, whereas pyridoxine

Table 4 Vitamin losses on canning of vegetables

Vitamin	*Typical losses (%)*
Vitamin C	30–90
Thiamin	16–83
Riboflavin	20–70
Nicotinic acid	20–70
Folic acid	30–80
Panthothenic acid	30–85
Vitamin B_6	30–90
Biotin	0–80
Vitamin A	0–80

Table 5 Vitamin C content of fresh and canned peas and carrots (mg per 100 g)

	Carrots		*Garden peas*		
	Uncooked	*Cooked*	*Uncooked*	*Cooked*	*Pea canning liquor*
Fresh	9	8	30	16	
Stored for 7 days at ambient	3	4	17	16	
Canned	2	3	20	12	11
Canned, stored for 6 months	2	2	15	2	10
Canned, stored for 12 months	2	3	8	5	5

(vitamin B_6) can be lost through heat degradation and leaching. Losses of these two vitamins during canning of fruits and vegetables range between 30 and 80%. Losses on canning of meat can be very considerable, with up to 90% losses.

Riboflavin (vitamin B_2) and nicotinic acid are both relatively heat-stable, with minimal losses on canning of meat products; however, losses of between 20 and 50% have been found on storage of canned meats. Losses seen during the canning of fruits and vegetables, ranging from 25 to 70%, are mainly attributed to leaching. Nicotinic acid and riboflavin both show very high retention levels during milk processing, but riboflavin is lost from bottled milk as it is very sensitive to sunlight.

Heat processing generally has a detrimental effect on most vitamin levels, but mild heating conditions can have a beneficial effect due to enzyme inactivation and the breakdown of binding agents, which increases the bioavailability of certain vitamins, e.g., biotin and nicotinic acid.

Conclusions

The canning of foods has resulted in a wider choice of nutritious, good-quality foods being available all year round in a convenient form for the consumer. Although, with better global links and improved agronomy, many foods are now available fresh all year round, canned foods still form an important part of the food marketplace.

When considering the quality of canned foods, it is important to compare them with fresh or frozen foods at the point of consumption. Many of the changes that occur in both sensory and nutritional aspects do so during any thermal process, whether it is conventional cooking, blanching, or canning. For most foods, the canning process replaces a conventional cooking process, and any mild reheating stage has no further significant effect on quality.

Losses in heat-labile nutrients such as vitamins can be significant. However, as canned products are usually produced from materials at optimum maturity and immediately after harvest, levels can often be as high as the 'fresh' material purchased from the greengrocers and prepared in the home.

See also: **Dietary Fiber**: Properties and Sources; **Fatty Acids**: Properties; **Oxidation of Food Components**; **Pectin**: Properties and Determination; **Phenolic Compounds**; **Starch**: Structure, Properties, and Determination

Further Reading

Bender AE (1978) *Food Processing and Nutrition*. London: Academic Press.

Hall MN, Edwards MC, Murphy MC and Pither RJ (1989) *Technical Memorandum No. 553. CFDRA. A Comparison of the Composition of Canned, Frozen and Fresh Garden Peas as Consumed*. Chipping Campden: CCFRA.

Harris RS and Kamas E (eds) (1988) *Nutritional Evaluation of Food Processing*, 3rd edn. West port, CT: AVI.

Lund D (1982) Effect of processing on nutrient content and nutritional value of food. In: *Heat Processing*.

Moyem T and Kvale O (eds) (1977) *Physical, Chemical and Biological Changes in Food Caused by Thermal Processing*. Essex: Applied Science Publishers Ltd.

Priestley RJ (ed.) (1979) *Effects of Heating on Foodstuffs*. Essex: Applied Science Publishers Ltd.

Rees JAG and Bettison J (eds) (1991) *Processing and Packaging of Heat Preserved Foods*. Glasgow, UK: Blackie.

Richardson T and Finley JW (eds) (1985) *Chemical Changes in Food During Processing*. AVI: Westpart, CT.

Canola *See* **Rape Seed Oil/Canola**

CARAMEL

Contents
Properties and Analysis
Methods of Manufacture

Properties and Analysis

P Tomasik, University of Zimbabwe, Harare, Zimbabwe

Introduction

Home cooks used to burn sugar to obtain so-called caramel for flavoring food. Commercial caramel differs from the domestic product by its method of preparation and consequently in many of its properties. Several products called caramel are available commercially. Tables presented in this article summarize their physical and chemical properties and analytical problems are also discussed.

Definition of the Product

Caramel is a brown product originating from various sugars when they are heated either dry or in concentrated solutions and alone or with certain additives. It is designed as a food additive or ingredient for coloring and/or flavoring food.

In some countries the name 'caramel' is reserved for products manufactured from saccharides in the absence of nitrogen-containing compounds. Such products are used as flavoring ingredients. The products resulting from heating sugars with nitrogen-containing additives are called 'sugar colors' and they serve as coloring additives. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Types and Standardization

The origin and properties of caramel are laid down in many producing countries. Also, the use of only a few, among many reported additives, is permitted in the manufacture of caramels. The World Health Organization and the Food and Agriculture Organization Joint Expert Committee on Food Additives distinguishes three general kinds of caramel: (1) caramel color, plain (CP), (2) caramel color, ammonia process or caustic (AC), and (3) caramel color, ammonium sulfite process (SAC), depending on the additive used in their manufacture. Both the European Technical Caramel Association (EUTECA) and the International Technical Caramel Association (ITCA) have standardized the properties of four classes and 10 types of caramel (Table 1).

4(5)-Methylimidazole in ammonia and ammonium sulfite caramel colors cause a special problem. This compound, which is formed from sugars and ammonia during manufacture of the caramel, is strongly neurotoxic. For this reason the daily intake of food containing ammonia caramel is controlled in some countries and in others ammonia caramel is banned.

Physical Properties

Caramel is polymeric in its character. Three main components of the plain type are called caramelan, caramelen, and caramelin. Ammonia caramels additionally contain melanoidin, which is much darker in color than the three other components mentioned above. Hence, ammonia caramels are the most intensively colored.

Depending on their isoelectric points (p*I*s) caramels may be roughly divided into positive (p*I* 5.0–7.0), negative (p*I* 4.0–6.0) and spirit (p*I* < 3.0) types. The p*I* determines the possibility of application of caramels. The average molecular weight of compounds in caramels is between 5 kDa (electropositive caramels) and 10 kDa (electronegative caramels). Other important properties are pH (Table 1), aqueous solubility (should be completely soluble), specific gravity (usually 1.315–1.345), color intensity (Table 1) or tinctorial strength, hue index (called 'redness') as well as flavor and aroma. These last two organoleptic properties consist of two components: taste arising from the acidity (modifiable) and a taste contribution attributed to the nature of caramel (nonmodifiable).

Chemical Nature

The products of caramelization are distributed among volatile and nonvolatile fractions. The volatile fraction consists mainly of water, carbon monoxide, carbon dioxide, formaldehyde, acetaldehyde,

Table 1 Classes of caramel color according to International Technical Caramel Association/European Technical Caramel Association

Class	*Sort*	*Type*	*Color intensity (ε_M at 610 nm)*	*EBC units ($\times 10^3$)*	*Dry residue (%)*	*4-Methylimidazole (mg kg^{-1})*	*Total nitrogen (%)*	*Total sulfur (%)*	*Ammonia nitrogen (%)*	*Sulfur dioxide (%)*
I	Caramel color, plain	CP-1	5–35	2–12	55–75	< 25	< 0.1	< 0.1	< 0.01	< 0.005 (< 0.1)
		CP-2	40–80	15–25						
II	Caramel color, caustic sulfite	CCS-1	40–80	15–25	62–82	< 25	< 0.1	0.15–2.5	< 0.01	< 0.15
III	Caramel color, ammonia	AC-1	60–90	16–24		< 200	0.5–3.0	(< 0.7)	< 0.5	< 0.015 (< 0.02)
		AC-2	100–140	27–37.5	55–75		0.5–5.0	< 0.3		
		AC-3	150–200	40–54			1.5–6.5			
IV	Caramel color, ammonium sulfite	SAC-1	35–70	8–16		0.1–1.3	0.3–2.0	0.3		0.08
		SAC-2	75–100	17.5–23	55–75	0.5–2.8	0.8–3.2	0.5 (< 0.5)		0.12 (< 0.1)
		SAC-3	105–150	22.5–37		0.8–2.8	1.0–4.0	0.8		0.15
		SAC-4	210–270	40–52	47–57	2.0–4.0	1.0–5.0	1.5		0.28

The values in parentheses are Food and Agriculture Organization/World Health Organization (FAO/WHO) proposals dating from 1980. Permissible content (in mg kg^{-1}) of the heavy metals in caramel according to FAO/WHO standards are: total, 25; copper, 20; lead, 5; astatine, 3; mercury, 0.1. ε_M, Molar extinction coefficient; EBC, European Brewery Convention.

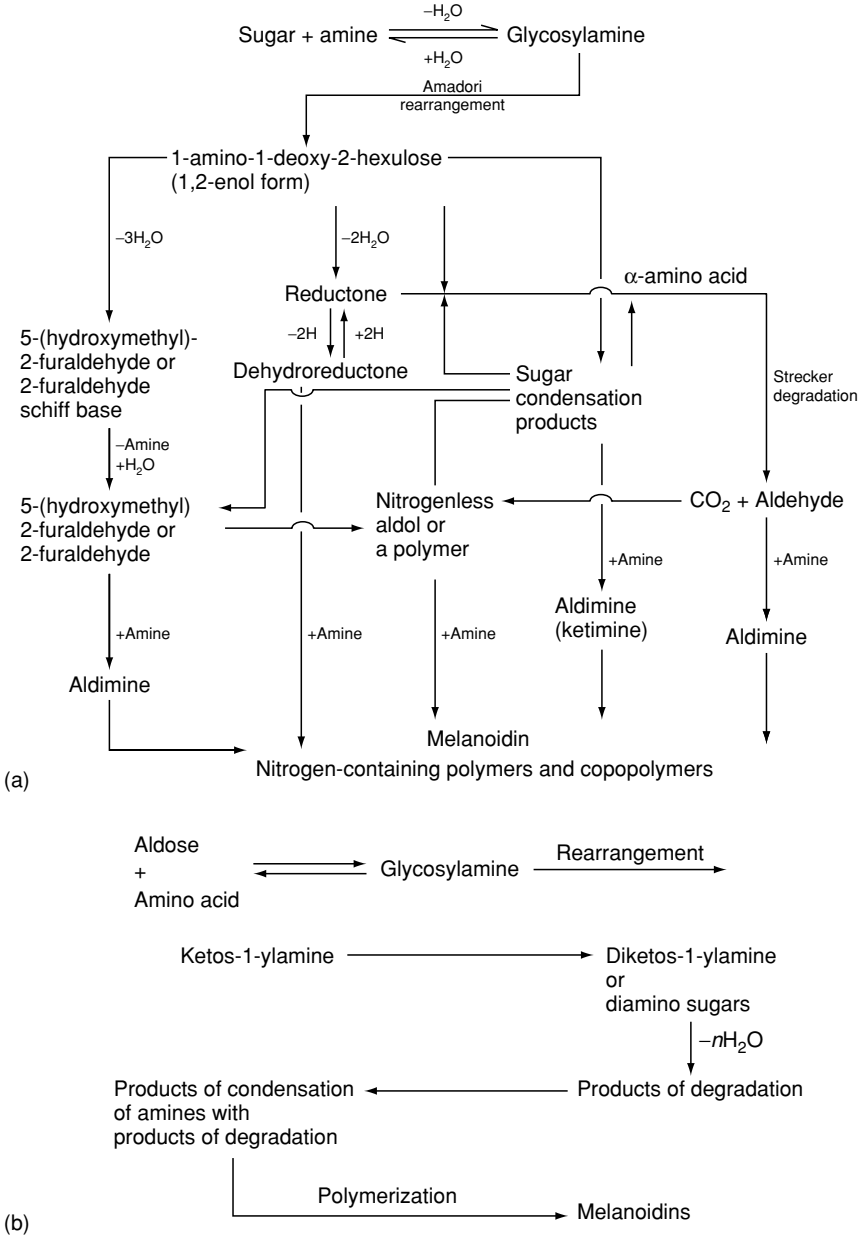

Figure 1 Theoretical schemes for the formation of melanoidins according to (a) Hodge and (b) Reynolds. Reproduced from Caramel/Properties and Analysis, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

methanol, and ethanol. The composition of the low-molecular-weight portion of the nonvolatile fraction depends on the type of caramel. The plain caramel nonvolatile fraction contains the following saccharides: D-fructose, D-glucose, kojibiose, isomaltose, nigerose, sophorose, laminarabiose, maltose, gentobiose, cellobiose, isomaltotriose, panose, and other oligosaccharides. There are also several carboxylic acids: formic acid and higher fatty acids, succinic, fumaric, pyruvic, levulinic, and furancarboxylic acids. Oxaheterocycles are represented by 2-furaldehyde, 5-(hydroxymethyl)-2-furaldehyde, 2,3-dihydro-4-hydroxy-5-methylfuran-3-one, 2,3-dihydro-4-hydroxy-2-(hydroxymethyl)-5-methylfuran-3-one, bis(5,5-formylfurfuryl) ether, and 2,3-dihydro-3,5-dihydroxy-6-methyl-4*H*-pyran-4-one. There are also various other aliphatic carbonyl compounds like glucoreductone (2-hydroxypropandial), pyruvaldehyde, and α,β-unsaturated carbonyl compounds. A number of other components which are products of reversion and polymerization remain uncharacterized. The low-molecular-weight fraction of caustic caramel

contains about 50 components. Ammonia caramel contains in this fraction a certain amount of 4(5)-methylimidazole, 2-acetyl-4(5)-(1,2,3,4-tetrahydroxybutyl)imidazole, and 3-hydroxy-6-methylpyridine, as well as at least 143 other components, including those mentioned above. (*See* **Carbohydrates**: Classification and Properties; **Fatty Acids**: Properties; **Fructose**; **Sucrose**: Properties and Determination.)

The high-molecular-weight, nonvolatile fraction of plain caramels is composed of 1,6-anhydro-α-D-glucose, caramelan (formed in the reaction shown in eqn (1)), caramelen (resulting from the reaction shown in eqn (2)), caramelin (being the product of reaction shown in eqn (3)), structures of which remain unknown, and polymers composed of furan ring units.

$$6C_{12}H_{22}O_{11} \xrightarrow{-12H_2O} 6C_{12}H_{18}O_9 \qquad (1)$$

$$6C_{12}H_{22}O_{11} \xrightarrow{-18H_2O} 2C_{36}H_{48}O_{24} \qquad (2)$$

$$6C_{12}H_{22}O_{11} \xrightarrow{-27H_2O} 3C_{24}H_{26}O_{13} \qquad (3)$$

The interpretation of carbon-13 cross-polarized magic-angle sample spinning nuclear magnetic resonance (^{13}C CP/MAS NMR) spectra of plain caramels leads to the conclusion that there are 8–9% carbonyl and aldehyde carbon atoms, 7–7.7% carbon atoms in ester groups, 33–31% heterocyclic carbon atoms, and 52.4–51.2% alkyl groups bonded to carbon atom and to oxygen. It may also be deduced that the furan system prevails among other heterocyclic systems present in caramel. (*See* **Spectroscopy**: Nuclear Magnetic Resonance.)

Two consecutive steps may be distinguished in the formation of caramels without participation of ammonia, ammonium sulfite, amines, or amino acids and their salts, which may also be used as the catalysts. They are degradation reactions which lead to colorless or yellow compounds. First, water and carbon dioxide are evolved and furan derivatives are formed (2-furaldehyde from furanoses and 5-(hydroxymethyl)-2-furaldehyde from pyranoses). This step is a function of both temperature and time; atmospheric oxygen plays some role in the final stage. This is followed by the second step, polymerization and condensation reactions forming highly colored compounds (caramelan, caramelen, and caramelin as well as furan polymers).

In caramels manufactured with basic nitrogen-containing additives, melanoidins are the most important components. Their structure is unknown, but there is evidence that they have, at least in part, an amide structure. Moreover, they contain unpaired electrons, i.e., they have a free radical character. There is a great similarity between this product and humic acids resulting from the Maillard reaction. Two consistent schemes for the formation of melanoidins are proposed by Hodge (Figure 1(a) and Reynolds (Figure 1(b)). Both theories differ in the number of steps involved in the route from the preliminary reaction between sugar and amine to give glycosylamine and the final product, melanoidin. Recent studies have revealed that the possibility of the Strecker degradation as one of the processes involved should be rejected. (*See* **Browning**: Nonenzymatic.)

Analysis of Caramels

The use of caramel is first of all limited by its type, which affects its use for a given food. For instance, a caramel may precipitate on contact with a product subjected to coloration, as a result of discharge of caramel micelles. The type and quality of caramel may easily be recognized by means of the citric acid, alcohol, and Lassaigne tests (Table 2). However, the complexity of caramel composition makes distinguishing between the different types difficult unless combined analytical methods are applied. These include chromatography of dilute solutions of caramel followed by thin-layer electrophoresis and size exclusion chromatography. Thus, full standardization of caramel may require many other determinations of chemical components as well as fractions, as shown in Table 3. This table presents several properties of certain types of caramel manufactured from sucrose. It is apparent that, for instance, two I CP-2

Table 2 Tests for particular types of caramel

Type of caramel[a]	*Test*		
	Citric acid[b,c]	*Alcohol*[b,c]	*Lassaigne*
Spirit caramel			
I A	+	−	−
I B	−	−	−
I C	+	−	−
Nonammonia caramel			
II	+	−	−
Beer caramel			
III A	−	±	+
III B	−	−	+
III C	−	+	+
Soft-drink caramel			
IV A	−	+	+
IV B	−	+	+

[a]Consult also Tables 1 and 3 for classification of types.
[b]Formation of a precipitate.
[c]Solubility in citric acid–monosodium phosphate at pH 2.5.
[d]Solubility in 13:7 (v/v) ethanol–water.

Table 3 Analytical characteristics of some classes, sorts, types, and subgroups of caramel colors

Properties	*Type (carbohydrate, additive)*						
	I A (sucrose, sodium carbonate)			*I B (sucrose, acid)*			*II (sucrose, sodium sulfite)*
	CP-2	CP-1	CP-2	CP-1	CP-1	CP-1	CCS
pH	3.7	3.1	3.1	3.6	3.9	3.5	3.0
Neutralization							
Citric acid test	+	+	+	+	−	−	+
Alcohol test	−	−	−	−	−	−	−
Color intensity (ε_{max})	65	32	71	29	38	29	100
Dry substance (%)	65.2	69.6	64.1	74.1	70.3	72.8	95.0
Ash (%)	0.55	1.16	0.47	0.09	0.38	0.06	0.07
Sodium (mg kg^{-1})	1825	2398	1856	126	1422	151	57
Sulfur dioxide (mg kg^{-1})	NF	NF	NF	NF	NF	NF	66
Sulfates in ash (%)	NF	NF	NF	NF	NF	NF	NF
Total nitrogen (%)	0.08	NF	NF	NF	NF	NF	NF
Basic nitrogen (mg kg^{-1})	NF	NF	NF	NF	NF	NF	
Formic acid (mg kg^{-1})	455	577	664	249	384	343	200
Glucose (%)	3.7	14.5	9.3	24.8	7.0	16.6	2.3
Fructose (%)	3.1	11.2	7.9	15.6	2.0	3.8	0.4
Sucrose (%)	0.6	NF	NF	NF	NF	0.8	NF
4-Methylimidazole (mg kg^{-1})	NF	NF	NF	NF	NF	NF	NF
Glucoreductone (%)	0.15	0.23	0.25	0.05	0.05	0.06	0.18
Gel permeation analysis							
Color between R_{gp}	0.5–1.75	0.39–1.66	0.36–1.65	0.06–1.13	−0.02–10.7	−0.02–1.02	−0.05–1.39
Maxima at R_{gp}	0.98	0.91	0.91	0.35	0.32	0.32	0
	1.32	1.16	1.16	0.81	0.61	0.84	0.55
				0.96	0.86		1.09
Minima at R_{gp}	1.18	1.02	1.02	0.74, 0.87	0.56, 0.77	0.73	0.91

Properties	*Type (carbohydrate, additive)*					
	III A (sucrose, ammonium carbonate)			*III B (sucrose, ammonia)*		
	AC 2	AC 3	AC 2	AC 1	AC 1	AC 2
pH	4.7	5.6	5.7	5.3	4.1	4.5
Neutralization		+	+	+	+	
Citric acid test	−	−	−	−	−	−
Alcohol test	−	+	+	−	−	−
Color intensity (ε_{max})	108	194	113	79	86	99
Dry substance (%)	63.4	93.6	70.5	60.9	62.6	64.1
Ash (%)	0.30	0.58	0.27	0.36	0.45	0.38
Sodium (mg kg^{-1})	1198	1688	−	1265	1355	953
Sulfur dioxide (mg kg^{-1})	NF	NF	NF	NF	NF	NF
Sulfates in ash (%)	NF	NF	NF	NF	NF	NF
Total nitrogen (%)	4.73	6.7	2.5	4.70	4.95	4.45
Basic nitrogen (mg kg^{-1})	240	1690	NF	250	180	220
Formic acid (mg kg^{-1})	673	491	296	542	415	671
Glucose (%)	23.0	3.5	10.0	25.8	23.7	16.1
Fructose (%)	10.9	NF	8.0	10.7	8.9	5.1
Sucrose (%)	1.7	NF	NF	NF	5.4	3.1
4-Methylimidazole (mg kg^{-1})	128	119	51	118	151	47
Glucoreductone (%)	0.54	0.53	0.44	0.54	0.62	0.54
Gel permeation analysis						
Color between R_{gp}	−0.99–1.66	−0.07–1.75	−0.09–1.69	0.41–1.80	0.39–1.73	0.54–1.61
Maxima at R_{gp}	0	0	0	0.84	0.91	0.97
	0.93	0.89	0.93	1.30	1.34	1.27
	1.23	1.18	1.23			
Minima at R_{gp}	0.07	0.25	0.09	1.02	1.04	1.11
	1.05	1.05	1.05			

Continued

Table 3 Continued

Properties	*Type (carbohydrate, additive)*		
	IVA (sucrose, ammonium sulfite)		
	SAC 2	SAC 3	SAC 2
pH	4.1	5.8	3.9
Neutralization		+	
Citric acid test	−	−	−
Alcohol test	+	+	+
Color intensity (ε_{max})	78	123	98
Dry substance (%)	65.0	68.3	68.4
Ash (%)	1.80	2.18	1.96
Sodium (mg kg^{-1})	5230	5307	5540
Sulfur dioxide (mg kg^{-1})	369	337	234
Sulfates in ash (%)	49.6	67.1	54.5
Total nitrogen (%)	1.35	1.17	1.17
Basic nitrogen (mg kg^{-1})	260	150	280
Formic acid (mg kg^{-1})	277	438	393
Glucose (%)	28.8	31.4	28.5
Fructose (%)	3.8	NF	NF
Sucrose (%)	12.0	NF	NF
4-Methylimidazole (mg kg^{-1})	117	40	37
Glucoreductone (%)	0.07	0.09	0.09
Gel permeation analysis			
Color between R_{gp}	−0.07–1.39	−0.04–1.34	−0.55–1.39
Maxima at R_{gp}	0	0	0
	0.30	0.27	0.27
	0.98	0.93	0.93
Minima at R_{gp}	0.09	0.07	0.09
	0.93	0.91	0.91

NF, not found; R_{gp}, retention factor on Sepharose C1-6B (calculated from retention volumes of caramel components relative to those of Blue Dextran and $NaHCO_3$).
Data abstracted from Hellwig E, Gombocz E, Frischenschlager S and Petuely F (1981) Detection and identification of caramel color by gel permeation chromatography. *Deutsche Lebensmitteln Rundschau* 77: 165–174.

Table 4 Methods of detection of caramel in food

Method	*Reagent*	*Appearance*
Jaegerschmidt	Resorcinol + hydrochloric acid, ether, or acetone	Red color in ether Violet red color in acetone
Amthor	Paraldydehyde + absolute alcohol	Brown precipitate after 24 h which reacts with phenylhydrazine hydrochloride to give a solid insoluble in hydrochloric acid but soluble in ammonia and alkali
Griessmeyer–Aubry	Ammonium sulfate in 96% ethanol	Yellow to brown color
Lichthard	Tannin + sulfuric acid	Brown solid within 24 h
Fradiss	Dry 1-pentanol	A precipitate
Crampton–Simons	Floridin, Tonsil or Fuller's earth	Decoloration of aqueous or ethanol solutions in caramel followed by colorimetric determination of resulting color
Straub	1% aqueous $SnCl_2$ + potassium acetate	Light yellow color and precipitate
Nessler–Carles	Fresh egg white	Brown to orange color
Ihl	Pyrogallic acid in hydrochloric acid	Dark red precipitate
Magalhaes	K_2SO_4 + cotton wool	Light orange color on boiling for 10 min
Schenck	Phenol or 2-naphtol	Red brown color immediately (phenol) or after 30 min (2-naphthol)

caramels listed therein differ from one another in the majority of their properties. It is a matter of discussion whether these findings will result in further attempts to standardize caramels within subtypes. These differences may also be considered as resulting from difficulties in the strict control of the caramelization process, which tends to be rather chimeric. (*See* **Chromatography**: Principles.)

Determination of the p*I* may be carried out by means of electrophoresis, the flocculation test with tannin, ionic surface-active agents (an industrial approximate measure), the gelatin test, and other more specific methods.

Isolation and separation of caramelan, caramelen, and caramelin may be achieved by dialysis, electrofiltration through an ion exchange membrane, or by determining solubilities in different solvents (84% aqueous ethanol, 1-propanol). Fractionation by gel filtration and adsorption on either charcoal or using ion exchangers are recommended mainly for elimination and separation of overall coloring matter from caramel. Anionic resins exhibit particular selectivity towards caramelan, which is adsorbed, leaving caramelen and caramelin in the unadsorbed state.

The determination of 4(5)-methylimidazole in ammonia caramel involves chromatographic techniques. Thus, extracts of caramels are developed on silica gel F_{254}-coated plates using a 4:1:1 ether–chloroform–methanol mixture (sodium nitrite with sulfanilic acid used as a spray). Gas–liquid chromatography involves columns packed with 10% carbowax 20 mol l^{-1} with 25% potassium hydroxide on CPLA (80–100 mesh). (*See* **Chromatography**: Gas Chromatography.)

There are several methods of detection of caramel in food. These are presented in Table 4. Furthermore, physical methods based on size exclusion chromatography and spectral measurements in the ultraviolet and visible region have recently been developed.

See also: **Browning**: Nonenzymatic; **Carbohydrates**: Classification and Properties; **Chromatography**: Principles; Gas Chromatography; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Fatty Acids**: Properties; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Fructose**; **Spectroscopy**: Nuclear Magnetic Resonance; **Sucrose**: Properties and Determination

Further Reading

FAO *Nutritional Meeting Report Series*, no. 57. Geneva: Food and Agriculture Organization.

Greenshields RN and Macgillivray AW (1972) Caramel. 1. Browning reactions. *Process Biochemistry* 7(12): 11–13, 16.

Hellwig E, Gombocz E, Frischenschlager S and Petuely F (1981) Detection and identification of caramel color by gel permeation chromatography. *Deutsche Lebensmitteln Rundschau* 77: 165–174.

ITCA/EUTECA (1979) *ITCA/EUTECA Specifications for Caramel Color, 3rd Meeting of EUTECA Sub-Committee Specifications*, 4th May. Paris: European Technical Caramel Association.

Tomasik P, Palasinski M and Wiejak S (1989) The thermal decomposition of carbohydrates. Part I. The decomposition of mono-, di, and oligosaccharides. *Advances in Carbohydrate Chemistry and Biochemistry* 47: 203–278.

Truhaut R, Vitte G and Lassale-Saint-Jean V (1962) Bibliographic study of caramel. *Bulletin de la Société Pharmaceutique Bordeaux* 101: 97–120.

WHO (1977) *WHO Technical Report Series*, no. 617. Geneva: World Health Organization.

Methods of Manufacture

P Tomasik, University of Zimbabwe, Harare, Zimbabwe

Introduction

Commercial caramels, even if manufactured from the same materials, may have different properties depending on the additives used and the conditions of caramelization. Caramelization itself is a very chimeric process and the manufacture of caramel is sometimes considered to be an art. Ready-made caramel is a rather unstable product and its storage is also a very important factor. The problems of production and storage of caramel are presented below.

Applications of Caramel

The classification of caramel into four classes results from the properties of the product and its intended food application. Thus, caramel of class I (CP-1 and CP-2) is designed as an additive to spirits, brandy, sweets, medicines, biscuits, pastries, aromas and spices. Caramel of class II (CSS-1) has very limited applications as it is used for special spirits as a flavoring rather than a coloring agent. Class III caramel (AC-1, AC-2, and AC-3) is used as brown color for beer, malt liquor, bread, biscuits, pastries, soups, sauces, canned food, meat, tobacco, and some spices. Finally, class IV caramel (SAC-1 to SAC-4) is the colorant of cola-type beverages, soft drinks, vermouths, and vinegar. The use of caramels for nonfood

purposes is of marginal importance. The data published for 1986 set the world production at the level of 60 000 tonnes: in the USA caramel of class IV was the only product manufactured to any significant extent (95% of overall production) whereas in the European Community the manufacture of both class IV (50%) and class III (45%) caramels was important (class I 4%; class II 1%). (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Flavor (Flavour) Compounds**: Structures and Characteristics.)

The isoelectric point and tinctorial strength are the most important criteria for selecting a class of caramel for a given purpose. However, flavor is also an important property. Incorrectly selected caramels can produce a haze in drinks and flocculation as well as nonuniform shades in finished products. Beer caramel (class III) has to withstand fermentation. The composition of caramel micelles, especially their calcium content, is another factor which can cause turbidity in some drinks finished with caramel. Caramel in brandy (0.2% v/v) accelerates its aging. Aspartame in drinks is stabilized by the addition of caramel. Oriental cuisines commonly utilize caramel for coloring and flavoring soups, gravies, and sauces, e.g., soy sauce (shoyu).

Sources for Manufacture

Some authors report that the quality of caramel depends, among other aspects, on its source. However, some authors express the opposite point of view, claiming that only parameters of caramelization (including catalyzing additives) are responsible for the quality of the final product. Undoubtedly the presence of amino acids, proteins, and hydroxy acids in materials used for caramelization contributes additional flavor and other specific organoleptic properties to the final product. As a matter of fact, such additives have a catalyzing role in the formation of the brown-colored components of caramel. There is also some relationship between the content of D-glucose in the stock and firmness of caramel. D-Glucose decreases the hygroscopicity and maltose has practically no effect on it.

Sucrose as well as D-glucose and D-fructose, both resulting from its hydrolysis, are prime sources for the manufacture of caramel. Reducing sugars caramelize more readily than nonreducing sugars. The mode of preparation of sugar for caramelization has some influence on the caramelization process. Sugar from carbonation is better than that from sulfination as residual carbonates catalyze the caramelization better than sulfites. (*See* **Fructose**; **Sucrose**: Properties and Determination.)

Other mono- and disaccharides have also been considered as sources of caramels, but they are only of theoretical importance. Molasses has attracted the attention of manufacturers as a relatively inexpensive source of caramel because of its brown components. A disadvantage in the the use of molasses is its high potassium content and the unfavorable viscosity of the resulting caramel. (*See* **Carbohydrates**: Classification and Properties.)

Many reasons, among them economic and political, make sucrose, invert sugar, and D-glucose rather unfavorable sources for the production of caramel. Oligo- and polysaccharides, which are hydrolyzed by acids, bases, or enzymes, provide a source of very stable caramels. Maize, cassava, sago, and potato starch as well as starch waste may be employed. Starch syrups from enzymatic hydrolysis deliver caramels with a higher tendency towards crystallization due to their higher content of dextrins after acid hydrolysis. Microwave heating of starch in a sealed vessel causes its hydrolysis accompanied by caramelization of the hydrolysate. Nonconventional sources such as malt and soya bean carbohydrates have also been paid some attention as the sugar syrups derived from them contain 70–85% reducing sugars. (*See* **Starch**: Sources and Processing.)

Additives and Catalysts of Caramelization

The caramelization of plain sugars produces flavoring rather than coloring caramels. Certain additives accelerate caramelization, influencing both flavor and tinctorial strength of caramel by being either reagents or catalysts. The use of the following additives has been published: acids – acetic, citric, phosphoric, sulfurous, sulfuric, and carbonic acids; bases – ammonia as well as hydroxides of sodium, potassium, and calcium; salts – carbonates, hydrogencarbonates, sulfates, sulfites or phosphates of ammonia, sodium, potassium, and calcium. Alkaline additives catalyze caramelization of furanoses more efficiently than pyranoses. Some sodium compounds, mainly biogenic amino acids and their sodium, potassium, magnesium, and calcium salts, taurine (2-aminoethanesulphonic acid) and sulfanilic acid, have also been tested. They may be of particular interest in view of the fact that the most effective catalyst, ammonia, produces caramel contaminated with the neurotoxin 4(5)-methylimidazole. Caramel sources as well as additives and catalysts are controlled by food laws of particular countries or economic unions. (*See* **Legislation**: Additives.)

Apart from chemical catalysts the possibility of catalysis of caramelization by ultraviolet, microwave

or γ-radiation or ultrasound has also been studied with inconclusive results. In particular, ultraviolet and γ-radiation introduce competing reactions such as the free radical decomposition of carbohydrate to water and carbon dioxide and a number of lower carbonyl compounds (aldehydes, ketones, carboxylic acids). In the case of caramels with poor tinctorial strength (class I CP-1 and CP-2) attempts have been made to increase their coloring ability by blending ready-made caramels with certain additives. Among possible additives enhancing the tinctorial strength of caramels the following have been tested: magnesium and calcium hydroxides, calcium phosphate as well as oxides of magnesium, calcium, zinc, and cobalt(II). Magnesium oxide appears to be a superior additive among those tested. Its application has, however, limited value. The increase of tinctorial strength of plain caramels has some limits because the most intensively colored melanoids are absent. The effect of magnesium oxide seems to be due to modification of the micellar structure of plain caramel. This effect enables further dehydration of caramelan to the darker caramelen and caramelin. Apart from such procedures, ultrafiltration, centrifugation combined with size exclusion chromatography as well as ion exchange columns have been proposed as methods to increase the tinctorial strength of caramels. Thus far, these methods have not achieved any application on an industrial scale. (*See* **Chromatography**: Principles.)

Preparation and Manufacture

The variability of the sources for caramelization causes a great deal of empiricism in this technology. Generally, the character of compounds constituting caramel depends on temperature of the process, its duration, and the concentration of reactants. Increase in the color of the product is proportional to the time of the process.

There are four aspects of thermolysis affecting caramelization and all of them have found practical industrial applications:

1. Thermolysis of plain saccharides above their melting temperatures. This can be carried out under normal, reduced, or enhanced pressure. The last approach is usually employed when syrups from the hydrolysis of starch are caramelized. The initiation of the process readily takes place, after which the pressure may be released. The process is allowed to continue open to the atmosphere to develop all the color, viscosity, and desired organoleptic properties. The temperature range is between 180 and 250 °C. Although it is commonly accepted that the contact of the reaction mixture with atmospheric oxygen does not play any role in the formation of caramel, one may see in Figures 1–4 that this is not so. Oxygen slows down the caramelization in later stages of the process. Its effect on the tinctorial strength of the final product is nonuniform and depends on the source. The elimination of nitrogen positively influences acid resistance and solubility of the final product.
2. Thermolysis in the presence of catalyst. More recent procedures allow the temperature of caramelization to be reduced to 120–130 °C. An increase of the temperature to above this range decreases the tinctorial strength of the caramel and develops an acid flavor.

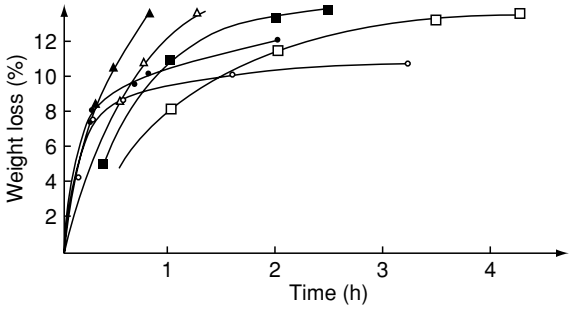

Figure 1 Course of caramelization of glucose (circles), sucrose (triangles) and maltose (squares) in air (open symbols) and under nitrogen (solid symbols). Reproduced from Caramel: Methods of Manufacture, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

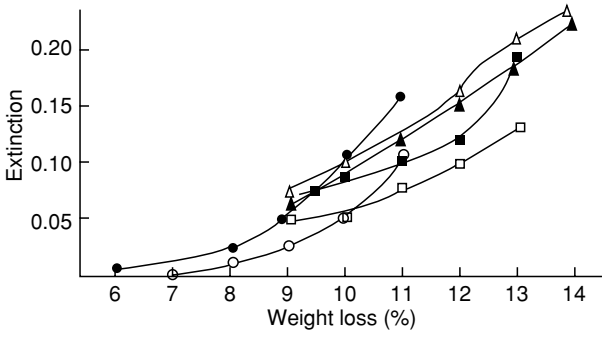

Figure 2 Extinction of the absorption band at 35 000 cm^{-1} in the ultraviolet/visible spectra of 0.1% aqueous solution of caramel from glucose (circles), sucrose (triangles), and maltose (squares). Open and solid symbols represent caramel prepared in air and under nitrogen, respectively. Reproduced from Caramel: Methods of Manufacture, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

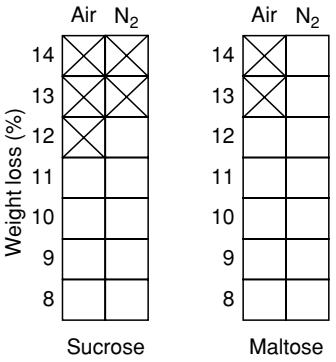

Figure 3 Acid resistance (the hydrochloric acid test) of caramel from sucrose and maltose prepared in air and under nitrogen. Crossed squares denote flocculation of caramel. Reproduced from Caramel: Methods of Manufacture, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

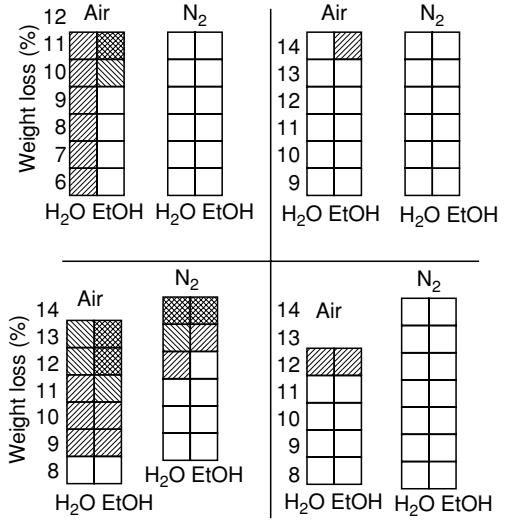

Figure 4 Solubility in water and in 96% aqueous ethanol of caramel from various sugars made in air and under nitrogen: □, solution is transparent; ▨, traces of turbidity; ▧, solution is turbid; ▩, caramel is sparingly soluble. Reproduced from Caramel: Methods of Manufacture, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

3. Thermolysis in the presence of either mineral acids or alkalis. These additives hydrolyze oligosaccharides which further caramelize. This process also requires lower temperature than these listed under (1). Increased pressures may also be used.
4. Thermolysis with ammonia, ammonium salts, amino acids and their salts as well as proteins and polypeptides. Such caramels contain nitrogen heterocyclic components (imidazoles, pyroles, pyrazines and pyridines) which enrich the flavor and aroma of the final product.

Thermolysis should be carried out in entirely stainless-steel equipment – kettles (open or pressurized), lines, storage tanks, fillers, agitators, valves, and so on.

Careful control of the process as it proceeds is very important. The parameters of the caramelization have to be precisely adjusted for a given source and to obtain the desired properties of the final product. Caramels from sucrose, D-glucose, and D-fructose with a distinct content of noncaramelized sugars have the best organoleptic quality. However, such caramels are quite hygroscopic and unstable. There are several techniques for the control of caramelization. One of the most recent developments is measuring the absorbance of free-flowing material in the near-infrared region. Lack of control of the process leads to loss of the micellar character of caramel and, in consequence, precipitation occurs. For this reason the isoelectric point has to be adjusted. This must be done at the beginning of the caramelization as attempts to change the isoelectric point in the course of the process are quite complicated and frequently unsuccessful. The pH of caramels constitutes an important property. A high pH may indicate incomplete caramelization or the presence of alkali. Above pH 6.0, caramel is readily attacked by molds and below pH 2.5 it quite easily resinifies.

The control of the viscosity of caramel is difficult. The rate of evolution of water (dehydration) significantly influences this property. The desired viscosity can be achieved by manipulation of the temperature and contact time with reagents.

Overburn caramel results from badly controlled temperature and from attempts to manufacture a highly colored product. This may occur particularly in the manufacture of ammonia caramel. The control of the temperature is important throughout the whole period of production, including the final stage of killing heat, i.e., the fast cooling of a caramel to about 30 °C.

The concentration and origin of the caramelized sources are of lesser importance. There is a relationship between the viscosity of caramel and its solubility. Less viscous caramels are usually more readily soluble and have more stable tinctorial strength, shelf-life, and retention of complete solubility. Such caramels are stored with the minimum of waste and effort.

For special use solid, dry caramels are manufactured. They are prepared by treating hot (120 °C) viscous caramel with ammonium carbonate followed by adding sucrose and orthophosphoric acid, cooling to 100 °C, and adding citric acid and sodium hydrogen carbonate. An alternative route involves addition of some cereal products, e.g., rye flour, and

conditioning of the mass at 80–85 °C at pH 3.5–5.5. Liquid caramel may also be thickened with a mixture of starch and dextrins. An extrusion of mono- and disaccharides at 150–300 °C also leads to solid caramels.

Storage

Undesirable properties of caramel may appear even if the product has been properly manufactured. Caramels are not fully stable and caramelization slowly progresses on storage. Therefore, caramel should be stored at low temperatures. Caramelization on storage may be catalyzed by metal ions. Hence tanks should be either plastic-lined or made of stainless steel. These precautions slow down resinification of the product into an amorphous gel which becomes useless as either an additive or ingredient for food and drinks. The stability of caramel stored in ideal conditions is estimated to be about 5 years.

See also: **Carbohydrates**: Classification and Properties; **Chromatography**: Principles; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Fructose**; **Legislation**: Additives; **Starch**: Sources and Processing; **Sucrose**: Properties and Determination

Further Reading

Greenshields RN (1973) Caramel. 2. Manufacture, composition and properties. *Process Biochemistry* 8(14): 17–20.

North RS (1973) Caramel versatile coloring. *Flavour Industry* 4(8): 337–338.

Smolnik MD (1987) Production and application of caramel from starch products. *Staerke* 39: 28–32.

Tomasik P, Palasinski M and Wiejak S (1989) The thermal decomposition of carbohydrates. Part I. The decomposition of mono-, di- and oligo-saccharides. *Advances in Carbohydrate Chemistry and Biochemistry* 47: 203–278.

CARBOHYDRATES

Contents

Classification and Properties

R F Tester and J Karkalas, Glasgow Caledonian University, Glasgow, UK

Background

Carbohydrates are the most abundant organic substances in the biosphere. Plants synthesize them from CO_2 and water, with energy provided by sunlight. The name originated a long time ago because of the erroneous assumption that they were hydrates of carbon having the empirical formula $(CH_2O)_n$. The simplest sugars are known as 'monosaccharides,' which can be polyhydroxyaldehydes or polyhydroxyketones. Disaccharides are formed from two monosaccharide molecules with the elimination of one molecule of water. Oligosaccharides consist of 2–10 monosaccharide units, whereas polysaccharides (glycans) may consist of a large number of monomers (e.g., over 10 000 glucose units). Oligo- and polysaccharides can be hydrolyzed by acid to the constituent monosaccharides. The structure and properties of carbohydrates are discussed in the following sections.

Configuration and Nomenclature

Chiral Molecules

The four valances of the carbon atom may be considered as directed towards the corners of an imaginary regular tetrahedron. In glyceraldehyde (2,3-dihydroxypropanal) (**Figure 1**) carbon atom 2 is

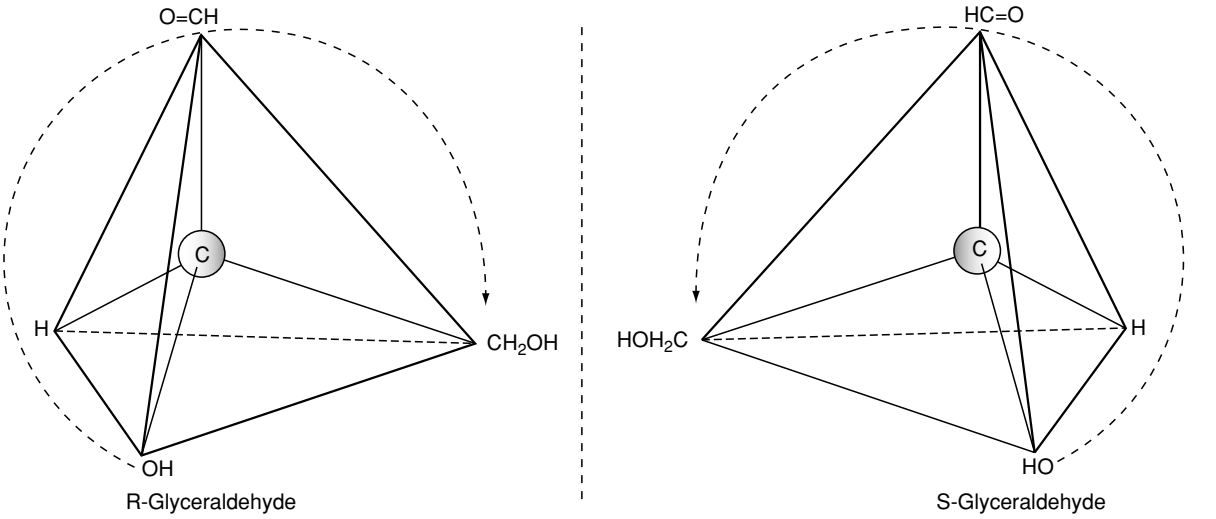

Figure 1 Enantiomers (mirror images) of glyceraldehyde drawn according to the Cahn–Ingold–Prelog rule. The direction of the arrows indicates diminishing priority of the groups OH > HC═O > CH_2OH (H always has the lowest priority). The carbon atom (chiral center) occupies the center of the tetrahedron. The four valences of carbon are directed towards the corners of the tetrahedron to which are bonded different groups. The OH groups are outside the plane of the paper directed towards the observer. If one of the figures is rotated around its vertical axis, its OH group will be moved behind the plane of the paper, thus confirming that the two figures are non-superimposable. See text for discussion.

situated at the center of the tetrahedron, and the four different substituents occupy its corners. Each molecule is asymmetric, i.e., it has no plane of symmetry that can divide it into two identical parts. The two possible molecular structures of 2,3-dihydroxypropanal are nonsuperimposable, and they are mirror images of each other. This property is termed 'stereoisomerism' (Greek *stereos* = three-dimensional or solid). The two chemical compounds are also known as 'stereoisomers,' because they have the same chemical formula and differ only in the spatial arrangement of atoms, or groups of atoms, around a stereocenter. In this example, they resemble a pair of human hands and are known as 'chiral molecules' (Greek *cheir* = hand). The asymmetric carbon atom is a chiral center or stereocenter. The two mirror images are also known as 'enantiomers' (Greek *enantios* = opposite). The position of the hydroxyl group in space determines the absolute configuration of each member of the chiral pair according to the rule explained in Figure 1. It should be noted that not all stereoisomers are enantiomers.

The letters R- and S- are derived from the Latin *Rectus* = straight (and, by extension, right) and *Sinister* = left. It should be noted that the R- and S-notation is used in modern organic chemistry to indicate the absolute configuration around a stereocenter. However, the classic D- and L-notation has been retained for the description of the chirality of carbohydrates (similarly for amino acids). D- is derived from the Latin *Dexter* (right) and L- from *Laevus* (left). The use of this notation is explained in Figure 2.

Properties of Enantiomers – Optical Isomerism

Enantiomers have identical chemical and physical properties and are indistinguishable from each other except for the direction of rotation of the plane of polarized light. They are described as optically active. One enantiomer in solution rotates the plane of polarized light to the right, whereas the other rotates it to the left. A chiral molecule that rotates light to the right is known as 'dextrorotatory,' and its chemical name is preceded by the plus sign, e.g., (+)-lactic acid. Its enantiomer, which rotates the plane of polarized light to the left, is described as 'levorotatory' and the minus sign is used, e.g., (−)-lactic acid. This property is known as optical isomerism. A solution containing equal concentrations of two enantiomers does not rotate the plane of polarized light (the two opposing effects cancel each other out) and is known as a racemic mixture. The optical properties of sugars are discussed below.

Polarimetry

The extent of rotation of the plane of polarized light is measured in angular degrees. A solution of an optically active substance is placed in a polarimeter tube (normally 20 cm long), both ends of which are made of optical glass. The tube is positioned in the light path of the polarimeter. If the solution rotates the plane of polarized light to the right, as viewed by

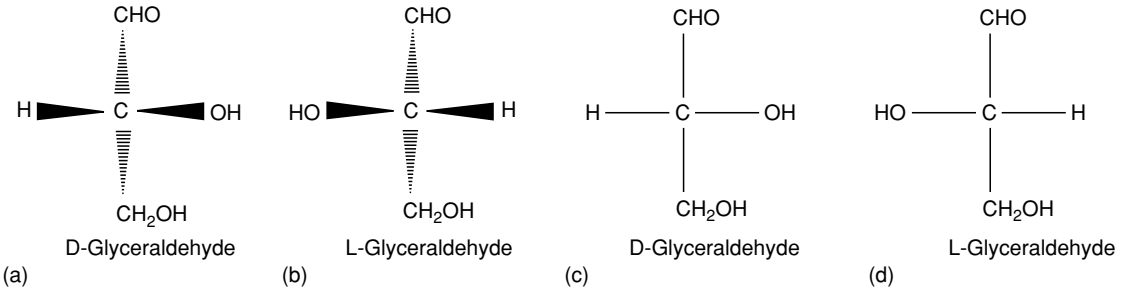

Figure 2 D- and L-notation conventionally used to indicate the position (right or left, respectively) of the OH group of glyceraldehyde (or any sugar), which is attached to the chiral center furthest from the aldehyde group. The carbon atoms are written vertically with the aldehyde group always at the top. In formulae (a) and (b), the carbon atom lies on the plane of the paper. The dark bonds with the respective groups are directed out of the plane of the paper towards the observer. Light bonds and respective groups are directed behind the plane of the paper. Formulae (c) and (d) are known as Fischer projections. They are commonly used to represent sugars, amino acids, organic acids, and other chiral compounds. The respective chiral molecules (mirror images) (a) and (b) or (c) and (d) cannot be superimposed. The rule is that these formulae may be rotated 180° on the plane of the paper only, but they should never be rotated around the imaginary axis that passes through the carbon atoms on the plane of the paper.

an observer facing the light source, the substance is dextrorotatory. The opposite effect is observed with a levorotatory substance. The following equation defines the specific rotation of a substance:

$$\text{Specific rotation} = [\alpha]_{\text{D}}^{\text{T}} = \frac{\alpha \times 100}{l \times c},$$

where: α = the observed angle of rotation (in degrees); l = length of the polarimeter tube in decimeters (dm); and c = concentration of the optically active substance in grams per 100 ml. The specific rotation of a pure substance of known concentration is measured with sodium light ($\lambda = 589.3$ nm; the subscript D stands for the sodium D line) at a specified temperature (superscript T), normally 20 °C. Specific rotation depends (to a lesser extent) on the concentration, which is usually reported in brackets. In addition, the polarimeter is used for the quantification of optically active substances by means of the equation given above. It is necessary, of course, to know the specific rotation of the substance being analyzed. All sugars are optically active, and therefore, their concentration can be conveniently measured by means of the polarimeter. However, it is not possible to measure mixed sugars by this technique.

Monosaccharides

Aldoses and Ketoses (D- and L-configuration)

D-Aldoses are sugars containing an aldehyde group and, for classification purposes, may be considered as derived from D-glyceraldehyde (a triose with three carbon atoms). As shown in Figure 3, sequential addition of carbon atoms (by the Kiliani–Fischer cyanohydrin reaction) leads to (aldo)tetroses, (aldo)-pentoses and (aldo)hexoses, with four, five, and six carbon atoms, respectively. Similarly, Figure 4 gives the formulae of D-ketoses derived from dihydroxyacetone. Simple sugars belong either to the D-series or the L-series. This is clear from the Fischer projection formulae, whereby the hydroxyl group attached to the asymmetric carbon atom furthest from the aldehyde (or ketone) group is projected to the right (shown in bold characters in the figures) or to the left.

The Fischer projection formulae (Figure 3 and 4) are widely used to represent the basic structure of sugars because of the convenience they offer. The aldehydic carbon atom (reducing end) is numbered 1, and the remaining carbon atoms are consecutively numbered (vertically) towards the nonreducing end of the molecule. Bonds in the vertical direction are behind the plane of the paper, whereas horizontal bonds are directed out of the plane of the paper. The Fischer formulae give an exact picture of the stereochemical arrangement of atoms in a carbohydrate molecule, as projected on the plane of the paper. The projected molecular structures may be rotated 180° on the plane of the paper but should never be rotated around their longitudinal axis by lifting them 'off the plane of the paper.' It should be stressed that D- and L-forms have nothing to do with the direction of rotation of the plane of polarized light, which is always denoted by (+) or (−) signs. Indeed, as seen in Figures 3 and 4, not all D-sugars are dextrorotatory. The fact that R-glyceraldehyde (D-glyceraldehyde) is dextrorotatory is coincidental. A clear picture of the architecture of sugars can be gained by the use of atomic models (preferably the ball-and-stick type). Computer-generated models are also useful, especially programs that allow interactive rotation of the molecules for three-dimensional viewing.

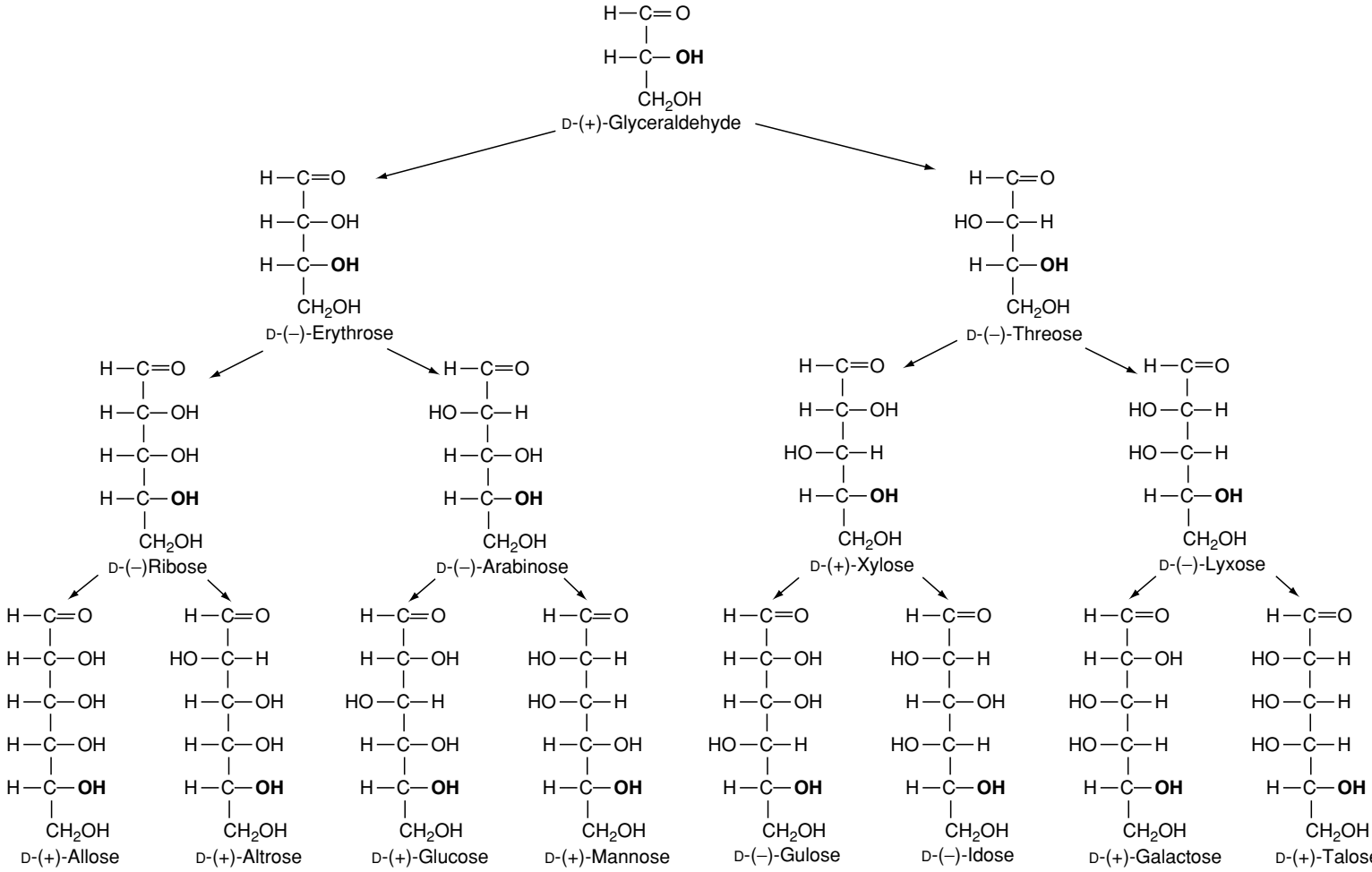

Figure 3 D-series of aldoses derived from D-glyceraldehyde (a triose). Sugars containing four, five and six carbon atoms are known as aldotetroses, aldopentoses, and aldohexoses, respectively. These Fischer projection formulae have the penultimate OH group (bold characters) attached to the highest numbered chiral center, on the right-hand side of the projection formula.

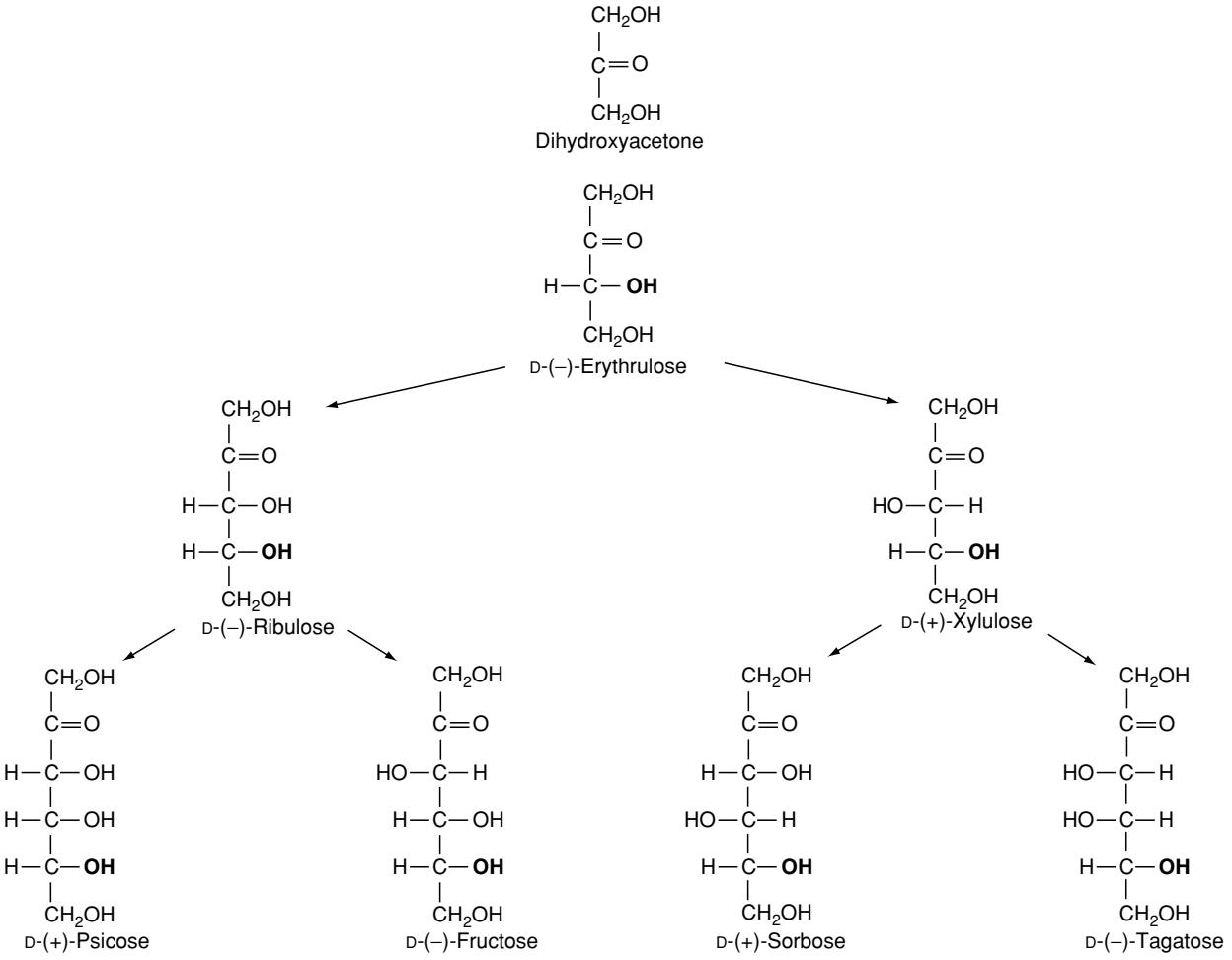

Figure 4 The D-series of ketoses derived from dihydroxyacetone. Sugars containing four, five, and six carbon atoms are known as ketotetroses, ketopentoses, and ketohexoses, respectively. These Fischer projection formulae have the penultimate OH group (bold characters) attached to the highest numbered chiral center, on the right-hand side of the projection formula.

There is confusion in the old literature with regard to the use of the letters D- and L- and also d- and l-. Both the upper-case and lower-case letters were used in the past to describe the configuration as well as the direction of optical rotation. The latter has also been denoted less ambiguously as (*dextro*-) and (*levo*-). The conventional notation currently in use is D- and L- for configuration and (+) and (−) for optical rotation.

Isomerism of Monosaccharides

Isomers are substances that have identical molecular formulas but differ in the way in which the atoms are bonded to each other. Stereoisomers differ only in the way the atoms are arranged in space as already explained. The number of stereoisomers for any monosaccharide is 2^n, where n is the number of chiral centers. Aldohexoses have four chiral centers and, therefore, $2^4 = 16$ isomers (Figure 3). There will be eight D-isomers and eight L-isomers. Therefore, the number of enantiomer pairs is 8 (2^{n-1}). Ketohexoses have only three chiral centers and thus eight isomers and four enantiomer pairs (Figure 4). It should be noted that the specific rotation of a pair of enantiomers has the same numerical value and differs only in sign. For example, the enantiomer of a D-(+)-glucose ($[\alpha]^{20} = +52.5°$) is L-(−)-glucose ($[\alpha]^{20} = -52.5°$). Biological systems can utilize either D- or L-molecules, but not both. The human body utilizes only D-glucose. By contrast, amino acids in the human diet belong to the L-series, and their D-form is not metabolized. However, L-arabinose, L-fucose, and L-rhamnose occur in plants, whereas their D-counterparts do not.

Epimers

Two sugars differing in configuration at a single asymmetric carbon atom are known as epimers. Glucose and mannose are C2 epimers, ribose and xylose

are C3 epimers, and gulose and galactose are also C3 epimers (Figure 3). D-Arabinose and L-xylose are C4 epimers, and so are D-glucose and D-galactose. In all cases (and there are several more), changing the position of H and OH on the same carbon atom (epimeric) of one sugar gives rise to the other sugar, i.e., its epimer. Epimers are also diastereoisomers, i.e., they have a similar structure but are not enantiomers (chiral or mirror images).

Mutarotation and Cyclic Hemiacetals

Glucose crystallized from methanol has a melting point of 147 °C. When dissolved in water, it has an initial specific rotation of +113°, which falls after several hours to +52.5°. Glucose crystallized from water at a high temperature (> 50 °C) has a melting point of 150 °C. Its initial specific rotation is +19° and rises gradually to the value of +52.5°. The change in optical rotation on standing is called mutarotation. This is explained by the fact that glucose gives rise to a cyclic 'internal' hemiacetal with the formation of a bond between carbon atom 5 (carrying an OH group) and carbon atom 1 (the aldehyde carbonyl group), as shown in Figure 5.

During ring closure, a new chiral center is formed (carbon atom 1), and the hydroxyl group will assume either the α- or the β-configuration (right or left respectively of the projection formula). The new chiral center is known as the anomeric carbon, and the resulting two structures as the α- and β-anomers. Anomers are not enantiomers and have different specific rotations. The planar ring structures, termed Haworth projections, are perpendicular to the plane of the paper with the thick lines directed towards the observer. The cyclic structures are then described as pyranoses (and their derivatives as pyranosides) and furanoses (or furanosides) by reference to the heterocyclic compounds pyran and furan. In aqueous solution, a dynamic equilibrium is established between the α- and β-glucopyranose structures. The equilibrium mixture of glucose at room temperature consists of about 36% of the α-anomer and 64% of the β-anomer. There is only a negligible amount of the open chain form. There is also a very small proportion (< 0.5%) of the α- and β-anomers of glucofuranose rings. In general, equilibrated solutions of most sugars contain different proportions of the two furanose and the two pyranose structures.

Fructose undergoes a similar rearrangement. Crystalline β-D-fructopyranose has an initial specific rotation of −133.5° and undergoes rapid mutarotation to − 92°. The equilibrated solution contains about 20% of the fructofuranose form. The four ring forms of fructose are shown in Figure 6.

Ring Conformations

So far, we have dealt with the spatial arrangement of atoms in various forms of monosaccharides. This is known as the configuration of molecules and involves the breaking of chemical bonds when converting, for example, from the D- to the L-form or from the α- to the β-anomeric form. Although the Haworth ring structure of sugars gives an exact view of the spatial arrangement of the hydroxyl groups of a particular sugar, it is misleading, because there is now ample

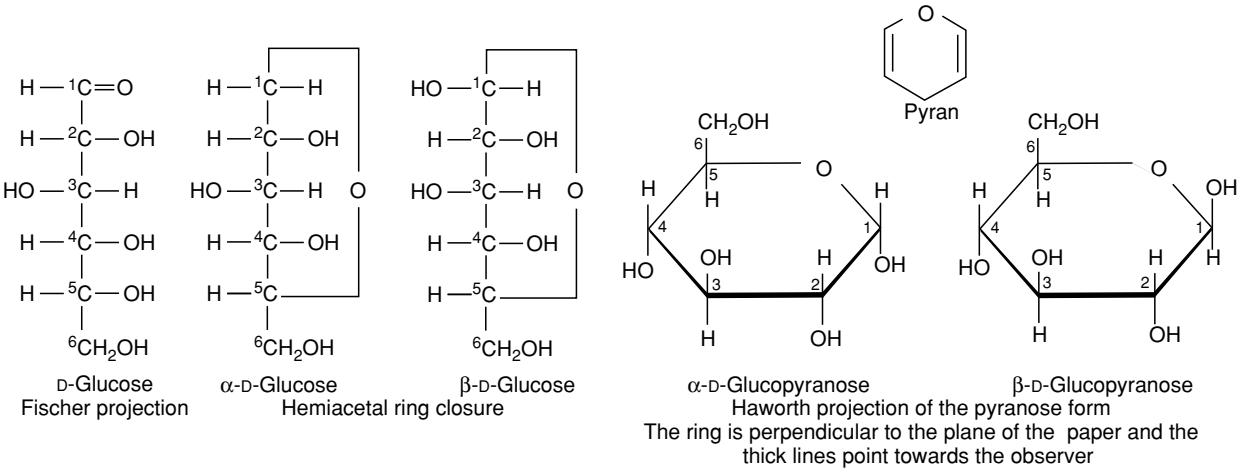

Figure 5 Cyclic hemiacetal formation between carbon 5 and carbon 1 of glucose. The anomeric carbon 1 is chiral, and two positions are possible for the OH group, which can be below (α-form) or above (β-form) the pyranose ring, or right and left, respectively, on the projection formula. The two forms are in equilibrium in aqueous solution. There is a small proportion (< 0.5%) of each of the two furanose forms and an almost negligible amount (0.003%) of the open-chain form.

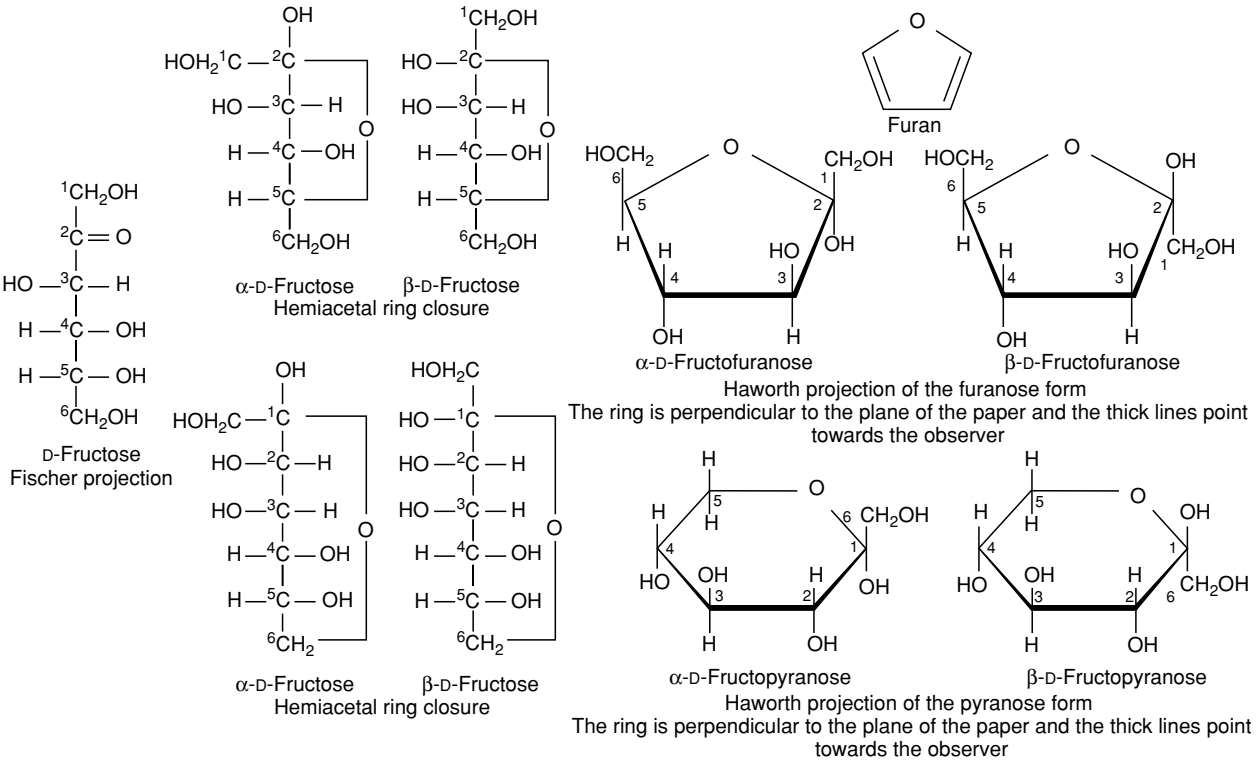

Figure 6 Cyclic hemiacetal (hemiketal) formation between carbon 5 and carbon 2 of fructose leads to the furanose structure. Hemiacetal formation between carbons 6 and 1 leads to the pyranose structure. The anomeric carbon atoms 2 and 1 are chiral, and two configurations are possible for the OH group. There are about 76% of the β-pyranose form, 20% of the β-furanose form, about 4% of the α-furanose form, and a negligible amount of the open-chain form.

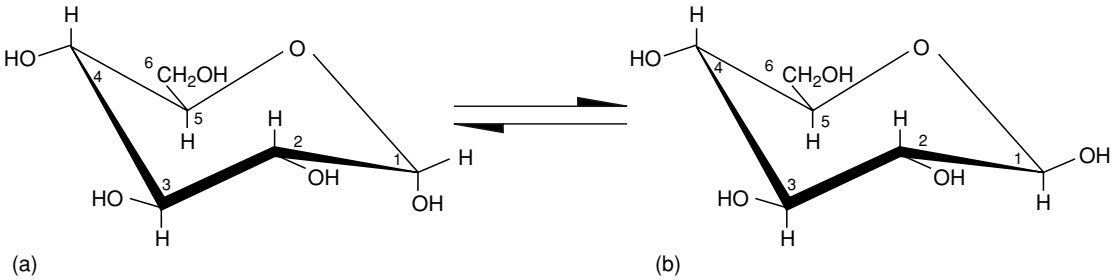

Figure 7 Cyclic hemiacetal form of glucose anomers. Chair [4C1] conformation of (a) α-D-Glucopyranose and (b) β-D-Glucopyranose. Note that all OH groups are equatorial, with the exception of OH on the anomeric carbon atom 1 in (a). In aqueous solution anomer (b) is more abundant (64%) than (a) (36%), because the all-equatorial conformation is thermodynamically favoured. The two forms are in equilibrium; there is a negligible fraction of the acyclic (open chain) form in solution.

evidence that the ring is not planar. The bonds in a planar ring are under considerable strain, which could be relieved if the ring adopted a nonplanar conformation to comply with the tetrahedral direction of valences. A change in the conformation of the ring structure of a sugar does not involve the breaking of bonds. It affects only bond angles and leads to an energetically favorable arrangement of atoms. Theoretical and experimental investigations have thrown considerable light on the conformation of the furanose ring in particular, both in the solid state and in solution. Complex formation, nuclear magnetic resonance spectroscopy and X-ray crystallography are the most useful techniques. Various theoretical schemes and energy minimization studies have also been useful in elucidating the conformation of carbohydrates. The preferred conformation for glucose is the so-called 'chair form' (Figure 7).

Glycosides

When an aldehyde or a ketone reacts with an alcohol in the presence of acid, a hemiacetal is formed first, and then an acetal with the addition of a second molecule of alcohol. The reaction is shown in Figure 8.

We have already seen that the cyclic forms of monosaccharides are hemiacetals. The hydroxyl group attached to the anomeric carbon atom can react with an alcohol (or the OH of another sugar). The resulting product is known as a 'glycoside' (from Greek *glykys* = sweet; gluco- is used specifically for glucose). For example, glucose (α-pyranose form) reacts with ethanol and gives rise to ethyl α-D-glucopyranoside, where the ethyl part of the molecule is known as the aglycone (see Figure 9).

There are many naturally occurring glycosides. Salicin is an example where the aglycone is *o*-(hydroxymethyl) phenol. Amygralin is another glycoside, which occurs in the kernels of bitter almonds, peaches, and apricots. Its composition is [(6-*O*-β-D-glucopyranosyl-β-D-glucopyranosyl)oxy]benzeneacetonitrile. It is hydrolyzed by β-glucosidase (the enzyme emulsin). Glycosides linked in the α-position are hydrolyzed by α-glucosidase, of which yeast is a principal source. Anthocyanins, the coloring matter of flowers and fruits, occur in nature as glycosides, which are hydrolyzed by acid to yield anthocyanidins and a number of sugars. Of course, oligo- and polysaccharides are all glycosides where the aglycone is any other sugar.

Representative Monosaccharides

Both glucose (*See* **Glucose**: Properties and Analysis) and fructose (*See* **Fructose**) occur widely in plants, particularly in fruits. They are also found in honey (*See* **Honey**). Glucose is an important item of commerce as crystalline dextrose monohydrate ($C_6H_{12}O_6.H_2O$), or as a component of glucose syrup and high fructose syrup (*See* **Syrups**). Other monosaccharides occur as components of polysaccharides rather than as free sugars (see below). The pentoses D-ribose and D-2-deoxyribose are components of nucleotides.

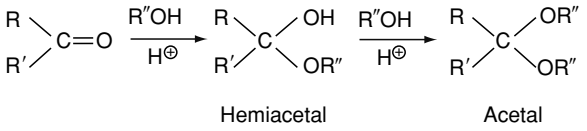

Figure 8 Addition of an alcohol to an aldehyde gives rise to a hemiacetal. Addition of a second molecule of alcohol gives an acetal (with the elimination of water).

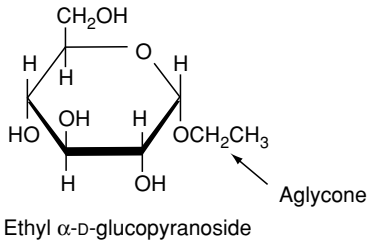

Figure 9 Formation of a 'glycoside' by the addition of ethanol and elimination of water.

Disaccharides

Nomenclature

Disaccharides can be reducing (having a free carbonyl group) or nonreducing. In the latter case, the two component monosaccharides are linked at their respective anomeric centers, and therefore, the carbonyl group is not available for reaction. Disaccharides are named as glycosides where the aglycone is another monosaccharide. Reducing disaccharides are named as substituted monosaccharides (Figure 10).

Sucrose (saccharose) is by far the most important disaccharide (*See* **Sucrose**: Properties and Determination; Dietary Importance; **Sugar**: Sugarcane; Sugarbeet; Palms and Maples; Refining of Sugarbeet and Sugarcane). Lactose (*See* **Lactose**) occurs in the milk of mammals but very rarely in the plant kingdom. Trehalose (α-D-glucopyranosyl α-D-glucopyranoside) (nonreducing) occurs in mushrooms and other fungi. Maltose is formed during the mashing of malt (*See* **Malt**: Malt Types and Products; Chemistry of Malting) in brewing and serves as a substrate for yeast in alcoholic fermentation. It is also a component of high-maltose syrup. Cellobiose (β-D-glucopyranosyl-(1→4)-D-glucose) (reducing) is formed by the enzymatic hydrolysis of cellulose.

Oligosaccharides

The so-called raffinose family of oligosaccharides comprises raffinose (trisaccharide), stachyose (tetrasaccharide), and verbascose (pentasaccharide), all of which occur in the seeds of legumes, as well as in different parts of plants. Verbascose (Figure 11) has three molecules of α-D-galactose attached to sucrose; stachyose has two, and raffinose one. They are all nonreducing. Invertase releases the fructose moiety and gives rise to reducing saccharides. Treatment of raffinose with invertase (β-fructosidase) gives fructose and melibiose, whereas α-galactosidase (from green coffee beans) gives sucrose and galactose. Lactase (β-galactosidase) has no effect.

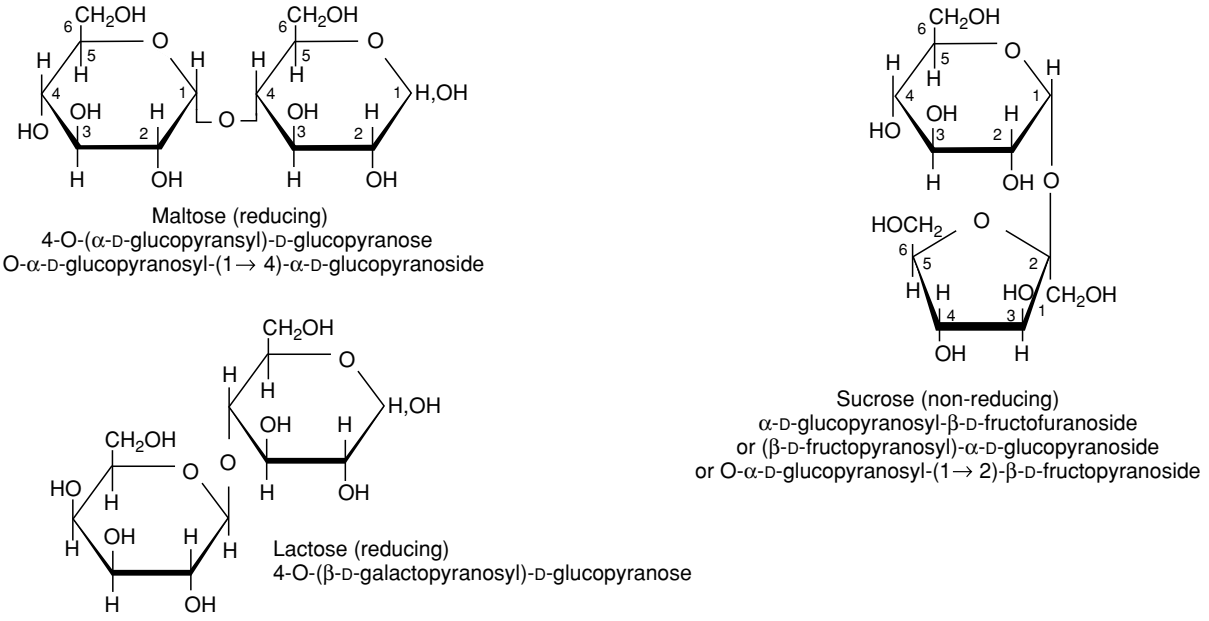

Figure 10 Formulae and nomenclature of three common disaccharides. Sucrose does not have a free carbonyl group. The OH group on the anomeric carbon atom of maltose and lactose can acquire either the α- or the β-configuration. Both sugars mutarotate when dissolved in water, and equilibrium is established after several hours. A few drops of ammonia accelerate the rate of mutarotation, and a constant specific rotation is rapidly attained.

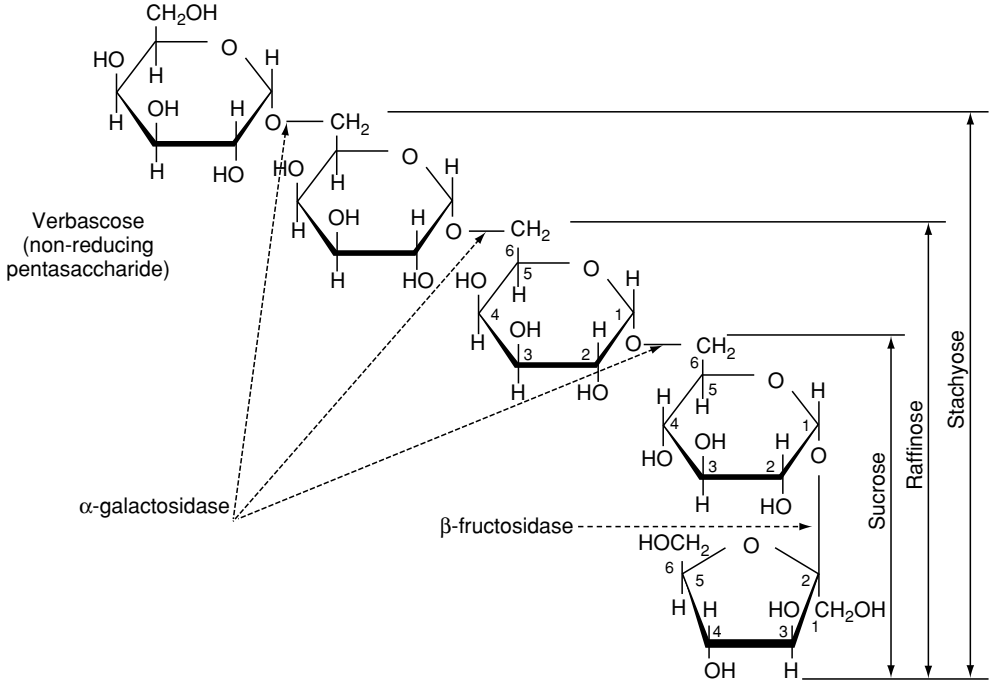

Figure 11 *O*-α-D-Galactopyranosyl-(1→6)-[*O*-α-D-galactopyranosyl-(1→6)]$_2$-*O*-α-D-glucopyranosyl-(1→2) β-D-frutofuranoside. Arrows indicate the points of hydrolysis by enzymes.

Polysaccharides

Nomenclature

Polysaccharides are also known as 'glycans.' Depending on the monomeric sugar, they are described as glucans, fructans, galactans, mannans, xylans, etc., where the first four letters refer to glucose, fructose, galactose, mannose, and xylose, respectively. Polysaccharides consisting of a single type of monomer are known as 'homopolysaccharides.' When two or more types of sugar monomers are involved, they are known as 'heteropolysaccharides.' The properties of polysaccharides are discussed in **Carbohydrates**: Interactions with Other Food Components.

Representative Polysaccharides

Cellulose (*See* **Cereals**: Contribution to the Diet) is the most abundant polysaccharide in nature. It consists of β-linked glucose units in the form of long chains, which associate parallel to each other by hydrogen bonding to form fibers, which are insoluble in water. Cellulose is a component of plant cell walls and occurs also in leaves and stems, often associated with lignin and xylans. The textural characteristics of leafy vegetables depend, in part, on cellulose. The related β-glucans (water-soluble) of malted barley cause filtration problems in brewing.

Starch (*See* **Starch**: Structure, Properties, and Determination; Sources and Processing; Functional Properties; Modified Starches; Resistant Starch) is the energy reserve polysaccharide of plants. It occurs in the form of microscopic granules (about 5–50 nm in diameter) in cereal grains, in roots, tubers, stems, and, to a lesser extent, in some fruit and leaves. The granules are insoluble in water at ambient temperature but absorb water and swell when heated to their gelatinization temperature (*c*. 50–70 °C depending on the origin). Starch consists of two polymers: essentially linear amylose with an average degree of polymerization (DP) of *c*. 1000, and branched amylopectin as shown in Figure 12. Amylopectin consists of α-(1 → 4)-linked glucose units (95%) and α-(1 → 6)-linked branched points (5%). The DP is 10^4–10^5. Glycogen (*See* **Glycogen**) is the reserve carbohydrate of the animal kingdom. It resembles amylopectin in structure but has more extensive multiple branching.

Dextran (*See* **Dextran**) is a polymer of dextrose synthesized from sucrose by *Leuconostoc dextranicum* or *L. mesenteroides*. It consists mainly of α-D-(1 → 6) linked glucose units to which are attached α-(1 → 3) and some α-(1 → 2) glucan branches. Low-molecular-weight products are water-soluble, whereas high-molecular-weight products are insoluble. Dextrins (*See* **Dextrins**) are products of partial acid hydrolysis of starch (very low acid at

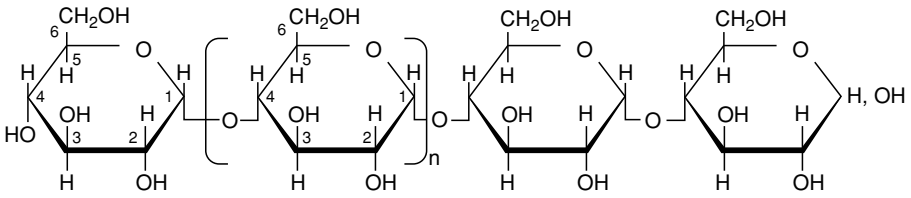

Amylose: α-(1→ 4)-glucan: average n = ca. 1000. The linear molecule may carry a few occasional moderately long chains linked α-(1→ 60).

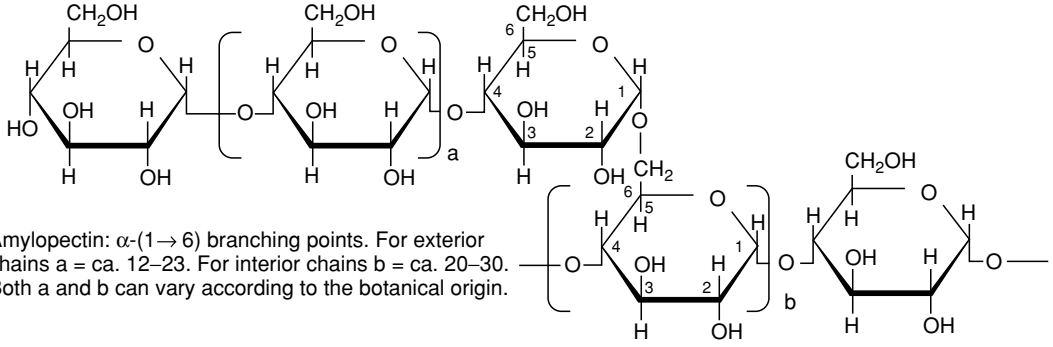

Amylopectin: α-(1→ 6) branching points. For exterior chains a = ca. 12–23. For interior chains b = ca. 20–30. Both a and b can vary according to the botanical origin.

Figure 12 Basic structure of amylose and amylopectin.

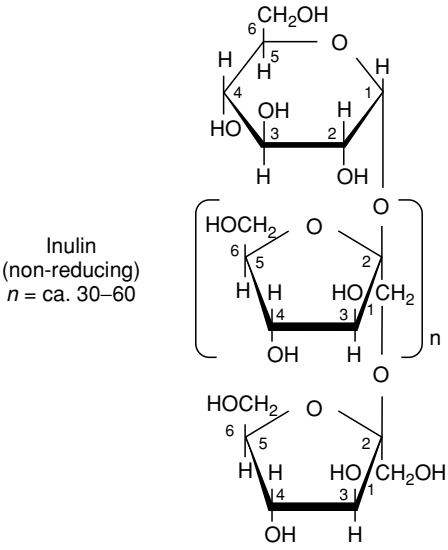

Figure 13 Structure of inulin. This polysaccharide is soluble in water and is not hydrolyzed by β-fructosidase.

high temperature). They are water-soluble and available in several grades. They are used mainly as adhesives, although food-grade dextrins are also known. Nonstarch, noncellulosic polysaccharides (formerly known as hemicelluloses) occur in cereals, roots, fruits, leaves, and stems of plants. They are the main components of dietary fiber (*See* **Dietary Fiber**: Properties and Sources; Determination; Physiological Effects; Effects of Fiber on Absorption; Bran and may consist of pentosans (composite polymers of pentoses), β-glucans, and pectic substances. Some are insoluble, whereas some are soluble in water. Rhamnogalacturonans, arabinogalactans, glucuronoxylans, and xyloglucans occur mainly in fruits and vegetables, whereas glucuronoarabinoxylans occur in cereals. Inulin (Figure 13) is a polymer of fructose, having one molecule of glucose at the start of the chain. It is soluble in water and is not hydrolyzed by invertase. Polysaccharides from various sources used as food additives are discussed in **Carbohydrates**: Interactions with Other Food Components.

Glycoproteins and Glycolipids

Some proteins may contain small amounts of carbohydrates. Casein contains 1% galactose, 1.2% galactosamine, and 2.4% acetyl neuramic acid. Egg white contains considerable amounts of carbohydrate (2–30% in the various fractions), the main sugars being galactose, mannose, and glucosamine. Glycolipids are found in various membranes in both plants and animals. The carbohydrate content is relatively minor.

Physical Properties

Most sugars are soluble in water and form syrups at high concentrations. Crystallization (*See* **Crystallization**: Basic Principles) occurs in saturated solutions. They are mainly used for their sweetness. Sucrose and fructose at a concentration of > 65% (w/w) have a preservative effect because of high osmotic pressure (*See* **Water Activity**: Principles and Measurement; Effect on Food Stability) that prevents the proliferation of microbes (*See* **Carbohydrates**: Interactions with Other Food Components). They also impart desirable textural characteristics to a wide range of foods like cakes, biscuits, and confectionery products. Polysaccharides are mostly used in foods to impart viscosity and induce gel formation (*See* **Carbohydrates**: Interactions with Other Food Components). Table 1 lists some of the properties of sugars.

Chemical Properties

Sugars undergo a variety of reactions and give rise to many derivatives. These are not all of immediate interest to the food scientist. Relevant interactions are discussed in **Carbohydrates**: Interactions with Other Food Components.

Reduction

Monosaccharides are easily reduced to alditols (known also as 'polyols') by sodium borohydride ($NaBH_4$) in weak alkaline solution. Large-scale reduction is carried out by catalytic hydrogenation. Alditols are linear molecules, do not form rings, and do not exist as anomeric forms. These are important properties used advantageously in sugar analysis by gas chromatography, since each alditol gives rise to a single peak on a chromatogram. Tetritols, pentitols, and hexitols have 2, 3, and 4 chiral centers, respectively. In general, they have a low specific rotation, and some are optically inactive (*meso*-forms); the latter are also known as internally compensated molecules, because they have a plane of symmetry (Figure 14).

It is noteworthy that when D-glucitol is rotated 180°, it is identical to D-gulitol, which is also identical to L-glucitol, the latter being the enantiomer (mirror image) of D-glucitol. L-Glucitol does not occur in nature, but D-glucitol (sorbitol) is fairly common (plums, berries). On hydrogenation of fructose, a new chiral center is created, and thus two alditols are produced, namely D-glucitol (75%) and D-mannitol (25%) (epimers). Galactose gives rise to D-galactitol (dulcitol) and maltose to maltitol. Lactitol is obtained by the hydrogenation of lactose. The free carbonyl end of oligo- and

Table 1 Some properties of carbohydrates

Commonly occurring forms	*Molecular weight*	*Specific rotation [α]p*[a]	*Melting point (°C)*[a]	*Solubility (grams per milliliter of water*[a]*)*	*Sweetness*[b,c] *Sucrose = 100*	*Hydrolyzing enzymes*	*Occurrence*
Pentoses $C_5H_{10}O_5$							
L-Arabinose	150.13	+173°→+105.1°	157–160	1			As arabinan in plant cell walls
D-Xylose	150.13	+92°→+18.6°	153–154	1.25	67–70		As xylan in plant cell walls
Hexoses $C_6H_{12}O_6$							
α-D-Glucose (dextrose)	180.16	+112.2°→+52.7°	146	0.91	70		Free in fruits
α-D-Glucose monohydrate	198.18	+102.0°→+47.9°	83	1			
α-D-Galactose	180.16	+150.7°→+80.2°	167	2	63		As galactan in plant cell walls
β-D-Galactose	180.16	+52.8°→+80.2°	167	0.6			
α-D-Mannose	180.16	+29.3°→+14.2°	133	Soluble	60		As mannan in plant cell walls and nuts
β-D-Mannose	180.16	−17.0°→+14.2°	132[d]	2.5			
β-D-Fructose (levulose)	180.16	−132°→−92°	103–105[d]	4	114–150		As fructan and free in fruits
Deoxyaldohexoses $C_6H_{12}O_5$							
α-D-Fucose	164.16	+127°→76.0°	144	Soluble			Plant cell walls
α-L-Fucose	164.16	−124°→−75.6°	140	Soluble			
α-L-Rhamnose	164.16	−7.7°→+8.9°	82–92	Soluble	33		Plant cell walls
Uronic acids $C_6H_{10}O_7$							
α-D-Galacturonic acid	194.14	+98.0°→+50.9°	159	Soluble			As pectin in fruits and plants
β-D-Glucuronic acid	194.14	+11.7°→+36.3°	165	Soluble			Plants
Disaccharides $C_{12}H_{22}O_{11}$							
Sucrose[e]	342.30	+66.5°	160–186[d]	2	100	β-Fructosidase	Sugar cane, sugar beet, fruits
Maltose	342.30	+111.7°→+130.4°	102–103	Soluble	40–46	α-Glucosidase	Free in malt extract
Lactose monohydrate	360.32	+92.6°→+52.3°	201–202	0.2	39	β-Galactosidase	Free in milk
Lactulose	342.30	−51.4°	169	3.2	60		Low concentration in UHT milk
Trehalose dihydrate[e]	378.34	+178°	96.5–97.5	Soluble		Trehalase	Free in mushrooms
Trisaccharide							
Raffinose pentahydrate[e]	504.44 anh	+105.2°	80	0.14	22	α-Galactosidase	Leguminous seeds
Sugar alcohols							
Xylitol	152.15	Meso-form	93–94.5	2	90–102		Metabolic intermediate
D-Glucitol (sorbitol)	182.17	−2.0°	110–112	4.9	51–60		Free in some fruits and berries
Mannitol	182.17	Inactive	166–168	0.2	50–69		Free in plant exudates
Lactitol	344.32	+14.0°	146	1.4	36		Reduction of lactose
Polysaccharides $(C_6H_{10}O_5)_n$							
Amylose	$(162.14)_n$	+200° ($CaCl_2$ solution)		‘Soluble’ (retrogrades)	0	α- and β-amylase, amyloglucosidase	*c.* 30% of starch in cereal grains, tubers, roots, and stems
Amylopectin	$(162.14)_n$	+200° ($CaCl_2$ solution)		‘Soluble’ (retrogrades)	0	As above and isoamylase[f] and pullulanase[f]	*c.* 70% of starch (as above)
Cellulose	$(162.14)_n$			Insoluble		Cellulase (β-glucanase)	Plant cell walls, wood, etc.

[a]From Merck Index, 12th edn. (1996).
[b]From Birch GG and Parker KJ (eds) (1982) *Nutritive Sweeteners*. London: Applied Science.
[c]Approximate relative sweetness varies with temperature, concentration, and pH.
[d]Decomposes.
[e]Nonreducing.
[f]Debranching enzymes.

polysaccharides can be reduced to -OH (alditol), as in hydrogenated glucose syrup, which is claimed to be less harmful to teeth and does not participate in the Maillard reaction. Xylitol and sorbitol are also considered as noncariogenic. They all taste sweet. Sorbitol, mannitol, and xylitol are absorbed by passive diffusion in the digestive system and subsequently metabolized. However, large doses have a laxative effect.

Oxidation

The aldehyde group of monosaccharides is easily oxidized (Figure 15) and gives aldonic acids. This occurs when the familiar Fehling's reagent is used for the detection of reducing sugars, whereby blue cupric ions in solution are reduced to insoluble red cuprous oxide. Strong oxidizing agents (e.g., nitric acid) oxidize both ends of the alsose molecule and give rise to aldaric acid. For example, glucose is oxidized by bromine water to gluconic acid, which is isolated as the δ-lactone. Nitric acid gives rise to glucaric (saccharic) acid, which forms a dilactone. A third type of acid is obtained when the aldehyde group is protected prior to oxidation. The OH group of carbon atom 6 is oxidized to give uronic acid. Uronic acids (*See* **Uronic Acids**) frequently occur in natural polymers (*See* **Pectin**: Properties and Determination; Food Use), and are discussed **Carbohydrates**: Interactions with Other Food Components.

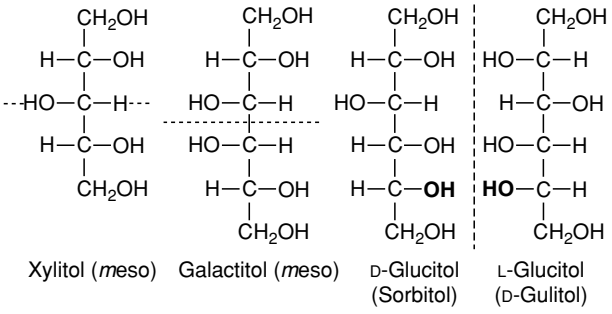

Figure 14 Some alditols and their stereochemical relationships.

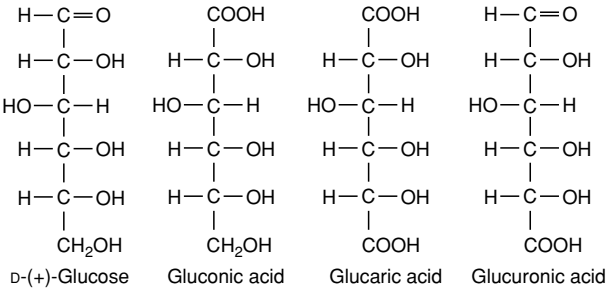

Figure 15 Glucose can be oxidized to give three different acids. The enzyme glucose oxidase gives rise to gluconic acid and H_2O_2. Cupric ions and ferricyanide ions also oxidize glucose to gluconic acid. Glucuronic acid is the monomeric component of pectin and occurs mostly as the methyl ester (methylated pectin).

Esterification

The hydroxyl groups of sugars are relatively easily esterified by anhydrides of organic acids. In sugar analysis, monosaccharides are firstly reduced to alditols and subsequently fully esterified with acetic anhydride; the resulting alditol acetates are then identified and quantified by gas chromatography. In biological systems, D-ribose esterified with phosphoric acid is the carbohydrate moiety of ribonucleotides and deoxyribose in deoxyribonucleotides.

Alkaline Rearrangement

In the presence of alkali, the sugars glucose, mannose, and fructose are interconvertible. This is known as the 'Lobry de Bruyn–Alberda van Ekenstein rearrangement.' The transformation is believed to occur by a 1,2-enolization reaction. However, prolonged contact with alkali leads to degradation of sugars.

Reaction with Amino Acids, Peptides, and Proteins

Reducing sugars readily interact with amino acids and give rise to Maillard reaction products, which lead to progressive browning and aroma formation. The color and aroma of dark beer, baked goods, toasted bread, and grilled foods are, at least in part, due to this reaction, which is initiated when the carbonyl group of a sugar reacts with an amino group (*See* **Browning**: Nonenzymatic). The initial products of the reaction are *N*-glycosylamines or *N*-fructosylamines, which give rise to intermediate products and final heterocyclization and polymerization.

See also: **Carbohydrates**: Interactions with Other Food Components **Cereals**: Contribution to the Diet; **Dietary Fiber**: Properties and Sources; Determination; Physiological Effects; Effects of Fiber on Absorption; Bran; **Fructose**; **Glucose**: Properties and Analysis; **Glycogen**; **Honey**; **Lactose**; **Malt**: Malt Types and Products; Chemistry of Malting; **Starch**: Structure, Properties, and Determination; Sources and Processing; Functional Properties; Modified Starches; Resistant Starch

Further Reading

Belitz H-D and Grosch W (1999) *Food Chemistry*, 2nd edn. Berlin: Springer.

Birch GG and Parker KJ (eds) (1982) *Nutritive Sweeteners*. London: Applied Science.

Collins PM (ed.) (1987) *Carbohydrates*. London: Chapman & Hall.
Collins PM and Ferrier RJ (1995) *Monosaccharides: Their Chemistry and their Roles in Natural Products*. Chichester, UK: Wiley.
Dey PM (ed.) (1990) *Methods in Plant Biochemistry Vol. 2 Carbohydrates*. London: Academic Press.
Ege S (1989) *Organic Chemistry*, 2nd edn. Lexington, KY: DC: Heath & Co.
Junk WR and Pancoast HM (1973) *Handbook of Sugars*. Westport, CT: AVI Publishing.
Lindhorst TK (1999) *Essentials of Carbohydrate Chemistry and Biochemistry*. Chichester, UK: Wiley.
Rao VSR, Qasba PK, Balaji PV and Chandrasekaran R (1998) *Conformation of Carbohydrates*. Amsterdam: Harwood Academic.
Stick RV (2001) *Carbohydrates: The Sweet Molecules of Life*. San Diego, CA: Academic Press.
Vollhardt KPC and Schore NE (1999) *Organic Chemistry, Structure and Function*, 3rd edn. New York: W. H. Freeman and Company.

Interactions with Other Food Components

R F Tester and J Karkalas, Glasgow Caledonian University, Glasgow, UK

Carbohydrate Interactions

Carbohydrates include simple sugars, oligosaccharides, and polysaccharides (*See* **Carbohydrates**: Classification and Properties). They can interact with themselves, or with noncarbohydrate substances. These interactions may be (1) physical (e.g., altered rheological properties, lowering of water activity), (2) ionic, (3) stable associations involving hydrogen bonds and van der Walls interactions in the form of helices, and (4) chemically bonded structures (covalent bonds) resulting from reactions of carbohydrates with other molecules. In addition, carbohydrates can be decomposed and oxidized at high temperatures. These interactions provide structure (e.g., gel formation), limit digestibility (e.g., with amylases), and may restrict bioavailability of trace elements. Sometimes, these interactions are deliberately created during food processing (e.g., generation of color or aroma), whereas at other times, they may be undesirable (e.g., browning of dehydrated foods, loss of nutrients), and steps are taken to prevent their occurrence.

Physical Associations and Interactions

Sugars

Sugars are generally very soluble in water. The saturation concentration of sucrose in water at 20 °C is 66.6% (by weight) or 1.994 kg in 1 l of water. Fructose is very soluble (> 70%), but glucose has limited solubility (about 48% by weight) at 20 °C. The viscosity of these sugars increases with increasing concentration, e.g., in cordials, golden syrup (a mixture of sucrose and invert sugar), and honey. The high solubility of sugars is due to the affinity of hydroxyl groups for water. Crystallization of sugars occurs when the solubility limit is exceeded. Sugar crystals, in common with all crystalline substances, have a highly ordered structure built from unit cells of the constituent molecules.

Fondant is a suspension of a very large number of minute sucrose crystals (< 20 μm particle diameter) in a saturated solution of sucrose. The creamy texture of this product is due to the very fine crystals of sucrose that are not detected on the palate. By contrast, crystallization of sucrose is prevented by the addition of glucose syrup or by partial inversion of the sucrose as in candies (hard-boiled sweets and fruit drops). (*See* **Sweets and Candies**: Sugar Confectionery.)

Lactose (milk sugar) has rather limited solubility (21.6% by weight at 25 °C) and crystallizes readily as the monohydrate ($C_{24}H_{22}O_{11}.H_2O$). This sugar forms gritty particles in sweetened condensed milk, because its solubility limit is exceeded during evaporation. By the judicious addition of a small quantity of exceedingly fine lactose crystals to the milk and sucrose mixture during evaporation, lactose is induced to crystallize as very fine crystals (< 30 μm diameter) that are not perceptible in the mouth.

Partially purified sucrose crystals (e.g., soft sugar, brown sugar, Demerara sugar, and Muscovado sugar) are coated with a thin film of molasses, which imparts a brown color and a characteristic flavor. These products are often considered nutritionally superior to white sugar. This is an erroneous assumption, because the protein and mineral contents are usually less than 1%. The 'unrefined' sugars contain only traces of vitamins. The cohesive nature of these products is due to a small quantity of inverted sugar and accompanying moisture (*c.* 1–5%), both of which originate from the molasses.

Apart from their widespread use as sweeteners, simple sugars have a preservative effect in concentrated solution. A 65% (by weight) solution of sucrose lowers the water activity (a_w) (*See* **Water Activity**: Principles and Measurement; Effect on

Food Stability) to about 0.8, and the product resists microbial growth because of the prevailing high osmotic pressure. Jams and preserves owe their stability at ambient temperature to this effect. The long shelf-life of intermediate-moisture foods (e.g., fruitcake) depends on the same principle. Glucose is not useful in this respect, because of its relatively low solubility. Conversely, invert sugar causes a considerable reduction in water activity as in honey, for example ($a_w = 0.75$). 'Set honey' is saturated with respect to glucose, which remains suspended as very small crystals in a viscous, almost solid, medium. Unlike sugars, polysaccharides do not cause any significant lowering of water activity, because they have a large molecular weight and because they are used at a very low concentration (normally < 5%).

Apart from sugar–sugar and sugar–water interactions as discussed above, sugars may interact chemically and form more extensive bonding (covalent bonds) with themselves and other molecules (below). However, certain sugars (like glucose) are also valuable in food systems because of their inherent reducing properties. These interactions may stabilize ionic groups in solutions and facilitate their absorption when ingested (e.g., maintenance of the ferrous form of iron and conversion from the ferric form).

Polysaccharides

Polysaccharides are macromolecules, many of which can adopt organized structures in foods. Most are polyhydric alcohols, acids and/or esters, which, depending on the monomeric sugar composition, are likely to interact with themselves, other polysaccharides, or other molecules. Polysaccharides may be insoluble in water, dissolve, and form true solutions, colloidal dispersions, and/or gels. In mixed polysaccharide systems, the polymers may intersperse and form a number of interactive states. These interactions may be described as:

1. *Additive*, where the measured property is the sum of the property of the individual polymers.
2. *Complementary*, where polymer one provides a characteristic that polymer two (or more) lacks.
3. *Synergistic*, where the effect of the combined polymer population is greater than the additive effect.
4. *Antagonistic* with respect to a ternary system, where there is phase separation as a consequence of solvent incompatibility. The phase separation process (coacervation) results in one phase being rich in one of the two solutes and the other phase being rich in the other solute. This process is utilized in, for example, encapsulation of flavors.

When polysaccharides are added to foods to stabilize structural form, they are described as 'stabilizers' (*See* **Stabilizers**: Types and Function; Applications) or 'gums' (*See* **Gums**: Properties of Individual Gums). Stabilization reflects thickening rather than emulsification in the true sense, although stabilizers are used in foods to prevent separation of emulsions and sedimentation of suspended solids. In many products (e.g., instant soups), they impart a desirable 'mouth feel.' Some act as gelling agents, which again impart desirable physical properties. With the exception of gelatine (a protein), all stabilizers are polysaccharides whose main function is to increase the viscosity of aqueous systems. Most of the polysaccharides listed in Table 1 are used as stabilizers in a wide range of manufactured foods. Nonmodified (native) starch is exceptional, because it is primarily a very important nutrient (together with sugars, it provides about 50% of food energy in the UK), which also happens to be a good stabilizer and texture modifier. According to Stokes' law, the velocity of sedimentation of particles is inversely proportional to the viscosity of the medium in which the particles are suspended.

Polysaccharides consist of chains of varying length, which generally occur as random coils in dilute aqueous systems. The chains can be branched or linear. They are surrounded by a cloud of water molecules, and this gives rise to internal friction and hence viscosity. They behave as imaginary spheres whose 'effective volume' or 'hydrodynamic volume' is related to the 'radius of gyration' of the polysaccharide chain. Branched polysaccharides have a lower viscosity than their linear counterparts of the same molecular weight because their hydrodynamic volume is smaller. Polysaccharides give rise to a substantial increase in viscosity, even at a very low concentration (< 1%). As the concentration increases, the random coils begin to overlap and a considerable increase in viscosity is observed as the chains interact with each other. Many polysaccharides, but not all, give rise to gels at a higher concentration (1–4%). The gel-forming properties of polysaccharides may be due to weak Van der Waals interactions, ionic interactions, hydrogen bonding, or helix formation. As a consequence, junction zones are established, which stabilize the three-dimensional structure that leads to self-supporting gels.

With the exception of starch, polysaccharides are not digested by the endogenous enzymes of the human digestive tract and are considered as dietary fiber. Nutritionally speaking, however, there may be a drawback, in so far as charged polysaccharides may bind minerals (cations) and make them less available to the human body.

Table 1 Polysaccharides used in foods

Polysaccharide	*Origin*	*Monomeric sugars*	*E-number*
Starch	Cereals, tubers, roots	α-(1→ 4)-D-Glucose [amylose] and α-(1 → 6)-D-glucose branches [amylopectin]	
Cellulose	Cell walls, wood, cotton	β-(1 → 4)-D-Glucose	E 460
Tree exudates			
Gum acacia (Arabic)	*Acacia senegal*	β-(1 → 3)-D-Galactose backbone with L-rhamnose, D-glucuronic acid, D-galactose, L-arabinose side-chains	E 414
Gum karaya (sterculia)	*Sterculia urens*	D-galacturonic acid, D-galactose, L-rhamnose, D-glucuronic acid in three-dimensional arrangement	E 416
Gum tragacanth	*Astragalus gummifer*	α-(1 → 4)-D-Galacturonic acid backbone with D-galactose -β-(1 → 2)-D-xylose, L-fucose-β-(1 → 2)-D-xylose and D-xylose-β-(1 → 3) side-chains	E 413
Seed gums			
Locust bean (carob) gum	*Ceratonia siliqua*	β-(1 → 4)-D-Mannose with α-(1 → 6)-D-galactose side-chains (Man:Gal = 4:1)	E 410
Guar gum	*Cyamopsis tetragonoloba*	β-(1 → 4)-D-Mannose with α-(1 → 6)-D-galactose side-chains (Man:Gal = 2:1)	E 412
Algal products			
Agar	*Gelidium* spp.	β-D-Galactose and 3,6-anhydro-D-galactose. Low esterification with sulfate groups	E 406
Alginic acid	*Ascophyllum nodosum, Laminaria hyperborea, Macrocystis pyrifera*	β-(1 → 4)-Mannuronic acid chains and α-(1 → 4)-guluronic acid chains, with alternating mannuronic–guluronic acid chains	E 400
Carrageenan	*Gigartina stallata*, *Chondrus crispus*, etc.	Galactose-sulfate, galactose disulfate and anhydogalactose chains; occurs in different forms: τ-, κ-, λ-, μ- and ν-carrageenan	E 407
Fruit products			
Pectin	Apple pomace, citrus peal	α-(1 → 4)-Galacturonic acid, largely esterified with methanol; some rhamnose and xylose may be present	E 440
Microbial products			
Curdlan	*Alcaligenes faecalis* var. *myxogenes*	β-(1 → 3)-D-Glucose	
Gellan	*Pseudomonas elodea*	β-(1 → 4)-D-Glucose-β-(1 → 4)-glucuronic acid-β-(1 → 4)-glucose-α-rhamnose-tetrameric repeat unit	E 418
Xanthan	*Xanthomonas campestris*	α-(1 → 4)-glucose backbone with mannose–glucuronic acid–mannose side-chains with terminal carboxyl groups	E 415
Root polysaccharides			
Inulin	*Helianthus tuberosus* (Jerusalem artichoke) and *Cichorium intybus* (chicory root)	β-(1 → 2)-D-Fructose	
Konjac	*Amorphophallus konjac*	β-(1 → 4)-D-Mannose, β-(1 → 4)-D-glucose	E 425

Homopolysaccharides

Homopolysaccharides (homoglycans) consist of a single type of monomer. Cellulose and starch are the best-known examples. However, they have very different properties. Cellulose (*See* **Cereals: Contribution to the Diet**) consists of β-(1 → 4)-linked glucose units arranged in a ribbon-type conformation in a zigzag pattern. Parallel chains fit closely to each other and associate with multiple hydrogen bonds to give rise to long fibers, which are totally insoluble in water and relatively inert. Although subject to some swelling in water, cellulose is completely unaffected by boiling in water. Natural cellulose (cellulose I) has considerable crystallinity (60–80%), as revealed by X-ray diffraction. Starch consists of two components, *viz*. amylose (essentially linear) and amylopectin (branched).

The α-(1 → 4)-linked glucose molecules in amylose allow considerable conformational freedom. In freshly prepared dilute solution, amylose is in the form of a random coil. On standing at ambient temperature, it 'retrogrades' (see below), i.e., the random coils form double helical associations, which are

insoluble and sediment as a white precipitate. At a high concentration, free settling is prevented, and gels are formed as the randomly oriented chains approach each other at several junction zones. Double helices are formed at the junctions, which do not extend the full length of the molecule. The net result is a three-dimensional structure, whose rigidity depends on the concentration of amylose and hence the number of junction zones per unit volume. Amylose gels are translucent to almost opaque, owing to light scattering. Double helices of retrograded amylose are dissociated at a temperature of about 160 °C.

Amylopectin consists of α-(1 → 4)-linked glucose units (95%) and α-(1 → 6)-linked branched points (5%) and is fairly soluble in water. The outer chains of amylopectin, consisting of about 15–25 glucose units, because of their close proximity readily form short double helices during starch biosynthesis and within food systems upon storage. Amylopectin is fairly stable in solution and forms soft gels at a high concentration. Waxy maize starch consists almost entirely of amylopectin and gives clear or slightly translucent weak gels.

When native starches (*See* **Starch**: Structure, Properties, and Determination; Sources and Processing; Functional Properties) are heated in excess water, the double helices progressively dissociate, water is imbibed, and the granules swell. This process of gelatinization is characteristic of the botanical origin. Typically, it occurs at around 65 °C and is irreversible, forming a solution, a colloid, or a gel, depending on the origin of starch and its concentration. Once gelatinized, double helices may reform in food systems (below).

Fructans are polymers of fructose linked by β-(1 → 2) bonds. The best-known example is inulin with 3–60 fructose units per chain. This forms clear viscous solutions in water. Inulin is not digested by the endogenous enzymes of the human digestive system. It is considered as dietary fiber and is used in some functional foods (*See* **Functional Foods**).

Curdlan is a microbial homopolysaccharide that is insoluble in water at ambient temperature. It forms a gel > 54 °C with an almost constant strength between 60–80 °C and increases up to, and beyond, 100 °C. It is an approved food additive in the USA.

Heteropolysaccharides

Heteropolysaccharides (heteroglycans) consist of more than one type of monomeric sugar. Indeed, the majority of polysaccharides used as food additives are heteropolysaccharides (Table 1).

Single Polymer Systems

Heteropolysaccharides can associate in a number of ways in food and nonfood systems forming solutions, colloids, gels, and sometimes associated helices. The physical properties are critical in terms of texturization in food systems and must be optimized for a given application. These structures are dependent on the nature of the polysaccharide and its concentration; agar, for example, can form relatively strong gels at the level of 1%, whereas starch concentrations of > 4% are required for gelling. Some, like carrageenan [β-(1 → 4)-linked galactose and galactose sulfate residues], form hydrogen-bonded 'junction zones' between two polymer chains that stabilize the gels. In agar [also β-(1 → 4)-linked galactose and galactose sulfate residues] and carrageenan, the polysaccharide chains form double helices, which associate further to form more complex junction structures. Pectin forms strong gels (e.g., jam-making) by a different mechanism. In pectin, the highly esterified polymers (methylated polygalacturonic acid) are heated in acid solution (pH about 3.5) in the presence of a high proportion (> 50%) of sucrose. The low pH suppresses ionization of the nonesterified galacturonic acid units, whereas the high degree of methylation favors the formation of junction zones by hydrophobic interactions. The sucrose further competes for water allowing junction zones to form without separation from water. However, when the gel is mechanically disturbed (broken), liquid is released as a result of shrinkage (syneresis).

Alkali metal salts of alginates are soluble in water, although the free acids are insoluble. These polymers are composed of mannuronic (M) and guluronic (G) acids. Three major arrangements of the sugars occur, namely linear regions consisting of M or G (M_n and G_n are known as periodically arranged monosaccharides) interspersed with regions of alternating M and G (MG_n, aperiodic segments). In the presence of divalent cations (especially calcium), the carboxylic groups of adjacent chains are linked to form the so-called 'egg-box' structure. The final three-dimensional structures give exceptionally rigid and stable gels. Addition of sodium phosphate or of sodium ethylenediaminetetraacetate destroys the gel structure by removing calcium ions.

Mixed Polymer Systems – Synergistic and Nonsynergistic Interactions

It is known that the viscosity of binary mixtures of some polysaccharides can be higher than the viscosity of the individual components measured at the same concentration. This synergistic interaction is particularly pronounced in mixtures of guar

(galactomannan) and xanthan gums. This interaction can be interpreted by the Flory and Huggins theory.

Three types of polymeric interactive behavior have been identified and characterized:

1. *Incompatibility*, where there is phase separation.
2. *Compatibility*, where a uniform homogeneous phase is created.
3. *Polymer associations*, where solid coacervates or gels are formed. Most mixed polymer systems are incompatible (especially above solution concentrations of 1–2%). If one of the polymers is charged, miscibility is promoted, but if both species are charged, immiscibility results.

In terms of strength of interactions, agarose–galactomannan and κ-carrageenan–galactomannan interactions are reduced as the galactose content of the galactomannan increases. This is because linear mannan regions form tight associations with double helices (below) of the other polymers. All these interactions can be very important in polysaccharide blends used in food systems.

Electrostatic Interactions

Polysaccharides carrying acidic groups have a strong tendency to bind cations. Neutral polysaccharides do not bind cations, whereas polycarboxylates like demethylated pectin, alginic acid, and xanthan readily bind divalent cations. Similarly, sulfated polysaccharides like κ-carrageenan also bind cations. Certain polysaccharides contain primary amine groups (like chitin and chitosan) and form associations with transition metals.

Polysaccharide–cation interactions lead to two transitions: sol–gel and phase separation. Where sol–gel transitions occur, viscous solutions are present below the transition concentration and above the transition temperature. Above the sol–gel transition concentration and below the sol–gel transition temperature, gels exist that exhibit viscoelasticity. High cation concentrations generally lead to phase separation in the form of precipitation of the polysaccharide.

There is a real nutritional concern (although there is some controversy regarding its significance) that minerals may be lost from the human body owing to polysaccharide–mineral interactions. Acidic conditions in the stomach will tend to discharge the minerals and make them potentially more easily absorbed lower in the intestine. In the almost neutral pH of the small intestine where the minerals would primarily be absorbed, however, re-association of cations with charged polymers will occur. The stability of these associations in this region of the gastrointestinal tract is uncertain as well as the restrictions imposed upon absorption. Additionally, the polysaccharides may be fermented in the large intestine, releasing the minerals for absorption in this region (assuming that there is little or no competition from the microflora, which will also tend to absorb the minerals).

Helical Structures

Polysaccharides can, as discussed above, assume a helical conformation. The helices may be single (one chain) where stabilization is achieved by the presence of a guest molecule (ligand) in the lumen of the helix. These structures are commonly referred to as 'inclusion complexes.' Starch polymers (amylose and amylopectin) are highly effective in forming inclusion complexes; especially amylose with its long linear α-(1 → 4) bonded chains. The formation of a blue color when iodine solution is added to starch is a familiar example, and it has been shown to be due to the formation of a helical complex. The starch–iodine reaction is used as a quantitative tool for the determination of amylose by colorimetry where a blue color is generated with a λ_{max} at 635 nm. This color is a function of the chain length. Amylopectin, with its shorter unit chain, gives rise to a red color with iodine and absorbs light in the region 570–580 nm.

Starch (particularly amylose) can also form inclusion complexes with organic guest molecules. Free fatty acids and their alkali metal salts, acyl monoglycerides of fatty acids, fatty alcohols, and surfactants containing linear acyl groups give rise to amylose–lipid complexes that are insoluble in water at pH < 7. These interactions occur during food processing and are common in extruded foods. Acyl monoglycerides are added to instant mash potato granules to prevent stickiness, and acetylated tartaric acid esters of acyl monoglycerides are added to bread to retard staling. Sodium and potassium salts of fatty acids are permitted additives in Dutch type rusks. Complexes are also possible with many other guest molecules, including flavoring agents, aldehydes, ketones, alcohols, vitamins, and dyes. However, many of these so-called 'complexes' have not been structurally characterized as being true inclusion complexes rather than unidentified associations.

It should be emphasized that amylose in solution occurs as a random coil, and not as a preformed helix. It assumes a helical conformation (left-handed helix) only in the presence of a hydrophobic guest (e.g., palmitic acid), because the interior of the helix is essentially hydrophobic with an affinity for the hydrocarbon chain of the ligand. Inclusion complexes (V-type helices) containing 6, 7, or 8 glucose residues per helical turn can be formed, depending on the

cross-sectional dimensions of the guest molecule. Adjacent helical turns (pitch 0.8 nm) are held together by hydrogen bonding, which is the main stabilizing force of the complexes. It has been shown that a stoichiometric relationship exists between the amylose and the guest molecules when the helix is completely saturated with the ligand. For instance, lauric acid (1.73 nm chain length, C12:0) is completely surrounded by 2.16 helical turns corresponding to about 13 anhydroglucose residues. Arachidic acid (2.74 nm chain length, C20:0) is surrounded by 3.42 helical turns corresponding to 20.5 anhydroglucose units. The long complexed helices fold upon themselves to form three-dimensional structures, which give characteristic X-ray diffraction patterns (amylose V-pattern), owing to the ordered packing of the helices. Complexes are also characterized by their distinct dissociation temperatures, which are mainly functions of the chain length of the ligand and the temperature at which the complexes are formed. The dissociation energy of the complexes is a measure of the energy required to disrupt intrahelical hydrogen bonds. Thermal properties of complexes are studied by differential scanning calorimetry.

Apart from single helices, which form around guest molecules, double helices may be formed from polysaccharide chains. Furthermore, these double helices may form ordered (registered) arrays and consequently generate crystalline regions (see below), as found in native starch granules. Double helices are stabilized by interchain hydrogen bonding and in this form are essentially insoluble.

Crystallinity and Retrogradation

Although helices or similar structures of polysaccharides may be formed from solutions/food systems, these structures need not necessarily be crystalline. For crystallinity to be present (as determined by X-ray diffraction, for example), the helical structures must be associated in ordered/registered structures. The ordered structures have unit cell repeats, which create characteristic diffraction patterns. Many polysaccharides associate to form true crystalline structures, although these are often 'contrived' by making specific derivatives of the polymers.

In food systems, polysaccharides would generally find it difficult to crystallize, because of interference from other molecular species. However, starch polymers (especially amylose) do crystallize relatively quickly in/from solutions and in highly hydrated systems like baked goods (bread, cakes, etc.). The consequence in terms of product texture is apparent to anyone who has eaten stale bread. This contains appreciable quantities of recrystallized or retrograded starch, which, being a form of 'resistant starch' (*See* **Starch**: Resistant Starch), imparts an unpleasant texture and potentially restricts amylase hydrolysis. In some products, there is a relatively high resistant starch content owing to retrogradation, and this is sometimes promoted as a positive dietary feature. Probably, the best-known example is corn flakes.

Formation of Covalent Bonds

Some carbohydrate interactions with other carbohydrates or other types of molecules incorporate true covalent bonding, as described below.

Caramelization

The process of caramelization (*See* **Caramel**: Properties and Analysis; Methods of Manufacture; **Carbohydrates**: Classification and Properties) involves the formation of brown-colored and odorous products from melted sugars. Many confectionery products incorporate steps that cause caramelization (e.g., toffee-making), although commercially available caramel is also added to products to impart a desirable color (e.g., colas and whisky). Hexoses are converted by heat to dicarbonyl compounds via an enediol intermediate. Loss of water leads to 5-hydroxymethylfurfural and related chemical species, which are very reactive. Pentoses give furfural. Maltol and isomaltol are also produced, both of which have a characteristic odor.

Nonenzymic (Maillard) Browning

Nonenzymic browning, as the name indicates, does not depend on the action of enzymes. The NH_2-groups of amino acids (especially the ε-amino group of lysine) in alkaline or neutral pH react with the aldehyde groups of sugars to form brown pigments. The N-substituted 1-amino-1-deoxy-ketoses (Amadori compounds) lead to the formation of a range of dicarbonyl compounds. These compounds may react further with amino acids (Strecker degradation) to form pyrazines, for example. Polymerization of the products of the amino acid sugar reactions causes the formation of very complex polymers (melanoidins), which impart color to foods. Flavor compounds are also generated as a consequence of these processes. Hence, in food systems, these browning reactions generate products with desirable sensory characteristics. However, there is a loss of nutritional quality as the amino acid availability progressively declines. This can be particularly undesirable in low-protein diets, where amino acids (especially essential) are made unavailable. The Maillard reaction can be minimized by lowering the pH, avoiding high temperatures and long storage times, and maintaining a low

water activity. The reaction is inhibited by SO_2 or busulfite salts and can be avoided in some foods by the use of nonreducing sugars.

Other Interactions

Sucrose is readily hydrolyzed to glucose and fructose (invert sugar) either by acids or by the enzyme β-fructosidase (invertase). In soft drinks containing acid (e.g., citric or phosphoric), gradual inversion of sucrose takes place, even at ambient temperature.

See also: **Caramel**: Properties and Analysis; Methods of Manufacture; **Carbohydrates**: Classification and Properties; **Cereals**: Contribution to the Diet; **Functional Foods**; **Gums**: Properties of Individual Gums; **Stabilizers**: Types and Function; Applications; **Starch**: Structure, Properties, and Determination; Sources and Processing; Functional Properties; Resistant Starch; **Sweets and Candies**: Sugar Confectionery; **Water Activity**: Principles and Measurement; Effect on Food Stability

Further Reading

Axelos MAV (1998) Polysaccharide–metal interactions. In: Walter RH (ed.) *Polysaccharide Association Structures in Food*. New York: Marcel Dekker.

Belitz H D and Grosch W (1999) *Food Chemistry*, 2nd edn. Berlin: Springer.

Ikan R (ed.) (1996) *The Maillard Reaction. Consequences for the Chemical and Life Sciences*. Chichester, UK: Wiley.

Karkalas J and Raphaelides S (1986) Quantitative aspects of amylose–lipid interactions. *Carbohydrate Research* 157: 215–234.

Karkalas J, Ma S, Morrison WR and Pethrick RA (1994) Some factors determining the thermal properties of amylose inclusion complexes with fatty acids. *Carbohydrate Research* 268: 233–247.

O'Brien J, Nursten HE, Crabbe MJC and Ames JM (eds) (1998) *The Maillard Reaction in Foods and Medicine*. Cambridge: Royal Society of Chemistry.

Tomasik P and Schilling CH (1998a) Complexes of starch with inorganic guests. In: Horton D (ed.) *Advances in Carbohydrate Chemistry and Biochemistry*. San Diego, CA: Academic Press.

Tomasik P and Schilling CH (1998b) Complexes of starch with organic guests. In: Horton D (ed.) *Advances in Carbohydrate Chemistry and Biochemistry*. San Diego, CA: Academic Press.

Walter RH (1998) Origin of polysaccharide supramolecular assemblies. In: Walter R H (ed.) *Polysaccharide Association Structures in Food*. New York: Marcel Dekker.

Williams PA and Phillips GO (1995) Interactions in mixed polysaccharide systems. In: Stephen AM (ed.) *Food Polysaccharides and Their Applications*. New York: Marcel Dekker.

Digestion, Absorption, and Metabolism

D H Alpers, Washington University School of Medicine, Washington, USA

Introduction

The assimilation of dietary carbohydrate into the body is a complex process involving secretion (salivary glands, stomach, and pancreas), hydrolysis (stomach, small intestine, and colon), and absorption (small intestine and colon; **Figure 1**). Abnormalities in each of these organs or processes can lead to diminished assimilation of carbohydrates.

Digestion

Dietary Carbohydrates

Western diets contain on average 200–300 g of carbohydrates day^{-1}, accounting for 40–50% of total food sources of energy but only 22% of total daily calories. The sources of this carbohydrate include complex polymers (e.g., starch), and simple sugars, including the disaccharides lactose and sucrose, and the monosaccharides glucose and fructose. Nearly all of these carbohydrates, except for lactose in milk and some muscle glycogen, are found in plants. Dietary fiber is the term used to describe polymeric substances (nearly all carbohydrate) that are not digested and absorbed in the small intestine. In addition to dietary

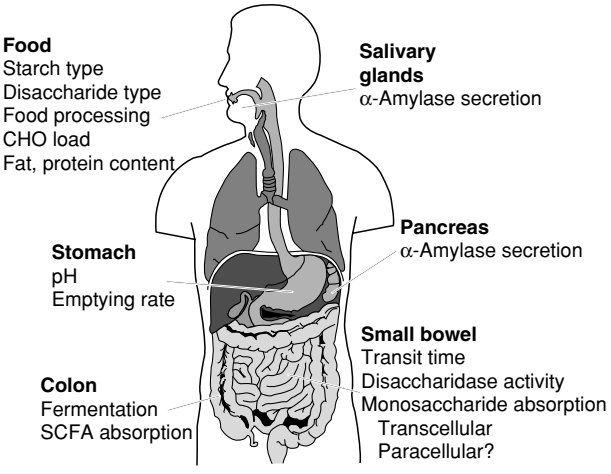

Figure 1 Factors affecting carbohydrate (CHO) assimilation. SCFA, short chain fatty acid. Reproduced from Carbohydrates: Digestion, Absorption and Metabolism. *Encyclopaedia of Food Science and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

fiber, plants contain oligosaccharides (3–8 sugars) mainly as α-galactosides in legumes (e.g., raffinose, stachyose) and as fructans (e.g., fructooligosaccharides) found in onions and other vegetables. Some of the undigested fiber and oligosaccharides (or disaccharides) can be hydrolyzed by colonic bacteria and their products absorbed.

Starch is the major storage form of polysaccharides in plants, along with cellulose (a β1–4 linked glucopolymer), hemicelluloses (e.g., galactans, arabinans), pectins, and gums. Starch is composed of polymers of glucose joined in straight-chain α1–4 linkages without (amylose) or with α1–6 branched chains (amylopectin). Even the starch is not completely digested, and some of the dietary fiber represents 'resistant starch.' Although the assay of 'resistant starch' is not well standardized, its presence is probably proportional to the amount of dietary fiber. The plant content of starch depends upon the type, part, and age of the plant. The efficiency of digestion of plant starch is inconsistent, as plant starches are enclosed in granules that are incompletely and variably disrupted during cooking or processing. Thus, actual dietary starch content may vary significantly from that contained in published tables.

The percentage of carbohydrate ingested as simple sugars has been increasing in recent decades. Sucrose is found in most fruits and vegetables and (in purified form) as table sugar and a sweetener for many foods. Sucrose accounts for about $40\,g\,day^{-1}$ (30% of carbohydrate intake, 9% of calories), two-thirds in the form of sweetener. Lactose is the major sugar in milk and processed foods, and can account for as much as 10% of total carbohydrate calories in non-lactase-deficient persons. Glucose and fructose are also found in fruits and vegetables, but most fructose is ingested as the sweetener, high-fructose corn syrup. Estimated intake for glucose is $20\,g\,day^{-1}$, but may be as high as $150\,g\,day^{-1}$ for fructose. The sugar alcohol D-glucitol (sorbitol), found in many fruits, especially prunes, pears, plums, and cherries, is slowly absorbed. It may cause diarrhea when it is added to foods as a noncaloric sweetener, if consumed in excessive amounts. Another sugar alcohol, xylitol, is tolerated somewhat more effectively by the digestive tract.

Starches and disaccharides are digested by secreted and by mucosal enzymes, respectively, in the mouth, stomach, and small intestine. The resistant portion of starch, undigested or malabsorbed disaccharides and monosaccharides, and the dietary fiber components are digested by bacteria in the lumen of the nonruminant colon. In the ruminant this digestion occurs in the forestomach.

Salivary and Gastric Digestion

Salivary amylase is probably important in initiating starch digestion, depending upon the time spent chewing. Human salivary amylase is 94% identical with pancreatic amylase, but is inactivated in the acid pH of the gastric lumen. Starch supplements are better tolerated in breast-fed than bottle-fed infants, because of the presence of human milk amylase. In humans before 1 year of age, α-amylase activity in the duodenum is fairly low, due to delay in development of full secretory capacity. Presence of starch or its hydrolytic products can protect the enzyme from acid denaturation, and in this way some salivary or milk amylase reaches the more neutral pH of the duodenal lumen.

Gastric juice contains no other carbohydrases than α-amylase. Acid nonenzymic hydrolysis of some carbohydrates can occur, presumably in the stomach, although the extent of this process *in vivo* is probably small. The rate of gastric emptying is inversely related to caloric load, limiting the delivery of undigested food to the duodenum so as not to exceed hydrolytic capacity.

Pancreatic Digestion

Although lipid and protein substrates are hydrolyzed by more than one gastric and pancreatic enzyme, duodenal fluid contains only α-amylase, derived largely from pancreatic secretion. Starch hydrolysis begins with random cleavage of internal α1–4 glucose bonds, proceeding to smaller fragments until only small linear glucopolymers (2–8 residues) or small α-limit dextrins (containing α1–6 branch chains) are left. These compounds require brush border disaccharidases for digestion.

There are many factors that affect the completeness of starch digestion in addition to the physical form and processing of the starch in foods. Meal composition may affect variation in transit time from meal to meal, and intestinal transit time is slowed by nutrient-triggered inhibitory feedback. This feedback is initiated in part by lipids, and mediated by peptide YY. Starch feeding can increase amylase secretion in animals and in newborn humans. Fat and protein may also increase the secretion of pancreatic amylase via release of cholecystokinin. These nutrients also delay gastric emptying, preventing too rapid delivery of starch into the duodenum. Amylase activity can be inhibited by proteins found in certain plants, although their role in normal assimilation of carbohydrates is unclear.

The release of glucose from maltose, resulting from starch digestion, elicits a rise in blood glucose. The amount of this rise (glycemic response) varies

depending upon the form and processing of the food starch, the presence of other nutrients in the meal, and the degree of secretion of the fat-responsive gastric-inhibitory polypeptide (GIP) that mediates the enhancement of the glucose-stimulated insulin release. Because of the variety of factors affecting the glycemic response to food carbohydrates, it is not surprising that this response is variable and often unpredictable.

Intestinal (Brush Border) Digestion

All the hydrolytic products of starch digestion, as well as ingested disaccharides, are hydrolyzed by the brush border carbohydrases, sucrase-isomaltase, lactase, maltase-glucoamylase, and trehalase (Table 1). These enzymes are all α-hydrolases, except for lactase, which cleaves β-bonds. The result of this digestion is monosaccharides that can be absorbed across the apical membrane of the enterocyte by specific transporters. These proteins are ectoenzymes present either as heterodimers (sucrase) or as single subunits (lactase, maltase, trehalase). They are synthesized as single peptide chains and modified by glycosylation with or without subsequent cleavage.

Only monosaccharides can be absorbed readily from the human intestine. Disaccharide absorption is very low, but does increase slightly relative to monosaccharides in conditions such as celiac sprue that produce mucosal damage. The disaccharidase activity is rate-limiting for absorption of glucose and galactose from lactose, and of glucose from hydrolysis of glucose oligomers when α-amylase activity is scarce (e.g., in the newborn child). The distribution of disaccharidases along the length of the gut is designed for efficient digestion, as digestion is nearly complete by mid-jejunum, and the enzyme content peaks in the jejunum. Expression along the vertical axis of the villi is also appropriate for hydrolysis of luminal contents. The enzymes are transcribed maximally in the lower and mid-villus, decreasing to the villus tip, and the enzyme content follows this pattern. Lactase mRNA in adults is more restricted to the lower villus, in contrast to its appearance in all villus cells before birth, but the protein persists in all villus enterocytes. The lower levels of enzymes near the villus tip may be related to their release from the membrane by pancreatic proteases.

The products of amylase digestion (maltotriose, maltose, and α-limit dextrins) cannot be absorbed without further digestion. Two of the disaccharidases, maltase-glucoamylase and sucrase-isomaltase, act sequentially to complete the hydrolysis of starch begun by α-amylase. Maltase-glucoamylase has maximal activity against linear polymers of 5–9 glucose residues, and cleaves all maltotriose, and a small amount of maltose and α-limit dextrins. Sucrase-isomaltase hydrolyzes most of the α-limit dextrins and maltose, providing the monosaccharides that can be efficiently absorbed.

Sucrase-isomaltase is anchored to the membrane by a transmembrane N-terminal fragment, and cleavage into sucrase and isomaltase subunits is mediated by trypsin on the external surface of the enterocyte. Sucrase-isomaltase is expressed in the colon from 12–28 weeks of gestation, but disappears at birth, along with the villus/crypt structure seen in small intestine. At birth in humans there is high sucrase activity in the small intestine that persists into adulthood. In rodents born with immature intestines, sucrase is hardly detectable at birth, and rises to adult levels after weaning. In rats, the high fasting levels of sucrase-isomaltase decline after eating, because of

Table 1 Human carbohydrases used in dietary carbohydrate digestion

Enzymes	*Size of protein (kDa)*			*Substrate specificity*	*Bonds cleaved*	*Products*
	Proprotein		*Mature*			
	−CHO	*+CHO*				
Amylase						
Salivary	[a]	[a]	55	$Glu_{n>6}$	Glcα1,4Glc	Maltose, maltotriose, α-limit dextrins
Pancreatic	[a]	[a]	60			
Sucrase-isomaltase	155	210[b]	130[e]+	Sucrose, α-limit dextrins	Glcα1,2Fru	Glucose, fructose
		245[c]	145[e]		Glcα1,4Glc	
Lactase-phlorizin hydrolase	210	215[b]	160[d]	Lactose, cellobiose, β-glycosides	Glcβ1,4Gal	Glucose, galactose
		225[c]	145[e]		Glcβ1,4Glc	hydrophobic aglycons
Maltase-glucoamylase	255	285[b]	335[c]	$Glc_{n=2-9}$	Glcα1,4Glc	Glucose
					Glcα1,6Glc	
Trehalase	65.5[f]	?75[f]	75[g]	trehalose	Glcα1,Glc	Glucose

[a]Not synthesized as proproteins; [b]high mannose form; [c]fully glycosylated form; [d]intracellular cleavage form delivered to brush border (human) or cleaved at brush border surface (rat); [e]final active form after surface cleavage by pancreatic trypsin; [f]predicted size (from cDNA) of rabbit enzyme; [g]mature form is linked by glycosylphosphatidylinositol, not by transmembrane segment as for other disaccharidases.

rapid enzyme turnover mediated by pancreatic proteases. Starvation decreases sucrase activity, and refeeding restores it. Sucrose and fructose more than starch induce sucrase and maltase activities. Sucrase deficiency in humans results from a number of different defects in glycosylation, intracellular transport, and possibly in altered catalytic activity.

The digestion and absorption of lactose, the major carbohydrate in milk, are often variable, and are related to lactase-phlorizin hydrolase levels that are rate-limiting for absorption, as well as the rate of gastric emptying, small intestinal transit rate, and the degree of colonic digestion and salvage. Lactase is anchored by a COOH-terminal sequence and then undergoes SH-sensitive intramolecular folding in the endoplasmic reticulum. Prior to arrival at the apical membrane, the 225-kDa lactase precursor undergoes cleavage to a 160-kDa intermediate form. On the apical cell surface final cleavage to the 145-kDa active enzyme occurs by the action of trypsin. Unlike the human, the rat lactase undergoes a several-step cleavage on the apical surface. Lactase activity is high at birth in most mammals, and declines in the human during childhood or adolescence, reaching levels about 5–10% those at birth. Initiation of this process is 'hard-wired,' but can be modified by hormones and nutrients, at least in animals. The age at which the fall in activity occurs varies according to the genetic background of the individual. Inherited lactase deficiency in humans involves transcriptional and posttranscriptional mechanisms. The phenotype in many Caucasians for persistence of lactase is inherited as an autosomal recessive trait. In all other mammals, lactase activity falls after weaning. Lactose feeding in humans does not increase lactase activity, and the absence of dietary lactose, in diets used for patients with galactosemia, does not cause a fall in activity. Lactase activity falls during malnutrition in children, but not in adults. A variety of infectious and inflammatory processes decrease lactase activity more than that of other disaccharidases. Activity may be slow to recover after an acute infectious process, and remains low in chronic inflammatory conditions, such as Crohn's disease and human immunodeficiency virus (HIV-1) infection. Table 2 summarizes the factors altering carbohydrase activity.

Maltase-glucoamylase accounts for about 2% of the brush border protein, and cleaves small glucose polymers and α-limit dextrins produced by α-amylase activity. Unlike sucrase and maltase in humans, the protein is inserted into the brush border membrane without any intra- or extracellular cleavage. In the pig and rat surface cleavage by pancreatic enzymes occurs. As all forms are enzymatically active; the significance of these structural differences is unknown. Also, unlike sucrase and lactase there is relatively little change in maltase levels at the time of birth, either in human or animal intestines.

Colonic (Bacterial) Digestion

Up to 70 g of carbohydrate (including the components of dietary fiber and other dietary carbohydrates that escape small bowel digestion) may reach the colon each day in humans. This carbohydrate is fermented by colonic bacteria to short-chain fatty acids (largely acetate, propionate, and butyrate), hydrogen gas, and methane. The overall fermentation reaction can be estimated thus:

$$34.5\ C_6H_2O_6 \rightarrow 48\ CH_3\,COOH + CH_3\,CH_2\,COOH + 5\ CH_3(CH_2)_2\,COOH + 23.75\ CH_4 + 34.25\ CO_2 + 10.5\ H_2O$$

In ruminants these fatty acids are produced in the forestomach, and in other mammals in the cecum and colon. Butyrate and caproate can be released by

Table 2 Factors affecting carbohydrase activity

Enzyme	*Factor*	*Enzyme content*	*Mechanism*
α-Amylase	Starch feeding	↑	↑d mRNA content
	Hydrolytic products	↑	Degradation inhibited
	Prenatal development	↑	↑d mRNA content
Disaccharidase	Hydrolytic products	No change	Active site
	Pancreatic enzymes	↓	Shortened *t*/2
	Pancreatic insufficiency	↑	Prolonged t/2
Sucrase-isomaltase	Single-meal feeding	↓	↑d mRNA content
	Sucrose or fructose feeding	↑	↑d mRNA content
	Prenatal development	↑	↑d mRNA content
	Diabetes mellitus	↑	Multifactorial
Lactase-phlorizin hydrolase	Postnatal development	↓	↓d mRNA
	Infection and inflammation	↓	↓d translation
	Malnutrition (children)	↓	Multifactorial ↓d mRNA

t/2, half life.

gastric lipase action on milk triacylglycerols, and may be nutritionally important in infants. The carbohydrate moieties of mucins and other glycoproteins are also used to support bacterial growth. Thus, endogenous carbohydrates are important in perpetuating colonic bacterial cultures, which in turn are responsible for salvaging much of the nutrient value of undigested dietary carbohydrate.

Absorption and Metabolism

General Characteristics of the Mucosal Barrier

The only carbohydrates absorbed to any extent across the intestinal mucosa are monosaccharides. The mechanism for absorption involves the use of either Na^+-coupled or non-Na^+ transporters (Table 3). The GLUT transporters mediate a facilitated diffusion process, whereas SGLT1-associated transport is an active process, involving the Na^+/K^+-ATPase located in the basolateral membrane of the enterocyte. This protein uses adenosine triphosphate (ATP) to maintain a low intracellular Na^+ concentration, providing a chemical gradient for Na^+ (and coupled glucose) from the lumen to the cell. The delivery of the solute to the cell surface is a function of the efficiency of stirring of the luminal contents. Although the thickness of this unstirred water layer in humans has been thought to be large, in the dog, where stirring can be vigorously conducted, the unstirred water layer is only 40 μm. Metabolism of disaccharides proceeds in two steps, with surface digestion by disaccharidases preceding absorption of the sugar. Thus, the delay in traversing this barrier should be brief, and the rate-limiting step for absorption is most likely due to the apical sugar transporters.

Water absorption is coupled with that of the monosaccharide and Na^+. The mechanism whereby this occurs is not well defined. In some epithelial cells, such as renal collecting ducts, there are specialized water transporters called aquaporins. Expression of these proteins in the intestinal mucosa is either nonexistent or sufficiently low to be functionally insignificant. A unique hypothesis suggests that SGLT is able to account for the water absorption, as about 260 water molecules are cotransported *in vitro* for every 2 Na^+ and every one sugar molecule. This capacity should be large enough to account for about 5 l of waterday^{-1} for a human ingesting a diet of average carbohydrate content. This hypothesis has yet to be tested *in vivo* in humans.

Another hypothesis suggested from data in small rodents that water movement in the intestine is paracellular, via the tight junctions. This route might also account for absorption of ions and other solutes for which there are no specific apical transporters. Moreover, the tight junctions are subject to regulation of their cross-sectional area, and could serve as a size-selective filter mechanism. Calculations have suggested that the rates of glucose absorption in humans ($> 200\,mmol\,h^{-1}$) exceed by an estimated 2–10-fold the predicted capacity for active sugar absorption. However, L-glucose, a driver for paracellular movement potentially equivalent to D-glucose, does not produce significant water absorption in either dog or rat intestines. Thus, there are insufficient data to suggest that paracellular movement of sugar or its coupled water is important in mammals.

Glucose and Galactose Absorption

Active hexose transport is driven by an electrogenic sodium gradient. The stoichiometry for Na^+-coupled glucose transport is 2:1. In intact brush border vesicles, both high- and low-affinity glucose uptake has been found, depending on the species. The low-affinity uptake would be suitable for the high intraluminal postprandial concentrations. These 'systems' vary independently depending on the developmental or fed status of the animal form which the membranes were made.

Despite these data suggesting more than one active hexose transport system in intestinal brush border membranes, only one Na^+-coupled transporter has

Table 3 Major cellular monosaccharide transporters

Transporter	*Na^+-coupled*	*Organ expression*	*Transport function*
GLUT1	No	Many tissues, especially RBC, blood–brain barrier	Glucose, across blood–tissue barriers
GLUT2	No	Liver, kidney, small bowel, pancreatic β-cells	Low-affinity, across basolateral membranes
GLUT3	No	Many tissues, especially brain	Glucose across basolateral membranes
GLUT4	No	Skeletal and cardiac muscle, adipose tissue	Insulin-dependent uptake
GLUT5	No	Small bowel, kidney, testis, adipose tissue, muscle	Fructose ≫ glucose across apical ≫ basolateral membranes
SGLT1	Yes	Intestine, kidney	High-affinity glu > gal, apical transport
SGLT2	Yes	Kidney	Low-affinity glu > gal apical transport

RBC, red blood cells.

been found in intestine, SGLT1. In humans the activity and expression of this transporter are maintained by the presence of luminal nutrients, as suggested by the brush border experiments in fed and fasted animals. The low-affinity SGLT2 is only found in kidney, and its kinetics can be explained by reducing the Na^+/glucose coupling from 2:1 to 1:1. SGLT1 is a glycoprotein of 75 kDa with 14 membrane-spanning regions, existing as a homotetramer in the membrane. The single *N*-linked carbohydrate side chain is not required for function. The relative specificity of the transporter is the same as that previously characterized for the active membrane transport system, i.e., D-glucose>α-methyl-D-glucose>D-galactose>3-*O*-methyl-D-glucopyranose>>>L-glucose. It is clear that SGLT1 accounts for all of the active glucose transport, because mutations in this protein account for the entire phenotype in patients with hereditary glucose-galactose malabsorption.

Glucose absorbed into the enterocyte is transported across the basolateral membrane by facilitated diffusion mediated by GLUT2. This high-capacity transporter has 12 membrane-spanning regions, and can transport fructose as well as glucose. Although glucose is metabolized by the enterocyte, the preferred energy substrates for this cell are amino acids, preferentially glutamate, glutamine, and aspartate. In the presence of amino acids, intestinal metabolism of glucose is decreased. After exit from the cell, glucose enters the portal vein, and is delivered to the liver and peripheral tissues (mostly muscle), in which tissues the glucose is extensively metabolized.

Fructose Absorption

Fructose transport occurs by an Na^+-independent, saturable system of lower capacity than that for glucose or galactose. The capacity for fructose absorption in humans is limited, although theoretical estimates of absorption capacity are relatively high. GLUT5 mediates all or most of fructose transport across the apical membrane of enterocytes. Human GLUT5 transports fructose alone, but the rat homolog recognizes both glucose and fructose. However, absorption of fructose in humans can be inhibited by the presence of glucose. Thus, it is possible that a second apical fructose transporter exists. Unlike the relatively wide tissue distribution in humans (Table 3), rat GLUT5 is expressed largely in the small bowel, kidney, and brain. Fructose is poorly metabolized in the enterocyte, and is transported from the cell by basolateral GLUT2, and in humans by basolateral GLUT5 as well. Expression of GLUT5 is increased in animals fed fructose. This adaptation accompanies the increase in sucrase-isomaltase found after fructose or sucrose feeding.

Short-chain Fatty Acid Absorption

Short-chain fatty acids are the major nutrients produced by bacterial fermentation. The usual starting substrates are carbohydrates. In humans the fermentation products are produced and absorbed in the colon. Both small bowel and colonic mucosa readily absorb unionized short-chain fatty acids. The transporter responsible for this uptake is most likely a member of the monocarboxylate-type transport proteins, perhaps by the anion exchanger AE2 found in apical membranes of intestinal mucosal cells. The anion gradient across the apical membrane is butyrate>bicarbonate>propionate>chloride. The capacity of this transporter is much lower than that for SGLT1, but is still sufficient to achieve some salvage of malabsorbed carbohydrate.

Unlike hexoses in the small intestine, short-chain fatty acids are partly metabolized in the colonic cells, and appear to be a major nutrient source. Most of the fatty acids are metabolized intracellularly to CO_2. Estimates of the contribution of short-chain fatty acid metabolism to the basal metabolic requirement vary from low (1–2% in the pig and 6–9% in humans) to high (30–40% in the rabbit). The importance of this pathway in humans increases in patients with sugar malabsorption, when delivery of nonabsorbed sugar is increased to the colonic lumen.

See also: **Carbohydrates**: Classification and Properties; **Dietary Fiber**: Properties and Sources; **Fructose**; **Glucose**: Glucose Tolerance and the Glycemic (Glycaemic) Index; **Lactose**; **Starch**: Structure, Properties, and Determination; Resistant Starch; **Sucrose**: Dietary Importance; **Sugar**: Refining of Sugarbeet and Sugarcane

Further Reading

Alpers DH (1994) Digestion and absorption of carbohydrates and proteins. In: Johnson LR (ed.) *Physiology of the Gastrointestinal Tract*, 3rd edn, pp. 1723–1749. New York: Raven Press.

Klein S, Cohn SM and Alpers DH (1999) Gastrointestinal function. In: Shils ME, Olson JA, Shike M and Ross AC (eds) *Modern Nutrition in Health and Disease*, 9th edn, pp. 605–629. Baltimore: Williams & Wilkinson.

Loo DDF, Zeuthen T, Chandy G and Wright EM (1996) Cotransport of water by the Na^+/glucose cotransporter. *Proceedings of the National Academy of Science of the USA* 93: 13367.

Levin RJ (1999) Carbohydrates. In: Shils ME, Olson JA, Shike M and Ross AC (eds) *Modern Nutrition in Health and Disease*, 9th edn, pp. 49–65. Baltimore: Williams & Wilkinson.

Thomson AB and Wild G (1997) Adaptation of intestinal nutrient transport in health and disease. *Digestive Diseases and Sciences* 42: 453–469, 470–488.

Traber PG (1999) Carbohydrate assimilation. In: Yamada T (ed.) *Textbook of Gastroenterology*, 3rd edn, pp. 404–428. Philadelphia: Lippincott-Raven Press.

Wright EM and Loo DD (2000) Coupling between Na^+, sugar, and water transport across the intestine. *Annals of the New York Academy of Science* 915: 54–66.

Requirements and Dietary Importance

I Macdonald, University of London, London, UK

Introduction

Dietary carbohydrates are the cheapest source of energy for metabolism by the body. They are cheap because they are produced by plants for energy metabolism and storage and consequently can be harvested in most temperate and humid climates. To produce a million calories for human consumption requires 0.08 ha for sucrose, 0.16 ha for potatoes, 0.4 ha for wheat and 6.88 ha for cattle. Thus a very important factor when considering requirement of dietary carbohydrate lies largely in economic necessity. Carbohydrates have long been the mainstay in world dietaries, with sucrose and starches being the major sources for human consumption. In regions where per capita incomes are low and the balance between food production and demand is close, consumption of carbohydrates as cereals or starchy root vegetables largely meets the need for energy intake. In some underdeveloped countries carbohydrates supply up to 90% of the dietary energy consumed and, of course, consumption at this level gives rise to concern about deficiencies of the other nutrients such as proteins, minerals, and vitamins. In industrial countries dietary carbohydrates provide about half the daily energy intake.

Although the total carbohydrate intake may be similar between some developing and developed countries, the nature of the carbohydrates eaten and the proportion they contribute to the energy intake may not be similar. Carbohydrates from staple foods such as cereals and roots are composed mainly of starch and may represent 85% of the carbohydrate intake in developing countries but only 62% in affluent ones. The difference is largely made up by carbohydrates from fruit, sucrose, corn syrups (mainly in North America) and, to a lesser extent, lactose.

Surveys in the 1980s have shown that the consumption of sucrose is changing, tending to fall in the more developed countries while rising in the other countries, and it has been suggested that the saturation point occurs at about $160\,g\,day^{-1}$ per person. In developing regions, starch consumption from cereals and roots seems to be related to income, rising when personal income rises. In developed regions, cereal consumption, and therefore starch intake, has fallen, especially in Japan and Eastern Europe, with little change in North America and Australia. These changes might also reflect income, for at higher income levels food choice is not limited to purchasing power and as income rises the proportion of energy derived from carbohydrate tends to fall.

Complex carbohydrates, like other insoluble components of the diet, have no taste, only 'mouth feel,' but carbohydrates with a comparatively smaller molecular weight are soluble and stimulate the taste buds for sweetness. This property of sugars, and especially sucrose, makes them an organoleptic requirement for many, especially the young. It has been suggested that this property of the simple carbohydrates is useful in making palatable those foods which are nutritionally desirable but may not be so pleasant to consume. (*See* **Cereals**: Contribution to the Diet; **Sucrose**: Dietary Importance.)

Metabolic Importance of Dietary Carbohydrate

The major carbohydrates that are found in the body tissues in humans are glycogen, glucose, fructose, galactose, and lactose (in lactation).

Glycogen, the storage form of carbohydrate, is found in the liver and muscle. Before being released into the blood for transport to other regions of the body, it has to be hydrolyzed to the monosaccharide, glucose. (*See* **Glycogen**.)

Glucose is the common metabolic currency of carbohydrate in the body and all cells are able to metabolize glucose to carbon dioxide and water with the consequent release of energy. However, glucose is not only used as a readily available source of energy but it can also be converted to glycogen or, if consumed in excess, to fat for storage. Although glucose can be utilized by all cells, it is only in a few organs that glucose is essential, and these include the brain and red cells of the blood. During pregnancy and growth, glucose is essential for the formation of cell constituents. (*See* **Glucose**: Function and Metabolism.)

Fructose, which largely comes from the sucrose in the diet and also from honey, does not seem to have a specific role in the body and the only site of production of fructose in humans seems to be in the seminal vesicles, where it is made from glucose.

Galactose is, along with glucose, one of the monosaccharides in lactose; practically the only dietary source of galactose is milk. Galactose is synthesized from glucose in the lactating mammary gland. (*See* **Galactose**.)

Metabolic Sources of Carbohydrate other than Dietary Carbohydrate

As glucose is essential to the biochemistry of the body it is not surprising that there are sources of glucose other than those in the diet. One of these is liver or muscle glycogen but these stores of 'animal starch' are very limited. After 24 h or so of complete starvation the stores are empty; yet, as long as water is supplied, a normal-weight person can survive complete starvation for 50–60 days, so alternative sources of glucose must exist.

Another source of glucose in the body is from the glycerol moiety of triglyceride. This part of the triglyceride molecule forms about 10% of its molecular weight, and when triglycerides are hydrolyzed the released glycerol can be converted to glucose. Another source of glucose within the body is from the so-called glucogenic amino acids, which can be metabolized to glucose. (*See* **Triglycerides**: Structures and Properties.)

Although dietary carbohydrates can be converted to fat in the body, fat cannot be converted to carbohydrate.

Metabolic Requirements for Dietary Carbohydrate

Apart from the fact that a diet with little or no carbohydrate would be most unpalatable, there is a metabolic need for dietary carbohydrate. Total deprivation of energy intake has been used in the treatment of obesity and this has provided the opportunity to study the effects of a zero intake of dietary carbohydrate. The effects of diets which are high in protein and fat and very low in carbohydrate have also been studied and the striking consequence of a low or zero carbohydrate intake is that the breakdown of fat in the body cannot go to completion. The final end product of fat metabolism is then a two-carbon chain remnant, existing in the blood as acetoacetic acid or β-hydroxy butyrate. The classical breath odor in this condition, known as ketosis, is due to acetone excretion by the lungs. Clinically, ketosis occurs in uncontrolled diabetes mellitus and after 24 h or more of dietary carbohydrate deprivation in otherwise healthy individuals. Thus carbohydrate is needed in order that the catabolism of fat can be completed to carbon dioxide and water. (*See* **Fats**: Digestion, Absorption, and Transport.)

There are two disadvantages of the ketotic state in an individual: (1) the judgment of such a person may be impaired and it could therefore be unwise to handle potentially dangerous machinery (e.g., a car) while ketotic; (2) the ketone bodies excreted in the breath and urine contain utilizable energy and thus represent an energy loss to the body and diminishing body stores. Production of ketone bodies in large quantities, as in uncontrolled diabetes mellitus, can lead to coma and death. Thus there is a need in all individuals for a minimal daily intake of carbohydrate that can supply the amount of glucose necessary to complete the breakdown of depot fat.

What is the minimum desirable intake of glucose or its equivalent? Under normal circumstances the adult brain needs about 140 g of glucose per day and the red blood cells need another 40 g day^{-1}. If diet contains no sugars or starch, about 130 g of glucose can be provided endogenously from the catabolism of protein and from the glycerol moiety of depot fat, thus leaving a shortfall of approximately 50 g day^{-1} to be obtained from the diet. It is therefore possible to state a minimum desirable intake of glucose or its equivalent for adults. After several days in the ketotic state the brain, a major consumer of glucose, adapts and can use, to some extent, the energy present in the ketone bodies, thus lessening the minimal daily requirement for dietary carbohydrate.

Metabolic Knock-on Effects of Dietary Carbohydrate

All dietary carbohydrates have to be broken down to their constituent monosaccharides (glucose, fructose, or galactose) before they can be absorbed from the intestine but only glucose stimulates the release of insulin, a hormone which not only accelerates the cellular uptake of glucose but also facilitates the uptake of amino acids. Insulin is, in general, an anabolic hormone, so that the glucose provided by the carbohydrate in the diet can have far-reaching effects on the metabolism of other dietary constituents, through its ability to bring about the release of insulin.

The amount of energy stored in the body as carbohydrate is minute when compared with that stored as fat or protein. The total quantity of carbohydrate in liver, muscle, kidney, and other tissues, plus the glucose that circulates in the blood, amounts to about 1800 kcal (7.56 MJ). However, it has been found that the carbohydrate stored in the skeletal muscle can be increased considerably by reducing the proportion of fat in the diet and replacing it with carbohydrate. This is of importance to those who compete in endurance sports and has led to the expression 'carbohydrate loading.'

Finally, there is what used to be called the protein-sparing effect of dietary carbohydrate. When the energy intake in the diet is below requirement, administration of carbohydrate (which raises insulin levels) reduces the breakdown of body protein, whereas dietary fat under comparable circumstances has a negligible effect on reducing protein breakdown. (*See* **Protein:** Interactions and Reactions Involved in Food Processing.)

See also: **Cereals**: Contribution to the Diet; **Fats**: Digestion, Absorption, and Transport; **Galactose**; **Glycogen**; **Protein**: Interactions and Reactions Involved in Food Processing; **Sucrose**: Dietary Importance; **Triglycerides**: Structures and Properties

Further Reading

British Nutrition Foundation (1987) *Sugars and Syrups*. London: British Nutrition Foundation.

British Nutrition Foundation (1990) *Complex Carbohydrates in Foods*. London: Chapman and Hall.

Food and Agriculture Organization/World Health Organization (1998) *Carbohydrates in Human Nutrition*. Food and Nutrition paper no. 66.

Metabolism of Sugars

I Macdonald, University of London, London, UK

Background

There are only three sugars that are normally present within the body, and these are the monosaccharides, glucose, fructose, and galactose. Glucose is the prime carbohydrate in the body; fructose is consumed mainly with glucose as the disaccharide sucrose and plays a relatively minor role; galactose is associated with lactation and milk ingestion, and is, along with glucose, one of the monosaccharides of the lactose molecule. (*See* **Fructose; Galactose; Glucose:** Function and Metabolism.)

Glucose

To produce energy, the glucose molecule must enter a cell from the blood and there be converted to glucose 6-phosphate in the cytoplasm. This compound can then be metabolized in a number of ways, depending on the needs and versatility of the cell. It can (1) be broken down to pyruvic acid/lactic acid, (2) go down the pentose phosphate pathway, or (3) form glycogen (**Figure 1**).

1. The most important way by which energy is released from the glucose molecule is by the splitting of the molecule to form two molecules of pyruvic acid (glycolysis). This end product of glycolysis may then enter the tricarboxylic acid cycle inside the mitochondria of the cell and be broken down completely to carbon dioxide and water with the release of energy.

   When quantities of pyruvic acid and hydrogen atoms become excessive, as in severe exercise, these two products inhibit glycolysis and react with each other to form lactic acid.
2. The pentose phosphate pathway accounts for as much as 30% of the glucose metabolism in the liver and more than this in the fat cells. This pathway supplies reducing power for fat synthesis from carbohydrate sources. Most enzymes involved in carbohydrate metabolism require vitamin B metabolites as essential co-factors.
3. When glucose is not immediately required for energy, the extra glucose that continually enters the cells is stored as glycogen or converted to fat. Glucose is preferentially stored as glycogen, and only when the cell approaches saturation with glycogen is the additional glucose then converted to fat. (*See* **Glycogen.**)

Fructose

Fructose in the bloodstream is utilized about twice as fast as blood glucose; the liver and, to a lesser extent, the kidney and small intestine are the main sites of

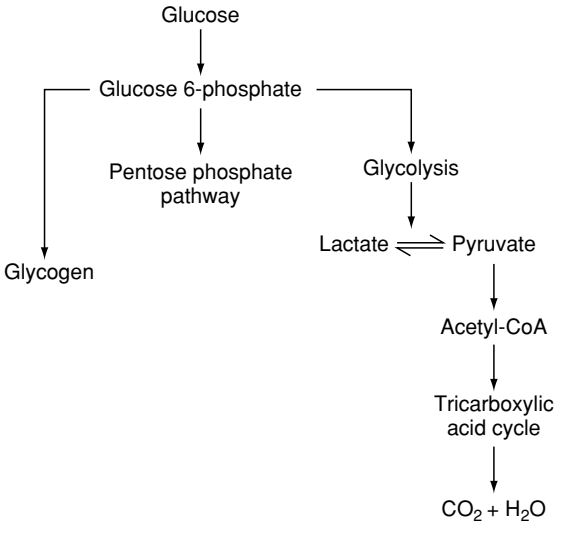

Figure 1 Metabolic options for glucose metabolism. Reproduced from Carbohydrates: Metabolism of Sugars, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

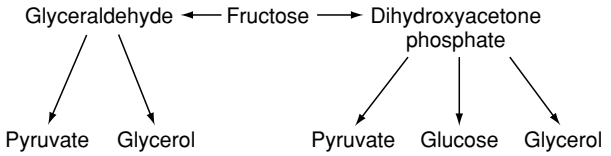

Figure 2 Metabolism of fructose. Reproduced from Carbohydrates: Metabolism of Sugars, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

fructose metabolism. The utilization of fructose by other peripheral tissues seems to be negligible. The first step in the metabolism of fructose is the formation of fructose 1-phosphate, which then splits to form two 3-carbon molecules, namely glyceraldehyde and dihydroxyacetone phosphate. As shown in Figure 2, these trioses can form pyruvate, glycerol, and, in the case of dihydroxyacetone phosphate, glucose.

Fructose is mainly converted to glucose and lactic acid, with perhaps up to 3% of the fructose being converted to triglycerides; ketone bodies; glycerol and sorbitol are minor end products.

Metabolism of Sugars in the Liver

Glucose

In the liver, glucose has no difficulty in crossing the cell wall; the main factors influencing the rate of entry are the concentration of glucose in the blood and the ability of the enzymes in the cell to dispose of the glucose. Much of the glucose that is absorbed from the intestine never reaches the peripheral circulation, as it is taken up in the first pass through the liver.

The liver is also capable of releasing glucose into the blood, either as a result of the breakdown of glycogen or protein, or as a result of synthesis from glycerol. The balance of input and output of glucose by the liver is entirely controlled by hormones, acting via the enzymes within cells:

1. Insulin, which is produced by the β cells of the pancreas, accelerates glycogen formation. In other cells in the body, insulin is solely concerned with the transfer of glucose into the cell, but in the liver, insulin influences the synthesis of enzymes.
2. Glucagon is produced by the α cells of the pancreas, and causes a rapid breakdown of liver glycogen.
3. Adrenaline, like glucagon, stimulates the breakdown of glycogen, but, unlike glucagon, it is able to do this in muscles as well; it does not stimulate insulin release.
4. Glucocorticoids are produced by the adrenal gland. These hormones maintain and aid the formation of glycogen. (*See* **Hormones**: Adrenal Hormones.)

Fructose

In man, most of the fructose absorbed from the intestine is taken up by the liver, hence the low blood levels of fructose after its ingestion. Compared with glucose, fructose has a greater ability to form lactate, but, unlike glucose, fructose administration leads to an increase in blood uric acid levels and, for this reason, has been used as a screening test for gout. Also in the liver, fructose is converted to the lipid triglyceride to a greater extent than is glucose; again, unlike glucose, whether given orally or intravenously, fructose speeds up the metabolism of ethanol by the liver.

Sugars and Muscle

The muscle mass, because of its size and sheer need, plays a principal role in carbohydrate metabolism. The uptake of glucose by muscle cells is insulin-sensitive, but the metabolism of fructose in the cell is probably minimal, and only the pyruvate and lactate formed from fructose in the liver would be of any support to the muscle cell.

Sugars and Depot Fat

It has been known since 1852 that dietary carbohydrate can be converted to depot fat, and this would appear to be in conflict (bearing in mind the current obesity epidemic) with the current dietary advice from several international authorities to reduce the fat intake and replace it by carbohydrate. However, when energy balance is in equilibrium, high-carbohydrate/low-fat diets do not result in appreciable *de novo* lipogenesis.

However, a high-carbohydrate diet can alter the blood lipid pattern in a direction that may be considered undesirable in terms of cardiovascular disease, but this change may be offset by the weight loss that usually occurs on switching to a high-carbohydrate diet.

There is no metabolic reason why sugars should be different in their ability to be converted to depot fat. The FAO/WHO 1998 consultation on Dietary Guidelines proposes no specific limit for sugar consumption, since any putative relationship of sugar consumption to obesity is offset by the inverse relationship between sugar and fat intake.

Fructose can enter the adipocyte, but its rate of transport is slow, and only at high concentrations of fructose do significant amounts enter the adipocyte.

Sugars and the Brain

The brain seems to be entirely dependent on glucose and oxygen, and a reduction in the supply of either will soon lead to irreversible damage. The brain

removes a fixed amount of glucose per unit time, irrespective of the blood concentration of glucose, and its uptake is not dependent on insulin. There has been a suggestion that the insulin release following glucose ingestion increases the level of serotonin in the cerebral tissue and that this compound diminishes the sensation of pain and produces a feeling of well-being.

Sugars and the Metabolic Rate

The increase in metabolic rate after the ingestion of various sugars is greater after sucrose or a mixture of fructose and glucose than after glucose alone, suggesting that the body 'handles' fructose more efficiently than glucose.

Sugars and the Fetus and Neonate

As lipid does not cross the placenta to any great extent, the fat present in the infant at birth must have been synthesized in the fetus from glucose or amino acid. Changes in the maternal blood glucose level are quickly reflected in the fetal blood.

In contrast to glucose, fructose cannot cross the placenta, although some animal species (not humans) can transform glucose to fructose in the placenta.

In the neonate, the sole dietary source of carbohydrate is the lactose in milk, but no specific role for lactose has been identified.

Factors Affecting the Metabolic Response to Dietary Sugars

Sex of the Consumer

The increase in blood triglycerides seen in men after a diet high in fructose is not seen in young women but is found in postmenopausal women. Although both sexes increase hepatic lipogenesis after fructose ingestion, it is possible that premenopausal women are able to remove the blood triglycerides more rapidly.

Type of Fat Accompanying the Carbohydrate

A synergistic effect of sucrose and animal fat on blood triglycerides has been shown, and the raised levels found after a diet high in sucrose are considerably reduced by polyunsaturated fat accompanying the sucrose.

Dietary Protein

The recovery of serum albumin after protein deficiency seems to be slower with sucrose in the diet than with starch, and the interplay of the metabolism of sugars and protein becomes more interesting when it is appreciated that the amino acids arginine and leucine stimulate the release of insulin. (*See* **Protein**: Interactions and Reactions Involved in Food Processing.)

'Sensitivity' of the Consumer

The extent to which sugars are converted to lipids, especially triglycerides, seems to vary between individuals. Those persons whose level of triglycerides in fasted blood is high, and who may be more prone to coronary heart disease, have a greater increase in these triglycerides after consuming carbohydrates than persons with normal lipid levels.

Species

There are not only within-species differences in metabolic responses to sugars but also more marked differences between the species. For example, rats can absorb fructose from the intestine very rapidly, and this will affect the metabolic handling of fructose when compared with humans.

See also: **Fructose**; **Galactose**; **Glucose**: Function and Metabolism; **Glycogen**; **Hormones**: Adrenal Hormones; **Protein**: Interactions and Reactions Involved in Food Processing

Further Reading

Dickens F, Randle PJ and Whelan WJ (1968) *Carbohydrate Metabolism and its Disorders*, vols. 1–3. London: Academic Press.

FAO/WHO (1998) *Carbohydrates in Human Nutrition*. Rome: FAO/WHO.

Macdonald I and Vrana A (eds) (1986) Metabolic effects of dietary carbohydrates. *Progress in Biochemical Pharmacology* 21.

Reiser S and Hallfrisch J (1987) *Metabolic Effects of Dietary Fructose*. Boca Raton, FL: CRC Press.

Determination

D A T Southgate, Norwich, Norfolk, UK

Introduction

The carbohydrates in foods are a heterogeneous group of substances ranging from monosaccharides, such as glucose and fructose, through to the complex polysaccharides found in the matrix of the plant cell walls in foods. In many cases the carbohydrate species may be limited and typical of the food or group of foods, such as lactose in milk and milk products, but

many foods contain complex mixtures which present many problems for the analyst.

As in so many aspects of food analysis, the choice of method is very much decided by the purpose of the analysis and the uses for which the analytical information is required. Thus, for quality control purposes a relatively simple empirical method may be adequate, whereas for nutritional research purposes the current trend is a requirement for relatively complete and chemically specific data.

The food matrix itself is frequently a deciding factor in the choice of method, as this will determine the extraction procedures required and the procedures needed to free the extracts of interfering materials. Thus many analytical procedures have been devised that are specific for a particular group of foods and the carbohydrates they contain.

The Carbohydrates in Foods

It is possible to classify the carbohydrates in foods according to a number of different principles. Table 1 illustrates a classification based primarily on structural terms which is linked to the physiological behavior of the carbohydrates once consumed: this is the most useful classification nutritionally.

Nutritionally it is convenient to consider the carbohydrates in foods as falling into three major groups:

1. Sugars, including mono- and disaccharides.
2. Oligosaccharides with three to nine monosaccharide units.
3. Polysaccharides with 10 or more monosaccharide units.

Oligosaccharides with three to nine monosaccharide units fall into two subgroups: the malto series, which are hydrolyzed by the brush border α-glucosidases, and the raffino and fructo series, which are poorly absorbed in the small intestine but are used as substrates by the microflora of the large intestine. The term short-chain polysaccharides has been suggested for this group because analytically it is difficult to separate some lower polysaccharides from conventional oligosaccharides.

Polysaccharides with 10 or more (usually very many more) monosaccharide units also fall into two subgroups: first, the α-glucans principally, the starches, but also including glycogen, and second, the nonstarch polysaccharides (NSP), which include a wide range of structural types of polysaccharides derived from plant cell wall materials in foods and a range of polysaccharide food additives.

Table 1 The major carbohydrates in foods

Number of monosaccharide units	*Class*	*Major examples*	*Physiological characteristics*
1	Monosaccharides	Glucose Fructose	Absorbed directly in the small intestine
2	Disaccharides	Sucrose Lactose Maltose	Hydrolyzed in the brush border by specific disaccharidases
3–9	Oligosaccharides	Malto- Fructo- Raffinose Stachyose Verbascose	Hydrolyzed in the brush border Poorly absorbed from the small intestine Fermented by the microflora in the large intestine
< 9	Polysaccharides		
	Starches		Hydrolyzed by gastric and pancreatic amylases and by brush border enzymes
		Amylose Amylopectin	
	Nonstarch polysaccharides		Not hydrolyzed in small intestine, fermented by the microflora in the large intestine
	Cell wall components	Cellulose Noncellulosic polysaccharides	
	Polysaccharide food additives	Gums Mucilages Algal Polysaccharides Modified starches, pectins, etc.	

Table 1 also illustrates the task facing the analyst asked to measure all the carbohydrates in foods, even where polysaccharides are grouped as suggested above. It also demonstrates the need for the analyst to consider the food matrix under examination because this will often simplify the procedures needed.

During the 19th century, food chemistry was in the early stages of development and the inadequacy of understanding carbohydrate chemistry led to the adoption of the so-called proximate method of food analysis. In this method, carbohydrates were determined by difference. Moisture, protein (from total nitrogen content), fat (by simple lipid solvent extraction) and ash (by incineration) were measured directly and deducted from 100% to give a value for carbohydrate. Although this method has effectively been superseded by the growth in knowledge and the developments in analytical chemistry, it is still in wide use and much food composition data are still based on this type of analysis.

Analytical Characteristics of Carbohydrates

The carbohydrates as a class, despite the wide range of species occurring in foods, present the analyst with a limited range of reactions that can be used analytically. As a class they are all polyhydric alcohols which undergo substitution reactions. These are very valuable in structural analysis but less useful analytically.

Physical Methods

These were developed within the sugar-processing industry, where analysts were primarily concerned with the analysis of solutions of sucrose and the products of its hydrolysis, glucose and fructose. The simplest methods were based on the very high solubility of sucrose in water and involved the measurement of specific gravity or refractive indices of the solutions which are very nearly linearly related to sucrose concentration. These methods are still in use for the analysis of syrups. The second method made use of the optical activity of the sugars: polarimetry provides a quick and reliable index of the proportions of sucrose and glucose and fructose in an aqueous solution.

Chemical Methods

The monosaccharides have aldehyde or keto functional groups which have powerful reducing properties by virtue of their ability to form enediol arrangements. These form the basis of a series of reductiometric procedures which are widely used. The Munson-Walker and Lane and Eynon procedures have formed the basic techniques for analysis in the sugar industry. These reductiometric techniques are highly empirical and require close attention to analytical protocols for consistent and accurate performance. Reduction of ferricyanide has also provided the basis for colorimetric procedures using this type of technique.

Nonreducing disaccharides such as sucrose require hydrolysis before the reducing sugar methods can be used and polysaccharides such as starch have also been measured as monosaccharides after hydrolysis. Each monosaccharide has, however, slightly different reducing strengths and the analyst concerned with mixtures of reducing sugars has to make adjustments for the composition of the mixture.

The polyhydric structure of the carbohydrate also renders them susceptible to dehydration and the formation of furan structures in strong acids, and these react with a number of chromogenic phenolic substances to form colors which are suitable for colorimetric measurement. These reactions tend to be relatively unspecific and careful control of conditions is required for satisfactory measurements.

Biochemical Methods

The sugars (monosaccharides and disaccharides) are involved in a series of biochemical reactions with highly specific enzyme systems. Most of these enzymes are available in a highly purified state and can be coupled with NADH or NADPH to provide a very specific and highly sensitive method for the measurement of specific sugars in complex mixtures. The clinical measurement of glucose uses an enzymatic method coupled with a colorimetric method. However, in many situations the development of chromatographic procedures has displaced these elegant methods.

Chromatographic Methods

The difficulty of precisely quantifying mixtures of sugars provided a major stimulus to the search of separation procedures. Paper chromatography and later thin-layer chromatography provided the first effective qualitative and semiquantitative analytical methods. Conventional ion-exchange techniques were difficult to apply, although some success was achieved in separating the weak borate complexes. The nonvolatile nature of sugars slowed the application of gas liquid chromatography until trimethyl silyl derivatives were developed which gave good separations. At present most analyses are based on the use of alditol acetate derivatives where sugars are reduced to alditols with borohydride and then acetylated.

Gas liquid chromatographic techniques are widely used, particularly in analysis of polysaccharides in plant cell walls and in the measurement of NSP. The derivatization stages are time-consuming and not

completely specific because fructose gives the same alditol as glucose (and partially as mannose).

High-performance liquid chromatography (HPLC) removes the need for the preparation of derivatives and provides separation of all the monosaccharides, disaccharides, and oligosaccharides. Initially the carbohydrate columns were rather erratic in performance but now consistent separations are usual. Detection systems initially used the refractive index but these were not very sensitive and pulsed amperometric systems have been introduced which have overcome this limitation.

Extraction and Clean-up

As can be seen from the above, a selection of methods for the end analyses of carbohydrates are used, most of which are designed to measure monosaccharides. Virtually all analytical methods for carbohydrate involve the preparation of an extract free from interfering substances.

The extraction and clean-up procedures are usually food-specific because this simplifies the procedure that needs to be used. This is especially important when the analysis is being carried out for quality control purposes, but applies across the whole range of purposes for which carbohydrates are measured.

Thus, in the sugar-processing industry the samples are often in aqueous solution and the interfering substances can be removed using precipitation with metal salts. Analogously, milks and many milk products only require the removal of proteins to provide a suitable extract for sugar analysis.

Foods which contain more complex mixtures of carbohydrates or when analyses of the various carbohydrate species are required need more detailed extraction and fractionation schemes. A large number of schemes have been, and are being, used, especially where the individual polysaccharides are being separated and measured but this article will focus on the principles of the most widely used approaches.

The major stages are illustrated in Figure 1.

Sample preparation As with all analyses, this is a critical stage. Sugars are sensitive to heat and polymers containing furanoside linkages are susceptible to acid hydrolysis under mild conditions. This means that the preparation of samples needs to minimize any possible changes due to heat and acid. Very acid samples should be neutralized before drying or extraction; freeze-drying is a suitably mild drying procedure. The sample also needs to be finely divided and care must be taken to prevent fractionation of the sample into different sizes of particles.

High-fat (>10%) foods need to be extracted with a lipid solvent to assist in later extraction procedures.

Extraction with Aqueous Alcohol The most common extractant is hot 80%v/v ethanol, but isopropanol has also been used, as has 85%v/v methanol. This provides a very good separation of polysaccharides from the other groups, but some lower polysaccharides such as inulin are incompletely precipitated by ethanol of this strength and some arabinans are also partially soluble.

Use of Supernatant This is usually suitable for many analytical procedures after dilution or removal of the ethanol by rotary evaporation. Reductiometric methods may call for clean-up with metal salts and colorimetric methods may require removal of strongly colored pigments using similar procedures. Dilution prior to chromatographic separation using HPLC is the preferred method for analyzing mixture of sugars and oligosaccharides. Table 2 presents the range of major options, which depend on the required objectives of the analysis.

Analysis of Total Carbohydrate and Sugars The colorimetric reactions with phenol/sulfuric and anthrone/sulfuric are general-purpose methods which give reasonable estimates of total sugars present and which, with suitable standard mixtures, can give acceptable accuracy.

Many of the classical reductiometric techniques depend for their accuracy on the existence of standard tables giving the reduction equivalents for a wide range of mixtures of glucose and fructose and of invert sugar and sucrose.

The specific enzyme procedures are based around reactions of the type illustrated below:

$$\text{Glucose} + \text{ATP} \xrightarrow{\text{Hexokinase}} \text{Glucose-6-phosphate} + \text{ADP}$$

$$\text{Glucose-6-phosphate} + \text{NADP} \xrightarrow{\text{Glucose-6-phosphate dehydrogenase}} \text{6-phosphogluconate} + \text{NADPH}$$

The amount of NADPH formed is proportional to the glucose present and can be measured by measuring in the ultraviolet range at 340 and 360 nm. Any fructose-6-phosphate produced by the hexokinase is converted to glucose-6-phosphate with phosphoglucose isomerase, which leads to formation of more NADPH, permitting the analysis of a mixture. Table 3 lists the enzymes used for the different carbohydrate species.

Oligosaccharides The ethanolic extracts contain the oligosaccharides and some polysaccharides with lower degrees of polymerization such as inulin. These will measure as total sugars with the colorimetric reactions in strong acid but the choice of another

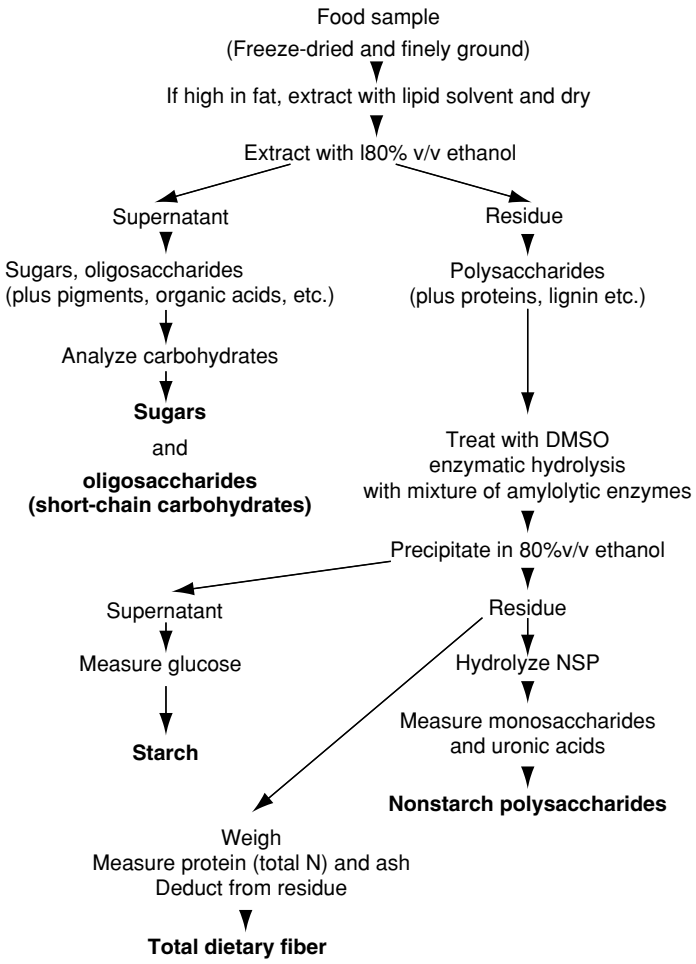

Figure 1 Typical scheme of fractionation used in the analysis of the major carbohydrates in foods. DMSO, dimethyl sulfoxide; NSP, nonstarch polysaccharides.

Table 2 Options for the analysis of the ethanolic extract[a]

Analytical objective	*Type of method*	*Choice of method*	*Limitations*
Total sugars	Colorimetric	Phenol//sulfuric	Color yield depends on sugars present
Total sugars	Reductiometric/colorimetric	Ferricyanide	Color yield depends on sugars present
Hexoses	Colorimetric	Anthrone	Color yield depends on sugars present
Reducing sugars	Reductiometric/volumetric	Lane and Eynon; Munson-Walker Copper salts in alkaline solution	Reduction depends on ratios of sugars present
Sucrose in invert sugar solutions	Reductiometric before and after inversion	Copper salts in alkaline solution	Reduction depends on ratio of invert sugar to sucrose
Individual sugars in mixture	Specific enzymatic procedures	Specific enzymes coupled with NADH and ultraviolet measurements	Availability of enzymes
Individual sugars in mixture	HPLC	Using either refractive index or pulsed amperometric detectors	Capital cost of equipment

[a]In most cases, removal of ethanol before analysis is necessary.
HPLC, high-performance liquid chromatography.

Table 3 Enzymes used in carbohydrate analyses

Enzyme class	*Substrate*	*EC number*	*Source*
Oxidoreductases			
Lactate dehydrogenase	Lactic acid	1.1.1.27	Pig muscle
6-Phosphogluconate dehydrogenase	6-Phosphogluconate	1.1.1.44	*Saccharomyces cerevisiae*
β-galactose dehydrogenase	D-galactose, L-arabinose	1.1.1.48	*Pseudomonas fluorescens*
Glucose-6-phosphate dehydrogenase	D-glucose 6-phosphate	1.1.1.49	*S. cerevisiae*
Glucose oxidase	D-glucose	1.1.34	*Aspergillus niger*
Diaphorase	NADH	1.6.4.3	Pig heart
Catalase	Hydrogen peroxide	1.11.1.6	Beef liver
Kinases			
Hexokinase	D-hexoses	2.7.1.1	*S. cerevisiae*
Gluconate kinase	D-gluconic acid	2.7.1.12	*Escherichia coli*
Hydrolases			
Amyloglucosidase	α-Glucans	3.2.1.3	*A. niger*
α-Amylase	α-Glucans	3.2.1.1	Pig pancreas
Pullulanase	1–6 α-Glucoside links	3.2.1.41	*Bacillus acidopullulyticus*
Thermamyl	α-Glucans (heat-labile)	3.2.1.1	*B. licheniformis*
Invertase	Terminal β-fructoside links	3.2.1.26	*S. cerevisiae*
Isomerases			
Phosphoglucose isomerase	D-glucose 6-phosphate	5.3.1.9	*S. cerevisiae*

Modified from Determination. In: Chaplin MF (1998) Encyclopedia of Food Science and Nutrition. 1st edition, pp. 684–691.

Table 4 Major carbohydrates in the nonstarch polysaccharides

Major group	*Components*	*Structural features*	*Distribution in foods*
Plant cell wall			
Cellulosic	Cellulose	Linear β-glucans	All cell walls
Noncellulosic			
	Pectic substances	Rhamnogalacturonans	Mainly in fruits and vegetables
		Arabinogalactans	Fruits and vegetables
	Hemicelluloses	Arabinoxylans	Cereals
		Glucuronoarabinoxylans	Cereals
		Glucuronoxylans	Fruits and vegetables
		Xyloglucans	Fruits and vegetables
		β-Glucans	Cereals

method must be based on the types present. The malto series will usually be hydrolyzed by the combination of amylase enzymes used to hydrolyze starch and, if no other oligosaccharides are present, may be hydrolyzed with dilute acid and the free sugars measured. Acid hydrolysis cannot be used if fructose-containing polymers are present because fructose is destroyed by acid; fructanase enzymes are available for selective hydrolysis. HPLC is the preferred method for the separation and analysis of oligosaccharides.

Treatment of the Residue: Hydrolysis of Starch The residue in most foods is a mixture of starch, NSP, proteins, the noncarbohydrate lignin, heat artifacts formed during the heat processing of foods and some inorganic matter.

Treatment with dimethyl sulfoxide (DMSO) renders any retrograded amylose soluble, and the starch can be completely hydrolyzed using a mixture of amylases. A heat-labile amylase is frequently used to give rapid solubilization of the starch followed by a mixture of pullulanase and amyloglucosidase or α-amylase to hydrolyze starch to glucose which can then be measured after precipitating the unchanged NSP with 80% v/v ethanol. The use of a glucose-specific method such as the colorimetric glucose oxidase method is preferred for analysis of glucose released by starch hydrolysis, but in practice any approach can be adopted.

Analysis of Nonstarch Polysaccharides NSPs form a complex mixture (Tables 4 and 5) and separation of the various types of polysaccharide is usually only carried out by those concerned with the study of the structure of the plant cell wall. A typical fractionation and analysis scheme is illustrated in outline in Figure 2.

Table 5 Polysaccharide food ingredients in nonstarch polysaccharides

Major group	*Components*	*Structural features*	*Use in foods*
Gums	Gum arabic	Arabinogalactans	Confectionery
	Guar, locust bean	Galactomannans	Soups as thickeners
Algal polysaccharides			
	Agar	Galactoanhydrogalactans	Powerful gelling agent
	Alginates	Guluronomannuronans	Powerful gelling agents
	Carageenans	Several different sulfated galactoanhydrogalactans	Interact with proteins to control physical properties
Modified starches		Phosphated and cross-linked, esters and ethers	Used to control starch retrogradation and as gelling agents
Cell wall extracts	Pectins		Gelling agent in jams
	Modified pectins (amidated)		Gelling agent
	Cellulose, esters, and ethers		

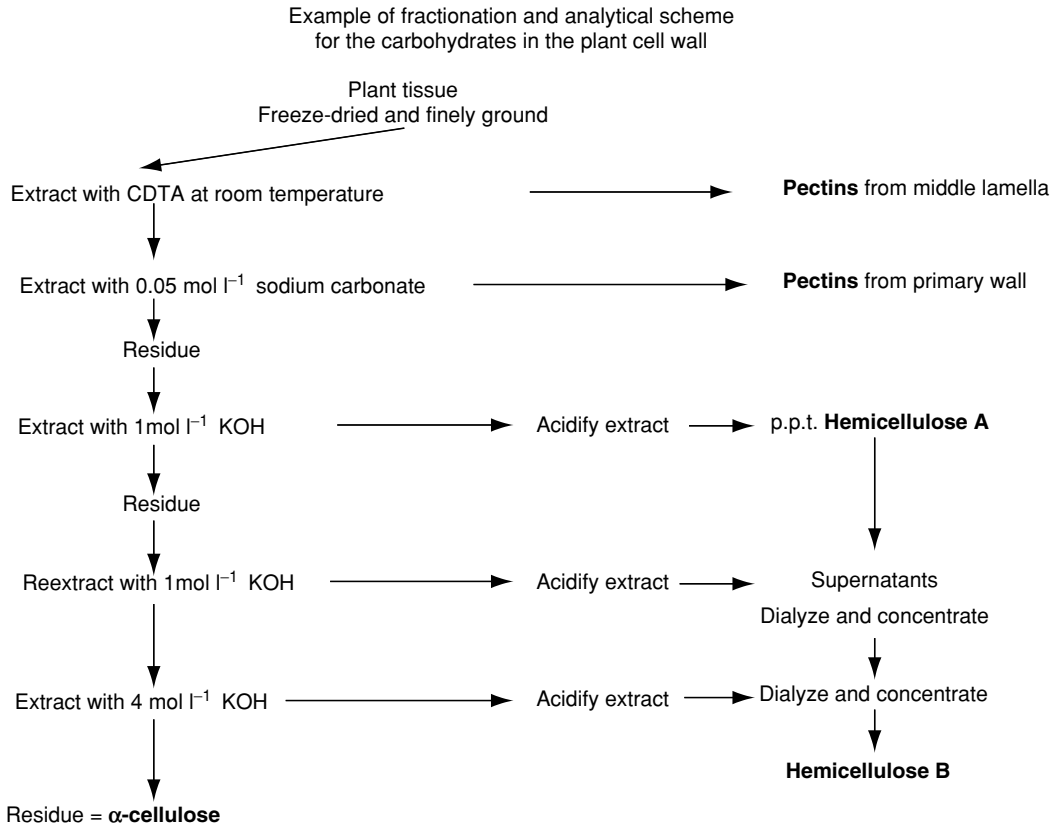

Figure 2 Typical scheme for fractionating and analyzing the carbohydrates in the plant cell wall. CDTA, cyclohexane diamine tetraacetate.

The polysaccharides in the extracts are concentrated and cleaned up by extensive dialysis and may occasionally be measured after freeze-drying the extract. More commonly, the extracts are further fractionated by size exclusion chromatography to separate the noncellulosic polysaccharides (pectins and hemicelluloses) into fractions based on molecular size. They may then be further characterized by methylation and hydrolysis for structural studies using chromatographic systems linked to mass spectrography. Monosaccharide composition is measured as described for the less detailed NSP analyses described below.

Most food analyses are restricted to the analysis of total NSP, although measurement of the cellulosic and noncellulosic fraction is often of value

nutritionally. Some procedures include treatment with a protease alongside the removal of starch, to reduce protein contamination before precipitating the NSP. This provides a procedure for measuring nonstarch components gravimetrically, and forms the basis of the official method of the Association of Official Analytical Chemists (AOAC) for measuring total dietary fiber. The residue contains the noncarbohydrate lignin and any heat-induced artifacts in processed foods. There is also some residual protein and inorganic matter and in the official method these are measured and deducted from the residue weight.

Measurement of the insoluble and soluble NSP fractions was until recently seen to be of nutritional value but this division is not now considered useful. This is because solubility is highly method-dependent in the choice of aqueous buffer used and the conditions under which it is measured and these have yet to be related to solubility under physiological conditions.

Hydrolysis of NSP Complete hydrolysis to the component monosaccharides is achieved using initially solubilization in $12\,\text{mol}\,l^{-1}$ H_2SO_4 followed by dilution to $2\,\text{mol}\,l^{-1}$ and heating to 100 °C for 1 h. The conditions do not release uronic acids which require enzymatic hydrolysis of the polysaccharides. In the Englyst method the uronic acids were measured colorimetrically and the monosaccharides by gas liquid chromatography (or HPLC). The noncellulosic polysaccharides are hydrolyzed in $2\,\text{mol}\,l^{-1}$ H_2SO_4 by heating at 100 °C for 1 h.

The monosaccharide constituents in both types of hydrolysates are measured after neutralization. For gas liquid chromatography the monosaccharides are reduced with sodium borohydride and then acetylated. It is necessary to use an internal standard such as allose to control these stages. Gas liquid chromatography is performed on a wide-bore capillary column with a flame ionization detector.

HPLC is controlled with an internal standard using a pulsed amperometric detector. Standard mixtures of the monosaccharides are analyzed in the same way and used to derive correction factors. Colorimetric procedures using the ferricyanide reaction may also be used to measure the total monosaccharides released by the $12\,\text{mol}\,l^{-1}$ H_2SO_4 hydrolysis. The standard mixture should reflect the ratios of the components in the samples under analysis to allow for differences in reduction equivalents.

Polyuronans which form part of the NSP are not completely hydrolyzed under these conditions and for this reason the uronic acids are measured separately colorimetrically in the $12\,\text{mol}\,l^{-1}$ hydrolysates using dimethyl phenol in $2\,\text{mol}\,l^{-1}$ H_2SO_4.

Modes of Expression

The carbohydrates in foods may be expressed in a number of ways. Total carbohydrate is often expressed as the value obtained by difference. This mode of expression is still widely used in food labeling. In the USA and the European Union (EU) this excludes dietary fiber. In the USA this implies total dietary fiber, as measured by the official AOAC method. In the EU, agreement on the procedure for dietary fiber has been slow to be reached and may be the AOAC value or a NSP value.

Total carbohydrate may also be expressed as the sum of the directly measured carbohydrates as they are present in the food, i.e., starch as starch or sucrose as sucrose. Once again, dietary fiber values are excluded. In the UK and some other nutritional databases, the carbohydrates are expressed as the monosaccharides. This can cause confusion if the mode of expression is not recognized because 100 g starch gives 110 g glucose and 100 g of disaccharide give 105 g monosaccharides.

Current international opinion on the measurement of carbohydrates in foods is that, wherever possible, direct and specific analytical methods should be preferred.

See also: **Dietary Fiber**: Properties and Sources; Determination; Physiological Effects; Effects of Fiber on Absorption; Bran; **Starch**: Structure, Properties, and Determination; Sources and Processing; Functional Properties; Modified Starches; Resistant Starch; **Sugar**: Sugarcane; Sugarbeet; Palms and Maples; Refining of Sugarbeet and Sugarcane

Further Reading

Anonymous (1989) *Methods of Biochemical Analysis*. Mannheim: Boehringer Mannheim.

Chaplin MF and Kennedy JF (eds) (1986) *Carbohydrate Analysis: A Practical Approach*. Oxford: IRL Press.

FAO/WHO Joint Expert Committee (1998) *Food and Nutrition Paper 66*. Rome: Food and Agricultural Organisation.

Nielson SS (ed.) (1998) *Food Analysis*, 2nd edn. Gaithersburg, MD: Aspen.

Southgate DAT (1991) *Determination of Food Carbohydrates*, 2nd edn. London: Elsevier Science.

Southgate DAT (1995) *Dietary Fibre Analysis*. Cambridge: Royal Society of Chemistry.

Sullivan DM and Carpenter DE (eds) (1993) *Methods of Analysis for Nutritional Labeling*. Arlington, VA: Association of Official Analytical Chemists International.

Sensory Properties

C S Setser, Kansas State University, Manhattan, KS, USA
G D Brannan, Customized Sensory Services, Wyoming, OH, USA

Background

The first sensory response expected of low-molecular-weight carbohydrates is sweetness. More than 100 substances are sweet and chemically identified as sugars, or nutritive carbohydrate sweeteners. Many other unrelated substances of diverse molecular geometry elicit the sweet taste, including some aliphatic and aromatic organic compounds, amino acids, and certain inorganic salts. However, not all sugars are equally sweet, and, moreover, the sensory properties of foods that are influenced by carbohydrates extend beyond sweetness. Flavors, in addition to the sweet taste, develop from browning reaction products. Color is another sensory property developed with browning reactions. Textural attributes are influenced in numerous ways: the low-molecular-weight carbohydrates contribute body and viscosity or interact with other components, including water and the high-molecular-weight carbohydrates, to influence sensory properties of food products; and the starches and gums contribute thickening and gel structure alone and by interactions with each other to alter other mechanical and geometrical texture characteristics.

Sweetness of Sugars

Stereochemical Similarities of Sweet Compounds

The common structural feature associated with all compounds possessing the sweet taste is a glycophore capable of concerted hydrogen bonding. A proton-donating group, often called the AH group, and a B group serve as a proton or hydrogen bond acceptor, respectively. The primary AH,B system of the common sugars is the α-glycol grouping, the portion farthest from the reactive, anomeric center; for example, the 3,4 α-glycol system of glucopyranose structures is responsible for its sweet taste. High-potency sweetness and bitterness can occur if a third component (X, or γ) functions as a 'dispersion' or hydrophobic site, and the electron-withdrawing group enhances the activity of the AH,B dipole. The sweet glycophores have AH–B distances of approximately 2.6 angstroms (Å), AH–X distances of 3.5 Å, and B–X distances of 5.5 Å.

A criticism of the AH,B system is that not all compounds that possess the AH,B system taste sweet, and some are actually bitter or tasteless. Symmetry principles help resolve this criticism. The charge on the dipole of the sweet substance must be bilaterally symmetrical, i.e., of opposite sign but nearly equal, for sweetness to be perceived.

Another issue is related to the enantiomers. Initially, the belief was that the D-sugars (and amino acids, for that matter) tasted sweet, but L-antipods did not. If the AH,B system were correct, both D- and L-sugars would be sweet. Later, it was learned that some L-sugars do taste sweet. The AH,B system of a stimulating compound would combine reversibly by intermolecular, antiparallel hydrogen bonding with a commensurate AH,B system on a proteinaceous receptor molecule of the tongue (Figure 1). The tripartite two-dimensional attributes of the faces of tastants determine the tastes of enantiomers. Such a fixed AH, B system would approach the receptor from only one direction, and L-forms of amino acids could be sterically restricted by side-chains from interaction with the receptor. This issue has not yet been resolved.

The AH,B model for the stimulus structure of sweet compounds suggests that only one receptor for sweetness is critical. However, current theories of sweet taste perception make it unlikely that a single receptor molecule can accommodate all sweet stimuli. Responses vary among species, individuals, and cells. Receptor sites for sugars also are not identical from one taste bud to the next.

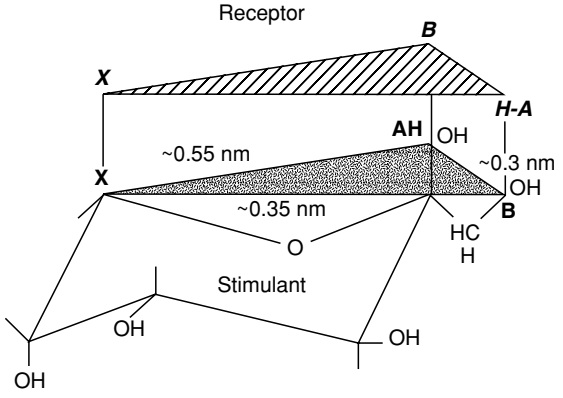

Figure 1 Representation of sweetness stimulant–receptor interaction for the AH-B-X theory using D-glucose as an example. The location of the AH-B-X units on the D-glucose template is shown in relation to a corresponding receptor site. Reproduced from Carbohydrates: Sensory Properties, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Sweet Taste Transduction

Perception begins with the contact of the sweet stimulus and a surface receptor molecule on the taste cell that leads to cellular changes via secondary messengers, which are translated into ion-channel events. Subsequently, the modified ion channels cause changes in cell polarization, resulting in the release of neurotransmitters to adjacent neurons and eventually an action potential. The theory of taste transduction, the process of translating sensory information into a signal useful to the nervous system indicates that high-intensity sweeteners, especially polyols, react with a seven-transmembrane receptor protein that is associated with a G-protein inside the cell. Several mechanisms on multiple receptor sites are possible; some, for example, use cyclic adenosine monophosphate (cAMP) as a second messenger, and others use inositol triphosphate (IP3). High-potency compounds apparently disrupt the transduction process at the G-protein. The number of sites involved and the effectiveness of the interaction have been related to the potency. Both sweet and bitter receptors share some of the same transduction components, such as IP3. These theories could help explain the close association between sweet and bitter tastes of some substances.

Sweeteners in Products

Sweetener synergy is critically important to the food processor. Synergistic effects are inferred if the sweetness of the mixture is greater than the sum of sweetness of the individual components. Synergism might result from solute–solute, solute–water, and solute–receptor interactions. The effects with any given sugar are not universal. Fructose, for example, shares a synergistic relationship with glucose and with saccharin, but not with other carbohydrate sweeteners. Recent evidence suggests that synergy depends on the compatibility of component hydration and the influence on water's structure. If components with identical types of hydration are mixed, increased water mobility in the proximity of the sweetener and a reduction in volume of the hydrated solute molecule result in synergy. Reduced water mobility in the medium has been observed with sweetness suppression of mixtures (sucrose–aspartame), and an increased water mobility led to enhancement (synergism) of sweet taste (sucrose–cyclamate and maltitol–acesulfame K mixtures), but no effect on water mobility and no appreciable synergy has been observed with alitame plus either sucrose or maltitol or with acesulfame K with sucrose.

Intensity of Sweetness

Sweetness generally decreases with the molecular weight of the sugar, although the reason is unknown. Perhaps, only one sugar residue in each oligosaccharide binds to the taste cell receptors. Lack of sweetness in some compounds is caused by steric interference of one or both of the axial hydroxyl groups, thus preventing binding to the taste cell receptor.

Food processors are interested in relative sweetness scores. Such numerical values attempt to relate overall sweetness of one substance to that of another substance. The value is defined as the 'ratio of concentration of substances matching in sweetness,' and sucrose is used as the reference (Table 1). The relative sweetness score has limited use, because it fails to recognize different solution properties of dissimilar substances and quality differences that occur because of a sweetener's multidimensional nature (onset, duration, intensity) is integrated into a single value. If the sweetener is in solution or in its crystalline form, the age of the test medium and the degree to which the sugar has mutarotated are further factors influencing relative sweetness.

Table 1 Relative sweetness of carbohydrate sweeteners

Sweetener	*Approximate sweetness*[a]
Bulk solids	
Fructose	1.2–1.7
Glucose	0.7–0.8
Invert sugar	1.0
Isomalt (Palatnit)	0.5–0.6
Lactitol	0.3
Lactose	0.2–0.4
Maltose	0.4–0.5
Mannitol	0.4–0.7
Mannose	0.6
Sorbitol	0.5–0.6
Xylitol	0.9–1.2
Xylose	0.7
Bulk syrups	
Corn syrup, unmixed, acid-converted	
30 DE[b]	0.3–0.35
36 DE	0.35–0.4
42 DE	0.45–0.5
54 DE	0.5–0.55
62 DE	0.6–0.7
High-fructose corn syrup	
42%	0.9–1.0
55%	1.0–1.1
90%	1.2–1.6
Hydrogenated glucose syrup (Lycasin)	0.4–0.75
Chloroderivatives of sucrose (Sucralose)	5–2000
Stevioside	300

[a]Sucrose = 1.0. Sweetness is a relative measurement dependent upon many external factors including concentration, temperature, pH, structural configuration, and degree of hydrolysis.
[b]DE, dextrose equivalent.

Concentration effects between 2 and 16% are approximately linear for sugars and sugar alcohols relative to sucrose standards. However, the psychophysical laws that govern responses indicate that increasingly larger stimulus increments are required with high concentrations to elicit measurable differences. An additional factor operating with concentration increases its adaptation. If a person is exposed repeatedly to a stimulus, the perceived intensity is reduced. The relative sweetness of sucralose, a potent sweetener, decreases with increases in the concentration at high concentrations. For most high-potency, noncarbohydrate sweeteners, the concentration response plots are hyperbolic. Concentration increases can negate or alter intensity differences or induce unpleasantness. The different behavior of polyols from high-potency sweeteners supports the tentative conclusion that at least two routes to receptor-cell activation exist.

Temporal effects and the sweetness response Onset time and duration of sweetness influence the acceptability of a sweetener. Consumers detect sucrose sweetness within 1 s, and that sweetness lasts about 30 s. Any sweetening agent that has a delayed initial sweetness or a sharp or prolonged sweetness sensation will taste unusual to many persons. A localized concentration of stimulus molecules at receptor sites is believed to govern the persistence or duration of sweetness. The lipophilic site could be responsible for directing potent sweeteners into localized concentrations at a nonspecific area of the cell membrane and account for persistence of sweetness. Thus, the physical length of the queue could determine the duration of sweetness by the length of time the receptor is supplied with stimuli.

Temperature effects and the sweetness response Temperature influences the sweet response, and not all sweeteners respond to temperature similarly. Heat can break intramolecular bonds of sweeteners, which could free more hydroxyl groups to participate in the AH,B system. Changes in structural form with bond breaking can result in isomers that do not have a sweetness equal to the original sweetener. An increase in temperature results in increased sucrose intensity, but decreased persistence. Temperature is believed by some to affect the approach of the stimulus molecules to the receptor site prior to the transduction process. However, others believe that the optimum temperature for detection of sweetness is near the temperature of the tongue, which is explained by maximal electrophysical responses at that point.

Synergism of Sugars with Other Food Components

Interactions of Components and Effects on Sweetness

The presence and concentration of other tastes, sweeteners, and flavors; the structural matrix, whether a solid, liquid, or gas; viscosity and solvent polarity of the medium; extreme temperatures of processing, storage, and preparation; and the microbes and enzymes present all influence and modify perception of sweetness from the sugars. Typically, sweet compounds are embedded in a complex matrix with bitter, sweet, sour, and salty tastes, and the flavor nuances of other ingredients. The 'taste modifiers' alter sweetness quality, increase or reduce intensity, or mask aftertaste probably by exerting a physical or chemical effect on the stimulus rather than the taste receptor. Sensory acuity to carbohydrate sweetness diminishes in the presence of other tastes, but the degree of suppression depends on the nature of the secondary agent, the concentration, and the intensity of taste.

Sweetness seems to be enhanced at low (< 0.4%) sodium chloride (NaCl) concentrations, but perceived sweetness declines if the concentration of NaCl is high. Sweetness of sucrose, fructose, and glucose is affected differently by various acids and the relative concentrations of each of the components. Sweetener–acid interactions, as well as interactions with fruit flavors, are important in the formulation of several beverages. Fruit flavorings alter the sweetness perception of aqueous systems, generally enhancing the sweetness of the carbohydrate sweetener. The sugars decrease the perception of bitterness in foods and beverages.

The relative initial and maximum sweetness of sugars is different in food products than in model systems. The texture and physical properties of a food affect its taste, because texture partially controls the amount and the rate which tastants reach the taste buds. The thresholds for the four basic tastes are higher in food products than in aqueous solutions. Sweetness is maximal when little or no interference from physical behavior of the taste medium exists. Viscous and oil solvents exhibit an interfering behavior that can raise the threshold and decrease the perceived sweetness intensity.

Sugar–Protein Interactions

Flavors and colors of foods are altered by the nonenzymatic browning of carbohydrates. Thermal decomposition of simple and complex carbohydrates usually occurs in combination with inorganic or

organic catalysts and results in caramelization, the key pathway for the formation of flavors associated with molasses, maple syrup, and caramel flavorings and colorings. The carbonyl-amine reaction involves a thermally induced interaction between an amino acid and a reducing sugar. Many pathways and reaction schemes follow ring opening and enolization of the sugars and their interactions with nitrogenous amino groups. The unique color and flavor compounds produced in roasted coffee, nuts, meats, chocolate, maple syrup, and bakery products result, in large part, from the carbonyl-amine reaction. Products of the carbonyl-amine reaction and interactions of the caramelization and carbonyl-amine reactions include pyrroles, pyridines, imidazoles, pyrazines, furfurals, furanones, oxazoles, and thiazoles, and they can form numerous brown to black polymers. Maltol and isomaltol, with their fragrant, caramel-like aromas, enhance both the flavor and sweetness of many foods. In micowaved and extruded foods, the time–temperature–moisture conditions result in a lack of flavor and color, because only low levels of the browning reaction products are formed. Strecker degradation and the fragmentation of carbonyl–amines are well-defined pathways that also can result in many off-flavor compounds in foods. During the final steps of the Strecker degradation reaction, formaldehydes and pyrazines evolve from amino ketone fragments of the reducing sugar. Food processors should optimize desirable caramel-type aromas but must minimize the burnt, bitter, and acrid notes that can be produced.

Sugar Interactions and Textural Changes in Foods

Carbohydrate sweeteners' colligative properties, which depend on the concentration of the solute, influence the body/viscosity of liquid and solid foods, and alter the texture of other foods by the effects on freezing point, boiling point, osmotic pressure, and vapor pressure. Increasing the number of sweetener molecules increases the boiling point and osmotic pressure, and decreases the freezing point and vapor pressure.

In addition to solute concentration, the viscosity of carbohydrate-sweetened solutions is governed by the temperature. With an increased temperature, molecular motion is increased, reducing the friction among the molecules and resulting in a decreased resistance to flow. The viscosities of most carbohydrate-sweeteners in solution are similar with equivalent molar weights except for glucose syrups. A large percentage of high-molecular-weight polysaccharides increase the relative viscosity of glucose syrups. Consequently, these syrups impart more cohesiveness, body, and adhesiveness to a food system than other types of sweeteners. The viscosity of carbohydrate sweetened solutions also is governed by solute concentration and temperature. With an increased temperature, molecular motion is increased, reducing the friction among the molecules and resulting in a decreased resistance to flow.

Low-dextrose equivalent syrups are viscous and increase the chewiness of products such as caramels and cookies more than an equal concentration of a higher-conversion, lower-viscosity syrup. The low-viscosity syrups decrease the viscosity of a candy mass of a given sucrose composition by decreasing air retention. The candy is brittle, and good flavor release in the mouth is promoted. The texture of confections is dependent upon dissolution and recrystallization of the sugar. If a sugar solution cools slowly and is not disturbed, large crystals are formed, as in rock candy. Rapid cooling with agitation results in fine crystals, as in creams.

The sharp, crystalline edges of sugars contribute to the aerated nature of chemically leavened bakery products by helping disperse the lipid portion of a batter in the initial creaming stages of multistage mixed batters and doughs. The creaming of the shortening allows the formation of many air cells, thus increasing the volume and tenderness. Sugars, particularly glucose, are efficient fermentation media, providing aeration in some bakery products. Sugars further tenderize bakery products by controlling starch hydration and by dispersing protein and starch molecules. Starch and protein molecules are separated by the sugars, as well as by lipids, and are prevented from forming a continuous mass. The structure remains flexible and pliable enough to allow for maximum expansion during leavening and before thermal setting. Excessive sugar can increase the fluidity of the batter, increase the coagulation temperature of the egg proteins, and increase the temperature at which starch gelatinization occurs to the extent that the structure is too weak to support its weight. The batter structure rises before it is thermal set, but then it collapses.

Regulating the amount and type of sugar controls cookie spread and surface cracking. Increasing the amount of sugar with a given amount of water generates more solution than is available with water alone. Small crystals increase cookie spread, because fine crystals dissolve more readily, generating more solution at a given time and temperature than do the coarser particles. Consequently, gelatinization and thermal set times are delayed, the dough is more fluid, the cookie spreads farther before the structure sets, and it is crisper. When cookies contain high levels of sucrose, a hard, sweet cookie with 'snap,' or a crispness and crunchiness, and surface cracks

results, because, after cooling, the sugar syrup recrystallizes to form an amorphous glass. Molasses, honey, glucose syrups, and high-fructose corn syrup inhibit crystallization and result in softer cookies that are chewy, rather than crisp.

Properties of Polysaccharides in Relation to Food Product Structure

Polysaccharides affect the sensory properties, particularly the texture, of food products by retarding the flow as well as by modifying the gelling characteristics. Nonstructured, amorphous regions of polysaccharides are highly hydratable and bind large quantities of water. In contrast, areas that contain a large portion of straight-chained, crystalline regions, such as retrograded starch gels, exclude water. Most natural and synthetic hydrocolloids, or gums, display pseudoplastic flow properties, that is the viscosity of the solution decreases as the shear rate increases. With high shear rates, viscosity decreases because the particles are oriented parallel to the shear field. In general terms, viscosity also decreases as the temperature increases.

Sensory Properties and Rheological Character

The predominant sensory property of many gum solutions is slimy characterized by a thick mouth coating and by being difficult to swallow material. Dispersions that approach Newtonian flow behavior are associated with a high degree of sliminess, and gums in solution that depart considerably from Newtonian flow and have a high degree of shear-thinning are less slimy in the mouth. Other sensory attributes used to describe gum solutions are adhesive, starchy, gummy, astringent, slippery, and oily. 'Starchy' is a flavor related to an undercooked product, but the other parameters are related to mouth feel and mechanical textural properties. Normally, blandness is the desired flavor quality for carbohydrate thickeners in food products.

Pseudoplastic behavior is associated with a less slimy mouth feel because the viscosity of the fluid decreases at the shear rate encountered in the mouth (approximately $50\,s^{-1}$). High-viscosity seed and exudate gums – guar, tragacanth, and locust bean – are more pseudoplastic than alginates and cellulose ethers, whereas gum arabic has a low viscosity and almost Newtonian flow properties up to a concentration of at least 40%. Dispersions of microcrystalline cellulose, a synthetic gum, show both thixotropic and plastic behavior. The behavior that dominates depends upon the concentration and shear history. Starches can be modified to alter the consistency; cross-linking contributes short, stringless consistency in waxy and tapioca starches. The unmodified starch dispersions will have a stringy, cohesive consistency. High levels of highly cross-linked waxy corn starch will increase the chewiness in a product.

Gel Properties

Several polysaccharides provide partial gel structure in some food products and total structure in others. High-methoxyl pectins in fruit jams and jellies; low-methoxyl pectins for dietetic gels; starches in puddings, agar in meat products; carrageenan for kosher fruit gels and milk puddings; and alginate in pie fillings, reformed vegetables, fruits, and meat chunks are all examples.

Gels can be brittle, rigid, springy, firm, soft, spreadable, cutable, rubbery, smooth, or grainy, depending upon the concentration and the degree of interactions of the polymers. Gelation balances polymer–polymer and polymer–solvent interactions to form a tertiary network or matrix; a gel is an intermediate state between a solution and a precipitate. Enhanced interactions among polymer molecules result as their solubility decreases and generally lead to firm, rubbery, springy, rigid gels. Gels tend to become more brittle as the concentration increases, but also faster rates of setting and decreased uniformity of matrices result in increased brittleness. A crisp or crunchy mouth feel for snack foods is obtained with high levels of amylose of slightly degraded waxy starch pregels that expand to tender structures with heating and drying. Rubbery structures are obtained with thin boiling corn starch in the presence of high concentrations of sugars. In other instances, substances influence the rigidity or strength by (1) competing with water for the binding loci, such as sugar-softening a starch gel, (2) competing with the solid phase for the liquid, as sugar does in high methoxyl pectin gels, (3) altering the pH, as acids do in pectin systems, (4) interacting chemically with either or both phases, as calcium ions forming cross-bridges with alginates or low-methoxyl pectins, and (5) altering the charge distribution on the polymer molecules.

Relationships of gel strength to interactive forces

The type and strength of the crystalline junction zones in polysaccharide gel networks govern the strength, elasticity, and flow behavior of the gel. If the junction zone is short, and chains are not held together strongly, the polysaccharide molecules will separate under physical pressure or with slight increases in temperature. One or more polysaccharides can be involved in the junction zone; the zones can involve multiple helices or aggregates of ordered ribbons. Agar forms one of the strongest gels known via bundles of associated double helices, and the gels

remain stable at temperatures from 30 to 85 °C. Helical junction zones also are involved in carrageenan and furcelleran gels. Low-methoxyl pectins, with less than 50% esterification, can form stable gels using divalent ions, such as calcium, to form cross-links. Alginate gelation occurs at room temperature in the presence of calcium or other di- or trivalent ions, or in the absence of ions at pH 3 or less. Gellan, in the presence of ions, gels similarly to alginate but it gives a similar brittleness, elasticity, and cohesiveness to agar gels, except that the gellan gels are more firm. In the case of carrageenans, ions are immobilized at the junction zones, although the primary gelation mechanism is believed to be hydrogen bonding. Kappa carrageenans form firm gels with potassium, but not with sodium ions, and iota carrageenan is calcium-sensitive. Kappa and iota carrageenan form a gel; however, lambda fractions are strongly anionic, do not complex readily with cations, and do not gel.

Polysaccharides with carbonyl groups also do not gel easily because of electrical repulsion between approaching chain segments. High methoxyl pectin (55–80% esterification) forms strong gels only if (1) the pH is adjusted to 2.0–3.5 to prevent charge repulsion and ionization of the carbohydrate groups and (2) a dehydrating material is used to increase the intermolecular interactions by hydrogen bonding among the polymers. A synergism between alginates and pectins for gel formation occurs if the system is first acidified to reduce electrostatic repulsion and permit molecular association.

Molecular configuration and gel character Branched molecules do not form strong junction zones, thus they do not form strong, elastic gels. Studies on cross-linking mechanisms of mixed gels suggest an ordered binding between extended ribbon conformations of smooth, unbranched areas on galactomannan chains and double helices of agarose or carrageenan. The smooth, unstructured regions of the mannan and xanthan gum are involved in thermally reversible gels.

Concentration effects Gel strength in simple alginate and pectin gels and in the more complex starch gel is strongly concentration-dependent. A critical concentration, specific for any given polymer–solvent pair, is required. Firm gels can be prepared from a few types of gums, such as pectins, at levels of 1% or less. Other colloidal gelling agents require up to 10%. The higher polymer concentrations necessary in some systems likely allow aggregate formation at low temperatures. The overall strength of starch gels depends upon the residual swollen granules to reinforce the strength of the amylose gel matrix. Soft gums can be obtained with high-amylose starches. Retrogradation of amylose is controlled to give textures ranging from a short, clean bite to a long, somewhat chewy bite in products such as gum drops, orange slices, and gummy bears. Thin boiling starches obtained by acid treatment to produce short chains in the amylopectin portion are used with high concentrations of starch and sugar for the manufacture of firm, rubbery gum candies.

Crystalline junction zones can occur to such a degree in polysaccharide gels that retrogradation and syneresis occur. Increased polyguluronate junction zones in alginate gels result in rigid, brittle gels with high syneresis, whereas few junction zones produce an elastic gel with less tendency to synerese. Generally, syneresis resulting from compression gives a feeling of juiciness in the mouth.

Synergism of Polysaccharides with other Food Components

Gel structures Mixed gels and filled or composite gels are utilized to obtain specific textural properties in food products. The texture is dependent upon the relative proportions of each component and the overall concentration. Locust bean gum with kappa-carrageenan forms brittle, elastic gels, but xanthan gum with kappa-carrageenan or low-methoxyl pectin with iota carrageenan forms soft, cohesive gels. Low levels of gellan gum can give similar gel characteristics to high levels of κ-carrageenan with locust bean gum. Myoglobin and bovine serum albumin are proteins that promote low-methoxyl pectin gelation, but they inhibit the formation of gels using alginate. Each gelling agent can contribute to the gel formation in mixed gels, or one of the gelling agents can be nonactive, modifying the characteristics but not interacting to form the gel network. For example, agar forms microgels in the gelatin matrix of an agar–gelatin mixed gel.

Flavor carriers and encapsulation Carbohydrates can enrobe, absorb, and retain flavor volatiles during processing, which has important implications in food systems. Gum arabic, guar gum, modified starches, and maltodextrins as well as sucrose and lactose are used for encapsulation by spray-drying or extrusion processes. Inclusion complexing is used for encapsulation by the β-cyclodextrins.

Glossary

Enantiomers – D and L forms of molecules;
Encapsulation – packaging of solids, liquids or gaseous materials in miniature, sealed capsules; that can

release the contents at controlled rates under specific conditions

Newtonian – fluid with a rate of deformation directly proportional to stress;

Pseudoplastic – fluid that undergoes thinning with increasing rates of shear;

Pyschophysical – study of the relationship between sensory stimuli and human responses;

Stereochemical – referring to spatial arrangement of atoms and groups in molecules;

Synergism – action of two or more substances to achieve effect of which each individually is incapable;

Syneresis – leakage of liquid from a gel;

Transduction – translating sensory stimuli information into signals useful to the nervous system

See also: **Biscuits, Cookies, and Crackers**: Nature of the Products; Chemistry of Biscuit Making; **Browning**: Nonenzymatic; **Cakes**: Nature of Cakes; Chemistry of Baking; **Flavor (Flavour) Compounds**: Production Methods; **Gums**: Food Uses; **Sensory Evaluation**: Appearance; Texture; Taste; **Starch**: Functional Properties; **Sucrose**: Properties and Determination; **Sweeteners**: Intensive; **Sweets and Candies**: Sugar Confectionery; **Syrups**

Further Reading

Acree TE, Shallenberger RS and Ebeling S (1998) Thirty years of the AH-B theory. In: Contis ET, Ho C-T, Mussinan CJ *et al.* (eds) *Food Flavors: Formation, Analysis and Packaging Influences*, pp. 1–13. New York: Elsevier Science.

Alexander RJ (ed.) (1998) Confections. In: *Sweeteners: Nutritive. Handbook Series*, pp. 63–77. St. Paul, MN: Eagan Press.

Birch GG (1980) Theory of sweetness. In: Koivistoinen P and Hyvönen L (eds) *Carbohydrate Sweeteners in Foods and Nutrition*, pp. 61–75. New York: Academic Press.

DuBois GE, Walters DE, Schiffman SS *et al.* (1991) Concentration–response relationships of sweeteners. A systematic study. In: Walters DE, Orthoefer FG and DuBois GE (eds) *Sweeteners, Discovery, Molecular Design, and Chemoreception. ACS Symposium Series 450*, pp. 261–276. Washington, DC: American Chemical Society.

Hutteau F, Mathlouthi M, Portmann MO and Kilcast D (1998) Physicochemical and psychophysical characteristics of binary mixtures of bulk and intense sweeteners. *Food Chemistry* 63: 9.

Koivistoinen P and Hyvönen L (eds) (1980) *Carbohydrate Sweeteners in Foods and Nutrition*. New York: Academic Press.

Lineback DR and Inglett GE (eds) (1982) *Food Carbohydrates*. Westport, CT: AVI Publishing.

McLaughlin S and Margolskee RF (1994) The sense of taste. *American Scientist* 82: 538.

Morris VJ (1986) Gelation of polysaccharides. In Mitchell JR and Ledward DA (eds) *Functional Properties of Food Macromolecules*, pp. 121–170. New York: Elsevier Science.

Oakenfull D (1987) Gelling agents. *CRC Critical Reviews in Food Science and Nutrition* 26: 1.

Reineccius GA (1989) Flavor encapsulation. *Food Reviews International* 5: 147.

Shallenberger RS (1998) Sweetness theory and its application in the food industry. *Food Technology* 52: 72.

Waller GR and Feather MS (1983) *The Maillard Reaction in Foods and Nutrition. ACS Symposium Series 215*. Washington, DC: American Chemical Society.

Carboxylic Acids *See* **Acids**: Properties and Determination; Natural Acids and Acidulants

CARCINOGENS

Contents

Carcinogenic Substances in Food: Mechanisms

R Mehta and T J Schrader, Sir Frederick Banting Research Centre, Ottawa, Ontario, Canada

Background

Cancer is a major disease resulting in more than 7 million deaths per year worldwide. The risk of cancer varies from region to region, suggesting a role for geographic, environmental, and cultural factors in cancer causation. Current figures suggest that although about 37% of cancers deaths in the USA are caused by cigarette smoking, diet is also a significant risk factor, estimated to be associated with about 35% of cancer deaths.

The human diet, whether derived from an animal or a plant source, is an extremely complex and a variable mixture of macro- and micronutrients, and naturally occurring and synthetic substances. Studies in experimental animals and human epidemiology have shown that cancer development is a multistep process consisting of three main stages: initiation, promotion, and progression (Figure 1). Each stage further involves numerous events, many of which have as yet not been defined precisely. Generally, during the initiation stage, normal cells are exposed to chemical, viral, or physical carcinogens. Such exposure may cause change(s) in the genome, providing a selective growth advantage to the altered cells. Promotion involves selective growth of these genetically altered cells leading to precancerous lesions and eventually tumors. In the progression stage, the tumor undergoes additional genetic and morphological changes, making it malignant with the ability to invade and spread (or metastasize) to other sites in the body. Natural or synthetic carcinogens in food may contribute to cancer development during either the initiation or promotion stage, depending on the type of reactivity with the genetic material.

Naturally Occurring Carcinogens in Food

Environmental exposures to naturally occurring chemicals result principally from the food we consume ($1.5\,kg\,day^{-1}$). Although air and water frequently contain at least trace levels of man-made contaminants, they are seldom a source of naturally occurring carcinogens. The numbers and amounts of naturally occurring substances in a food far exceed the synthetic compounds, and include macronutrients (fat, carbohydrates, and proteins), micronutrients (vitamins and minerals), and nonnutrient constituents.

Both macro- and micronutrients have been shown to modify cancer incidence in experimental animals. In man, the modulation of cancer by macro- and micronutrients is less definitive, though there is some evidence that breast cancer rates are associated with the amount and type of fat as well as with a high proportion of calories as fat. In experimental models, total caloric intake, as opposed to a single macronutrient, is also known to have an impact on age-related changes in tumor incidence. For example, decreasing calorie intake results in a reduction in cancer incidence.

The naturally occurring chemical nonnutrients, also termed natural food contaminants, may originate in the food supply as a result of (1) physiological and biochemical processes inherent in the food organisms themselves; (2) byproducts formed during stress, storage, processing, and preparation of food; (3) infection or spoilage during storage by bacteria or fungi; (4) materials introduced into the food product through the dietary source consumed by the animal destined for food; and (5) naturally occurring substances isolated from raw plants or animal sources and intentionally added back to the food, for example, as a flavor or color enhancer. Examples of carcinogens naturally occurring in normal human diet under each of these categories are shown in Table 1.

Synthetic Carcinogens in Food

Synthetic chemicals in foods occur as (1) direct or intentional additives where synthetic chemicals are added deliberately to enhance the appearance, nutri-

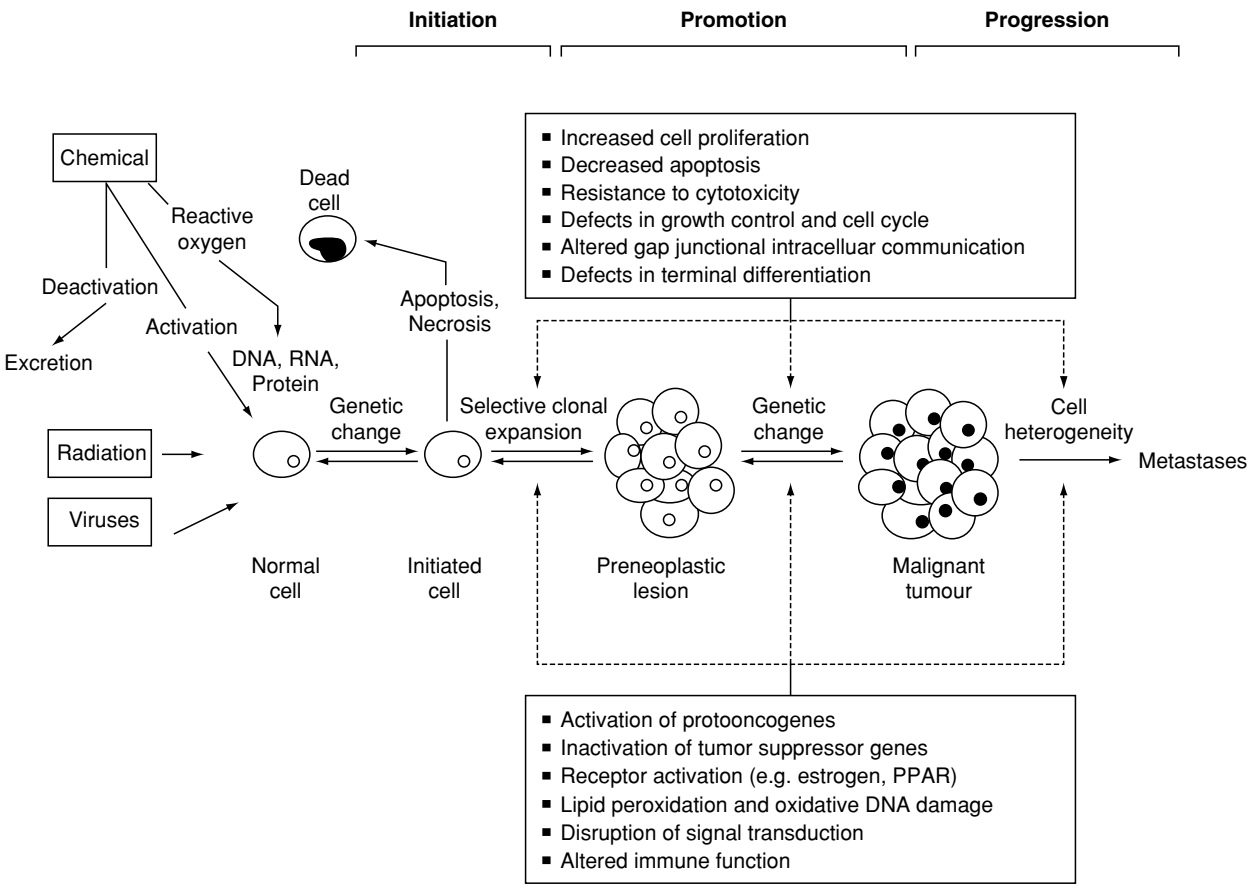

Figure 1 General scheme for sequential steps in multistage carcinogenesis.

Table 1 Naturally occurring substances that may be present in US diets

Origin of natural substance in food	*Examples (partial list)*
(1) Physiologically/biochemically produced through processes inherent in the food organism	Acetaldehyde, caffeic acid, estrone, estradiol, projestrone, safrole, styrene, testosterone, 8-methoxypsoralen (with UV light exposure)
(2) Byproducts formed during stress, storage, food processing, and preparation.	Polycyclic hydrocarbons (benzo[a]pyrene), nitrosoamines (*N*-nitrosodimethylamine), heterocyclic aromatic amines (PhIP, MeIQx, IQ, glu-P-1), methylmercury compounds
(3) Infection or spoilage during storage by bacteria or fungi	Aflatoxins, ochratoxin A, fumonisins, sterigmatocystin
(4) Introduction into food through dietary source consumed by the organism destined for food	Arsenic, cadmium, benz[a]anthracene, benzo[a]pyrene, seafood toxins, indeno(1,2,3)pyrene, toxol in snakeroot, aflatoxin from cows' milk
(5) Natural substances isolated from raw plants or animal sources and intentionally added back to the food	Sucrose, glucose, isolated soy protein flavors extracted or distilled from spices, gums, corn, or tapioca starches

tional quality, or shelf-life of a food product by, for example, adding color, a nutritional supplement, or a preservative; and (2) indirect additives where synthetic chemicals are present as a result of the use of, for example, pesticides, packaging materials, antibiotics, or growth-promoting substances during food production, processing, or storage. Indirect additives may also be termed synthetic food contaminants. Table 2 lists examples of carcinogens under these categories.

Exposure and Risk Evaluation

The identification, monitoring, and categorization of actual carcinogenic risks in relation to the types and amounts of natural and synthetic substances in the

Table 2 Synthetic substances that may be present in US diets

Origin of synthetic substance in food	*Examples (partial list)*
(1) Directly or intentionally added	*Appearance modifiers*: Glazes, waxes, colors (tartrazine, sunset yellow), surface finishing agents (polyvinyl-pyrrolidone)
	Curing and pickling agents: Sodium nitrite, ascorbic acid, sodium tripolyphosphate
	Nutrient replacements/supplements: Sucrose polyesters, all essential vitamins, minerals and fatty acids
	Sweeteners: Saccharin, aspartame, sucrose, glucose
	Product stability aids: Antioxidants (butylated hydroxyanisole), antimicrobials (sodium benzoate, potassium sorbate)
	Flavoring agents: Triethyl citrate, monosodium glutamate, ethyl maltol
(2) Indirectly or unintentionally added	*Pesticides*: Toxaphene, chlordane, organotins, triazines, dinitranilines, arsenicals and fluorides, DDT, chlordecone
	Growth promoters: Estradiol, progesterone, reserpine, testosterone
	Antibiotics: Sodium *o*-phenyphenate
	Packaging materials: Acrylamide, di-2-ethylhexylphthalate, carbon tetrachloride, chloroform, styrene, styrene oxide, dimethylformamide

food are complicated by (1) the large variety of such materials that are present in the human diet and (2) the ability of such substances to manifest both carcinogenic and anticarcinogenic properties with different mechanisms of action.

When comparing possible carcinogenic risks of natural relative to synthetic substances, it is therefore necessary to take into account (1) their relative exposure through food and (2) any theoretical differences in chemical and biological properties of both types of chemicals. Traditional foods (e.g., grains, fruits, vegetables, and meat) comprise the bulk of the diet. Items such as sugar and salt, generally recognized as safe, are the most highly used direct food additives. Other direct additivies such as artificial sweeteners, colors, preservatives, spices, and flavors are used in much smaller amounts. In contrast, indirect food additives, such as pesticides and substances from packaging materials, represent over 2000 chemicals, many of which may be present in food below the level of detection.

In terms of possible differences in chemical properties, one commercially desirable property for a synthetic food substance is chemical stability, which is often achieved by adding a chlorine group to the synthetic chemical. This may then give rise to persistence of such chemicals in the environment, and accumulation in the plant and animal food chains.

An excellent example of such persistent indirect food additives is pesticides, which, among the general public and regulatory agencies, have raised the most concern compared to other indirect food additives. Pesticides in food may originate from any of the following sources: agricultural residues persisting on the foodstuffs or in the environment, chemicals used during storage, or water used in food preparation. While animal studies provide evidence for the carcinogenicity of many pesticides, most of the data for an association between pesticides and human cancer have been obtained from occupational exposure studies in farmers. The findings of residues of pesticides and their metabolites in blood, urine, breast milk, and adipose tissues of individuals in the general population, however, have sparked concerns over health effects of pesticides. Certain subpopulations, in particular, may be susceptible to higher than normal rates of exposure through the food supply. For example, children are one vulnerable subpopulation that may be exposed to significantly greater amounts of pesticide residues through placental transfer and breast milk. Another subpopulation of current concern is that of the Arctic whose marine food chain has accumulated persistent residues of organochlorine pesticides that were heavily used in southern parts of the North American continent until the 1980s. Such migration of persistent, indirect food additives across continents may result in adverse health consequences for the Arctic indigenous populations that consume marine mammals as part of their traditional diet.

An example of a group of naturally occurring carcinogens in food that have been most studied and regulated include the mycotoxins such as aflatoxins, ochratoxins, and fumonisins. Aflatoxins, produced primarily by strains of *Aspergillus flavus*, have been found as fungal contaminants in peanuts, corn, wheat, rice, and foods made from these commodities. An association between dietary aflatoxin exposure and high incidence of liver cancer has been observed in a number of African and South-east Asian countries. Of all the aflatoxins, aflatoxin B_1 is the most potent liver carcinogen. Recent studies have shown that other factors such as hepatitis B viral (HBV) infection, alcohol consumption, smoking, and other

environmental variables also play a role in liver cancer induction by aflatoxin B_1. In Western countries, human exposure to aflatoxin B_1 is mainly from nuts and nut products, and for children from peanut butter, estimated to be 1–2 ng per kilogram body weight (bw) per day according to Canadian estimates (1985–1987). This intake level is about 10-fold higher than the Canadian set tolerable daily intake (TDI) of 0.11–0.19 ng per kilogram bw per day. This TDI was based on a carcinogenic risk level of 1:100 000, as estimated from epidemiological data obtained from Asian and African populations. However, such an exposure level is not considered a health risk for Canadians because HBV infection, another associated factor in liver cancer development by aflatoxin B_1, is not endemic in Canada, unlike the Asian and African countries.

Another mycotoxin, ochratoxin A, produced by *Aspergillus ochraceus* and related species, has been found to occur in foods of plant origin such as wheat, barley, oats, corn, dry beans, and animal-derived food products. It has also been detected in human blood, milk, and tissues from European and Canadian populations, suggesting widespread human exposure. However, various Western countries have regulations to ensure that levels in food are controlled at levels below the estimated TDI of 1.5–5.7 ng per kilogram bw per day, which is based on a National Toxicology Program carcinogenicity study. Ochratoxin A has been implicated in endemic nephropathies in European populations as well as livestock, and has been shown to cause renal cancer in two rodent species.

Fumonisins, produced by *Fusarium moniliforme* and *F. proliferatum*, are fungal contaminants primarily found in corn and corn-based food products. Six fumonisins have been isolated and characterized from *F. moniliforme*, and are designated as fumonisin B_1, B_2, B_3, B_4, A_1, and A_2. Only fumonisin B_1 and B_2 have been studied to a significant extent. The consumption of corn contaminated with fumonisins has been associated with a high incidence of esophageal cancer in human populations in the Transkei region of South Africa and Henan Province of Northern China. Fumonisin B_1 has also been shown to cause liver cancer in rats and female mice, and renal tubule neoplasms in rats. Levels of fumonisin B_1 vary from year to year, based on fungal growth conditions (e.g., moisture during the growing season), but are consistently in the 0.5–2 p.p.m. range in US cornmeal and have been reported to be as high as 150 p.p.m. in corn for human consumption in South Africa. The levels of this mycotoxin in some foods also change as a result of the cooking process through hydrolysis of the tricarballylic acid groups, or by reacting with reducing sugars producing (carboxymethyl)-fumonisin B_1. Overall, it has been difficult to determine the accurate average daily consumption of fumonisin B_1 in the USA because this will depend on the level of contamination, the diet portion that is corn or corn-based foods, and the cooking habits. In terms of risk management, the TDI for these group of mycotoxins has not been established owing to a lack of sufficient toxicity data, although Switzerland has a regulatory limit of 1 p.p.m. in human food.

Examples of naturally occurring carcinogens that may originate in the food supply as byproducts during food processing and preparation include the nitrosamines, polycyclic aromatic hydrocarbons (PAHs), heterocyclic amines (HCAs), and lipid oxidation products. These compounds are present in some cooked foods, particularly fried, grilled, or barbecued meat and fish products. Exposure to these products can also arise from food processing: curing, drying, smoking, roasting, refining, and fermentation. The most commonly detected PAHs are benzo[a]pyrene, benzo[a]anthracene and dibenzo[a,h]anthracene, of which benzo[a]pyrene is the most carcinogenic. The intake of PAHs in Europe, the USA, and Japan has been estimated to be around 1 mg per person per year, and for benzo[a]pyrene, the values range from 0.01 to 0.61 mg per person per year. Of this, only about 25–30% of the PAHs originates from food, with the majority coming from air pollution and smoking. PAHs have been considered probable human carcinogens based on animal and *in-vitro* mutagenicity data. There is no definitive evidence for a possible relationship between ingestion of PAH-contaminated food and human cancer.

The main forms of *N*-nitroso compounds in food are nonvolatile, including a large number of compounds that could be potentially formed, e.g., proteins containing *N*-nitrosated peptide linkages, such as *N*-nitrosoproline. Although nonvolatile *N*-nitroso compounds have not been shown to be carcinogenic or mutagenic, they could be precursors to the volatile carcinogenic nitrosamines. For example, during food processing, preparation, or preservation, added nitrite and/or nitrate, naturally occurring nitrite, or nitrogen oxides resulting from heating or drying of foods with combustion gases can react with amino compounds to produce *N*-nitroso compounds. Examples of the volatile carcinogenic nitrosamines that occur in food include *N*-nitrosodimethylamine, (NDMA), *N*-nitrosodiethylamine, *N*-nitrosopyrrolidine, and *N*-nitrosopiperidine. The evidence that some *N*-nitroso compound are carcinogenic is overwhelming from animal studies. In epidemiological studies, positive correlations have been observed between the high intake of nitrosamines, and incidences of esophageal and nasal cavity cancers in

China, stomach cancer in Japan and Norway, and colorectal cancer in Finland.

A variety of HCAs are produced as pyrolysis products of sugars, amino acids, creatinine, and other constituents when fish and meat products are subjected to high temperatures during normal cooking processes. The extent of HCA exposure (estimated at nanogram to microgram levels per day per person) is influenced by factors such as type of fish or meat ingested, frequency of consumption, portion size, method of cooking, cooking temperature and duration, and the amounts of the pan residues and gravy ingested. The different groups of HCAs identified include pyrido-imidazoles or indoles (Trp-P-1, Trp-P-2, AαC, MeAαC); derivatives of imidazoquinolines or IQ compounds (IQ, MeIQ); imidazoquinoxalines (Iqx, MeIQx, DiMeIQx); and the imidazopyridines (PhIP). Among the various HCA contained in cooked meat, the content of PhIP is highest, followed by MeIQx in the Western diet. In urine samples from Japanese volunteers, however, the levels of MeIQx are higher than those of PhIP. Long-term animal studies have shown multisite carcinogenic effects with major target organs being the liver, colon, and mammary gland in mice and rats for MeIQx and PhIP, respectively, and liver for IQ in cynomolgus monkeys. Epidemiological studies have focused on meat consumption and several types of human cancer, but a major challenge has been to differentiate the effects of HCAs from that of meat by itself and other confounding factors. DNA adducts in human tissues have also been detected.

Lipid oxidation products, particularly cholesterol oxide products, are common in foods exposed to severe cooking conditions such as deep-fat frying and pan frying following long-term storage in a freezer or in dehydrated form. The formation of cholesterol oxide products is greatly accelerated by autooxidation of coexisting unsaturated triacyglycerols, resulting in oxidized forms or epoxides of cholesterol. Several studies, for example, have reported high levels (6.8–58.8 p.p.m.) of cholesterol oxide products in french fries cooked in animal fats or mixtures of vegetable oils and animal fats. In terms of risk, the relationship between long-term consumption of lipid oxidation products and human health is not clear. Data from animal studies are not reliable, because high levels (10–15% of diet) of lipid oxidation products are required in the diet of the laboratory animal to produce an effect.

Following a number of evaluations, the International Agency for Research on Cancer has concluded that several of the nitrosamines, PAHs, and HCAs are probable or possible human carcinogens based on both long-term animal studies and *in vitro* mutagenicity tests. However, in the absence of sufficient evidence that these mutagens really cause human cancer, no limits have been set for their presence in foods. The regulatory agencies in most developed countries recommend minimizing their occurrence in foods by changing cooking practices.

In the USA, UK, Canada, and many other countries, food-safety legislation is based on traditional food, added substances, and unintended added substances. It would seem, therefore, that food additives and pesticides do not pose a major cancer threat in view of their rigorous regulation, and hence, only limited exposure is expected from below 'tolerance level' quantities being permitted in foods. In the USA, the Food Additive Amendment in 1958 contained the 'Delaney Clause', which states that no natural of synthetic additive be allowed in food in any amount if it has been shown to produce cancer in animal studies or other appropriate tests. There is no legislation, however, to regulate naturally occurring carcinogens from plants or cooking practices, or the macrocomponents of the diet which with improved analytical techniques have been shown in recent years to be a cancer risk, either directly or by modulating the disease process. Attempts to estimate cancer risks from naturally occurring carcinogens are further complicated because of the difficulties in differentiating between the effects of these substances from those of other confounding factors such as individual differences within human populations with respect to rates of exposure, the ability to activate and detoxify these substances, the presence of other protective mechanisms, and lifestyle. An in-depth understanding of the biological mechanisms by which these diverse carcinogens exert their action is therefore necessary for the development of methods and biological markers that allow (1) reliable extrapolation of biological mechanism-related data from appropriate experimental models to humans; (2) accurate biomonitoring in human populations for exposure to specific food borne carcinogens and their health effects; and (3) estimations of cancer susceptibility to specific foodborne substances in individuals and subpopulations. Such measures would then provide essential data for assessment of potential human cancer risk as well as perhaps avenues for legislative control of these substances.

Mechanisms of Carcinogenesis

Although the detailed molecular mechanisms of cancer induction by carcinogens in food remain to be discovered, several factors required for the multistep process have been identified: two or more genetic events, either inherited or resulting from point

mutations, chromosomal rearrangements, insertions or deletions of genes, and gene amplification; cell replication to incorporate the altered DNA into the genome; cell proliferation to selectively expand the altered cell population into a benign tumor; and additional genetic events leading to a malignant tumor with the ability to invade and metastasize to other sites in the body. Some of the genetic events at various stages during cancer development include overexpression and/or activation of cellular protooncogenes, and deletion or inactivation of tumor suppressor genes (Figure 1). Such alterations in the expression of protooncogenes and tumor suppressor genes may then cause the dysregulation of the cell cycle either directly or indirectly. The cell cycle is a complex process involved in the growth and proliferation of cells, and involves numerous regulatory proteins that direct the cell through a specific sequence of events concluding with mitosis and the production of two progeny cells. Many of the protooncogenes and tumor-suppressor genes also play a key role in the regulation of natural cell death or apoptosis. Over the past decade, it has become increasingly clear that cancer is the disease not only of increased cell replication but also of inhibition of apoptosis. The relationship between cell replication and normal cell death (or loss) determines the rate of tumor growth. In addition, the immune system is known to play a role in recognition of cancer cells and may contribute to cancer prevention by killing cancer cells. Therefore, any inhibition of the immune surveillance mechanism may compromise its protective capacity and thus result in tumor development.

Natural or synthetic carcinogens in food can generally be divided into those that directly affect DNA (genotoxic) and those that do not (nongenotoxic), and may be classified according to their role in triggering events during one or more of the stages of cancer induction (Table 3). Genotoxic substances chemically react with DNA and are usually implicated during the initiation stage of carcinogenesis. Nongenotoxic carcinogens, which do not have the capacity to react with the DNA directly, may alter DNA through indirect or epigenetic mechanisms and are, therefore, expected to contribute during the promotion or progression stages. Some chemicals cause indirect DNA damage and genomic alteration through oxygen radical generation in a variety of ways. Certain nongenotoxic carcinogens, such as peroxisome proliferators and estrogens, interact with specific receptor proteins to produce protein–ligand complexes, which then interact with, and

Table 3 Possible mechanisms of carcinogenicity by foodborne substances

Carcinogen classification according to cancer stage of most likely activity	*Capacity to*	*Possible mechanism(s) of action*	*Selected examples*
Initiating agent	Initiate cells only	Electrophile generation through spontaneous breakdown or metabolic activation of chemical with resultant DNA damage and two or more genomic mutations	Mycotoxins (aflatoxin B_1) Nitrosamines (NDMA) Heterocyclic amines (PhIP, IQx) Pesticides (toxaphene)

Promoting agent	Selectively expand the initiated cell population
	Convert an initiated cell or a promoted cell to a malignant cell
		One or more of:
		• enhanced cell proliferation
		• decreased apoptosis
		• epigenetic effects via:
		  • lipid peroxidation and oxidative DNA damage
		  • receptor protein interaction–peroxisome proliferation, hormonal action
		  • altered signal transduction pathways
		  • compromised immune function
			Mycotoxins (fumonisin B_1)[a]
			Pesticides (chlordane, organochlorines, chlorophenols)
			Packaging materials (di-(2-ethylhexyl)phthalate)

Complete carcinogen	Induce cancer from normal cells usually exhibiting the properties of initiating, and promoting agents	Electrophile generation, DNA damage, genomic mutations, and any or all of the above tumor-promoting mechanisms	Mycotoxins (aflatoxin B_1, fumonisin B_1)[a] Nitrosamines Heterocyclic amines Pesticides

[a]Some promoting agents may also exhibit weak initiating capacity with sufficient promoting ability to produce tumors, in which case, they may also be classified as complete carcinogens, e.g., Fumonisin B_1.

modify, the expression of critical regions of genomic DNA. Other substances do not appear to exert any of the above effects, yet greatly enhance cell proliferation, and are usually active during the promotion stage (Table 3).

Metabolic Activation and Electrophile Generation

Many of the genotoxic foodborne carcinogens described above undergo metabolic activation and transformation in the body. Through oxidative metabolism primarily by the cytochrome P450 family of mixed-function oxidase enzymes, highly reactive electron-deficient species can be produced, which can then depurinate or deaminate DNA, or form covalent adducts with cellular macromolecules, including DNA. Electrophiles may also be produced through flavine-dependent enzymes, the prostaglandin endoperoxide synthetase system, and other peroxidases. These electrophiles may be further metabolized by reaction with sulfate, glutathione, or glucuronic acid (phase II reactions), resulting in detoxification and excretion. In some cases, phase II reactions may also lead to carcinogenic metabolites. Therefore, the overall dose of the carcinogen that interacts with DNA (biologically effective dose) is often dependent on the presence of a number of competing metabolic pathways in any given organ or animal species. The following examples illustrate the diversity in organ specificity and mechanisms of metabolic activation leading to DNA interaction among naturally occurring carcinogens originating from fungal contamination and cooking processes.

Mycotoxins Aflatoxin B_1, a liver carcinogen, is metabolized via the cytochrome P450-catalyzed oxidation to various hydroxylated derivatives, and two unstable, highly reactive epoxide isomers. The epoxide, aflatoxin B_1-8,9-oxide is considered to be the active metabolite, which covalently binds to guanine in the DNA to form the major adduct 8,9-dihydro-8-(N^7-guanyl)-9-hydroxyaflatoxin B_1 (AFB_1-N^7-GUA) (Figure 2).

This epoxide can also react with RNA, histones, and albumin. The balance between the various primary steps, the relative rates of detoxification of the hydroxylated metabolites via glutathione or glucuronic acid conjugation, and the extent of formation and repair of the DNA adducts determine the carcinogenicity of aflatoxin B_1 in different species. The AFB_1-N^7-GUA formed may lead to persistence of repair-resistant adducts and depurination or error-prone DNA repair, which then result in single-strand breaks, base-pair substitution, or frameshift mutations. Mispairing of the adduct induces both transversion and transition mutations. Such interactions may also result in the activation of oncogenes and/or inactivation of tumor suppressor genes. Thus, recent human and experimental studies have indicated a high frequency of $G \rightarrow T$ transversions, and the prevalence of the codon 249 mutation in the tumor suppressor p53 gene in liver tumors obtained from populations where aflatoxin B_1 is the causative agent.

Ochratoxin A, a renal carcinogen, contains an isocoumarin moiety linked by a peptide bond to phenylalanine and can be hydroxylated via P450 or degraded by peptidase, as indicated in Figure 3.

Recent studies have provided clear evidence of renal carcinogenicity of ochratoxin A in two rodent species, and ochratoxin A–DNA adducts have been detected in kidney, liver, and spleen of mice exposed to the mycotoxin. Ochratoxin A was initially found not to be mutagenic in various microbial and mammalian gene mutation assays, although weak genotoxic activity was observed to mammalian cells. While the mechanism of carcinogenesis is unknown, one toxic mechanism is the inhibition of protein synthesis by competition with phenylalanine in the phenylalanyl-tRNA synthetase-catalyzed reaction, and inhibition of other enzymes that use phenylalanine as a substrate such as phenylalanine hydroxylase. The effect of ochratoxin A on protein synthesis is followed by an inhibition of RNA synthesis. Ochratoxin A has also been found to enhance lipid peroxidation both *in vitro* and *in vivo*.

Nitrosamines Through mechanistic studies, it is evident that nitrosamines are the most potent initiating carcinogens as well as complete carcinogens. This is because nitrosamines, through metabolic activation, have the capacity to cause point mutations leading to protooncogene activation by alkylation and base substitution of critical sites in DNA. As an example, the metabolic activation of NDMA is illustrated (Figure 4).

N-Nitrosodimethylamine is oxidized by the cytochrome P450 monooxygenase enzymes to hydroxymethyl-methylnitrosamine, which is unstable and further decomposes to formaldehyde and nitrosomethylamine. The latter then gives rise to nitrogen-separated ion pairs and the unstable intermediate methyldiazoniumhydroxide. The proximate species that alkylate sites in protein, RNA, and DNA are produced either directly (methyl cabonium ion) or after loss of water (diazomethane) from methyldiazoniumhydroxide. More than 15 different sites susceptible to alkylation of bases in DNA have been identified. Alkylation at different sites appears to have variable probabilities of generating specific genotoxic events. For example, the miscoding O^6-alkylguanines are likely the predominant sources

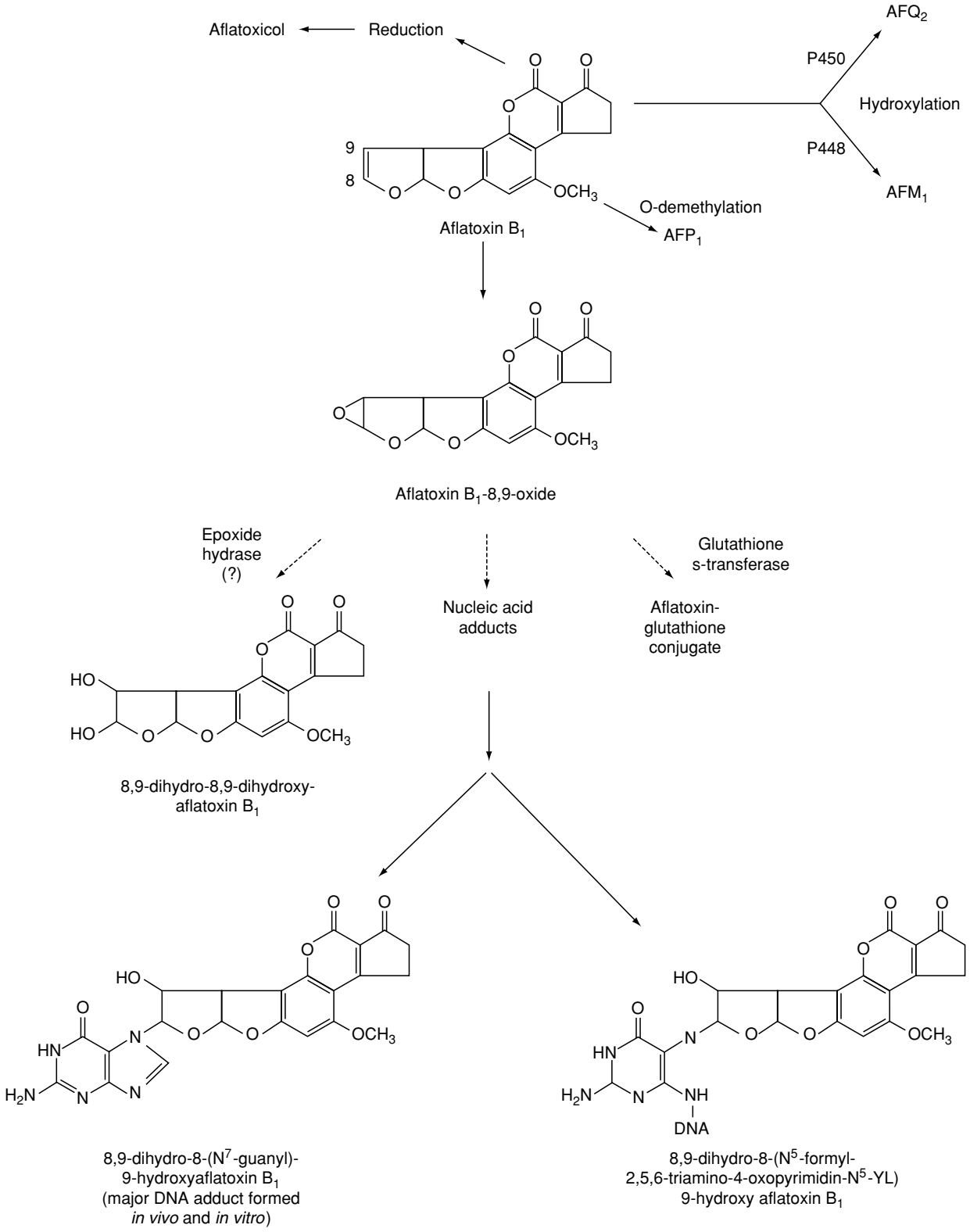

Figure 2 Metabolic activation of aflatoxin B_1.

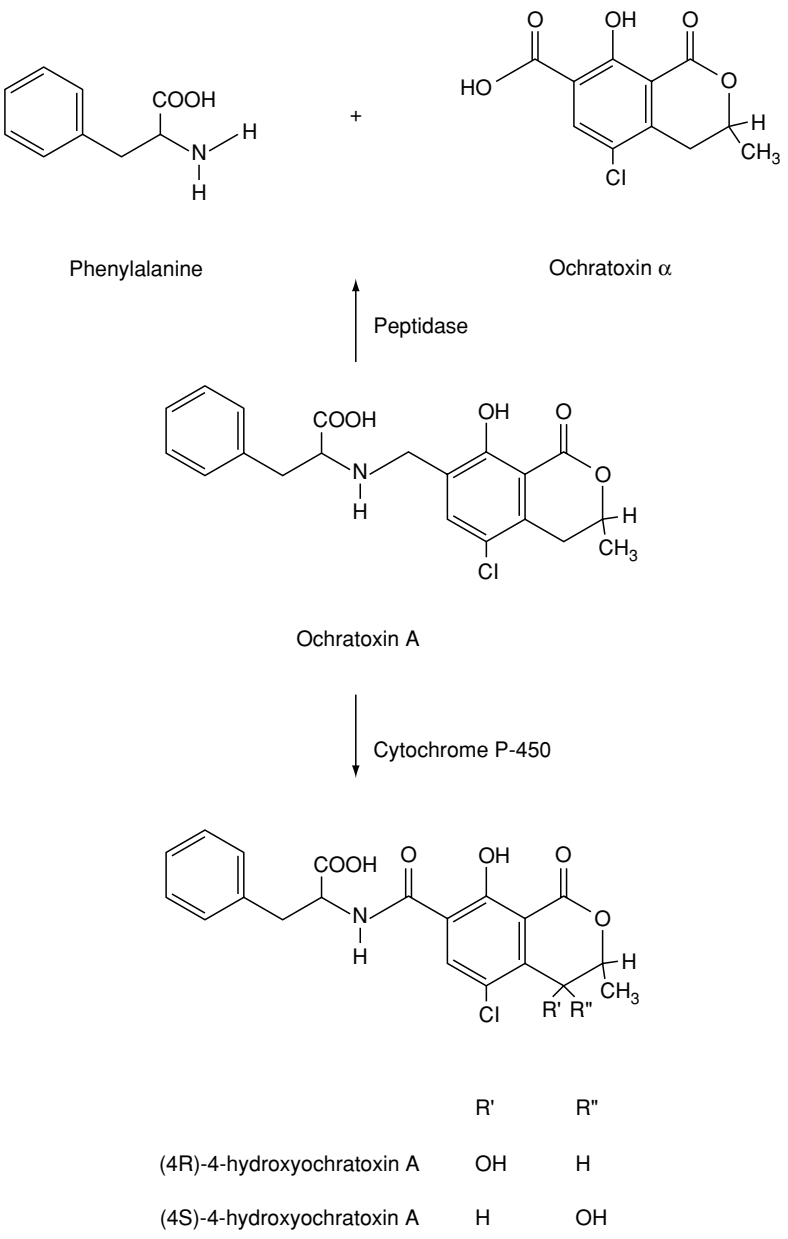

	R'	R"
(4R)-4-hydroxyochratoxin A	OH	H
(4S)-4-hydroxyochratoxin A	H	OH

Figure 3 Ochratoxin A and its metabolism.

for mutations leading to activated oncogenes. However, other adducts may also contribute at a lower frequency or may enhance the effects of the O^6-alkylguanine adducts through DNA repair processes that lead to incorporation of incorrect bases.

Heterocyclic amines Studies in rodents and nonhuman primates indicate that HCAs are rapidly absorbed, widely distributed across tissues, extensively metabolized and excreted via urine or feces. The principal sites of HCA metabolism are shown in Figure 5 with MeIQx as an example. These compounds can be metabolized by cytochrome P450-mediated ring oxidation or direct phase II conjugation reactions leading to detoxification. The cytochrome P450-mediated N-oxidation of the exocyclic amine nitrogen produces reactive metabolites that bind to DNA, RNA, and protein in almost all tissues studied. Figure 6 illustrates the structures of MeIQx adducts with guanine in the DNA. Further activation of the *N*-hydroxy HCA metabolites may occur by phase II enzymes to reactive ester derivatives, which are not stable and spontaneously rearrange into electrophilic arylnitrenium ion intermediates that react with DNA, RNA, or protein, and are mutagenic in *in vitro* assays.

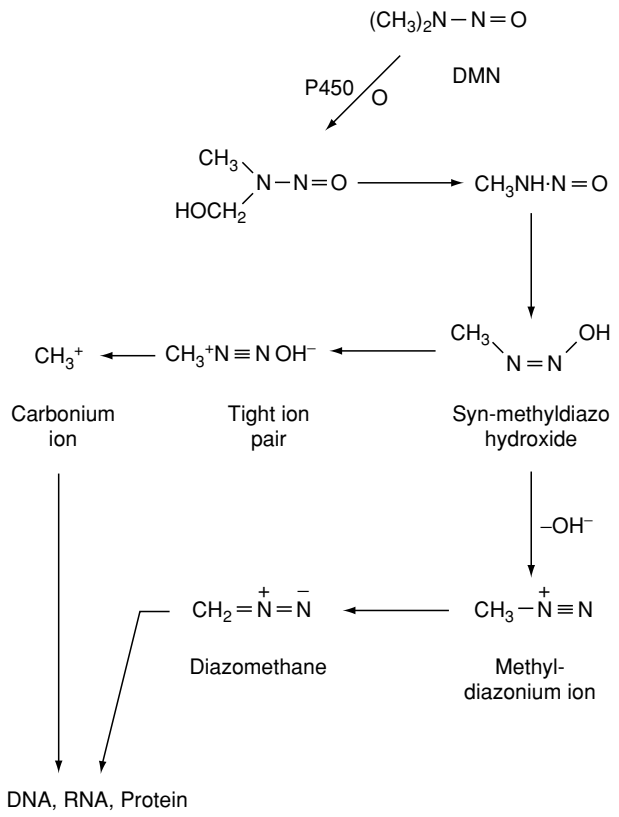

Figure 4 *In vivo* conversion of dimethylnitrosamine.

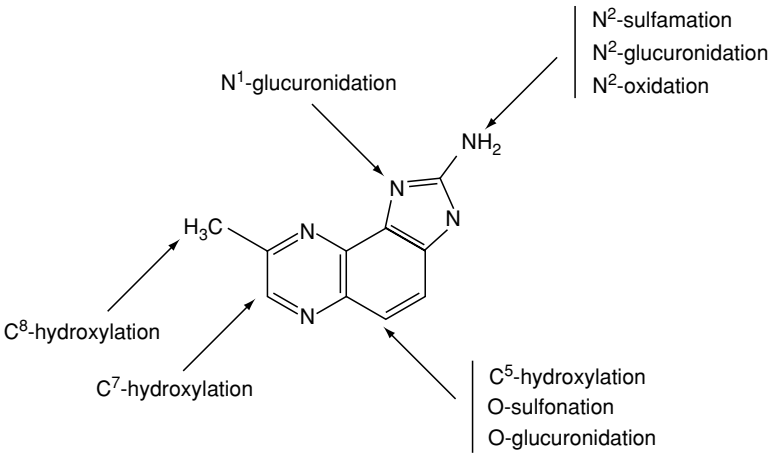

Figure 5 Sites of heterocyclic amine metabolism (MeIQx is shown as an example) by cytochrome P450 and phase II enzymes. The *N*-hydroxy HCA derivatives may undergo further activation by phase II enzymes.

The mutagenicity of HCAs has also been attributed to the DNA adducts formed from the arylnitrenium ion intermediates. Interspecies, interindividual, and tissue-specific differences in the expression of various enymatic pathways influence DNA adduct formation by HCAs. The levels of DNA adduct formation do not always correlate with target organ specificity, suggesting that, as with other genotoxic carcinogens, other factors are necessary for tumor induction by HCAs. These substances generally enhance cell proliferation in their target organs: for example, PhIP, a colon carcinogen in colon, but not liver or kidney; and MeIQx, a hepatocarcinogen, in liver but not colon or kidney. Inhibition of apoptosis, and

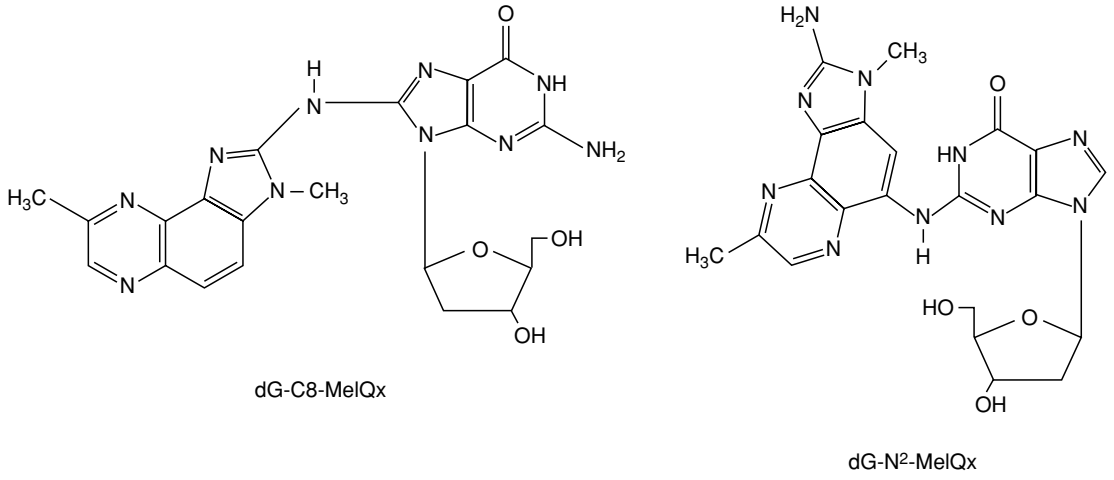

Figure 6 Chemical structures of IQ and MeIQx-DNA adducts formed at the C8 and N^2 atoms of guanine.

mutations in oncogenes and tumor suppressor genes have been reported in various types of HCA-induced tumors. Recent evidence in animal models suggests that spontaneously or artificially induced inherited mutations in cancer-related genes and certain dietary factors increase the susceptibility to cancer induction by HCAs.

Nongenotoxic Carcinogens and Epigenetic Mechanisms

The diversity of epigenetic mechanisms that may contribute to cancer causation by nongenotoxic foodborne carcinogens can be illustrated with various pesticides, and the mycotoxin, fumonisin B_1, as examples (Table 3). The carcinogenic pesticides can be found amongst each category of pesticide, that is, herbicides, insecticides, or fungicides, and belong to several chemical classes such as triazines, organophosphates, and organochlorines. Animal studies indicate that some pesticides are classic genotoxic agents that are metabolically activated to react with DNA and cause mutations or other DNA alterations, as shown above for other groups of foodborne carcinogens. Other pesticides are largely nongenotoxic and function as tumor promoters through the differing mechanisms described below. With more available evidence, it is being realized that, although these diverse epigenetic mechanisms may be implicated only during the promotion or progression stages, this does not necessarily preclude their role(s) during the initiation stage or throughout the carcinogenesis process as well.

Enhanced cellular proliferation and inhibition of apoptosis Some substances have been classified as tumor promoters because they greatly increase cell proliferation in one or more tissues without necessarily exerting any other effects at a significant rate. Studies in rodent liver carcinogenesis models have shown that cell proliferation in response to a chemical carcinogen may be classified as cytotoxic or mitogenic. Cytotoxic carcinogens induce cell death, which is then followed by cell proliferation to replace lost cells. A continued cell proliferation can result in an increase in levels of mutations through DNA replication errors. Mitogenic chemicals in the rodent liver have been observed to produce a sustained increase in organ weight during exposure to the mitogen. However, cell proliferation rates return to normal once the organ has reached its new size. Mitogenic agents provide a selective growth advantage to initiated cells by decreasing apoptosis, inhibiting normal liver cell proliferation or increasing cell proliferation in precancerous cells. Hence, depending on the type of carcinogen and the cell population involved, cell proliferation may affect the initiation of a tumor, whereas proliferative changes in initiated cell populations may influence tumor development and progression.

Similarly, in recent years, it has become known that programmed cell death or apoptosis contributes during multiple stages of cancer development. The number of DNA-damaged cells may be increased by mitosis or lost through apoptosis. Thus, chemicals that decrease the rates of apoptosis may favor initiation, in addition to tumor promotion and progression, or vice versa. For example, agents such as peroxisome proliferators, which induce a transient mitogenic activity, also inhibit apoptosis. Tumor promoters such as *2,3,7,8*-tetrachloro-dibenzo-*p*-dioxin may promote initiated cell growth by inhibiting apoptosis in the initiated cell populations. Current data suggest that responses of cells to proliferation and apoptosis signals may change throughout cancer

development and that, at each stage, a delicate balance between cell proliferation and apoptosis is necessary to prevent cancer induction.

Oxidative DNA damage Recent evidence implicates oxidative DNA damage as having similar consequences for the genome as electrophiles generated through metabolic activation. Fat metabolism, for example, produces a number of reactive oxygen species of which the hydroperoxide radicals ($RO^{\bullet}$ and $ROO^{\bullet}$) are stable and reactive enough to interact with DNA. About 20 different DNA lesions caused by these radicals have been described, including single- and double DNA strand breaks, and the formation of mono- or dihydroxylated derivatives of the DNA bases. Of the adducts, the most studied to date is the 8-hydroxy-2′-deoxyguanosine, which can lead to base mispairing at the same site or at an adjacent base in the DNA. Nitrogen radical-induced damage has also been reported but less frequently than oxygen radical damage. Even though carcinogens producing oxidative DNA damage are similar to the electrophile-generating substances in causing mutations, these chemicals cannot be identified as genotoxic by the current battery of genotoxicity tests, because they do not contain the necessary enzyme functions necessary to permit their detection. Among the foodborne carcinogens, the peroxisome proliferator class of pesticides (chlorophenols and di(2-ethylhexyl)phthalate) and the mycotoxin fumonisin B_1 are likely to generate lipid peroxidation and oxygen radicals, and therefore may be expected to operate through the oxidative DNA damage pathway. Much work is required before this can be confirmed.

Gap junctional intracellular communication Some tumor-promoting organochlorines perhaps operate through mechanisms such as inhibition of gap junctional intercellular communication (GJIC). Gap junctions are clusters of intercellular aqueous channels that allow communications between adjoining cells. They are formed by two semichannels, called connexons, each coming from two adjacent cells. Each connexon consists of six identical or similar transmembrane protein molecules called connexins, which are oligomerized within a plasma membrane. Gap junctions are the only means by which cells in organized tissue can share different messengers of low molecular weight and therefore maintain tissue homeostasis. The nature of these messengers can range from ions, nutrients, nucleotides and metabolites, either of endogenous or exogenous origin. During cancer induction, it has been found that certain tumor cells show high levels of communication with each other but not with surrounding normal cells, thus facilitating rapid selective clonal expansion and malignant transformation of abnormal cells. Two basic mechanisms of altered connexin functions may contribute to carcinogenesis at the genetic and functional levels. Genetic alteration of GJIC may occur at either initiation stage when genotoxic carcinogens induce mutations in connexin or other GJIC-related genes, or at later stages of tumor progression as a result of cumulative genetic alterations due to widespread genetic instability. Findings in cell cultures suggest that GJIC can affect cell-cycle-related gene expression supporting the earliest hypothesis that the role of GJIC may be to provide a channel for distribution of an intracellular factor to control cell growth.

Hormonal mimicry Other pesticides act by mimicking hormones, leading to disruption in endocrine responsive organs such as mammary gland, testis, prostrate, and uterus. For example, the nongenotoxic herbicide atrazine is associated with an increased incidence of mammary tumors and uterine adenocarcinomas in animals This is probably related to atrazine's ability to lengthen the estrous cycle, thus prolonging exposure to endogenous estrogens. Several organochlorines, including chlordecone and DDT, also have demonstrated estrogenic activity. A receptor-mediated mechanism has been suggested to explain the role of hormonal effects in cancer induction. Estrogens stimulate DNA replication and increase the mitotic rate of target cells. This is achieved through interaction of estrogens or estrogen-mimetic substances with the estrogen receptor protein, which may then react with the target cell DNA to modify the expression of genes involved in cell growth and differentiation. Sustained occupation of the estrogen receptors by estrogenic compounds is necessary for alteration of the genome leading to enhanced cell proliferation. Progestins oppose this effect by decreasing the concentration of the estrogen receptors. There is evidence that estrogenic substances may further generate oxidative DNA damage and lipid peroxidation.

Peroxisome proliferation Peroxisome proliferation is another nongenotoxic, receptor-mediated mechanism whereby some pesticides (e.g., chlorophenols and phenoxyacetic acid herbicides) and a packaging material residue (e.g., di(2-ethylhexyl)phthalate) have been shown to induce cancer in rodents. Peroxisomes are subcellular organelles responsible for fat metabolism in various tissues. For the peroxisome proliferator class of compounds, the primary target organ is liver where they cause proliferation of

peroxisomes, induce hypertrophy and hyperplasia of liver, stimulate growth factors, activate oncogenes, and increase the activity of the enzymes catalase, cytochrome P450 monooxygenases, epoxide hydrase, glucuronyl transferase, and those involved in fatty acid β-oxidation. While the precise mechanism of cancer induction by these chemicals is unknown, the pleitropic response appears to be related to activation of a novel steroid hormone receptor, the peroxisome proliferator-activated receptor (PPAR). In addition, two more theories of cancer induction have been proposed. First, induction of fatty acid β-oxidation enzymes by the peroxisome proliferators results in the generation of hydrogen peroxide and reactive oxygen species that interact with DNA to form an oxidative DNA damage-related adduct. In support of this theory, increases in levels of 8-hydroxy-2′-deoxyguanosine have been detected in peroxisome proliferator treated rats. The second theory relates to an increase in liver size and weight that follows peroxisome proliferator administration. This effect has been demonstrated to be due to both enhanced cell replication and a decrease in apoptosis. The possibility then exists that the process of DNA replication itself causes mutations. However, peroxisome proliferators may promote the growth of cells that are already initiated either spontaneously or by other environmental stimuli. Operationally, it is likely that some or all of these mechanisms play a role in cancer induction by peroxisome proliferators. The role of peroxisome proliferators in human cancer is not clear, since peroxisome proliferation does not occur in humans, although they respond to these agents in hyperlipidemia, for example. Preliminary data suggest that PPARs are expressed at low, variable levels in human liver and may also be polymorphic.

Disruption of the sphingolipid pathway The mycotoxin, fumonisin B_1, unlike aflatoxin B_1 and ochratoxin A, does not undergo extensive metabolism. In orally treated rats, the majority of fumonisin B_1 is excreted in the feces in unmetabolized form. In female vervet monkeys, products of hydrolysis of the tricarballylic acid groups of fumonisin B_1 have been observed in feces, suggesting that the only metabolism of fumonisin B_1 is by ester hydrolysis to the less polar aminopentol. Fumonisin B_1 persists in livers and kidneys of rats longer than any other tissues, and this is consistent with the observed toxicity and carcinogenic effects of this mycotoxin in these organs. Although fumonisin B_1 is a liver carcinogen, it lacks mutagenic activity in the Ames assays, and genotoxic effects in *in vivo* and *in vitro* DNA repair assays. This mycotoxin is a weak initiator, as demonstrated in *in vivo* initiation-promotion assays, but has effects

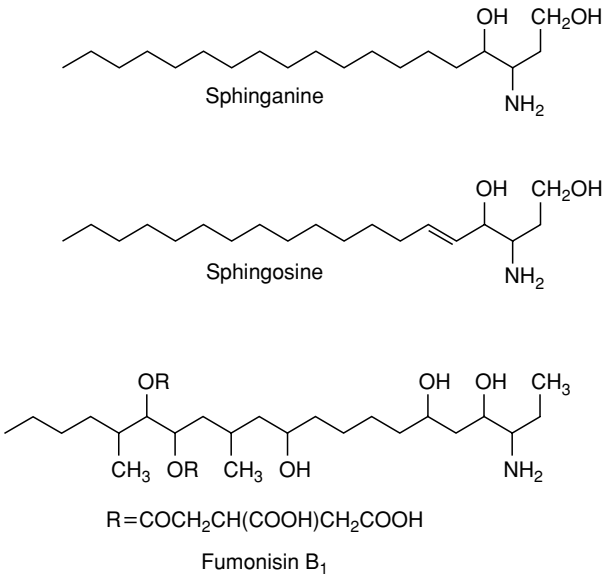

Figure 7 Structures of sphinganine, sphingosine, and fumonisin B_1.

on tumor promotion and selection of initiated cells. Lipid peroxidation and oxidative damage may be involved in the genesis of fumonisin B_1-induced initiated cells. While the molecular mechanism of action of fumonisins is not completely known, the chemical structural similarity of these compounds to sphingosine and sphinganine (Figure 7) has suggested the following mechanism. Sphingolipids are believed to be involved in the regulation of cell growth, differentiation, and cell transformation through effects on cell–cell communication, cell-membrane interactions, cell receptors, and signaling systems (Figure 8).

It has been postulated that fumonisins may exert their effects by disrupting sphingosine metabolism, and current data support this hypothesis. The main target for fumonisins is the inhibition of the enzyme sphinganine (sphingosine) *N*-acyltransferase (or ceramide synthase), resulting in the accumulation of free sphingoid bases in serum, and in the target tissues, which in the rat are the liver and kidney. The exact mechanistic role of the sphingolipid pathway in liver-cancer induction in rats remains to be elucidated.

Altered immune function Alteration of immune function is a known cause of many human cancers. Many pesticides and other foodborne carcinogens are also known to disrupt immune function, which may additionally contribute to their carcinogenic potency. For example, the insecticide, DDT, causes thymus atrophy, and 2,4-dichlorophenoxyacetic acid enhances the T- and B-cell immune response in the mouse and reduces lymphocyte subsets and natural

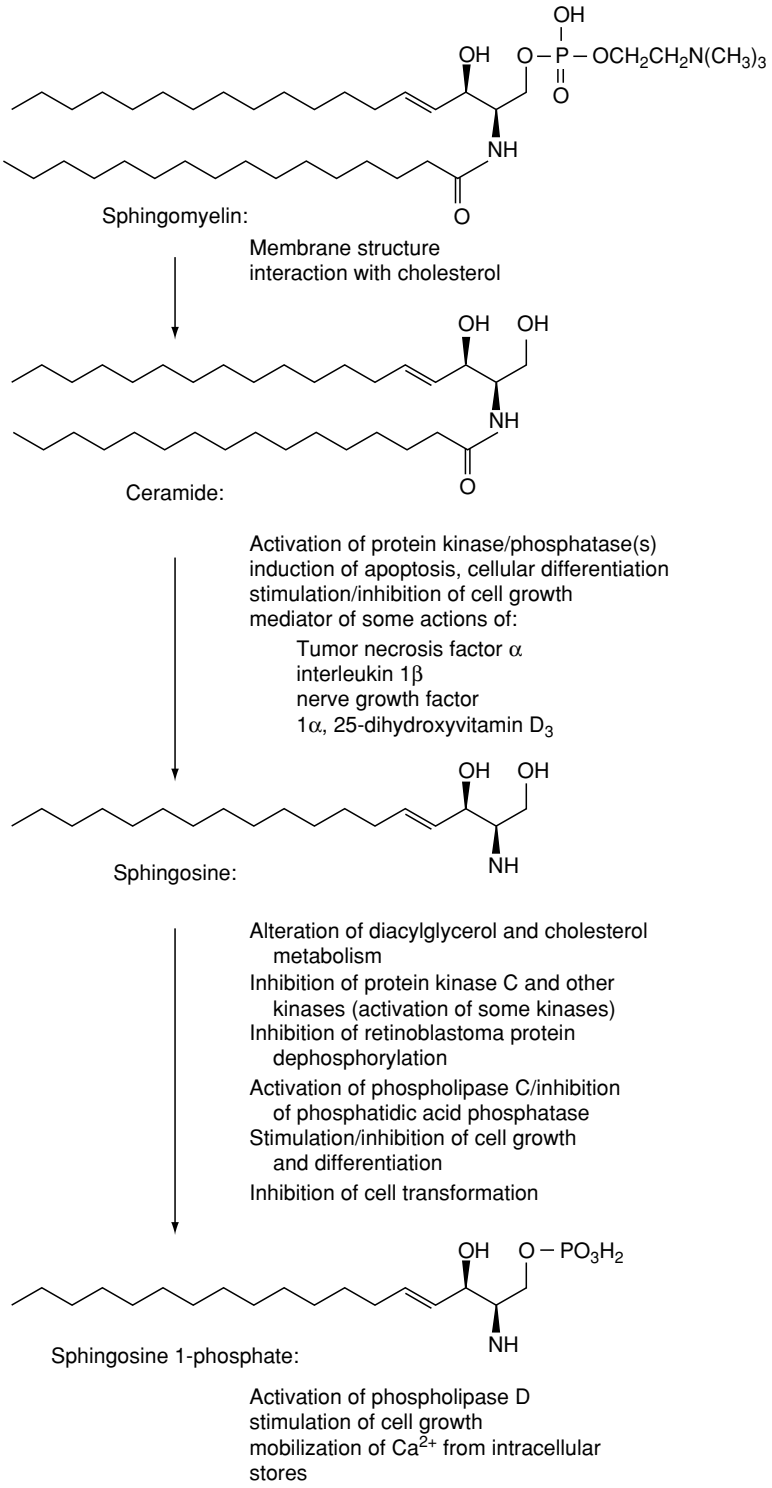

Figure 8 Examples of the products of sphingolipid hydrolysis and some of the biological activities that have been attributed to these compounds.

killer cells in farmers. Organophosphates may play a role in cancer induction through inhibition of serine esterases, which are critical components for the cytolytic activities of T-lymphocytes and natural killer cells. Ochratoxin A, *N*-nitrosodimethylamine, 7,12-dimethylbenz[a]anthracene, benzo[a]pyrene and the antioxidant butylated hydroxytoluene have also been observed as immunosuppressants in animal studies.

Future Considerations

It is clear that the majority of evidence for the types of specific mechanisms in cancer induction is from animal studies. Evidence to date suggests that the most identified genotoxic human carcinogens react with DNA through electrophile generation, but for nongenotoxic carcinogens, the mechanisms are unclear. Much research is required (1) to understand the precise role(s) of the currently known mechanisms, and (2) to discover any new mechanistic pathways that may contribute to human cancer induction by foodborne carcinogens. In addition, there is a need to define the relationship to cancer induction by factors that may interact with carcinogenic substances in food such as macro- and micronutrient levels in the diet, concurrent environmental exposure to multiple chemicals, and genetic differences in responses to these substances.

See also: **Amines**; **Carcinogens**: Carcinogenicity Tests; **Food Additives**: Safety; **Mutagens**; **Mycotoxins**: Classifications; Occurrence and Determination; Toxicology; **Nitrosamines**

Further Reading

Ames BN, Gold LS and Willett WC (1995) The causes and prevention of cancer. *Proceedings of the National Academy of Sciences of the United States of America* 92: 5258–5265.

Clayson DB (2001) *Toxicological Carcinogenesis*. Boca Raton, FL: Lewis Publishers.

Food and Agricultural Organisation of the United Nations (1996) *Worldwide Regulations for Mycotoxins*. Rome: FAO.

Keifer MC (1997) Human health effects of pesticides. In: *Occupational Medicine: State of Art Reviews*, vol. 12. Philadelphia, PA: Hanley & Belfus.

Kitchin KT (1999) *Carcinogenicity – Testing, Predicting and Interpreting Chemical Effects*. New York: Marcel Dekker.

Lawley PD (1990) *N*-nitroso compounds. In: Cooper CS and Grover PL (eds) *Chemical Carcinogenesis and Mutagenesis I*, pp. 409–470. New York: Springer Verlag.

Lockshin RA, Zakeri Z and Tilly JL (1998) *When Cells Die. A Comprehensive Evaluation of Apoptosis and Programmed Cell Death*. Chichester, UK: Wiley-Liss.

Merrill AH, Schmelz E-M, Dillehay DL *et al.* (1997) Sphingolipids, the enigmatic lipid class: Biochemistry, physiology and pathophysiology. *Toxicology and Applied Pharmacology* 142: 208–225.

Nagao M and Sugimura T (2000) *Food Borne Carcinogens. Heterocyclic Amines. Current Toxicology Series*. Chichester, UK: Wiley-Liss.

Schafer KA (1998) The cell cycle: a review. *Veterinary Pathology* 35: 461–478.

Scheuplein RJ (1990) Perspectives on toxicological risk – an example: Food-borne carcinogenic risk. In: Clayson DB, Munro IC, Shubik P and Swenberg JA (eds) *Progress in Predictive Toxicology*, pp. 351–371. Amsterdam: Elsevier.

Carcinogenicity Tests

T J Schrader, Sir Frederick Banting Research Centre, Ottawa, Ontario, Canada

Introduction

Carcinogens may be present in foodstuffs as: (1) natural components; (2) byproducts of cooking, heating, or processing; (3) food additives (preservatives, colors, flavors); (4) leachates from packaging materials; and (5) pesticides. The testing of new foods and food-associated chemicals for carcinogenicity is an important aspect of an overall toxicological assessment for determining product safety and allowable exposure limits. Government regulatory decisions on existing products can also be revisited and modified to include new knowledge gained in the understanding of carcinogenic processes and the results of new testing methods.

In excess of 200 assays have been developed to screen substances for carcinogenic potency and to characterize their modes of action. However, many of the processes involved in carcinogenesis remain poorly understood and a perfect assay for the identification of human carcinogens remains elusive. Confounding factors include: (1) the long latency period (frequently decades) required for metastatic tumor development in humans; (2) the number and types of molecular mechanisms governing the many cellular changes involved in carcinogenesis; (3) the mechanisms involved in tissue-specific susceptibilities, often showing species differences; (4) variations in genetic make-up and metabolic capabilities between individuals in a population not reflected within inbred strains of animals or tissue culture cell lines; (5) the effects of different exposure routes and timing of exposure; and (6) the contribution of coexposure to other chemicals in mixtures or other agents which could modify the carcinogenicity of a test chemical.

A tiered approach can be used to screen chemicals for carcinogenic activity and obtain supportive evidence. The first clues may arise from an examination of the test chemical for structural features which are

common with known carcinogens. Other evidence may arise from epidemiological studies relating human cancer incidence with chemical exposure. Experimental evidence may be gained from a variety of short-term chemical and cell-based (*in vitro*) assays used for screening test chemicals for DNA damage and repair, mutagenesis, chromosomal aberrations, and morphological transformation. However, while these assays can be important in identifying chemicals positive for their respective endpoints and for providing important mechanistic data, they are not representative of the entire carcinogenic process and may lack important endpoints or metabolic processes. Therefore, testing may proceed to mutagenesis and chromosomal aberration assays in animal (*in vivo*) models. Ultimately, the most validated and reliable assay for the evaluation of human carcinogens is the long-term rodent carcinogenicity assay, in which animals are exposed to the test chemical over a large part of their lifespan (generally 18 months for mice and 2 years for rats). Even so, the rodent carcinogenicity assay suffers some major drawbacks which could lead to the misinterpretation of results: (1) species differences in chemical metabolism which could cause the chemical to act in a way not possible in humans; (2) difficulties in determining the human risk of chemicals which affect the formation of tumors which otherwise occur spontaneously with age in some rodent species; and (3) errors which are introduced in extrapolating or estimating the risk to humans at normal exposure limits when chemicals are tested at far higher levels. Therefore, data from the rodent carcinogenicity assay are usually combined with that from some complementary shorter-term tests to confirm the classification of a test chemical. A number of testing procedures can also be used on human blood and urine samples to monitor if exposure to certain classes of carcinogens has occurred. Of course, this tiered approach may also work in the opposite direction, where a chemical found positive *in vivo* is then subjected to a number of *in vitro* tests to obtain information on its mechanism of action.

DNA Damage and Repair Assays

Many assays have been developed to characterize the DNA damage induced by carcinogen exposure as well as the induction of DNA repair and changes in cellular gene expression. With slight modifications, most of the assays can be used for samples arising from both *in vitro* and *in vivo* studies.

Strand Breakage Assays

Comet assay The comet assay has become one of the most convenient and widespread assays for DNA strand breakage. Cells exposed to test agent are embedded in agarose which is spread over a microscope slide to form a thin gel. The cells are then lysed and the DNA electrophoresed under either neutral conditions to detect double strand breaks or alkaline conditions to detect single and double strand breaks as well as apurinic/apyrimidinic sites. When subsequently visualized by staining with a fluorescent dye, the DNA appears as a 'comet' with a compact 'head' of larger-molecular-weight DNA and a 'tail' of smaller fragments, the size of which depends upon the amount of strand breakage present. Various imaging and software packages are available to quantitate the distribution of DNA in the images. The assay can be extended to detect specific DNA adducts and specific forms of oxidative damage by using damage-specific DNA endonucleases to cleave the DNA at the damage sites, thus creating strand breaks.

Other assays for strand breakage include: alkaline gel electrophoresis in which fragmented genomic DNA separates as a smear instead of the discrete band of high-molecular-weight material observed with intact DNA; alkaline elution, in which the time taken for alkaline solutions of DNA to pass through filters is related to the extent of fragmentation; and alkaline unwinding techniques, in which the fluorescent dye ethidium bromide is used to intercalate into the DNA, causing it to unwind. The unwinding in turn causes a decrease in fluorescence intensity which can be related to DNA fragmentation. As well, the induction of strand breakage in compact supercoiled plasmids causes a conformational change to an open circular form, which can be detected as a molecule with slower mobility by agarose gel electrophoresis.

Adduct Assays

^{32}P-Postlabeling The ^{32}P-postlabeling method is extremely sensitive for detecting chemical adducts and is theoretically sensitive to any DNA-damaging agent. Isolated DNA is first broken down into normal and damaged 3′ monophosphate deoxynucleosides. A 5′ radioactive phosphate residue is then added by the enzyme T4 polynucleotide kinase to allow detection of the deoxynucleosides. The radioactive products are separated by multidirectional anion exchange thin-layer chromatography and the presence of adducts detected and quantitated.

Chemical adducts in DNA can also be detected by a variety of other assays. If the chemical is available as a radioisotope, the presence of adducts may be detected as radioactivity copurifying with DNA from treated cells. The adducts may then be isolated and characterized using the radioactive signal as a tracer. As well, particular types of DNA damage can be recognized by damage-specific antibodies which

can be used to quantitate the amount of damage or isolate the damaged sites for further study. Some adducts can be identified by their altered fluorescent properties or mobilities when separated by high-pressure liquid chromatography or gas chromatography. Further chemical identification may also be made by mass spectrometry.

Unscheduled DNA Synthesis or Repair Replication

Unscheduled DNA synthesis refers to the DNA synthesis which occurs in mammalian cells during the repair of DNA damage by the excision repair pathway. During excision repair, damage is removed in the form of an oligonucleotide which must be replaced by a 'repair patch' synthesized using the opposite undamaged strand as a template. This form of repair can be followed in treated cell cultures by including a labeled nucleotide precursor, such as tritiated thymidine or bromodeoxyuridine, in the culture medium. Tritiated thymidine incorporation can be followed by autoradiography of the labeled nuclei or by liquid scintillation counting while bromodeoxyuridine incorporation can be quantitated by immunostaining with antibromodeoxyuridine antibodies.

The repair of DNA damage can be followed in specific genes and comparisons made in transcribed and nontranscribed strands using a variety of molecular biological techniques. Damage-specific endonucleases can be used to cleave the site of damage which can be detected by its increased mobility during gel electrophoresis and different hybridization patterns. Other techniques including strand-specific quantitative polymerase chain reaction (PCR) in which the amount of product is proportional to the undamaged template and strand-specific ligation PCR which allows sites of damage to be identified with DNA sequencing gels.

Gene Expression Assays

Assays which measure carcinogen-induced changes in gene expression are becoming an increasingly important method for characterizing test chemical effects and a variety of assays have been developed. DNA arrays are available which make the analysis of gene expression changes in hundreds or thousands of genes possible by using gene-specific complementary oligonucleotides arranged on either nylon membranes or glass slides. Several reporter gene vectors have been developed in which the expression of a variety of genes coding for easily measured products (such as firefly luciferase, or green fluorescent protein) is controlled by the promoter sequences of genes important for the control of carcinogenesis. A chemical-induced change in the expression of the reporter gene then indicates possible carcinogenic activity. Nuclease protection assays can also be used to quantitate changes in the expression of specific genes involved in carcinogenesis.

Mutagenicity Assays

In vivo Mutagenesis Assays

The mouse specific locus test The mouse specific locus test is used to detect and quantitate mutations in the germ line of a mammalian species. Animals are chosen for mating which differ at certain easily assayed genetic loci such that mutations will produce an offspring which differs from that expected. Males are usually exposed to the test chemical. Three variations of the method have been used for detecting the induction of point mutations in mouse germ cells: (1) a visible specific locus test examining either five or seven loci; (2) a biochemical specific locus test surveying up to 20 enzymes; and (3) a test for histocompatibility loci mutations.

Transgenic mouse mutagenesis assays Transgenic mouse strains have been developed which have had several copies of either *LacI* or *LacZ* gene sequences from *Escherichia coli* incorporated into their genomes, either as part of a shuttle vector or recombinant λ-phage backbone. Animals are exposed to the test chemical and the DNA isolated from tissues. For shuttle vector-based systems, the shuttle vector DNA is excised, recovered, religated, and introduced into the proper *E. coli* host which are then plated on selective medium allowing growth of mutant colonies. Reconstitution of the λ-based systems is somewhat simpler and more efficient, where the λ-DNA can be cleaved and repackaged into phage using commercially available extracts. The phage are used to infect the proper *E. coli* host, which are then plated in the presence of a selective agent. In a positive selection version of the *LacZ* system, only mutant plaques will grow while mutant plaques in the *LacI* system are identified by a blue/colorless selection procedure. Mutation frequencies can be examined in all somatic tissues as well as germline tissues using these transgenic systems and the defined sequences provide a well-characterized target for sequence analysis of the mutations.

Assays in Bacteria

Ames *Salmonella* test The Ames *Salmonella* test remains the most validated mutagenesis assay for reasons of speed, economy, and dependability. A series of *Salmonella* strains have been developed which possess different mutations in the operon

responsible for histidine synthesis, making it impossible for the bacteria to grow in the absence of histidine. The *Salmonella* test is a reverse mutation assay in which back-mutation of the *his* operon will allow the bacteria to grow in the absence of histidine. Bacteria are exposed to the test chemical with and without a metabolic activation system (usually rat liver S9, a cofactor-supplemented postmitochondrial preparation of rat liver microsomes prepared from rats exposed to Aroclor 1254, a PCB (polychlorinated biphenyl) inducer of the cytochrome P450 hydroxylation enzymes) and spread on to minimal medium agar plates. After a suitable period of incubation, revertant colonies are counted and compared to those found on control plates containing either vehicle alone or a known mutagen. Several strains with different sensitivities are available with TA98 and TA100 detecting frameshift and basepair mutations, respectively, while strains TA102 and TA104 respond to oxidative damage. Other frameshift-specific strains such as TA1535 and TA1537 are also available. Recently, related strains have been developed which are each reverted by specific basepair changes, allowing a mutational spectrum of test chemical activity to be generated. Specialized strains have also been created which express various metabolic activation enzymes to examine specific classes of chemicals. A luminescent version of the assay is also commercially available which substitutes luminescence in place of colony formation as a measure of revertant cell growth.

***E. coli* tryptophan (*trp*) reversion assay** The *E. coli trp* reversion assay using *E. coli WP2* and related strains is a microbial assay which measures trp^- to trp^+ reversion induced by mutagenic chemicals. Various strains are available which are deficient in different aspects of DNA repair, making them more susceptible to reversion. As well, both frameshift and basepair mutations can be detected using different strains. The assays are carried out in a similar manner to the Ames *Salmonella* test.

Assays in Somatic Cells in Culture

Mammalian cell culture systems may also be used to detect mutations induced by chemical substances. Widely used cell lines include L5178Y mouse lymphoma cells as well as the Chinese hamster ovary (CHO) and V-79 lines of Chinese hamster cells. The most commonly used systems measure mutation at the thymidine kinase (TK, L5178Y cells), hypoxanthine-guanine-phosphoribosyl transferase (HGPRT, CHO, and V-79 cells) and Na^+/K^+ ATPase (V-79) genetic loci. The TK and HGPRT mutational systems detect basepair mutations, frameshift mutations, and small deletions, while the Na^+/K^+ ATPase assay detects only basepair mutations.

Cells are exposed to the test agent with and without metabolic activation for a suitable period of time. They are then subcultured to determine cytotoxicity and to allow phenotypic expression prior to mutant selection. The cultures are analyzed for mutant frequency at the end of the expression time by seeding known numbers of cells in medium with and without a selective agent which is lethal to the parental cells but ineffective against the mutant cells. After a suitable incubation time, cell colonies are counted. The number of mutant colonies in selection medium is adjusted by the number of colonies in normal medium to derive the mutant frequency. Cell lines are now available which express a variety of metabolic enzyme activities introduced by genetic engineering. These cell lines can be used to probe the effects of various metabolic activities on carcinogen activation and to avoid the need for adding an exogenous metabolic activation system.

Fungal Mutagenesis Assays

A number of fungal systems have been developed for mutagenicity assays, including *Aspergillus nidulans* and *Neurospora crassa*, which can be used to detect both forward and reverse gene mutations by nutritional, biochemical, or morphological changes in the treated population. Similar methods have been developed using the yeast *Saccharomyces cerevisiae* to measure forward and reverse mutations as well as mitotic gene conversion, the nonreciprocal exchange of DNA between sister chromatids.

Characterization of Mutations

Mutations in specific DNA sequences can be detected by heteroduplex analysis. Heteroduplexes are formed when wild-type and mutated sequences are annealed to form double-stranded mismatched structures. Heteroduplexes can be distinguished from perfectly matched homoduplexes by techniques such as denaturing gradient gel electrophoresis, single-stranded conformation polymorphism, constant denaturant gel electrophoresis, chemical mismatch cleavage, Rnase A cleavage, and cleavage with bacteriophage resolvases. The ligase chain reaction is one method which allows the detection of specific DNA sequences and point mutations. Primers to specific wild-type and mutated gene sequences are designed such that they anneal side by side when the complementary sequence is present. The primers are then ligated together and subsequent cycles of annealing and ligating amplify the amount of product produced, which can then be detected by several means.

Chromosomal Analysis

Detection of Chromosomal Aberrations in Bone Marrow

Assays for chromosomal aberrations can detect changes in either the structure or number of chromosomes. However, in practice, cells are analyzed at the first mitosis after treatment, and therefore number changes are not usually found. Chromosome preparations are made from the bone marrow cells of exposed animals, stained and examined under a microscope for metaphase cells which are scored for chromosomal aberrations. A more recent development is fluorescent *in situ* hybridization (FISH) in which chromosomal aberrations are detected with fluorescent-labeled DNA probes which hybridize to specific chromosomal regions.

Micronucleus Assay

The micronucleus test is a mammalian *in vivo* and *in vitro* test which detects chromosomal or mitotic apparatus damage. Micronuclei are small cytoplasmic particles consisting of acentric fragments of chromosomes or entire chromosomes, which become separated from the nuclei of daughter cells. Polychromatic erythrocytes in the bone marrow of rodents form a useful model for this assay since the nuclei are lost during development from erythroblasts. The lack of a nucleus makes the scoring of micronuclei, either by microscope or flow cytometry, easier in these cells.

Rodent Dominant Lethal Assay

A dominant lethal mutation is one which occurs in a germ cell and results in the death of the fertilized egg or developing embryo. Dominant lethals are generally accepted to be the result of both structural and numerical chromosomal aberrations. In this assay, male animals are usually exposed to the test substance and mated to untreated females. The females are sacrificed after an appropriate period of time and the numbers of live and dead embryos in the uteri determined. The ratio of dead to live embryos from the treated group compared to the ratio of dead to live embryos from the control group is used as a measure of dominant lethality. Several treatment protocols are available. The most widely used require either single administration of the test substance or treatment on five consecutive days.

The Rodent Heritable Translocation Assay

In this assay, males are treated with the test chemical, mated, and the male progeny screened either for chromosomal translocations, or for sterility, which is indicative of translocations between nonhomologous chromosomes in the treated male gametes.

Sister Chromatid Exchange Assay

The sister chromatid exchange (SCE) assay detects the ability of a chemical to enhance the exchange of DNA between two sister chromatids of a duplicating chromosome. The test may be performed *in vitro*, using continuous cell lines, or *in vivo* using animal models such as mice, rats, and hamsters. For *in vitro* assays, cell cultures are exposed to test chemicals, and allowed to replicate in the presence of bromodeoxyuridine. The cells are then treated with colchicine or colcemid to arrest cells in a metaphase-like stage of mitosis, and following harvesting chromosome preparations made. Preparations are stained and metaphase cells analyzed for SCEs using a microscope.

Transformation Assays

Both *in vivo* and *in vitro* assay systems have been developed which assay chemicals for their abilities to induce morphological transformation, changing the morphology of cells from normal or near-normal monolayer growth-inhibited cultures to multilayered altered foci resembling neoplastic cell growth. Depending upon the assay, test chemicals may be tested for activity as complete carcinogens, tumor initiators, or tumor promoters.

Rodent Skin *in vivo* Assay

When chemicals are tested for activity as complete carcinogens in the rodent skin transformation assay, they are applied repetitively over several weeks to a shaved area on the backs of the animals, and the area observed for tumor formation. Tumor initiators can be detected by a single application followed by repetitive application of a known tumor promoter such as benzoyl peroxide or chrysarobin. Similarly, tumor promoters can be assayed by first applying a known initiator such as 3-methylcholanthrene to the shaved area followed by repeated applications of the test chemical.

In vitro methods for assaying cellular transformation include: the Syrian hamster embryo (SHE) cell transformation assay, using primary cultures derived from 2-week-old SHEs, epithelial cells, the BALB/c-3T3 and C3H-10T½ cell lines and a number of cells which have been infected by a transforming virus, such as Rauscher leukemia virus-infected Fisher rat embryo cells or simian adenovirus SA7-infected SHE cells. While the SHE cells maintain may of their metabolic capabilities, many of the other cell types require the addition of a metabolic activation

system such as primary hepatocytes and microsomal fractions. Many of these systems can also be used for identifying tumor promoters by first treating the cells with a low concentration of known initiator.

Gap Junctional Communication

Adjacent cells form contact points called gap junctions which allow the exchange of small molecules in a form of communication. The exposure of cells to carcinogens causes a blockage of gap junctional communication. Presumably the loss of communication with neighbors alters cells in such a way that they act in isolation and ignore neighboring signals to limit growth, a process thought to be involved in the development of carcinogenesis. The activity of gap junctions can be measured by several systems.

Metabolic Cooperation

Wild-type and 6-thioguanine-resistant V79 Chinese hamster cells are cultured together in the presence of 6-thioguanine. With active gap junctions, the lethal metabolites formed from 6-thioguanine in the wild-type cells can disperse into the otherwise resistant cells, causing cell death. If gap-junctional communication is blocked, treated resistant cells survive to form colonies, which can be enumerated.

Dye Transfer

The fluorescent dye lucifer yellow, a fluorescent dye, is introduced into cells either by preloading or through a scratch in the monolayer. In the presence of intercellular communication, the dye spreads to neighboring cells in the culture and the fluorescent area increases. When chemical treatment causes gap junctions to be blocked, fluorescence remains concentrated at the original site.

Gene Expression

The blockage of intracellular communication can be reflected by the altered expression of mRNA coding for gap junction components, such as connexin 43, which is involved in gap junction formation.

Peroxisome Proliferation

Peroxisomes are cytoplasmic organelles responsible for the oxidative degradation of fatty acids and other molecules. It has been observed that several nonmutagenic rodent liver carcinogens, including some lipid-lowering drugs and plasticizers, cause liver enlargement and the proliferation of numbers of peroxisomes within hepatocytes. The enumeration of peroxisomes in liver tissue is sensitive for several nongenotoxic rodent hepatocarcinogens but peroxisome proliferation does not appear to occur in humans and its relevance is not known.

In vivo Assays for Tumor Formation

A number of animal models have been used to assess test chemical carcinogenicity. However, considerations of time, cost, and ease of handling have favored the use of rodents instead of other animals such as dogs or primates since assays require animals to be exposed to the test agent over a significant portion of their lifespan. Although animal models provide the most complete and dependable surrogates for the identification of human carcinogens, they are not entirely representative of the human situation due to a number of factors, including species differences in metabolism and susceptible tissues, differences in time and route of exposure as well as amount of chemical tested, and lack of exposure to other chemicals which may act as cocarcinogens. The interpretation of results obtained when chemicals are tested in rodents at the maximum tolerated dose (MTD, the largest dose possible which does not shorten the lifespan or produce clinical signs of toxicity) may be difficult to relate to human exposures. Many chemicals identified as rodent carcinogens have been found to cause inflammation and oxidative damage leading to tissue death and regeneration only at these elevated levels far removed from any expected human exposure.

Rodent Carcinogenicity Assay

The rodent carcinogenicity assay remains the most reliable and validated system for the identification of possible human carcinogens. With minor variations, the methods used for the rodent carcinogenicity assay have been standardized both nationally and internationally, and by law all new chemicals introduced for human use must be tested for long-term effects using the assay. In order to address possible species differences in carcinogenic response, a chemical is tested using two species, usually rats and mice, for comparison, although hamsters may sometimes be substituted. Animals of both sexes are treated with increasing doses of chemical, which are included in the diet or given orally by gavage for food-related studies. Each dose group may contain 50 or more animals and the dosing is maintained over most of the animal's remaining lifespan (18 months for mice and 2 years for rats), after which tissues are collected at necropsy and examined by a pathologist for malignant tumors and preneoplastic lesions.

The duration of the study and number of animals required make the rodent carcinogenicity assay an extremely expensive and time-consuming undertaking. Several other *in vivo* assays have been developed

to reduce the number of animals, length of time, and costs involved. At present, none of these tests has been sufficiently validated to replace the full carcinogenicity test.

Neonatal Mouse Assay

The exposure of newborn mice to carcinogens has been found to provide a sensitive assay for tumor production, primarily in the liver. Since the newborn liver is rapidly replicating, tumors develop relatively quickly compared to adults. Neonatal mice are exposed to two treatments of test chemical in the first and second weeks after birth, and allowed to grow for 1 year, when the study is ended. This assay is sensitive to genotoxic chemicals which form adducts with DNA.

Transgenic Mouse Carcinogenicity Assays

Several mouse models which display faster tumor formation compared to parental strains have been created through genetic engineering. These include the $p53^{+/-}$ heterozygous mouse model which contains one active copy of the wild-type *p53* tumor suppressor gene, the CB6F1-TG-rasH2 hemizygous mouse model containing five to six copies of the human c-Ha-ras oncogene, the Tg.AC transgenic mouse model produced by introducing the *v-Ha-ras* oncogene linked to a fetal zeta-globin promoter and containing 40 copies *v-Ha-ras* per allele, and the $XPA^{-/-}$ knockout mouse, which lacks the XPA component of the nucleotide excision repair pathway. Each treatment group can contain 15 animals and studies are continued for 26 weeks, after which the background of spontaneously occurring mutations begins to increase. The $p53^{+/-}$ and $XPA^{-/-}$ strains generally respond to genotoxic carcinogens while the CB6F1-TG-rasH2 and Tg.AC transgenic strains can respond to both genotoxic and nongenotoxic carcinogens. Bigenic strains combining more than one transgenic alteration, such as $XPA^{-/-}$. $p53^{+/-}$ and $p53^{+/-}$. Tg.AC, have also been produced to increase the responsiveness of the assays.

Preneoplastic Foci

Preneoplastic foci are small, contained proliferative lesions that precede the development of either benign or malignant neoplastic tumors. They may be detected as proliferating regions by measuring uptake of bromodeoxyuridine or tritiated thymidine into the DNA by immunohistochemistry or autoradioraphy, respectively. Preneoplastic foci can also be identified by the use of immunohistochemistry to measure expression of proliferating cell nuclear antigen or the placental form of glutathione *S*-transferase. Preneoplastic foci can be enumerated in various tissues, but the liver is examined most frequently. Tissues can be examined for preneoplastic foci at the close of the rodent carcinogenicity assay or animal studies can be carried out specifically for this endpoint. More preneoplastic foci are usually found than would form eventual tumors, a finding which allows assessments to be carried out in a shorter time period with fewer animals compared with the complete rodent carcinogenicity assay.

Assays for Monitoring Human Population Exposures

Obviously, techniques for monitoring human exposure to carcinogenic compounds must be relatively noninvasive and a number of assays have been applied to samples of blood and urine. The presence of DNA adducts and oxidized DNA bases or nucleosides such as 8-hydroxyguanine or 8-hydroxyguanosine can be detected in urine and the DNA of leukocytes (white blood cells) using specific antibodies or instrumental analysis. The comet assay can be used to detect strand breakage in leukocyte DNA and levels of repair replication can be measured as well. The ^{32}P-postlabeling assay has proven to be very sensitive to examine adduct formation. Levels of mutagenesis in lymphocytes can be detected using a variation of the HGPRT mutagenesis assay and chromosome aberrations and sister chromatid exchange can also be assayed.

See also: **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **Mass Spectrometry**: Principles and Instrumentation; Applications; **Mutagens**; **Nucleic Acids**: Properties and Determination; Physiology; **Pesticides and Herbicides**: Toxicology

Further Reading

Clayson DB (2001) *Toxicological Carcinogenesis*. Boca Raton: Lewis.

Douglas JF (1984) *Carcinogenesis and Mutagenesis Testing*. Clifton: Human Press.

Franks LM and Teich NM (eds) (1997) *Introduction to the Cellular and Molecular Biology of Cancer*, 3rd edn. Oxford: Oxford University Press.

Friedberg EC, Walker GC and Siede W (1995) *DNA Repair and Mutagenesis*. Washington, DC: ASM Press.

Henderson DS (1999) *DNA Repair Protocols. Methods in Molecular Biology*, vol. 113. Totowa: Humana Press.

Kilbey BJ, Legator M, Nichols W and Ramel C (1984) *Handbook of Mutagenicity Test Procedures*. Amsterdam: Elsevier Science.

Kitchin KT (ed.) (1999) *Carcinogenicity: Testing, Predicting, and Interpreting Chemical Effects*. New York: Marcel Dekker.

Klug WS and Cummings MR (2000) *Concepts of Genetics*, 6th edn. Upper Saddle River: Prentice Hall.
Milman HA and Weisburger EK (eds) (1994) *Handbook of Carcinogen Testing*, 2nd edn. Park Ridge: Noyes Publications.
Pfeifer GP (ed.) (1996) *Technologies for Detection of DNA Damage and Mutations*. New York: Plenum Press.
Twyman RM (1998) *Advanced Molecular Biology: A Concise Reference*. Oxford: BIOS Scientific.

CAROTENOIDS

Contents

Occurrence, Properties, and Determination

D B Rodriguez-Amaya, Universidade Estadual de Campinas, Campinas, SP, Brazil

Introduction

Current knowledge on the structures, distribution, and properties of food carotenoids is presented. Their role as food colorants and changes during processing and storage of foods are discussed. Trends in the inherently difficult analysis of these compounds are described, pointing out sources of errors, as well as means to guarantee the reliability of results.

Structures and Occurrence in Foods

Among the naturally occurring pigments, carotenoids are notable for their wide distribution, structural diversity, and varied functions and actions. In nature, about 100 million tonnes of these compounds are produced annually. More than 600 carotenoids, exclusive of *cis* and *trans* isomers, have now been isolated and characterized from natural sources. This remarkable number includes the enormous variety of carotenoids in algae, bacteria, yeast, and fungi. In foods, the number is much more restricted, but the carotenoid composition can still be complex.

Food carotenoids are generally C_{40} tetraterpenoids formed from eight C_5 isoprenoid units joined head to tail, except at the center, where a tail-to-tail linkage reverses the order and results in a symmetrical molecule. Lateral methyl groups near the center are separated by six carbon atoms and the others by five (**Figure 1**). An important feature is a centrally located, extended double-bond system. The π electrons are highly delocalized over the entire polyene chain, although electron density appears to be greater at or towards the end of the chain. The basic skeleton may be modified in many ways, such as cyclization, hydrogenation, dehydrogenation, introduction of oxygen functions, rearrangement, chain shortening, or combinations thereof, resulting in an immense array of structures.

A semisystematic nomenclature that conveys structural information has been devised for carotenoids. In this chapter, the semisystematic names are given in parentheses in the figures and table; those of other carotenoids will be cited with the trivial names when first mentioned in the text. Thereafter, only the better known trivial names will be used.

Hydrocarbon carotenoids (e.g., β-carotene, lycopene) are known as carotenes; oxygenated derivatives are called xanthophyls. Common oxygen substituents are: hydroxy (as in β-cryptoxanthin), keto (as in canthaxanthin), epoxy (as in violaxanthin), and aldehyde (as in β-citraurin) groups. Carotenoids may be acyclic (e.g., ζ-carotene, lycopene), monocyclic (e.g., γ-carotene), or bicyclic (e.g., α- and β-carotene). Cyclization is limited to the formation of a six-membered (occasionally five-membered) ring at one or both ends of the molecule.

In nature, carotenoids exist primarily in the more stable all-*trans* (or all-*E*) form, but *cis* (or *Z*) isomers do occur. The first C_{40} compound in the biosynthetic pathway has the 15-*cis* configuration in most natural sources. Bixin, the principal pigment of the colorant annatto, also occurs naturally in the *cis* form (**Figure 2**) Small amounts of *cis* isomers of other carotenoids have been increasingly reported. Commonly encountered *cis*-β-carotenes are shown in **Figure 3**.

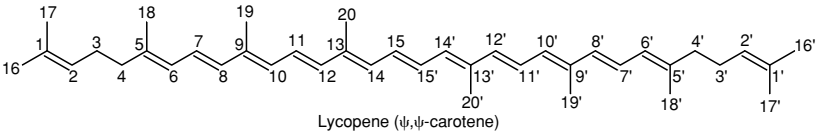

Lycopene (ψ,ψ-carotene)

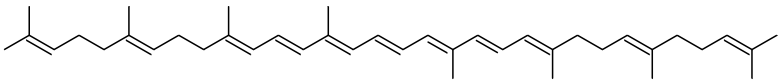

ζ-Carotene (7,8,7',8'-tetrahydro-ψ,ψ-carotene)

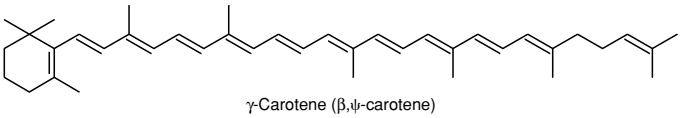

γ-Carotene (β,ψ-carotene)

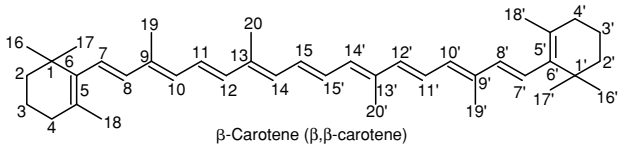

β-Carotene (β,β-carotene)

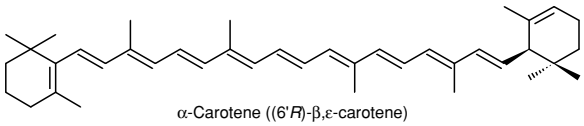

α-Carotene ((6'*R*)-β,ε-carotene)

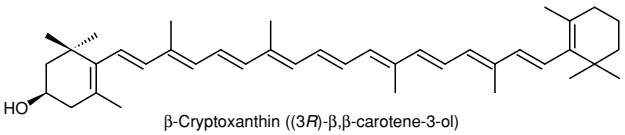

β-Cryptoxanthin ((3*R*)-β,β-carotene-3-ol)

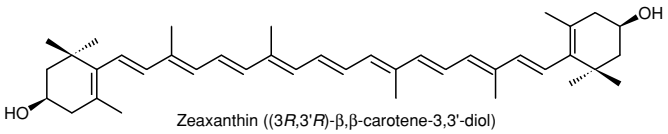

Zeaxanthin ((3*R*,3'*R*)-β,β-carotene-3,3'-diol)

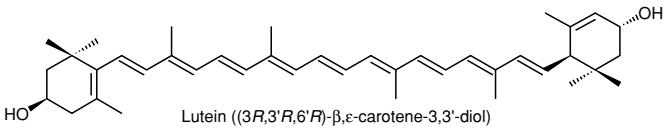

Lutein ((3*R*,3'*R*,6'*R*)-β,ε-carotene-3,3'-diol)

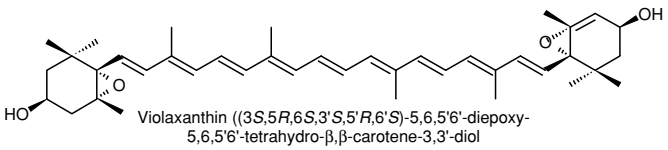

Violaxanthin ((3*S*,5*R*,6*S*,3'*S*,5'*R*,6'*S*)-5,6,5'6'-diepoxy-5,6,5'6'-tetrahydro-β,β-carotene-3,3'-diol

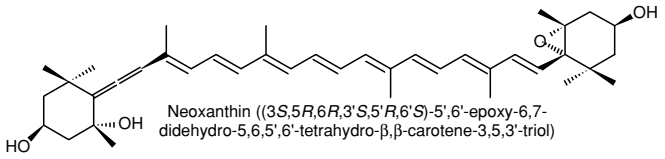

Neoxanthin ((3*S*,5*R*,6*R*,3'*S*,5'*R*,6'*S*)-5',6'-epoxy-6,7-didehydro-5,6,5',6'-tetrahydro-β,β-carotene-3,5,3'-triol)

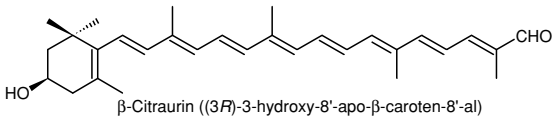

β-Citraurin ((3*R*)-3-hydroxy-8'-apo-β-caroten-8'-al)

Figure 1 Structures of common carotenoids in foods of plant origin.

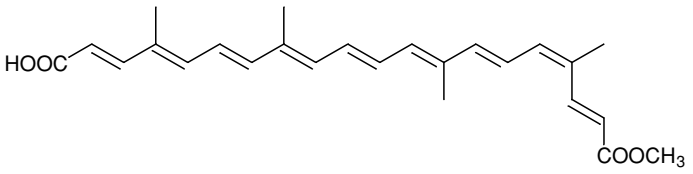

Bixin (methyl hydrogen 9'-*cis*-6,6'-diapocarotene-6,6'-dioate)

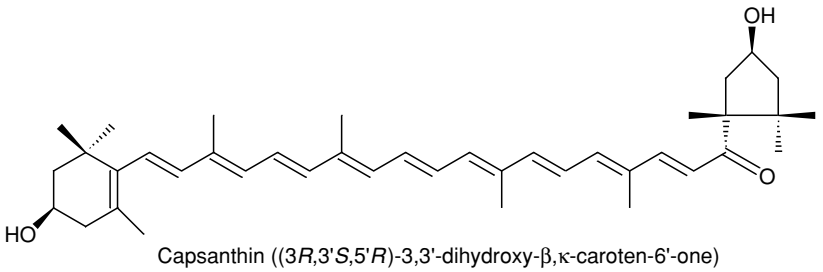

Capsanthin ((3*R*,3'*S*,5'*R*)-3,3'-dihydroxy-β,κ-caroten-6'-one)

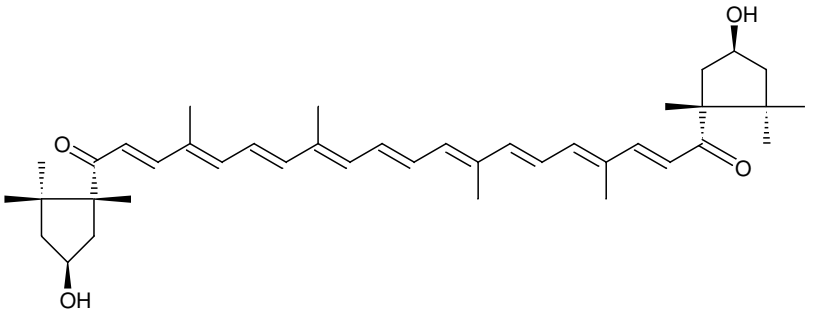

Capsorubin ((3*S*,5*R*,3'*S*,5'*R*)-3,3'-dihydroxy-κ,κ-carotene-6,6'-dione)

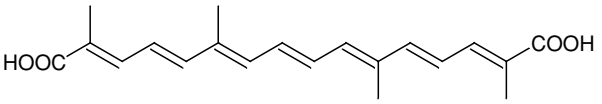

Crocetin (8,8'-diapocarotene-8,8'-dioic acid)

Figure 2 Principal carotenoids of natural extracts used as food colorants.

The major structural differences that account for variations in properties and biological activities of carotenoids, therefore, are: length and rigidity of the molecule, length of the conjugated double-bond system, cyclized or acyclic nature of the end groups, and presence of polar substituents in the predominantly hydrocarbon molecule.

Although masked by the green chlorophyls, carotenoids are found universally in the chloroplasts of photosynthetic tissues. Leaves of all species studied revealed the presence of the same major carotenoids: lutein, β-carotene, violaxanthin, and neoxanthin. Small amounts of α-carotene, α-cryptoxanthin ((3′*R*,6′*R*)-β,ε-caroten-3′-ol) or β-cryptoxanthin, zeaxanthin, antheraxanthin ((3*S*,5*R*,6*S*,3′*R*)-5,6-epoxy-5,6-dihydro-β,β-carotene-3,3′-diol), and lutein-5,6-epoxide ((3*S*,5*R*,6*S*,3′*R*,6′*R*)-5,6-epoxy-5, 6-dihydro-β,ε-carotene-3,3′-diol) may also be found. The xanthophyls are unesterified, and the relative ratios are fairly constant, but the absolute concentrations vary considerably. Lettuce also has lactucaxanthin ((3*R*,6*R*,3′*R*,6′*R*)-ε,ε-carotene-3,3′-diol) at levels comparable with, or slightly higher than, neoxanthin.

In ripe fruits carotenoids are located in chromoplasts and the hydroxycarotenoids are mostly esterified with fatty acids. The composition varies dramatically from fruit to fruit, but eight major patterns can be discerned: (1) insignificant levels of carotenoids; (2) small amounts generally of chloroplast carotenoids (e.g., grape); (3) considerable amounts of lycopene (e.g., tomato, watermelon, red-fleshed guava and papaya); (4) predominance of β-carotene and/or β-cryptoxanthin (e.g., apricot, peach, loquat); (5) large amounts of epoxides (e.g., mango,

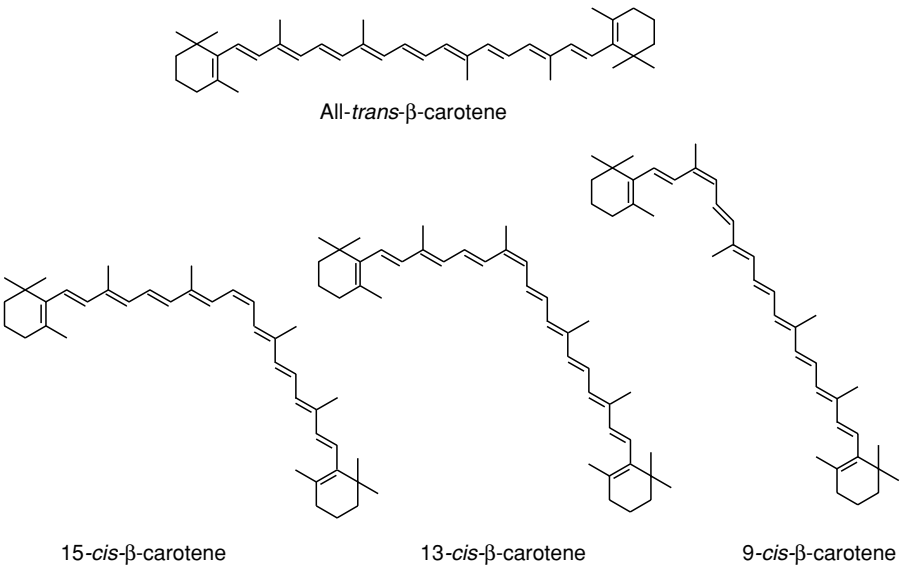

Figure 3 Common geometrical isomers of β-carotene.

carambola); (6) preponderance of unusual or species-specific carotenoids (e.g., red pepper); (7) substantial amounts of poly-*cis*-carotenoids (e.g., tangerine tomato); (8) significant levels of apocarotenoids (carotenoids with shortened carbon skeleton) (e.g., citrus species). Some merging of these patterns can be seen in some fruits.

In the few carotenogenic roots (e.g., carrot, sweet potato), the carotenes are preponderant. In corn (seed), the xanthophyls predominate.

Since plants are able to synthesize carotenoids *de novo*, the composition of plant foods is enriched by the presence of trace amounts of biosynthetic precursors along with derivatives of the main components (Figure 4, Table 1). In a given food, compositional variations occur as a consequence of varietal differences, climatic conditions, agricultural variables, stage of maturity, postharvest handling, conditions during storage, and transportation. Ripening of fruits is generally accompanied by enhanced carotenogenesis, the carotenoids increasing in number and quantity.

Carotenoids are not as widely distributed in animal foods, and the total content is much less. Incapable of carotenoid biosynthesis, animals depend on dietary carotenoids, which are selectively or unselectively absorbed, converted to vitamin A, deposited as such or slightly altered to form carotenoids typical of animal species (Figure 5). Astaxanthin is the major carotenoid in most crustaceans, either free, esterified, or as carotenoprotein complexes. β-Carotene, echinenone and canthaxanthin are the other pigments usually encountered. When found in fish, carotenoids are located in the skin and flesh. The xanthophyls predominate over the carotenes; astaxanthin is the most common, followed by lutein (dominant in freshwater fishes) and tunaxanthin (ε,ε-carotene-3,3′-diol) (characteristic of marine fishes).

Avian species preferentially accumulate certain xanthophyls, which provide the color of eggs, body, skin, and fat. Cattle absorb β-carotene efficiently, but not xanthophyls; thus, diet-derived β-carotene predominates in milk.

Physical and Chemical Properties

The conjugated double-bond system constitutes the light-absorbing chromophore that gives carotenoids their attractive color and provides the visible absorption spectrum, which serves as the basis for their identification and quantification. A few examples of the structure–spectrum relationship will be cited, referring to carotenoids in petroleum ether. Most carotenoids absorb maximally (λ_{max}) at three wavelengths. The greater the number of conjugated double bonds, the higher the λ_{max} values. The acyclic lycopene, with 11 conjugated double bonds, is red and absorbs at the longest wavelengths (λ_{max} at 444, 470, and 502 nm). At least seven such bonds are needed for a carotenoid to have a perceptible color. Thus, ζ-carotene is light yellow, and, being also acyclic, its spectrum has three well-defined peaks but at much shorter wavelengths (378, 400, and 425 nm). The two carotenoids that precede ζ-carotene in the desaturation biosynthetic pathway, phytoene (three conjugated double bonds) and phytofluene (five conjugated double bonds), are colorless and absorb maximally at 276, 286, and 297 nm and 331, 348, and 367 nm, respectively. Cyclization results in steric

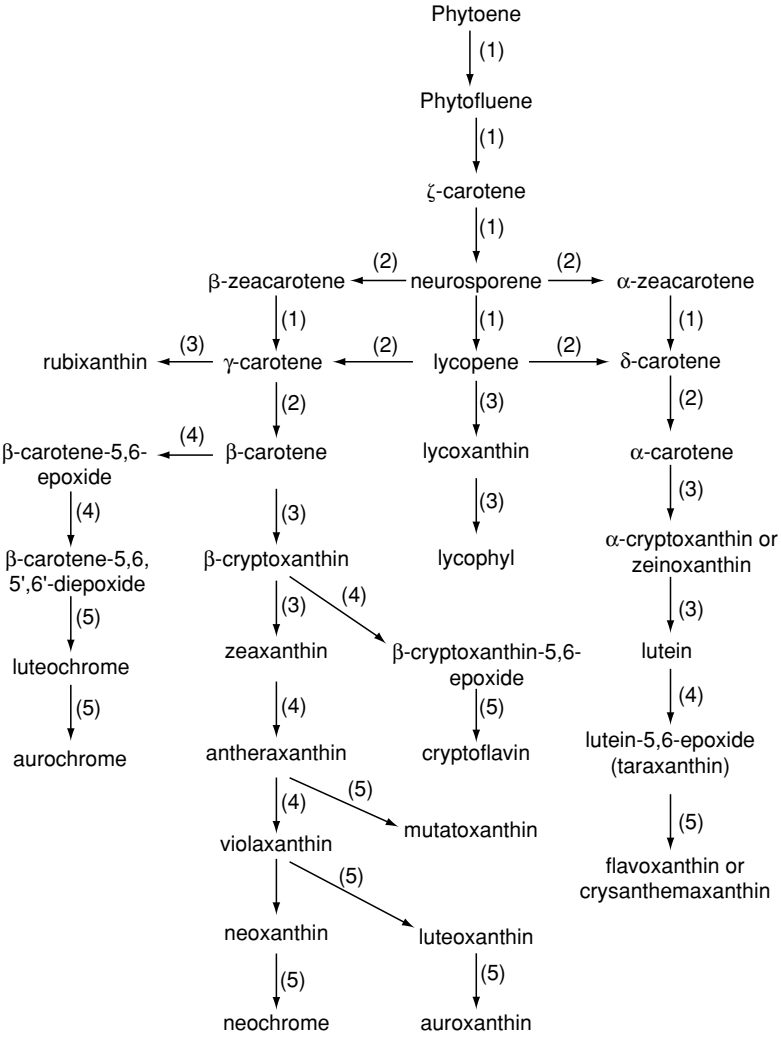

Figure 4 Later stages of carotenoid biosynthesis and possible transformations of carotenoids. Reactions: (1) desaturation, (2) cyclization, (3) hydroxylation, (4) epoxidation, and (5) epoxide–furanoxide rearrangement. From Rodriguez-Amaya DB (1993) Nature and distribution of carotenoids in foods. In: Charalambous G (ed.) *Shelf Life Studies of Foods and Beverages. Chemical, Biological, Physical and Nutritional Aspects*, pp. 547–587. Amsterdam: Elsevier Science, with permission.

hindrance between the ring methyl group at C-5 and the hydrogen at C-8 of the polyene chain, taking the π electrons of the ring double bond out of plane with those of the chain. Consequently, a hypsochromic shift, hypochromic effect, and loss of spectral fine structure are observed. Thus, bicyclic β-carotene, although possessing the same number of bonds as lycopene, is yellow–orange and has a λ_{max} at 450 and 477 nm and a mere inflection at 425 nm. Monocyclic γ-carotene is red–orange and exhibits a spectrum intermediate between those of lycopene and β-carotene. The double bond in the ε-ring of α-carotene is out of conjugation, so it is light yellow, and the absorption peaks are slightly more defined and at slightly shorter wavelengths (422, 445, and 473 nm) compared to β-carotene. Conjugated carbonyl groups cause a bathochromic shift and loss of fine structure to the extent that the three-maxima spectrum is replaced by a broad curve, asymmetrical with a maximum at 458 nm and a shoulder at 482 nm for echinenone (orange) and symmetrical with the maximum at 466 nm for canthaxanthin (red–orange). Hydroxy substituents result in virtually no change in color and absorption spectrum. Thus, lutein resembles α-carotene, and β-cryptoxanthin and zeaxanthin are similar to β-carotene in spectral properties. *Cis* isomerization of one of the chromophore's bond causes slight loss in color, small hypsochromic shift and hypochromic effect, accompanied by the appearance of a *cis* peak in the ultraviolet region, the intensity of which is greater as the *cis* double bond approaches the center of the chromophore. The 5,6-mono- and

Table 1 Carotenoid distribution in some foods

Food	*Major carotenoid concentration (μg/g)*	*Other carotenoids*
Carrot	β-Carotene (40–120) α-Carotene (20–50)	ζ-Carotene, γ-carotene, δ-carotene, lycopene, lutein, neurosporene, β-zeacarotene
Egg yolk	Lutein (6)	Zeaxanthin, isolutein, β-cryptoxanthin, β-carotene, α-carotene, neoxanthin, ζ-carotene
Guava (pink-fleshed)	Lycopene (47–60)	β-Carotene, ζ-carotene, γ-carotene, zeinoxanthin, β-cryptoxanthin, β-carotene-5,6,5′,6′-diepoxide, phytofluene, rubixanthin, cryptoflavin, lutein, neochrome
Loquat	β-Carotene (6–8) β-Cryptoxanthin (5–7)	γ-Carotene, neurosporene, lutein, violaxanthin, neoxanthin, β-cryptoxanthin-5,6-epoxide, phytofluene, ζ-carotene, luteoxanthin, neochrome
Mango	β-Carotene (6–75) Violaxanthin (18–22)	Luteoxanthin, neoxanthin, zeaxanthin, auroxanthin, α-cryptoxanthin, β-cryptoxanthin, ζ-carotene, phytoene, phytofluene, antheraxanthin
Papaya (red-fleshed)	Lycopene (21–40) β-Cryptoxanthin (7–10)	β-Carotene, ζ-carotene, β-cryptoxanthin-5,6-epoxide, γ-carotene, β-zeacarotene, cryptoflavin, antheraxanthin
Peach	β-Cryptoxanthin (4–6)	β-Carotene, ζ-carotene, violaxanthin, luteoxanthin, auroxanthin, lutein, zeaxanthin
Tomato	Lycopene (30–50)	Phytoene, phytofluene, β-carotene, ζ-carotene, γ-carotene, neurosporene
Tree tomato	β-Cryptoxanthin (14) β-Carotene (8)	ζ-Carotene, β-carotene-5,6-epoxide, lutein, zeaxanthin

Semisystematic names: neurosporene (7,8-dihydro-ψ, ψ-carotene); δ-carotene ((6*R*)-ε,ψ-carotene); β-zeacarotene (7′,8′-dihydro-β, ψ-carotene); isolutein ((3*S*, 5*R*, 6*S*, 3′*S*,6′*R*)-5,6-epoxy-5,6-dihydro-β,ε-carotene-3,3′-diol); zeinoxanthin ((3*R*,6′*R*)-β,ε-caroten-3-ol; β-Carotene-5,6,5′,6′-diepoxide (5,6,5′,6′-diepoxy-5,6,5′,6′-tetrahydro-β,β-carotene); rubixanthin ((3*R*) -β,ψ-caroten-3-ol); β-cryptoxanthin-5,6-epoxide (5,6-epoxy-5,6-dihydro-β,β-caroten-3-ol); cryptoflavin (5,8-epoxy-5,8-dihydro-β,β-caroten-3-ol); neochrome (5′,8′-epoxy-6,7-didehydro-5,6,5′,8′-tetrahydro-β,β-caroten-3,5,3′-triol); auroxanthin ((*3S*,*5R*,*8RS*, 3′*S*,*5′R*,*8′RS*)-5,8,5′,8′ diepoxy-5,8,5′,8′-tetrahydro-β,β-caroten-3,3′-diol); luteoxanthin (5,6,5′,8′-diepoxy-5,6,5′,8′-tetrahydro-β,β-carotene-3,3′-diol); phytoene (7,8,11,12,7′,8′,11′,12′-octahydro-ψ,ψ-carotene); phytofluene (7,8,11,12,7′,8′-hexahydro-ψ,ψ-carotene).

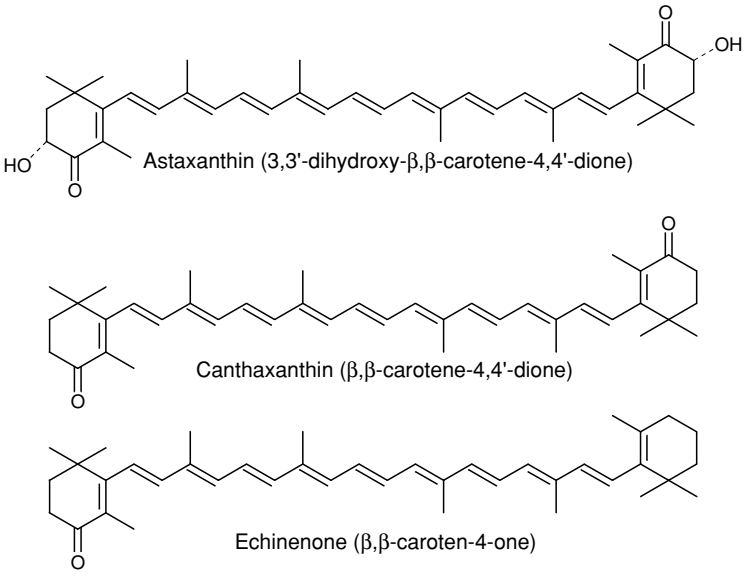

Figure 5 Some carotenoids typical of foods of animal origin. Astaxanthin occurs as a mixture of the (3*S*, 3′*S*), (3*R*, 3′*R*) and (3*R*, 3′*S*) forms.

5,6,5′,6′-diepoxide, having lost one and two ring double bonds, respectively, absorb maximally at wavelengths some 5 and 10 nm shorter and are lighter colored than the parent compounds. On transformation to the 5,8-furanoid oxide, two double bonds (one ring, one chain) are lost. Thus, the 5,8-mono- and 5,8,5′8′-diepoxide are 20–25 and 50 nm lower, respectively, than those of the parent compounds. Solvent effects can be pronounced; the λ_{max} values are higher relative to petroleum ether by 2–6 nm in acetone, 10–20 in chloroform and dichloromethane, and 18–24 nm in toluene.

The intensity and hues of plant foods depend on which carotenoids are present, their concentrations,

and physical state. In animals, complexation of carotenoids with proteins extends the color to green, purple, blue, or black. A well-known example is the blue carotenoprotein crustacyanin of the lobster carapace, an astaxanthin complex. On denaturation of the protein (e.g., heating), astaxanthin is released, and its vivid red color ensues.

Carotenoids are lipophilic, generally insoluble in water but soluble in organic solvents such as acetone, tetrahydrofuran, ethyl ether, and chloroform. Carotenes dissolve well in hexane and petroleum ether, whereas the more polar xanthophylls dissolve more readily in methanol and ethanol. In plants and animals, carotenoids occur as crystals, in solutions in fat depots, in colloidal dispersions, or combined with proteins in the aqueous phase.

Because of the many double bonds, carotenoids are prone to *trans–cis* isomerization. A large number of *cis* isomers are theoretically possible for each carotenoid. Only a few are actually formed because the *cis* configuration in some double bonds creates steric hindrance between nearby hydrogen atoms and methyl groups, making the isomers unstable. The steric hindrance is small when it occurs between hydrogen atoms, so isomers with *cis* double bonds in this situation are relatively stable and are easily formed (e.g., 9-*cis*, 13-*cis*, 15-*cis*) (**Figure 3**).

The highly reactive, electron-rich polyene chain is also subject to oxidative degradation. Carotenoids apparently have different susceptibilities, ζ-carotene, lutein, and violaxanthin being cited as more labile. In contrast to the wealth of information on the oxidation of lipids, carotenoid oxidation is not well understood. It involves initially epoxidation and the formation of apocarotenoids. The preferred site of epoxidation is the terminal double bond; oxidative chain cleavage appears to commence at the C-7,8 position in β-carotene. Subsequent fragmentations presumably yield low-mass compounds similar to those obtained in fatty acid oxidation.

Also related to the conjugated double-bond system is the antioxidant property attributed to carotenoids more recently. The primary mode of action is quenching of singlet oxygen and interaction with free radicals. The oxygen quenching ability is maximal, with carotenoids having nine or more double bonds. The acyclic lycopene was found to be more effective than the bicyclic β-carotene. In a free radical-initiated system, canthaxanthin and astaxanthin, both with conjugated keto groups, were also shown to be better antioxidants than β-carotene and zeaxanthin. At elevated oxygen pressures, however, carotenoids are reported to act as prooxidants.

Xanthophyls undergo specific group reactions which serve as simple chemical tests in the determination of the structure. For example, primary and secondary hydroxy groups are acetylated by acetic anhydride in pyridine. Allylic hydroxyls, isolated or allylic to the chromophore, are methylated with acidified methanol. Epoxy groups in the 5,6- or 5,6,5′6′-positions are easily detected by acid-catalyzed conversion to the furanoid derivatives. Aldehyde and keto carotenoids undergo reduction with $LiAlH_4$ or $NaBH_4$.

Use as Food Colorants

Carotenoids as food colorants find their way into food products by direct addition or indirectly through an animal's feed. Commercial formulations are of two types: natural extracts and synthetic nature-identical carotenoids.

Annatto, paprika, and saffron as dry powders or extracts have been used for years. Annatto is a series of red coloring preparations all based on the extracts of *Bixa orellana* seeds, where the pigments are concentrated in the thin seed coat. The apocarotenoid bixin (**Figure 2**) is the main component of oil-soluble formulations, and its saponification product, norbixin, the major coloring matter of water-soluble preparations. Oleoresin of paprika is the oil extract of *Capsicum annuum*, which imparts a pinkish yellow to crimson red color to foods, the predominating pigments being capsanthin and capsorubin. Saffron consists of the dried stigma of *Croccus sativus* and is used as a spice and yellow coloring agent. It contains mainly crocin, the digentiobioside of crocetin, a diapocarotenedioic acid. Other commercial sources of carotenoids are lutein-rich marigold petals for poultry feed and β-carotene-rich microalgae. Industrial production of natural carotenoids by biotechnology is gaining more interest. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments.)

The first carotenoid prepared by chemical synthesis, β-carotene, was introduced commercially by Roche in 1954. It was followed by β-apo-8′-carotenal (8′-apo-β-caroten-8′-al) in 1960, β-apo-8′-carotenoic acid ethyl ester (ethyl 8′-apo-β-caroten-8′-oate) in 1962, and canthaxanthin in 1964. In 1968, BASF introduced citranaxanthin (5′,6′-dihydro-5′-apo-18′-nor-β-caroten-6′-one), and in 1984, Roche launched (3*RS*, 3′*RS*)-astaxanthin as feed additives. Crystalline carotenoids suffer from problems that render their commercialization in this form impractical: instability, insolubility in water, and limited solubility in fats and oils. To satisfy the needs of the food industry, special application forms have been developed through sophisticated physicochemical operations. Micronized oil suspensions are the major marketable

forms for coloring fat-based foods. For water-based foods, water-dispersible emulsions or colloidal preparations are available.

Advantages cited for carotenoids as food colorants are: natural connotation, high tinctural potency, unaffected by reducing conditions, noncorrosive, good stability in the pH range of most food products, provitamin A activity, and other beneficial effects on human health. Their disadvantages are: limited color range, higher price compared with synthetic dyes, sensitivity to oxidative degradation, solubility problems. The last two disadvantages have been overcome in the application forms mentioned above.

Stability on Processing and Storage

Naturally present or added carotenoids are subject to isomerization and oxidation during food processing and storage. Isomerization to *cis* isomers, provoked by the release of constituent acids during slicing and pureeing of foods, heat treatment and exposure to light, results in some loss of color and vitamin A activity. Oxidation depends on: availability of oxygen; the carotenoids present and their physical state; water activity; presence of antioxidants (e.g., tocopherols and ascorbic acid); exposure to light; presence of metals, enzymes and peroxides; severity of the processing treatment (i.e., destruction of the cellular structure that protects the carotenoids, increase of surface area, duration and temperature of heat treatment); packaging material; and storage conditions. Completely losing their color and biological activities, the carotenoids give rise to volatile compounds that contribute to the aroma/flavor, desirable in tea and wine and undesirable in dehydrated carrot. (*See* **Retinol**: Properties and Determination.)

Reports on carotenoid retention during processing and storage of foods appear conflicting, some claiming no loss or increase in carotenoid content and others showing considerable reductions. Claimed increases are likely to be artifacts of the analytical process due to loss of carotenoids in fresh samples because of enzymatic activity, greater extractability of carotenoids from processed samples, unaccounted loss of water, and leaching of soluble solids. On the other hand, care must be taken so as not to attribute carotenoid losses during analysis to processing and storage effects.

Notwithstanding the diverging results, processing under good manufacturing practices should have a small effect on carotenoids; stability is good to excellent in frozen and heat-sterilized foods throughout the normal shelf-life. Oxygen content is minimized by hot packing, vacuum packing and oxygen scavenging with ascorbic acid. Stability in dehydrated and powdered fruits and vegetables is generally poor, unless the product has been carefully processed and stored in inert atmosphere in packaging impermeable to oxygen and light. A considerable portion of the carotenoids may be physically removed on peeling and juicing of fruits and by milling of seeds and grains. The time lag between peeling, cutting or puréeing and processing should be kept to a minimum so as not to allow enzymatic oxidation of carotenoids, which can be a more serious problem than thermal decomposition. Crude red palm oil contains a substantial amount of carotenoids, which are degraded on refining.

Analysis

Although more reliable data on food carotenoids are being acquired, erroneous data persist in the international literature. This reflects the inherent difficulty in carrying out this type of analysis. The principal confounding factors are: (1) the existence of a large number of naturally occurring carotenoids, (2) qualitative and quantitative variation in the carotenoid composition among foods, (3) a wide range of carotenoid concentrations in a given food, and (4) the instability of carotenoids.

A major problem in carotenoid analysis arises from their instability. Thus, whatever the method chosen, precautionary measures to avoid artifact formation and quantitative losses must be taken, such as completion of the analysis within the shortest possible time, exclusion of oxygen, protection from light, avoidance of high temperature and contact with acids, and use of high purity solvents, free from damaging impurities (e.g., peroxides in ether and tetrahydrofuran).

The general procedure for carotenoid analysis consists of sampling and sample preparation, extraction, partition, or transfer to a solvent compatible with the subsequent chromatographic step, saponification and washing (for some samples), chromatographic separation, identification, and quantification. Common errors are: samples not representive of the food lots under investigation; inefficient extraction; physical losses during the different steps; incomplete chromatographic separation; erroneous identification, quantification, and calculation; and isomerization and oxidation (enzymatic or nonenzymatic) during analysis.

Carotenoid analysis has been carried out to different extent, depending on the information sought. For a long time, only the major provitamin A carotenoids were determined. With the mounting evidence on the importance of carotenoids in reducing the risk of degenerative diseases and the recognition that this

role is not related to the provitamin A activity, determination of nonprovitamin A carotenoids has also been pursued. The complete carotenoid composition is the ultimate aim of carotenoid analysis. However, considering that the carotenoid composition of foods typically consists of one to four principal components, with a series of carotenoids in trace amounts, it is doubtful that the added information is worth the greater complexity, higher cost, and longer analysis time. The determination of major carotenoids appears adequate for the generation of data for food databases.

Because of the various factors that affect the carotenoid composition of foods, proper sampling and sample preparation to obtain representative and homogenous samples for analysis are of paramount importance. Errors in these initial steps can easily surpass those of the analysis *per se*. In addition, results should be accompanied by pertinent information, such as the variety, stage of maturity, part of the plant analyzed, season, and geographic origin.

The classical technique for separating carotenoids is open-column chromatography (OCC). MgO: Hyflosupercel and neutral alumina are the preferred adsorbents. Fractions are eluted successively with solvents of increasing polarity (e.g., increasing proportion of ethyl ether and acetone in hexane or petroleum ether), the separation being monitored visually. Reproducibility and efficiency of separation depend heavily on the analyst's skill and experience.

Reversed-phase high-performance liquid chromatography (HPLC) on a C_{18} column is currently the method of choice. Reasons for the popularity of the C_{18} column are: its weak hydrophobic interaction with the carotenoids (expected to be less destructive than the polar forces in normal-phase OCC), compatibility with most carotenoid solvents, and the polarity range of carotenoids, and wide commercial availability.

Most carotenoid separations have been carried out with 5-μm C_{18} spherical particles packed in a 250 × 4.6 mm column. Some laboratories are already using shorter and narrower (narrow bore) columns, smaller particles (3 μm) and C_{30} stationary phase (especially for separating geometric isomers).

Monomeric phases are simpler to use and are more reproducible. Polymeric C_{18} phases, however, have an excellent selectivity for structurally similar carotenoids (e.g., geometric isomers, lutein and zeaxanthin). However, the peaks tend to be broader, and columns from different production lots are more variable than the monomeric columns

The most important properties to be considered in selecting the mobile phase are polarity, viscosity, volatility, and toxicity. In addition, it must be inert with respect to the carotenoids. Many solvent systems have been suggested as mobile phases for carotenoids, but the primary solvents are acetonitrile and methanol, and most systems are actually slight modifications of some basic combinations. Acetonitrile has been widely used because of its lower viscosity and slightly better selectivity for xanthophyls when a monomeric C_{18} column is used. However, higher recoveries of carotenoids were reported with methanol-based solvents. Carotenoid recovery with acetonitrile-based solvents can be improved with the addition of ammonium acetate and triethylamine. Methanol is also more available, less expensive, and less toxic than acetonitrile. Small amounts of other solvents (e.g., tetrahydrofuran, ethyl acetate) are added to obtain the desired retention, increase solubility, and improve resolution.

Gradient elution should only be used when the analysis cannot be done isocratically. Isocratic separation is rapid, can be performed with simple equipment, and results in a stable baseline and more reproducible retention times. It is usually sufficient for the determination of provitamin A carotenoids or the principal carotenoids.

Gradient elution offers a greater resolving power, improved sensitivity, and elution of strongly retained compounds. It is more likely to resolve the whole range of carotenoids found in a given food. However, it has several disadvantages: an increased complexity, requirement for more sophisticated and expensive equipment, a need for column reequilibration between runs, a greater differential detector response, and often a poor reproducibility.

Because of the qualitative and quantitative variation of the carotenoid composition of foods, it is doubtful that a single chromatographic condition can be established for the different foods. At least some modification of the mobile phase is needed when changing from one food to another. Analysts should also guard against undue confidence that modern instruments inadvertently give. It is very easy to make errors in HPLC. Unsatisfactory and variable recoveries from the HPLC column, especially of lycopene, have been reported.

The injection solvent must be compatible with the HPLC mobile phase. A chromatogram with peak tailing and broad, doubled peaks is obtained when the carotenoids are much more soluble in the injection solvent than in the mobile phase. However, the sample will not dissolve completely if the injection solvent is too weak.

The retention times and the ultraviolet–visible absorption spectra (λ_{max} and fine structure) provide the first clues for the identification of carotenoids. The availability of the photodiode array detector allows

the spectra to be obtained on-line. However, the widespread use of these two parameters as the only criteria has led to misidentifications. Thus, it has been recommended that the following minimum criteria be fulfilled for identification: (1) the visible (or ultraviolet for shorter chromophores) spectrum in at least two different solvents must be in agreement with the chromophore suggested; (2) chromatographic properties must be identical in two systems, preferably TLC (R_F) and HPLC (t_R), and cochromatography with the authentic carotenoid should be demonstrated; and (3) a mass spectrum should be obtained, which allows at least confirmation of the molecular mass. The requirement of a mass spectrum, however, would limit carotenoid analysis to a very few laboratories around the world, precluding its execution in areas where it is very much needed. Moreover, major well-known carotenoids can be conclusively identified by the judicious and combined use of chromatographic data, absorption spectra and specific chemical reactions to confirm the type, location, and number of functional groups. (*See* **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography.)

Mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR) are indispensable in the elucidation of unknown or inconclusive structures of carotenoids and are increasingly used with food carotenoids. MS gives the molecular mass, and typical fragmentations provide information about the structure. NMR recognizes new end groups and determines the location of *cis* double bonds. The chemical shift data for many end groups and *trans*/*cis* isomers are available from tables for both ^{1}H-NMR and ^{13}C-NMR. (*See* **Spectroscopy**: Nuclear Magnetic Resonance.)

The quantification step in OCC is straightforward. The separated carotenoid fractions are collected and quantified spectrophotometrically through the use of tabulated absorption coefficients. HPLC quantification is carried out by means of internal or external calibration, for which the concentrations of the standard solutions are also determined spectrophotometrically. Unfortunately, few carotenoid standards are available commercially and are of a widely varying purity. Other carotenoids have to be isolated and purified from natural sources by OCC or by accumulating separated fractions from several HPLC runs. Both procedures are time-consuming and require experience and patience. The instability of the carotenoid standards is another serious problem.

Notwithstanding the inherent difficulties and the many possible sources of errors, reliable data on food carotenoids can be obtained in the hands of careful and well-informed analysts.

See also: **Analysis of Food**; **Antioxidants**: Natural Antioxidants; **Cancer**: Diet in Cancer Prevention; **Chromatography**: High-performance Liquid Chromatography; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Guavas**; **Mangoes**; **Mass Spectrometry**: Applications; **Oxidation of Food Components**; **Papayas**; **Peaches and Nectarines**; **Ripening of Fruit**; **Tomatoes**; **Vegetables of Temperate Climates**: Carrot, Parsnip, and Beetroot

Further Reading

Britton G, Liaaen-Jensen S and Pfander H (eds) (1995) *Carotenoids*, vol 1A. *Isolation and analysis*. Basel: Birkhauser.

Britton G, Liaaen-Jensen S and Pfander H (eds) (1995) *Carotenoids*, vol. 1B. *Spectroscopy*. Basel: Birkhauser.

Goodwin TW and Britton G (1988) Distribution and analysis of carotenoids. In: Goodwin TW (ed.) *Plant Pigments*, pp. 61–132. London: Academic Press.

Gross J (1987) *Pigments in Fruits*. London: Academic Press.

Gross J (1991) *Pigments in Vegetables. Chlorophylls and Carotenoids*. New York: Van Nostrand Reinhold.

Krinsky NI, Mathews-Roth MM and Taylor RF (eds) (1990) *Carotenoids: Chemistry and Biology*. New York: Plenum Press.

Packer L (ed.) (1992) Carotenoids part A. Chemistry, separation, quantitation, and antioxidation. *Methods in Enzymology* 213: 3–538.

Pfander H (ed.) (1987) *Key to Carotenoids*, 2nd edn. Basel: Birkhauser.

Rodriguez-Amaya DB (1993) Nature and distribution of carotenoids in foods. In: Charalambous G (ed.) *Shelf Life Studies of Foods and Beverages. Chemical, Biological, Physical and Nutritional Aspects*, pp. 547–587. Amsterdam: Elsevier Science.

Rodriguez-Amaya DB (1999a) *A Guide to Carotenoid Analysis in Foods*. Washington, DC: International Life Sciences Institute Press.

Rodriguez-Amaya DB (1999b) Changes in carotenoids during processing and storage of foods. *Archivos Latinoamericanos de Nutrición* 49: 36S–47S.

Physiology

H C Furr, Craft Technologies, Wilson, NC, USA
R M Clark, University of Connecticut, Storrs, CT, USA

Background

The carotenoids are polyisoprenoid compounds (usually tetraterpenes); most common carotenoids have 40 carbon atoms, although shorter (apocarotenoids) and longer forms exist. Frequently, carotenoids have

six-membered rings at each end of the polyprenoid chain, although lycopene (with no end rings) is a prominent exception (Figure 1). β-Carotene ($C_{40}H_{46}$, MW 536), with two β-ionone rings connected by a conjugated polyene chain, is the carotenoid prototype; formally, it can be considered to be composed of eight isoprene units joined head-to-tail, except in the center, so that the molecule is symmetrical about the center. The large number of conjugated carbon–carbon double bonds is responsible for carotenoids' high light-absorbing ability and intense red–yellow colors (wavelengths of maximum light absorption are typically 450–500 nm, molar extinction coefficients typically 132 000–185 000 $mol^{-1}l^{-1}$). Effective photoprotection requires at least nine conjugated double bonds, and absorption of visible light requires seven to 13 conjugated double bonds. Higher carotenoids (C_{45} and C_{50}), of bacterial origin, are products of addition of C_5 units to give a ring structure substituted at the 2-position.

The two major carotenoid subgroups are the carotenes (hydrocarbon carotenoids, including lycopene, α-carotene, and β-carotene) and the oxycarotenoids, also called xanthophyls (hydoxy-, oxy-, epoxy-, and furanoxy derivatives, such as lutein, canthaxanthin, and astaxanthin) (Figure 1). The prefix 'neo-' was previously used to designate carotenoids with at least one *cis* configuration in the polyisoprene chain, and the prefix 'pro-' designates some poly-*cis*-carotenoids. The prefix 'apo-' designates a chain-shortened structure (e.g., β-apo-8′-carotenal). Because the basic structures can be further metabolized by the introduction of hydroxyl, keto, or methoxyl groups, and by hydrogenation, cyclization, isomerization, and oxidative degradation, a great variety of structures exist; it has been estimated that more than 600 carotenoids have already been identified in nature. Good reviews on the structures and physical properties of carotenoids have been provided by Isler *et al*. and by Bauernfeind, and in a series edited by Britton, Liaaen-Jensen, and Pfander.

Carotenoids are synthesized by higher plants and by certain fungi, algae, and bacteria; they are not synthesized by animals but may be biochemically modified by them. Carotenoids are ubiquitous in the plant kingdom (no naturally occurring green plant lacks carotenoids), which suggests that their two functions are irreplaceable: as protectants against reactive singlet oxygen and as accessory photopigments. In photoprotection, chlorophyl and other cellular components are protected against irreversible destruction by preferential reaction of reactive singlet-state oxygen (electronically excited) with carotenoids. As accessory photopigments,

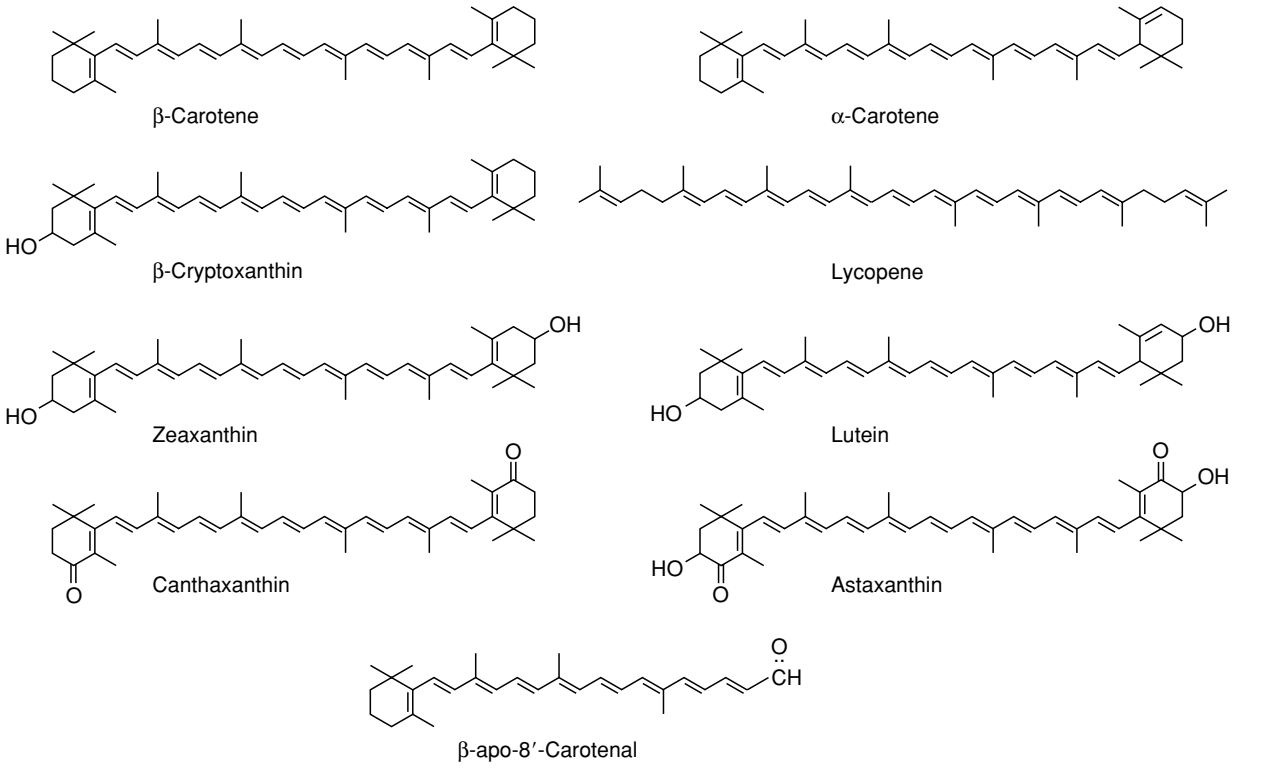

Figure 1 Structures of some representative carotenoids.

carotenoids increase the efficiency of photosynthesis by absorbing light in regions of the visible spectrum where chlorophyls do not absorb efficiently, then transferring the captured energy to other photopigments; carotenoids in photosystems are associated with proteins, and the carotenoid–protein interactions can shift the wavelength of maximum light absorption. Carotenoids are particularly effective in both processes, because their singlet electronic states are higher in energy, and their triplet states are lower in energy than those of chlorophyls, thus allowing both photoprotection and light-harvesting to be energetically favorable.

The total carotenoid production in the biosphere has been estimated at 10^8 tonnes per year; this is mostly due to fucosanthanin (characteristic of many marine algae) and the main carotenoids of green leaves, namely lutein, violaxanthin, and neoxanthin. β-Carotene, although quantitatively less important, occurs widely and is of particular interest as a major precursor of vitamin A for animal and human nutrition.

Absorption and Bioavailability of Carotenoids

There is considerable species variability in the efficiency of absorption and in metabolism of carotenoids. Humans are apparently unusual in absorbing, transporting, and storing both carotenes and xanthophyls well. Rats do not accumulate β-carotene, but instead metabolize almost all dietary β-carotene to vitamin A. In contrast, cattle show serum and milk accumulation of β-carotene but not of xanthophyls. Animals with yellow fat absorb and store carotenoids, whereas animals with white fat do not. Ferrets, gerbils, preruminant calves, and nonhuman primates have been used as animal models for human carotenoid absorption; rats and mice are also useful for studies of noncleavable carotenoids and for studies of pharmacological doses of carotenoids used for cancer prevention or treatment. Some fish (for example, salmonids) and birds absorb xanthophyls particularly well and use them as pigments.

Intestinal absorption of carotenoids in animals requires the presence of bile acids; human subjects with impaired bile flow (biliary atresia) show low levels of liver carotenoids compared with normal subjects. The presence of other dietary lipid promotes expansion of the mixed bile salt micelle and enhances carotenoid absorption. Animal and human feeding studies have shown that the biological matrix of food affects absorption of dietary carotenoids, and dietary carotenoids in plant tissues have a lower bioavailability than those in oil solutions. Hence, the plant source and method of preparation of foods affect bioavailability of β-carotene and other carotenoids: processes that break down the biological matrix (e.g. grinding or chopping, mild cooking) improve carotenoid absorption. Castenmiller and West (1998) have characterized the factors affecting carotenoid absorption as Species of carotene; Molecular linkage (e.g., esterification of xanthophyls); Amount of carotene consumed in a meal; Matrix in which the carotenoid is incorporated; Absorption modifiers (including lipid content of the meal); Nutrient status of host (in particular, vitamin A status); Genetic factors; Host-related factors; and Interactions (e.g., competition among carotenoids for absorption).

Unequivocal estimation of the efficiency of absorption of dietary carotenoids is very difficult because of the ubiquitous presence of carotenoids in the diet (and thus in plasma and other tissues), slow absorption kinetics, partial metabolism of carotenoids, and considerable individual variation in the extent of absorption and metabolism.

Transport, Distribution, and Storage of Carotenoids

As with other lipids, freshly absorbed carotenoids are transported in the lymph via chylomicra. It is believed that the carotenoids, as with other lipid components of the chylomicron remnants, are taken up by the liver and then released into the blood stream as components of very-low-density lipoproteins (VLDL). The VLDL are taken up by extrahepatic tissues, their components processed into low-density lipoproteins (LDL), and eventually into high-density lipoproteins (HDL). Early studies showed that in human plasma, the hydrocarbon carotenoids β-carotene and lycopene are predominantly associated with LDL (75%) and HDL (25%); the xanthophyl lutein is approximately equally distributed between HDL and LDL. In cattle, where the predominant plasma lipoprotein is HDL, more than 80% of total plasma β-carotene is transported in this fraction. The predominant human serum carotenoids in the USA are β-carotene, α-carotene, lycopene, β-cryptoxanthin, lutein, and zeaxanthin, and also the colorless carotenoids phytoene and phytofluene; serum carotenoid profiles and concentrations are highly dependent on dietary intake and so depend on food habits. It is estimated that serum carotenoids represent about 1% of total body carotenoids in the human.

Adipose tissue (80–85% of total body carotenoids) and liver (8–12%) are the major sites of carotenoid deposition in humans when the total mass of tissue carotenoid is considered. Typical total human carotenoid levels of 100–150 mg are reported from

autopsy analyses. Human adipose carotenoid profiles are similar to those of serum. Reported values for total human liver carotenoids range between 0 and 97 μg per gram of liver (i.e., 0–0.18 μmol g^{-1}; total carotenoids were expressed as β-carotene in older studies), and include lutein, lycopene, α-carotene, and β-carotene. Other tissues that are known to contain high concentrations of carotenoids include adrenals (20 μg g^{-1} (37 nmol g^{-1}) tissue, mostly as β-carotene) and human macular pigment (containing predominantly the xanthophyls lutein and zeaxanthin), as well as kidney, ovaries, pituitary, placenta, testes, and thymus in cattle. Chick retinal oil droplets accumulate astaxanthin, which can be formed from dietary zeaxanthin or canthaxanthin. It has been suggested that organs with high numbers of LDL receptors and high rates of LDL uptake show higher tissue levels of carotenoids.

The specific accumulation of certain carotenoids in particular tissues may furnish clues to distinctive physiological roles of carotenoids. The macula of the primate eye accumulates the xanthophyls lutein and zeaxanthin, with lutein being found predominantly in the center of the macula; loss of these xanthophyls is associated with macular degeneration, a disease characteristic of aging. Bovine corpus luteum amasses xanthophyls (as much as 60 μg of total carotenoids per gram, 105 nmol per gram of tissue), and it has been suggested that carotenoids play some important role in bovine reproduction; however, porcine corpus luteum does not accumulate carotenoids. Bovine pineal gland has been found to store β-carotene (as much as 1 μg per gram, 2 nmol per gram of wet tissue), with perhaps small amounts of xanthophyls also present; the absence of other carotenoids was noteworthy. Levels of approximately 1 μg of total carotenoid per gram of bovine pituitary gland have been reported. The physiological importance of these accumulations of carotenoids is, however, not clear.

Carotenoid–protein interactions have been studied in plants, bacteria, and marine invertebrates. The strength of specific protein–carotenoid binding depends, not surprisingly, on carotenoid structure. Distinctive interactions between a binding protein and its carotenoid ligand are evident by the spectral changes of the carotenoid on binding; these spectral changes may reflect physiological functions as well as producing color polymorphism among species. A notable example of these carotenoid–protein interactions is the binding of the xanthophyl astaxanthin (red-colored in solution) to the protein α-crustacyanin in the lobster, giving a pigment that is typically blue–green; on boiling the animal, the carotenoid is released from its binding protein to yield a bright red color.

Retinol Equivalents of Different Carotenoids

Of the more than 600 known carotenoids, only about 60 have been reported to be precursors of vitamin A. In order to serve as a precursor of vitamin A, a carotenoid must have at least one unsubstituted 'β-ionone' ring (2,6,6-trimethyl-1-cyclohexen-1-yl) with a polyene side chain of at least 11 carbon atoms. Thus, α-carotene (with one β-ionone ring) has half the biological activity of β-carotene, and canthaxanthin (with keto substitutions on both rings; **Figure 1**) has no provitamin A activity. Absolute vitamin A activity, however, depends on a number of other factors, not all of which are well understood. In general, it seems that the efficiency of the conversion of carotenoids to vitamin A depends on the vitamin A status: high intakes of preformed vitamin A result in poor efficiency of conversion to vitamin A, perhaps due to metabolic control of the cleavage enzyme(s). Frank vitamin A deficiency is also associated with impaired conversion efficiency, perhaps because of damaged intestinal epithelial function in vitamin A inadequacy. Because conversion of β-carotene to vitamin A is impaired in vitamin A deficiency, vitamin A deficiency disease is more rapidly cured by providing preformed vitamin A than by providing dietary carotenoids. Excessive vitamin E intake may impair carotenoid cleavage (or interferes with intestinal absorption); vitamin E deficiency decreases vitamin A formation, perhaps because adequate vitamin E is needed to protect carotenoids and vitamin A from oxidation. Conversion efficiency of β-carotene and other carotenoids to vitamin A is also decreased at high carotenoid intakes, perhaps due to impaired intestinal absorption or to metabolic control of the cleavage enzyme(s). Protein deficiency impairs carotenoid cleavage, suggesting that protein malnutrition in humans exacerbates vitamin A deficiency.

Common dietary sources of provitamin A carotenoids are carrots, yellow squash, dark-green leafy vegetables, yellow maize, tomatoes, papaya, and oranges. Cereal grains and white maize contain little or no provitamin A. Red palm oil is one of the richest sources of provitamin A carotenoids, containing approximately 0.5 mg of α-carotene + β-carotene per milliliter; as little as 7 ml of palm oil per day may provide adequate vitamin A for the preschool child. On a global basis, β-carotene is the most important vitamin A precursor, both because of its greater provitamin A activity and because of its wide distribution in plant products. The recent incorporation of the enzymatic pathway of carotenoid biosynthesis into rice ('Golden Rice') promises to help alleviate vitamin A deficiency in parts of the world where rice is

Table 1 Vitamin A activity of common carotenoids relative to β-carotene[a]

Carotenoid	*Activity (%)*	*Occurrence (partial list)*
β-Carotene	100	Green plants, vegetables, carrots, yellow sweet potatoes, squash, tomatoes, red and yellow fruits
α-Carotene	50–54	Green plants, carrots, squash, maize, green peppers
γ-Carotene	42–50	Carrots, sweet potatoes, maize, tomatoes, algae, some fruits
β-Carotene 5′,6′-monoepoxide	21	Plants, potatoes, red peppers
β-Carotene 5,6:5′,6′-diepoxide	Active?	Plants
β-Carotene 5′,8′-monofuranoxide (mutatochrome, citroxanthin, flavacin)	50	Orange peel, red peppers, tomatoes, sweet potatoes, cranberries
4-Keto-β-carotene (4-oxo-β-carotene, echinenone, aphanin, myxoxanthin)	44–54	Algae, sea urchins, *Daphnia*, *Hydra*, red sponges, brine shrimp, crustaceans
3-Keto-β-carotene (3-oxo-β-carotene)	52	
3-Hydroxy-β-carotene (β-cryptoxanthin)	50–60	Yellow maize, green peppers, lichens, persimmons, papayas, lemons, oranges
4-Hydroxy-β-carotene (isocryptoxanthin)	48	Brine shrimp
3,4-Dehydro-3′-hydroxy-β-carotene (anhydrolutein, deoxylutein)	Active, *c.* 10	Alfalfa meal, acidulated soybean soapstock
3-Hydroxy-4-keto-β-carotene (hydroxyechinenone)	Probably active	Algae, bacteria, flowers
β-Apo-2′-carotenal	Active	Citrus fruit
β-Apo-8′-carotenal	36–72	Citrus fruit, green plants
β-Apo-10′-carotenal	Active	Citrus fruit, green plants, alfalfa meal
β-Apo-12′-carotenal	120	Alfalfa meal
β-Apo-8′-carotenoic acid	Active	Maize, animal tissue
β-Apo-8′-carotenoic acid ethyl ester	25–78	
Lycopene	Inactive	Tomatoes, carrots, green peppers, pink citrus fruit
3,3′-Dihydroxy-β-carotene (zeaxanthin)	Inactive	Spinach, paprika, yellow maize, green peppers, fruits
3,3′-Dihydroxy-α-carotene (lutein)	Inactive	Green leaves, yellow maize, potatoes, spinach, green peppers, carrots, tomatoes, fruits
3,3′-Dihydroxy-4,4′-diketo-β-carotene (astaxanthin)	Inactive	Oranges, crustaceans, lobster, fish, algae, *Daphnia*, trout, salmon
4,4′-Diketo-β-carotene (canthaxanthin, aphanicin)	Inactive	Mushrooms, trout, *Daphnia*, *Hydra*, microorganisms, algae, crustaceans, brine shrimp
Capsanthin	Inactive	Red peppers, paprika
Capsorubrin	Inactive	Red peppers, paprika
Bixin	Inactive	Annatto seeds

[a]Adapted and shortened from Bauernfeind JC, Adams CR and Marusich WL (1981) Carotenes and vitamin A precursors in animal feed. In: Bauernfeind JC (ed.) *Carotenoids as Colorants and Vitamin A Precursors*, pp. 563–743. New York: Academic Press.

eaten extensively. Other provitamin A carotenoids, such as α-carotene and β-cryptoxanthin and β-apo-carotenals, can be nutritionally important sources of vitamin A from particular foods (Table 1). Other carotenoids, such as lycopene (which has no β-ionone rings) and the dihydroxycarotenoids lutein and zeaxanthin (which are substituted on both ionone rings), may be major carotenoids in particular foods and may have other important physiological functions but have no provitamin A activity.

Estimation of provitamin A content of foods from food composition tables is problematic: carotenoid composition of raw fruits and vegetables varies with species, growth conditions, and mode of storage. Furthermore, older tables often express total carotenoids as 'β-carotene,' ignoring the different provitamin A activity of different carotenoids. Biological activity of carotenoids in foodstuffs may differ from the analyzed content because of the mode of preparation, binding of carotenoids within the foodstuff, and the nature of the meal (presence or absence of fat), as discussed above. Conversion factors, recently revised to account for apparent bioavailability of carotenoids from foods, are given in Table 2.

Enzymatic Conversion to Vitamin A

By comparison of the structure of β-carotene with that of vitamin A, it would appear that β-carotene could be cleaved in its center (between the 15 and 15′ carbon atoms) to give two molecules of vitamin A per molecule of carotenoid. Such a mechanism was first proposed by Karrer. However, in addition to this central (symmetric) cleavage, a random (asymmetric) cleavage has been proposed. Glover first suggested that asymmetric cleavage of the symmetric β-carotene molecule could be followed by stepwise shortening to vitamin A. In support of this hypothesis, small

Table 2 Conversion factors for estimating vitamin A value from carotenoid composition[a]

Factors
1 retinol activity equivalent (RAE)
= 1 μg all-*trans* retinol
= 12 μg all-*trans* β-carotene
= 24 μg other provitamin A carotenoids
= 3.33 IU vitamin A
(Note that 1 RAE = 1.15 μg of retinyl acetate (MW 328) = 1.83 μg of retinyl palmitate (MW 524), i.e., molar equivalence)
1 International Unit of vitamin A (IU_A)
= 0.3 μg all-*trans* retinol
= 0.3 RAE
= 3.6 μg all-*trans* β-carotene
= 7.2 μg other provitamin A carotenoids
Formulas
(1) Retinol and β-carotene given in μg, RAE = μg retinol + (μg β-carotene/12)
(2) β-Carotene and other provitamin A carotenoids given in μg, RAE = (μg β-carotene/12) + (μg other provitamin A carotenoids/24)

[a]Adapted from Olson JA (1987) Recommended dietary intakes (RDI) of vitamin A in humans: Appendix. *American Journal of Clinical Nutrition* 45: 704–716 and from Panel on Micronutrients, Subcommittees on Upper Reference Levels of Nutrients and of Interpretation and Use of Dietary Reference Intakes, and the Standing Committee on the Scientific Evaluation of Dietary Reference Intakes (2001) *Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc*. Washington, DC: National Academy Press. Available online at http://books.nap.edu/.

amounts of labeled β-apo-8′-, β-apo-10′-, and β-apo-12′-carotenal (and their alcohols and acids) have been found in rat tissues after feeding radioactive β-carotene. These apo-carotenals can be formed by chemical reaction of oxidizing agents (hydrogen peroxide, potassium permanganate) with β-carotene *in vitro*, and they are found in small amounts in plants and in animal-feeding experiments. Enzymatic asymmetric cleavage activity has been demonstrated *in vitro*, but its relative importance has not been established. However, *in vitro* experiments (initially carried out by Olson and Hayaishi and by Goodman and Huang) using partially purified intestinal homogenates (from rat, rabbit, or guinea-pig) have demonstrated symmetric cleavage of β-carotene, with retinal (vitamin A aldehyde) as the only detectable product. This enzymatic activity requires the presence of O_2, and has been named '15,15′-β-carotenoid dioxygenase' (EC 1.13.11.21). The enzyme has recently been cloned from mammalian sources, and its further characterization promises to explain much about the factors affecting provitamin A activity of β-carotene and related carotenoids. This enzymatic activity has been demonstrated in intestine, liver, kidney, and several other tissues *in vitro*. β-Carotene is not a source of vitamin A in the cat, and it has been reported that 15,15′-dioxygenase activity is not present in cat intestine, consistent with the 'central cleavage' hypothesis. The possiblity remains that excentric cleavage is a source of retinoic acid, which is an important modulator of gene expression.

Some fish and birds can also convert the xanthophyls astaxanthin, canthaxanthin, and isozeaxanthin to vitamin A; some freshwater fish can convert lutein to 3,4-didehydroretinol (vitamin A_2). Keto carotenoids can be reduced to the corresponding alcohols and esterified with long-chain fatty acids.

Toxicity of Carotenoids

There is virtually no toxicity associated with high intakes of carotenoids, in marked contrast to vitamin A. Excessive human ingestion of β-carotene can cause high serum carotenoid levels and yellowing of light-colored skin, but such over-intake is very unusual, and no harmful physiological effects from even such high intakes of β-carotene have been reported. The effects disappear when high carotenoid intakes cease. Ingestion of large amounts of carotenoids does not give rise to hypervitaminosis A, probably because of the decreased efficiency of conversion of carotenoids to vitamin A at high intakes. Teratogenicity due to carotenoids has never been suggested. β-Carotene supplementation in humans does not produce elevated serum triacylglycerols or cholesterol. However, prolonged intake of high doses (20 mg per day) of β-carotene has been associated in epidemiologic studies with increased risk of lung cancer in cigarette smokers, perhaps because of the prooxidant effects of β-carotene at relatively higher oxygen concentrations in the lungs. Although ingestion of 3–10 mg of carotenoid per day from foods and dietary supplements appears to be completely safe and advisable, there are concerns that higher doses of β-carotene (20–30 mg per day) may not be safe in smokers.

The only carotenoid that has been clearly associated with adverse effects is canthaxanthin (4,4′-diketo-β-carotene), a carotenoid that has been used in treatment of patients with erythropoietic protoporphyria

and related skin disorders. High daily doses (50–100 mg) have resulted in canthaxanthin retinopathy (crystalline deposits in the retina leading to impaired night vision); the effect is slowly reversible on cessation of intake. Pharmacological doses of canthaxanthin have depressed liver accumulation of dietary vitamin A in rats. Human ingestion of large amounts of canthaxanthin gives a yellowish coloration to light-colored skin; this is the basis of some 'skin tanning' pills.

Therapeutic Role of Carotenoids

In addition to the important role of some carotenoids as vitamin A precursors, there is evidence that carotenoids can quench singlet oxygen and can serve as radical-trapping antioxidants; it has also been shown that carotenoids may have a role (apart from being precursors of retinoids) in enhancing the immune function in animals. Singlet-oxygen trapping and antioxidant capabilities are shown by carotenoids that have no provitamin A activity (such as lycopene and astaxanthin) as well as by provitamin A carotenoids.

Both β-carotene and canthaxanthin have been used in large dose regimens to ameliorate photosensitivity associated with erythropoietic protoporphyria (a genetic disease of porphyrin metabolism). β-Carotene also prevents photosensitivity resulting from quinidine ingestion, but not photosensitivity to ultraviolet light, although it has been reported that human serum carotenoid levels decrease after repeated exposure to ultraviolet light.

Epidemiological evidence strongly suggests a role for carotenoids in human cancer prevention, although such studies may be confounded by other factors. The strongest evidence is for a protective effect of carotenoids against lung cancer; tomato products (which provide lycopene) may be more effective than carrots and squash (sources of β-carotene). In particular, consumption of tomato products (high in lycopene) has been associated epidemiologically with a reduced risk of prostate neoplasms. Protective effects of carotenoids against laryngeal cancer, gastric cancer, invasive bladder cancer, and cervical dysplasia and cervical cancer have been suggested also; these effects are not shown by vitamin A, suggesting a direct role of carotenoids. Neither carotenoids nor vitamin A protect against cancers of the esophagous or gastrointestinal tract, breast, head, or neck. A variety of cancers in animal models are prevented or delayed by carotenoids, including skin tumors and mammary tumors. In studies in cells in culture, carotenoids can prevent malignant transformation, sister chromatid exchange, and mutagenic effects in bacterial systems.

A high consumption of carotenoid-containing foods is associated with a reduced risk of coronary heart disease, although there is a lack of consistency in identifying the responsible compounds.

Carotenoids have been found to enhance both specific and nonspecific immune functions, including proliferation of T- and B-lymphocytes, induction of specific effector cells capable of killing tumor cells, and increased production of tumor necrosis factor, and secretion of factors required for communication between cells (prostaglandins and leukotrienes). It has been suggested that the mechanism(s) may involve quenching singlet oxygen and trapping free radicals (generated by neutrophils to kill invading cells), or maintaining cell-membrane fluidity (important for the function of membrane receptors and for the release of immunomodulating factors). Again, some of the immunologic functions of carotenoids seem to be independent of any provitamin A activity, although retinoids are also important in immune function.

The xanthophyls lutein and zeaxanthin are accumulated specifically in the macula of the primate eye, and high dietary intakes of these two carotenoids (and elevated plasma concentrations) are associated with a reduced risk of age-related macular degeneration.

Elucidation of these biochemical effects of carotenoids and further explanation of the metabolism of conversion of carotenoids to vitamin A remain central issues in the study of carotenoid function.

See also: **Carotenoids**: Occurrence, Properties, and Determination; **Protein**: Deficiency; **Retinol**: Properties and Determination; **Tocopherols**: Properties and Determination

Further Reading

Bauernfeind JC (ed.) (1981) *Carotenoids as Colorants and Vitamin A Precursors*. New York: Academic Press.

Beyer P, Al-Babili S, Ye X *et al.* (2002) Golden Rice: Introducing the beta-carotene biosynthesis pathway into rice endosperm by genetic engineering to defeat vitamin A deficiency. *Journal of Nutrition* 132: 506S–510S.

Britton G, Liaaen-Jensen S and Pfander H (eds) (1995a) *Carotenoids. 1A. Isolation and Analysis*. Basel: Birkhauser.

Britton G, Liaaen-Jensen S and Pfander H (eds) (1995b) *Carotenoids. 1B: Spectroscopy*. Basel: Birkhauser.

Castenmiller JJ and West CE (1998) Bioavailability and bioconversion of carotenoids. *Annual Review of Nutrition* 18: 19–38.

Isler O, Gutmann H and Solms U (eds) (1971) *Carotenoids*. Basel: Birkhauser Verlag.

Lee CM, Boileau AC, Boileau TW *et al.* (1999) Review of animal models in carotenoid research. *Journal of Nutrition* 129: 2271–2277.

Straub O (1987) In: Pfander H (ed.) *Key to Carotenoids*. Basel: Birkhauser.

Subcommittees on Upper Reference Levels of Nutrients and Interpretation and Uses of Dietary Reference Intakes, and the Standing Committee on Scientific Evaluation of Dietary Reference Intakes (2000) *Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium and Carotenoids: A Report of the Panel on Dietary Antioxidants and Related Compounds*. National Academy Press, Washington, D.C.

Von Lintig J and Wyss A (2001) Molecular analysis of vitamin A formation: cloning and characterization of beta-carotene 15,15′-dioxygenase. *Archives of Biochemistry and Biophysics* 385: 47–52.

CASEIN AND CASEINATES

Contents
Methods of Manufacture
Uses in the Food Industry

Methods of Manufacture

C R Southward, Fonterra Research Centre, Palmerston North, New Zealand

This article is reproduced from *Encyclopedia of Dairy Sciences*,

Introduction

The commercial production of casein, the principal protein in cows' milk, has occurred for most of the 20th century. During the period to 1960, the major proportion of casein was used in technical (or non-food) applications. More recently, however, there has been a significant change from technical to edible uses for casein products and this has been reflected in the introduction of requirements for pasteurization of milk intended for casein manufacture (*c*. 1970) and the greater number of specifications for microbial quality and freedom from impurities. At the present time, the major producers of casein include the European Community (in particular, Denmark, France, Germany, the Irish Republic, and The Netherlands) and New Zealand, with Poland, Australia, and India producing smaller quantities. World production of casein products is currently estimated at 220 000–250 000 tonnes. This article describes the manufacture of the main casein products.

General

The proteins that exist in milk can be broadly divided into two groups – casein and whey proteins. Caseins may be considered as those proteins that are precipitated when unheated (raw) milk is acidified to pH 4.6 (the isoelectric point of casein) whereas whey proteins remain in solution. Commercial casein is a mixture of four different caseins – α_{s1}-, α_{s2}-, β- and κ-casein, and may sometimes be referred to as whole casein. As a phosphoprotein, casein belongs to a relatively rare class of proteins. It contains 0.7–0.9% phosphorus, covalently bound to the casein by a serine ester linkage. Casein exists in milk in combination with calcium, inorganic phosphate, and citrate as a colloidal suspension of complex micelles and accounts for 2.6–2.9% by weight of whole milk.

Manufacture

Casein may be precipitated from skimmed milk to produce several products such as acid casein, rennet casein, or coprecipitate. All these products are insoluble in water after precipitation. However, addition of alkali to acid casein yields water-soluble caseinate.

Acid Casein

Acidification

When milk (pH 6.6) is acidified, the calcium and inorganic phosphate are removed from the casein micelles, the net charge on the micelles decreases, and the micelles become less and less stable until the casein precipitates. Complete precipitation of the casein occurs at the isoelectric point, pH 4.6. Acidification of the milk may be carried out by one of the following processes:

1. Inoculation of milk with lactic acid-producing bacteria such as *Lactococcus lactis* subsp. *lactis*

or *cremoris*. These bacteria, commonly known as 'starters,' convert some of the lactose in the milk to lactic acid during the period of incubation (about 16–18 h). This is the most commonly employed method of manufacture in New Zealand (Figure 1).

2. Direct addition of dilute acid to skimmed milk. Hydrochloric acid, sulfuric acid, phosphoric acid, or lactic acid (or, occasionally, other organic acids) may be used for this purpose. The most common method of producing acid casein in countries other than New Zealand is by means of hydrochloric acid.
3. Indirect acidification of skim milk. A number of alternative processes have been patented in which skimmed milk is acidified by one or more of the following:
   - contact by means of ion exchange resins in the acid (cation) form
   - addition of acidified whey
   - electrodialysis

It is believed that these processes have relatively minor commercial significance.

Cooking/Acidulation

In the manufacture of acid casein, acidification of the milk is followed, or occasionally preceded, by heating of the mixture. Heating promotes agglomeration of the casein curd particles which subsequently shrink (syneresis) to expel whey. At the same time, the curds become firmer and are able to withstand the mechanical processing that follows. Heating of the acidified milk is sometimes termed 'cooking' (as used in cheese manufacture) and may be carried out, usually in the temperature range 45–55 °C, by:

1. injection of steam into the pipeline carrying the acidified milk
2. indirect heating by means of a heat exchanger
3. a combination of both – preheating through a heat exchanger with steam injection to complete the heating process (Figure 1)

In each case, the cooked curds and whey are held in a 'cooking pipe' for a period of about 10–20 s before they emerge into an acidulation vat. The curds and whey may remain here for a period varying from 30 s to about 15 min, during which time the curds are agitated gently in the whey until equilibrium between the calcium in the curds and that in the whey is attained.

Alternative processes may use a syneresis tube; the cooked curds and whey are held in a large-diameter tube for a period of several minutes (also undergoing acidulation or equilibration).

Dewheying and Washing

Following acidulation, the curds and whey are discharged by gravity to a dewheying screen above the first washing vat. The whey is removed and the curds fall into the vat. Alternatively, the curds and whey may be separated more completely using a horizontal solid-bowl centrifuge (decanter) or a casein-dewatering press before the curds are transferred to the first wash. A combination of screening and decanter dewheying can be used to reduce the hydraulic load on the decanter. The purpose of washing is to remove whey (containing mainly lactose) from the curds so that the casein produced is relatively pure. The temperature of the wash water may be varied, depending upon particular requirements. Casein is usually subjected to multiple washes and these are operated in a counterflow to the direction of the curds, with the purest curd meeting the cleanest water.

Dewatering

After washing, the curds are mechanically dewatered to remove excess water before drying. As drying is a relatively expensive operation, it is worthwhile removing as much water as possible from the curd before transferring the casein to the drier. The texture of the curd is affected by temperature; as the temperature of the wash water is increased, the curd releases more water during dewatering but becomes firm and more plastic and is consequently harder to break up and dry. Therefore, it is necessary to regulate carefully the temperature of the last wash to optimize the conflicting requirements of minimum water content and maximum friability of the curd.

Equipment for dewatering casein curd consists of roller or belt presses, decanters, and screen-bowl centrifuges. The roller press, used for many years, is designed to reduce the moisture content of curd to about 55%. The belt press will do a similar duty. Solid-bowl decanters or screen-bowl centrifuges are capable of reducing the moisture content of acid casein curd to about 52%.

Drying

Drying of casein curd is most commonly carried out using horizontal vibrating fluid-bed driers. These driers have two or more perforated stainless-steel decks (Figure 1). The combined effect of vibration of the decks and the flow of hot air (typical temperature range 75–115 °C) up through the holes in the decks causes the casein curd to become fluidized and materially helps in the removal of moisture from the particles. Most of the water is removed during the early stages of drying of the casein as it is evaporated from the surface of the particles. The later stages

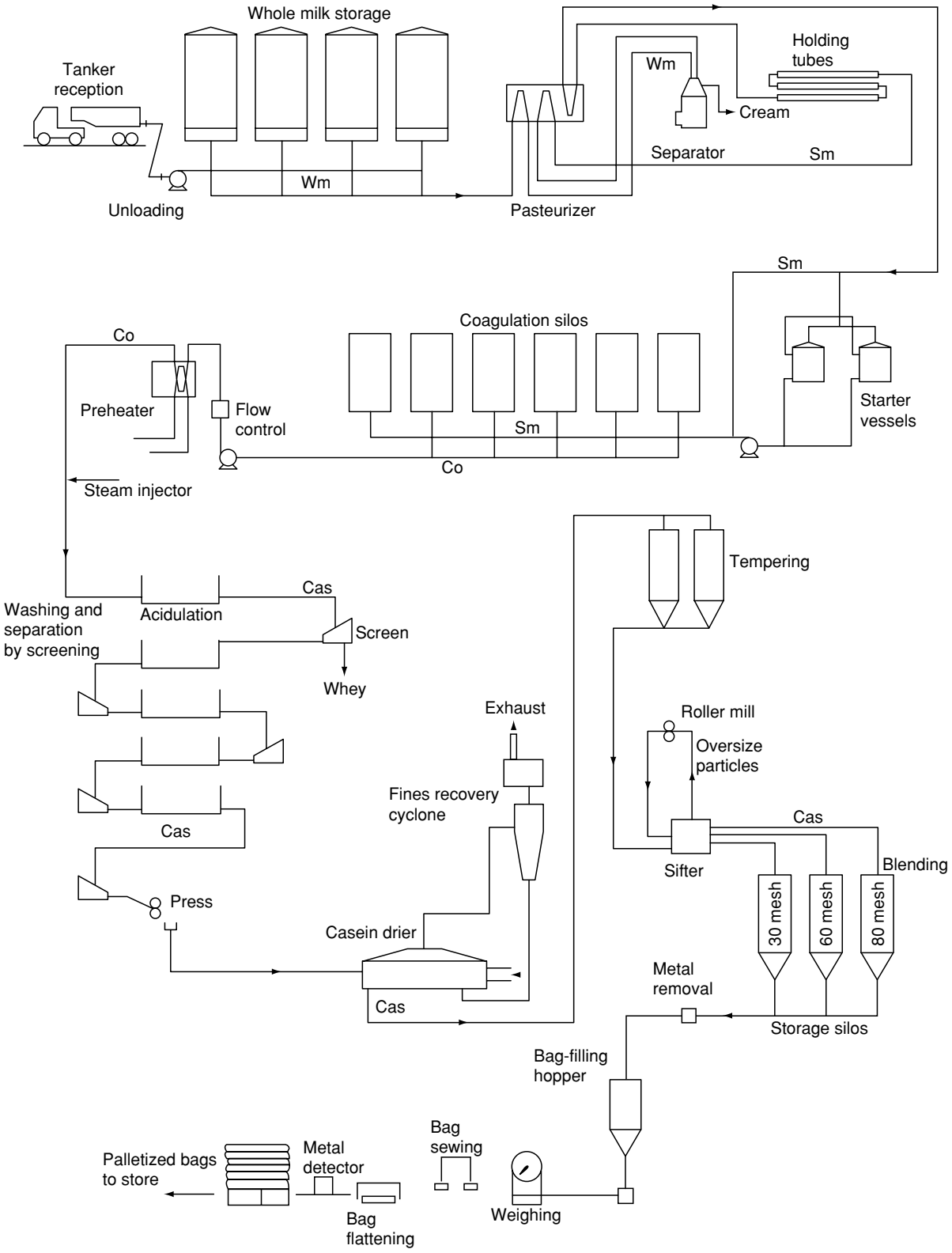

Figure 1 Outline of the manufacturing steps involved in producing lactic acid casein from skim milk. Wm, whole milk; Sm, skimmed milk; Co, coagulum; Cas, casein.

of drying require the transfer of moisture from the center to the surface of the particle, and this is a much slower process.

Pneumatic-conveying ring driers and attrition driers are also used for drying casein. These employ inline milling and tend to produce finer casein products than those dried in fluid-bed driers.

Cooling, Tempering, Milling, Sifting, Blending, Packing

Casein that is produced using fluid-bed driers is warm and soft and unsuitable for grinding immediately in some mills such as roller mills. Consequently, the casein may be cooled and then transferred to tempering bins where equilibration of moisture can occur in and between all the particles during a period of 8–24 h.

The casein may then be ground and sieved, using multideck, gyrating screens, into various particle sizes, usually < 600 μm. Very fine casein (e.g. < 150 μm) is generally produced using pin mills.

Following blending of the ground casein, it is packed into multiwall paper bags equipped with plastic liners and stored. The typical composition of acid casein is shown in Table 1.

Rennet Casein

Action of Chymosin

Casein may also be precipitated from milk by the action of a proteolytic enzyme, such as chymosin, which is present in calf rennet. Chymosin (and other enzymes from animal or microbial sources that cause milk to clot) splits off a portion of the κ-casein (referred to as glycomacropeptide or GMP) from the micelles. As a result, the micelles are no longer stable in the presence of calcium ions and form a three-dimensional gel. This process is essential in the production of most types of cheese.

Clotting of Skimmed Milk

Pasteurized skimmed milk at a temperature of 29 °C (or lower) is mixed with calf rennet (or other milk-clotting enzyme) in the approximate ratio (by volume) of 1:7500 rennet to the milk. If a lower setting temperature is used, renneting time must be correspondingly increased. It is also possible to reduce the quantity of rennet added under these conditions and consequently allow a longer time for rennet action to occur.

Cooking

The usual technique for cooking rennet casein involves the injection of steam into a cooking line of clotted milk pumped from a vat. However, the vat cooking technique (similar to that used in cheese manufacture) is also practiced in some countries. The cooking temperature used in making rennet casein usually varies from about 50 to 60 °C.

Dewheying, washing, dewatering, and drying of the curd then proceed in a manner similar to that used for acid casein, and the dried casein is also treated as outlined previously.

Where indirect cooking of rennet casein is used, a tubular heat exchanger may be used to cook the curds and whey to a temperature similar to that used in the direct cook (steam injection) process. Other processing steps are similar to those described above but no acidulation step is required after cooking of the rennet casein.

The typical composition of rennet casein is shown in Table 1.

Table 1 Typical composition and properties of casein products

Component	*Acid casein*	*Rennet casein*	*Sodium caseinate (spray dried)*
Composition	Amount per 100 g		
Moisture (g)	11.5	11.5	4.5
Fat (g)	1.1	0.4	1.0
Protein (g) (nitrogen × 6.38)	86.2	81.3	92.0
Ash (g)	1.8	8.2	3.6
Lactose (g)	0.1	0.1	0.1
	Amount per 1 kg		
Copper (mg)	2	2	2
Lead (mg)	< 1	< 1	< 1
Iron (mg)	5–20	5	5–10
Physical properties			
Color	Creamy white	Creamy white	White
Flavor	Bland, clean	Bland, clean	Bland, clean
Solubility in water	Insoluble	Insoluble	Soluble

Coprecipitates

Coprecipitates are combinations of casein and whey proteins that are coprecipitated from heated (skimmed) milk. When milk is heated to a temperature greater than 70 °C, some of the whey proteins are heat-denatured and may interact with some of the caseins. When a casein precipitant (such as acid or calcium chloride) is added to the heated milk, the casein and whey proteins coprecipitate together. Depending on the pH of precipitation (which may vary from 6.6 to 4.5), the coprecipitate will contain different quantities of calcium (high-calcium at pH 6.6 to low-calcium at pH 4.5). As whey proteins have a higher nutritional value than casein (as measured by several different biological and chemical techniques), coprecipitates also have an enhanced nutritional value compared with casein. The yield of coprecipitate from skimmed milk is usually 5–20% greater than that of acid casein.

Caseinates

Caseinates are produced by the neutralization of acid casein with alkali. All caseinates are substantially water-soluble and are typically prepared as a solution of about 20% solids prior to spray drying. Roller-dried caseinates may be prepared from more concentrated solutions. It is also possible to prepare granular, partly soluble, or semidispersible forms of caseinate in which the casein and alkali have only partly reacted. Sodium caseinate is the most common form of this class of product and is prepared by mixing a solution of sodium hydroxide, bicarbonate, or carbonate with acid casein curd or dry acid casein that has been suspended in water and then drying the resultant solution (Figure 2). The dried powder dissolves completely in water to produce a viscous, sticky, straw-colored solution.

Calcium caseinate, on the other hand, produces a thin, opaque, white colloidal dispersion in water, similar in appearance to milk. Other caseinates, such as those of potassium and ammonium, are similar in general properties to sodium caseinate. Magnesium caseinate has properties that are intermediate between those of sodium and calcium caseinates. However, relatively insignificant commercial quantities of these products are manufactured at the present time.

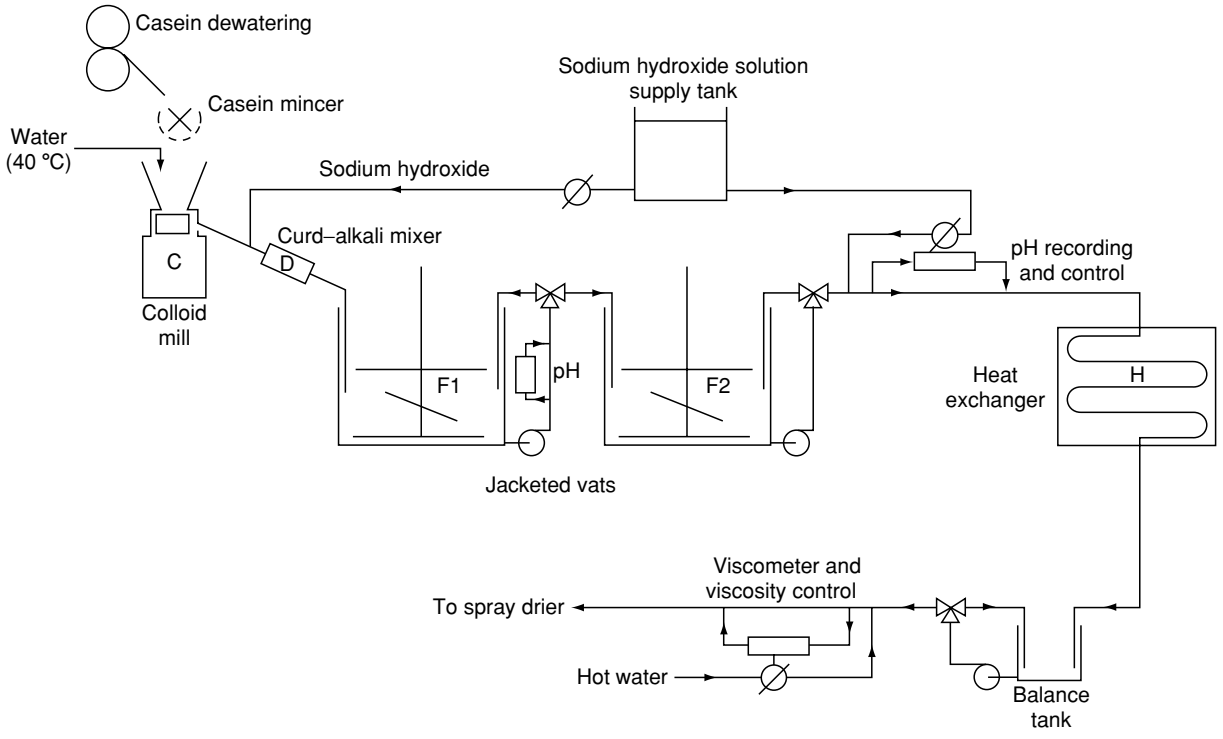

Figure 2 Suggested plant layout for dissolving casein for sodium caseinate manufacture. Dewatered casein curd is minced, mixed with water, and finely milled in a colloid mill (C). The curd–water slurry is then mixed (D) with dilute sodium hydroxide solution and transferred to the first of two jacketed dissolving vats (F_1), each equipped with an agitator. The sodium caseinate solution is subsequently pumped from the second dissolving vat (F_2) and extra alkali is added, if necessary. The solution is heated by means of a tubular heat exchanger (H) and then pumped via a balance tank to the spray (or roller) drier for drying. The viscosity of the solution may be monitored and reduced, if necessary, by addition of hot water. From Southward CR (1985) Manufacture and applications of edible casein products. I. Manufacture and properties. *New Zealand Journal of Dairy Science and Technology* 20: 79–101, with permission.

The typical composition of sodium caseinate is shown in Table 1. Acid (low-calcium) coprecipitates can be dissolved in alkali in a similar manner to that used for acid casein. Both rennet casein and high-calcium coprecipitate (precipitated at a pH of 6 or greater) are usually rendered soluble by means of complex phosphates, such as sodium tripolyphosphate, to produce translucent solutions with a viscosity somewhat higher than that of the corresponding caseinates.

In the great majority of applications, in both edible and technical (nonfood) uses, casein must first be made soluble before it can be used in its final application. Although some users convert the dry casein to caseinate themselves, others tend to purchase the caseinate directly from the producer (often as a 'fresh-curd' caseinate). For applications in foods. (*See* **Casein and Caseinates:** Uses in the Food Industry.)

See also: **Casein and Caseinates**: Uses in the Food Industry; **Cheeses**: Starter Cultures Employed in Cheesemaking; Chemistry of Gel Formation; **Drying**: Fluidized-bed Drying; **Lactic Acid Bacteria**; **Pasteurization**: Principles; **Starter Cultures**; **Whey and Whey Powders**: Production and Uses; Protein Concentrates and Fractions; Fermentation of Whey

Further Reading

Australian Society of Dairy Technology (1972) *Casein Manual*. Parkville, Melbourne: Australian Society of Dairy Technology.

Farrell HM Jr (1988) Physical equilibria: proteins. In: Wong NP, Jenness R, Keeney M and Marth EH (eds) *Fundamentals of Dairy Chemistry*, 3rd edn, pp. 461–510. New York: Van Nostrand Reinhold.

Muller LL (1971) Manufacture and uses of casein and coprecipitates: a review. *Dairy Science Abstracts* 33: 659–674.

Muller LL (1982) Manufacture of casein, caseinates and coprecipitates. In: Fox PF (ed) *Developments in Dairy Chemistry*, vol. 1. *Proteins*, pp. 315–337. London: Applied Science.

Mulvihill DM (1989) Caseins and caseinates: manufacture. In: Fox PF (ed) *Developments in Dairy Chemistry*, vol. 4, *Functional Milk Proteins*, pp. 97–130. London: Elsevier Applied Science.

Mulvihill DM (1992) Production, functional properties and utilization of milk protein products. In: Fox PF (ed) *Advanced Dairy Chemistry*, vol. 1, *Proteins*, pp. 369–404. London: Elsevier Applied Science.

Southward CR (1985) Manufacture and applications of edible casein products. I. Manufacture and properties. *New Zealand Journal of Dairy Science and Technology* 20: 79–101.

Southward CR (1994) Utilization of milk components: casein. In: Robinson RK (ed) *Modern Dairy Technology*, 2nd edn, vol. 1, *Advances in Milk Processing*, pp. 375–432. London: Chapman and Hall.

Southward CR and Walker NJ (1980) The manufacture and industrial use of casein. *New Zealand Journal of Dairy Science and Technology* 15: 201–217.

Spellacy JR (1953) *Casein, Dried and Condensed Whey*. San Francisco, California: Lithotype Process.

Whitney RMcL (1988) Proteins of milk. In: Wong NP, Jenness R, Keeney M and Marth EH (eds) *Fundamentals of Dairy Chemistry*, 3rd edn, pp. 81–169. New York: Van Nostrand Reinhold.

Uses in the Food Industry

C R Southward, Fonterra Research Centre, Palmerston North, New Zealand

Introduction

Although casein, as it exists in milk, has been consumed as 'food' for thousands of years, the extracted form of casein had very little application in foods prior to 1960. Instead, from early in the 20th century, and in some cases before then, it was used as an adhesive in wood glues and paper coating, and in paints, fibers, plastics, and leather finishing. The introduction of coffee whiteners and whipped toppings, in particular, by the food industry in the USA during the 1960s played a significant part in the establishment of casein products in foods. These two product groups were based on vegetable fat and contained casein in a water-soluble form (usually sodium caseinate), together with carbohydrate, emulsifiers, and stabilizers. They were promoted as 'non-dairy' foods. (As casein was derived from milk, it was not considered to be a dairy product. Furthermore, it had been classified as a chemical because of its long-established use in nonfood technical applications in adhesives, paints, etc., as described above.) Nondairy coffee whiteners and whipped toppings were thus presented as alternatives to the traditional dairy products of milk, cream, and whipping cream. They offered advantages of lower price (because the (imported) raw materials from which they were made were cheaper than the domestic milk solids they were replacing), convenience, and shelf stability (several of the products were sold in powder form). Other nondairy products followed, many of them containing casein, e.g., so-called imitation milks, imitation cheese, and salad dressings. Food supplements and dietary products were also produced and a number of these were based on casein. This article reviews the use and function of casein products in food.

General

The main applications of casein products in foods, together with their functions in such foods, are shown in Table 1. As a high-quality protein, casein can provide nutrition in foods, although this is often not the main reason for its use. It is frequently incorporated into food products for its ability to impart so-called 'functional' properties of whipping, foaming, fat emulsification, film formation, water binding, and thickening. Consequently, casein can have a significant effect on the texture or consistency of a food.

Animal Feeding and Pet Foods

During the middle of the 20th century, many experiments were undertaken in the USA and, to a lesser extent, in Europe and elsewhere, to determine the effect of feeding cows with iodine and iodinated casein (sometimes referred to as 'thyroprotein') on their nutrition and milk yield. A number of studies suggested that milk yield was increased, but there was sufficient conflicting evidence for the practice not to be widely adopted. Extensive nutritional feeding trials for ruminants were undertaken in the late 1960s and 1970s to determine the effect – on milk production and wool growth, respectively – of feeding 'protected' casein to cows and sheep. The casein was protected by reaction with formaldehyde, which prevented it from being broken down by the rumen microflora. It could pass undigested to the abomasum where the acidic conditions were able to promote the digestion of the casein. These studies were also extended to determine how the feeding of polyunsaturated oils, encapsulated in formalin-treated caseins, to ruminants affects the quantity of polyunsaturated fat in meat and milk. Again, however, this was not adopted commercially.

Casein (generally in the form of sodium caseinate) may be used as a nutritional supplement in pet foods and calf milk replacers, and it has also been reported as an ingredient in fish food. Some recent reports in the literature indicate that casein can be used to produce novel textured pet foods, as well as the more traditional nutritional foods.

Bakery Applications

The water-binding properties of casein (as well as nutrition) are important for bakery applications in which casein products are incorporated. Soluble forms of casein (such as caseinates) tend to bind too much water and may make some foods too sticky or 'doughy.' Accordingly, insoluble or partly soluble casein products are often used, because these have less water binding than the fully soluble caseinates. Some bakery products that have been made with caseins include doughnuts, waffles, cake mixes, and bread. A number of examples of the use of casein products in bakery applications, taken from the recent literature, are shown in Table 2. Figure 1 illustrates the use of casein products in biscotti.

Cheese-like Foods

Casein has been used as an extender and texture modifier in processed cheese, and even to increase the yield of cheese from cheese milk (e.g., in Europe). However, the most significant use of casein in cheese products has probably been in the production of

Table 1 Use and function of casein products in foods

Application	*Principal functions*
Animal and pet foods	Nutrition, binder for moisture and fat
Bakery products	Nutrition, texture, water binding
Beverages, including soups	Emulsifier, stabilizer, nutrition
Coffee whiteners and creamers	Resistance to feathering, emulsifier
Confectionery	Nutrition, texture
Cultured products	Stabilizer, emulsifier, consistency, nutrition
Extruded snack foods	Texture
High-fat powders and shortenings	Emulsifier, whipping
Icecream, mousse, and desserts	Whipping, foaming, emulsifier
Imitation cheese and other cheese-like foods	Texture, emulsifier
Infant foods	Nutrition
Instant breakfasts and dietary preparations	Nutrition
Meat products	Emulsifier, water binding
Pasta	Nutrition, texture
Pharmaceutical products	Nutrition, therapeutic, film formation
Spreads	Texture, stabilizer
Whipped toppings	Stabilizer, emulsifier

Table 2 Use of casein products in bakery applications

Examples	*Casein product used*	*%*	*Use/function of casein product*
Bread	Sodium caseinate	2–30	Emulsification
Glazing agent	Casein, casein salt	5–15	Film formation
Blini (pancake)	Sodium caseinate	2.9–5	Increased storage life, friability
High-protein wafer	Casein	≥ 50	Texture, nutrition

Figure 1 (see color plate 17) Biscotti, representing baked products that can be made using high-calcium milk protein. Courtesy of NZMP (North America) Inc.

Figure 2 (see color plate 19) Pizza, representing the use of imitation Mozzarella cheese that can be made using rennet casein. Courtesy of NZMP (Wellington) Ltd.

so-called 'imitation' or synthetic cheese (often referred to as cheese analogs). This product, similar in properties to processed cheese, is made from water, casein (e.g., acid casein, rennet casein, or sodium or calcium caseinate, singly or in mixtures), vegetable fat, stabilizers, and emulsifiers. The food industry in the USA was first to develop imitation cheese (at least in name), which could be seen as a similar development in the cheese industry to the much earlier production of margarine as an 'imitation' butter. Just as coffee whiteners and whipped toppings (and artificial milks) were developed in the US food industry in the 1960s as cheaper and/or convenience foods to compete with milk and cream (see elsewhere in this article), so imitation cheese appeared. The commercial production of imitation cheese first occurred in the USA in the early 1970s, and the quantity produced grew, during a period of several years, to a maximum of about 5% of the total cheese sales in the USA. Imitation cheese found its application as a cheaper alternative to natural cheese, particularly in fast-food outlets, in frozen pizza and hamburgers, for instance. Figure 2 illustrates the use of rennet casein in an imitation Mozzarella cheese, used on a pizza. In consideration of its nutritional properties, imitation cheese was also approved for use in the American school lunch program. The main function of the casein product in this application is to provide the required body and texture, with emulsification of fat, as well as the melting properties of the finished cheese in pizza. This last-named function is achieved by careful formulation, using food acids and stabilizing salts to achieve the correct pH and mineral balance in the cheese that will cause it to melt during the cooking of the pizza to provide the expected 'stretch' and 'stringing' properties required for pizza cheese. Although the use of imitation cheese in the UK and Japan has been mentioned in the literature, its main area of commerical significance is in the USA.

Other applications of casein in the cheese food industry include surface coating of whole cheese and cheese granules, using the natural film-forming properties of casein products. Several examples of the use of casein products in cheese-like foods, taken from the recent literature, are shown in Table 3.

Beverages

Imitation Milks

In the USA, so-called 'filled milk', which contains vegetable fat and skimmed milk solids, has been

Table 3 Use of casein products in cheese-like foods

Examples	*Casein product used*	*%*	*Use/function of casein product*
Simulated cheese	Rennet casein	≥25	Texture, nutrition
Coating of cheese pieces	Casein	2	Free-flow agent
Fat-free cheese analogs	Rennet casein	15–35	Texture
Protective protein coating for cheese	Sodium caseinate/ casein	3–14	Film coating
Imitation cheese	Calcium caseinate	12	Texture

Table 4 Use of casein products in beverages, including coffee whiteners and creamers

Examples	*Casein product used*	*%*	*Use/function of casein product*
Soluble tea product	Acid casein	10–40[a]	Inhibits tea cream formation
Powder for cream liqueur	Sodium caseinate	4–14[a] 0.8–2.8[b]	Emulsion stabilizer
Foaming coffee whitener	Caseinate	1–2.5	Whiteness and foaming
Stabilizer for wine	Alkali caseinate	5–60[c]	Stabilizer against changes in color or taste
Coffee whitener	Casein	0.1–10	Emulsion stabilizer
Iced chocolate	Calcium caseinate	5.6–6.7	Texture/fat replacer

[a]Dry basis.
[b]In powder.
[c]Expressed as a percentage of the stabilizer.

sold for many years. In the late 1960s and 1970s, a new class of milk – 'imitation milk' – appeared on the market. This contained vegetable fat and various other ingredients, including protein as sodium or potassium caseinate or from soy beans, and a carbohydrate source such as corn syrup solids. Artificial milks were also reported in the UK and the USSR. The lower cost of the ingredients and the absence of lactose (for which some people show intolerance) were no doubt significant factors in the establishment of such products. However, concerns were voiced over the generally lower nutritional quality (e.g., lack of vitamins and minerals and, sometimes, lower protein content) and often poorer flavor compared with 'normal' cows' milk. Casein products have also been used to fortify fresh milk, though this has occurred in Europe rather than in North America.

Several examples of the use of casein products in beverages, taken from the recent literature, are shown in Table 4.

Cream Liqueurs

Casein, in the form of sodium caseinate, has found application recently in cream liqueurs, especially in the UK and the Irish Republic, where they have been, apparently, one of the fastest-growing markets for cream. Other ingredients include sugar, alcohol, and trisodium citrate. In this application, the sodium caseinate acts mainly as an emulsifier or emulsion stabilizer.

Coffee Whiteners and Creamers

As shown in Table 1, the main functions of casein (as sodium caseinate) in coffee whiteners are to provide emulsification and to promote resistance to 'feathering.' It has been claimed that sodium caseinate (or other protein) also imparts body, provides some whitening (though this is mainly due to the fat) and improves flavor. Although both liquid and dry (powdered) coffee whiteners have been produced, the most popular by far appear to be the latter because, as mentioned in the introduction to this article, they have the advantages of being shelf-stable and convenient to use. In addition, the low price of the major ingredients (vegetable fat and carbohydrate, usually as corn syrup solids) compared with that of (US domestic) milk solids makes coffee whitener an attractive alternative to milk or cream. Figure 3 shows the use of a coffee whitener that contains high-calcium milk protein.

Soups, Gravies, etc.

When casein products are used in soups, their purpose may be for nutrition (e.g., as calcium caseinate) or for increasing the consistency or viscosity of the mixture (possibly in the form of sodium caseinate), as illustrated in Figure 4. Occasionally, hydrolysates of casein products may be used in soups and gravies for flavor enhancement.

Figure 3 Coffee whitener/creamer, representing imitation dairy products that can be made using high-calcium milk protein. Courtesy of NZMP (Wellington) Ltd.

Sour and Cultured Milk Products

In the manufacture of yogurt, casein products have been incorporated for protein enrichment, and improvement in consistency and stability; they give less syneresis than standard yogurt products.

Imitation sour cream products have been produced from vegetable fat, sodium caseinate, and other ingredients. The purposes of adding sodium caseinate to these products are to act as a stabilizer, to increase the consistency, and to emulsify the fat – functions that sodium caseinate performs in mayonnaise as well. Although some publications have claimed nutritional fortification as a reason for incorporating casein/caseinates in cultured milk products, this does not seem to be as significant a reason for their use in this context as the 'functional' properties mentioned earlier. Some examples of the use of casein derivatives in cultured and sour milk products, taken from the recent literature, are shown in Table 5.

Figure 4 Canned creamed soup, representing convenience foods that can be made using sodium caseinate. Courtesy of NZMP (North America) Inc.

Desserts, Icecream, Whipped Toppings, and Confectionery

Sodium caseinate functions in frozen desserts and icecream as a stabilizer (Figure 5), improves the whipping properties of the mix and imparts 'body' (also known as consistency or thickening). Other properties of sodium caseinate that are important in these products include emulsification, foam stability, and film formation.

The incorporation of casein products in icecream has not been commercially extensive, generally because of various legal restrictions. Several icecream substitutes, however, did appear in the USA in the 1960s (e.g., Mellorine) and some of these products incorporated casein (usually as sodium caseinate) into their formulations.

As described in the introduction to this article, whipped toppings which contained vegetable fat, water, and sodium caseinate (together with other ingredients) appeared in the USA in the 1960s as specific competitors for whipping cream. In this application, sodium caseinate is used in forming a film to trap aerating gases. It also functions in fat encapsulation, as a bodying agent and as a stabilizer. Hydrolyzed caseinates, with enhanced whipping and foaming properties, are also produced for food products like these. Such products are now manufactured in other countries, including the UK, Italy, Israel, and

Table 5 Use of casein products in sour and cultured milk products

Examples	*Casein product used*	*%*	*Use/function of casein product*
Fat-free cream cheese	Caseinate	8–12	Fortification, (mouth feel)
Mayonnaise	Casein	0.7	Emulsifier
Crème fraiche-type	Casein	9.5–40[a]	Texture, stabilizer
Mayonnaise	Calcium caseinate	5–6.6	Texture, fat replacer
Salad dressing	Sodium caseinate	0.9[b]	Restricts oxidation

[a]Dry basis.
[b]Expressed as a percentage of the emulsion.

Figure 5 (see color plate 20) Frozen dairy dessert, representing dairy products and analogs that can be made using milk protein isolates. Courtesy of NZMP (North America) Inc.

Japan. A number of examples of the use of casein products in desserts and whipped toppings, taken from the recent literature, are shown in Table 6.

Casein products are used much less in confectionery than are other milk products (e.g., nonfat dry milk, whole milk powder, and sweetened condensed milk). Nevertheless, their use has been described in toffees, caramels, and fudges, where they can form a tough, chewy body, and in marshmallow and nougat to produce heat-stable whips and foams. Casein proteins have also been used to produce high-protein chocolate snacks and confectionery sticks or bars (Figure 6).

Spreads and Fat Replacers

In the wealthier and industrialized countries, in particular, concern over health and nutrition has been focused, at least to some extent, on maintaining a healthy diet and weight. 'Fat,' both in the sense of appearance and in food intake, has taken on a generally undesirable image, and butterfat (with its generally high content of saturated fats), in particular, appears to have been a high-profile target of the health sector of most of these countries. One of the responses to these concerns has been the introduction of lower-fat spreads to complement or replace the traditional fats. In some of these products, casein (usually in the form of sodium caseinate) has been used to stabilize the water phase and to improve the texture of the product. In addition, casein has been used in some (processed) cheese spreads, presumably for similar (functional) reasons.

Table 6 Use of casein products in desserts, icecream and whipped toppings

Examples	*Casein product used*	*%*	*Use/function of casein product*
Jelly food base	Casein solution	2–20	Water binding
Acid dessert	Sodium caseinate	0.01–0.3	Stabilizer
Whipping cream	Casein	≤ 5	Stabilizer
Gelling agent	Casein	37.1	Water binding/replacer for gelatin
Chocolate mousse	Calcium caseinate	4–5	Texture, fat replacer

Figure 6 (see color plate 21) Candy bars, representing confectionery products that can be made using modified milk proteins and milk protein isolates. Courtesy of NZMP (North America) Inc.

From a different perspective, the food industry (again starting in the USA) developed some fat-replacer foods that exhibited the mouth feel and other desirable properties of fat-based products, without containing any fat. These fat replacers were based on water-insoluble proteins or carbohydrates, with about half the energy–weight ratio of fat. They were very finely milled to minimize the detection of any individual particles by the tongue but, nevertheless, exhibited the desirable 'creamy' mouth feel found in fat-based foods, such as cream, for example. Products such as Simplesse™ were introduced to the market, and generated considerable interest. A range of alternative products that contained various fat replacers came on to the market with claims for helping weight reduction. A number of foods that are claimed to provide fat replacement are shown in Table 7. These examples have been extracted from the recent literature.

Meat and Fish Products

Comminuted Meats

The reason for using casein products (usually as sodium caseinate) in comminuted meat products such as sausages and mincemeat is to emulsify the fat, bind water, and generally improve consistency. Sometimes, milk protein may be added for nutritional purposes. The use of casein in meat products is strictly controlled by legislation in most countries and the quantity that may be added is generally less than 5% of the weight of the meat. One estimate suggests that the potential world market for the use of high-protein milk products in processed meats is about 100 000 tonnes.

Textured Protein

Meat analogs (with a texture resembling the fibrous nature of meat) have been prepared from casein, either using a spinning technique (similar to methods used for producing textile fibers in the 1930s, in which a protein solution is forced through spinnerets into a coagulating bath) or by production of chewy, meat-like gels. The general disadvantage of using casein in this form is that, when the fibers are heated in a moist environment, they tend to melt together and lose their individual fibrous structure (which is why casein readily forms a smooth plastic or molten mass in the production of processed cheese – or casein plastic for buttons). However, the addition of heat-denaturable proteins (such as egg white or whey protein) can overcome that disadvantage.

Fish Products

Just as casein is used for fat emulsification and water binding in comminuted meat products, so it has found some application in a number of fish products, especially those formed from minced fish and in paste-like forms. The examples in Table 8, which have been taken from the recent literature, show some uses of casein or caseinates in meat and fish products that utilize a number of properties of casein other than those commonly ascribed to casein in this application.

Pasta, Snacks, and Fried Foods

When casein is used in pasta products, it is mainly to enhance their nutritional quality because the high

Table 7 Use of casein products in fat replacers

Examples	*Casein product used*	*%*	*Use/function of casein product*
Fat-free cream cheese	Caseinate	8–12	Fortification, (mouth feel)
Fat substitute in foods	Caseinate	50–60[a]	Structure ('unctuous' texture)
Coffee whitener	Sodium caseinate	10–15[b]	Fat substitute
Fat-free cheese analogs	Rennet casein	15–35; 20–30[c]	Texture

[a]Expressed as a percentage of the total protein.
[b]Wet basis.
[c]Preferred range.

Table 8 Use of casein products in real and imitation meat and fish products

Examples	*Casein product used*	*%*	*Use/function of casein product*
Fibrous meat substitute	Calcium caseinate	≥ 30[a]	Texture
Powdered oil flavor	Sodium caseinate	1–10[b]	Emulsion stabilizer
Meat patties	Sodium caseinate	2–4	Increases storage life
Minced fish	Calcium caseinate	8	Cryoprotectant

[a]w/w.
[b]Wet basis.

Table 9 Use of casein products in snacks, fried foods, pasta, etc.

Examples	*Casein product used*	*%*	*Use/function of casein product*
Emulsion for dough for frying	Casein	1–5	Reduces oil absorption
Instant food products	Caseinate	≤ 5	Water binding
Coating for potatoes	Caseinate	0.05–10	Emulsifier
Powder for fried food in microwave oven	Casein	10–50	Coating
Noodles	Casein	20–80	Oil absorption

lysine content of the casein complements the low lysine- and high sulfur-containing amino acids of wheat and other vegetable proteins present in such products. In some cases, casein plays an important part in forming a suitable texture matrix for the pasta product. Thus, casein products have been reportedly used in macaroni, spaghetti, rice, and noodles.

Casein products are used in extruded (high-protein) snack foods to produce the required texture and for nutritional fortification. In fried foods, it appears that the film-forming properties of casein may also be used, either as coatings or to reduce oil absorption during the cooking/frying stage in preparing the food. A number of examples of the use of casein in pasta and fried foods, taken from the literature, are shown in Table 9.

Edible Films

The film-forming properties of soluble casein products, such as sodium and calcium caseinates, have been referred to earlier in Tables 1 and 3. This specific property of casein products was recognized many years ago, when casein was used in technical, non-food applications such as paper coating, and in paints and leather finishing. Commercially, the film-forming ability of caseinates in foods is seen at the oil–water interface in the formation of emulsions in coffee whiteners, desserts, and whipped toppings, and at the air–water interface in whipping and foaming products (such as whipped toppings). These same properties have been used in the production of sausage skins. More recently, caseinates have been used to coat fruit and vegetable products as a means of maintaining their freshness. As a hydrophilic (water-loving) substance, casein produces films that, on their own, are not very resistant to the passage of moisture vapor. However, they do appear to provide good mechanical strength, and to act as an oxygen barrier. Research has therefore included the study of combined films of casein/caseinate and lipid material in which the lipid repels or resists the passage of moisture vapor and the caseinate provides the matrix and mechanical strength, with resistance to the passage of oxygen. Some examples from the recent literature are shown in Table 10.

Microencapsulation

Encapsulation of oils was practiced in feeding trials for animals in the early 1970s, as referred to earlier in this article. This technique was used as a means of conveying a nutritional supplement, in a protected environment, through a part of the animal's digestive tract (where it would otherwise have been broken down) to the area where the supplement would have the intended benefit. Often referred to as microencapsulation, this principle has been used now for many years as a means of applying drug therapy in humans and in animals, and as a means of preventing undesirable odors and flavors of beneficial compounds from being rejected as having an unacceptable smell or producing a foul taste. Although not as widely used for microencapsulation as gelatin, casein products do find application in these fields, as indicated in the examples from the recent literature that are shown in Table 11.

Pharmaceutical and Medical Applications

Various casein preparations have been used in a range of pharmaceutical applications. These include casein

Table 10 Use of casein products in edible films

Examples	*Casein product used*	*%*	*Use/function of casein product*
General			
Edible film	Lactic acid casein, rennet casein	37.5–59	Strength, moisture and oxygen barrier
Irradiated edible film	Sodium or calcium caseinate	33–50	Mechanical strength, flexibility
Ultrasound-treated edible film	Sodium caseinate	100	Mechanical strength, moisture barrier
Fruits and vegetables			
Coating for apples and celery sticks	Calcium caseinate	50[a]	Moisture barrier
Edible coating for zucchini	Calcium caseinate	15–45[a]	Barrier for moisture
Film for fruit and vegetables	Sodium caseinate	0–50[a]	Moisture barrier
Food containers			
Biodegradable food containers	Casein	0.1–10	Water resistance
Edible film for foods (coconut, cereal, peanuts, almonds)	Casein/caseinate	83–86	Barrier for liquids and flavors
Biodegradable flexible packaging film	Casein/caseinate	33–50	Barrier to water, gases, and microorganisms

[a]Expressed as a percentage (dry basis) of the coating weight.

Table 11 Use of casein products in microencapsulation

Examples	*Casein product used*	*%*	*Use/function of casein product*
Nutrition			
Microparticles containing physiologically active substances	Casein	NS	Forms 'shell' and acts as edible binder
Microencapsulated oil or fat for nutrition and infant formula	Caseinate	NS	Matrix formation
Odors and flavors			
Film for liver powder	Caseinate	'Major part'	Odor barrier
Film for garlic	Caseinate	NS	Flavor and odor barrier
Film for horseradish paste	Casein	NS	Odor barrier
Film for fruit	Caseinate	NS	Color and flavor barrier

NS, not stated.

Table 12 Pharmaceutical and medical applications of casein products

Examples	*Casein product used*	*%*	*Use/function of casein product*
Coated cytochrome-c drug for intestinal absorption	Casein	NS	Coating
Nitrogen source for patients	Casein/caseinate	7–12 (of protein)	Nutrition
Preventing agents for hyperlipidemia	Casein/caseinate	$\leq$100	Binds fat or cholesterol
Hard capsules	Casein	20–80	Matrix
Ferro-succinyl casein complex	Casein	[100]	Nutrition

NS, not stated.

hydrolysates and solutions of caseinates, TMP® and/or milk protein hydrolysates for intravenous nutrition in intestinal disorders, veterinary medicine, and disorders involving protein metabolism. Some preparations of casein have been used for feeding patients following surgery, as a therapeutic agent in dressing wounds, in cosmetics, in toothpaste for inhibition of dental caries, and in hair shampoos. The potential for producing specific drugs, such as opiates, from casein has been described, and other extracts from casein have been claimed as being suitable for treatment of arthrosis (pain in a joint) and gastric ulcers, for enhancement of calcium and iron absorption, and to augment immunity. Table 12 illustrates some recent published claims for applications of casein products in this field.

Health and Nutrition

As indicated earlier in this article, a significant proportion of the population in the wealthier and industrialized countries has some concern with regard to their health and nutrition. Such concern has been seen in the great proliferation of nutritional supplements for loss or gain of weight. Other supplements are sold to provide a certain percentage of the recommended dietary allowance of various vitamins and minerals, fiber, protein, etc. Preparations that contain casein

Table 13 Use of casein products in health and nutrition, weight loss, and satiation

Examples	*Casein product used*	*%*	*Use/function of casein product*
Trace mineral composition	Caseinate	≥90	Complexing agent
Beverage/health drink	Caseinate	1–7	Accelerates muscle activity
Magnesium or calcium caseinate	Casein	[100]	Promotes absorption of Mg and Ca
Rice-based infant cereal	Casein phosphopeptide	0–2 g/serving	Ca and Zn absorption promoter
Satiation			
Food composition	Casein/caseinate	33–67[a]	Satiety
Appetite suppressant	Casein/sodium caseinate/calcium caseinate	2.5–4.8	Prevents overeating

[a]Dry basis.

Figure 7 (see color plate 18) Chocolate drink, representing products for health and nutrition that can be fortified using milk protein isolates and calcium, potassium, or magnesium caseinates. Courtesy of NZMP (North America) Inc.

products are among these supplements and a selection of these, taken from the published literature, is included in Table 13. Figure 7 shows a chocolate-flavored health drink that contains calcium sodium caseinate and/or milk protein isolate (consisting of caseins and whey proteins).

Conclusion

In the majority of the applications described above, where casein is used to provide a specific functional or physical effect, it is incorporated in the food in relatively small quantities, i.e., generally below 10% by weight of the product. Unlike other dairy products, such as milk and milk powder, butter and cheese, it is not consumed as a food *per se*. However, when casein is used for nutrition, it may indeed represent a significant proportion of the total weight of the food. In pharmaceutical and medical applications, there are some indications that derivatives of casein may be important therapeutic agents in the future.

See also: **Cakes**: Methods of Manufacture; **Cheeses**: Processed Cheese; **Fats**: Fat Replacers; **Liqueurs**: Cream Liqueurs; **Low-fat Foods**: Types and Manufacture; Low-fat Spreads; **Meat**: Sausages and Comminuted Products; **Milk**: Dietary Importance; **Pasta and Macaroni**: Methods of Manufacture; **Protein**: Functional Properties; Digestion and Absorption of Protein and Nitrogen Balance; **Satiety and Appetite**: The Role of Satiety in Nutrition; Food, Nutrition, and Appetite; **Yogurt**: The Product and its Manufacture

Further Reading

Centre National du Commerce Extérieur (1970) *The United States Market for Edible Caseins and Caseinates*. Paris: Centre National du Commerce Extérieur.

Jelen P and Lutz S (1998) Functional milk and dairy products. In: Mazza G (ed.) *Functional Foods: Biochemical and Processing Aspects*, pp. 357–380. Lancaster, Pennsylvania: Technomic.

Manson W (1980) The use of milk and milk constituents in pharmaceutical preparations. *Factors Affecting the Yields and Contents of Milk Constituents of Commercial Importance*. International Dairy Federation. Bulletin, document 125, pp. 60–65. Brussels: International Dairy Federation.

Meisel H and Schlimme E (1996) Bioactive peptides derived from milk proteins: ingredients for functional foods? *Kieler Milchwirtschaftliche Forschungsberichte* 48: 343–357.

Muller LL (1971) Manufacture and uses of caseins and co-precipitate. *Dairy Science Abstracts* 33: 659–674.

Mulvihill DM (1992) Production, functional properties and utilization of milk protein products. In: Fox PF (ed.) *Advanced Dairy Chemistry*, vol. 1, pp. 369–404. London: Elsevier Applied Science.

Schuette HA (1939) Alimentary and medicinal uses of casein. In: Sutermeister E and Browne FL (eds) *Casein*

and its Industrial Applications, 2nd edn, pp. 366–390. New York: Reinhold.
Southward CR (1989) Uses of casein and caseinates. In Fox PF (ed.) *Developments in Dairy Chemistry*, vol. 4, pp. 173–244. London: Elsevier Applied Science.
Southward CR (1994) Utilisation of milk components: casein. In: Robinson RK (ed.) *Modern Dairy Technology*, 2nd edn, vol. 1, *Advances in Milk Processing*, pp. 375–432. London: Chapman and Hall.
Southward CR and Walker NJ (1982) Casein, caseinates and milk protein co-precipitates. In: Wolff IA (ed.) *CRC Handbook of Processing and Utilization in Agriculture*, vol. I, pp. 445–552. Boca Raton, Florida: CRC Press.
Webb BH (1970) Miscellaneous products. In: Webb BH and Whittier EO (eds) *Byproducts from Milk*, pp. 285–330. Westport, Connecticut: AVI.

CASHEW NUTS AND CASHEW APPLES

J F Morton†, University of Miami, Coral Gables, FL, USA

Introduction

The cashew is one of the two costliest nuts marketed in quantity worldwide. Formerly second to the macadamia nut in retail value, it is now more expensive because of short supply. Few people in the northern hemisphere question why the cashew nut is never sold in the shell as peanuts, pecans, walnuts, and almonds often are. There is a good reason for this. The cashew, by nature, is much more complex than other nuts and a technical explanation of its processing is required.

Description

The cashew tree, *Anacardium occidentale*, of the family Anacardiaceae, is short-trunked, up to 13 m high and normally with a very broad crown, although it is often stunted and bushy on coasts. There is an enormous, very old, wild cashew tree in Natal, Brazil, that covers 0.75 ha and is a great tourist attraction. Its leathery, evergreen leaves are clustered at the branch tips. The small, yellow-and-red flowers are borne in open sprays. The true fruit of the tree is the kidney-shaped, hard-shelled nut that is at first green, later turning an ashy brown; it grows to 3 cm or more in length and develops at the tip of a fleshy stalk. As the 'nut' matures, the stalk inflates to form a showy, pear-shaped, smooth-skinned, succulent, juicy pseudofruit (false fruit), which is bright-red, orange, yellow, or two-tone and is usually viewed, and utilized, as a fruit (**Figure 1**). The weight of the expanded pseudofruit causes the nut to fall to the ground at its peak of maturity. A caustic oil in the honeycomblike cells within the double-layered shell of the 'nut' protects it from being destroyed by foragers that feed on the 'apple.' However, the oil seriously complicates the processing of the 'nut' and extraction of its kernel for food use.

Origin and Distribution

The cashew tree is native from southern Mexico to Peru, Brazil, and the West Indies. Wild stands extend along the coast of Brazil from Pará to Rio de Janeiro, and the nut and apple have always been of importance to the indigenous people. The cashew was one of the first tropical American fruit trees to be introduced by early Portuguese and Spanish voyagers into the tropics of the Old World. In the sixteenth century the Portuguese planted it, especially on the west coast of India and the east coast of Mozambique, in order to halt soil erosion and, in time, it formed extensive forests.

The tree soon became commonly cultivated and naturalized at low altitudes in East, Central, and West Africa, South-east Asia, Sri Lanka, Malaysia, the Philippines, Mauritius, and the Seychelles. In addition, it has been planted around villages in most of the Pacific Islands, including Hawaii, and is occasionally grown in dooryards in southern Florida.

The name 'cashew' is derived from the Brazilian acaju, usually abbreviated to caju. In the Orinoco region, the tree is known as pauji. In all the Spanish-speaking countries of Latin America and the Caribbean area, the common name is marañon, except in Venezuela where it is called merey. In French it is cajou or acajou.

Climate and Soil

The tree requires a tropical or subtropical climate. It is highly sensitive to frost when young, but later is able to withstand brief cold spells. It is well adapted to sandy soil, intolerant of heavy clay, and is drought- and

†Deceased.

Figure 1 The cashew nut, shaped like a boxing glove, develops at the tip of a fleshy, juicy, fruitlike stalk (peduncle), called cashew apple, which enlarges as the nut matures, becomes red, orange, yellow, or two-tone, and is eaten raw or preserved. Reproduced from Cashew Nuts and Cashew Apples, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

salt-tolerant, succeeding on land too poor for most other crops. It thrives from sea level to an altitude of 1000 m.

Commercial Development

India pioneered the domestication and commercialization of the cashew and, for a long time, augmented supply by importing raw nuts from Mozambique. The 1950s saw large-scale planting of cashew trees in Africa, India, and Brazil. In north-eastern Brazil there was first a private plantation of about 1500 ha. Then, in 1957, a programme was begun to plant a million trees in the state of Ceará, the government supplying the seeds and a bonus for each tree set out in plots of 100 or more. From 1957 to 1972, a company in Pacajus set out 2000 ha. Another surge of government-stimulated plantings totalled nearly 136 000 ha by 1972, mostly in Ceará. Later, 18 000 ha were developed in Rio Grande do Norte. By 1986, more than 40 000 ha of cashew had been added in reforestation programs in Ceará, Piaui, Rio Grande do Norte, and Bahia.

There have been sporadic attempts to develop cashew industries in Venezuela, Colombia, Peru, Guatemala, Belize, and Jamaica. Some have been moderately successful. The primary problem is that cashew trees are still mainly grown from seed and there is great variation in yield, and great variability in the form and size of the nuts from seedling trees (Figure 2). The nuts may range from 3 to 32 g in weight and the apples (mostly orange or yellow) from 15 to 650 g. Trees bearing small nuts tend to produce heavily, while trees with large nuts and apples are less productive and may bear in alternate years.

Much research in Mozambique and India has been devoted to techniques of vegetative propagation (by cuttings, grafting, budding, or air layering) to achieve uniformity, but it is still not widely practiced, mainly because of the costs involved. India has accomplished much in rejuvenation of old groves by top-working beheaded, unproductive trees. Side-veneer grafting has been recommended in Venezuela. In Trinidad and India air-layered trees have poor form and tend to blow over in the wind. Once a method of propagation has been adopted, there must be a search for the best types to be grown. An experimental planting of grafted trees was initiated in Ceará but did not prosper because of a prolonged drought. In 1980, a selected clone of a 'dwarf' cashew was grafted on to *Anacardium microcarpum* and the 40 trees set out fruited early and were highly drought-resistant. This trial encouraged experiments with other rootstocks and selection of high-quality clones, but low yields

Figure 2 Cashew nuts range from 2.5 to 5 cm in length and this variability, together with the odd shape, contributes to the difficulties of opening them mechanically. They are never sold in the shell. Reproduced from Cashew Nuts and Cashew Apples, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

have caused such work to languish. There are still no commercial orchards of selected, vegetatively propagated trees.

Varieties

Generally, named varieties have been distinguished merely by color and size of the apple, such as the Vermelho (yellow) and Amarello Gigante (large yellow) of Brazil, and the Marañon Amarillo (yellow) and Marañon Rosado (red) of Colombia. Natural hybrids in Colombia have been named La Gigante (huge) del Rio Magdelena, La Larga (long) de Nazareth, and La Pequeña (small) del Meta. Usually the nut and apple are larger in the yellow forms than in the red, and the apple is less astringent. One name pertaining not to the apple but to the nut is the Jumbo grown by the late Edward B Smith, Crescent Estate, Trinidad, who has supplied air layers to people from South America wishing to establish this large-seeded type in their countries.

In evaluating cashew trees, preference should be given to a cashew tree of slender, compact form because the cashew needs light to flower and fruit and bears only on its outer circumference. The interior of a broad-spreading tree is devoid of foliage and fruit. Likewise, if cashew trees are planted too close together, the branches that overlap or shade each other will be nonbearing and may die back. From an experimental plantation of compact trees, the grower should select those that produce a high yield of large nuts and multiply them vegetatively, rather than increasing labor costs by simply planting more and more seedling trees with yields varying from 5 to 100 kg per tree. There is variation also in the color of the young leaves of cashew trees. On some, the new shoots at the branch tips are purplish-red, while on others they are more or less yellow. The yellow leaves are heavily attacked by leaf-eating insects while the red, being richer in phenolic compounds, are naturally insect-resistant.

Diseases and Pests

In humid atmospheres and in seasons of excessive rainfall, cashew trees are subject to attack by the fungus *Colletotrichum gloeosporioides*, which causes anthracnose, a disease affecting the foliage, especially young shoots, the twigs, flowers, the apple, and the developing nut. A parasitic fungus, *Oidium anarcardii*, appears on the leaves, shoots, and flowers when there are periods of alternating sun and rain. They become coated with a powdery white substance and this is followed by blackening, shrivelling, and shedding. The nut may be ruined by the fungus, *Nematospora* spp., which invades it through minute perforations caused by sucking insects. A large wasp sucks the juice from immature apples, causing them to shrivel and blacken while the nut is still green. Yellow, red-banded thrips (*Selenothrips rubrocinctus* Giard) often infest the foliage. Cashew apples are commonly eaten by birds, bats, monkeys, and squirrels.

Harvesting and Drying

Harvesting of cashew nuts is extremely labor-intensive. The entire crop does not ripen at one time, and harvesting may extend over a period of 40–75 days. Ideally the nuts should be allowed to fall (with the apple attached) when perfectly mature. Some people unwisely shake the branches to bring them down or knock them off with a pole. This results in slightly immature nuts of high moisture content being mixed with partly dried mature nuts

and makes it impossible to assure uniform drying, thereby adversely affecting all subsequent steps in processing. In Brazil, cashews are plucked from the tree only if perfect apples are desired for dessert use. If allowed to fall, the apple will remain in good condition on the ground for 2–3 days. If it is not to be utilized, it is twisted off and left on the ground for cattle or pigs to consume. If it is wanted for preserving, it is salvaged and transported to a packing plant. The nuts are conveyed separately to be spread out in the sun and dried, with constant raking, and then stored if they are not to be immediately shelled. Thoroughly dried nuts can be stored for 1 or 2 years.

Processing

In Latin America and the West Indies there have been crude precracking practices such as boiling the nuts in steel barrels, simply drying them in the sun for 2 or 3 days, or sweating them in cement bins for several months. There was much loss because many nuts were found to be infested with insects or fungi on opening. In some cases the nuts were cut open by machete wielders with their hands wrapped in layers of plastic. The half-kernels were then picked out and sent to local clubs or bars, but these nuts were too contaminated with the cashew nut shell liquid (CNSL) to be fit for consumption (Figure 3).

Figure 3 Cut open with a machete, these cashew nut halves show the caustic, glistening cashew nut shell liquid (CNSL) which is usually expelled by roasting before the nuts are opened and the kernel is extracted, peeled, and reroasted before consumption. Reproduced from Cashew Nuts and Cashew Apples, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Sanitary processing of cashews in the past was best performed in India by roasting the nuts in shallow pans over open fires to expel the caustic CNSL, some of which was caught in jars below for various local uses. Then they rubbed the nuts with sand or ashes before handshelling, at which the Indian female workers were especially adept. But the fumes from the roasting were extremely irritating to eyes, nose, throat, and skin, and the shellers frequently had to dip their hands in lime, ash, or linseed or castor oil to minimize skin inflammation. A merciful improvement was a change to roasting in rotating cylinders over furnaces with chimneys to carry off the toxic fumes. The next advance, brought about by Mortimer Harvey, of New Jersey, who had investigated the chemistry and potential industrial uses of the shell oil, was a method of heating the nuts in baths of the oil kept at 188–194 °C. This method resulted in greater recovery of higher-quality oil.

Semimechanized shelling was initiated in Brazil in 1946. Dried raw nuts were first steamed to moisten the shells and then fed into a device, operated by hand and footpedal simultaneously, which cut the shell into two halves, leaving the kernel whole or cut in half. Facing the cutter, a seated operator pries the kernel loose from the cut shells if it adheres. This system is still used in Ceará for shelling very large nuts.

When East African countries achieved independence from colonial governments, they wanted to cease exporting their raw cashews to India in order to gain more revenue by processing the nuts themselves. A simple, raw-cashew opening-machine was invented by a mining engineer in Tanzania in 1960. The operator protected his hands with barrier cream. There followed the development of complete factory installations of cashew-shelling equipment in Italy, France, and Germany. Mozambique obtained its first full-scale factory in 1964 and 2 years later set a world record in cashew production; most of the crop went to India for shelling, while the rest (26 000 t) of the locally shelled kernels was exported to the USA, West Germany, and South Africa.

The various systems of cracking by pressure, centrifugal cracking, and cutting or sawing were evaluated by the Tropical Products Institute of London (now the Natural Products Institute), and the Sturtevants Engineering Company began producing and exporting the centrifugal shelling equipment in 1970. This system is now used by Brazil Oiticica, and modified versions are used in north-eastern Brazil. An Italian factory installation of a series of machines, designed by the Instituto Agronomico per l'Oltremare, involves 15 steps. The dried nuts are cleaned, roughly sorted according to size, wetted for several

days to the desired moisture content, roasted in CNSL, centrifuged to eliminate any oil residue, cooled, and mechanically calibrated to separate eight sizes. The nuts are then machine-cut, and the shell separated from the kernel and removed by pneumatic tubes for use as a fuel. The kernels are dried for 48 h in the sun or in ovens until the testa wrinkles and is removed mechanically. The kernels are rehumidified to resist breaking; they are then sorted into wholes and pieces. Lastly, kernels that still have clinging testa must be hand-peeled. Japanese processors freeze the nuts instead of roasting them before shelling.

However, none of these or other factory systems have given ideal results, mainly because of the variability of the raw material. There is a great need for a portable, simple cashew sheller that can take dehydrated, raw nuts of any size, splitting them without contamination of the kernel. The shells would be sent to facilities for total extraction of the oil for industrial purposes, and the split kernels of Jumbo cashews should not be inferior to whole kernels from smaller nuts. Such shelling has been achieved experimentally with explosive decompression, but further refinement of the method is needed.

The latest development in India is a cashew kernel drier, electronically controlled, with timer and alarm to avoid scorching, and adjustable dampers for air inflow; the drier turns out higher-quality kernels in 60–90 min.

Economic Status

In the past, the cashew ranked as India's second dollar earner. After mechanization, India found that the total crop of its cashew plantations could not keep its big factories busy as the productivity of the trees declined after they reached 35 years of age. In Kerala, the leading growing area, cashew orchards have been cut down and replaced by more profitable rubber plantations. From 1980 to 1987, the World Bank assisted India in a cashew development project to improve old plantations and establish new ones. In 1990, adverse weather conditions in India and other major cashew areas seriously affected the supply of cashews for export. Indian production was reduced by 20–25%, causing a sharp increase in prices. Nevertheless, the demand for cashew kernels has increased, especially with the opening of markets in eastern Europe. In 1989, world export of cashew kernels from India, Brazil, Mozambique, Kenya, Tanzania, and other areas amounted to 90 000 t. If horticultural improvement could be brought up to the level of cashew engineering, this figure could be doubled or tripled, and cashews would be less expensive and more widely enjoyed.

Inspection, Grading, and Packing

Before export, cashew kernels go through rigid inspection systems, are classified according to grade – wholes, butts, splits, large and small pieces and baby bits, also white, scorched, and dessert – and are carefully vacuum-packed in 11.3-kg capacity metal containers, two to a crate. They are then held in cold storage at temperatures below 9 °C. For retail sale, cashew kernels are usually reroasted in olive, peanut, or other oil, an adhesive and salt are applied and the kernels are packed in glass jars or cans. In recent years, many have been dry-roasted without salt. High grades have a minimum of split or broken kernels. Lower grades are mainly utilized by the bakery or confectionery industries. They are familiar ingredients in cookies, chocolate bars, and other products, and are made into cashew butter.

Cashew Apple

Because of the difficulty of cracking cashew nuts, many Latin Americans have habitually discarded the nut and consumed the cashew apple or, in Spanish, manzana del marañon, large heaps of which are seen on native markets at the height of the season. As it is fibrous, the apple is massaged and the juice is squeezed into the mouth as a thirst-quencher, especially out in the countryside. Inferior types that are still astringent when ripe are eaten with salt so that they will not irritate the throat. In the kitchen, cashew apples may be sliced and stewed or used to make juice, fruit soup, jam, jelly, paste, or chutney. The best cashew apples are preserved whole in syrup in glass jars. The less perfect are candied – cooked thoroughly in heavy syrup until well wrinkled, and then dried and sometimes rolled in granulated sugar. This is an excellent product resembling candied figs but it is, of course, seedless.

In India, the Central Food Technological Research Institute at Mysore has found that the astringency and acridity of cashew apple juice, due to 35% tannin content and 3% of an oily substance, can be eliminated by pressure steaming for 5–15 min and thorough washing. Then the juice is mechanically extracted. In the Philippines, the Bureau of Plant Industry, Manila, designed a cashew-apple-crushing machine for this purpose. To free the juice of some other undesirable elements, casein, gelatin, pectin, or lime juice may be added before straining or centrifuging. After the addition of sugar and citric acid to arrive at 15° Brix and 4% acidity, the juice is boiled for 1 min and then bottled or canned, alone or with other fruit juices.

In Brazil, it is estimated that less than 10% of the total cashew apple crop is processed, though the juice

Table 1 Food value of cashew nuts, oil-roasted and dry-roasted

Nutrients and units[a]	*Mean amount per kg, edible portion*	
	Oil-roasted	*Dry-roasted*
Proximate		
Water (g)	39.1	17.0
Food energy (kcal)	5760	5740
(kJ)	24 090	24 020
Protein (N × 5.30) (g)	161.5	153.1
Total lipid (fat) (g)	482.1	463.5
Carbohydrate, total (g)	285.2	326.9
Fiber (g)	12.7	7.0
Ash (g)	32.1	39.5
Minerals (mg)		
Calcium	410	450
Iron	41.0	60.0
Magnesium	2550	2600
Phosphorus	4260	4900
Potassium	5300	5650
Sodium[b]	170	160
Zinc	47.5	56.0
Copper	21.7	22.2
Manganese	8.07	
Vitamins[c]		
Ascorbic acid (mg)	0.0	0.0
Thiamin (mg)	4.24	2.00
Riboflavin (mg)	1.75	2.00
Niacin (mg)	18.00	14.00
Pantothenic acid (mg)	11.90	12.17
Vitamin B_6 (mg)	2.50	2.56
Folacin (μg)	677.0	692.0
Vitamin B_{12} (μg)	0	0
Vitamin A (RE)	0	0
(IU)	0	0
Lipids (g)		
Fatty acids		
Saturated, total	95.26	91.57
4:0		
6:0		
8:0	1.37	1.32
10:0	1.37	1.32
12:0	8.16	7.84
14:0	3.61	3.47
16:0	45.26	43.51
18:0	30.91	29.72
Monounsaturated, total	284.15	273.17
16:1	3.31	3.18
18:1	278.86	268.08
20:1	1.44	1.39
22:1		
Lipids (g)		
Fatty acids		
Polyunsaturated, total	81.52	78.36
18:2	79.68	76.60
18:3	1.67	1.61
18:4		
20:4		
20:5		
22:5		
22:6	0	
Cholesterol		0
Phytosterols		1.58
Amino acids (g)		
Tryptophan	2.50	2.37
Threonine	6.25	5.92
Isoleucine	7.71	7.31
Leucine	13.56	12.85
Lysine	8.62	8.17
Methionine	2.89	2.74
Cystine	2.99	2.83
Phenylalanine	8.35	7.91
Tyrosine	5.18	4.91
Valine	10.97	10.40
Arginine	18.37	17.41
Histidine	4.20	3.99
Alanine	7.40	7.02
Aspartic acid	15.87	15.05
Glutamic acid	38.24	36.24
Glycine	8.47	8.03
Proline	7.28	6.90
Serine	8.96	8.49

[a]28 g = approximately 14 large, 18 medium, or 26 small kernels.
[b]Value based on data for product without added salt. Oil-roasted product with added salt contains 626 mg sodium per 100 g; dry-roasted contains 640 mg of sodium per 100 g.
[c]α-Tocopherol = 0.57 mg per 100 g of dry-roasted product.
From McCarthy MA and Matthews RH (1984) *Composition of Foods: Nut and Seed Products (Raw, Processed, Prepared)*. Agriculture Handbook no. 8–12. Washington: US Department of Agriculture.
One retinol equivalent (RE) = 6 μg β-carotene or 10 international units (IU) vitamin A activity from β-carotene.

Table 2 Food value of cashew apple, raw caju, *Anacardium occidentale* L.

Nutrients and units	*Mean amount per kg, edible portion*
Nutrients and minerals (g)	
Food energy (kcal)	460.0
Moisture	871.0
Proteins	8.0
Lipids	2.0
Glycides	116.0
Fiber	15.0
Ash	3.0
Calcium	0.04
Phosphorus	0.18
Iron	0.01
Vitamins (mg)	
Retinol	0.40
Vitamin B_1	0.30
Vitamin B_2	0.30
Niacin	4.0
Ascorbic acid	2190.0

From Anonymous (1981) *Tabelas de Composição de Alimentos, Estudo Nacional da Despesa Familiar – Endef*. Rio de Janeiro: Secretaria de Planejamento da Presidência da República Funacao Instituto Brasileiro de Geografia e Estatística.

is locally popular and, with suspended pulp or clear, it is pasteurized, preservatives are added, and it is bottled and exported. The unclarified juice may be concentrated and bottled as nectar, or made available as frozen concentrate. Cashew apple juice is also made into vinegar. In Cuba, Costa Rica, and the Philippines, and formerly in Brazil, it has been

fermented into wine. Cashew apple brandy is subject to government control in East Africa and Goa. In the late nineteenth century, in Mozambique, the cashew was denounced as a source of vice and ruin because of the highly intoxicating liquor distilled from the apple. Cultivation of the tree was banned for a while; later a tax was imposed on plantations as a possible deterrent.

Food Value of Cashew Nut and Cashew Apple

See Tables 1 and 2. Refer to individual nutrients.

Cashew Nut Shell Liquid

This toxic oil is highly heat- and friction-resistant and is valuable for many industrial uses. When it was first used in aircraft paint-stripping products, mechanics suffered skin reactions. Mortimer Harvey promptly patented processes of detoxification and its uses have multiplied. It has been standard material for automobile clutch facings and brake linings and in insulation for electrical tools. It is incorporated into marine paints, water-resistant plywood, resins utilized in laminating, varnishes, floor tiles, cold-setting cements, molding powders for plastics, and many other products. In addition, the testa, which is removed from cashew kernels, has a high level of condensed tannin (as much as 25%) and the extract has been used for tanning leather.

See also: **Brazil Nuts**; **Peanuts**; **Walnuts and Pecans**; **Walnuts and Pecans**

Further Reading

Agnoloni M and Guiliani F (1977) *Cashew Cultivation*. Florence, Italy: Instituto Agronomico per l'Oltremare, Ministry of Foreign Affairs.

Cashew Export Promotion Council of India (1990) World Bank assisted cashew development scheme. *Indian Cashew Journal* 10(3): 18–20.

Khan MM, Hegde M, Mallik B *et al*. (1987) Rejuvenation of old cashew trees by top-working. *Indian Cashew Journal* 17(3): 9–25.

Morton JF (1967) Marañon! Necesidad urgente de un programa para la seleccion de variedades. *La Hacienda* 62(9): 38–43.

Morton JF and Venning FD (1972) Avoid failures and losses in the cultivation of the cashew. *Economic Botany* 26(3): 245–254.

Ohler JG (1979) *Cashew*. Communication 71. Amsterdam: Department of Agricultural Research, Koninklijk Instituut voor de Tropen.

Santos Lima V de PM (eds) (1988) *A Cultura do Cajueiro no Nordeste do Brasil*. Fortaleza, Brazil: Banco do Nordeste do Brasil.

Wilson RJ (1975) *The Market for Cashew-Nut Kernels and Cashew-Nut Liquid*. London: Tropical Products Institute.

CASSAVA

Contents

The Nature of the Tuber

C C Wheatley, Nelson, New Zealand
G Chuzel and N Zakhia, CIRAD-AMIS, Montpellier, France

Introduction

The cassava plant is a highly efficient producer of carbohydrate, mainly in the form of starch. It is the fourth most important source of calories in the human diet in tropical regions of the world and is consumed in a wide variety of forms. This article will review the origin and current distribution of the crop, the anatomy of the edible root, and its chemical composition. Special emphasis will be placed on the starch component of the root and on the presence of cyanide-containing glucosides.

Cassava Production

Crop Distribution

Cassava originated in the Americas, although the exact center of origin is disputed: northern Brazil,

northern Colombia/Venezuela, Paraguay, and southern Mexico have been suggested. The crop was widely grown in the Amazon basin by 2000 BC and remains a traditional staple of many tribal groups. Cassava was taken by Spanish and Portuguese traders to Africa, India, and South-east Asia in the sixteenth and seventeenth centuries but did not become widespread in Africa until the late nineteenth century. It is currently cultivated in almost all tropical and subtropical countries, the major producers being Thailand, Indonesia, India, Brazil, Zaire, and Nigeria. The total world production in 1998 reached 162×10^6 t, of which the Americas, Africa, and Asia accounted for 17%, 56%, and 27%, respectively.

Cassava has found acceptance over a diverse range of agricultural and food systems due to its tolerance of poor soil and harsh climatic conditions. Yields in excess of 10 t ha^{-1} $year^{-1}$ per year are achieved with minimal inputs under conditions of acid soils (pH< 4), low rainfall (< 1000 mm $year^{-1}$) and nutrient deficiency. Cassava has thus gained a reputation as a food security and famine reserve crop and as such has become especially important to small farmers living in marginal areas of the tropics. In addition, the efficiency of carbohydrate production, owing to cassava's unique combination of C_3 and C_4 photosynthetic pathways, has promoted its agroindustrial use in the last 30 years, especially in Thailand, Indonesia, and southern Brazil, for starch extraction and as a component of dried animal feed.

This short-daylength, perennial crop can be grown from 30° N to 30° S latitude, and up to an altitude of 2000 m at the equator, where the mean annual temperature is at least 17 °C in the high-altitude tropics, or 20 °C in lowland areas. Established plants can withstand over 6 months of drought but cannot tolerate prolonged flooding or saline conditions. In the humid lowland tropics, roots can be harvested after 6–7 months' growth. This increases to 18 months at higher altitudes, and 2 years in the subtropics. Propagation is by woody stem pieces called stakes. When grown in high-fertility soils with good agronomic practices, up to 50 t ha^{-1} can be produced in 12 months. On a world-wide basis, the crops most frequently interplanted with cassava are maize, cowpea, sorghums, and millets.

Traditional varieties tend to be of restricted agroecological range, having been selected by small farmers over many years. Many countries have introduced new varieties over the last decade, especially Thailand, Indonesia, Colombia, and Nigeria. Varietal improvement has focused on yield, adaptation to environmental and biotic stresses, and, recently, root quality.

Morphology and Anatomy of the Storage Root

The cassava root is anatomically a true root, not a tuberous root. The root cannot serve for vegetative propagation. Root size and shape depend on the variety and environmental conditions. The variability in size within a variety is greater than that found in other root crops. Cassava roots are generally 15–100 cm long and 3–15 cm wide. They are cylindrical, conical, or oval, with a coffee, pink, or cream-colored peel that is covered by a thin brown bark. The parenchyma is generally white, cream, or yellow. Plants produce five to 10 roots weighing 0.5–2.5 kg each.

The root is composed of three distinct tissues: bark (periderm), peel, and parenchyma. The parenchyma is the edible portion of the fresh root and comprises approximately 85% of the total weight. The parenchyma consists of xylem vessels radially distributed in a matrix of starch-containing cells. A central fibrous vascular bundle becomes progressively larger as the roots mature. Other fibrous bundles may develop throughout the root. The peel layer comprises sclerenchyma, cortical parenchyma, and phloem and constitutes 12% of the root weight, with the periderm layer comprising another 2%.

Harvesting, Root Deterioration, and Storage

Harvesting

The cassava root has no fixed period of optimum maturity: the woody plant is perennial, and starch deposition will continue for many years in most ecosystems. Harvest time is therefore determined by a combination of factors relating to yield, quality (starch content, etc.), the farmer's need to clear the field, climatic factors, and market prices. Climate affects starch content, fresh root eating quality, and hence market acceptability. Roots left beyond their optimum time for the fresh market are often used for animal feed or industrial processing at a later date, giving the farmer great flexibility of harvest time.

Harvesting is largely manual, although the use of mechanical harvesters is expanding in Brazil and South-east Asia where large-scale production and processing are located. In most countries, cassava is planted in small plots, often on a sloping terrain unsuited to mechanical operations. Manual harvesting represents one of the major production costs, along with weeding, often accounting for 30% of total cost. The ease of harvesting depends on the variety, soil type, and moisture content. The aerial part of the plant is frequently removed before harvesting. At harvest, roots destined for the fresh market are

sorted, based on size, into commercial and noncommercial classes. For industrial processing, all roots are collected.

Deterioration

Freshly harvested cassava roots have the shortest postharvest life of any of the major staple food crops. Roots become inedible within 24–72 h after harvest due to a rapid physiological deterioration process. This deterioration is a major constraint for industrial processing of the fresh roots and for marketing them to distant urban centers. Deterioration is due to the rapid, *de novo*, postharvest synthesis of simple phenolic compounds (catechins, coumarins, leucoanthocyanins) that polymerize to form blue, brown, and black pigments (condensed tannins). The accumulation of the coumarin, scopoletin, is especially rapid, reaching 80 mg hg^{-1} (dry weight) in 24 h. Scopoletin has intense blue fluorescence in UV light and can be confused with aflatoxins B_1 and B_2 which also have blue fluorescence and similar R_f values under some chromatographic systems: many reports of aflatoxins in fresh and processed cassava should therefore be treated with caution. (*See* **Mycotoxins**: Occurrence and Determination; **Phenolic Compounds**.)

Rapid physiological deterioration is an oxidative process. Tissue dehydration, especially at sites of root mechanical damage, encourages the rapid onset of this phenomenon. Care at harvest to reduce mechanical damage is beneficial. Tissue discoloration is initiated at damage sites and rapidly spreads through the entire root, starting in the vascular system, but later spreading to parenchymal cells. Within 3 days after harvest, a ring of discolored tissues arranged around the outer portion of the parenchyma (Figure 1) typically appears. In advanced cases, dehydrated, brown–white parenchyma tissues result. (*See* **Oxidation of Food Components**.)

A secondary deterioration can follow physiological or primary deterioration 5–7 days after harvest. This is due to microbial infection of mechanically damaged tissues, and results in the same tissue discoloration with vascular streaks spreading from the infected tissues. Generalized rotting and fermentation of tissues follow.

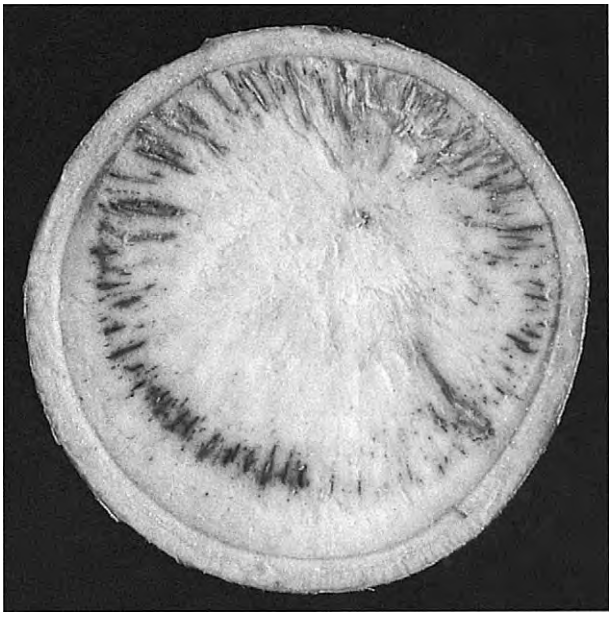

Figure 1 Transverse section through cassava root showing symptoms of physiological deterioration, 3 days after harvest. Reproduced from Cassava: The Nature of the Tuber, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Storage

Conditions of high temperature (30 °C) and high relative humidity (85%) favor the rapid wound healing of mechanically damaged tissues (curing). Under these conditions, the physiological processes described above are localized in the vicinity of the wound itself. The cured tissues represent a barrier to further oxygen entry, thus preventing the oxidative reactions that lead to the formation of condensed tannins. Encouragement of this curing process is the basis for several simple storage methods for fresh cassava.

Traditionally, cassava roots are preserved in the fresh state by reburying them in soil or moist sand. Field clamps, using straw and roots in layers below a soil-covered mound, with adequate ventilation, were developed in the 1970s. Packing roots in boxes with moist sawdust, coconut fiber or other locally available materials has also been successful. Packing roots in perforated polyethylene bags (with optional use of a thiabendazole-based fungicide to protect against microbial deterioration) is effective and has undergone commercial trials in South America and West Africa. With all these storage methods, root quality is maintained for 2–3 weeks. Beyond this time, starch breakdown to free sugars gives roots an unacceptable sweet taste. (*See* **Fungicides**; **Storage Stability**: Mechanisms of Degradation.)

For export markets, paraffin-coating of fresh roots provides an artificial barrier to oxygen entry and hence to deterioration. Costa Rica and the Dominican Republic export paraffin-coated roots to the USA and Europe. Peeled, frozen, or vacuum-packed cassava pieces are also marketed widely in Latin America and exported. No economically viable method for fresh root storage at the plant level prior to industrial processing is available. Fresh roots

should be processed within 2 days of harvest to ensure an acceptable quality and yield: starch extracted from deteriorated roots is light brown in color and of a poor functional quality, and the starch yield decreases by 10% for each day beyond the initial 2 days of storage. Excellent links between processing plant and cassava producers are essential to avoid deliveries of roots of unacceptable quality. The lack of a large-scale storage method for fresh cassava limits processing-plant operations to cassava harvest periods. However, in many areas, cassava can be harvested all year round, especially if irrigation is available. If necessary, fresh roots can be chipped and sun-dried: the dried chips can be used for starch extraction at a later date, although starch yield and quality are lower.

Chemical Composition of the Cassava Root

The chemical composition of cassava roots is shown in Tables 1 and 2. The composition of peel and parenchyma is different: the peel has more protein, fiber, sugars, and cyanogens than the parenchyma, and less dry matter and starch. A wide range of values is reported in the literature for each root component. Reports of whole-root dry-matter contents vary from 20% to 40%, and between 70% and 90% of dry matter is composed of starch, with average values between 80% and 85%. Total carbohydrates, including starch, free sugars, and other cellulose or hemicellulose components, together make up over 90% of parenchyma dry weight. The protein content is uniformly low (below 2% on a fresh-weight basis), as are fats and ash content. The fiber contents are more variable, and increase with plant age. A great variation in total cyanogen content is also found, with parenchyma values of 30–100 mg kg^{-1} common for low-cyanogenic cultivars destined for direct consumption, compared with 1350 mg kg^{-1} in industrial varieties used for processing. (*See* **Carbohydrates**: Classification and Properties; **Starch**: Sources and Processing.)

Table 1 Constituents of cassava root parenchyma and peel

Constituent	*Percentage of dry weight*	
	Parenchyma	*Peel*
Dry matter (%, fresh weight)	23–44	15–34
Starch	70–91	44–59
Total sugars	1.3–5.3	5.2– 7.1
Crude fiber	3.0–5.0	5.0–15.0
Ash	1.0–2.5	2.8–4.2
Protein	1.0–6.0	7.0–14.0
Fat	0.3–1.5	1.5–2.8

Table 2 Vitamin, mineral, and cyanide constituents of cassava root parenchyma and peel

Constituent	*Milligrams per kilogram dry weight*	
	Parenchyma	*Peel*
Total cyanide	30–1350	60–50
Calcium	480–920	
Phosphorus	770–150	
Potassium	6000–10 000	
Iron	5–25	
Vitamin A	0–70	
Vitamin C	380–900	

A variation in chemical composition is found between cassava varieties, and also within varieties at different plant ages, and is due to changing environmental and biotic stress factors. For example, parenchyma dry-matter content may decrease from 35% to 28% after a period of drought followed by the onset of rainfall. Under these conditions, starch is rapidly hydrolyzed to free sugars, especially sucrose, which are used by the plant to initiate regrowth of foliage. These large fluctuations in root quality over relatively short periods must be taken into account in planning harvesting times, both for direct consumption and for processing. (*See* **Starch**: Modified Starches.)

Cyanide in Cassava Roots

All known *Manihot* species contain cyanogens. In cassava, cyanide is synthesized in the leaf and transported to the roots, where it is partitioned between peel and parenchyma. Some 85% of the cyanide occurs as cyanoglucosides, mainly linamarin but also lotaustralin. Linamarin is broken down by the enzyme linamarase, also found in cassava tissues. In intact roots, the compartmentalization of linamarase in cell walls and linamarin in cell vacuoles prevents the formation of free cyanide. On processing, the disruption of tissues ensures that the enzyme comes into contact with its substrate, resulting in rapid production of free cyanide via an unstable cyanhydrin intermediary.

Cassava varieties differ in their total cyanogen content, and although the variation in content is continuous, two distinct types can be discerned in many regions of the world: low- and high-cyanogen varieties. Low-cyanogen varieties are often called sweet, although confusion with elevated free sugar contents in some conditions may occur. These are grown for fresh consumption, or for simple processing. High-cyanogen varieties are called bitter, although the phenolic compounds involved in root physiological deterioration can also produce a bitter taste. These are invariably grown for processing, especially for farinha in Brazil and gari in Africa. It is interesting that low-cyanogen varieties are rarely

used for processing: the reasons for a preference for high-cyanogen varieties are unclear. Contrary to anecdotal evidence, high-cyanogen varieties do not yield more than low-cyanogenic varieties. There is some evidence that cyanide may protect against rodent pests and robbery, and also that there may be a link between end-product quality (e.g., farinha texture) and cyanogen content.

Cyanide is concentrated in the root peel, especially in low-cyanogen varieties where peel:root ratios of 40:1 are found. In high-cyanogen varieties, ratios of 1.6:1 are common. The use of whole or peeled roots for processing may therefore result in great differences in end-product cyanogen content. In root parenchyma tissues, free cyanide – hydrogen cyanide (HCN) plus cyanhydrin – normally accounts for 15% of total cyanide; the remainder is bound as linamarin or lotaustralin. All cassava varieties characterized to date contain these cyanogens. Projects involving genetic manipulation are currently under way to reduce the toxicity problems of cassava, through regulation of cyanide synthesis and increasing the activity of linamarase.

Starch

Cassava starch has many food and industrial uses, which are linked to its functional properties. Although the basic properties of this starch are known, much research is required to complete our knowledge, especially as regards varietal differences in composition and functional properties. Cassava starch granules are round with a truncated end and a well-defined hilum. The granule size is between 5 and 35 μm. The starch has an A-type X-ray diffraction pattern, usually characteristic of cereals, and not the B type found in other root and tuber starches. The C-type spectrum, intermediate between A and B types, has also been reported. The nonglucosidic fraction of cassava starch is very low: the protein and lipid content is below 0.2%. There is thus no formation of an amylose complex with lipids in native starch. Amylose contents of 8–28% have been reported, but most values lie within the range of 16–18%. (*See* **Starch**: Structure, Properties, and Determination.)

Figure 2 shows Brabender amylographs of starch from four cassava varieties. Although a significant varietal variation is present, all curves follow the same general pattern and are similar to all high-amylopectin starches. The starch gelatinizes at relatively low temperatures. Initial and final gelatinization occurs at 60 °C and 80 °C, respectively. The peak viscosity is high and is reached rapidly, within 2 min. At high temperatures starch viscosity decreases greatly. Viscosity increases again on cooling (setback), and then remains stable. Cassava starch has a low retrogradation tendency (i.e., viscosity is stable over time following the cooling episode) and produces a very stable and clear gel. The swelling power of the starch is also very high: 100 g of dry starch will absorb 120 g of water at 100 °C. At this temperature, over 50% of the starch is soluble. (*See* **Starch**: Functional Properties.)

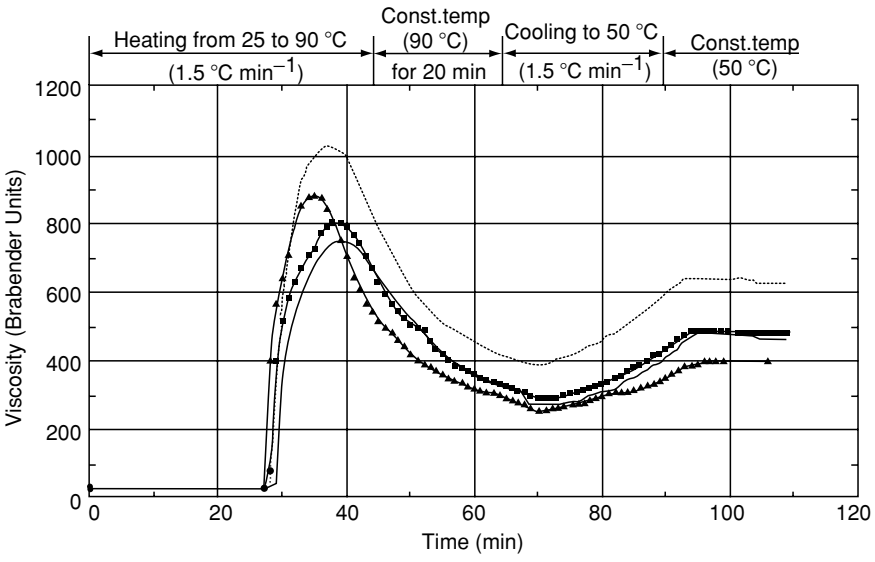

Figure 2 Brabender viscoamylograms of starch from four cassava varieties (4% w/v paste). —— CMC-40; —■— CM 523-7; . . . . . MCol 8; —▲— MCol 1684. Source: CIAT (International Centre for Tropical Agriculture), Cali, Columbia. Reproduced from Cassava: The Nature of the Tuber, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

See also: **Carbohydrates**: Classification and Properties; **Fungicides**; **Mycotoxins**: Occurrence and Determination; **Oxidation of Food Components**; **Phenolic Compounds**; **Starch**: Structure, Properties, and Determination; Sources and Processing; Functional Properties; Modified Starches; **Storage Stability**: Mechanisms of Degradation

Further Reading

Bradbury JH and Holloway WD (1988) Chemistry of tropical roots crops: significance for nutrition and agriculture in the Pacific. *Australian Center for International Agricultural Research Monograph No. 6*.

Cock JH (1985) *Cassava: New Potential for a Neglected Crop*. Boulder, CO: Westview Press.

Cooke RD and Coursey DG (1981) Cassava: a major cyanide containing food crop. In: Conn E, Knowles EJ, Vennesland B, Westley E and Wissung F (eds) *Cyanide in Biology*, p. 193. New York: Academic Press.

Kay DE (revised by Gooding EGB) (1987) *Crop and Product Digest No. 2 – Root Crops*, 2nd edn. London: Natural Resources Institute XV.

Richard JE and Coursey DG (1981) Cassava storage. Part 1: storage of fresh cassava roots. *Tropical Science* 23: 1–32.

Wheatley CC and Best R (1991) How can traditional forms of nutrition be maintained in urban centers: the case of cassava. *Entwicklung & Landlicher Raum* 1: 13–16.

Uses as a Raw Material

C C Wheatley, Nelson, New Zealand
G Chuzel and N Zakhia, CIRAD-AMIS, Montpellier, France

Background

The fresh cassava root is an important staple in many rural areas of the tropics, especially in sub-Saharan Africa, where it is the principal carbohydrate source for 40% of the population. More recently, the crop has been used as a raw material for food and other industries. This article briefly reviews the wide range of products made from cassava roots and then focuses on four of the most important products.

Use as a Staple Food

The fresh, peeled root is commonly boiled for 15–30 min, either alone or with other ingredients in a soup. Raw or boiled root pieces can also be fried. Boiled cassava should have a soft, floury texture, but is sometimes hard and glassy, leading to rejection by consumers. The causes of glassy texture are unknown. The bitter taste is frequently due to residual cyanide remaining after boiling, although the phenolic compounds involved in physiological deterioration can also produce a bitter taste. Cultivars with a low cyanogen content in the parenchyma are invariably used for fresh consumption. Consumption per capita of fresh cassava may reach very high levels (e.g., 405 kg per year in Zaire, and over 200 kg per year in rural areas of Paraguay). A variety of foods made with grated or mashed, boiled cassava are prepared at the household level. Examples of other staple foods based on cassava are gari in Nigeria and farinha in Brazil. These processed products will be considered in the next section. (*See* **Phenolic Compounds**.)

Cassava leaves form an important part of the diet in Zaire and other African countries, and are of minor importance in South-east Asia. In Zaire, cassava leaves comprise nearly 70% by weight of all vegetables consumed. They are normally chopped and boiled, a process that greatly reduces the level of residual cyanogens.

Cyanide Toxicity and Processing

Tissue disruption (chipping, grating or fermentation) is effective in bringing linamarase into contact with the linamarin substrate, thereby liberating hydrogen cyanide (HCN). Processes such as sun-drying of large root pieces or boiling are less effective. Toxicity problems may result when poorly processed cassava products are consumed. Acute cyanide toxicity appears when more than 30 mg is consumed over a 24 h period. The body's natural thiocyanate detoxification mechanism requires a good supply of sulfur-containing amino acids: a protein-deficient diet in combination with cyanide consumption can therefore result in health problems. In addition, thiocyanate interferes with iodine uptake by the thyroid gland, leading to goiter and, in extreme cases, cretinism. Epidemic spastic paraparesis has been reported on several occasions in Africa since 1981, affecting a total of 5000–10 000 people with leg paralysis. The exact metabolic processes involved are unknown, but each outbreak is clearly associated with famine-induced consumption of poorly processed cassava and a lack of protein in the diet. (*See* **Iodine**: Iodine-deficiency Disorders.)

The toxicity of processed cassava products is due to the residual linamarin and cyanhydrin remaining after the action of linamarase ceases. Once consumed, linamarin reaching the gut may be broken down by microbial enzymes or by ingested linamarase. This process is incomplete, since linamarin can also be detected in the urine. The percentage of linamarin

that may be degraded in the gut has received little research attention; however, individual sensitivity to cyanide toxicity seems to vary with nutritional status. Adequate cassava processing (grating, fermentation, slow sundrying), combined with sufficient dietary protein, is necessary for populations at risk in order to take advantage of the food security offered by cultivating high-cyanide cassava varieties. The use of cassava varieties low in cyanogens also has potential.

Overview of Cassava Processing

Processing of cassava has several objectives: to produce a more stable product, capable of being stored in tropical environments for extended periods with no decrease in quality or deterioration; to reduce the content of cyanide to innocuous levels; to reduce bulk and hence lower transport costs; to diversify the food and other uses of cassava, leading to market expansion; and to provide food and other industries with a low-cost raw material for further processing.

Cassava roots may be processed whole, after removal of the peel, or after washing, a process that also removes most of the outer barky layer. The roots can then be grated, chipped, or sliced. Grated roots may be pressed to remove excess water and then toasted to produce the flour called farinha, or a flat bread called cassabe. These are staple foods in several Latin American and Caribbean countries. If a fermentation stage is included, using peeled roots, the product is gari, a staple in West Africa. Alternatively, starch can be extracted from the grated roots, and can be either dried directly or fermented to produce a naturally modified starch (sour starch). Cassava chips are usually sun-dried to produce a dried chip (gaplek), which is used for human consumption in Indonesia, and for animal consumption in Latin America. Larger root pieces are sun-dried in Africa. (*See* **Starch**: Sources and Processing.)

A high-quality cassava flour can be produced from washed or peeled roots, chipped and sun-dried on trays, or artificially dried to avoid microbial contamination. This flour can replace wheat and other imported flours in tropical countries. In Africa, pieces of cassava roots are fermented and dried to produce a flour called lafun. In Indonesia, root pieces are fermented using an innoculum of *Amylomyces* and *Endomycopsis*, to produce a moist product called tape. In Africa, whole roots are retted, grated, and boiled to produce a fermented dough called chickwangwe or fu-fu.

A great diversity of products are prepared from cassava roots in rustic, small-scale facilities, using traditional methods. Over recent decades, medium-scale processing of these traditional products has developed on all continents. This had led to the design of more efficient processing equipment, and to improvements and standardization of quality. In recent years, large-scale operations for starch and farinha have been developed to supply urban and export markets, especially in Brazil, Thailand, and Indonesia. (*See* **Traditional Food Technology**.)

Important Products

Gari

Gari, a fermented, cooked and dehydrated cassava meal, is widely consumed in West Africa. It is a stable, ready-to-use foodstuff, well-suited to urban markets. Gari consumption is increasing, even in countries where it is not a traditional food. Gari is principally consumed in the main meal with vegetable sauces and meat. It can also be eaten as a snack soaked in cold water or milk, with roasted peanuts or coconut. Gari contributes up to 60% of the total calorie intake in West Africa, where an average of 150 g per person per day is consumed. Gari is mainly produced on a small scale and marketed by women; it provides an important source of income in many rural areas. Although many process improvements have been designed, diffusion of equipment has been patchy, especially for the roasting stage, so today, most of the gari consumed is still manufactured using the traditional process.

In traditional preparation of gari (**Figure 1**), the roots are hand-peeled, then grated into a pulp, often using a roughly perforated iron sheet. This pulp or mash is placed in hessian sacks, under heavy weights (logs, rocks) to squeeze out excess water. The mash is left to drain for 1–6 days, during which natural fermentation occurs. The pressed cake is shredded by hand and then rubbed through a woven palm-frond sieve to remove fiber and large lumps. This sieved mash is toasted (garified) in clay or iron vessels over a wood fire until the starch gelatinizes, and the moisture content is reduced to 10–15%. The mash must be stirred constantly to avoid sticking or burning. Palm oil may be added to facilitate this operation and imparts a yellow color to the final product.

Great variability exists in consumer perception of gari quality, and it is difficult to establish optimum values for the different quality parameters. However, in general, consumers prefer a crisp, fine-grained, slightly sour product with good swelling power in water, and a lightly toasted color. Although all the processing steps are important in determining end-product quality, some of them are only mechanical (peeling, grating, sieving, dewatering), whereas others involve some physical and chemical modifications that

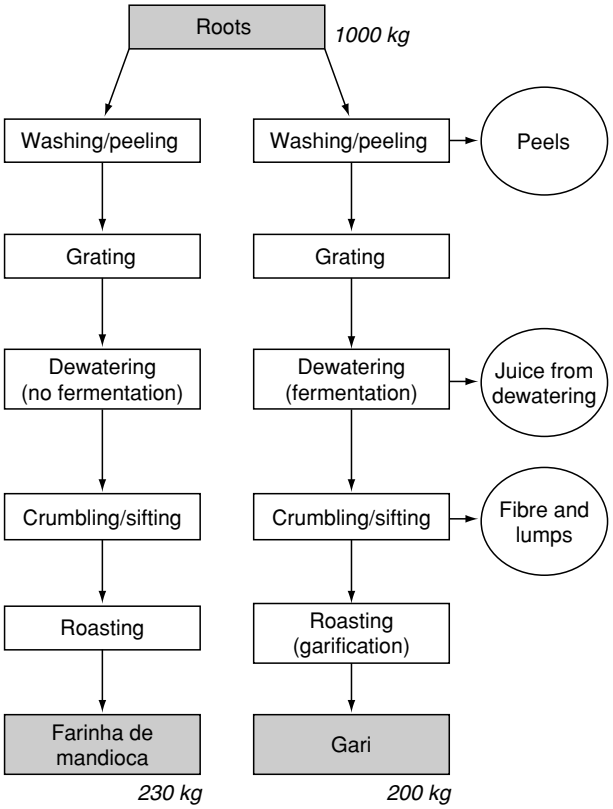

Figure 1 Flow diagram of the traditional processes for production of gari (West Africa) and farinha de mandioca (Brazil). Reproduced from Cassava: Uses as a Raw Material, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

improve the digestibility and reduce the cyanide toxicity (fermentation, garification). The fermentation step is crucial to the development of the characteristic aroma and sour flavor of gari, imparted mainly by lactic acid bacteria, which produce lactic acid and volatiles such as aldehydes, diacetyl esters, and ethanol. Fermentation also helps to detoxify the mash, whereas grating allows linamarase to come into contact with cyanoglucoside substrates. During fermentation, free HCN is released. The pH falls from 6.8 to 4.5 after 24 h of fermentation. Because the cyanohydrin intermediary is quite stable in acidic conditions, and linamarase activity is also reduced, less free cyanide is produced, and cyanhydrins accumulate in the fermented mash. Cyanide detoxification of the mash occurs in three ways: enzymic hydrolysis of cyanoglucosides, solubilization of cyanide in the water lost during dewatering, and volatilization of free cyanide during garification, at temperatures over 26 °C. The combination of these is not enough to completely remove all cyanide from the end product: total cyanogen contents of 15–33 mg kg^{-1}, with a free cyanide content of 5–25 mg kg^{-1}, have been reported. (*See* **Lactic Acid Bacteria**.)

The garification step is essential to obtain good-quality gari, for which 60–80% of starch should be gelatinized. This also improves starch digestibility. The starch is not completely gelatinized during garification because of the low initial moisture content of the dewatered mash (50–55%) and the temperatures used (60–85 °C). Gelatinization proceeds until the moisture content is reduced to about 40%, after which only drying occurs. The efficiency of manual stirring during garification is an additional variable affecting the final quality.

Farinha

Farinha de mandioca is a major dietary staple in Brazil, especially in the northeast region of the country, where the per capita consumption is over 40 kg per year or 1881 J (450 cal) per person per day. This represents 25% of the total calories consumed – more than rice and wheat combined. In Brazil as a whole, farinha contributes an average of 752 J (180 cal) per person per day to the diet. Farinha is a traditional cassava product and is still the major source of dietary energy for many of the indigenous peoples of the Amazon basin; one study of the Colombian Vaupes

region found that cassava provided over 80% of dietary calories and that farinha was the major product consumed.

The process for the production of farinha is similar to that of gari (Figure 1). It is probable that gari is an adaptation of farinha to African tastes, which occurred when cassava was introduced from Brazil to Africa. Unlike gari, however, a typical farinha preparation does not include a fermentation step. The operation proceeds directly from grating through pressing and sieving to toasting. Another difference is the use of washed but unpeeled roots. The end product is thus a cream/yellow-colored meal with no fermented taste. The final moisture content is the same as in gari, but the residual cyanogen content is lower ($10\,mg\,kg^{-1}$). Since no fermentation occurs, acid conditions do not develop, and the cyanhydrin intermediate rapidly breaks down to free cyanide.

Farinha is an integral ingredient in many traditional dishes. Mixed with beans, farinha provides a thicker texture. Farinha is also consumed with meats. Consumption decreased markedly between 1960 and 1980, as government subsidies on wheat reduced the competitiveness of farinha. However, over the last few years, elimination of wheat subsidies has resulted in an increase in the attractiveness of farinha. As a storable, rapidly prepared, and well-accepted food, farinha integrates easily into urban food habits.

Considerable applied research has resulted in the design, testing, commercial construction and use of efficient medium- and large-scale farinha production facilities in Brazil. Plants with a capacity of 50–200 tonnes of cassava roots per day are in operation. Alongside this, many thousands of small-scale traditional 'casas de farinha' exist, especially in the northeast, providing employment over vast areas of this region, the poorest in Brazil. Over 460 000 such plants exist, often run as communal operations at the village level. The range of scales of operation and the regional differences in cassava varieties and the process itself result in large variations in end-product quality. The variation in quality and its effect on consumption patterns need to be studied further. Research aimed at improving and standardizing quality has yet to produce results.

Starch

Although cassava is an efficient starch producer, only 5% of total starch traded worldwide is from cassava, representing about $0.8–1.0 \times 10^6$ tonnes per year. The main industrial-scale producers of cassava starch are Thailand, Brazil, China, and Indonesia. Production of cassava starch has increased dramatically in Thailand: by 1994, 50% of the 19×10^6 tonnes of cassava roots produced were processed into starch. Exports, mainly to East Asia, totaled 1.15×10^6 tonnes, and domestic use comprised 0.9×10^6 tonnes. In Thailand, starch is developing as a feedstock for further processing: in 1994, 31% of starch was modified, 12% was used for sweeteners, and 12% was used for MSG (monosodium glutamate). In addition to large-scale industries, many small-scale starch extraction plants operate in South and South-east Asia and in Latin America. Here, cassava starch is used in the local food industry to make a wide range of traditional food products (tapioca or sago in India; krupuk in Indonesia; chipa in Paraguay; maltose in Vietnam). In Brazil and Colombia, moist starch is fermented before drying, and a naturally modified starch is obtained, which has functional properties distinct from those of native cassava starch, namely a spontaneous expansion capability during baking. This starch is used in several traditional cheese breads (pandebono, pandequeso, bizcocho) and, more recently, in novel snack foods.

Cassava roots are washed and peeled; sometimes, only the outer bark is removed and then grated to release starch granules (Figure 2). The starch is extracted under running water and separated from the fiber and other root components by screening. Solid starch is separated from the starch/water slurry by sedimentation or centrifugation. The resultant starch is dried to a final moisture content of 12–14%. Sun-drying in small-scale operations and flash-drying in large-scale plants are typical. To obtain sour, fermented starch, the sedimented starch is left in tanks for 20–30 days before drying. Whereas a small-scale extraction plant may process only 2–10 tonnes of cassava per day, large-scale operations in Brazil and Thailand can handle 200 tonnes per day. Starch extraction rates are usually 18–22% for small-scale plants and 20–25% in large operations, depending on the process efficiency and initial root starch content. In small-scale plants, operations are manual or partially mechanized, resulting in less efficient starch extraction: in Colombia, 45% of the starch contained in the roots is lost in the waste water from the process, or remains in the peel and fiber byproduct, which is dried and used for animal feed. Waste water from processing plants can present severe environmental contamination problems, owing to significant concentrations of starch and cyanide. More efficient use of water in the process, and simple water treatment systems can reduce pollution. Regulatory pressure, especially on large-scale factories, is increasing.

In India, Brazil, and Malaysia, partially dried native cassava starch is used to make tapioca or sago: the still moist starch is globulated on a vibrating

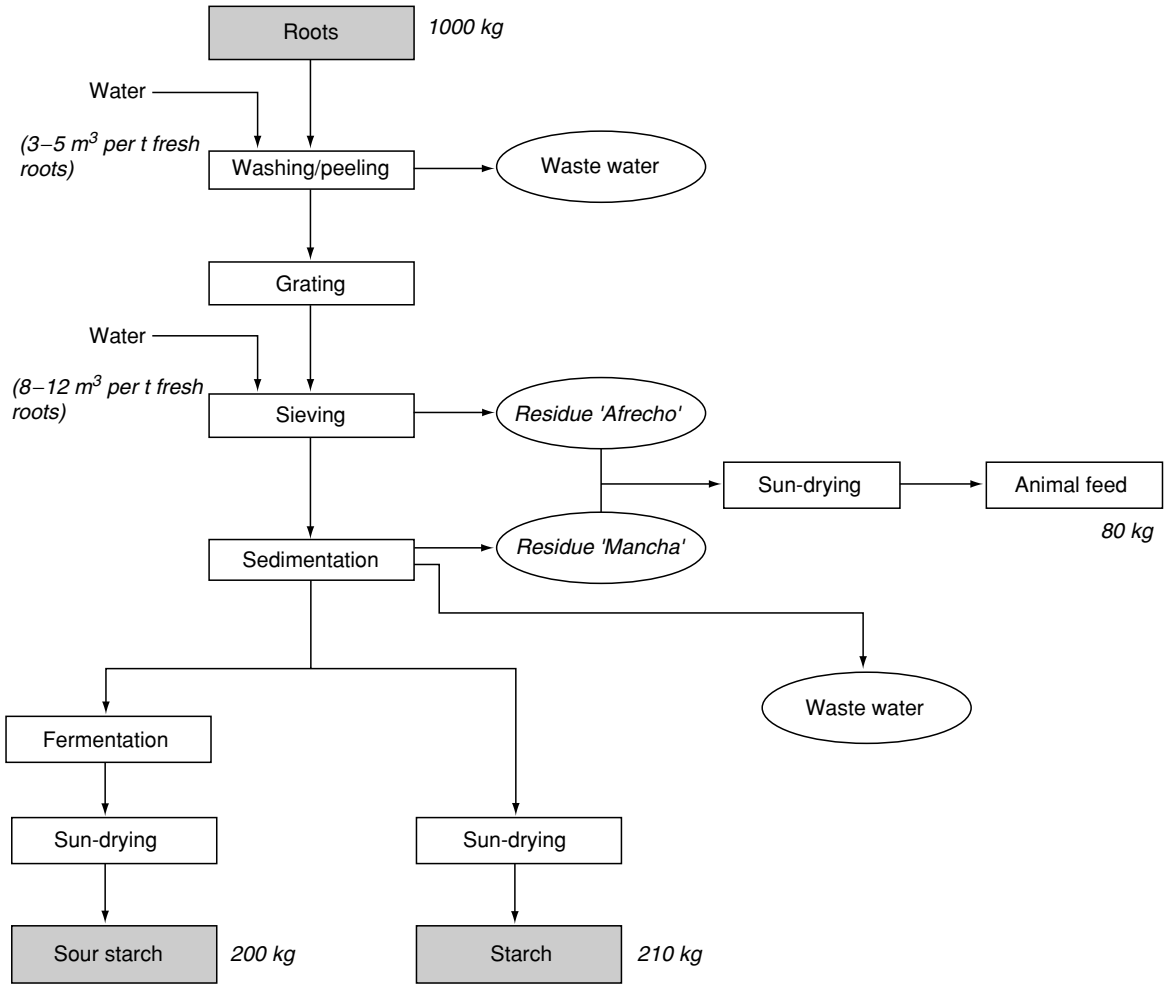

Figure 2 Flow diagram of the process for production of fermented (sour) and nonfermented (sweet, native) cassava starch. Reproduced from Cassava: Use as a Raw Material, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

surface, and the resulting small spheres are then steamed or baked to gelatinize the surface layers of starch. The end product, tapioca pearls, is used to make a variety of desserts. In Indonesia and Malaysia, the native starch is mixed with water, fish, or prawn flavoring and coloring. This dough is steamed, then sun-dried to make krupuk, or prawn crackers. When fried, these expand to several times their original size. They are a major snack food in Java and elsewhere in the region.

Fermented or sour starch has excellent expansion power; thus, the volume of a dough containing this starch increases greatly during baking. Although many factors affect the expansion power of sour starch, including environmental conditions, cassava variety and water quality, the major factors are fermentation and sun drying. During fermentation, amylolytic organisms produce reducing sugars, which are immediately consumed by lactic acid bacteria to produce organic acids (lactic, propionic, butyric, acetic, etc.) and carbon dioxide (CO_2). The high lactic acid content in sour starch correlates with a good expansion power and a good-quality final baked product. The action of the amylolytic enzymes and the acidification of the fermentation medium (pH 3.2) result in a lower viscosity, improved ease of baking, and reduced gelling ability of sour starch compared with native cassava starch. During baking of doughs containing sour starch, the moisture content loss (from 50% to 8–12%) together with volatilization of lactic acid and CO_2 results in an increased expansion power and reduced gelling ability, giving the end product its characteristic texture. A simple method for assessing the expansion power of sour starch is now available and in use with processors in Colombia.

Flour

In several regions of Africa and in Indonesia, cassava roots are chopped into pieces, sun-dried, and then pounded into a coarse meal or flour. This flour serves as a base for the production of many dishes at the household level. More recently, in Thailand, sun-dried cassava chips have been produced on a large scale, over 6×10^6 tonnes per year, for export to Europe as an animal feed. The chips are ground to a meal, which is then converted under steam and pressure to hard pellets. Dried cassava for animal feed expanded in Latin America during the 1980s as an alternative to imported grains for internal use. Globalization and cheap imports of feed grains reduced the feed use of cassava in the early to mid-1990s, but is now becoming increasingly competitive again. Potential also exists to produce high-quality cassava flour to replace imported wheat flour in bakery and other products in many tropical countries. One such program in Brazil was halted during the 1970s by wheat flour subsidies, which made it impossible for cassava flour to compete. The food industry use of root crop flours depends largely on international wheat grain prices: if these continue to decrease, the potential for substitution will be limited to those uses where cassava flour offers some specific functional advantages, or in remote areas where the transport costs for wheat flour are high (e.g., the Peruvian Amazon, where cassava flour is currently a commercial product).

The process for the production of high-quality cassava flour requires the use of cleaned roots with the bark removed. Peel removal is optional. Roots are chipped and sun-dried on trays or artificially dried (60 °C for 8–10 h) to obtain a hygienic product. The dried chips, with 12% moisture content, can be reduced in size (premilled) sufficiently to enable them to be used in standard wheat flour roller mills. Conversion rates in excess of 90% can be obtained in this way, compared with 72% for wheat grains. If roots are not peeled before chipping and drying, the peel and fiber fractions can be removed efficiently during the milling and grading process. A conversion rate of 3:1 is possible with cassava flour, compared with 4.5–5:1 for starch. The approximate composition of the flour is as follows: moisture 12%; carbohydrate 75%; protein 3%; fiber 5%; and lipids 2%. The total cyanogen content should be below 50 mg kg^{-1}.

Industry trials with cassava flour in Colombia have demonstrated that some specific attributes of the flour are especially suited to certain products. A 40–50% substitution of cassava for wheat flour produces biscuits with superior texture and color compared with 100% wheat flour controls. In Indonesia, a wide range of bakery products are being manufactured from cassava flour, at the village level as well as the food-industry level. (*See* **Flour**: Analysis of Wheat Flours; Dietary Importance; Roller Milling Operations.)

See also: **Flour**: Roller Milling Operations; Analysis of Wheat Flours; Dietary Importance; **Iodine**: Iodine-deficiency Disorders; **Lactic Acid Bacteria**; **Phenolic Compounds**; **Starch**: Sources and Processing; **Traditional Food Technology**

Further Reading

Alarcon F and Dufour D (1998) *Almidón Agrio de Yuca en Colombia. Tomo 1 Producción y Recommendaciones. CIAT/CIRAD, 35 P. Publicación CIAT No. 268, Cali, Colombia*.

Balagopalan C, Padmaja G, Nanda SK and Moorthy SN (1988) *Cassava in Food, Feed and Industry*. Boca Raton, FL: CRC Press.

Campbell-Plat G (1987) *Fermented Foods of the World*. London: Butterworth.

Cereda MP (1987) Tecnologia e qualidade do polviho azedo (Technology and quality of sour starch). *Informe Agropecuario Belo Horizonte* 13: 145.

Chuzel G, Cereda MP and Griffon D (1996) Alimento fermentes a base de manise. In: Bougeor CM and Iarfent JP (eds) *Microbiologie Alimentaire*, 2nd edn., pp. 249–273. Paris: Technique de Documentation.

Lancaster PA, Ingram JS, Lim MY and Coursey DG (1982) Traditional cassava based food: survey of processing techniques. *Economic Botany* 36: 12–45.

Muchnik J and Vinck D (1984) *La Transformation du Manioc: Technologies Autochtones*. Paris: Presses Universitaires de France.

Onyekwere OO, Akminrele IA, Koleojo OA and Heys G (1989) Industrialisation of gari fermentation. In: Steinkraus KH (ed.) *Industrialisation of Indigenous Fermented Foods*, pp. 363–408. New York: Marcel Dekker.

Rosling H (1988) *Cassava Toxicity and Food Security: A Report for UNICEF African Household Food Security Programme. International Child Health Unit, University Hospital, Uppsala, Sweden*.

Wheatley CC and Best R (1991) How can traditional forms of nutrition be maintained in urban centers: the case of cassava. *Entwicklung Landlicher Raun* 1: 13–16.

Zakhia N, Dufour D, Chuzel G and Griffon D (1996) Review of sour cassava starch production in rural Colombian areas. *Tropical Science* 36: 247–255.

CATERING

Contents
Catering Systems
Nutritional Implications

Catering Systems

A G Smith and A West, The University of Huddersfield, Huddersfield, UK

Background

Catering systems are complex sociotechnical organizations involving both people and machinery in the production and service of food. Such systems have the key purpose of transforming a diverse combination of 'inputs' into desired 'outputs.' Systems need to maximize their interdependence with the environment within which they exist. The objective for the management of any system is to find ways of ensuring its long-term survival and growth, often by seeking gains in efficiency and effectiveness in producing its outputs. Different types of catering systems have evolved in recent years as a result of efforts to achieve these goals. These developments have brought with them further technological challenges related to the provision of safe, nutritious, quality food products.

Emergence of Catering Systems

It has been noted that the 'systems' approach to management involves viewing organizations 'holistically' in order to gain a better insight into these complex situations, rather than merely focusing on individual parts or problems. Systems theory has been applied to a wide range of organizational contexts and has proven valuable in solving problems particularly where the situation is unpredictable and issues causing concern are ill defined or vague.

It has been suggested that 'systems' terminology was first applied to food production and service operations during the 1950s. The terms 'catering systems,' in common use in the UK, and 'food service system,' more frequently used in the USA, can be said to have been coined as a result of the 'systems approach' being applied to those operations that undertook the production of food and its service to consumers.

A simple definition of the term 'catering system' that has been proposed is that it refers to 'a particular method of organizing the production and service of food.' It is contended that the application of a systems approach involves defining and describing food production and service operations using 'systems concepts.' This enables the component parts, or subsystems, of an operation to be identified and the efficiency of their interaction assessed. Undertaking such a 'systems analysis' facilitates the application of 'problem-solving' methodologies, which can generate design solutions for the optimization of system performance. Utilization of these techniques, and awareness of the nature of organizations as being complex groups of 'subsystems' that need to be monitored and modified according to how efficiently they fulfil their specified objectives, can hence be referred to as 'systems management.'

Catering systems may be described as systems that have objectives relating to the production and/or service of food products to specified groups of consumers. Such systems normally receive a combination of inputs, which include:

- fully, part, or unprocessed food items;
- adequate numbers of appropriately skilled people;
- sufficient equipment and machinery;
- any necessary financing.

A simple 'model' of catering systems can be represented diagrammatically, as in Figure 1. The inputs form the components of a complex set of 'subsystems' that act to produce outputs required by the environment within which the system exists. The activity of the subsystems must be 'managed' effectively to ensure that they exhibit unity of purpose and exploit potential synergies within the system. The broad aim of any system is to fulfill demand for its outputs with maximum efficiency on a continuous basis. Through effective management that takes account of feedback from the market and environment, and also of 'feed-forward' from the manager's knowledge and experience of similar operations, the system's continued existence and growth are ensured.

Different views have been put forward by management and systems theorists over the years about the most important focus for managers of sociotechnical

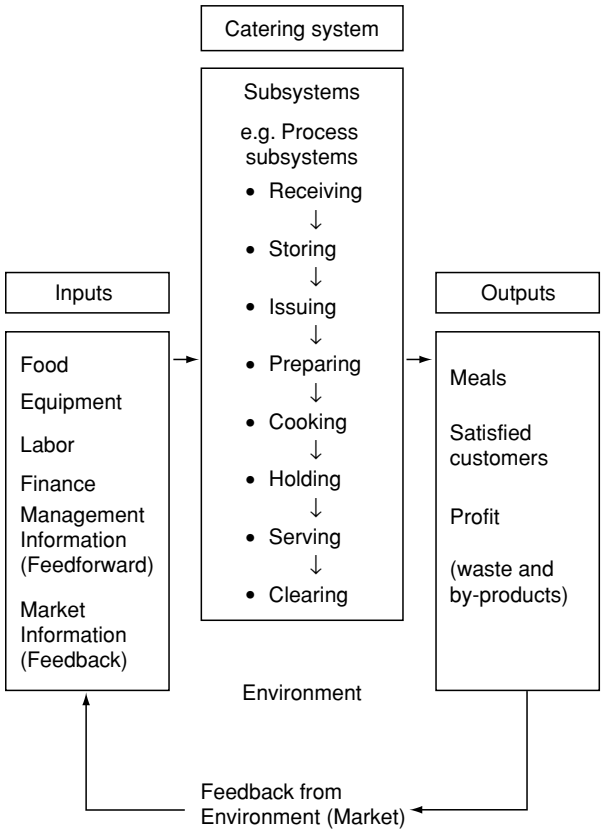

Figure 1 Simple model of a catering system.

systems. In essence, these contentions may be summarized as follows:

- The long-term survival of a system will depend on how well it 'interacts' with its environment, that is to say, how efficiently it acquires appropriate inputs and how ideally the output it produces match the current and potential demands of the environment.
- The efficiency with which a system works, and therefore thrives, will depend on ensuring that its subsystems work harmoniously towards common goals that contribute to, and are congruent with, the objective/s of the organization.

Catering systems, therefore, have been defined either in terms of the markets that they serve, that is against the context of their environment, or in terms of the differentiation in their internal processes or subsystems. There follows a brief summary of the classifications that can be applied to catering systems in relation to these perspectives.

Catering Systems Defined in Terms of Their External Environment

Catering organizations, as with all 'systems,' interact with their environment. In an ideal world, a catering system will produce the exact outputs the 'environment' demands in terms of both quantitative and qualitative requirements. A key task in managing such systems involves matching systems output to the requirements of the environment in such a way as to maximize the systems potential to do work. Thus, catering systems have typically been classified in terms of the sector of the environment they serve.

The two major sectors into which catering systems are divided are the 'Cost Sector,' which, broadly speaking, incorporates all not-for-profit catering activities, and the 'Profit Sector,' which includes all profit-orientated organizations.

The 'Cost Sector' typically includes all catering systems in the following subsectors:

- Healthcare, e.g., hospitals;
- Education, e.g., schools and colleges;
- Business and Industry, e.g., staff-feeding operations;
- Public Services, e.g., the police and armed services.

In 1996, the value of purchases (which equate with sales) in this sector, in the UK, was almost £1.5 billion, 'Healthcare', 'Education,' and 'Business and Industry' sectors each having similar values of £460–475 million.

The 'Profit Sector' includes catering systems in the subsectors listed below:

- Hotels;
- Restaurants;
- Fast Food;
- Cafes and Takeaways;
- Public Houses;
- Travel Organizations (e.g., on-board and in-flight catering);
- Leisure Operations.

In 1996, the value of purchases in the profit sector was over £6 billion, and sales were in excess of £18 billion. The Hotel sector alone accounted for £4.7 billion sales, with Public Houses being the second largest subsector having sales just less than £4 billion.

These figures serve to confirm the economic importance of catering systems to the UK economy, which reflects a global pattern in the developed world.

Catering Systems Defined by Their Subsystemic Differentiation

Catering systems may be defined in terms of the differences in the subsystems they contain. Subsystems can be defined in many ways, depending on the analysis being undertaken. In the example of 'process subsystems' alone, the possible permutations that can be identified mean that there are potentially innumerable variations in operational systems. Attempts have been made to resolve these variations into viable 'generic' groups. There are essentially three major classifications of catering systems into which the minor 'generic' variations may all be fitted. These can be summarized as follows, and their key characteristics are outlined in Figure 2.

- 'Cook–serve'/Integratedfood-servicesystems,where both the preparation of food and its service are an integral function carried out in a single operation, and there is little delay between preparation and food service. The majority of the catering industry still operates conventional systems of this kind.
- Food-manufacturing systems, where the production of food and meals is separated or 'decoupled' from the service of meals. Such systems encompass the use of chilling and freezing methods to preserve the food and in the past have been referred to as 'technological catering systems.'
- Meals-assembly/food-delivery systems, a recent systems development, in which little or no actual food preparation takes place in the system and the operation focuses on the assembly, regeneration, and service of meals.

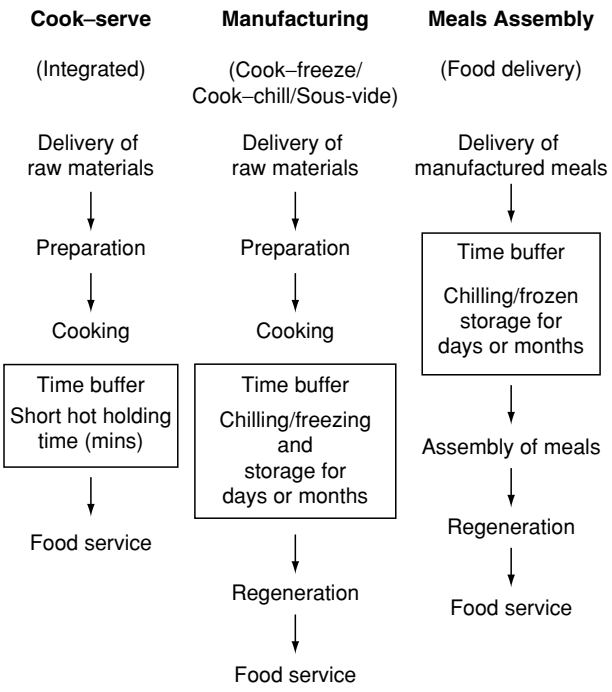

Figure 2 Major types of catering systems.

Cook–Serve Catering Systems

The cook–serve or integrated catering system represents the conventional approach to catering. These systems operate over a short time scale with the minimum feasible period or 'time buffer' between cooking and food service (see Figure 2). Where food is cooked to order, or produced in small batches to suit expected demand, it is quite easy to minimize the 'time buffer,' but this can be more difficult where food is made in bulk quantities for large numbers such as in institutional food services. The period, over which food is held hot, which, according to UK legislation, must be at a temperature above 63 °C, is largely dictated by the logistics of serving individual customers in each situation.

In cook–serve systems, should food need to be held hot for any appreciable period, time- and temperature-dependent changes occur that reduce both the sensory and nutritional properties of food. Adverse sensory changes can affect the color, texture, and flavor of foods. Losses in heat-labile vitamins such as thiamin and vitamin C also occur.

In research, the loss of vitamin C (as measured using the 'Fluorimetric method') has frequently been used as an indicator of food quality, as its destruction is time- and temperature-dependent and is accompanied by losses in sensory qualities. The effect of different cooking periods and hot holding times on the retention of vitamin C in Brussels sprouts is shown

Table 1 Comparison of the percentage overall losses of vitamin C in Brussels sprouts at the end of different periods of hot holding, in four different catering systems

System	*Overall percentage vitamin C loss at end of stages during hot holding (min)*				
	15	*30*	*45*	*60*	*90*
Cook–serve					
'Ideal' cook–serve	17.2	26.9	*	*	*
'Abused' cook–serve	25.3	32.3	*	42.7	51.4
Cook–chill					
2-day storage	34.5	41.0	50.7	*	*
3-day storage	39.8	48.1	54.3	*	*
4-day storage	46.6	53.6	60.0	*	*
5-day storage	55.7	61.9	67.5	*	*
Cook–chill sous vide					
4-day storage	32.9	38.6	45.8	*	*
5-day storage	32.7	41.4	48.3	*	*
Meals assembly	43.8	55.1	63.8	*	*

*no data available.
'Ideal' represents 15-min steaming.
'Abused' represents 30-min steaming.
From West A, Walker A and Lawson JM (1998) The effects of food processing in hospital catering systems. In: Edwards JSA (ed.) *Culinary Arts and Sciences 2. Global and National Perspectives*, pp. 283–289. London: Computational Mechanics Publications.

in Table 1. Extended cooking times and hot holding period cause substantial losses not only in cook–serve systems but also in the other systems discussed.

It can be imputed that cook–serve systems in many areas of catering have their limitations. This is particularly the case in large-scale institutional catering such as in the healthcare sector. Large hospital food-service operations formerly using traditional cook–serve methods and experiencing associated quality problems owing to extended hot holding before service, have been at the forefront of the change to food-manufacturing and delivery systems.

Food-manufacturing Systems

Food-manufacturing systems avoid prolonged hot holding of foods as they incorporate a substantial 'time buffer' between cooking and food service. This is achieved by chilling or freezing the prepared meals followed by appropriate storage. In this case, a 'time buffer' is effectively used to 'decouple' the production subsystems of the operation from food assembly and service (see Figure 2).

In food-manufacturing systems, food items are prepared in quantity in a large kitchen or centralized production unit, then chilled or frozen usually in multiportion packs, in effect preserving food until required. Products can be regenerated in the required amounts in a satellite kitchen, or using specialized reheating equipment close to the point of service (e.g., in large institutional operations) or within individual units (e.g., in an aeroplane's galley kitchen in in-flight catering). This approach minimizes the time food is actually hot-held as it should take only the time required to serve food to customers immediately after regeneration, which will minimize potential deterioration.

Meals-assembly/Food-delivery Systems

Food-delivery systems take the concept of 'decoupling' to a logical conclusion in that the manufacturing subsystems are completely removed. Food products are 'bought in' from specialist manufacturers producing frozen menu items for assembly and regeneration in simple kitchens consisting essentially of reheating equipment, or close to the consumer using specially designed heated trolleys. These catering systems, which do not require the traditional kitchen, may be referred to more aptly as 'meals assembly systems' as they only consist of the storage, assembly, regeneration, and service subsystems (see Figure 2).

The length of the 'time buffer' in these systems is determined by the preservation technique used, being up to 5 days for standard chilled foods and a year for frozen foods. Thus, the time buffer decouples production from service and potentially preserves food quality. This potential can be lost if the food is subjected to another period of hot holding after regeneration. UK guidelines recommend that it should take no longer than 15 min for meals to be served after regeneration is complete. Table 1 shows the damaging effect on vitamin C content if extended periods of hot holding do occur.

Whilst the food-manufacturing approach aims to improve efficiency by decoupling the major subsystems, 'meals assembly' actually simplifies a catering system by removing the preparation and cooking

subsystems completely. This allows the caterer to focus on food service and the monitoring of food safety and quality.

Preservation Methods in Food-manufacturing and Meals-assembly Catering Systems

Cook–Freeze

Freezing is a reliable method of preservation as it includes the barrier of the latent heat plateau seen in any time/temperature graph during the thawing of frozen food. This phase change provides a safety factor against temperature abuse, although any partially thawed foods at the surface could potentially cause food-safety and sensory problems.

A variation of this is the cook–freeze–thaw system introduced because, once thawed, the food can be treated as for cook–chill. However, the system can be potentially hazardous unless the temperatures are tightly controlled.

Cook–Chill

Cook–chill systems have gained acceptance through consumer preference for minimally processed foods. These are, however, more susceptible to temperature abuse in only having a small barrier to temperature change. Precise temperature control is necessary throughout the whole system. The time and temperature requirements currently specified under UK legislation are identified in italics in Figure 3.

Sous Vide

Sous vide is a specialist form of cook–chill system (see Figure 3) in which product shelf-life is extended from the normal 5 days up to as long as 42 days. Sous vide cooking is semicontinuous when operated on a large scale but discontinuous in smaller operations. There are particular concerns that all systems employing this method must have appropriate technical expertise to ensure that all the processing equipment and operating procedures are adequately designed so that products receive sufficient heat treatment.

Questions about food safety and quality in sous vide operations have been raised. Food is cooked in sealed plastic bags under a partial vacuum, which slows the rate of heat transfer and could result in underprocessing. This has led to concerns that psychrotrophic pathogens, such as *Listeria monocytogenes*, *Yersinia enterocolitica*, *Clostridium botulinum*, and the like could survive. The most hazardous of these is *Clostridium botulinum*, some strains of which produce a toxin at temperatures as low as 3.3 °C. Although the vegetative cells of

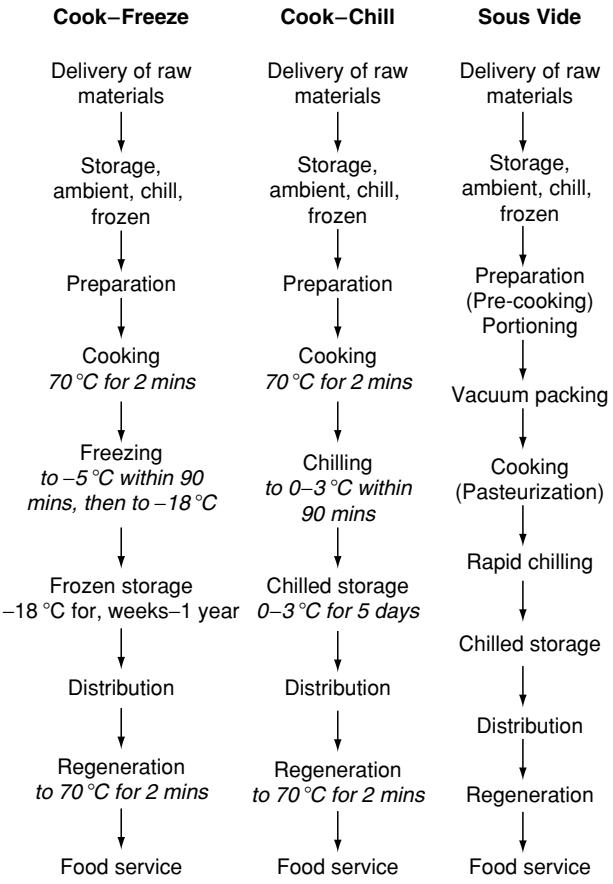

Figure 3 Stages involved in food-manufacturing systems.

Clostridium botulinum are destroyed by heat treatment of not less than 70 °C for 2 min at the coldest spot, foods should be heated to 90 °C for 10 min to ensure destruction of spores.

Concerns about the safety of vacuum-packaged and modified atmosphere-packaged products have led the Advisory Committee on the Microbiological Safety of Food in the UK to issue a Code of Practice that recommends the use of additional hurdles and the application of the HACCP technique in the manufacture of all such products.

The retention of most of the original juices within the package after cooking has led to claims that sous vide processed foods have enhanced sensory qualities. This claim has not been fully supported by research findings to date.

Further Developments in Cook–Chill

As there is doubt that chilled storage alone can assure the safety of sous vide products, recent work has focused on the formulation of products with additional hurdles. This involves the combination of several factors together, which collectively ensure

the microbial safety of food, even though each hurdle on its own might be insufficient to maintain safe food.

Examples of hurdles are low water activity, low pH, use of modified atmospheres (*See* **Chill Foods**: Effect of Modified-atmosphere Packaging on Food Quality) irradiation, added organic acids and protective cultures. The incorporation of such adjuncts provides hurdles to microbial growth and can also impact positively on safety, nutritional, and sensory aspects.

Research indicates that cook–chill produces a product that is inferior in sensory qualities to that produced in cook–serve systems in ideal conditions, but better than that produced in the nonideal or abused situations often found in catering systems.

Meals-assembly Techniques

The main stages in food delivery or meals assembly systems are shown in Figure 2. In this type of system, menu items are prepared and processed by food manufacturers to a specification prepared by a caterer. Typically, these menu items will include main courses, vegetables, and puddings, plus accompaniments such as sauces. The meals are chilled or frozen, and packaged in bulk foil containers suitable for use in commercial catering operations. After delivery to the catering unit, food is stored appropriately until requisitioned. Meals are then assembled, regenerated usually using forced-air convection ovens, and served.

Meals assembly represents a radical change in catering systems. By the year 2000, such systems were being used by almost a fifth of UK NHS hospitals as well being more widely adopted within the profit sector, particularly by pub and fast-food operations. This type of system has proven popular in healthcare catering, because it offers cost savings by reducing the need for skilled labor and expensive plant and equipment.

The meals-assembly system allows the catering manager to concentrate on food-service techniques and monitoring methods, in particular to control time and temperature factors. A quality-management approach is essential for both the manager's own operation and for their suppliers, the food manufacturers.

The final quality of food produced using these new systems is governed by the control of time–temperature parameters throughout the system. Potentially, regeneration is the stage that can affect food quality most adversely, as primary cooking is common to all types of system.

Regeneration

In a meals-assembly system, regeneration usually involves a standard reheat cycle in which a mixed load of chilled or frozen meals is regenerated in a specially designed oven. This means that food can be under- or over-heated as different foods have different thermal characteristics. Owing to food-safety concerns food, is often effectively overheated, with core food temperatures at the end of regeneration in excess of 80 °C being usual. The time spent at this temperature level is particularly detrimental to food quality, causing both heat-labile vitamin losses and sensory losses. There is often potential to reduce the severity of the regeneration process whilst still maintaining safety, yet preserving sensory and nutritional quality, but reliable control and monitoring methods are vital.

Various types of equipment can be used for regenerating assembled meals, the most common currently being forced-air convection ovens. Forced-air convection technology gives a fast regeneration time owing to constant air velocity and even temperature distribution. Hot-air convection currents being forced to circulate the oven cavity by a fan achieve this. These currents remove the steam layer from the surface of the food quickly and enable heat to reach the food directly, thus causing a rapid rise in temperature.

The heating effect can be enhanced further by the incorporation of steam injection into the oven cavity. The latent heating effect of condensing steam on the surface of the food package promotes faster reheating.

The use of microwaves in regeneration is less common but can be successfully combined with forced-air convection technology. Microwaves alone are better for reheating chilled, rather than frozen, foods because of the different absorbency of microwaves by ice and water, which can cause thermal runaway, leading to hot spots and cold spots in the product. This can still occur in chilled products because of different dielectric properties in foods of varying compositions. (*See* **Chill Foods**: Effect of Modified-atmosphere Packaging on Food Quality; **Chilled Storage**: Principles; Attainment of Chilled Conditions; Quality and Economic Considerations; Microbiological Considerations.)

The newer systems of food manufacture and food delivery require strict control throughout heat processing to ensure food safety and maintain consistent quality. New approaches to equipment design include the application of computational fluid dynamics and the use of model-based control design.

Food-safety Issues in Catering Systems (*See* Food Safety)

Temperature control in catering systems is imperative, and monitoring systems have consequently

been developed in recent years. Larger units can have computer-controlled temperature-monitoring devices, whereas smaller operators generally rely on making manual checks and keeping manual records.

Failure to control time and temperature remains a problem responsible for the majority of food-poisoning cases caused by the catering industry. In the UK, there is a mandatory requirement for caterers to identify the steps in the system where hazards can arise and put appropriate control and monitoring measures in place. This relates to the implementation of Hazard Analysis Critical Control Point (HACCP) techniques. Caterers are encouraged to develop modified HACCP approaches based on the Assured Safe Catering methods, but this is expected to be a slow process that will take place over several years. (*See* **Hazard Analysis Critical Control Point**.)

There is a need to address the safety of individual foods as caterers continue to cause outbreaks of food poisoning owing to the use of contaminated fresh shell eggs. The UK Government has recommended that caterers should make use of heat-treated egg products, pasteurized to eliminate *Salmonella enteritidis*. (*See* ***Salmonella***: Properties and Occurrence; Detection; Salmonellosis.)

Systems used in the catering industry can be seen as a continuum of the techniques used in the food-manufacturing industry and therefore can employ similar methods of control. For example, the concept of quality management can be usefully applied. As more food-manufacturing techniques are adopted, and encouraged by regulatory authorities, these techniques should accompany the HACCP approach so that safety and quality issues are both addressed.

Catering Systems in Conclusion

There is a need for further research to confirm claims of improved sensory qualities, safety, and nutritional values in food manufacturing and meals assembly systems. Quality changes can occur rapidly, but many variables affect the end result that is often product-specific. The lack of control and standardization in both the catering industry and experimental methods makes comparisons between studies difficult. The fact that technological innovations in catering systems have not introduced new methods of food production but rather, as some have noted, introduced systems of food preservation has not been proven to have had a positive impact on food quality predicted, as a matter of course.

Generally cook–serve catering systems still predominate throughout the food-service industry. It has been found that increased size, certainly in the healthcare sector, tends to dictate whether or not food-manufacturing systems have been adopted. However, recent surveys have indicated that around three-quarters of hospitals in the USA and almost the same proportion in the UK still operate conventional food-service systems.

In recent years, the emergence of food-delivery/meals-assembly systems has offered the opportunity for many smaller catering operations to adopt this type of system. Whilst this development means that catering systems can provide a wider menu with less skilled staff and reduced equipment needs, it has been observed that quality suffers if regeneration of products is poorly controlled and if the aesthetics of meals assembly and customer service expectations are overlooked.

It is pertinent to recall that the term 'catering system' emerged from the application of systems theory to the management of food-service operations. It should not be forgotten, therefore, that the systems approach is 'holistic,' and that any innovation that affects one or more subsystems will also affect the rest of the system, its inputs, and the efficiency and effectiveness with which it produces its outputs.

See also: **Chilled Storage**: Principles; **Chill Foods**: Effect of Modified-atmosphere Packaging on Food Quality; **Chilled Storage**: Packaging Under Vacuum; **Food Poisoning**: Classification; Tracing Origins and Testing; Statistics; Economic Implications; **Food Safety**; **Hazard Analysis Critical Control Point**; **Legislation**: History; International Standards; Additives; Contaminants and Adulterants; Codex; **Quality Assurance and Quality Control**

Further Reading

Creed PG and Reeve W (1998) Principles and applications of sous vide processed foods. In: Ghazala S (ed.), pp. 25–51. *Sous Vide and Cook–Chill Processing for the Food Industry*. Gaithersburg, MD: Aspen.

Department of Health (1999) *Assured Safe Catering. A Management System for Hazard Analysis*. London: The Stationery Office.

Department of Health and Social Security (1989) *Chilled and Frozen – Guidelines on Cook–Chill and Cook–Freeze Catering Systems*. London, UK: HMSO.

Johns N and Jones P (1999/2000) Parts 1, 2 and 3 – Systems and management. *The Hospitality Review* July 1999: 43 *et seq*., Oct. 1999: 40 *et seq*., Jan. 2000: 47 *et seq*.

Jones P (1994) *Catering Systems*. In: Davis B and Lockwood A (eds) *Food and Beverage Management – A Selection of Readings*, pp. 131–144.

Kirk D and Pine R (1998) Research in hospitality systems and technology. *International Journal of Hospitality Management* 17: 203–217.

Reeve WG and Edwards JSA (1996) Catering systems revisited. In: Edwards JSA (ed.) *Culinary Arts and Sciences*

Global and National Perspectives, pp. 117–127. London: Computational Mechanics Publications.

West A, Walker A and Lawson JM (1998) The effects of food processing in hospital catering systems. In: Edwards JSA (ed.) *Culinary Arts and Sciences 2. Global and National Perspectives*, 283–289. London: Computational Mechanics Publications.

Nutritional Implications

A G Smith and C Hunt, University of Huddersfield, Queensgate, Huddersfield, UK

Introduction

The catering or food service industry incorporates a wide range of organizations engaged in the provision of food products for customers within either the commercial or welfare sectors (*See* **Catering**: Catering Systems). Catering operations have been defined as: 'establishments whose main business comes from the sale of meals . . . refreshments . . . alcoholic drinks . . . or the provision of accommodation.' The following article focuses specifically on the aspect of the provision of meals by food service systems. The process subsystems involved in typical catering operations are identified and the consequent nutritional implications, both positive and negative, that these processes might have for the final food products is then considered.

It is the preparation and cooking of raw food ingredients that renders the nutritional benefit of food available to the consumer. In this sense, the food service system may be seen as both positive and necessary for the acquisition of necessary nutrition by human consumers within today's society, where an increasing number of meals are consumed outside the domestic setting. However, there are many opportunities for an inadequately managed food service system to contribute to the avoidable deterioration of the nutritional content of food items.

General Nutritional Considerations

Nutrients can be subdivided into macronutrients (fats, protein, and carbohydrate) and micronutrients (minerals and vitamins).

Macronutrients are relatively stable during storage and heat processing. Commodities such as meats and dairy products provide macronutrients such as proteins and fats. They may also provide fat-soluble vitamins. These nutrients are relatively stable, although they can be broken down by certain activities, such as the natural digestive process, whereby they can be utilized by the body.

Cereal products are key sources of carbohydrate in the diet. Processed cereal products such as pasta and breads hence are also a major provider of this essential energy fuel. Carbohydrates are relatively stable and, if commodities such as those outlined are stored and handled correctly, then little deterioration should occur whilst the products are processed by a food production or service system.

Of the micronutrients, minerals, which occur in a variety of foods, are heat-stable and will not be destroyed by cooking, Nevertheless, some loss can occur by water leaching.

Vitamins vary considerably in terms of stability. The fat-soluble vitamins (A, D, E, and K) found in meats, nuts, and yellow fats, for instance, are relatively stable. The B-complex of vitamins and vitamin C, which are largely supplied by fruit and vegetables, are the nutrients most vulnerable to loss during storage and processing. These water-soluble vitamins are also generally unstable to heat and liable to oxidation.

A Catering System – Defined in Process Terms

Amongst the many ways in which catering systems can be defined (*See* **Catering**: Catering Systems), perhaps the most useful of these, when considering the production elements of catering, is to view a catering system in terms of process subsystems (**Figure 1**).

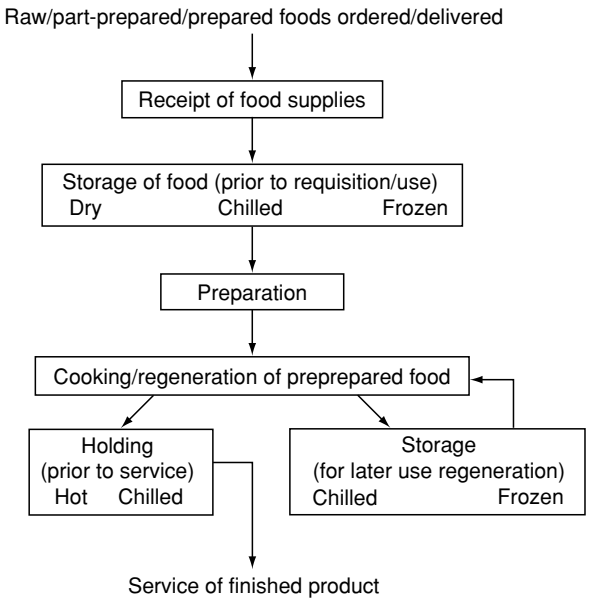

Figure 1 Process subsystems within a typical catering (production) system.

To consider the nutritional impact on food items as they are transformed from raw products into finished dishes for the consumer, each process stage is discussed below.

Ordering and Receipt of Food Supplies – Nutritional Considerations

The first process undertaken by all food service systems is generally that of ordering and receiving the raw materials or part-processed goods that will be used to make finished products. Whilst there are aspects of efficiency in the undertaking of this process that should be considered in order to insure that foods retain their maximum nutritional value, it is apposite to consider initially the nutritional quality of the foodstuffs when received by the caterer.

Many factors affect the incoming nutritional value of food materials, which may ultimately have more impact on the final nutritional profile of manufactured food products than the following preparation activities might have. These influences vary depending on the commodity, so that animal-based raw materials such as meats may display different nutritional values for different reasons to plant-derived foodstuffs such as vegetables.

Key influences on the nutritional content of these major commodity areas are listed in Table 1.

In order to insure that goods received are of appropriate quality, it is important that food production and service systems draw up comprehensive purchase and product specifications. Even so, it will not be possible to avoid the substantial effects on nutrient values that certain natural occurrences or agricultural practices instigate.

Receiving Procedures – Impact on Nutrition

Receiving procedures for food items arriving at a food service operation can have a negative impact on nutritional value if not planned and executed efficiently. All food service systems should have procedures that incorporate staff accountability and immediate checking and transfer to appropriate storage facilities for foodstuffs at the time of delivery.

Accountability is important for retracing the chain of events if there should be any problems with food materials found in storage, hence a responsible member of staff should be appointed to undertake the receipt of deliveries.

If food is held at the point of receipt prior to appropriate storage then time- and temperature-related factors can lead to rapid onset of deterioration. Foods that are not properly stored immediately on receipt may be open to contamination or infestation, leading to hygiene and food safety-related problems and an adverse impact on nutritional value.

Table 1 Factors influencing nutritional value of food comÙmodities prior to processing

Animal origin (e.g., meats)	*Plant origin (e.g., vegetables)*
Breed/strain/genetic make-up of animal	Variety of plant
	Climate variations
Age of animal at slaughter	Seasonal factors
Feedstuffs used during growth	Quality of soil
Quality of grazing pasture	Use and nature of fertilizers
Other seasonal variations	Growing conditions
Use of growth stimulators	Degree of maturity of crop

Nutritional Implications of Storage of Food

There are essentially three key forms of storage of food items that most, if not all, food production and service systems utilize for keeping food prior to preparation into finished products. Items may be held as:

- dry goods
- chilled goods
- frozen goods

Each of these categories of stored goods may be subject to nutritional deterioration whilst being held prior to use. The amount of deterioration that takes place will largely depend on the following factors:

- the food item itself
- the nature of the item's nutritional make-up
- the type of storage
- the length of time it is held in storage

Given that the nutritional content for key commodities can vary on delivery due to many factors whilst stored pending use, in most cases further deterioration will be largely time- and temperature-dependent. The higher the temperature stored and the longer the item is held, the more the nutritional value of some foods will decline. The likelihood that food will exhibit nutritional deterioration depends on the nature of the major nutrients that food contains.

As mentioned earlier, vitamin C is the nutrient which is most susceptible to loss, so measurement of the extent of its retention during storage or processing has come to be used as an indicator of the efficiency of food production systems. If the proportion of vitamin C retained in the food is good, then this indicates that the levels of the less unstable nutrients will be also be good. (*See* **Catering**: Catering Systems.)

The major food commodities that supply vitamin C are fruit and vegetables. Fruit and vegetables should not be stored above 8 °C, with optimal storage temperature within the range 0–3 °C. Many fruit and

vegetable products are stored prior to delivery in a controlled atmosphere where the environmental gas mixture contributes to the lengthening of shelf-life. Products delivered to food service systems are often supplied in modified-atmosphere packaging (MAP). MAP should be retained intact until the stocks are issued, as the gas mixture used in the packing process can increase shelf-life of fruit and vegetables from between 2 and 7 days to up to 35 days, depending on the item. The inhibition of the ripening process and enzyme activity preserves the organoleptic quality and nutritional value of foods held in storage pending use. (*See* **Chill Foods**: Effect of Modified atmosphere Packaging on Food Quality.)

For food items that provide a source of the relatively stable nutrients, the nutritional profile of the foods will be adversely affected if storage is inappropriate such that loss due to spoilage renders the food unfit to eat. This means that the nutrient value of the food would no longer be available for safe consumption. For instance, over time poorly stored foods which contain fats will become rancid. This will occur more rapidly if temperatures are too high, and the product is poorly wrapped, allowing oxidation to occur. Protein-rich foods such as meats quickly become subject to contamination, hence unsafe to eat, if stored at temperatures and in conditions that allow the establishment of bacterial colonies that thrive on the same sources of protein as humans.

In the case of carbohydrate sources, such as cereals, grains, pasta, and flour, it is essential that these items are kept dry and in containers that are impervious to insect pests. The nutrient value *per se* will not deteriorate but poor storage can lead to spoilage or contamination, rendering the food unfit to eat.

Food Preparation Techniques and Impact on Nutrient Content

Many food items delivered to catering systems require preparation before they can be utilized. Fruit and vegetables will need washing and peeling and may be further prepared, by chopping or slicing, for example, prior to cooking. Raw meats may require differing amounts of preparation depending on how much has already been done by the butchers prior to delivery and according to what use is to be made of the item.

Whilst the purpose of preparation of foodstuffs is either to remove any inedible parts or generally to enhance the eating quality of the ingredient, any preparation process can lead to a reduction in overall nutritional value of the food.

The process of peeling fruit and vegetables, whilst enhancing visual appeal and possibly eating quality, will normally result in the reduction of overall nutrient content. If any proportion of the preparation waste would originally have been edible, overall nutrient value will be reduced. It has also been shown that peeling vegetables can lead to increased vitamin loss as a result of leaching during the cooking process.

In the preparation of meat it is increasingly likely that, due to healthy-eating advice, the larger proportion of surface fat will be removed from raw meats during the preparation stages. In effect, the overall potential nutritional value of the meat is reduced by doing this. However this might be viewed as acceptable nutritional loss in the light of the potential contribution this makes towards achieving healthier diets.

It can be argued that the consequent improvement in final eating quality of a product achieved by preparing ingredients appropriately could result in more of the total finished product being consumed than if such preparation had not been undertaken. This implies that good preparation may lead to more efficient consumption of the nutritional potential of foods items made for consumption. However, for the manager of a food-manufacturing or catering operation it is imperative that good staff training emphasizes the importance of minimizing waste during the preparation stages in food production. This will mean that both nutrient contents and potential profits are preserved.

Cooking Processes and Nutritional Implications

In general, cooking has the biggest nutritional implications of any single process undertaken within a typical catering system. The impact of cooking on nutrients varies according to the method of cookery, the length of time food is cooked for, and the temperatures employed.

Broadly, the shorter the cooking time and the more quickly food is consumed after it is prepared, the less the unstable nutrients will have declined. However, the methods of cookery that are used in most systems are dependent upon the nature of the foods themselves. The nutritional implications of the methods of cookery are briefly considered below.

Dry-Heat Cooking

These cooking methods are largely applicable to proteinaceous foods and to certain carbohydrate nutrient sources. Meats, poultry, and fish, as major protein sources, can be baked, roast, or grilled, for instance. Cakes and bread are baked, whereas potatoes are often baked or roasted: all these are important sources of carbohydrates.

These dry-heat cooking methods generally require quite high temperatures (150–400 °C) and the period for which heat is applied is directly related to the size and density of the food being cooked. However, proteins and carbohydrates are quite stable and efficient cooking results in positive changes taking place during cooking that increase the edibility and digestibility of these nutrients.

When carbohydrate-rich foods are dry-cooked, the starches are turned to easily digested sugars called dextrins which darken as they cook. This accounts for the browning of toast and the surface of potatoes, as dextrins caramelize, and which shows when a food is cooked. When such foods are overcooked, the sugars char, which can impart unpleasant visual and flavor characteristics, making foods less palatable and resulting in the destruction of the carbohydrates that have been burnt. B-vitamins can also be lost in the toasting process, e.g., toasting bread results in considerable loss of thiamin.

Moist-Heat Cooking

These methods are generally applied to:

1. Delicate foods such as vegetables that contain a high percentage of water in themselves, and must be kept moist to insure that they are edible and digestible. This also applies to delicate protein foods, such as fish. The final eating quality of these foods can benefit from gentle moist cooking for a relatively short period of time.
2. Tougher protein foods such as stewing meat and boiling fowl, which require moderate temperatures for longer times to retaining moistness and achieve a tender, palatable product.

For vegetables this method of cookery can be problematic because they are often a source of the least stable vitamins, such as vitamin C and B-vitamins such as folate. Vegetables commonly lose around half of their vitamin C content on boiling, mostly due to the vitamin dissolving into the cooking liquor by leaching out of the cut or peeled vegetables. Similarly, almost a third of folate is lost on boiling. In contrast, it has been shown that steaming vegetables considerably reduces such losses (vitamin C 10–20%; folate 0–10%) because it involves a lot less water so there is less opportunity for leaching to occur.

Nutritional analyses have shown that vegetables lose considerably higher proportions of these vulnerable vitamins when they are peeled and when they are chopped up. New potatoes lose only about 5% of vitamin C on boiling when their skins are left on, as opposed to up to 50% loss when peeled. As the skin acts as a natural barrier to vitamin leaching, caterers should consider retaining edible peels whenever possible during cooking. It is also advisable to minimize the chopping of vegetables within reason, as whole items retain vitamins much more successfully because the cellular structure of the vegetables remains intact and cell contents are less likely to leak into the cooking liquor.

Although heat in combination with exposure to oxygen in the air will result in permanent destruction of a proportion of vitamins during the cooking of vegetables, by far the greater amount of 'lost' vitamins will simply be dissolved into the cooking liquor. It is for this reason that caterers should make it common practice to incorporate cooking liquors into stocks, sauces, and gravies wherever possible as this will insure that such 'lost' vitamins are retained for potential consumption.

Many vegetables also provide important carbohydrates and fiber in the diet. Moist cooking methods help to gelatinize carbohydrates, which aids their digestion. Fiber, in the form of cellulose, also needs to be softened by such cooking to make the vegetables more palatable. The more cellulose there is, the longer cooking time is required. Hence carrots need longer cooking than some delicate green vegetables. This has implications for the caterer who wishes to achieve the optimum balance of color, flavor, and digestibility, along with retention of maximum nutritional value.

Delicate protein foods such as fish may be poached or steamed to achieve moist and palatable results. Flavors will not intensify as much as they might during dry cooking, which effectively dehydrates foods and hence concentrates flavor elements. Natural mineral salts are also dissolved into the cooking liquor and these are not only important nutrients but also key to the development of flavor. There may also be some minor loss of proteins into the cooking liquor. Once again, using any resultant stock, after cooking is complete, to make sauces to be served with the dish will recoup the majority of these nutrients.

When moist cooking methods are applied to appropriate cuts of meat in the process of making stews, pie fillings, or casseroles, for example, this helps to turn the tough collagen in the meat to gelatin. Gelatin is more easily digested because it is a soluble protein, hence the cooking process not only improves palatability but also makes the nutrients more readily available to the consumer of the product. Once again, it is imperative that the cooking liquor becomes part of the dish, in the form of sauce or gravy, because flavorsome, nutritionally valuable natural mineral salts as well as some proteins will be dissolved or suspended in it.

Nutritional Impact of Regeneration and Warm-Holding of Foods

Traditionally catering systems either operated on a basis of 'making-to-order,' where there is no delay between cooking and service of food, such as in quality restaurant operations, or bulk production, when food may be held warm for considerable periods prior to service, as in hospital and canteen feeding operations. In relatively recent times interrupted catering systems, such as cook-chill and cook-freeze operations, have been introduced (*See* **Catering**: Catering Systems) where foods are cooked, then rapidly chilled or frozen, to be reheated later just prior to use. These different methods have implications for retention of nutritional value of foods that must be considered by caterers.

Traditional Catering Methods and Warm-Holding Foods

Where food is made-to-order and served, and presuming quality stock has been purchased originally and stored, prepared, and cooked efficiently, then it is likely that consumers will receive food with optimal palatable, digestible, and nutritional characteristics. However, for those large-scale operations where it is necessary to hold food warm for any length of time, the outcome can be significantly different.

Whilst some unstable nutrients are irredeemably destroyed during the cooking process, as discussed, due to the action of heat and oxidation, this may be seen as a necessary trade-off to achieve improved digestibility and hence give access to the vast majority of nutrients remaining in the food, and liquors produced during cooking. Should food have to be held warm for any period after cooking and before consumption, the only nutritional effect is adverse.

Maintaining foods at safe temperatures (above 63 °C, and usually much higher) and further exposure to atmospheric oxygen results in rapid and unrecoverable depletion of unstable nutrients. Warm-holding of cooked vegetables, for instance, which are the key dietary sources of vulnerable nutrients such as vitamin C, can result in typical permanent losses of up to 20% h^{-1} and 10–15% h^{-1} for folic acid. Prolonged warm-holding also results in deteriorating organoleptic (flavor, texture, aroma) and visual properties. This reduction in the palatability of foods may lead to less food being consumed, hence further reducing utilization of remaining nutritional potential.

Cold-Holding and Regeneration of Foods

In interrupted catering systems, food is chilled rapidly immediately upon completion of cooking. The process of chilling itself has no adverse impact on nutritional value of food. However, if foods are only chilled and not frozen, a significant reduction of unstable nutrients continues to occur during chilled storage. Vitamin C levels have been shown to fall by around 10% per day during such holding, although this varies depending on the particular food. If foods are frozen then further vitamin losses are virtually halted, although losses may occur during the thawing process, if foods are not regenerated from frozen, when liquid containing vitamins may run off.

Frozen or chilled foods will need to be regenerated prior to service and brought to a safe temperature in excess of 63 °C. During this second heating, further permanent vitamin losses will occur, as in primary cooking, due to the effects of heat and oxidation. Foods that have been regenerated must then be served and consumed as swiftly as possible, to reduce to a minimum the warm-holding effects previously discussed.

The Importance of Catering in Relation to Nutritional Intakes

Whilst there are many influences that can affect the final nutritional value of a meal or food item before it reaches the consumer, it should be noted that the contribution that catering makes to the overall nutrition of individuals varies greatly.

Catering systems operate either as welfare-related concerns within the cost sector, or as commercial operations within the profit sector (*See* **Catering**: Catering Systems). Where customers are captive, such as in hospitals, schools, or prisons, then the catering provision may be an important contributor to daily nutritional intakes. Just how important this becomes to individuals will depend on the length of time for which they are reliant on catering provisions for all or a major part of their daily diet. Long-stay hospital patients, boarders at schools, and those retained at Her Majesty's Pleasure for prolonged periods will need to acquire satisfactory nutritional benefit from the food provided by the caterers in their respective environments.

In the UK the National Food Survey, conducted by the Ministry of Agriculture, Fisheries and Food (MAFF), indicated that households were spending almost £24 per person per week on average on food and drink in 1999. Of this sum, just over £7 per person per week was spent on meals and beverages eaten out – around 30% of total expenditure. Between 1998 and 1999 expenditure on eating out rose by just over 5%, whereas the total spent on food and drink at home remained almost static. Whilst the economic climate is generally positive it is likely that such a tend will continue, as consumers

may not feel the need to 'tighten their belts.' In terms of nutrition, the implication must be that, as the proportion of expenditure on food and drink eaten outside the home increases, so the relative importance of catering and food service operations to the national diet increases.

The National Food Survey indicated that the pattern of consumption of food away from the home varies considerably between socioeconomic groups in respect of total amount spent and nature of foods consumed. Gender and age differences also influence individuals, food consumption patterns. This means that it is difficult to make any generalizations about the impact of commercial catering on the nutrition of the community at large. However, some observations relating to the nutritional value of food and drink consumed include the fact that fats and alcohol provide a considerably higher proportion of the total energy (caloric value) of food eaten out. Overall, food eaten away from the home accounted for just over 12% of total energy intake for the average person in the UK in 1999.

There are increasing concerns about the increasing prevalence of obesity in developed countries such as the UK. Medical practitioners and nutritionists regularly confirm this growing problem, exemplified by the fact that nearly one in five adults in the UK are classified as clinically obese. Because obesity is recognized as contributing to a range of health problems such as cardiovascular disease, it is suggested that caterers should make any efforts that they can to contribute to the general dietary goals of reducing fats, sugar, and salt by examining and modifying recipes whenever they can.

See also: **Catering**: Catering Systems; **Chill Foods**: Effect of Modified-atmosphere Packaging on Food Quality

Further Reading

Department of Health (1995) *Nutrition Guidelines for Hospital Catering – Prepared by Nutrition Task Force of Health of the Nation*. London: HMSO.

Hunt C (1984) Nutrient losses in cook-freeze and cook-chill catering. *Human Nutrition Applied Nutrition* 38A: 50–59.

Hunt C (1990) Vitamins in institutional feeding. *Journal of Micronutrient Analysis* 7: 193–206.

MAFF (2000) *National Food Survey 1999 – Annual Report on Food Expenditure, Consumption and Nutrient Intakes*. London: Stationery Office.

West A, Walker A and Lawson JM (1998) The effects of food processing in hospital catering systems. In: Edwards JSA (ed.) *Culinary Arts and Sciences 2. Global and National Perspectives*, pp. 283–289. London: Computational Mechanics Publications.

Williams PG, Ross H and Brand Miller JC (1995) Ascorbic acid and 5-methyltetrahydrofolate losses in vegetables with cook-chill or cook/hot-hold foodservice systems. *Journal of Food Science* 60(3): 541–546.

CELIAC (COELIAC) DISEASE

P D Howdle, St. James's University Hospital, Leeds, UK

Introduction

Celiac disease, or gluten-sensitive enteropathy, is a common condition in which the principal abnormality is in the small intestine. This abnormality is related to the ingestion of a cereal protein, gluten, and leads to a variety of clinical problems. The pathogenesis of the abnormality is thought to be an immune reaction to gluten in a genetically predisposed individual.

The clinical presentation varies from a normal state to a severely malnourished state, and there are a number of associated diseases.

Treatment is with a gluten-free diet and, in the majority of patients, is very effective.

Epidemiology

The most comprehensive data on the prevalence of celiac disease are from Western Europe. The prevalence of clinically apparent disease ranges from one in 300 to one in 1000. The disease has thus been described principally in Caucasian peoples, although cases have been diagnosed in all other ethnic groups. Any variation between ethnic groups may be due to a genetic variation between different populations, the availability of diagnostic facilities, and the consumption of the relevant cereals, since the disease can only present in wheat-eating populations.

In recent years, serological antibody tests have been used to screen for the disease in the general population of several European countries. Such studies suggest that the true prevalence may be as high as one in 140, although the majority of the subjects have no apparent symptoms of the disease.

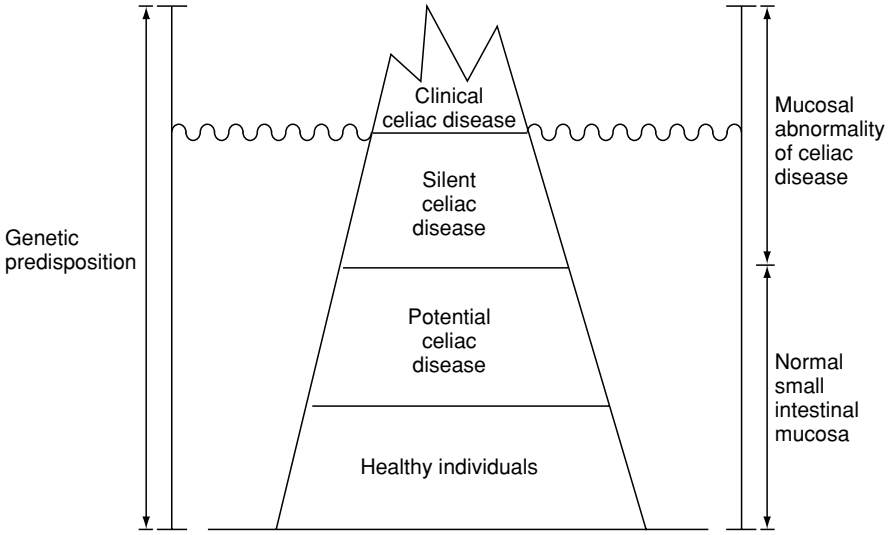

Figure 1 'Celiac iceberg' showing the spectrum of disease and small intestinal abnormality.

Such screening studies have led to the concept of the celiac iceberg (see Figure 1). This suggests that a minority of patients have clinical evidence of the disease, others have the small intestinal abnormality but are asymptomatic (silent disease), and others have the genetic predisposition but have not yet developed the pathological changes that could lead to symptoms (potential disease). There are normal individuals who have the genetic predisposition as described at present, but who never develop the disease. There is presumably an as-yet unidentified factor that distinguishes these normal individuals from the potential celiac patient.

Pathogenesis

Genetics

Celiac disease has been recognized as being a familial condition for many years. Approximately 10% of first degree relatives have the condition. The mode of inheritance is not understood.

It is also known that a particular haplotype in the MHC region of chromosome 6 is necessary in order to develop the disease. The extended haplotype concerned is B8, DR3, DQ2 but it is the DQ2 haplotype that is specifically associated with the disease, being present in 95% of patients. The 5% of patients who are DQ2-negative are DQ8-positive. It should be noted that other genes must be involved, since approximately 30% of the general population carry these genes, and the majority of these do not develop the disease. Hence, DQ2 or DQ8 is necessary, but not sufficient in terms of the genetic constitution of an individual in order to develop the disease. For this reason, other genes are being sought elsewhere in the genome that might predispose to the disease, but to date, no consistent results have been produced.

Gluten

The term 'gluten' is the generic name applied to the storage proteins found in wheat, rye, and barley. Gluten is responsible for the visco-elastic properties of these cereals, which make them good for baking. It is the insoluble material (a complex mixture of proteins, with small amounts of lipids, sugars, and minerals) obtained when flour is washed in tap-water to remove the soluble substances (albumins, globulins, and starch). Gluten is thus an ill-defined mixture that varies between different flours and is even dependent upon the solubilizing properties of the tap-water used. Gliadin is the ethanol-soluble component of wheat flour, usually solubilized in 70% ethanol; the same component can be extracted from wheat gluten. Gliadin, although a complex mixture of proteins, is thus more homogenous than gluten. The comparable ethanol-soluble components from rye and barley are termed secalin and hordein, respectively. Collectively, these proteins are prolamins. All these prolamins are related phylogenetically and are rich in the amino acids proline and glutamine. All are capable of provoking the pathological changes in the small intestine characteristic of celiac disease.

Most is known about wheat gluten or gliadin, and the terms are often used interchangeably when referring to their involvement in celiac disease. Gliadin comprises approximately 50% of wheat proteins, and a single variety of wheat may have 45 different gliadins. Their complexity has made them difficult to

investigate, but some gliadins have been sequenced, and synthetic or chemically derived peptides from gliadin have been tested in *in-vitro* or *in-vivo* experiments in celiac disease. It appears that short peptides, approximately 10–20 amino acids in length, which are particularly rich in glutamine, are capable of producing small-intestinal damage. It is believed, therefore, that ultimately, several peptides from the prolamins of wheat, rye, and barley, with similar amino acid sequences, will be found to be responsible for initiating the disease process in predisposed individuals.

Oats are a major related cereal, but the prolamin in oats (avenin) only accounts for 5–15% of the oats' protein. It is thought that because of this low prolamin content and possible dissimilarities between the amino acid structures of avenins and the other prolamins, oats are nontoxic to the vast majority of celiac patients.

Immunology

There is much evidence of enhanced immunological activity in celiac disease. In the small-intestinal mucosa where the characteristic pathology develops, there is an increase in T-lymphocytes, principally of the helper (CD4) type in the lamina propria, resulting in the production of pro-inflammatory cytokines, and of the cytotoxic-suppressor (CD8) type in the epithelium, the so-called intra-epithelial lymphocytes. Concomitantly, the plasma cell numbers increase, leading to the production of immunoglobulins, particularly antibodies to gliadin and tissue auto-antigens. Recently, tissue transglutaminase (tTG), a ubiquitous enzyme present particularly in the lamina propria, which catalyzes the cross-linking of proteins, has been reported to be an auto-antigen in celiac disease. The antibody against this antigen is termed the 'anti-endomysial antibody' and is very specific for the disease when measured in the serum of untreated patients. Interestingly, gliadin is an excellent substrate for tTG because of its high glutamine content. It is currently hypothesized that in celiac patients, tTG reacts with gliadin, thus forming a deaminated gliadin and a gliadin–tTG complex. This neo-antigen could then be involved in the initiation of an immune reaction in a predisposed individual.

Pathophysiological pathway

A possible pathway leading to the small intestinal mucosal damage in celiac disease is shown in Figure 2. This scheme takes into account the pathogenetic factors discussed above. The products of the

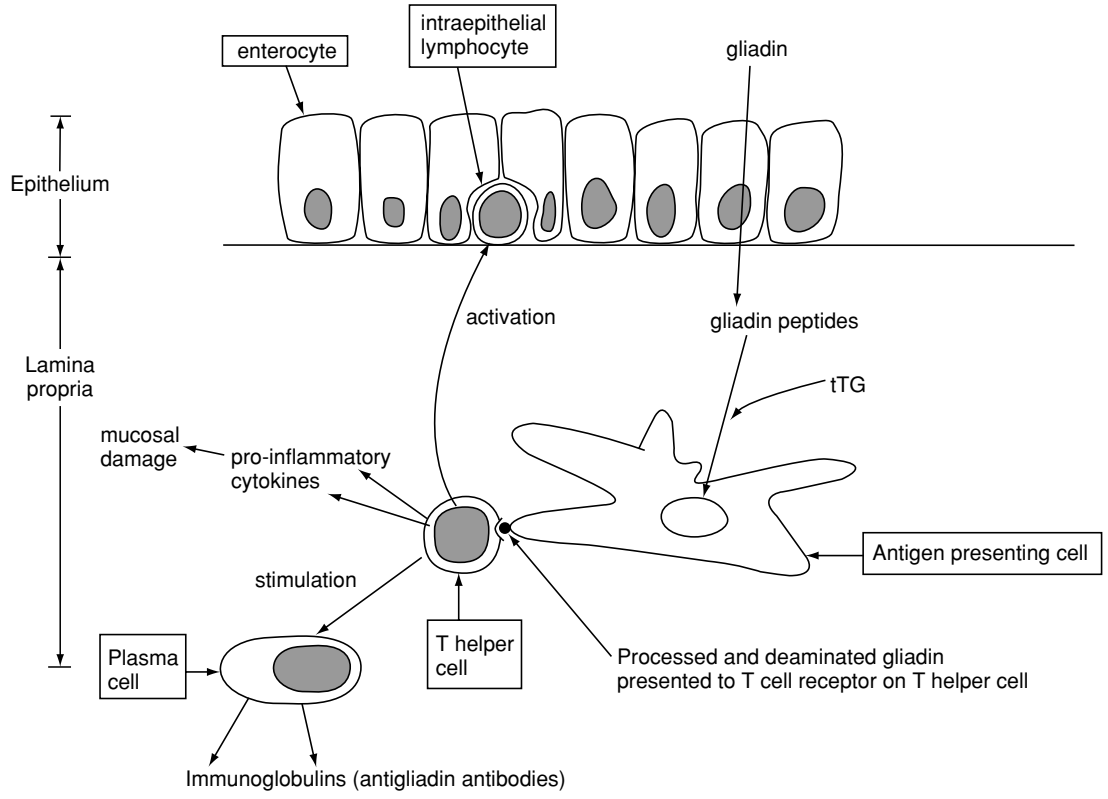

Figure 2 Schematic representation of the immune mechanisms leading to the mucosal abnormalities in celiac disease.

HLA-DQ2 genes are proteins located on the membrane of antigen-presenting cells (e.g., macrophages in the lamina propria). These HLA molecules form a groove in which short peptides (i.e., digested gliadins) can be specifically bound. Deamination by tTG may unmask neo-antigens, which bind more effectively to the HLA-DQ2 molecules. The bound peptides are presented to the T-cell receptor on the CD4 lymphocytes, and the lymphocytes are activated to proliferate and to produce pro-inflammatory cytokines. This promotes the inflammatory cascade and leads to an immune reaction, with resulting tissue damage.

Mucosal pathology

The normal appearance of the small intestinal mucosa has finger-like villi (see Figure 3a), which, on microscopic sections, appear as tall villi covered with a layer of regular epithelial cells supported by a lamina propria (Figure 3b). In untreated celiac disease, the villi are lost, and the surface looks 'flat' (Figure 4a). Microscopically, there are no villi, the crypts between the villi are deeper, and the lamina propria has a vastly increased number of inflammatory cells (Figure 4b). Careful studies of serial biopsies have shown that these changes develop in a sequential fashion related to the amount of gluten in the diet. Hence, in celiac patients, one can see a range of abnormalities from the severe, as can be seen in Figure 4, to the normal, as can be seen in Figure 3 when the patient is well treated, or in potential patients before the mucosal abnormalities develop.

Clinical Features

Celiac disease can present at any age, in both adults and children. The clinical features are protean (see Table 1), although since it is a disease affecting the small intestine, symptoms of gastrointestinal disease predominate in classical descriptions of the disease. Patients may have diarrhoea, anorexia, nausea,

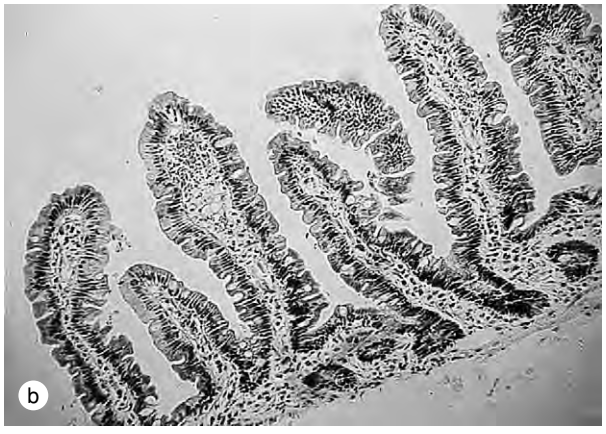

Figure 3 (see color plate 22) (a) Dissecting-microscopic appearance of normal mucosal biopsy showing finger-like villi; (b) light-microscopic appearance of normal mucosal biopsy.

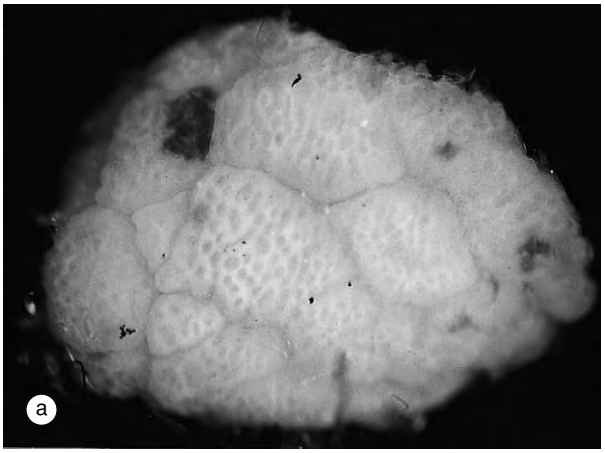

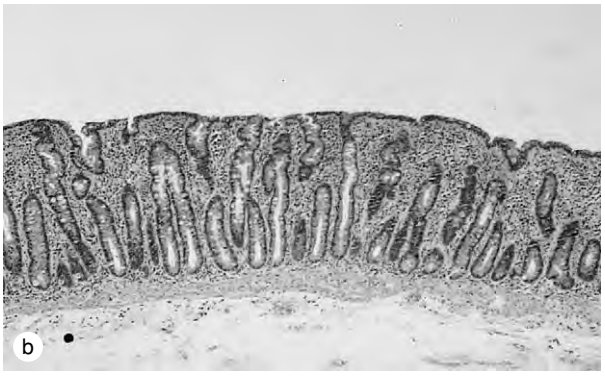

Figure 4 (see color plate 23) (a) Dissecting-microscopic appearance of biopsy from untreated celiac patient with total absence of villi and a 'flat'-looking mucosal surface; (b) light-microscopic appearance of untreated celiac mucosa with absence of villi, deep crypts, and an increase in inflammatory cells.

vomiting, abdominal distention and pain, flatulence, a sore tongue, or mouth ulcers. As well as these gastrointestinal symptoms, patients commonly have constitutional symptoms of lassitude and malaise, weakness, and weight loss. Children may fail to thrive as infants and be underdeveloped, not reaching their expected height. Apart from these general symptoms, there are symptoms that arise as a result of poor absorption of various vitamins and minerals due to the small intestinal mucosal abnormality, which causes 'malabsorption'. Such symptoms include features of anemia (malabsorption of iron and/or folic acid), a bleeding tendency (vitamin K) cramps, paraesthesiae, rickets, and proximal muscle weakness (vitamin D). Finally, there are rare symptoms that may result from the chronicity of the disease or malabsorption of undiscovered factors; these include disorders of the nervous system, reproductive system, and skin.

It is important to stress that many of these symptoms are described in untreated patients with celiac disease, and almost all are reversible with treatment. It is also of note that as physicians have become more aware of the disease, the diagnosis is being made more quickly, and patients nowadays do not present with such gross features. In fact, the majority of patients are diagnosed with mild symptoms and minor biochemical or hematological abnormalities. Recently, it has become apparent that previously unrecognized osteopenia and osteoporosis are present in a significant minority of celiac patients.

When untreated patients are examined, the findings on physical examination reflect the symptoms referred to above. For example, a patient with major symptoms may be thin and wasted, with muscle weakness, bruising of the skin, a sore tongue, and a distended abdomen. Patients diagnosed at an early stage may be apparently normal on physical examination.

Table 1 The presenting symptoms encountered in patients with celiac disease

General	Lassitude, malaise, weakness, weight loss
Gastrointestinal	Diarrhea (± features of steatorrhea)
	Anorexia
	Nausea and vomiting
	Flatulence and abdominal distention
	Abdominal pain
	Constipation
	Glossitis/aphthous ulcers
Metabolic	Features of anemia
	Bleeding tendency
	Edema
	Cramps/tetany
	Paresthesiae
	Nocturnal diuresis
Musculoskeletal	Increased risk of fracture
	Bone pain
	Rickets
	Proximal myopathy
Neuropsychiatric	Peripheral neuropathy
	Cerebrospinal degenerations
	Anxiety/depression
Reproductive	Menstrual irregularities
	Reduced fertility (male and female)
	Relapse associated with pregnancy
Skin	Variety of rashes
	Dermatitis herpetiformis

Associated Disorders

Many disorders have been described to occur in association with celiac disease. The explanation may be that the underlying predisposition to celiac disease enhances the possibility of other disorders occurring, particularly those that have an immunological etiology. The most common of the associated disorders are shown in Table 2.

Diagnostic Tests

Hematological investigations

The majority of untreated patients are anemic. Anemia is caused by iron or folic acid deficiency, or both, and is a result of the malabsorption of these factors. Vitamin B_{12} may be low but rarely causes the features of B_{12} deficiency. Unexplained anemia may therefore alert the physician to the possibility of celiac disease.

Biochemical investigations

Similarly, other nutritional factors are often low in the serum of untreated celiac patients, e.g., the fat-soluble vitamins A, E, D, and K, deficiency of the latter two being responsible for the features of osteomalacia and a bleeding tendency referred to above. Calcium and magnesium, as well as water-soluble vitamins such as Vitamin C, may be low, although scurvy is rare in celiac disease. Albumin is often low, owing to loss into the gastrointestinal tract from the

Table 2 Disorders known to be associated with celiac disease

Insulin-dependent diabetes mellitus
Autoimmune liver disease
Thyroid disorders
Pulmonary alveolitis
Inflammatory bowel disease
Connective-tissue disorders
Immunoglobulin A (IgA) deficiency
Down's syndrome
Osteoporosis
Epilepsy

blood, rather than malabsorption. This can lead to peripheral edema in the patient. Immunoglobulin A (IgA) deficiency may be found since this is 10 times more common in celiac disease than in the general population. None of these measurements is specific for celiac disease, and therefore, they are not used in diagnosis.

Tests of malabsorption

Apart from finding low levels of various nutrients in the serum, as outlined above, absorption can be assessed more formally. For example, the absorption of fat can be assessed by measuring the dietary intake and fecal output over a period of days. A more indirect method measures the serum levels and urinary excretion of D-xylose, a carbohydrate not normally found in the diet. Such tests can indicate malabsorption but are not specific for celiac disease and are rarely used routinely.

Serological screening tests

IgA and immunoglobulin G (IgG)-class antigliadin antibodies are present in the serum of many patients with celiac disease. Their measurement has been used particularly as a screening test. The sensitivity and specificity of these tests approach 90%. False positive results limit their use as a definitive diagnostic test.

IgA-class anti-endomysial antibodies are also found. As outlined above, tTg is the auto-antigen responsible for anti-endomysial antibody positivity. This antibody test has more than 95% sensitivity and specificity in celiac disease and thus is an ideal screening test. It cannot, however, be regarded as definitely diagnostic.

tTg itself has been used as the substrate in an enzyme-linked immunosorbent assay (ELISA) test, and tTg antibodies are now coming into use as another serological screening test for celiac disease, with a similar accuracy to anti-endomysial antibodies.

These serological tests are useful as screening tests in case-finding studies, particularly in at-risk groups such as patients with disorders associated with celiac disease, or first-degree relatives. They are also useful for follow-up of patients to monitor treatment and to indicate the timing of follow-up intestinal biopsies, since the tests usually become negative when treatment is being effective (see below).

Small-intestinal mucosal biopsies

Celiac disease is defined as that disease in which there is an abnormality of the small intestinal mucosa, made manifest by contact with the gluten fraction of wheat and related cereals. The cornerstone of diagnosis must therefore be the examination of a biopsy of this mucosa. Such a biopsy can be obtained using a biopsy tube or fiber-optic endoscope, which the patient swallows, usually under sedation. Careful examination of the biopsy will reveal the typical pathological features, although these can vary in severity. Since there are other, rarer causes of the same abnormality, the diagnosis is only definite when a second biopsy is obtained after a period (usually 6–12 months) on a gluten-free diet, and the abnormality has been shown to improve and return towards normal (**Figures 3** and **4**). If such an improvement in the pathological features is seen, the diagnosis of celiac disease can be firmly made. Such a scheme for diagnosis is suggested in the definition of the disease, since this indicates that the mucosal abnormality must be related to the presence of gluten in the diet.

Management

Treatment: a gluten-free diet

The cornerstone of treatment is a gluten-free diet, which should be taken for life. Patients should omit foods containing wheat, rye, and barley from their diet. In severely symptomatic untreated patients, initial treatment may also include replacement of nutritional deficiencies such as iron, folate, calcium, and Vitamins D and K. Rarely, severely ill patients may require parenteral nutrition.

Once treatment with a gluten-free diet is established, patients begin to respond quickly and usually note a remarkable improvement within weeks. Patients vary in their sensitivity to gluten in the diet. Some can eventually consume some gluten and appear to remain clinically well; others have to adhere strictly to their diet, even to the extent of avoiding Communion wafers. They may need to omit oats from their diet. If patients do consume gluten, they can produce some damage to the small intestinal mucosa, which may take months or years to produce clinical symptoms. There is evidence that the malignancy that can occur in celiac disease can be reduced by strict adherence to a gluten-free diet.

Osteopenia will improve with a gluten-free diet, but patients should have an adequate calcium intake and take regular exercise.

If patients are to adhere to a gluten-free diet, they need help from an experienced dietitian. Some patients have problems once dietary treatment has started. Initially, some put on weight, presumably as a result of improved absorption. These patients need advice about weight reduction. In this situation, diabetes may be unmasked.

Constipation sometimes becomes a problem, reflecting the fact that bran from most cereals is omitted from the diet. Defatted rice bran or soya bran has

been shown to help, and bulking agents such as ispaghula husk or methyl cellulose may also help. The dietitian can help with all these problems and advise about a diet based around fresh meat, vegetables, fish, and fruit, with the gradual introduction of other safe foods as the patient becomes more confident about the gluten-free diet.

Prognosis

The prognosis is good for celiac patients. Nearly all respond clinically and pathologically to treatment with a gluten-free diet, there being an improvement in symptoms and in the small intestinal mucosa. Lifelong treatment with the diet is recommended, as is follow-up in an out-patient clinic with the help of a dietitian. Younger patients, particularly teenagers, need sympathetic follow-up since they often find the restrictions of the diet quite onerous.

Some physicians recommend further intestinal biopsies (after the first two that are necessary for diagnosis) every few years. This provides an objective assessment of the patient's condition and helps to motivate patients to maintain a strict diet.

Occasionally, the mucosa does not respond to a gluten-free diet, and the physician must consider the other, much rarer, causes of an abnormal small intestinal mucosa (see Table 3). However, by far the most common cause of failure to respond to the gluten-free diet is inadequate dietary gluten exclusion. This may be conscious or inadvertent. Dietary compliance needs to be assessed by a well-informed dietitian. Rarely, other foods (e.g., milk, eggs and soya) need to be excluded as well as gluten, either temporarily or permanently. Other causes preventing a mucosal response to gluten exclusion may be pancreatic insufficiency, or contamination of the small intestine by bacterial overgrowth. Such causes should be sought and treated in this situation. Finally, nonresponse may be associated with serious complications of the disease, which, fortunately, occur only rarely. These are the development of ulcers and strictures in the small intestine and also malignancy, particularly small intestinal lymphoma. Such complications may be a cause not only of nonresponse but also of deterioration in a previously well patient who has been adequately treated for some time with a gluten-free diet. As already suggested, there is good evidence that the incidence of malignancy in celiac disease may be reduced in patients who maintain a strict gluten-free diet.

If no underlying cause is found for nonresponse to a gluten-free diet, and the diagnosis has been carefully considered, patients may be treated with oral steroids in order to induce a remission. This should only be done under careful supervision, and there are no long-term studies, but the side-effects of steroid therapy suggest that such a course of treatment has a limited role in celiac disease.

Table 3 Causes of an abnormal small-intestinal mucosa

Celiac disease (CD), including:
Non-responsive CD
Complications of CD (ulceration, malignancy)
Cow's milk protein intolerance
Soya protein intolerance
Immunodeficiency syndromes
Eosinophilic gastroenteropathy
Immunoproliferative small-intestinal disease (IPSID)
Protein-energy malnutrition
Intractable diarrhea of infancy
Gastroenteritis in children
Infections:
Parasites, e.g., giardiasis
Tuberculosis
Human immunodeficiency virus
Contaminated bowel syndrome
Whipple's disease
Tropical sprue
Arterial disease of small intestine
Drug and radiation damage

See also: **Anemia (Anaemia)**: Iron-deficiency Anemia; **Calcium**: Properties and Determination; **Cereals**: Contribution to the Diet; **Folic Acid**: Properties and Determination; **Immunology of Food**; **Iron**: Properties and Determination; Physiology; **Magnesium**; **Vitamin K**: Properties and Determination; **Vitamins**: Overview; **Wheat**: The Crop

Further Reading

Dieterich W, Ehnis T, Bauer M *et al.* (1997) Identification of tissue transglutaminase as the autoantigen of coeliac disease. *Nature Medicine* 3: 797–801.

Feighery C (1999) Coeliac disease. *British Medical Journal* 319: 236–239.

Hin H, Bird G, Fisher P *et al.* (1999) Coeliac disease in primary care: case finding study. *British Medical Journal* 318: 164–167.

Holmes GKT and Catassi C (2000) *Coeliac disease*. Health Press, Oxford.

Holmes GKT, Prior P, Lane MR *et al.* (1989) Malignancy in coeliac disease – effect of a gluten-free diet. *Gut* 30: 333–338.

Howdle PD (1995) *Coeliac Disease. Baillière's Clinical Gastroenterology*, vol. 9. London: Baillière Tindall.

Janatuinen E, Pikkarainen PH, Kemppainen TA *et al.* (1995) A comparison of diets with and without oats in adult coeliac disease. *New England Journal of Medicine* 333: 1033–1037.

Maki M and Collin P (1997) Coeliac disease. *Lancet* 349: 1755–1759.

Marsh MN (1992a) *Coeliac Disease*. Oxford: Blackwell Scientific.

Marsh MN (1992b) Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). *Gastroenterology* 102: 330–354.

Trier JS (1998) Diagnosis of celiac sprue. *Gastroenterology* 115: 211–216.

CELLS

P McGrath, University of Newcastle, Callaghan, Australia

Structure and Function of Human and Animal Cells

The cell is the fundamental functional unit of all living organisms. Cells perform all of the body's vital physiological functions. Cells are ultimately responsible for activities such as breathing, moving, combating infection, digesting food, thinking – in fact, all bodily functions. These activities require energy. It is inside cells that fuels are used to produce the energy needed to drive these activities. This article examines the structure inside the cell and the combined activities of cells that make all of this possible.

Structure and Organization of Human Animal Cells

Structure

Common to nearly all cells are three compartments, namely the cell membrane (also called the plasma membrane), the cytoplasm, and the nucleus (mature, human red blood cells have no nucleus). This basic structure is evident in a wide variety of cells even though they may differ dramatically in size and shape (**Figure 1**). The plasma membrane not only encloses the contents of the cell but performs many vital functions. The cytoplasm is that compartment between the nucleus and the cell membrane. It is composed of a soluble component (the cytosol) in which are immersed various microscopic components called organelles. Among the many important biochemical reactions that take place in the cytosol is glycolysis, an energy-releasing process involving the breakdown of glucose to pyruvic acid. (*See* **Glucose:** Function and Metabolism.)

The cellular organelles are a diverse group of structures that perform many different functions within the cell. The nucleus of the cell is a membrane-bound structure that contains within it the genetic material of the cell. It is here that the genes reside in the form of DNA (deoxyribonucleic acid) molecules grouped into structures called chromosomes.

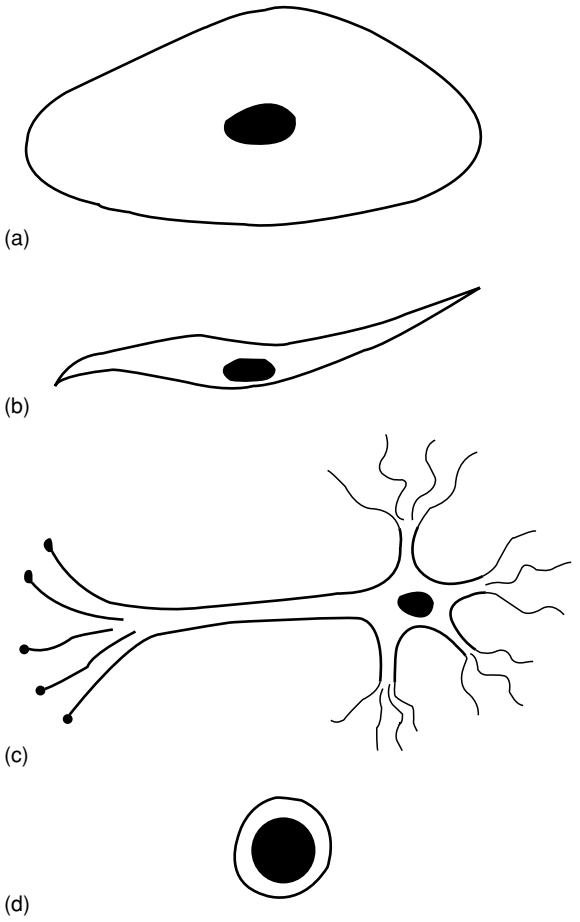

Figure 1 Different shapes of cells. The nucleus is darkly shaded; the plasma membrane is represented by the outline of the cell; the cytoplasm is represented by the clear space between the plasma membrane and the nucleus. (a) Epithelial cell; (b) smooth muscle cell; (c) neuron; (d) lymphocyte. Reproduced from Cells, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Organization

In the body, cells often function in combination with other cells of the same type and are held together with varying degrees of tenacity and with varying amounts of extracellular material between them. Clusters of cells of the same type are called tissues. Four different types of tissues are recognized:

1. Epithelial tissue, in which the cells are held very tightly together, and there is very little extracellular material between them. This tissue forms glands and the linings of the body cavities and surfaces.
2. Connective tissue, in which the cells are not bound closely to one another. There is much extracellular material in such tissue, and the cells may be held rigidly in place by this material, as in the case of bone, or they may be freely moving, as in blood.
3. Muscle tissue, in which the cells may be bound to each other, as in cardiac muscle and smooth muscle, or separate from each other but gathered into bundles by connective tissue, as in skeletal muscle. (*See* **Exercise**: Muscle.)
4. Nervous tissue, which is composed of neurons (nerve cells) that conduct electrical impulses from one neuron to another. These cells are not physically coupled but are held in place by specialized connective tissue cells.

The organs of the body are in turn composed of different types of tissues. For example, the heart is composed mainly of muscle tissue but also includes connective tissue, epithelial tissue, and some nervous tissue.

Cellular Components

There are numerous chemical activities occurring simultaneously inside the cell. Eukaryotic cells (cells that have membrane-bound compartments) are able to minimize interference of one chemical reaction with another by confining certain chemical activities to specific membrane-bound components of the cell. This design also facilitates cellular specialization. Prokaryotic cells (cells that lack membrane-bound compartments) such as bacteria are disadvantaged in this regard. These membrane-bound components and other specific structures inside the cell are called organelles. Although some of these (e.g., the nucleus) can be seen under the light microscope, it has been necessary to employ electron microscopy to clearly visualize most of them. Not all of the organelles are found in every cell, and the relative abundance of a particular type of organelle varies from cell to cell, depending on the activity of the cell. Figure 2 is a drawing of a generalized cell with various organelles represented.

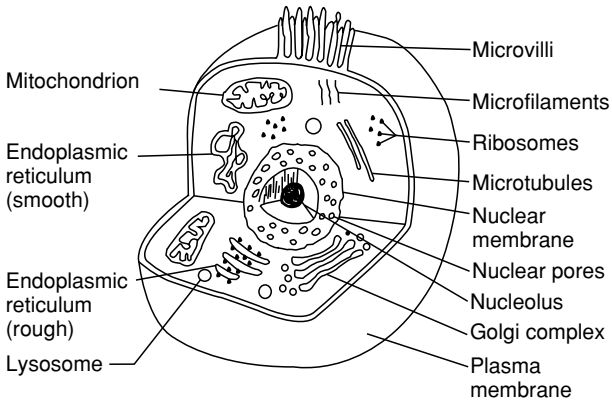

Figure 2 Drawing of a 'generalized' cell showing many of the organelles found in human and animal cells. Reproduced from Cells, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Nuclear Components

The nucleus contains the chromosomes. In humans, there are 46 chromosomes (23 pairs). The number in other animals varies from species to species. The chromosomes are very long, thin strands of DNA. They also contain proteins called histones. They are not easily seen under the light microscope until just prior to cell division, when they shorten and coil up, making them visible following appropriate staining. The nucleus also contains other structures called nucleoli. These are roughly spherical structures containing strands of RNA (ribonucleic acid) molecules, DNA, and proteins. The nucleoli are responsible for manufacturing rRNA (ribosomal RNA) of the ribosome. The nucleus is enclosed by a membrane that is perforated by holes (nuclear pores) that are large enough to allow large molecules, such as RNA, to pass through. There are several different types of RNA, and each type has a distinct role in the manufacture of proteins (protein synthesis). Protein synthesis takes place in the cytoplasm. (*See* **Nucleic Acids**: Physiology.)

Cytoplasmic Components

There are many organelles or structures within the cytoplasm that are involved with specific cell functions. These are listed below with a brief description of the structure and function of each.

Ribosomes Ribosomes are tiny granules (about 25 nm in diameter) composed of rRNA and protein. A ribosome consists of two subunits, one about twice the size of the other. Ribosomes play an integral role in protein synthesis.

Endoplasmic reticulum The endoplasmic reticulum is actually a series of double-membrane channels distributed throughout the cytoplasm. Ribosomes may be attached to the membranes and are then referred to as the rough endoplasmic reticulum. Without ribosomes attached, it is called smooth endoplasmic reticulum. This organelle serves several important functions in the cell, among which are involvement in the synthesis and storage of molecules, providing a system of channels for distribution and transport of materials throughout the cell, release of calcium ions into the cytosol that initiate contraction in muscle cells, and providing some structural support for the cell.

Golgi complex The Golgi complex consists of flattened membranous sacs stacked upon each other with expanded areas at their ends. The main function of this organelle is in the sorting and packaging of various molecules, particularly proteins, for distribution to various parts of the cell. The Golgi complex is particularly extensive in cells with high secretory activities.

Mitochondria Mitochondria are double membrane–bound organelles that may have a variety of shapes. The inner of the two membranes is thrown into folds or plates, called cristae. This arrangement provides a large surface area for chemical reactions to take place. Many enzymes involved in energy-releasing reactions are located on the cristae. These reactions are collectively known as cellular respiration and include the reactions of the tricarboxylic acid cycle and the electron transport chain. Since much of the cell's energy is produced through these reactions, mitochondria are sometimes called the 'powerhouse' of the cell. Cells that have a high energy expenditure, e.g., muscle, liver, and kidney tubule cells, have large numbers of mitochondria. The fact that mitochondria have some structural similarities with bacteria, have their own DNA, and are capable of reproducing themselves has led some scientists to speculate that mitochondria may have evolved from bacteria. According to this idea, the bacteria were incorporated into the cell during evolution.

Lysosomes Lysosomes are membrane-bound, spherical structures containing powerful digestive enzymes and are formed from vesicles budding off the Golgi complex. These enzymes are capable of digesting bacteria and other solid matter that may be engulfed by a cell. Leucocytes, specifically neutrophils and monocytes that engulf bacteria and other foreign particles (a process known as phagocytosis), contain large numbers of lysosomes.

Cytoskeleton Helping to maintain the shape of the cell and supporting various organelles within the cell is a scaffolding-like assembly of filaments and tubules called the cytoskeleton. The filaments and tubules are composed of proteins similar to those found in the contractile machinery of muscle. They are rod-shaped structures and vary in length and thickness. Microtubules average about 24 nm in diameter and may also provide channels for the transport of materials within the cell. Microfilaments are about 6 nm in diameter and may also play a special role in movement of cells, e.g., movement of phagocytes. Intermediate filaments, diameters ranging from 8 to 12 nm, have also been discovered inside cells.

Plasma Membrane

The plasma membrane (Figure 3) is composed mainly of lipid molecules. The lipid is arranged into two layers, the hydrophobic tail regions of the lipid molecules pointing to the inside of the bilipid layer, and the hydrophilic heads of the lipid pointing to the outside. Proteins and carbohydrates float in this lipid membrane. The plasma membrane facilitates contact and communication with other cells, mediates the entry and exit of materials into and out of the cell, and is the site of many important biochemical reactions.

The plasma membrane is selectively permeable: it will allow the transit of some substances, but not of others, and furthermore, some substances are allowed across the membrane more readily than others (differentially permeable). There are very specific 'channels' in the membrane that allow the passage of specific substances. These 'channels' are in fact proteins.

All of these properties of the plasma membrane contribute to the differences that are seen in the composition of the fluid inside the cell (intracellular fluid) compared with that outside (extracellular fluid). A very important difference to note in terms of ion concentrations is that of potassium and sodium ions. Intracellular fluid has a much higher concentration of potassium ions than the extracellular fluid. The reverse is true of sodium ions. It is extremely important to the function of all cells in the body that these concentration differences (or gradients) are maintained. At times, the concentration differences will be disturbed. The role of certain proteins found in the plasma membrane is to act as ion pumps to restore the original concentration gradients. Of particular importance is the so-called sodium/potassium pump. When the concentration gradient of sodium and potassium ions is decreased, these pumps expend energy to pump sodium ions out of the cell and at the same time pump potassium ions into the cell. The importance of this mechanism to normal cell function

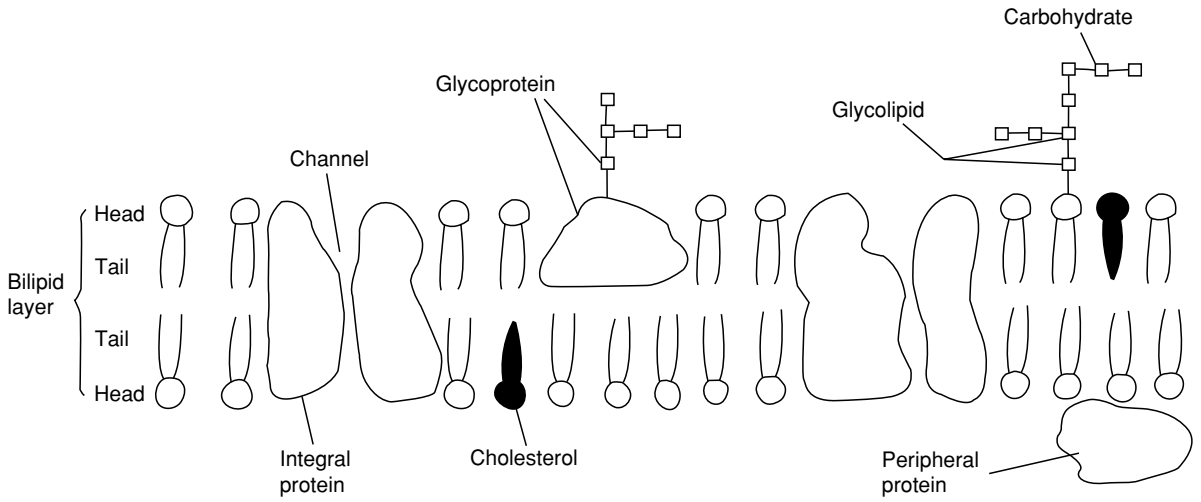

Figure 3 Arrangement of molecules in the plasma membrane. Reproduced from Cells, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

is markedly demonstrated by the nerve cell (the neuron). In order to conduct electrical impulses, sodium ions must enter the cell and potassium ions move out. The sodium/potassium pump works vigorously in the neuron to try to maintain the original gradients. Indeed, the neuron expends 80% of its energy on this pump.

These properties of the plasma membrane contribute not only to chemical differences between the inside and the outside of the cell but also to an electrical difference. A small voltage (of the order of tens of millivolts) can be measured across the plasma membrane. The inside of the cell is electrically negative with respect to the outside. This electrical difference, or membrane potential as it is called, is particularly important when considering the functions of neurons and muscle cells, both of which can conduct electrical impulses.

The cell membrane in some cells is thrown into many small folds that project from the bulk of the cell and are called microvilli. These greatly increase the surface area of the cell and provide a much larger surface for the absorption of materials across the cell membrane. Such specializations are seen in cells lining the small intestine and in tubule cells in the kidneys. In some cells, similar types of projections of the plasma membrane are seen with underlying contractile machinery of the cytoplasm. These form cilia. The cilia move in unison, providing a sweeping motion that moves material over the surface of the cell. Cells bearing cilia line many of the larger respiratory passages, and the mucus produced here traps dust and foreign particles and is swept up to the throat region by this mucociliary escalator.

Requirement of Cells

For cells to carry out their normal functions properly, certain conditions need to apply. For human and most animal cells, environmental conditions such as temperature, pH, and ionic strength of solutions bathing the cell need to be kept fairly constant. Many of the body's mechanisms are geared towards this homeostatic control. Cells also need to be supplied with nutrients. They need the basic building blocks to manufacture complex organic molecules, such as amino acids for the synthesis of proteins. They need fuels to supply the substrates for generating energy, and oxygen to oxidize the fuels and allow energy to be released. The fuel requirements of some cells are very exacting. While several types of fuel molecules may be available to cells, neurons need to use glucose almost exclusively. If the glucose cannot be supplied to these cells, the brain will stop functioning properly. Such is the importance of glucose to neurons that many body mechanisms operate to maintain adequate levels of glucose in the blood at all times. Other fuels used by cells are fatty acids and, to a limited extent, proteins. When glucose is in short supply, many cells turn to a greater utilization of fatty acids to derive energy. (*See* **Fatty Acids**: Metabolism; **Protein**: Synthesis and Turnover.)

Importance of Specialized Cells

While all cells are able to (and need to) carry out certain basic functions, such as protein synthesis and energy metabolism, many cells develop special intracellular machinery to carry out specific functions. The

complex behavior of animals is possible only because of this cellular specialization. All somatic cells (body cells, as opposed to the gametes, or sex cells) contain the full complement of genetic information. With cellular specialization, however, the capacity to develop in a variety of ways (sometimes called cellular plasticity) is sacrificed. In some cases, cells become so constrained and devoted to their special task in the body that they even lose the ability to divide and provide more of the same type. Such is the case with neurons; these cells make up the main controlling and information-transfer system of the body. Neurons detect various environmental stimuli and pass this information on to centers in the nervous system that interpret this information. Having determined the appropriate response, the nervous system sends messages to the muscles and glands of the body to effect this response. All this is achieved by transmitting electrical signals around the body. When neurons are injured and die, they are not replaced. Muscle cells, too, are incapable of dividing. These cells devote their energies to building contractile machinery within them. When these cells are activated, they contract, bringing about movement of various organs, limbs, etc., of the body. Adipocytes, or fat cells, are specialized to store large amounts of fat to such an extent that the nucleus of the cell becomes confined to a small part of the cell squashed against the cell membrane. It is thought that mature adipocytes do not divide. Erythrocytes are packed with the protein, hemoglobin, which is responsible for transporting the blood gases, particularly oxygen. It enhances the oxygen-carrying capacity of the blood about 60-fold. Human erythrocytes are peculiar cells in that they have no nucleus, losing their nucleus during maturation. They have a limited lifespan of about 120 days. Under normal circumstances human erythrocytes are produced in the body at a rate of 2×10^6 per second. Hepatocytes are specialized cells of the liver. These cells contain large amounts of specific enzymes involved in the metabolism and detoxification of many different molecules in the body. These are just some of the many specialized cells in the body; there are numerous others, including osteocytes (bone cells), endocrine cells (hormone-producing cells), and chondrocytes (cartilage cells). (*See* **Adipose Tissue**: Structure and Function of White Adipose Tissue; Structure and Function of Brown Adipose Tissue.)

See also: **Adipose Tissue**: Structure and Function of White Adipose Tissue; Structure and Function of Brown Adipose Tissue; **Exercise**: Muscle; **Fatty Acids**: Metabolism; **Glucose**: Function and Metabolism; **Fats**: Classification; **Nucleic Acids**: Physiology; **Protein**: Synthesis and Turnover

Further Reading

de Duve C (1984) *A Guided Tour of the Living Cell*, vols. 1 and 2. New York: Scientific American Library.

Fawcett DW (1981) *The Cell*, 2nd edn. Philadelphia, PA: WB Saunders.

Ganong WF (1999) *Review of Medical Physiology*, 19th edn. Upper Saddle River: Prentice Hall.

Keeton WT and Gould JL (1986) *Biological Science*, 4th edn. New York: WW Norton and Co.

Martini FH (1998) *Fundamentals of Anatomy and Physiology*, 4th edn. Upper Saddle River: Prentice Hall.

Reid RA and Leech RM (1980) *Biochemistry and Structure of Cell Organelles*. Glasgow: Blackie.

Tribe MA, Morgan AJ and Whittaker PA (1981) *The Evolution of Eukaryotic Cells*. London: Edward Arnold.

CELLULOSE

M T Holtzapple, Texas A & M University, TX, USA

Background

Cellulose is the world's most abundant biological material. An estimated 10^{11} tonnes are produced annually, a daily production of about 45 kg per person. Because of its rigidity, cellulose provides structure to plants. It is almost always associated with other components as a composite material termed 'lignocellulose.' (*See* **Hemicelluloses**; **Lignin**.)

The most-studied celluloses are from cotton, ramie, *Valonia* algae and *Acetobacter xylinum* bacteria.

Structure

Polymer Structure

Cellulose is a polymer of the monomer glucose (see **Figure 1a**). The free-aldehyde glucose conformation is very unstable, so it cyclizes into a six-membered pyranose ring. (In aqueous solutions at 25 °C, only about 0.02% of the glucose is in the free aldehyde form.) The C1 hydroxyl of the cyclic ring can be in the equatorial β

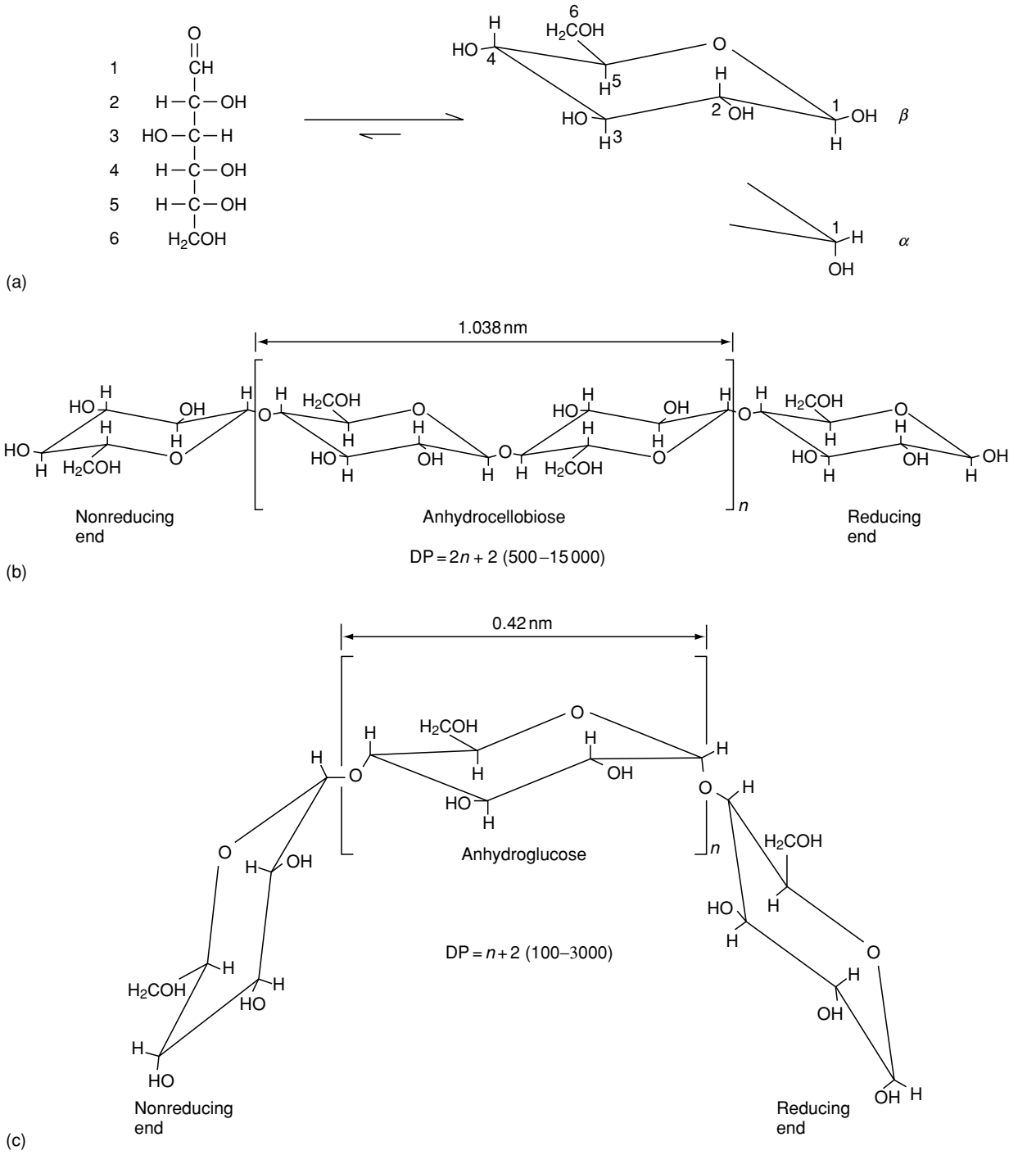

Figure 1 (a) Glucose, (b) cellulose, and (c) amylose starch. Reproduced from Cellulose, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

position or the axial α position. The β position is favored thermodynamically and accounts for 62% of the glucose with the remaining 38% in the α position.

Cellulose is a linear, unbranched polymer of anhydroglucose joined with ether linkages between C1 and C4 (see Figure 1b). Cellulose is polymerized from β-glucose, whereas starch (see Figure 1c) is polymerized from α-glucose. In the rightward anhydroglucose unit, C1 retains its hydroxyl group, so it may potentially form free-aldehyde glucose, which has reducing power. Hence, this terminus is called the 'reducing end.'

β-Linked cellulose has dramatically different properties than α-linked starch. Starch is helical,

water-soluble, and easily hydrolyzed by enzymes, whereas cellulose is planar, water-insoluble, and difficult to hydrolyze. Whereas starch is a widely utilized animal food, cellulose is used only by a select few, ruminants being the most prominent example. In starch, the repeating unit is anhydroglucose. For cellulose, anhydrocellobiose is the repeating unit because adjacent anhydroglucoses are rotated 180° with respect to their neighbors. This rotation causes cellulose to be highly symmetrical, because each side of the chain has an equal number of hydroxyl groups, whereas starch is unsymmetrical. The cellulose degree of polymerization (DP) ranges from 500 to 15 000. When fully stretched, a single cellulose molecule could extend as long as 7 μm. Amylose starch has a lower DP (100–3000) and could extend as long as 1 μm if fully stretched. Cellulose is completely unbranched, whereas amylopectin starch has branches at C6. (*See* **Starch**: Structure, Properties, and Determination.)

Crystalline Structure

The cellulose hydrogen atoms are all in the axial position, whereas the hydroxyl groups are all equatorial. These equatorial hydroxyl groups can hydrogen-bond with their nearest neighbors, allowing cellulose to crystallize. The monoclinic crystalline unit cell for cellulose I (native cellulose) is shown in Figure 2. The hydrogen bonds run in the *a* direction and are medium-strength (15 kcal mol^{-1}). In the *c* direction, the structure is held by weak van der Waals forces (8 kcal mol^{-1}). Covalent bonds run in the *b* direction and give cellulose its strength (50 kcal mol^{-1}). A continuous cellulose strand is about four to

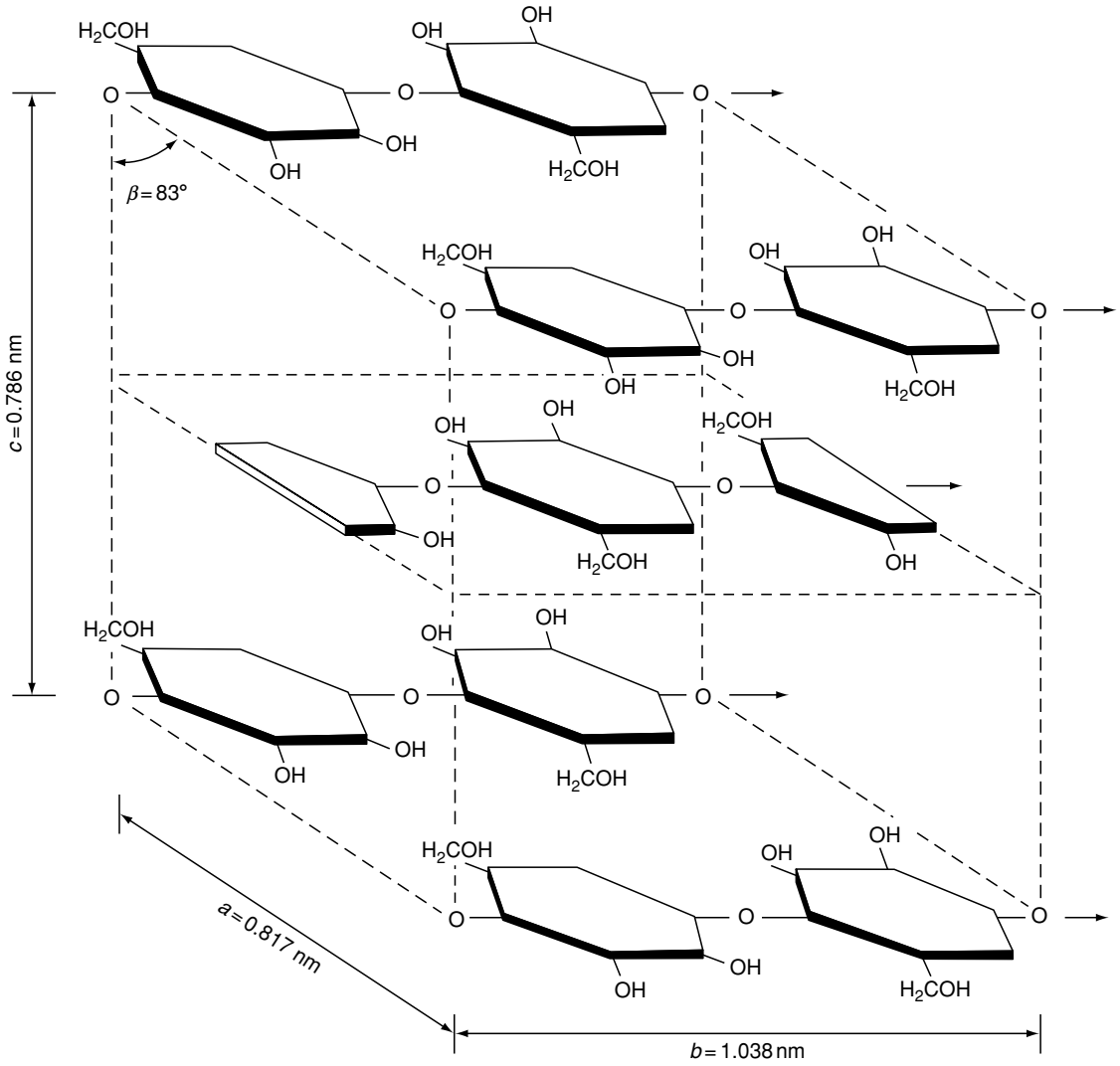

Figure 2 Parallel cellulose I unit cell. Reproduced from Cellulose, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

five times stronger than steel with the same cross-section. Cellulose I is parallel; that is, all the cellulose molecules run in the same direction from nonreducing to reducing ends (see Figure 3a).

Native cellulose (cellulose I) can be converted to other crystalline forms. Cellulose II is formed by (1) treating cellulose with sodium hydroxide (mercerization), (2) precipitating from solutions of alkali/salt (e.g., cuprammonium hydroxide), or (3) removing the added functional groups from cellulose derivatives (i.e., regenerated cellulose). Cellophane and rayon are both forms of cellulose II. Table 1 shows that the unit cell dimensions are slightly expanded in the *c* direction and compressed in the *a* direction. Of course, the *b* direction is essentially the same, because it is the covalent bond. Cellulose II is the most thermodynamically stable form of cellulose because it can always be produced from cellulose I, but not vice versa. The stability may result from hydrogen bonds extending in the *c* direction, which normally has only van der Waals bonds. There is general agreement that cellulose II is antiparallel (see Figure 3b) with three to four anhydroglucose moieties required to make the bend. Precipitation of cellulose II from solution appears to favor the antiparallel conformation, as occurs with many synthetic polymers.

Cellulose III is formed by soaking cellulose in cold (about $-80\,^{\circ}C$) liquid anhydrous ammonia, which is subsequently removed by evaporation. Cellulose I is transformed into cellulose III_1 and cellulose II is transformed into cellulose III_2. When rehydrated, cellulose III reverts back to its original form.

Cellulose IV is formed by soaking cellulose in hot (about $200\,^{\circ}C$) glycerol, with subsequent removal by washing with 2-propanol and water. Cellulose I is transformed to cellulose IV_1, and cellulose II is transformed to cellulose IV_2.

Native cellulose (Figure 3a) forms crystalline regions (40% bacteria, 60% cotton, 70% *Valonia*) interspersed with amorphous regions. The amorphous regions are more porous than the crystalline regions, allowing water or dyes to penetrate and increasing the reactivity to acid or enzymatic hydrolysis. When purified cellulose fibers are subjected to dilute acid hydrolysis, the amorphous regions selectively hydrolyze, leaving the more resistant crystalline regions, which have a 'levelling-off DP' of 100–300 in the case of cotton.

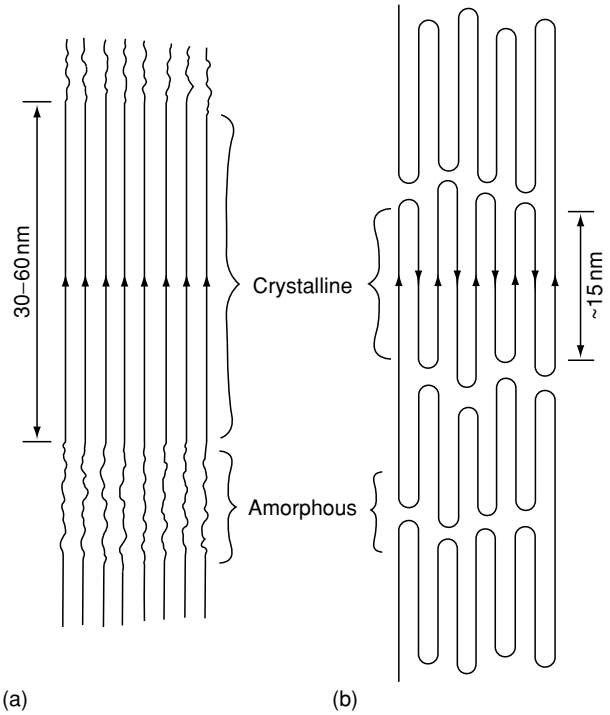

Figure 3 (a) Cellulose I parallel and (b) cellulose II antiparallel structures. Reproduced from Cellulose, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 1 Unit cell dimensions of cellulose I and II

Cellulose	*a (nm)*	*b (nm)*	*c (nm)*	β *(°)*
I	0.817	1.038	0.786	83.0
II	0.801	1.036	0.904	62.9

From Blackwell J, Kurz D, Su M-Y and Lee DM (1987) X-ray studies of the structure of cellulose complexes. In: Atalla RH (ed.) *The Structures of Cellulose. ACS Symposium Series*, No. 340, pp. 199–213. Washington, DC: American Chemical Society.

Cellular Structure

Young plant cells have only the primary wall to resist osmotic pressure. This wall is composed principally of hemicellulose and pectin with small amounts of cellulose and protein and is sufficiently flexible to accommodate cell growth. When cell growth ceases, the secondary wall is formed in three layers (S_1, S_2, and S_3) with S_2 the thickest (see Figure 4). (*See* **Protein**: Chemistry.)

All cell wall layers are composed of 7–30-nm-diameter microfibrils, which are visible using an electron microscope. In the primary cell wall, the fibrils are randomly oriented. In the S_1 layer, the microfibrils are oriented helically, with each successive layer alternating between right-handed and left-handed helices. In the S_2 layer, which provides much of the plant strength, the microfibrils are arranged in steep right-handed helices nearly oriented along the cell axis. The S_3 layer is another shallow helix. All higher plants (softwoods, hardwoods, herbaceous) are thought to have cell structures similar to that shown in Figure 4.

The plant cell diameters range from 15 to 80 μm, depending on the species and time of year. Spring cell

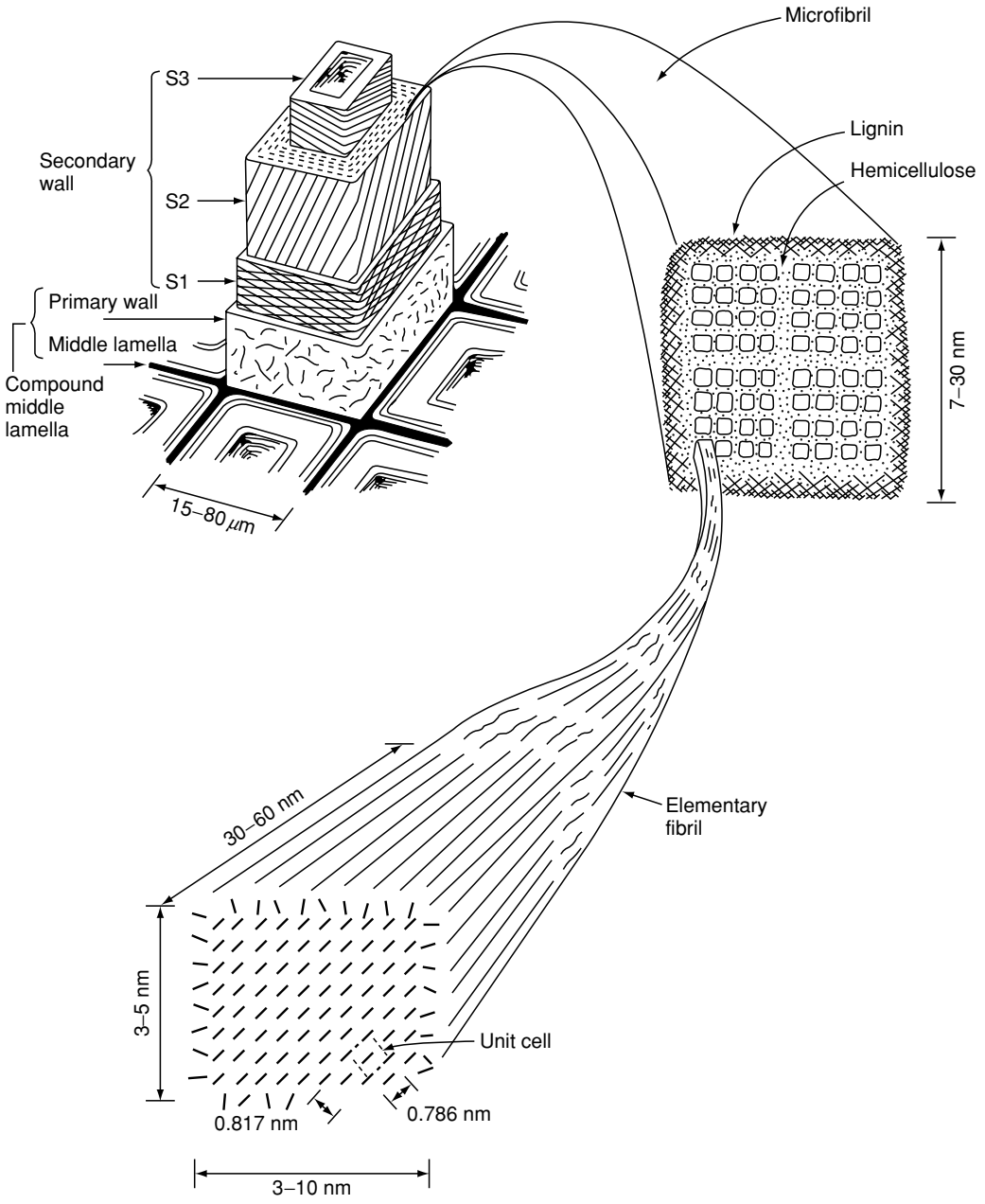

Figure 4 Schematic of plant cell, microfibril, and elementary fibril. Reproduced from Cellulose, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

diameters are larger than summer diameters, giving rise to the familiar growth rings in trees. The wall thickness is typically about 2–5 μm, with thicker walls formed in the summer. Softwood cells are about 3–4 mm long, which makes them particularly useful for paper pulp, whereas hardwood cells tend to be shorter (0.7–2 mm long). Cotton cells suitable for textiles are about 25 mm long.

Microfibrils are composed of elementary fibrils of pure cellulose embedded in a matrix of hemicellulose. The fringes of the elementary fibrils are paracrystalline and intermingle with tightly adsorbed hemicellulose. (An exception to this is cotton in which the fringes contain only cellulose.) In Fengel's model of the microfibril, the elementary fibrils are clustered into four 4×4 arrays. The lignification process occurs late in cell life, so lignin is located primarily on the microfibril exterior, where it covalently bonds to the hemicellulose. Lignification is initiated in the middle lamella where lignin constitutes about 70% of

the wall. The wall interior is less lignified, with only about 15% lignin. The hemicellulose content is fairly constant (20–30%), with cellulose the remaining 10–50% of the wall. The cells in fruit pulp are composed mainly of primary walls containing approximately 34% pectin, 24% hemicellulose, 23% cellulose, and 19% protein.

Properties

Physical

Table 2 shows the physical properties of cellulose and cellulosic materials. The listed heat of combustion is the 'higher heat' (i.e., the combustion water is 25 °C liquid). Cellulose is stable up to 300 °C.

The DP for various types of cellulose is shown in Table 3. For cotton, the DP may be as high as 15 000. Cellulose is insoluble in water because of the strong hydrogen bonding in its crystal lattice. However, it is soluble in a number of solvents, including concentrated acids (e.g., 85% phosphoric, 72% sulfuric, and 40% hydrochloric acids) and inorganic salt solutions (e.g., cuprammonium hydroxide, cadoxen (i.e., cadmium ions and ethylenediamine)). The viscosity of cellulose dissolved in salts is often used for molecular weight determinations.

Chemical

Cellulose may be hydrolyzed to glucose by acids or enzymes. Acid hydrolysis produces a number of degradation products, such as 5-hydroxymethylfurfural, formic acid, levulinic acid, formaldehyde, furfural, and resins. Enzymatic glucose production requires a cellulase system containing *endo*-cellulase (to produce nonreducing ends from the chain interior), *exo*-cellulase (to produce cellobiose generally from the nonreducing end), and cellobiase (to hydrolyze cellobiose to glucose). Up to about 10% of cellulose is enzymatically digested by microbes in the human large intestine, so it is not completely noncaloric.

Cellulose is fairly stable to base, provided that oxygen is excluded. The reaction terminates after approximately 50 anhydroglucose units are 'peeled off' from the cellulose reducing end; D-glucoisosaccharinate is the soluble product. Under severe conditions (e.g., 1 M sodium hydroxide, 170 °C), alkaline hydrolysis readily occurs.

Cellulose is more stable to oxidants than lignin, a property exploited in pulp bleaching and analytical methods, which selectively oxidize lignin. However, oxidants (e.g., chromic acid, permanganate, and hypochlorite) can damage cellulose by cleaving the chain or by inserting carbonyl functional groups.

The three cellulose hydroxyl groups are very reactive, allowing ready formation of the cellulose derivatives described later.

Specialty Celluloses

Microcrystalline cellulose (Avicel, FMC Corporation) is prepared by acid hydrolysis of cellulose using 2 M hydrochloric acid at 105 °C for 15 min. The highly reactive amorphous regions selectively hydrolyze, releasing the crystallites, which are subsequently

Table 3 Cellulose polymer length

Material	*Degree of polymerization*	*Molecular weight*
Native cellulose	3500–10 000	600 000–1 500 000
Chemical cottons	500–3000	80 000–500 000
Wood pulps	500–2100	80 000–340 000
Rayon filaments	350–450	57 000–73 000

From Hamilton JK and Mitchell RL (1964) Cellulose. In: Standen A (ed.) *Kirk–Othmer Encyclopedia of Chemical Technology*, 2nd edn., vol. 4, pp. 593–616. New York: Wiley.

Table 2 Physical properties of cellulose and cellulosic materials

Property	*Compressed cellulose*	*Paper*	*Pine*	*Oak*
Specific gravity[a]	1.47	0.70–1.15	0.43–0.67	0.64–0.87
Heat capacity (kJ kg^{-1} K^{-1})[b]	1.3	1.3	2.8	2.4
Heat of combustion (MJ kg^{-1})[b]	17.6	17.6	20.4	19.2
Thermal conductivity (W m^{-1} K^{-1})[b]				
With grain		0.13	0.35	
Across grain		0.13	0.15	0.21
Tensile strength (MPa)[a]				
With grain			62	99
Across grain			2.8	5.5
Water diffusion constant (cm^2 s^{-1})[c]			5.3×10^{-7}	1.9×10^{-7}

[a]From Baumeister T, Avallone EA and Baumeister T, III (1978) *Mark's Standard Handbook for Mechanical Engineers*, 8th edn. New York: McGraw-Hill.
[b]From Graboski M and Bain R (1981) Properties of biomass relevant to gasification. In: Reed TB (ed.) *Biomass Gasification: Principles and Technology*, pp. 41–71. Park Ridge: Noyes Data.
[c]From Summitt R and Sliker A (1980) *CRC Handbook of Materials Science: Wood*, vol. IV, pp. 26–27. Boca Raton, FL: CRC Press.

mechanically dispersed. Aqueous suspensions of microcrystalline cellulose have constant viscosities over a wide temperature range, are heat-stable, and have good mouth-feel properties. Avicel is used to extend starches, stabilize foams, and control ice crystal formation. Avicel has found widespread acceptance in the food industry for meringue, whipped toppings, confections, and ice cream, and is also used as a binder in pharmaceutical tablets and in cosmetics.

Bacterial cellulose (Cellulon, Weyerhaeuser Co.) is produced by selected strains of *Acetobacter xylinum*, which maintain their ability to produce cellulose in agitated submerged fermentors. The cellulose fibers are about 0.1 μm in diameter, which is substantially smaller than softwood pulp fibres (about 30 μm diameter). Cellulon is a potential noncaloric food thickener or texturizer.

Modified Celluloses

Alkali cellulose is prepared by soaking cellulose in concentrated sodium hydroxide (> 14%) in the 'mercerization' process invented by John Mercer in 1844. The sodium ions are incorporated into the cellulose structure according to the following reaction (eqn (1)):

$$\underset{\text{Alkali cellulose}}{R_{cell}ONa} + H_2O. \qquad (1)$$

The cellulose structure swells, allowing easy penetration by dyes or reagents for the manufacture of cellulose derivatives.

Cellulose xanthate is formed by reacting alkali cellulose with carbon disulfide (eqn (2)):

$$\underset{\text{Cellulose xanthate}}{R_{cell}O\overset{\overset{S}{\|}}{C}SNa}. \qquad (2)$$

Regenerated cellulose is made by dissolving cellulose xanthate in 4–7% sodium hydroxide and contacting with aqueous sulfuric acid. These steps convert the cellulose xanthate back into cellulose, which may be spun into viscose rayon or cast into films. The fibers are used in textiles (artificial silk), tyre cords, and V belts. The films are used in packaging (Cellophane) or sausage casings. Weiner casings (70% regenerated cellulose, 12% glycerol, and 18% water) are peeled away after the meat emulsion is cooked. Hemp paper casings (23% paper, 46% regenerated cellulose, 21% glycerol, and 10% water) are used in bologna, salami, pepperoni, summer sausage, and liverwurst.

Cellulose hydroxyl moieties are highly reactive, allowing a variety of esters and ethers to be manufactured. Because each anhydroglucose has three hydroxyl groups, the maximum degree of substitution (DS) is three. Purified wood pulp or cotton linters (short fibers) are the industrial sources of 'chemical cellulose.'

Cellulose Ethers

Sodium carboxymethyl cellulose (CMC) is formed by reacting sodium chloroacetate with alkali cellulose (eqn (3)):

$$R_{cell}ONa + ClCH_2\overset{\overset{O}{\|}}{C}ONa \rightarrow \underset{\text{CMC}}{R_{cell}OCH_2\overset{\overset{O}{\|}}{C}ONa} + NaCl. \qquad (3)$$

Commercially available CMC has a DS range of 0.38–1.4, with 0.65–0.85 more common. The negatively charged carboxyl group makes CMC soluble in both hot and cold water. The solution viscosity decreases as the temperature increases. CMC is 'generally recognized as safe' and is used as a thickener in many foods such as cheese, frozen desserts, and salad dressings. It is not metabolized, so it is used in low-calorie foods.

Methyl cellulose is formed by reacting alkali cellulose with methyl chloride (eqn (4)):

$$R_{cell}ONa + ClCH_3 \rightarrow \underset{\text{Methyl cellulose}}{R_{cell}OCH_3} + NaCl. \qquad (4)$$

Methyl cellulose (DS 1.8) solutions form a firm gel when heated to 50–55 °C and return to solution when cooled. Methyl cellulose is added to salad dressings, jams and preserves, soda water, and meat patties as a binder.

Ethyl cellulose is produced by reacting alkali cellulose with ethyl chloride. In commercial products, the DS ranges from 2.0 to 2.6. It is water-insoluble and may be incorporated into inks used for marking foods and in vitamin tablet binders.

2-Hydroxypropyl methyl cellulose is formed by reacting alkali cellulose with mixtures of methyl chloride and 2-hydroxypropyl chloride. It forms gels like methyl cellulose, but has a higher gelation temperature. It may be used as an emulsifier, film former, stabilizer, or thickener in foods such as salad dressings, sherbet, pie fillings, fried foods, whipped toppings, breading batters, and baked goods.

2-Hydroxyethyl cellulose (HEC) is produced by reacting cellulose with ethylene oxide using a sodium hydroxide catalyst at 30–35 °C for about 4 h (eqn (5)):

$$R_{cell}OH + \underset{\diagup\ \diagdown}{CH_2{-}CH_2}\overset{\overset{O}{}}{} \rightarrow \underset{\text{HEC}}{R_{cell}OCH_2{-}CH_2OH}. \qquad (5)$$

Because the side-chain also has a hydroxyl group, ethylene oxide can continue to react and form a

side-chain with several units. HEC is soluble in both hot and cold water. The solution viscosity decreases as the temperature increases. HEC is not permitted as a direct food additive, but it may be used in food-packaging adhesives and coatings.

2-Hydroxypropyl cellulose is produced using propylene oxide, rather than the ethylene oxide used for HEC. It has a thermal gel point, like methyl cellulose, and is used in food coatings and glazings.

Cellulose Esters

Cellulose acetate is the most important cellulose ester. The cellulose is first 'activated' in aqueous acetic acid to ensure uniform acetylation. It is then dehydrated and reacted with acetic anhydride using a catalyst (e.g., sulfuric acid) in a solvent (e.g., anhydrous acetic acid) (eqn (6)).

$$R_{cell}OH + CH_3\overset{\overset{O}{\|}}{C}O\overset{\overset{O}{\|}}{C}CH_3 \rightarrow R_{cell}O\overset{\overset{O}{\|}}{C}CH_3 + CH_3\overset{\overset{O}{\|}}{C}OH. \quad (6)$$

Cellulose acetate

The resulting cellulose triacetate (DS ~ 3) product may be subsequently acid-hydrolyzed to lower the DS. Cellulose triacetate is water-insoluble and hydrophobic, whereas cellulose monoacetate is water-soluble. Cellulose acetate is used in fibers, plastics, photographic films, lacquers and reverse osmosis or dialysis membranes.

Other esters (e.g., cellulose formate, cellulose propionate, cellulose butyrate) may be formed, but they do not have the widespread commercial applications of cellulose acetate. Also, mixed esters (e.g., cellulose acetate propionate, cellulose acetate butyrate, cellulose acetate phthalate) may be produced.

Cellulose nitrate is made by reacting cellulose with nitric acid/sulfuric acid for 20–30 min. The acids are then removed by washing with water. The water must be removed with great care because dry cellulose nitrate is extremely explosive. It is often shipped wetted with water or alcohol. Highly nitrated (DS 2.4–2.6) forms are used as explosives (gun cotton). Less nitrated forms (DS 2.1–2.3) find applications in plastics, films, and inks.

Composition of Cellulosic Materials

Table 4 shows the composition of many cellulosic materials. The cellulose content in vegetables ranges from 1 to 21%; fruits range from 0.6 to 4.2%; seeds range from about 2 to 12%; agricultural residues range from 31 to 59%; wood ranges from 41 to 53%; flax and hemp have about 70% cellulose; and cotton (the purest natural cellulose) has about 95% cellulose. The free sugars shown in Table 4 are the sum of glucose, fructose, and sucrose. The 'crude fiber' is reported as the sum of cellulose and lignin in the case of the seeds. The indigestible portion of some fruits, called 'dietary fiber,' is reported as the sum of cellulose, lignin and hemicellulose. (*See* **Dietary Fiber**: Properties and Sources; **Fructose**; **Sucrose**: Properties and Determination.)

Cellulose Isolation and Analysis

Gravimetric Methods

Van Soest procedure The raw plant material is contacted with dilute acid and washed with solvents to remove starch, pectin, hemicellulose, fats, oils, protein, free sugars, and soluble minerals. The residue, called 'acid detergent fiber,' contains cellulose, lignin, and insoluble minerals (mainly silica).

Air-dried plant material (1 g), ground to particles of less than 1 mm, is placed in a beaker with 100 ml of acid-detergent solution (49 g l^{-1} H_2SO_4, 20 g l^{-1} cetyl trimethylammonium bromide). Decahydronaphthalene (2 ml) is added, and the contents are slowly boiled for 1 h while the reflux condenser maintains a constant liquid volume in the beaker. Then, the beaker contents are filtered through a tared Gooch crucible, washed with boiling water, acetone, and hexane (optional). The residual acid detergent fiber (ADF) is dried at 100 °C and weighed.

The cellulose content in ADF can be measured by two methods: (1) cellulose removal with acid or (2) lignin removal by oxidation.

Method 1 The ADF prepared above is placed in a tared Gooch crucible containing an equal volume of asbestos as a filter aid. The crucible contents are contacted with room-temperature 72% sulfuric acid for 3 h and then thoroughly washed with hot water. The sample is dried at 100 °C for 8 h and weighed. The weight loss corresponds to the cellulose content.

Method 2 The ADF (0.5–1.0 g) is placed in a tared Gooch crucible. A saturated potassium permanganate solution (50 g l^{-1} $KMnO_4$, 0.05 g l^{-1} $AgSO_4$) and buffer solution (6 g l^{-1} $Fe(NO_3)_3 \cdot 9H_2O$, 0.15 g l^{-1} $AgNO_3$, 500 ml l^{-1} glacial acetic acid, 5 g l^{-1} potassium acetate, 400 ml l^{-1} *t*-butyl alcohol) are mixed in a 2:1 ratio (by volume). This mixture (25 ml) is added to the crucible for 90 min at room temperature to oxidize the lignin in the ADF. The spent reagent is then removed by vacuum filtration. The crucible contents are soaked for 5 min in demineralizing solution

Table 4 Composition of cellulosic materials (g per 100 g of dry matter)

	Cellulose	*Lignin*	*Hemicellulose*	*Pectin*	*Starch*	*Free sugars*
Vegetables[a]						
Leafy						
Broccoli	7.2	0.26	24		1.7	15
Brussels sprouts	9.04	2.1	26		8.7	18
Cabbage	8.9	4.3	26		4.3	47.3
Cauliflower	13.4	Tr	13		5.0	29
Lettuce	20.6	Tr	9.2		0	17
Legumes						
Beans, haricot	5.3	0.9	22.0		51.3	
Beans, runner	17	3	21		2.9	39.6
Peas	14	2	36		66	6.9
Root						
Carrot	12.9	Tr	19		0.9	24.8
Turnip	11	Tr	23		Tr	47.7
Fruiting vegetables						
Pepper	3.5	Tr	10		Tr	34.2
Tomato	9.1	5.3	11		Tr	46.4
Tuber						
Potato	1.2	Tr	9.2		84.6	1.5
Fruits[a]						
Apples	2.9	Tr	5.8	2.3	1.8	63.9
Apricots	15			3.3	0	56.6
Banana	1.3	0.93	3.83		10.4	55.8
Blackberries	44			1.9	0	34.1
Cherries, sweet	1.2	0.3	4.5	0.4	0	65.4
Grapefruit	0.6	0.9	4.9		0	67.7
Lemons	35			3.4	0	21.1
Oranges	14			3.3	0	59.6
Peaches	1.8	5.1	12.2	3.3	0	74
Pears	4.2	2.7	8.2		Tr	63.5
Pineapples	7.64			0.25	0	74.1
Strawberries	3.6	8.4	10	3.5	Tr	61.8
Seeds[b]						
Barley	5.3				64.1	
Corn	2.4				71.8	1.9
Grain sorghum	2.7		2.5		70.2	1.4
Oats	11.9				60.6	
Peanut	2.8		2.5		4.0	4.7
Wheat	2.1				74	
Agricultural residues[c]						
Barley straw	44	7	27			
Cottonseed hulls	59	13	15			
Oat straw	41	11	16			
Rice straw	33	7	26			
Sorghum straw	31	11	30			
Sugarcane bagasse	40	13	29			
Wheat straw	39	10	36			
Trees[d]						
Hardwood						
Aspen	53.3	16.3	26.2			
White birch	41.0	18.9	36.2			
Red maple	44.1	24.0	29.2			
Softwood						
Balsam fir	44.8	29.4	23.6			
Jack pine	41.6	28.6	25.6			
White spruce	44.8	27.1	26.1			
Bast fibers[e]						
Flax	71.2	2.2	18.6	2.0		
Leaf fibers[e]						
Manila hemp	70.2	5.7	21.8	0.6		
Seed fibers[e]						
Crude cotton	95.3	0	0	1.0		

(50 g l^{-1} oxalic acid dihydrate, 700 ml l^{-1} 95% ethanol, 50 ml l^{-1} 12 M HCl), which is then removed by filtration. The filter contents are washed with ethanol and acetone. The crucible is dried at 100 °C for 8 h. The crucible contains cellulose and insoluble ash. The crucible may be placed in a 500 °C oven for 3 h; the weight loss corresponds to the cellulose content.

A simpler approach to removing the lignin from the ADF has been described by C.S. Edwards. The ADF is soaked in activated trigol (i.e. 6.3 ml of 32% hydrochloric acid dissolved in 1 l of pure triethylene glycol) in a 121 °C autoclave for 60 min. The sample is then washed with 95% ethanol and acetone. The residue contains cellulose and insoluble ash.

Other gravimetric methods Cross and Bevan cellulose is the lignocellulose portion remaining after removing hemicellulose with two boilings in NaOH and removing lignin with chlorine and bleach.

Monoethanolamine cellulose is the lignocellulose portion remaining after monoethanolamine treatment, chlorination, and bleaching.

Norman–Jenkins cellulose is the lignocellulose portion remaining after sodium sulfite boiling, bleaching, and acid treatment.

α-Cellulose is the cellulose fraction insoluble in room temperature 17.5% sodium hydroxide wash water (DP > 90).

β-Cellulose is the cellulose fraction in the alkaline wash water that precipitates upon neutralization (15 < DP < 90).

γ-Cellulose is the cellulose fraction soluble in the neutralized alkaline wash water (DP < 15).

Holocellulose is the residue that remains after lignocellulose is defatted with hot, azeotropic benzene/ethanol and delignified with hot chlorous acid ($HClO_2$). Holocellulose contains the cellulose and hemicellulose of the original plant material.

Colorimetric Method

Hexosan assays may be used to measure cellulose, the dominant hexose polymer in starch-free plant materials (other than softwoods). Holtzapple describes an assay in which the hexosan sample is placed in a sealed test tube with 1530 g l^{-1} sulfuric acid and 20 g l^{-1} chromotropic acid, and boiled for 1 h. Hexose C6 degrades to form formaldehyde, which reacts with the chromotropic acid to form a purple compound that is measured spectrophotometrically. There is minor interference by pentosans that may be reduced by lowering the chromotropic acid concentration to 1 g l^{-1} and shortening the reaction time to 20 min. There is also some interference with lignin, so the most accurate results are obtained with holocellulose, rather than the original lignocellulose. (*See* **Spectroscopy**: Visible Spectroscopy and Colorimetry.)

Chromatographic Method

The cellulose content can be estimated from the glucose composition, because cellulose is the main source of glucose in plant materials (assuming free sugars and starch are not present). (*See* **Chromatography**: Principles.)

See also: **Chromatography**: Principles; **Dietary Fiber**: Properties and Sources; **Fructose**; **Hemicelluloses**; **Lignin**; **Protein**: Chemistry; **Spectroscopy**: Visible Spectroscopy and Colorimetry; **Sucrose**: Properties and Determination

Further Reading

Atalla RH (1987) *The Structures of Cellulose. ACS Symposium Series, No. 340*. Washington, DC: American Chemical Society.

Bogan RT, Kuo CM and Brewer RJ (1979) Cellulose derivatives, esters. In: *Kirk–Othmer Encyclopedia of Chemical Technology*, 3rd edn., vol. 5, New York: Wiley.

Edwards CS (1973) Determination of lignin and cellulose in forages by extraction with triethylene glycol. *Journal of the Science of Food and Agriculture* 24: 381–388.

Fengel D (1970) Ultrastructural behavior of cell wall polysaccharides. *Tappi* 53(3): 497–503.

Goering HK and Van Soest PJ (1970) *Forage Fiber Analysis. Agricultural Handbook, No. 379, Jacket No. 387–598*. Washington, DC: US Department of Agriculture.

Greminger GK (1979) *Kirk–Othmer Encyclopedia of Chemical Technology*, 3rd edn., vol. 5. New York: Wiley.

Holtzapple MT and Humphrey AE (1983) Determination of soluble and insoluble glucose oligomers with chromotropic acid. *Analytical Chemistry* 55: 584–585.

Jayne G (1971) Cellulose and cellulose derivatives. In: Bikales NM and Segal L (eds) *High Polymers*, 2nd edn., vol. V, part IV. New York: Wiley Interscience.

Plunguian M (1943) *Cellulose Chemistry*. New York: Chemical Publishing.

Southgate DAT (1976) *Determination of Food Carbohydrates*. London: Applied Science.

Turbank AF, Durso DF, Battista OA *et al*. (1979) Cellulose. In: *Kirk–Othmer Encyclopedia of Chemical Technology*, 3rd edn., vol. 5. New York: Wiley.

Tr, trace.

[a]From Southgate DAT (1976) *Determination of Food Carbohydrates*, pp. 91–93. London: Applied Science.

[b]From Zaborsky OR (1981) *CRC Handbook of Biosolar Resources*, vol. II. Boca Raton, FL: CRC Press.

[c]From Marsden WL (1986) *CRC Critical Reviews in Biotechnology*, vol. 3, issue 3, p. 235. Boca Raton, FL: CRC Press.

[d]From Timell TE (1957) Carbohydrate composition of ten North American species of wood. *TAPPI* 40: 568.

[e]From Turbank AF, Durso DF, Battista OA *et al*. (1979) Cellulose In: *Kirk–Othmer Encyclopedia of Chemical Technology*, 3rd edn., vol. 5. New York: Wiley.

CEREALS

Contents

Contribution to the Diet

C MacEvilly, Food Safety Promotion Board, Cork, Ireland

Background

Cereals are grown for their highly nutritious edible seeds which are often referred to as grains. They remain the staple food in most diets, even given their decline in importance as a country becomes more affluent. Current nutritional advice highlights the benefits of cereal foods as a source of complex carbohydrates and as foods that should replace some of the energy currently provided by fat in the diet.

Cereals supply a variety of nutrients and other food components like phytochemicals. Although cereals make an important contribution to the diet, they cannot alone support life because they are lacking in vitamins A, D, and C and iodine, for example. Also cereal proteins are deficient in certain amino acids, notably lysine. Refined cereal products and unrefined cereals have certain advantages and disadvantages. With refinement, some nutrients and fiber are removed, but the body is better able to make use of certain nutrients. The role of fortified cereals as a means of increasing levels of specific vitamins has been recognized. This chapter will discuss the role that cereal plays in both in the developed and developing world and the implications that the type of cereal has on the diet.

Cereals as a Global Food Source

The world has over 50 000 edible plants and just three of them – rice, maize, and wheat – supply 60% of the world's food energy intake. Table 1 presents an overview of world cereal production for 1999.

Rice feeds almost half of humanity. Per capita rice consumption has generally remained stable, or risen slightly since the 1960s. However, it has declined in recent years in many of the wealthier rice-consuming countries, such as Japan, the Republic of Korea, and Thailand, because rising incomes have enabled people to eat a more varied diet.

Although there are over 10 000 species in the Gramineae (cereal) family, few have been widely introduced into cultivation over the past 2000 years. Cereals constitute the 'starchy staples' in the human

Table 1 World cereal production – forecast for 2000 (million tonnes)

	Wheat		*Coarse grains*		*Rice (paddy)*		*Total*	
	1999	*2000*	*1999*	*2000*	*1999*	*2000*	*1999*	*2000*
Asia	260.1	251.2	218.1	190	553.1	540.4	1031.2	981.6
Africa	14.9	13.5	76.1	79.1	17.7	17.5	108.7	110.1
Central America	3.1	3.4	28.7	28.7	2.3	2.4	34.1	34.5
South America	19.7	18.6	58.6	61.7	21.1	20.1	99.5	100.4
North America	89.5	86.7	290.5	305.6	9.3	8.7	389.3	401.0
Europe	178.3	188.2	202.4	194.8	3.2	3.1	383.9	386.2
Oceania	24.3	20.3	8.9	9.9	1.4	1.1	34.6	31.2
World	589.8	581.8	883.4	869.7	608.2	593.4	2081.4	2044.9
					406[a]	397[a]	1880[b]	1848[b]
Developing countries	276.6	268.4	369.3	345.2	582.1	568.1	1228.0	1181.8
Developed countries	313.2	313.3	514.1	524.4	26.1	25.3	853.4	863.1

From FAO/GIEWS Food Outlook no. 5, p. 4, with permission.
[a]Rice in milled terms.
[b]Including rice in milled terms.

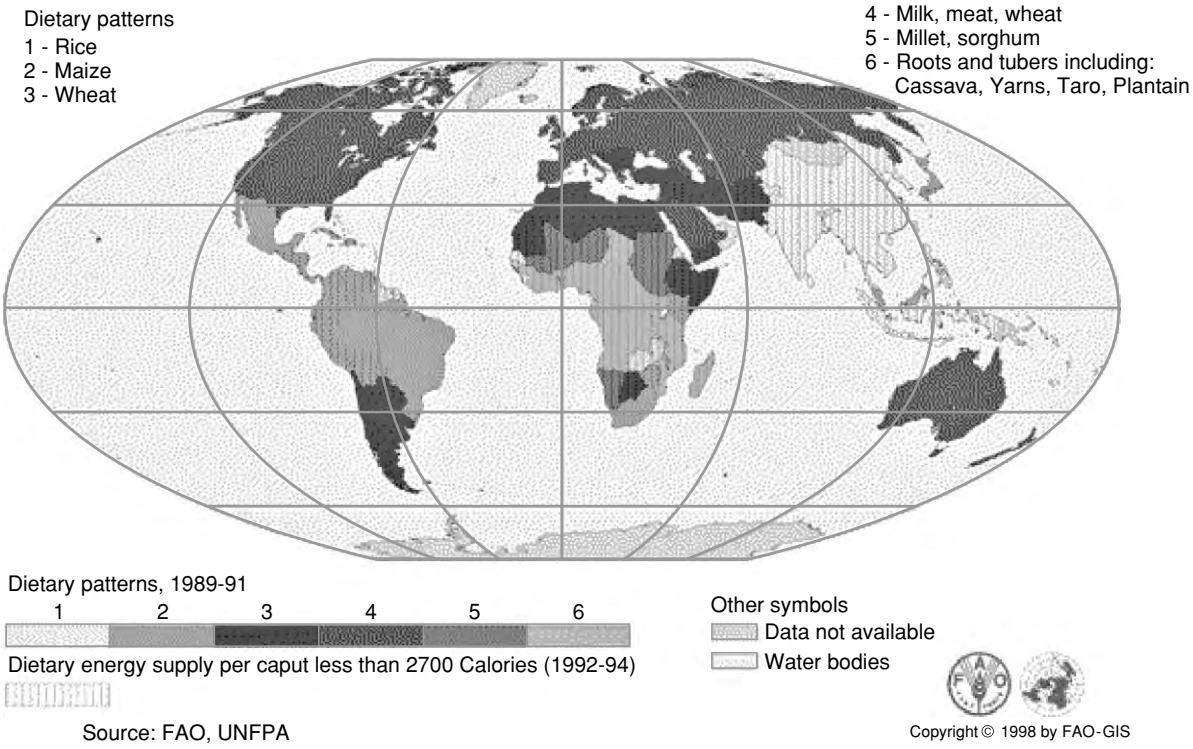

Figure 1 Dietary patterns.

diet and as such are the primary source of dietary carbohydrate throughout the world. The contribution of these crops to the diet varies from continent to continent, and even among countries within continents. The consumption of starchy staples throughout the world has changed over the last three decades. Figure 1 shows the world picture of dietary patterns, with countries highlighted according to the starchy staple providing the highest proportion of energy. As can be seen, large parts of the world depend on wheat and rice, with more selected areas dependent on maize, sorghum, and millet.

The distribution of the production, and to a great extent the importance, of the different cereals is in part due to the climatic constraints on the growth of the different species. Wheat production is limited to temperate zones; rye has a wider northern range and can be grown in the Arctic circle, and rice can be grown in standing water in tropical and subtropical climates and less commonly on dry uplands. Barley and oats have a similar range to wheat, although oats thrive on poorer soils. Maize and millets are major crops in the tropics and subtropics; much is also grown in temperate climates where it is also important as animal feed.

As a food, cereals have many benefits and there are several reasons why cereals have become the staple food of mankind. They do not deteriorate easily if kept dry, they are easily stored and transported, they are cheap to produce, and they have a high nutrient density. (*See* **Wheat**: The Crop; **Rice**; **Barley**; **Oats**; **Maize**; **Rye**; **Millets**.)

The Nutritional Composition

In westernized countries approximately 50% of the food energy available is derived from cereals and in the developing world it provides two-thirds of energy and protein. Cereals provide 1400–1600 kJ per 100 g whole cereal. The nutritional composition of the main types of cereal grains and some products is shown in Table 2. Cereal grains contain, on average, 75% carbohydrate, 10–15% protein, 2% fat, and 10–15% water.

Cereals as an Energy and Protein Source

Cereals are classified as 'carbohydrate foods' and although this is compositionally correct, such a classification does not take into consideration the importance of the proteins present in cereals. They constitute a rich source of both nonstarch polysaccharides (dietary fiber) and starch, which together comprise 70–77% of the grain. Protein accounts for 6–15% and the limiting amino acid is lysine, although maize is additionally low in tryptophan. Protein quality is highest in the outer layers of the endosperm (a starchy

Table 2 Composition of cereals, grains, and some products

Product	*Water (%)*	*Protein (%)*	*Fat (%)*	*Carbohydrates (%)*	*Energy value (kJ per 100 g)*
Wheat	14	12.7	2.2	63.9	1318
Flour (plain)	14	9.4	1.3	77.7	1450
Bran	8.3	14.1	5.5	26.8	872
Germ	11.7	26.7	9.2	(44.7)	1276
Bread (white sliced)	40.4	7.6	1.3	46.8	926
Raw macaroni	9.7	12.0	1.8	75.8	1483
Rye (grain and wholemeal)	15	8.2	2.0	75.9	1428
Rice (raw white)	11.4	7.3	3.6	85.8	1630
Oatmeal	8.2	11.2	9.2	66.0	1567

Data from Holland B, Welch AA, Unwin ID, Buss DH, Paul AA and Southgate DA (1991) *The Composition of Foods*, 5th edn. London: Royal Society of Chemistry and MAFF.

Table 3 Consumption of cereal grains (1998) and their contribution to the estimated intake of dietary energy and protein in selected countries

Country	*Total consumption of cereals (kg per person per year)*	*Energy*	*Protein*
Argentina	127.1	926	24.4
Australia	90.2	724	24.3
Chile	135.8	1102	29.6
China	186.6	1623	35.5
Denmark	112.4	875	26.8
Finland	104.6	842	26.0
Germany	96.9	764	23.2
Greece	149.9	1056	33.3
India	168.4	1545	36.5
Iran	174.7	1446	40.1
Nigeria	157.2	1320	34.9
Spain	102.9	758	23.5
UK	85.0	656	21.3
US	118.8	885	26.1
Yugoslavia	112.5	872	25.0

From FAO Food Balance Sheets 1994–1998. Rome: FAO.

core with a protein matrix) and in the germ (the grain's embryonic root and shoot) because that is where the lysine-rich albumins and globulins (salt-soluble proteins) are located and as a result are more affected by processing.

Gluten is the major protein in wheat and rye. The special properties of this protein make it particularly important in breadmaking; isolated gluten is also used as a method for assessing suitability of cereal proteins in gluten intolerance (celiac disease), which will be discussed later. Oryzenin is the major protein in rice. Table 3 shows the role of cereals in the supply of dietary energy and protein in various countries.

The data are presented as food supply statistics (sometimes referred to as food balance sheets or food disappearance statistics). This type of population level data is relatively easy to collect compared with information based on the dietary records of individuals, which requires meticulous attention to detail. However, there are also important drawbacks: data exist only for basic commodities, they cannot show patterns in different sectors of the population, and they substantially overestimate the amounts of food actually eaten as they do not take into account wastage, for example.

Lipids

In themselves, cereals are low-fat foods, their endogenous fats making only a very minor contribution to total intake and being overshadowed by the amount and composition of the fats used in the cooking and preparation of the cereal foods commonly consumed. Although lipids comprise only about 1.5–7% of cereal grains, some lipid components appear to have special properties that contribute to added health benefits. Most (approximately 80%) of the lipid in cereals is present as triglycerides (fatty acids esterified in a glycerol backbone). Linoleic (C18:2) acid is the major unsaturated fatty acid in cereals except brown rice, which contains more oleic acid (C18:1). Palmitic acid (C16:0) is the major saturated fatty acid in cereals and, except in barley, is present in lower concentrations than linoleic and oleic acids. However, cereals also contain a broad class of unsaponifiable lipid material (about 2–10% of the total) that includes tocols, carotenoids, sterols, and high-molecular-weight alcohol. Tocols and carotenoids have antioxidant properties and sterols have been extensively studied for their role in the reduction of blood cholesterol levels.

Mineral Elements and Vitamins

Whole grains are a good source of thiamin and provide significant amounts of many other B vitamins. However, factors affecting the bioavailability of these vitamins from grain products are still under investigation. The germ is rich in vitamin E; however, vitamin E is not stable to light, heat, and air, so almost all processing methods will adversely affect its concentration. As grains mature, the niacin they contain becomes progressively more bound to indigestible components of the seeds. Cereal grains are deficient in vitamin A, C, and D. The ascorbic acid that is customarily added to white flour to improve its baking properties has no nutritional significance. In addition, low levels of vitamins D and K are found in the lipids of most cereals. Recent results from the National Food Survey in the UK suggest that fortified cereals provide almost 32% of the average intake of folic acid.

Cereals consist of about 1.5–2.5% of minerals. This includes minerals such as potassium, phosphorus, magnesium, iron, and zinc. Selenium enters the food chain via plants, which obtain it from soil. Wide international variations in intakes have been observed, in part because of wide differences in selenium soil content and differences in selenium bioavailability from different foods. In 1997, 12.2% of total selenium intakes in the UK were obtained from breads and 10% from cereals. A wide variability exists in the levels of selenium present in grain and cereal products. Among the cereal grains, rice contains the highest levels (10–13 μg $100\,g^{-1}$) while recent analysis of levels in bread (6–9 μg $100\,g^{-1}$) suggest that levels have decreased over the last 20 years and are significantly lower than selenium levels in US bread samples (32–44 μg $100\,g^{-1}$). This is because wheat is a good source of selenium in North America, as the wheat is grown in soils rich in selenium. In Europe, although bread and cereals make a significant contribution to selenium intake, due to frequency of consumption (22% of total selenium in the UK), European wheat is not a good source of selenium because of the low bioavailability of selenium from European soils. Wheat is not a good natural source of iron in Europe but many cereals are fortified with iron.

The differences in selenium intake between the USA and Europe are reflected in serum and whole-blood selenium concentrations. Although serum values of $100\,\mu g\,l^{-1}$ are the suggested level required for optimal enzyme activity, current levels in the EU appear to be below this level with mean serum selenium levels in seven European countries calculated as $79\,\mu g\,l^{-1}$.

Mineral composition of grains can be affected by genetic factors, as well as environmental ones, such as soil composition, crop fertilization practices, and climatic conditions. The values reported for any grain can vary over a wide range. Refer to individual vitamins and minerals.

Complex Carbohydrates and Dietary Fiber

Cereals are composed of up to 85% (dwb; dry weight basis) of complex carbohydrates. These are defined by their sugar residues and the linkages between them. Cereals also contain cellulose, β-glucans and dietary fiber. Fiber is the sum of the plant polysaccharides plus lignin that is not digested by enzymes in the intestine. Cereals are a unique source of dietary fiber and in countries where whole-grain products, such as wholemeal bread, are popular, cereal products may contribute more than 40% of the average dietary fiber intake (data taken from the UK National Diet and Nutrition Survey).

Fiber has two forms – soluble and insoluble. Although to some extent their function overlaps, they do have different effects on the body and so it is important to eat a variety of high-fiber foods to get both types.

Both types of fiber – soluble and insoluble – can help with weight control. In the stomach, they absorb water and some delay the emptying of the stomach, instigating a feeling of fullness.

Insoluble fibers found in wheat bran, whole-grain breads, and cereals hold water in the colon, thus increasing bulk which stimulates the muscles of the digestive tract so that they retain their health and tone. The toned muscles can move waste product more easily through the colon for excretion. This helps prevent constipation, hemorrhoids, and diverticulosis.

Soluble fibers found in oat bran are credited with reducing the risk of heart and artery disease – atherosclerosis – by lowering the level of cholesterol in the blood. It appears that the products of bacterial digestion of soluble fiber in the colon are absorbed into the body and may inhibit the body's production of cholesterol as well as enhancing the clearance of cholesterol from the blood. Blood cholesterol levels may also decrease if food sources of soluble fiber are used as part of a low-fat, low-cholesterol diet. Certain fibers also bind to cholesterol compounds in the gut and carry them out of the body so that the whole-body content of cholesterol is lowered.

In January 1997, the US Food and Drug Administration (FDA) allowed the first food-specific health claim for a whole-grain cereal, oats. That regulation was amended in February 1998 to include soluble fiber from psyllium seed husks. In the USA, foods containing whole oats or psyllium seed husks as ingredients can now make a health claim linking consumption with a reduced risk of coronary heart disease when they are part of a diet low in saturated fat and cholesterol. The FDA acknowledged that soluble fiber from sources other than oats and psyllium probably also affects blood lipids and reduces the risk of heart disease, but that the sources would need to be considered on a case-by-case basis.

Soluble fibers also improve the body's handling of glucose, even in people with diabetes, perhaps by slowing down the digestion or absorption rate of carbohydrates. Blood glucose levels therefore stay moderate, helping to prevent symptoms of diabetes such as hypoglycemia.

Phytochemicals

During the last decade, data from experimental and epidemiological studies have been accumulated showing that grains contain a large variety of potentially health-promoting substances called 'plant chemicals' or 'phytochemicals.' The explosion of new information related to the exciting approach of

phytochemicals is certainly only a prelude to their inclusion in our daily diet.

Because foods are complex mixtures of various phytochemicals, studying them is a difficult task. A major group of phytochemicals in foods are the phenolic compounds, which include plant flavonoids like catechins. In cereals, flavonoids are present in small quantities only. Barley contains measurable amounts of catechins and some di- and trimer procyanidins. Work mainly in experimental animals and *in vitro* has suggested that flavonoids possess numerous biochemical and pharmacological qualities, including anticarcinogenic, antiinflammatory, and antiallergic properties.

Other phytochemicals found in cereals include phytoestrogens such as lignans, isoflavones, and coumestans. The precursors of these compounds are found in fiber-rich unrefined grain products and cereals. Lignans are potent antioxidants and exert anticancer effects in experimental studies. Lignan contents vary greatly between cereals. The amounts in cereals are low compared to that in linseed and high compared to those in vegetables. However, since the consumption of cereals is considerable, high amounts of lignans and other bioactive phytochemicals are provided in the daily diet.

Further nutritive antioxidants present in cereals are small amounts of carotenoids, tocopherols, and tocotrienols, especially in the lipid-rich germ. Grains are a unique group in that they contain more tocotrienols than other food products. Tocotrienols possess vitamin E activity.

Fortification

No one cereal can provide adequate amounts of all nutrients to meet the nutritional requirements of a child or adult. Because of the stable nature of cereal products, their contributions to the diet of the populations of the world's nations are frequently perceived by governments as an important means of insuring adequate nutritive standards, not only through their natural composition but also through the addition of nutrients from other sources. Various terms are applied to such additions, including restoration, fortification, and enrichment.

The nutritional quality of sorghum and millets, especially the former, is poor. Therefore attempts have been made to fortify these cereals to make nutritionally superior and acceptable products. Cost, availability of ingredients, and marketability must be taken into consideration if fortification is to be implemented successfully on a sustained basis. In the UK, bread and breakfast cereals are currently voluntarily fortified with folic acid and this in part may explain the importance of cereal sources to the folic acid intake of populations. Other vitamins and minerals that are currently fortified include thiamin, iron, riboflavin, niacin, and calcium. In the USA, mandatory fortification was enforced on 1 January 1998.

Wheat flour and other common cereal products are very suitable vehicles for fortification because there are virtually no major population subgroups which would not be able to use cereal grains and because cereals are consumed by all socioeconomic subgroups. In the UK and Canada, the addition of nutrients to flour, where addition is required, is made at the flour mill. In the USA enrichment is permitted either at the mill or at the bakery. In most developing countries refined cereal products are not vitamin-enriched, but 10 nations are currently adding iron to wheat or maize flours and many other countries are considering doing so.

For any supplementation initative to succeed it is necessary that the added nutrient is physiologically available; it does not create an imbalance of essential nutrients, e.g., by competing for absorption with another nutrient; it does not adversely affect the acceptability of the product; and there is reasonable assurance that intakes will not become excessive. (*See* **Food Fortification**.)

Effects of Processing and Cooking

Changes resulting from cooking are complex: in only a few cases, such as boiled rice and pasta, are cereal products cooked substantially on their own. More frequently they are included in a recipe, so that differences between the raw cereal ingredient and the final product reflect not only changes due to cooking but also to dilution and interaction with other ingredients.

In general, fiber, protein, vitamin, and mineral concentrations in cereals decrease from the outer layer to the inner part of the seed, with the decline being steeper in some cereals than others. For that reason, and because of the removal of the aleurone (outer) fraction of the cereal, milling results in large losses of many nutrients. When wheat is milled to white flour, the niacin content drops to about 20% of its original level, thiamin to 27%, and riboflavin to 19%. After milling of wheat to white flour and then bleaching, almost 98% of vitamin E is lost.

Many cereal-based foods are subjected to heat processing in baking, frying, and extrusion cooking, and these have effects on more labile vitamins. In processes involving boiling the losses are of the order of 40% for most of the B-group vitamins; losses of folates are slightly higher at around 50%. In baking the losses are generally lower except for the folates. The losses during the more vigorous conditions of

extrusion cooking are still of the same order, primarily because the cooking times are generally shorter than with more conventional cooking procedures.

Probably the most significant enzymic changes that occur during processing are those involved in the conversion of starch into sugars and the subsequent fermentation mediated by microorganisms, converting sugars to alcohol or lactic acid. Another important example is the reduction in phytate content during fermentation and the proving of bread doughs.

Negative Attributes

In spite of the fact that cereals are among the safest and most important foodstuffs, there are some antinutritious aspects of their composition (some of these affect only a small proportion of consumers). There are hazards associated with their storage and handling also.

The negative nutritional attributes of phytate derive from the fact that it forms insoluble complexes with minerals such as calcium and iron, possibly reducing their bioavailability and leading to failure of their absorption in the gut of animals and humans and thus to mineral deficiencies. The reduced bioavailability of minerals depends upon several factors, including the nutritional status of the consumer, concentration of minerals and phytate in the foodstuff (and the diet as a whole), ability of endogeneous carriers in the intestinal mucosa to absorb essential minerals bound to phytate and other dietary substances, and the digestion or hydrolysis of phytate by phytase during processing and in the intestine.

Tannins are phenolic compounds of the flavonoid group and are also considered to have negative attributes. It is alleged that they reduce protein digestibility through phenol–protein complexing. There is also no doubt that they and their derivatives can adversely affect flavor and color, thus reducing palatability. Tannins in seeds are concentrated in their outer layers, therefore milling to remove the bran and testa is an effective means of reducing the tannin content. However, because nutrient content and protein quality tend to decrease from the outer to the inner regions of cereal grains, tannin reduction by milling is achieved at the expense of nutrient loss.

Adverse effects on health can arise from the consumption of cereals contaminated by toxic substances. At least 25% of the world's crops may be affected by mycotoxins, even small quantities of which can potentially harm humans. These are toxic secondary metabolites produced by fungi. Aflatoxins, often found in wheat and rice, as well as corn, are some of the most potent naturally occurring carcinogens. (*See* **Mycotoxins**: Classifications.)

Celiac disease is a permanent condition, requiring a lifelong strict gluten-free diet, and is the main form of wheat intolerance. Older literature suggests that celiac disease has a prevalence of around one case in 1500 people in the UK. However, the advent of sensitive serological screening tests has allowed studies which suggest that the true prevalence may be higher, both in the UK (although a population-based study has yet to be conducted) and abroad. Studies in Italy, the USA, and The Netherlands have suggested that one in 300 of the general population in these countries may be gluten-sensitive. Classically, celiac disease inflicts many "negative attributes" on its sufferers, including diarrhea, weight loss, and anemia. However, these may be absent and many individuals may suffer from mild or nonspecific symptoms, which may be ignored for many years. Delayed diagnosis can result in osteoporosis, infertility, or even cancer. The condition is triggered by the ingestion of wheat, rye, and barley. Oats are probably safe for consumption by gluten-sensitive individuals. (*See* **Food Intolerance**: Types; Food Allergies.)

In conclusion, cereals are important sources of many nutrients, but they are limited in others. In the past few years, a steady stream of exciting new developments in cereal nutrition has flowed from scientists to consumers. Sometimes data seem contradictory but gradually the physiological benefits of cereal-rich diets are becoming clearer and indications are good that more health benefits of grains and grain components will soon be revealed.

See also: **Barley**; **Food Fortification**; **Food Intolerance**: Types; Food Allergies; **Maize**; **Millets**; **Mycotoxins**: Classifications; **Oats**; **Rice**; **Rye**; **Wheat**: The Crop

Further Reading

Andlauer W and Furst P (1998) Antioxidative power of phytochemicals with special reference to cereals. *Cereal Foods World* 43: 356–360.

British Nutrition Foundation (BNF) (1994) *Briefing Paper. Starchy Foods in the Diet*. London: British Nutrition Foundation.

British Nutrition Foundation (BNF) (1999) *Briefing Paper. Nutrition and Food Processing*. London: British Nutrition Foundation.

British Nutrition Foundation (BNF) (2001) *Briefing Paper. Selenium and Health*. London: British Nutrition Foundation.

British Nutrition Foundation (BNF) (2002) *Task Force Report. Health Effects of Plant Derived Foods and Drinks*. London: British Nutrition Foundation.

Blackwood AD, Salter J, Dettmar PW and Chaplin MF (2000) Dietary fibre, physicochemical properties and their relationship to health. *Journal of the Royal Society for Promotion of Health* 120: 242–247.

Carl Hoseney R (1994) *Principles of Cereal Science and Technology*, 2nd edn. Minnesota: American Association of Cereal Chemists.
FAO/GIEWS Food Outlook No 5, November 2000, page 4.
FAO/WHO Experts consultation. Carbohydrates in human nutrition, 2nd edn. (1998) FAO. Food and Nutrition Paper. Rome: FAO.
Garrow JS, James WPT and Ralph A (2000) *Human Nutrition and Dietetics*, 10th edn, pp. 333–348. London: Churchill Livingstone.
Henry RJ and Kettlewell PS (1996) *Cereal Grain Quality*. London: Chapman and Hall.
Holland B, Welch AA, Unwin ID, Buss DH, Paul AA and Southgate DA (1991) *The Composition of Foods*, 5th edn. London: Royal Society of Chemistry and MAFF.
Kent NL and Evers AD (1994) *Technology of Cereals and Introduction for Students of Food Science and Agriculture*, 4th edn. Oxford: Pergamon.
Kulp K and Ponte JG (2000) *Handbook of Cereal Science and Technology*, 2nd edn. New York: Marcel Dekker.
Ministry of Agriculture, Fisheries and Food (1997) *Total Diet Study – Aluminium, Arsenic, Cadmium, Chromium, Copper, Lead, Mercury, Nickel, Selenium, Tin and Zinc*. Food Surveillance Information Sheet no. 191. (http://archive.food.gov.uk/maff/archive/food/infsheet/1999/no191/191tds.htm)
Ministry of Agriculture, Fisheries and Food (2000) *National Food Survey*. London: The Stationery Office.

Bulk Storage of Grain

D Richard-Molard, INRA, Nantes, France

Introduction

At the beginning of the 21st century, world production of cereal grains is about 1600 million tonnes. This represents by far the greatest production of staple food material, the second most important source, i.e., tuber-producing crops, being far behind, with a total production of less than 100 million tonnes of edible dry matter. A simple arithmetic calculation shows that cereal crops will be able to provide more than the 1.3 kg of bread or rice per day for all of the 6000 million world population, provided that postharvest losses are minimized.

Whatever the political problems and ethnic questions, two essential facts must be taken into account: cereals (mainly wheat and maize) are produced in excess in only a limited number of countries, and only during limited periods in the year. Good and safe storage would allow consumption to be delayed and grain to be exported anywhere throughout the year and is, therefore, of primary economic, political, and social importance.

Existing Losses of Cereal Crops

It is difficult, if not impossible, to calculate accurately the postharvest losses resulting from poor storage conditions and inadequate practices. However, in 1976 it was estimated that losses of grain could represent a commercial value up to US$4000 million.

Food and Agriculture Organization (FAO) reports generally indicate that percentage losses remain at rather low levels in industrialized countries, probably never more than 1 or 2%, whereas they can reach very high levels, maybe 50% or even more, in countries where climatic conditions are unfavorable and modern storage equipment is lacking. It is important, however, to realize that, in developed countries, low percentage levels of loss involve enormous quantities of grain, leading to problems of significant economical importance. By contrast, high levels of losses in developing areas apply in most cases to only limited quantities of stored cereals, raising questions of vital importance only for the populations concerned.

It must also be emphasized that losses in grain quality can occur without significant losses of dry matter, e.g., decreases of nutritional value, losses of vitamins or essential fatty acids, appearance of toxicity, off-flavors or discolorations due to microbiological activity and insect pest infestation. The assessment of such hidden losses remains impossible at the present time.

In every situation it is of primary importance to have the best understanding of local environmental, economic, climatic, and ethnic conditions, and appreciation of possible causes of grain deterioration during storage, in order to minimize losses.

Some basic questions still lack satisfactory technical or economic answers but, usually, it can be said that postharvest losses are mostly due to lack of adequate training of the people who handle and store the grain. Most significant losses generally result from insect infestation or from the growth of microorganisms. To a lesser extent, biochemical changes in the grain itself can also be important.

The causes and extent of grain losses may differ widely from one place to another. They depend on many parameters but three major environmental factors determine the extent of damage in a given situation: the water content of the grain, the mean temperature of the bulk, and the duration of storage. Of these factors, water is by far the most important and also the most interesting because it can be managed; whereas it is very difficult to control temperature in practice. (*See* **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability.)

Main Biological Causes of Losses

Unless cereal grains are sufficiently dry at harvest, they must be considered as perishable biological products that will be damaged by microorganisms or insects, depending on the environmental conditions. Surprisingly, the major role played by microorganisms, and particularly by molds, has only been recognized recently, but it is now understood that artificial drying of moist grains after harvest is essential for the avoidance of mold growth during storage, rather than for the prevention of development of insects or biochemical changes in the kernel itself. (*See* **Spoilage**: Molds in Spoilage.)

Respiration of the grain itself is a secondary phenomenon in most situations, when germination is not taking place. Respiration of the germ seems to remain more or less constant and at a very low level, compatible with seed conservation, whatever the hydration or the gaseous environmental conditions.

To initiate germination of the grain, very large moisture contents, higher than 70–80% (wet basis), i.e., sufficient for the full hydration of the grain, are necessary. When much moisture migration occurs in metallic bins, intense condensation is produced, generally on the roof of silos, and grain sprouting at the top of farm silos or in railway wagons frequently occurs. (*See* **Water Activity**: Effect on Food Stability.)

Role of Microorganisms in Grains

The microorganisms which make up the microflora of cereal grain and their products are now well known in regard to their ecological requirements and the damages they are able to produce during storage.

As soon as the relative humidity in equilibrium with the grain is sufficient for the growth of microorganisms (with freely available oxygen), the greater the humidity and temperature, the more intense and rapid the activity of microorganisms (Table 1). Fortunately, there exists an absolute hydration limit below which microbial activity cannot take place. Expressed in terms of relative humidity of the air in equilibrium with the stored grain, this absolute limit is about 65% for all cereal species of interest for humans. It corresponds to a moisture content of about 13% (wet basis) for wheat at 25 °C. Below this limit, nothing of importance can occur in the kernel for many years, with the exception of possible weak evolution of the lipids, as explained below. Above this hydration limit, which is in fact very low, some mold species belonging to the genera *Aspergillus, Eurotium* (Figure 1) and *Monascus*, among others, can develop. With the extraordinary physiological ability to grow in extremely dry conditions, these molds invariably do develop unless additional barriers such as cooling, antifungal agents, or modified atmospheres are utilized. Such molds contaminate cereal grains all over the world, but it must be emphasized that the growth of these so-called xerotolerant or xerophilic species is very slow, and usually not accompanied by mycotoxin formation or significant deterioration of nutritional value. For their growth, molds utilize lipids and carbohydrates from the grain (as do humans) and produce carbon dioxide, water, and heat in exchange, following the simplified scheme shown in eqn (1).

$$\text{Grain components} + \text{oxygen from air} \quad \text{I} \quad \text{CO}_2 + \text{H}_2\text{O} + \text{energy} + \text{heat} \qquad (1)$$

The main consequence is a local increase in relative humidity and temperature and it is generally accepted, but not easily demonstrated, that this progressively allows the growth of molds with greater water requirements, such as *Penicillium* spp., other *Aspergillus* spp., or even *Fusarium* spp. These kinds of molds are able to destroy the grains and to produce extremely dangerous mycotoxins, such as the hepatotoxic aflatoxin B_1 or the immunosuppressive T_2 toxin. (*See* **Mycotoxins**: Occurrence and Determination.)

In practice, chiefly for economic reasons, grains are most often stored at moisture contents around 15–16% (wet basis); this allows storage without visible molding for some months, as shown in Table 1,

Table 1 Microorganisms able to grow and approximate safe storage periods for various moisture contents of wheat

Moisture (% wet basis)	*ERH (%)*	*Microorganisms*	*Safe storage period*
12	50	None	Indefinite
14	65	*Aspergillus glaucus*	Years
15	68	*A. candidus, Penicillium* spp.	Years
16	75	*A. ochraceus*, most penicillia	Year
17	77	Most storage fungi	Months
18	80	Some yeasts, some lactic bacteria	Month
20	85	Most yeasts, most fungi (*Fusarium* spp.)	Week
30	95	Most fermentative bacteria	Days

ERH, Equilibrium relative humidity. Data mainly from the cited literature.

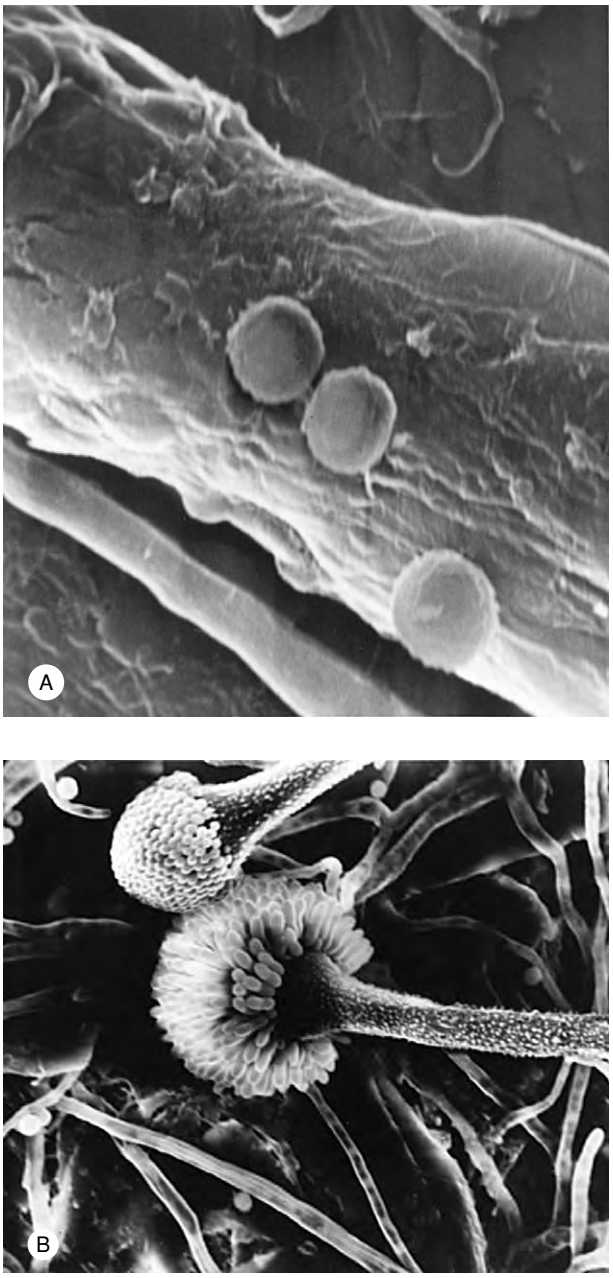

Figure 1 Scanning electron micrographs of *Eurotium chevalieri*, a xerophilic species common on stored grains: (A) × 1400, round conidia on the surface of the kernel; (B) × 280, conidial heads, responsible for vegetative multiplication and dissemination). Courtesy of INRA, LMTC, Nantes.

but also allows xerophilic species to develop slowly. At such hydration levels, a slight increase in water content, due for example to water vapor migration or to insect respiration in a limited pocket, can be sufficient to cause a significant acceleration in microbiological activity, leading to the formation of 'hot spots' in the bulk.

In hot spots, molds grow actively, producing further heat and water, leading to moisture content which is eventually sufficient to permit the growth of yeasts and bacteria. Grain dust and small particles accumulate at the center of the pile of grain in bulk silos, because grain tends to separate into heavier and lighter components when poured into a bin. Such accumulations of dust and broken kernels greatly increase the risk of hot spot formation by molds and insects, because they reduce the circulation of dry or cool air through the grain and promote the accumulation of heat and moisture. In such situations, it becomes urgent to cool and dry the grain by forced ventilation in order to avoid further problems throughout the bulk of the silo. (*See* **Spoilage**: Bacterial Spoilage; Yeasts in Spoilage.)

Role of Insects in Stored Grains

Among animal pests (i.e., mites, insects, rodents, and birds), insects cause by far the most damage to stored cereal. They are also the most feared because, even if they do not produce clearly toxic substances or strong off-flavors as do molds, their presence depreciates the value of grain even in cases of light infestations. Insects secrete some metabolic residues like quinones or uric acid that sometimes produce objectionable smells and tastes. They may also transmit pathogenic bacteria to the grain or leave fragments (including feces) in flours and other cereal products. Most consumers are strongly repelled by this filth, and one can say that, due to their contamination, insects ruin much more grain than they consume. (*See* **Insect Pests**: Insects and Related Pests; Problems Caused by Insects and Mites.)

Infestation of grain by insects is a constant phenomenon and several species are particularly well adapted to the grain ecosystem. They exist permanently in machines and buildings for crop reception and must always be regarded as a permanent, potential pest danger, leading to sytematic grain treatments by chemical contact insecticides or by fumigation.

Storage insects belong mainly to the orders Coleoptera and Lepidoptera. Most are adapted forms of crop-destroying species. The most injurious species belong to genera *Sitophilus, Rhyzopertha, Trogoderma, Sitotroga, Cryptolestes*, and *Tribolium*, and are very flexible and adaptable with only minimal nutritive requirements: for example, they tolerate movements of grain during intersilo transfer or temperature fluctuations induced by heating and ventilation. Most important species have a classical development cycle: the females lay fertilized eggs which lead to worm-like larvae, and then to adults through successive stages and a final metamorphosis via a pupal stage.

Adults usually live for several months, sometimes more than a year, depending on storage conditions.

Insects do not like low temperatures and cannot survive subzero temperatures for more than weeks. For this reason, damage caused by insects is generally only a secondary problem in cold or temperate climates.

Like molds, insects are aerobic organisms, only able to 'burn' with oxygen the dry matter they eat, producing carbon dioxide, heat, and water. They can initiate hot spots in dry grains, but such dry heating generally remains below 40 °C, which is the lethal temperature for most grain-infesting insects. By contrast, heating due to aerobic microorganisms can exceed 65–70 °C and can even lead to autocombustion of the bulk, when the moisture content becomes sufficient for the development of thermophilic fermentative bacteria and, subsequently, the occurrence of exothermic chemical reactions.

Despite their absolute requirement for oxygen, insects are able to stop breathing for several hours, days, and even weeks, depending on the development stage. However, they are also very sensitive to high concentrations of carbon dioxide and this is why it is possible to control insect infestations with efficiency by the use of modified atmospheres. (*See* **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs.)

They are remarkably tolerant to dryness, and commonly develop in grains with 11–14% water content (wet basis) but cereal weevils cannot multiply below 9% moisture content and adults cannot survive long in such dry grains. On the other hand, flour beetles seem to be able to develop in completely dry flour. At least some storage insects seem capable of reutilizing their own metabolic water (produced by respiration, as with molds) when surviving in extremely dry grains (below 6% moisture content).

Moisture Migration in Silos

In silos, elevators, and other installations for storage and transportation, the temperature fluctuates as a consequence of solar heating and daily or seasonal fluctuations in ambient temperature. The outermost layers of grain are most affected by such temperature variations. In a dark-green metallic silo, the wall temperature on the sunny side can frequently change from 15 to 80 °C on a sunny summer day in northern Europe. This produces measurable temperature changes in the outermost 30–40 cm of grain, and large oscillations occurring in the first 5–10 cm. Such fluctuations are less pronounced in concrete silos.

When heated, grain loses water vapor and dehydrates. Due to convective movements of the intergranular atmosphere, the lost water vapor is transported to the coldest zones in the bulk where it condenses, increasing the water content of the cold grain. The main consequence is always an increase in microbial activity.

In nonhermetic silos, when the intergranular atmosphere is not controlled, oxygen is always available, at least in the outer layers, for respiration of molds, which in turn produce heat and moisture and an acceleration of microbial activity, which may cause carbon dioxide to accumulate deeper in the bulk. This chain reaction always occurs, everywhere in the world, under all climatic conditions. The main difference from one place to another is the speed and intensity of this general phenomenon but, whatever the initial moisture content of the grain, sunshine will more or less rapidly destroy the stored cereals it previously allowed to produce. (*See* **Oxidation of Food Components**.)

The only way this process can be slowed is by removing heat and moisture using forced ventilation of the bulk, or by transfer of the grain from one bin to another to mix the hot and dry grain with cold and moist grain from other parts of the silo. Underground storage under airtight conditions, where temperature fluctuations and moisture migration are minimized, provides an interesting alternative to bin storage, especially when germinative capacity is not the main characteristic to be preserved.

Losses Due to Rodents

Many species of small rodents cause grain losses throughout the world. Rats and mice, which are found everywhere, are probably the most important, causing damage mainly at farms where grain is stored in loose structures. Control measures against rats and mice include traps, baits containing rodenticides, and fumigation similar to that used against insects. In many cases, it is important to know which kind of rodent is concerned in order to take its feeding preferences into account for poisoning, or to find the best way to trap it. Rodent control in larger grain stores requires a sound concrete construction that is able to resist rat gnawing, and the silo must be maintained in good condition. (*See* **Fumigants**.)

Biochemical Changes During Storage

For dry grains, i.e., with moisture contents permitting only very slow development of molds, slight changes in biochemical composition can be observed over several months. The lipid fraction of the grain deteriorates most rapidly during open-air storage under the influence of microbial enzymes, leading to the accumulation of fatty acids. Unsaturated acids can be further oxidized by the oxygen in air to produce volatile compounds with rancid odors and off-

flavors. When moisture contents exceed 15–16% (wet basis), amylases from both grain and microorganisms start to hydrolyze starch into dextrins and maltose, resulting in losses of dry matter and a deterioration in quality. Very small changes seem to occur in the protein fraction of stored grains when molds cannot develop significantly in a few months, i.e., below 15% moisture content. However, the aggregative properties of some proteins may be modified after several years of storage, leading to a decrease in the water-soluble fraction. This may result in slightly lowered digestibility or, for wheat, in losses in bread-making value.

In order to verify good storage of cereal grains, several simple laboratory examinations can be performed. Arguably the most sensitive is the germinative capacity of the grain, which clearly indicates the presence or absence of mold growth during storage in the open air. Fat acidity is another good criterion which correlates well with the metabolic activity of storage fungi and the quality of the grain after a long period of storage.

In practice, in business transactions, the most important criterion is to insure the absence of any live insect. In most cases, it is too often still necessary to fumigate the grain or treat it with insecticide before dispatching from the elevator. Modern consumers are more and more suspicious of chemical residues in food and the coming years will probably see a revision of the risk assessment of chemical residues and of the real benefit of insect destruction in grains.

See also: **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Fumigants**; **Insect Pests**: Insects and Related Pests; Problems Caused by Insects and Mites; **Mycotoxins**: Occurrence and Determination; **Oxidation of Food Components**; **Spoilage**: Bacterial Spoilage; Molds in Spoilage; Yeasts in Spoilage; **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability; **Water Activity**: Effect on Food Stability

Further Reading

Calderon M and Barkai-Golan R (eds) (1990) *Food Preservation by Modified Atmospheres*. Boca Raton: CRC Press.

Christensen CM (ed.) (1982) *Storage of Cereal Grains and their Products*, 3rd edn. St Paul: American Association of Cereal Chemists.

Multon JL (ed.) (1988) *Preservation and Storage of Grains, Seeds and their By-Products*. Paris: Lavoisier.

Highley E, Wright EJ, Banks HJ and Champ BR (eds) (1994) Stored products protection. *Proceedings of the 6th International Working Conference on Stored-products Protection*. Canberra: CAN International.

Handling of Grain for Storage

D Richard-Molard, INRA, Nantes, France

Background

The conservation of cereal grains is of utmost importance for the survival of human populations.

Different technologies have been used to conserve grain, such as storage in waterproof or airtight buried silos.

In the sixteenth century, in Europe, the first texts explaining the importance of drying, ventilation, or modified atmospheres for the good conservation of grains appeared, but the scientific bases explaining the reasons for these practices were absent.

Today, with the development of scientific knowledge on the grain ecosystem, and progress in the fields of building materials, sensors, and measurement techniques, grain storage technologies have been mastered.

However, with continuous, and increasing, requirements of technological and sanitary quality, particularly at the present time, increasing world trade of grains, and the resulting increased commercial competition mean that this area of industry is very dynamic and responsive to new innovative technologies.

Storage Facilities

Storage facilities may take many forms, from a simple pile of unprotected grain on the ground in a farm building to the most sophisticated terminal elevator designed to handle grain for long-distance exportation. At the farm level, grain may be stored temporarily in small concrete or round steel bins of small capacity (25–100 tonnes), which receive the crop immediately after harvest and before transport to a better long-term storage facility or in order to wait for the best market situation.

Grain may also be taken from the farm directly to a collecting center, a small storage installation with several small bins providing an average capacity of about 1000 tonnes. This does not generally have machinery for grain cleaning or drying, since the grain is usually transferred soon after to an elevator (**Figure 1**).

Country elevators (5000–50 000 tonnes), so called because they are filled with grain elevated into them by rolling belts with buckets, receive grain from the individual producer or from small collecting centers. Their capacity is adapted to the seasonal production of the area, and their main role is to keep crops in good condition before and during storage, and to

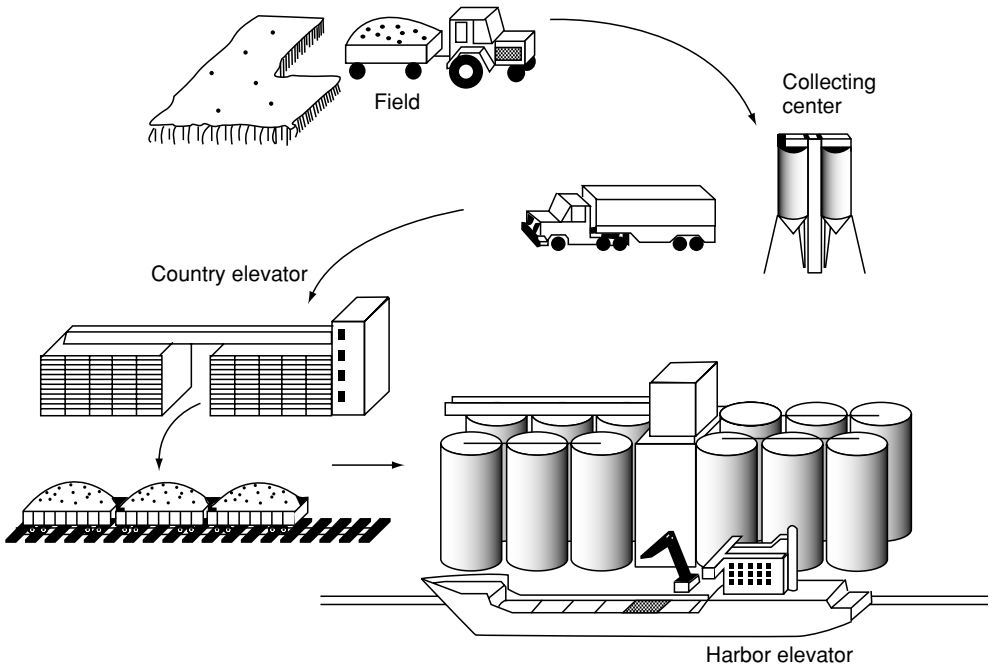

Figure 1 Grain collection, from field to terminal elevator, a long trip for the grain.

reload it into trucks or rail cars for transportation to terminal silos, export elevators, or industrial users.

There is no general rule for determining the best dimensions of a farm bin or a country elevator. Usually, the cost of storage increases sharply as the size of the silo decreases, so there is a tendency to build large structures. However, the grains are managed in multiple silos, where grains of different grades or water contents can be separated easily.

The terminal elevators are generally located close to trade centers and/or transportation terminals, such as harbors (Figure 2). Terminal elevators commonly offer storage capacities of 5000–500 000 tonnes in the USA or northern Europe.

Equipment for grain handling and control is often basic at the farm level, with only very simple devices for grain elevation and ventilation. By contrast, country and terminal elevators have handling capacities adapted to their transportation facilities and usually have sufficient capacity for peak load handling during the short period just following harvest. Today, most are fully equipped with remote control and automation, so that grain can be automatically transferred from one silo to another or to ships and railway cars. Weighing, cleaning, sampling, and even several grain sample tests are now electronically controlled.

When possible, gravity discharge from grain bins is ideal, but the cost of the corresponding structures and equipment is high. Consequently, flat storage of cereals is also commonly carried out in many situations, both at the farm level and in terminal elevators, to provide additional storage capacities when necessary. Such flat-bottomed silos are emptied through a central outlet, where a rotating screw-conveyor draws the grain away beneath the floor. Many other sophisticated devices have been developed for moving the grain using air blown through metallic networks of ducts below the bins, or with suction blowers as back-up devices.

Quality Maintenance During Storage

When the farmer brings freshly harvested cereals to the elevator, several tests are performed to determine the appropriate treatment for safe storage, such as cleaning, dust removal to avoid very dangerous dust explosions, drying, cooling, chemical treatments against insects, etc.

As previously emphasized, the major parameter to be controlled is water content. Even with mature dry grain harvested at about 12–13% moisture content, serious problems will be encountered within a few months if moisture cannot be properly managed to prevent the heat and moisture transfer that can occur in bulk during storage and transportation. It is, therefore, important to store grains with a sufficiently low initial moisture content to prevent mold development and biochemical changes, and to maintain this low temperature as constant as possible everywhere in the silo throughout the storage period. This can be achieved by artificially drying grains with excessive humidity at harvest, like maize in Europe, and by periodic forced ventilation during storage to remove

Figure 2 (see color plate 24) Harbor elevator in Rouen, France, with a total storage capacity of 300 000 tonnes. This is able to load grains into vessels at a rate of 5000 tonnes h^{-1}. Photo courtesy of J. Pfeiffer.

excess water where condensation has occurred. (*See* **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability.)

Moisture Levels for Safe Storage

Determination of the correct moisture level for safe storage of a given crop in a given situation and for a given storage time is difficult. Very general and theoretical rules can be given, but one should bear in mind that a large mass of cereal grain is never homogeneous in temperature, moisture, apparent density, thermal conductivity, etc., and so continuous monitoring with modern techniques, such as silothermometry, and periodical sample examination is highly recommended in every situation.

The thermodynamic activity of water, which represents the availability of water in the grain, is the most important physical parameter governing grain stability during storage, determining both biochemical changes and microbial growth. The availability can be estimated (and measured) by the equilibrium relative humidity (ERH) of the intergranular air in equilibrium with the grain. A nonlinear relation relates ERH to moisture content, as shown in Figure 3 for wheat and maize. This curve is named a 'sorption isotherm' and depends on both the average temperature and the biochemical composition of the grain. It differs from one cereal to another, but its relationship to the level of deterioration by different causes, such as microorganisms or enzymes, remains constant to a first approximation. Wheat stored at 16% moisture content is in equilibrium with an ERH of about 80% and will undergo the same kind of degradation at the same rate as maize with 15% moisture content, which also has an ERH of 80% at the same temperature. As the temperature increases, the sorption isotherm is shifted to the left, i.e., at the same water content, there is a higher ERH, and spoilage is faster. (*See* **Water Activity**: Effect on Food Stability.)

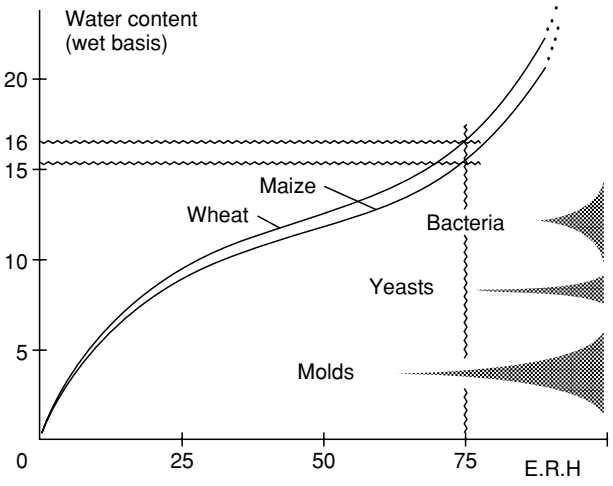

Figure 3 Sorption isotherms for wheat and maize at 20 °C and zones of activity for the main causes of grain degradation during storage (the wider the zone representing a cause of spoilage, the more intense and rapid the spoilage).

Silothermometry is a sophisticated technique, utilizing thermocouples placed in the silo able to detect slight variations in grain temperature, which indicate the beginning of deterioration through an aerobic process, resulting from a localized increase in humidity. Modern sensors can detect temperature changes

of less than 0.5 °C. In most situations, this gives sufficient sensitivity and overcomes the problem of the very low thermal conductivity of cereal grains, which allows temperature differences to be transmitted only over short distances in the bulk. With this method it is also possible to determine the best moment for drying and ensure the completion of grain cooling by nocturnal ventilation.

Drying Methods and Alternatives

The use of adequately dried grains for storage in silos is a mandatory condition (Figure 3). Many types of driers, such as full-bin driers, layer driers, continnous-flow driers, etc., are presently used on farms or in country elevators, but all utilize the same basic principle: heated dry air is blown through the grain to remove excess water. Figure 4 represents a farm bin equipped with a low-heat drier capable of drying grains with moisture contents of less than 20% (wet basis). (*See* **Drying**: Theory of Air-drying.)

When large quantities of moist grain are to be dried rapidly, high-temperature drying is necessary. This gives a higher extractive capacity but also can cause thermal degradation of the grain, if used improperly, as often happens with maize, which is sometimes harvested with a water content as high as 35% (wet basis) in Europe. If the air for drying is too hot, the functional proteins are denatured, and the value of the crop is decreased. However, high drying temperatures destroy all forms of insects in the bulk. (*See* **Drying**: Hygiene.)

When possible, a combination of different drying and cooling practices often gives the best technical and economic results. The essential principle is to decrease the moisture content rapidly to a level, thus decreasing the rate of mold development, i.e., to about 18% moisture in a first step, then to decrease the residual humidity to 15–16% by forced aeration when the ambient air is cool and dry enough to be efficient.

Natural Drying Methods

Alternative natural methods for grain drying have often been used, and at least two must be mentioned. The first, used mostly for maize, uses gentle drying by the wind of cobs stored outside in cribs or other forms of mesh silos. This is satisfactory, provided that the temperature remains sufficiently low to inhibit the growth of such toxinogenic fungi as *Fusarium* species.

The second type of natural drying method uses solar driers. These have been tested in efforts to save energy, but their use can be recommended only in regions where sufficient energy is provided by the sun to achieve drying in a time short enough to avoid spoilage by microorganisms. Unless drying to 80–82% ERH can be achieved in less than 7–10 days, the grain will certainly be damaged. (*See* **Drying**: Drying Using Natural Radiation.)

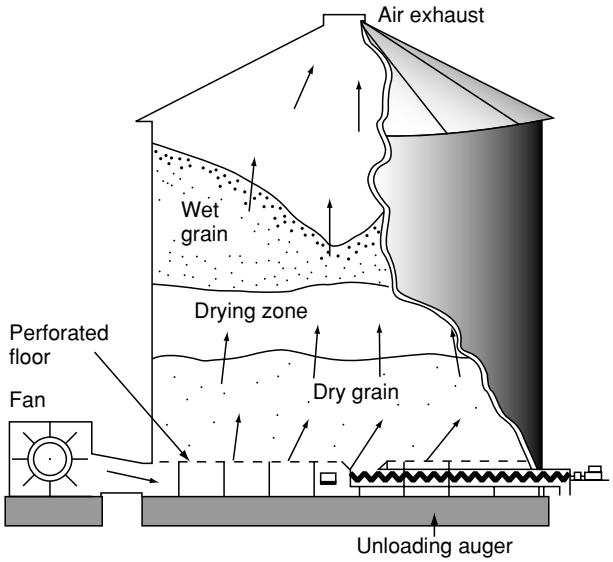

Figure 4 Farm bin for low-temperature drying.

Modified-atmosphere Storage

When cereals are not intended for uses requiring particular properties, such as a high germination ability for malt production or functional properties of proteins for breadmaking, it is possible to store them with a higher water content than usual, provided the oxygen is removed from the system.

Underground storage and storage in anoxic atmospheres in hermetically sealed silos, or silos continuously flushed with nitrogen, for example, are quite possible. For example, it was recently shown that the nutritional value of wheat for animal feeding could be successfully maintained over several months by storing it, after grinding, in airtight conditions at 21% water content and 15–20 °C. Storage of wet (35% water content or more) maize for pig feed in hermetic silos is now popular in several countries. Because of the important demand for intergranular oxygen at such a high humidity, especially through respiration of yeasts and lactic acid bacteria, the concentration of available oxygen remains extremely low in the silo. The growth of molds is inhibited and, at least until the silos are nearly empty, that is to say when free oxygen again appears and concentrations of carbon dioxide are decreasing, no mycotoxins can be produced. With such hermetic silos, it is not necessary to grind the humid grain in order to produce silage, and the commodity can be handled more easily than with traditional practices. (*See* **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Spoilage**: Molds in Spoilage.)

Control of Insects

Storage of dry or wet grains under modified atmospheres is an excellent way to kill insects, and modern techniques for storage of dry wheat under high concentrations of carbon dioxide have yielded very good results in Australia in recent years. Most often, if all living insects are to be killed, chemicals need to be used when infestations are detected in a silo or a ship. (*See* **Insect Pests**: Insects and Related Pests; Problems Caused by Insects and Mites.)

Insecticides can be divided into two main classes: contact insecticides, which kill insects and prevent reinfestation owing to their remanent effect, and fumigants, which destroy insects without leaving any significant residues. Organochlorine insecticides, which belong to the first class, were extensively used in the past but are now forbidden because the toxicity of their residues is very high for humans.

Pyrethroids like deltamethrin, permethrin bifenthrin, and organophospates like malathion, dichlorvos or methyl pirimifos, are also contact insecticides, used at concentrations of about 4–8 g of active substance per tonne of grain. They are very fast-acting and highly toxic by contact and ingestion, but their residues necessitate careful consideration of the number and timing of treatments. Their relative inefficiency against the hidden forms of insect (eggs, larvae) that live within the kernels should also be mentioned. (*See* **Pesticides and Herbicides**: Residue Determination.)

The only current fumigant is hydrogen phosphide, which is used to rapidly kill all live insect stages, including adults. It is used in airtight structures equipped with systems that allow the gas to be introduced and removed safely. (*See* **Fumigants**.)

Apart from modified-atmosphere storage under nitrogen or carbon dioxide, only persistent insecticides can provide good long-term protection for grain. It is then of primary importance to ensure that there is no residue in byproducts like flour and brans used for human and animal food. At the present time, there is a general tendency to decrease the concentrations of toxic substances used against insect infestations. A possible approach would be to combine the use of carbon dioxide with an active substance, such as methyl bromide or hydrogen phosphide, which would permit a considerable reduction in the dose of fumigant used because of synergism with carbon dioxide, which increases the penetration of the active substance in the respiring stages.

The Future

All over the world, safe storage of cereal grains is a vital but expensive activity. Too many cereal grains are still used in the food industry despite the poor quality resulting from inappropriate storage conditions. Current scientific and technical knowledge in this domain is sufficient to answer the main questions about handling and storage of grain, but new trends are appearing: consumer demand for safe products, i.e., free of chemical residues, is growing fast. This probably means that the use of contact insecticides and even fumigants should be replaced by more acceptable techniques like physical techniques in the coming years.

Another trend seems to be a clear demand for identified grains, nongenetically modified grains, and grains from a precise geographic origin or a well-specified technological quality, for example. Such an evolution would certainly introduce very important changes in the way in which the grains are harvested, collected, stored, and distributed. Loading speeds as high as 5000 tonnes per hour are often reached today in harbor elevators, but this is not the most convenient technology if small bulks of 50–100 tonnes of well-certified grains are to be prepared.

See also: **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Drying**: Theory of Air-drying; Drying Using Natural Radiation; Hygiene; **Fumigants**; **Insect Pests**: Insects and Related Pests; Problems Caused by Insects and Mites; **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability; **Water Activity**: Effect on Food Stability

Further Reading

Christensen CM (ed.) (1982) *Storage of Cereal Grains and their Products*, 3rd edn. St. Paul, MN: American Association of Cereal Chemists.

Navarro S and Donahaye E (eds) (1993) Controlled atmosphere and fumigation in grain storages. In: *Proceedings of the International CAF Conference, Winnipeg, Canada, June 1992*. Jerusalem: Caapit Press.

Richard-Molard D (1990) Conservation of humid grains in controlled atmosphere storage. In: Calderon M and BarkaiGolan R (eds) *Food Preservation by Controlled Atmosphere*, pp. 57–82. Boca Raton, FL: CRC Press.

Breakfast Cereals

R S Kadan, United States Department of Agriculture, New Orleans, LA, USA
E F Caldwell, American Association of Cereal Chemists, St Paul, MN, USA

Introduction

The US breakfast cereals (BC) industry, in its history of over 110 years, has emerged as an important segment of the food industry, with annual sales of about \$9.5 billion (9.5×10^9). This remarkable progress has been achieved through a combination of knowledge gained in human nutrition, innovative progress in cereal processing technology, and imaginative marketing. The chief driving force has been the continued market for new, convenient, nutritious, and fun foods for a discriminating, affluent, and mobile society. Any large grocery store in the USA, Canada, or the UK is likely to stock 100 or more different BC. In the UK, annual sales are about £1 billion (1×10^9). The relatively lower penetration of BC into other European and other developing countries increased significantly during the 1990s, and this is forecast to continue over the next decades.

BC have been defined as processed grains for human consumption. The major grains used in the manufacturing of BC are corn, rice, wheat, oats, and barley. Cereals can be divided into those that are ready-to-eat (RTE) or ready-to-cook, also known as hot cereals (HC). Alternative breakfast products based on cereal grains include cereal bars, toaster pastries, waffles, French toast, and bagels. RTE cereals are frequently made from mixtures of one or several grain components with other ingredients; they require extensive processing, are usually fortified with vitamins and minerals, and are specially packaged to protect their flavor, texture, and nutrition during storage. On the other hand, HC are typically made from a single grain component by using relatively simple processing and packaging technologies. The difference between the RTE and HC products is reflected in the price at the consumer level. The typical single serving (28–35 g or 1 oz) of RTE cereal costs about \$0.20 (\$0.08–0.40) in the USA and represents about 90% of the business. A typical 28–35-g or 1-oz serving of HC costs from about \$0.06 to \$0.24, for bulk-packaged product to instant single-serving pouches, respectively or up to \$0.30 for instant farina.

Annual sales in the USA of alternative breakfast products amounted to over \$2 billion (2×10^9) by 1999, but the per capita annual consumption of RTE cereals had still increased from 3.8 to 5.0 kg (8.3–11.0 lb) since 1970. On the other hand, HC per capita annual consumption had declined from 1.1 to less than 0.6 kg (2.4–1.2 lb) over the same period. The latest per capita annual consumption in UK was 5.9 kg (13.0 lb) and 0.6 kg (1.4 lb) for RTE and HC respectively. (*See* **Barley; Oats; Rice; Wheat:** The Crop.)

Methods of Manufacture

Ready-to-eat Cereals

RTE breakfast cereals are primarily grain formulations, suitable for human consumption without further cooking. Manufacturing technology varies according to type of cereal.

Flaked cereals Flaked RTE breakfast cereals can be divided into two groups: flakes made from whole grains or parts of the whole grains, and flakes made from more finely ground materials that must be first extruded into pellets that can then be rolled into appropriate-size flakes.

The whole grains (wheat and rice) or major components (such as grits from degermed yellow maize or corn) are cooked with flavorings such as sugar, salt, and malt; they are then dried and tempered to a firm but slightly plastic state, flaked by passing between rolls, and toasted or dried to a final specified moisture content. Cooking of grits or whole grains for traditional flakes is usually done in batches. The moisture content of the cooked mass at the end of cooking is usually about 28%. The cooked material should not be mushy, soft, or sticky. After cooking, the mass of material is cooled, dried at about 120 °C, and tempered (held) for several hours at a final equilibrated moisture content of 10–18%, depending upon the grains. The tempering process is important, not only to allow equilibration of moisture within the cooked grains or pellets, but also to allow sufficient starch recrystallization or retrogradation to provide a suitable texture for flaking (or for shredding, as discussed next).

Flaked RTE cereals can also be made from extruded pellets rather than cooked grits or whole grains. A variety of floury or finely ground grain products such as whole wheat or oat flour can be mixed with sugar, salt, and malt syrup or other flavoring and coloring ingredients to form a dough. This is then extruded to form pellets about the same size as cooked grits or whole grains, the cooking having taken place in the extruder. Pellet moisture is in the range of 18–24%. Conditioning or tempering prior to flaking may or may not be required.

Whether consisting of cooked grits or whole grains or of extruded pellets, the tempered material is then

flaked. Tremendous pressures are necessary to flatten the prepared material into thin flakes, which traditionally have been toasted by keeping them suspended in hot air between 270 °C and 330 °C for about 90 s. Some properties of the finished flakes can be modified by a heating or steaming step immediately before flaking (e.g., infrared heating immediately before flaking will increase the tendency of flakes to crinkle rather than remain flat), while other properties, such as blistering, color, crispness, tenderness, and flavor of the final product, are influenced by toasting conditions. The moisture content of the finished product is about 1–3%.

Puffed cereals The predominant grains used for making puffed cereals are wheat and rice. The milled rice used for puffing is either short- or medium-grain white rice. Such rice usually requires no pretreatment other than the hull and bran removal that occurs during milling or pearling. Wheat, on the other hand, is usually decorticated (bran removed), either by pretreatment with concentrated brine solution or by pearling. In the pearling operation, the grain is passed through a revolving cylinder inside which are mounted silicon carbide or carborundum stones.

Usually hard wheat, preferably durum, is used for puffing. A grain flour mixture can also be formed into desired shapes by an appropriate extrusion process and dried to a moisture content of 9–12% for subsequent gun-puffing.

Puffing guns are vessels capable of holding very high temperature and high-pressure steam. The critical feature of any puffing gun vessel or barrel is a quick opening lid that seals the vessel. Operation of a simple single-shot gun consists of placing the grain in the vessel, closing the lid and raising the pressure and temperature in the vessel to about 1380 kPa (200 psi). The quick-opening lid is then activated, causing rapid pressure drop, which in turn instantaneously causes moisture in the grain to 'flash off' as steam, and the kernels of grain or preformed pellets to expand to as much as 16 times their original volume. The guns can also be multiple-shot automatic or continuous in operation.

Another group of BC is made by oven puffing. Specially prepared rice or corn grains or their mixtures are oven-heated to 290–340 °C in order to puff them. Puffed rice cereal sold in the USA can be either oven-puffed (e.g., Rice Krispies, prepared from short-grain rice) or gun-puffed (e.g., Puffed Rice). Whether gun- or oven-puffed, the grains are then screened to remove any unpuffed kernels, loosened bran, fine dust and broken kernels, and the product is heated to lower the final moisture to 1–3%. Processes such as coating or vitamin enrichment can be carried out during this drying operation.

Shredded cereals Shredded wheat, one of the oldest mass-produced breakfast cereals, is still an important product. Soft white wheat is the primary grain used. Other varieties of wheat, as well as other cereals such as rice, corn, and oats or mixtures thereof, can also be used to make shredded products. The grains are cleaned, cooked in boiling water, partially dried, and tempered. The shredding operation consists of passing the tempered grain between two rolls. One roll is grooved and the other is smooth. Strong forces in the range of 7000 kPa (1000 psi) along the roll length are often required to maintain the rolls in tight contact. The tempered kernels are squeezed into the grooves of the roll and emerge as strands roughly 1 mm in diameter. These strands are accumulated in multiple layers below multiple pairs of rolls, and formed into biscuits or bite-sized pieces. A refinement of the shredding process is the use of a roll with added cross-grooves to provide lateral strands that can tie the shreds together in a coherent sheet. Mixtures of ground cereal products with sugar, salt, malt, and flavoring can be cooked by extrusion, formed into pellets, tempered, and then shredded in the same way as cooked whole grain kernels.

The formed pieces are oven-dried and baked to the desired color, texture, and moisture. The baking procedure is done in a specialized manner to impart fluffy texture, particularly for products containing rice or corn. Other minor ingredients, such as flavorings, nutrients, or antioxidants, can also be added prior to drying and baking.

Granola cereals The main grain used to make granola cereals is either regular or quick-cooking rolled oats. The cereal products are mixed with other desired ingredients such as nut pieces, coconut, corn syrup, brown sugar, honey, malt extract, dried milk, dried fruits, and/or vegetable oil. Liquid and dry ingredients are prepared separately, then mixed together and spread in a uniform layer on the band of a continuous oven. Baking temperatures are in the range of 150–220 °C, until the product is uniformly toasted to a light brown color and the moisture is reduced to about 3%. After baking, the product is broken into small pieces and packaged.

Extruded expanded cereals Successful application of extrusion technology for producing cereal products from grain flours or meal mixtures by formulating, continuous preconditioning, extrusion cooking, and forming has been an important achievement of the BC processing industry. The extruders can be single-screw, twin-screw, low-shear, high-shear, and combinations of these or other features. Extrusion technology was developed to produce flaked, puffed,

shredded, and expanded breakfast cereals. For extruded expanded cereals, cereal flour mixtures having about 25–30% moisture are preconditioned, cooked, and extruded. Perforations in the die and design of the cut-off knives at the end of the extruder, as well as temperature, pressure, time, and shear in the extruder barrel, all influence the shape, size, and texture of the final product. The sudden change from high pressure and temperature to ambient conditions as the product exits the extruder imparts the necessary expansion, texture, and other characteristics. The extruded products may be coated with solutions of vitamins and minerals, and with sweetening and flavoring materials, and are then oven-dried to desired moisture levels.

Hot Cereals

HC are products of a single grain or a simple mixture that require cooking or heating in water before consumption. The main such products are rolled oats and related oat products, farina and other wheat fractions, and corn grits.

After cleaning, oat kernels are dehulled by an impact process. The resulting groats are steamed, primarily to deactivate lipolytic enzymes, then rolled, dried, and packaged. Quick-cooking rolled oats are made by steel-cutting groats into four or five pieces before the steaming and flaking process. Instant oatmeal is made by subjecting groats to a special process that results in rapid cooking of the flakes. Incorporation of 0.1–1.0% of an edible gum during processing helps to achieve this objective. Distribution of the gum and necessary salt without damaging the extra thin flakes is accomplished by portion-control packaging. Such instant products require only the addition of boiling water, stirring, and a short standing time to make the product ready for consumption. The industry has successfully incorporated dried fruits, nuts, and other flavorings to make instant oatmeal products more attractive to consumers.

Wheat farina is another important HC. In the USA, farina and enriched farina have to meet respective legal standards of identity. Farina is basically wheat endosperm, preferably obtained from hard red spring or winter wheat, the granules of which stay intact during cooking at home. Farina is normally obtained by drawing off chunks of endosperm during milling of wheat into bread flour. In the UK it can also be called semolina, but in the USA that term is reserved for a similar product from durum wheat intended for macaroni and other pasta products. Farina is heated at around 60 °C for 15 min before packaging, for flavor and to head off infestation. Quick and instant farina products are prepared by the use of disodium phosphate, enzymes, and other treatments. Similarly, wheat germ, bran, malted barley, cocoa, and other flavorings can be mixed into farina to enhance the consumer appeal.

Other important HC products include wheat products such as cracked wheat and bulgur, and corn grits. Grits and bulgur are generally used as cereal accompaniment to other meal menu components such as eggs, sausage, meat and related preparations, whereas cracked wheat products are cooked for breakfast. Grits usage is mostly in southern parts of the USA, garnished with butter and salt rather than with sugar and milk as in traditional breakfast cereals. Grits are produced by dry milling of white degermed corn and can be fortified with vitamins and minerals. Bulgur or parboiled wheat is one of the oldest cereal-based foods and has been consumed for centuries in Turkey, Syria, Jordan, Lebanon, and Egypt. To prepare it, common hard or soft wheat varieties are soaked in water and heated to about 90 °C for 2–3 h. The cooked wheat is then cooled, dried, moistened, peeled (optional), redried, cleaned and sized. Bulgur has a good shelf-life because cooking destroys insects present in harvested crops.

Significance in the Diet and Fortification

BC are one of the most highly fortified foods in both vitamins and minerals. One serving typically can supply 25% of the daily values of most vitamins and trace minerals as these are defined in the USA for labeling purposes. A few BC supply 100% of all established daily values, making them as much nutritional supplements as breakfast foods. The cereal industry justifies these practices on the basis that they fulfill a genuine need to improve the overall nutritional needs of consumers. Food consumption surveys have shown that some of the US population had inadequate intakes of several vitamins and minerals, particularly vitamin A, folic acid, calcium, and iron, and possibly vitamin B_6, zinc, and magnesium. (*See* **Calcium**: Physiology; **Food Fortification**; **Iron**: Physiology; **Magnesium**; **Retinol**: Physiology; **Vitamin B6**: Properties and Determination; **Zinc**: Physiology.)

The practice of fortification of cereal foods started in 1941, when the US Food and Drug Administration (FDA) established a standard of identity for enriched flour to include the addition of thiamin, niacin, and iron. Riboflavin was subsequently included, and addition of calcium and vitamin D was optional. Such recommendations developed after extensive discussions by the various committees of the American Medical Association, hearings sponsored by the FDA, and the Food and Nutrition Board. As a result of these meetings, criteria for fortification of foods were developed:

1. The intake of nutrient is below the desirable level in the diet of a significant number of people.
2. The food used to supply the nutrient is likely to be consumed in quantities that will make a significant contribution to the diet of the population in need.
3. The addition of nutrient is not likely to create an imbalance of essential nutrients.
4. The nutrient added is stable under proper conditions of storage.
5. The nutrient is physiologically available from the food.
6. There is reasonable assurance against excessive intake to a level of toxicity.

Various studies have confirmed that fortification of BC foods is appropriate and has contributed significantly to improving overall nutrition. Fortification of flour and other grain-based foods with folic acid became a requirement in the USA in 1998 because of evidence that it may help prevent neural tube defects.

Many nutritionists consider eating breakfast regularly as one of the important health habits associated with subsequent favorable health status and reduced mortality. Breakfast consumers were found to do significantly more work in the morning than those who did not. Children and the elderly (above 65 years old) are the largest consumers of BC. Half of the school-age children in the USA prepare their own breakfast because most of their mothers work outside their homes. These groups are more likely to skip breakfast. Studies have indeed confirmed that BC consumers are likely to skip fewer breakfasts and that they obtain higher levels of vitamins and minerals than others. Most BC also contain up to 5 g of dietary fiber per serving. Some high-fiber products can provide 10 g or more of fiber and thus help consumers to improve their dietary fiber intake and to meet recommendations of 25–30 g daily dietary fiber intake. Oat-based BC containing sufficient beta-glucan soluble fiber to provide 3 g or more of that substance daily, or other BC containing sufficient psyllium to provide 7 g of soluble fiber from that source daily, have been shown to decrease blood cholesterol by a small but significant amount and in the USA may carry label health claims to that effect.

BC, with few exceptions, are not high in protein. The protein in some cereals, such as gun-puffed or flaked products, may have little or no biological value, due to excessive Maillard browning. Some RTE cereals may also be high in sugar and/or salt (sodium) and may contain added saturated fat. Nevertheless, BC with milk and fruit make an inexpensive, nutritious, and easy-to-prepare breakfast. (*See* **Browning**: Nonenzymatic; **Dietary Fiber**: Properties and Sources.)

Packaging, Storage, and Spoilage

The BC industry, over the years, has developed efficient processing, packaging, storage, and shipping technologies to supply consumers with cereal foods that routinely have 6 months' or longer shelf-life. It is indeed quite a feat to have such an array of vitamins and minerals, which often interact with each other, in a highly processed food and yet end up with a wholesome product with a desirable flavor and texture. In the process, the industry has developed novel techniques for vitamin preservation and for the incorporation of antioxidants directly or via the packaging material. BC packages not only protect the product against moisture, oxygen, and insects during shipping and storage, but also provide attractive consumer appeal at the point of purchase. The industry has accomplished this by following good manufacturing practices and other quality control measures, and by instituting Hazard Analysis and Critical Control Point (HACCP) systems for food safety purposes. Computer control of processing and attention to environmental issues also characterize most BC manufacturing operations, and certainly the newer ones.

Future Developments

The Nutritional Labeling and Education Act of 1990 and subsequent amendments in the USA undoubtedly contributed to the reevaluation of product development, marketing, and advertising strategies by the BC industry as a result of stricter control of nutrient claims, ingredient listing, and other label data. Among other moves, the BC industry has endeavored to develop new products made from whole grains that are also low in fat, sugar, and salt (sodium). Encouragement was provided by FDA approval of a health claim being made for products containing one or more whole grains as a majority constituent. Other product developments or modifications of BC in the USA or elsewhere are likely to be influenced by such regulatory developments as well as by advances in technology or nutritional knowledge.

See also: **Barley**; **Browning**: Nonenzymatic; **Calcium**: Physiology; **Dietary Fiber**: Properties and Sources; **Food Fortification**; **Iron**: Physiology; **Magnesium**; **Maize**; **Oats**; **Retinol**: Physiology; **Rice**; **Wheat**: The Crop; Grain Structure of Wheat and Wheat-based Products; **Zinc**: Physiology

Further Reading

Fast RB (1999) Origins of the US breakfast cereal industry. *Cereal Foods World* 44: 394–397.

Fast RB and Caldwell EF (eds) (2000) *Breakfast Cereals and How They Are Made*, 2nd edn. St Paul, Minnesota: American Association of Cereal Chemists.

Hoseney RC (1994) *Principles of Cereal Science and Technology*, 2nd edn. St Paul, Minnesota: American Association of Cereal Chemists.

Slade L and Levine H (1991) Beyond water activity: recent advances based on an alternative approach to the assessment of food quality and safety. *Critical Reviews in Food Science and Nutrition* 30: 115–360.

Dietary Importance

S O Serna Saldivar, Instituto Tecnologico y de Estudios Superiores de Monterrey, Monterrey N.L., Mexico

Introduction

Cereal-based foods are by far the major source of energy, protein, B vitamins, and minerals for the world population. In most countries, diets have a single cereal as the primary staple. The most widely used cereals are rice, wheat, and maize, which provide 93% of the total cereal calories. These grains constitute the main staple for Asia, Europe, and America, respectively. In Africa and India, sorghum and millets are widely grown and consumed. (*See* **Maize**; **Rice**; **Sorghum**; **Wheat**: The Crop.)

According to the Food and Agriculture Organization (FAO), the amount and proportion of food energy and protein provided by cereals in human diets in 1997 were 1384 kcal (5796 kJ), nearly 50% of the average per capita caloric intake. Likewise, cereals provided 33.8 g of protein of the total estimated daily intake of 73.9 g of protein. Cereal grains are considered primarily as caloric or starchy foods and, more recently, as an important source of dietary fiber. However, their protein quality, especially for infants, is marginal. In developing countries, cereals are usually consumed with legumes and pulses, thus significantly increasing protein intake and, more importantly, complementing their essential amino acid composition. Protein and/or calorie malnutrition (i.e., marasmus, kwashiorkor) is prevalent among groups of people who have inadequate food intake or rely solely upon cereals and/or starchy roots as their source of protein. (*See* **Energy**: Measurement of Food Energy; **Kwashiorkor**; **Malnutrition**: The Problem of Malnutrition; **Marasmus**.)

Plant breeders have developed maize, sorghum, and barley genotypes with high lysine, which results in an improved protein quality and nutritional value. A high lysine and tryptophan corn, called quality protein maize, with satisfactory grain yield and structure is being commercially grown in different places around the world (Brazil, China, Ghana, and other African countries). (*See* **Maize**.)

Carbohydrates

Starch is the most abundant carbohydrate in cereals and the main contributor of calories (**Table 1**). It is composed of molecules of branched amylopectin and linear amylose. Most cereals contain 70–75%-amylopectin. Waxy maize, rice, and sorghum contain more than 95% amylopectin. Most food processes partially or totally gelatinize the starch granules, making the molecules more prone to amylase hydrolysis. (*See* **Starch**: Structure, Properties, and Determination.)

Starch is highly digestible by the human digestive system. Practically all starch disappears in the gastrointestinal tract. A small portion of the starch (1–5%) resists enzymatic hydrolysis when cereal foods are thermally abused. This residual starch can be quantified in the soluble dietary fiber residue and is highly susceptible to fermentation in the large intestine. (*See* **Starch**: Resistant Starch.)

Other factors that can decrease starch digestibility are excessive amounts of fiber and enzyme inhibitors (i.e., tannins). The digestible energy of cereals is negatively and positively related to the grain's fiber and lipid contents, respectively. Refined grains contain more digestible energy than whole products (**Table 1**). As a result of its high fat and low fiber content, maize contains high quantities of digestible energy.

Cereals contain small quantities of soluble carbohydrates (i.e., glucose, fructose, maltose, sucrose); they are therefore suitable foods for diabetics. Highly refined cereals have a higher glycemic index than whole-grain products. (*See* **Glucose**: Glucose Tolerance and the Glycemic (Glycaemic) Index.)

Whole cereal grains are considered a rich source of fiber (**Table 1**). However, foods from grains have marked differences in the amount and type of dietary fiber. Fiber components are concentrated in the outer tissue of the kernel (pericarp). The dietary fiber content in cereal-based food varies greatly, depending on the extent of milling. Refined flours lose most of their fiber during decortication and milling. All cereals are

Table 1 Nutrient composition of cereal grains[a]

Proximate composition									*Dietary fiber*	
Cereal	*Protein (%)*	*Crude fat (%)*	*Crude fiber (%)*	*Ash (%)*	*NFE (%)*	*Starch (%)*	*Digestible energy*		*Total (%)*	*Soluble (%)*
							kcal kg^{-1}	*kJ kg^{-1}*		
Wheat										
Hard	14.4	2.3	2.9	1.9	78.5	64	3 865	16 181	12.1	1.7
	11.5–17	1.8–2.8	2.8–3	1.8–2	75.2–82.1					
Soft	9.9	2.8	2.7	1.7	82.9	69	3 865	16 181	12.1	1.7
	8–12	2.6–2.9	2.5–2.8	1.8–1.9	80.4–85.1					
Durum	13.2	2.8	2.8	2	79.2	70.2	4 056	16 981	12.1	1.7
	12–15.6	1.8–3.8	2.4–3.1	1.8–2.1	75.4–82					
Rice										
Paddy	7.5	2.4	10.2	4.7	75.2		2 821	11 810		
	6.7–9	1.7–2.7	8.4–12.1	3.4–6	70.2–79.8					
Brown	9.2	2.5	0.9	1.5	85.9	77.2	4 321	18 091	3.7	0.9
	8.3–10.1	1.8–3.3	0.7–1.2	1.2–1.8	83.6–88					
Milled	7.8	0.5	0.4	0.6	90.7	90.2	3 938	16 488	1.3	0.4
	7.3–8.3	0.3–0.6	0.2–0.6	0.3–0.9	89.6–91.9					
Maize										
Dent	9.1	4.4	3	1.7	81.8	71.8	4 056	16 982	12.8	1.1
	8.1–11.5	3.9–5.8	2.4–3.5	1.4–2	77.2–84.2					
Flinty	11.1	4.9	2.2	1.7	80.1		4 056	16 982		
	9.5–12.8	4–5.8	1.6–2.8	1.4–2	76.6–83.5					
Popcorn	12.1	4.6	2.3	1.8	79.2	62.3			13.1	0.4
	11–13.2	4–5.3	2.2–2.4	1.6–1.9	77.2–81.2					
Sweet	13.2	4.6	2.7	2.3	77	54.1	4 145	17 354	9.4	1.2
	12.1–14.2	3.7–9	2.2–3.2	2–2.7	70.9–80.1					
Barley	11.5	2.2	5.6	2.9	77.8	58.5	3 543	14 833	15.4	3.9
	7.5–15.6	1.8–2.6	5.3–5.9	2.6–3.1	72.8–82.8					
Rye	13.4	1.8	2.1	2	80.7	68.3	3 794	15 885	16.1	3.8
	12.6–14.5	1.6–2.2	1.6–2.6	1.7–2.2	78.5–82.5					
Oats										
Whole	17.1	6.4	11.3	3.2	62	52.8	3 058	12 803		
	12.4–24.4	4.5–10.3	10.4–14.3	2.9–3.4	47.6–69.8					
Groats	16.9	7.4	1.6	2.1	72	65	4 316	18 070	12.5	6.6
	13.8–22.5	5.9–8.4	1–3.3	1.9–2.4	65.2–77.4					
Sorghum	11	3.2	2.7	1.8	81.3	73.8	3 880	16 245	11.8	1
	7.3–15.6	0.5–5.2	1.2–6.6	1.1–4.5	68.1–89.9					
Millets										
Pearl	14.5	5.1	2	2	76.4	60.5	2 822		7.0	0.6
	8.6–19.4	1.5–6.8	1.4–7.3	1.6–3.6	62.9–86.9					
Finger	8	1.5	3	3	84.5	59				
	6–10.9	1–4.6	2–6.8	2.3–3.9	73.8–88.7					
Italian	11.7	3.9	7	3	74.2	59.1	3 395			
	6–14	1.2–5.2	2.6–8.6	1.5–3.6	68.3–88.7					
Proso	11	3.5	9	3.6	72.9	56.1	3 636			
	6.4–12.8	2.9–4.9	4.6–12	1.4–5	65.3–84.7					
Japanese	11.8	4.9	14.3	4.9	64.1	62				
	11.2–12.7	2.5–6.3	13.9–14.7	4.7–5	61.3–68					
Fonio	8.7	2.8	8	3.8	76.6					
	5.1–10.4	2.1–5.2	4.6–11.3	1.8–6	67.1–86.4					
Kodo	10.4	3.7	9.7	3.6	72.6	72				
	6.2–13.1	3.2–4.9	8.4–11	3–4.1	66.9–79.2					
Teff	10.9	2.4	2.4	2.2	82.1					
	7.9–12.6	2.3–2.5								

[a]All values are expressed on a dry-matter basis. Protein conversion factors: for wheat, 5.7; rice, 5.95; other cereals, 6.25; NFE = nitrogen-free extract.

considered a rich source of insoluble dietary fiber (i.e., cellulose, insoluble hemicellulose). Insoluble fiber binds water, speeds up intestinal transit and binds some carcinogens. Epidemiological evidence has related a high-fiber diets to a lower incidence of diabetes, obesity, and diverticular disease. A high dietary fiber consumption benefits diabetics because of the reduced diffusion of glucose in the intestinal mucosa and the diminished insulin secretion rate. (*See* **Dietary Fiber**: Properties and Sources; Determination; Effects of Fiber on Absorption.)

Oats, barley, and rye are considered goods sources of soluble dietary fiber (i.e., pentosans, β-glucans, soluble hemicellulose), which has been associated with increased excretion of bile salts and dietary cholesterol, thus lowering blood cholesterol levels.

Protein

Cereal grains provided 45% of our total protein intake in 1997. Genotype, environment, and growing conditions affect the amount of protein in the kernel. In general, brown rice and oats contain the lowest and highest protein content, respectively (Table 1). Protein quality is mostly dictated by the amino acid profile and digestibility (Table 2). The apparent protein digestibilities in cereals range from 80% to 90%. Sorghum (especially kernels with condensed tannins), whole barley, rye, and oats are consistently lower in digestibility than other cereals. Milling, decortication, fermentation, and germination all increase protein digestibilities because of the removal of fiber and the enzymatic breakdown of proteins. (*See* **Protein**: Chemistry; Quality.)

The storage protein in cereals (prolamines) contains small amounts of essential amino acids, especially lysine. The albumin and globulin proteins, mainly located in the germ, contain the best essential amino acid profile. High-lysine cultivars of maize, sorghum, and barley are genetically modified to contain lower amounts of prolamines and higher amounts of the other protein fractions. Thus, they have a more balanced protein. (*See* **Amino Acids**: Metabolism; **Protein**: Quality.)

For all cereals, the most limiting amino acid is lysine. Oats, rice, rye, barley, and the high-lysine cultivars contain a more favorable essential amino acid composition or amino acid score than the rest of the cereals (Table 2). Lysine deficiency is observed particularly in postweaned infants who suffer from kwashiorkor and rely on cereals and/or starchy tubers as the only food source. The supplementation of cereal-based foods with small quantities of legumes or animal foods improves protein intake and protein quality. The next most limiting amino acids for maize and the rest of the cereals are tryptophan and threonine, respectively.

Decortication or milling partially or totally removes the pericarp and germ tissues, therefore improving the digestibility and reducing the lysine level. Malting improves protein quality due to enzymatic degradation of proteins. Fermentation improves protein quality via improved digestibility and *de-novo* synthesis of lysine by the fermenting biomass.

Lipids

Lipids are relatively minor constituents in cereal grains (Table 1). However, cereal lipids are rich in the essential fatty acid, linoleic acid (18:2 Δ 9, 12; 30–60% of total fatty acids), and practically devoid of saturated fatty acids. Cereals contain trace quantities of phytosterols; they do not have any cholesterol. Cereals with yellow endosperm (i.e., yellow corn, yellow sorghum, durum wheat) contain some provitamin A activity (β-carotenes). Various types of tocopherols are responsible for the vitamin E activity of cereal grains. Degermination of cereal grains greatly reduces the content of lipids and tocopherols.

Minerals and Vitamins

The pericarp, germ, and aleurone layer are rich in vitamins and minerals. Refined cereal products therefore lose part of these important nutrients. The enrichment of cereal-based foods is aimed toward the replacement of minerals (Fe and recently Zn) and vitamins (thiamin, riboflavin, niacin, and recently folic acid) lost during milling. In some countries, the enrichment of cereal-based foods is required by law. (*See* **Anemia (Anaemia)**: Iron-deficiency Anemia; **Folic Acid**: Properties and Determination; **Iron**: Properties and Determination; **Niacin**: Properties and Determination; **Riboflavin**: Properties and Determination; **Thiamin**: Properties and Determination; **Zinc**: Properties and Determination; .)

In general, cereals are a very poor source of calcium, except for finger millet and teff (Table 3). Some food processes, such as nixtamalization of maize for tortilla production, greatly increase calcium. The bioavailability of this mineral in tortillas is high. (*See* **Tortillas**.)

Phosphorus is the most abundant mineral (Table 3). Unfortunately, its availability is questionable because it is complexed with phytic acid. The phytic acid has the property of binding strongly with cations, thus decreasing their bioavailability. Phytic acid decreases significantly during fermentation, due to activation of phytases. These processes considerably improve the

Table 2 Amino acid composition of cereal grains[a]

Amino acid (g per 100 g of protein)	*Wheat*		*Maize*		*Rice*		*Sorghum*		*Barley*	*Rye*	*Oats*	*Groats*	*Millets*						
	Hard	*Durum*	*Normal*	*High-lysine*	*Brown*	*Milled*	*Normal*	*High-lysine*					*Pearl*	*Finger*	*Proso*	*Japanese*	*Italian*	*Kodo*	*Teff*
Essential																			
Phenylalanine	4.6	4.1	4.8	4.3	5.2	5.2	5.1	4.9	5.2	5	5.4	4.2	5.2	5.2	5.2	5.9	5.5	5.8	5.7
Histidine	2	1.9	2.9	3.8	2.5	2.5	2.1	2.3	2.1	2.4	2.4	2.2	2.2	2.2	2.2	1.9	2.9	1.8	3.2
Isoleucine	3	3.6	3.6	3.4	4.1	4.5	4.1	3.9	3.6	3.7	4.2	3.9	4.4	4.4	4.6	4.5	5.9	3.1	4
Leucine	6.3	7	12.4	9	8.6	8.1	14.2	12.3	6.6	6.4	7.5	7.4	11	9.5	12.9	11.5	14.1	8.6	8.5
Lysine	2.3	2.2	2.7	4.3	4.1	3.9	2.1	3	3.5	3.5	4.2	4.2	2.9	2.9	2.2	1.7	2.2	3.2	3.5
Methionine	1.2	0.9	1.9	2.1	2.4	1.7	1	1.6	2.2	1.6	2.3	2.5	2	3.1	2	1.8	2.6	1.7	4.1
Threonine	2.4	2.9	3.5	3.9	4	3.7	3.3	3.3	3.2	3.1	3.3	3.3	3.9	4.2	3.3	2.7	4.3	2.9	4.3
Tryptophan	1.5		0.5	0.9	1.4	1.3	1	0.9	1.5	0.8		1.3	2.3	1.5	0.9	1	1.4	0.8	1.4
Valine	3.6	4.6	4.9	5.6	5.8	6.7	5.4	5.1	5	4.9	5.8	5.3	5.7	6.6	5.1	6.1	5.1	4.2	5.5
Nonessential																			
Aspartic acid	4.7	4.7	6.4	7.7	9.3	9.8	6.4	7.5	6	6.7	9.2	8.9	8.6	6.5	5.7	6.1	6.9	6.3	6.6
Glutamic	30.3	32.3	19.2	17.1	17.3	19.3	20.6	20.1	25.5	24.7	21.6	23.9	20.7	20.3	20.4	23.9	18.8	23.1	24.8
Alanine	3.1	4.8	7.7	6.3	5.8	5.8	8.6	8.4	2.1	2.4	5.1	5	8.5	6.2	10.7	9.3	8.9	5.5	5.7
Arginine	4	3.5	4.8	6.9	9.5	8.8	3.5	4.5	4.6	5.9	6.4	6.9	5.3	4.5	3.2	3.6	2.8	3.6	5
Cysteine	2.8		1.4		2.3	2.2	1.6	1.5	1.8	2	1.7	1.6	2.1	2.6	1.6	2.7	1.4	1	0.9
Glycine	3.8	6.5	3.8	5	4.8	4.8	2.9	3.5	3.9	4	5.1	4.9	3.3	4	2.2	2.3	2.9	3.8	3.8
Proline	10.1	13.4	8.2	9.1	5	4	7.9	7.6	11.6	9.1	5.7	4.7	6.6	7	7.2	10.1	10.6	7.2	5.5
Serine	4.2	5.7	4.6	4.7	5.3	4.3	4.1	4.2	3.8	4.1	4	4.2	4.9	5.1	6.3	5.6	5.8	4.1	5.2
Tyrosine	2.7	2	4.2	3.5	4.2	5	3.2	4.2	2.8	2.6	2.6	3.1	3.2	3.6	2.4	2.4	2.6	3.8	3.9
Amino acid score (%)	42.3	40.4	49.6	79	75.4	71.7	38.6	55.1	64.3	64.3	77.2	77.2	53.3	53.3	40.4	31.2	40.4	58.8	64.3
Protein digestibility	89.8	87.9	88.2	88.9	88.5	89	85.9	88.9	84.9	86.7	86.2	90.6	89		89.9		89.8		

[a]Essential amino acid requirements for infants (g per 100 g of protein): lysine, 5.44; methionine and cysteine, 3.52; threonine, 4; isoleucine, 4; leucine, 7.04; phenylalanine and tyrosine, 6.08; histamine, 1.4; tryptophan, 0.96; tyrosine and cysteine are not essential amino acids, but they can spare the requirements for phenylalanine and methionine, respectively.

bioavailability of minerals. (*See* **Phytic Acid**: Properties and Determination.)

Cereals are considered a good source of potassium and are practically devoid of sodium. Whole grains provide a significant amount of magnesium, iron, zinc, and copper, which is reduced by degermination, decortication, and milling.

Cereals are also considered an important source of B vitamins (except B_{12} or cobalamin), but dried-matured grains do not contain vitamin C. The B vitamins are concentrated in the aleurone layer. Beriberi (a thiamin-deficiency disease), endemic in eastern and southern Asia, is prevalent among people who consume milled rice. Milled rice contains about 10% of the thiamin of brown rice (Table 3). (*See* **Beriberi**.)

Niacin is found in a free and bound form and can be synthesized from tryptophan. The alkali treatment of maize for tortilla production considerably improves niacin bioavailability because the glucosidic bond that renders it unavailable is alkali-labile. Niacin deficiency produces pellagra, which causes dermatitis, diarrhea, and dementia and has been prevalent in regions of Southern Africa where people rely on maize as the main food source. (*See* **Niacin**: Properties and Determination; **Pellagra**.)

Toxins and Contaminants

Naturally Occurring Antinutritional Compounds

All cereals contain relatively high levels of phytic acid, which binds minerals (cations). Germination and/or fermentation improves mineral bioavailability due to the enzymatic activity of phytases and bacteria.

Brown sorghums contain condensed tannins in their testa. The brown sorghums are grown to avoid bird damage and grain deterioration due to sprouting and weathering. The tannins bind hydrophobically with, and precipitate, proteins in food systems, thus lowering digestibility. Germination and the treatment of grain with calcium oxide (CaO), potassium carbonate (K_2CO_3), ammonium bicarbonate (NH_4HCO_3), or sodium bicarbonate ($NaHCO_3$) detoxifies the grain and consequently improves its nutritional value. In Tanzania, Magadi soda is used to detoxify brown sorghum in small villages. (*See* **Sorghum**; **Tannins and Polyphenols**.)

A higher incidence of goiter among pearl millet eaters in West Africa (Sudan) has been attributed to a goitrogenic compound. The goitrogen has been identified as a thioamide, mainly found in the bran and endosperm. Heat treatment (autoclaving) apparently detoxifies the grain. Trypsin inhibitors and saponins have been isolated in pearl millet. (*See* **Goitrogens and Antithyroid Compounds**.)

Contaminants

Cereal grains are susceptible to mold attack, especially when storage conditions are inappropriate. Some molds produce many toxic compounds that can severely affect human health or even cause death. The most important mycotoxin is that produced by *Aspergillus flavus*. Aflatoxins have been found in most cereal grains, but they are most prevalent in maize. The toxin is a potent hepatocarcinogen at concentrations in parts per billion. The most prevalent aflatoxins in cereal grains are B_1 and B_2. For optimum growth and aflatoxin production, the mold requires a relative humidity of 85% and a temperature range of 27–30 °C and at least 17.5% moisture content in the cereal grain. The alkali treatment of maize for tortilla production significantly reduces the amount of aflatoxins. Most of the aflatoxins end up in the steep and wash waters. (*See* **Mycotoxins**: Occurrence and Determination; **Spoilage**: Molds in Spoilage; **Tortillas**.)

Ochratoxins are produced predominantly by the storage fungus, *Aspergillus ochraceus*. Toxicity is characterized by nephropathy, mild degeneration of the liver, and enteritis in swine, poultry, and humans. Endemic ochratoxicosis has been observed in the Balkans (Bulgaria, Yugoslavia, Romania) since 1950.

Species of *Fusarium* produce toxins that affect humans and animals. An example is scab wheat and barley contaminated with *F. graminearum*. The preparation of bakery products with contaminated grain causes vomiting and inebriation in humans. Zeralenone, another *Fusarium* toxin, is the most frequent mycotoxin in cereal grains. It is mainly found in maize. Zeralenone produces vomiting in monogastrics, and diarrhea, hemorrhage, swelling of genitals, and infertility in swine. The fumonisins are the most recently discovered family of *Fusarium* toxins. Their isolation was prompted by the death of horses from encephalomalacia and esophageal cancer. Recent epidemiological data have related consumption of fumonisin-contaminated maize with esophageal cancer in humans.

Ergot (*Claviceps purpurea*) is a fungus that infects rye and, less often, wheat, barley, and oats. Honey disease of millet and sorghum is caused by ergot. The toxicity of ergot has been documented for several centuries. Epidemics of poisonings have occurred in North America, the UK, Europe, Russia, and other places. Ergot toxicity produces a disease commonly called *Saint Anthony's Fire*, which produces gangrenous necrosis, hallucinations, and convulsions. The fungi produce toxicity via the formation of alkaloids (ergotamine, ergotoxine, and ergometrine). Growth of the fungus is promoted by moist, warm climatic conditions.

Table 3 Mineral and vitamin composition of cereal grains

Nutrients	*Wheat*		*Maize*	*Rice*		*Sorghum*	*Barley*	*Rye*	*Oats*	*Groats*	*Millets*						
	Hard	*Durum*		*Brown*	*Milled*						*Pearl*	*Finger*	*Proso*	*Italian*	*Kodo*	*Teff*	*Fonio*
Minerals																	
Ca (%)	0.03	0.04	0.03	0.03	0.02	0.04	0.04	0.05	0.11	0.08	0.01	0.33	0.01	0.01	0.01	0.17	0.03
P (%)	0.35	0.51	0.29	0.25	0.12	0.35	0.56	0.36	0.38	0.51	0.35	0.24	0.15	0.31	0.32	0.45	0.18
Phytic acid (%)	0.97		0.71	0.56		0.77	1.06	0.97	1.80				0.32				
K (%)	0.36	0.49	0.37	0.17	0.10	0.38	0.50	0.47	0.47	0.44	0.44	0.43	0.21	0.27	0.17	0.31	0.16
Na (%)	0.04		0.03	0.03	0.00	0.05	0.02	0.01	0.02		0.01	0.02	0.01	0.01	0.01	0.02	0.02
Mg (%)	0.14	0.17	0.14	0.19	0.03	0.19	0.14	0.11	0.13	0.14	0.13	0.11	0.12	0.13	0.13	0.18	0.40
Fe (ppm)	40.1	47.8	30	28	19	50	36.7	38	62	47.2	74.9	46	33.1	32.6	7	14.9	36
Co (ppm)	0.05		0.1	0.07	0.01	3.1	0.04		0.05		0.50	0.10				0.06	3.30
Cu (ppm)	4.9	5.6	4	4.2	2	10.8	15.1	9	4.7	4.8	6.2	0.3	8.3	9.2		4.4	15
Mn (ppm)	40.0	33.5	5	24	12	16.3	18.9	58.4	45	46	18	7.5	18.1	21.9		2.5	30
Zn (ppm)	30.9	41	20	18	10	15.4	23.6	32.2	37	35.8	29.5	15	17.2	21.4		6.7	30
Vitamins																	
Thiamin (mg per 100 g)	0.57	0.67	0.38	0.34	0.07	0.46	0.44	0.69	0.77	0.72	0.38	0.48	0.63	0.48	0.32	0.45	0.30
Riboflavin (mg per 100 g)	0.12	0.11	0.14	0.09	0.03	0.15	0.15	0.26	0.14	0.16	0.22	0.12	0.22	0.12	0.05	0.10	0.10
Nicotinic acid (mg per 100 g)	7.40	11.10	2.80	4.62	1.60	4.84	7.20	1.52	0.97	0.91	2.70	1.30	1.32	3.70	0.70	2	3
Pyridoxine (mg per 100 g)	0.35	0.43	0.53	0.92	0.45	0.59	0.44	0.34	0.12	0.21							
Pantothenic acid (mg per 100 g)	1.36		0.66	1.35	0.75	1.25	0.57	0.73	1.36	1.23	1.09		1.10	0.82			
Biotin (mg per 100 g)	0.01		0.01	0.01		0.02	0.01	0.01	0.02								
Folacin (mg per 100 g)	0.04	0.04	0.03	0.02	0.02	0.02	0.04	0.05	0.06					0.02			
Carotenes (mg kg^{-1})	0.2	0.2	29.5			15.7	1				5.4						
Vitamin E (mg kg^{-1})	12.8	28	24	1.7	0.14	13.8	24.8	16.6	16.7	17	19	22		31			

See also: **Dietary Fiber**: Properties and Sources; Determination; Effects of Fiber on Absorption; **Energy**: Measurement of Food Energy; **Kwashiorkor**; **Maize**; **Malnutrition**: The Problem of Malnutrition; **Marasmus**; **Pellagra**; **Rice**; **Sorghum**; **Spoilage**: Molds in Spoilage; **Starch**: Resistant Starch; **Wheat**: The Crop

Further Reading

FAO (2000) Electronic page: http//apps.fao.org/. Rome: Food and Agriculture Organization.

Finley JW and Hopkins DT (1985) *Digestibility and Amino Acid Availability in Cereal and Oilseeds*. St Paul, MN: American Association of Cereal Chemists.

Hulse JH, Laing EM and Pearson OE (1980) *Sorghum and the Millets: Their Composition and Nutrition Value*. New York: Academic Press.

Lorenz KJ and Kulp K (1991) *Handbook of Cereal Science and Technology*. New York: Marcel Dekker.

Miller JD and Trenholm HL (1994) *Mycotoxins in Grain. Compounds other than Aflatoxin*. St. Paul, MN: Eagan Press.

Mirocha CJ and Christensen CM (1982) Mycotoxins. In: Christensen CM (ed.) *Storage of Cereals, Grains and their Products*. St. Paul, MN: American Association of Cereal Chemists.

Serna Saldivar SO (1996) *Química, Almacenamiento e Industrialización de los Cereales*. Mexico, DF: AGT Editor.

Champagne *See* **Wines**: Types of Table Wine; Production of Table Wines; Production of Sparkling Wines; Dietary Importance; Wine Tasting

CHAPATIS AND RELATED PRODUCTS

P Haridas Rao and R Sai Manohar, Central Food Technological Research Institute, Mysore, India

Introduction

The chapati, also known as the roti, is a flat unleavened baked product that forms the staple food of the majority of the population in the Indian subcontinent. The consumption of this traditional product is increasing and has become popular even in areas where, traditionally, rice has been the staple diet. Chapatis are normally made and served hot at breakfast, lunch, or dinner, and are eaten along with other adjuncts in households, restaurants, and industrial canteens. There are several culinary variations of the chapati – the important ones being the parotha, stuffed parotha, tandoori roti, phulka and puri. The other culinary variations include nan, South Indian parotha and roomali roti.

Either whole-wheat flour – generally known as atta and obtained by grinding wheat in a plate/disc mill – or 'resultant atta', a byproduct of the roller flour-milling industry, simulating whole-wheat flour, is used for the preparation of chapatis. Sometimes, whole-wheat flour is first sieved to remove 3–5% of the coarse bran, and then used for chapati-making. The flour, after having been mixed with water to form the dough, is sheeted to a circular shape, and baked on a heated iron plate or on an electric hotplate. The major differences between chapatis and similar products arise mainly from differences in processing conditions. Puri is prepared similar to chapati, but it is deep-fat-fried instead of being baked. The nan is a lactic fermented flat product. South Indian parotha is a nonfermented, laminated product. Roomali roti is a very thin product baked in an inverted heated convex iron plate. The nan, South Indian parotha, and roomali roti are basically made from refined wheat flour obtained from roller flour-milling operations. (*See* **Flour**: Roller Milling Operations; **Wheat**: The Crop.)

Characteristics of Products

The chapati and its culinary variations come under the category of unfermented flat breads; consequently, they are dense in texture, comprising mostly crust with little or no crumb (**Figure 1**). But, in a few products, viz., nan, south Indian parotha, and tandoori roti, the crumb-to-crust ratio is higher. They are

usually round, but may occasionally be made triangular or rectangular in shape. The diameter varies from 10 to 20 cm, and the thickness from 3 to 10 mm. The desirable characteristics for these products are a soft, smooth, and pliable texture, light creamish brown color, slight chewiness, and wheaty aroma. However, there are very few distinct differences in the product characteristics of the various forms of chapatis.

Chapati

The chapati is circular with a diameter of 10–15 cm and a thickness of 2–3 mm. The dough sheet is baked on a heated iron plate and puffed on a live coal fire or flame. The thinner chapati, with a thickness of 1–2 mm, is known as 'phulka'.

Complete and full puffing is one of the important quality attributes of a chapati, as it results in a soft and pliable texture. The chapati should have an attractive appearance without any cracks on the surface. The surface should have uniformly distributed lightbrown spots. The texture should be soft, smooth, and pliable, and these characteristics should be retained for at least 2–3 h of storage.

The chapati should have a sweetish, wheat taste, and a baked wheaty aroma, and it should not be perceived as being tough and leathery when chewed.

Parotha

Like the chapati, the parotha is also made from whole-wheat flour or resultant atta. Parotha dough contains salt and fat. The dough is sheeted and laminated at least three times; oil is applied during lamination as well as during baking.

The major difference between the chapati and the parotha is that the latter contains four to eight discrete layers, and has a more pronounced taste and flavor of oil. In view of the presence of oil, its keeping quality is better than that of chapatis. The parotha is slightly thicker (4–6 mm) than the chapati and is normally triangular in shape. Its color is creamish brown and it is slightly brittle on chewing. The parotha can be stuffed with cooked vegetables, e.g., potato. Other types of parotha can be made from a dough containing leafy spices, like methi. They are either baked or fried.

Tandoori Roti

The tandoori roti is similar to the chapati in composition, but it is somewhat thicker and larger. It is made in restaurants and other roadside eating places known as 'dhabas'. It is rarely made at home as it requires a special type of oven, called a tandoor, which is essentially an enclosed cavity dug out in a layer of clayey soil, and heated by live coal. In some urban areas, ovens of this type can be seen in centralized places; customers normally bring the dough sheet and get the rotis made and baked in the tandoor for a nominal charge. The thickness of a tandoori roti is greater than that of a chapati; it may be 8–10 mm thick, and its diameter may range between 15 and 20 cm. The crumb-to-crust ratio is greater than that of a chapati. A tandoori roti is less pliable than a chapati; its surface is rough and exhibits several large blisters. The moisture content of a tandoori roti is also lower.

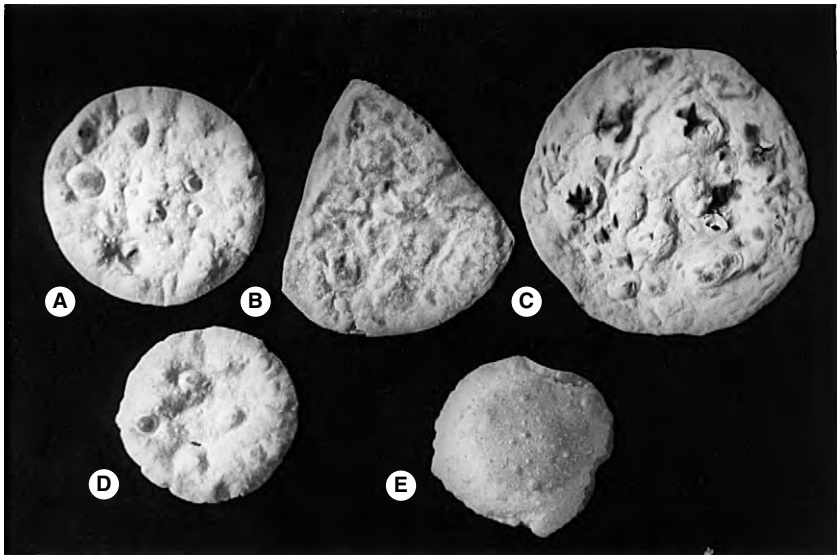

Figure 1 Chapati and its culinary variations. (A) Chapati, (B) parotha, (C) tandoori roti, (D) phulka, and (E) puri.

A typical baked aroma and slightly brittle and chewy texture are considered desirable.

Puri

Puri is also made from whole wheat flour but of 85–90% extraction. Like chapati, it has a thickness of 1.5–2.0 mm and a diameter of 10–12 cm. It is a deep-fat-fried product. The puri should completely puff on frying. A light golden brown color with full puffing, as well as a soft and pliable texture, are considered desirable.

Phulka

Phulka is very similar to chapati, except that it is very thin, with a thickness of only 0.80–1.2 mm and a diameter of 10–12 cm. The product is very light and fully puffed. It is made from whole-wheat flour and water, with fat and salt as optional ingredients. In some places, *phulka* made with fat is preferred, as it can be made into a very thin sheet and yields a highly pliable product. The sheeted dough is baked on a hotplate and puffed on live coals or on a fire.

The following products (Figure 2) are prepared from wheat flour obtained from a roller flour mill.

South Indian Parotha

South Indian parotha is made from refined wheat flour. This type of parotha is thick, with a thickness ranging from 5.0 to 7.5 mm and has a diameter of 15–20 cm. It is a laminated product. Since oil is applied during lamination, discrete layers are seen in the product. It is a soft and chewy product, creamish white in color, and slightly pliable.

Roomali Roti

Roomali roti is a very thin product made from wheat flour, milk powder, oil, and water. It is a very thin product with a thickness of 0.5–0.8 mm but has a large diameter of 60–90 cm. It is served after folding and is eaten along with other side dishes.

Nan

Nan is made using flour, salt, milk, milk powder, curd, baking soda, etc. Egg is an optional ingredient used to improve the color and texture of the product. The nan is circular or ellipsoidal, with a thickness of 4–6 mm, and is baked in a special tandoor oven similar to that used for tandoori roti. The product is creamish in color with dark spots on the surface with a typical baked flavor. Unlike chapati, crumb-to-crust ratio is higher. A soft texture is desired in the product.

Choice of Ingredients

Whole-wheat flour and water are the major ingredients used in the preparation of chapatis and other similar products. Salt and oil form the optional ingredients in products like chapatis and tandoori rotis and phulka, while they are essential in the preparation of parothas and stuffed parotha. Wheat flour, water, and oil are the major ingredients of South Indian parotha and roomali roti. In addition to wheat flour, milk and curd are also essential ingredients in the preparation of nan. Curd is used as a source of lactic acid bacteria.

Whole-wheat Flour

Whole-wheat flour, obtained by grinding wheat in a plate/disk mill, is preferred for the preparation of chapatis. The mill consists of two chilled-case corrugated iron plates placed vertically. One of the plates is stationary, while the other rotates. The wheat entering from the hopper is ground between the plates due to the shearing action as well as the friction. The particle size of the flour is adjusted by varying the clearance between the plates. It is desirable to grind wheat to a fine particle size, so that 80–90% can pass through a 60-mesh sieve. It is normal practice in households to remove 3–5% of the coarse bran from whole-wheat flour by sieving through a 30-mesh sieve. The resultant atta is also used for chapati making, as it is available at a lower price, being the byproduct of the roller flour-milling industry. There is little difference in the composition of resultant atta and whole-wheat flour, except that the former has lower contents of ash and damaged starch, while its water absorption capacity is appreciably less (see Table 1).

The higher water absorption capacity of whole-wheat flour is attributed to its greater damaged starch content. The resultant atta has 1.5 times as much thiamin and riboflavin and twice as much niacin as that of whole-wheat flour. (*See* **Niacin**: Properties and Determination; **Riboflavin**: Properties and Determination; **Starch** Structure, Properties, and Determination; **Thiamin**: Properties and Determination.)

Whole-wheat flour, obtained from medium-hard aestivum wheat, and with the quality characteristics listed in Table 2, is suitable for chapati-making. The dough should show an alveograph stability value of 100–200 mm, a strength of 15–20 $\times$ 10^{-4} m^2, and an extensibility of 0.20–0.27 m.

In the case of phulka, a very fine ground wheat is required, and normally, 100% will pass through a 60-mesh sieve. The grinding is normally done in a stone mill, which is similar to a plate mill, except that it has a pair of emery stones. The damaged starch content in such flour is very high, ranging from 18 to 20%. The

Figure 2 Wheat flour-based culinary variations of chapati. (A) South Indian parotha, (B) nan, and (C) roomali roti.

quality of the flour used for tandoori roti is similar to that used for chapati. The flour used for puri, however, has a lower ash content because 10–15% of coarse branny materials are removed by sieving; alternatively, whole-wheat flour can be blended with 20–25% of refined wheat flour.

Wheat Flour

Products such as roomali roti, nan, and South Indian parotha make use of wheat flour obtained by roller flour-milling of medium-hard wheat. The flour should have 9.5–10.5% protein and 0.5–0.7% ash.

Composite Flour

Part of the whole-wheat flour can also be replaced by other (i.e., non-wheat) cereal flours and flours obtained from legumes or millets, either to improve the taste and texture, or to improve the nutritional quality. Among the legume products used in this way, the most common is bengal gram flour, which is normally added at the 20–30% level; the chapati so made is locally known as a mesi roti. Other flours that are used, or could be used, in chapati making at levels ranging from 10 to 20% are defatted soya, groundnut, or cottonseed flours or refined flours from barley, bajra, sorghum, maize, cassava, and sweet potato. Triticale, a man-made cereal, was found to be unsuitable for chapati-making, but it can be used after blending with wheat flour in a ratio of 1:1. Soft wheat, which is also not suitable for chapati-making, can be improved by blending with hard wheat, or extra-hard durum wheat. (*See* **Barley**; **Cassava**: Uses as a Raw Material; **Sorghum**; **Soy(Soya) Beans**: The Crop; **Vegetables of Tropical Climates**: Root Crops of Uplands.)

Salt

Salt is generally dispersed in water and added to the dough. Hence, edible common salt of any particle size can be used.

Oil

Refined oil is used during dough preparation and during rolling and lamination of the dough sheet for the preparation of parotha and South Indian parotha. Solid fat (hydrogenated oil) can also be used, after it has been melted to make it easier to use during lamination. Solid or plastic fat can also be used in the preparation of the dough.

Egg

Egg is used as an optional ingredient in the preparation of nan. Fresh egg is normally used in the preparation. It is used to improve the color, texture, and flavor of nan.

Skimmed Milk Powder (SMP)

SMP is an optional ingredient used in the preparation of nan, while milk or SMP is an essential ingredient in the preparation of roomali roti. It is added to improve the texture, flavor, and color.

Table 1 Characteristics of resultant atta and whole-wheat flour

Characteristics[a] *(%)*	*Resultant atta*	*Whole-wheat flour*
Ash	1.31	1.68
Protein	10.80	10.60
Dry gluten	9.90	9.90
Ether extractives	3.20	2.30
Crude fiber	2.09	2.24
Starch	52.20	54.50
Damaged starch	7.00	14.50
Water-absorption capacity[b]	62.80	73.50

[a]Values are the average of 18 samples and are expressed on a 14% moisture basis.
[b]Required to prepare chapati dough of optimum consistency.

Table 2 Characteristics of whole-wheat (aestivum) flour suitable for chapati making

Characteristics	*Level*
Ash (%)	1.2–1.5
Protein (%)	9.0–10.5
Sedimentation value (ml)	25–35
Pelshenke value (min)	120–150
Total sugars (%)	2.5–4.0
Diastatic activity (mg/10 g flour)	250–375
Damaged starch (%)	16–18
Water-absorption capacity (%)[a]	72–78

[a]Required to prepare dough of optimum consistency.

Water

There is no information on the type of water suitable for chapati-making. Generally, any potable water, with a medium hardness and a pH value close to neutrality, is suitable. This is also true for other variants of chapati.

Influence of Ingredients

Whole-wheat Flour

Since whole-wheat flour is the major ingredient in the preparation of chapati and other products, its quality has a considerable influence on the quality of the end products. Whole-wheat flour obtained from medium-hard aestivum wheat is preferred, as it yields a chapati with the desired quality characteristics. The use of a strong wheat flour, with a protein content of more than 11.0%, results in a chapati with a tough and leathery texture, whereas a soft wheat, with less than 9% protein, yields a chapati with a stiff and brittle texture. Such chapatis also tend to become hard and brittle within an hour of storage. A wheat flour of a similar quality is also well suited for making tandoori roti, puri, and phulka.

The quality of a chapati also depends on the relative amounts of gliadin, glutenin, and residue protein present in the flour. A flour with equal amounts of the above proteins gives fully puffed chapatis with a soft texture, whereas a flour that has more residue proteins, or gliadin, gives a stiff and leathery chapati. A flour with a high content of total glutenin has been found to yield tandoori roti of a highly acceptable quality. The higher levels of polar lipids in flour result in chapatis with a soft texture.

In addition to the type of wheat, the content of damaged starch, the diastatic activity, and the sugars present in the flour affect the quality of a chapati. The higher the damaged starch content, the higher will be the water absorption capacity and hence, the softer the texture of the chapati. A significant direct correlation exists between damaged starch content and water absorption ($r = 0.89$, $P < 0.001$) and texture of chapatis ($r = 0.56$, $P < 0.001$). A flour with a high diastatic activity and sugar content yields a chapati with a sweetish taste.

The characteristics of chapati dough, which are influenced by the quality of the flour, affect the sensory qualities of the chapati. The sensory texture of a chapati is related significantly to farinograph consistency ($r = 0.47$, $P < 0.01$), extensograph extensibility ($r = -0.51$, $p < 0.01$) and resistance to extension ($r = 0.52$, $P < 0.01$), and textureometer cohesiveness ($r = -0.57$, $P < 0.01$).

Wheat with a low polyphenolase activity is desirable to avoid an excessive brownish color in the chapati. These enzymes are found to be low in Indian wheat varieties, whereas Mexican hybrid wheat varieties, now grown in India, have a high content of this enzyme. The adverse effect of enzymatic browning is greater when the chapati dough is rested for a longer period, or when the baked chapati is stored for a few hours.

Among the different *Triticum* species, the wheat, *T. aestivum*, which yields a nonsticky dough with good machinability, and a fully puffed chapati with soft and pliable texture, is most suitable for chapati preparation. When the species *T. durum* and *T. dicoccum* are used for chapati making, dry and hard-textured products result, which are perceived to be leathery when chewed. They are also sweeter due to the higher damaged starch content. In addition to the type and species of wheat, the severity of grinding and particle size of the flour play important roles in influencing the quality of chapatis and other similar products. Among the different types of mills (plate mill, hammer mill, pin mill, roller mill, etc.), the flour obtained from plate mills has been found to be the best for making chapati, tandoori roti, phulka, and puri. This is because of the greater severity of

grinding, which results in a flour with a desirable high damaged starch content, and hence a higher water absorption, resulting in a soft and pliable chapati. The soft and pliable characteristics are retained in products over a long storage period. The heat developed in such a mill during grinding tends to yield a chapati with a better flavor. Among the plate mills, the power- or water-driven mills are found to be better than hand-driven mills, as the latter yield flour with a low maltose figure of 1.72–1.99%, and a water-absorption capacity of just 66–69%, in contrast to the desired higher values of 2.6–4.65% for maltose content and 72–78% for water absorption.

A finely ground flour is desirable for chapati and tandoori roti-making. However, for phulka, the grinding should be much finer. The coarse flour, with more than 70–80% overtailing on a 10XX sieve (130 μm), yields a chapati with a poor appearance and tough texture. The shear force required to cut such chapatis, which is related to the sensory texture ($r = 0.67$, $P < 0.001$), is as high as 18–20 kg, compared with the force of 5–6 kg required for chapatis made from normal flour (20–30% overtailing). The quantity of water required to prepare the chapati dough of desired consistency is significantly related to the fraction passing through a 10XX sieve ($r = 0.89$, $P < 0.01$) and the protein content of the flour ($r = 0.53$, $P < 0.05$).

The extraction rate of flour has a considerable influence on the chapati quality. The quality of chapati is not affected by removing 4–6% of the coarse bran by sieving (extraction rate 94–96%), as is the practice in households. However, removing higher amounts of bran particles by sieving, or using a flour of a lower extraction rate, makes the chapati tough and leathery. Hence, to make good chapatis, a considerable amount of bran needs to be present in flour.

The chapati made from resultant atta is slightly tough and has a bland flavor; the latter is attributed to a lower damaged starch content, as well as to the negligible heat developed during grinding. The temperature reached while grinding in a roller flour mill can be as low as 30–40 °C, as against 90–95 °C in a plate mill. The high temperature reached during the milling operation seems to favor impartment of a wheaty aroma in chapatis. This is true for other products like tandoori roti, parotha, and puri.

If wheat is infected by kernel bunt, at infection levels above 3%, chapatis will have a dark color and an undesirable flavor. However, these adverse affects can be minimized by soaking or washing the infected grains before grinding. Storing whole-wheat flour for more than 3 months at a high temperature adversely affects the quality of chapatis.

Wheat flour

Roller-milled flour obtained from medium-hard wheat is used for the preparation of South Indian parotha, nan, and roomali roti. The protein content and the damaged starch in flour are related to the quality of South Indian parotha. The protein content of 9.5–10.5% and damaged starch of 10–12% are desirable (Table 3). The dough used for roomali roti and South Indian parotha should be quite extensible, but, at the same time, it should have sufficient strength. The optimum extensograph *R/E* value ranges between 4 and 5 for flours used in the above products. In nan, the optimum *R/E* value is 3–4.

Salt

Incorporation of salt in the dough reduces the stickiness and, hence, improves the sheeting characteristics of chapati dough and other similar products (doughs). The texture of the products is also improved considerably.

Oil

Incorporation of oil in the chapati dough improves the texture and pliability. It also keeps the chapati soft and pliable during storage. Puffing characteristics are also slightly improved on incorporation of very small amounts of oil in the dough. Higher levels of fat or shortening in the dough reduce the puffing quality of chapatis.

Discrete layers in a parotha are due to the application of oil during lamination. Oil is also used during the dough preparation to improve the rolling property. Oil is used extensively during stretching and rolling in the preparation of South Indian parotta, while it is used in dough as well as stretching in the preparation of roomali roti. Application of oil during lamination helps to produce discrete laminations in South Indian parotha. However, in the case of roomali roti, the oil helps to produce a very thin sheet of dough.

Water

The amount of water required to prepare a dough of optimum consistency for chapati preparation ranges from 70 to 80%. The amount depends on several factors, such as the type of wheat, grinding method, extraction rate, particle size distribution, severity of grinding, etc. The higher the water requirement, the softer will be the texture of the chapati and the sweeter the taste. Such chapatis keep well on storage. The amount of water required for phulka is higher and ranges from 75 to 80%. The dough for the prep-

aration of tandoori roti requires slightly higher levels of water as the dough is made slightly loose to enable sheeting between two palms. The South Indian parotha and roomali roti require less water since they are made from wheat flour milled in a roller flour mill with a low damaged starch content; moreover, oil in dough reduces the water absorption capacity.

Additives

The incorporation of yeast (0.5%) or yogurt (10%) softens the texture, which is also retained during storage. Emulsifiers or surfactants, such as glycerol monostearate, sucrose esters, and sodium stearoyl lactylate (SSL), and the enzyme, amylase, also improve the texture of a chapati as well as its keeping quality. The use of glycerol monostearate (GMS) in nan improves its texture. Incorporation of SSL and DATEM (diacetyl tartaric acid esters of monoglycerides) improves the softness of South Indian parotha. Cysteine, HCl at a level of 50 ppm further improves the machinability of the dough and texture of South Indian parotha. (*See* **Emulsifiers**: Organic Emulsifiers; **Enzymes**: Uses in Food Processing.)

Methods of Production

Chapatis and their culinary variations are generally freshly made and served, while still hot, in households, restaurants, and industrial canteens. Fresh, hot chapatis are preferred, as they have a better flavor and soft texture. Consequently, chapatis produced on a large scale and marketed in unit packs have not become popular among consumers.

Chapati

The method of preparation of chapatis involves three important steps, *viz*. preparation of the dough, sheeting, and baking-cum-puffing (Figure 3).

The main ingredients used in chapatis are whole-wheat flour, or resultant atta, and water. The water level varies from 65 to 75%, depending on the type of wheat and the milling method. In households, and even in restaurants, the dough is kneaded by hand to a consistency that is stiffer than bread dough, but enables easy sheeting without sticking to the base of the rolling pin. In large industrial canteens, however, either a planetary vertical mixer with a U-shaped blade, or a horizontal mixer similar to that used for bread dough, is used. The dough is mixed for 10–15 min, for complete development of the gluten, rested for a period ranging from 30 to 120 min depending on the convenience. Normally, dough is rested for proper hydration and for stress relaxation, which enables easy sheeting with minimal stickiness. The optimum resting time required is shorter (15–20 min) for a dough mixed in a mixer but longer for hand-mixed dough (30–120 min).

The dough is divided into 35–40 g pieces and then rounded between the palms. The rounded dough is sheeted (using a rolling pin) to a circular shape of about 15 cm in diameter and 1.5–2.0 mm in thickness. Dusting with flour is normally done to reduce the stickiness. The rolled circular dough sheet is then baked on each side on a heated shallow iron plate (200–220 °C), for a total time of 1–2 min. At the final stages of baking on the second side, the surface of the chapati is gently pressed with a clean dry cloth for puffing. Alternatively, puffing can also be done by placing the chapati immediately after baking on a gas flame or live, glowing coals for a few seconds. The hot chapati is served either as such, or after smearing the surface with fat, preferably ghee or butter.

As yet, little work has been carried out on the mechanization of chapati production, particularly with respect to sheeting or baking. However, in some countries, the equipment generally used for the manufacture of tortillas is also used for chapati sheeting

Table 3 Quality characteristics of flour suitable for chapati and its variants

Characteristics	*Whole-wheat flour*				*Wheat flour*	
	Chapati	*Phulka*	*Tandoori roti*	*Puri*	*South Indian parotha*	*Nan*
Ash (%)	1.2–1.5	1.2–1.5	1.2–1.5	1.0–1.2	0.5–1.0	0.5–1.0
Protein (%)	9.0–10.5	9.0–10.5	9.0–10.5	9.0–10.0	10.0–11.0	8.5–9.5
SV (ml)	25.0–35.0	25.0–35.0	25.0–30.0	20.0–30.0	25.0–35.0	25.0–35.0
PV (min)	120–150	120–150	110– 140	100–120	130–160	125–150
DA (mg/10 g flour)	250–375	375–400	225– 350	200–300	175–250	150–205
Damaged starch (%)	16–18.0	18–20	13–15	12–14	8–10	7–9
Water absorption (%)[a]	72–78	75–80	70–75	65–70	62–65	52–60
Particle size (throughs of 130 μm)	80–90	85–95	75–85	70–80	100	100

[a]Required to prepare dough of optimum consistency.
SV, sedimentation value; PV, Pelshenke value; DA, diastatic activity.

and baking. A few mechanized chapati-sheeting units have been developed, and are being used in restaurants and industrial canteens. (*See* **Tortillas**.)

Sheeting of dough using a rolling pin is time-consuming and requires experienced persons to produce chapatis with a circular shape as well as uniform size and thickness.

Shaping by Pressing

Simple machines are now available to shape chapati dough. A 'Hillif' chapati press with a capacity to produce 100–200 chapatis per hour is used in some households and restaurants for sheeting. It consists of two circular cast iron disks connected by a hinge. The contacting surfaces are covered with Teflon sheets. A lever-type handle is attached to the bottom disk. About 30–40 g of dough are placed at the center of the bottom disk and covered by the top disk, and pressure is applied to the top disk for a few seconds with the help of the lever.

The pressed chapati is removed carefully and baked. Oil is applied to the surface of the disk to avoid sticking.

Another mechanical device available for flattening chapati dough is based on the principle of the can flanker, and consists of two disks, one fixed, the other movable. A toggle mechanism is used to press the dough, which is placed in between the disks, as it develops a high pressure with little effort. The thickness of the chapati can be varied by adjusting the clearance between the disks. The bottom plate, on which 40–50 g of dough have been placed, is made to press onto the top disk, with the help of a pedal connected to the toggle. This machine has the capacity to produce 200–300 chapatis per hour. This device is either hand- or foot-operated.

Shaping by Sheeting

All the equipment described above is of the batch type. Hence, little effort has been made to develop continuous shaping machines for shaping by sheeting, which is accomplished by feeding the dough through a hopper and passing it through four sets of roller to attain the desired thickness. The sheet then passes under a rubber-lined drum, and is cut (or stumped) with a Teflon-coated cutter of 18 cm diameter. The scrap dough is fed back to the hopper. The machine has the capacity to produce 1200–1800 chapatis per hour. Some machines, based on continuous sheeting and rotary cutting, are now available for the shaping of chapati dough.

Shaping and Baking

A new device, known as a 'roti maker', is now on the market. It can be used for shaping, baking, and puffing. The device is similar to the 'Hillif' chapati press, but the two disks are heated through the electric coil attached to the disk. About 30–40 g of dough are placed at the center of the bottom disk. The pressure is applied at the top disk for a few seconds

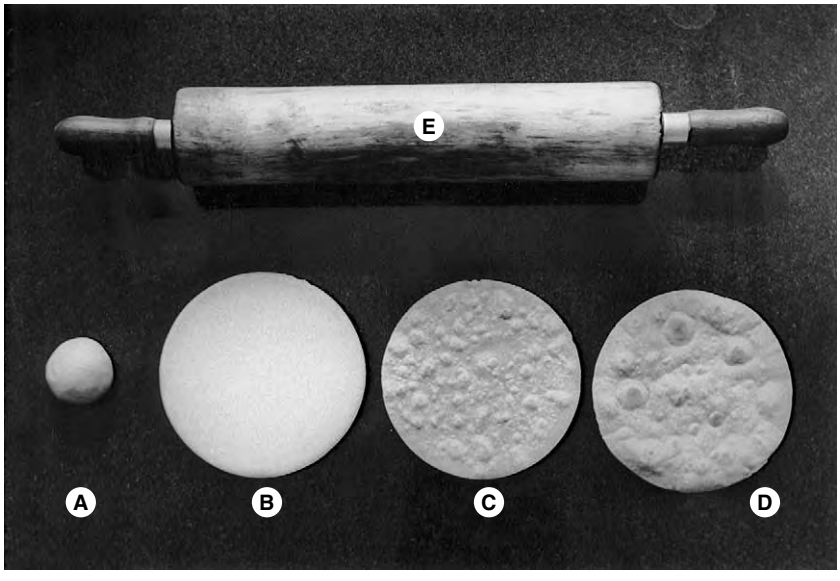

Figure 3 Method of preparation of chapati. (A) Dough ball, (B) sheeted dough, (C) partially baked chapati, (D) puffed chapati, and (E) rolling pin.

and allowed to bake. The chapati gets baked and lifts the top disk while puffing, and the puffed chapati is removed and served.

Continuous Shaping, Baking, and Puffing Units

A continuous chapati-making machine, designed and developed by Central Food Technological Research Institute, Mysore, India, is being fabricated by different machine manufacturers, and is now on the market (Figure 4). The machine has the capacity to produce 600–800 chapatis per hour. The process involves preparation of the dough in a planetary mixer. The dough is continuously conveyed to the extruder. A fish tile die moves the dough sheet to the endless belt conveyor. The sheet travels on the conveyor and passes through a rotary cutter and gets uniformly cut into a circular shape. These chapati sheets enter the baking chamber, while the remaining/cut portions of the sheet are trapped and returned to the extruder. The fuel used for baking is liquid propane gas. The baking chamber consists of a three-tier endless conveyor system. The conveyors are heated by ribbon burners at the bottom. At the third stage, the chapatis are exposed to a direct flame to yield a puffed product, which is collected in a tray.

Phulka

The preparation method, ingredients, and consistency of the dough are similar to those of chapatis. However, the flour requires higher amounts of water (75–80%) to form the dough. About 20–25 g dough are sheeted using a rolling pin to a very thin circular sheet (0.8–1.0 mm) of about 8–12 cm in diameter. The dough sheet is baked on each side on a shallow heated iron plate or hot plate maintained at 200–220 °C for 1–1.5 min and puffed on glowing coal or a live fire. In view of the lighter weight, puffing of phulka is complete. Hot phulka is consumed after smearing with oil, ghee, or butter before serving.

Phulka is also prepared using salt (0.5%) and oil (4–6%) in the dough. This helps in sheeting the dough very thinly, and the resulting phulka will have a softer texture. Phulka, particularly that made with salt and oil, can remain mold-free for about 5–6 days.

Puri

Puri is a deep-fat-fried product made from a whole-wheat flour of low extraction rate 85–90%. The dough made by kneading flour and water is sheeted to a thickness of 1.5–2 mm and a diameter of 10–12 cm using a rolling pin. The sheeted dough is deep-fat-fried in oil maintained at 180–200 °C for about 1 min with intermittent turning. The frying is continued till the puri puffs fully and results in a light golden brown color. Intermittent pressing of puri while frying with perforated ladle helps in puffing. Puri is then removed from the oil bath with a perforated ladle. It is normally consumed hot, along with boiled and spiced vegetables.

Parotha

The dough made by kneading whole-wheat flour, salt (1.0–2%), fat (2–5%) and water (65–70%) is divided into pieces of 35–45 g and sheeted, using a rolling pin, to a circular shape of 10 cm diameter. Then, oil is applied to the top surface (1–2 g) and spread uniformly. The sheet is folded into a semicircle, and oil is again uniformly applied over the surface (0.5–1.0 g); the semicircle is then folded into a quarter circle. This laminated piece of dough is then sheeted (by rolling) into the shape of a triangle with a thickness of 2–3 mm, and sides of 15–18 cm in length. The sheeted dough is baked on an electric hotplate or heated shallow iron plate, for about 1–2 min until the color becomes light brown. Oil is applied to both sides (1–2 g) while baking (Figure 5).

No attempt has been made so far to mechanize the sheeting and lamination processes in parotha-making.

Stuffed Parotha

A parotha stuffed with vegetables and spices, so that it may be eaten as such, without any side dishes, is known as ‘stuffed parotha’. The dough for stuffed parothas is made in the same manner as that for parothas. A filling is prepared by adding chopped onion (25 g), salt (1 g) and chilli powder (0.5 g) to

Figure 4 Chapati-making machine.

boiled, peeled, and mashed potato (100 g), and mixing the ingredients well. A small portion of filling (about 20 g) is placed on the center of the sheeted dough (5 cm diameter). The dough is folded over the filling to make a ball again. The dough is then rolled into a circular shape (12–15 cm) and baked on a hotplate by applying oil or butter to both sides.

The ingredients of the filling may be changed, depending on the availability of vegetables and preferences of consumers. Stuffed parothas can be prepared using radish, cauliflower, fenugreek leaves, etc. Sometimes, vegetables are mashed and mixed along with dough and parothas prepared in the normal manner.

Tandoori Roti

The ingredients used are the same as those used for a chapati, except that salt (1.25–1.5%) is sometimes added. The quantity of dough used for a tandoori roti is higher (40–60 g) than for a chapati (35–40 g). The dough is flattened by beating and pressing between the hands to a thickness of 4–6 mm and a diameter of 15–18 cm. The sheeted dough is baked in an oval-shaped oven, the walls of which are plastered with clay (a tandoor, Figure 6). The oven is heated by burning wood or coal in the bottom. The sheeted dough is placed on a cloth pad, and pasted to the heated walls of the tandoor. Depending on the temperature of the tandoor, the roti is baked in 60–90 s. When the roti has been baked properly, it falls from the side of the wall to the bottom surface. The baked roti is then taken out, using a long L-shaped iron rod. The major heat transfer in a tandoor oven is mainly by radiation, though some heating also occurs through conduction and convection.

South Indian Parotha

The dough is prepared by mixing wheat flour (100%), salt (1.0%), sugar (0.5%), egg (0.1%) and water (55–58%). It is then rested for 30 min, divided into 75 g portions, rounded and again rested for 10 min. Next, the dough is sheeted with a rolling pin, and oil is applied while sheeting. The thin dough sheet thus obtained is folded to form multiple layers and then coiled (Figure 7). The coiled dough is rested for 10 min, then sheeted again to a final thickness of 5 mm using a rolling pin. The dough sheet is then baked at 230–240 °C for 2 min on a hotplate, or on a shallow iron plate heated by gas or coal, and turned every 15 s. Oil is applied during baking, and the parotha is served hot and consumed along with spiced boiled vegetables.

Nan

Nan is prepared from refined wheat flour. The dough is made from wheat flour (100%), milk (6%), curd (12%), salt (1%), baking soda (0.5%), fat (4%), and egg (3%). The dough is fermented for 4 h; thereafter, 80 g of dough are sheeted to a thickness of 2.5 mm, using a rolling pin, and baked in a tandoor oven maintained at 350–365 °C for 60 s. Nan is served hot, and it is normally made in restaurants or

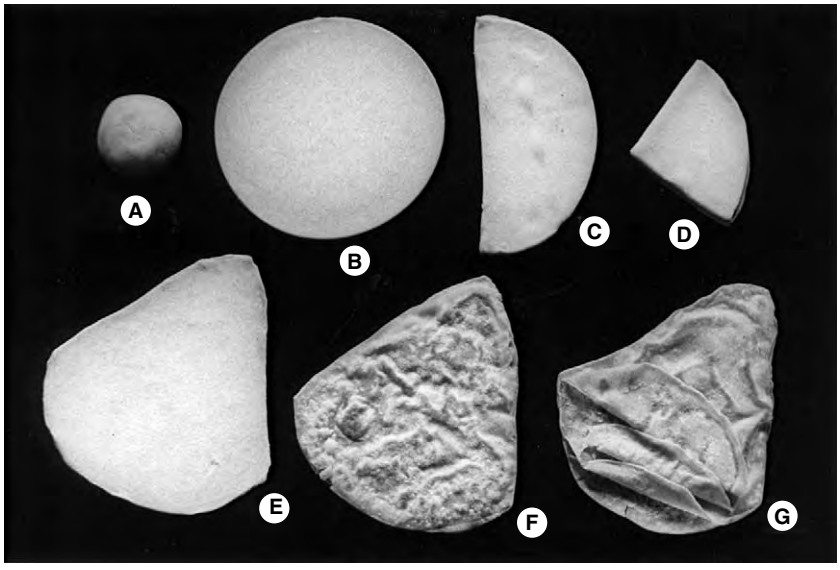

Figure 5 Method of preparation of parotha. (A) Dough ball, (B) circular dough sheet, (C) first fold, (D) second fold, (E) sheeted parotha, (F) baked parotha, and (G) different layers of parotha.

Figure 6 Tandoor oven.

small shops known as 'Dhabas', using a tandoor oven similar to that used for the preparation of tandoor roti.

Roomali Roti

Roomali roti is prepared using refined wheat flour, milk powder, fat, salt, and water. The dough made by mixing these ingredients is rested for about 30 min. About 100–200 g of the dough are stretched through fingers to a very thin transparent sheet, to a thickness of 0.5–0.8 mm and a diameter of 60–90 cm. The dough sheet is then placed on a heated (350–360 °C), inverted shallow iron plate and baked for 60–90 s. The baked roomali roti is semi-folded, quarter-folded, and served hot.

Handling, Storage, and Distribution

As already mentioned, chapatis and other similar products are normally consumed fresh and hot. On storage, they stale, becoming brittle, and hard, and lose their typical flavor; hence, they are not liked by consumers. The very short keeping quality of chapatis poses serious problems when they have to be served to a large number of consumers at one time, as in industrial canteens or in restaurants. A good chapati would be expected to retain its soft, pliable texture for at least 2–3 h. No serious mechanization attempts have been made to enable large-scale preparation and distribution of such products. Persistent research and development efforts are also being made in different parts of the world to improve upon the storage and distribution of chapatis. However, in some parts of the world, the baked chapati is now being marketed in unit packs. These chapatis need to be stored in a refrigerated condition. They are heated or warmed in a microwave oven before consumption.

The chapati, being a high-moisture product, containing 28–30% moisture, stales like bread or cake, becoming less acceptable on storage. Compared with other bakery products, a perceptible staling is observed on cooling. Further storage tends to make chapatis hard and brittle. In addition to these changes, mold growth also occurs after 3–4 days of storage.

Sorbic acid is normally used as an antifungal agent for preserving chapatis/parothas. The maximum level that could be incorporated without affecting the taste is 0.1%. The mold growth is delayed by 2–3 days at this level. However, the use of a higher level of sorbic acid (0.3%) along with 1.5% salt, and packing in polythene (200 gauge), or in an aluminum foil/polythene laminate, can delay mold growth for as long as 180 days. Sorbic acid is found to be a better preservative than its potassium salt. The level of sorbic acid could be reduced further either by including 0.4% citric acid, 3% sugar, and 2.5% salt in the recipe, or by heating the packed chapatis for 2 h at 90 °C. The chapatis made for defense personnel are preserved by in-pack heat sterilization after being packed in paper/foil/polylaminate. Such chapatis can be preserved for 180 days. Subjecting packed chapatis/parothas to γ radiation of 1 Mrad also delays the mold growth for 180 days. Parothas can be preserved well for 10 months by using 0.19% sorbic acid and 1.6% salt; the inner pack, made of MST (moisture-proof sealable and transparent) cellophane, is repacked in paper/foil/polythene pouches. (*See* **Spoilage**: Molds in Spoilage.)

A shelf-life of one year can be obtained for products such as chapatis/parothas/stuffed parothas by packing under a moderate vacuum of about 560 mmHg in sanitary cans, and heating in an oven at 115–120 °C for 1 h.

For civilian consumption, polypropylene or polythene metallized polyester laminate pouches are

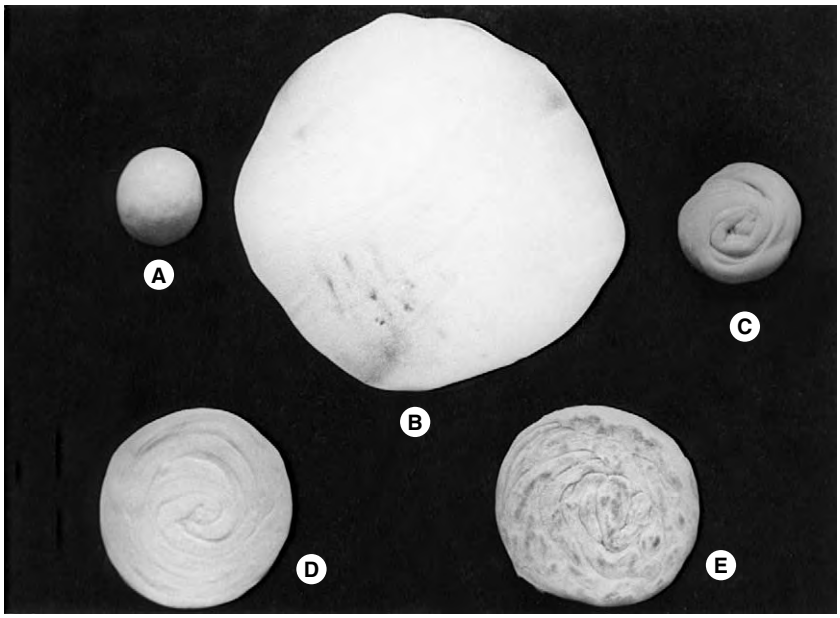

Figure 7 Method of preparation of South Indian parotha. (A) Dough ball, (B) circular dough sheet, (C) coiled dough sheet, (D) sheeted South Indian parotha, and (F) baked South Indian parotha.

found to be quite adequate. Four chapatis are packed and sealed in each pouch. The individual packets are then packed in fiberboard boxes for transportation.

The puri can be stored for a slightly longer period of 4–5 days, but it becomes hard and brittle within 2–3 days. South Indian parotha remains mold-free only for 3–4 days. Its life could be extended using sorbic acid at 0.15% level, but the parotha can become hard and brittle. The tandoor roti and nan also can be stored for the same period as a chapati. No serious efforts have been made to market these products in unit packs.

See also: **Barley**; **Emulsifiers**: Organic Emulsifiers; **Enzymes**: Uses in Food Processing; **Flour**: Roller Milling Operations; **Niacin**: Properties and Determination; **Riboflavin**: Properties and Determination; **Sorghum**; **Soy (Soya) Beans**: The Crop; **Spoilage**: Molds in Spoilage; **Starch**: Structure, Properties, and Determination; **Thiamin**: Properties and Determination; **Tortillas**; **Vegetables of Tropical Climates**: Root Crops of Uplands; **Wheat**: The Crop; **Cassava**: Uses as a Raw Material

Further Reading

Arya SS, Vidhyasagar K and Parihar DB (1977) Preservation of chapati. *Lebensmittel-Wissenschaft und Technologie* 10: 208–210.

Austin A and Ram A (1971) *Studies on Chapati Making Quality of Wheat*. Technical Bulletin, No. 31, New Delhi: Indian Council of Agricultural Research.

Faridi H (1988) Flat breads. In: Pomeranz Y (ed.) *Wheat: Chemistry Technology*, vol. II, pp. 457–506. St. Paul, MN: American Association of Cereal Chemists.

Haridas Rao P (1982) *Studies on Chapati and Similar Traditional Foods*. PhD thesis, University of Mysore.

Indrani D (1998) *Rheological Characteristics of Wheat Flour Dough in Relation to Quality of Parotha – an Indian Traditional Food*. PhD thesis, University of Mysore, Mysore, India.

Qarooni J (1996) Flat breads. In: Qarooni J (ed.) *Flat Bread Technology*. New York: Chapman & Hall, International Thomson Publishing.

Saxena DC (1996) *Studies on Tandoori Roti – An Indian Traditional Food*. PhD thesis, University of Mysore, Mysore, India.

Sidhu JS and Seibel W (1988) Effect of flour milling conditions on the quality of Indian unleavened flat bread (chapati). *Journal of Food Science* 53: 1563–1565.

CHEESES

Contents

Types of Cheese

N F Olson, University of Wisconsin, Madison, WI, USA

Definition of Cheese

Like many fermentation-derived food products, cheese cannot be easily and succinctly defined. Early versions of cheese were likely a soft gel/curd formed through acidification of milk from a mammal by contaminating lactic acid bacteria. A general definition of cheese has recently been accepted by international food regulatory bodies, which illustrates the complexity of defining this complex food. It is defined as the ripened or unripened soft or semi-hard, hard or extra-hard product, which may be coated, and in which the whey protein:casein ratio does not exceed that of milk, obtained by: (1) coagulating wholly or partly the protein of milk, skimmed milk, partly skimmed milk, cream, whey cream or buttermilk, or any combination of these materials, through the action of rennet or other suitable coagulating agents, and by partially draining the whey resulting from such coagulation; and/or (2) processing techniques involving coagulation of the protein of milk and/or products obtained from milk which give an end product with similar physical, chemical and organoleptic characteristics defined under the Codex Committee entry in the Further Reading section.

History of Cheese

The large number of different names (varieties) of cheeses resulted from the long history of cheese manufacturing, the use of milk from various animals, and development of unique technologies usually by accident but also dictated by environmental conditions. These differences were further compounded by the isolation of cheese manufacturing regions and differences in languages that resulted in cheeses that were similar but were given different names. Regional and nationalistic pride institutionalized the plethora of names of cheeses, which is estimated to exceed 1000. Subsequent trade between regions spread the notoriety and consumption of certain cheese varieties, presumably based upon their acceptable characteristics but also their ability to retain desirability during transport.

It is generally accepted that cheese as a distinct product evolved in the Middle East, specifically in the Fertile Crescent between the Tigris and Euphrates rivers. Domestication of animals such as goats and sheep led to their use as sources of meat, milk, and hides. These early herdsmen must have deduced that suckling animals derived satisfaction and growth promoting factors from the mothers' milk. Gaining sustenance from a product without sacrificing the animal for meat would likely have led to taking a portion of the mothers' milk for human consumption. The highly nutritious blend of components in milk promoted the growth of various microorganisms, including lactic acid-producing bacteria that

would be common contaminants from the environment. The soured milk was too precious to be disposed, leading to an acquired taste for this product. Handling of the clotted milk caused separation of a serum (whey) and yielded the first acid coagulated cheeses (*See* **Cheeses**: Soft and Special Varieties; Quarg and Fromage Frais). The practice of storing and transporting milk in animal skins, including stomachs, also clotted the milk by enzymes such as chymosin and pepsin that leached from the stomach lining, which yielded a gel that differed in characteristics from the acid-coagulated curd (*See* **Cheeses**: Chemistry of Gel Formation). Jostling in the bag caused separation of whey that yielded curd, the precursor of a wide range of enzyme (rennet)-clotted cheeses.

The release of the whey from the curd and the hardening of the cheese probably prompted the existing artisans to further dehydrate the curd by techniques similar to the primitive practices used for air-drying of meats. Curds produced by the action of the stomach enzymes were more conducive to dehydration, which greatly extended the storage and transportation of these cheese types.

The evolution of using sodium chloride and heat to preserve various foods was applied to the acid- and rennet-coagulated curds and created more diversity in the characteristics of cheeses that were given unique names. Modern cheese varieties such as highly salted Feta and Domiati cheeses (*See* **Cheeses**: White Brined Varieties) and the heat-plasticized Provolone and Mozzarella cheeses are examples of this evolution.

Travel and trade introduced cheeses with other valued goods to widespread regions and cultures. Introduction of cheeses into different climatic conditions altered the type of microorganisms and their rate of growth in the cheese curd and, consequently, the characteristics of the resulting cheeses. Lower temperatures minimized the need for high sodium chloride levels and use of higher heat treatments as means of preservation. The requirement for low moisture contents was also lessened. Artisans in Central Europe undoubtedly recognized that cooler temperatures in the spring and autumn yielded cheeses of better quality. They also would have known about cool caves in their localities and started the practice of storing cheese in caves or in cooler areas of their dwellings. Under these conditions that imposed fairly consistent temperature and humidity, various proportions of fungal and bacterial growth occurred that imparted unique and desired characteristics. Surface-ripened cheeses such as Camembert, Roquefort, and Munster are current varieties that evolved from such conditions (*See* **Cheeses**: Surface Mold-Ripened Cheese Varieties). Artisans in the mountainous areas of Europe, bordering on Mediterranean regions, would have been familar with cheese varieties that were subjected to a higher temperature during manufacturing. Storing such cheeses or derivatives of such cheeses at lower mountainous temperatures and with an accidental or deliberate reduction of the sodium chloride content would promote the growth of eye-forming bacteria, resulting in present-day Swiss-type cheeses (*See* **Cheeses**: Cheeses with 'Eyes').

The evolution of varietal cheese development has been, and still is, a continuous process. Varieties introduced within the past century have generally been variants of recognized cheese varieties. None of the existing major cheese varieties was the result of deliberate responses to consumer demands and by technologies developed specifically to achieve the consumer-driven demand. All were the results of accidental geneses and subsequent derivations. Science and technology have been vital in optimizing the quality, uniformity, and ability to modify defined characteristics either within the variety or to produce derivatives of the original variety (*See* Chemistry of Gel Formation; **Cheeses**: Chemistry and Microbiology of Maturation). Science and technology will become increasingly important in controlling the functional characteristics of cheeses in the future. Traditional cheese varieties will be altered to attain unique flavors and physical properties that will differ substantially from the original variety. These derived types are used as food ingredients to enhance and modify the food, to which the cheeses are added, a trend that will increase in the future.

Classification of Cheese

Cheese technologists and merchandizers of cheese have attempted to categorize cheeses based upon various criteria. Segregating cheeses by national origins may be useful for marketing purposes, but this approach has few technological bases for categorization, and the production of a given variety may be globally dispersed. A good example is Cheddar cheese, which originated in England, but now, more is produced in North America and in Oceania than the country of origin. Similarly, production of Swiss cheese and surface-ripened French cheeses occurs throughout the world.

Scientists and technologists have generally categorized cheeses based upon composition, manufacturing methods, and systems for maturing or aging cheese. One scheme differentiates varieties by the type of microbial primary and secondary starters (*See*

Cheeses: Starter Cultures Employed in Cheese-making) used in manufacturing and maturation of cheeses and by the moisture content or, more specifically, the ratio of moisture to protein in cheeses. This scheme would differentiate cheeses on the basis of flavor and physical characteristics but would include only natural cheeses, which are generally defined as those varieties derived from milk or milk components. A similar approach suggested that there were only 18 distinct types of natural cheeses that could be categorized into eight families based on firmness of the varieties and the method of maturation (Table 1).

Slightly more comprehensive classifications differentiated varieties by composition, firmness, and maturation/flavoring agents (Table 2) or by manufacturing and maturation processes (Table 3). The system used in Table 2 expands the differentiation of natural cheeses, includes cheeses made from whey and introduces a new category of cheeses that attribute their dominant flavor to added spices or ingredients. Cheese types are divided into two major categories, natural and process cheeses, and the natural cheeses are divided into those in which the curd is formed either by acid or by enzymes in Table 3. This classification introduces the category of process cheeses, which are produced by heating and melting natural cheeses with the aid of calcium-chelating emulsifying salts (*See* **Cheeses:** Processed Cheese).

The previous classification systems were integrated into a comprehensive categorization shown in Figure 1. Although process cheeses are not included, a detailed differentiation of natural cheeses based upon methods of clotting milk, maturation agents, and composition gives a good overview of the diversity of cheeses. There is also merit in the initial division into four 'superfamilies' based upon differences in curd formation. The method of coagulating milk or milk components (curd formation) yields distinctly different end products because of the chemistry of the coagulum (*See* **Cheeses:** Chemistry of Gel Formation). Although there is heterogeneity within the families of cheeses formed by acid coagulation, heat/acid coagulation, and concentration/crystallization, these differences are less than those between families. The most diverse family contains cheeses formed by rennet coagulation. Dividing this family into classes based on the location and major types of microorganisms that distinguish the classes is useful from technological and marketing standpoints. The most diverse category is the surface-ripened cheeses that have a heterogeneous surface flora that can include molds. However, these cheeses are distinct from the surface mold-ripened cheeses. The major cheese varieties, in terms of breadth of type and of amounts produced, are included in the internal bacterially ripened cheese category. The extra-hard, hard, and semihard cheeses vary in moisture contents (moisture:protein ratios) but are similar in containing a mixture of bacteria within the cheese to create the flavors and desired maturity of the particular type (*See* **Cheeses:** Manufacture of Extra-hard Cheeses; Manufacture of Hard and Semi-hard Varieties of Cheese). The extra-hard types are generally distinguished by high numbers of heat-tolerant lactic acid-producing bacteria that evolved from the practice of heating the curd and whey to high temperatures during manufacturing. The severe heat treatments were not used in the manufacturing of the hard and semihard types, resulting in higher moisture contents; different types of lactic-acid producing bacteria dominating the flora of such cheeses during manufacturing and early stages of maturation (*See* **Cheeses:** Starter Cultures Employed in Cheese-making). These same effects of heat treatments resulted in two classes of cheese with eyes. Higher temperatures and lower concentrations of salt promoted the heat-tolerant lactic bacteria plus the eye-forming propionibacteria in Swiss-type cheeses. Lower temperatures during manufacturing and early maturation of Dutch-type cheeses created the environment for the mesophilic lactic bacteria and lactic bacteria that form eyes through metabolism of citrate in the cheese. The flavor and physical properties of the types with eyes usually differ, but variants of these types can be quite similar, as exemplified by the recently developed Maasdam cheese and highly aged Gouda cheese. The classification in Figure 1 also segregates varieties that evolved from the early practices of preservation by using high levels of sodium

Table 1 Classification of cheeses based on firmness and maturation processes

Very hard (grating)
Ripened by bacteria (e.g., Parmesan)
Hard
Ripened by bacteria that form eyes in cheese (e.g., Swiss)
Ripened by bacteria without eye formation (e.g., Cheddar)
Semisoft
Ripened principally by bacteria (e.g., Gouda)
Ripened by bacteria and surface microorganisms (e.g., Limburger)
Ripened principally by blue mold in the interior of the cheese (e.g., Roquefort)
Soft
Ripened (e.g., Brie)
Unripened (e.g., cottage)

Fox PF, Guinee TP, Cogan TM and McSweeney PL (2000) *Fundamentals of Cheese Science*, pp. 1–18, 388–428. Gaithersburg, MD: Aspen.

Table 2 Classification of cheeses according to composition, firmness, and maturation agents

Soft Cheese (50–80% moisture)
- *Unripened – low-fat*
  - Cottage
  - Quark
  - Baker's
- *Unripened – high-fat*
  - Cream
  - Neufchâtel
- *Unripened stretched curd or pasta filata cheese*
  - Mozzarella
  - Scamorze
- *Ripened by external mold growth*
  - Camembert
  - Brie
- *Ripened by bacterial fermentation*
  - Kochkse
  - Handkse
  - Caciotta (ewe or goat)
- *Salt-cured or pickled*
  - Feta – Greek
  - Domiati – Egyptian
- *Surface-ripened*
  - Liederkranz

Semisoft Cheese (39–50% moisture)
- *Ripened by internal mold growth*
  - Blue
  - Gorgonzola
  - Roquefort (sheep's milk)
- *Surface-ripened by bacteria and yeast (surface smear)*
  - Limburger
  - Brick
  - Trappist
  - Port du Salut, St. Paulin
  - Oka
- *Ripened primarily by internal bacterial fermentation but may also have some surface growth*
  - Mnster
  - Bel Paese
  - Tilsiter
- *Ripened by internal bacterial fermentation*
  - Pasta Filata
  - Provolone
  - Low-moisture Mozzarella
- *Internally ripened by bacterial fermentation plus CO_2 production resulting in holes or 'eyes'*
  - Gouda
  - Edam
  - Samsoe

Hard Cheese (maximum 39–40% moisture)
- *Internally ripened by bacterial fermentation*
  - Cheddar
  - Colby
  - Caciocavallo
- *Internally ripened by bacterial fermentation plus CO_2 production resulting in holes or 'eyes'*
  - Swiss (Emmental)
  - Gruyere
- *Internally ripened by mold growth*
  - Stilton

Very hard cheese (maximum 34% moisture)
- Asiago Old
- Parmesan, Parmigiano, Grana
- Romano
- Sardo

Whey cheese
- *Heat and acid denaturation of whey protein*
  - Ricotta (60% moisture)
- *Condensing of whey by heat and water evaporation*
  - Gjetost (goat milk whey, 13% moisture)
  - Myost, Primost (13–18% moisture)

Spiced cheese
- Caraway – caraway seeds
- Noekkelost – cumin, cloves
- Kuminost – cumin, caraway seeds
- Pepper – peppers
- Sapsago – hard grating, clover

From Olson NF (1995) Cheese. In: Rehm H-J and Reed G (eds) *Biotechnology*, 2nd edn., vol. 9, pp. 355–384. VCH: Weinheim.

Table 3 Classification of cheeses by manufacturing and maturation processes

Natural cheeses
- Cheese varieties in which milk is clotted by acid:
  - Cottage cheese
  - Baker's cheese
  - Cream cheese
  - Neufchâtel cheese
- Cheese varieties in which milk is clotted by proteases:
  - Cheddar cheese
  - Colby and stirred curd (granular) cheese
  - Surface-ripened cheeses – Brick cheese, Limburger cheese, Port du Salut, Bel Paese, Tilsit cheeses
  - Other semisoft cheeses – Edam, Gouda, Monterey, Mnster cheeses
  - Cheeses with eyes – Swiss, Gruyère, Samsoe
  - Italian type
    - Very hard (grating) – Parmesan, Romano
    - Other hard – Asiago, Fontina
    - Pasta Filata – Provolone, Mozzarella
  - Mold-ripened
    - Blue, Roquefort
    - Cheese with surface mold – Camembert, Brie, Coulommiers

Process cheese
- Processed Swiss, processed Cheddar, etc.
- Cold-pack cheese

From Olson NF (1995) Cheese. In: Rehm H-J and Reed G (eds) *Biotechnology*, 2nd edn., vol. 9, pp. 355–384. VCH: Weinheim.

chloride or high heat treatments of the curd (*pasta-filata* varieties). The need for preservation has lessened so that modern versions of the high salt cheeses may contain salt concentrations sufficient to satisfy tastes of consumers. The *pasta-filata* types, especially Mozzarella cheese, have become global favorites as ingredients in a variety of foods, especially pizza. It is unlikely that the originators of Mozzarella cheese envisioned that heat preservation would yield a product that possessed the functional stability and melt properties that fit the requirements as an ingredient in pizza. Acute perception of

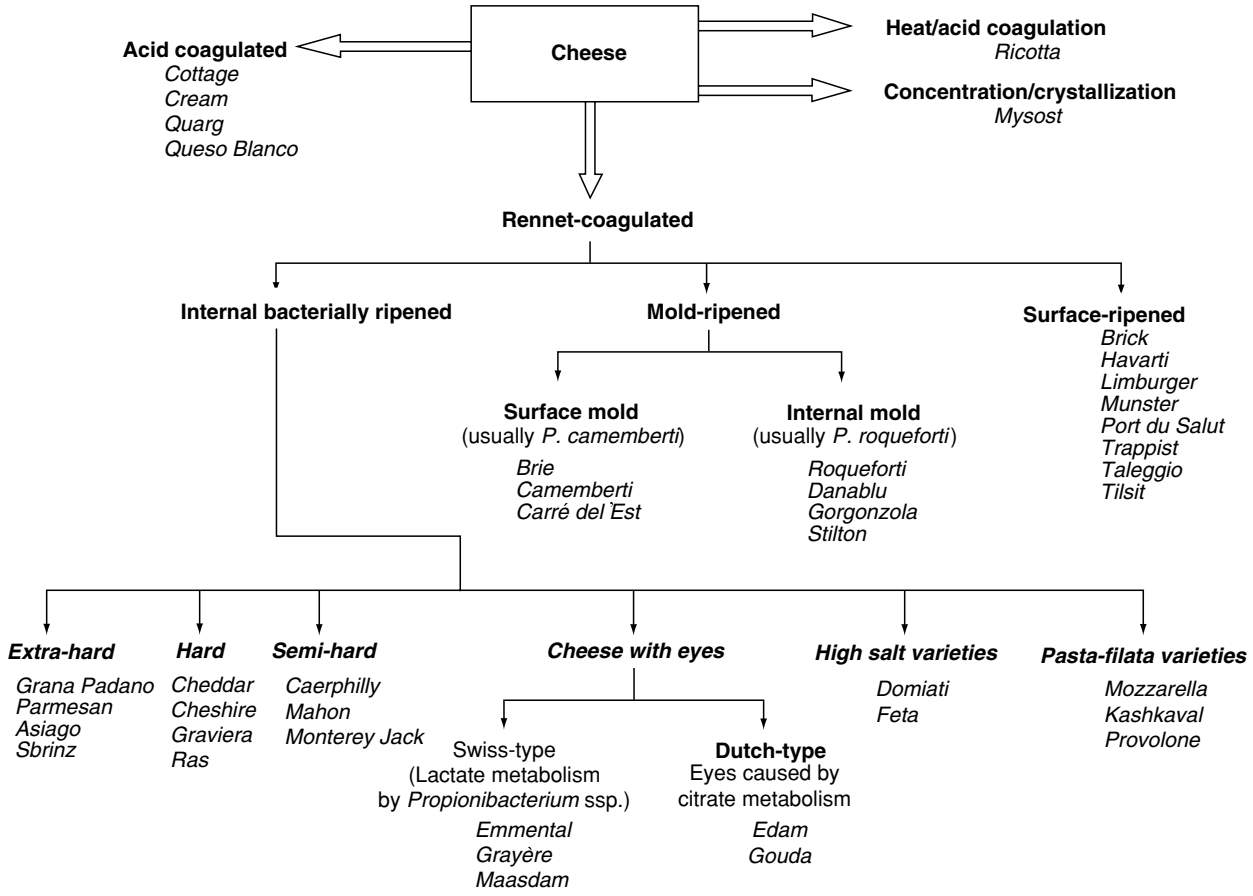

Figure 1 Classification of cheese into superfamilies based on the method of coagulation of milk or milk components with further subdivision based on the principal ripening agents and/or characteristic technology. From Fox PF, Guinee TP, Cogan TM and McSweeney PL (2000) *Fundamentals of Cheese Science*, pp. 1–18, 388–428. Gaithersburg, MD: Aspen, with permission.

accidents in food preparation and subsequent science and technology created this unexpected and widespread usage.

Homogeneity and Diversity

In spite of the great number of cheese types, there are common elements in the manufacture of all varieties of cheese. These are clotting of milk to yield a curd, expulsion of serum/whey from the curd, forming the grains of curd into a final shape, and storage either for very limited periods for unaged cheeses or for varying lengths of time to attain the desired flavor and physical characteristics (*See* **Cheeses:** Chemistry and Microbiology of Maturation). The organization of these elements and the flow of the cheesemaking process are illustrated in Figure 2. The bases for diversity in cheeses are also evident in Figure 2 starting with the type, composition, chemistry, and biology of the milk. Agents and conditions at each of the steps in cheese manufacturing impart specific effects that create further diversity to create a cheese type with unique flavor, texture, and functional properties. Additional discussion of the microbiology, chemistry, and technologies of important cheese types are given in the succeeding articles.

See also: **Cheeses**: Starter Cultures Employed in Cheesemaking; Chemistry of Gel Formation; Chemistry and Microbiology of Maturation; Manufacture of Extra-hard Cheeses; Manufacture of Hard and Semi-hard Varieties of Cheese; Cheeses with 'Eyes'; Soft and Special Varieties; White Brined Varieties; Quarg and Fromage Frais; Processed Cheese; Surface Mold-ripened Cheese Varieties; Surface Mold-ripened Cheese Varieties

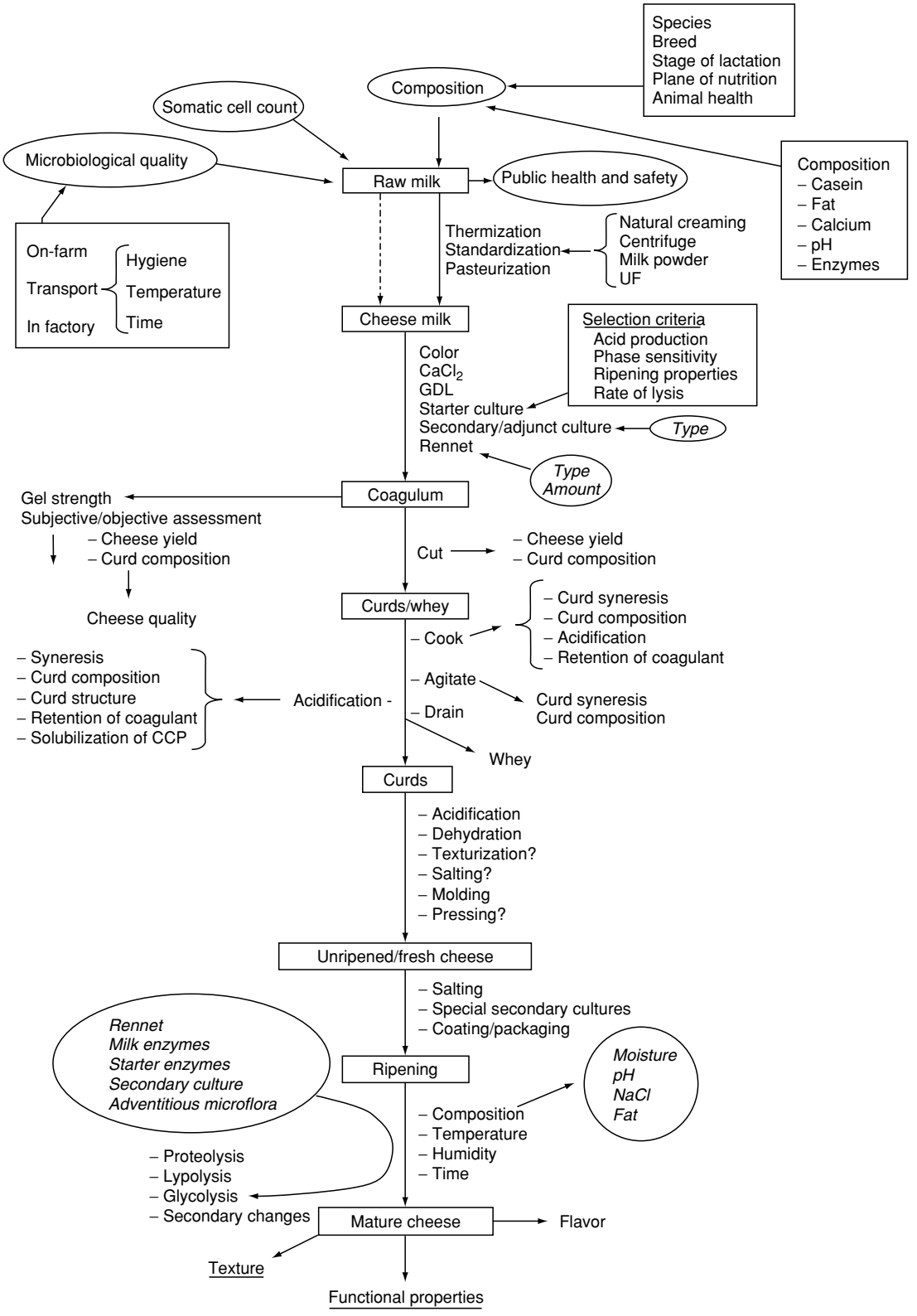

Figure 2 Flow diagram of cheese manufacturing to show the common elements for all types and the ingredients, factors, and technologies that create the diversity of cheese types. From Fox PF, Guinee TP, Cogan TM and McSweeney PL (2000) *Fundamentals of Cheese Science*, pp. 1–18, 388–428. Gaithersburg, MD: Aspen, with permission.

Further Reading

Codex Committee on Milk and Milk Products (2000) *Proposed Draft Amendment to the Codex Standard for Cheese*. 4th Session of the Codex Committee on Milk and Milk Products, Appendix IV. Wellington, New Zealand, February 28 to March 3, 2000.

Fox PF (1999) Cheese: an overview. In: *Cheese: Chemistry, Physics and Microbiology*, pp. 1–35. Gaithersburg, MD: Aspen.

Fox PF, Guinee TP, Cogan TM, and McSweeney PL (2000) *Fundamentals of Cheese Science*, pp. 1–18, 388–428. Gaithersburg, MD: Aspen.

Olson NF (1995) Cheese. In: Rehm H-J and Reed G (eds.) *Biotechnology*, 2nd edn. vol. 9, pp. 355–384. Weinheim: VCH.

Starter Cultures Employed in Cheese-making

G Stanley, Rhodia Food UK Ltd, Woodley, Stockport, UK

Background

The bacteria used in cheesemaking are commonly referred to as 'lactic acid bacteria' (LAB), having lactic acid as their main end product of lactose and other sugar fermentations. It is this acid that plays a major role in the preservation of cheese. Cheesemaking LAB have a 'Generally Regarded As Safe' (GRAS) status and are roughly divided into mesophilic and thermophilic types. Commercial cheese cultures are used for the production of the majority of the developed world's cheeses and for the most part are balanced blends of selected single strain cultures. Such preparations are offered in either a direct vat inoculation (DVI) or a bulk starter (BS) format.

Development of Starter Cultures

LAB occur naturally in raw milk, and initially, cheesemaking relied on these indigenous bacteria to promote acid production; however, this must have resulted in very variable fermentations. Cheesemakers soon developed the technique of inoculating fresh milk with whey from the previous day's cheesemaking – a crude form of starter inoculation, but a technique that is still used in some cheesemaking operations today. The first commercial cultures were prepared as powdered starters by Chr. Hansen Laboratories (Denmark) just before the end of the nineteenth century. These were likely to have been mixtures of strains of unknown composition, commercial undefined starters, and for the first half of the twentieth century, such commercial mixtures supplied to creameries in liquid, frozen, or dried form were the dominant starter types. The discovery of the bacterial nature of starters owes much to the work of Pasteur (mid 1800s), Lister, who isolated *Bacterium lactis* (now called *Lactococcus lactis*) in 1873, and Orla-Jensen, who published his thesis on the nature of starter cultures in 1919. Whitehead (New Zealand), in the 1930/40s, pioneered the concept of single strains, while identifying bacteriophage (phage) as an important causative agent in slow cheese fermentations. This use of (blended) single strains, well characterized for phage resistance, acid production, and flavor delivery has now been widely adopted.

Concentrated starter cultures, developed in the 1960s and 1970s, were prepared in freeze-dried or deep-frozen form, which required simple inoculation into a bulk starter vessel (containing, traditionally, milk, or reconstituted milk powder or a phage-inhibitory whey-based phosphated medium). Improvements to this concentration process have led, in the last 10–20 years, to the development and commercial acceptance of DVI cultures that require no subculture or activation prior to use. In the UK in 2000, more than 50% of all cheese was made using DVI systems.

Microbiology of Starter Cultures

LAB can be divided broadly into mesophiles and thermophiles, which have an optimum growth temperature of approximately 32 or 45 °C, respectively.

Mesophilic Starters

Lactococcus lactis (*Lc. lactis*) is the dominant acidifying mesophile species used, with *Leuconostoc mesenteroides* ssp. *cremoris*, a weakly acidifying culture, being used to impart flavor in certain (fresh) dairy fermentation. *Lc. lactis* has two subspecies, namely *cremoris* and *lactis*, and also a biovariant *diacetylactis*, now referred to as *Lc. lactis* citrate+, which is able to catabolize citrate (found in milk at approximately 0.2%) to carbon dioxide and the flavor compound, diacetyl – both of these are technologically important products. Traditionally, strains of *Lc. lactis cremoris* have been regarded as producing the best flavors in cheese, whereas some strains of *Lc. Lactis* ssp. *lactis* are noted for promoting certain off-flavors (bitter, malty, fruity). However, better selection techniques have identified many *Lc. Lactis* ssp. *lactis* strains suitable for delivering good flavors and,

because of their faster rates of lactic acid production and better yields in concentrated culture preparation, are now regularly used in commercial starters, usually blended with *Lc. lactis* ssp. *cremoris* strains.

Thermophilic Starters

The most important thermophilic LAB are *Streptococcus salivarius* ssp. *thermophilus* (*Sc. thermophilus*) and species of *Lactobacillus* (Lb): *bulgaricus*, *helveticus* and *lactis*. They are used in those technologies where a temperature >40 °C is used, such as yogurt, Mozzarella, Emmenthal, and Gruyere. Generally, these fermentations use blends of strains of *Sc. thermophilus* and a *Lactobacillus* (particulary *Lb. bulgaricus*), there being an associative growth relationship between them: *Sc. thermophilus* produces formate that stimulates *Lb. bulgaricus*, and *Lb. bulgaricus* produces amino acids that stimulate *Sc. thermophilus*.

Some differentiating characteristics of mesophilic and thermophilic acidifying starter cultures are given in Table 1.

Function of Starter Cultures

Three main technological functions can be attributed to LAB starter cultures, namely acidification, texture enhancement, and flavor development. Acidification, the metabolism of lactose to lactic acid, is the major attribute of LAB; the method of this metabolism differs in the various species.

Lactococci actively transport the lactose across the cell wall membrane as lactose-phosphate, which is hydrolyzed to glucose and galactose-6-phosphate; the glucose molecule is then metabolised to L(+) lactate by the glycolytic pathway, and the galactose is metabolized *via* the tagatose-6-phosphate pathway. The thermophilic LAB also use the glycolytic pathway, but they differ from lactococci in several respects, namely the lactose uptake mechanism, and that *St. thermophilus* and *Lb. bulgaricus* are not able to metabolize galactose, whereas *Lb. helveticus* metabolizes it *via* the Leloir pathway. The isomeric form of lactate produced also varies between the different species (see Table 1). It should be noted that for starters, the fermentation of lactose is their energy-producing mechanism, and lactate is merely a waste product. The LAB starters are referred to as 'homofermentative cultures,' i.e. they produce one main end product (lactate); in contrast, *Leuconostocs* can ferment lactose to lactate, ethanol, and carbon dioxide *via* the phosphoketolase pathway, and are regarded as heterofermentative bacteria.

The process of acidification enhances the expulsion of whey from the curd during the cheesemaking process and promotes the development of the texture of the cheese. The final pH obtained in the cheese will dictate to a large extent the texture of the curd: in very acid curds, e.g., Stilton, Cheshire (pH 4.6–4.8), the texture is described as short, noncohesive and crumbly due to the low level of calcium and the small size of the casein aggregates, whereas high pH cheeses (e.g., Edam/Gouda, pH 5.3–5.5) have a springy or plastic texture with a high level of calcium and larger casein aggregates.

Some LAB are able to polymerize galactose and glucose to give polysaccharide molecules, these can be important in modifying the texture of fermented dairy products – they are particularly useful in yogurts, and in low and reduced fat products, as they can improve texture and act as a fat replacer. Strains of *Sc. thermophilus* have been particularly useful in this respect. Eye formation is important in certain cheese (e.g., Dutch varieties), and occurs as a result of carbon dioxide produced by the fermentation of citrate by citrate+ *Lc. lactis* ssp. *diacetylactis* and *Leuconostoc* strains in the starter.

LAB are important in flavor development due to both the end products of fermentation (particularly diacetyl from citrate fermentation) and the slow proteolysis of the milk proteins to peptides and amino acids by the protinases and peptidases of the LAB that are released into the curd matrix as the cells die and autolyze during the cheese maturation.

A fourth function often attributed to LAB is that of the health benefits, which are further discussed under the probiotic section of secondary cheese cultures.

Selection, Production, and Use of LAB

With the increasing use of LAB cultures in a concentrated form, emphasis has been placed on the selection process for isolating strains, the functions of which match the cheesemaking requirements. Raw milk, traditional mixed starters, and dairy products, in general, are a valuable source of natural LAB; selection tests that are conducted on strain isolates from these sources include:

1. flavor promotion, absence of off-flavors;
2. acidification rates at different temperatures;
3. phage insensitivity;
4. compatability with other strains (absence of bacteriocin/inhibitor production);
5. temperature sensitivity;
6. salt sensitivity;
7. proteolytic activity;
8. special attributes, e.g., polysaccharide production;
9. cell yield in industrial fermentation;

Table 1 Some differentiating characteristics of starter lactic acid bacteria

Type	*Species*	*Shape*	*Growth at*			*Lactate isomer*	*Citrate metabolism*	*Galactose metabolism*	*Ammonia from arginine*	*Percentage lactic acid produced in milk*	*Percentage salt inhibition*	*Important metabolic products*
			10 °C	*40 °C*	*45 °C*							
Mesophilic	*Lactococcus lactis* ssp. *lactis*	Cocci	+	+	−	L (+)	−	+	+	0.8	4.0–6.5	Lactate
	Lactococcus lactis ssp. *lactis* (−cit+)	Cocci	+	+	−	L (+)	+	+	+	0.8	4.0–6.5	Lactate, diacetyl, CO_2
	Lactococcus lactis ssp. *cremoris*	Cocci	+	−	−	L (+)	−	+	−	0.8	2.0–4.0	Lactate
	Leuconostoc mesenteroides ssp. *cremoris*	Cocci	+	−	−	D (−)	+	+	−	0.2	2.0–4.0	Lactate, diacetyl, CO_2
Thermophilic	*Streptococcus thermophilus*	Cocci	−	+	+	L (+)	−	−	−	1.2	< 2.0	Lactate, acetaldehyde
	Lactobacillus bulgaricus	Rods	−	+	+	D (−)	−	−	−	1.8	< 2.0	Lactate, acetaldehyde
	Lactobacillus helveticus	Rods	−	+	+	DL	−	+	−	2.0	< 2.0	Lactate
	Lactococcus lactis	Rods	−	+	+	D (−)	−	−	+/−	1.8	2.0–4.0	Lactate

10. stability during cell concentration;
11. stability in freeze-dried or frozen format;
12. acid-producing activity as in DVI cultures.

Genetic techniques developed for bacteria have allowed the further modification of selected cultures to improve their functionality for cheesemaking purposes. Initial studies in this area concentrated on improving the phage resistance of cultures, with the knowledge that certain plasmids in LAB coded for functions relating to the phage infection process. Other areas that have been targeted for such development include food safety (ability of LAB to inhibit pathogen growth), and flavor (transfer of the genes of protein/peptide degrading enzymes, transfer of the citrate fermentation genes coded on plasmids). Although much research has been conducted on the genetics of LAB, its transference into commercial applications has received a setback, particularly in Europe, with the strength of the antigenetic modification lobby.

Once a strain has been selected as a suitable commercial culture, it will be stored as seed cultures in a culture bank (freeze-dried or deep-frozen) in multiple copies in an easy to use form ready for future commercial productions. The seed cultures are activated and grown in a small volume of a selected sterile nutritive medium, then inoculated into an industrial fermenter and grown under strict time, temperature, and pH conditions. The use of pH control (usually at pH 6.2–6.5 for mesophilic and pH 6.0–5.5 for thermophilic LAB) to neutralize the lactic acid produced allows the culture to attain a level of 10e10 bacteria per milliliter. Following concentration by either centrifugation or membrane techniques, the final cell numbers achieved are between 10e11 and 10e12. The cultures are then standardized for their acid-producing activity or cell numbers and frozen (in pellet form for DVI cultures) or freeze-dried; often, cryoprotective agents (e.g., sugars such as sucrose and lactose) are added at this stage to enhance culture stability. Finally, the cultures are blended in predetermined ratios and packaged in sachets or cartons and enter commercial stock once they have passed the appropriate bacteriological/activity quality control tests. Culture blends usually contain between two and six strains and are prepared on the basis of activity, flavor promotion, and phage relationships of the individual components. Mesophilic LAB are used for the majority of cheese fermentations (Dutch, Cheddar, French mold-ripened, cottage, soft cheeses), with thermophilic cultures being used mainly for Italian (Parmasan, Mozzarella, and Pizza) cheeses, Swiss cheese varieties, and yogurt. However, there is an increasing trend to blend mesophilic and thermophilic LAB for certain technologies such as Cheddar and Brie cheeses. The inclusion of *Sc. thermophilus* in mesophilic cultures, together with an increase in the vat milk temperature, produces a 'stabilized' Brie that has a more uniform texture and longer shelf-life than traditional products. In Cheddar DVI cultures, the inclusion of *Sc. thermophilus* enhances culture activity, reduces bitter notes, and increases phage durability; DVI cultures containing *Sc. thermophilus* now account for at least 30% of UK Cheddar. *Lb. helveticus* strains are also being used commercially for Cheddar to enhance and modify flavor profiles, and these strains are being offered either as adjunct cultures or preblended with the DVI culture.

The preparation of bulk starter in the cheese plant involves inoculating the bulk starter culture (usually freeze-dried, in a sachet, or deep-frozen, in a ring-pull can – in the latter case, partial thawing of the culture is required prior to inoculation) into the precleaned and sterilized bulk starter vessel. The growth medium normally used is either skim milk/reconstituted skim milk powder (10–12%) or a blended powder (6%), containing a protein source (whey/milk powder), a sugar source (whey powder/lactose), a bacterial growth stimulant (yeast extract), and a buffering/phage inhibitory agent (phosphate salts). The inoculated culture is grown for a predetermined time at a constant temperature (e.g., 18 h at 22–24 °C for mesophilic cultures, and 5–6 h at 40–42 °C for thermophilic cultures). During incubation, the pH evolution may be monitored, and also pH control may be used, whereby alkali (ammonia) is automatically dosed into the vessel to maintain the pH at the optimum level for the culture (approximately pH 6.2–6.5 for mesophilic cultures and pH 6.0–5.5 for thermophilic cultures). Once the required end point has been reached (as determined by the pH/acidity level, or quantity/rate of alkali addition), the bulk starter is ready for use, ideally after rapid cooling.

Secondary Cheese Cultures

In addition to the mesophilic and thermophilic starters described above, other microbial cultures are used in cheese to produce the flavors, textures, and appearance typical for that cheese type.

Molds

For Coulommier and Camembert type cheese, the white mold, *Penicillium camemberti*, is used to give a white felt on the surface of the cheese. Strains have been isolated that exhibit different mycelial growth characteristics, and in general, four types have been identified:

1. Neufchâtel form (rapid growth that produces a thick mat of yellow–white mycelia);
2. short, flat mycelial form (rapid growth and tight flat thallus);
3. long, tall mycelial form (slow growth, loose high thallus);
4. fluffy woolly mycelial form (white thallus that turns gray on aging, originally called *Penicillium album*);

The mold is added to the cheese as a spore suspension either in the vat or sprayed on at the end of cheesemaking (or both).

Geotrichum candidum is another surface mold, which produces a very short, fine mycelium. It is generally used together with *P. camemberti* and exerts a synergistic effect on its growth. Some strains of *G. candidum* produce a yeast-like appearance on the surface of the cheese, often referred to as 'toad-skin.'

Penicillium roqueforti is the mold culture used for blue-veined cheeses, such as Roquefort, Stilton, Gorgonzola, and Danish Blue. Unlike *P. camemberti*, it generally grows inside the cheese, but like *P. camemberti*, it requires oxygen to grow; therefore, after cheesemaking, the cheese must be pierced to allow sufficient oxygen to penetrate for spore germination and for the carbon dioxide to be expelled.

Yeasts

These form an important part of the microflora of many surface ripened cheeses, such as blue cheeses, soft mold-ripened cheeses, and washed-rind cheeses, e.g., St. Nectaire, St. Paulin, Cantal. *Kluyveromyces lactis*, *Saccharomyces cerevisiae*, *Candida utilis*, *Debaryomyces hansenii*, and *Rodosporidium infirmominiatum* are the most common yeasts used. They colonize the surface of the cheese at an early age, are able to grow in 4% NaCl, and have a strong neutralizing activity by virtue of their ability to metabolize lactate. They are both proteolytic and lipolytic and produce a range of volatile and peptide/amino acid flavor components.

Bacteria

Groups of bacteria that are important in the secondary flora of cheese are *Micrococcus*, *Lactobacillus*, *Corynebacterium*, and *Propionibacterium*.

Brevibacterium linens (*Corynebacteria*) is responsible for the red–orange color and characteristic aroma (originating from the catabolism of sulfur-containing amino acids) of surface-smear and washed-rind cheeses, such as Munster, Maroilles, Pont L'Eveque and St. Paulin.

Micrococci cultures are facultative anaerobes and are used to promote texture and flavor in both hard, semihard and mold-ripened cheeses.

Mesophilic *Lactobacillus* species are used in washed-rind cheeses and in some hard and semi-hard cheeses; they are of value because of their aminopeptidase activity, which promotes flavor and can reduce bitterness during cheese maturation.

The microbiology of the surface smear and washed rind cheeses is complex; combinations of Corynebacteria, Micrococci, molds, and yeasts are present. The Corynebacteria grow once the pH has increased due to the degradation of lactate by the yeasts and molds.

Propionibacteria, principally the species *Proprionibacterium freudenreichii* and *Shermanii*, are used in Swiss cheese technology (Emmenthal, Gruyere, Appenzell). They convert the lactate produced by the *Sc. thermophilus* and *Lb. helveticus* starters to the flavor components, propionic and acetic acids, and carbon dioxide, which is responsible for eye formation.

Probiotic bacteria form an addition group that are used for their potential health-giving properties. The main cultures used are certain *Lactobacillus* species (particularly *Lb. acidophilus*) and *Bifidobacterium* species, and are generally considered to be natural inhabitants of the gut. They are usually added with starter LAB, particularly the thermophilic LAB, since they do not themselves normally produce sufficient lactic acid for standard fermentations. They are used mainly in the production of fresh fermented dairy products (e.g., yogurts). Health benefits associated with them include improved digestion, stimulation of the immune system, reduction of tumor incidence, and reduction of serum cholesterol. However, it has proved difficult to obtain medical evidence in this respect that is acceptable to health authorities.

Factors that Affect Culture Growth

The growth of LAB in milk can be affected by many parameters. For example, variations in milk composition, either seasonal or cattle feed-influenced, can have a stimulatory or inhibitory effect on LAB. Excessive entrainment of oxygen in the milk, as a result of agitation, can cause inhibition due to the bacterial conversion of oxygen into the inhibitory substance hydrogen peroxide. LAB generally do not have the enzyme catalase to degrade this. Milk may contain antibiotic residues following the administration of antibiotics for cow mastitis; however, strict quality-control measures have significantly reduced this problem of recent times. Residues of cleaning and sterilizing chemicals may be present in the cheesemaking vats

and could inhibit starters if present at sufficiently high concentrations.

Nonadherence to the time/temperature regimes of the cheesemaking process can also affect acidification, e.g., overscalding by 1–2 °C in Cheddar cheesemaking would cause significant slowing of the rate of acidification for a culture containing predominantly *Lc. lactis* ssp. *cremoris* strains.

Phage is probably the most important destructive agent in cheesemaking. Phage is found in milk and can lie dormant (lysogenic state) within some wild type and starter LABs. These are viruses with a hollow protein head containing DNA and a hollow protein tail with a plate assembly at the end. Phage attack LAB by attaching themselves to the bacterial surface by means of the plate assembly, injecting their DNA through the tail and into the host cell; once inside, their DNA directs the host to produce more phage particles, and, when mature (30–50 min), lysin is produced to dissolve the bacterial cell wall and release the new phage; up to 200 particles can be produced in one such cycle (in the same time span, one LAB cell would have divided to give two cells). Phage are present in most, if not all, cheesemaking plants, and it is only through the use of good manufacturing practices and hygiene control, together with the selection of suitable LAB, that phage is maintained at low, noninvasive levels. Culture techniques aimed at reducing/eliminating phage problems, which have already been touched on in this article, include:

1. selection of naturally resistant strains;
2. genetic modification (e.g., plasmid transfer), where legally permitted;
3. blending several phage resistant/unrelated strains;
4. use of culture rotation, three or four blends (a different blend used on different days);
5. inclusion of *Sc. thermophilus* in mesophilic blends (used in Cheddar technology).

See also: **Lactic Acid Bacteria**; **Starter Cultures**; **Yeasts**

Further Reading

Cogan TM and Hill C (1993) Cheese starter cultures. In: Fox PF (ed.) *Cheese Chemistry Physics and Microbiology*, vol. 1, pp. 193–255. London: Chapman & Hall.

Davies FL and Law BA (eds) (1984) *Advances in the Microbiology and Biochemistry of Cheese and Fermented Milk*. Barking, UK: Elsevier Applied Science.

Gasson MJ and de Vos WM (eds) (1994) *Genetics and Biotechnology of Lactic Acid Bacteria*. London: Chapman & Hall.

Salminen S and von Wright A (eds) (1993) *Lactic Acid Bacteria*. New York: Marcel Dekker.

Stanley G (1998) Microbiology of fermented milk products. In: Early R (ed.) *The Technology of Dairy Products*, pp. 2–80. London: Chapman & Hall.

Tamime AY (1983) Microbiology of starter cultures. In: Robinson RK (ed.) *Dairy Microbiology*, vol. 2, pp. 113–156. Barking, UK: Applied Science.

Chemistry of Gel Formation

J M Banks and D S Horne, Hannah Research Institute, Ayr, UK

Background

The basic ingredient for cheese is milk, usually from cows, sheep, or goats, although the milk of yak, reindeer, camels, mares, and buffaloes is also used in local, indigenous products in some regions of the world. In nature, milk is produced by the mother mammal to feed her newborn offspring; milk, therefore, has to provide all elements essential for their growth and nourishment, namely protein, fat, carbohydrate, minerals, and water, but in proportions relevant to the needs of the animal species. Recognizing the nutritional benefits, man initially sought to preserve this fundamental food for the lean times of low milk availability. Cheese manufacture, which retains the majority of the protein, fat, and some of the minerals, provides one such outcome of this exercise of his ingenuity, with a wide range of products to suit all tastes and desires in the modern portfolio of the cheesemaker's art.

Cheesemaking capitalizes on the curdling of milk. Curds can be produced by either acidification of milk to pH 4.6 or the use of a protease in conjunction with acidification to pH 6.4. The former method is used to produce fresh high-moisture cheeses, e.g., cottage cheese, and the curd is formed by the isoelectric precipitation of milk casein. The latter approach is the basis for producing hard ripened cheeses, such as Cheddar, Emmental, or Parmesan, which have a low moisture content. Gel formation is initiated by proteolysis of casein, and the destabilized casein aggregates to form a gel. The gel produced at the higher pH using a protease differs markedly from that produced by acid alone. The gel is more elastic, and in the presence of increasing acidity and the application of heat, the gel shrinks and expels moisture by a process termed syneresis. To understand the process of curd formation, we must look more closely at the individual components of milk to discern their role,

particularly the fat and protein, and also the agents used in proteolysis.

Fat

Fat in milk exists as small globules surrounded by me mbrane proteins and in a size range dependent on breed, lactational status, and diet of the cow. The fat in milk contributes to flavor development in mold-ripened cheeses with a strong lipolytic flavor, e.g., blue cheeses. In cheeses such as Cheddar or Gouda, in which proteolysis is dominant during maturation, the fat acts as a solvent for flavor compounds. The fat globules are trapped in the protein network created during curd formation; hence, their size and the network mesh size interact in determining overall yield of cheese. Otherwise, the fat globules play no active role in determining curd properties.

Protein

Two types of protein are found in milk: the globular whey proteins, which are soluble in the serum phase, and the caseins, which exist in a stable colloidal suspension of aggregates known as casein micelles. Cheesemaking exploits the destabilization mechanisms nature has built into this colloidal system by using the natural enzyme, chymosin, originally extracted from the stomach of the calf, to hydrolyze the κ-casein and induce the destabilization of the micelle to form the gel, which goes forward through to the finished cheese. Nowadays, a number of alternatives to the calf stomach extract of chymosin are commercially available for cheese manufacture. These include microbial rennets and cloned chymosin produced using yeast or fungi as the host organism.

Casein Micelle Structure

The physicochemical properties of the different casein proteins dictate how the casein micelle is assembled, and they govern the response of the micelle to destabilizing influences. The caseins and casein micelles have been studied extensively over a long period. Several different models of micelle assembly have emerged. Criticisms have been leveled at most of these models, though a more recent dual-binding model answers most of these points and plausibly explains micellar assembly and stability.

In bovine milk, the casein micelles have an average molecular weight of $\sim 10^8$ Da and a mean diameter of ~ 100 nm (range 50–600 nm). The micelles are very open, highly hydrated structures with a typical hydration value of 2–3 g of H_2O per gram of protein, depending on the method of measurement. They contain almost all of the casein protein and a large fraction of the mineral calcium phosphate found in milk. The casein proteins are four distinct gene products, designated α_{s1}-, β-, α_{s2}-, and κ-casein, and are present in a ratio of approximately 40:35:10:12 in cows' milk. Two posttranslational modifications of the proteins newly synthesized in the mammary gland, have a major impact on their physicochemical, functional, and assembly properties. These reactions are glycosylation and phosphorylation. In bovine milk, only κ-casein is found glycosylated, with several threonine and occasionally serine residues in the hydrophilic C-terminal end of the κ-casein molecule carrying relatively short sugar chains. These chains increase the hydrodynamic bulk and hydrophilic character of this end of the κ-casein molecule.

The second posttranslational reaction is phosphorylation. All of the caseins are phosphorylated at serine, or rarely threonine, to varying extents. The phosphorylation reaction requires a particular sequence template, —Ser —X —A —, where X is any amino acid, and A is Glu, SerP, or rarely Asp. The pattern of serine residues along the molecular sequences of α_{s1}-, α_{s2}-, and β-casein ensures that most of the phosphorylated residues are found in clusters in these molecules. κ-Casein is unique amongst the caseins in the absence of phosphoseryl clusters. Most κ-casein molecules carry only one phosphoseryl residue. Some have two or three, but all are singlets and, again, are located in the hydrophilic C-terminal region. The presence or absence of these phosphoseryl clusters has important consequences for micellar assembly and structure.

The caseins are not globular proteins locked into a secondary conformational structure and are not truly random coil polymers. They have been described as rheomorphic proteins, indicating that the conformational structure adopted is dictated by, and responsive to, the molecular environment. The caseins have little α-helical structure, no denaturation temperature, and a high hydrodynamic volume.

When categorized as nonpolar hydrophobic or polar/charged hydrophilic groups, the amino acids of the caseins show distinct segregation into different regions of the molecular chains. Each casein possesses its own pattern (Figure 1), and this helps rationalize their selfassociation behavior and the conformations adopted at a hydrophobic interface. Together with the phosphoseryl clusters in the hydrophilic loops/tails of α_{s1}-, α_{s2}-, and β-caseins, these sequence patterns allow modeling of the micellar assembly as the polymerization process described below.

In the dual-binding model, two distinct forms of bonding take part in micellar assembly and growth. These are cross-linking through hydrophobic regions

of the caseins or through phosphoseryl cluster bridging across calcium phosphate nanoclusters, small mineral crystallites the precipitation of which is regulated by the presence of the caseins. These cross-linking mechanisms are depicted in Figure 2. Each casein molecule effectively functions as a block copolymer, as detailed in Figure 1, with each block possessing different and possibly multiple functionality for the two distinct cross-linked paths, which can be traced through the model micellar network. Thus, α_{s1}-casein can chain-polymerize through the hydrophobic blocks, B, giving a worm-like chain appearance, as highlighted in Figure 2. Further hydrophobic blocks can also link to these duplexes to create branches, but limitations on growth are likely because of electrostatic repulsion from the neighboring contiguous hydrophilic regions of these molecules. That is unless, as postulated for the casein micelle, the chains also polymerize through the negatively charged phosphoseryl cluster cross-linking to a positively charged calcium phosphate nanocluster, with the potential for multiple protein binding to each nanocluster allowing the second different pathway to be extended. β-Casein, with only two blocks, the phosphoseryl cluster (P) and the hydrophobic region (B), can form links into the network through both and thus facilitate chain growth through these. α_{s2}-Casein is envisaged in this model as having two of each block (Figure 1), two phosphate clusters, and two hydrophobic regions. It is only a small fraction of total casein, but by being able to sustain growth through all its blocks, it is likely to be bound tightly into the network. κ-Casein is the most important of the caseins in this dual-binding model of micellar assembly. It can link into the growing chains through its hydrophobic N-terminal block, but its carboxy-terminal block, denoted C in Figure 1, has no phosphoserine cluster and therefore cannot extend the polymer chain through a nanocluster link. Chain and network growth are therefore terminated here. The occurrence of this event on all the growing chains leaves the casein micelle network with a surface layer of κ-casein molecules, limiting micellar growth and ensuring the stability and integrity of the system.

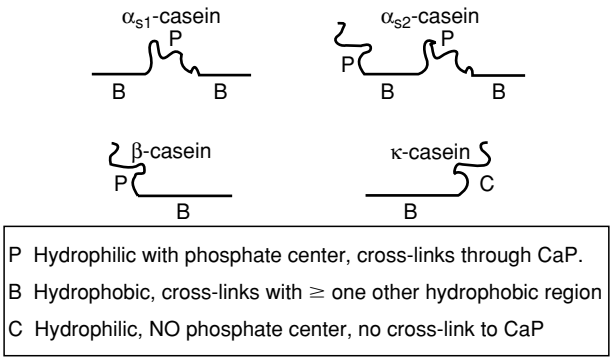

Figure 1 Schematic description of segregation of hydrophobic/hydrophilic residues in the caseins. Hydrophobic regions (B) are represented by the bar structures, which do not imply any form of rigidity; hydrophilic loop regions containing phosphoserine clusters are denoted P and the hydrophilic carboxy peptide of κ-casein by the letter C.

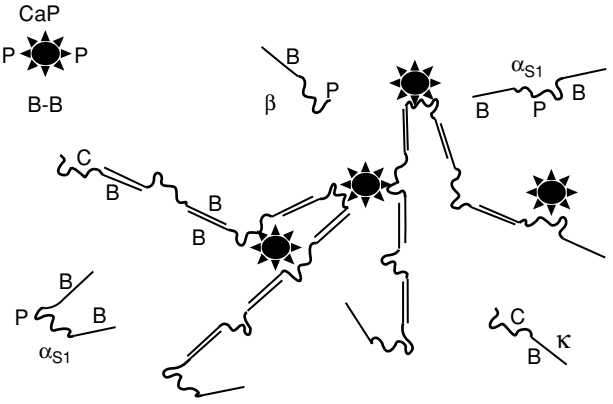

Figure 2 Part of the polymer network built up in the casein micelle according to the dual-binding model. Bridging of casein molecules across calcium phosphate nanoclusters, cross-linking of caseins via hydrophobic interactions, and chain termination by κ-casein are all illustrated. The hydrophobic pathway follows along the straight bars; the nanocluster cross-link points are indicated by the starred circles. Other chains can attach at these points to extend the network or terminate it in the case of κ-casein attaching to a hydrophobic region.

Such a surface location for κ-casein is demanded by experimental evidence linking micellar size and casein composition. Casein micelles can be fractionated from the initial wide size distribution found in milk by subjecting the milk to successive centrifugation steps of increasing speed and duration. A short, low-speed centrifugation sediments the largest micelles. Subjecting the supernatant to a higher-speed, longer-duration stage produces a second pellet of slightly smaller micelles and a supernatant for the next stage, and so on until a clear supernatant is obtained. The end result is a series of pellets, containing largely monodisperse slices of the micellar size distribution. The pellets can be analyzed for protein and mineral composition. Their resuspension in milk ultrafiltrate allows measurement of their size by dynamic light scattering. Such experiments demonstrate that the proportion of α_{s1}- and α_{s2}-caseins is constant with micelle size and that κ-casein is inversely proportional to size. For an object such as a sphere, the surface-to-volume ratio is inversely proportional to the radius,

so such results imply that the κ-casein component resides predominantly on the micellar surface where it controls micellar surface area and hence micellar size. Model and experiment are fully consistent in the role and location of κ-casein, which also readily explains the destabilization of the micelle system on proteolysis of the κ-casein by chymosin. This proteolysis of the κ-casein is termed the primary enzymatic phase of gel formation, and it is followed by a secondary nonenzymic aggregation stage.

Primary Enzymatic Phase of Curd Formation

The κ-casein molecules provide a steric stabilizing layer with their hydrophilic C-terminal peptides protruding into the aqueous phase. Gel formation is a direct consequence of the proteolysis of the κ-casein molecules, the casein being cleaved specifically at the Phe_{105} — Met_{106} bond, releasing the hydrophilic peptide, termed the caseinomacropeptide, into the serum phase, leaving the N-terminal region bound into the micelle network, and thus destabilizing the micelle.

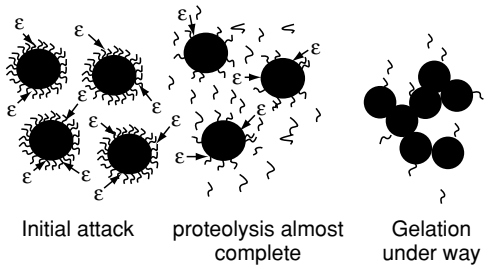

Figure 3 Depiction of the various stages involved in curd formation following proteolysis of κ-casein by chymosin.

Secondary Aggregation Stage of Gel Formation

The gradual loss of the caseinomacropeptide is accompanied by a decrease in the micellar zeta potential from ~−20 to ~−10 mV. The overall process of destabilization and aggregation to a gel network is depicted schematically in Figure 3. Approximately 80% of the κ-casein hairs need to be cut before significant aggregation to define a coagulation time becomes detectable. During this lag phase, the lower levels of κ-casein proteolysis are accompanied by a measurable decrease in micellar hydrodynamic size, as the hairs are lost.

During the development of the gel, the micelle continues to behave as if it were a hard sphere that had become sticky or adhesive with the loss of the stabilizing κ-casein hairs. The gel is therefore initially a network of aggregated particles. As such, it can be predicted that the gel stiffness should be inversely proportional to the cube of the micelle size and directly proportional to protein concentration. Such behavior has been observed experimentally in gels made from micellar fractions, pelleted as described above.

This mechanistic model also helps to explain the influence of the genetic variant of κ-casein on the cheesemaking properties of milk. The C-terminal sequences of the major genetic variants, denoted A and B, of κ-casein are listed in Figure 4. κ-Casein B differs from κ-casein A only at two residues with an isoleucine substituted for a threonine, a potential glycosylation site, at position 136 and an alanine substituted for an aspartic acid residue at position 148. Milk from cows which are homozygous for the κ-B variant have a shorter coagulation time, show a faster rate of gel firming and give a stiffer gel than that from κ-AA cows. All of these changes mean that higher cheese yields are recorded for κ-BB milks. The genetic modifications reduce the net charge of the hydrophilic

- 101 105↓106
  Pro-His-Leu-Ser-Phe-Met-Ala-Ile-Pro-Pro-Lys-Lys-Asn-Gln-Asp-LysThr-Glu-Ile-Pro-
- 121 Ile (Variant B)
  Thr-Ile-Asn-Thr-Ile-Ala-Ser-Gly-Glu-Pro-Thr-Ser-Thr-Pro-Thr- -Glu-Ala-Val-Glu-
  Thr (Variant A)
- 141 Ala (Variant B)
  Ser-Thr-Val-Ala-Thr-Leu-Glu- -SerP-Pro-Glu-Val-Ile-Glu-Ser-Pro-Pro-Glu-Ile-Asn-
  Asp (Variant A)
- 161 149
  Thr-Val-Gln-Val-Thr-Ser-hr-Ia-Val.OH

Figure 4 Primary sequences of the major genetic variants of κ-casein. Only the C-terminal region is given, since modifications are found only in this section of the molecule.

C-terminal peptide, making it a less efficient steric stabilizer, leading to a shorter coagulation time and a faster rate of gel firming. The firmer gels are more difficult to explain because the modified hydrophilic peptide is lost to the serum phase and plays no further role in curd network formation and growth. Even more remarkable is that the increase in cheese yield is due to a greater efficiency in fat retention in these curds. The explanation is that κ-BB milk also has a higher content of κ-casein. This means they have smaller micelles on average, and the increased gel firmness results from this difference, in line with the predictions of the particle gel model, and not from the genetic modifications themselves.

Coagulants and their Activities in Curd Manufacture

Proteases capable of initiating the required proteolysis of κ-casein, which is the first step in gel formation, are aspartic proteinases (EC 3.4.23) and originate from a number of sources. Milk-clotting enzymes were historically derived by extraction from the stomach of ruminants, and calf rennet is used widely in commercial cheese manufacture today. Rennets can be extracted from other species such as sheep, goat, and pig, and generally, the clotting activity of these rennets is best suited to milk of the species from which the rennet is derived. The active clotting enzyme in these rennets is chymosin, and the proportion of chymosin in rennet extracts varies with the age of the animal at slaughter. Calf rennet extract contains approximately 90% chymosin with a small amount of pepsin, whereas adult bovine rennet may contain only 10% chymosin and 90% pepsin. Alternative sources of rennet have been sought to sustain the growth in commercial cheese manufacture worldwide. This has led to the development of milk-clotting agents from microbial sources. While the term rennet is applied to clotting extracts derived from ruminants, clotting enzymes from nonruminant sources are termed coagulants.

Commercially produced microbial coagulants for cheese manufacture are produced by fungal fermentation. The most widely used microbial coagulants on the market today are produced from *Rhizomucor miehei*, but *Rhizomucor pusillus* and *Cryphonectria parasitica* are also used.

Since 1991, pure chymosin coagulants produced by fermentation of a genetically modified organism have been available for commercial use. These are termed fermentation-produced chymosins and are produced from either of two genetically modified host organisms, the yeast, *Kluyveromyces lactis* or the fungus *Aspergillus niger*. Following fermentation, the chymosin extract is treated with acid which destroys any residual DNA or RNA from the host organism.

Gel Formation with Rennets and Coagulants

The formation of a rennet coagulum in curd manufacture takes place in two stages as described above. The primary phase is enzyme-mediated and involves the hydrolysis of micellar κ-casein at the Phe_{105}—Met_{106}. This bond is also hydrolyzed by the aspartic proteases of the microbial coagulants produced by *Rhizomucor miehei* and *Rhizomucor pusillus*. However, with *Cryphonectria parasitica*, the Leu_{104}—Phe_{105} is hydrolyzed preferentially.

Approximately 6% of the chymosin added during cheese manufacture is retained in curd. For the microbial coagulants, retention levels are approximately 2–3%. The level of retention of chymosin is influenced by the pH of the curd at cut, and the greater the acidity at cut, the greater the retention of chymosin in curd. Retention of microbial coagulants is not pH-dependent. The temperature profile during cheese manufacture also affects chymosin activity in curd during ripening. The high scald temperature used in the manufacture of Swiss-type cheeses inactivates residual chymosin activity in curd.

The rennet used in cheese manufacture should have a high ratio of milk clotting to nonspecific proteolytic activity to maximize cheese yield. While hydrolysis of the Phe_{105}—Met_{106} is essential for clotting, other nonspecific proteolysis processes in the cheese vat would result in reduced cheese yield owing to a weak coagulum and high losses of fat and protein in the whey. Fermentation-produced chymosin and calf rennet have the lowest level of nonspecific protease activity, whereas increasing levels of nonspecific protease are seen in bovine rennet, *Rhizomucor miehei*, *Rhizomucor pusillus*, and *C. parasitica* coagulants.

Differences in the specificity of proteases in coagulants and their overall proteolytic activity may impact on cheese flavor development, and in producing particular types of cheese, there may have to be a compromise between yield and flavor potential of the coagulant. Residual chymosin activity in Cheddar has been researched extensively and characterized. Chymosin retained in curd rapidly hydrolyzes α_{s1}-casein at the Phe_{23}—Phe_{24} bond producing α_{s1}-casein (f1–23) and α_{s1}-casein (f24–199). This results in softening of the cheese texture in the early stages of ripening. In mature Cheddar (>6 months), this hydrolysis is complete. During the early stages of ripening, the concentration of α_{s1}-casein (f24–199) increases initially, but as ripening progresses, it is

further hydrolyzed by chymosin, predominantly at the Leu_{101}—Lys_{102} and, to a lesser extent, at the Phe_{32}—Gly_{33} and Leu_{109}—Glu_{110}. Peptides resulting from these activities including the α_{s1}-casein (f24–199), α_{s1}-casein (f33–199), α_{s1}-casein (f102–199), and α_{s1}-casein (110–199) have been identified in Cheddar. Further hydrolysis of the α_{s1}-casein peptides is achieved by action of the indigenous milk protease, plasmin, and the lactococcal cell envelope-associated proteinase from the starter culture, and these activities result in the formation of small peptides and amino acids, which may be further catabolized to generate flavor components. The specificity of rennet substitutes on the caseins during maturation has not been characterized but is known to differ from that resulting from chymosin action.

Factors Influencing Curd Formation with Rennet

Factors influencing curd formation include rennet concentration, pH, temperature, and ionic calcium. Each has the potential to influence either the primary enzymatic phase or the secondary phase of aggregation, but some influence both.

Increasing rennet concentration results in an increase in the rate of the primary enzymatic phase. The secondary aggregation phase of renneting also proceeds more rapidly at higher rennet concentrations. Generally, the quantity of rennet is selected to ensure adequate gel formation within 20–40 min during commercial cheese manufacture.

Rennet coagulation of milk is pH-dependent, with optimum activity for hydrolysis of the Phe_{105}—Met_{106} in κ-casein at pH 6.0. Decreasing the natural pH of milk from 6.7 to 6.0 results in solubilization of calcium phosphate and a decrease in the charge on the casein micelles, which also encourages gel formation. The net effect of a decrease in pH is therefore an increase in both the primary enzymatic and secondary aggregation phase of coagulation.

The higher the temperature at renneting, the greater is the rate of aggregation. At a temperature of less than 15 °C, the primary enzymatic phase proceeds normally, whereas the aggregation phase does not proceed at a low temperature. Also, research studies indicate that gel firmness falls off above 40 °C.

Natural seasonal deficiencies in the calcium levels in milk can be overcome by the addition of calcium chloride, which causes a slight decrease in pH, which will promote the primary enzymic phase of renneting, while the secondary stage of aggregation is enhanced by calcium addition. Processing treatments applied to milk, e.g., excessive cold storage or high heat treatment of milk, can inhibit gel formation. These effects can be overcome by the addition of calcium chloride or reduction of the pH of milk.

Although a great deal of research effort has been directed into methods for the determination of optimum gel strength during cheese manufacture, in normal circumstances, there is a window extending to 10 min during which time the coagulum can be cut without any detrimental effects on cheese yield. However, the manufacture of cheese from abnormally weak gels, produced for example from milk deficient in calcium or with a high proportion of the κ-AA casein variant, results in reduced yield. The appropriate time for cutting the curd is dependent on the manufacturing plant, and the time of cutting must be adjusted to allow for seasonal variations in milk composition.

See also: **Calcium**: Physiology; **Casein and Caseinates**: Methods of Manufacture; Uses in the Food Industry; **Cheeses**: Starter Cultures Employed in Cheese-making; **Enzymes**: Uses in Food Processing; **Milk**: Analysis; **Cheeses**: Dutch-type Cheeses

Further Reading

Dalgleish DG (1992) The enzymatic coagulation of milk. In: Fox PF (ed.) *Advanced Dairy Chemistry 1. Proteins*, 2nd edn. pp. 579–619. London: Elsevier Applied Science.

DeKruif CG (1999) Casein micelle interactions. *International Dairy Journal* 9: 183–188.

Farrell HM Jr. (1988) Physical equilibria: proteins. In: Wong N (ed.) *Fundamentals of Dairy Chemistry*, pp. 461–510. New York: Van Nostrand Reinbold.

Fox PF and McSweeney P (1997) Rennets: their role in milk coagulation and cheese ripening. In: Law BA (ed.) *Microbiology and Biochemistry of Cheese and Fermented Milk*, pp. 1–40. London: Blackie Academic & Professional.

Harboe M and Budtz P (1999) The production, action and application of rennets and coagulants. The formation of cheese curd. In: Law BA (ed.) *Technology of Cheesemaking*, pp. 66–92. Sheffield, UK: Sheffield Academic Press.

Holt C (1992) Structure and stability of the bovine casein micelle. In: Anfinsen CB, Edsall JD, Richards FR and Eisenberg DS (eds) *Advances in Protein Chemistry*, Vol. 43, pp. 63–151. San Diego, CA: Academic Press.

Holt C and Horne DS (1996) The hairy casein micelle: evolution of the concept and its implications for dairy technology. *Netherlands Milk and Dairy Journal* 50: 85–111.

Horne DS (1992) Ethanol stability. In: Fox PF (ed.) *Advanced Dairy Chemistry 1. Proteins*, 2nd edn, pp. 657–689. London: Elsevier Applied Science.

Horne DS (1998) Casein interactions: casting light on the black boxes, the structure of dairy products. *International Dairy Journal* 8: 171–177.

Lomholt SB and Qvist KB (1999) The formation of cheese curd. In: Law BA (ed.) *Technology of Cheesemaking*, pp. 33–65. Sheffield, UK: Sheffield Academic Press.

Sousa MJ, Ardo Y and McSweeney PLH (2001) Advances in the study of proteolysis during cheese ripening. *International Dairy Journal* 11: 327–345.

Chemistry and Microbiology of Maturation

N Y Farkye, California Polytechnic State University, San Luis Obispo, CA, USA

Cheese Maturation (Ripening)

Most rennet-coagulated cheeses are ripened for at least 4 weeks before consumption. When manufactured from raw (unpasteurized milk), the cheese must be ripened for at least 60 days at not less than 1.7° C before consumption. Cheese ripening involves several complex and dynamic biochemical processes that result in textural and flavor changes, unique for each variety. Notably, a 'green,' rubbery, tough cheese with a bland taste is converted into a 'mature,' firm, elastic or soft-body cheese with a characteristic flavor.

The major biochemical processes that occur during ripening include metabolism of lactose/lactate, proteolysis, and lipolysis. The biochemical changes are results of the activities of residual rennet, enzymes from starter, starter adjuncts and nonstarter microorganisms, indigenous milk enzymes, and exogenous enzymes added to the milk or cheese for specific functions. The composition of cheese curd (especially moisture, salt, and pH) and ripening conditions (temperature and humidity) affect enzyme activity and, consequently, the rate and extent of ripening.

Rennets

The term 'rennet' was originally used to describe the milk-clotting enzyme preparation from calf stomach, which contains the active digestive enzyme called chymosin (rennin). At present, the term 'rennet' is used broadly to describe milk-clotting enzymes. Commercial milk-clotting enzymes currently available are listed in **Table 1**.

Chicken pepsin and papain are used in some countries for cheese-making. Also, rennets from kid and lamb stomachs are available. Rennet action during cheese-making results in the conversion of milk to cheese curd. The principal reaction that causes coagulation of milk is the cleavage of Phe_{105}–Met_{106} bond in the milk protein, κ-casein. A portion (0–30%) of the rennet used in cheese-making is retained in the cheese, depending on the type. The amount of rennet retained depends on the pH of the milk at setting, the type and amount used, and its ability to survive cooking temperatures used in cheese-making. The retention of chymosin, bovine pepsin, or porcine pepsin in cheese increases as the pH of milk is reduced at the time of rennet addition (setting). However, the pH of milk at setting does not affect the amount of *Mucor miehei*, *Mucor pusillus*, or *Cryophonectria parasitica* proteinases retained in the cheese curd.

Table 1 Commercial milk-clotting enzymes and their sources

Commercial name	*Systematic name*	*Source*
Calf rennet	Chymosin	Calf stomach
Bovine rennet	Pepsin A	Adult bovine stomach
Porcine pepsin	Pepsin A	Pig stomach
Microbial rennet	Mucorpepsin	*Mucor pusillus Lindt*
Microbial rennet	Mucorpepsin	*Mucor miehei*
Microbial rennet	Enthothiapepsin	*Cryophonectria* (*Endothia*) *parasitica*
Fermentation-produced rennet	Chymosin	*Escherichia coli* K12 *Kluyveromyces lactis* *Aspergillus niger*

Starter and Nonstarter Microorganisms

The term, 'starter' refers to a culture of lactic acid bacteria used for acid production, through the fermentation of lactose, during cheese-making. **Table 2** contains a list of starter and nonstarter microorganisms important in the ripening of major cheeses.

Mesophilic starters (*Lactococcus lactis* ssp. *cremoris* (formerly *Streptococcus cremoris*) or *Lc. lactis* ssp. *lactis* (formerly *S. lactis*)) are important in the ripening of Cheddar and Dutch-type (Gouda, Edam) cheeses. Thermophilic starters (*Streptococcus thermophilus* (formerly *S. salivarius* ssp. *thermophilus*) or *Lactobacillus delbruekii* ssp. *bulgaricus* (formerly *Lb. bulgaricus*)) are important in the ripening of Emmental-type (Swiss, Gruyère) and Italian-type (Romano, Parmesan, and Provolone) cheeses.

Other lactic acid bacteria (*Leuconostoc* spp., citrate$^+$ *Lc. lactis* ssp. *lactis* (formerly *Lc. lactis* ssp. *lactis* biovar. *diacetylactis* or *S. diacetylactis*) or *Lb. helveticus*) play a role in the ripening of Dutch- and Emmental-type cheeses. Microorganisms, other than lactic acid bacteria, significant to cheese ripening include *Propionibacterium shermanii* for Emmental-type cheese, molds (*Penicillium roqueforti*, *P. glaucum*) for blue-veined varieties, *P. camemberti* for

Table 2 Microorganisms of importance in cheese ripening

Cheese category and examples	*Moisture content (% max)*	*Starter*	*Secondary starter or other microorganism*
Hard grating			
Parmesan	32	*Lactobacillus delbrueckii* ssp. *bulgaricus*	*Pediococcus* sp.
Asiago	32	*Streptococcus thermophilus*	*Micrococcus* sp.
Romano	34		*Propionibacterium* sp.
Hard	39	*Lc. lactis* ssp. *cremoris*	*Lb. casei*, *Lb. brevis*, *Lb. plantarum*, *Pediococcus* sp. *Micrococcus* sp.
Cheddar		*Lc. lactis* ssp. *lactis*	
Gruyère	39	*S. thermophilus*	
		Lb. delbruekii ssp. *bulgaricus*	
		Lb. helveticus	
Semisoft	41	*Propionibacterium shermanii*	
Emmental (Swiss)			
	45	*Lc. lactis* ssp. *cremoris*	
Gouda	45	*Lc. lactis* ssp. *lactis*	
Edam		Citrate$^+$ *Lc. lactis* ssp. *lactis*	
		Leuconostoc sp.	
Surface-ripened by bacteria and yeast			
Limburger	50	*Lc. lactis* ssp. *cremoris*	Yeast
Brick		*Lc. lactis* ssp. *lactis*	*Brevibacterium linens*
Port Salut			*Artherobacter* spp.
			Corynebacterium spp.
Internally ripened by mold growth			
Blue	46	*Lc. lactis* subsp. *cremoris*	*Penicillium roqueforti*
Roquefort	45	*Lc. lactis* ssp. *lactis*	Yeast, micrococci
Gorgonzola	42	Citrate$^+$ *Lc. lactis* ssp. *lactis*	
Stilton	42	*Leuconostoc* sp.	
Soft ripened	50–80		
Brie		*Lc. lactis* ssp. *cremoris*	*P. camemberti*, *P. caseicolum*, *P. candidum*, *B. linens*
Camembert		*Lc. lactis* ssp. *lactis*	

Camembert and Brie, and *Brevibacterium linens* for cheeses such as Limburger, Brick, and Port Salut.

Probably all cheese varieties contain nonstarter lactic acid bacteria (NSLAB), such as lactobacilli (*Lb. casei*, *Lb. plantarum*, *Lb. brevis*), pediococci (*P. pentosaceus*), and micrococci. Some of the biochemical processes that occur during cheese ripening are due to the activities of NSLAB.

Microbiological Changes

Starter bacteria multiply from about 10^6–10^7 colony-forming units (cfu) per milliliter milk to 10^9 cfu per gram of fresh cheese curd. There is a decline in starter population as ripening progresses. The rate of decline varies among cheeses and is dependent on the elimination of lactose (the primary source of energy), inhibition by salt, and/or autolysis. Starter activity is inhibited when the salt-in-the-moisture (S/M) of cheese exceeds 5%.

In Cheddar and Gouda cheese, the starter population declines to $< 10^3$ cfu g^{-1} within the first few weeks, whereas in varieties like Provolone and Parmesan, cell densities remain high ($> 10^4$ cfu g^{-1}), even after 12 months of ripening.

In Emmental cheese, the maximum cell density of the lactic streptococci (10^9 cfu g^{-1}) at the periphery and 10^8 cfu g^{-1} in the center) is reached after about 3 h of pressing. The lactic acid produced by the streptococci stimulates the growth of lactobacilli, which reach cell densities of about 10^9 cfu g^{-1} after 10–20 h of pressing. The populations of both streptococci and lactobacilli decline after pressing and brining. When cheese is placed in a warm room (20–25 °C and 80–85% relative humidity, RH), the propionibacteria multiply rapidly, reaching cell densities of about 10^9 cfu g^{-1} in the center of the cheese in 3 weeks.

The growth of starter bacteria is inhibited in blue-veined cheeses because of the high salt content (10% S/M). However, growth of *P. roqueforti* is induced after the cheese is pierced with needles to allow air into its interior, causing blue veining. Ripening of blue-veined cheeses occurs at 10 °C and 96% RH, and growth of *P. roqueforti* reaches a maximum in 90 days.

In many varieties, growth during NSLAB occurs of ripening. The cell densities of nonstarter lactobacilli reach 10^7 cfu g^{-1} in 10 weeks. Also, there is growth of *Pediococcus*, sp. and cell densities of *Micrococcus* sp. reach about 100 cfu g^{-1} during the same period.

Biochemical Changes

Metabolism of Lactose, Lactate, and Citrate

Lactic acid bacteria use either the phosphoenolpyruvate-dependent phosphotransferase or the lactose permease system to transport lactose into the cell. In lactococci, lactose is phosphorylated to lactose-P and transported into the cell, where it is hydrolyzed by phospho-β-galactosidase to glucose and galactose-6-phosphate. The glucose is fermented via the Embden–Meyerhof–Parnas (glycolytic) pathway to produce lactic acid. The galactose-6-phosphate is fermented via the tagatose phosphate pathway to lactic acid. In *S. thermophilus* and *Lb. bulgaricus*, lactose is not phosphorylated before it is transported into the cell. In the cell, it is hydrolyzed by β-galactosidase (or lactase) to glucose and galactose. The glucose moiety is converted to glucose-6-phosphate and fermented via the glycolytic pathway. The galactose moiety is converted to glucose-6-phosphate via the Leloir pathway.

Citrate is metabolized by citrate$^+$ *L. lactis* ssp. *lactis* and *Leuconostoc* species. Both organisms have specific pH-dependent permeases to transport citrate across the cell membrane into the cell, where it is converted into acetate, pyruvate, and CO_2. Pyruvate is converted into other products such as diacetyl, which is an important aroma compound in dairy products.

The concentration of lactose in fresh cheese (1 day old) ranges from < 0.1% in Dutch- and Emmental-type cheeses to about 1% in Cheddar. At S/M levels < 5%, starter and NSLAB metabolize residual lactose to lactate, which may be produced in the L(+) or D(−) isomeric form, depending on the organism (Table 3).

Lactic acid (1.5%, in Cheddar) reduces the initial pH of cheese to < 5.3, and is a source of energy for some microorganisms, thereby serving as a precursor of flavor compounds. L(+)-Lactate is metabolized by propionibacteria, in the pH range 5.0–5.3, to propionic acid, acetic acid, and CO_2 (eqn. (1)) in Emmental-type cheese. Some of the CO_2 produced accumulates in the cheese and forms holes, which are commonly called 'eyes.'

$$3CH_3\text{-}CHOH\text{-}COOH \longrightarrow 2CH_3\text{-}CH_2\text{-}COOH + CH_3\text{-}COOH + CO_2 + H_2O. \quad (1)$$

Lactic acid — Propionic acid — Acetic acid

In Cheddar cheese, NSLAB are capable of metabolizing lactose to produce ethanol and formic acid, both of which are undesirable in large quantities. The NSLAB may also oxidize lactate to acetate and CO_2. Pediococci and lactobacilli convert L(+)-lactate to D(−)-lactate, resulting in a racemic mixture of both isomers after 6 months of ripening. D(−)-Lactate forms an insoluble calcium salt that may crystallize in cheese and appear as undesirable white specks on cut surfaces. The ingestion of high levels of D(−)-lactate by humans causes metabolic disturbances. *Lb. casei* and *Lb. plantarum* oxidize citrate to acetate and CO_2, resulting in a gradual increase in acetic acid during Cheddar cheese ripening.

In Dutch-type cheeses, pyruvate, produced from citrate metabolism, is oxidized to diacetyl and CO_2 (eqn. (2)) by citrate$^+$ *L. lactis* ssp. *lactis* and *Leuconostoc* species.

$$2CH_3\text{-}CO\text{-}COOH + O_2 \longrightarrow CH_3\text{-}CO\text{-}CO\text{-}CH_3 + 2CO_2 + H_2O. \quad (2)$$

Pyruvic acid — Diacetyl

The catabolism of lactate and the formation of alkaline N-compounds (due to proteolysis) result in an overall increase in the pH of some varieties during ripening. After 6 months of ripening, the pH of Emmental-type cheese increases from ~5.3 to ~5.9, and that of Camembert and Blue increases from ~4.8 to ~7. The pH of Cheddar cheese, however, increases only slightly (0.1 pH unit), because the concentration of lactic acid remains high (1.2–1.9%), even after 12 months of ripening.

Table 3 Isomer of lactic acid produced by lactic acid bacteria

Lactate isomer	*Organism*
L(+)	*Lc. lactis* ssp. *cremoris*
	Lc. lactis ssp. *lactis*
	Citrate$^+$ *Lc. lactis* ssp. *Lactis*
	Leuconostoc spp.
	S. thermophilus
D(−)	*Lb. delbruekii* ssp. *bulgaricus*
	Lb. lactis
DL	*Lb. helveticus*
	Lb. plantarum
	Lb. brevis
	Lb. casei
	P. pentosaceus

Proteolysis

Of the major milk proteins, α_{s1}-, α_{s2}-, and β-caseins predominate in cheese. Proteolysis involves the breakdown of these proteins and polypeptides by residual rennet, indigenous milk proteases, and/or proteinases/peptidases of starter and nonstarter microorganisms. Figure 1 shows the effects of different rennets on the release of water-soluble nitrogen in cheese. In general, the lowest level of proteolysis

occurs in cheese made with porcine pepsin, whereas the most extensive proteolysis occurs in cheese made with microbial rennets.

Chymosin hydrolyzes the Phe_{23}–Phe_{24} (and Phe_{24}–Val_{25}) bond of α_{s1}-casein to produce α_{s1}–I [α_{s1}–CN (f24/25–199)] peptide. The hydrolysis of this bond is probably the most important reaction responsible for the initial softening of cheese. Subsequent degradation of the α_{s1}–I casein by rennet occurs during ripening. The peptide, α_{s1}–CN(f1–24/25), is rapidly hydrolyzed by starter cell wall-associated proteinases.

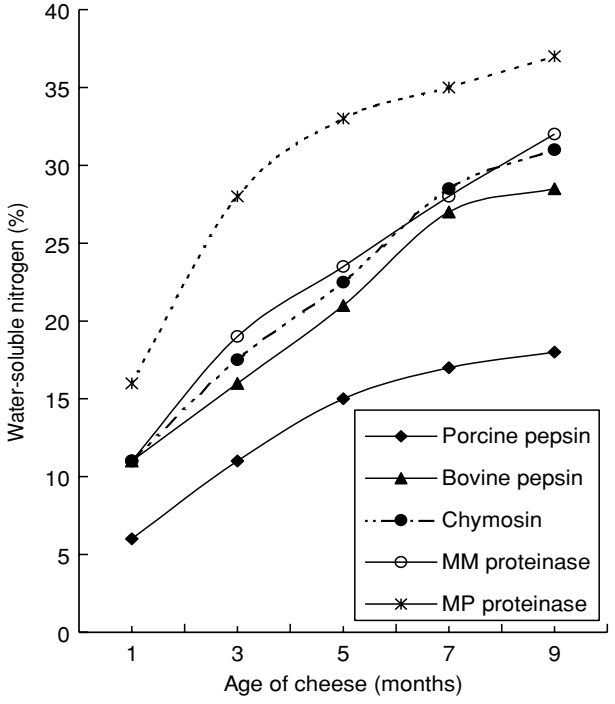

Figure 1 Water-soluble nitrogen (percentage of total nitrogen) during ripening of Cheddar cheese made with identical milk-clotting activities of different enzymes. MM, *Mucor miehei*; MP, *Mucor pusillus*.

Proteolysis of β-casein is less extensive than that of α_{s1}-casein in cheeses made with chymosin, bovine pepsin, or porcine pepsin, but more extensive breakdown of β-casein occurs in cheeses made with proteases from *M. miehei*, *M. pusillus*, and *C. parasitica*.

Plasmin, an indigenous milk proteinase, hydrolyzes all the caseins except κ-casein. Specifically, plasmin hydrolyzes β-casein to γ-caseins [β-CN(f29–209, 106–209, and 108–209)] and proteose peptones. The activity of plasmin is high in cheeses like Romano and Emmental, for which high cooking temperatures are used during manufacture, and in which the coagulant is denatured.

Starter and nonstarter bacteria have very limited proteolytic abilities towards whole caseins in cheese, although proteinase-positive strains of lactococci have cell envelop associated serine proteinases that hydrolyze caseins. Starter lactococci also contain a variety of peptidases (Table 4) that hydrolyze oligopeptides and di- and tripeptides resulting from the action of rennet and/or indigenous milk proteinases on caseins to amino acids essential for the development of flavor. The most common free amino acids in cheese include glutamic acid, methionine, asparagine, histidine, alanine, valine, phenylalanine, leucine, and lysine. In Emmental-type cheese, proline is also present.

Proteolytic enzymes from *Penicillium* species and *Brevibacterium* species are major contributors to proteolysis in mold-ripened cheeses. Depending on the

Table 4 Peptidases in Lactococci

Enzyme	*Other name/abbreviation*[a]	*Specificity*
Endopeptidases	PepO	Hydrolyzes glucagon
	LEPI	Hydrolyzes α_{s1}-CN(f1–23)
General aminopeptidases	AMP	Broad specificity. Hydrolyze several di-, tri-, and tetrapeptides
Aminopeptidase A	PepA	Hydrolyzes N-terminal Glu- and Asp residues
Glutamyl aminopeptidase A	GAP	N-terminal Glu- and Asp-containing peptides
Aminopeptidase C	PepC	Active on Lys, Phe, His, Glu, and Leu
Aminopeptidase N	PepN	Substrates with N-terminal Lys, Arg and Leu
X-prolyl-dipeptidyl aminopeptidase	PepX (PepXP)	Hydrolyzes with N-terminal X–Pro in X–Pro–Y-containing peptides
Pyrrolidonyl carboxyl peptidase	PCP	Removes N-terminal pyroglutamate residues from peptides
Aminopeptidase	PepP	Hydrolyzes N-terminal residues from X–Pro–Pro–Y
General aminopeptidase	PepT	Hydrolyzes Pro–Gly–Gly
Dipeptidase	DIP	Hydrolyzes dipeptides
Prolidase	PRD	Hydrolyzes dipeptides containing C-terminal proline (X–Pro)
Tripeptidases	TRP	Cleaves N-terminal amino acids. Do not hydrolyze tripeptides containing Pro as second amino acid
Proline iminopeptidase	PIP	Hydrolyzes di- and tripeptides containing N-terminal Pro residues

[a]The abbreviation 'Pep' is used for peptidases that have been genetically characterized. Others denote peptidases that have not been genetically characterized.

strain of *P. roqueforti* used, water-soluble N in Blue cheese may be as high as 50% of the total N after 3 months of ripening.

A summary of the sequence of proteolytic events that occurs in cheese during ripening is shown below.

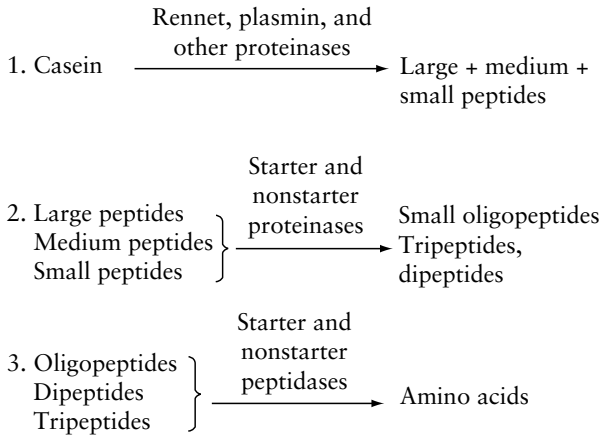

Breakdown of Amino Acids

Amino acids can be converted into various products during the ripening of cheese. Amino acids can be decarboxylated to CO_2 and an amine; the resulting amine undergoes oxidative deamination to give NH_3 and an aldehyde from which a primary alcohol or acid is formed. An aldehyde can also be formed by oxidative deamination of an amino acid. For example, deamination of glycine and alanine produces methanol and ethanol, respectively; decarboxylation of tyrosine and histidine results in tyramine and histamine, respectively. Amino acids can be converted to other amino acids by transamination. Sulfur-containing amino acids such as methionine can be converted to products such as methanethiol (CH_3SH), dimethysulfide, etc. Methanethiol and H_2S contribute to Cheddar cheese flavor. Trytophan can be degraded to indole, pyruvic acid, and NH_3. Tyrosine can be degraded to phenol.

Lipolysis

The action of lipase on triglycerides produces fatty acid(s), monoglycerides, and diglycerides. Lipolysis in cheese is due to the activities of milk lipoprotein lipase, exogenous lipases (from rennet pastes), or lipases from microflora in cheese milk. In pasteurized milk cheeses, the activity of milk lipoprotein lipase is small, as under normal pasteurization conditions, 10–15% of the activity in raw milk is left. Some starter and nonstarter bacteria contain lipases that hydrolyze mono- and diglycerides but have a weak lipolytic activity towards unhydrolyzed milk fat (triglycerides). Therefore, the concentration of free (C_4 to C_{18}) fatty acids (FFA) in bacteria-ripened cheeses like Cheddar, Emmental-type, and Dutch-type cheeses is low (0.1% in 3-month-old Cheddar, increasing to < 0.2% after 12 months of ripening). Cheddar cheese containing > 0.2% FFA usually has a rancid flavor.

In some Italian cheese varieties (e.g., Romano, Parmesan and Provolone) lipolysis is more than threefold greater than that in Cheddar as exogenous lipases are added during cheese making. The level of lipolysis in blue-veined cheeses (e.g., Blue, Roquefort, Gorgonzola, and Stilton) is high and is due to the activity of one (the acid enzyme) of the two (acid or neutral) types of lipases produced by *P. roqueforti*.

The sequential process of catabolism of fat during cheese ripening is as follows:

1. Liberation of FFA, mono- and diglycerides from milk fat by lipase.
2. Oxidation of FFA to β-keto acids (β-oxidation).
3. Decarboxylation of β-keto acids to produce methyl ketones (e.g., acetone from butyric acid).
4. Reduction of methyl ketones to secondary alcohols (e.g., reduction of acetone to 2-propanol).

Other lipolytic reactions include formations of esters by esterases (e.g., esters of alcohol, methional, and phenol), formation of lactones from unsaturated fatty acids, and the formation of oct-1-en-3-ol from esters of linoleic acid or arachidonic acid. The short-chain (C_4–C_{10}) fatty acids, methyl ketones (mostly 2-heptanone, followed by 2-nonanone and 2-pentanone), and secondary alcohols contribute to the characteristic flavor and aroma of cheese.

See also: **Amino Acids**: Metabolism; **Cheeses**: Starter Cultures Employed in Cheese-making; Dutch-type Cheeses; **Lactic Acid Bacteria**

Further Reading

Andrews AT and Varley J (eds) (1994) *Biochemistry of Milk Products*. Cambridge: Royal Society of Chemistry.

Farkye NY and Fox PF (1990) Objective indices of cheese ripening. *Trends in Food Science and Technology* 1: 37–40.

Fox PF (ed.) (1993) *Cheese: Chemistry, Physics and Microbiology*, 2nd edn. London: Elsevier Applied Science.

Fox PF, Lucey JA and Cogan TM (1990) Glycolysis and related reactions during cheese manufacture and ripening. *Critical Reviews in Food Science and Nutrition* 29: 237–253.

Robinson RK (ed.) (1990) *Dairy Microbiology*, vol. 2, 2nd edn. London: Elsevier Applied Science.

Wong NP, Jenness R, Keeney M and Marth EH (eds) (1988) *Fundamentals of Dairy Chemistry*. New York: Van Nostrand Reinhold.

Manufacture of Extra-hard Cheeses

V Bottazzi, Universita Cattolica del Sacro Cuore, Piacenza, Italy

Classification, Definition, and Examples

The extra-hard (or very hard) cheeses are a group of varieties characterized by a long ripening time, and by the production in large pieces weighing from 18 kg to more than 100 kg per whole cheese; they are produced principally in Mediterranean Europe with sheeps' or cows,' milk. Grana cheese produced in the Po Valley can have the mark 'Parmigiano Reggiano' or 'Gran Padano,' depending on the areas of production. The technology of production is thousands of years old, and the hard cheeses are characterized by having a low moisture content. (*See* **Sheep**: Milk.)

Once they are ripe, their taste can be harsh and strong or delicate and fragrant, their structure is compact and friable, and they are used mainly for grating as a condiment. They play an important role in the Mediterranean diet.

The following cheeses are included in the group of extra-hard (or very hard) cheeses: Grana, Pecorino Romano, Provolone Piccante and Caciocavallo, Sapsagno, and Spalen.

Each grana (whole cheese) weighs 34–38 kg, and it is produced only in northern Italy (Po Valley) using partially skimmed cows' milk. It is the best known cheese among those used for grating as a condiment and, in order to maintain all its fragrance, it should be grated just before use. Annual production is 190 000 tonnes, which means using more than 35% of the Italian whole-milk production.

Pecorino Romano cheese is one of the most ancient cheeses: a description of the cheese-making process dating back to before Christ is available. It is produced in Sardinia and in the area around Rome, using only sheeps' milk coagulated by means of rennet paste from lambs; whole cheeses range between 16 and 22 kg in weight, and its taste is typically strong.

Provolone Piccante cheese and the Caciocavallo cheese are typical stirred-curd cheeses, originating in, and very widespread in, the Mediterranean area. The large Provolone Piccante cheese is cylindrical and can weigh up to 100 kg: it is produced with raw milk and rennet paste from kids, and once ripe, its taste is sharp, strong, and very pleasant.

As far as the description of the manufacturing technology, both traditional and modern, of extra-hard cheeses is concerned, we are going to describe only that of the grana cheese, as it is the most well known.

Traditional Methods of Manufacture

The traditional method is not very different from methods used today, for it is a feature of hard cheeses to apply technological rules that do not vary very much with the passing of time.

Up to about 100 years ago, grana cheese was very often manufactured in cheese factories designed following particular rules aimed at obtaining well-defined goals. Such a factory was a single, isolated building with an octagonal base, and walls free from windows. The height of the building was about 3 m, and the roof was supported by a wooden structure. Architecturally, such buildings are pleasant in appearance.

Inside, all the cheese-making operations were carried out using direct-fire heating of the milk, and the typical structure allowed natural aeration.

According to tradition, raw milk obtained the evening before manufacture was poured into shallow vats with a large surface area, so as to facilitate separation of the cream; this process was necessary since grana cheese is produced with partially skimmed milk. The vats were made of either wood or tin-plated copper of sufficient volume to hold half of the milk necessary to produce a whole grana cheese, i.e., 250 l. The creaming took place overnight, and in the morning, after the skimming operation, the milk was mixed with fresh milk, and cheese manufacture in a large copper vat commenced.

Until the end of the nineteenth century, no starter cultures were added, and in order to reach the desired acidity, the contaminating microflora acquired during milking and in the cheesemaking operation was exploited; the vat was then left for a prolonged period, which allowed a remarkable bacterial flora to develop. (*See* **Starter Cultures**.)

Rennet prepared from the stomach of sucking calves was always used to coagulate the milk. After breaking the coagulum, very small curd granules, which had a very low moisture content, were obtained. The granules were smaller in dimension than rice grains, and the expulsion of whey was through the high 'cooking' temperature (53–54 °C).

Originally, the cheese vat had a capacity of 500 l (so as to produce one cheese at a time), had a truncated cone shape (inverted bell), was made of copper, and had direct-fire heating.

Salting, using brine, was started 2 or 3 days after production and continued for 25–28 days in separate brining rooms.

The ripened cheeses were carried out in stores at the prevailing temperature, i.e., under the influence of the seasons, temperatures varied from about 5 to 28 °C, for at least 18 months. During this long period, and especially during the first months, many manual operations were necessary in order to avoid the development of molds on the cheese surfaces, and to obtain a smooth, even rind. Towards the end of the ripening period, the cheese was covered with a mixture of carbon black, umber, and grape seed oil, so as to produce a black color and a characteristic shiny, bright appearance.

Modern Systems of Production

The current manufacturing technology of grana cheese shows, in comparison with the traditional technology, several differences.

The current procedure uses raw milk, natural creaming and a high cook temperature, but nowadays, the technology also includes the addition of a starter culture, 'double cheese-making,' a rapid cheese-making time, 'twins' production, controlled ripening as well as other changes, which will be described later.

The whey starter culture was introduced at the end of the nineteenth century, and has a very complex microflora composition. The prevailing bacteria are *Lactobacillus helveticus* together with *Lb. delbrueckii* ssp. *lactis* and ssp. *bulgaricus* and *Lb. fermentum*; thermophilic streptococci are generally absent. The culture is phage-resistant, has a high acidifying capacity, is thermophilic, has a very good viability, and is very easy to prepare. It is affected by the microbial characteristics of the milk and the environment where manufacture takes place, and it is the result of the action of several technological factors typical of each environment; because of its characteristics, it turns out to be an unrepeatable culture that cannot be replaced by the association of different strains in the laboratory. The activity of the natural whey culture is very important: primarily to produce an acid environment in the cheese (thereby inhibiting gas-producing bacteria), to take part in the hydrolysis of proteins, as well as to take part in the hydrolysis of proteins, as well as to take part in taste and fragrance formation. (*See* **Lactic Acid Bacteria**; **Whey and Whey Powders**: Production and Uses.)

The 'double cheese-making' process involves the transformation of milk from two milkings, and brings about the creaming of all the milk employed. From a practical point of view, this is a very important change, because by creaming at a temperature of 12–15 °C, not only is the milk skimmed, but also many bacteria are carried out with the cream; an effect similar to that obtained through pasteurization.

This cold reduction in bacterial numbers leaves the cheese-making properties of raw milk unchanged, a point of primary importance in grana cheese production.

This system is widespread in the whole production zone of grana Padano; the grana Parmigiano Reggiano cheese-makers keep to the traditional method of manufacture.

The 'rapid cheesemaking time' used today in grana cheese production refers to the time between the addition of rennet to the milk and the end of the curd cooking. At the end of the nineteenth century, the time was rather long – sometimes even 50–60 min – but today, it is just 18–22 min.

The reasons for these changes include both the different conditions of milk production, and the introduction of thermophilic cultures of lactic acid bacteria. From a practical point of view, it allows better work organization and more regular cheese-quality standards.

The 'twins' production started with the replacement of the 500-l vats with 1000-l vats; the shape of the copper vat is unchanged and is typical for this kind of cheese; the maximum capacity is 1000 l of milk. As a consequence, two whole cheeses can be obtained each time, and these are called twins, since they are the result of the division into two even parts of the curd obtained from 1000 l of milk.

The number of grana cheese factories is still more than 1100 today, but a progressive transformation towards medium-to-large factories is taking place. Several copper vats may be found in one factory, but because the manufacture involves separate batches of milk, each of 1000 l, there is neither real mechanization nor automation (not even cheese pressing is carried out mechanically). None the less, the whole cheese-making process in the vat is carried out in a short time, with a regular and precise sequence, even in the case of large factories. Today, the coagulum is broken first manually, but further breaking up and stirring of the curds/whey to complete cooking are carried out using simple mechanical means.

Ripening is carried out in stores with controlled temperature (16–18 °C), humidity (85%), and continuous air change without creating currents of air.

The cheese is ripened for 14–24 months (the grana Padano needs a shorter time than Parmigiano Reggiano), and during this period, the whole cheeses are often turned upside down once a week during the first months, and less frequently thereafter. Maintenance of the cheeses (inverting and brushing) is carried out mechanically by means of automatic systems.

The modern system for Grana cheese production provides for milk cooling at 8 °C for 12 h in sheds, then mixing with warm milk from the subsequent milking, thus allowing a single collection per day. If necessary, and in particular in areas where there is a high consumption of ensiled fodder, irregular fermentations can be avoided through the partial elimination of bacteria by centrifugation. The main steps in Grana cheese production can be described briefly as follows:

1. Milk is poured into 1000-l vats, with dimensions of $4.50 \times 1.90 \times 0.25$ m, for creaming, so that the milk reaches the copper, truncated cone-shaped cheese vats with a fat content ranging, depending on the region, from 2.00 to 2.30%.
2. After transfer of the partially skimmed milk to the 1000-l copper vats, about 3% of natural whey culture is added to give an acidity (in lactic acid) in the mixture (milk + whey culture) of 0.19–0.20%; the milk is heated to 33 °C and rennet (titer 1 : 100 000) at a level of 2 g per 100 l.
3. Coagulation occurs in 9–10 min, and the coagulum is broken for 3 min to obtain curd granules having the dimensions of rice grains.
4. The curds/whey are heated to reach the cook temperature of 54 °C in 7–8 min.
5. The curd is allowed to settle and form a compact layer at the bottom of the vat, and the curd is then kept in the hot whey for about 45 min.
6. The bed of curds is divided into two portions, each of which is scooped from the mold using a coarse cloth, and placed in the unfixed wooden molds; at the start of the pressing operation, the cheeses are pressed by loading each cheese with 20-kg weights.
7. During the first 10 h, the cheeses are turned three or four times, with cloth changes; after 15–20 h, the cloths are removed, and the wooden mold is replaced by a metal mold.
8. On the third day, the cheeses are brine-salted (density 22 °Bé and temperature 16–17 °C) for 24–28 days.
9. At the end of the salting and once the whole cheeses are dry, they are transferred to the ripening store and kept at a temperature of 16–18 °C and a humidity of 85%.

Maturation and Storage

Grana cheese, after ripening, has the following average composition (%):

- protein: 33.20
- fat: 27.50
- ash: 4.80
- moisture: 32.00
- sodium chloride: 1.60
- lactic acid: 1.30

Owing to the high content of protein and low fat content, grana cheese can be defined as a 'half-fat' cheese having a high protein content.

During the first hours of pressing, an intense fermentation process occurs in grana cheese, transforming the lactose into lactic acid. After 6 h, the pH is about 5.50 and after 16 h, 5.0; the lactic acid is distributed as indicated in Table 1.

After 5 h of fermentation, the lactose has decreased to about 1.6%, the accumulated galactose is slightly higher than 1%, and the concentration of glucose is about 0.30%.

During the long ripening period, even the casein undergoes extensive proteolysis. β-Casein is hydrolyzed rapidly during the first part of the ripening cycle, and its enzymatic degradation finishes within the first 12 months of ripening. However, α_{sl}-casein is degraded much more slowly. With reference to the total amount of hydrolysis products, about 30% are from β-casein, while those from α_{sl} casein form about 18%.

The release of amino acids increases progressively until about the 15th month of ripening and then becomes stable. When the cheese is ready to use, the amino acids form, on average, 22% of the total crude nitrogen value and 7% of the cheese; as a consequence, grana cheese is one of the cheeses having a high content of free amino acids. In cheeses of first-class quality, serine increases progressively throughout the ripening time and finally reaches about 1.5% of free amino acids; glutamine reaches 1% in the first 12 months and disappears once the ripening is finished, and arginine, released during the ripening period, is degraded into ornithine. The presence of γ-aminobutyric acid indicates that anomalous fermentation processes have taken place. (*See* **Amino Acids**: Properties and Occurrence.)

The hydrolysis of fat in grana cheese does not play an important role, in spite of the long ripening period. There is also a slight loss of vitamin A, a greater remarkable loss of vitamin E, and a serious reduction of β-carotene from 640 to 100 μg per 100 g of cheese. (*See* **Carotenoids**: Occurrence, Properties, and

Table 1 Isomeric distribution of lactic acid with time in a grana cheese (values are in g^{-1})

0 h		*2 h*		*4 h*	
D(−)	0.90	D(−)	2.85	D(−)	3.99
L(+)	1.80	L(+)	4.47	L(+)	6.35
DL	2.7	DL	7.32	DL	10.34

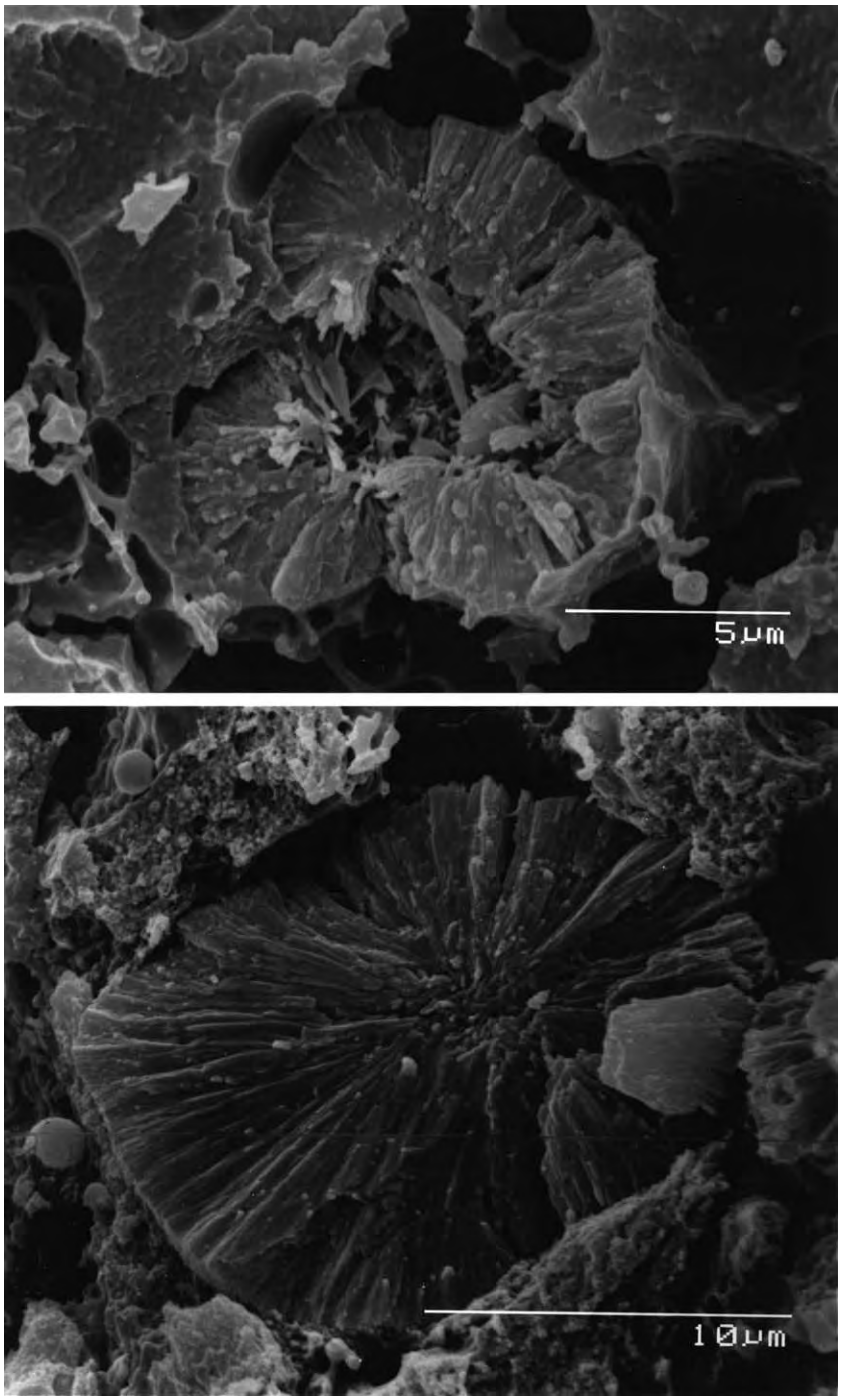

Figure 1 Calcium phosphate microcrystals in 24-h grana cheese. (A) Microcrystal internal structure during the formation phase. (B) Microcrystal internal structure at the end of the formation process.

Determination; **Retinol**: Properties and Determination; **Tocopherols**: Properties and Determination.)

The mineral component, composed of macro- and microelements, takes on a very precise character: calcium, phosphorus, and magnesium are the elements present at the highest levels: 1150, 680, and 430 mg per 100 g respectively. The calcium:phosphorus ratio is high (1.70), i.e., much higher than that found in many other cheeses. (*See* **Calcium**: Properties and Determination; **Magnesium**.)

The average content of sodium in grana cheese is 650 mg per 100 g. Among the microelements, the content of zinc is very high. (*See* **Sodium**: Properties and Determination; **Zinc**: Properties and Determination.)

The microbiological composition of grana cheese is characterized by two separate steps. The first concerns the very young cheese (first hours), and the second covers the first months, and continues even until the end of ripening.

The lactic acid bacteria are numerous during the first period, and the flora is that of the starter, mainly composed of *L. helveticus*, usually reaching, after 12–14 h, $0.8–1 \times 10^9$ cells per gram, and *L. fermentum*, which reaches a maximum in about 24–28 h with $120–140 \times 10^6$ cells per gram. In this first period, even *L. delbrueckii* ssp. *bulgaricus* and ssp. *lactis* are active. The development of this homo- and heterofermentative thermophilic lactic acid flora plays a very important role in the commercial success of the final product, since it:

- ensures a rapid acidification of the fresh cheese;
- regulates whey drainage;
- accumulates enzymes of the proteolytic system;
- causes the formation of small holes;
- produces volatile compounds;
- contributes to the formation of the cheese structure.

During the second period, a nonstarter lactic acid flora develops, consisting of coccoid lactic acid bacteria, e.g., *Pediococcus acidilactici* and rod-shaped bacteria, e.g., *L. casei* ssp. *casei*, ssp. *pseudoplantarum* and ssp. *rhamnosus*.

During the period of maximum development, i.e., 25–40 days after production, this lactic acid flora reaches $60–80 \times 10^6$ cells per gram, but, for the remaining ripening period, the number decreases progressively. Pediococci are the most resistant forms and are still present in the cheese at the time of consumption.

The metabolic activity of the mesophylic lactic acid bacteria contributes to the ripening process, and its contribution to casein hydrolysis is very important.

Finally, there are the propionic bacteria, which are useful for the development of the organoleptic characteristics of the cheese, so long as their development is limited to a few million cells per gram.

Formation of Micro- and Macrocrystals

The formation of individual or grouped crystals occurs during ripening of the cheese (Figures 1–3). In grana cheese, the first type of microcrystals appears at the end of the lactic fermentation driven by a thermophilic lactic microflora, during the first hours of ripening. The crystals are spherical. The external side is porous, showing an irregular surface, with a diameter of 8–20 μm. The internal portion, as shown in Figure 1A, is well defined, with a compact peripherical ring and laminar structure converging in the centre. Crystal formation is limited to the first 48 h

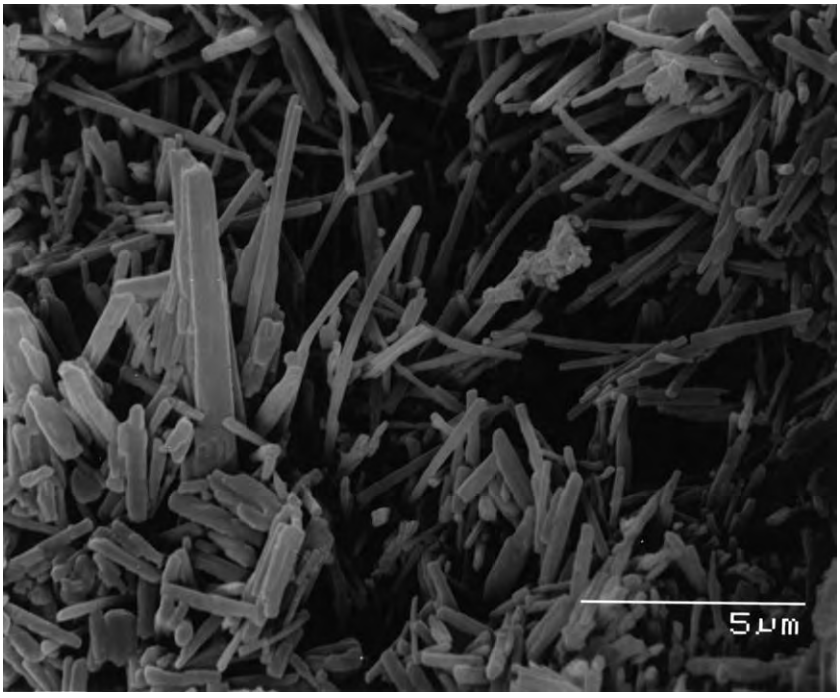

Figure 2 Needle-shaped calcium lactate crystal aggregates in ripened grana cheese.

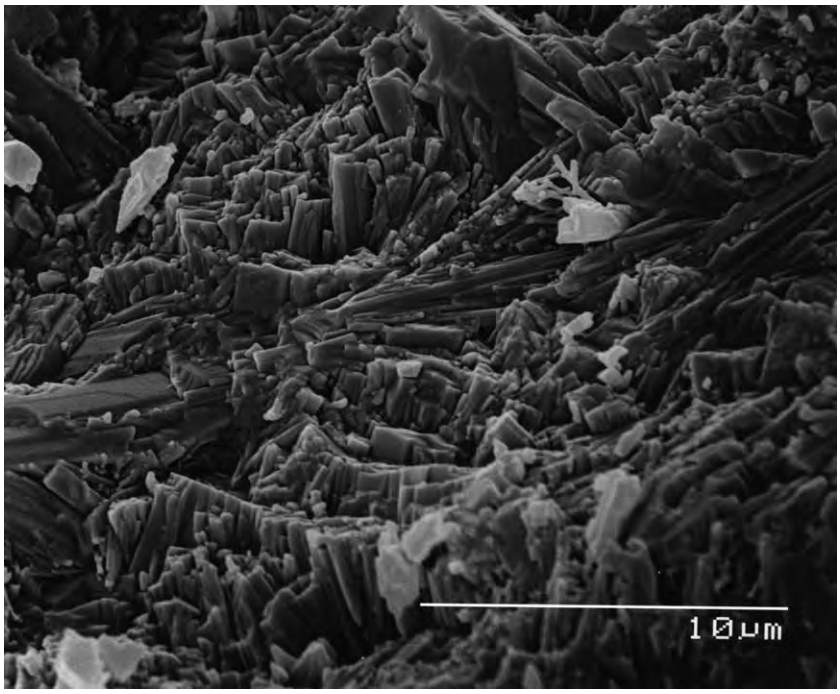

Figure 3 Internal structure of a tyrosine macrocrystal in ripened grana cheese.

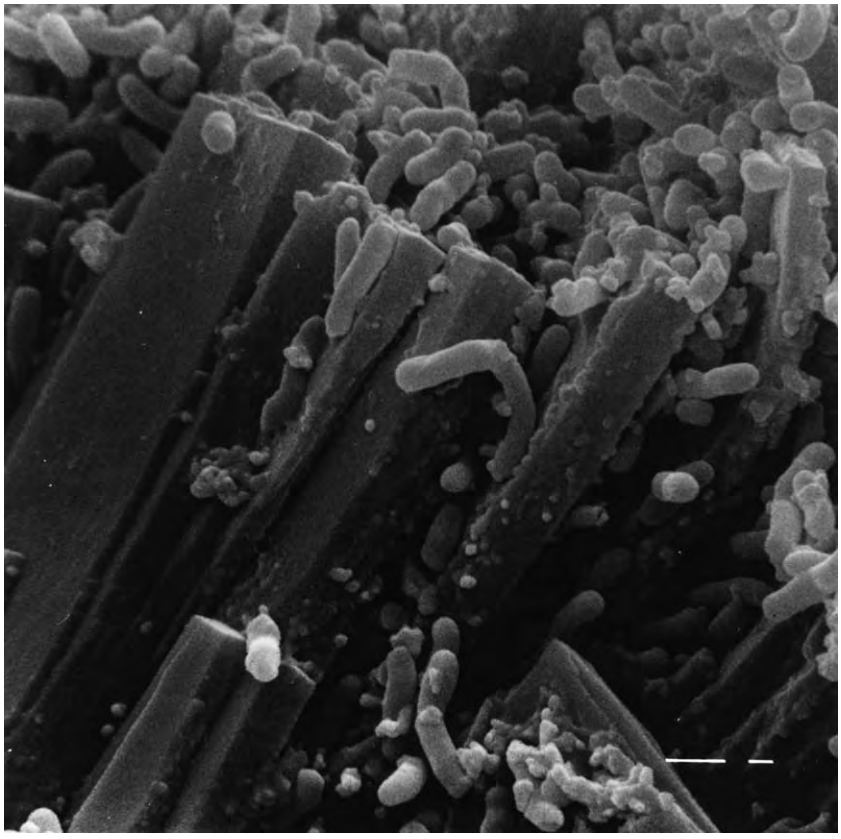

Figure 4 Sodium chloride microcrystals in ripened grana cheese.

in grana cheese. During further cheese ripening, no further increase in volume or number of crystals can be seen. These crystals are composed of calcium phosphate.

Figure 2 shows a second type of crystal: needle-shaped and generally grouped. At the end of ripening, these can be observed in the cheese paste as white, friable, sandy granules. These crystals are typical of grana Padano cheese and are generally associated with a high-quality product (if not too numerous) and are composed of calcium lactate.

At the end of ripening, a gray crystalline formation becomes evident (0.3–0.5 mm in diameter). These crystals are formed by tyrosine and impart a bitter taste. Their internal part is shown in Figure 3. Aggregates of NaCl rarely appear in cheese at the end of the ripening process. Figure 4 shows these aggregates observed under a scanning electronic microscope.

Specification and Standards

Grana cheese is a half-fat hard cheese, cooked and slowly ripened, with a light straw color yellow and a compact, granular and friable structure with flake-shaped radial fractures and a fragrant and delicate aroma. For its production, raw milk, half-skimmed through natural creaming, is used. To the milk is added a natural starter culture of thermophilic lactic acid bacteria developed in whey, and the mixture is coagulated by means of rennet from sucking calves. The ripening lasts 18–24 months for Parmigiano Reggiano and 13–15 months for grana Padano.

The ripe product, ready to eat, has:

- a cylindrical form with convex (lateral) sides, and slightly convex flat sides;
- dimensions varying between 35 and 45 cm in diameter, and between 18 and 24 cm in height on the lateral side;
- a whole-cheese weight between 32 and 34 kg;
- a hard, smooth, bright surface, with a natural uniform color.

See also: **Amino Acids**: Properties and Occurrence; **Calcium**: Properties and Determination; **Carotenoids**: Occurrence, Properties, and Determination; **Lactic Acid Bacteria**; **Magnesium**; **Retinol**: Properties and Determination; **Sheep**: Milk; **Sodium**: Properties and Determination; **Starter Cultures**; **Tocopherols**: Properties and Determination; **Whey and Whey Powders**: Production and Uses; **Zinc**: Properties and Determination

Further Reading

Bosi F, Bottazzi V, Vescovlo M *et al.* (1990) Batteri lattici per la produzione di formaggio grana. I parte: caratterizzazione tecnologica di ceppi di lattobacilli termofili. *Scienza e tecnica Lattiero Casearia* 41: 105–136.

Bottazzi V (1979) *Microbiologia del Fermenti Lattici*. Modena: Aitel.

Fortina MG, Parini C, Manachini PG *et al.* (1990) Genotypic and phenotypic correlationships among some strains of *Lactobacillus helveticus*. *Biotechnology Letters* 12(10): 765–770.

Parisi O (1958) *Il Formaggio Grana*. Mucchi, Modena: Stem.

Various authors (1979a) *Il formaggio grana. Tomo 1. Latteria Didattica P. Marconi, Thiene (VI).*

Various authors (1979b) *Il formaggio grana. Tomo 2. Latteria Didattica P, Marconi, Thiene (VI).*

Various authors (1988) *Atti giornata di studio 'Ricerca triennale sulla composizione e su alcune peculiari caratteristiche del formaggio Parmigiano Reggiano'. Reggio Emilia 28 marzo 1988. Consorzio del Formaggio Parmigiano Reggiano.*

Manufacture of Hard and Semi-hard Varieties of Cheese

P F Fox and P L H McSweeney, University College, Cork, Ireland

Background

Approximately 500 varieties of cheese are recognized by the International Dairy Federation; in addition, numerous minor cheese varieties are manufactured locally. For various reasons, a number of attempts have been made to classify cheese varieties into meaningful groups. Traditional classification schemes have been based principally on moisture content, i.e., extra-hard, hard, semihard/semisoft, or soft. Although used widely, this scheme suffers from serious limitations, since it groups cheeses with widely different characteristics, e.g., Cheddar and Emmental are classified as hard cheeses, although they have quite different textures and flavors, they are manufactured by very different technologies, and their microbiology and biochemistry are very different. This composition-based scheme is made more discriminating by including information on the source of the milk, coagulant, principal ripening microorganisms, and cook temperature. Based on the method of milk coagulation, cheeses may be divided into four super-families:

- rennet-coagulated cheeses: most major cheese varieties;
- acid-coagulated cheeses, e.g., Cottage, Quarg, Cream;
- heat/acid-coagulated, e.g., Ricotta;
- concentration/crystallization, e.g., Mysost.

Rennet-coagulated cheeses represent ~ 75% of the total world production and include all ripened varieties; owing to the great diversity of these cheeses, further classification is desirable, e.g., based on the characteristic ripening agent(s), e.g., internal bacteria, internal mold, surface mold or surface smear (bacteria), or manufacturing technology. The most diverse family of rennet-coagulated cheeses is that containing the internal bacterially ripened varieties, which include most hard and semihard cheeses. The composition of these cheeses is not defined precisely, and for the present purpose, cheese containing about 35–45% moisture will be included. Traditionally, cheeses develop a rind through which moisture evaporates; hence, the composition of cheese changes as it ages, and there is a moisture gradient from the surface to the center. The moisture content of long-ripened cheese may decrease by 5–10% during ripening. Obviously, the differentiation between hard and semihard is arbitrary.

Hard Varieties

Several subgroups of hard/semihard cheeses are discussed in other articles, i.e., Swiss-type, Dutch-type, pasta filata, and ripened-in-brine varieties. This article will focus on Cheddar-type cheese with brief descriptions of British Territorial cheeses and several other varieties which are generally similar to the above.

Cheddar-type Cheese

Hard cheeses have a moisture content of 35–40% and are manufactured by a generally similar technology, including renneting at ~ 30 °C, cutting the coagulum into small pieces, cooking to ~ 40 °C, drainage of the whey and pressing the curd. In some cases, e.g., Cheddar and other British varieties, the curds are textured in the vat ('cheddared'), and the blocks of curd are milled and dry-salted when sufficient acidity has developed. The salted curds are molded, pressed for 12–16 h, and matured for 3–12 months, or longer. Hard cheeses include Cheddar, Cheshire, Derby, Gloucester, and Leicester (British), Cantal (French), Friesian and Leiden (Netherlands), Graviera and Kefalotiri (Greece), Manchego, Idiazabal, Roncal, and Serena (Spain), Sâo Jorge (Azores, Portugal), and Ras (Egypt).

Cheddar cheese, which originated around the village of Cheddar, Somerset, England, is one of the most important cheese varieties worldwide, representing ~ 30% of total cheese production; it is produced on a large scale in the USA, the UK, Australia, New Zealand, Canada, and Ireland. Although small amounts of Cheddar cheese are made from raw milk, mainly at the farmhouse level, the vast majority is produced from pasteurized cows' milk standardized to a casein:fat ratio of ~ 0.7:1, intended to give a fat-in-dry matter (FDM) of at least 48%. With the objective of reducing the intake of dietary fat, reduced-fat Cheddar is produced commercially, but because its flavor and especially its texture (which is hard and crumbly) are inferior to those of full-fat Cheddar, it has only a minor share, perhaps 5%, of the market. However, research is continuing on ways of improving the flavor and texture of reduced-fat Cheddar and related varieties.

Some consumers prefer an intensely colored cheese, and the orange pigment, annatto (E160b), is sometimes added to the milk for Cheddar and other British cheeses. Annato is produced from the berries of the tropical plant, *Bixa orellana*, and contains two apocarotenoid pigments, bixin and norbixin.

A key feature of the manufacture of all cheeses is a decrease in pH to ~ 5, usually through the production of lactic acid in the curd. Traditionally, the lactic acid was produced by the adventitious microflora, especially lactic acid bacteria (LAB), but this was very variable and unpredictable, and hence the rate of acidification varied, with adverse effects on the quality of the cheese. The practice of adding a culture of selected LAB, referred to as a starter, was introduced at the end of the nineteenth century, and it is now standard practise to inoculate milk for Cheddar-type cheese with a culture of mesophilic LAB, i.e., *Lactococcus lactis* ssp. *cremoris* and/or *Lc. lactis* ssp. *lactis*. Initially, these cultures were undefined and contained a mixture of strains. Defined-strain starters were introduced in New Zealand in the 1930s, initially as a component of a phage-control program, and are now used widely in large Cheddar factories; their main attraction is that, if properly managed, they give a very reproducible rate of acid production and hence cheese quality, although the flavor of the cheese may be rather mild, due to the lack of microbial enzyme diversity. Today, the thermophilic organism, *Streptococcus thermophilus*, is fairly widely used as a component of starters for Cheddar cheese, mainly to increase stability to phage, to permit a higher cook temperature, and to intensify and modify the flavor of the cheese.

Sufficient rennet is added to the milk at ~ 30 °C to give a firm coagulum after about 35 min. Traditionally,

rennet prepared by extracting the stomachs of young, milk-fed calves with NaCl brine was used for Cheddar, but owing to increased production of cheese and a dwindling supply of young calves, several rennet substitutes have been introduced over the past 40 years, especially bovine and porcine pepsins, acid proteinases from *Rhizomucor meihei, R. pusillus*, and *Cryphonectria parasitica* and chymosin from genetically engineered microorganisms. An aqueous extract of the flowers of the cardoon thistle, *Cynara cardunculus* L., which is used in Portugal for the manufacture of Sera da Estrêla, is used in the western parts of Spain for the manufacture of Serena, a hard cheese.

When the coagulum is sufficiently firm, it is cut into cubes of about 1 cm. A rennet-induced milk gel is quite stable if left undisturbed, but if cut or broken, the pieces of curd contract (synerese), expelling whey. By controlling the rate and extent of syneresis, the cheese-maker can control the composition of the cheese curd and hence the rate and direction of ripening and the quality and stability of the cheese. Syneresis is promoted by increasing the temperature, reducing the pH or agitating the curds/whey. In the case of Cheddar and related cheeses, the curds/whey mixture is cooked from 30 to 37–39 °C over 30 min and held at this temperature for about 1 h. At this stage, the pH should be ~ 6.1, and the whey is drained off; if the pH has not decreased sufficiently, removal of the whey may be delayed. The curds for Cheddar and related varieties are 'cheddared,' a process unique to, and characteristic of, these varieties. The traditional cheddaring process involves forming the drained curds into beds at both sides of the vat, separated by a trough for whey drainage. After about 15 min, the beds of curd are cut into blocks, which are inverted and piled at intervals. The cheddaring process allows time for acidity to develop in the curd (from *c.* pH 6.1 to 5.4) and subjects the curds to gentle pressure, which assists in whey drainage and causes the blocks of curd to flow and spread. During cheddaring, the curd granules fuse, and the texture becomes similar to that of cooked chicken-breast meat at the end of cheddaring. This textural change was considered to be the characteristic feature of Cheddar cheese manufacture, but the principal function of cheddaring is to promote curd syneresis and to allow development of sufficient acid as a result of which colloidal calcium phosphate dissolves, leaches out of the curd, and is lost in the whey. The calcium:-casein ratio has a major effect on the textural properties of cheese, including stretchability, pliability, and meltability; these features are particularly important in Cheddar and *pasta filata* cheeses.

When the pH reaches *c.* 5.4, the curd blocks are milled into chips and salted at a level of ~ 3%, to give ~ 2% salt in the finished cheese. The salted curds are 'mellowed,' during which the salt dissolves in moisture on the surface of the chips. Mellowing is critical for the proper control of the concentration and distribution of salt in the cheese. The curds are then molded, traditionally as 10–40 kg cylinders (typically, 20 kg), and pressed overnight at ~ 200 kN m^{-2}. Cheddar cheese curd is matured at 6–12 °C for a period ranging from 3 to 24, or more, months, depending on the maturity desired. Traditionally, the wheels of Cheddar were dried slowly to form a rind and then dipped in yellow or black wax.

Although the traditional manufacturing process described above, and summarized in **Figure 1**, is still practiced in small factories and on a farmhouse scale, most Cheddar cheese is now manufactured in large (20–50 000 tonnes per annum), highly automated factories (e.g., **Figure 2**). The principal features of automated Cheddar production include the use of large (30 000 l) enclosed vats in which cheese-making commences at 30 min intervals to provide a semi-continuous supply of curd. All operations in the vat (addition and mixing of starter and rennet, cutting of the gel, and cooking of the curds) are automated and computer-controlled. After cooking, the curds/whey mixture is pumped on to an inclined screen, on which the curds and whey are separated. The drained curds are moved pneumatically to a tower, *c.* 5 m high, or to a belt system on which restricted flow and slight pressing occurs. The pH of the curds should have decreased to ~ 5.4 as they exit the tower or belt. Since it is not possible in large automated factories to hold the curds until the pH decreases to 5.4, consistent production of acid is critical and is one of the principal advantages of using defined-strain highly selected starters.

Large blocks of curd are sliced off the body of curd as it exits the tower or belt and are milled and salted automatically on a belt system. The salted curds are conveyed pneumatically to the top of a 'block former,' a large tower in which the salted curds are compressed by their own weight and subjected to a slight vacuum. As the curds exit the block former, after ~ 30 min, 20-kg blocks are cut off by a guillotine, vacuum-packaged in plastic bags, placed in cardboard boxes, stacked on a pallet (usually 1 tonne per pallet), and transferred to ripening rooms. In many large factories, the boxed cheeses are cooled rapidly in a forced-air cooling tunnel before palleting; the objective is to retard the growth of nonstarter LAB, which may cause defects in flavor and texture.

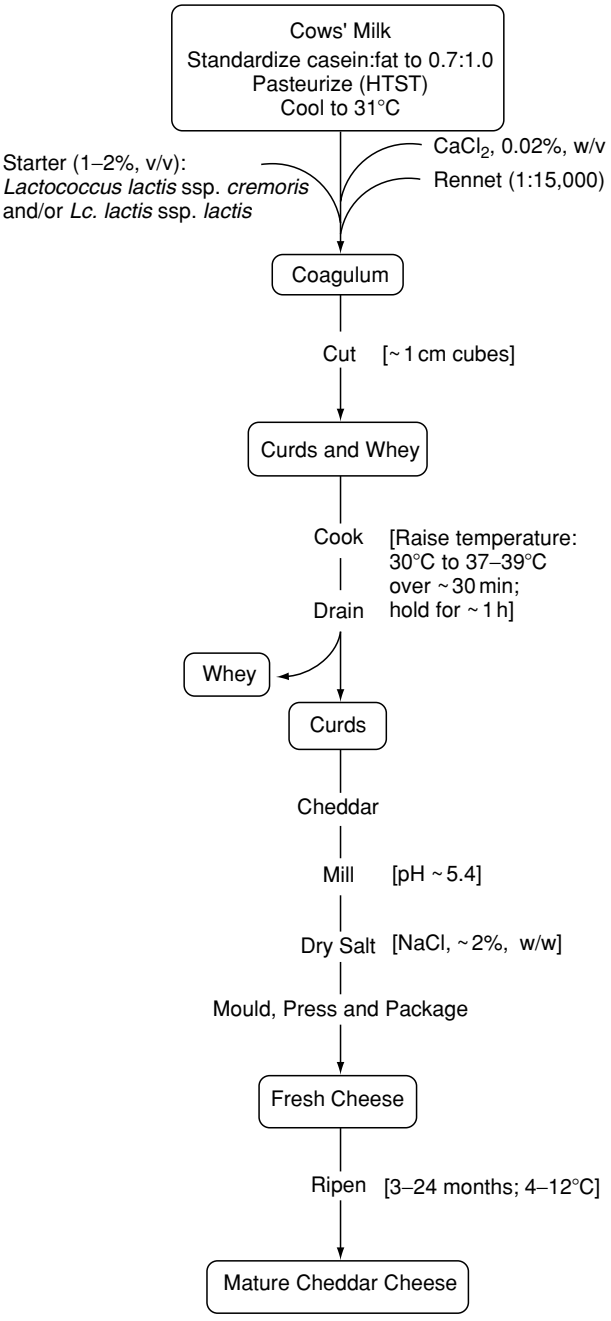

Figure 1 Protocol for the manufacture of Cheddar cheese.

The possibility of using ultrafiltration (UF) to concentrate milk to 'precheese' for Cheddar cheese has been investigated, but the texture and flavor of the product are unsatisfactory. However, UF retentate is used to standardize the fat and protein content of cheese-milk.

Other Hard Cheeses

British Territorial varieties are a group of cheeses which originated in various parts of Great Britain. The most important of these is Cheshire, which has a crumbly, hard texture and several mechanical openings. The milk is inoculated with ~4% of a mesophilic starter, and the curds are cooked at 32–35 °C, which results in a lower pH, a shorter manufacturing time, a higher moisture content, a lower calcium content (due to the lower pH at whey drainage), and a more crumbly texture than for Cheddar. The curds are held on the bottom of the vat after whey drainage, and the bed is broken occasionally to prevent the development of an extensive structure. Cheshire curds are dry-salted at ~3%, placed in molds, drained overnight, pressed, packaged, and matured at 6–8 °C.

Leicester is similar to Cheshire but is normally colored with annatto. The curds are cooked at ~37 °C and pressed after whey drainage; blocks of curd are placed on draining racks, turned, and cut to promote whey drainage. The cheeses are dry-salted and matured for 4–8 months at 10–15 °C. Derby is a little-known English variety with properties generally similar to those of Leicester; it contains 48% FDM and about 42% moisture.

Gloucester is manufactured by a process similar to that for Cheddar and is colored with annatto. The curds are cooked to 35–38 °C, textured, milled twice, dry-salted and pressed into cylinders about 40 cm in diameter and 6–8 cm high for 'Single' Gloucester and 15–20 cm high for 'Double' Gloucester. The cheeses are matured for 4–6 months at 10–13 °C.

Cantal, a hard cheese from the Auvergne region of France, is manufactured by a process somewhat similar to that used for British Territorial varieties. The milk is coagulated using calf rennet and acidified by a mesophilic lactic starter. The curds/whey mixture is not cooked, but the drained curds are cheddared and pressed to promote whey drainage. The blocks of curd are milled, dry salted, molded, pressed, and matured at 8–10 °C for 3–6 months.

Kefalotiri is a Greek cheese made from pasteurized sheep's or goats' milk standardized to *c*. 6.0% fat. The milk is inoculated with a thermophilic culture (usually *Sc. thermophilus* and *Lactobacillus delbrueckii* ssp. *bulgaricus*) and coagulated by calf rennet. The curds are cooked to 43–45 °C, transferred to molds lined with cheesecloth, and pressed at an increasing pressure. The pressed cheeses are brine-salted initially and then rubbed with dry salt for a few days to give a final salt content of *c*. 4%. The cheeses are washed occasionally with a brine-soaked cloth to control microbial growth on the surface and ripened for *c*. 3 months at ~15 °C. Kefalotiri has a hard texture and a strong, salty flavor.

Graviera is also produced in Greece, principally from ewes' milk, which is acidified by a mixed culture

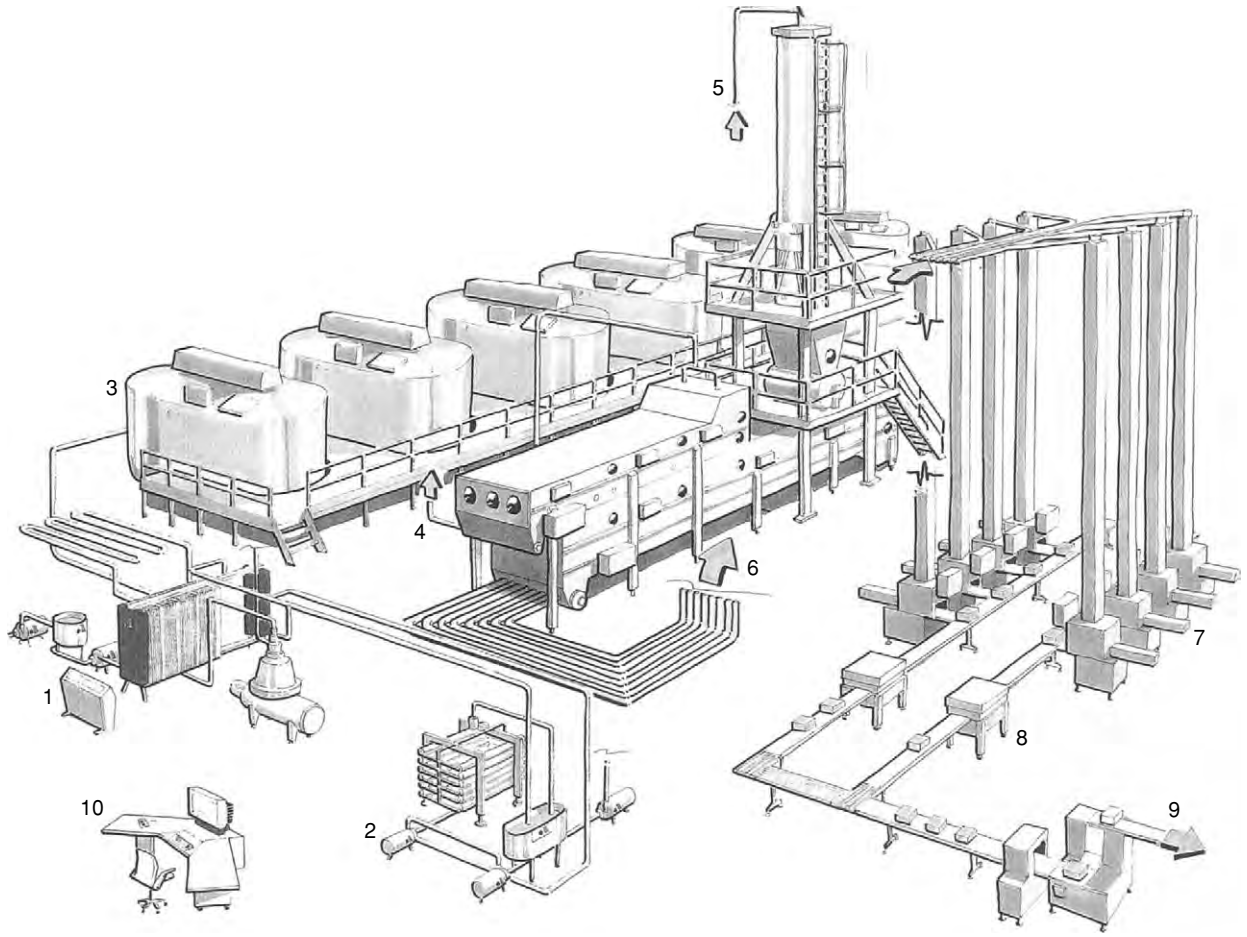

Figure 2 Large-scale Cheddar cheese-making factory incorporating the CheddarMaster 3 system with a Cheddaring tower. 1, Pasteurization and fat standardization; 2, protein standardization using UF; 3, cheese-making; 4, draining conveyor; 5, cheddaring tower; 6, salting/mellowing conveyor; 7, block former; 8, vacuum packaging; 9, cheese block packing; 10, main process control panel. Courtesy of APV Nordic Cheese, Denmark.

of mesophilic (1%; *Lc. lactis* ssp. *lactis* or *Lc. lactis* ssp. *cremoris*) and thermophilic (0.1%; *Sc. thermophilus* and *Lb. helveticus*) organisms. The curds are cooked to ~ 50 °C and, after whey drainage, molded and pressed at an increasing pressure. The cheeses are salted by frequent application of dry salt to the surface for 2–3 weeks and ripened for 3–4 months.

Ras, the principal hard Egyptian cheese, is produced from cows' milk, standardized to 3% fat, by a process generally similar to that for Kefalotiri. The curds are cooked to 45 °C, salted at a level of 1% after whey drainage, molded and pressed as for Kefalotiri. The cheeses are brined for 24 h, rubbed with dry salt daily for several weeks, and washed with brine.

Manchego, a cheese with *appellation d'origine contrôlée* (AOC) status, is probably the most important Spanish variety. It is made from ewes' milk, although generally similar cheeses (but without AOC status) are manufactured from goats' or cows' or mixed milk. Manchego is produced on an artisanal scale from raw milk, without culture addition, or commercially from pasteurized milk, inoculated with a mesophilic starter. The milk is coagulated with standard calf rennet. The curds are cooked to ~ 38 °C, molded, pressed for 12–16 h, brine-salted and matured for at least 2, and up to 12, months, depending on the maturity required, at 10–15 °C and 85% equilibrium relative humidity (ERH).

Idiazábal, an AOC cheese, is produced in the Basque region of northern Spain from raw ewes' milk coagulated at 38 °C. The coagulum is allowed to cool to 25 °C, then broken and the curds ladled into molds. The cheeses are salted by brining or by the application of dry salt and matured in caves for 2 months at ~ 9 °C. The cheeses may be smoked in beechwood kilns and matured further for up to a year.

Roncal is another AOC cheese made in northern Spain from raw ewes' milk acidified by the adventitious microflora. The curds/whey mixture is cooked to ~ 37–40 °C and the curds then allowed to settle to the bottom of the vat. The whey is removed slowly and the curds pressed against the sides of the vat, molded, pressed, and dry-salted. Roncal is smoked and ripened at 6–8 °C and 100% ERH for 45–50 days.

La Serena is a hard cheese made in western Spain from ewes' milk. Traditionally, raw milk is used, but pasteurized milk is now used for large-scale production. The milk is coagulated at ~ 30 °C with rennet extracted from the thistle, *Cynara cardunculus*. The flat cylindrical cheeses are dry-salted for ~ 24 h and ripened for about 2 months at ambient temperature (10–18 °C).

São Jorge is an AOC cheese produced on a small scale from raw cows' milk on the island, São Jorge, in the Azores. The milk is coagulated using calf rennet, and acidification of the curd depends on the activity of the indigenous adventitious microflora. The pressed cheeses are ripened at 10–12 °C for up to 12 months and, being made from raw milk, develop a strong, piquant flavor. The cheese, which is considered to resemble Cheddar, has been the subject of very little study.

Leyden and Friesian are hard cheeses produced in The Netherlands. Leiden is made from partially skimmed milk, has a flat cylindrical shape, weighs 8–12 kg, and is matured for 6–12 months. Its texture is dry and hard (artisanal, ~ 30% FDM; commercial, 40% FDM), and it is flavored with cumin seeds. Traditional Friesian cheese is similar to Leiden and is made from partially skimmed raw milk. It has a very hard texture and ~ 20% FDM. It may be made without spices, with cumin, or with cumin and cloves. Today, most Friesian cheese is made from fresh, pasteurized, partly skimmed milk, acidified using a mesophilic culture.

Semihard Varieties

The 'semihard' group of cheeses is rather arbitary and heterogeneous, and the distinction between it and other groups, e.g., hard cheeses, smear-ripened varieties or *pasta filata* cheeses, is not clear-cut. Semihard cheeses include Colby and Monterey (stirred-curd Cheddar-type cheeses), some British Territorial varieties (Caerphilly, Lancashire, and Wensleydale), Bryndza (Slovakia), and Mahon and Majorero (Spain).

Caerphilly, which originated in Wales, is a crumbly, acidic cheese made from pasteurized cows' milk using calf rennet and a mesophilic starter. The curds are cooked to 32–34 °C and held at this temperature for about 1 h. The whey is drawn off and the curds held in the vat where rapid acidification occurs. Salt (1%) is added to the curds before molding and pressing overnight. The pressed curds are brine-salted for 24 h, packaged, and ripened at 10–13 °C for 10–14 days.

Lancashire is made from cows' milk using calf rennet and a mesophilic starter. The curds and whey are not cooked, but the drained curds are cheddared and held overnight for acidification. The next day, fresh curds are mixed with the acidified curds and the mixture milled to promote uniformity, dry-salted, molded, pressed, and ripened at 13–18 °C for 3–12 weeks.

Wensleydale cheese, which originated in Yorkshire, England, is made from cows' milk, inoculated with a mesophilic starter; the curds are cooked at 32–34 °C. The bed of drained curds is broken to assist whey drainage, dry-salted, molded, and pressed lightly. Wensleydale, which is matured for *c.* 1 month at ~ 12 °C, has mechanical openings and a mild, acidic taste. A blue variant is produced also.

Two stirred-curd variants of Cheddar cheese are recognized: Colby and Monterey. The manufacture of Colby, which originated in the USA, follows a protocol similar to Cheddar until after cooking, when some whey is removed and replaced by cold water, which reduces the lactose content and pH, and increases the moisture content of the cheese. The curds/whey mixture is stirred and most of the whey removed; the curds and remaining whey are stirred vigorously and the remaining whey drained off. The drained curds are stirred until the pH reaches ~ 5.4, when they are dry-salted, molded, pressed, and ripened for 2–3 months at 3–4 °C. Colby contains less than 40% moisture and $\geq$ 50% FDM.

Monterey (Monterey Jack) cheese was first made in California and is similar to Colby. The cheese, which contains < 44% moisture and $\geq$ 50% FDM, is allowed to form a rind before waxing or packaging in a plastic film. Monterey, which is ripened for 5–7 weeks at ~ 4 °C, has many mechanical openings.

Bryndza is made from sheep's milk coagulated with rennet (with significant lipase activity) and acidified by the adventitious microflora of the milk or by a mesophilic starter. The curds are allowed to settle to the bottom of the vat, most of the whey is removed, and the curds are consolidated manually into lumps, which are placed into cloth bags and stored until sufficient acidity develops (about 3 days). The fresh cheese is known as Hrudka and may be sold locally. However, most Hrudka is transferred to a central factory, where it is broken up, salted, passed between granite rollers to a smooth paste, packed in polythene-lined wooden tubs, and matured.

Mahon is produced in the Balearic Island of Minorca from raw cows' milk acidified by its adventitious microflora. Mahon is brine-salted and ripened at 18 °C for *c.* 2 months. Although the texture of Mahon is semihard, its moisture content is reported to be about 32%. Majorejo is made from goats' milk on the Canary Island, Fuerteventura. Milk is acidified by a mesophilic starter and coagulated using rennet extract, although artisanal cheesemakers rely on the indigenous microflora of the milk and use rennet paste as coagulant. The curds are molded in braided palm leaves (which give its surface a characteristic pattern), pressed lightly, and dry-salted. Majorejo develops a strong flavor during ripening.

Ripening

Although rennet-coagulated cheese curd may be consumed at the end of manufacture, it is usually ripened (cured, matured) for a period ranging from ~ 3 weeks to > 2 years; generally, the duration of ripening is inversely related to the moisture content of the cheese. Many varieties may be consumed at any of several stages of maturity, depending on the flavor preferences of consumers and economic factors.

The unique characteristics of the individual cheeses develop during ripening as a result of a complex set of biochemical reactions, which are largely predetermined by the manufacturing process, i.e., by composition, especially moisture, NaCl, pH, level of coagulant activity in the curd, type of starter, and an intentional or adventitious secondary microflora. The ripening agents in cheese are derived from one or more of the following:

- coagulant;
- indigenous milk enzymes, especially proteinase and lipase;
- starter bacteria and their enzymes;
- secondary microorganisms, both intentional and adventitious, and their enzymes.

The characteristics of many cheeses are dominated by the metabolic activity of the secondary microflora, but traditionally, Cheddar and the other varieties discussed here were not inoculated with a secondary culture, although an adventitious secondary microflora of nonstarter LAB (NSLAB), mainly mesophilic lactobacilli, grows to 10^7–10^8 CFU g^{-1}. The NSLAB are more heterogeneous in cheese made from raw milk than in that made from pasteurized milk, which probably explains the more intense, but more variable, flavor development in the former. The interior of Cheddar cheese is a rather selective environment: it has a low pH (~ 5) and a relatively high salt content, is anaerobic, lacks a fermentable carbohydrate, and may contain bacteriocins produced by the starter bacteria; hence, the rather limited diversity of NSLAB.

The primary biochemical changes in ripening cheese are the metabolism of residual lactose and the catabolism of lactate and citrate (often referred to collectively, but erroneously, as glycolysis), lipolysis and proteolysis, which are followed and overlapped by secondary catabolic changes to the products of these primary pathways, including deamination, decarboxylation, and desulfurylation of amino acids and β-oxidation of fatty acids. Considerable progress has been made on elucidating the primary reactions in Cheddar. Unfortunately, only limited information is currently available on the ripening of the other hard and semihard cheeses discussed here, but since the ripening agents and the composition of all these cheeses are generally similar, the overall features of ripening are probably similar. However, differences in flavor and texture between members of this group are readily apparent, due partially to the degree of maturity of the cheeses, but also reflect the effect of small differences in composition and manufacturing technology, which affects the microflora of the cheese.

Glycolysis and Related Events

Most ($\sim 98\%$) of the lactose in milk is removed in the whey as lactose or lactic acid but fresh curd for hard and semihard cheese contains 1–2% lactose. The catabolism of lactose and resulting lactic acid is understood at the molecular level. For most cheese varieties, the residual lactose is metabolized by the starter bacteria within about 12 h. However, owing to the low pH (~ 5.4) of curd for Cheddar and similar cheeses at molding and the practise of dry-salting the curd for these cheeses, which rapidly gives an inhibitory level of NaCl throughout the cheese, residual lactose is metabolized slowly, normally by residual starter activity to L(+)-lactic acid; young Cheddar cheese contains $\sim 1.5\%$ lactic acid. The curds for Colby and Monterey are washed to reduce the lactose content of the curd (and also to increase its moisture content).

In Cheddar and probably in related varieties, L(+)-lactate is racemized to DL-lactate. Racemization has no significant effect on cheese flavor, but if the concentration of D-lactate is too high, it may form undesirable crystals on the cheese surface. Oxidation of lactic acid to acetic acid occurs to an extent dependent on the concentration of O_2 in the curd, which is strongly affected by the gas permeability of the packaging material.

The pH of most cheeses increases during ripening, but the pH of Cheddar changes little. Cheese has a strong buffering peak at pH ~ 5.2, and if the pH of cheese falls below this value, as for Cheddar,

insufficient NH_3 and amines may be produced to cause an increase in pH. However, if the lactose content of Cheddar curd is reduced by washing or whey replacement, residual lactose in the cheese is metabolized rapidly, and the pH increases during ripening. Low-lactose cheese has a clean, mild flavor, whereas lactose-supplemented cheese develops a strong, harsh flavor, probably due to the low pH.

Cheddar-type cheese contains ~ 2 g of citrate per kilogram, which is fermented slowly by NSLAB, principally to formic acid and CO_2; the latter may cause an undesirable open texture.

Lipolysis

Lipolysis is quite limited in most cheese varieties, including Cheddar and related types. A low concentration of fatty acids is important, probably essential, for cheese flavor, but even slightly excessive lipolysis causes rancidity or unbalanced flavor in hard and semihard cheeses. Considerably more lipolysis occurs in raw milk than in pasteurized milk Cheddar, suggesting that the indigenous lipoprotein lipase in milk and/or NSLAB is significant. Although LAB are weakly lipolytic, they do possess esterases/lipases, which release a low level of fatty acids during a long ripening period. Fatty acids contribute directly to cheese flavor and also serve as substrates for several flavor-generating reactions (Figure 3). These secondary reactions are quite limited in hard and semihard, internal bacterially ripened cheese but probably do contribute to overall flavor intensity and balance, especially the lactones and thioesters. Esters, especially ethyl butyrate, ethyl hexanoate, and ethyl octanoate, are responsible for a fruity flavor, which is usually regarded as a defect in Cheddar-type cheese. The oxidation of fatty acids to alkan-2-ones (*n*-methyl ketones) is a key flavor-generating reaction in mold-ripened cheese, especially blue cheeses, but is very

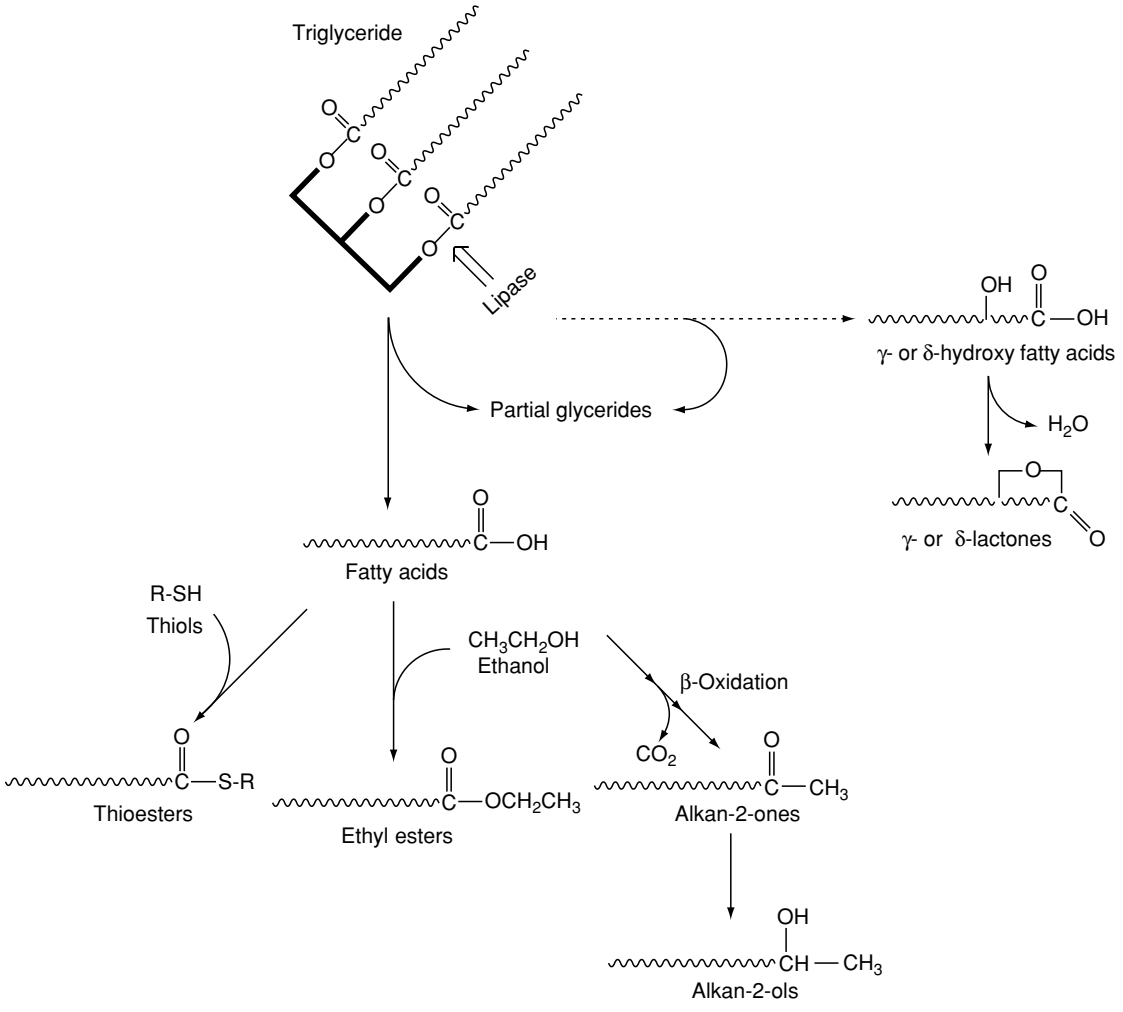

Figure 3 General pathway for the production and catabolism of fatty acids in cheese during ripening.

limited in hard and semihard bacterially ripened cheeses.

Proteolysis

Proteolysis is the most complex and perhaps the most important of the three primary ripening events, especially in internal bacterially ripened cheese. It is primarily responsible for changes in cheese texture and makes a significant contribution to flavor, especially background flavor *via* the formation of amino acids and small peptides. Catabolism of amino acids leads to many sapid and aromatic compounds, which are probably major contributors to cheese flavor and may be rate-limiting in flavor development in hard and semihard cheese. Owing to its perceived importance to the rate of ripening and to the quality of the mature cheese, proteolysis and related events have been the subject of intense research in recent years.

Cheddar-type cheese contains a broad range of proteinases and peptidases which originate from the coagulant, milk, starter, and NSLAB. The proteinases and peptidases from most of these sources have been isolated and characterized, and their contribution to cheese ripening established.

The sequence of proteolysis in Cheddar and similar varieties is summarized in Figure 4. Primary proteolysis is catalyzed mainly by the coagulant and is principally responsible for the hydrolysis of α_{s1}-casein which, in mature Cheddar cheese, is completely hydrolyzed at Phe_{23}–Phe_{24}, at least 50% at Leu_{101}–Lys_{102}, and, to a lesser extent, at other bonds. In cheese, β-casein is hydrolyzed only slightly by most coagulants, but about 50% is hydrolyzed by plasmin to γ-caseins and proteose peptones. Primary proteolysis contributes little or nothing to the flavor of cheese but is mainly responsible for the changes in cheese texture which occur during the early stages of ripening.

The large C-terminal peptides produced from α_{s1}- and β-caseins by the coagulant and plasmin, respectively, and *para*-κ-casein produced during milk coagulation, accumulate during ripening and are present even in extra-mature cheese. However, the complimentary N-terminal peptides from α_{s1}- and β-caseins are hydrolyzed by lactococcal proteinases and peptidases to smaller peptides and amino acids. The NSLAB contribute mainly at the level of amino acid production. About 200 peptides have been isolated from Cheddar cheese and characterized.

Small peptides and amino acids contribute to the background flavor of cheese, but an excessive concentration of small hydrophobic peptides may cause bitterness, which is probably the most common flavor defect in cheese. More importantly, amino acids can be converted to a range of flavor compounds, including amines, acids, carbonyls, NH_3, sulfur compounds, and hydrocarbons by enzymes derived from the starter or NSLAB or through chemical reactions. The enzymatic catabolism of amino acids is summarized in Figure 4 and elaborated in Figures 5–7. These reactions and the enzymes involved have attracted considerable attention during the past few years. The principal amino acid-derived flavor compounds in cheese are listed in Table 1.

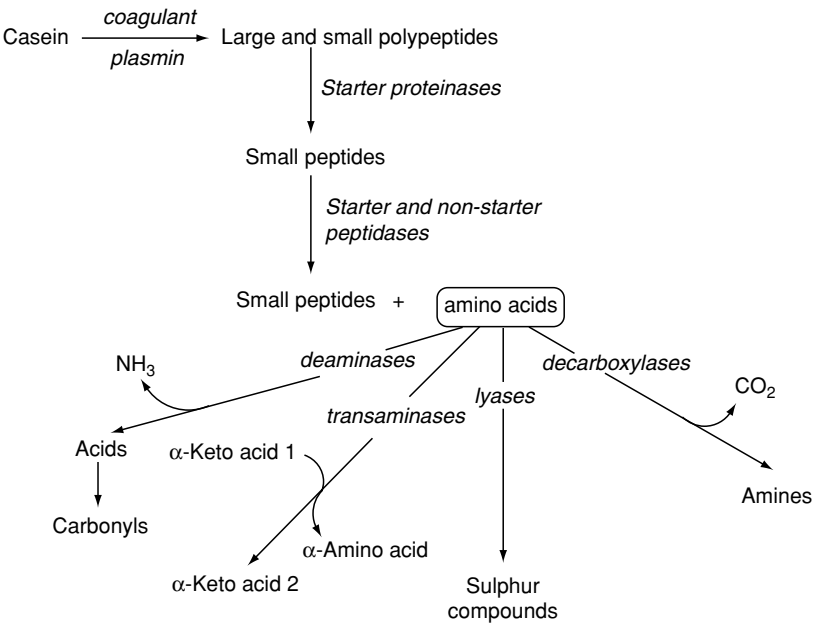

Figure 4 Summary of proteolysis and amino acid catabolism in cheese during ripening.

Table 1 Major aroma compounds derived from branched chain and aromatic amino acids and methionine in cheese during ripening

Amino acid	*Carboxylic acids*[a]	*Amines*	*Aldehydes*	*Alcohols*	*Thiols/others*
Valine	2-Methyl propionic acid	2-Methyl propylamine	2-Methyl propanal	2-Methyl propanol	
Leucine	3-Methylbutanoic acid	3-Methyl butylamine	3-Methyl butanal	3-Methyl butanol	
Isoleucine	2-Methylbutanoic acid	2-Methyl butylamine	2-Methyl butanal	3-Methyl butanol	
Phenylalanine	Phenylacetic acid	Phenylethyl amine	Phenylacetaldehyde, benzaldehyde	Phenylethanol	
Tyrosine	*p*-Hydroxyphenylacetic acid	*p*-Hydroxyphenylethyl amine (tyramine)	*p*-Hydroxyphenylacetaldehyde, *p*-hydroxybenzaldehyde	*p*-Hydroxyphenylethanol	*p*-Cresol, phenol
Tryptophan	Indole-3-acetic acid	Indole-3-ethyl amine (tryptamine)	Indole-3-acetaldehyde Indole-3-aldehyde	Indole-3-ethanol (tryptophol)	Skatole, indole
Methionine	3-Methylthiopropionic acid	3-Methylthiopropylamine	3-Methylthiopropanal (methional)	3-Methylthiopropanol	Methanethiol

[a]Further carboxylic acids can be produced from amino acids by deamination.

From Yvon M and Rijnen L (2001) Cheese flavour formation by amino acid catabolism. *International Dairy Journal* 11: 185–201.

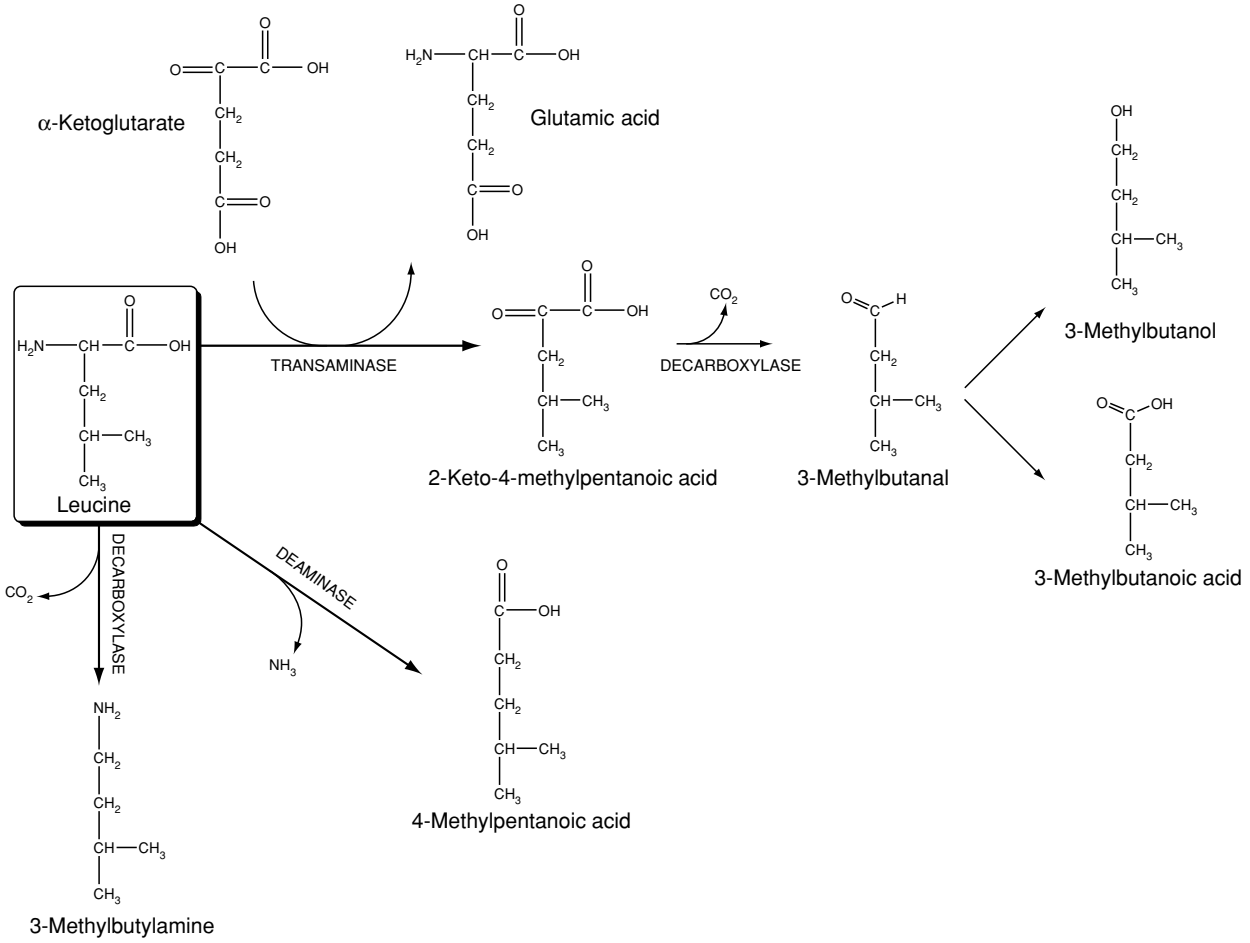

Figure 5 Catabolism of leucine. Similar pathways apply for the other branched-chain amino acids, isoleucine, and valine.

The Strecker reaction between an amino acid and a dicarbonyl (Figure 8) is an important chemical flavor-generating reaction involving amino acids in cheese, but the Maillard reaction between an amine (e.g., amino acid) and a monocarbonyl (e.g., a reducing sugar) may be significant to cheese flavor (and color) under certain circumstances, e.g., low-moisture grated cheese or cooked cheese.

Adjunct Cultures

The use of pasteurized milk for cheese-making dates from the beginning of the twentieth century but did not become widespread until about 1950. The principal objective of pasteurizing milk is to kill pathogenic bacteria, but most other microorganisms, including desirable NSLAB, are killed also, and many indigenous enzymes are inactivated. For public-health reasons and because the quality of cheese made from pasteurized milk is more consistent and uniform than that made from raw milk, it is very unlikely that raw milk will be used for the manufacture of Cheddar cheese in the very large, highly mechanized factories which are now common. However, there is a general perception that the flavor of modern Cheddar has become too mild, possibly owing to the use of very high microbiological quality cheese milk (pasteurization of which renders it essentially free of NSLAB), the use of highly active starters containing only a few strains of *Lactococcus*, and the manufacture of cheese in enclosed equipment which reduces contamination from the environment. While most consumers today prefer mild Cheddar, there are *niche* markets for more mature, intensely flavored cheese. Also, some retailers and cheese manufacturers wish to offer for sale cheese with a flavor identifiable with that company. The use of adjunct cultures (which are not required for acid production) is considered to be the most effective way of achieving this; essentially, these cultures diversify the microflora of the cheese and are an attempt to reproduce the microflora of raw milk cheese in a controlled way. Since the

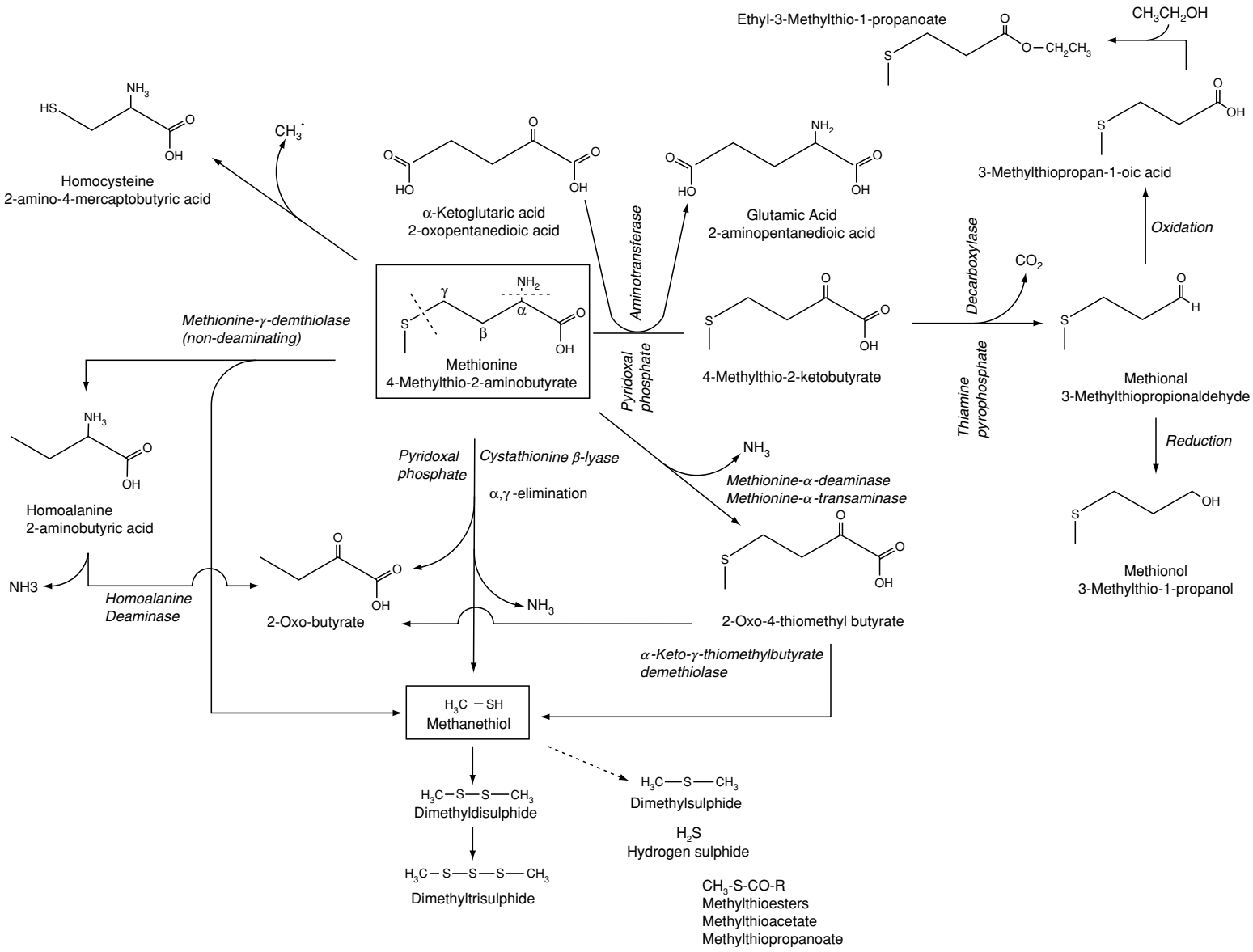

Figure 6 Suggested pathways for the catabolism of methionine in cheese during ripening.

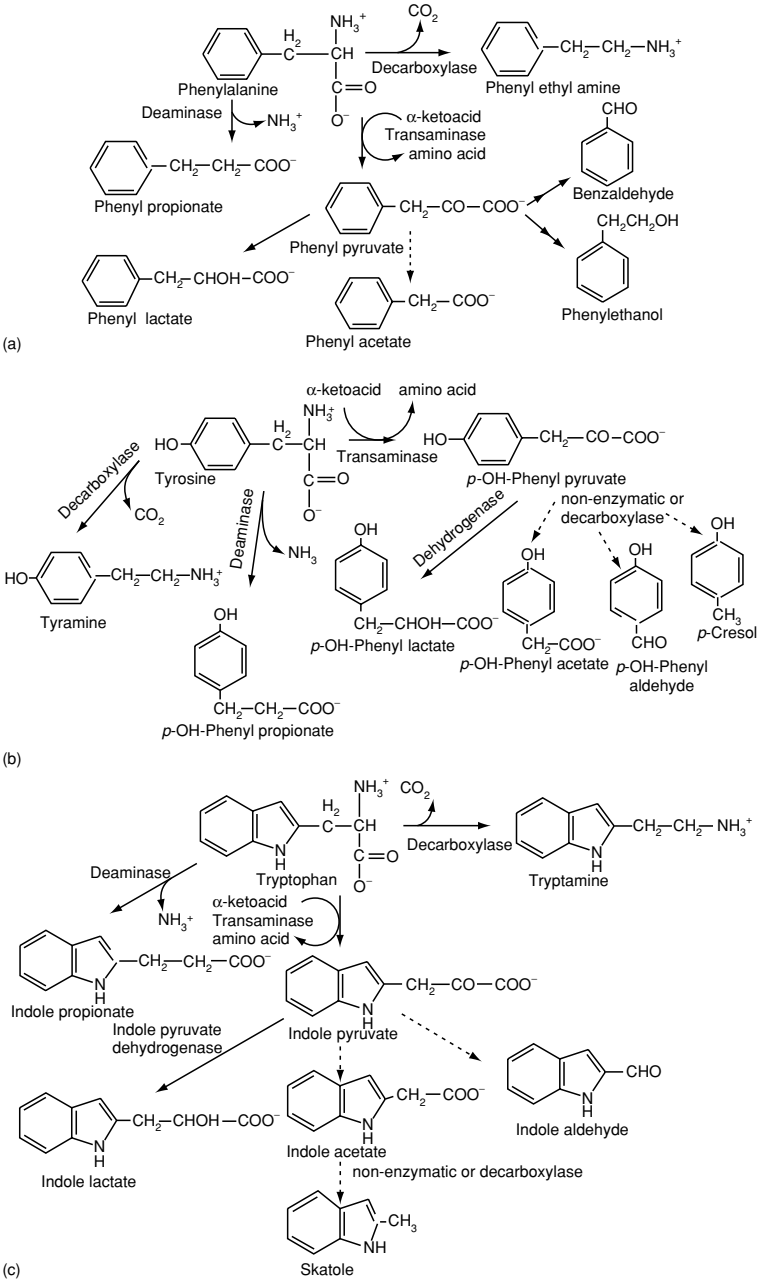

Figure 7 Suggested pathways for the catabolism of (a) phenylalanine, (b) tyrosine, and (c) tryptophan in cheese during ripening.

NSLAB in Cheddar cheese are mainly mesophilic lactobacilli, especially *Lb. casei* and *Lb. paracasei*, most studies on the use of adjunct cultures for Cheddar cheese have concentrated on mesophilic lactobacilli; most investigators report positive results. Thermophilic lactobacilli are more effective than mesophilic lactobacilli, possibly because the former die rapidly in cheese, releasing their intracellular enzymes, while the latter do not. However, the number of thermophilic lactobacilli added must be controlled carefully – if too numerous, a Swiss cheese-type flavor will develop.

Acceleration of Cheese Ripening

Some high-moisture cheeses develop an intense flavor through a very active secondary microflora, e.g., internal blue mold, external white mold, or a bacterial surface smear; these cheeses ripen rapidly. High-moisture internal bacterially ripened cheeses also mature rapidly but develop a low flavor intensity; if

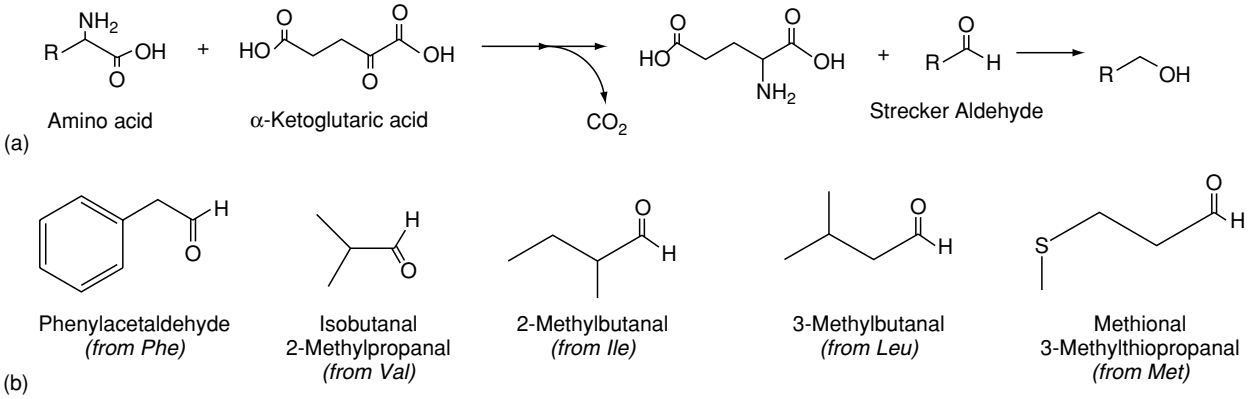

Figure 8 (a) Strecker reaction between an amino acid and an α-keto acid (e.g., α-keto glutaric acid) and (b) structures of some Strecker aldehydes found in cheese.

the ripening of such cheeses is extended, they will probably develop off-flavors. It is possible to develop an intense flavor in internal bacterially ripened cheeses, such as hard and semihard varieties, only if the moisture content is low and if they are ripened for a long period, e.g., 2–3 years. Owing to the high cost of ripening facilities and stocks, the ripening of extra-mature, low-moisture cheese is expensive, and there is commercial interest in accelerating the ripening process. Most of the work on accelerating cheese ripening has been on Cheddar.

Methods for accelerating cheese ripening fall into six categories:

- elevated ripening temperature;
- exogenous enzymes;
- chemically or physically modified cells;
- genetically modified starters;
- adjunct starters.

Ripening at an elevated temperature, e.g., up to 15 °C, is the simplest and cheapest method for accelerating the ripening of Cheddar cheese, but modified primary starters and adjunct cultures appear to be very promising. Techniques for accelerating ripening may also be applicable to low-fat cheeses, which tend to ripen slowly.

See also: **Cheeses**: Types of Cheese; **Enzymes**: Functions and Characteristics; **Goat**: Milk; **Lactic Acid Bacteria**; **Milk**: Physical and Chemical Properties; **Pasteurization**: Pasteurization of Liquid Products; **Sheep**: Milk; **Starter Cultures**

Further Reading

Eck A and Gilles JC (2000) *Cheesemaking: From Science to Quality Assurance*, 2nd edn. Paris: Technique et Documentation (Lavoisier).

Fox PF (ed.) (1993) *Cheese: Chemistry, Physics and Microbiology*, 2nd edn., vols. 1 and 2. London: Chapman & Hall.

Fox PF, Guinee TP, Cogan TM and McSweeney PLH (2000) *Fundamentals of Cheese Science*. Gaithersburg, MD: Aspen.

Fox PF and Wallace JM (1997) Formation of flavour compounds in cheese. *Advances in Applied Microbiology* 45: 17–85.

Fox PF, Wallace JM, Morgan S *et al.* (1996) Acceleration of cheese ripening. *Antonie van Leeuwenhoek* 70: 271–297.

Kosikowski FV and Mistry VV (1997) *Cheese and Fermented Milk Foods*, 3rd edn., vols. 1 and 2, Westport, CT: F.V. Kosikowski LLC.

McSweeney PLH and Sousa MJ (2000) Biochemical pathways for the production of flavour compounds in cheese during ripening. *Lait* 80: 293–324.

Robinson RK (ed.) (1995) *A Colour Guide to Cheese and Fermented Milks*. London: Chapman & Hall.

Robinson RK and Wilbey RA (1998) *Cheesemaking Practice, R. Scott*, 3rd edn., Gaithersburg, MD: Aspen.

Scott R (1986) *Cheesemaking Practice*. London: Elsevier Applied Science.

Smit G, van Hylckama Vlieg JET, Smit BA, Ayad EHE and Engels WJM (2002) Fermentative formation of flavour compounds by lactic acid bacteria. *Australian Journal of Dairy Technology* 57: 61–68.

Yvon M and Rijnen L (2001) Cheese flavour formation by amino acid catabolism. *International Dairy Journal* 11: 185–201.

Cheeses with 'Eyes'

P L H McSweeney, University College, Cork, Ireland

Synopsis

Mechanical openings (resulting from the incomplete fusion of curd pieces) are common in many cheese varieties where they may be considered desirable (e.g., Monterey) or a defect (e.g., Cheddar). However, some internal bacterially ripened varieties are characterized by the development of 'eyes' caused by gas (CO_2), produced by bacterial metabolism, being trapped in the curd. The development of eyes in cheese is governed by the rate of gas production by bacteria and the ability of the curd to retain the gas. There are two main families of cheeses with eyes: Dutch types (Edam, Gouda, and related varieties), which have small eyes, and Swiss types, which are characterized by large eyes. In the case of Edam and Gouda cheeses, CO_2 is produced by the action of *Leuconostoc* spp. or citrate-utilizing (Cit^+) lactococci on citrate (*See* **Cheeses: Dutch-type Cheeses**), whereas in Swiss varieties, CO_2 is produced by *Propionibacterium freudenreichii* from lactate. These two families of cheeses are quite different in terms of their manufacturing protocol, ripening conditions, and flavor and appearance of the final product.

The most popular Swiss-type cheese is Emmental (Figure 1), which originated in the Emm river valley in Switzerland, but is now made world-wide. This article will concentrate on the manufacturing protocol for, and the ripening of, Emmental. A number of other related Swiss-type cheeses will be discussed briefly.

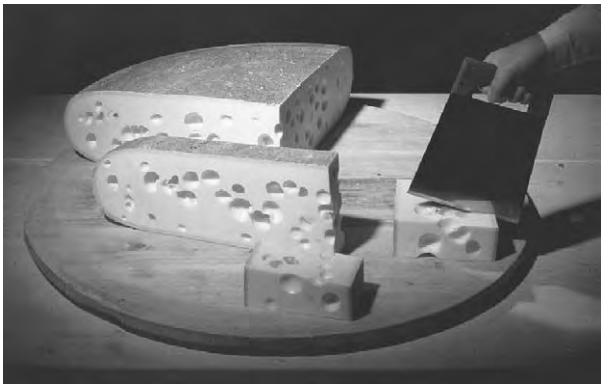

Figure 1 Emmental cheese. Reprinted with permission from Roginski H, Fuquay JW and Fox PF (eds) (1992) *Encyclopedia of Dairy Sciences*. Academic Press: London.

General Characteristics of Emmental Cheese

Typically, Emmental contains *c*. 30.5% fat, 35.5% moisture, 27.5% protein, 0.5–0.7% salt, and 3.5% ash, and has a pH of *c*. 5.6. It has a mild, sweet, and somewhat nutty flavor. Emmental cheese is characterized by large eyes that are caused by CO_2 produced by *P. freudenreichii*, which metabolizes lactate to propionate, acetate, and CO_2. Propionibacteria do not grow in the milk during cheesemaking but grow in the curd during the early stages of maturation when the cheeses are transferred to a 'hot room' ($\sim$20 °C) to promote their growth. The texture of Emmental is quite rubbery and entraps the CO_2 (which migrates through the curd until it reaches a fissure or weakness at which an eye develops). The texture of Emmental cheese results from a high cooking temperature ($\sim$55 °C), which inactivates much of the coagulant, and a high pH at draining, leading to a high Ca^{2+} content in the curd. Emmental cheese is brine-salted.

The manufacturing protocol for Emmental is shown Figure 2. Traditionally, Emmental is made from raw cows' milk, which is acidified through the action of a mixed thermophilic starter consisting of *Streptococcus thermophilus* and a *Lactobacillus* sp. *Lb. helveticus* was used traditionally, but *Lb. delbrueckii* sp. *bulgaricus* and *Lb. delbrueckii* sp. *lactis* are now common also. Propionibacteria are added to the milk (usually to *c*. 10^3–10^5 cfu ml^{-1}) or gain access to the milk from the environment (which is often the case in traditional manufacture). The milk, at 30 °C, is coagulated using calf rennet, and the coagulum is cut into small pieces and cooked to $\sim$55 °C until the curd pieces are of the desired firmness. The curds–whey mixture is transferred into molds where the whey is drawn off. The mold is sufficiently large to give a cheese wheel weighing up to *c*. 100 kg and with a diameter of up to 1 m. The size of the Emmental wheel is far from cosmetic. The large size of the wheel means that Emmental has a relatively low surface area-to-volume ratio. This is of significance, since it determines the rate of cooling of the curd (and thus the activity of the starter) and salt diffusion throughout the cheese mass, and helps to trap gas within the cheese and allows gas pressure to increase within the cheese until eye formation occurs. Over the next 1–2 days, the wheels are pressed and turned frequently. During this time, the curd cools, and acid production by the starter (which was dormant during the scalding step) recommences. Lactic acid fermentation in Emmental takes *c*. 12 h. After pressing, the cheese wheels are brine-salted. Swiss cheeses have the lowest NaCl content of all cheese varieties, permitting the growth of its

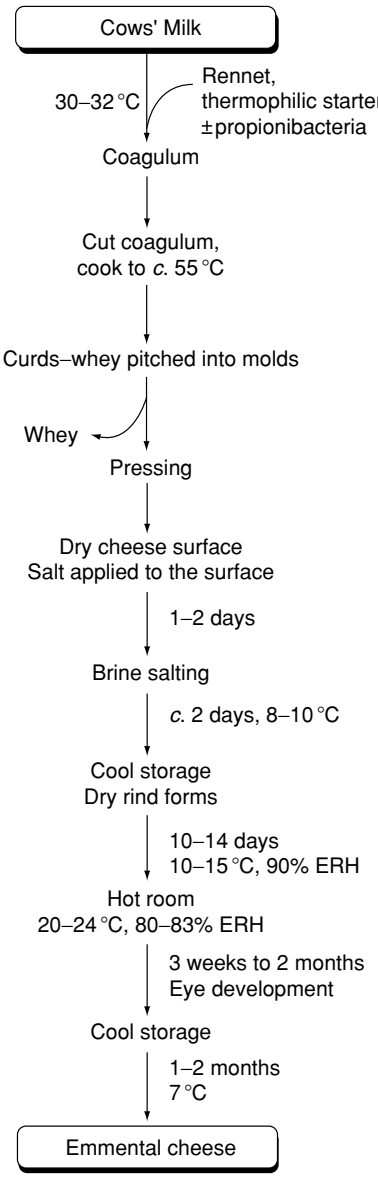

Figure 2 Traditional manufacturing protocol for Emmental cheese. Reprinted with permission from P.F. Fox, T.P. Guinee, T.M. Cogan and P.L.H. McSweeney. Fundamentals of Cheese Science, © 2000, Aspen Publishers, Gaithersburg, MD.

secondary flora during ripening (*P. freudenreichii*). *Propionibacterium freudenreichii* is very sensitive to NaCl and will not grow at a_w values below 0.96 (*c.* 1.2% NaCl).

Although Emmental is manufactured in industry by this traditional process, a rindless product is now produced in large factories and, in many countries with a developed dairy industry, is the major type of Emmental produced. Rindless, or 'block,' Emmental is manufactured by a process generally similar to that described for traditional Emmental, but this cheese is manufactured from milk with a lower fat content (*c.* 2.5%) and cooked to a lower temperature. The cheese is molded in the form of blocks that are wrapped in plastic film; thus, no rind develops during ripening.

Ripening of Swiss Cheese

In common with most, if not all, rennet-coagulated cheeses, immediately after manufacture, Swiss cheese has a bland flavor and a texture and appearance quite different from the mature cheese. The characteristic flavor, texture, and appearance of Swiss cheese develop during ripening. The ripening of Swiss cheese typically starts with storage in a cool room and 10–15 °C and 90% equilibrium relative humidity (ERH) for 10–14 days; in traditional practice, the cheeses are brushed, dry salt is rubbed on to the surface, and the cheeses are turned daily until a smooth rind develops. The cheeses are then transferred to a hot room (20–24 °C, 80–83% ERH) and held there for 3 weeks to 2 months. The cheeses are held in the hot room until adequate eye formation occurs (which is assessed traditionally by tapping the cheese and listening to the hollow sound produced); cheeses are then transferred to another ripening room at about 7 °C and matured for 1 or 2 months for further flavor development.

During the ripening of Swiss cheese, protein and, to a lesser extent, lipid in the cheese are degraded to flavor compounds, lactate is metabolized with important consequences for flavor and eye development, and changes occur to the microflora of the cheese. The ripening agents in Swiss cheese originate from the milk (particularly the principal indigenous proteinase in milk, plasmin), from the microflora of the cheese (starter, nonstarter bacteria, and propionic acid bacteria in addition to a minor contribution from the coagulant (usually chymosin). The ripening of Swiss cheese is summarized in Figure 3.

Starter bacteria reach maximum numbers in Swiss cheese soon after manufacture. Thereafter, their numbers decline rapidly as a result of the lack of a readily fermentable carbohydrate and low pH. Numbers of viable *Sc. thermophilus* decline at a slightly faster rate than those of the *Lactobacillus* sp. However, the importance of the starter to cheese ripening does not end with their death. After death, cells lyse at a strain-dependent rate, releasing their intracellular enzymes, which are of great importance in the ripening of Swiss cheese. Starter activity during acidification also has a major indirect effect on ripening since it is their action, together with the buffering capacity of the cheese, that controls pH and oxidation–reduction potential (E_h) which, in turn, have a very important effect on the rate of

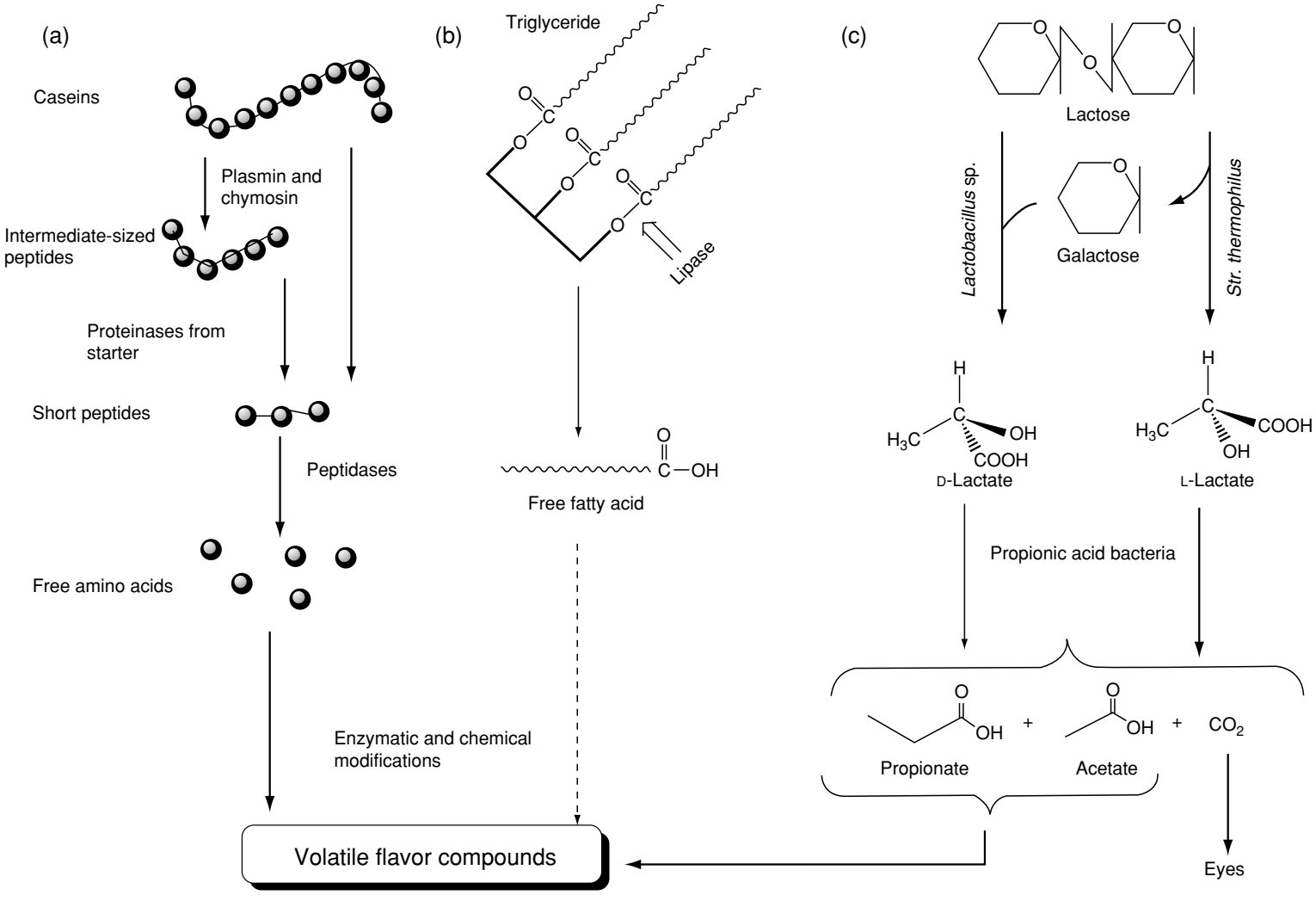

Figure 3 Schematic representation of the principal biochemical events occurring during the ripening of Swiss-type cheese. (a) Proteolysis and catabolism of amino acids, (b) lipolysis and (c) metabolism of lactose and lactate.

many ripening reactions. Most, if not all, ripened cheeses develop an adventitious secondary flora during ripening. These 'non-starter lactic acid bacteria' (NSLAB), typically mesophilic lactobacilli, are present in the raw milk or gain entry from the cheese-making environment; in the case of cheese made from pasteurized milk, they may survive pasteurization, perhaps in a heat-shocked state, or gain entry from the environment. NSLAB are quite different from the thermophilic lactobacilli used as starters in the manufacture of Swiss cheese, and include species of lactobacilli such as *Lb. casei* and *Lb. paracasei*. As with all other cheeses, the NSLAB flora of Swiss cheese is uncontrolled, but increases from very low numbers (particularly in the case of cheeses made from pasteurized milk) to *c.* 10^8 cfu g^{-1} cheese. Studies on the role of the NSLAB flora in Swiss cheese suggest that they contribute to the differences in flavor between raw and pasteurized milk cheeses and play a minor role in ripening.

The level of lactose in Swiss cheese decreases rapidly during the latter stages of manufacture and early in ripening, reaching very low levels after 24 h. Soon after manufacture, Emmental cheese typically contains ~1.7% lactose which is metabolized rapidly by *Sc. thermophilus* to L-lactate; after 24 h, the level of lactose in Swiss cheese is usually very low (Figure 4). However, *Sc. thermophilus* metabolizes only the glucose moiety of lactose and thus galactose levels in the curd increase initially, reaching a maximum *c.* 10 h after manufacture; galactose is then metabolized by the lactobacilli to D/L-lactate. It is important to use a Gal^+ *Lactobacillus* sp. in the starter mix or galactose will accumulate in the curd, the metabolism of which by NSLAB can lead to off-flavors. Metabolism of galactose is usually complete after 14 days. During the early stages of ripening, there is significantly more L-lactate than D-lactate in the curd.

Lactate metabolism is more important in Swiss cheese than in any other variety, with the exception of surface mold-ripened cheeses such as Camembert and Brie. *Pr. freudenreichii* added to, or contaminating, the milk for Swiss-type cheeses grow once the cheeses are transferred to the warm room after brining and metabolize L-lactate during ripening *via* a complex pathway to propionate, acetate, CO_2 and H_2O (Figure 5). The pathway for lactate metabolism by *Propionibacterium freudenreichii* (Figure 6) involves two separate cycles, one producing propionate and the other acetate. In general, propionic acid bacteria are able to metabolize both L- and D-lactate, but usually, they metabolize the former preferentially to CO_2, and so levels of L-lactate decrease with a concomitant increase in the levels of propionate and acetate (Figure 4). Later in ripening, the propionic acid bacteria utilize D-lactate, at which point, the concentration begins to decline. Carbon dioxide produced is essential for eye development, and the propionate and, to a lesser extent, acetate contribute to the flavor of these cheeses. Eyes are a characteristic sensory feature of Swiss-type cheeses and are a parameter used by the consumer to identify these cheeses.

An undesirable pathway for lactate metabolism in Swiss and many other cheeses is its anaerobic fermentation to butyrate and H_2 by *Clostridium* spp., resulting in a serious defect known as 'late gas blowing.' Spores of clostridia, which survive pasteurization of milk and cooking of Swiss cheese, may be removed from milk by bactofugation or microfiltration. Other approaches to avoid this problem include the use of lysozyme or NO_3^- (which degrade the cell

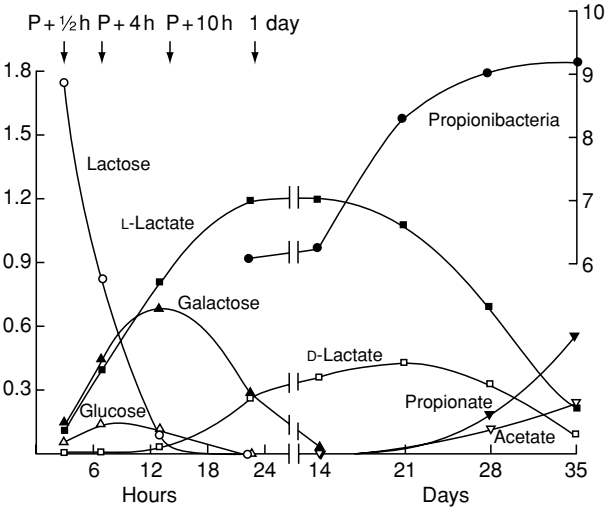

Figure 4 Lactose and lactate metabolism in Swiss-type cheese manufactured using *Streptococcus thermophilus* and a galactose-positive strain of *Lactobacillus helveticus* as starters. Lactose (○), glucose (△), galactose (▲), D-lactate (□), L-lactate (■), acetate (▽), propionate (▼) and numbers of propionic acid bacteria (●). Reprinted with permission from P.F. Fox, T.P. Guinee, T.M. Cogan and P.L.H. McSweeney. Fundamentals of Cheese Science, © 2000, Aspen Publishers, Gaithersburg, MD.

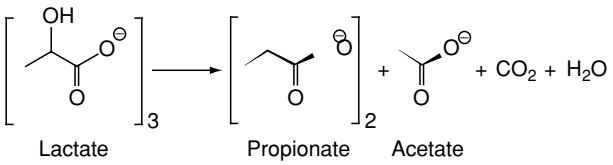

Figure 5 Overall equation for the metabolism of lactate by *Propionibacterium freudenreichii* in Swiss-type cheeses during ripening. Reprinted with permission from P.F. Fox, T.P. Guinee, T.M. Cogan and P.L.H. McSweeney. Fundamentals of Cheese Science, © 2000, Aspen Publishers, Gaithersburg, MD.

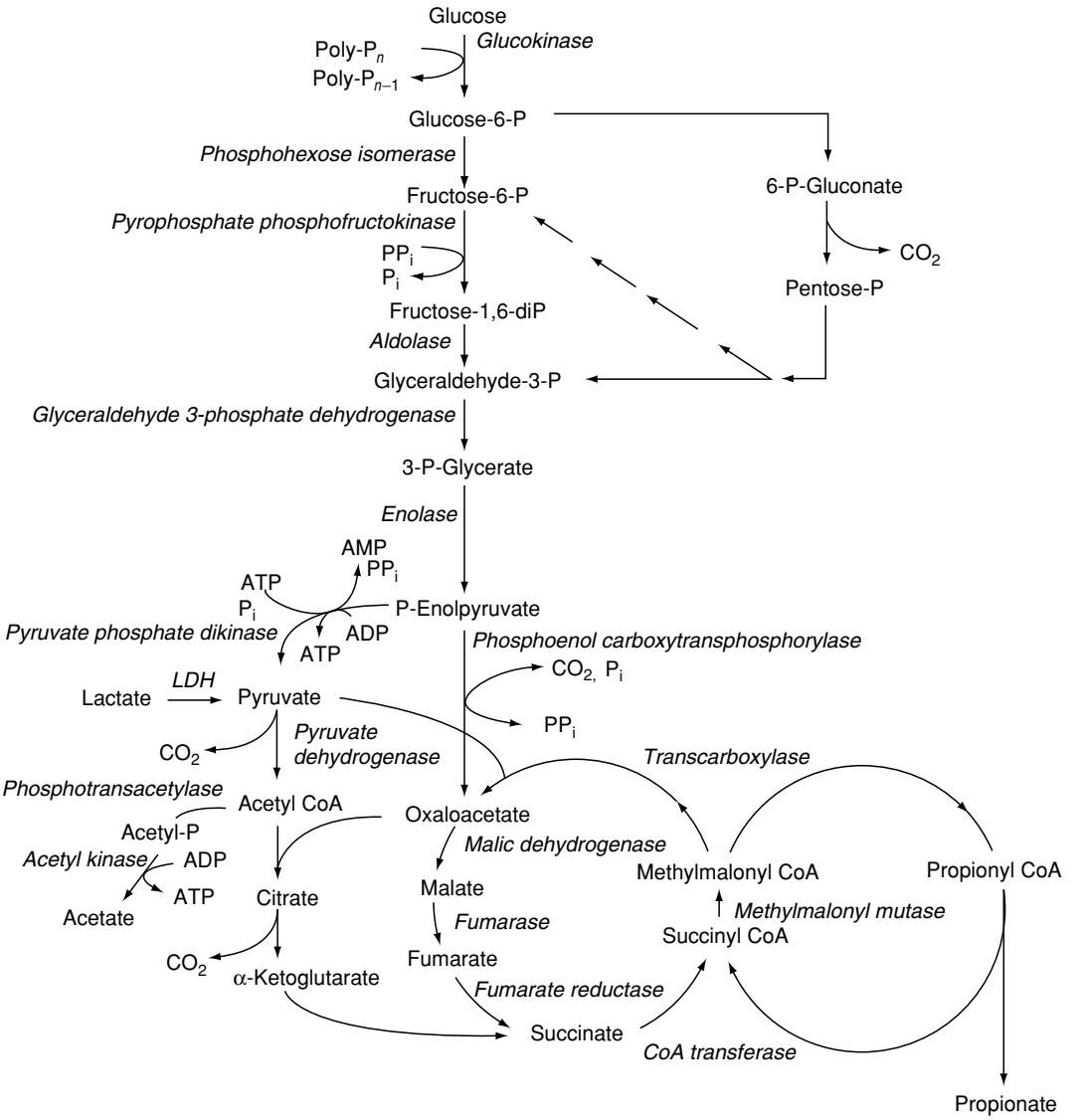

Figure 6 Cycles involved in the propionic acid fermentation by *Propionibacterium* sp. (LDH, lactate dehydrogenase; Poly-PN, polyphosphate; Ppi, pyrophosphate.) For reasons of clarity, only the pyrophosphate-dependent conversion of fructose-6-P to fructose-1,6-diP is shown, and the generation of ATP by the electron transfer system is omitted. All reactions are directed towards production of propionate, even though the reactions are reversible. Reprinted with permission from P.F. Fox, T.P. Guinee, T.M. Cogan and P.L.H. McSweeney. Fundamentals of Cheese Science, © 2000, Aspen Publishers, Gaithersburg, MD.

wall of Gram-positive bacteria and inhibit germination of spores, respectively) and good hygiene. In particular, feeding silage to cows, whose milk will be used in the manufacture of Swiss cheese, is forbidden in some regions.

Because of the absence of strongly lipolytic agents, lipolysis is quite limited in Swiss cheese. However, lipolysis does contribute to the flavor of Swiss cheese. Short-chain fatty acids contribute directly to cheese flavor, and low levels are desirable in Swiss cheese. However, high levels of free fatty acids give an undesirable, rancid off-flavor to Swiss cheese. Fatty acids also act as precursors for a range of further reactions (e.g., the formation of fatty acid lactones, hydroxyacids, esters, thioesters, and methyl ketones) that contribute to cheese flavor. Lipolysis in Swiss cheese is catalyzed by the indigenous milk lipoprotein lipase (LPL; particularly in the case of raw milk cheese, since this enzyme is extensively inactivated by pasteurization). The action of LPL is probably less important in raw milk Swiss cheese than in other raw milk varieties, since it is partially inactivated during the cooking step. Other sources of lipases and esterases in Swiss cheese are the starter and NSLAB; *Pr. freudenreichii* has also been shown to possess a lipase.

Proteolysis is a major series of events during the ripening of Swiss cheese. Proteolysis in Swiss cheese is catalyzed by enzymes indigenous to the milk and from the cheese microflora and, to a much lesser extent, from the coagulant. The principal indigenous proteinase in milk, plasmin, is particularly important for proteolysis in Swiss-type cheese. Plasmin (fibrinolysin, EC 3.4.21.7) is a serine proteinase with an alkaline pH optimum that is formed from an inactive precursor, plasminogen, by limited proteolysis. The activation of plasminogen and plasmin activity in milk are controlled by a system of plasminogen activators, plasmin inhibitors, and inhibitors of plasminogen activators. Plasmin is a particularly heat-stable proteinase, and hence its activity relative to other proteolytic enzymes (especially chymosin) is higher in Swiss cheese than in other varieties, since it survives the cooking step. Plasmin acts during ripening to degrade β-casein to γ-caseins; the latter accumulate to greater levels in Swiss cheese than in most other varieties. Chymosin, or other enzymes from the coagulant, are the most important enzymes for the initial degradation of the caseins during the ripening of most internal, bacterially ripened varieties. However, since chymosin is denatured extensively during the cooking step of Swiss cheese, its action is less important in this variety than in varieties such as Cheddar.

Intermediate-sized peptides produced by the action of plasmin or chymosin are hydrolyzed by proteinases and peptidases from the cheese microflora. Lactic acid bacteria are auxotrophic for a number of amino acids and therefore possess a range of proteolytic enzymes capable of liberating amino acids from the caseins. Although less well studied than the lactococci, thermophilic lactobacilli and streptococci used as starters in Swiss cheese possess generally similar proteinase and peptidase systems, including a cell envelope-associated proteinase that is responsible for degrading many of the intermediate-sized peptides produced by plasmin and chymosin to shorter peptides. Lactic acid bacteria, including those used as starters in Swiss cheese, possess a wide range of intracellular peptidases. These enzymes are released into the cheese matrix following cell lysis, which occurs during ripening, and are essential for the liberation of amino acids (AA). *Pr. freudenreichii* is weakly proteolytic, although it does possess intracellular peptidases that contribute to ripening.

Proteolysis results ultimately in the liberation of AA. AA contribute directly to the background flavor of cheese; some are sweet, some are sour, and many are bitter. However, the major role of AA in the development of cheese flavor is now thought to be as precursor molecules for a range of poorly understood catabolic reactions that lead to the formation of many volatile flavor compounds. Although amino acid catabolism has not been studied extensively in Swiss cheese, it is highly likely that it follows the same general pathways as in other internal, bacterially ripened varieties such as Cheddar and Gouda. Amino acid catabolic reactions include deamination (to yield NH_3 and α-keto acids), decarboxylation (to yield amines), formation of alkyl pyrazines, and catabolism of sulfur-containing and aromatic amino acids.

Other Swiss-type Cheeses

Although Emmental is the Swiss-type variety produced in the largest amount world-wide, there are several related Swiss cheese varieties. Gruyère is a popular Swiss-type cheese that differs from Emmental in being smaller, with a somewhat stronger flavor and fewer eyes, and is characterized by the development of a surface flora. This surface flora (similar to that which develops on smear-ripened varieties such as Tilsit and Limburger) is encouraged by ripening for 2–3 weeks at 10 °C and for 2–3 months at 15–20 °C and 90–95% ERH. During the hot-room stage, the cheeses are rubbed with a brine-soaked cloth. Further ripening at 12–15 °C is required before retail; the cheese is ripe in 8–12 months. Varieties similar to Gruyère include Raclette (which is manufactured from raw milk and is acidified by the natural flora of the milk) and Gruyère de Comté, which is produced in eastern France. Comté has smaller eyes than Emmental as a consequence of a low ripening temperature (*c.* 18 °C). The surface of this cheese is covered by an orange smear ('*morge*'); cell numbers at the surface are typically 10^{10} cfu cm^{-2}, and the microflora is very complex, consisting of coryneform bacteria, micrococci, and yeasts. Beaufort is a French variety that is similar to, but larger than (*c.* 45 kg), Gruyère and usually develops fewer eyes (as a result of the growth of mesophilic starter acting on citrate at a low ripening temperature) than Gruyère. Appenzeller, which originated in Switzerland, is a small cheese (*c.* 30 cm) with a soft texture and propionic acid fermentation that results in the development of a few eyes. The curds are cooked at 43–45 °C. Appenzeller is immersed in cider or spiced wine or coated with a mixture of salt and spices during ripening, which imparts a distinctive flavor to the cheese. Maasdammer, a variety developed recently in The Netherlands, is characterized by the use of a mesophilic starter and extensive propionic acid fermentation, which gives large eyes and a domed appearance to the cheese wheels. Jarlsberg is a Swiss-type cheese produced in Scandinavia, which is characterized by large eyes.

See also: **Cheeses**: Chemistry and Microbiology of Maturation; Dutch-type Cheeses

Further Reading

Christensen JE, Dudley EG, Pederson JA and Steele JA (1999) Peptidases and amino acid catabolism in lactic acid bacteria. *Antonie van Leeuwenhoek* 76: 217–246.

Fox PF (ed.) (1993) *Cheese: Chemistry, Physics and Microbiology*, 2nd edn. vols 1 and 2, London: Chapman & Hall.

Fox PF, Guinee TP, Cogan TM and McSweeney PLH (2000) *Fundamentals of Cheese Science*. Gaithersburg, MD: Aspen.

Fox PF, Lucey JA and Cogan TM (1990) Glycolysis and related reactions during cheese manufacture and ripening. *Critical Reviews in Food Science and Nutrition* 29: 237–253.

Fox PF and McSweeney PLH (1996) Proteolysis in cheese during ripening. *Food Reviews International* 12: 457–509.

Kosikowski FV and Mistry VV (1997) *Cheese and Fermented Milk Foods*, 3rd edn. vols 1 and 2, Westport, CT: F.V. Kosikowski.

Kunji ERS, Mierau I, Hagting A, Poolman B and Konings WN (1996) The proteolytic system of lactic acid bacteria. *Antonie van Leeuwenhoek* 70: 187–221.

McSweeney PLH and Sousa MJ (2000) Biochemical pathways for the production of flavour compounds in cheeses during ripening: A review. *Lait* 80: 293–324.

Polychroniadou A (2001) Eyes in cheese: a concise review. *Milchwissenschaft* 56: 74–77.

Robinson RK and Wilbey RA (1998) *Cheesemaking Practice. R. Scott*. Gaithersburg, MD: Aspen.

Steffen C, Eberhard P, Bosset JO and Rüegg M (1993) Swiss-type varieties. In: Fox PF (ed.) *Cheese: Chemistry, Physics and Microbiology*, 2nd edn. vol. 2, pp. 83–110. London: Chapman & Hall.

Turner KW, Morris HA and Martley FG (1983) Swiss-type cheese II. The role of thermophilic lactobacilli in sugar fermentation. *New Zealand Journal of Dairy Science and Technology* 18: 117–123.

Soft and Special Varieties

R C Chandan, New Brighton, MN, USA

Introduction

This article contains terse description of cottage cheese, cream cheese, and cheese varieties commonly used in South Asia and Latin America. The typical composition of these products, compared with fresh Cheddar cheese, is shown in **Table 1**.

Table 1 Composition (%) of some soft cheeses

Cheese	*Moisture*	*Protein*	*Fat*	*Lactose*
Cottage cheese curd	79.8	17.3	0.4	1.9
Creamed cottage cheese	79.0	12.5	4.5	2.7
Neufchatel cheese	62.9	10.1	26.1	3.3
Cream cheese	53.8	7.6	34.9	2.7
Ricotta (whole milk)	71.7	11.3	12.9	3.1
Chhana – cow milk	53.4	17.4	24.8	2.6
Paneer – buffalo milk	51.3	17.3	27.0	2.6
Queso Blanco	55.0	23.0	15.0	2.7
Cheddar cheese curd	37.0	25.0	32.2	1.3

Cheddar is included for comparison.
Sources: USDA (1998). Nutrient Database for Standard Reference, Release 12, Nutrient Data Laboratory. www.nal.usda.gov/fnic/foodcomp, and Kosikowski FV and Mistry VV (1997). *Cheese and Fermented Milk Foods*, 3rd edn. Westport, CT: FV Kosikowski LLC.

Cottage Cheese

Cottage cheese belongs to the class of natural, unripened, soft cheeses. Since it has a significantly lower fat content than most cheeses and is a good source of protein, it is a popular part of low-calorie diets. This cheese is made from pasteurized skim milk, concentrated nonfat milk or reconstituted nonfat dry milk. Coagulation of skim milk is accomplished by fermentation with *Lactococcus lactis* ssp. *lactis* or *Lc. lactis* ssp. *cremoris*. Alternatively, food-grade acids may be used to set the milk, but the product must be labeled 'Direct Acid set.' It is marketed as a small-curd (< 4 cm diameter) or large-curd (> 8 cm diameter) product.

Cottage cheese is a major dairy product in North America. In 2000, total sales of cottage cheese in the US were 333.8 million kilograms, with average per caput sales of 1.18 kg. The market value exceeds one billion US dollars. The US Food and Drug Administration standards define cottage cheese as the product containing moisture not exceeding 80% and milk fat not less than 4%. Cottage cheese may be labeled as reduced fat if it contains at least 25% less fat than creamed cottage cheese. To qualify for a low-fat label, the product must contain not more than 3 g of fat per serving of 28.35 g. Fat-free or nonfat cheese must contain less than 0.5 g of fat perserving of 28.35 g. Certain preservatives, including sorbates, are allowed to extend the shelf-life of Cottage cheese. Cottage cheese may contain fruits, seafood, meats, or vegetables.

Cottage cheese is a major ingredient of salads and is used as a filling in the preparation of breakfast food, blintze. Good-quality cottage cheese should have a clean, creamy, cultured milk flavor, a natural, creamy color, and a meaty, soft texture. It should have a reasonably uniform curd size and should not be pasty or too firm. It should have a uniform cream layer around the curd particles with a little free cream.

Manufacturing Process

Depending on the plant schedule, a short-set, medium-set, or long-set method may be used. The long-set method involves using 1% starter, incubation at 22 °C for 12–16 h, whereas the short-set method uses 5–7% starter at 32 °C for 4–6 h.

The manufacturing operation may involve the use of simple cheese vat and equipment for whey removal, curd washing, draining, and creaming. The creamed curd is pumped to a filler for packaging. However, large-scale operations use automated systems to continuously carry out the various steps outlined below.

Setting Raw skim milk is pasteurized at 73 °C for 16.5 s. It is desirable to avoid previously heated milk because a minimal heat treatment is needed to avoid the formation of weak coagulum and subsequent shattering of the curd. The pasteurized milk is cooled to 32 °C and pumped into a cheese vat, where the jacket temperature is adjusted to maintain the milk temperature at 32 °C. Bulk culture is added at the level of 5–7% and mixed thoroughly. The titratable acidity is measured at this point. Stirring is continued at intervals of 30 min for 1.5 h. The increase in acidity at this point should be 0.05–0.07%. If not, 1% more culture is added for each 0.01% increment below 0.05%. Next, rennet, diluted 1 to 40 with water, is added at a level of 40 ml for a large curd and 12–20 ml for a small curd per 454 kg of skim milk. After mixing thoroughly, the vat is covered for about 2 h for fermentation, during which time the pH should drop to 4.6–4.7.

Cutting When the coagulum is ready, it is cut first along the length of the vat using appropriate horizontal knives to yield the desired small- or large-curd product. Wire knives with a 0.62-cm spacing are used for the small-curd product and with 1.86 cm spacing for the large-curd product. Next, the coagulum is cut lengthwise with the vertical knife and then crosswise with the vertical knife.

Cooking After cutting, the curd is allowed to heal for 10 min. The jacket water temperature is raised to 44–50 °C to initiate cooking of the curds. The temperature of curd is raised slowly to 54 °C over 1.5 h. The curd-whey is hand-stirred every 5–10 min, and a mechanical agitator is used after the temperature has reached 38–41 °C. The agitation speed is increased slowly to avoid matting of curd particles. Small curd particles cook faster than large particles. Agitation is continued until the curds become reasonably firm; the curd is adequately cooked when a handful of water-chilled curd springs apart when squeezed together with moderate pressure.

Washing After completion of cooking, the jacket water is drained off. Whey is removed until the curd begins to appear above the surface. Wash water, acidified with phosphoric acid to pH 4.6 and chlorinated to 10–20 mg kg^{-1} is then added to a volume equal to that of the whey removed. The temperature of curd should decrease to 27–30 °C. After agitation for 10 min, the water is drained off. Washing is repeated twice with chilled water to reduce the temperature of the curd to 13–16 °C and finally to 5 °C or lower. The curd is trenched and allowed to drain for 30–60 min before creaming.

Creaming The curd is blended with cream dressing in a blender or in the cheese vat. The dressing is formulated to contain 12.5% fat, 8.5% solids-not-fat (SNF), 2.7% salt, and 0.25% stabilizer. The stabilizer is especially designed for the dressing to achieve a high viscosity and to avoid wheying off. The dressing is pasteurized at 75–77 °C for 30 min and homogenized at 57 °C at a pressure of 13.8 MPa, single stage, followed by cooling to 4 °C or below. One part of the dressing is blended with two parts of curd to give at least 4% fat in creamed cottage cheese. The dressing can be treated further to enhance the shelf-life of cottage cheese to control the growth of spoilage psychrotrophic organisms. Cultures inhibitory to spoilage organisms may be added and the dressing itself may be fermented. Several bacteriocin-containing preservatives are commercially available to extend the shelf-life of cottage cheese. The typical shelf-life of cottage cheese packaged in moisture-barrier containers is 3–4 weeks at 4 °C.

The yield of cottage cheese curd is typically around 15.5% kg per 100 kg of skim milk of 9% SNF. A fortified skim milk containing 12% SNF should give 21.6 kg of curd per 100 kg of starting material.

Quarg, also called Quark, Kwarg, Koarg, Twarog, Tvorog, Taho, or Topfen, is a popular fresh cheese of central Europe. It has a slightly sour flavor owing to its low pH. It is generally made from skim milk. In some cases, some variations may be made from homogenized full-fat milk. Buttermilk is also used as a raw material. In a process similar to cottage cheese, the milk is cultured with selected strains of *Lactococcus* and *Leuconostoc* species. After fermentation, the curd is collected by centrifugation in a special separator. It may be also be produced by direct acidification with lactic acid, lemon juice, or vinegar. The curd is blended with cream to yield low-fat or full-fat Quarg.

Cream Cheese

Cream cheese contains at least 33% fat and not more than 55% moisture. It is a soft, unripened lactic acid-coagulated cheese, made by a process similar to that for cottage cheese. It has a mild, acid, and creamy flavor with a buttery aroma. The manufacturing procedures are similar for cream and Neufchatel cheese, but, compared with a fat content of 35% in cream cheese, Neufchatel cheese contains 23% fat.

Cream cheese is an important variety in North America. In 2000, cream and Neufchatel cheese production in the USA exceeded 331.8 million kilograms. Cream cheese is mostly used as a spread on bagels and toasted bread or as an ingredient of cheesecakes. At present, cream cheese is marketed in flavors like strawberry and other fruits. It is also flavored with vegetables, condiments, spices, and herbs. Low-fat versions are available.

Manufacturing Procedure

Several processes are used to produce cream cheese. Typically, the process involves standardizing cream to 11% fat, followed by pasteurization at time–temperature combinations (e.g., 68 °C/30 min) similar to normal pasteurization. Using a starter consisting of *Lactococcus lactis* ssp. *lactis/cremoris* and *Leuconostoc cremoris* at a level of 5–6%, the cream is cultured at 30–32 °C for 5 h. A long-set process uses 1% starter and incubation for 16 h at 22 °C. *Leuconostoc* sp. imparts a buttery aroma. In some plants, rennet is added at a low level. Fermentation is continued until a pH of 4.4–4.5 is attained.

In the traditional process, the cultured cream is drained in muslin bags by hanging overnight in a cold room. However, most modern plants are equipped with centrifuges to remove the whey. Cream cheese collects in the bowl. Membrane processes are also used for separation of whey. The cheese is subsequently packed by a cold-pack procedure without additives or heat. Cold-pack cream cheese has more aroma and flavor and lacks a pasty/sticky body than the hot-packed product. However, it lacks the long shelf-life obtained by the more commonly used hot-pack procedure. This procedure involves heat treatment of cream cheese in a kettle or scraped surface heat exchanger. Cream cheese is mixed mechanically with a stabilizer (0.35% locust bean gum) and 1% salt and brought to 70 °C. The hot mixture is homogenized at 13.8 MPa and transferred to the packaging machine for hot packaging. The hot process imparts a shelf-life of at least 2 months under refrigerated storage.

Cheeses Produced by Direct Addition of Acid and Heat

Worldwide, curd formation during cheese making is effected by the use of coagulating enzymes isolated from various sources in combination with or without a fermentation step. Alternatively, curd can be formed by direct acidification of hot milk with food-grade acids. Generally, the directly acidified cheeses are consumed in fresh, unripened form. The cheese varieties included in this category are quite diverse. From the standpoint of commercial importance, Ricotta, Chhana, Paneer, and Latin American cheeses will be discussed briefly.

Ricotta Cheese

Like cottage cheese, Ricotta cheese is high-moisture, unpressed cheese. Its composition varies, depending on whether it is made exclusively from whey or from a blend of whey and milk. Ricotta is also prepared from whole milk, with no addition of whey. When made from a blend of 95% sweet whey and 5% milk, Ricotta contains 68–73% moisture, 16% protein, 4–10% fat, and 4% lactose.

Ricotta has a bland to slightly cooked, but pleasing, flavor. Its texture is soft and creamy. It is consumed as such as a spread and may be used as a replacement for cream cheese or sour cream in dips. It is basically a nonmelting cheese. Its major use is in Italian cuisine, as in Lasagna and Ravioli, for example. (*See* **Whey and Whey Powders**: Production and Uses.)

Manufacturing Procedure Most Ricotta is produced by a batch process. Traditionally, open kettles are used. Heating may be direct, or steam jackets may provide heat transfer. Sweet whey from Italian cheese manufacture is suitable as long as its pH is 6.2 or higher. Ten to 25% milk is often blended to neutralize acid in the whey, enhance yield, and curd cohesiveness. The mixture is heated in a kettle to 82–93 °C, followed by the addition of sufficient food-grade acid, e.g., lactic, acetic, or citric, to reduce the pH of the mixture to 5.9–6.1. In some plants, cultured milk/whey may be used as a source of lactic acid. This pH range is crucial to maintain the sweet flavor of the cheese. As a result of heat and acidity, the proteins are denatured, and a foam-type curd ascends to the surface. The curd is then dipped with a perforated ladle and collected in muslin bags. The bags are allowed to drip and cool in a cold room. Alternatively, the curds are drained in perforated stainless hoops and allowed to dry. The curd is soft, fragile, and grainy, and may be pressed slightly to achieve cohesiveness. The yield of Ricotta is low, around 5–6% if no milk is mixed

with whey. Mechanized production for the continuous manufacture of Ricotta cheese involves adjustment of the pH of the whey–milk mixture to 6.9–7.1 with NaOH, heating the blend to 88–92 °C, and injecting in-line an appropriate quantity of salt and acid. Again, the target pH is 5.9–6.1. The hot acidified whey–milk blend is pumped into the bottom of a V-shaped vat, and the resulting curd is collected mechanically from the top of the vat into nylon mesh for the curd to drain. It is than transferred to perforated hoops, cooled, and packaged for sale.

Chhana

Chhana is a Ricotta-like cheese produced in India from hot milk by direct acidification. Although Chhana is an intermediate material for producing Paneer and various confectionery foods, the product standards for Chhana and Paneer are identical. According to Indian Pure Foods Act, Chhana (or Paneer) is defined as a product obtained from cow or buffalo milk or a combination thereof by precipitation with sour milk, lactic acid, or citric acid; it should contain not more than 70% moisture, and the milk fat content should not be less than 50% of the dry matter. Skim milk Chhana or skim milk Paneer is the product obtained from cow or buffalo skim milk by precipitation with sour milk, lactic acid, or citric acid; it should contain not more than 70% moisture, and the milk fat content of the product should not exceed 13.0% of the dry matter.

Manufacturing method The production of Chhana basically involves the precipitation of casein along with fat and entrapped water-soluble components of milk (lactose, whey proteins, minerals, vitamins) by the addition of an acidulant to milk at a near-boiling temperature, followed by removal of whey from the curd. Basically, Chhana is made by a batch process, but a continuous process may be adapted from the continuous process for Ricotta, as described above.

The texture of Chhana varies from smooth and pasty to crumbly. It is an intermediate starting material for the preparation of indigenous Indian dairy confectionery products. The preferred milk for Chhana is cows' milk. In Chhana production, pressing of the curd is restricted to pressing caused by gravity after draining of the whey in a cheesecloth/muslin bag. A longer draining time yields a hard product. Depending on its end use, Chhana may be soft or hard. For example, the confectionery product, Rasogolla, requires soft Chhana, whereas the confection, Sandesh, requires the hard variety. These two confections utilize the majority of Chhana production, and optimum functionality is required in each case.

Chhana is made in kettles. Cows' milk is brought to near boiling, and an appropriate acidulant is added quickly to the hot milk to bring the pH of the mixture to 5.4. The curd settles to the bottom, and whey is removed by filtration through a strainer lined with cheesecloth. The curd is then cooled in running tap water. Chhana can be used to make sweetmeats and other Indian confections. The coagulants used are lime or lemon juice, vinegar, citric acid, lactic acid, fermented milk, or whey. A solution of citric acid (0.5–1.5%) or lactic acid (1–2%) is appropriate to lower the pH of hot milk to 5.4. An interaction between the temperature of coagulation and final pH is observed in terms of obtaining the best yield and quality of Chhana. For cows' milk, a temperature of 80 °C at pH 5.4 is optimum, whereas for buffalo milk, 70 °C at pH 5.7 is desirable. A higher coagulation temperature imparts graininess and hardness to the texture, whereas lower temperatures result in a sticky Chhana, which drains too slowly. Chhana is an extremely perishable food product, and at ambient temperature, its quality deteriorates within a day or two and is used soon after production. Under refrigeration, the shelf-life is extended to 6 days.

Two important confections based on Chhana are Sandesh and Rassogolla.

Preparation of Sandesh Chhana is thoroughly kneaded to obtain a homogeneous mass. Ground sugar (30 parts) is added to Chhana (100 parts), and the mixture is cooked in a vessel until the confection stops sticking to the surface of a griddle. At this point, the hot mass is poured into molds or into a pan to cool and set. Cardamom and other flavors may be added prior to setting. Upon cooling, the surface of the confection may be sprinkled with cracked pistachio nuts and cut into the desired shapes and sizes. The typical composition of Sandesh is 24% moisture, 0.45% titratable acidity, 14.9% protein, 10.7% fat, and 47% sugar.

Preparation of Rasogolla Rasogolla, a widely popular sweetmeat in India, is the result of a puffing phenomenon in which milk protein, primarily the caseins, acquires a foam-type structure under special cooking conditions in sugar syrup. Accordingly, the product is a spongy, round ball, 2–4 cm in diameter, suspended in sugar syrup. Occasionally, the shape may be cylindrical. In this process, Rasogolla absorbs sugar syrup, providing a succulent, juicy texture. The ultrastructure of the product is stable for years, as evidenced in canned products, and the spongy texture may even be accentuated to give a variety called Sponge Rasogolla.

Most large manufacturers of Rasogolla in India prefer to use only Chhana to knead and prepare the dough. However, some sweetmakers knead together 100 g of Chhana, 0.7 g of cream of wheat, 5.2 g of white wheat flour, and 12 mg of baking soda. The dough is portioned and rolled into balls (10–20 g each) and cooked in a medium consisting of 50% sugar, 17% whey, and 33% water. The pH of the cooking solution is adjusted to 6.8 with $Ca(OH)_2$. Cooking is completed in 25 min. The cooked balls are conditioned in water at 30–35 °C for 10 min, immersed initially in 59–60 ° Brix syrup and finally in 50 ° Brix syrup, and then refrigerated. They may be canned to enhance their shelf-life and to extend marketing range. Rasogolla contains 38–55% moisture, 37–48% sugar, 5.3–5.7% fat, and 5.0–5.5% protein.

Paneer

It is estimated that about 4–5% of the total milk produced in India is converted into Paneer. The distinctive texture of Paneer is obtained by the application of low pressure to the cheese in hoops, followed by immersion of the blocks in chilled water. The cooling step results in reabsorption of some water, which imparts a characteristic springy and rubbery body to Paneer blocks. Paneer offers outstanding nonmelt functionality. Its ability to withstand frying temperatures makes it suitable for use in Indian cuisine. Paneer is suitable for the preparation of fried cheese snacks. It may be included as an entrée item in a meal, with or without a suitable batter and breadcrumbs. Such items resemble fish sticks, chicken nuggets, or tofu chunks.

Manufacturing Method Buffalo milk containing 6% fat and 9.5% SNF is preferred for Paneer production. The milk is heated in a plate heat exchanger to 85–90 °C and pumped into a water/steam jacketed cheese vat. By circulating cold water in the jacket, milk is cooled to 70–75 °C in about 10 min prior to acid blending. Sufficient citric acid (1% solution), warmed to 70 °C, is gently but quickly blended with the milk (within 1 min) to reduce the pH to 5.3–5.4. Normally, it requires about 1.8–2.0 g of citric acid to coagulate 1 kg of milk. At this point, clear greenish yellow whey separates, and the curd sinks to the bottom. The coagulation temperature influences the moisture content of Paneer; a temperature of 70 °C gives the best organoleptic quality as well as the most desirable frying quality in terms of shape retention, softness, and integrity maintenance. The coagulated milk is allowed to separate into curd and clear whey. After 10 min, the curd chunks sink to the bottom. At this point, a strainer is fitted into the outlet of the vat. The whey drainage valve is then opened, and the whey flows into a surge tank partitioned by a strainer to retain any curd fines that escape from the vat. Hot whey accumulates in the second section of the surge tank from which it is pumped out, cooled to 4 °C, and stored in a whey tank for further processing. The curd is transferred to molds/hoops lined with cheesecloth. For culinary purposes, unflavored and unsalted Paneer is traditionally used. The paneer hoops are pressed with a hydraulic press to exert a low pressure (10 kPa) on the curd. The pressing time is generally 10–15 min, after which time, the wrapped block of paneer is ejected from the hoop for quick cooling by immersion for 1–2 h in cold water (4 °C) containing 3–5 mg per kilogram of chlorine. Cooled blocks are then mechanically cut into convenient retail size and packaged. A yield of 18–20 kg of Paneer per 100 kg of standardized milk is expected. (*See* **Buffalo**: Milk.)

Latin American Cheeses

Latin American White (LAW) cheeses, known by various names, are produced throughout Mexico, Central and South America, and the Caribbean Islands. These cheeses are made from whole milk, skim milk, cream, or a mixture. Queso blanco and Queso del pais are the only varieties produced by direct acidification of hot milk with vinegar or fruit juice. Rennet is used for the coagulation in Queso de prensas in El Salvador, Mexico, and Venezuela, Queso estera in Colombia, and queijo Minas in Brazil. Starter cultures are also used and allowed to generate a high acidity. LAW cheeses are white, creamy, highly salted, and acidic in flavor. They have a body and texture similar to young high-moisture Cheddar, and can be sliced for sandwich use. The fresh or dry 'Criollo' product is grated or macerated to accompany beans, rice, or tortilla (flat corn bread). These cheeses can be used as a snack, in salads and as a cooking cheese in casserole dishes, grated for use in pizza and other foods, or included as an ingredient in the manufacture of processed cheese. Like paneer, LAW cheese can be made to be resistant to melting at frying temperatures. However, these are highly salted products, unlike paneer.

Manufacturing method Directly acidified Queso blanco is produced from whole/part skim milk by heating to 80–88 °C, holding for 15 min, followed by quick addition of diluted vinegar/citric acid solution to bring the pH to 5.3–5.7. The acidified milk is stirred gently for 2–3 min, and the curd is allowed to settle to the bottom of the vat. As for paneer, both casein and heat-denatured whey proteins are retained in the curd. After 15 min, the settled curd is recovered

by whey drainage, and up to 5% salt is mixed in. The salted curd is soft and dry, and forms large chunks; it is pressed overnight in molds to yield a more cohesive product. The blocks are refrigerated and packaged.

The yield of Queso blanco varies from 11 to 14.6%, depending on the moisture content. It typically contains 51–57% moisture, 18–22% protein, 15–17% fat, and 3.5–4.5% salt. A drier variety suitable for maturation may have 35–39% moisture, and a hard variety may have an even lower moisture (27–32%).

See also: **Acids**: Properties and Determination; Natural Acids and Acidulants; **Cheeses**: Types of Cheese; Starter Cultures Employed in Cheese-making; Quarg and Fromage Frais; Dietary Importance

Further Reading

Aneja RP (1992) Traditional delicacies: A compendium. In: Gupta PR (ed.) *Dairy India*. Edited by Delhi: Dairy India Yearbook.

Chandan RC (1982) Fermented dairy products. In: Reed G (ed.) *Prescott & Dunn's Industrial Microbiology*, 4th edn. Wetport, CT: AVI.

Chandan RC (1991) Cheeses made by direct acidification. In: Robinson RK and Tamine AY (eds) *Feta and Related Cheeses*. London: Ellis Horwood.

Chandan RC (1997) *Dairy-based Ingredients*. St. Paul, MN: Eagan Press.

Emmons DB and Tuckey SL (1967) *Cottage Cheese and Other Cultured Milk Products. Pfizer Cheese Monographs*, vol. III. New York: Chas. Pfizer & Co.

Kosikowski FV and Mistry VV (1997) *Cheese and Fermented Milk Foods*, 3rd edn. Westport, CT: FV Kosikowski LLC.

Mathur BN, Rao KVSS and Zanzad PN (1993) Modern trends in paneer manufacture. In: Potty VH and Mulky MJ (eds) *Food Processing*. New Delhi: Oxford and IBH Pulishing Co.

Torres N and Chandan RC (1981) Latin American White cheese – A review. *Journal of Dairy Science* 64: 552–558.

White Brined Varieties

M H Abd El-Salam, National Research Centre, Dokki, Cairo, Egypt

Introduction

It is evident from ancient literature (Homer, 1184 BC) and archeological findings (tomb Hor Aha, Saqqara, Egypt) that this group of cheeses is probably the origin of various types of cheeses known now.

The production of brined (pickled) cheeses has long been limited to the Mediterranean and southeast European Countries. Nowadays, the international recognition of brined cheeses and the dynamic growth in their demand have extended their production to other countries, including Denmark, Ireland, the UK, the USA, New Zealand, and Australia. Feta cheese is the most significant type of brined cheese marketed in the world. The consumption of brined cheese varies between countries, but the highest per- capita consumption of Feta cheese is recorded in Greece (12 kg).

Characteristic Properties

Preservation in brine (pickle) is the main characteristic of this group of cheeses. The brine serves to preserve the cheese and prevent it drying out. Consequently, the composition and properties of these cheeses and those of the brine are interrelated.

Brined cheeses are generally rindless and are produced in the form of blocks of various shapes and sizes (cubes, bricks, or segments) covered in salt brine or whey.

Brined cheeses are either ripened in wooden barrels and then marketed in tins or packed and ripened directly in retail packages. They are marketed fresh or after storage of up to one year.

Their clean, acidic and salty taste, and sharp flavor characterize brined cheeses. The high salt content of the cheese and brine allows for a good keeping quality in hot climates without the need for cold storage.

The color of brined cheeses is pure white, as they are traditionally made from ewes', goats' or buffaloes' milk, which are known for their whitish color.

Although all brined cheeses are subjected to the same practice of storage in pickle for extended periods, they differ significantly in their manufacturing technologies, composition, and properties.

Classification

As shown in **Table 1**, brined cheeses include soft and semihard varieties, and nearly all of them are made by rennet coagulation. In general, brined cheeses can be made from different kinds of milk, but sheeps' milk is preferred for most of these varieties. (*See* **Buffalo**: Milk; **Goat**: Milk; **Milk**: Liquid Milk for the Consumer; **Sheep**: Milk.)

Methods of Manufacture

Traditional Methods

Feta cheese Normally, ewes' milk is used, but a mixture with goats' milk (up to 30% goats' milk in

Table 1 Classification of brined cheeses

Cheese	*Country of origin*	*Milk used*
Rennet-coagulated		
Soft cheeses		
Feta type (salting of curd)		
Feta	Greece	S, S plus G, C
UF-feta	Denmark	C, R
Telemea	Romania	S, B, C
White pickled	Eastern Europe	S, C
Bli-Sir-O-Kriskana	Former Yugoslavia	S
Bajalo (Belo Salamurene Sirene)	Bulgaria	S, C
Brinza	Former USSR, Bulgaria	C
Chenakh	Former USSR	S
Akiavi	Syria, Turkey	S, G
Domiati type (salting of milk)		
Domiati	Egypt	B, C, R
UF-Domiati	Egypt	C, R
Dani	Egypt	S
Semihard cheese		
Halloumi	Cyprus	S, G, C
Medaffara/Magdola	Sudan, Syria	S, B, C, R
Shenkalish	Syria	S
Acid-coagulated		
Mish	Egypt	Skim milk

S: sheep; G: goat; C: cow; B: buffalo; R: recombined milk; UF: ultrafiltration.

mixture) can be used. The milk is heated to 65 °C for 30 min (optional) and then cooled to 32–33 °C; 1–3% starter (*Lactococcus lactis* plus *Lactococcus lactis* subsp. *cremoris*) is then added (optional); 20–30 ml of rennet (preferably lamb or kid rennet) per 100 l are used to coagulate the milk in 50–150 min; the curd is then cut into 2–3-cm cubes and left in the whey for 10–15 min, after which it is scooped into metal hoops, 20 cm deep, and left for 1 h; the hoops are transferred to a room (16–18 °C), and the surface of the curd is lightly salted; the cheese is turned upside down and left for 20 h; the curd is then cut into pieces (different sizes), placed in wooden barrels in layers, and dry salted with 3% coarse salt over a period of 3 days; and the barrels are topped up with 5% brine and left for 15 days, prior to being transferred to cold storage (4 °C) until marketed. (*See* **Lactic Acid Bacteria**; **Starter Cultures**.)

Telemea cheese Whole milk (usually cows' milk) is heated to 60–68 °C for 20–30 min, 0.5% starter (mesophilic bacteria plus *Lactobacillus casei*) is added at 28–31 °C, followed by 25–30 ml of rennet per 100 l of milk, and held at 30–34 °C for 60 min, and the curd is transferred into cheesecloth in layers. The resultant curd is cut crossways with a long knife and pressed for 2 h. The cheese is cut into blocks and immersed in a salt bath (18–22% NaCl) at 15–16 °C overnight, and layered in tins, which are then completed (filled) with acid (1.3–2% lactic acid), salted (4–5% NaCl) whey for 1 month, then stored at 5–10 °C. (*See* **Whey and Whey Powders**: Production and Uses.)

White pickled cheese Whole milk is flash-heated to 70 °C or 68 °C for 10 min (optional), and a starter (sour milk or mesophilic bacteria 0.5–1.0%) is added followed by rennet (25–30 ml per 100 l of milk at 30–31 °C) to complete coagulation in 60–70 min; the curd is ladled into a cheesecloth, which is laid over a wooden frame, placed over a drainage table; the cheese is pressed by tightening the cloth and by the use of weights; the curd is cut into cubes (10 cm^3) and immersed in saturated brine overnight; and the cheese is packed into barrels or tins filled with salted whey (8–12% NaCl).

Brinza Milk is coagulated by rennet at 25–32 °C for 40–90 min, after which the curd is transferred to the drainage table for cutting, tuning, and pressing (whey is removed in 2 h), cut into rectangular blocks (1 kg) and salted in brine at low temperature (10–12 °C). The cheese is then packed into wooden barrels or lacquered tins filled with brine or salted whey and stored at 4 °C or 12–20 °C.

Bli-Sir-O-Kriskana Starter (0.1% *Lactococcus lactis* plus *Lactobacillus casei*) is added to the milk, followed by rennet (20–30 ml per 100 l at 30–32 °C) in order to achieve coagulation in 70–90 min. The

curd is cut horizontally and vertically, and left for 15–20 min, recut into smaller cubes and ladled into a cheesecloth, pressed for 1–4 h, then cut into cubes (10 cm^3) for dry salting or immersion in brine (20–25% NaCl) for 10–20 h at 12–15 °C. The cheese is packed in parchment-lined casks or tins completed with brine (10–12% NaCl) and held at 12–15 °C to ripen in one month and then stored at 10 °C.

Bjalo (Belo Salamureno Sirene) Milk is heated at 68 °C for 10 min, then cooled at 31 °C; 0.1–0.3% starter (*Lactococcus lactis* plus *Lactobacillus casei*) is added followed by 20 ml of rennet per 100 l (coagulation time: 60–90 min). The curd is cut into cubes (2–3 cm^3), which are ladled into a cheesecloth, then cut into strips, and the cloth tied over the curd. It is then left to drain, pressed for 4–5 h at 14–16 °C, cut into pieces (12 × 12 cm), immersed in saturated brine or dry-salted for 24 h, and packed into metal or plastic drums filled with brine (10–12% NaCl). The cheese is ripened for 2–4 weeks and stored at 4–8 °C.

Domiati Dry salt (5–15%) is added to the milk and stirred until dissolved. The salted milk is then strained through cheesecloth, heated to 37–42 °C, and rennet (30–40 ml per 100 l) added to achieve coagulation in 2–3 h. The curd is scooped into wooden frames lined with cheesecloth, placed over a wooden board on a draining table and allowed to drain for 3–4 h. The curd is then broken and pressed to achieve complete drainage in 16–24 h. The cheese is cut into blocks, wrapped in cheesepaper, and packed into 2–20-l tins filled with salted whey (from the same cheese) and stored at room temperature.

Halloumi A sufficient quantity of rennet is added to sheep's milk to coagulate the milk at 34 °C within 7–8 min. The curd is cut after 30 min into grains (1 cm^3) and left for 10 min, stirred gently for 10 min, scalded at 60 °C for 15 min, transferred to hoops, then pressed with weights (3 kg per kilogram of cheese) for 35 min. The pressed curd is cut into small blocks (10 × 10 × 3 cm), transferred to deproteinized whey, then heated to 90–92 °C for 30 min. The cheese is then drained and sprinkled with dry salt and dried *Mentha viridis* before conservation in brine.

Medaffara Rennet is added to raw milk at 32–37 °C for 40–60 min and left to ripen (30–60 min). After the curd cutting and partial removal of the whey, the cubes are left to ripen in the whey until they form a slim block about 2 m long. The ripened curd is then kneaded in hot water at 80 °C for 10 min, after which black cumin spice (*Nigella sativa*) is added; the curd is then shaped, while hot, to form a cord that is braided and then brined.

Industrial Scale Methods

Equipment and process development Rapid developments in the technology of brined cheeses have been achieved during the last two decades; the application of ultrafiltration (UF) in the manufacture of feta cheese is one of the most outstanding developments. Industrial-scale production of brined cheeses is based on modifications of traditional methods. The main differences between the traditional and industrial scale methods are:

- Type of milk: Cows' milk is usually used on an industrial scale. This requires the treatment of milk fat in order to retain the white color of brined cheeses. Blue–green dye (e.g., chlorophyl) is added mainly for this purpose. Decolorization of pure milk fat by heat treatment/use of oxidants is practiced to produce a special milk fat for recombined brined cheeses.
- Heat treatment: Milk is normally pasteurized before cheese making to ensure public safety, to minimize quality defects, and to allow the cheese to retain more moisture.
- Homogenization: This step is essential for the manufacture of brined cheeses from recombined milk. It also improves the fat retention and smoothness of cheese from fresh milk. Cheese from homogenized milk appears whiter in color.
- Additives: Mesophilic and thermophilic lactic starters, or a combination of these, are used. Also, lipases, acidulants (glucono-δ-lactone), and calcium chloride are frequently used.
- Ultrafiltration: Most of the progress in brined cheeses has been achieved through the use of ultrafiltration techniques. Such techniques result in a substantial increase (20–30%) in cheese yield, saving manpower, additives, and energy consumption. Also, the compositional quality of the product can be easily controlled.
- Packaging: New packaging materials and equipments are used. In addition to tins, laminated paper cartons (tetrabricks), special polymeric films, and rigid plastic containers are now being used in retail packaging.

Several systems have been developed for the continuous manufacture of brined cheeses from milk or partially ultrafiltered milk using all or part of the following processes and mechanized equipment.

- Dosing equipment. For continuous blending of milk/UF retentate with various additives, i.e., a starter culture, coagulant, or decolorizing agent.

Such equipment consists of metering devices and in-line static mixers.
- Coagulator/cutting devices.
- Molding and dewheying.
- Bulk packaging, salting, and brining.

UF-cast feta cheese Two systems are in use for the production of UF-cast feta cheese:

1. Standardized pasteurized milk (12.5% total solids (TS), 3.5% fat) is ultrafiltered to give a retentate of 37–38% TS. The retentate is then pasteurized, homogenized, and pumped from a balance tank by means of a finely controlled, positive pump to the filling, mixing, and dosing aggregate. At the same time, precise amounts of rennet, lipase, NaCl, and acidulant (glucono-δ-lactone and blue phosphoric acid) are dosed into the concentrate before it reaches the filling head. The concentrate is filled into laminated 'tetra bricks' (250–500 g), or it can be packed into a special polymeric film and encased in a rectangular carton.
2. The UF-milk retentate (37–38% TS) is pasteurized, homogenized, and directed into an intermediate tank. After cooling to 30 °C, 3% mesophilic starter and lipase (4–8 g of pregastric lipase powder/100 kg) are added. No time is allowed for acid development to take place. Rennet is dosed in-line into the retentate before casting in tins (10/20 l). Filling takes place in three stages as each layer is coagulated separately. The tins are then taken to the portioning section, where the cheese is cut by a cutting machine equipped with vertical knives. The cheese is salted with 4–5% NaCl, which is dosed and spread automatically on the top of the cheese; the tins are then closed and stored in a cold room until marketed.

UF-structure feta Continuous coagulation of retentate with 26% TS takes place in a coagulator, followed by continuous cutting of the coagulum. The cheese grains are treated in the permeate (from the UF process), in which syneresis takes place. The cheese grains are filled into molds, whey is drained off, and the finished cheese is cut into blocks, filled into tins, and salted with dry salt, as described for cast feta.

UF-Domiati Standardized pasteurized milk (12.5% TS, 3.5% fat) is concentrated by UF to 35% TS. The retentate is homogenized, heated to 75 °C for 1 min, and cooled to 34 °C; 1% mesophilic starter, rennet (3 g of rennet powder/100 kg), and 3–4% NaCl are added and treated as follows:

- The retentate is poured into stainless steel trays (50 × 50 × 10 cm) and left to coagulate. The cheese is then cut manually into cubes (0.5 kg), wrapped in polyethylene sheet and packed into a 1-kg rigid plastic container (consumer package) or in 10/20-1 tins; the containers or tins are filled with 5% salted permeate and then closed.
- The retentate is cast in plastic containers (750 g) placed in a moving conveyer in a constant temperature tunnel (40 °C), in which there is complete coagulation before the package leaves the tunnel. Salted brine (5% NaCl) is dosed, and packages are sealed with aluminum foil and covered with a plastic lid.

Recombined feta/Domiati cheeses Recombined feta/Domiati cheeses are made from recombined milk using either low-heat skim milk powder or skim milk powder retentate and anhydrous milk fat. Recombination is carried out in batches or continuously depending on the scale of production. Immediately after mixing, the recombined milk is homogenized, heated, cooled, and made into cheese as for fresh milk. Commercial pregastric lipase is added to improve the acceptability of the product.

Changes in the Composition of Brined Cheeses During Storage

Table 2 shows the average composition of several brined cheeses. General factors affecting the composition of other cheese varieties apply also to brined cheeses. These include the type of milk, heat treatment received, and conditions and duration of storage. In addition, storage in brine plays a special and important role in determining the changes in composition of brined cheeses. This can be understood from the following:

- brine provides a very high water activity for stored cheeses;
- there is a continuous equilibrium in the distribution of soluble constituents between cheese and brine;
- there are ion-exchange equilibria between Na^+Cl^- in the brine and the ions in the cheese;
- brine partially solubilizes the protein matrix of cheese;
- brine controls the cheese microflora; the surface microflora are of limited significance.

In general, the changes occurring in brined cheeses can be described as follows.

Table 2 Average composition of some brined cheeses

Cheese	*Moisture (%)*	*Fat/dry matter (%)*	*Total protein (%)*	*NaCl (%)*	*pH*
Feta	48–54	48–52		4–5	4.3–4.6
Telemea	55	48–51		3–5	4.4–4.7
Bjalo	58–60	27–31	10–12	3–4	
Brinza	58	45–50		4–10	
Domiati (fresh)	60	40		8–10	6.6
Domiati (3 months)	54	45–50		6–8	3.7–4.2
Halloumi	35–48	38–50	24–30	2–5	5.3–6.1
Medaffara	53–54	17–22	21–24		5.1

Moisture Content

Depending on the initial moisture content, pH, and concentration of brine, the moisture content of brined cheeses decreases on storage. This is obvious in Domiati cheese, where a 5–10% decrease in cheese moisture is normally observed during storage.

Nitrogenous Constituents

The different cheese proteins are subject to variable proteolysis: α_S-casein is extensively hydrolyzed, while β- and *p*-χ-casein are less affected. An increase in the soluble nitrogenous constituents is normally observed during storage of brined cheeses. These compounds include peptides, amino acids, and ammonia, but the formation of biogenic amines is very limited.

Milk Fat

The rennet used in brined cheeses traditionally contained pregastric lipases, and in developed technologies, lipase is added. Lipolysis is therefore one of the characteristics of brined cheeses. The mono- and diglyceride content of the cheese increases on storage with the formation of variable free fatty acids. However, the volatile fatty acids of brined cheeses contain a high percentage of acetic acid, originating from microbial fermentations. (*See* **Fatty Acids**: Properties.)

Lactose

Fresh brined cheeses retain variable quantities of lactose, and those stored in salted whey have access to a continuous supply of lactose during storage. It is not surprising, therefore, to find measurable quantities of lactose in brined cheeses throughout the storage period. The presence of galactose in stored Domiati cheese has been reported.

Mineral Constituents

The calcium content of brined cheeses decreases as a result of acid development and ion exchange with NaCl in the brine; the phosphate content of cheese also decreases during storage. Changes in Na^+ and Cl^- contents of cheese are dependent on: (1) the concentration of the brine, (2) the initial NaCl content of cheese, and (3) any decrease in cheese moisture during storage.

Vitamins

Vitamin A is stable, but thiamin, riboflavin, and nicotinic acid show a variable decrease in cheese stored in tins. (*See* **Niacin**: Properties and Determination; **Retinol**: Properties and Determination; **Riboflavin**: Properties and Determination; **Thiamin**: Properties and Determination.)

Volatile Flavor Compounds

Brined cheeses develop a mixture of short-chain fatty acids. Acetic acid is the major volatile fatty acid in Domiati cheese, whereas in feta cheese, butyric acid is present in significant quantities, particularly in traditional feta. In feta cheese, several alcohols and carbonyl compounds increase in concentration on prolonged storage.

Texture and Structure

The textures of cast UF and traditional brined cheeses are significantly different. However, the texture of both is affected by several factors including homogenization, amount of rennet used, pH, NaCl content, and storage period/conditions. Generally, brined cheeses retain a softer texture during ripening that is related mainly to proteolysis and the level of whey proteins in cheese.

Electron microscopic examination of feta and Domiati cheeses indicates that the internal structure of fresh cheese is composed of a framework of spherical casein aggregates, held together by bridges and enclosing fat. During storage in brine, the casein aggregates in Domiati cheese disintegrate into smaller spherical particles, forming a loose structure. The casein particles in Feta cheese coalesce as particles cling to one another, forming larger clumps than in Domiati cheese. The changes in the protein matrix

during storage are generally responsible for the smooth body of ripened brined cheeses. These changes are likely to arise from (1) the partial loss of calcium from the cheese matrix into the brine and (2) continuous proteolysis of α_s-casein, known for its important role in linking the cheese protein network.

Microbiology

Microorganisms Present

Lactic acid bacteria are dominant in Domiati cheese, Streptococci during the early stage of pickling, and later Lactobacilli. *Streptococcus faecalis*, *Lactococcus lactis*, *Lactobacillus casei*, *Lactobacillus plantarum*, *Lactobacillus brevis*, and *Lactobacillus fermeti* are the most common Streptococci and Lactobacilli present in Domiati cheese. Several nonlactic bacteria and yeasts are also found.

The predominant lactic acid bacteria in other white brined cheese are Streptococci (*S. faecalis*, *S. faecium*, *S. durans*), Leuconostoc (*L. dextrerincum*) and Lactobacilli (*Lb. plantarum*, *Lb. casei*, *Lb. brevis*). The principal psychrotrophs found are from the genera *Pseudomonus*, *Aeromonas* and *Rhodotorula*. Salt-resistant yeasts growing on the surface of feta cheese during ripening may contribute to ripening.

Effect of Manufacturing and Storage Conditions

Raw-milk cheese generally has higher bacterial counts than the pasteurized-milk cheese during early stages of ripening, but both cheeses have similar counts thereafter. The total bacterial count of Domiati cheese increases rapidly, reaching a maximum after a week of manufacture, and then declines. Streptococci are dominant in the early stages of pickling and then replaced with Lactobacilli, or Lactobacilli and Micrococci, in the later stages of ripening, depending on the level of NaCl in cheese. Micrococci and Lactobacilli dominate in Domiati cheese with a high salt content.

The number of aerobic, lactic, and proteolytic bacterial counts of feta cheese reaches a maximum after 5, 20, and 45 days, respectively, and then decreases markedly at 3–4 months of storage. The lipolytic count shows a slight increase during ripening and storage.

The fresh Hallomi cheese has very limited bacterial counts, but these increase significantly during storage. Spore formers contribute to this flora.

Survival of Harmful Organisms

Coliform bacteria are frequently found in white brined cheeses in large numbers mainly due to the use of raw milk and poor sanitary conditions during small-scale cheese making. No less than 9.5% NaCl should be added to the milk to suppress the growth of coliforms in Domiati cheese from raw milk. The presence of coliforms is responsible for the blowing defect in white brined cheeses.

Staphylococcus aureus can tolerate 2.5–15% NaCl in Domiati cheese. Also, *Salmonella typhi* can survive for up to 16 days in Domiati cheese from milk containing 10% NaCl. Brined cheeses are also good media for the survival of *Listeria monocytogenes* and *Yersinia enterocolitica*. *Nocardia*-like organisms dominate the microflora of the surface slime defect observed in Domiati cheese from raw milk. (*See* **Spoilage:** Bacterial Spoilage; Yeasts in Spoilage.)

See also: **Buffalo**: Milk; **Fatty Acids**: Properties; **Goat**: Milk; **Lactic Acid Bacteria**; **Milk**: Liquid Milk for the Consumer; **Sheep**: Milk; **Spoilage**: Bacterial Spoilage; Yeasts in Spoilage; **Starter Cultures**; **Whey and Whey Powders**: Production and Uses

Further Reading

Abd El-Salam MH, El-Shibiny S and Fahmi AH (1976) Domiati cheese. A review. *New Zealand Journal of Dairy Science and Technology* 11: 57–61.

Abou-Donia SA (1986) Egyptian Domiati soft white pickled cheese. *New Zealand Journal of Dairy Science and Technology* 21: 167–190.

Fox PF (ed.) (1993) *Cheese: Chemistry, Physics and Microbiology*, vol. 2. London: Chapman & Hall.

IDF (1986) *Proceedings of the IDF Seminar 'Production and Utilization of Ewe's and Goat's Milk'* Doc. 202. Brussels: International Dairy Federation.

IDF (1990) *Recombination of Milk and Milk Products. Special Issue No. 9001*. Brussels, Belgium: International Dairy Federation.

Lloyd G and Ramshaw EH (1979) The manufacture of Bulgarian style feta cheese. A review. *Australian Journal of Dairy Technology* 34: 180–183.

Ottosen N (1988) *Feta by the Application of UF*. Silkeborg, Denmark: APV Pasilac A/S.

Papademas P and Robinson RK (1998) Hallomi cheese: the product and its characteristics. *International Journal of Dairy Technology* 51: 98–103.

Renner E and Abd El-Salam MH (1991) *Application of Ultrafiltration in Dairy Industry*. London: Elsevier Science.

Robinson RK and Tamime AY (1991) *Feta and Related Cheeses*. London: Ellis Horwood.

Scott R (1986) *Cheese Making Practice*, 2nd edn. London: Elsevier Science.

Quarg and Fromage Frais

H Klostermeyer, Lehrstuhlfür Chemie der Biopolymere, Freising-Weihenstephan, Germany

Introduction

Acid fresh, unripened cheeses are produced throughout the world, but consumption is extremely variable, even within individual countries. Besides Israel (*c.* 11 kg per head per annum), high consumption is observed in central Europe (Poland, 10 kg; Germany, 9 kg), Iceland (7 kg), former Czech Republic and France (5 kg), and Hungary and Austria (4 kg) (for comparison: UK, 0.1 kg). About 30–40 different types of unripened cheeses are mainly made from cows' milk, but also from ewes' milk and especially from goats' milk, these latter products being rather popular in the Mediterranean countries. A common criterion for all cheeses of this class is the low pH value and a calcium content of 0.6–0.9% in the dry matter.

Historically, the lactic curd cheeses are mainly byproducts of butter manufacture, i.e., they are produced from partly skimmed milk. Before 1888, the separation of cream occurred by gravitation. The milk was set at rest in a cool place in shallow pans, during which the cream and the skimmed milk became acid. After skimmed milk had clotted, the coagulum was scrambled and drained of whey, either in cloths or in bags. Sometimes the curd was pressed or even washed to remove most of the whey. Handling of the lactic curd (quarg) was different in individual countries. In central Europe, for example, small cheeses were sometimes prepared from salted quarg, either mold-ripened (Korbkäse, Graukäse, Schabziger, etc.) or smear-ripened with yellow-reddish smear-forming, wild-type coryneformes (Harzer Käse, Mainzer Käse, Olmützer Quargel, etc.), or heated with a mild base (e.g., sodium hydrogen carbonate) and spiced to form cooked cheese. Originally, all these types of cheese had a variable content of fat, owing to the fact that creaming was more or less incomplete before the acid coagulation of the milk took place in the vats.

After introduction of the milk centrifuge (separator) in 1888, sweet skimmed milk became available, which could be used for the production of lactic as well as rennet-coagulated cheeses. However, in most countries, the consumption of lactic curd cheeses decreased constantly because of their very limited shelf-life (less than 2 days), their often unclean flavor, bitterness due to proteolytic activity, and short consistency. In addition, the rather high fat content of the traditional products had made them much more tasty than the new products made from pure acid casein. Even in central Europe, where quark had a long tradition, the product remained a homemade one.

About the time of World War I, the taste of quarg was improved by adding about 2% buttermilk (which contains homofermentative lactic starters, and heterofermentative, aroma-producing lactic acid bacteria) as fermentation culture, e.g., starter to skimmed milk, and by combining the acid-induced formation of the curd with coagulation by renneting. This not only caused a change in the consistency (acid casein gels are hydrophilic, rennet-induced gels are hydrophobic with a tendency to syneresis, especially when heated (maximum 65 °C)) but also permitted a certain heat treatment of the curds which led to products with shelf-lives of up to a week (Dauerquark) at temperatures below 15 °C. This development enabled the production and distribution of quarg outside the home. At the same time, emigrants from Europe introduced into the north-west of the USA a variant of drained quark, namely cottage cheese, which became very popular, at least in North America.

Up to that time, fresh cheeses were prepared from unheated milk to avoid the precipitation of denaturated whey proteins on to the casein, a change which induces a higher water-binding capacity in the protein, and gives a very soft, smooth consistency to the product. In former Federal Republic of Germany in 1951 it was required by law that milk used for unripened cheeses must be pasteurized (72–74 °C for 20 s) in order to avoid health risks. This led to a further reduction in the level of dry matter in quarg, and the new product, Speisequark, became very popular, particularly when enriched with cream. Up to that time, quarg had been mainly used for cooking and baking, but it then lost its cheese image and became a lactic acid dairy product for direct consumption, either sweet with fruit, or salty with herbs and spices.

The problem of limited shelf-life was overcome by the development of special separators (up to 10 000 l h^{-1}) for quarg in the 1950s. It became possible to produce Speisequark on a large scale in enclosed sanitary production lines, avoiding recontamination during packaging. The shelf-life of this product was extended to 22 days (12 °C), even without adding preservatives. (In some countries, preservatives such as sorbic acid and hydrocolloids are used to prolong the shelf-life of fresh cheeses.) Within a 10–15-year period, consumption of the new type of quarg increased in Germany about 10-fold, and production lines handling 200 000 kg of milk per day made the product a rather low-priced one.

Consequently, new developments were necessary to overcome the resulting low profits of the dairies.

In the middle of the 1970s, a product (Thermoquark) was developed which contained, besides the caseins, all of the whey proteins of milk. They were denatured by high-temperature heat treatment (82–92 °C for 5–6 min) of the milk prior to fermentation, and special centrifuges were developed for the Thermoquark. Some manufacturers also started to add denaturated whey protein concentrates – separated from heated whey either by the 'centriwhey' process or by ultrafiltration – to quarg. Since some dairies added up to 30% whey proteins to fresh cheeses, the International Dairy Federation and the German government limited the use of whey in fresh cheeses 'up to the natural whey protein content of the milk', i.e., 16–18% of the total milk protein. (*See* **Casein and Caseinates**: Uses in the Food Industry; **Membrane Techniques**: Principles of Ultrafiltration; **Whey and Whey Powders**: Protein Concentrates and Fractions.)

Direct concentration of sweet milk up to 18% dry matter by ultrafiltration, fermentation, and curd formation without whey drainage was studied extensively for the production of quarg, but without success, since calcium ions (Ca^{2+}), which are not removed from sweet milk by ultrafiltration (UF), lead to the development of a bitter flavor in the product during storage. For this reason UF-Quark is now produced by a method similar to Thermoquarg, mainly from high-temperature-treated milk and not from a milk concentrate, but the separation of the curd from the whey is achieved mechanically by concentrating the fermented milk by UF with stable membranes instead of a special separator. The benefit for the producer is that the dry matter of UF-quarg contains more lactose and ash in the dry matter compared to Thermoquarg, and thus there is a saving in expensive protein. As a result of this development, the law in Germany was revised, and Speisequark must now contain at least 12% total protein; otherwise, the product must be named 'fresh cheese.'

Industrially produced fresh cheeses containing high amounts of whey proteins are heat-treated at about 60 °C before separating the whey by centrifugation or ultrafiltration, otherwise it is impossible to obtain a dry-matter content of 18% and more. Lactic acid bacteria and their enzymes are inactivated at this temperature, as are the enzymes produced by psychrotrophs. The shelf-life of unripened cheeses is nowadays not limited by bacterial contamination, but rather by a decline in sensory properties, e.g., from light-induced off-flavors. (*See* **Lactic Acid Bacteria**; **Pasteurization**: Principles.)

A special group of products, e.g., double cream cheeses, is made from quarg containing high levels of cream. Sometimes, they are stabilized with hydrocolloids (not allowed in Germany) and pasteurized by heating to 60–80 °C for 2–5 min. Hot packaging guarantees a shelf-life of at least 12 weeks. During the last few years, some rather unique products of this type, containing herbs and spices, have been produced by techniques and equipment used for the manufacture of processed cheese. Products with a very high fat content are made by adding emulsifiers (e.g., phosphates and citrates). Italy seems to be the leading country in the development of this type of product. (*See* **Emulsifiers**: Phosphates as Meat Emulsion Stabilizers.)

Within the last 100 years, the character and composition of fresh cheeses have changed completely, at least where the quarg-type products are concerned. However, besides the newly developed varieties of product, the traditional products are still available, e.g., in shops for alternative food or in Italian specialty shops, so that the range of fresh cheeses available to modern-day consumers has expanded greatly during recent decades. Table 1 shows changes in the composition of quarg within the last century.

Production

Milk

The raw material for most fresh cheeses is (partly) skimmed milk, but some types are produced from (homogenized) full-fat milk or even from fat-enriched

Table 1 Mean composition of quarg from skimmed milk in former Federal Republic of Germany within the last century

Product	*Year*	*Dry matter (%)*	*Lactose (%)*	*Lactic acid (%)*	*Salts (%)*	*Protein (%)*	*Ratio of whey protein to casein*	*Milk content (kg) per kg of quarg*
Traditional quarg	1871	32–35	2	1.5	1	27	0.02:1	10
Dauerquark	1910	23.3	3.5	0.9	0.9	18	0.04:1	6.7
Quark	1947	21	3.0	1.0	0.7	16.0	0.07:1	5.8
Speisequark	1951	17	3.0	0.8	0.7	12.5	0.04:1	4.8
Thermoquark	1977	17	3.2	1.0	0.8	12.0	0.19:1	3.9
UF-Quark	1985	17	3.7	1.3	1.3	10.7	0.20:1	3.4

Mean values from different sources.

milk (up to 12% fat). In the case of Speisequark, it is usual to add the desired amount of cream to the low-fat curd after separation of the whey. This is impossible in the case of highly structured products such as full-fat cottage cheese. Besides milk, considerable amounts of buttermilk are used nowadays to produce fresh cheeses. Owing to the fact that most fresh cheeses have a soft consistency, reconstituted milk from low-heat powder is also a rather good raw material; however, some $CaCl_2$ should be added. Some varieties of lactic cheese are prepared from milk–whey mixtures.

As already mentioned, heat treatment of the milk varies between zero and high heating, according to the individual products and the technology in use. However, it should be remembered that fresh cheeses do contain a low amount of their own flavor compounds and they are very sensitive to off-flavors from the milk used. Only raw milk of the best quality yields fresh cheeses with the typical and delicate flavor.

Acidification

Some (Mediterranean) fresh cheeses are prepared by direct acidification with lemon juice or vinegar, but nowadays commerically available lactic acid may be used, and even phosphoric acid has been mentioned as an acidulant. However, most fresh cheeses are produced by acidification of the milk with lactic acid bacteria, i.e., by fermentation. In the northern countries, mesophilic starters are used and in oriental areas with warm climates, thermophilic microorganisms, or even pure yogurt cultures, are used.

Industrial production of fresh cheeses involves the use of selected strains of *Lactococcus lactis*, and *L. lactis* subsp. *cremoris* to develop lactic acid, and sometimes *Lc. lactis* spp. *lactis* (*L. lactis* biovar. *diacetylactis*) or *Leuconostoc mesenteroides* subsp. *cremoris* to form flavor compounds. To avoid proteolysis, which causes loss in yield and initiates off-flavors, *Lc. lactis* spp. *lactis* may be removed from some selected starters.

With thermophilic starters (optimum 43 °C) it is easier to limit acidification; *Lactobacillus acidophilus* is a typical species in use. For example, cultures with limited acidification not lower than pH 4.8–5.0 are produced by Biogarde® containing *Streptococcus thermophilus, Lactococcus lactis*, *Lc. lactis* biovar. *diacetylactis, Leuconostoc mesenteroides* subsp. *cremoris, Lactobacillus acidophilus* and even *Bifidobacterium bifidum*. Labaneh and labneh, which are very popular in the Middle East, are acidified with thermophilic yogurt starters. Industrially used starters should not produce gas (CO_2); this avoids problems in the production line, especially in the separators. (*See* **Starter Cultures**.)

Normally, 0.5–1% starter is added to the milk at about 30 °C. When pH 6.3 is reached (after about 1.5 h), rennet may be added (in this case, the milk is stirred well to distribute the enzyme), and in about 16 h the desired acidity of pH 4.5–4.55 is reached. During acid development, the milk should not be stirred. Clearly, acid development is strongly dependent on some important factors, which are part of the typical and individual technology for the different unripened cheeses.

Curd Formation and Treatment

Rennet, if any, is used in small amounts, e.g., 0.5–1.0 ml (1:10 000) per 100 l of milk for quarg.

Pure acid casein curds are rather sensitive to heat, a temperature around 43 °C and a pH of 4.3 seem to be optimal for whey drainage. If the temperature is only a few degrees higher, the curd changes from dry to smeary. After cutting, traditionally prepared lactic curds are therefore normally filled into cheesecloths and drained at room temperature, and afterwards at a low temperature (4–5 °C) for many hours.

In the case of renneted acid curds, heating to 50–55 °C for 1.5 h (cottage cheese) or even 60–64 °C for 4–5 min (quarg) improves syneresis of the coagulum. Owing to these effects, cottage cheese is prepared in a granular form; the curd is washed three times with water at decreasing temperature (30 °C, 16 °C, 4 °C) to remove lactose and to stabilize the pH and shelf-life.

Centrifugation of quarg curds is performed at about 40–44 °C; ultrafiltration is performed over the same range of temperatures or even higher. Cooling of quarg takes place in special tubular coolers, and the product is transported through the line by positive pumps. At the end of the line, a quarg mixer may be installed to enrich the product with cream, salt, fruits, stabilizers, etc. Fresh cheeses are normally packaged in plastic containers, glass jars, etc., preferably under aseptic conditions. Some types of fresh cheese with a high content of dry matter are sold as small, round, or wedge-shaped bodies packaged in foil.

This is certainly true for the high-fat cream cheeses. They are produced from homogenized milk containing 10% or more fat (with 9% fat, the curd has the same specific weight as whey, so that centrifugal separation is impossible!). The acidified milk is heated to 80 °C before separation. The total solids content of the hot drained curd is 44%, and after salting and cooling, it becomes a viscous mass due to crystallization of fat.

Miscellaneous Products

Cottage cheese represents the most structured type of lactic cheese. 'Dry curd' quality (< 0.5% fat) is produced from skimmed milk; cottage cheese (low-fat) contains 0.5–2% fat, and cottage cheese itself usually contains > 4% fat. It is normally produced by treating the grains with a cream dressing. According to the size of the grains, it is distinguished as Californian-style, popcorn-style, and country-style, as well as large- and small-curd style. Block-pressed cottage cheese is called farmer's cheese.

Although cottage cheese is slightly salted (1% NaCl), the shelf-life is often below 2 weeks, owing to the rather high pH value and sometimes poor bacteriological quality of the wash water. In Germany, products with a fat in dry matter up to 20% are made, and manufacturing technologies involving ultrafiltration of the milk are also applied.

Quarg (Quark, Topfen, Koarg, Kwarg, Twarog, Tvorog, Taho, lactic cheese, fromage frais, etc.) is available with different levels of fat, up to 45% in the dry matter. With increasing fat content, the consistency changes from crumbly-dry to smooth, but this texture can be influenced by technology (e.g., dry-matter, casein, and whey protein content). A foamed quarg is named Bresso. Whipped quarg with not more than 15% dry matter is fromage frais battu maigre. Baker's cheese is characterized by a high dry-matter content (22–26%). Schichtkäse, like Cambridge and York, is prepared from layers of low-fat and high-fat lactic curd. Danish smoked quarg is named Rygeost.

Typical lactic curd, fresh cheeses with a high fat in dry-matter content are cream cheese (70% fat in dry matter), Petit-Suisse (> 70% fat in dry matter) and Demisee (40% fat in dry matter), Neufchâtel (55% fat in dry matter), Doppelrahmfrischkäse (< 85% fat in dry matter) and double cream cheese (> 65% fat in dry matter). Some of these products are stabilized with hydrocolloids, and the milk is homogenized. Products such as Philadelphia, Gervais, Boursin, and Le Tartar (with herbs and spices) are internationally distributed.

A very-high-fat Italian fresh cheese is mascarpone (Mascherpone). It is produced from full-fat cream, heated to 90 °C, and acidified with citric acid. Whey drainage is performed at 8–10 °C for 12–18 h. In the new Sordi process, cream (40–60% fat) is pasteurized, mixed with milk concentrate from ultrafiltration, and acidified with citric acid in heated vats. After careful stirring, the product is packaged and stored for about 12 h in the cold to crystallize the fat. Mascarpone is very popular for sweet desserts.

Ricotta was prepared originally by the direct acidification of heated cheese whey from ewes' milk (ricotta di pecora). Many varieties of ricotta are now available, mainly from cows' milk (ricotta di vacca). The consistency varies between soft and dry (ricotta secca). Salted varieties (ricotta salata), as well as the smoked cheeses, are used in the kitchen, whereas the unsalted product (ricotta tipo dolce), like the baked ricotta (ricotta salata al forno), are consumed directly as appetizers or desserts. Ricotta is named requesón in Spain and requeijao in Portugal.

Queso blanco (queso del país, queso fresco) is also available in many varieties in Spanish-speaking countries, but is now also manufactured in the USA. It is produced by a combined heat-acid process (82 °C and pH 4.6–4.7). Citric acid is said to give the best results, but acetic acid is the most popular acidulant. This cheese can be fried without melting, and is used in the preparation of snack foods; with spices and tomato sauce or chilli sauce, it is consumed fresh. The organoleptic qualities can be improved using yogurt starters instead of mesophilic starters. The moisture content is about 50–54%. (*See* **Acids**: Natural Acids and Acidulants.)

Yogurt starters are used exclusively in the production of labaneh and labneh. Labaneh is the pure acid curd from skimmed milk, and it can be stored for a long time in olive oil. Labneh is prepared from full-cream milk by acidification and by rennet coagulation. The dry-matter content is approximately 40%. These products are the Middle East equivalents of quarg.

Last, but not least, the Swiss Zieger (named Seirass in Piemont, Cérat in Savoy) should be mentioned. Originally prepared by heating acidified whey, they are now also produced from mixtures of milk and whey, or even from skimmed milk. The cheese is white or somewhat yellow, spreadable, and crumbly-soft.

See also: **Acids**: Natural Acids and Acidulants; **Casein and Caseinates**: Uses in the Food Industry; **Emulsifiers**: Phosphates as Meat Emulsion Stabilizers; **Lactic Acid Bacteria**; **Membrane Techniques**: Principles of Ultrafiltration; **Pasteurization**: Principles; **Starter Cultures**; **Whey and Whey Powders**: Protein Concentrates and Fractions

Further Reading

Drewes K (1952) *Die Herstellung und Beurteilung von Sauermilchquark*. Hildesheim: Mann-Verlag.

Eck A and Gilles JC (eds) *Le Fromage: de la Science à l'Assurance-Qualité*, 3rd edn. Paris: Lavoisier.

Lehmann HR, Dolle E and Bücker H (1984) *Prozesslinien zur Herstellung von Frischkäse*. Oelde: Westfalia Separator.

Spreer R (1998) *Milk and Dairy Product Technology*. New York: Marcel Dekker.

Stieger W (1920) *Anleitung zur Quarkbereitung und zur Handkäsefabrikation*, 3rd edn. Berlin: Paul Parey.

Processed Cheese

A Gouda and A Abou El-Nour, Suez Canal University, Ismailia, Egypt

Introduction

Cheese in general, including processed cheese, is rich in different nutrients such as proteins, fats, minerals, and vitamins. It is reported that utilization of proteins of processed cheese is better than proteins of natural cheese.

Processed cheese, compared to natural cheese, is a new product which originated at the beginning of the twentieth century. Spoilage of cheese exported to distant places was the motive behind the trials to improve the shelf-life of cheese. Some of these trials succeeded in prolonging the keeping quality of soft or even semihard cheese by applying heat treatment, but the situation was not the same for hard cheese, due to shrinkage of the cheese protein by heat and separation of the water and fat phases. In 1912–13, the Swiss workers, Walter Gerber and Fritz Settler, solved these problems by heating hard cheese to which sodium citrate was added. This treatment was the real invention of processed cheese. Later, phosphate salts were introduced as emulsifying salts, and the development of the processed cheese industry continued and different patents were granted. In 1916 Kraft established the first commercial factory for processed cheese in the USA. Production and consumption of processed cheese increased steadily and, today, large tonnages of natural cheeses are converted to processed cheese. The production and consumption of processed cheese products showed a constant increase during the period from 1990 to 1994 (**Table 1**), notably in the European Union, the USA, and Japan. Production of processed cheese affects – directly or indirectly – dairy production in general in the following ways:

1. It encourages cheese production as a base for the manufacture of processed cheese.
2. It makes it possible to use second-grade cheese or cheese with mechanical or surface defects.
3. At peak times of cheese production, young cheese could be precooked and stored until it is used for processed cheese. In this way it is possible to control ripening and save the cost of storing fresh cheese for long periods in expensive stores.

Processed cheese products have many advantages over natural cheese:

1. In most cases, processed cheese can be stored without refrigeration.
2. Processed cheese can be produced in different shapes, flavors, and physical properties, e.g., soft, firm, spreadable, in variable and attractive packages, and it has a relatively long shelf-life.
3. It is free from pathogenic microorganisms.

Table 1 Production (A) and consumption (B) of processed cheese (1000 tonnes)

	1990		*1992*		*1994*	
	A	*B*	*A*	*B*	*A*	*B*
Belgium	42.8	14.6	54.9	16.1	51.2	17.8
Denmark	12.5	3.6	13.4	5.2	15.2	5.6
Germany	152.7	112.8	164.5	119.5	164.1	116.6
Spain	32.0	31.8	36.2	35.6	39.2	38.1
France	99.1	57.2	105.7	63.2	119.4	62.4
Ireland	8.6		9.6			
Italy	20.3		20.3		20.2	
Netherlands	33.1	11.6	32.1	10.8	31.2	9.1
Portugal	8.0		3.5			
United Kingdom	20.4	44.7	19.8	47.9	19.7	51.5
European Union	424.5	309.5	460.0	369.3	474.0	376.6
USA	1070.1		1101.3		1147.7	
Japan	74.0	76.0	85.0	87	91.0	94.0

Reproduced from Hetzner E and Richarts E (1996) The international market for processed cheese. *Bulletin of the International Dairy Federation* 316: 16–26, with permission.

Principles of Processed Cheese-Making

Processed cheese is made by heating natural cheese of different types, age, and maturity in the presence of suitable emulsifying salts and with the help of mechanical agitation. In making processed cheese, the insoluble calcium *para*-casein gel of the natural cheese, under the influences of heating and the action of emulsifiers and agitation, is changed to a *para*-casein sol – a homogeneous, flowing mass. This sol is changed again to a gel by the influences of cooling and polymerization forces resulting in processed cheese. By understanding the role of different factors related to processed cheese manufacture, it is possible to control and select the characteristics of the end product – soft, firm, sliceable, spreadable, etc.

Types of Processed Cheese

Processed cheese can be grouped into three main types.

Pasteurized Processed Cheese

Pasteurized processed cheese is made from one or more lots of cheese of the same or different varieties. Water, NaCl, coloring, and flavoring materials may be added. The moisture content of the processed cheese may exceed by not more than 1% the maximum permitted in the natural cheese from which it is made, while the fat content may not be less than that of the natural cheese. Processed cheese bears the name of the natural cheese from which it is made. The processing temperature varies from 80 to 85 °C and the pH of the end product ranges from 5.4 to 5.6. Good processed cheese has a smooth, compact, firm body, and can be cut into slices.

Pasteurized Processed Cheese Foods

Processed cheese food resembles the above processed cheese but contains more moisture and less fat. Optional ingredients may be added, including dried whey, dried skim milk, lactose, and organic acids. The moisture content should not exceed 44% and the fat content should be not less than 23%. The processing temperature is normally 85–90 °C and the pH of the processed product is 5.6–5.8. Processed cheese food has a softer body and milder flavor than processed cheese. (*See* **Lactose**; **Whey and Whey Powders**: Production and Uses.)

Pasteurized Processed Cheese Spread

The constituents of pasteurized processed cheese spread are similar to those used for making processed cheese food, but it contains more moisture in order to achieve a soft body and spreading properties. In many standards, the moisture content should not exceed 60% and the fat should be not less than 20%. This cheese is normally processed at a high temperature (85–95 °C) in traditional cookers, and has a pH between 5.7 and 5.9. Binding agents may be added in amounts not exceeding 0.8% of the finished product to improve water retention. Processed cheese spreads are characterized by a mild flavor, and a smooth, soft body that is easy to spread at room temperature.

Processed Cheese Analog

Recently, a new version of processed cheese has been introduced under the name of imitation or analog processed cheese: the natural cheese and fat are replaced by caseinates, vegetable proteins such as soya bean protein, and a suitable vegetable oil. The processing conditions, i.e., emulsifiers, heating temperature, and pH, are quite similar to that reported for processed cheese spread. (*See* **Casein and Caseinates**: Uses in the Food Industry; **Soy (Soya) Beans**: Processing for the Food Industry; **Vegetable Oils**: Types and Properties.)

Processed cheese analog may be produced either as block or spread and should have good melting properties and good shredding or slicing behavior. A small amount of free oil may be acceptable, particularly for use in pizza making. Rennet casein is commonly used in making processed cheese analog as it gives ideal properties to the cheese, especially casein, which contains 0.1–0.3% lactose and has a particle size of about 30 mesh.

Analog processed cheeses have some advantages compared to processed natural cheese:

- Cheaper ingredients may be used.
- A wide range of suitable proteins is available.
- Ingredients can be stored for a long time without spoilage.
- The properties of the final product can be controlled.
- It is more stable and has greater flavor.

The limits for moisture and fat in processed cheese are quite flexible and may vary from country to country according to specific standards implemented in each country.

Raw Materials for Processing

The main materials used for the manufacture of processed cheese include natural cheese, emulsifiers, water, and other additives.

Cheese

Two main factors should be considered in selecting the cheese for processing – type and degree of maturity.

Type of cheese In general, all rennet-curd cheeses, i.e., soft, semihard and hard cheese, can be used for making processed cheese, but it is common practice to use only hard and semihard cheeses; soft cheeses are only used for flavoring. The selected cheese should be checked for dry matter, fat, pH, protein content, age, and degree of maturity. In many countries, Cheddar cheese is preferred as a main base for processed cheese. A single type of natural cheese can be used but it is more common to use more than one variety of cheese with the aim of giving the desired body, texture, and flavor to the final product. It is worth mentioning that good-quality natural cheese is essential for producing high-quality processed cheese. Second-grade cheese or cheese with mechanical defects can be used, but cheese with a putrid or rancid flavor must not be used, even in small quantities, as the fault will appear in the processed cheese.

Degree of maturity In general, hard cheeses may be grouped according to the degree of maturity:

- Fresh, green or young cheese – 1–2 weeks old.
- Medium ripened cheese – 2–4 months old.
- Ripened cheese – over 4 months old.

Natural cheeses of different degrees of maturity are selected for processing to obtain the required composition, physical properties and flavor in the processed cheese. Young cheese has a high level of intact casein as most of the protein has not been hydrolyzed to soluble components; the ratio between insoluble casein nitrogen and the total nitrogen is called relative casein. The higher the relative casein (90–95% in young cheese), the more stable the resultant processed cheese will be. Processing young cheese results in processed cheese with a long structure, and the body tends to be smooth or firm depending on the moisture content. It is difficult to use only young cheese in processing, as it will result in a processed cheese of 'flat' flavor.

However, the long structure of young cheese can be altered by the action of various factors during processing. These factors are mainly related to the use of suitable emulsifying salts together with prolonged agitation during the thermal process. Changing the long structure of young cheese to a short one with good spreadability is known as creaming. During the creaming process, the large casein particles, with a low level of hydration in young cheese, are split into small casein particles with increased surface area, hydration, and good spreadability. However, excessive hydrolysis of cheese casein when making processed cheese will lead to an overcreaming defect in the cheese texture. Thus, young cheese, with proper processing, is convenient as the main base for making processed cheese spreads containing a high level of fat-in-dry-matter (60–70%); young cheese also contains a high level of intact casein necessary for stability. It is reported that stable processed cheese should contain not less than 12% intact casein. In a recent study in the authors' laboratory processed cheese spreads containing 60–65% fat in dry matter, 60% moistures and pH 5.85 were made from cheese blends containing 9.8–10.2 of intact casein. The resulting spreads had a good body and texture and could be packed in triangular portions without sticking in the foil. The spread was stable (without fat or water separation) through the storage period. As the ripening cheese progresses, its relative casein content decreases as a result of proteolysis; hence, fully ripened cheese is normally added to the processed cheese blend at a level of 10–20%, mainly to give the desired flavor. It is too difficult to use fully ripened cheese as the main base for processing, as its intact casein is already degraded and is not able to form a stable emulsion.

Also, use of too much of a very mature cheese, which contains considerable amounts of low-solubility amino acids as tyrosine, results in precipitation of these free amino acids as white crystals within the processed cheese structure.

Normally medium-ripened cheeses are blended with young cheese in different ratios to make processed cheese spread. If a high percentage of medium-ripe cheese is used in a processed cheese spread blend, there is no need for extensive processing to achieve the necessary creaming properties as the intact casein is, to some extent, already hydrolyzed. Different ratios of young medium, and fully ripened cheese have been reported for the manufacture of processed cheese products. These ratios are not fixed figures, and may vary according to the type, characteristics, and composition of the natural cheese; they may also vary if other dairy products, such as milk or whey powder, are used in the blend. Some typical blends are presented in Table 2.

The storage of cheese for ripening is rather expensive; hence, many attempts have been made to render young cheese, or even rennet curd, suitable for processing a few days or weeks after manufacture. The addition of lipolytic and/or proteolytic enzymes has been investigated, cooking the curd in a lactic acid solution. There has been success in using a hard cheese with a low pH (4.8–5.0) a few days after

Table 2 Suggested blends of natural cheeses for producing processed cheeses

Processed cheese types	*Natural cheese (%)*		
	Young	*Medium-ripened*	*Fully ripened*
Block	50–60	20–30	10–20
Processed cheese foods	30–40	50–60	10
Spread	50–60	30–40	10

manufacture as the main base for a good processed cheese spread, aiming at the release of calcium from the curd. This process has been employed on a commercial scale. (*See* **Enzymes**: Uses in Food Processing.)

Also, ultrafiltered retentates have been used as a part of the blend for processed cheese. Many studies reported that 60–80% of processed cheese blends could be replaced by ultrafiltered retentate.

Emulsifying Salts

Natural cheese is basically an oil-in-water emulsion, stabilized by cheese protein. Heating and agitation, or changing the pH, will affect the protein which may lose all or part of its ability to effect stabilization; defective texture, water and fat separation will occur. In the presence of emulsifying salts, the separated water and fat are reincorporated into the cheese mass, resulting in a homogeneous mixture. As the emulsifiers (melting salts) work mainly on the cheese protein (casein) they have a direct effect on the physical properties of processed cheese. The action of emulsifiers is through the displacement of the calcium of the insoluble *para*-caseinate by sodium ions, resulting in a more soluble casein. The role of emulsifying salts in processed cheese manufacture can be summarized as follows:

1. Removal of calcium from the protein system
2. Dispersion of the protein
3. Hydration and swelling of the protein
4. Control and stabilization of the emulsion system and the pH of the cheese
5. Control of spoilage of the cheese

(*See* **Emulsifiers**: Phosphates as Meat Emulsion Stabilizers; Uses in Processed Foods.)

The major emulsifying salts which are used on a large scale are the sodium salts of citrates, monophosphates, diphosphates, and polyphosphates. Sodium aluminum phosphate is also used. Some emulsifying salts are listed in Table 3.

It is common practice to use a mixture of emulsifying salts rather than one salt to obtain the general desired characteristics. The amount of emulsifiers used in processed cheese is normally 2–3%, depending on type, maturity, and amount of natural cheese in the blend. It has been reported that some emulsifying salts, especially phosphate, have a bacteriostatic effect on some microorganisms and accordingly improve the shelf-life of the processed cheese.

Table 3 Some emulsifying salts used in processed cheese

Type	*pH (1% solution)*
Monosodium citrate	3.75
Trisodium citrate	8.55
Monosodium phosphate	4.5
Disodium phosphate	8.9–9.1
Trisodium phosphate	11.9
Disodium pyrophosphate	4–4.5
Trisodium pyrophosphate	6.7–7.5
Tetrasodium polyphosphate	10.2–10.4
Tetrasodium polyphosphate	9.0–9.5
Sodium hexametaphosphate	6.0–7.5

Water

Water is very important for producing a stable emulsion. Emulsifying salts need water to dissolve and act properly on the casein. Water is added to the blend to give the required water content in the processed cheese. In calculating the amount of water, it should be borne in mind that heating by direct steam injection will add some condensed water to the blend. Water may be added at the start of processing, as in the case of block-processed cheese, or in portions, as for processed cheese spread. The addition of only part of the water at the start of processing increases the concentration and action of the emulsifiers on the casein.

Additives

Optional ingredients may be added to the processed cheese blend for economic reasons, or to improve the shelf-life and quality of processed cheese.

Dried skim milk and dried whey Dried skim milk or whey can be used to replace part of the cheese solids in the blend for processed cheese food or spread. Skim-milk powder and whey powder promote the creaming process and improve spreadability, but the amount must not exceed 10% of the blend in order to avoid a sweet-salty taste, especially when young cheese is used and to prevent browning and crystallization owing to excess lactose in the processed cheese.

Fats Different fats may be added to the blend to increase the fat-in-dry-matter content of the processed cheese. High-fat cream, butter or butter oil may be used as a source of fat. For processed cheese

spread of defined total solids, increasing the fat content causes a decrease in the solids-nonfat (SNF) in the blend; hence, young cheese with a high intact casein should be used as the main base to give a stable emulsion; caseinates may also be used. The emulsifying salts used in processed cheese have no direct effect on the fat. (*See* **Fats**: Uses in the Food Industry.)

Precooked cheese Precooked cheese may be used to improve the stability of processed cheese. It is normally produced at peak times of cheese production when the volume of cheese exceeds the capacity of the ripening rooms, or when the cheese is not fit for maturation. Precooked cheese made from young cheese has a long structure and can be added to a processed cheese blend containing overripened cheese to improve its stability. Precooked cheese made from mature cheese has a short structure and can be used to promote the creaming property of a processed cheese spread. The amount of precooked cheese added to a blend varies according to the degree of maturity of the cheese in the blend and the type of processed cheese to be manufactured.

Flavoring materials Flavoring materials may be added to give a particular flavor to the processed cheese, or to improve the flavor, especially when young cheese is used. Among these materials, meat, ham, wine, fruit, spices, and essences are commonly used in processed cheese. Flavoring materials should be of good microbiological quality, and this can be attained by thermal treatment before addition to the blend. The physical properties of the processed cheese are not affected by the use of such additives. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Binding agents Binding agents or stabilizers are sometimes added to processed cheese spread to absorb some of the water and improve the stability of the cheese. The amount of stabilizers, if permitted, should not exceed 0.8% of the processed cheese. Gum arabic, locust bean gum, gelatin, pectin, carboxymethyl cellulose, and agar are examples of such materials. (*See* **Stabilizers**: Types and Function; Applications.)

Preservatives The addition of preservatives to processed cheese is prohibited in many countries. However, preservatives such as benzoic and sorbic acids are used to overcome the blowing of processed cheese. In addition, biological substances, such as nisin, are used as inhibitory agents to prevent the growth of anaerobic spore-formers (*Clostridium*), the main cause of blowing in processed cheese. Nisin is a polypeptide produced by cultures of some strains of *Lactococcus lactis*, and is produced commercially in powder form. Although nisin is used successfully in many processed-cheese factories, there is some doubt concerning its efficiency against some strains of *Clostridium*. (*See* **Preservation of Food**.)

Manufacture of Processed Cheese

The steps involved in the manufacture of processed cheese can be summarized as follows.

Selection and Calculation of the Raw Materials

Two important factors must be considered in selecting the cheese for the blend: (1) the characteristics of the natural cheese, i.e., type, age, maturity, pH, fat, total solids, and the physical properties; (2) the desired properties for the processed cheese, i.e., firmness and spreadability.

The amount of cheese and other ingredients in the blend are calculated according to their fat and dry-matter contents, so that the product conforms to the legal standards of the specific type of processed cheese.

Cleaning

Hard and semihard cheeses selected for processing are subject to cleaning before use. Very hard rind, waxes, or wrapping materials used to cover rindless cheeses and surface defects must be removed manually or mechanically before processing.

Cutting, Mincing, and Milling

The selected blocks of cheese are cut manually, or mechanically with special knives, into small slices suitable for the mincing machines. The slices are then minced or shredded through special mincing machines into fine granules, which are ground through roller mills to make them soft, smooth, and free from small, hard particles. Mincing and milling of the cheese enable the emulsifying salts to be in contact with the cheese, and facilitate the absorption of water and dispersion of the cheese protein.

Processing

Milled cheese and other ingredients of the blend are placed in the processing kettle. Heating may be carried out directly by steam injection, indirectly, or by both methods to a temperature normally not less than 75 °C to insure complete pasteurization of the processed cheese. Agitation during processing is important for the complete emulsification of the cheese blend. Processing kettles are operated either batchwise or continuously. They are available in various

shapes and capacities. A normal kettle consists of two pans equipped with one movable cover which bears an agitator, steam injection nozzles, vacuum tubes, and gages for temperature and pressure. While processing is carried out in one of the pans, the other one is loaded with raw materials ready for processing.

The temperature and duration of processing are determined by the type of processed cheese and by the condition of natural cheese in the blend. In general, a period of 2–4 min is required to reach the processing temperature, and the temperature is kept constant for a further 5–10 min. Processed cheese spread is processed at 80–85 °C for 8–10 min while block-processed cheese is processed at 85–95 °C for 8–15 min in a traditional kettle. The duration of processing is related to the speed of agitation. Stephan kettles are equipped with very fast agitators, so that a period of 4–5 min is sufficient for processed cheese spread. In modern continuous processed cheese-making, the processing is carried out in horizontal or vertical tubes in which the blend is heated by direct steam to 135–145 °C for a few seconds.

Homogenization

Homogenization is an optional operation used for processed cheese spreads, especially those with a high fat content, and is conducted directly after processing. It improves the consistency, stability, and appearance of the processed cheese. Contamination during homogenization should be avoided.

Packaging

The hot, flowing mass of processed cheese is transported manually in stainless-steel buckets or by special pumps to the filling machines. Keeping the processed cheese quite hot in the filling machines eases the sealing of the packaging materials. Processed cheese is packed in thin aluminum foil coated with a special lacquer, in plastic membranes, or in tins; it is packed in different shapes in amounts varying from 13 g to 3 kg.

Storage

Directly after packing, the processed cheese is still quite warm (40–60 °C), but it should be cooled to room temperature in a time that varies according to the type of processed cheese. As a rule, processed cheese spread should be cooled quickly (30–60 min) to prevent the creaming phenomenon; otherwise, overcreaming will cause fat and water separation and a defective texture. In cooling processed cheese spread, a space should be left between the boxes to facilitate the movement of cool air. For block-processed cheese, a long time (10–15 h) is needed to cool it down, but the slow cooling, especially in the presence of lactose, may result in browning or firming-up of the cheese, and may allow the growth of spore-forming microorganisms. After the cheese has been cooled, it should be kept in a cold store (5–10 °C). It is advisable to keep processed cheese above 0 °C – to prevent the freezing of the product – and below 20 °C; otherwise, surviving microorganisms, especially *Clostridium*, may grow and cause defects. (*See* **Spoilage**: Molds in Spoilage.)

Defects in Processed Cheese

The two main defects in processed cheese are microbiological defects and those of physicochemical origin.

Microbiological Defects

Microbiological defects are mainly related to the blowing of the processed cheese as a result of gas formation by anaerobic spore-formers (*Clostridium*). Spongy texture and bad odor are indicators of a heavy contamination with *Clostridium*.

In addition, molds may grow on the surface of poorly sealed packs of processed cheese. However, proper processing and the addition of permitted antimicrobial materials should minimize or eliminate these defects.

Defects of Physicochemical Origin

Too firm a block results from one or more of the following:

- low moisture content
- high casein content
- use of the wrong emulsifiers
- strong overcreaming
- low pH

Too soft a block may occur as a result of:

- high moisture content
- unsuitable or insufficient emulsifiers
- too high pH
- excessive processing
- addition of milk or whey powder
- use of too much mature cheese

Inhomogeneous cheese could be the result of:

- insufficient milling, especially of cheese with a hard rind
- improper processing time, temperature, or agitation
- unsuitable or insufficient emulsifiers
- use of raw materials of too low pH, particularly an acid-precipitated cheese or curd

Gum-like spread could be caused by:

- use of too much young cheese
- insufficient processing to achieve proper creaming
- absence of creamed, precooked cheese
- addition of all the water at one time

'Sticky spread' is the term used when the processed cheese adheres to the aluminum foil. This defect may be attributable to:

- excess water content or all the water added at once
- the blend contains too much young cheese
- proper creaming is not achieved
- no well-creamed, precooked cheese has been added
- the level of intact casein in the blend is insufficient

Reasons for brittle spread are:

- overcreaming during processing
- addition of too much overcreamed, precooked cheese
- decrease of pH after processing or during storage

Gas formation As result of a chemical reaction between the aluminum foil and the cheese, the resultant hydrogen can cause gas 'holes' on the cheese surface without odor formation. This defect can be attributed to the absence of, or use of poor-quality, lacquering materials or to the use of an emulsifying salt which is too acidic or too alkaline.

Browning (Maillard reaction) The normal color of processed cheese is between white and pale yellow. Browning may occur directly after processing, or during storage as a result of a reaction between amino compounds and reducing sugars. This defect may be caused by:

- Use of too high a processing temperature for a long time, especially in the presence of lactose.
- Storage of the processed cheese at a high temperature, particularly when the pH is high.

This defect is more common in processed cheese spread than in block-type cheese. (*See* **Browning**: Nonenzymatic.)

Crystallization Crystallization is a defect which can be seen inside or on the surface of the cheese; a 'sandy' texture may result. The reasons for this defect may be:

- precipitation of calcium monophosphate, diphosphate, polyphosphate or citrate, particularly when excess emulsifier has been used
- undissolved particles of emulsifier, owing to poor storage of the emulsifier or improper processing
- use of precooked cheese with a sandy texture
- precipitation of lactose crystals as a result of too much dried milk or whey in the blend
- use of very mature cheese may result in white precipitates of some amino acids, such as tyrosine

Water separation Water separation can take the form of small droplets inside the cheese, or wetting of the cheese surface. It is caused by the following factors:

- too low pH of the processed cheese
- changes in the cheese structure (overcreaming)
- unsuitable storage conditions, e.g., the cheese is subjected to mechanical pressure

Fat separation Fat separation could be the result of the following:

- use of too much mature cheese
- insufficient or excess emulsifiers
- too low a pH of the processed cheese
- storage of the cheese for a long time at a high temperature

Flavor Defects

Examples of the more common flavor defects are:

- sharp flavors, owing to use of too much mature cheese
- flat flavor, owing to use of too much young cheese
- salty taste, owing to use of salty cheese or excess emulsifiers
- putrid taste, owing to use of putrid cheese or the growth of *Clostridium*
- rancid taste, owing to use of rancid cheese and/or butter, or cheese ripened by molds
- chemical taste, owing to use of impure emulsifiers, addition of certain preservatives or stabilizers, or use of too salty cheese

See also: **Browning**: Nonenzymatic; **Casein and Caseinates**: Uses in the Food Industry; **Cheeses**: Types of Cheese; **Emulsifiers**: Phosphates as Meat Emulsion Stabilizers; Uses in Processed Foods; **Enzymes**: Uses in Food Processing; **Fats**: Uses in the Food Industry; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Lactose**; **Preservation of Food**; **Spoilage**: Molds in Spoilage; **Stabilizers**: Types and Function; Applications; **Vegetable Oils**: Types and Properties; **Whey and Whey Powders**: Production and Uses

Further Reading

Abou El-Nour A, Scheurer GJ, Omar MM and Buchheim W (1996) Physiochemical and rheological properties of block-type processed cheese analogue made from rennet casein and total milk protein. *Milchwissenschaft* 51: 684–687.

Caric M, Ganter M and Kalab M (1985) Effect of emulsifying agents on the microstructure and other characteristics of processed cheese – a review. *Food Microstructure Vollum* 4: 297–312.

Fox PF (ed.) (1989) *Cheese: Chemistry, Physics and Microbiology*, vol. 2. London: Elsevier Science.

Gouda A, Ahmed HA, El-demerdach HA and Abd El-Fadeel MG (1998) Effect of using acidified Ras cheese on processed cheese quality. *Annals of the Agricultural Science of Moshtohor* 36: 2273–2279.

Gupta VK and Reuter H (1992) Processed cheese foods with added whey protein concentrates. *Lait* 72: 201–212.

Hetzner E and Richarts E (1996) The international market for processed cheese. *Bulletin of the International Dairy Federation* 316: 16–26.

Kosikowski FV (1978) *Cheese and Fermented Milk Foods*. Ann Arbor, Michigan: Edwards Brothers.

McCarthy J (1990) Imitation cheese products. *International Dairy Federation Bulletin* 249: 45–52.

Muir DD, Tamime AY, Shenana ME and Dawood AH (1999) Processed cheese analogues incorporating fat-substitutes. I. Composition, microbiological quality and flavour changes during storage at 5 °C. *Lebensmittel Wissenschaft und Technologie* 35: 41–49.

Rayan AA, Kalab M and Ernston CA (1980) Microsturucture and rheology of processed cheese. *Scanning Electron Microscopy* III: 635–643.

Shimp LA (1985) Processed cheese principles. *Food Technology* 39(5): 63–70.

Thomas MA, Newell G, Abad GA and Turner AD (1980) Effect of emulsifying salts on objective and subjective properties of processed cheese. *Journal of Food Science* 45(3): 458–460.

Dietary Importance

J Buttriss, British Nutrition Foundation, London, UK

Introduction

Cheese is the oldest way of preserving the nutrients in milk and is essentially a concentrated form of milk. Most cheese is made from cows' milk, but other milks including sheep's, goats', and buffalos' can be used also. The process used to make the cheese tends to determine the nutritional value. This article will review the nutritional value of different types of cheese, its role in the diet for different groups of the population, and trends in cheese consumption around the world.

Hard Cheeses

Hard cheeses such as Cheddar and Cheshire retain most of the protein, fat, calcium, and some other minerals, and vitamins such as vitamin A, riboflavin (vitamin B_2), and vitamin B_{12}. **Table 1** shows the nutrient composition of some hard cheeses. Refer to individual nutrients.

Hard cheese contains very little lactose as most is either lost with the whey during the cheesemaking process or converted to lactic acid. The major protein in cheese is casein, which is a high-quality protein containing all the essential amino acids in roughly the proportions required by the body for health. Minerals in cheese, such as calcium and zinc, are particularly well absorbed and utilized (high bioavailability).

Low-fat hard cheeses are available in the UK. They typically contain about half the quantity of fat found in the traditional cheese and remain a valuable source of essential nutrients.

Soft Cheeses

Soft cheeses vary widely in nutritional composition, depending on whether they belong to the ripened or unripened varieties. Fresh, unripened cheeses, such as cottage cheese, are low in fat, relatively low in calcium, high in moisture and contain unfermented lactose. Very-low-fat versions are also available. However, a surface-mold-ripened cheese, such as Brie or Camembert, contains a high proportion of fat and protein and less water (**Table 2**). Mold-ripened soft cheeses have a slightly lower fat content than traditional hard cheeses. The addition of salt is an essential part of the cheese-making process. The salt is added to help preserve the cheese and to bring out the flavor.

Cheese from Unpasteurized Milk

Most of the cheese produced in the UK is made from pasteurized milk. However, there is a demand for cheese made from unpasteurized milk, as it is believed by some to have a superior flavor. Provided that the milk used has come from farms operating to the highest standards, so that the milk is free of pathogens, and that the creamery operates to the same high standards, the resulting cheese will not pose a health hazard.

Table 1 Nutritional composition (per 100 g) of some hard cheeses

	English Cheddar	*Cheshire*	*Blue Stilton*	*Edam*	*Reduced-fat Cheddar*
Energy					
(kcal)	412	379	411	229	261
(kJ)	1708	1571	1701	957	1091
Protein (g)	25.5	24.0	22.7	32.6	31.5
Carbohydrate (g)	0.1	0.1	0.1	Trace	Trace
Sugars (g)	0.1	0.1	0.1	Trace	Trace
Fat (g)	34.4	31.4	35.5	25.4	15.0
Saturates (g)	21.7	19.6	22.2	15.9	9.4
Monounsaturates(g)	9.4	9.1	10.3	7.4	4.4
Polyunsaturates (g)	1.4	0. 9	1.0	0.7	0.4
Sodium (mg)	670	550	930	1020	670
Dietary fiber (g)	Nil	Nil	Nil	Nil	Nil
Vitamin A (μg)	336	387	386	200	182
Thiamin (mg)	0.03	0.03	0 .03	0.03	0.03
Riboflavin (mg)	0.42	0.48	0.43	0.35	0.53
Nicotinic acid (mg)	0.09	0.11	0.49	0.07	0.09
Potential nicotinic acid from tryptophan (mg)	6.00	5.64	5.34	6.12	7.41
Vitamin B_6 (mg)	0.10	0.09	0.16	0.09	0.13
Folic acid (μg)	37	40	77	40	56
Vitamin B_{12} (μg)	1.2	0.9	1.0	2.1	1.3
Pantothenic acid (mg)	0.38	0.31	0.71	0.38	0.51
Biotin (μg)	3.1	4.0	3.6	1.8	3.8
Vitamin C (mg)	Trace	Trace	Trace	Trace	Trace
Vitamin D (μg)	0.26	0.24	0.27	0.19	0.11
Vitamin E (mg)	0.54	0.70	0.61	0.48	0.39
Vitamin K (μg)	N/A	N/A	N/A	N/A	N/A
Calcium (mg)	740	560	320	770	840
Chlorine (mg)	1010	830	1410	1570	1110
Copper (mg)	Trace	0.13	0.18	0.05	0.05
Iodine (μg)	46	(46)	(46)	N/A	N/A
Iron (mg)	0.2	0.3	0. 3	0.4	0.2
Magnesium (mg)	26	19	20	39	39
Phosphorus (mg)	490	400	310	530	620
Potassium (mg)	79	87	130	97	110
Selenium (μg)	12	(11)	(11)	N/A	15
Zinc (mg)	2.3	3.3	2. 5	2.2	2.8

N/A: no reliable figures available. ( ): estimate.
Source: Holland B, Unwin ID and Buss DH (1989) Milk Products and Eggs. The fourth supplement of McCance and Widdowson's *The Composition of Foods* 4th edn. Cambridge: Royal Society of Chemistry, and Ministry of Agriculture, Fisheries and Food, with permission.

Tyramine in Cheese

In common with foods such as red wine, hung game, soused herrings, and yeast extracts, certain cheeses (e.g., mature Cheddar, Roquefort, and Gruyère) contain an amino acid derivative known as tyramine.

In sensitive individuals, tyramine can provoke migraine and skin rashes. This may occur, for example, in those taking drugs that block the enzyme monoamine oxidase, which metabolizes tyramine. (*See* **Migraine and Diet.**)

Trends in Cheese Consumption

Table 3 shows the consumption of cheese in different countries. Compared with most European countries, cheese consumption in the UK is low. Consumption is particularly high in France and The Netherlands.

In the UK, cheese consumption has risen from 4.7 kg per person per year in 1966 to 9.8 kg per person per year in 1998, according to National Dairy Council estimates (Figure 1). Data from the National Food Survey, conducted by the UK Ministry of Agriculture, Fisheries and Food, indicate that cheese consumption is greater in higher income groups and smaller households. There are no strong regional differences.

Importance of Cheese for Different Groups

As a product of milk, cheese confers similar nutritional benefits. In particular, it is a rich source of protein, calcium and vitamins A, B_2 (riboflavin), and B_{12}. It is also a useful source of highly bioavailable

Table 2 Nutrient composition (per 100 g) of soft cheese

	Soft, fresh				*Surface-mold-ripened*	
	Cottage cheese: plain	*Fromage frais: plain*	*Fromage frais: plain, very low fat*	*Quark*	*Brie*	*Camembert*
Energy						
(kcal)	98	113	58	74	319	297
(kJ)	413	469	247	313	1323	1232
Protein (g)	13.8	6.8	7. 7	14.6	19.3	20.9
Carbohydrate (g)	2.0	5.4	6 .5	3.8	Trace	Trace
Sugars (g)	2.0	5.4	6. 5	3.8	Trace	Trace
Fat (g)	3.9	7.1	0. 2	Trace	26.9	23.7
Saturates (g)	2.4	4.4	0.1	Trace	16.8	14.8
Monounsaturates(g)	1.1	2.1	0.1	Trace	7.8	6.9
Polyunsaturates (g)	0.1	0.2	Trace	Trace	0.8	0.7
Sodium (mg)	380	31	(33)	45	700	650
Dietary fiber (g)	Nil	Nil	Nil	Nil	Nil	Nil
Vitamin A (μg)	46	100	3	2	320	283
Thiamin (mg)	0.03	0.04	(0.03)	(0.04)	0.04	0.05
Riboflavin (mg)	0.26	0.40	(0.37)	0.30	0.43	0.52
Nicotinic acid (mg)	0.13	0.13	(0.14)	0.19	0.43	0.96
Potential nicotinic acid from tryptophan (mg)	3.24	1.59	1.81	3.43	4.53	4.91
Vitamin B_6 (mg)	0.08	0.10	(0.07)	0.08	0.15	0.22
Folic acid (μg)	27	15	(15)	45	58	102
Vitamin B_{12} (μg)	0.7	1.4	(1.4)	0.7	1.2	1.1
Pantothenic acid (mg)	0.40	N/A	N/A	0.44	0.35	0.36
Biotin (μg)	3.0	N/A	N/A	3.0	5.6	7.6
Vitamin C (mg)	Trace	Trace	Trace	1.0	Trace	Trace
Vitamin D (μg)	0.03	0.05	Trace	Trace	0.20	0.18
Vitamin E (mg)	0.08	0.02	Trace	Trace	0.84	0.65
Vitamin K (μg)	N/A	N/A	N/A	N/A	N/A	
Calcium (mg)	73	89	(87)	120	540	650
Chlorine (mg)	550	100	(89)	110	1060	1120
Copper (mg)	0.04	Trace	(0.01)	0.06	Trace	0.07
Iodine (μg)	N/A	N/A	N/A	4	N/A	N/A
Iron (mg)	0.1	0.1	(8)	Trace	0.8	0.2
Magnesium (mg)	9	8	(110)	11	27	2.1
Phosphorus (mg)	160	110	(8)	200	390	310
Potassium (mg)	89	110	(110)	140	100	100
Selenium (μg)	(4)	(2)	(2)	N/A	N/A	N/A
Zinc (mg)	0.6	0.3	(0.3)	0.9	2.2	2.7

N/A: no reliable figures available. ( ): estimate.

Source: Holland B, Unwin ID and Buss DH (1989) Milk Products and Eggs. The fourth supplement to McCance and Widdowson's *The Composition of Foods* 4th edn. Cambridge: Royal Society of Chemistry, and Ministry of Agriculture, Fisheries and Food, with permission.

zinc. In addition, cheese is a very versatile food that can be incorporated into many different dishes. It keeps well if properly stored in the refrigerator, wrapped in greaseproof paper or foil. Hard cheeses such as Cheddar freeze well and can be kept for up to 3 months. On thawing, best results are achieved if the cheese is allowed to defrost gradually, preferably in the refrigerator.

Mild hard cheeses and fresh soft cheeses such as cottage cheese can be introduced into a child's diet from the age of 6 months. Cheese can be a useful snack food for children. It is one of the few foods that do not contribute to dental caries, and it may even have a protective effect against dental decay, perhaps by controlling the pH at the surface of the tooth. Some dentists recommend a small piece of cheese at the end of a meal for this reason. (*See* **Dental Disease:** Role of Diet.)

Calcium requirements are particularly high during pregnancy, and cheese, an excellent source of calcium, is therefore a valuable food for pregnant women. However, as a precaution, pregnant women are advised to avoid mold-ripened soft cheese such as Brie and Camembert because of the increased chance of their containing *Listeria monocytogenes*, the bacterium that causes listeriosis. Most healthy people are unaffected by listeriosis, but pregnant women are at a greater risk. Fresh soft cheeses, e.g., cottage cheese,

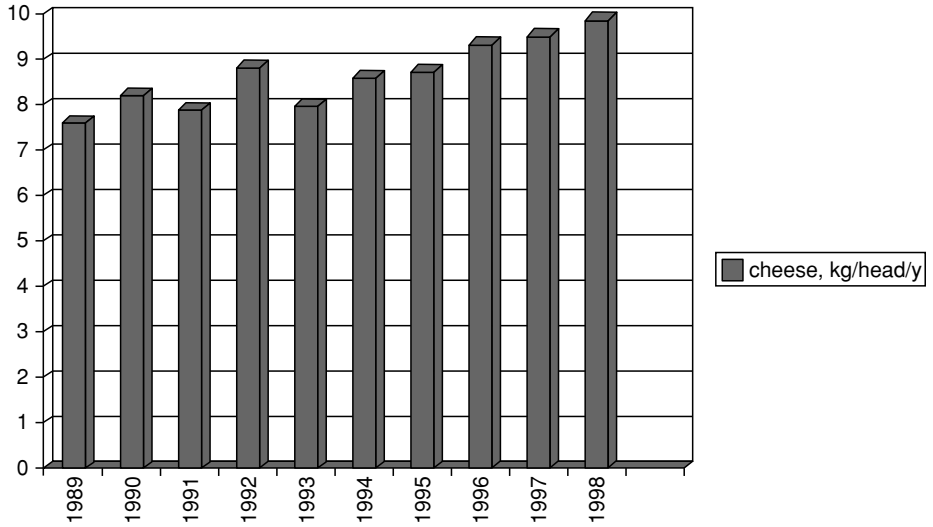

Figure 1 Trends in consumption of cheese in the UK. Source: National Dairy Council (1999) *Diary Facts and Figures*. London: NDC.

Table 3 Estimates of cheese consumption in a selection of countries, 1998

	Cheese consumption (kg per person per year) in 1998
Australia	8.9
Canada	10.8
Argentina	10.9
Hungary	7.1
Germany	20.5
Denmark	16.4
Finland	17.0
France	23.6
Ireland	8.3 (1997 figure)
Israel	16.5
Iceland	18.8
Italy	19.0
Japan	1.6 (1997 figure)
The Netherlands	16.6
Norway	14.2
South Africa	0.9
Spain	8.4
Sweden	16.8
UK	9.7
USA	13.1

Source: International Dairy Federation (1999) *The World Dairy Situation 1998*. Bulletin of the IDF 339, 52, with permission.

and hard cheeses such as Cheddar, are considered safe in this respect. The pH of Cheddar and other such cheeses limits the growth of this bacterium, which prefers more alkaline conditions. *See* ***Listeria***: Listeriosis.

For vegetarians who include dairy products in their diet, cheese can be an important source of protein, vitamin B_{12} and minerals. Vegans, who exclude all foods of animal origin, will require vitamin-B_{12}-fortified foods or supplements, since the vitamin is found only in dairy foods and other foods of animal origin, such as meat. (*See* **Vegetarian Diets**.)

Hard cheeses, such as Cheddar, have a very low lactose content and can be a useful source of calcium and other essential nutrients for those people who are lactose-intolerant. (*See* **Food Intolerance**: Lactose Intolerances.)

See also: **Dental Disease**: Role of Diet; **Food Intolerance**: Lactose Intolerance; ***Listeria***: Listeriosis; **Migraine and Diet**; **Vegetarian Diets**

Further Reading

International Dairy Federation (1999) The World Dairy Situation 1998. *Bulletin of the IDF* 339: 52

Ministry of Agriculture, Fisheries and Food (1999) *Household Food Consumption and Expenditure, 1998*. London: The Stationery Office.

National Dairy Council (1999) *Dairy Facts and Figures*. London: NDC.

Renner E (1983) *Milk and Dairy Products in Human Nutrition*. Munich: Volkswirtschaftlicher.

Renner E (1989) *Micronutrients in Milk and Milk-based Food Products*. London: Elsevier Applied Science.

Scott R (1981) *Cheesemaking Practice*. London: Applied Science.

Mold-ripened Cheeses: Stilton and Related Varieties

P A Atkinson, Formerly of Long Clauson Dairy Ltd, Melton Mowbray, UK

Stilton and Related Varieties

Varieties of internally mold-ripened, hard-pressed cheeses include Stilton, Blue Shropshire, Blue Cheshire, Blue Wensleydale, Buxton Blue, Blue Vinney, Danish Blue, Roquefort, and Gorgonzola. All these varieties of blue cheese are considered to be hard cheeses, although only the Blue Cheshire and Blue Wensleydale are actually pressed during manufacture. The other varieties are turned regularly during manufacture, the only pressure being the weight of the curd.

Specifications and Standards

The 1970 Cheese Regulations in the UK require that all full-fat hard cheese shall contain not less than 48% milk fat in dry matter, and not more than 48% moisture. In addition, the standards for Stilton are stated in Schedule 2 of these regulations as follows:

	Fat in dry matter	Moisture
Blue Stilton	48%	42%
White Stilton	48%	46%

The Stilton Cheese Makers Association (SCMA) has registered the name of Stilton and defined Blue Stilton as follows:

> A fully matured, blue-veined Blue Stilton Cheese is an English made cheese produced within the Counties of Leicestershire, Derbyshire and Nottinghamshire with no applied pressure, forming its own crust or coat and made in cylindrical form from full cream milk produced in English and Welsh Dairyherds.

The SCMA also describes flavor, internal and external body, and shape. These will be discussed in more detail below when grading is described.

Method of Manufacture – Stilton Cheese

Stilton cheese is made at nine factories within the three counties stipulated in the SCMA schedule, and the overall process is shown in Table 1.

It is made from cows' milk produced locally, except in times of peak production when cows' milk may be used from anywhere within England and Wales. The milk is collected in bulk and it is now all pasteurized (71.6 °C/15 s); the last factory stopped using raw milk in 1989. After pasteurization the milk is cooled to between 26 and 32 °C according to season and individual factory variations, and put into cheese vats. (*See* **Milk**: Processing of Liquid Milk.)

Table 1 Flowchart of the manufacture of Stilton

	Time/age	*Temperature*
Milk pasteurization	0	
Starter addition	0	26–32 °C
Rennet addition	1–2 h	26–32 °C
Cutting	1–2 h	26–32 °C
Whey separation	1–2 h	26–32 °C
Milling and salting	24 h	26–32 °C
Mold filling and turning		
Maturation	5–7 days	26–30 °C
Coating		26–30 °C
Maturation: white store	5–10 days	13–15 °C
Maturation: blue store	6–12 weeks	13–16 °C
Piercing	6–7 weeks	13–16 °C
Grading	9–10 weeks	
Extended maturation	As directed by flavor	5 °C
Packing and distribution		8 °C

The temperature of milk in the vat at this stage has a very strong influence on the texture of the curd formed and, as a result, will have a marked effect on the openness of the curd when the cheese is made. Seasonal and geographical variations in milk composition require variations in the cheese-making technique to give consistent cheese quality; one of these is variation in milk temperature.

Starter cultures are added to the milk in the vats at varying rates from 50 ml to 2 l per 4500 l (1000 gallons), or as recommended by the supplier if concentrated (direct vat inoculation, DVI) cultures are used. Considerable variations in starter culture are used by the different manufacturers of Stilton, but one consistent factor is slow, but persistent, acid development throughout the early stages of the cheese-making process. The presence of *Lactococcus lactis* biovar. *diacetylactis* and *Leuconostoc* spp. in low numbers in the culture can help to achieve this. (*See* **Cheeses**: Starter Cultures Employed in Cheese-making.)

Spore suspensions of *Penicillium roqueforti* are added to the vat at the same time as the starter culture. The strains used vary with different manufacturers and can have an effect on the final flavor of the mature cheese. The cheese-makers' judgment in the selection of mold strains is very important in ensuring consistent product quality and market acceptance.

Rennet is added to the milk after a period of ripening, usually of 1–2 h. As the starter addition rate is very low, there is a negligible rise in acidity. This necessitates an increase in the quantity of rennet

used to give a satisfactory coagulation. The rate of rennet addition is between 1.18 and 1.48 l per 4500 l of milk (0.025–0.03%). This compares with 0.88–1.03 l per 4500 l for normal cheddar cheese manufacture, where a very much higher starter inoculation is used, and higher acidities are achieved. Animal rennet is generally used, although a 'vegetarian Stilton' is now being produced in which microbial rennet is used. Some manufacturers have suggested that cheese of satisfactory quality can be made with microbially derived coagulants, but that the texture of the cheese is not the same as that made from calf rennet. (*See* **Cheeses**: Chemistry of Gel Formation.)

The curd is cut using American knives or nylon grids which give cubes with 1.25 cm sides. The cutting time is determined by the experience and skill of the cheese-maker and is influenced by the firmness of the curd. After cutting, the curds and whey separate and the curd sinks to the bottom of the vat. Over a further period of usually 1–2 h, syneresis continues and, when the curd occupies about half the volume of the original milk, the curd may be stirred to aid subsequent whey drainage. A further period is then allowed for whey separation. At this stage the whey is removed progressively, taking 12–18 h before the last of the whey is drained away. As the whey is drained, the curd is cut into blocks to aid drainage. At the end of the draining period the titratable acidity will have risen to between 1.1 and 1.3% lactic acid, and the curd texture should be firm but crumbly (easily broken).

The curd is then passed through a peg mill which breaks it into pieces 1.25–2.5 cm in diameter. Salt is added to the curd and mixed evenly at the rate of 2½% w/w. After salting, the curd is put into cylindrical, open-ended molds (22.8–25 cm diameter), with 9–11 kg of curd per mold, depending on the final weight of each cheese. At this stage the curd is put into the molds with no pressure. The presence of the salt increases syneresis and curd drainage. (*See* **Cheeses**: Chemistry and Microbiology of Maturation.)

After 1–2 h drainage the molds are turned over. This first turn is quite critical and failure to turn at this stage can result in uneven drainage of the cheese at later stages in the maturing process. The cheeses are then turned daily for a further 5–7 days, allowing migration of the salt through the curd and further whey drainage. During this period the curd becomes much firmer and, at the end of this period, the cheese will slide easily within the mold when it is turned.

The mold is then removed and the cheeses are rubbed with the flat side of a knife to give a smooth coat over the whole surface. This is thought to prevent ingress of oxygen at too early a stage in the ripening process.

The temperature and humidity from filling the vat with milk until the cheeses have been rubbed or coated are maintained at 26–30 °C and 90% relative humidity (RH). This insures that the curd does not get cold and does not dry out and become discolored. It also influences the microflora of the finished cheese by encouraging the growth of starter organisms, particularly the slower-growing *Lactococcus lactis* biovar. *diacetylactis*.

After coating, the cheeses are transferred to the first maturing room or 'white room', so called because the cheeses remain in this room for 5–10 days, and at the end of this time those produced for White Stilton (without the blue mold being added) may be sold. Cheeses are turned daily at this stage.

The humidity in the white room is maintained at 85–90% RH, but the temperature is reduced to 13–15 °C. This allows the cheese to cool and the coat to dry. Humidity control at this stage is important as too low an RH would cause excessive drying and the coats would crack. Too high an RH would result in wet coats which would encourage the wrong microflora and could result in off-flavors developing; for example, *Brevibacterium linens* frequently occurs on wet-coated cheese and can cause very uncharacteristic flavors to develop. Also, as the natural coat flora begins to develop and get thicker, a condition known as 'slip coat' can occur. In this case, proteolysis occurs just beneath the coat and causes the whole coat to slip down the cheese and fall off, so making the cheese unsaleable.

After 5–10 days in the white room, the cheeses are transferred to a blue maturing room where turning continues daily, but may become less frequent as the coat develops and the cheeses start to dry out. The temperature and humidity levels are controlled at 13–16 °C and 90% RH in these stores.

At approximately 6–7 weeks after production (from the day of rennetting), the cheeses are pricked, spiked, or skewered. This operation involves piercing the cheese with stainless-steel needles to allow air into the cheese, thus encouraging growth of the *Penicillium* spores put into the milk.

Up to this stage the pH within the cheese has initially dropped to 4.8–5.0, but, as the mold grows, the pH rises. In addition, the temperature rises as a result of the metabolic activity within the cheese. Lipolysis and proteolysis occur, releasing carbon dioxide and ammonia. The ammonia can frequently be detected in the atmosphere in the maturing rooms at levels up to 30–40 p.p.m. When this occurs it is necessary to increase the ventilation in order to reduce it to an acceptable level.

As the maturing process continues, the piercing is usually carried out twice more with between 100 and

300 needle insertions at each skewering. In order to insure a good blue mold development, it is essential to have created an open-textured curd which will allow easy penetration of the mold hyphae between the curd particles. This will give the characteristic blue veining effect which radiates out from the center of the cheese where the spike holes are most dense.

After a period of 2–3 weeks from the onset of mold growth, the blue veining should have radiated out to the edge of the cheese. At this stage the cheeses are transferred to a cool store at 2–5 °C where mold growth will be somewhat retarded. Flavor and texture continue to develop and the byproducts of fat and protein degradation accumulate in the cheese.

With the advent of the Food Hygiene Amendment Regulations (1989) and the need to store mold-ripened cheese at temperatures below 8 °C (5 °C from 1993), the flavor development in cold storage is slowed down, but still continues and can give a strong mature flavor (and texture) after a further 2–6 weeks of storage.

Grading

Grading usually takes place at 8–9 weeks and involves examination of every cheese for the following characteristics.

External

- Shape – should be regular and uniform.
- Coat – should be dry, firm, slightly wrinkled, intact, drab in color, and free from surface mold blemishes.

Internal

- Flavor should be clean and mild.
- Body should be uniform creamy white except for blue-green mold radiating from the center.
- Texture should be velvety or flaky and open with an absence of discoloration, gas holes, or chalkiness.

The variability within Stilton makes it necessary to examine each cheese individually. It is not possible to take a random sample out of a larger batch, as is normally done with the other UK hard cheeses.

Modern Approaches to Manufacture

The current trend in the marketplace is towards a more mature-flavored Stilton, which has been described as 'vintage' or 'extra mature.' This is Stilton which has been allowed to develop a stronger flavor either in the maturing stores or in cold stores.

In recent years, some manufacturers have moved away from the manual handling of the curd at the stage where it is being made in the vat. Fully enclosed vats have been used which allow mechanical cutting of the curd followed by very gentle transfer of the curds and whey into drainers where the curd is separated. Another alternative involves vats that open completely at one end, allowing the curd to move on to draining tables. However, the most successful innovation in this area has been the transfer of curds and whey from vats to drainers through pipes.

Mold-turning systems have been tried in various forms ranging from very sophisticated pallet-handling systems to more simple pallet inverters. They all work on the principle that it is quicker and cheaper to turn pallets containing 64 cheeses than to turn each individual cheese by hand.

Packing

Traditionally, Stilton was always sold as a whole cheese wrapped in vegetable parchment. However, in recent years, the proportion of prepacked cheese for supermarket sales has increased to the point at which prepacked Stilton accounts for more than 50% of sales.

After cutting into triangular portions from the round cheese, prepacked Stilton is almost always wrapped in polyvinylidene chloride (PVDC) film; it is always sold with the crust on. In recent years, there has been some interest shown in alternative packing, and modified-atmosphere packing (MAP) has been tried. Alternative films with greater oxygen permeability have also been tried. (*See* **Chill Foods**: Effect of Modified-atmosphere Packaging on Food Quality; **Chilled Storage**: Use of Modified-atmosphere Packaging.)

The problem with Stilton cheese is that the absence of oxygen eventually causes the loss of the blue color, which progressively changes through green to yellow and then disappears. This means that vacuum packing in barrier bags has a very limited application. The other alternative is an oxygen-permeable barrier or a loose wrap which allows oxygen in and thus maintains the blue mold. This will also support the growth of any other aerobic organisms which may be present; this includes *Penicillium roqueforti*, which is quite acceptable as blue veining but is unacceptable as adventitious growth on the cut surface of the cheese. This latter growth is usually seen after 4–5 weeks under refrigeration.

The reduction in product temperature during storage has had the effect of extending the shelf-life slightly, but it only slows the adventitious growth of the mold, it does not stop it.

Another method of packing Stilton that has gained considerably in popularity in recent years is putting it into ceramic jars with a wax or plastic seal. This involves breaking the cheese into small pieces and pushing them into the jars. Alternatively, some companies are using an extrusion technique to fill the jars. It has always been found necessary to put a note on the pot/jar advising the customer that the blue veining will have disappeared in storage, and that removal of the wax seal will result in it returning after about 1 h.

Hygiene

Increased awareness of the need for high standards of hygiene in food production and distribution has resulted in a great deal of legislation being introduced. The food industry has also introduced a number of Codes of Practice to guide manufacturers through the pitfalls of *Listeria* and *Salmonella*. (*See* ***Listeria***: Properties and Occurrence.)

Stilton cheese is covered by the Code of Practice for the Production of Soft and Mould-Ripened Cheese which has been produced by the UK Dairy Trade Federation.

The result of manufacturers' increasing awareness of the problems of bacterial contamination has had the effect of making dramatic changes in the construction of buildings and the training and control of staff. Some aspects of the manufacturing procedures have come under very close scrutiny; for example, the last factory producing cheese from raw milk has now decided to pasteurize its milk. It is argued that, in some cases, higher hygiene standards are being achieved at the expense of a loss of 'character' and 'individuality' in some of the speciality cheeses, such as Stilton.

Related Varieties

Although a number of other internally mold-ripened cheeses can be found worldwide, the most famous are Roquefort and Gorgonzola. The origins of Roquefort can be traced back with some degree of certainty to the fifteenth century, and the name is now internationally recognized for a cheese made from sheeps' milk and matured in huge limestone caves in the region of Aveyron, France. The method of manufacture is broadly similar to that of Stilton, except that the milk is not pasteurized. The strain of *P. roqueforti* may also be unique to the region, but the differences in flavor between Roquefort and Stilton arise mainly from the higher content of short-chain fatty acids associated with sheeps' milk. (*See* **Sheep**: Milk.)

Gorgonzola derives its name from a village in Italy where it is supposed to have originated around 1000 years ago. It is made from cows' milk, and again derives its veining from a strain of *Penicillium*.

More recent in origin is Danublu or Danish Blue cheese, which is extensively manufactured in Denmark, North America, and Australia. The process is, in essence, a more industrial version of Stilton manufacture, in that mechanization is heavily involved, and a more limited ripening period insures a rapid throughput of the product; as a result, connoisseurs claim that it lacks the unique character of the more traditional blue cheeses. (*See* **Cheeses**: Dietary Importance.)

See also: **Cheeses**: Starter Cultures Employed in Cheesemaking; Chemistry of Gel Formation; Chemistry and Microbiology of Maturation; Dietary Importance; **Chill Foods**: Effect of Modified-atmosphere Packaging on Food Quality; **Chilled Storage**: Use of Modified-atmosphere Packaging; ***Listeria***: Properties and Occurrence; **Milk**: Processing of Liquid Milk; **Sheep**: Milk

Further Reading

Davis JG (1976) *Cheese*, vols 1–4. London: Churchill Livingstone.

Robinson RK (ed.) (1990) *Dairy Microbiology*, 2nd edn. London: Elsevier.

Scott R (1981) *Cheesemaking Practice*. London: Applied Science.

Seth RJ and Robinson RK (1988) Factors contributing to the flavour characteristics of mould-ripened cheese. In: Robinson RK (ed.) *Developments in Food Microbiology*. London: Elsevier.

Surface Mold-ripened Cheese Varieties

M J Sousa, University College, Cork, Ireland

Surface Mold-ripened Varieties

Surface mold-ripened cheeses represent a relatively small proportion of world cheese production, but they are becoming increasingly popular with consumers. In surface mold-ripened cheese varieties (e.g., Brie and Camembert), mold grows as a mat on the surface, whereas blue-veined varieties (*See* **Cheeses**: Mold-Ripened Cheeses: Stilton and Related Varieties) are characterized by the growth of *Penicillium roqueforti* in fissures throughout the cheese. Hybrid

mold-ripened cheeses are also produced (e.g., Cambazola); these have a growth of white mold on the surface and blue mold in the interior.

Surface mold-ripened cheeses are soft varieties characterized by the growth of white mould (*Penicillium camemberti*) on the surface of the cheese (Figure 1). The first microorganisms that become established on the surface are yeasts, including *Debaromyces* spp., and *Kluyveromyces* spp. *Geotrichum candidum* also becomes established at this time, although its growth may be limited if the level of salt is high. Growth of *P. camemberti* is observed after 6–7 days of ripening, and this mold forms the characteristic mat on the surface of the cheese. After the surface of the cheese has been neutralized by yeast and moulds (~15–20 days), aerobic bacteria (particularly micrococci and coryneforms), which are inhibited by the low initial pH, begin to grow.

Surface mold-ripened varieties originated in France a few hundred years ago; Camembert and Brie, which originated in the eighteenth century, are prime examples of this type of cheese. Camembert is a small flat cylindrical (~10 cm diameter, 5 cm thick and ~250 g) surface mold-ripened cheese that originated in Normandy, France; it has a moisture content of 46–47% and fat-in-dry-matter (FDM) of at least 40%. Traditionally, Camembert has been manufactured from raw milk (e.g., Camembert de Normandie), but nowadays, industrial Camembert is manufactured from pasteurized cows' milk in large, automated factories. Camembert has a close, smooth texture that usually becomes soft and spreadable, depending on the degree of maturation. Young Camembert has a mild, slightly aromatic taste, but as ripening progresses, it develops a distinctive, piquant flavor. Brie is a flat cylindrical surface mold-ripened cheese with a larger diameter than Camembert, which it resembles closely in flavor, texture and manufacturing protocol. Carré de l'Est is a square (10 cm side and 2.5 cm thick) surface mold-ripened cheese that originated in eastern France; it has a moisture content of 50–53% and an FDM of 45–50%. It resembles Camembert and Brie, although its flavor is normally milder. Neufchâtel cheese is manufactured near Rouen, France, at the farm level, in a wide variety of shapes and dimensions; the ripened cheese has a moisture content of 51–53% and an FDM of 45–51%. Tomme de St. Marcellin is a flat, cylindrical (8–10 cm diameter and 2–3 cm thick) cheese that was manufactured traditionally from goats' milk but is now made from cows' or sheep's milk or a mixture of these milks. There are many minor goats' milk cheese varieties that develop a surface mold growth during ripening, but the microflora of these minor varieties is often uncontrolled. The flavor of such cheeses is usually strong (mainly due to the presence of short-chain fatty acids), giving the products a characteristic flavor, and their rinds are often colored as a result of the growth of bacteria.

Figure 1 Examples of surface mold-ripened cheeses (Maison du Lait, Les patés molles à croute fleurie, Centre Interprofissional de Documentation et d'Information Laitiéres, www.cidil.fr).

Production of Surface-mold Cheeses

There are several basic steps in the manufacture of cheese. *See* **Cheeses: Chemistry and Microbiology of Maturation**. Acidification, dehydration, shaping, and salting, in the usual order of occurrence, are common to most varieties; the reason for the individuality of each variety is due mainly to differences in detail (Figure 2).

These five operations, along with the chemical composition of the milk, determine the degree of dehydration. The flavor, aroma, and texture of the finished product are determined by the moisture content, salt level, pH, and microflora, and by regulating and controlling the biochemical changes that occur during the ripening process.

The manufacture of surface-mold cheese begins by selecting milk of high quality, both chemically and microbiologically. Raw milk is still sometimes used for both commercial and farmhouse cheesemaking; however, cheese milk is now usually pasteurized or heat-treated immediately before use. In addition to killing pathogens, pasteurization eliminates certain other indigenous microorganisms, and hence cheese of a more uniform quality can be produced. The surface microflora is often more complex than that of the core, particularly in cheeses made from raw milk by traditional technology.

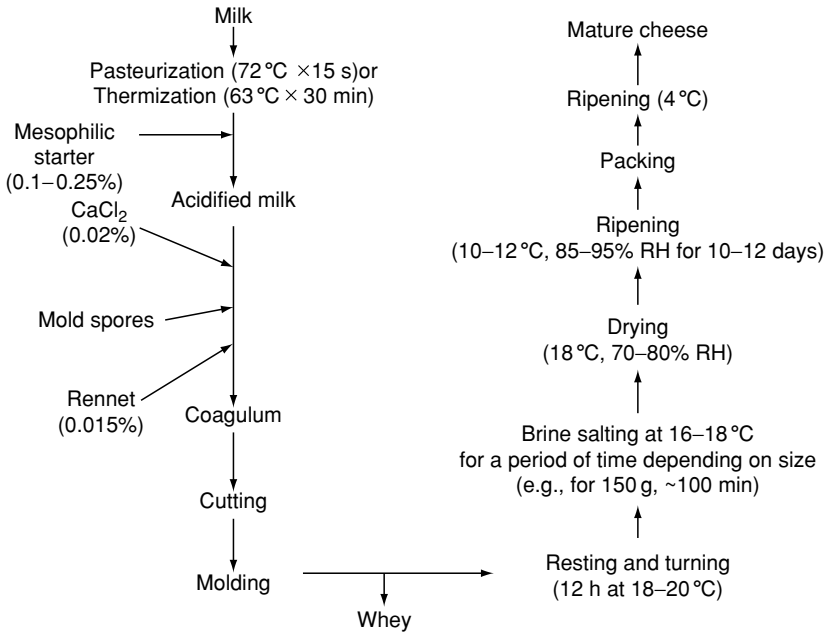

Figure 2 Manufacturing protocol for Camembert or Brie cheese.

Once the milk has been selected and pretreated (if desired), the acidification process is started and continues for up to 24 h. It is now almost universal practice to use a starter culture of known lactic acid-producing bacteria, the predominant function of which is to produce acid at the appropriate rate during cheesemaking. The lactic starters used for Camembert and other surface-ripened cheeses are generally homofermentative mesophiles (e.g., *Lactococcus lactis* ssp. *lactis* and *Lc. lactis* ssp. *cremoris*) that produce lactic acid by the hexose diphosphate pathway. The rate of acidification for Camembert cheese is slower than for hard varieties (e.g., Cheddar; *See* **Cheeses**: Manufacture of Hard and Semi Hard varieties of Cheese) owing to the quantity of starter added to cheese milk, usually ~0.1% and 1–2% starter for Camembert and Cheddar, respectively.

At the same time, the milk can be inoculated with microorganisms that will later make up the surface flora (i.e., a suspension of spores of *P. camemberti*). Usually, $CaCl_2$ (0.02%) is added to the milk to promote coagulation.

Coagulation of the casein component of the milk protein system involves forming a gel, which entraps the fat, water, salts, and lactose. Chymosin is the principal proteinase in traditional rennet used for cheesemaking, and its main role is to specifically hydrolyze the Phe_{105}–Met_{106} bond of the micelle-stabilizing protein, κ-casein, leading to coagulation in the presence of Ca^{2+} at a temperature > 20 °C.

A rennet-induced gel or coagulum is quite stable until it is cut or broken, which causes syneresis. Brie and Camembert are high-moisture cheeses (45–55% moisture), and to prevent excessive syneresis, the coagulum is cut or broken into very large pieces, and the curds are not cooked and only very slightly stirred in the whey, thus discouraging dehydration. Traditionally, an alternative means of whey and curd separation, called 'dipping,' is practiced, in which the uncut coagulum is scooped from the vat into perforated molds, where drainage occurs. To facilitate industrial-scale manufacture, the coagulum for Brie or Camembert is first cut into large cubes and then transferred to molds without cooking.

After molding, the body of curds in the molds is turned at intervals over a period of 3 h and allowed to cool overnight; Brie and Camembert are pressed under their own weight only. Shortly after manufacture, the pH of the cheese is low (4.6–4.9).

Surface mold-ripened cheeses are generally salted by immersion in brine for long enough to give an overall salt content of 1.5–2.5%. Generally, salting stops the production of acid in the curd, preventing a further decrease in pH. Along with having a direct effect on cheese flavor, the salt encourages curd syneresis and therefore has an effect on the moisture content; it reduces the water activity in the curd and influences the activity of rennet, indigenous milk enzymes, and other enzymes. The Na^+ of the salt exchanges with the calcium in the curd matrix, causing dissociation of proteins, and thus enhances the emulsification of fat by the protein matrix.

In surface mold-ripened varieties, mold spores are introduced as a secondary starter during the

cheesemaking process. Generally, *P. camemberti* spores are sprayed on to the surface of each newly shaped cheese, although it is now industrial practice to inoculate the cheese milk with spores before coagulation. Under suitable conditions [10–12 °C and 85–95% relative humidity (RH) for 10–12 days], these spores result in a white mold developing on the outside of each cheese. This mold gives the cheese its characteristic appearance and produces enzymes that play a major role in the hydrolysis of the lipids and proteins during ripening. After mold development, cheeses are packaged in waxed paper and placed in wooden or cardboard boxes prior to final ripening at 4 °C for a further 7–10 days. In traditional Camembert, in which ripening continues for up to 12 weeks, the mold is subsequently overgrown by *Brevibacterium linens* and other coryneform bacteria to form the yellow crust, a distinct sulfur flavor, and the very soft texture of extra-mature Camembert.

Biochemical Pathways during Ripening

Three main biochemical events occur during cheese ripening: metabolism of lactose and lactic acid (generally called glycolysis), lipolysis, and proteolysis (Figure 3). (*See* **Cheeses:** Chemistry and Microbiology of Maturation.)

Cheese ripening can be a relatively slow process, but in the case of Camembert and other soft cheeses, ripening occurs very rapidly, owing to their moisture content and the rapid growth of surface mold.

Metabolism of Lactose and Lactate

During the manufacture of cheese curd, the lactose in the milk is converted mainly to L-lactate by starter bacteria. In surface mold-ripened cheeses, the lactic acid produced by the starter culture is utilized by the secondary microflora, first by *G. candidum* and yeast, principally *D. hansenii*, followed by *P. camemberti*, all of which catabolize lactate to CO_2 and H_2O and hence have a major role in the deacidification of the cheese. There is clear evidence of this event in the dramatic increase of the surface pH during the sixth and seventh days, as the mold grows. Deacidification occurs initially at the surface, resulting in a pH gradient from the surface to the core, which causes lactate to diffuse outwards. Neutralization of the surface causes precipitation of Ca^{2+} at the surface as $Ca_3(PO_4)_2$, causing minerals in the cheese curd to migrate to the surface, developing a gradient within the cheese, which has a major effect on cheese texture. This loss of calcium, as a consequence of neutralization, leads to considerable softening of the body of Camembert, as calcium is necessary for the integrity

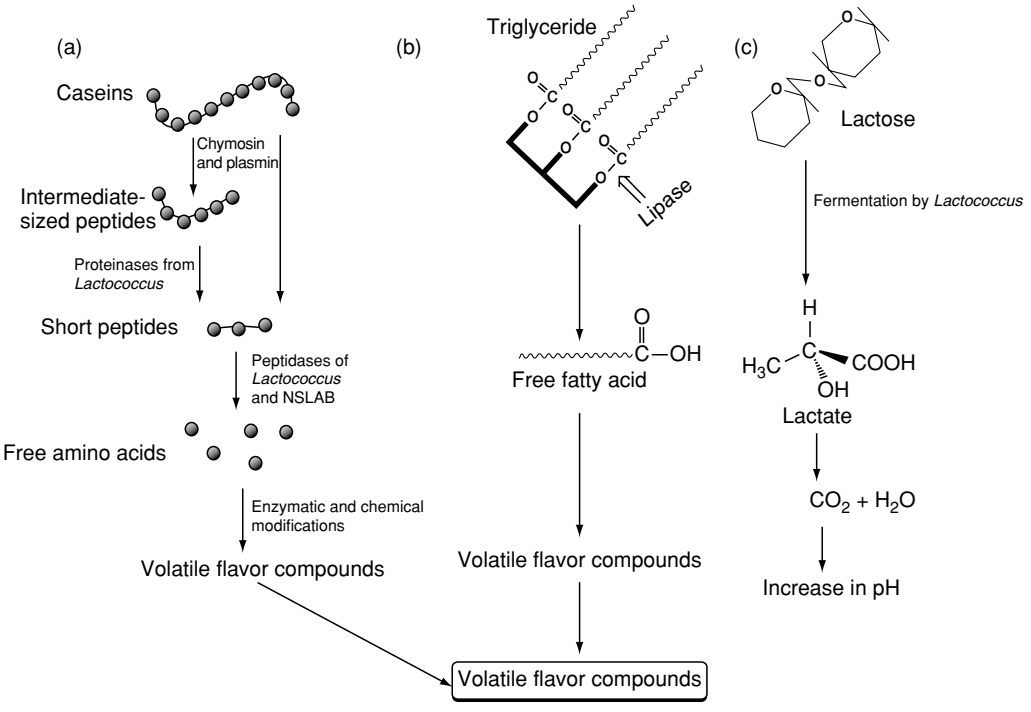

Figure 3 Summary of the biochemical pathways involved in the ripening of surface mold cheeses; (a) proteolysis, (b) lipolysis and (c) glycolysis and their role in flavor biogenesis. Adapted from McSweeney PLH and Sousa MJ (2000) Biochemical pathways for the production of flavour compounds in cheeses during ripening. A review. *Lait* 80: 293–324.

of the protein network. When the lactic acid has been exhausted, *P. camemberti* catabolizes amino acids released from the caseins with the production of NH_3, which diffuses inward, further increasing the pH. Deacidification of the cheese surface has a major impact on the growth of coryneform bacteria (*Br. linens* does not grow below pH 5.8), plasmin activity (which is optimally active at alkaline pH), and coagulant activity (the activity of which decreases with increasing pH).

Lipolysis and Catabolism of Fatty Acids

In most cheese varieties, relatively little lipolysis occurs during ripening. Lipases and esterases in cheese can originate from milk, rennet preparation, starter bacteria, secondary nonstarter bacteria and secondary cultures (e.g. *P. camemberti*). The limited lipolysis is a consequence of the fact that there is normally no significant exogenous or endogenous lipases in pasteurized cheese milk, and the starters used are only weakly lipolytic. The extent of lipolysis in surface mold ripened cheeses (e.g., Brie and Camembert) is generally between 6 and 10% of the total fatty acids. However, significantly lower levels of free fatty acids are present in surface-mold cheeses than in blue-veined cheeses, in which lipases have more contact with fat because of the development of molds in the interior of the cheese and their longer ripening time. Lipolysis tends to be highest towards the surface, suggesting that *Penicillium* is the principal lipolytic agent in Camembert. *P. camemberti* produces a high level of extracellular lipase, which is optimally active at about pH 9.0 and 35 °C, resulting in an increase in lipase activity as the mold develops during ripening. Most free fatty acids between C_4 and C_{20} are produced via lipolysis of triglycerides by fungal lipases, but a smaller proportion of acids between C_2 and C_6 originate from the degradation of lactose and amino acids.

In addition to their direct impact on cheese flavor, free fatty acids act as precursors for series of catabolic reactions, which lead to the production of other flavour compounds (Figure 4). Molds play an important role in the degradation of fatty acids to methyl ketones and their reduction products (secondary alcohols) by a pathway related to β-oxidation. *P. camemberti* and *G. candidum* possess an enzymatic system that oxidizes free fatty acids to β-ketoacyl-coenzyme A, and the action of a thiolase yields a β-ketoacid that is rapidly decarboxylated by a β-keto-acyl-decarboxylase to give a methyl ketone with one less carbon than the initial fatty acid.

A complex array of compounds are produced as a result of lipolytic activity, which contributes to flavor. However, lipolysis is not accompanied by a rancid taste, probably because the fatty acids are neutralized on elevation of the pH.

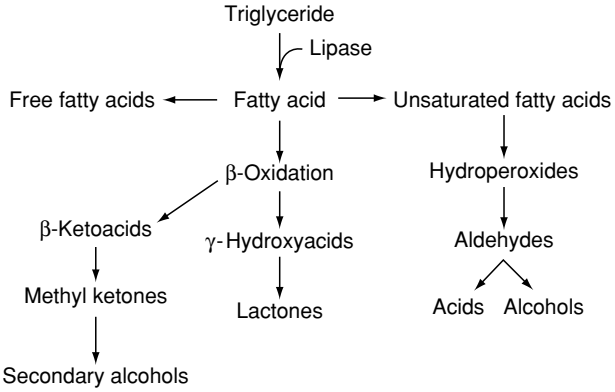

Figure 4 Lipolysis and catabolism of fatty acids during ripening of surface mold-ripened cheeses.

Proteolysis

Proteolysis makes a direct contribution to the flavor, and perhaps to off-flavor (e.g., bitterness), of cheese through the formation of peptides and amino acids, liberation of substrates (amino acids) for catabolic changes (e.g., deamination, decarboxylation, transamination, desulfuration, catabolism of aromatic compounds such as phenylalanine, tyrosine, and tryptophan, and reactions of amino acids with other compounds), and textural changes in the cheese curd, owing to breakdown of the protein network, a decrease in a_w through water binding by formed carboxyl and amino groups and an increase in pH.

Surface mold-ripened cheeses undergo considerable proteolysis during ripening due to the activity of proteinases from starter microorganisms, rennet, milk and, especially, the proteinases and exo- and endopeptidases secreted by the *Penicillium* species. The progression of proteolysis can be summarized as follows: initial breakdown of caseins, which is catalyzed mainly by the residual coagulant, plasmin, and perhaps cathepsin D and other somatic cell proteinases, followed by the production of large water-insoluble polypeptides and intermediate-sized water-soluble peptides, which are degraded by proteinases and peptidases and peptidases from the starter, nonstarter bacteria, and secondary microflora (*P. camemberti*) of the cheese to small water-soluble peptides and amino acids.

As in other cheeses, coagulant is considered to be essential for the development of the correct texture in surface mold-ripened cheeses. Since plasmin is not inactivated by pasteurization, proteolytic activity owing to plasmin is high in the surface of these cheeses owing to their high pH (~7) and higher

than in cheese varieties without molds. the increase in pH may also increase the proteolytic activity of secondary microorganisms.

While rennet, plasmin, enzymes from starter or nonstarter microflora have an effect on proteolysis, enzymes of *Penicillium* spp. are the principal proteolytic agents in surface mold-ripened cheeses. These molds have very potent proteolytic systems, which include an aspartyl proteinase, a metalloproteinase, an acid carboxypeptidase, and an alkaline aminopeptidase. The aspartyl proteinase is the principal enzyme produced by cultures at grown pH 4.0; its optimum pH is ~5.0 on casein as substrate. It is generally stable between pH 3.5 and 5.5. It is active on α_{s1}-, β- and κ-caseins with different pH optima, but it tends to be most active on α_{s1}-casein. It releases a number of short peptides (soluble in 12% trichloroacetic acid, TCA) from α_{s1}-casein, whereas it releases predominantly 12% TCA-insoluble peptides from β-casein. The metalloproteinase has a pH optimum of 5.5–6.0 and is stable over the range pH 4.5–8.5, which suits the conditions in Camembert and other surface mold-ripened cheeses. The metalloproteinase hydrolyses α_{s1}-casein better than β- or κ-casein. Quantitative studies show that the aspartyl proteinase is more active than the metalloproteinase during ripening. *Penicillium* spp. also have carboxy- and aminopeptidase activities. There are two carboxypeptidases, one with an acid and the other with an alkaline pH optimum. The carboxypeptidase activity shows differences in behavior, depending on whether its origin is intra- or extracellular; the extracellular carboxypeptidase is a serine enzyme optimally active on X–Glu–Tyr substrate at pH 3.5, whereas the intracellular carboxypeptidase is optimally active on X–Gly–Val at pH 6. Owing to their broad specificity, the acid carboxypeptidases of *Penicillium* spp. may contribute to the breakdown of bitter peptides in surface mold-ripened cheese. The aminopeptidase has an optimum pH around 8.0–8.5; it is a metalloenzyme with the ability to release apolar amino acids, but peptides with glycine in the penultimate position or N-terminal are hydrolyzed slowly.

Proteolytic activity at the center of Camembert and other surface-ripened cheeses is low, but it is high at the surface, indicating that the migration of *Penicillium* proteinases in the curd is limited. However, peptides produced by the aspartyl proteinase have been detected inside the cheese at a depth greater than 7 mm at the end of ripening. This is probably due to the migration of peptides produced by the aspartyl proteinase, rather than of the enzyme.

The final products of proteolysis are free amino acids, and their concentration at any stage of cheese ripening is the net result of the liberation of amino acids from casein and their catabolism to a number of compounds, including ammonia, amines, aldehydes, phenols, indole, and alcohols (Figure 5), all of which contribute to the flavor of all cheese varieties, especially surface mold-ripened cheeses.

Flavor of Surface Mold-Ripened Cheeses

Numerous compounds contribute to the flavor of surface mold-ripened cheeses including fatty acids, methyl ketones, ketones, alcohols, sulfur compounds, amines, lactones, esters, aldehydes and pyrazines, derived mainly from the three metabolic pathways described above.

Fatty acids are important to the aroma of surface mold-ripened cheeses, not only directly, but also as precursors of methyl ketones, alcohols, lactones, and esters. Short- and medium-chain, even-numbered fatty acids (C_4 to C_{12}) have low perception thresholds, and each has a characteristic flavor. Acetic acid has a typical vinegar odor, butyric has a rancid, cheesy aroma, isobutyric and isovaleric acids have a mild aroma, suggestive of sweet or rotten fruit, and

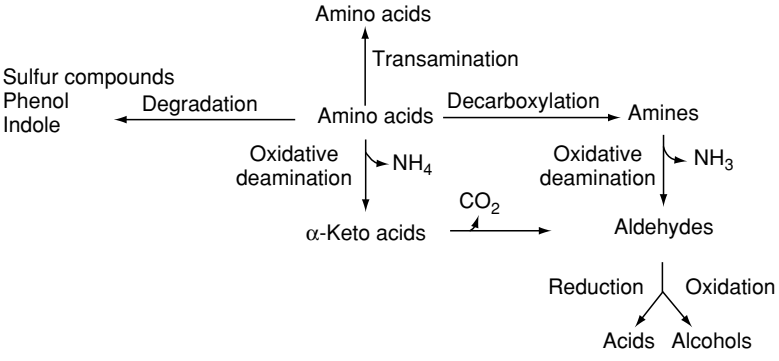

Figure 5 Pathways for the catabolism of free amino acids in surface mold-ripened cheeses.

octanoic, 4-methyloctanoic and 4-ethyloctanoic acids have a goaty aroma. Long-chain free fatty acids ($> C_{12}$) play a minor direct role in cheese flavor because of their high perception thresholds.

Ketones have different perception thresholds and different odor characteristics. The most important ketone in Camembert cheese is octan-1-en-3-one, which has mushroom note; octan-2-one, nona-2-one, and decan-2-one are described as having a musty note, and diacetyl and acetoin are well known for their buttery aromas.

Methyl ketones (C_3 to C_{15}, odd chain) are the most abundant aromatic compounds in surface mold-ripened cheeses and are present in Camembert and Brie cheeses from the eighth day of ripening. Several methyl ketones (e.g., 3-methylpentan-2-one, 4-methylpentan-2-one, non-1-en-2-one, and undec-1-en-2-one) have been identified in Camembert, but the two major methyl ketones are nonan-2-one and heptan-2-one, produced by *P. camemberti* from long-chain fatty acids by β-oxidation, and their concentrations increase during ripening *G. candidum* produces methyl ketones, such as pentan-2-one, hepta-2-one, nonan-2-one, undeca-2-one, and pentan-3-one. Octan-2-one, nonan-2-one, decan-2-one, undecan-2-one, and tridecan-2-one are described as having fruity, floral, and musty notes, whereas heptan-2-one has a blue cheese note. Methyl ketones with even-numbered carbon chains (except butan-2-one) appear later during ripening and are never present in large amounts, except in extra-ripe cheese.

Primary and secondary alcohols, along with ketones, are considered to be the most important compounds in the aroma of surface mold-ripened cheeses. 3-Methylbutan-1-ol is present at a relatively high concentration in Camembert cheese and gives it an alcoholic floral aroma. The principal secondary alcohols present are heptan-2-ol and nonan-2-ol, which are derived from their respective methyl ketones. Oct-1-en-3-ol, which is well known for its raw mushroom aroma, is produced by *P. camemberti* during ripening and represents 5–10% of the aroma compounds in ripened Camembert. This alcohol has a low perception threshold and is one of the key compounds in the aroma of Camembert. Phenyl-2-ethanol is one of the major compounds in Camembert after 7 days of ripening; it exhibits a characteristic rose floral note, and its ester, phenylethylacetate, plays an important role in the flavor of raw milk Camembert.

Sulfur compounds (2,4-dithiopentane, diethyldisulfide, 2,4,5-trithiohexane and 3-methylthio-2,4-dithiapentane) also play important roles in the flavor of surface mold-ripened cheeses; they cause the strong garlic notes characteristic of the mature cheese. Coryneform bacteria are thought to be the key contributors to the formation of sulfur compounds in surface-mold-ripened cheeses. *Br. linens* can produce methanethiol, because these bacteria have a methionine-γ-lyase, which produces methanethiol from methionine, although methanethiol can also be produced from methionine by cystathionine-β-lyase, methionine-α-deaminase or methionine-α-transaminase. *P. camemberti* and *G. candidum* produce methanethiol and other sulfur compounds.

Numerous volatile amines have been identified in Camembert cheese (e.g., methylamine, ethylamine, dimethylamine, isopropylamine, or *n*-butylamine). Ammonia is also an important contributor to the aroma of surface mold-ripened cheese, and *P. camemberti, G. candidum*, and *Br. linens* play major roles in its production. Nonvolatile amines (i.e., tyramine, histamine, tryptamine, and putrescine) have also been identified in this cheese type.

The lactones, γ-decalactone, δ-decalactone, γ-dodecalactone and δ-dodecalactone, are present in Camembert cheese; they are characterized by very pronounced fruity notes (peach, apricot, and coconut).

Many esters are present in surface mold-ripened cheeses; they are characterized by fruity, floral notes (e.g., pineapple, banana, apricot, pear, floral rose, honey, and wine). 2-Phenylethyl acetate and 2-phenylethyl propanoate are quantitatively important in Camembert.

The main aldehydes in Brie and Camembert cheeses are hexanal, heptanal, nonanal, 2-methylbutanal, 3-methylbutanal, and benzaldehyde, and they appear in trace amounts as early as the first week of ripening. Hexanal gives the green note of immature fruit, nonanal is described as having an aromatic note resembling orange, while benzaldehyde is described as having an aromatic note of bitter almond.

Compounds like styrene, which have a very strong plastic odor, have been found in trace amounts in Camembert, when the degradation of linoleic and linolenic acids by *P. camemberti* is higher than usual. Pyrazines (e.g., 2,5-dimethylpyrazine and 2-methoxy-3-isopropylpyrazine) are also present in Camembert. The former imparts a toasted hazelnut note, whereas the latter is responsible for the rotten soil, raw potato aroma note, which is regarded as a defect.

Bitterness in surface mold-ripened cheese has been correlated with the concentration of pH 4.6-soluble nitrogen and thus with the formation of peptides. Bitterness could be eliminated by ripening cheese in an atmosphere containing NH_3, which results in slower growth of *P. camemberti* on the cheese surface

and therefore less proteolysis and a lower level of bitter peptides, or by inoculating with secondary microorganisms (*P. camemberti, G. candidum*, and *Br. linens*) with a higher aminopeptidase activity.

The characteristic appearance of surface mold-ripened cheese is due to the presence of mold (e.g., *P. camemberti*), which produces enzymes with high lipolytic and proteolytic activities, which play a major role during ripening. These molds also lead to more complex ripening patterns than in other varieties of cheese with simpler flora. Much progress has been made during the last 15 years in elucidating the mechanisms of ripening in surface mold-ripened cheeses. However, further work is required to understand the enzyme systems and the mechanisms involved in the conversion of the primary products to volatile flavor compounds.

Acknowledgments

The author would like to thank Prof. P.F. Fox and Dr. P.L.H. McSweeney for useful comments on the manuscript.

See also: **Cheeses**: Chemistry and Microbiology of Maturation; Manufacture of Extra-hard Cheeses; Manufacture of Hard and Semi-hard Varieties of Cheese; Mold-ripened Cheeses: Stilton and Related Varieties

Further Reading

Fox PF, Guinee TP, Cogan TM and McSweeney PLH (2000) *Fundamentals of Cheese Science*. Gaithersburg, MD: Aspen.

Gripon JC (1993) Mould-ripened cheeses. In: Fox PF (ed.) *Cheese: Chemistry, Physics and Microbiology*, 2nd edn, vol. 2, pp. 111–136. London: Chapman & Hall.

McSweeney PLH and Sousa MJ (2000) Biochemical pathways for the production of flavour compounds in cheeses during ripening. A review. *Lait* 80: 293–324.

Molimard P and Spinnler HE (1996) Review: compounds involved in the flavour of surface mould-ripened cheeses: origins and properties. *Journal of Dairy Science* 79: 169–184.

Schlesser JE, Schmidt SJ and Speckman R (1992) Characterization of chemical and physical changes in Camembert cheese during ripening. *Journal of Dairy Science* 75: 1753–1760.

Sousa MJ and McSweeney PLH (2001) Studies on the ripening of Cooleeney, an Irish farmhouse Camembert-type cheese. *Irish Journal of Agricultural and Food Research* 40: 83–95.

Robinson RK (1995) Cheese ripened with moulds. In: Robinson RK (ed.) *A Colour Guide to Cheese and Fermented Milks*, pp. 11–127. London: Chapman & Hall.

Dutch-type Cheeses

G van den Berg, NIZO Food Research, Ede, The Netherlands

Introduction

Dutch-type cheeses represent the most important kinds of semihard cheese. They are manufactured from bovine milk, although small amounts of similar cheeses are made from ovine or caprine milk. Other semihard cheese types are also manufactured according to a largely similar technology but their characteristics are different. For this reason, Jarlsberg or Maasdam and Tilsiter or Havarti cheese will not be treated here.

Gouda and Edam cheese are the primary Dutch-type cheese varieties and they are made all over the world. However, in many countries a similar technology is used for related cheese types manufactured under various names. The majority of all these cheese varieties contain 40–52% fat in dry matter, and the water content in the (unripened) fat-free cheese ranges from 55 to 63%. They are made in sizes between 0.2 and 20 kg. Their shape can be a flat cylinder, often with somewhat bulging sides, a sphere, a rectangular block, or a loaf. The latter are often used for foil ripening (rindless cheese). The standard cheeses are naturally ripened under drying conditions.

Dutch-type cheeses are ripened for some weeks or longer, sometimes for more than a year, and are often covered with a plastic coating. The consistency varies from rather firm and smooth to semisoft and changes during natural ripening to a firmer and more brittle structure. The flavor also changes from mild to strong during such long ripening times. The interior of the main cheese types shows some round eyes about the size of a pea. Substances like cumin seeds or other spices may be added. There is essentially no surface flora.

Production Statistics

The worldwide production volume of Dutch-type cheeses can only be estimated because from many countries there are no official figures for individual cheese types. Moreover, all related cheese types are not always well described in different countries. So a worldwide picture can only be estimated.

World cheese production in 1998, excluding processed cheese but including fresh cheese and quark, is estimated as $12\,800 \times 10^3$ t. The share of the Dutch-type varieties is estimated to be approximately 15%

of this volume, which is based on a mix of real figures for various European countries and estimates for others. The production of these cheese types in The Netherlands in 1999 was 535×10^3 t and in Germany 415×10^3 t as the two countries with the highest production, while the whole European Union produced approximately 1300×10^3 t. Dutch-type cheese is usually made from pasteurized milk; only 10×10^3 t, mainly Gouda cheese, is farm-made from raw milk in the Netherlands and will not be considered further.

Technology

A typical manufacturing scheme for Gouda cheese is given in Figure 1.

The milk is thermized and standardized for its fat : protein ratio in order to control the fat content in the dry matter of the cheese. It is cold-stored until cheese-making can take place. Then, the milk is pasteurized and pumped into the cheese vat. During one of these heat treatments, the milk is often bactofugated to reduce the number of spores of butyric acid bacteria. The total heat load must be sufficient to inhibit pathogenic microorganisms and some spoilage organisms but not xanthine oxidase (EC no. 1.2.3.2). The denaturation of whey proteins should be limited.

For the manufacture of Dutch-type cheeses, the milk is coagulated at approximately 30 °C by the action of renneting enzymes, usually a chymosin (EC 3.4.23.4) preparation (calf rennet). This is added together with the mesophilic starter and calcium chloride to control the renneting process. Other additives, such as nitrate and coloring agents, like anatto (E160b) or carotene (EE160a), are often used. The resulting coagulum consists of a continuous network of strands of (*para*)casein micelles that extends throughout the milk volume. (*See* **Cheeses: Starter Cultures Employed in Cheese-making.**) In general, this gel is not as firm when cut as is the case for soft cheese-making. It is divided into fairly small particles but not so fine as is the case for hard cooked cheese.

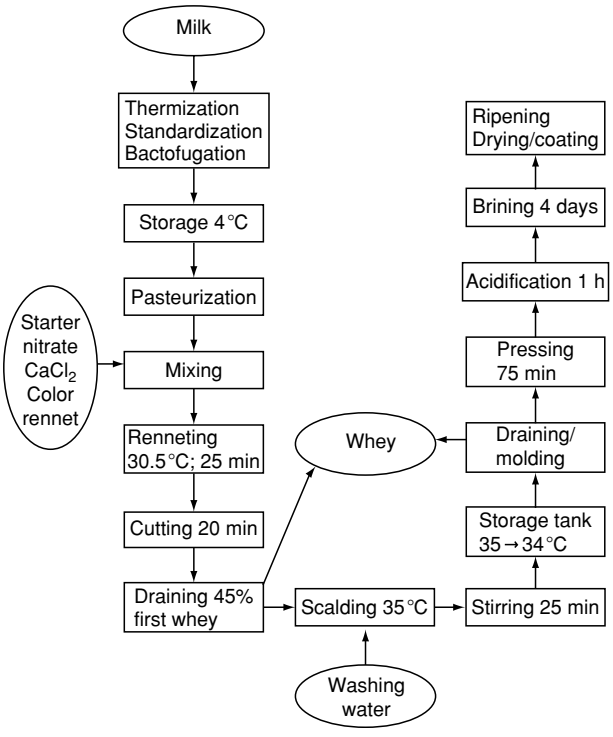

Figure 1 Main process steps of a modern manufacturing scheme for 12-kg Gouda cheese.

After cutting, a part of the (first) whey already liberated from the curd particles by syneresis (shrinkage and expulsion of whey) is removed and warm water is added. This scalding of the curd serves to reduce the lactose content and to enhance syneresis of the curd during further stirring. The amount of this curd wash water is related to the desired pH of the final cheese. This parameter is a result of the ratio between the amount of lactose that is fermented to lactic acid and the amount of the (*para*)caseinate–phosphate complex (the acid buffering substance). So in this respect, the lactose content of the milk and the water content of the nonfat matter after pressing/before brining are important. This water content is important because it is equal to that of the center of the cheese when all lactose has been fermented. The higher these parameters, the more lactose remains in the curd and the more water that has to be added to maintain a constant pH.

The scalding temperature is primarily used to influence the dry-matter content of the cheese because it affects syneresis. This is an intrinsic property of the (*para*)casein gel that leads to the protein strands, breaking whereupon new bonds will be easily formed because the whole surface of the casein micelle can be considered as reactive. The network is rearranged when the moisture with dissolved components (whey) is expelled. Increasing the surface of the coagulum (cutting), acidification, temperature increase, and stress (by stirring the curd or, during a successive stage, by taking the curd out of the whey) all enhance syneresis. The increase of the scalding temperature is limited because of the use of a mesophilic starter.

After making curd, the curd particles are collected under the whey to exclude air. Subsequently, the free whey between the curd particles must be expelled by drainage in order to obtain a closed texture in the final cheese. However, even after pressing, some whey still remains in the small interstices between the curd particles, but during later stages, the temperature drops and this whey is readsorbed by the curd so

that fusion can be completed. In fact, a section of the cheese will show a closed texture only after some days.

The blocks of curd are pressed in perforated molds covered with a lid, allowing whey drainage. A rind of a few millimeters thick is created by deformation of the curd around threads of the liner or edges of perforations in the mold. This rind formation is due to strong local syneresis and further fusion of the curd particles, and more bonds between the casein strands are created than elsewhere in the cheese. In essence, syneresis, fusion, and rind formation are similar processes.

A closed rind is necessary, for the natural ripening cheese in particular, to withstand the tension caused by brining and to protect the cheese from penetration of undesired microorganisms during brining and from molds during ripening. (*See* **Cheeses**: White Brined Varieties.)

During curing the integrity of the rind has to be maintained. Coating the cheese makes it much easier to maintain the rind and gives a smooth and glossy appearance when the optimal climatic conditions, air flow, and cheese treatment scheme in the ripening room are applied. The polyvinylacetate-based coating layer reduces water evaporation from the cheese surface to some extent. However, too strong drying will make this hydrophilic coating layer too brittle and mold will penetrate into cracks. On the other hand, too little drying will easily allow undesired development of bacteria that make the surface sticky, and the presence of natamycin will not be sufficient to prevent the growth of yeasts and molds. The normal ripening temperature is rather moderate, approximately 13 °C. In the case of ripening in foil, or in wax, this temperature is normally lower and the moisture content is often somewhat lower to keep the cheese in better shape and to prevent it from being too soft after curing.

It is obvious that cheese composition has to be controlled carefully. During the manufacture of Gouda cheese, the casein is concentrated at the end of curd preparation by a factor of approximately 4.5 and in the final cheese by approximately 10. In parallel with this, the dry-matter content of the curd mass increases to approximately 32% at the end of curd preparation, to 42% after drainage, to 53% after pressing, and to 57% after brining. The composition of the cheese directly influences the product yield. However, milk composition is the primary factor to reckon with. Of the components to be concentrated, casein is the most expensive one. The fat in the whey can be recovered in the cheese by adding the whey cream to the cheese milk whereas the casein (curd fines) cannot when making quality products. So it is good to express yield as kg per kg casein of the milk. However, in a quantitative sense, water is the most important cheese component but cheese quality demands optimal water content, before brining in particular, which depends on the cheese type involved. This is important for the desired consistency of the young cheese and for the conditioning of the ripening processes.

Nowadays, the cheese-making process is carried out in factories manufacturing 3–5 tonnes of cheese per hour, and mechanization and automation based on a good understanding of the technology are well developed.

Starters and Lactose Fermentation

The role of starter bacteria is discussed elsewhere (*See* **Starter Cultures**). The starters for Dutch-type cheese have always been mesophilic lactic acid bacteria (LAB). The commonly used starters consist of combinations of acid-producing *Lactococcus lactis* subsp. *lactis* and *cremoris* strains and citrate-fermenting (and carbon dioxide-producing) *Leuconostoc lactis* and/or *L. mesenteroides* subsp. *cremoris* (L-starters) or *Lactococcus lactis* var. *lactis* biovar *diacetylactis* and *Leuconostoc* strains (DL-starters). DL-starters are normally used when stronger eye formation is desired because they ferment citrate faster and to a greater extent than do L-starters. When in particular cases no eye formation must occur, then a starter without citrate-fermenting bacteria is often used (O-starter). Mixed-strain starters are mainly used in Europe for Dutch-type cheese. This is also true in the case of an O-starter. Manufacturers, especially the bigger factories, prefer to use their own bulk starter, to realize lower starter costs and a better yield in comparison with the use of a DVS (direct to the vat) system. Such a bulk starter system is provided with the means to avoid bacteriophage contamination, such as a bacteriophage-free mother starter concentrate and a starter propagation tank with an overpressure of bacteriophage-free air. Such starters have to be well selected because of their major influence on cheese quality.

The amount of bulk starter of normal activity used to inoculate cheese milk is 0.5–1.0%. During cheesemaking, starter LAB grow until a level of 10^9 colony-forming units (cfu) g^{-1} has been reached. In Gouda and Edam cheese, the pH at that time is approximately 5.7 and the cheese has already been pressed. A picture of the usual pH course during cheesemaking is given in Figure 2.

During curd preparation, the pH remains rather high until most of the whey is separated by drainage in order to obtain the desired cheese consistency and

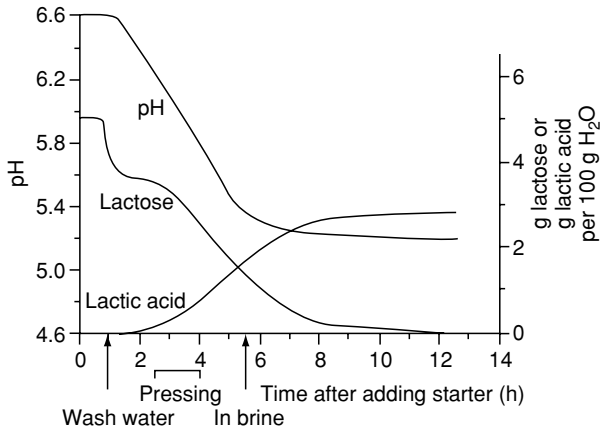

Figure 2 Lactose fermentation, acid production, and the pH of curd and cheese during the manufacture of Gouda cheese after starter addition.

yield. Moreover, less chymosin is bound to the casein and the danger of bitterness after ripening is diminished. Before lactose fermentation has finished, the cheese has already been put into the brine. The moisture diffusing out of the cheese still contains lactose and so does the brine, which is used for a long time. As a result, after brining, the rind of the cheese still contains some lactose, which disappears within the first few weeks of ripening. For foil-ripened block cheese, in particular, the time between pressing and brining is often still shorter after pressing and these effects are stronger.

There is a strong interest in acceleration of ripening and product differentiation, not only from differences in packaging and shape, but also in flavor. To this end, attenuated LAB starters are introduced as an extra means of enhancing flavor development. They are often thermophilic and propagated or treated in such a way that they practically do not grow or contribute to acidification, so that normal manufacturing procedures can be maintained. In the case of low-fat cheese, the ability to enhance protein degradation is of special interest in order to achieve a more acceptable consistency and a stronger flavor of the final product.

Ripening Patterns

After fermentation of lactose and citric acid, proteolysis is the main phenomenon in the manufacture of Dutch-type cheese, followed by amino acid degradation. Within 24 h the pH of the cheese is 5.1–5.2, increasing during the first 2 weeks by approximately 0.15, and thereafter only slightly during maturation. The redox potential in the cheese has been reduced by the lactic acid fermentation to approximately −140 mV. In the center of Gouda cheese the original water content in the fat-free cheese is approximately 65%, but this decreases gradually by salt diffusion and water evaporation. Under these conditions, casein is degraded by chymosin (not inactivated by high scalding temperatures), and to a lesser extent by plasmin, into larger peptides. These are further degraded, finally into amino acids, by peptidases of the starter LAB that also provide amino acid-converting enzymes (AACEs) for the production of various volatile (flavor) components.

Under the gradually changing conditions in normal Gouda cheese, only a very global quantitative picture can be given of how degradation of the (total) casein proceeds during ripening. To this end, the nitrogen-containing fraction of the cheese is separated into a fraction which is soluble under cheese conditions (SN) and a fraction which is still soluble in 12% trichloroacetic acid (AN). The latter fraction of < 1400 molecular weight contains only very small peptides, amino acids, and further converted components.

Figure 3 gives the average increase of these fractions (expressed as a percentage of total nitrogen) in normal Gouda cheese during ripening at 13 °C. Generally, the SN fraction is mainly dependent on chymosin activity and the AN fraction on activities of enzymes associated with the starter bacteria. However, the production of peptides by the former stimulates activity by the latter.

The use of attenuated thermophilic starters in addition to the normal culture causes a greater increase of the AN fraction while hardly affecting the SN fraction. In fact, the peptidase and AACE activity is primarily increased.

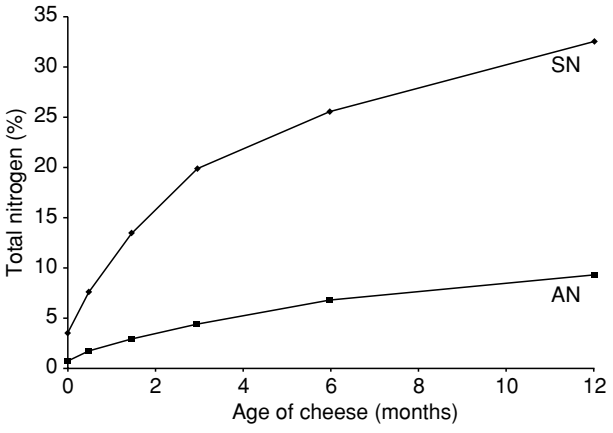

Figure 3 Increase of the soluble nitrogen-containing fraction (SN) and the amino acid-containing fraction (AN) during ripening of normal Gouda cheese as a percentage of total nitrogen fraction in the cheese.

As noted, Figure 3 gives an overall picture of the cheese. The center shows a greater degradation because of the higher moisture content and the (initially) lower salt content. In the rind, proteolysis stops within 3 months because of the low water content. It will be obvious that when water evaporation is prevented by packaging in foil or wax one may speak of rindless cheese, because then the internal ripening processes also continue in the rind zone. However, such cheese is usually ripened at a low temperature.

Lipolysis is very limited in Dutch-type cheese.

Textural Characteristics

Figure 4 gives an impression of the desired eye formation in normal Gouda cheese.

For eye formation, a certain gas pressure in the very young cheese is necessary when the consistency of the cheese is still very elastic. To this end, saturation of the milk with air, in particular with nitrogen – because oxygen disappears due to the activity of the starter bacteria – and citrate fermentation by the starter bacteria during the first weeks after manufacture are crucial. The partial pressure of nitrogen in the young cheese is approximately 0.9 bar and that of carbon dioxide 0.4 bar, which together are sufficient for eye development in Gouda cheese. Without the nitrogen, the cheese would be 'blind.' However, without 'nuclei,' eye formation does not take place either, because then the pressure necessary to start the development of an eye can never be reached. Various deviations in the structure of a curd block may serve as such 'nuclei,' like tiny air bubbles or remaining whey pockets, but also foreign particles with certain apolar surface properties. Faster citrate fermentation by DL-starters can give more or bigger eyes, dependent on the 'nuclei' available, because of the increased gas pressure before the carbon dioxide can diffuse through the rind. A sufficiently elastic texture is necessary because too low a fracture stress will result in cracks instead of eyes.

Such a short consistency is to be expected at low pH. However, the desired consistency also depends on the correct moisture and fat content and the presence of sufficient calcium phosphate. Therefore the pH before pressing must be high. Proteolysis turns the consistency during ripening from rather elastic to smooth within 1–2 months and later to short. In the meantime, the water content has considerably decreased and the consistency will also be rather hard after a year. The rind zone increases in thickness during ripening; it is tougher and harder than the interior and becomes somewhat translucent. The increase in 'shortness' of Gouda cheese during ripening is illustrated in Figure 5 by the shift in the stress–strain curves.

In Gouda cheeses ripened for more than 1 year, tyrosine crystals are often visible as a result of the progress of proteolysis. If attenuated thermophilic starters are added, these crystals will be present after a shorter period.

Flavor Characteristics

Primary products produced during cheese-making, like lactic acid, diacetyl, and carbon dioxide and related fermentation products, are essential for the basic flavor of the young cheese. The same holds for the very small amount of free fatty acids but the desired flavor is based on the right balance of all components. However, ripening starts immediately after manufacture and causes development of a strong flavor in the matured cheese. In this process, all components are probably involved in numerous conversions. So it is understandable that the flavor of the young cheese depends more directly on the type of

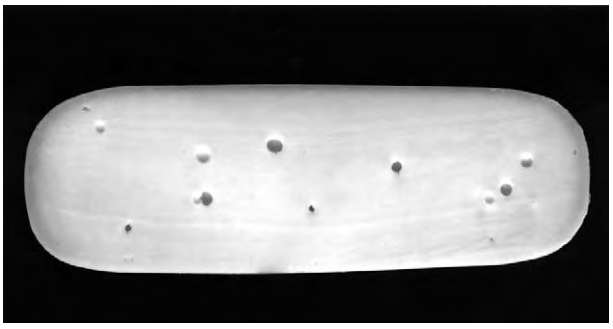

Figure 4 Section through a normal Gouda cheese.

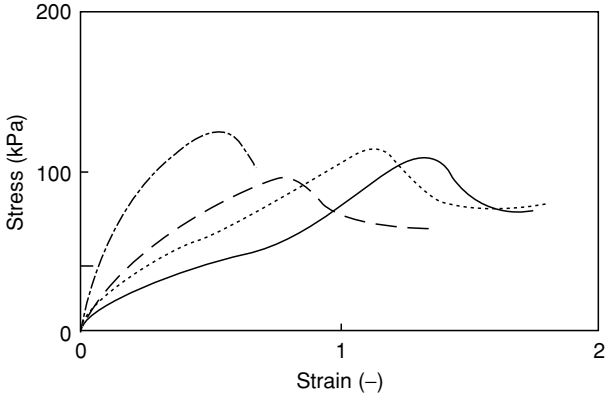

Figure 5 Stress–strain curves obtained for Gouda cheese in a compression test at 2 (———), 6 (·····), 13 (— —), and 26 (—·—·-) weeks. From Zoon P, NIZO Food Research.

starter. Cheeses made using an O-starter have a less creamy flavor than do those with an L-starter and certainly those with a DL-starter. For the more mature cheese, these differences disappear but other starter-related differences are then observed, such as more or less sweetness or fruitiness.

During ripening, the effect of proteolysis becomes more dominant, including amino acid conversion into a great number of volatile flavor compounds. Pathways from protein to these compounds are becoming better known. In good cheese, made with a mixed-strain starter, they are quantitatively well balanced. Some are more characteristic for a particular cheese type than others. This is illustrated in Figure 6, which shows the gas chromatogram area with the most important differences between a Gouda cheese and a similar cheese with an extra attenuated thermophilic starter APS (Proosdij cheese).

All volatile flavor components are already present in small amounts in a 6-week-old cheese but the different flavor is due to the difference in concentration of the key components. Proosdij cheese has larger amounts of peak 8 (butanone), 12 (3-methylbutanal), 14 (2-pentanone), and 22 (2-heptanone), while Gouda cheese shows more of peak 7 (diacetyl), 11 (2-methylpropanol), 13 (1-butanol), 16 (3-methylbutanol), and 21 (ethyl butyrate). The taste of Proosdij cheese has more sweet, caramel, bouillon, and fruity notes than the taste of Gouda cheese.

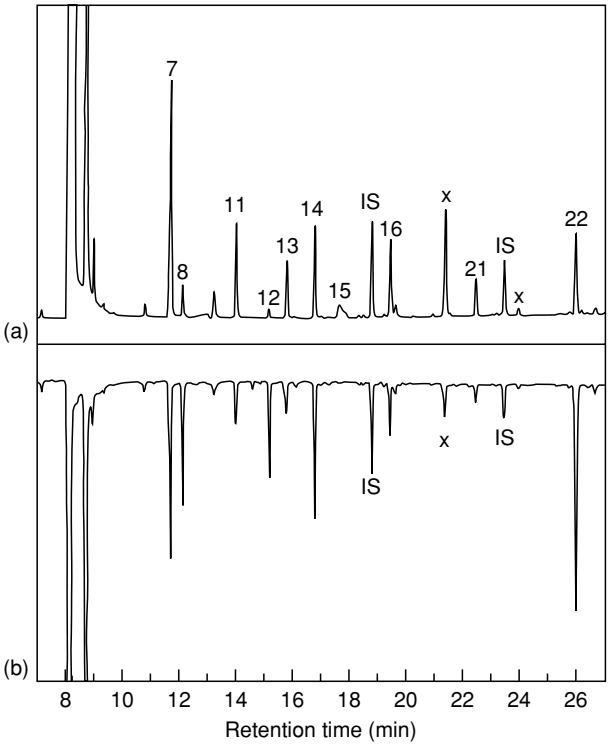

Figure 6 Gas chromatograms (relevant parts) of (a) Gouda cheese (positive) and (b) Proosdij (negative), both after 6 months' ripening. IS, internal standard; x, contaminant. Intensities of peaks have been normalized upon the first IS. From Neeter R, De Jong C, Teisman HGJ and Ellen G (1996) Determination of volatile components in cheese using dynamic headspace techniques. In: Taylor AJ and Mottram DS (eds) *Flavour Science, Recent Developments*. Special publication no. 197, pp. 293–296. Cambridge, UK: Royal Society of Chemistry, with permission.

Defects

Process control is necessary to avoid a number of defects in the cheese. The most important ones are listed, with the main measures required to avoid them.

Butyric Acid Fermentation

This 'late blowing' (large eyes and/or cracks with a sweet and butyric off-flavor) is caused by the growth of the anaerobic bacterium, *Clostridium tyrobutyricum*, originating from spores in the milk, in the 1–5-month-old cheese. The use of nitrate and bactofugation of the milk mainly prevent it.

Mesophilic Lactobacilli

Various strains of *Lactobacillis plantarum, L. casei*, or *L. brevis* are able to grow in cheese and to produce various off-flavors (gassy, putrid, fruity, etc.) and/or carbon dioxide, causing cracks in the mature cheese. Contamination and growth during cheese-making must be avoided as much as possible. Some of these bacteria are salt-tolerant and may contaminate the cheese during brining. So the rind has to be closed during pressing and the bacterial counts of the brine must be controlled.

Thermoresistant Streptococci

In particular, strains of *Streptococcus salivarius* subsp. *thermophilus* may be responsible for excessive carbon dioxide production and some sweet off-flavor in 1–2-month-old cheese. Counts of this organism in the cheese milk have to be controlled, especially by regularly cleaning the plate heat exchangers, because the streptococci grow on the surface of the plates in the regenerator section.

Slimy Rind

Growth of coryneform bacteria on the surface of the cheese should be avoided by good acidification during cheese-making and good curing conditions. Otherwise they may cause gas production under the wax layer which is applied before selling the cheese. A slimy rind without bacterial growth is possible when the calcium content of the brine is too low.

Mold Growth

Mold growth on the surface of the cheese must be avoided by the right curing conditions and turning and coating program. Under normal conditions, the use of natamycin in the coating inhibits mold growth.

Bitterness

This is a ripening defect that can be caused by using too much rennet or a low pH before pressing so that more rennet is entrapped in the cheese, but mainly by using the wrong starter with insufficient peptidase activity to break down bitter peptides.

Texture Defects

Besides the formation of cracks or over-large holes, it is also possible to have too many small eyes or even pinholes. The renneting milk should not contain many very fine air bubbles and air inclusion during drainage of the curd must be avoided. Curd lumps that do not lose their whey properly during drainage mainly cause nesty holes.

See also: **Cheeses**: Starter Cultures Employed in Cheesemaking; White Brined Varieties; Mold-ripened Cheeses: Stilton and Related Varieties; Surface Mold-ripened Cheese Varieties; **Starter Cultures**

Further Reading

Akkerman JC, Walstra P and Van Dijk HJM (1989) Holes in Dutch cheese. 1. Conditions allowing eye formation. *Netherlands Milk Dairy Journal* 43: 453–476.

Exterkate FA, Alting AC and Slangen CJ (1995) Conversion of α_{s1}-casein-(24-199)-fragment and β-casein under cheese conditions by chymosin and starter peptidases. *Systematic and Applied Microbiology* 18: 7–12.

Fox PF (ed.) (1993) *Cheese: Chemistry, Physics and Microbiology*, vol. 1, *General Aspects*, 2nd edn. London: Chapman and Hall.

Neeter R, De Jong C, Teisman HGJ and Ellen G (1996) Determination of volatile components in cheese using dynamic headspace techniques. In: Taylor AJ and Mottram DS (eds) *Flavour Science, Recent Developments*. Special publication no. 197, pp. 293–296. Cambridge, UK: Royal Society of Chemistry.

Smit G, Verheul A, Van Kranenburg R, Ayad E, Siezen R and Engels W (2000) Cheese flavour development by enzymatic conversions of peptides and amino acids. *Food Research International* 33: 153–160.

Van den Berg G (2000) Semi-hard cheeses. In: Tamime AY and Law BA (eds) *Mechanisation and Automation in Dairy Technology, Sheffield Food Technology*, vol. 7. Sheffield: Academic Press.

Van den Berg G and Exterkate FA (1993) Technological parameters involved in cheese ripening. *International Dairy Journal* 3: 485–507.

Van den Berg MG, Van den Berg G and Van Boekel MAJS (1996) Mass transfer processes involved in Gouda cheese manufacture, in relation to casein and yield. *Netherlands Milk Dairy Journal* 50: 501–540.

Visser J (1991) Factors affecting the rheological and fracture properties of hard and semi-hard cheese. *Bulletin of the IDF* 268: 49–61.

Walstra P, Geurts TJ, Noomen A, Jellema A and Van Boekel MAJS (1999) Part IV: cheese. In: Walstra P (ed.) *Dairy Technology*, pp. 539–708. New York: Marcel Dekker.

Walstra P, Noomen A and Geurts TJ (1993) Dutch-type varieties. In: Fox PF (ed.) *Cheese: Chemistry, Physics and Microbiology*, vol. 2, *Major Cheese Groups*, 2nd edn. London: Chapman and Hall.

Chemical Analysis *See* **Analysis of Food**

CHERRIES

W Loescher, Michigan State University, East Lansing, MI, USA

Background

In some temperate regions of the world, i.e., Eastern Europe, Western Asia, and parts of North America, sweet and sour cherries are second only to apples in their economic importance as tree fruits. Sweet cherries are primarily grown for fresh consumption, but sour cherries are almost entirely processed, i.e., dried, canned, or frozen. These differences profoundly affect production, harvest, handling, shipping, and marketing. Although cultivated since ancient times, both sweet and sour cherries have undergone relatively little advanced breeding. Improvement has primarily been through selection, but there is the potential to deal with cultural, pest, disease, and quality problems that now limit production and economics of both types. There are significant

opportunities to expand markets for fresh sweet cherries and for some sour cherry products.

Cherry Taxonomy and Types

Commercially, the two most important cherry species are: sweet cherry, *Prunus avium*, L. and sour cherry, *Prunus cerasus* L., both tree fruits native to southeastern Europe and western Asia. They are closely related, graft-compatible, and hybridize to form interspecific (Duke) cultivars. Sweet cherry (diploid, with a base chromosome number of 8, and a somatic number of 16) probably originated somewhere between the Black Sea and the Caspian Sea, but it spread into Europe in ancient times. Sour cherry (tetraploid, with a base chromosome number of 16, somatic chromosome number of 32) is native to the same areas as sweet cherry, and there is good evidence that crosses between *Prunus avium* and the ground cherry, *P. fruticosa* Pall, gave rise to sour cherry. There are several other cherry species, but most, e.g., Nanking cherry (*Prunus tomentosa*), have limited commercial value as fruits.

Sweet cherries can be divided into two major types based on fruit characteristics. Heart-type cherries are ovoid or heart-shaped with a relatively soft flesh, often ripening early. Most of the commercially important cultivars, however, are of the Bigarreau type with firmer, crisp-fleshed fruit, ripening mid to late season. Fruit flesh may be red or yellow, and the skin may be dark (red to nearly black) or light (yellow–red to yellow–white).

Many sweet cherry cultivars grown throughout the world originated in Europe, but a number of important cultivars were selected or bred in local cherry districts. European cultivars grown in the USA are Napoleon (Royal Ann), Black Tartarian, Eagle, Early Purple, Early Rivers, Elkhorn, Hedelfingen, Knight's Early Black, Lyon, and Schmidt. The cultivars Windsor, Van, Sam, Vista, Victor, Sue, Vega, Summit, and Stella were developed in Canada. Chinook and Rainier were developed in Washington. Bing, Lambert, Black Republican, Corum, and Hoskins were selected and developed in Oregon. Chapman, Burbank, Bush Tartarian, and the new cultivars Mona, Larian, Jubilee, Berryessa, and Bada originated in California. Recent introductions include Ulster and Hudson from New York, and Angela from Utah.

The most important sweet cherry cultivars in the western USA, where about 80% of the US crop is produced, have been dark-fruited, crisp-fleshed cultivars: Bing (the leading cultivar in North America), Van, and Lambert. But others may be available because of their use as pollinators or as the result of recent fresh market demand for large, light-colored, and crisp fleshed fruits from cultivars like Rainier. Firmness, size, color, and soluble solids are all important market considerations, and growers in regions where summer rains are prevalent, e.g., the eastern USA and eastern Europe, are at a disadvantage because the main cultivars are often the softer fleshed, rain-cracking resistant types, e.g., Emperor Francis, Hedelfingen, and Schmidt. In these regions, light-fleshed cultivars, Rainier, Napoleon (Royal Ann), Corum, and Emperor Francis, are best for making into maraschino cherries (because the pigment is undesirable), but a few are nevertheless grown for the fresh market. Napoleon is also used for canning. Bing is mainly a fresh-market cultivar, and Lambert is used both for canning and fresh market. Black Republican and other very firm, dark cherries are good for freezing.

Sour cherry fruit are generally soft, juicy, and depressed-globose in shape, but colors may range from the Morello types, with red to dark red flesh and juice, to the Amarelle types with an almost colorless juice and flesh.

Although new cultivars are being tested, there are only a few sour cherry cultivars commonly grown in North America, ranging from the light red Early Richmond, the medium red-skinned Montmorency, to the late dark red English Morello, but Montmorency is still the standard. In western Europe, Schattenmorelle and Sternsbaer are common, but many others are grown in Russia, the former Yugoslavia, Romania, and Hungary. Most, unlike sweet cherry, are more or less self fertile and generally do not require pollinators. Almost all of those grown in the USA and western Europe are harvested mechanically and sold for processing, primarily as a frozen or canned ingredient for use in manufactured food products such as pies but more recently as a dried fruit product, and in Europe and other areas, several other uses have been developing for juice, liqueur, and marmalade production, and combinations with yogurt.

Production Areas

The USA, Russia, and Germany are large producers of both sweet and sour cherry (Table 1). In some areas of northern Europe, sour cherries are, after apples, the second most important fruit grown. Otherwise, within Europe and western Asia, sour cherry production is concentrated in the Ukraine, Russia, Poland, Turkey, and Germany, whereas sweet cherry production is concentrated in Iran, Turkey, Italy, and Germany. Sweet cherry production is increasing in the Southern Hemisphere, in New

Table 1 Production and utilization of sweet and sour cherries internationally (derived from FAO 1996–2000 statistics)

Country	*1000s of tons per year*		*Percentage destined for processing*	
	Sweet	*Sour*	*Sweet*	*Sour*
Argentina	6.0		30–35	
Armenia	1.8	3.0	0	0
Australia	6.9		0	
Austria	24.7	4.5	10–15	20–30
Azerbaijan	10.8	5.2	0	0
Belgium/Luxemburg	6.3		0	
Bulgaria	45.3	11.9	40–50	90–100
Byelorussia	15.8	29.4	0	0
Canada	5.4	6.4	20–25	100
Chile	23.4		35–50	
People's Republic of China	1.3	0		
Czech Republic	12.6	7.3	10–20	30–40
Denmark	4.4		0	
Estonia	0.9	0.1	0–5	0–5
France	63.8		25–30	
Georgia	17.0	23.3	0	0
Germany	122.4	72.0	30–35	10–15
Greece	47.8	2.9	10	100
Hungary	20.6	53.8	60–70	80–90
India	4.0		10	
Iran	217.7	42.1	0	
Israel	1.1		0	
Italy	132.7	6.9	5–10	
Japan	17.0		35–50	
Kazakhstan	5.4		0	
Kyrgyzstan	3.5	9.0	0	0
Latvia	6.0		0	
Lithuania	17.0	2.6	10–20	90–100
Moldavia	6.7	9.2	20	90
New Zealand	1.3		0	
Norway	4.5		0	
Poland	36.8	143.8	10–30	90
Portugal	6.9	0.6	15–20	
Romania	78.1		20	80
Russia	64.2	154.0	25–35	65–75
Slovak Republic	5.7	1.5	70	90
Spain	81.7	1.5	5–10	
Switzerland	20.5		60–70	
Tajikistan	3.5	3.4	0	0
Turkey	200.0	124.0	20–25	95–100
Ukraine	61.3	193.4	30–40	95–100
UK	1.5		0	
USA	188.3	128.8	45–55	100
Uzbekistan	10.9	14.0		0
Former Yugoslavia	26.9	68.9	5–10	95–100

Zealand, Australia, and Chile, for shipment to northern markets in their winter season. In the USA, sweet cherry production is mostly in the west, primarily in Washington, but also in Oregon and California. Most USA sour cherry production occurs near the Great Lakes, primarily in Michigan with 70–75% of the US crop, which, with New York, Wisconsin, and Pennsylvania totals 90–95% of the US crop.

Maximum world production values for sweet and sour cherry are approximately 1 450 000 and 1 600 000 tonnes, respectively. Wide annual supply fluctuations, especially regionally, in both sweet and sour cherries characterize production and create high risks in product availability and price change for producers, processors, and marketers. The annual US sour cherry production has ranged, for example, from 63 600 to 141 000 tonnes, but there has been a gradual downward trend in average production since the mid-1960s to about 125 000 tonnes during recent years, with a farm value of about $50 million, but the processed value is at least three times that. Sweet cherry production, however, is increasing, especially as markets develop in Japan and the Far East of the Pacific Rim for fresh cherries grown in the western USA and elsewhere.

Growth and Management

Flowering and Fruit Set

Sweet cherry flowers are in clusters of two to four, usually borne laterally on short spurs on 2-year-old twigs or near the base of longer 1-year-old shoots. Floral initiation takes place in July, after the crop is harvested, and only on buds where the subtending leaves opened relatively early in summer. Flower buds are of the unmixed type and do not give rise to a lateral or bourse shoot. As a result, flowering spurs, unlike those on apples and pears, do not remain productive. Flowers generally have a single pistil but, in very hot summers, may form two pistils that result in undesirable double fruits. With few exceptions, e.g., 'Stella' (and its progeny), and the new 'Lapins' and 'Sweetheart' cultivars, commercial sweet cherries are self-sterile (self-incompatible), and therefore require another cultivar for pollination. There are, however, intrasterile groups, where none of the group will cross-pollinate any other member of the group. Bing, Lambert, and Napoleon are one such group.

Sour cherry flowers develop much like sweet cherry, with buds of two to four flowers on either spurs or lateral buds. Sour cherry cultivars range from compatible to self-incompatible. Montmorency, for example, is only partially compatible, but is always grown without a pollinator. Fruit set, however, clearly limits yield or the fully incompatible cultivars, but overcropping may occur in other cultivars with excessive flowering or fruit set resulting in too few leaves or leaf buds to develop fruit of an adequate size and quality.

In both sweet and sour cherry, flower development and fruit set may frequently be harmed by late frosts,

although sour cherry is hardier and generally blooms later than sweet cherry. Wide annual supply fluctuations are consequently common in major growing areas for both species owing to spring frosts or to low midwinter temperatures where a lack of wood hardiness is a contributing factor. Sweet cherries are less hardy than apples, but some sour cherry cultivars may be as hardy as the apple cultivars McIntosh or Northern Spy. In addition, the best-quality sweet cherry cultivars tend to be more susceptible to rain cracking, and do best in regions with dry summer growing conditions. Fruit also develop a good quality in regions often too cool for peaches or apricots. The climate also dictates that cherries be grown where winter chilling temperatures (about 1000 h for most sweet cherries, longer for sour cherry) are adequate to break rest, and so cherry culture is generally limited to cooler temperate regions.

Tree Size and Rootstocks

Tree size plays a central role in production of quality fruit. Dwarf trees have many advantages: light penetrates better, favoring photosynthesis, and the tree produces more and better fruit; spraying can be done more efficiently, usually with a reduced use of chemicals; and dwarf trees are easier to harvest. Cherries are no exception, but dwarfing rootstocks have not been available for either sweet or sour cherry. The common rootstocks, 'Mazzard' and 'Mahaleb,' only slightly affect tree size, if at all. 'Colt' is similar, and may be somewhat drought- and cold-susceptible. Recently, however, dwarfing rootstocks have been developed in several breeding programs, and these are being tested. For example, of 17 cherry rootstocks developed in Giessen, Germany, most produce relatively large trees, but two of these rootstocks give trees about 25% of the standard, and several rootstocks developed in Belgium may also be promising.

Harvesting and Handling

Sweet cherries are almost all hand-harvested, particularly those intended for the fresh market. Avoiding pitting and bruising throughout harvest, sorting, and packing is a major problem in delivering high-quality fruit to the fresh market. The susceptibility of some white- or yellow-fleshed cultivars to bruising may even require field packing to minimize loss. Mechanical harvesting of sour cherries for processing, however, has been a major technological development that substantially reduces grower costs. A grower and their family plus a group of six to eight high-school students can harvest mechanically as much as 200–300 hand pickers formerly did. The results include huge savings in direct labor costs, large reductions in housing costs, plus substantial savings in direct labor fringe costs. The US sour cherry industry became essentially completely mechanically harvested during the 1970s, although there have since been additional improvements in equipment and techniques.

Some aspects of cherry processing have also changed substantially, which has led to greater efficiencies and product quality in the cherry industry. Almost all sour cherry processors have adopted electric-eye sorting equipment that substantially reduces in-plant sorting labor, and destemming equipment efficiently removes the stems from mechanically harvested cherries. Although picking of sweet cherries for the fresh market is still by hand, subsequent handling has been improved dramatically very recently with the substitution of hydraulic flumes for conveyor belts to reduce bruising and pitting throughout sorting and packing.

Quality Factors

Soluble solids (primarily hexose sugars and sorbitol) and fruit color (depending on the type) are the best indicators of quality for both sweet and sour cherry, although the fruit acid level may be important in sour cherry. Except for soluble carbohydrates, vitamins A and C, and certain flavonoids that may be important as antioxidants in some cultivars, cherries are relatively low in nutrients, but calcium, iron, magnesium, phosphorus, and copper contents are high compared with apple, peach, grape, and strawberry. High-quality sour cherry fruit typically have at least 15% soluble solids, whereas sweet cherries should be nearer 20% (or higher). Standards for harvest and marketing may, however, often be lower. Optimum conditions also often vary with use. To facilitate brining (bleaching in sulfur dioxide solutions for maraschino cherries), fruit may be picked prematurely before color and soluble solids are adequate for the fresh market. Stem-fruit removal force is carefully monitored for sour cherries that will be mechanically harvested, and abscission can be brought on by treatment with ethephon, which releases ethylene, expediting abscission and fruit drop in response to mechanical shaking.

Disorders, Diseases, and Pests

Disorders

In processing brined sweet and sour cherries, the solution pocket problem involves subepidermal splits in the flesh that fill with brine solution and ruptured

cell contents. The time of harvest, degree of turgidity at brining, temperature, or any procedures that reduce either the sugar or water content of the fruit will tend to reduce the problem.

Rain cracking (swelling followed by rupture of the epidermis) of sweet cherries occurs mostly during the harvest period when the fruit is mature or almost mature and has been wet with rain for some time. The primary cause is absorption of water directly through the skin of the fruit, and not through the root system. Cultivar-cracking susceptibility has been tested extensively. In testing, Bing, one of the best quality cultivars, was worst, followed by Napoleon, Lambert, Emperor Francis, Giant, Schmidt, Yellow Spanish, and Montmorency, which did not crack. In another ranking, Bing invariably cracked very badly, and was followed by Lambert, Giant, Gil Peck, and Hedelfingen. In yet another test, Van, Merton Glory, Vega, and Vista were very susceptible, whereas Emperor Francis, Schmidt, and Sam were less susceptible, and Sue, Kristin, Ulster, and Early Rivers were least susceptible. From the long-range viewpoint, breeding programs under way ultimately may produce desirable crack-resistant cherries. Cracking may be reduced by some chemical treatments, e.g., rain activated calcium sprayers, but results with hormones, auxin (NAA, naphthalene acetic acid) and gibberellin (GA_3) applications are equivocal. In some parts of the world, e.g., Norway, covering trees with plastic film has been widely used to avoid cracking, and the method is being tested elsewhere despite the high costs.

Pitting of sweet cherry is a condition in which areas near the surface of the fruit become sunken, forming dimples or pits, and may occur before or after harvest, and there are at least three different sources: usually from bruising during handling, but also from feeding by sucking insects such as the soldier bug, and perhaps from physiologic injuries, e.g., adverse low-temperature stress during postharvest cooling or growing conditions.

Diseases

Bacterial canker, one of the most important sweet and sour cherry pathogens, is caused by two different pathogens, *Pseudomonas syringae* and *P. morsprunorum*, and is characterized by oozing of gum (gummosis) at infection sites. Disease development is most prevalent during the cool, wet periods of early spring. Crown gall, caused by *Agrobacterium tumefaciens*, can affect sweet and sour cherry rootstocks and is characterized by galls forming usually near infection sites caused by wounds, sometimes man-made, e.g., cultivation injuries, or resulting from damage from subterranean chewing insects or rodents. Some rootstocks, however, are only moderately susceptible, and some hybrids may be tolerant. Brown rot, caused by the fungi *Monilinia fructicola* or *M. laxa*, affects both sweet and sour cherries and reduces yield in infecting and decaying blossoms, twigs, and fruit. Fruit decay after harvest is also a problem. The fungi persist in mummified fruit on the tree and the ground, and infection continues from these inocula the following spring. Growing regions with cooler and drier summers provide some relief. Cherry leaf spot, *Blumeriella jaapii* (*Coccomyces hiemalis*), is the most serious disease affecting sour cherry and most ground cherries. Infection occurs in the spring on expanding leaves and continues throughout the season under favorable conditions, e.g., high humidity. Severely infected leaves become chlorotic and abscise, and if defoliation is severe, fruit may not ripen properly, and tree vigor and hardiness are reduced. Powdery mildew (*Podosphaera ocyacanthae*) is similar. Other fungal pathogens may sometimes be important, causing blights, crown or root rots, and replant (orchard reestablishment) problems. Cherry dieback is thought to be a complex of several disorders, one of which may be mycoplasma. X-disease, leafhopper-transmitted and often devastating, is also due to mycoplasma.

Several viruses cause poor vegetative growth, reduce yields, and may even result in tree death, but others are symptomless. Among the most severe is Prunus Necrotic Ringspot Virus, which is pollen transmitted and present in all cherry growing areas of the world. Little Cherry (Prune Dwarf) disease is another pollen-transmitted virus and very destructive. *Prunus* stem pitting disease is caused by the tomato ringspot virus and is spread by nematodes.

Pests

Bird damage can be very serious, and some areas may require protective netting to reduce predation. The major insect pest, the cherry fruit fly passes the winter in the soil as a pupa, adult flies emerge in late spring, and females feed on the surfaces of leaves and fruit and lay eggs in the nearly ripe fruit. On hatching, the larvae (maggots) feed on the fruit flesh. The larvae are easily killed by holding fruit near 0 °C, but fumigation, until recently carried out most often with methyl bromide, may be required to meet quarantine restrictions for shipping overseas. Other pests include black cherry aphid, plum curculio, European red mite, peach tree borer, and two-spotted mite.

Economic Problems and Future Developments

Although the sour cherry industry is facing a serious problem of excessive productive capacity in some years and persistent oversupplies, this industry has adopted new technologies and practices in the last 10 years that substantially improve its cost efficiency and productivity. Much of the newly planted acreage uses efficient trickle irrigation and closely planted orchard systems that also involve hedging techniques. These recent new techniques, especially in combination, provide large increases in yields per hectare and hence substantial reductions in costs.

Considerable genetic diversity still exists in Eastern Europe and Russia, the center of origin for both sour and sweet cherries. Although breeding programs have been limited, exploiting that diversity should do much to overcome the growing, handling, and processing problems that face growers using the industry standards, the sweet 'Bing' and the sour 'Montmorency' in the USA. Sweet cherry growers especially need dwarfing rootstocks and spur types for growth control, and all growers need disease and pest resistance, less self-sterility, and cultivars with a range of maturities so that there are longer seasons for fresh markets. Sweet cherry growers also need raincracking resistance and, for postharvest fresh-market quality, resistance to bruising. Sour cherry growers need new cultivars for diversifying and strengthening marketing options, e.g., fresh and frozen juice products, dyes for cosmetics and the food processing industry, and dry stem scars and small freestone pits to facilitate handling and processing. A combination of new marketing strategies and products for sour cherries and advances in breeding of both sweet and sour cherries would clearly benefit the economic potential of the entire cherry industry.

See also: **Preservatives**: Classifications and Properties

Further Reading

Brettin TS, Karle R, Crowe EL and Iezzoni AF (2000) Chloroplast inheritance and DNA variation in sweet, sour, and ground cherry. *Journal of Heredity* 91: 75–79.

Brown SK, Iezzoni AF and Fogle HW (1996) Cherries. In: Janick J and Moore JN (eds) *Fruit Breeding, Vol. I: Tree and Tropical Fruits*, pp. 213–255. New York: John Wiley.

Iezzoni A, Schmidt H and Albertini A (1990) Cherries (*Prunus* spp.). *Acta Horticultura*.

Perry R (1987) Cherry rootstocks. In: Rom RC and Carlson RF (eds) *Rootstocks for Fruit Crops*, pp. 217–264. New York: Wiley Interscience.

Webster AD and Looney NE (eds) (1996) *Cherries: Crop Physiology, Production and Uses*, p. 464. New York: Oxford University Press.

Westwood MN (1978) *Temperate Zone Pomology*, p. 428. San Francisco, CA: W.H. Freeman and Co.

CHESTNUTS

N A M Saleh, National Research Centre, Cairo, Egypt

Global Distribution

Castanea is a genus of about 12 species native to the north temperature regions, belonging to the family Fagaceae with close botanical relationship to birch. The sweet chestnut (*Castanea sativa* Mill. synonyms: *C. vulgaris* Lam., *C. vesca* Gaertn., and *C. castanea* Karst.) is native from southern Europe, north Africa, and Asia Minor to China. The main center of origin of chestnut is thought to be located in the Ponto-Caucasian region. It is cultivated in many parts of the Himalayas, especially in Punjab and the Khasia Hills. Naturalized in central, western, and northern Europe, and introduced on the Pacific Coast of the USA, Australia, and New Zealand, it is commonly known as European, Spanish, Italian, Eurasian, or sweet chestnut.

Description

Chestnut is a deciduous tree, which reaches heights of up to 30 m. The shiny leaves, of about 20 cm in length, are the most characteristic part of the tree. They are lanceolate, with serrated margins, and have 15–20 pairs of straight, parallel veins. The female flowers are green and rounded, growing at the base of male catkins or in separate round catkins. The male flowers are long, slender, creamy-white and grow upright in catkins. The flowers appear in June to early July. Fruits are in groups of one to three single-seeded nuts, often 2–3.5 cm in diameter. Shiny brown, with paler bases, they are thickly pubescent at the tip, bearing a short-stalked perigynium with its

persistent styles. The seed has a large embryo but no endosperm. Nuts of the European chestnut are generally dark brown in color and often lightly striped. The nuts contain an edible kernel, enclosed in a thin, tough, and astringent skin (pellicle) and surrounded by a spinous, cup-like organ termed a cupule. The latter is a unique feature of the family. The kernel of European chestnut is often deeply grooved, with the fibrous pellicle folded into the grooves, making it difficult to separate this bitter covering from the kernel. There are several cultivars of chestnut, differing in the quality of kernels; the forms grown in India, for example, are inferior to those cultivated in Europe.

Cultivars

In the past, only two cultivars were known: Asplenii-folia with narrow leaves, often linear and irregularly lobed, and Macrocarpa with large fruit. Later, the following examples were cultivated: Marron Combale, Marron Noursillard and Marron Quercy, which originated in France. All have very large, light- to dark-brown nuts and are very productive. Numbo and Paragon are the most frequently grown cultivars in the USA: they have medium-large, roundish nuts of fair quality, and bear regularly. Ridgely, which originated in Dover, Delaware, has fair-sized nuts, of very good quality and flavor, with two to three nuts per bur; it is vigorous and productive. Rochester and Comfort are grown to a limited extent. A number of factors are involved in choosing a given cultivar: (1) resistance to chestnut blight (caused by *Cryphonectria parasitica* Murr.); (2) size of fruit, as large-sized nuts are more profitable to sell; and (3) the ease with which the 'skin' or pellicle can be removed from the kernel.

Commercial and Industrial Importance

Uses

Chestnuts are grown as ornamentals and for the edible nuts. Trees often bear nuts 2 or 3 years after plantation. Propagation is by budding or grafting on chestnut stocks. They are naturally tolerant of acid soils and present no special cultural difficulties on well-drained land. The chestnut blight may be injurious or fatal to susceptible species, and weevils may be troublesome.

The Fagaceae family is the source of some of the most important hardwood timbers of the world, the most notable being oak, beech, and chestnut. The timber of the sweet chestnut is almost equal to oak in strength and durability, and the one is often substituted for the other. Coppiced chestnut is in great demand for fencing, which lasts for very many years, even when not treated with creosote. The relatively hard, durable, fine-grained wood is easy to split but not easy to bend (density, 593–865 $kg\,m^{-3}$). It is used for general carpentry, railway ties, and the manufacture of cellulose. The bark and wood are used for tanning. Wood and burs may be used for firewood or for the production of charcoal.

European chestnuts are grown for the kernels of the nuts, and are extensively eaten by humans and animals. The nuts are used as a vegetable, boiled, roasted, steamed, puréed, or in a dressing (stuffing) for poultry and meats. In some European mountainous regions, chestnuts replace wheat and potatoes in the form of chestnut flour bread, and mashed chestnuts. Flour made of ground chestnuts is said to have provided a staple ration for the Roman legions. In Italy, chestnuts are prepared like a stew with gravy. Dried nuts are used for cooking purposes in the same way as the fresh nuts, or eaten like peanuts. They are also used in brandy, in confectionery, in desserts, as a coffee substitute, for thickening soups, and as a source of oil. The largest nuts, called marron, command the highest price and are used to make the famous French delicacy, *marrons glacés*. Culled chestnuts are safely used for fattening poultry and hogs, as well as for feeding cattle.

Yields

Chestnut trees begin to produce at an early age, and can produce two tons or more of fresh nuts per hectare. Yields average from 45 to 136 kg of nuts per tree. In 60–80-year-old stands in Russia, yields average 770 $kg\,ha^{-1}$ and up to 1230 $kg\,ha^{-1}$ in better stands. Italy reports *c.* 1100 $kg\,ha^{-1}$, France *c.* 1500–2200 $kg\,ha^{-1}$, and Spain *c.* 2800 $kg\,ha^{-1}$. In the best years, 5000 $kg\,ha^{-1}$ are reported. Nuts are marketed to a limited degree, but are mostly locally cultivated and used.

Biological Properties

The Greek physician Dioscorides noted the *Castanea sativa* as a tree was a fruitful source of medicinal remedies. The leaves, twigs, bark, flowering catkins, and spiny cases of the nuts are good astringents, which cause mucous membranes and skin to contract. They were therefore used to control bleeding, aid in healing wounds, and against diarrhea. The leaves, which contain demulcents, were used in teas to soothe irritated throats and in relieving symptoms of coughs and colds. The nuts are also reported in folk remedies for the treatment of fever, hematochezia, infections, sores, inflammation, wounds, sclerosis, kidney ailments, myalgia, nausea, and stomach ailments.

Sweet chestnut (*C. sativa*) is related to the beech family (Fagales) with close botanical relationship to birch, which is known as a major pollen allergen in Europe. It is not surprising that *Castanea* pollen can cause allergies – particularly in areas where *Castanea* trees are frequently found. Southern Switzerland (Lugano) is reported to experience the highest *Castanea* pollen loads and hence high *Castanea* allergy rates. It is reported that patients allergic to pollen are affected by oral allergy syndrome (allergy to food resulting from contact between food and the oral mucosa). A reported case of oral allergy syndrome to chestnut appears to be a manifestation of immediate immunoglobulin E-dependent hypersensitivity. It was also reported in a second study that a patient with latex allergy has a high risk of contact urticaria or even anaphylaxis to the related fruits such as chestnut. Class I chitinases are relevant allergens of avocado and chestnut, and could be the panallergens responsible for the latex-fruit syndrome.

Chemistry

General

C. sativa leaves proved to contain two flavonoid aglycones (quercetin and myricetin) along with their glycoides. A study of *Castanea* species indicated that catechol tannins were more abundant in buds at the tip and base, while bud scales contained about 80% of the tannis of the whole bud. The leaves, bark, and wood also contain tannins. High concentrations of ellagic acid were found to be present in the hardwood of European chestnut (89 mg g^{-1} wood) as compared to those of a number of red and white oaks (19–63 mg g^{-1} wood). After mimosa and quebracho extracts, chestnut extract is the third most important vegetable tannin used for leather production. It is only the wood (8–13% tannins), however, that is used in the tanning industry and for commercial preparation of the chestnut extract. The chestnut bark contains a number of polyphenols based on ellagitannins, galloylesters, phenolic glycosides, and hamamelitannins. Countries manufacturing the extract are France, Italy, and the USA. Commercial liquid extracts contain 29–49% tannin, while solid extracts contain 56–76% tannin.

Chestnut seeds are desiccation-sensitive and have high moisture contents in the mature stage. For this reason, they are possibly better targets for pathogenic fungi than nonrecalcitrant seeds, which might explain in turn why chitinases are abundant and stable proteins in chestnut cotyledons. A study of class Ib endochitinase from *C. sativa* cotyledons showed *in vitro* antifungal properties when assayed against *Trichoderma viride*. (*See* **Tannins and Polyphenols**.)

Nutritional Composition

Recent reports of the US Department of Agriculture for nutritional composition of unpeeled and roasted chestnut seeds are outlined in Table 1. The chemical composition is similar to that of wheat, and the starch is easily digested after cooking. Spanish chestnuts are reported to contain 2.87–3.03% ash, 9.61–10.96% total protein, 2.55–2.84% fiber, 73.75–77.70% total nitrogen free extract, and 7.11–9.58% fat. Chestnuts from 19 natural stands in southern Yugoslavia showed a total fat content of 4–5%; oleic and linoleic acids predominated, followed by palmitic acid. (*See* **Wheat**.)

Handling and Storage

Harvesting

The nuts fall to the ground within about a week after they have become ripe in the hulls. Once fallen, they are subject to depredation by birds, squirrels, and mold. Traditionally, mature chestnuts have been gathered from the ground after they fall. Some cultivars stick, and shaking or jarring the limbs is useful. In other cultivars, burs open, and nuts fall to the ground. Burs that fall and do not open can be made to shed their nuts by pressure of the feet, or by striking with a small wooden mallet. Some harvesters use heavy leather gloves and twist nuts out of burs by hand.

Unopened burs stored in a humid, cool 12–18 °C location will continue to mature and open in a week. In gathering nuts, the collector usually has two pails or containers, one for first-grade, perfect nuts, the other for culls. Some sort of harvest mechanization is essential for the development of a chestnut orchard industry.

Storing

Fresh chestnuts contain 40–50% carbohydrate, mostly in the form of starch, about 5% oil, 5% protein, and 40% moisture. They are highly perishable because they lose water rapidly at normal room temperature and humidity, causing the kernel to become hard and incapable of germinating. Numerous fungi and bacteria attack the nuts, causing decay and spoilage. High temperatures, as a result of remaining in the sun, cause rapid kernel deterioration. The nuts should be stored at 0–4 °C under conditions of high humidity but no free moisture. (*See* **Storage Stability: Mechanisms of Degradation**.)

Table 1 Nutrients of unpeeled and roasted chestnuts

Nutrient	*Units*	*Value per 100 g of edible portion (unpeeled)*	*Value per 100 g of edible portion (roasted)*
Proximates			
Water	g	48.650	40.480
Energy	kcal	213.000	245.000
Energy	kJ	891.000	1025.000
Protein	g	2.420	3.170
Total lipid (fat)	g	2.260	2.200
Carbohydrate, by difference	g	45.540	52.960
Fiber, total dietary	g	8.100	5.100
Ash	g	1.130	1.200
Minerals			
Calcium, Ca	mg	27.000	29.000
Iron, Fe	mg	1.010	0.910
Magnesium, Mg	mg	32.000	33.000
Phosphorus, P	mg	93.000	107.000
Potassium, K	mg	518.000	592.000
Sodium, Na	mg	3.000	2.000
Zinc, Zn	mg	0.520	0.570
Copper, Cu	mg	0.447	0.507
Manganese, Mn	mg	0.952	1.180
Selenium, Se	mg		1.200
Vitamins			
Vitamin C, ascorbic acid	mg	43.000	26.000
Thiamin	mg	0.238	0.243
Riboflavin	mg	0.168	0.175
Niacin	mg	1.179	1.342
Pantothenic acid	mg	0.509	0.554
Vitamin B_6	mg	0.376	0.497
Folate	mcg	62.000	70.000
Vitamine B_{12}	mcg	0.000	0.000
Vitamin A, IU	IU	28.000	24.000
Vitamin A, RE	µg_RE	3.000	2.000
Vitamin E	mg αTE		1.200
Lipids			
Fatty acids, saturated	g	0.425	0.414
4.0	g		0.000
6.0	g		0.000
8.0	g		0.000
10.0	g		0.000
12.0	g		0.000
14.0	g	0.010	0.010
16.0	g	0.384	0.373
18.0	g	0.021	0.021
Fatty acids, monounsaturated	g	0.780	0.759
16.1	g	0.021	0.021
18.1	g	0.749	0.728
20.1	g	0.010	0.010
22.1	g		0.000
Fatty acids, polyunsaturated	g	0.894	0.869
18.2	g	.798	0.776
18.3	g	0.095	0.093
20.4	g		0.000
20.5	g		0.000
22.5	g		0.000
22.6	g		0.000
Cholesterol	mg	0.000	0.000
Phytosterols	mg	22.000	
Amino acids			
Tryptophan	g	0.027	0.035
Threonine	g	0.086	0.113
Isoleucine	g	0.095	0.125
Leucine	g	0.143	0.188
Lysine	g	0.143	0.188
Methionine	g	0.057	0.075
Cystine	g	0.077	0.101
Phenylalanine	g	0.102	0.134
Tyrosine	g	0.067	0.088
Valine	g	0.135	0.178
Arginine	g	0.173	0.227
Histidine	g	0.067	0.088
Alanine	g	0.161	0.212
Aspartic acid	g	0.417	0.549
Glutamic acid	g	0.312	0.410
Glycine	g	0.124	0.164
Proline	g	0.127	0.167
Serine	g	0.121	0.159

Nuts should be picked up every morning and stored in sacks, if they are to be shipped at once. If they are to be kept for a while, they should be piled on the floor to sweat. The pile should be stirred twice a day for 2 days; then the nuts should be sacked. Nuts should always be stored in a manner so that air can circulate freely. A practical method for storing small quantities of nuts is to mix the freshly harvested nuts with dry peat moss in plastic bags, close the bags, and refrigerate. Properly stored, they will keep satisfactorily for 6 months and have been kept for as long as 3 years.

Nuts to be eaten raw should be 'cured' for several days by allowing the kernel to dry to a spongy texture. The amount of free sugar and sweetness is maximized in this way.

Seed Propagation

Chestnut seeds require a moist cold treatment of at least 1–2 months at 0–4 °C to insure good and uniform germination. Stored nuts should not be subjected to temperatures much below freezing (irreversible damage begins to occur around − 4 °C) or above 7 °C.

See also: **Storage Stability**: Mechanisms of Degradation; **Tannins and Polyphenols**; **Wheat**: Grain Structure of Wheat and Wheat-based Products

Further Reading

Allona I, Collada C, Casado R, Paz-Ares J and Aragoncillo C (1996) Bacterial expression of an active class Ib chitinase from *Castanea sativa* Cotyledons. *Plant Molecular Biology* 32: 1171–1179.

Antico A (1996) Oral allergy syndrome induced by chestnut. *Annals of Allergy, Asthma, and Immunology* 76: 37–40.

Bailey LH and Bailey EZ (1976) *Hortus Third. A Concise Dictionary of Plants Cultivated in the United States and Canada*, pp. 230–231. New York: Macmillan.

Bianco MA, Handaji A and Savolainen H (1998) Quantitative analysis of ellagic acid in hardwood samples. The Science of the Total Environment 222: 123–126.

Diaz-Perales A, Collada C, Blanco C *et al.* (1998) Class I chitinases with hevein-like domain, but not class II enzymes, are relevant chestnut and avocado allergens. *Journal of Allergy and Clinical Immunology* 102: 127–133.

Duke JA (1989) *CRC Handbook of Nuts*, pp. 90–92. Boca Raton, Florida: CRC Press.

Jaynes RA (ed.) (1979) *Nut Tree Culture in North America*, pp. 111–127. Northern Nut Growers Association.

Oraguzie NC, McNeil DL and Thomas MB (1998) Examination of graft failure in New Zealand chestnut (*Castanea* spp) selections. *Scientia Horticulturae* 76: 89–103.

Stebbins RL (1990) Requirements for a United States Chestnut Industry. In: Janick J and Simon JE (eds) *Advances in New Crops*, pp. 324–327. Portland, Oregon: Timer Press.

Tomitaka A, Matsunaga K, Akita H, Suzuki K and Ueda H (2000) Four cases with latex allergy followed by anaphylaxis to chestnut (article in Japanese). *Arerugi* 49: 327–334.

Villani F, Pigliucci M and Cherubini M (1994) Evolution of *Castanea sativa* Mill. in Turkey and Europe. *Genetical Research of Cambridge* 63: 109–116.

Chicken *See* **Poultry**: Chicken; Ducks and Geese; Turkey

CHICORY BEVERAGES

K Löhmar and V Theurillat, Nestlé Product Technology Centre Orbe, Orbe, Switzerland

Introduction

Chichorium intybus L., the wild chicory, has been adapted to a 2-year vegetation cycle. The cultivation of chicory started in several European countries at the end of the eighteenth century and spread from there to other countries, e.g., the USA, India, and South Africa. Only the root of the chicory plant is used for making various beverages; the plant is shown in **Figure 1**. This article describes the cultivation of chicory, the processing of the roots to roast and ground blends and to soluble beverage powders in which chicory is an ingredient. Furthermore, the composition of chicory and changes taking place during processing are discussed.

Figure 1 Photograph of a harvested chicory plant. Reproduced from Chicory Beverages, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Cultivation, Drying, and Roasting

In central Europe, sowing is at the end of April or at the beginning of May. Three types of seed are

employed: ordinary naked seed, seed coated with dyes and fungicides, and seed pelleted with nutrients, fungicides, and insecticides. Coated and pelleted seeds facilitate the control of the right sowing depth. Protection and the conditions for germination of the seed embryo are optimal in pelleted seed.

In order to produce chicory roots of a suitable size for industrial processing, a space of about 600 cm^2 per plant is recommended. The distance between the seed rows is usually 45 cm; pelleted seeds are drilled within the row at a distance of 6 cm. After emergence and when the plants have reached the three- or four-leaf stage, they are thinned out to distances of about 14 cm in the row. The plant density should not exceed 160 000 plants per hectare in order to attain the right root size and a good field yield.

Chicory requires a vegetation period of about 160 days. In Europe, the chicory harvest is from the end of September to the middle of November. Machines are used which carry out the entire harvesting operation: cutting of leaves, lifting of the roots, their collection in containers, and transfer to trailers. The yield of chicory roots varies from 30 to 45 t ha^{-1}, depending on the climate, weather, and soil conditions. The harvested chicory roots are delivered to the drying plant, where the quantity of roots is determined by weighing the trailer loaded and unloaded. The farmers are paid for the root weight reduced by the amount of soil present. The chicory roots are stored until processing in long, extended heaps with heights of up to 3 m.

Figure 2 illustrates the processing of the fresh chicory roots to soluble powder. The drying of chicory is a seasonal operation which starts in central Europe at the beginning of October. The installation for washing the roots comprises several compartments; it operates on the counterflow principle and rinses the roots with clean water during discharge. The roots are cut into slices (6×22 mm cross-section and of varying length) and cubes (lengths of sides between 12 and 16 mm). Chunks are also produced, having a thickness of more than 16 mm.

Drying of chicory reduces the water content from about 75 to 12%. Dried chicory with a water content of less than 12% is considered to be stable and can be kept in storage for several years. A rotary drum drier is the most frequently used equipment for drying. It is equipped with baffles inside and works continuously. Only those driers furnished with gas firing allow direct contact of the combustion gases with the cut material. The chicory moves in the slowly rotating drum in the same direction as the hot air is drawn through. The temperature of the hot air at the inlet reaches 400–500 °C; at the outlet it is about 120 °C. The dried chicory leaves the drum at a temperature of about 80 °C. The necessary cooling and the final drying are accomplished on well-ventilated conveyor belts. The chicory fragments produced through breakage and abrasion are sifted off.

Chicory is roasted in batches using drum roasters and long cycle times. Since smaller chicory pieces are roasted faster than larger ones, a narrow size distribution of the dried chicory should be used to obtain a homogeneous roast. The rotating drum is heated by a gas burner and charged with chicory.

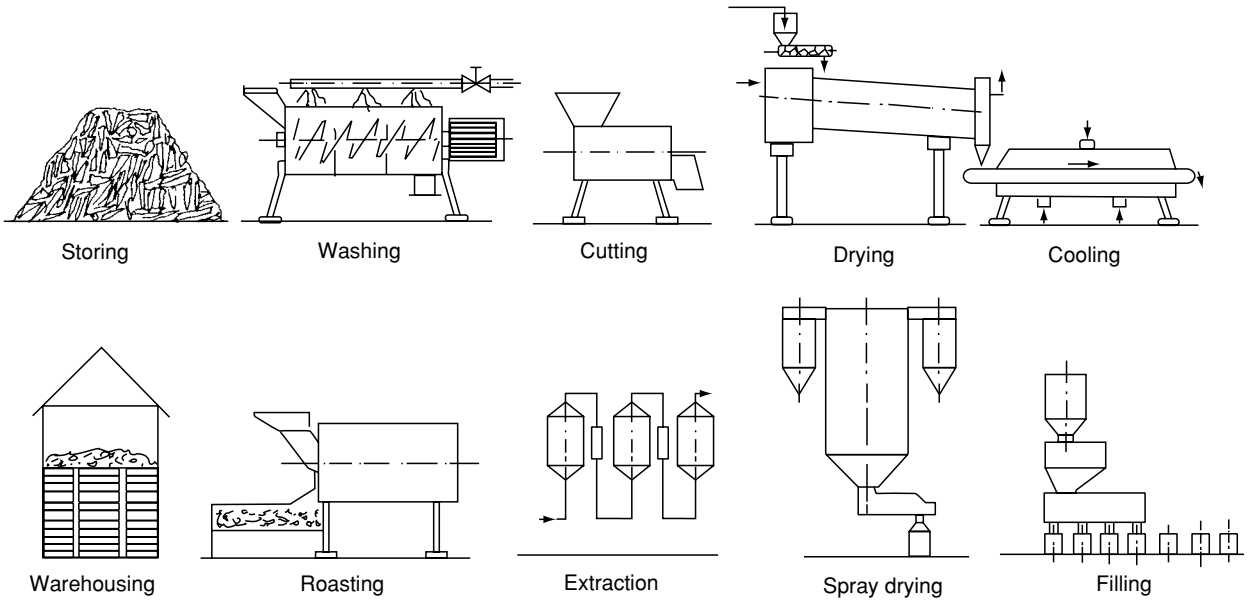

Figure 2 Processing of fresh roots to soluble powder. Reproduced from Chicory Beverages, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Vegetable oil in a quantity of about 1% of the chicory load is added to bind the dust and to prevent sticking of chicory to the roaster wall and baffles. The combustion gases are first directed around the drum and are then drawn through the drum. The moisture in the chicory is driven off and is removed by the combustion gases. In the later stages of roasting, the heat supply is reduced by steps and the heating gases are only drawn outside around the drum. The actual roasting takes place at temperatures of 170–180 °C. After a roasting time of 60–80 min the chicory reaches the required color and is discharged into the cooling screen. The total roast loss is in the range of 15–25%.

Roast and ground chicory is usually blended with coffee or cereals and marketed as coffee mixture or coffee substitute. (*See* **Coffee**: Roast and Ground.) Otherwise, it may be extracted and dried.

Extraction and Spray Drying

Roasted chicory is more frequently extracted in blends with roasted coffee or roasted cereals than alone by itself. This extraction with coffee or cereals is done in percolation batteries which consist of about six columns connected in series. Hot water is pumped into the column which contains the most spent roasted material, flows in a counterclockwise fashion through the plant and finally enters the column with fresh roasted material. The concentration of the draw-off extract varies from 15 to 30% and depends on, among other factors, the composition of the roast blend. Each column with an exhausted blend is replaced by one with a new roast blend. The feed water temperatures may be up to 180 °C with blends of coffee and chicory, or up to 140 °C with blends of cereals and chicory. Heat exchangers placed between the columns lower the extract temperature in stages down to 90 °C in the column with the fresh roast blend. The extraction yield and the draw-off concentration increase with higher temperature profiles. Since chicory contains a high proportion of soluble carbohydrate, it can be extracted efficiently at temperatures below 100 °C.

The extraction of roasted chicory by itself is performed in twin-screw conveyors or hydraulic piston presses. Pure chicory extracts are, for the most part, subsequently mixed with coffee or cereal extracts. The liquid extracts are sometimes concentrated before spray drying or mixed with glucose syrup in order to improve the drying properties of the extract. (*See* **Coffee**: Instant.)

The spray drying of the liquid extracts with soluble solid contents of 30–45% is conducted in tall towers in which the extract is pumped under high pressure through nozzles and is dispersed into small droplets. In the upper part of the tower the water in the droplets evaporates in a flow of hot air; in the lower part the exhaust air is drawn off and the separated powder is collected in bins. The instant powders, which have a moisture content of about 3%, are filled into water vapor-tight containers such as jars with sealable lids.

Composition and Varieties

The composition of two chicory varieties based on dry matter is given in Table 1. The water content of fresh chicory roots is in the range 70–80% and is influenced by the variety, the soil, and weather conditions. The solid matter of finely ground chicory roots is extractable in boiling water to an extent of 80%. Chicory is characteristic in that the main constituent of the roots is the polysaccharide inulin. This reserve carbohydrate with a molecular weight of approximately 6000 comprises up to 65% of the dry matter of the root. About 35 fructose molecules are linearly linked together with one glucose molecule at the end. The polymer chain of inulin is completed by a sucrose molecule and, therefore, it has no reducing power. In addition, chicory roots contain further carbohydrates, including usually contents of about 5% sucrose and about 15% fiber substance. During extraction, the fiber remains insoluble in the residue. Nitrogenous compounds amount to about 6% and comprise protein and free amino acids. The content of organic acids is about 3%, of which malic acid and citric acid are about 30% each. Mineral substances are present at about 5% in the root, the most important components being potassium, calcium, phosphate, and sulfate. Furthermore, the roots contain lactucin and lactucopicrin in a total amount of about 0.4%. These sesquiterpene lactones are

Table 1 Composition (grams per 100 g) of chicory varieties

	Variety superior in	
	Root quality	*Field yield*
Dry matter[a]	26.9	21.7
Extractable matter[b]	82.2	80.0
Composition[b]		
Inulin	64.1	57.8
Sucrose	5.3	7.8
Protein	4.5	5.5
Amino acids	1.1	1.5
Organic acids	2.6	3.2
Fiber	13.6	16.1
Minerals	4.0	5.0
Others	4.8	3.1

[a]Based on root weight.
[b]Based on dry weight.

Table 2 Yields (tonnes per hectare) of chicory varieties

	Variety superior in	
	Root quality	*Field yield*
Fresh roots	40.7	46.9
Dry matter[a]	10.9	10.2
Soluble acids[a]	9.0	8.1

[a]Calculated with the values of Table 1.

responsible for the bitter taste of fresh and dried chicory. Both lactones are completely decomposed during roasting. (*See* **Acids**: Natural Acids and Acidulants; **Amino Acids**: Properties and Occurrence; **Carbohydrates**: Classification and Properties; **Minerals – Dietary Importance**; **Protein**: Chemistry.)

It is known from screening tests of chicory varieties that a negative correlation exists between the root quality and the field yield. Breeding of chicory aims at combining both attributes. A variety with a superior root quality has a higher content of inulin, and dry and soluble matter, and, conversely, a variety with a superior field yield has a high root weight, and a high content of protein and minerals. For these two opposing chicory varieties, yields of fresh roots and calculated yields of dry matter and soluble solids are given in Table 2.

Changes in Composition During Processing

Higher temperatures in the root heaps during longer storage periods induce the start of hydrolysis of the polysaccharide by the enzyme inulinase present in the roots. The generated sugars favor loss of material as a result of respiration of the roots and also by decay through microbial activity. Fresh roots should, therefore, not be stored for long periods at high temperatures in order to minimize inulin hydrolysis and other causes of losses. The ideal storage conditions would be temperatures of about 5 °C and air humidities of about 95%, which normally cannot be achieved in practice.

Very high temperature profiles during drying cause puffing of the wet chicory pieces. Hollow spaces remain inside after drying which reduce the firmness of the cubes or slices and increase the tendency to break more easily during transport. Chicory pieces with larger dimensions usually leave the drum drier still with moist centers. During storage, enzymatic hydrolysis of the inulin proceeds at a rate dependent on the moisture content, and the sugar content increases. Dried chicory should normally show a bright color, a low content of reducing sugar, and a high content of extractable solids and inulin. Darker outer colors and burnt edges of dried chicory enhance water uptake during storage and make it more difficult to obtain a homogeneous roast. Higher sugar contents of dried chicory are sometimes accepted in view of the taste of the product, but they impair further processing due to the higher hygroscopicity.

Roasting of chicory results in the formation of specific color, flavor, and aroma compounds which are characteristic of the product. The constituents of chicory, such as inulin, sucrose, protein and free amino acids, partially or totally undergo various changes. Most alterations in the chemical composition are connected with two browning reactions – the Maillard reaction and the pyrolysis of the inulin. The Maillard reaction takes place between the reducing sugars and the amino groups of free amino acids and proteins, giving the melanoidins (brown polymeric materials) and volatile aldehydes. These give chicory its typical flavor. The pyrolytic degradation of inulin yields high-molecular-weight caramel compounds and low-molecular-weight carbonyl compounds, such as aldehydes, ketones, and organic acids. These compounds likewise contribute to the color and the taste of roasted chicory. (*See* **Browning**: Nonenzymatic.)

The analytical values in Table 3 illustrate the changes in the carbohydrate composition and the free acidity during chicory roasting. The decomposition of inulin and the formation of free acids steadily proceed with increasingly darker roasting. The contents of reducing sugars, of fructose, and of glucose first increase during roasting, reach a maximum, and decrease again. The content of reducing sugars in, for example, cubes or slices reaches a maximum value of 15%. Insoluble matter is formed by the strongly exothermic pyrolysis of inulin and lowers the amount of extractable material, especially of dark roasted chicory. Light and medium roasted chicory are

Table 3 Characteristic values (grams per 100 g of dried matter) of dried and roasted chicory

		Roasted		
	Dried	*Light*	*Medium*	*Dark*
L-Color value of powder	77.0	46.4	40.0	32.5
Extractable matter	83.1	82.8	82.2	78.3
pH value[a]	5.7	4.7	4.6	4.4
Acid degree[a,b]	12	27	31	40
Reducing sugars	1.9	8.4	14.1	12.1
Free fructose	0.5	2.4	4.6	3.4
Free glucose	0.1	1.0	2.2	1.9
Inulin and sucrose	67.1	52.6	39.1	24.8

[a]Measured in a 1% extract solution.
[b]Defined as millimoles of sodium hydroxide per 100 g of soluble solids.

Table 4 Increase of free acids and free sugars during storage of a liquid chicory extract (23% dry matter, 61% inulin and sucrose) for 1 h at different temperatures

		Treated at				
	Original	*60 °C*	*80 °C*	*100 °C*	*120 °C*	*140 °C*
pH value of extract	4.7	4.7	4.6	4.5	4.1	3.6
Acid degree[a]	21	22	23	26	34	68
Free fructose[b]	2.7	3.0	3.7	8.2	54.3	44.7
Free glucose[b]	1.2	1.2	1.4	1.7	5.3	8.8

[a]Defined as millimoles of sodium hydroxide per 100 g of soluble solids.
[b]Grams per 100 g of dry matter.

extractable in boiling water to about 80%, whereas the extractable content of dark roasted chicory is lower. In addition, the heat generated inside the chicory pieces by the pyrolytic reaction accelerates the roasting and leads to charring of the cores.

During extraction, free acids already present in roasted chicory hydrolyze inulin to monosaccharides. Fructose, itself thermally unstable, is subjected to further decomposition, the extent of which depends strongly on the extraction conditions. Shorter extraction times and lower water temperatures are considered favorable to minimize the amount of free fructose and free acids in the chicory extracts.

The analytical values stated in Table 4 show that temperatures below 100 °C should be applied for the extraction of chicory.

Fructose has a steeper sorption isotherm than inulin and is very hygroscopic. As a result, light roasted chicory with more inulin and less fructose is preferably used for the production of instant powders. Chicory extracts have a higher hygroscopicity and an increased stickiness, in contrast to coffee or cereal extracts. This means that, compared to pure coffee, the output capacity of the spray drier is reduced by extracts produced from roasted coffee–chicory or cereal–chicory blends and, in particular, by pure chicory extracts. (*See* **Fructose**.) In addition, spray drying of extracts containing chicory-soluble solids is more difficult and requires certain precautions.

Taste and Physiological Effects

Infusions of light roasted chicory have a sweetish and mild flavor. In combination with coffee or cereals, the chicory imparts to the beverage a smooth and round taste. With darker roasting of the chicory, the sweet taste diminishes and the strength, astringency, acidity, and bitterness of the brew increase.

Information concerning the physiological effects of roasted chicory is exceedingly scarce. A lot of practical advice compiled over the years, concerning natural medicine, is found in old literature. According to this knowledge, infusions of roasted chicory are beneficial and contribute to the well-being of the body. Chicory brews are supposed to be mild diuretics and are said to have sedative effects. It has been known for a long time that chicory promotes digestion by stimulating the secretion of the gastrointestinal glands and especially by increasing bile production. Some investigators claim that the digestion of milk is facilitated by the addition of chicory because the finer milk particles formed in the stomach are more easily digested. Modern scientific studies confirm these conclusions. Tests performed on animals and humans substantiate the fact that chicory is perfectly safe and has no adverse effects. Some physiological effects were thought to be related to the sesquiterpene lactones contained in fresh or dried chicory. These lactones are decomposed during roasting and therefore are absent from roasted chicory.

Soluble chicory powders can contain up to 60% inulin or oligomer compounds derived from it. Inulin and fructose oligomers are soluble dietary fibers. This means that they are not hydrolyzed in the small intestine. In the intestine, however, they are fermented by the microflora, so a considerable amount of bacterial mass is produced and bifidobacteria development is promoted. These fibers are well tolerated. As inulin is not hydrolyzed in the stomach and the small intestine, it has no influence on the blood glucose and can be used in diabetic diets. (*See* **Coffee**: Physiological Effects.)

Use as an Adulterant and Analysis

Roast and ground coffees or instant coffee powders may be adulterated by mixing chicory with coffee after roasting, before extraction or even after drying. Microscopic, physical, and chemical methods are applied to detect and to analyze the adulteration of roasted or soluble coffee with chicory. Cell structures from chicory can be identified by the examination of suspicious samples of roasted coffee under the microscope. The extractability of roasted coffee and roasted chicory differ enormously: finely ground coffee is extractable in boiling water to about 30%, chicory to about 80%. From the increase in coffee extractability, the portion of chicory in roast and ground coffee can be estimated. Several methods depend on the analysis of specific constituents present in either coffee or chicory, but not both. The determination of the content of total fructose appears most suitable, since the content of fructose in chicory is relatively high compared to the content of caffeine and chlorogenic acids in coffee. Although all these methods detect the adulteration of coffee, it is

difficult to specify the blend quantitatively. (*See* **Adulteration of Foods:** History and Occurrence.)

See also: **Acids**: Natural Acids and Acidulants; **Adulteration of Foods**: History and Occurrence; **Amino Acids**: Properties and Occurrence; **Browning**: Nonenzymatic; **Carbohydrates**: Classification and Properties; **Coffee**: Roast and Ground; Instant; Physiological Effects; **Fructose**; **Minerals – Dietary Importance**; **Protein**: Chemistry

Further Reading

Beitter H and Schröder C-H (1970) *Handbuch der Lebensmittel-Chemie*, vol. VI, pp. 96–138. Berlin: Springer Verlag.

Clarke RJ and Macrae R (1987) *Related Beverages, Coffee*, vol. 5. London: Elsevier.

Maier HG (1981) *Kaffee*, pp. 146–194. Berlin: Parey.

Niness KR (1999) Inulin and oligofructose: what are they? *Journal of Nutrition* 129: 1402S–1406S.

CHILDREN

Contents

Nutritional Requirements

B Caballero, Johns Hopkins University, Baltimore, MD, USA

Nutritional Requirements

Assessment of Nutritional Status

Normal growth is one of the most important indicators of nutritional status. Monitoring of weight and height (or recumbent length when appropriate) should be performed at regular intervals, with frequency depending on the age of the child. Infants are usually measured monthly during the first 6 months of life, and at bimonthly intervals during the remainder of the first year.

However, it is difficult to determine the adequacy of children's diet only by physical examination. Many clinical signs of nutrient deficiencies only appear in the advanced stages. Thus, interrogation on food habits and intake, and even use of dietary records may be necessary when the possibility of inadequate intake is considered.

Nutrition from 1 to 5 Years of Age

The first 5 years of life are among the periods of more rapid growth. Major changes also occur in the relative size of the body regions and in the proportions of water, fat, and lean mass tissues. Over these years, nutrition must accommodate a changing pattern of energy, protein, and micronutrient needs. (*See* **Infant Foods:** Weaning Foods; **Infants:** Weaning.)

The growth pattern after the first year of life shows periods of both rapid and slow gain in stature and weight. Children tend to grow 12 cm during the second year, 9 cm in the third, and 7 cm or less thereafter. (*See* **Growth and Development.**)

After the first year, healthy children should be able to eat virtually the same as adult members of the family. Indeed, it is at this stage that family dietary practices have the most significant impact on the child's preferences and food intake patterns. It is thus important for parents to be aware of their role in educating and forging sound dietary habits in their children.

The feeding behavior of children at this age is much more variable than previously. Children have a much wider repertoire of eating behaviors, and should be allowed, within reason, to practice them. As a rule, a healthy child should be allowed to eat as much or as little as he or she wishes. Forced feeding or threats around food do not lead to good nutrition, but rather to sometimes long-term confrontation, to the detriment of actual food intake. Furthermore, there is evidence that healthy children have a quite accurate mechanism of maintaining energy balance by compensating for deficient or excessive food intake over periods of days or even weeks. (*See* **Infants:** Feeding Problems.)

Nutrition over 5 Years of Age

The growth pattern between 5 and 10 years of age is usually more stable than during earlier years,

averaging about 5 cm per year. Weight gain is around 2 kg per year at 5–6 years of age, and increases rapidly as puberty approaches, reaching as much as 4 kg per year by age 10. Physical activity also increases markedly during school years, as children become more involved in sports and outdoor activities in general.

Recommended Nutrient Intake

Gender differences in nutrient requirements are usually not substantial until preadolescence and menarche. At that time, genders begin to diverge in terms of body fat and skeletal muscle accumulation, which are to be reflected in nutrient requirements.

Protein and Energy Requirements

Dietary energy needs of infants are usually estimated from the *ad libitum* energy intake of healthy children growing at a normal rate. In older children estimates can be derived, as for adults, from equations based on resting energy expenditure and body size. Since recommended energy intakes should match energy expenditure (with normal growth and adequate physical activity), the best approach to estimate energy needs would be from measurements of daily energy expenditure. Until recently, the number of studies measuring daily energy expenditure in free-living individuals has been small, due mainly to the cost and complexity of the doubly labeled water technique. But the increasing number of studies with that technique is now approaching a critical number that will allow ongoing revisions of recommended energy intakes to be based on direct energy expenditure measurements. Until these revisions (by the US Institute of Medicine, the World Health Organization, and others) are released, existing recommendations are presented in Table 1.

For protein, estimates of requirements continue to be based primarily on patterns of dietary protein intakes of normal infants and children, with the consideration of nitrogen balance data when available. Requirements for indispensable amino acids have also considered the information obtained by stable isotope tracer studies (Table 2). (*See* **Protein**: Requirements.)

Vitamins and Minerals

Recommendations for most vitamins and minerals were recently revised by the US Food and Nutrition Board (*See* **Dietary Reference Values**), and are presented in Table 3. For nutrients and age groups for which there are no sufficient data to determine the traditional recommended dietary allowance (RDA), the committee introduced the adequate intake (AI). This variable indicates observed intakes in healthy groups, and thus is assumed to be adequate. It can be used as an RDA, but it is obviously less accurate.

Food Additives

Over the years, several investigators have proposed that certain food additives could have a major behavioral impact on young children. But in spite of many years of research and controversy, there is no compelling evidence supporting those claims. Nevertheless, there are documented intolerances to some additives such as sodium glutamate and sulfites in adults, which can also occur in children.

Dental Development

Dental development is another important milestone during school age. Dental caries can appear early, and some surveys found that over 50% of children aged 3–4 years of age had dental caries. Major nutrients

Table 1 Recommended energy and protein intake in children aged 1–10 years

Age (years)	*Energy ($kcal\,kg^{-1}\,day^{-1}$)*	*Protein ($g\,kg^{-1}\,day^{-1}$)*
1–2	105	1.20
2–3	100	1.15
3–5	95	1.10
5–7		1.0
Boys	90	
Girls	85	
7–10		1.0
Boys	78	
Girls	67	

Data from World Health Organization (1985) *Energy and Protein Requirements. FAO/WHO/UNU Expert Consultation. WHO Technical Report Series 724*. Geneva: WHO.

Table 2 Essential amino acid requirements during childhood

Amino acid ($mg\,g^{-1}$ crude protein)	*Preschool child (2–5 years)*	*School child (5–19 years)*	*Adult*
Histidine	19[a]	19[a]	16
Isoleucine	28	28	13
Leucine	66	44	19
Lysine	58	44	16
Methionine and cystine	25	22	17
Phenylalanine and tyrosine	63	22	19
Threonine	34	28	9
Tryptophan	11	9[a]	5
Valine	35	25	13

Data from World Health Organization (1985) *Energy and Protein Requirements. FAO/WHO/UNU Expert Consultation. WHO Technical Report Series 724*. Geneva: WHO.

[a]Values interpolated from requirement versus age curves.

Table 3 Recommended daily intakes of vitamins and minerals for children aged 1–10 years[a]

	Age of child (years)	
	1–3	*4–8*
Vitamin A (μg day^{-1})	**300**	**400**
Vitamin D (μg day^{-1})[a,b]	5*	5*
Vitamin K (μg day^{-1})	30	55
Vitamin C (mg day^{-1})	**15**	**25**
Vitamin E[e] (mg day^{-1})	**6**	**7**
Thiamin (mg day^{-1})	**0.5**	**0.6**
Riboflavin (mg day^{-1})	**0.5**	**0.6**
Niacin (mg day^{-1})[c]	**6**	**8**
Vitamin B_6 (mg day^{-1})	**0.5**	**0.6**
Folate (μg day^{-1})[d]	**150**	**200**
Vitamin B_{12} (μg day^{-1})	**0.9**	**1.2**
Pantothenic acid (mg day^{-1})	2*	3*
Iron (mg day^{-1})	**7**	**10**
Zinc (mg day^{-1})	**3**	**5**
Iodine (μg day^{-1})	**90**	**90**
Calcium (mg day^{-1})	500*	800*
Phosphorus (mg day^{-1})	**460**	**500**
Magnesium (mg day^{-1})	**80**	**130**
Selenium (μg day^{-1})	**20**	**30**

This table presents recommended dietary allowances (RDAs) in **bold** type and adequate intakes (AIs) in ordinary type followed by an asterisk (*). RDAs and AIs may both be used as goals for individual intake. RDAs are set to meet the needs of almost all (97–98%) individuals in a group.
[a]As cholecalciferol. 1 μg cholecalciferol = 40 IU vitamin D.
[b]In the absence of adequate exposure to sunlight.
[c]As niacin equivalents (NE). 1 mg of niacin = 60 mg of tryptophan; 0–6 months.
[d]As dietary folate equivalents (DFE). 1 DFE = 1 μg food folate = 0.6 μg of folic acid from fortified food or as a supplement consumed with food = 0.5 μg of a supplement taken on an empty stomach.
[e]As α-tocopherol. α-Tocopherol includes RRR-α-tocopherol, the only form of α-tocopherol that occurs naturally in foods, and the 2R-stereoisomeric forms of α-tocopherol (RRR-, RSR-, RRS-, and RSS-α-tocopherol) that occur in fortified foods and supplements. It does not include 2S-stereoisomeric forms of α-tocopherol (SRR-, SSR-, SRS-, and SSS-α-tocopherol), also found in fortified foods and supplements.

related to dental development are fluoride, and vitamin D, calcium and phosphorus (enamel formation), vitamin A and C, zinc, and folate (gum integrity). Adequate intake of these nutrients is usually not a problem in developed societies. Of more practical importance are, perhaps, a number of dietary habits which may favor dental caries: excessive use of sugar and other fermentable carbohydrates, the nursing bottle syndrome (the baby is allowed to fall asleep while sucking a sweetened beverage, or even human milk), and poor dental hygiene. (*See* **Dental Disease:** Etiology of Dental Caries.)

A Prudent Diet for Children

There is growing awareness of the importance of diet in the development of chronic diseases, e.g., obesity and coronary heart disease. A number of changes in the current dietary practices of developed societies have been proposed, although the long-term impact of these changes is still unclear. There is also the question of how early in life a 'prudent' diet should be adopted. Special issues of nutrition in childhood, such as nutrient density, protein digestibility, and micronutrient requirements, need to be recognized. For example, there must be awareness of the potential negative impact that high-fiber or all-vegetable diets may have on protein digestibility, mineral bioavailability, and energy density. (*See* **Dietary Fiber:** Effects of Fiber on Absorption.)

See also: **Dental Disease**: Etiology of Dental Caries; **Dietary Fiber**: Effects of Fiber on Absorption; **Dietary Reference Values**; **Growth and Development**; **Infant Foods**: Weaning Foods; **Infants**: Weaning; Feeding Problems; **Protein**: Requirements

Further Reading

Caballero B (1985) Food additives in the pediatric diet. *Clinical Nutrition* 4: 200–206.

Committee on Nutrition, American Academy of Pediatrics (1998) *Pediatric Nutrition Handbook*. Elk Grove Village, IL: American Academy of Pediatrics.

US Food and Nutrition Board, Institute of Medicine (1997–2001) *Dietary Reference Intakes. Calcium and Related Nutrients* (1997); *Thiamin, B-complex Vitamins* (1999); *Vitamins C, E, Selenium* (2000), *Vitamins A, E, K, and others* (2001). Washington, DC: National Academy Press.

World Health Organization (1985) *Energy and Protein Requirements. FAO/WHO/UNU Expert Consultation. WHO Technical Report Series* 724. Geneva: WHO.

Nutritional Problems

D Bentley, Imperial College of Science Technology and Medicine, London, UK

Rickets

Deficiency of vitamin D in the growing child leads to rickets (**Figure 1**), a metabolic disorder common in English and American polluted cities during the Industrial Revolution. The term indicates a failure in mineralization of growing bone or osteoid tissue. The typical clinical picture was well described in the seventeenth century and the cause was secondary to poor nutrition, arising from both financial and educational poverty as well as the result of a hazy, smoggy atmosphere with little exposure to sunshine or ultraviolet (UV) radiation. Adequate sunshine

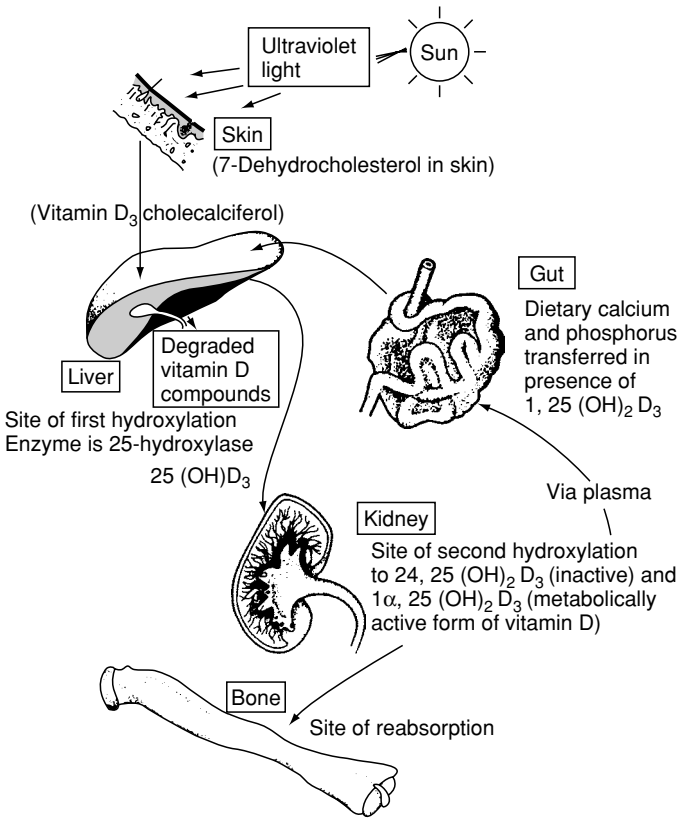

Figure 1 Vitamin D is derived from the diet and from the effect of ultraviolet light on the skin. It is hydroxylated in the liver ($25(OH)D_3$) and undergoes a second hydroxylation in the kidney to the biologically active form, 1,25-dihydroxycholecalciferol ($1\alpha 25(OH)_2D_3$), which acts on the gut and bone to increase gut absorption. $24,25(OH)_2D_3$ is 24,25-dihydroxycholecalciferol. Courtesy of Professor P. Byrne.

is not the entire explanation because rickets was common in the 1940s in southern California as it is today in India. Furthermore, from October to March no UV light reaches the earth's surface in the UK and consequently vitamin D is not formed in the skin during that period. The incidence of rickets declined after the 1927 discovery that UV irradiation of ergosterol, a vitamin D plant derivative, increased its antirachitic effect.

Skin photochemical synthesis is the major source of vitamin D, and tanning prevents further formation. Black infants have lower circulating vitamin D levels than white children. This is because white skin allows vitamin D-inducing wavelengths to pass through but black skin filters UV radiation under wavelengths of 436 nm. With the development of vitamin D-fortified milks and milk products, rickets has decreased in incidence. However, in some countries, cows' milk is not routinely fortified. Human milk contains almost 20 times more of this vitamin than unmodified whole cows' milk, yet in the USA a revival in breast-feeding has been linked to an increase in cases of nutritional rickets.

Rickets was described in Asian immigrants in Glasgow, Scotland in 1956. An estimate of the 1977 incidence of clinical overt rickets among UK Asians aged 0–16 years was 4 cases per 1000 in London and 10 per 1000 in the Midlands and the north of England. Rickets seemed to increase in the UK from 1960 to 1973, then decreased but has returned to the West Midlands in the UK (2002). Pregnant Asian women with osteomalacia risk having infants with neonatal convulsions as a result of hypocalcemia. Typically rickets is seen in breast-fed premature and growth-retarded infants and is especially prevalent in those born to Asian mothers in the UK. Phytic acid is a modifier of calcium absorption and one source is unrefined cereals such as chapattis (unleavened breads). This has been implicated in the causation of Asian rickets. High-phytate diets will increase vitamin D requirements. Infants on macrobiotic diets (e.g., unpolished rice, pulses – rich in fiber) and who do not receive vitamin D supplements are at risk. Rickets has also been described in the fetus. Adolescent Asian rickets, characteristically in children 8–14 years old, may present with gross genu valgum

(knock knee) and aching joints. Other clinical features include leg bowing, poor linear growth, delayed dental eruption, and impaired closure of the skull's fontanelle. Other skeletal deformities may develop.

Causes

The causes of rickets in children are as follows:

1. Inadequate exposure to sunlight
2. Inadequate vitamin D intake
3. Malabsorption of vitamin D, e.g., celiac disease, short-bowel syndrome, bacterial overgrowth in the bowel, deficiency of bile salts, cystic fibrosis
4. Increased enzyme activity; anticonvulsants (phenobarbital and/or phenytoin) will increase metabolic degradation of $25(OH)D_3$ by stimulating hepatic microsomal enzymes
5. Decreased enzyme activity: vitamin D-dependent rickets (pseudovitamin D deficiency; deficiency of the renal enzyme, 25-hydroxyvitamin D_1 α-hydroxylase)
6. Acquired deficiency of 1α-hydroxylase, e.g., chronic renal failure
7. Failure of 25-hydroxylation, e.g., chronic liver failure
8. Fanconi's syndrome – phosphaturia: the principal component of the disorder results in hypophosphatemic rickets

Diagnosis

Early manifestation of rickets includes:

1. Craniotabes (skull feels like a ping pong ball)
2. Frontal bossing and thickening of the skull (hot cross bun appearance)
3. Rachitic rosary (enlarged costochondral junctions of the ribs)
4. Widening of the epiphyses at the wrists

In active rickets, there is cupping and fraying of the distal ends of the forearm and parts of the skull are demineralized. There is an increased distance between distal ends of the forearm bones and the metacarpals.

Biochemical Findings

The following biochemical findings are characteristic of rickets:

1. Increased plasma alkaline phosphatase (isoenzyme originating in bone)
2. Phosphate reduced or normal
3. Calcium reduced or normal
4. Serum parathyroid hormone normal or increased

Treatment

Adequate amounts of dietary calcium and phosphate should be insured with vitamin D therapy.

Obesity

It is somewhat ironic that the commonest nutritional problem in affluent countries is obesity, in sharp contrast to the undernutrition in less privileged societies. Overweight and obesity are important public health problems and are increasing at an alarming rate. Obesity is the presence of excessive body fat and the visual identification of severe obesity is readily made. Weight in itself is an unsatisfactory way to assess fatness and particularly so in younger children where there is a poor relationship between weight and fat. (*See* **Obesity**: Etiology and Diagnosis.)

In adults one indicator of obesity is Quetelet's index (QI), which determines the body mass, or $(\text{weight})/(\text{height})^2$, but in children it is less useful because of a variation with age. The weakness of the QI is that it fails to take account of the distribution of body fat which influences health risks. However, this parameter can be related to a theoretical child with height and weight upon the 50th centile, and the measure then gives a percentage value; 90% indicates underweight, 90–110% normal weight, 110–120% overweight, and > 120% obese.

Clinical examination needs to include an assessment of pubertal development and anthropometry, which must be carefully evaluated. In addition, the observer ought to be aware of particular disorders associated with obesity, e.g., Down's syndrome, Prader–Willi syndrome, Klinefelter's syndrome and the Laurence–Moon–Biedl syndrome, as well as some endocrine diseases (hypothyroidism, Cushing's syndrome).

Body fatness can be assessed with a skinfold calliper at a number of sites, a tape measure, a stadiometer, and weighing apparatus. Bioelectrical impedance (BIA) and dual X-ray absorptiometry (DEXA) are the most widely used of the more elaborate options. Other techniques include underwater weighing and densitometry, near-infrared spectroscopy (NIR), and measurement of naturally occurring isotopes. Centile charts for triceps skinfold thickness, in boys and girls, are available. Values exceeding the 97th centile are indicative of obesity. (*See* **Obesity**: Fat Distribution.)

Prevalence

As many as one-third of the population of some industrialized nations are obese; in North America obesity affects 25% of children and adults, yet in

other communities, such as Finland, it is seen in only 3%.

Etiology

It is uncommon for a child to be obese because of an underlying pathological condition (e.g., Prader–Willi syndrome or the Laurence–Moon–Biedl syndrome). Obesity develops only if there is a positive energy balance; this is contrary to some views – there are opinions which imply that not all obese children eat excessively. Overweight children are less active than their leaner counterparts and this pattern of reduced physical activity is of importance.

Many factors are involved in the etiology of fatness, including behavioral problems, feeding patterns within the family, nature of newborn feeding, attitude of parents and siblings to obesity, and physical activity. Fetal and neonatal growth factors are also associated with the pathogenesis of obesity.

Recent discoveries regarding obesity are related to the hormone leptin and the identification of certain genes in animal models. Leptin precursors and the hormone itself are produced by placenta, fetal tissues, gastric mucosa, and hepatic stellate cells, which can participate in many physiological functions.

Management

Specialized obesity clinics have many advantages, particularly if located at school. Serial measures of height, weight, and skinfold thickness need to be recorded. Whatever the cause of obesity, management must include a reduction in intake of energy. Regular exercise also aids weight loss. Behavioral therapy has been shown to have a role. Anorectic drugs are only a transient aid because the original obesity will reappear after a time. (*See* **Obesity**: Treatment.)

A multidisciplinary support group made up of parents, peers, nutritionist, pediatrician, and a therapist or clinical psychologist has much to offer. Food fads or crash diets are potentially dangerous and must be avoided in childhood, if not in adulthood too. The aim of treatment is to correct faulty eating habits and not to induce ketosis.

Parents of children under 5 need advice to avoid offering a high intake of fat and sugar. Incidences of moderate obesity can be managed by allowing the children to grow into their present weight.

As 450 g of excess body weight represents 14.7 MJ, a daily reduction of 2 MJ is needed to lose 450 g of weight a week. Dietary restriction is not recommended for infants under 18 months of age. Greater affluence will inevitably mean than an ever-increasing number of youths will be obese, with all the associated phenomena. Determined educational and nutritional efforts must be adopted to avoid this scourge of life in privileged communities.

Protein–Energy Malnutrition (PEM)

Formerly known as protein–calorie malnutrition, PEM is a major global problem and one which arouses much controversy even in terms of its classification. Although PEM is rare in adults, it is responsible for a high mortality and indeed morbidity amongst children in developing countries. At one end of the spectrum PEM includes marasmus, and at the other the distinct syndrome of kwashiorkor, with intermediate deficiencies between both polarities. (*See* **Protein**: Deficiency.)

Within this band there are deficiencies of specific dietary constituents, particularly folate and vitamin A. Many children with severe PEM have infestations due to *Giardia lamblia, Strongyloides stercoralis*, or *Ascaris lumbricoides*. Tuberculosis, bronchopneumonia, and measles are significant secondary complications.

Anorexia plays a dominant role in PEM, as does the presence of small-bowel mucosal atrophy. These phenomena may be accompanied by the presence of both a protein-losing enteropathy and iron deficiency. (*See* **Anorexia Nervosa**.)

These disorders can be grouped together. Some workers regard marasmus as a successful adaptation to nutritional stress, and kwashiorkor as the failure of such an attempt. A comparison of kwashiorkor and marasmus is given in Table 1. (*See* **Kwashiorkor**; **Marasmus**.)

Malnutrition and its associated infections cause the death of many millions (Figure 2).

PEM is not limited to nonindustrialized countries. Fad and strict vegetarian diets as well as medically unsupervised elimination regimes all have the potential for causing PEM. (*See* **Vegetarian Diets**.)

Table 1 Comparison of marasmus and kwashiorkor

	Marasmus	*Kwashiorkor*
Weight	↓↓↓	↓↓
Edema	−−−	+++
Depigmentation	−−−	↑↑↑
Hair changes	±	++
Vitamin levels	↓↓↓	↓↓
Small intestine – villous atrophy	++	+++
Lactase, sucrase, maltase	↓↓	↓↓↓
Pancreatic enzymes	↓	↓
Fatty liver	+	+

↓, slightly reduced; ↓↓, moderately reduced; ↓↓↓, markedly reduced.
Modified from Gryboski J and Walker WA (1983) *Gastrointestinal Problems in the Infant*, 2nd edn. Philadelphia: WB Saunders.

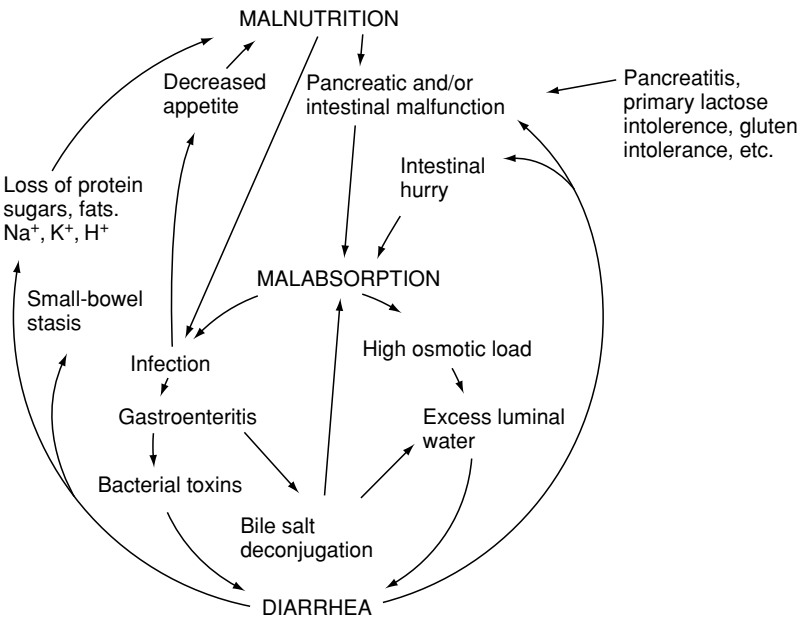

Figure 2 The vicious cycle of diarrhea and malnutrition. Courtesy of WF Balistreri.

Lack of nutritional knowledge and financial resources within a household, the community or society at large is responsible for the appalling childhood mortality and morbidity seen in PEM. Malnutrition ought to be regarded as the consequence of adverse social and economic conditions rather than a primary medical problem. Simple measures, such as the encouragement of breast-feeding, personal hygiene, improved sanitation, and the establishment of adequate and safe drinking water, would do much to prevent the many potentially hazardous enteric infections seen in malnourished infants and children.

Food Fads

Food fads are not only potentially dangerous for adults but also particularly hazardous in childhood. Apart from human milk for babies and young infants, there is no other single food which supplies all nutritional needs, yet food faddism seems increasingly attractive to many seeking 'healthy foods' and a diet free of all alleged contaminants.

As food technology and supermarket retailing become more sophisticated and food processing is seen as an adulteration of our food, some will seek alternatives via the use of 'natural, organic, and health foods' – all loose terms. There is an organic food myth and we do not even have an acceptable definition of 'organic foods' – a term introduced by the late JI Rodale. Consumers presume such food is grown in soil free of artificial fertilizer, is pesticide-free, and does not contain additives, preservatives, hormones, or antibiotics. In northern Europe, 70% or more of all foods now consumed is processed, preserved, and/or packaged. There is no evidence to support the contention that 'organic foods' are healthier for the consumer.

Synthetic vitamins are as effective as natural vitamins. Food grown in rich soil is not nutritionally dissimilar to that found in poor soil. Unfortunately, disinformation by the healthy-food movement has facilitated the public acceptance of food mythology as a reality. Nutrition is a subject that is ill understood. (*See* **Organically Farmed Food**.)

Tragically, children have died from malnutrition when fed weird and unconventional diets. Growth failure and anemia were found in a small group of infants fed a formula of barley water, corn syrup, and whole milk.

Faddism can result in many diverse and potentially dangerous outcomes. There are those who expect their children to survive solely on fruit (fruitarians); others are committed to a diet of raw foods. The discovery of fire 40 000 years ago by primitive humans via our ancestors brought an end to reliance on diets comprised solely of raw food. Some uncooked foods are toxic and unsafe, e.g., meat might become contaminated by bacteria, and raw foods may contain substances that interfere with vitamin B_1 (antihistaminases). Raw red kidney beans possess natural toxins (hemagglutinins). The Zen-macrobiotic diet has been associated with deaths. The ever-increasing restrictive nature of this diet is such that

ultimately only cereals are used, with disastrous consequences. (*See* **Macrobiotic Diets**.)

Embroiled in this widening saga of unconventional diets is the use of vitamin supplements in megadoses. It is now increasingly evident that such therapy is metabolically hazardous, particularly for children.

Food Intolerance and Aversion

There are few subjects that are so enigmatic, confusing, and clouded with anecdotal clutter as that of food adversion (intolerance) and aversion (psychological intolerance or avoidance). It has been suggested that the term 'allergy' which implies an unpleasant and immediate response should be applied only to those with immediate allergic reactions to food and with the presence of specific immunoglobulin E (IgE) antibodies. Mast cells, basophils, eosinophils, macrophages, and platelets can all be activated by IgE antibodies, causing the release of preformed or instant mediators. (*See* **Food Intolerance**, Types; Food Allergies; Lactose Intolerance.)

Food Intolerance

Clinical manifestations of food intolerance can be diverse and may include diarrhea, abdominal pain, failure to thrive, urticaria, eczema, migraine, behavioral disturbance, and food-sensitive colitis. Food intolerance occurs in approximately 5% of the pediatric population and can be caused by the following:

1. Allergic reaction(s). This mechanism is mediated via IgE antibody, a T-cell-mediated reaction, or the systemic formation of antigen–antibody complexes.
2. Pharmacological factors, e.g., tyramine. There is a high concentration of tyramine in fermented cheese and this can elevate the blood pressure and produce symptoms. Caffeine in coffee and tea, in the hypersensitive, can cause symptoms such as anxiety and tachycardia.
3. Enzyme deficiency, e.g., lactose intolerance caused by hypolactasia. It has been suggested that a deficiency of the enzyme phenolsulfotransferase may have a role in food-induced migraine.
4. Toxic effect. In many foods there are naturally occurring toxins, e.g., mycotoxins from storage of foods contaminated with molds.
5. Food fermentation. Bacteria in the colon, particularly *Proteus morganii*, can produce decarboxylase, an enzyme that will convert dietary histidine to histamine. However, as histamine does not cause an immunological reaction, its effect is pseudoallergic.
6. Other intolerances: tartrazine, an azodye, is a common coloring substance contained in many foods, soft drinks, and pharmaceutical products; benzoates or sulfur dioxide preservatives and butylated hydroxyanisole or butylated hydroxytoluene antioxidants may cause urticaria and asthma.

Acetylsalicylic acid (ASA), present in some vegetables and a variety of fruits, can cause a range of symptoms in sensitive patients, including asthma, urticaria, and rhinitis; synthetic salicylates are found as flavoring products.

Food Aversion

Food aversion is the result of either psychological intolerance or food avoidance.

Pathogenesis

The permeability of the gastrointestinal tract mucosa to food allergens is believed to be a major factor in the pathogenesis of antigenic food reactions. It is known that intestinal permeability is increased following an acute gastroenteritis. Temporary changes in the gut have been demonstrated after rotavirus enteritis by studying the differential permeation of the sugars, lactose, lactulose, and rhamnose. Studies using the inert probe, polyethylene glycol, have shown increased intestinal permeability in allergic children. Moreover, using a radioimmunological method for measuring serum concentration of human α-lactalbumin, preterm infants have an increased absorption of macromolecules compared with that seen at term.

The integrity of the bowel mucosa is maintained in part by secretory IgA and the glycocalyx (a glycoprotein coating the bowel enterocytes) as well as the mucosa-associated lymph tissues (MALTs). It has been suggested that following food challenge there is the deposition of immune complexes in different target organs and sites which can result in migraine, arthalgia, or eczema.

Treatment

Treatment is often unsatisfactory because of difficulties in identifying and then removing the offending allergen(s).

Food Additives, Salicylates, and Hyperactivity

Pediatricians and nutritionists seem to polarize themselves into either protagonists or vehement opponents of this theory which postulates that hyperkinesis can

be managed in 30–50% of children by eliminating artificial food-coloring matter, preservatives, and other additives from the diet, and by eliminating fruits and vegetables containing natural salicylates. The major symptoms of this psychological disorder are overactivity, distractibility, restlessness, impulsive behavior, and a short attention span, identifying this syndrome in part as a conduct disorder. It is probably not a single entity and the etiology is multifactorial. Unfortunately, it is now accepted that as many as 50% of all children with this disorder show some of its features in adulthood.

See also: **Anorexia Nervosa**; **Dahi**; **Food Intolerance**: Types; Food Allergies; Lactose Intolerance; **Kwashiorkor**; **Macrobiotic Diets**; **Marasmus**; **Obesity**: Etiology and Diagnosis; Fat Distribution; Treatment; **Organically Farmed Food**; **Protein**: Deficiency; **Vegetarian Diets**

Further Reading

Bentley D and Lawson M (1988) *Clinical Nutrition in Paediatric Disorders*. London: Baillière Tindall.

Bentley D, Lifschitz C and Lawson M (2002) *Pediatric Gastroenterology and Clinical Nutrition*. London: Remedica.

Callaghan AL, Booth IW, Moy RGD, Shaw NJ and Debelle G (2002) *Archives of Disease in Childhood* 86 (suppl. 1): A35–A36.

Ellis KJ and Nicholson M (1997) Leptin levels and body fatness in children: effects of gender, ethnicity and sexual development. *Pediatric Research* 42(4): 484–488.

Garrow JS, James WPT and Ralph A (2000) *Human Nutrition and Dietetics*. Edinburgh: Churchill Livingstone.

Mann J and Truswell S (2000) *Essentials of Human Nutrition*. Oxford: Oxford University Press.

Stockman JA (2002) *Yearbook of Pediatrics*. St Louis: Mosby.

Walker AF (ed.) 1990 *Applied Human Nutrition for Food Sciences and Home Economists*. Chichester: Horwood (Ellis).

Whitaker RC, Wright JA, Pepe MS, Seidel KD and Dietz WH (1997) Predicting obesity in young adulthood from childhood and parental obesity. *New England Journal of Medicine* 25 337(13): 869–873.

World Health Organization (1998) *Obesity: Preventing and Managing the Global Epidemic*. Geneva: World Health Organization.

CHILL FOODS

Effect of Modified-atmosphere Packaging on Food Quality

A G Smith, The University of Huddersfield, Huddersfield, UK
B P F Day, Campden & Chorleywood Food Research Association, Gloucestershire, UK

Introduction

One form of food preservation technology that has grown in importance in recent years, often in combination with chilled foods, is modified-atmosphere packaging (MAP). The market for products using MAP has grown rapidly because of the benefits that this technique offers, particularly in respect to maintaining the quality of food throughout the extended shelf-life that the application of the method achieves. Estimates of sales indicate that around 1.6 billion modified-atmosphere (MA) packs of food were sold in 1990; this had risen to 2.6 billion by 1997 and is expected to exceed 3.3 billion in 2002.

MAP can be described as a food preservation technique in which the composition of the atmosphere surrounding the food, inside its hermetically sealed packaging, is different from the normal composition of air. Within the food packages air is replaced by a blanket of gas, a mixture of gases, or even no gas at all in the case of vacuum-packing, which is a form of MAP. Under MAP the growth of bacteria, molds, and yeasts that cause food to deteriorate is inhibited and other forms of physiological decay are slowed down. When applied effectively, MAP extends the shelf-life of food – the period of time for which it might be considered to be of saleable quality.

Consumers increasingly demand the natural qualities of fresh foods in raw, part-prepared, and even ready-to-eat products, whilst also preferring them to be free from additives. Retailers wish to provide the widest possible range of products for the consumer to choose from but need to be assured that items will have a long enough life on the display shelf to make them viable to stock. MAP contributes towards the fulfillment of these diverse demands by extending the shelf-life of food items whilst maintaining the product's original quality.

Unlike controlled-atmosphere bulk storage, where precise control of the gaseous components in a food warehouse atmosphere is possible, in MAP there is no way of varying the concentration of these components once the packaging has been sealed. Highly perishable foods will only achieve their full potential in terms of extended life when packed in MA if they are handled hygienically throughout any processing and, in most cases, also held chilled.

Chilled products already sold in MAP include cheeses, cooked and cured meats, cooked and dressed vegetable products, crustaceans, dairy and bakery goods, fish, kebabs, pasta, pizza, red (uncooked) meats, ready meals, whole and prepared fresh fruit and vegetables (Figure 1), and this range continues to expand rapidly.

MAP cannot in itself enhance the quality of food, be it raw or prepared in some way. The technology can only be used to extend the reasonable life of a food product because it can preserve food at or very near to its original quality due to its ability to inhibit the various forms of deterioration that naturally occur when products are stored in air. Any positive impact that MAP has on maintaining food quality will be limited by the need for foods:

1. to be of prime quality in all respects in the first instance
2. to be handled in an appropriate manner that minimizes risk from any form of contamination or deterioration prior to and throughout the packing process
3. to be held at appropriate temperatures after packing and prior to sale (ideally between 0 and 3 °C for chilled foods).

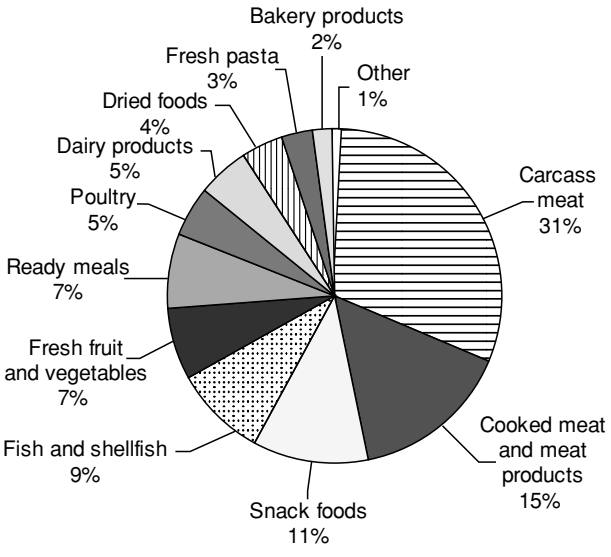

Figure 1 Relative importance of major food sectors to the modified-atmosphere packaging retail market (UK) 1997. Data from MSI Data Report (1998).

With these issues in mind, this chapter aims to outline the effects of MAP on the quality of food purveyed to consumers. The typical gases used in MAP and their applications are briefly considered and the impact of this technique on the quality of the major categories of chilled foods is discussed.

The Gases Used in MAP

It is necessary for manufacturers to utilize food-grade gases for the safe and hygienic MAP of food products. These high-purity gases are produced in hygienic conditions applicable to food production systems and are stored in dedicated containers made and maintained to the standards required in any food-handling environment.

The gases most commonly used for food MAP are those found naturally in air: carbon dioxide (CO_2), nitrogen (N_2), and oxygen (O_2). The effectiveness and safety of other gases that might be used for MAP are being examined and gases such as argon, helium, nitrous oxide, and sulfur dioxide are already permitted for food use in the European Union (EU). Other gases, such as carbon monoxide, hydrogen, and chlorine, are also being investigated, although their commercial application may be limited due to negative effects on food quality and regulations relating to use.

Air consists of N_2 (78.09%), O_2 (20.95%), argon (0.93%), and CO_2 (0.03%), and in the majority of MAP applications, N_2, O_2, and CO_2 are currently the only gases used. The ratio of each gas employed in MAP mixtures varies significantly from air according to the food commodity being packed. It is the careful maintenance and monitoring of the proportions of each gas and the utilization of correct mixtures of gases for specific products that are important to insure that the benefits of MAP are fully exploited.

Carbon Dioxide

Whilst CO_2 provides particular benefits in concentrations that can enable the shelf-life of food to be extended considerably, it can also have some disadvantageous effects on food quality. CO_2 effectively inhibits most bacteria (aerobic strains) and molds that need O_2 to survive. Generally, the greater the concentration of CO_2 around packaged foods, the longer its shelf-life can be extended.

CO_2 is easily absorbed into fats or water which are key constituents of many perishable foods. The mild carbonic acid that forms from the absorption of CO_2

into the food's natural moisture has further preservative effects. However, this factor can lead to 'pack collapse,' as the package's gas volume reduces due to such CO_2 absorption, and, although foods might still be edible, it can also result in other undesirable changes in the sensory properties of some foods.

The determination of the proportion of CO_2 used in MAP gas mixes will relate to achieving an acceptable balance between extending shelf-life and avoiding the potentially undesirable effects. However, the concentration of CO_2 should not normally be below 20% if it is required to inhibit mold growth effectively (Table 1).

Nitrogen

This inert gas is useful for excluding or reducing the relative proportion of O_2 in the atmosphere around food. It is also used to counter the effects of pack collapse caused by the absorption of CO_2 by fatty or moist food items, which is particularly likely to occur under chilled storage.

Whilst N_2 does not support the growth of aerobic microorganisms, it is neither absorbed by foods nor does it have the mildly acidifying effect of CO_2. This means that on its own it is not as effective as CO_2 for inhibiting microorganisms. However, N_2 has little impact on the sensory qualities of food which makes it particularly useful for gas-packing foods that might be adversely affected in this way by CO_2, such as creams and other dairy products (Table 1). N_2 is also used in the head space of beverage containers, to flush out O_2, and in packaging dry food which does not require the bacteria and mold-inhibiting properties of CO_2 yet needs to be protected from oxidative deterioration.

Oxygen

It is the O_2 in the air that causes most of the spoilage problems that MAP seeks to reduce. Not only is oxygen needed for survival by many microorganisms, it can also cause oxidative deterioration in foods. Hence, MAP largely originated from the objective of excluding O_2 from around foods to obviate potential quality problems. However, there are applications when a controlled amount of oxygen in MA gas mixes is necessary when packing certain food commodities.

For example, O_2 can maintain acceptable color quality in raw meats and is needed to avoid anaerobic respiration in fruits and vegetables. O_2 is also used in packing white fish and seafood products because it inhibits *Clostridium botulinum*, a highly toxic anaerobic pathogen. These benefits mean that carefully

Table 1 Recommended gas mixtures for retail modified-atmosphere packaging of a selection of food commodities

Food item	*CO_2 (%)*[a]	*O_2 (%)*[a]	*N_2 (%)*[a]
Most raw red meats (beef, lamb, pork)	30	70	
Venison/wild boar	20	80	
Raw offal	20	80	
Raw poultry/game[b]	30		70
Raw fish and seafood crustaceans/molluscs	40	30	30
Raw oily fish	40		60
Cooked, cured, and processed meats	30		70
Cooked, cured, and processed fish/seafood	30		70
Cooked, cured, and processed poultry and game products	30		70
Ready meals and cook-chill products	30		70
Fresh pasta products	50		50
Bakery products	50		50
Dairy products – hard cheeses (not mold-ripened varieties)[c]	100		
Dairy products – grated and soft cheeses (not mold-ripened varieties)[c]	30		70
Dairy products – other products, including butter, cream, custard, margarine, yogurts[d]			100
Dried foods			100
Combination products (pies, pastries, pizzas, battered products)	30		70
Cooked/sauced/stuffed vegetable products	30		70
Most liquid food and beverages (beers, cider, cordials, yogurt drinks, oils, wines)			100
Carbonated drinks	100		
Fresh whole and prepared fruit and vegetables	5	5	90

[a]Gas mixtures relate to retail packaging only. Bulk and primal storage require different optimum gas mixes.
[b]Dark poultry meat such as skinned and diced turkey leg meat requires a gas mixture similar to red meats to retain acceptable color quality.
[c]CO_2 and N_2 inhibit mold growth, therefore mold-ripened cheeses are not normally modified-atmosphere packaged.
[d]Aerosol creams excepted, which are packaged under nitrous oxide (N_2O).
Adapted from Air Products (1995) *The Freshline Guide to Modified Atmosphere Packaging (MAP)*. Basingstoke, Hants: Air Products, with permission.

monitored amounts of O_2 may be used in MAP for certain major food commodities.

Recent research has indicated that a novel form of MAP utilizing high levels of O_2 (over 70%) has the potential to maintain both quality and food safety, particularly for respiring fruit and vegetable items and where products involve a combination of respiring and nonrespiring chilled foods, such as sandwiches, pizzas, and ready meals.

Recommended gas mixtures to be used for maximizing shelf-life whilst maintaining sensory quality of a range of major food commodities being packed for retail sale are summarized in Table 1.

Other Gases

As noted, numerous other gases are being examined for use in MAP. Some are already permitted for specified food use in the EU such as argon (E938), helium (E939), nitrous oxide (E942), and sulfur dioxide (E220). Of these, argon (Ar) appears to inhibit enzyme activity and tissue fermentation more effectively than N_2. Ar may also be better at displacing O_2 from the cells of food than N_2, so might further reduce the occurrence of oxidative deterioration. Nitrous oxide (N_2O) and Ar are also claimed to enhance the effect of antimicrobial agents used in food products. Research is continuing into the development of novel forms of MAP using such alternative gases.

Major Chilled Food Categories and the Effect of MAP

Red Meat

Today's consumers demand bright red meat. A generation or more ago, butchers' meat was expected to have 'hung' well, and, if anything, exhibit a deep red color, which was seen as an indicator of tenderness and flavor, but most meat purchasers today would probably infer that such meat was 'past its best.' Retailers are largely responsible for promoting and displaying meat of a fresh red hue and this is now perceived as a key quality indicator. MAP has a major contribution to make in respect to extending the life of meat in a desired red color. This is because gas mixtures containing high levels of O_2 react with the natural postmortem blood chemistry to maintain a 'consumer-attractive' appearance for longer. Typically, raw meat will have a shelf-life of up to 4 days, but packaged in a modified gas mix, this can be doubled.

MAP producers acknowledge that gas mixes that are high in O_2 contribute to the maintenance of color quality preferred by today's consumers. In fact, if meat is packed in a CO_2-rich atmosphere, excluding O_2 which many microorganisms need to thrive, shelf-life could be extended for considerably longer. However, the consequent innocuous darkening of the meat color that would occur in the presence of CO_2 is regarded as an undesirable quality characteristic and hence this is not done.

Raw meats require cooking before they are edible. If the cooking process is undertaken effectively, there should be minimal risk of food poisoning. It is therefore considered that, whilst O_2-rich atmospheres might support aerobic microflora, the risks are acceptably small if meat is hygienically handled prior to packing and held chilled prior to sale. Thus, MAP protects the meat from contamination and enables the maintenance of visually appealing fresh color during the period of extended shelf-life, a quality characteristic which is highly valued by both the retailer and the consumer.

Cooked and Cured Meats

Cooked meats are considered a 'high-risk' food since they do not undergo further processing prior to consumption. It is therefore imperative that producers of such products employ hazard minimization techniques during handling, preparation, cooking, and packaging of such foods. If this is done then MAP has an important role to play in the safe extension of product shelf-life whilst simultaneously insuring that the product exhibits all the expected qualities of a freshly cooked item.

As can be seen in Table 1, the gas mixture advised for MAP of cooked meats and poultry excludes O_2. The cooking process will kill any vegetative pathogens and also fix color, unlike raw meat. Without the need for O_2 for the maintenance of visual quality, a mixture of N_2 and CO_2 is used for MAP, affording the advantage of increased inhibition of microbial deterioration. The maximal extension of shelf-life is dependent on appropriate chilled storage, ideally between 0 and 3 °C.

Some lightly cured meats, such as bacon, may require cooking before consumption but because they are usually treated with salts such as nitrites, their pink coloring is assured, hence the same CO_2/N_2 gas mixture as for cooked meats is recommended. Keeping quality of these products may be adversely affected if lower salt formulations are developed however. It is also imperative that appropriate storage temperatures are rigorously maintained below 8 °C and preferably between 0 and 3 °C.

Other cured products, such as salami, do not need further cooking and the curing process is a traditional preserving method that may even be effective enough

to insure considerable shelf-life, even if stored in ambient conditions. However, fattier meat products may be subject to oxidative rancidity, so packing in an MA without O_2 can be useful to inhibit this process.

Although vacuum-packing has been successfully employed in the production of retail cooked and cured meat products with enhanced shelf-life, gas-filled MAP provides a better-looking product that is easier to handle after removal of the packaging. However, vacuum skin packaging (VSP), where a lightly applied, highly ductile plastic barrier laminate forms itself around the food during the vacuum process, without squashing it, has been seen to be a particularly useful innovation by retailers for some sliced cooked meat products. VSP not only achieves the benefit of easy separation of meat slices but also enables packs to be stacked vertically for more attractive and efficient merchandising.

Fish

MAP is particularly useful for extending the shelf-life of raw fish which normally deteriorates rapidly after exposure to air. The controlled proportions of gases used in MAP can contribute towards reducing the formation of peroxides and other chemicals, which create characteristic off-odors and taints in fish.

Fish are particularly susceptible to oxidative rancidity; however, MAP gas mixtures cannot contain too high a concentration of CO_2 alone as absorption into the moist tissues will cause packs to collapse. CO_2 absorption also exacerbates drip and affects color and flavor adversely, especially in seafood. O_2 must also be incorporated into the gas mixture used in MAP of white fish, shellfish, and crustaceans to insure sensory quality and also to inhibit the growth of toxin-forming anaerobes, such as *Clostridium botulinum*. Oily fish, however, such as mackerel, herring, tuna, and salmon, need an MAP gas mixture consisting of CO_2 (40%) and N_2 (60%) alone. This is because oxidative deterioration is considered the major potential quality problem.

From catch through to MAP, fish must be handled hygienically and rigorously stored in temperature-controlled conditions. The fish species and season, affecting such factors as fat content, will also influence the keeping quality of the product. The extension of shelf-life and the maintenance of acceptable food quality using MAP can only be fully exploited if the product being packaged is initially of prime quality and has been handled appropriately at all times prior to packaging. Overall shelf-life and acceptable quality can be prolonged by MAP for up to 6 days – around twice as long as fish would keep held chilled in air.

Fruit and Vegetables

Fresh fruit and vegetables continue to respire and ripen after harvesting. MAP has an important role in extending the life of these commodities in that it can be used to reduce the rate of respiration and delay ripening. By slowing textural softening, chlorophyl degradation, and other ripening processes and by reducing the impact of chilling injury and other physiological problems, product quality can be maintained for up to five times longer than if stored in air. This means that shelf-life can be increased from 2 to 5 days for highly perishable items, and increased from up to 7 to around 35 days for hardier fruit and vegetables. The potential increase in shelf-life that can be attained through the use of MAP is inversely proportional to the rate of respiration naturally occurring.

There are certain technical issues that must be considered when packaging fruit and vegetables in MAs. The natural respiration process tends to deplete O_2 and cause CO_2 levels to increase. A minimum level of O_2 is required in the gas mixture to avoid anaerobic respiration occurring, as this would result in the development of undesirable sensory qualities in the food. Without any O_2 present there would also be an increased risk of food poisoning by pathogenic anaerobes such as *Clostridium* species, the hardy spores of which, should they be present in any soil remnants, might germinate if storage temperatures are not well controlled at any time.

The packaging itself must be carefully selected to insure that the correct level of gas permeability can occur. In order to achieve optimal conditions, the MA within the packaging must be maintained such that the gas mix contains between 2 and 10% of both O_2 and CO_2. With commodities such as beansprouts, broccoli, leeks, mushrooms, and peas, which exhibit a high rate of respiration, traditional MAP films are not permeable enough to maintain the gas mixture. With these foodstuffs, a microperforated film is more appropriate for their level of respiration, but has the consequent disadvantage of potentially allowing ingress of microorganisms and other forms of organoleptic deterioration to occur.

Recent research has shown that freshly prepared vegetables packed in high O_2 atmospheres (> 70%) may offer a solution to the disadvantages of gas mixtures currently in use. High O_2 MAP has been shown to prevent the formation of anaerobic fermentation byproducts and inhibit both enzymic deterioration and the growth of microorganisms. Research is currently continuing in this area.

Dairy Products

Ideal gas mixtures for the range of dairy products vary according to the likely types of deterioration to which specific items are most susceptible (Table 1).

As hard-cheese quality is most likely to be affected by mold growth, an atmosphere consisting solely of CO_2 is most appropriate for MAP. Acceptable eating quality can be maintained and shelf-life trebled from around 4 weeks to up to 3 months by using MAP.

Softer cheeses are more likely to deteriorate by separation and can be affected by yeasts, bacteria, and oxidative rancidity. A mixture of CO_2 and N_2 is recommended to counteract the likely pack-deforming absorption of CO_2 by these moister products. For the most perishable products, MAP can help to maintain quality whilst doubling normal shelf-life to around 2 weeks and for many items this could be considerably longer.

It is because of the mold-inhibiting properties of CO_2 and N_2 that cheeses that incorporate natural molds are not normally packed using MA technology.

Creams and similar dairy products are packed solely in N_2 because flavor quality would be adversely affected by CO_2 absorption. N_2 would also be recommended for packaging butter and yogurts should MAP be applied to such products.

Ready Meals

The range of chilled products available from the major retailers has increased greatly in recent years. Major retailers have developed their own brands, often in collaboration with large manufacturers who have the technical capability to produce cook-chill products safely and hygienically.

Within the UK, current regulations stipulate that the maximum shelf-life for cook-chill products utilized in commercial operations is limited to a maximum of 5 days. This does not apply to retail foods for the domestic market, which means that, by using MAP, or by cooking under vacuum, as in the sous-vide process, shelf-lives of ready meals and the like can be safely extended considerably further.

Whilst proper storage is essential, ideally below 3 °C, food safety and eating quality can be maintained for up to 10 days without any appreciable risk when using MAP technology or sous-vide methods to produce such foods.

Temperature control is imperative throughout the production process, be it to insure proper cooking, or to insure the minimization of risk from bacterial growth after chilling. This is particularly important for preventing the germination of hardy spores of

Table 2 Examples of active packaging systems and their applications

Active packaging system	*Active element*	*Actual/potential food applications*
O_2 scavengers	1. Iron-based 2. Metal/acid 3. Metal catalyst (e.g., platinum) 4. Ascorbate/metallic salts 5. Enzyme-based	Bread, cakes, cooked rice, biscuits, pizza, pasta, cheese, cured meats and fish, coffee, snack foods, dried foods and beverages
CO_2 scavengers/emitters	1. Iron oxide/calcium hydroxide 2. Ferrous carbonate/metal halide 3. Calcium oxide/activated charcoal 4. Ascorbate/sodium bicarbonate	Coffee, fresh meats, and fish, nuts and other snack foods products and sponge cakes
Ethylene scavengers	1. Potassium permanganate 2. Activated carbon 3. Activated clays/zeolites	Fruit, vegetables and other horticultural products
Preservative releasers	1. Organic salts 2. Silver zeolite 3. Spice and herb extracts 4. BHA/BHT antioxidants 5. Vitamin E antioxidants	Cereal, meats, fish, bread, cheese, snack foods, fruit and vegetables
Ethanol emitters	1. Encapsulated ethanol	Pizza crusts, cakes, bread, biscuits, fish and bakery products
Moisture absorbers	1. PVA blanket 2. Activated clays and minerals 3. Silica gel	Fish, meats, poultry, snack foods, cereals, dried foods, sandwiches, fruit and vegetables
Flavour/odor absorbers	1. Cellulose triacetate 2. Acetylated paper 3. Citric acid 4. Ferrous salt/ascorbate 5. Activated carbon/clays/zeolites	Fruit juices, fried snack foods, fish, cereals, poultry, dairy products and fruit

BHA, butylated hydroxyanisole; BHT, butylated hydroxytoluene.

Bacillus or *Clostridium* species which might remain after cooking.

The quality of cook-chill meals has improved dramatically in recent years, but MAP can only maintain the quality of the product as originally produced. This means that retailers must insure that product formulation incorporates efficient organoleptic assessments as well as risk control methods such as Hazard Analysis Critical Control Points (HACCP) if they wish to guarantee that consumers receive products of acceptable quality in all respects.

Conclusion – The Future of MAP

Innovations in MAP and related packaging technologies are continuously being made. In recent years these innovations have improved the utility and function of such packaging. This has resulted in MAP providing additional benefits that have a positive impact on the quality of food as it reaches the end consumer, particularly through retailers.

Innovations in film materials have allowed the development of vacuum MAP methods such that VSP (vacuum skin packaging) can be formed around food without squashing it, unlike earlier vacuum pack technology.

Active packaging refers to systems and devices which incorporate certain additives, such as oxygen scavengers, carbon dioxide emitters, ethylene scavengers, moisture absorbers, and other 'active' elements (Table 2).

The overall aim of these developments is to act as 'freshness enhancers' and hence to achieve even longer extensions in the shelf-life of foods, whilst maintaining sensory and nutritional quality and microbial safety. Advances in biotechnology, the science of materials and packaging methods are continuously being made and hence will influence the future development of MAP.

Future innovations in packaging in general may be driven by environmental concerns such as the implications arising from the increasing problem of disposing of packaging waste. Consumers' desire for easier-to-use packs, along with other added-value features, may also influence future developments in food packaging. However, the potential to prolong the shelf-life of food whilst retaining desirable fresh qualities is likely to remain the key benefit and purpose of employing MAP.

See also: **Chilled Storage**: Principles; Attainment of Chilled Conditions; Quality and Economic Considerations; Microbiological Considerations; Use of Modified-atmosphere Packaging; Packaging Under Vacuum; **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; Effects on Fruit and Vegetables; **Sensory Evaluation**: Sensory Characteristics of Human Foods; Food Acceptability and Sensory Evaluation; **Spoilage**: Bacterial Spoilage; Fungi in Food – An Overview; Molds in Spoilage; Yeasts in Spoilage

Further Reading

Air Products (1995) *The Freshline Guide to Modified Atmosphere Packaging (MAP)*. Basingstoke, Hants: Air Products.

Day BPF (1997) Modified-atmosphere packaging markets, Europe. In: Brody AL and Marsh KS (eds) *The Wiley Encyclopedia of Packaging Technology*, 2nd edn, pp. 656–659. John Wiley.

Day BPF (1999) Principles – chilled storage of foods. In: Robinson R, Batt C and Patel P (eds) *Encyclopedia of Food Microbiology*, pp. 403–410. London: Academic Press.

Day BPF (1999) Applying MAP to chilled ready meals. In: *Savoury Foods Industry Guide – Europe*, pp. 32–34. Uxbridge, Middlesex: Tusset RAI.

Day BPF (2000) Chilled food packaging. In: Stringer M and Dennis C (eds) *Chilled Foods – A Comprehensive Guide*, 2nd edn, pp. 135–150. Cambridge, UK: Woodhead.

Farber JM and Dodds KL (eds) (1995) *Principles of Modified-Atmosphere Packaging and Sous Vide Product Packaging*. Lancaster, Pennsylvania: Technomic.

O'Connor-Shaw RE and Reyes VG (1999) Use of modified-atmosphere packaging – chilled storage of foods. In: Robinson R, Batt C and Patel P (eds) *Encyclopedia of Food Microbiology*, p. 410. London, UK: Academic Press.

CHILLED STORAGE

Contents

Principles

S James, University of Bristol, Langford, North Somerset, UK

Basic Theory

The complete chill chain for a food will contain many of the following unit operations: prechilling preparation, primary or secondary chilling, storage, transportation, and display. Storage, transport, and display are temperature maintenance operations. Chilling results in a substantial decrease in the mean temperature of the product, whereas during the preparatory treatment, there can be a range of temperature responses from a large gain to a small decrease in the temperature of the foodstuff.

The initial and final temperatures, and the maximum and minimum desirable rates of temperature reduction during chilling, depend on the food being processed. In all cases, the final product temperature must be above the initial freezing point of the food, but in a number of products, biological factors dictate higher storage temperatures.

The principle factors that control the rate of heat and mass transfer during cooling are independent of the foodstuff. Heat can be lost from the surface of a body by four basic mechanisms: radiation, conduction, convection or evaporative cooling. (*See* **Heat Transfer Methods**.)

To determine the rate of heat transfer (Q_r) by radiation from food, the following approximation may be applied:

$$Q_r = \varepsilon A \sigma (T_s^4 - T_e^4), \qquad (1)$$

where ε is the emissivity, A is the surface area of the food, σ the Stefan–Boltzman constant, and T_s and T_e the temperature of the surface and the enclosure, respectively. To achieve substantial rates of heat loss by radiation, large temperature differences between the product and the enclosure are required. Such differences are not normally present during food cooling operations, except in the initial chilling of bakery products. However, researchers in Russia have used banks of refrigerated plates to enhance radiant heat transfer during the chilling of beef carcasses.

The rate of one-dimensional heat transfer (Q_{cd}) by conduction is given by:

$$Q_{cd} = k A \frac{\partial T}{\partial x}, \qquad (2)$$

where k is the thermal conductivity of the medium through which the heat is passing, and $\partial T/\partial x$ is the temperature gradient.

Physical contact between the product and a cooler surface is required to extract heat by conduction. The irregular shape of most foodstuffs precludes this mechanism in many applications. Plate coolers are routinely used to freeze fish and offal, which have soft pliable surfaces. Belt freezers are used for steaks, fish fillets, and other thin products. However, the rate at which heat can be conducted away from the surface is not the sole criterion that governs the time taken to cool a product. Heat must also be conducted from within the product to its surface before it can be removed. Most foodstuffs are poor conductors of heat, and this imposes a severe limitation on attainable chilling times for either large individual items or small items cooled in bulk.

The rate of mass transfer (M) from the surface of an unwrapped food is described by the equation::

$$M = mA(P_s a_w - P_m), \qquad (3)$$

where m is the mass transfer coefficient, A the area, P_s the saturated vapor pressure at the surface, a_w the water activity, and P_m the vapor pressure above the food surface. Q_e, the heat loss due to evaporation, can be obtained by multiplying M by the appropriate latent heat. The heat lost in the evaporation of water from the surface of the product is a minor component of the total heat loss for most foodstuffs. In

spray-cooling systems, the rate of heat removal is enhanced by evaporation from the surface. However, the main advantage of spray cooling is a reduction in weight loss. In a number of specific cases, i.e., vacuum cooling, evaporation is the primary cooling agent.

Most food-chilling systems rely on convection as the principal means of heat removal. The rate of one-dimensional heat transfer (Q_c) by convection is given by::

$$Q_c = h_c A(T_s - T_a), \quad (4)$$

where h_c is the convection or film heat transfer coefficient, and T_a is the temperature of the cooling medium. The most common media are air or water, although sugar solutions, salt brines, and other refrigerants have been used. Each combination of product and cooling system can be characterized by a specific surface heat-transfer coefficient. The value of the coefficient depends on the shape and surface roughness of the foodstuff and, to a much greater degree, on the thermophysical properties and velocity of the medium. Values range from $5\,W\,m^{-2}\,K^{-1}$ for still air to over $500\,W\,m^{-2}\,K^{-1}$ for agitated water.

Chilling Methods for Solid Foodstuffs

Moving Air

This is the most widely used method as it is economical, hygienic, and relatively noncorrosive to equipment. Systems range from the most basic, in which a fan blows the cooled air around an insulated room, to products on conveyors passing through blast chilling tunnels or spirals. Low rates of heat transfer are attained in air-cooling systems, but this is not important if conduction within the product is the rate-controlling factor. A major disadvantage is excessive dehydration from the surface of any unwrapped product, whilst the need to avoid surface freezing limits the lowest air temperature that can be used. In practice, air distribution is a major, often overlooked, problem.

A batch system where warm food is placed in refrigerated rooms is the most common method of chilling. Individual items such as carcasses, tuna, or bunches of bananas are hung from rails, smaller products are placed on racks or pallets, and bulk fruits and vegetables are placed in large bins.

Conveying products overcomes problems of uneven air distribution since each item is subjected to the same velocity/time profile. In the simplest continuous systems, the food is suspended from an overhead conveyer and moved through a refrigerated room. This process is often used in air chilling of poultry or in prechilling of pork carcasses. Some small cooked products are continuously chilled on racks of trays, which are pulled or pushed through a chilling tunnel using a simple mechanical system. In more sophisticated plants, the racks are conveyed through a chilling tunnel in which the refrigeration capacity and air conditions can be varied throughout the length of the tunnel.

In larger operations, it is more satisfactory to convey the cooked products through a linear tunnel or spiral chiller. Linear tunnels are far simpler constructions than spirals but require more space.

Hydrocooling/Immersion Cooling

Hydrocooling is probably the least expensive method of achieving rapid cooling in small products. The product is immersed in, or sprayed with, cool water, either at ambient or near 0 °C. Practical systems vary from simple unstirred tanks to plants where the product is conveyed through agitated tanks or under banks of sprays. Such systems are typically used for celery, asparagus, peas, sweet corn, carrots, peaches, etc.

Immersion chilling also has application with larger products. Most poultry to be frozen is initially cooled by immersion in chilled water or an ice/water mixture. The procedure is very fast, and the birds gain weight during the process. The maximum weight gain is controlled by legislation.

An alternative system to air or immersion is spray chilling. Practical spray-chilling systems use a combination of air and sprays for the initial part of the chilling period and then air only for the rest of the chilling cycle. The sprays of cold water at 2–3 °C are applied not continuously but in short bursts. The main advantage claimed for these systems is reduced weight loss.

Plate Cooling

With thinner materials, a plate-cooling system has the potential to halve the cooling time required in an air-blast system. Continuous horizontal and rotary plate freezing systems have been produced commercially and could be modified to operate at higher temperatures. They tend to be expensive, especially if automatic loading and unloading are required, but have low running costs. Designs have been produced for continuous belt coolers similar to those used for belt freezing.

Ice or Ice/Water Chilling

Chilling with crushed ice or an ice/water mixture is simple, effective, and commonly used for fish cooling. The individual fish are packed in boxes between layers of crushed ice, which extract heat from the fish and consequently melt. The temperature of the

'coolant' remains at a constant 0 °C until all the ice has turned into water. The process is labor-intensive, although automatic filling systems have been developed. Cabbage and root crops are also cooled with crushed ice.

Vacuum Cooling

Solid products with a large surface area-to-volume ratio and an ability to release internal water are amenable to vacuum cooling. The products are placed in a vacuum chamber, and the resultant evaporative cooling removes heat from the food. In general terms, a 5 °C reduction in product temperature is achieved for every 1% of water that is evaporated. Prewetting is commonly applied to facilitate cooling without any loss of weight.

Vacuum cooling is rapid and economical to operate because of low labor costs, but the capital cost of the large vacuum vessels is high. Large amounts of lettuce, celery, cauliflower, green peas, and sweetcorn are vacuum-cooled.

Cryogenic Cooling

Avoiding surface freezing of the product is the main problem in using liquid nitrogen or solid carbon dioxide for chilling. Continuous chilling systems using liquid nitrogen either immerse the product in the liquid, spray the nitrogen on to the surface, or vaporize the nitrogen in a forced draught and pass it over the surface of the foodstuff.

Direct spraying of liquid nitrogen on to a food product, whilst it is conveyed through an insulated tunnel, is the most common method. Surface freezing is still a problem, but an extra refrigeration effect is obtained by precooling the food with the cold gas produced by the vaporization.

Chilling Methods for Liquid Foodstuffs

The majority of liquid foodstuffs require cooling after a heat-processing operation such as cooking, pasteurization, or sterilization. Milk is cooled at the point of collection to maintain its quality, unpasteurized fruit juices are cooled immediately after production, whilst fermented beverages are often cooled during primary and secondary fermentation, and before storage. (*See* **Fermented Milks**: Products from Northern Europe; **Milk**: Processing of Liquid Milk; **Pasteurization**: Principles.)

In simple or small-scale processes, containers of hot liquids are allowed to cool in ambient air or placed inside chill rooms. Other cooling systems for liquid foods rely on direct expansion refrigeration, the use of a secondary refrigerant that is passed through or around the foodstuff, vacuum cooling, or a combination of liquid and vacuum.

Batch coolers for liquid foods range in capacity from 100 to 10 000 l with the foodstuff contained in a stainless steel vessel. The cooling medium may circulate through the jacket of the vessel, through a coil immersed in the liquid, or both. Most vessels are provided with agitators to improve convective heat transfer and stop temperature stratification. One common method used to decrease cooling times in a closed vessel is to apply a vacuum to produce evaporative cooling.

Continuous cooling used to be achieved in falling film or surface coolers in which the hot liquid was pumped over the top of a horizontal bank of refrigerated coils and flowed down over the cooled surfaces. These systems have now been replaced by totally enclosed coolers. Double-pipe coolers have also been employed in specialized applications but have a limited heat-transfer surface. Multiplate coolers are extensively used for liquid foods. They have the highest available heat-transfer surface, lowest material requirements, and maximum efficiency, and are very flexible in operation and easy to clean. In certain applications, such as beer and wine cooling, multitube coolers that have a much higher resistance to pressure and can use primary refrigerants have advantages over multiplate coolers.

Scraped surface heat exchangers can have advantages in the cooling of very viscous liquid foods and where surface fouling is a potential problem. Currently, there is interest in the use of pumpable ice in the form of an ice 'pig' to clean heat exchangers and pipeline systems for liquid foods.

Chilled Storage

Theoretically, there are clear differences between the environmental conditions required for chilling, which is a heat removal/temperature reduction process, and storage where the aim is to maintain temperature. However, in many air-based systems, chilling and storage take place in the same chamber, and even where two separate facilities are used, in many cases, not all the required heat is removed in the chilling phase. With some fruits and vegetables, the rate of respiration during storage is sufficient to require heat removal if product quality is to be maintained. (*See* **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability.)

Bulk Storage Rooms

Most unwrapped meat and poultry and all types of wrapped foods are stored in large air-circulated

rooms. To minimize weight loss, air movement around the unwrapped product should be the minimum required to maintain a constant temperature. With wrapped products, low air velocities are also desirable to minimize energy consumption. However, many storage rooms are designed and constructed with little regard to air distribution and the maintenance of low localized velocities over products. Horizontal throw refrigeration coils are often mounted in the free space above the racks or rails of product, and no attempt is made to distribute the air around the products. Using air socks, it is claimed that an even air distribution can be maintained with localized velocities not exceeding $0.2\,m\,s^{-1}$.

Controlled-atmosphere Storage Rooms

Controlled-atmosphere storage rooms were developed for specialized fruit stores, especially those for apples. Interest is growing in the application of this technique to other commodities including meat and fish. In addition to the normal temperature control plant, these stores also include special gas-tight seals to maintain an atmosphere, which is normally lower in oxygen and higher in nitrogen and carbon dioxide than air. An additional plant is required to control the CO_2 concentration, generate nitrogen and consume oxygen. (*See* **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs.)

There is growing interest in the use of controlled atmosphere retail packs to extend the chilled storage and display life of meat and meat products. Since the packs insulate the products, efficient precooling before packaging is important.

Transportation

Developments in temperature-controlled transportation systems for products have been one of the main factors leading to the rapid expansion of the chilled food market. The sea transportation of chilled meat from Australasia to European and other distant markets, and road transportation of chilled products throughout Europe and the Middle East, are now common practice. Air freighting was initially used for high-value perishable products such as strawberries, asparagus, and live lobsters. It is now increasingly used to provide consumers with a year-round supply of locally out-of-season products such as green beans, mangetout, and green on carrots.

Overland Transport

Overland transportation systems range from 12-m refrigerated containers for long-distance road or rail movement of bulk chilled products to small uninsulated vans supplying food to local retail outlets. The majority of current road transport vehicles for chilled foods are refrigerated using either mechanical, eutectic plates or liquid nitrogen cooling systems.

Many types of independent engine and/or electric motor-driven mechanical refrigeration units are available for lorries or trailers. One of the most common is a self-contained 'plug' unit mounted in an opening provided in the front wall of the vehicle. The condensing section is on the outside and the evaporator on the inside of the unit, separated by an insulated section, which fits into the gap in the wall. Units have one or two compressors, depending upon their capacity, which can be belt-driven from the vehicle but are usually driven direct from an auxiliary engine. This engine may use petrol from the vehicle's supply or an independent tank, or liquid petroleum gas. Many are equipped with an additional electric motor for standby use or for quiet running, e.g., when parked or on a ferry. Irrespective of the type of refrigeration equipment used, the product will not be maintained at its desired temperature during transportation unless it is surrounded by air or surfaces at or below that temperature. This is usually achieved by a system that circulates moving air, either forced or by gravity, around the load. Inadequate air distribution is probably the principal cause of product deterioration and loss of shelf-life during transport. Conventional forced air units usually discharge air over the stacked or suspended products either directly from the evaporator or through ducts towards the rear cargo doors. Because air takes the path of least resistance, it circulates through the channels, which have the largest cross-sectional area. These tend to be around, rather than through, the product. If products have been cooled to the correct temperature before loading and do not generate heat, they only have to be isolated from external heat ingress. Many trucks are now being constructed with an inner skin that forms a return air duct along the side walls and floor, with the refrigerated air being supplied via a ceiling duct.

Sea Transport

Recent developments in temperature control, packaging, and controlled atmospheres have increased substantially the range of foods that can be transported around the world in a chilled condition. Control of the oxygen and carbon dioxide levels in shipboard containers has allowed fruits and vegetables such as apples, pears, avocado pears, melons, mangoes, nectarines, blueberries, and asparagus to be shipped from Australia and New Zealand to markets in the USA, Europe, Middle East, and Japan. With conventional vacuum packing, it is difficult to achieve a shelf-life in excess of 12 weeks with beef and 8

weeks for lamb. However, a study has found that a shelf-life of up to 23 weeks at −2 °C can be achieved in cuts of lamb that are individually packed in evacuated bags of linear polyethylene, and then placed in a foil laminate bag that is gas-flushed and filled with a volume of CO_2 approximately equal to that of the meat. Similar storage lives are currently being achieved with beef primals transported from Australia and South Africa to the EU. Assuming good standards of preparation and prompt cooling, the times given in Table 1 could be used as approximate guidelines for long-distance meat shipment.

These times rely on the meat being at or below the storage temperature before loading. The 2–4-week advantage of transporting meat at −1.5 °C rather than 0 °C will be lost if the meat is loaded at a temperature above 0 °C. Cooling in the center of a load of meat is very slow, and the meat will be well into its journey before the desired temperature is achieved.

International Standard Organization (ISO) containers for food transport are 6 or 12 m long, hold up to 26 tonnes of product and can be 'insulated' or 'refrigerated.' The refrigerated containers incorporate insulation and have refrigeration units built into their structure. Insulated containers either utilize plug-type refrigeration units or may be connected directly to an air-handling system in a ship's hold or at the docks. Refrigerated containers are easier to transport over land than the insulated types, but have to be carried on deck when shipped because of problems in operating the refrigeration units within closed holds. On board ship, they are therefore subjected to much higher ambient temperatures and consequently larger heat gains, which make it far more difficult to control product temperatures. Close temperature control is most easily achieved in containers that are placed in insulated holds and connected to the ship's refrigeration system.

Air Transport

Although air freighting of foods offers a rapid method of serving distant markets, there are many problems because the product is unprotected by refrigeration for much of its journey. Up to 80% of the total journey time is spent on the tarmac or transport to and from the airport. Perishable cargo is usually carried in standard containers, sometimes with an insulating lining and/or dry ice, but is often unprotected on aircraft pallets.

Studies in Australia have led to the following recommendations for air transport of chilled foods:

1. Insulated containers should always be used to reduce heat gain.
2. Product should always be precooled and held at the required temperature until loading.
3. With products that deteriorate after any surface freezing, dry ice should not be used.
4. Containers should be filled to capacity.
5. A thermograph should accompany each consignment.

Table 1 Recommended transportation life of meat

	Vacuum pack 0 °C	*Vacuum pack −1.5 °C*	*CO_2 −1.5 °C*
Pork	6 weeks	8 weeks	
Lamb	7 weeks	10 weeks	> 12 weeks
Beef	10 weeks	14 weeks	

Retail Display

The retail display of chilled food is probably still the weakest link in the cold chain. In the UK, the Food Hygiene (Amendment) Regulations 1990 required food retailers to maintain the temperature of certain chilled foods below 5 °C and others 8 °C during storage, transport, and display. This legislation, which contained severe financial penalties for temperature infringements, produced a marked improvement in the design, construction, and operation of retail display cabinets. However, the legislation has since been relaxed, and there is currently less emphasis on the maintenance of food temperatures.

The required retail display life and consequent environmental conditions for wrapped chilled products differ from those for unwrapped products. The desired display life for wrapped meat, fish, vegetables, and processed foods ranges from a few days to many weeks and is primarily limited by microbiological considerations. Retailers of unwrapped fish, meat, and delicatessen products normally require a display life of one working day.

Display cabinets for delicatessen products are available with gravity or forced convection coils, and the glass fronts may be nearly vertical or angled up to 20 °. In the gravity cabinet, cooled air from the raised rear-mounted evaporator coil descends into the display well by natural convection, and the warm air rises back to the evaporator. In the forced-circulation cabinets, air is drawn through an evaporator coil by a fan and then ducted into the rear of the display. It then returns to the coil after passing directly over the products, or forming an air curtain, via a slot in the front of the cabinet and a duct under the display shelf.

Although the same cabinets can be used for wrapped foods, most are sold from multideck cabinets with single or twin air curtain systems. It is important that the front edges of the cabinet shelves do not project through the air curtain since the

refrigerated air will then be diverted out of the cabinet. However, if narrow shelves are used, the curtain may collapse, and ambient air can be drawn into the display well. External factors such as the store ambient temperature, the siting of the cabinet, and poor pretreatment and placement of products substantially affect cabinet performance. Computational fluid dynamics is increasingly being used as a valuable modeling tool to improve air distribution and temperature control in retail display cabinets and retail stores themselves.

See also: **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Fermented Milks**: Products from Northern Europe; **Fish**: Fish as Food; **Heat Transfer Methods**; **Milk**: Processing of Liquid Milk; **Pasteurization**: Principles; **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability

Further Reading

Anonymous (1994) *ASHRAE Handbook. Refrigeration*. New York: American Society of Heating, Refrigeration and Air Conditioning Engineers.

Anonymous (2000) *Recommendations for Chilled Storage of Perishable Produce*. Paris: International Institute of Refrigeration.

Gormley TR (1990) *Chilled Foods – The State of the Art*. Barking, UK: Elsevier Applied Science.

James SJ (1999) *Food Refrigeration 1990 to 2020 A.D*. Institute of Refrigeration Centenary Publication.

James C and James SJ (1996) *Past, Present and Future Methods of Cooling Foods*. Publication of MAFF Fellowship, University of Bristol.

James SJ and James C (2002) *Meat Refrigeration*. Cambridge, UK: Woodhead Publishing.

Jowitt R, Escher F, Hallstrom B *et al*. (1983) *Physical Properties of Foods*. Barking, UK: Elsevier Applied Science.

Perry RH and Chilton CH (1973) *Chemical Engineers' Handbook*, 5th edn. Tokyo: McGraw-Hill.

Stringer M and Dennis C (2000) *Chilled Foods: a Comprehensive Guide*, 2nd edn. Cambridge, UK: Woodhead Publishing.

Attainment of Chilled Conditions

S James, University of Bristol, Langford, North Somerset, UK

Background

The purpose of the chill chain is first to reduce the temperature of a food to below a set temperature and henceforth maintain it at or below that temperature. Since the growth of both pathogenic and food spoilage organisms is very temperature-dependent, the rapid attainment and maintenance of low temperatures are important to food safety and shelf-life. In this article, the cooling rates and temperatures that can be attained and maintained are discussed together with the limiting physical factors.

The speed at which a food can be cooled can be limited by either the rate of heat removal from its surface or internal conduction. Heat removal from the surface is a direct function of the surface heat transfer coefficient (h). Typical values of h range from $5\,W\,m^{-2}\,°C^{-1}$ for slow-moving air to $500\,W$ $W\,m^{-2}\,°C^{-1}$ for agitated water. **Table 1** shows that, at low values of h, a 10-fold increase, 5 to $50\,W\,m^{-2}\,°C^{-1}$, makes a substantial, 3.2- to 4.2-fold reduction in cooling time. A further 10-fold increase from 50 to $500\,W\,m^{-2}\,°C^{-1}$ decreases the cooling time threefold for a 2-cm-thick slab but only results in a 60% reduction at a thickness of 8 cm. This indicates that in the thicker material internal heat conduction is becoming rate-controlling.

Primary Chilling

Primary chilling after harvest or slaughter is usually applied to large individual items or small items cooled in bulk. Air cooling is often used in primary chilling because internal conduction controls the rate the products can be cooled. Data from the cooling of beef sides (**Figure 1**) illustrate the relative influence of air temperature and air velocity. The data have been presented as a plot of the logarithm of temperature against time and, provided air temperatures are chosen to avoid substantial surface freezing, it is quite feasible to determine the cooling time for any other air temperature. The fractional unaccomplished temperature on the y-axis can be replaced by the meat temperature calculated from $y = (t-t_f)/(t_i-t_f)$, where t is the meat temperature, t_i the initial meat temperature, and t_f the air temperature.

Cooling in air at a constant 4 °C, compared with 0 °C, at $3\,m\,s^{-1}$ will increase the time to reach 7 °C in

Table 1 Predicted cooling time (h) from 40 °C to 2 °C at the center

Cooling method	h *($W\,m^{-2}\,°C^{-1}$)*	*Meat thickness (cm)*		
		2.0	*4.0*	*8.0*
Air (still)	5	5.0	11.0	24.0
Air ($5\,m\,s^{-1}$)	50	1.2	2.8	7.4
Plate	360	0.7	1.8	5.5
Immersion	500	0.4	1.2	4.4

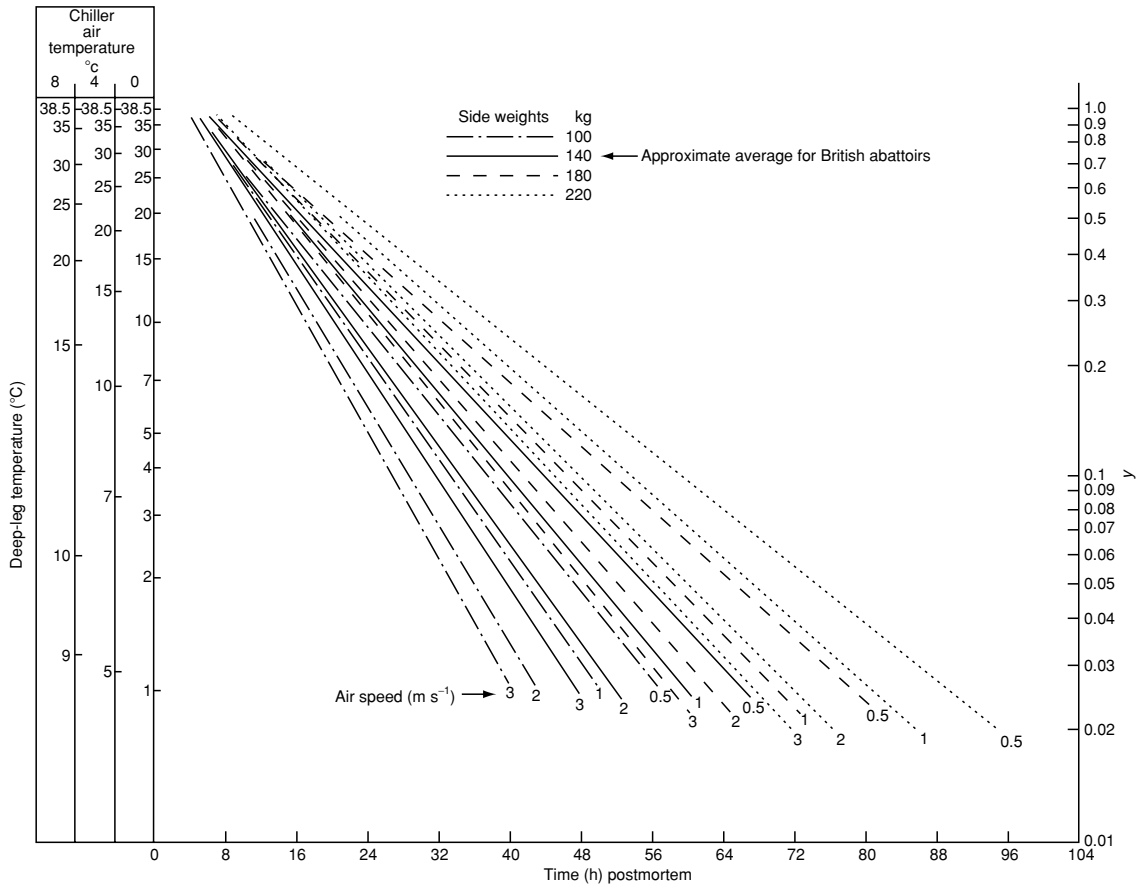

Figure 1 Relationship between deep-leg temperature and cooling time for beef sides. *y* is the fractional unaccomplished temperature which can be used to calculate cooling under different chiller temperatures.

the deep leg of a 100-kg side from 20.3 to 27.7 h (36% increase). At 0.5 m s^{-1} the same rise in air temperature for a 220-kg side will increase the chilling time from 45.9 to 68.3 h (49% increase).

The shelf-life of vegetables is largely governed by the amount of water that is lost from the vegetables. Leafy vegetables have a relatively short shelf-life because of the potential for high weight of evaporation. Their shelf-life can be greatly improved if the field heat is removed immediately after harvesting and the temperature of vegetables subsequently maintained at their optimum. For most winter vegetables, which have their origins in temperate climates, this temperature is close to 0 °C. Some tropical fruits and vegetables are damaged if the temperature is reduced too low and they should, therefore, be chilled and maintained at warmer temperatures.

Vegetables are cooled in many ways. One convenient bulk system is to pack the vegetables into large crates which are then stacked in front of a chiller and air is forced to flow through the crates back to the chiller in order to effect a fast rate of cooling. Weight loss will be rapid during the first parts of the process but when the temperature of the surface of the vegetables approaches that of the air then the relative humidity of the air will become important. Artificial ways of increasing the humidity of the air are common for vegetable cooling. The most normal system utilizes an ice bank from which chilled water at 0 °C is drawn. This is passed over packing to extend the surface area of water through which air from the chill room is drawn. The resulting air is at very high humidity and with a temperature of close to 0 °C. By controlling the temperature of the water and bypassing the air, the bulk humidity and temperature of the air can be controlled to lower relative humidities and warmer temperatures than these.

Vegetables can be readily cooled with running water. This reduces the evaporation of water to zero and even small weight gains can be achieved. The water can also be used to wash the vegetables. Subsequent storage should be in very high humidities to prevent further deterioration.

Leafy vegetables can be cooled in a vacuum cooling process. In this process, the vegetables are placed in a vacuum chamber and a vacuum is drawn. The

evaporation of water from the plant is used to promote rapid cooling. The small loss, 2% in the case of lettuces, is considered acceptable, given that the subsequently chilled lettuces have a longer shelf-life than those chilled by conventional means.

Secondary Chilling

Cooking rarely eliminates all food-poisoning organisms and a number survive as spores which will germinate and grow if cooling rates are slow. Rapid reduction in surface temperature retards microbial growth and consequently extends shelf-life. This is especially important when chilling cooked products that will eventually be consumed cold or in a warm reheated state. Rapid rates of cooling are required without surface freezing. The problem is complicated in two-compartment ready-meal consumer packs. These typically contain rice or pasta in one compartment, and a meat- or fish-based product in the second. The thermal properties of the two items are often very different and they may be filled to different depths. Air at $-10\,^{\circ}C$ and $5\,m\,s^{-1}$ produces a cooling time of 34 min, but substantial quantities of the product are frozen. At an air velocity of $0.5\,m\,s^{-1}$ the cooling time is doubled but only a small area of the rice is frozen. With higher air temperatures, the extent of freezing is reduced. A cooling time of approximately 0.75 h is achieved at $-5\,^{\circ}C$, $5\,m\,s^{-1}$ with acceptable freezing.

The importance of achieving a minimum required air velocity around small products is clearly demonstrated by data obtained from cooling pork pies (Figure 2). To guarantee that all the crust remained above $-2\,^{\circ}C$ on the unwrapped 400-g (70-mm high, 95-mm diameter) pies, an air temperature of $-1.5 \pm 0.5\,^{\circ}C$ was used. At this temperature, a small increase in air velocity from 0.5 to $1.0\,m\,s^{-1}$ reduced the cooling time by 85 min (almost 30%). Even at very high velocities ($> 6.0\,m\,s^{-1}$) appreciable reductions in cooling time were still being achieved.

Rapid rates of temperature reduction can be achieved in trays of cooked product such as mince, baby foods, and poultry portions when cooled under vacuum.

Slow cooling of large hams and cooked meat has been implicated in a number of food-poisoning outbreaks. It is now recommended that large uncured joints are cooled in < 8 h and cured joints in < 10 h. In laboratory studies three different cooling methods – forced air at $0\,^{\circ}C$, $1.2\,m\,s^{-1}$, immersion in water at $0\,^{\circ}C$, and cooling under vacuum – for large (6.8–7.3 kg) hams, and other meats were compared (Table 2). Although cooling times are much faster under vacuum, weight losses (typically 7.5–9.5%) are much higher than those in air (typically 2–3%).

Storage, Transport, and Display

In general, after initial chilling, as a chilled product moves along the chill chain it becomes harder to control and maintain its temperature. Temperatures of bulk packs of chilled product in large storerooms are far less sensitive to small heat inputs than single consumer packs in transport or open display cases. If primary and secondary cooling operations are

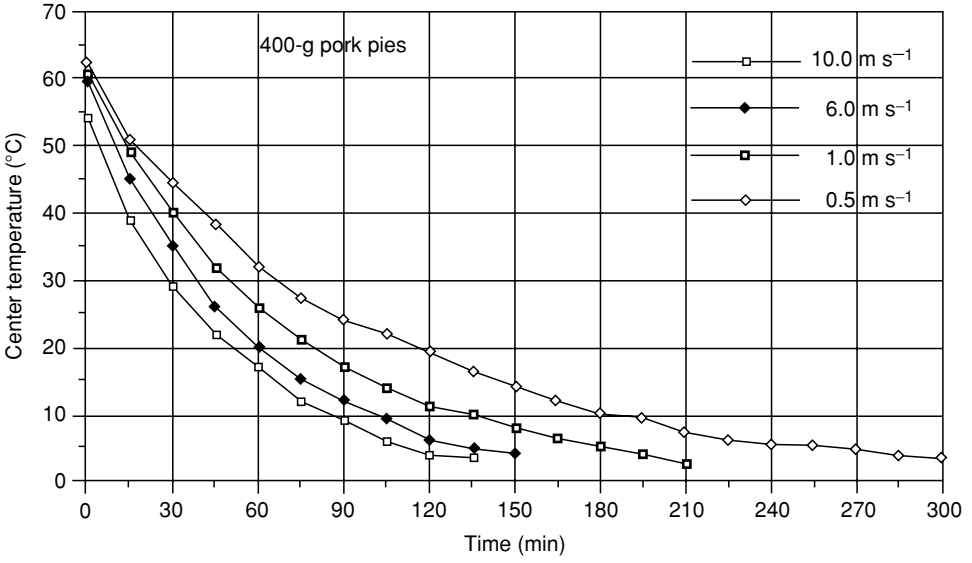

Figure 2 Temperature at slowest cooling point in 400-g pork pies at $-1.5\,^{\circ}C$ air temperature and velocities of 10, 6, 1, and $0.5\,m\,s^{-1}$.

Table 2 Cooling time to 10 °C for different cooked meats

Type of meat	*Weight (kg)*	*Initial temperature (°C)*	*Cooling time (min) to 10 °C in*		
			Air	*Immersion*	*Vacuum*
Beef					
M. semitendinosus	0.9	75	150	100	52
Forequarter	2.7	75	390	282	43
Silverside	2.7	75	338	240	61
Boneless turkey	6.4	85	526	411	36
Boned out ham	7.1	80	761	456	57

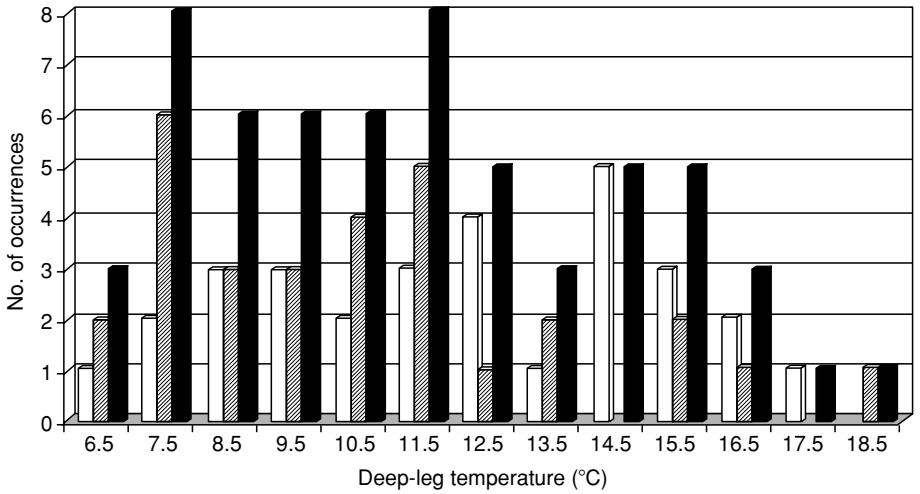

Figure 3 Number of occurrences of deep-leg temperatures in different ranges when loaded into railway wagons or road trailers. Open columns, railway wagons; cross-hatched columns, road trailers; filled columns, road and rail.

efficiently carried out, then the food will be reduced below its required temperature before it is placed in store. In this situation the store's refrigeration system is only required to extract extraneous heat that enters through the walls, door openings, etc.

Efficient cooling is also important before transportation. A study of different transportation methods found that the deep temperature in beef sides and quarters at the time of their loading into transport vehicles in three USA plants ranged from 6 to 18 °C (Figure 3). In rail wagons the surface temperature declined during the first 24 h and was subsequently maintained at a temperature of 0 ± 1 °C. In the road vehicles the surface temperature fell slowly during the whole journey and had not attained a steady minimum value when unloaded. On average the deep temperature of sides in rail wagons reached 1 °C after 72 h. Temperatures in quarters in road vehicles were still above 2 °C after 120 h.

Even when temperature-controlled dispatch bays are used, there is a slight heat pick-up during loading. In bulk transportation, the resulting temperature rise is small and the refrigeration system of the vehicle rapidly returns the product to the required temperature. Larger problems exist in local multidrop distribution to individual stores. There are substantial difficulties in maintaining the temperature of chilled foods transported in small refrigerated vehicles that conduct multidrop deliveries to retail stores and caterers. The vehicles have to carry a wide range of products and operate under diverse ambient conditions. During any one delivery run, the chilled product can be subjected to as many as 50 door openings, where there is heat ingress directly from outside and from personnel entering to select and remove product. The design of the refrigeration system has to allow for extensive differences in load distribution, dependent on different delivery rounds, days of the week, and the removal of product during a delivery run. A refrigeration system's ability to respond to sudden demands for increased refrigeration is often restricted by the power available from the vehicle. All these problems combine to produce a complex interactive system. Computer programs such as CoolVan are now available to aid the design and operation of local delivery systems.

Surveys carried out in the UK, Denmark, and Sweden (Figure 4) have revealed a very wide range

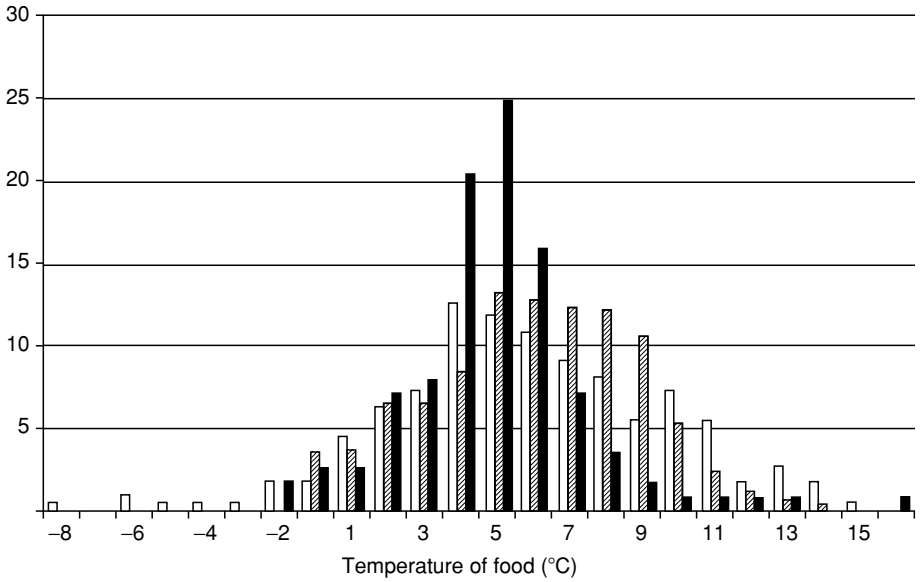

Figure 4 Temperature of chilled foods in retail display cabinets in the UK (open columns), Denmark (cross-hatched columns), and Sweden (filled columns). Data from [a]Bøgh-Sørensen L (1980) Product temperatures in chilled cabinets. In: *26th European Meeting of Meat Research Workers*, Colorado, paper 22, [b]Olsson P (1990) Chill cabinet surveys. In: *Processing and Quality of Foods. Volume 3. Chilled Foods: The Revolution in Freshness*, pp. 3.279–3.288. London: Elsevier Applied Science and [c]Rose SA (1986) Microbiological and temperature observations on pre-packaged ready-to-eat meats retailed from chilled display cabinets. In: *Recent advances and developments in the refrigeration of meat by chilling*, pp. 463–470. Paris: International Institute of Refrigeration.

Table 3 Maximum temperature (°C) measured and increase in bacterial numbers (generations) during 1 h in car followed by 5 h in a domestic refrigerator

Product	*Conditions*	*Maximum temperature*	Pseudomonas	Clostridium
Pâté	Ambient car	25	1.5	0.4
	Cool box in car	13	< 0.4	0
Raw chicken	Ambient car	24	1.6	0.2
	Cool box in car	4	0	0
Cooked chicken	Ambient car	28	1.8	0.7
	Cool box in car	12	0	0
Prawns	Ambient car	37	1.3	1.6
	Cool box in car	14	0	0
Brie cheese	Ambient car	28	2.2	0.8
	Cool box in car	12	0	< 0.1

of temperatures within foods on retail display. Legislation on maximum temperatures for the display of chilled foods is in place in many European countries.

Domestic

Although legislation such as that introduced in the UK should insure that food producers and retailers maintain acceptable product temperatures during the distribution chain, they lose control when the product leaves the retail store. After a chilled product is removed from a display cabinet it spends a period outside a refrigerated environment whilst it is carried around the store and then transported home.

Investigations have compared unprotected transportation in a car boot with that in an insulated box. Initial product temperatures measured when the food reached the car ranged from 4 to over 20 °C. Some product temperatures on samples placed in the boot rose to approaching 40 °C (Table 3) during the 1-h car journey whilst most of the samples placed in the insulated box cooled during the car journey, except for a few at the top of the box which remained at their initial temperature. Small products, i.e., prawns, showed the highest temperature changes during transport. Thicker products like cooked chicken were less influenced. After being placed in the domestic refrigerator it required approximately 5 h before the temperature was reduced below 7 °C.

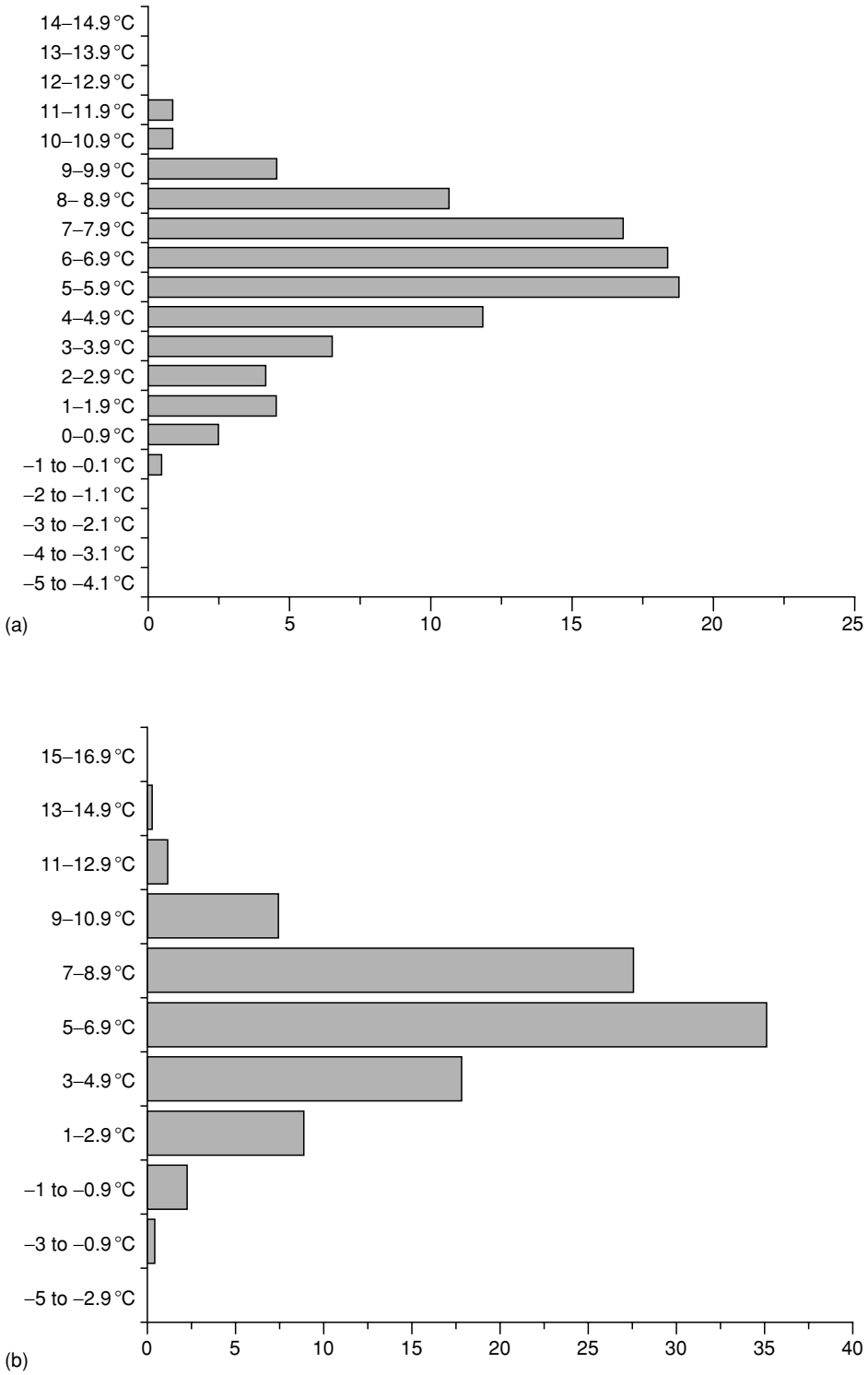

Figure 5 (a) Food temperatures and (b) air temperatures measured in domestic refrigerators in 252 households in the UK. (a) percentage of food with temperature in stated range; (b) percentage of refrigerators with overall average temperatures within stated range.

Predictions made using a mathematical model that calculated bacterial growth from temperature/time relationships indicated that increases of up to 1.8 generations in bacterial numbers (Table 3) can occur during this transport and domestic cooling phase. The model assumes that bacteria require a time to acclimatize to a change in temperature (the lag phase) and that no acclimatization had occurred during display. If this rather optimistic assumption was not made, then up to 4.2 generations of *Pseudomonas* and growth of both *Salmonella* and *Listeria* were predicted. Very small increases in bacterial numbers (< 0.4 generations: Table 3) were predicted when the insulated box was used, due to the lower product temperatures.

Chilled foods spend a period between a few minutes and many days stored in domestic refrigerators. The average food temperature measured in a survey of refrigerators in 252 UK households was approximately 6 °C (Figure 5a). However, in over 15% of the refrigerators the average food temperature was above 8 °C. In a typical UK refrigerator, a chilled food would spend over 35% of its time in an air temperature above 7 °C (Figure 5b).

See also: **Chilled Storage**: Principles; Quality and Economic Considerations; **Vegetables of Temperate Climates**: Leaf Vegetables

Further Reading

Anonymous (1972) *Meat Chilling – Why and How? MRI Symposium no. 2*. Bristol: Langford.

Anonymous (1986) *Proceedings of the International Institute of Refrigeration. Recent Advances and Developments in the Refrigeration of Meat by Chilling*. Bristol.

Anonymous (1997) *Proceedings of Meat Refrigeration – Why and How? EU Concerted Action Programme CT94 1881*. Bristol, Langford: FRPERC, University of Bristol.

Anonymous (2001) *Proceedings of the International Institute of Refrigeration. Rapid Cooling – Above and Below Zero*. Paris: International Institute of Refrigeration.

Bøgh-Sørensen L (1980) Product temperatures in chilled cabinets. In: *26th European Meeting of Meat Research Workers*, Colorado, paper 22.

Cutting CL (1972) *Meat Chilling – Why and How? MRI Symposium No 2*. Bristol: Langford, FRPERC, University of Bristol.

Evans JA, Stanton JI, Russell SL and James SJ (1991) *Consumer Handling of Chilled Foods: A Survey of Time and Temperature Conditions*. London: Ministry of Agriculture, Fisheries and Food.

Gigiel AJ, James SJ and Evans JA (1998) *Controlling Temperature During Distribution and Retail. EU Conference on Better and Safer Food*. Kalsrhue: Bundesforschungsanstalt für Emährang.

James SJ and Evans JA (1990) Temperatures in the retail and domestic chilled chain. In: *Processing and Quality of Foods*, vol. 3. *Chilled Foods: The Revolution in Freshness*, pp. 3.273–3.278. London: Elsevier Applied Science.

James SJ and James C (2002) *Meat Refrigeration*. Cambridge: Woodhead Publishing Limited.

Olsson P (1990) Chill cabinet surveys. In: *Processing and Quality of Foods. Volume 3. Chilled Foods: The Revolution in Freshness*, pp. 3.279–3.288. London: Elsevier Applied Science.

Ramila A, Valin C and Taylor AA (1988) *Accelerated Processing of Meat*. London: Elsevier Applied Science.

Rose SA (1986) Microbiological and temperature observations on pre-packaged ready-to-eat meats retailed from chilled display cabinets. In: *Recent advances and developments in the refrigeration of meat by chilling*, pp. 463–470. Paris: International Institute of Refrigeration.

Quality and Economic Considerations

S James, University of Bristol, Langford, North Somerset, UK

Introduction

Quality and economic considerations are often the most important factors that govern the choice of a chilling system and the subsequent chilled chain for a particular foodstuff. Since the microbiological aspects are covered elsewhere, quality considerations in this chapter relate to organoleptic and nutritional changes. Chemical and biochemical changes within foods cause modifications in the appearance, taste, or texture that can limit its high quality shelf life. These changes are not always detrimental to the eating quality of the food. For example, biochemical changes, referred to as 'conditioning' or 'aging', that occur in meat after slaughter lead to improvements in both its texture and taste.

In general with meat- and fish-based foods, the longest high-quality shelf life is achieved by rapidly reducing the temperature of the food and then maintaining it at a temperature very close to its initial freezing point. However, this is not true of all foods. For example, tropical fruits such as bananas suffer discoloration at temperatures below 12 °C, whereas salad vegetables such as cucumbers lose their textural properties at temperatures below 6 °C. Different cultivars of tomato suffer damage, called watersoaking and softening at temperatures below 7 or 10 °C, pitting, and russeting occur in beans below 7 °C and brown core in apples below 2 °C.

Primary and Secondary Chilling

Quality Considerations

In the majority of foods, fast cooling is beneficial, but there are instances where excessively rapid chilling rates cause quality problems in particular foods. For example, a serious defect known as 'woolly texture' can be produced in rapidly cooled peaches. Biochemical constraints lead to toughening if the lean tissue of beef or lamb is reduced to 10 °C or below within 10 h of slaughter. Due to differences in the biochemistry of pork muscle, the same temperature has to be achieved within 3 h to cause toughening. (*See* **Meat**: Preservation.)

Different foodstuffs exhibit particular quality advantages because of rapid chilling. In meat, the pH starts to fall immediately after slaughter, and protein denaturation begins. The result of this denaturation is a pink, proteinaceous fluid, commonly called 'drip', often seen in pre-packaged joints. The rate of denaturation is directly related to temperature, and it follows therefore that the faster the chilling rate, the less the drip. With both pork (Table 1) and beef, the use of rapid rates of chilling can halve the amount of drip loss. Fish passing through rigor mortis above 17 °C are largely unusable because the fillets shrink and become tough. A relatively short delay of an hour or two before chilling can demonstrably reduce shelf life. In a different commodity area, freshly harvested sweetcorn loses 5, 20, or 60% of its sugar content after 24 h in air at 0, 10, and 30 °C, respectively. Prompt cooling is clearly required if this vegetable is to retain its desirable sweetness. Similarly, the ripening of fruit can be controlled by rapid cooling, the rate of ripening declining as temperature is reduced and ceasing below about 4 °C. (*See* **Fish**: Processing; **Ripening of Fruit**.)

Rapid cooling is also often desirable with cooked products to maintain quality by eliminating the overcooking that occurs during slow cooling. For vitamin retention, the time taken to reduce the center temperature from 80 to 15 °C is a critical factor. The vitamin C content in nonpasteurized meals is reduced by 1 to 12% if cooling is carried out in 0.5 h, by 2 to 17% if the time is increased to 2 h, and by 10 to 38% when cooling takes 5 h. (*See* **Ascorbic Acid**: Properties and Determination.)

Vacuum-cooling is a very effective method of cooling cooked beef joints and hams. However, the process results in high weight losses and some reduction in eating qualities of the meat. Intermittent spraying to replace the water loss is practiced in vacuum-cooling of vegetables. It could have an application with meat.

Table 1 Drip loss after 2 days' storage at 0 °C, from leg joints from different breeds of pig cooled at different rates

Breed	*Drip loss (percent by weight)*	
	Slow	*Quick*
Landrace	0.47	0.24
Large White	0.73	0.42
Wessex × Large White	0.97	0.61
Pietrain	1.14	0.62

Economic Considerations

It is clear from restrictions already detailed that attempts to increase chilling rates are complicated by many factors, but there are a number of clear advantages in production economics if faster cooling can be achieved. Most foods are of high value, and any increase in rate of product throughput will improve cash flow and utilize expensive plants more efficiently. For example, the cooling time of 400-g pies can be reduced by 10 min if the air velocity is raised from 6 to 10 m s^{-1}. In high-throughput baking lines (> 1000 items per hour), the 7% increase in throughput could justify the higher capital and running costs of larger fans. The power required by the fans to move the air within a chill room increases with the cube of the velocity. A fourfold increase in air velocity from 0.5 to 2 m s^{-1} results in a 4.4-h (18%) reduction in chilling time for a 140-kg beef side, but requires a 64-fold increase in fan power. Increasing air velocity to 3 m s^{-1} only achieves an extra 6% reduction in chilling time. In most practical situations, where large items (e.g., meat carcasses, tuna, bins of vegetables) are being cooled, it is doubtful whether an air velocity greater than 1 m s^{-1} can be justified.

Efficient chilling produces a reduction in weight loss, which results in a higher yield of saleable material. Most foods have a high water content, and the rate of evaporation depends on the vapor pressure at the surface. Vapor pressure increases with temperature, and thus any reduction in the surface temperature will reduce the rate of evaporation. The effect of air temperature and velocity on evaporative weight loss during chilling is dependent upon the endpoint of the chilling process (Figure 1a and b). When chilling for a set time, weight loss increases as the temperature decreases and velocity increases. The opposite effect is found when chilling to a set temperature.

If specified cooling schedules are to be attained, refrigeration machinery must be designed to meet the required heat extraction rate at all times during the chilling cycle. Heat enters a chill room via open doors, from personnel, through the insulation, from lights and cooling fans, and from the cooling

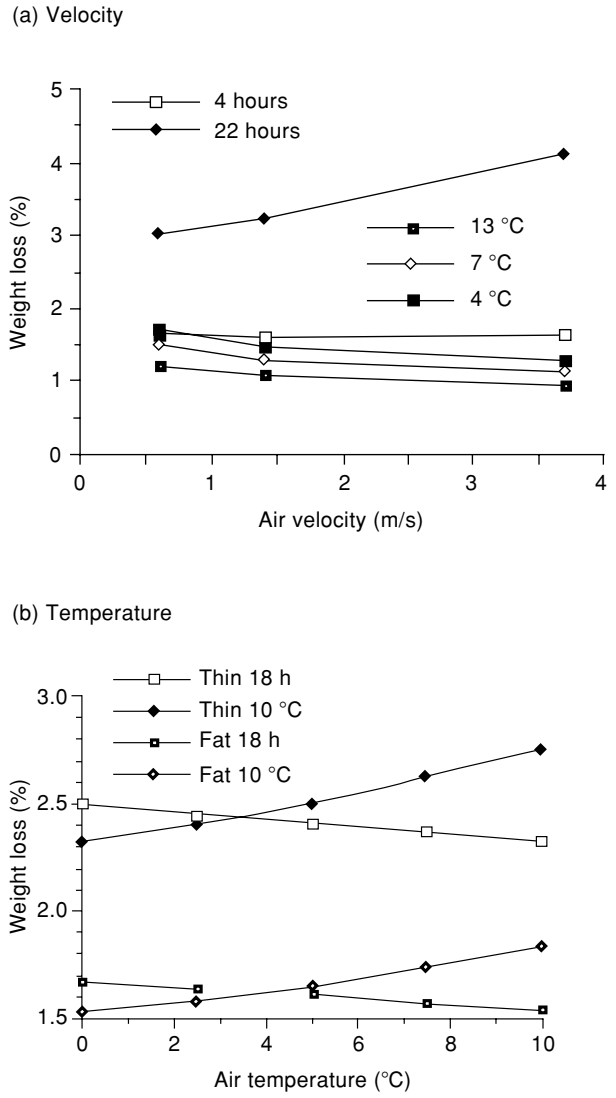

Figure 1 Relationship between weight loss and (a) air velocity for lamb carcasses and (b) air temperature for beef sides when chilling either for a set time or to a final temperature.

products. The product load is the major component of the total heat to be extracted from a fully loaded chill room. The rate of heat release from a food varies with time. It is at a peak immediately after loading and then falls rapidly.

The peak value is primarily a function of the environmental conditions during chilling. In commercial systems, the peak load imposed on the refrigeration plant is also a function of the rate at which hot food is introduced into the chill room. Increasing the air velocity, decreasing the air temperature, or shortening the loading time increases the peak heat load. In beef-chilling, for example, there is a fourfold difference in peak load between a chill room operating at 8 °C, 0.5 m s^{-1} loaded over 8 h and the same room operating at 0 °C, 3 m s^{-1} and loaded over 2 h.

Alternative chilling treatments can produce large increases in throughput and reductions in weight loss over conventional treatments (Table 2). The cash savings that result from the reduced weight losses can substantially increase the overall profits of many slaughtering operations (Table 2).

Commercial Storage, Retail Display, and Domestic Storage

Packaged Food

Most packaging systems substantially reduce evaporative weight loss from the surface of the product, and many retard the rate of chemical degradation. The use of modified atmospheres or opaque packaging can limit color changes during the chilled storage of meat, fish, and fruits.

Losses in vitamin content tend to be reduced at lower storage and display temperatures. Vitamin C losses from lettuce average 4.8, 5.6, and 6.6% per day at temperatures of 1, 5, and 10 °C, respectively.

Table 2 The weight loss, saving (£) over conventionally chilled controls, total work done at 24 h post mortem, cooling time to 10 °C for each treatment, drip loss from loin chops and annual saving for each treatment compared with the conventional chill used as a control for an abattoir slaughtering an average of 1000 pigs per week (74 kg dead weight at £0.90 kg^{-1})

Chilling treatment	*Weight loss at 24 h (%)*	*Cooling time to 10 °C*		*Drip (%)*	*Texture work (J)*	*Saving (£)*
		Deep leg (h)	*Surface leg (h)*			
(1) Ultrarapid						
Side	1.13 (0.85)	> 7.0	5.0[a]	2.3	0.18	28 300
Whole carcass	1.10 (1.01)	> 8.0	4.0[a]	0.9	0.21	33 600
(2) Ultrarapid two-stage	2.00 (0.29)	7.0		0.2	0.23	9 600
(3) Immersion	0.32 (2.08)	5.4	3.6	1.5	0.30	69 200
(4) High humidity	1.96 (0.21)	12.9	8.7	0.8	0.20	7 000
(5) Delay+high humidity	1.93 (0.24)	14.8	9.1	1.0	0.20	8 000
(6) Delay+spray	0.95 (1.22)	14.1	9.2	1.0	0.20	40 600
(7) Rapid+high humidity	1.53 (0.64)	11.6	1.3	0.9	0.21	21 300
(8) Rapid+conventional	1.67 (0.50)	11.5	1.6	1.0	0.24	16 600

[a]Measured at intersection of lean and fat.

Larger effects of temperature are found in French beans with equivalent losses of 1.9, 3.0, and 5.1% per day. No losses were reported in carrots at the three temperatures.

Oranges and pineapples are examples of foods that have an optimum temperature for vitamin C retention. In oranges, a 10% loss per month at 10 °C increases to 12% at 5 °C and 26% at 0 °C. Raising the storage temperature above 10 °C also increases the rate of loss to 13% at 15 °C and 22% at 30 °C. A similar optimum at 10 °C is found in pineapples.

Vitamin B_6 retention is also a function of both storage temperature and species. Loss from lettuce and French beans increases twofold and threefold, respectively, as the temperature is increased from 1 to 10 °C. No change was reported in parsley or carrots. (*See* **Vitamin B_6**: Properties and Determination.)

Unwrapped Food

Changes in appearance are normally the criteria that limit storage and display of unwrapped foods, commercial buyers or domestic consumers selecting fresh or newly loaded product in preference to that displayed or stored for some time. Deterioration in appearance has been related to degree of dehydration in red meat (Table 3), and similar changes are likely to occur in other foods. Apart from any relationship to appearance, weight loss is of considerable importance in its own right. The direct cost of evaporative loss from unwrapped foods in chilled display cabinets in the UK is in excess of £5 million per annum.

In the equation that governs weight loss, already described, the mass-transfer coefficient, m, is a function of air velocity. Typical values range from approximately $2 \times 10^{-8}\,\mathrm{kg\,m^{-2}\,s^{-1}\,Pa^{-1}}$ at $0.25\,\mathrm{m\,s^{-1}}$ to $14 \times 10^{-8}\,\mathrm{kg\,m^{-2}\,s^{-1}\,Pa^{-1}}$ at $1.5\,\mathrm{m\,s^{-1}}$. The vapor pressure difference ($P_s a_w - P_m$) is a function of surface temperature and dryness, and the temperature and relative humidity of the air.

In storage rooms at temperatures in the range of 0–2 °C, 76–90% RH, the average weight loss from cauliflower, French beans, and green peas is in the range of 0.1–0.4% per day. Under conditions typical of domestic refrigeration, 4–8 °C and 70–90% RH, weight losses from the same products range from 0.3 to 3.0% per day. In a nonrefrigerated ambient, 16–24 °C and 50–70% RH, losses are even higher in the range of 1.0–4.0% per day.

The relative effects of air temperature, velocity, and humidity on weight loss in retail display are demonstrated in Figure 2. Changes in relative humidity have a substantial effect with a reduction from 95 to 40%, increasing weight loss over a 6-h display period by a factor of between 14 and 18. The effect of air velocity on weight loss is confounded by that of relative humidity. Raising the air velocity from 0.1 to $0.5\,\mathrm{m\,s^{-1}}$ has little effect on weight loss at 95% RH but

Table 3 Relationship between change in appearance and evaporative weight loss ($\mathrm{g\,cm^{-2}}$)

Evaporative loss	*Change in appearance*
Up to 0.01	Red, attractive, and still wet; may lose some brightness
0.015–0.025	Surface becoming drier; still attractive but darker
0.025–0.035	Distinct obvious darkening; becoming dry and leathery
0.05	Dry, blackening
0.05–0.10	Black

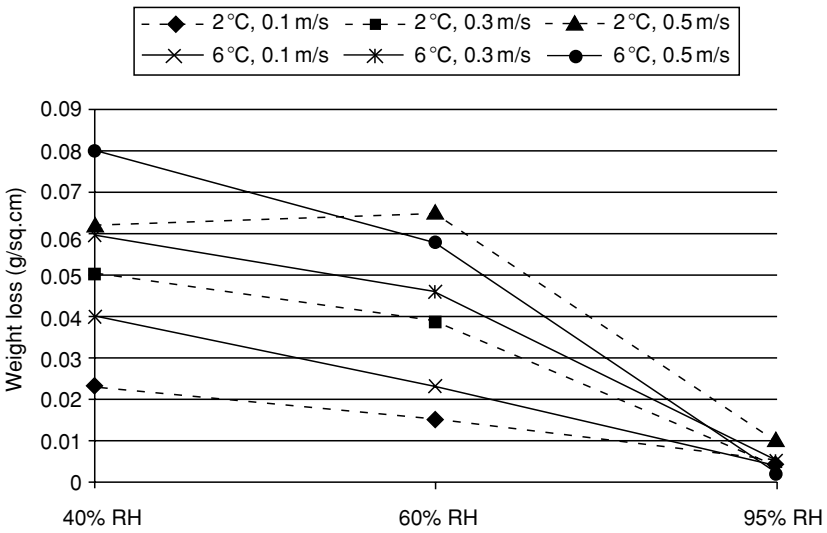

Figure 2 Weight loss from samples of beef steak under simulated display conditions.

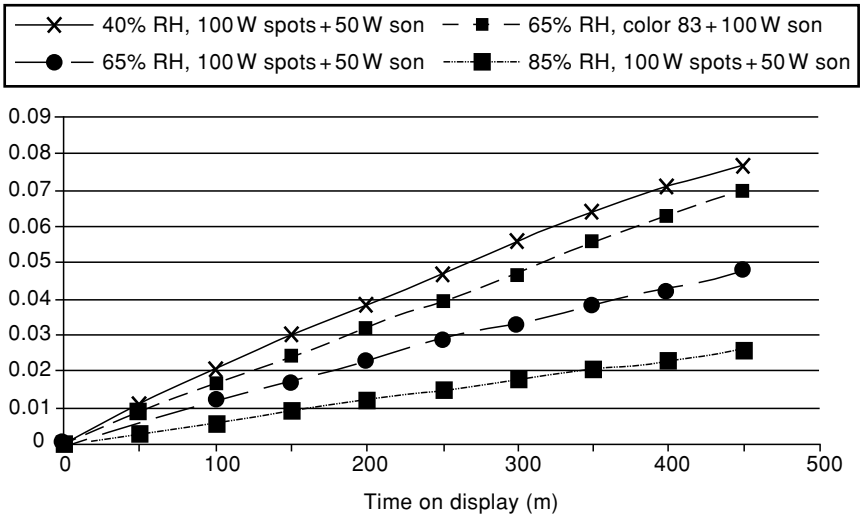

Figure 3 Weight loss from delicatessen products under different humidity and lighting conditions.

increases the loss by a factor of between 2 and 2.4 at 60% RH. A temperature change from 2 to 6 °C has a far smaller effect on weight loss than changes in either relative humidity or velocity.

In further work, a model developed to predict the rate of weight loss from unwrapped meat under the range of environmental conditions found in chilled retail displays showed that it was governed by the mean value of the conditions. Fluctuations in temperature or relative humidity had little effect on weight loss, and any apparent effect was caused by changes in the mean conditions.

There is a conflict between the need to make the display attractive and convenient to increase sales appeal and the optimum display conditions for the product. High lighting levels increase the heat load, and the air temperature in the cabinet rises. This rise increases the temperature difference across the evaporator coil, and the air entering the cabinet is dehumidified. Consequently, the rate of evaporative weight loss from foods on display is increased.

Studies have shown that changing a lighting combination of 50 W sons and 100 W halogen lights to 100 W sons and a colour 83 fluorescent significantly increased the weight loss (Figure 3). The increase was similar in magnitude to that produced by 20% reduction in relative humidity. On average, the rate of weight loss under the combination of 50 W sons and 100 W halogen (spot) lights was approximately 1.4 times less than the 100 W sons and colour 83 fluorescent lighting.

See also: **Ascorbic Acid**: Properties and Determination; **Fish**: Processing; **Meat**: Preservation; **Ripening of Fruit**; **Vitamin B_6**: Properties and Determination

Further Reading

Anonymous (1974) *ASHRAE Handbook and Product Directory. Applications*. New York: American Society of Heating Refrigeration and Air Conditioning Engineers.

Food Refrigeration & Process Engineering Research Centre (FRPERC) (1997) *Meat Refrigeration – Why and How? EU Concerted Action Programme CT94 1881*. Langford, UK: FRPERC.

FRPERC (2000) Publications on the refrigeration and thermal processing of food. University of Bristol. http://www.frperc.bris.ac.uk.

Gormley TR (1990) *Chilled Foods – The State of the Art*. Barking, UK: Elsevier Applied Science.

International Institute of Refrigeration (1985) *Technology Advances in Refrigerated Storage and Transport*. Orlando, FL: International Institute of Refrigeration – Commissions D1, D2 & D3.

International Institute of Refrigeration (1986) *Recent Advances and Developments in the Refrigeration of Meat by Chilling*. Bristol, UK: International Institute of Refrigeration Commission C2.

International Institute of Refrigeration (1990) *Progress in the Science and Technology of Refrigeration in Food Engineering*. Dresden: International Institute of Refrigeration – Commissions B2, C2, D1, D2/3.

International Institute of Refrigeration (2000) *Recommendations for Chilled Storage of Perishable Produce*. Paris: International Institute of Refrigeration ISBN 2-913149-09-X

James SJ (1999) Food refrigeration and thermal processing at Langford, UK: 32 years of research. *Transactions of the IChemE Food and Bioproducts Processing* 77(C): 261–279.

Zeuthen P, Cheftel JC, Eriksson C, Gormley TR, Lonko P and Paulus K (1990) *Chilled Foods: The Revolution in Freshness*. Barking, UK: Elsevier Applied Science.

Microbiological Considerations

S J Walker, Campden & Chorleywood Food Research Association, Chipping Campden, Gloucestershire, UK

Background

During storage, many foods and their raw materials will deteriorate and become unacceptable to the end user. Food preservation techniques aim to prevent or retard this deterioration. Chilled storage can be rarely relied upon to prevent spoilage, but can be used to extend the time until this becomes apparent. Such storage has major effects on the rates of biochemical and biological changes in foods, although this article will focus on microbiological issues in chill stored foods and ingredients.

Foods or ingredients intended for chilled storage may initially contain a wide variety of microorganisms derived from the material itself or contaminants arising from harvesting, storage, and processing. The organisms that spoil chilled foods will be those that are best adapted to these temperatures.

Effects of Chilled Storage on Microorganisms

If the temperature of chilled storage is below the minimum growth temperature for a microorganism, no growth will occur. However, microbial growth has been reported for a number of microorganisms at temperatures below 0 °C. Consequently, chilled storage cannot be relied upon for microbial stability. The major effects on microbial growth are to extend the time until an increase in numbers is apparent (the lag time) and to reduce the rate of growth when this occurs.

Cardinal Temperatures

Microorganisms may be characterized by their abilities to grow at specific temperatures, commonly referred to as the cardinal temperatures.

With chilled foods, the factor of most concern is the *minimum growth temperature* (MGT), which represents the lowest temperature at which growth of a particular microorganism can occur. If the MGT of a microorganism is greater than 10 °C, this microorganism will not grow during chill storage. Care is needed with published MGT values. For example, if the time period for the investigation reporting this value were too short, the resultant value would be overestimated. As the MGT is affected by other factors, including the pH, salt, preservatives, and previous heat treatments, a true estimate can be determined only when other factors are optimal for growth.

Although gradual death may occur if a microorganism is stored below its MGT, this should not be considered to be a lethal process, as, in many cases, growth will resume if the temperature is subsequently raised.

The *optimum growth temperature* represents the temperature at which the biochemical processes governing growth of a particular microorganism are overall operating most efficiently. At this temperature, the lag phase before growth is minimized and the growth rate maximized. As the temperature rises above the optimum, the rate of growth decreases until the *maximum growth temperature* is reached. In general, the maximum growth temperature is only a few degrees higher than the optimum. At temperatures just above the maximum for growth, cell injury starts to occur. However, if the temperature is subsequently reduced, growth may resume, although a period of time may be required to permit cell repair. At higher temperatures, the inactivation of one or more critical enzymes in the microorganism becomes irreversible, and cell damage occurs, leading to cell death.

Based on the relative positions of the cardinal temperatures, microorganisms can be divided into four main groups, viz. psychrophiles, psychrotrophs, mesophiles and thermophiles (Table 1). With chilled foods, the groups of most concern are the psychrophiles and psychrotrophs. Whilst these terms have been used synonymously in the past, it is now accepted that the major spoilage microorganisms of chilled foods are psychrotrophic in nature. True psychrophiles are rare in food microbiology and generally limited to some microorganisms from deepsea fish. The minimum growth temperatures for many of the microorganisms relevant to chilled foods are given in Table 2.

Microbial Growth in Chilled Foods

Whilst microbiological spoilage of chilled foods may take diverse forms, it generally manifests itself in a change in the sensory characteristics. In the simplest form, this may be due to the production of visible

Table 1 Cardinal temperatures for microbial types

	Temperature (°C)		
	Minimum	*Optimum*	*Maximum*
Psychrophiles	< 0–5	12–20	20–22
Psychrotrophs	< 0–5	20–35	30–40
Mesophiles	10	30–40	40–45
Thermophiles	30–40	45–65	60– > 80

Table 2 Minimum growth temperatures of microorganisms (Adapted from Martens, 1999)

Microorganism	*Minimum temperature (°C)*[a]
Bacillus cereus	
Mesophilic	15
Psychrotrophic	4
Bacillus species (spoilage)	0–15
Brochothrix thermosphacta	4
Campylobacter jejuni	32
Clostridium botulinum:	
Mesophilic (proteolytic)	10
Psychrotrophic (nonproteolytic)	3
Clostridium perfringens	12
Coliforms	2–8[b]
Escherichia coli	4
Escherichia coli O157	6.5
Lactobacillus species	0–10[b]
Listeria monocytogenes	0
Micrococcus species	1–10[b]
Pseudomonas species	−3
Salmonella	7
Staphylococcus aureus	6 (10 for toxin)
Vibrio parahaemolyticus	5
Yeasts	−10
Yersinia enterocolitica	−1

[a]Under otherwise optimal conditions. These figures are indicative only and not necessarily representative of all strains of microorganisms in foods, and will also depend on the composition of the food and the storage conditions.
[b]Depending on the strain.
From Martens T (1999) *Harmonisation of Safety Criteria for Minimally Processed Foods: Rationale and Harmonisation Report. European Commission FAIR CT-96-1020*.

growth, and this is common in molds that produce large, often pigmented colonies. Bacteria and yeasts also may produce visible (sometimes pigmented) colonies on foods. Other forms of spoilage include the production of gases, slime (extracellular polysaccharide material), diffusible pigments, and enzymes that may produce softening, rotting, off-odors and off-flavors from the breakdown of food components.

When only a few spoilage microorganisms are present, the consequences of growth may not be apparent. If, however, the microorganisms multiply during storage, the production of deleterious compounds or deterioration in structure in the food may become unacceptable. Some of the enzymes produced by spoilage bacteria may remain active, even when a thermal process has destroyed the causative microorganisms in the food.

The relationship between microbial numbers and food spoilage is complex and depends on the number, type, and activity of the microorganisms present, the type of food, the intrinsic properties of the food (e.g., pH, water activity, preservatives) and the extrinsic conditions (e.g., temperature of storage, atmosphere).

Spoilage is often most rapid and pronounced in proteinaceous chilled foods such as red meats, poultry, fish, shellfish, milk, and some dairy products. These products allow relatively good microbial growth as they are highly nutritious and have a high moisture content and relatively neutral pH value. However, care is needed to distinguish between those microorganisms present in spoiled food and those responsible for the spoilage defect (often called specific spoilage organisms), which may be only a fraction of the total microflora. Consequently, the relationship between sensory spoilage and microbial numbers is often only poorly correlated.

Spoilage Microorganisms

For this article, spoilage microorganisms have been arbitrarily divided into six categories: Gram-negative (oxidase positive) rod-shaped bacteria, coliform/enteric bacteria, Gram-positive spore-forming bacteria, lactic acid bacteria, other bacteria, and yeasts and molds.

Gram-negative (oxidase-positive) rod-shaped bacteria This group generally comprises the most common spoilage microorganisms of fresh, chilled proteinaceous products stored in air. The minimum growth temperatures are often −3 to 0 °C, and they grow relatively rapidly at 5–10 °C. Although they may represent only a small proportion of the initial microflora, they may grow and rapidly dominate the microflora of fresh, proteinaceous, chilled, stored foods.

Within this general group, the genus *Pseudomonas* is most common, although other genera include *Acinetobacter*, *Aeromonas*, *Alcaligenes*, *Alteromonas*, *Flavobacterium*, *Moraxella*, *Shewanella*, and *Vibrio* species. These microorganisms are common in the environment, particularly in water, and so may easily contaminate foods. Often, they may proliferate on inadequately cleaned surfaces of food processing plant or equipment.

These organisms may spoil products by the production of pigments, slime material on surfaces, and enzymes that result in food rots, off-flavors and off-odors. Some of the enzymes produced by *Pseudomonas* species are extremely heat-resistant and may produce long-term defects (e.g., rancidity or age gelation) in thermally processed (i.e., sterile) products.

Although this group is well adapted to grow at chill temperatures, it tends to be sensitive to other factors such as the presence of salt or preservatives, lack of oxygen, low pH (< 5.5) and a low water activity (< 0.98). Should these conditions prevail in a food, these bacteria will compete less well, and so other

microbial groups may cause spoilage. Overall, this group is not heat-resistant and is readily removed by mild thermal treatments. Consequently, their presence in heat-processed chilled foods is usually as a consequence of postprocess contamination.

Coliform/enteric bacteria This bacterial group also consists of Gram-negative rods, but these may be distinguished from the above group by a negative oxidase reaction. Traditionally, microbiologists have tended to examine for these groups separately, as their sources, significance, and factors affecting growth may differ. This group is frequently used as an indicator of postprocess contamination, inadequate processing or potential fecal contamination.

Compared with the Gram-negative (oxidase positive) rod-shaped bacteria, the coliform-enteric group is generally less well adapted to growth at temperatures of less than 5–10 °C, although some may grow at lower temperatures. However, they often dominate the flora of foods stored at temperatures of 8–15 °C and are less sensitive to changes in pH compared with the previous group. They are, however, generally sensitive to low water activity, preservatives, salt and thermal treatments. The coliform/enteric group does not necessarily require the presence of oxygen for growth.

This group may break down carbohydrates to give acids, which may result in souring of milk and other foods. Other types of spoilage include the production of pigmented growth, gases, slime, off-odors and off-flavors; these have often been described as 'grassy,' medicinal, unclean, and fecal.

Typical spoilage genera include *Citrobacter*, *Escherichia*, *Enterobacter*, *Hafnia*, *Klebsiella*, *Proteus*, and *Serratia*. These microorganisms are widely disseminated in the environment, including in animals. For animal products, poor slaughter and dressing practices may contribute to their presence in foods.

Gram-positive spore-forming bacteria These are of particular significance as they can produce heat-resistant spores that can survive many thermal processes that destroy vegetative cells, leaving these organisms to grow and dominate the microflora. The minimum growth temperatures are often quoted as 0–5 °C, although growth is frequently slow below 8 °C.

The genera of concern in this group are *Bacillus* and *Clostridium*. Again, these are common in the environment, and spores may survive for tens and possibly hundreds of years. The most common form of spoilage is the production of large quantities of gas that may result in pack or product blowing. The heat resistance of psychrotrophic strains is considered to be lower than that of mesophilic strains, but the former group may be of concern in chilled pasteurized foods.

Lactic acid bacteria At chill temperatures, lactic acid-producing bacteria grow more slowly than many of the bacterial groups mentioned previously. Consequently, they are only responsible for spoilage if growth of most other bacterial species is inhibited. This group is more tolerant of low pH than other spoilage bacteria and may multiply at pH values below 4.0. The lactic acid bacteria are also more resistant than the previously discussed spoilage bacteria to slight reductions in the water activity (a_w), and some strains are salt-tolerant. Lactic acid bacteria usually predominate in vacuum-packed products and in some modified-atmosphere-stored foods.

This bacterial group comprises both rod- and coccus-shaped Gram-positive bacteria, and typical genera include *Carnobacterium*, *Lactobacillus*, *Leuconostoc*, *Pediococcus*, and *Streptococcus*. Spoilage is generally by the production of acid, which results in souring, with or without concomitant gas production.

The presence of lactic acid-producing bacteria is not necessarily a problem, as some strains are deliberately added during the manufacture of some chilled foods (e.g., cheese, yogurts, and some salamis) and are essential for the development of the desired product characteristics. In addition, there is much interest in their potential use as a novel preservation system, as many produce natural antimicrobial compounds in addition to acids.

Other bacteria Depending on the food type and preservation system operating, other microorganisms also may cause problems in specific chilled foods. For example, *Brochothrix thermosphacta* is occasionally present on raw meats but does not normally create a spoilage problem. However, it can grow in atmospheres with a low oxygen level and/or high carbon dioxide concentration and so may cause problems in vacuum-packed or modified-atmosphere-packed meat products and produces an objectionable pungent 'cheesy' odor.

Micrococcus species are Gram-positive cocci that can grow in the presence of high salt concentrations. They grow more slowly than many other spoilage bacteria at chill temperatures but can cause souring and slime production when growth of competing organisms is inhibited.

Yeasts and molds Yeasts and molds generally grow more slowly than bacteria in foods, permitting good growth and so are generally out-competed.

Therefore, this group is seldom responsible for the spoilage of fresh proteinaceous foods. When the conditions in the food are altered to restrict bacterial growth, the role of yeasts and molds may become more significant. They are generally more resistant than bacteria to low pH, reduced a_w values, the presence of preservatives and low temperatures. Molds tend to require oxygen for growth, whereas many yeasts can grow in the presence or absence of oxygen. Most yeasts and molds are not heat-resistant and are readily destroyed by a thermal process, although some (e.g., *Byssochlamys*) may produce relatively heat-resistant ascospores that may survive food pasteurization treatments.

Air movements may be an important vector of transmission, especially with mold ascospores and so result in widespread contamination.

Typical spoilage yeasts include *Candida*, *Debaryomyces*, *Hansenula*, *Kluyveromyces*, *Rhodotorula*, *Saccharomyces*, *Torula*, and *Zygosaccharomyces* species. Molds that may be isolated from spoiled chilled foods include *Aspergillus*, *Cladosporium*, *Geotrichum*, *Mucor*, *Penicillium*, *Rhizopus*, and *Thamnidium* species. Fungal spoilage may be characterized by the production of highly visible, often pigmented, growth, slime, fermentation of sugars to form acid, gas or alcohol, and the development of off-odors and off-flavors, often described as yeasty, fruity, musty, rancid, and ammoniacal.

As with the lactic acid bacteria, yeasts and molds are sometimes deliberately added to food products (e.g., mold on ripened cheeses such as Brie).

Pathogenic Microorganisms

Foods may be considered to be microbiologically unsafe owing to the presence of microorganisms that may invade the body (e.g., *Salmonella*, *Listeria monocytogenes*, *E. coli* O157:H7 and *Campylobacter*) or those that produce a toxin that is ingested with the food (e.g., *Clostridium botulinum*, *Staphylococcus aureus*, and *Bacillus cereus*). The growth of pathogenic microorganisms in foods may not necessarily result in spoilage, and so the absence of deleterious sensory changes cannot be relied upon as an indicator of microbial safety. Furthermore, some toxins are resistant to heating and so may remain in a food after viable microorganisms have been removed.

In general, with human pathogens, the greater the number of cells consumed, the greater the chance of microbial invasion, as a larger number of cells may be able to evade/swamp the body's defense mechanism. Consequently, control, and preferably inhibition, of growth in foods is essential. However, with some invasive pathogens (e.g. *Campylobacter*), growth in the food may not be necessary as the infectious dose is low.

For discussion in this article, the pathogenic bacteria of concern for chilled foods can be arbitrarily divided as follows.

Microorganisms capable of growth at temperatures below 5 °C

These microorganisms are potentially of greatest concern as they continue to multiply, even with 'good' refrigeration temperatures. Although growth may continue, temperature control is critical, and the growth rate becomes increasingly slow as the temperature is reduced. In addition, temperature control can interact effectively with other factors (e.g., pH, salt, preservatives) to prevent or greatly limit growth.

Listeria monocytogenes A wide range of foods including meat, poultry, dairy products, seafoods and vegetables have been reported to be contaminated with *L. monocytogenes*. Although the total absence of *L. monocytogenes* from raw meats, poultry, and vegetables is difficult to ensure, the bacterium has been isolated from products that have undergone a thermal process designed to eliminate this bacterium. Such isolations are of concern, as many of these chilled foods may be consumed without further heating. The presence of *L. monocytogenes* on cooked foods suggests that postprocess contamination may have occurred. Several studies have shown that this bacterium has been isolated from a wide range of sites in several types of factory and may even be spread by poorly controlled cleaning procedures. Environmental control of *Listeria*, particularly in key areas of production (e.g., after cooking), is crucial to the prevention of product contamination.

The major concern with *L. monocytogenes* is its ability to grow at low temperatures, and a minimum growth temperature of −0.4 °C has been reported. Temperature control will, however, retard the rate of growth. Conversely, temperature abuse during storage of a food can exacerbate problems. *Listeria monocytogenes* is more resistant than many other vegetative bacteria to some, but not all, of the preservation mechanisms used in food manufacture (e.g., chilling, reduced water activity). Whilst resistance may be noted to these preservation systems when examined individually, foods are complex, and the combined effects of two or more preservation factors may effectively prevent growth.

L. monocytogenes is not considered to be a classically heat-resistant bacterium. It is generally accepted that conventional high-temperature, short-time milk pasteurization (71.7 °C/15 s) eliminates this

microorganism when freely suspended in milk. In other foods, a minimum process of 70 °C for 2 min (or the thermal equivalent) ensures the effective elimination of this bacterium.

The symptoms of disease are protean and range from a mild flu-like illness to meningitis, septicemia, stillbirths, and abortions. In general, the major symptoms of disease are restricted to the pregnant mother, fetus, elderly, and immunocompromised, and with these groups, the mortality level can be high.

Yersinia enterocolitica Outbreaks of disease caused by *Y. enterocolitica* are relatively rare, but have implicated milk, tofu, and chocolate milk. Although the reported incidence of *Y. enterocolitica* in gastrointestinal samples is generally low, it has increased, but this may be due not only to a true underlying increase but also to a greater awareness of this bacterium, recognition of symptoms, and improved methods. In some European countries, disease cause by *Y. enterocolitica* has surpassed that of *Shigella* and even rivals that of *Salmonella*.

A wide variety of foods have been reported to be contaminated with *Y. enterocolitica*, including many chilled products, i.e., raw and cooked meats, poultry, seafoods, milk, dairy products, and vegetables. Care is needed as the isolates responsible for disease generally belong to a few specific (bio-sero) types, whereas those from foods and the environment belong to a wide range of (bio-sero) types. Therefore, the pathogenic significance of food isolates should be ascertained before the food is considered as a health risk. The types responsible for human disease may be isolated from pigs and occasionally pork products.

The minimum reported growth temperature for *Y. enterocolitica* is −1.3 °C, and the bacterium grows relatively well at chill temperatures. Storage at refrigeration temperatures interacts with other preservation factors present in foods and help prevent growth of this bacterium. *Y. enterocolitica* is a heat-sensitive bacterium and can be readily eliminated from foods by heating. It has, however, been isolated from cooked meats, seafoods, and pasteurized dairy products, which is indicative of postprocess contamination. Greater attention is required for the environmental control of *Y. enterocolitica* in food-manufacturing establishments.

The symptoms of human yersiniosis are also variable. Overall, acute gastroenteritis is the most common symptom, particularly with children, and is characterized by diarrhea, abdominal pain, fever, and, less commonly, vomiting. With adolescents, abdominal pain may be localized and misdiagnosed as appendicitis. The mortality rate from human yersiniosis is low, and, other than cases involving appendectomies, the symptoms are generally self-limiting and rarely require treatment. In some cases, secondary symptoms may occur several weeks after the typical gastrointestinal symptoms disappear, e.g., postinfectious polyarthritis and erythema nodosum.

Aeromonas hydrophila The role of *A. hydrophila* as an agent of foodborne disease is still a matter of controversy. This bacterium, however, does possess many of the characteristics of other pathogenic bacteria. Incidents of foodborne disease implicating *A. hydrophila* have included oysters and prawns.

The minimum reported growth temperature for *A. hydrophila* is −0.1 to 1.2 °C. The bacterium is considered to be heat-sensitive and so may be readily eliminated from foods. Little has been published on the presence of *A. hydrophila* in the processing environment, but it is likely to be isolated particularly from wet areas.

Bacillus cereus The role of *B. cereus* as a spoilage bacterium of chilled foods is well recognized, and some strains may grow at temperatures as low as 1 °C. The minimum reported growth temperatures of pathogenic strains is usually considered to be 10–15 °C, although some isolates from outbreaks that involved chilled products have been able to grow and produce toxins at 4 °C. In addition, psychrotrophic, presumptively enterotoxigenic strains have been isolated from pasteurized milks and some cook–chill meats.

Bacillus cereus may be of particular significance in foods that have been heated or pasteurized, as the heat treatment may have eliminated other competitor microorganisms. During subsequent chilled storage, spores that survive the heat treatment may germinate and grow. Although relatively little published information is available, the heat resistance of psychrotrophic *B. cereus* (and other related species) is generally lower than that of the mesophilic strains.

Clostridium botulinum Human botulism is caused by the ingestion of a potent neurotoxin preformed in the food. Based on the antigenic analysis of this, seven types can be distinguished (named A–G). The strains responsible for disease can be divided into two main groups. First, mesophilic types A and some strains of B and F are proteolytic and so often cause putrefaction of foods if substantial growth occurs. Second, psychrotrophic types E and others of B and F are nonproteolytic, and so the consequences of growth and toxin production can occur with no discernible change to the food's appearance or smell. The

minimum growth temperature of the mesophilic proteolytic strains is considered to be 10 °C, and so these are of limited significance with chilled foods. Type E *C. botulinum* is able to grow and produce toxin after incubation at 3.3 °C for 32 days. It is now recognized that nonproteolytic strains of types B and F are also capable of growth and toxin production at 5 °C or less. Therefore, these nonproteolytic strains may grow, albeit slowly, in chilled foods.

'Sous vide' processing consists of packing foods under vacuum in impermeable sealed bags that are then heat-processed and stored chilled for extended periods. Whilst the time and temperature of cooking are specific to the food type, it will destroy vegetative microbial cells but may not be sufficient to destroy bacterial spores. The growth of nonproteolytic *C. botulinum* is of particular concern, because in the absence of air, spores may germinate and grow during storage if temperatures are not carefully controlled below the minimum for growth (3 °C).

It should be noted that the heat resistance of the psychrotrophic nonproteolytic strains is considerably lower than that of the mesophilic proteolytic strains. The risk of botulism from the former group can be minimized by the use of appropriate heating, controlled chilled storage and/or alterations in the product formulation to prevent growth.

Microorganisms capable of initiating growth at temperatures of 5–10 °C

These are a number of other pathogenic bacteria, which, although unable to grow at temperatures below 5 °C, may grow if temperature abuse occurs. These include *Salmonella* species, *Escherichia coli*, and *Staphylococcus aureus*, with generally accepted minimum growth temperatures of 5–7 °C. Even at temperatures up to 10 °C, the growth rate of these bacteria is generally slow, but these bacteria do, however, cause foodborne disease, frequently implicating chilled foods. Psychrotrophic strains of salmonellae have very occasionally been reported.

Several types of *E. coli* are well recognized as agents of foodborne disease. At present, the type of most concern is *E. coli* 0157:H7 and other verocytotoxigenic *E. coli*, which may produce severe hemorrhagic colitis. Limited growth of some strains may occur at 5–10 °C.

Whilst *Staphylococcus aureus* may grow at temperatures as low at 6 °C, disease is caused by the ingestion of a preformed toxin. The minimum temperature for toxin production is greater than for growth and is considered to be 10–14 °C.

Overall, the above bacterial species do not grow at temperatures below 5 °C, but may survive at these temperatures. Pathogens, and spoilage bacteria, often survive adverse conditions (e.g., low pH or high salt) better at refrigeration temperatures compared with higher temperatures. Therefore, if the infectious dose of the bacterium is low and/or growth of the pathogen has already occurred (e.g., during slow cooling), growth during chilled storage may not be a prerequisite for disease.

Microorganisms capable of initiating growth at temperatures above 10 °C

These species include mesophilic *C. botulinum*, mesophilic *B. cereus* and other *Bacillus* species, *C. perfringens* and *Campylobacter* species. In general, these do not grow below 10 °C, and growth is limited at temperatures between 10 and 15 °C.

Of particular concern in this bacterial group are the *Campylobacter* species that comprise the most commonly reported cause of gastrointestinal disease in many developed countries. Although many of the reported cases are sporadic, outbreaks have frequently implicated the consumption of raw milk and undercooked chicken. This bacterial group is unusual, as the minimum temperature for growth is 25–30 °C, and so it will not grow on most foods. However, as the infectious dose of the microorganism is very low, growth may not be necessary for disease to occur.

Whilst disease caused by the mesophilic spore-forming bacteria has implicated chilled foods, this is usually as a consequence of poor temperature control during cooling after cooking. These bacteria may grow extremely rapidly during a long slow cooling regime after cooking and then persist during chilled storage.

Conclusions

Chilled foods cover a wide range of commodities, containing a large number of ingredients. The number and types of microorganisms present are affected by the indigenous microflora, microorganisms contaminating before and after processing, the growth rates and abilities of the microorganisms present, the intrinsic properties of the food, the effects of processing and packaging, and the time and temperatures of storage. Consequently, the microbial safety and spoilage of chilled foods is very complex. Considerable efforts have been made by the food chain to minimize these effects through careful raw-material selection, product design and formulation, processing and packaging conditions, and good temperature control from processing through distribution, retail display, and even consumer storage. All of these have helped ensure the safety and quality of this important food sector.

See also: **Chilled Storage**: Principles; Attainment of Chilled Conditions; Quality and Economic Considerations; Microbiological Considerations; Use of Modified-atmosphere Packaging; Packaging Under Vacuum; **Chill Foods**: Effect of Modified-atmosphere Packaging on Food Quality; **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant; Overall Approach; Modern Systems; **Food Poisoning**: Classification; Tracing Origins and Testing; Statistics; Economic Implications; **Spoilage**: Bacterial Spoilage

Further Reading

Alcock SJ (1987) Growth characteristics of food-poisoning organisms at suboptimal temperatures. *Campden Food Preservation Research Association Technical Memorandum No. 440*.

Bell C and Kyriakides A (1998) Listeria: *A Practical Approach to the Organism and its Control in Foods*. London: Blackie Academic & Professional.

Betts GD and Everis L (2000) Shelf-life determination and challenge testing. In: Stringer MF and Dennis C (eds) *Chilled Foods: A Comprehensive Guide*, 2nd edn, pp. 259–286. Cambridge, UK: Woodhead.

Borch E, Kant-Muermans M-L and Blixt Y (1996) Bacterial spoilage of meat and cured meat products. *International Journal of Food Microbiology* 33: 103–120.

Cousin MA (1982) Presence and activity of psychrotrophic microorganisms in milk and dairy products: a review. *Journal of Food Protection* 45: 172–207.

Dainty RH (1996) Chemical/biochemical detection of spoilage. *International Journal of Food Microbiology* 33: 19–34.

Filtenborg O, Frisvad JC and Thrane U (1996) Moulds in food spoilage. *International Journal of Food Microbiology* 33: 85–102.

Gould GW and Russell NJ (1991) Major food-poisoning and food-spoilage microorganisms. In: Russell NJ and Gould GW (eds) *Food Preservatives*. London: Blackie Academic & Professional.

Huis in't Veld JHT (1996) Microbial and biochemical spoilage of foods: an overview. *International Journal of Food Microbiology* 33: 1–18.

Martens T (1999) *Harmonisation of Safety Criteria for Minimally Processed Foods: Rationale and Harmonisation Report*. *European Commission FAIR* CT96–1020.

Mitscherlich E and Marth EH (1984) *Microbial Survival in the Environment*. Berlin: Springer-Verlag.

Pitt JI and Hocking AD (1985) Spoilage of fresh and perishable foods. In: *Fungi and Food Spoilage*. Sydney: Academic Press, pp. 365–382.

Walker SJ and Betts GD (2000) Chilled foods microbiology. In: Stringer MF and Dennis C (eds) *Chilled Foods: A Comprehensive Guide*, 2nd edn., pp. 153–186. Cambridge, UK: Woodhead.

Whitfield FB (1998) Microbiology of food taints. *International Journal of Food Science and Technology* 33: 31–51.

Use of Modified-atmosphere Packaging

T R Robertson, Massey University, Palmerston North, New Zealand

Definition of Terms

'Modified atmosphere packaging' (MAP) is the term applied to different packaging technologies that use an internal atmosphere composition that has a different composition to air (20.6% oxygen/78% nitrogen) to increase the shelf-life of food products. The internal atmosphere is usually made up of mixtures of oxygen (O_2), carbon dioxide (CO_2) and nitrogen (N_2), with the possible addition of very small amounts of other gases such as carbon monoxide (CO), ethanol, sulfur dioxide (SO_2) and noble gases (e.g., argon). An alternative often used to describe MAP is 'gas packaging.'

'Controlled atmosphere packaging' is sometimes used incorrectly to describe modified atmosphere packages, as it is not possible to control the atmosphere inside a primary package (the package that is directly in contact with the product), but it could be used to describe the controlled storage conditions inside specialist shipping containers.

'Vacuum packaging (VP)' involves the 'complete' removal of air from the package, usually resulting in O_2 levels of below 1%. The product is placed in an oxygen barrier film, the air evacuated, and the product sealed. The evaluated pack collapses around the product. In the case of 'skin packaging' systems, the packaging material forms a second skin around the product. Skin packaging involves placing a product on to a substrate material such as a thermoformed plastic tray. A specialized polymer film is heated to the required softening temperature then draped over the product and tray, and a vacuum draws the film down and around the product and the tray to make a secure and attractive product. This packaging system has been used very successfully to retail shelled shellfish and marinated steak products.

'Active packaging' is the inclusion of specific additives into the packaging film or within the packaging to modify the headspace atmosphere and extend the product shelf-life. Active packaging systems include oxygen absorbents, carbon dioxide adsorbents or emitters, ethanol emitters, and ethylene absorbents. This is a developing technology and will gain more acceptance as system costs are reduced, although the use of oxygen absorbents is well established in Japan.

Although the benefits of MAP can be numerous, it is important that the microbiological safety considerations of using anaerobic conditions are understood. Anaerobic conditions can promote the growth of pathogens, and hurdles must be in place to insure the safety of the products. This is especially true of high-moisture foods, meat, fish, and ready-to-eat meals, where temperature control is critical to insure safe MAP foods. The use of good manufacturing practices, to insure that the packaged product has a low initial microbial count, and the use of MAP can provide a useful extension to the product shelf-life. The control of the temperature during the storage and distribution cool chain is equally important to MAP fruit and vegetables where temperature abuse can cause increased biochemical activity, thereby also increasing the rate of product-quality deterioration.

Principal Factors in MAP

The principal factors in a successful MAP operation are the selection of:

1. a gas or gas mixture to minimize the selected deteriorative reactions in the food to insure the required shelf-life;
2. a suitable packaging material to maintain the desired gas composition; and
3. an appropriate packaging machine.

For MAP to provide the stated shelf-life, the critical factor is control of the chill chain (refrigeration systems in the processing, storage, distribution, and retailing chain) to insure that no temperature abuse (i.e., temperatures outside the specified range for a particular product) occurs.

Choice of Gas

The choice of gas, or gas mixture, used to replace the air in the package requires knowledge of the nature of the food and its principal mode(s) of deterioration. Oxidative reactions and the growth of microorganisms are the two common deteriorative modes, and both are affected by the oxygen present in the package atmosphere. Reduction or removal of oxygen will reduce the oxidation reactions and prevent the growth of some microorganisms, whereas addition of CO_2 will inhibit the growth of other microorganisms. However some microbes, including some pathogens are not affected by high CO_2 and low O_2 levels and must be controlled by good manufacturing practices to insure low microbial numbers and temperature control.

Oxygen is a key component of oxidative degradation reactions occurring in food, combining readily with fats and oils causing rancidity, and other food components causing color, flavor, and odor changes. Oxygen is, however, a necessary requirement for normal respiration metabolism of whole and minimally processed fresh fruit and vegetables. Oxygen levels below the minimum required (as low as 1–2%) by the produce can result in anaerobic respiration and the development of off-flavors and aromas.

Carbon dioxide is present in the atmosphere in very low levels (~0.03%). It is highly soluble in water, and will be absorbed by the food until equilibrium is attained. For example, at 0 °C and a partial pressure of 101 kPa of CO_2, the solubility of CO_2 is 3.4 g per kilogram of water. The absorption of the CO_2 by the food in a sealed package creates a partial pressure difference between the inside and outside of the packaging, resulting in partial or complete collapse of the package around the food. This CO_2 absorption can also result in off-flavors, excess purge by muscle foods, and discoloration of fresh produce.

Nitrogen is used to purge air from a package to achieve sufficiently low levels of oxygen to prevent aerobic microbial spoilage. It is also used as a filler gas in MAP to reduce the concentration of other gases in the package and to keep the pack from collapsing as the CO_2 is absorbed by the product. The percentage of N_2 used in gas mixtures in combination with CO_2 often has to be determined experimentally to obtain the necessary positive effects of the CO_2 while maintaining the appearance of the pack with the level of N_2.

Packaging-material Selection

Most MAP relies on the correct choice of packaging material. A low water-vapor transmission rate and a high gas barrier are usually required to maintain the modified atmosphere, except for fresh produce where specific barrier properties are required to maintain the specified gas mixture, often using quite permeable materials. Most MAP are based on monolayer or multilayer thermoplastic polymers, and/or polymer trays to maintain the product and atmospheric integrity. High barrier multilayer films will often incorporate either metalized film or an aluminum foil layer, and the cost of the film usually increases with the increasing barrier properties. These films will still allow some gas transmission, even at chill temperatures. Thus, over the shelf-life of the MAP food, there will be movement of gases across the package. The movement of these gases can be mathematically modeled to allow accurate prediction of this gas movement, although there is a comparative lack of gas permeability data for thermoplastic polymers over the range of temperatures 0–40 °C and at the high humidity levels common in chilled food storage conditions.

The packaging material needs to have the necessary tensile strength to withstand machine handling, and subsequent storage, distribution, and retailing. The package must also be designed so that the end user can open, heat (if required), and consume the food easily.

Applications of MAP

Red Meats

Fresh red meat undergoes both oxidation, resulting in color and flavor changes, and microbial deterioration. The color of the meat is determined by the oxidation state of the pigment myoglobin. The oxygenated state, oxymyoglobin, is a bright red color associated by the consumer with fresh, high-quality red meat, whereas the reduced form, deoxymyoglobin, is a purple color. The surface of the meat is contaminated during the slaughtering process with microorganisms, and the presence of O_2 encourages the spoilage bacteria, especially the *Pseudomonas* species, which will produce putrid, unpleasant smelling compounds. The systems used for MAP of fresh meat can be classified into the following groups:

- VP involves the exclusion of air from the pack atmosphere, so the lack of oxygen prevents the oxidation deteriorative reactions and suppresses the growth of spoilage bacteria, especially the *Pseudomonas* species. The delay of these forms of deterioration through packaging significantly extends the shelf-life of chilled fresh meat. VP has been used for 2–9-kg primal cuts and subprimal cuts of meat for international and local distribution to retail outlets. These packs use high-gas-barrier films that often include shrink characteristics. The meat cut is placed into the bag, with any necessary protection to prevent the bone piercing the bag, vacuum-sealed, and then passed under a hot-water spray to shrink the bag tightly over the cut. Shrinking the bag produces a second skin effect on the product, eliminates folds and capillaries, improves seal strength and oxygen barrier, and reduces purge or drip from the meat. However, VP without the shrink will still provide a useful extension to fresh chilled meat shelf-life, for beef (10–12 weeks) and lamb (6–8 weeks).
- High-oxygen MAP systems have atmosphere gas mixtures of about 30% CO_2 and up to 70% O_2 to extend the color stability and delay the microbial spoilage of display packaged meat. This system is very suitable for centralized production of retail display meat packages where the oxygen insures that the meat has the required red color of fresh meat. These systems are not suitable for long storage times, where the high levels of oxygen can lead to off-odors and rancidity.
- Low-oxygen MAP includes mixtures of CO_2 and N_2, with less than 1% oxygen, and achieves a shelf-life similar to that of VP. One-hundred percent nitrogen flush can be used and has been shown to be as effective as VP with reduced purge.
- High-CO_2 MAP has very high levels of CO_2, and in some commercial operations includes 100% CO_2 atmosphere. The absence of O_2 retards the formation of brown metmyoglobin in the meat, and the CO_2 suppresses the growth of spoilage bacteria. However, because CO_2 is very soluble in both water and fat, especially at chiller temperatures, excess CO_2 must be added to the package to allow for solubility. Thus, the package must be overflushed with CO_2, then allowed to remain in the chiller for 24 h to allow for the CO_2 adsorption and pack-size reduction, before the final cartoning operation can be completed. These 100% CO_2 packages have resulted in a chilled shelf-life for lamb and beef of up to 16 weeks.

Poultry

Poultry are usually contaminated with large bacterial populations that consist primarily of spoilage bacteria but can include significant numbers of pathogens. Poultry has a higher pH compared with red meat and so provides an ideal environment for the growth of these bacteria. Vacuum packages are difficult to form around whole dressed birds because of their irregular shape. Thus, the shelf-life of vacuum-packed poultry is usually short, limited to about 2 weeks. There are now an increasing range of cut-up portions (e.g., breasts, wings, fillets) and other further processed products, both raw and cooked, available. High-CO_2 atmospheres will increase the product shelf-life. For chicken meat, 20% CO_2/70% N_2 can be as effective as 80% CO_2/20% N_2 mix over 1 week's storage at 2 °C and thus is very useful for retail packs. However, if a longer shelf-life is required, as in bulk packs, higher levels of CO_2 are needed. A storage life of up to 35 days has been achieved under certain closely controlled conditions, although commercially, 10–21 days at −1 to +2 °C would be expected using an 80–100% CO_2/0–20% N_2 mixture.

Fish and Shellfish

Fish encompasses a great diversity of species and habitats. Fish differ from terrestrial foods in that the

interior of their muscles is not sterile, they undergo rapid enigmatic breakdown of their proteins, even in low-oxygen environments, and they often grow in uncontrolled environments. Thus, fish are only as clean as the water from which they are taken. Thus, the commercial use of MAP to extend the shelf-life of fish has been limited by the potential of *Clostridium botulinum* nonproteolytic types B and E, *Vibrio parahaemolyticus*, and *Listeria monocytogenes* growth and toxin production in refrigerated MAP fish, without any evidence of spoilage.

The gas mixtures used vary according to the fish species, with low oxygen concentrations being used with fatty fish to suppress oxidative rancidity. Vacuum and MAP using high CO_2 have been used to extend the short shelf-life of various kinds of fish, e.g., gas mixtures of 30% O_2, 40% CO_2 and 30% N_2 for nonfatty fish and 40% CO_2 and 60% N_2 for smoked and fatty fish.

As inhibition of normal spoilage bacteria of fish is limited, extensions of shelf-life are not usually as dramatic as that which can be achieved with meat, and the need to maintain low temperatures throughout processing, storage, distribution, and retailing is not diminished. If the effort that is widely applied to controlling temperatures and turnover during production, distribution, and storage of MAP had been applied to earlier packaging systems, these may have provided similar shelf-life extensions as the MAP system.

Fruits and Vegetables

Fresh, whole, and minimally processed fruits and vegetables are living, respiring tissue, and therefore, under MAP conditions, care must the taken to insure that the gas mixture in the atmosphere does not become depleted of oxygen or contain too much CO_2. The increased shelf-life or storage time for produce comes with a combination of lower temperatures and lower O_2 levels, thereby slowing down the rate of respiration. Reduced respiration leads to a reduced depletion of carbohydrate reserves, hence less weight loss, slower ripening, and longer storage times. This suppression of the rate of respiration continues down to 2–4% O_2, depending on the produce species and storage temperature. At lower levels, the respiration becomes anaerobic, resulting in fermentation metabolism producing off-flavors, off-odors, and undesirable volatiles. The increasing CO_2 level in the gas mixture also helps suppress the rate of respiration in some produce. Reduced O_2 levels and increased CO_2 levels together can reduce respiration more than either alone. The lower limits for O_2 and the upper limits of CO_2 vary between different fruit and vegetable species and can vary within one produce species as a result of different growing and handling conditions. Temperature control of fruit and vegetable MAP is very important because the rate of change in the gas permeability of the polymeric films and the rate of respiration of the produce are different with the same temperature changes. The produce will use up the available oxygen very rapidly and start anaerobic respiration. To increase the complexity of fresh produce MAP, the selection of the packaging film is difficult as it has to release CO_2 at a rate similar to the production and let oxygen into the pack at the rate at which it is used by the produce. Thus, different produce will require a film with different gas transmission rates to insure the correct maintenance of the desired gas mixture. Despite almost 60 years of research and hundreds of publications reporting many experiments that have been conducted in this area, there is surprisingly little commercialization of MAP for horticultural produce. Current areas of research and commercialization include the development of models to aid the design of perforated film packaging to minimize weight loss (moisture loss) while controlling the development of a modified atmosphere. The use of high-oxygen 30–80-kPa atmospheres and noble gases, specifically argon modified atmosphere to extend the produce shelf-life. Current research indicates that exposure to superatmospheric O_2 concentration may stimulate, have no effect on, or reduce the rates of respiration and ethylene production, depending on the commodity, maturity and ripeness stage, O_2 concentration, storage time and temperature, and concentrations of CO_2 and C_2H_4 present in the atmosphere. The rapid growth of fresh-cut or minimally processed produce including retail salads, baby peeled carrots, apple slices, and many fruit and vegetables prepared for the service industry has been possible due to the improved quality and shelf-life of cut produce in MAP.

There are two methods of creating modified atmosphere (MA) conditions within packages of fruit and vegetables: passive MA and active MA. Passive MA involves sealing the produce in the bag and letting the product respiration and the film gas transmission rate develop the required gas mixture. Alternatively, the desired gas mixture can be added during bag sealing, and the respiration and film gas transmission rates will maintain this gas mixture. However, it is the produce respiration rate, package oxygen and carbon dioxide gas transmission rates, and storage temperature that determine the composition of the pack atmosphere, not the initial gas mixture used in the bag.

Ready-to-eat Meals, 'Meal Solutions,' Delicatessen Foods

Ready-to-eat meals, or prepared food products, are a rapidly increasing supermarket category involving chilled meals that only require the minimum of preparation, if any, before consumption. These products can be referred to as 'delicatessen' and include multicomponent food products. For a product to be eaten hot, the only preparation is heating, whereas for products like salads and sandwiches, no preparation is required. These ready-to-eat meals can range in quality from cheap meals, or snacks to expensive gourmet-style food dishes, and include pasta, pizza, precooked meats, sandwiches, precooked French fries, complete dishes, meat based dishes, pasta based dishes, and chilled bakery products, and includes 'sous vide' or cook/chill products. MAP is being used to increase the shelf-life of these products, along with control of the storage, distribution, and retail chill chain to prevent temperature abuse.

The shelf-life of these products is usually limited by two factors: microbial growth and oxygen sensitivity of the product. Thus, the packaging system used for ready-to-eat products should exclude O_2 and provide some microbial suppressing action, and usually involves a mixture of CO_2 and N_2. These atmospheres provide several advantages, including reduction of oxidative rancidity, lack of growth of aerobic spoilage organisms, suppression of mold growth by CO_2, minimal moisture loss through the package, and reduced oxidative breakdown of flavor and aroma volatiles.

Prepackaged sandwiches have grown in popularity over the last decade, and their sale has extended from large retail outlets to snack bars, lunch bars, and petrol service stations. An increase in shelf-life would be of great advantage in allowing more flexibility in the distribution of these products. The hygiene and safety of these MAP products during manufacture and the use of good manufacturing practices, combined with the use of good refrigeration practice, are critical. Typical gas compositions include 30% CO_2/70% N_2, 5% CO_2/95% N_2 or a minimum of 70% CO_2. The higher rate of of CO_2 can cause pack collapse, as the gas is absorbed by the product, hence the use of higher levels of N_2. The specific combination used will be a compromise, delaying microbial growth without causing collapse of the packs.

'Sous Vide' involves vacuum packaging of a prepared food in a multilayer composite polymer film pouch, cooking the packaged product, followed by rapid cooling, and then chilled storage. These products can have a shelf-life of 2–3 weeks under optimum chiller conditions. An alternative style of high-quality ready-to-eat meals involves precooking the individual food components and then assembling these components into the package before gas packaging with the desired atmosphere of CO_2/N_2 and chilling.

Modified atmosphere packaging for fresh pasta products is common. Typical gas compositions include 100% N_2 or 70–80% CO_2 and 20–30% N_2 mixtures resulting in a shelf-life, if stored under 4 °C, of up to 4 weeks.

See also: ***Clostridium***: Occurrence of *Clostridium botulinum*; **Contamination of Food**; **Convenience Foods**; **Fish**: Spoilage of Seafood; **Meat**: Preservation; **Oxidation of Food Components**; **Packaging**: Packaging of Liquids; Packaging of Solids; Aseptic Filling; **Spoilage**: Bacterial Spoilage

Further Reading

Barry-Ryan C and O'Briene (1998) Novel high oxygen and nobel gas modified atmosphere packaging for extending the quality and shelf-life of fresh prepared produce. *IIR Proceedings Series 'Refrigeration Science and Technology'* 6: 417–424.

Blakistone BA (1998) *Principles and Applications of Modified Atmosphere Packaging of Foods*, 2nd edn. London: Blackie Academic & Professional.

Brody AL (ed.) (1989) *Controlled/Modified Atmosphere/ Vacuum Packaging of Foods*. Trumball, CT: Food and Nutrition Press.

Farber JM and Dodds KL (eds) (1995) *Principles of Modified-Atmosphere and Sous Vide Product Packaging*. Lancaster, PA: Technomic.

Gill CO (1990) Controlled atmosphere packaging of chilled meat. *Food Control* 2: 74–78.

Inns R (1987) Modified atmosphere packaging. In: Paine FA (ed.) *Modern Processing, Packaging and Distribution Systems for Food*. Glasgow: Blackie & Son.

Kader AA and Ben-Yehoshua S (2000) Effects of superatmospheric oxygen levels on postharvest physiology and quality of fresh fruits and vegetables. *Postharvest Biology and Technology* 2: 1–13.

Kader AA, Zagory D and Kerbel (1988) Modified atmosphere packaging of fruits and vegetables. *CRC Critical Review of Food Science and Nutrition* 28(1): 1–30.

Matz SA (1989) *Bakery Technology: Packaging, Nutrition, Product Development, Quality Assurance*. Barking, UK: Elsevier Science.

Robertson GL (1993) *Food Packaging: Principles and Practice*. New York: Marcel Dekker.

Rooney ML (1995) *Active Food Packaging*. London: Blackie Academic & Professional.

Zagory D (1997) Modified atmosphere packaging. In: Brody AL and Marsh KS (eds) *The Wiley Encyclopedia of Packaging Technology*. New York: Wiley.

Packaging Under Vacuum

M Rossi, Packaging Technical Center, Sealed Air s.r.l., Passirana di Rho, Milano, Italy

The Role of Oxygen in Food Spoilage

The presence of oxygen is one of the major factors of spoilage of foods and causes:

1. oxidation reactions, damaging vitamins, fatty substances, pigments, flavoring substances that are often catalyzed by enzymes;
2. growth and activity of aerobic microorganisms (aerobic bacteria, yeasts, molds). (*See* **Oxidation of Food Components**.)

It is therefore essential, in order to prolong the freshness of foodstuffs, to eliminate the presence of oxygen in contact with the foodstuff itself and to prevent further access of oxygen during storage. Vacuum packaging is one of the simplest and most widely used systems to achieve this objective. This type of packaging has enjoyed increasing success since the late 1950s, in parallel with the development of the technology of plastics owing also to the changes in the distribution chain of perishable foods requiring increased hygiene and storage life.

This article outlines the principles of vacuum-packaging technology and the features of the packages utilized, and describes the main fields of application of vacuum packaging to chilled foods.

Vacuum Packaging: A Definition

'Vacuum packaging' is a term improperly but commonly used to define a packaging system that implies the reduction of the partial pressure of atmospheric gases (oxygen being 20% of them) inside a package. A vacuum inside a package can be obtained mainly through two systems:

- steam flushing of the headspace;
- sucking of air from the package headspace by means of equipment (vacuum chamber, nozzle) based on a vacuum pump.

The former system is mainly utilized in hot packaging of shelf-stable products, which are generally packaged in rigid containers. The elimination of air is achieved by means of a steam flush that replaces atmospheric gases inside the package; steam condensation in the package headspace after chilling reduces the inner gaseous pressure. The latter system is commonly utilized for packaging of perishable foodstuffs that have to be stored in chilled conditions. In this case, packaging equipment based on a vacuum pump is generally utilized in combination with flexible packages, which, after evacuation, are closed hermetically (by means of a seal or sometimes a tight clip) and collapse on the packaged product once the package is exposed to atmospheric pressure. Depending on the type of equipment employed, a residual pressure of atmospheric gases as low as 500 Pa can be obtained in the package.

Figure 1 illustrates the main systems utilized for evacuating flexible packages:

1. nozzle (Figure 1a);
2. single vacuum chamber (Figure 1b);
3. divided vacuum chamber (Figure 1c).

Packaging Materials Utilized for Vacuum Packaging

Plastics are the main raw material utilized for manufacturing the flexible materials employed in vacuum packaging of chilled foods. These materials must have the following properties:

- flexibility;
- mechanical resistance to various forms of abuse (abrasion, puncture, flex cracking);
- gas-barrier properties adequate to the application requirements (generally expressed as gas permeation rates);
- thermosealability or clippability;
- good optics (haze and gloss);
- printability;
- suitable dimensional behavior (dimensional stability to heat, formability after heating, shrinkability after heating).

To combine and balance to the required level the above properties, it is often necessary to mix different types of individual materials. Different resins can be mixed together (resin blends), resin additives (such as plasticizers, pigments, stabilizers, slip agents, etc.) can be used, and, more commonly, a multilayer material is produced, each layer being composed of a distinct individual material that imparts its properties to the overall structure.

Individual components most commonly used in manufacturing flexible materials utilized for vacuum packaging are listed below, together with their abbreviations and main properties:

- polyethylene (PE): sealability, formability, moisture barrier, low cost;
- polypropylene (PP): moisture barrier, thermal resistance, dimensional stability;

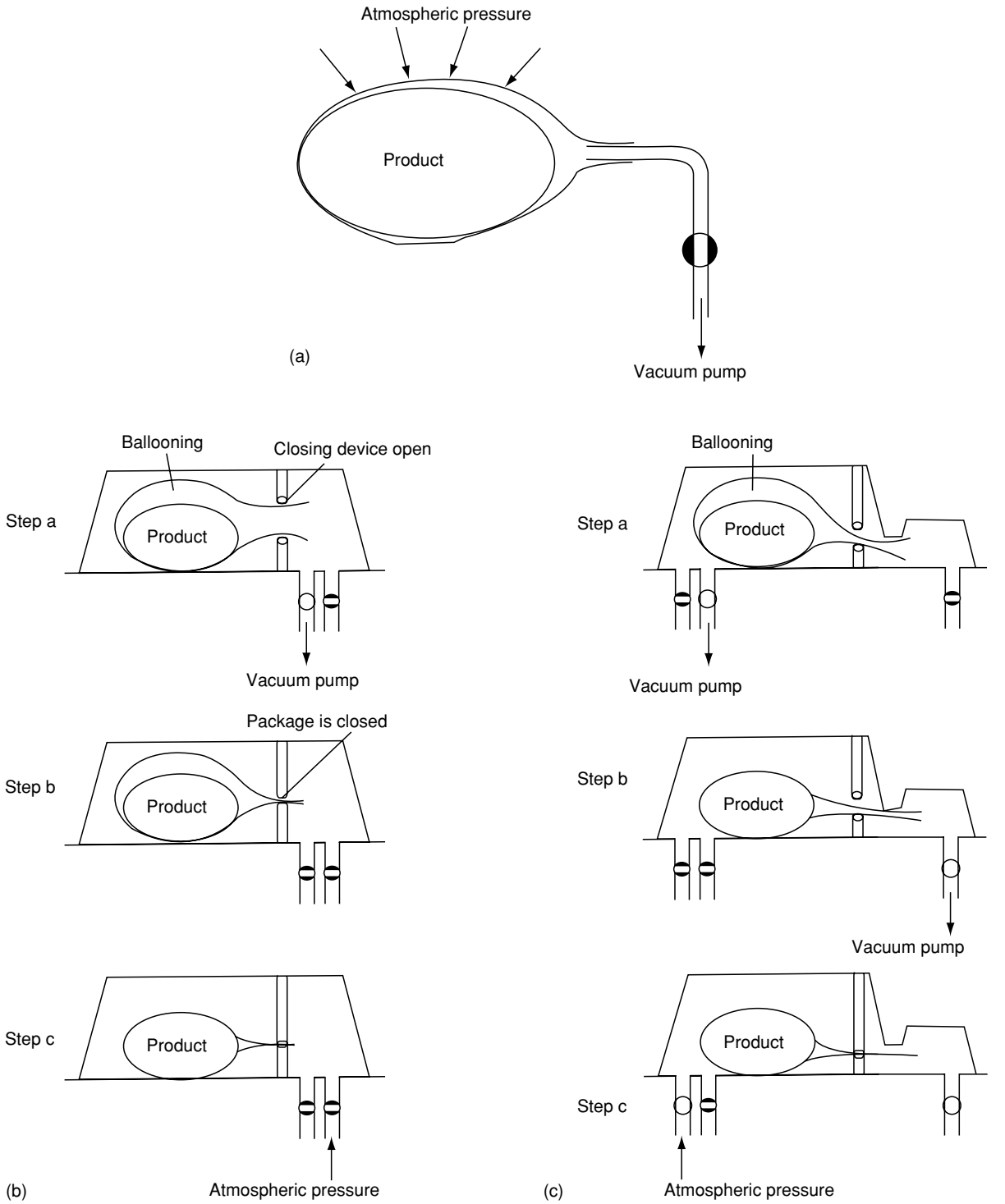

Figure 1 Main systems utilized for obtaining vacuum packages. (a) Nozzle system. Air is extracted from the package (a bag or a pouch) through a nozzle, and then the package is closed. This is the simplest way of extracting the air, but it does not allow high levels of vacuum in the package. (b) Most common system utilized for evacuating bags, pouches, and thermoformed packages. (c) Divided vacuum chamber. This allows a better evacuation of the package headspace by using two separate chambers where a vacuum is applied sequentially. Reproduced from Chilled Storage: Packaging Under Vacuum, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

- ethylene–vinylacetate copolymer (EVA): easy sealability, thermal shrink properties;
- ionomers: mechanical strength, easy sealability, grease resistance, formability;
- polyamides (PA): mechanical strength, gas barrier, formability;
- polyesters (PET): mechanical resistance, heat resistance, gas barrier;
- ethylene–vinylalcohol copolymer (EVOH): gas barrier, easy processability in coextrusion;
- polyvinylidenechloride (PVDC): gas barrier, grease barrier.

PE, PP, EVA, and ionomers are classified in the wide family of resins called polyolefins. Aluminum foils and vacuum-metallized plastic films are also utilized because of the excellent gas-barrier properties of aluminum.

Multilayer materials are manufactured using various techniques:

1. coextrusion: the molten resins are combined in the final structure by extruding them through a round or flat extrusion die, which keeps them separate in discrete layers;
2. lamination: previously extruded plastic films and, sometimes, aluminum foil, are joined together by means of glues (glue lamination) or with resins that have adhesive properties (extrusion lamination);
3. coating: pre-extruded films are coated with a layer of molten or dissolved resin (a latex).

Package Configurations Used for Vacuum Packaging of Chilled Foods

These can be classified into four main categories: shrink bags such as Cryovac*, pouches, thermoformed packages, and skin packages such as Darfresh*.

Shrink Bags

Shrink bags are available in the form of premade bags that can be prepared in different packages (e.g., taped bags) to allow their utilization on automatic equipment. The main feature of these bags is their ability to shrink when exposed for a short time to heat (for instance, a few seconds at 90 °C). This behavior is the consequence of treatment imparted to the packaging material during its production (oriented polymeric chains that retain a built-in tension, making them able to shrink when relaxed by heating).

Shrinking increases the packaging material thickness (which varies between 40 and 120 μm), improving mechanical resistance and gas-barrier properties, determines a tighter package around the product, limiting dripping out of juices in moist products such as meat, and improves the appearance by eliminating excess of packaging material around the product. Shrink bags, at the onset of their introduction on the market in the 1950s, were monolayer PVDC materials, but subsequent technological evolution gave rise to complex coextruded multilayer structures having polyolefins as the main components and gas-barrier layers made of PVDC or EVOH.

Some materials are electronically cross-linked to improve mechanical properties.

Shrink bags are used for vacuum packaging of industrial units of fresh meat, processed meat and cheese, and for consumer units of processed meat and cheese.

Pouches

Pouches can be utilized in two forms:

- premade plastic envelopes of different sizes that are loaded with the product, then evacuated and sealed in vacuum chambers;
- in-line prepared pouches from machines using rollstock material (horizontal form fill seal and vertical form fill seal machines).

The above machines, starting from a flat web, produce a film tubing that can be either horizontal or vertical (hence the different definitions), which is filled with the product, sealed transversely, and cut into final packages.

The machines can be equipped with a vacuum chamber to evacuate the pouch before final sealing. Form fill seal machines are widely employed in food packaging; however, their utilization for vacuum packaging of perishable foods is rather limited as thermoforming is preferred whenever a high packaging output and automation of packaging operation are required.

Premade pouches are commonly utilized for vacuum packaging of industrial units of fresh red meat such as whole primal cuts, processed meat (ham, bacon, salami, bologna) and cheese. Typical structures of materials for pouches, obtained by means of glue or extrusion lamination, are based on bioriented PA6 or PET (biorientation plus heat setting enhances dimensional stability and mechanical resistance), which can be coated with PVDC to increase the gas-barrier properties and are laminated to suitable sealing layers (PE, EVA, or PP and ionomers when heat resistance for pasteurization or cooking in the package is requested). For long storage-life

Cryovac* and Darfresh* are registered trademarks of Cryovac Inc., a subsidiary of Sealed Air Corporation.

applications (pasteurized cooked ham or sausages) where maximum gas-barrier properties are needed, aluminum foil is also used.

The total thickness of the above materials varies between 70 and 250 μm, the higher thickness being employed for heavy and hard products.

Thermoformed Packages

These are obtained on continuous thermoforming machines, which use rollstock materials. Two rolls are used, one for the bottom web, which is unwound, heated by a warm plate and formed into cavities where the product to be packaged is subsequently loaded, and a second one, the top web, which is sealed onto the bottom web inside a vacuum chamber from which the air has been removed. Thermoforming is widely applied to packaging of perishable foodstuffs because of the flexibility of the process, the packaging speed, and the ease of automation.

A wide variety of packaging materials is utilized. Bottom webs are generally based on PA laminated or coextruded with all types of polyolefins. PVDC or EVOH are used when high gas-barrier properties are needed. Some bottom webs are also heat-shrinkable after thermoforming.

Top webs are generally laminated structures similar to those used for pouches, the thickness of which seldom exceeds 100 μm. Metallized materials are often used in the top web formulation.

Thermoforming is utilized for packaging many kinds of perishable products, both consumer and industrial units, including whole ham and fresh meat primal cuts, with the exception of the biggest units of cheese and processed meat.

Skin Packages

Skin packaging represents an evolution of thermoforming where the top web is softened by means of heat and subsequently formed onto the packaged product.

The bottom web can be either flexible or rigid and can be preformed into a tray or a product-sized cavity into which the product to be packaged is loaded.

In skin packaging, the top film conforms tightly to the product shape with advantages in terms of appearance, limited product crushing due to vacuum, and the possibility of contour sealing around the product, limiting the exudation of liquids from the product. Because of its appealing appearance and the high cost of the packaging materials, skin-packaging utilization is limited to consumer or family units of meat (both fresh and processed), fish, prepared meals, and cheese.

Top webs are usually based on ionomeric resins or on coextruded structures of polyolefins. PVDC or EVOH provides the necessary gas barrier.

Bottom webs are generally either laminated or coextruded structures similar to those used for thermoforming.

When a rigid tray is required, polyvinylchloride (PVC), polystyrene (PS) or PET is employed.

Influence of Vacuum Packaging on the Storage Behavior of Chilled Foods

Fresh Meats and Poultry

Meats contain an abundance of nutrients necessary for the growth of microorganisms; they are particularly rich in soluble organic substances such as carbohydrates, amino acids and nucleotides. Therefore, spoilage of fresh meat and poultry during chilled storage is mainly due to the growth of aerobic psychrophyllic bacteria belonging to the Pseudomonadaceae family and to the *Moraxella–Acinetobacter* group. Growth of these bacteria results in development of off-odors (due to hydrogen sulfide, ammonia, amines, and indole) and bacterial slimes, which contribute to meat discoloration. (*See* **Spoilage**: Bacterial Spoilage.)

Vacuum packaging in oxygen-impermeable materials limits oxygen supply to the typical aerobic spoilage microflora, providing conditions suitable only for the slower-growing lactic acid bacteria, which, in chilled conditions, require several weeks to produce off-odors. (*See* **Lactic Acid Bacteria**.)

In addition, vacuum packaging has an impact on meat color, which is mainly due to the presence in the muscle tissue of myoglobin, a conjugated protein where the protein moiety (globin) is bonded to a heme group.

The iron atom of the hematin nucleus can form complexes with different ligands and can be in either the ferrous (Fe^{2+}) or ferric (Fe^{3+}) oxidation state. The globin can be in either the native or denatured state. Among the different complexes of heme, globin and ligands, three are important in fresh meat.

- oxymyoglobin, with oxygen as the ligand and iron in the ferrous state, characterized by a bright red color;
- myoglobin, with water as the ligand and iron in the ferrous state, characterized by a purplish red color;
- metmyoglobin, with water as the ligand and iron in the ferric state, characterized by a brown color.

Oxymyoglobin is the pigment normally present on the surface of meat exposed to air, and gives the meat its bright and attractive color.

During storage, as a consequence of the growth of aerobic bacteria that reduce the availability of oxygen on the meat surface, and of the reduction in the capability of the meat to reduce its own metmyoglobin level, metmyoglobin tends to predominate, imparting its brown color to the meat surface, which contributes to consumer rejection of the product.

In vacuum-packaged meat, as a consequence of preventing oxygen access to the meat surface, the pigment is of the myoglobin color, and the meat appears darker than a sample exposed to air or packaged in modified atmospheres. Displaying of vacuum-packaged consumer units of fresh meat can create problems of consumer acceptance because of the purplish color of the meat surface. In this case, proper consumer warning has to be given to explain the origin of the color and the advantages of vacuum packaging in terms of prolonged storage life. (*See* **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs.)

The storage life at 0–2 °C of fresh primal meat vacuum packaged in bags or pouches is 4–8 weeks, allowing full aging of meat types such as beef that require maturation. Consumer units in the form of meat slices have a storage life of 2–3 weeks.

Processed Meats

It is useful to classify the many existing types of processed meats into two main categories of products:

- products having high-activity water, the production processes of which often imply cooking (frankfurters, patties, bologna, fresh sausages, cooked ham, luncheon meats);
- cured products with low-activity water (raw ham, salami, bacon).

The spoilage of the former category of products can be due to surface drying, fat oxidation, discoloration or to microbiological factors (sliminess, souring, greening). The latter category of products is less susceptible to spoilage caused by bacteria. Rancidity, discoloration and mold growth are the most common limiting factors for storage life. (*See* **Meat**: Structure; **Spoilage**: Molds in Spoilage.)

All the above products can benefit from vacuum packaging, which avoids surface drying, mold growth and greatly slows down the oxidation of fat and pigments. Pasteurizable packages are also commonly used for microbial stabilization of some products (cooked ham, patties, and frankfurters). Furthermore, cook-in-the-package technology has been developed for production of cooked ham, allowing optimization of the production process, hygienic quality and yield.

The chilled storage life of vacuum-packaged processed meats is very variable, depending primarily upon product composition and packaging material characteristics and can vary from a couple of weeks (certain types of fresh sausages) to several months (cured products and pasteurized or cook-in-the-package products).

Cheese

Vacuum packaging is mainly applied to hard and semihard types of cheese, for both industrial and consumer units, and its effect in storage-life extension is due mainly to:

- elimination of surface drying;
- slowing down of fungal growth;
- limitation of oxidation of fatty substances.

An interesting application of vacuum packaging is the curing in the package technology that is utilized for certain types of cheeses (Emmenthal, Gouda, Edam). The cheese is vacuum-packaged at an early curing stage and cured inside the package. This technique improves yield (limiting rind formation and water loss) and allows a certain product standardization. Curing in the package is particularly demanding in terms of gas transmission properties of the packaging material; a good compromise is necessary between a sufficiently high carbon dioxide permeability to allow the escape of gas formed inside the cheese during curing and a sufficiently low oxygen transmission rate to avoid mold growth. The gas-transmission rate requirements can vary from cheese type to cheese type and also depend upon production conditions, which are often variable among different dairies. (*See* **Cheeses**: Cheeses with 'Eyes'.)

Fish

Wet fish is probably the most perishable type of foodstuff because of the high content of soluble substances in the flesh, many of which contain nitrogen, and triglycerides characterized by polyunsaturated fatty acids (in fatty types of fish such as clupeids, salmonids and scombroids). In addition, fish flesh is characterized by a high level of enzymatic activity, which plays a major role in the early stages of spoilage. (*See* **Fish**: Spoilage of Seafood.)

Vacuum packaging retards the growth of aerobic spoilage bacteria and limits the oxidation of fat. However, vacuum-packaged fish is very perishable and has to be carefully stored at temperatures below 2 °C, the normal storage life being 5–7 days. Prepackaged consumer units of wet fish, utilizing skin packages, have recently been introduced at a commercial level due to the need for a distribution system capable of handling in a relatively easy way a product

characterized by bad smells, dripping out, and hygienic and preparation problems at the consumer level. For processed fish (smoked, salted, pickled), vacuum packaging is commonly utilized because it limits the spoilage factors of this kind of product (mainly fat oxidation and mold growth). The storage life of vacuum-packaged processed fish depends upon the water activity level, ranging between 2 and 3 weeks for slightly salted fish and a few months for highly salted fish.

Prepared Foods

Vacuum packaging is also utilized for a wide range of prepared foods (delicatessen products, cooked meats and poultry, prepared meals, salads).

The spoilage mechanism and perishability of these products are related to their chemical composition (pH, additives) and heat treatment (pasteurization), and therefore, their storage life is very variable, ranging from 2 weeks to several months.

An interesting application of vacuum packaging to prepared foods is the cook-in-the-package technique, which implies packaging under vacuum of raw or partially cooked foods that are cooked inside the package, pasteurized (when required), chilled and stored, and reheated and unpackaged at the time of consumption. This technology allows the rationalization of meal preparation in central units (e.g., in institutional kitchens and restaurant chains) because of the storage life of the prepared meal (1 week or more) and it has also been introduced to food-processing plants to prepare meals to be distributed chilled at the retail level.

See also: **Cheeses**: Cheeses with 'Eyes'; **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Lactic Acid Bacteria**; **Meat**: Structure; **Oxidation of Food Components**; **Spoilage**: Bacterial Spoilage; Molds in Spoilage

Further Reading

Brody AL (ed.) (1989) *Controlled/Modified Atmosphere/ Vacuum Packaging of Foods*. Trumball, CT: Food and Nutrition Press.

Farber JM and Dodds KL (ed.) (1995) *Principles of Modified Atmosphere and Sous Vide Product Packaging*. Lancaster, PA: Technomic.

Kadoya T (ed.) (1990) *Food Packaging*. San Diego, CA: Academic Press.

Mathlouthi M (1994) *Food Packaging and Preservation*. London: Blackie Academic & Professional.

Renerre M and Labadie J (1993) Fresh Meat Packaging and Meat Quality. Proceedings of the 39th International Congress of Meat Science and Technology. Calgara.

Robertson GL (1992) *Food Packaging – Principles and Practice*. New York: Marcel Dekker.

CHLOROPHYL

H G Daood, Central Food Research Institute, Budapest, Hungary

Introduction

The green color in nature results from the presence of chlorophyll pigments that occur in the chloroplasts of plant tissues. The chloroplasts hold chlorophylls close to the cell wall and carry out photosynthesis on biomembranes. In coordination with other pigments like carotenoids, chlorophylls play an important role in the metabolism of light energy and catalyze synthesis of carbohydrates, the key metabolites of the metabolism of living cells.

Because of their widespread occurrence in a variety of plant products, chlorophylls play a vital role in the acceptability of food commodities. Chlorophyll-containing products have the potential to be used as a natural food colorant and nutritional supplement. The changes in chlorophyll content and composition can be used as an indication for some physiological process and orders. Recently, biological studies have postulated the anticarcinogenic and antimutagenetic effect of chlorophyll-related compounds. This article covers topics concerning biosynthesis, biochemical conversion, response to technological factors and the antioxidant–oxidant role of food chlorophylls.

Biosynthesis

Chlorophylls in plant foods are synthesized from δ-aminolevulinic acid (ALA), whose role is demonstrated in the biosynthesis of the tetrapyrrole nucleus. Once ALA is formed, two molecules condense to form porphobilinogen (PBG) by converting an aliphatic compound into an aromatic one. The head-to-tail condensation of four molecules of PBG results in the formation of the first tetrapyrrole intermediate,

a linear hydroxymethylbilane porphyrin. This linear molecule is enzymatically closed to form the first cyclic tetrapyrrole, uroporphyrinogen III. By the action of decarboxylase-catalyzed decarboxylation of the acetic acid groups on the pyrrole rings A, B, C, and D, uroporphyrinogen III is converted to coproporphyrinogen III. Propionic acid group oxidation and aromatization of the oxidized intermediate forming protoporphyrin IX follow the last step.

The next phase of chlorophyll biosynthesis starts with the chelation of protoporphyrin IX, mediated by Mg chelatase. This is followed by methylation of one of the propionic acid residues to form Mg-protoporphyrin-*n*-monomethyl ester (Figure 1). The last phase is characterized by the conversion of Mg-protoporpyrin-Me to protochlorophyllide and the protochlorophyllide to chlorophyllide through reduction of the vinyl substitutuent in the side chain of the pyrrole ring B, followed by oxidation and photoreduction of the D ring. The final step is esterification of the propionate substituent on pyrrole ring D with geranyl geraniol, which is then reduced to phytyl.

The first enzyme involved in tetrapyrrole synthesis, aminolevulinic acid synthetase, regulates chlorophyll biosynthesis. During maturation of some plant seeds chlorophyll biosynthesis may be stimulated by ethylene in the dark.

Concerning biosynthesis of chlorophyll b, old and more recent investigations gave convincing evidence that chlorophyll b is formed from chlorophyll a. The kinetics of ^{14}C pulse labeling indicate that chlorophyll a is the precursor of chlorophyll b. In light pulse experiments, chlorophyll b formation was shown to take place stoiciometrically at the expense of chlorophyll(ide) a. In homogenate of plant leaves, in the presence of NADP+, labelled chlorophyll b is formed from [^{14}C]chlorophyll a. The label locates in the tetrapyrrole and the phytol portion of the chlorophyll b, ruling out the possibility that the only carbon transfer is by transphytylation reaction.

Biochemical conversion

Degradation of chlorophyll usually happens all the way from ripeness to processing and storage of plant-derived foods. Although the mechanism of degradation is fragmentary, chlorophylls are degraded to colorless products.

Initially, chlorophylls are degraded to the phytyl-free chlorophyllides by the enzyme chlorophyllase

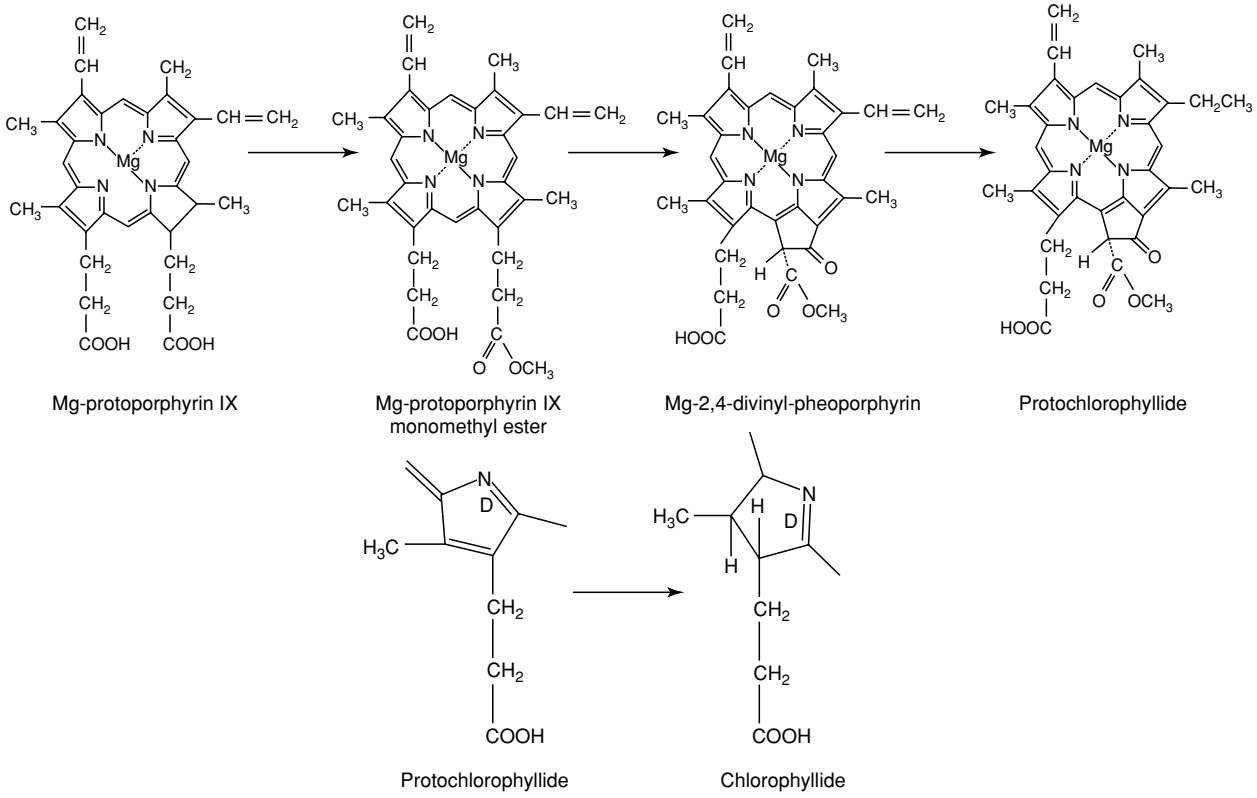

Figure 1 Last steps of chlorophyll biosynthesis in higher plants. Adapted from Castelfranco PA and Beale SI (1981) Chlorophyll biosynthesis. In: Hatch MD and Boardman NK (eds) *The Biochemistry of Plants*. New York: Academic Press.

(chlorophyll chlorophyllide hydrolase, EC3.1.1.14). The enzyme is located in the lipid envelope of the thylakoid membranes as an intrinsic membrane glycoprotein. Although chlorophyllase activity increases during ripening of some fruits and vegetables and parallels the respiratory climacteric, it is not affected by the endogenous ethylene. The enzyme activity remains even during storage and after processing of some plant products such as green tea. Generally, enzyme activity reaches a maximum at both the beginning and end of the vegetative growth phase.

Small amounts of chlorophyllide a and b can be found during the initial growth period, which coincides with a phase of great chlorophyll synthesis. Later, the synthetic and degeneration mechanisms may overlap, making the detection of phytyl-free compounds impossible. In ripening fruits and vegetables, the presence of allomerized chlorophyll derivatives suggests that, in addition to dephytylation by chlorophyllase, chlorophylls can be degraded by other oxidative systems involving active oxygen.

A chlorophyllide a-degrading enzyme is present in green tissues. It catalyzes the breakdown of chlorophyllide a in the presence of H_2O_2 and 2,4-dichlorophenol.

Biochemical conversion of chlorophylls to pheophytins or pheophorbides is initiated by the coordinating function of Mg-dechelatase (the enzyme catalyzes dechelation of Mg ion from chlorophyllins to form different pheophytins) and decarbomethoxylase. It is evident from the occurrence of 13-2-hydroxy-chlorophylls, hydroxy-lactone-chlorophylls, and unknown intermediates in fresh or stored green-colored plant products that more than one enzyme system is involved in the chlorophyll degradation mechanism.

Of the biological oxidants, lipoxygenase (EC1.12.13.11) contributes to the oxidative degradation of chlorophyll. The enzyme is well known to have pigment-bleaching activity of its isoenzyme-1 and isoenzyme-2, with the former being more active. Lipoxygenase cooxidizes chlorophylls through the oxidation of 1,2-pentadiene containing unsaturated fatty acids and results in the rapid formation of free radicals and, in the presence of molecular oxygen, in the creation of very reactive peroxy radicals. The results of many biochemical studies confirm that the chlorophyll-bleaching reaction requires an intermediate formed during peroxidation of polyunsaturated fatty acids by the isoenzyme. This type of degradation is responsible for the low storage stability of a variety of plant products. Furthermore, lipoxygenase-catalyzed chlorophyll bleaching is a characteristic change in some legumes and cereal products.

Due to various bioconversions of chlorophylls, the dephytylated polar intermediates, chlorophyllides, pheophorbides, and probably hydroxy chlorophylls appear and disappear during the ripening and storage of food commodities. The results of several investigations suggest that chlorophyll is first degraded to chlorophyllide by chlorophyllase, followed by oxidative degradation by peroxidase, lipoxygenase, or chlorophyll oxidase. The simultaneous action of dechelase may raise the diversity of chlorophyll extract by producing the corresponding Mg-free derivatives of the aforementioned chlorophyll degradation products. It should be noted that further degradation of chlorophylls and their derivatives to colorless compounds of low molecular weight is a function of oxidation processes which is not yet well described.

Degradation during Processing

Except for enzymatic conversions, chlorophylls undergo light-, heat-, and acid-catalyzed alterations that lead to a marked shift in the greenness of stored and processed foods.

Acid-catalyzed removal of Mg ion upon release of endogenous acids (in case of mechanical injuries and processes), acid formation via fermentation (brining, pickling, etc.), and dressing with an acidic ingredient are responsible for chlorophyll-to-pheophytin conversion. The acids produced in the fermentation that occurs naturally and spontaneously in brined or pickled fruits induce degradation of chlorophylls and chlorophyllides to the corresponding pheophytins and pheophorbides. Although physically acid-catalyzed conversion occurs inside the chloroplast, the media in which it takes place is the fermentor, since it is the diffusion across the membranes by osmosis that leads to fermentation. The intracellular pH is thus altered by the pH of the brining solution. The change in content of chlorophylls and pheophytins as a function of fermentation time and pH of the brine of green olive fermentation is shown in Table 1. The kinetic equation describing the degradation of the pigments is:

$$-\mathrm{d}c/\mathrm{d}t = k[\mathrm{H}^+]^n[\text{pigment}]^n \qquad (1)$$

In brining starting with alkaline treatment, when the interior of the fruit reaches a pH of 8, the concentration of chlorophylls is greater than that of hydrogen ions. In this case, if pheophytinization occurs, its kinetics would be of second order. When the concentration of hydrogen ions in the fermentation media, compared with the existing chlorophyll, is in excess and can thus be considered to be constant, the kinetic reaction is of pseudoorder and can be expressed as:

Table 1 Rate constant (k) for degradation of chlorophyll a, chlorophyll b, and total chlorophyll content in broccoli juice due to a thermal or a combined pressure–temperature treatment

Pressure (MPa)	*Temperature (°C)*	*Chlorophyll a* k (10^{-2} *min*$^{-1}$)	*Chlorophyll b* k (10^{-2} *min*$^{-1}$)	*Total chlorophyll* k (10^{-2} *min*$^{-1}$)
0.1	80	2.51 ± 0.26[a]	0.57 ± 0.03[a]	1.59 ± 0.19[a]
	90	4.41 ± 0.24	1.20 ± 0.05	2.95 ± 0.20
	100	8.84 ± 0.24	2.46 ± 0.11	6.15 ± 0.31
	110	14.13 ± 0.95	4.73 ± 0.27	10.29 ± 0.65
	120	23.60 ± 0.35	9.07 ± 0.78	17.66 ± 0.36
200	60	0.45 ± 0.02	ND	0.33 ± 0.02
	70	1.60 ± 0.15	0.57 ± 0.01	1.13 ± 0.09
	80	4.11 ± 0.43	1.59 ± 0.05	2.76 ± 0.19
500	60	0.52 ± 0.02	0.10 ± 0.01	0.38 ± 0.02
	70	1.48 ± 0.09	0.56 ± 0.03	1.13 ± 0.04
	80	4.04 ± 0.24	2.91 ± 0.07	3.45 ± 0.13
700	60	0.64 ± 0.03	0.23 ± 0.01	0.50 ± 0.02
	70	1.67 ± 0.06	0.84 ± 0.06	1.37 ± 0.05
	80	3.99 ± 0.41	2.78 ± 0.16	3.36 ± 0.23
800	50	0.20 ± 0.02	ND	0.16 ± 0.02
	60	0.57 ± 0.04	0.19 ± 0.02	0.44 ± 0.03
	70	1.91 ± 0.05	0.96 ± 0.04	1.45 ± 0.04
	80	3.48 ± 0.02	3.67 ± 0.14	3.94 ± 0.27

[a]Asymptotic standard error of regression.
ND, not determined.
Reproduced from Van Loey A, Ooms V, Weemaes C *et al*. (1998) Thermal and pressure degradation of chlorophyll in broccoli (Brassica oleraecea) juice. A kinetic study. *Journal of Agriculture and Food Chemistry* 46: 5289, with permission.

$$-\mathrm{d}c/\mathrm{d}t = k'[\text{pigment}]^n \qquad (2)$$

where $k' = k[\mathrm{H}^+]^n$.

At the first step of brining, when the pH is suitable for chlorophyllase activity a portion of the chlorophyll is converted to chlorophyllides, followed by acid-catalyzed formation of the corresponding Mg-free derivatives (pheophytins and pheophorbides). The decrease in the concentration of chloropyll a and b as well as the accumulation of Mg-free forms as a function of time follows first-order kinetics. A rapid accumulation of pheophytins in cooked or canned foods is a result of the combined action of heat and acid release after cell rupture.

At neutral or alkaline pH values, high-temperature short-time heat treatment, as in blanching, often results in an epimerization on carbon-10 of chlorophyll molecule, giving rise to the formation of the so-called a′ and b′ epimers that are a brighter green color than their original pigments. With an extended period of thermal treatment the rate and final products of chlorophyll degradation are changed. In canning the function of time, temperature, and pressure determines the kinetics of chlorophyll degradation. Chlorophyll degradation during thermal processing is frequently described as a first-order reaction. For such kinetics, the integrated form of eqn (2) for a decay process at constant temperature and pressure is given by:

$$X/X_0 = \exp(-kt) \qquad (3)$$

where X_0 is the response value at $t = 0$ (i.e., the concentration of chlorophyll at $t = 0$) and X is the residual response value after treatment (i.e., the residual concentration of chlorophyll).

The temperature dependence of the degradation rate constant (k) at atmospheric pressure can be adequately described using activation energy E_a, as given in the Arrhenius relationship:

$$k = k_{\mathrm{ref}}\exp[E_a/R(1/T_{\mathrm{ref}} - 1/T)] \qquad (4)$$

where k_{ref} is the rate constant at reference temperature T_{ref}, E_a is the activation energy, and R is the universal gas constant.

The pressure dependence of the rate constant (k) at a certain temperature is commonly described as an activation volume V_a, as given in the Eyring relation:

$$k = k_{\mathrm{ref}}\exp[V_a/RT(P - P_{\mathrm{ref}})] \qquad (5)$$

where k_{ref} is the rate constant at reference pressure P_{ref}, V_a is the activation volume at a certain temperature, T is the absolute temperature, and R is the universal gas constant.

In order to interpret the validity of a first-order reaction of chlorophyll degradation, eqn (3) is linearized using a logarithmic data transformation [ln (X/X_0)]. A log linear plot of relative chlorophyll a and b retention versus degradation time is depicted in Figure 2. These plots show that the rate of chlorophyll degradation follows a first-order kinetic model.

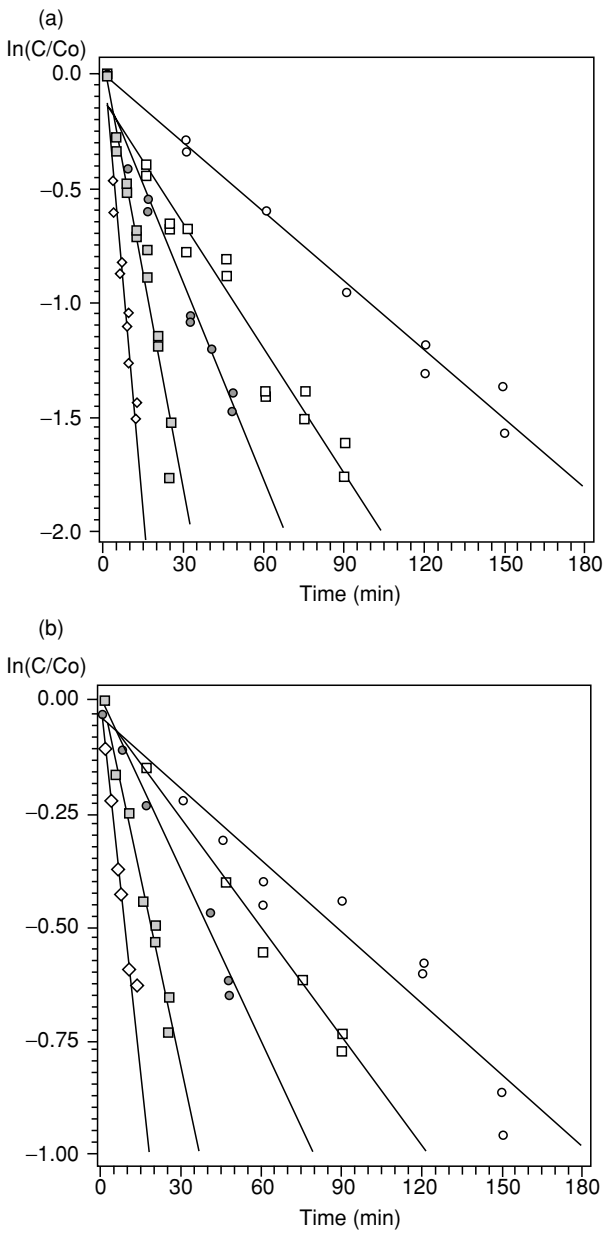

Figure 2 First-order thermal degradation of (a) chlorophyll a and (b) chlorophyll b in broccoli juice at (○) 80, (□) 90, (⊙) 100, (⊡) 110, and (◇) 120 °C. Reproduced from Weemaes CA, Ooms V, Van Loey M, Handrickx ME (1999) Kinetics of chlorophyll degradation and color loss in heated broccoli. *Journal of Agriculture and Food Chemistry* 47: 2404, with permission.

With regard to the ranking in heat stability of chlorophylls a and b, it is evident that chlorophyll a is less heat-stable than chlorophyll b, and hence the former degrades more quickly. The higher thermal stability of chlorophyll b is attributable to the electron-withdrawing effect of its C-3 formyl group.

Chlorophyll exhibits an extreme stability toward pressure processing at ambient or relatively low temperatures (not higher than 50 °C). This can be ascribed to the stability of the covalent structure of chlorophyll to high pressure and to the slight compressibility of covalent bonds.

From semilogarithmic plots of chlorophyll retention in broccoli juice as a function of treatment time at constant pressure and temperature, it can be seen that the pressure–temperature-induced degradation process follows a first-order kinetics. Rate constants for chlorophyll degradation due to thermal or combined pressure–temperature treatment are shown in Table 2.

Severe heat processing enhances the decarbomethoxylation reaction, leading to formation of pyro derivatives of chlorophylls. Heat treatment of 121 °C for more than 15 min is required for the appearance of pyropheophytin a and b. The rate of decarbomethoxylation reaction as a function of time and temperature follows first-order degradation kinetics.

Chlorophyll to pheophytin conversion has been recognized for more than 50 years in frozen foods stored at lower than −18 °C. This conversion is responsible for color change from a bright green to a dull olive green in freeze-stored fruits and vegetables. The rate of such a conversion is shown to be a first-order model with respect to acid concentration. A linear relationship is expected between the appearance and pheophytin formation for frozen plant products.

In drying processing, the rate of chlorophyll degradation is affected by some factors, including drying temperature and time, water activity (a_w), and the storage conditions of dried foods. At a_w values greater than 0.32 and ambient storage with nitrogen gas, pheophytins are expected to be the predominant pigment of green-colored dry vegetables. It is well demonstrated that chlorophyll a undergoes degradation more rapidly than does chlorophyll b by a factor of 2.5–3.0. A linear relationship between a_w and log time for a 20% loss of chlorophyll is recognized for dry fruit and vegetables. Recent studies indicate that the main reason for chlorophyll degradation in heat-dried foods (sweet peppers, kidney beans, Chinese onions, etc.) is the free radicals initiated in the food system, particularly in the presence of unsaturated fatty acids. It is recommended to increase the content of effective scavengers in the product before drying to neutralize free radicals and prevent pigment degradation.

Formation of pheophytins in edible oils during extraction and refining is a well-known photo-catalyzed degradation. At different temperature and luminous energies, this conversion follows first-order kinetics. From the Arrhenius plot display, it appears that the

Table 2 Qualitative and quantitative changes in chlorophylls during table olive processing as a function of pH

Time (days)	Pigment concentration (μmol kg^{-1})													
	pH		a series						b series					
	Fruit	Brine	Chl	Phy	Chl + phy	Chld	Pho	Chld + pho	Chl	Phy	Chl + phy	Chld	Pho	Chld + pho
0	6.08		49.97		49.97				12.37		12.37			
4	8.44	8.28	40.73		40.73	7.04		7.04	7.62		7.62	4.52		4.52
6	7.28	6.93	33.09	8.47	41.56	4.70	2.70	7.40	7.29	0.37	7.66	3.86	0.85	4.71
8	6.67	5.49	26.18	15.44	41.62	3.18	3.90	7.08	6.61	0.72	7.33	3.28	1.43	4.71
10	5.17	5.49	25.01	17.20	42.21	2.60	4.95	7.55	6.58	1.06	7.64	2.58	2.12	4.70
14	5.40	4.56	15.12	26.52	41.64	1.40	5.61	7.01	5.95	1.43	7.38	2.11	2.35	4.46
19	5.41	4.76	13.91	27.63	41.54	0.70	6.10	6.80	4.86	2.43	7.29	1.41	3.10	4.51
20	5.08	4.65	12.59	29.01	41.60	0.28	6.80	7.08	4.83	2.48	7.31	1.05	3.40	4.45
26	4.83	4.57	11.76	29.53	41.29		7.02	7.02		3.79	2.74	6.53	4.14	5.10
30	4.69	4.60	6.16	35.40	41.56		7.05	7.05	3.99	3.65	7.64	0.35	4.36	4.71
33	4.55	4.52	5.78	35.32	41.10		7.23	7.23	3.66	3.17	6.83		4.61	4.61
40	4.61	4.42	3.69	36.99	40.68		7.16	7.16	3.14	4.42	7.56		4.60	4.60
50	4.53	4.46	1.52	38.50	40.02		7.18	7.18	2.42	5.20	7.62		4.66	4.66
54	4.44	4.35	1.54	38.24	39.78		7.18	7.18	2.24	5.70	7.97		4.70	4.70
60	4.43	4.29	0.90	39.10	40.00		7.15	7.15	1.88	6.18	8.06		4.70	4.70
70	4.41	4.32	0.37	40.53	40.90		7.20	7.20	1.47	6.30	7.77		4.71	4.71
89	4.33	4.21	0.23	40.82	41.05		7.12	7.12	0.95	6.30	7.25		4.78	4.78
104	4.45	4.25		40.77	40.77		7.14	7.14	0.61	6.59	7.20		4.58	4.58
117	4.37	4.26		40.69	40.69		7.15	7.15		6.57	6.57		4.90	4.90
161	4.21	4.10		40.85	40.85		6.94	6.94		6.41	6.41		4.80	4.80
203	4.34	4.19		40.81	40.81		6.97	6.97		6.68	6.68		4.51	4.51
287	4.27	4.08		40.76	40.76		7.17	7.17		6.59	6.59		4.67	4.67

a series: chlorophyll a and derivatives; b series: chlorophyll b and derivatives; Chl, chlorophyllide; Phy, pheophytin; Pho, pheophorbide.
Reproduced from Mínguez-Mosquera M, Candul-Rojas B, Mínguez-Mosquera J (1994) Mechanism and kinetics of the degradation of chlorophyll during the processing of green table olives. *Journal of Agriculture and Food Chemistry* 42: 1089, with permission.

incident luminous energy does not change activation energy, but increases the reaction frequency factor.

Effect of Controlled Storage and Atmosphere

Storage in controlled or ethylene-containing atmosphere can affect, to a considerable extent, the chlorophyll content of food commodities. Recently, such storage has been widely applied in food technology either to speed up chlorophyll loss in some fruit (citrus fruits, bananas, mangoes, pears, etc.) or to maintain the greenness of stored crops.

Ethylene can be used directly as gas to treat the product in the package or in closed stores. Also it is indirectly applicable either by treating trees or immersing fruit in diluted solutions of ethylene, releasing plant growth regulators such as ethrel and hydrel (150–250 p.p.m.). To accelerate the ripening of some pears, a combination of 500 p.p.m. ethylene and 100 p.p.m. acetylene gas is recommended.

Ethylene treatment enhances degreening of the peel of some fruit via *de novo* synthesis of chlorophyllase and chloroplast-dependent enzyme that regulates chlorophyllase activity. However, in spinach leaves accelerated color loss by ethylene is not associated with increased content of dephytylated derivatives. This difference may be due to differences in the degradation pathway in the different products. It is stated in the literature that ethylene treatment can also enhance the synthesis of chlorophyll-oxidative enzymes in cotyledons held in the dark. An inverse correlation between the content of chlorophyll of ethylene-stored radish cotyledons and the activity of chlorophyll-oxidizing enzymes indicates that chlorophyll oxidation is an important step in the degreening process.

Storage under controlled atmosphere (CA) retards chlorophyll decomposition in a wide variety of agricultural crops. CA of 10% O_2 and 10% CO_2 is very effective in reducing the rate of chlorophyll degradation to the extent that the shelf-life would be extended about 20% longer than that in air-held samples (**Figure 3**).

Low O_2 (3%) and high CO_2 (20%) treatment, as in CA of tomato storage, maintains firmness, reduces chlorophyll loss, and slows lycopene and carotenoid development. In CA storage, ethylene-mediated responses are impaired by low O_2 and high CO_2 and exogenous ethylene and low O_2/high CO_2 concentrations act antagonistically. Controlled storage may include vacuum cooling and prestorage treatment with ethylene-reducing and respiration rate-lowering agents. Of these agents, ethanol vapor is used to maintain freshness and keep the chlorophyll

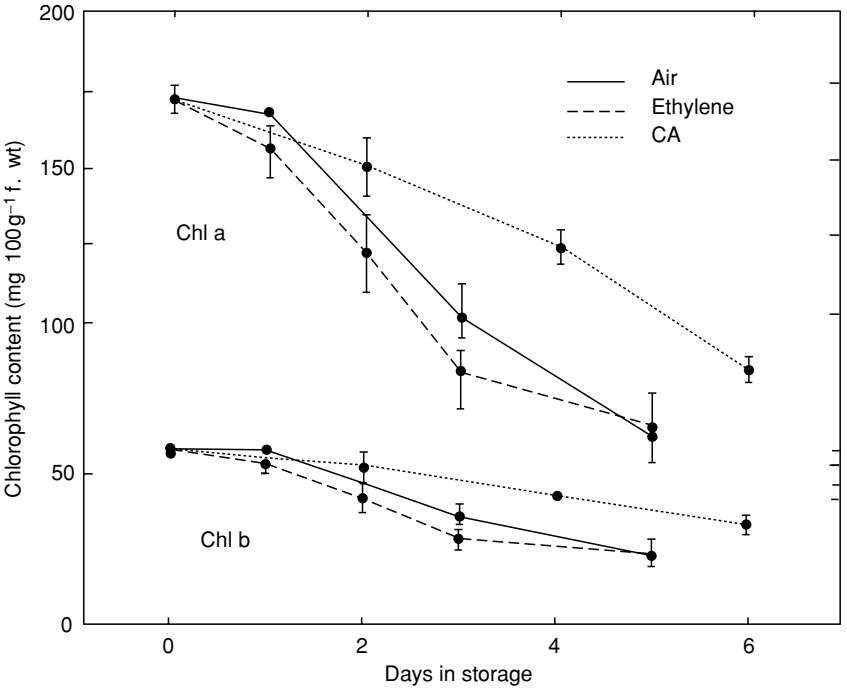

Figure 3 Change in content of chlorophylls of parsley leaves stored in air with or without 10 p.p.m. C_2H_4 or controlled atmosphere of 10% O_2 and 10% CO_2. Reproduced from Yamauchi N and Watada AE (1993) Pigment changes in parsley leaves during storage in controlled and ethylene containing atmosphere. *Journal of Food Science* 58: 616, with permission.

content at high levels in some horticultural crops. A prestorage treatment of special interest is the short-term immersion of broccoli florets in ozonated water (10 p.p.m.)for 10–50 min. Although the treatment significantly reduces ethylene formation and total soluble proteins by the end of cold storage, it causes the vegetables to lose their green color rapidly.

Antioxidant–Oxidant Role

Due to their chemical nature and hydrophobic properties, chlorophylls are able to interfere with the chains of lipid oxidation in foods. From industrial practices and *in vitro* experiments, chlorophylls and their degraded products can play the role of either antioxidants or prooxidants. Factors which are most likely to determine the antioxidative or oxidative activity of chlorophylls include variety, ripeness stage, and chemical composition of food where they exist, the presence or absence of effective oxidants or antioxidant, and the climate.

When tested by ferric thiocyanate and other assays, chlorophylls and their derivatives exhibit marked antioxidant activity. Recent studies show that acidic fractions from plant containing pheophytins and related compounds have an antioxidant effect which is higher than that of α-tocopherol and comparable to that of BHT (butylated hydroxytoluene) (Figure 4).

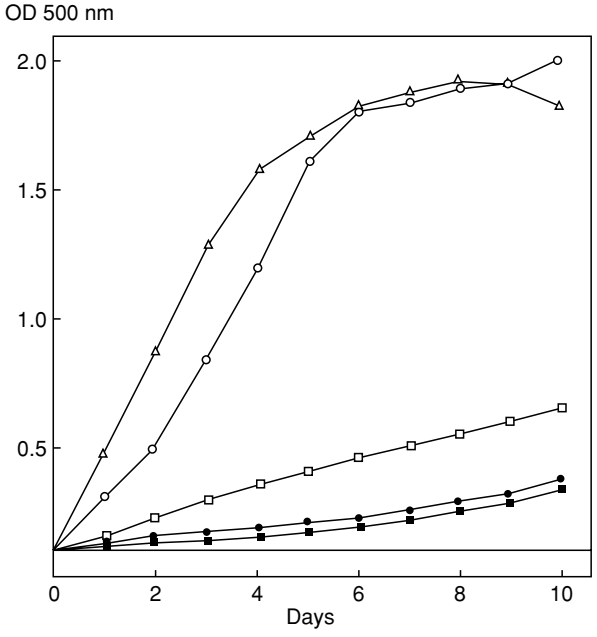

Figure 4 Antioxidative effect of fractionated extracts from marine algae by the ferric thiocyanate method. Sample concentration 0.02%; BHT concentration 0.01%. □, neutral fraction; •, weakly acidic fraction; ■, strongly acidic fraction; △, basic fraction; continuous line, BHT; ○, control. Reproduced from Cahyana AH, Shuto Y, Kinoshita Y (1992) Pyropheophytin a as antioxidative substance from the marine alga, arame (Eisenia bicyclis). *Bioscience Biotechnology Biochemistry* 56: 1533, with permission.

The antioxidant activity of chlorophyll-derived compounds is attributable to their interference with free radical cycles as scavengers to neutralize highly reactive oxygen-free forms such as superoxides. The porphyrin ring system of chlorophylls seems to be important for the antioxidant activity.

Treating some foods with oil containing chlorophyll between 200 and 800 mg% significantly reduces peroxide value during storage in the dark. This gives convincing evidence on the antioxidant effect of chlorophylls. In dry, oiled, and toasted laver, for example, the high lipid oxidative stability is ascribed to the presence of chlorophylls in the oil used. A positive correlation between oil stability and chlorophyll content is recognized in the processing of virgin oil.

The antioxidant activity of chlorophylls can be enhanced by ascorbic acid and α-tocopherol which exists at low level as 10^{-6} mol l^{-1} for each in the food system. This suggests that synergism between chlorophylls and other effective bioantioxidants is possible and of great importance from the point of view of oil oxidative stability.

Cooling green colored vegetables modifies their antioxidative ability due to chlorophylls. In cooked juices from broccoli, cabbage, green pepper, and leek, the superoxide scavenging ability is higher than that of the original raw materials. Moreover, the antioxidant effect of cooked spinach containing considerable amounts of chlorophyll a and b is much lower than that estimated for other vegetable extracts having no chlorophyll remains. It is, therefore, convincingly evident that Mg-free derivatives of chlorophylls have greater chemical capacity and ability as oxidation barriers than their origins. Also, as many recent *in vitro* studies have indicated, the antioxidant effect is more marked for chlorophyll a than chlorophyll b.

In oil-in-water food emulsions which have been stored in the dark, the addition of chlorophylls from plant sources such as dried leaves insures oxidation prevention equal to that provided by 80 p.p.m. of propyl gallate. However, the effect of green plant extract on the stability of oil-in-water emulsions is not the same when the emulsions are exposed to light. In contrast, under various conditions the prooxidant effect of chlorophylls is recognized.

As a practice in the processing of rapeseed oil, the chlorophyll content of raw oil should not exceed 25 mg kg^{-1} to maintain top-grade oil and extend shelf-life. The oxidation stability of the refined oil is reduced as the preprocessing content of pheopytins exceeds 30 p.p.m. The prooxidant effect of chlorophyll-derived compounds is increased, to a high extent, when the oil is stored in the light.

The degree of unsaturation of oil is another factor which modifies the prooxidant ability of chlorophylls. In highly unsaturated marine oil, the addition of chlorophylls from green tea extracts at less than 25 p.p.m. impairs oil oxidative stability and, moreover, removing chlorophylls from the tea extract gives it an antioxidant effect which is even higher than that of synthetic antioxidants at the same concentration. It can therefore be concluded that, in highly unsaturated oils, the prooxidant capacity of chlorophyll-related compounds is so high that oxidation is scarcely inhibited by spontaneous antioxidants like flavonoids and tocopherols.

Another type of chemical decay in foods is the photo-catalyzed oxidation of unsaturated lipids. Chlorophylls contribute to photooxidation of oils exposed to light or irradiation via a photosensitization reaction. The mechanism of light-induced photosensitization may involve free radical initiation and incorporation of singlet oxygen.

Fat-soluble oxygen quenchers and free radical scavengers such as carotenoids, ascorbyl palmitate, and tocopherols can interfere to minimize chlorophyll-sensitized oxidation of fatty acids and oils. Ascorbyl palmitate is recommended to be applied as an antioxidant with photo-oxidation-preventive properties (total quenching rate = 10^{-8} mol s^{-1}) in fatty foodstuffs. Tocopherols are efficiently used to minimize chlorophyll-sensitized oxidation. The activity of tocopherol analogs to prevent photosensitization is in the order of $\delta > \gamma > \beta > \alpha$. Carotenoid-type pigments actively interfere with photosensitized oxidation in food systems. The antiphotosensitization ability of carotenoids is attributed to the rapid transfer of radicals from chlorophyll to carotenoid molecules as well as the ability of carotenoids to quench oxygen free forms. In photosensitized reactions, carotenoid pigments undergo *trans*-to-*cis* isomerization on their structure. This may explain the presence of *cis* carotenoids in photosynthetic tissues.

It should be noted that all chlorophyll-derived compounds are able to induce photoisomerization of carotenoids in light-exposed food systems. Different carotenoids vary in their rate and capacity of interaction with photosensitizing agents. When tested by the headspace and rancimat methods, capsanthin has the highest rate and ability to minimize chlorophyll-induced photosensitization, followed by β-carotene and lutein.

Metal–Chlorophyll Interaction

Bright green color in drastically heat-processed vegetables is associated with the formation of metal–chlorophyll complexes. Copper and zinc, present as

Table 3 Reaction rate constants (k), half-life values ($t_{1/2}$) and energy activation (E_a) for the reaction of chlorophyll a and derivatives with zinc (II) ion at 20, 25, 30, and 35 °C

Pigment	*Temp. (°C)*	*Slope* ± SD *(× 10^{-3}) (min^{-1})*	k ± SD *(min^{-1} m^{-1})*	*Correlation coefficient*	$t_{1/2}$ ± SD *(min)*	E_a *(kcal mol^{-1})*
Pheophytin	20	−0.99 ± 0.02	0.035 ± 0.001	0.998	305 ± 7	22.09
	25	−1.91 ± 0.05	0.067 ± 0.002	0.999	157 ± 3	
	30	−3.37 ± 0.16	0.12 ± 0.01	0.995	89 ± 4	
	35	−6.36 ± 0.16	0.22 ± 0.01	0.998	47 ± 1	
Pyropheophytin	20	−1.73 ± 0.01	0.061 ± 0.004	0.998	174 ± 1	22.60
	25	−3.39 ± 0.01	0.119 ± 0.002	0.999	89 ± 1	
	30	−6.41 ± 0.01	0.23 ± 0.01	0.998	47 ± 1	
	35	−11.45 ± 0.01	0.40 ± 0.01	0.999	26 ± 1	
Ethyl pheophorbide	20	−1.76 ± 0.01	0.062 ± 0.001	0.998	171 ± 1	21.40
	25	−3.12 ± 0.01	0.110 ± 0.001	0.998	96 ± 1	
	30	−5.65 ± 0.01	0.20 ± 0.01	0.998	53 ± 1	
	35	−10.55 ± 0.01	0.37 ± 0.01	0.998	29 ± 1	
Methyl pheophorbide	20	−1.81 ± 0.01	0.064 ± 0.001	0.999	166 ± 1	21.36
	25	−3.28 ± 0.01	0.115 ± 0.001	0.999	92 ± 1	
	30	−6.11 ± 0.01	0.21 ± 0.01	0.998	49 ± 1	
	35	−10.72 ± 0.01	0.38 ± 0.01	0.995	28 ± 1	
Pheophorbide	20	−3.36 ± 0.01	0.118 ± 0.001	0.998	90 ± 1	21.52
	25	−5.73 ± 0.01	0.20 ± 0.01	0.999	53 ± 1	
	30	−10.31 ± 0.01	0.36 ± 0.01	0.997	29 ± 1	
	35	−20.47 ± 0.01	0.72 ± 0.02	0.998	15 ± 0.5	
Pyropheophorbide	20	−5.57 ± 0.01	0.20 ± 0.01	0.998	54 ± 1	22.06
	25	−10.68 ± 0.01	0.38 ± 0.01	0.998	28 ± 1	
	30	−19.33 ± 0.01	0.69 ± 0.01	0.998	16 ± 0.5	
	35	−35.50 ± 0.01	1.25 ± 0.01	0.998	9 ± 0.5	

SD, standard deviation for duplicate determination.
Reproduced from Tonucci LH and von Elbe H (1992) Kinetics of zinc complexes of chlorophyll derivatives. *Journal of Agriculture and Food Chemistry* 40: 2341, with permission.

contaminants in processing solutions, are the most familiar metals reacting with chlorophylls during the thermal processing and storage of processed foods. Such metal complexes can be found in canned green beans, Brussel sprouts, spinach and pea purées, table olives, candies, and chewing gum. The formation of Zn–chlorophyll complexes is of greater interest because of the toxicity nature of copper complexes.

The formation of green-colored metal complexes of chlorophyll derivatives is the basis of efforts by processors to preserve the desired green color of canned vegetables. This process is called regreening. Spontaneous regreening of vegetables during processing, which has long been known by canners as a nonuniform defect, is attributed to Zn-containing pigments.

Copper–chlorophyll complexes are formed more rapidly than Zn complexes. Clorophyll a-related derivatives form metallocomplexes more rapidly than chlorophyll b derivatives. Furthermore, pyropheophytins and pheophorbides have a markedly high tendency to react with metal ions. Steric hindrance due to the lack of a carbomethoxy group at C-10 is one possible reason for the rapid interaction of pyropheophorbides and metals. The other explanation is that the distribution of charge in the aromatic system is affected by the carbomethoxy group, which is strongly electron-withdrawing. In the absence of a carbomethoxy group, the pyrrole nucleus could become slightly more negatively charged, resulting in an increased reaction rate with positively charged ions. The kinetic parameters of Zn–chlorophyll reaction are shown in Table 3.

Food additives such as sugars, salts, and other preservatives may influence metal complexes of chlorophylls. While sucrose, glucose fructose, chlorides, sulfates, lactates, acetates, and propionates have no influence on metal–chlorophyll complex formation, malate, tartarate, citrate phosphate, and other metal-chelating agents such as ethylenediaminetetraacetic acid significantly decrease the rate of metal–chlorophyll reaction. In contrast, thiocyanate, benzoate, oleate, and caprylate ions accelerate the chemical interaction of chlorophyll with Cu and Zn.

The presence of active anionic compounds in canning solutions facilitates the formation of metal complexes of chlorophyll-derived compounds by adsorbing on to chloroplast membranes, thereby increasing the negative surface charge and giving rise to a higher affinity to bind positively charged ions. Increasing cation concentration in the food promotes the complex formation of chlorophylls. The mechanism for this promotion includes a direct common ion effect, which increases the rate constant

of the reaction of cations and chlorophylls, and an indirect pH-lowering effect of cation.

The pH of the media is an important factor affecting chlorophyll–metal complex formation. The rate of Zn–complex formation increases between pH 4.0 and 8.0, reaching a maximum between 6.0 and 8.0. Raising the pH to values higher than 8.0 decreases the amount of metal–chlorophyll complexes formed. At high pH values, chlorophyll a and b are very stable and the amount of derivatives available for the reaction is reduced.

See also: **Antioxidants**: Natural Antioxidants; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Copper**: Physiology; **Freezing**: Structural and Flavor (Flavour) Changes; **Oxidation of Food Components**; **Storage Stability**: Mechanisms of Degradation; **Zinc**: Physiology

Further Reading

Castelfranco PA and Beale SI (1981) Chlorophyll biosynthesis. In: Hatch MD and Boardman NK (eds) *The Biochemistry of Plants*. New York: Academic Press.

Cross J (1987) *Pigments in Fruits*. London: Academic Press.

Czygan FC (1980) *Pigments in Plants*. New York: Gustav Fischer.

Eskin NAM (1990) *Biochemistry of Foods*. San Diego: Academic Press

Fennema OR (1985) *Food Chemistry*. New York: Marcel Dekker.

Goodmin TW (1976) *Chemistry and Biochemistry of Plant Pigments*, 2nd edn. New York: Academic Press.

Heaton JW and Marangoni AG (1999) Chlorophyll degradation in processed foods and senescent plant tissues. *Trends in Food Science and Technology* 7: 8–15.

Hendry GAF (1996) *NERC Unit of Comparative Plant Ecology*. Sheffield: Press of Sheffield University.

Hulme AC (1970) *The Biochemistry of Fruits and Their Products*. London: Academic Press.

Jen JJ (1989) *Quality Factors of Fruits and Vegetables. Chemistry and Technology*. Washington: Press of American Chemical Society.

Murata N (1992) *Progress in Photosynthesis Research*. Dordrecht: Kluwer Academic.

Schwartz SJ and Lorenzo TV (1990) Chlorophylls in foods. In: Clydesdale FM (ed.) *Critical Reviews in Food Science and Nutrition*. Boca Raton: CRC Press.

Chocolate *See* **Cocoa**: Chemistry of Processing; Production, Products, and Use

CHOLECALCIFEROL

Contents
Properties and Determination
Physiology

Properties and Determination

D C Woollard, AgriQuality NZ, Auckland 1, New Zealand
H E Indyk, Anchor Products, Waitoa, New Zealand

Physical Properties

Cholecalciferol (9,10-seco(5Z,7E)-5,7,10(19)-cholestatriene-3β-ol), commonly referred to as vitamin D_3, exists in the pure form as white crystalline needles. Referred to as a 'fat-soluble' vitamin, it is insoluble in water but is readily soluble in most organic solvents, notably hydrocarbons, chlorinated hydrocarbons, and alcohols. A closely related substance, ergocalciferol (9,10-seco(5Z,7E)-5,7,10(19),22-ergostatetraene-3β-ol), will also be occasionally referred to in the text. Described more simply as vitamin D_2, it is physically, chemically, and nutritionally similar to cholecalciferol. The term 'vitamin D' usually implies collectively both cholecalciferol and ergocalciferol together with any other active isomers and metabolites. Relevant physical properties are listed in **Table 1**.

Table 1 Physical properties of the calciferols

	Cholecalciferol	*Ergocalciferol*
Molecular weight	384.62	396.63
Empirical formula	$C_{27}H_{44}O$	$C_{28}H_{44}O$
Melting point (°C)	84–85	115–118
λ_{max} (nm)	264.5	264.5
Extinction coefficient, $E^{1\%}_{1cm}$ in hexane	485	459
Optical rotation in chloroform	+52°	+52°

Chemical Properties

Cholecalciferol is described in terms of steroid nomenclature and numbering (Figure 1). The '9,10-seco' prefix is added to denote the bond cleavage and the opening of the typical steroidal ring structure, an essential feature necessary to impart antirachitic activity. This cleavage is generally induced by ultraviolet (UV) irradiation of the precursor compound, 7-dehydrocholesterol (provitamin D_3), and occurs in the skin during exposure to sunlight. The previtamin D_3 so formed subsequently undergoes temperature-dependent equilibration to vitamin D_3. This process, along with the requirement for further hydroxylation challenges its original classification as vitamin; it is currently more accurately regarded as a prohormone.

Any homolog which possesses antirachitic activity is referred to as a D vitamin and each of the several compounds sharing this property elicits a unique and selective biological response (Figure 1). Common structural features of these substances are the β stereochemistry of the 3-hydroxy substituent and the *cis* conformation of the double bond at C5. While the 3-hydroxy substituent does not have an overwhelming influence on biological activity, other structural features, such as the A ring configuration and side-chain length, appear to be more critical. Thus, while alterations to the side chain result in diverse activity, only vitamins D_2 and D_3 are of prominence therapeutically and commercially. They are usually obtained via chemical synthesis, although vitamin D_3 has traditionally been extracted from fish liver oils.

It is widespread practice to express vitamin D concentration in food in international units (IU), rather than on a weight basis (1 IU is equivalent to 0.025 μg of either calciferol, although this equivalence in humans is occasionally challenged). This can be a useful concept because it reflects the nutritional status where several components of different biopotency coexist within a product. Nevertheless, there is some move back to mass units, particularly in supplemented foods where the contribution of cholecalciferol is dominant.

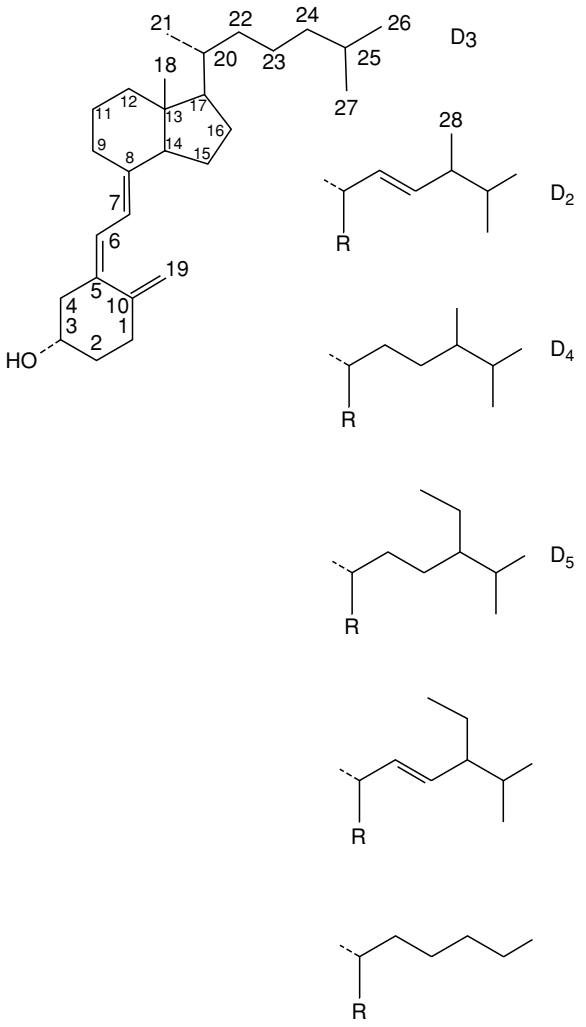

Figure 1 The chemical structure of cholecalciferol (vitamin D_3), indicating the carbon-numbering system of the molecule. Related calciferols with different side-chain configurations are also given, including ergocalciferol (vitamin D_2).

The stability characteristics of cholecalciferol have received considerable attention because they play a major part in the shelf-life of marketable products. Under conditions of low thermal, photochemical, and oxidative stresses, cholecalciferol is stable for several years. When added to foods and exposed to industrial and domestic processes, degradation is minimal. However, excessive light (UV) causes significant losses through the production of inactive substances such as toxisterol, suprasterol, lumisterol, and tachysterol, in a process which is accelerated by heating (Figure 2). Cooking temperatures above 100 °C, even in the absence of light and air, will cause isomerization through ring closure to the pyrocholecalciferols. Cholecalciferol is also sensitive to low pH and, if subjected to an acidic environment, will irreversibly

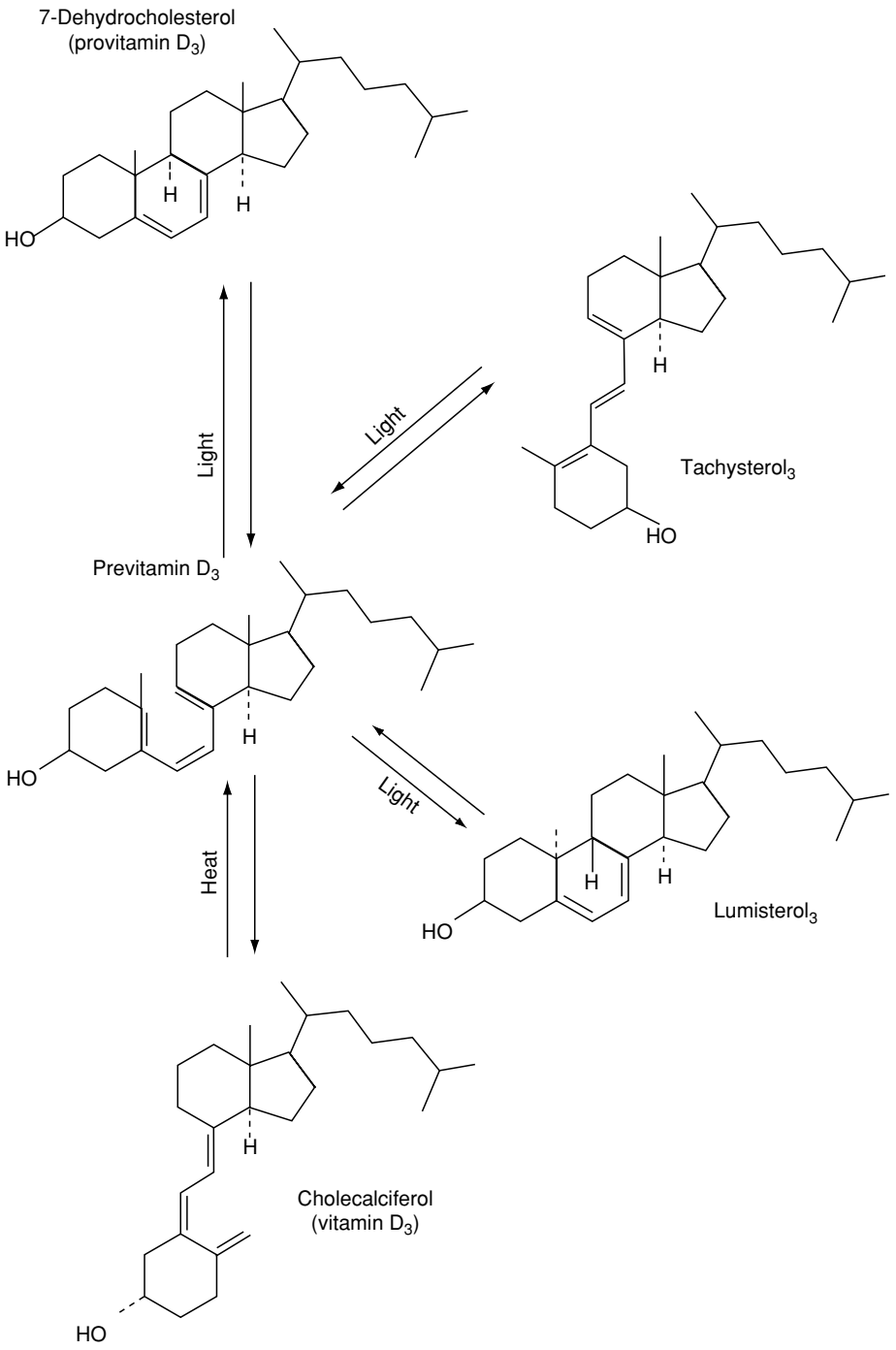

Figure 2 The major photochemical reactions of cholecalciferol. Ergocalciferol undergoes similar reactions. Overirradiation products are not shown.

rearrange to the inactive isotachysterol via the 5,6-*trans* isomer. These reactions (illustrated in Figure 3) are complex and will occur to an extent determined by the overall environment to which cholecalciferol is exposed. The reactions largely involve alterations to ring structure and so affect ergocalciferol and other D vitamins in a similar manner. It is generally acknowledged that ergocalciferol is less stable than cholecalciferol. This may imply that the double bond in the side chain of ergocalciferol imparts additional lability to the molecule.

Losses during storage are also known to occur in foods, and vary considerably between food types and conditions of storage. Thus, low temperatures and

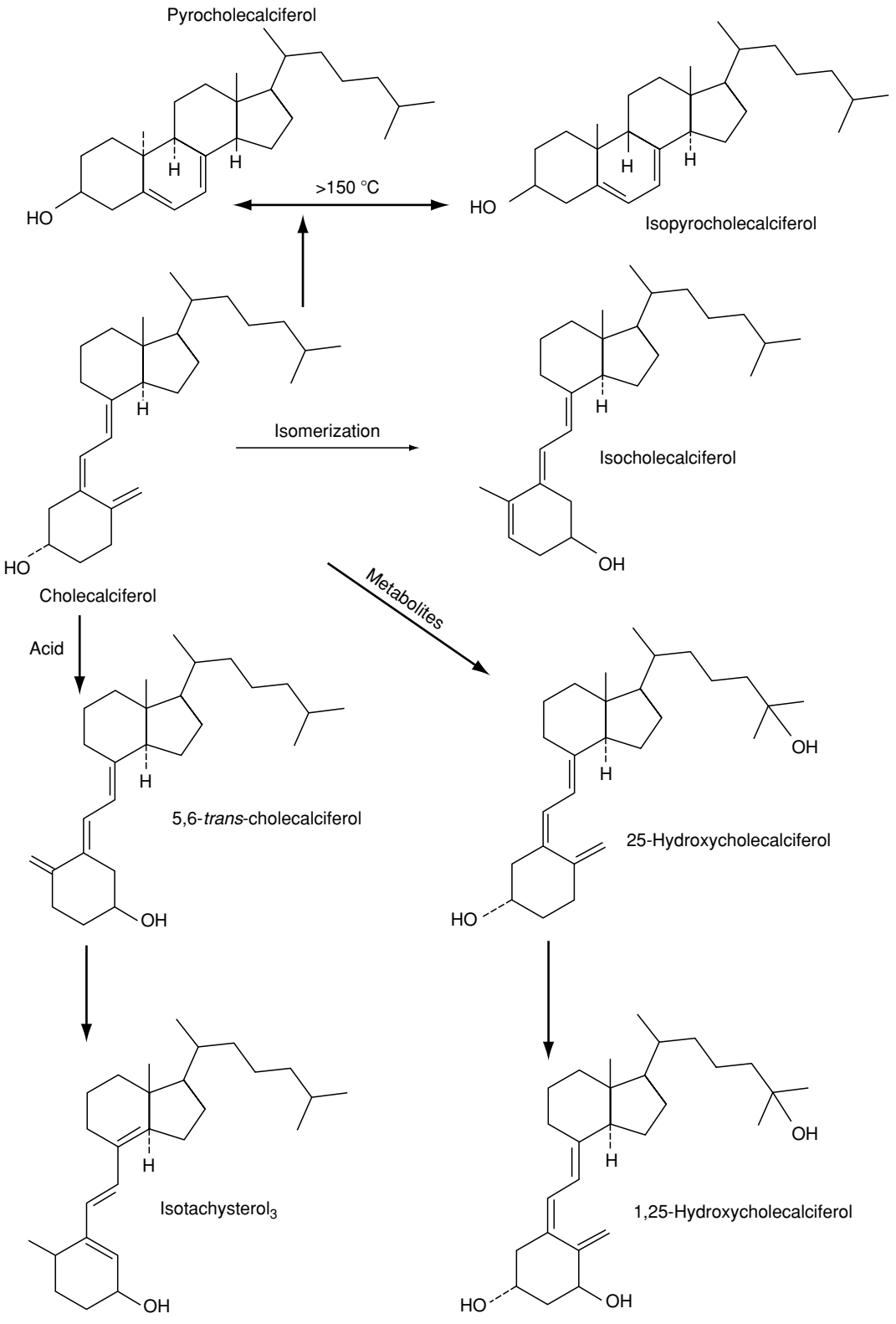

Figure 3 The major degradation reactions of cholecalciferol as relevant to foods.

absence of light will minimize losses in manufactured foods, as will vacuum packaging or nitrogen flushing. Foods containing natural, or added fat-soluble antioxidants such as vitamin E and carotenoids will, in general, exhibit superior cholecalciferol conservation. (*See* **Antioxidants**: Natural Antioxidants; **Antioxidants**: Synthetic Antioxidants.)

Occurrence and Forms in Foods

Most natural foods are limited in their content of active vitamin D components. 7-Dehydrocholesterol (provitamin D_3) and ergosterol (provitamin D_2) are widely distributed within the animal and plant kingdoms and supply a potentially good dietary source of vitamin D for humans. In particular, provitamins are abundant in fish, eggs, yeast, liver, milk, and some vegetables such as cabbage. It has been suggested that ergosterol only enters the food chain following fungal contamination. Consequently, mushrooms are good sources of ergosterol, particularly in the wild.

Vitamin D itself is generally present at low levels in unfortified foods derived from animal sources. Eggs, fish, and fish liver oils possess significant amounts, while meat and dairy products contain lesser quantities. Plants and vegetable oils contain negligible levels (< 2% recommended dietary allowance). Table 2 provides a list of the richest food sources of vitamin D and their approximate concentrations.

In addition, previtamin D isomers invariably coexist with cholecalciferol and ergocalciferol. Previtamin concentrations are proportional to the calciferols and are influenced by thermal conditions during food processing and storage. Higher temperatures increase the previtamin-to-vitamin ratio. Although the previtamins are biologically active precursors, they are not always accounted for in nutritional tables due to analytical difficulties in their determination.

Some foods also contain small but significant quantities of hydroxylated metabolites. These compounds are highly bioactive and are found in edible tissues (meat, liver) and fluids (milk) as a result of their biosynthesis within the live animal. The most predominant and biologically important compounds are 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D, while other trihydroxy metabolites may also be present.

Table 2 Typical vitamin D content of various foods

	Vitamin D content of edible portion (μg per 100 g)[a]
Natural food products	
Fish liver oils	150–3800
Fish	2–30
Egg yolk	5–8
Mammalian liver	0.5–4
Mushrooms (cultivated)	1–3
Mushrooms (wild)	10–30
Butter	1.5–2
Meat	0.2–2
Cheese	0.1–1
Liquid whole milk (raw)	0.05–0.15
Green vegetables (typical)	0.005 (approx.)
Supplemented food products	
Dietary formulae (milk- and soya-based)	3–14
Whole-milk powders	6–12
Margarine	8–10
Infant formulae	5–9
Liquid milk	0.75–1.25

[a]To convert to IU per 100 g, multiply by 40.

Use in Food Fortification

The supplementation of food products with vitamin D is a contentious issue because of the toxicological consequences of overdose. While adequate exposure to sunlight will generally negate the need for this practice, modern western lifestyles (including the recent trend to reduce the intake of dietary fat) and climatic variables justify its continuation. (*See* **Food Fortification**.)

As many diets will not supply the 10 μg (400 IU) of vitamin D required per day, some additional source is generally recommended by nutritionists. The balance of vitamin can be obtained from specially prepared natural sources (concentrated fish oils) or from consumption of artifically fortified food products. In this way, the onset of diet-induced bone disease is minimized, particularly in infants, the sick, and the elderly. Legislation in many countries governs fortification strategies.

Cholecalciferol is the most common D vitamin additive, ergocalciferol being less frequently used for human nutrition. Difficulties potentially arise in delivery of the vitamin to the consumer as a consequence of its intrinsic hydrophobic and labile nature. Common vehicles are margarine, milk, and milk powders, while a number of cereal products and various dietetic formulations contain vitamin D as part of their primary purpose.

Cholecalciferol can be easily incorporated into fat and oil-based foods by simple dissolution. It is common to add both vitamin D and vitamin A (retinol) to these foods. A protective phenolic antioxidant (e.g., 3-tert-butyl-4-hydroxyanisole, tert-butylhydroquinone, tocopherol) is usually incorporated at the same time. For powdered foods, cholecalciferol is added in an encapsulated 'beadlet' form, protected by a gelatine or acacia barrier and incorporating

stabilizers and carrier. If the blending of the vitamin and food is performed by dry mixing, then cholecalciferol is maintained in its protected environment, ensuring enhanced shelf-life. However, attaining uniform dispersion remains a problem and particle sizes must be carefully controlled to avoid redistribution during packaging, transport and storage. Alternatively, these difficulties can be overcome using 'wet-blending' procedures, but the vitamin is inevitably released from within its stabilized form into direct contact with the bulk food. Consequently, the receiving environment must be designed to minimize the potentially rapid degradation of the cholecalciferol supplement. (*See* **Retinol**: Physiology.)

Analytical Considerations

While it is clinically important to measure the hydroxylated metabolites, food scientists have been largely concerned with vitamin D itself, at both endogenous and supplemental levels.

The quantification of the parent secosteroid in foods is complicated by several factors, including low concentration, overwhelming excesses of other lipophilic components, limiting spectral properties (nonselective λ_{max}, low absorptivity), absence of native fluorescence, thermal instability, and the requirement to differentiate cholecalciferol from ergocalciferol.

Historically, curative and prophylactic bioassay techniques, based on the antirachitic quality of vitamin D-containing food, have been used extensively and have the advantage of estimating the true (species-specific) physiological response. Although sensitive, disadvantages of time, cost, and poor precision moderate against the use of bioassays in routine food analysis.

Competitive protein-binding radioassay techniques, utilizing a vitamin D receptor, have been exploited more recently, especially in clinical assays. Such biochemical recognition methods offer advantages in sensitivity and specificity, but extensive sample purification is mandatory in order to avoid endpoint interference from food artifacts. Lack of binding discrimination between vitamins D_2 and D_3 as well as the need for radioactive tracers and lengthy incubation periods remain problematic within the food industry. (*See* **Immunoassays**: Radioimmunoassay and Enzyme Immunoassay.)

Physicochemical determination by UV spectrophotometry or colorimetry without prior separation is clearly impracticable except for highly simplified food matrices, since the spectral properties of the parent or derivative calciferol species are insufficiently characteristic. Even high-potency pharmaceutical preparations still require scrupulous clean-up procedures in order to minimize spectral interference.

The development of instrumental chromatographic techniques has revolutionized the task of vitamin D estimation in both food and clinical samples. Although novel detection methods exploiting selective UV, fluorescence derivatization, or electrochemical techniques are presently under investigation, contemporary UV approaches still require the inclusion of rigorous purification procedures, since this detection mode remains inherently nonspecific.

Isolation, Extraction, and Clean-up

Vitamin D is vulnerable to oxygen, light, low pH and is also subject to reversible isomerization when heated. Therefore, adequate precautions are essential throughout any analytical procedure to avoid loss of the target analytes.

All chemical extraction techniques exploit the inherent lipophilic property of this vitamin and a vitamin-rich fraction is separated from other food components either by saponification or by total lipid extraction.

Saponification

Alkaline hydrolysis is a convenient way to eliminate the bulk of neutral lipids and is particularly favored in high-fat foods and those of an intractable nature. This popular technique is also advantageous for products containing encapsulated supplements and when relatively large sample quantities are needed for reasons of assay sensitivity or analyte heterogeneity.

While some authors report the use of high-temperature saponification, strategies such as direct measurement or use of conversion factors are then needed to account for the consequent elevated levels of previtamin D. These concerns can significantly complicate the assay and may be avoided through the use of overnight hydrolysis at ambient temperature, which additionally offers operational simplicity. It is considered mandatory during extraction to include a protective antioxidant and to purge with inert gas. Vitamin D, along with the other fat-soluble vitamins, sterols, carotenoids and hydrocarbons, remains in the nonsaponifiable fraction.

Enzymatic hydrolysis with lipase has been suggested as an alternative to saponification, thereby minimizing possible degradation and facilitating concurrent recovery of the alkali-unstable vitamin K, if required.

The nonsaponifiable fraction containing the calciferol is rapidly and conveniently partitioned into a solvent mixture of hexane and diethylether with good recovery, although other solvent systems have

also been cited in the literature. Following washing to remove hydrophilic remnants and excess alkali, the organic phase is dried and evaporated to recover an enriched crude extract. This will generally require further clean-up prior to quantification, dependent on the vitamin D concentration and level of coextractives.

Total Lipid Extraction

The risks of thermal equilibration with previtamin D and potential degradation of vitamin D may be avoided through an initial total lipid extraction. A variety of solvent systems have been recommended depending upon sample type and fat content. A consequence of employing such a technique for foods is the need to isolate the vitamin components from high-molecular-weight fractions (triglycerides, phospholipids) by gel permeation and/or adsorption chromatography. Various protocols have been advocated which, while successful, contribute complexities to the assays which are difficult for most quality control laboratories to manage.

This extraction route has been commonly applied to low-fat clinical specimens (such as plasma), and fat-reduced fortified foods (e.g., skim milk). It is also useful when vitamin D and its metabolites are to be estimated concurrently. The literature consensus seems to support the view that saponification is more appropriate when complex and poorly characterized foods are to be assayed.

The concentrations of cholecalciferol are generally low, even after isolation of the crude extracts. Further clean-up to remove substantial quantities of sterols and other unsaponifiable material is usually incorporated prior to analysis. Such strategies may be as simple as cholesterol precipitation or as complex as multistage semipreparative chromatography, depending on both sample type and the sophistication of subsequent analytical techniques. Often at this stage, analysts may expediently utilize the convenience of prepacked solid-phase extraction cartridges. Usually, the adsorption mode with activated silica is selected, producing an extract of higher vitamin D content and fewer potential interferences. These disposable cartridges are now widely replacing the earlier techniques of open-column or thin-layer chromatography (although the latter is still occasionally used for clean-up or qualitative 'spot testing').

Chromatography

Quantification of enriched vitamin extracts is achievable by applying gas–liquid chromatography (GLC) techniques, although the thermal instability of vitamin D at operating temperatures results in formation of pyro and isopyro peaks of both parent or derivative forms. This problem has been successfully avoided by prior conversion to the thermostable isotachysterol isomers. Modern GLC now benefits from the use of capillary columns and ultrasensitive mass-selective spectrometric detectors, but the extensive manipulative procedures still make GLC less attractive than other technologies for routine food analysis. (*See* **Chromatography**: Gas Chromatography.)

High-performance liquid chromatography (HPLC) has led to continuing improvements in the assay of vitamin D (and its metabolites) in foods and has now superseded GLC. While additional clean-up is not always unavoidable, derivatization procedures are unnecessary in the majority of schemes. Furthermore, the ambient and nondestructive features inherent in HPLC are more compatible with the lability properties of this, and other, vitamins. Recently, variants of micellar capillary electrophoresis (CE) have been applied to fat-soluble vitamin separations. Although representing an alternative to HPLC techniques, the application of CE to the hydrophobic vitamins is currently in its infancy. (*See* **Chromatography**: High-performance Liquid Chromatography.)

In the absence of useful native fluorescence or stable electrochemical viewing modes and the current infancy of commercial online liquid chromatography–mass spectroscopy (LCMS) interfaces, the use of UV spectrophotometric detection is universal.

The modest spectral properties of vitamin D often require a semipreparative HPLC fractionation step before an analytical HPLC stage can be undertaken. Spectral selectivity is further gained by the judicious use of either wavelength rationing or full-spectrum (e.g., diode array) detection. Alternative choices to attain additional specificity and selectivity include the use of offline competitive protein-binding assay or the precolumn conversion of cholecalciferol to its bathochromically shifted isotachysterol.

In view of the differing selectivities of normal-phase and reversed-phase HPLC, many researchers have concluded that the assay benefits by combining the two modes. This multidimensional technique exploits the fact that cholecalciferol and ergocalciferol are unresolved on silica columns yet are completely separated using C_{18} reversed-phase columns. There are a number of reported methods which employ reversed-phase chromatography during clean-up, either low-pressure or HPLC, followed by normal-phase quantitation. This regimen is generally of use where both calciferols are known to coexist in the sample or where the sample is not well defined in terms of its vitamin D content. Alternatively, normal-phase silica chromatography can be used during the preparative step and the two vitamins

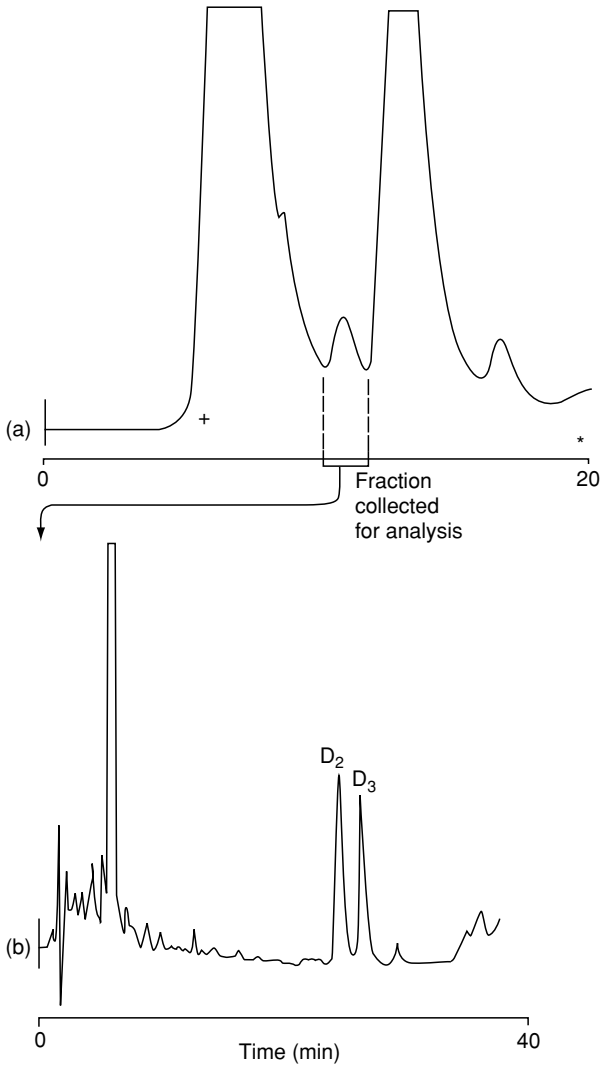

Figure 4 High-performance liquid chromatography (HPLC) chromatograms showing the analysis of cholecalciferol in halibut liver oil. An internal standard of ergocalciferol (D_2) was used throughout the analysis. (a) The semipreparative clean-up stage. A Brownlee (25 cm × 10 mm) silica column was used with a mobile phase of 1% 2-propanol in hexane, flowing at 3 ml min^{-1}. Detection was by ultraviolet at 280 nm (0.04 aufs). The unsaponifiable fraction from 2.5 g of fish oil was dissolved in 1.0 ml of mobile phase and injected. The fraction collected and evaporated for quantitative analysis is shown. The elution positions of previtamin D_3 (+) and provitamin D_3 (*) are shown. (b) A reversed-phase C_{18} column (Waters 5 µm Radial-PAK) was used for analyte quantification. A mobile phase of methanol:water:tetrahydrofuran (90:8:2 v/v/v) at a flow rate of 2.0 ml min^{-1} was used. The injection volume was 100 µl, representing the cholecalciferol (D_3) content of 0.25 g of oil. Detection was at 280 nm (0.005 aufs).

collected in a single fraction before application to a reversed-phase column. This allows one vitamin to be used as an internal standard for the other, greatly decreasing the cumulative assay errors introduced though potential isomerization and manipulative losses. The method can be used where only one vitamin is present, as is usually encountered in the food industry. Chromatograms of a fish oil taken through this analytical procedure are shown in Figure 4. The endogenous vitamin is cholecalciferol, allowing internal standardization with ergocalciferol.

The elution positions of previtamin D_3 and provitamin D_3 are shown in Figure 4a. They are well separated from vitamin D_3, thus allowing potential differentiation of the species. Other calciferols shown in Figures 2 and 3 can be similarly resolved. On silica columns, using hydrocarbon/alcohol mobile phases, the major isomers elute in the sequence previtamin D_3 < *trans*-vitamin D_3 < lumisterol$_3$ < isotachysterol$_3$ < vitamin D_3 < tachysterol$_3$ < provitamin D_3 < hydroxymetabolites. The pyrocalciferols elute close to lumisterol$_3$. The most likely interfering isomer is tachysterol$_3$ which elutes close to vitamin D_3 in both normal and reversed-phase analyses. Interestingly, the retention sequence of previtamin D_3, vitamin D_3, and provitamin D_3 is the same, regardless of which separation mode is used.

See also: **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **Food Fortification**; **Immunoassays**: Radioimmunoassay and Enzyme Immunoassay; **Retinol**: Physiology

Further Reading

Ball GFM (1988) *Fat-Soluble Vitamin Assays in Food Analysis. A Comprehensive Review*. London: Elsevier.

Bates CJ (2000) Vitamins: fat and water soluble: analysis. In: Meyers RA (ed.) *Encyclopedia of Analytical Chemistry*, pp. 1–35. Chichester: John Wiley.

Belitz H-D and Grosch W (1987) *Food Chemistry*, pp. 305–319. Berlin: Springer-Verlag.

Collins ED and Norman AW (1991) Vitamin D. In: Machlin LJ (ed.) *Handbook of Vitamins*, pp. 59–98. New York: Marcel Dekker.

Eitenmiller RR and Landon WO (1999) Vitamin D. In: Eitenmiller R and Landon WO (eds) *Vitamin Analysis for the Health and Food Sciences*, pp. 77–107. Boca Raton, FL: CRC Press.

Friedrich W (1988) Vitamin D. In: *Vitamins*, pp. 143–216.Berlin: de Gruyter.

Indyk H and Woollard DC (1984) The determination of vitamin D by high performance liquid chromatography. *New Zealand Journal of Dairy Science and Technology* 19: 1–6.

Jones G, Seamark DA, Trafford DJH and Makin HLJ (1985) Vitamin D: cholecalciferol, ergocalciferol and hydroxylated metabolises. In: De Leenheer AP, Lambert WE and De Ruyter MGM (eds) *Modern Chromatographic Analysis of the Vitamins*, pp. 73–128. New York: Marcel Dekker.

Lawson DEM (1985) Vitamin D. In: Diplock AT (ed.) *Fat Soluble Vitamins. Their Biochemistry and Applications*, Chapter 2, pp. 76–153. London: Heinemann.

Luque de Castro MD, F-Romero JM, O-Boyer F and Quesada JM (1999) Determination of vitamin D_3 metabolites: state-of-the-art and trends. A review. *Journal of Pharmaceutical and Biomedical Analysis* 20: 1–17.

Macrae R (1988) *HPLC in Food Analysis*, 2nd edn. London: Academic Press.

Mattila PH, Piironen VI, Uusi-Rauva EJ and Koivistoinen PE (1996) New analytical aspects of vitamin D in foods. *Food Chemistry* 57: 95–99.

Parrish DB (1979) Determination of vitamin D in foods. *CRC Critical Reviews in Food Science and Nutrition* 12: 29–57.

Strohecker R and Henning HM (eds) (1965) Vitamin D. In: *Vitamin Assay, Tested Methods*, pp. 254–282. Verlag Chemie.

Physiology

A W Norman, University of California-Riverside, Riverside, CA, USA

Background

Vitamin D is essential for life in higher animals. Classically, it has been shown to be one of the most important biological regulators of calcium homeostasis. It has been established that these important biological effects are only achieved as a consequence of the metabolism of vitamin D into a family of daughter metabolites, including the two key kidney-produced metabolites $1\alpha,25(OH)_2$-vitamin D_3 ($1\alpha,25(OH)_2D_3$) and $24R,25(OH)_2$-vitamin D_3 ($24R,25(OH)_2D_3$). $1\alpha,25(OH)_2D_3$ is considered to be a steroid hormone and there is increasing evidence that $24R,25(OH)_2D_3$ is also a steroid hormone.

It has become increasingly apparent since the 1980s that $1\alpha,25(OH)_2D_3$ also plays an important multidisciplinary role in tissues not primarily related to mineral metabolism, e.g., the hematopoietic system, effects on cell differentiation and proliferation including important interactions with keratinocytes and cancer cells, and participation in the processes of parathyroid hormone and insulin secretion. The purpose of this chapter is to provide a succinct overview of our current understanding of the important nutritional substance vitamin D and the mechanisms by which its daughter hormone $1\alpha,25(OH)_2D_3$ mediates biological responses.

Historical Review

The first scientific description of rickets, which is the hallmark of a vitamin D-deficiency, was provided in the seventeenth century by both Dr. Daniel Whistler (1645) and Professor Francis Glisson (1650). The major breakthrough in understanding the causative factors of rickets was the development of nutrition as an experimental science and the appreciation of the existence of vitamins. Considering the fact that now we accept that the biologically active form of vitamin D is a steroid hormone, it is somewhat ironic that vitamin D, through a historical accident, became classified as a vitamin. It was in 1919/20 that Sir Edward Mellanby, working with dogs raised exclusively indoors (in the absence of sunlight or ultraviolet light), devised a diet which allowed him to establish unequivocally that rickets was caused by a deficiency of a trace component present in the diet. In 1921, he wrote, 'The action of fats in rickets is due to a vitamin or accessory food factor which they contain, probably identical with the fat-soluble vitamin.' Furthermore, he established that cod-liver oil was an excellent antirachitic agent; this ultimately led to the antirachitic factor being classified as a vitamin.

The chemical structures of the vitamins D were determined in the 1930s in the laboratory of Professor A. Windaus at the University of Gottingen. Vitamin D_2, which could be produced by ultraviolet irradiation of ergosterol, was chemically characterized in 1932. Vitamin D_3 was not chemically characterized until 1936, when it was shown to result from the ultraviolet irradiation of 7-dehydrocholesterol. Virtually simultaneously, the elusive antirachitic component of cod-liver oil was shown to be identical to the newly characterized vitamin D_3. These results clearly established that the antirachitic substance vitamin D was chemically a steroid, more specifically a seco-steroid (see below).

The modern era of vitamin D began during the period of 1965–70 with the discovery and chemical characterization of $1\alpha,25(OH)_2D_3$ and its nuclear receptor, the VDR_{nuc}.

Chemistry of Vitamin D

Vitamin D_3 is the naturally occurring form of vitamin D and is produced from 7-dehydrocholesterol (see **Figure 1**). Vitamin D_2 is a synthetic form of vitamin D that is produced by irradiation of the plant yeast steroid, ergosterol.

The structures of vitamin D_3 (cholecalciferol) and its provitamin 7-dehydrocholesterol are presented in **Figure 1**. Vitamin D is a generic term and indicates a molecule of the general structure shown for rings A, B,

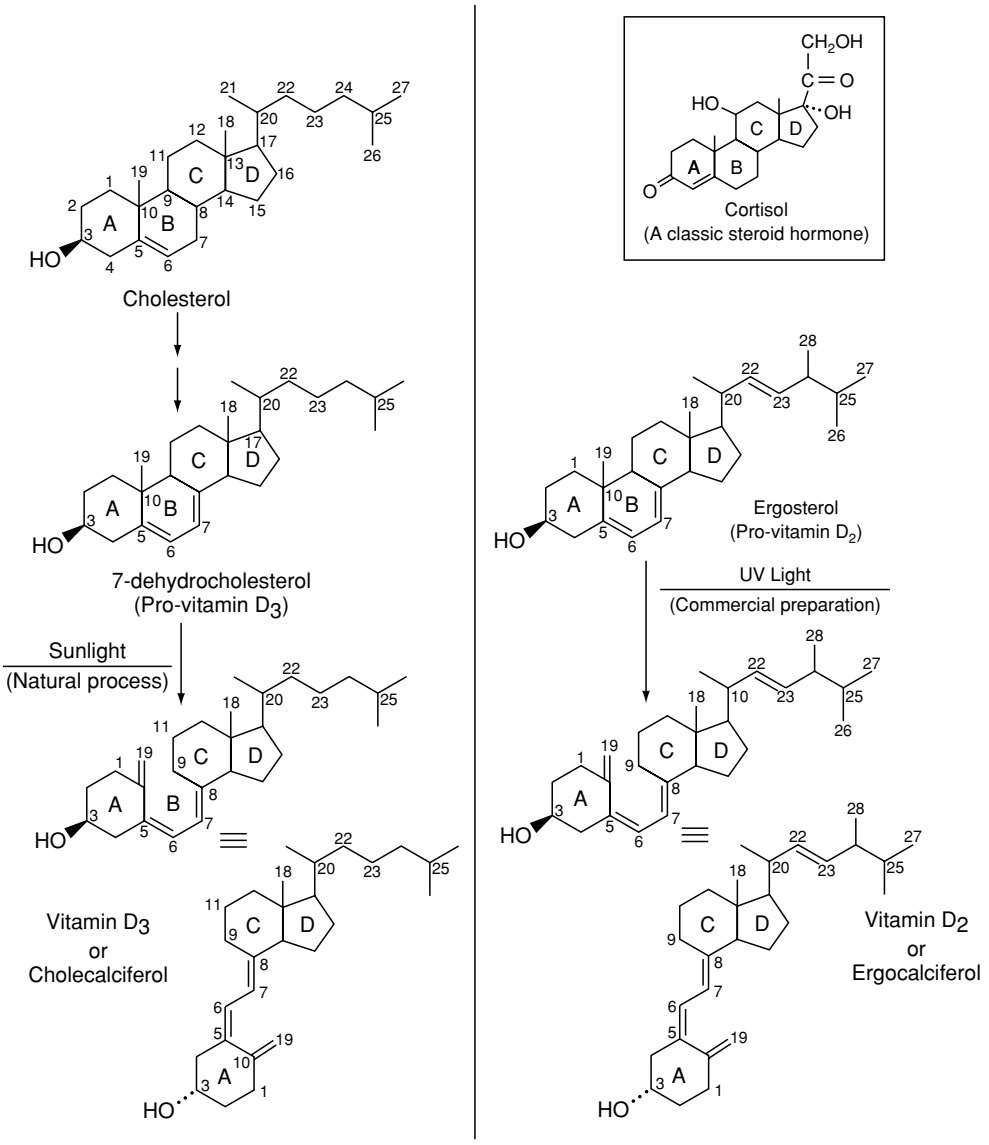

Figure 1 Structural relationship of vitamin D_3 (cholecalciferol) and vitamin D_2 (ergocalciferol) with their respective provitamins (7-dehydrocholesterol and ergosterol), cholesterol, and a classic steroid hormone, cortisol (see inset box). The two structural representations presented at the bottom for both vitamin D_3 and vitamin D_2 are equivalent; these are simply different ways of drawing the same molecule. It should be noted that vitamin D_3 is the naturally occurring form of the vitamin; it is produced from 7-dehydrocholesterol, which is present in the skin, by the action of sunlight. Vitamin D_2 (which is equivalently potent to D_3 in humans and many mammals, but not birds) is produced commercially by the irradiation of the plant sterol ergosterol with ultraviolet light.

C, and D with differing side-chain structures. The A, B, C, and D ring structure is derived from the cyclopentanoperhydrophenanthrene ring structure for steroids. Technically, the steroid vitamin D is classified as a seco-steroid. Seco-steroids are those in which one of the rings has been broken; in vitamin D, the 9,10 carbon–carbon bond of ring B is broken, and it is indicated by the inclusion of '9,10-seco' in the official nomenclature.

Vitamin D (synonym calciferol) is named according to the revised rules of the International Union of Pure and Applied Chemistry (IUPAC). Because vitamin D is derived from a steroid, the structure retains its numbering from the parent compound cholesterol (see Figure 1). Asymmetric centers are designated by using the R,S notation; the configuration of the double bonds is indicated as E for 'entgegen' or *trans*, and Z for 'zusammen' or *cis*. Thus, the official name of vitamin D_3 is 9,10-seco (5Z, 7E)-5,7,10(19)cholestatriene-3β-ol. Vitamin D_2 differs from D_3 by virtue of the presence of a 22-ene and 24-methyl group in the side-chain. The official

name of vitamin D_2 is 9,10-seco(5Z,7E)-5,7,10(19),22-ergostatetraene-3β-ol.

Vitamin D_3 can be produced photochemically by the action of sunlight or ultraviolet light from the precursor sterol 7-dehydrocholesterol that is present in the epidermis or skin of most higher animals. The chief structural prerequisite of a provitamin D is that it be a sterol with a Δ5–7 diene double bond system in ring B (see Figure 1). The conjugated double bond system in this specific location of the molecule allows the absorption of light quanta at certain wavelengths in the UV range; this can be readily provided in most geographical locations by natural sunlight. This initiates a complex series of transformations (partially summarized in Figure 1) that ultimately result in the transformation into vitamin D_3. Thus, it is important to appreciate that vitamin D_3 can be endogenously produced and that as long as the animal (or human) has access on a regular basis to sunlight, there is no dietary requirement for this vitamin.

Physiology and Biochemistry

Vitamin D Endocrine System

Vitamin D_3 is not known to have any intrinsic biological activity itself. It is only after vitamin D_3 is metabolized first into 25(OH)D_3 in the liver and then into 1α,25(OH)$_2D_3$ and 24R,25(OH)$_2D_3$ by the kidney, that biologically active molecules are produced. *In toto*, some 37 vitamin D_3 metabolites have been isolated and chemically characterized. Figure 2 illustrates the concept of the vitamin D endocrine system. The elements of the vitamin D endocrine system include the following: (1) in the skin, photoconversion of 7-dehydrocholesterol to vitamin D_3 or dietary intake of vitamin D_3; (2) metabolism of vitamin D_3 by the liver to 25(OH)D_3, which is the major form of vitamin D circulating in the blood compartment; (3) conversion of 25(OH)$_2D_3$ by the kidney (functioning as an endocrine gland) to produce the two principal dihydroxylated metabolites,

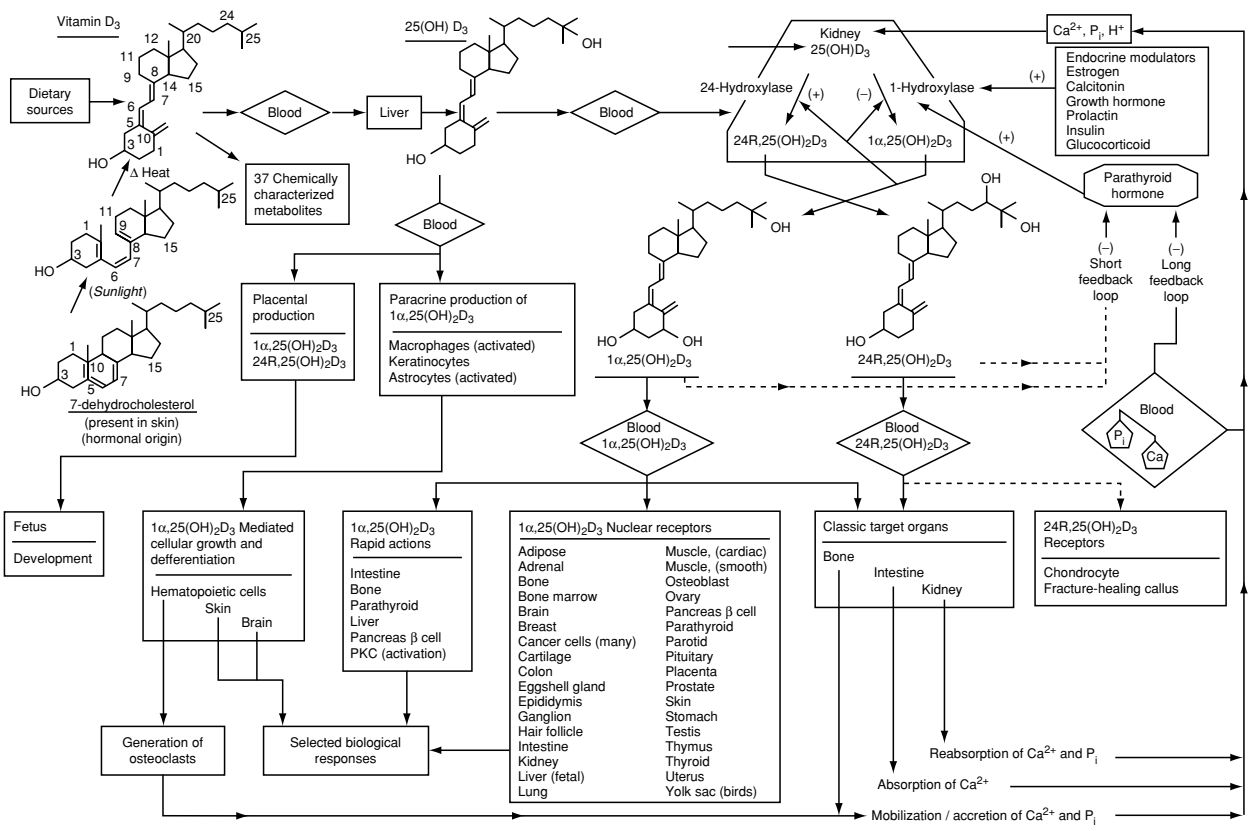

Figure 2 Summary of the vitamin D endocrine system. In this system, the biologically inactive vitamin D_3 is activated, first in the liver to generate 25(OH)D_3, which is then converted by the endocrine gland, the kidney, to the hormones 1α,25(OH)$_2D_3$ and 24R,25(OH)$_2D_3$. P_i, inorganic phosphate. There is currently much research being conducted to understand structure–function relationships with respect to chemical synthesis of analogs of 1α,25(OH)$_2D_3$ and their interaction with the vitamin D endocrine system. The objective is to develop new drug forms of 1α,25(OH)$_2D_3$.

$1\alpha,25(OH)_2D_3$ and $24R,25(OH)_2D_3$; (4) systemic transport of the dihydroxylated metabolites $24R,25(OH)_2D_3$ and $1\alpha,25(OH)_2D_3$ to distal target organs; (5) binding of the dihydroxylated metabolites, particularly $1,25(OH)_2D_3$, to either a nuclear receptor or membrane receptor at the target organs followed by the subsequent generation of appropriate biological responses (see also **Figure 3**). An additional key component in the operation of the vitamin D endocrine system is the plasma vitamin D binding protein, which carries vitamin D_3 and all its metabolites to their various target organs.

The three enzymes responsible for the conversion of vitamin D_3 into its two key daughter metabolites include the hepatic vitamin D_3-25-hydroxylase and the two kidney enzymes, the $25(OH)D_3$-1α-hydroxylase and the $25(OH)D_3$-24R-hydroxylase. All three enzymes have been demonstrated to be cytochrome P450 mixed-function oxidases. Both renal enzymes are localized in mitochondria of the proximal tubules of the kidney. The $25(OH)D_3$-1α-hydroxylase has been cloned and the specific site of mutations (which result in the appearance of vitamin D-resistant rickets, type I) identified. Also, the $25(OH)D_3$-24R-hydrodroxylase and vitamin D_3-25-hydroxylase have been cloned.

The most important point of regulation of the vitamin D endocrine system occurs through the stringent control of the activity of the renal $25(OH)D_3$-1α-hydroxylase. In this way, the production of the hormone $1\alpha,25(OH)_2D_3$ can be modulated according to the calcium and other endocrine needs of the organism. The chief regulatory factors are $1\alpha,25(OH)_2D_3$ itself, parathyroid hormone (PTH), and the serum concentrations of calcium and phosphate. Probably the most important determinant of the 1α-hydroxylase is the vitamin D status of the animal. When the circulating concentration of $1\alpha,25(OH)_2D_3$ is low, production of $1\alpha,25(OH)_2D_3$ by the kidney is high; when the circulating concentration of $1\alpha,25(OH)_2D_3$ is high, the output of $1\alpha,25(OH)_2D_3$ by the kidney is sharply reduced.

The process of nuclear receptor-mediated regulation of gene transcription is exquisitely dependent upon the complimentary relationship between the unoccupied receptor and its cognate ligand. Thus, the unoccupied receptor is largely incompetent to engage in a productive fashion with the

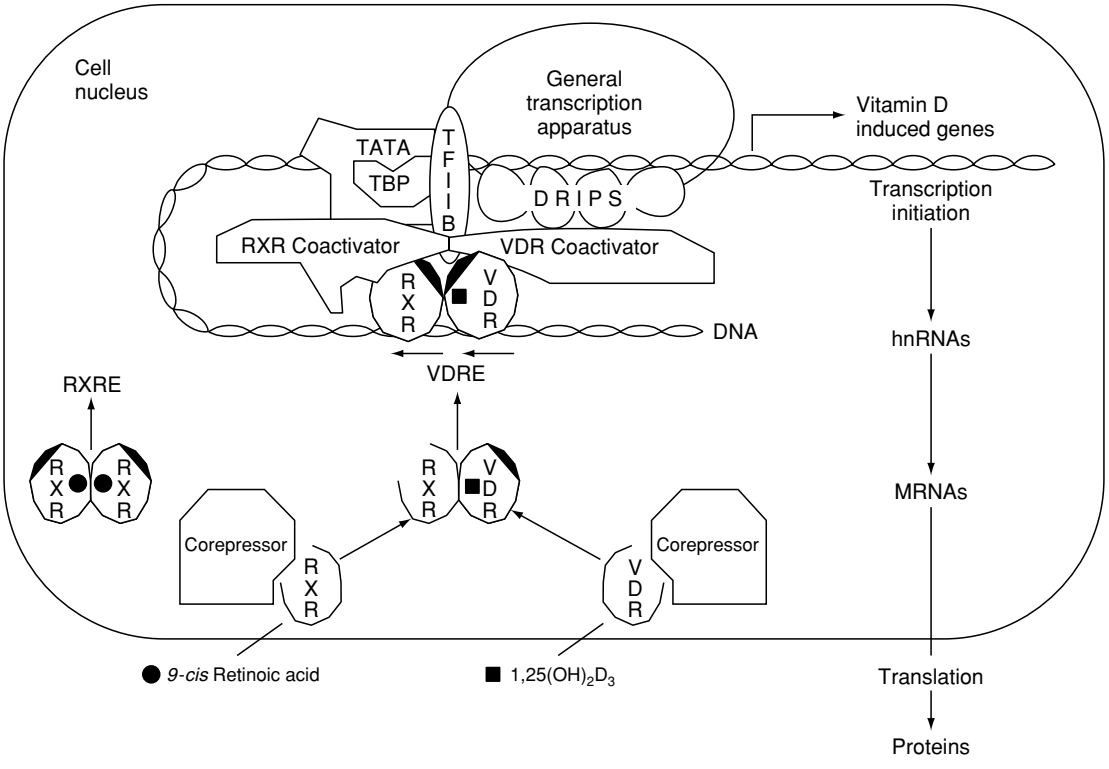

Figure 3 Model of $1\alpha,25(OH)_2D_3$ and VDR_{nuc} activation of transcription. The VDR, after binding its cognate ligand $1\alpha,25(OH)_2D_3$, forms a heterodimer with the retinoid X receptor. This heterodimer complex then interacts with the appropriate vitamin D response element on the promoter of genes (in specific target cells) which are destined to be up- or downregulated. The heterodimer – DNA complex then recruits the necessary coactivator proteins, such as TATA, TBP, TFIIB and other proteins to generate a competent transcriptional complex capable of modulating mRNA production.

transcriptional machinery to effect meaningful regulation of gene transcription. It is only after the ligand–receptor complex has formed (which results in conformational changes in the receptor protein) that a functional receptor protein is generated. Thus acquisition of a detailed understanding of the complementarity of the ligand shape with that of the interior surface of the nuclear VDR receptor ligand binding domain is believed to be the key not only to understanding the structural basis of receptor action and its formation of heterodimers and interactions with coactivators (see **Figure 3**), but also to designing new drug forms of the various hormones, including $1\alpha,25(OH)_2D_3$.

The receptors for all steroid hormones (estrogen, progesterone, testosterone, cortisol, and aldosterone) and the nuclear receptors for $1\alpha,25(OH)_2D_3$, retinoic acid, and thyroid hormone are all members of the same super gene family (*See* **Hormones**: Steroid Hormones, Figure 5); accordingly, there is a high level of conservation in their amino acid sequences, particularly in their DNA and ligand binding domains. The X-ray crystallographic structures of the ligand-binding domains (LBD) of the thyroid hormone receptor, the retinoic acid receptor, the estrogen receptor, the ligand binding and coactivator assembly of the peroxisome proliferator-activated receptor-γ, and the progesterone receptor have all been determined with their respective ligands bound. Also, the crystal structure of the nuclear receptor for vitamin D bound to its natural ligand $1\alpha,25(OH)_2D_3$ has been determined at a 1.8-Å resolution. The VDR_{nuc} LBD structure, like the other nuclear receptors, consists of 12 α-helices that are arranged to create a three-layer sandwich that completely encompasses the ligand $1\alpha,25(OH)_2D_3$ in a hydrophobic core. The secondary and tertiary structural features of this group of proteins have been found to be remarkably similar.

Nutritional Aspects

Recommended Dietary Allowance

The World Health Organization (WHO) has responsibility for defining the International Unit (IU) of vitamin D_3. Their most recent definition, provided in 1950, stated that 'the International Unit of vitamin D recommended for adoption is the vitamin D activity of 0.025 μg of the international standard preparation of crystalline vitamin D_3.' Thus, 1.0 IU of vitamin D_3 is 0.025 μg, which is equivalent to 65.0 pmol. With the discovery of the metabolism of vitamin D_3 to other active seco-steroids, particularly $1\alpha,25(OH)_2D_3$, it was recommended that 1.0 Unit of $1\alpha,25(OH)_2D_3$ be set equivalent in molar terms to that of the parent vitamin D_3. Thus, 1.0 Unit of $1\alpha,25(OH)_2D_3$ has been operationally defined to be equivalent to 65 pmol.

The vitamin D requirement for healthy adults has never been precisely defined. Since vitamin D_3 is produced in the skin after exposure to sunlight, the human does not have a requirement for vitamin D when sufficient sunlight is available. Man's tendency to wear clothes, to live in cities where tall buildings block adequate sunlight from reaching the ground, to live indoors, to use synthetic sunscreens that block ultraviolet rays, and to live in geographical regions of the world that do not receive adequate sunlight all contribute to the inability of the skin to biosynthesize sufficient amounts of vitamin D_3. Thus, vitamin D does become an important nutritional factor in the absence of sunlight. It is known that a substantial proportion of the US population is exposed to suboptimal levels of sunlight. This is particularly true during the winter months. Under these conditions, vitamin D becomes a true vitamin, and therefore must be supplied in the diet on a regular basis.

Since vitamin D_3 can be endogenously produced by the body and since it is retained for long periods of time by vertebrate tissue, it is difficult to determine with precision the minimum daily requirements for this seco-steroid. The requirement for vitamin D is also known to be dependent on age, sex, degree of exposure to the sun, season, and the amount of pigmentation in the skin.

In the USA, adequate amounts of vitamin D can be readily obtained from the diet and/or from casual exposure to sunlight. The UV exposure can be as little as three 20-min periods of exposure of the face and hands to ambient sunlight per week. However, in some parts of the world where food is not routinely fortified and sunlight is often limited during some periods of the year, obtaining adequate amounts of vitamin D becomes more of a problem. As a result, the incidence of rickets in these countries is higher than in the United States.

Many countries as well as the WHO have prepared recommendations on the daily intake of essential nutrients, including vitamin D. The current 'adequate intake' allowance of vitamin D recommended in 1998 by the Food and Nutrition Board of the US Institute of Medicine is 200 IU per day (5 μg per day) for infants, children, and adult males and females up to age 51. For males and females aged 51–70 or >70, the adequate indicated level is set at 400 IU per day (10 μg per day) or 600 IU (15 μg per day), respectively. The adequate allowance for pregnancy and lactation is set at 200 IU per day (5 μg per day). The recommendations of the UK and Canada for adults are about 50% lower than of the USA. Australia, which is generally

sunny, does not recommend daily supplements of vitamin D.

Food Sources

Animal products constitute the bulk source of vitamin D that occurs naturally in unfortified foods. Marine fish, such as herring, salmon, and sardines, and fish liver oils are good sources of vitamin D_3. Small quantities of vitamin D_3 are also derived from eggs, veal, beef, butter, and vegetable oils, while plants, fruits, and nuts are extremely poor sources of vitamin D. In the USA, artificial vitamin D_3 fortification of foods such as milk (both fresh and evaporated), margarine and butter, cereals, and chocolate mixes help in meeting the recommended dietary allowance. Vitamin D_2 was used in the period of 1940–60 as a food supplement for vitamin D activity.

Functions of Vitamin D

Classical target tissues In the classical target tissues such as intestine, bone, and kidney, $1,25(OH)_2D_3$, largely in conjunction with PTH, serves to regulate mineral homeostasis such that serum calcium and phosphorus levels are maintained within a physiological range that can support normal mineralization of bone.

In the intestine, $1,25(OH)_2D_3$ primarily stimulates the active transport of Ca^{2+} and inorganic phosphate, P_i, via mechanisms that involve a calcium-binding protein, termed calbindin D. Available evidence suggests that calbindin D may be involved in protecting the cell (as a buffer) against large fluxes of Ca^{2+} which result from active transport induced by $1,25(OH)_2D_3$.

In addition to the genomic effects of $1,25(OH)_2D_3$, recent studies have shown that $1,25(OH)_2D_3$ can stimulate the rapid (2–4 min) transport of calcium, termed transcaltachia, via a receptor-mediated process, albeit one independent of gene activation.

In the kidney, $1,25(OH)_2D_3$ functions in concert with PTH to enhance renal Ca^{2+} reabsorption, besides regulating its own biosynthesis by feedback inhibition of renal 1-hydroxylase.

$1\alpha,25(OH)_2D_3$ plays an important role in bone growth, development, and differentiation, and supports bone mineralization indirectly by supplying the minerals calcium and phosphorus via their enhanced intestinal absorption. In addition, $24,25(OH)_2D_3$ has been shown to promote the bone-mineralization process and to participate in the process of fracture-healing.

Nonclassical target tissues The $1,25(OH)_2D_3$ hormone promotes differentiation of cells in the hemopoietic system. Such effects of $1,25(OH)_2D_3$ offer a therapeutic prospect for leukemia.

In the immune system, $1,25(OH)_2D_3$ acts as an immunomodulator, regulating the functional performance of cells involved in the immune response.

Insulin production by the endocrine pancreas is influenced by vitamin D status in that the blunted secretion of insulin and impaired glucose tolerance seen in vitamin D-deficient conditions are corrected by treatment with vitamin D_3 and/or $1,25(OH)_2D_3$.

Myopathy and abnormalities in muscle contraction seen in patients afflicted with metabolic bone disease are amenable to correction with vitamin D therapy.

$1\alpha,25(OH)_2D_3$ induces differentiation of keratinocytes in skin and exerts antiproliferative effects on these epithelial cells. Such an effect of $1,25(OH)_2D_3$ in the skin has prompted the use of $1,25(OH)_2D_3$ analogs for treatment of psoriasis, which is a hyperproliferative disease of the skin.

Disease States in Humans Related to Vitamin D

In humans, diseases related to vitamin D can arise because of (1) altered availability of vitamin D; (2) altered hepatic conversion of vitamin D_3; (3) impaired renal metabolism of $25(OH)D_3$, or (4) variation in end organ responsiveness to $1,25(OH)_2D_3$ (see Figure 4).

Renal Disorders

Chronic renal failure, also known as renal osteodystrophy, is characterized by impaired renal production of $1,25(OH)_2D_3$ and intestinal malabsorption of calcium, which can often lead to derangements in skeletal metabolism and hyperparathyroidism. These symptoms are alleviated by $1,25(OH)_2D_3$ administration. In 1977, in the USA, the Food and Drug Administration approved the prescription use of $1,25(OH)_2D_3$ (calcitriol or Rocaltrol®) for renal osteodystrophy.

Vitamin D-resistant Rickets

Also known as familial X-linked hypophosphatemic rickets, vitamin D-resistant rickets is characterized by a primary phosphate leak in the kidney, skeletal deformities, and hypophosphatemia. A combination of oral phosphate and $1,25(OH)_2D_3$ is effective in treating these patients.

Vitamin D-dependent Rickets

Vitamin D-dependent rickets is also referred to as hereditary hypocalcemic vitamin D-resistant rickets, and is classified into Type I and Type II disease states.

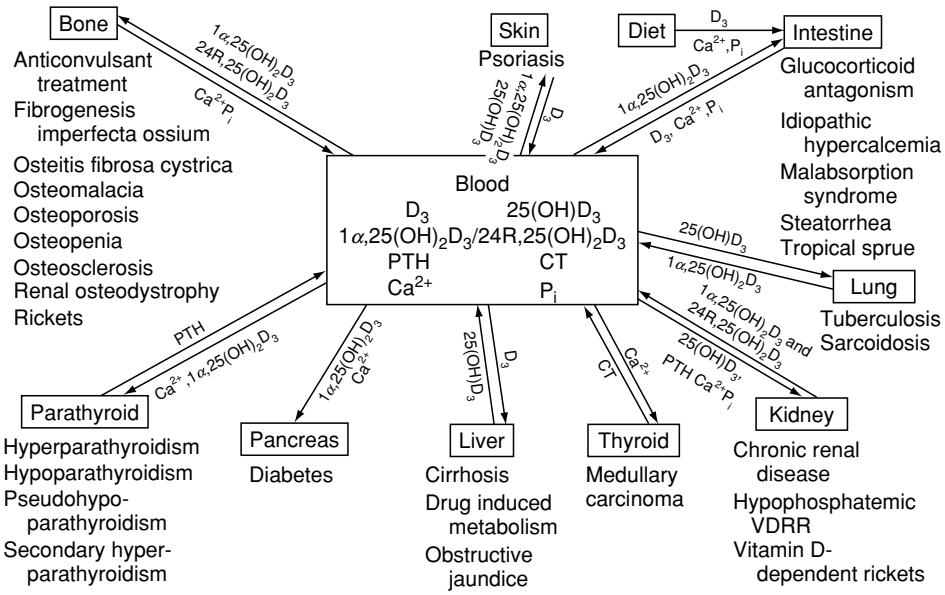

Figure 4 Human disease states related to the vitamin D endocrine system. Under the boxed headings (e.g., parathyroid, liver, bone, etc.) are listed disease states occurring in man which have been shown, or are believed, to have some functional linkage between some aspect of the vitamin D endocrine system and that particular organ. The information associated with the arrows indicates the direction of flow of calcium, phosphate or the calcium-regulating hormones (CT (calcitonin), PTH (parathyroid hormone), vitamin D_3, 1α, $25(OH)_2D_3$, 24R, $25(OH)_2D_3$). Ca^{2+}, calcium; P_i, inorganic phosphate; VDRR, vitamin D-resistant rickets.

Type I rickets is believed to arise as a result of an inborn error in the renal 1-hydroxylase enzyme. The clinical features include hypocalcemia, hypophosphatemia, and several rachitic lesions. These symptoms can be treated with pharmacological doses of vitamin D_3 or low doses of $1,25(OH)_2D_3$.

Point mutations in the vitamin D receptor gene have been shown to be responsible for the defective receptors seen in these children with Type II rickets. The clinical manifestations are defective bone mineralization, decreased intestinal calcium absorption, hypocalcemia, and increased serum $1,25(OH)_2D_3$ levels. To date, a point mutation in a steroid receptor gene resulting in the loss of functional activity has been demonstrated only for the vitamin D receptor and is therefore unique in this respect.

Diseases of Parathyroid

Hypoparathyroidism Hypoparathoidism exhibits hypocalcemia as a major clinical consequence and is corrected with large doses of vitamin D or physiological doses of $1,25(OH)_2D_3$.

Hyperparathyroidism Increased $1,25(OH)_2D_3$ levels, enhanced intestinal absorption of calcium (contributing to hypercalciuria), and nephrolithiasis are typical of this disorder.

Pseudohypoparathyroidism Pseudohypoparathyroidism results from a state of resistance to PTH. The biochemical abnormalities are hypocalcemia, hyperphosphatemia, elevated serum PTH, and decreased serum $1,25(OH)_2D_3$ levels.

Bone disorders Clinically, a deficiency in vitamin D manifests as rickets in children and osteomalacia in adults. Hypocalcemia, hypophosphatemia, increased serum alkaline phosphatase, and decreased serum $25(OH)D_3$ levels are some of the salient biochemical abnormalities, all of which can be ameliorated by vitamin D administration.

Conclusion

Current evidence supports the concept that the classical biological actions of the nutritionally important fat-soluble vitamin D in mediating calcium homeostasis are supported by a complex vitamin D endocrine system which coordinates the metabolism of vitamin D_3 into $1α,25(OH)_2D_3$ and $24R,25(OH)_2D_3$. It is now clear that the vitamin D endocrine system embraces many more target tissues than simply the intestine, bone, and kidney. Notable additions to this list include the pancreas, pituitary, breast tissue, placenta, hematopoietic cells, skin, and cancer cells of various origins. Key advances in understanding the mode of action of the $1α,25(OH)_2D_3$ have been made by a thorough study of nuclear receptors as well as emerging studies describing a membrane receptor for this steroid hormone. Integral to

these observations are efforts to define the signal transduction systems, which are subservient to the nuclear and membrane receptors for $1\alpha,25(OH)_2D_3$, and to obtain a thorough study of the tissue distribution and subcellular localization of the gene products induced by this steroid hormone. There are clinical applications for $1\alpha,25(OH)_2D_3$ or related analogs for treatment of the bone diseases of renal osteodystrophy and osteoporosis, psoriasis, and hypoparathyroidism; other clinical targets for $1\alpha,25(OH)_2D_3$ currently under investigation include its use in leukemia, breast, prostate, and colon cancer as well as use as an immunosuppressive agent.

See also: **Cholecalciferol**: Properties and Determination; **Dietary Reference Values**; **Hormones**: Steroid Hormones; **Renal Function and Disorders**: Nutritional Management of Renal Disorders

Further Reading

Bikle DD (1995) Vitamin D: New actions, new analogs, new therapeutic potential: update 1995. *Endocrine Reviews* 4: 77–83.

Bouillon R, Okamura WH and Norman AW (1995) Structure–function relationships in the vitamin D endocrine system. *Endocrine Reviews* 16: 200–257.

Haussler MR, Whitfield GK, Haussler CA *et al*. (1998) The nuclear vitamin D receptor: Biological and molecular regulatory properties revealed. *Journal of Bone and Mineral Research* 13: 325–349.

Norman AW (2001) Vitamin D. In: Bowman B and Russell R (eds) *Present Knowledge in Nutrition (PKN)*, 8th edn., pp. 134–143. Washington, DC: International Life Sciences Institute.

Norman AW and Litwack G (1997) Calcium-regulating hormones: Vitamin D, parathyroid hormone, and calcitonin. In: *Hormones*, 2nd edn., pp. 251–280. San Diego, CA: Academic Press.

Norman AW, Henry HL, Bishop JE *et al*. (2001) Different shapes of the steroid hormone $1\alpha,25(OH)_2$-vitamin D_3 act as agonists for two different receptors in the vitamin D endocrine system to mediate genomic and rapid responses. *Steroids* 66(3–5): 147–158.

Reichel H, Koeffler HP and Norman AW (1989) The role of the vitamin D endocrine system in health and disease. *New England Journal of Medicine* 320: 980–991.

Rochel N, Wurtz JM, Mitschler A, Klaholz B and Moras D (2000) The crystal structure of the nuclear receptor for vitamin D bound to its natural ligand. *Molecular Cell* 5(1): 173–179.

Weatherman RV, Fletterick RJ and Scanlon TS (1999) Nuclear receptor ligands and ligand-binding domains. *Annual Review of Biochemistry* 68: 559–582.

CHOLESTEROL

Contents

Properties and Determination

A J Sheppard and R G O'Dell, United States Food and Drug Administration, Washington DC, USA
J A T Pennington, National Institute of Health, Bethesda, MD, USA

Background

Cholesterol is a well-known and commonly determined lipid component and an important intermediate in the synthesis of steroid hormones. It is a sterol ($C_{27}H_{45}OH$) that occurs notably in animal fats and oils, bile, gallstones, nerve tissues, blood, brain, plasma, and egg yolk. Cholesterol is the most common animal sterol and is also found in trace amounts in vegetable fats and oils, seaweeds, and green leaves. Cholesterol was first found in gallstones and derives its name from the Greek *kholé* (bile) and *stereos* (solid). The determination of cholesterol in serum and foods is of significance because of the implication of cholesterol in the etiology of arteriosclerosis and coronary heart disease. (*See* **Atherosclerosis; Coronary Heart Disease**: Intervention Studies.)

Structure

A systematic study of the chemistry of cholesterol began at the end of the nineteenth century. The classic work of Wieland, Vindaus, Diels, Rosenheim, and King led to the formulation of the structure of cholesterol in 1932. The fundamental carbon skeleton of the cholesterol molecule is the cyclopentanoperhydrophenanthrene ring. The structure of cholesterol is shown in Figure 1. The hydroxyl group on C3 is connected by a 'solid' bond (β orientation) and the hydrogen by a 'dashed' bond (α orientation), depicting the naturally occurring β-cholesterol. Atoms connected by solid bonds (β orientation) are regarded as projecting in front of the plane of the steroid ring, and those connected by dashed bonds (α orientation) as lying behind the plane. The molecular weight of cholesterol is 384.64. Cholesterol consists of 83.87% carbon, 11.99% hydrogen, and 4.145% oxygen by weight. The formal chemical name of the molecule is cholest-5-en-3β-ol. Because the cholesterol nucleus contains eight centers of asymmetry, approximately 240 isomers of the molecule are possible. However, only two carbon centers (C3 and C5) appear to be involved in naturally occurring cholesterol isomers. In some of the earlier scientific literature, cholesterol is referred to as 'cholesterin.'

Chemical Characteristics

Cholesterol is a glistening, white, soapy, crystalline substance that is practically insoluble in water (about 0.2 mg per 100 ml). It is slightly soluble in alcohol (1.29% w/w at 20 °C) and more soluble in hot alcohol (100 g of saturated 96% alcoholic solution contains 28 g of cholesterol at 80 °C). One gram of the compound dissolves in 2.8 ml of ether, 4.5 ml of chloroform, or 1.5 ml of pyridine. Cholesterol is also soluble in benzene, hexane, petroleum ether, oils, fats, and aqueous solutions of bile salts. It crystallizes easily from absolute alcohol, acetic acid, ether, and similar solvents as colorless rhombic plates with one or more characteristic notches in the corners. Because cholesterol has an unsaturated bond, it will accept up to two halogen atoms. Cholesterol is not saponifiable.

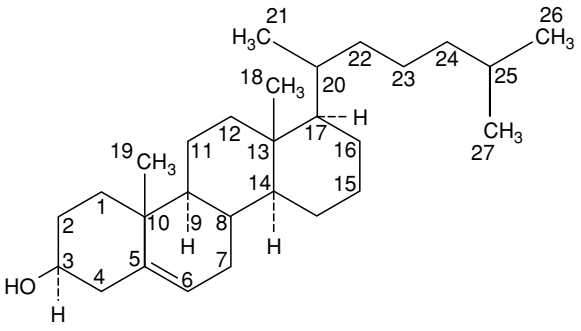

Figure 1 Structure of β-cholesterol showing carbon atom numbering. Reproduced from Cholesterol: Properties and Determination, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Cholesterol gives a number of color reactions that are useful to test for the molecule both qualitatively and quantitatively. The Salkowski reaction produces a series of colors when a chloroform solution of cholesterol is stratified over concentrated sulfuric acid. The acid assumes a yellowish color with a green fluorescence, whereas the chloroform layer first becomes bluish red and then gradually changes to a violet–red. If the chloroform layer is decanted into a porcelain evaporating dish, it changes from violet–red to violet, to green, and then to yellow. Another test, the Liebermann–Burchard reaction, which involves adding acetic anhydride and concentrated sulfuric acid (under conditions as nearly anhydrous as possible) to a chloroform solution of cholesterol, results in an initial blue to violet color that changes to emerald green. Under carefully controlled conditions, the intensity of the green color produced is proportional to the amount of cholesterol present.

Free cholesterol unites with digitonin, a glycosidic saponin, to form cholesterol digitonide; cholesterol esters do not form such compounds. Cholesterol digitonide is insoluble in petroleum ether; cholesterol esters are freely soluble in petroleum ether. This difference in solubilities is useful to test both qualitatively and quantitatively for free versus esterified cholesterol.

Fieser showed that the melting point of cholesterol, which had been purified by recrystallization from acetic acid, was 149.5–150.0 °C. Radin and Gramza indicated that the acceptable melting point of recrystallized cholesterol was 149.3–151.3 °C. The *Merck Index* indicates that the melting point of anhydrous cholesterol is 148.5 °C, and the boiling point is 233 °C at 0.5 mmHg and 360 °C at 1 atm (760 mmHg). At 360 °C, some decomposition occurs.

Primary Cholesterol Standard

The requirements for primary standards become progressively more stringent as methods are developed that are more sensitive and more compound-specific; such is the case for cholesterol. Primary standards are chemical substances that, by virtue of their purity, can be weighed directly for the preparation of solutions with known concentrations. Primary cholesterol standards are expected to be at least 99% pure.

Cholesterol stored at room temperature, unprotected by a nitrogen cover, will undergo autoxidation over an extended period. In addition, ultraviolet light will produce structural changes in cholesterol unless amber glass containers are used. Consequently, various amounts of cholesterol oxidation products may be present in what was initially pure cholesterol. Among the cholesterol oxides that have been identified in stored cholesterol are 7-ketocholesterol, 20-hydroxycholesterol, and 24-, 25-, and 26-hydroxycholesterol. Therefore, primary standards that have been stored must be checked and may need to undergo repurification before use. Crystalline standards should be stored in small amounts over a desiccant such as silica gel at $-20\,^{\circ}C$ in the dark.

Cholesterol may be recrystallized from ethanol or acetic acid, or as the dibromide. A cholesterol preparation may be added to absolute ethanol, which is gently heated until the cholesterol dissolves and is then cooled to room temperature. The precipitated cholesterol is collected on a filter, washed with a small volume of diethyl ether, dried overnight in air, and dried in an oven at $90\,^{\circ}C$ for 2 h. Cholesterol is recrystallized from boiling glacial acetic acid solution, which is then cooled to room temperature in an ice bath. The crystals are collected on a filter, washed with acetic acid and methanol, and dried as described above.

The recrystallization of cholesterol by the dibromide method (Schoenheimer) is a more arduous task than the recrystallization from either ethanol or acetic acid, but it produces a superior product. A bromine solution is added to the cholesterol solution. A white paste is produced that is then transferred to a filter and washed with acetic acid until the filtrate is colorless. Zinc dust is then added to a suspension of the white material in diethyl ether and glacial acetic acid (750:10 v/v); the reaction produces evolution of hydrogen. The resultant white precipitate of zinc salts is dissolved in water, and the ether solution is decanted. The water contains any excess solid zinc. The ether solution is washed in a separatory funnel with acid solution and neutralized with sodium hydroxide solution. Methanol is added to the ether solution, and most of the ether is evaporated on a steam bath as the purified cholesterol begins to crystallize. The crystallization proceeds more rapidly as room temperature is approached. The product is collected on a filter and dried as described above.

The cholesterol purified by the above methods may be characterized by the color reactions discussed above (Salkowski and Liebermann–Burchard). Other color tests, such as the formaldehyde–sulfuric acid test, may be used. In this test, formaldehyde–sulfuric acid solution is added to a solution of cholesterol dissolved in chloroform. The solution, which turns cherry red, is poured into another tube, and two to three drops of acetic anhydride are added. A blue color develops. These color tests have been adapted to form the basis of spectrophotometric measurements. The results of the purified products are determined by reference to a standard of known purity.

The purity of cholesterol standards may be assessed by using melting point and boiling point determinations, microscopic comparison of the crystals with a pure reference material, and infrared and ultraviolet–visible spectra. The spectrum of the cholesterol standard is compared with that of a pure crystalline reference material.

Classic colorimetric tests, microscopic examinations, and melting point determinations have been supplemented with more modern techniques, such as mass spectrometry, nuclear magnetic resonance spectrometry, gas chromatography (GC) and high-performance liquid chromatography (HPLC) for determining the purity of prepared crystalline cholesterol. No one test in itself is sufficient to determine purity. Confirmation of results requires a minimum of two tests, which should preferably be chemically or physically unrelated. (*See* **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **Mass Spectrometry**: Principles and Instrumentation; **Spectroscopy**: Nuclear Magnetic Resonance.)

A serum reference material for determining serum cholesterol is normally used for both manual and automated methods. A serum reference material may be prepared by filtering pooled human serum through clarifying and 'sterilizing' filters. Stable serum preparations of cholesterol, with concentrations ranging from 100 to $400\,mg\,dl^{-1}$, may be made by adding an alcohol-precipitated cholesterol-rich protein from human serum to bovine, horse, or human serum. Aliquots of the preparation should be stored in sealed vials or ampoules at a temperature of $-20\,^{\circ}C$ or below. The stability of these preparations is similar to that of human serum. These sterile preparations may be shipped at room temperature for periods of up to 5 days. The concentrations of some commercial reference sera may vary considerably if the Abell–Kendall method is used as the reference assay. The Abell–Kendall procedure includes an initial step in which the serum is treated with alcoholic potassium hydroxide to liberate the cholesterol from the lipoprotein complexes and to saponify the cholesterol esters. The total cholesterol is extracted into a measured volume of petroleum ether. The cholesterol in an aliquot of the petroleum ether extract is measured by means of the Liebermann–Burchard color reaction. All commercial reference sera should be

checked against serum reference materials and standardized by the reference cholesterol method (Abell–Kendall method).

Colorimetry-based Analytical Methods

Routine lipid testing in clinical settings generally includes serum determination of triglycerides and cholesterol, and a more recent trend has been to include lipoprotein–cholesterol determination. The literature contains more than 200 methods or modifications for measuring serum cholesterol. This number is perhaps an indication of the difficulties of developing a reliable assay. Free cholesterol and cholesteryl esters may be determined separately, but it is common practice to determine the two together as 'total cholesterol.' (*See* **Spectroscopy**: Visible Spectroscopy and Colorimetry.)

The Liebermann–Burchard reagent has had a central role in much of the methodological development of cholesterol measurement systems. Since Grigant introduced a procedure for the quantitative estimation of cholesterol in 1910 by using the Liebermann–Burchard reagent (developed between 1885 and 1890), numerous modifications of the method have appeared in the literature. Most of the deviations from earlier methods consist of changes in extraction and color development.

The reaction of cholesterol with sulfuric acid in acetic anhydride (Liebermann–Burchard) to form a colored product has provided the basis for many subsequent methods. These methods may be categorized into three groups: (1) direct, in which the serum is added directly to the color reagent; (2) extraction, in which cholesterol is extracted into an organic solvent before it is added to the color reagent; and (3) hydrolysis, in which the esterified cholesterol is hydrolyzed before solvent extraction and color development. Many methods utilize acetic anhydride, acetic acid, or ferric chloride to develop color. Specificity is enhanced by prior extraction techniques. Values for cholesterol are usually higher by direct and automated methods than by manual, extraction, or hydrolysis methods. The color development of these various procedures obeys the Lambert–Beer law, which states that the absorbance is directly proportional to the concentration.

The direct methods are simple and convenient, but are subject to interference from compounds normally present in serum, such as bilirubin and proteins. Much of this interference is eliminated by extracting the cholesterol with a solvent before the color development reaction is initiated. Esterified cholesterol in serum should be hydrolyzed before a total cholesterol determination is made because more color is produced by esterified cholesterol than by free cholesterol. Failure to hydrolyze results in an overestimation of total cholesterol. Standard solutions of cholesterol must be reacted with the color reagent to determine when the color has developed its maximum intensity. Color reagents should be added to test solutions and standard solutions at fixed intervals, for example, every 30 or 60 s. When the color intensity of the standard solutions has reached its maximum, the color intensities of the test solutions should be read according to the preparation schedule, for example, every 30 or 60 s. The time required for the color to develop its maximum intensity is influenced by the ambient temperature and the composition of the individual batches of prepared reagents.

The American Association of Clinical Chemistry has suggested three methods for small laboratories that need a manual procedure for estimating total serum cholesterol. These procedures are an enzymatic method, the iron–uranyl acetate method (Parekh–Jung), and the Liebermann–Burchard reagent method. The modified Liebermann–Burchard reagent method, in both manual and automated forms, is widely used by small clinical laboratories. This discussion focuses on the manual manipulation of the method for clarity of the chemistry. The Liebermann–Burchard color reaction was discussed above with regard to primary standards. In the modified method, the cholesterol is extracted from serum test samples into 2-propanol to eliminate any interfering substances. A measured aliquot of the extract is evaporated before the Liebermann–Burchard color reagent is added. If the reagents are added directly to the extract, it is difficult to control the rate of color reaction under manual conditions. The heat produced from the exothermic reaction of 2-propanol with sulfuric acid cannot be controlled sufficiently to allow reproducible measurements. However, under stringent automated conditions, this problem can be circumvented. After a timed incubation and color-development period, the absorbance is measured at 630 nm. In this method, reference serum and serum test samples are treated in the same manner. According to the American Association for Clinical Chemistry, recoveries of cholesterol added to serum were 98–101% over a linear range of cholesterol concentrations of 0.8–4.0 g^{-1}. This method may be used to determine cholesterol in food and tissue extracts that have been cleaned in a separating funnel before being dried and redissolved in 2-propanol. Crystalline cholesterol is used as the primary standard for these test samples.

An estimation of the relationship of unesterified cholesterol to cholesterol ester can be made with another modification of the Liebermann–Burchard method. Unesterified cholesterol in serum is

precipitated as the digitonide from ethanol–ether solution. The solvent is evaporated, and the esters are extracted from the residue by adding petroleum ether, bringing the solution to a boil, cooling and centrifuging, and collecting the supernatant. The supernatant is processed as in a total cholesterol determination. The resultant value is a measure of the esterified cholesterol in the serum test sample.

The Parekh–Jung manual method is used in many clinical laboratories. This method is based on the precipitation of proteins and associated substances with ferric acetate–uranyl acetate reagent. The mixture is centrifuged, and the resultant supernatant is treated with sulfuric acid–ferrous sulfate color reagent. After a 20-min incubation and color-development period, the absorbance is measured at 560 nm.

The enzymatic method for the determination of total serum cholesterol is frequently used both manually and in an automated setting. The enzyme method is of limited value when either aqueous or pure alcohol standards are used. The method may be calibrated accurately with a homogeneous and stable serum pool. The method provides a direct measure of serum cholesterol. The enzymatic analytical determination is calibrated by using serum labeled with a target value assigned by an accepted reference method (e.g., the Abell–Kendall method). (*See* **Enzymes**: Uses in Analysis.)

Although cholesterol in serum is primarily free, cholesterol associated with lipoproteins is esterified. Cholesteryl esters are freed from the lipoproteins and enzymatically hydrolyzed by cholesterol-ester hydrolase to free cholesterol and fatty acids. The free cholesterol is then oxidized by cholesterol oxidase to cholest-4-en-3-one and hydrogen peroxidase. The peroxide, in the presence of peroxidase, oxidatively couples with phenol and 4-aminoantipyrine to produce a quinoneimine dye. The absorbance values at 560 nm are proportional to the concentration of total cholesterol in the serum test sample. The absorbance obeys the Lambert–Beer law over a wide concentration of up to about 5.5 g of cholesterol per liter. An estimate of the relationship of unesterified cholesterol to cholesterol ester is made by subtracting the free cholesterol value (the enzyme cholesterol esterase is withheld from the working test reagent) from the total cholesterol value. The enzymatic method has also been used successfully to determine the cholesterol content of foods. For this purpose, the cholesterol is extracted with a chloroform–methanol solvent, and crystalline cholesterol is used for calibration. A survey of the literature indicates that the coefficient of variation for this method is about 1–3%.

The Abell–Kendall method is generally accepted by clinical chemists as the total cholesterol reference method. The results obtained by other methods are nearly always compared with those obtained by the Abell–Kendall method. The method, however, does not lend itself well to analysis of large numbers of test samples. A saponification step generally precludes the development of an automated version of the method. The serum or plasma is saponified with alcoholic potassium hydroxide. The free cholesterol (from both unesterified and esterified cholesterol) is extracted into petroleum ether and dried; the cholesterol is determined photometrically by a modified Liebermann–Burchard reagent (acetic anhydride–sulfuric acid–acetic acid) at 620 nm. The results of this method are in good agreement with those obtained by the Schoenheimer–Sperry method.

No discussion of colorimetric cholesterol methodology is complete without a brief discussion of the separate determination of blood cholesterol and cholesterol esters that was originally developed by Bloor and Knudson in 1916. Practically all methods for separating free and esterified cholesterol emulate their work. Bloor and Knudson adapted the Windaus cholesterol digitonin precipitation method to separate cholesterol from its esters in small amounts of blood. The method consists of the determination of total cholesterol in an aliquot of an alcohol–ether extract of blood and the determination of cholesterol esters in another aliquot after precipitation of the free cholesterol by digitonin. The difference between the two values represents free cholesterol. The color reagent and color development were previously used by Liebermann–Burchard.

Gas Chromatographic (GC) Analysis

GC is used to measure cholesterol as free cholesterol, as the trimethylsilyl ether derivative, or as cholesteryl butyrate. Although colorimetric methods have been used in the past to measure the cholesterol content of foods, body fluids, and tissues, they have been, for the most part, cumbersome and not compound-specific in many applications. Agricultural chemists, biochemists, nutritionists, and food scientists began to explore the use of GC for measuring sterols in the mid-1970s. Although blood cholesterol determinations by GC can be fairly easy to accomplish, GC is not widely used as a routine clinical technique for determining cholesterol. The most probable reasons for this are the difficulties of automating the preparative stages and the fact that adequate colorimetric assays currently exist. Most chromatographic determinations of cholesterol are found in the research environment, in food and nutrition analytical laboratories, and in government regulatory laboratories.

All GC methods use a column stationary phase that is classified as nonpolar, typically methyl silicone (SE-30). In one method, the test sample is saponified, and the unsaponifiable materials are analyzed by GC. This method performs well for most products, but not for those that contain measurable amounts of α-tocopherol in addition to cholesterol. The retention times of α-tocopherol and cholesterol are nearly identical and, therefore, an analysis for either compound is not feasible in the presence of the other. The α-tocopherol may be present naturally, or may be added as a nutritional supplement or antioxidant. (*See* **Tocopherols**: Properties and Determination.)

Another approach is to saponify the test sample, extract the unsaponifiables, and derivatize the unsaponifiable compounds before GC analysis. The official method of the Association of Official Analytical Chemists (AOAC) uses this approach to form trimethylsilyl ethers (Punwar method). Another well-established method uses the classic butyric anhydride–pyridine reaction to attach a C_4 ester at position 3 on the A ring of the sterol to form cholesteryl butyrate. During GC analyses, both derivatives function equally with respect to response, reproducibility, and reliability. The trimethylsilyl ethers create problems over an extended series of analyses. The detector sensitivity drops drastically, and the detector must be disassembled, thoroughly cleaned, and reinstalled to restore sensitivity. The butyrate derivatives burn cleanly in the hydrogen flame ionization detector and do not degrade detector sensitivity. The procedure for preparing the trimethylsilyl derivative is somewhat less tedious than that for preparing the buyrates.

The GC official method of the AOAC determines cholesterol as the trimethylsilyl ether on a 2.4 m × 3 mm internal diameter silanized glass column packed with 0.5% Apiezon L on 80–100 mesh Gas-Chrom Q (Alltech Associates/Applied Science, Deerfield, IL). An alternative column may be used that consists of a 1.8 m × 4 mm internal diameter glass column packed with 1% SE-30 on 100–120 mesh Gas-Chrom Q (Alltech Associates/Applied Science, Deerfield, IL). In this method, one column is maintained at 230 °C, and the carrier gas flow is adjusted to elute the trimethylsilyl ether of cholesterol in 9–11 min. An internal standard of 5α-cholestane is used, and test samples containing 0.5–1 g of fat are extracted with chloroform–methanol–water (50:100:40 v/v/v).

A GC method based on steryl butyrates has been incorporated into a total-lipid analytical system (Sheppard–Hubbard system) for determining total lipids, fatty acid composition, and cholesterol in foods from extraction of a single test portion. The system is widely used in university, food quality control, commercial, and regulatory laboratories to determine the lipid components required for fatty acid and cholesterol labeling. The system has also been used to identify adulterated foods. This analytical system uses an aliquot of the fatty acid methyl ester preparation to prepare cholesteryl butyrate and other sterols. The fatty acid methyl esters are not affected by the reaction that forms the butyrate derivatives of the sterols. On the SE-30 column, the fatty acid methyl esters elute near the solvent front and before the appearance of the butyrates of the tocopherols and the sterols. In this system, external calibration is used. However, internal standards can be used if the internal standard peak is situated very close to the peak being measured and if there are no interfering peaks from compounds such as squalane, squalene, or cholestane in the test sample. Generally, in this system, there are too many peaks from the matrix to have a clear retention time available for internal standards.

The relative retention times for some sterols are 1.0 for cholesteryl butyrate, 1.15 for brassicasteryl butyrate, 1.3 for campesteryl butyrate, 1.4 for stigmasteryl butyrate, and 1.6 for sitosteryl butyrate. Free cholesterol and other sterols can be easily separated with a 15 m × 0.242 mm, SE-54, wall-coated, fused-silica capillary column operated at 250 °C with a helium flow rate of 0.74 ml min^{-1}. The sequence of compound appearance is the same as that for SE-30 packed columns.

High-performance Liquid Chromatography (HPLC)

HPLC is increasingly being used to determine cholesterol and other sterols in foods and tissue extracts. However, like GC, HPLC is not generally used in routine clinical analyses performed with automated clinical multiple analysis systems based on colorimetric or fluorometric assays. Sterols that can be separated by GC usually cannot be separated by an HPLC system. One HPLC system in widespread use determines the benzoate ester of cholesterol on a 300 × 3.9 mm internal diameter μBondapak C_{18} column with a 100% methanol mobile phase and a variable wavelength ultraviolet detector set at 230 nm. GC and HPLC determinations of cholesterol in a variety of foods show that the two techniques yield statistically identical results. Amounts as low as 10 ng of cholesterol benzoate can be determined using HPLC.

See also: **Atherosclerosis**; **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **Coronary Heart Disease**: Intervention Studies; **Enzymes**: Uses in Analysis; **Mass Spectrometry**: Principles and Instrumentation;

Spectroscopy: Nuclear Magnetic Resonance; Visible Spectroscopy and Colorimetry; **Tocopherols**: Properties and Determination

Further Reading

Brumley WC, Sheppard AJ, Rudolf TS *et al*. (1985) Mass spectrometry and identification of sterols in vegetable oils as butyryl esters and relative quantitation by gas chromatography with flame ionization detection. *Journal of the Association of Official Analytical Chemists* 68: 701–709.

Christie WW (1987) *High-performance Liquid Chromatography and Lipids: A Practical Guide*, 1st edn. New York: Pergamon Press.

Faulkner WR and Meites S (eds) (1982) *Selected Methods for the Small Clinical Chemistry Laboratory*, vol. 9, pp. 157–183. Washington, DC: American Association of Clinical Chemistry.

Fieser LF and Fieser M (1949) *Natural Products Related to Phenanthrene*, 3rd edn. New York: Reinhold.

Macrae R (1988) *HPLC in Food Analysis*, 2nd edn. London: Academic Press.

Sheppard AJ, Newkirk DR, Hubbard WD and Osgood T (1977) Gas–liquid chromatographic determination of cholesterol and other sterols in foods. *Journal of the Association of Official Analytical Chemists* 60: 1302–1306.

Windaus A (1932) Uber die Konstitution des Cholesterins und der Gallensauren. *Zeitschrift für Physiologische Chemie* 213: 147–187.

Yasaei P, Sheppard AJ, Brumley WC, Mazzola EP and Aldridge MH (1989) Structural proof of cholesterol isolated from plants. II. Identification and verification by ^{13}C nuclear magnetic resonance spectroscopy and mass spectrometry. *Journal of Micronutrient Analysis* 5: 259–267.

Yasaei P, Sheppard AJ, Rudolf TS, Shen C-SJ and Aldridge MH (1989) Structural proof of cholesterol isolated from plants. I. Isolation and preliminary identification by chromatography and infrared spectroscopy. *Journal of Micronutrient Analysis* 5: 238–245.

Absorption, Function, and Metabolism

D R Arnold and P O Kwiterovich Jr, Johns Hopkins University, Baltimore, MD, USA

Cholesterol Metabolism

Cholesterol metabolism in humans is complex. Cholesterol is either supplied from the diet (exogenous) or synthesized *de novo* by many cells of the body (endogenous). The major factors in the diet that may increase the blood cholesterol level are high intakes of cholesterol itself, or of saturated fats and excessive calories. The liver is one of the major sites of endogenous cholesterol synthesis. The pool of cholesterol in the liver is tightly regulated and reflects the input of cholesterol from the diet, the biosynthesis of cholesterol, the secretion and uptake of cholesterol from plasma lipoproteins, the conversion of cholesterol into bile, and the reuptake of biliary cholesterol and bile acids from the intestine to the liver. The metabolic exogenous and endogenous pathways of cholesterol metabolism are described in this chapter. The concentration of cholesterol in the blood in fasting normal humans is the result of the metabolism of cholesterol from exogenous and endogenous sources. Environmental factors such as dietary fatty acids, and metabolic perturbations such as diabetes and obesity as well as genetic factors also influence the level of cholesterol in blood. Consequently, after several decades, hypercholesterolemia causes atherosclerotic vascular disease. Dieting, exercise, and several drugs are available as tools to lower blood cholesterol levels, and the mechanisms by which these interventions lower elevated cholesterol levels are discussed here.

Cholesterol – Chemical Structure and Importance

Cholesterol (cholest-5-en-3-ol (3-beta) cholesterol) and cholesterol metabolites and esters are major components of the plasma membrane and of many other cellular organelles in animals. On average, 1.5 g of cholesterol is present per kilogram of body weight; some tissues like fat, brain, and liver contain higher amounts of cholesterol (e.g., liver contains 4.5 g kg^{-1}). Cholesterol is a water-insoluble lipid, made of a so-called sterol backbone with one hydroxy group and one double bond, together with a side-chain made up of eight carbon atoms. Cholesterol is chemically distinctly different from two other lipids, triacylglycerols and phopholipids. The steroid backbone of cholesterol also constitutes the fundamental structure of bile acids (with cholic acid and chenodeoxycholic acid being the primary bile acids), certain fat-soluble vitamins, such as vitamin D_3, steroid hormones (cortisone and aldosterone), and sex hormones (estradiol and testosterone, **Figure 1**).

Cholesterol is primarily transported through blood on low-density lipoproteins (LDL). When present in elevated amounts, LDL cholesterol causes 'hardening of the arteries' or atherosclerotic cardiovascular disease (ASCVD). The evidence for a causal relationship

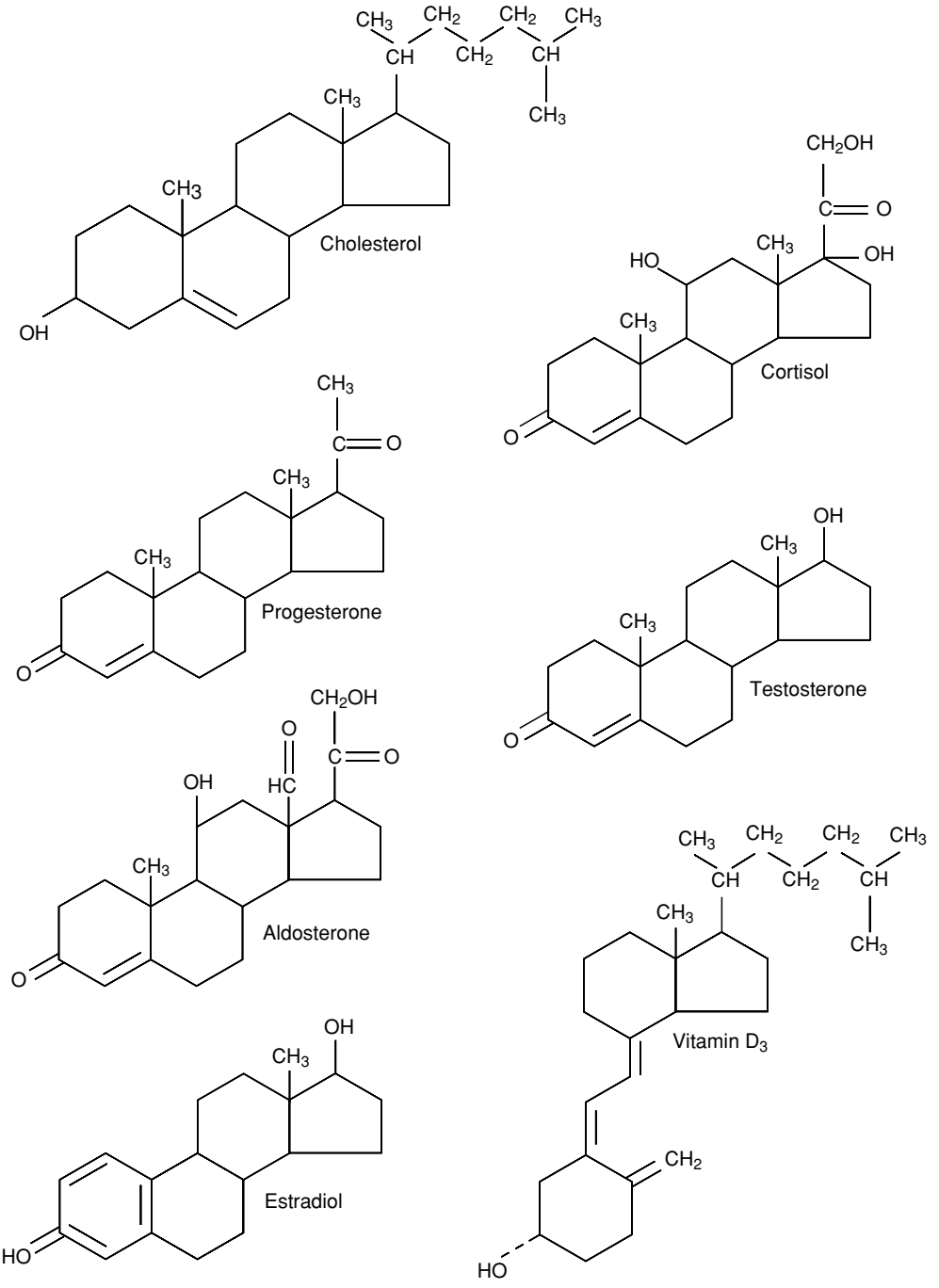

Figure 1 Biochemical structure of cholesterol in comparison with other hormones derived from cholesterol. Cholesterol is homologous to several steroid hormones: cortisol, a stress hormone, progesterone, testosterone and estradiol (sex hormones), aldosterone (important for blood pressure control and salt–fluid balance) and vitamin D_3 (important for calcium absorption in the gut and homeostasis).

between elevated blood LDL cholesterol and ASCVD comes from many research studies in various disciplines (feeding studies of animals, epidemiological surveys, studies of genetic abnormalities of lipid disorders, and large, placebo-controlled, double-blind clinical trials).

History

Poulletier de la Salle discovered cholesterol in gallstones in 1769, but Chevreul gave it the name 'cholesterine' in 1815. Ignatowsky first demonstrated in 1908 the connection between intake of cholesterol in

the diet of rabbits and the development of arteriosclerosis. Anitschkow demonstrated in 1913 that it was cholesterol feeding only that caused atherosclerotic changes in the blood vessels of rabbits.

Cholesterol De-novo Production

Cholesterol can be either absorbed from food or made *de novo* in multiple human tissues in varying amounts. The biosynthesis of cholesterol starts with acetyl coenzyme A (acetyl CoA). Three molecules of acetyl CoA are condensed to form 3-hydroxy-3-methylglutaryl CoA (HMG CoA). The conversion of HMG CoA to mevalonate by HMG CoA reductase is the rate-limiting step of cholesterol synthesis. A schematic diagram summarizing the *de-novo* biosynthesis of cholesterol is shown in Figure 2. While most of the sterol pool of the body is made *de novo* in extrahepatic tissues, the liver is a major organ involved both in the synthesis of cholesterol and

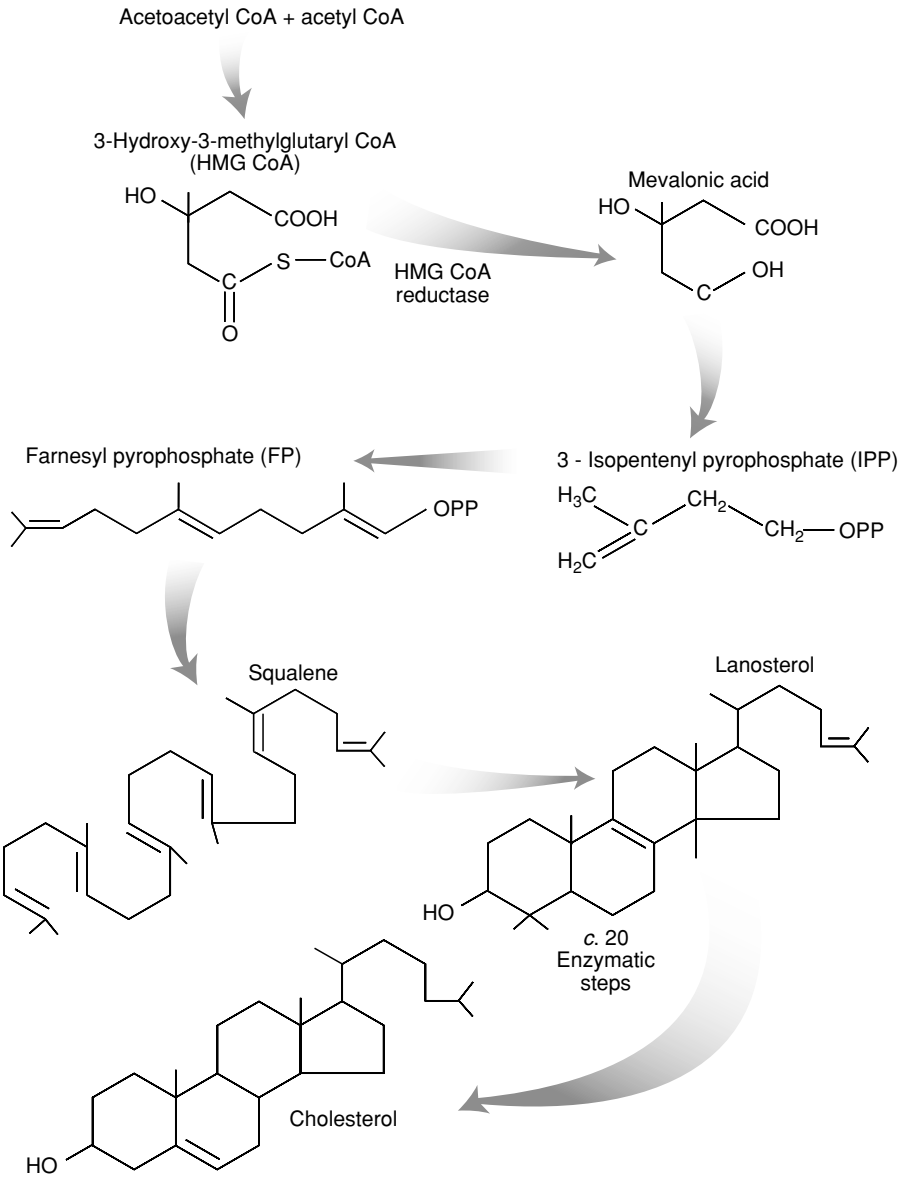

Figure 2 *De-novo* biosynthesis of cholesterol. The biosynthesis of cholesterol begins with the condensation of three acetyl molecules to form 3-hydroxy-3-methylglutaryl CoA (HMG CoA). The acetyl-CoA can be derived from glucose, fatty acids, or amino acids. Reduction of HMG CoA produces mevalonate, which gets further processed to isopentenyl pyrophosphate (IPP). Three molecules of IPP are condensed to make up farnesyl pyrophosphate (FP). Two molecules of FP are used to make up squalene, which gets converted by three enzymatic steps to lanosterol. Approximately 20 additional chemical steps are needed to make up cholesterol.

in the processing and repackaging of cholesterol and other lipids absorbed from the diet. By using H^3-labeled water, the average production of cholesterol in man has been estimated to about 9 mg per kilogram per day. Using this technique, the most active cholesterol processing tissue is the liver followed by the gastrointestinal tract, skin, striated muscle and bone marrow while minor amounts are made in the adrenal glands, and other endocrine organs like ovaries and testis.

Cholesterol from Diet and Factors Influencing Cholesterol Absorption/ *de novo* Synthesis

The amount of synthesis of cholesterol in mammalian liver is tightly controlled by multiple factors. Modulating dietary factors that affect plasma cholesterol levels include energy restriction in the form of decreased lipid and carbohydrate intake (diet), meal frequency, dietary fat type (saturated, unsaturated or trans), and cholesterol content of the food. Plants make sitosterol but not cholesterol, and an average American diet contains about 100–200 mg of this plant sterol per day. In the intestine, inhibition of the uptake of cholesterol occurs with intake of these so-called phytosterols, and this can lower the blood cholesterol level. Diseases of the liver, or exposure to drugs and toxins such as alcohol, can easily disturb the ability of the liver to handle cholesterol and other fats, and may result in the initiation of 'fatty degeneration,' thus demonstrating further the importance of the liver in the metabolic processing of lipids.

Cholesterol in the intestinal lumen comes either from the diet (250–500 mg) or the bile (*c.* 600–1000 mg). While bile acids, such as cholic acid and chendeoxycholic acid, are nearly completely recirculated, only about 50% of the dietary cholesterol is absorbed at an average intake of 300–600 mg per day. Formulas developed by Keys and by Hegstedt and coworkers predict the average increase in the blood cholesterol level that will result from a given change in the amounts of dietary cholesterol, saturated and unsaturated fatty acids. The Hegstedt formula is as follows:

$$< \text{Blood cholesterol (mg dl}^{-1}) = 2.16 < \text{S} - 1.65 < \text{P} + 0.068 < \text{C},$$

where < symbolizes a change in; S = dietary saturated fatty acids (% of total calories), P = dietary polyunsaturated fatty acids (% total calories) and C = dietary cholesterol (mg per day). Examples of polyunsaturated fats include linoleic acid (C18:2), linolenic acid (C18:3), and arachidonic acid (C20:4). Linoleic acid is the essential fatty acid, which can be elongated and desaturated to form other polyunsaturated fats. Fish oils include eicosapentaenoic acid (EPA: C20:5) and docosahexaenoic acid (DHA: C22:6).

Without drug treatment, dietary cholesterol and saturated fats, and the *de novo* production and removal of LDL, determine the blood LDL-cholesterol level. Saturated fatty acids, particularly with 14 and 16 carbons, suppress the receptor-mediated uptake of LDL and increase the LDL cholesterol level, whereas unsaturated fatty acids like 9-*cis* 18:1 increase receptor-mediated LDL uptake and lower the LDL cholesterol level.

Excretion of Cholesterol

The human does not have an enzyme that can degrade the sterol ring of cholesterol; the backbone/sterol ring of cholesterol thus has to be excreted in the bile and stool. All cholesterol metabolites, including hormones and vitamin metabolites, are excreted as bile acids or metabolites in the intestine, but they are partially reabsorbed in the intestine via the enterohepatic pathway. The sterol ring of hormones and vitamins are metabolized (hydroxylated/sulfonurated) by multiple enzymes in the liver to make the sterol more water-soluble, which promotes excretion in bile and the urine as well. The delicate balance of intake and excretion of cholesterol and cholesterol products is tightly regulated. Evolutionarily, excess cholesterol must be avoided inside cells as it can be toxic. Excess cholesterol in the bloodstream is also undesirable, as it causes hardening of the arteries (atherosclerosis) and consequently, ASCVD.

Lipoproteins

To make lipids soluble in blood, they usually are packaged inside a lipid-rich core, around which are wrapped amphipathic proteins, whose nonpolar amino acid groups are exposed to the lipid core, whereas their polar or charged amino acids are oriented to the outer surface, making the entire lipoprotein particle water-soluble. The proteins associated with lipoproteins are called apolipoproteins and act as surface components of the particle, as enzymes key to lipid metabolism in blood, or as ligands to cell-surface receptors promoting the binding and uptake of lipoproteins. A list of these apolipoproteins explaining their functions are listed in Table 1. Multiple apolipoproteins can be present on the surface of the lipoprotein particle and provide information about the tissue of origin and the function of the lipoprotein. For example, apoB-48, a

Table 1 Characteristics of apolipoproteins: their origins, functions and distribution in plasma lipoprotein classes

Apolipoprotein	*Origin*	*Function*	*Present in lipoprotein classes*	*Molecular weight (kDa)*
A-1	Liver, intestine	Activator of LCAT	CM, HDL	29 016
A-2	Liver, intestine	Inhibits HL and VLDL hydrolysis	CM, HDL	17 414
A-4	Liver, intestine	Activates LCAT	CM, HDL	44 465
B	B100 liver	Ligand for LDL receptor	B100→VLDL, LDL	512 723
	B48 intestine	Necessary for secretion of intestinal and hepatic fat	B48→CM	241 000
C-1	Liver	Cofactor for LCAT	CM, VLDL, HDL	6 630
C-2	Liver	Activates LPL, cofactor for LCAT	CM, VLDL, HDL	8 900
C-3	Liver, intestine	Inhibits LPL/inhibits binding of ApoE to the LDL receptor	CM, VLDL, HDL	8 800
D	Many sources	Reverse cholesterol transport	HDL	19 000
E (2,3, and 4)	Liver, peripheral tissues	Ligand for LDL receptor and LRP receptor	CM, VLDL, HDL	34 145

CM, chylomicron; LDL, low-density lipoprotein; HDL, high-density lipoprotein; VLDL, very-low-density lipoprotein; LPL, lipoprotein lipase; HL, hepatic lipase; LCAT, lecithin cholesterol acyltransferase; LRP, low-density lipoprotein receptor-like protein.

Table 2 Characteristics of key enzymes and transfer proteins of lipoprotein particles in the blood

Enzyme	*Origin*	*Function*	*Lipoprotein classes involved*	*Molecular weight (kDa)*
LPL	Capillaries of muscle and adipose tissue	Hydrolyzes triglycerides and phospholipids	CM, large VLDL	50 394
HL	Liver	Hydrolyzes triglycerides and phospholipids	Small VLDL, IDL, HDL	53 222
LCAT	Liver	Converts free cholesterol from cell membranes to cholesterol esters	HDL	47 090
CETP	Liver, spleen, adipose tissue	Transfers cholesteryl esters from HDL to ApoB containing lipoproteins	HDL, LDL, VLDL	74 000
PTP	Pancreas, adipose tissue, lung	Transfers the majority of phospholipids in plasma	HDL	81 000

CM , chylomicron; LDL , low-density lipoprotein; HDL , high-density lipoprotein; VLDL , very-low-density lipoprotein; LPL , lipoprotein lipase; HL , hepatic lipase; LCAT, lecithin cholesterol acyltransferase; CETP, cholesterol transfer protein; PTP, phospholipid transfer protein.

truncated version of apolipoprotein B, is only made in the intestine and is present only on chylomicrons and chylomicron remnants, whereas apoB-100 is made in the liver and is the major apolipoprotein on very-low-density lipoproteins (VLDL) and the products of VLDL catabolism, VLDL remnants and intermediate density lipoproteins (IDL), and LDL. Multiple enzymes are present on the capillary walls and on the lipoprotein particles that metabolize/transfer the lipids present in the particles (see Table 2 and also below).

There are principally five major classes of lipoprotein particles in the blood: chylomicrons (CM), VLDL, IDL, LDL, and high-density lipoproteins (HDL), based on their physicochemical properties. This classification is based on a separation via a salt density gradient (chylomicrons are the lightest particles, whereas HDL particles are the heaviest; see Figure 3 for further explanations).

Chylomicrons and Transport of Dietary (Exogenous) Fat

Dietary cholesterol and triacylglycerol are emulsified in the intestinal lumen by bile salts; the cholesteryl esters are hydrolyzed into unesterified cholesterol and free fatty acids (FFA), and the triacylglycerol into 2-monoglycerol and FFA. These constituents are absorbed into the intestinal epithelial cells where they are repackaged into chylomicrons. Chylomicrons are large (*c.* 1 μm) and their lipid-containing core is surrounded on the surface by apoproteins which are amphipathic molecules that allow the lipid core to be soluble in aqueous lymph fluid and blood. These particles contain the apolipoproteins B-48, A-1, A-2, and A-4. The secretion of the chylomicrons from the intestinal cells into the lymph requires the presence of apoB-48. The lymph from the gut collects in the thoracic duct and enters the blood stream via the right subclavian vein. Even before the chylomicrons reach the systemic circulation, the apoA-1 content is decreased, and apoE, apoC-1, apoC-2, and apoC-3 are acquired as a result of HDL particles interacting with chylomicrons in the intestinal lymph capillaries.

Chylomicrons then interact with endothelial cells in the arterioles lining a number of tissues. The triacylglycerols in the core of the chylomicrons, through the interaction of apoC-2 with the enzyme lipoprotein lipase (LPL) bound to the surface of endothelial

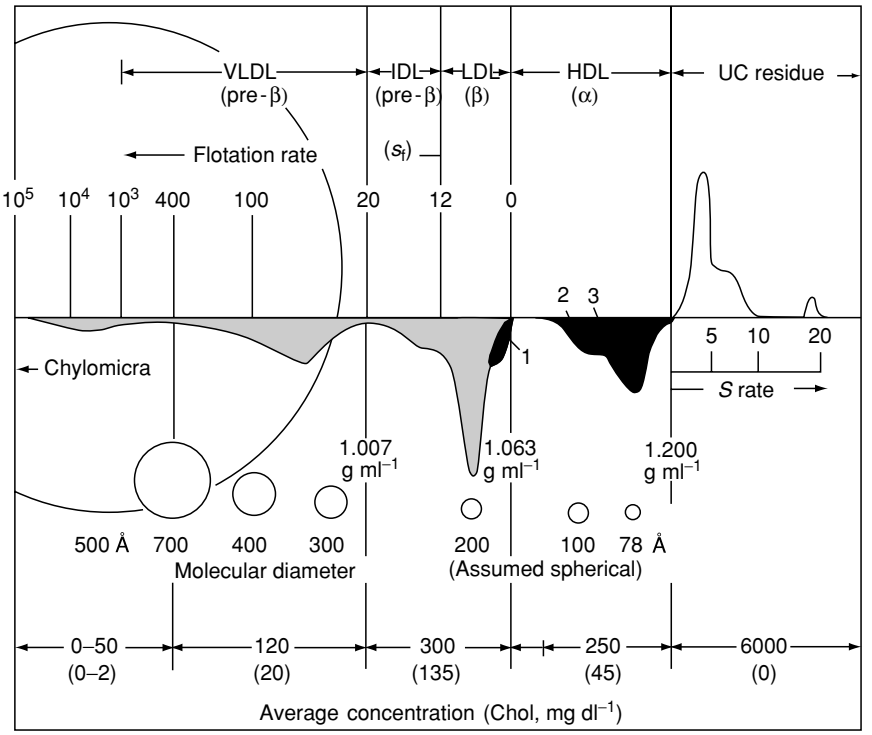

Figure 3 Characteristics of the major lipoproteins present in normal human plasma. Flotation rates are measured in S_f units. The flotation of VLDL, IDL and LDL is performed at a salt gradient of 1.063 g ml^{-1}, whereas a density of 1.200 g ml^{-1} is used for HDL. During ultracentrifugation at 200 000 g for 1 h, a light beam is directed through an aperture in the cell and the changes in diffraction recorded to outline the boundaries caused by the floating lipoproteins. The class of lipoprotein is indicated at the top, corresponding to abbreviations and a description given in the text. 'UC-residue' describes nonlipid plasma proteins that sediment at a density of 1.200 g ml^{-1}. The amount of each lipoprotein class is proportional to the area under each curve shown numerically at the bottom. There are three subclasses of HDL (1, 2, and 3), and these are indicated by numbers and colored black. HDL1 occurs only in hyperlipemic plasma. The approximate diameter of each lipoprotein class particle is shown in Ångströms (1 Å = 0.1 nm), and the average amount of plasma cholesterol present in each class of lipoprotein is shown in parentheses (for example, 45 mg dl^{-1} for HDL lipoproteins). From Olson RE (1998) Discovery of the lipoproteins: their role in fat transport and their significance as risk factors. *Journal of Nutrition* 128(Supplement 2): 439S–443S, with permission.

cells, are hydrolyzed into FFA and glycerol, which are then absorbed into adipocytes and other cells. During hydrolysis, phospholipids and apolipoproteins are removed from the surface of the chylomicrons and transferred to HDL. By removing apoC-2, an activator of LPL, the process starts to slow down. The resultant chylomicron remnant, which contains less triacylglycerols and relatively more esterified cholesterol, is removed rapidly from the circulation by its interaction with the chylomicron receptor (low-density, lipoprotein receptor-like protein (LRP)) on the surface of the liver cells. The LDL receptor and proteoglycans can also facilitate the uptake of chylomicron remnants into the liver. ApoE is the ligand that permits the binding and uptake of the chylomicron remnant.

Very-low-density Lipoproteins and Transport of Hepatic (Endogenous) Fat

In the liver, cholesterol, triacylglycerols, and other fats are packaged into the VLDL particle that is surrounded by apoB-100. Triacylglycerol and cholesteryl ester production is necessary for VLDL synthesis by the liver. The availability of apoB-100 affects both the assembly and secretion of the VLDL. The apoB-100 gene is constitutively expressed.

The amount of apoB-100 available is regulated by the rate of proteolysis of apoB-100. When excess lipid is available, proteolysis is decreased, and more apoB-100 is available for VLDL production. When lipid is depleted, apoB-100 proteolysis increases. The triacylglycerols and apoB-100 interaction requires the presence of the microsomal triacylglycerol transfer protein (MTP). MTP is a heteromer that consists of two subunits, a protein widely distributed in many cells, called disulfide isomerase, and a unique 97-kDa large subunit. A mutation in the 97-kDa protein prevents the association of lipid with apoB-100, leading to little, if any, secretion of VLDL and no production of IDL and LDL. At least a dozen mutations in the 97-kDa protein have been described in patients with the recessive disorder, abetalipoproteinemia.

VLDL is secreted from the liver into the bloodstream surrounded by apoB-100, apoE, and the apoCs. Once in the bloodstream, the apoC-2 on VLDL binds to the same LPL site as chylomicrons in the capillaries surrounding various tissues.

Intermediate-density lipoproteins

As the core triacylglycerol in VLDL is hydrolyzed, a remnant VLDL is formed, and the released FFA are taken up by tissues. Upon further hydrolysis of triacylglycerol, the final remnant, IDL, is formed. The liver takes up some IDL through the interaction of apoE with the LDL (B, E) receptor. The rest of the IDL is modified at the surface of the liver by hepatic lipase (HL), producing the cholesteryl ester-enriched LDL.

Low-density Lipoproteins – Production, Uptake, and Catabolism

The majority of LDL is formed from VLDL. The major apolipoprotein of LDL is apoB-100, which serves as the ligand for the binding of LDL to the LDL (B, E) receptor in clustered pits on the surface of liver and peripheral cells. The bound LDL and its receptor are internalized into the cell. LDL are taken up and degraded in the lysosomes, while the LDL receptor is released and recycled back to the cell surface. Cholesteryl esters in the core of LDL are hydrolyzed into unesterified (free) cholesterol and FFA. The enzyme acyl cholesterol acyl transferase (ACAT) controls the cellular levels of unesterified (free) cholesterol versus cholesteryl esters. Monounsaturated fat such as oleic acid is a preferred substrate for ACAT, whereas saturated FFA is not. In the presence of excess dietary saturated FFA, less cholesterol is esterified, and the excess cholesterol inhibits the release of the transcription factor from the nuclear membrane. Thus, less of this factor is available to go into the nucleus where it ordinarily upregulates the transcription of both the LDL receptor gene and the gene for HMG CoA reductase, the rate-limiting enzyme of cholesterol production (see below).

High-density Lipoproteins and Reverse Cholesterol Transport

HDL play a paramount role in the reverse cholesterol transport pathway, which mediates the removal of cholesterol from peripheral cells for transport back to the liver. HDL are secreted from the liver and intestine as nascent, disc-shaped particles that consist primarily of phospholipids and apoA-1, with very little cholesterol. The nascent HDL interacts with the ABCA1 transport protein on the surface of peripheral cells such as macrophages and removes unesterified cholesterol. In the bloodstream, the free cholesterol is esterified by lecithin cholesterol acyl transferase (LCAT) and its cofactor, apoA-1, producing a spherical HDL particle with a cholesteryl ester in its core. The core cholesteryl ester in HDL can be selectively taken up in liver and other steroidogenic tissues such as the adrenal gland, ovaries, and testes through the interaction of HDL with the SR-B1 receptor. Alternatively, the cholesteryl esters in HDL can be exchanged for the triacylglycerols in VLDL and IDL through the action of cholesteryl ester transfer protein (CETP). Such a cholesteryl ester is eventually removed by the interaction of IDL or LDL with the LDL receptor on the surface of liver or peripheral cells.

Important Enzymes Responsible for the Processing of Lipoprotein Particles in the Bloodstream and Peripheral Tissues

Apolipoproteins present on the surface of lipid particles serve as ligands for receptors as well as activators or inhibitors of enzymes important in the lipoprotein-processing pathway. The names and functions of enzymes playing a major role in the processing of lipid particles are summarized in Table 2. Functions of some of the important apolipoproteins and enzymes will be discussed briefly.

LPL catabolizes the triacylglyceride in the core of many lipoprotein particles. LPL converts chylomicrons into chylomicron remnants, and VLDL particles into VLDL remnants and IDL. LPL is bound to heparan sulfate proteoglycans on the surface of capillaries, and production of this enzyme is upregulated by insulin. Heparinase-induced proteoglycan deficiency (heparinase cuts off proteoglycans from the cells) impairs the function of LPL. Both chylomicrons and chylomicron remnants are cleared from the bloodstream by the liver through the binding of the ligand, apoE, either to the LRP receptor or to the LDL (B,E) receptor. This initiates internalization via coated pits and lysosomal degradation of the apoE-coated lipoprotein particles.

LCAT is made in the liver and is present on HDL particles. Initially, HDL has very little cholesteryl ester but is loaded up with cholesteryl ester by the action of LCAT. This HDL particle then undergoes exchange of cholesteryl esters for triacylglycerol on the triacylglycerol-rich lipoproteins through the action of the CETP. The triacylglycerol-rich lipoproteins are then processed by HL, present on the surface of liver cells. HL hydrolyzes triacylglycerols and phospholipids. HL is similar to LPL and belongs to the lipase superfamily. Both enzymes, CETP and HL, are needed for reverse cholesterol transport as

demonstrated by transgenic experiments/substitution experiments. Estrogen is an important negative regulator of HL. Phospholipid transfer proteins (PTP) are made in multiple tissues. PTPs play an important role in the maintenance of plasma HDL content and remodeling of HDL in the circulation.

Role of Transcription Factors in the Regulation of Cholesterol Metabolism

Recently, transcription factors (proteins that bind to DNA and control transcription of proteins important in lipid metabolism) have been described that tightly control the amount of cholesterol present inside the cell.

Uptake of cholesterol in liver cells is mediated by receptors such as the LRP receptor and the LDL receptor. Brown, Goldstein, and coworkers have described transcription factors called sterol regulatory element-binding proteins (SREBPs). SREBPs control the transcription of enzymes such as HMG CoA reductase, essential in cholesterol biosynthesis, and of the LDL receptor, which mediates the uptake of LDL by the liver. It has been known for years that cholesterol in the diet shuts down the synthesis of new cholesterol in the liver. It does so by downregulating the rate-limiting enzyme in the biosynthesis of cholesterol, HMG CoA reductase. It is now known that an increase in hepatic cholesterol levels inhibits the sequential release of an active form of SREBP by two proteases from the nuclear membrane. Conversely, a decrease in hepatic cholesterol levels promotes the proteolysis and release of SREBP, allowing its translocation into the nucleus where SREBP binds to the promoter of the LDL receptor gene and activates its transcription (Figure 4). The increased production of the LDL receptor molecule results in an increased uptake of LDL cholesterol by the cells. Also, SREBP interacts with the gene for HMG CoA reductase, enhancing its transcription and increasing the number of enzyme molecules. This increases the biosynthesis of cholesterol. Thus, both an increased uptake of cholesterol on LDL from the blood and enhanced production of cholesterol synthesis occur when the level of cholesterol in the liver is decreased. Conversely, as the cholesterol pool in the liver is increased, the release of SREBP is inhibited, the LDL receptor gene and HMG CoA reductase genes are downregulated, and an equilibrium is achieved (see Figure 4).

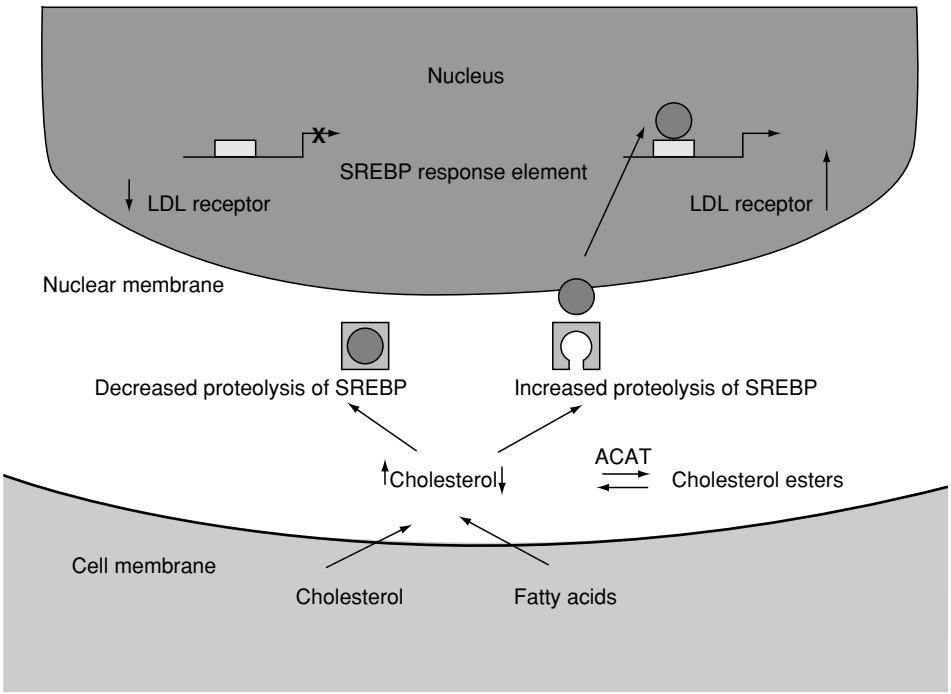

Figure 4 LDL-receptor/SREBP pathway. Liver cells adsorb cholesterol and fatty acids, and the amount of cholesterol versus cholesterol ester in the cell cytoplasm is controlled by the enzyme acyl cholesterol acyl transferase (ACAT). If the level of free cholesterol is decreased (on the right), a proteolytic process is activated to free the steroid response element-binding protein (SREBP) which then translocates to the cell nucleus and activates the transcription of the LDL-receptor gene. This leads to an increased expression of LDL-receptors on the surface of the liver cell and to enhanced clearance of LDL from the bloodstream. The reverse is true (on the left) if the intracellular cholesterol level is increased.

Other enzymes, such as the intracellular ACAT, also play a role in this regulatory pathway; for example, as ACAT converts free cholesterol into cholesteryl esters, less unesterified cholesterol is available to inhibit the proteolysis of SREBP, leading to increased release of SREBP and the upregulation of both the LDL receptor and HMG CoA reductase. If ACAT is less active, as might occur in the presence of saturated fats (which are not preferred substrates for the enzyme), more free cholesterol will be available to inhibit the release of SREBP, leading to a downregulation of LDL receptors and HMG CoA reductase. If the inhibition of the LDL receptors is greater than the inhibition of HMG CoA reductase, as occurs in about one in four humans, an increase in the blood cholesterol and LDL cholesterol occurs.

Another important protein involved in the control of cholesterol synthesis is the SREBP cleavage activation protein (SCAP). The NH_2-terminal protein domain of SCAP resembles the NH_2-terminal protein domain of HMG CoA reductase, and it is thought that both function as sterol-sensing domains. SCAP is a required activator of SREBP cleavage, as cleavage is usually abolished in the presence of sterols. The family of SREBP transcription factors also appears to be important in the regulation of fatty acid synthesis. acetyl CoA carboxylase, fatty acid synthase, stearyl CoA desaturase-1 (a necessary enzyme for the generation of unsaturated fatty acids), and even the enzyme LPL appear to be modulated by SREBP.

Inherited Mechanisms of Disorders of Lipid and Lipoprotein Disorders

Apolipoproteins are very important in cholesterol metabolism as they act as enzymes and ligands for receptors mediating lipid modification and absorption (see also above). Most apolipoproteins are made in the liver and intestine, but a small amount of apoB can be made in heart tissue and apoE is made by macrophages. All tissues can degrade apolipoproteins, but most of them are degraded in the liver.

One of the most serious inherited abnormalities of the cholesterol metabolism is a condition called familial hypercholesterolemia (FH). In FH, the LDL receptor is defective, and liver cells have a reduced capacity to bind and take up LDL cholesterol from blood. FH can occur in a severe homozygous form where patients have few, if any, functional LDL receptors. Homozygotes for FH are rare, about one in a million, whereas patients with one mutant allele and one normal allele for the LDL receptor, FH heterozygotes, occur in about one in 500 persons. The ligand for the LDL receptor, apoB-100, on the surface of LDL particle can also be defective, and this may also lead to inherited high total cholesterol and LDL cholesterol levels similar to those with FH.

More than 150 mutations in the LDL receptor gene have been described, and grouped into five different classes: the class I mutations do not permit the LDL receptor to be synthesized; class II mutations interfere with the transport of the receptor to the surface of the cell; class III mutations produce a receptor that does not bind normally to its ligand, apoB-100; in class IV mutations, the LDL receptor binds to LDL but it can not be internalized; and in class V mutations, the receptor cannot be reused by recycling.

Treatment of the homozygous FH individuals is problematic, since it is difficult in many patients to induce functional LDL receptors, either with a statin or with a bile acid sequestrant. Niacin and a high dose of statins may decrease the production of LDL by inhibiting to a degree the formation and secretion of VLDL in liver. FH homozygotes often die from ASCVD before the age of 20 years. The atherosclerosis also affects the aortic valve and can lead to life-threatening aortic stenosis. Transplantation of a liver with functional LDL receptors was once used but has fallen out of favor because of the significant morbidity and mortality associated with this procedure. LDL phoresis can often decrease the LDL levels by half; however, the procedure must be repeated every 2 weeks, and sufficient blood flow can be a problem in very young FH homozygoyes. There has been a strong interest in gene therapy as a potential successful treatment, but it is still experimental. Heterozygous FH individuals can be successfully treated by a combination of diet, exercise, and drugs.

Treatment of Lipid Disorders

Lipid disorders have traditionally been classified according to the increased amount of lipoprotein particles present in blood (Fredrickson's class I–V, Table 3).

With increasing knowledge, mutations in receptors, apolipoproteins, and enzymes have been identified as the cause of these abnormalities, but many defects are still not known, and often, their expression is modulated by environmental and genetic factors as well. Many lipid disorders are secondary to other diseases. For example, the LPL is upregulated by insulin, and diabetic patients who do not produce enough insulin often manifest a hypertriacylglycerolemia secondary to their diabetes and relative LPL deficiency. Other medical conditions such as hypothyroidism can cause an elevated LDL cholesterol level, and in such cases, the other medical condition must be treated first. Certain classes of drugs can also

Table 3 Classification of lipid disorders by phenotype (after Fredrickson)

Group	*Lipoprotein in excess*	*Plasma cholesterol*	*Plasma triacylglycerol*	*Plasma appearance*
I	Chylomicrons	Elevated or very elevated	Very elevated	Clear plasma with creamy layer on top
IIa	LDL	Elevated	Normal	Clear
IIb	LDL+VLDL	Elevated	Moderately elevated	Clear, turbid
III	β-VLDL, IDL	Elevated	Elevated	Turbid
IV	VLDL	Normal or moderately elevated	Elevated	Turbid
V	Chylomicrons+VLDL	Elevated or very elevated	Very elevated	Turbid plasma, creamy layer on top

Table 4 Intervention National Cholesterol Education Program (NCEP) guidelines

LDL-cholesterol level with	*Initiation of diet*	*Added drug therapy*
Less than two risk factors	$>160\,mg\,dl^{-1}$	$>190\,mg\,dl^{-1}$
More than two risk factors	$>130\,mg\,dl^{-1}$	$>160\,mg\,dl^{-1}$
ASCD		$>100\,mg\,dl^{-1}$
Diabetes mellitus		$>100\,mg\,dl^{-1}$

Patients with diabetes mellitus or ASVD have a treatment goal for LDL cholesterol of equal or less than 100 mg/dl independent of the presence or absence of risk factors.
From National Cholesterol Education Panel (NCEP) (1993) Adult treatment panel II. *Journal of the American Medical Association* 269: 3015–3023, with permission.

Table 5 Atheroclerotic vascular disease (ASVD) risk factors after the National Cholesterol Education Program (NCEP) guidelines

	Positive risk factor	*Negative risk factor*
HDL cholesterol value	$<40\,mg\,dl^{-1}$	$>60\,mg\,dl^{-1}$
Age	Men >45 years of age or older Women >55 years of age or older	
Family history of premature ASVD	Present	
Current cigarette smoking	Present	
Hypertension	Present	

Each person's risk factors are to be determined and used as a guideline for treatment. Patients with diabetes mellitus or ASVD have a treatment goal for LDL cholesterol of equal or less than $100\,mg\,dl^{-1}$ independent of the presence or absence of risk factors.
From National Cholesterol Education Panel (NCEP) (1993) Adult treatment panel II. *Journal of the American Medical Association* 269: 3015–3023, with permission.

contribute to lipid abnormalities and, if so, can be discontinued if medically feasible.

Elevated LDL-cholesterol levels are strongly correlated with development of ASCVD, whereas a high HDL-cholesterol level is usually protective against ASCVD. The National Cholesterol Educational Program (NCEP) uses measurements of LDL-cholesterol as guidelines for treatment with diet and drug therapy in addition to multiple risk factors (Tables 4 and 5).

The first step in treatment of a primary disorder of lipid and lipoprotein metabolism is a diet reduced in total fat, saturated fat, and cholesterol (NCEP step I and step II diets; see Table 6). Weight loss is particularly important, especially in those with hypertriacylglycerolemia, low HDL cholesterol, glucose intolerance, and insulin resistance. Drug treatment is usually based on the particular lipoprotein that is elevated starting with the atherogenic LDL. For every individual, positive and negative risk factors as outlined by the NCEP III guidelines determine whether the aim of treatment with diet, exercise, and drug treatment has been reached (Tables 4 and 5). Guidelines have also been proposed for hypertriacylglycerolemia and low HDL cholesterol. A plasma triacylglycerol level in adults above $400\,mg\,dl^{-1}$ is elevated, $200–400\,mg\,dl^{-1}$ is borderline, and below $200\,mg\,dl^{-1}$ is normal. An HDL cholesterol less than $40\,mg\,dl^{-1}$ is too low.

The drugs of choice to treat an elevated LDL cholesterol level are the HMG CoA reductase inhibitors, or the so-called 'statins.' Inhibition of this enzyme causes a reduction in the hepatic cholesterol pool, leading to an induction of LDL receptors and a fall in blood LDL cholesterol levels (see above). In 1976, a fungal metabolite, called mevinolin or mevastatin, was isolated, which inhibited HMG CoA reductase. Most of the currently available cholesterol-lowering drugs, which target these enzymes, are chemically modified fungal metabolites or completely synthetic analogs of this compound. These include: atorvastatin, fluvastatin, lovastatin, pravastatin, and simvastatin. When given in sufficiently high doses, these drugs can also produce a significant fall in triacylglycerols, presumably by decreasing hepatic VLDL production and by removing the IDL molecules. HDL levels often increase with statin therapy, but the mechanism is not currently known.

Other drugs that are given when the LDL cholesterol level is elevated are the bile acid sequestrants or resins (cholestyramine or colestipol) and niacin.

Table 6 Characteristics of step one and step two diets for patients with dyslipidemias

Nutrient	*Step one diet*	*Step two diet*
Total fat (% of calories)	Average of no more than 30	Same
Saturated fatty acids (%)	< 10	< 7
Polyunsaturated fatty acids (%)	Up to 10	Same
Monounsaturated fatty acids (%)	Remaining total fat calories	Same
Cholesterol (mg dl^{-1})	< 300	< 200
Carbohydrates (% of calories)	About 55	Same
Proteins (% of calories)	About 15–20	Same

From National Cholesterol Education Panel (NCEP) (1993) Adult treatment panel II. *Journal of the American Medical Association* 269: 3015–3023, with permission.

The bile acid sequestrants bind bile acids in the intestine promoting their excretion in the stool. This interferes with the reabsorption of the bile acids, causing more cholesterol to be converted into bile acids in the liver. The hepatic cholesterol level falls, and LDL receptors are induced, causing a fall in the blood LDL cholesterol levels.

Niacin is vitamin B_3 and is present in food at an average level of about 50 mg per day. When niacin is given in amounts 20- to 40-fold as much as this, it can decrease FFA mobilization, leading to a decrease in lipid in the liver, increased proteolysis of apoB-100, and decreased VLDL production. Thus, both the triacylglycerol-rich VLDL, and the product of VLDL catabolism, the cholesterol-enriched LDL, fall as the result of niacin treatment. Niacin has a separate effect on raising HDL cholesterol, presumably by decreasing the catabolism of HDL and its major apolipoprotein, apoA-1, in the liver. Niacin can be given as a regular short-acting preparation or in an extended or sustained release preparation.

Fibric acid derivatives, such as gemfibrozil and fenofibrate, are often used as drugs of choice for marked triacylglycerol elevations. They act by inducing the gene for LPL and repressing the gene for apoC-3. This leads to an increased hydrolysis of triacylglycerol by LPL and a decreased inhibition of LPL by apoC-3.

Combination therapy of HMG CoA inhibitors with fibrates can cause a serious side-effect, called myositis, which can lead to life-threatening rhabdomyolysis, and a specialist should carry out such combined therapy.

Knowledge of lipid disorders including abnormalities in cholesterol metabolism and the consequences of such has grown considerably, and many dyslipidemias are largely treatable with the proscribed means. It is expected that in the 21st century, ASCVD will be reduced considerably in developed countries, providing an increased quality of life and a prolonged life span to humans. In addition, greater attention will need to be paid to developing countries to prevent them from assuming the rates of ASCVD in developed nations.

See also: **Cholesterol**: Properties and Determination; Factors Determining Blood Cholesterol Levels; Role of Cholesterol in Heart Disease; **Diabetes Mellitus**: Problems in Treatment; Secondary Complications; **Fish Oils**: Composition and Properties; Dietary Importance; **Phospholipids**: Properties and Occurrence; Determination; Physiology; **Stress and Nutrition**; **Triglycerides**: Structures and Properties; Characterization and Determination; **Vegetable Oils**: Composition and Analysis; Dietary Importance

Further Reading

Brown MS and Goldstein JL (1997) The SREBP pathway: Regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. *Cell* 89: 331–340.

Cooper AD (1999) Role of the enterohepatic circulation of bile salts in lipoprotein metabolism. *Gastroenterology Clinics of North America* 28: 211–229.

Dietschy JM (1997) Theoretical considerations of what regulates low-density-lipoprotein and high-density-lipoprotein cholesterol. *American Journal of Clinical Nutrition* 65: 158S–159S.

Dietschy JM (1998) Dietary fatty acids and the regulation of plasma low density lipoprotein cholesterol concentrations. *Journal of Nutrition* 128 (Supplement 2): 444S–448S.

Grundy SM (1990) Cholesterol and atheriosclerosis. Diagnosis and treatment. Philadelphia, PA: JB Lippincott Company; New York: Gower Medical Publishing.

Jones PJ (1997) Regulation of cholesterol biosynthesis by diet in humans. *American Journal of Clinical Nutrition* 66: 438–446.

Kritchevsky D (1995) Dietary protein, cholesterol and atherosclerosis: a review of the early history. *Journal of Nutrition* 125 (Supplement 3): 589S–593S.

Kwiterovich PO (1998) The antiatherogenic role of high-density lipoprotein cholesterol. *American Journal of Cardiology* 82: 13Q–21Q.

National Cholesterol Education Panel (NCEP) (1993) Adult treatment panel II. *Journal of the American Medical Association* 269: 3015–3023.

Olson RE (1998) Discovery of the lipoproteins, their role in fat transport and their significance as risk factors. *Journal of Nutrition* 128 (Supplement 2): 439S–443S.

Ordovas JM (1999) The genetics of serum lipid responsiveness to dietary intervention. *Proceedings of the Nutrition Society* 58: 171–187.

Rodriquez-Oquendo A and Kwiterovich PO. Jr. (2000) Disorders of lipid and bile acid metabolism. In: Fernandes J, Saudubray J-M and van den Berghe G (eds) *Inborn Metabolic Disease*, 3rd edn. pp. 321–336. Berlin: Springer.

Factors Determining Blood Cholesterol Levels

S M Grundy, University of Texas, Dallas, TX, USA

Introduction

A high blood (serum) cholesterol level is a major risk factor for atherosclerotic coronary heart disease (CHD). Consequently, there has been much interest in the causes of elevated serum cholesterol concentrations. Although the serum cholesterol can be measured as a single entity, in fact cholesterol is carried in the blood stream by several independent entities called lipoproteins. Each lipoprotein has its own characteristics and the concentrations of each are affected by different factors. Several of these factors are related to diet, i.e., dietary cholesterol, certain fatty acids, and energy imbalance resulting in obesity. Other factors also modify lipoprotein metabolism including advancing age, the postmenopausal state in women, and genetics. Consideration of each of the factors regulating serum cholesterol concentrations first requires a description of the different lipoprotein species.

Serum Lipoproteins

Lipoproteins are macromolecular complexes that consist of discrete particles and are composed of both lipids and proteins. The lipids include cholesterol, phospholipids, and triacylglycerols (TAG). A portion of serum cholesterol is esterified with a fatty acid; the remainder is unesterified. The protein components go by the name of apolipoproteins. The major forms of apolipoproteins and their functions are listed in **Table 1**. Four categories of lipoproteins that carry cholesterol in the serum include chylomicrons, very-low-density lipoproteins (VLDL), low-density lipoproteins (LDL), and high-density lipoproteins (HDL). The characteristics and metabolism of each lipoprotein will be reviewed briefly.

Table 1 Apolipoproteins of serum lipoproteins

Apolipoprotein	*Function*
A-I	Major apolipoprotein of HDL Activator of LCAT
A-I	Structural apolipoprotein of HDL (other functions unknown)
A-IV	Apolipoprotein of chylomicrons (other functions unknown)
B-48	Chylomicron assembly and secretion
B-100	VLDL assembly and secretion Ligand for LDL receptor unknown
C-I	Unknown
C-II	Activator of LPL
C-III	Inhibitor of LPL
E	Apolipoprotein of remnant lipoproteins Ligand for LDL receptor Promotes hepatic uptake of remnants

HDL, high-density lipoproteins; LDL, low-density lipoproteins; VLDL, very-low-density lipoproteins; LCAT, lecithin cholesterol acyl transferase; LPL, lipoprotein lipase.

Chylomicrons

Dietary cholesterol enters the intestine together with fat, which is predominantly TAG. The latter undergoes hydrolysis by pancreatic lipase and releases fatty acids and monoacylglycerols. In the intestine, these mix with bile acids, phospholipids, and cholesterol from the bile. The mixture of hydrolyzed lipids associates with phospholipids and bile acids to form mixed micelles. Fatty acids, monoacylglycerols and cholesterol are taken up by the intestinal mucosa. In the mucosal cells, the fatty acids and monoacylglycerols are recombined by enzymatic action to form TAG, which are incorporated with the cholesterol into lipoprotein particles called chylomicrons. Most of the cholesterol in chylomicrons as esterified with a fatty acid. The major apolipoprotein of chylomicrons is apo B-48; other apolipoproteins – apo Cs, apo Es and apo As – attach to the surface coat of chylomicrons and aid in metabolic processing. In the mucosal cells, microsomal lipid transfer protein (MTP) facilitates the transfer of TAG and cholesterol ester into chylomicron particles. The presence of MTP is required for the secretion of chylomicrons from mucosal cells.

Chylomicrons are secreted by intestinal mucosal cells into the lymphatic system. From here they pass through the thoracic duct into the systemic circulation. When chylomicrons enter the peripheral circulation they come into contact with an enzyme, lipoprotein lipase (LPL), which is located on the

endothelial surface of capillaries. LPL is activated by apo C-II on chylomicrons; this process is modulated by apo C-III, an inhibitor of LPL activity. None the less, most chylomicron TAG is hydrolyzed by LPL; a residual lipoprotein particle, named chylomicron remnant, is released into the circulation and is rapidly removed by the liver. Hepatic uptake of chylomicron remnants is believed to be mediated by binding of remnants with a glycoprotein on the surface of liver cells. Almost all newly absorbed cholesterol thus enters the liver in association with chylomicron remnants.

Very-Low-density Lipoproteins

The liver also secretes a TAG-rich lipoprotein called VLDL. Fatty acids used in the synthesis of TAG in the liver are normally derived from circulating nonesterified fatty acids (NEFA); even so, the liver has the capacity to synthesize fatty acids when the diet contains mainly carbohydrate. MTP inserts TAG into newly forming VLDL particles. The surface coat of VLDL contains unesterified cholesterol, phospholipids and apolipoproteins. The major apolipoprotein of VLDL is apo B-100. Other apolipoproteins, notably apo Cs and apo Es, are also present. As VLDL circulate they acquire cholesterol esters from HDL. Circulating VLDL particles lose TAG through interaction with LPL in the peripheral circulation; in this process, VLDL are transformed into VLDL remnants. The latter can have two fates: hepatic uptake or conversion to LDL. Hepatic uptake of VLDL remnants may occur via two mechanisms: interaction with glycoproteins or interaction with LDL receptors. Both glycoproteins and LDL receptors are located on the surface of liver cells.

Low-density Lipoproteins

Conversion of VLDL remnants to LDL appears to be largely the result of hydrolysis of remaining TAG by hepatic triacylglycerol lipase (HTGL). Normally about two-thirds of cholesterol is carried by LDL; most of this LDL cholesterol exists in the form of esters. The only apolipoprotein in LDL is apo B-100. LDL is removed from the circulation largely by hepatic LDL receptors. The level of expression of LDL receptors is a major determinant of serum LDL cholesterol concentrations. The synthesis of LDL receptors is regulated in large part by the liver's content of cholesterol. An increase in hepatic cholesterol content suppresses LDL receptor synthesis and raises serum LDL cholesterol; conversely, a decrease in hepatic cholesterol stimulates receptor synthesis and lowers serum LDL cholesterol. The mechanism whereby hepatic cholesterol controls LDL receptor synthesis is through a regulatory protein called sterol regulatory element-binding protein (SREBP). When hepatic cholesterol content falls, SREBP is activated and stimulates the synthesis of LDL receptors.

The regulatory form of cholesterol in the liver cell is unesterified cholesterol, not cholesterol ester. The hepatic content of unesterified cholesterol depends on several factors, including the amounts of cholesterol derived from chylomicrons and other lipoproteins, hepatic synthesis of cholesterol, secretion of cholesterol into bile, conversion of cholesterol into bile acids, esterification of cholesterol, and secretion of cholesterol into serum with VLDL. Factors that influence each of these processes can alter serum LDL cholesterol concentrations by modifying the hepatic content of unesterified cholesterol and thereby expression of LDL receptors.

High-density Lipoproteins

HDL consist of a series of lipoprotein particles of relatively high density, all of which contain apo A-I. A proportion of HDL particles also contain apo A-II. Some HDL species (HDL_3) are more dense than others (HDL_2). HDL particles are composed largely of byproducts of catabolism of TAG-rich lipoproteins. The surface coats of HDL particles contain phospholipids and unesterified cholesterol, apo A-I with or without apo A-II, and other apolipoproteins (apo Cs and apo Es). Their particle cores consist largely of cholesterol esters, although small amounts of TAG are also present. The cholesterol esters of HDL are formed by esterification with a fatty acid through the action of an enzyme, lecithin cholesterol acyl transferase (LCAT); the substrates for this reaction derive from either unesterified cholesterol released during lipolysis of TAG-rich lipoproteins or from the surface of peripheral cells. After esterification of cholesterol, the cholesterol esters of HDL are then transferred back to TAG-rich lipoproteins and eventually are removed by the liver through direct uptake of remnant lipoproteins or LDL. Whether whole HDL particles can be directly removed from the circulation is uncertain. Some investigators believe that the HDL components are dismantled and removed piecemeal.

Dietary Regulation of Serum Lipoproteins

A large body of research has shown that diet has a major impact on the concentrations and composition of serum lipoproteins, and hence on serum cholesterol concentrations. Three major factors affect cholesterol and lipoprotein concentrations: dietary cholesterol, the macronutrient composition of the diet, particularly dietary fatty acids, and energy

balance, as reflected by body weight. The influence of each of these factors can be considered.

Dietary Cholesterol

All dietary cholesterol is derived from animal products. The major sources of cholesterol in the diet are egg yolks, products containing milk fat, animal fats, and animal meats. Many studies have shown that high intakes of cholesterol will increase the serum cholesterol concentration. Most of this increase occurs in the LDL cholesterol fraction. When cholesterol is ingested, it is incorporated into chylomicrons and makes its way to the liver with chylomicron remnants. There it raises hepatic cholesterol content and suppresses LDL receptor expression. The result is a rise in serum LDL cholesterol concentrations. Excess cholesterol entering the liver is removed from the liver either by direct secretion into bile or by conversion into bile acids; also, dietary cholesterol suppresses hepatic cholesterol synthesis. There is considerable variability in each of these steps in hepatic cholesterol metabolism; for this reason the quantitative effects of dietary cholesterol on serum LDL cholesterol levels vary from one person to another. For every 200 mg of cholesterol per day in the diet, serum LDL cholesterol is increased on average by about 6 mg dl^{-1} (0.155 mmol l^{-1}).

Macronutrient Composition of the Diet

Dietary fat and fatty acids Most of the fat in the diet consists of TAG that are composed of three fatty-acid molecules bonded to glycerol. The contribution of TAG to total energy intake varies among individuals and populations, ranging from 15% to 40% of total nutrient energy. The fatty acids of TAGs are of several types: saturated, *cis*-monounsaturated, *trans*-monounsaturated, and polyunsaturated fatty acids. All fatty acids affect lipoprotein levels in one way or another. Table 2 lists the major fatty acids of the diet and denotes their effects on serum lipoproteins. Also shown are the effects of carbohydrates, which also influence serum lipoprotein metabolism. It should be noted that all lipoprotein responses are compared and related to those of *cis*-monounsaturated fatty acids which are widely accepted to be neutral, or baseline.

Saturated fatty acids The saturated fatty acids are derived from both animal fats and plant oils. Rich sources of dietary saturated fatty acids include butter fat, meat fat, and tropical oils (palm oil, coconut oil, and palm kernel oil). Saturated fatty acids are straight-chain organic acids with an even number of carbon atoms (Table 2). All saturated fatty acids that have from eight to 16 carbon atoms raise the serum LDL cholesterol concentration when they are consumed in the diet. In the USA and much of Europe, saturated fatty acids make up 12–15% of total nutrient energy intake.

The mechanisms whereby saturated fatty acids raise LDL cholesterol levels are not known, although available data suggest that they suppress the expression of LDL receptors. The predominant saturated fatty acid in most diets is palmitic acid ($C_{16:0}$); it is cholesterol-raising when compared with *cis*-monounsaturated fatty acids, specifically oleic acid ($C_{18:cis1\ n\text{-}9}$), which

Table 2 Macronutrient effects on serum lipoprotein cholesterol

Nutrient	*Symbol*[a]	*VLDL cholesterol*	*LDL cholesterol*	*HDL cholesterol*[b]
Fatty acids				
Saturated				
Palmitic	$C_{16:0}$	–[b]	↑↑	–
Myristic	$C_{14:0}$	–	↑↑↑	↓
Lauric	$C_{12:0}$	–	↑	–
Caproic	$C_{10:0}$	–	↑	–
Caprilic	$C_{8:0}$	–	↑	–
Stearic	$C_{18:0}$	–	–	– or ↓
trans-Monounsaturated	*trans* $C_{18:1\ n\text{-}9}$	–	↑ or ↑↑	↓
cis-Monounsaturated	*cis* $C_{18:1\ n\text{-}9}$	–	–	–
Polyunsaturated				
n-6[c]	$C_{18:2\ n\text{-}6}$	– or ↓	– or ↓	– or ↓
n-3[c]	DHA, EPA	↓↓↓	– or ↓	–
Carbohydrate		↑↑↑	–	↓↓

[a]First number denotes number of carbon atoms; second number denotes number of double bonds.
[b]The dash (–) indicates that there is no change in level compared to that produced by *cis*-monosaturated fatty acids (oleic acid) ($C_{18:1\ n\text{-}9}$). All the lipoprotein responses to oleic acid are considered 'neutral', i.e., no effect.
[c]The letter 'n' and number indicate at which carbon atom, numbered from the terminal methyl group, the first double bond appears.
VLDL, very-low-density lipoproteins; LDL, low-density lipoproteins; HDL, high-density lipoproteins; DHA, docosahexanoic acid ($C_{22:6\ n\text{-}3}$); EPA, eicosapentanoic acid ($C_{20:5\ n\text{-}3}$).

is considered to be 'neutral' with respect to serum cholesterol concentrations. In other words, oleic acid is considered by most investigators to have no effect on serum cholesterol or lipoproteins. Another saturated fatty acid, myristic acid ($C_{14:0}$), apparently raises LDL cholesterol concentrations somewhat more than does palmitic acid, whereas other saturates – lauric ($C_{12:0}$), caproic ($C_{10:0}$), and caprylic ($C_{8:0}$) acids – have a somewhat lesser cholesterol-raising effect. On average, for every 1% of total energy consumed as cholesterol-raising saturated, fatty acids, compared with oleic acid, the serum LDL cholesterol level is raised about 2 mg dl^{-1} (0.025 mmol l^{-1}).

One saturated fatty acid, stearic acid ($C_{18:0}$), does not raise serum LDL cholesterol concentrations. The main sources of this fatty acid are beef tallow and cocoa butter. The reason for its failure to raise LDL cholesterol concentrations is uncertain, but may be the result of its rapid conversion into oleic acid in the body.

Trans-monounsaturated fatty acids These fatty acids are produced by hydrogenation of vegetable oils. Intakes of *trans*-monounsaturates vary from one country to another depending on consumption of hydrogenated oils. In many countries they contribute between 2% and 4% of total nutrient energy intake. A series of *trans* acids are produced by hydrogenation; most are monounsaturated. For many years, it was accepted that *trans*-monounsaturated fatty acids were neutral with respect to LDL cholesterol concentrations. However, recent studies have shown that they raise LDL cholesterol concentrations to a level similar to that of palmitic acid when substituted for dietary oleic acid. In addition, they cause a small reduction in serum HDL cholesterol concentrations. Thus, *trans*-monounsaturates must be placed in the category of cholesterol-raising fatty acids.

Cis-monounsaturated fatty acids The major fatty acid in this category is oleic acid ($C_{18:cis1\ n\text{-}9}$). It is found in both animal and vegetable fats, and typically is the major fatty acid in diet. Intakes commonly vary between 10% and 20% of total energy. Oleic acid intake is particularly high in the Mediterranean region where large amounts of olive oil are consumed. Other sources rich in oleic acid are rapeseed oil (canola oil) and high-oleic forms of safflower and sunflower oils. Peanuts and pecans are also high in oleic acid. Animal fats likewise contain a relatively high percentage of oleic acid among all their fatty acids; even so, these fats also tend to be rich in saturated fatty acids. When high-carbohydrate diets are consumed, the human body can synthesize fatty acids; among these, oleic acid is the predominant fatty acid produced.

As indicated before, oleic acid is generally considered to be the 'baseline' fatty acid with respect to serum lipoproteins levels, i.e., it does not raise (or lower) LDL cholesterol or VLDL cholesterol concentrations, nor does it lower (or raise) HDL cholesterol concentrations. It is against this 'neutral' fatty acid that responses of other fatty acids are defined (Table 2). For example, when oleic acid is substituted for cholesterol-raising fatty acids, the serum LDL cholesterol concentration will fall. None the less, oleic acid is not designated a cholesterol-lowering fatty acid, but instead, this response defines the cholesterol-raising potential of saturated fatty acids.

Polyunsaturated fatty acids There are two categories of polyunsaturated fatty acids: n-6 and n-3. The major n-6 fatty acid is linoleic acid ($C_{18:2,n\text{-}6}$). It is the predominant fatty acid in many vegetable oils, e.g., corn oil, soya bean oil, and high linoleic forms of safflower and sunflower seed oils. Intakes of linoleic acid typically vary from 4 to 10% of nutrient energy, depending on how much vegetable oil is consumed in the diet. The n-3 fatty acids include linolenic acid ($C_{18:3,n\text{-}3}$), docosahexanoic acid (DHA) ($C_{22:6,n\text{-}3}$), and eicosapentanoic acid (EPA) ($C_{20:5,n\text{-}3}$). Linolenic acid is high in linseed oil and present in smaller amounts in other vegetable oils. DHA and EPA are enriched in fish oils.

For many years, linoleic acid was thought to be a unique LDL cholesterol-lowering fatty acid. Recent investigations suggest that earlier findings overestimated the LDL-lowering potential of linoleic acid. Even though substitution of linoleic acid for oleic acid in the diet may reduce LDL cholesterol levels in some people, a difference in response is not consistent. Only when intakes of linoleic acid become quite high do any differences become apparent. At high intakes, however, linoleic acid also lowers serum HDL cholesterol concentrations. Moreover, compared with oleic acid, it may reduce VLDL cholesterol levels in some people. Earlier enthusiasm for high intakes of linoleic acid to reduce LDL cholesterol levels has been dampened for several reasons: for example, its LDL-lowering ability does not offset potential disadvantages of HDL lowering; and other concerns include possible untoward side-effects such as promoting oxidation of LDL and suppressing cellular immunity to cancer.

The n-3 fatty acids in fish oils (DHA and EPA) have a powerful action to reduce serum VLDL levels. This action apparently results from suppression of the secretion of VLDL by the liver. The precise mechanism

for this action is not known. However these fatty acids do not reduce LDL cholesterol concentrations relative to oleic acid. They have been used for one treatment of some patients with elevated VLDL concentrations, although drug treatment is generally employed when it is necessary to lower serum VLDL levels.

Carbohydrate When carbohydrates are substituted for oleic acid in the diet, serum LDL cholesterol levels remain unchanged. However, VLDL cholesterol concentrations usually rise and HDL cholesterol concentrations fall on high-carbohydrate diets. Thus, a lack of difference in total serum cholesterol concentrations during the exchange of carbohydrate and oleic acid is misleading. The two categories of nutrients have different actions on lipoprotein metabolism. The differences in response to dietary carbohydrate and oleic acid provide a good example of how measurements of serum total cholesterol fail to reveal all of the changes that are occurring in the lipoprotein fractions.

Energy Balance

Obesity When energy intake exceeds energy expenditure, the balance of energy is stored in adipose tissue in the form of TAG. When the TAG content of adipose tissue becomes excessive (body mass index 30 or above), a state of obesity is said to exist. In some obese persons, excessive accumulations of TAG occur in other tissues than adipose tissue. Two such tissues are skeletal muscle and liver. High contents of TAG in muscle and liver arise due in large part to continuous leakage of excessive quantities of NEFA from adipose tissue. In the presence of desirable body weight, normal insulin levels are sufficient to suppress hydrolysis of TAG in adipose tissue, and NEFA release is low. On the other hand, in obese persons NEFA release is excessive and skeletal muscle and liver are flooded with high serum NEFA concentrations. The result is engorgement of these organs with TAG. When skeletal muscle is overloaded with TAG, insulin-mediated glucose uptake is impaired. This condition is called insulin resistance. When liver is packed with TAG, hepatic metabolism is altered and insulin action on the liver is deranged. As a result, there is an overproduction of VLDL; this leads to high VLDL cholesterol concentrations, and because LDL is a product of VLDL, to higher LDL cholesterol levels. In addition, obesity is accompanied by a reduction in HDL cholesterol concentrations. Thus obesity is responsible for multiple alterations in lipoprotein metabolism; it has significant effects on three major lipoprotein species – VLDL, LDL, and HDL. These changes appear to be the result of a combination of excessive hepatic TAG as a substrate for VLDL formation and failure of insulin to exert its usual action to curtail VLDL secretion.

Exercise Many of the adverse metabolic effects of obesity are reversed by exercise. Increased energy expenditure through regular and sustained exercise helps to prevent accumulation of excessive quantities of TAG in adipose tissue. In addition, increased muscle metabolism produced by exercise burns off NEFA and prevents TAG accumulation in the liver. Hence, increased and sustained energy expenditure favorably modifies the lipoproteins, particularly by lowering VLDL cholesterol concentrations and raising serum HDL cholesterol. The effects of exercise on LDL cholesterol concentrations are more modest, but in some people exercise produces a reduction.

Other Factors Affecting Serum Lipoproteins

Advancing Age

Between the ages of 20 and 50 years, there is a gradual rise in serum cholesterol concentrations. In the USA, for example, the serum cholesterol increases on average about $50\,\mathrm{mg\,dl^{-1}}$ ($1.295\,\mathrm{mmol\,l^{-1}}$). This change may be related in part to increasing obesity, according to the mechanisms described above. However, even in people who do not gain weight with advancing age, serum cholesterol concentrations usually rise to some extent. Available evidence indicates that this rise results from a decrease in expression of LDL receptors. The reasons for a decline in receptor synthesis with aging are not known, but may reflect 'metabolic' aging. However, in men, after age 50 years, there is little further rise in serum cholesterol. This observation suggests that the impact of weight gain, which occurs mostly between the ages of 20 and 50 years, may be greater than generally recognized.

Postmenopausal State in Women

In women, there is a further rise in serum cholesterol concentrations which occurs after age 50 years. This rise is believed to be due largely to loss of estrogens after the menopause. Estrogens are known to stimulate the synthesis of LDL receptors and, consequently, receptor expression declines after the menopause. This increment in cholesterol levels can be largely reversed by estrogen replacement therapy.

Genetics

Family studies and research in twins indicate that about 50% of the variation of serum cholesterol concentrations in the general populations can be explained by genetic polymorphisms. Presumably

this variation is related to factors that regulate lipoprotein concentrations. In some cases, specific genetic defects are severe, resulting in marked changes in lipoprotein concentrations. When this occurs, the affected individual is said to have a monogenic disorder. In other cases, multiple genetic modifications are present that combine to alter lipoprotein concentrations. When a few modifications are present, the condition is called oligogenic, but when many modifications combine to change lipoprotein concentrations, the condition is named polygenic. Several monogenic disorders have been identified; a few oligogenic conditions have been described; but there are very few instances in which complex polygenic traits have been unraveled. A question of great interest is whether nutritional and genetic factors ever interact synergistically to alter lipoprotein concentrations. Undoubtedly, dietary factors and genetic changes can be additive in their effects on serum lipoproteins; but synergistic interaction has been difficult to prove. In what follows, consideration will be given to the impact of modification of some of the key gene products regulating lipoprotein metabolism.

LDL Receptors The most severe elevations in LDL cholesterol levels occur in patients who have mutations in the gene encoding for LDL receptors. About one in 500 people are heterozygous for these mutations. Their condition is called heterozygous familial hypercholesterolemia. LDL cholesterol concentrations are essentially twice the normal level in this condition. Very rarely patients are homozygous for mutation in the LDL receptor gene, and thus have homozygous familial hypercholesterolemia. Their LDL cholesterol levels are approximately four times normal. Individuals with this condition develop severe, premature atherosclerosis.

Many other people appear to have a reduction in LDL receptor expression on a genetic basis, but they do not have as severe elevations of serum LDL cholesterol as occur in patients with familial hypercholesterolemia. Presumably, these people have genetic modifications in factors that regulate transcription of the LDL receptor gene. Although such genetic modifications may be relatively common, they are poorly defined. Again, an important but unanswered question is whether some people are genetically susceptible to the cholesterol-raising effects of dietary cholesterol and saturated fatty acids. If so, they may possess modifications in the genetic control of LDL receptor expression.

Apolipoprotein B-100 Structure About one in 500 people also have a mutation in the primary structure of apo B that interferes with its binding to LDL receptors. This mutation gives rise to the disorder called familial defective apolipoprotein B-100. The consequence is an elevation of LDL cholesterol concentrations, and the clinical pattern resembles that of familial hypercholesterolemia.

Apolipoprotein B Synthesis Rare patients have mutations in the gene encoding for apo B that impair the synthesis of this apolipoprotein. Such patients usually have very low LDL cholesterol concentrations. These individuals are said to have familial hypobetalipoproteinemia. In other rare cases, the intracellular TAG transport protein called MCT (medium chain triglyceride) is genetically absent; when this occurs, no lipoprotein particles containing apo B can be formed. LDL cholesterol is absent from serum, and the disorder is called familial abetalipoproteinemia.

Some researchers speculate that serum elevations in VLDL cholesterol and LDL cholesterol can result from excessive synthesis and/or secretion of apo B-containing lipoproteins by the liver. When this occurs on a genetic basis, the disorder is designated familial combined hyperlipidemia. However, a monogenic basis of this clinical phenotype has never yet been identified. Therefore, most investigators have concluded that familial combined hyperlipidemia most likely represents an oligogenic or polygenic disorder. In this disorder, lipoprotein elevations appear to be worsened by nutritional factors, particularly by obesity.

Apolipoprotein E This apolipoprotein is present on TAG-rich lipoproteins and it facilitates removal of remnant lipoproteins by LDL receptors in the liver. When apo E is affected by mutation, this enabling action is curtailed and hepatic uptake of remnant lipoproteins is impaired. The result is an accumulation of chylomicron remnants and VLDL remnants in the circulation. The accumulation is accentuated by the coexistence of other disorders of metabolism of TAG-rich lipoproteins. When remnant accumulation occurs on a genetic basis, the disorder is called familial dysbetalipoproteinemia.

Apolipoprotein C There are two forms of apo C – apo C-II and apo C-III. Apo C-II is required for activation of LPL; when it is genetically absent, affected patients develop severe elevations of TAG-rich lipoproteins. Apo C-III inhibits the activity of LPL. In certain metabolic disorders, notably insulin resistance, synthesis of apo C-III is increased; an elevated apo C-III can lead to impaired function of LPL and increases in serum concentrations of TAG-rich lipoproteins.

Apolipoprotein A-I This is the major apolipoprotein of HDL. Rare patients have mutations in apo

A-I which results in very low concentrations of HDL cholesterol. However, most people in whom HDL cholesterol concentrations are moderately reduced show increased catabolism of apo A-I. The mechanism for this change has not been fully determined, but one important cause may be an overexpression of HTGL.

Lipoprotein Lipase This enzyme is required for lipolysis of TAG in TAG-rich lipoproteins. Rare patients are homozygous for mutations in LPL that impair its function. In such patients, serum concentrations of chylomicrons are markedly increased. The accumulation of chylomicrons in serum is greatly accentuated by the presence of fat in the diet. Only by severe dietary fat restriction is it possible to prevent severe TAG elevations in serum.

Genetic Regulation of HDL Cholesterol Family and twin studies reveal that about 50% of the variation in serum HDL cholesterol levels in the general population is explained by genetic factors. However, the regulation of HDL cholesterol concentrations is complex, and HDL cholesterol levels are determined by many factors, e.g., serum TAG concentrations, activity of HTGL, production rates of apo A-I, and activities of cholesterol ester transfer protein (CETP) and LCAT. Genetic factors undoubtedly affect each of these regulating factors.

See also: **Eggs**: Dietary Importance; **Exercise**: Muscle; **Fatty Acids**: Properties; Metabolism; Gamma-linolenic Acid; Analysis; Dietary Importance; *Trans*-fatty Acids: Health Effects; **Meat**: Nutritional Value; **Obesity**: Etiology and Diagnosis; **Poultry**: Chicken; Ducks and Geese; Turkey

Further Reading

Bonanome A and Grundy SM (1988) Effect of dietary stearic acid on plasma cholesterol and lipoprotein levels. *New England Journal of Medicine* 318: 1244–1248.

Cater NB, Heller HJ and Denke MA (1997) Comparison of the effects of medium-chain triacylglycerols, palm oil, and high oleic sunflower oil on plasma acylglycerol fatty acids and lipid and lipoprotein concentrations in humans. *American Journal of Clinical Nutrition* 65: 41–45.

Connor WE (1988) Effects of omega-3 fatty acids in hypertriglyceridemic states. *Seminars in Thrombosis and Hemostasis* 14: 271–284.

Denke MA, Sempos CT and Grundy SM (1993) Excess body weight: an underrecognized contributor to high blood cholesterol levels in white American men. *Archives of Internal Medicine* 153: 1093–1103.

Dietschy JM, Turley SD and Spady DK (1993) Role of liver in the maintenance of cholesterol and low density lipoprotein homeostasis in different animal species, including humans. *Journal of Lipid Research* 34: 1637–1659.

Ericsson S, Eriksson M, Vitols S, Einarsson K, Berglund L and Angelin B (1991) Influence of age on the metabolism of plasma low density lipoproteins in healthy males. *Journal of Clinical Investigation* 87: 591–596.

Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (SM Grundy, chairman) (1994) National cholesterol education program: second report of the expert panel on detection, evaluation, and treatment of high blood cholesterol (adult treatment panel II). *Circulation* 89: 1329–1445.

Grundy SM (1986) Comparison of monounsaturated fatty acids and carbohydrates for lowering plasma cholesterol. *New England Journal of Medicine* 314: 745–748.

Grundy SM (1991) Multifactorial etiology of hypercholesterolemia: implications for prevention of coronary heart disease. *Arteriosclerosis and Thrombosis* 11: 1619–1635.

Grundy SM and Denke MA (1990) Dietary influences on serum lipids and lipoproteins. *Journal of Lipid Research* 31: 1149–1172.

Hegsted DM, McGandy RB, Myers ML and Stare FJ (1965) Quantitative effects of dietary fat on serum cholesterol in man. *American Journal of Clinical Nutrition* 17: 281–295.

Innerarity TL, Mahley RW, Weisgraber KH *et al*. (1990) Familial defective apolipoprotein B-100: a mutation of apolipoprotein B that causes hypercholesterolemia. *Journal of Lipid Research* 31: 1337–1349.

Keys A, Anderson JT and Grande F (1965) Serum cholesterol response to changes in the diet. IV. Particular saturated fatty acids in the diet. *Metabolism* 14: 776–787.

Mattson FH and Grundy SM (1985) Comparison of effects of dietary saturated, monounsaturated, and polyunsaturated fatty acids on plasma lipids and lipoproteins in man. *Journal of Lipid Research* 26: 194–202.

Mensink RP and Katan MB (1990) Effect of dietary trans fatty acids on high-density and low-density lipoprotein cholesterol levels in healthy subjects. *New England Journal of Medicine* 323: 439–445.

Role of Cholesterol in Heart Disease

R M Shireman, University of Florida, Gainesville, FL, USA

Background

The compound cholesterol (3β-cholest-5-en-3-ol) ($C_{27}H_{46}O$) (**Figure 1**) was first isolated from gallstones in the late eighteenth century. Its crystalline structure has been determined by X-ray diffraction.

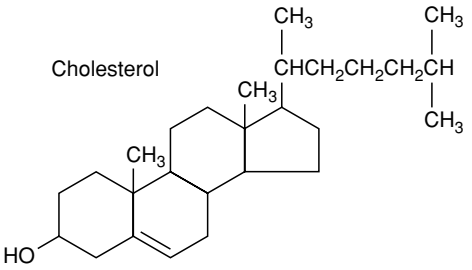

Figure 1 Structure of cholesterol.

More Nobel prizes have been awarded to scientists who worked on cholesterol synthesis, metabolism, and function than on any other compound. Much of the work on the complex biosynthetic pathway was done in the 1950s by Konrad Bloch, Feodor Lynen, John Cornforth, and George Popják, yet many questions remain unanswered. Vertebrates and most other animals can synthesize cholesterol *de novo*. Similar sterol compounds called phytosterols are synthesized by the same pathway in plant cells. Cholesterol is essential in all animal cells, and either too much or too little is detrimental. A familiar example of cholesterol accumulation occurs in the disease atherosclerosis, in which build-up occurs in the lining of the lumen of arteries. A less familiar example is Wolman's disease, a hereditary deficiency in an enzyme, cholesteryl ester hydrolase, in which cholesterol esters rapidly accumulate to very high levels in fetal cells, with ensuing mental retardation and other problems, leading to infantile death. Conversely, Smith–Laemli–Opitz syndrome is characterized by an inherited enzyme defect in the last step in cholesterol synthesis, resulting in very low cholesterol levels in cells, severe mental retardation, and abnormal skeletal development. The major consequence of these enzyme deficiencies that produce very high vs. very low intracellular cholesterol is severe mental retardation; this illustrates the importance of the strict regulation of cholesterol concentrations in cells. Cholesterol is present in high amounts in the brain and nervous system; in fact, about a quarter of the body's cholesterol is present in myelin. Table 1 lists some of the important functions of this sterol.

It is estimated that the adult human body contains about 100 g of cholesterol. The average adult requires approximately 1.1 g of cholesterol per day. About 15–25% of this is normally derived from the diet, while the rest is from endogenous synthesis, mainly in the liver and intestine. This explains why it is rather difficult to reduce blood cholesterol levels to any great extent by decreasing the intake of dietary cholesterol. Cholesterol absorption varies widely among individuals and also in the same individual,

Table 1 Major functions of cholesterol

Function	*Comments*
Membrane structure and integrity	Structural and protective functions; helps maintain the semipermeable barrier; especially important in myelin structure
Precursor to steroid hormones	Required for synthesis of progesterone, testosterone, estrogen, corticosteroids, and aldosterone
Precursor to vitamin D	Cholesterol in skin cells is converted to a precursor form under the influence of UV irradiation (sunlight)
Precursor to bile acids	Cholesterol and bile acids are emulsifiers in bile and aid in digestion and absorption of dietary fat

depending on other constituents in the diet. Cholesterol absorption is very high in about 20% of the population; these individuals have been referred to as 'hyper-responders.' Plant sterols are not absorbed well, but they do decrease cholesterol absorption. A number of factors have been reported to affect absorption besides phytosterols, including vitamin C, several different metal ions, and various types of protein or fiber, but there is little general agreement about the real physiologic effects, if any, of these agents. Unesterified cholesterol that enters the enterocyte (intestinal cell) is esterified before packaging into chylomicrons. Chylomicrons are large lipoprotein particles assembled in enterocytes and transported first in the lymph and then by the bloodstream to the liver. If insufficient cholesterol is absorbed to provide for the packaging of other absorbed fats into chylomicrons, enterocytes synthesize the needed cholesterol. Cholesterol biosynthesis is regulated somewhat differently in different cells, but feedback inhibition of synthesis occurs in most cells except enterocytes. The exact regulation of this process in enterocytes is still unclear, but it is evident that increased dietary fat absorption increases the intestinal biosynthesis of cholesterol.

Cholesterol Metabolism

Cholesterol is not metabolized for energy. It is removed from the body via desquamating cells and by the excretion of cholesterol and its products, including bile acids and steroid hormone catabolites. A relatively constant level is maintained in the body by controlling the rate of biosynthesis, rather than by conversion to products. Although most cells are capable of synthesizing cholesterol, the liver is responsible for much of the cholesterol subsequently delivered to cells such as neurons. Important clues to the synthesis of sterols came from the work of

Konrad Block in the early 1940s, when he fed (U)[^{14}C]acetate to rats and found all 27 carbons of cholesterol to be derived from acetate. The discovery that mevalonic acid is an intermediate allowed great progress in defining this pathway, which has a huge number of enzymatic steps. An abbreviated pathway is shown in Figure 2. Several important compounds in addition to cholesterol are synthesized in this pathway, commonly called the isoprenoid pathway.

Numerous genetic factors, as well as levels of hormones and nutrients, influence cholesterol homeostasis. Although the isoprenoid pathway may be controlled at several separate enzymatic steps, the key regulatory enzyme is hydroxymethylglutaryl-coenzyme A reductase (HMG-CoA). HMG-CoA reductase is an endoplasmic reticulum enzyme subject to a high degree of regulation by several and varied means. The enzyme is regulated acutely in response to extracellular signals through a cAMP-mediated phosphorylation cascade. Phosphorylation decreases the activity of this enzyme. Longer-term control depends on the regulation of transcription of the gene for reductase. Sterol accumulation normally decreases mRNA synthesis of reductase and may accelerate reductase degradation. A sterol regulatory element (SRE-1) is the recognition site for a specific DNA-binding protein whose activity is influenced by intracellular sterol levels. The SRE of the reductase gene promoter shares homology with those of HMG-CoA synthetase and low-density lipoprotein (LDL) receptor, suggesting coordinate control of these pathways by sterol effectors. Coordinate control of the three proteins allows very fine regulation of this synthetic pathway. Thus, when the intracellular concentrations of cholesterol or oxysterol become elevated, decreased transcription of two major genes in the cholesterol biosynthetic pathway leads to a decline in cholesterol biosynthesis by that cell, and decreased transcription of the LDL receptor protein leads to a decrease in cellular uptake of cholesterol via LDL. Reductase synthesis and activity are influenced also by hormones, especially insulin and glucagon.

The rate of synthesis of cholesterol is increased in proliferating normal tissues and tumors. Inhibition of cholesterol synthesis inhibits cell growth. In addition to the needs for cholesterol, other compounds from this pathway have important functions in dividing cells. For example, the farnesyl moiety activates some of the proteins involved in regulation of cell growth, and farnesyl and other prenyl groups are used in modifying certain membrane proteins. Inhibition of HMG-CoA reductase can affect the synthesis of all products of the isoprene pathway. A large number of pharmacologic inhibitors, mostly competitive inhibitors of reductase, are now on the market. Obviously, cholesterol-lowering drugs that inhibit reductase should be used with great caution in growing children, and not at all during pregnancy. *In vitro* studies have shown that compounds such as the monounsaturated fatty acid oleic acid inhibit reductase activity to some extent. Certain xenobiotic compounds, such as phenobarbital, which induce enzymes of the cytochrome P450 system, also affect some of the enzymes of cholesterol metabolism. Other enzymes affecting cholesterol metabolism are

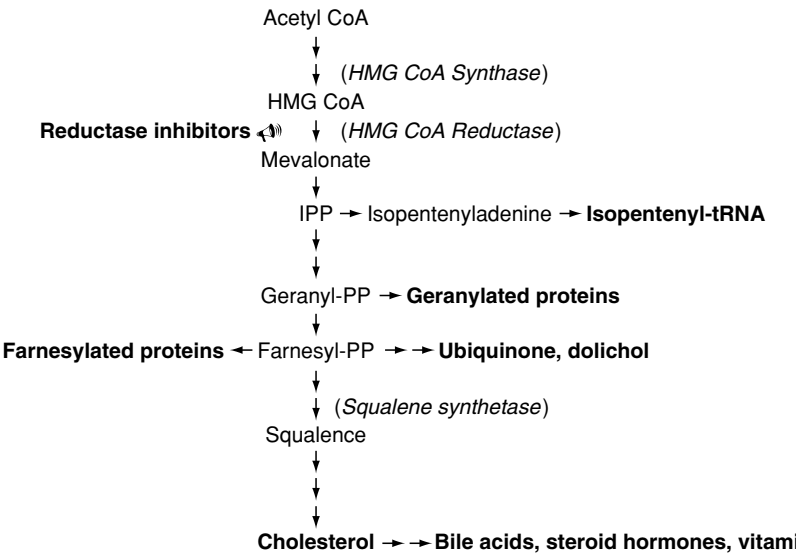

Figure 2 Major steps in the biosynthetic pathway for cholesterol. The major regulated enzymes are shown in italics.

7α-hydroxylase, acyl CoA cholesterol acyl transferase (ACAT) and lecithin cholesterol acyl transferase (LCAT). The 7α-hydroxylase catalyzes the first step in the conversion of cholesterol to bile acids, while ACAT and LCAT are responsible for esterifying cholesterol in cells and plasma, respectively.

Cholesterol and Lipoprotein Transport

Cholesterol is transported along with other lipids by the plasma lipoproteins. These are classified according to their size and density, which are functions of their compositions (Table 2). Each class is characterized by specific proteins (apoproteins), some of which have known important functions (Table 3). Mutations in apoproteins are involved in several types of hyperlipidemia. Although the liver contains a large amount of cholesterol because of immense amounts of membraneous organelles, it does not store extra cholesterol. Normal nascent (newly secreted by the liver) very-low-density lipoprotein (VLDL) has a cholesterol ester (CE)/triacylglycerol (TG) ratio of < 1:4, but CE-rich VLDL may have a ratio of 1:1. CE-rich VLDL is synthesized by the liver in some experimental animal models fed very high levels of dietary cholesterol. Such VLDL has extra apoenzyme E (apoE). This may be important because macrophages have not only the apoenzyme (apoB)/LDL receptor, but also a B/E receptor, which will bind apoB and/or apoE lipoproteins, as discussed below.

Circulating VLDL loses much of its triglyceride to tissues and some of its other components to high-density lipoprotein (HDL); in the process, VLDL becomes transformed to the cholesterol-rich LDL particle. LDL binds to its specific receptor on cells and is carried into the cell by receptor-mediated endocytosis. The major function of the LDL receptor is delivery of lipids to cells, but in the liver, it helps remove LDL and small VLDL from the bloodstream. The receptor is subject to transcriptional control in which high [cholesterol] leads to decreased receptor synthesis, thus downregulating the entry of apoB-containing lipoproteins. The LDL receptor is expressed in most cells and is basically the same in all tissues, but the number of receptors varies greatly between tissues. The greatest density of receptors occurs in the adrenal cortex, which requires enormous amounts of cholesterol for steroid hormone synthesis. It is expressed at low levels in nondividing cells, presumably because there is less need for cholesterol for membrane synthesis. Cholesterol that is not immediately used in the cell is esterified by

Table 2 Classification and composition of major human plasma lipoproteins

Class	*Density ($g\,ml^{-1}$)*	*Percentage protein*	*Percentage cholesterol*	*Percentage cholesterol ester*	*Percentage triacylglycerol*
Chylomicrons	< 1.00	2	1	3	85
Very-low-density lipoprotein	0.95–1.006	10	7	12	50
Low-density lipoprotein	1.006–1.063	23	8	37	10
High-density lipoprotein	1.063–1.210	55	2	15	4

The other major constituents of these lipoproteins are phospholipids; minor components are sphingolipids, carotenoids, and other hydrophobic compounds.

Table 3 Major human apolipoproteins

Apoprotein	*Associated lipoprotein*	*Function*
ApoA-1	HDL	LCAT activation; possibly involved with reverse cholesterol transport
ApoA-II	HDL	Transport of HDL lipids; unknown
ApoA-IV	HDL, chylo	Unknown; possibly antiatherogenic
ApoB-48	Chylo	Lipid transport
ApoB-100	VLDL, LDL	Lipid transport; LDL receptor interaction
ApoC-1	VLDL, HDL	Unknown; may function with A-I to activate LCAT
ApoC-II	VLDL, HDL, chylo	Activatation of lipoprotein lipase
ApoC-1II	VLDL, HDL, chylo	Inhibition of lipoprotein lipase
ApoD	HDL	Unknown
ApoE	VLDL, HDL, Chylo	Clearance of VLDL and chylomicrons from plasma; abundant in brain and cerebrospinal fluid, modifies risk of Alzheimer dementia

Apo, apoenzyme; chylo, chylomicron; HDL, high-density lipoprotein; LDL, low-density lipoprotein; VLDL, very-low-density lipoprotein.

ACAT, but normally, there are only small amounts of the ester within cells. Among the other factors thought to regulate LDL receptor expression are thyroid hormone, insulin, and platelet-derived and other types of growth factors.

Once LDL is delivered to cells, no degradation of cholesterol can occur in peripheral cells, and removal of cholesterol appears to be a function of another lipoprotein. HDL seems to function in 'reverse cholesterol transport,' in which it helps remove cholesterol from cells and return it to the liver for excretion in the bile. The exact mechanism and sequence of events are not yet understood. High levels of HDL cholesterol appear to indicate that more reverse cholesterol transport is occurring. This topic is discussed elsewhere in this volume.

Oxidation of LDL

The term oxidized LDL (oxLDL) encompasses a variety of particles, ranging from slight oxidation of LDL fatty acids to particles with oxidation of fatty acids, cholesterol and even apoB. LDL that has undergone certain modifications such as peroxidation or chemical acetylation or maleylation of apoB is taken up avidly by the modified-LDL receptor, also called the scavenger receptor. The expression of this receptor is independent of cell [cholesterol], and there is no downregulation of receptor number. Cholesterol accumulation can occur in macrophages and other cells that have this receptor. In addition, fatty acid oxidation fragments such as malondialdehyde may react with lysine residues of apoB, which may increase its ability to accumulate in macrophages. OxLDL is not a good ligand for the native LDL receptor, but it is more atherogenic than native LDL. After the discovery of the LDL receptor by Mark Brown and Joseph Goldstein in the mid-1970s, it gradually became evident that cholesterol accumulation in arterial lesions involves some processes other than uptake of native LDL by its receptors, since subjects with homozygous familial hypercholesterolemia, whose cells lack the receptors, accumulate cholesterol in 'foam' cells. Furthermore, 'foam' cells develop from macrophages/monocytes and smooth muscle cells, but *in vitro* experiments show that neither of these cell types accumulates much cholesterol from native LDL. Also, when native LDL delivers cholesterol and other lipids to cells via receptor-mediated endocytosis, feedback mechanisms decrease the number of receptors on the cell surface. This is called downregulation of the receptor, a process that normally helps maintain cholesterol homeostasis. Further experimentation has shown that the macrophages take up chemically modified LDL by a separate, nonregulated mechanism that allows accumulation and cellular damage (Table 4).

Table 4 Detrimental effects attributed to oxidized LDL

Avidly binds scavenger receptors, leading to accumulation of LDL lipids in arterial macrophages
Cytotoxic to arterial endothelial cells *in vitro* and possibly *in vivo*
Stimulates release of mitogenic factor(s) for macrophages and smooth muscle cells
Stimulates recruitment of additional monocytes/macrophages from bloodstream into intima
Is immunogenic and leads to autoantibody production
Antagonizes normal nitric oxide-induced vasodilation of arteries

It was soon learned that *in vitro* incubation of LDL with cultured endothelial cells creates oxidatively modified LDL that is readily bound and accumulated by the macrophages. Addition of antioxidants *in vitro* prevents this modification of LDL. We now know that the uptake of oxidized LDL occurs via a distinct receptor, called scavenger receptor A (SRA), on macrophages/monocytes and smooth muscle cells. Because this receptor is not downregulated, and LDL lipids accumulate in these cells, eventually they become 'foam cells.' Even slightly oxidized LDL, often called 'minimally modified,' can be recognized and taken up by the SRA. More than one type of scavenger receptor is present on macrophages, and other types of modification of LDL may also contribute to increased uptake by the macrophage. The sequence of events thought to be involved in atherogenesis is discussed below. Studies with experimental animals and *in vitro* cell-culture studies suggest that the use of antioxidants may help reduce or inhibit production of oxLDL and therefore may help inhibit atherogenesis. However, although epidemiologic data offer evidence of an inverse relationship between dietary antioxidant intake and risk of coronary heart disease (CHD), effectiveness in preventing atherosclerosis has not yet been conclusively demonstrated in humans.

Atherogenesis

Atherosclerosis (Greek athērē 'gruel' + sklērōsis 'hardness') occurs within major arteries when cholesterol-laden lesions build up in the lining and wall, eventually resulting in narrowing and sometimes occlusion of the lumen. In addition to decreasing the flow through the narrowed vessel, the lesions, called fibrous plaques, can cause thrombosis (blood clots), which may lead to CHD (also called myocardial infarction (MI) or heart attack), decreased oxygen delivery to the brain or other organs, or thrombotic

stroke. Atherosclerosis is usually described as a multifactorial disease with numerous risk factors (Table 5). Although genetics seems to be a predominant risk factor, attention has focused on those factors that can be changed, namely hypertension and blood cholesterol levels. Risk factors appear to be multiplicative rather than merely additive. Risk varies between races, and in America, blacks have a higher incidence of CHD than whites of the same age. Blood cholesterol levels do not differ between the two groups, and the higher incidence of CHD is attributed to the greater incidence of hypertension among blacks.

Theories on the etiology of atherosclerosis have gradually coalesced into the prevailing 'endothelial injury' hypothesis, which encompasses most *in vivo* and *in vitro* observations (Table 6). The endothelium lining of arteries plays a major role in the development of atherosclerosis. The response to endothelial injury is similar to the inflammatory response to injuries elsewhere. This may occur at the bifurcation of blood vessels where flow is turbulent; it may be in response to the constant insult of hypertension or to chemicals such as benzo(a)pyrene in cigarette smoke. White blood cells invade the area. If high levels of LDL are present, oxidation of LDL, possibly by endothelial cells, results in the formation of a modified (oxidized) LDL. OxLDL triggers a series of events: it causes more endothelial injury, and endothelial cells synthesize endothelial-leukocyte adhesion molecules (ELAMs). The latter attract monocytes and T lymphocytes that adhere to the endothelium. These move below the endothelium and not only engulf more oxLDL but also secrete cytokines and growth factors that stimulate movement of smooth muscle cells and macrophages to the area where all cells begin to accumulate increasing amounts of lipid, particularly cholesterol esters. Over a number of years, the process results in pronounced lesions (Table 7).

Table 5 Primary and secondary risk factors for coronary heart disease

Primary risk factors for CHD
Serum [cholesterol] $> 5.2\,\text{mmol}\,l^{-1}$
[HDL cholesterol] $< 0.9\,\text{mmol}\,l^{-1}$
Hypertriglyceridemia
Heredity
Hypertension
Smoking
Age (risk increases at 45 + years age for men, 55 + for women)
Diabetes mellitus
Obesity
Physical inactivity
Stress
Secondary risk factors for hyperlipidemia
Nephrotic syndrome
Alcoholism
Various drugs
Certain hormonal disorders (e.g., Cushing's disease)
Systemic lupus erythematosis
Pregnancy[a]

[a]A normal physiologic response to pregnancy.

Table 6 Likely sequence of events and relationship of LDL and atherogenesis

Dietary, genetic, and environmental factors
↓
Increased blood levels of LDL lipids (cholesterol, triglycerides)
↓
Greater adherence of monocytes to endothelial lining of arteries and more LDL entering lining of arteries
↓
Oxidative modification of LDL to MM-LDL by cells in the intima (lining)
↓
Stimulation of release of MCP-1 from endothelial cells by oxLDL; and simultaneous stimulation of release of MCSF
↓
Movement of more monocytes into the area from the bloodstream; differentiation of these cells into macrophages with increased SRAs
↓
Increased uptake of modified LDL by SRA and accumulation of cholesterol esters
↓
Recruitment of smooth muscle cells from arterial media to intima; induction of SRA expression in these cells
↓
Formation of lipid-laden foam cells from macrophages and smooth muscle cells
↓
Formation of fatty streaks
↓
Continuation of above sequence, leading eventually to localized cell necrosis, atheroma, and finally plaque formation

MCP-1, monocyte chemoattractant protein-1; MCSF, macrophage colony-stimulating factor; MM-LDL, minimally modified low-density lipoprotein; oxLDL, oxidized LDL; SRA, scavenger receptor-1.

Studies in molecular genetics are proving helpful in establishing the relationships between lipoproteins, lipids, apolipoproteins, membrane proteins, and numerous other factors involved in cholesterol metabolism and atherogenesis. Also, the use of transgenic technology has provided animal models for studying atherogenesis. There now exist transgenic animals (mice or rabbits) for overexpression of most of the apolipoproteins, as well as LCAT, cholesterol ester transfer protein, and scavenger receptor B-1 protein. In addition, 'knockout' animal models, in which certain apolipoprotein genes are missing, are being studied to help determine the functions of the proteins.

Intervention Trials and Current Status

While epidemiologic studies long ago suggested the relationship between high blood cholesterol levels and increased incidence of atherosclerosis and CHD, much of the progress in understanding the etiology, pathology, and relationships has been made in the last two to three decades. The positive correlation between blood cholesterol levels and risk of CHD-related deaths was established in the 1970–80s by the Framingham Heart Study and the Lipid Research Clinics Cornary Primary Prevention Trial. In the mid- to-late 1980s, health agencies in many countries published ranges of 'desirable' vs. 'undesirable' cholesterol levels, along with recommendations on modifying high levels of total and LDL-cholesterol levels (Table 8). But whether deliberately lowering cholesterol levels could decrease the risk of CHD-related events remained questionable until recent years. Several intervention studies have now provided good evidence that reducing plasma cholesterol levels in patients with established CHD with drugs and/or diet does in fact reduce the incidence of both fatal and nonfatal heart attacks. The efficacy of early studies with some of the statin drugs appeared questionable because, although the incidence of CHD deaths declined, overall mortality did not. Newer studies, however, have shown no increase in noncardiovascular disease deaths. Table 9 summarizes some of these studies.

Current thinking is that there should be aggressive cholesterol management in most patients with CHD, irrespective of the etiology of the condition. Very-low-fat diets diets help to reduce total plasma cholesterol levels in some hypercholesterolemic individuals, but recent studies have shown that HDL cholesterol is lowered by such diets, particularly in women. This is of concern and deserves further work. Cholesterol levels increase in older people, but the correlation of high [blood cholesterol] and CHD appears to be less prominent in the elderly population. Nevertheless, it has been reported recently that high total cholesterol levels do increase CHD risk in the elderly population. In the USA, current guidelines, supported by the American Heart Association and the American College of Cardiology, and based on experimental evidence and clinical trials, suggest that LDL cholesterol should be reduced to $\leq 100\,\mathrm{mg\,dl^{-1}}$. Since LDL cholesterol is a calculated value and since the measurement of total and HDL cholesterol may be subject to some error, the consensus opinion is that two or three fasting lipoprotein measurements on consecutive days should be made for an accurate baseline. All patients with CHD and LDL cholesterol $> 100\,\mathrm{mg\,dl^{-1}}$ should immediately reduce the daily intake of saturated fats to $\leq 7\%$ of total calories and of cholesterol to < 200 mg per day. In addition, weight loss is advised for those who are overweight, and appropriate physical activity is suggested. If baseline LDL cholesterol in those with CHD is $> 130\,\mathrm{mg\,dl^{-1}}$, drug therapy in addition to the diet is normally advised immediately.

Table 7 Stages in atherogenesis

Stage	*Age*	*Characteristics*
One	Childhood	Development of 'fatty streaks' in the aorta
Two	Early and Mid-adulthood	Fatty deposition under intima; migration of smooth muscle cells from arterial media to intima; macrophages and smooth muscle cells become lipid-laden 'foam cells'
Three	Middle age and seniors	Formation of frank plaques; calcification of plaques

Table 8 Classification of cholesterol levels

Type	*Desirable*	*Borderline high*	*High risk*
Total serum cholesterol	< 5.2 mM ($< 200\,\mathrm{mg\,dl^{-1}}$)	5.2–6.2 mM (200–239 $\mathrm{mg\,dl^{-1}}$)	> 6.2 mM ($> 240\,\mathrm{mg\,dl^{-1}}$)
Low-density lipoprotein cholesterol	< 3.35 mM ($< 130\,\mathrm{mg\,dl^{-1}}$)	3.35–4.1 mM (130–159 $\mathrm{mg\,dl^{-1}}$)	> 4.15 mM ($> 160\,\mathrm{mg\,dl^{-1}}$)
High-density lipoprotein cholesterol	> 0.9 mM ($> 35\,\mathrm{mg\,dl^{-1}}$)		< 0.9 mM ($< 35\,\mathrm{mg\,dl^{-1}}$

Based on recommendations of the USA National Cholesterol Education Program.

Table 9 Summary of recent intervention trials

Name of study	*Subjects, characteristics*	*Treatment*	*Outcome*
WOSCOPS (1995)	6500 males, [Ch] = 249–295 mg dl^{-1}, no CHD	Low fat diet plus statin drug or placebo	20% ↓ [total Ch] 31% ↓ CHD risk
AFCAPS (1998)	Healthy men and women with average [Ch] ~ 221 mg dl^{-1}	Low-fat diet plus statin drug or placebo	16% ↓ [total Ch] 40% ↓ MI risk
SSSS (1994)	Hypercholesterolemic patients with diagnosed CHD	Diet plus statin drug or placebo	24% ↓ in MI
CLAS (1991)	Middle-aged men with atherosclerosis (4 years)	Colestipol + niacin + diet vs. placebo + diet	16.2% of subjects had lesion regression compared to 2.4% of placebo group

AFCAPS, Airforce Coronary Atherosclerosis Prevention Study; Ch, cholesterol; CHD, coronary heart disease; CLAS, Cholesterol-lowering Atherosclerosis Study; MI, myocardial infarction; SSSS, 4S, Scandinavian Simvastatin Survival Study; WOSCOPS, West of Scotland Coronary Prevention Study.

The choice of drugs depends largely on the lipoprotein patterns (e.g., concomitant high [triglyceride], low [HDL-C], and baseline [LDL-C]). Drugs most often used individually or in combination are statins, bile acid sequestrants, and nicotinic acid. Sequestrants increase the excretion of bile acids in feces and reduce blood cholesterol levels by conversion of more cholesterol to bile acids. Statins inhibit HMG-CoA reductase and thereby decrease the synthesis of cholesterol and other products of the isoprenoid pathway. Nicotinic acid appears to decrease the synthesis of LDL and increase the synthesis of HDL. In patients with a [LDL-C] of 100–129 mg dl^{-1} and low [HDL-C] (< 35 mg dl^{-1}), nicotinic acid is probably the first-choice drug. The efficiency, cost, and side-effects of these drugs are usually important factors in decisions about therapy. Each of these groups of drugs may have undesirable side-effects in some patients, and compliance is often a problem, but recent advances in drug therapy have resulted in safer, but effective, pharmocologic reductions in blood cholesterol levels. Reports from many of the recent intervention trials have indicated that in addition to the decreases in risk, number of CHD deaths, and overall mortality, there was also a reduction in the need for bypass surgery or angioplasty. Furthermore, the interventions have proven effective in reducing high blood cholesterol levels in both genders, in middle-aged and elderly, and in those with or without established CHD.

See also: **Antioxidants**: Natural Antioxidants; **Atherosclerosis**; **Coronary Heart Disease**: Etiology and Risk Factor; Antioxidant Status; Intervention Studies; Prevention; **Fats**: Digestion, Absorption, and Transport; **Hypertension**: Physiology; Hypertension and Diet; **Lipoproteins**

Further Reading

American Heart Association Scientific Statement; AHA Dietary Guidelines, Revision (2000) A statement for healthcare professionals from the Nutrition Committee of the American Heart Association. *Circulation* 102: 2284.

American Heart Association Task Force on Risk Reduction (1997) When to start cholesterol-lowering therapy in patients with coronary heart disease. *Circulation* 95: 1683–1685.

Blake GH and Triplett LC (1995) Management of hypercholesterolemia. *American Family Physician* 51(5): 1157–1166.

Haber E (ed.) (1995) *Molecular Cardiovascular Disease*. New York: Scientific American Medicine.

Jacobson TA (2000) 'The lower the better' in hypercholesterolemia therapy: a reliable clinical guideline? *Annals of Internal Medicine* 133: 549–554.

Rye KA, Clay MA and Barter PJ (1999) Remodelling of high density lipoproteins by plasma factors. *Atherosclerosis* 145: 227–238.

Silver DL, Jiang XC, Arai T, Bruce C and Tall A (2000) Receptors and lipid transfer proteins in HDL metabolism. *Annals of New York Academy of Science* 902: 103–112.

Steinberg D (1997) Low density lipoprotein oxidation and its pathobiological significance. *Journal of Biological Chemistry* 272: 20963–20966.

CHOLINE

Contents
Properties and Determination
Physiology

Properties and Determination

S H Zeisel, University of North Carolina, Chapel Hill, NC, USA

Background

Choline is a common constituent of foods and is an essential component of the human diet. This quaternary amine is important for the structural integrity and signaling functions of cell membranes; it is the major source of methyl groups in the diet; it directly affects cholinergic neurotransmission; and it is required for lipid transport/metabolism. Most choline in human tissues is found in phospholipids such as phosphatidylcholine and sphingomyelin. In some components, such as human milk, the choline esters phosphocholine and glycerophosphocholine are the predominant forms of choline. Though representing a smaller proportion of the total choline pool, important metabolites of choline include platelet-activating factor, acetylcholine, choline plasmalogens, lysophosphatidylcholine, and betaine (Figure 1).

Measurement of Choline

Careful collection of biological samples to be analyzed for choline is exceedingly important. In blood samples, choline concentrations double every 15 min in the sample at room temperature, probably because membrane phosphatidylcholine is hydrolyzed. Therefore, it is important that blood be kept on ice until plasma is separated by centrifugation. For similar reasons, tissues and foods should be stored at −70 °C. At this temperature, samples are stable for at least one year. Acetylcholine is readily hydrolyzed by basic conditions, or by cholinesterases present in tissues. If acetylcholine is to be determined, rapid deactivation of tissue enzymes is essential, and microwave irradiation is commonly used. When acetylcholine is not an important analyte (e.g., for tissues other than brain), it is sufficient to freeze-clamp the tissue using tongs cooled in liquid nitrogen.

There are multiple forms of choline in biological tissues, as discussed above. Assay procedures are available for the most common forms. For all the assays, it is important to realize that purified choline salts are very hygroscopic, and this makes assessing the true weight of internal standards difficult. Unesterified choline can be measured in plasma and red blood cells using a radioisotopic procedure, but this difficult assay has largely been supplanted by a procedure that uses high-pressure liquid chromatography (HPLC), an enzyme reactor, and electrochemical detection after a simple sample pretreatment. An alternative method using HPLC and fluorometric detection has also been used. These methods are accurate, reliable, and inexpensive, but are only suitable for measurement of the choline (unesterified) and acetylcholine, and they do not permit the use of internal standards. A gas chromatography/mass spectrometry method requires much more expensive equipment, but has the advantage that it has been adapted to measure other choline esters. The original method has been enhanced by preseparation of the various aqueous soluble choline metabolites using HPLC. This procedure allows analysis of choline, acetylcholine, phosphocholine, and glycerophosphocholine in a single sample at minimum detection levels of 5–10 pmol applied to the GC/MS (e.g., the amount of choline present in 1 µl of plasma, but owing to the complex extraction and derivatization procedures, 25 µl of plasma is the minimum amount that can be analyzed). It is a demanding procedure and requires that the separate HPLC peaks be collected, the metabolite in question be converted to choline, and then this choline be injected on to the gas chromatogram/mass spectrometer. It does allow the inclusion of an isotopically labeled internal standard for each metabolite, thereby insuring a greater accuracy and reliability when widely divergent matrices are to be analyzed (e.g., tissues and foods).

Neither the electrochemical nor the mass spectrometric assays detect betaine. Betaine can be isolated using HPLC and derivatized with 4′-bromo-phenacyl triflate so that it can be detected and quantified using UV absorbance. Phosphatidylcholine and sphingomyelin can be isolated by thin-layer chromatography

Compound	*Chemical structure*	*Biologic function*
Choline	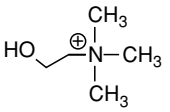	
Betaine	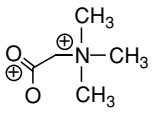	Methyl group donor; renal osmolyte
Acetylcholine	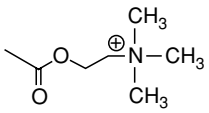	Neurotransmitter
Phosphatidylcholine	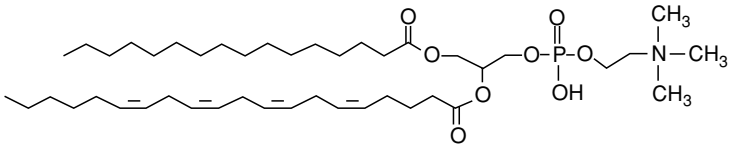	Necessary building block of biomembranes; very-low–density lipoprotein component necessary for hepatic very-low–density lipoprotein secretion
Platelet-activating factor	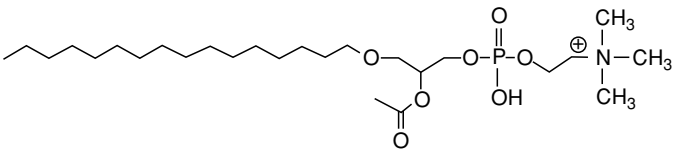	Hormone
Sphingomyelin	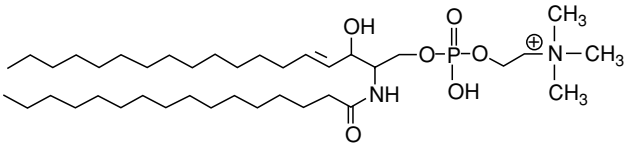	Necessary building block of biomembranes
Glycerophosphocholine	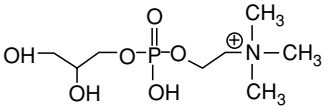	Intracellular storage pool of choline
Phosphocholine	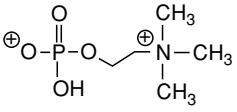	Intracellular storage pool of choline

Figure 1 Biologically important metabolites of choline.

and quantified using choline or phosphorus determination, or isolated by HPLC and quantified with ultraviolet or fluorometric detection. The advantage of these last methods is the ability to differentiate phospholipids by fatty acid composition.

A method has recently been devised, using liquid chromatography/electrospray ionization/mass spectrometry (LC/ESI/MS), which greatly reduces the number of steps in the analysis (eliminates the transfer to the gas chromatograph), but the equipment costs still remain very high. This method requires no derivatization and allows the use of internal standards. After sample extraction, the aqueous and organic phases are separated and analyzed separately

using straight-phase liquid chromatography. This method requires 2 pmol of choline to be applied to the LC/MS (e.g., the amount of choline in 0.2 μl of plasma).

Dietary Sources of Choline

Many foods eaten by humans contain significant amounts of choline and esters of choline, but there is currently no comprehensive database of the choline content of foods. Some of this choline is added during food processing (especially when preparing infant formula). The average dietary intake of choline and choline esters in the adult human is estimated to be 7–10 mmol per day. When humans are switched from a diet of normal foods to a defined diet containing 5 mmol per day, plasma choline and phosphatidylcholine concentrations decrease in most subjects. Thus, the average dietary intake of choline seems to exceed this level in adults. Human, commercially available infant formulas and bovine and rat milk contain approximately 1–2 mmol per liter of choline and choline esters. Human milk has a significantly higher phosphocholine concentration, the same or lower glycerophosphocholine concentration, and similar phosphatidylcholine and sphingomyelin concentrations compared with either bovine milk or bovine-derived infant formulas. Soy-derived infant formulas have lower glycerophosphocholine and sphingomyelin concentrations and higher phosphatidylcholine concentrations than do either human milk or bovine milk-derived formulas. By assuming that a newborn infant drinks 150 ml of milk per kilogram per day, their choline intake would be approximately 200–250 μmol per kilogram per day, two to three times that ingested by the adult human.

Intestinal Absorption

The extent to which dietary choline is bioavailable depends upon the efficiency of its absorption from the intestine. In adults, some ingested choline is metabolized before it can be absorbed from the gut. Gut bacteria degrade it to form betaine and to make methylamines. The free choline surviving these fates is absorbed all along the small intestine. No other component of the diet has been identified as competing with choline for transport by intestinal carriers. Both pancreatic secretions and intestinal mucosal cells contain enzymes (phospholipases A_1, A_2, and B) capable of hydrolyzing phosphatidylcholine in the diet. The free choline that is formed enters the portal circulation of the liver.

In infants, there are differences in the bioavailability of the water soluble, choline-derived compounds (choline, phosphocholine, and glycerophosphocholine) and the lipid-soluble compounds (phosphatidylcholine and sphingomyelin) present in milk.

See also: **Phospholipids**: Properties and Occurrence

Further Reading

Ceder G and Schuberth J (1977) *In vivo* formation and post-mortem changes of choline and acetylcholine in the brain of mice. *Brain Research* 128: 580–584.

Cheng W-L, Holmes-McNary MQ, Mar M-H, Lien EL and Zeisel SH (1996) Bioavailability of choline and choline esters from milk in rat pups. *Journal of Nutritional Biochemistry* 7: 457–464.

Damsma G and Flentge F (1988) Liquid chromatography with electrochemical detection for the determination of choline and acetylcholine in plasma and red blood cells. Failure to detect acetylcholine in blood of humans and mice. *Journal of Chromatography* 428: 1–8.

FASEB Life Sciences Research Office (1981) *Effects of Consumption of Choline and Lecithin on Neurological and Cardiovascular Systems*, *Report # PB-82-133257*. Washington, DC: Bureau of Foods, Food and Drug Administration, Department of Health, Education, and Welfare.

Fox JM, Betzing H and Lekim D (1979) Pharmacokinetics of orally ingested phosphatidylcholine. In: Barbeau A, Growdon JH and Wurtman RJ (eds) *Nutrition and the Brain*, vol. 5, pp. 95–108. New York: Raven Press.

Goldberg AM and McCaman RE (1973) The determination of picomole amounts of acetylcholine in mammalian brain. *Journal of Neurochemistry* 20: 1–8.

Holmes-McNary M, Cheng WL, Mar MH, Fussell S and Zeisel SH (1996) Choline and choline esters in human and rat milk and infant formulas. *American Journal of Clinical Nutrition* 64: 572–576.

Ikarashi Y, Sasahara T and Maruyama Y (1985) Determination of choline and acetylcholine levels in rat brain regions by liquid chromatography with electrochemical detection. *Journal of Chromatography* 322: 191–199.

Institute of Medicine, and National Academy of Sciences USA (1998) *Dietary Reference Intakes for Folate, Thiamin, Riboflavin, Niacin, Vitamin B_{12}, Panthothenic Acid, Biotin, and Choline*, vol. 1. Washington DC: National Academy Press.

Jenden DJ, Roch M and Booth RA (1973) Simultaneous measurement of endogenous and deuterium-labeled tracer variants of choline and acetylcholine in subpicomole quantities by gas chromatography–mass spectrometry. *Analytical Biochemistry* 55: 438–448.

Koc H, Mar M-H, Swenberg JA and Zeisel S (2001) Quantitation of choline and its metabolites in various tissues and food items by liquid chromatography/electrospray ionization/isotope dilution mass spectrometry. *Analytical Chemistry*, in press.

Mar MH, Ridky TW, Garner SC and Zeisel SH (1995) A method for the determination of betaine in tissues using high performance liquid chromatography. *Journal of Nutritional Biochemistry* 6: 392–398.

Ou Z, Ogamo A, Guo L, Konda Y, Harigaya Y and Nakagawa Y (1995) Identification and quantitation of choline glycerophospholipids that contain aldehyde residues by fluorometric high-performance liquid chromatography. *Analytical Biochemistry* 227: 289–294.

Pomfret EA, daCosta K, Schurman LL and Zeisel SH (1989) Measurement of choline and choline metabolite concentrations using high-pressure liquid chromatography and gas chromatography–mass spectrometry. *Analytical Biochemistry* 180: 85–90.

Postle A, Al M, Burdge G and Hornstra G (1995) The composition of individual molecular species of plasma phosphatidylcholine in human pregnancy. *Early Human Development* 43: 47–58.

Riccny J, Tucek S and Vins I (1992) Sensitive method for HPLC determination of acetylcholine, choline and their analogues using fluorometric detection. *Journal of Neuroscience Methods* 41: 11–17.

Sheard NF and Zeisel SH (1986) An *in vitro* study of choline uptake by intestine from neonatal and adult rats. *Pediatric Research* 20: 768–772.

Zeisel SH, daCosta K-A, Franklin PD *et al.* (1991) Choline, an essential nutrient for humans. *FASEB Journal* 5: 2093–2098.

Zeisel SH, daCosta KA, Youssef M and Hensey S (1989) Conversion of dietary choline to trimethylamine and dimethylamine in rats: dose–response relationship. *Journal of Nutrition* 119: 800–804.

Zeisel SH, Story DL, Wurtman RJ and Brunengraber H (1980) Uptake of free choline by isolated perfused rat liver. *Proceedings of the National Academy of Sciences, of the USA* 77: 4417–4419.

Physiology

S H Zeisel, University of North Carolina, Chapel Hill, NC, USA

Metabolism of Choline

There have been several comprehensive reviews of the metabolism and functions of choline that describe its role in the synthesis of the phospholipids in cell membranes, methyl metabolism, cholinergic neurotransmission, transmembrane signaling, and lipid-cholesterol transport and metabolism. Choline can be acetylated, phosphorylated, or oxidized.

Choline, methionine, and folate metabolism interact at the point that homocysteine is converted to methionine (**Figure 1**). Betaine:homocysteine methyltransferase catalyzes the methylation of homocysteine using the choline metabolite betaine as the methyl donor. Elevated plasma homocysteine is an independent risk factor for cardiovascular disease and stroke in humans. Treatment with betaine or choline also lowers elevated plasma homocysteine in humans. In an alternative pathway, 5-methyl-tetrahydrofolate:homocysteine methyltransferase regenerates methionine. In addition, tetrahydrofolate is needed to scavenge one-carbon groups when betaine is metabolized. Perturbing metabolism of one of the methyl donors results in compensatory changes in the other methyl donors owing to the intermingling of these metabolic pathways. Rats ingesting a choline-deficient diet have diminished tissue concentrations of methionine and *S*-adenosylmethionine, doubled plasma homocysteine concentrations, and diminished hepatic methyl-folate content. These effects are reversible by refeeding choline. Rats fed diets deficient in both methionine and choline for 5 weeks have hepatic folate concentrations that are half of those present in controls. Rats treated with the antifolate, methotrexate, have diminished pools of choline metabolites in liver. Severe folate deficiency, induced in rats by feeding an amino acid-defined diet containing no folate and succinylsulfathiazole for 4 weeks have resulted in hepatic choline and phosphocholine concentrations that were 65 and 80% lower, respectively than in controls.

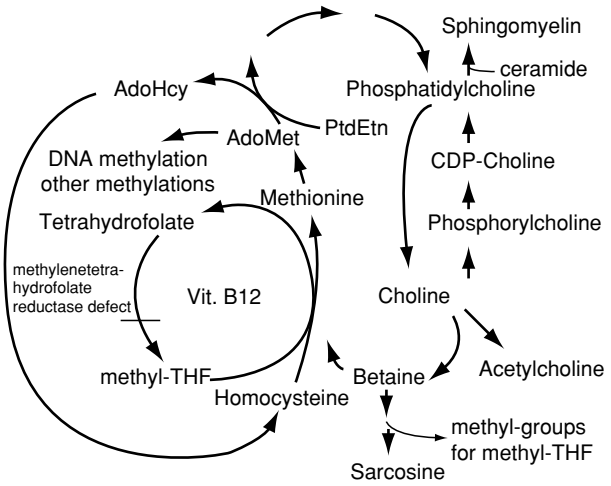

Figure 1 Choline, folate, and homocysteine metabolism are closely inter-related. The pathways for the metabolism of these three nutrients intersect at the formation of methionine from homocysteine. PtdEtn, phosphatidylethanolamine; AdoHcy, *S*-adenosylhomocysteine; AdoMet, *S*-adenosylmethionine

Dietary Requirements

Though there is a pathway (in all tissues, but most active in the liver) for the *de novo* biosynthesis of the choline moiety via the sequential methylation of phosphatidylethanolamine using *S*-adenosylmethionine

as the methyl donor (Figure 1), only some of the demand for choline can be met by using methyl groups derived from one carbon metabolism. Animals and humans fed a choline deficient diet deplete choline stores and develop liver dysfunction. Healthy male humans with normal folate and vitamin B_{12} status fed a choline-deficient diet have diminished plasma choline and phosphatidylcholine concentrations, and develop liver damage (elevated plasma alanine aminotransferase). Liver cell death occurs in choline deficiency, because hepatocytes initiate programmed cell death (apoptosis) when deprived of choline. Because methyl supplementation with betaine, methionine, folate, or vitamin B_{12} does not prevent apoptotic death induced by choline deficiency in hepatocytes, it must be that depletion of intracellular choline moieties rather than depletion of methyl groups is the critical parameter involved in induction of apoptosis. Some humans (male and female) fed with total parenteral nutrition solutions devoid of choline, but adequate for methionine and folate, develop fatty liver and liver damage that resolves when a source of dietary choline is provided. Fatty liver occurs because choline is required to make the phosphatidylcholine portion of the very-low-density lipoprotein particle. Animals fed a choline-deficient diet may also develop growth retardation, renal dysfunction and hemorrhage, or bone abnormalities. A metabolite of choline, betaine, is used in the kidney glomerulus as an osmolyte, and for this reason, choline-deficient animals have trouble excreting concentrated urine.

Human studies of choline requirements in women, children, or infants have not been conducted. Thus, we do not know whether choline is needed in the diet of these groups. Female rats are less sensitive to choline deficiency than are male rats perhaps because estrogen enhances females' capacity to form the choline moiety *de novo* from *S*-adenosylmethionine. Pregnancy may be a time when dietary supplies of choline are especially limiting. Though female rats are resistant to choline deficiency, pregnant rats are as vulnerable to deficiency, as are males. During pregnancy, large amounts of choline are delivered to the fetus across the placenta, and this depletes maternal stores of the various forms of this nutrient. At birth, humans and other mammals have plasma choline concentrations that are much higher than those in adults. Also, the need for choline is increased during lactation, because so much must be secreted into milk; lactating rats are more sensitive to choline deficiency than are nonlactating rats.

The Institute of Medicine (IOM) recently made recommendations for choline intake in the diet. There were insufficient data with which to derive an estimated average requirement for choline, and so only an adequate intake can be estimated. The IOM report cautioned, 'this amount will be influenced by the availability of methionine and methyl-folate in the diet. It may be influenced by gender, and it may be influenced by pregnancy, lactation, and stage of development. Although AIs are set for choline, it may be that the choline requirement can be met by endogenous synthesis at some of these stages.' The IOM recommendations are shown in Table 1.

Table 1 Institute of Medicine recommendations for choline intake in the diet

AI for infants	0–6 months	125 mg per day, 18 mg per kilogram
	6–12 months	150 mg per day
AI for children	1–3 years	200 mg per day
	4–8 years	250 mg per day
	9–13 years	375 mg per day
AI for males	14–18 years	550 mg per day
	19 and older	550 mg per day
AI for females	14–18 years	400 mg per day
	19 years and older	425 mg per day
AI for pregnancy	All ages	450 mg per day
AI for lactation	All ages	550 mg per day

AI = adequate intake level.

The recent report by Jacob that folate deficiency in humans exacerbates choline deficiency highlights why studies of choline–folate–homocysteine interrelationships in humans are important for the future refinement of these recommendations.

Plasma choline concentration varies in response to diet and can rise as much as twofold after a two-egg meal. Fasting plasma choline concentrations vary from 7 to 20 μM, with most subjects having concentrations of 10 μM. Individuals that have starved for up to 7 days have diminished plasma choline, but levels never drop below 50% of normal. Plasma phosphatidylcholine concentration also decreases in choline deficiency, but these values are also influenced by factors that change plasma lipoprotein levels. Fasting plasma phosphatidylcholine concentrations are approximately 1–1.5 mM.

Choline and Brain Development

Choline availability during embryogenesis and perinatal development may be especially important. There is a sensitive period in rodent brain development during which treatment with choline (about three times the dietary intake) produces long-lasting enhancement of spatial memory that is lifelong. This period occurs during embryonic days 12–17 in the rat (rats give birth on day 21; mice give birth one day earlier and probably have a slightly different period of susceptibility). Choline supplementation

during these critical days elicits a major improvement in memory performance at all stages of training on a 12-arm radial maze. Choline deficiency during the critical timeframe reduces memory performance. The choline-induced spatial memory facilitation correlates with changes in the birth, death, and migration of cells in the hippocampus during fetal brain development, with altered distribution and morphology of neurons, biochemical changes, and electrophysiological changes in the hippocampus. Human and rat brains mature at different rates; rat brain is comparatively more mature at birth than is the human brain, but in humans, hippocampal development may continue for months or years after birth.

Choline and Cancer

Dietary deficiency of choline in rodents causes development of hepatocarcinomas in the absence of any known carcinogen. Choline is the only single nutrient for which this is true. It is interesting that choline-deficient rats not only have a higher incidence of spontaneous hepatocarcinoma but also are markedly sensitized to the effects of administered carcinogens. Several mechanisms are suggested for the cancer-promoting effect of a choline-devoid diet. A progressive increase in cell proliferation that is related to regeneration after parenchymal cell death occurs in the choline-deficient liver. Cell proliferation and its associated increased rate of DNA synthesis could be the cause of the heightened sensitivity to chemical carcinogens. Methylation of DNA is essential to the regulation of expression of genetic information, and the undermethylation of DNA observed during choline deficiency (despite adequate dietary methionine) may be responsible for carcinogenesis. Choline-deficient rats experience increased lipid peroxidation in liver. Lipid peroxides in the nucleus are a possible source of free radicals that could modify DNA and cause carcinogenesis. Choline deficiency activates protein kinase C signaling, usually involved in growth factor signaling in hepatocytes. Finally, a defect in cell-suicide (apoptosis) mechanisms may contribute to the carcinogenesis of choline deficiency.

Summary

Choline in the diet is important for many reasons. As our understanding of the importance of folate and homocysteine nutrition increases, there should be increased interest in how choline interacts with folate and homocysteine metabolism. Recent findings about choline in brain development should stimulate comparable studies in humans. During the next few years, it is likely that food composition data will be available for choline, and this will make it possible to examine interactions between choline, folate, and methionine when considering epidemiological data. In addition, we should learn more about choline requirements in women. For these reasons interest in choline as a nutrient for humans should be sustained.

See also: **Cancer**: Diet in Cancer Prevention; **Cobalamins**: Properties and Determination; **Phospholipids**: Properties and Occurrence; **Riboflavin**: Properties and Determination; **Vitamin B_6**: Properties and Determination

Further Reading

Albright CD, Friedrich CB, Brown EC, Mar MH and Zeisel SH (1999a) Maternal dietary choline availability alters mitosis, apoptosis and the localization of TOAD-64 protein in the developing fetal rat septum. *Brain Research. Developmental Brain Research* 115: 123–129.

Albright CD, Lui R, Bethea TC *et al.* (1996) Choline deficiency induces apoptosis in SV40-immortalized CWSV-1 rat hepatocytes in culture. *FASEB Journal* 10: 510–516.

Albright CD, Tsai AY, Friedrich CB, Mar MH and Zeisel SH (1999b) Choline availability alters embryonic development of the hippocampus and septum in the rat. *Brain Research. Developmental Brain Research* 113: 13–20.

Anonymous (1997) Betaine for homocystinuria. *Medical Letters on Drugs and Therapeutics* 39: 12.

Banni S, Corongiu FP, Dessi MA *et al.* (1989) Free radicals and lipid peroxidation in liver of rats kept on a diet devoid of choline. *Free Radical Research Communications* 7: 233–240.

Bremer J and Greenberg D (1961) Methyl transfering enzyme system of microsomes in the biosynthesis of lecithin (phosphatidylcholine). *Biochimica et Biophysica Acta* 46: 205–216.

Buchman AL, Dubin M, Jenden D *et al.* (1992) Lecithin increases plasma free choline and decreases hepatic steatosis in long-term total parenteral nutrition patients. *Gastroenterology* 102: 1363–1370.

Buchman A, Dubin M, Moukarzel A *et al.* (1995) Choline deficiency: a cause of hepatic steatosis during parenteral nutrition that can be reversed with intravenous choline supplementation. *Hepatology* 22: 1399–1403.

Buchman AL, Moukarzel A, Jenden DJ *et al.* (1993) Low plasma free choline is prevalent in patients receiving long term parenteral nutrition and is associated with hepatic aminotransferase abnormalities. *Clinical Nutrition* 12: 33–37.

Cermak JM, Blusztajn JK, Meck WH *et al.* (1999) Prenatal availability of choline alters the development of acetylcholinesterase in the rat hippocampus. *Developmental Neuroscience* 21: 94–104.

Chandar N, Amenta J, Kandala JC and Lombardi B (1987) Liver cell turnover in rats fed a choline-devoid diet. *Carcinogenesis* 8: 669–673.

Chawla RK, Wolf DC, Kutner MH and Bonkovsky HL (1989) Choline may be an essential nutrient in malnourished patients with cirrhosis. *Gastroenterology* 97: 1514–1520.

Christman JK (1995) Dietary effects on DNA methylation: do they account for the hepatocarcinogenic properties of lipotrope deficient diets? *Advances in Experimental Medicine and Biology* 369: 141–154.

da Costa K, Cochary EF, Blusztajn JK, Garner SC and Zeisel SH (1993) Accumulation of 1,2-*sn*-diradylglycerol with increased membrane-associated protein kinase C may be the mechanism for spontaneous hepatocarcinogenesis in choline deficient rats. *Journal of Biological Chemistry* 268: 2100–2105.

da Costa K-A, Garner SC, Chang J and Zeisel SH (1995) Effects of prolonged (1 year) choline deficiency and subsequent refeeding of choline on 1,2,-*sn*-diradylglycerol, fatty acids and protein kinase C in rat liver. *Carcinogenesis* 16: 327–334.

Drouva SV, LaPlante E, Leblanc P *et al*. (1986) Estradiol activates methylating enzyme(s) involved in the conversion of phosphatidylethanolamine to phosphatidylcholine in rat pituitary membranes. *Endocrinology* 119: 2611–2622.

Finkelstein JD, Martin JJ, Harris BJ and Kyle WE (1983) Regulation of hepatic betaine-homocysteine methyltransferase by dietary betaine. *Journal of Nutrition* 113: 519–521.

Garcia-Perez A and Burg MB (1991) Role of organic osmolytes in adaptation of renal cells to high osmolality. *Journal of Membrane Biology* 119: 1–13.

Glueck CJ, Shaw P, Lang JE *et al*. (1995) Evidence that homocysteine is an independent risk factor for atherosclerosis in hyperlipidemic patients. *American Journal of Cardiology* 75: 132–136.

Goddijn-Wessel T, Wouters M, van de Molen E *et al*. (1996) Hyperhomocysteinemia: a risk factor for placental abruption or infarction. *European Journal of Obstetrics, Gynecology, and Reproductive Biology* 66: 23–29.

Goshal AK, Ahluwalia M and Farber E (1983) The rapid induction of liver cell death in rats by a choline-deficient methionine-low diet. *American Journal of Pathology* 113: 309–314.

Gwee MC and Sim MK (1978) Free choline concentration and cephalin-*N*-methyltransferase activity in the maternal and foetal liver and placenta of pregnant rats. *Clinical and Experimental Pharmacology and Physiology* 5: 649–653.

Handler P and Bernheim F (1949) Choline deficiency in the hamster. *Proceedings of the Society of Experimental Medicine* 72: 569.

Hershey JM and Soskin S (1931) Substitution of 'lecithin' for raw pancreas in a diet of depancreatized dog. *American Journal of Physiology* 93: 657–658.

Hoffbauer FW and Zaki FG (1965) Choline deficiency in the baboon and rat compared. *Archives of Pathology* 79: 364–369.

Holmes-McNary M, Cheng WL, Mar MH, Fussell S and Zeisel SH (1996) Choline and choline esters in human and rat milk and infant formulas. *American Journal of Clinical Nutrition* 64: 572–576.

Holmes-McNary MQ, Loy R, Mar M-H, Albright CD and Zeisel SH (1997) Apoptosis is induced by choline deficiency in fetal brain and in PC12 cells. *Developmental Brain Research* 101: 9–16.

Horne DW, Cook RJ and Wagner C (1989) Effect of dietary methyl group deficiency on folate metabolism in rats. *Journal of Nutrition* 119: 618–621.

Institute of Medicine and National Academy of Sciences USA (1998) *Dietary Reference Intakes for Folate, Thiamin, Riboflavin, Niacin, Vitamin B_{12}, Panthothenic Acid, Biotin, and Choline*, vol. 1. Washington DC: National Academy Press.

Jacob R, Jenden D, Okoji R, Allman M and Swendseid M (1998) Choline status of men and women is decreased by low dietary folate. *FASEB Journal* 12: A512.

Kang S (1996) Treatment of hyperhomocyst(e)inemia: physiological basis. *Journal of Nutrition* 126: 1273S–1275S.

Kim Y-I, Miller JW, da Costa K-A *et al*. (1995) Folate deficiency causes secondary depletion of choline and phosphocholine in liver. *Journal of Nutrition* 124: 2197–2203.

Kuksis A and Mookerjea S (1978) Choline. *Nutrition Reviews* 36: 201–207.

Lieber CS, Robins SJ, Li J *et al*. (1994) Phosphatidylcholine protects against fibrosis and cirrhosis in the baboon. *Gastroenterology* 106: 152–159.

Lindblad L and Schersten T (1976) Incorporation rate *in vitro* of choline and methyl-methionine into human hepatic lecithins. *Scandinavian Journal of Gastroenterology* 11: 587–591.

Locker J, Reddy TV and Lombardi B (1986) DNA methylation and hepatocarcinogenesis in rats fed a choline devoid diet. *Carcinogenesis* 7: 1309–1312.

Lombardi B and Smith ML (1994) Tumorigenesis, proto-oncogene activation, and other gene abnormalities in methyl deficiency. *Journal of Nutritional Biochemistry* 5: 2–9.

Loy R, Heyer D, Williams CL and Meck WH (1991) Choline-induced spatial memory facilitation correlates with altered distribution and morphology of septal neurons. *Advances in Experimental Medicine and Biology* 295: 373–382.

Lyman RL, Sheehan G and Tinoco J (1971) Diet and $^{14}CH_3$-methionine incorporation into liver phosphatidylcholine fractions of male and female rats. *Canadian Journal of Biochemistry* 49: 71–79.

McCully K (1996) Homocysteine and vascular disease. *Nature Medicine* 2: 386–389.

Meck W and Williams C (1997) Simultaneous temporal processing is sensitive to prenatal choline availability in mature and aged rats. *Neuroreport* 8: 3045–3051.

Meck WH and Williams CL (1999) Choline supplementation during prenatal development reduces proactive interference in spatial memory. *Brain Research. Developmental Brain Research* 118: 51–59.

Meck WH, Smith RA and Williams CL (1988) Pre- and postnatal choline supplementation produces long-term

facilitation of spatial memory. *Developmental Psychobiology* 21: 339–353.

Mudd SH, Ebert MH and Scriver CR (1980) Labile methyl group balances in the human: the role of sarcosine. *Metabolism* 29: 707–720.

Newberne PM and Rogers AE (1986) Labile methyl groups and the promotion of cancer. *Annual Review of Nutrition* 6: 407–432.

Pogribny IP, Basnakian AG, Miller BJ *et al*. (1995) Breaks in genomic DNA and within the p53 gene are associated with hypomethylation in livers of folate/methyl-deficient rats. *Cancer Research* 55: 1894–1901.

Pomfret EA, da Costa K and Zeisel SH (1990) Effects of choline deficiency and methotrexate treatment upon rat liver. *Journal of Nutritional Biochemistry* 1: 533–541.

Pyapali G, Turner D, Williams C, Meck W and Swartzwelder HS (1998) Prenatal choline supplementation decreases the threshold for induction of long-term potentiation in young adult rats. *Journal of Neurophysiology* 79: 1790–1796.

Savendahl L, Mar M-H, Underwood L and Zeisel S (1997) Prolonged fasting results in diminished plasma choline concentration but does not cause liver dysfunction. *American Journal of Clinical Nutrition* 66: 622–625.

Selhub J, Seyoum E, Pomfret EA and Zeisel SH (1991) Effects of choline deficiency and methotrexate treatment upon liver folate content and distribution. *Cancer Research* 51: 16–21.

Shapira G, Chawla RK, Berry CJ *et al*. (1986) Cysteine, tyrosine, choline and carnitine supplementation of patients on total parenteral nutrition. *Nutrition International* 2: 334–339.

Sheard NF, Tayek JA, Bistrian BR, Blackburn GL and Zeisel SH (1986) Plasma choline concentration in humans fed parenterally. *American Journal of Clinical Nutrition* 43: 219–224.

Shin OH, Mar MH, Albright CD *et al*. (1997) Methyl-group donors cannot prevent apoptotic death of rat hepatocytes induced by choline-deficiency. *Journal of Cellular Biochemistry* 64: 196–208.

Tessitore L, Sesca E, Greco M, Pani P and Dianzani M (1995) Sexually differentiated response to choline in choline deficiency and ethionine intoxication. *International Journal of Experimental Pathology* 76: 125–129.

Varela-Moreiras G, Ragel C and Perez de Miguelsanz J (1995) Choline deficiency and methotrexate treatment induces marked but reversible changes in hepatic folate concentrations, serum homocysteine and DNA methylation rates in rats. *Journal of the American College of Nutrition* 14: 480–485.

Varela-Moreiras G, Selhub J, da Costa K and Zeisel SH (1992) Effect of chronic choline deficiency in rats on liver folate content and distribution. *Journal of Nutritional Biochemistry* 3: 519–522.

Welsch F (1976) Studies on accumulation and metabolic fate of (*N*-Me^{3}H)choline in human term placenta fragments. *Biochemical Pharmacology* 25: 1021–1030.

Zeisel SH Albright CD, Shin O-K *et al*. (1997) Choline deficiency selects for resistance to p53-independent apoptosis and causes tumorigenic transformation of rat hepatocytes. *Carcinogenesis* 18: 731–738.

Zeisel SH and Blusztajn JK (1994) Choline and human nutrition. *Annual Review of Nutrition* 14: 269–296.

Zeisel SH, daCosta K-A, Franklin PD *et al*. (1991) Choline, an essential nutrient for humans. *FASEB Journal* 5: 2093–2098.

Zeisel SH, Epstein MF and Wurtman RJ (1980) Elevated choline concentration in neonatal plasma. *Life Sciences* 26: 1827–1831.

Zeisel SH, Growdon JH, Wurtman RJ, Magil SG and Logue M (1980) Normal plasma choline responses to ingested lecithin. *Neurology* 30: 1226–1229.

Zeisel SH, Mar M-H, Zhou Z-W and da Costa K-A (1995) Pregnancy and lactation are associated with diminished concentrations of choline and its metabolites in rat liver. *Journal of Nutrition* 125: 3049–3054.

Zeisel SH, Zola T, daCosta K and Pomfret EA (1989) Effect of choline deficiency on *S*-adenosylmethionine and methionine concentrations in rat liver. *Biochemical Journal* 259: 725–729.

CHROMATOGRAPHY

Contents

Principles

E Forgács and T Cserháti, Chemical Research Center, Institute of Chemistry, Budapest, Hungary

Background

Chromatography is a common name for techniques based on the partition of the molecules to be analyzed between a mobile and a stationary phase. Separation is the result of different partitions of molecules between the two phases. Because the best separation of any solutes can be obtained under equilibrium conditions, analytical chemists prefer to use chromatographic systems that are as near to the equilibium state as possible. However, in the case of preparative chromatography, where the main objective is not the optimal separation of solutes but the maximum yield of one or more solutes at a given purity, the situation is entirely different. Preparation chromatographic separations are generally not equilibrium processes. The high sensitivity, selectivity, and reproducibility of chromatographic methods have been extensively exploited in food and nutrition science and technology.

This chapter overviews the theoretical backgrounds, principles, and fundamental technical characteristics of chromatographic separation processes and instrumentations.

Basic knowledge of the principles of chromatography may help analytical chemists solve any separation problems, using chromatography as a tool.

Gas Chromatography

Principle

Gas chromatography (GC) is the common name for chromatographic methods in which the mobile phase is gas, and the stationary phase is solid or liquid (gas–solid chromatography (GSC) or gas–liquid chromatography (GLC)). There are limitations to the type of compound suitable for GC analysis. A compound needs to have an appreciable vapor pressure at temperatures below 350–400 °C and has to be vaporized rapidly without decomposing or reacting with the components of the stationary and mobile phase (instability at high temperatures usually renders a compound unsuitable for GC analysis).

The common GC equipment consists of a carrier gas system, injector, gas chromatographic column, detector, and data-processing unit (Figure 1). The carrier gas is generally a permanent gas with a low or negligible adsorption capacity, i.e., hydrogen, helium, argon, or nitrogen. The nature of the carrier gas may influence the separation characteristics of the GC system and can modify the sensitivity of the detection.

As the stability and reproducibility of the carrier gas flow rate are the prerequisites for a successful GC analysis, they have a considerable influence on both the efficacy of the separation and the quantification of the results.

Injectors deliver the sample to the head of the GC column and can be classified into two major groups: vaporization and on-column injectors. Vaporization injectors use high temperatures to rapidly vaporize a liquid sample. Usually, a syringe is used to introduce the sample into the thermostated injector. In this case, the sample rapidly vaporizes, mixes with the carrier gas, and is transported into the column. On-column injectors deposit the sample directly into the column without relying on vaporization of the sample and its subsequent transport into the column. Separation of volatile compounds of the injected sample is performed in the GC column.

The columns used for gas chromatography can be divided into two distinct groups: packed and capillary columns. A packed column is a rigid metal or glass column filled with small particles that are often

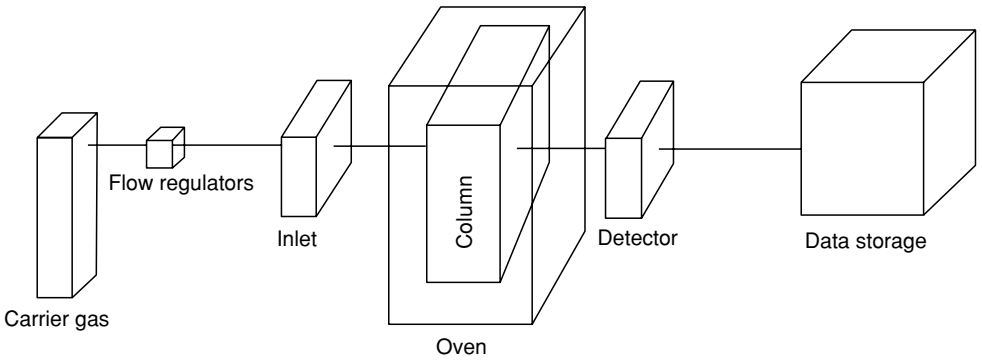

Figure 1 Block diagram of a typical gas chromatogram.

coated with a thin layer of high-molecular-mass polymer. The most common solid supports are diatomaceous earth, fluorocarbons, graphitized carbon black, and glass beads. The stationary liquid phase has to comply with the following requirements: low vapor pressure, high chemical stability, and relatively low viscosity at the temperature of analysis; selectivity for the sample components under investigation; and a good wetting capacity for the surface of the inert support or for the possibly inert wall of the column. A capillary (open tubular) column is a glass or fused-silica tube of very small internal diameter (generally between 0.20 and 0.53 mm). The inner surface of a capillary column is coated only with a thin layer of stationary phase, so it is still possible for the analytes to come into contact with the inner wall of the tubing. Stationary phases can be cross-linked and covalently bonded to the inner wall of the capillary. The stationary phase in a capillary column is characterized by its film thickness (0.1–5 μm). The advantage of capillary columns is the considerably higher separation capacity than that of packed columns. The decrease in lateral diffusion, easier availability of the separation liquid, and increase in column length account for the enhanced separation efficacy of capillary columns. Because of the better separation performance, capillary columns have been used more frequently in gas chromatography than packed columns.

In order to achieve effective and reliable separation, a gas chromatographic column has to be kept at a constant temperature (isothermal separation mode), or the column temperature can be changed according to a temperature program (temperature gradient). As linear temperature programs consisting of more than one step are usually sufficient for the effective separation of analytes, nonlinear temperature programs are rarely used in the analysis of food constituents. Detectors interact with the solute molecules as they exit the column. This interaction is converted into an electrical signal that is sent to a recording or data-storage device. A chromatogram, i.e., a plot of the intensity of the signal versus elapsed time, is then created. The primary differences between the detectors are in the lowest amount of a compound that is detectable (sensitivity) and the compounds that produce the strongest detector response (selectivity). Many different detectors (flame ionization detectors (FID), nitrogen–phosphorus detectors (NPD), flame photometric detectors (FPD), electron capture detectors (ECD), thermal conductivity detectors (TCD), atomic emission detectors (AED), electrolytic conductivity detectors (ELCD), chemiluminescence detectors, mass spectrometry (MS) detectors, etc.) and data-handling devices have been developed for the sensitive and selective detection and quantification of sample components. The FID uses a hydrogen flow mixed with the carrier gas. The mixture is ignited, the analytes are burned, and the ions formed during the burning process are collected in a cylindrical electrode at a high voltage applied between the jet of flame and the electrode. The resulting current is amplified and detected. The NPD is similar to the FID in its design. It contains rubidium or cesium beads inside a heater coil near the hydrogen jet. The partially combusted nitrogen and phosphorus molecules absorb on to the surface of the beads, thereby reducing the emission of electrons, which increases the current. The FPD especially detects sulfur and phosphorus compounds. Analytes are burned in the flame, and because of the excitation in the flame, light is emitted at 392 nm (sulfur) and 526 nm (phosphorus). A filter selects the wavelengths reaching a photomultiplier tube. The ECD employs a low-energy β-ray source for the production of electrons and ions. Electron-capturing molecules (halogenated compounds) entering the detector decrease the electron current, which can be amplified and registered. The TCD responds to the changes in thermal conductivity and specific heat, via a filament under a current applied to the carrier gas flow.

Changes in the thermal conductivity and/or specific heat of the current gas caused by the analytes change the potential across the filament. The AED is suitable for detecting selected atoms or groups of atoms, and the ELCD can be used for detecting Cl-, N-, or S-containing analytes. The chemiluminescence detector is used mainly for detecting sulfur compounds. The MS ionizes molecules at low pressures and then separates them to create a spectrum of ion intensity vs. mass-to-charge ratio. The good application parameters of capillary GC/MS using the temperature gradient is illustrated on the chromatograms of sterol esters and free sterols in rapeseed and sunflower oils (Figure 2). (*See* **Chromatography**: Gas Chromatography.)

Fundamentals

The distribution of solute molecules between the stationary and mobile phases is defined by the distribution constants (K_D), i.e., the ratio of the concentration of the solute molecules in the stationary phase to that in the mobile phase:

$$K_D = \text{compound concentration of stationary phase/compound concentration of mobile phase.} \quad (1)$$

The general expression describing retention in terms of column temperature and solute properties can be expressed by:

$$\ln K_D = -\Delta G^0/RT, \quad (2)$$

where, ΔG^0 is the change in Gibbs free energy for the evaporation of the compound from the stationary phase, T is the column temperature, and R is the ideal gas constant. Equation (2) shows that the differences in the Gibbs free energy for the evaporation of the solutes from the stationary phase result in different degrees of solute retention.

The amount of time that the compound remains in the stationary phase is called the retention time (t_R).

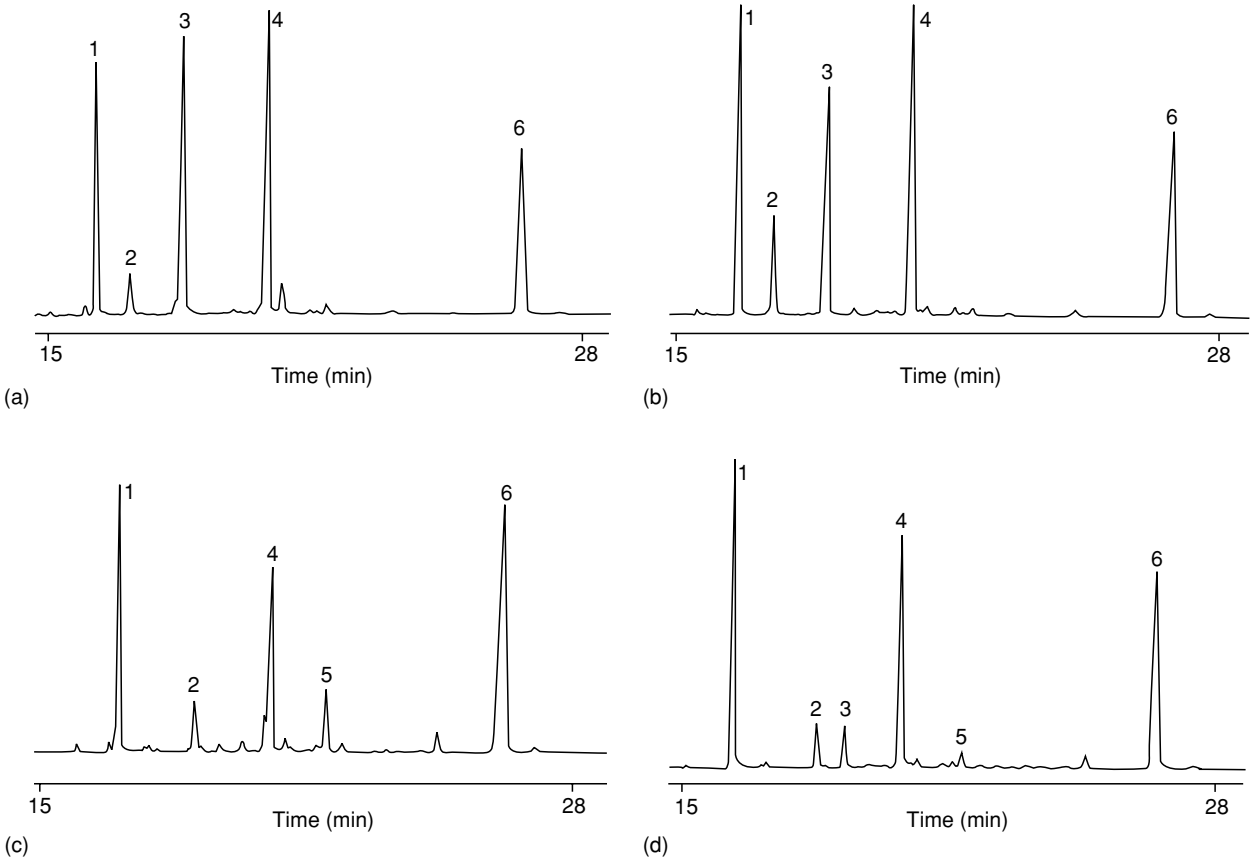

Figure 2 Gas chromatograms of sterol esters (a, c) and free sterols (b, d) in rapeseed (a, b) and sunflower oils (c, d). Peak identification: 1, cholesterol (internal standard); 2, brassicasterol; 3, campesterol; 4, sitostesterol; 5, Δ^7-stigmastenol; 6, betulin (secondary internal standard). Reprinted with permission from Kalo P and Kuuranne T (2001) Analysis of free and esterified sterols in fats and oils by flash chromatography, gas chromatography and electrospray tandem mass spectrometry. *Journal of Chromatography* 935: 237–248.

Dead time (t_0) is the time required for a nonretained solute to travel through a column.

The partition ratio, also known as the capacity factor (k') is the amount of time a compound remains in the stationary phase relative to the mobile phase.

$$k' = (t_R - t_0)/t_0. \quad (3)$$

The resolution number (R) is used to describe the efficacy of the separation of two peaks.

Resolution numbers can be calculated using either of two equations, as illustrated in Figure 3:

$$R = 1.18(t_{R2} - t_{R1})/(w_{h1} + w_{h2}) \quad (4)$$

$$R = 2(t_{R2} - t_{R1})/(w_{b1} + w_{b2}) \quad (5)$$

where t_{R1} and t_{R2} are the retention times of peaks 1 and 2, respectively, w_{h1} and w_{h2} are the peak widths at half height of peaks 1 and 2, respectively, and w_{b1} and w_{b2} are the peak widths at the base of peaks 1 and 2, respectively.

The theoretical plate number (N) characterizes the separation capacity of a chromatographic column. The number of theoretical plates is calculated by the following equation:

$$N = 5.545(t_R/w_h)^2. \quad (6)$$

Liquid Chromatography

The term liquid chromatography (LC) encompasses chromatographic methods in which the mobile phase is liquid and the stationary phase is organic or inorganic solid. According to the shape of the solid-phase support, liquid chromatographic methods can be divided into planar and column chromatography. Moreover, LC methods can be classified as normal (adsorption or direct) and reversed-phase (RP) separation modes. When the chromatographic system consists of a nonpolar mobile phase and a polar stationary phase, it is called normal chromatography. In the reversed-phase separation mode, the mobile phase is polar, and the stationary phase is nonpolar. (*See* **Chromatography**: High-performance Liquid Chromatography.)

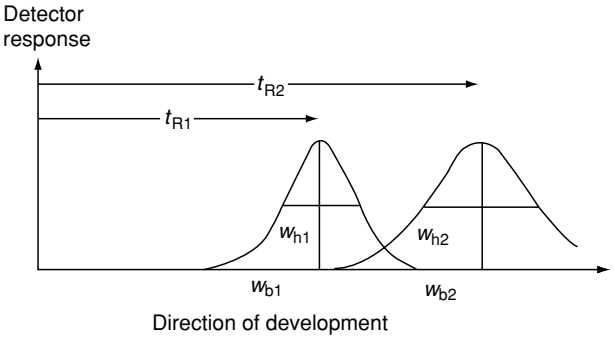

Figure 3 Schematic illustration of the mode of calculation of resolution numbers in gas and high performance liquid chromatography.

Thin-layer Chromatography

Thin-layer chromatography (TLC) is a planar chromatographic technique extensively used as a rapid and straightforward analytical tool. The success of TLC is due to its rapidity, cost-efficient optimization of separation (mobile and stationary phases can be changed easily), the high sample throughput in a short time, and the suitability for screening tests. A wide variety of inorganic and organic supports have been used for separation in TLC; however, silicas with different physicochemical characteristics (e.g., various surface pHs and pore diameters) and silicas with hydrophobic ligands covalently bonded to the surface (silanized silica, C_2, C_8, C_{18} alkyl bonded silica, amino, diol, cyano) are commonly used in TLC. Besides silica and modified silica stationary phases, other sorbents such as alumina (basic, neutral, and acidic), diatomaceous earth (Kieselguhr), celluloses and cellulose derivatives, polyamides have also been used in TLC, because their separation characteristics differ from those of silica-based layers.

Sample application Samples prepared for TLC are dissolved in an appropriate solvent and applied in the form of spots or narrow bands to the plates.

Development of chromatogram The most frequently used developing technique is ascending TLC, which is carried out in a chamber. Besides linear ascending development, linear horizontal development and anticircular development have sometimes been employed. The use of multistep gradient elution and two-dimensional development enhance the efficacy of TLC separation.

Detection in TLC is generally based on the visible or ultraviolet absorption of the solutes or on the use of various detection reagents. One of the main advantages of TLC is that a large number of detection reagents can be used that are more or less selective. In the last decade, many new instruments have been developed for the quantitative evaluation of TLC plates. Up-to-date chromatogram spectrophotometers (densitometers) are suitable for the determination of the visible, ultraviolet, and fluorescence intensity of solute spots. Characteristic densitograms obtained after TLC separation of cholesterol in liquid egg yolk are shown in Figure 4. (*See* **Chromatography**: Thin-layer Chromatography.)

Fundamentals The retention factor, R_f, characterizes the position of spots in TLC and is determined

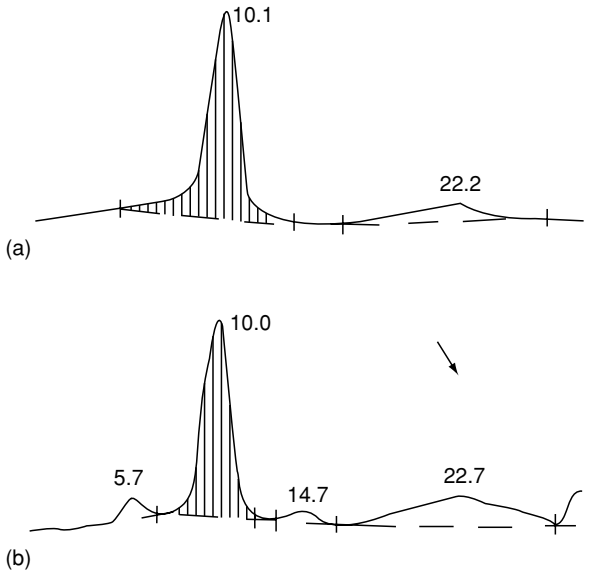

Figure 4 Typical densitogram obtained after TLC of cholesterol: (a) standard solution ($c = 400\,\mu g\,ml^{-1}$); (b) cholesterol from egg yolk. Reprinted with permission from Pozar M, Wondra AG and Prosek M (1999) TLC and HPLC determination of levels of cholesterols before and after its reduction in liquid egg yolk. *Journal of Planar Chromatography – Modern Thin-layer Chromatography* 12: 416–424.

by dividing the distance between the center of the spots and the start line (z_s) by the distance of the eluent front from the start line (z_f):

$$R_f = z_s/z_{f.} \qquad (7)$$

Other chromatographic parameters can be calculated, as described above.

High-performance Liquid Chromatography

High-performance liquid chromatography (HPLC) is the most frequently used liquid-chromatographic technique. Separation is carried out using a liquid mobile phase and a solid stationary phase filled in columns. Similar to TLC, normal-phase and reversed-phase separation modes can be employed. Supports with different separation mechanisms, such as ion exchange, size exclusion, gel-permeation chromatography, etc. can be used in HPLC.

Separation in ion-exchange HPLC is based on the interaction of the charged solute with the oppositely charged surface of the stationary phase. Ion-exchange chromatography has been increasing in popularity for the separation of inorganic species. Size-exclusion chromatography (SEC) is a common name for liquid-chromatographic methods that separate analytes based on their effective shape and bulkiness at relatively low pressures. SEC methods can be differentiated according to the character of the mobile phase.

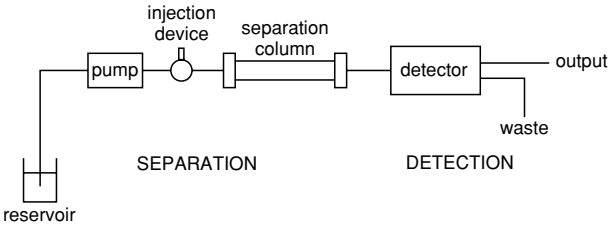

Figure 5 Schematic diagram of a typical HPLC system.

Gel-permeation chromatography (GPC) uses organic mobile phases, and gel-filtration chromatography (GFC) uses aqueous phases. The stationary phases applied in this separation mode are porous particles with closely controlled pore sizes. The molecules of the solute may be able to enter the pores of the particles of the stationary, depending on their size and shape. The choice of a mobile phase for this separation mode is simpler than for other HPLC techniques, because generally, only one solvent is required. HPLC systems can be divided into two different parts: one part is for separation, and the other part is for detection, including one or more detectors and a signal output device (Figure 5). The part of the HPLC responsible for separation consists of an injection device, a column, and the mobile-phase delivery system.

Injectors There are several approaches to injection, using a combination of a septum and syringe or valve injection.

Columns The stationary phases for adsorption HPLC are silica, alumina, porous graphitized carbon, porous glass, etc., the mobile phase generally being a mixture of relatively nonpolar organic solvents. The RP mode requires the combination of a polar mobile phase and a nonpolar stationary phase. Similar to TLC, mainly silica with hydrophobic ligands covalently bonded to the surface have been used as the stationary phase. The RP technique is popular because of the reproducibility associated with the use of the hydrocarbon-bonded phase. Octyl- and octadecyl-bonded silica stationary phases are relatively stable, with most aqueous eluents having a pH value below 8. RP HPLC uses less toxic solvents (water and an organic modifier miscible with water), thus causing less environmental pollution. Typical organic modifiers are methanol and acetonitrile.

The mobile-phase delivery system includes a pump with a filter, a degasser and transfer tubing. The pump delivers the mobile phase into the separation system with a constant, precise, and reproducible flow rate. Two different types of elution modes are used in

HPLC: isocratic and gradient elution. The isocratic elution mode uses a mobile phase of constant composition. The composition of the mobile phase is changed according to a predetermined program in gradient elution. The degasser removes the residue of gases from the mobile phase. Gases can form bubbles in the mobile phase, impairing separation and sometimes precluding the reliable detection of analytes.

Detectors The ideal HPL detector has a high sensitivity, a universal or specific response, a wide linear dynamic range, and a stable response with temperature and flow rate. The most commonly used detectors in HPLC are: ultraviolet-visible (UV-vis), fluorescence, refractive index, and electrochemical detectors.

The characteristic chromatographic parameters (separation factor, resolution, theoretical plate number) in HPLC can be calculated in the same way as in gas chromatography. (*See* **Chromatography**: High-performance Liquid Chromatography.)

Supercritical Fluid Chromatography

Principle

Supercritical fluid chromatography encompasses all chromatographic methods in which the mobile phase is supercritical under the conditions of analysis, with the solvating properties of the fluid having a measurable effect on the separation. The physical state of these conditions are illustrated in Figure 6 for CO_2, the most commonly used supercritical solvent. SFC has several advantages over GC and HPLC: it extends the molecular-weight range of GC, thermally labile compounds can be separated at lower temperatures, compounds without chromophores can be sensitively detected, and the use of open-tubular and packed columns is feasible. Figure 7 shows a schematic layout of a SFC system.

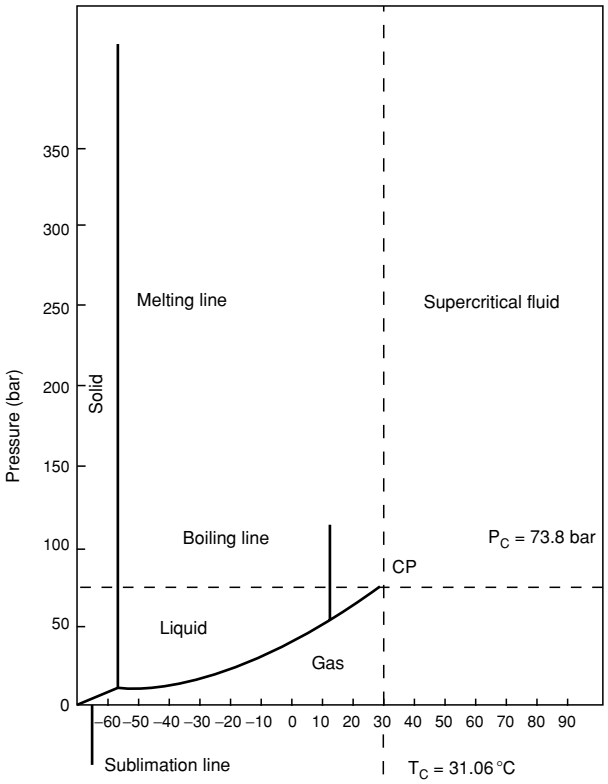

Figure 6 Phase diagram (pressure–temperature) for CO_2.

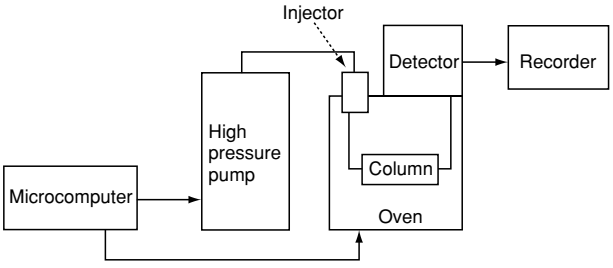

Figure 7 Schematic diagram of a SFC system.

Injectors

The most common injectors in SFC are high-pressure valve injectors similar to those used in HPLC. The method used for injection for open tubular columns is split injection. Several split-injection methods have been employed, dynamic flow split being the simplest and most popular.

Columns

Open-tubular coated capillaries (50–100 μm internal diameter (ID)), packed capillaries (100–500 μm ID), and packed columns (1–4.6 mm ID) have been used for SFC. Many types of stationary phases developed for either GC or HPLC have been adopted for use in SFC.

The separation process in SFC can be influenced by changing the pressure or the density (pressure or density programming). The programmed addition of mobile phase modifiers can also be used to change the solute retention. The mobile phase travels from the pump to the heated zone, where it becomes supercritical, and then to the injector and column.

Detectors

A major advantage of SFC over GC and HPLC is that SFC is 'detector-friendly.' Both HPLC and GC

detectors can be successfully interfaced to SFC. The most frequently used SFC detectors are flame ionization, electron capture, ultraviolet, and mass-spectrometric detectors. (*See* **Chromatography**: Supercritical Fluid Chromatography.)

Fundamentals

Carbon dioxide with or without a modifier is the mobile phase used preferentially in SFC. CO_2 is often used as the mobile phase in SFC, possibly because its physicochemical characteristics can be changed easily under moderate temperature and pressure conditions. The retention of solutes can be changed by an order of magnitude by increasing the fluid density during density programming.

The chromatographic parameters (separation factor, resolution, theoretical plate number, etc.) in SFC can be calculated in a similar way to that in GC.

Capillary Electrophoresis

Principle

Separation is based on the different electromigration velocities of ionic solutes under the influence of a high voltage in narrow-bore capillaries. The use of a capillary has numerous advantages, particularly with respect to the detrimental effects of Joule heating. The high electrical resistance of the capillary enables very high electrical fields (100–500 $\mathrm{V\,cm^{-1}}$) to be used, with only minimal heat generation. Moreover, the large surface area-to-volume ratio of the capillary efficiently dissipates any heat that is generated. The separation efficiency, which is often in excess of 10^5 theoretical plates number, is due in part to the plug profile of the electroosmotic flow, an electrophoretic phenomenon that generates the bulk flow of the solution within the capillary. This flow also enables the simultaneous analysis of all solutes, regardless of charge. The ideal properties of the capillary material are its chemical and electrical inertness, UV-visible transparency, flexibility, and low cost.

CE is characterized by its simple experimental setup. A schematic diagram of a basic CE instrument is shown in Figure 8. Both ends of a capillary filled with buffer are dipped into two buffer reservoirs that are maintained at the same level. A high voltage is then applied to this capillary.

Various modes of capillary electrophoresis can be performed using the same intruments, including capillary zone electrophoresis (CZE), capillary gel electrophoresis (CGE), micellar electrokinetic capillary chromatography (MEKC), capillary isoelectring focusing, and capillary isotachophoresis. Figure 9 shows the different modes of capillary electrophoresis. MEKC is the second most popular CE separation technique (after CZE), because it can separate neutral analytes too. A surfactant above its critical micellar concentration is added to the background electrolyte. Solutes with varying degrees of hydrophobicities partition between the hydrophobic core of the micelles and aqueous buffer. The interaction between the micelle and the neutral solutes leads to separation. The separation mechanism of neutral solutes in MEKC can be described using modified chromatographic relationships. The ratio of the total number of moles of solute in the micelle (that is in the pseudostationary phase) to those in the mobile phase, the capacity factor k', is given by:

$$k' = (t_r - t_0)/[t_0(1 - t_r/t_m)] = K(V_s/V_m), \qquad (8)$$

where t_r is the retention time of the solute, t_0 is the retention time of unretained solute moving at the EOF rate, t_m is the micelle retention time, K is the partition coefficient, V_s is the volume of the micellar phase, and V_m is the volume of the mobile phase.

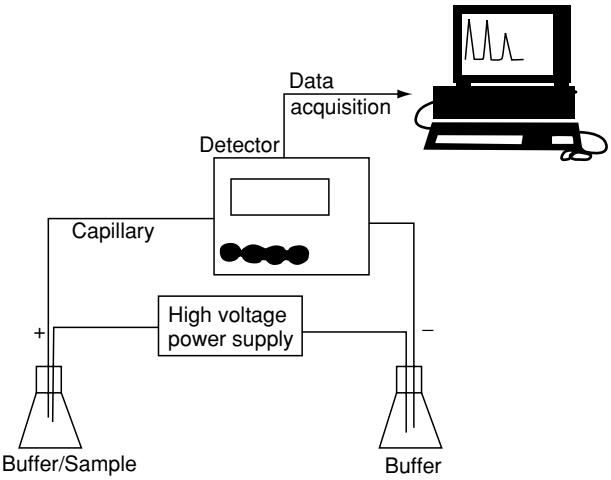

Figure 8 Schematic diagram of a CE system.

Injectors

Two different injection modes can be used: hydrodynamic and electrokinetic injections. With hydrodynamic injection, a pressure difference is applied for a short period of time between the two ends of the capillary. A pressure difference can be introduced, depending on the design of the instrument, by pressure, vacuum, or simply the height difference between the levels of the buffer and sample reservoirs. Electrokinetic injection is performed by replacing the injection-end reservoir with the sample vial and applying the voltage.

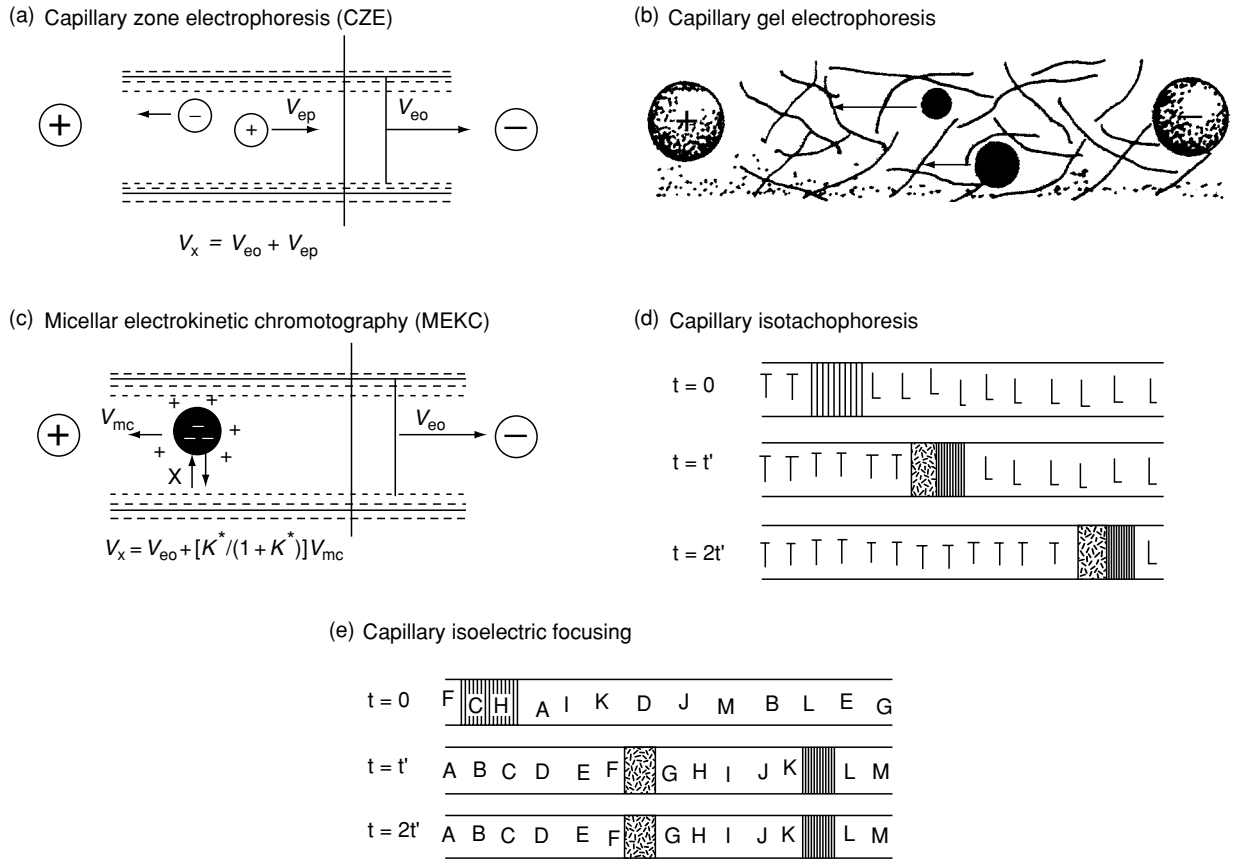

Figure 9 Different modes of capillary electrophoresis. In (a), V_x is the electromigration velocity, and V_{eo} is the electroosmotic velocity. In (b), there are obstructing strands of gel, and large ions move more slowly through the pores than the small ions. In (c), V_x is the electromigration velocity of the solute, V_{eo} is the electroosmotic velocity, V_{mc} is the migration velocity of the micelle, and K^* is the phase capacity ratio. In (d), L is the leading electrolyte, and T is the terminating electrolyte.

Detection

Detection is generally carried out by measuring the ultraviolet or visible absorbance at the end of the capillary. Other detection systems such as laser-induced fluorescence, mass spectrometry, and conductivity detection have also been employed.

Fundamentals

The basic phenomena in capillary electrophoresis are electroosmotic flow (EOF) and electrophoretic flow (μ_{ep}). Electroosmotic flow is bulk flow of a solvent in a capillary under an applied potential. Electrophoretic flow is the flow of ions due to a certain charge, with cations being attracted to the cathode and anions being attracted to the anode.

The magnitude of the EOF can expressed by

$$\mu_{EOF} = (\varepsilon \cdot \zeta)/\eta, \qquad (9)$$

where μ_{EOF} = EOF mobility; ε = dielectric constant; ζ = zeta potential; η = bulk viscosity.

According to eqn (9), the mobility is independent of the applied electric field. The zeta potential is essentially determined by the surface charge on the capillary wall. Since it is strongly pH-dependent, the magnitude of the EOF varies with pH. Depending on the specific conditions, the EOF can vary by more than an order of magnitude between pH 2 and 12. The zeta potential is also dependent on the ionic strength of the buffer.

The separation parameters of capillary electrophoresis can be described in similar terms to those for GC and HPLC.

The time required for a solute to migrate to the point of detection is called the migration time, and is given by the quotient of migration and velocity. The migration time and other experimental parameters can be used to calculate the apparent solute mobility, using

$$\mu_a = 1/tE = 1L/tV \qquad (10)$$

and

$$\mu_a = \mu_e + \mu_{EOF}, \qquad (11)$$

where l is the effective capillary length (to the detector), L is the total length, t is the migration time, and E is the electric field. In the presence of EOF, the measured mobility is called the apparent mobility (μ_a). The effective mobility (μ_e) can be calculated from the apparent mobility by independently measuring the EOF using a neutral marker that moves at a velocity equal to the EOF.

See also: **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; Combined Chromatography and Mass Spectrometry

Further Reading

Cserháti T and Forgács E (1999) *Chromatography in Food Science and Technology*. Lancaster, PA: Technomic.
Forgács E and Cserháti T (1997) *Molecular Basis of Chromatographic Separation*. Boca Raton, FL: CRC Press.
Grob RL (1995) *Modern Practice of Gas Chromatography*. New York: John Wiley.
Grossmann PD and Colburn JC (eds) (1992) *Capillary Electrophoresis – Theory and Practice*. San Diego, CA: Academic Press.
Hill HH and McMinn DG (1995) *Detectors for Capillary Chromatography*. New York: John Wiley.
Jennings W (1987) *Analytical Gas Chromatography*. Orlando, FL: Academic Press.
Jinno K (1992) *Hyphenated Techniques in SFC and Extraction*. New York: Elsevier Science.
Li SFY (1992) *Capillary Electrophoresis – Principles, Practice and Applications, Journal of Chromatography Library*. Amsterdam: Elsevier Scientific.
Maguire KL and Denyszyn RB (1985) In: White CM (ed.) *Modern Supercritical Fluid Chromatography*, p. 45. New York: Springer.
Sherma J and Fried B (eds) (1991) *Handbook of Thin Layer Chromatography*, vol. 55. New York: Marcel Dekker.
Snyder LR (1968) *Principles of Adsorption Chromatography*. New York: Marcel Dekker.
Snyder LR, Glajch JL and Kirkland JJ (1988) *Practical HPLC Method Development*. New York: John Wiley.
Vindevogel J and Sandra P (1992) *Introduction to Micellar Electrokinetic Chromatography*. Heidelberg: Hürtig.

Thin-layer Chromatography

D R Tocher, University of Stirling, Stirling, UK

Introduction

Thin-layer chromatography (TLC) is a type of liquid chromatography performed on a stationary phase in the form of a thin layer usually on a glass, plastic, or aluminum support. A solution of sample is applied as a small spot or narrow band to the thin layer of adsorbent that has been spread uniformly over the support. The mobile phase (a solvent mixture) then passes through the adsorbent by capillary action, and the sample is resolved into discrete components. The separated components are then located by either physical methods or chemical staining reagents after the solvent has evaporated. Quantitation by methods such as densitometry may also then be possible, depending upon the precise conditions utilized. Although high-performance liquid chromatography (HPLC) and gas chromatography (GC) are often regarded as superior, especially with regard to quantitation, TLC has many advantages including low capital costs, simplicity of operation, the availability of many sensitive and selective reagents for detection and confirmation without interference of the mobile phase, and in-system calibration for quantitative analysis. Therefore, TLC should be regarded as highly complementary to the other techniques and particularly suited to low-cost screening of many samples.

Apparatus

TLC is very cost-effective as the basic apparatus required is simple and relatively cheap, and many samples can be analyzed on a single plate with low solvent usage. However, more sophisticated and expensive apparatus are available particularly for automated sample application and quantitation.

Adsorbents and Supports

The predominant adsorbent used in TLC is silicic acid, a partially hydrated silicon dioxide, often termed silica gel, but others are also available, including cellulose, aluminum oxide, celite (kieselguhr), magnesium oxide, and zeolites. Modified versions of these include C_8- or C_{18}- (reverse-phase) and 3-aminopropyl-silica gel, polyamide- and polyethylene imine (PEI)-cellulose. The particle size of the adsorbents is usually in the range 1–25 μm.

TLC plates were often made in the laboratory using specific apparatus. However, home-made plates could be variable, and commercially produced precoated plates, which are available in various sizes and supports with almost every adsorbent, are now the first choice due to their much greater reproducibility and consistent quality. Adherence of the adsorbent to the support is often enhanced in precoated plates by the inclusion of binders such as calcium phosphate ($\sim$10% w/w) or organic binders like polyacrylic acid.

Silica gel is the most commonly employed adsorbent in TLC, particularly that with a pore size of 6 nm. Precoated analytical plates frequently use silica gel of particle size 10–12 μm coated in a layer of 0.25 mm thickness. Plates coated with this type of silica provide 1000–2000 theoretical plates per 5 cm of migration. Modern high-performance TLC (HPTLC) uses silica gel of particle size 5–7 μm in layers of 0.20 mm thickness. The reduced particle size provides 5000–10 000 theoretical plates per 5 cm of migration distance and increases the separation efficiency. In comparison with conventional TLC, HPTLC is a more efficient, instrumentalized, and quantitative method. The narrower particle size distribution, thinner layers, and shorter development distances of HPTLC lead to a greater separation efficiency, faster separations, and enhanced sensitivity through improved detection limits such that quantitative HPTLC can produce results comparable with GC and HPLC when optimally performed. HPTLC plates cannot be prepared easily in the laboratory and are usually purchased ready-made.

The adsorption properties of silica gel can be modified by impregnation with various complexing reagents, including silver nitrate, urea, boric acid, oxalic acid, and ethylenediaminetetraacetic acid (EDTA) to aid the separation of particular classes of compounds. Reverse-phase plates can be produced by using silanized silica gel in which the surface silanol groups of the silica are silylated with chlorosilanes. Silanized silica has a low carrying capacity that can be increased by impregnating the silanized silica plates with hexadecane or paraffin oil. It is now more common to utilize ready-made reverse-phase plates with C_2, C_8, or C_{18} hydrocarbon chains, or diphenyl groups chemically bonded to the silica. Aminopropyl bonded silica gel plates are also available. Kieselguhr or celite, a diatomous earth, is used less often than synthetic silica gel on account of its natural variability. Other adsorbents include cellulose and its derivatives, such as diethylaminoethyl (DEAE)-cellulose for ion-exchange chromatography, magnesium oxide, and alumina (aluminum hydroxide), which is manufactured in three types, acidic, basic, and neutral, according to its pH.

Glass is the support employed most commonly for adsorbents on grounds of rigidity, flatness, and inertness. Layers of adsorbents can also be obtained precoated on flexible aluminum or plastic (polyester) sheets, which have the advantage of being able to be cut to smaller size, but care must be taken in insuring their compatibility with solvent systems and subsequent detection reagents. The standard sizes of TLC and HPTLC plates are 20×20 cm and 10×10 cm, respectively, although other sizes such as 10×20 cm are available (Figure 1).

Sample Applicators and Development Chambers

The application of the sample solution to the adsorbent layer is usually performed manually using microsyringes or disposable glass capillaries. However, a wide range of instruments, including microprocessor-controlled instruments, are available for the automated application of samples either as spots or streaks. Rectangular glass chambers with lids are usually employed for the linear development of chromatograms in solvent mixtures. Sizes are available for standard 20×20 cm plates and 10×10

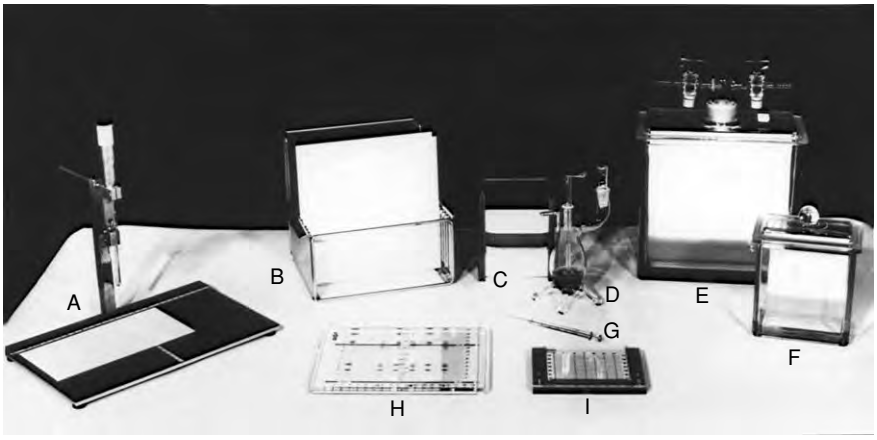

Figure 1 Typical apparatus used in TLC. (**A**) Applicator with micrometer-controlled syringe containing 10×20 cm HPTLC plate; (**B**) drying/storage rack containing standard 20×20 cm TLC plates; (**C**) 10×10 cm HPTLC plates in drying/storage rack; (**D**) atomizer for spraying plates; (**E**) 20×20 cm TLC plate being developed in a standard chamber; (**F**) 10×10 cm HPTLC plate being developed in HPTLC chamber; (**G**) glass microsyringe; (**H**) sample application template/R_f calculator overlayed on the standard plate; (**I**) HPTLC plate contained in the sample application template.

HPTLC plates. Chambers can be purchased with vertical grooves in the side and end walls to allow the simultaneous development of several plates (Figure 1).

Equipment for Detection and Quantification of Components

Atomizers operated by compressed air or hand bellows are required for spraying developed chromatograms evenly with detection reagents. An oven is necessary for heating plates sprayed with certain reagents. For the quantitation of separated components on developed chromatograms, modern computer-controlled scanning densitometers are available along with sophisticated software packages for evaluation and data storage. Densitometry is particularly useful for one-dimensional chromatography, but programs can be adapted for two-dimensional chromatograms. The location and quantitation of radiolabelled compounds can be achieved using specialized radioactivity detectors that can scan the whole chromatogram, although these instruments are rather expensive. However, less elaborate equipment is required for the location of radiolabelled compounds on chromatograms by autoradiography. Fluorography can be performed with similar equipment. New detection apparatus that may prove useful include the charge coupled device array detector for HPTLC and matrix-assisted laser desorption/ionization mass spectrometry.

Specialized Systems

Several techniques that rely on the same chromatographic principle as conventional TLC, but which require the use of specialized equipment, have been developed. The most durable of these specialized systems is probably TLC with flame ionization detection (TLC–FID), which was developed to ally the quantitative benefits of FID to TLC. In the 'Iatroscan' system, samples are applied to silica gel fused to thin quartz rods, which are then developed conventionally in mobile phase. After drying, the developed 'chromarods' are passed through a flame ionization detector to determine the mass of individual components by quantifying the amount of ionizable carbon. The 'Iatroscan' has been relatively successful, particularly in certain areas such as lipid metabolism. Another reasonably successful system is automated multiple development (AMD), which permits the multiple development of TLC plates in the same or different solvent systems. This repeated development of TLC plates can improve the resolution of components.

Other specialized systems have not had such a durable or widespread appeal but have had limited use. In over-pressured TLC (OPTLC), the TLC plate is held beneath a flexible membrane under hydrostatic pressure, and solvent is forced through the adsorbent using a pump. This system allows compounds to be eluted completely from the TLC plate for collection or detection using detectors normally associated with HPLC. Centrifugal TLC also involves the forced flow of solvent. In this case, solvent is fed into the center of a rapidly rotating TLC or HPTLC plate and is forced through the adsorbent by centrifugal force. In radial-development TLC, the sample is applied to the center of a plate and the solvent supplied through a hole in the plate via a wick that dips into a solvent reservoir. Another system allowed the application of solvent under pressure to HPTLC plates allowing very short development times. (*See* **Chromatography**: High-performance Liquid Chromatography.)

Methodology

Sample Application

In conventional TLC and HPTLC, the sample is applied as a small volume of solution on to the adsorbent near the bottom of the plate. The point of application is termed the origin. A maximum of about 80 or 10 μg for TLC and HPTLC, respectively, can be applied as a spot on adsorbent layers of silica gel G 0.20–0.25 mm thick. Samples can also be applied as a narrow streak along the origin. Standards are applied as separate spots on the same plate.

Development

The choice of solvent system depends on the components to be resolved and the type of adsorbent utilized (see Table 1). Regardless of the components, in conventional TLC, the solvent mixture is placed in the development chamber. In the case of solvent systems containing large proportions of polar solvents, the chamber can be lined with filter paper to aid saturation of the atmosphere. The plate is placed within the chamber so that the bottom edge is immersed in the solvent mixture, and the chamber is sealed with a lid (Figure 1). When the solvent has migrated to within 1 cm of the top edge of the plate, the plate is removed from the chamber. Although ascending chromatography, as described above, is the most common form of TLC, chromatograms can also be developed by descending chromatography using a wick to feed the solvent on to the adsorbent layer.

TLC plates can also be subjected to multiple development in which the plate is developed partially in one solvent system and removed from the

Table 1 Substances separated by selected one-dimensional TLC systems

Adsorbent	*Substances separated*	*Solvent system*[a]	*Detection reagent*
Alumina	*Alkaloids*	$CHCl_3$	Dragendorff reagent
	Amino acids (dicarboxylic)	2 M HAc	Ninhydrin
	Aromatic hydrocarbons	CCl_4	10% tetracyanoethylene in C_6H_6
	Carotenes	Hexane/C_6H_6/EtOH (100:100:1)	None required
	Food dyes (fat-soluble)	Hexane/EtAc (98:2)	None required
	Food dyes (water-soluble)	H_2O/EtOH/*n*-BuOH (1:1:5)	None required
	Neutral lipids	Hexane/Et_2O/HAc (94.5:5:0.5)	0.05% 2′,7′ dichlorofluorescein in MeOH
	Sterols	C_6H_6/EtOH (95:2)	10% phosphotungstic acid in 90% EtOH
	Steroids	$CHCl_3$/EtOH (99:1)	Anisaldehyde/H_2SO_4/HAc (1:2:100), UV
	Vitamins (fat-soluble)	Toluene	1.3% $FeCl_3$ in 2 M HCl/0.7% K_3FeCN_6 (1:1)
Cellulose	*Amino acids*	*n*-BuOH/Me_2CO/Et_2NH/H_2O (10:10:2:5)	Ninhydrin
	Antibiotics (water-soluble)	PrOH/pyridine/HAc/H_2O (15:10:3:12)	Ninhydrin
	Food dyes (fat-soluble)	*n*-PrOH/EtAc/H_2O (6:1:3)	None required
	Nucleotides	*n*-BuOH/Me_2CO/HAc/5% NH_4OH/H_2O (4.5:1.5:1:1:2)	20% $SbCl_5$ in CCl_4
	Nucleosides and free bases	H_2O	20% $SbCl_5$ in CCl_4
DEAE cellulose	*Amino acids*	*n*-BuOH/HAc/H_2O (4:1:5)	Ninhydrin
	Nucleotides	Isobutyric acid/NH_4OH/H_2O (33:1:16)	20% $SbCl_5$ in CCl_4
PEI cellulose	*Nucleotides*	0.25, 1.0, 1.6 M LiCl in H_2O (in successive developments)	20% $SbCl_5$ in CCl_4
Celite (Kieselguhr)	*Sugars*	EtAc/*iso*-PrOH/H_2O (130:57:23)	0.2% naphthoresorcinol in EtOH/10% H_3PO_4
	Oligosaccharides	*Iso*-PrOH/EtAc (65:35)	0.2% naphthoresorcinol in EtOH/10% H_3PO_4
	Food dyes (fat-soluble)	Cyclohexane	None required
Magnesium oxide	*Carotenoids*	Pet. ether/C_6H_6 (3:1) or pet. ether/Me_2CO (97:3)	None required
Polyamine	*Anthocyanins*	HAc/HCl/H_2O (10:1:3)	10% oxalic acid in Me_2CO/H_2O (1:1), UV
	Antioxidants	MeOH/Me_2CO/H_2O (6:1:3)	10% phosphomolybdic acid in EtOH
	Flavonoids	Me_2CO/95% EtOH/H_2O (2:1:2)	25% PbAc in basic aqueous solution
Sephadex	*Proteins*	0.5 M NaCl	1% naphthalene black in MeOH/H_2O/HAc (5:4:1)
Silica gel	*Alkaloids*	$CHCl_3$/Me_2CO/Et_2NH (5:4:1)	Dragendorf reagent
	Amino acids	*n*-BuOH/HAc/H_2O (4:1:1)	Ninhydrin
	Antibiotics		
	Macrolides	$CHCl_3$/MeOH/H_2O (80:20:2.5)	20% phosphomolybdic acid in EtOH
	Penicillins	Me_2CO/MeOH (1:1)	20% phosphomolybdic acid in EtOH
	Streptomycins	H_2O/Na citrate/citric acid (100:20:5)	20% phosphomolybdic acid in EtOH
	Tetracyclines	BuOH/HAc/H_2O (2:1:1)	20% phosphomolybdic acid in EtOH
	Antioxidants	$CHCl_3$	20% phosphomolybdic acid in EtOH
	Bile acids	Me_3pentane/*iso*-PrOH/HAc (60:20:0.5)	5% phosphomolybdic acid in EtOH/Et_2O (1:1)
	Food dyes (fat-soluble)	Hexane/Et_2O/HAc (70:30:1)	None required

Continued

Table 1 Continued

Adsorbent	*Substances separated*	*Solvent system*[a]	*Detection reagent*
	Food dyes (water-soluble)	HAc/*iso*-BuOH/H_2O (2:5:2)	None required
	Gangliosides	$CHCl_3$/MeOH/0.3% KCl (30:18:4) to purify $CHCl_3$/MeOH/0.25% $CaCl_2$ (55:45:10) to resolve classes	0.5% orcinol in 20% sulfuric acid 0.5% orcinol in 20% sulfuric acid
	Insecticide (Imidacloprid)	$CHCl_3$/Me_2CO/MeOH (23:1:1)	UV
	Lipids (neutral)	Hexane/Et_2O/HAc (80:20:2)	3% CuAc in 8% H_3PO_4
	Lipids (polar)	MeAc/*iso*-PrOH/$CHCl_3$/MeOH/ 0.25% aq. KCl (25:25:25:10:9)	3% CuAc in 8% H_3PO_4
	Lipopolysaccharides	*iso*-PrOH/H_2O/$CHCl_3$/NH_4OH Et_3NH (120:60:16:4:1)	0.1% orcinol in 30% aq. MeOH
	Mycotoxins	$CHCl_3$/Me_2CO (90:10)	*p*-Anisaldehyde/MeOH/HAc/ H_2SO_4 (0.5:70:10:5)
	Opiates (acetate or methoximine derivatives)	CH_2Cl_2/*iso*-PrOH (88:12) + 1.5% NH_4OH	Conc. Sulfuric acid
	Plasticizers	CH_2Cl_2	4 M H_2SO_4/20% resorcinol (1:1)
	Mono-, di-, and trisaccharides	MeCN/H_2O (85:15)	0.5% $KMnO_4$
	Polysaccharides	*n*-BuOH/MeOH/H_2O (50:25:20)	0.5% $KMnO_4$
	Shellfish toxin (domoic acid)	BuOH/HAc/H_2O (3:1:1)	UV and ninhydrin
	Steroids	$CHCl_3$/EtOH (92:8)	10% phosphomolybdic acid in EtOH
	Synthetic sweeteners	$CHCl_3$/HAc (90:10)	0.2% 2′,7′ dichlorofluorescein in EtOH
	Terpene alcohols	CH_2Cl_2	0.06% diphenylpicryl hydrazyl in $CHCl_3$
	Terpene aldehydes	$CHCl_3$	10% $SbCl_5$ in CCl_4
	Vitamins (fat-soluble)	Cyclohexane/EtAc (75:25)	UV
	Vitamins (water-soluble)	HAc/Me_2CO/MeOH/C_6H_6 (5:5:20:70)	UV
3-Aminopropyl silica	*Oligosaccharides*	MeCN/10 mM Me_3NH_2Ac (3:2)	Fluorography after spraying with 4-methylenaphthalene enhancer
$C_{8/18}$-silica (reverse-phase)	*Alkaloids*		
	Antifungals	MeOH/H_2O (90:10)	UV
	Carotenoids	Pet.ether/MeCN/MeOH (2:4:4)	None required
	Quinones	Me_2CO/H_2O (19:1)	10% phosphomolybdic acid in EtOH
	Vitamin K	CH_2Cl_2/MeOH(70:30)	UV

[a]Solvent proportions by volume.

Ac, sub > /sub > CO_{2^-}; Bu, C_4H_9-;Et, C_2H_5-; Me, CH_3-; Pr, C_3H_7-; DEAE cellulose, diethylaminoethyl cellulose; PEI cellulose, polyethylene imine cellulose; UV, ultraviolet. For health and safety reasons, C_6H_6 can usually be replaced with toluene, and hexane can be replaced with *iso*-hexane.

chamber. After evaporation of solvent from the developed plate, the plate is developed fully or partially in another solvent system. Multiple developments enable improved resolution of components. Development of the chromatogram with the mobile phase in one direction only is known as one-dimensional TLC (Figure 2). In instances where one-dimensional TLC/HPTLC does not allow the complete resolution of all the components of a mixture, improved separation can be achieved with two-dimensional TLC/HPTLC. In this technique, the sample is applied as a spot at one corner and developed fully in one direction in the first solvent system. The plate is removed from the chamber and, after evaporation of the solvent, is then developed in a second solvent system in a direction at 90° to that of the first development (Figure 3).

Identification of Separated Components

Colored compounds are visible on developed chromatograms, but colorless compounds require detection by physical or chemical means. Detection by chemical means usually involves spraying the adsorbent layer of the developed chromatogram with a derivatizing reagent that reacts with the separated components to produce a colored derivative *in situ*. Nonspecific reagents such as sulfuric acid and iodine will detect a wide range of compounds. Specific reagents react with specific functional groups and will

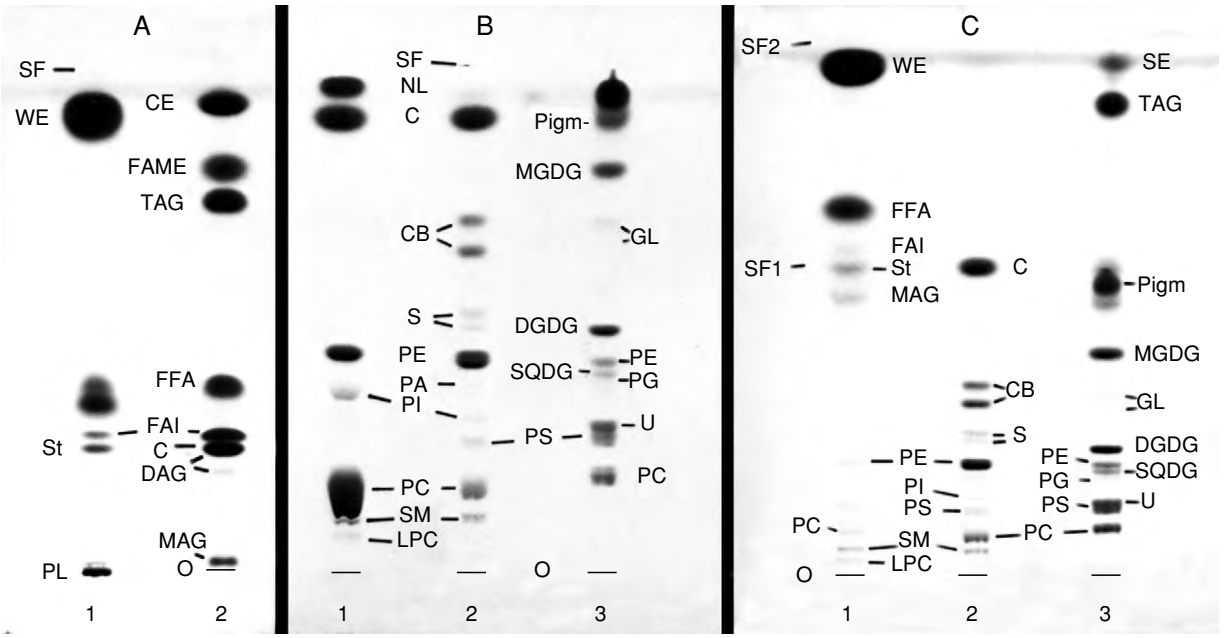

Figure 2 HPTLC chromatograms of lipid mixtures developed in various systems. (**A**) Single development in hexane/Et_2O/HAc (80:20:2 by volume): 1, zooplankton total lipid; 2, standard authentic mixture, all based on oleic acid. (**B**) Single development in MeAc/*iso*-PrOH/$CHCl_3$/MeOH/0.25% aqueous KCl (25:25:25:10:9, by volume): 1, cod roe total lipid; 2, rat brain total lipid; 3, algal (*Croomonas salina*) total lipid. (**C**) Double-development with development to half final distance in solvent system of (**B**) followed by full development in solvent system of (**A**): 1, Zooplankton total lipid; 2, rat brain total lipid; 3, algal (*C. salina*) total lipid. All chromatograms were stained by charring after spraying with 3% copper acetate in 8% orthophosphoric acid. C, cholesterol; CB, cerebrosides; CE, cholesteryl ester; DAG, diacylglycerol; DGDG, digalactosyl diacylglycerol; FAI, fatty alcohol; FAME, fatty acid methyl ester; FFA, free fatty acid; GL, glycolipid; LPC, lyso-phosphatidylcholine; MAG, monoacylglycerol; MGDG, monogalactosyl diacylglycerol; NL, neutral lipid; O, origin; PA, phosphatidic acid; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PI, phosphatidylinositol; Pigm, pigment; PL, polar lipid; PS, phosphatidylserine; SF, solvent front; SM, sphingomyelin; SQDG, sulfoquinovosyldiacylglycerol; St, sterol; TAG, triacylglycerol; U, unknown; WE, wax ester.

only detect compounds containing the group, e.g., ninhydrin for amino groups. The use of specific reagents can be used as an aid to identification and characterization of components.

Physical methods of location commonly involve ultraviolet light. Commercial plates are available coated with adsorbents containing an indicator that absorbs light at 254 nm and reemits or fluoresces light at the green end of the spectrum. When separated on these plates, compounds that absorb ultraviolet light appear as dark spots on a green fluorescent background. Plates with adsorbents not containing fluorescent indicator can be sprayed with 2′,7′-dichlorofluorescein with similar results when viewed under ultraviolet light. Radiolabelled compounds can be located on TLC plates physically by autoradiography or *in situ* measurement of radioactivity using a radioscanner.

A useful aid to the identification of separated components is the R_f value, which is defined as the ratio of the distance traveled by the compound to the distance traveled by the solvent. Although the R_f value of a given compound in a defined mobile phase and adsorbent is very characteristic, many factors, including the thickness and moisture content of the adsorbent and development distance, affect the reproducibility of the value. For this reason, the R_f value is only an indication of the identity of a compound, and confirmation of identity should be obtained by other means. Identification of separated components can be aided by comparison with authentic standards run alongside the samples, if such standards are available, and by staining patterns to different reagents.

Quantification of Separated Components

Once located, it is often useful and necessary to quantify the components in a mixture, and this has been the most difficult aspect of TLC. However, various methods can be employed for the quantitation of separated components.

Areas of adsorbent that contain components can be scraped from the support and the compounds eluted from the adsorbent with suitable solvents. The recovered compounds can be estimated gravimetrically after evaporation of the solvent, or by using a

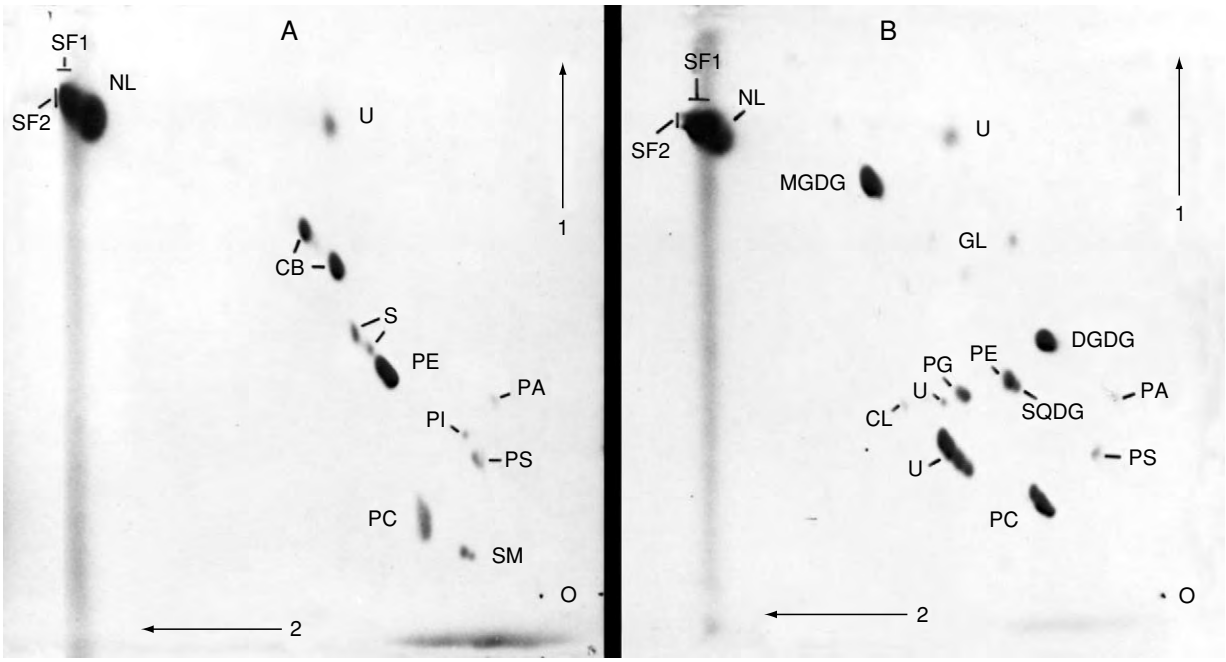

Figure 3 Two-dimensional HPTLC of total lipid from (**A**) rat brain and (**B**) the alga *C. salina*. Plates were first developed in direction 1 with MeAc/*iso*-PrOH/$CHCl_3$/MeOH/0.25% aqueous KCl (25:25:25:10:9, by volume) and then in direction 2 with $CHCl_3$/MeOH/7M NH_4OH (65:35:5, by volume). Developed chromatograms were stained by charring after spraying with 3% copper acetate in 8% orthophosphoric acid. Abbreviations: as in **Figure 2**.

specific assay such as the measurement of phosphorus or sugar content.

Developed chromatograms can be subjected to scanning densitometry and the separated components quantified on the basis of transmittance, reflectance intensity, or fluorescence, depending on the nature of the compounds and the detection reagent/stain used. Modern techniques of imaging in conjunction with sophisticated computer software packages for image analysis and data processing have increased the accuracy of these techniques. As well as locating radiolabeled compounds, radioscanners can also quantify the amount of radioactivity present in individual components. Alternatively, bands of adsorbent that contain compounds can be scraped from the support into scintillation vials and measured directly for radioactive content by liquid scintillation counting after addition of suitable scintillation fluid.

Preparative TLC

Although TLC is mostly used as an analytical technique, one-dimensional TLC can be scaled up to a preparative scale. To this end, sample can be applied as a streak across the origin of an analytical plate, the length of the streak being determined by the amount of sample. Special preparative TLC plates coated with thicker layers of adsorbent can be obtained, which can accommodate relatively large amounts of sample. However, these plates lack the resolving power of the analytical plates. After nondestructive visualization of the separated components, the bands of adsorbent containing the compound or compounds of interest are scraped from the support and eluted with suitable solvents.

Modes of Chromatography

TLC conforms to the basic principles of liquid chromatography. The most common mode of chromatography used in TLC is adsorption chromatography, although other modes of chromatography can be employed. With silica, celite, keiselghur, and cellulose, the mechanism is by adsorption chromatography if the adsorbent on the plate is completely free of water, and the solvent system is a nonpolar mixture. However, if water is present on the adsorbent or if the mobile phase contains a highly polar solvent, then separation will be by partition chromatography. In addition, reverse-phase TLC plates also separate by partition chromatography. TLC plates are also available precoated with reverse-phase silica gel that has been impregnated with a chiral reagent and copper ions. These can be used to separate optically active isomers, e.g., amino acids, by chiral chromatography on the basis of ligand exchange.

Alumina separates components by adsorption but, depending upon the nature of the surface and the solvent system, can also function as an ion exchanger.

Modified cellulose such as DEAE-cellulose can be used for ion-exchange separations on TLC. Separation by size-exclusion chromatography can be carried out with TLC plates coated with Sephadex gel but is slower and significantly more difficult than other forms of TLC. Polyamides such as polyhexamethylenediaminoadipate can be used as the adsorbent to separate components that interact with it by hydrogen bonding. (*See* **Chromatography**: Principles.)

Applications

The actual method employed for preparing samples from foodstuffs for analysis by TLC depends on the nature of the substances being investigated. Many methods involve the extraction of the food with a suitable solvent followed by precipitation and filtration steps to remove classes of compounds that are not of interest. (*See* **Analysis of Food**.)

The choice of adsorbent and solvent system used in TLC is dictated by the nature of the sample to be analyzed. Examples of broad groups of substances that have been successfully resolved by TLC methods are shown in Table 1, along with suggested solvent systems and detection reagents commonly used. However, for most substances, there can be a variety of different solvent systems and detection reagents that are suitable, depending on the precise composition of the samples to be resolved.

See also: **Chromatography**: Principles; High-performance Liquid Chromatography; Gas Chromatography

Further Reading

Ackman RG (1982) Flame ionization detection applied to thin-layer chromatography on coated quartz rods. *Methods in Enzymology* 72: 205–252.

Henderson RJ and Tocher DR (1992) Thin-layer chromatography. In: Hamilton RJ and Hamilton S (eds) *Lipid Analysis: A Practical Approach*, pp. 65–111. Oxford: Oxford University Press.

Jork J, Funk G, Fischer G and Wimmer CJ (1989) *Thin-layer Chromatography: Reagents and Detection Methods*. Cambridge: VCH.

Kirchner JG (1978) *Thin-layer Chromatography. Techniques of Chemistry*, vol. XIV. New York: Wiley.

Sherma J (1996) Thin layer chromatography in food analysis. *Food Testing and Analysis* 2: 12–15.

Touchstone JC (1988) Instrumentation for thin-layer chromatography: a review. *Journal of Chromatographic Science* 26: 645–649.

Wilson ID (1996) Thin layer chromatography: a neglected technique. *Therapeutic Drug Monitoring* 18: 484–492.

High-performance Liquid Chromatography

M V Moreno-Arribas and M C Polo, Instituto de Fermentaciones Industriales (CSIC), Madrid, Spain

Introduction

High-performance liquid chromatography (HPLC) is an instrumental form of liquid chromatography that employs stationary phases consisting of small particles, thereby achieving more efficient separations than those used in conventional liquid chromatography. Since its origin in the late 1960s, it has been known by several different names, including high-pressure liquid chromatography, because of the high pressures required to force the mobile phase or solvent through the stationary phase, and high-resolution liquid chromatography, because of the good resolution achieved using this technique.

This article describes the equipment needed to carry out HPLC and summarizes some of the applications of this technique in food analysis. Two techniques related to HPLC, fast-protein liquid chromatography (FPLC) and supercritical fluid chromatography (SFC), are also considered. The basic theory of the chromatographic process and the factors that affect separation efficiency are discussed in the section on Chromatography. (*See* **Chromatography**: Principles; Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; Supercritical Fluid Chromatography; Combined Chromatography and Mass Spectrometry.)

Separation Modes

There are many different stationary and mobile phases that can be used in HPLC, and for this reason there is a great variety of separation mechanisms. The treatises on HPLC use different criteria in their attempt to classify the modes of this technique: type of stationary phase, predominant separation mechanism, type of groups of compounds it is aimed at. These criteria sometimes overlap; that is to say, it is possible to work in several different chromatographic methods with the same stationary phase, and the same group of compounds can be separated using several different stationary phases, and for this reason it is not easy to classify HPLC techniques into specific groups. Generally speaking, it can be said that there are three large groups

that dominate the field of chromatographic separations: reversed-phase chromatography, ion exchange chromatography and size exclusion chromatography. There are also other types of chromatography that are rapidly gaining in importance and cannot be included exactly in any of these three groups: ion pair chromatography, affinity chromatography, and chiral separations.

Reversed-Phase Chromatography

This is the most commonly used method for chromatographic separations, so it could in fact be called 'normal-phase' except that it was developed later than the latter. In reversed-phase chromatography, the stationary phase is of a nonpolar nature and the mobile phase is more polar than the stationary one. Among the most widely used stationary phases are silica-based ones with cyano, phenyl, or alkyl functionality and those of a polymeric nature, above all polystyrene-divinylbenzene (PSDVB). The mobile phase is usually water with an organic modifier (methanol, acetonitrile, or tetrahydrofuran). Ternary or quaternary mixtures of these components can be used to improve selectivity.

There is controversy over the mechanism that controls reversed-phase separations. The three basic models are the partitioning model, the solvophobic effect, and the adsorption model. It is also necessary to consider the possibility that a mixed mechanism might exist.

In the partitioning mechanism it is accepted that the retention of a solute in reversed-phase liquid chromatography is due to its partitioning between the mobile phase and the stationary phase. This mechanism assumes that the stationary phase behaves in a similar way to a liquid, a situation that may occur in certain conditions, such as when working with monomeric phases and high percentages of water in the mobile phase, and the size of the hydrocarbonated chains is large enough for them to be associated. The adsorption mechanism accepts that retention takes place through the interaction of the solute to the organic molecules bonded to the surface of the packing matrix. The solvophobic effect theory accepts that the solute is expelled from the mobile phase towards the stationary phase.

Ion Exchange Chromatography

This is used for the analysis of electrically charged substances in working conditions. Stationary phases with electrically charged functional groups are used for this method of chromatography. The support can be silica or a polymer. The mobile phase contains a buffer which controls the loading of the solutes, and a salt that competes with the solutes in their interaction with the charges of the stationary phase. The charges permanently bonded to the chromatographic support are compensated by counterions, that is to say, by free ions with the opposite charge. Passage through the column of a sample or of the mobile phase with ions of the same sign as the counterions of the column may result in the displacement of the counterions and the retention of the ions of the sample or of the stationary phase.

The ion exchangers can be classified as cationics and anionics. Within each group it is possible to differentiate strong and weak exchangers. Strong cation exchangers (SCX) usually contain sulfonic groups, while weak ones have carboxylic groups. Quaternary ammonium salts are the most usual functional groups in strong anion exchangers (SAX), and primary amines are used in weak anion exchangers.

Size Exclusion Chromatography

This separation method is based on the size of the molecules. For many years it was known as gel permeation chromatography (GPC) or gel filtration chromatography (GFC). The retention of the molecules depends on their size: the larger molecules that do not enter the pores of the packing elute first, while the smaller molecules, whose diameter makes it possible for them to enter and exit the pores of the chromatographic packing, take a longer time.

Ion Pair Chromatography

This chromatographic method enables the separation of solutes that are ionized in separation conditions. On most occasions the separations are carried out on the same columns as those used in reversed-phase separations without the formation of ion pairs. The mobile phase is made up of a water buffer and an organic modifier, to which mixture is added a counterion with an opposite charge to that of the solute. Quaternary ammonium salts are often used as counterions for the analysis of anion substances, while *n*-alkyl sulfate salts are among the most commonly used for the analysis of cation substances.

Affinity Chromatography

This is based on the specific interaction between a molecule, or group of molecules, and a ligand, which is a molecule that is attached to the stationary phase to interact with those of the solute. The growing importance of affinity chromatography is due to the development of bioaffinity chromatography in which the interactions between ligands and molecules are based on recognitions of a biological type.

Chiral Separations

These can be approached mainly in two ways; by transforming enantiomers into chromatographically separable diasteroisomers, which does not necessitate the use of chiral phases, or by preparing chiral phases. Chiral phases can be obtained by covering the packings by making a solution pass through the column, or by ionic or covalent bonding of the chiral reagent with the column packing material.

Instrumental Configurations

The basic equipment consists of a column packed with a stationary phase, a driving force to propel the solvent through the column (pump), a system (injector) for introducing the sample on to the column, a system (detector) for measuring a physical property of the solutes being analyzed that differs from the properties of the solvent of a property of the mobile phase which is altered by the presence of the solute, and a system for recording the detector signals and converting them into graphic traces or chromatograms.

A single solvent is often used to carry out the separation (isocratic elution), but differing proportions of various solvents are also often used (gradient elution), in which case a gradient device is needed. A variety of accessories, such as pressure controllers, valves for switching solvents, valves for switching column, and ovens for heating the columns, are also commonly employed. Today most chromatographs are controlled by a computer, which is also used for data collection. This provides greater quality of quantitative data and enables automation of the system.

Solvents

The nature of the solvent will depend upon the mode of chromatographic separation employed, but a series of precautions that are common to all types of HPLC must be taken when preparing solvents. Because the columns have frits at the ends to hold the packing in place, the solvents must be devoid of particles and consequently must be filtered through membranes with a pore size of 0.5 μm or smaller, prior to use. Bubble formation must also be avoided, since bubbles may cause variations in the flow rate if they reach the pump or perturbations in the chromatogram if they reach or form in the detector cell. Solvents must therefore be degassed by immersing the bottle containing the solvent in an ultrasonic bath or by flushing the solvent with a stream of helium or nitrogen before delivering it to the chromatograph. A small stream of helium is commonly bubbled through the solvent during the chromatographic procedure to prevent uptake of air. To prevent bubble formation in the detector cell with depressurization, a restrictor is sometimes attached to the outlet of the detector cell.

Pumps

The pump is the system for delivering the solvent from the solvent reservoir to the column through the injector. Basically, two types of pumps are available: constant-pressure pumps and constant-flow-rate pumps. The latter are more frequently used in HPLC.

Constant-pressure pumps are less expensive and easy to operate, but the flow rate may vary with changes in the viscosity of the mobile phase caused by temperature fluctuations or by the accumulation of undissolved sample components in the column. These variations in flow rate affect retention times, and they may also affect resolution, increasing the difficulty of both qualitative and quantitative analysis.

Constant-flow-rate pumps afford the advantage of maintaining retention times irrespective of changes in solvent viscosity. This type of pump includes syringe pumps, which consist of a cylinder containing the mobile phase, which is expelled by a piston. The piston is driven by a motor, so as to supply a constant flow devoid of pulses. This type of pump can achieve relatively high pressures, but maintenance and changing solvents are complicated.

Reciprocating pumps, also a type of constant-flow-rate pump, are most commonly employed. Their price varies depending on their complexity of design, and their main drawback is that they generate pulses that may cause noise in the detector. Single-piston pumps are the least expensive and have a rotating eccentric cam, which drives the plunger, discharging the liquid through a one-way valve. Duplex pumps have two plungers driven by a single motor by means of a shared cam. This arrangement means that, while one of the plungers is in the discharge phase, the other is in the intake phase, thereby superimposing the two flow-rate profiles and considerably reducing pulsation. In this type of pump, delivery from the solvent reservoir is steepled and changing solvents is rapid.

Injectors

Introducing the sample on to the column is one of the most critical steps in HPLC. Ideally, the sample should reach the column in the form of a tiny droplet that does not undergo diffusion, which would broaden the chromatographic bandwidth and thus lower resolution.

Several methods are employed to deliver the sample on to the column. In on-column injectors the

sample is introduced through a syringe which traverses a septum and enables the desired quantity of sample to be deposited at the column inlet. In this method the mobile phase flows continuously through the column. The stop-flow injector is a variation of this type of injector. In this method the pump is stopped before the syringe is inserted in the injector, and injection is effected when column pressure has dropped to atmospheric pressure. The advantage of these injectors is that they are inexpensive and of simple construction, yet they do not lower efficiency. However, reproducibility is poor, they are not suitable for high working pressures, and they are complicated to operate.

Valve injectors are most commonly used. In these injectors the sample is delivered on to a pressurized column with no appreciable interruption in flow. The sample is deposited by a syringe into an external loop, the valve selector is turned, and the mobile phase passes through the loop on its way to the column. The valve thus has two positions, a load position and an injection position. High injection reproducibility can be achieved with these injectors. The band-broadening effect is comparable to or somewhat higher than that obtained using on-column syringe injectors. The drawback afforded by these injectors is that there is a split-second interruption in mobile-phase flow that may damage the column. To avoid this problem, a bypass may be attached, such that there is always a constant flow of mobile phase from the pump to the column.

Automatic injectors that are capable of running analyses on up to 100 samples without operator attention have been designed on the basis of the same mechanism as valve injectors. There are also injectors equipped with a system of valves connecting them to several columns, which enables columns to be switched without stopping the flow.

Columns

The columns most commonly utilized in HPLC consist of stainless-steel, plastic, or glass tubes, measuring 15–25 cm in length and packed with small-diameter particles (3–20 μm). Internal column diameter is normally between 2 and 5 mm.

The use of small-internal-diameter (microbore) columns has become increasingly important from the early 1980s; they are similar to the columns described above, but their internal diameter is between 0.5 and 2 mm, while they range in length from 10 to 25 cm. Packing particle size normally ranges from 3 to 5 μm. These columns are suitable for use when only small samples are available or when low solvent consumption is required.

The length of both conventional and microbore columns is mainly limited by the pressure needed to drive the solvent through the column, which is inversely proportional to the size of the particles used in the packing. Minimum column diameter is limited by column-wall effects that cause solute molecules flowing next to the wall to move more slowly than those flowing through the center of the column, resulting in an increase in chromatographic bandwidth.

Open capillary columns similar to those employed in gas chromatography are the most recent to have come into use. The column tube is made of glass and is 20–50 μm in diameter and several meters long. The stationary phase is chemically bonded to the wall of the tube. Capillary columns are also prepared by packing a tube with particles 5–30 μm in size and subsequently heating and drawing the tube to an internal diameter of 50–125 μm.

In addition to the column on which the separation is carried out, two other types of column are also utilized in HPLC to protect the analytical column. These are precolumns and guard columns. Precolumns are placed between the pump and the injector to saturate the mobile phase with the stationary phase and thus prevent dissolution of the stationary phase in the analytical column. Guard columns are placed between the injector and the main column in order to retain components in the sample that might otherwise become permanently adsorbed on the analytical column, thereby affecting column efficiency and permeability. Both precolumns and guard columns are normally made of a material similar to that used in the analytical column.

Chromatography is customarily performed at ambient temperature, but it may be necessary to regulate the column temperature or to carry out the separation at a temperature other than room temperature. In such cases thermostat-equipped compartments (ovens) and systems for heating or cooling the columns are required.

Detectors

In order to be suitable for use in HPLC, detectors must meet a number of requirements. First and foremost, detector design must prevent broadening of chromatographic bandwidth to insure that the separations achieved on the column do not deteriorate in the detector. In addition, response time must be short and the response must be linear over a sufficiently broad range of concentrations.

The detectors most frequently employed are the refractive index detector, the photometric detector, and the fluorescence detector.

Refractive index detectors measure the difference between the refractive index of the mobile phase and that of the column eluate. They are universal detectors that are highly sensitive to small changes in the mobile phase and even to small variations in temperature or pressure. This sensitivity means that to achieve a suitable signal-to-noise ratio they are only capable of detecting solute concentrations in the order of micromoles. In addition, they are unsuitable for working with gradient conditions.

Photometric detectors measure absorbance in the ultraviolet (UV) or visible regions of all the components in the column eluate. They are less universal than refractive index detectors but by the same token are more specific. This type of detector is normally capable of detecting nanomoles, provided that the compound contains a strong chromophore. The three types of photometric detector most frequently used are fixed-wavelength detectors, variable-wavelength detectors, and diode array detectors. This last type of detector is capable of performing complete spectral analysis of the column eluate on a continuous basis, i.e., without stopping the flow.

Fluorimeter detectors are more specific and more sensitive than photometric detectors, but their linear range is smaller. Detection limits are in the order of picomoles for suitably fluorescent compounds, and they are very useful in trace component analysis.

Electrochemical detectors are also widely employed in HPLC. These come in two types: amperometric and conductometric detectors. Amperometric detectors are highly sensitive but are only applicable to analyses that can be oxidized or reduced; conductometric detectors are moderately sensitive and are applicable for detecting anions and cations. This type of detector is the detector most commonly used in ion exchange chromatography.

Derivatization of the components being analyzed may sometimes be employed to increase the detection limit or specificity.

Mass spectrometry (MS) is being used more and more as an online detection system in HPLC. The use of HPLC-MS coupling has spread since 1980 due to the improvements introduced in the different types of interface used. Several coupled MS-HPLC systems are available commercially, and improvements are continually being introduced in existing equipment. At the same time, new equipment is appearing all the time. MS has been considered to be the ideal detector, since it furnishes information on component structure. Microbore HPLC is particularly useful for online HPLC-MS, which requires low sample volumes. MS in combination with HPLC may become a standard technique in a few years' time. (*See* **Mass Spectrometry**: Principles and Instrumentation.)

Selected Applications

The use of HPLC in food analysis is growing daily, and it is now routinely applied in many laboratories. The many different types of columns and detectors that are currently available commercially make it possible to apply HPLC in analyzing nearly all the nonvolatile components in foods, be they present naturally or added artificially. The techniques employed for certain groups of food components are summarized below by way of example.

Carbohydrates

Nearly all chromatographic modes may be used in separating carbohydrates. For example, ion exchange on strongly or weakly basic anionic resins or on cationic resins, and partition chromatography on ion exchange resins, on chemically bonded cyano, amino, propyl-amino, or combined amino-cyano phases, or on silica gel or gel permeation are all possible. Differential refractometry is the conventional detection system, although direct detection, using short UV wavelengths or by forming derivatives detectable at longer wavelengths or fluorescent derivatives, is also employed. (*See* **Carbohydrates**: Determination.)

Acids

A variety of chromatographic modes are also applied to this group of components. Certain workers have employed ion exchange chromatography on strongly acid cationic resins or strongly or weakly basic anionic resins. Reversed-phase chromatography and ion pair chromatography have also been used. Detection is carried out by refractometry, photometry using UV, or visible wavelengths, as in the case of carbohydrates. (*See* **Acids**: Properties and Determination.)

Amino Acids and Amines

Most separations of amino acids and amines are performed using reversed-phase chromatography of dansyl chloride, orthophthaldialdehyde (OPA), or phenyl dithioisocyanate derivatives. 9-Fluorenylmethyl chloroformate (FMOC) is a suitable reagent for the analysis of secondary amino acids. Detection is carried out by means of fluorescence or UV absorption. The derivatizing, highly fluorescent, reagent 6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate (AQC) reacts rapidly and easily with both primary and secondary amines and its use is currently spreading. (*See* **Amines**; **Amino Acids**: Determination.)

Peptides

Diverse techniques are used to separate peptides on account of the broad range of molecular weights

of these components. Conventional reversed-phase columns are applied for peptides with molecular weights of less than 3 kDa, while reversed-phase columns packed with particles with large pore sizes or size exclusion columns are used for larger peptides. Detection is performed at 214 nm. Figure 1 shows a chromatogram of the peptides with molecular weights higher and lower than 700 from a sparkling wine, as an example of this type of analysis. (*See* **Peptides**.)

Proteins

Practically all known modes have been used in the separation of proteins, e.g., separations based on molecule size (gel permeation chromatography), on charge (ion exchange chromatography), on hydrophobicity (reversed-phase chromatography and hydrophobic–interaction chromatography), and even combinations of these mechanisms. The combination of HPLC and MS analysis is commonly used for the sequence analysis of glycoproteins. Detection is carried out at 214 or 280 nm. (*See* **Protein**: Determination and Characterization.)

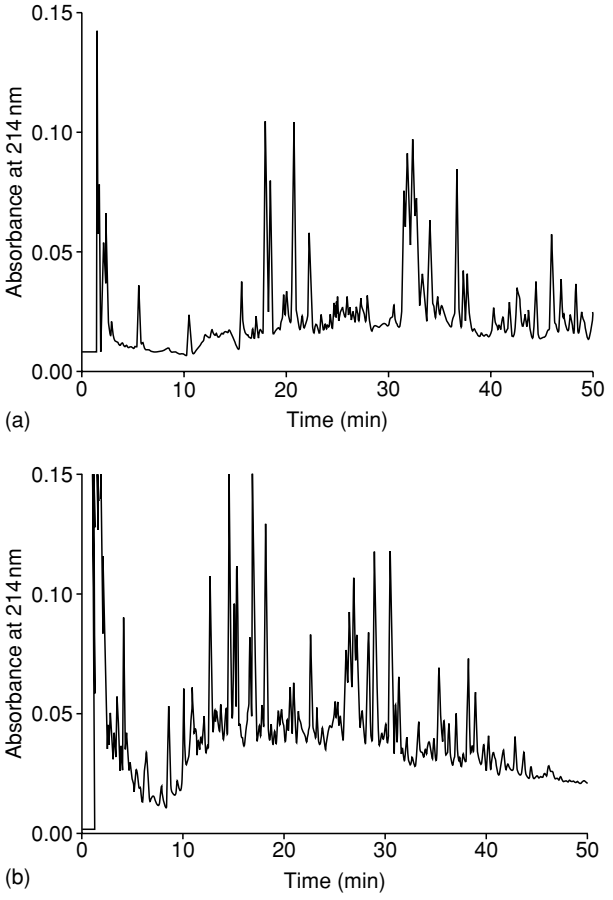

Figure 1 Chromatogram of peptides with molecular weights (a) higher and (b) lower than 700 Da from a sparkling wine. Column: Nova Pak C_{18}. Solvents: (a) TFA/water; (b) TFA/acetonitrile. Gradient elution. Detection at 214 nm. Reproduced from Moreno-Arribas MV, Bartolomé B, Pueyo E and Polo MC (1998) Isolation and characterization of individual peptides from wine. *Journal of Agriculture and Food Chemistry* 46: 3422–3425, with permission.

Other Compounds

In addition to the major groups of food components mentioned above, HPLC has found application in many other areas of food analysis, and details may be found in the relevant articles for the following compounds or groups of compounds: lipid components, phospholipids, triglycerides, vitamins, colors, pesticides, drug residues, polycyclic aromatic hydrocarbons, and nitrosamines. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments; Properties and Determinants of Synthetic Pigments; **Fatty Acids**: Analysis; **Nitrosamines**; **Pesticides and Herbicides**: Types, Uses, and Determination of Herbicides; **Phospholipids**: Determination; **Polycyclic Aromatic Hydrocarbons**; **Triglycerides**: Characterization and Determination; **Vitamins**: Determination.)

Related Techniques

FPLC designates a fast chromatographic method, developed by Pharmacia, which is similar to HPLC and yields high resolution. Since FPLC needs only a relatively low backpressure to drive the high flow rates at which the separations are performed, the risk of denaturation caused by shearing forces is reduced. Moreover, the mechanical components are resistant to corrosive buffers, and there is no contamination or inactivation or the components of interest.

Given the range of columns available in the market, a variety of separation modes can be applied using this technique: size exclusion, hydrophobic interaction, chromatofocusing, ion exchange, and reversed-phase chromatography.

This method was developed to separate and purify biomolecules and is very useful in separating isoenzymes and molecular species with similar charge characteristics. It is also used to distinguish between different types of meat or grains.

SFC is another technique, related to HPLC, which uses as the mobile phase a supercritical fluid, i.e., a fluid at a pressure and temperature above the critical point. The properties of supercritical fluids are intermediate between those of gases and those of liquid. Thanks to their higher diffusivity and lower viscosity as compared to liquids, high efficiencies are achievable with shorter analysis times than those customarily employed using HPLC.

The basic advantage of SFC with respect to gas chromatography (GC) is the possibility of analyzing

components that span a broad range of volatilities as well as heat-labile components. At the same time, SFC is also compatible with many of the detectors commonly used in GC or HPLC and SFC-MS coupling is easy to carry out.

The critical temperature of the supercritical fluids employed as the mobile phases between 0 °C and 200 °C and the critical pressure of these fluids should not be too high. Carbon dioxide, nitrous oxide, alkanes (such as *n*-pentane), and xenon, all of which are nonpolar, are most often used. Ammonia can be used to elute polar solutes, but mixtures of phases, e.g., a nonpolar mobile phase containing a small quantity of a polar organic solvent, known as a modifier are normally employed.

SFC can be performed on capillary, packed, and micropacked columns. Stationary phases should be cross-linked; otherwise the supercritical fluids, which are excellent solvents for polymers, could extract the stationary phase. Enantiomers can be resolved using chiral phases.

The equipment used in SFC is similar to that used in HPLC and basically consists of a high-pressure syringe pump, an injector, and a restrictor or post-column valve to keep the mobile phase in a supercritical condition inside the chromatographic column.

Fluid density is commonly programmed to adjust mobile-phase selectivity, in as much as the physicochemical properties of supercritical fluids (solvation strength, viscosity, diffusion) are all dependent upon density.

This method has been applied in food analysis (oils, cheeses, coffee, etc.) Some of the most interesting applications include separations of acids, alcohols, lipids, carbohydrates, vitamins, and terpenes. (*See* **Acids:** Properties and Determination; **Alcohol:** Properties and Determination; **Analysis of Food; Carbohydrates:** Determination; **Coffee:** Analysis of Coffee Products; **Vitamins:** Determination.)

See also: **Acids**: Properties and Determination; **Alcohol**: Properties and Determination; **Amino Acids**: Determination; **Chromatography**: Principles; **Coffee**: Analysis of Coffee Products; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; Properties and Determinants of Synthetic Pigments; **Fatty Acids**: Analysis; **Nitrosamines**; **Peptides**; **Phospholipids**: Determination; **Polycyclic Aromatic Hydrocarbons**; **Protein**: Determination and Characterization; **Triglycerides**: Characterization and Determination; **Vitamins**: Determination

Further Reading

Charalambous G (1984) *Analysis of Foods and Beverages. Modern Techniques*. Orlando: Academic Press.

Dabrio MV, Blanch GP, Cifuentes A *et al.* (2000) *Cromatografia y Electroforesis en Columna*. Barcelona: Springer-Verlag Ibérica.

Gonzalez-Llano D, Polo C and Ramos M (1990) Update on HPLC and FPLC analysis of nitrogen compounds in dairy products. *Lait* 70: 255–277.

Gruewedel D and Whitaker JR (1987) *Food Analysis. Principles and Techniques*. New York: Marcel Dekker.

Henschen A, Hupe KP, Lottspeich F and Voelter W (1985) *High Performance Liquid Chromatography in Biochemistry*. Weinheim: VCH Press.

Macrae R (1987) *HPLC in Food Analysis*, 2nd edn. London: Academic Press.

Moreno-Arribas MV, Bartolomé B, Pueyo E and Polo MC (1998) Isolation and characterization of individual peptides from wine. *Journal of Agriculture and Food Chemistry* 46: 3422–3425.

Oliver RWA (1989) *HPLC of Macromolecules. A Practical Approach*. Oxford: IRL Press.

Shaw LP (1988) *Handbook of Sugar Separations in Foods by HPLC*. Boca Raton. FL: CRC Press.

Smith RM (1988) *Supercritical Fluid Chromatography*. RSC Chromatography Monographs. Cambridge: Royal Society of Chemistry.

Gas Chromatography

I Martínez-Castro, J Sanz and M V Dabrio, Instituto de Química Orgánica General (CSIC), Madrid, Spain

Background

Gas chromatography was born in 1951, when James and Martin developed a new separation technique in order to analyze a mixture of 17 fatty acids. This technique includes different chromatographic modes that use a gas as the mobile phase: the separation process takes place in a chromatographic column. Like other separation techniques, gas chromatography (GC) can be employed on the preparative scale, but more frequently, it is used as a powerful analytical technique.

Instrument

The analytical instrument consists of a mobile phase supply (usually a gas cylinder) regulated by a control system, a sample introduction system (usually called injector), a thermostatically controlled oven

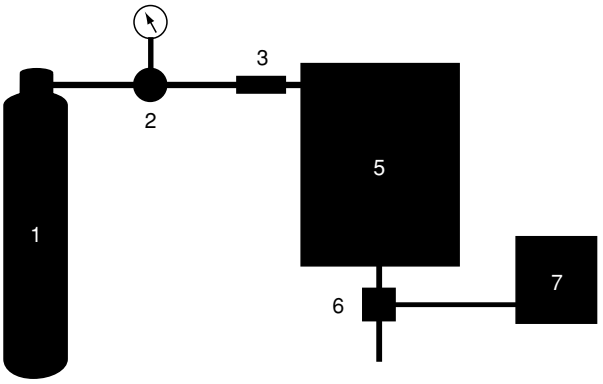

Figure 1 Scheme of a gas chromatograph. 1, gas cylinder; 2, pressure control system; 3, injector; 4, oven; 5, column; 6, detector; 7, recorder/data system.

containing the column, a detector, and a data-acquisition system (Figure 1).

Mobile Phase

The separation process is mainly based on the interaction of the analytes with the stationary phase; the mobile phase is called the 'carrier gas,' since its role is merely to transport the analytes through the column. Helium, nitrogen, hydrogen, and argon are usually employed as mobile phases. High-purity gases are necessary because traces of impurities, such as oxygen or water, can spoil the stationary phase or disturb the detector response. Helium is recommended to achieve faster separations, and nitrogen is a less expensive alternative, whereas the use of hydrogen requires special care.

Carrier gas pressure is between 5 and 700 kPa at the inlet (p_i), while the outlet pressure (p_o) is usually ambient pressure. The density, pressure, and velocity (u) of the carrier gas vary nonlinearly through the column, and it is necessary to specify at which column point they are measured or calculated. A factor that corrects for the mobile phase compressibility in the column can be calculated as:

$$j = \frac{3}{2}\frac{(p_i/p_o)^2 - 1}{(p_i/p_o)^3 - 1}. \quad (1)$$

The average linear carrier gas velocity can be obtained by dividing the column length (L) by the retention time of an unretained compound:

$$\overline{u} = L/t_M \quad (2)$$

t_M is called the 'hold-up time.'

The flow rate at the column outlet (F_c) can be easily measured and related to the average flow rate through $\overline{F} = jF_c$. It is also related to linear velocity:

$$\overline{F} = ku_o = k\frac{\overline{u}}{j}. \quad (3)$$

The retention time of a compound is defined as the length of time between sample injection and the emergence of the compound peak maximum. It is related to the volume of mobile phase entering the column during this time (V_R, retention volume) by the following equation:

$$t_R = V_R/F_c. \quad (4)$$

The net retention volume, V_N, takes into account the compression correction factor and the hold-up time:

$$V_N = jF_c(t_R - t_M). \quad (5)$$

The specific retention volume at the column temperature (V_g^θ) is the net retention volume per gram of stationary phase (W_S):

$$V_g^\theta = V_N/W_S. \quad (6)$$

The specific retention volume at 0 °C is:

$$V_g = \frac{V_N}{W_S}\frac{273.17}{T_c}. \quad (7)$$

The variation of efficiency with the carrier-gas velocity is described by the Van Deemter equation:

$$H = A + \frac{B}{u} + (C_S + C_M)u, \quad (8)$$

where C_S and C_M, respectively, represent the contribution of the stationary and mobile phases to the zone spreading caused by the resistance to mass transfer and other diffusion phenomena. When the column is an open capillary, Golay's equation can be used:

$$H = \frac{B}{u} + (C_S + C_M)u. \quad (9)$$

In most cases, C_S is negligible when compared with C_M.

Stationary Phase

Stationary phase type (solid or liquid) determines the two main gas chromatographic modes: gas–solid chromatography (GSC) and gas–liquid chromatography (GLC). System selectivity, which depends on the compounds to be separated and on the stationary phase, is estimated by the separation factor, α:

$$\alpha = V_{N2}/V_{N1} = (t_{R2} - t_M)/(t_{R1} - t_M). \quad (10)$$

Solid Phases

These consist of adsorbent particles; the separation mechanism is mainly physical adsorption. Since

sorption/desorption processes are usually faster than those of solution/vaporization (occurring in GLC), transfer kinetic constants are higher, and C_S values are lower in GSC, allowing better efficiencies. Selectivities can be high enough to separate stereoisomers and isotopes. Other characteristics of GSC are a wide temperature operation range, isotherms with a narrow linearity range, and high retention values. These characteristics make GSC the technique of choice for the separation of permanent gases and small polar molecules. Solid phases are formed by small particles of uniform size, with a high specific surface area. Commercial solid phases are based on silica (Spherosil, Porasil), graphitized carbon (Carbopack, Graphpac), or porous organic polymers (Porapak, Chromosorb Century series, Tenax).

Liquid Phases

These are usually inert and thermally stable polymeric molecules. When the stationary phase is a liquid, the main separation mechanism is partition, although other mechanisms like chirality, mesomerism, and complex formation may also contribute.

Partition Compounds dissolved in the liquid phase are in equilibrium (Henry's law) with the vapor; the elution order may be proportional to the boiling points or very different, depending on the activity coefficients.

Chiral interactions Enantiomer pairs always coelute when chromatographed on achiral phases. However, liquid phases possessing chiral groups can show a greater interaction with one of the pair members, which is then relatively more retained. As chiral interaction energies are small, α values are usually close to 1, and a high efficiency is required, such as that provided by capillary columns. Chiral phases are based on peptides, amide-substituted polysiloxanes, or substituted cyclodextrins.

Interactions with mesophases (liquid crystals) These are formed by elongated or planar molecules. In a temperature range from the melting point to the anisotropic transition, they are liquids ordered in a mono- or bidimensional orientation. These phases present a high selectivity for long or planar isomers.

Complex formation Organometallic compounds possessing metal ions are able to interact selectively with solutes having electron-donor groups.

Characteristics of liquid phases The liquids used as stationary phases have to be thermally stable over a wide temperature range. The lower limit is marked by the melting point, and the so-called maximum allowable operation temperature (MAOT) is determined by the increase in the detector signal caused either by an appreciable value of vapor pressure or by thermal degradation of the phase: in both cases, some molecules from the phase arrive continuously at the detector. The most important characteristic of a stationary phase is its polarity, which is usually defined by various empirical scales, the most popular being that of McReynolds, which uses as probes 10 solutes with different polar groups and a series of saturated, linear hydrocarbons. The molecular interactions are supposed to be additive and are described by several parameters (McReynolds constants) deduced from the retention observed in the studied phase and in squalane, which was defined as the zero on this polarity scale. Five McReynolds constants, which appear in most commercial catalogs, are directly related to phase polarity. Cross-linking improves the phase stability.

Table 1 shows some characteristics of the most commonly used liquid phases. Hydrocarbons (like squalane or apolane) are used to obtain reference data. Polysiloxanes (with methyl, phenyl, trifluoropropyl, or cyanoalkyl substituents, showing a wide polarity range) are commonly called 'silicones' and are nowadays the most popular liquid phases, since they have a good thermal stability and high permeability to solutes. Polyethylene glycols with different chain lengths (from a molecular weight of 150 to around 4.1×10^6) are another set of highly used phases.

Columns

Columns are usually classified as open or packed. Figure 2 shows a cross-section of the main types of columns.

In open columns, the stationary phase is distributed on the inner wall, and the mobile phase circulates through a central channel. Packed columns consist of a glass or metal tube packed with solid particles (100–250 μm), whose surface adsorbs the analyzed solutes (GSC) or is covered with a thin film of liquid stationary phase (GLC). Table 2 lists the main characteristics of the analytical columns.

Packed columns are mainly used for permanent gases analysis by GSC; they are cheap, robust, and easy to handle. The most frequently used GLC columns are open tubular and are usually referred to as 'capillary' since their diameter is very small. Since the system efficiency depends on the dimensions of the analytical columns, they should be chosen in order to achieve the desired resolution in the minimum time. Efficiency is always directly related to

Table 1 Common liquid stationary phases

Composition	*Commercial names*	*McReynolds constants (120 °C)*					*Operating temperature range (°C)*
		X	Y	Z	U	S	
2,6,10,15,19,23-Hexamethyl-tetracosane	Squalane	0	0	0	0	0	20/120
24,24-Diethyl-19,29-dioctadecyl-heptatetracontane	Apolane-87	21	10	3	12	25	35/260
Poly(dimethylsiloxane) 100%	OV-1, SE-30, OV-101, SP-2100, CP-Sil 5CB, DB-1, SPB-1, BP-1	16	55	44	65	42	30/330
Poly (94% methyl, 5% phenyl, 1% vinyl siloxane)	SE-54, DB-5, CP-Sil 8CB, SPB-5, BP-5, AT-5	33	72	66	98	67	50/300
Poly (86% methyl, 7% phenyl, 7% cyanopropyl siloxane)	OV-1701, CP-Sil 19CB, BP-10, DB-1701, HP-17	82	170	157	236	160	30/250
Poly (50% methyl, 50% phenyl siloxane)	OV-17, DB-17, SP-2250, HP-50, SPB-50	119	158	162	243	202	20/350
Poly (50% methyl, 50% trifluoropropyl siloxane)	OV-210, SP-2401, DB-210	146	238	358	468	310	20/275
Poly (50% methyl, 25% phenyl, 25% cyanopropyl siloxane)	OV-225, XE-60, DB-225, CP-Sil 43CB, BP-15	228	369	338	492	386	20/250
Poly (ethylene glycol)	Carbowax, Superox, Supelcowax, DB-Wax, CP-Wax 52CB, BP-20	322	536	368	572	510	60/225
Poly (ethylene glycol) modified with nitroterephthalic acid	FFAP, SP-1000, AT-1000, OV-351, Nukol, CP-Wax 58CB, BP-21	340	580	597	602	627	50/250
Poly(biscyanopropyl siloxane)	OV-275, CP-Sil 88, SP-2340, Silar 10C	629	872	763	1106	849	30/250

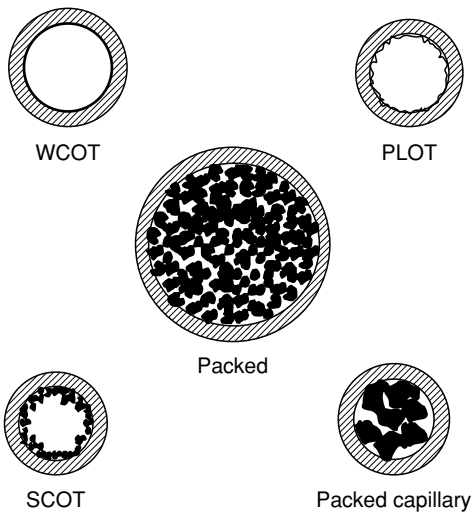

Figure 2 Cross-section of columns. PLOT, porous layer open tubular; SCOT, support-coated open tubular; WCOT, wall-coated open tubular. Adapted from Dabrio *et al*. (2000) *Cromatografia y electroforesis en columna*, p. 122. Barcelona: Springer with permission.

column length, as shown in eqn (2). The efficiency of a packed column is mainly related to the particle diameter, which appears in two terms (A and C_M) of the van Deemter equation (eqn (8)). Column diameter has a large influence on the efficiency of capillary columns, as Golay's equation (eqn (9)) shows. Open tubular capillary columns require more sophisticated equipment than packed columns, but their efficiency is very high, since it is possible to use very long columns with a moderate pressure drop (between 0.1 and 7 kg cm^{-2}).

Capillary columns with the inner wall covered by a porous layer open tubular (PLOT, Figure 2) or by a support-coated open tubular layer (SCOT, Figure 2), where the stationary phase is coated, have both a high sample load capacity and a high efficiency.

Operating Conditions

The flow rate and column temperature can be easily modified, enabling the best conditions to be selected for an adequate separation.

The *flow rate* is usually controlled through either a pressure regulator or a flow controller. According to eqn (4), t_R is inversely proportional to the flow rate for a given compound. A linear velocity of carrier gas close to the van Deemter optimum (eqn (8)) affords the best efficiency, but higher flow rates reduce the analysis time.

Temperature has a decisive influence on retention. When the temperature increases, V_R and t_R decrease (eqn (7)). The temperature operation should be selected in such a way that the retention factor (k) values lie between 1 and 15. When a mixture contains

Table 2 Some characteristics of GC columns

Type	*Name*	*Internal diameter (mm)*	*Length (m)*	d_f (μm)	*Particle size (μm)*
Open	WCOT (Wall-coated open tubular)	0.05–0.75	10–100	0.1–0.5	–
	PLOT (Porous layer open tubular)	0.5–1	10–100	0.1–5[a]	–
	SCOT (Support-coated open tubular)	0.5–1	10–100	0.1–5[a]	–
Packed	Micropacked	0.8–1	2–7	5–15[b]	80–200
	Packed	2–6	2–7	5–15[b]	80–500

[a]Bed thickness (particles + liquid phase).
[b]percentage of liquid phase (referred to support + phase).

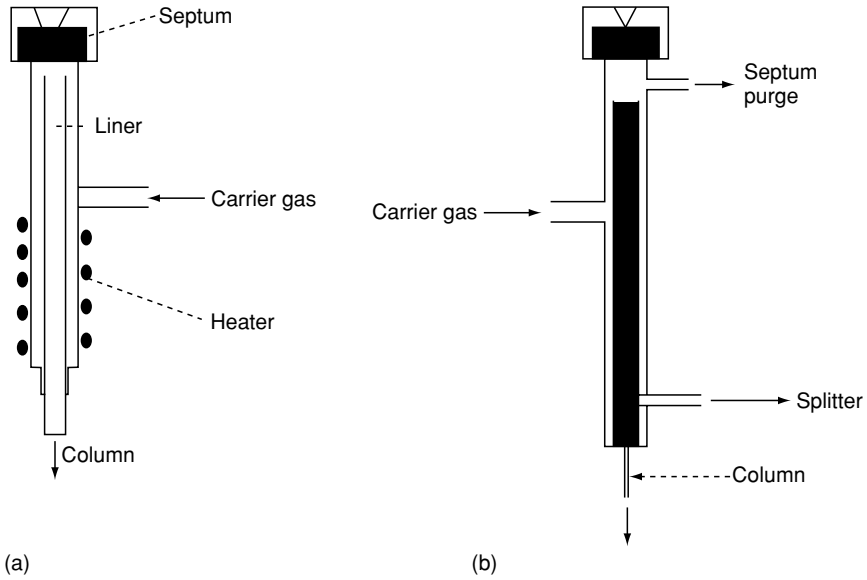

Figure 3 Scheme of an injector. a) for direct injection; b) for split/splitless injection. Adapted from Dabrio MV *et al*. (2000) *Cromatografia y electroforesis en columna*, p. 130. Barcelona: Springer with permission.

compounds with very different volatilities, it is not possible to find an optimum temperature for separating all components. Since the oven temperature is easily controlled, it can be changed during the chromatographic run ('programmed') in such a way that each compound traverses the column within an optimum temperature range.

Sample Introduction

The objective of injection is to introduce the sample as a narrow band, whose composition should correspond closely to that of the original mixture. No universal system is available to handle all the different samples encountered. Permanent gases may be introduced with gas-tight syringes or using multiport valves. Liquid samples and solutions are usually introduced by microsyringe at the injection port, which is a heated chamber where the carrier gas enters under controlled flow or pressure. The sample is vaporized and transferred to the column inlet by the carrier gas stream. A schematic diagram of an injector is shown in Figure 3.

The chamber is provided with an inert glass or quartz liner to minimize sample contact with hot metallic surfaces. It is closed with a mechanical seal or, more usually, with an elastomer membrane called an septum, with can be pierced with a microsyringe to introduce the sample. Depending on the column dimensions, the quantity of sample can be varied from nanoliters to a few microliters, containing nanograms or micrograms of analytes. There are special devices capable of transferring several hundreds of microliters of very diluted samples, the solvent excess being discarded before it enters the analytical column.

Injection Modes

There are several different ways of injecting samples. The more common modes are detailed below.

Direct injection The sample is introduced through the septum in the hot injector, where the liquid

rapidly vaporizes; carrier gas sweeps the vapor into the column inlet.

Split injection This was designed for the introduction of samples into capillary columns. The basic design used for direct injection was modified as shown in Figure 3. The carrier gas, controlled by a pressure regulator or a combination of a flow controller and a back-pressure regulator, is divided into two streams: one is used to eliminate the possible impurities from the septum membrane; the other enters the vaporization chamber. This gas stream, after sweeping the vaporized sample, is split before it enters the column inlet to attain a high flow rate. The sample is also proportionally divided, so only a small part reaches the column. Since quantitative analysis presents some problems related to the split process, the use of internal standards is recommended.

Splitless injection When very dilute solutions need to be injected, the amount of sample entering the column can be increased, stopping the split process by using a solenoid valve. The split valve is closed, the sample is introduced, and after a controlled delay (usually 15–60 s), the split valve is opened again. Most of the sample enters the column while the split is closed, while the rest is purged out when the valve opens. Since the sample enters the column as a broad band, it is a necessary focusing step. Usually, the column is kept at a low temperature during injection, and sample vapors condense at the column inlet in a narrow band (*cryofocusing*); next, the temperature is raised in order to elute the solutes correctly. Sample solvent condensation can also be used (*solvent effect*) to retain the solutes as a narrow band at the column inlet.

Programmed temperature vaporizer (PTV) With the same basic design of a split/splitless detector, it can be heated and cooled very fast. The injector is cold when the sample is introduced and deposited on inert packing; then, the injector is heated to the desired temperature so that the vaporization is homogeneous, and sample discrimination is reduced.

On-column injection A liquid sample is deposited directly at the column head, without previous vaporization. The injectors are very simple, having a mechanical stop valve instead of a septum, and the microsyringes usually have very thin needles (less than 0.3 mm outer diameter). This system reduces the discrimination, thermal decomposition, and degradation of labile compounds. Its main drawback is possible column contamination with nonvolatile sample impurities.

External Sampling Devices

Pyrolysis A solid sample is thermally decomposed; the volatile fragments resulting from its breakdown can be chromatographically separated and detected, and their chromatographic pattern related to that of the original sample.

Headspace sampling The sample is thermostatized in a closed vial, and the vapor is sampled after equilibrium is attained. Usually, the process is automatic, under either atmospheric or high-pressure conditions, and the technique is frequently used for aroma analysis.

Purge and trap A stream of inert gas passes over or through the sample, and the stripped volatiles are trapped on a solid sorbent. These can be recovered either by a solvent or by thermal desorption (see below). The enrichment factors obtained are very high; matrix interferences are largely reduced. This technique is useful when very low concentrations of volatile substances have to be analyzed.

Thermal desorption The injector chamber can be ballistically heated to the desired temperature while it is purged with carrier gas. When a solid matrix is placed in this chamber, the volatile substances present are released and stripped towards the column; they can then be cryofocused before arriving at the column inlet. In the *direct* mode (DTD, direct thermal desorption), the solid matrix is a sample (e.g., a spice or a herb). In the *indirect* mode, volatiles have been previously stripped from the sample and trapped on an adsorbent.

Multidimensional Gas Chromatography

This usually refers to a combination of columns with different selectivity, where a small fraction from a complex sample is preseparated on the first column and introduced into the second column for further separation. Whereas the first column may be packed or may be a capillary, the second column is always a capillary and more selective. A special pneumatic switching arrangement is necessary in order to provide a sharp sample band.

Detectors

A high number of detectors have been designed for GC. The most frequently used detectors can be classified into four types.

Thermal Conductivity Detectors (TCD)

These are universal detectors, with a medium sensitivity, and are concentration-dependent. The column effluent passes through a chamber containing a hot filament. When a compound appears in the effluent, the thermal conductivity of the mixture solute/carrier gas, and hence the heat dissipation velocity, varies; this causes a temporal variation in filament temperature and a change in its electric resistance, which is very easy to measure. The best carrier gases for this detector are hydrogen and helium. TCD is appropriate for substances such as permanent gases, and can also be coupled in tandem with other detectors.

Ionization Detectors

These induce sample ionization by using thermal, chemical, electromagnetic, or radioactive energy. Solute ions in the carrier gas produce changes in the electric field between two electrodes. This is shown schematically in Figure 4.

Flame-ionization detectors These are the most commonly used detectors in GC and are almost universal, with a high sensitivity to organic carbon-containing compounds, a wide linear range, and an excellent baseline stability. They are also very reliable, with a negligible dead volume and a fast response time. The ionization energy is supplied by a flame, obtained by a hydrogen flow, which is mixed with the column effluent. The organic compounds are burnt, producing H_2O and CO_2; the response is proportional to the solute mass flow and is similar for most organic substances. Molecules with less C—H bonds than hydrocarbons give a lower response than hydrocarbons. These detectors can be used with all carrier gases and are easy to calibrate.

Electron capture detectors (ECD) These are very specific and sensitive to compounds with a high electron affinity. The ionization source is a foil (usually ^{63}Ni or ^{3}H), which emits β-radiation. Nitrogen and an argon/methane mixture are suitable carrier gases for ECD. A low potential is established between electrodes, producing a current that decreases when a solute capable of accepting ('capture') electrons appears in the effluent. The relative response is very different for compounds that have groups with different electronic affinities: e.g., a polychlorinated compound has a response 10^6 times higher than that of a saturated hydrocarbon. Thus, it is comparatively easy to detect less than 1 pg of polyhalogenated compounds in the presence of other molecules that have only C, H, and O atoms. Since the ECD response is dependent on the sample concentration and electron affinity of each analyte, calibration using standards is necessary for quantitative analysis.

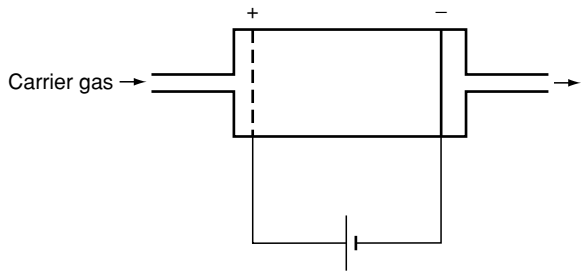

Figure 4 Scheme of an ionization detector.

Electrochemical Detectors

These are designed mainly for the environmental analysis of substances with heteroatoms like S, N, or halogens, and these detectors are very specific and sensitive.

Spectroscopic Detectors

Several spectroscopic devices can be coupled with GC in order to obtain structural information about the eluting peaks. The most commonly used detectors are those based on mass spectrometry (MS) (*See* **Chromatography**: Combined Chromatography and Mass Spectrometry; **Mass Spectrometry**: Applications; Principles and Instrumentation); and Fourier transform infrared (FTIR); *See* **Spectroscopy**: Infared and Raman.) The former have a high selectivity and sensitivity. The column effluent is continuously fed into the ionization chamber of a mass spectrometer, and a mass spectrum is recorded every few seconds. GC–MS coupling can be used in two modes: scan, in which all the ion fragments are collected and registered, giving structural information about all chromatographic peaks, and selected ion monitoring (SIM), in which only a few mass fragments are recorded, improving the signal-to-noise ratio and significantly reducing the detection limit.

Applications

GC has been used to analyze the main constituents of foods, such as lipids, carbohydrates, or amino acids, and it is especially useful in the study of minor constituents such as aromas or pollutants.

Food Constituents

Lipids GC has been the technique of choice for the study of fats. It is possible to study the degree of lipolysis by analyzing the free fatty acid (FFA) profile or the authenticity of fats and oils by determining the

glyceride fraction, the fatty acid composition, and the unsaponifiable fraction. FFAs can be studied using modified polar stationary phases, although fatty acids are usually determined as fatty acid methyl ester derivatives; these derivatives are volatile and easy to resolve on polar phases such as polyesters, polyethylene glycols and cyanoalkyl silicones. There is considerable interest in the separation of geometric isomers of unsaturated fatty acids: the use of very long capillary columns coated with cyanoalkyl silicones is the best choice. Direct analysis of triacylglycerols, initially problematic because of their low volatility, has improved since the development of fused silica capillary columns with a thin layer of stationary phase with high MAOT values; to date, very good separations have been reported, based on the number of carbon atoms and even on unsaturations. The injection of high-molecular-weight triacylglycerols requires the use of a cold-on column or PTV injection.

Carbohydrates GC is used to separate monosaccharides, disaccharides, and oligosaccharides. Although the preparation of volatile derivatives is necessary, the advantage that GC affords over HPLC is its greater sensitivity and higher resolution. The most popular derivatives are silyl ethers, oximes, methyl ethers, and acetates. Silicones ranging from methyl to cyanoalkyl are the preferred stationary phases. These methods have been used for juices, milk, molasses, honeys, and beverages.

Amino acids These have to be derivatized to achieve a sufficient volatility. The carboxylic group is usually esterified, and the amino group can be acylated or silylated.

Minor constituents Sterols are determined in order to assess food authenticity, especially with fats and oils, since the sterol composition is characteristic of each species. They can be analyzed, either as free sterols or as trimethyl silyl ethers, using packed columns (methyl or phenyl silicones), but they are better resolved on capillary columns. Naturally occurring wax esters (from C_{24} to C_{48}) have been analyzed using stable thin-film capillary columns.

Volatiles Volatile components are difficult to analyze because they are usually present as very complex mixtures at very low concentrations. Thus, the first requirement is the use of high-resolution columns: these are long, well-deactivated capillaries, coated with nonpolar (methyl or methyl vinyl silicones) or medium polar (polyethylene glycols) phases in order to elute hydrocarbons, alcohols, esters, pyrazines, aldehydes, acids, etc. High-resolution gas chromatography coupled with MS and sometimes with other spectrometric detectors like FTIR detectors allows the successful analysis of complex mixtures and the identification of many new compounds. A data-acquisition system is suitable for profile analysis and computer-based pattern recognition. A prior isolation and concentration step is usually necessary in order to suppress the matrix. Different off-line or on-line techniques can be used: these are based on extraction, static headspace, thermal desorption, and purge-and-trap devices.

Chiral compounds GC affords excellent separations of enantiomeric pairs, due to the high efficiency of capillary columns. Commercial columns are prepared with D- or L-valine *tert*-butylamide derivatives, metal complexes, or substituted cyclodextrins. They have been used to determine the racemization of amino acids, to detect adulterations in several food products by analyzing the enantiomeric purity of marker compounds, and to assess the aging or thermal processing of different foods.

Permanent gases CO_2 and SO_2, present in certain beverages or wines, can be determined using packed or SCOT columns and TCDs.

Pollutants

GC is the basic method used to detect and analyze the presence in foods of pollutant residues, such as pesticides, toxaphenes, polychlorinated biphenyls (PCBs), polycyclic aromatic hydrocarbons (PAHs), chlorinated solvents, polychlorodibenzodioxins (PCDDs), and polychlorodibenzofurans (PCDFs). Multiresidue procedures are used more frequently than specific methods. Analysis of these substances, which are present at p.p.b. or p.p.t. levels, requires several extraction, clean-up, and purification steps. High-purity solvents and reagents must be used, and glassware has to be cleaned with special care to avoid introducing new residues. High-efficiency capillary columns are recommended; this is especially true for PAHs, which can be resolved efficiently using liquid crystal phases, and for chiral PCBs. Selective detectors are required for all of these compounds. Chlorinated compounds are detected using ECDs; organophosphorus compounds are better detected using thermoionic or flame-photometer detection. MS detectors working in the SIM mode are useful for all pollutants. Those present at very low levels, such as PCDDs and PCDFs, require high-resolution MS or MS–MS systems to be selectively characterized. (*See* **Contamination of Food**; **Pesticides and Herbicides**: Residue Determination; **Pesticides and Herbicides**: Types of Pesticide.)

Other

GC allows analysis with very low detection limits of different substances such as illegal drugs used in cattle fattening, N-nitrosamines, residues in food deriving from plastic packaging, additives like antioxidants, preservatives, certain colorants, and other food ingredients

See also: **Chromatography**: Combined Chromatography and Mass Spectrometry; **Contamination of Food**; **Mass Spectrometry**: Principles and Instrumentation; Applications; **Pesticides and Herbicides**: Types of Pesticide; Residue Determination; **Spectroscopy**: Infrared and Raman

Further Reading

Biermann CJ and McGinnis GD (1989) *Analysis of Carbohydrates by GLC and MS*. Boca Raton, FL: CRC Press.
Christie WW (1989) *Gas Chromatography and Lipids*. Ayr, UK: The Oily Press.
Grob K (1986) *Classical Split and Splitless Injection in Capillary GC*. Heidelberg: Hüthig.
Guiochon G and Guillemin CL (1988) *Quantitative Gas Chromatography*. Amsterdam: Elsevier.
IUPAC (1993) Nomenclature for chromatography. *Pure and Applied Chemistry* 65: 819–872.
König WA (1992) *Gas Chromatographic Enantiomer Separation with Modified Cyclodextrins*. Heidelberg: Hüthig.
Marsili R (1997) *Techniques for Analyzing Food Aroma*. New York: Marcel Dekker.
Rotzsche H (1991) *Stationary Phases in Gas Chromatography*. Amsterdam: Elsevier.
Sandra P (1985) *Sample Introduction in Capillary Gas Chromatography*, vol. 1. Heidelberg: Hüthig.
Scott RPW (1996) *Chromatographic Detectors*. New York: Marcel Dekker.
Shibamoto T (1998) *Chromatographic Analysis of Environmental and Food Toxicants*. Amsterdam: Elsevier.
Wittkowski R and Matissek R (1993) *Capillary Gas Chromatography in Food Control and Research*. Lancaster, PA: Technomic.

Supercritical Fluid Chromatography

J W King, Los Alamos National Laboratory, Los Alamos, NM, USA

Background

Supercritical fluids are unique substances that exhibit physical properties that are intermediate between those of a gas or liquid. Their density can be adjusted by regulating the pressure or temperature of the fluid, thereby allowing the fluid's density to range from that of a dilute gas to liquid-like values. Such a variation in density accounts for the adjustable, liquid-like properties that these fluids exhibit when used for extraction and for separation purposes. Mass transport properties of supercritical fluids (SFs) also exhibit values that are in between those of gases and liquids, but in general, properties like solute diffusion coefficients in SFs are more gas-like than liquid-like in magnitude. These favorable characteristics allow SFs to readily penetrate samples, thereby allowing for rapid dissolution and high flux rates of solutes into the SFs.

The supercritical fluid region is usually defined in the upper right region of a pressure and temperature phase diagram by a substance's critical points, the critical pressure (P_c) and critical temperature (T_c). This is shown in **Figure 1** for carbon dioxide (CO_2), the substance usually utilized in supercritical fluid extraction (SFE) and supercritical fluid chromatography (SFC). Above a pressure of 7.38 MPa (1070 psig) and a temperature of 31 °C, carbon dioxide will always be in its supercritical state (SC-CO_2) no matter what external pressure is applied to CO_2. This relatively low T_c, CO_2's high nonideality, and its relative environmentally benign nature, make CO_2 the fluid of choice for SFC and SFE.

The application of SFC to foods and nutritional analysis came about naturally owing in part to the early application of SC-CO_2 extraction in the food industry, i.e., for the extraction of coffee, hops, and similar food items used routinely by the consuming public. As will be demonstrated, SFC is particularly applicable for the analysis of lipid-containing

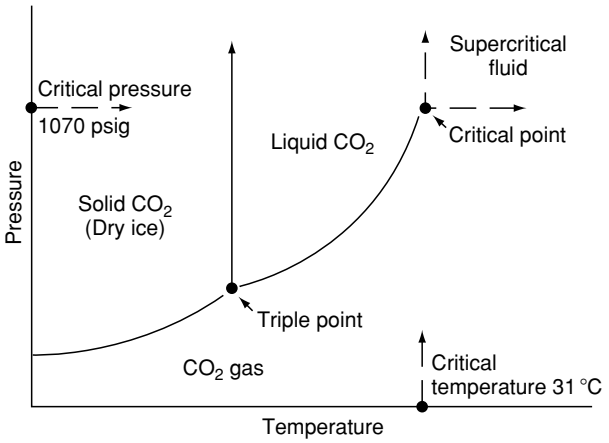

Figure 1 Pressure and temperature phase diagram for carbon dioxide.

materials, owing to relative high solubilities exhibited by these solutes (analytes) in SC-CO_2. Analysis and detection of ultratrace components in foodstuffs, e.g., pesticides or drugs, has not generally been successful owing to the problems in routinely interfacing and using sensitive detectors, such as the electron capture detector with SFC, owing to the change in mobile phase characteristics with respect to time during analysis. However, the ability to routinely use the flame ionization detector (FID) with SFC has provided the analyst with a useful technique to detect an array of components, or a specific moiety in a complex food matrix.

With respect to the chromatographic technique utilized, it was initially the capillary mode of SFC that was cited more often than packed column SFC for the analysis of foods. This has changed somewhat in recent years, however, particularly with the advent of using packed column SFC for the analysis of chiral compounds. This has created more opportunities for applying SFC using packed columns in food analysis, since a knowledge of the chirality of a compound can be of importance even in the food industry (e.g., flavor esters). Figure 2 shows the basic schematic for a capillary SFC (a) and a packed column SFC (b) instrument. The components making up the instrumentation for both modes of SFC are very similar: dual pumps for the SFs and cosolvent (or modifier), respectively; columns contained inside a thermostatted oven and detectors external to the oven. However, in the case of the capillary SFC, the fluid is depressurized directly into the detector, whereas in the packed column case, the detector is usually maintained under pressure, before it is discharged to waste (W). In the packed column mode, the SFC can be interfaced with a ultraviolet detector (UV), evaporative light-scattering detector (ELSD), or a mass spectrometer. In general, the coupling of analytical supercritical fluid extraction (SFE), on-line with SFC, has not been adopted to any considerable extent by food analysts, owing to the lack of an interface that permits routine coupling for use of the SFE/SFC mode. However, preparative and production-scale SFC have been utilized for specialized applications in the food-production industries and will probably see increased use owing to the current interest in producing high-value nutraceutical components, in a 'natural' and environmentally benign manner.

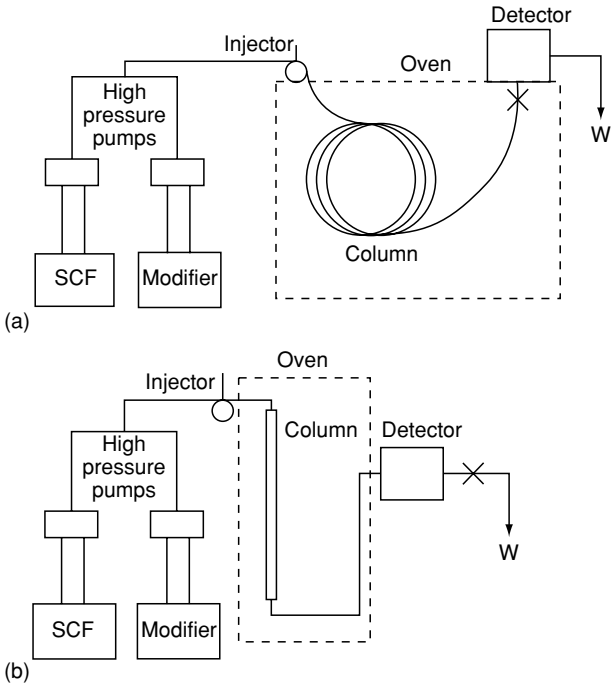

Figure 2 Schematic diagram of (a) capillary SFC, and (b) packed column SFC.

SFC is perceived as a niche technique in the food industry, so it is critical to recognize when and where it can be used to advantage relative to what can be achieved using other forms of chromatography, such as gas chromatography (GC) and high-performance liquid chromatography (HPLC). (*See* **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography.) Some of these opportunities are as follows:

- reduction of the use of organic solvents relative to HPLC;
- direct analysis of samples avoiding sample preparation step;
- deformulation of commercial food products;
- detection of product adulteration or deterioration;
- support of food engineering extraction/reaction process development.

With respect to nonpolar solutes, pressure- or density-programmed SFC provides the capability to analyze compounds having molecular weights approaching 1200 Da using one chromatographic analysis. Separation of these compounds by SFC is a function of their solubility in the mobile phase, their respective vapor pressures, and the miscibility pressure/temperature of the solute in the fluid phase. For example, in Figure 3, a diverse number of components have been separated using the capillary SFC method, that traditionally would have required the use of either GC or HPLC, as well as derivatization of some of the analytes. Utilizing SFC allows the analyst to avoid using either of both GC or HPLC, and to directly analyze the sample, obtaining a 'snapshot' of the entire molecular composition of the sample. These characteristic elution patterns produced using a SFC can be used as means of identifying the presence

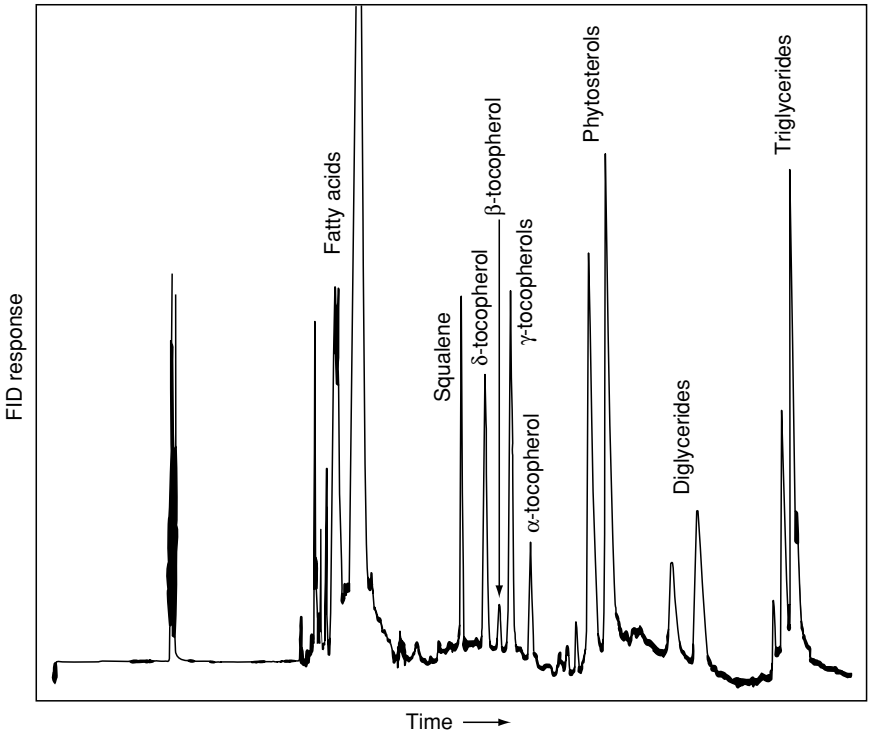

Figure 3 Supercritical fluid chromatography analysis of deodorizer distillate with a SB-octyl-50 column using flame ionization detection.

or absence of a particular analyte in a food sample, thereby providing valuable information for a food product formulator, to match or alter, in developing new and competitive food-related products.

Increasing concerns about minimizing or eliminating the use of hazardous organic solvents in the laboratory also bodes well for use in SFC. By incorporating SFC for the separation and detection of food-related solutes, one not only eliminates most of the traditional solvent needs associated with HPLC, but potentially solvents utilized in the extraction or sample workup steps prior to analysis. In this regard, SFC is an excellent tool for monitoring the end result of an extraction or reaction of a food component using supercritical fluid media. Also, by using SFC, food-related analytes that are thermally labile or susceptible to degradation via oxidation are not exposed to the harsh conditions that often accompany their analysis by GC or HPLC. Such an advantage can be attributed to the protective action of SC-CO_2, which excludes oxygen, and the low temperatures used when separating components via SFC.

Selecting and Optimizing SFC Separation Conditions for Food Components

For the SFC analysis of food-related samples, the analyst will undoubtly want to start with a general fluid programming sequence to separate as many components in the sample matrix as possible. These programs are executed for an extended time to assure optimum resolution and detection of the unknown or target analyte(s). Run times of 1.5 h in length are not unusual in this beginning stage of method development for SFC separations. After the target analyses have been identified by retention time matching with standards, or via an independent method such as mass spectrometry (MS), the original program can be modified to reduce the analysis time or improve the resolution within the chromatogram. Usually, changes in the mobile phase program will be executed to hasten the elution of early or late eluting components that are of no importance for the analysis.

Because of the molecular complexity exhibited by many food ingredients or compositions, it is not unusual to have a temperature gradient with respect to time, superimposed on the mobile phase pressure program during the SFC run. For example, separation of the like-carbon number triglycerides in soybean oil is not possible by pressure or density programming alone, but by superimposing a temperature gradient during the analysis, these oil components can be separated quite well. SFC analysis under isobaric conditions is limited in application when analyzing foods, but it should not be overlooked, since it can yield often the most precise and accurate results.

By far, the FID has been the most used detector used to date in SFC. FID sensitivity to food components is lower than that obtained using GC, since expansion of the mobile phase dilutes the detector signal substantially. However, the FID signal can be amplified to permit analysis down to the p.p.m. level, provided that baseline shifts can be compensated for. Analytes with chromaphoric properties are amenable to analysis using the ultraviolet (UV) detector in conjunction with SFC. The absorption maxima of components shifts as a function of fluid density or pressure, but most detector units constructed for operation at these elevated pressures allow for stop flow or *in situ*, on-the-fly scanning of peaks to determine absorbance maxima shifts. Bathochromic shifts of 15–20 nm have been recorded for carotenoids in SC-CO_2 over a 25-MPa pressure interval. The mating of the ELSD detector with SFC has been reported by several investigators, but the day-to-day stability is inferior to that obtained with HPLC–ELSD couplings when applied to food analysis. Heteroelement specific detectors, such as the electron capture or flame photometric detector, have been utilized mostly in research studies using SFC and have not been adopted for routine analysis. The detector sensitivity and stability under SFC conditions limit their sensitivity at best to the p.p.m. range.

Many of the promising applications of capillary SFC have utilized nonpolar bonded phases utilizing siloxane functionalities as methyl, octyl, phenyl, and biphenyl in a polymeric form. The weak elutropic strength of neat SC-CO_2 has favored the use of short-chain-length, monomeric silance-modified columns having C_1, C_4, C_{18}, phenyl, amino, and diol phases for packed column SFC. The choice of these phases is not so much related to their selective interaction with food-related solutes, but to their surface-modifying properties, which reduce peak tailing and solute interaction with the base silica matrix. Polymer resin columns have also been utilized, but they are susceptible to voiding unless specifically packed for use under supercritical fluid conditions.

Table 1 Food components separated and analyzed by SFC

Carbohydrates:	Derivatized corn syrups, mannose glycans
Chiral compounds:	Monoterpenes, pyrazines, clenbuterol
Drugs/antibiotics:	Caffeine, erythromycin, polycyclic ether antibiotics, sulfonamides, assorted steroids
Hydrocarbons:	Sesquiterpenes, squalene, waxes and wax esters
Lipids:	Fatty acids, fatty acid esters, monoglycerides, diglycerides, triglycerides, sterol esters, sterols (cholesterol), fat-soluble vitamins, tocopherols, phospholipids (lecithin), lipid hydroperoxides, glycolipids
Nutraceuticals:	Valeriana, gingolides, sawtooth palmetto berry
Oils/fats:	Celery oil, coconut, fish, soybean, wheat germ, palm oil, rice oil, milk/cheese triglycerides
Packaging/film components:	Polypropylene oligomers, polyvinyl chloride, phenolic antioxidants, low-molecular-weight polystyrene
Pesticides:	Halogenated, organophosphorus, carbamate, pyrethrins, acidic phenoxy herbicides, sulfonyl ureas
Pigments:	Carotenoids, xyanthophyls
Speciality ingredients:	Hops components
Spices/flavors:	Capsicum, cardoman, coumarin, curry, garlic components, marjoram, rosemary, vanillan
Terpenes/essential and fruit oils:	Grapefruit oil, limonenes, mint, lemon
Other toxicants:	Mycotoxins, nitrosamines, polycyclic aromatic hydrocarbons

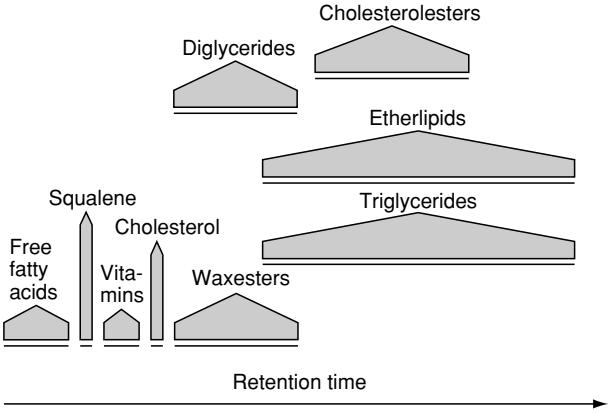

Figure 4 General elution order (retention time) of lipids in SFC on nonpolar stationary phases.

Types of food Ingredients Analyzed by SFC

A myriad of food-related components and matrices have been analyzed by SFC, as indicated by the partial listing in Table 1. These include naturally occurring ingredients such as fats/oils, spices, etc., minor unwanted constituents like pesticides, antibiotic drugs, and mycotoxins, and specific food components, including nutraceuticals and flavoring aids. Inspection of Table 1 indicates a preponderance of applications in the lipid analysis area. Indeed, SFC is tailor-made for lipid analysis, although somewhat lacking in the high-resolution capabilities demonstrated by high temperature GC. The retention pattern for typical lipid solutes in SFC as shown in Figure 4, follow a distinct pattern governed approximately by the solute's molecular weight/volatility characteristics. Elution of the following classes of lipids occurs in the following order: fatty acid methyl esters, free fatty acids, hydrocarbons, vitamins, sterols, wax esters, mono- and then diglycerides, followed by triglycerides and steryl esters. Although there is some

overlap between individual classes of the above solutes, owing to the overlapping molecular weight ranges (e.g., triglycerides and steryl esters), this solute separation pattern has proven very useful in tracking conversion of lipid species undergoing reactions, as well as in the quality control of food ingredients.

Triglyceride-based oils/fats are also readily amenable to analysis by SFC. Separation of the individual components is once again governed by molecular weight considerations, allowing SFC to facilitate the separation of the major triglyceride species, i.e., T_{50}, T_{52}, T_{54}, etc. For some oils, such as coconut oil, picturesque chromatograms result (Figure 5), as shown by the separation of the C_{18} to C_{54} saturated triglycerides as equally spaced peaks. For other oils such as soybean oil, there is some overlap between the saturated and unsaturated triglyceride species, requiring, as noted previously, a superimposition of a temperature gradient along with the pressure gradient program to achieve adequate resolution. However, even without high resolution, the rapid SFC analysis can be used to advantage in quality control, where speed, rather than optimal separation, is often desired.

Detection of minor components in foods is limited by the detector sensitivity and stability as noted previously, but those components that can be detected by using FID, UV, or ELSD, are often analyzed more rapidly by SFC, owing to the time savings afforded by avoiding elaborate preparation of the sample prior to analysis. In addition, SFC analysis provides a more detailed profile of the entire sample in addition to detecting the target analyte. This allows an accurate assessment of the total contribution of the minor constituent to the entire ingredient profile, e.g., the presence of sterol esters in sawtooth palmetto berry extracts, where the fatty acids and triglycerides are the major constituents.

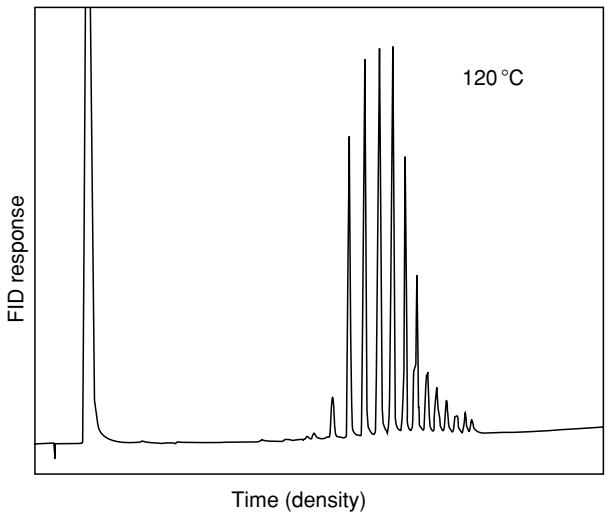

Figure 5 Capillary supercritical fluid chromatogram of coconut oil triglycerides.

Other food sample types that can be readily analyzed by SFC are the fat-soluble vitamins, essential and flavor oil ingredients, spice extracts, hop components, as well as functional food components, i.e., nutraceuticals. Some SFC-based separations require the use of a cosolvent (usually 5–20 vol.%), in addition to the SC-CO_2, to modify the mobile phase. For example, phospholipids are only sparingly-soluble in neat SC-CO_2, but these polar lipid compounds can be chromatographed successfully on packed silica columns by incorporating ethanol and/or water as a modifier into the mobile phase. Likewise, carbohydrate moieties, which exhibit limited or no solubility in SC-CO_2 or SC-CO_2 /cosolvent mobile phases, can be derivatized so that they can be analyzed by SFC.

Selected Applications of SFC in Food and Nutritional Analysis

In this section, several brief examples will be given to illustrate the utility and potential of SFC for food and nutritional analysis. As noted previously, both high and low resolution of solutes are possible using SFC. SFC separation and detection of α-tocopherol and cholesterol in a fish oil capsule can be achieved on a capillary SB-methyl column at 120 °C, using a density program with SC-CO_2 from 0.28 to 0.66 g ml^{-1} and a density programming rate of 0.0006 g ml^{-1}. Although the analysis took 90 min to perform, it allowed the analyst to avoid any sample preparation, other than diluting the oil from the capsule in small quantity of solvent, and injecting it into the SFC. Therefore, no derivatization of the sample was required, and there was sufficient resolution between the α-tocopherol and cholesterol to quantify these analytes. This was made possible by adjustment of the elution conditions, i.e., the background components (fish oil triglycerides), that were of no interest in this analysis, and were programmed off the column without resorting to a prefractionation of the sample prior to SFC analysis.

Not all applications of SFC require the above high-resolution separation. Packed column SFC (5 μm, C_8 – Deltabond) has been used to 'clean up' samples prior to other types of chromatographic analysis (GC). In this case, organochlorine- and phosphorus pesticides extracted by SFE with SC-CO_2 from a meat sample were separated from the coextracted fat moieties using the packed SFC column. Hence, by 'heart cutting' the appropriate elution fraction, a lipid-free, pesticide-containing fraction was provided for residue analysis by GC.

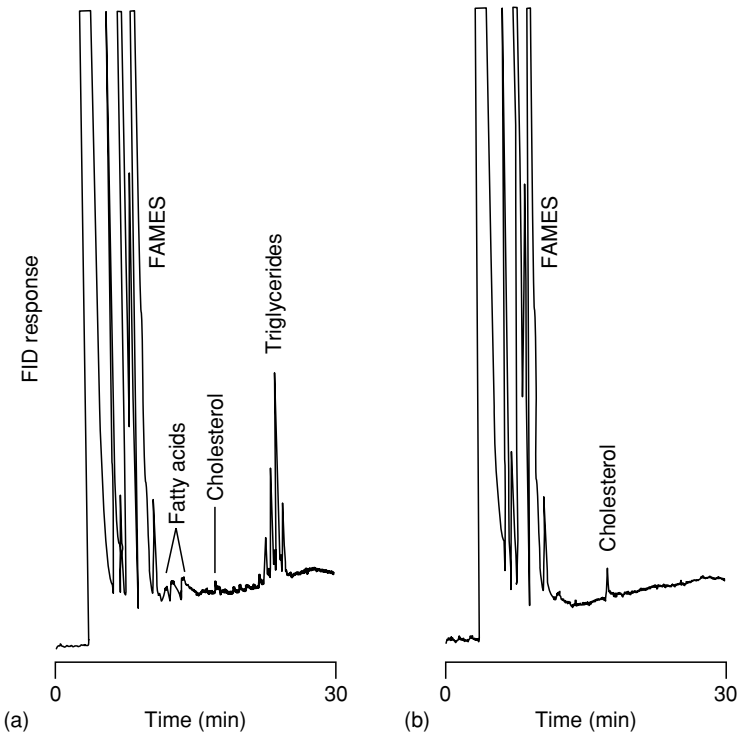

Figure 6 SFC analyses of reaction products from the transesterification of ground beef lipids: (a) neat ground beef sample; (b) freeze-dried ground beef sample.

SFC is an excellent technique to monitor reaction chemistry between lipid species, since it avoids the need to employ more than one analytical technique or analyte derivatization. Further, it permits the successful chromatography of all of the relevant reactants and products using one chromatographic analysis. Examples where SFC has been applied are for the esterification or transesterification of lipids, glycerolysis reactions, and randomization of fats/oils. Figure 6 shows two SFC profiles for the enzymatic-catalyzed transesterification reaction between the triglycerides from a ground beef sample and methanol, to produce the corresponding fatty acid methyl esters (FAMEs) for nutritional labeling analysis. The 30-min chromatograms show the effect of the moisture content of the ground beef sample on the yield of FAMEs, (a) being the yield from the neat ground beef sample, and (b) being the conversion attained by freeze-drying the sample matrix prior to methanolysis. Note that freeze-drying of the ground beef was crucial to obtaining a quantitative yield of FAMEs for fat analysis required in nutritional labeling. The transesterification reaction in this case was run in a flow extractor-reactor in which the ground beef triglycerides were extracted using SC-CO_2 and carried over a bed of supported lipase in the presence of methanol to produce the desired FAMEs.

Table 2 Analysis of lipid species in a soybean oil glycerolysis mixture

Method of analysis	*MAG*	*DAG*	*TAG*	*FFA*
SFC – FID	75.3	23.2	< 0.1	1.5
GC – FID	76.2	21.8	< 0.1	2.0

Table 2 shows the analysis of the glyceride content of a soybean oil glycerolysis mixture by both HPLC-FID and SFC-FID. The information desired was the relative amounts of free fatty acids (FFA), mono- (MAG), di-(DAG), and triacylglycerols (TAG) in the reaction mixture. A comparison of the SFC-FID results with those obtained by HPLC-FID shows a relatively good agreement between the two methods. However, the SFC method does not require the time and effort for sample preparation associated with the alternative technique, and in addition saves on the cost of solvents and chemical reagents.

Although mentioned previously, preparative or production-scale SFC is now being used as a separation technique in the food industry. Fractionation and isolation of food components, such as tocopherols and phospholipids, or the ω-fatty acids/esters from fish oils, have been cited in the literature.

Recently, a production plant for the separation of fish oil ethyl esters has been constructed in Spain to produce ≥95% pure polyunsaturated fatty acids for the nutraceutical market. The basic separation premise upon which this production scale plant is designed was based on chromatographic fractionations initially developed using analytical-scale, packed SFC columns.

See also: **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography

Further Reading

Anton C and Berger C (eds) (1997) *Supercritical Fluid Chromatography with Packed Columns*. New York: Marcel Dekker.

Caude M and Thiebaut D (eds) (1999) *Practical Supercritical Fluid Chromatography and Extraction*. Amsterdam: Harwood Academic.

Chester TL (1997) Chromatography from the mobile-phase perspective. *Analytical Chemistry* 69: 165A.

Dean JR (ed.) (1993) *Applications of Supercritical Fluids in Industrial Analysis*. Boca Raton, FL: CRC Press.

King JW (1990) Applications of capillary supercritical fluid chromatography – supercritical fluid extraction to natural products. *Journal of Chromatographic Science* 28: 9.

King JW (ed.) (1996) Supercritical fluid extraction and chromatography. *Seminars in Food Analysis* 1: 101–116, 133–144, 163–165.

King JW, Hill HH, Jr and Lee ML (1993) Analytical supercritical fluid chromatography and extraction. In: Rossiter BW and Baetzold RC (eds) *Physical Methods of Chemistry, Supplement and Cumulative Index*, 2nd edn., vol. X. New York: John Wiley.

King JW and List GR (eds) (1996) *Supercritical Fluid Technology in Oil and Lipid Chemistry*. Champaign, IL: AOCS Press.

Lee ML and Markides KE (eds) (1990) *Analytical Supercritical Fluid Chromatography and Extraction*. Provo, UT: Chromatography Conferences.

McDonald RE and Mossoba MM (eds) (1997) *New Techniques and Applications in Lipid Analysis*. Champaign, IL: AOCS Press.

Markides KE and Lee ML (eds) (1988, 1989) *SFC Applications*. Provo, UT: Brigham Young University Press.

Nam KS and King JW (1994) Coupled SFE/SFC/GC for the trace analysis of pesticide residues in food samples. *Journal of High Resolution Chromatography* 17: 577.

Saito M, Yamauchi Y and Okuyama T (eds) (1994) *Fractionation by Packed Column SFC and SFE*. New York: VCH.

Smith RM (ed.) (1988) *Supercritical Fluid Chromatography*. London: Royal Society of Chemistry.

Wenclawiak B (ed.) (1992) *Analysis with Supercritical Fluids: Extraction and Chromatography*. Berlin: Springer.

Combined Chromotography and Mass Spectrometry

F A Mellon, AFRC Institute of Food Research, Norwich, UK

Introduction

Techniques of combined chromatography/mass spectrometry are widely used in food analysis and have been applied to the solution of a very diverse range of problems. These include flavor analysis, determination of vitamins and other biologically active naturally occurring substances in foods, analysis of lipids, amino acids, peptides and proteins, analysis of toxicants, and determination of pesticide and veterinary drug residues. These are a small sample of a very wide range of applications of combined gas chromatography–mass spectrometry (GC-MS), liquid chromatography–mass spectrometry (LC-MS), capillary electrophoresis–mass spectrometry (CE-MS), supercritical fluid chromatography–mass spectrometry (SFC-MS), and capillary electrochromatography–mass spectrometry (CEC-MS).

The general principles of combined chromatography–mass spectrometry are the same for all hyphenated mass spectrometry techniques: the chromatograph separates the mixture components and these are transferred to the mass spectrometer, often via a suitable interface (**Figure 1**). An analog detector (for example, diode array or fluorescence) may also be incorporated, particularly in LC-MS instruments. This may be placed in series with the mass spectrometer interface or the eluent flow can be spilt between the two types of detector. Many mass spectrometer data systems are capable of recording and displaying mass spectra and (for example) ultraviolet data simultaneously, adding an extra dimension to the information obtained.

The mass spectrometer is scanned continuously and cyclically (e.g., one scan per second), so that the spectra of all components, even minor ones, are recorded. The data are plotted as a total ion current (TIC) chromatogram or a mass chromatogram (**Figure 2**).

The TIC is simply the sum of the intensities of all the ions in each mass spectrum plotted against retention time and/or scan number. A mass chromatogram (e.g., *m/z* 270 in **Figure 2**) comprises a plot of the intensity of a particular ion in a mass spectrum against time. This technique is often useful in identifying particular molecular ions or fragment ions that are characteristic of particular structural features. For example, the ion at *m/z* 74 in **Figure 2** is a McLafferty rearrangement ion characteristic of fatty acid methyl

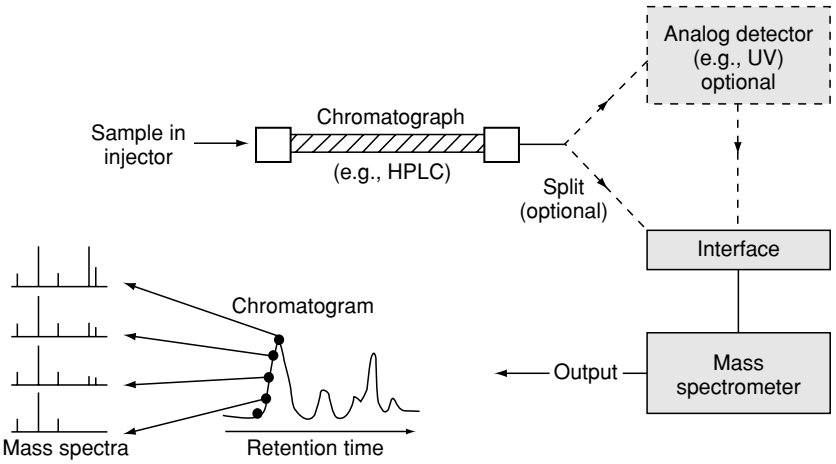

Figure 1 Schematic diagram of the chromatography–mass spectrometry process. HPLC, high-performance liquid chromatography; UV, ultraviolet. Copyright Institute of Food Research, reproduced with permission.

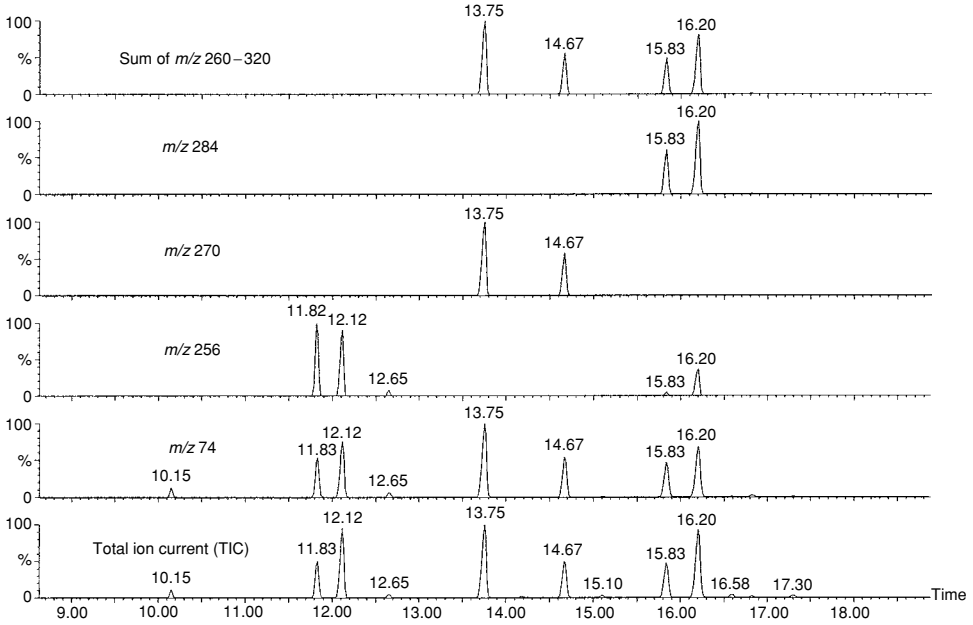

Figure 2 Total ion current (TIC), mass, and summed mass chromatograms of selected ions in the gas chromatography/mass spectrometry of fatty acid methyl esters. Copyright Institute of Food Research, reproduced with permission.

esters containing the moiety $-C(H)\text{-}CH_2\text{-}CH_2\text{-}COOMe$. Similarly, the ions at 270 and 284 are characteristic molecular ions of methyl palmitate and margaric acid (methyl heptadecanoate) and isomers of these molecules. A mass chromatogram is different from a selected ion-monitoring (SIM) trace (see below) because the ions of interest are part of a complete mass spectrum: the entire mass spectrum is still available for examination. In addition to TIC and mass chromatogram plots, ion intensities may be summed over a preselected mass range and plotted against time (for example, m/z 260 – 320 in Figure 2). This type of display is sometimes referred to as a Reconstructed ion chromatogram and is useful for eliminating intense background ions (for example, those that arise from solvent peaks during an LC-MS analysis), or for searching for characteristic ions that appear in a limited mass range.

Whichever chromatographic plotting techniques are used, each time point in the chromatogram represents a single mass spectrum that may be examined or processed either by the operator or automatically

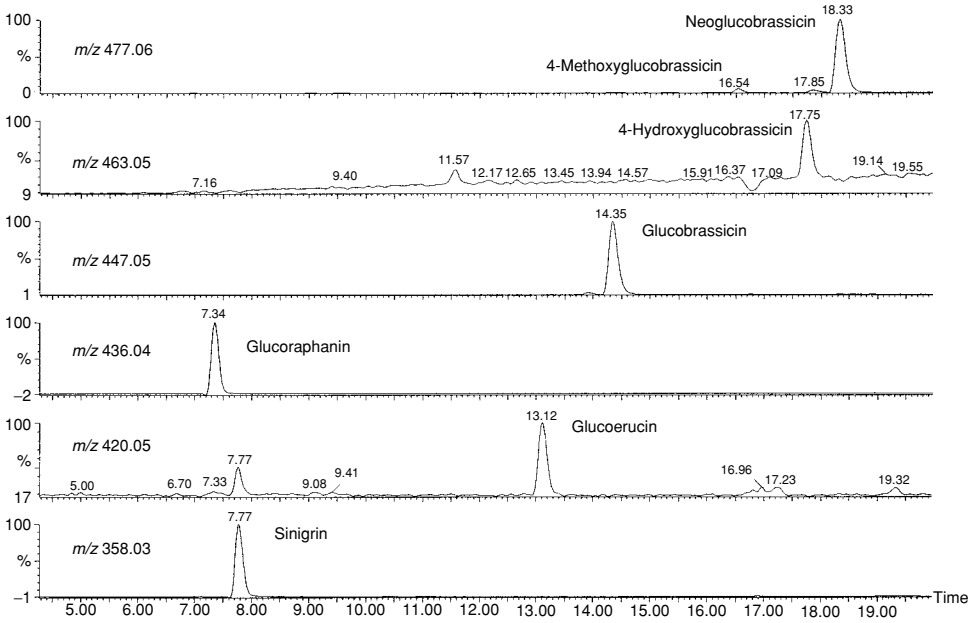

Figure 3 Negative-ion electrospray liquid chromatography–mass spectrometry (LC-MS) selected ion-monitoring (SIM) chromatogram of intact glucosinolates extracted from broccoli. Copyright Institute of Food Research, reproduced with permission.

using appropriate software (for example, library search or quantification routines).

In the special case of SIM, preselected masses are scanned in preference to the entire spectrum. This yields much greater sensitivity and is used for quantifying analytes. Figure 3 shows an LC-MS SIM chromatogram of the molecular anions of the glucosinolates glucoraphanin, glucoerucin, glucobrassicin, 4-hydoxyglucobrassicin, 4-methoxyglucobrassicin, and neoglucobrassicin extracted from broccoli.

Sinigrin (a glucosinolate not present in broccoli) is used as a quantitative spike. Peak areas are integrated and the ratios of these areas against that of the sinigrin internal standard are used to generate quantitative data by comparison with a calibration curve.

Gas Chromatography–Mass Spectrometry

GC-MS was the first combined chromatography–mass spectrometry technique to enjoy widespread acceptance in food science (one of the earliest published applications of GC-MS was in flavor analysis). GC-MS is still used widely in food science, particularly in flavor analysis. User-friendly and cheap benchtop instruments are now available for conducting these types of analysis. Despite their compactness and simplicity, these instruments are sensitive and equipped with powerful data systems that are used not only to control and acquire data from the gas chromatograph and mass spectrometer, but also to identify flavor components by automated matching against reference libraries of spectra of known flavorants. More sophisticated GC-MS instruments are also used in many other areas of food analysis, especially in the identification of toxic residues (e.g., dioxins). Advances in column technology (particularly the widespread use of flexible fused silica columns) and improved mass spectrometer vacuum pumps mean that coupling a gas chromatograph to a mass spectrometer is now routine and does not require any special form of interface. Capillary columns generally operate at helium gas flows of a few milliliters per minute, and modern turbomolecular vacuum pumps are easily able to cope with these flow rates and still maintain ion source pressures (for electron ionization (EI)) of $< 10^{-3}$ Pa. One of the main recent advances in GC-MS technology has been the introduction of 'fast' GC-MS techniques in which the speed of analysis available from some types of modern GC capillary columns have been complemented by fast-scanning mass spectrometers (time-of-flight instruments, for example). GC-MS techniques are now routine and will not be dwelt on at any length here. LC-MS is now providing some of the most interesting and versatile new methodologies in food analysis and food science generally and this article will focus mainly on this technology.

Liquid Chromatography–Mass Spectrometry

LC-MS presents far more technical problems than GC-MS because of the incompatibility of flowing

liquids with the high vacuum necessary for mass spectrometers to function. A typical liquid chromatograph eluent flow of 1 ml min^{-1} becomes over 1000 times this flow of gas when converted to a vapor, far in excess of the gas flow that a conventional mass spectrometer ion source can cope with and still maintain vacuum. Furthermore, the expansion from liquid to gas results in dilution of analyte molecules entrained in the solvent, potentially reducing sensitivity by a very large factor. Consequently, special interfaces and ion sources have been developed to remove excess solvent vapor and insure that the analytes are not diluted excessively.

LC-MS Interfaces and Ion Sources

Several different techniques, or interfaces, for combining liquid chromatography and mass spectrometry were developed in the 1970s and 1980s. These included the direct liquid introduction (DLI) interface, the moving belt interface, dynamic FAB (fast atom bombardment) or dynamic LSIMS (liquid secondary ion mass spectrometry), and thermospray ion source. All of these techniques had some disadvantages in terms of the range of compounds amenable to analysis, robustness and reliability, sensitivity and, in some cases, limited flow rates.

Electrospray and Atmospheric Pressure Chemical Ionization

The interfaces mentioned above have been superseded by electrospray (ESI) and atmospheric pressure chemical ionization (APCI) ion sources. Both of these ion sources can be coupled directly to liquid chromatographs, with very few restrictions on flow rate or solvent composition (*See* **Mass Spectrometry**: Principles and Instrumentation). Both techniques have been used very widely in food and nutrition studies. Examples of SIM chromatograms of 1 and 20 pmol of β-carotene (**Figure 4**) and a mass chromatogram (**Figure 5a**) and spectrum (**Figure 5b**) of quercetin 3′ glucoside are shown. These types of data are useful in quantitative and qualitative metabolic studies.

The Particle Beam Interface

A third type of interface, the particle beam, completes the trio of modern LC-MS technology, providing a suite of techniques capable of handling nearly all types of analysis. A schematic diagram of a particle beam interface is shown in **Figure 6**.

The incoming liquid stream (i.e., the high-performance liquid chromatography (HPLC) eluent) enters the device on the left of the diagram. The liquid is nebulized by helium gas flowing, forming a fast-moving spray of droplets that enter the heated desolvation chamber. Solvent begins to evaporate from the spray droplets which shrink, thus concentrating any dissolved sample molecules. The partially evaporated spray then passes through a narrow orifice (skimmer 1 in the diagram) and across a region that is maintained under vacuum by a rotary pump. A supersonic molecular beam is formed in this region and, because

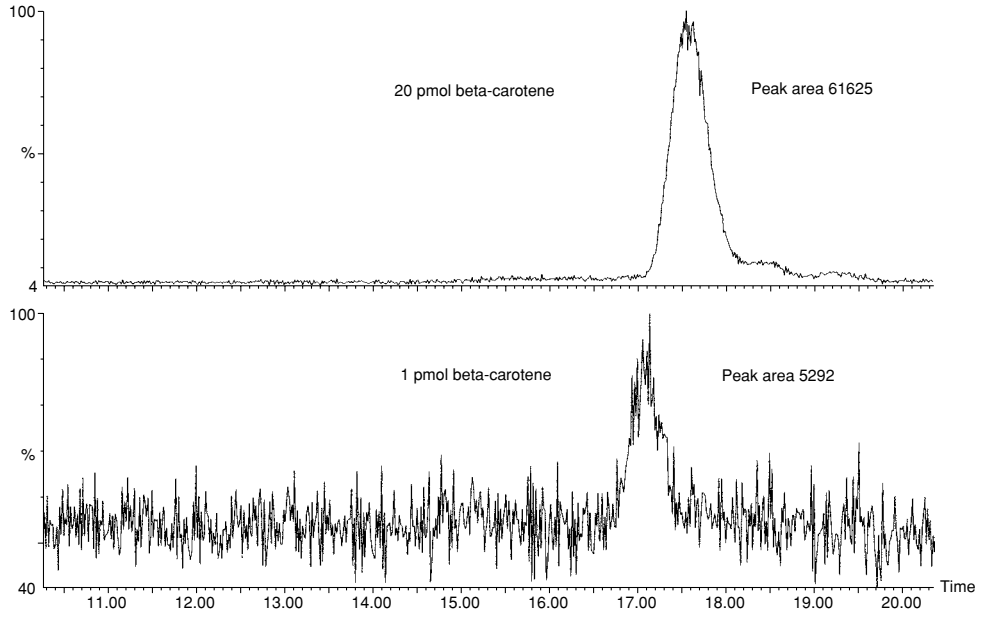

Figure 4 Positive-ion atmospheric pressure chemical ionization liquid chromatography–mass spectrometry selected ion-monitoring chromatogram of the molecular ion of β-carotene (1 and 20 pmol injections, respectively). The slight retention time shift between the two traces is caused by temperature drift affecting the chromatography. These types of data are useful for quantifying carotenoids in food extracts and in metabolic experiments. Copyright Institute of Food Research, reproduced with permission.

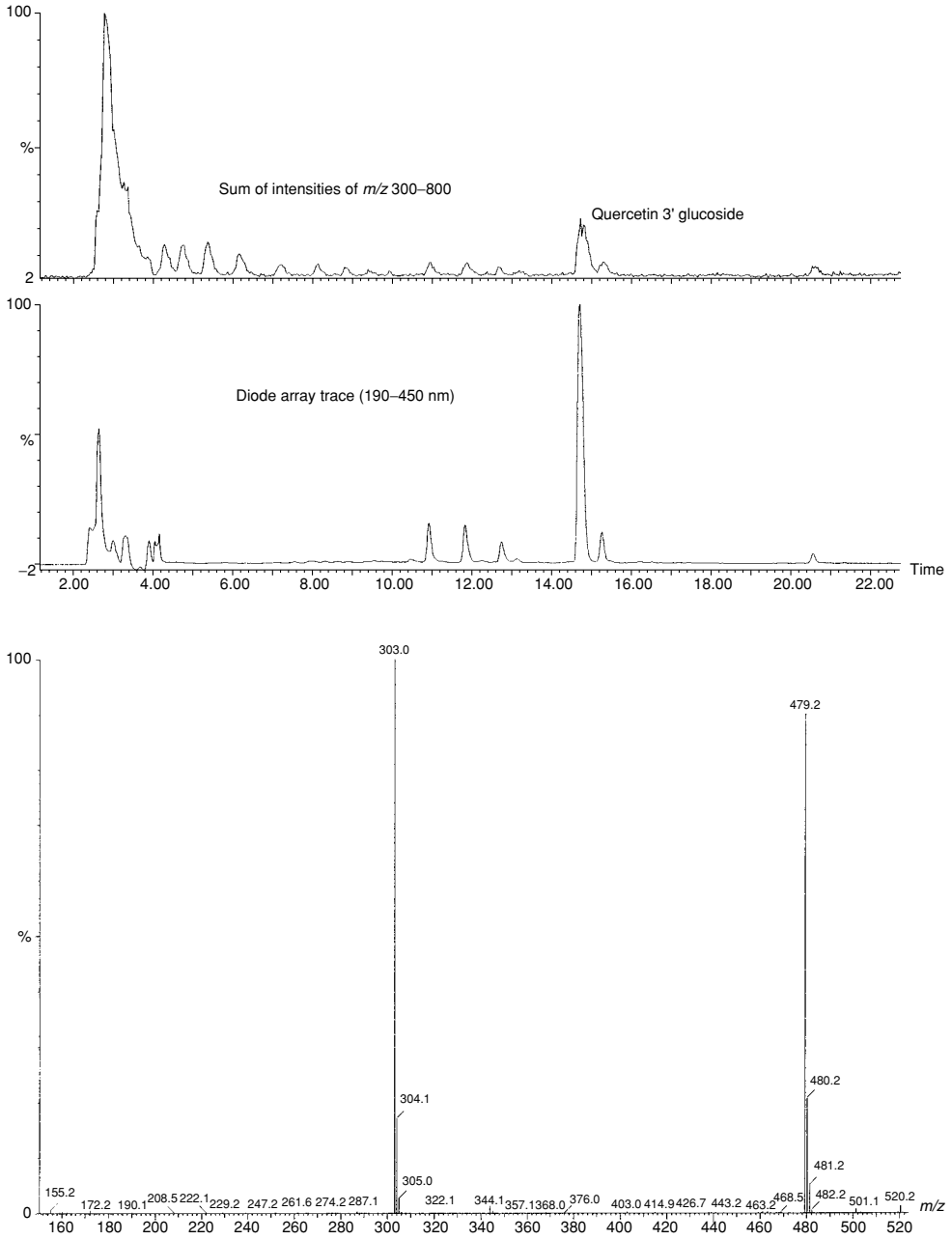

Figure 5 Positive-ion electrospray ionization liquid chromatography–mass spectrometry (a) chromatogram and (b) mass spectrum of the important food flavonoid quercetin 3′ glucoside. Copyright Institute of Food Research, reproduced with permission.

the solvent molecules and helium atoms are significantly lighter than sample molecules, they diffuse away more rapidly and are pumped out of the system preferentially. This leaves the heavier solute molecules concentrated in the beam. A second skimmer removes yet more solvent molecules and allows particles of sample to enter the EI or chemical ionization (CI) ion source. The interface can handle flows of between 0.1 and 1 ml min^{-1}, depending on the volatility of the solvent.

The great advantage of the particle beam system is that it can generate CI and, especially, EI spectra that are highly reproducible and comparable with spectra obtained by other types of EI and CI inlet systems. These spectra are therefore searchable against databases of known spectra. The major difference between the particle beam interface and other LC-MS systems is that the device processes the HPLC eluent in such a way that sample molecules can be ionized by the conventional mass spectrometric techniques of EI

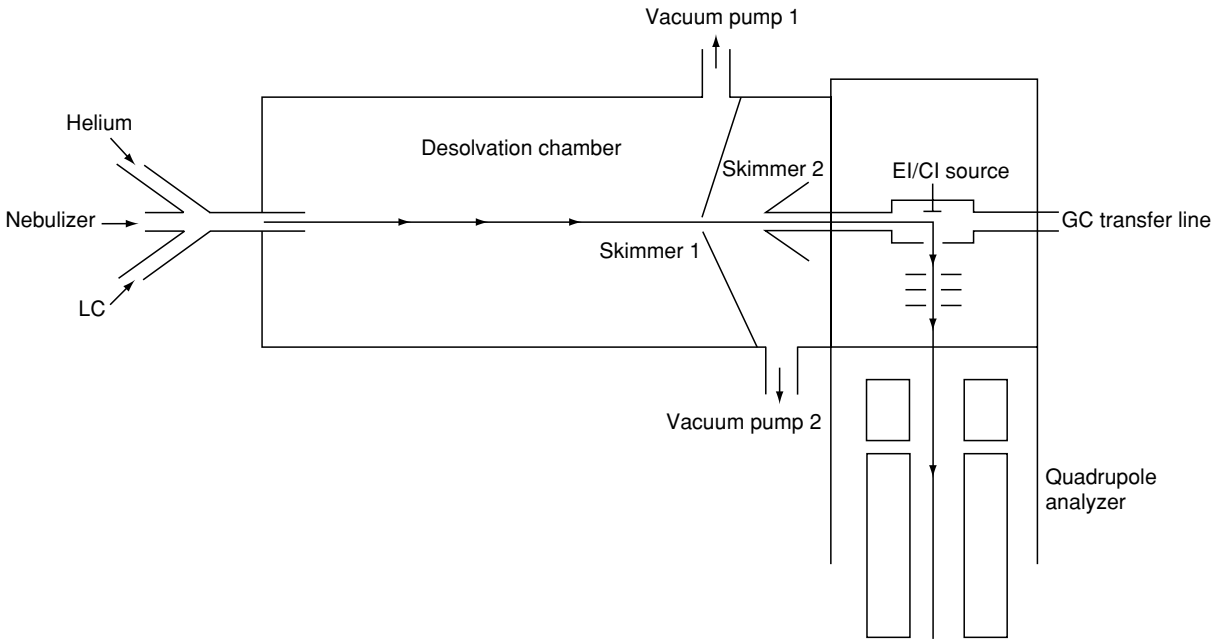

Figure 6 Schematic diagram of a particle beam liquid chromatography–mass spectrometry interface. LC, eluent exiting liquid chromatograph; EI/CI, electron impact/chemical ionization; GC, gas chromatography. Reproduced from Mellon FA, Self R and Startin JR (2000) *Mass Spectrometry of Natural Substances in Foods*. London: Royal Society of Chemistry, with permission.

and CI. APCI and ESI are effectively ionization techniques that generate anions or protonated molecules with few fragment ions under normal operating conditions. The particle beam method does not have the versatility and sensitivity of APCI and ESI but is useful for analyzing nonthermally labile molecules that are insufficiently volatile for GC-MS analysis.

Capillary Electrochromatography–Mass Spectrometry

CEC is a high-resolution separation technique that has been described as a hybrid of HPLC and CE. The high efficiency available from CEC makes it an ideal partner for coupling to mass spectrometry as the narrow chromatographic peaks result in increased mass flow per unit time into the mass spectrometer ion source, and thus greater sensitivity as well as selectivity. Many of the technical problems associated with coupling CEC to mass spectrometry (principally Joule heating and consequent bubble formation) are being tackled successfully, to the extent that automated sample introduction systems may be used with CEC-MS systems. Although the technique has not yet caught on in food science and nutrition, CEC-MS has great potential in studies requiring separation of complex mixtures and high-sensitivity/high-selectivity analysis.

Capillary Electrophoresis–Mass Spectrometry

Many of the difficulties associated with combining the ultrahigh-resolution separation technique of CE with MS have been overcome. However, a significant number of publications devoted to interfacing CE with MS still appear annually, indicating that an ideal solution is still not available. Despite this, several mass spectrometer manufacturers now supply special probes for interfacing CE with MS, indicating that the technology is now commercially viable. The ionization technique most compatible with CE is electrospray and the vast majority of CE-MS applications reported make use of this ionization technique. Recently the use of microfabricated devices ('CE on a chip') with mass spectrometers has been reported and promises significant advances in sensitivity and reproducibility. As with CEC, there are few reported applications of CE-MS in the food and nutrition sciences; nevertheless, the potential of technique in these fields is very high.

Supercritical Fluid Chromatography–Mass Spectrometry

Techniques for coupling SFC to MS were first described in the late 1970s. Although a number of different SFC-MS interfaces were developed in the

intervening years, APCI, with specially adapted SFC probes, is currently the main method of choice. The APCI source is particularly suited to SFC because it is unaffected by the increasing flow rates resulting from pressure programming of the supercritical mobile phase and because of the ease of connection with

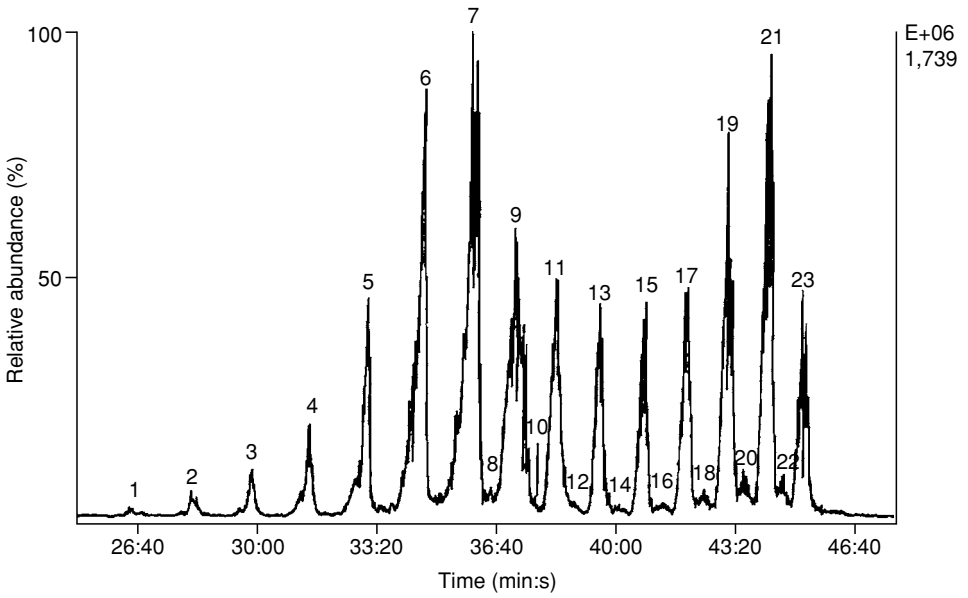

Figure 7 Reconstructed ion chromatogram of milk fat triacylglycerols analyzed by capillary supercritical fluid chromatography–mass spectrometry. Samples ionized by atmospheric pressure chemical ionization with ammonia as reactant ion. Reproduced from Laakso P and Manninen P (1997) Identification of milk fat triacylglycerols by capillary supercritical fluid chromotagraphy atmospheric pressure chemical ionization mass spectrometry. *Lipids* 32: 1285–1295, with permission.

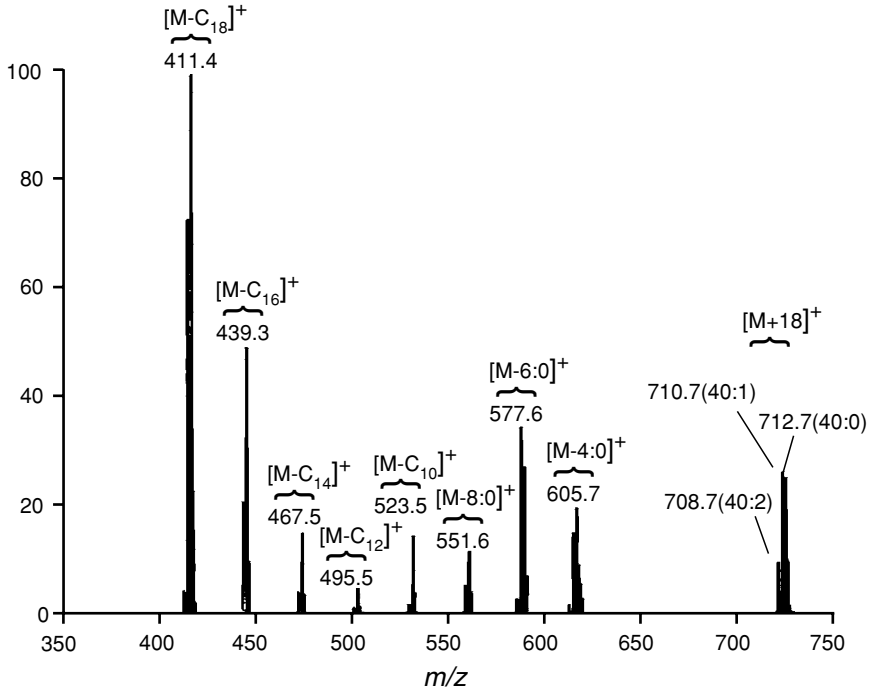

Figure 8 Mass spectrum of peak 9 in Figure 7. Samples ionized by atmospheric pressure chemical ionization with ammonia as reactant ion. Reproduced from Laakso P and Manninen P (1997) Identification of milk fat triacylglycerols by capillary supercritical fluid chromotagraphy atmosphere pressure chemical ionization mass spectrometry. *Lipids* 32: 1285–1295, with permission.

SFC systems. Although SFC-MS is much less popular than LC-MS in the food sciences, it is particularly suited to certain types of analysis. A good example is provided by the identification of milk fat triacylglycerols by capillary SFC-MS. Samples were ionized, in APCI mode, by admitting methanol or 0.5% ammonia in methanol as reagent gas. A reconstructed ion chromatogram of the milk fat triacylglycerols is shown in Figure 7.

Individual mass spectra yielded abundant $[M+18]^+$ and $[M-RCOOH]^+$ ions when ammonia was present in the reagent gas, thus defining molecular weight and the nature of the fatty acid constituents of the triacylglycerols. As can be seen in Figure 7, complete chromatographic resolution of all triacylglycerols was not achieved, However, the specificity of the mass spectrometric data allowed identification of different components of unresolved chromatographic peaks where fatty acid chain lengths yielded characteristic fragment ions. The mass spectrum of peak number 9 in Figure 7 is shown in Figure 8.

Three different chromatographically unresolved triacylglycerols, 40:0, 40:1, and 40:2, are present. This study demonstrated the value of SFC-MS in the analysis of both saturated and unsaturated triacylglycerols.

The interpretation of the mass spectrum in Figure 8 was complicated by the occurrence of several different-molecular-weight species in a single chromatographic peak. However, the $[M-RCOO]^+$ ion region contained ion clusters that indicated the presence of C_{18} and C_{16} fatty acid moieties in the triacylglycerols.

See also: **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; Supercritical Fluid Chromatography; **Mass Spectrometry**: Principles and Instrumentation; Applications

Further Reading

Chapman JR (1995) *Practical Organic Mass Spectrometry: A Guide for Chemical and Biochemical Analysis*, 2nd edn. Chichester: John Wiley.

Cole RB (1997) *Electrospray Ionization Mass Spectrometry*. New York: John Wiley.

James P (2000) *Proteome Research: Mass Spectrometry (Principles and Practice)*. Berlin: Springer Verlag.

Johnstone RAW and Rose ME (1996) *Mass Spectrometry for Chemists and Biochemists*, 2nd edn. Cambridge: Cambridge University Press.

McMaster MC and McMaster C (1998) *GC/MS: A Practical User's Guide*. New York: John Wiley.

Mellon FA, Self R and Startin JR (2000) *Mass Spectrometry of Natural Substances in Food*. London: Royal Society of Chemistry.

Siuzdak G (1996) *Mass Spectrometry for Biotechnology*. San Diego: Academic Press.

Willoughby R, Sheehan E and Mitrovich S (1998) *A Global View of LC/MS: How to Solve Your Most Challenging Analytical Problems*. Pittsburgh: Global View.

CHROMIUM

Contents
Properties and Determination
Physiology

Properties and Determination

R Farré and M J Lagarda, University of Valencia, Burjassot, Spain

Physical and Chemical Properties

Chromium is a hard, brittle, white metal of the first transition series, with an atomic number of 24, a relative atomic mass of 51.996 amu, an electronic configuration $[Ar]3d^5 4s^1$, and a relative density of 7.19. The melting and boiling points are 1857 °C and 2672 °C, respectively.

The resistance of chromium to attack from a variety of chemicals at normal temperature makes it useful in protecting other more reactive metals, e.g., it is a component of stainless steel. However, at high temperatures it reacts with many chemicals.

As a typical transition metal element, chromium forms many compounds that are colored and paramagnetic. It has oxidation states from −2 to +6, but

the most common and stable oxidation states are +2, +3, and +6. Since Cr(II) is a very strong reducing agent, it is not found in biological systems. All Cr(VI) compounds, except the hexafluoride (CrF_6), are oxo compounds; chromium occurs predominantly as either chromate (CrO_4^{2-}) or dichromate ($Cr_2O_7^{2-}$). Cr(VI) compounds are strong oxidizing agents and are therefore readily reduced to Cr(III) in acidic solutions. The most stable and important oxidation state is +3.

One of the outstanding features of Cr(III) chemistry is the ability of trivalent chromium to form coordination complexes, most of which are hexacoordinated. These coordination complexes are relatively inert. In aqueous solution the hexahydrate ion $[Cr(H_2O)_6]^{3+}$ exists in an octahedral configuration. In aqueous solution the ligand and displacement reactions have half-lives of several hours.

Chromium can form bridges with hydroxyl groups (olation). Olation is enhanced by alkalis and temperatures up to 120 °C, but can be prevented and reversed by oxalate ions and other strong ligands, whereas weaker ligands can only prevent the reaction. In biological systems pyrophosphate, methionine, serine, glycine, leucine, lysine, and proline therefore inhibit the olation of Cr(III). In such systems chromium is able to react because its solubility is maintained by weaker organic and inorganic ligands.

Determination in Foods and Beverages

There are a wide range of techniques available for the analysis of chromium and other trace elements. Chromium is one of the most difficult elements to determine accurately because of the very low levels in foods.

The determination of chromium content in biological matrixes has long been a controversial problem because of the many errors that can arise and invalidate the data. Therefore, much of the earlier work should be viewed critically. Reasons for this skepticism arise from a number of problem areas:

- The low chromium content of normal (noncontaminated) biological samples. In food samples the concentration is on the order of nanograms per gram and, therefore, it is essential to apply very sensitive measurement techniques.
- Chromium losses by volatilization during sample dry ashing above 450 °C.
- Contamination at any point through sample collection and handling, before the content is measured.
- Errors inherent in the applied analytical method, such as, for instance, matrix interferences in electrothermal atomic absorption spectrometry (ETA-AAS) and spectral in inductively coupled plasma–atomic emission spectrometry (ICP-AES).

Studies should report precise data on all steps of the analytical procedure, including sample collection, homogenization, and matrix reduction. An accurate determination is a challenge because of problems that originate from the risk of chromium losses by volatilization or adsorption and/or enrichment by contamination. Interlaboratory assays therefore give rise, in the case of chromium analysis, to very variable results.

Of all of the proposed analytical methods, AAS is the one most often used for chromium determination in biological samples. When applied with a graphite furnace (ETA-AAS) it is one of the most sensitive techniques, but it is open to a wide range of interferences, which depend on the matrix and can lead to erroneous results. The advances in background correction procedures have contributed to overcome these limitations.

ICP-AES is also applied to determine chromium in foods and beverages. Although the detection limits are higher than those obtained with ETA-AAS, it offers the advantages of simpler sample preparation and the possibility of carrying out multielemental analysis.

Determination of total chromium content may not be a valid indicator of the nutritional benefit of a food because not all chromium in foods has activity as a glucose tolerance factor. Some authors point out that the alcohol-extractable fraction of chromium seems to be a more reliable index. The speciation of chromium is of great interest because of the differences in bioavailability, essentiality, and toxicity between Cr(III) and Cr(VI).

Precautions to be Taken in Chromium Determination in Foods

As mentioned above, contamination is one of the main problems encountered in chromium determination in foods. Indeed, it is the explanation why food chromium contents reported some years ago are higher than those obtained recently, as better measures have now been adopted to avoid contamination.

The possibility that small amounts of chromium can be retained by the surface of vessels makes it necessary to decontaminate the material used, whether glass or plastic. In order to decontaminate the containers, all material used has to be soaked in diluted nitric acid (10–20% v/v) for 12–24 h and then rinsed several times with deionized water. A mixture of 4 mol l^{-1} nitric acid and 4 mol l^{-1} perchloric acid

(1:1, v/v) or a solution of 2,4-pentanedione or dithizone/carbon tetrachloride has also been proposed.

To avoid losses by adsorption on to container walls, it is advisable to store digested samples and to prepare standard solutions in diluted acid (nitric or hydrochloric acid) below pH 1.5. However in acidic conditions reduction of Cr(VI) to Cr(III) is favored.

Since the acids used in chromium determination, especially nitric acid, are a source of contamination, it is advisable to use reagents of high quality or to distil them previously. Reagent blanks must be used to monitor such contamination.

Contact with materials that can give up chromium, i.e., metallic surfaces such as stainless steel, should be avoided. Therefore, knives, mixers, homogenators, or mills which are made from stainless steel should be avoided and replaced by others made from polymeric materials, quartz, or titanium. Colored polymeric materials such as screw-caps for bottles, dispensers, and pipette tips can contain chromium and should be controlled for contamination.

Chromium determinations below the $1\,\mu g\,l^{-1}$ level require strict contamination control and clean-room facilities.

Sample Digestion

The choice of the digestion method is important in trace chromium determination. The use of dry methods carries the risk of the formation of acid-insoluble Cr(III) oxide and can lead to chromium losses by adsorption on the walls of the crucible or by volatilization after formation of chromyl chloride at temperatures higher than 550 °C. These losses are minimized when the temperature is increased gradually and does not exceed 450–500 °C. Nevertheless, some authors have reported losses of chromium by volatilization at 450 °C and obtained better results at lower temperatures together with oxygen plasma. In the latter procedure, the amount of sample is limited and relatively large amounts of hydrogen peroxide are needed as an ashing aid. A dry-ashing in Pt crucibles with a temperature ramp at a rate of $50\,^{\circ}C\,h^{-1}$ up to 450 °C is the digestion method adopted by the Nordic Committee on Food Analysis as its official methodology for chromium determination in foodstuffs.

In wet digestion procedures, mixtures of acids and oxidizing agents are used, the most commonly used ones being nitric, sulfuric, and perchloric acids with hydrogen peroxide or vanadium pentoxide. When foods have a high fat content, mixtures of nitric, perchloric, and sulfuric acids, or nitric and sulfuric acids, are recommended.

The high boiling point of sulfuric acid helps the activity of oxidizing agents and, with the addition of nitric acid, the disadvantage of forming sulfate compounds of low solubility is overcome. Digestion, however, with nitric and sulfuric acids requires good refluxing to prevent losses.

Significant losses of chromium during wet ashing occur when perchloric is present and the sample is concentrated close to dryness because of chromyl chloride (CrO_2Cl_2) formation, which boils at 117 °C. Formation of chromyl chloride can be minimized by the addition of sulfuric acid.

The main risk of wet digestion is contamination through the reagents, especially when high volumes are used. Digestions in closed systems and/or in microwave systems require less time and reagents. Therefore, the risk of contamination decreases and losses by volatilization are also reduced to a minimum.

Microwave digestion is now widely used to measure chromium contents in foods, because it is faster, safer, and provides more reproducible conditions than conventional methods, resulting in better analytical precision.

In some vegetable species chromium is retained by an insoluble residue of silica. When this occurs, treatment with hydrofluoric acid is necessary to overcome the problem of low recovery values.

Extreme care must be taken in all wet ashing procedures where strong acids and strong oxidizing agents are used. The formation of explosive nitro compounds has been reported when fat-rich food samples have been digested with nitric acid.

Determination

Molecular Absorption Spectroscopy

The traditional spectrophotometric method of 520 nm, based on the violet complex formed with 1,5-diphenyl-carbazide (DPC), is still widely used for determining chromium in water, because of its high sensitivity. Chromium must be present as Cr(VI) and the diphenylcarbazide is oxidized to the diphenylcarbazone. Iron, copper, molybdenum, and vanadium can interfere. To remove interference and to increase the sensitivity of the method, both cation exchange resins (iron and copper) and cupferron (iron, copper, molybdenum, and vanadium) have been used.

This method has been applied to chromium determination in foods after destroying the organic matter and removing the interferences. Detection limits of $10\,ng\,g^{-1}$ have been reported.

The complex chromium hematoxiline, with a maximum absorbance between 360 and 390 nm, has also been applied to foods, although less often than the diphenylcarbazide method.

Flame Atomic Absorption Spectrophotometry (FAAS)

The low levels of chromium in foods and the complexity of the matrix make it difficult to measure chromium directly in digested samples by FAAS. With an air-acetylene flame and at a wavelength of 357.9 nm it is possible to obtain detection limits of 0.1 $\mu g\,g^{-1}$. It has been well established that chromium atomization is dependent on flame stoichiometry, with higher sensitivity with reducing flames (rich in acetylene). A 10-fold difference in sensitivity between fuel-lean and fuel-rich flames has been reported. However, the composition of the flame affects the behavior of the interfering compounds, and often the conditions for maximum signal-to-background are quite different from those required for minimizing interferences.

This determination is not free of interelement interferences. Iron depresses chromium absorbance in fuel-rich air-acetylene flames. It can be separated from the analyte by extraction with acetylacetone/chloroform (1:1, v/v) at pH 1–2, or by extraction, as Fe(III), from 5 $mol\,l^{-1}$ hydrochloric acid into isobutyl acetate or 4-methyl-2-pentanone (MIBK).

Direct chromium determination by FAAS in digested samples is not often applied, owing to interferences and the low concentration of chromium. For this reason it is more usual to enrich and to remove interference by chelating chromium and then extracting with an organic solvent.

Chromium can be chelated as Cr(III) with 2,4-pentanedione at pH 7.5 and heating to nearly 80 °C. The chelate is then extracted with chloroform or MIBK. With the former, a backextraction to an aqueous solution should be carried out before measuring by FAAS. To remove metal interferences the same chelating reagent or ammonium pyrrolidine dithiocarbamate (APDC) is added at a lower pH value and then extraction with chloroform is carried out.

Alternatively, chromium can be chelated with APDC as Cr(III) (pH 5 at 80–90 °C) or as Cr(VI) (pH 3–9 at room temperature). Therefore, to work under less strict conditions it is advisable to oxidize chromium to Cr(VI), and to eliminate the excess oxidizing agent. The procedure is not free of interferences, because manganese and tin are also chelated. In order to remove interfering substances and to increase the stability of the element in solution, some authors have proposed that backextraction should be carried out.

Another method that has been applied to chromium determination in foods is based on the formation of a compound by ionic association between Cr(VI) and the system hydrochloric acid/MIBK (0 °C). Prior to this, chromium is oxidized to Cr(VI) with potassium permanganate. This procedure offers the advantage that it is not necessary to remove the excess oxidizing agent.

Electrothermal Atomization Atomic Absorption Spectrophotometry

ETA-AAS is considered one of the techniques of choice for chromium determination in biological matrices and it is the technique most often used to determine chromium in foods. It is not free of interferences, but it is simpler than FAAS solvent-extraction methods. In foods, detection limits of 1 $\mu g\,kg^{-1}$ have been reported.

If organic matter is present, chromium is found mainly in the form of Cr(III), and in this oxidation state it easily forms refractory oxides during heating in graphite tubes and then reacts with graphite to form a carbide. To avoid the resulting decrease in the signal, pyrolytically or specially coated graphite tubes should be used. Pyrolytically coated tubes increase sensitivity threefold with respect to uncoated ones.

When the furnace temperature reaches 450 °C, breakdown products of chromium complexes are formed. Some of these are volatile and, therefore, there is a risk of losses of the element. Alternatively, other breakdown products can prevent complete atomization of chromium, because of the strong bonding with oxygen, nitrogen, or carbon. To achieve complete atomization of chromium, temperatures higher than 2300 °C should be used. The usual ranges of temperatures are: drying (100–200 °C), ashing (600–1600 °C), atomization (2300–2700 °C).

Diammonium hydrogen phosphate, magnesium nitrate, and palladium have been used successfully as matrix modifiers. The use of magnesium nitrate as a matrix modifier extends the possible charring temperature to 1600 °C, thus insuring a more complete remove of inorganic salts.

Since chromium is measured at a wavelength of 357.9 nm, it is better to correct background absorption with the Zeeman effect or a tungsten lamp rather than with a deuterium lamp. The emission intensity of the deuterium arc is low at the chromium resonance line of 357.9 nm. The hollow cathode lamp current has to be reduced in order to balance the source and background beams. The Zeeman effect can compensate for high absorbances, such as those produced by biological matrices.

The standard addition method is generally applied, even when prior digestion of organic matter is carried out. In the case of liquid samples, e.g., beer or wine, it is possible to determine chromium by injecting the

diluted sample directly into the graphite furnace and then applying the standard addition method. This direct procedure has also been applied to suspensions of vegetables and milk samples (whole, low-fat, skim, condensed, evaporated, and powdered). In these cases slurries of the samples were prepared with Triton X-100. Chromium contents below $0.2\,\mu g\,g^{-1}$ are measurable.

Inductively Coupled Plasma Spectrometry

The detection limits of ICP with atomic emission (ICP-AES) are lower than those obtained by direct FAAS, but not commensurate with ETA-AAS. Moreover, they can be seriously affected by spectral interferences, which are much more usual in ICP-AES than in AAS. A wavelength of 267.716 nm is usually chosen, though some overlap with emission lines of manganese and phosphorus can be observed. Other emission lines used are 283–563, 284–325, and 357–869 nm, but all are prone to spectral interference. Algorithms are generally used to correct for such interferences.

One of the main advantages of ICP is that sample preparation is simpler than in ETA-AAS, because elements can be measured when organic residues are present and several elements can be measured simultaneously. This explains why it is used in chromium determination in foods.

Online coupling mass spectrometry (MS) with ICP improves the detection limits 10–100-fold, obtaining, in optimized conditions, values below $0.1\,\mu g\,l^{-1}$ and therefore close to those of ETA-AAS. ICP-MS has the advantage of multielement capability. An ICP-MS method was proposed as the Association of Official Analytical Chemists (AOAC) official method for chromium determination in drinking, groundwater and waste waters. On the other hand, the lower background and greater sensitivity of the double-focusing instruments at low-resolution mode make it possible to obtain better detection limits than with quadrupole-based instruments.

Chromium determination in food samples by flow-injection ICP-MS has been described. The digests obtained in a wet high-pressure microwave system are delivered at a previously established flow rate to the ICP-MS instrument. FIA gives good recovery values and these are more reproducible than when a conventional method is used.

Electrochemical Methods (Polarography, Voltammetry)

Differential pulse polarography (DPP) is used to measure chromium in water and biological materials (including foods). Chromium is only detectable as Cr(VI). A reduction peak is observed at -0.80 V, when sodium citrate is used as a buffer (pH 9–10.5) and ammonia (ammonium chloride) as a support electrolyte. Detection limits as low as $0.12\,\mu g\,l^{-1}$ can be obtained.

DPP is mainly used for chromium determination in water. Preparation of the sample depends on its complexity and the expected chromium concentration.

Nitric acid should not be used to acidify the sample, since the formation of nitrous acid can partially reduce Cr(VI) to Cr(III). If water contains sulfites, these can also reduce Cr(VI) to Cr(III), so Cr(VI) must be chelated with sodium diethyldithiocarbamate (pH 4, diluted sulfuric acid/ potassium acid phthalate) and extracted with MIBK. Sulfuric acid is added and the solution is heated to remove the solvent. Hydrogen peroxide and sulfuric acid are added to insure that chromium is present entirely as Cr(VI).

In the case of samples with a high chloride or organic matter content, sodium sulfite is added to prevent the formation of chromyl chloride.

Samples having chromium levels lower than $1\,\mu g\,g^{-1}$ can be enriched by passing the solution through an ion exchange resin.

Voltammetric chromium measurements have been applied to wet digested biological samples after a controlled adsorptive preconcentration step on a hanging-mercury-drop electrode. Optimal conditions were a preconcentration potential of -1 V, a preconcentration period of 60 s, and a solution pH of 6.3. A current peak was recorded at -1.1 V.

Electrochemical determinations can also be useful to differentiate between chromium species.

Gas-Liquid Chromatography (GLC)

The great sensitivity of GLC allows the detection of extremely small amounts of chromium. GLC has been applied mainly to chromium determination in biological fluids, such as blood and urine, and to chromium determination in water, when speciation is wanted. It is applied less frequently to foods, because this method requires prior destruction of the organic matter and chromium levels are higher in foods than in biological fluids.

Since GLC can only be applied to volatile compounds, volatile chromium chelates are formed with fluorine substitutes of acetylacetone derivatives. Chromium complexes with β-diketones are volatile and stable, and trifluoroacetyllactone (TFA) is the chelating agent most frequently reported in practical applications, because of the enhanced sensitivity of halides using electron capture detection (ECD). This determination requires a purification step to remove the excess chelating reagent and other interfering

substances from the detector. An ion exchange resin can be used for this purpose. (*See* **Chromatography**: Gas Chromatography.)

ECD permits the detection of 0.1 pg of chromium injected as $Cr(TFA)_3$ chelate but, at these levels, serious complications arise in the separation of the metal chelate from interfering electron capture impurities. For this reason, several other detectors have been evaluated, among them the nitrogen-phosphorus detector, atomic absorption spectroscopy and MS. The hyphenated GLC-MS techniques provide a positive identification for the very small amounts of chromium compounds. (*See* **Mass Spectrometry**: Principles and Instrumentation.)

GLC application to foods is more difficult than in biological fluids. For instance, in liver, many of the metals present, such as iron, copper, cobalt, and nickel, tend to emerge shortly after the solvent peak in large trailing peaks. These are probably thermal decomposition products, and any chromium peak present is masked.

The introduction of the mass spectrometer as a detector in food testing meant a substantial improvement over the ECD, in particular the compatibility of monitoring a peak which can be identified with a single metal, and freedom from interferences. The detection limit is around 0.5 pg.

Neutron Activation Analysis

Neutron activation analysis (NAA) is based on the formation of radioactive isotopes, when a sample is irradiated with neutrons from a nuclear reactor or from other neutron sources. The elements of interest are identified and quantified by γ-spectroscopy.

It is generally used as a reference method to test the accuracy of other analytical methods, because it is highly specific and accurate. One of its main advantages is that it is a nondestructive method. The chief limitations to widespread use of this technique are the need for a neutron source and the very high cost of the analysis, that preclude the use of this method for routine analysis.

Moreover, in the case of chromium, the low abundance of ^{50}Cr (4.31%), a small thermal neutron section (17 barns), a low level of γ-decay and a half-life for ^{51}Cr of 27.8 days make it necessary to apply long irradiations with a high-density flux of electrons, especially when chromium concentrations are low. ^{51}Cr is separated by distillation of chromyl chloride.

NAA has also been used to investigate the behavior of chromium during drying or when the matrix is subjected to high temperatures in dry ashing, because it allows the direct determination of the sample without previous destruction of the organic matter.

Another application of NAA (radiochemical method) is to measure chromium in biological reference materials for certification purposes.

Chromium (III) and (VI) speciation

Metal speciation refers to the identification and quantification of organometallic, chelated, or free metal ion forms or their oxidation states.

As has already been mentioned, differences between the toxicity and bioavailability of Cr(III) and Cr(VI) make it necessary to ascertain not only the total chromium content but also the species present. Cr(III) plays an essential role in the maintenance of the glucose tolerance factor, while Cr(VI), because of its high oxidation potential and its relatively small size, enables it to pass through cell membranes, is toxic, and carcinogenic.

Trace metal speciation is achieved by combining two different techniques: one providing an efficient and reliable separation procedure and the other detection and quantification. Since liquid chromatogaphy is suitable for the separation of ionic, polar, and nonpolar compounds, it is recognized to be specifically efficient in the speciation of organometallics. At trace levels the chromatographed species are usually quantified by atomic absorption or emission techniques. Among the latter, most applications are based on ICP-AES or ICP-MS. When coupled with separation procedures, ICP-MS is one of the most powerful techniques for speciation analysis, and, when ICP is used as a detector for high-performance liquid chromatography (HPLC), it offers good sensitivity, a dynamic range of over five orders of magnitude and multielement detection capabilities. Moreover, in comparison with ICP-AES, ICP-MS offers the advantages of lower limits of detection (sub-ng to sub-pg levels), wide linear ranges, isotope analysis capability, and good precision values (0.1–5.0% RSD). AAS is also a sensitive detector after chromatographic separation.

A review of the literature reveals a large number of procedures for chromium speciation in water and soils. These include the Cr(VI) determination methods based on three main procedures – a selective chelation/extraction technique, coprecipitation and column separation, together with Cr speciation by hyphenated HPLC methods.

However, chromium speciation has rarely been performed in complex matrices such as foodstuffs. In these products total chromium is usually determined and then Cr(VI) or Cr(III) is separated and measured. Examples of these determinations have been reported in milk and in solid food samples.

A method has been reported for determining total chromium directly by ETA-AAS in milk after the addition of a surfactant and Pd-Mg as a chemical modifier. To extract Cr(VI) quantitatively and selectively a solid-phase extraction method using a silica column modified with NH_2 is applied to the supernatant obtained by protein precipitation.

On solid samples alkaline digestion is followed by coprecipitation with iron (III) hydroxide, redissolution with HCl and determination of Cr(III) by FAAS. This is used together with total chromium determination after a wet digestion of the sample. The amount of Cr(VI) is then calculated by substracting Cr(III) from total Cr. A positive interference was reported when the content of Cr(VI) was 10 times larger than Cr(III).

Conclusion

ETA-AAS with Zeeman background correction and matrix modification appears to be the most useful method for routine, and relatively rapid determination of the total chromium content of foods. Special precautions should be adopted to avoid sample losses and/or contamination throughout the period of analysis, i.e., from sampling to measurement. Therefore, accurate chromium determination requires the use of clean room facilities, filtered air and water, and ultrapure reagents to control environmental pollution. The daily use of appropriate reference materials is essential to check analytical accuracy. At the moment NAA remains useful as a reference method.

In the future the hyphenated HPLC techniques may contribute to chromium speciation.

See also: **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **Fish Farming**; **Mass Spectrometry**: Principles and Instrumentation

Further Reading

Boumans PWJM (ed.) (1987) *Inductively Coupled Plasma Emission Spectroscopy. Part II: Applications and Fundamentals*. New York. John Wiley.

Fishbein L (1984) Overview of analysis of carcinogenic and/or mutagenic metals in biological and environmental samples. 1. Arsenic, beryllium, cadmium, chromium and selenium. *International Journal of Environmental Analytical Chemistry* 17: 113–170.

Hoenig H and De Borger R (1983) Particular problems encountered in trace metal analysis of plant material by atomic absorption spectrometry. *Spectrochimica Acta* 38B: 873–880.

Maslowska J (1996) Metal contamination. In: Nollet LML (ed.) *Handbook of Food Analysis*, vol. 2, pp. 1665–1682.

Rollinson CL (1975) In: Trotman-Dickenson AF (ed.) *Comprehensive Inorganic Chemistry*, vol. 3, pp. 631–637. Oxford: Pergamon Press.

Rubio R, Sauquillo A, Rauret G and Quevauvillier PH (1992) Determination of chromium in environmental and biological samples by atomic-absorption spectroscopy: a review. *International Journal of Environmental Analytical Chemistry* 47(2): 99–128.

Sarzanini C (1999) Review: liquid chromatography: a tool for the analysis of metal species. *Journal of Chromatography A* 850: 213–228.

Slavin W (1984) *Graphite Furnace AAS: A Source Book*. Perkin Elmer.

Sperling M (1995) Chromium. In: Townshend A (ed.) *Encyclopaedia of Analytical Science*, vol. 2, pp. 729–743. Academic Press.

Torgrimsen T (1982) Analysis of chromium. In: *Biological and Environmental Aspects of Chromium*, pp. 65–97.Amsterdam: Elsevier.

Physiology

R A Anderson, Beltsville Human Nutrition Research Center, Beltsville, MD, USA

Background

Chromium in the trivalent form is an essential nutrient that functions in carbohydrate, lipid, and nucleic acid metabolism. Dietary intake of Cr by humans and farm animals is often suboptimal. Insufficient dietary intake of Cr is associated with increased risk factors associated with type 2 diabetes mellitus (DM) and cardiovascular diseases. Chromium functions in glucose and insulin metabolism primarily via its role in the regulation of insulin. Adequate dietary Cr leads to a normalization of insulin leading to decreased blood glucose in subjects with elevated blood glucose, increased blood glucose in subjects with low blood sugar, and no effect on subjects with near optimal glucose tolerance. Improved insulin function is also associated with an improved lipid profile. Chromium in foods and dietary supplements is trivalent, whereas Cr often found in paints, welding fumes, and other industrial settings is hexavalent and is severalfold more toxic than the trivalent nutritional Cr.

Absorption, Transport, Storage, and Excretion

Absorbed Cr is excreted primarily in the urine, and only small amounts of Cr are lost in the hair,

perspiration, and bile. Therefore, urinary Cr excretion can be used as an accurate estimation of absorbed Cr. At normal dietary Cr intakes (10–40 μg per day), Cr absorption is inversely related to dietary intake. Chromium absorption is approximately 0.5% at a daily intake of 40 μg and increases to 2% when intake drops to 10 μg. Therefore, the amount of absorbed Cr over this range is approximately 0.2 μg and is reflected in the urinary Cr losses of approximately 0.2 μg per day. This inverse relationship of Cr intake and absorption appears to be a basal control mechanism to maintain a minimal level of absorbed Cr. Intakes above 40 μg result in corresponding increases in total Cr absorbed.

Chromium absorption in young and old normal subjects is similar, but people with type 1 DM absorb two- to fourfold more Cr than other groups of subjects tested. People with diabetes appear to have an impaired ability to convert inorganic Cr to a useable form. Diabetic mice also lose the ability to covert Cr to a useable form. People with diabetes require additional Cr, and the body responds with increased absorption, but the absorbed Cr cannot be utilized effectively and is excreted in the urine. The chromium content of tissues of people with diabetes is also lower.

Chromium absorption and incorporation into tissues are also dependent on the form of Cr ingested. An accurate estimation of Cr absorption and utilization in animal studies can be achieved by measuring Cr incorporation into tissues. The tissue with the greatest Cr concentration is the kidney, followed by the spleen, liver, lungs, heart, and skeletal muscle.

Tissue Cr is an accurate method to assess Cr absorption and utilization and is also a measure of Cr storage. The kidney, which is one of the primary sites of tissue Cr storage, is also one of the best sources of insulin potentiating forms of Cr. Chromium is transported to the tissues primarily bound to transferrin, the same protein that transports iron. There are two metal-binding sites on transferrin, one primarily for iron and a second involved in Cr transport. During conditions of high iron excess or iron overload such as in iron storage diseases (hemosiderosis; hemochromatosis), all the metal transport sites on transferrin are occupied by iron. This may explain the high incidence of diabetes in hemochromatosis patients that may be induced by Cr deficiency.

Metabolic Functions and Essentiality

The essentiality of Cr in human nutrition was documented in 1977 when diabetic signs and symptoms of a patient on total parenteral nutrition (TPN) were reversed by supplemental Cr. Diabetic symptoms, in addition to elevated blood glucose, included weight loss, impaired nerve conduction, brain disorders and abnormal respiratory quotient that were refractory to exogenous insulin. Upon daily addition of supplemental Cr to her TPN solution for 2 weeks, diabetic symptoms were alleviated, and exogenous insulin requirement dropped from 45 units per day to zero.

Table 1 Signs and symptoms of Cr deficiency

Function	*Animals*
Impaired glucose tolerance	Human, rat, mouse, squirrel monkey, guinea-pig, cattle
Elevated circulating insulin	Human, rat, pig, cattle
Glycosuria	Human, rat
Fasting hyperglycemia	Human, rat, mouse
Impaired growth	Rat, mouse, turkey
Hypoglycemia	Human
Elevated serum cholesterol and triglycerides	Human, Rat, mouse, cattle, pig
Increased incidence of aortic plaques	Rabbit, rat, mouse
Increased aortic intimal plaque area	Rabbit
Nerve disorders	Human
Brain disorders	Human
Corneal lesions	Rat, squirrel monkey
Ocular eye pressure	Human
Decreased fertility and sperm count	Rat
Decreased longevity	Rat, mouse
Decreased insulin binding	Human
Decreased insulin receptor number	Human
Decreased lean body mass	Human, pig, rat
Elevated percentage body fat	Human, pig
Impaired humoral immune response	Cattle
Increased morbidity	Cattle

Adapted from Anderson RA (1998) Chromium glucose intolerance and diabetes. *Journal of American College of Nutrition* 17: 548–555.

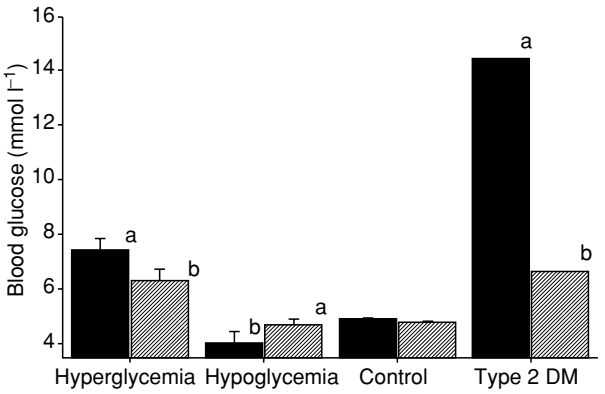

Figure 1 Response to supplemental Cr of people with hyperglycemia, hypoglycemia, optimal glycemia (control), and type 2 DM. The minimal amount of Cr usually showing beneficial effects in people with high or low blood sugar is 200 μg per day. People with diabetes require 400 to 600 μg per day or more.

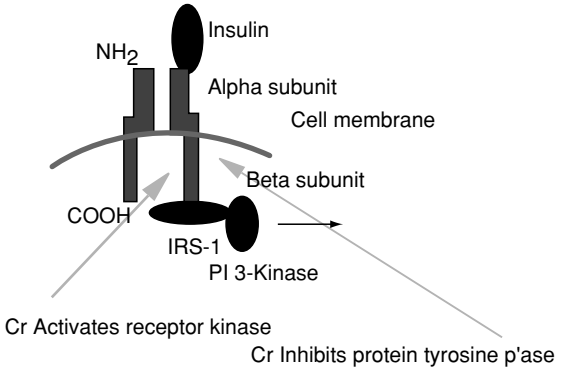

Figure 2 Postulated mechanism of action of chromium. When insulin binds to the α subunit, phosphorylation is increased, leading to increased insulin sensitivity and action. Chromium increases insulin binding to cells by increasing the insulin receptor number, which varies depending upon the type of cell, and also increases insulin sensitivity by increasing insulin receptor phosphorylation. IRS-1 denotes the insulin receptor substrate 1, and PI 3-kinase denotes phosphoinositol-3-kinase. For original references and details see Anderson RA (1998) Chromium, glucose intolerance and diabetes. *Journal of American College of Nutrition* 17: 548–555 with permission.

These findings have been repeated and documented in the scientific literature on several occasions.

Signs and symptoms of Cr deficiency listed in Table 1 are not limited to subjects on total parenteral nutrition. Improvements in glucose and/or lipid concentrations have been reported in children with protein-calorie malnutrition, the elderly, people with type 1 and type 2 DM, hypoglycemia, and marginally impaired glucose tolerance, and numerous animal species have also been shown to be improved by increased Cr intake.

The hallmark sign of marginal Cr deficiency is impaired glucose tolerance. The effects of Cr on people with high, low, and normal glucose tolerance as well as diabetes are illustrated in Figure 1. Chromium leads to a decrease in blood glucose in people with elevated blood sugar and an increase in those with low blood sugar owing to its role in normalizing insulin. In the presence of Cr in a physiologically active form, insulin is more efficient, and much lower levels of insulin are required. During periods of elevated blood glucose, more efficient insulin leads to a decrease in blood glucose. In people with low blood sugar, reactive hypoglycemia, more efficient insulin leads to a rapid rise in response to a glucose challenge and a more rapid return to baseline values. This leads to less of a drop or a raising of the hypoglycemic glucose values. Supplemental Cr also leads to increased insulin binding and increased insulin receptor number, and recent evidence suggests that Cr may be involved in the phosphorylation–dephosphorylation of the insulin receptor proteins (Figure 2). Chromium activates insulin receptor kinase, the enzyme that phosphorylates the insulin receptor leading to activation of insulin function, and appears to inhibit the phosphatase enzyme that deactivates insulin function.

Assessment of Chromium Status

There are no accurate methods for reliable a priori detection of marginal Cr deficiency. Chromium

Table 2 Daily intakes of chromium from various food groups

Food group	*Average daily intakes (μg)*	*Comments*
Cereal products	3.7	55% from wheat
Meat	5.2	55% from pork; 25% from beef
Fish and seafoods	0.6	
Fruits, vegetables, nuts, and mushrooms	6.8	70% from fruits and berries
Dairy products, eggs, and margarine	6.2	85% from milk
Beverages, confectionaries, sugar, and condiments	6.6	45% from beer, wine, and soft drinks
Total	29.1	

From Anderson RA (1988) Chromium. In: Smith K (ed.) *Trace Minerals in Foods*, pp. 231–247. New York: Marcel Dekker.

Table 3 Chromium content of selected foods

	Cr ($ng\,g^{-1}$)[a]	μg per serving
Dairy products		
Whole milk, 1 cup = 244 g	< 0.50	< 0.12
Skim milk, 1 cup = 244 g	< 0.50	< 0.12
Butter, 1 pat = 5 g	12 ± 2	0.06
Margarine, 1 pat = 5 g	3 ± 1	0.02
American cheese, 1 oz. = 28 g	20 ± 1	0.56
Eggs, 1 egg = 50 g	4 ± 1	0.20
Meats, poultry, and fish		
Beef cubes, 3 oz. = 85 g	24 ± 2	2.0
Chicken breast, 3 oz. = 85 g	6 ± 1	0.50
Turkey breast, 3 oz. = 85 g	20 ± 1	1.7
Turkey ham, 3 oz. = 85 g	122 ± 19	10.4
Ham, 3 oz. = 85 g	42 ± 4	3.6
Haddock, baked, 3 oz. = 85 g	7 ± 1	0.60
Grain products		
Waffles, 1 waffle = 75 g	89 ± 6	6.7
Whole wheat bread, 1 slice = 25 g	39 ± 1	0.98
Whole wheat roll, 1 roll = 26 g	23 ± 1	0.60
English muffin, WW, 1 muffin = 100 g	36 ± 1	3.6
Dinner roll, 1 roll = 25 g	25 ± 3	0.62
Rye bread, 1 slice = 25 g	37 ± 6	0.92
Bagel, egg, 1 bagel = 55 g	46 ± 7	2.5
Crackers, trix, 2 crackers = 14 g	9 ± 2	0.12
Spaghetti, 1 cup = 140 g	2 ± 1	0.28
Rice, white, 1 cup = 165 g	7 ± 1	1.2
Rice, brown, 1 cup = 165 g	4 ± 1	0.66
Fruits and vegetables		
Banana, peeled, 1 med = 126 g	8 ± 2	1.0
Grapes, 10 grapes = 50 g	2 ± 0	0.10
Apple, peeled, 1 med = 135 g	3 ± 1	0.40
Apple, unpeeled, 1 med = 150 g	9 ± 1	1.4
Orange, peeled, 1 med = 131 g	3 ± 1	0.39
Juice, grape, 1 cup = 250 g	30 ± 1	7.5
Juice, orange, 1 cup = 248 g	9 ± 2	2.2
Peas, 1 cup = 160 g	5 ± 1	0.80
Tomato, 1 med = 135 g	7 ± 1	0.94
Lettuce, 1 wedge = 135 g	13 ± 2	1.8
Celery, 1 stalk = 40 g	3 ± 1	0.12
Green beans, 1 cup = 135 g	16 ± 0	2.2
Carrots, raw, 1 carrot = 72 g	4 ± 1	0.29
French fries, 10 strips = 50 g	24 ± 1	1.2
Potatoes, mashed, 1 cup = 210 g	18 ± 1	2.7
Broccoli, 1 cup = 185 g	118 ± 0	22.0
Juice, tomato, 1 cup = 240 g	6 ± 1	1.5
Condiments		
Dressing, Thousand Island, 1 tbsp = 16 g	11 ± 1	0.18
Dressing, French, 1 tbsp = 15 g	12 ± 0	0.18
Pepper, 1 packet = 0.25 g	145 ± 15	0.04
Salt, 1 packet = 1 g	5 ± 1	0.01
Mayonnaise, 1 tbsp = 15 g	2 ± 1	0.03
Mustard, 1 tsp = 5 g	48 ± 6	0.24
Ketchup, 1 tbsp = 17 g	58 ± 0	0.99
Barbecue sauce, 1 tbsp = 16 g	108 ± 8	1.73
Maple syrup, 1 tbsp = 21 g	25 ± 3	0.52
Jelly, grape, 1 packet = 14 g	14 ± 1	0.20
Vinegar, 1 tbsp = 15 g	25 ± 4	0.38
Sugar, 1 packet = 6 g	3 ± 1	0.03
Miscellaneous		
Fruit punch, 1 cup = 200 g	3 ± 1	0.60
Orange sherbet, 1 cup = 193 g	7 ± 1	1.4
Apple pie, 1/7 = 136 g	11 ± 1	1.5
Cookies, oatmeal raisin, 4 cookies = 52 g	18 ± 3	0.94
Cookies, vanilla sand., 4 cookies = 42 g	16 ± 4	0.64
Cookies, choc. chip, 4 cookies = 42 g	82 ± 30	3.4
Peanut butter, 1 tbsp = 16 g	38 ± 3	0.61
Soybeans, 1 tbsp = 9 g	22 ± 6	0.20

[a]Values are mean ± SD of triplicate analyses.
From Anderson RA, Bryden NA, and Polansky MM (1992) Dietary chromium intake – freely chosen diets, institutional diets and individual foods. *Biological Trace Element Research* 32: 117–121.

concentrations in hair, urine, blood, tissues, etc. can be used to assess recent Cr exposure but are not long-term measures of Cr status. The only reliable indicator of Cr status is to monitor glucose, insulin, lipid and/or related variables before and after Cr supplementation. Response in blood glucose can often be seen in 2 weeks or less, whereas effects on blood lipids may take longer. One study reported no effects of supplemental Cr on blood lipids after 3 months but significant effects after 7–16 months.

Requirements

The estimated safe and adequate daily dietary intake (ESADDI) for chromium for children 7 years to adult of 50–200 μg per day was established by committees of the US National Academy of Sciences in 1980 and affirmed in 1989. The ESADDI is similar to an RDA and is usually established prior to the RDA. The Food and Drug Administration proposed a Reference Dietary Intake for Cr effective in 1997 of 120 μg per day. However, the new committee of the National Academy of Sciences representing North America has proposed that the normal intake of Cr should serve as the adequate intake, which is 20 μg for women and 30 μg for men over 50 and 25 for women and 35 μg for men aged 19–50 years. The French Conseil National d'Etudes et de Recherche sur la nutrition et l'Alimentation has proposed daily intakes of 55 μg for adult French women 19–65 years (60 μg per day for those over 65 years) and 65 and 70 μg, respectively, for the men.

Chromium requirement is related to the degree of glucose intolerance and diabetes. Glucose tolerance of subjects with mild glucose intolerance declines when consuming diets containing less than 20 μg per day of Cr for 5 weeks, whereas glucose tolerance of subjects with near optimal glucose tolerance is not altered by these low Cr intakes. People with diabetes also have a higher requirement for Cr. Chromium supplementation with 200 μg of Cr per day to people with mild glucose intolerance results in significant improvements, whereas the blood glucose of people with diabetes often does not respond. However,

daily Cr supplementation of 400 μg or more results in significant improvements in many people with diabetes.

Dietary Sources and High Intakes

Dietary Cr content of foods varies widely, and there are no comprehensive databases to calculate dietary Cr intake. The chromium content of foods is often erroneously high owing to Cr contamination during collection and analyses. For example, stainless steel blender blades are often used in the homogenization of foods, but stainless steel is approximately 18% Cr. In the presence of acidic foods, more Cr may leach from the blender blades than is present originally in the foods. Chromium is present in several food groups but at low levels. The distribution is similar among fruits, vegetables, dairy products, beverages, and meats, with lesser amounts from cereal products and small amounts from fish and seafood (Table 2). Specific Cr contents of selected foods are listed in Table 3. The chromium content of foods is a combination of the endogenous Cr present in the foods and the Cr introduced during the various stages of growing and processing. For example, fruit juices are often high in Cr, since Cr may leach from containers during processing and storage under acidic conditions. Processed foods such as meats also often contain elevated levels of exogenous Cr (Table 3).

Trivalent Cr, the form of Cr found in foods and in nutrient supplements, is one of the least toxic nutrients. The reference dose established by an expert panel of the US Environmental Protection Agency is 350 times the upper limit of the estimated safe and adequate daily dietary intake. The reference dose is defined as "an estimate (with uncertainty spanning perhaps an order of magnitude) of a daily exposure to the human population, including sensitive subgroups, that is likely to be without an appreciable risk of deleterious effects over a lifetime." This conservative estimate of safe intake has a much larger safety factor for trivalent Cr than almost any other nutrient. The ratio of the EPA reference dose to the ESADDI for chromium of 350 compares to roughly two to five for other trace elements such as zinc and manganese and five to seven for selenium. Chromium in the form of both Cr chloride and Cr picolinate fed to rats at several thousand times the ESADDI for humans (based on body weight) resulted in no detectable signs of toxicity.

In summary, dietary intake of Cr is suboptimal for most humans and farm animals. Increased intake of trivalent Cr leads to improved glucose and lipid metabolism. The physiological role of Cr appears to be primarily though the improved function of insulin. Increased intake of Cr leads to increased insulin binding, increased insulin receptor number, and increased phosphorylation of the insulin receptor proteins leading to increased insulin sensitivity and function. There have been no documented negative effects of supplemental Cr in any of the more than 40 Cr supplementation studies involving daily Cr intakes of 1000 μg or more.

See also: **Coronary Heart Disease**: Etiology and Risk Factor; Antioxidant Status; Intervention Studies; **Diabetes Mellitus**: Etiology; **Glucose**: Glucose Tolerance and the Glycemic (Glycaemic) Index; **Trace Elements**

Further Reading

Anderson RA (1988) Chromium. In: Smith K (ed.) *Trace Minerals in Foods*, pp. 231–247. New York: Marcel Dekker.

Anderson RA (1998) Chromium, glucose intolerance and diabetes. *Journal of American College of Nutrition* 17: 548–555.

Anderson RA, Bryden NA and Polansky MM (1992) Dietary chromium intake – freely chosen diets, institutional diets and individual foods. *Biological Trace Element Research* 32: 117–121.

Anderson RA, Cheng NC, Bryden NA *et al.* (1997) Elevated intakes of supplemental chromium improve glucose and insulin variables of people with type II diabetes. *Diabetes* 46: 1786–1791.

Lukaski HC (1999) Chromium as a supplement. *Annual Review of Nutrition* 19: 279–302.

Morris B (1999) Chromium action and glucose homeostasis. *Journal of Trace Elements in Experimental Medicine* 12: 61–70.

Ravina A, Slezak L, Mirsky N, Bryden NA and Anderson RA (1999) Reversal of corticosteroid-induced diabetes mellitus with supplemental chromium. *Diabetic Medicine* 16: 164–169.

Vincent JB (2001) The bioinorganic chemistry of chromium(III). *Polyhedron* 20: 1–26.

CIDER (CYDER; HARD CIDER)

Contents

The Product and its Manufacture

B Jarvis, Ross Biosciences Ltd, Ross-on-Wye, UK

A Brief History of Cider

The fermentation of apple juice to produce an alcoholic beverage is believed to have been practiced for over 2000 years. Cider is recorded as being a common drink at the time of the Roman invasion of England in 55 BC. Celtic mythology revered the 'sacred apple,' and in his famous natural history (AD 77), Pliny the Elder refers to a drink made from the juice of the apple. Cider was drunk throughout Europe in the third century AD, and in the fourth century, St. Jerome used the term 'sicera' (whence the name cider was possibly derived) to describe drinks made from apples. Reputedly, cider was a more popular drink than beer in the eleventh and twelfth centuries in Europe. Cider is produced throughout the temperate regions of the world where apple trees flourish. Such localities include Austria, Belgium, England, France, Germany, the Northern coastal area of Spain, Switzerland, and, more recently, Argentina, Australia, Finland, New Zealand, South Africa, Sweden, and the USA.

There are many references to cider in writings from the Middle Ages. The popularity of cider in fourteenth century England was such that William of Shoreham reflected the Church's concern for the niceties of sacramental rites by stating that 'young children were not to be baptised in cider'! William Langdon in *Piers Plowman* and Shakespeare in *A Midsummers Night's Dream* make reference to the consumption of cider; Daniel Defoe observed that Hereford people 'boast the richest cider in all Britain,' and Samuel Pepys noted in his Diary that on 1 May 1666 he 'drank a cup of Syder.' From the seventeenth century onwards, cider was praised in numerous poems and other literary works as an aid to good cheer and a homely cure for almost every ailment known to man.

Prior to the twentieth century, cider was a popular rural drink, cheaper than beer and often more potent (at c. 7% alcohol by volume (abv)). Farm workers often had part of their wages paid as truck (i.e., in kind), and every farmer would make his own cider for consumption by his workers and his own family and guests – although it is recorded that the best ciders were retained for his personal use. Most farms in the West of England and the West Country had their own cider press, or used the services of a travelling cider press that was hauled by horse from farm to farm. My own cider mill and press (dated 1717) stand outside my window as I write this account!

Commercial cider production commenced during the nineteenth century in England (e.g., H Weston & Sons of Much Marcle, 1880; H P Bulmer of Hereford, 1887; Whiteways Cyder of Whimple in Devon, 1894), although a few farms produced cider commercially from as early as 1727 (e.g., Symonds Cider and English Wine Co. Ltd., Stoke Lacey, UK). Cider production in England was estimated as 250×10^6 l (55×10^6 gallons) in 1900. By 1920 the level of cider production had decreased significantly; although some 73×10^6 l (16×10^6 gallons) were still produced on farms only 23×10^6 l (5×10^6 gallons) came from factory producers. By the late 1980s, cider sales had risen to over 273×10^6 l (60×10^6 gallons) per year and, in the 12 months to September 1999, almost 500×10^6 l (109×10^6 gallons) were sold in the UK by members of the National Association of Cider Makers, a trade association formed originally in 1920; a further 23×10^6 l (5×10^6 gallons) of cider were exported from the UK. Since 1999, UK cider sales have remained constant at 500×10^6 l (109×10^6 gallons) p.a.

The significant increase in the volume of cider produced commercially each year in the UK up to 1999 followed a period when growth had been inhibited by the imposition of Excise Duty in 1976 and results both from the introduction of new cider products, linked to increased marketing activity, and from changes in drinking habits, especially of younger adults. In 2002, it is possible to obtain cider either on draught or prepackaged in glass and PET bottles and in cans. Products range in alcoholic strength from less than 0.5 to 8.5% (abv), and in sweetness from very dry to sweet. In addition, cider products now include 'white' ciders, naturally colored ciders, and blends of

cider with fruit juice or other ingredients. Some cider, made from apples of a single crop, may be sold as a defined year Vintage Cider, whilst others are made from a single apple cultivar, e.g., Kingston Black.

Because of the significant growth in cider sales in recent years, imitation ciders have appeared on the market in several countries. Such products are produced by compounding distilled alcohol with apple juice and artificial flavors. Such products are not cider. In 1998, the European Association of Cider and Fruit Wines Industries of the EU (AICV) defined the acceptable procedures for production of cider and related products. Fortification of cider or apple juice by the addition of distilled alcohol is not an approved process for a product to be sold as cider.

Cider Fruits

Traditional ciders are made from the juice of cider apples, believed to have been imported into Britain by the Normans, although it is recorded in Gaulmier's *Traite du Sidre* (1573) that a Spaniard named Dursus de l'Etre brought apple trees into France in 1486! Traditional cider apples are of four main types: bitter sweet – low in acidity but high in tannin; bitter sharp – high acidity and high tannin; sharp – high acidity but low tannin; and sweet varieties – low acidity and low tannin. Examples of some cider fruit cultivars are listed in Table 1.

In some areas, especially Kent, Suffolk, and Norfolk in the UK, cider is made primarily from culinary fruit varieties such as the Bramley, although blends of bittersweet and culinary juices are frequently used to develop particular flavor profiles in commercial cider blends.

Table 1 Characteristics of cider apple varieties

Type	*Typical varieties*	*Typical juice composition*	
		Acidity (g per 100 ml)	*Tannin (g per 100 ml)*
Sweet	Northwood Sweet Alford Sweet Coppin	< 0.45	< 0.2
Bitter sweet	Ashton Brown Jersey Dabinett Michelin Yarlington Mill	< 0.45	> 0.2
Sharp	Brown's Apple Frederick Reinette Obry	> 0.45	< 0.2
Bitter sharp	Bulmer's Foxwhelp Brown Snout Chisel Jersey Kingston Black	> 0.45	> 0.2

Cider Orchards

The traditional farm orchard of standard trees still exists, and although generally declining in acreage, many hundreds of such orchards provide both an apple crop and grazing for livestock.

Modern cider apple orchards are largely intensive bush orchards, where the trees are frequently planted as closely as 2.4 m apart in rows some 5.5 –6 m apart. After planting, staking, and protecting against rabbits, using a wire guard, the grass under the trees is treated with a suitable herbicide to reduce competition for nutrients and water. Some growers cover the herbicide strips with a mulch of straw or other suitable material to ensure maximum retention of moisture. Although it has become standard practice to retain the herbicide strip into the productive years of an orchard, this practice is currently the subject of research to assess whether a grass sward may be better in relation to the quality of harvested fruit.

During the growing season, it may be necessary to spray the trees against pests, such as red spider mite, and diseases such as mildew, apple scab, and canker. In years of heavy potential cropping, chemical thinning (e.g., using paclobutrazol, 'Cultar') is frequently recommended, to reduce stress on the tree and thereby minimize the risk of biennialism, a condition that occurs significantly in cider fruit orchards.

Harvesting

In traditional orchards, fruits are generally allowed to fall naturally, or shaken from the trees using long poles (lugs), and then picked up either by hand or by machine. Traditionally, fruit from standard trees would be raked into piles or filled into sacks that were stored under the trees until the fruit was in a suitably ripe condition to be milled and pressed.

In intensive bush orchards, it is normal to shake the trees mechanically to cause the fruit to fall. Such shaking does not harm the tree and permits fruits to be harvested mechanically immediately after falling, so reducing the risk of rots that may occur if fruit are left for any length of time on the ground. The collected fruit is normally washed and then transported by road to the cider mill where, after weighing, the fruit is tipped into a fruit canal or on to a concrete pad for holding prior to milling and juicing.

Processing the Fruit

Milling and Pressing

The fruit is generally transferred into the mill using either a water flume, which provides additional washing, on a conveyor belt or using screw conveyors.

Detritus such as twigs, leaves, and stones is removed mechanically, and the fruit is milled using a grater, slicer, or hammer mill. The apple pulp is then conveyed to a mechanical press where the juice is extracted.

Traditionally, the presses consist of a frame containing a slatted board covered by a cloth into which a measured amount of the pulp is transferred. The corners of the cloth are folded over to form an envelope, the frame is removed, and the next slatted board is added together with the frame and another cloth. This procedure is repeated some 10 –20 times to build a 'cheese,' which is then pressed hydraulically to remove the juice, under a ram pressure of some 14 MPa (*c*. 2 MPa on the 'cheese'). This method removes up to 80% of the juice. Sometimes, the pomace residue (see below) is soaked in a small volume of water and repressed to remove further quantities of fermentable sugar.

Such processes are very labor intensive and, except for farmhouse production, have now largely been replaced by automatic presses such as the Bücher-Guyer, a horizontal cylinder press, or the Bellmer continuous belt press. Such presses are more efficient, often permit countercurrent extraction of the pomace with warm water, and have little demand for labor once properly set up. Processes such as electro-acoustic dewatering have been claimed to release a higher yield of juice by passing an electric current through the pulp prior to pressing but have not found application in the cider industry.

Pomace

Apple pomace can be used for the extraction of the natural gelling agent, pectin, used in the manufacture of jam, dairy products, and other foodstuffs, and in certain medical applications.

Pomace, either before or after depectinization, can be fed to animals either as a wet slurry or after drying; it can also be spread on to agricultural land as an organic mulch or fertilizer. Although it aids the breakdown of clay soils, its acidity often poses problems.

Apple Juice

The extracted juice is rich in malic acid, tannins, and sugar; it will also be contaminated with a variety of microorganisms derived from the fruit itself, from the orchard floor, from the harvesting equipment, and from the environment of the cider mill. Immediately after pressing, juice is treated with sulfite, either as sulfur dioxide (SO_2) or in the form of potassium metabisulfite. Sulfite acts both as an antioxidant, to prevent browning of the juice (due to polyphenol oxidase and chemical browning reactions), and as an antimicrobial that destroys 'wild' yeasts and bacteria. The sulfite-treated juice will normally be stored for some 24 h before being pumped into fermentation vessels. The level of sulfite will normally be adjusted to give 10–30 mg of free SO_2 per liter at the time of pitching with yeast.

Generally, the juice will not be clarified prior to fermentation. However, if the juice is to be concentrated by thermal evaporation, it will be treated with a mixture of enzymes (pectinases and amylases) and then clarified prior to concentration. Failure to destroy pectins and starch results in a thick viscous mass that neither concentrates effectively nor stores adequately.

Fermentation of Cider

Fermentation Vessels

Traditionally, the apple juice would have been fermented in oak vats. Although many such vats are still used, they have been largely replaced in commercial practice by vats of mild steel with a ceramic or resin lining, by lined fiber-glass resin vats and, more recently, by stainless steel vats. A few cider makers use redundant brewing vessels such as tall conico-cylindrical vats. However, there is evidence that tall conicocylindrical vats inhibit effective fermentation by creating excessive hydrostatic pressure on the yeast, although squat conicocylindrical vats are excellent. The Unitank-style vessel used frequently for the fermentation of wine is also excellent for cider fermentation. Figure 1 illustrates some of the styles of fermentation vessels used in cider making.

Fermentation Substrate

Traditional cider making fermented the whole juice, often with much of the apple solids remaining. Such solids included the pips, which contain glycosides, and so the cider would contain small quantities of cyanide derivatives.

In many countries, the practice of chaptalization has become increasingly common. Chaptalization supplements the sugar content of the apple juice with a suitable fermentation sugar (e.g., brewers' dextrose) and permits the production of strong cider with an alcohol content of 11.5–12% abv. By comparison, fresh juice fermentations, depending upon the sugar content, will rarely exceed 5.5–6.5% abv. In France, where the production of traditional cider is constrained by legislation, only the product of fermentation of fresh apple juice or fresh juice reinforced with juice from concentrate, can be called '*cidre*.' However, it is now permissible in France to produce a cider

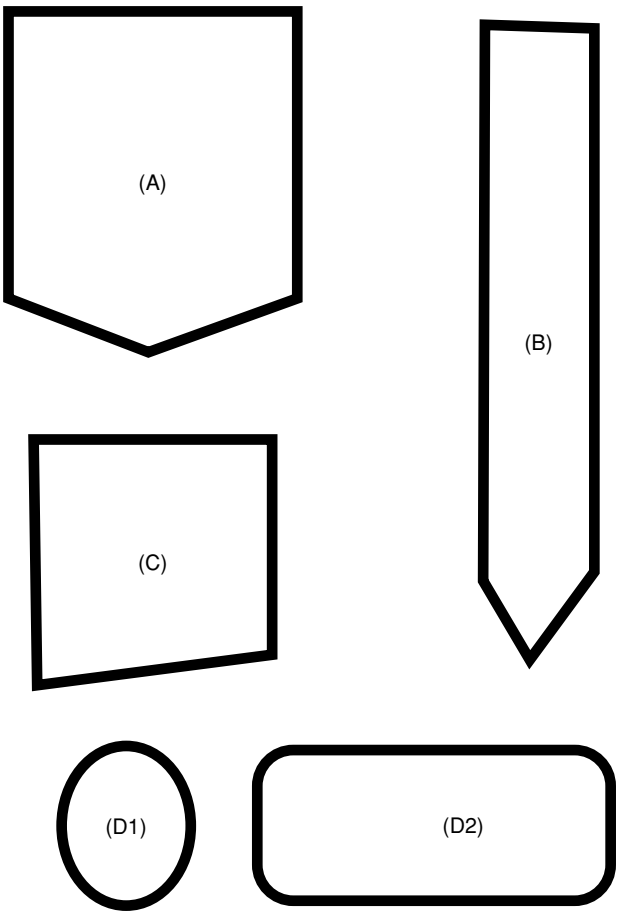

Figure 1 Schematic design of cider fermentation vessels. (A) Shallow angle conicocylindrical fermenter suitable for cider. (B) Typical steep angle, tall conicocylindrical fermenter as used for brewing beer. (C) Unitank fermentation vessel, with sloping base. (D) Oval cross-section (D1) horizontal fermentation vessel (D2).

from juice and fermentation sugar, which must be described as '*un boisson alcoolisée à base de pomme*' (*sic* an alcoholic beverage derived from apples) to differentiate it from traditional *cidre*.

In the preparation of cider, the apple juice may be fresh or reconstituted from apple juice concentrate; in many countries, pear juice (fresh or reconstituted) may be added to a maximum of 25% of the apple juice content. Other than fermentation sugars, the only other primary addition will be a suitable yeast culture. Fermentations originating wholly or largely from concentrate normally need the addition of yeast nutrients, such as ammonium phosphate, sodium pantothenate and/or thiamine.

Fermentation Process

Where a natural fermentation of the apple juice is required, the pressed and sulfite-treated juice is transferred to suitable fermentation vessels and allowed to ferment naturally. The process is monitored by measurement of the specific gravity (SG) of the juice; the fermentation is deemed to have gone to completion when the SG is 1.0000 or less.

However, most commercial cider is inoculated with a selected strain, or a mixture of strains, of a wine yeast capable of growth at elevated alcohol levels. The process consists of partially filling the vat with juice at a temperature at or below 18 °C, followed by inoculation with the specific starter culture of yeast. The use of a slightly reduced initial temperature is to prevent excessive temperature development during the initial exothermic growth of the yeast prior to onset of fermentation. The starter culture may be grown in the laboratory from a pure culture or added either as a commercial dried yeast preparation or as an aerated slurry of dried yeast – the latter is the most efficient for commercial processing. Once fermentation has started, additional substrate (i.e., apple juice and/or fermentation sugar) is added from time to time until the vat is filled. Although most cider vats are typically of $45 \times 10^3 - 9 \times 10^5$ l ($10 \times 10^3 - 2 \times 10^5$ gallons) capacity, much larger vats do exist.

At H P Bulmer's plant in Hereford, UK, the world's largest storage container for alcoholic beverages holds some 7.3×10^6 l (1.6×10^6 gallons) of product.

The fermentation typically continues with or without control of temperature, pH, or other parameters until all fermentable sugar has been metabolized into alcohol. This process can take from 10 days to 12 weeks, depending upon conditions! The modern cider fermentation plant generally has a facility to control the temperature, although ambient temperature fermentation still occurs widely (e.g., in farmhouse cider making). There is no doubt that good control of temperature can give a more rapid and more consistent fermentation. In France, cider is often fermented at temperatures below 15 °C. (*See* **Fermented Foods**: Origins and Applications.)

Maturation

When fermented to dryness, the cider is frequently left for a few days on the lees to permit the yeasts to autolyze, thereby adding cell constituents such as enzymes, amino and nucleic acids, etc. to the brew. However, if it is left for too long on the lees, there is a serious risk of development of off-flavors in the cider. The cider is separated from the lees (or tank bottoms) and transferred either directly, or after centrifugal clarification or filtration, into storage vats (traditionally made of oak). The actual process of maturation is generally uncontrolled, although, increasingly, modern commercial cider makers seek to control the storage temperature and the secondary malolactic fermentation. The maturation process is only slowly being understood: malolactic fermentation (carried out by various lactic acid bacteria) reduces the acidity by conversion of malic acid to lactic acid. Many other microbial and biochemical conversions also take place including modification of the tannins and esterification, e.g., of lactic acid to form ethyl lactate. Some of the chemical markers of maturation are now being identified, although judgment as to the extent of maturation and the suitability of the cider for use is still an art vested in the cider maker, who will have many years of experience in judging the quality of the product.

Final Processing

When required for use, different batches of cider, generally made from juices of different apple cultivars, will be blended by the cider maker to provide specific flavor attributes. The raw cider may be fined using agents such as bentonite, gelatin, or chitin and filtered to give a bright product with no haze. Modern processing refinements include the use of microfiltration systems to obviate the need for fining and to speed the process.

If the cider has a high alcohol content, it may be 'broken back' to final product strength using water or dilute apple juice. Other permitted ingredients may be added such as sugars, intense sweeteners (e.g., saccharin), color and/or additional preservative (normally limited to sulfite, although in some countries, benzoic or sorbic acids may be used). The finished blend may be treated with filter aids, such as kieselguhr (diatomaceous earth), to give an optically bright product or it may again be microfiltered. The most modern commercial plants use computer-controlled automated blending and microfiltration to convert the strong raw cider into commercial product blends.

Cider Packaging

Although a small market exists for 'live' cask-conditioned cider, i.e., cider in a wooden or plastic barrel, to which a small quantity of sugar has been added, together with a further yeast inoculum, the majority of commercial cider is carbonated and pasteurized, or sterile filtered, prior to filling into kegs, bottles, or cans. (*See* **Packaging**: Packaging of Liquids.)

Keg Cider

The cider is carbonated and pasteurized in line, through a continuous-flow plate heat exchanger. It is filled into stainless steel kegs in a plant that rinses, washes, and sterilizes the kegs prior to filling. The cider is dispensed in on-trade outlets using either carbon dioxide, or carbon dioxide and nitrogen, over-pressure through a cooling system designed to deliver the product into a glass at a temperature of 10 ± 0.5 °C. This process is similar to that used for dispensing keg beer.

Bottled Ciders

Cider to be filled into glass bottles may either be carbonated and flash pasteurized, or carbonated and then in-pack-pasteurized after filling. Since it is generally not possible to pasteurize PET bottles, the product will be flash-pasteurized prior to filling. Bottles will be sealed using either crown closures or tamper-evident metal or plastic screw caps. Glass bottles range in size from 25 cl to 1.13 l and PET bottles from 25 cl to 5 l. Carbonation pressures generally range from 2.5 to 3.5 bar, the higher initial pressures being used in PET bottles, which, due to gaseous diffusion, lose carbonation during storage. Glass bottled ciders have a shelf-life in excess of 2 years (provided that they are not opened!), whereas PET bottled ciders generally have a shelf life of 9–12 months, because of loss of carbonation.

Can Cider

The inside of cans for cider is always lacquered to prevent the product from attacking the metal. Cans are either of extruded aluminum or of mild steel, generally with a retained tag. Since sulfur dioxide is very corrosive to metal, especially if minute pinholes occur in the lacquer, cider for canning is generally prepared with little ($< 35\,mg\,l^{-1}$ total) sulfite. Such products must, of necessity, be prepared in much more controlled conditions than general blend ciders, since at these low sulfite levels, microbial contamination can lead to the formation of undesirable flavours. Cider filled into cans is always bed-pasteurized, generally at a process level of 30–40 pasteurization units (PUs). Control of dissolved oxygen levels for can ciders is also most important to prevent the development of oxidation off-flavors. Can ciders generally have a shelf-life of 9 months or less. (*See* **Canning**: Principles.)

Secondary Packaging

Bottles and cans are increasingly packaged using trays and shrink wraps, although the higher-value products (e.g., vintage and high-strength cider in glass bottles) are frequently packaged in carboard boxes with or without dividers.

Labeling of Cider Products

Throughout Europe, cider packs are required to conform to EU legislation on food labeling. At the present time, ingredient and nutritional labeling of alcoholic products is not required within the EU, although certain ingredients (e.g., intensive sweeteners) must be labeled. All alcoholic products, including cider, must display the alcohol content (as % abv) and the volume. The labeling requirements in other countries are dependent upon the national legislation.

Special Ciders

Vintage and Single Cultivar Ciders

Vintage ciders are made only from fresh juice from a named year. Some vintage and other ciders are made from the juice of a single named apple cultivar.

Sparkling Ciders

Sparkling cider is generally carbonated to a level of 3.5–4 bar pressure. Such products are filled into 'champagne-style' bottles with wired closures (generally plastic-mushroom stoppers). The product is normally sterile-filtered prior to bottling. Traditionally, sparkling cider received a secondary 'in-bottle' yeast fermentation (*méthode champanoise*), but such processing is rarely seen nowadays. A process used sometimes is that of *cuvée close*, in which a secondary yeast fermentation is done within a sealed tank, thereby developing a natural carbonation in the cider prior to bottling. Under EU legislation, it is illegal to refer to sparkling ciders as 'champagne cider.'

White cider

White cider is prepared by fermenting decolorized apple juice, or the fermented cider is itself decolorized by treatment with activated charcoal or other suitable decolorizing agent (e.g., PVPP) prior to final blending. The term '*white*' merely indicates that the product has little or no color – it is not 'white' in the sense that gin or vodka is 'water white.'

De-alcoholized and Low-Alcohol Ciders

De-alcoholized ciders are prepared by removing the alcohol from strong cider, by thermal evaporation, reverse osmosis, or other suitable technology to give a product with an alcohol content not in excess of 0.5% abv. De-alcoholized cider lacks body and flavor and is not sold commercially. Low-alcohol cider (< 1.2% abv) is prepared either using a stopped fermentation or by fortification of de-alcoholized cider with apple juice and/or other ingredients to provide a product with a flavor and aroma close to that of normal alcoholic cider.

Organic cider

Following consumer interest in organic foods, a number of cider makers now offer cider made only from fruit grown in accordance with EU Regulations governing organic horticultural practices. The fruit is pressed separately from nonorganic fruit, and at present, it can be treated only with gaseous sulfur dioxide. The other ingredients included in the product must also conform with current legislation on organic products.

See also: **Alcohol**: Properties and Determination; **Barrels**: Wines, Spirits, and Other Beverages; **Canning**: Principles; **Lactic Acid Bacteria**; **Preservatives**: Classifications and Properties; **Tannins and Polyphenols**; **Yeasts**

Further Reading

AICV (2000) *Code of Practice for the Production of Cider, Perry and Fruit Wine in the EU*. Revised 2000. Brussels: AICV.

Beech FW (1972) English cidermaking: technology, microbiology and biochemistry. In: Hockenhull DJD (ed.) *Progress in Industrial Microbiology*, vol. 11, pp. 133–213. Edinburgh: Churchill Livingstone.

Beech FW and Davenport RR (1983) New prospects and problems in the beverage industry. In: Roberts TA and Skinner FA (eds) *Food Microbiology: Advances and Prospects* (S.A.B. Symposium Series No. 11), pp. 241–256. London: Academic Press.

Charley VLS (1949) *The Principles and Practice of Cidermaking*. London: Leonard Hill.

Jarvis B (2001) Cider, perry, fruit wines and other alcoholic beverages. In: Arthey D and Ashurst PR (eds) *Fruit Processing*, 2nd edn. pp. 111–148. Gaithersburg, MA: Aspen Publishers Inc.

Jarvis B, Forster MJ and Kinsella WP (1995) Factors affecting the development of cider flavour. In: Board RG, Jones D and Jarvis B (eds) *Microbial Fermentations: Beverages, Foods and Feeds* (SAB Symposiuim Series No. 24) *Journal of Applied Bacteriology* 79 (supplement): 5S–18S.

Lea AGH (1995) Cidermaking. In: Lea AGH and Piggott JR (eds) *Fermented Beverage Production*, pp. 66–96.London: Blackie Academic & Professional.

Williams RR (ed.) (1991) *Cider and Juice Apples: Growing and Processing*. Bristol: University of Bristol.

Chemistry and Microbiology of Cidermaking

B Jarvis, Ross Biosciences Ltd, Ross-on-Wye, UK

Introduction

The fermentation of apple juice to cider can occur naturally through the metabolic activity of the yeasts and bacteria present on the fruit at harvest, which are then transferred into the apple juice on pressing. Other organisms, arising from the milling and pressing equipment and the general environment, can also contaminate the juice at this stage. Unless such organisms are inhibited, e.g., through the use of sulfur dioxide, the resulting mixed fermentation will yield a product that varies considerably from batch to batch, even if the composition of the apple juice for fermentation is identical.

Hence, control of the indigenous and adventitious microorganisms, followed by deliberate inoculation with a selected strain of yeast, is the preferred commercial route for the production of cider. Transfer of the fermented juice into (traditionally oak) maturation vessels will result in a secondary malolactic fermentation by microorganisms that occur naturally in these vats. Such organisms may produce beneficial or detrimental changes in the chemical and organoleptic properties of the final cider.

Microbiology of Apple Juice and Cider

Freshly pressed apple juice will contain a variety of yeasts and bacteria, many of which will be incapable of growth at the acidity of the juice. Examples of organisms often present in juice are shown in **Table 1**, together with an indication of their susceptibility to sulfur dioxide and ability to grow at the pH of the juice. (*See* **Microbiology**: Detection of Foodborne Pathogens and their Toxins.)

Role of Sulfur Dioxide in Apple Juice

The use of sulfur dioxide as a preservative in cider making is controlled by legislation, in most countries the maximum level permitted in the final product being 200 mg kg^{-1}.

The addition of sulfur dioxide to apple juice results in the formation of so-called sulfite addition compounds through the binding of sulfite to carbonyl compounds. The extent of sulfite binding is dependent upon the nature and origin of the carbonyl compounds present in the juice (see below). Similarly, if sulfur dioxide is added to an actively fermenting juice, there is a rapid combination with yeast metabolites such as acetaldehyde. Such juices will require a higher quantity of sulfite addition, if wild yeasts and other microorganisms are to be controlled effectively. Consequently, all additions of sulfur dioxide must be completed immediately after pressing the juice,

Table 1 Typical microorganisms of freshly pressed apple juice

Type	*Typical species*	*Ability to grow at the acidity of apple juice*[a]	*Sensitivity to sulfite*[b]
Yeast	*Saccharomyces cerevisiae*	++++	± or −
	S. uvarum	++++	± or −
	Saccharomycodes ludwigii	++++	−
	Kloeckera apiculata	++++	+++
	Candida mycoderma	++++[c]	++++
	Pichia spp.	++++[c]	++++
	Torulopsis famata	++++	++
	Aereobasidium pullulans	++++	+++
	Rhodotorula spp.	++++	++++
Bacteria	*Acetobacter* spp.	++++[c]	++
	Pseudomonas spp.	+	++++
	Escherichia coli	−/+	++++
	Salmonella spp.	−	++++
	Micrococcus spp.	+	++++
	Staphylococcus spp.	+	++++
	Bacillus spp.	− (cells)	− (spores)
	Clostridium spp.	− (cells)	− (spores)

[a]++++, capable of good growth; +, capable of some growth; −/+, strain-dependent; −, no growth.

[b]−, insensitive; ±, relatively insensitive; ++, +++, ++++, increasingly sensitive.

[c]Only in the presence of air (e.g., on the surface of the cider).

although, provided the initial fermentation by 'wild' yeasts is inhibited, further additions to give a desired level of free sulfur dioxide can be made during the following 24h.

When dissolved in water, sulfur dioxide and its salts set up a pH-dependent equilibrium mixture of 'molecular sulfur dioxide,' bisulfite and sulfite ions (Figure 1). The antimicrobial activity of sulfur dioxide is believed to be due to the molecular sulfur dioxide moiety of that part which remains unbound (the so-called 'free' sulfur dioxide). Less sulfur dioxide is needed in juices of high acidity, for instance 15 mg l^{-1} of free sulfur dioxide at pH 3.0 has the same antimicrobial effect as 150 mg l^{-1} at pH 4.0.

Fermentation Yeasts

The fermentation process is carried out by strains of *Saccharomyces* spp., especially *S. cerevisiae* and *S. uvarum*, which are added to the sulfite-treated juice as a pure culture. The starter culture will be prepared in the laboratory from freeze-dried or liquid nitrogen frozen cultures, which are resuscitated in broth and then cultivated through increasing volumes of a suitable culture medium to give an inoculum for use in a starter propagation plant. The nature of the cultivation medium used will vary but is often based on sterile apple juice supplemented with appropriate nitrogenous substrates (e.g., yeast extract) and sometimes with vitamins such as pantothenate and thiamine. In order to insure that the starter culture has both a high viability and a high vitality, it is normal to aerate the yeast culture during the propagation stage.

Increasingly, commercial dried or frozen yeast cell preparations are used, either for direct vat inoculation or as inocula for the yeast propagation plant. The availability of such preparations enables the cider maker to use different strains of yeast for different cider fermentations without the need to maintain a wide range of cultures in the laboratory. They also reduce the risks from mutation and/or contamination of the starter culture.

The choice of culture is dependent upon many criteria, such as flocculation characteristics, ability to ferment efficiently at a range of temperatures, alcohol- and sulfur dioxide tolerance, lack of ability to produce hydrogen sulfide, etc. One desirable characteristic is the ability to produce fusel oils (e.g., higher alcohols), which affect both the flavor and aroma of the cider (Table 2).

Dependent upon temperature, the fermentation process typically takes some 10 days to 12 weeks to proceed to dryness (i.e., specific gravity (SG) 0.990–1.000) at which time all fermentable sugars will have been converted to alcohol, carbon dioxide and other metabolites (Tables 2 and 3). After inoculation, the starter yeast together with any sulfite-resistant wild yeasts from the juice will increase in numbers from an initial level of about $2–5 \times 10^5$ colony-forming units (CFU) ml^{-1} to $1–5 \times 10^7$ CFU ml^{-1}. Following an initial aerobic growth phase, the resulting oxygen limitation and high carbohydrate levels in the media trigger the onset of the anaerobic fermentation process.

In controlled fermentations, a maximum temperature of 25 °C will generally be tolerated, although fermentations controlled at 15–18 °C are not uncommon in many countries. Because of the exothermic

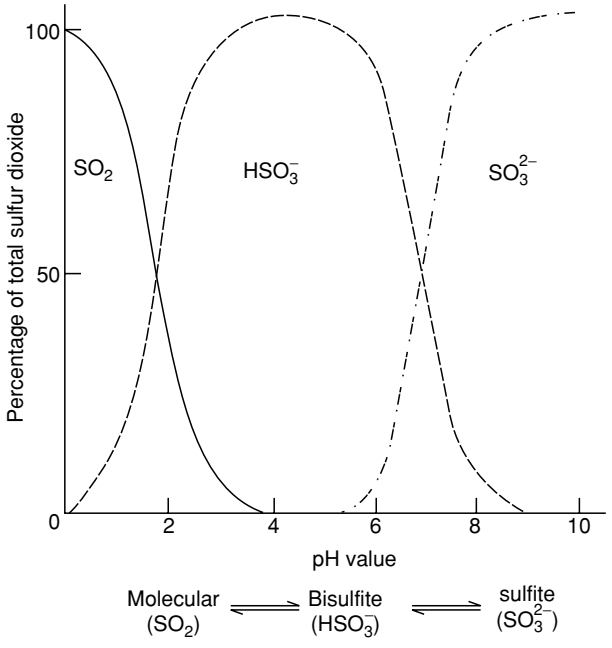

Figure 1 Percentage of sulfite, bisulfite and molecular sulfur dioxide as a function of pH in aqueous solution. From Hammond SM and Carr JG (1976) The antimicrobial activity of SO_2 – with particular reference to fermented and non-fermented fruit juices. In: Skinner FA and Hugo WB (eds) *Inhibition and Inactivation of Vegetative Microbes* (*S.A.B. Symposium Series No. 5*), pp. 89–110. London: Academic Press, with permission.

Table 2 Higher alcohols in apple juices and ciders

	Concentration range (mg l^{-1})	
	Apple juices	*Ciders*
n-Propanol	0.2–2	4–200
n-Butanol	3–24	4–32
iso-Butanol		14–74
iso-Pentanol	0.1	42–196
sec-Pentanol	0.1–2	16–39
n-Hexanol	1–2	2–17
2-Phenylethanol		7–260

nature of the initial stages of fermentation, it is normal to cool the fermentation base to a temperature < 18 °C prior to inoculation, otherwise excessive temperatures (30 °C or above) can be reached. In tropical countries, it is not uncommon for temperatures as high as 35–40 °C to occur in the fermentation vat in the absence of effective temperature control.

Such high temperatures are undesirable since activity of the starter yeast strain may be inhibited – leading to 'stuck' fermentations and the growth of undesirable thermoduric yeasts and spoilage bacteria. Stuck fermentations can sometimes be restarted by 'rousing' (bubbling carbon dioxide through the vat), and/or addition of glucose, a nitrogen source (10–50 mg l^{-1}), usually diammonium phosphate, and thiamine (0.1–0.2 mg l^{-1}); yeast 'hulls' may also be added.

At the end of fermentation, the yeast cells will flocculate and settle to the bottom of the vat. During this process, a certain amount of cell autolysis occurs, which liberates cell constituents into the cider. The raw cider will be drawn (racked) off the lees as a cloudy product and transferred to storage vats for maturation. In some plants, the cider may be centrifuged or rough filtered at this time. If the cider is left too long on the lees, the extent of autolysis may become excessive leading to a build-up of nitrogenous materials that will act as substrates for subsequent undesirable microbial growth and the development of off-flavors in the product. (*See* **Yeasts**.)

Maturation and Secondary Fermentation

The maturation vats are filled with the racked-off cider and either provided with an 'over-blanket' of carbon dioxide or sealed to prevent ingress of air, which would stimulate the growth of film-forming yeasts (e.g., *Brettanomyces* spp., *Pichia membranefaciens, Candida mycoderma*) and aerobic bacteria (e.g., *Acetobacter xylinum*). Growth of such yeasts will produce precursors of an unpleasant flavor compound, believed to be 1,4,5,6-tetrahydro-2-aceto-pyridine, which is responsible for a 'mousey' flavor defect. Growth of *Acetobacter* spp. will produce acetic and other volatile acids, which impart a vinegary note. Of course, deliberate acetification of cider can be used to produce cider vinegar.

During the maturation process, growth of lactic acid bacteria (LAB) (e.g., *Lactobacillus pastorianus* var. *quinicus, L. mali, L. plantarum, Leuconostoc mesenteroides*, etc.) causes a malolactic fermentation (see below). (*See* **Lactic Acid Bacteria**.)

Spoilage and Other Microorganisms in Cider

Bacterial pathogens, such as *Salmonella* spp., *Escherichia coli*, and *Staphylococcus aureus*, may occasionally occur in apple juice, having been derived from the orchard soil, farm, and factory process equipment or human sources. Normally, the acidity of the product prevents growth, and such organisms do not survive for long in the fermenting product. In recent years, there have been several reports of food poisoning in the USA due to *E. coli* O157:H7 in 'cider'. These references to 'cider' refer not to fermented (or 'hard' cider) but to fresh pressed apple juice. It has been established that although these highly acid-tolerant strains of *E. coli*, associated with the US outbreaks, can survive for long periods, and may even grow, in apple juice, they are extremely sensitive to alcohol and die within 1–2 h in fermenting cider.

In 1994, a report of a serious outbreak of cryptosporidiosis in the USA from consumption of 'cider' was again associated with freshly pressed apple juice, *not* fermented cider. However, the report highlights a potential risk of contamination of apples, juice and cider by oocysts of *Cryptosporidium* spp. if apples are harvested from an orchard sward following grazing by animals. *Crytosporidium* oocysts are sensitive to pasteurization; furthermore, the filtration processes used in commercial cider production would remove any contaminant oocysts.

Bacterial spores from species of *Bacillus* and *Clostridium* can survive for long periods and are frequently found in cider but do not create a spoilage threat, because of the acidity, although their presence may be indicative of poor plant hygiene.

The juice from unsound fruits and juice contaminated within the process plant may show extensive contamination by microfungi, such as *Penicillium expansum, P. crustosum, Aspergillus niger, A. nidulans, A. fumigatus, Paecilomyces varioti, Byssochlamys fulva, Monascus ruber, Phialophora mustea*, and by species of *Alternaria, Cladosporium, Botrytis, Oospora*, and *Fusarium*. None are of particular direct concern in cider making, except that heat-resistant species (e.g., *Byssochlamys* spp.) can survive pasteurization and grow in cider if it is not adequately carbonated. (*See* **Microbiology**: Classification of Microorganisms; **Food Poisoning**: Classification.)

The occurrence of the mycotoxin 'patulin' in apples infected with *Penicillium expansum* may result in carry-over of patulin in the apple juice base used for cider fermentation. Although patulin initially inhibits growth of the fermentation yeasts, the organisms rapidly become tolerant to patulin, and once growth is initiated, the patulin is rapidly metabolized to form ascladiol and smaller amounts of other metabolites. Hence, if apple juice is contaminated with patulin, the fermentation may be slow to start, but the patulin will be destroyed within a few hours. Claims from France that patulin has been found in fermented cider are

believed to be associated with post-fermentation sweetening of the cider by addition of patulin-contaminated apple juice. (*See* **Mycotoxins**: Occurrence and Determination; Toxicology.)

In addition to *Brettanomyces* spp. and *Acetobacter* spp., which can cause oxidative spoilage of cider during fermentation and maturation, the yeast *Saccharomycodes ludwigii* can be a major spoilage organism. *Sacc. ludwigii*, which is often resistant to sulfite levels as high as 1000–1500 mg l^{-1}, can grow slowly during all stages of fermentation and maturation and is often an indigenous contaminant of cidermaking premises. Its presence in bulk stocks of cider does not cause overt problems. However, if it is able to contaminate 'bright' cider at bottling, its growth will result in a butyric flavor and the presence of flaky particles that spoil the appearance of the product. Although the organism is sensitive to pasteurization, it is not unknown for it to contaminate products at the packaging stage, either as a low level contaminant of clean but nonsterile bottles or from the packaging plant and its environment.

Environmental contamination of final products can occur also with wild strains of yeasts such as *S. cerevisiae, Zygosaccharomyas bailii*, and *S. uvarum*, which will metabolize any residual or added sugar to generate further alcohol and, more importantly, to increase the concentration of carbon dioxide. Process-plant strains of these organisms are frequently resistant to low levels of sulfite. In bottles of cider inoculated with such fermentative organisms, carbonation pressures of up to 9 bar have been recorded. For this reason, it is essential to maintain an adequate level of free sulfite (typically 30–50 mg l^{-1}) in the final product, particularly in multiserve containers that may be opened and then stored with a reduced volume of cider; alternatively, a second preservative such as benzoic or sorbic acid can be used, where permitted by legislation. This precaution is not necessary for products packaged in single-serve cans and small bottles that are in-pack pasteurized after filling.

Special Secondary Fermentation Processes

Conditioned draught cider Traditional 'conditioned' draught cider results from a live secondary fermentation process. After filling into barrels, a small quantity of fermentable carbohydrate is added to the cider, followed by an active inoculum of alcohol-resistant yeasts. The subsequent growth is accompanied by a low-level fermentation, during which sufficient carbon dioxide is generated to produce a 'pettilance' in the cider together with a haze of yeast cells. Such products have a relatively short shelf-life in the barrel.

Double fermented cider The cider is initially fermented to a lower than normal alcohol content (e.g., 5% abv) by restricting the total amount of sugar present. The liquor is racked off as soon as the cider has fermented to dryness and either sterile-filtered or pasteurized prior to transfer to a second sterile fermentation vat. Fermentation sugar and/or apple juice is added, and a secondary fermentation is induced following inoculation with an alcohol-tolerant strain of *Saccharomyces* spp. Such a process permits the development of very complex flavors in the cider.

Sparkling ciders Sparkling ciders are normally prepared nowadays by artificial carbonation to a pressure of 3.5–4 bar. Traditionally, sparkling ciders were prepared according to the 'methode champagnoise.' After bright filtration, the fully fermented dry cider is filled into bottles containing a small amount of sugar and an appropriate Champagne yeast culture. The bottles are corked and wired and laid on their side for the fermentation process, which will last from 1 to 2 months at 15–18 °C. Following this stage, the bottles are placed in special racks with the neck in a downwards position. The bottles are gently shaken each day to move the deposit down towards the cork, a process that can take up to 2 months. The disgorging process involves careful removal of the cork and yeast floc but without loss of any liquid (sometimes the neck of the bottle is frozen to aid this process). The disgorged product is then topped up using a syrup of alcohol, cider, and sugar prior to final corking, wiring, and labeling. It is not difficult to understand why this process is rarely used nowadays! An alternative process, known as '*cuvée close*,' involves a secondary fermentation using champagne yeast in a sealed vat. The naturally produced carbon dioxide is retained in the cider, which is filtered and bottled under a positive pressure.

Chemistry of Cider

The chemical composition of cider is dependent upon the composition of the apple juice, the nature of the fermentation yeasts, malolactic bacteria, microbial contaminants, and their metabolites, and the nature of any additives used in the final product.

Composition of Cider Apple Juice

Apple juice is a mixture of sugars (primarily fructose, glucose, and sucrose), oligosaccharides, and polysaccharides (e.g., starch) together with malic, quinic, and citromalic acids, tannins (i.e., polyphenols), amides and other nitrogenous compounds, soluble pectin, vitamin C, minerals, and a diverse range of esters

that give the juice a typical apple-like aroma (e.g., ethyl- and methyl-*iso*-valerate). The relative proportions will be dependent upon the variety of apple, the cultural conditions under which it was grown, the state of maturity of the fruit at the time of pressing, the extent of physical and biological damage (e.g., mold rots), and, to a lesser extent, the efficiency with which the juice was pressed from the fruit.

Treatment of the fresh juice and/or cider with sulfite results in the complexing of carbonyl compounds to form stable hydroxy sulfonic acids. If the apples contained a high proportion of mold rots, then appreciable amounts of carbonyls such as 2,5-dioxogluconic acid and 2,5-D-threo-hexodiulose will occur, which bind sulfite effectively. Sulfite-binding also occurs with sugars, such as glucose and xylose, and with yeast metabolites such as acetaldehyde and pyruvate. Sulfur dioxide is important also as an antioxidant that prevents enzymic and nonenzymic browning reactions of the polyphenols.

Products of the Fermentation Process

The primary objective of fermentation is the production of ethyl alcohol from fruit sugars with the associated formation of carbon dioxide. The biochemical pathways that govern this process are well recognized (Embden–Meyerhof–Parnass pathway).

The various intermediates in this metabolic pathway can also be converted to form a diverse range of other metabolites, including glycerol (up to 0.5%). Diacetyl and acetaldehyde may also occur, particularly if the conversion of pyruvate to ethanol is inhibited by excess sulfite and/or if uncontrolled lactic fermentation occurs. Other metabolic pathways operate simultaneously with the formation of long- and short-chain fatty acids, esters, lactones, etc. Methanol will be produced in small quantities (10–100 mg l^{-1}) as a result of demethylation of pectin in the juice. Table 3 illustrates some of the volatile compounds found in a normal and spoiled cider blend.

If LAB are also present in the fermentation, these can convert malic and quinic acids to lactic and dihydroshikimic acids, respectively, thereby reducing the acidity of the cider. These reactions are accompanied by further diverse, but not widely understood, chemical and biochemical changes that result in subtle, yet important, flavor changes in the final product. Lactic and acetic acids can also be formed by metabolism of residual sugars and ethanol; great care needs to be taken to avoid excessive production of volatile acids in cider.

It has long been believed that most tannins in cider do not change significantly during fermentation, other than the reduction of the chlorogenic, caffeic and *p*-coumaryl quinic acids to dihydroshikimic acid and ethyl catechol, respectively. However, recent work has shown that several other very important quantitative and qualitative changes occur, especially during the malolactic fermentation. Such changes modify the organoleptic perception of bitterness and astringency in cider.

The nitrogen content of cider juice includes a range of amino acids, the most important of which are asparagine, aspartic acid, glutamine, and glutamic acid; small amounts of proline and 4-hydroxy-methyl-proline also occur. Aromatic amino acids are virtually absent from apple juices. With the exception of proline and 4-hydroxy-methyl-proline, the amino acids are largely assimilated by the yeasts

Table 3 Volatile compounds in a normal and a diacetyl-spoiled cider

Compound	*Normalized peak area*[a]	
	Normal cider	*Spoiled cider*
Ethyl acetate	86.1	89.2
Diacetyl	0	3.4
Ethyl-2-methylbutyrate	10.3	12.9
2-Methylpropanol	35.4	97.3
iso-Amyl alcohol	305	213
2- and 3-Methyl-butan-1-ol	503	456
Ethyl hexanoate	233	179
Hexyl acetate	2	6.3
Octanol	0.7	1.2
Ethyl lactate	45.9	37.9
Hexan-1-ol	35.7	29.5
Nonanol	0.8	0.9
Unknown ester or acetal (relative molecular mass 172)	25.6	77.7
Ethyl octanoate	280	226
Heptan-1-ol	1.3	0.8
Ethyl octanoate	3.4	11.5
Decan-2-one	3.6	0.9
Benzalaldehyde	1.4	1.4
Ethyl-2-hydroxy-4-methyl pentanoate	4.2	7
Ethyl decanoate	57.6	53.4
Decanal	7.5	0.4
Ethyl benzoate	4.7	5.7
Diethyl succinate	5.2	2.9
Unknown ester	3.2	4.9
Methionol	1.3	0.6
Undecanal	2.2	1.2
2-Phenylethyl acetate	11.4	6.4
Hexanoic acid	12.6	12.6
Ethyl dodecanate	2.1	1
2-Phenylethanol	69	61.6
Heptanoic acid	0	3.4
d-Decalactone	4.4	3.4
Ethyl guaiacol	2.5	4.5
Octanoic acid	34.6	29.4
Nonanoic acid	1	1.2

[a]Based on gas chromatography–mass spectroscopy analysis of headspace volatiles.

during fermentation. However, leaving the cider on the lees for an appreciable length of time will significantly increase the amino nitrogen content as a consequence of the release of cell constituents during autolysis.

Inorganic compounds in cider are derived largely from the fruit and will depend upon the conditions prevailing in the orchard. These levels do not change significantly during fermentation. Small amounts of iron and copper may occur naturally, but the presence of more than a few milligrams per liter will result in significant black or green discolorations and flavor deterioration. The discolorations are due to the formation of iron or copper tannates from traces of metal ions derived from equipment and/or from the use of rotten fruit.

Cider Maturation

The aroma and flavor of freshly fermented cider are quite harsh. During maturation, significant changes occur in the composition of the cider that are due to microbial and biochemical activity, producing diverse compositional changes. More than 200 metabolites have been identified in mature cider; some produce desirable aroma and flavor characteristics, whereas others may be responsible for undesirable characteristics, especially if present in excessive amounts. The 'malolactic fermentation' causes a reduction in the acidity of the cider and imparts subtle flavor changes that generally improve the flavor. Much of the lactic acid, produced by decarboxylation of malic acid, is esterified with the formation of ethyl-lactate and other esters, which impart a smooth 'creamy' flavor to the product. Malolactic fermentation also results in the production of important volatile metabolites such as 1-hexanol, 2- and 3-methyl-butanol, and 2-phenyl-ethanol. The occurrence in mature cider of a compound with molecular mass 172 has been attributed variously to the occurrence of 1-ethoxyoct-5-en-1-ol, ethenylthio-octane, 5-chloro-salicylic, and other metabolites. It has been suggested that this compound may be responsible for the typical 'cider' flavor and aroma. Changes to the polyphenols (tannins) result in subtle changes to the perception of astringency and bitter characteristics.

However, in certain circumstances, metabolites of the lactic acid bacteria may damage the flavor and result in spoilage, e.g., excessive production of diacetyl (and its vicinyl-diketone precursors), the 'butterscotch-like' taste of which can be detected in cider at a threshold level of about 0.6 mg l^{-1}.

Commercial Ciders

A final blending of ciders is made to attain specific characteristics of sweetness, dryness, alcohol content, flavor, and aroma. Sugars or intense sweeteners, such as saccharin (which increases the perception of asringency), may be added. Artificial colors, ascorbic acid (as an antioxidant), and additional chemical preservatives (e.g., sulfur dioxide or sorbic acid) may also be added. Prior to final packaging, the cider is carbonated to give the product a petillance or sparkle.

See also: **Cider (Cyder; Hard Cider)**: The Product and its Manufacture; **Lactic Acid Bacteria**; **Mycotoxins**: Classifications; **Preservatives**: Classifications and Properties; **Spoilage**: Bacterial Spoilage; Yeasts in Spoilage; **Tannins and Polyphenols**; **Yeasts**

Further Reading

Beech FW (1972) English cidermaking: technology, microbiology and biochemistry. In: Hockenhull DJD (ed.) *Progress in Industrial Microbiology*, vol. 11, pp. 133–213. Edinburgh: Churchill Livingstone.

Beech FW and Davenport RR (1983) New prospects and problems in the beverage industry. In: Roberts TA and Skinner FA (eds.) *Food Microbiology: Advances and Prospects (S.A.B. Symposium Series No. 11)*, pp. 241–256. London: Academic Press.

Charley VLS (1949) *The Principles and Practice of Cidermaking*. London: Leonard Hill.

Hammond SM and Carr JG (1976) The antimicrobial activity of SO_2 – with particular reference to fermented and non-fermented fruit juices. In: Skinner FA, Hugo WB (eds) *Inhibition and Inactivation of Vegetative Microbes (S.A.B. Symposium Series No. 5)*, pp. 89–110. London: Academic Press.

Jarvis B (2001) Cider, perry, fruit wines and other alcoholic beverages. In: Arthey D and Ashurst PR (eds) *Fruit Processing*, 2nd edn. pp. 111–148. Gaithersburg, MA: Aspen Publishers Inc.

Jarvis B, Forster MJ and Kinsella WP (1995) Factors affecting the development of cider flavour. In: Board RG, Jones D and Jarvis B (eds) *Microbial Fermentations: Beverages, Foods and Feeds (SAB Symposium Series No. 24)*. *Journal of Applied Bacteriology* 79(supplement): 5S–18S.

Jarvis B and Lea AGH (2000) Sulphite binding in ciders. *International Journal of Food Science and Technology* 35: 113–127.

Lea AGH (1995) Cidermaking. In: Lea AGH and Piggott JR (eds.) *Fermented Beverage Production*, pp. 66–96. London: Blackie Academic & Professional.

Williams RR (ed.) (1991) *Cider and Juice Apples: Growing and Processing*. Bristol: University of Bristol.

CIRRHOSIS AND DISORDERS OF HIGH ALCOHOL CONSUMPTION

I Bitsch, Institut für Ernährungswissenschaft der Justus-Liebig Universit, Giessen, Germany

Introduction

It is well documented that heavy drinking of alcoholic beverages carries an increased risk of morbidity and mortality from diseases affecting many organs and may lead to psychic and physical dependence on alcohol. The risks depend on the amount of alcohol consumed, the drinking pattern, and the individual sensitivity. For these reasons, no generally valid threshold value for safe alcohol consumption can be given. Most authors agree that an upper limit of 80 g of ethanol per day should not be exceeded. A moderate drinker is considered as one who consumes 5–25 g of ethanol per day and a light drinker as one who consumes 0.2–5 g per day. Definitions of moderate drinking vary among studies. The US Department of Agriculture and the US Department of Health and Human Services define moderate drinking as not more than 24 g pure alcohol per day for men and not more than 12 g per day for women. Women appear to be more vulnerable than men to many adverse consequences of alcohol use. Women achieve higher concentrations of alcohol in the blood and become more impaired than men after drinking equivalent amounts of alcohol. Research also suggests that women are more susceptible than men to alcohol-related organ damage. Compared with men, women develop alcohol-induced liver disease over a shorter period of time and after consuming less alcohol. In addition women are more likely than men to develop alcoholic hepatitis and to die from cirrhosis.

Alcoholism

Three main patterns of chronic alcohol abuse are described in the *Diagnostic and Statistical Manual of Mental Disorders* of the American Psychiatric Association (1994):

- regular drinking of large amounts
- regular heavy drinking but limited to weekends
- episodic binges of heavy daily drinking lasting weeks or months, interrupted by long periods of sobriety

These criteria for the diagnosis of alcoholism also include an impairment in social or occupational functioning. Alcoholism is further distinguished from alcohol abuse by tolerance and physical dependence. Tolerance is defined as the 'need for markedly increased amounts of alcohol to achieve the desired effect, or a markedly diminished effect with regular use of the same amount.' Dependence is defined as 'the development of alcohol withdrawal syndrome after cessation of or reduction in drinking.' (*See* **Alcohol**: Metabolism, Beneficial Effects, and Toxicology; **Alcohol**: Alcohol Consumption.)

Excessive alcohol intake is often associated with malnutrition. This results from the limited intake of nutritionally adequate foods and from the impairment of digestion, absorption, transport, storage, metabolism, and excretion of many nutrients by direct or indirect action of ethanol. The role of nutrient deficiencies in the initiation and progression of the medical complications of alcoholism is still controversial.

Protein malnutrition (kwashiorkor-like) and protein-energy malnutrition (marasmus-like) are frequent in alcoholic patients with liver disease and, in some studies, the prevalence of the malnutrition correlates closely with the severity of organ failure. On the other hand, many patients who drink to excess are clearly not protein-energy-malnourished. Ethanol has appreciable effects on amino acid metabolism. Intestinal absorption and transport of isoleucine, arginine, and methionine are impaired by high concentrations of ethanol. Branched-chain amino acids and α-amino-*N*-butyric acid are increased in the plasma of alcoholics. (*See* **Protein**: Deficiency.)

Vitamin deficiencies are frequent in alcoholics and especially in those with liver disease. Most common is an insufficient folate supply, characterized by megaloblastic anemia and macrocytosis of the intestinal epithelium. Many factors contribute to folate deficiency in alcoholics. These include dietary deficiency, intestinal malabsorption, impairment of uptake and/or storage in the liver, and increased urinary excretion. Also ethanol metabolism perturbs folate metabolism through inhibition of methionine synthase, resulting in dysregulation of nucleotides and enhanced cancer risk. Also, the acetaldehyde product of ethanol metabolism was shown *in vitro* to trigger oxidative catabolism of the folic acid molecule. Poor dietary intake is undoubtedly the major cause, with the exception of heavy beer drinkers: 2 l of beer contain nearly 50% of the recommended daily allowance of folate for an adult man. The combined effect of low folate concentration in the diet and heavy alcohol

drinking leads to damage of intestinal mucosal epithelial structures, often associated with diarrhea, and results in malabsorption of folates and other nutrients. Symptoms of vitamin B_{12} deficiency are much less common than those of folate deficiency in alcoholics. This is probably attributable to the large stores of vitamin B_{12} in the body and the reserve capacity for absorption. In most studies the circulating levels of vitamin B_{12} in alcoholics are not different from normal controls. (*See* **Cobalamins**: Physiology; **Folic Acid**: Physiology.)

As with folate, poor dietary intake and an impairment of absorption are the main causes for insufficient thiamin supply in alcoholics. Alcohol inhibits active transport of thiamin across the intestinal mucosa, due to inhibition of Na,K-ATPase at the basolateral membrane of enterocyte, whereas the passive transport remains unimpaired. In addition, hepatic storage of thiamin may be reduced owing to fatty infiltration of the liver, hepatocellular damage, or cirrhosis in chronic alcoholic patients. Extreme thiamin deficiency is responsible for the Wernicke–Korsakoff syndrome, beriberi, and polyneuropathy in alcoholics. (*See* **Thiamin**: Physiology.)

The incidence of pyridoxine deficiency in alcoholic patients, as defined by low plasma levels of its circulating form, pyridoxal-5-phosphate (PLP), is more than 50% in different studies. Clinical data indicate that malnutrition, rather than the amount and duration of alcohol abuse, is the major determinant of pyridoxine deficiency. Pyridoxine absorption is primarily passive and is affected only by very high concentrations of ethanol. PLP in erythrocytes and liver is more rapidly destroyed in the presence of acetaldehyde, the first metabolite of ethanol oxidation, perhaps by displacement of PLP from protein and its exposure to phosphatases. This results in high pyridoxic acid excretion in the urine. Clinical signs of pyridoxine deficiency in alcoholics are infrequent. Sideroblastic bone marrow changes often occur in alcoholics with low plasma PLP values, but most of these patients also suffer from liver disease and folate deficiency. (*See* **Vitamin B_6**: Properties and Determination.)

Studies of vitamin A deficiency in alcoholism have mainly concerned patients with established cirrhosis, who may have impaired storage or transport of vitamin A because of an inadequate synthesis of retinol-binding protein. Most of them also have inadequate dietary intake. Other complications of alcoholism, e.g., pancreatic and biliary insufficiency, lead to malabsorption of vitamin A because this fat-soluble vitamin requires for its absorption adequate quantities of pancreatic lipases and bile salts in the small intestine. Another possible mechanism for low hepatic vitamin A level is increased hepatic metabolism of retinoic acid to polar metabolites through the action of microsomal enzymes, which are inducible by ethanol consumption. The clinical consequences of insufficient vitamin A supply in alcoholics are increased incidence of night blindness, follicular keratosis, and corneal ulcerations. Vitamin A deficiency may compromise immune function and may be procarcinogenic through effects on epithelial cell metaplasia in the oropharynx. Alcoholics with these complications often require both vitamin A and zinc treatment to correct visual dysfunction, because of an association between zinc and vitamin A metabolism. Zinc is an essential cofactor in the conversion of retinol to retinaldehyde in the retina. Alcohol appears to stimulate the release of zinc from hepatic stores, and its urinary excretion. Zinc levels in plasma and red blood cells are often reduced in humans after chronic alcohol ingestion. Also zinc pools are redistributed with greater tissue binding in response to various cytokine mediators of alcoholic liver injury. The potential consequences of zinc deficiency include acrodermatitis, altered taste and smell, and night blindness. (*See* **Retinol**: Physiology; **Zinc**: Physiology.)

Vitamin D intake, absorption, and metabolism seem to be impaired in alcoholics, and there are abnormalities of phosphorus, calcium, and magnesium homeostasis. Alcoholic patients often suffer from decreased bone mass and an increased incidence of fractures. (*See* **Calcium**: Physiology; **Cholecalciferol**: Physiology; **Magnesium**.)

Alcoholic Cirrhosis

Alcohol is the most common cause of cirrhosis in western countries. The incidence of this disease in alcoholics depends upon the mean daily intake of alcohol and the mean duration of alcohol consumption. However, only about 20% of heavy drinkers develop cirrhosis, and liver disease may progress to cirrhosis after cessation of ethanol ingestion. This fact suggests that other factors superimposed on alcohol drinking are involved in the pathogenesis of this severe form of liver injury. These include gender, genetic or immunological variables, other hepatotoxins, nutrition, viral hepatitis, and others. Individual predisposition is an important physiological factor in the development of this disease. Understanding the mechanism of the differences in susceptibility to cirrhosis may help clinicians to identify and treat patients at increased risk. (*See* **Liver**: Nutritional Management of Liver and Biliary Disorders.)

Morphologically, cirrhosis of the liver is a diffuse process, characterized by an excessive proliferation of connective tissue and the deposition of structurally

abnormal nodules. Alcoholic cirrhosis is mostly micronodular in type, with a size of nodules from 1 to 5 mm. Macronodules with a size ranging from 5 to 50 mm may occur, especially in the late phases of the disease. The loss of normal liver architecture, with separation of the portal tracts and the central zones of the liver by septa of fibrotic tissue, results in alterations of the vascular supply and a disturbance of the intrahepatic blood circulation.

Morbidity and mortality resulting from alcoholic cirrhosis are related principally to the loss of liver cell function, to derangements in the vascular system of the liver, or to both. The onset of cirrhosis is often insidious and associated with nonspecific symptoms such as fatigue, anorexia, weight loss, nausea, and abdominal discomfort. As the disease progresses, signs of hepatocellular failure became prominent. The most severe complications are iron overload, hepatic encephalopathy, and portal hypertension with ascites and bleeding from esophageal varices.

In association with cirrhosis of the liver, hepatocellular carcinoma may develop. The pathogenesis of the carcinomatous transformation is still unclear, especially because these tumor forms may also occur in noncirrhotic livers.

Ethanol leads to a number of metabolic and structural alterations in the liver that predispose this organ to derangements in its functional integrity. These are as follows: an increase in the ratio of NADH (the reduced form of nicotinamide adenine dinucleotide, or NAD) to NAD; interactions of ethanol with lipid and protein metabolism; stimulation of fibrosis with deposition of collagen; inhibition of liver cell regeneration; humoral and cellular immunological alterations; and excessive production of free radicals and cytokines. Alcohol induces cell death and inflammation, which can result in scarring that distorts the liver's internal structure and impairs its function. Acetaldehyde, the first oxidation product of ethanol metabolism, may exert some toxic effects of its own in liver tissue; it activates stellate cells directly and promotes liver scarring in the absence of inflammation.

All stages of liver injury can be produced in the baboon fed high-protein and vitamin-supplemented diets. This evidence suggests that toxic effects of alcohol, and not malnutrition, are the principal causes for the development of cirrhosis in chronic alcoholics. But alcohol abuse establishes only the conditions for the generation of cirrhotic lesions, which require the addition of some independent factors emerging over time.

Abstinence from alcohol is the essential factor for prevention and treatment of alcoholic cirrhosis. If irreversible liver damage is already established, some complications of cirrhosis can be alleviated by nutritional treatment. Portal-systemic encephalopathy, for example, often responds to an application of amino acid mixtures, enriched with branched-chain amino acids, whereas ascites responds favorably to sodium restriction.

Wernicke's Encephalopathy and Korsakoff's Syndrome

Wernicke's encephalopathy and Korsakoff's psychosis are diseases of the central nervous system (CNS) secondary to alcoholism. They represent a continuum of the same neuropathological process and develop in about 2–3% of alcoholics. Wernicke's disease is often followed by Korsakoff's syndrome. Severe alcoholics may have both diseases, but some of them show Korsakoff's psychosis without preceding Wernicke's encephalopathy. Wernicke's disease is an acute or chronic encephalopathy with a triad of clinical abnormalities: ophthalmoplegia, ataxia, and mental confusion. Korsakoff's syndrome is a psychosis with marked abnormalities in cognitive function: the cardinal symptoms are anterograde amnesia, disorientation, learning deficits, and confabulations. The relationship between the two diseases is not entirely clear. The neuropathological changes seen in autopsy materials of Wernicke–Korsakoff patients consist of circumscribed, symmetrical lesions in the diencephalon and brainstem. Most affected are the mammillary bodies and the dorsomedial nuclei of the thalamus. In acute cases the lesions are widespread and severe. In chronic cases the lesions are more restricted and show great variations in extent and severity within the affected area. There are corresponding variations of the clinical symptoms.

Extreme deficiency of thiamin, induced by malnutrition and interaction of ethanol with thiamin absorption and metabolism, is regarded as the primary cause of this syndrome. Patients with Wernicke's disease often have a high energy intake, consisting mainly of ethanol and/or carbohydrates, without sufficient protein and vitamins. Beyond that a direct toxic effect of alcohol on the brain has also been implicated. The Wernicke–Korsakoff's disease is not confined to alcoholism, but is also present in other conditions associated with thiamin malnutrition, e.g., hyperemesis gravidarum, Hodgkin's lymphoma, carcinoma of the stomach, and anorexia nervosa. Thiamin application can reverse Wernicke's syndrome in many but not all patients, at least when administered early in the course of the disease. In contrast, Korsakoff's patients often show poor response to thiamin therapy. Additional factors, of as yet unknown origin, seem to be necessary for the full development of

Korsakoff's psychosis. There is evidence to suggest that a genetic predisposition is involved in the etiology of the Wernicke–Korsakoff syndrome.

In some patients a variant form of transketolase with a low affinity for its coenzyme thiamin pyrophosphate has been diagnosed. This isoenzyme requires much larger amounts of thiamin to function than the normal enzyme. Some authors suggest that a variant transketolase and thiamin deficiency together contribute to the pathogenesis of the brain damage of the Wernicke–Korsakoff syndrome. In other studies this hypothesis has not been confirmed. (*See* **Anorexia Nervosa**.)

Alcoholic Polyneuropathy, and Beriberi in Alcoholics

Symmetric and predominantly distal polyneuropathy is probably the most common sequela of chronic alcohol abuse. It occurs in about 20% of all alcoholics. First symptoms include paresthesia, dysesthesia and pain sensations affecting primarily the lower extremities, accompanied by reduced or absent reflex activity. These symptoms are usually followed in the more advanced stages of the disease by marked motor impairments, such as weakness and atrophy of the anterior tibial muscles. Electrophysiological slowdown of the sensory and motor conduction velocities of the peripheral nerves, as well as myopathic changes, are usually observed. The predominant pathological abnormalities are distal pronounced axonal degenerations, mainly of the large fibers with a gradient of changes toward the extremities ('dying-back degeneration'). Segmental demyelination appears as a secondary phenomenon. Alcoholic polyneuropathy is clinically indistinguishable from beriberi ('dry beriberi') neuropathy. As in the Wernicke–Korsakoff syndrome, thiamin deficiency is regarded as the primary cause. Besides that, a long-lasting low protein-energy supply seems to be one of the main pathogenic factors for the development of the typical lesions of peripheric nerves. In addition, insufficient intake or absorption of pyridoxine, folates, B_{12} and other B vitamins contributes to the development of the disease, and direct toxic effects of ethanol and its metabolites are involved in the destruction of nerve ultrastructure and function. Recovery from alcoholic peripheral neuropathy is slow and often incomplete. The most important therapeutic factors are abstinence from alcohol and improvement of the overall nutritional status. Additional supplementation with high doses of B-complex vitamins is necessary. With this therapeutic schedule a slow diminution of symptoms may be expected in many but not all cases.

Pancreatitis

The association between alcohol consumption and pancreatitis is well recognized. In general, about 60–90% of all cases of chronic pancreatitis are alcohol-related. As with alcoholic cirrhosis, there has been an increasing prevalence of alcoholic pancreatitis in industrialized countries in the last 30 years. This disease often develops without obvious clinical symptoms. However, in autopsies of patients with a history of alcoholism, marked pancreatic lesions are frequently observed. Patients with severe or complicated forms of this disease may rapidly develop marked generalized malnutrition. (*See* **Gallbladder**.)

A favored hypothesis regarding the mechanism of action of chronic alcohol consumption on the pancreas is the observation that ethanol increases the protein content of pancreatic juice, with a concomitant decrease in water and electrolytes. This induces a precipitation of protein plugs within the pancreatic ducts, followed by retraction and calcification, resulting in pancreatic stones, atrophy of the duct epithelium, and proliferation of the connective tissue. The consequences are stenosis or dilatations of the ducts, cysts and pseudocysts, and progressive disappearance of the pancreatic exocrine tissue which is replaced by fibrosis. Pathologically, the gland is edematous and hemorrhagic. Besides chronic overconsumption of alcohol, both a high-fat, high-protein diet and, paradoxically, malnutrition have been implicated in the pathogenesis of the disorder.

Alcoholic pancreatitis tends to be recurrent and progressive and to result in pancreatic exocrine insufficiency. Some 1–2% of alcoholics suffer from an acute form. This complication can be quite severe, with violent epigastric pain, nausea, and vomiting. Cases of sudden death from acute attacks have been reported.

Even heavy drinking does not modify the risk for pancreatic cancer. Several epidemiologic studies do not support an association between alcohol consumption and incidence of pancreatic cancer.

Effect of Chronic Alcohol Consumption on the Brain

Ethanol – in common with other centrally acting agents – rapidly diffuses across the blood–brain barrier and equilibrates in brain tissue. Its concentration depends upon the plasma concentration, vascularization, local blood flow, and water content of the concerned brain area. Chronic ethanol consumption is associated with a variety of deficits in brain function over a wide range of doses and is often followed by clinically obvious brain damage.

Alcoholic Dementia

Alcoholics often suffer from disorders in cognitive processes, varying from a relatively slight memory reduction to deep dementia. The deficits in cognitive function are similar to those seen in patients with Alzheimer's disease or multiinfarct dementia. Histopathologically, the brain structures that are possibly involved do not show any lesion or atrophy, but there is considerable evidence that alcoholic dementia may potentially result from progressive loss of some neurons, synapses, and/or their associated receptors. Prolonged abstinence has a considerable reducing effect on dementia symptoms, but in most subjects the reversibility is only partial. Alcoholic dementia seems to be a severe form of Korsakoff's syndrome.

Alcohol Tolerance and Dependence

Functional tolerance and physical dependence are inexorably linked phenomena reflecting adaptive changes in the brain which compensate for the inhibition of functions by ethanol. Functional tolerance means a decrease in the sensitivity of the CNS produced by chronic alcohol intake. Physical dependence means hyperexcitability of the CNS following alcohol withdrawal after adaptation. The molecular mechanisms underlying both phenomena are as yet unknown. Possibly the effects are primarily related to changes in the microorganization and physical state of brain membranes.

Brain Shrinkage

Cerebral shrinkage seems to be one of the most frequent brain lesions induced by chronic alcohol consumption. It can be demonstrated by computerized tomography (CT) scanning in patients with alcoholism and even in 'social drinkers.' Cerebral shrinkage relates to a loss of white matter rather than of gray matter. In alcoholics the shrinking results in a lower brain weight compared to age- and sex-matched controls. After cessation of alcohol drinking cerebral shrinkage is often partially reversed.

Cerebellar Atrophy

Many alcoholics show cerebellar atrophy with progressive ataxia of stance and gait and, occasionally, an impairment of speech and ocular motility. This cerebellar degeneration is characterized by great uniformity of the clinical and pathological manifestations. In the majority of cases the disorder develops over several weeks or months, followed by years of stability. The pathological changes consist of degeneration of neuronal elements of the cerebellar cortex, particularly of the Purkinje cells. The dendritic networks of these cells show alterations in length parameters under the influence of ethanol in rats. Topographically, the lesions of the cerebellum are restricted to the anterior and superior vermis and to the hemispheres.

Although cerebellar atrophy is most frequently encountered in chronic alcoholic patients, it has also been reported in old people and in malnourished individuals, who allegedly did not drink. It has been postulated that alcohol and the aging process superimposed upon malnutrition may be responsible for the damage caused to the cerebellum.

Fetal Alcohol Syndrome

Prenatal exposure to alcohol produces various morphological, physiological, and behavioral abnormalities in the newborn, termed fetal alcohol syndrome (FAS). It has been observed in 1 in every 1000 children born to alcoholic mothers, and is the most common known cause of mental retardation. The Fetal Alcohol Study Group of the Research Society on Alcoholism has defined three criteria for the diagnosis of FAS:

- growth retardation
- abnormalities of the CNS
- abnormal facial features

The mechanisms by which alcohol produces these effects are not yet fully elucidated. Possibly ethanol affects NMDA and γ-aminobutyric acid receptors in the developing brain, leading to neuron death. The minimum level of alcohol consumed during pregnancy that causes FAS has not been established. Animal models have demonstrated conclusively that ethanol can cross the placenta and is teratogenic in several species. The effect is dose-dependent and the type of malformation observed depends upon the stage of development when the exposure occurred. Besides the toxic effects of ethanol, another potential cause of FAS is maternal and/or fetal malnutrition induced by impaired placental transfer of essential nutrients.

Among the neuroanatomical alterations observed in humans and animal models with FAS are decreased brain weight, delays in dendritic development, decreased numbers of pyramidal neurons in the hippocampus, and sprouting in the dentate gyrus. The clinical symptoms of the CNS dysfunction are mental retardation, hyperactivity, and learning disabilities. Despite nutritional rehabilitation, the physical and mental development of the children remains impaired. To date, a safe intake of alcohol for pregnant women has not been established. (*See* **Pregnancy**: Safe Diet.)

See also: **Alcohol**: Metabolism, Beneficial Effects, and Toxicology; Alcohol Consumption; **Anorexia Nervosa**; **Calcium**: Physiology; **Cholecalciferol**: Physiology; **Cobalamins**: Physiology; **Folic Acid**: Physiology; **Gallbladder**; **Liver**: Nutritional Management of Liver and Biliary Disorders; **Magnesium**; **Pregnancy**: Safe Diet; **Retinol**: Physiology; **Thiamin**: Physiology; **Vitamin B_6**: Properties and Determination; **Zinc**: Physiology

Further Reading

Bitsch I (1983) Perspektiven der Alkoholforschung. *Ernährungs-Umschau* 30: 132–135.

Bitsch R (1987) Alkohol und Vitaminstoffwechsel. *Ernährungs-Umschau* 34: 161–166.

Eriksson K, Sinclair JD and Kiianmaa K (eds) (1980) *Animal Models in Alcohol Research*. London.

Hall PM (ed) (1995) *Alcoholic Liver Disease: Pathology and Pathogenesis*, 2nd edn. London.

Kalant H, Khanna JM and Israel Y (1991) *Advances in Biomedical Alcohol Research*. Oxford.

Lindros KO, Ylikahri R and Kiianmaa K (1987) *Advances in Biomedical Alcohol Research*. Oxford.

National Institute on Alcohol Abuse and Alcoholism (1999) Rockville, Alcohol Alert No 42, 46.

Shils ME, Olson JA and Shike M (eds) (1999) *Modern Nutrition in Health and Disease*, 9th edn. Baltimore: Lippincott, Williams & Wilkins.

Torvik A (1987) Brain lesions in alcoholics: neuropathological observations. *Acta Medica Scandinavica Supplementum* 717: 47–54.

Victor M, Adams RD and Collins GH (1989) *The Wernicke–Korsakoff Syndrome*, 2nd edn. Philadelphia.

Ziegler EM and Filer LJ (eds) (1996) *Present Knowledge in Nutrition*, 7th edn. Washington, DC: ILSI.

Citric Acid Cycle *See* **Tricarboxylic Acid Cycle**; **Oxidative Phosphorylation**

CITRUS FRUITS

Contents

Types on the Market

J Forsyth, Formerly of New South Wales Agriculture, Orange, NSW, Australia
J Damiani, Australian Citrus Growers Inc., Mildura, Victoria, Australia

Background

Citrus is the main fruit tree crop grown throughout the world. It is made up of many species that vary in importance due to different uses, markets, growing conditions, and climatic zones. The large production has significance in local and world trade for both fresh and processed products. Brief details on the classification of the species, growing regions and conditions, production statistics, with the utilization of the main and also lesser known types are discussed.

Classification

The taxonomic classification of the *Citrus* species is complex and diverse, and not universally agreed upon, but those used by Swingle and Reece will be followed in this article. Citrus, as a social and cultural

fruit crop, also has countless local names and identities throughout the world, often for an identical cultivar.

Citrus trees belong to the plant family Rutaceae, subfamily Aurantioideae, which comprises 33 well-known and described genera and 203 species. In addition, many natural and man-made hybrids also exist that have resulted in new edible cultivars. Many genera contain unusual or remote relatives of citrus with nonedible fruit, but of ornamental value like *Merrillia* and *Murraya*, or have genetic importance in breeding programs like *Poncirus* and *Severinia*.

True citrus fruit trees, which have a berry fruit called a hesperidium, belong to six genera: *Citrus, Fortunella, Poncirus, Microcitrus, Eremocitrus*, and *Clymenia*. Only the *Citrus* and *Fortunella* genera have fresh fruit cultivars of commercial importance. Both these genera are evergreen and unifoliate, and the genus *Citrus* provides nearly all of the commercial cultivars grown throughout the world.

Main Types on the Market

Citrus fruits are principally marketed according to how they are consumed:

1. As fresh or dessert fruits – sweet oranges, mandarins (eaten out of hand), grapefruit, or pummelo (spooned). Juice, slices, segments, rind, and leaves are also used to garnish food.
2. As processed products – juice (fresh, chilled, frozen, canned, blended, or concentrated), syrups and cordials, segments and rind oil, or essence. After the juice is extracted, there remain residues that can be a source material from which over 300 valuable byproducts can be produced. Some specialized types are also candied, dried, or used for marmalade manufacture. (*See* **Citrus Fruits**: Composition and Characterization; Processed and Derived Products of Oranges.)

The main types of fresh citrus fruits on the export or main local markets, and also used for processing, are normally broadly grouped as oranges, mandarins, lemons and limes, and grapefruit. There is also a wide range of other minor or lesser known species and types that may have regional or local significance. Some of these are mentioned in a later section. The following brief comments are made on the main citrus types, as listed in available statistics and found on world markets.

Sweet Oranges

The sweet orange types (*Citrus sinensis* [L.] Osbeck) are the most widely grown citrus fruits throughout the world and provide the greatest fruit marketing production. The many known cultivars can be subdivided into three main groups as the acidless or sugar oranges do not contribute to world trade:

- Common oranges (also known as blond or white oranges). The Valencia is the main cultivar grown, with harvesting from spring to autumn, depending on local climates. The fruit is dual purpose and suitable for both eating as fresh fruit or processing, while seedless types are also grown, e.g., Delta and Midknight. Other common oranges are often seedy and mainly suitable for processing, e.g., Pera, Hamlin, and Pineapple.
- Navel oranges are also widely grown as winter seedless eating fruit (**Figure 1**). The Washington navel with numerous clones or selections is the most important cultivar, while there are many others navels (over 50 distinct varieties or clones) available that spread the maturity period from early (like Fukumoto, Navelina, Leng, and Newhall) to late (Lane and Navelate), and to what are now being called 'summer navels,' following the location of many new very late maturing cultivars in Australia, e.g., Powell, Chislett, and Barnfield. A new pink/red fleshed navel – Cara Cara – is also creating marketing interest.
- Blood or pigmented oranges have been popular in Mediterranean countries for their distinctive flavor and both rind, flesh, and juice pigments, e.g., Tarocco, Sanguinello, and Moro.

Figure 1 (see color plate 25) Navels are the most important sweet, juicy, and seedless orange marketed for fresh fruit consumption.

Mandarins

Mandarins are also referred to as 'tangerines' in some countries. The mandarin group and related hybrids is very diverse (with over 100 cultivars), but can be classified into a number of main groups, some of which contain a large number of subgroups. This fruit is more seasonal, with each cultivar maturing and suitable for harvesting only over a short period of time, e.g., 6–8 weeks. Some are now available for the main types that can supply fresh fruit from very early to very late maturity. Mandarins production is increasing with a consumer trend towards easier to peel, segment, and seedless lunch/eaten out of hand fruit, while limited quantities are processed into juice or canned segments.

- Satsuma mandarins (*C. unshiu* Marcovitch) are also known as 'Unshiu mikan' in Japan and China. This is an important expanding seedless group that includes Clausellina, Miho, Miyagawa and Okitsu.
- Common mandarins (*C. reticulata* Blanco) are a very extensive group of different cultivars widely grown and available in world markets, e.g., Imperial, Nova, Ponkan, Sunburst, and Fortune. The group also includes the seedless clementines (like Marisol, Clemenules, and Hernandina), and two hybrid groups, e.g., tangors, which are hybrids of the mandarin and orange (like Murcott and Ellendale), and tangelos, hybrids of the mandarin and grapefruit or pummelo (Minneola and Seminole).
- Mediterranean mandarin (*C. deliciosa* Tenore) is also known as 'Willowleaf' and includes Avana clones.

Lemons

The lemon (*C. limon* [L.] Burm. f.) is an important fresh fruit group in world markets. Even though they are not eaten fresh, they are widely used for their acid juice content and slices, and for processing. The main cultivars are dual purpose and include Eureka, Lisbon, Meyer (a hybrid), Fino, and Verna. A minor related group of sweet or acidless cultivars and hybrids also exists.

Limes

Limes are also a varied group with a distinctive flavor and aroma, consisting mainly of two broad subgroups:

- Acid limes are small fruited (*C. aurantifolia* Swing) with West Indian, Mexican, or Key lime as the main cultivar, and often also processed into cordial; the large fruited (*C. latifolia* Tan) with Persian, Tahiti, or Bearss the main cultivar, is normally marketed as a fresh seedless fruit (Figure 2).
- Sweet limes (*C. limettioides* Tan), also known as the Palestine or Indian sweet lime, are insipid and not normally attractive to Western palates.

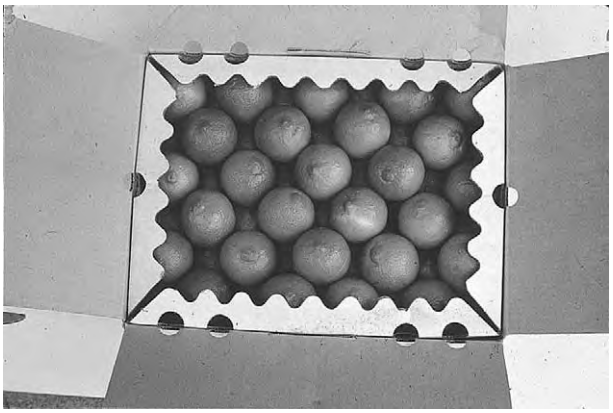

Figure 2 (see color plate 26) Limes are normally harvested for fresh marketing when mature but with green or light green to silver-colored rinds.

Grapefruit

Grapefruit is one of the newer types (*C. paradisi* Macfadyen) of citrus. It contains several distinct groups according to the rind or flesh-colored pigments:

- Common or white fleshed, with Marsh (seedless) the main cultivar.
- Pigmented, where the older cultivars were considered 'pink'-fleshed (with Thompson and Ruby the main cultivars), while newer selections, mutations or cultivars bred in the USA are now classified as 'red'-fleshed, due to the deeper pigmentation in both the rind and flesh while also being seedless. From this group, Star Ruby, Flame, and Rio Red are becoming important in world markets (Figure 3).

A recent breeding program in California, where an acidless pummelo was crossed with a white seedy grapefruit, has resulted in the selection and release of two new low-acid seedless grapefruit type cultivars – Melogold and Oroblanco (also called 'Sweetie' in Israel and some other countries).

Growing Regions and Conditions

Production regions are located in a wide range of climatic conditions, including the humid tropics, arid subtropics and intermediate climates. Commercial production tends to be located in two narrow belts in the subtropics and between 20 and 40° latitude north and south of the equator.

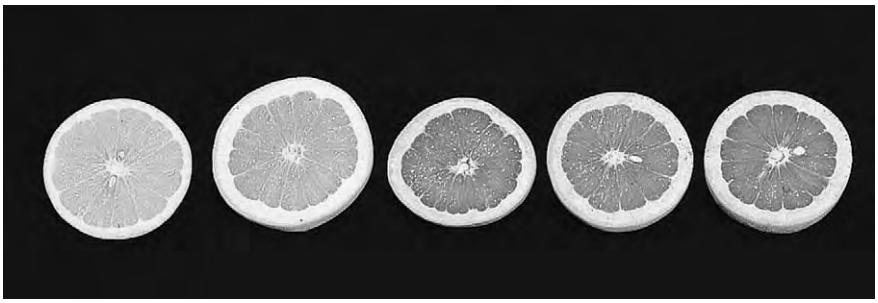

Figure 3 (see color plate 28) A range of internal pigments are now available in grapefruit, with consumer interest in the newer red fleshed cultivars (left to right: Marsh, Thompson, Rio Red, Flame, and Star Ruby).

Table 1 Fresh citrus crop production and utilization (with respective percentages), and the main producing countries of each for the major types produced throughout the world in 1998–1999

Citrus type	*Total production ($\times 10^3$ t) %*	*Exports ($\times 10^3$ t) %*	*Processed ($\times 10^3$ t) %*	*Main producing countries*
Oranges	53 644 65	4 327 48	22 784 82	Brazil, USA, Mexico, Spain, China
Mandarins	14 463 18	2 191 24	1 206 5	China, Spain, Japan, Brazil, Korea
Lemons and limes	9 446 11	1 362 15	1 991 7	Argentina, Mexico, Brazil, Spain, Egypt
Grapefruit	4 614 6	1 135 13	1 682 6	USA, Israel, Cuba, China, Argentina
Total world	82 166	9 015 11	27 664 34	Brazil, USA, China, Spain, Mexico

Citrus grows best in cooler, frost-free, Mediterranean-type climates, provided that the soils are suitable and rainfall is at least 1200 mm, distributed evenly throughout the year, or supplemented by irrigation. Climate can significantly affect yield, fruit quality, and tree health. In particular, minimum temperatures and degrees of frost restrict commercial production within many countries of the world.

Some cultivars have been bred or selected that extend production into marginally cooler climates, e.g., satsuma mandarins in Japan and China. Limes and pummelos grow well in tropical areas, where the normal rind color of oranges does not develop. High-quality grapefruit tend to be grown in hot desert areas.

Production Statistics

Commercial citrus production has been recorded by the FAO in over 100 different countries and through out six regions – Africa, North America, South America, Asia, Europe (Mediterranean), and Oceania. The world's citrus production has been gradually increasing through a 3-year average of $87\,980 \times 10^3$ t in 1995–1998, to a slightly smaller total crop of $82\,166 \times 10^3$ t in 1998–1999, as shown in Table 1. The production for the main types for 1998–1999 is shown in Figure 4. 'Other' citrus types totaled only 5124×10^3 t, with the majority recorded in Asia. The 4-year 1995–1999 production trends and percentage of the main citrus types grown are shown in Figure 5.

The largest growing areas are located in the northern hemisphere, where 68% of world production occurs, and include the important Mediterranean region with 20%, e.g., Spain, Egypt, Italy and the USA, China, and Mexico, whereas the southern hemisphere (32%) includes the largest producer in the world – Brazil, together with Argentina and South Africa. Nonspecified areas of citrus production with just over $16\,000 \times 10^3$ t in 1998–1999 would mainly occur in the subcontinent (especially India) and Asia.

Some regions like Japan, China, India, Mexico, and Asia have important local or domestic markets for their fresh citrus fruits, whereas others like Spain, USA, South Africa, Morocco, Turkey, and Argentina depend on exports as a major outlet for much of their production. Table 1 indicates total world fresh fruit exports of 11% in 1998–1999 for the main citrus types and with FAO recording 92 countries exporting some citrus.

In the two largest producing countries (Brazil and the USA), there has been an increasing trend for a large percentage of the production to be processed into frozen concentrated orange juice for storage and/or export throughout the world. The processing of some citrus production in Spain, Argentina, and Italy is also an important fresh juice market (Figure 6). Overall, in 1998–1999, $27\,664 \times 10^3$ t were processed, or 34% of total citrus production

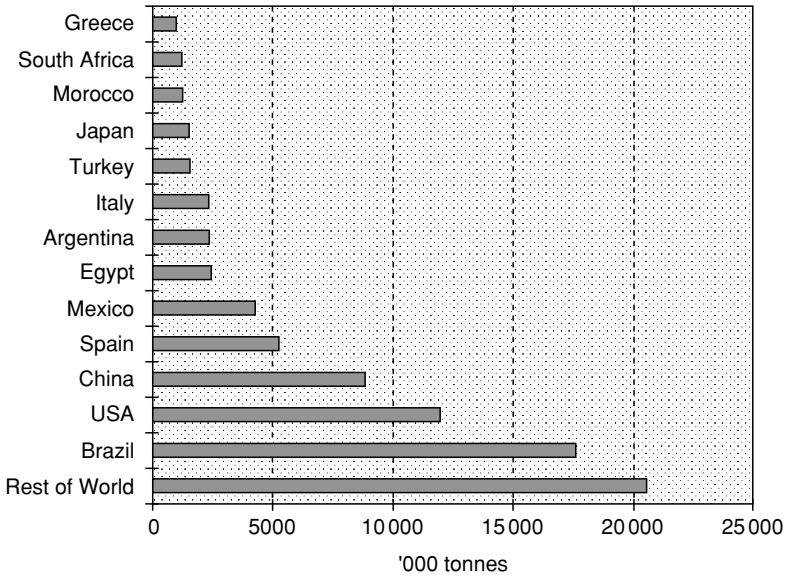

Figure 4 World citrus production, 1998–1999 for major producers and the rest of the world (which includes Cuba, Israel, Australia, Venezuela, South Korea, Lebanon, Uruguay and Cyprus).

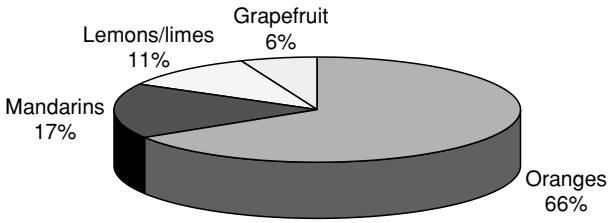

Figure 5 (see color plate 27) Average production percentages for the main citrus types grown in the world between 1995 and 1996 and between 1998 and 1999.

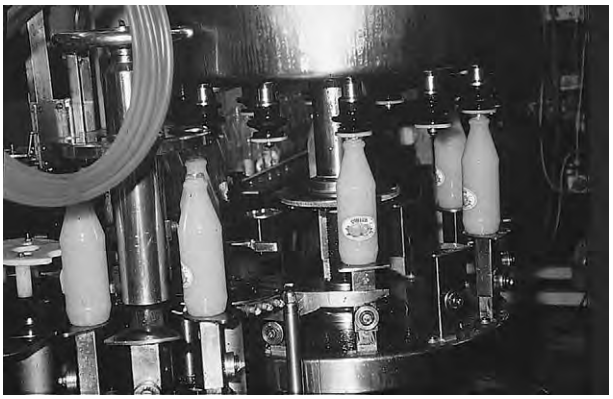

Figure 6 Juicing of oranges into various products is an important market in many countries. Fresh juice and citrus drinks are normally sold in consumer convenience packs.

(Table 1), whereas Brazil is reported to have processed 73% of their total orange production in 1989–1999, and the USA 88%.

Utilization of Lesser-Known Citrus Types

Some minor and lesser-known citrus fruit types that have fresh or processing uses and have some market importance in different countries or regions of the world are listed in alphabetical order. In many situations, local names exist for some of these types.

Bergamot (*C. bergamia* Risso)

This appears to be a variant of the sour orange and is mainly grown for its unique aromatic rind oil or essence in the coastal region of Calabria, a province of Italy. This essence is widely used in the food trade and in confectionery, and as an aroma for liqueurs, teas, sweets, and candied fruit.

Calamondin (*C. madurensis* Loureiro)

This resembles the mandarin and is grown extensively in the Orient, China, Taiwan, Japan, and the Philippines (where it is known as the 'Calamonding' or 'Kalamansi'). The fresh fruit is sour but is widely used for processing and as an ornamental tree.

Citrons (*C. medica* L.)

Like the other members of the acid group of lemons and limes, these also have two classes, the acid and the sweet, each with several cultivars. Citron are principally used for candied peel, but the etrog citron is also used in Jewish religious ceremonies. The Fingered Citron or Buddha's Hand, has an unusual appearance, with a distinctive and pervasive aroma,

Figure 7 Fresh or dried Kaffir lime leaves are an important garnish in many Asian recipes. (Market packaged dried leaves (L) and freshly picked leaves showing their distinctive shape (R)).

and is also used in religious ceremonies, perfuming rooms, and clothing for its medicinal values, and serves to highlight fruit-bowl arrangements.

Kaffir Lime (*C. hystrix* DC)

This belongs to the subgenus *Papeda* and is not a true 'lime,' but as the Kaffir lime is widely grown in Thailand, Philippines, Malaysia, Indonesia, and generally throughout Asia. It has unusual looking leaves (having a large winged petiole, see Figure 7) that are used (fresh or dried) with the fruit rind (as the fruit is normally not eaten) as a seasoning, flavoring, and garnishing in Asian food or recipes.

Kumquats (*Fortunella* spp.)

Also known as 'cumquats,' these are the smallest of the true citrus fruits. They can be eaten as fresh fruit or marmalade, or can be brandied. Several distinct cultivars that bear edible fruits are known:

- Nagami or oval kumquat (*F. margarita* [Lour.] Swing)
- Marumi or round kumquat (*F. japonica* [Thumb.] Swing)
- Meiwa or large round kumquat (*F. crassifolia* Swing).

Several minor kumquats are also grown mainly for ornamental purposes, e.g., Hong Kong (*F. hindsii* Swing) and Malayan (*F. polyandra* Tan).

Pummelo (*C. grandis* [L.] Osbeck)

This is also known as 'shaddock' or 'pomelo' and is the largest of the citrus fruits. Because pummelo hybridizes very readily, this has resulted in acid, acidless or sweet, pigmented or nonpigmented flesh, seedy or seedless cultivars being grown. This citrus type is common in some Asian countries, e.g., Thailand, China, and Indonesia, but is not common in Western countries or in world markets. It is usually eaten as a spooned desert fruit.

Sour oranges (*C. aurantium* L.)

Also known as bitter oranges, sour oranges are normally used for processing (marmalade or rind oil to flavor soft drinks and liqueurs). Most cultivars can be classified into three groups:

- Common, bitter, or sour oranges, of which the Seville (rough or smoothed skinned) is the most important.
- The bittersweet orange is a subgroup of the above common sour orange but has a lower acidity and better flavor.
- Variant bitter oranges or Bouquetiers are an important perfumery cultivar.

Conclusion

Many types of citrus fruits and processed products are now widely available throughout the world from local production or importation to supplement local supplies or in non-producing countries. Citrus is mainly consumed as fresh fruit, with their distinctive refreshing appeal and flavor, while also being a healthy, enjoyable, convenient and nutritious food. The fruit is also an excellent source of vitamins, minerals, dietary fiber and still low in kilojoules. Some fruit are used extensively to garnish food, while many processed products especially juices, cordials and marmalades are also widely available and consumed.

See also: **Ascorbic Acid**: Properties and Determination; Physiology; **Citrus Fruits**: Composition and Characterization; Processed and Derived Products of Oranges; **Jams and Preserves**: Methods of Manufacture; Chemistry of Manufacture

Further Reading

Cottin R (coordinator) (1997) *Citrus of the World A Citrus Directory*. San Giuliano, France: SRA INRA-CIRAD.

FAO (1999) *Citrus Fruit Fresh and Processed, Annual Statistics*, Tables 1–10 and 16–20. Rome: Food and Agriculture Organization.

Reuther W, Webber HJ and Batchelor LD (eds) (1967) *The Citrus Industry*, vol. 1. Berkley, CA: University of California.

Saunt J (1990) *Citrus Varieties of the World*. Norwich, UK: Sinclair International.

Sinclair WB (1984) *The Biochemistry and Physiology of the Lemon and other Citrus Fruits*, pp. 711–718. Berkley, CA: University of California.

Swingle WT and Reece PC (1967) *The Botany of Citrus and its Wild Relatives*. In: Reuther W, Webber HJ and Batchelor LD (eds), vol. 1. *The Citrus Industry*. Berkley, CA: University of California.

Young RH (1986) *Fresh Citrus Cultivars*. In: Wardowski WF, Nagy S and Grierson W (eds), pp. 110–125. *Fresh Citrus Fruits*. Westpoint, CT: AVI Publishing.

USDA (1999) *World Fresh Citrus Situation*. In: *World Horticultural Trade and US Export Opportunities, Circular Series FHORTH*, pp. 18–32. USA: Foreign Agricultural Service.

Composition and Characterization

L Izquierdo, Formerly of Consejo Superior de Investigaciones Científicas, Valencia, Spain

J M Sendra, Instituto de Agroquímica y Technología de Alimentos, Valencia, Spain

Introduction

Citrus fruits are one of the largest fruit crops in the world. About 30% of citrus fruits is processed to obtain various products, mainly juice. Similarly, the citrus industry is also the second largest fruit-processing industry, surpassed again by the grape industry, which mainly produces wine. Neither orange juice nor wine can be considered essential foods but they do have an important role in our lives.

Although citrus fruits have been consumed since ancient times, citrus processing, as it is known today, was not possible until thermal treatment (to inactivate enzymes and microorganisms) and concentration processes were commercially available. Since then, the citrus industry has developed rapidly, becoming prominent among food industries.

Although consumption of fresh citrus fruits is popular in all producing countries, processed products must still be considered almost as luxury products. Breakfast with orange juice is only common in developed countries. Thus, citrus industries process value-added products whose quality, nutritional characteristics, and purity are appreciated. Since these three aspects are closely related to composition, the analysis of citrus constituents is a frequent subject of research work, supported by governments and industries.

This article covers the most important aspects of citrus fruit composition, its relationship to nutritional value, and its importance for product authentication. Several books have been published on these subjects and the Further reading section lists some of them as well as published composition tables.

Constituents

The genus *Citrus* has many species and the differences among them are of great interest to specialists. However, from a general point of view, the similarities are greater than the differences, which is not surprising when considering species of the same genus. The economic importance also differs among species and compositional studies of the main species are more frequent. Thus, data from *Citrus sinensis* (sweet orange) are more extensive than from *C. reticulata* (tangerine), *C. limon* (lemon), or *C. paradisi* (grapefruit), and data from these species are more comprehensive than from other *Citrus* species. Within each species some varieties are better known than others. Valencia orange is the best-studied sweet orange, since it is the most important variety for juice extraction, the main citrus product. Thus, most information given here will refer to the juice rather than to the fruit, since juice, accounting for about half of the total weight of an orange, is the most important part of the fruit, and is the part of the fruit mainly consumed by humans.

Orange peel constitutes most of the other half of the fruit, but peel is of much less importance than juice. Although some byproducts (cattle feed, molasses) are obtained from peel, it is more a question of removing residues and avoiding pollution than of economic interest. Only peel oil (obtained before or during juice obtention) and pectins (obtained only from suitable species and varieties) are important peel products for human consumption, perfumery, and cosmetics.

Our knowledge of the chemical composition of juice and fruits is being continuously improved. The efficiency of instrumental methods of analysis allows the rapid identification of more and more minor constituents. But the basic major constituents have been well known since the application to citrus research of classical methods of analysis, which are still of interest for some rapid determinations. Both types of methods will be considered in this article, but it must be pointed out that analysis is a dynamic discipline and methods (mainly instrumental) are continuously being improved. Many of those which seem almost perfect today will look old-fashioned tomorrow.

Citrus fruit parts are represented schematically in **Figure 1**, and the approximate constitution of oranges is shown in **Table 1**. Juice vesicles, located in the endocarp, contain the juice, which constitutes about 50% of the total weight of a typical orange fruit. The peel is formed by the flavedo (epicarp and outer mesocarp) and the albedo (inner mesocarp). Flavedo and albedo account, respectively, for about 10 and

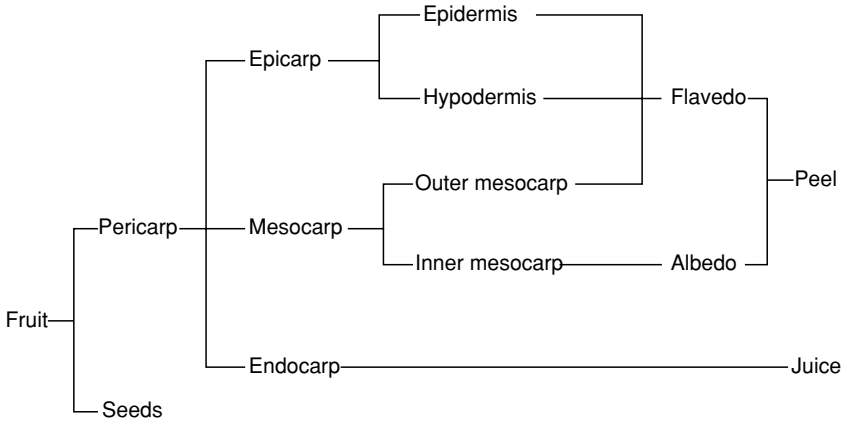

Figure 1 Structure of citrus fruits. Reproduced from Citrus Fruits: Composition and characterization, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 1 Approximate composition of oranges

Overall composition (%)	
Juice	40–50
Flavedo	8–10
Albedo	15–25
Rag, pulp, and seeds	Rest
Main constituents (%)	
Water	85–90
Sugars	6–9
Acids	0.5–1.5
Pectins	0.5–1.5
Minerals	0.5–0.8
Essential oils	0.2–0.5
Fiber	0.5–1.0
Protein	0.5–0.8
Fat	0.1–0.2

25% of the whole fruit. The flavedo contains peel oil in the oil sacs and the albedo most of the pectins. After juice extraction, rag, pulp, and seeds account for the rest of the total fruit weight. Most of the fruit (almost 90%), is water. The rest is mainly formed by sugars and acids. Minerals, amino acids, aroma compounds, and pectins are present in smaller proportions. These percentages vary greatly in other citrus fruits, but they also vary within orange fruits, depending on variety, ripeness, area, and culture practices. Concentrations, distribution, and methods of analysis of these constituents will be discussed in more detail in the following sections.

Sugars and Acids

Soluble solids of citrus juices consist mainly of sugars and acids. Sugars are more abundant (70–80% of total solids) in oranges, tangerines, and grapefruits, whereas citric acid is predominant in ripe lemons and limes. Soluble solids are usually expressed in degrees Brix (for easy and rapid evaluation with refractometers), referring to all juice solids as sucrose. Total acidity is expressed in grams of citric acid per 100 ml of juice, approximately equivalent to percentage.

The total solids increase with ripeness in all citrus species. In most cases, this is due to the increase in sugar content but, in lemons, citric acid increases with ripeness. Ranges from 9 to 14°Brix and from 1.5 to 0.5% acidity can be considered usual in orange juices processed in citrus plants. Commercial juices, prepared by blending juices of different varieties at different degrees of ripeness, show less variations: from 11 to 12°Brix and from 0.8 to 0.9% acidity are normal ranges. (*See* **Ripening of Fruit**.)

The degrees Brix-to-acidity ratio, also called the maturity index, is a fundamental measure of quality of citrus juices. In commercial orange juice it usually has values from 12 to 14 and no orange juice is considered of high quality if it shows a ratio lower than 11. The importance of the degrees Brix-to-acidity ratio as a quality index lies in its relationship with the equilibrium between sweetness and sourness, one of the most appreciated sensory characteristics of citrus fruits. As discussed above, degree Brix value depends mainly on sugars.

Only three sugars are important in citrus fruits, sucrose, glucose, and fructose, which produce juice (or fruit) sweetness. Other sugars, such as galactose, heptulose, mannose, rhamnose, xylose, or trehalose, have only been detected as traces. (*See* **Carbohydrates**: Classification and Properties; **Fructose**; **Sucrose**: Properties and Determination.)

The analysis of citrus sugars has been approached by many different methods. A classical standard method is the determination of reducing sugars (fructose and glucose) by copper reduction, followed by

the inversion of sucrose to fructose and glucose, which can then also be determined as reducing sugars. By this method, the concentrations of the main citrus sugars (reducing and nonreducing) are determined. When the sugar composition must be known in more detail, separative (chromatographic) methods are used. Numerous analyses of citrus sugars have been performed by paper chromatography, gas chromatography, and high-performance liquid chromatography. Traces of the different minor sugars mentioned above have been detected by these techniques. The analyses of citrus sugars have shown that the typical ratios of sucrose to glucose to fructose are 2:1:1 in orange juice, 1:4:4 in lemons and 1:3:3 in limes. Typical contents of total sugars are 8% of total weight in orange and grapefruit juices, and 2% in lemon juice, although these percentages can vary greatly with variety and ripeness. (*See* **Chromatography**: Principles; High-performance Liquid Chromatography; Gas Chromatography.)

The major acid in citrus fruits is citric acid. Total acidity, which is determined by titration, is expressed as grams of citric acid per 100 ml of juice. Malic acid is the second most important acid, present in variable but much lower amounts than citric acid. Other organic acids, such as succinic, malonic, lactic, oxalic, phosphoric, tartaric, adipic, isocitric, etc., are present in even smaller amounts and have been detected by separative methods, such as gas chromatography of methyl ester derivatives. (*See* **Acids**: Properties and Determination.)

Acid content, like total solids, varies greatly with ripeness. Ranges from 0.5 to 1.5% in orange juice, from 0.8 to 2.5% in grapefruits, and from 4 to 8% in lemons and limes are usual.

Polysaccharides

Citrus fruits contain cellulose, hemicelluloses, and pectic substances, mainly in the peel. The juice contains only small amounts of these compounds, although they play an important role. From the nutritional point of view, they provide dietary fiber. Concerning quality, pectin acts as a colloidal stabilizer, protecting juice cloud. Pectins are polymers of galacturonic acid and exist in juices at levels of about 0.5%. In the form of the methyl ester, polygalacturonic acid maintains juice turbidity. If deesterification occurs, the colloid degrades and the juice clarifies. The main reason for thermal treatment of commercial juices is not only to destroy microorganisms, but to inactivate the enzyme pectinesterase, which produces pectin degradation. Citrus peel, mainly from lemon, is an important source of commercial pectins, which is of great interest in the food industry because of their gelling power. (*See* **Dietary Fiber**: Properties and Sources.)

Nitrogenous Compounds

About 0.7% of citrus juices consists of nitrogenous compounds. Free amino acids are the most important part of the nitrogenous fraction (70%); the rest consists of vitamins, proteins (mainly enzymes), and inorganic compounds. Although not important as nutrients due to their low concentration, amino acids are extensively used for characterization purposes and to detect adulteration in fruit juices. Total nitrogen determination (closely related to total amino acid content) is a standard method of purity control. Nevertheless, the nitrogen content can be easily manipulated by adding inorganic salts. For this reason, separative techniques allowing the determination of individual amino acids are of considerable interest. Gas chromatography of volatile derivatives has often been used for this purpose, as has ion exchange chromatography. Reversed-phase liquid chromatography is an increasingly accepted technique. Figure 2 shows chromatograms obtained by this method from the dansyl derivatives of orange juice amino acids. (*See*

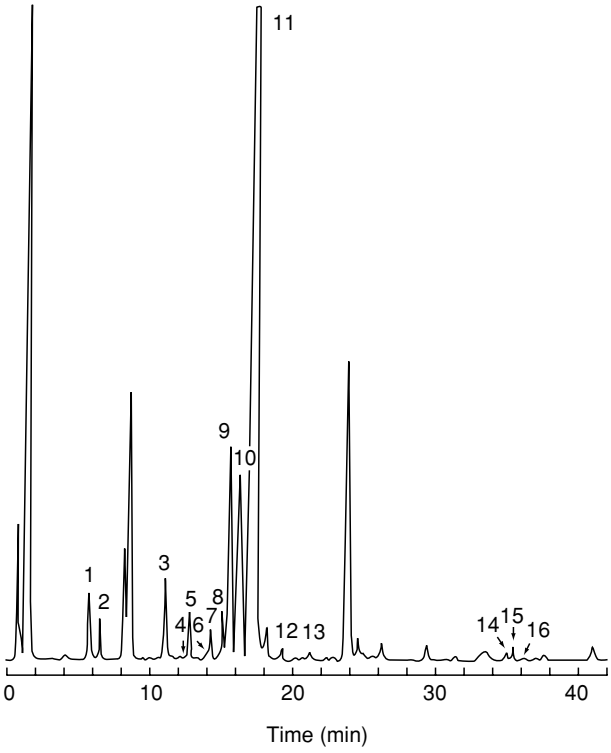

Figure 2 Liquid chromatogram of orange juice amino acids: 1, aspartic acid; 2, glutamic acid; 3, asparagine; 4, glutamine; 5, serine; 6, threonine; 7, glycine; 8, alanine; 9, arginine; 10, γ-aminobutyric acid; 11, proline; 12, valine; 13, methionine; 14, ornithine; 15, lysine; 16, histidine. Reproduced from Citrus Fruits: Composition and characterization, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Amino Acids: Determination; **Chromatography**: Thin-layer Chromatography.)

Bitter Compounds

Two main types of bitterness, caused by two different types of compounds, occur in citrus fruits. Flavanone neohesperidosides, as naringin in grapefruit and neohesperidin in sour oranges, produce the typical bitterness of fruits and juices from these species. The other type of bitterness, which constitutes an extremely negative quality factor in some orange juices, is produced by limonin, a triterpene derivative of the limonoid group. Limonin bitterness is known as 'delayed bitterness' since it is not detected in fresh fruits or freshly extracted juices, but develops during juice storage or with heat treatment. The reason is that fresh fruits do not contain limonin, but a nonbitter precursor which converts into limonin after juice preparation. Limonin is detected by taste at concentrations of about 6–8 mg l^{-1} in orange juice. Citrus industries avoid exceeding these concentrations in commercial juices by using only proper fruits or by blending different raw materials to obtain the desired final products. The limonin content of orange juice depends on variety and ripeness. It is extremely high in Navel varieties (still about 15 mg l^{-1} at a degrees Brix-to-acidity ratio of 12), which makes them very unsuitable for juice processing. Generally, early varieties produce more bitter juices than late varieties but country/region of origin can also affect limonin contents. Problems of bitterness are worse in temperate areas (California, Australia, Spain) than in tropical or subtropical regions (Brazil, Florida). (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Reduction of bitterness has been attempted by many means, involving changes in cultivation practices (rootstock, fertilization) and juice treatments. Debittering of processed juices seems to be the most promising approach and some citrus industries are already equipped with debittering devices.

Limonin analysis is essential in quality control. Thin-layer chromatography or immunoassay methods as well as liquid chromatography techniques are used in its determination.

Minerals

Amounts and proportions of minerals contained in citrus juices are interesting for nutrition. Total mineral content, referred to as 'ash,' is about 0.4% of juice weight. Ash composition varies in relation to the species, ripeness, growth area, and methods of cultivation. Generally, the major inorganic components of citrus juices are potassium and nitrogen, each accounting for about 40% of the total. Calcium, magnesium, and phosphorus are less abundant (about 5% each). Sodium, iron, sulfur, and chlorine are also important. About 20 more minerals have been detected in smaller proportions or only as traces. Some of these minor constituents have been proposed for ascertaining the geographic origin of orange juice. Minerals are very satisfactorily determined by atomic absorption spectrometry or plasma emission techniques, where available. Refer to individual minerals.

Aromas

The typical citrus aroma depends on many minor constituents. Their individual contribution and relative proportions, as well as interactions among them, give the characteristic odor to each species. These constituents are present in juice vesicles of the endocarp and, in larger amounts and different proportions, in the oil sacs of the flavedo. When extracting the juice, oil sacs break and part of the oil joins the juice. Peel oil is also obtained as an independent product in citrus plants. One of its principal uses is as a flavoring ingredient for soft drinks.

Odor compounds constitute a part of the volatile fraction of citrus juices, which also contain nonodorous compounds. The distinction between both types of constituents and the evaluation of their relative importance in juice aromas are important research objectives in this field. Table 2 shows the main types of compounds forming the volatile fraction of orange juice. Several examples of individual constituents, including some which cause off-flavors, are listed in the table. Major constituents of the volatile fraction are ethanol (about 0.5 ml l^{-1} in juice) and limonene (0.15 ml l^{-1}). None of these are important for aroma, since their concentrations are lower than detection thresholds. Chromatography, preferably gas chromatography using capillary columns of high resolution coupled to a mass spectrometer, is the most suitable method for determining either individual compounds or quite large groups of volatile compounds in a single analysis. Due to the low concentrations in the juice of most volatiles, a concentration step may be necessary for detailed studies. This step can be performed by solvent extraction of the juice or of a juice distillate, followed by solvent removal.

For rapid determinations, necessary in routine quality control, the total analysis of recoverable oil or the specific analysis of some groups of constituents, such as alcohols, aldehydes, and ketones, or unsaturated constituents, is also used. (*See* **Alcohol**: Properties and Determination; **Essential Oils**: Properties and Uses.)

Other Minor Constituents

Juice color in all citrus fruits is due to carotenoids, with the exception of blood oranges in which

Table 2 Composition of the volatile fraction of orange juice

Types of compound	*Approximate number of identified constituents of the type*	*Examples*
Alcohols	54	Linalool
		α-Terpineol[a]
		4-Vinylguaiacol[a]
		Citronellol
		Nerol
		Octanol
		Geraniol
		Methanol
		Ethanol
Aldehydes	41	Acetaldehyde
		Hexanal[a]
		Citronellal
		Geranial
		Neral
Ketones	16	Carvone[a]
		Nootkatone[a]
		Acetone
Esters	39	Ethyl butyrate
		Methyl butyrate
		Ethyl acetate
		Linalyl acetate
Hydrocarbons	51	α-Pinene
		Terpinolene[a]
		Valencene
		Myrcene
		Limonene
Acids	10	Acetic
		Butyric
Others	12	Ethyl butyl ether
		Linalool oxides

[a]Some constituents produce off-flavor.

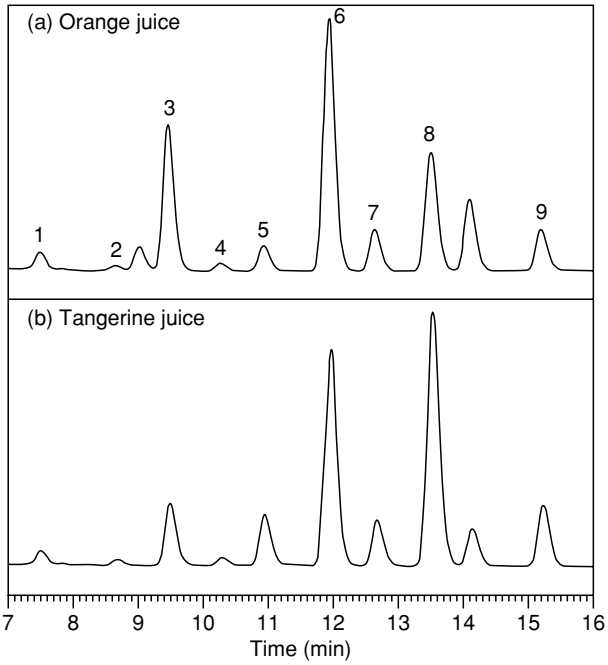

Figure 3 Liquid chromatogram of fully methoxylated flavones of (a) orange juice and (b) tangerine juice. 1, Isosinensetin; 2, gossypetin; 3, sinensetin; 4, isoscutellarein; 5, quercetagetin; 6, nobiletin; 7, scutellarein; 8, heptamethoxyflavone; 9, tangeretin. Reproduced from Sendra JM, Navarro JL, and Izquierdo L (1988) C_{18} solid-phase isolation and high-performance liquid chromatography/ultraviolet diode array determination of fully methoxylated flavones in citrus juices. *Journal of Chromatographic Science* 26: 443–448.

anthocyanins, types of flavonoid compounds, are responsible for their typical color. β-Carotene is present in orange juice in concentrations of about 0.5 mg l^{-1}, representing a small but significant source of vitamin A. Color is one of the principal quality attributes of citrus juices and several methods, either visual or instrumental, have been used for its evaluation. The Hunter colorimeter must be mentioned since it was specifically developed for citrus fruits. Basically it determines three parameters: *L*, *a*, and *b*, respectively related to lightness, green-to-red, and blue-to-yellow preponderance. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments.)

Lipids constitute about 0.07% of orange juice and consist mainly of fatty acids, 90% of them being palmitic, palmitoleic, oleic, linoleic, and linolenic acids. The rest of the lipid fraction includes a range of both polar and nonpolar compounds. Lipids can be partially responsible for off-flavor development since, even though they contribute little to flavor, they are precursors of malodorous compounds. Independently, lipids are important for taxonomy because each different species has its own profile of lipid components. (*See* **Fatty Acids**: Analysis.)

Fully methoxylated flavones, belonging to the flavonoid group, are considerably more abundant in the peel than in the juice, where their concentration increases with extracting pressure. Consequently, these compounds can be considered as a potential index of extractor performance. Figure 3 shows chromatograms obtained by liquid chromatography of fully methoxylated flavones of orange and tangerine juices.

Vitamins are very important minor constituents of citrus fruits, and will be discussed in the following section.

Nutritional Value

Although citrus fruits cannot be considered as basic foods, they are excellent complementary foods to appropriate diets. During the eighteenth century, citrus fruits were the only known source of vitamin C able to prevent scurvy on long sea journeys, but this was a long time ago. Nevertheless, the contribution to the diet of vitamins, minerals, and dietary fiber from citrus fruits is still significant. (*See* **Scurvy**.)

Orange juice is an important source of vitamins. Composition tables show that it contains (per 100 ml) about 100 μg thiamin, 30 μg riboflavin, 40 μg vitamin B_6, 20 μg vitamin E, 300 μg niacin, 200 μg pantothenic acid, 300 μg folacin and 50 mg vitamin C. Thus, a glass of orange juice covers the daily requirements of vitamin C, about 25% of those for folacin, and 5% for vitamin B_6. This contribution to dietary needs increases in interest if the low energy content of orange juice (about 50 cal per 100 ml) is considered as this allows higher consumption of orange juice than of other foods richer in vitamins (but also in energy). On the other hand, commercial orange juice needs less heat treatment for microorganism inactivation than other less acid foods, so vitamin losses are lower. Refer to individual vitamins.

Lemon juice contains similar amounts of vitamins C and B_6, but less folacin (100 μg per 100 ml), whereas tangerine juice is poorer in vitamin C (30 mg per 100 ml), and grapefruit juice has a lower content of all these three vitamins.

Concerning minerals, citrus juices are rich in potassium (150–200 mg l^{-1}) but low in sodium (1–2 mg l^{-1}), which makes them very suitable for people receiving treatment with diuretics.

Finally, cellulose, hemicelluloses, and pectins contained in citrus fruits are a source of dietary fiber, but which will make only a small contribution to daily intake for most people. (*See* **Dietary Fiber**: Properties and Sources.)

Authentication

Citrus processing is an important world industry and some products reach high prices, which may make the idea of adulteration attractive. Although typical adulteration of citrus juices (by dilution, addition of sugar, blends of juices from different species without declaration) is not harmful to health, the consumer is paying for a substandard product, and unfair competition against honest producers is always involved. (*See* **Adulteration of Foods**: Detection.)

The methods for detecting adulteration generally consist of comparing the values of some selected characteristics (i.e., constituent concentrations) of suspicious samples with the known ranges of values in pure products. The effectiveness of the method will depend mainly on the success in selecting those characteristics which are specific to the pure product, easy to determine accurately, and difficult to mask. For a sound knowledge of the characteristics of pure products, sufficient data, representative of the possible raw materials, must be collected. Data variability is usually high due to ripeness, area, or fruit variety. Consequently, values show wide ranges, which complicates the detection of adulteration. Mean values and ranges of several characteristics of Spanish orange juices are presented in Table 3.

Table 3 Mean values and ranges of some characteristics of Spanish orange juices (values in mg l^{-1}, except for absorbances)

	Mean	*Minimum*	*Maximum*
Aspartic acid	165.6	26.9	412.8
Glutamic acid	86.5	18.9	205.8
Asparagine	434.0	113.2	933.4
Glutamine	33.3	5.3	132.1
Serine	124.9	33.4	272.4
Threonine	20.9	8.8	50.0
Glycine	26.8	12.4	57.6
Alanine	89.5	49.7	182.6
Arginine	408.0	88.9	953.2
γ-Aminobutyric acid	250.7	83.4	470.4
Proline	1549.0	617.8	3973.0
Valine	14.8	3.4	62.4
Methionine	15.3	1.8	42.2
Ornithine	11.9	3.1	96.8
Lysine	32.8	11.1	62.4
Histidine	8.4	3.0	42.5
Acidity	11 000.0	5090.0	18 400.0
Absorbance at 280 nm	1.46	0.44	3.07
Absorbance at 325 nm	0.960	0.295	2.23
Absorbance at 443 nm	0.131	0.015	0.487
Sucrose	36 920.0	17 900.0	46 300.0
Glucose	20 880.0	13 190.0	29 240.0
Fructose	24 620.0	15 000.0	44 160.0
Potassium	1296.0	562.4	1872.0
Magnesium	98.4	74.6	146.7
Calcium	118.8	62.2	212.0
Ash	3165.0	2100.0	4248.0
Isocitric acid	117.1	40.6	207.2
Isosinensetin	0.129	0.02	0.61
Gossypetin	0.059	0.0	0.22
Sinensetin	1.62	0.07	7.15
Isoscutellarein	0.042	0.0	0.39
Quercetagetin	0.359	0.01	1.77
Nobiletin	2.24	0.0	9.12
Scutellarein	0.569	0.1	2.95
Heptamethoxyflavone	1.01	0.15	4.4
Tangeretin	0.321	0.02	2.81

Detection methods may be based on the determination of one, few, or many characteristics. Instrumental analysis, computers, and statistics are useful in all cases but essential where multiple data must be obtained and processed from each sample.

Several types of adulteration are possible in citrus products. The most important will be mentioned and the methods developed for their detection discussed.

Since water, sucrose, and citric acid are the main constituents of juices and can be obtained from cheaper sources, an easy type of adulteration consists of their addition to the juice in proportions that keep the solid content and acidity at normal juice levels. In this simple method, the adulteration should be easy to detect, since concentrations of the remaining con-

stituents decrease with dilution level. However, these concentrations can also be manipulated. Detection methods have increased in complexity by successively determining minerals, individual sugars, total nitrogen, nitrogen from amino acids, specific amino acids, and so on. Masking procedures also increased in their complexity by successively adding inorganic salts, blends of sugars, ammonium salts, glycine, amino acid blends, and so on. The ideal final situation to prevent adulteration will be achieved when masking procedures are either more difficult or more expensive than producing pure products.

The methods mentioned above work on the basis that all artificially added constituents, even from external sources, are normal constituents of juices. So, only through the modification of the natural proportions between added and nonadded constituents can fraud be detected. A different approach is to consider that raw materials from extraneous sources must show some differences. For example ratios of carbon (^{13}C:^{12}C) or oxygen (^{18}O:^{16}O) isotopes can be used to determine if any sugar or water comes from sources other than the citrus fruit.

Blends of juices from different species can be detected through the knowledge of the pattern distribution of certain compounds in each species. A method of detecting the addition of grapefruit juice to orange juice consists of analyzing the profiles of flavanone glycoside concentrations.

Finally, ultraviolet absorption is useful for determining the addition of pulpwash solids to orange juice.

See also: **Acids**: Properties and Determination; **Adulteration of Foods**: Detection; **Alcohol**: Properties and Determination; **Amino Acids**: Determination; **Carbohydrates**: Classification and Properties; **Chromatography**: Principles; Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Dietary Fiber**: Properties and Sources; **Essential Oils**: Properties and Uses; **Fatty Acids**: Analysis; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Fructose**; **Ripening of Fruit**; **Scurvy**; **Sucrose**: Properties and Determination

Further Reading

Feinberg M, Favier JC and Ireland-Ripert J (eds) (1991) *Répertoire Géneral des Aliments. Table de Composition*. Paris: Institut National de la Recherche Agronomique.

Kawaii S, Tomono Y, Katase E *et al*. (2001) Quantitative study of fruit flavonoids in citrus hybrids of King (*C. nobilis*) and Mukaku Kishu (*C. Kinokuni*). *Journal of Agricultural Food Chemistry* 49(8): 3982–3986.

Lota ML, de Rocca Serra D, Tomi F, Jacquemond C and Casanova J (2002) Volatile components of peel and leaf oils of lemon and lime species. *Journal of Agricultural Food Chemistry* 50(4): 796–805.

Nagy S and Attaway J (eds) (1980) *Citrus Nutrition and Quality*. Washington, DC: American Chemical Society.

Nagy S, Shaw P and Veldhuis M (eds) (1977) *Citrus Science and Technology*. Westport: AVI.

Nagy S, Attaway J and Rhodes M (eds) (1988) *Adulteration of Fruit Juice Beverages*. New York: Marcel Dekker.

Sendra JM, Navarro JL and Izquierdo L (1988) C_{18} solid-phase isolation and high-performance liquid chromatography/ultraviolet diode array determination of fully methoxylated flavones in citrus juices. *Journal of Chromatographic Science* 26: 443–448.

Oranges

S Pao, CREC, Lake Alfred, FL, USA
P J Fellers, Winter Haven, FL, USA

Background

Of all the various fruits in the world, the sweet orange (*Citrus sinensis* Osbeck) is perhaps the most cherished. The refreshing natural flavor of orange flesh or juice from mature fruit is characterized as uniquely pleasing, with just the right mix of sweetness and tartness. Furthermore, the fruit's bright sunny color is especially appealing and often complements other food items on a dining table. Much effort has been expended in the attempt to imitate orange flavor and color.

The orange is the most widely grown and popular citrus species worldwide. Its origin is believed to be the tropical and subtropical regions of Asia. The mention of oranges in Chinese literature dates to 2200 BC, long before they were introduced to Europe – about 1400 AD Continental America probably received its first oranges in the early-to-mid 1500s with the arrival of the explorers Bernal Diaz in Mexico and Ponce de Leon in Florida. Brazil probably received its first oranges with the coming of the Portuguese in the mid-1500s. Oranges entered interstate commerce in the USA in the early 1800s with the shipping of fruit, usually in barrels, from coastal Florida locals to northern cities. Today oranges and orange products are shipped all over the world from areas of production.

Production and Utilization

Oranges are cultivated in both tropical and subtropical regions around the world. The quality of a

Table 1 Major orange-producing countries and amount ($\times 10^3$ t) exported and processed for 1998–99

Country	*Production*	*Fresh*[a]	*Fresh export*	*Fruit processed*
Brazil	15912	4162	82	11750
USA	8986	1456	245	7530
Mexico	2903	2543	50	360
China	2724	2595	5	129
Spain	2624	2034	1214	590
Egypt	1442	1340	215	102
Italy	1422	961	75	461
South Africa	989	764	500	225
Morocco	867	789	361	78
Greece	801	584	260	217
Turkey	780	722	102	58
Argentina	660	530	75	130
Australia	524	294	125	230
Cuba	450	110	10	340

[a]As calculated from production amount less the amount of fruit processed. From Food and Agriculture Organization of the United Nations (2000) *Citrus Fruit, Fresh and Processed. Annual Statistics*.

specific variety of fruit produced is largely influenced by rootstock, climate, types of soil, grove practice, and harvesting time. Commercial propagation of citrus is usually accomplished by bud-grafting a desired variety on to a suitable rootstock. Major rootstocks used in Florida for sweet orange varieties are Carrizo citrange, sour orange, Swingle citrumelo, Cleopatra mandarin, Volkameriana lemon, rough lemon, Milam lemon, and Palestine sweet lime. Orange varieties with different ripening periods are often cultivated to extend the availability of fruit from a particular producing region.

World production of oranges in 1998–99 amounted to 53×10^6 t, up 17×10^6 t from 1988–89. Of all the countries in the world, Brazil and the USA produce the lion's share of oranges (Table 1). Some lesser-producing countries not listed include Algeria, Cyprus, Israel, Japan, Lebanon, Tunisia, Uruguay, Former USSR, and Venezuela. About 58% of the world's orange production is consumed fresh. The USA and Brazil process the bulk of their crops, while countries like China and Egypt process very little of their crops. Spain is the world's largest exporter of fresh oranges, while Brazil is the largest exporter of processed orange juice. The leading orange fruit-importing countries in 1998–99 were Germany (467×10^3 t), The Netherlands (424×10^3 t), and France (423×10^3 t).

Anatomy and Varieties

A citrus fruit is botanically classified as a type of berry called a hesperidium that has a thick, leathery rind, with numerous oil glands, and a large flesh portion composed of several wedge-shaped sections. Unlike pome fruit, such as the apple, the citrus fruit is derived from a superior ovary, an ovary completely separate from the calyx. In general, the development of citrus fruit does not create a cavity at the blossom end of the fruit at any maturation stage.

Oranges are most often round, in contrast to the various shapes of many other citrus species. Generally, grapefruit are larger than oranges, while mandarins (especially tangerines), lemons, and limes are smaller. The major structural features of an orange from the outside to the center of an equatorial section are as follows: peel composed of the epidermis (outer layer of cells); flavedo, a layer of tissue containing plastids which are dark green when immature, and oil glands dispersed throughout; albedo, a white pulpy material surrounding the segments which in turn hold the individual juice vesicles or juice sacs; and seeds (if any) arranged around the pulpy core. Each segment is encased by a rather tough segment membrane or wall which, along with the juice vesicles, is ruptured when the fruit is extracted for the juice. Examples of seedy orange varieties are Pineapple and Parson Brown, whereas Hamlin, Washington navel, and Valencia are examples of seedless or nearly seedless varieties.

Oranges may be classified as sweet or bitter (sour), the latter having generally little commercial importance for the fruit, but great importance for the use as a rootstock. Sweet orange varieties are further classified, as follows: (1) common orange, represented by many important varieties, including Valencia (perhaps the major variety worldwide), Hamlin (important in Florida and Brazil), Pera and Natal (both of importance in Brazil), Pineapple (important in Florida), Parson Brown, Shamouti (of importance in Israel and several other Middle Eastern countries); (2) sugar or acidless orange, possessing a bland flavor; (3) blood or pigmented orange, important in the Mediterranean area of Europe, especially Spain; and (4) navel orange, especially Washington navel (of great importance in California). In addition, the Temple orange, a natural tangor, has economic significance and is being cultivated in quantity in Florida.

Composition and Nutritional Value

An orange is composed of about 88% water. The chief soluble solids in orange juice are the soluble carbohydrates or sugars (*c.* 10%), while citric acid is the main soluble solid in lemons and limes. The chief sugars in orange juice – glucose, fructose, and sucrose – occur in the approximate ratio of 1:1:2. Total sugars (especially the sucrose fraction) generally increase

with advancing maturity, and the stylar half of the fruit possesses more than the stem half.

Fresh oranges and grapefruit contain dietary fiber, being mostly soluble fibers such as pectic substances. Fiber aids in digestion and elimination, and when part of a low-fat diet rich in fruits and vegetables, the fiber therein may help reduce the risk of some cancers. Also, low-fat diets rich in fruits and vegetables containing soluble fiber may reduce the risk of heart disease. Pectin or pectic substances, made up of complex carbohydrate derivatives or polysaccharides, are important in terms of juice quality for imparting body or viscosity and the cloudy appearance. Pectic substances are a major component of primary cell wall and middle lamella in all fruit. In the orange, up to 30% of the albedo, on a dry-weight basis, may be pectin, but only 0.01–0.13% occurs in the juice. Without heat stabilization to inactivate pectinesterase enzyme, clarification or loss of cloud may occur in single-strength orange juice due to the settling out of pectins and other solids, while in concentrate a pectin gel may form, resulting in a nonpourable product. The pectinesterase in orange fruit is mainly associated with fruit peel, rag, and juice sac tissue (pulp). It is introduced to the juice from the ruptured cell membrane or wall tissue during the extraction process. In addition to pectinesterase, numerous other enzymes have been identified, including 18 oxidoreductases, nine lyases, nine transferases, seven hydrolases, three isomerases, and three ligases.

The acidity of orange juice is due primarily to the organic acid, citric acid and, to a much lesser extent, malic and succinic acids. Orange juice with about 0.5–1.0% acid (pH of about 3.5) is less acidic than that of grapefruit, and much less acidic than lemon and lime. Perhaps more so than any other quality factor, it is the favorable ratio of sugar to acid (sweetness to tartness), along with the unique orange flavor, that gives orange juice its universal high consumer acceptance. Orange flesh and orange juice are excellent sources of ascorbic acid (vitamin C). One medium-sized orange (about 154 g) may provide 120% of required daily vitamin C value. Vitamin C is an important antioxidant and scientists now believe that antioxidants may counteract harmful free radicals, which may contribute to the onset of several major diseases. Folate (folate acid), a B-vitamin commonly found in orange juice and green leafy vegetables, has been shown to help in the production of mature red blood cells. Recent research shows that women of child-bearing age can reduce the risk of having babies with birth defects of the brain and spine by consuming ample quantities of fruits and vegetables to maintain adequate levels of folate.

About 70% of the nitrogen in citrus juices occurs in free amino acids which constitute about 0.1% (w/w) in oranges. Proline is the major amino acid with less quantities of asparagine, aspartic acid, arginine, and gamma-aminobutyric acid.

The most abundant mineral in orange juice is potassium, with other minerals of significant quantity being calcium, iron, magnesium, phosphorus, and sulfur. Other elements of note are chlorine and nitrogen. Orange flesh and orange juice are considered to be excellent sources of potassium which is necessary to maintain fluid balance in the body, maintain cell strength and structure, and aid in nerve transmission.

Limonin, a triterpene derivative, because of its bitter nature, is an important constituent in certain citrus fruits, such as navel orange and grapefruit. When the juice of navel oranges is extracted, an undesirable bitter taste develops in the juice after a period of time. This is a problem of great significance in several large citrus-producing areas around the world, unless a debittering treatment is used. Limonin exists in the major sweet orange varieties (Hamlin, Pineapple, and Valencia) but generally in low enough levels not to produce significant bitterness problems.

Most of the lipid content of oranges occurs as fatty acids, with most being in the seeds. The five major fatty acids are palmitic, palmitoleic, oleic, linoleic, and linolenic. Phosphatidylcholine and phosphatidylethanolamine comprise the bulk of the phospholipid fraction of Florida and Calfornia orange juice. Orange essential oils are found primarily in the peel, with significantly less in the juice vesicles. One study shows that the amount of oil in fruit ranged from a low 0.29% in Hamlins to a high of 0.94% in Valencias. Orange oil contains at least 150 volatile essential oil components. Of these, *d*-limonene is most abundant, comprising about 95% by weight of orange oil, but it is the oxygenated aldehydes and esters that are considered the most important contributors to orange flavor. Many oxygenated alcohols are present in orange oil but not ketones. Coumarins and flavonoids comprise most of the 1% of the nonvolatile constituents in orange oil. Generall, total yield of orange oil decreases with increasing maturity.

An unusually large number and quantity of flavonoids, C_{15} compounds, occur in citrus. Compounds have been identified in oranges representing the three major flavonoid types: (1) flavanones (including flavanonols) generally occurring as glycosides in oranges, (2) flavones (including flavonols), and (3) anthocyanins, of importance in citrus only as red pigments in blood oranges.

The orange color in peel generally associated with mature fruit is due primarily to carotenoids, chiefly β-citraurin, while the green peel color generally

associated with immature oranges is due to chlorophyl. The main pigment in orange juice producing the orange color is the carotenoid, β-crytoxanthin.

Harvesting and Packinghouse Procedures

Oranges, like most other nonclimacteric fruits, do not continue to ripen after separation from the trees. Therefore, the fruit are picked only after they have reached maturity. In most citrus-producing areas, maturity standards must be met prior to the fruit being utilized in the fresh or processed markets. In the USA, the US Department of Agriculture and various state agencies develop and oversee the maturity standards. Internal standards generally consider a minimum ratio of Brix to percentage of acid (below which the juice will taste too acidic or tart), total soluble solids, total acid, and juice content. External grade standards include skin color, texture, discoloration and blemishes, fruit form, and firmness.

Harvesting is accomplished primarily by hand, and the fruit is placed in a picking bag or other container. A considerable effort has been expended on developing mechanical harvesting methods, especially with fruit that is destined for processing. In the 1999–2000 season, an estimated 6500 acres of oranges were mechanically harvested in Florida. The harvesting systems varied from those which could shake fruit to the ground from the oldest and tallest trees ($\sim$10 trees per hour) to those which could be continuously shaken with fruit caught on smaller uniformly spaced and skirted trees ($\sim$500 trees per hour). In most of the heavy citrus-producing areas, oranges harvested for fresh market are placed in pallet boxes (a standard pallet box in Florida holds 408.2 kg or 900 lb of oranges) for transport to the packinghouses. Otherwise, fruit may be loaded on to semitrailer trucks for bulk transport to the processing plants.

Most fresh oranges are marketed in the USA in fiberboard boxes holding 14.5–19.0 kg (32.0–41.9 lb) net weight of fruit. Preparation of oranges for market may include degreening, washing, applying fungicides, grading, waxing, and sizing. These procedures are usually performed in a citrus packinghouse near the groves. Figure 1 is a common process flowchart for a citrus packinghouse.

Some early-season oranges such as Hamlins and Washington navels are mature internally but may still be greenish or patchy green externally. Consumers generally believe greenish fruit to be immature, thus it is desirable to degreen the fruit. Degreening may be accomplished through use of either ethylene, or color enhancement utilizing a dye or a combination of the two treatments. In Florida, degreening with ethylene gas, a natural plant hormone, is generally at

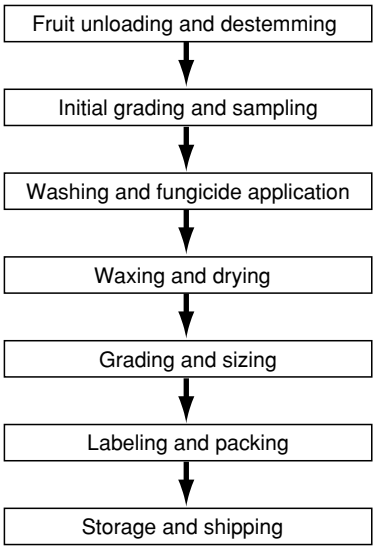

Figure 1 Common process flowchart for a citrus packinghouse.

1–5 p.p.m. in a room maintained for 36 h or less between 29 and 30 °C, 90–96% relative humidity, and having at least one change of air per hour. Use of much longer holding periods such as 72 or 96 h is not uncommon at times. Degreening procedures vary with citrus growing area, cultivar, fruit condition, etc. Color enhancement of the peel may be used in certain cases. In Florida, for example, for degreened (or naturally pale) oranges, temples and tangelos having a final yellow appearance, these fruit, under government supervision, may be flooded in a tank containing Citrus Red no. 2 for orange color enhancement, with 2 p.p.m. residue being allowed following rinsing. California and many other countries do not allow coloring of citrus.

Fruit may be treated with fungicides to control postharvest decay. Important postharvest fungal diseases are stem-end rot caused by *Diplodia natalensis* or *Phomopsis citri*, green mold caused by *Penicillium digitatum*, and sour rot caused by *Geotrichum candidum*. Fungicides, each with its own mode of application, used to control decay include sodium *ortho*-phenlyphenate (SOPP), thiabendazole (TBZ), and imazalil. Oranges to be exported from the USA to other citrus-producing areas are either fumigated with methyl bromide or subjected to cold treatment at 1.1 °C for 17 days to rid the fruit of potentially harmful fruit fly larvae and eggs.

In the washing operation, natural waxes are lost and rind permeability is increased, allowing moisture loss, and a decrease in sugar content, flavor, texture, quality, and ascorbic acid. Waxing of fruit tends to minimize the negative side-effects of washing plus enhancing the appearance of the rind and, in some

cases, provides a vehicle for antimicrobial agents. There are numerous formulations of waxes and wax-like materials available for application on oranges, but the most commonly used are water waxes, either emulsion-based or as a resin solution. During storage, waxes on fruit may cause the reduction of fruit internal oxygen levels, and increase internal carbon dioxide and ethanol levels. Recent studies show that the shellac-based waxes which give high surface shine are not the optimal choice for extended cold storage of oranges due to their relatively poor control of water loss and their low gas permeability that lead to off-flavor development within 5–10 weeks of storage. Polyethylene- and carnauba-based waxes, on the other hand, may provide better weight loss control, allow sufficient gas exchange, and maintain good internal quality under the extended storage conditions.

Storing citrus in the refrigerator keeps it better-tasting, better-looking, and firmer than citrus stored at room temperature. Cold storage of oranges at 0–1 °C for Florida fruit, and 3–9 °C for California fruit, will generally provide several weeks of marketability; otherwise, fruit held at room temperature has a significantly shorter shelf-life. Controlled-atmosphere storage, using increased carbon dioxide and decreased oxygen at cool temperature, does not work for extending the shelf-life of citrus fruit.

Fresh and Minimally Processed Citrus Products

'Fresh' or 'fresh-squeezed' citrus juice is pure juice obtained from mature citrus fruit and has not been further pasteurized, frozen, or concentrated after extraction. It is valued by many consumers for its superior natural taste and fresh quality. Although less processing is involved within its production, it is generally marketed at higher prices than pasteurized juice. Commercially, fresh orange juice is mechanically extracted from graded and washed fruit, and is commonly distributed for the retail market once packaged and refrigerated. Both chemical and thermal antimicrobial treatments may be applied to fruit surfaces prior to juice extraction to achieve a desired decontamination. In the USA, federal regulations also mandate the application of Hazard Analysis and Critical Control Point (HACCP) principles for the safe and sanitary processing of juices. Mechanical juice extractors that are designed to separate the juice from the fruit peel, seeds, and large pieces of pulp simultaneously are commonly utilized in the fresh juice industry. Small-scale extractors squeeze one fruit at a time and are often favored by local retailers. They are portable, simple to clean, and specially designed to attract the attention of shoppers in store or roadside operations. At juice-processing plants, largescale extractors are used. These machines have the capacity to extract hundreds of fruit per minute. Many of these extractors are adjustable with different types of cutters and strainer tubes to accommodate fruit of various sizes, and allow desired amounts of oil and pulp to pass into the juice. Good mechanical extraction of oranges yields about 45–55% juice by weight. Juice-blending programs are used by some processors following extraction. The program often extends the season by mixing juice from two or more varieties of available fruit to enhance nonblended juice color and/or taste. After processing, juice is packed and marketed at 0–4 °C. Freshly squeezed orange juice packed in plastic containers and held at −1 to 4 °C loses somewhat less than 1% ascorbic acid per day. The loss of ascorbic acid in the juice of whole oranges during the normal marketing life is expected to be less than 10%. (*See* **Hazard Analysis Critical Control Point**.)

'Fresh-cut produce' is an established food category that incorporates old and new processing techniques to prepare fresh, convenient, ready-to-eat products. Although the laborious task of removing citrus peel for personal consumption has been reduced by peelers such as modified knives and scoops, a growing segment of consumers desire no preparation. Thus, preparing citrus fruit to ready-to-eat forms would encourage fruit consumption at more dining occasions and by greater numbers of consumers. Both hand and mechanical peeling methods have been used in commercial preparation of citrus products for retail and food service outlets. The traditional peeling processes are often facilitated by steam treatment to loosen the peel and/or chemical digestion to remove peel remnants. High-temperature treatments and caustic chemicals damage the fruit and render it unsuitable for the fresh market. The resulting product is one that is chemically preserved, canned, or frozen before consumption. Citrus fruit may also be mechanically peeled with knives without steam or chemical treatment, but wastage is typically high and juice vesicle damage is often excessive. Recent efforts have been focused on minimizing damage due to peeling to maintain the integrity of fruit during storage. It has been found that the removal of citrus peel may be facilitated by vacuum or pressure infusion of an enzyme mixture (consisting of pectinase and/or pectinase and cellulase) or water into the peel. The fluids digest or solubilize the pectin-rich albedo tissue of the peel, and thus, the peel is loosened from the flesh. The peel may then be manually or mechanically removed, and the flesh may be packaged intact or cut into sections for marketing and consumption.

See also: **Antioxidants**: Natural Antioxidants; **Ascorbic Acid**: Properties and Determination; Physiology; **Chilled Storage**: Principles; **Dietary Fiber**: Properties and Sources; **Enzymes**: Functions and Characteristics; **Folic Acid**: Properties and Determination; Physiology; **Potassium**: Physiology; **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage; Yeasts in Spoilage; **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability; Parameters Affecting Storage Stability

Further Reading

Fellers PJ (1985) Citrus: sensory quality as related to rootstock, cultivar, maturity and season. In: Pattee H (ed.) *Evaluation of Quality of Fruits and Vegetables*, pp. 83–128.Westport, Connecticut: AVI Publishing.

Fellers PJ (1993) Citrus fruits: oranges. In: Macrae R, Robinson RK and Sadler MJ (eds) *Encyclopaedia of Food Science, Food Technology and Nutrition*, vol. 2. London: Academic Press.

Food and Agriculture Organization of the United Nations (2000) *Citrus Fruit: Fresh and Processed. Annual Statistics*.

Food and Drug Administration (2001) 21 CFR Part 120: Hazard analysis and critical control point (HACCP); procedures for the safe and sanitary processing and importing of juice; final rule. *Federal Register* 66: 6138–6202.

Kefford JF and Chandler BV (1970) *The Chemical Constituents of Citrus Fruits*. New York: Academic Press.

Nagy S (1980) Vitamin C contents of citrus fruit and their products: a review. *Journal of Agriculture and Food Chemistry* 28: 8–18.

Nagy S and Attaway JA (eds) (1980) *Citrus Nutrition and Quality*. Washington, DC: American Chemical Society.

Nagy S, Shaw PE and Veldhuis MK (eds) (1977) *Citrus Science and Technology*, vols I and II. Westport, Connecticut: AVI Publishing.

Pao S, Petracek PD and Ismail MA (1997) Advances in preparing peeled fresh-cut citrus. *Food Technology International* 39–42.

Pao S, Davis CL and Parish ME (2001) Microscopic observation and processing validation of fruit sanitizing treatments for the enhanced microbiological safety of fresh orange juice. *Journal of Food Protection* 64: 310–314.

Peeples WW, Albrigo LG, Pao S and Petracek PD (1999) Effects of coatings on quality of Florida Valencia oranges stored for summer sale. *Proceedings of the Florida State Horticultural Society* 112: 126–130.

Redd JB, Shaw PE, Hendrix CM Jr and Hendrix DL (eds) *Quality Control Manual for Citrus Processing Plants*, vol III. Auburndale, Florida: Agscience.

Tetra Pak Processing Systems AB (1998) *The Orange Book*. Lund, Sweden: Tetra Pak.

Wardowski WF, Nagy S and Grierson W (eds) (1986) *Fresh Citrus Fruits*. Westport, Connecticut: AVI Publishing.

Processed and Derived Products of Oranges

C M Lanza, University of Catania, Catania, Italy

It seems that the sky had scattered gold dust and the earth had moulded this into shining spheres. (Abù l-Hasan Ahmad – Arabian poet in Sicily)

Background

Wine and the olive in the west, and the mango and litchi in the east, have played a determining role in the culture in which they have flourished, whereas citrus fruits, with their high quality and high content of vitamin C, have become well established across all continents, societies, and cultures, from man's first interests in agriculture through to the most sophisticated technology to date. At the beginning of the last century, countries dedicated to citrus fruit cultivation processed one or two products per species, discarding the rest (for example, lemon growing in Sicily was mainly used for essential oil and citric acid production). Today, such countries must concentrate their efforts on an industrial system providing the best utilization of the fruit to minimize costs and to offer a wide range of products on the basis of the consumer's economic status. High-value products are aimed at wealthy markets, whereas a large proportion of the less expensive byproducts can be added to the diets of consumers in economically weak countries. This new perspective is not really a technological problem, because equipment for the citrus industry is being continually developed. Rather, the problem concerns the political decisions to organize the citrus agricultural sector for production for industrial utilization. This is, perhaps, more difficult for those countries that have a citrus fruit-growing history. Traditionally, there has often been a lack of flexibility and holistic management in dealing with production problems.

Sweet orange constitutes by far the most important class of commercial citrus fruits grown in tropical and subtropical regions around the world. Approximately 55 million tonnes of orange are produced per year worldwide. About 40% of the total tonnage is processed, the rest being consumed as fresh fruit. Their quality, and the quality of the extracted juice, is greatly influenced by grove practices and pedoclimatic conditions (**Table 1**).

As a whole fruit, the orange is widely recognized for its nutritional (**Figure 1**) and sensorial characteris-

Table 1 Quality parameters of orange juices

°Brix	Min 10.0	Potassium	1300–2500 mg l^{-1}
Density	Min 1.040	Magnesium	70–160 mg l^{-1}
L-Ascorbic acid	Min 200 mg l^{-1}	Titratable acid	90–240 meq
Volatile acids	Max 0.4 g l^{-1}	Citric acid	6.3–17.0 g l^{-1}
Volatile oils	Max 0.3 mg l^{-1}	L-malic acid	0.8–3.0 g l^{-1}
Ethanol	Max 0.3 g l^{-1}	Calcium	60–150 mg l^{-1}
Lactic acid	Max 0.5 g l^{-1}	Phosphorus	115–210 mg l^{-1}
D-Malic acid	Absent	Nitrates	Max 10 mg l^{-1}
Sulfurous acid	Absent	Sulfates	Max 150 mg l^{-1}
Hydroxy methyl furfurol	Max 20 mg l^{-1}	Formol index	15–26
Arsenic	Max 0.1 mg l^{-1}	Esperidin	Max 1000 mg l^{-1}
Lead	Max 0.2 mg l^{-1}	Total pectin	Max 700 mg l^{-1}
Copper	5.0 mg l^{-1}	Soluble pectin	Max 500 mg l^{-1}
Zinc	5.0 mg l^{-1}	Carotenoids	Max 15 mg l^{-1}
Iron	5.0 mg l^{-1}	Glucose	20–50 g l^{-1}
Tin	1.0 mg l^{-1}	Fructose	20–50 g l^{-1}
Mercury	0.01 mg l^{-1}	Saccharose	10–50 g l^{-1}
Cadmium	0.02 mg l^{-1}	Ethanolamine	Max 36.6 mg l^{-1}
Sodium	Max 30 mg l^{-1}	Ammonium	Max 25.5 mg l^{-1}

Source: Association of the Industry of Juices and Nectars from Fruits and Vegetables of the EEC (24/1/1996).

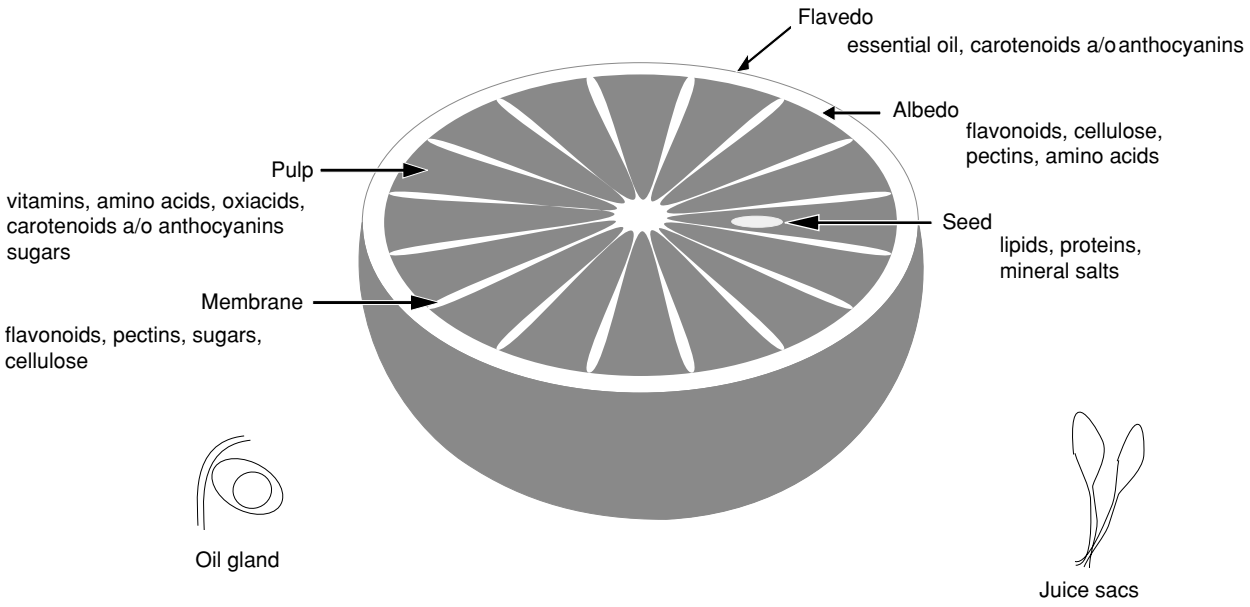

Figure 1 Transverse section of orange fruit with constituents and enlarged views of juice sacs and oil gland.

tics of freshness, juiciness, crispness and sugar/acid content.

To extend the availability of orange fruit during the year and utilize better processing equipment, varieties with different ripening periods are grown. Together, Brazil and the USA grow 50% of the world's oranges and produce around 90% of the world's orange juice. Some Mediterranean regions, such as Italy, are oriented to the fresh fruit market, because of more profit. Italian orange production, concentrated primarily in the south of Italy, comprises 70% of pigmented orange ('Tarocco,' 'Sanguinello,' and 'Moro'), and the remainder blonde ('Navel' and 'Valencia late').

Blonde oranges contain relatively large amounts of a mixture of carotenoids, whereas the red color of blood oranges is due to (besides carotenoids) anthocyanins, water-soluble pigments such as cyanidin-3-glucoside and delphinidin-3-glucoside in Moro oranges. Many of the biological properties of these compounds have been correlated with their ability to scavenge oxygen-generated free radicals and to

inhibit lipid peroxidation. With regard to therapeutic properties, anthocyanins are largely used in the pharmaceutical industry because of their ability to prevent capillary fragility. The most important objective at the moment in Italy is the qualitative improvement of pigmented oranges and their juice, which has created a niche market for export in Northern Europe (Denmark, Germany, Holland, Sweden, Switzerland) (Figure 2).

Juice

Orange juice, the predominant product of the citrus industry, has the sensory and nutritional properties of the fresh fruit (Table 2), but, being a complex beverage sensitive to treatments, it is very difficult to preserve the same level of quality of orange juice squeezed directly from fresh oranges. The maximum juice yield from an orange is 40–60% by weight, depending on the fruit variety and local climate.

Even if orange fruit is not climacteric, its quality begins to deteriorate after picking, and so it must be processed into juice and other products as quickly as possible to insure that the microbiological, nutritional and sensorial qualities are maintained (Figure 3).

Oranges unloaded by hydraulic lifts from storage silos on the waiting conveyors undergo a first washing to remove dirt, leaves, and stems, and then, in the first grading process, damaged or substandard fruit is removed. The oranges are then conveyed to the final washing and grading area, where they are inspected and automatically sorted by size for entry to the extractor.

Two extraction systems are used: (1) squeezing and (2) reaming, either whole or halved, under mechanical pressure. In the first system as shown in Figure 4 the fruit are often lined up (up to 15 per line). The cups (upper and lower) are brought together as the corresponding cutters cut two holes in the fruit (Step 1). As the cups continue to come together, the peel is separated from the fruit, which is forced into the prefinisher tube (Step 2). The peeled fruit moves into the orifice tube through the holes of the prefinisher tube, forcing the juice to flow into the manifold (Step 3). The juice is instantaneously separated from the seeds, and the rest of the fruit, including the oil emulsion, is sent to the centrifugal processing part.

In the reamer type of extractor, the fruit is cut in half, and the halves are penetrated and rotated by a series of nylon reamers mounted on a rotating turntable. The juice, pulp, rag, and seeds are separated from the peel by a strainer, and the juice and pulp are separated from the rag and seeds and passed to the

Table 2 Chemical composition and energy of orange fruit and orange juice

	Orange fruit	*Orange juice*
Edible part (%)	80	100
Water (g)	87.2	89.3
Protein (g)	0.7	0.5
Lipid (g)	0.2	0
Carbohydrate	7.8	8.2
Fiber (g)	1.6	
kcal	34	33
kJ	142	138
Sodium (mg)	3	3
Potassium (mg)	200	200
Iron (mg)	0.2	0.2
Calcium (mg)	49	15
Phosphorus	22	17
Thiamin (mg)	0.06	0.05
Riboflavin (mg)	0.05	0.03
Niacin (mg)	0.2	0.4
Vitamin A retinol equivalents (μg)	71	38
Vitamin C (mg)	50	44

Source: National Institute of Nutrition, Rome.

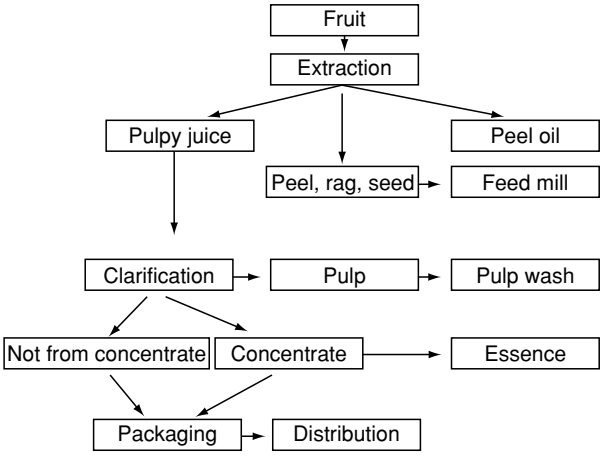

Figure 3 Flow chart of orange processing.

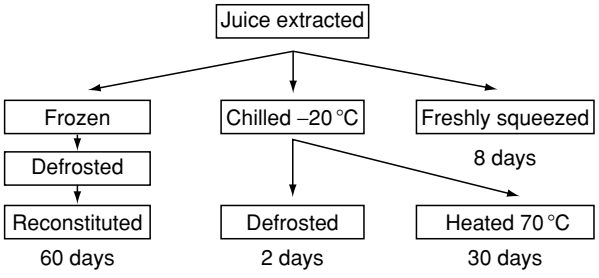

Figure 2 Orange juices in the European market. Shelf-lives shown (in days) are applicable to a storage temperature of 4 °C. Source: Maccarone E, Campisi S, Cataldi Lupo MC, Fallico B and Nicolosi Asmundo C (1996) Effetti dei trattamenti tecnologici sui costituenti del succo di arancia rossa di Sicilia. *Industrie delle Bevande* XXV: 335–341.

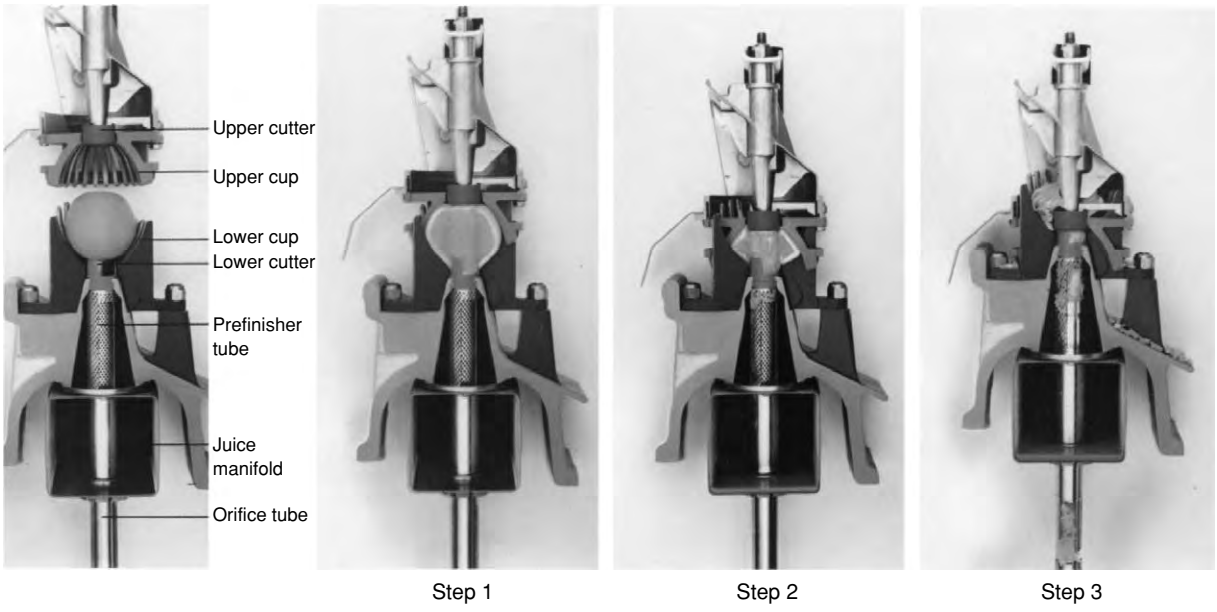

Figure 4 (see color plate 29) FMC juice extractor principle. Courtesy of FMC Food Tech Citrus System, Catania (Sicily) Italy.

finishers. In this system, peel oil can be recovered in a separate step by puncturing the oil sacs of the flavedo and washing the oil out.

Another extraction system is the rotary press, popular in the Mediterranean area since the turn of the twentieth century. This machine consists of a combination of two units arranged in a cascade, one to extract essential oil by rasping the outer surface of the whole fruits and the other to extract juice. Fruits introduced into the machine are pushed forward by paddles along couples of rotating rasping cylinders. The essential oil is then gathered by sprays of water, the mixture flows into a finisher to eliminate any solid particles, and essential oil is recovered by centrifugal separators. After oil extraction, the fruits are cut into two halves and squeezed against a stainless steel screen, and then seeds, pulp, and peels are removed from the juice. Juice obtained by the different extraction processes has a high content of pulp derived from juice vesicles and membrane material. These represent the principal part of the segments, which, during processing, are destroyed by enzymes and microorganisms in the peel. *Lactobacillus* and *Leuconostoc* are the most common genera of lactic acid bacteria that cause spoilage of orange juice, producing diacetyl and, consequently, off-flavors. The yeasts *Saccharomyces cerevisiae*, *Rhodotorula*, and *Zygosaccharomyces* can impart a fermented taste, transforming the delicate aroma of orange juice into an unpleasant smell. Only rarely are molds responsible for juice spoilage, because, even if they grow in juice, they are generally sensitive to heat treatment. (*See* **Yeasts**.)

The pulpy juice therefore must be finished and clarified by centrifugation to remove the oil, if necessary, before undergoing further treatments. A recent improvement of clarification has been introduced, using ultrafiltration to separate the serum from the pulp. The latter is then subjected to thermal treatment before being remixed with the serum to obtain a product with a better flavor and nutritional value. After this step, the juice takes different paths, leading to different products:

- Freshly squeezed juice: This has a shelf-life of 5–6-day and is a product that is not thermally processed, thus satisfying increasing demands for natural products. The raw materials are selected with great care, and washed, squeezed, aseptically packaged, and distributed for retail.
- Not-from-concentrate juice: Minimal thermal treatment is used for this juice. It is processed and stored frozen, or quickly pasteurized and stored in aseptic form. Most of these juices also consist of a blend, differing in quality in terms of color, sugar content, etc., being extracted throughout the picking season and/or in different producing countries.
- Frozen concentrate orange juice: This form is very advantageous for storage and freight, requiring less than five to six times volume than fresh juice. It provides a year-round supply to consumers and consists of a concentrate of a blend of specified quality from different orange varieties. Analytical control of blends uses the following characteristics: °Brix, per cent acid, Brix/acid ratio, per cent oil, per

Figure 5 (see color plate 31) Model thermally accelerated short-time evaporator. Courtesy of FMC Food Tech Citrus System, Catania (Sicily) Italy.

cent pulp and color. Today, a 'thermally accelerated short-time evaporator' is often used (Figure 5), which yields a concentration of up to 65/75 °Brix in a very short time, thus insuring a high product quality. (*See* **Heat Treatment:** Chemical and Microbiological Changes.)

During a preheating cycle, juice is pasteurized to deactive pectinase enzymes (which cause cloudiness) and prevent spoilage, thus prolonging the shelf-life. It then undergoes a variable number of stages under vacuum until the required concentration is obtained. The disadvantage is that the heat treatment reduces the oxidative balance and removes some of the flavor components. Particularly in blood-orange juice, a considerable change in color is ascribed to the degradation of anthocyanins and the different distribution of carotenoids between the serum and pulp structure in juices reconstituted from concentrate. There is also a remarkable loss of aroma components and generation of off-flavors such as *p*-vinylguaiacol, *p*-vinylphenol, and decarboxylation products of hydroxycinnamic acids in red juices processed and stored for up to 4 months.

Byproducts

The use of whole fruit in citrus processing is necessary to reduce costs and justify new investments in equipment, providing a correct solution for the pollution problem connected with the process. Any improvement in the recovery process would be of great interest, especially if this were obtained using technologies respectful of product quality such as the use of enzymes, peeling of the fruit, reduction of bitterness juices, reduction in the viscosity of pulp wash, and hydrolysis of peel solids to produce fermentable sugars.

Comminuted juice

Originally developed in the UK, this byproduct is made by milling the whole fresh fruit or by mixing juice concentrate with milled peel. The comminuted citrus base has a stronger flavor and provides more cloud than pure orange juice, thus imparting a good orange flavor in fruit drinks of low fruit content.

Peel Oil

Valuable oil recovered (about 0.3% of the fruit intake) from peel is used to produce various flavor compounds in the beverage, chemical and cosmetic industries. 'Cold-pressed peel oil' is obtained by mechanical pressure of oil glands (using different methods), whereas a less valuable oil can be recovered by distillation of waste peel residue. Among the numerous compounds, the major component is D-limonene, which is also used in the electronics industry and in the manufacture of synthetic resins and adhesives. Other hydrocarbons produced include myrcene, α-pinene, sabinene, and valencene. Important aromatic compounds include aldehydes, esters, alcohols, ketones, and phenols, and the nonvolatile residues include waxes, coumarins, flavonoids, fatty acids, and sterols. To avoid the problem of dissolved wax from peel, the oil is stored at 1 °C (winterization) to allow successive sedimentation of wax.

Colored essential oils rich in carotenoids can be obtained from orange peels, for use as colorants in the drink industry. The increased use of citrus essential oil to aromatize baking products promotes the technology of encapsulation by drying spray obtaining powders capable of maintaining the aroma intensity during cooking. Another interesting technology is the extraction of essential oil using supercritical fluid,

especially CO_2. Among the citrus peels, the most important essential oil is extracted from bergamot (*C. bergamia*, Risso) hybrid, a close relative of the sour orange cultivated only in the narrow strip of land of Southern Italy (Calabria). It is indispensable to the international perfume industry and is also used in the pharmaceutical industry as an antiseptic and as an aroma for liqueurs, tea, sweets, and candied fruits.

Essence

The recovery of volatile components (essence) from the evaporation process constitutes an interesting use of this component. Two phases are separated: the first containing the water-soluble compounds (essence aroma) and the second (essence oil), containing fruit flavor. The latter is more desirable than peel derivatives in the manufacture of flavor mixtures for the beverage and other food industries.

Hesperidin

In sweet orange, the predominant flavonoid is the tasteless and odorless hesperidin, located in the membrane and peel, which is not soluble in neutral solutions but slightly soluble in acidic solution as citrus juice. After extraction of the juice, the hesperidin forms crystals that appear as white flakes; these are more evident during evaporation and concentration of juice. The recovery of this product is a profitable investment for its therapeutic properties in the treatment of capillary fragility. The traditional extraction method, with a very low recovery, treats triturated peel with alkaline compounds to dissolve the flavonoids, which, after separation of solid residue, are precipitated by acidification. An innovative method to recover hesperidin from orange peel and waste water utilizes (at a pH of around 7) styrene-divinylbenzene as the absorbing phase and removing the flavonoid with alcoholic NaOH. The results may be a useful starting point for industrial applications.

Pectin

Pectin is a product found mainly in the albedo and juice sac, and has a high molecular weight (100 000–200 000 Da). Very important for the stability of juice cloud, pectin consists of long chains of polygalacturonic acid units linked together by α (1–4) glycosidic bonds with side chains of rhamnose, arabinans, galactans, and xylose. Many of the carboxyl groups are esterified with methanol to form methoxy groups; these determine the gelling ability of the pectin used in jam, marmalade, jelly, and preserve production. Lemon peel is the most common source for extraction of commercial pectin. After oil extraction, the peel is shredded and washed with water to remove sugars and other soluble material. The peel is heated to deactivate the pectinase enzymes and then treated with hot mineral acid to extract pectin. To the concentrated solution (containing a low content of pectin) alcohol is then added to 50–70% by weight, resulting in a precipitate of pectin in a gelatinous mass, which can be washed and dried as a powder. (*See* **Pectin**: Food Use.)

Peel in Brine and Candies

The peel in this case must be processed by a soft system, as 'sfumatura,' or peeling to extract the oil. First, the peel is hardened using lime, dried, and dipped for 24 h in tanks containing 3% sodium chloride. Then, the peel is dripped, stored in containers, and transferred to candy factories. Immersion in hot water for a few minutes softens the peel. The peel is cut by machine to undergo a cold or warm liquid sweetening process, depending on the technology chosen. The former process requires more time, because the diffusion of sugar into peel is slow, but the sensory characteristics of these candies are more valuable than those obtained by the latter process.

Fruit Sections

Very popular in the USA and Israel, fruit sections have good nutritional characteristics and a long shelf-life. Prior to processing, orange should be held in storage bins to allow the fruit to soften. Hot-water treatment at 100 °C for a few minutes facilitates peeling by hand or machine. The fruit is then subjected to a hot caustic spray (about 2.0% NaOH) for a few seconds and water sprays to remove the basic solutions. The peeled fruit is then chilled in water to facilitate sectioning of the fruit. A sweetener solution containing water, juice, and sugar is finally added to the fruit sections.

Pulp

Pulp recovered after the extraction process usually can be added back to juice and juice drinks to improve the mouth feel and give a natural appearance to the product. After the different components are separated by an in-line extractor, it is possible to recover whole juice sacs (premium pulp) by the premium pulp recovery system (Figure 6) (FMC Food Tech Citrus System).

Specially designed extractor components separate the juice, pulp, core, and seeds into a single stream. The core material is gently separated from the seeds, juice, and pulp and sent to the core wash system or

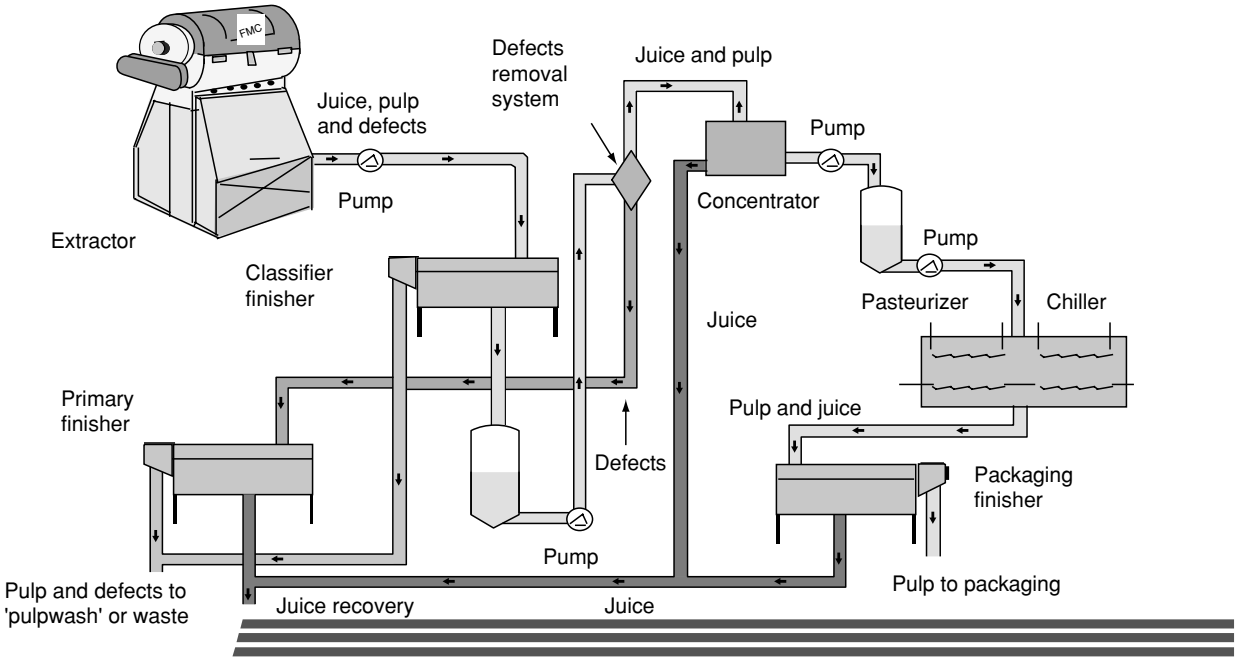

Figure 6 (see color plate 30) Premium pulp recovery process. Courtesy of FMC Food Tech Citrus System, Catania (Sicily) Italy.

waste. The core stream is finished to separate and recover any remaining juice. The remaining juice, seeds, and pulp are sent to a classifying finisher, which separates the seeds from the pulpy juice stream. Minor impurities are then removed, and the premium pulp, finally separated from the juice, is ready for pasteurization and packaging.

Pulp Wash

Product referred to as 'water-extracted soluble orange solids' is reclaimed from the pulp stream and is often used in fruit drinks as a source of sugars. It is also added, where the law permits, as a clouding agent to provide body and mouth feel because of its pectin content. The diffusion techniques currently in use for the preparation of pulp wash are optimized using a citrus finisher (Figure 7).

The unfinished juice feeds into the finishing cavity, a space created by a spiral with specially designed flights inside 360° cylindrical screens, resulting in a finishing action. Once inside the finisher cavity, the product is immediately carried outward on to the screw's helical spirals, thus making full use of the entire screen area, which results in high-capacity finishing. As the screw rotates, air restriction at the discharge end places pressure on the product causing the liquid to move toward the lower-pressure area outside the screen. Once through the screens, the liquid flows into a fully enclosed pan and discharges. The pulp discharges through the plug valve assembly.

Animal Feed

Unprocessed citrus peel, rag, pulp, and seed constitute waste products, but they can provide substantial profits if they are processed as feed. As the product is susceptible to spoilage (owing to its 78–90% moisture content), it must be consumed quickly and presents expensive transportation costs. The principal qualitative parameters are the amounts of protein, crude fiber, fat, ash, and nitrogen-free extract, primarily carbohydrates, which varies somewhat. The total quality of the feed is generally estimated from the total digestible nutrients: nitrogen-free extract, protein, crude fiber, and fat. The type of animal, effect of these products on food, economics involved, and storage capabilities must all be considered before using these waste products.

The high acidity of orange peel requires blending with lemon peel, usually by ensilage with straw and/or other material. Research carried out in 1960 showed, in mixtures of orange pulp and straw (10%) ensiled for 2 months, good fermentation and digestibility characteristics without any effluent being released.

Pressed Peel

To reduce transportation costs, the peel is limed, shredded, and pressed to improve the texture and composition. The reaction between lime and peel removes (by pressing) about 10% of the peel moisture. After this

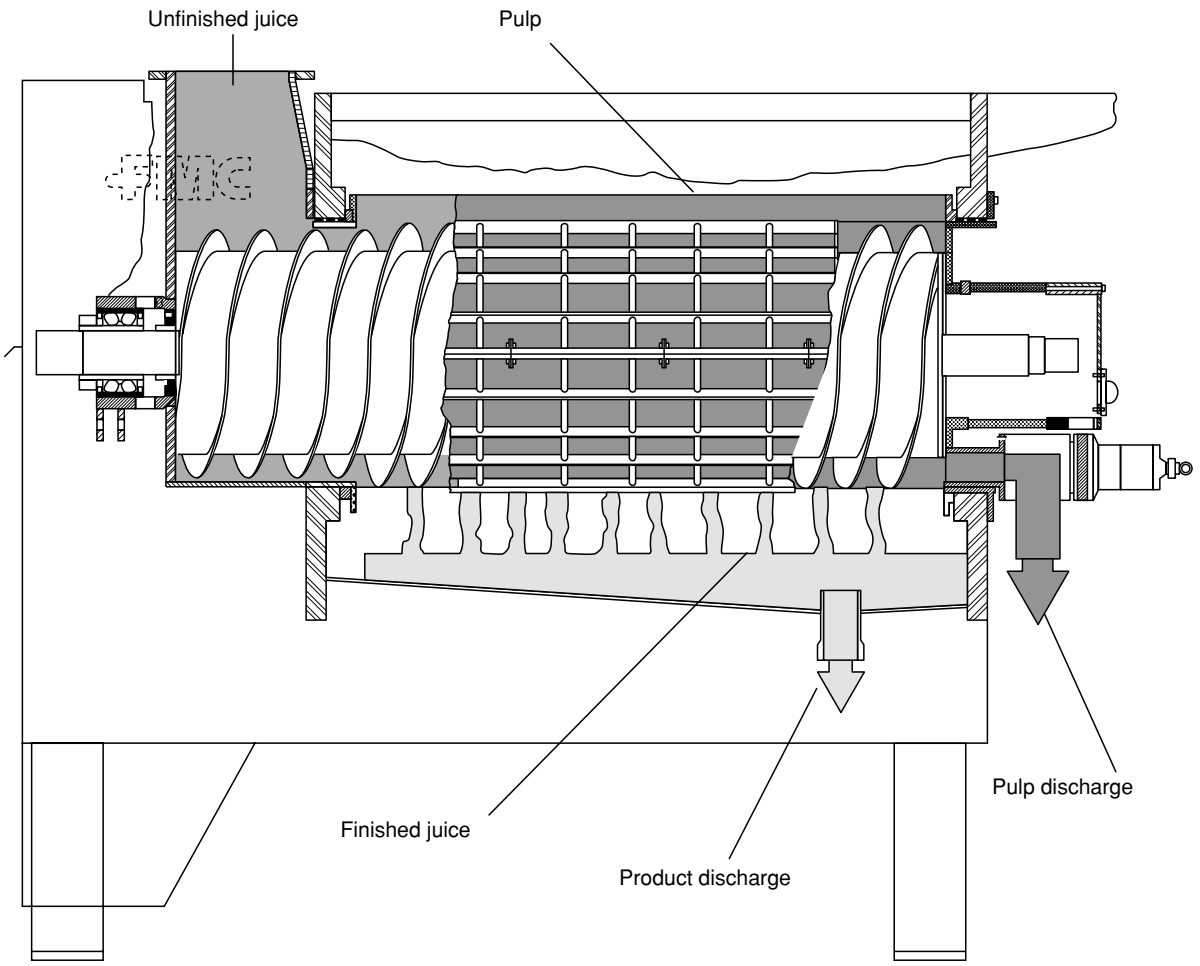

Figure 7 Citrus finisher. Courtesy of FMC Food Tech Citrus System, Catania (Sicily) Italy.

treatment, the peel is shredded and then conveyed to the peel press. Warm molasses is added to the peel (1:3), along with some press liquor, and the slurry is pumped to the feed mill ('pumped peel'). Using a supercharger and a sealed screw air pockets are removed, thus allowing the peel to feed into the press under pressure and to leave the press at a moisture content as low as 60%.

Dried Peel and Pulp

Dried citrus peel is one of the most common feeds. It is manufactured by pressing peel through a rotary dryer and adding citrus molasses to help the drying process and help prevent the peel from burning. The moisture content of dried peel must be below 10%. Many experiments published in the 1970s have shown that dried orange pulp, partially or completely replacing cereals in concentrate mixtures, are particularly useful in reducing feeding costs in dairy cows, have no influence on production, and have a good palatability. Dried pulp has also been used in swine, which have been shown to utilize it at a ratio of up to 20–25%. Besides its use as a substitute for maize, up to 20% in diet has no influence on the growth and production of laying hens. The dried pulp can be pelletized and is consumed more easily by ruminants with advantages of storage, shipping, and microbial spoilage. Pellets made from dried pulp have different dimensions, and several factors affect their characteristics, such as the energy used in pelletizing and the proportions of citrus molasses (about 5–15% of the total weight gives excellent results) used as binding agents.

Citric Molasses

The solution obtained from shredded and limed peel (press liquor) containing 10% sugar is principally concentrated to citric molasses. It is used as an additive to animal feed and for the production of citrus alcohol by fermentation, and also can be mixed with

pressed peel to facilitate drying. In the manufacture of citric molasses, antifoaming agents are used, or the suspended solids are removed by centrifugation, to prevent foaming from microbial spoilage and alkaline processing. The advantages of citric molasses in terms of press liquor are a reduction in cost during storage and a better microbial protection owing to the sugar content.

Alcohol Production

Citrus press liquor and citric molasses can be used as a feedstock for commercial alcoholic fermentations carried out by yeasts in continuous fermentators. The presence of peel oil, which is toxic to most microorganisms, must be at levels of no more than 0.08%. Press liquor contains at least 0.2% oil (based on 11.8 °Brix), whereas citric molasses usually contain much less oil stripped out during evaporation. Liquor is removed from the citrus press before fermentation, often using direct steam injection. Also, the oil removed in this way (steam-distilled oil) can be sold as a fuel, herbicide, or feedstock in chemical manufacture.

Final Consideration

In industrialized countries, the citrus industry cannot be static, but must apply scientific and technological know-how to increase the value added to the market prices of products. It is necessary to develop new mild technology to limit the damage during processing, to re-evaluate minor products, and to obtain a new ingredient better suited to consumer requirements: fresher products that retain their nutritional characteristics, especially the active components (vitamins, antioxidants, fermentation agents), have a longer shelf-life, and are more convenient for ecologically compatible cultivation and animal rearing systems.

See also: **Citrus Fruits**: Types on the Market; Composition and Characterization; Oranges

Further Reading

AIJN (1996) *Code of Practice for Evaluation of Fruit and Vegetable Juices*. Brussels: Association of the Industry of Juices and Nectars of the Economic Community.

Arena E, Fallico B and Maccarone E (2000) Influence of carotenoids and pulps on the color modification of blood orange juice. *Journal of Food Science* 65: 458–460.

Fallico B, Lanza CM, Maccarone E, Nicolosi Asmundo C and Rapisarda P (1996) Role of hydroxycinnamic acids and vinylphenols in the flavour alteration of blood orange juices. *Journal of Agricultural and Food Chemistry* 44: 2654–2657.

Fox K (2000) New technology in citrus processing. *Fruit Processing* 10: 94–101.

Kimball DA (1999) *Citrus Processing: A Complete Guide*, 2nd edn. Gaithersburg, MD: Aspen.

Maccarone E, Campisi S, Cataldi Lupo MC, Fallico B and Nicolosi Asmundo C (1996) Effetti dei trattamenti tecnologici sui costituenti del succo di arancia rossa di Sicilia. *Industrie delle Bevande* XXV: 335–341.

Manthey JA and Grohmann K (1996) Concentrations of hesperidin and other orange peel flavonoids in citrus processing byproducts. *Journal of Agricultural and Food Chemistry* 44: 811–814.

Parish ME (1991) Microbiological concerns in citrus juice processing. *Food Technology* 4: 128–133.

Redd JB and Hendrix CM (1993) Processing of natural citrus oils and flavors. In: Nogy S *et al*. *Fruit Juice Processing Technology*. Auburndale FL: Agscience.

Rouseff RL and Nagy S (1994) Health and nutritional benefits of citrus fruit components. *Food Technology* 48: 125–126, 128–129, 132.

Tetra Pak Processing Systems (1998) *The Orange Book*. Lund: Tetra Pak.

Wang H, Cao G and Prio L (1997) Oxygen radical absorbing capacity of anthocyanins. *Journal of Agricultural and Food Chemistry* 45: 304–309.

Lemons

R Goodrich, University of Florida, Lake Alfred, FL, USA

Background

Lemon (*Citrus limon* (L.) Burm.) a yellow or pale yellow prolate fruit with five to 10 seeds, botanically a berry, is known throughout the world, and is used in numerous foodways and cuisines. The juice of the fruit is commonly used as a food ingredient in both commercial and home recipes, and is valued for its tart, tangy, and fresh character. The exceptionally fragrant rind is used as a garnish and flavoring and is a major source of commercial essential oil and aroma compounds. The lemon flower is sweetly scented, with an aroma similar to other citrus flowers. The essential oils from the flowers, twigs, and fruit can be distilled or otherwise extracted to obtain materials suitable for use in flavors and fragrances.

The lemon tree is a small, thorny tree that grows well throughout the tropical and subtropical regions of the world. Lemons are cultivated in modest amounts for home use almost everywhere that citrus trees can survive, but large-scale commercial production is limited to the subtropical regions since more humid tropical environments enhance pest and

disease problems for both fruit and tree. Lemon trees grow to about 3–6 m and are sparsely foliaged. Compared to orange and grapefruit, the trees are rather cold-sensitive. Because of the lemon's lower sugar content, the fruit can freeze at temperatures that would not affect other citrus.

The lemons has existed for so long that its true origin is not known, although experts believe it was probably a hybrid of the citron (*C. medica*). The first recorded habitat of the lemon was South-east Asia (Myanmar, and southern China); from there, it was introduced to Persia and the Middle East, were it was well established by the twelfth century. Around 1150, Arabs took the lemon to Spain and North Africa. The lemon traveled from the Mediterranean to the New World with Columbus on his second voyage in 1493. Lemon culture spread from Haiti through the West Indies and then to South and North America. Plantings existed in Florida by the 1500s and were established in the mission gardens of California and Baja California by the mid-eighteenth century.

Table 1 Lemon production (tonnes) by selected countries for crop years 1995–96 and 1999–2000

	1995–96	*1999–2000*
Argentina	720	1050
Spain	457	899
USA	900	797
Italy	699	590
Turkey	418	500
Greece	133	155
South Africa	71	110

From USDA, Foreign Agricultural Services www.fas.usda.gov.

Table 2 Leading lemon (*Citrus limon*) varieties

Variety	*Primary growing location(s)*
Bearss (Sicilian)	Florida (USA)
Eureka	California (USA), Australia, Argentina, Israel
Lamas	Turkey (early season)
Lisbon	California, Arizona (USA), Australia
Meyer Lemon (not a true lemon)	USA, New Zealand
Verna (Berna)	Spain

Worldwide Production

Leading producers are Argentina, Spain, California (USA), and Italy. Table 1 summarizes lemon fruit production for the 1995–96 and 1999–2000 seasons. Production from the top four countries accounted for about 75% of the world's total output. Argentina and Spain in particular have experienced significant growth in lemon production in the past few decades; in 1965, the USA and Italy alone accounted for about 75% of the world's lemons. From a global perspective, understanding the system of coordination of production, harvesting, and marketing of lemon fruit is complicated due to factors such as multiple harvesting seasons, the numerous commercial varieties, and the substantial regional differences in practices and markets throughout the world. The following discussion summarizes the commercially important lemon varieties and harvesting periods in the major lemon producing regions.

The principle varieties of lemon are listed in Table 2, along with their associated growing regions. Different varieties of lemons can be difficult to identify by the fruit alone, especially as the everbearing trait results in the presence of fruits of various size, color, and appearance on a single tree at the same time. Most lemon fruits are oval to elliptical in shape, with characteristic necks (stem end) and nipples (blossom end). The peel is yellow at maturity (generally) but can vary substantially in thickness and smoothness. The flesh is pale yellow (straw-colored) and very acidic, and can have a variable number of seeds. As a further complication, the same variety can present visually different fruit under different climate conditions and horticultural practices. Lemons are often identified by tree characteristics, which are less variable.

Most of the US production is from California, although there is some from Arizona. In coastal California, the primary growing region, the lemon tree blooms and sets fruit essentially year round, but there is a heavier set in the spring and fall season. This multiple flowering and fruiting phenomenon can be utilized to provide a steady stream of fruit with pickings occurring every 6–8 weeks throughout the year. In south-central California and Arizona, the other major lemon-producing areas in the USA, harvesting is usually limited to late fall to early winter (October through January).

One of the oldest commercial varieties is the Eureka lemon, and it is also one of the most widely grown varieties outside the Mediterranean area. It is a major part of the crop in California, Australia, South Africa, and Israel. It is also a significant part of Argentina's crop. The Eureka has achieved predominance due to its quality and productivity, and because a large proportion of fruit matures during late spring and summer when demand and prices are greatest. New plantings of the Lisbon lemon have replaced Eureka in central California due to the former's dense foliage, making the fruit less susceptible to

sunburn and possibly freeze damage. The Eureka tree, however, is less thorny than the Lisbon and thus easier to pick. The Eureka tree also appears to be more susceptible to cold damage, making it most suitable for the coastal areas of California and Australia, where freezes seldom occur. The two fruits, however, are very similar, and processors in particular have difficulty in differentiating the two.

The Mesero lemon, of Spanish origin but grown more commonly in Italy, is considered one of the best processing lemons; it has a smooth, thin, somewhat pale rind, but contains a large amount of juice. Conversely, the Interdonato is primarily a fresh cultivar, with a lower juice content and a smooth, thin rind. The fruit itself is large, and the flesh contains few seeds.

Winter is the main harvest for Verna fruit grown in Spain, Algeria, and Morocco. However, in Spain, where Verna is the primary cultivar, a significant supply of fruit is kept available for harvest throughout much of the summer, a peak season for lemon demand. The Femminello, a name for an everbearing group of cultivars grown in the Mediterranean countries, has its main crop reaching maturing in the late winter and early spring but is also marketed throughout the year. Interestingly, in Italy, each particular lemon crop has its own name, and is marketed according to season instead of by the varietal name. Thus, the *Primofiore* is from the September through November harvest, the *Limoni* is from December to May, the *Bianchetti* is from the April to June harvest, and the *Verdelli* is from the June to September harvest. Collectively, this group accounts for about 75% of the lemons grown in Italy. Most (over 90%) of the lemons grown in Italy are produced in the coastal area of Sicily, a warm, dry climate that is highly favorable for lemon production The Interdonoto, another Mediterranean cultivar, is harvested in the fall and early winter. This cultivar is a significant portion of Turkey's lemon crop.

The Genova is grown principally in South America and is the leading variety in Chile. The fruit has a good internal quality (high acid, flavor, and juice yield), similar to Eureka, and a characteristic spherical shape. The origin of Genova is Italy, and it is actually considered a strain of Eureka in California, where it was first imported from Italy in the late nineteenth century.

Compared to other citrus, the projected production growth of lemons is forecast to be only modest, with an estimated rate of 1.5% per annum into 2005. This is about half the growth of the previous decade. The geographic production will also shift, with the strongest production growth forecast for Italy and Argentina, and with possible crop reductions in California.

Fruit Structure and Chemistry

The structure of the lemon fruit is similar to other citrus and consists of the peel and the interior edible portion (endocarp or pulp). The peel is composed of two distinct layers; the outside layer is termed the flavedo (epicarp), and the interior of the peel (the white spongy part) is the albedo (mesocarp). The many chemical constituents of the lemon fruit are distributed among the various tissues, usually in differing concentrations. The flavedo is composed of mainly carotenoid pigments, vitamins, and essential oils. The green appearance of immature lemon flavedo is due to the predominance of chlorophyl over the carotenoid pigments prior to complete ripening. The albedo contains celluloses and soluble carbohydrates, pectin and protopectin, flavonoids, amino acids, and vitamins. The oil glands, containing the valuable essential oil, are embedded primarily in the flavedo, although larger oil glands may also extend through to the upper part of the albedo.

The pulp of the lemon fruit, like other citrus, is divided into segments called carpels. Each segment is composed of up to several hundred units called juice sacs or vesicles. These vesicles are the compartments, which contain the actual lemon juice. The walls of the vesicles are composed of celluloses, hemicelluloses, pectin, protopectin, sugars, flavonoids, and other minor components such as amino acids and vitamin C. When intact, the vesicles effectively contain the juice and limit microbial and enzymatic degradation of the juice.

Comparisons of California fruit (navel and Valencia oranges, and lemons) show that, in relation to the oranges, lemons have a somewhat lower protein content, a substantially lower sugar content, and a much higher acid content. These biochemical measurements of the fruit, confirmed by obvious taste characteristics, lead to the categorization of citrus by horticulturalists as either low-acid (oranges, mandarins and grapefruit) or high-acid (lemons and limes) fruit. Lemon juice contains about 4.3–4.8 g citric acid per 100 ml, as compared to around 0.8–1.0 g per 100 ml for mandarin, orange, or grapefruit juices.

Fresh lemon juice has approximately the following vitamin levels per 100 g: β-carotene, 12 μg; thiamin, 0.3 mg; riboflavin, 0.01 mg; vitamin B_6, 0.05 mg; folate, 13 μg; and vitamin C, 36 mg. These figures vary depending on the source of the fruit and processing and storage conditions.

The peel and seeds of all citrus are relatively rich in phenolic compounds such as flavonoids and phenolic acids. Phenolic compounds are known, potent antioxidants. Citrus flavonoids include polymethoxylated flavones and glycosylated flavanones; both are

Table 3 Volatile compounds identified in lemon oils and essences

Acetone	Myrcene
3-Buten-2-ol	Neral
Ethanol	Octanal
Geranial	Octanol
Geranyl acetate	β-Pinene
Limonene	Terpinene-4-ol
Linalool	α-Terpineol
Methanol	

unique to citrus and are thought to have bioactive properties such as anticancer, antiinflammatory, and antiviral effects, among others. Methanol extracts of lemon seeds are rich in eriocitrin and hesperidin, while peel extracts contain significant amounts of neoeriocitrin, naringin, and neohesperidin. There are many ongoing research programs investigating the functional properties of these flavonoids, and other bioactive compounds in all citrus, including lemon.

All fruits and fruit products contain a range of volatile, low-molecular-weight (< 250) compounds that are important in producing the characteristic flavor and aroma of the specific fruit. The predominant compounds responsible for lemon aroma number are relatively few (two to five) compared to other citrus. The most important component is citral, an aldehyde that is a *cis–trans* mixture of the isomers neral and geranial. Table 3 summarizes some volatile compounds found in commercial lemon oils and essences.

Postharvest Handling and Fruit for the Fresh Market

In the past, the USA, Italy, and Spain dominated total lemon production, and, as noted in Table 1, Argentina has become a major producer. For fresh fruit markets, Spain has the major and profitable role in supplying fresh lemon fruit to Western Europe and other markets; the USA, Argentina, and Italy continue to process about half of their fruit and market the other half as fresh fruit.

Two terms must be understood in order to discuss citrus fruit quality, markets, and economics: grade and maturity. The grade refers to external characteristics such as size, presence of blemishes, and appearance, while maturity refers to mostly internal characteristics such as juice content, taste, total soluble solids (TSS), acid content (titratable acidity; TA) and the ratio of TSS to TA. As with most produce destined for the fresh market, appearance is the primary attribute of quality for fresh market lemons, and grading is critical and standardized in a given country or region. Consumers may make repurchasing decisions based upon internal quality, but generally will not choose a fruit initially if it presents exterior faults. Areas with hot, dry summers and cool, wet winters (Mediterranean climate) typically produce citrus fruit with brighter, thicker peel than areas with more humid growing seasons and warmer winter nights. The Mediterranean climate is also associated with generally better exterior fruit appearance (at the same level of inputs such as fertilizer and pesticides) than fruit grown in a humid climate and is yet another reason for the predominance of fresh market lemon production in the Mediterranean-type climate. Quality indices include color intensity and uniformity; size (larger lemons are preferred), visually pleasing shape and smoothness; freedom from obvious decay and damage; and freedom from exterior blemishes, splotches, and rind discoloration. Fresh-fruit pricing in most lemon-producing countries is based on these factors, along with market supply and demand.

Lemons for both the fresh and processed markets are harvested manually. Pickers use bags, ladders, clippers, and gauges to judge the size of the fruit. Fruit picked early in each variety's particular growing season may be quite green, even though they may be deemed physiologically mature. Lemons picked while still green will generally have a longer postharvest shelf-life than the fully yellow lemons that need to be handled gently and marketed more rapidly. Young lemons can be degreened at the packing house, if necessary. Degreening is accomplished by exposing the lemons to 1–10 p.p.m. ethylene for 1–3 days at relatively high temperatures (20–25 °C). Ethylene is a natural product of plant metabolism and acts as a plant hormone by regulating many aspects of growth, development, and senescence of the fruit.

At the packing house, the lemons are unloaded, washed, and graded, degreened as described above, if necessary, sized, waxed, and packed in cartons or other shipping containers. Each packing house may be set up slightly differently, but the major unit operations in packing fresh lemons are the same as for other citrus fruit.

Citrus fruits, lemons included, are nonclimacteric, and their respiration and ethylene production rates are low. Postharvest changes are therefore low in relation to commodities that are climacteric, such as avocados, bananas, and berries. Most lemon cultivars can be, and are, stored for a moderate amount of time; this requires no special controlled atmosphere (as with apples) if the environmental conditions of humidity and temperature are carefully controlled, and the fruit has been properly handled. The

optimum temperature range for lemon storage is 12–14 °C, depending on cultivar, maturity stage at harvest, and intended distribution and transport time. The relative optimum humidity for storage is 90–95%.

An interesting and specific example of long-term lemon storage is the traditional practice of harvesting lemons (primarily in the coastal California area) in mid- to late winter, and storing them until the summer marketing season. These lemons, when picked, are green, thick-skinned, and not very juicy but after storage (termed curing) are yellow, juicy, and thin-skinned. The thinning of the peel is due to water transpiration loss, while the increase in juice yield is probably due to physiological changes in which the water in the juice vesicles is released into the vesicles from its gel-bound form at the membrane.

As noted, lemon fruit for fresh markets must be attractive and wholesome in appearance; physiological disorders, defined as the breakdown of fruit tissue not caused by microbes or mechanical damage, must be prevented or minimized. Adverse preharvest or postharvest conditions, or nutritional deficiencies during growth and development, can cause physiological disorders in lemons. Fresh citrus fruit are more prone than pome fruits (apples, pears) to a particular type of physiological disorder, termed chilling disorder. In lemons, storage at suboptimal temperatures can lead to symptoms such as pitting of the flavedo, membrane browning, and red blotches. Lemons should not be stored at temperatures below 10 °C.

Major Processed Products and Byproducts

Lemon juice is a commonly used ingredient in both commercial and home kitchens, as well as a component of many processed food and beverage formulas. The following discussion focuses on the commercial production of lemon juice and concentrate.

After the fruit has been delivered to the processing plant, it is unloaded, washed, and graded. The juice is extracted in the same manner as other citrus fruit; Brown and FMC extractors are the most commonly used in the western hemisphere, although there are other citrus juice extractors available. Either prior to (Brown system) or during (FMC system) extraction, the lemon oil in the peel is recovered by washing with water to form an emulsion that is used to produce cold-pressed (CP) lemon oil (see below). After extraction, the juice is separated from the rag, seeds, and ruptured juice sacs in a process known as finishing. The juice sacs, along with a small amount of juice, may be stabilized and packaged to produce a product termed pulpy juice, which can be added to juices, beverages, and other food products. Juice that requires even more depulping can be centrifuged to reduce the insoluble solids level even further.

Most lemon juice is then heat-stabilized and concentrated in special evaporators developed for the citrus industry. These evaporators operate under a moderate vacuum and thus lower temperatures to minimize undesirable thermal effects. While all citrus juices are particularly susceptible to thermal abuse during processing, lemon juice is especially prone to reactions such as browning and acid hydrolysis, which are accelerated by the very high TA of the juice. Most fruit and vegetable juices are concentrated to a standard °Brix level (°Brix is a measure of soluble solids, which is correlated to sugar content or sucrose equivalent). Since the major soluble solid in high-acid fruit (lemons and limes) is citric acid, not sugar, lemon and lime concentrates are produced and sold on the basis of grams of equivalent citric acid per liter of solution (GPL). Lemon concentrate of 400 ± 5 GPL is most commonly produced. The concentrate is then either stored at low temperature in large stainless steel tanks or packed in drums or pails and frozen.

Single-strength lemon juice can be bottled or otherwise packaged for consumer, food service, or industrial use. Most often, the juice is reconstituted to a standardized GPL from lemon concentrate and purified water. Approved preservatives, such as sodium benzoate at very specific levels, are often added. Some not-from-concentrate (NFC) lemon juice is produced and packaged, but its initially higher quality suffers unless the product is quickly pasteurized, packaged, and then frozen (-10 to -20°C) throughout warehousing, distribution, and marketing.

Lemon juice and concentrate are most often used throughout the world not as pure juice but as an ingredient in juice beverages. In many parts of the world, a sweetened, diluted lemon drink, known as lemonade, is the most popular citrus drink. Lemonade is specifically defined in the USA under the Code of Federal Regulations and must have the requisite GPL of citric acid from the lemon juice (not added) to be marketed under that name. Elsewhere, standards may differ, although almost all lemonade is composed of lemon concentrate, water, and sweetener(s), and possibly added flavors and colors, in slightly differing proportions.

Lemon oil is a valuable byproduct and, along with lime oil, has the greatest economic value of all of the CP citrus oils. It has been made for centuries in Sicily by traditional processes, and the methods and oil characterization have been described in the classical essential oil literature. This discussion focuses on

current recovery methods in commercial use in lemon processing facilities.

As the water–oil emulsion is produced just before or during the juice extraction step, it enters the oil-recovery process. The emulsion may contain small particles of peel and other insoluble solids, and may be sent to a finisher for their removal. In both the Brown and FMC systems, the emulsion, containing about 0.5–3.0% oil goes through a two-step centrifugation process. The first centrifuge, known as the desludger, concentrates the mixture to about 50–60% oil, removing water and insoluble solids. The second step, known as polishing, removes the remainder of the water, along with any remaining solids. Depending on the specifications of the oil that is being produced, the lemon oil may be dewaxed. This process is known as winterization and refers to the removal of natural waxes from the peel, by exposing the oil to low temperatures and letting the wax solidify and settle out. The oil is then packaged in containers that are inert and that exclude air.

The typical yield of oil from lemon peel is somewhat less than the total amount available (60–70%) and is dependent on the efficiencies of the operations, fruit condition, and even economic incentives. Lemon oil recovery processes usually try to recover as much oil as possible, due to the value of CP lemon oil. Recovery of lemon oil from Californian lemons ranges from about 4 to 10 kg of oil per tonne of fruit.

Lemon oil quality is related to the aldehyde content, specifically citral, which ranges from about 2.0 to 5.5%, depending on horticultural factors. D-Limonene, the major constituent (> 95%) of orange and grapefruit CP oils, is significantly less in lemon oil but still the major volatile component of CP lemon oil. The largely flavorless, but oxidatively unstable, D-limonene may be removed through various concentration techniques to produce a folded lemon oil that is more stable and more highly concentrated in desirable flavor compounds.

The albedo portion of spent lemon peel is an important source of raw material for commercial pectin production and contains about 35–40% pectin on a dry-weight basis. Pectin is a functional ingredient in many processed food products, where it is used extensively as a gelling agent and mouth-feel enhancer. Pectin exists primarily a polymer of galacturonic acid units, some of which are esterified with methanol, and has a molecular weight of between 100 000 and 300 000. As most pectin is sold as a dry powder, there is a complex recovery scheme to obtain the pectin from the spent peel and process it into a commercially useful product. Figure 1 is a flow diagram outlining the major processing steps in refining citrus pectin, which is extracted from the peel through an acid treatment, then recovered by precipitation with alcohol. Lemon, grapefruit, and lime albedo have a higher level of pectins than that of orange or tangerine, neither of which is typically used for commercial pectin production.

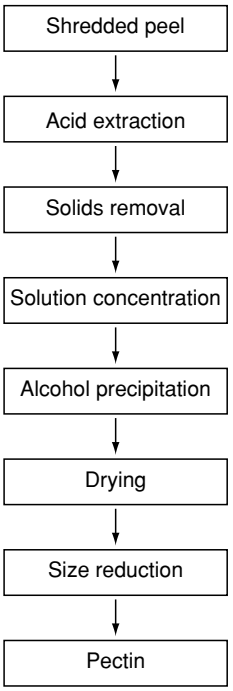

Figure 1 Commercial citrus pectin production.

See also: **Citrus Fruits**: Types on the Market; **Controlled-atmosphere Storage**: Effects on Fruit and Vegetables; **Essential Oils**: Properties and Uses; **Flavor (Flavour) Compounds**: Structures and Characteristics; Production Methods; **Packaging**: Packaging of Liquids; **Pectin**: Properties and Determination

Further Reading

Braddock RJ (1999) *Handbook of Citrus By-products and Processing Technology*. New York: John Wiley.

Carter BA (1993) Lemon and lime juices. In: Nagy S, Chen CS and Shaw PE (eds) *Fruit Juice Processing Technology*. Auburndale, FL: AgScience.

Guenther E (1949) Individual essential oils of the plant Families *Rutaceae* and *Labiatae*. In: *The Essential Oils*, vols. 1 and 3. Van Nostrand: New York.

Kimball D (1996) Grapefruit, lemons, and limes. In: Somogyi LP, Barrett DM and Hui YH (eds) *Processing Fruits: Science and Technology*, vol. 2, pp. 305–336. Lancaster, PA: Technomic.

Saunt J (2000) *Citrus Varieties of the World*. Norwich, UK: Sinclair International Limited.

Sinclair WB (1984) *The Biochemistry and Physiology of the Lemon*. Oakland, CA: University of California.

Grapefruits

J Forsyth, Citrus Consultant, Sydney, NSW, Australia

Background

Grapefruits, with 6% of world production, are the fourth most important commercial group within the edible *Citrus* species. They consist of a small number of cultivars that vary in importance due to market requirements, growing conditions, and climatic zones. Stable production over the last 20 years is now being influenced by new cultivars with lower acid levels and red-pigmented flesh. These factors are creating renewed interest by growers, marketers, and consumers for fruit or processed products.

They are mainly consumed as fresh fruit because of their refreshing, but distinctive, flavor, and they are desirable as a chilled breakfast fruit served in halves or as a juice. The 1970s 'grapefruit diet' created considerable interest in this fruit. There are several health benefits associated with fresh citrus consumption, such as the low levels of energy (calories), fat, and sodium and good sources of vitamin C, dietary fiber, folate, and potassium.

Origin

As with other *Citrus* species, the origin and significance of the grapefruit name remain obscure. It is considered unique as it is the only citrus species known to have originated in the New World and not Asia. Almost certainly, the grapefruit originated in the West Indies/Caribbean region during the seventeenth and eighteenth centuries, before being introduced to Florida in the USA early in the nineteenth century (Figure 1).

Classification

The taxonomic classification of the *Citrus* species is complex and diverse, and not universally agreed upon, in this article grapefruit will be classified by *C. paradisi* (Macfadyen). Citrus as a social and cultural fruit crop, also has local names throughout the world, often for an identical group, e.g., Pamplemousse in French, Pompelmo in Italian, or Pomelo in Spanish. Note that Pomelo (*C. grandis* (L). Osbeck) is also known as shaddock or pomelo/pummelo and, while similar in many characteristics (like shape and rind color, but larger fruits), horticulturally, the pummelos and grapefruit comprise separate classes.

Grapefruit Groups

The grapefruit group and related hybrids group of cultivars is limited, with important commercial but very diverse cultivars that are classified into two groups: common (or white fleshed) and pigmented. The fruit is medium to large in size and more seasonal, with each cultivar maturing and suitable for harvesting over a relatively short period of time. Different cultivars are available that can supply fresh fruit from very early to very late maturity.

Figure 1 (see color plate 34) Grapefruit are often borne in clusters, like a small bunch of grapes and this has been suggested as the origin of the fruits name.

Common or White-fleshed

Marsh is the main cultivar with a number of other, now minor, seedy types including Duncan, Davis, and Wheeny.

Marsh (seedless) This originated as a chance seedling in Florida, selected and commercially propagated in 1886, as the first seedless grapefruit. The cultivar spread throughout the world to become the most important grapefruit grown. In more recent times, planting has declined in favor of the new red-fleshed cultivars.

The fruit is medium in size with few or no seeds. The rind color is pale to light yellow at marurity, medium thin, and very smooth. The flesh is buff-colored, tender, and juicy, with a good flavor. Because of the high heat requirements, successful Marsh production is restricted commercially to very hot climates (Figure 2).

White hybrids A breeding program by the University of California in 1958 crossed an acidless pummelo with a white seedy grapefruit, resulting in the selection and release of two new patented low-acid seedless grapefruit hybrids: Oroblanco and Melogold. The classification is as for grapefruit *C. paradisi*. They have many similar characteristics (thick rinds) but some differences.

Oroblanco This was released in 1980, with major plantings limited to California and Israel. The fruit shape is similar to that of Marsh, but it is larger and slightly flatter. The rind remains greener far longer (even when internally mature) and is thicker. The flesh is paler with an open core. Oroblanco matures about 6 weeks before Marsh, and the flavor is sweeter due to higher total soluble solids (TSS) and lower acid. This internal quality, together with the lack of bitterness, when Marsh is grown in cooler areas, created early interest when it was first released to industry. Initial plantings on vigorous rootstocks like rough lemon and *C. volkamerianna* adversely affected the fruit quality – thick rinds, reduced juice content, and poorer flavor. Oroblanco is considered by growers to be slightly better than Melogold, partly because of the better TSS: acid ratio and smaller fruit (Figure 3).

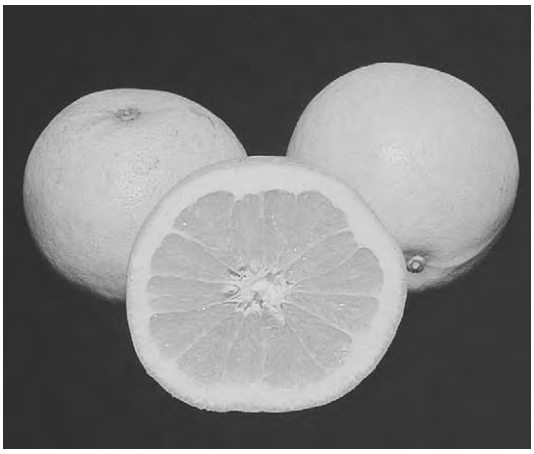

Figure 2 Marsh is one of the oldest, most important and widely grown cultivars.

Melogold This was released in 1985. The fruit shape is similar to that of other grapefruit except for a slightly tapering neck (sheepnose), and the fruit often initially produces significantly larger fruit, but as the trees become older, the average fruit size is reduced. The rind color is more normal than the greener Oroblanco and the taste inferior, although Melogold has a thinner skin and smoother texture, and matures slightly later in the season. The fruit does not peel easily, and the segments do not separate as with traditional breakfast grapefruit or pummelo. Initially, growers in California preferred Melogold, but this is now being replaced.

Pigmented

While the older cultivars were considered 'pink'-fleshed with Thompson and Redblush the main cultivars; newer selections are classified as 'red'-fleshed, due to the deeper pigmentation in the rind and flesh while also being seedless. From this group, Star Ruby, Flame, and Rio Red are becoming more important on world markets, with a consumer trend away from the older white-fleshed cultivars.

Pink-fleshed Foster originated as a limb sport and the first recognized pink-fleshed cultivar located in Florida (1907). The pink or red rind and flesh pigments in grapefruit are lycopene and carotene, and are not due to anthocyanins, as in blood oranges. Thompson (Pink Marsh) was found as a limb sport on a Marsh tree in 1913 and replaced Foster as a commercial cultivar.

Redblush (Ruby, Ruby Red) This originated as a limb sport of Thompson in Texas in 1931, while Ruby is thought to have originated from the same source but in a different locality in 1929. Both are so similar as to be indistinguishable for all practical purposes and are considered identical. The fruit is similar to Thompson (which it generally replaced), except for a much deeper pigmentation in the flesh and albedo (but not the juice) and a crimson blush on the rind, especially at the points of contact between

Figure 3 (see color plate 35) Oroblanco is a new low acid white-fleshed hybrid grapefruit.

fruits. This was the first pink grapefruit that could be recognized externally without having to cut the fruit, due to a rind blush. Having been replaced in some localities by the newer red-fleshed cultivars, Redblush remains an important cultivar in Florida, California, and South Africa (Figure 4).

Red-fleshed The following cultivars are relative new to the industry, originating from more recent natural mutations or seedlings, breeding, or irradiation programs in the USA from the 1970s:

Star Ruby This was selected as a seedling from a Texas seed irradiation program in 1959 and released to industry in 1970. The fruit has a yellow, smooth, fine, and thin rind, with red blushes. The flesh is redder than that of Redblush, with a later maturity, higher TSS and acid but few seeds (rarely more than two). The juice content is high and as sweet as, or sweeter than, Marsh and other new pigmented cultivars.

The tree shows herbicide, nematicide, heat, and cold sensitivity. These deleterious effects have been widely recorded, as Star Ruby appears the most difficult of the new grapefruits to grow successfully. It is also more susceptible to *Phytophthora* diseases and to *Alternaria* brown spot. These problems lead to erratic bearing, small fruit size, low tree vigour, and short longevity.

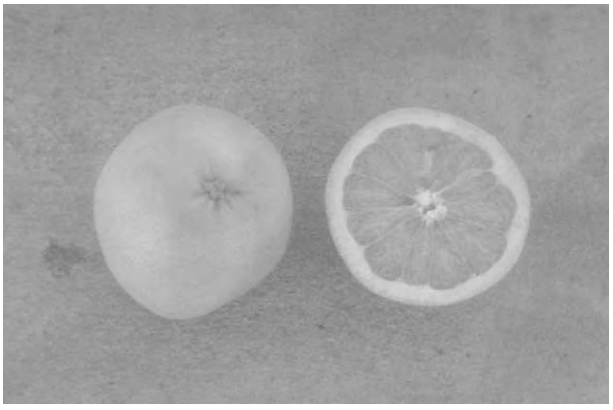

Figure 4 Redblush was the first cultivar to have a deeper red flesh and some rind pigmentation.

Subsequent nucellar clones from the original Star Ruby, now available in Israel, Southern Africa, and Australia, produce high-quality fruit with less of the above-mentioned cultural problems (Figure 5).

Flame This is now thought to have originated from a natural mutation of Ruby Red grapefruit seed collected in Texas (1973) and, after selection and evaluation, was released in 1987. During the last decade, more has become known about Flame's performance, especially from Florida and Argentina.

Trees are vigorous and not as cold- or herbicide-sensitive as Star Ruby. The larger round fruit has an attractive appearance with a very smooth thin rind; an external blush is slightly less than Rio Red or Star Ruby. The background rind color tends to be light bronze (similar to Star Ruby) rather than yellow like Rio Red. Internally, the color is of a uniform intensity, similar to Rio Red, and uniform across the segments. Fruit hangs on the tree, with a maturity similar to that of Star Ruby, but earlier than Rio Red. Flame is now widely planted in Florida and is expected to perform well in similar climates.

Figure 5 (see color plate 36) Star Ruby is a newer grapefruit cultivar with a deep red pigmented flesh.

Rio Red This originated in Texas from irradiated budwood sourced from a Ruby Red tree, which, in 1976, produced a natural bud mutation that was then evaluated before release in 1988. The fruit shape is generally good but tends to be slightly sheepnose when grown under semiarid and drier conditions.

The Rio Red fruit is the latest maturing of the new cultivars, and a deep blush develops where adjacent fruits touch. The flesh color may be patchy, especially early and late in the season, but for much of the marketing period is only slightly lighter than Star Ruby. The thicker rind is compensated for by a softer flesh and higher juice content.

Henderson and Ray Ruby These are two other new red-fleshed varieties from Texas that have attracted only a minor growing or marketing interest. All these new red cultivars have been available for a relatively short period. Little comparative information is available for the red-fleshed cultivars, on the effect of different climatic growing conditions in many citrus-growing regions of the world, i.e. the effect on tree longevity, yield, fruit quality and pigmentation. Table 1 summarizes several important fruit quality parameters.

Growing Regions and Conditions

Grapefruit, with a range of total heat requirements, is grown in varying climatic conditions and regions, including the humid tropics, arid subtropics, and intermediate climates. Commercial production tends to be located in two narrow belts of the subtropics and between 20 and 40° latitude north and south of the equator.

Resistance to heat is outstanding and in cold tolerance is similar to that of sweet oranges. Their very high heat requirement for the production of good-quality fruit restricts successful commercial culture to hotter climates. The sweetest, juiciest, and most bitter-free fruit is grown in semitropical summer rainfall areas such as Florida, Texas, and areas of Southern Africa. This makes grapefruit almost equally well adapted to hot desert and humid semitropical or tropical climates. In the cooler drier Mediterranean climates, there is a tendency for the fruit to have a thicker rind, lower sugar, high acid levels, and some bitterness.

Under desert growing conditions, the color is brighter and deeper with a rich flavor more pronounced than in humid climates, where smaller fruit sizes, a less oblate shape, and a lower juice content

Table 1 Some internal and external grapefruit quality parameters

Variety	*Percentage juice*[a]	*Percentage acid*[a]	*TSS:acid ratio*[a]	*Number of seeds*[a]	*Rind*	*Flesh*
Oroblanco	47	0.6	15.9	0	Nil	Nil
Melogold	40	0.8	14.5	6	Nil	Nil
Marsh	39	1.5	7.5	2	Nil	Nil
Thompson	na	na	na	3	Nil	1
Redblush	40	1.3	6.6	2	2	2
Henderson	36	1.9	5.8	1	4	3
Ray Ruby	35	1.8	5.8	2	4	3
Rio Red	43	1.3	7.1	4	4	4
Flame	41	1.3	7.7	3	4	4
Star Ruby	46	1.4	8.1	2	5	5

[a]University of California sampling date December (various years for comparison) and rind or flesh pigmentation 1 pale/poor to 5 deepest.
na, not available.

Table 2 World grapefruit production trends

Year	*Total production* ($\times 10^3$ t)	*Exported* ($\times 10^3$ t)	*Processed* ($\times 10^3$ t)
1980/81 to 1988/89 average	4640	920	1815
1995/96	5007	1175	1986
1996/97	5056	1177	1919
1997/98	4726	1082	1733
1998/99	4614	1135	1682
Four-year average	4851 (38%)	1142 (24%)	1830 (38%)

From FAO (1999) *Citrus Fruit Fresh and Processed, Annual Statistics, Tables 1–10 and 16–20*. Rome: Food and Agriculture Organization; USDA (1999) World fresh citrus situation. In: *World Horticultural Trade and US Export Opportunities, Circular Series FHORTH*. Foreign Agricultural Service.

may occur. For example, Florida leads the world in grapefruit production, with a warm to hot and humid climate, resulting in a grapefruit of excellent internal quality, with a thin peel and attractive shape.

For the newer red-fleshed cultivars, the trees are resistant to heat but sensitive to cold. They can be grown under a range of climatic conditions, but very hot and warm night temperatures are necessary to produce high-quality fruit. The newer low acid cultivars like Oroblanco may be harvested and marketed before the full yellow rind color develops, with green or only pale colored rinds, although the fruit is mature to eat with a TSS: acid ratio of 8:1 and less than 1% acid.

Site selection and soil management will be influenced by climate, while some cultural practices may modify harmful effects, including cold protection, irrigation, and windbreaks. However, they cannot completely overcome major unfavorable climatic effects on tree growth or fruit quality. Some of these effects include regular severe frosts, poor rind color/pigment development, variable maturity, high acidity or naringin levels, poor fruit form, or pest and disease incidence.

Although susceptible to the normal range of citrus pests and diseases, they are particularly susceptible to grapefruit stem pitting caused by certain strains of Citrus Tristeza Virus (CTV). The disease will cause sudden or 'quick tree decline,' pits in the tree and small, lopsided or malformed fruits. CTV is spread in infected budwood and the brown citrus aphid *Toxoptera citricida*. This disease is endemic in many countries, but the planting of tolerant rootstocks like Troyer and Carrizo citrange, together with only propagating nursery trees with budwood that has been preimmunized and carries a mild strain cross-protection, will limit losses.

Production Statistics

Although grapefruit is a relatively new citrus fruit, having been marketed for just over 100 years, grapefruit is now the fourth most important citrus with 6% of total production in 1998–99, after oranges (with 66%), mandarins (17%), lemons, and limes (11%).

The world's grapefruit production has been almost stable from the 10-year average of 4640×10^3 t in 1980/81 to 1988/89, to only slightly larger average total crops (allowing for annual variations) in the four 1995/96–1998/99 seasons of 4851×10^3 t, as shown in Table 2. About 20% of this production is estimated to be pummelo fruit.

Although production is forecast to increase slightly to 4930×10^3 t by 2005, this is still below the slightly

higher production recorded in 1995–97 (Table 2). The latest increases will mostly be in pummelo production, especially in Asia, China, the Philippines, Indonesia, and India. Only small increases in are grapefruit production are expected in Mexico, Argentina, Cuba, Brazil, and South Africa.

The largest growing regions are in the northern hemisphere, with 86% (Table 3) of world production, and include countries like the USA, Israel, Cuba, and China, while the southern hemisphere, with only 14%, is relatively minor but expanding, mainly in Argentina and South Africa.

Unlike other citrus types, the Mediterranean region (apart from Israel) produces only minor quantities of grapefruit (14% of the world production, Table 3), mainly due to local climatic conditions. There is, however, interest in producing new red grapefruit cultivars, especially in Spain, Cyprus, and Turkey.

In the two largest citrus-producing countries (Brazil and the USA), grapefruit are relatively minor with only 4% in Brazil, but due to a wider range of climatic regions especially in Florida and Texas, a more significant 19% of USA total production. The relative importance of grapefruit producing in some countries is outlined below.

Table 3 Fresh grapefruit crop production and utilization (with respective percentages), and the main producing countries throughout the world in 1998–99

Region/country	*Total production* $\times 10^3$ *t (%)*	*Exports* $\times 10^3$ *t (%)*	*Processed* $\times 10^3$ *t (%)*
Northern hemisphere			
USA	2285	425	1200
Mediterranean region			
Italy	2		
Spain	28	22	2
Israel	321	123	184
Algeria	1		
Morocco	2		
Tunisia	50		
Cyprus	52	33	16
Turkey	85	63	11
Subtotal (Mediterranean)	631 (14%)	235	217
Cuba	270	40	155
Mexico	180		35
China	255		
Subtotal (northern)	3979 (86%)	950	1607
Southern hemisphere			
Argentina	210	30	40
Brazil	62		
Australia	17		
South Africa	189	145	35
Subtotal (southern)	634 (14%)	186	75
Total (world)	4614 (39%)	1135 (25%)	1682 (36%)

From FAO (1999) *Citrus Fruit Fresh and Processed, Annual Statistics, Tables 1–10 and 16–20*. Rome: Food and Agriculture Organization; USDA (1999) World fresh citrus situation. In: *World Horticultural Trade and US Export Opportunities, Circular Series FHORTH*. Foreign Agricultural Service.

USA

Florida is the principal state with about 17% (or 56 630 ha) of citrus plantings devoted to grapefruit, with Marsh the main cultivar (50% of planting) because of its use in both fresh and processing markets. Older, pink-fleshed varieties like Thompson and Redblush make up 37% and were the industry standards. Recent development has revolved around the new 'seedless' red-fleshed cultivars. Since 1982, many new plantings were Flame and Ray Ruby and, more recently, Rio Red. Some Star Ruby has been planted, but cultural problems have limited grower acceptance, despite prices usually being higher than for other red cultivars.

Interest in the newer cultivars is based on current consumer preference for red-fleshed grapefruit. Because of the long marketing season, retaining flesh color is critical. Flame and other deep red-fleshed fruit begin and end the season with a better color than that of Redblush. A rind blush also enhances consumer appeal, and despite perceptions, there are no substantial quality differences between white and red-fleshed grapefruit.

White grapefruit are easily processed (in the early 1990s, about 70% of the crop), but a relatively high TSS:acid ratio juice is required while a smaller part of the red types are processed (30–40%). Fresh fruit exports are important, e.g., for white seedless cultivars to Japan, but the French market favors the red-fleshed seedless.

In Texas, before the freezes in the late 1980s, nearly 66% of all citrus plantings were devoted to grapefruit, mainly the red-fleshed varieties – Redblush, Rio Red, Star Ruby, Ray Ruby, and Henderson. Extensive replanting in the early 1990s indicated that 61% were grapefruit, with 87% Rio Red (replacing the inconsistent bearing and cultural problems of Star Ruby) and 5% Henderson and Ray.

In California and Arizona, grapefruit are less important (9% of plantings), but Marsh and Redblush are grown in selected areas. Oroblanco was being grown (1000 ha planted by 1999) for the Japanese export market and was preferred by growers over Melogold, but in recent years, returns have been reduced markedly. Problems with low production, small fruit sizes, and thick rinds are now reported, and only limited plantings are expected.

Israel

In the 1980s, Israel's grapefruit declined from a 60% dominance of the European import market to 30%,

with shelves left available for red grapefruits from Florida. Following restructuring of the industry's marketing activities, success has been achieved in the export of 'Sweetie' (Oroblanco) grapefruit. This grew successfully in Israel and was found to be highly suited to the Japanese palate. In recent times, most of the 'Sweetie' grapefruit produced has been exported to Japan and sold as green-skinned, early grapefruit or later in the season with yellow skin as 'Golden Sweetie.'

Grapefruits have expanded rapidly, resulting in the planting of 27 000 ha, with a low 13%, nonbearing. About half are red-fleshed cultivars, like 'Sunrise' (Star Ruby) and 'Sweetie' (Oroblanco).

Cuba

Given its climate, citrus are widely grown in Cuba, with grapefruit production increasing to 270 000 t in 1998/99. Following the loss of some European export markets in the 1990s, the markets declined from a 10-year average of 134 900 t to only 43 750 t during the four years of 1995/96 to 1998/99. An increased percentage of the crop is processed, up from the 10-year average in 1980/81 to 1988/99 of 32 800 t to an average of 158 000 t during the 4 years of 1995/96 to 1998/99. The main cultivar grown has been the traditional white Marsh.

China

Although listed as the fourth largest producer of 'grapefruit' in the world, most production is pummelo. Owing to favorable climatic conditions and the very large domestic market, production will increase generally in tropical Asian countries.

Southern Africa

Although only the sixth largest grapefruit producer, Southern Africa's production has increased significantly from the 10-year average of only 76 900 t in 1970/71 to 1978/79 to 189 000 t in 1998/99. In the southern hemisphere, this production is only slightly smaller than the fifth producing country, Argentina (210 000 t, Table 3). Grapefruit makes up 17% of plantings with 5% red-fleshed cultivars, 10% white, and 2% rose or pink, with greatest expansion for Star Ruby. Of all the other red-fleshed cultivars, their degree of pigmentation is far less than that of Star Ruby. Total exports have increased rapidly since 1984 from 24 000 to 548 000 cartons.

Harvesting and Marketing

The growing, harvesting, and marketing of grapefruit can present unique problems, which are best overcome by specialist growers in selected climatically suitable localities. Present-day production focuses on careful cultivar selection to best suit conditions and cultural practices, in order to achieve a successful marketing outcome.

Maturity

Early in the marketing season, grapefruit are especially prone to poor palatability and low consumer acceptance due to the high acidity and bitterness from the glucoside naringin and/or limonin. This problem is worse where fruits are harvested from marginal growing climates and poorly chosen rootstocks. Most regions now have maturity standards to insure that all marketed grapefruits are palatable and acceptable to consumers, e.g., in Southern Africa, export standards have been set at 50% minimum juice, 90° TSS, 1.75 maximum acidity and 6:1 minimum TSS:acid ratio. As grapefruit mature on the tree, the TSS increases, and the acidity decreases, but once harvested, these do not change.

The estimated maturity and main marketing periods for some grapefruit in Southern Africa (southern hemisphere) are:

- Marsh: end of April to end of August
- Ray Ruby, Flame, Star Ruby: all maturing mid-May to end of August
- Rio Red: early June to end of September.

Only a limited number of cultivars may be grown in any locality, but there is a small difference in optimum maturity (see above) that can be taken into account when selecting those cultivars to be grown and helping to insure a continuous supply of fresh fruit for the available markets. In Florida, all commercial cultivars reach their legal maturity in September or October (northern hemisphere), but can be stored on the tree (as described below) for several months and up to the end of May, when they begin to fall, and seeds start germinating within the fruits.

Degreening

Early in the season, mature grapefruit may not be fully colored. Fruit may be degreened by exposure to ethylene gas for short periods in special chambers.

Preharvest Drop

Grapefruit are prone to preharvest drop and severe crop losses (Figure 6). A common cultural practice is to apply a registered growth regulator (like 2,4-D amine salts) at color change. This spray has the beneficial effect of delaying the formation of the abscission cell layer at the junction of the fruit stem and button. It greatly reduces fruit drop over a period of

Figure 6 Preharvest fruit drop off near mature grapefruit.

several months and helps retain green buttons after harvest. Fruit can be 'stored on the tree' for up to 7 months past normal maturity and fruit drop, enabling more orderly harvesting and marketing over much longer periods of time than would be normal without treatment.

Rind Aging

Another cultural practice is the application of a registered growth regulator, gibberellic acid, to slow down the rind aging process without affecting fruit quality, by reducing the rind thickness and puffiness in late-hanging fruit.

Cold Storage

Apart from the normal marketing periods, grapefruit are not held for long-term cold storage. Fruit may be stored for 5–8 weeks at 15.5–10.0 °C but are very susceptible to common storage breakdown problems and chilling injury, like peel pitting and oil-gland darkening.

Crop Utilization

Grapefruits are principally marketed according to how they are consumed, with local fresh fruit and exports totaling 62% and 38% processed (compared with 34% of oranges) as summarized in Table 2. All major producing countries process a significant proportion of their crops; the USA is the largest processor with 53%, followed by Israel and Cuba (both 57%). See Tables 2 and 3.

In the world fresh grapefruit export fruit trade in 1998/99, the following regions/countries were the major importers: Japan, the Mediterranean region, the Netherlands, France, and Belgium.

Fresh Fruit

Grapefruit cannot be considered a 'sweet dessert' fruit, but because of its refreshing flavor and mild bitterness, fresh grapefruit is desirable as a healthy breakfast fruit. Chilled halves with the sections loosened from the peel, flavored with sugar or honey to taste and the pulp spooned out make a refreshing start to the day.

As an appetizer before dinner, grapefruit halves may be similarly sweetened, lightly boiled, and served hot, often topped with a maraschino cherry. Sections are commonly used in fruit cups or fruit salads, in gelatins, puddings, or tarts. Fresh (chilled) juice makes a refreshing healthy drink any time of the day. Segments are suitable for garnishing and decorating salads or topping desserts. The fruit and juice can be used in dishes such as grapefruit and ginger chicken.

Grapefruit and other edible citrus fruits derive their unique qualities from about 20 principal factors, which can mainly be conveniently grouped as internal and external. The internal or chemical ones can also be measured like sugars or carbohydrates (sucrose, glucose and fructose as total soluble solids – TSS), organic acids (citric and malic measured as titratable acidity – TA), sugar:acid ratio – TSS; TA, juice content and color, are of vital concern for consumer acceptance and preference. All of these factors change

Table 4 Selected food values per 100 g of some edible grapefruit portions[a]

Chemical composition	*Pulp (raw)*	*Juice (raw)*
Calories	34.4–46.4	37–42
Moisture	87.5–91.3 g	89.2–90.4 g
protein	0.5–1.0 g	0.4–0.5 g
Fat	0.06–0.20 g	0.1 g
Carbohydrates	8.07–11.5 g	8.8–10.2 g
Fiber	0.14–0.77 g	trace
Sodium	1.0 mg	1.0 mg
Potassium	135 mg	162 mg
Vitamin A white	10 IU	10 IU
pink/red	440 IU	440 IU
Ascorbic acid	36–49.8 mg	36–40 mg

[a]According to analyses made in California, Texas, Florida, Cuba and Central America.
Adapted from Morton J (1987) Fruits of warm climates grapefruit, pp. 152–158, Miami, Florida.

as fruit matures and with grapefruit can be extended for over six or seven months on the tree, before becoming overmature and unacceptable.

See Table 4 for some food values of normal edible grapefruit portions – pulp and juice. The bitterness factors of limonin are normally found in the peel oil, while glycoside naringin occurs in the pulp segments.

Processed Products

Processed grapefruit products include white, pink, or red segments in syrup or natural juice, juice products (fresh or frozen, canned, dehydrated as a powder, or concentrated), and rind oil or essence as often used in aromatherapy, for massage, or in baths. After the juice is extracted, there remain residues that can be a source material for other valuable byproducts. Some specialized types are also candied, dried, or used for marmalade, jelly, and pectin manufacture.

White-fleshed varieties are mainly processed into higher-priced juice products. A smaller percentage of red-fleshed varieties are juiced, one reason being because if the processing fruits and therefore the juices are mixed the red lycopene pigment turns the white juice an undesirable brown or muddy color, and juice prices are lower for normal red juice. With the trend towards increased red-fleshed varieties, this has resulted in problems of disposing of processing or non-fresh marketable fruit. Whether significant volumes of red juice and other processed products can be successfully marketed will depend on innovative companies with new products like red grapefruit juice cocktails and aggressive promotion and marketing.

See also: **Citrus Fruits**: Types on the Market; Composition and Characterization

Further Reading

Castle WS (1994) A production and marketing perspective on Florida grapefruit and mandarin cultivars. In: *Proceedings of the Second Horticultural Industry Technical Conference, NSW.* Sydney: AUS HS Wine Titles.

Cottin R (coordinator) (1997) *Citrus of the World. A Citrus Directory.* San Giuliano, France: SRA INRA-CIRAD.

FAO (1999) *Citrus Fruit Fresh and Processed, Annual Statistics, Tables 1–10 and 16–20.* Rome: Food and Agriculture Organization.

Reuther W, Webber HJ and Batchelor LD (eds) (1967) *The Citrus Industry*, vol. 1. Berkley, CA: University of California.

Saunt J (2000) *Citrus Varieties of the World. An Illustrated Guide*. Norwich, UK: Sinclair International.

Sinclair WB (1984) *The Biochemistry and Physiology of the Lemon and other Citrus Fruits*, pp. 711–718. Berkley, CA: University of California.

Swingle WT and Reece PC (1967) The botany of citrus and its wild relatives. In: Reuther W, Webber HJ and Batchelor LD (eds) *The Citrus Industry*, vol 1. Berkley, CA: University of California.

USDA (1999) World Citrus Situation. In: *World Horticultureal Trade and US Export Opportunities Circular Series* FHORTH, pp. 18–32. USA: Foreign Agricultural Service.

Webber HJ and Batchelor LD (eds) *The Citrus Industry*, vol 1. Berkley, CA: University of California.

Young RH (1986) Fresh citrus cultivars. In: Wardowski WF, Nagy S and Grierson W (eds) *Fresh Citrus Fruits*, pp. 110–125. Westpoint, CT: AVI.

Limes

R E Berry, Citrus/Tropical Fruit Products, Winter Haven, FL, USA

Introduction

Limes constitute one of the general types of citrus fruits long known and consumed throughout the world. Relative values of parts of the fruit are closely related to relative amounts of acid and sugars, and to the amount and composition of the peel oil. Several main cultivars and fruit types of limes are quite different in appearance as well as physical and chemical composition. Lime juices and oils are prepared by several different processes, and their economic value is directly related to the method of preparation, as well as the fruit source.

History and General Characteristics

Limes are known world-wide for their tart, tangy-flavored juice and especially for their unique flowery, characteristic aromas. They were introduced to many parts of the world through the explorations of the British navy, where they were probably first used for clinical and nutritional purposes. Their extensive use on-board ship to prevent or treat attacks of 'scurvy' (caused by vitamin C deficiency) earned British sailors the distinctive name 'Limeys.' Most evidence indicates that limes probably originated in northeastern India and migrated via the trade routes into the Middle East and on to the Western world through sailing vessel trade markets.

The juice of most types of limes is too acidic to be consumed undiluted. However, limes are popular for use in juice mixtures, -ades (mixes with sugar, water and/or other juices) and carbonated beverages, and as a component of alcoholic drinks and mixes. In some countries, they are used in pickling, culinary, and medical applications. A main economical value of limes comes from the highly valued lime oil, used in products as diverse as cosmetics (shave lotions, perfumes, and colognes) and household cleaning products (washing powders, soaps, furniture polish, and bathroom deodorants). Lime trees are extremely cold-sensitive and will be productive only in subtropical to tropical climates where temperatures seldom go below the freezing point.

Limes are the most acidic of all citrus fruits and lower in vitamin C content than lemons. Like lemons, they are highly valued for their juice and acid content. Most limes consist of 40% or more juice by weight or volume. The remainder consists of seeds, vascular bundles, parts of albedo (white inner part) adhering to the skin or flavedo (outer, colored part), and parts of fruit segments, collectively referred to as 'rag.' The juice of most cultivars consists of about 90–91% water, 0.2–0.3% protein, 0.7% (in acid types) to 9.0% (in sweet types) sugars, 0.1% fat, and 0.3% minerals (ash). The juice (100 ml) supplies 110–140 kJ (26 kcal) of energy, 50 mg of ascorbic acid (vitamin C) and a trace of dietary fiber. A wide range of vitamin C levels has been reported, and, as with most citrus fruits, the level is higher in the peel or rind (63–121 mg%) than in the fruit or juice (30–70 mg%).

Quite different from the basis of value for other types of citrus, the value of limes (and lemons) is based primarily on the weight of acid (as citric, although there are traces of many others) per unit volume of juice. This is in contrast to the ratio of sugar to acid used for oranges, grapefruits, tangerines, etc. Computer programs and automatic titrator instruments provide such data at a high rate in citrus processing plant quality-control and inspection laboratories. The values reported for lime juice acid content (as citric) range from about 6.0 to 8.5% (average 7.20%).

Cultivars and Fruit Types

The small-fruited, acid lime (*Citrus aurantifolia* Swingle) is the predominant type in the world. It is widely grown in India, Mexico, Egypt, the West Indies, China, California, the southern tip of the Florida Peninsula, and the Florida Keys (USA), where it is known as the 'Mexican' or 'Key' lime. Although found in Brazil as the 'Gallego' lime, it is not as well known there. The fruit is small, round, somewhat short-elliptical, and bright to greenish yellow when mature. Moderately seedy, it has a thin rind (smooth and leathery), and the flesh is very juicy, highly acid, and has a unique, distinctive flowery aroma.

The large-fruited, acid limes (main cultivar *Citrus latifolia* Tan.) are more oblong in shape, light to dark green at maturity, and two to five times larger, on average, than the small-fruited limes. These limes, often called Tahitian or Persian, were brought to California from Tahiti, but do not seem to have any connection with Persia. About the size and shape of the common Bearss lemon, they may be a hybrid of the citron and/or the lemon, but the parentage and true source are not definite. This lime is grown mostly in South Florida (and, to some extent, in Australia and California) because of its better tolerance of cold and its lower requirement for heat to reach maturity. Other types of lime – the Bearss, from California, as well as the Sahesli of Tunisia and the Pond of Hawaii – seem certain, from horticultural evidence, to be the same as the Persian. The juice has a high-quality characteristic lime flavor, and the aroma is typical of that commonly described as 'lime-like' throughout most of the world.

The sweet lime (*Citrus limettioides* Tan.) is widely consumed in India, the Middle East, and Latin America. It is known as the Palestine or Indian sweet lime, and is used in those countries for medicinal purposes. The fruit, medium in size and round to slightly oblong, is greenish yellow to orange-yellow at maturity. The flesh is yellowish; the rind has a distinctive, slightly medicine- or turpentine-like aroma, and the juice is so nonacid as to be insipid to most tastes. These limes are seldom processed except in very small-scale, simple home-type juicing and bottling operations. (*See* **Atherosclerosis**.)

Lime Juice

Lime juice can be prepared and concentrated in large amounts at high rates in the same general manner as

for orange and grapefruit juices, using similar equipment. For juice extraction, however, smaller (than usual for oranges) cups are used on one type of extractor, which individually cuts and presses each fruit. Another type of extractor presses and disrupts the entire fruit between rotating converging disks, and the fruit, pulp, and rag mixture is then separated from the juice in a screw press/screen finisher. Single-strength lime juice may be blended with sugars, syrups, water, or carbonated water and bottled in glass or sealed in metal cans. These types of juice products are usually preserved with a small amount of added sulfite or bisulfite. They may remain stable without refrigeration up to a year or longer.

The juice can be concentrated on a thermal accelerated short time evaporator (TASTE), widely used throughout the world citrus processing industries. However, since lime juice contains a higher percentage of organic acids than of carbohydrates, the degree of concentration is determined as grams per liter of acid rather than the degrees Brix (approximate percentage of sugars) commonly used for other citrus concentrates. However, Brix measurements with a commonly available Brixometer remain necessary, to determine the bulk density from which the GPL can be calculated. Computer programs and automated instrumentation provide dependable determinations at a high rate. Most lime concentrates are evaporated to about 400–500 $g\,l^{-1}$. The uncorrected, observed, desired Brix value (roughly analogous to the acid concentration desired) must be known since evaporation control is based on in-line Brixometers. The approximate desired Brix value must be determined based on the acid and Brix of the incoming single strength juice, using standard methods.

Lime Oil

Lime oil, a major commercial product in world markets is, in most processes, more valuable than the juice, which may be considered a byproduct of the lime oil process. It may be prepared by allowing the liquid, separated from the pressed mixture of fruit parts described above, to stand until the oil 'layers' off at the top, and then decanting or distilling off the layer. The character and composition of the decanted oil vary considerably from those of the distilled oil. As a byproduct of this process, the bottom juice layer is a clarified, sparkling clear, light green-colored juice. Special clarifying additives are often used in this process to enhance the clarity of the resultant juice. This product is distributed widely around the world, especially in the UK and the Caribbean area, as a mixer base for alcoholic and soft drinks.

Lime oil may also be prepared by a long-used conventional 'cold-pressing' process, whereby the flavedo and adhering fruit parts are passed into a conical pressure filtration press. The 'peel juice' is pressed through a steel screen, and the oil is separated in a high-speed centrifuge. This process produces a clear, green-colored oil with a fine aromatic quality and stability of aroma and flavor. The juice or concentrate resulting from this type of process has a stable, well-suspended 'cloud' and is used in bases for juice drinks, -ades, and soft drinks. (*See* **Soft Drinks**: Production.)

The oil is located in oval-shaped oil sacs situated rather evenly throughout the flavedo or outer layer of the peel. The oil sacs function as a natural toxic barrier to microorganisms and insects that might otherwise attack the fruit. The oil is composed of a base sesquiterpene hydrocarbon (D-limonene, more than 50%) and a mixture of other mono- and sesquiterpenes. From lime oil analyses many organic components have been identified, including 12 alcohols, seven aldehydes, four esters, one ketone, and 22 hydrocarbons, with 7% nonvolatiles (mainly coumarins).

To control and classify the quality and value of citrus oils, the National Research Council, officially recognized by the US Food and Drug Administration, has published and regularly updates the 'Food Chemicals Codex' (FCC). A current summary of lime oil specifications and standards can be readily obtained from that source.

Storage

Because of their extremely high acidity and low pH, limes are more stable than many other citrus fruits. However, problems may be encountered at certain temperatures and humidities with growth of acid-tolerant surface fungi. Control has been accomplished using biological growth regulators. In Persian limes, 2,4-dichlorophenolindophenol (24D) and gibberellic acid (GA) have been successful in reducing the growth of '*Penicillium*'-type molds, yellowing, mottling and discoloration of the rinds during storage. Such problems have not been solved through the use of controlled-atmosphere storage, but can be greatly controlled through the use of optimum storage conditions for fresh fruit, i.e., 10 °C and 90–95% relative humidity where a storage life of up to 4 weeks or longer has been reported without any notable injury, pitting, or discoloration of fruit.

Value

The great majority of limes in the world are produced in India, Indonesia, and Mexico, predominantly for

local fresh fruit markets, and thus statistical data on value and extent of the crop and the fresh market are unavoidably biased and not highly accurate. Few dependable agricultural statistics are available for local fresh markets in most geographic areas. Although the USA is a notable and important producer of limes and lime products, it is not the major world producer of the fresh fruit, but the most dependable statistics on production and value of products are those for the USA. The total annual production of all types of limes in the USA ranges around 70 000–80 000 t. In general, the amount of those processed has been about 50–60% of the total crop. The annual value varies considerably and is highly influenced by climate, local weather, outbreaks of plant diseases, and insect-control measures. The USA imports five to 10 times more limes than it exports, and the total USA annual crop has been equal to about 25% of the world crop during the past decade. Data on US production of limes, lime juice products and lime oil are available from the US Department of Agriculture statistics department, and Florida fruit data are available from the Florida Department of Citrus, Lakeland, FL.

See also: **Ascorbic Acid**: Properties and Determination; Physiology; **Controlled-atmosphere Storage**: Effects on Fruit and Vegetables; **Essential Oils**: Properties and Uses; **Scurvy**; **Spoilage**: Fungi in Food – An Overview; **Storage Stability**: Mechanisms of Degradation

Further Reading

Florida Department of Citrus (1998) *Annual Citrus Statistical Data*. Lakeland, FL: Florida Department of Citrus Headquarters.

Hicks D (ed.) (1990) *Production and Packaging of Noncarbonated Fruit Juices and Fruit Beverages*. New York: Van Nostrand Reinhold.

Kimball DA (1999) *Citrus Processing, A Complete Guide*. Gaithersberg, MD: Aspen Publishers.

Pattee HE (1985) *Evaluation of Quality of Fruits and Vegetables*. Westport, CT: AVI Publishing.

Sinclair WB (1984) *The Biochemistry and Physiology of the Lemon and Other Citrus Fruits*. Davis, CA: University of California, Division of Agriculture and Natural Resources.

US Department of Agriculture (1998) *Agricultural Statistics*. Washington, DC: US Government Printing Office.

US Food and Drug Administration (1999) *Food Chemicals Codex (FCC)*. Washington, DC: US Government Printing Office.

Wardowsky WF, Nagy S and Grierson W (eds) (1986) *Fresh Citrus Fruits*. Westport, CT: AVI Publishing.

Mandarins

J Forsyth, Citrus Consultant, Sydney, NSW, Australia

Background

Mandarins, which are also known as tangerines[1] in some parts of the world, and their related hybrids, are the largest and most variable group within the edible *Citrus* species. They consist of many cultivars that vary in importance according to local and export market needs, growing conditions, and climatic zones. The large and increasing production has significance in local and world trade, especially for fresh fruit. Some brief details on mandarin classification and origin of the species, growing regions and conditions, production statistics, harvesting marketing, and crop utilization are discussed.

Fruits range from oblate to globose to subglobose in shape, and small to medium in average size. The rind varies from thin to medium in thickness, with a smooth to rough or pebbled surface texture that may be easy or tight to peel. The color of both the rind and flesh may be yellow, or light to dark orange. Maturity ranges from very early to late, depending on the cultivar and growing conditions.

Mandarins are mainly consumed as fresh fruit. There are several health benefits associated with fresh citrus consumption: citrus fruits are a source of many essential nutrients and compounds, and a low-energy (low calories), fat-free source of vitamin C, dietary fiber, folate, potassium, and several phytochemicals. These multiple nutritional benefits from citrus, and especially the very convenient mandarins group, are only possible through the consumption of fresh fruit and juices and currently cannot be obtained from supplements.

Classification

The taxonomic classification of the *Citrus* species is complex and diverse, and not universally agreed upon, but those used by Swingle and Reece will be followed in this article. Citrus, as a social and cultural fruit crop, also has countless local or regional names and identities[2] throughout the world, often for an

[1]Tangerine was initially adopted in the USA, to better describe those mandarin cultivars that had a deep reddish orange rind like the Dancy cultivar.

[2]Ortanique originated in Jamaica but is also known as Topaz (in Israel), Tambor (Southern Africa), Mandore (Cyprus), Ortaline (Morocco), and Ormanda (Honduras).

identical cultivar. (*See* **Citrus Fruits**: Types on the Market.)

The mandarin group and related hybrids are very diverse (with over 100 cultivars), but can be classified into a number of main groups, some of which contain a large number of subgroups. This fruits is more seasonal, with each cultivar maturing and suitable for harvesting only over a relatively short period of time, e.g., 6–8 weeks on average. Several cultivars are now available for the main types that can supply fresh fruit from very early to very late maturity. Mandarin production is increasing, with a consumer trend towards fruit that is easier to peel and segment and is seedless. Only limited quantities are processed into juice or canned segments.

- Common mandarins (*Citrus reticulata* Blanco) are a very extensive group of different cultivars now widely grown and available in world markets; e.g., Imperial, Kara, Emperor, Ponkan, Dancy, Kinnow, Sunburst, and Fortune are all old cultivars, with some examples of newer common mandarins including Michal, Nova, Fallglo, Encore, and Daisy. This large group also includes several other commercially important subclasses.

  The clementine may have originated in Algeria, as a natural hybrid located and growing there in the 1890s.
  Two hybrid groups exist, e.g., tangors, which are hybrids of the mandarin and orange *C. reticulata* × *C. sinensis* [L] Osbeck like Murcott, Afourer, Mor, Ortanique, Temple, Tankan, and Ellendale. Tangelos are hybrids of the mandarin and grapefruit or pummelo *C. reticulata* × *C. paradisi* Macfadyen with Minneola, Orlando, Page, and Seminole as cultivar examples.
- Satsuma mandarins (*C. unshiu* Marcovitch) are also known as *Unshiu mikan* in Japan and China. They were initially mainly grown in Japan and Spain up to the 1980s but are now an important expanding seedless group throughout the world, because of their cold hardiness. This Satsuma group now includes many newer selections, like Hashimoto, Clausellina, Miho, Owari, Miyagawa, and Okitsu.
- The Mediterranean mandarin (*C. deliciosa* Tenore) is also known as Willowleaf (because of its drooping branches and small narrow leaf characteristics) and includes Avana clones Apireno and Tardivo.
- King mandarins (*C. nobilis* Loureiro) are a small and relatively minor group like the King/King of Siam and the Kunenbo. They are mainly grown in South-east Asia and are extremely late maturing. There is only minor world interest in King hybrids like Encore, Kinnow, and Wilking cultivars.
- Small-fruited mandarins remain old cultivars mainly in China and Asia, where they are sold on local markets, e.g., Bendizau, Nanfengmiju, and Hungchien. Some are also used as specialist citrus rootstocks, i.e., Cleopatra mandarin (*C. reshni*) from India and Sunki mandarin (*C. sunki*) from China.

Origin

The main common mandarins are thought to be native to northeastern India and southern China and were grown there in very early times (twelfth century BC). The clementine may have originated in Algeria, as one of the first cultivars grown in the 1890s bore this name. Other distinctive groups (satsuma types) are believed to have originated in Japan or China, the Mediterranean mandarins are thought to have been introduced mainly to Italy from China in 1805 (via England and Malta) and the minor King group appears to have originated in Indo-China.

Some current cultivars are very old and originated in ancient cultures or growing regions. However, a large number of current commercially important mandarin are of much more recent origin, especially in the 1900s through natural mutations or sports, chance seedlings, or the result of specific breeding programs and irradiation treatment of seeds or budwood, e.g.:

- Ellendale originated as a seedling in about 1878 in Queensland, Australia.
- Kara is a hybrid of Owari satsuma and King mandarin in 1915, but not released to the USA industry until 1935.
- Marisol is a bud mutation found in 1970 on a Oroval clementine tree in Spain.
- Miho and Okitsu are nucellar selections derived from Miyagawa satsuma by controlled pollination in 1940 in Japan and released in 1953.
- Fallglo is a complex hybrid of Bower (clementine × Orlando) and Temple in 1962, but not released in the USA until 1987.

For the remainder of this article, the term 'mandarin' is used without qualification and should be taken to cover and include the whole mandarin group and related hybrids mentioned above.

With recent improvements in virus or other transmissible disease elimination like shoot–tip grafting *in vitro*, heat therapy, and postentry plant quarantine procedures, many more recent and modern cultivars

are now available for evaluation and local culture throughout many countries of the world.

There is also a trend for new cultivars to be protected and patented, so that the breeder can help offset the high cost of breeding, evaluating and multiplying new releases by receiving royalty payments. For example Gold Nugget in the USA, Loretina in Spain, Sunset in Australia, Robyn in South Africa, Afourer in Morocco, Cami in Italy, and Mor in Israel have all become available and patented only in the last 10 years.

Many countries throughout the world have major citrus germplasm collections – for cultivar evaluation, breeding programs and/or sources of propagation material. One of these major germplasm collections by the Institute National de Recherche Agronomique in Corsica has 716 scion cultivars recorded in 1997. These included a total of 271 (varieties, hybrids, selections, or clones) or 38% as mandarins, collected from around the world (mainly common mandarins 78 cultivars, clementine 83, satsuma 31, and tangelo 31).

Growing Regions and Conditions

Growing regions for mandarins, with a range of total heat requirements are like all citrus in general, being produced in a wide range of climatic conditions, including the humid tropics, arid subtropics, and intermediate climates. Commercial production also tends to be located in two narrow belts in the subtropics and between 20 and 40° latitude north and south of the equator.

Citrus grows best in cooler, frost-free, Mediterranean-type climates, provided that the soils are suitable, and rainfall is at least 1200 mm, distributed evenly throughout the year or supplemented by irrigation. Climate can have a significant effect on yield, fruit quality, and tree health. In particular, minimum temperatures and degrees of frost restrict commercial mandarin production within specific areas of many growing countries of the world.

However, mandarin trees are generally the most cold-resistant of all the citrus species of commercial importance, but the fruit is more susceptible to cold damage than oranges or grapefruit, owing to its small size and thin soft rind. Many cultivars are still very specific to their individual climatic requirements.

Some cultivars have been bred or selected that extend production into marginally colder subtropical climates; e.g., satsuma mandarins in Japan, China, Korea, and New Zealand are the hardiest. Clementines have a low heat requirement, and many current cultivars mature early to mid-season, but the King group is the most tender, with a high heat requirement, and tends to mature late.

At the other extreme, mandarins are amongst the most heat-resistant of citrus fruits, comparing favorably with grapefruit. For the development of good flavor, some hotter weather at the end of the growing season seems essential. However, exposed fruit can become sunburnt.

As with other yellow/orange rind citrus, mandarins grown in tropical climates are less successful commercially than other types and will not develop a normal full rind color but are usually harvested and marketed with green or only pale-colored rinds, although the fruit may be sufficiently mature to harvest and eat.

While site selection and soil management are influenced by local climate, some cultural practices also may modify harmful effects, including cold protection, soil drainage, irrigation, and windbreaks. However, these practices cannot completely overcome major differences and unfavorable climatic effects. Some of these effects include regular severe frosts, poor rind color development, variable maturity, fruit form, and pest and disease incidence, when production is under extreme or regular unfavorable production conditions.

Production Statistics

Commercial citrus production has been recorded by the FAO in over 100 different countries and throughout six regions – Africa, North America, South America, Asia, Europe (Mediterranean), and Oceania. The world's mandarin production has been gradually increasing from a 10-year average of 8745×10^3 tonnes in 1980–81 to 1988–89, to a significantly larger average total crop (allowing for annual seasonal variations, especially the common alternate cropping) in the four 1995/96–1998/99 seasons of $14\,865 \times 10^3$ tonnes, as shown in Table 1.

The largest mandarin growing areas are located in the northern hemisphere, where 90% of recorded world production occurs, and include the important growing countries like China, the Mediterranean region, with 26% (e.g., Spain, Italy, Morocco, Egypt, and Turkey), Japan, and Korea. The southern hemisphere, with only 10%, is a relatively minor growing region, as in Brazil and Argentina, with smaller plantings in Australia and South Africa. Omitted from the FAO statistics is the known and projected mandarin production in Pakistan of 547×10^3 tonnes in 2005.

Many countries like Japan, China, India, Mexico, and Asia have important local or domestic markets for their fresh mandarin fruits, overall in 1998–99

with 77% of total world production, with only 15% exported and 8% processed. However current and future export marketing will be important markets for Spain (currently the largest exporter), Morocco, Turkey, and China (Tables 1 and 2).

It is interesting to note that in the two largest citrus-producing countries (Brazil and the USA), mandarins are a relatively minor type produced with only 4 and 3%, respectively, of their total production. In other countries, mandarins are significantly more important, and these are discussed below.

Table 1 World mandarin production trends

Year	*Total production* $\times 10^3$ *tonnes*	*Exported* $\times 10^3$ *tonnes*	*Processed* $\times 10^3$ *tonnes*
1980–81 to 1988–89 average	8745	1236	1055
1995–96	14350	2052	876
1996–97	14133	2205	1256
1997–98	16516	2406	1508
1998–99	14463	2191	1206
Four-year average	14865 (82%)	2214 (12%)	1212 (6%)

Source: FAO (1999) *Citrus Fruit Fresh and Processed, Annual Statistics, Tables 1–10 and 16–20*. Rome: Food and Agriculture Organization; USDA (1999) World Fresh Citrus Situation, pp. 18–32. In: *World Horticultural Trade and US Export Opportunities, Circular Series FHORTH, Foreign Agricultural Service*.

China

Plantings have expanded rapidly in China (Figure 1) over the past three decades, and current nonbearing planting will increase production further. China is now the largest producer of mandarins in the world (totaling 68% of its total citrus production). Newer plantings are mainly satsuma cultivars (with over 40% including Wenzhou mikan), replacing older and previously more popular Tankan, Ponkan, Hongju, Zhuju, and small fruit local cultivars Nanfengmiju and Bendizau. Over 90% of mandarin production is consumed in the huge domestic market as fresh fruit (sometime ex-local storage), but with an increasing 4% processed (canned segments) and 3% exported. These exports are expected to increase in the long term, especially to Pacific Rim countries.

Spain

Spain is the largest citrus producer in the Mediterranean region, and the fourth largest in the world, and for mandarins, it is the second largest producer (after

Table 2 Fresh mandarin crop production and utilization (with respective percentages), and the main producing countries throughout the world in 1998–99

Region/country	*Total production* $\times 10^3$ *tonnes (%)*	*Exports* $\times 10^3$ *tonnes (%)*	*Processed* $\times 10^3$ *tonnes (%)*
Northern hemisphere			
USA	401	15	175
Mediterranean region			
Greece	85	25	3
Italy	443	26	83
Spain	1749	1102	296
Israel	73	30	18
Algeria	111		
Morocco	425	222	10
Tunisia	38		
Egypt	360	10	5
Turkey	372	137	20
Subtotal (Mediterranean)	3827 (26%)	1632	430
Japan	1439	5	127
Cuba	6		
Mexico	190	2	19
China	6022	190	250
Korea	508	8	5
Subtotal (northern)	13031 (90%)	2061	1006
Southern hemisphere			
Argentina	340	30	30
Brazil	762	8	170
Uruguay	77		
Subtotal (southern)	1431 (10%)	130	200
Total (world)	14463 (77%)	2191 (15%)	1206 (8%)

Data from: FAO (1999) *Citrus Fruit Fresh and Processed, Annual Statistics, Tables 1–10 and 16–20*. Rome: Food and Agriculture Organization; USDA (1999) World Fresh Citrus Situation, pp. 18–32. In: *World Horticultural Trade and US Export Opportunities, Circular Series FHORTH, Foreign Agricultural Service*.

Figure 1 Mandarin plantings in China on terraced hillsides.

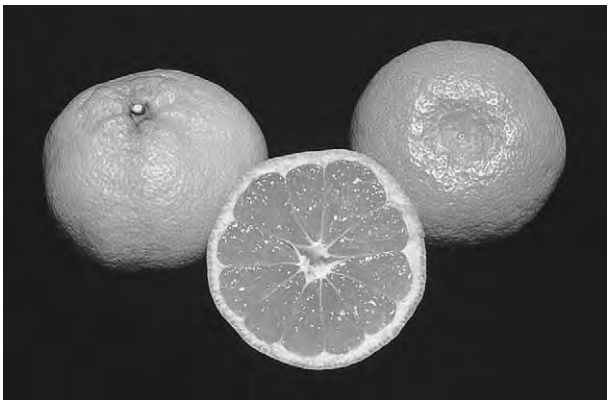

Figure 2 Okitsu satsuma has become an important new cultivar in many countries.

China), but the largest exporter of fresh fruit. Plantings in Spain continue to expand (totaling 270 000 ha in 1995). Main citrus types are oranges 55%, mandarins 31%, and lemons 13% of these plantings. Of the 21.7 million trees planted between 1990 and 1995, the main mandarin cultivar demand trends (in thousands of trees planted) were:

Satsuma	Okitsu	1575
	Owari	219
	Clausellina	131
Clementine	Marisol	1925
	Clemenules	1900
	Hernandina	991
Mandarins	Nova	1391
	Fortune	1122
	Ellendale	443.

See Figures 2 and 3.

Japan

Japan is currently third in world mandarin production, although it has experienced a significant decline in recent decades. Mandarins total 92% of Japanese citrus production, with only minor plantings of other citrus types. Most of this production is now in the form of satsuma cultivars (unshiu or mikan) and has generally replaced other older cultivars like Natsumikan, Hassaku, and Iyokan. As the satsuma group has evolved, a large number of distinct new cultivars have emerged and become important initially locally, but more recently also throughout the world. A large range of different satsuma cultivars are now available, varying from very early (Goko Wase) to midseason and late-maturity periods, i.e., Hashimoto, Miyagawa, Okitsu (Figure 2) and Miho (Wase); Hayashi, Owari, Nankan No. 20, Seto (mid); Aoshim, Juman and Otsu No. 4 (late). This enables satsuma fruit to be harvested over about 3 months, although culture in plastic/heated tunnels will advance the season by up to 4 months. Coupled with

Figure 3 (see color plate 32) Harvesting Fortune mandarins for export in Spain.

specialized storage procedures, this means that local satsuma fruit can be available for up to 8 months of the year.

Harvesting and Marketing

The growing, harvesting, and marketing of mandarins as soft fruit present many problems, which can be best overcome by specialist growers and carefully selected growing localities. They are comparatively higher in all these costs, when compared with other types of citrus production. Present day mandarin production focuses on careful type and cultivar selection to best suit local climatic conditions and cultural practices, in order to achieve a successful product and outcome.

Mandarins are mature and are ready for harvesting for a much shorter period of time than most other citrus types. Fruit left on the tree for more than a few weeks quickly pass the stage of optimum maturity and marketability, usually becoming insipid (reduced juice acidity), puffy, soft, and dry.

Early in the season, some cultivars may be internally mature before the rind color has fully developed, like Marisol clementine. Most respond to selective initial harvesting for larger fruit sizes or the required rind color development. All mandarins require careful harvesting, transport, and grading for size in order to reduce rind blemish and damage. Some cultivars are more difficult to harvest and require clipping to reduce plugging, where a small piece of rind is left attached to the stalk that remains on the tree.

Unlike many other citrus fruits, which may be available the whole year, mandarins are normally available as fresh fruit only in late autumn, winter, and spring. Within this period, a range of modern cultivars are now available for the main commercial types (common, clementine, and satsuma) that permit the production and marketing of fresh fruit from very early, early, mid (season), late, and very late maturity periods. See **Figure 4** for an example of the maturity range and sequence for some important cultivars in the Mediterranean region. All selection and breeding programs have a basic need to develop new cultivars that fit into a particular maturity period, when no cultivar is currently available or some major problem exists with the current cultivar (alternate cropper, poor fruit size, pest- or disease-susceptible).

Apart from the normal marketing periods, all mandarins are unsuitable for long-term cold storage. They are more perishable than sweet oranges or grapefruit and very susceptible to common storage breakdown problems and chilling injury. In normal export trade, careful storage is often necessary for several weeks depending on the shipping distance between producing countries, such as southern hemisphere producers like Argentina, Southern Africa, and Australia, exporting to the northern hemisphere's major markets in Europe, South-east Asia, Japan, and North America.

Crop Utilization

Mandarins are the second most important citrus type grown through the world with 17% of total production in 1998–99, after oranges (with 66%) but ahead of lemons and limes (11%), and grapefruit (6%).

As with all citrus fruits, mandarins are principally marketed according to how they are consumed, with domestic or local fresh fruit and exports totaling 92% of world production, with only a minor 8% processed – compared with 34% of oranges processed in 1998–99 (**Figure 5**).

In the world mandarin fresh export fruit trade in 1997–89, the following regions/countries (as generally noncitrus-producing) were the major importers – Mediterranean, Germany, France, UK, the former USSR, and The Netherlands.

As fresh or dessert fruits, mandarins are usually eaten out of hand, as they are generally easier to peel and separate into segments (than other citrus types) and have a distinctive, pleasant flavor. This is why mandarins are also known and promoted by a range of other descriptive terms, i.e., easy peelers, zipper-skin, TV fruit, soft fruit, kid glove citrus, picnic, or lunch box fruit. In some cultures, the presentation and eating of citrus, for example during the Chinese New Year, are considered important. Fresh mandarin segments are also suitable for garnishing and decorating salads or topping desserts.

In the home, mandarins can be widely used for juicing; cooking (as with mandarin bran muffins or mandarin-glazed spare ribs), sweets (like mandarin trifle and chocolate coating), marmalade, drying, glace, or crystallizing.

Mandarins are used as commercial processed products, mainly as segments in syrup or natural juice, with only limited production of juice products (often because of the limonene bitterness factors) and rind oil or essence, as often used in aromatherapy. After the juice has been extracted, there remain residues that can be a source material from which over 300 valuable byproducts can be produced. Some specialized types are also candied (the peel), dried, or used for marmalade and pectin manufacture.

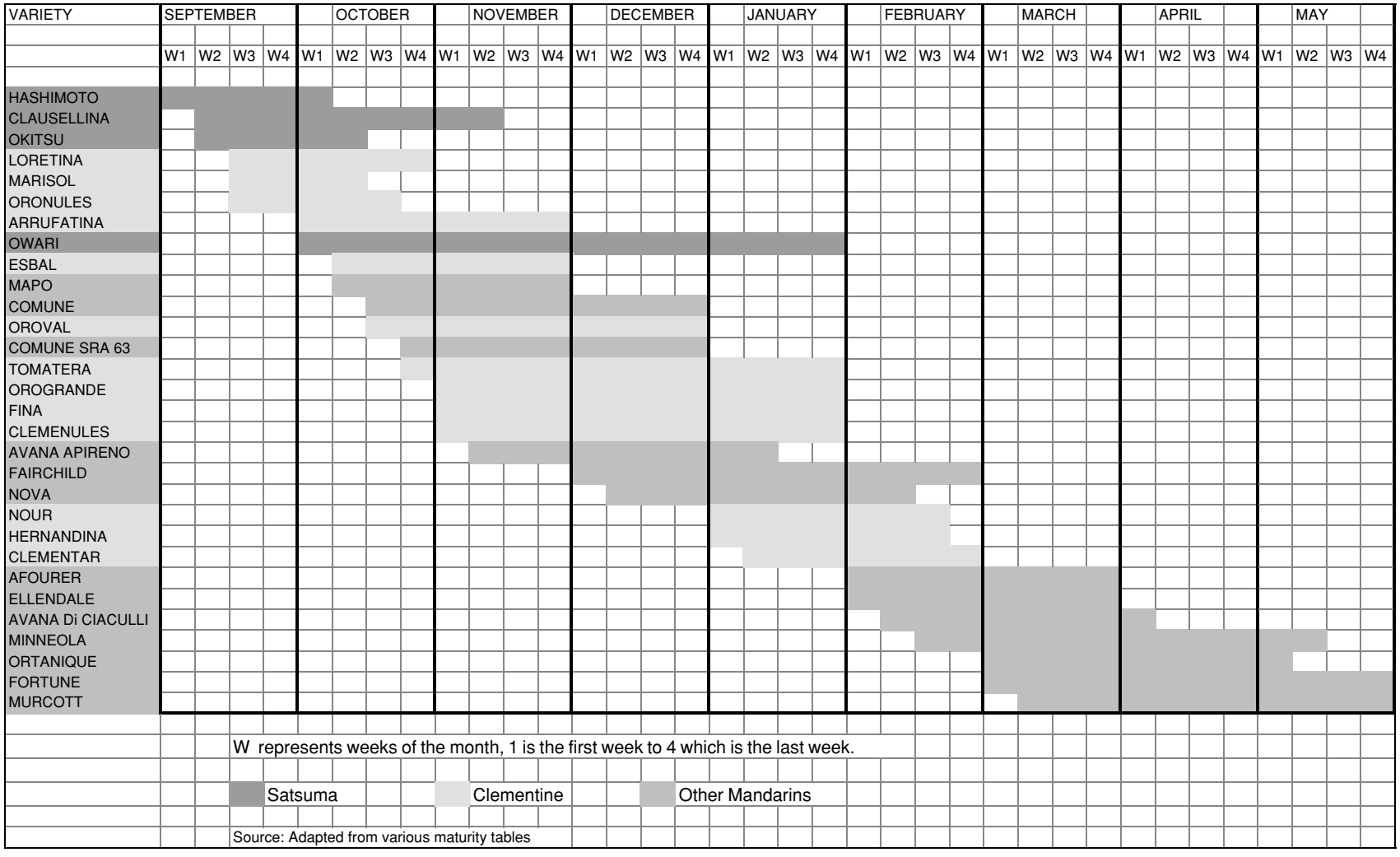

Figure 4 Some typical mandarin maturity periods in the Mediterranean region.

Figure 5 (see color plate 33) Moroccan clementine exports in open 10-kg packages.

(*See* **Citrus Fruits**; Composition and Characterization; Processed and Derived Products of Oranges.)

Over 400 different constituents have been isolated from fruit of different citrus species including mandarins. The principal components of non-acid fruits (like lemons and limes) are carbohydrates, which include mainly sucrose, glucose and fructose. These constitute about 75% of the total soluble solids (TSS) in juice. Organic acids, mainly citric and malic constitute less than 10% of the TSS and free amino acids, nitrogenous bases and glutathione about 6%. The remaining components include inorganic ions (about 3%), vitamins (2.5%), flavonoids and volatiles (1.2%) and lipids (1.2%).

See also: **Citrus Fruits**: Types on the Market; Composition and Characterization; Processed and Derived Products of Oranges

Further Reading

Cottin R (coordinator) (1997) *Citrus of the World: A Citrus Directory*. San Giuliano, France: SRA INRA-CIRAD.

FAO (1999) *Citrus Fruit Fresh and Processed, Annual Statistics*, Tables 1–10 and 16–20. Rome: Food and Agriculture Organization.

Reuther W, Webber HJ and Batchelor LD (eds) (1967) *The Citrus Industry*, vol. 1. Berkeley, CA: University of California.

Saunt J (2000) *Citrus Varieties of the World an Illustrated Guide*. Norwich, UK: Sinclair International Limited.

Sinclair WB (1984) *The Biochemistry and Physiology of the Lemon and other Citrus Fruits*, pp. 711–718. Berkeley, CA: University of California.

Swingle WT and Reece PC (1967) The botany of *Citrus* and its wild relatives. In: Reuther W, Webber HJ and Batchelor LD (eds) *The Citrus Industry* vol. 1. Berkeley, CA: University of California.

USDA (1999) World fresh citrus situation. In: *World Horticultural Trade and US Export Opportunities, Circular Series FHORTH*, pp. 18–32. USA: Foreign Agricultural Service.

Young RH (1986) Fresh *Citrus* Cultivars. In: Wardowski WF, Nagy S and Grierson W (eds) *Fresh Citrus Fruits*, pp. 110–125. Westpoint, CT: AVI Publishing.

Clams *See* **Shellfish**: Characteristics of Crustacea; Commercially Important Crustacea; Characteristics of Molluscs; Commercially Important Molluscs; Contamination and Spoilage of Molluscs and Crustaceans; Aquaculture of Commercially Important Molluscs and Crustaceans

CLEANING PROCEDURES IN THE FACTORY

Contents

Types of Detergent

H M J van Eijk and F A Majoor, Unilever Research Vlaardingen, Vlaardingen, The Netherlands

Background

A wide variety of chemicals can be used in the cleaning process, and these are generally classified as:

- water, which acts as the transport medium for chemicals and soil as well as delivering thermal and mechanical energy;
- cleaning agents, chemicals whose function is to remove soil.

Water

Water in general makes up the great majority of the cleaning solution (> 90%), and must therefore be given careful consideration. In most cases, the water used for making up the cleaning solution is also used for intermediate and final rinses and must therefore be of suitable microbiological quality. Particular attention should be also given to the content of inorganic material. If the hardness of the water is high (> 60 mg of Ca^{2+} per liter), it might be less suitable for cleaning purposes, and partial softening should be considered.

Cleaning Agents

Although water is important during cleaning, it cannot always remove all soils, particularly those encountered in the food industry, and therefore, detergents are required. Since no one detergent or combination of detergents possesses all the properties required for the removal of all types of soil at minimum cost, a selection must be made of the most appropriate for the application being considered. Materials used must be acceptable to regulatory agencies, especially with regard to biodegradability and the effect on the environment when discharged into surface waters. Typical components of a detergent formulation will include some of the following:

- alkalinity;
- surfactant/antifoam;
- sequestrant;
- threshold agents;
- hydrotropes;
- specialties;
- acids.

Many cleaning formulations are currently being reviewed in light of environmental concerns. It is probable that there will be an increasing use of low-phosphate, completely biodegradable products and a reduced use of sequestrants such as ethylene diaminetetraacetic acid (EDTA). Such changes in cleaning chemicals may well affect the other cleaning parameters such as the time and temperatures used.

Alkalis

Alkaline compounds are effective for the dissolution of proteins and removal of fats. Strong alkalis result in hydrolysis of the protein, forming water-soluble compounds. Examples of alkalis are sodium hydroxide (caustic soda) and potassium hydroxide (caustic potash). These compounds, being hazardous to personnel, are more widely used in cleaning in-place than in open-plant cleaners and, where used, should be in combination with automatic dosing equipment.

Further strong alkalis are:

- sodium orthosilicate;
- sodium metasilicate.

Milder alkalis include:

- sodium triphosphate;
- triethanol amine;
- sodium carbonate.

The milder alkalis possess some degree of protein hydrolysis and fat-removing capabilities but mainly

act as builders or sequestrants to soften water. In this respect, alkalis act to suspend and disperse the soil deposit in solution.

Surfactants/Wetting Agents

Surfactants are chemicals that lower the surface tension of water, and thus aid the 'wetting' of surfaces. For this reason, they are commonly called wetting agents. They also aid the penetration of the cleaning solution into the deposit.

Surfactant molecules consist of a hydrophobic (water hating) tail and a hydrophillic (water loving) head, as shown in Figure 1. The hydrophobic part of the molecule has an affinity for fats and oils, whereas the hydrophilic part of the molecule has an affinity for water. These molecules therefore emulsify and disperse oils, fats, waxes, and pigments. Some will precipitate at high pHs (> 12) and are unsuitable for use in combination with strong alkalis.

The polar nature of water molecules creates a high surface tension, and as a result, water forms droplets on a surface (Figure 2a). When a surfactant is added, the hydrophillic heads are attached by the water at a water/air or water/substrate interface (Figure 2b). The orientation of the surfactant molecules at the interface lowers the surface tension and results in spreading of the water droplet (Figure 2c). In a formulated detergent, these effects of the surfactant result in better penetration of that detergent into the soil deposit.

At a water/substrate interface, the same orientation of the surfactant molecules also occurs. This allows surfactants to solubilize or emulsify fats. On contact with a fatty soil, the hydrophobic ends of the surfactant enter the fat. As more such hydrophobic ends enter the fatty soil, the soil comes to occupy the center of a micelle molecular array, resulting in solubilization of the fat (Figure 3).

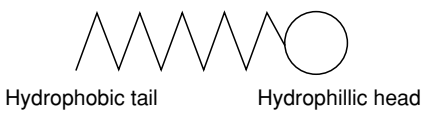

Figure 1 Surfactant molecule.

Surfactants are classified into three groups based on the charge of the active part of the molecule (in solution).

Anionic Surfactants

The active part of an anionic surfactant molecule carries a negative charge. Included in this group are the fatty acid salts and the sulfated and sulfonated derivatives of fatty acids. Anionic surfactants act in the same way as soaps, except that they are less affected by hard water ions and are generally high foaming.

Cationic Surfactants

The active part of a cationic surfactant molecule carries a positive charge. Included in the cationic surfactants are the quaternary ammonium compounds (quats). These compounds have limited detergent characteristics but have found an effective use as disinfectants.

Nonionic Surfactants

The nonionic surfactants do not ionize, and therefore, no charged species are formed. A wide range of nonionics exist, and consequently, these surfactants have a wide range of chemical properties. Nonionics can be low foaming and very penetrative wetting agents, making them particularly effective at removing oils.

Sequestrants

Sequestrants serve to enhance the effectiveness of the detergent by forming chelate complexes with metal ions in the deposit, or in the water itself. The metal ions mostly encountered are calcium and magnesium. If these ions are not complexed, they will readily form deposits on surfaces and will make the removal of fats more difficult.

Sequestrants commonly used are:

- phosphates;
- EDTA;
- nitrilotriacetic acid (NTA);
- gluconate.

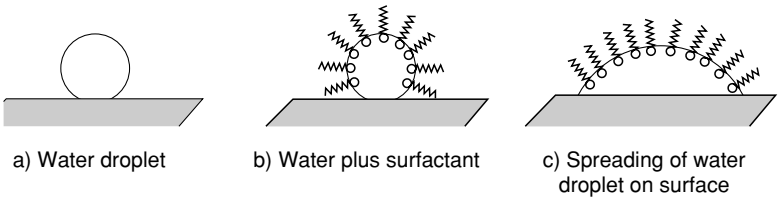

Figure 2 Wetting effect of surfactants.

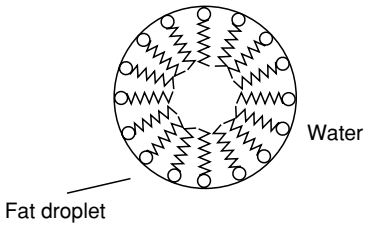

Figure 3 Solubilization of fat in water.

The sequestering power of EDTA is sufficiently high that even calcium salts can be redissolved.

Threshold Agents

Threshold agents prevent the deposition of hard water salts by disrupting the (calcium carbonate) crystal growth. These agents are added to detergents, not to prevent the deposition of hard water salts during cleaning (this is done by the sequestrant) but during the rinse stage. Threshold agents are used at very low concentrations (5–20 mg l^{-1}), which is sufficient to prevent deposition of hard water salts.

Hydrotropes

Hydrotropes are used to stabilize the neat product to ensure an adequate storage life. Liquid detergents generally contain a significant number of chemically active species, and special chemicals are necessary to prevent separation or sedimentation during storage.

Specialties

Further additions to detergents consist of corrosion inhibitors, tracers, and dyes.

Acids

Acids, both organic and inorganic, are commonly used for the removal of mineral deposits such as hard water scale or milkstone. Examples of inorganic acids are phosphoric, nitric, and sulfamic acid, and examples of organic acids are gluconic, hydroxyacetic, citric, and formic acid.

Whilst extremely effective at solubilizing mineral salts, acids are potentially corrosive to construction materials and must be used with care. Hydrochloric acid and fluoric acid are too corrosive (even to stainless steel), and they should not be used for cleaning in the food industry. A summary of the main functions of the chemicals is given in Table 1.

Table 1 Main functions of chemicals in detergents

Type	*Chemical*	*Function*
Strong alkalis	Sodium hydroxide	Removal of fat and protein
	Potassium hydroxide	Removal of fat and protein
	Orthosilicate	Emulsification
	Metasilicate	Corrosion inhibition
Mild alkalis	Sodium triphosphate	Dispersion
	Sodium carbonate	Alkalinity
Surfactants	Anionics	Wetting and foam
	Cationics	Corrosion inhibition
	Nonionics	Wetting and antifoam
Sequestrants	Polyphosphate	Prevention of Ca deposits
	Gluconate	Prevention of Ca deposits
	Na.NTA	Prevention of Ca deposits
	Na.EDTA	Dissolution of Ca deposits
Threshold agents	Phosphonates	Prevention of Ca deposits during rinse
	Polyacrylates	
Inorganic acids	Phosphoric acid	Removal of inorganic scale
	Nitric acid	
	Sulfamic acid	
Organic acids	Hydroxyacetic acid	Removal of inorganic scale
	Gluconic acid	
	Citric acid	
	Formic acid	

See also: **Cleaning Procedures in the Factory**: Types of Disinfectant; Overall Approach; Modern Systems; **Factory Construction**: Materials for Internal Surfaces

Further Reading

Imholte TJ (1984) *Engineering for Safety and Sanitation*. Crystal, MN: Technical Institute of Food Safety.

Karlsson CA-C (1999) *Fouling and Cleaning of Surfaces – The Influence of Surface Characteristics and Operating Conditions*. Lund, Sweden, Lund University.

Lelieveld HLM (2000) Hygienic design of factories and equipment. In: *The Microbiological Safety and Quality*, pp. 1656–1690. Gaithersburg, MD: Aspen.

Shapton DA and Shapton NF (eds) (1998) *Principles and Practices for the Safe Processing of Foods*. Cambridge, UK: Woodhead.

Wilson DI, Fryer PJ and Hasting APM (eds) (1999) *Fouling and Cleaning in Food Processing '98. Proceedings of a Conference at Jesus College, Cambridge*. Brussels: Directorate-General Science, Research and Development.

Types of Disinfectant

J Fisher, Diversey Ltd, Riddings, UK

Background

In food factories, the detergent stage is normally followed by a disinfecting stage. The detergent stage is required to remove all the soil, leaving a chemically clean surface. The disinfectant stage is used as an extra guarantee of cleanliness and to prevent recontamination in some cases. It does not compensate for a bad detergent stage or badly designed process or cleaning equipment.

Disinfection

The definition of disinfection taken from BS5283 (1986) states 'The destruction of microorganisms, but not usually bacterial spores. It does not necessarily kill all microorganisms but reduces them to a level acceptable for a defined purpose, for example a level which is harmful neither to health nor to the quality of perishable goods.' The acceptable level of microbial contamination on a surface or piece of equipment has to be determined; obviously, no pathogens should be found. (*See* **Spoilage:** Bacterial Spoilage; Molds in Spoilage; Yeasts in Spoilage.)

The state of sterility is defined as free from all living microorganisms. This is not achievable in the food factory by using acceptable chemicals ('acceptable' meaning safe for humans, plant materials and products).

Disinfectants are used after the detergent application in cleaning-in-place (CIP) operations where the term 'terminal sterilant' may be used. They are also used after hand cleaning. Equipment should be left in a soak bath until it is ready to be used, thus ensuring that it remains free from recontamination.

There are a wide range of disinfectants available. The choice of disinfectant depends on the user's requirements, the type of processing and cleaning equipment, the method of use, and, to some extent, the personal preference of the user.

Types of Disinfectant

Disinfectants can be split into two broad groups, oxidizing and nonoxidizing. Oxidizing disinfectants include the halogens, chlorine, iodine, bromine, and chlorine dioxide, and oxygen-releasing materials such as peracetic acid and hydrogen peroxide. Nonoxidizing disinfectants are as follows: quaternary ammonium compounds, amphoterics, biguanides, and acid anionics.

Physical and Chemical Properties

Oxidizing Disinfectants

Halogens Chlorine and iodine have been used as terminal disinfectants for many years. More recently, bromine and chlorine dioxide have been introduced.

Chlorine Chlorine was first used as a gas for fumigation in hospitals in 1791, but this application has one obvious drawback – chlorine gas is toxic. Active chlorine is available from two types of material:

1. Inorganic compounds containing hypochlorite ions either as a liquid, e.g., sodium hypochlorite (NaOCl), or as a powder, e.g., chlorinated trisodium phosphate
$$((Na_3PO_4.11H_2O)_4NaOCl, NaCl).$$
2. Powdered organic chlorine release agents, e.g., trichloroisocyanurate (Figure 1).

In solution, both types hydrolyze to produce hypochlorous acid and/or hypochlorite ions, depending on the pH.

Acid				Alkaline
Cl_2	$\rightleftharpoons$	HOCl	$\rightleftharpoons$	OCl^-
Chlorine gas		Hypochlorous acid		Hypochlorite ion

In the food industry, sodium hypochlorite is used as a general-purpose disinfectant. It is most stable in a slightly alkaline solution, and it is for this reason that the concentrate is supplied stabilized with sodium hydrocide at a pH of up to 12. An in-use solution of between 50 and 300 p.p.m. will have a pH between 7 and 9. The optimum pH for disinfection is pH 5.0, but

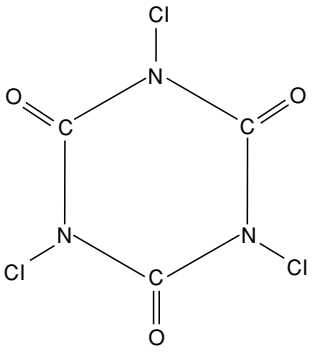

Figure 1 Trichloroisocyanuric acid. Reproduced from Cleaning Procedures in the Factory: Types of Disinfectant, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

the solution is less stable. Below pH 5.0, chlorine gas will be produced.

Applications for sodium hypochlorite in the food industries are CIP, soak, and spray. Sodium hypochlorite has many advantages: it is nonfoaming; it is not affected by water hardness; it does not leave an active residue, and it has a wide antimicrobial spectrum, which includes activity against bacterial spores and viruses. It is also fast-acting and cheap.

However, it also has numerous disadvantages: it can be corrosive to a wide range of components, including stainless steel; it is irritating to the skin and eyes; the in-use solution is unstable; it is inactivated by organic materials, and it may give rise to taint problems.

Chlorine dioxide Chlorine dioxide (ClO_2) is an unstable and toxic gas that is soluble in water. When chlorine dioxide is generated in solution, as shown below, it is a very effective water disinfectant at point of use.

$$5NaClO_2 + 4HCl \rightleftharpoons 4ClO_2 + 5NaCl + 2H_2O.$$

Chlorine dioxide at use concentrations (0.5–1 p.p.m.) overcomes some of the disadvantages of hypochlorite in that it is nontainting, noncorrosive, and nontoxic. Its sole use at present is in water disinfection.

Iodine Iodine itself is not very soluble in water, and the vapor is irritating to the eyes, making it difficult to handle. Iodine is a very reactive element, and it is this reactivity that makes it a good disinfectant.

Iodine compounds used in the food industry contain iodine complexed with polyvinylpyrrolidone and other surface active agents, usually in an acid solution. These are known as iodophors and were first introduced in 1949.

The complexes formed between iodine and carrier molecule are water-soluble and overcome the handling difficulties of iodine whilst retaining the disinfecting power. On dilution, the iodophors release iodine gradually, and it is the free iodine that acts as the disinfecting agent. The optimum pH for microbial activity is pH 5.0.

$$\underset{\substack{\text{Greatest}\\\text{activity}}}{\overset{\text{Acid}}{I_2}} \rightleftharpoons \underset{\substack{\text{Some}\\\text{activity}}}{HOI} \rightleftharpoons \underset{\text{Inactive}}{OI^-} \rightleftharpoons \underset{\text{Inactive}}{\overset{\text{Alkaline}}{IO^-}}.$$

The surface-active agents provide better wetting and organic soil penetration, thus making iodophors less affected by soil than hypochlorite. The choice of surface-active agent may lead to foam generation in applications such as CIP.

Iodophors have a broad antimicrobial spectrum that is similar to hypochlorite, although they are less active against bacterial spores. In common with sodium hypochlorite, they are fast-acting but are more expensive. Iodophors are used in soak baths and spray application at up to 10 p.p.m. available iodine. In solution, iodophors are yellow-brown in color. This color can be an advantage: in a soak bath application, the color indicates the presence of iodine; the in-use solutions are unstable, so that as the iodine dissipates, the solution will become colorless.

Staining may be a problem, especially with some plastics, and this may also result in taint problems. Iodophors can be corrosive; it is therefore necessary to ensure that the correct dilution is used; otherwise, damage to plant and personnel may occur.

Bromine Bromine itself is not used as a disinfectant, mainly because of its handling difficulties. Bromochlorodimethylhydantoin is supplied as a powder or a solid. In solution, it releases hypobromous and hypochlorous acids.

Oxygen-releasing compounds

Peracetic acid Peracetic acid was introduced in 1955. The material is supplied as an equilibrium mixture:

$$\underset{\text{Peracetic acid}}{CH_3C(=O)OOH} + \underset{\text{Water}}{H_2O} \rightleftharpoons \underset{\text{Acetic acid}}{CH_3C(=O)OH} + \underset{\text{Hydrogen peroxide}}{H_2O_2}.$$

It is soluble in water and is completely biodegradable, breaking down to harmless products:

$$2CH_3C(=O)OOH \rightarrow 2CH_3C(=O)OH + O_2$$

As supplied, peracetic acid is corrosive and has a very irritating smell, similar to vinegar; because of these properties, it is unpleasant to handle, and manual use is not recommended. It is suitable for CIP, as it is nonfoaming.

Peracetic acid is a highly reactive material. As an in-use solution, it is not very stable and will react with organic materials. Peracetic acid may attack plant materials, such as rubber gaskets, and at higher concentrations, corrosion may be a problem.

Peracetic acid has a wide antimicrobial spectrum, which includes bacterial spores and viruses. This activity is fast and is maintained at temperatures lower than ambient.

Hydrogen peroxide Hydrogen peroxide (H_2O_2) was introduced as a disinfectant in 1887. It is supplied in solution, which has a tendency to decompose:

$$2H_2O_2 \rightarrow 2H_2O + O_2.$$

Manual use of hydrogen peroxide is not recommended, but it is used in spray applications such as aseptic packaging. Hydrogen peroxide is both bactericidal and fungicidal. Some bacteria and fungi are less sensitive because of catalase activity, which destroys H_2O_2. Hydrogen peroxide is slow-acting, so that a long contact time or elevated temperature is required for effective disinfection.

Nonoxidizing Disinfectants

Quaternary ammonium compounds Quaternary ammonium compounds (QACs) were first introduced in 1917 and are probably the best known cationic surface-active agents. Their general formula is as follows:

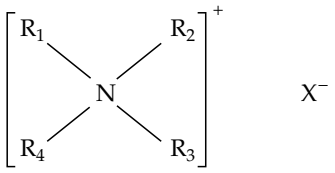

X is usually a halide but sometimes a sulfate ion. R_1, R_2, R_3, and R_4 may be a variety of alkyl or aryl groups.

QACs are generally poor detergents but good wetting agents. In solution, they ionize to produce a cation, the substituted nitrogen part of the molecule, which provides the surface-active property. The length of the carbon chain in the R groups affects the disinfectant ability; usually, C_8 to C_{18} are the most effective.

The surface-active nature of these molecules tends to make them too high-foaming for CIP use, but they can be used for soak and manual cleaning at 200–400 p.p.m. active. The optimum activity is around neutral pH, but QACs are active between pH 3.0 and 10.0. Activity may be inhibited by water hardness.

QACs are noncorrosive and are stable at in-use dilution. Their major disadvantages are that they are affected by organic soil and that they tend to cling to surfaces, so that they may be difficult to rinse off, resulting in possible taint problems.

The antimicrobial range of QACs is less than that of the oxidizing disinfectants. They are less effective against Gram-negative bacteria than against Gram-positive bacteria. They also have limited activity against bacterial spores and very little activity against viruses. To be effective against yeasts and molds, a higher concentration is required.

Biguanides Biguanides with antimicrobial activity were first reported in 1933. The biguanides are derivatives of guanidine, a naturally occurring substance found in vegetables such as turnips and cereals:

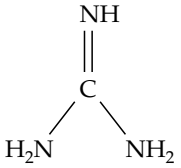

Biguanides are usually supplied as polymers in the salt form, mostly as the hydrochloride. Optimum activity lies between pH 3.0 and pH 9.0. Below pH 3.0, activity is suppressed, whilst above pH 9.0, they are precipitated.

They are cationic in nature but are not regarded as surface-active. Biguanides do not foam and are, therefore, suitable for CIP; they may also be used for soak and manual cleaning. They are noncorrosive but taint may be a problem if not properly rinsed. The in-use solution is stable but is affected by organic soil and, to some extent, by hard water.

Most biguanides have equal antibacterial activity against Gram-positive and Gram-negative microorganisms. They are less effective against molds and yeasts, and are ineffective against bacterial spores and viruses.

Amphoterics Amphoterics have been in use as disinfectants since the early 1950s. They are based on a substituted amino acid, usually glycerine. The term ampholyte is often used to describe them because in solution they ionize to produce cations, anions, or zwitterions, depending on the pH:

$$R-\overset{+}{N}H_2-CH_2-CH_2-C(=O)OH$$
Acid cation

$$R-\overset{+}{N}H_2-CH_2-CH_2-C(=O)O^-$$
Zwitterion

$$R-NH-CH_2-CH_2-C(=O)O^-$$
Alkaline anion.

Only certain amphoterics have a disinfecting ability and surface activity. The disinfecting ability appears to increase with the increase of basic groups.

Amphoterics tend to be viscous liquids that are freely soluble in water. They are generally too high-foaming for use in CIP, but are suitable for soak, spray, and hand use. Amphoterics are equally effective against Gram-negative and Gram-positive bacteria; they are less effective against yeasts and molds, and have very little effect against bacterial spores and viruses. The optimum activity lies between pH 3.0 and 9.0.

Properties such as soil tolerance and corrosion vary with the amphoteric concerned. Corrosion is not usually a problem. The in-use solution, usually 1000 p.p.m. active, is stable.

Acid anionics The active molecule in acid anionics varies considerably. There are two main types: those based on carboxylic acids, which include fatty acids and derivatives, and those based on anionic surfactants combined with mineral acid.

Acid anionics tend to be formulated products with additions to aid activity or solubility. Properties will vary with product, but they tend to have some detergent and wetting ability. The higher-foaming products are unsuitable for CIP, so that their general use is for spray. They are not suitable for hand use, since a pH of 2 is required for optimum antimicrobial activity.

The antimicrobial activity is against Gram-negative and Gram-positive bacteria, but they are less effective against bacterial spores and viruses. Certain carboxylic acid types are active against yeasts and molds. Both types are affected by organic soil and water hardness, but again, both properties will vary with the product. The in-use solutions are stable.

Effluent Problems

The oxidizing disinfectants are degraded very easily by organic soil to ineffective products. Peracetic acid and hydrogen peroxide break down to the products that have been described earlier.

The breakdown products from the halogens vary with pH of the effluent, but in general, halide ions will be produced. The halogens should not be mixed with acid products, as chlorine will react with organic chemicals to produce organo-chloro compounds, which may be carcinogenic. The nonoxidizing disinfectants in modern products tend to consist of biodegradable compounds. The cationic products will adsorb on to organic material. The biguanides are incompatible with alkaline chemicals and will form a precipitate. (*See* **Effluents from Food Processing**: On-Site Processing of Waste; Disposal of Waste Water; Composition and Analysis.)

Analysis of Disinfectants in Waste Water

Obviously, detection will tend to depend on the concentration of disinfectant present. The oxidizing disinfectants are unlikely to be detected. Halide ions can be detected but cannot be identified as coming from the disinfectant.

Using an available chlorine probe, chlorine may be detected up to 200 p.p.m., but because of the presence of organic material and other chemicals, the presence of available chlorine may not be detected. As with chlorine, available iodine can be detected using a probe, but iodine is converted very quickly to iodide. The breakdown products of peracetic acid and hydrogen peroxide are unlikely to be detected.

For the quaternary ammonium compounds, the biguanides and the amphoterics, it would be necessary to know the specific active molecule to be able to quantify these activities in effluent. The active content could then be determined by HPLC.

Acid anionics can be detected in the effluent by determining the anionic content.

Comparison with Steam

There is no chemical suitable for use as a disinfectant in the food factory that can compete with steam. It is effective against bacteria, molds, yeasts, bacterial spores, and viruses. It is not affected by soil and hard water. There are no corrosion or stability problems, and it leaves no residues. The drawbacks are that it cannot be used with heat-sensitive plant materials, and it needs careful use to avoid human contact.

See also: **Effluents from Food Processing**: On-Site Processing of Waste; Disposal of Waste Water; Composition and Analysis; **Spoilage**: Bacterial Spoilage; Molds in Spoilage; Yeasts in Spoilage

Further Reading

Block S (1977) *Disinfection, Sterilisation and Preservation*. Philadelphia, PA: Lea & Febiger.

Hugo WB (1971) *Inhibition and Destruction of the Microbial Cell*. London: Academic Press.

Hugo WB (1991) A brief history of heat and chemical preservation and disinfection. *Journal of Applied Bacteriology* 71: 9–18.

Sykes G (1972) *Disinfection and Sterilisation (ICMSF) International Commission on Microbiology Specification for Foods*. London: Chapman & Hall.

Overall Approach

H M J van Eijk and F A Majoor, Unilever Research Vlaardingen, Vlaardingen, The Netherlands

Background

Plant and equipment is generally cleaned for one of two reasons.

1. Microbiological
   - defined by required final product microbiological quality.
2. Process

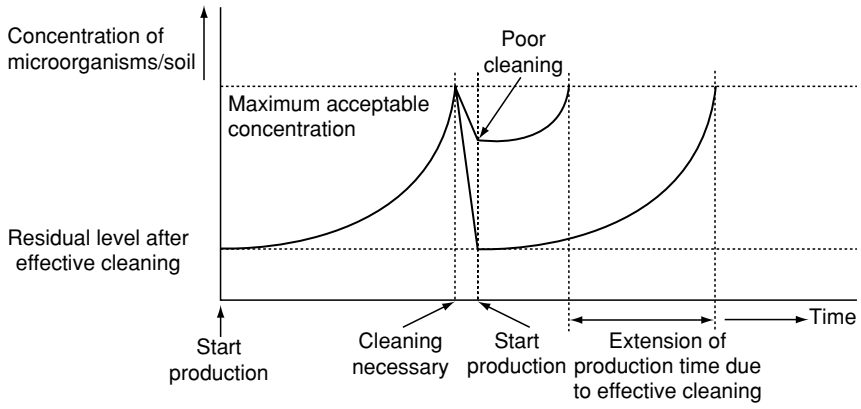

Figure 1 Influence of cleaning efficiency on available production time.

- formation of deposits that may adversely affect the operation of equipment.
- changeover to a different product;
- production shutdown over an extended period.

An example of a *microbiological requirement* for cleaning would be a food process line where the microbiological quality of the final product decreases during production until the point where shutdown and cleaning is required.

An example of a *process-related* requirement for *open* plant cleaning would be a freezer in which ice build-up on the evaporator surfaces takes place, eventually resulting in a failure to maintain the desired air temperature and flow velocity. The effect of this would be to reduce the capacity of the freezer and eventually reach a point at which the products are not cooled to the required temperature, thus requiring production to be terminated. A *process-related* requirement for cleaning of a *closed* system would be a pasteurizer, in which fouling of the heat transfer surfaces takes place, eventually resulting in both physical blockage of the heat exchanger and failure to maintain the desired pasteurizing temperature. The microbiological effect of the pasteurization process would therefore be reduced below that required for product safety, thus requiring production to be terminated.

There are a variety of mechanisms by which surfaces may become fouled, as shown in Table 1, and the nature and composition of the resultant soil can therefore also vary widely. Failure to clean effectively may have serious implications in terms of product quality and storage characteristics as well as process operation and control. Figure 1 shows how a properly cleaned process line can operate for a much longer period of time before the microbiological quality of the product becomes unacceptable, and the line must be shut down and cleaned.

Table 1 Major fouling mechanisms

1. Chemical reaction	Deposits formed at the surface by chemical reactions in which the construction material itself is not a reactant
2. Crystallization	Deposition/formation of crystals on a surface
3. Particulate	Accumulation of solid particles from the fluid stream on the surface
4. Biological	Deposition and growth of microorganisms on a surface
5. Corrosion	Corrosion of the surface itself which may in turn promote the attachment of other soil
6. Solidification	Solidification of fluid components on a surface

Clean equipment leaves microorganisms exposed to disinfection agents such as hot water, steam, and chemical disinfectants. Microbiological methods of evaluation may be used to validate the condition of the system after cleaning, and several countries have recommended hygiene standards based on colony counts per unit surface area. In addition to standard microbiological techniques (contact plates, swabs) a variety of more direct methods may be used to assess the cleanliness and/or hygienic status of surfaces.

Scope and Philosophy of Cleaning Operations

Cleaning operations within any manufacturing environment are likely to involve a variety of surfaces, materials of construction, and cleanliness requirements. Patterns of cleaning will also vary considerably according to the environment, equipment, and products. An initial classification is used to differentiate between the types of surface that are to be cleaned, product contact and nonproduct contact, which may be defined as:

Product contact surfaces These include all equipment surfaces that intentionally or unintentionally (e.g., resulting from splashing) come into contact with the product, or from which product or condensate may drain, drop, or be drawn into the main product or product container. This includes surfaces that may indirectly cross-contaminate product contact surfaces or containers.

Nonproduct contact surfaces These include all other exposed surfaces. These may include surfaces associated with equipment such as support structures, control panels, and external surfaces. They also include surfaces related to the manufacturing environment.

It is important to appreciate the importance of the manufacturing environment as a potential source of pathogens and/or spoilage organisms. Drains, drain channels, walls, floors, and roofs near air intakes have all been implicated as significant sources of microorganisms.

In certain situations, environmental cleaning can only be carried out during a production shut-down period and therefore can be difficult and costly. Because of this and the trend towards extended production runs within the food industry, intermediate cleaning is being considered more frequently. An example of this is cleaning a line, while production continues on an adjacent line in the same area.

A further subdivision of cleaning operations is whether the process is carried out dry or wet.

Dry cleaning

Dry cleaning methods are primarily used where the products are hygroscopic or where water can react to form hard deposits that are difficult to remove. The principal safety risk is that failure to control moisture can permit the growth of pathogens, e.g., *Salmonella* spp., in the food processing environment. Environments typically dry-cleaned include plants producing flour, cocoa, dry milk products, dry soup and snack mixes, and dry infant formulae.

Dry cleaning is essentially the mechanical removal of soils using sweeping, brushing, wiping, and vacuuming. Vacuuming is in principle a desirable system for removing dust and dirt. Disinfection following dry cleaning may pose difficulties, although alcohol-based disinfectants can be applied and allowed to dry off before equipment is reassembled.

Additional wet cleaning in a dry area needs great care, and it is usually best to remove equipment or parts requiring wet cleaning to a separate area. The wet-cleaned parts must be throughly dried before returning to the dry area for reassembly.

Wet Cleaning

Wet cleaning, as its name implies, involves the application of fluids, usually water based, to achieve the desired cleaning result. This can be carried out in place (CIP), where the plant remains assembled as for production.

The other alternative is to clean using open plant cleaning (OPC). In this case, the surfaces to be cleaned have to be made accessible to the cleaning fluids, which may involve the use of delivery systems to enable the operator to deliver the necessary cleaning fluids, e.g., lances, sprays, or dismantling of the equipment to allow the necessary access. In addition, some plant components may be physically removed from the production area and cleaned separately in a designated cleaning area. This is termed cleaning out of place (COP).

More complex process equipment may involve a combination of OPC and CIP; for example, a pump may be cleaned in place along with the rest of the line, but if it is known that CIP is not fully effective, it may be stripped down and cleaned manually before production restarts.

Manufacturing environments may require a combination of dry and wet cleaning. Dry cleaning using scrapers, shovels, or brushes may be more appropriate, particularly immediately after production, to remove quantities of gross soil and thus reduce the challenge to the wet cleaning process.

Factors Affecting Cleaning of Equipment

The equipment to be cleaned may be considered as a set of unit components that may vary widely in terms of volume, configuration, and geometry. There are several interrelated factors that have a direct bearing on the effectiveness and ease of cleaning process plant (Figure 2). A key point to appreciate is that a system of poor mechanical or process design may be uncleanable, irrespective of the cleaning process imposed on it. It is therefore essential that consideration of cleaning should be integrated into the mechanical and process design at an early stage rather than a final process to be incorporated into an already fully specified plant.

The factors that can contribute to the efficiency of a cleaning process are mechanical, thermal and chemical activity – and the time for these to work. The cleaning process is invariably a combination of these factors, and different cleaning processes involve a different balance of these factors, which has implications for the type of detergent to be used. For example, a cleaning process carried out at a low temperature with little mechanical activity would require a more powerful detergent than one carried

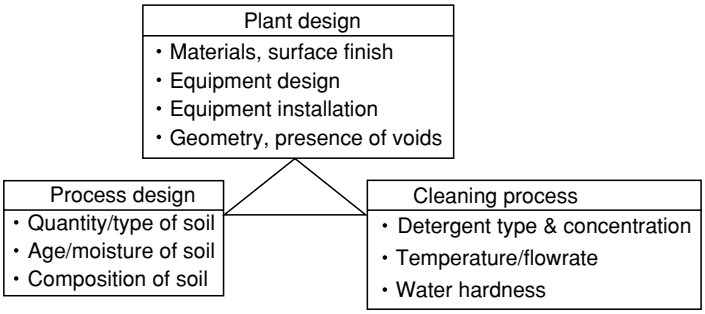

Figure 2 Factors affecting process plant cleaning.

out at a high temperature with a large amount of agitation.

Plant-cleaning Options

Open Plant Cleaning (OPC)

The function of OPC process is to clean and, where necessary, disinfect all surfaces to an acceptably high and consistently reproducible standard. A process plant cleaned in this way is disassembled and/or moved from its normal location for cleaning. This also applies to plant that is opened up to make it more accessible for cleaning manually or using some form of delivery system such as a hose.

The cleaning process consists of a number of stages:

Preparation for cleaning Prior to cleaning, all food and packaging material has to be removed from the area, including pallets, crates, bins, etc. All food material that can be reused further should be stored away from the process area and packaging material returned to the packaging material store. Only if all this material is removed, will thorough and complete cleaning be possible.

Dismantle equipment Equipment should be dismantled in order to expose the surfaces to be cleaned. Small items or components may be removed at this stage for cleaning in a dedicated wash area.

Gross soil removal Before application of the detergent, the gross soil has to be removed from the surface. Detergent has little effect on soil layers several millimeters in thickness or on large lumps.

Gross soil removal can be achieved by different methods as scraping, brushing, or spraying, but it is not necessary for the line to be 100% visibly clean after this step. It should be borne in mind that collecting gross soil by scraping may be more effective than flushing it to drain by pressure rinsing in terms of reducing the loading and cost of effluent treatment. Pressure rinsing could take more time and cause extra problems of drain blockage.

Detergent application Detergent is applied to the surfaces using mechanical aids to assist soil removal including brushes or high- or low-pressure water hoses as appropriate. The selection of detergent should be done in conjunction with the supplier.

Rinsing Rinsing of the detergent residues and soil can best be done under the same conditions as the prerinse. The extra demand for this rinse is that, after rinsing, the equipment should be 100% visibly clean. The rinse water should be of potable quality

Disinfection In many applications, food contact surfaces have to be disinfected after cleaning. The choice of the correct disinfectant depends on a number of factors, including:

- type of microorganisms involved;
- construction material of the equipment;
- effluent system;
- application of final rinse after disinfection.

Final Rinse It depends on the local situation as to whether a final rinse after disinfection is necessary. In some countries, adequate draining after disinfection with approved disinfectants at approved concentrations is sufficient. If a final rinse is required, the water must be of potable quality.

Drying Liquids must be completely removed by draining and/or drying, before the next startup can be done.

Clearing Away the Cleaning Tools After cleaning, the tools must be cleaned and (optionally) disinfected

and stored in a designated area (according to local rules). It is helpful to have color coded cleaning tools for special applications/areas (for example, black-bristled brushes for drain cleaning).

Cleaning In-place (CIP)

CIP may be defined as the circulation of chemicals and/or water through the plant that remains assembled as for production, such that all product contact surfaces are cleaned and, where necessary, disinfected to an acceptably high and consistently reproducible standard.

This process consists of up to three main stages:

1. Product recovery/rinsing – removal of bulk product from the system immediately after production, or cleaning residues such as used detergents and disinfectants after cleaning.
2. Cleaning – removal of soil from product contact surfaces.
3. Disinfection – further reduction of relevant microorganisms in the plant by chemical or thermal means, to achieve the desired hygiene standards within the plant.

Hence, CIP may be taken to mean processes including some or all of the above stages.

Product recovery should be employed whenever possible to reduce product losses and decrease the effluent chemical oxygen demand (COD). Product recovery may be achieved by simple purging with water or, with a viscous product, more effectively by pigging.

For disinfection, there are several options depending on the microbiological requirements, construction materials, and plant design:

- pasteurization, a thermal process that will inactivate most vegetative organisms, including pathogens, but not heat-resistant bacterial spores;
- sterilization, a more severe thermal process that destroys all microorganisms, including heat-resistant bacterial spores;
- chemical disinfection, which, depending on the chemical and the temperature used, may kill all or only some of the microorganisms present.

It is possible to clean a plant by applying a simple 'use and dump' procedure, whether or not making use of additional equipment for preparing and pumping cleaning liquids. However, considerable savings in running costs may be achieved by installing an appropriate CIP system, which enables recovery and reuse of rinse water and detergents and a more efficient utilization of energy.

Economic and Environmental Implications of Cleaning

Economic Implications of Cleaning

Cleaning costs are often closely associated with the cost of the cleaning chemicals alone and less attention paid to the other costs of the process, namely:

- steam;
- electricity;
- water;
- effluent;
- labor;
- plant downtime.

It is difficult to be specific as regards the cost of cleaning, as this will depend on the type of plant being cleaned, the cleaning process used, and the relative cost of services, which will vary widely from site to site. It is, however, generally accepted that in open plant cleaning, the major cost factors are plant downtime and labor. Where the plant downtime cost is not included in the costing, the labor cost may be 60–85% of the total.

Although manual cleaning is widely accepted in some industries as the only way of achieving reasonable standards of cleanliness where vessel and pipeline cleaning is required, its inherent disadvantages made the eventual implementation of CIP techniques inevitable, and their use has confirmed the significant benefits anticipated. CIP may incur additional capital costs for equipment, but this can be minimized for new lines by consideration of CIP during the design stage.

1. Improved hygiene. Consistent, reproducible cleaning schedules may be carried out and monitored with a corresponding improvement in plant cleaning performance and subsequent product quality.
2. Cost savings. Optimal use of water, chemicals, and steam results in lower cleaning costs.
3. Improved plant utilization and reduced plant downtime. Tanks and pipelines can be cleaned as soon as they are empty and refilled immediately afterwards.
4. Reduced labor cost. A reduction in the need for manual disassembly of equipment, coupled with an increased use of automation leads to a dramatic reduction in labor requirements.
5. Improved occupational safety. Less opportunity for operators to come into contact with aggressive chemicals.
6. Reduced effluent, minimizing volumes of waste water. CIP systems can be used to monitor and manage the reuse of cleaning waters and chemicals.

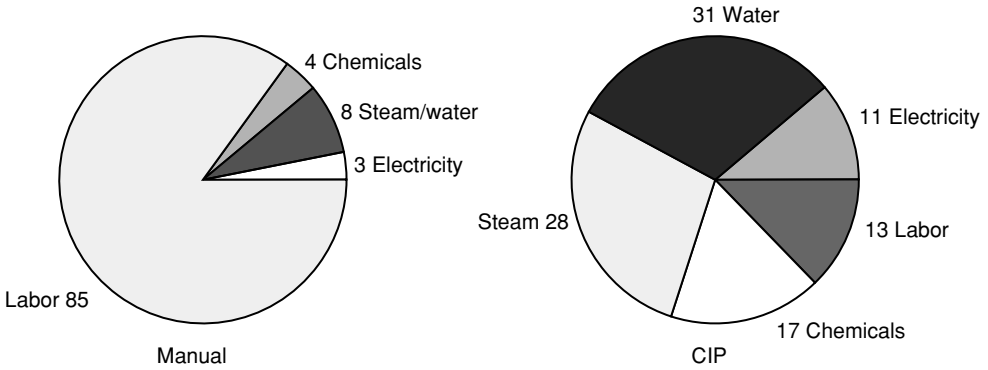

Figure 3 Typical percentage cost breakdown for manual and CIP cleaning. (Figures are percentages of total cost, excluding depreciation. Total costs of the manual and CIP operations are *not* the same.)

The cost of cleaning will depend on the plant being cleaned, the cleaning process used, and the cost of services, which will vary widely from site to site. However, the additional costs incurred with manual cleaning can be seen clearly in Figure 3, a cost breakdown of a labor-intensive manual operation and a highly automated CIP process.

The costs in Figure 3 are all shown relative to the total cost of the single stage clean, which is taken to be 1. The depreciation costs are very variable, and the total costs of manual and CIP cleaning in this figure are not the same.

With a predictive costing model, it is possible to compare the (total) cleaning costs for different cleaning procedures and thus optimize the cleaning process, which automatically results in water, effluent, and energy savings.

Environmental Implications of Cleaning

Cleaning in the food industry will inevitably bring a high amount of organic material into the effluent system. In general, more than 95% of the organic load consists of the soil, which is removed. Scraping away gross soil and sending it to destruction, instead of rinsing it to drain, could substantially reduce the organic load that goes into the effluent system.

Any environmental assessment of cleaning should be considered in the context of the complete manufacturing operation and also include all types of cleaning operations, both OPC and CIP. These are likely to indicate that the majority of the biochemical oxygen demand/chemical oxygen demand load on the effluent plant is associated with product losses, both during production, e.g., product leakage as well as cleaning. The detergents used should contain only biodegradable surface active agents and, if possible, biodegradable sequestrants. For inorganic materials, they should not contain fillers such as chlorides and sulfates, which do not have any positive effect on the cleaning. The technical information sheet of the detergent should provide information concerning the environmental effects of the detergent. In addition, automatically preparing detergent solutions has a cost-reducing effect, and instead of manual preparation, it prevents overdosing.

The most attractive way of reducing the environmental implications for closed systems is product recovery by pigging. The so called 'pig' is sent, by means of compressed air or water as the driving force, through the pipeline and expels most of the product from the system.

Conclusions

All forms of cleaning and disinfection are known to play an increasingly important role in manufacturing in all stages of processing right up to the packaging of the final product. Because there is a trend towards improved hygiene standards in food manufacturing, cleaning/disinfection forms an essential component of such improvements.

See also: **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant; Modern Systems; **Factory Construction**: Materials for Internal Surfaces

Further Reading

Imholte TJ (1984) *Engineering for Safety and Sanitation*. Crystal, MN: Technical Institute of Food Safety.

Karlsson CA-C (1999) *Fouling and Cleaning of Surfaces – The Influence of Surface Characteristics and Operating Conditions*. Lund, Sweden, Lund University.

Lelieveld HLM (2000) Hygienic design of factories and equipment. In: *The Microbiological Safety and Quality*, pp. 1656–1690. Gaithersburg, MD: Aspen.

Shapton DA and Shapton NF (eds) (1998) *Principles and Practices for the Safe Processing of Foods*. Cambridge, UK: Woodhead.

Wilson DI, Fryer PJ and Hasting APM (eds) (1999) *Fouling and Cleaning in Food Processing '98. Proceedings of a Conference at Jesus College, Cambridge*. Brussels: Directorate-General Science, Research and Development.

Modern Systems

H M J van Eijk and F A Majoor, Unilever Research Vlaardingen, Vlaardingen, The Netherlands

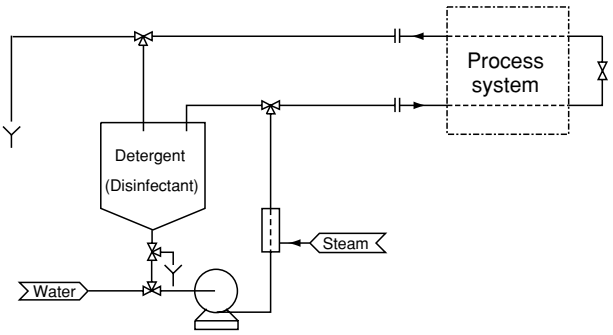

Figure 1 Flow diagram of a 'single-use' CIP system (types of valves depending on specific local requirements).

Introduction

In the design stage of a process line, the cleaning requirements should already be taken into account. A CIP-able plant should be designed in such a way that cleaning and disinfection can take place in the same line configuration as that during processing. Unfortunately, in practice, this target cannot always be reached; it is possible that some parts are not cleanable in-place and thus require dismantling and manual cleaning. Those parts have to be physically separated from the system, and it should be realized that reinstalling may result in contamination of the cleaned process line. Implementation of a CIP system depends on the type of process and whether the equipment is suitable to be cleaned in-place.

Resulting from this discussion, normally more options can be defined. If some of those solutions are equivalent, a detailed cost comparison should follow, taking into account investment in equipment, running costs, and, increasingly importantly, plant down time.

Options for central CIP systems

In some cases, CIP of a plant can be done without the addition of extra equipment. This could be appropriate for small plants with only a few tanks and short pipe runs. Cleaning liquids can be prepared in the production equipment itself, e.g., in balance tanks of pasteurizers. Possible disadvantages are that running costs may be high because of losses of cleaning chemicals and process down time. Therefore, it should be investigated in all cases whether possible savings in running costs can justify the investments in additional CIP equipment, normally consisting of extra tanks, pumps and pipe runs.

'Single-use' CIP system

Such a system (**Figure 1**) consists of one tank that is used to prepare a certain cleaning solution at the desired temperature. This can already be done when the line is still in production, provided that there is a good physical separation between the cleaning system and the production system. After connection to the process line, a centrifugal pump circulates the cleaning solution through the part of the process plant that has to be cleaned. The pipe system between tank and pump contains a manifold through which optionally cold or hot water or steam can be introduced into the circuit. Prerinsing and postrinsing are once-through operations; after passing the process equipment, the rinse is discharged directly to drain. The cleaning liquid may be circulated over the tank for a certain period of time, but is also discharged to drain at the end of the cleaning procedure.

The automation level of such systems can be kept low; because of their simplicity they can be easily operated manually. The advantage of these systems is that time is gained during preparation of the cleaning solution(s). The disadvantage of possibly high running costs is still the same.

Single-use System with Partial Recovery

This system can be used under the same conditions as above, but now the option of recovering fluids has been added by the use of two buffer tanks (**Figure 2**). Also, here, the detergent solution is normally dumped after use, but when, after the prerinse, the residual soiling is light, recovery of the cleaning solution in one of the tanks may be considered. This may be used one or more times for subsequent cleaning procedures. When a disinfection step is required, there might also be recovery of disinfectant in one of the buffers. If detergent cannot be recovered, there is still the possibility of recovering postrinse water, which can be used for the next prerinse.

Optimal Reuse System with Maximum Recovery (Centralized System)

This system contains sufficient tanks that several fluids and solutions used in cleaning have their own

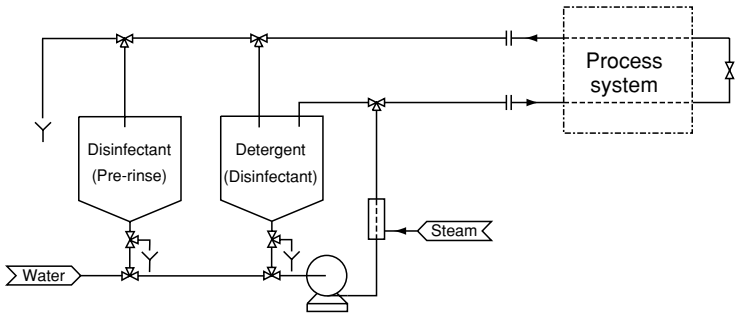

Figure 2 Flow diagram of a 'single-use' CIP system with partial recovery (types of valves depending on specific local requirements).

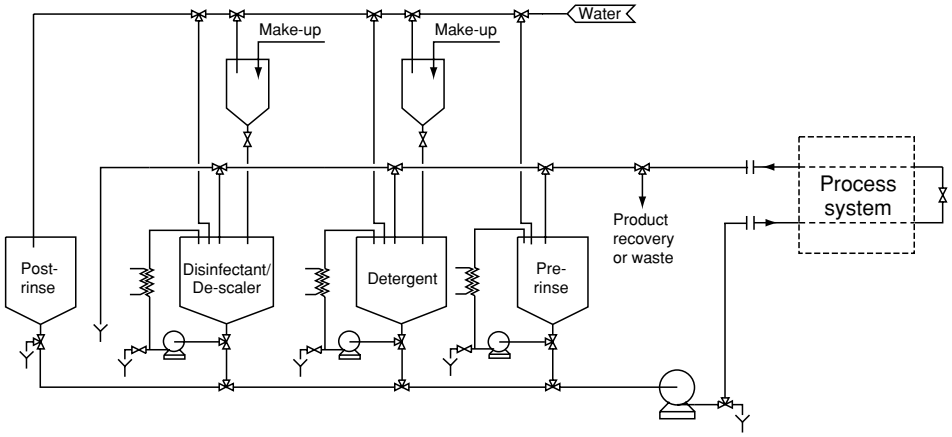

Figure 3 Flow diagram of an optimal, centralized reuse CIP system with the possibility of maximum recovery (types of valves depending on specific local requirements).

dedicated buffer (Figure 3). In the simplest form, there are at least two tanks; one for prerinse water and one for detergent solution. Postrinse may be recovered in the prerinse tank, and the detergent can be recovered (after a certain waiting time to prevent dilution) in its own buffer. Extra tanks might be necessary for buffering, e.g., a disinfectant, descaling solution or postrinse water.

The benefits are clear; cleaning can start immediately after production stops, and the cleaning procedure can be executed without any interruptions. Preparation of cleaning fluids of the right strength and temperature can be done with separate circuits to avoid any delay between cleaning stages. Moreover, there is a cost saving from the recovery of water, chemicals, and energy, and, in some cases, by the possibility to recover product (if this can be reworked).

An additional advantage is a better control over the effluent costs. To run such systems in a safe and efficient way, a higher level of automation is highly preferred. This implies higher investment costs.

Partially Decentralized System

When, in a factory, a centralized CIP system (as described before) is situated a long distance from the various process systems, the efficiency of such a system might be affected by the following factors:

- large volume of liquid in long pipe runs, thus requiring larger tanks and more energy input;
- long time needed to purge out a liquid/replace it by another (resulting in chemical loss);
- when more cleaning loops are necessary, the costs of pipe work increase dramatically.

To overcome this, a central CIP system can be extended with satellites, which only serve one distinct part of a process system (Figure 4). These satellites can be rather simple, but should be able to receive the required cleaning liquids from the central plant within a reasonably short time, e.g., by means of a main loop system with relatively small pipe diameters.

Depending on the situation, a decision can be made on whether or not to recover liquids to be sent back to

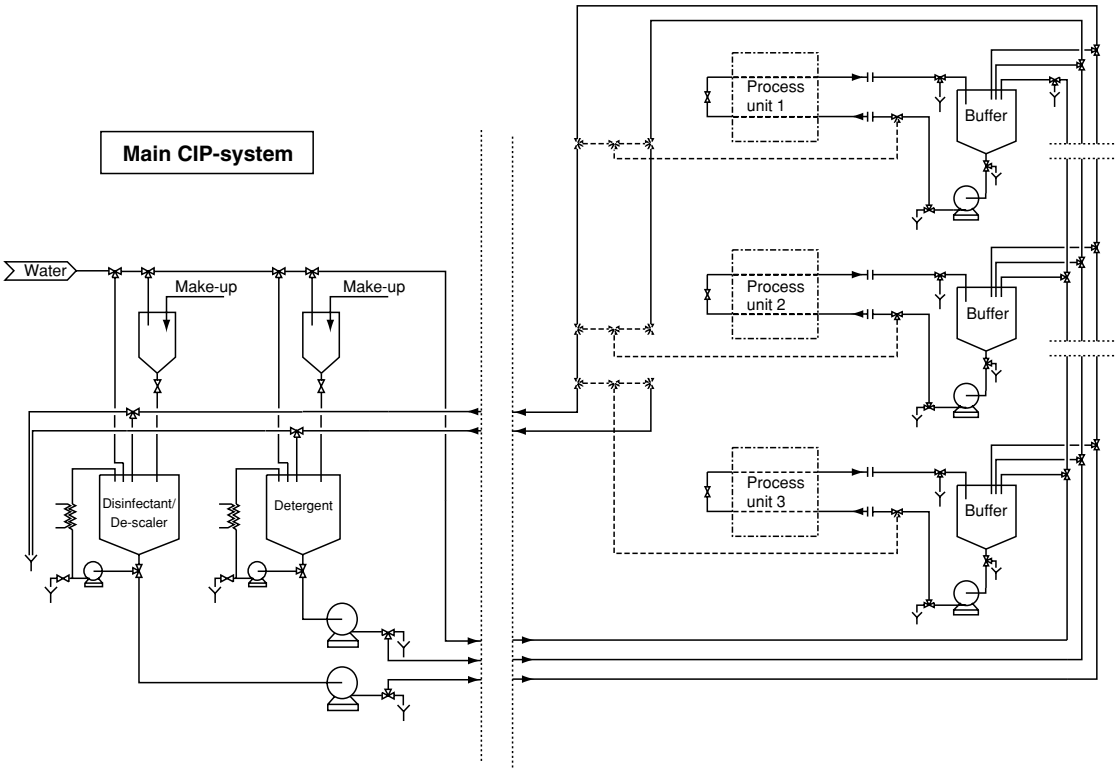

Figure 4 Flow diagram of a partially decentralized system with satellites (types of valves depending on specific local requirements).

the central system. Such a system may offer particular advantages, e.g., when preparation stages of a process can be cleaned while, downstream, other process equipment and filling machines are still in use. In such cases, a centralized CIP system will require far more safety measures (e.g., leak protected valves) or should make partial cleaning unacceptable. Also, parts that require more frequent cleaning could be cleaned by satellite systems.

Logistic Aspects Affecting the Layout of the CIP System

Normally, a process plant will be cleaned at the end of a (series of) production run(s). A logical sequence is to start with complete removal of all loose product from the process line by purging (optionally for recovery) and prerinsing (to waste and/or to drain). This means that the complete prerinse of the whole plant is done before cleaning liquid is introduced. Usually, such a rinsing procedure follows the same routing as the process; starting from the introduction point of raw materials, the process line is followed down stream to the filling machine. Later on, the same sequence is repeated with subsequent cleaning fluids. From a control point of view, this is a good and simple option, and moreover, only a few safety measures are required. The only safety precaution necessary is to physically separate the central CIP system immediately after the feed pump from the process system during actual production. Another advantage is that switching over from one stage to the next (e.g., prerinse to detergent cleaning), has to take place only once, so that any losses by mixing of liquids in the CIP circuit are minimized. A typical example of such a 'sequential' system is presented in Figure 5.

A disadvantage of such a system is that, because of this sequential procedure, the cleaning time might be rather long. It can be regarded as a waste of time that a prerinsed preparation vessel has to wait for the next cleaning cycle because the finished product buffer still has to be rinsed. Moreover, it is common practice to start cleaning of an ingredient preparation department directly after preparing the last batch. Frequently, at that moment, processing and especially filling of the product will still continue for a considerable time.

For these reasons, a process system might be split up in individual smaller units. As mentioned previously such units should all be able to be physically separated from the CIP system and from each other. By doing this, the parts that are already out of production can be immediately cleaned, resulting in a considerable gain of time. An example of such a 'unit cleaning' system is presented in Figure 6. This

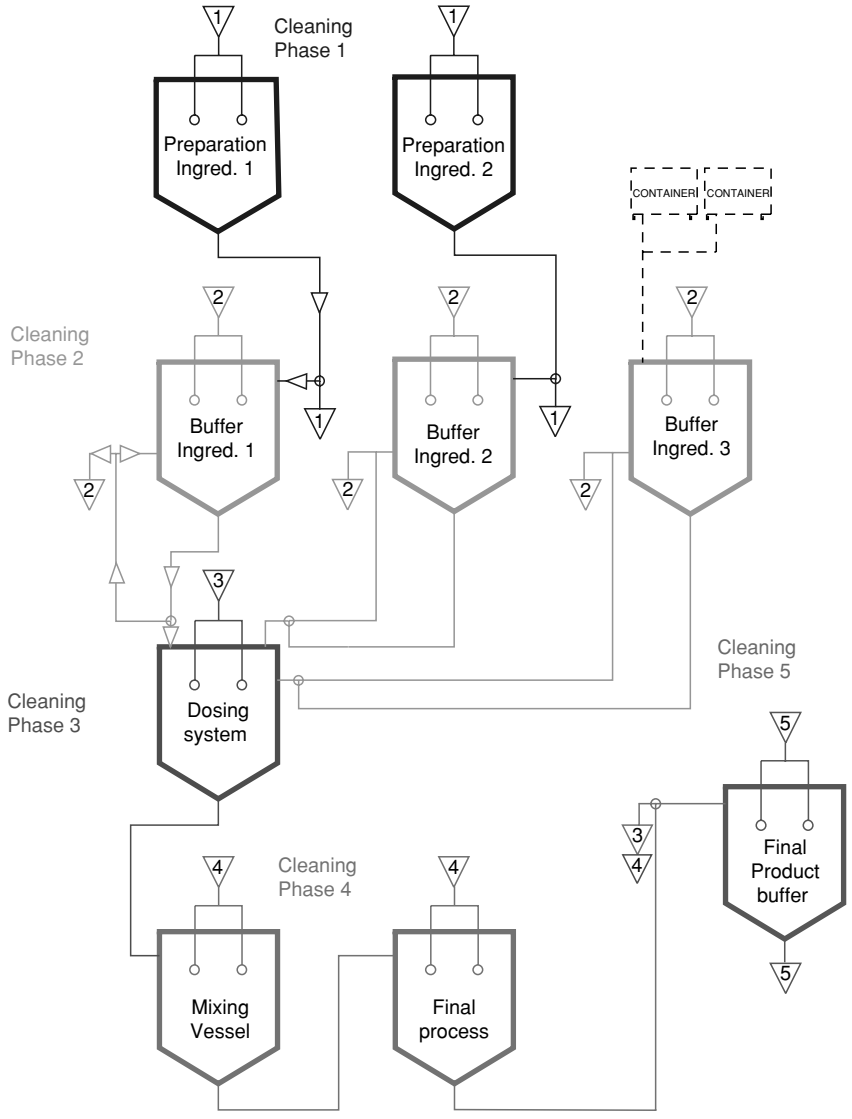

Figure 5 Typical example of a process line that is cleaned 'sequentially'. All CIP inlets and outlets (triangles) are connected to the same CIP circuit.

may result in overlapping of cleaning cycles of various parts of the plant, which are not compatible with each other (e.g., one part requires detergent, another post-rinse water). Again, this may give rise to undesired waiting times. A solution is then to install more cleaning circuits with feed and return pumps connected to the same buffer tanks. Each of these circuits is connected to a defined and separated part of the process. This latter option is also the optimal practice when parts of the process require a much higher cleaning frequency than the total plant. To avoid an extensive, complete safeguarding of the total plant, such parts can be provided with their own dedicated cleaning loop, connected to the same central CIP-buffer tanks, with only one physical separation in supply and return.

Systems for Open Plant Cleaning (OPC)

A wide range of systems are available for the delivery of pressurized water and the application of detergents in open plant cleaning. These systems vary in terms of sophistication and investment. The selection will be based on a number of factors, which can be broadly classified on the basis of whether the operation being carried out is a cleaning, disinfection, or rinsing step.

Hosing is one of the most usual ways of delivering water, which is used as a cleaning agent. Pressure systems are generally characterized as being either low pressure–high volume or high pressure–low volume.

Normally, for (warm) rinsing water, a ring main system is available with tap points through the

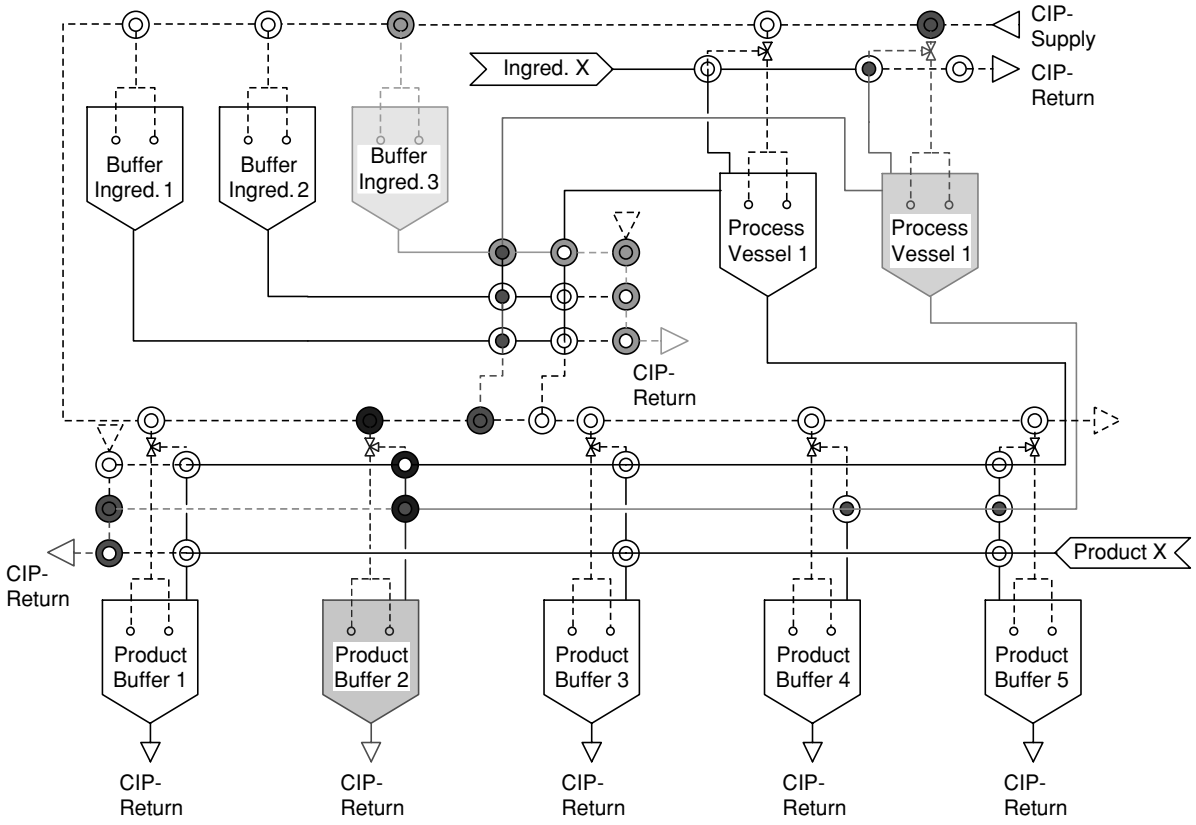

Figure 6 Example of a 'unit cleaning' system. The various degrees of shading indicate some of the units that can be cleaned in isolation of the rest of the plant, where production may continue during the cleaning operation. The circles indicate the places where block-and-bleed systems are required. The individual units may be connected to different CIP circuits.

factory. If the gross soil removal is to be carried out by pressure rinsing, warm water (50–55 °C) at a pressure of up to 20 bar is typically used. The temperature is high enough to melt fats but not high enough to denature proteins, which could result in burn on. A pressure of 20 bar (with a flow of 20–30 l min^{-1}) is sufficiently high to remove most of the gross soil whilst minimizing risk of damage. The pressure is low enough to prevent excessive aerosol formation occurring, which could lead to microorganism-containing aerosols being created, with the risk of contaminating adjacent areas of the factory.

Foam cleaning is a process that involves applying detergent to a dirty surface as a foam. Provided that the foam is stable and not too wet, it will cling to the surface for a relatively long time (5–10 min), during which time, physical/chemical attack can soften the soil.

The foam and loosened dirt are then rinsed away from the surface. This is preferably done with a high-pressure jet, provided that there is no danger of spraying electrical switchgear or similar water-sensitive equipment. In this case, a low-pressure spray or, occasionally, a vacuum can be used to remove the foam.

Another cleaning method is *gel cleaning*. Gel cleaning detergents are chemicals that are of low viscosity as a 100% product. On dilution, however, they gelate, and by this, they will increase the contact time of the detergent on the soil (10–30 min for gels). This prolonged contact time gives a much stronger effect and therefore better cleaning results, or equal results with rinsing at reduced pressure (5–20 bar).

Mobile Equipment

Detergent and disinfection can be carried out using mobile equipment. Mobile foam or gel equipment generally consists of a pressure vessel. The vessel (50–100 l) is filled with detergent solution. The detergent solution is forced by compressed air into a mixing chamber and mixed with foam or gel. The foam is applied to the surface via a hose and foam lance provided with a vee-jet nozzle. The vee-jet nozzle allows the application of an even and thin foam layer. After the required contact time, the

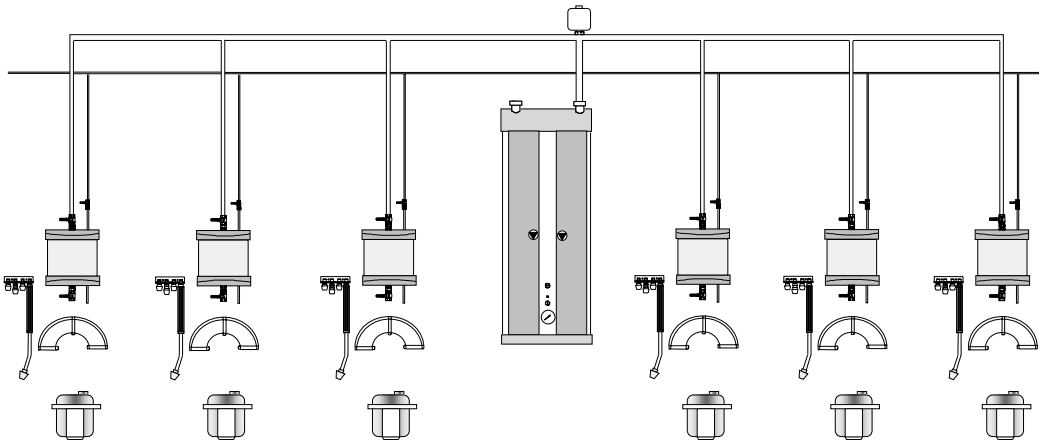

Figure 7 Decentralized OPC installation with booster pump and satellites.

foam/gel and soil are rinsed away, again using the pressurized water supply.

Foam cleaning is particularly useful for cleaning the outside of complex machinery, walls, and other large surfaces, as these surfaces can be easily and quickly covered with a foam layer.

A wide range of detergents from strongly alkaline to acidic are suitable for use with foam cleaning, so that most soiling in the food industry can be dealt with. Foam cleaning is typically not suited for surfaces covered with a thick soil layer (> 1 mm), as the foam will not be able to penetrate thick layers completely and reach the surface.

Advantages and Disadvantages

Foam cleaning is similar to the presoak approach of pressure jet cleaning, the only difference being related to the form of the detergent solution that is used for presoaking. The foam approach has important benefits:

- it is a relatively safe way of applying aggressive, e.g., caustic based detergents on to open surfaces;
- it is less likely to cause irritant aerosol problems;
- it is easy to see where the detergent has been applied;
- it strongly appeals to operators.

Against these benefits, it has the following disadvantages:

- it requires special, although not very expensive, equipment to produce the foam;
- it could give a false feeling of security because everything is covered with a white clean blanket, also if the soil cannot be attacked effectively.
- the detergent solution has to be prepared manually.

Disinfection can be done with the same equipment, except that a small tank (15 l) provided with a tube and spray-lance with a vee-jet is used. The tank is filled with 10 l of disinfectant solution and is then pressurized with 5-bar compressed air. The disinfectant is applied to the surface using the tube and the spray-pistol. The advantages of this system are its simplicity and (therefore) low cost, compared with more advanced systems.

Booster Pump with Satellites

A more advanced system consists of a booster pump (10–25 bar) that provides satellites, with pressurized (warm) water (Figure 7). The satellite is also connected to a compressed air supply. Using a venturi system, detergent or disinfectant can be diluted to suit the application concentration. Because of the connection with a compressed air supply, the detergent solution can be applied as a foam. Each satellite can be used for prerinse, detergent application, rinse, disinfectant application, and final rinse. Different satellites, connected to the same booster pump, can carry out different options (foaming, rinsing, or disinfecting). The advantages of this system are that it uses an automatic detergent solution make-up and fixed pressure and that there is no mobile equipment floating around. The disadvantage of this system is the higher capital investment compared with the previous system.

Booster Pump with Central Make-up for Detergent and Disinfectant

For large plants, with many take-off points (Figure 8), the detergent and disinfectant solution preparation can be done centrally, using dosing pumps (and flow meter measurements). Not only pressurized water,

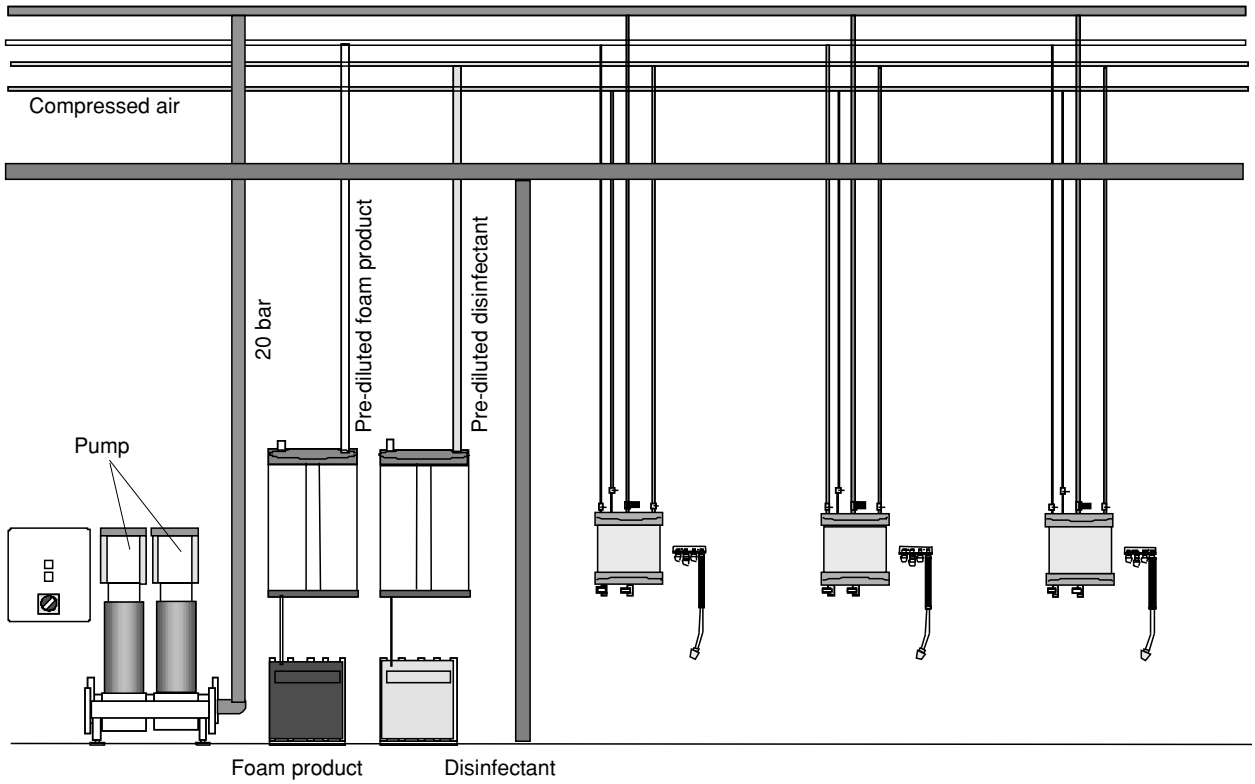

Figure 8 Central OPC installation with central booster pump, satellites, and central make-up of media.

but also detergent solution, disinfectant solution, and compressed air flow to each satellite.

The advantages of this system are that it has higher capacities and does not use concentrated chemicals in the production area. The disadvantage is that the same detergent and disinfectant have to be used on all take-off points.

See also: **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant; Overall Approach; **Factory Construction**: Materials for Internal Surfaces

Further Reading

Imholte TJ (1984) *Engineering for Safety and Sanitation*. Crystal, MN: Technical Institute of Food Safety.

Karlsson CA-C (1999) *Fouling and Cleaning of Surfaces – The Influence of Surface Characteristics and Operating Conditions*. Lund, Sweden, Lund University.

Lelieveld HLM (2000) Hygienic design of factories and equipment. In: *The Microbiological Safety and Quality*, pp. 1656–1690. Guithersburg, MD: Aspen.

Shapton DA and Shapton NF (eds) (1998) *Principles and Practices for the Safe Processing of Foods*. Cambridge, UK: Woodhead.

Wilson DI, Fryer PJ and Hasting APM (eds) (1999) *Fouling and Cleaning in Food Processing '98. Proceedings of a Conference at Jesus College, Cambridge*. Brussels: Directorate-General Science, Research and Development.

CLOSTRIDIUM

Contents

Occurrence of *Clostridium perfringens*

R G Labbé, University of Massachusetts at Amherst, Amherst, MA, USA

Introduction

Clostridium perfringens is probably the most widespread of all pathogenic bacteria. There are several toxigenic types: A, B, C, D, and E. Type A is primarily associated with human illness. The other types are associated with diseases of domestic animals. In very specific situations, type C is occasionally involved in human illness. *Clostridium perfringens* is an anaerobic, spore-forming organism commonly found in fresh meat and poultry products. Spores of the organism can survive many food processing procedures. Because of its ability to grow over a wide temperature range, it is often implicated in human food poisoning. With regard to human illness in general, it should be noted that, historically, *C. perfringens* have been most closely associated with gangrene and wound infections. In this article, discussion will be limited to its role in human food poisoning.

Occurrence in Humans, Foods, and the Environment

Clostridium perfringens is part of the normal intestinal flora in humans and animals and also occurs widely in soil. It is the most commonly found *Clostridium* in clinical specimens. Because of its abundance in feces, the organism is also found in sewage-polluted water. Water authorities in certain localities use its presence as an index of water quality.

Clostridium perfringens has also been found in the intestinal tract of virtually every animal examined, with a wide variation within and between species. Although the levels of *C. perfringens* in healthy adults are relatively small compared with other strict anaerobes, *C. perfringens* can be isolated from the gut of virtually all humans. In infants, adult levels are established by 6 months of age.

Early studies on the incidence of *C. perfringens* focused on the isolation of so-called heat-resistant strains (those whose spores could survive – and be activated by – heating at 100 °C for 60 min) since it was thought that this group was more likely to survive cooking than less heat-resistant, i.e., 'heat-sensitive' spore strains. By the mid-1960s, it became apparent that heat-sensitive strains were equally capable of causing outbreaks of food poisoning. In epidemiological investigations, no distinction is now made between the two groups.

There is a wide variation in the total *C. perfringens* count in human feces. However, in healthy adults, the values are usually 10^3–10^5 per gram (Table 1). Patients in outbreaks carry 10^6–10^8 per gram. The level of spores of this organism in feces is within one log of the total count. The procedure used to obtain the spore count – heating the sample at 75–80 °C for 10–20 min – also eliminates competing microflora. The fecal spore count is one of several laboratory criteria for investigating outbreaks caused by this organism (see below).

It has become apparent that the elderly, although healthy, often carry relatively high (total or spore) numbers of *C. perfringens*, often above 10^6 per gram. This phenomenon is not attributable to ingestion of elevated levels of *C. perfringens* since surveys have been conducted in extended-care facilities where the daily intake was monitored. Such high levels in asymptomatic elderly limit the usefulness of examining stools of patients for elevated levels of *C. perfringens*. In such situations, other criteria for confirming outbreaks, discussed below, are available.

Early studies on the incidence of *C. perfringens* in raw foods focused on heat-resistant strains, thus understating true levels of the organism. Representative results of many market and slaughterhouse surveys, conducted over the years, are presented in Table 2. No distinction is made between vegetative

Table 1 Representative surveys of *C. perfringens* cells and spores in feces from various population of various countries

Population	*Country*	*Cell type*	*Levels (per gram)*
Healthy adults	USA, UK	Total viable count	10^3–10^4
Young patients	UK	Total viable count	3 of 6, $\geq 10^4$ 0 of 10, $\geq 10^6$
Elderly adults	Japan	Total viable count	5 of 30, $\geq 10^7$
Elderly patients	UK	Total viable count	10 of 11, $\geq 10^4$ 5 of 11, $\geq 10^6$
Elderly mental patients	UK	Total viable count	6 of 10, $\geq 10^4$ 3 of 10, $\geq 10^6$
Health adults	Canada	Spore count[a]	10^3–10^4
Food-poisoning patients	UK	Spore count[a]	56 of 66, 10^6–10^8
Food-poisoning patients	USA	Spore count[a]	2.0×10^4–4.0×10^8 Mean, 10^7
Food-poisoning patients	USA	Spore count[a]	$< 10^3$–2.2×10^5[b]

[a]Fecal samples heated at 80 °C for 10 min.
[b]30 days after illness.

Table 2 Results of surveys of the incidence of *C. perfringens* in food and feeds

Raw food types	*Incidence (%)*
Meat and poultry	
Poultry carcass	58
Frozen chicken	63
Beef carcass	26
Pork carcass	66
Lamb carcass	85
Ground beef	50–70
Beef liver	26–50
Veal	82
Pork	37
Pork sausages	39
Lamb, mutton	52
Fish and shellfish	
Fish, body surface	84
Fish, alimentary tract	82
Vacuum-packed fish	67[a]
Oysters	100
Trout	0
Miscellaneous	
Spices and herbs	42
Dehydrated soups and sauces	18
Animal feeds	35

[a]Incidence of *Clostridium*; predominant species is *C. perfringens*.

cells and spores. The latter would obviously be more able to withstand subsequent cooking procedures. The data for meat and poultry were from surveys conducted in North America and the UK. Results from Japan are consistently lower. It should also be noted that such surveys have been carried out using a variety of methods for enrichment, heat selection, selective plating, and confirmation. It is clear from Table 2 that the organism is abundant in raw, protein-rich foods. The source of the organism is the intestinal contents of these animals. As noted above, the feces of all animals examined, domestic and wild, contain *C. perfringens*. In the case of fish and shellfish, there are wide fluctuations in isolation rates, presumably depending on the degree of water pollution. Such products are rarely involved in outbreaks of *C. perfringens* food poisoning.

The organism is also found in many types of processed foods, although at a low level. However, in some cases, such as soups and sauces, only short heating times are required for preparation. Other items, such as herbs and spices (well known for their high general bacterial spore levels, including *C. perfringens*) are often added to large amounts of cooked foods. Slow cooling or inadequate reheating of such foods can result in the large number of *C. perfringens* necessary to cause food poisoning. Oxygen is driven off during cooking, creating ideal conditions for growth of this organism.

Clostridium perfringens is part of the microflora of soil and is present at levels of 10^3–10^4 per gram. Even in Antarctica, moist soil samples examined contained this organism. In view of its widespread presence in soil, its presence in air and dust (including kitchen dust) is not surprising. In the case of marine sediments, there is a close relationship between the amount of fecal pollution and the number of *C. perfringens*.

Food Poisoning

Food poisoning attributable to *C. perfringens* usually occurs 8–24 h after the ingestion of temperature-abused food containing large numbers of vegetative cells. Symptoms last 1–2 days and generally include diarrhea and severe abdominal cramps. Vomiting is not common, and fever is rare. Type A cells are usually responsible. A more severe type of illness, caused by type C, occurs among young adults of the

highlands of New Guinea. It is necrotizing (necrosis: death of areas of tissue surrounded by healthy parts), hemorrhagic jejunitis (inflammation of the jejunum, the second portion of the small intestine extending from the duodenum to the ileum) which is often called 'pig-bel' (enteritis necroticans) because it follows traditional pig feasting.

It should be noted that ingestion of low levels of microbial spores, including *Clostridium botulinum* and *C. perfringens*, is a common occurrence and not a public safety issue for adults. Only when these spores have been allowed to germinate and grow in food products do they pose a health threat.

Mechanisms of Entry into the Food Chain

The vegetative cells and spores of *C. perfringens* are common surface contaminants of fresh meat and poultry carcasses. This is not surprising in view of the common occurrence of the organisms in the intestine of these animals. They can be easily disseminated during processing steps such as skinning, evisceration, and scalding. Unlike the case for *Salmonella*, the absence of this organism from fresh meat and poultry is an unreasonable expectation. Furthermore, the mere presence of *C. perfringens* (as spores) surviving cooking will not cause outbreaks of foodborne illness. For the latter to occur, gross mishandling and temperature abuse must always be involved. (*See* **Meat**: Eating Quality; Hygiene; **Poultry**: Chicken; Ducks and Geese; Turkey.)

Fate During Processing and Storage

Temperature is the single most important determinant of the survival and multiplication of *C. perfringens* subsequent to slaughter and packaging. Generation times as low as 8–10 min have been reported for this organism at its optimum growth temperature. Other considerations affecting growth include absence of oxygen, water activity, pH, and salt content. However, alterations of these usually involve further processing steps, and epidemiological investigations have consistently implicated fresh meat and poultry as sources of the organism. (*See* **Meat**: Preservation.)

The fate of *C. perfringens* during processing and storage depends on the form of the organism, i.e., vegetative cell or spore. Both are present in fresh meat and poultry, but each requires different considerations with regard to immediate or potential hazard.

Vegetative cells can grow over the temperature range of 15–50 °C, with optima between 43 and 46 °C. Even between 60 and 70 °C, viability may be maintained, but vegetative cells are rapidly inactivated at 75 °C. Considering the short generation time of the organism, the slow attainment of a safe interior temperature can actually increase the initial number of organisms and permit more cells to survive. Thus, the rate at which the interior temperature is attained may also influence the thermal survival of vegetative cells. For example, rump roast cooked to an internal temperature of 77 °C in 2.25 h has been shown to retain significant numbers of viable *C. perfringens* cells. However, experiments with chicken breast and thigh have shown complete killing of 10^8 vegetative cells when the pieces were cooked in water at 82 °C, and the internal temperature of 77 °C was attained in 20 min or less. The standard dictum that cooked meat should be kept above 62.8 °C or below 10 °C will insure safety of properly heated food.

Most *C. perfringens* spores isolated from meat and poultry are of the heat-sensitive variety. These are killed in a few minutes at 100 °C. Unfortunately, spores of the heat-resistant variety are also present in lower numbers. These have D_{100} (decimal reduction value at 100 °C) values of 6–17 min and can survive cooking procedures (which themselves drive off oxygen), germinate, and resume vegetative cell growth given the proper conditions, principally a suitable temperature.

The effect of low-temperature storage on *C. perfringens* cells is important because food safety with regard to this organism is based largely on proper refrigerated holding. *Clostridium perfringens* vegetative cells are sensitive to low temperature, e.g., refrigerated storage. Slow die-off occurs under these conditions. Similarly, long-term (several weeks) freezing slowly inactivates vegetative cells. The initial freezing step reduces the population approximately 10-fold. Surprisingly, vegetative cells die more rapidly at −5 °C than at −20 °C. As one would expect, spores are considerably more resistant. They are virtually unaffected by refrigerated storage and only somewhat inactivated by freezing. Indeed, frozen storage in the spore state is routinely used for culture carriage.

Before spores can resume vegetative cell growth, they must germinate. Proper nutrients must be available, and these are readily available in meat and poultry products. Viable bacterial spores are traditionally measured by heating a culture at an elevated temperature (75–80 °C, depending upon the strain) for 10–20 min and performance routine plating procedures. This procedure 'activates' the spore population (and inactivates any vegetative cells). In the case of raw food, this function is effectively achieved by routine cooking procedures. Optimal temperatures for germination are similar to those for vegetative cell growth, in a pH range of 5.5–7.0.

It is difficult to specify the time required for cells of *C. perfringens* to multiply in foods to attain toxic

numbers, but it has been observed that meats stored at a 'warm' temperature for at least 2 h after cooking were common factors in many outbreaks. The hazard is magnified when such food is allowed to cool slowly for several hours, e.g., overnight at room temperature, as has occurred with large turkeys or large bulk of other meats.

The multiplication of bacteria is a logarithmic function. The rate at which a product may accumulate harmful numbers of cells will depend to a large extend on the size of the inoculum. The temperature at which cooked foods is held or stored is the other highly dependent variable. As mentioned above, that is especially true with *C. perfringens* in view of its ability to grow at relatively elevated temperatures.

See also: **Beef**; ***Clostridium***: Detection of *Clostridium perfringens*; Food Poisoning by *Clostridium perfringens*; **Meat**: Preservation; **Pork**; **Poultry**: Chicken; Ducks and Geese; Turkey; **Sheep**: Meat

Further Reading

Labbé R (2000) *Clostridium perfringens*. In: Lund B, Baird-Parker T and Gould G (eds) *The Microbiological Safety and Quality of Food*, vol. II, pp. 1110–1135. Gaithersburg, MD: Aspen.

Labbé R (2001) *Clostridium perfringens*. In: Downes FP and Ito K (eds) *Compendium of Methods for the Microbiological Examination of Foods*, 4th edn, pp. 325–330. Washington, DC: American Public Health Association.

McClane B (2000) *Clostridium perfringens*. In: Doyle M, Beuchat L and Montville T (eds) *Food Microbiology: Fundamentals and Frontiers*, 2nd edn. pp. 351–372. Washington, DC: American Society for Microbiology Press.

Smith L and William B (1984) *The Pathogenic Anaerobic Bacteria*, 3rd edn. Springfield, IL: Charles Thomas.

Stringer M (1985) *Clostridium perfringens* type A food poisoning. In: Borriello S (ed.) *Clostridia in Gastrointestinal Disease*, pp. 117–141. Boca Raton, FL: CRC Press.

Detection of *Clostridium perfringens*

R G Labbé, University of Massachusetts at Amherst, Amherst, MA, USA

Laboratory Criteria for Confirming Outbreaks

Laboratory confirmation of an outbreak of *Clostridium perfringens* food poisoning is based on one of five criteria: (1) more than 10^5 of the organism per gram of food; (2) more than 10^6 spores of the organism per gram of the feces of ill person; (3) the presence of the same serotype in most of the ill patients; (4) the presence of the same serotype in the incriminated food and feces of the patients; and (5) detection of enterotoxin in feces. Detailed procedures for serotyping are available from citations in the Further Reading section at the end of the next article.

Detection in Raw and Processed Foods

Detection of *C. perfringens* in raw and processed foods is performed by similar methods. Such foods, properly handled, would not normally contain more than 100 *C. perfringens* cells or spores per gram, usually much less. In such situations, most probable number (MPN) test tube procedures can be used for enumerating low numbers in foods. Iron-containing milk (iron milk medium, or IMM, i.e., 10 ml of homogenized milk containing 0.2 g of iron powder) has been used for this purpose. When incubated at 46 °C, *C. perfringens* produces a typical 'stormy fermentation' in IMM. This is defined as the production of an acid curd (caused by lactic acid fermentation) with subsequent disruption of the curd by large volumes of gas. NonMPN enrichment media (trypticase–glucose–yeast extract broth), with incubation at 37 °C followed by selective plating on trypticase–sulfite–cycloserine (or neomycin blood agar) agar plating, can also be used for enumeration of very low numbers. Confirmation (see below) is required in either method.

Detection of Cells in Suspected Food Poisoning

Foods implicated in outbreaks of human food poisoning would normally contain large numbers of vegetative cells and relatively few spores. Such food should be chilled and processed rapidly because of the susceptible of the cells to cold shock. For delayed analyses, the highest counts are obtained when foods (finely chopped if necessary) are mixed 1:1 with 20% glycerol and kept at −20 °C or, if shipping is necessary, placed in a container of dry ice. (*See* **Food Poisoning**: Tracing Origins and Testing.)

Selective plating methods are used for enumeration of viable cells. Enrichment techniques are unnecessary for examination of food containing large numbers of cells. Most plating media depend on the ability of *C. perfringens* to reduce sulfite to sulfide, which, in the presence of an iron salt, results in the formation of black colonies owing to ferrous sulfide. Collaborative analyses have indicated that

pour-plated tryptose sulfite cycloserine (TSC) without egg yolk is the medium of choice. More consistent blackening of surface colonies can be obtained by overlaying plates with sterile media. TSC is commercially available from Unipath (Oxoid). After anaerobic incubation for 24 h at 37 °C, representative black colonies (usually 10) must be confirmed. This is achieved by inoculating a liquid medium, such as trypticase peptone glucose yeast extract broth or fluid thioglycollate medium, and incubating at 46 °C for 4 h or overnight at 37 °C. Tubes of lactose–gelatin and motility–nitrate are inoculated from each and incubated at 37 °C for 24 h. *Clostridium perfringens* is nonmotile, ferments lactose, liquefies gelatin, and reduces nitrate to nitrite. The number of *C. perfringens* per gram is determined by multiplying the presumptive plate count by the ratio of colonies confirmed as *C. perfringens*. Fecal spore levels are determined in the same manner, except that the sample is heated at 75 °C for 20 min. The elevated-temperature MPN methods mentioned above are not recommended for quantification of *C. perfringens* in outbreak stools.

Surface-plated neomycin blood agar plates are often used in the UK and can be prepared well in advance. This medium also provides information on the hemolytic activity of isolates. However, because recovery of certain heat-resistant strains may be no more than 10% on this medium, its use is limited to outbreak stools or food samples containing large numbers of *C. perfringens*. Neomycin blood agar is not recommended for examining normal food samples in which the organism is present in low numbers.

Detection of Enterotoxin in Suspected Food Poisoning

The ingestion of large numbers of vegetative cells in incriminated food is followed by multiplication of the cells in the small intestine. When they sporulate, there is an accompanying formation of enterotoxin. Lysis of the sporangia to release the mature spore also results in the release of enterotoxin.

Serum values of antienterotoxin are of little value in the diagnosis of *C. perfringens* food poisoning, and enterotoxin detection in foods is not a practical approach. However, detection of the enterotoxin in stools is of significant diagnostic importance since the toxin is not detectable in the feces of healthy adults. Of the criteria listed above, there are occasions when only detection of enterotoxin in feces is conclusive; for example, when no food is available, when the strains are not typeable, or when the incidents concern geriatric patients who may carry large numbers of the same serotype or spores without symptoms of food poisoning. Most fecal specimens from food poisonings incidents have enterotoxin concentrations exceeding 1 μg per gram of feces.

Two procedures for detection of enterotoxin in feces have found widespread use and are effective when used within 2 days of onset of symptoms. They are the enzyme-linked immunosorbent assay (ELISA) and the reversed passive latex agglutinations assay (RPLA). The latter is available as a kit from Unipath (Oxoid). Although expensive (if obtained commercially) for multiple samples, the RPLA method is the simpler of the two. In this method, latex beads that have been sanitized (treated) with enterotoxin antiserum are exposed to serial dilutions of enterotoxin-containing material. After overnight incubation, the agglutination titer is determined. Expensive equipment, such as a microplate reader (needed for ELISA), is unnecessary. However, nonspecific agglutination can occur at very low (near the detection limit) dilutions. The ELISA method is preferable when more than occasional samples are to be assayed. Some half-dozen different ELISA procedures have been proposed with sensitivities of 2–5 ng per gram of feces. (*See* **Immunoassays:** Radioimmunoassay and Enzyme Immunoassay.)

Statistics

As mentioned above, meat and poultry products are most commonly involved in cases of human food poisoning attributed to *C. perfringens*. The organism has complex nutritional requirements that are easily satisfied by such foods. However, cured meats are rarely implicated. Bacon and ham, for example, are seldom involved, presumably owing to the presence of curing salts and the lowered water activity, both of which inhibit vegetative cell growth. (*See* **Food Poisoning:** Statistics.)

Mass feeding establishments are consistently cited as the source where implicated food was eaten. Examples of these have included restaurants, cafeterias, prisons, schools, and hospitals. All such sites prepare large amounts of food well in advance of serving. Opportunities for mishandling of food in such setting are plentiful.

Detection of *Clostridium perfringens*

As with other agents of human food poisoning, the number of outbreaks of food poisoning attributable to *C. perfringens* is greatly underreported. This is particularly true with *C. perfringens* because of the relatively mild and short-lived nature of the symptoms. In addition, in some countries, e.g., the USA,

Table 1 Incidence of confirmed *C. perfringens* foodborne illness in selected countries

	Canada[a]		*USA*[b]		*England and Wales*[c]		*Japan*[d]	
Year	*Outbreaks*	*Cases*	*Outbreaks*	*Cases*	*Outbreaks*	*Cases*	*Outbreaks*	*Cases*
1980	18	753	25	1463	55	1054	13	5178
1981	15	399	28	1162	46	918	23	3482
1982	18	1420	22	1189	69	1455	11	896
1983	14	324	5	353	68	1624	16	4571
1984	20	888	8	882	68	1716	9	971
1985	13	390	6	1016	64	1466		
1986	21	354	3	202	51	896	22	3258
1987	18	369	2	290	51	1266	9	288
1988	42	456			57	1312	19	2671
1989	13	381	7	436	55	901	24	3316
1990	4	84	11	1240	53	1442	24	2503
1991			10	1213	44	733	21	3691
1992			12	912	36	805	17	1086
1993			15	534	38	562	9	1077
1994			12	517			16	1821
1995			14	455				
1996			10	1011				
1997			6	255			16	1821

[a]Todd (personal communication).
[b]Centers for Disease Control and Prevention, Atlanta, USA; yearly annual summaries.
[c]R Gilbert (personal communication)
[d]T Uemura (personal communication).

medical personnel are not required to report incidences of outbreaks to central public health officials. Thus, the data in Table 1 represent only a fraction of the true number of cases and outbreaks. In Western countries, the organism ranks second or third behind *Salmonella*, *Campylobacter*, or *Staphylococcus aureus* in terms of the number of cases of human food poisoning caused by bacteria. (*See* ***Campylobacter***: Properties and Occurrence; ***Staphylococcus***: Properties and Occurrence.)

See also: ***Campylobacter***: Properties and Occurrence; **Food Poisoning**: Tracing Origins and Testing; Statistics; **Immunoassays**: Radioimmunoassay and Enzyme Immunoassay; ***Staphylococcus***: Properties and Occurrence

Further Reading

Labbé R (2000) *Clostridium perfringens*. In: Lund B, Baird-Parker T and Gould G (eds) *The Microbiological Safety and Quality of Food*, vol. II, pp. 1110–1135. Guithersburg, MD: Aspen.

Labbé R (2001) *Clostridium perfringens*. In: Downes FP and Ito K (eds) *Compendium of Methods for the Microbiological Examination of Foods*, 4th edn, pp. 325–330. Washington, DC: American Public Health Association.

McClane B (2001) *Clostridium perfringens*. In: Doyle M, Beuchat L and Montville T (eds) *Food Microbiology: Fundamentals and Frontiers*, 2nd edn, pp. 351–372. Washington, DC: American Society for Microbiology Press.

Smith L and Williams B (1984) *The Pathogenic Anaerobic Bacteria*, 3rd edn. Springfield, IL: Charles Thomas.

Roberts D, Hooper W and Greenwood M (1995) *Practical Food Microbiology*. London: Public Health Laboratory Service.

Stringer M, Watson G and Gilbert R (1982) *Clostridium perfringens* type A: serological typing and methods for the detection of enterotoxin. In: Corry J, Roberts D and Skinner F (eds) *Isolation and Identification Methods for Food Poisoning Organisms*, pp. 111–135. London: Academic Press.

Food Poisoning by *Clostridium perfringens*

R G Labbé, University of Massachusetts at Amherst, Amherst, MA, USA

Introduction

Historically *Clostridium perfringens* is best known for its role in gas gangrene. During the 1940s and 1950s its association with foodborne illness was suggested. In the 1970s the responsible enterotoxin was isolated and its general mode of action described. In recent years work has focused on the subcellular mode of action of this toxin as well as the molecular genetics of its production.

Clinical Features and Characteristics

It is now well established that an enterotoxin is responsible for symptoms of *Clostridium perfringens* food poisoning. The symptoms are typically diarrhea and severe abdominal cramps (fever and vomiting are unusual) which occur 8–24 h after ingestion of food containing large numbers of vegetative cells. Sufficient numbers of cells survive stomach passage and sporulate in the small intestine.

The sequence of events can be duplicated in the laboratory by inoculating vegetative cells into a suitable sporulation medium. Many different types of sporulation media have been developed for this, but no single one is suitable for all strains. Figure 1 shows that about 3 h after inoculation of the sporulation medium, heat-resistant spores develop, followed closely by the intracellular accumulation of enterotoxin. Maximum numbers of spores are obtained after 7 h and free spores can be detected after 10–12 h. With the liberation of the mature spores from the sporangia, enterotoxin is released and the concentration of enterotoxin in the cell extract therefore decreases. The concentration of extracellular enterotoxin increases in parallel with the increase in free spores. In humans this corresponds to the release of enterotoxin into the lumen of the small intestine.

Site and Mode of Action

The enterotoxin of *C. perfringens* causes fluid accumulation in ligated small intestinal loops (sections), and overt diarrhea in a large number of experimental animals. The colon is not affected by the enterotoxin since, as least in rabbits, there is no change in the transport of fluid of electrolytes in this tissue. To determine the mode of action of the toxin, rabbit small intestines have been perfused with various electrolytes and nutrients after exposure to the toxin. The effects on intestinal transport and structure were determined. These procedures indicate that the enterotoxin causes a net secretion of water, sodium, and chloride. Glucose absorption is inhibited, whereas potassium and bicarbonate absorption are unaffected. The sensitivity of the rabbit's small intestine to the toxin increases from the upper duodenum downward, with the terminal ileum being the most responsive. Histological studies have shown that there is destruction of intestinal epithelial cells at the tips of

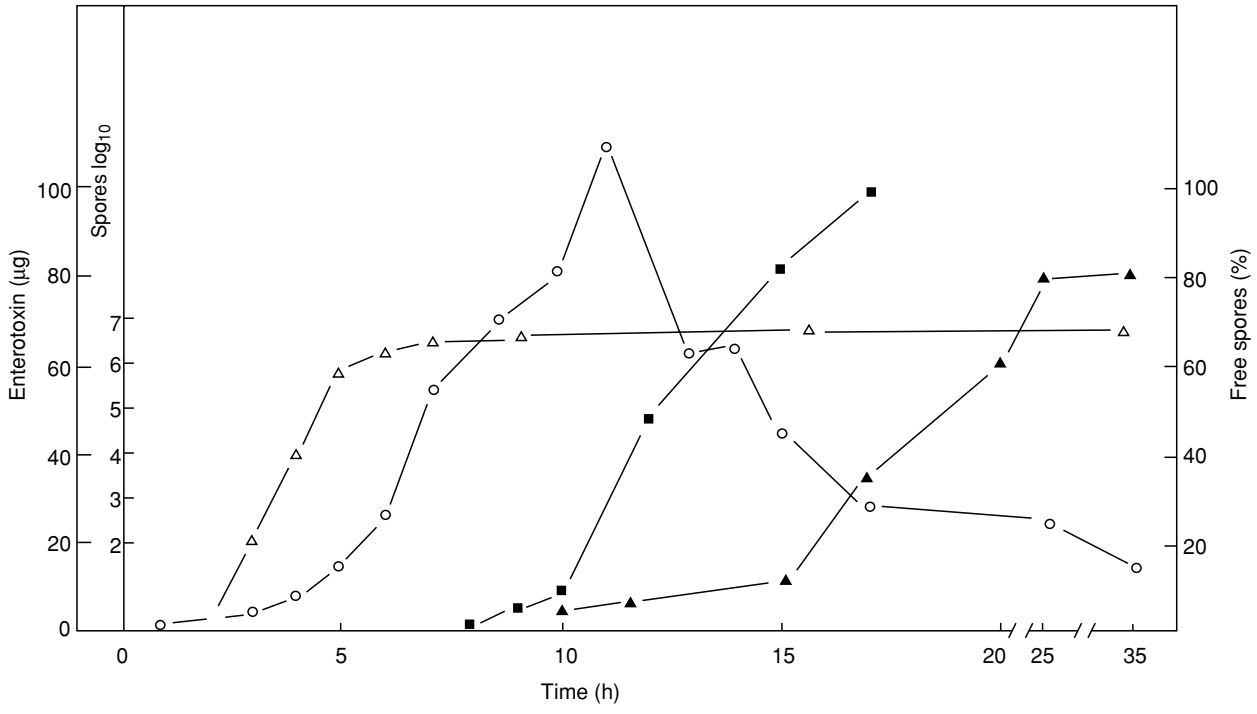

Figure 1 Time course of intracellular enterotoxin formation and release during sporulation of *Clostridium perfringens* type A. Open circles, content (μg) of biologically active enterotoxin per mg of cell protein extract; filled squares, μg of biologically active enterotoxin per ml of culture filtrate; open triangles, heat-resistant spores per ml; filled triangles, percentage of refractile spores free from sporangia. Adapted from Labbé R (1989) *Clostridium perfringens*. In: Doyle M (ed.) *Foodborne Bacterial Pathogens*, pp. 191–234. New York: Marcel Dekker.

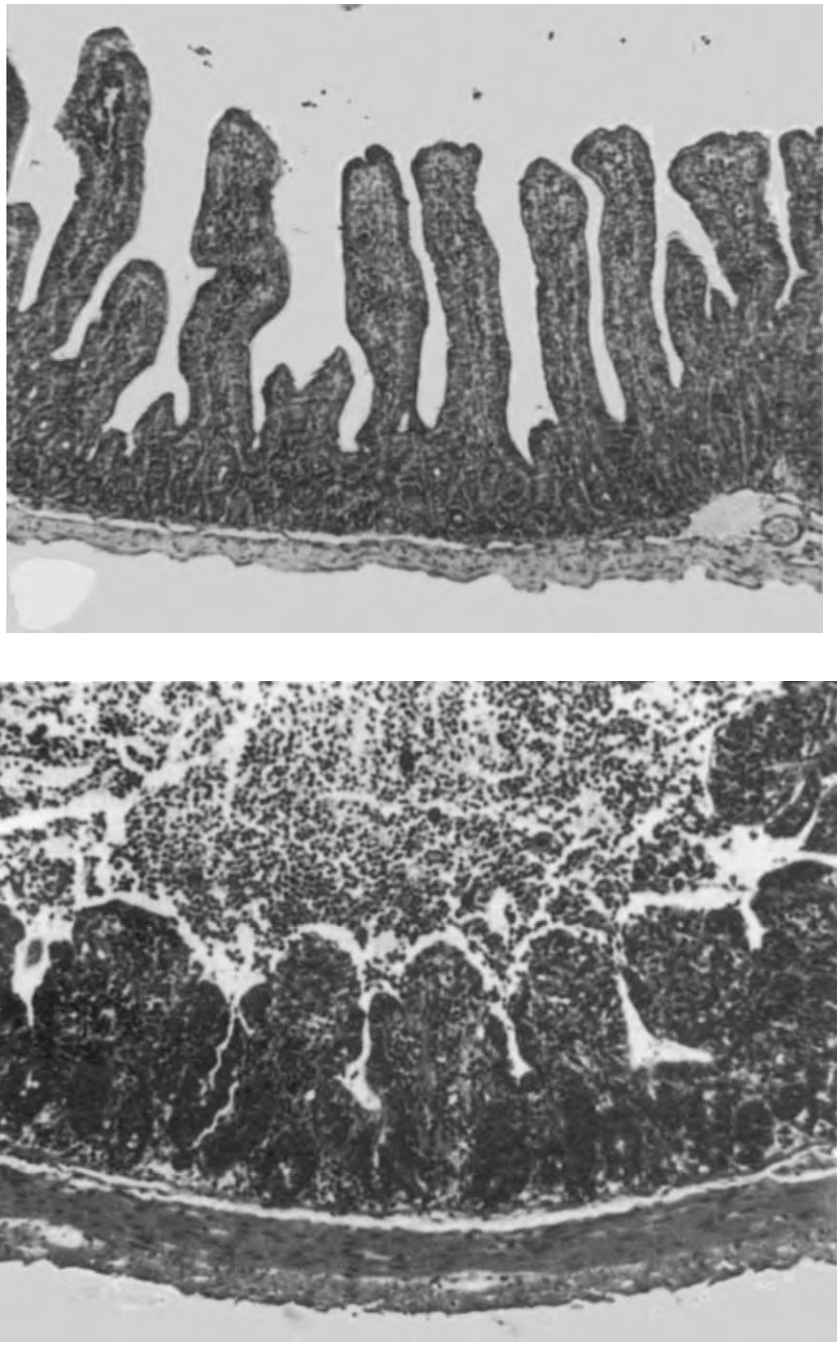

Figure 2 Effect of *Clostridium perfringens* enterotoxin on rabbit ileum. (a) Control showing typical villous morphology. (b) Ileum treated with enterotoxin for 90 min showing shortened villi denuded of epithelial cells. Reprinted from Laboratory Investigation, Baltimore, USA, with permission.

villi (Figure 2). Intestinal brush border membranes lose their characteristic folded configuration and large quantities of membrane and cytoplasm are lost to the lumen. Similar morphological changes occur after intravenous injection of the enterotoxin. The brush border (microvillous membrane) of the villus tip epithelial cells is considered to be the primary site of action of the enterotoxin.

Food Poisoning by *Clostridium perfringens*

In contrast to the effects of cholera toxin and *Escherichia coli* heat-lable enterotoxin, *C. perfringens* enterotoxin does not increase levels of cyclic adenosine monophosphate (cAMP) in intestinal mucosa that is actively secreting fluid.

Our understanding of the mode of action of the enterotoxin at the molecular level has been greatly facilitated by the use of *in vitro* model systems using various cell lines. Vero (African green monkey kidney) cells are particularly sensitive to the toxin. Rat liver, HeLa, small intestinal epithelial cells, and other cell lines have also been used as models.

The enterotoxin causes rapid membrane bleb formation in cultured Vero cells (Figure 3). Total inhibition of amino acid transport and deoxyribonucleic acid (DNA), ribonucleic acid (RNA), and protein synthesis occurs within 30 min of exposure. Within 40 min, 75% of the Vero cells detach (from the plastic culture flask) and nearly 50% are nonviable.

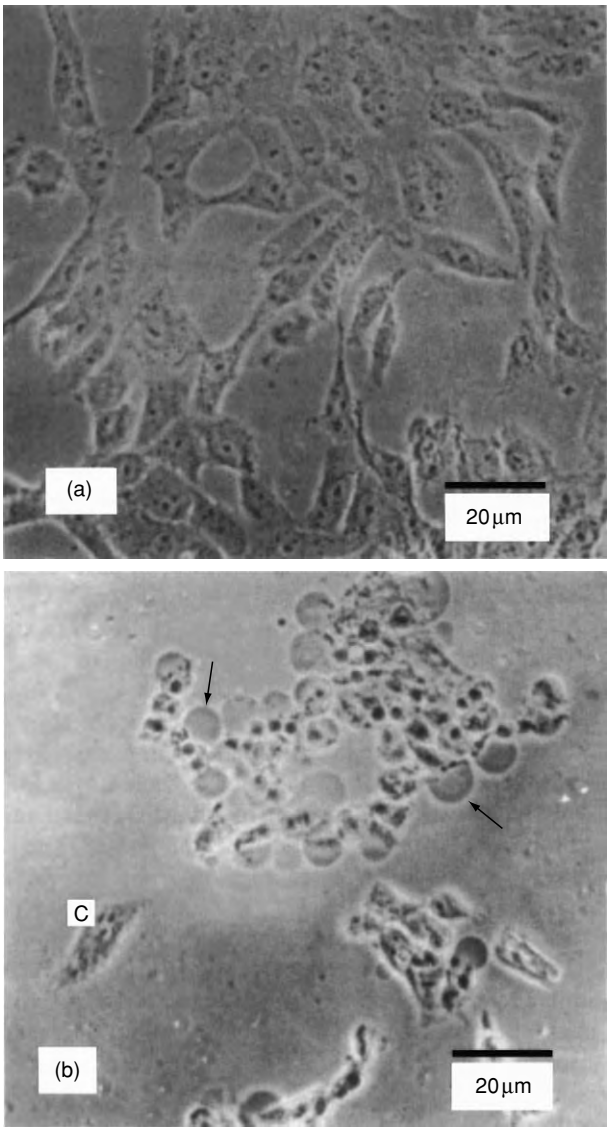

Figure 3 Effect of enterotoxin on morphology of Vero cells. (a) Control culture showing typical Vero cell morphology. (b) Vero cells after 30 min exposure to enterotoxin. Note the spherical morphology and formation of blebs (arrows). Reprinted from McClane B and McDonel J (1979) The effect of *Clostridium perfringens* enterotoxin on morphology, viability and macromolecular synthesis in Vero cells. *Journal of Cellular Physiology* 99: 193, with permission.

Disruption of the plasma membrane is the specific action of the enterotoxin in Vero cells. Binding to the cell membrane must occur and is an early event. This does not occur in naturally resistant cell lines or in cells selected for resistance to the toxin. A protein of molecular weight 50 000 has been extracted from rabbit intestinal brush border membranes. This protein inhibits biological activity of enterotoxin in Vero cells, suggesting that it has a role in enterotoxin binding. No role has been found for ganglioside GM_1 (gangliosides constitute a family of acidic glycolipids which are important membrane components) from rabbit intestinal brush border membranes in binding *C. perfringens* enterotoxin. This is in contrast to the role ganglioside GM_1 plays in binding cholera toxin.

The enterotoxin causes functional holes of defined size in Vero cell membranes. The result is an alteration of membrane permeability, causing rapid water and ion flux changes immediately after binding. Osmotic stabilizers protect against such enterotoxin-induced morphological and permeability changes, further suggesting the membrane as the site of action. With a net influx of water and ions, the membrane stretches and bleb formation results. This is followed by the loss of essential precursors through leakage. With the interruption of macromolecular synthesis the cell eventually dies. While such specific events in an *in vitro* cell model cannot be directly extrapolated to the effects of the enterotoxin in human intestinal cells, such evidence is highly suggestive that the membrane of the human small intestine is the primary site of action of *C. perfringens* enterotoxin.

Treatment of Illness

As in most cases of enterotoxin-mediated human food poisoning, symptoms are self-limiting. In the case of *C. perfringens*, symptoms subside within 1 or 2 days. Death is uncommon but has occurred in institutionalized individuals, especially the elderly. In such cases fluid and mineral replacement therapy is essential.

Prevention and Control

The factors contributing to outbreaks of foodborne illness by *C. perfringens* were identified long ago. Those most frequently cited are: (1) preparation of food too far in advance; (2) inadequate cooling; (3) storage at ambient temperature; and (4) inadequate reheating. Multiple factors are frequently cited, but

failure to refrigerate properly large portions of previously cooked foods, especially in institutional settings, invariably summarizes the events leading to an outbreak.

Spores on raw meat and poultry can survive cooking which effectively heat-activates the spores to promote germination when the product reaches a suitable temperature during cooling. Rapid, uniform cooling is therefore imperative. Gravies, broth and large pieces of meat should be cooled to 10 °C within 2–3 h. The organism can multiply rapidly in the slowly cooling masses of meat and poultry. Cooking also drives off oxygen and thereby promotes anaerobic conditions; this is especially important in liquid foods, and in rolled meats where the contaminated outside surface is rolled into the middle. Cooked, chilled foods should be reheated to a minimum internal temperature of 75 °C immediately before serving, in order to destroy vegetative cells. As noted in the preceding article, cooked meat should be kept above 62.8 °C or below 10 °C.

Human carriers are not a danger. It is impossible to prevent human carriers from handling food because most people harbor *C. perfringens* in their intestinal tract. Similarly, the organism is present in a wide variety of foods. Hence preventive measures depend largely on knowledge of proper food preparation and storage techniques, especially temperature control. Certain trade organizations related to the food service industry offer short courses dealing with proper food-handling procedures. Clearly, education of food handlers is a key aspect in the prevention of human food poisoning in general, and that caused by *C. perfringens* in particular.

See also: ***Clostridium***: Occurrence of *Clostridium perfringens*; Detection of *Clostridium perfringens*; **Contamination of Food**; **Diarrheal (Diarrhoeal) Diseases**; **Food Poisoning**: Classification

Further Reading

Labbé R (1989) *Clostridium perfringens*. In: Doyle M (ed.) *Foodborne Bacterial Pathogens*, pp. 191–234. New York: Marcel Dekker.

McClane B (2001) *Clostridium perfringens*. In: Doyle M, Beuchat L and Montville T (eds) *Food Microbiology: Fundamentals and Frontiers*, 2nd edn., pp. 351–372. Washington, DC: American Society for Microbiology Press.

Smith L and Williams B (1984) *The Pathogenic Anaerobic Bacteria*, 3rd edn. Springfield, IL: Charles Thomas.

Stringer M (1985) *Clostridium perfringens* type A food poisoning. In: Borriello S (ed.) *Clostridia in Gastrointestinal Disease*, pp. 117–141. Boca Raton, FL: CRC Press.

Occurrence of *Clostridium botulinum*

J W Austin, Bureau of Microbial Hazards, Ottawa, Canada

Introduction

Clostridium botulinum is a Gram-positive, anaerobic, rod-shaped, spore-forming bacterium that produces the most potent biological toxin known, botulinum neurotoxin. Strains of *C. botulinum* are separated into seven types, A through G, based on the serological specificity of the neurotoxin produced. Human botulism, including foodborne, wound, and infant botulism, is associated with types A, B, E, and, very rarely, F. Types C and D cause botulism in animals. To date, there is no direct evidence linking type G to disease. Based on physiological differences and 16s rRNA gene sequencing, the species is also divided into four groups: (I) all type A and proteolytic strains of types B and F; (II) all type E and nonproteolytic strains of types B and F; (III) type C and D strains; (IV) type G strains. The emphasis here will be on groups I and II since they are involved in human illness. This article will discuss the distribution of botulinum spores in the environment and in foods, the different factors affecting growth in foods, and detection of the organism and its neurotoxin in foods.

Presence of *C. botulinum* in the Environment

Spores of *C. botulinum* are commonly present in soils and sediments, but their numbers and types vary, depending on the location (**Table 1**). The possibility of contamination of food with *C. botulinum* depends on the distribution and incidence of spores in the environment where the agricultural products are grown.

C. botulinum spores are widely distributed in North America, but the spore load varies considerably, as does the predominating type. Soils in the USA east of the rise of the Rocky Mountains usually contain type B spores, while type A spores predominate in the western USA. Most US type B strains are proteolytic. Type C spores are localized in soils along the Gulf of Mexico coast. Overall, type E is found infrequently, and only in damp to wet locations. However, in the region around the Great Lakes, and particularly around Green Bay of Lake Michigan, high numbers of type E are found in shoreline and sediment samples. *C. botulinum* type E is also found in the coastal areas of the states of Washington and Alaska, and along the

Table 1 Incidence of *Clostridium botulinum* in soils and sediments[a]

Location	*Sample size (g)*	*% Positive samples*	*MPN per kg*[a]	*Type (%)*[b]				
				A	*B*	*C/D*	*E*	*F*
Eastern USA, soil	10	19	21	12	64	12	12	0
Western USA, soil	10	29	33	62	16	14	8	0
New York – Florida coast	1	4	36	8	17	42	33	0
Florida–Texas coast	1	6	62	12	4	38	46	0
Great Lakes, sediment	1	2	24	0	0	0	100	0
Green Bay, Wisconsin, sediment	1	77	1280	0	0	0	100	0
California coast north of 36 °N	5	16	29	5	5	0	90	0
California coast south of 36 °N	5	10	20	44	50	0	0	6
Alaska, coast	1	49	660	0	0	0	100	0
Canada, St Lawrence River estuary, sediment River	10	75	140	0	0	0	100	0
Canada, Gulf of St Lawrence,	10	8	8	0	0	0	100	0
Laurentian Trough, sediment Canada, Gulf of St Lawrence,	10	19	21	0	0	0	100	0
Magdalen Shelf, sediment Britain, soil	50	10	2	0	100	0	0	0
Britain coast, sediment	2	4	18	0	100	0	0	0
Ireland, soil	50	18	4	0	100	0	0	0
Iceland soil	10	3	3	0	100	0	0	0
Faroe Islands, coast	10	1	1	0	0	0	100	0
Greenland, coast	10	37	46	0	0	0	100	0
Skagerrak, Kattegat sediment	6	100	> 780	0	0	0	100	0
Finland								
Fresh water	300	61	372	0	0	0	100	0
Baltic, offshore	500	88	1020	0	0	0	100	0
Baltic, coast	300	58	500	0	0	0	100	0
Trout farm, sediments	200	68	2020	0	0	0	100	0
Netherlands, soil	0.5	94	2500	0	22	46	32	0
Norway, coast	5	7	7	0	0	0	100	0
Scandinavian coast, sediment	6	100	> 780	0	0	0	100	0
Sweden, miscellaneous soil	6	29	30	0	5		95	0
Sweden, inland	6	47	110	0	4	0	96	0
Sweden, coast	36 590	84	410	0	0	1	99	0
Danish, coast	10	67	110	0	0	0	100	0
Denmark, miscellaneous soil	10	13	15	0	93	7	0	0
Poland, Baltic coast	5	32	76	3	0	0	97	0
Czechoslovakia, soil		42		0	0	100	0	0
Switzerland, soil	12	44	48	28	83	6	0	27
Rome, Italy, soil	7.5	1	2	86	14	0	0	0
Bangladesh, sediment	10	37	19	0	0	100	0	0
China, Sinkiang, soil	10	70	25 000	47	32	19	2	0
Former United Soviet Socialist Republic, soil								
North-eastern		2		0	14	0	86	0
South-eastern		13		8	17	0	73	0
Central		14		16	77	0	6	0
North-western		4		14	0	0	86	0
Far East		14		0	8	5	87	0
Indonesia, sediment	50	21	5	0	0	100	0	0
Iran, Caspian sea, sediment	2	17	93	0	8	0	92	0
Japan, Hokkaido, soil	5–10	4	4	0	0	0	100	0
Japan, Ishikawa, soil	40–50	56	16	0	0	100	0	0
Japan, Honshu, river sediment	100	73	18	0	10	12	78	0
Java, coast	10	13	14	0	33	67	0	0
Taiwan, soil	10	78	30	34	37	4	20	5
Thailand, sediment	10	3	3	0	0	83	17	0
Argentina, soil		34		67	20	0	0	5
Brazil, soil	5	35	86	57	7	29	0	7
Costa Rica, soil	1	23	260	14	14	72	0	0
Paraguay, soil	5	24	10	14	0	14	0	71
Kenya, soil	5	25	33	89	0	11	0	0
South Africa, soil	30	3	1	0	100	0	0	0
New Zealand, sediment	20	55	40	0	0	100	0	0

Pacific, Arctic and Atlantic coastlines of Canada. The distribution of types on the Pacific coast changes with latitude; south of 36 °N, the prevalent types shift from E to A and B.

Type B predominates in the terrestrial environments of the UK, Ireland, Iceland, Denmark, and Switzerland. It is associated with most botulism outbreaks in Spain, Portugal, Italy, France, Belgium, Germany, Poland, the former Czechoslovakia, Hungary, and the former Yugoslavia, indicating wide distribution of this type in the European environment. Type B also predominates in the aquatic environments of the UK. Most European type B strains are nonproteolytic. The predominant serotype from other aquatic environments is E. The highest numbers of type E spores are found in Scandinavian waters, particularly in the Sound and Kattegat, between Denmark and Sweden. The highest overall spore concentrations detected in Europe were reported in The Netherlands, where a mixture of types was found.

Type E spores also predominate in most parts of the former Soviet Union, except for the central portion, where type B spores predominate. In general, surveys of Asia report lower numbers, with the exceptions of a high incidence of type E spores around the Caspian Sea, and a high incidence of all types in the Sinkiang district of China.

Fewer surveys have been carried out in the southern hemisphere. Spores of all types have been detected in South America; type E has been found in fish, oyster, and shrimp samples taken off the Brazilian coast. Type A spores predominate in Brazilian and Argentine soils. In the tropical regions of Asia, types C and D replace type E as the predominant type in aquatic environments.

The distribution of spore types in fish and shellfish is similar to that in the sediments from their respective areas. Because the methods of sampling fish and shellfish differ widely, it is difficult to compare numbers.

In summary, type A spores predominate in soils in the western USA, China, Brazil, and Argentina, and type B spores in the eastern USA, the UK, and much of continental Europe. However, most US type B strains are proteolytic, while most European strains are nonproteolytic. Type E is the predominant type in northern regions, and in most temperate aquatic regions and their surroundings. Types C and D are found more frequently in warmer environments. The reasons for this distribution pattern are not well understood. Type A appears to favor neutral to alkaline soils with low organic content, consistent with its virtual absence in the highly cultivated soils of the eastern USA and Europe. Type E is psychrotolerant, which undoubtedly plays a role in its prevalence in the north and in many aquatic environments. Similarly, the higher optimum growth temperature for types C and D is consistent with their presence in warmer environments.

Presence of *C. botulinum* in Foods

The sequence of events resulting in foodborne botulism begins when *C. botulinum* contaminates a food. This often occurs during growth or harvesting, and is most likely to occur when a product originates in an environment with a high incidence of spores. However, contamination can also occur during or after processing. There have been considerably fewer surveys of foods for contamination with *C. botulinum* than environmental surveys, and they have focused primarily on fish, meats, and infant foods, particularly honey.

Fish may become contaminated with spores of *C. botulinum* in their environment, or during processing and handling. The presence of *C. botulinum*, mostly type E, in fish is readily demonstrated, although the incidence is lower than in environmental surveys. The average level of contamination of prepared fish in the USA appears to be higher than in Europe and Asia. The highest contamination level reported was in salted fish from the Caspian Sea.

The level of contamination of meats is generally low. It appears lower in North America, where the average most probable number (MPN) is about 0.1 spore kg^{-1}, than in Europe, where the average MPN is about 2.5 spores kg^{-1}. The types most often associated with meats are A or B.

C. botulinum is common in soil and organic fertilizers. *C. botulinum* types A or B may be present on fruits and vegetables, particularly those in close contact with the soil. One product of particular concern is cultivated mushrooms, in which up to 2.1×10^3 type B spores kg^{-1} have been detected.

Spores in honey and other infant foods pose a unique hazard because, in some infants, the spores are able to colonize the intestines, produce toxin, and cause infant botulism. By 1984, honey had been implicated as the likely source of botulinum spores in 20 cases of infant botulism in California. Surveys for the presence of *C. botulinum* spores in honey

[a]Most probable number (MPN) calculated using the Halvorson-Zeigler equation $\log_e (n/q)$, where n = the total number of samples analyzed and q = the number of nontoxic samples.

[b]As percentage of the types identified.

Adapted from Hauschild AHW (1989) *Clostridium botulinum*. In: Doyle MP (ed.) *Foodborne Bacterial Pathogens*, pp. 111–184. New York: Marcel Dekker.

suggest that the botulinum spore level in random samples of honey is in the order of 1–10 spores kg^{-1}. However, in honey samples associated with infant botulism the level is approximately 10^4 spores kg^{-1}. Other infant foods have also been examined. While *C. botulinum* has been detected in samples of corn syrup and rice cereal, exposure of infants to botulinum spores via these foods seems to be minimal as the spore levels are low, and spores are unlikely to multiply during manufacture and storage.

Other foods examined, including dairy products, vacuum-packed products, and convenience foods, show a very low incidence of *C. botulinum* spores.

Factors Affecting Growth and Toxin Production in Foods

The main factors affecting growth of *C. botulinum* in foods are temperature, pH, water activity (a_w), redox potential, added preservatives, and other microorganisms. The growth of types A, B, E, and F, of groups I and II, has been intensively studied. Traditionally, food microbiologists have established maximum and/or minimum limits for these parameters which would permit growth of *C. botulinum* (Table 2), and these limits have often been used in the control of *C. botulinum*. However, these factors seldom function independently; usually, they act in concert, often having synergistic or additive effects. Models incorporating pH, a_w, and temperature are available to predict the growth of *C. botulinum* under various combinations of conditions.

Temperature

Because foods are generally stored at low temperatures, studies have focused on determining the minimum temperatures permitting growth. The established lower limits are 10 °C for group I and 3.0 °C for group II. However, these limits apply to relatively few strains and depend on otherwise optimum growth conditions. Production of toxin generally requires several weeks at the lower temperature limits.

Table 2 Properties of group I and II *Clostridium botulinum*

Property	*Group I*	*Group II*
Toxin types	A, B, F	B, E, F
Minimum temperature for growth (°C)	10	3
Maximum temperature for growth (°C)	48	45
Minimum pH for growth	4.6	5
Inhibitory (NaCl) (%)	10	5
Minimum a_w for growth	0.94	0.97
$D_{100\,°C}$ of spores (min)	25	< 0.1
$D_{121\,°C}$ of spores (min)	0.1–0.2	< 0.1

The optimum growth temperature is in the range 35–40 °C for group I organisms, and in the range 25–30 °C for those of group II. The upper temperature limits for group I and group II organisms are approximately 45–50 °C and 40–45 °C, respectively.

pH

It is generally accepted that the minimum pH permitting growth of *C. botulinum* group I is 4.6, and many regulations worldwide use this limit. For group II, the limit is about pH 5.0. The upper pH limits for growth are in the pH range 8–9, but are of little practical consequence. Many fruits and vegetables are sufficiently acidic to inhibit *C. botulinum* by their pH alone, while other products, such as marinated vegetables, are preserved by added acidulants. Several factors influence the acid tolerance of *C. botulinum*, including strain, substrate, temperature, nature of the acidulant, the presence of preservatives, a_w, and redox potential. The growth of acid-tolerant microorganisms such as yeasts and molds may raise the pH in their immediate vicinity to a level that permits growth of *C. botulinum*. *C. botulinum* can also grow in some acidified foods if excessively slow pH equilibration occurs. While high concentrations of proteins in laboratory media appear to protect *C. botulinum* and permit growth at pH levels below 4.6, this does not occur in foods preserved by acidity. Current regulations stipulating a minimal acidity of pH 4.6 for the control of *C. botulinum* are therefore valid.

Salt and a_w

Salt (NaCl) is one of the most important factors controlling *C. botulinum* in foods. Its inhibitory effect is primarily due to the depression of a_w and consequently to its concentration in the aqueous phase, also called the brine concentration (%brine = $NaCl \times 100/(\%H_2O + \%NaCl)$). Under otherwise optimal conditions, the growth-limiting brine concentrations are about 10% for strains of group I and 5% for strains of group II. These concentrations correspond well to the limiting a_w of 0.94 for group I and 0.97 for group II in foods where NaCl is the main a_w-depressant. The type of solute used to control a_w may influence these limits. Generally, NaCl, potassium chloride, glucose, and sucrose show similar patterns, while the use of glycerol reduces the growth-limiting a_w level by up to 0.03 units. The limiting a_w may be raised significantly by other factors, such as increased acidity or preservatives.

Redox Potential

C. botulinum grows optimally at an oxidation–reduction potential (E_h) of −350 mV, but growth

initiation may occur in the E_h range of +30 to +250 mV. The presence of other inhibitory factors lowers this upper limit. Once growth is initiated, the E_h declines rapidly.

Modified atmosphere packaging is being used increasingly to extend the shelf-life and improve the quality of foods. Depending on the atmosphere and the food, growth of *C. botulinum* may be inhibited or stimulated. Many studies have shown that *C. botulinum* will grow as well in foods packed in air as in vacuum-packed foods; the oxygen in the package head space can be rapidly depleted by respiration of background microorganisms or, in the case of modified-atmosphere packaging vegetables, by the plant tissue itself. The safety of different atmospheres with respect to *C. botulinum* should be carefully investigated before use.

Preservatives

Nitrite is important in producing the characteristic color and flavor of cured food products, but its most important role is the inhibition of *C. botulinum*. It is more effective with decreasing pH, increasing NaCl content, and addition of ascorbate or isoascorbate to the food product. Nitrite reacts with many cellular constituents and appears to inhibit *C. botulinum* by more than one mechanism, one of which is probably its reaction with essential iron-sulfur proteins to inhibit the phosphoroclastic system which supplies the cell with energy. The reactions of nitrite, or nitric oxide, with secondary amines in meats to produce nitrosamines, some of which are carcinogenic, has led to regulations limiting the amount of nitrite used.

Other compounds which are active against *C. botulinum* include sorbates, parabens, nisin and other bacteriocins, phenolic antioxidants, polyphosphates, ascorbates, ethylenediaminetetraacetic acid (EDTA), metabisulfite, *n*-monoalkyl maleates and fumarates, and lactate salts. Extracts of certain foods, including mace, bay leaf, and nutmeg extracts, oils from garlic, onion, black pepper, clove, cinnamon, and origanum have all been shown to inhibit the growth of *C. botulinum*. The use of natural or liquid smoke has a significant inhibitory effect against *C. botulinum* in hot smoked fish, reducing the inhibitory water-phase NaCl concentration from 3.7% to 2.9–3.0%.

Other Microorganisms

Other microorganisms have a significant role in the control of *C. botulinum* in foods. Acid-tolerant yeasts and molds may make the environment more favorable for growth of *C. botulinum*. Other microorganisms may inhibit *C. botulinum*, either by changing the environment, or by producing specific inhibitory substances, or both. Lactic acid bacteria, including *Lactobacillus, Pediococcus*, and *Streptococcus* spp., can inhibit growth of *C. botulinum* in meat products, largely by reducing the pH, perhaps also by the production of bacteriocins. Growth of *Bacillus* spp. that produce antimicrobials has been shown to inhibit growth of group II *C. botulinum* in foods. The use of lactic acid bacteria and a fermentable carbohydrate, the 'Wisconsin process,' has been permitted for producing bacon with a decreased level of nitrite in the USA. The growth of other microorganisms may also protect the consumer by causing spoilage that would make a toxic product less likely to be consumed.

Thermal Inactivation

C. botulinum spores of group I are very heat-resistant. *D* values (the time required to inactivate 90% of the population at a given temperature) vary considerably among *C. botulinum* strains and depend on how the spores are produced and treated, the heating environment, and the recovery system. Spores of types A and B are the most heat-resistant, having $D_{121\,^\circ C}$ values in the range 0.1–0.2 min. These spores are of particular concern in the sterilization of canned low-acid foods, and the canning industry has adopted a *D* value of 0.2 min at 121 °C as a standard for calculating thermal processes. The *z* value is defined as the temperature change necessary to bring about a 10-fold change in the *D* value. For the most resistant strains, *z* values are approximately 10 °C, which has also been adopted as a standard. Actual *z* values may vary by several degrees. Despite the variations in *D* and *z* values, the adoption of a 12*D* process ('botulinum cook') as the minimum thermoprocess applied to commercial canned low-acid foods by the canning industry has insured the production of safe products. The botulinum cook will produce a product that is safe but not always sterile, depending upon the presence and levels of other, more heat-resistant, spores. The thermal process of commercially canned foods usually uses the more heat-resistant organisms, such as *Bacillus stearothermophilus*, as target organisms for the process.

Although strains of group 2 are considerably less heat-resistant ($D_{100\,^\circ C} < 0.1$ min) than those of group 1, their survival in pasteurized, refrigerated products is of concern because of their ability to grow at refrigeration temperatures. $D_{82\,^\circ C}$ values of type E in neutral phosphate buffer are generally in the 0.2–1.0 min range. While the *D* values are often higher in foods, the *z* values are essentially the same.

Inactivation by Irradiation

C. botulinum spores are probably the most radiation-resistant spores of public health concern. *D* values (irradiation dose required to inactivate 90% of the population) of group 1 strains at −50 to −10 °C are in the range 2.0–4.5 kGy in neutral buffers and in foods. Spores of type E are only marginally more sensitive, having *D* values in the range 1–2 kGy. The goal of radappertization is to reduce the number of viable spores of the most radiation-resistant *C. botulinum* by 12 $\log_{10}$ cycles. *D* values are affected by any pretreatment of spores, the presence of oxygen, irradiation temperature, and irradiation and recovery environments. Generally, spores are more sensitive in the presence of oxygen or preservatives and at temperatures above 20 °C.

Detection of *C. botulinum* and its Toxins in Foods

It is not necessary to isolate *C. botulinum* in pure culture from foods in order to demonstrate its presence. Usually, the sample is inoculated into a non-selective enrichment medium. If the neurotoxin is present in the culture after incubation, the toxin-producing organism must have been present originally.

Botulinum neurotoxins are recognized by their lethal action in mice and neutralization with specific antisera. While several *in vitro* tests have been developed, they are generally less sensitive and less specific than the mouse bioassay which remains the standard. The sample, or an extract prepared by homogenizing it in a slightly acidic buffer, is clarified by centrifugation and normally filter-sterilized. Trypsin treatment may be required to activate low levels of toxin from nonproteolytic strains. The prepared sample is injected intraperitoneally into mice with and without neutralization with antitoxin. Typical signs of botulism are ruffled fur, pinched waist, laboured breathing, limb paresis, and general paralysis before death. Definitive results are obtained if mice injected with untreated sample display symptoms and die within 72 h, while mice injected with neutralized sample do not display symptoms and survive.

Common enrichment media for detecting viable *C. botulinum* are cooked meat medium (CMM), CMM glucose, chopped meat glucose starch (CMGS) medium, and trypsin–peptone–glucose–yeast extract (TPGY) broth, to which trypsin may be added (TPGYT). Trypsin is necessary to activate the toxin produced by group II organisms, and may also inactivate potential inhibitors of *C. botulinum* such as bacteriocins in mixed cultures. While foods may be inoculated directly, the sediments of centrifuged samples are preferred because potential growth inhibitors are removed. At least two tubes of media are inoculated. One is heated at 75–80 °C or 60 °C, depending on whether the suspected type belongs to group I or II, to select for spores. Alternatively, spores of group II may be selected by holding samples in 50% alcohol for 1 h before inoculation. The other tube is incubated without any heating to allow development of vegetative *C. botulinum* cells in case few or no spores are present. Adding lysozyme to the medium may increase recovery of heat-injured spores. *C. botulinum* is identified after incubation of the enrichment medium by toxin analysis of the supernatant fluid, as outlined above.

See also: **Canning**: Principles; **Chill Foods**: Effect of Modified-atmosphere Packaging on Food Quality; **Chilled Storage**: Microbiological Considerations; Packaging Under Vacuum; Use of Modified-atmosphere Packaging; **Fermented Foods**: Origins and Applications; Fermentations of the Far East; **Heat Treatment**: Chemical and Microbiological Changes; **Irradiation of Foods**: Basic Principles; **Lactic Acid Bacteria**; **Microbiology**: Classification of Microorganisms; Detection of Foodborne Pathogens and their Toxins; **Preservatives**: Classifications and Properties; Food Uses; **Water Activity**: Principles and Measurement

Further Reading

Austin JW, Dodds KL, Blanchfield B and Farber JM (1998) Growth and toxin production by *Clostridium botulinum* on inoculated fresh-cut packaged vegetables. *Journal of Food Protection* 61: 324–328.

Collins MD and East AK (1998) Phylogeny and taxonomy of the food-borne pathogen *Clostridium botulinum* and its neurotoxins. *Journal of Applied Microbiology* 84: 5–17.

Daifas DP, Smith JP, Blanchfield B and Austin JW (1999) Effect of pH and CO_2 on growth and toxin production by *Clostridium botulinum* in English-style crumpets packaged under modified atmospheres. *Journal of Food Protection* 62: 1157–1161.

Daifas DP, Smith JP, Blanchfield B and Austin JW (1999) Growth and toxin production by *Clostridium botulinum* in English-style crumpets packaged under modified atmospheres. *Journal of Food Protection* 62: 349–355.

Dodds KL and Austin JW (1997) *Clostridium botulinum*. In: Doyle MP, Beuchat LR and Montville TJ (eds) *Food Microbiology. Fundamentals and Frontiers*, pp. 288–304. Washington: ASM Press.

Graham AF, Mason DR, Maxwell FJ and Peck MW (1997) Effect of pH and NaCl on growth from spores of non-proteolytic *Clostridium botulinum* at chill temperature. *Letters in Applied Microbiology* 24: 95–100.

Hatheway CL and Ferreira JL (1996) Detection and identification of *Clostridium botulinum* neurotoxins.

Advances in Experimental Medicine and Biology 391: 481–498.
Hauschild AHW (1989) *Clostridium botulinum*. In: Doyle MP (ed.) *Foodborne Bacterial Pathogens*, pp. 111–184. New York: Marcel Dekker.
Hauschild AHW and Dodds KL (1993) *Clostridium botulinum. Ecology and Control in Foods*. New York: Marcel Dekker.
Hielm S, Hyytia E, Andersin AB and Korkeala H (1998) A high prevalence of *Clostridium botulinum* type E in Finnish freshwater and Baltic Sea sediment samples. *Journal of Applied Microbiology* 84: 133–137.
Lyver A, Smith JP, Austin J and Blanchfield B (1998) Competitive inhibition of *Clostridium botulinum* type E by *Bacillus* species in a value-added seafood product packaged under a modified atmosphere. *Food Research International* 31: 311–319.
Lyver A, Smith JP, Nattress FM, Austin JW and Blanchfield B (1998) Challenge studies with *Clostridium botulinum* type E in a value-added surimi product stored under a modified atmosphere. *Journal of Food Safety* 18: 1–23.
Roberts TA, Baird-Parker AC and Tompkin RB (eds) (1996) *International Commission on Microbiological Specifications for Foods: Microorganisms in Foods*, vol. 5. *Microbiological Specifications of Food Pathogens*. London: Blackie Academic and Professional.

Botulism

J W Austin, Bureau of Microbial Hazards, Ottawa, Canada

Introduction

Botulism is a neuroparalytic disease resulting from neurotoxin-mediated inhibition of acetylcholine release from peripheral nerve synapses, including neuromuscular junctions. This results in a descending symmetrical flaccid paralysis beginning with the cranial nerves. Death may result from asphyxiation caused by paralysis of the diaphragm. Botulinum neurotoxin is produced by *Clostridium botulinum* and, rarely, by *C. butyricum* or *C. barati*. Human botulism is classified into four categories. Foodborne botulism, the most common form worldwide, is caused by ingestion of food contaminated with preformed botulinum neurotoxin, usually type A, B, or E. Infant botulism, first recognized in 1976 and now the most common form of botulism in the USA, is caused by ingestion of viable spores which colonize the intestinal tract of infants less than 1 year of age and produce toxin locally. The environment appears to be the most common source of spores for infant botulism, and honey is the only food which has been associated with infant botulism. Would botulism is due to infection of a wound with spores of *C. botulinum* which grow and produce toxin *in situ*. Wound botulism associated with injection drug use has increased dramatically since 1994. Most cases of wound botulism occur among heroin users in California who inject black tar heroin subcutaneously. The fourth category, adult infectious botulism, includes adult cases with intestinal colonization and toxemia, and resembles infant botulism. This entry will discuss the epidemiology, clinical aspects, and prevention of foodborne botulism, and give a brief overview of the structure and mode of action of the neurotoxin.

Epidemiology

Since epidemiological data depend upon the reporting systems, and the thoroughness of these systems varies between countries, these data are scarce in large areas of the world. **Table 1** shows the data for countries with frequent outbreaks. Unrecognized and misdiagnosed cases of botulism do occur, as shown by a 1985 outbreak in Vancouver, Canada, where the initial diagnoses for 28 patients included psychiatric illness, viral syndrome, laryngeal trauma, and overexertion.

Incidence in North America

In Canada, since 1985, 4.4 outbreaks of foodborne botulism occur annually, with an average of 2.75 cases per outbreak. Most botulism outbreaks in Canada have occurred in northern and west-coast native communities. The foods involved were mainly raw meats from marine mammals, fermented meats such as muktuk (meat, blubber, and skin of the beluga whale), raw fish, or fermented salmon eggs. Type E was implicated in almost every case. Toxin production in raw meats can occur because the meats are often held at ambient temperature for some time. The problem with so-called fermented products is that the level of fermentable carbohydrates is too low to ensure a sufficiently rapid pH reduction to prevent growth of *C. botulinum*. Commercial products have been implicated in four incidents in Canada since 1971: bottled marinated mushrooms imported from the USA; bottled garlic in oil, also imported from the USA but temperature-abused locally; inhouse bottled mushrooms; and temperature-abused pâté produced in Quebec.

In the USA, 9.4 outbreaks of botulism occur annually, with an average of 2.5 cases per outbreak. Alaska accounts for 16.2% of all the outbreaks in the USA occurring from 1950 to 1996. All of the Alaskan outbreaks have involved native Alaskans; they

Table 1 Recorded outbreaks of foodborne botulism

Country	*Period*	*No. of outbreaks*	*No. of cases*	*Average cases per year*	*Predominant type*	*Predominant food implicated*
Russia	1988–1992	1327	2300	460	B	Home-preserved mushrooms, fish
Poland	1993–1997	432	565	113	B	Home-preserved meat
China	1958–1989	745	2861	90	B	Homemade fermented bean curds
Italy	1988–1998		412	37	B	Homemade vegetable preserves in oil or water
USA	1950–1998	476	1134	23	A	Vegetables
Hungary	1987–1996	117	191	19	B	Meats
Germany	1988–1998	NA	177	16	B	Meats
Argentina	1979–1997		277	15	A	Vegetables
Japan	1951–1987	97	479	13	E	Fish
Canada	1985–1999	66	179	12	E	Fermented marine mammal meat
Spain	1988–1998	NA	92	8	B	Home-preserved vegetables
France	1988–1997	93	72	7.2	B	Home-cured ham
Portugal	1970–1984	13	50	3	B	Home-smoked ham
Norway	1975–1997	10	22	1	E	*Rakfish* – traditional fermented fish
Belgium	1988–1998		10	1	B	Meats

NA, not available.

involved raw or fermented native foods, and type E was usually implicated. In continental USA, the situation is quite different. Most of the implicated foods were home-preserved vegetables. The relative prevalence of type A and type B outbreaks was about equal in the east, but type A outbreaks were much more prevalent in the west, where type A spores predominate in the environment. Most botulism outbreaks in the USA are caused by home-processed foods. Commercially processed foods account for only 7% of the botulism cases occurring in the USA since 1970. Temperature abuse of either food ingredients or the final product was often the problem.

Incidence in Europe

Poland reported by far the highest number of cases annually. This probably reflects both a high local incidence of botulism and a very thorough surveillance system. In Poland and several other European countries, including Italy, France, Spain, Germany, Hungary, Portugal, Czechoslovakia, and Belgium, the predominant type involved was B (Table 1). In Poland, the use of weck jars (weckglas) to seal cooked meats (mostly pork) hermetically under a vacuum led to a significant number of botulism cases. Other common vehicles of botulism in Poland include sausages, ham, headcheese (brawn), liver sausage, and bacon. The incidence of botulism was highest in Poland during the period of social change in the 1980s, reaching a peak of 738 cases in 1982. The number of botulism cases in Poland is seasonal, with the greatest incidences in May to August and in December. This may correspond to increased slaughtering of pigs in private farms and increased preparation of home-preserved 'wecked' products. Commercial foods, usually canned meat and fish, were responsible for 26% of cases in Poland.

The foods most frequently implicated, except in Italy and Spain, were home-preserved meats such as ham, fermented sausages, or canned products. Of the implicated foods in France, 12% were of commercial origin, but the manufacturers were generally small local establishments. In Italy and Spain, the most commonly implicated foods were home-preserved vegetables, most of which in Italy had been preserved in oil or water; in Spain, all of the incriminated vegetables had been canned. Commercial products were implicated in several outbreaks in Italy and included canned tuna fish in oil, dairy products (mascarpone cheese), low-acid canned vegetables, and pasteurized eggplant in oil. Scandinavian countries recorded fewer outbreaks, and these were mainly associated with fish and type E. Type E botulism has been reported in Greenland after consumption of raw seal meat and raw seal intestines. An outbreak of type B botulism, involving 12 cases, caused by commercial salted raw pork occurred in Switzerland between 31 December 1993 and 12 January 1994. Botulism outbreaks have been rare in the UK; however, a major outbreak in 1989 involved 27 cases, with one death. Type B toxin had been produced in a hazelnut purée which was then used to flavor a yogurt produced by a local dairy. Other European countries reported either few or no botulism outbreaks.

Incidence in the Russian Federation

The Russian Federation recorded 1327 outbreaks of botulism, involving 2300 cases and 182 deaths from 1988 to 1992. Most of these outbreaks were caused by type B, and most outbreaks were caused by

home-preserved mushrooms. Other foods commonly involved in botulism outbreaks include fish and pork. During this period, commercial foods, usually canned meats and vegetables, were implicated in one to five outbreaks annually. Incidents associated with fish occurred mainly in regions around the Black Sea, the Sea of Azov, the Caspian Sea, and Lake Baikal.

Incidence in Asia

Only a few countries in Asia reported outbreaks of foodborne botulism. Israel reported an outbreak in 1987 which affected eight people, with one fatality. The incriminated food was kapchunka (salted, uneviscerated whitefish) contaminated with type E toxin. All the kapchunka consumed by the patients had been purchased in New York City. While the high salt concentration of the kapchunka would have prevented growth of *C. botulinum* in the flesh of the fish, *C. botulinum* likely grew in the intestines of the fish and neurotoxin may have diffused into the flesh.

Iran reported a high incidence of botulism. Of the 314 cases recorded between 1972 and 1974, 170, comprising 63 outbreaks, were investigated. The majority of outbreaks (97%) were associated with type E and with fish or fish products. However, fleshy portions of fish were responsible for only 10% of the outbreaks; the other 90% were caused by fish eggs. These eggs (ashbal) are salt-cured for several months and then eaten without further treatment. A large outbreak, caused by locally made cheese stored under oil and involving 27 cases with one fatality, occurred in a northern district of Iran in 1997.

China has recorded 745 botulism outbreaks from 1958 to 1989. Most were associated with type A, followed by types B and E. The north-western province of Xinjiang recorded the majority of outbreaks, which were usually type A. Typically, the incriminated food was fermented bean curd. In the northern provinces of Ningxia and Hebei, outbreaks associated with type B predominated. Type E predominated in Xizang province. Neurotoxigenic *C. butyricum* was the cause of an outbreak of foodborne type E botulism in Guanyun, Jiangsu province, China, in January 1994. Six cases of clinically diagnosed botulism were reported after consumption of salted and fermented paste made of soybeans and wax gourds. Type E toxin was detected in the implicated food and *C. butyricum*, but not *C. botulinum*, was isolated. Subsequently, type E neurotoxin-producing *C. butyricum* was isolated from soil from four sites in an area near to the sites where the type E outbreak occurred. *C. butyricum* has also been isolated from sevu (crisp made of gram flour) implicated in a 34-case outbreak of botulism in India in September 1996.

The majority of Japanese outbreaks occurred in northern areas and were associated with type E and fish or fish products. Izushi was the food most often implicated. To prepare izushi, fleshy pieces of fish are soaked in water for a few days, and then packed tightly into a tub with cooked rice, vegetables, salt, vinegar, and spices and left to ferment, often for 3 weeks or longer. Izushi is eaten without further cooking. Two outbreaks in Japan were associated with commercial food. A type A outbreak was caused by vacuum-packaged, stuffed lotus rhizome and involved 36 cases with 11 deaths. An outbreak caused by imported bottled caviar with 21 cases and three deaths was due to type B toxin. Taiwan has also reported a few botulism outbreaks, but reports from other Asian countries are rare.

Incidence in Other Areas

Argentina is the only country in the southern hemisphere which reported a substantial number of botulism outbreaks. These occurred mostly in the provinces of Mendoza and Buenos Aires, and at a latitude between 30° and 40° S. Most outbreaks were associated with type A and the implicated foods were usually vegetables. A recent outbreak of type A botulism among bus drivers in Buenos Aires was caused by consumption of matambre, a meat roll containing vegetables and eggs. The matambre was wrapped in a heat-shrunk plastic wrap and stored with inadequate refrigeration. Mexico, Guatemala, Venezuela, Peru, Brazil, and Chile have all reported few outbreaks.

Chad, Egypt, Kenya, Madagascar, and Zimbabwe are the only African countries with confirmed outbreaks. In April 1991, an outbreak with 91 reported cases of type E botulism and 18 reported deaths occurred in Egypt. The food implicated in this outbreak was faseikh (uneviscerated, salted mullet fish) originating from a single shop in Cairo. In Kenya, two outbreaks were caused by native foods: one by sour milk prepared in a gourd, and one by consumption of raw termites. One of the two outbreaks reported from Madagascar was unusual; the numbers involved were high, and type E toxin was associated with a meat product. About 60 people were involved, with 30 deaths, and locally manufactured bologna was the vehicle.

Since 1942, Australia has recorded only six outbreaks of foodborne botulism, and none since 1991. One outbreak of type A botulism was recorded in New Zealand, caused by home-bottled fermented mussels and watercress, a traditional Maori dish.

Unconfirmed Outbreaks

Several outbreaks of foodborne botulism have been reported in the media and distributed over the

internet by ProMED-mail, the Program for Monitoring Emerging Diseases. ProMED-mail is a program of the International Society for Infectious Diseases (ISID) and acts as a global electronic reporting system for outbreaks of emerging infectious diseases. These outbreaks are all unconfirmed and the only source for these outbreaks is media reports.

Two large unconfirmed botulism outbreaks have been reported in the north African countries Algeria and Morocco. Algerian state-run radio was cited as reporting that an outbreak of botulism killed 17 people in July 1988 in eastern Algeria. One of the foods suspected in this outbreak was a processed meat known as kashir. Another unconfirmed outbreak was reported by Associated Press, stating that mortadella sausage was responsible for a suspected botulism outbreak in Morocco during September 1999 with at least 10 deaths.

Reuters news agency had reported an unconfirmed botulism outbreak involving 90 cases with four deaths from consumption of fish at a wedding reception in the former Soviet republic of Azerbaijan. The Russian news agency Itar-Tass has reported an outbreak of botulism in southern Siberian Buryat, with 48 cases and four deaths after consumption of fish from Lake Baikal. More than 70 cases of botulism, including eight deaths, had been reported in Buryatia in 1999.

Conclusions Regarding Epidemiology

Most botulism outbreaks have occurred in the northern hemisphere, particularly in countries north of the Tropic of Cancer. Argentina is the only country in the southern hemisphere which has reported several outbreaks. Seasonal trends in outbreak frequencies are apparent in some areas. In Canada, Alaska, and Poland most outbreaks occur from May until October, whereas in China most outbreaks occur in the late winter and spring. There is a strong association between the prevalent outbreak type and the prevalent environmental type, as well as the implicated food. In colder regions, Canada, Alaska, Greenland, Scandinavia, parts of Russia, Iran, and northern Japan, type E causes most botulism outbreaks and is the prevalent environmental type. The implicated foods are usually fish or marine mammals. In central Europe, type B causes most outbreaks and is the prevalent environmental type. Meats, particularly home-cured smoked ham, are the major cause of botulism. In the western USA, Argentina, and China, type A causes the majority of outbreaks and is the prevalent environmental isolate. In these areas, the most frequently implicated foods are vegetables.

Clinical Aspects

Symptoms

The disease may vary from a mild illness which may be overlooked or misdiagnosed to a serious disease which may be fatal within 24 h. The onset of symptoms typically occurs 12–36 h after ingestion of toxin, with a range from a few hours to 14 days. Usually, the earlier symptoms appear, the more serious the disease. The first symptoms are generally nausea and vomiting. Mainly neurological signs and symptoms appear next, including visual impairments (blurred or double vision, ptosis, fixed and dilated pupils), loss of normal mouth and throat functions (difficulty in speaking and swallowing, dry mouth, throat and tongue, a sore throat), general fatigue and lack of muscle coordination, and respiratory impairment. Other gastrointestinal symptoms include abdominal pain, diarrhea, and constipation. Diarrhea occurs relatively early in the course of the disease, whereas constipation persists in the advanced stages. Respiratory failure and airway obstruction are the main causes of death. Fatality rates in the first half of the century were about 60%, but with the availability today of antisera and modern respiratory support systems, they have decreased to about 10%.

Treatment

Initially, treatment of botulism is directed toward removing or inactivating the toxin by (1) neutralizing circulating toxin with antiserum; (2) enema or treatment with cathartics to remove residual toxin from the bowel; and (3) in the absence of vomiting, gastric lavage or treatment with emetics. Induced emesis or gastric lavage is recommended if exposure has occurred within several hours. Treatment with antiserum is most effective in the early stages of the illness. The impact of antiserum is obvious from the Chinese data; before the availability of antisera in 1960, the death rate in China was approximately 50%, but it was only 8% in the nearly 4000 patients who received antitoxin. Subsequent treatment is mainly to counteract the paralysis of the respiratory muscles by artificial ventilation.

Diagnosis

The initial diagnosis of foodborne botulism is based on the patient's signs and symptoms, and perhaps food history. It must be confirmed by detecting toxin or viable *C. botulinum* in a suspect food or clinical sample, or by epidemiological association with a laboratory-confirmed case. Serum, feces, enema fluid, stomach contents, and autopsy sections of the small and large intestines, and of the liver, are

suitable specimens for neurotoxin detection. Except for serum, these specimens are also suitable for detecting viable *C. botulinum*. The methods for detecting the neurotoxin and *C. botulinum* are given in the previous article ***Clostridium***: Occurrence of *Clostridium botulinum*. Other than neutralizing any acidic samples, little special treatment is required. Occasionally, extracts prepared from feces cannot be filter-sterilized. In that case, tetracycline should be added to 200 p.p.m. to control infection of the mouse.

Case History

One of the most severe, nonfatal cases of botulism was documented in a patient in the UK. Within approximately 10 h of eating the suspect food, he had blurred and double vision followed by nausea and vomiting, difficulty in swallowing and talking, dryness of the mouth, and arm weakness. He suffered a respiratory arrest on arrival at the hospital and was intubated. A progressive flaccid paralysis ensued. He was given polyvalent antitoxin, a tracheostomy was done and total parenteral nutrition (TPN) was started. There was little change in his condition for 100 days. His ptosis did not resolve until 46 days after onset of symptoms. TPN was stopped on day 158, and respiratory support was withdrawn after 173 days. He was not discharged until after 237 days.

Prevention

Usually, the preservation of high-moisture foods is geared toward control of *C. botulinum*, which usually involves inhibition rather than destruction. Such control generally also insures control of other foodborne pathogens and of many spoilage microorganisms. The effects of different factors on the growth of *C. botulinum* in foods are described in the previous entry ***Clostridium***: Occurrence of *Clostridium botulinum*. Control of *C. botulinum* in commercial foods is generally achieved by one of the following methods:

1. Low-acid shelf-stable canned foods are preserved by a full thermal process.
2. Shelf-stable canned cured meats are preserved by a combination of thermal processing and addition of salt and nitrite.
3. Canned acid foods are preserved by a pasteurizing thermal process and acidity.
4. Products such as dry fermented sausages are preserved by reduced water activity (a_w) and pH, and added nitrite.
5. Packaged raw or cooked meats or fish and shellfish are preserved by refrigeration alone < 3°C.
6. Many meat and fish products are preserved by a combination of added salt and refrigeration.
7. Many cured meat products are preserved by a combination of added salt and nitrite, and refrigeration.
8. Vacuum-packaged smoked fish are preserved by a combination of thermal processing, added salt, smoking, and refrigeration.

Foodborne botulism, caused by home-prepared foods, can be prevented by strict adherence to proper home-canning instructions. The type of container, levels of acidity, and salt and heat treatments are especially important for home-canning. All home-canned foods should be brought to a boil before they are eaten. An effective botulinal toxoid is available from the US Centers for Disease Control for immunization. Immunization of high-risk populations, such as Alaskan and Canadian Inuits, has been repeatedly proposed, but never implemented. At present, only laboratory workers at risk are immunized.

Neurotoxin

As previously stated, seven serologically different neurotoxins are produced by various strains of *C. botulinum*. Some strains of *C. butryricum* produce type E neurotoxin and some strains of *C. barati* can produce type F neurotoxin. In general, the designation of the strain is that of the toxin it produces. Strains of subtypes. AB, AF, BA, and BF are rare and produce, in addition to the first toxin type, lesser amounts of toxin of the second type. Type C and D strains often produce small amounts of D and C toxin, respectively. They may also produce a toxin, C2, which is distinctly different from the neurotoxin and will not be discussed here.

Structure

The neurotoxins are all very similar proteins with a molecular weight of approximately 150 kDa. They are synthesized as a single-chain protein with relatively low toxicity, which is activated when it is 'nicked' by a number of proteases, including proteases of group I C. *botulinum*, into a double-chain molecule which is held together by a disulfide bond. The two components of the nicked toxin, a light (L) and heavy (H) chain, have molecular weights of approximately 50 and 100 kDa, respectively. Individually, the two components are not toxic, but toxicity can be restored by reestablishing the disulfide bond.

The neurotoxins exist as complexes of four molecular sizes: 7S, 12S, 16S, and 19S, designated S (small), M (medium), L (large), and LL (extra large) toxins, respectively. The molecular weights range

from 150 to 900 kDa. The M form is the most common natural form, found in foods and cultures, along with the L form. It is a complex of the S form with an atoxic component. In the L form, a hemagglutinin is also part of the complex. The LL form is only known for type A, and was the first form of the botulinum toxin to be purified and crystallized. It has not been found in cultures and is probably an artificial aggregate. The M and L forms are referred to as progenitor toxins. They dissociate under mild alkaline conditions into the S, or derivative toxin, which is the single-chain protein.

Mode of Action

Botulinum neurotoxins cause paralysis by blocking acetylcholine release at neuromuscular junctions. Cholinergic systems are affected the most, but adrenergic systems may also be affected by high concentrations of toxin. Paralysis appears to be caused by a four-step process. First, the toxin binds to a type-specific receptor on the presynaptic membrane. This step is mediated by the carboxy-terminus of the H chain. Second, the toxin is internalized into the nerve cell by an initial receptor-mediated endocytosis. Acidity in the endosome is believed to induce the amino-terminus of the H chain to form channels in the endosome membrane which permit the L chain to enter the cytoplasm. Finally, the light chain acts as a zinc endopeptidase and hydrolyzes the SNARE proteins responsible for synaptic vesicle fusion: synaptobrevin, syntaxin, or SNAP-25. Cleavage of the SNARE proteins results in inhibition of acetylcholine release, ultimately resulting in flaccid paralysis.

Inactivation

The most effective means of inactivating botulinum toxins in foods is by heat. Heat inactivation curves are biphasic, with an initial steep decline that levels off with time. Foods, especially those high in protein, colloidal components, or ionic strength, have a protective effect. The toxin is most stable between pH 4 and 5. For the safe thermal inactivation of toxin at concentrations up to 10^5 LD_{50} per gram, time/temperature combinations of 20 min at 79 °C or 5 min at 85 °C have been recommended. Other means of toxin inactivation include treatment with chlorine or ozone.

See also: **Canning**: Principles; **Chilled Storage**: Microbiological Considerations; Packaging Under Vacuum; ***Clostridium***: Occurrence of *Clostridium botulinum*; **Fish**: Spoilage of Seafood; **Heat Treatment**: Chemical and Microbiological Changes; **Lactic Acid Bacteria**; **Marine Foods**: Marine Mammals as Meat Sources; **Meat**: Preservation; **Microbiology**: Detection of Foodborne Pathogens and their Toxins; **Nisin**; **Nitrates and Nitrites**; **Nitrosamines**; **Smoked Foods**: Principles; **Water Activity**: Principles and Measurement

Further Reading

Centers for Disease Control and Prevention (1998) *Botulism in the United States, 1899–1966. Handbook for Epidemiologists, Clinicians, and Laboratory Workers*. Atlanta, GA: Centers for Disease Control and Prevention.

Dodds KL and Austin JW (1997) *Clostridium botulinum*. In: Doyle MP, Beuchat LR, and Montville TJ (eds) *Food Microbiology. Fundamentals and Frontiers*, pp. 288–304. Washington: ASM Press.

Galazka A and Przybylska A (1999) Surveillance of foodborne botulism in Poland: 1960–1998. *Eurosurveillance* 4: 69–72.

Gao QY, Huang YF, Wu JG, Liu HD and Xia HQ (1990) A review of botulism in China. *Biomedical and Environmental Sciences* 3: 326–336.

Hauschild AHW and Dodds KL (1993) Clostridium botulinum: *Ecology and Control in Foods*. New York: Marcel Dekker.

Pellizzari R, Rossetto O, Schiavo G and Montecucco C (1999) Tetanus and botulinum neurotoxins: mechanism of action and therapeutic uses. *Philosophical Transactions of the Royal Society of London Series B – Biological Sciences* 354(1381): 259–268.

Proulx JF, Milot-Roy V and Austin J (1997) Four outbreaks of botulism in Ungava Bay, Nunavik, Quebec. *Canada Communicable Disease Report* 23: 30–32.

Shapiro RL, Hatheway CL and Swerdlow DL (1998) Botulism in the United States: a clinical and epidemiologic review. *Annals of Internal Medicine* 129: 221–228.

Smith LDS and Sugiyama H (1988) *Botulism: The Organism, Toxins, the Disease*, 2nd edn. Toronto: Academic Press.

St Louis ME, Peck SH, Bowering D *et al.* (1988) Botulism from chopped garlic: delayed recognition of a major outbreak. *Annals of Internal Medicine* 108: 363–368.

Therre H (1999) Botulism in the European Union. *Eurosurveillance* 4: 2–7.

Villar RG, Shapiro RL, Busto S *et al.* (1999) Outbreak of type A botulism and development of a botulism surveillance and antitoxin release system in Argentina. *Journal of the American Medical Association* 281: 1334–1338, 1340.

Weber JT, Hibbs RJ Jr, Darwish A *et al.* (1993) A massive outbreak of type E botulism associated with traditional salted fish in Cairo. *Journal of Infectious Diseases* 167: 451–454.

COBALAMINS

Contents
Properties and Determination
Physiology

Properties and Determination

M Jägerstad and K Arkbåge, Swedish University of Agricultural Sciences, Uppsala, Sweden

Discovery

The history of cobalamin (vitamin B_{12}) started almost two centuries ago, when a fatal anemia, resulting from some disorder of the digestive and assimilative organs, was described. The symptoms of this anemia, named pernicious anemia, was further outlined during this century, but a successful treatment of this fatal disease was reported first in 1926, when raw liver was introduced as therapy. In 1929, it was shown that the intestinal absorption of the antipernicious anemia principle in liver (extrinsic factor) required prior binding to an 'intrinsic factor' secreted by the stomach. The search for the active principle in liver culminated with the isolation of vitamin B_{12} in 1948 by a team of Merck scientists in the USA, which was confirmed by a British team 3 weeks later. Injections of the isolated vitamin B_{12} into muscular tissues induced a dramatic beneficial response in patients with pernicious anemia. Hodgkin and coworkers elucidated the chemical structure of cyanocobalamin (vitamin B_{12}) by X-ray crystallography in 1964, for which they were rewarded the Nobel Prize in chemistry (the second Nobel Prize designated for this vitamin).

Structure, Nomenclature, and Physical Properties

Cobalamins have a complex chemical structure based on a corrin ring system. Four reduced pyrrole rings are linked together to form a macrocyclic ring designated corrin, which is chelated by the four pyrrole nitrogens. Corrin resembles the heme of hemoglobin, but has one less α-methene bridge and has cobalt (Co) instead of iron at the center. All the compounds containing this ring are designated corrinoids. The fifth co-ordinate covalent bound to cobalt is a nitrogen of a dimethylbenzimidazole moiety. Thus, as with nucleic acids, cobalamins contain a nucleotide, in this case 5,6-dimethylbenzimidazole, rather than various purine or pyrimidine bases. And the sugar, ribose, has an α-glycosidic linkage, unlike the β-linkage in the nucleic acids. The D-1-amino-2-propanol moiety of the molecule is esterified to the nucleotide and joined in amide linkage to the porphyrin-like nucleus. The sixth position of cobalt may be occupied by different anionic ligands, (*R*) according to Figure 1 and Table 1. This ligand can be one of the following types, e.g., –CN(cyanocobalamin), –OH (hydroxocobalamin), $-H_2O$ (aquacobalamin), $-NO_2$ (nitritocobalamin), $-SO_3$ (sulfitocobalamin), $-CH_3$ (methylcobalamin), or –5′-deoxy-5′-adenosyl (adenosylcobalamin). Approximately 20 naturally occurring analogs of cobalamins have been identified. Some of these have no biological activity in mammals, whereas others exhibit at least partial vitamin activity but are often poorly absorbed. Human serum contains not only the active cobalamins (mainly the hydroxo-, methyl-, and adenosyl-cobalamins) but also a variable number of analogs that are biologically inactive noncobalamin corronoids. These vitamin B_{12} analogs have their base of nucleotide, 5,6-dimethylbenzimidazole replaced by other nucleotides, e.g., adenine, guanine, 5-methyl-adenine, or 5-hydroxybenzimidazole.

According to the nomenclature of the corrinoids, a cobalamin is a cobamide in which 5,6-dimethylbenzimidazole is the aglycon attached by a glycosyl link. Consequently, the systemic name of cobalamin given by IUNS (International Union of Nutritional Sciences) and IUPAC (International Union of Pure and Applied Chemistry) is: Coα-[α-(5,6-dimethylbenzimidazolyl)]-Coβ-'*R*' cobamide, where '*R*' denotes 'cyano,' 'aqua,' 'hydroxo,' 'nitrito,' 'adenosyl,' 'methyl,' etc) IUNS and IUPAC state that the term 'vitamin B_{12}' should be used as the generic description for all corrinoids exhibiting qualitatively the biological activity of cyanocobalamin, whereas the term 'corrinoid(s)' should be used as the generic descriptor for all compounds containing the corrin nucleus and thus chemically related to cyanocobalamin. The term 'vitamin B_{12}' has two meanings.

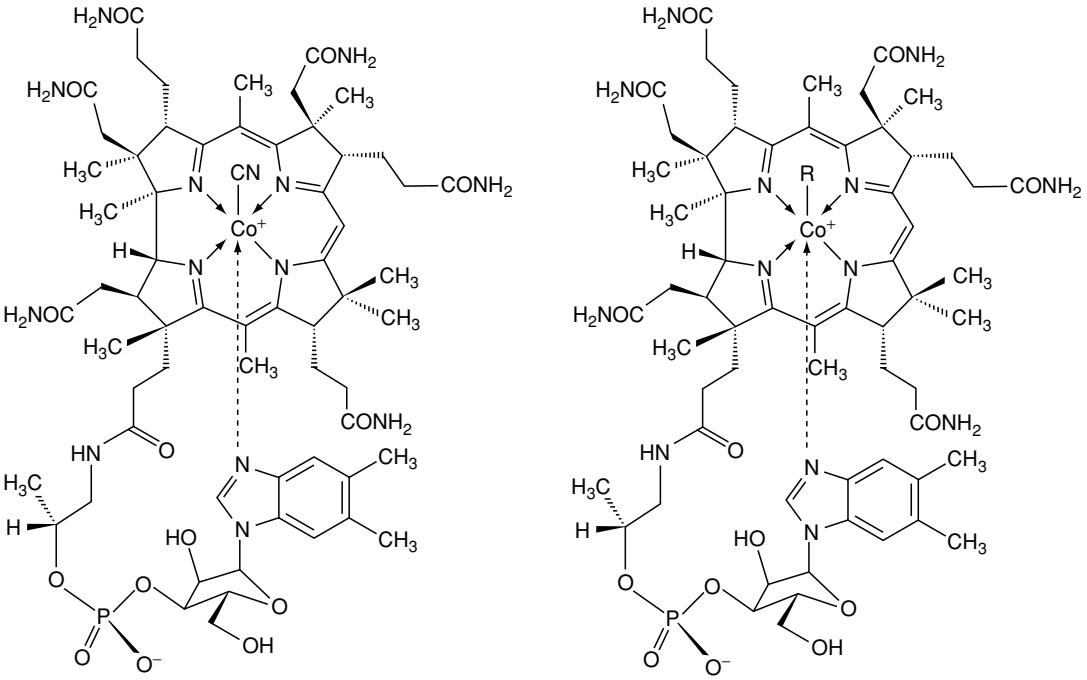

Figure 1 *Left*: Chemical structure of cyanocobalamin; *right*: Coenzyme B_{12}/adenosylcobalamin ($R = 5'$-deoxy-$5'$-adenosyl); methylcobalamin ($R =$ methyl); aquacobalamin ($R = H_2O^+$); hydroxocobalamin ($R =$ OH), with kind permission of Kräutler.

To the chemist, it means only cyanocobalamin, whereas in the nutrition and pharmacology literature, vitamin B_{12} is a generic term for all the cobamides exerting biological activity in humans. So far, all the cobamides found to play a role in human metabolism have been cobalamins. Table 1 summarizes the generic or trivial names for different cobalamins, together with their abbreviations, CAS numbers, and physical properties.

Chemical Properties, and Stability

Cyanocobalamin

Cyanocobalamin, a synthetic form of vitamin B_{12}, is commercially available for medical use, food fortification, and nutrient supplements. The original identification of vitamin B_{12} as cyanocobalamin involved its formation as an artefact of the isolation procedure. Little or no naturally occurring cyanocobalamin exists in foods. Thus, cyanocobalamin is normally not present in humans except in very small amounts in smokers. It is a tasteless, odorless, red crystalline substance with good water solubility. The color of the compound may pose a limitation in the possible addition of cyanocobalamin to certain foods, especially light-colored products (e.g., white bread).

The cyanide group of cyanocobalamin can be replaced by other ions to form hydroxycobalamin, chlorocobalamin, nitritocobalamin, thiocyanatocobalamin, and others; for example, bisulfite ions cause conversion of aquacobalamin to sulfitcobalamin, and similar reactions can occur to form cobalamins substituted with ammonia and nitrite. These reactions have little influence on the net vitamin B_{12} activity of foods, since all the cobalamines are readily reconverted to cyanocobalamin after treatment with cyanide, except methylcobalamin.

Cyanocobalamin appears to be the most stable of the various vitamin B_{12} analogs, being stable in air and in dry form, and is even relatively stable at 100 °C for a few hours. Aqueous solutions at pH 4–7 can be autoclaved at 120 °C. Cyanocobalamin is slowly decomposed by ultraviolet or strong visible light. The cyano group is split off, yielding hydroxocobalamin, which is also a biological active form of vitamin B_{12} (and is available for medical use and used in fortification by the food industry). Prolonged exposure to light, however, causes irreversible decomposition and inactivation.

Mild acid hydrolysis of cyanocobalamin induces the removal of the nucleotide, and additional fragmentation occurs as the severity of the acidic conditions increases. Exposure to alkaline conditions causes hydrolysis of amides, yielding biologically inactive carboxylic acid derivatives of vitamin B_{12}. In solution, thiamin and nicotinamide, or nicotinic acid, destroy cyanocobalamin slowly, whereas the addition of small amounts of iron or thiocyanate appear to protect it. However, the literature is not definitive.

Table 1 Nomenclature and physical properties of cobalamins (vitamin B_{12})

Generic name	*Ligand (R)*	*Abbreviation*	*Molar mass*	*Formula*	*Solubility*	*Crystal form*	λ_{max} *(nm)*	*Absorption* (E_{1cm})	$\varepsilon \times 10^{-3}$	*Solvent*
Cyanocobalamin (vitamin B_{12})	$-CN$	CNCbl	1355.38	$C_{63}H_{88}CoN_{14}O_{14}P$	1 g in 80 ml of water	Dark red hygroscopic	278	115	[15.6]	Water
							361	204	[27.6]	Water
CAS No. 68-19-9[a]							551	64	[8.7]	Water
Aquacobalamin (vitamin B_{12a})	$-H_2O$	H_2OCbl^+	1347.0	$C_{62}H_{90}CoN_{13}O_{15}P$			274	[153]	20.6	Water
							351	[197]	26.5	Water
CAS No. 13422-52-1							499	[60]	8.1	Water
Hydroxocobalamin (vitamin B_{12b})	$-OH$	HOCbl	1346.37	$C_{62}H_{89}CoN_{13}O_{15}POH$	Moderately soluble in water; insoluble in acetone, ether, petroleum ether, benzene	Dark red orthorombic	279	[141]	19.0	Water
							325	[85]	11.4	Water
CAS No. 13422-51-0							359	[153]	20.6	Water
							516	[66]	8.9	Water
							537	[71]	9.5	Water
Nitritocobalamin (vitamin B_{12c})	$-NO_2$		1374.6	$C_{62}H_{88}CoN_{14}O_{16}P$		Red crystalline solids	352	153	[21.0]	Water
							528	60	[8.4]	Water
CAS No. 20623-13-6										
Sulfitocobalamin	$-SO_3$		1409.5	$C_{62}H_{89}CoN_{13}O_{17}PS$			275	328	[46.2]	Water, pH 7
CAS No. 15671-27-9							365	130	[18.3]	Water, pH 7
							516	61	[8.6]	Water, pH 7
Adenosylcobalamin (cobamide or coenzyme B_{12})	5′-deoxy-5′-adenosyl	AdoCbl	1579.6	$C_{72}H_{100}CoN_{18}O_{17}P$	Soluble in ethanol, phenol; insoluble in acetone, ether, dioxane	Yellow–orange six-faced crystal	288	[115]	18.1	Water, pH 7
							340	[78]	12.3	Water, pH 7
							375	[60]	10.9	Water, pH 7
CAS No. 13870-90-1							522	[51]	8.0	Water, pH 7
Methylcobalamin (methyl B_{12})	$-CH_3$	MeCbl	1344.4	$C_{63}H_{91}CoN_{13}O_{14}P$		Bright red	266	[148]	19.9	Water
							342	[107]	14.4	Water
CAS No. 13422-55-4							522	[70]	9.4	Water

[a]CAS No., Chemical Abstract Service Number.

Values in brackets are calculated from the corresponding ε value or $E_{1\,cm}$ value.

Modified from Eitenmiller.

Coenzyme B_{12}

The coenzyme forms of vitamin B_{12} are methylcobalamin and adenosylcobalamin. Methylcobalamin functions coenzymatically in the transfer of a methyl group (from 5-methyltetrahydrofolate to homocysteine) in methionine biosynthesis, whereas 5′-deoxyadenosylcobalamin serves as the coenzyme in an enzymatic rearrangement reaction catalyzed by methylmalonyl-CoA mutase. Both forms of vitamin B_{12} coenzymes are converted to the biological active aquacobalamin upon light exposure. In fact, vitamin B_{12} coenzymes exhibit extreme photosensitivity, which necessitates the use of subdued lighting and low actinic glassware during vitamin B_{12} analyses. The photochemical degradation of vitamin B_{12} coenzymes interferes with experimental studies of vitamin B_{12} metabolism and function, but this conversion has no influence on the total vitamin B_{12} activity of foods, because aquacobalamin retains vitamin B_{12} activity. Surprisingly, methylcobalamin and adenosylcobalamin are relatively stable in neutral aqueous solution in the dark and can be heated for 20 min at 100 °C.

Adenosylcobalamin is unstable under acidic and alkaline conditions, whereas methylcobalamin is stable in the presence of dilute acid or alkali. In general, the natural forms of cobalamins (hydroxo-, aqua-, methyl-, or adenosylcobalamin) are inactivated by severe alkaline conditions, heavy metals, strong oxidizing agents, or strong reducing agents (like ascorbate). A possible interaction with ascorbic acid has been reported to be detrimental to the vitamin, especially under heating conditions. The significance of these observations for the processing of food remains unclear (see further discussion below). The mechanism of vitamin B_{12} degradation has not been fully determined, in part because of the complexity of the molecule and the very low concentration in foods.

Cobalamins in Food

Only microorganisms are able to synthesize cobalamin, and therefore, higher plants are devoid of vitamin B_{12} unless processed microbiologically or contaminated by fecal material, e.g., from fertilizer. Certain legumes have been reported to absorb small amounts of vitamin B_{12} produced by bacteria associated with root nodules, but little enters the seeds. Food from animal origin, especially organs like liver and kidney, contain > 30 μg vitamin B_{12} per 100 g. Other rich sources are meat muscles, fish, dairy products, and eggs. Bovine milk contains about 0.4 μg of vitamin B_{12} per 100 g (range 0.24–0.74 μg per 100 g). Human milk contains only about 15% as much vitamin B_{12} as bovine milk. The predominant source of vitamin B_{12} in bovine milk is biosynthesized by microorganisms in the rumen. Hydroxocobalamin is the predominant form of vitamin B_{12} in bovine milk. There may also be some of the light-sensitive forms of vitamin B_{12}, adenosylcobalamin and methylcobalamin, present. In milk, most of the vitamin B_{12} is protein-bound. The significance of this is still unknown, although a protective effect against microbial breakdown has been suggested. The vitamin B_{12} in most animal tissues consists mainly of the coenzyme forms, methylcobalamin and adenosylcobalamin, in addition to aquacobalamin. Some foods or microorganisms contain other corrinoids, with varying biological activity, some of which may have antivitamin activity. Foods are classified according to their vitamin B_{12} content, as shown in Table 2.

Process and Storage Losses

Under most conditions of food processing and preservation, there is little nutritionally significant loss of vitamin B_{12}. Cyanocobalamin added to breakfast cereal products has been reported to undergo an average loss of 17% during processing, with an additional 17% loss during storage for 12 months at ambient temperature. When milk is pasteurized for 2–3 s, it loses 7% of its available vitamin B_{12}, and boiling for 2–5 min destroys 30%. Sterilization in a bottle for 13 min at 119–120 °C causes a loss of 77%, and rapid sterilization (3–4 s) with superheated steam at 143 °C, i.e., ultrahigh temperature (UHT) destroys about 10% of the vitamin.

There are also processes like fermentation used for manufacture of certain dairy products, e.g., yogurt

Table 2 Grouping of food according to vitamin B_{12} content

Food	*Vitamin B_{12} (μg per 100 g edible weight)*
Very rich sources: organ meats (liver, kidney, heart), bivalves (clams and oysters)	> 10
Moderately rich sources: some fish, crabs, egg yolk	3–10
Moderate sources: muscle meats (beef), some fish, cheese, egg, milk powder	1–3
Good sources: milk, yogurt, muscle meats (pig)	< 1
Poor sources: vegetables, cereals, fruits and berries	n.d.

n.d., not detectable.

and cheeses, which might result in either increased or decreased vitamin B_{12} concentrations depending on the vitamin B_{12}-producing or -consuming microorganisms.

Studies indicate that the storage temperature and time have a significant influence on the rate of breakdown. Thus, if UHT milk is stored at a low temperature (+7 °C), as for pasteurized milk, the losses of vitamin B_{12} are low for up to at least half a year. However, if storage is at ambient temperature, which is a normal condition for an aseptic product such as UHT milk, breakdown of the vitamin may be considerable even within a few weeks. Fermented dairy products might lose or gain vitamin B_{12} during storage for the reason given above.

Like all water-soluble vitamins, vitamin B_{12} is subject to losses through leaching into the cooking water. This indicates the potential for substantial losses of vitamin B_{12} during prolonged heating of foods at or near neutral pH. Typical oven heating of commercially prepared convenience dinners has been shown to yield 79–100% retention of vitamin B_{12}.

Ascorbic acid has long been known to accelerate the degradation of vitamin B_{12}, especially during high-temperature heating, although this may be of little practical significance, because foods containing vitamin B_{12} usually do not contain significant amounts of ascorbic acid. In curing solutions for ham, neither ascorbate nor erythorbate has been found to have any influence on vitamin B_{12} retention, probably since this process does not include high-temperature heating.

Determination of Cobalamins

General

For detection purposes, cobalamins are all UV absorbers. Unfortunately, their absorption spectra vary between different congeners, and detection is complicated by the low concentration at which they occur normally in foods. Available methods for assay of cobalamins developed over the years include both general methods, e.g., polarographic, spectrophotometric, various chromatographic procedures (paper, thin-layer, and open-column chromatography), and more specific and/or sensitive methods, e.g., microbiological, protein-binding, and high-performance liquid chromatography (HPLC) procedures.

Almost all available data on vitamin B_{12} in food have been obtained by microbiological assay (MA) using *Lactobacillus delbrueckii* as the test organism. Still, the microbiological assay is the reference method for assaying cobalamins in foods. For clinical tissues and serum samples, rapid radioprotein-binding assays with ^{57}Co-labeled cyanocobalamin are routine. However, radioassay kits for clinical assay are not reliable for assay of food unless carefully validated for matrix effects. A more recent development of this protein-binding assay is the use of a biosensor technique to quantify cobalamins in food. Although the various forms of vitamin B_{12} can be separated chromatographically, HPLC methods are not readily suitable for food analysis, even if preconcentration steps are used, because of the very low concentrations typically found. Below, only methods currently in use for the assay of naturally occurring cobalamins in foods and/or fortified foods are presented and discussed.

Extraction

All methods for cobalamin analyses of food samples need a preparative step prior to analysis. Since animal products are the sole food sources of vitamin B_{12}, the vitamin must be extracted from protein-rich matrices under conditions that will not lead to its destruction. As a result, extraction procedures often involve digestion with protease followed by extraction in hot (80 °C) ethanol. For quantitation as total vitamin B_{12}, heating in a cyanide- or sulfate-containing buffer has been used to extract the endogenous vitamin B_{12} forms, thereby converting them to the more stable cyano- or sulfitocobalamin forms. The AOAC official method (952.20) recommends mixing approximately 1 g or 1 ml of the sample in $\geq$25 ml of a freshly prepared solution containing sodium phosphate, citric acid, and metabisulfite. The extraction is completed by homogenizing the sample in the extraction solution, autoclaving the mixture at 121–123 °C for 10 min, adjusting the cooled mixture to pH 6.0, filtering and/or centrifugation, and diluting with water to a vitamin B_{12} concentration of approximately 0.2 ng ml^{-1}. However, most extraction procedures described in the literature add either sodium cyanide (NaCN) or potassium cyanide (KCN) to 0.1 M acetate buffer, pH 4.5–4.7, during homogenization and extraction. The heating time varies between 3 and 30 min, and a heating temperature of either 100 °C (boiling water) or 121 °C (autoclavation) is used. The heat-extraction procedure including centrifugation and/or filtration are enough of sample preparation for microbiological assays and protein-binding assays (immunoassays). In the use of HPLC, further clean-up work might be necessary, e.g., ultracentrifugation, ultrafiltering, affinity chromatography, or adsorption on to charcoal or chromatography on anion exchange resin and/or solid-phase extraction on C_{18} cartridges.

The extraction conditions have been shown to affect significantly the efficiency of the overall extraction

process. It has therefore been suggested that the extraction conditions should be optimized for each type of food sample in order to achieve reproducible and good recoveries of the complete analysis.

Microbiological Assay

Over the years, the growth response of various microorganisms has been used to analyze cobalamins in foods and other biological material. The microorganisms that have been used are *Lactobacillus delbrueckii* (previously, *leichmanni*), *Escherichia coli, Euglena gracilis*, and *Ochromonas malhamensis*. Below, only the AOAC (Association of Official Analytical Chemists) Official method (reference method) based on *Lactobacillus delbrueckii* is presented. The other microorganisms have different disadvantages compared with *Lactobacillus delbrueckii*, i.e., less sensitive and/or specific also towards less active biological analogs. In addition, *Euglena gracilis* and *Ochromonas malhamensis* grow more slowly and are thus more painstaking.

The AOAC Official Method 952.20: *Cobalamin (Vitamin B_{12}-activity) in Vitamin preparations* use a microbiological assay with *Lactobacillus delbrueckii* (ATCC 7830) (45.2.02). This method was originally collaborated for use on vitamin preparations, but the AOAC Task Force on Methods for Nutrition Labeling recommended the procedure for use on all food matrices. The *Lactobacillus delbrueckii* growth response is sufficiently sensitive to quantify cyanocobalamin at concentrations approaching 1.0 pg per milliliter of assay growth medium (corresponding to less than 0.5 μg per 100 g). It has a variable response to various cobalamins. A similar growth response has been reported for cyanocobalamin, hydroxocobalamin, sulfitocobalamin, nitritocobalamin, and dicyanocobalamin. However, adenosylcobalamin produces a greater response, and methylcobalamin a lower growth response. This will generally not cause a problem, since excess of cyanide or sulfite is used during the extraction procedure prior to analysis, thereby converting all different native forms of cobalamins except methylcobalamin into one stable form.

An important aspect in all methods of analysis is their specificity. *Lactobacillus delbrueckii* can utilize vitamin B_{12} analogs, e.g., deoxyribonucleotides and deoxyribonucleosides, in addition to biologically active cobalamins. Older literature suggests that dilution of deoxyriboside levels (e.g., thymidine) to less than 1 μg per milliliter of the assay medium will eliminate the effect.

Radioprotein-Binding Assays

Radioprotein-binding assays have been used routinely for blood and tissue analysis since early work by Rothenberg and several others during the 1960s. The principle of this assay is based on the competition of added isotope-labeled cyanocobalamin (^{57}Co) versus cyanocobalamin from the unknown sample for a limited number of binding sites on cobalamin-binding proteins (often intrinsic factor from hog). The amount of labeled cobalamin bound to the binding protein is quantified in a gamma counter, and the concentration of cobalamin in the unknown sample is inversely proportional to the concentration of labeled cobalamin, and calculated from a calibration curve.

Early types of radioprotein-binding assays for determination of vitamin B_{12} in clinical specimens and foods were often inaccurate because the binding protein could bind to active forms of vitamin B_{12} as well as biologically inactive analogs. The specificity of such assays has been greatly improved through the use of a vitamin B_{12}-binding protein (generally porcine intrinsic factor, IF) that is specific for the biologically active forms of the vitamin. Radioligand binding methods based on monospecific antisera that eliminate any cross-reactions with other cobalamins are now available and should be used routinely in all clinical laboratories.

The relative binding affinities of porcine IF for various cobalamins have also been studied. An equivalent affinity has been found for cyano-, dicyano, nitrito-, and methylcobalamins, whereas significantly different binding affinities have been found for hydroxo-, sulfito, and adenosylcobalamins. Therefore, with cyanocobalamin as the calibration standard, it is necessary to add cyanide to the samples prior to quantification during the extraction procedure to convert all native forms to cyanocobalamins.

Studies by several research groups indicate that radioprotein-binding assays with IF can be used for food analysis. When optimized extraction procedures have been used, radioprotein binding assays and microbiological procedures show agreement for most foods; however, differences have been noted often enough to conclude that the two methods are not universally interchangeable for the assay of all foods. Thus, recent work has shown that the radioprotein binding assay could be reliable for food analysis, providing careful method validation including control of selectivity. One should bear in mind that not only the food matrix itself but also substances used during the extraction procedure, e.g., buffer strength, pH, cyanide concentration, and added enzymes needed to release vitamin B_{12} bound to the food matrix, might cause nonspecific binding and therefore must be accounted for in order to achieve reliable results.

An alternative to the protein-binding assay is an enzyme-linked protein-binding assay that appears to be useful for fortified foods in which cyanocobalamin is the predominant cobalamin. The procedure uses an *R*-protein–enzyme conjugate and microtiter plate techniques for quantitation of cyanocobalamin. Since *R*-proteins bind cobalamins and their analogs, the assay has not been applied to nonfortified foods. Another drawback is that the *R*-protein enzyme conjugate is not commercially available. Applications to

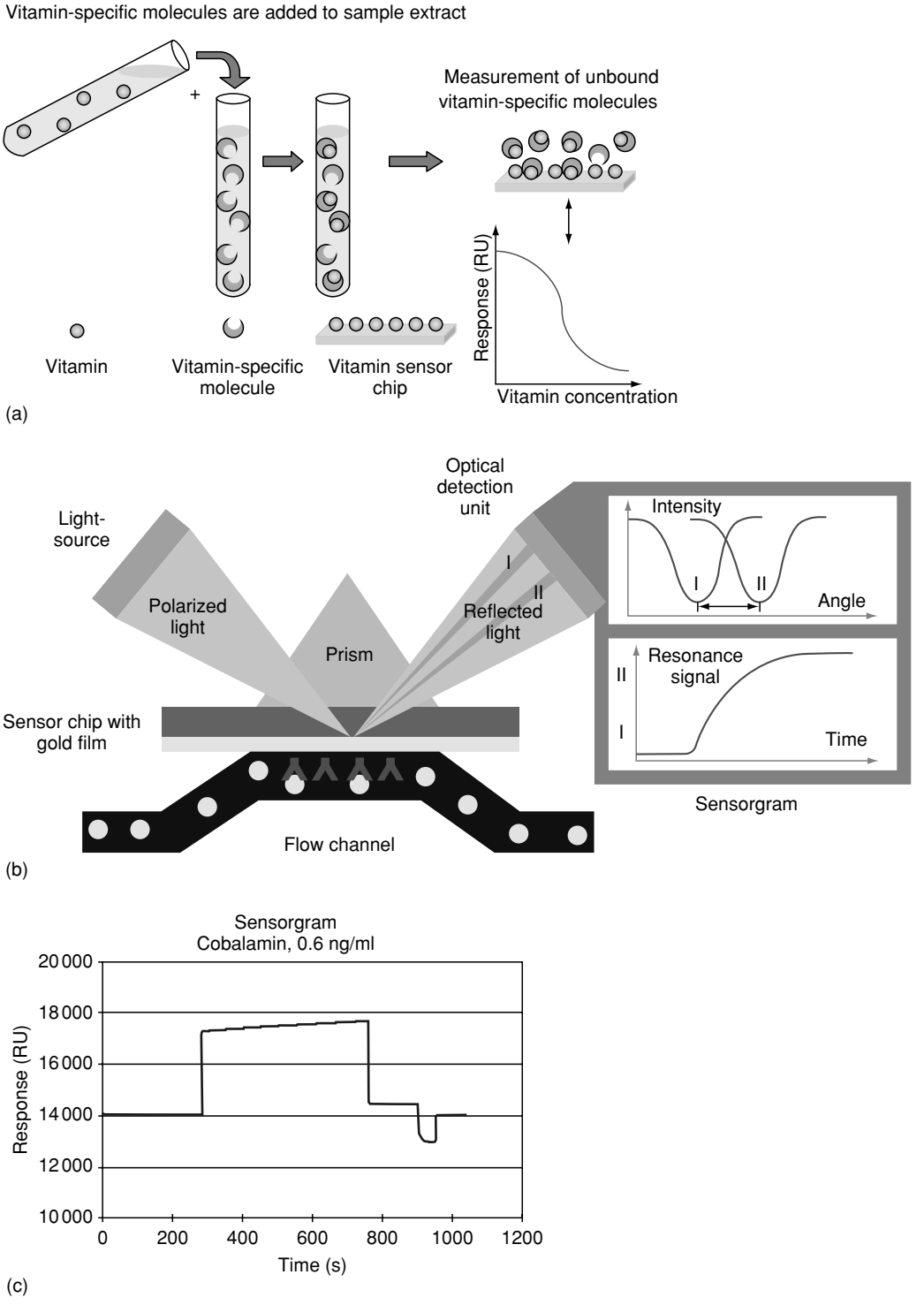

Figure 2 Assay principle for determination of vitamin B_{12} in food using optical biosensor technology (BIACORE).

fortified breakfast cereals have given a limit of detection of 9 pg per well and a quantification limit of 0.09 μg per 100 g.

Optical Biosensor-based Immunoassay

This assay is the latest methodological development for vitamin B_{12} in foods and is based on biospecific recognition of the analyte, cyanocobalamin, by a vitamin B_{12} binding protein. The event of recognition is transduced to a signal by means of an optical sensing mechanism. Real-time biomolecular interaction analysis recently developed by BIACORE is a label-free technology for monitoring biomolecular interaction as they occur. A special kit, Qflex™ Kit Vitamin B_{12}, is commercially available and uses an inhibition assay/indirect assay for the analyses. In an inhibition assay, the analyte or an analog is immobilized onto the surface of a sensor chip. A high-molecular-weight detecting molecule such as a binding protein in a defined concentration is added to the sample. The binding proteins bind to the analyte, but at equilibrium, some binding proteins will remain in solution free to bind to the sensor surface. A standard curve is prepared from solutions with known concentrations. The concentration of analyte in an unknown sample is then derived from a standard curve.

The detection principle relies on surface plasmon resonance (SPR), an electronic charge-density wave phenomenon that arises at the surface of a metallic film, usually based on gold, when light is reflected at the film under specific conditions. The resonance is a result of energy and momentum being transformed from incident photons into surface plasmons and is sensitive to the refractive index of the medium on the opposite side of the film from the reflected light. A scheme of the principle of the optical biosensor technology is shown in Figure 2.

Method validation data using milk-based infant formulas fortified with cyanocobalamin have given the following results; precision (RSD_r): 4.4; recovery: 93–106%; detection limit (LOD $-3 \times$ SD): 0.07 ng ml^{-1} and quantification limit (LOQ $-10 \times$ SD): 0.17 ng ml^{-1}. Samples can be quantified in the range of 0.1–4 ng ml^{-1}, corresponding to an LOQ of $\sim$ 0.5 μg per 100 g, depending on sample preparation.

Samples analyzed by MA and Biacore Q show a good agreement, but the BIACORE method is less sensitive so far. This restricts its use at the moment to fortified foods, e.g., infant milk formulas and gruels, and pharmaceutical vitamin mixtures. The biosensor-based BIA:SPR technique provides a convenient, label-free, and automated quantitative analysis amenable to routine compliance monitoring in the food laboratory. Advantages of this technique include the speed and simplicity of operation for routine compliance analyses and an enhanced precision.

HPLC Methods

Liquid chromatography can be conveniently applied to the assay of vitamin B_{12} in high-concentration supplements and pharmaceutical preparations (usually cyano- or hydroxocobalamin forms). A few such HPLC methods have been published since 1980. It can also effectively resolve natural cobalamins. However, owing to the low levels of analytes, radioligand-binding assays are necessary to quantify eluted cobalamins. For milk, which is a good source of vitamin B_{12} (containing around 4 μg of vitamin B_{12} per liter), an HPLC method based on UV detection (550 nm) and with a detection limit of 0.2 μg of vitamin B_{12} per liter of milk is available. Recently, a novel HPLC method for the determination of vitamin B_{12} (cyanocobalamin) with fluorescence detection has been reported. The method has been used successfully to determine vitamin B_{12} in pharmaceutical preparations, e.g., vitamin B_{12}-containing multivitamin tablets and fermentation medium. The recovery varies from 94 to 102%, and the relative standard deviation is in the range of 1.8–4.1%.

Certified Reference Material

Two Bureau Commission Reference materials, milk powder (CRM 421) and lyophilized pig's liver (CRM 487) are available by the European Commission in Brussel with a certified vitamin B_{12} content analyzed by microbiological assay (*Lactobacillus leichmanni*, ATCC 7830) and radioprotein-binding assay. The two different methods show similar vitamin B_{12} concentrations in these reference materials.

See also: **Biosensors**; **Chromatography**: High-performance Liquid Chromatography; **Cobalamins**: Physiology; **Coenzymes**

Further Reading

AOAC International (1999) *Official Methods of Analysis*, 17th edn. Arlington, VA: AOAC.

Eitenmiller RR and Landen WO Jr. (1999) *Vitamin Analysis for the Health and Food Sciences*, pp. 467–478. New York: CRC Press.

Ellenbogen L and Cooper BA (1991) Vitamin B_{12}. In: Machlin LJ (ed.) *Handbook of Vitamins*, 2nd edn, pp. 491–505. New York: Marcel Dekker.

Finglas PM, Scott KJ, Witthöft CM, van den Berg H and de Froidmont-Görtz I (1999) *The Certification of the Mass Fractions of Vitamins in Four Reference Materials:*

Wholemeal Flour (CRM 121), Milk Powder (CRM 421), Lyophilized Mixed Vegetables (CRM 485) and Lyophilized Pig's Liver (CRM 487). European Commision. BCR Information/Reference materials: EUR 18320 EN.

Gregory JF (1996) Vitamins. In: Fennema OR (ed.) *Food Chemistry*, 3rd edn. pp. 603–606. New York: Marcel Dekker.

Herbert V (1996) Vitamin B_{12}. In: Ziegler EE and Filer LJ (eds) *Present Knowledge in Nutrition*, 7th edn, pp. 191–205. Washington, DC: ILSI Press.

Kräutler B, Arigoni D and Golding BT (1998) *Vitamin B_{12} and B_{12}-Proteins*. Weinheim, Germany: Wiley-VCH.

Lindemans J and Abels J (1992) Cobalamins. In: De Leenheer AP, Lambert WE and Nelis HJ (eds) *Modern Chromatographic Analysis of Vitamins*, 2nd edn. New York: Marcel Dekker.

Machlin LJ and Hüni JES (1994) Vitamin B-12. In: *Vitamins Basics*, pp. 40, 56. Basel, Switzerland: La Roche.

Muhammad K, Briggs D and Jones G (1993a) The appropriateness of using cyanocobalamin as calibration standards in competitive binding assay of vitamin B_{12}. *Food Chemistry* 48: 423–430.

Muhammad K, Briggs D and Jones G (1993b) Comparison of a competitive binding assay with *Lactobacillus leichmanni* ATCC 7830 assay for the determination of vitamin B_{12} in foods. *Food Chemistry* 48: 431–439.

Physiology

C J Bates, MRC Human Nutrition Research, Cambridge, UK

Introduction

The importance of vitamin B_{12} in human nutrition has been dominated by the potentially fatal disease, pernicious anemia, which results from a failure of secretion of the specific cobalamin-binding protein, intrinsic factor, by the parietal cells of the stomach. This protein is essential for the subsequent recognition, binding, and transport of vitamin B_{12} at the ileal wall. Transcobalamin II then carries it to the portal circulation. The treatment of pernicious anemia with massive amounts of dietary liver, followed by the isolation, characterization, and X-ray crystallographic structure determination of vitamin B_{12}, were major achievements of the first half of the 20th century. During the past few years, new evidence has been accumulating to suggest that more subtle, insidious and easily overlooked forms of vitamin B_{12} deficiency may occur in modern-day western society, and that the challenges of their recognition and treatment are by no means entirely solved. The fact that prolonged vitamin B_{12} depletion can lead to irreversible neural damage, coupled with the prediction that defects of absorption can arise slowly and remain entirely unnoticed as people grow older may be of special relevance at a time when the quality of life in old age is becoming a major priority issue. (*See* **Anemia (Anaemia)**: Megaloblastic Anemias.)

Dietary intake by omnivores is much greater than by vegetarians or vegans (who depend to a large extent on contamination sources, unless they deliberately take B_{12} supplements). Some B_{12}-like corrinoids in certain seaweeds have limited B_{12} activity. *In vitro*, vitamin B_{12} may be destroyed by contact with vitamin C (ascorbic acid), but *in vivo* this probably does not occur, which suggests that the catalysis of destruction may depend on free transition metal ions which are not usually present in living tissues. (*See* **Vegetarian Diets**.)

Although B_{12} intakes are much lower in Third World countries than in western countries, severe deficiency is surprisingly rare in the Third World. However, a moderate degree of deficiency has recently been reported from parts of South America, currently attributed to a combination of low intakes and impaired absorption due to gastrointestinal parasites. Fully breast-fed infants of mothers with very low intakes, or of mothers with vitamin-B_{12} malabsorption syndromes (usually pernicious anemia), may develop overt B_{12} deficiency. Older omnivores usually accumulate considerable amounts in their tissues, and in these subjects the effect of a switch to diets with low contents only becomes apparent several decades after the B_{12} supply to the internal milieu has virtually ceased.

There are still some problems with the assay of vitamin B_{12} levels in food, and 5–30% of the reported B_{12} content may be biologically inactive corrinoids. Small amounts of B_{12} arising from bacteria in the ileum can be absorbed, but B_{12} produced by bacteria in the colon is not. An efficient enterohepatic circulation of B_{12} occurs in normal people, and if this cycle is interrupted, as it is in cases of impaired intestinal absorption, this results in an additional source of losses for subjects with impaired B_{12} absorption. (*See* **Microflora of the Intestine**: Role and Effects.)

Large-scale industrial production of vitamin B_{12} makes use of the capacity of organisms such as *Propionibacterium* spp. to synthesize as much as 40 mg B_{12} per liter of growth medium. World consumption was around 5000 kg in the early 1980s.

Absorption and Tests of Status

The key stages in vitamin B_{12} utilization from food sources are depicted in **Table 1**.

Table 1 Stages of normal vitamin B_{12} absorption in humans

- Release from dietary protein by gastric acid and pepsin (this can often become impaired in older people)
- Binding to 'R-binders': glycoproteins secreted by saliva
- Transfer to intrinsic factor (IF) when R-binders are degraded by pancreatic enzymes in the intestine
- Attachment of the B_{12}–IF complex to ileal receptors, followed by endocytotic transfer across the intestinal wall. This has very precise specificity requirements and is the rate-limiting step in absorption. It is most efficient in the range of 1–2 μg B_{12}, whereas passive absorption occurs with only *c.* 1% the efficiency, but over a much wider range of intakes, so that useful amounts are absorbed when intakes are high
- Circulation in the plasma in several forms: transcobalamins (TC) I, II and III, and R-proteins (haptocorrins), of which TC II is the main tissue uptake precursor, with a half-life of only *c.* 6 min
- Storage, mainly in the liver; recycling by enterohepatic circulation; turnover at a rate of *c.* 0.1% of the body pool per day

Table 2 The Schilling test for vitamin B_{12} absorption efficiency

- Measures percentage recovery of an oral dose of labeled vitamin B_{12}, using an intravenous dose to flush the newly absorbed vitamin largely into the urine
- Cobalt-labeled vitamin B_{12} is used, and there are three usable, γ-emitting isotopes: ^{60}Co (half-life 5 years); ^{57}Co (half-life 270 days), and ^{58}Co (half-life 71 days)
- After a parenteral injection of unlabeled B_{12}, normal absorbers excrete *c.* 15% of the labeled dose within 24 h, whereas those with impaired absorption excrete less than 5%
- By using two different isotopes of cobalt simultaneously, it is possible to distinguish the different types of malabsorption, e.g., by comparing free B_{12} against B_{12} bound to intrinsic factor, R-binders, or binding proteins in food

In addition to classical pernicious anemia (lack of secretion of intrinsic factor due to atrophy of gastric mucosa or the production of autoantibodies to intrinsic factor or its intestinal binders), there are several other medical conditions that can result in impaired absorption, as well as a variety of congenital abnormalities of transport of the vitamin, which require both specialized detection procedures and specialized therapeutic measures. These include:

1. surgical procedures such as total gastrectomy, removal of ileum, or presence of blind loops of the small gut with an abnormal bacterial flora
2. the effects of certain parasites, drugs, or substances like nitrous oxide
3. the Imerslund–Gräsbeck syndrome which appears to result from lack of ileal receptors
4. pancreatic insufficiency leading to failure of digestion of R proteins
5. lack of circulating transcobalamin II

The classical procedure for the recognition of impaired absorption, specially designed to detect pernicious anemia, is the urinary excretion (Schilling) test (Table 2). Around 57% of people with pernicious anemia also have circulating antibodies to intrinsic factor in their blood, and their detection can provide further corroborative evidence of the presence and type of pernicious anemia. In conjunction with a macrocytic anemia and a standard assay of circulating plasma vitamin B_{12} or of holo-transcobalamin II, the measurement of these antibodies can be a useful screening procedure to be performed before the more complex Schilling test is undertaken.

Serum vitamin B_{12} levels are usually measured nowadays by radioisotope dilution assay, generally using pure intrinsic factor as the specific binder. Normal levels are above 200 pg ml^{-1}, and levels below 160 pg ml^{-1} (118 pmol l^{-1}) are indicative of deficiency. Many commercial kits are now available for the radioisotope dilution assay. (*See* **Immunoassays:** Radioimmunoassay and Enzyme Immunoassay.)

An older functional test of vitamin B_{12} or folate status is the deoxyuridine suppression test, usually performed *in vitro* on bone marrow aspirates. This looks directly at the cells' ability to synthesize thymidine from deoxyuridine. For B_{12}-deficient cells, B_{12} cofactor is the only addition which will partially overcome the block in conversion. This test has the potential advantage of being a functional test and of probing a specific tissue site, but it is time-consuming, and it is now mainly of research, rather than diagnostic, importance. (*See* **Coenzymes**.)

Another functional test involves the metabolism of an oral load of valine: if B_{12} deficiency is present, then methylmalonyl CoA mutase activity falls (see below) so that methylmalonic acid accumulates, and increased amounts of this byproduct are found in the plasma and excreted in the urine. Moderately raised plasma levels of methylmalonic acid can also arise in subjects with marginal vitamin B_{12} status, even without the loading dose, and this is becoming the basis for a useful diagnostic procedure for marginal status. The most commonly used assay procedure for serum or urinary methyl malonic acid is based on mass spectrometry, although other methods, such

as capillary electrophoresis, are now also becoming available. The assay is still too demanding of skill and expensive equipment for routine clinical chemistry laboratories, but it is becoming the index of choice for the demonstration of mild functional vitamin B_{12} deficiency. Circulating levels of homocysteine can provide further evidence, but since folate and vitamin B_6 status can also affect homocysteine economy, the evidence that it provides is less specific, and requires the investigation of the other vitamins.

Biological Functions of Vitamin B_{12}

Three vitamin B_{12}-dependent enzymes occur in animals. The first, methionine synthase or synthetase, has the Enzyme Commission (EC) official name 5-methyltetrahydrofolate-homocysteine *S*-methyltransferase, EC 2.1.1.13, and requires methylcobalamin as the cofactor or cosubstrate. The other two vitamin B_{12}-dependent enzymes both require adenosyl cobalamin: they are methylmalonyl-coenzyme A (CoA) mutase EC 5.4.99.2, responsible for conversion of methylmalonyl-CoA to succinyl CoA during the oxidation of propionate (see above), and leucine 2,3-aminomutase, EC 5.4.3.7, which converts L-α-leucine to aminoisocaproate as the first step in leucine degradation or synthesis. There is a rare congenital disorder in which the enzyme methylmalonyl-CoA mutase is almost absent, and this is characterized by methylmalonic aciduria, homocysteinuria, muscle cramps, mental retardation, and lethargy. Large parenteral doses of vitamin B_{12} may be helpful.

The most important biochemical reaction catalyzed by vitamin B_{12} in humans and animals is the conversion of homocysteine to methionine. However, the most obvious result of deficiency *in vivo* is a block in DNA synthesis, leading characteristically to megaloblastosis, and macrocytic anemia, which is indistinguishable, at least at the cytological level, from the megaloblastosis and anemia of folate deficiency. A good test for response of a probable case of pernicious anemia, or other B_{12}-deficiency syndrome, is an increase in circulating reticulocyte numbers following B_{12} treatment. (*See* **Amino Acids**: Metabolism; **Anemia (Anaemia)**: Megaloblastic Anemias.)

The most generally accepted explanation for the connection between B_{12} deficiency and the failure of cell division leading to anemia is the 'methylfolate trap hypothesis', in which methylene tetrahydrofolate polyglutamate, the intracellular cofactor for thymidine synthesis, becomes depleted by excessive conversion (i.e., reduction) to methyltetrahydrofolate polyglutamate. The latter is unable to transfer its carbon unit (methyl group) to homocysteine and thus complete the folate cycle back to the methylene form, because of the absence of B_{12} cofactor (methylcobalamin). At the same time, short-chain methylfolate glutamates from the diet are poorly utilized and therefore excreted, so that intracellular folate levels decline. (*See* **Folic Acid**: Physiology.)

An alternative interpretation, the 'formate starvation hypothesis', maintains that because B_{12} deficiency causes a reduction in methionine levels, an important consequence of its deficiency is a reduced conversion of the methyl group of methionine to formyltetrahydrofolate, which is a good precursor of folate polyglutamates and hence of active methylene groups for thymidine synthesis. In the absence of vitamin B_{12}, formate accumulates in blood, liver, and brain, and there is increased formate excretion in the urine. In support of this hypothesis, formylfolate, but not tetrahydrofolate, can correct the functional evidence of B_{12} deficiency, and the organism is still able to oxidize the methyl group of methylfolate to methylene- and formyl-folates. One very useful experimental tool has been the use of nitrous oxide, which can chemically inactivate vitamin B_{12} *in vivo*, e.g., in experimental animals. This procedure has permitted the rapid induction of tissue B_{12} deficiency, without the tedious and protracted use of B_{12}-deficient diets.

An important functional result of vitamin B_{12} deficiency is the failure of maintenance of nerve tissue myelin, which accounts for the irreversible neurological damage seen after prolonged B_{12} deficiency. This lesion may also be connected biochemically with the block in methionine formation (and hence in phospholipid metabolism), but its etiology has not yet been fully elucidated. There may also be other disturbances in lipid metabolism in B_{12}-deficient subjects.

Human Vitamin B_{12} Requirements

In the absence of the metabolic abnormalities that can affect B_{12} absorption and utilization, human dietary requirements for the vitamin are small. A purely nutritional deficiency may, however, occur in strict vegans and lifelong vegetarians such as strict Hindus who have very low intakes because they do not eat any animal products. Studies of the amount of B_{12} needed to cure this, together with studies of populations with low intakes who, nevertheless, do not show deficiency signs, have indicated that the requirement is generally less than $1\,\mu g\,day^{-1}$ for adults, and is probably in the range $0.1–1\,\mu g\,day^{-1}$. There is little firm evidence for a major increase in requirement during pregnancy and lactation, but modest increments in the recommended dietary allowances (RDAs) are usually applied to these physiological

states. The amount secreted in breast milk ranges from 0.1 to 3 μg day^{-1}. Infants who were breast-fed by mothers who were themselves deficient and were secreting milk with a very low content of the vitamin have in several reported instances been diagnosed with B_{12}-responsive megaloblastic anemia. (*See* **Lactation**: Human Milk: Composition and Nutritional Value; Physiology.)

The early recognition of B_{12} deficiency in certain immigrant vegetarian groups, especially Asians, merits attention, especially in western countries where bacterial contamination sources of the vitamin are eliminated more completely than in developing countries.

Typical B_{12} intakes by omnivorous people in western countries are 3–30 μg day^{-1}.

Food table values of B_{12} contents of foods have been obtained mainly by microbiological assays and some further verification is needed. Vitamin B_{12} is stored mainly in the liver (60%) and muscles (30%). Methylcobalamin is the most abundant form in human plasma, whereas in most human tissues, deoxyadenosyl cobalamin is the most abundant, with aquacobalamin coming second.

At high intakes, there is little evidence of toxicity, and injections of as much as 3 mg day^{-1} have been used in an attempt to treat fatigue and neurological disorders. However, there is no evidence that such high doses can have any benefit for normal subjects and they may occasionally result in allergic reactions. Degradation of B_{12} to B_{12} antagonists in multinutrient preparations has been claimed.

The RDAs from the World Health Organization (WHO) currently stand at 1 μg day^{-1} for adults. In the UK the reference nutrient intake for vitamin B_{12} is 1.5 μg day^{-1} for adults, and in the USA, the dietary reference intake is 2.4 μg day^{-1}. Thus there is a rather large variation between the different sources of reference values. Values for pregnancy and lactation are generally slightly higher than for nonpregnant, nonlactating adults. A typical treatment schedule for pernicious anemia would be 500 μg by injections every 2–3 months.

Vitamin B_{12}: Increased Vulnerability in Older People

The aging process is characterized by progressively increasing vulnerability to serious impairment of vitamin B_{12} absorption, and hence the development of functional deficiency. The most widely known example of this is pernicious anemia, often resulting from autoantibodies which interfere with the functions of intrinsic factor, and which can be detected by blood antibody analysis. Another common cause is atrophic gastritis, which reduces gastric acid production and thus interferes with the liberation of the vitamin from the protein binders in food. Several recent studies have highlighted the evidence for a relatively common, mild vitamin B_{12} deficiency in older people living in western countries. There is recent evidence of racial differences in vulnerability to vitamin B_{12} deficiency in older Americans: white and Latin subjects are more vulnerable than black or Asian ones. Raised methylmalonic acid levels and lowered levels of holotranscobalamin II are found. As absorption becomes more severely impaired, the normal B_{12} enterohepatic cycle is also affected, which increases the rate of loss of B_{12} (normally *c.* 0.1% day^{-1}). The longer a deficiency state goes undetected, the greater the danger of permanent neurological damage. In the absence of the highly efficient intrinsic factor-catalyzed absorption pathway, the alternative passive absorption route is only about 1% as efficient; therefore an intake at least 100-fold greater than the normal requirement is needed. At around 100–200 μg day^{-1} this is not available (in the UK) in nonprescribed vitamin supplements; therefore diagnosis and treatment are of critical importance. There is some concern that recent public health interventions designed to increase the level of food fortification with folate may mask the anemia of incipient vitamin B_{12} deficiency, and thus delay diagnosis.

See also: **Amino Acids**: Metabolism; **Anemia (Anaemia)**: Megaloblastic Anemias; **Coenzymes**; **Folic Acid**: Physiology; **Immunoassays**: Radioimmunoassay and Enzyme Immunoassay; **Lactation**: Human Milk: Composition and Nutritional Value; Physiology; **Macrobiotic Diets**; **Microflora of the Intestine**: Role and Effects; **Vegetarian Diets**

Further Reading

Chanarin I (1990) *The Megaloblastic Anaemias*, 3rd edn. Oxford: Blackwell.

Mann NJ, Li D, Sinclair AJ *et al.* (1999) The effect of diet on plasma homocysteine concentrations in healthy male subjects. *European Journal of Clinical Nutrition* 53: 895–899.

Nilsson-Ehle H (1998) Age-related changes in cobalamin (vitamin B_{12}) handling. Implications for therapy. *Drugs and Aging* 12: 277–292.

Savage DG and Lindenbaum J (1995) Neurological complications of acquired cobalamin deficiency: clinical aspects. *Bailliere's Clinical Haematology* 8: 657–678.

Scott JM (1997) Bioavailability of vitamin B_{12}. *European Journal of Clinical Nutrition* 51: (suppl. 1): S49–S53.

Scott JM (1999) Folate and vitamin B_{12}. *Proceedings of the Nutrition Society* 58: 441–448.

Seetharam B (1999) Receptor-mediated endocytosis of cobalamin (vitamin B_{12}). *Annual Review of Nutrition* 19: 173–195.

Stabler SP, Lindenbaum J and Allen RH (1997) Vitamin B_{12} deficiency in the elderly. *American Journal of Clinical Nutrition* 66: 741–749.

Woodson RD (ed.) (1990) New frontiers in vitamin B_{12} metabolism. *American Journal of Hematology* 34: 81–139.

Zittoun J and Cooper BA (eds) (1989) *Folates and Cobalamins*. Berlin: Springer-Verlag.

COBALT

A MacPherson and J Dixon, The Scottish Agricultural College, Auchincruive, Ayr, UK

Background

Cobalt was first isolated by a Swedish chemist, Brandt, in 1735 and characterized more fully by Bergman in 1780. However, it had been used as an ore for millennia to give a bright blue color to Egyptian pottery and Iranian glass from 2600 to 2250 BC, respectively. The name is derived from the German word *kobold*, meaning an evil spirit or gnome, as copper miners working in the Harz mountains believed that these spirits delighted in exposing ores that looked like copper but yielded none when smelted.

Cobalt is widely distributed in the earth's crust but it is only 30th in order of abundance and thus less common (at 25 mg kg^{-1}) than all other elements in the first transition series with the exception of scandium (22 mg kg^{-1}). There appears to be some overall loss of cobalt during soil formation as the world mean value is quoted as 12 mg kg^{-1} and Scottish soils at 8 mg kg^{-1}. World consumption of cobalt was 31 400 tonnes in 1997 and is projected to increase by 50–100% by 2012. Asia, including China, is the largest consuming region, and the chemical sector is the largest end user. The use of cobalt in chemicals for rechargeable batteries is having a significant impact on total consumption.

The metal itself is silvery in appearance with a bluish tinge and considerably harder than iron. It has only one naturally occurring isotope, ^{59}Co, but this can be converted to radioactive ^{60}Co by bombardment with thermal neutrons. This has a half-life of 5.271 years and decays to nonradioactive ^{60}Ni. The radioactive isotope ^{60}Co is used medically in cancer therapy, where it can selectively destroy abnormal areas in the brain, such as tumors or arteriovenous malformations, while leaving normal tissue unaffected. This procedure has been termed 'radiosurgery.' Nitrosylcobalamin shows potential as a biologically compatible and selective chemotherapeutic agent as tumor cells have surface receptors for vitamin B_{12}. ^{60}Co also affords a concentrated source of radiation for research purposes. ^{57}Co is used in the assay of vitamin B_{12} by radioisotope dilution.

Like iron, cobalt is a ferromagnetic element, but is considerably less chemically reactive. It is stable to oxygen, and the range of oxidation states is considerably fewer than for earlier members of the first transition series. The commonest are $+2$ and $+3$, but, as the latter is a strong oxidizing agent, it decomposes rapidly in aqueous solution with the production of oxygen. As a result, only a few Co^{3+} salts exist, and these tend to be unstable. However, Co^{3+} forms a large number of coordination complexes, particularly with nitrogen donor ligands, and this is important in its biological role as a structural component of vitamin B_{12}. This structure is characterized by a 'corrin' ring, where the cobalt is coordinated to four coplanar nitrogen atoms with another (imidazole) nitrogen in the fifth position. The sixth position, however, makes vitamin B_{12} unique as, here, the cobalt is bonded to carbon, making it to date the only naturally occurring organometallic vitamin.

The incorporation of cobalt into the corrin ring affects its reduction potential and gives it the three consecutive oxidation states shown in eqn (1). The reduced Co^{+} species is a highly reactive compound and can liberate hydrogen from water. Both vitamin B_{12r} and vitamin B_{12s} are very labile in the presence of air and are instantly oxidized to the Co^{3+}

$$\underset{\text{Red}}{\text{Vitamin } B_{12}(Co^{3+})} \underset{O_2}{\overset{NaBH_4,H^+}{\rightleftharpoons}} \underset{\text{Brown}}{\text{Vitamin } B_{12r}(Co^{2+})}$$

$$\underset{O_2}{\overset{NaBH_4,OH^-}{\rightleftharpoons}} \underset{\text{Blue–gray green}}{\text{Vitamin } B_{12s}(Co^{+})} \qquad (1)$$

compound cobalamin. The sixth coordination site is crucial to the biological role of vitamin B_{12}, where it is involved as a hydrogen carrier in substitution reactions of the type shown in eqn (2):

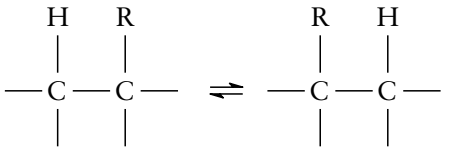

(2)

Occurrence in Soil

This is primarily dependent on the geological nature of the parent material from which the soil is derived. Thus, soils formed from old red sandstone, granitic, and other acid igneous, limestone, and predominantly sandy rocks are liable to be cobalt-deficient.

The availability of soil cobalt and hence its uptake by pasture and crops are affected by the natural drainage conditions of the soil, and so cobalt deficiency is most likely to be found on freely drained soils. Improvement of upland pasture by the removal of heather and the introduction of more productive grass species, raising the soil pH by liming, and improving the drainage tend to lower the availability of soil cobalt and thus induce cobalt deficiency in the grazing animal where none existed previously. Soil maps have been produced based on the cobalt content of the B horizon and these categorize the concentrations (in $mg\,kg^{-1}$) as follows: low <5, medium 5–15 and high >15. Areas falling into the 'low' category have the highest risk of cobalt deficiency. Soils that have a high content of secondary manganese or iron oxides adversely affect cobalt availability, as they strongly absorb native cobalt.

Concentrations in Food and Dietary Intakes

Cobalt, like other trace elements, tends to be concentrated in young actively growing parts of plants, with green leafy material often having the highest concentrations. Concentrations tend to be lower in stems, roots, tubers, and cereal grains and also tend to diminish as maturity is reached. Typical values for some foods are given in Table 1 but are scarce. A dietary and nutritional survey of British adults conducted in 1995 concluded that average dietary intakes of cobalt were lower than those reported previously. The Committee on Toxicology stated that levels of cobalt in the UK diet are not a cause for toxicological concern.

There is, however, a wide margin of safety between the cobalt concentration of ordinary diets and toxic levels, which are around 250–300 $mg\,kg^{-1}$ and thus more than 1000 times normal dietary concentrations. High cobalt levels have, however, been implicated as a triggering factor in cases of severe cardiac failure in heavy beer drinkers (up to 12 l daily). Cobalt had been added as a foaming agent in concentrations of 1.2–1.5 $mg\,l^{-1}$, but high alcohol and cobalt intakes, together with poor-quality diets, combined to produce a distinctive cardiomyopathy, and so this practice has been discontinued. Cardiomyopathy has also been reported in humans following industrial exposure to cobalt. Studies in Germany have suggested that

Table 1 Range of cobalt concentrations in foods

Food	*Co ($mg\,kg^{-1}$)*[a]	*Co ($mg\,kg^{-1}$)*[b]	*Co ($mg\,kg^{-1}$)*[c]
Green leafy vegetables	0.20–0.60	0.009	0.01–0.12
Dairy products	0.01–0.03	0.004	0.0005–0.001
Liver and kidney	0.15–0.25	0.06	0.004–0.47
Muscle meat	0.06–0.12	0.004	0.001–0.007
Cereal grains	0.01–0.04	0.01	0.02–0.085
Sugar	0.01–0.03	0.03	
Fish	<0.10–0.03	0.01	0.004–0.06
Milk	0.50 $\mu g\,l^{-1}$	0.002	0.0008
White bread		0.02	
Wholemeal bread			
Fruit		0.004	0.002–0.15
Eggs		0.002	0.06–0.35
Peas/beans		0.009	0.03–0.08
Root vegetables		0.006	0.006–0.02
Poultry		0.003	0.019–0.02
Nuts		0.09	0.09–0.34

[a]Values (dry matter) from Mertz W (1987) *Trace Elements in Human and Animal Nutrition, 5th edn*. London: Academic Press.
[b]Data (fresh weight) from Ministry of Agriculture, Fisheries and Food Total Diet Study (1994) and published by Ysart G, Miller P, Crews H *et al*. (1999) *Food Additives and Contaminants* 16: 391–403.
[c]Data (fresh weight) from Souci SW, Fachmann W and Kraut H (1994) *Food Composition and Nutrition Tables, 5th revised edn*. Stuttgart: Scientific Publishers.

solutions used for total parenteral nutrition (TPN) are virtually all contaminated with cobalt, but, even so, several preparations include supplemental cobalt in addition. Serum cobalt concentrations in preterm babies on TPN at $1.44 \pm 0.48\,\mathrm{ng\,l^{-1}}$ have been found to be significantly higher than those ($0.67 \pm 0.47\,\mathrm{ng\,l^{-1}}$) of age-matched controls. No toxic consequences have been reported, but any clinical consequences of long-term cobalt surplus have yet to be studied systematically. Most species studied have a high tolerance to cobalt, particularly sheep, but recent evidence suggests that this does not extend to lambs < 6 weeks old. Cattle are more susceptible to toxic effects from excess cobalt intakes.

Cobalt is present in very low concentration in body fluids and tissues, with the total content of an adult human body being < 1.5 mg, with liver, heart, and bone containing the highest concentrations. Reference levels are said to be 0.22, < 0.16, 0.06, 0.11, and < 0.08 mg of cobalt per kilogram (dry matter) in liver, spleen, kidney, heart, and pancreas, respectively. Cobalt deficiency in ruminants has been diagnosed on the basis of liver concentrations of $< 0.06\,\mathrm{mg\,kg^{-1}}$ (dry matter) with values $> 0.1\,\mathrm{mg\,kg^{-1}}$ being regarded as adequate. Deficiency during pregnancy will result, with the offspring being born with inadequate concentrations, and this has an adverse effect on their viability. Supplementation of the dams diet or by direct treatment overcomes this problem.

Cobalt differs from all other essential trace elements in that it is required by the body in a preformed compound, vitamin B_{12}, whereas the other elements are required in ionic form and then converted into their metabolically active species. Animals and humans are unable to synthesize vitamin B_{12}. This ability is found only in some bacteria and algae. Some of these occur in ruminants, so that, provided there is an adequate supply of dietary cobalt salts ($> 0.1\,\mathrm{mg\,kg^{-1}}$, dry matter), vitamin B_{12} is synthesized in the rumen and then absorbed lower down the intestinal tract. Humans derive their essential vitamin B_{12} through the consumption of animal-related foods. It has been thought that animals do not produce or require these cobalt-related compounds, and strict vegetarians and vegans are at risk from pernicious anemia, unless they take synthetic supplements. Recent evidence suggests that this may not be strictly true, but it remains practically so, as the amounts involved would be of little nutritional significance.

The cobalt atom in vitamin B_{12} catalyzes the reactions of two B_{12} coenzymes – adenosyl cobalamin and methyl cobalamin – which are essential cofactors in the activity of methylmalonyl-coenzyme A mutase and methionine synthetase. The unique nature of the cobalt–carbon bond in adenosylcobalamin means that it can undergo reversible homolytic splitting with the production of free radicals, which are stabilized by the cobalt atom. This allows the intramolecular rearrangements of the type referred to earlier to take place. Methylcobalamin has a similar cobalt–carbon bond that is involved in the synthesis of methionine. The mechanism has not yet been fully established, but it is thought that vitamin B_{12s} (in the Co^{+} form) is involved, as it can undergo rapid addition and substitution reactions.

Analytical Methods

Sample Preparation

Since many of the available analytical methods require cobalt to be in solution, solid samples such as foods and clinical and biological materials need to be subjected to a primary dissolution process. Most commonly, this involves either acid digestion or dry-ashing procedures, which mineralize cobalt through destruction of the organic matter present. The most frequently used ashing technique is by muffle furnace at temperatures within the range 450–550 °C. Alternatively, there are low-temperature ashers that operate between 100 and 200 °C under low pressure using a gaseous oxidant, usually oxygen or an oxygen/tetrafluoromethane mixture; the oxidant is activated by a high-frequency electromagnetic field. Dissolution of cobalt from the ashed material is usually achieved using dilute acid.

There are a number of different combinations of acids and oxidizing chemicals that have proved to be effective digestion mixtures, namely nitric acid/perchloric acid, nitric acid/perchloric acid/sulfuric acid, and sulfuric acid/hydrogen peroxide. These operations have been traditionally performed in an open Kjeldahl-type flask; however, much more rapid digestion procedures have recently been developed using Teflon bomb vessels and microwave heating.

Dissolution of cobalt from soils and related materials can be performed using strong acid digestion, normally hydrofluoric acid or a hydrochloric acid/nitric acid mixture (aqua regia), whereas the more readily available fractions of cobalt are extracted with dilute acetic acid. Water samples and biological fluids such as urine and blood plasma can be analyzed directly using one of the more modern sensitive techniques such as electrothermal atomization atomic absorption spectrometry (ETAAAS).

Where limited sample preparation of solid material, other than drying and milling, is required, a solid sample presentation can be made using direct current arc emission spectrometry (DCarcES), direct current

plasma atomic emission spectrometry (DCPAES) and ETAAAS; presentation in slurry form can also be made with the last two techniques. Nondestructive analysis can be performed using X-ray fluorescence (XRF).

Analytical Techniques

Traditional gravimetric and titrimetric methods have much poorer levels of detection for cobalt than modern instrumental techniques and are subject to interferences. Two of the more commonly used precipitating reagents are 1-nitroso-2-naphthol (interference from copper and iron) and anthranilic acid (interference from iron, nickel and zinc). Typical titrants are potassium cyanide (interference from copper, mercury, and zinc), potassium hexacyanoferrate (interference from manganese), and ethylenediaminetetra acetic acid (EDTA) (interferences from nickel, zinc, and copper).

Prior to the rapid expansion of flame atomic absorption spectrometry (FAAS) in the 1960s, the analysis of cobalt at the microgram level was performed routinely by colorimetric techniques using spectrophotometry in the ultraviolet and visible wavelength range (300–700 nm). However, this technique is also subject to chemical interference from a wide range of other elements. These have to be eliminated either by the addition of masking agents or buffers, or by solvent extraction of the cobalt complex. The more widely used complexing agents are ammonium thiocyanate (interferences from iron and copper), nitroso-*R*-salt (interferences from iron, chromium, nickel, vanadium, and copper), 1-nitroso-2-naphthol (interference from copper and iron), and sodium diethyldithiocarbamate (interferences from iron, aluminum, chromium, titanium, manganese, copper, and nickel). By applying modern chemometric procedures to spectrophotometric measurements, cobalt can be determined simultaneously with other trace elements. Flow injection analysis has automated many cobalt colorimetric methods. Recently, sensitive spectrophotometric methods have been developed that depend on the catalytic effect of the cobalt iron on the oxidation reaction between hydrogen peroxide and *o*-dihydroxybenzene derivatives such as tiron (disodium 1,2-dihydroxybenzene-3,6-disulfonate), catechol, quinazarin, and gallocyanine; detection limits at the picogram level are obtainable. A more recent novel utilization of catalysis incorporates chemiluminescence sensors that measure the catalytic activity of cobalt on the reaction between luminol and certain oxidizing agents such as hydrogen peroxide or periodate salts.

The following spectrographic techniques have been used for the analysis of cobalt (approximate detection limits have been quoted to an order of magnitude): FAAS (detection limit, $0.01\,\mu g\,g^{-1}$), ETAAAS (detection limit, $0.0002\,\mu g\,g^{-1}$), inductively coupled plasma atomic emission spectrometry ICPAES, detection limit, $0.01\,\mu g\,g^{-1}$), inductively coupled plasma mass spectrometry (ICPMS, detection limit, 0.0002 $\mu g\,g^{-1}$), DCarcES (detection limit, $0.3\,\mu g\,g^{-1}$), spark source mass spectrometry (SSMS, detection limit, 0.005 $\mu g\,g^{-1}$) and DCPAES (detection limit, $0.01\,\mu g\,g^{-1}$).

FAAS has been the most extensively used technique for the routine determination of cobalt at the microgram level in food, agricultural, and biological materials. The technique is relatively free from chemical and spectral interferences, and a large number of preconcentration chemical procedures have been developed to improve detection levels. Ammonium pyrrolidinedithiocarbamate and sodium diethyldithiocarbamate are the two principal chelating agents that have been used for multitrace element complexation prior to solvent extraction, although a more cobalt-specific chelate is 1-nitroso-2-naphthol. A range of different organic solvents have been used for extraction of the chelated cobalt compounds, the two most common being methyl isobutyl ketone and chloroform. Ion-exchange resins such as chelex-100 or C-18 sorbant materials can be used as a method of on-line preconcentration in FIA systems for FAAS.

ETAAAS has proved to be sufficiently sensitive for the determination of cobalt at the nanogram level in clinical and biochemical samples. The technique is more susceptible to sample matrix interferences than FAAS, but these have been significantly reduced by using FAAS-type preconcentration techniques as separation procedures for the removal of major elements. Alternatively, matrix modifiers such as palladium, magnesium, or certain inorganic salts such as ammonium phosphate or thiocyanate have proved effective. Graphite tubes are normally used for atomization, but more recently, tungsten tubes have been found to be equally effective. Automatic on-line preconcentration systems have been developed that utilize a microcolumn packed with Muromac A-1 chelating resin. Analysis of solid samples is now possible through the incorporation of minature cups that can hold powdered samples. The resonant absorption wavelength for cobalt lies in the ultraviolet region at 240.7 nm. Measurements are therefore subject to background interferences that occur to a greater degree in electrothermal rather than flame atomization. All modern AAS instruments have automatic background correction; there are four different types: Zeeman, deuterium continuum, Smith–Hieftje, and xenon continuum in the simultaneous

multielement atomic absorption continuum instruments. The normal detection limit for cobalt as determined by ETA-AAS has been further improved through the coupling of laser technology with electrothermal atomization. This is either in the form of laser-excited atomicfluorescence spectrometry or with the laser-enhanced ionization technique.

The 1980s saw a rapid growth in the number of commercial ICPAE spectrometers being used for multielement analysis. Since the detection limit for cobalt by ICPAES is similar to that for FAAS, similar preconcentration techniques have been applied. The two most prominent emission lines are almost equal in sensitivity; however, since the line at 238.8 nm is subject to strong spectral interference from iron, the cobalt line at 228.6 nm has been preferentially used, but this suffers from a significant spectral interference from titanium and slight spectral interference from nickel and iron. The exact magnititude of the interfering spectral overlap depends on the spectral resolution of each particular type of spectrometer. Interfering elements that produce strong spectral overlap used to be removed by chemical separation prior to instrumental analysis, whereas those that exhibit a small degree of spectral overlap can be corrected for by appropriate interelement spectral subtraction procedures.

Since the mid 1980s, there has been a steady growth in the application of ICPMS to multielement ultratrace analysis of clinical and biochemical samples. Detection limits are about the same as for ETAAAS, and therefore, cobalt can be determined at the nanogram level. Interferences occur from the production of polyatomic ions within the plasma that have the same atomic mass as ^{59}Co, e.g., calcium–oxygen and argon–sodium. These are corrected for by measuring the background levels of such ions in simulated blank solutions. Automatic on-line preconcentration systems have also been developed for ICPMS.

DCarcES, SSMS, and XRF have mainly been applied to the direct analysis of cobalt in solid samples such as soils and related materials; with these techniques, matrix matching of samples with calibration standards and choice of internal standards are the most critical aspects of the determination. The application of the DCarcES to plant and agricultural materials has been largely superceded by AAS and ICPAES.

Electrochemical techniques have proved to be very effective in the determination of cobalt at the microgram and nanogram levels in food, biological, and environmental samples, being among the most sensitive techniques for the determination of cobalt in natural waters. During the 1980s, the technique of adsorptive stripping voltametry was developed as the most sensitive of the polarographic methods for trace metal analysis. For the determination of cobalt, the square wave stripping is more sensitive than the differential stripping mode. Electrolytic preconcentration of cobalt at the mercury electrode is carried out with chelated cobalt, since the free cobalt ion would be irreversibly adsorbed. The most widely used chelating agent has been dimethyl glyoxime; adsorption of the cobalt dimethyl glyoximate occurs at -0.7 V, and the voltammetric peak at -1.12 V in a cell system using a silver/silver chloride/potassium chloride electrode as the nonworking electrode. Interference from zinc can be masked by the addition of nitrilotriacetic acid or sodium iminodiacetate; nickel interferes if present in great excess, but this is reduced by using alternative chelating agents such as 1,2-cyclohexanedione dioxime or α-furil dioxime for complexing the cobalt. More recently, cathodic stripping voltammetry coupled with FIA has become a successful addition to the range of electrochemical techniques used for cobalt analysis.

High-performance liquid chromatography (HPLC) has been successfully applied to the determination of cobalt and other trace metals in plant materials, natural waters, and environmental samples, and detection levels of nanogram quantities are obtainable. The use of disubstituted dithiocarbamates such as sodium diethyldithiocarbamate, ammonium pyrrolidinedithiocarbamate, and ammonium bis(2-hydroxyethyl)dithiocarbamate has been favored, since they rapidly form stable complexes with a wide range of trace metals. These metal complexes can be separated chromatographically using alkyl-bonded silica columns and quantitatively determined with ultraviolet/visible diode array detectors or electrochemical detectors. Precolumn derivitization can be done off-line using solvent extraction for water-insoluble complexes or in-line using a presentation column loaded with an appropriate ion-pair reagent such as cetyltrimethylammonium bromide-dithiocarbamate. On-column derivatization can be performed by including the dithiocarbamate reagent in the mobile phase, usually a water/methanol mixture; however, this process is less sensitive than precolumn derivatization. Other complexing chelating agents that can be used include EDTA or 4-(2-pyridylazo)resorcinol. Metal ion chromatography can be applied to cobalt analysis using an appropriate complexing chelating agent such as pyridine-2-carboxyaldehyde phenylhydrazone as a stationary phase immobilized on a silica matrix.

HPLC has been successfully interfaced with a wide range of spectroscopic instruments (FAAS, ETAAAS, ICPAES, and ICPMS) to provide powerful techniques

for cobalt speciation analysis. The cyanocobalamin form of cobalt has been measured in blood serum, seminal fluid, and milk serum. Vitamin B_{12} analogs have also been determined in animal carcass tissue. These speciation techniques should see intense development in future.

Neutron activation analysis can be used to determine the amount of cobalt in biological and environmental samples; typical detection levels are $0.03\,\mu g\,g^{-1}$. This technique usually requires a post-irradiation period of around 3 weeks before γ-ray emission from the ^{60}Co isotope can be measured. Radiation flux levels are usually around 10^{12}–10^{13} neutrons $cm^{-2}s^{-1}$. Concentration techniques using dithiocarbamate complexation are usually applied to improve detection levels.

See also: **Chromatography**: High-performance Liquid Chromatography; **Cobalamins**: Properties and Determination; Physiology; **Mass Spectrometry**: Principles and Instrumentation; Applications

Further Reading

Barceloux DG (1999) Cobalt. *Journal of Toxicology – Clinical Toxicology* 37: 201–216.

Brown S and Savory J (1984) *Chemical Toxicology and Clinical Chemistry of Metals*. New York: Academic Press.

Schneider Z and Stroinski A (1987) *Comprehensive B_{12}*. Berlin: Walter de Gruyter.

Smith IC and Carson BL (1981) *Trace Metals in the Environment*. Cobalt, vol. 6, pp. 1202. Ann Arbor, MI: Ann Arbor Science.

Smith KA (1990) *Soil Analysis, Modern Instrumental Techniques*, 2nd edn. New York: Marcel Decker.

Smith RM (1987) Cobalt. In: Mertz W (ed.) *Trace Elements in Human and Animal Nutrition*, 5th edn. London: Academic Press.

Underwood EJ and Suttle NF (1999) *The Mineral Nutrition of Livestock*, 3rd edn. Wallingford, UK: CAB International.

Ysart G, Miller P, Crews H *et al*. (1999) Dietary Exposure estimates of 30 elements from the UK Total Diet Study. *Food Additives and Contaminants* 16: 391–403.

Cockles *See* **Shellfish**: Characteristics of Crustacea; Commercially Important Crustacea; Characteristics of Molluscs; Commercially Important Molluscs; Contamination and Spoilage of Molluscs and Crustaceans; Aquaculture of Commercially Important Molluscs and Crustaceans

COCOA

Contents
Chemistry of Processing
Production, Products, and Use

Chemistry of Processing

B Biehl, Technische Universität Braunschweig, Germany
G Ziegleder, Fraunhofer-Institut für Verfahrenstechnik und Verpackung, Freising, Germany

Introduction

The most valuable constituents in cocoa are cocoa butter and flavors. Both are specific for seeds of *Theobroma cacao* L. Whilst the quality of cocoa butter is not changed during processing, the particular flavor of cocoa depends on profound chemical reactions which take place during processing and manufacturing. The unprocessed cocoa seed would not give the characteristic flavor when roasted unless it was first fermented and dried. The sequence of raw cocoa processing in the tropics and then manufacturing is essential since reactions during fermentation depend on water and enzymes of the living seed, whilst manufacturing comprises nonenzymatic reactions in a dry lipid phase. Raw cocoa beans are the link between processing (normally in the tropics) and manufacturing (normally in the consuming

countries). A large number of essential constituents of cocoa flavor are produced during roasting of raw cocoa from flavor precursors formed during tropical processing. Reactions during fermentation and drying resulting in raw cocoa beans and subsequent reactions taking place during manufacturing are described in this first article.

Chemistry of Raw Cocoa Processing

Both the seeds and the pulp are essential components in raw cocoa processing. However, the reactions which are essential for raw cocoa quality take place within the seeds alone and not in the pulp. Enzymes of the seeds are directly involved but microorganisms only indirectly, in that bacteria and fungi are responsible for the degradation of the pulp, and their metabolites, especially acetic acid, which are instrumental in the subsequent changes in the seeds. Organisms participating in the direct attack on the nibs during the late stage of fermentation and overfermentation are also important.

Processing depends on the chemical composition and physiological state of the beans, their enzymes, and their pre- and postmortem subcellular structures. They greatly affect the final product quality.

Biology of the Seed

Two commercially exploited types of the species *T. cacao*, Criollo and Forastero, are distinguished on the basis of fruit morphology and geographic origin. Criollo gives fine-flavor cocoas, although in low yields, while the more vigorous, high-yielding Forastero subtypes and hybrids predominate in plantations worldwide. Cocoa pods contain about 35–45 seeds consisting of the embryo and the shell, which is covered when ripe by a mucilaginous pulp, the endocarp (Figure 1). The main parts of the embryo are two folded cotyledons connected by a small embryonic axis. A rudimentary, skin-like endosperm covers the surface of the embryo. After fermentation the shell, covered with residues of pulp, comprises about 12–16% of the seed dry weight, depending on bean size and pulp degradation during fermentation. The dry weight of a seed varies considerably, but is approximately 1.0–1.2 g. Only the cotyledons are used for cocoa and chocolate manufacturing and in these the leaf mesophyll is the dominating tissue, consisting of two types of cells which differ in composition.

About 80% of the cells store lipid and protein, the major volume of which is occupied by a large number of individual lipid bodies of constant size (about 2 μm) surrounding one or more branched protein storage vacuoles and amyloplasts. The remaining 20% of the cells are polyphenol storage cells. Their lumen is almost entirely occupied by one central vacuole containing all the stored polyphenols and purines.

Figure 1 Unfermented cocoa seeds showing the pulp, the testa and the cotyledons. Reproduced from Cocoa: Chemistry of Processing. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Subcellular structures in the lipid and protein storage cells control pre- and postmortem reactions during fermentation, which are not the same as in cell free systems. On the subcellular level *in vivo*, the situation can roughly be described as a dispersed lipid phase (lipid bodies) in a continuous water phase (cytoplasm; Figure 2). During fermentation, the cells are killed by heat and acetic acid, but postmortem structures are maintained. In the cell lumen, with low concentrations of acetic acid, the lipid bodies fuse *in situ*, and they form, more or less, a continuous lipid phase which separates cytoplasmic inclusions as a hydrophilic dispersed phase. In contrast, in the presence of a high concentration of acetic acid, fusion of the lipid bodies is more extensive, causing complete

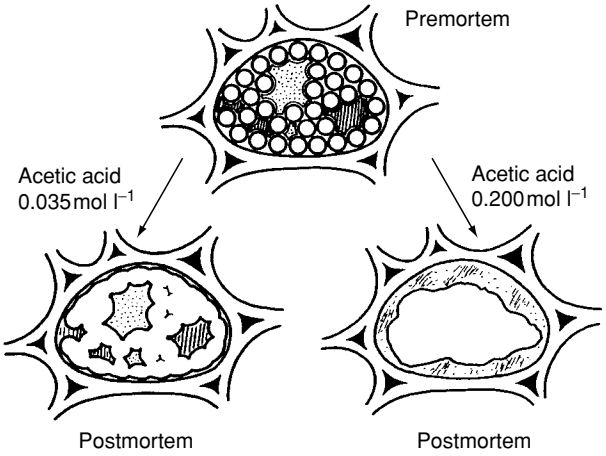

Figure 2 Schematic representation of electron microscopic aspects of postmortem changes in mesophyll cells caused by different concentrations of acetic acid during fermentation-like seed incubation. Spherical lipid globules in the living cell fuse *in situ* (left) or form a central bulk of lipid (right). Hydrophilic compartments are indicated by dots or dashes. Reproduced from Cocoa: Chemistry of Processing. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

segregation of the lipid in the cell center (Figure 2). In the first case, intracellular diffusion of water-soluble compounds is reduced by lipid barriers. In the latter case, diffusion is not restricted. Proteolysis, browning reactions, acid diffusion, and flavor precursor formation mutually depend on these structures.

Composition of Unfermented Cocoa Seeds

The prominent secondary compounds are flavonoids and purines (theobromine and caffeine). The storage reserves include predominantly cocoa butter but also protein and starch. All these compounds are stored in the cotyledons. The composition of ripe seeds is shown in Table 1. Differences of cocoa butter, proteins, and pigments are found between Forastero and Criollo types.

Cocoa Butter

Fat, the main storage component, on average comprises 53–58% of the cotyledon dry weight. It contains about 95% triacylglycerols, 2% diacylglycerols, <1% monoacylglycerols, 1% polar lipids, and 1% free fatty acids (as percentages of lipids). Triacylglycerols consist of about 37% oleic (O), 32% stearic (S), 27% palmitic (P) and 2–5% linoleic (L) acids (as percentages of total fatty acids). Other saturated and monounsaturated fatty acids do not contribute more than about 2% of the fatty acids. These values are typical of fermented beans but do vary between samples, and between genetic and geographic origins. This is also true of the complex fatty acid composition in triacylglycerols. In cocoa butter, the disaturated triacylglycerols, mainly POS, SOS, and POP predominate by far over monosaturated triacylglycerols,

Table 1 Analyses of unfermented West African cocoa

Constituent	*Dried beans (%)*
Cotyledons	89.60
Shell	9.63
Embryonic axis	0.77
Fat	53.05
Water	3.65
Ash (total)	2.63
Nitrogen	
Total nitrogen	2.28
Protein nitrogen	1.50
Ammonia nitrogen	0.028
Amide nitrogen	0.188
Theobromine	1.71
Caffeine	0.085
Carbohydrates	
Fructose	(0.09)
Glucose	0.30 (0.07)
Sucrose	0 (2.48)
Starch	6.10
Pectins	2.25
Fiber	2.09
Cellulose	1.92
Pentosans	1.27
Mucilage and gums	0.38
Tannins (total phenolics)	7.54 (13.5)
Acids	
Acetic (free)	0.014
Oxalic	0.29

After Rohan (1963) with permission of the Food and Agriculture Organization of the United Nations, except data in parentheses for sugars in Bahia from Berbert (1979). Total phenolics from Swain (1954) as cited by Rohan (1963) assuming 55% fat; see Further Reading.

mainly POO and SOO. The melting temperature of cocoa butter at 31–34 °C is an essential characteristic. Differences of fat hardness are reflected in the iodine value of cocoa butter and the ratio of disaturated to monosaturated triacylglycerols (Table 2B, C). South American cocoas are the softest compared with cocoa from Asia and Africa. Storage fat accumulation during seed development undergoes characteristic changes in fatty acid composition (Table 2A). Climate, demonstrably the ambient temperature and stress due to heat or drought, affects biosynthesis and thus the final composition of triacylglycerols, and the melting and crystallization characteristics of cocoa butter. Whilst the amount of fat per seed depends on both, genetics and physiology of pod development, effects due to genetic differences or the process of fermentation on cocoa butter quality are not unequivocal. (*See* **Fatty Acids**: Properties; **Triglycerides**: Structures and Properties.)

Polyphenols

Flavonols are the major components. The order of quantities in Forastero types is (approximate weight percentages of total polyphenols): proanthocyanidins (58–65%), catechins (29–38%), then anthocyanins (1.7–4.0%). Specific compounds detected include: anthocyanins (3-α-L-arabinosidyl cyanidin and 3- β-D-galactosidyl cyanidin); catechins ((−)-epicatechin, (+)-catechin, (+)-gallocatechin, (−)-epigallocatechin). Additionally, several (up to eight) proanthocyanidins (di- and oligomeric) and proanthocyanidinglycosides (predominantly based on (−)epicatechin) have been isolated and identified. Three flavonols have been found as minor constituents: quercitin, quercitin-3-glucoside, and quercitin-3-galactoside. Up to 17 phenolic acids and esters have also been reported. The total amount of seven of them comprise not more than 23 p.p.m. of the seed dry weight (phloroglucinol, protocatechuic acid, vanillic acid, *o*-hydroxyphenylacetic acid, *p*-coumaric acid, caffeic acid, ferulic acid). (*See* **Tannins and Polyphenols**.)

Anthocyanins and their aglyca are lacking in Criollo-type seeds but provide the characteristic color of Forastero seeds. They are also taken as an indicator of the 'degree of fermentation.' Colorless proanthocyanidins and catechins are strongly astringent and are effective tannins. Their enzymatic oxidation and polymerization cause browning reactions after cell wounding and reduction of astringent taste. Cocoa polyphenols are strong antioxidants which are increasingly considered as antioxidative protectors in food. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments.)

Table 2 Data reflecting differences of cocoa butter hardness

A *Ratio of saturated[c] to unsaturated[d] fatty acids in triacylglycerols in developing cocoa seeds*

Days after pollination	*Mass ratio*
105	0.98
130	1.06
170	1.63

B *Ratio of disaturated[e] to monosaturated[f] triacylglycerols in cocoa butter from different origins*

Provenance	*Mass ratio*
South America	9.43
Africa	12.32
Asia	19.45

C *Iodine values in cocoa butter from different origins*

Provenance	*Iodine value*
Ghana[b]	36.5 ± 0.78
Ivory Coast[a]	35.5 ± 1.15
Brazil[a]	37.5 ± 1.31
Ecuador[a]	36.7 ± 0.04
Malaysia[a]	34.7 ± 0.34

A Data derived from Griffiths G, Harwood JL (1991) The regulation of triacylglycerol biosynthesis in cocoa (*Theobroma cacao*) L. *Planta* 184: 279–284; **B** and **C**[a]derived from Chaiseri S, Dimick PS (1989) Lipid and hardness characteristics of cocoa butter from different geographic regions. *Journal of the American Oil Chemists Society* 66: 1771–1776; **C**[b] from *Cocoa Bean Test 1961/62* (1962) Hamburg: Gordian. In **A** cocoa butter is from unfermented seeds; in **B** and **C** it is from fermented seeds.[c] Palmitic acid (P) plus stearic acid (S); [d]oleic acid (O) plus linoleic acid (L); Arachidic acid (A). [e]POP, POS, SOS, PLP, PLS, SLS. [f]POO, SOO, SOA.

Anthocyanins and their aglyca are lacking in Criollo-type seeds but provide the characteristic color of Forastero seeds. They are also taken as an indicator of the 'degree of fermentation.' Colorless proanthocyanidins and catechins are strongly astringent and are effective tannins. Their enzymatic oxidation and polymerization cause browning reactions after cell wounding and reduction of astringent taste. Cocoa polyphenols are strong antioxidants which are increasingly considered as antioxidative protectors in food. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments.)

Alkaloids

The composition of methylxanthines in cocoa seeds is genetically variable but characteristic of cocoa. Theobromine (1–2%, dry weight) and caffeine (0–2%, dry weight), together with traces of theophylline and 7-methylxanthine, are found. They are not metabolized during fermentation. Theobromine plays a role in the bitter taste of cocoa. Several phenolic amines and alkaloids derived from tyrosine and tryptophan have been detected at very low levels (3–40 $\mu g\,g^{-1}$) in both unroasted and roasted cocoas. (*See* **Alkaloids**: Toxicology.)

Sugars and Acids

Besides a total of around 12% (dry weight) of polysaccharides (Table 1), the free sugar content in cocoa seeds is not more than 2–4% (dry weight). Sucrose is the major component (about 90% of total sugars), followed by fructose plus glucose (about 6%). In addition, galactose, raffinose, melobiose, sorbose, mannotriose, xylose, arabinose, mannitol, and inositol have been detected by several investigators in

minor quantities. (*See* **Carbohydrates**: Classification and Properties.)

Volatile acids are not found in fresh cocoa beans, but nonvolatile acids are found in low concentrations. Phosphoric, malic plus tartaric acids were found to make up 0.32% (dry weight), oxalic acid 0.35%, and citric acid 0.73%. These acids in unfermented seeds are not metabolized during processing nor do they contribute to cocoa quality.

Proteins and Amino Acids

Correct analyses of proteins strictly depend on suitable protection against tanning, especially against protein–quinone reactions. The protein content in the cotyledons of ripe seeds varies between 10 and 16% (dry weight). Criollo cocoa seeds have been reported to contain less protein than Forastero seeds. The proteins of Forastero seeds reveal about 16 peptide bands in sodium dodecyl sulfate–polyacryl amide gel electrophoresis (SDS-PAGE): four of them predominate in amounts. They are subunits of vacuolar storage proteins. One of these polypeptides is an albumin of 21 kDa, characterized as a Kunitz protease inhibitor (52% (w/w) of total protein). Three – the 47-kDa, the 31-kDa, and the 16-kDa polypeptides – make up about 43% (w/w) of total protein. They are subunits of the vicilin-type (7*S*) globulin, a glycoprotein, each of them consisting of multiple pI-forms. The 31-kDa and the 16-kDa subunits are posttranslational processing products of the 47-kDa subunit. The native (7*S*) globulin is a trimer (139-kDa) composed of 47-kDa and (31-kDa7 + 16-kDa) subunits. It is almost unsoluble at pH < 5.0. The amino acid sequences of the major polypeptides (47, 31 and 21 kDa) and their precursors have been derived from the nucleotide sequences of the corresponding cDNAs and genomic clones from immature cocoa beans. The sequence of cocoa (7*S*) globulin is unique but similar to that of cotton seeds and is different from those of other plants. Cocoa seeds do not contain legumin-like (11*S*) globulin.

The fresh cotyledons from ripe seeds contain about 5 mg g^{-1} (dry weight) of free amino acids. Acidic amino acids dominate. (*See* **Amino Acids**: Properties and Occurrence.)

Enzymes

Various enzymes have been identified in cotyledons of fresh, ungerminated cocoa seeds which in part are active during postmortem fermentation or drying, including glycosidases, invertase, proteases, polyphenoloxidase, and peroxidase. Several proteases have been established:

1. An aspartic endoprotease: its activity optimum is at pH 3.5. It consists of two polypeptides, which are obviously derived from a 42-kDa polypeptide by self-digestion. It splits native (7*S*) globulin, and its specific activity is 20–50-fold higher in recalcitrant cocoa seeds compared with that in orthodox seeds of several higher plants.
2. A cysteine endoprotease: its specific activity is considerably lower than that of the aspartic endoprotease. In contrast to an additional cysteine endoprotease which is induced during germination, it does not split the native (7*S*) globulin.
3. A seryl exopeptidase: this carboxypeptidase reveals optimum pH around pH 5.8 and substrate specificity, preferentially liberating hydrophobic amino acids.
4. A leucine-*p*-nitroanilide cleaving seryl exopeptidase displaying maximal activity at pH 6.8. (*See* **Enzymes**: Functions and Characteristics.)

Glycosidases are responsible for hydrolysis of flavonoid glycosides. Their highest activity in cocoa beans was found between pH 3.5 and 4.5. Polyphenoloxidase (*o*-diphenol: O_2 oxidoreductase) in unfermented seeds displays a high activity which is strongly reduced after fermentation. Under a wide range of pH optima reported by several authors, pH 6.0–6.4 is the most probable range of highest activity. In wounded plant tissues it effectively oxidizes polyphenols, and the resulting quinones react spontaneously to give brown polymerization products.

Composition of Pulp

The mucilaginous, sweet pulp firmly adhering to each of the individual seeds is formed during pod development from an endocarp meristem. Pulp from ripe seeds contains high but variable amounts of sugars (Table 3) which impair seed germination. Citric acid is responsible for the low pH value (pH 3.5–3.8). The pulp variably binds a lot of water due to pectins. Additionally hemicelluloses, cellulose, and proteins are constituents which are insoluble in water and which resist hydrolysis in strong mineral acids.

Besides pectins in the pulp, the seed testa also contains mucilage: both swell considerably after removal of acid.

Changes During Fermentation

Changes in the seeds (cf. Figure 3 in **Cocoa**: Production, Products and Uses) start in the organized subcellular structure of the living seed. They proceed to postmortem reactions which are caused by heat and diffusion of acetic acid from the fermenting pulp

Table 3 Analysis of cocoa pulp mucilage

Fraction	*Unit*	*Quantity*
Sugars		
Hexoses	mg seed^{-1}	104 (51)[b]
Others (including sucrose)	mg seed^{-1}	61 (33)[b]
Polysaccharides		
Pectins	mg seed^{-1}	12–13[a]
Water-insoluble		
Polysaccharides (hemicelluloses and cellulose)	mg seed^{-1}	23–31[a]
Nitrogen		
In the pectin fraction	%w/w	0.35
In the fraction of water-insoluble polysaccharides	%w/w	2.28
Pulp fresh weight	mg seed^{-1}	900–1200[a] (800–1000)[b]
Pulp dry weight	mg seed^{-1}	200–240[a] (120–180)[b]

Unpublished data from Dziggel M and Biehl B (1986).
Pulp of seeds from Amelonado-type and in some cases Malaysian hybrids from [a]ripe pods and in some cases from [b]overripe pods were analyzed giving one example of strongly variable quantities, depending on pod physiology and ripeness. Compare Table 1 in Chapter Cocoa – Production, Products and Uses and see Biehl B, Meyer B, Crone G, and Pollmenn L (1989) *Journal of the Science of Food and Agriculture* 48: 189–208.

(Figure 2). These changes are controlled by the effective concentration of acetic acid and the nib pH value, which drops to pH 5.0–4.0.

Once the plasma membranes and tonoplasts are destroyed, proteins, polyphenols, and other compounds are released from their compartments and come into contact with enzymes. The pH value established in the cells controls the postmortem activity of acidic hydrolases (proteases, glycosidases, invertase) under anaerobic conditions.

As a result of cell destruction, variable amounts of soluble constituents, flavonoids, amino acids, purines, and sugars are lost from the seeds due to exudation. A loss of 10–40% of soluble substances modifies the original composition of cocoa seeds. Reported changes in the content of fat are in accordance with this loss of soluble substances.

Sucrose disappears to a large extent from the seeds during fermentation, due to invertase activity. However, the amount of fructose plus glucose in the fermented beans is not more than 50% of the original amount of sucrose. Since their proportions are variable, they obviously undergo further reactions.

The flavonoid glycosides are hydrolyzed by acidic glycosidases. The violet aglycon cyanidin subsequently undergoes bleaching by nonenzymatic transformation to a colorless pseudobase. Leucocyanidins and proanthocyanidins are transferred nonenzymatically to more complex forms. Although no reactions of the catechins have been established, the loss of 50–75% suggests nonenzymatic oligomerization. These strongly tanning compounds are released from polyphenol storage cells but they do not prevent proteolysis in the absence of oxygen.

Proteolysis in the beans is crucial for the formation of cocoa-specific aroma precursors. It starts after uptake of acetic acid. The extent and type of proteolysis, and thus the resulting aroma potential in fermented beans, strongly depend on the pH value established in the beans during this stage of proteolysis. The significant details have been found by comparing changes during fermentation with aseptic incubations of seeds, acetone dry powder, and of purified cocoa seed proteins and proteases (Figure 3).

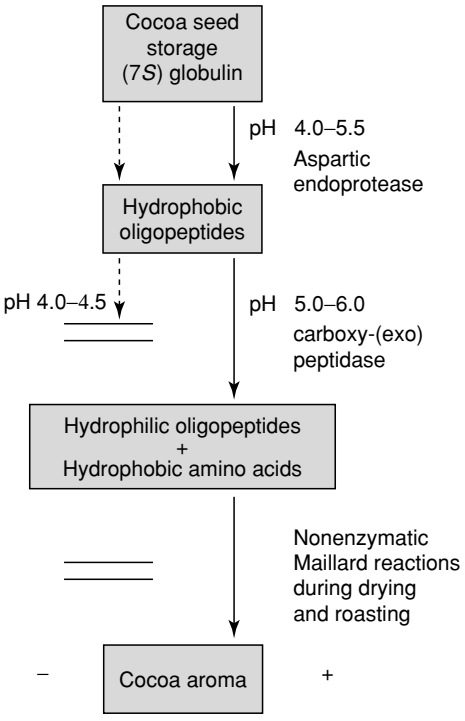

Figure 3 Proteolytic formation of cocoa-specific flavor precursors during fermentation and their transformation during drying and roasting.

All cocoa seed proteins, especially albumin and (7*S*) globulin, are degraded by cocoa aspartic endoprotease after strong acidification (pH ≤ 4.0). However, at higher pH values (pH 5.0–5.5), the (7*S*) globulin is the only protein attacked by the aspartic endoprotease. Cocoa-specific aroma precursors comprise characteristic patterns of hydrophilic oligopeptides and free hydrophobic amino acids resulting from suitable degradation exclusively of cocoa (7*S*) globulin. Neither cocoa albumin nor (7*S*) globulins from other plants (so far as these have been investigated) give the characteristics degradation pattern and cocoa-specific aroma precursors depends on the substrate specificity and the specific cooperation of both the aspartic endoprotease and the carboxyexopeptidase. In the pH range 4.0–5.5, the aspartic endoprotease cleaves (7*S*) globulin at the site of hydrophobic amino acids in its chain, giving a considerable number of hydrophobic oligopeptides. At pH ≥ 5.0, carboxyexopeptidase acts near to its pH optimum and specifically removes hydrophobic amino acids from the hydrophobic oligopeptides. In this way they are further converted to hydrophilic oligopeptides and free hydrophobic amino acids, especially leucine, alanine, phenylalanine, and tyrosine. After incubation or fermentation, their proportions are significantly higher than in the proteins (Table 4).

Consequently, during strongly acidic fermentation (around pH 4.0), almost no hydrophilic oligopeptides and few free amino acids are generated and cocoa aroma potential in fermented beans is low in contrast to moderate nib acidification during fermentation. At pH ≥ 5.0 both enzymes are sufficiently active.

During fermentation, about 2 days after seed death, the rate of proteolysis decreases significantly. This decrease is not due to tanning since it is also found in the absence of polyphenols during acetone-dry powder autolysis.

The liberated amino acids and peptides in the beans do not seem to be metabolized during fermentation unless the seed takes up oxygen during the final stages of fermentation. They may now be exposed to direct aerobic microbial attack, especially during subsequent prolonged (over)fermentation.

Changes in the pulp are considered below in Microbiology of Pulp Fermentation.

Changes During Drying

Reactions during drying are an oxidative continuation of fermentation. They are important for quality. Only about 10–20% of catechins and proanthocyanidins present in unfermented beans are found in raw cocoa after fermentation and drying. Polyphenols are oxidized by molecular oxygen: the resulting quinones undergo polymerization to give brown products. Oxidation is brought about by residual activities of the polyphenoloxidase. However, as it is optimum at pH 6, in acidic beans (pH 4.0–5.0) browning is impaired during drying. At pH > 7.0 polyphenols are oxidized nonenzymatically. Thus, overfermented cocoa easily turns brown, giving an incorrect appearance in the cut test of perfectly fermented cocoa beans. Since

Table 4 Amino acids in cocoa seeds

	Free amino acids in cotyledons of cocoa seeds			*Protein-bound amino acids in unfermented seeds*	
Amino acid fraction	*In cotyledons after standard fermentation*[a]	*In cotyledons after aseptic seed incubation*[b]	In vitro *after enzymatic digest of purified cocoa (7S) globulin*[c]	*Total cotyledon protein*[d]	*Enriched (7S) globulin*[e]
Acidic[f]	15.9	20.5	21.8	30.4	26.0
Hydrophobic[f]	58.3	58.6	56.7	34.1	41.3
Portion of Leu, Ala, and Phe under hydrophobic[f]	40.9	41.3	36.0	18.8	23.6
Others[f]	25.8	21.0	21.5	35.5	32.7

[a]From 5 days heap fermentation in Ghana (Amelonado).
[b]Unfermented ripe beans from Malaysia (hybrids) underwent aseptic incubation at pH 4.5 in acetic acid containing media for 20 h at 40 °C and 40 h at 50 °C under nitrogen.
[c]Purified cocoa (7*S*) globulin was incubated in dilute acetic acid at pH 5.2 at 50 °C for 16 h in the presence of partially purified cocoa aspartic endoprotease. Resulting peptide mixtures were incubated at pH 5.8 at 45 °C for 16 h in the presence of partially purified cocoa carboxypeptidase.
[d]Acetone dry powder from unfermented Malaysian seeds were hydrolyzed in hydrochloric acid.
[e]Proteins in pH 8.9 buffer extract of acetone dry powder from unfermented Malaysian seeds were separated on a Sephadex G150 column. The fraction containing the (7*S*) globulin was hydrolyzed in hydrochloric acid.
[f]Values are mol% of total estimated free amino acids or protein-bound amino acids respectively.
Data from Kirchhoff PM, Biehl B and Crone G (1989) Peculiarity of the accumulation of free amino acids during cocoa fermentation. *Food Chemistry* 31: 295–311; and Voigt J, Heinrichs H, Voigt G and Biehl B (1994) Cocoa-specific aroma precursors are generated by proteolytic digestion of the vicilin-like globulin of cocoa seeds. *Food Chemistry* 50: 177–184. High-performance liquid chromatography was used for estimation of amino acids in polyphenol-free extracts of cotyledons, lyophilized after treatments.[a–d]

enzymatic browning requires high oxygen concentrations, browning does not take place before drying. A consequence of polyphenol oxidation and quinone polymerization is the reduction of the astringent flavor, which is due to the tanning property of catechins and procyanidins, especially the oligomeric forms. Oxidation and polymerization transfer these mono- and oligomeric flavonoids to nonastringent brown polymers. (*See* **Browning**: Enzymatic – Biochemical Aspects.)

Thus, the extent of these polymerizations is controlled by the nib pH, the permeability of the shell to oxygen, and the destruction of subcellular compartmentation. Roughly, the internal color of raw cocoa beans is related to astringency. Slaty beans reveal a strongly astringent taste while entirely brown beans and, especially, overfermented beans give an insipid taste.

The extent of the browning reaction also affects the residual seed proteins. Free amino- and sulfhydryl groups of proteins readily react with quinones and thus participate in the production of brown polymers. Amino acids and oligopeptides are less reactive. However, they participate in the formation of Maillard products during drying. The spontaneous reaction of proteins and quinones is also responsible for the toxic effects of quinones on bacteria and fungi. Therefore, overfermentation is suppressed during browning reactions in the course of drying. However, the resulting polymers are not toxic and would not impair microbial growth on brown raw cocoa beans during storage and on fully brown beans in particular.

Microbiology of Pulp Fermentation

The events taking place during pulp fermentation correspond to a succession of microorganisms metabolizing the pulp. The wide range of organisms in the wild inoculum is similar in different cocoa-growing countries. When starting fermentation, the low pH value and the high sugar content of the pulp allow anaerobic fermentation by yeasts and lactic acid bacteria. Ethanol is produced, and after exhaustion of pulp sugars, acetic acid bacteria predominate. Acetic acid and lactic acid are the most significant metabolites penetrating the beans. Several authors have isolated more than 30 species from 13 genera of yeasts. The size of the population of lactic acid bacteria is subordinate, and is further repressed when acetic acid bacteria subsequently become active after exhaustion of pulp sugars. Several homofermentative and heterofermentative lactobacilli have been isolated. (*See* **Lactic Acid Bacteria**.)

A number of species of the genus *Acetobacter* which have been found in fermented cocoa not only oxidize ethanol to form acetic acid but also oxidize this acid, causing a decrease in acetic acid during late stages of fermentation. Additionally, *Gluconobacter* spp. were found.

The sizes of these populations correlate with the amount of products formed (ethanol, lactic acid, acetic acid). However, there is no information available about the more detailed competition and significance of all these individual species.

More recent experiments have been described studying the effect during usual fermentation of pure cultures of selected microorganisms either in addition to or in the absence of wild inoculum. In total, from different observations with respect to changes in the pulp, temperature increase, acid production, and final bean quality, it can be deduced that a limited number of species, yeasts, lactic acid, and acetic acid bacteria, which are adapted to the conditions in the cocoa pulp, can completely substitute for the numerous organisms in fermenting cocoa.

Besides pulp sugar and acid metabolism, pectinolytic activity is of interest: this assists in drainage of pulp in the early stage of fermentation. *Kluyveromyces marxianus* was found to have highest pectolytic (endopolygalacturonase, EC 3.2.1.15) activity but additionally *Saccharomyces chevalieri, Torulopsis candida, Candida norvegensis*, and *Kluyveromyces fragilis* were reported to reveal significant endopolygalacturonase activity.

The last stage of fermentation is characterized by the development and dominance of aerophilic bacteria, depending on aeration. Up to 14 species of the genus *Bacillus* have been isolated which, in part, are facultative anaerobes. Further species of other genera have been reported. There is no direct evidence as to how they contribute to cocoa fermentation in detail. Most probably they participate by increasing the shell permeability and later on in overfermentation. *Bacillus stearothermophilus, B. subtilis, B. circulans, B. licheniformis* and *Streptococcus thermophilus* have been found as dominating organisms. Species like *B. megaterium, B. subtilis, B. coagulans, B. cereus, B. polymyxa* or *Enterobacter aerogenes* have been isolated, which metabolize carbonic acids, proteins, and amino acids, producing low-molecular-weight fatty acids. Propionic, butyric, and isovaleric acids are increased to 0.1–1.0% of dry weight each in overfermented cocoa beans, compared to 0–0.02% in controls. These findings strongly indicate bacterial degradation of amino acids (and flavor precursors) during overfermentation. (*See* ***Bacillus***: Occurrence.)

A large number of different filamentous fungi develop on the seeds in well-aerated niches, especially when less acidic and strongly aerated during the later stages after a steep pH increase. They may develop

further during drying when the humidity is high. Xerophilic fungi like *Aspergillus fumigatus*, as well as *A. glaucus, Penicillium* spp., *Mucor* sp., *Paecilomyces* sp. and *Geotrichum* sp., which have been found to develop during late stages of fermentation, have also been reported to occur on stored raw cocoa beans.

Flavor Compounds and Flavor Precursors Formed during Fermentation and Drying

Some characteristic flavor compounds in cocoa products are original constituents of the unfermented seeds but they are changed in amount during fermentation and drying: theobromine tastes bitter. Astringency is due to tanning flavonoids, and their oligomeres. Further oligomerization during fermentation increases. Oxidation and polymerization during drying decrease their taste intensity. Acidic taste is due to lactic and acetic acid, which diffuse into the beans during fermentation and persist in variable amounts after drying and roasting. When present in high concentrations, acetic acid may mask the impression of cocoa aroma.

Cocoa flavor precursors comprise compounds in fermented beans which react during drying and manufacturing, especially roasting, to give flavor. Specific cocoa aroma precursors are not original components of the seed but are formed during fermentation and drying: reducing sugars, free amino acids, and oligopeptides result from enzymatic hydrolysis. Their formation and significance is described above, in the section on changes during fermentation. They undergo Maillard reactions during drying and roasting in order to generate cocoa aroma.

After *in vitro* proteolysis under suitable conditions of cocoa (7*S*) globulin in the presence of cocoa seed proteases, the methanol extract of aroma precursors gives a typical cocoa aroma when roasted in the presence of reducing sugars. Synthetic mixtures of amino acids, adapted to the composition of free amino acids in fermented cocoa beans in the presence of reducing sugars but in the absence of the fraction of hydrophilic oligopeptides, do not generate cocoa aroma on roasting.

Intermediate Amadori products and several pyrazines have been found in fermented and dried beans before roasting; pyrazines (mono-, di-, tri-, and tetramethyl-pyrazines) in part are already generated during fermentation. Polyphenols (flavonoids), although important for the final astringent taste, do not contribute to the cocoa aroma. Some authors found that they enhance cocoa aroma.

In addition to cocoa aroma precursors, oligopeptides and theobromine are flavor precursors: during roasting, N-terminal amino acids of several hydrophobic peptides produce diketopiperazines which give the characteristic cocoa bitter principle in a 1:2 complex with theobromine on roasting.

Chemistry of Manufacturing Processes

Cocoa flavor is chemically very complex because it arises from fermentation and roasting and is influenced by further processing steps. The flavor of plain chocolate is modified but chemically very similar to cocoa. The flavor sensations of cocoa products and chocolates are based on a great number of volatile aroma compounds and their interaction with nonvolatile constituents which affect taste and tactile characteristics.

Cocoa Flavor Development

Flavor of unroasted cocoa Raw cocoa has an acid taste and a flat but characteristic aroma. As a result of the fermentation process it contains aroma precursors, mainly free amino acids, monosaccharides, and their first reaction products (Amadori compounds), peptides, monomeric flavonoids, and methylxanthines. The first aroma compounds were developed in an early Maillard reaction during the drying of the fermented seeds or absorbed from the fruit pulp. The major volatiles in the concentration range 0.1–1 mg kg^{-1} are aldehydes (produced via Strecker degradation of amino acids), alcohols, acetates, and acids, which derive from valine, leucine, isoleucine, and phenylalanine (Tables 4 and 5). Tetramethylpyrazine is formed by microorganisms. Special flavor-grade cocoas, mainly harvested in Venezuela, Trinidad, and Ecuador (Arriba), reveal a flowery and tea-like aroma and contain significant concentrations (0.5–2 mg kg^{-1}) of linalool and further terpenoids which contribute to this valuable note. On the other hand, basic cocoas from West Africa, Malaysia, or Brazil (Bahia) carry a fairly strong inherent flavor and have very low concentrations of linalool. The sharp acidity of raw cocoas arises from acetic acid, the astringency and bitterness from soluble polyphenols (mainly epicatechin), tannins, theobromine, and caffeine. A smoky flavor, which is a flavor defect of some cocoas, is produced by the absorption of volatile phenols from the smoke of the firewood used for drying.

Roasting and Aroma Formation Aroma is formed from its precursors in raw cocoa beans. In contrast to amino acids, glucose and fructose are consumed to a large extent during roasting (up to 90%). While the composition of triacylglycerols in cocoa butter remains unchanged during roasting and further

Table 5 Typical aroma compounds in unroasted cocoas and some of their precursors

Acids		**Carbonyls**
Acetic		Acetone
2-Methylpropanoic	Val	2,3-Butandione
3-Methylbutanoic	Leu	Acetophenone
2-Methylbutanoic	Ile	Dihydrohydroxymaltol
Benzoic	Phe	
Phenylacetic	Phe	
Aldehydes		**Heterocyclic compounds**
2-Methylpropanal	Val	Tetramethylpyrazine
3-Methylbutanal	Leu	2-Acetylpyrrole
2-Methylbutanal	Ile	
Benzaldehyde	Phe	
Phenylacetaldehyde	Phe	
Alcohols		**Terpenes**
2-Methylpropanol	Val	Linalool
3-Methylbutanol	Leu	Linalool oxides
2-Methylbutanol	Ile	
Benzyl alcohol	Phe	
2-Phenylethanol	Phe	
Acetates		
2-Methylpropyl acetate	Val	
3-Methylbutyl acetate	Leu	
2-Methylbutyl acetate	Ile	
Benzyl acetate	Phe	
2-Phenethyl acetate	Phe	

Val, valine; Leu, leucine; Ile, isoleucine; Phe, phenylalanine.

processing steps, the entire cocoa aroma is formed by careful roasting of fermented and dried seeds at temperatures between 120 and 140 °C. Using modern technology, cocoas can be roasted at different particle sizes, such as whole beans, as nibs, which are coarsely ground and broken beans, or as liquid cocoa mass, which is produced by a fine grinding of cocoa and liquefying within its own fat. The roasting of smaller particles has the advantages of better controlled roasting, partial exhaustion of abundant acetic acid, and reduced roasting time (beans about 30 min, nibs 12 min, mass 2 min).

About 600 compounds have been identified in cocoa aroma. In common with all roasting aromas there is a large number of carbonyl and heterocyclic compounds. The main carbonyl groups are acids and esters (each about 50 compounds), alcohols, aldehydes, and ketones (each between 30 and 40 derivatives). The main heterocyclic compounds are pyrazines and chinoxalines (together about 80 compounds), furans, pyrones, and lactones (together 40 compounds), diketopiperazines, phenols, pyrroles, and oxazoles (each group about 10 compounds). Derivatives of phenylalanine are prominent volatiles after roasting. As a result of the Maillard reaction, pyrazines, pyrroles, phenylacetaldehyde, phenylalk-2-enals, pyrones, furanones, and furans increase during cocoa roasting, whereas alcohols, esters, and acids remain more or less unchanged.

Simple aldehydes arise by Strecker degradation of free amino acids. Their immediate flavor characteristics are not very striking, but they act as most important reactants. Some phenylalk-2-enals are generated via the aldol condensation, for example, which carry a typical flowery odor fairly reminiscent of chocolate or cocoa. Many pyrazines with different substitutes and chinoxalines with desirable flavor properties contribute to the roasted note. It was found that methylpyrazines are formed during roasting according to a specific rate depending on substitution. Some pyrones and furanones, as dihydrohydroxymaltol, hydroxymaltol, and furaneol, arise from degradation of monosaccharides and reach remarkable concentrations in roasted cocoas. Characteristic of cocoa aroma are abundant phenyl derivatives which arise from phenylalanine and contribute to the intensive sweet aromatic note. The typical intense bitterness of roasted cocoa is to a great extent induced by adducts of such diketopiperazines with theobromine and caffeine. In the course of roasting, monosaccharides react to the extent of about 60–90%, while free amino acids only react to about 20%. Therefore, in several roasters the unroasted nibs or masses are treated with aqueous sugar solutions to increase the concentration of the carbohydrate precursors and to improve the aroma yield. In addition, during the moist thermal treatment of cocoa, the level of Amadori compounds may be raised and the aroma may be intensified. (*See* **Caffeine**.)

Technologically Influenced Flavor Changes Roasted cocoa nibs and mass are important semimanufactured goods and are used for the production of cocoa powder, cocoa butter, or chocolate mass. The cocoa flavor is modified and improved by the subsequent processing, mainly by alkalization of nibs, degassing of cocoa mass, or conching of chocolate mass.

Conching of chocolate The conversion of the bitter cocoa flavor to the finer chocolate flavor occurs in the conche. There, a mixture of cocoa mass, sugar, cocoa butter, and milk powder (for milk chocolate production), fine ground in rolling mills, is stirred vigorously and treated for about 8–24 h at 50–60 °C. In addition to the aroma improvement, this process gives the desired consistency to the chocolate by covering all the solid particles with fat. Aroma and consistency are important for the flavor of the finished product. There are two main functions of the conche: degassing and mechanical–thermal treatment of the chocolate mass. Low-boiling volatiles are removed, together with steam. Acetic, propionic, isobutyric, and isovaleric acids are reduced by about 30%. About 20–40% of the unspecified aroma compounds,

such as abundant aldehydes and alcohols, as well as of aroma-damaging substances like volatile phenols, are also removed. As a result the typical and slightly volatile aroma becomes more pronounced.

Sorption phenomena of aroma materials at particle surfaces of the heterogeneous chocolate mass of fat and fat-free portions have considerable influence on the flavor.

Thin-layer treatment of cocoa mass During modern chocolate processing the cocoa liquor is degassed before being put into the chocolate recipe. By using special equipment, degassing can be performed more effectively than in the conche. Thin-layer vaporizers are generally used, which may shorten the remaining conching time considerably. The process conditions have to be coordinated with the quality of the cocoa liquor, so that the valuable aroma is not removed together with the more volatile unknown materials, thereby flattening the taste of the product. A reduction of 10–30% of acids and low-boiling compounds seems to be optimal.

Alkalization of cocoa Alkalization ('Dutching') is mainly used to intensify the color, to modify the flavor, and to improve the dispersability of cocoa powder in water or milk. In this process, beans, nibs, or mass are treated with solutions or suspensions of alkali, usually potassium or sodium carbonate, sometimes hydroxides or ammonia. The permitted maximum is generally 2.5–3 parts of potassium carbonate (or equivalent alkali) per 100 parts of cocoa. The pH of a natural cocoa is about 5.5, whereas the pH of commercially available Dutch process cocoas will fall between 6.8 and 8.0.

Color and flavor changes during alkalization are mainly based on modifications of polyphenolic substances. Dutch processing results in a milder and altered flavor, and a darker color. Under alkaline conditions, soluble monomeric flavonoids (mainly epicatechin) are oxidized and condensed to give polymer tannins. Red and brown pigments are developed and the astringency is reduced. Volatile acids are partially neutralized.

Cocoa Aroma Analysis

The cocoa aroma volatiles only occur in milli- or microgram per kilogram quantities. They are dissolved in the fat phase or physically bound to a heterogeneous matrix of constituents of differing polarity. Isolation and enrichment techniques are necessary to obtain the flavor volatiles in sufficiently high concentrations. Distillation methods, headspace techniques, and extractions, or combinations of these techniques are used. The separation of the volatiles is performed by means of capillary gas chromatography (GC) or high-performance liquid chromatography (HPLC). (*See* **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography.)

Isolation and Analytical Techniques

Extraction Extraction techniques have several disadvantages (time-consuming and the possibility of introducing artifacts) but have the advantage of quantitatively isolating volatiles independent of their boiling points. Usually, cocoa is defatted with petroleum ether and the residue is extracted with a polar solvent such as water, ethanol, acetonitrile, ethyl ether, or subcritical carbon dioxide. By means of extraction, specific high-boiling compounds have been isolated from cocoa: diketopiperazines, pyrones, amines, amides, aromatic acids and phenols (these compounds are mainly for subsequent GC separation), and theobromine, caffeine, and monomeric flavonoids (for HPLC measurement).

Distillation Most of the known aroma volatiles have been isolated from cocoa by means of distillation (steam distillation, vacuum condensation, simultaneous steam distillation–extraction (SDE), head-space enrichment). Nonvolatile impurities can thus be avoided and a defatting step is not necessary, but high-boiling compounds are partially lost and one has to be aware of possible reactions during the heat treatment.

The aqueous distillates are usually extracted with small portions of ethyl ether for performing GC or may be injected directly into an HPLC system.

Table 6 shows typical volatiles of roasted cocoa. The flavor chromatograms of different cocoas are rather similar, but the quantitative distribution of the single constituents and their total amounts vary distinctly with origin, fermentation, drying, roasting, degassing, alkalization, or conching. Distillates of chocolate contain similar compounds in lower concentration and altered distribution.

Analysis of indicative compounds and key odorants

Objective analytical data are necessary to relieve the industrial quality control of time-consuming sensory tests. For that purpose suitable indicators are needed to evaluate the flavor. Several examples of indicative compounds in cocoa aroma follow, together with methods suitable for their estimation (GC; HPLC; ultraviolet (UV) detector; electrochemical detector (ELCD)): Fermentation degree – epicatechin (HPLC–ELCD), tetramethylpyrazine (GC or HPLC–UV); flavor-grade cocoas linalool (GC); roasting intensity ratios of 2,5-dimethyl- to tetramethylpyrazine

Table 6 Selected aroma compounds identified in extracts of roasted cocoas

Compound	*Concentration range*
3-Methylbutanal	IV
2,3-Butandione	III
3-Methyl-butan-1-oe	IV
Acetone	II
2-Methylpyrazine	III
2,5-Dimethylpyrazine	III
2,6-Dimethylpyrazine	II
2-Ethylpyrazine	I
2,3-Dimethylpyrazine	II
Methyltrithiomethane	I
2-Ethyl-5-methylpyrazine	II
2-Ethyl-3-methylpyrazine	II
Acetic acid	V
Trimethylpyrazine	III
Linalool oxide (*cis*-furanoid)	I
2,5-Dimethyl-3-ethylpyrazine	II
2-Furfural	II
Linalool oxide (*trans*-furanoid)	I
Tetramethylpyrazine	IV
Pyrrole	I
Benzyaldehyde	III
Linalool	II
2,5-Diethyl-3-methylpyrazine	I
5-Methyl-2-furfural	I
3-Methyl butanoic acid	V
Phenylacetaldehyde	IV
Acetophenone	III
Furfuryl alcohol	I
1-Phenylethanol	I
2-Phenethyl acetate	II
2-Hydroxy-3-methyl-2-cyclo penten-1-one (cyclotene)	II
Guajacol	I
Benzylalcohol	II
2-Phenylethanol	III
2-Phenylbut-2-enal	III
2-Phenethylamine	II
2-Acetylpyrrole	III
4-Methyl-2-phenylpent-2-enal	II
Phenol	I
3-Hydroxy-2-methyl-4-pyron (maltol)	II
4-Hydroxy-2,5-dimethyl-3-(2*H*)-furanone (furanol)	III
5-Methyl-2-phenylhex-2-enal	III
2-Pyrrolecarbaldehyde	II
5-Methyl-2-pyrrolecarbaldehyde	II
2,3-Dihydro-3,5-dihydroxy-6-methyl-4-pyrone (dihydrohydroxymaltol)	IV
3,5-Dihydroxy-6-methyl-4-pyrone (hydroxymaltol)	II
5-(2-Hydroxyethyl)-4 -methylthiazole	III
1,2-Benzenediol	III
Benzoic acid	II
2-Phenylacetic acid	IV
2-Phenylacetamide	IV

Usual concentration range (mg kg^{-1}): I, < 0·1; II, 0·1–0·5, III, 0·5–2; IV, 2–10; V, > 10. Succession of gas chromatographic retention times on a DB-wax capillary column.

or trimethyl- to tetramethylpyrazine (GC or HPLC–UV), ratio of 5-methyl-2-phenyl-hex-2-enal to 2-phenylethanol (GC), dihydrohydroxymaltol (GC or HPLC–ELCD–UV); intensity of degassing of cocoa masses – 3-methylbutanal (head-space GC).

Aroma extract dilution analyses (AEDA) is a potent tool to screen the most odor-active compounds in foods. The application of AEDA on the cocoa volatiles revealed about 40 main odorants in cocoa mass.

Sensory Evaluation

Sensory evaluation is extremely important for cocoa process control and for the quality assurance of cocoa and chocolate products. Based on a tasting panel, which consists of trained and selected members, different sensory tests (flavor profile, difference, and scoring tests) are used and the results are statistically evaluated. (*See* **Sensory Evaluation**: Taste.)

The flavor profile test This test is essential to obtain an assessment of the development of taste and flavor within the mouth as a result of individual sensory perceptions. The varying intensities of the specific characteristics are observed and evaluated. An evaluation plan is used which considers the flavor and texture characteristics and is completed by a numeric scale. Some characteristics of cocoa mass are raw, acid, flavorful, typical cocoa bitterness, burnt/bitter, astringent, off-taste (cooked, moldy, smoky, ham, chemical, etc.). In addition, descriptors for plain chocolate are sweetness, aromatic, cocoa intensity, harmonic, persistent, caramel, vanilla, off-taste (cardboard, metallic, rancid, etc.). Intensities range from 0 to 4: absent, barely perceptible, weak, medium, strong. Texture descriptors are sticky/neat, coarse/smooth, dry/mellow, snappy/soft. (*See* **Sensory Evaluation**: Descriptive Analysis.)

Difference tests These are preference tests to compare a sample against a control: in a simple form it is a paired test or, with one sample duplicated, a triangular test. Preferences and the intensity of deviations are reported. Triangular tests are used to check samples which have slight deviations from the standard.

Scoring test This test is used for production quality control and for assessing chocolate samples after periods of storage under different conditions. Sensory evaluation: 9–7, optimal to good; 6–4, satisfactory to sufficient; 3–1, lacking to unsatisfactory. (*See* **Sensory Evaluation**: Sensory Rating and Scoring Methods.)

Tasting method Solid cocoa masses are chopped and tasted with a spoon. Cocoa powders are

suspended to 10% in warm water, plain chocolates are broken into pieces and tasted.

See also: **Amino Acids**: Properties and Occurrence; **Browning**: Enzymatic – Biochemical Aspects; **Caffeine**; **Carbohydrates**: Classification and Properties; **Chromatography**: Gas Chromatography; High-performance Liquid Chromatography; **Cocoa**: Production, Products, and Use; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Enzymes**: Functions and Characteristics; **Fatty Acids**: Properties; **Lactic Acid Bacteria**; **Fats**: Classification; **Sensory Evaluation**: Sensory Rating and Scoring Methods; Descriptive Analysis; Taste

Further Reading

Berbert PRF (1979) Contribuição para o conhecimento dos açucares componentes da améndoa e do mel do cacau. *Revista Theobroma (Brasil)* 9: 55–61.

Biehl B, Heinrichs H, Ziegeler A, Berghausen H, Hammoor M and Senyuk V (1993) The proteases of ungerminated cocoa seeds and their role in the fermentation process. *Angewandte Botanik (Journal of Applied Botany)* 67: 59–65.

Fincke A (1989) Analytische und technologische Entwicklungen auf dem Gebiet der Kakaoerzeugnissc. *Lebensmittelchemie und gerichtliche Chemie* 43: 49–55.

Forsyth WGC and Quesnel VC (1963) The mechanism of cocoa curing. *Advances in Enzymology* 25: 457–495.

Hardy F and Rodriguez G (1953) Quantitative variations in nitrogenous components of the cacao bean: effect of genetic type and soil type. *Annual Report on Cacao Research, Imperial College of Tropical Agriculture (Trinidad)*, 1945–51: 89–91.

Lehrian DW and Patterson GR (1983) Cocoa fermentation. In: Rehm HJ and Reed G (eds) *Food and Feed Production with Microorganisms. Biotechnology*, vol. 5, pp. 529–575. Weinheim: Verlag Chemie.

Martin RA Jr (1987) Chocolate. *Advances in Food Research* 31: 211–342.

Oberparleiter S and Ziegleder G (1997) Amadori-Verbindungen als Aromavorstufen in Kakao. *Nahrung* 41: 142–145.

Rohan TA (1963) *Processing of Raw Cocoa for the Market*. FAO Agricultural Studies, no. 60. Geneva: Food and Agriculture Organization.

Schnermann P and Schieberle P (1997) Evaluation of key odorants in milk chocolate and cocoa mass by aroma extract dilution analysis. *Journal of Agriculture and Food Chemistry* 45: 867–872.

Schwan RF (1998) Cocoa fermentations conducted with a defined microbial cocktail inoculum. *Applied and Environmental Microbiology* 64: 1477–1483.

Spencer ME and Hodge R (1992) Cloning and sequencing of a cDNA encoding the major storage proteins of Theobroma cacao: identification of the proteins as members of the vicilin class of storage proteins. *Planta* 186: 567–576.

Voigt J and Biehl B (1995) Precursors of the cocoa-specific aroma components are derived from the vicilin-class (7*S*) globulin of the cocoa seeds by proteolytic processing. *Botanica Acta* 108: 283–289.

Ziegleder G (1991) Composition of flavor extracts of raw and roasted cocoas. *Zeitschrift für Lebensmittel Untersuchung und Forschung* 192: 521–525.

Ziegleder G and Biehl B (1988) Analysis of cocoa flavour components and flavour precursors. In: Linskens HF, Jackson JF (eds) *Analysis of Nonalcoholic Beverages. Modern Methods of Plant Analysis*, new series, vol. 8, pp. 321–393. Berlin: Springer-Verlag.

Production, Products, and Use

B Biehl, Technische Universität Braunschweig, Braunschweig, Germany
G Ziegleder, Fraunhofer Institut für Verfahrenstechnik und Verpackung, Freising, Germany

Introduction

This article describes raw cocoa processing and the nature, the technologies, and the use of cocoa powder and chocolate. These are made from defatted and nondefatted raw cocoa beans, respectively, which come from the fermented seeds of the tropical tree *Theobroma cacao* L. 'Cocoa' here is used as the overall term covering the finished products (cocoa powders and chocolate), the semifinished product (the raw cocoa beans), and the original seeds in the pods on the tree.

Although cocoa products contain valuable nutrients, their special nature lies in the unique flavor which is not obtained from other plants. The unprocessed seeds, however, do not have this desired flavor, but instead an astringent unpleasant taste. Hence the constituents expressed by the unique genetic potential of the cocoa seeds are profoundly changed by processing and production techniques, and these are responsible for realizing or destroying the flavor potential, as described in the previous article.

'Processing' covers all steps from harvesting the pods up to the stage where the raw cocoa beans come to the industrial manufacturer's plant for 'production' of the final goods.

Raw Cocoa Processing

General Outline

Processing usually includes the harvesting and breaking of the pods, extraction of the wet beans, transport

(of pods or wet beans), fermentation in covered heaps or containers for some days, drying of the fermented beans, sorting, grading, sacking, and storage. 'Beans' is a traditional term used commercially for seeds of cocoa at any stage of their processing. Wet beans are the seeds covered by the mucilaginous, sweet and acidic pulp or its remainder during fermentation. Raw cocoa beans are fermented and dried. The term 'fermentation' is used to describe the overall process of heaping wet cocoa beans for several days, microbial pulp degradation, and also endogenous postmortem reactions, both enzymatic and nonenzymatic, that take place in the seeds and which are induced by pulp fermentation.

At the cocoa plantations, fermentation and drying may primarily be looked upon as a curing process in order to stabilize the fresh beans by microbial degradation of the firmly adhering, perishable pulp, and by drying.

From this point of view the resulting qualities, especially the flavor potential, may be subordinate aspects. However, the quality is essential for subsequent manufacture of finished products. Raw cocoa quality may vary considerably depending on the manner of processing. First grades command an extra premium. However, since grading takes into account defects of raw cocoa, but not flavor intensities or flavor notes, there is no way of providing an extra premium for raw cocoa with superior flavor potential. Certain raw cocoa origins are given preference (and a higher price) by the industry and the market in virtue of experienced quality differences compared to less desirable origins.

Origin of Flavor Differences

Some experts assume flavor differences to be due to edaphic, climatic, or plant genetic particularities in the cocoa-growing countries. However, these differences may partially or even predominantly be due to traditional processing practices which are not uniform in the various producing countries.

Soil nutrients and climate generally affect the physiology of plants. Climatic changes have influence on pod growth, seed size, the amount of seed constituents, and thus on the starting material used for cocoa processing. Apart from the variable response of cocoa cultivars to climate, there is inadequate knowledge about the significance of genetic diversity of the planted cocoa trees and harvested seeds with respect to its influence on flavor quality. Criollo (fine) and Forastero (bulk) raw cocoas reveal flavor differences which are attributed to genetic influences. A huge number of different genotypes and hybrids, derived from one or both of these varieties and their forms, are planted. During the last decade genotype-depending flavors in raw cocoa beans have been reported which were harvested from trees of different vegetative clones and which were processed under empirical standard conditions. Unequivocal conclusions are complicated since flavor determinant constituents in raw cocoa (like acids, catechins, and peptides) underwent profound transformations during processing. Physiological effects like variation of the seed's individual pulp layers also interfere. In the case of clonal effects it is likely that flavor differences do not depend on the genetic expression of the cotyledons (nibs), but rather on the genotype of the vegetative tree controlling pod growth and the property of pulp and testa. All vegetative parts of a cloned hybrid represent the same genotype but the generative progeny, the embryos, does not.

Practice of Fermentation

Handling and care during processing significantly affect raw cocoa quality and flavor. To improve raw cocoa quality, repeated attempts have been made to modify the methods. Few met general application. The different methods of cocoa fermentation are mostly developed from local traditions under particular conditions in the various cocoa-growing countries. They will not be described in this article. The reader is referred to earlier reviews cited below for detailed and valuable information. Instead, representative processing methods as practiced in Brazil, Ghana, and Malaysia, variable courses of fermentation, and effective modifications will be described. Most of the common practices correspond to the description given in the following sections. Heap fermentation and box fermentation as the prevailing methods will be described briefly.

In both cases, ripe pods are harvested and the husks are opened by means of a cutlass or wooden billet. The wet seeds are removed from the husk and placenta and are collected for transport. Abnormal, black, infested or clustered beans and foreign matter are removed. The wet cocoa is piled up in heaps or boxes of variable capacity, allowing drainage of sweatings through the bottom. Care is taken that mixing can be carried out by turning the fermenting cocoa upside down and inside out. The wet cocoa is covered with banana leaves or jute bags and is left by itself for 5–7 days, interrupted by one or more turnings after 24, 48, or 72 h. After a given time, or following an assessment of its appearance and odor, the fermented cocoa is transferred to a drier.

Standardized methods, such as the heap and box fermentations, are used at well-established farms or estates but they are not in common use elsewhere. Any kind of box or other receptacle such as troughs, frames, baskets, boats, or bags can be found to serve

for curing widely varying quantities of wet cocoa. Alternatively, the wet cocoa is left on the ground or on a drying platform where a partial fermentation may occur, depending on the depth of the layer, occasional heaping, and weather conditions. These irregular types of curing probably play a subordinate role in most of the raw cocoa-producing countries, but they contribute to the nonuniform low-quality cocoas as they appear on the market. However, some special methods are used to produce high-quality cocoas, as is reported from Ecuador, where in some estates the wet beans are spread at day but heaped overnight. This reminds us that cocoa fermentation is still basically an empirical procedure.

Heap fermentation (Figure 1) The typical succession of microbial pulp degradation occurs whenever wet cocoa is piled up for several days. Heap fermentation is used in West Africa. The heap is prepared outdoors on banana leaves layered on wooden sticks to allow better drainage. The cocoa is covered by the same leaves by folding up the peripheral ends in such a way that rain is kept off. The amount of wet cocoa (80–1000 kg) depends on the harvest. Due to the flat, conical shape of the heap compared to boxes, the bottom layer and the surface layer are large, which promotes not only aeration but also cooling. Thus, at least one turning after 48 h is necessary. The highest temperature may normally be found 5 cm beneath the surface.

Box fermentation (Figure 2) Boxes are placed under roofs and are made from locally indigenous timbers. They are jacked up and their bottom (sometimes also the walls) are perforated to allow drainage. On large plantations the fermentary would consist of batteries of several open boxes ($1.0 \times 1.0 \times 1.0$ m) which facilitate turning from one box to the next. Alternatively, long boxes (e.g., 10 m length) can be divided by partitions. Depending on the harvest, the cocoa is filled up to 100 cm in depth.

Compared to average-sized heaps (about 50–80 cm top height), the deep box (100 cm) depth requires a longer time for fermentation because of reduced aeration (almost exclusively from the top).

In Malaysia, shallow boxes have been introduced to overcome the lack of uniformity (see below and Figure 2). Cascades of shallow boxes are used in large estates, each of which may have a capacity of several tonnes, but a height not exceeding 35 cm.

Revolving barrels which are provided with perforations for aeration and with some device for effective mixing during gentle rotations allow quite uniform fermentations. Aeration, pulp loss, and water evaporation are significantly increased by frequent turnings. Whether this is favorable or not depends on the ripeness and constitution of the pulp.

Standard Course of Fermentation

The progress of fermentation is characterized by visual changes and changes of odor. Most typical are

Figure 1 Preparing a heap fermentation (Cocoa Research Institute of Ghana, Ghana). Reproduced from Cocoa: Production, Products and Uses. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Figure 2 Experimental shallow box fermentation (Malaysian Agricultural Research and Development Institute, Malaysia). Reproduced from Cocoa: Production, Products and Uses. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

color changes and the accumulation of a red sap in the beans penetrating the shell when the beans die. The internal appearance of dried beans taken at regular stages from the fermenter also indicates the progress of fermentation. Seeds appear slatey as long as anthocyanins have not yet been released from their intact storage cells. After seed death they would dry to compact beans, violet in color, with the shells sticking strongly to the nibs. When fermentation is completed the nibs would turn brown and the beans retain a plump, porous structure, with a loosened shell. This is the type of bean required for marketing. For grading on the market, the internal appearance of beans is taken to estimate the so-called 'degree of fermentation' in a cut test. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments.)

After harvested pods have been broken, the wet cocoa is inoculated by spores of an ubiquitous range of airborne microorganisms. The course of fermentation may be subdivided into four stages (I–IV). The succession of events is generally as follows (cf. solid line in Figure 3).

Stage I

Pulp In the mucous, acidic (pH 3.5–4.0) pulp, anaerobic, microbial metabolism of sugar starts immediately, with alcoholic fermentation first dominating over the lactic acid fermentation. Pulp is drained off. Carbon dioxide is produced and this displaces air from the box. The odor changes from sweet fruits to ethanol.

Seeds Most seeds at this stage are still alive (germination test).

Stage II

Pulp As more pulp has been drained off and sugar metabolized, less carbon dioxide is produced. Consequently, more air is taken up by the pulp, giving rise to oxidative formation of acetic acid, an exothermic process heating up the cocoa from the ambient temperature to 45–52 °C. Aeration would now be randomized and accelerated by turning. There is an odor of acetic acid.

Seeds Both the temperature increase to > 45 °C and the uptake of acetic acid kill the seeds. A maximum level of acetic acid produced in the pulp is followed by a maximum amount of acetic acid absorbed in the beans. The average pH value in the nibs drops from 6.4 to 4.0–4.7. Oxygen is quantitatively consumed in the fermenting pulp, keeping the nib under anaerobic conditions and allowing postmortem reactions in the absence of oxygen.

Stage III

Pulp As acetic acid production slows down because of exhaustion of substrate, bacterial oxidation of acetic acid causes a slow increase in pulp pH. The temperature may drop because of lack of substrates and reduced microbial activity in the pulp, which is still acidic (about pH 5.0). Lactic acid formation, which has been suppressed during stage II, may increase once more. The odor is still acidic but changes to a bread-like impression.

Seeds Most of the postmortem enzymatic and nonenzymatic reactions in the seeds are completed by this time.

Stage IV

Pulp Now, after pulp degradation, a buoyant air flow from the bottom to the top of the hot cocoa

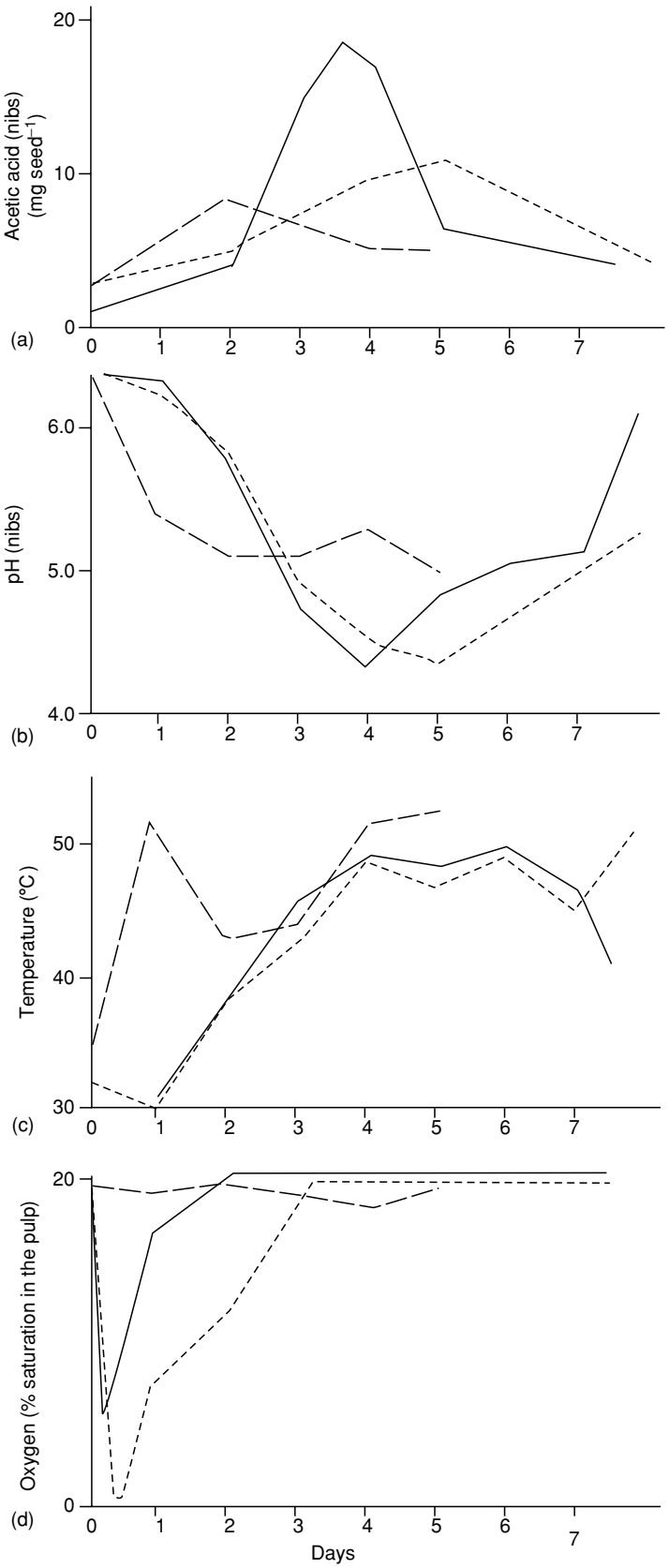

Figure 3 Differences in the course of shallow box fermentations. Solid lines, well-aerated top layer; seeds from unstored ripe pods (up to 8 days of fermentation). Dotted lines, less aerated middle layer; seeds from unstored ripe pods (up to 8 days of fermentation).

provides the beans with access to air. A second temperature increase indicates a new aerobic microbial activity on the shells, which will spread to the nibs if drying is delayed.

Seeds Stage IV seems to be important to insure browning of the beans and to harmonize flavor during subsequent drying. Although no experimental explanation is available, it is probable that developing aerophilic organisms may now help to increase shell permeability. (*See* **Lactic Acid Bacteria**.)

Although this fourth stage cannot clearly be distinguished from the preceding one, it merits special attention for correct termination of fermentation to avoid subsequent overfermentation.

Overfermentation subsequent to stage IV is characterized by a steep increase in pH value, first at the surface of the beans, and later in the nibs. It is accompanied by a pronounced darkening or even blackening of the beans and a characteristic hammy off-flavor which persists in the raw cocoa to an extent which depends on how severely this process took place before drying. Overfermentation may be looked upon as a direct aerophilic microbial attack on the nibs, destroying the cocoa flavor potential. Depending on the bean surface moisture at stage IV, the transition to overfermentation may be slow or very rapid. Wet beans may turn black overnight and may evolve a typical strong hammy off-flavor. It appears that the oxygen concentration in the fermenting mass at this stage is high enough for saturating respiration of aerobes and microaerophilic organisms, but too low for sufficient polyphenoloxidase-triggered browning reactions which would impair microbial growth.

Variables in the Course of Fermentation

Rate of fermentation The above description may give the incorrect impression of a uniform process running on a constant timescale. The stages described in the previous section and illustrated in Figure 3, however, are not bound to a fixed time but may be considerably shorter or longer.

The main reason is aeration: the more the wet cocoa is aerated in stages I and II, the quicker the temperature increases, the quicker the peak of acetic acid appears first and decreases later on (solid and dotted lines in Figure 3). These differences are much more pronounced in deep boxes. The succession of events described is not changed. Wet cocoa containing >1.0 ml of pulp per seed would pass through an ethanol fermentation even under continued aeration during stage I. In contrast, if no aeration is allowed at stage I, the pulp of wet cocoa may appear unchanged for several days. Lactic acid fermentation would slowly acidify the seeds at a low temperature. Later on, forced aeration at stages III or IV causes a quick pH increase in the nibs.

Thus, if fermentations are terminated at a prefixed time, e.g., after 6 days, the cocoa may be at stage III, stage IV, or overfermented, revealing a low or a high pH value in the nibs, respectively. In practice, aeration is intensified by turning, batch size, type of fermenter, and pulp volume per seed: the smaller the batch size, the larger is the specific surface for air uptake. The surface area is also influenced by the type of fermenter. A small conical heap (e.g., 250 kg) is much more aerated than a deep box (100 cm depth). A shallow box (30 cm depth) helps to increase surface aeration.

Nonuniformity in the course of fermentation Not only variation in rates of fermentation but also local differences in the fermenting heaps or boxes are due to inhomogeneous aeration. This applies especially to deep boxes. At stage I, oxygen is only accessible at the surface layer, where the process is accelerated, while the lower layers are still under a carbon dioxide atmosphere. The hot zone of pulp oxidation and acetic acid production slowly moves down to the bottom. If pulp volume is large and the cocoa is not turned, the anaerobic, cold situation (stage I) in the bottom layer may persist during the time the top layer goes through stages I–IV. After drying, the bottom layer would produce slatey or violet beans, whereas the top layer would consist of properly fermented or overfermented brown beans. Turning after 24 and 48 h reduces this inhomogeneity, but not enough to eliminate this effect entirely. As shown in Figure 3, even in shallow boxes acetic acid accumulation and consumption are delayed in a less aerated middle layer compared to a top layer. While this type of nonuniformity is found during fermentation of wet cocoa beans with a fully developed pulp layer, the situation may be different with beans from overripe pods covered by a thin layer of pulp. Too many turnings would reduce the capacity of the pulp to fulfill the entire sequence of stages I–IV and a premature overfermentation would result. Additionally, beans in unturned, aerated niches of

Dashed lines, middle layer; seeds from ripe pods stored for 10 days (up to 5 days of fermentation). (a) Acetic acid in the nibs; (b) pH values in the nibs; (c) temperature in the fermenting mass; (d) available oxygen between the beans. Fermentations were performed with wet beans from Malaysian hybrids. Simplified curves: turnings in (a) after 44, 92, and 139 h; (b) and (c) after 67 and 120 h. Data derived from Meyer B, Biehl B, Said M, Samarakoddy RJ (1989) Post harvest pod storage: a method for pulp preconditioning to impair strong nib acidification during cocoa fermentation in Malaysia. *Journal of Science and Food Agriculture* 48: 285–304.

the batches in boxes or heaps may undergo abnormal changes due to heat loss, drying out, or fungal attack.

Effect of harvest and postharvest treatment on fermentation and raw cocoa quality Controlled aeration is helpful in overcoming nonuniformity but does not help in avoiding strong nib acidification (Figure 3). The minimal pH value in the beans during enzymatic reactions at stage II is essential for the formation of flavor precursors rather than the nib pH when terminating fermentation. With a large pulp volume per seed, an excess of aeration during the first stages accelerates acidification, but does neither limit acetic acid production nor decrease of nib pH values to less than pH 4.5 (Figure 3, solid line). However, nib acidification is moderate if the pulp volume per seed is small (Figure 3, dashed line).

The composition and volume of the pulp depend on pod ripeness at harvest and postharvest treatments. Ten days of open-air postharvest pod storage or several hours of bean spreading for pulp drying after breaking the pods were found significantly to reduce pulp volume, pulp water, and pulp sugar per seed (Table 1). These methods are particularly helpful prior to fermentation of small seeds with a strong layer of pulp. Subsequent shallow box fermentation proceeds with well-aerated pulp during stage I. Correspondingly, there is an early steep temperature increase, but the decrease of nib pH value and the formation and accumulation of acetic acid in pulp and nibs is considerably reduced (Figure 3). It is assumed that early aeration of a thin pulp surface layer enhances respiration of sugars by yeasts and reduces alcoholic fermentation beneath the surface, in contrast to a voluminous pulp.

The important pulp/seed relation depends on both genetic and physiological parameters: large seeds bear less pulp per seed surface area than small seeds. However, seed size does not only depend on genetics. Pods and seeds may be considerably smaller than usual when growing under adverse conditions, e.g., during severe drought. Furthermore, the pulp-to-seed ratio is significantly reduced during ripening. Large pulp volumes per seed from unripe seeds cause badly aerated slimy fermentations. In overripe pods the pulp volume is reduced as with stored pods, leading to well-aerated, rapid fermentations. Raw cocoa flavor characteristics with particular origins may depend to a large extent on these physiological stages of wet cocoa due to local traditions in postharvest treatments. In average, bean size and ripeness of pods harvested in Malaysia or in Ghana are quite different.

Seasonal effects on fermentation and raw cocoa quality are often due to these differences in the pulp. In practice, it depends on the circumstances discussed above whether or not pulp preconditioning is advisable.

Finally, if wet cocoa is exposed to rain, subsequent fermentation may lack sugar for heating and acetic acid production, and would proceed in an erratic way.

Drying

Enzyme-controlled browning reactions which do not occur before drying are essential for quality.

For sun-drying, the beans are spread on the ground, on mats or on wooden floors raised from the ground and protected against rain. Care is taken to move the drying cocoa. The layers of cocoa are made shallow or high to regulate the time of drying. Five to seven days are usually necessary to reduce the water content to less than 7.0%, a limit for fungal growth. (*See* **Drying**: Drying Using Natural Radiation.)

For artificial drying, wood fires or oil burners are connected to a flue under a floor of closely spaced slats. There are many devices of this type for passing a stream of hot air, but not of smoke, through the layer of cocoa beans. The industry rejects smoky beans but the smell of smoke may be confused with a hammy off-flavor of overfermented beans. On large estates, different types of industrial driers are used for

Table 1 Changes in the pulp during pod storage and bean spreading

Wet beans	*Pulp volume (ml seed^{-1})*	*Pulp water (g seed^{-1})*	*Sucrose + glucose (mg seed^{-1})*[e]	*Reducing sugars (mg seed^{-1})*[f]
Malaysia, ripe (from unstored pods)[a]	1.19 ± 0.14	0.95 ± 0.12	121.3 ± 57.1	
Malaysia, ripe (from pods stored for 10 days)[b]	0.74 ± 0.06	0.58 ± 0.05	39.9 ± 5.4	
Malaysia (from unstored pods after spreading)[c]	0.46 ± 0.06	0.34 ± 0.12	79.2 ± 32.3	
Ghana, ripe (from unstored pods)[d]	0.84	0.69		91.2
Ghana (from pods stored for 7 days)[d]	0.55	0.41		68.6

Mean values and standard deviations from samples of [a]10 and [b]four experiments with 1000 pods each.
[c]Samples from four experiments with 130 kg of wet cocoa, each of different ripeness, sun spreading and surface drying for several hours.
[d]From one experiment with 10 pods each.
[e]Estimated polarographically.
[f]3,5-dinitrosalicylic acid method (Biehl B, Meyer B, Crone G, Pollmann L (1989) Chemical and physical changes in the pulp during ripening and post harvest storage of cocoa pods. *Journal of Science and Food Agriculture* 48: 189–208).

effective, short-time drying. (*See* **Drying**: Theory of Air-drying.)

The first stage of drying should be effective in order to avoid overfermentation or fungal growth on the wet cocoa, if the internal pH value is high. At this stage, there is no danger in using hot air (e.g., 100 °C), although the temperature of the beans should not exceed 40–60 °C. In a shallow layer of cocoa beans, water evaporation will keep the beans cool. However, hot air flow through a deep layer ($\geq$ 10 cm) would heat the beans excessively. After external dryness of the beans has been attained, the second stage of drying can be continued slowly to facilitate enzymatic browning in the nibs and to allow moisture equilibrium in the beans from the wet core to the dry surface. Artificial drying may be interrupted for a while by a resting period. In a third stage, the cocoa should effectively be dried to < 7.5% water (w/w), preferably in a stream of warm (< 60 °C) but not hot air.

Rewetting, polishing, and drying raw cocoa are practiced locally to improve the external appearance, especially of moldy beans, but shell breaking and reinfection of the nibs may result. Rewetting also increases browning of violet beans.

After drying, the raw cocoa is sorted by removing flat, broken, externally moldy beans and foreign matter and by sieving (in the case of nonuniform bean size) to meet the quality standards. Locally, batches from different origins may be blended to reach the maximum percentage of the particular visible bean defects which are allowed in the official grading system.

Storage

Provided there are good storage conditions, properly processed raw cocoa can be stored for years without any unwanted, serious sensory or analytical indication of loss of quality or of spoilage. However, once processed, raw cocoa usually travels a long way from the tropical origin to the final manufacturing plant, being subjected to several risks underway. These risks are primarily water and high air humidity, insect infestation, and consequences of improper processing. (*See* **Storage Stability**: Parameters Affecting Storage Stability.)

In addition to general rules for storage in tropical and temperate climates, the hygroscopy of raw cocoa beans must be considered. Dry raw cocoa containing < 7.5% water will not be attacked by fungi. According to moisture absorption isotherms, this bean moisture would not be exceeded at < 80% relative humidity. However, at a high level of air humidity, any reduction in air temperature (e.g., during shipping) would cause damage by water condensation and mold growth. Residual humidity from fast drying must be considered. Repeated moisture determinations are necessary in every lot. Fungal spores are ubiquitous. However, moldy beans resulting from fermentation and drying increase the danger of fungal growth at a critical level of moisture. Thermophilic and xerophilic fungi such as *Aspergillus glaucus, A. fumigatus, Penicillium* spp. and *Mucor* sp. have been found to occur during both processing and storage. Internal mold is further facilitated on broken beans and on injured shells. Brittle shells, particularly from overfermented or rewetted cocoa, not only increase the danger of mold but also increase the amount of broken beans during transport and stacking of bags. (*See* **Spoilage**: Molds in Spoilage.)

Out of nine insect species listed by Rohan, four were mentioned to be important threats to stored raw cocoa in the tropics: *Cadra cautella, Lasioderma serricorne, Araecerus fasciculatus* and *Tribolium castaneum*. *Ephistia elutella*, the cocoa moth, is only found in temperate climates. In tropical countries, insects are controlled by fumigation of raw cocoa using the gases methyl bromide or phosphine applied to stacks of stored cocoa under a gas-proof and waterproof polythene sheet for 1.5 h. The gas is then extracted by fans. The sheets are left to protect the cocoa against uptake of moisture and insects. (*See* **Fumigants**; **Insect Pests**: Problems Caused by Insects and Mites.)

Production of Cocoa Powder and Semifinished Products

Cocoa Products

Two products are derived solely from cocoa beans: chocolate and cocoa powder. These products are interrelated, and two intermediate products play an important role – cocoa mass and cocoa butter. This relationship is made clear in Figure 4. The system

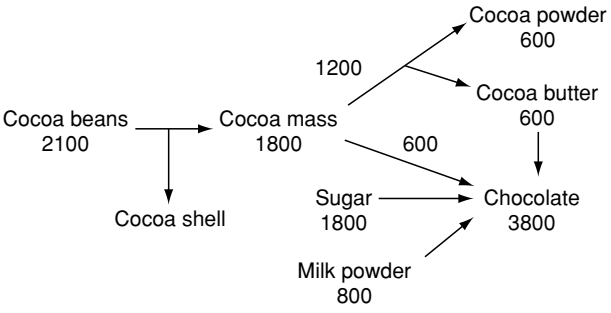

Figure 4 Products derived from cocoa beans. The figures are approximate and in 1000 tonnes. Reproduced from Cocoa: Production, Products and Uses. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

shown in Figure 4 is not fully in balance. The chocolate industry needs a certain quantity of cocoa butter and this implies that a fixed quantity of cocoa powder is also produced. However, the demand for cocoa powder does not always meet this volume. As a result, the price of cocoa butter is usually substantially higher than that of cocoa powder. The relatively high price of cocoa butter has made it attractive to lipid chemists to look for cheaper fats which could replace cocoa butter. This is not easy, as cocoa butter has unique properties, especially hardness and melting behavior. In the course of time, cocoa butter substitutes (CBSs) were developed, mostly based on palm kernel fat. These are used in combination with cocoa powder for manufacturing imitation chocolate, or coatings (for biscuits, cakes, icecream, etc). Recently, ways have been found to separate the desired triacylglycerols from other fats and to put them together in the right proportion, thus obtaining a fat that closely resembles cocoa butter. Such a fat is called a cocoa butter equivalent (CBE) and can replace cocoa butter in chocolate (Figure 5).

Quality Aspects

Cocoa powder is sold in retail packages to consumers. However, most cocoa powder reaches the consumer as a color and flavor in other products, like dessert powders, sterilized chocolate milk, chocolate cake, icecream, etc. Most of the cocoa powder produced is sold as an ingredient to other food manufacturers. Such a product should have the necessary quality aspects to meet the needs of the user, regarding microbiology, consistency, purity, and other factors. In the case of cocoa powder this can only be fulfilled when maintaining strict good manufacturing practices, the reason being that there are a number of problems which put cocoa powder in a high-risk category:

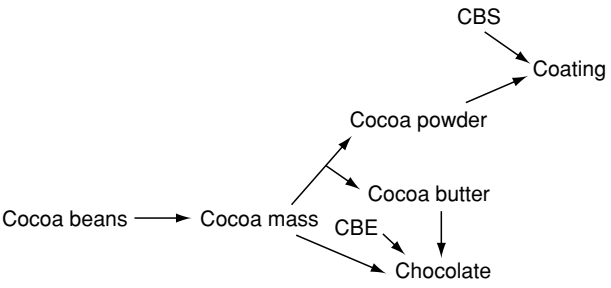

Figure 5 The use of alternative fats to cocoa butter. CBS, cocoa butter substitute; CBE, cocoa butter equivalent. Reproduced from Cocoa: Production, Products and Uses. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

- Cocoa beans undergo a fermentation process, which leaves very high bacteria counts on the bean shell.
- During fermentation the temperature goes up to 50 °C, which leaves thermoresistant spores on the bean. These may interfere with the sterilization process of chocolate milk.
- After fermentation, the cocoa beans are dried in the open air, allowing contamination with bird droppings. *Salmonella* bacteria can thus be found on crude cocoa beans.
- Bacteria which are present in cocoa mass are surrounded by fat during the grinding process. The fat gives them good protection and it has been shown that *Salmonella* can live for months in chocolate. Also, the bacteria will be protected in the stomach against its natural acidity when chocolate is eaten.
- Cocoa powder is used in many different foodstuffs, made with different manufacturing processes, requiring different shelf-lives, and having different moisture contents. As the cocoa manufacturer does not always know how the powder will be used, all the product must meet the strictest hygiene specifications. (*See* **Spoilage**: Bacterial Spoilage.)

Environment

The cocoa-processing industry is not one of the major polluters of the environment. The raw materials and the products are simply too expensive to be discarded. Of course, there are the general problems of each factory, such as noise and high, ugly buildings. How these can be avoided is well known and is taken into consideration when new chocolate factories are built. The main problem comes from the smell of exhaust gases – air blown through the grinding systems and combustion gas from the roasters. These gases also contain dust. To collect the dust, cyclones are used on a large scale in the cocoa industry. Biological filters are being tested for the removal of odors but they have not been completely successful. The present method of solving the problem is to build high chimneys. Cocoa bean-processing gives one waste product – the shell, about 10% of the weight of the beans. This loose material is usually ground to reduce its volume and sold to fertilizer or cattle feed manufacturers. Burning the shell could be worthwhile when energy costs are high, but introduces the problem of removing large quantities of dust from the combustion gas.

Production of Cocoa Mass

All cocoa beans are initially turned into cocoa mass. The necessary processes are illustrated in Figure 6. The word 'nib' is used in all languages to indicate the

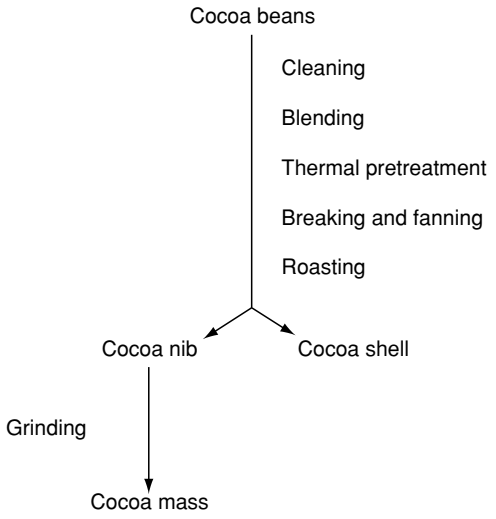

Figure 6 The production of cocoa mass. Reproduced from Cocoa: Production, Products and Uses. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

broken pieces of the kernel of the cocoa bean. Cocoa mass contains approximately 55% cocoa butter. The fat is liquid after the roasting and grinding procedures, which make the whole mass liquid. It can be pumped and transported in tank lorries.

Cleaning Cocoa beans, from silo storage or directly from their burlap bags, first pass through cleaning machines – screens, magnets and controlled air streams. Extraneous materials removed include sticks, stones, string, and metal objects.

Blending Substantial variety in flavor exists among cocoa beans from different countries. It is seldom that one batch of beans will be used exclusively in a formula. One reason is that blending provides the opportunity to obtain a certain flavor. Another reason stems from the inconsistency of cocoa beans in flavor and other quality aspects. Batches of cocoa beans show variations from one locality to another, due to differences in the weather during growth and ripening, from differences in fermentation and other variations in processing conditions. Blending will contribute to uniformity.

Thermal pretreatment An important quality aspect in cocoa bean processing is removal of the cocoa shell. The shell, as the covering of the kernel, is always greatly polluted with sand, high bacteria counts, and pesticide residues. Depending on the fermentation, the shell often sticks to the kernel, which prevents easy removal. Recently machines have been introduced which loosen the shells by means of a thermal shock. This is achieved with hot air, saturated steam, or infrared radiation.

Breaking and fanning To remove the shell, the beans are first broken between adjustable toothed rollers. The broken pieces are subsequently separated in fractions by sieving. Each fraction is treated with a stream of air which carries away the light shell pieces. This breaking and fanning process is often referred to as 'winnowing.'

Roasting The roasting process at 120–140 °C is needed for the development of the typical flavor of cocoa. Using modern technology, cocoas can be roasted at different particle sizes, such as whole beans, or as liquid cocoa mass, which is produced by a fine grinding of cocoa and liquefying within its own fat. The roasting of smaller particles has the advantages of better controlled roasting, partial exhaustion of abundant acetic acid, and reduced roasting time (beans take about 30 min, nibs 12 min, mass 2 min). The type of roasting technique may also influence microbiological aspects. The bacteria present are on the shell and by removing the shell first the material enters the roasting (= sterilization) process with a lower initial bacteria count.

Grinding and refining The next step is grinding of the cocoa nib particles. The nib consists of 55% cocoa butter and 45% solid material coming from the plant cells. The latter must be ground very finely. When eating chocolate or drinking chocolate milk, no grittiness should be felt in the mouth. For both cocoa powder and chocolate it is important that the final particle size distribution is narrow. This facilitates the pressing operation and improves the rheological properties of chocolate. The desired fineness (15–70 μm) and particle size distribution are reached in several successive grinding steps. Pregrinding of the large and hard nib particles is done in disk mills, hammer mills, or pin mills. To reach the ultimate fineness, mass for pressing is refined in modern versions of the triple stone mill, or in vertical ball mills (attrittors). In the chocolate industry, roller mills with five rolls are most commonly used for refining the mass. These are also used to grind the chocolate mass, after adding preground sugar and milk powder. (*See* **Milling**: Characteristics of Milled Products.)

Types Cocoa mass may be used for the production of chocolate, or for pressing, i.e., in the production of cocoa butter and cocoa powder. The difference between these two types of cocoa mass are the bean

blend, fineness, and microbiological condition, among others.

For chocolate, many manufacturers prefer beans from Ghana and Nigeria on account of their desirable flavor attributes. For the highest qualities, South American so-called flavor beans may also be added to the blend.

Although for all purposes mass needs to be completely safe from the bacteriological point of view, cocoa powder requires, moreover, that thermoresistant spores are absent. This condition must already have been reached in the mass, which tolerates higher temperatures than cocoa powder.

Mass for pressing is also already alkalized. This process will be discussed below.

Packaging and shelf-life When cocoa mass has to be stored or transported, it is usually packed into 30-kg cardboard boxes with a plastic liner or bag inside. In Europe and the USA most of the transport takes place in liquid form.

Cocoa mass has a very good shelf-life. The solid particles are protected by a fat with a high degree of saturation. Moreover, cocoa contains powerful natural antioxidants. These conditions also apply to cocoa beans. If the moisture content of the beans is well under control, their shelf-life will be several years. (*See* **Antioxidants**: Natural Antioxidants.)

Production of cocoa butter Cocoa mass contains 55% cocoa butter. Part of this can be extracted by mechanical pressing. The solids which stay in the press contain 22% fat, or even as low as 10%. These press cakes are ground to cocoa powder. The cocoa butter is filtered and part of it is deodorized. The production process is illustrated in Figure 7.

Pressing The cocoa mass is pumped to horizontal hydraulic presses, with pots lined up face to face, each equipped with very fine mesh metal filter screens. When all the pots are filled with mass at a temperature of 90 °C, the hydraulic ram is set in motion and the cocoa butter begins to flow through the screens on both sides of each pot. The pressure then increases to 400 bars. The hard cocoa cakes remaining are then discharged from the press. The cocoa butter collected from the press is not 'clean,' having carried with it a small amount of tiny nonfat particles. It is therefore filtered through filter paper.

Deodorization and blending Crude cocoa butter has a strong flavor. This is desirable in dark chocolate. However, in milk chocolate much more cocoa butter is used and the cocoa flavor would become too strong, suppressing the milk flavor. Thus, buyers are stipulating a cocoa butter with a weak flavor, or no flavor at all. This is reached by steam deodorization of the cocoa butter and blending crude and deodorized butter to the desired flavor. This is controlled by sensory testing, or a chemical method, the determination of the flavor profile. Crude cocoa butter, being a natural product, has a rather inconsistent flavor strength. Even if crude cocoa butter is asked for, some blending with deodorized butter is performed

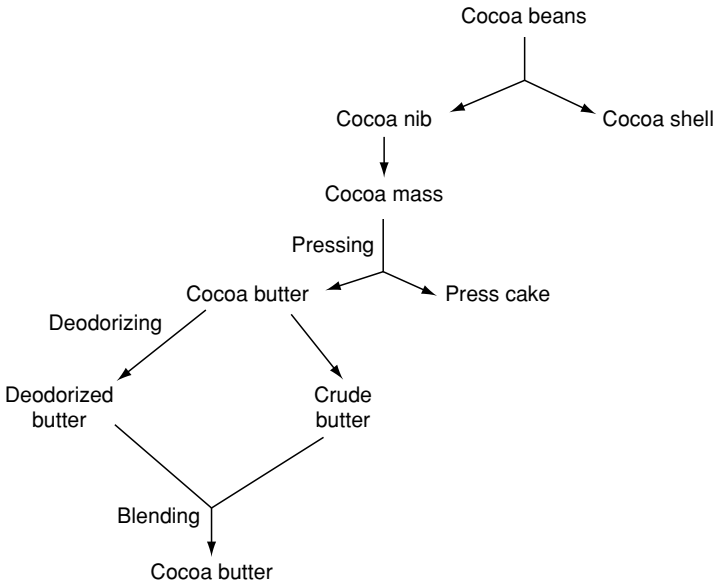

Figure 7 The production of cocoa butter. Reproduced from Cocoa: Production, Products and Uses. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

to standardize the flavor. (*See* **Sensory Evaluation**: Taste.)

Properties Cocoa butter is one of the most expensive natural fats. This is due to its unique melting behavior. Cocoa butter has a much narrower melting range than any other fat. This quality is fundamental to chocolate. Chocolate should be hard – even in hot weather – and it should not stick to the fingers. At room temperature, about 70% of the triacylglycerols in cocoa butter are solid. On eating, the cocoa butter in the chocolate should melt completely in the mouth and cause a pleasant taste sensation due to the consumed melting enthalpy. The relatively narrow melting range of cocoa butter is important in the production of chocolate. When the chocolate is cooled and the fat solidifies, a considerable decrease in volume takes place. This contraction makes it easy to release chocolate articles from their molds.

Types The above discussion has been restricted to the processing of sound cocoa beans, from which the shell has been properly removed. The fat derived from this raw material is called prime pure pressed cocoa butter. In practice, not all cocoa beans that are harvested are of a good quality. Even if the beans are unripe, moldy, or smoky, they contain a fat that can be made fully fit for human consumption by refining. When processing subgrade cocoa beans, it is difficult to remove their shells. Such beans are processed whole in continuous expeller presses. The resulting press cake still holds another 10% of fat, which can be obtained by solvent extraction. The remaining solids are, of course, not edible and are best put back on the land as a fertilizer. Although the fat will be edible after refining, the quality is inferior to the prime pure pressed butter. It is less hard and shows a smaller volume contraction. The following types of subgrade cocoa fat are recognized by the Codex Alimentarius:

1. Expeller cocoa butter: the fat extracted by means of mechanical pressing from material which has approximately the composition of whole cocoa beans.
2. Solvent-extracted cocoa butter: the fat obtained from cocoa beans or cocoa waste materials by means of extraction with permitted solvents.
3. Refined cocoa butter: any of the fats obtained in the ways mentioned above and afterwards fully refined according to the standard processing techniques for edible oils and fats.
4. Cocoa fat: fat extracted from waste materials and having a quality below certain standards.

The above processing options are summarized in Figure 8.

Packaging, transport, and shelf-life Much like cocoa mass, cocoa butter is packed in 30-kg slabs or transported and stored in a liquid form. The shelf-life

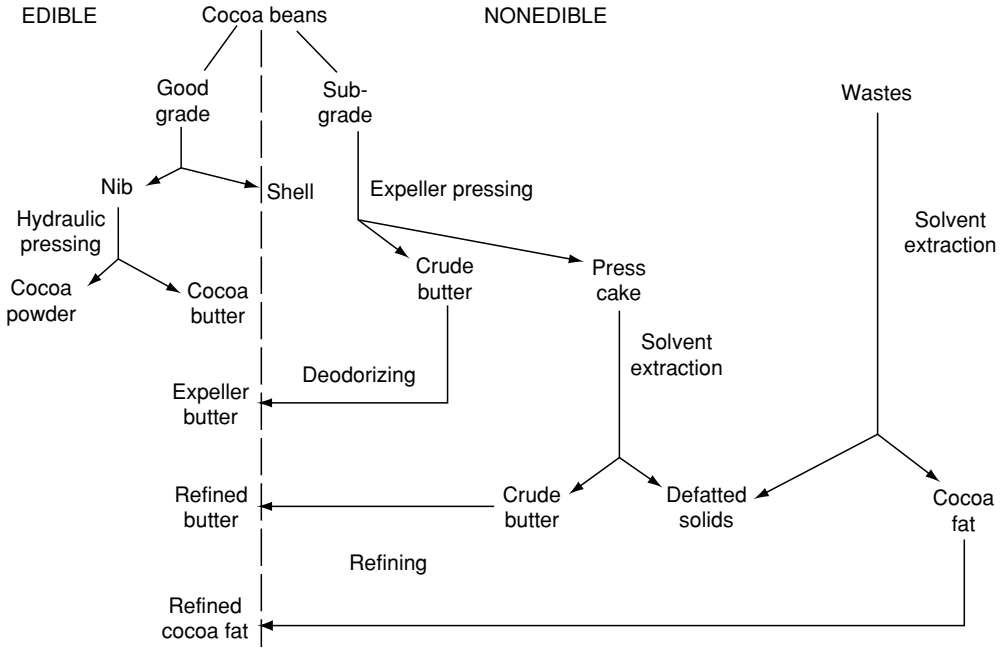

Figure 8 Types of cocoa butter. Reproduced from Cocoa: Production, Products and Uses. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

of cocoa butter is good, due to its high degree of saturation, and the shelf-life of solid cocoa butter is several months to a year, if properly packed and stored.

Production of Cocoa Powder

The press cakes are ground to powder, which is packed after complete cooling. The alkalizing process can be used to create many types of cocoa powder with differing colors. The production of cocoa powder is shown in Figure 9.

Alkalizing The alkalizing process was invented in the first half of the 19th century in the Netherlands. It improves the quality of cocoa powder in two ways:

1. It takes away the slightly acid taste of cocoa.
2. It makes the color darker. This is due to condensation reactions forming high-molecular-weight colored products. Careful control of the reactions can lead to different shades of color: orange, red, brown, and even black is possible.

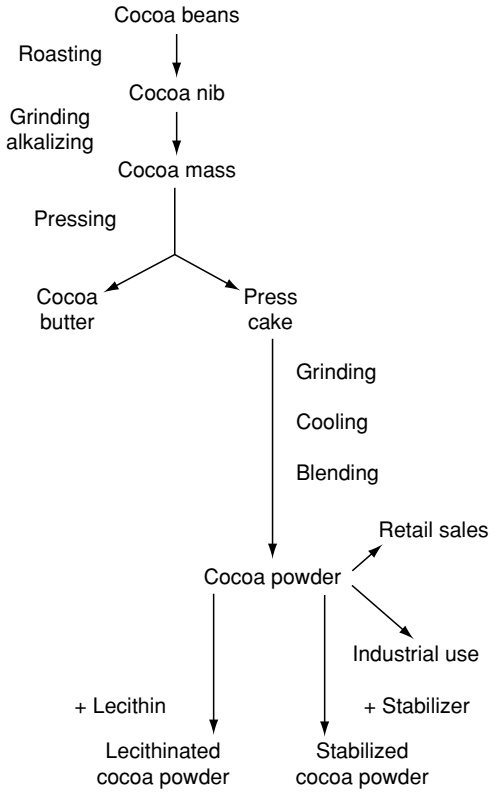

Figure 9 The production of cocoa powder. Reproduced from Cocoa: Production, Products and Uses. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Alkalizing consists of treating the cocoa with a solution of an alkali, mostly potash. Legally, the maximum amount of potash which may be used is 3%, calculated on the nib. The treatment can be performed with the nib, mass, press cake, or powder. The specific color is reached by choice of the reaction conditions: temperature, time, amount and concentration of the alkaline solution, and other factors.

Grinding, cooling, and blending of cocoa cake The press cakes are broken and then ground to powder in pin or other mills. The cocoa powder leaves the mill hot. It should be completely cooled before packing, otherwise the setting fat will turn it into hard lumps inside the package.

Blending of pieces of broken cake before grinding allows standardization of the color, or the preparation of blends with intermediate colors.

Types Variations in the two main processes lead to the formation of hundreds of different types of cocoa powder. The pressing can be performed to a fat content in the cake of 20% (giving cocoa powder) or to 10% (giving low-fat cocoa powder), while the alkalizing process creates many different colors. Cocoa powder with 20% fat is the common household type. The food industry uses mostly low-fat cocoa powder (10%), in many different colors.

For use in certain cocoa-flavored products, some special types of cocoa powder have been developed. Lecithinated cocoa powder contains 5% of soya lecithin. Owing to its fat content, cocoa powder is difficult to wet and to disperse in water. Even in hot water or milk, lumps will be formed easily. Lecithin can improve these properties by its action as a wetting agent. Five percent of soya lecithin is intensively mixed with cocoa powder and the resulting lecithinated powder is agglomerated with sugar. This gives so-called instant cocoa, which can be put directly into cold milk. (*See* **Emulsifiers**: Organic Emulsifiers.)

Stabilized cocoa powder contains about 2% carrageenan. This is a polysaccharide derived from seaweed. It prevents the cocoa powder from settling in sterilized chocolate milk. (*See* **Stabilizers**: Types and Function.)

Packaging, transport, and shelf-life Paper or tin retail packaging is well known. For bulk deliveries, 25-kg (or 50-lb) multiple paper bags are commonly used. Loose bulk delivery or storage of cocoa powder is difficult. The fat-containing fine powder tends to stick and to block pipelines for pneumatic transport. Only large users, in cooperation with large suppliers, have developed workable bulk systems. The shelf-life of cocoa powder is excellent. Even after 10 years the

flavor is good when packed in air- and moisture-proof containers. Only some fading of the color may be noticed, comparable with fat bloom on chocolate. This disappears when the cocoa is used in milk.

Production of Chocolate

The major ingredient of chocolate is cocoa mass, which is mixed with sugar and, in the case of milk chocolate, also with milk powder. The conching process then follows, which is very important for the development of the full chocolate flavor. Melted cocoa butter is added, the chocolate mass becomes liquid, and is cooled, tempered, and poured into molds to form chocolate products (Figure 10).

Mixing of ingredients and refining Sugar, cocoa mass, and milk powder are intensively mixed, forming a dry powder. This powder is preground in various types of mills and then finely ground in a five-roll roller refiner. Small operators also use vertical ball mills. The sizes of particles in chocolate are between 15 and 70 μm.

Conching The flavor which has been formed during roasting is rounded off into the typical chocolate flavor by the conching process. This process also contributes to the physical properties of the chocolate, and hence its eating characteristics. Conching is a mechanical treatment of the chocolate mass in large containers fitted with rollers, paddles, or a variety of other devices. Chemical and physical changes take place under the influence of air which is brought into the mass, at a temperature of about 60 °C, and of rubbing and shearing forces. The result is the formation or liberation of flavor components which give the delicate chocolate note. The chemistry of this process is largely unknown. Physically the dry, crumbled mass is converted into a flowable liquid suspension. The sharp edges of the sugar particles are rounded off, which will give the chocolate a smooth feeling in the mouth.

Three phases can be distinguished in conching:

1. Dry phase: the powdered chocolate mass loses moisture and less desirable volatile flavor components like acetic acid.
2. Pasty phase: during dry conching the mass becomes pasty. Flavor development takes place under the influence of shearing forces.
3. Liquid phase: the last component, cocoa butter, is added. The chocolate mass becomes liquid. Homogenizing takes place under the influence of intensive stirring and shearing.

The final degree of viscosity is very important: the melted chocolate must be sufficiently thin to fill all cavities of the molds. The melted cocoa butter is the liquid phase. However, this is the most expensive component of chocolate. A substantial saving of butter can be obtained by using an emulsifier. Soya lecithin (0.3%) is the most commonly used emulsifier which is added during the third conching phase.

A special way to produce milk chocolate is from crumb. The sugar and cocoa mass are brought into liquid milk and the whole is dried. During the lengthy drying a caramel flavor is developed. Dry milk chocolate crumb has a good shelf-life, due to the natural antioxidants of cocoa. To make chocolate, only cocoa butter has to be added. Crumb is much used in the UK.

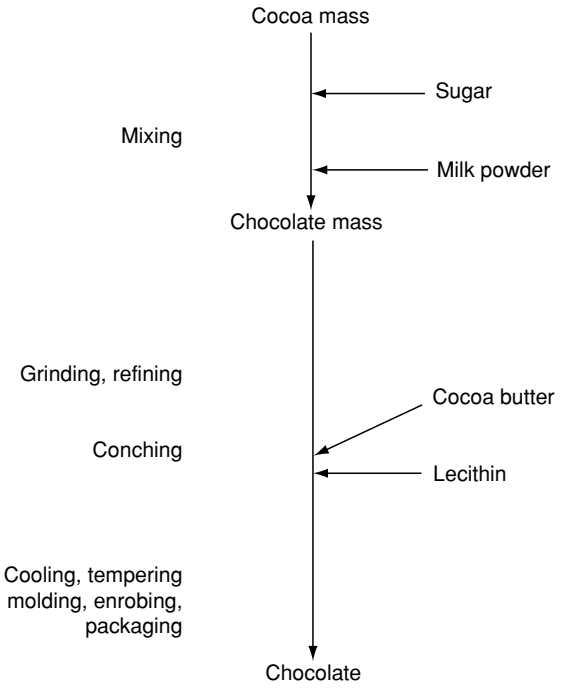

Figure 10 The production of chocolate. Reproduced from Cocoa: Production, Product and Uses. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Cooling and Tempering

Critical to the appearance, gloss, shelf-life, and mouth feel of chocolate is the way it is cooled before solidifying in molds. Cocoa butter triacylglycerides can set in several polymorphic forms. Some are unstable and recrystallization into the stable form will occur after some hours or days. This will result in a loss of gloss and the formation of white fat crystals on the surface of the chocolate – the so-called 'fat bloom.' To obtain stable crystals, stable seed crystals must be formed first. This is done in a process called

tempering. The liquid chocolate at a temperature of 45–50 °C is cooled to 32 °C, then to 27–27.5 °C and treated by special shear elements. Shear forces enhance the nucleation of stable cocoa butter crystals. The temperature is now raised to 29–31 °C, causing any crystals of unstable form to melt. The exact temperatures required for this process depend on the type of chocolate. After tempering, the chocolate is ready to be poured into the molds. It is still liquid but contains 0.1–1% of crystallized fat as seeding material.

Molding and Enrobing

Chocolate articles are made in two distinct ways.

In molding, the chocolate is poured into a mold, which then moves on a belt through a cooling tunnel. After setting, the mold is turned upside down and as a result of the volume contraction of cocoa butter, the article will easily fall out of the mold. Solid bars are then ready for packaging. It is also possible partly to fill the mold. After filling, the mold is immediately emptied, and a thin layer remains on the bottom and wall of the mold. The mold is then cooled, and a filling is poured in. After cooling, a further chocolate layer is put on the filling: this becomes the bottom of the candy when the mold is turned. Large filled bars are made this way, as also is smaller confectionery.

To make Easter eggs and other hollow products, a small quantity of chocolate is poured into a split mold. The mold is then closed and put in a shaking machine to cool. The chocolate sets against the inner wall.

In enrobing, the liquid chocolate is poured over a solid center. Many different products are made this way, like candy bars, biscuits, cakes, and icecream. The viscosity of the chocolate determines the thickness of the layer. This can be controlled with the fat content of the chocolate, which is usually slightly higher than that of molding chocolate (Table 2).

Packaging, Storage, and Shelf-life

Packaging should protect chocolate products against moisture and against odors, which can easily be picked up by the chocolate fat. Furthermore, of course, the outside of the packaging must radiate the delicacy and luxury which is the image of chocolate. Up to now these requirements have resulted in multiple packages: aluminum foil first, paper or heat-sealed foil, finally a cardboard box. Often, the combination of materials prevents reuse and it is foreseeable that less complicated packaging will be developed in the near future.

Chocolate, with its very low moisture content (below 1%) and natural antioxidants, has a very good shelf-life. However, rats, insects, and molds like it just as much as humans. Thus, proper storage is important. Another danger arises from high and irregular storage temperatures. The latter causes the formation of fat bloom. Although the white color of the fat resembles mold growing on the product, the phenomenon is completely harmless.

Food Uses of Cocoa Products

Uses of Cocoa Powder

Cocoa powder is used as a natural color and flavor in a wide variety of sweet foodstuffs. These may be solid, semisolid, or liquid, and consumed frozen or hot. These products can be grouped in the following way:

1. Aqueous systems, consumed frozen. This is the category to which icecream belongs. The cocoa powder used is mostly low-fat (10%). A great variety of colors is used to attain effects resembling dark or milk chocolate.
2. Aqueous systems, consumed at room temperature. Examples are milk products like chocolate milk drinks and desserts. Lightly alkalized cocoa with 10% fat is mostly used.
3. Aqueous systems, consumed hot. Hot chocolate is the product in which lightly alkalized cocoa powder with 20% fat is used most frequently.
4. Fat systems. These are coatings based on cocoa butter substitutes, made with 10% fat cocoa and often not alkalized (real chocolate is also not alkalized). Cocoa butter substitutes may be used for the sake of economy and, indeed, their use may be unavoidable for technical reasons. This is the case with coatings on icecream. At the low temperature of icecream, chocolate becomes too brittle and softer fats are needed.

Table 2 Indicative composition of some types of chocolate

	Cocoa mass (%)	*Cocoa butter (%)*	*Milk powder (%)*	*Sugar (%)*
Dark chocolate	40	10	—	50
Milk chocolate	10	20	15	55
Enrobing chocolate	40	15	—	45
White chocolate	—	25	25	50

Uses of Chocolate

Chocolate can be consumed as such, but it is also much used as an ingredient on or in other foodstuffs. Enrobing of biscuits and other items has already been mentioned. Chocolate drops go into biscuits, and chocolate is used in between wafers. The latter is an example of the possible use of chocolate in tropical or subtropical countries. Even if the chocolate melts, it does not spoil the outside appearance of the product, nor does it stick to the fingers.

Legislation of Cocoa Products

General directives, serving as a guideline for national food laws, are laid down by the Codex Alimentarius. Based on this, the European Union has developed harmonized directives for its member countries. The Food, Drug, and Cosmetics Act in the USA gives similar rules. In most of these countries the name 'chocolate' may only be used if no other fat is present but cocoa butter. CBE fats, up to 5% of the total chocolate content, may be used in the European Union. In Japan three types of chocolate are distinguished: pure chocolate, chocolate, and semichocolate. Only the first of these may not contain fats other than cocoa butter. *See* **Legislation**: International Standards.

See also: **Antioxidants**: Natural Antioxidants; **Cocoa**: Chemistry of Processing; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Drying**: Theory of Air-drying; Drying Using Natural Radiation; **Emulsifiers**: Organic Emulsifiers; **Lactic Acid Bacteria**; **Legislation**: International Standards; **Fats**: Classification; **Milling**: Characteristics of Milled Products; **Sensory Evaluation**: Taste; **Spoilage**: Bacterial Spoilage; Molds in Spoilage; **Stabilizers**: Types and Function; **Storage Stability**: Parameters Affecting Storage Stability

Further Reading

Beckett ST (1988) *Industrial Chocolate Manufacture and Use*. London: Blackie.

Biehl B (1995) Post harvest processing in Asia: approaches towards flavour improvement. In: *Seminar Rencontres Cacao – les Différents Aspects de la Qualité*. pp. 75–89.Montpellier, France: AFCC/CIRAD.

Biehl B, Meyer B, Crone G and Pollmann L (1989) Chemical and physical changes in the pulp during ripening and post harvest storage of cocoa pods. *Journal of Science and Food Agriculture* 48: 189–208.

Cook LR and Meursing EH (1982) *Chocolate Production and Use*. New York: Harcourt Brace Jovanovich.

Drouven H, Fabry I and Göpel G (1996) *Süßwarentechnologie I: Schokolade*. Stolberg-Vicht: Drouven & Fabry.

Fincke A (1965) *Handbuch der Kakaoerzeugnisse*. Berlin: Springer-Verlag.

Hunter JR (1990) The status of cacao (*Theobroma cacao*, Sterculiaceae) in the western hemisphere. *Economic Botany* 44: 425–439.

Knapp AW (1937) *Cacao Fermentation: A Critical Survey of its Scientific Aspects*. London: Bale and Curnow.

Lockwood G and Eskes AB (1995) Relationship between cocoa variety and quality. In: Cocoa Meetings – The Various Aspects of Quality (Seminar Rencontres Cacao – les Différents Aspects de la Qualité), pp. 53–73, 159–168. Montpellier, France: AFCC/CIRAD.

Martin RA (1987) Chocolate. *Advances in Food Research* 31: 211–342.

Meursing EH (1983) *Cocoa Powders for Industrial Processing*. Koog aan de Zaan, The Netherlands: Cacao de Zaan.

Meyer B, Biehl B, Said M and Samarakoddy RJ (1989) Post harvest pod storage: a method for pulp preconditioning to impair strong nib acidification during cocoa fermentation in Malaysia. *Journal of Science and Food Agriculture* 48: 285–304.

Roelofsen PA (1958) Fermentation, drying and storage of cocoa beans. *Advances in Food Research* 8: 225–296.

Rohan TA (1963) *Processing of Raw Cocoa for the Market*. FAO Agriculture Studies no. 60. Geneva: Food and Agriculture Organization.

Wood GAR and Lass RA (1985) *Cocoa*, 4th edn. Harlow: Longman.

Ziegleder G (1985) Improved crystallization of cocoa butter under shearing. *Zeitschrift für Lebensmittel Technologie und Verfahrenstechnik* 36: 412–416.

COCONUT PALM

J J Lal and C V Sreeranjit Kumar, Centre for Development of Science and Technology, Thrissur, Kerala, India
M Indira, University of Kerala, Thiruvananthapuram, Kariavattom, India

Background

This is a graceful palm, 20–30 m in height, and native to the tropical regions of the world. Its requirement of high humidity, abundant rainfall (60 cm) and mean annual temperature of about 33 °F makes the sea coast its usual habitat. Characterized by tall slender and uniformly thick stem and massive crown with a large number of paripinnate leaves bearing bunches of nuts in their axils, the palm enhances the natural panoramic beauty of countries where it is extensively cultivated. It is a long-lived plant that may live as long as 100 years (Figure 1).

The fact that the coconut palm could be found throughout the tropics, even before the advent of European explorers, renders the tracing of its origin quite difficult. Coconut was first mentioned by an Egyptian monk, Cosmos Indicopleustes in 'Topographia Christiana' in AD 545. The most important regions of coconut production are the Philippine islands, Malaysia, India, Indonesia, and Sri Lanka.

There are several different varieties of coconut palms, including tall and dwarf. The common tall varieties are Ceylon tall, Indian tall, Jamaica tall, Malayan tall, Java tall, and Laguna. Malayan dwarf, Dwarf green, Dwarf orange, and Dwarf Fijian are some of the dwarf varieties. Tall varieties are commonly cultivated for commercial purpose. Tall varieties are slow to mature and produce the first flowers 6–10 years after planting. The life span of tall varieties is about 60–100 years. Dwarf varieties may grow to a height of 7.5–9 m, produce the first flowers after 3 years, and have a life span of about 30 years. Dwarf varieties are thought to have originated as a result of a mutation of tall types. D × T (dwarf × tall) and T × D (tall × dwarf) hybrids are early bearers and are found to be superior to their parents in terms of yield under favorable conditions. There are also several unusual types of coconut palms in different parts of the world. *Lodoicea maldivica*, a palm very closely related to coconut is known as double coconut or Maldive nut. Its nut is considered to be the largest plant seed in the world and looks like two coconuts joined together side by side. Macapuno is a Philippine variety of the coconut palm that does not have any water inside the coconut shell. Its 'meat' is a soft jelly-like substance that is used in popular Filipino sweets.

Coconut palm adapts to almost all types of well-drained tropical soils like coastal sand, red loam, laterite, alluvial, and the reclaimed soil of the marshy low land. A water table that is too high and remains static for long periods is harmful, but the palm can tolerate salinity and a wide range of pHs from 5.0 to 8.0.

Coconut palm is propagated with the help of seeds. Seeds are selected from healthy, high-yielding, and regularly bearing palms. One-year-old healthy seedlings with a minimum of six leaves are selected for planting. Nine-month-old seedlings with a minimum of four leaves can also be used for planting. The square system of planting, with a spacing of 7.5–9 m accommodates, 124–177 palms per hectare. The

Figure 1 (see color plate 37) Coconut palm.

triangular method with a spacing of 7.6 m accomodates 198 palms per hectare. In the hedge system of planting, larger number of plants can be planted in a unit area, ensuring sufficient light on each palm.

Calopagonium muconoides, *Mimosa invisa*, *Stylosanthes gracilis*, etc. are cultivated as cover crops in coconut gardens. *Crotalaria juncea*, *Tephrosia purpurea*, *Indigofera hirsuta*, etc. are grown as green manure crops in coconut gardens. Under rain-fed conditions, tuber crops (tapioca, yams, ginger, turmeric, etc.), banana, pineapple, and pulses can be grown as intercrops in adult coconut plantations without affecting the fertility of the soil and productivity of the palms. Under irrigated conditions, plants like cocoa form crop combinations with coconut. Pepper, cinnamon, clove, nutmeg, etc. are ideal for mixed cropping with coconut. Mixed farming, including the rearing of cattle on fodder grasses and legumes raised as intercrops with coconut gardens, in turn yield dung, and urine is a very remunerative form of mixed husbandry suited to farmers with small coconut holdings. It increases the fertility of the soil by recycling organic wastes, enhances the yield of palms, and generates additional income and job opportunities for the farmers.

Once the palm has attained the normal bearing stage, it continues to bear a bunch of nuts in every leaf axil at monthly intervals throughout its life extending over 50–60 years. The nuts mature 1 year after fertilization. Harvesting is done once every 45 or 60 days. Tender nuts are harvested at the age of 6–7 months for delicious soft drinks. Nuts for culinary purposes and copra are harvested at 11 months. The average yield per hectare varies from 10 000 to 14 000 nuts per annum. In Asia, experienced harvesters often pick the nuts from the tree after observing the color of the nuts, position of the bunch, quantity of water inside the nut, and texture of the husk. But in certain regions of the world, the ripe nuts are collected only after they have fallen from the tree. Coconut harvesting after picking from tall trees is a tedious operation, so various techniques such as coconut-climbing machines have been adopted in different regions.

Harvesting of immature nuts causes the production of rubbery copra with a high moisture content. At the same time, permitting nuts to fall naturally without picking the nuts from the tree may result in a 10% loss in the total harvest because of over-ripe nuts or germinated nuts. Losses resulting from nuts that are hidden or covered by thick weeds or shrubs could also range from 5 to 10% of the total harvest, if the coconuts are not harvested regularly.

The value-added products obtained from the palm and its wide utility make it one of the most useful trees in the world. Each and every part of the palm is useful to humanity in one way or the other. Hence, in Sanskrit, it is called the 'Kalpa Vriksha,' which means that it is a tree that provides all the necessities of life. It constitutes the part and parcel of all religious and cultural activities and forms a major source of income of the people in coconut-growing countries. It has a variety of uses centering around the nut, whether matured or green. Most parts of this tree are used medicinally especially in the Ayurvedic system, but the documented reports are scanty. The know-how is passed down from generation to generation by word of mouth, and its usage varies from place to place (**Figure 2**).

The palm plays an important role in the functioning of the local populations through its sustainable production of food, drink, and shelter. More than 100 products are available directly or indirectly from coconut, varying from simple cooking utensils to high-added-value products like cocochemicals and charcoal. The major cocochemicals are methylesters, fatty alcohols and glycerine.

Parts used

These include the root, trunk, leaves, inflorescence, bud, flower, fruit, and germinating nut.

Roots

Decoctions prepared from the roots are known for their astringent, diuretic, and antipyretic activities. It is used in mouthwashes and gargles, and is also used to treat uterine diseases. Roasted and ground roots can be used as a coffee substitute.

Trunk

The trunk is used for its starch. The juice that flows from the trunk when a coconut tree is cut down can be used as a drink, and the bark is used in some places for curing toothache and earache. The ash of the bark is sometimes sued as an antiseptic and is sometimes prescribed in scabies. The fibers of the trunk are recommended as a diuretic and are sometimes used in the treatment of tapeworm and inflammation of the throat.

Leaves

Coconut palm leaves are used for cattle as a fodder, after removing the midrib. The plaited leaves are used for thatching houses, fencing, religious ceremonies, and decorations, etc.

Bud

The terminal bud or growing point of the tree consists of a group of tightly packed yellow–white

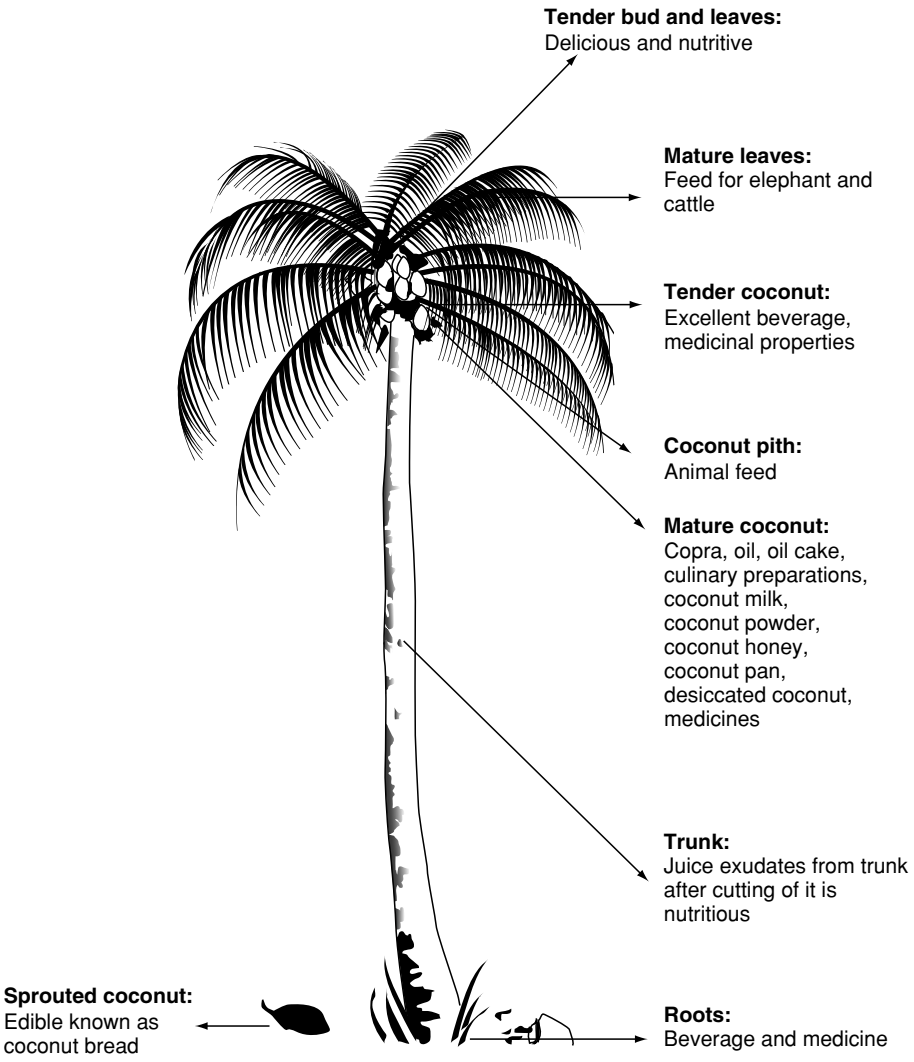

Figure 2 Coconut palm: the tree of life.

cabbage-like leaves. This forms a tasty food item known as 'millionaire's salad.' It is nourishing and digestible, and is considered to be cooling and diuretic.

Inflorescence

The inflorescence of coconut (Figure 3) is very attractive in appearance and has many medicinal properties. It is commonly used for decorations, religious functions, and several cultural and traditional rituals. It forms an ingredient of a large number of traditional medicinal preparations. Tender inflorescence as such is used for the preparation of many traditional food items, which is believed to act like a nutraceutical. It is tapped as a source of sap and is used for the production of coconut palm wine, vinegar, jaggery, and arrack.

Coconut palm wine In India and Sri Lanka, coconut palm wine is known as 'toddy,' in Philippines and Mexico as 'tuba,' and in Indonesia as 'tuwak.' It is an alcoholic beverage obtained by the natural fermentation of a sugar-containing juice from coconut palms and is obtained by tapping the unopened inflorescence of the palm. The methods of tapping the coconut palm are very ancient and vary from place to place (Figure 4). The tapping process can be summarized as follows. When the palm has reached the normal bearing stage, every leaf axils produce a spadix or inflorescence. The tapping begins when the inflorescence is 1 month old. This unopened flower inflorescence is prepared by slightly bruising it by gentle tapping with a small round piece of wood or bone (mallet). To prevent it from opening, the inflorescence is tightly bound with fiber or coconut leaves. When the inflorescence is nearly ready to produce juice, which is about 3 weeks, a length of about 5–8 cm is cut from the end. During the preparation,

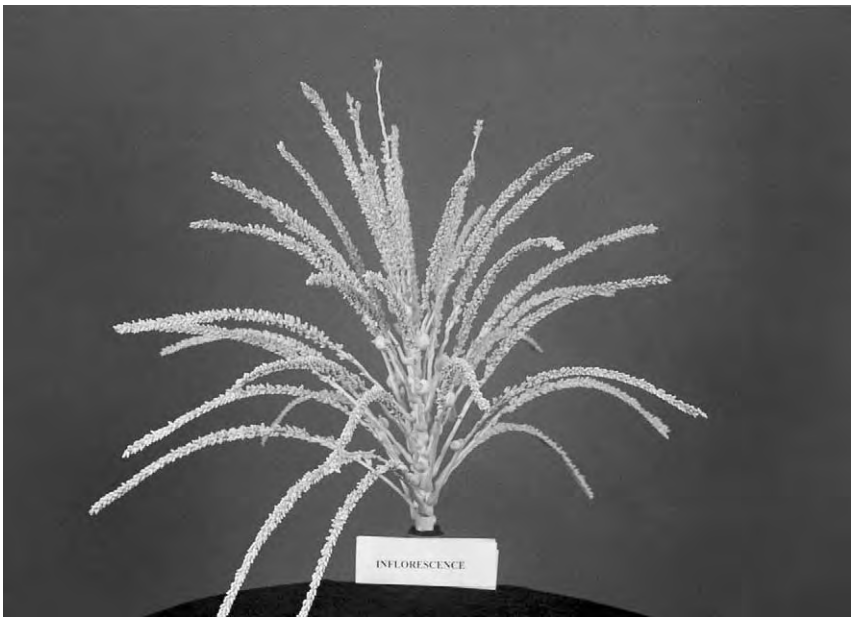

Figure 3 (see color plate 38) Inflorescence.

the inflorescence is gradually bent over, so that when the coconut palm juice flows from the vascular bundles, an earthen receptacle containing the crust of microorganisms formed from the previous fermentation is placed for its collection. The flow of juice increases gradually, and the pot is changed twice daily, at the same time shaving a thin slice from the end of the inflorescence, tapping slightly with the mallet, and smearing on a mixture of bruised leaves to stimulate the flow of palm wine from the inflorescence. When the collection of sap is over, fermentation sets in immediately, and the sap is converted into fresh coconut palm wine or sweet toddy. This product is a milky white and effervescent liquid (Figure 4a). Fermentation continues for few more hours after the sap has been pooled from various pots. The average yield of toddy per inflorescence for a tapping period of about 1 month is 18 l. Fresh coconut palm wine is sweet in taste and is considered to be a high-energy health drink. It contains a considerable amount of the natural antioxidant ascorbic acid, or vitamin C, and the amount of ethyl alcohol is 2–5% (Table 1). When fermentation proceeds, the nutritive value of fresh toddy and concentration of ascorbic acid decrease, while the concentrations of alcohol and acetic acid increase (Table 2).

Palm syrup Palm syrup is made by boiling toddy until it is thick. It can be consumed as a beverage after dilution with water. It is also a major ingredient in a number of dairy products.

Jaggery Jaggery is a brown solid substance obtained by boiling sweet toddy up to the crystallization point (Figure 4a). It is used as a sweetening agent in bakeries and is also used in the preparation of several food items and in dairy products. The yield of jaggery from sweet toddy is about 12–15%. Owing to its flavor, color, taste, and nutritive value, nowadays, it is used as a substitute for cane sugar in several developed countries.

Fermented coconut palm wine The main fermentable carbohydrate in fresh coconut palm wine is sucrose (12–15%), and a mixture of natural microflora is used for fermentation. Isolation and culturing of microflora have revealed the presence of *Streptococcus* sp., *Lactobacillus* sp., *Acetobacter* sp., etc. The main fermentation products are lactic acid, ethanol, and acetic acid. The production of lactic acid reduces the pH from around 7.2 to 5.5 within 8 h. Acetic acid is produced from ethanol by *Acetobacter*. The low pH results in high levels of ethanol, thus providing more substrate for acetic acid production, and hence prolonged standing results in the production of acetic acid from ethyl alcohol. The concentration of ethyl alcohol in fermented toddy is 5.2–5.7%.

Different types of toddies are marketed, namely, mineral toddy, ultra toddy, total toddy, oxy toddy, etc. Techniques have also been developed to can toddy. The toddy is sterilized and canned immediately, to preclude any further fermentation and thus increasing its shelf-life to approximately 8 months.

Very few research studies have been conducted on the effects of ingestion of toddy. Studies on the lipid

(a)

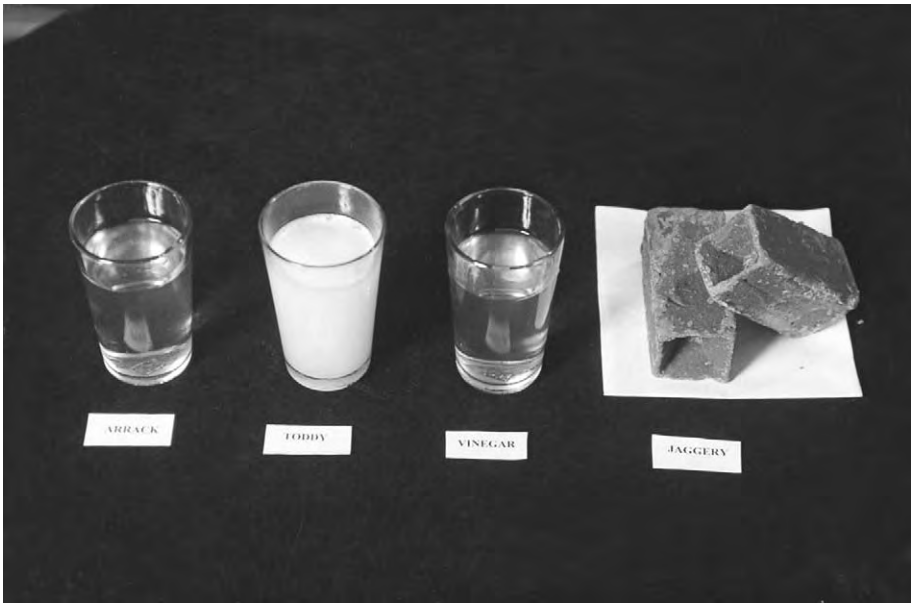

(b)

Figure 4 (a) Toddy tapping. (b) Toddy products.

and carbohydrate metabolism in female rats fed commercially available toddy in India and its equivalent quantity of ethyl alcohol during gestation and lactation have shown that the biochemical alterations produced in fetuses and pups are different. The nonethanolic portion (congeners) of the toddy interferes with metabolic processes and increases the ethyl alcohol-induced toxicity.

Toddy yeast Freshly fermented toddy is a favorable medium for the propagation of baker's yeast. Hence, toddy is used as a source of yeast in the bread-baking industry. Various traditional food preparations like 'toddy bread' or 'kallappam' (a preparation of rice paste and toddy after overnight fermentation), 'vattayappam' or round bread, etc. are very common in coconut-growing areas.

Coconut vinegar Coconut vinegar (**Figure 4a**) is produced throughout South Asia, particularly Sri Lanka. It is a clear liquid with a strong acetic acid flavor and a hint of coconut flavor. It is made from

Table 1 Composition and specific gravity of fresh coconut palm wine

Ethanol	2–5 ml per 100 ml
Ascorbic acid	16–30 mg per 100 ml (little change during fermentation)
Total solids	15.2–19.7 g per 100 ml
Sucrose	12.3–17.4 g per 100 ml
Ash	0.11–0.41 g per 100 ml
Protein (N × 6.25)	0.23–0.32 g per 100 ml
Specific gravity at 29 °C	1.058–1.077

Table 2 Composition and specific gravity of 12-h-fermented coconut palm wine

Ethanol	5–5.7%
Acidity (as acetic)	0.32–0.67%
Water	90.3–96.3%
Protein	0.2%
Ether extractives	0.1%
Carbohydrates	1.3%
Mineral matter	0.01%
Calcium	< 0.01%
Phosphorus	0.01%
Vitamin B_1	< 5 IU per 100 ml
Specific gravity at 29 °C	0.998–1.033

toddy on a commercial scale. Toddy is stored in acetifying vats, permitting free access to air. Acidification usually takes about 10–14 days. The alcohol is converted to acetic acid by acetic acid-producing bacteria (*Acetobacter*) naturally present in toddy. Vinegar usually contains 4–5% acetic acid.

Arrack Arrack (Figure 4a) is obtained by distilling fermented toddy and is common in many countries, notably in the Philippines and Sri Lanka. It is also known as coconut fenny in the Indian State of Goa. In these regions, arrack obtained from toddy is considered to be superior to the arrack from molasses and hence is recommended as a potential agent, even for postnatal care.

Flower

The male and female flowers are separate and are arranged on the same inflorescence. The spadix of a mature palm produces on average 300 female flowers. These flowers are used in many Ayurvedic preparations. Pollen grains also have medicinal properties. The flowers are astringent and are used in the treatment of diabetes, dysentery, leprosy, and urinary discharges. Opened flowers in the inflorescence provide a good source of pollen for bees.

Fruit

The fruit is 25 cm or longer in length, ovoid, and obtusely triangular in cross-section. The tough fibrous outer husk (mesocarp) encloses a spherical nut comprising a hard, bony shell (endocarp) within which is a layer of fleshy meat or kernel (endosperm) just under 1 cm in thickness (Figures 5–7). The kernel is high in oil and protein when compared with other oil seed crops. Inside the shell is the liquid endosperm of coconut or coconut water. If the nut is allowed to germinate, the cavity of the fruit is filled with a spongy mass called 'bread.' This bread is eaten raw or toasted in the shell in coconut-growing regions. Sprouting nuts are also delicious edible items. The husks and small nuts from a variety of coconut trees found in the Pacific are also eaten. Religiously, the fruit plays an important role as an offering to God.

Husk Coconut husk is widely used in horticulture, especially for growing orchids. Experiments have also proved that coir pith from green fruits can be used as feed for pigs.

Coir pith Coir pith is a waste product obtained during the process of extraction of coir fiber from the husk. It is very light, highly compressible, and highly hygroscopic. It is used as a soil conditioner, surface mulch/rooting medium and as a desiccant. Composted coir pith is an excellent organic manure for indoor plants as well as for horticulture crops.

Shell Tarry oil prepared from the shell of the nut is used externally for the treatment of ringworm. It is also used to tenderize meat while cooking. It can also be made into attractive vessels, spoons, flower vases, etc. In rubber-growing regions, it is used for collecting latex in plantations. Being a natural product, nowadays, it is considered to be the best natural alternative for food packing.

Kernel Coconut kernel is consumed in the daily diet as a culinary ingredient in several parts of the world. It is either used in the normal state or ground to a paste. Tender coconut kernel is very delicate and is a refreshing and high-energy item. Several preparations with tender coconut kernel are popular in the coconut-growing areas of the world. Raw coconut kernel and fried coconut kernel are delicious items in several parts of the world and are used for several Ayurvedic preparations.

Factors such as the variety of the palm, stages of maturity of the nut, and conditions of growth determine the composition of the kernel. The moisture, protein, and ash content of the kernel decreases as the nut matures, whereas the fat content increases.

Since the coconut kernel contains 40% oil on average, the general notion is that its consumption could elevate cholesterol levels. Besides coconut oil, the

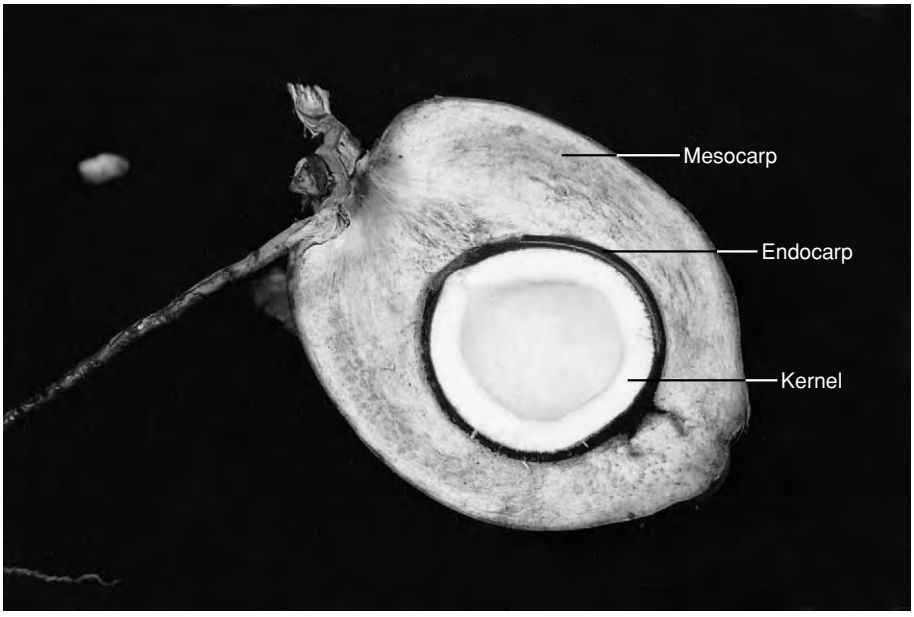

Figure 5 Vertical section: coconut.

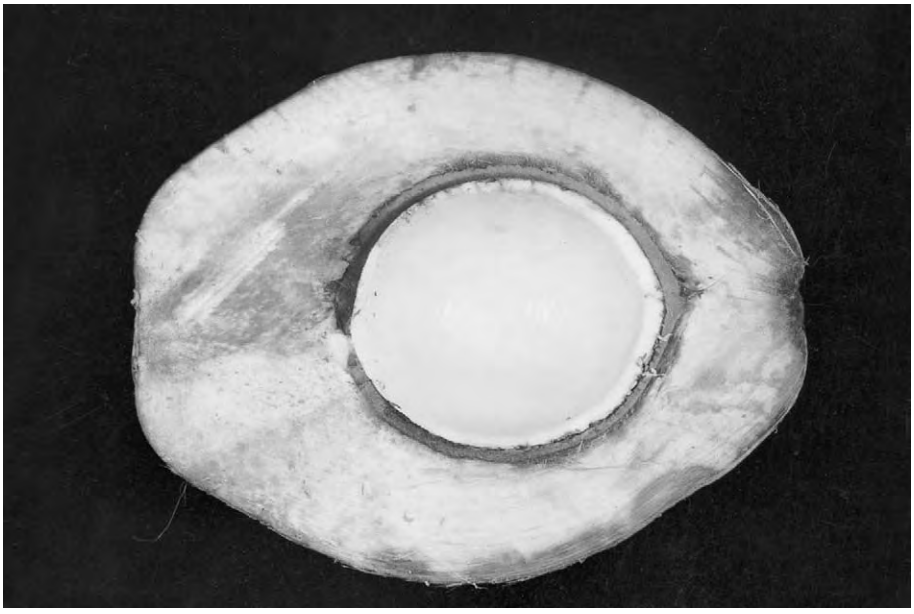

Figure 6 Vertical section: tender coconut.

kernel contains 3.6% dietary fiber (Table 3). Research studies undertaken in rats at the University of Kerala have strongly indicated a significant hypolipidemic effect of the neutral detergent fiber isolated from coconut. This effect is attributed to the presence of a hemicellulose component. Proteins isolated from coconut kernel have also been found to have a lipid-lowering effect in experimental animals. This beneficial effect of the kernel protein is attributed to its very low lysine/arginine ratio (2.13% lysine and 24.5% arginine). Recent reports have shown that feeding coconut kernel along with coconut oil in human volunteers reduced the total serum cholesterol and low-density lipoprotein-C (LDL-C) when compared with feeding coconut oil alone. Neutral detergent fiber extracted from coconut has been shown to decrease the levels of lipids and lipid peroxides in rats fed a fat-rich diet.

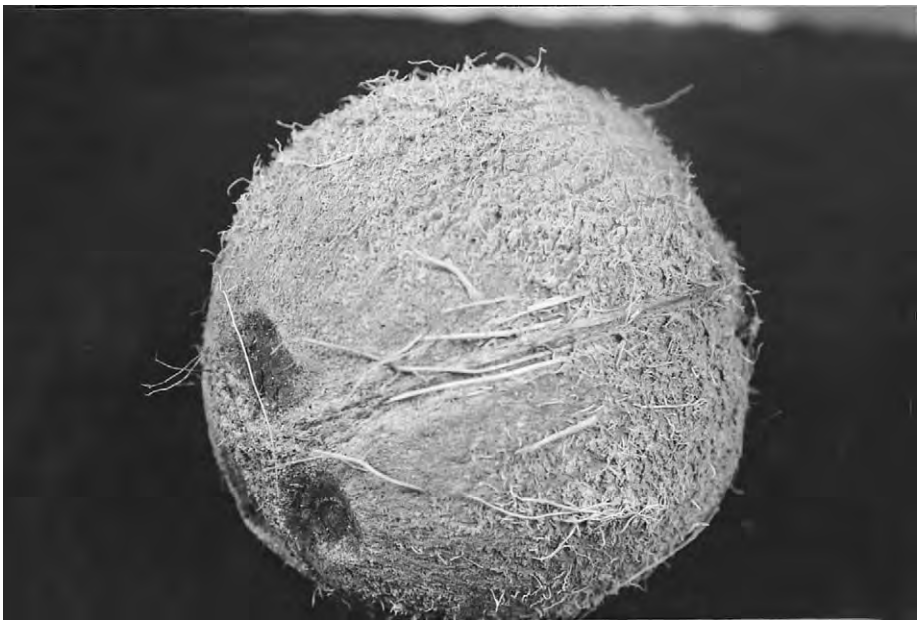

Figure 7 Dehusked coconut.

Table 3 Composition of fresh coconut kernel

Water	36.3 ml per 100 g
Protein	4.5 g per 100 g
Fat	41.6 g per 100 g
Carbohydrates	13.0 g per 100 g
Fiber	3.6 g per 100 g
Minerals	1.0 g per 100 g
Among the minerals present:	
Ca	10 mg per 100 g
P	24 mg per 100 g
Fe	1.7 mg per 100 g

Desiccated coconut Desiccated coconut is the disintegrated and dehydrated coconut kernel or pulp largely comprising the white portion of the kernel, after removing the brown testa or parings. It is referred to as 'farane de coco' in France. The process of preparation involves initial sterilization by washing in boiling water, followed by disintegrating or shredding and ultimately by drying at 60–75 °C to a moisture level of less than 2.5%. Desiccated coconut was first manufactured from imported nuts in England and in the USA in the early 1880s. It is pure white in color, crisp, and has a fresh taste. It contains 68–72% oil and has a nitrogen content of around 1.05%. The chief protein is globulin, the sugar present is mainly sucrose (6–7%), and the crude fiber amounts to 4%.

Coconut oil is important in the confectionery and bakery industry. It is mainly used in biscuits, confectionery, bakery, and other food industries, and also for household purposes in countries where coconut is not grown.

Byproducts of manufacture of desiccated coconut (brown testa or parings) The parings, which, under normal working conditions, amount to 12–15% of the kernel, are dried and pressed for oil (yield 55%). Parings oil differs somewhat in composition from ordinary coconut oil. Its iodine value ranges from 19 to 25, its saponification value ranges from 237 to 244, and its acid value is higher than that of commercial coconut oil. Parings oil is mostly used for soapmaking.

Coconut milk Coconut milk is the cold aqueous extract of coconut kernel (Figure 8) and is used in cooking. Techniques have been developed to preserve coconut milk, and so now it is available commercially. Various techniques have also been developed to manufacture coconut cream and spray-dried coconut milk powder.

Coconut milk after bottling can be used as a substitute for cows' milk. In Philippines, sugar and citric acid are added to coconut milk to produce coconut syrup, which is used in the confectionery. Coconut honey is also prepared from coconut milk after the addition of jaggery and glucose. This has a nutty flavor and can be used as a substitute for honey.

Copra Dried or dehydrated coconut is called copra (Figure 8). On average, 1000 nuts will produce about 180 kg of copra. Copra is used as a culinary ingredient and also for the production of oil. Copra processing is a widespread activity involving dehusking, splitting the nuts into two equal halves, and sun

Figure 8 Coconut products.

drying. Sun drying takes about 4–7 days, depending on the sunshine, to dry copra to the desired moisture level (below 6%). Improved drying methods to reduce the duration of drying include smoke drying and hot-air drying. Smoke drying, requiring a drying time of 36 h, and hot-air drying reduce the contamination of copra and fungal attack, thus improving the quality of the oil.

Feeding of copra to pigs and poultry has produced good results, but it is an expensive animal feed. As the fat in copra contains only small amounts of unsaturated fatty acids, its consumption leads to firm body fat and good flavor.

Table 4 Fatty acid profile of coconut oil

Common name	*Fatty acid*	*Percentage*
Caproic acid	6:0	0.2–0.5
Caprylic acid	8:0	5.4–9.5
Capric acid	10:0	4.5–9.7
Lauric acid	12:0	44.1–51
Myristic acid	14:0	13.1–18.5
Palmitic acid	16:0	7.5–10.5
Stearic acid	18:0	1.0–3.2
Arachidic acid	20:0	0.2–1.5
Oleic acid	18:1*n*-9	5.0–8.2
Linoleic acid	18:2*n*-6	1.0–2.6

Coconut oil Coconut oil (Figure 8) constitutes the major commercial product of coconut. Coconut oil is used as a cooking fat, hair oil, body oil, and industrial oil. Refined coconut oil is prepared exclusively for industrial purposes and is widely used in the manufacture of biscuits, chocolates, icecreams, margarine, and confectionery items. It is also used for the manufacture of paints and pharmaceutical agents. Desirable properties such as a low melting point, resistance to rancidity, pleasant flavor, and easy digestibility make it an ideal ingredient in the food industry. Coconut oil is a source of many oleo chemicals such as fatty acids, methyl esters, and fatty alcohol. For cooking and toiletry purposes, it is commonly used in the form of filtered coconut oil. Virgin coconut oil, which is a high-quality oil, is prepared from the milk extracted from the raw kernel. This type of coconut oil is most suitable as a massage oil for babies.

Coconut oil is a rich source of saturated fatty acids, and short- and medium-chain fatty acids account for 70% of these fatty acids (Table 4). It has a low content of unsaturated fatty acids with a negligible content of both *n*:6 and *n*:3 polyunsaturated fatty acids and a low *n*:6/*n*:3 ratio (< 4). The highly resistant nature of coconut oil to oxidative rancidity is attributed to its high concentration of saturated fat and low unsaturated fatty acids. This quality makes it suitable for storage without deterioration.

Copra, which is obtained by drying coconuts, is the source of coconut oil. Power-driven rotaries and expellers are used for extracting oil from copra. This oil extraction is immediately followed by the separation of cake residue and mucilage by filtering or by settling.

Because of the unique qualities of lauric acid (C:12) present, coconut oil is widely used in soaps and

cosmetic manufacturing industries. Lauric acid is known to possess antiviral, antibacterial, and antiprotozoal qualities. It is converted to the monoglyceride monolaurin in the human or animal body. Monolaurin is antiviral, antibacterial, and antiprotozoal. Reports indicate that monolaurin is capable of destroying lipid-coated viruses such as HIV, herpes, cytomegalovirus, influenza, various pathological bacteria, including *Listeria monocytogenes* and *Helicobactor pylori*, and protozoa such as *Giardia lamblia*. It is synthesized in babies from the lauric acid of mother's milk. Capric acid, another fatty acid found in coconut, also has antimicrobial activities.

Tocopherols are the natural antioxidants present in coconut oil. The volatile flavor constituent of crude coconut oil includes ketones, lactones and δ-lactones of which δL_8 to δC_{10} with undecan-2-1 as the major component at 290 p.p.m. and δ-decalactone as the major lactone component at 97 p.p.m. The flavor and aroma of coconut oil are attributed to δ-octalactone. Ketones are derived from the microbiological dissociation of fatty acids. The digestibility coefficient of coconut oil is higher (with 91.0% assimilable glycerides) than any other fat, including butter, and so it is digested more rapidly than any other fats. This easy digestibility makes it an essential ingredient for many ghee substitutes.

The consumption of saturated oil could hasten the onset of cardiac problems, as suggested by certain research studies. Coconut oil, being a saturated oil, caused concern that adversely affected the prospects of the coconut industry. The major fatty acids of coconut oil are medium-chain fatty acids. A shorter chain length allows fatty acids to be metabolized without the use of a carnitine transport system. Since the short- and medium-chain fatty acids can be rapidly oxidized, they are less conducive to fat deposits compared with long-chain fatty acids. Early studies conducted in experimental animals fed a synthetic diet containing coconut oil as the source of fat have shown it to be atherogenic, since it is deficient in essential fatty acids. Essential fatty acid deficiency is known to facilitate the development of atherosclerosis. But under normal conditions, the possibility of essential fatty acid deficiency as such is quite remote, since their presence in other food items will offset any deficiency in coconut oil. Thus, the increase in lipogenesis observed in earlier studies was due to the faulty design of the experiments. Feeding coconut oil at normal levels along with other fats adequately supplemented with linoleic acid renders coconut oil neutral in terms of atherogenecity. Epidemiological studies also support this. The University of Kerala conducted a study in 64 volunteers and found no statistically significant alterations in the serum total cholesterol, high-density lipoprotein (HDL) cholesterol, LDL cholesterol, HDL-C/total cholesterol ratio, LDL-C/total cholesterol ratio, and triglycerides from the baseline values. Feeding coconut oil results in an increase in HDL cholesterol.

Recent studies have shown that the presence of natural coconut fat in the diet leads to a normalization of body lipids, protects alcohol damage to the liver, and improves the immune system's antiinflammatory response.

Coconut oil is needed for the good absorption of fat and calcium from infant formulas. Hence, it has been recommended in infant formulas.

Coconut Cake

This is the residue left after the extraction of oil from copra and is widely used as cattle feed (Figure 8). The residual oil in the cake can be extracted by the process of solvent extraction, and the resulting defatted cake, containing less than 1% oil, is used in cattle-feed manufacture. The extracted residual oil is largely used for industrial purposes.

Uses of coconut cake Coconut cake, which is commonly used as a cattle and poultry feed, is soaked in water prior to feeding. Since it imparts a firm texture to butter and provides harder body fat, it is useful for dairy and fattening cattle, respectively. Field and laboratory investigations have shown that coconut cake can be used as a fertilizer (Table 5). Coconut cake is rather rich in fiber, and hence, its inclusion in pig diets is restricted. The maximum safe quantity of coconut cake with dairy cows is 1.5– 2 kg daily; feeding larger quantities may result in tallowy butter. Beef cattle can consume much more without any impairment in carcass quality.

Coconut flour This is essentially coconut cake in a more attractive and edible form. It is prepared by extracting oil from desiccated coconut, drying, and grinding the residue to a suitable degree of fitness. Its

Table 5 Composition of coconut cake after removing oil by the expeller method and solvent-extraction method

Composition	*Method of extraction*	
	Expeller (%)	*Solvent extraction (%)*
Moisture	7.0	8.9
Fat	6.7	2.4
Protein (N × 6.25)	21.2	21.4
Nitrogen-free Extract	47.4	47.4
Fiber	11.2	13.3
Mineral matter	6.5	6.6

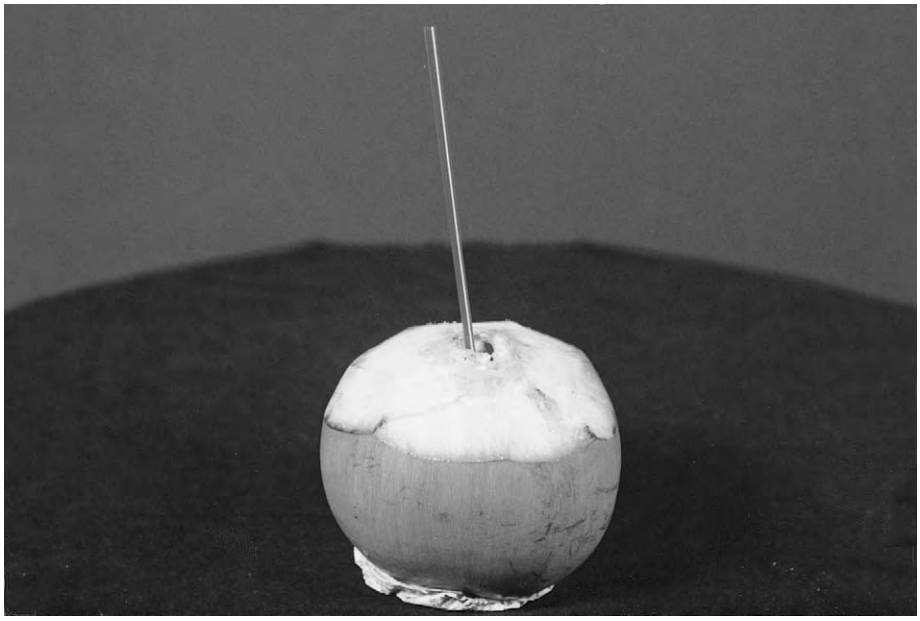

Figure 9 Tender coconut.

keeping qualities are not good, and it may be expensive.

Coconut water This is the liquid endosperm of coconut (Figure 9). The composition of coconut water changes during the course of ripening of the fruits and in germination. The concentration of total solids during the initial days is about 2.5 g per 100 ml. This increases gradually as the nut ripens, reaches a maximum of about 6 g per 100 ml at about the seventh month, and then declines. During germination, water completely disappears.

Sugars are the most important constituent. During the early days, only inverted sugars are present, and their concentration in the coconut water increases to a maximum by the fifth and sixth month. Thereafter, the level of sucrose rises, though the concentration of the total sugars declines rapidly, until its content reaches 2 g per 100 ml approximately.

Tender coconut water is commonly consumed as a beverage. It is enriched with various minerals such as potassium, phosphorus, magnesium, sulfur, chlorine, sodium, calcium, copper, iron, etc. Various proteins and amino acids such as arginine, cystine, alanine, etc. and the B vitamins are also present. Moreover, it is recommended as a useful alternative to glucose saline in cases of gastroenteritis and urinary diseases. It has a caloric value of 0.073 kJ per 100 g, potassium constitutes the principal inorganic constituent, and malic acid is the predominant organic acid (Table 6).

Experimental studies have indicated the presence of certain unidentified substances or combination of substances in coconut water that are capable of inducing rapid proliferative changes in plant tissues. Nate-de-coco is a gelatinous delicacy formed by the action of the microorganism *Acetobactor xylinium* in the culture medium of coconut water. Coconut water is also an excellent medium for the production of protein foods using a culture of the food yeast *Saccharomyces fragilis*.

A patent granted by the UK to the FAO involving a new technology permits manufacturers to bottle biologically pure and tasty coconut water that is rich in salts, sugars, and vitamins, as used by athletics. This technology employs a cold sterilization processing method that prevents the loss of flavor and other nutritional characteristics. It is also used in medicinal preparations. In the Ayurvedic ointment 'ilaneer-kuzhmbu' (from the Sanskrit terms 'ilaneer,' meaning tender coconut water, and 'kuzhambu,' meaning

Table 6 Composition of ascorbic acid and B vitamins in coconut water

Vitamin	*Amount*
Ascorbic acid	2.2–3.7 mg per 100 ml
Nicotinic acid	0.64 μg ml^{-1}
Pantothenic acid	0.52 μg ml^{-1}
Biotin	0.02 μg ml^{-1}
Riboflavin	< 0.01 μg ml^{-1}
Folic acid	0.003 μg ml^{-1}
Thiamine	Trace
Pyridoxine	Trace

paste) for the eye, coconut water is used as the principal ingredient.

Coconut water increases the blood circulation in the kidneys and causes profuse diuresis. It is also a urinary antiseptic and eliminates poisons through the kidneys in case of mineral poisoning.

Coconut water vinegar Despite its low sugar content, coconut water is considered to be a good base for vinegar. For this, coconut water is allowed to ferment for 7 days after adding sugar, as a result of which sugar is converted to alcohol.

Heart of the Palm

In some places, the heart of the palm is used raw or boiled for use in salads.

See also: **Alcohol**: Properties and Determination; **Amino Acids**: Properties and Occurrence; **Antioxidants**: Synthetic Antioxidants; **Biotin**: Properties and Determination; **Fats**: Digestion, Absorption, and Transport; **Fatty Acids**: Dietary Importance; **HIV Disease and Nutrition**; **Protein**: Food Sources; **Sugar Alcohols**; **Wines**: Production of Table Wines; **Yeasts**

Further Reading

Child R (1974) *Coconut, Tropical Agriculture Series*. London: Longman.

Nair MK, Khan HH, Gopalasundaram P and Rao EVVB (1993) *Advances in Coconut Research and Development*. Oxford: Oxford & IBH Publishing.

Parsley GJ (1992) Replanting the Tree of Life: Towards an International Agenda for Coconut Palm Research. Wallingford, UK: CAB International.

Silas EG, Aravindakshan M and Jose AI (1991) *Coconut Breeding and Management*. Vellanikkara, India: Kerala Agricultural University.

Thampan PK (1981) Handbook on Coconut Palm, 3rd edn. Oxford: Oxford and IBH Publishing.

Cod *See* **Fish**: Introduction; Catching and Handling; Fish as Food; Demersal Species of Temperate Climates; Pelagic Species of Temperate Climates; Tuna and Tuna-like Fish of Tropical Climates; Pelagic Species of Tropical Climates; Demersal Species of Tropical Climates; Important Elasmobranch Species; Processing; Miscellaneous Fish Products; Spoilage of Seafood; Dietary Importance of Fish and Shellfish; **Fish Farming**; **Fish Meal**

Codex *See* **Legislation**: History; International Standards; Additives; Contaminants and Adulterants; Codex

COENZYMES

J H Freeland-Graves, University of Texas at Austin, Austin, TX, USA
C Bavik, University of Sheffield, Sheffield, South Yorkshire, UK

Background

Life as we know it is made possible by enzymes, highly specific proteins that facilitate biochemical reactions. The term holoenzyme refers to an active enzyme complex. An apoenzyme is the protein portion of the active unit. The term prosthetic group is used to refer to minerals, activated vitamins or other nonprotein compounds that are required for full enzyme activity. Cofactors are nonprotein substances, typically mineral ions and activated vitamins that are required for the function of certain enzymes and other regulatory proteins. Some cofactors are essential at the active site of a reaction, while others help maintain the structural integrity of an enzyme or protein. Coenzymes are activated vitamins that participate in reactions with a stoichiometry equal to substrate. (*See* **Enzymes**: Functions and Characteristics.)

An overview of how vitamins and minerals participate as cofactors in the overall processes of

metabolism is illustrated in Figure 1. Cofactors are essential in numerous biochemical pathways, including the breakdown, or catabolism, of nutrients and the synthesis, or anabolism, of biological compounds. The vitamin and mineral cofactors complex with enzymes to convert nutrients into usable energy and produce biomolecules that are the basis of life. (*See* **Minerals – Dietary Importance**; **Trace Elements**; **Vitamins**: Overview.)

Nutrients as Coenzymes and Cofactors

Without the required vitamins and minerals, cofactor-dependent enzymes could not mediate metabolism or maintain normal cell function and biological processes that are essential for cell division, differentiation, growth, and repair. Nutrient cofactors are also necessary for the structural integrity of certain hormones and regulatory proteins.

Vitamins

All of the water-soluble vitamins and two of the fat-soluble vitamins, A and K, function as cofactors or coenzymes. Coenzymes participate in numerous biochemical reactions involving energy release or catabolism, as well as the accompanying anabolic reactions (Figure 1). In addition, vitamin cofactors are critical for processes involved in proper vision, blood coagulation, hormone production, and the integrity of collagen, a protein found in bones. (*See* **Retinol**: Physiology.)

The active coenzyme form of thiamin, vitamin B_1, is thiamin pyrophosphate (TPP) (Figure 2a). TPP is involved in oxidative decarboxylation and transketolase reactions. An example is the decarboxylation (removal of —COO^{-}) of three-carbon pyruvate to two-carbon acetyl coenzyme A (CoA), an important step in carbohydrate breakdown. (*See* **Thiamin**: Physiology.)

The active forms of riboflavin, vitamin B_2, are the coenzymes flavin mononucleotide (FMN; Figure 2b) and flavin adenine dinucleotide (FAD). These coenzymes serve as hydrogen carriers for oxidation reactions that affect energy nutrients in the citric acid cycle and in the electron transport system. (*See* **Riboflavin**: Physiology.)

The coenzyme forms of nicotinic acid are nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP). These compounds assist dehydrogenase enzymes in the catabolism of fat, carbohydrates, and amino acids, and in the enzymes involved in synthesis of fats and steroids and other vital metabolites. (*See* **Niacin**: Physiology.)

Pyridoxal phosphate (PLP; Figure 2c) and pyridoxamine phosphate (PMP) are the coenzyme forms of vitamin B_6. These are cofactors for approximately 120 enzymes, such as the transaminases, racemases, decarboxylases, cleavage enzymes, synthetases, dehydratases, and desulfydrases. Both PLP and PMP participate in the metabolism of amino acids, including transamination, racemization, deamination, and desulfhydration, and the conversion of tryptophan to nicotinic acid. (*See* **Vitamin B_6**: Physiology.)

Pantothenic acid (PA) is a B vitamin that is a component of coenzyme A (Figure 2d). Coenzyme A is necessary for the metabolism of carbohydrates, amino acids, fatty acids, and other biomolecules. As a cofactor of the acyl carrier protein, pantothenic acid participates in the synthesis of fatty acids. (*See* **Osteoporosis**.)

The coenzyme forms of vitamin B_{12} are methylcobalamin (Figure 2e) and deoxyadenosylcobalamin. These assist in the conversion of homocysteine to the amino acid methionine, the oxidation of amino acids and odd-chain fatty acids, and the removal of a methyl group from methyl folate, which regenerates tetrahydrofolate. (*See* **Cobalamins**: Physiology.)

Biotin as the coenzyme biocytin functions in carboxylation reactions that convert odd-carbon-numbered amino acids and fatty acids to even-carbon-numbered compounds, which can then be metabolized. Biocytin is also necessary for the synthesis of pyrimidines and the formation of urea. Some holoenzymes containing biotin act as carboxylases to convert acetyl CoA to cholesterol precursors, and as transcarboxylases and decarboxylases in other important reactions. (*See* **Biotin**: Physiology.)

A coenzyme of folate is tetrahydrofolate (THF), a carrier of one-carbon units, such as methyl groups (—CH_3). One-carbon units arise primarily from the metabolism of amino acids. They are needed to interconvert amino acids and to synthesize purines and pyrimidines for the formation of RNA and DNA. (*See* **Nucleic Acids**: Physiology.)

Vitamin C (ascorbic acid) is a cofactor for the hydroxylases. Some examples are the hydroxylation of proline and lysine to create cross-links from intramolecular hydrogen bonds that are critical to the structural integrity of collagen, the hydroxylation of cholesterol to form bile acids, and the hydroxylation of tyrosine to form the hormone norepinephrine (noradrenaline). (*See* **Ascorbic Acid**: Physiology.)

The aldehyde form of vitamin A, retinal, is a cofactor for apoproteins in the eye called opsins. Opsins are responsible for dim-light vision in the rods (rhodopsin) and are involved in color and bright-light vision in the cone of the retina (iodopsin). When light strikes the retinal bound to opsin, the conformation of the retinal is changed (photoisomerization) such that photoreceptor cell membranes are

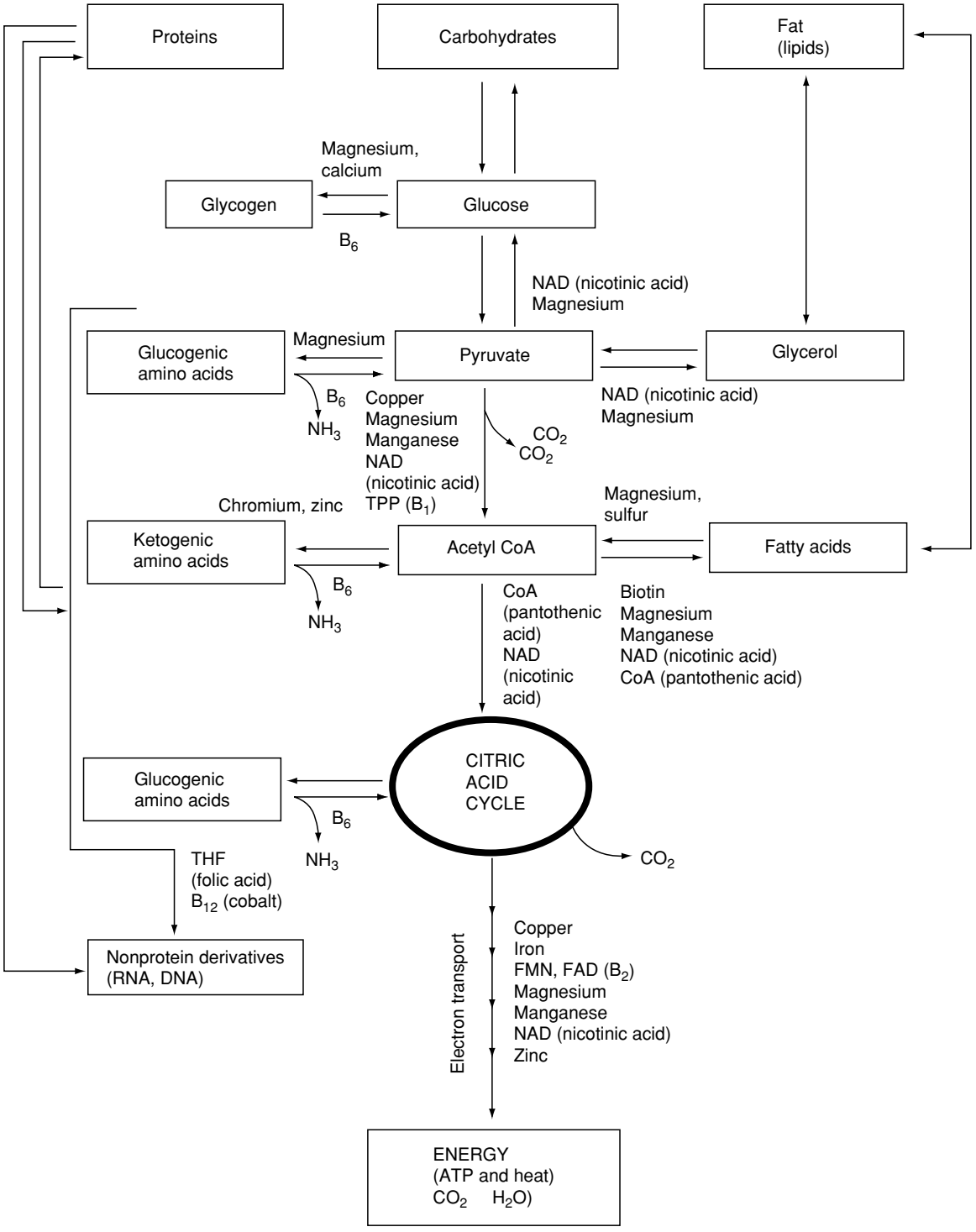

Figure 1 Overview of carbohydrate, protein, and lipid (fat) metabolism. Vitamins and minerals play crucial roles as coenzymes and cofactors in both the energy-releasing, catabolic pathways and the anabolic pathways involved in the synthesis of proteins, lipids, carbohydrates, and nucleic acids. NAD, nicotinamide adenine dinucleotide; TPP, thiamin pyrophosphate; CoA, coenzyme A; THF, tetrahydrofolate; FMN, flavin mononucleotide; FAD, flavin adenine dinucleotide; ATP, adenosine triphosphate. Reproduced from Coenzymes, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK, and Sadler MJ (eds), 1993, Academic Press.

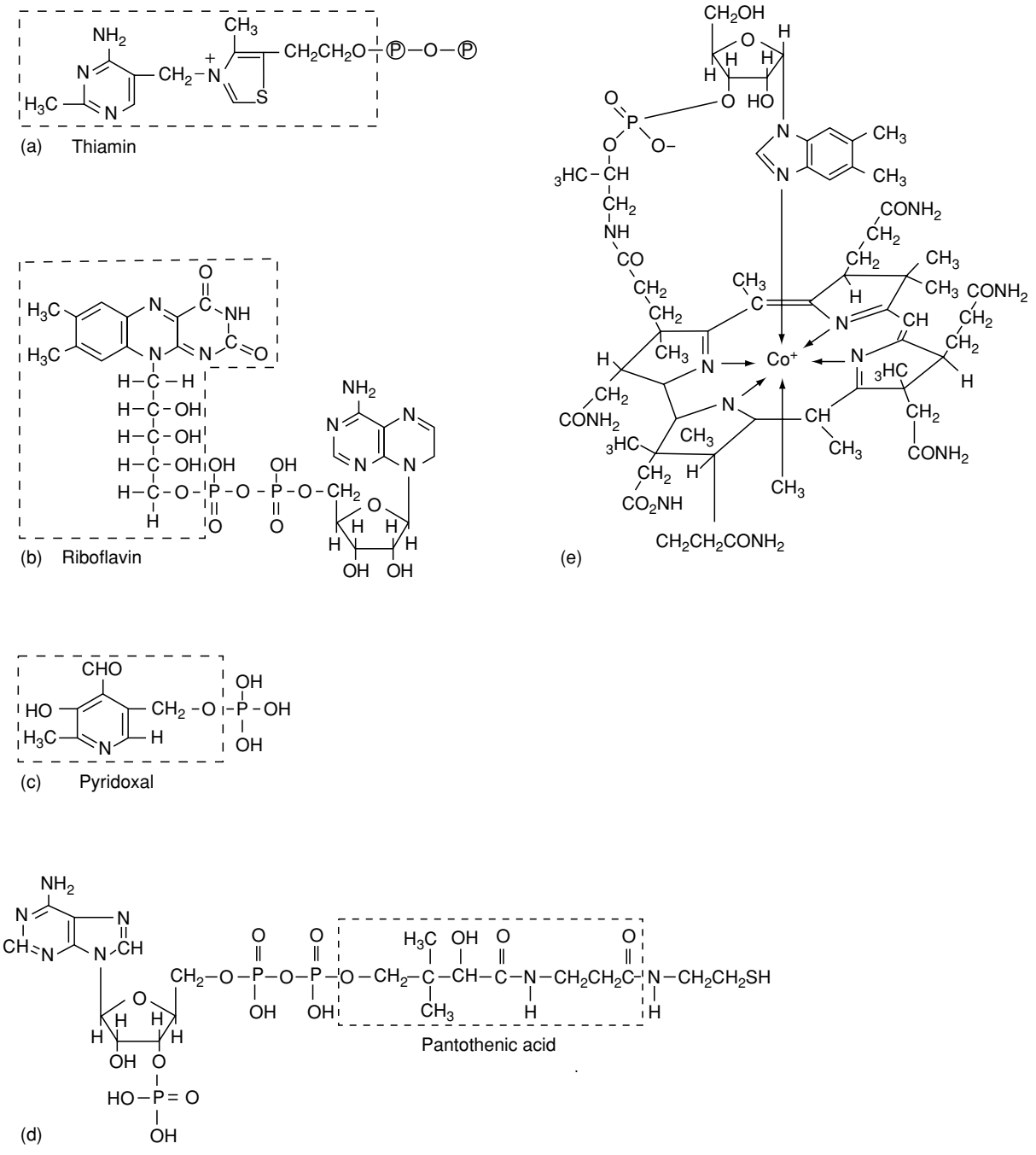

Figure 2 Selected examples of vitamins as coenzymes: (a) thiamin pyrophosphate; (b) flavin mononucleotide; (c) pyridoxal phosphate; (d) coenzyme A; and (e) methylcobalamin or coenzyme B_{12}. Reproduced from Coenzymes, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK, and Sadler MJ (eds), 1993, Academic Press.

hyperpolarized and the optic nerve transmits signals to the brain interpreted as vision. Retinoic acid is the metabolite form of vitamin A that regulates genes. It binds to proteins called retinoic acid receptors (RARs) and retinoid X receptors (RXRs). These proteins are transcription factors belonging to the steroid/thyroid hormone receptor superfamily of proteins and are found throughout the body. The RAR/RXR proteins bind to and regulate the transcription of numerous target genes important for cell development.

Vitamin K acts as a coenzyme for γ-carboxylases, enzymes that transfer —CO_2 groups. The resulting carboxylic acid groups are available for calcium

binding. Gamma-carboxylation is necessary for the formation of osteocalcin, a protein important in bone remodeling, and prothrombin, a coagulation factor (II) involved in blood clotting (*See* **Vitamin K**: Physiology).

Minerals

Minerals participate as both catalysts and cofactors in biological reactions. As catalysts, minerals are not part of an enzyme or substrate, but accelerate the reaction between the two. As cofactors, they become a structural component that is essential for the function of an enzyme or protein. Minerals that play critical roles as cofactors for enzymes include magnesium, manganese, molybdenum, and selenium. Other minerals, such as calcium, cobalt, phosphorus, and iodine, act as essential cofactors for nonenzymatic proteins. Zinc, copper, and iron are cofactors for both enzymatic and nonenzymatic proteins.

Mineral Cofactors in Enzymatic Reactions

Some examples of how minerals serve as cofactors for enzymes involved in metabolism are as follows.

Magnesium (Mg) is required as a cofactor for over 300 enzyme reactions. One critical function is the stabilization of the structure of adenosine triphosphate (ATP). Energy provided by magnesium-dependent ATP hydrolysis is required during the catabolism of carbohydrates (glycolysis and the citric acid cycle) and fatty acids (β-oxidation) and the anabolism of proteins. Magnesium also plays a cofactor role in enzymes involved in the synthesis of DNA, and it helps maintain the double helical structure of DNA. (*See* **Magnesium**.)

Manganese (Mn) has been identified as an essential cofactor in several metalloenzymes. Some examples are: (1) superoxide dismutase, a mitochondrial enzyme which catalyzes the breakdown of superoxide free radicals to hydrogen peroxide and water, thereby protecting cells from free radical damage; (2) arginase, which helps in the production of nitric oxide and the urea cycle; and (3) phosphoenol-pyruvate carboxykinase, which participates in carbohydrate metabolism. Manganese is also important (but not essential) in activating the glycosyltransferases, which are necessary for the formation of glycoproteins and proline depeptidase, which catalyzes the final step in the breakdown of collagen.

Molybdenum (Mb) is a cofactor for several oxidation enzymes. Xanthine oxidase is necessary for the production of uric acid from purines; sulfite oxidase converts sulfite to sulfate; and aldehyde oxidase is involved in the hydroxylation of heterocyclic nitrogen compounds, such as nicotinic acid.

Selenium (Se) functions as a component of enzymes involved as antioxidants (glutathione peroxidase) and in thyroid hormone metabolism (5′-deiodinases). Although the metal is needed for activity, it is not a cofactor since it is incorporated in protein as the amino acid selenocysteine. (*See* **Selenium**: Physiology.)

Zinc (Zn) is an essential component for more than 100 enzymes. Examples of zinc-containing enzymes are found in all known classes of enzymes, including transferases, hydrolases, oxidoreductases, lyases, isomerases, and ligases. Zinc is a cofactor in key biochemical reactions in the body, such as carbohydrate, lipid, and protein metabolism, stabilization of membranes, and synthesis and catabolism of DNA and RNA. Consequently, zinc is an important contributor to the processes of replication (synthesis of new DNA), transcription (synthesis of messenger RNA), and translation (synthesis of enzymatic and nonenzymatic proteins). Some of the general biological functions depending upon the cofactor functions of zinc include cell replication, tissue growth and repair, bone formation, skin integrity, and cell-mediated immunity. (*See* **Zinc**: Physiology.)

Mineral Cofactors in Nonenzymatic Molecules

Minerals also serve as integral structural components of a variety of nonenzymatic proteins, as well as for hormones and vitamin B_{12}.

Calcium (Ca) functions as a cofactor when it forms a complex with two structurally related proteins – calmodulin and troponin C. Calmodulin is a protein with two globular lobes, each having two binding sites for calcium (Figure 3a). When calcium ions bind to calmodulin, a variety of calcium-dependent enzymes are activated, including membrane phosphorylase kinases and some forms of cyclic nucleotide phosphodiesterases and adenylate cyclases. These enzymes change the three-dimensional conformation of the target protein and influence the activity of signaling pathways whereby cell surface receptors transmit extracellular signals into cellular responses. For example, the hormones epinephrine (adrenaline) and glucagon react with their respective cell surface receptors to signal the cell to utilize glycogen (the storage form of glucose). These 'signals' induce phosphorylation of the enzymes involved in the synthesis (glycogen synthase) and degradation (glycogen phosphorylase) of glycogen.

Troponin C is a muscle protein that is structurally similar to calmodulin. When this protein is activated by calcium binding, it enhances interactions between actin and myosin, proteins involved in muscle contraction. (*See* **Calcium**: Physiology.)

Cobalt (Co) is a central atom in the structure of vitamin B_{12} (Figure 2e). This vitamin is essential for

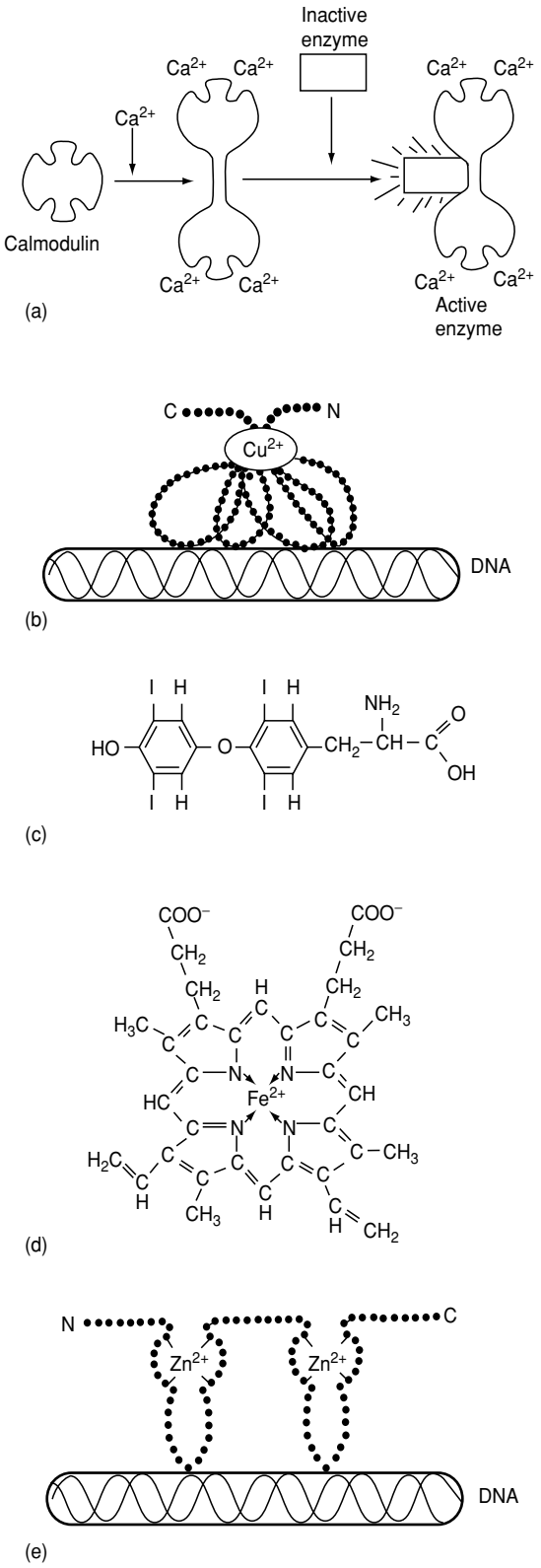

Figure 3 Selected examples of minerals as cofactors: (a) calmodulin; (b) copper-fist motif; (c) thyroxine or T_4; (d) heme; and (e) zinc-finger motif. Reproduced from Coenzymes, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

carbon transfer reactions involved in the synthesis of DNA and regeneration of methionine. (*See* **Cobalt**.)

Copper (Cu) ions play a critical role in the structure of some transcription-regulating proteins. In the presence of copper ions, certain transcription factors acquire a loop-like structure which forms a cluster close around the copper ions (Figure 3b). This complex is called a 'copper fist' since it appears to be similar to a fist clutching a small object. The loop 'knuckles' of the fist are thought to bind to a regulatory region (promoter) of the metallothionein gene. Once the transcription factor–copper complex (copper fist) is bound to the promoter region, another part of the transcription factor stimulates gene transcription. The translated metallothionein protein regulates copper levels and prevents toxicity. (*See* **Copper**: Physiology.)

Iodine (I) forms part of the thyroid hormones, thyroxine (Figure 3c), and thyronine. Both hormones help regulate the basal metabolic rate of organisms. (*See* **Iodine**: Physiology.)

Iron (Fe) is a critical constituent of heme (Figure 3d) which forms part of the hemoglobin and myoglobin molecules. Hemoglobin transports oxygen to and carbon dioxide away from cells in the body; myoglobin stores oxygen in muscles. (*See* **Iron**: Physiology.)

Phosphorus (P) forms part of the energy-storage compound ATP. The removal of phosphate (dephosphorylation) from ATP to form adenosine diphosphate (ADP) releases considerable biochemical energy. Adding the phosphate (phosphorylation) to ADP to form ATP again permits the body to store energy. The storage and release of energy via interconversions of ATP and ADP is one way the major components of food (carbohydrates, fats, and protein) ultimately provide energy for the body. Other types of reversible phosphorylation help regulate the conformation and activity of certain proteins such as enzymes.

Zinc ions are also needed for the structure of some transcription factors. The transcription of certain genes is regulated by DNA-binding proteins that contain important functional domains characterized as 'zinc fingers' (Figure 3e). Zinc stabilizes the folding of the transcription factor into a 'finger loop' which is capable of site-specific binding to double-stranded DNA. Zinc finger loops are present in the DNA-binding domains of receptors for glucocorticoids, mineralocorticoids, estrogen, progesterone, thyroid, 1,25-dihyroxy-vitamin D_3, and retinoic acid.

Effect of Nutrient Deficiencies

A primary deficiency of an essential vitamin or mineral cofactor is caused by inadequate amounts of the nutrient in the diet. A secondary deficiency occurs because

of something other than diet, such as a disease state or metabolic alteration, which leads to decreased absorption, impaired transportation, or increased requirements, utilization, or excretion. If a deficiency of a nutrient continues, body stores begin to diminish.

A continued insufficiency of a nutrient cofactor will impair or inhibit biochemical functions of the dependent enzyme. The lack of a functional enzyme–cofactor complex produces an accumulation of the substrate and a deficit of the enzyme product. At this point, measurements of biochemical parameters may indicate a problem that is not yet evident by a physical examination. This condition is known as a subclinical deficiency state. Eventually the deficiency state is developed to the point at which it can be observed physically and produces the classical clinical symptoms of a nutrient deficiency.

An example of this process is seen when a dietary deficiency of iron produces iron-deficiency anemia. Inadequate dietary intake of the minerals leads to declining stores in the body. A sensitive clinical test that measures the amount of the body's iron-carrying protein, transferrin, and the amount of iron it is carrying can detect a developing iron deficiency before many of the symptoms of anemia are observed. As body stores continue to be depleted, a lack of sufficient iron impairs the production of heme, a prosthetic group necessary for the formation of hemoglobin in red blood cells. The production of red blood cells declines and the reduced cell number can be determined by a simple clinical blood test called a hematocrit. When the decreased number of red blood cells cannot transport enough oxygen to the peripheral tissues, the result is fatigue, a clinical symptom associated with iron-deficiency anemia. (*See* **Anemia (Anaemia)**: Iron-deficiency Anemia.)

See also: **Anemia (Anaemia)**: Iron-deficiency Anemia; **Ascorbic Acid**: Physiology; **Calcium**: Physiology; **Cobalamins**: Physiology; **Enzymes**: Functions and Characteristics; **Iron**: Physiology; **Magnesium**; **Minerals – Dietary Importance**; **Niacin**: Physiology; **Selenium**: Physiology; **Thiamin**: Physiology; **Trace Elements**; **Vitamin K**: Physiology; **Vitamins**: Overview; **Zinc**: Physiology

Further Reading

Berdanier CD (1998) *Advanced Nutrition Micronutrients*. Boca Raton, Florida: CRC Press.

Bowman BA and Russell RM (2001) *Present Knowledge in Nutrition*, 8th edn. Washington, DC: International Life Science Institute Press.

Groff JL and Gropper SS (2000) *Advanced Nutrition and Human Metabolism*, 3rd edn. Belmont, California: Wadsworth.

O'Dell BL and Sunde RA (eds) (1997) *Handbook of Nutritionally Essential Mineral Elements*. New York: Marcel Dekker.

Shils ME, Olson JA, Shike M and Ross AC (1999) *Modern Nutrition in Health and Disease*, 9th edn. Baltimore, Maryland: Williams & Wilkins.

COFFEE

Contents

Green Coffee

R J Clarke, Donnington, Chichester, West Sussex, UK

Introduction

Green or raw coffee comprises green coffee beans; though, in trading practice, it may also contain small amounts of extraneous matter, derived from the harvested and processed coffee cherry and other foreign matter such as stones.

Classification of Green Coffee Beans

A green coffee bean, as defined in the International Standard, ISO 3509–1989 is 'a commercial term designating the dried seed of the coffee plant.' The coffee

plant or tree (not shrub) belongs botanically to the *Coffea* genus in the family Rubiaceae, with subdivisions and some 80 separate species, of which only two species are commercially important for green coffee; these are *C. canephora* (known in the trade as *C. robusta*), and *C. arabica* L. In each of these two species, there are a number of true botanical varieties, but also cultivars, developed by horticultural research and used in plantations for various agronomic advantages. In recent years, a number of interspecies hybrids have been developed, notably arabusta in the Ivory Coast, from the crossing of the arabica and canephora (robusta) species, but also with other lesser-known species growing wild in the hope of conferring advantage in respect of disease resistance, etc. None of these hybrids has yet developed much commercial success. Two particular 'original' varieties of arabica have been generally recognized, *C. arabica* var. *arabica* (syn. var. *typica*) and *C. arabica* var. *bourbon*. Cultivars are usually intraspecific by breeding/selection, such as caturra, mundo novo and catuai in Central/South America amongst the arabicas. Varieties in the *C. canephora* species are less precise, originally found in Africa; but *C. canephora* var. *kouillensis* is important and planted also in Indonesia and latterly in Brazil (where it is known as *Conillon robusta*); and also *C. canephora* var. *nganda*, especially found in Uganda. Whilst these varietal/cultivar names are not normally used in the trade, the different coffee types will contribute, along with other factors, to differences in flavor quality (after roasting/brewing).

Genetically speaking, most of the coffee species are diploid, as is *C. canephora*, but *C. arabica* is tetraploid, that is, arabica has $4 \times 11 = 44$ chromosomes in its genome, unlike *C. canephora*, which has 22. This phenomenon has given rise to problems in interspecific breeding. Arabica plants are self-pollinating, though they can be crossed, whereas canephora (robusta) plants are self-sterile and require cross-pollinating for seed development. Robusta plants are generally propagated by use of cuttings (Fr. *bouterage*), whereas arabica plants are generally grown from seeds in nurseries, and then transplanted. These two species differ in their optimal environment for growing. Robusta will grow at low altitudes, will tolerate high temperatures and heavier rainfalls, and requires a higher soil humus content than arabica, and is generally more resistant to diseases and pests (hence its common name). Whilst arabica is grown at higher altitudes (with quality connotations in respect of height above sea level) the plants are particularly susceptible to frost damage, which can occur from time to time, particularly in Brazil.

In general, coffee plants are only grown in those countries between the tropics. Arabica is generally believed to have originated in Ethiopia, and was first cultivated for large-scale export from the Yemen in about 1600 by the Turks. From the Yemen, seedlings were transported by Europeans to other parts of the world, so that it is now found mainly in Central/South America, and also in India, Kenya, Tanzania, and other countries. Robusta derives from the rain forests of Central Africa, but was only really discovered and commercialized from about 1880. It is now mainly grown in plantations in West Africa, and also in Uganda and Indonesia.

The flavor quality (after roasting/brewing) of robusta is generally considered to be inferior to arabica. It is certainly less expensive per unit weight of green coffee, and now constitutes about 25% of the world trade (imports into consuming countries). Its particular characteristics have been found favorable in the manufacture of some instant coffees, but robusta is also widely consumed as regular brewed coffee in countries such as France, Italy, and Spain, and often features in espresso coffees.

A further classification of coffee beans relevant to both arabica and robusta coffee is into (1) flat beans, a term characterizing the majority of beans produced, with their single flat side with a central cleft, and (2) peaberries, which are small rounded beans resulting from a false embryony within the original cherry. The latter have a specialty roaster interest, as do so-called Maragogype, with an abnormally large-size arabica bean, found in some parts of Brazil.

A further basis of classification is described in the next section.

Green Bean Processing

Since coffee is originally harvested in the various growing countries, as 'cherries' or 'berries' with a fleshy interior usually carrying two seeds, and an outer skin, a sequence of operations is carried out in those same countries, in order to remove their seeds and present them as beans (dried seeds) or the 'clean coffee' of commerce. Figure 1 illustrates the sequences.

Removal of Beans from within Coffee Cherries

Two procedures have been developed – one, called dry processing, and the other wet processing (for washed and pulped coffee). The first, also in historical order, requires the sun-drying of the coffee cherries laid out in layers (about 30 mm thick), which need to be periodically turned over during a period of some 3 weeks, until the moisture content is brought down at least below 13% w/w. This time period may be substantially reduced by the alternative use of specialized

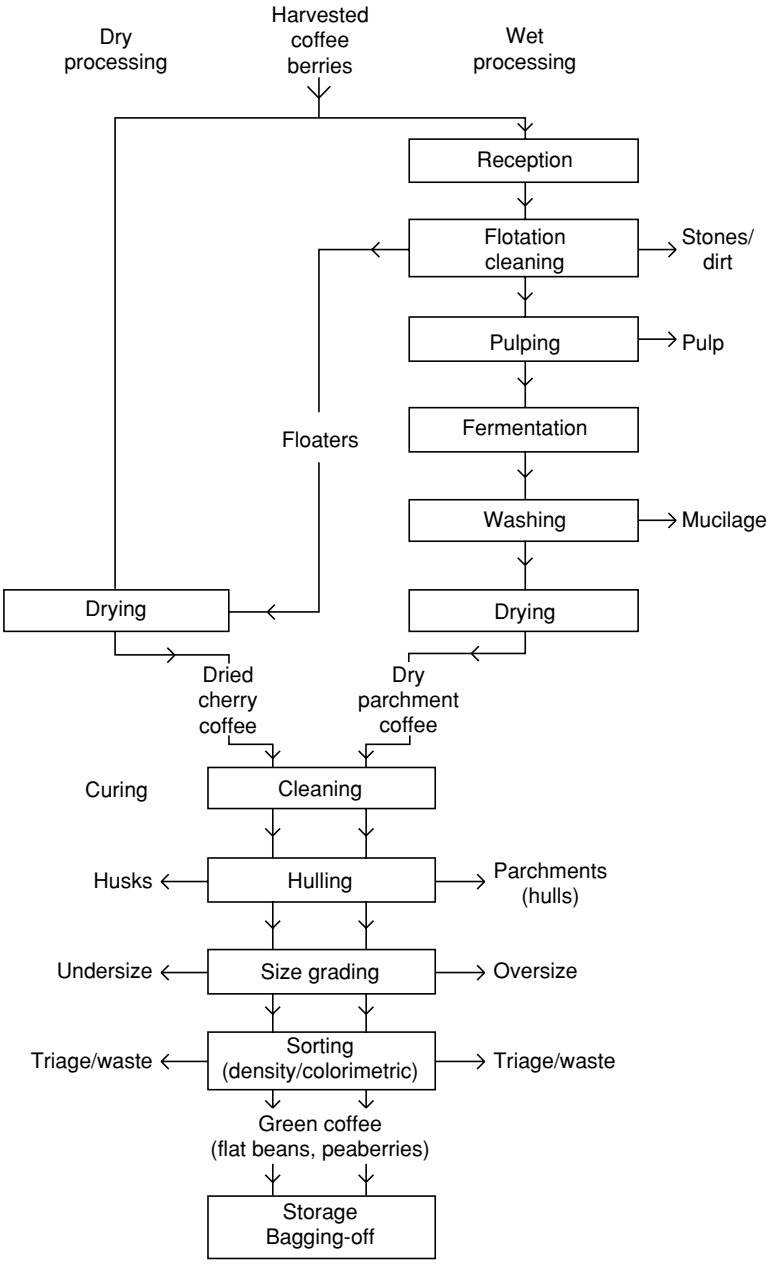

Figure 1 Flow sheet illustrating the stages of wet and dry processing. Reproduced with permission from Clarke RJ and Macrae R (1987) Coffee, vol. 2. Technology, p. 2.

machine air driers. This type of process is used for nearly all robusta coffee production in the world, but it is still used for most of the arabica coffee in Brazil and also in Ethiopia and Haiti. The product is now in the form of so-called husk coffee, that is, dried coffee cherries, carrying the dried seeds and all outer covering, which await the next stages known as 'curing,' usually carried out at a central large-scale curing station, which takes in consignments from various outlying plantations, both large and small. In curing, after cleaning, the beans are first separated from their outer coverings by dehusking machines, based upon a screw principle, of which there are a number of commercial designs. The final presentation as clean coffee is described in a subsequent paragraph. (*See* **Drying:** Drying Using Natural Radiation.)

The second procedure, known as wet processing, is more sophisticated than that of the first. In this procedure, it is important that the coffee cherries be first graded for ripeness, preferably by harvesting only those judged to be ripe and of a red skin color; water-flotation methods may be used for those

overripe cherries and dried on the tree, but not underripe cherries, which are equally, if not more, undesirable. These cherries are then fed into pulping machines of various commercial designs, which tear off and separate the skins and fleshy pulps (exocarp and mesocarp, respectively) in the presence of much water. Such machines will, however, leave a portion of the mesocarp, a mucilaginous layer adhering to the pericarp or a parchment layer directly surrounding the beans. The next stage in the process is therefore a means of removal by firstly loosening with fermentation, which is carried out in tanks. There are variants of detail of this operation in different countries, according to local 'know-how,' climatic conditions, and location height above sea level. Fermentation proceeds through the action of microorganisms or enzymes from within the semipulped coffee, over a period of about 24 h under either wet or dry (no added water) conditions. Care has to be taken that excess acidity does not develop, nor that of taints. After a dry fermentation, a water soak has been recommended (Kenya). The loosened mucilage is then completely washed off, by use of large quantities of water as the product is allowed to flow along long concrete channels, after which it is drained of excess water. An alternative system to pulping/fermentation/washing uses the Aquapulper; the manufacturer claims it will achieve all these operations within one machine. The resulting product is now known as 'wet parchment coffee,' which has to be dried. Air-drying is usually practised, that is, the parchment coffee is laid out on supported trays, with movable coverings that can be used during the hottest hours of the day, or again during cold nights. The drying time will be between 10 and 15 days to reach a desired moisture content of 11% w/w. Alternatively the parchment coffee may be machine-dried in a shorter time, or for part of the time. In either method, great care is necessary for quality reasons, and optimal conditions have been the subject of much study in research stations in Kenya, Colombia, and elsewhere, with many published papers. It is now necessary to remove the dry parchment layer to uncover the beans, by hulling machines similar in principle to those used in the dry process, already described, though the percentage amount of dried coverings to be disposed of is clearly much less.

Preparation of the Clean Coffee for Export

The next stages of curing from either of the two processes described above, after cleaning, involve firstly a size-grading stage, that is, by machines with rotating cylinders on a horizontal axis, fitted along their length with punched hole screens, of about three different hole sizes. Size grading is particularly needed for dry processed coffees (Figures 2 and 3). Each of the size grades is then subjected to a series of sorting operations. The first is a density separation, which enables residual extraneous matter originating from the cherry, such as pieces of husk, parchment, etc. (dependent upon the wet or dry process previously used), abnormally light density beans, and indeed of foreign matter, to be removed as much as possible. Separation methods generally rely on air levitation principles, for which a number of machine types are available. A second sorting based upon the color of beans has traditionally relied upon hand-picking as the beans move along a traveling belt, which also enables some other defects such as malformed beans and other residual defective matter to be removed. Electronic sorting is, however, becoming very widely used, with monochromatic light used to sort out, in particular, 'black beans,' regarded as especially unfavorable to quality, which may otherwise occur frequently in dry processed coffees. The target is to achieve substantially less than 1% by weight of black beans (or fewer than five per 300 g sample). Electronic sorting may also use multichromatic light, which enables a wider range of discolored beans to be discarded, though such sophisticated machines are more usual in the consuming countries. For wet processed arabica there has also been a considerable growth in the

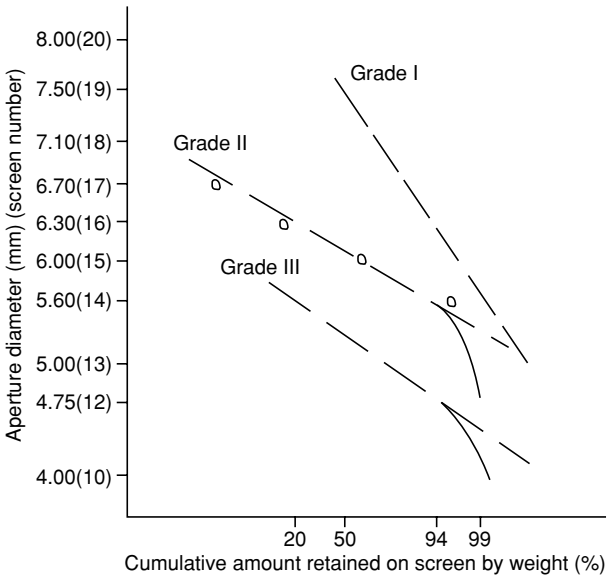

Figure 2 Typical screen analyses for different size grades of Ivory Coast robusta green coffee beans according to specifications (including allowable tolerances), plotting percentage cumulative amount by weight (probability scale) held at each screen size (aperture diameter, mm) for screens according to ISO 4150. o—o, actual experimental data for grade II sample. Reproduced with permission from Clarke RJ and Macrae R (1987) Coffee, vol. 2. Technology, p. 41.

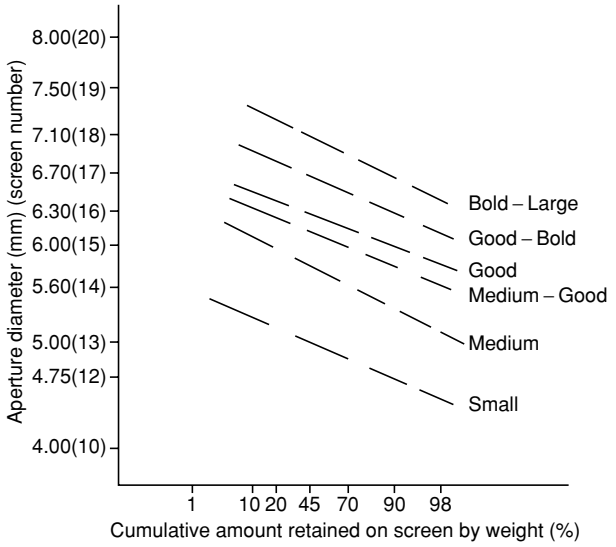

Figure 3 Typical screen analyses for different size grades of Brazilian arabica green coffee beans according to specifications (including allowable tolerances). Percentage cumulative amount by weight held at each screen size (screens according to ISO 4150) plotted on a probability scale. Reproduced with permission from Clarke RJ and Macrae R (1987) Coffee, vol. 2. Technology, p. 43.

use of sorters, in the producing countries, based upon fluorescence differences, which are designed to remove any so-called 'stinker beans,' difficult to detect easily otherwise but especially undesirable to flavor quality if present, even in very small numbers.

The object of all the foregoing stages is therefore to provide 'clean coffee' at the right moisture content, which can then be bagged for export (or internal consumption) and documented according to origin, type, and grade. Commercial green coffees are further subdivided into wet-processed arabicas, known as 'milds' in the trade, dry-processed arabicas, and robustas (dry-processed).

Marketing, Grades, Shipping, and Storage

The marketing of coffee is a complex commercial operation, with an underpinning by the International Coffee Organization, based in London, which, for example, through international governmental agreements in force from time to time can organize quota systems for price stability. For the various green coffees under their control for export, the marketing authorities in different countries issue specifications, which may be either detailed or brief, without necessarily direct reference to flavor (i.e., when roasted/brewed). Test methods are available, many from the International Standards Organization on representative samples from consignments.

Grades and Types of Green Coffee

An important grade criterion is that of bean size (range), as can be characterized by a screen analysis using a number of internationally recognized screens with specific hole diameters (ISO 4150–1980, revised 1991). Different shorthand terms are used to express bean size distribution, e.g., letters such as A, B, C, etc., in Kenya, words (large–small with intermediates) in Brazil, and numbers (1, 2, 3) in robustas from the Ivory Coast. From Colombia, virtually only one size grade (excelso) is exported, all defined on screen no. 15 (6 mm) and most through no. 18 (7.5 mm hole diameter).

A second criterion is that of type, specifically defining the number of defects present per sample, that is, defective beans of various kinds, extraneous and foreign matter. Different systems are in force, though most adopt a black bean equivalency system, where one black bean per 300-g (or 1-1b) sample equals one defect, and where other kinds of defect are assessed as numbers required to equal one black bean. Total numbers of defects define the type, which may be expressed in words, e.g. supérieure in the Ivory Coast for robusta, or terms, e.g. NY numbers (1–8) in Brazil. Such type numbers are not, however, used for wet-processed coffees from Kenya, Colombia, and some other countries, where the number of defects in most of their exported coffee is very small (e.g., < 13 per 300-g or 1-lb sample). Numbers of particular defects such as moldy or insect-damaged beans are rigorously controlled in many countries (both importing/exporting); especially in the USA by the Federal Drug Administration with legislative backing, primarily for health and hygiene reasons.

A specification will, of course, refer to country of origin, maybe also growing area/port of embarkation and species/green bean processing method where this may be otherwise uncertain; and whether new crop or old crop. Purchase or otherwise may be primarily on the basis of exchange of samples, where the opportunity of inspecting for flavor quality and appearance is available.

Shipping and Storage

Some storage of green coffee is inevitable at various times from source to roaster, including of course during transit by ship (whether in bags in holds, or now more usually bags or loose in containers). It is generally recognized that green coffee should not be allowed to reach a moisture content in excess of 13% w/w; otherwise, mold growth will start to occur, which increases rapidly with increasing moisture content, causing flavor deterioration and the possibility of mold toxin formation. It should be noted that

green coffee at a moisture content of 12% w/w is in equilibrium with ambient air of a relative humidity of about 65% at 28 °C. Deterioration is also markedly accelerated by temperature, so that it has been stated that the temperature should not exceed 26 °C or even 20 °C. For a shelf-life of 6 months, the air humidity should also be kept low, and the green coffee at <11% w/w moisture content. However, there is some evidence that some storage of green coffee is desirable for flavor quality, with enzymatic generation of favorable aroma precursor before roasting. Holding of crop from one harvesting season to another, i.e., from new crop to old crop, even under favorable conditions, will give rise to changes in flavor and also in color (fading or whitening on long storage), so that the crop year needs to be known. (*See* **Spoilage**: Molds in Spoilage; **Storage Stability**: Mechanisms of Degradation.)

Compositional Data: Chemical Changes During Processing

The fully detailed composition of green coffee, which is quite complex, has been gradually unraveled by both classic methods of analysis and the more modern laboratory techniques, such as high-performance liquid chromatography. Green coffee, is, of course, especially characterized by its content of caffeine, trigonelline, and chlorogenic acids; otherwise, its composition is similar to that of other comparable vegetable substances with their protein, carbohydrate, vegetable oil, and mineral content. However, the carbohydrate portion consists mainly of polysaccharides, the exact nature of which has only recently been determined. A particular constituent is a mannan (of low degree of polymerization), responsible for the observed physical hardness of green coffee beans. The two main species of the *Coffea* genus, described previously, differ in composition in a number of respects, e.g., caffeine content, on a dry basis: averages are 1·2% in arabica and 2·2% in robusta. Like all natural substances, variations from an average can be found, to which can be added those lesser ones accompanying green processing and storage effects. It may be noted, however, that commercial shipments of green coffee of a particular origin often have a surprisingly uniform composition, resulting from bulking procedures over the range of producing centers arranged by marketing authorities in a given country. (*See* **Caffeine**; **Carbohydrates**: Classification and Properties; **Chromatography**: High-performance Liquid Chromatography.)

Table 1 shows the average composition of the two main species, in broad terms.

In finer detail, the quantity of free amino acids, whilst small, may well be of significance as an aroma precursor in subsequent Maillard reaction during roasting. The content of free amino acids and that of reducing sugars may well increase from zero during green processing through storage; though data are not available, primarily on account of difficulties in carrying out the sophisticated types of analysis on site in producing regions. (*See* **Browning**: Nonenzymatic.)

There are a number of well-characterized enzymes present in green coffee, such as the polyphenol oxidases, which may vary in activity according to their denaturation, relatable to quality by some authorities, though discounted by others. (*See* **Enzymes**: Functions and Characteristics.)

The distribution of these components within the bean has been a subject of study, facilitated by scanning and transmission electron microscopy. A distinction between cell wall polysaccharides and reserve carbohydrates can be observed, and the occurrence of a coffee wax at the surface of the bean can be seen. However, the majority of the lipid material is bound as a globular membrane between the cell walls and the cytoplasm.

Table 1 Green coffee composition

Component	*Typical average content (%, dry basis[a])*	
	Arabica	*Robusta*
Alkaloids (caffeine)	1.2	2.2
Trigonelline	1.0	0.7
Minerals (as oxide ash; 41% K and 4% P)	4.2	4.4
Acids		
Total chlorogenic	6.5	10.0
Aliphatic	1.0	1.0
Quinic	0.4	0.4
Sugars		
Sucrose	8.0	4.0
Reducing	0.1	0.4
Arabinogalactan, mannan and glucan[b]	44.0	48.0
Others[c]	1.0	2.0
Lignin[c]	3.0	3.0
Pectins[c]	2.0	2.0
Proteinaceous		
Protein	11.0	11.0
Free amino acids	0.5	0.8
Lipids		
Coffee oil (triglyceride with unsaponifiables)	16.0	10.0
Total	100.0	100.0
Also small quantity of volatile organic compounds		

[a]Data updated from Clarke RJ and Macrae R (1987) Coffee: Chemistry, vol. 1; (1987) Technology, vol. 2. Barking: Elsevier, and Clarke RJ (1987) Coffee technology. In: Herschdoefer SH (ed.) *Quality Control in the Food Industry*, vol. 4. London: Academic Press. p. 145.
[b]Data from Bradbury AGW and Halliday DJ (1987) Polysaccharides in green coffee beans. In: *Proceedings of the 12th ASIC Colloquium*, pp.265–269. Paris: ASIC.
[c]Only very limited data available.

See also: **Browning**: Nonenzymatic; **Caffeine**; **Carbohydrates**: Classification and Properties; **Chromatography**: High-performance Liquid Chromatography; **Drying**: Drying Using Natural Radiation; **Enzymes**: Functions and Characteristics; **Fats**: Classification; **Spoilage**: Molds in Spoilage; **Storage Stability**: Mechanisms of Degradation

Further Reading

Bradbury AGW and Halliday DJ (1987) Polysaccharides in green coffee beans. In: *Proceedings of the 12th ASIC Colloquium*, pp. 265–269. Paris: ASIC.

Clarke RJ (1987) Coffee technology. In: Herschdoefer SH (ed.) *Quality Control in the Food Industry*, vol. 4. London: Academic Press.

Clarke RJ and Macrae R (eds) (1985) *Coffee: Chemistry*, vol. 1; (1987) *Technology*, vol. 2. Barking: Elsevier.

Clarke RJ and Vitzthum OG (2001) *Coffee: Recent Developments*. Oxford: Blackwell.

Clifford MN and Willson KC (eds) (1985) *Coffee: Botany, Biochemistry, Production of Beans and Beverage*. London: Chapman and Hall.

Roast and Ground

R J Clarke, Donnington, Chichester, West Sussex, UK

Introduction

Green coffee itself has no comestible value for humans, and must first be roasted before use as a flavorful and stimulant aqueous beverage. Furthermore, roasted whole beans must be ground, either by the housewife or by the manufacturer, after which brewing in any one of a number of appliances, manual or automatic, is required.

Roasting Processes

Roasting is a time–temperature-dependent process, whereby chemical changes are induced by pyrolysis within the coffee beans, together with marked physical changes in their internal structure. In particular, a whole range of different volatile organic compounds are generated, many of which are responsible for the flavor/aroma of a prepared coffee beverage, and for the headspace aroma of the dry product; though it is still not entirely certain which compounds, either singly or in combinations, are crucial. The required changes will take place with a bean temperature from about 190 °C upwards, subsequently needing to be controlled on account of exothermic reactions occurring, so that bean temperatures up to 240 °C may be reached in a time that should be less than 12 min. The particular time–temperature chosen will determine the degree of roast from very light to very dark (with intermediate subdivisions) assessed in the first place from visual bean color. The choice of degree of roast is primarily a matter of consumer preference, with medium most usual for US and UK tastes.

In commercial roasting practice, green coffee beans under movement are subjected to heat by conduction from hot metal surfaces, or convection from hot air, or more generally a mixture of both methods of heat transfer, together with a contribution by radiation. Roasters may be broadly classified as in **Figure 1**. The modern types of roaster maximize the use of convective heating, whereas older household machines will rely mainly on conductive heating. The most widely used type has been the batch-operated horizontal rotating drum roaster (with either solid or perforated walls) in which hot air from a furnace/burner is passed through the tumbling green coffee beans. A typical example is shown in **Figure 2**. Dependent upon the air (with combustion gases) flow rate, the temperature of the hot gases/air has to be substantially above the bean temperatures to effect satisfactory heat transfer. In a batch operation, the roasted beans have to be quickly discharged at the end of the required roasting period into a cooling car, or vessel, allowing the upward passage of cold air. In addition, water may also be sprayed from within the

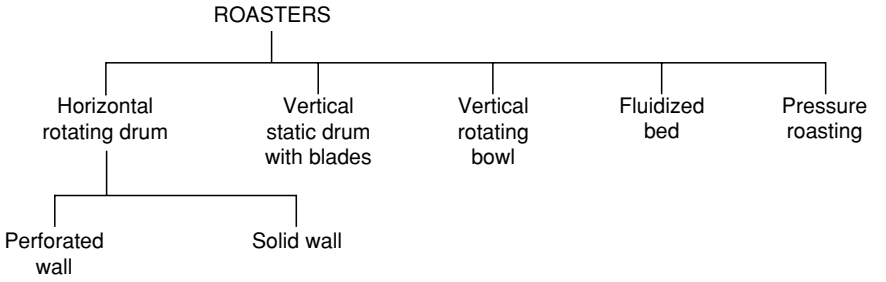

Figure 1 Mechanical principles in roasting methods. Reproduced from Coffee: Roast and Ground, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

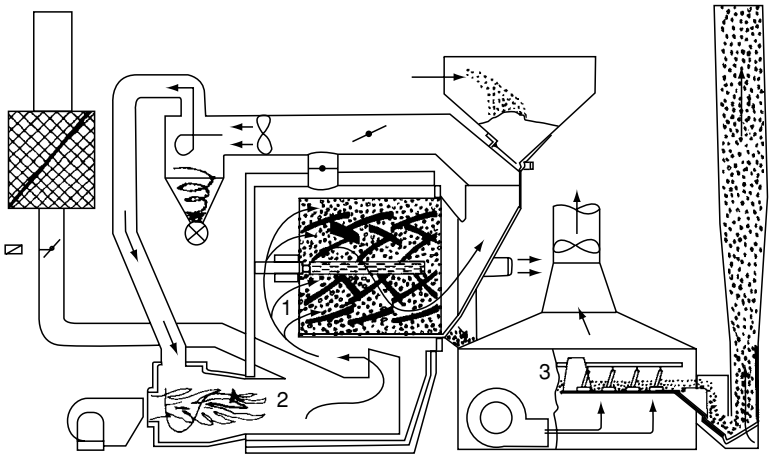

Figure 2 Probat batch roaster, type R, showing (1) solid-wall roaster drum, (2) furnace, (3) cooling car, and ancillaries. Courtesy of Probat-Werke GmbH. Reproduced from Coffee: Roast and Ground, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Kluwer Academic, with permission.

rotating drum, just before the end of the roast, the so-called water quenching. Not only does this operation assist necessary cooling, but it also adds a small percentage of water by weight to the roasted beans, which has been found to assist uniformity of particle size in subsequent grinding. Generated gases, mainly carbon dioxide, from the pyrolysis often carry solid particulate matter (chaff), which has to be continuously vented away and, with current environmental concerns, needs to be 'after-burned' before actual discharge to the outside air. After-burning requires very high temperatures (450 °C), and therefore further energy consumption, though lower-temperature catalytic convertors are also available. In recent decades, there has been a general movement towards incorporation of recirculation methods for the hot gases, in which they are passed back, boosted with additional air, through the furnace, though with controlled partial release to the atmosphere, still including some after-burning or catalytic conversion. A typically sized roaster of this type would hold 240 kg of green coffee, with an out-turn (charging to discharging) of 15 min. The furnace/burner will be either oil- or gas-fired, with a net energy requirement of about 320 kcal per kilogram green coffee (excluding after-burning). Recirculation techniques provide higher velocities of air flow, increasing heat transfer coefficients, and therefore shortening the roast time per unit weight of green coffee but do not necessarily offer energy savings per unit weight of coffee.

The rotating drum principle was first applied to continuous operation in about 1950 of Jabez Burns 'Thermalo' roasters. Since then, roasters based on other mechanical principles have been developed and widely used, e.g., the rotating bowl, vertical drum (with mixing blades) and fluidized air roasters of various designs.

A recent sophisticated type of fluidized bed roaster is the Nepro Vortex Fluidat Equipment, with three separately controlled roasting zones and two, cooling with continuous action, manufactured in Germany by Nolte GmbH.

The general improvement in heat transfer rates has now enabled shorter roast times (i.e., of the order of 3–5 min) in many of these roasters, with claimed advantages for the resultant coffee, so-called fast-roasted coffee of lower bulk density and 'high yield' on brewing.

Chemical and Physical Changes Taking Place on Roasting

Green coffee contains sucrose, polysaccharides, and proteins, together with free amino acids. Considerable potential therefore exists for the occurrence of Maillard-type reactions (essentially between reducing sugars, derived from partial inversion of sucrose during roasting, or already present, and amino acids), leading to the formation of both volatile compounds and high-molecular-weight polymers. Caramelization of sucrose also leads to the production of both volatile and nonvolatile compounds of sensory significance. Under roasting conditions, reactions are necessarily complex and difficult to unravel, with the effect of heat on proteins/carbohydrates in providing reactive centres and molecules to be also considered. (*See* **Sucrose**: Properties and Determination.)

To date, some 800+ different volatile compounds in roasted coffee have now been identified by modern gas chromatography–mass spectrometry (GC–MS)

techniques by numerous investigators since 1960. Many of these compounds have also been quantitatively determined by GC–MS methods, and their total amount found to match closely the actual 700–800 mg per kilogram of roast coffee by steam distillation or other extractive methods. Some 6000 mg of semivolatile acids per kilogram, e.g., acetic and formic acids, are also present. It is clear that amounts of individual compounds present will be in the milligram or microgram per kilogram range. It has been reported recently that the volatile complex comprises by weight 38–45% furan derivatives, 25–30% pyrazines, 3–7% pyridines, 3–5% benzenoid aromatics, 1% aliphatics, and 0.5 alicyclics and 1% of various sulfur compounds (which may well be particularly important for flavor/aroma) in a medium-roast arabica coffee. The mode of formation can be demonstrated by model heating experiments of selected components. Some compounds may also be generated by straight pyrolysis of single compounds, e.g., trigonelline in forming mainly pyridines, chlorogenic acids in generating phenols and of coffee oil in forming small amounts of aldehydes and hydrocarbons. Considerable work has been carried out in recent years on roast coffee volatile compounds with particular reference to those that really are significant for coffee aroma. About 27 are now believed to be important. (*See* **Chromatography**: Gas Chromatography; **Flavor (Flavour) Compounds**: Structures and Characteristics.)

After roasting, substantial amounts of the compounds present may be found to be unchanged, dependent upon the degree of roast by direct analytical determinations. However, there is a newly formed residuum of about 25% by weight of the roast coffee, which is of uncertain composition but variously described as melanoidins/humic acids. These are caramelized sugars/condensation products of carbohydrates and proteins, linked with breakdown products of chlorogenic acids or even these acids themselves. The protein content (assessed by hydrolysis into amino acids) holds up quite well, though the term 'protein' is more applicable, similarly, the polysaccharide content (assessable by hydrolysis into sugars), but the coffee oil is practically unaffected, as is the caffeine. Since there is a loss of mass on roasting, which will range on a dry basis from some 2–3% for a light roast up to about 12% for a very dark roast; indeed, the percentage content of coffee oil and of caffeine will show a slight increase from green to roasted coffee on a percent dry basis. The change of chlorogenic acid content (together with that of its individual component acids) will be marked on roasting; thus, there is an overall 40% residual content for a medium roast (with differential figures for the individual components), which observations can be used as an analytical measure of degree of roast. The fate of much of the chlorogenic acids destroyed during roasting remains obscure. Table 1 shows a typical average final composition for medium roast coffee. (*See* **Caffeine**.)

Physical changes in the roasted coffee can be equally significant, including, of course, color. Scanning electron and light microscopy studies have been undertaken of cross-sections of coffee beans during roasting; cell walls are generally intact, but with a softening of internal structure, and the formation of cavities/cracking of the surface.

There is a marked change in individual bean density through swelling during roasting, which will be reflected in bulk density measurements (both of the whole roasted beans, and of the roast and ground). For a medium-roast bean, the void volume percentage is about 47% compared with virtually zero in the green bean. The cavities are initially filled with gaseous products of pyrolysis, mainly carbon dioxide, which can amount to some 2% by weight of the roasted bean. On standing, this carbon dioxide is slowly released and of course, as a result of subsequent grinding, more rapidly.

Table 1 Roast coffee composition

Component	*Typical average content (%)*[a]	
	Arabica	*Robusta*
Alkaloids (caffeine)	1.3	2.4
Trigonelline (including roasted byproducts)	1.0	0.7
Minerals (oxide ash)	4.5	4.7
Acids		
Residual chlorogenic	2.5	3.8
Quinic	0.8	1.0
Aliphatic	1.6	1.6
Sugars		
Sucrose	0.0	0.0
Reducing	0.3	0.3
Polysaccharides (unchanged from green)	33.0	37.0
Lignin[b]	3.0	3.0
Pectins[b]	2.0	2.0
Proteinaceous		
'Protein'	7.5	7.5
Free amino acids	0.0	0.0
Lipids (coffee oil with unsaponifiables)	17.0	11.0
Caramelized/condensation products (melanoidins, etc.) by difference	25.5	25.5
Total	100.0	100.0
Also 0.7–0.8% volatile substances (other than acids)		

[a]Dry basis figure for a medium-roast coffee.
[b]Limited data.
Data updated from Clarke RJ and Macrae R (eds) (1987) *Coffee: Technology*, vol. 2. Baking, UK: Elsevier, with permission.

Grinding

Grinding of whole roast coffee beans may be conducted either as a small-scale household operation or on a large scale (e.g., typically of the order of 1000 kg h^{-1}). For the first, a variety of single-stage machines, either manually or electrically operated, have been available for some time. The use of a rotating serrated disk against a similar static disk with an adjustable gap is the most usual mechanical principle adopted. On the larger scale, however, multistage twin horizontal rollers are employed, primarily to insure a more uniform particle size distribution than would otherwise be possible. Up to four stages may be used, the first two essentially cracking or crushing the beans into smaller units, followed by stages for progressively finer grinding. The type of serrations on the rollers is important; the most successful type was originated and patented in 1905 by Lepage and still features in Gump grinders. The fast roller of each of the first three pairs has slanting U-shaped corrugations running lengthways from end to end, whereas the slow rollers have a straight U-shaped groove, 'ring-around' or helically cut. The gaps are again adjustable. Simpler corrugations are used in the fourth pair as required for fine grinding.

Different degrees of grind are defined from the results of screen or sieve analysis, with three or four different screens in a set or nest. The nests of screens need to be shaken/tapped under standardized conditions for a given time (e.g., 5 min for a 100-g sample). Especially important is the use of standard wire mesh screens with precisely defined mesh sizes or even certified screens, as described in BS 410-1976 or later, ISO 565-1983 and corresponding other national standards. Relating to roast and ground coffee, the number of different screen sizes numbered by aperture size within the range of about 1400 mm to 250 μm is quite large, so that it will be noted that there are a principal and two supplementary series described in ISO 565, which may result in some confusion. Care is needed in selection of appropriate screen sizes to give suitable weight distributions at each size. In addition, in the USA, there is another set of sieve sizes in the Tyler series. Plotting the results of test screenings will show a good straight line relationship between aperture size (plotted on a linear scale) and the cumulative percentage weight at that size (probability scale), except at the ends (Figure 3). The approximate average size of the particles will be given at the 50% cumulative weight point. More sophisticated statistical methods may be employed to assess average particle size. Newer methods of sizing by laser beams are becoming available.

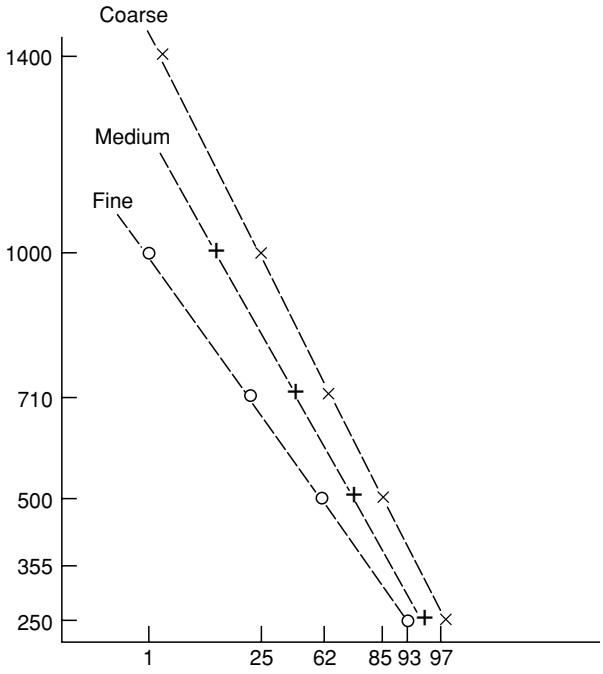

Figure 3 Screen analyses for different degrees of grind of roast and ground coffee, plotting percentage cumulative amount by weight held at each screen size (aperture in μm) according to the R.20/3 series of screens (ISO 565 or BS 410). ×–×, coarse; +–+, medium; ○–○, fine. Data from tabulation in BS 3999 part 8 (1982). Reproduced from Coffee: Roast and Ground, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Kluwer Academic, with permission.

The grind size required is related to the subsequent method of brewing to be adopted, and whether for home use or subsequent large-scale extraction. Grinds are usually defined as coarse – medium, fine, very fine – though the screen analysis that determines these may differ in, say, the UK to the USA. Coarse grinds were used for household percolators, with an average particle size of 850 μm (Europe), 1130 μm (USA), and fine grinds, as used for the now more popular filter machines, are 430 μm (Europe) and 800 μm (USA).

Even finer grinds are required for making Espresso coffees, around 30 μm average, and optimum conditions have been reported for their brewing.

Packaging

Roasted whole coffee beans (RWB) and roasted and ground (R and G) coffee are not really finished products until they have been packaged for sale to the consumer. The type of packaging needed depends upon the time interval between manufacture and sale/consumption. In the case of small speciality retail shops, with in-house roasting, simple packets of

paper or plastic are all that is necessary to maintain quality, provided the coffee is freshly roasted and ground and consumed shortly thereafter, say within a week for R and G coffee (or 10–12 days or longer for RWB).

However, packaged RWB and especially R and G coffee from large-scale manufacture have been available for some considerable time, accelerated by the advent of supermarkets. There are now twin problems in the inevitably longer time scale (e.g., 3–12 months or more) between manufacture and consumption. The first problem is that roast coffee, especially R and G coffee, deteriorates quite rapidly, in respect of headspace aroma and of flavor, in the presence of air/oxygen/moisture, and the second problem is that roasted coffee gradually releases substantial quantities of its entrapped carbon dioxide gas (and other minor gaseous components), R and G coffee much more quickly than RWB. In the case of the former, however, about half will be released in the course of grinding. The amount of gas still retained plotted against time is asymptotic in character to zero, which may well show some 1000 h to near completion for RWB but substantially less time for R and G coffee after grinding.

A fully closed package containing freshly ground roast coffee will therefore rapidly develop a high internal pressure, maybe sufficient to cause the package to burst, and is thus dangerous. RWB will cause a similar or greater problem, though the rate of development of internal pressure will be much slower.

In practice, these twin packaging problems have been solved in a number of ways: firstly, packaging under vacuum, allowing a low percentage oxygen content in the headspace to be established within the package (e.g., 25 mmHg pressure means less than 1% oxygen headspace level), and also accommodating release of carbon dioxide, so that the final pressure in the package may eventually just reach atmospheric; secondly, by allowing the R and G coffee in bulk to degas over a sufficient period of time to a low level, followed by gas purging whilst individual packages are being filled. In the first method, tin cans have been used for many years, and more recently plastic packages, closed by heat sealing and then fitted into cardboard boxes. Such plastic packages have raised a number of additional problems, not least of seal integrity due to the hazard of coffee dust, but also due to a smaller headspace volume than is possible in a tin can, to take up residual carbon dioxide, and some prior degassing may be necessary. The package must not be allowed to become soft, thus restricting the final internal pressure that is tolerable (around 500 mmHg abs. max.). These packages are so-called 'hardpacks,' though in fact a package that has gone soft does not necessarily indicate *per se* that the coffee has deteriorated in quality.

In the second method, where no vacuum is employed, various packaging materials can be used and heat-sealed, to give so-called 'soft packs.' Gas purging is used to ensure that, again, the residual percentage oxygen content in the headspace gas is preferably less than 1%. In both these methods, care is necessary in the selection of plastic packaging material (usually laminates) in consideration of gas ingress/egress, both of oxygen and carbon dioxide and also of volatile aroma compounds, and, importantly, high resistance to water-vapor ingress.

A third method that is especially popular in Europe is in the use of plastic packages to which a nonreturn valve (e.g., Goglio valve) has been securely attached, which allows release of excess carbon dioxide when the internal pressure exceeds a certain predetermined level. In this way, the amount of prior degassing is minimized, and the use of a vacuum is not necessary.

All of these methods allow a much longer shelf-life than is possible with simple air packs, though it should be appreciated that shelf-life is shortened once any package is opened. Stability is discussed in the next section.

Deterioration on Storage

As will have been evident from the previous section, both oxygen and moisture are the agents that are primarily responsible for flavor-quality deterioration together with the effect of temperature. It should be noted that like all foodstuffs, deterioration is continuously occurring from the time of harvesting and, in the case of roasted coffee, from immediately after roasting. Deterioration of flavor quality is, however, a subjective phenomenon, though accompanying chemical changes, only some of which may be directly related, can of course be detected by laboratory analytical techniques. Nevertheless, trained panels with statistical treatment of results can be used to assess deterioration in terms of rating scales/words. Considerable work has been undertaken at the Munich Technological University and others in recent years to study the deterioration of both RWB and R and G coffee under different conditions. They found it useful to rate on a scale (10→4→1), with word subdivisions of quality as 'high' (close to fresh) 'medium' (satisfying) and 'low' (still acceptable without emergence of stale or off flavours). Table 2 shows some of these findings. Other data are available for different conditions but are not directly comparable, showing the markedly increased rate of deterioration at (1) higher moisture contents of the R and G coffee beyond 4%, and (2) percentage oxygen headspace

Table 2 Shelf-life of roasted and ground coffee

Storage conditions			*Allowable time[a] before reaching quality rating of*			
Headspace oxygen (%)	*Moisture content (%)*	*Temperature (°C)*	*High 9→8*	*Medium 6–5 Satisfying*	*Low 4 Acceptable*	*Roast coffee type*
0.5	< 4	21–23	6	12–17	20–25	R and G
1.0	< 4	21–23	4	9–17	14–20	R and G
21.0	< 4	21–23			10 days	R and G
21.0	< 4	21–23	21 days	41 days	74 days	RWB

[a]Time in months, except where otherwise stated.
Data from sources fully cited in Clarke RJ and Macrae R (eds) (1987) *Coffee: Technology*, vol. 2. Barking, UK: Elsevier, with permission.

levels beyond 1% (though an air pack is already included in Table 2). The effect of temperature on stability has been estimated by saying that each 10 °C decrease doubles the shelf-life and vice versa. The advantages of storing roast coffee in a refrigerator is therefore clearly indicated, provided the moisture content does not rise, even if the package has been opened. (*See* **Storage Stability**: Parameters Affecting Storage Stability.)

Various attempts have been made to correlate flavor quality changes with determinable compositional parameters. It may be noted that the chemical changes occurring are primarily related to the volatile organic compounds, called staling; only after a very considerable storage time do the changes start to affect the coffee oil by the development of rancidity. Headspace aroma deterioration may be found to be more rapid than detectable flavor deterioration, but does not necessarily mean a corresponding deterioration in the latter.

Domestic and Catering Methods of Brewing

It has already been stated that roast coffee must be ground before brewing. There are three further factors in brewing: coffee-to-water weight ratio, the appliance used, and the temperatures employed. Some 14–28% w/w of soluble substances are extractable from roasted coffee by hot or boiling water, the so-called 'yield,' though the concentration of these solubles in the final brew (which will be between 0.75 and 4%, and optimally at 1.2% for US/UK tastes), most markedly affects the flavor. Brewing is the extraction not only of soluble substances contributing only to basic tastes but also of the volatile substances for overall flavor, which also is not necessarily exhaustive (40–80%). The less polar substances will tend to reside in the coffee oil within the roast coffee and will be more difficult to extract than the more polar substances but are likely to be the more important for coffee flavor.

Apart from the brewing conditions, the actual flavor quality will be determined by the choice of blend used, and the degree of roast. The choice of grind is also important. Different brewing devices or appliances (for home and catering use) have been developed over the centuries, from both the functional and artistic point of view, leading to a present apotheosis in the automatic electric filter machines. The appliances have been ascribed different names, often according to the maker, but the mechanical operation involved essentially is a means of separating the undesired so-called spent coffee grounds from the required brew formed by sufficient contact with water. The brew should contain as little of the spent ground particles as possible. At the same time, the brew must be presented hot (between 50 and 55 °C), having been prepared using water at or somewhat below boiling temperatures, ostensibly to minimize loss of volatiles to the air. Two main mechanical principles can be identified: (1) 'steeping' (slurrying) of the R and G coffee with water, with or without agitation, followed by sedimentation or filtration or both, and (2) percolation in fixed beds of R and G coffee held in an open or closed container. In the second method, the water may be passed through in a single pass under gravity, now perhaps the most popular, or under pressure (including steam, as in espresso making), or in multipass, in which the water may then include increasing amounts of soluble substances as in so-called household percolators. In the first method, two separate methods can be identified: infusion and decoction (using boiling water). Of the components of roasted coffee, only some will be extracted completely with variable amounts of the others to reach about 28% w/w total maximum, and about 21% optimum under household brewing conditions.

Some comparative studies have been carried out on exactly the same blend/roast degree of coffee in different types of appliance, with one study suggesting a flavor advantage for an automatic filter machine.

See also: **Caffeine**; **Chromatography**: Gas Chromatography; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Storage Stability**: Parameters Affecting Storage Stability; **Sucrose**: Properties and Determination

Further Reading

Clarke RJ and Macrae R (eds.) (1987) *Coffee: Technology*, vol. 2. Barking, UK: Elsevier.

Clarke RJ and Vitzthum OG (2001) *Coffee: Recent Developments*. Oxford: Blackwell Science.

Grosch W (2001) Volatile compounds. In: Clarke RJ and Vitzthum OG (eds) *Coffee: Recent Developments*. pp. 68–89 Oxford: Blackwell Science.

Illy A and Viani R (1995) *Espresso Coffee – The Chemistry of Quality*. London: Academic Press.

Silwar R, Kampschroer H and Tressl R (1986) Gaschromatische-massenspektrometrische Untersuchungen des Röstkaffeearomas. *Chemische Mikrobiologie Technologie Lebensmittel* 10: 176.

Vitzthum OG (1999) Thirty years of coffee flavour research. In: Taranishi R, Wick EL and Hornstein I (eds) *Flavour Chemistry – Thirty Years of Progress*, pp. 117–133. New York: Kluwer Academic/Plenum.

Instant

R J Clarke, Donnington, Chichester, West Sussex, UK

Introduction

Instant coffee is the dried soluble portion of roasted coffee, which can be presented to the consumer in either powder or granule form for immediate make-up in hot water, whilst the insoluble parts, or spent coffee grounds, are left behind at the factory, for the manufacturer to dispose of. Instant coffee is the name by which this product is now generally known, though there are two synonyms, soluble coffee and dried coffee extract, to express in different ways what the product is.

Definitions and Composition

Most countries have made a point of insisting that instant coffee be made from roasted coffee using water only (which includes steam), though allowance is needed for the commercial reality of added aroma oils prepared by mechanical pressing of roasted coffee or by other nonorganic solvent methods. This view is expressed in a standard of the International Standards Organization, ISO 3509–1989 (originated in 1970), which carries a simple definition, 'the dried water-soluble product, obtained exclusively from roasted coffee by physical methods using water as the only carrying agent which is not derived from coffee.' The European Economic Community (EEC) Directive for coffee and chicory extracts, 1977 and as amended 1983, in legislative force in all member countries has a somewhat longer definition, 'coffee extracts are the products in any concentration obtained by extraction from roasted coffee using water only as the medium of extraction and excluding any process of hydrolysis involving the addition of an acid or base and,

1. containing the soluble and aromatic constituents of coffee
2. which may contain insoluble oils derived from coffee, traces of insoluble substances derived from coffee, and insoluble substances not derived from coffee or from the water used for extraction.'

Instant coffee is then the dried product of such an extraction, and can only be labelled as such (or with synonyms, or foreign language equivalents) when the product so conforms. Products may of course contain other soluble beverage constituents, but then need to be appropriately labelled and all ingredients declared. Some soluble mineral constituents (e.g., 0–1000 p.p.m. of sodium or calcium ions, etc.) may be present from the water used in the extraction.

Historically, most instant or soluble coffees first contained added carbohydrates (about 50% w/w) such as corn syrup solids, since it was found that a simple aqueous extract of roasted coffee, extracted under atmospheric conditions (100 °C temperature), could not be dried (usually spray dried) to a satisfactory free-flowing low hygroscopic powder. It was only from about 1950 that instant coffee at 100% pure coffee solids became commercially available on a large scale. This step forward resulted from the discovery that the desirable carbohydrate (polysaccharide) substances could be obtained from the roast coffee itself, by a further aqueous extraction at temperatures up to 175 °C and addition to the simple extract before drying, and thus providing a powder of satisfactory physical properties. (*See* **Carbohydrates:** Classification and Properties.)

Instant coffees therefore have a higher yield of soluble substances from the roast coffee than that of a normal household brew prepared at 100 °C or below. The additional substances, though extracted or solubilized at temperatures above 100 °C, are fully soluble in hot or even cold water. It is often asked what is the difference in composition between cups of brewed and of instant coffees, when in fact there is a continuum, since they contain essentially the same substances, though the proportions or percentages of individual components will be different according to solubilization achieved for each. Commercial yields can differ considerably from one instant product to another and, for a short time, were subject to legislative control in the European Community, but this was abandoned in 1983. It is now recognized that

yield is only tenuously related to 'quality'; the extraction and retention of aromatic substances in the total manufacturing process is much more important in determining flavor quality. The yield (on a dry basis) of soluble solids from roasted coffee will be around 21% w/w in brewed coffee, up to 32% in exhaustive extraction at 100 °C, and typically 40–55% for instant coffees, dependent upon blend/roast color of the originating roasted coffee. A highly water-soluble substance such as caffeine will be extracted to about 85–100% under household brewing conditions, and to 100% in instant coffee manufacture, but the actual percentage content in the soluble solids of the former will be substantially lower in the latter. In a cup of coffee, the quantity present will be determined by its strength, typically from 2 g of instant coffee dissolved in 150–170 ml of water in a cup. Table 1 shows a comparison of approximate composition by the main individual components for a typical filter coffee brew and a corresponding instant coffee beverage. The manufacture of instant coffee is accompanied by some slight hydrolysis of the polysaccharides in the roast coffee, which is reflected in the slightly increased reducing sugar content (i.e., arabinose, mannose, and galactose) and probably assists solubilization of these polysaccharides, not otherwise easily possible at 100 °C. The quantity of volatile organic substances present in an instant coffee, which are so important to the flavor of a coffee beverage, will depend upon the particular process used in its manufacture, as described below. The moisture content of instant coffee is controlled by the drying stage, which should be less than 5% w/w. (*See* **Caffeine**.)

Physical Forms of Instant Coffee

In its earlier marketing, instant coffee was almost entirely sold as a spray-dried powder, light-to-dark brown in color, free-flowing, and, importantly, with a bulk free-flow density of between 180 and 220 g l^{-1}. The latter two properties enable easy spooning out from a container, such that a typical semiheaped teaspoon would carry about 2 g of instant coffee to a 150–170-ml cup, giving a strength typically representative of average UK and US consumer taste preference.

From about 1965 onwards, instant coffee in granule form became available, though having the same free flow and bulk density characteristics, but generally somewhat darker in color than the corresponding powder. With improved retention of aromatics also, such products were compared with brewed coffee far more favorably by most consumers than before.

Table 1 Approximate typical composition of brewed and instant coffees[a]

		Filter brew[b]			*Instant*[c]		
Component	*Roast coffee (%, dry basis)*	*R and G basis*[d] *(%, dry basis)*	*Extract basis (%, dry basis)*	*Per cup (mg)*	*R and G basis (%, dry basis)*	*Extract basis (%, dry basis)*	*Per cup (mg)*[e]
Caffeine	1.3	1.1[f]	5.3	110	1.3	2.9	58
Trigonelline (including roasted byproducts)	1.0	0.9	4.3	90	1.0	2.2	44
Minerals (as oxide ash)	4.5	4.0[g]	19.0	400	4.2	9.3	186
Acids							
Residual							
Chlorogenic	2.5	2.5	12.0	250	2.5	5.5	110
Quinic	0.8	0.8	3.8	80	0.8	1.8	36
Aliphatic	1.6	1.6	7.6	160	1.6	3.5	70
Reducing sugars	0.3	0.3	1.4	30	2.2	4.9	98
Polysaccharides (unchanged from green)	31.0	3.2	15.1	320	12.9	28.7	574
Lignin/pectins	5.0						
'Proteins'	10.0	2.0	9.5	200	5.5	12.2	244
Lipids	17.0	0.01	0.04	1	0.02	0.04	1
Caramelized/condensed compounds (by difference)	25.0	4.6	22.0	460	13.0	28.9	578
Totals	100	21.0	100.0	2100	45.0	100.0	2000
Volatile compounds (excluding acids)	0.08	0.04–0.07		4–7	Variable according to process		

[a]Assumed both made from a medium roast arabica coffee.
[b]Brew prepared using 10 g of roasted and ground coffee per cup (150–170 ml) at 21% yield of soluble solids from roasted coffee.
[c]Extract manufactured at a 45% yield of soluble solids from roasted coffee, both on a dry basis; and instant coffee without added coffee oil.
[d]R and G, roasted and ground.
[e]Beverage made up using 2 g of instant coffee per cup (150–170 ml).
[f]Assumed extraction efficiency of 85%.
[g]Assumed extraction efficiency of 90%.

Whilst freeze drying can produce powders from liquid feeds, its use for instant coffee was specially developed to provide a granular product. In a subsequent attempt to match the general appearance of freeze-dried granules, spray-dried powders were granulated by agglomeration methods (either simultaneously with spray drying, or subsequent to it), and now form a high percentage of instant coffees sold in the market place. (*See* **Freeze-drying**: Structural and Flavor (Flavour) Changes.)

Dilute or concentrated aqueous extracts of roasted coffee would also be 'instant,' but their stability at room temperature is poor. Nevertheless, recent reports show canned coffee liquid to be very popular in Japan. Frozen granules or chunks (−20 °C) would be quite stable, though they have only been commercialized in a very limited way; spoonable frozen instant coffee has been developed but requires the presence of gel-like additives.

Manufacturing Processes

The initial stages in manufacture, selection of green coffees, roasting, and grinding (coarse grind) are as for roasted coffee. The subsequent basic processes needed are extraction, drying, and packing, but with ancillary optional processes of concentration of extract, separate handling of aromatics and agglomeration. The know-how and patents are largely in the hands of large international companies. A typical scheme (simplified) is shown in Figure 1.

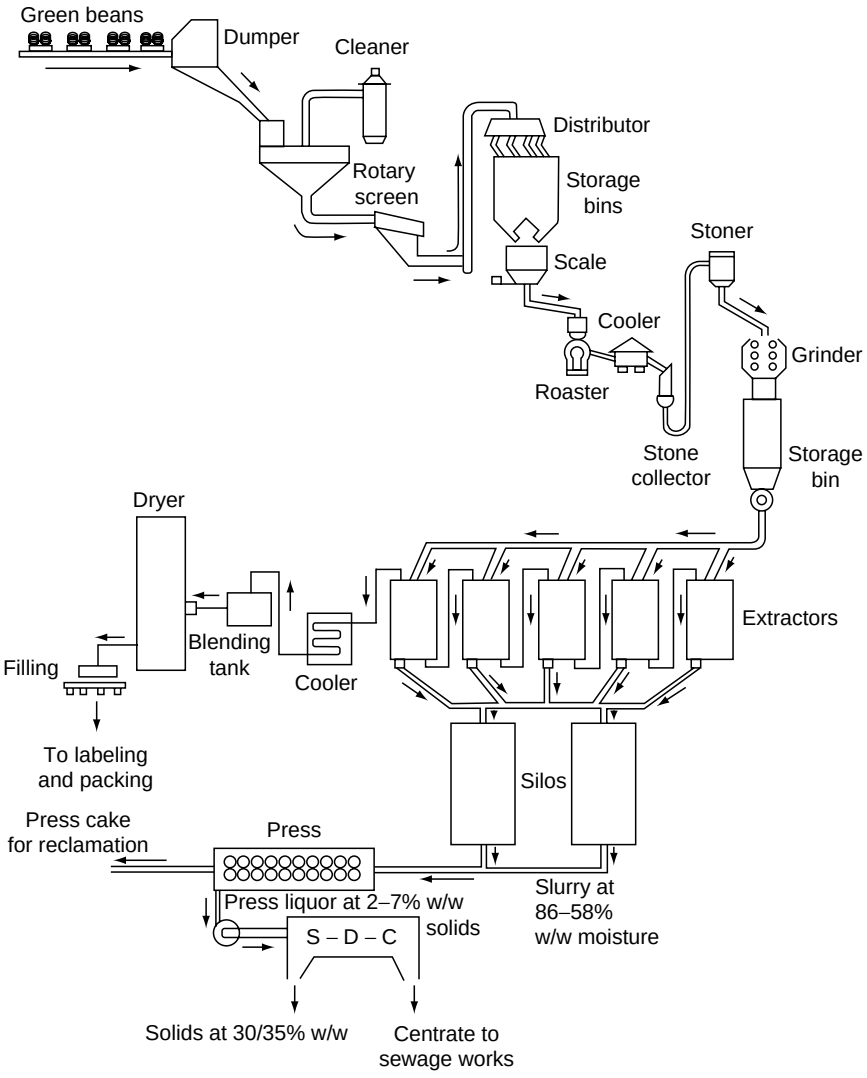

Figure 1 Instant coffee manufacturing process with spent grounds recovery (omitting volatile compound handling). Reproduced from Clarke RJ and Macrae R (1987), with permission from Kluwer Academic.

Extraction

Percolation batteries with five to eight interconnected columns holding roast coffee in different stages of exhaustion is the most widely used system for extraction to obtain coffee extracts of an economic soluble solids concentration (20–25% w/w) for subsequent drying. Its operation and variants are fully described in the literature, though it should be emphasized that the operation is intermittent (for draw-off of extract), countercurrent (in respect of water–coffee), and under pressure for all columns except the first (to keep the system hydraulic at temperatures in excess of 100 °C up to 175 °C). The first column contains fresh roasted coffee extracted with liquor at 100 °C from which draw-offs are made, whilst the most spent is contacted with pure feed water at the highest temperature. The operation is outlined in Figure 2. Recent patents suggest the use of ever higher temperatures for very short times in the final stages of extraction.

Concentration of Extract

The removal of water from a percolated coffee extract is more economically achieved if concentrated extract is fed to the driers. Evaporation is widely used and again, for reasons of heat economy, conducted in various types of multistage units. Short-contact-time evaporators are favored, such as plate and centrifugal film evaporators, coupled with close attention to operating temperatures to minimize undesirable changes in the extract. However, evaporation of water is always accompanied by evaporation and loss of organic volatile substances contributing to flavor. Certain prestripping techniques are available to overcome this problem.

Freeze concentration can be used alternatively, which has the marked advantage of substantially retaining the volatile substances whilst the water is removed as ice, but has the disadvantage of only allowing a final concentration of about 38% w/w soluble coffee solids, due to high viscosity problems.

Spray Drying

Extracts from a percolation battery as described above can be directly spray dried, though an intermediate filtration stage (centrifuges) may be included. To provide dried particles of an average size around 300 μm, needed for a satisfactory bulk density and flowability, specialized spray driers are used. The main features are:

1. tall drying chambers, e.g., of height 7.5 m or more;
2. use of centrifugal pressure nozzles rather than spinning disks for feed spraying, with their inherent narrow angle of discharge;
3. internal chamber separation arrangements between desired product and dust, though not essential.

Spray drying is conducted also to give powder of desired colour, and a final moisture content of less than 5% w/w.

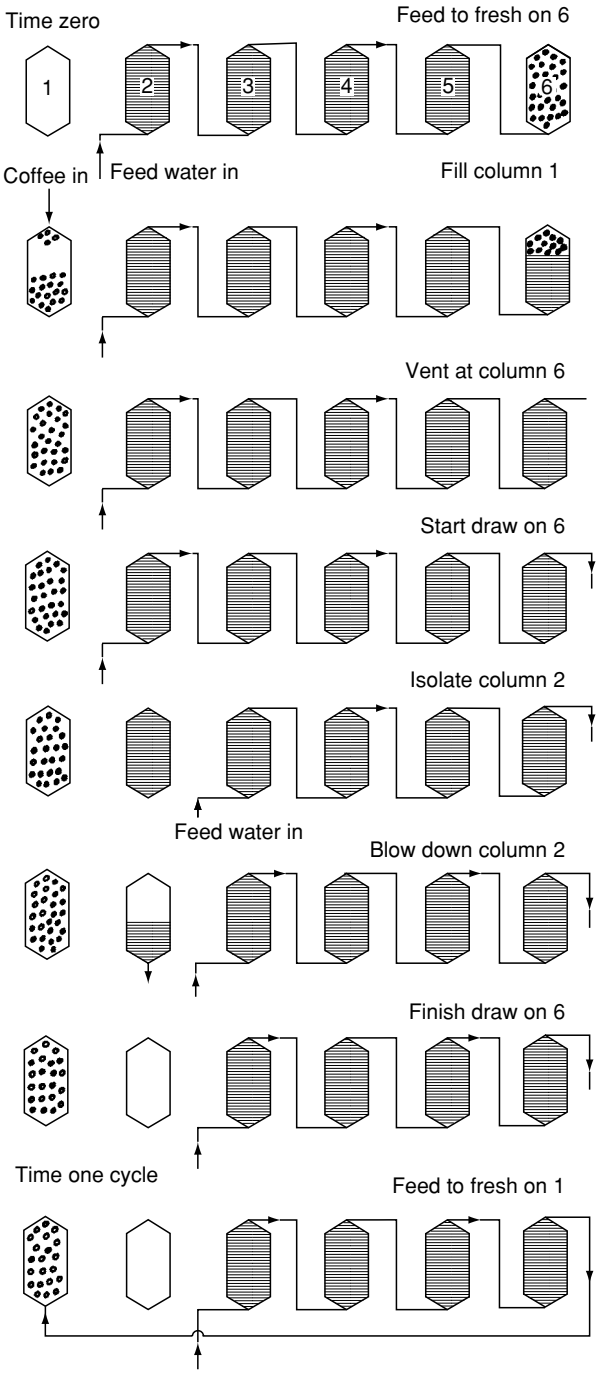

Figure 2 Typical sequence of events in the operation of a percolation battery for instant coffee. The shape, size, and number of columns are diagrammatic. Reproduced from Clarke RJ and Macrae R (1987), with permission from Kluwer Academic.

More concentrated extracts, say up to 60% w/w soluble solids, may also be spray dried, though some controlled foaming of extract with carbon dioxide or nitrogen gas is then generally required, if the desired physical properties are to be maintained. (*See* **Drying**: Spray Drying.)

Freeze Drying

Freeze drying has been particularly successful for coffee extracts, when its use for other foodstuff liquids/solids has markedly declined in the last decades. However, a specialized technique has been developed, first patented around 1965, whereby the coffee extract is first frozen, and then the slabs are granulated, whilst still frozen, to particles approximately the same size as desired in the finished dried product. Oversize/undersize particles are recycled. There are a number of designs of freeze drier available, which will generally handle the frozen granules in trays resting on heated shelves in a batchwise manner. The amount of time of freeze drying required is up to 7 h under a very high vacuum (approx 0.4 torr), and a carefully controlled supply of heat to the drying granules, by conduction and/or radiation.

Whilst satisfactory product can be obtained by freeze-drying extracts direct from a percolation battery, it is more usual for economic and other reasons for concentrated extracts up to 40% w/w by freeze concentration (q.v.) or by evaporation (40% w/w or higher) to be freeze dried. If a favorable bulk density of 180–220 $g\,l^{-1}$ is required, it is additionally necessary to foam the extract in the slush-frozen form, before full freezing and freeze drying. (*See* **Freeze-drying**: The Basic Process.)

Separate Volatile Compound Handling

In recent decades, instant coffee manufacture has become increasingly sophisticated, through greater attention to methods of maximizing extraction and retention of volatile compounds responsible for flavor/aroma, which earlier procedures of simple extraction and spray drying could not accomplish. The first method concerns headspace aroma of the finished powder, due to the presence of very highly volatile substances that cannot be retained by any method of spray or freeze drying. A convenient vehicle was found to be coffee oil, which can be sprayed or plated (at a level of about 0.5% or less) on to the powder/granules as already described, and will give an aroma similar to that of dry roasted and ground coffee when sniffed. The coffee oil can be that obtained by mechanical expression of part of the roast coffee blend to be percolated, or a spent grounds coffee oil purified and enhanced with suitable aromatics from other stages of the manufacturing process. The coffee oil and its aromatics are very susceptible to oxidation, hence the need for gas-packed product, as already described. Its application (so-called 'aromatization') will play only a small role in the flavor of the made-up beverage product, so that other methods are needed, of which there are many available, mostly patented. (*See* **Oxidation of Food Components**.)

Battery extraction, designed primarily for the extraction of soluble solids, can be fairly successful also for the extraction of volatile compounds, so that, when followed by freeze concentration and freeze drying, a flavorful product is formed. However, volatile compounds can be prestripped by steam from the roasted coffee before aqueous extraction, in the form of an aqueous condensate to be reincorporated later. Similarly, a percolated extract can be stripped of the important volatile compounds by partial evaporation (say 10–20% of the water), again to give an aqueous essence condensate. The stripped extract can then be further evaporated as required, with the condensate discarded.

The final solution after evaporation containing a high concentration of solubles, can then be slightly diluted back with either or both of the aqueous essence condensates, and used as feed for either spray or freeze drying. The retention of volatile substances in spray drying is known to be very markedly improved when feed extracts of high soluble solids concentration are spray dried. This will be true also of freeze drying, though, additionally, it has been found that parameters of the freezing itself are also very important (e.g., slow freezing to large ice crystals favors subsequent volatile retention).

Agglomeration

Agglomeration was a process first used for dusty skim milk powders to increase average particle size and, more importantly, for ease of rapid dissolution in water, so-called 'instantization.' Whilst the latter is not required for a satisfactory instant coffee powder, a similar agglomeration process, using steam/water to rewet the surface of particles followed by drying, is used to manufacture instant coffee granules that are not in fact freeze-dried. Precise details of processes differ, as the numerous patents will indicate. The diminution of volatile compound content from spray drying to granule formation will only be small.

Packing

Instant coffee was originally packed into tins but, since 1960, has now been almost entirely packed into glass jars. Under EEC prescribed weight

directives, allowable packed weights for retail sale are 50, 100, and 200 g, and certain multiples, accompanied by the average weight system to be assessed over a stipulated number of containers. Filling machinery is now very high-speed (e.g., 250 per minute for 50 g jars), operating by volumetric fill, and adjustable by vacuum level according to the bulk density of the product being packed. Two other operations may also be incorporated, simultaneous spraying of small amounts of coffee oil into the powder/granules for headspace aroma purposes and a final provision of a carbon dioxide/nitrogen atmosphere within the jar to lower the headspace oxygen level to 4% v/v or preferably less.

General Comments

In recent decades, all these process stages have been investigated fundamentally through the disciplines of chemical (food) engineering by a number of people, especially the late Professor Thijssen of Eindhoven Technical University, to provide considerable mathematical insight into their mechanisms.

Storage and Stability

Instant coffee is relatively hygroscopic, easily picking up moisture from the atmosphere and caking at about 7–8% w/w moisture content. For example, to keep instant coffee below 5% (dry basis) moisture content, the relative humidity of the air with which it is in contact must be below 35–40%, though the precise value depends upon the nature of the instant coffee in question, primarily due to differences in porosity of the particular particles made. It is necessary, therefore, that jars of instant coffee be well sealed prior to sale.

A simple spray-dried product, provided the dry matter content is kept above 95% w/w (i.e., < 5% w/w moisture content), will have a shelf-life of several years. The modern sophisticated products, containing substantial amounts of volatile compounds (especially plated coffee oils), need to be packed in jars also with no more than 4% in-pack oxygen content and preferably less, so that the shelf-life to acceptable quality is maintained up to 18 months.

See also: **Caffeine**; **Carbohydrates**: Classification and Properties; **Drying**: Spray Drying; **Evaporation**: Basic Principles; Uses in the Food Industry; **Freeze-drying**: The Basic Process; Structural and Flavor (Flavour) Changes; **Oxidation of Food Components**

Further Reading

Clarke RJ (1987) Coffee technology. In: Herschdoefer S (ed.) *Quality Control in the Food Industry*, 2nd edn, vol. 4. London: Academic Press.

Clarke RJ and Macrae R (1985–1988) *Coffee: Chemistry*, vol. 1, *Technology*, vol. 2, *Commercial and Technico-legal Aspects*, vol. 6. Barking, UK: Elsevier.

Clarke RJ and Vitzthum OG (2001) *Coffee: Recent Developments*. Oxford: Blackwell Science.

Clifford MN and Willson KC (eds) (1985) *Coffee: Botany, Biochemistry, Production of Beans and Beverage*. London: Chapman & Hall.

Analysis of Coffee Products

L C Trugo, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil

Introduction

The composition of the coffee seed is complex, as is the case for other natural products. After roasting at unusually high temperatures many chemical reactions take place, with degradation of the original compounds as well as polymerization of intermediate products, which substantially increases its complexity. These reactions are indispensable for the characteristic coffee beverage. Analysis of the different coffee products may be relatively simple, involving inexpensive equipment such as that used for proximate analysis, or may demand a great deal of analytical expertise and sophisticated apparatus, such as chromatography coupled to mass spectrometers, infrared spectrometers, and capillary electrophoresis to assess coffee flavor composition and authenticity.

Proximate Analysis

Proximate analysis in food products is the analysis of the major constituents which together comprise nearly 100% of the food composition. This generally includes water, ash, total protein, and carbohydrates. Proximate analysis is the first approach for food product characterization and control and it is not dependent on expensive or sophisticated equipment. It is easily carried out by quality control laboratories, including coffee-processing plants.

Water

The control of the water content of coffee at different stages of processing, from green beans to instant coffee, is an important action to achieve a desirable final product. In the first stages of processing the ripe coffee cherries, containing 60–65% water, must be dried to a level of 10–12% which is normally found in the green beans, and to prevent undesirable

microbial growth which can alter the sensory quality of the final product. After roasting, the water level drops to about 5% and after industrial extraction and drying to produce instant coffee, the powder should present values in the range 2–4% of water. The simplest method for general moisture determination in food is oven drying at 105–110 °C until constant weight is reached, and this may be applied to coffee products. However, some inaccurate data may be obtained, particularly when roast or instant coffee is used, due to water formation by the browning reaction or by loss of volatile components at these conditions.

Empirical time–temperature procedures have been proposed, including a two stage drying process for green coffee samples, involving a drying period at 130 °C for 6 h, followed by another of 4 h after a rest period of 15 h in a desiccator. This procedure is recognized as giving reproducible data; however, it may not be convenient for routine analysis. Vacuum oven-drying methods are also recommended, particularly for roast and instant coffee, because it is possible to use lower temperatures of around 60 °C. However, in practice the overnight oven-drying method at 105 °C is still adopted. Nondestructive methods based on nuclear magnetic resonance, near infrared spectroscopy, and microwave techniques may also be used to assess the water content of coffee samples but their wide application may be limited due to the relative high cost. The alternative for more accurate water determination is the use of the Karl–Fischer method based on a specific quantitative chemical reaction of the water from the sample. Automated equipment based on the Karl–Fischer reaction is commercially available and may be useful for routine purposes.

Ash

Ash determination can easily be achieved by dry ashing the coffee sample at 550–580 °C for at least 6 h and usually up to 17 h to obtain constant ash weight. The ash content of green coffee is in the range 3.0–4.5%. This range does not alter significantly in roasted coffee but increases considerably in instant coffee, with values in the range 9–10%. The ash content gives a rough idea of the total mineral amount and it has been a useful parameter for coffee standardization and to detect instant coffee adulteration by the addition of coffee substitutes.

Crude Protein

The crude protein content of coffee products is usually assessed by the universally used Kjeldahl method for total nitrogen determination. Together with other data of proximate analysis this may be useful for quality control but is of very little use as an indicator of a food source of protein since coffee products are not relevant as protein sources in the diet; however, they are important for coffee aroma formation during roasting. The total protein content of green coffee is in the range 11–13% and this value is increased after roasting to 13–15% and 16–21% in the instant coffees. Although these data may be useful for quality control, one must consider that total nitrogen includes non-nitrogen proteins, particularly from trigonelline and caffeine which will inflate the actual protein results. Instant coffees analyzed only for protein nitrogen can present final protein values in the range 8–13% and if the actual protein value is needed, the non-nitrogen content should be subtracted from the crude protein content.

Oil

Total lipids as determined by organic solvent extraction are present in high levels in green and roasted coffees. Significant differences are found in different species, with Arabica coffee containing about 12–13% and Robusta 7–10%. They are important vehicles for coffee aroma and are relatively stable during roasting, showing a relative increase in roasted coffees: Arabica presents 14–20% and Robusta 11–16%. However, only minute amounts are found in the beverage or even in instant coffee since they are not very soluble during domestic or industrial hot-water extraction. The saponifiable fraction of the green coffee oil is the major component, with triglycerides accounting for 70–80% and free fatty acids 0.8–2%. The traditional Soxhlet procedure, based on organic solvent extraction, for example with petroleum ether, is usually applied for routine analysis; however, for green coffee samples it may be necessary to carry out a previous acidic hydrolysis for liberation of structurally bound lipids in order to obtain more accurate results.

Total Carbohydrates

The carbohydrate composition of coffee products varies considerably during processing. In the green coffee bean large amounts of both sugars and polysaccharides are found. In roasted coffee only polysaccharides are found since sucrose is destroyed by caramelization during roasting. However, in instant coffees, free sugars are again present, and these are mainly monosaccharides, due to hydrolysis of polysaccharides during industrial extraction. Colorimetric methods such as the phenol-sulfuric or the anthrone procedures may be applied to assess the total amount of carbohydrates, in green, roasted, or instant coffees; however, a great degree of interference is expected, particularly in roasted and instant

coffee, due to complex browning components formed during processing. Due to this general difficulty many laboratories simply calculate the total carbohydrates by substracting the total proximate results from 100%.

Protein and Amino Acids

Coffee presents a relatively high level of protein in the range 11–16% in the green bean with less than 0.5% corresponding to free amino acids. However, coffee is not considered to be a source of protein in the diet since most of this material will remain in the insoluble residue after water extraction to prepare the coffee beverage. The nitrogen-protein determination will give only a crude idea of the protein content. More accurate results may be obtained by amino acid analysis. The study of specific protein fractions may be achieved by gel permeation or ion exchange chromatography, hydrophobic and reverse-phase chromatography, or by electrophoresis. The importance of protein and free amino acids is related to their contribution to coffee aroma compounds formation during roasting and also to pigment formation though the Maillard reaction. The degradation of amino acids during roasting may lead to the formation of important volatile compounds such as pyrazines, oxazoles, and pyrroles. The analysis of amino acids by acid hydrolysis of protein will provide more accurate true protein data in green coffee. However, for roasted and instant coffees acid hydrolysis will liberate amino acids from Maillard condensed material with little relation to true protein composition. For tryptophan quantification a specific colorimetric method after basic or enzymatic hydrolysis should be applied, since this amino acid is not resistant to the current acid hydrolysis. One less common amino acid, pipecolic acid, has been used for coffee species characterization since it is present in *Coffea arabica* but not in *C. robusta*.

Amino acid analysis may be carried out using dedicated amino acid analyzers based on ion exchange chromatography followed by postderivatization with ninhydrin and detection by colorimetry. More recently, high-performance liquid chromatography (HPLC) using reverse-phase columns with prederivatization has been used. The protein hydrolysate may be treated with phenyl isothiocyanate to form phenylthiocarbamyl derivatives followed by chromatography and ultraviolet detection or submitted to reaction with *o*-phthaldialdehyde (OPA) to produce OPA derivatives for fluorimetric detection. Capillary electrophoresis has also been applied and shows great potential as a competitive technique in relation to sensitivity and resolution. Free amino acid analysis may be important as aroma precursors and their determination can be achieved by deproteination with 10% trichloroacetic acid and the amino acid analysis of the supernatant, as described above. (*See* **Amino acids**: Properties and Occurrence.)

Free Sugars and Polysaccharides

More sophisticated techniques are needed for a more detailed carbohydrate analysis. The determination of individual free sugars has become very important, particularly in instant coffee samples, since the sugar profile seems to be an accurate indicator of coffee authentication. Sucrose is the main sugar present in green coffee and it may easily be determined by HPLC with normal-phase chromatography (generally silica-amino), acetonitrile (70–80%) in water as mobile phase and refractive index detection, after hot water or aqueous ethanol (80%) extraction, in the ground and defatted beans. The extract should be cleaned up by means of Carrez solutions or activated charcoal and filtered before chromatography. The extract may be concentrated if necessary, particularly when roasted and instant coffees are used. However, due to rapid degradation of sucrose during roasting and the appearance of smaller amounts of other free sugars due to polysaccharide hydrolysis during instant coffee manufacturing, the analysis of sugars in both roasted and instant coffees is much more complicated. Chromatographic analysis may be difficult with the use of refractive index detector due to its low sensitivity. The application of evaporative mass detectors and pulsed amperometric detectors has produced better results.

The polysaccharide fraction of coffee is important as a substrate for pigment and aroma formation and also as a source of free sugars, particularly in instant coffee production. The industrial extraction at high temperatures during instant coffee manufacture promotes some degree of polysaccharide hydrolysis, liberating constituent monosaccharides and oligosaccharides from the polysaccharide backbone. Consequently, arabinose, galactose, mannose, and glucose from arabinogalactan, galactomannan, and cellulose are typical of instant coffees. The analysis of polysaccharides is mainly studied after separation of fractions by solvent extraction or by gel permeation chromatography and the subsequent analysis of the constituent monosaccharides after acid hydrolysis ($1\,\mathrm{mol\,l^{-1}}$ sulfuric acid) by gas chromatography or HPLC. Table 1 shows the sugar composition of the water-insoluble fraction from green coffee which contains most of the polysaccharides, before and after being submitted to roasting conditions. Arabinose shows greater lability to heating and this may be explained by the presence of short side chains with

Table 1 Composition of water-insoluble polysaccharide from Arabica coffee before and after roasting

Sugars in the polysaccharide fraction	*Nonroasted g% (dry basis)*	*roasted g% (dry basis)*
Arabinose	11.0	6.0
Galactose	24.8	23.9
Mannose	38.0	41.7
Glucose	20.0	22.0

From De Maria CAB, Trugo LC, Aquino Neto FR, Moreira RFA and Alviano CS (1996) Composition of green coffee water-soluble fractions and identification of volatiles formed during roasting. *Food Chemistry* 55: 203–207.

terminal arabinose in the arabinogalactan structure. Due to this characteristic, arabinose from the polysaccharide fraction decreases significantly during roasting and is found in relatively high levels in instant coffee.

Lipids

Lipids are present in green coffees at relatively high levels (7–13%); higher levels are usually found in Arabica coffee. They are relatively stable during roasting, showing a relative increase in roasted coffee (11–20%). The saponifiable fraction is predominant, with triacylglycerols accounting for over 70% of the coffee oil. The analysis of constituent fatty acids is achieved by usual transmethylation as a derivatization procedure of triacylglycerols, prior to gas chromatography of the fatty acid methyl esters. The fatty acid composition of coffee oil shows linoleic acid as the main component (45% of total fatty acids) followed by palmitic (35%), oleic (8%), stearic (7%), arachidic (3%), and linolenic (2%). The nontriacylglycerol fraction is quite different from other oil seeds. High levels of diterpenes and diterpene mono-fatty acid esters are found, particularly cafestol and kahweol, which account for about 20% of the oil fraction. Apparently, kahweol is present in Arabica but absent in Robusta coffee and for this reason it has been used for identification of coffee species. Cafestol and kahweol have received much attention recently since it has been confirmed that they can increase plasma cholesterol levels, as observed experimentally and by epidemiological studies. Curiously, the correlation between coffee consumption and high levels of cholesterol has been found only in the Scandinavian population drinking unfiltered coffee. It was observed that unfiltered coffee has small amounts of a floating oil material on the surface. This material was determined to be mainly cafestol and kahweol, which were proved to be cholesterol-raising factors in experimental animals. Coffee filtration appears to be sufficient to retain all the diterpenes in the filter, explaining why the correlation between high cholesterol and coffee drinking is not verified in consumers of filtered coffee. The isolation of diterpene alcohols and their esters has been carried out by paper, thin-layer, or column chromatography. The qualitative detection of kahweol is based on the reaction with potassium iodide (KI) in acetic acid medium; this produces a blue color after heating. Sterols are also important components of the unsaponifiable fraction, accounting for some 20% and consisting mainly of sitosterol, stigmasterol, and campesterol. Terpenes and sterols may be analyzed by gas chromatography after deesterification. To aid in structure elucidation, the use of gas chromatography–mass spectrometry may be helpful. The analysis of coffee wax has been carried out to detect the presence of 5-hydroxytryptamine, which appears to be an irritant of the gastric mucosa. The use of reverse-phase HPLC seems to be adequate for the analysis of this class of compound. Because coffee beverage is an aqueous extract, only a small amount of lipid will be dispersed in the liquid and only those with specific biological activity will have a greater impact for the consumer.

Minerals

More than 30 minerals have been determined in coffee products. An unusually high potassium content is found in both Arabica and Robusta coffees, representing about 40% of total mineral content. Particularly interesting is the fact that Arabica coffees present consistently a higher amount of manganese when compared to Robusta, indicating a genetic difference. In instant coffees the mineral profile is similar to green coffee; however, due to salt solubility in water the total mineral content will be higher than green and roasted coffees. The control of ash and also potassium contents of instant coffees may be useful as an indicator of extraction yield. The analysis of potassium may be easily carried out by inexpensive flame photometers. A more complete assessment of mineral composition may be achieved using more comprehensive techniques such as atomic absorption, thermal neutron activation, and plasma spectrometry or inductively coupled plasma mass spectrometry (ICP-MS), after sample mineralization.

Trigonelline and Niacin

Trigonelline is a pyridine derivative present in several species of fruits and seeds, including coffee. It is present in both Arabica and Robusta coffees in the average amount of 1%. It is extensively degraded during coffee roasting and when dark roasting conditions are used, only about 0.1–0.2% remains in the roasted

Table 2 Loss of trigonelline during coffee roasting and the correspondent increase in niacin

Type of coffee	*Type of roasting*	*Trigonelline g% (dry basis)*	*Niacin mg% (dry basis)*
Arabica	Light	0.9	4.9
	Medium	0.7	10.2
	Dark	0.4	14.2
	Very dark	0.2	15.0
Robusta	Light	0.7	9.9
	Medium	0.6	15.0
	Dark	0.3	23.4
	Very dark	0.1	27.0

From Trugo, LC (1984) PhD thesis. University of Reading, UK.

coffee (Table 2). Due to the high solubility in water, trigonelline is readily extracted from instant coffees, appearing in the range 0.9–1.7%. The final concentration in instant coffee depends on the extraction conditions used as well as on the original material and on the roasting degree. Many volatiles are formed by trigonelline degradation, such as pyrrols and pyridines which will contribute to the aroma fraction of the coffee. Nicotinic acid is another important product formed by demethylation of trigonelline. Nicotinic acid or niacin is a vitamin which is absent in green coffee and in fact is produced during coffee roasting, being present in both roasted and instant coffees, in the range 10–26 and 20–50 mg%, respectively. Because niacin is produced during coffee processing, the vitamin is highly bioavailable in the coffee beverage, in contrast with natural sources where it is present in the bound form.

The analysis of trigonelline in coffee products may be carried out by isolating it from coffee extract using a celite column followed by measurement by ultraviolet (UV) spectrophotometry, by paper chromatography followed by elution and UV measurement of the spot, or by HPLC. This last technique has been shown to produce more reliable results and is presently widely used. A simple reverse-phase method using a C_{18} column, water with a small amount of methanol as solvent, and monitoring at 265–272 nm has been quite adequate for routine analysis. Clarification of coffee extracts may be achieved by solid-phase extraction or precipitation of proteins and pigments by means of lead acetate solution or other adsorbent material except Carrez solutions, since trigonelline is strongly adsorbed by zinc ferrocyanide.

Nicotinic acid may be determined by microbiological, colorimetric (with cyanogen bromide and *p*-aminoacetophenone), or chromatographic methods. However, HPLC has frequently been the technique of choice. In this case better results have been achieved using reverse-phase procedures with a C_{18} column, and a buffered ion-paired mobile phase to attenuate the nicotinic acid polarity and increase its retention in the column and detection at 254 nm. As for trigonelline analysis, some clarification procedure of the coffee extract must be used.

Caffeine

Caffeine is widely consumed in many countries as a constituent of many food items such as coffee products, tea, cocoa products, and cola beverages. Due to its physiological and pharmacological effects, caffeine is the most studied compound present in coffee. Caffeine is the major alkaloid present in coffee, although low levels of theobromine and theophylline, which are other purine alkaloids, may also be found. Caffeine is the component responsible for the stimulant properties of coffee beverages. Green Robusta coffee has a higher content of caffeine (1.5–2.5%) than Arabica (1.0–1.2%) and it is relatively stable to heat, showing a small loss during the roasting process. Due to its solubility in hot water, caffeine is highly extracted during instant coffee production. The amount present in instant coffee strongly depends on the green coffee blend used and less so on extraction rate. Robusta coffee is widely used as a blend component in instant coffee manufacture and because it has considerably more caffeine than Arabica the final content depends on the proportion of these types of coffees. Usual figures would fall in the range 2.5–4.8% for regular instant coffee and 0.2% for decaffeinated coffee. (*See* **Caffeine**.)

The analysis of caffeine was carried out for many years using the method based on its extraction from the coffee matrix with chloroform, followed by extensive purification of the extract with celite or alumina and subsequent measurement of caffeine absorption at 272 nm. However, extensive extract purification procedures are needed to avoid the presence of significant amounts of interfering materials. The application of chromatographic techniques, particularly HPLC and gas chromatography, has produced more reliable results. HPLC has been widely adopted due to its simplicity, rapidity, and reliability.

Reverse-phase HPLC methods using a C_{18} column and aqueous methanol as mobile phase with UV monitoring at 272 nm have proved to be perfectly adequate for fundamental and quality control analyses. The water solution obtained from the coffee matrix may be easily purified by solid-phase extraction or by clean-up using Carrez solutions prior to chromatography. When analysis of a large amount of samples is continuously needed, an HPLC method may be relatively easily automated.

Chlorogenic Acids

Chlorogenic acid is a trivial name used to define the major phenolic components found in coffee. This name originates from early observations that green coffee extracts produced a green color after addition of a solution of ferric chloride. In fact, chlorogenic acids present in coffee are mainly esters of quinic acid with caffeic, ferulic, or coumaric acids. Significant amounts of three isomers (in positions 3, 4, and 5 of quinic acid) of each of these groups of caffeoylquinic, di-caffeoylquinic, and feruloylquinic acids are usually found in green and processed coffees. They are not resistant to heat and are degraded during roasting, contributing to the aroma fraction. Consequently, their levels in the coffee beverage are dependent on the roasting degree, with less than 10% of the original green coffee content found in very dark roasted coffees (Table 3).

Chlorogenic acids and their derivatives have received much attention due to recent investigations showing different biological properties of this class of compounds. A lactone of ferulic acid presented affinity to bind to opioid receptors in the brain. The antioxidant activity of caffeoylquinic acid has been demonstrated and di-caffeoylquinic acids presented strong inhibitory activity against integrase enzymes, which are important for virus replication. Although the biological significance of the ingestion of coffee chlorogenic acids is not well defined, their *in vitro* activities observed in different studies are sufficient to draw attention to the chemical and biological characteristics of these molecules.

Colorimetric and chromatographic methods have been applied to chlorogenic acid analysis. The first approach is used for determination of the caffeoylquinic group by reaction with molybdate or total determination using the metaperiodate reagent. The latter is useful for individual chlorogenic acid isomer analysis. Although gas chromatography has been applied to chlorogenic acid analyses, HPLC appears to be the preferred technique. Chlorogenic acid extraction is critical for obtaining accurate results. Different solubilities of different isomers can make quantitative extraction difficult. Hot solvent extraction appears to give more reproducible results since it will stop possible action of endogenous phenolases during green coffee solubilization. Comparison between different extraction methods showed that the use of aqueous methanol (40%, v/v), followed by precipitation of colloidal material using the Carrez solutions (sol. I, 21.9 g zinc acetate and 3 ml of glacial acetic acid dilute in distilled water up to 100 ml; sol. II, 10.6 g potassium hexacyanoferrate (II) in 100 ml distilled water). Reverse-phase HPLC with an ordinary C_{18} column, gradient elution from a pH 2.5 solution of 0.01 mol l^{-1} tripotassium citrate or trifluoracetic acid to methanol, increasing from 20% to 70% and UV detection at 325 nm, has become a popular procedure. The difficulties of commercial standards may be partially overcome by the isomerization of the commercially available 5-caffeoylquinic acid solution at pH 8 with ammonia and heating in a boiling-water bath. Using this procedure a mixture of the three caffeoylquinic isomers is obtained. The same procedure may be applied to other isolated fractions of feruloylquinic acid to help peak assignment. Quantification may be achieved by calculation based on extinction coefficients of different isomers found in the literature.

Other Organic Acids and Coffee Acidity

Together with chlorogenic acid, other organic acids such as citric, malic, oxalic, and tartaric acids are found in green coffee. Other acids may be generated

Table 3 Content of chlorogenic acid groups in green, roasted, and instant coffees

Chlorogenic acid groups	*Green Arabica*	*Medium roasted Arabica*	*Green Robusta*	*Medium roasted Robusta*	*Instant coffee (average of 13 samples)*
CQA	5.8	2.0	6.8	1.8	4.3
diCQA	0.9	0.2	1.4	0.1	0.4
FQA	0.3	0.1	0.6	0.2	1.1
Total	7.0	2.3	8.8	2.1	5.8

Results expressed in g% dry basis.
CQA, caffeoylquinic acid; diCQA, di-caffeoylquinic acid; FQA, feruloylquinic acid.
Adapted from Trugo LC (1984) PhD thesis. University of Reading, UK.

during coffee roasting, contributing to the acidity of the beverage. Consequently, organic acids such as lactic, pyruvic, succinic, glutaric, and others, although present only in small amounts, will contribute to the overall coffee acidity and flavor. Acidity is very important for the taste of coffee beverage. The term 'sour' is frequently used to define the undesirable taste of over-acid coffees. However, their harmonic composition also contributes to the desirable taste of good coffees. Coffee acidity can be assessed by pH measurement or by titration with an alkali solution. Undissociated molecules of acids may have specific flavor which is not exactly related to acidity and in some cases the analysis of individual acids may be useful for quality control. In this case, ion chromatography with electrochemical detection or gas chromatography after derivatization seems to be a more adequate analytical technique. The use of electrochemical detection with continuous-flow injection analysis may be useful for quality control involving a large amount of samples.

Aroma Analysis

The volatile composition of roasted coffee is extremely complex. The green coffee shows relatively simple volatile composition which is completely modified during roasting. Many classes of volatiles are generated during coffee roasting (Table 4). Over a thousand compounds have already been identified in the volatile fraction of roasted coffee but the composition of the characteristic coffee aroma is not yet completely established. The analysis of the headspace composition of coffee by gas chromatography is the most common approach. The use of mass spectrometry coupled to gas chromatography (GC-MS) has proven to be very useful for structure elucidation of the aroma composition. Sometimes it is necessary to concentrate the volatile fraction, which may be too diluted to be detected in the chromatogram. In those cases codistillation procedures using the Likens–Nickerson apparatus or entrapment in solid sorbents such as tenax and porapack may be very useful. Some volatiles, although present in only minute amounts, may have very low sensory thresholds and be extremely relevant for coffee flavor. Other useful purge-and-trap methods may involve freezing at very low temperatures in liquid nitrogen environment of the coffee volatiles carried by an inert gas in capillary fused silica tubes and thermal purging of the concentrated volatiles direct to the gas chromatography column.

Table 4 Volatile compounds important for coffee aroma and their possible precursors formed during roasting

Class of components	*Precursors*
Pyrrole and furan	Reaction products of amino acids + carbohydrate
Furanone	Sucrose
Furan aldehyde	Saccharose, arabinogalactan
Pyrazine	Hydroxy amino acids
Pyridine	Trigonelline, hydroxy amino acids
Phenol	Chlorogenic acid

Adapted from De Maria CAB (1995) PhD thesis. Federal University of Rio de Janeiro.

Due to the complex gas chromatograms obtained from volatile coffee analysis, some research has been carried out using the sniffing port technique. In this approach, part of the column effluent is diverted from the detector to the sniffing person who will register all characteristics of the compound corresponding to the detected chromatographic peak. This sensorial evaluation (aromagram) will provide very useful information about the sensory impact of specific components. Further chemical and structure research will then be directed to the selected peaks which are relevant for aroma composition. Electronic chemical gas sensor systems, called electronic noses, have recently been developed in order to assess the sensorial impact of specific volatile compounds. This kind of equipment has been applied to the analysis of food, cosmetic, and other related materials and, although producing some controversial results, it appears to present some potential for coffee and other food analysis. The gas sensors commercially available are mainly based on metal oxide semiconductors, organic conducting polymers, and piezoeletric crystal sensors. Sensors based on fiberoptic, electrochemical, and bimetal principles are due to become commercial. Statistical methods such as principal components analysis, cluster analysis, canonical discriminant analysis, artificial neural network, and radial basis function are generally used to treat the electronic nose data.

Roasted coffee is considered to be one of the more complex matrices regarding aroma composition and has been chosen to validate some commercial instruments. The instruments appeared to be capable of differentiating between Arabica and Robusta coffee. Sensor output and roasting level were found to produce highly correlated data submitted to another electronic nose equipment. However, it is apparent that there is a need for further investigation and development in this sort of equipment for it to be significant in coffee quality control.

Detection of Coffee Contaminants

The characteristic thermal conditions used for coffee roasting at very high temperatures make the coffee

beverage a product with very low risk of microbial contamination. However, during harvesting, drying, and storage, some extraneous material may be incorporated to the product, such as in any stage of food production. Inadequate storage conditions may favor fungi development, producing mycotoxins which may be resistant to coffee roasting. Ochratoxin A appears to be the most reported mycotoxin detected in coffee products. Although having potent nephrotoxin activity and being immune-suppressant, teratogenic, and carcinogenic, its eventual presence in coffee products does not appear to represent a public health risk due to the very low levels found ($< 8\mu g\ kg^{-1}$). In these cases the use of HPLC or enzyme-linked immunosorbent assay techniques is recommended for analysis. HPLC analysis usually needs extensive concentration and clean-up by means of solid-phase extraction, with polar, nonpolar, and affinity phases, reverse-phase chromatography, and fluorimetric detection (333 nm excitation and 470 nm emission). Atomic absorption spectrometry may be applied for the investigation of possible contamination with metals. HPLC and GC techniques may be indicated for the analysis of polycyclic aromatic hydrocarbons due to the use of direct fired roasters, or to the analysis of phenols and pesticide residues due to contaminated sacks or disinfection treatments. (*See* **Contamination of Food.**)

Sensory Analysis

Sensory analysis is still very important for coffee classification because there is no objective approach to define a clear correlation of chemical data with sensorial attributes. Expert coffee tasters are still important professionals in the coffee industry since they produce important information on coffee flavor directed to consumers' expectations. The use of sensory analysis has increased in both applied and fundamental investigations, coupled with the use of different statistical methods. Some statistical methods, such as analysis of variance (ANOVA), principal components analysis (PCA), cluster analysis (CA) and flash analysis, have been extensively used to assess coffee sensory attributes and consumer preferences. (*See* **Sensory Evaluation**: Sensory Characteristics of Human Foods.)

Coffee Authentication

Different approaches and techniques have been used to assess coffee authenticity. The application of unusually high processing temperatures for coffee roasting followed by grinding to produce roasted and ground coffee make it extremely difficult to compare the physical and chemical characteristics of the final product with the original green coffee beans used. This is further complicated during elaboration of soluble coffees where the roasted and ground material is extracted with hot water under pressure followed by concentration and water evaporation to obtain the dried solubles. Chemical analysis becomes more complex by the generation of a wide range of pigments which interfere with sample preparation.

Different groups of compounds have been used as recognition patterns to differentiate coffee species and varieties and also adulteration by the addition of extraneous material during processing. The analyses of specific lipids such as diterpene mono-fatty esters, particularly cafestol and kahweol which appear to be typical of coffee beans, have been used. Additionally, kahweol is only found in Arabica coffee, which may be useful for species characterization. The fatty acid composition and also the triacylglycerol profile are dependent on species and may produce discrimating data after statistical treatment by principal component analysis, cluster analysis, linear discriminant analysis, or other pattern recognition method. The determination of mineral composition may be used to compare coffee varieties and also as an aid to detect adulteration due to modification of the typical mineral composition of coffee. The volatile fraction may also be useful for coffee authentication due to the presence of typical coffee aroma from specific compounds such as furfurylthiol and kahweofuran. The combination of results from volatile and nonvolatile analyses coupled with statistical treatments also presents potential for coffee authentication. Results from gel permeation profile, chlorogenic acid composition, caffeine, and trigonelline combined with head-space analysis show a potential field for development of authentication methods. The appearance of free sugars in instant coffees from polysaccharide hydrolysis during industrial extraction has been explored as an approach for instant coffee authentication. The typical presence of arabinose, fructose, mannose, glucose, and galactose, with the clear predominance of arabinose, proved to be useful indicators of coffee authenticity. Adulteration of coffee with other plant materials will considerably alter the typical sugar profile. For example, the addition of chicory during instant coffee manufacture will abnormally increase the fructose content and the use of byproducts from green coffee production (husks, parchments, etc.) will show the appearance of mannitol and higher levels of fructose and sometimes xylose. After hydrolysis, pure soluble coffee is characterized by high amounts of total galactose and mannose, while the product adulterated with coffee husks and parchments will present high total glucose

and xylose. The analysis of these free sugars is not always easy and the application of both HPLC and gas chromatography may be necessary. Better HPLC results have been obtained by using pellicular anion exchange chromatography with detection using pulsed amperometry, and this has been adopted as an international standard procedure.

More recently, infrared spectroscopy has been studied as a possible tool for coffee authentication and it has been claimed to be an easy, rapid, and relatively inexpensive technique. The application of near- and mid-infrared spectroscopy by means of a Fourier transform infrared spectrometer to the analysis of Arabica and Robusta coffees produced differentiated spectra. The chemometric treatment of the generated data showed the potential for species identification. Further application of this approach seems to be useful to investigate adulteration of instant coffee, since spectra variation is observed depending on sugar composition of the coffee extract.

Enantiomeric separation of chiral components may be explored in greater detail in the near future for coffee authentication. The analyses of chiral volatile components or their precursors by means of chiral gas chromatography or HPLC may be extensively used for green coffee characterization and to analyze the sensorial impact of enantiomeric volatiles formed during coffee roasting. Enantiomeric profiles may then be useful to verify coffee authenticity and adulteration.

See also: **Amino Acids**: Properties and Occurrence; Determination; Metabolism; **Caffeine**; **Contamination of Food**; **Sensory Evaluation**: Sensory Characteristics of Human Foods

Further Reading

Clarke RJ and Macrae R (1985) *Coffee – Chemistry*, vol. 1. London: Elsevier Applied Science.

Clifford MN and Wilson KC (1985) Coffee – Botany, *Biochemistry and Production of Beans and Beverage*. London: Croom Helm.

De Maria CAB, Trugo LC, Aquino Neto FR, Moreira RFA and Alviano CS (1996) Composition of green coffee water-soluble fractions and identification of volatiles formed during roasting. *Food Chemistry* 55: 203–207.

De Roos B and Katan MB (1999) Possible mechanisms underlying the cholesterol-raising effect of the coffee diterpene cafestol. *Current Opinion on Lipidology* 10: 41–45.

Downey G, Briandet R, Wilson RH and Kemsley EK (1997) Near- and mid-infrared spectroscopies in food authentication: coffee varietal identification. *Journal of Agriculture and Food Chemistry* 45: 4357–4361.

Ky Chin-Long, Noirot M and Hamon S (1997) Comparison of five purification methods for chlorogenic acids in green coffee beans. *Journal of Agriculture and Food Chemistry* 45: 786–790.

Redgewell RJ, Trovato V, Cuti D and Fischer M (2002) Effect of roasting on degradation and structural features of polysaccharides in Arabica coffee beans. *Carbohydrate Research* 337: 421–431.

Schaller E, Bosset JO and Escher F (1998) Electronic noses and their application to food. *Lebensm. Wiss.u.-Technol.* 31: 305–316.

Schlich P (1998) What are the sensory differences among coffees? Multi-panel analysis of variance and FLASH analysis. *Food Quality and Preference* 9: 103–106.

Decaffeination

R J Clarke, Donnington, Chichester, West Sussex, UK

Introduction

Decaffeinated roasted coffees have been available since about 1905, especially in Germany, and have become a substantial sector in the sales of both roasted and instant coffee in the last two decades.

Definitions and Composition

The International Standards Organization in ISO 3509–1989 (and earlier versions) has defined decaffeinated coffee as 'coffee from which caffeine has been extracted. NB. A maximum residual caffeine content would usually be stated in a specification for decaffeinated coffee.' In fact, there is considerable national legislation specifying this maximum residual amount. In the UK and most other European countries, the maximum caffeine content for decaffeinated roasted coffees is set at 0.1% (dry basis); in the USA there is no specific legislation but manufacturers generally claim that more than 97% of the original caffeine has been removed. There is no particular market for partially decaffeinated coffees. For decaffeinated instant coffee, in European Community countries, the maximum caffeine content is set at 0.3% (dry basis) by the European Community Coffee and Chicory Products Directive of 1977 (and as amended 1983). This figure is also generally accepted elsewhere in the world, whether by legislation or otherwise, except again in the USA, where a 97% elimination figure is usual. It should be pointed out that in commercial practice, both of these figures (0.1% and 0.3%) are in

fact higher than those normally obtainable, especially for decaffeinated instant coffee, so that legal enforcement problems rarely arise. The relationship between the maximum figures for roasted and instant coffee was originally based upon a purely nominal extraction yield figure of 33% soluble solids in extracting roast coffee. It can be seen therefore that a cup of brewed coffee (from 10 g of roasted coffee) will contain not more than 10 mg of caffeine, and of instant coffee (using 2 g of product) 6 mg. (*See* **Caffeine**.)

Decaffeination is the name of the process whereby caffeine is removed. Almost entirely in commercial practice the process is applied to green coffee, after which the decaffeinated green coffee is roasted and ground, or converted to instant coffee exactly as for the corresponding nondecaffeinated products. The composition of these decaffeinated products, apart from caffeine content, will therefore be almost correspondingly identical. However, there will be slight differences, and also of flavor, depending upon the particular decaffeination process employed.

Caffeine is the most studied physiologically active component of coffee. Though this activity is generally weak, it has been the subject of numerous publications and much investigative work. According to the US Food and Drug Administration in 1984, 'the evidence received does not suggest that caffeine at present levels of consumption poses a hazard to public health.'

Decaffeination Processes

The various decaffeination processes in use can be classified with subdivisions in various ways. However, the original process, still used, is based upon direct organic solvent extraction of the green beans; subsequent to that an indirect solvent process was devised, in which water is first used to remove the caffeine from the beans, and the aqueous extract is then treated with the same kind of organic solvent as before. Since 1970, a variant of direct solvent extraction has become available in which the solvent is supercritical carbon dioxide. In all these methods, it is necessary to be able to recover the caffeine from the extracting liquids, and generally to refine it to a pure form for sale. It is also necessary to eliminate all but traces of organic solvent from the decaffeinated coffee beans, which should finally have a normal moisture content of, say, 11% w/w for sale or further processing. In commercial practice, residues of organic solvent will be exceedingly small, less than $1\,mg\,kg^{-1}$ in decaffeinated green coffee and, in a recent survey by the US Food and Drug Administration of commercial coffees, less in roasted coffee ($11–640\,\mu g\,kg^{-1}$) and its brews, and even less in instant coffee ($0.49\,\mu g\,kg^{-1}$). No particular legislation for residues has been adopted by European Community countries, though it has been under consideration for a considerable time now, except that of the choice of organic solvent which may be used.

Decaffeination processes and their offshoots have been the subject of much patenting activity, since conventional processes in use can be time-consuming and complex, and furthermore have led to developing environmental concerns. Though it can be seen that residual solvent amounts are negligible in respect of consumer exposure, use of organic solvents at manufacturing sites has to be carefully controlled on account of various potential hazards. This situation has prompted the development of alternative processes, solvents, and caffeine adsorbents.

Direct Solvent Decaffeination

The first commercial process was developed and patented in Bremen, Germany, in 1905, and sold as Café Hag, which name is still in use. The process used benzene as the solvent upon previously steamed green beans. However, benzene is both flammable and toxic, and became replaced by chlorinated hydrocarbons as they became available, and cheaper. Trichloroethylene was particularly favored, though in 1976 it became the subject of US Food and Drug Administration investigations and was gradually phased out and replaced by methylene chloride, which was affirmed for use in 1985 by the above regulatory body. A number of other organic solvents have been proposed, though only ethyl acetate and vegetable oils (including coffee oil and purified spent grounds coffee oil) have been or are believed to be used in commercial practice.

It was early found that a dry organic solvent extracted relatively little caffeine, or only very slowly, from green coffee beans, even though the solubility of pure caffeine in methylene chloride is reported, for example, to be 19 g per 100 g of solvent at 33 °C or 1.82 g in trichloroethylene at 15 °C. By raising the moisture of the green beans, first by steaming and then soaking in warm water, to 20–55% w/w (optionally 42% for methylene chloride) at, say, 67 °C, the transfer of caffeine into the solvent is markedly expedited. The explanation for this effect is twofold: a swelling of the coffee beans to assist diffusion, and destabilization of caffeine–chlorogenate complexes in the coffee bean by heat/water.

The extraction itself is carried out in percolation batteries of five to eight columns (similar to aqueous extraction in the manufacture of instant coffee, though not under pressure), though the time of contact can still be as long as 10 h. About 4 kg methylene

chloride is required per kilogram of coffee. The column containing coffee at the required decaffeination level is then isolated and drained, and steamed to remove traces of solvent for about 1.5 h. The decaffeinated coffee is then dried. The methylene chloride with dissolved caffeine from the most spent column is sent to a refinery, where the caffeine is extracted and the caffeine-free liquid sent back to the battery. A typical operation is shown in Figure 1.

Methylene chloride is known to be quite selective for caffeine amongst the components of green coffee, e.g., trigonelline is very poorly soluble. There will be some loss of weight in decaffeination, apart from that due to the caffeine, but some of this may be accounted for by normal process losses. No published data are available. Vegetable oils, such as coffee oils, are also a suitable solvent for decaffeination use, from which the caffeine can be removed.

Indirect Solvent Decaffeination or Water Decaffeination

A process was patented in 1941 in which water was used to remove caffeine from the green coffee beans. However, there are other water-soluble constituents (up to about 20%) present, so to prevent their extraction the extracting water has to contain equilibrium quantities of these noncaffeine solubles, so-called 'green' extract, but little or no caffeine.

The decaffeination process is again conducted in a percolation battery. After an initial start-up, the caffeine-rich water extract at about 0.5% caffeine content is contacted countercurrently (e.g., in a rotary disk contactor) with an organic solvent (such as methylene chloride, also described under direct decaffeination) at around 80 °C in order to reduce its caffeine content to below 0.05%. This 'green' water extract is then stripped of its residual dispersed and dissolved organic solvent, and is then recycled to the battery to extract further caffeine, and so on. The decaffeinated beans from the most spent column are dropped to a container, after they are washed with water on a screen to remove adhering soluble material, with the wash water then being added to the caffeine-free water extract before recycling. The washed decaffeinated beans are then dried. The methylene chloride stream containing the caffeine is evaporated to leave the caffeine which is then refined, and the solvent returned for reuse in the contactor.

This process is somewhat more complex than the direct method, but has the claimed advantages of a faster (about 8 h) and higher extraction rate of caffeine, less heat treatment of coffee, retention of surface waxes, and purer recovered caffeine. Though no published data are available, it is probable that there is a slightly higher loss of noncaffeine water-soluble substances, including some aroma precursers like free

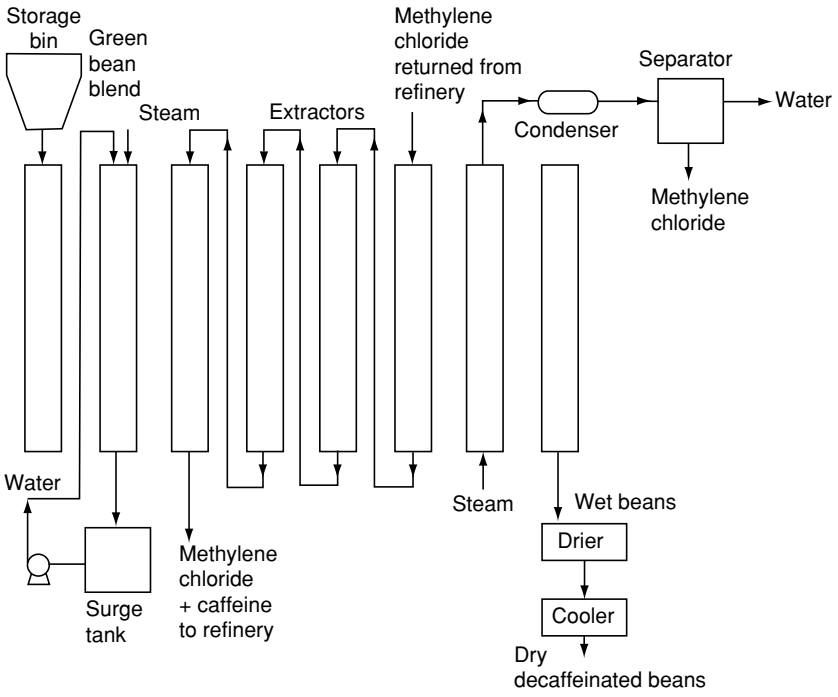

Figure 1 Solvent decaffeination. Reproduced with permission from Clarke RJ and Macrae R (eds) (1987) *Coffee*, vol. 2. *Technology*. Barking: Elsevier with permission.

amino acids, which may affect flavor/aroma on roasting.

Other Water Decaffeination Methods

An important aspect of both of the foregoing processes is the need to separate and recover solvent from the caffeine, usually for low-boiling solvents by multistage evaporation, with a high degree of efficiency. The caffeine itself in an impure form needs to be refined.

Recently, methods have been patented and some commercialized based on the removal of the caffeine from the aqueous 'green' extract by adsorbents of various kinds. The Coffex Company of Amsterdam, for example, proposed the use of an activated carbon which is preloaded with other coffee extract substances or with substitute substances of similar molecular structure or size, especially with carbohydrates, so that the charcoal will take up as few extracted substances as possible, other than caffeine.

The caffeine is eventually desorbed, and the charcoal may then be reactivated for reuse, or disposed of. These types of processes therefore avoid the use of organic solvents in any way.

Supercritical Carbon Dioxide Decaffeination

Certain fluid substances are known to have superior solvent powers in the supercritical state than when in the liquid state, of which carbon dioxide is perhaps the best known example. The supercritical state, however, is only achieved after reaching a minimum critical pressure (e.g., 75.2 bar for carbon dioxide) and critical temperature (e.g., 31.4 °C for carbon dioxide). Numerous diagrams have been published, plotting temperature against pressure and indicating in particular the supercritical region for carbon dioxide and other fluids, and incorporating lines of equal density (Figure 2). Superior solvent powers are associated with increased fluid density, though other factors will also dictate the particular temperature/pressure combination usable and economic in practice.

Zosel, in Mülheim, Germany, from 1970 published a number of patents relating to the supercritical use of carbon dioxide in extracting caffeine from green coffee beans. The claimed conditions were a temperature range of 40–80 °C and a pressure range of 120–180 bar, from which it may be seen that the single fluid phase has a density of the order of 0.5–0.6 g ml^{-3} compared with a gaseous density below the critical pressure of 0.1 g ml^{-3}. Carbon dioxide, being a nonpolar solvent, is quite selective for caffeine. As with organic solvents, it is still necessary to bring the green coffee beans before decaffeination to a moisture content of 30–50%

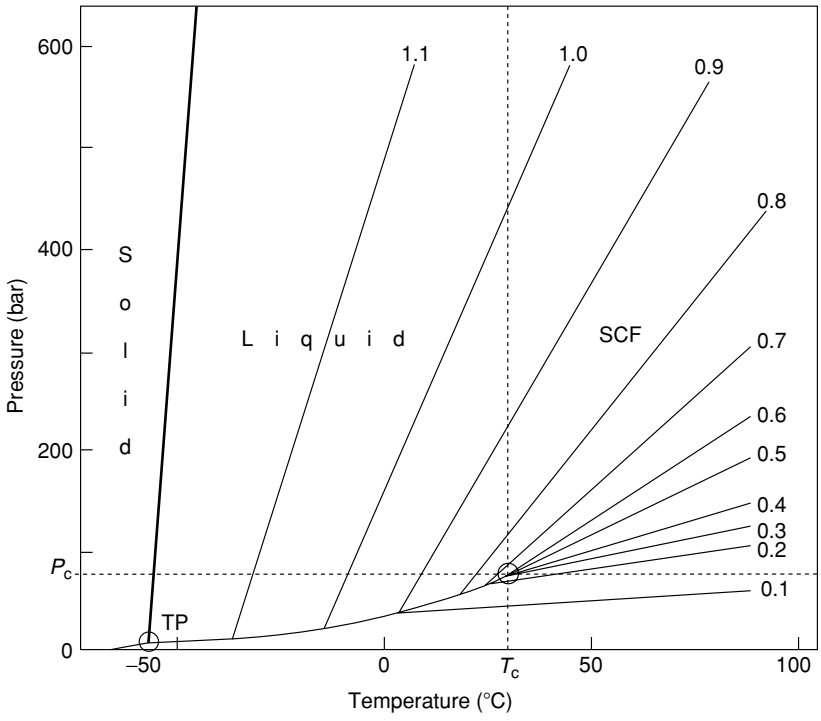

Figure 2 Pressure–temperature diagram for carbon dioxide, showing liquid, gaseous, and supercritical fluid regions (SCF), from P_c and T_c, and isodensity lines. TP, triple point; CP, critical point. Reproduced from Coffee: Decaffeination, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

w/w by steaming/wetting for the same reasons. The removal of the caffeine from the enriched supercritical carbon dioxide can be accomplished by a number of methods: (1) reduction of temperature/pressure; (2) adsorption in a closed circulating system on to activated charcoal; and (3) washing with water. The second method is to be favored in that a given quantity of supercritical carbon dioxide is loaded into and passed through an extractor vessel holding the moistened green beans; and then through a bed of activated charcoal held in a separate vessel, and recirculation performed in a closed-pressure system until the caffeine is sufficiently extracted from the coffee (Figure 3). At the end of a predetermined time the carbon dioxide is taken out to a holding vessel, and both the decaffeinated coffee beans and adsorbent are discharged. The decaffeinated coffee beans are then dried, whilst the caffeine is removed from the adsorbent, which may be reactivated for further use. A decaffeination process along these lines was commercialized by the Kaffee Hag Company, in Bremen, Germany. It can be appreciated that the high pressures necessary involve high capital expense, and the design and running procedures must be carefully considered to minimize energy and other running costs. Nevertheless, the process and the product have many advantages, especially from the marketing and consumer perception points of view; and, with the high selectivity of this solvent, it is claimed that roasted and soluble products from it are very similar in flavor quality to the corresponding nondecaffeinated products.

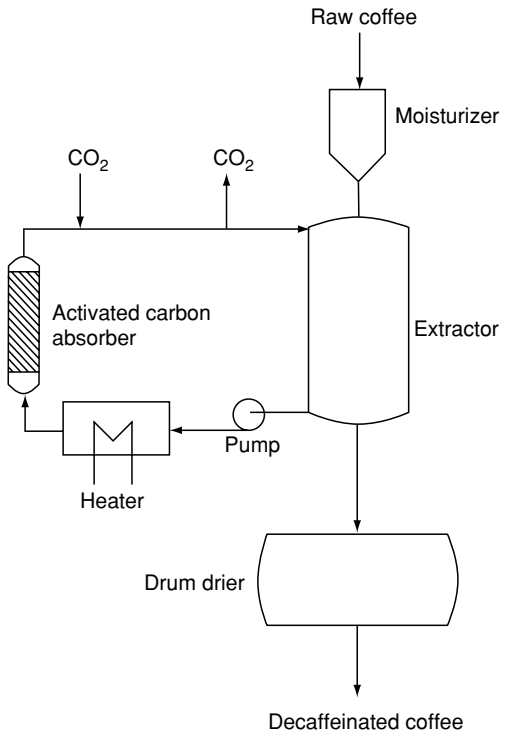

Figure 3 Carbon dioxide decaffeination process, schematic. From Clarke RJ and Macrae R (eds) (1987) *Coffee*, vol. 2. *Technology*. Barking: Elsevier with permission.

There are published and patented variants of this basic supercritical carbon dioxide decaffeination process, including use of other fluid such as nitrous oxide and hydrocarbon gases, and other modified systems of operation. Numerous patent methods have also been proposed for the effective removal of caffeine absorbed in the carbon.

Decaffeination of Roasted Coffee and Liquid Coffee Extracts

There are a number of patents relating to decaffeination of coffee products, rather than green coffee, including use of organic solvents and supercritical carbon dioxide. It would be necessary to remove first all the organic volatiles which are responsible for aroma/flavor before decaffeination; as indeed is already widely practiced in the manufacture of non-decaffeinated instant coffees, and add them back before or after drying.

Effective removal of organic solvent, though not carbon dioxide or vegetable oils, together with emulsification problems are likely to be deterrents to commercialization.

Caffeine Refining

Caffeine represents a valuable byproduct of decaffeination, though this substance can also be manufactured by synthetic chemical methods. Byproduct value may differ in economic terms from time to time. Pure caffeine has an outlet in soft drinks of the cola type, and also for medical purposes.

From the solvent decaffeination processes described, the caffeine will be in a state of 70–85% purity, so that a refining process is required to produce BP (British Pharmacopaedia) or USP (US Pharmacopedia) grade by eliminating coffee waxes and oils as well as water-soluble or other materials that impart a dark color. The series of interconnected and recycle steps involve use of activated carbon filtrations, and repeated recrystallizations from water. The final product consists of white, needle-shaped crystals, with a melting point of 236 °C but subliming at the much lower temperature of 178 °C. Recent studies have shown it to be, when crystallized from aqueous solution, a 4/5 hydrate with 6.95% water content. It has some unusual solubility characteristics in water, in that above 52 °C it is anhydrous caffeine which is stable in contact with aqueous solution and,

conversely, below 52 °C only the hydrate is stable. However, the interconversion is not very rapid, so that true solubility determinations must be obtained with the appropriate form of caffeine according to temperature. At 40 °C, caffeine will have a solubility of 4.6 g per 100 g of water, and at 70 °C, 13.50 g. Of special reference to decaffeination is its reasonable solubility in a number of organic solvents (though very low in carbon tetrachloride and aliphatic/petroleum ethers), and its capacity as a weak base for forming unstable salts, such as acetates and chlorogenates. In addition, it can form soluble complexes with polynuclear hydrocarbons, which may be found in roasted coffee and other roasted substances; this property can be used in their selective extraction.

See also: **Caffeine**

Further Reading

Clarke RJ and Macrae R (eds) (1985–1988) *Coffee: Chemistry*, vol. 1, *Technology*, vol. 2, *Physiology*, vol. 3, *Commercial and Technico-Legal Aspects*, vol. 6. Barking: Elsevier.

Coughlin JR (1987) *Methylene Chloride: A Review of its Safety in Coffee Decaffeination*. Proceedings of the 12th International Colloquium on Coffee. Paris: ASIC, pp. 127–40.

Heilmann W (2001) Decaffeination of Coffee. In: Clarke RJ and Vitzthum OG, eds. *Coffee: Recent Developments*. Oxford: Blackwell Science.

Physiological Effects

R Viani, Corseaux, Switzerland

Consumption Levels in Different Countries

Approximately 1.7×10^9 cups of coffee are drunk each day in the world. The USA is the largest consumer country, and consumption, which had declined to less than two cups per capita per day, is now again increasing thanks to specialty coffee and in particular to espresso coffee. The Nordic countries and The Netherlands have the highest per capita consumption, with four to five cups per day.

Production and consumption figures are shown in **Table 1**. Patterns of consumption vary widely and evolve in time.

Table 1 Coffee production and consumption, average over period 1993–98

	10^6 tonnes per year	*Green coffee (kg) per person per year*
Production	5.75	
Local consumption		
Brazil	0.54	3.7
Colombia	0.08	2.8
Ethiopia	0.07	1.9
Guatemala	0.02	2.3
Mexico	0.61	0.8
Thailand	0.04	0.2
Consumption in importer countries		
European Community	2.12	
Austria	0.11	9.7
Belgium/Luxembourg	0.07	6.2
Denmark	0.05	9.8
Finland	0.06	11.2
France	0.32	5.6
Germany	0.61	7.4
Greece	0.02	2.3
Ireland	0.01	1.7
Italy	0.29	5.0
Netherlands	0.14	9.2
Portugal	0.04	3.4
Spain	0.17	4.4
Sweden	0.09	9.7
UK	0.15	2.5
Japan	0.36	2.9
Norway	0.06	9.8
Switzerland	0.05	7.5
USA	1.10	4.1

Brewing Techniques

The main coffee brewing techniques are:

- Boiled coffee – brew is prepared by boiling coarsely ground light roasted coffee in water (50–70 g l^{-1}); the infusion (1 cup = 150–190 ml) is drunk without separation of the grounds.
- Cafetière coffee – brew is prepared by pushing water through a bed of medium ground coffee with a plunger (6–7 g per cup).
- Espresso – brew is prepared by extracting very finely ground medium to dark roasted coffee (6–7 g per cup) with water at 92–95 °C and 8–12 bar (1 cup = 20–35 ml in Italy, up to 120 ml elsewhere).
- Filter coffee – brew is prepared by pouring boiling water over finely ground light to dark roasted coffee (30–80 g l^{-1}) in a paper filter or automatic drip machine (1 cup = 150–190 ml).
- Instant coffee – brew is prepared by dissolving 1.5–3.0 g of soluble coffee into 80–190 ml of hot water.
- Liquid coffee – ready-to-drink coffee mixture often containing sweeteners and creamers, consumed either hot or cold, mainly in Japan.

- Percolated coffee – brew is prepared by recirculating boiling water through coarsely ground light to medium roasted coffee (30–60 g l^{-1}; 1 cup = 150–190 ml).
- Mocca coffee – brew is prepared by forcing just overheated water through a bed of very finely ground medium to dark roasted coffee (6–7 g per 40–120 ml cup).
- Greek/Turkish coffee – brew is prepared by bringing to a gentle boil extremely finely ground dark roasted coffee (4–6 g) in water (50–60 ml) and sugar (5–10 g per 30–50 ml cup).

Espresso coffee consumption is also increasing in countries such as the USA, where more diluted types of coffee were common.

Boiled coffee consumption, previously quite common in the northern part of the Nordic countries, is now diminishing after the epidemiological link between its consumption and an increase in serum cholesterol has been confirmed with the identification of the responsible factor.

Nutritional Value of Coffee/Effects on Availability of Nutrients in Diet

The coffee brew is naturally poor in digestible proteins, fats, carbohydrates, and sodium, and is considered a nonnutritive dietary component, drunk for sensory pleasure and for its mild stimulatory effects. Its use as a vehicle for nutritious additives such as milk and sugar, and its contribution to the total water intake must, however, not be neglected.

Among the micronutrients found in coffee, niacin (nicotinic acid), formed from trigonelline during roasting, present at levels of 1–3 mg per cup, which corresponds to 5–20% of the recommended daily intake, has been shown to play a role in preventing pellagra in populations with a marginal diet. Animal studies have indicated that trigonelline itself can be transformed into nicotinic acid. (*See* **Niacin**: Physiology.)

Amounts of soluble dietary fiber (sum of the indigestible carbohydrates and of carbohydrate-like components formed at roasting) of the order of 10–25% of the total coffee solids present in the brew may explain, on one hand, the protective role of coffee against colorectal cancer, and, on the other, together with chlorogenic acids, the reduction in absorption of nonheme iron when coffee is consumed with or just after a meal. (*See* **Cancer**: Diet in Cancer Prevention; **Dietary Fiber**: Physiological Effects.)

The hypothesis that the phytate content of coffee (1–20 mg per cup) significantly lowers the gastrointestinal absorption of zinc needs verification. (*See* **Phytic Acid**: Nutritional Impact; **Zinc**: Physiology.)

Potassium, present at levels of 80–160 mg per cup, may contribute up to 10% of the daily intake for an adult. (*See* **Potassium**: Properties and Determination.)

The intake of magnesium and of manganese from coffee is significant. (*See* **Magnesium**.)

The importance of coffee in the calcium balance of the bone is still unclear. (*See* **Calcium**: Properties and Determination.)

Physiologically Active Components

Coffee is consumed for its characteristic flavor and the mild stimulation produced by caffeine, and all its proven behavioral effects appear to be related to its caffeine content. (*See* **Caffeine**.)

Flavor Constituents

Of the 230 components identified in green coffee aroma, most survive roasting and certainly contribute either positively – the aroma – or negatively – the defects, from poor processing – to the flavor of roasted coffee. Most of the volatile constituents formed during roasting of coffee are common to all roasted foods, and none of them alone can explain the aroma of freshly roasted or brewed coffee, so that the organoleptic appeal of the brew is still partially unexplained.

In order to evaluate the contribution of specific components to the aroma, the aromatic complex above a cup or from the cup is separated in a high-resolution gas chromatogram, and the composition of each peak is identified by mass spectrometry, while its smell is evaluated by sniffing at the exit port of the chromatograph, and described by notes such as cocoa-like, floral, or buttery for the aroma, and medicinal, earthy, or moldy for the defects. Some intensely aromatic substances, which alone have objectionable odors – like 3-mercapto-3-methylbutylformate, which pure has a catty-blackcurrant smell – contribute to the roasty character of a brew.

The aromatic threshold of each aroma component is then measured by sniffing at more and more dilute concentrations until only a few substances are perceptible to the nose. By this technique the most potent odorants in a brew can be selected among the about 1000 volatile substances present (Table 2).

The bitter taste of coffee is not just due to the presence of caffeine, itself a bitter substance, but accounting for no more than 10% of the total bitterness of the brew. The chlorogenic acids, like caffeine, are higher in robusta than in arabica coffees, and participate in the bitter taste. Bitterness increases with the degree of roasting, with the formation of bitter

Table 2 Concentrations of the most potent odorants in arabica and robusta

Odorant	*Arabica ($\mu g\,l^{-1}$)*	*Robusta ($\mu g\,l^{-1}$)*	*Odor threshold in water ($\mu g\,l^{-1}$)*
2-Furfurylthiol	19.1	39.0	0.01
E-β-Damascenone	1.3	1.5	0.00075
3-Mercapto-3-methylbutylformate	5.5	4.3	0.0035
3-Methylbutanal	550	925	0.35
Methylpropanal	800	1380	0.7
Methanethiol	210	600	0.2
5-Ethyl-4-hydroxy-2-methyl-3(^{2}H)-furanone	840	670	1.15
2-Methylbutanal	650	1300	1.3
2,4,6-trichloroanisole	If present, a defect		0.001
2-methylisoborneol	If present, a defect		0.0025
Geosmin	If present, a defect		0.005

volatile aromatic substances (pyrazines) and brown pigments, the melanoidins, from the pyrolysis of carbohydrates, polyphenols, and proteins. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Caffeine

Caffeine content per cup The amount of caffeine present in a cup depends on the type of coffee used and on its mode of preparation, and may vary between 1 and 5 mg for a cup of decaffeinated coffee, physiologically an insignificant amount, and 50 mg to more than 150 mg for a cup of regular coffee, corresponding to an intake of 1–3 mg per kilogram of body weight (Table 3).

These figures are well below those producing a urinary concentration of caffeine of $12\,mg\,l^{-1}$, defined by the Olympic Committee as the acceptable upper limit for competing athletes; such a limit could be reached only after a single oral intake of 900–1000 mg.

Physiological effects of caffeine Ingested caffeine is absorbed and distributed throughout all the tissues in the body within minutes, and is eliminated in a few hours, up to 4% as such in the urine, and the rest is metabolized, with a half-life of 2–6 h for a healthy adult. The half-life is increased during pregnancy and in those with an impaired liver function, like newborn babies and patients suffering from liver disease, and shortened in smokers.

The variation in the physiological response to the consumption of equivalent levels of caffeine could be explained by different rates of stomach emptying, as a function of the content of the stomach, or by genetic differences in the metabolic clearance of caffeine between slow and fast acetylators.

Consumption of caffeinated coffee increases the time needed to fall asleep and decreases sleep duration, particularly in older subjects. Caffeine shortens reaction time and prolongs the amount of time during which an individual can maintain auditory and visual vigilance, during boring tasks or when performing physically exhausting work. An improvement in visual acuity of as much as 40% in humans reported after doses of 180 mg of caffeine might be indirectly explained by an increased vigilance after ingestion. The effect of caffeine on short-term memory, if any, is slight. The decrease in hand steadiness (tremor) observed in some people after consumption of caffeine does not affect fine motor control. A link between caffeine consumption and anxiety or panic attacks has not been demonstrated, with the possible exception of psychiatric patients. The relationship between blood pressor and central stimulatory behavioral effects of caffeine at the concentration of a cup of coffee needs further investigation.

At the doses associated with coffee consumption, caffeine produces a thermogenic effect with an immediate increase of about 10% in the metabolic rate and elimination of carbon dioxide; a delayed lipolytic effect with an increase in the plasma level of free fatty acids has been observed in young lean subjects. Caffeine increases muscular oxygen consumption and glycogen–glucose transformation.

Caffeine in coffee has a rapid and short-lasting diuretic action with increase in urinary volume and sodium in subjects kept on a methylxanthine-free diet.

Table 3 Caffeine content of coffee products

Product	*Portion size*[a] *(ml)*	*Caffeine*[a] *(mg)*
Boiled coffee	200	130–170
Cafetière coffee	150–200	40–100
Filter coffee	100–200	40–180
Filter decaffeinated	100–200	3–5
Espresso coffee	20–60	20–60
Instant coffee	100–200	30–120
Instant decaffeinated	100–200	2–3
Mocca coffee	40–80	40–80
Percolated coffee	150–200	40–170
Ready-to-drink coffee	225–285	60–200

[a]Indicative.

Epidemiological evidence shows no casual relationship between caffeine consumption and miscarriages, low infant birth weight, or short gestation period, particularly if smoking is taken into account. The US Food and Drug Administration feels that 'there is insufficient evidence to conclude that caffeine adversely affects the reproduction functions in humans.'

Since caffeine is the most widely used psychoactive substance consumed by humans, the question of a possible caffeine dependence and whether it should be considered a drug of abuse is debated. The four major criteria for drug dependence are withdrawal, tolerance, reinforcement (viz., the ability of the drug in controlling a behavior depending on delivery of the drug), and dependence:

- Caffeine withdrawal symptoms, mainly headaches, usually starting within 24 h, with a peak after 20–48 h and lasting a few days, possibly unrelated to the quantity of caffeine ingested, are well documented.
- Doses of caffeine of at least 200 mg affect sleep by prolonging sleep latency and shortening its total duration. It is still under debate if the difference among individuals in the effect of caffeine on sleep must be attributed to interindividual differences or to tolerance.
- Doses of caffeine at the lower limit of those present in a cup of coffee (25–50 mg) have reinforcing effects, while higher doses (50–100+ mg) reduce the frequency of caffeine intake, and higher doses (400–600 mg) are avoided. The reinforcing activity of coffee may come from its flavor and be unrelated to caffeine.
- In contrast to caffeine, common drugs of abuse, such as amphetamines, cocaine, or nicotine, induce a dopamine release and an increase in glucose utilization in the shell of the nucleus accumbens, indicating a fundamental difference of action.

The consensus in the scientific community is that the relative risk of addiction to caffeine is low, even if it fulfills some of the criteria for drug dependence.

Contaminants

Contamination of green coffee beans by low levels of ochratoxin A (OTA) or, very seldom, aflatoxin B, has occasionally been observed. The presence of OTA has been linked with poor processing practices in the producer countries, particularly drying, and its presence is mainly concentrated in the outer husks. Its formation can also occur during transport or decaffeination, whenever there is a noncontrolled moisture pick-up. Decomposition of up to 85% the amount present in the green beans has been shown to occur during processing, most of it at roasting. OTA exposure from coffee consumption is estimated at around 7%, behind cereals (67%), and beer (10%), and followed by wine (6%).

Adulteration of soluble coffee with worthless husks considerably increases the risk of soluble coffee contamination by OTA. Although the discussion on OTA toxicity is not yet closed, experimental data obtained using combinations of mammalian biotransformation enzymes showed a lack of mutagenicity and that the formation of DNA-binding intermediates is unlikely. (*See* **Mycotoxins**: Occurrence and Determination.)

Formation of polycyclic aromatic hydrocarbons and, in particular, benzo[*a*]pyrene may occur, especially at nonoptimal roasting conditions. The analysis of brewed and soluble coffee indicates that these strongly lipophilic substances are, anyway, not released, but are retained in the spent grounds, both in home brewing and in industrial extraction. Thus, coffee does not constitute a significant source of dietary intake of polycyclic aromatic hydrocarbons. (*See* **Polycyclic Aromatic Hydrocarbons**.)

Health Implications of Coffee Consumption

Botanical species (arabica, robusta), roasting degree (light, dark), types of coffee (regular, decaffeinated), modes of consumption (boiled coffee, soluble coffee, etc.) or specific constituents (methylglyoxal, dicarbonyls) and contaminants (mycotoxins, benzo[*a*]pyrene) have occasionally been associated with different adverse symptoms: the only links clearly established, both in epidemiological and clinical studies, are the positive correlation between the presence on boiled coffee as drunk in the Nordic countries of sufficient amounts of cafestol (and kahweol) to explain the increase in serum cholesterol levels in high consumers, and the caffeine-withdrawal headaches. The polyphenol content of coffee (200–550 mg per cup), either free as chlorogenic acids or linked in the macromolecular melanoidins, may contribute to positive physiological effects, repeatedly shown by clinical and epidemiological studies (such as a protection against alcohol-derived liver disease), attributed to its antioxidant properties. Ongoing research would indicate that 3,4-diferuloyl-1,5-quinide (DIFEQ) (Figure 1), and similar compounds formed at roasting by chlorogenic acids inhibit the adenosine transporter at the concentrations of the brew. Further studies are, however, necessary to verify the suggestion that DIFEQ-like compounds in coffee are able to modulate the stimulant effect of caffeine, and may contribute to health-related effects of coffee.

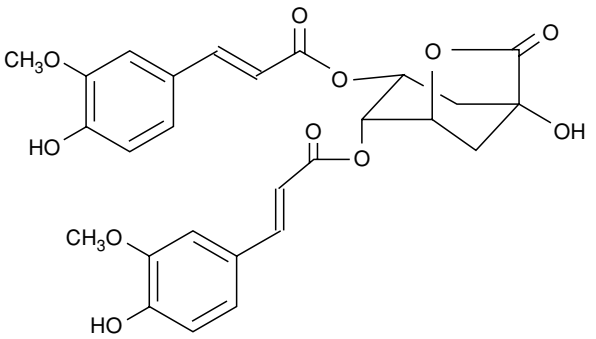

Figure 1 DIFEQ.

Cancer

There is no conclusive evidence from experimental and epidemiological studies that coffee and caffeine are carcinogenic. A doubt still remains on the possibility of a weak link between coffee consumption and cancer of the bladder and urinary tract. Conversely, more and more evidence is collected that coffee consumption may have a protective effect on colorectal cancer. In both cases confirmation is considered necessary.

***In vitro* mutagenicity tests** Hydrogen peroxide and methylglyoxal present in brewed, instant, and decaffeinated coffee are mutagenic in various *in vitro* tests on microorganisms in the absence of the *S*9 fraction containing mammalian microsomal enzymes. Mutagenic heterocyclic amines related to 2-amino-3,4-dimethylimidazo[4,5-f]quinoline (MeIQ) can only be extracted from roasted coffee beans in laboratory experiments under basic conditions. No MeIQ could be found, however, in either brewed or instant coffee.

Animal studies Several lifetime studies in rodents consistently failed to show any correlation between coffee consumption and cancer.

Green, roasted and instant coffee and coffee constituents, the diterpenes cafestol and kahweol and antioxidants, have been shown in animal models to have cancer chemopreventive activity.

Human studies There may be a weak positive relationship between coffee consumption and bladder/urinary tract cancer, but the results of the epidemiological studies are inconsistent, and a residual confounding effect of cigarette smoking or another bias cannot be ruled out. More and more studies, some of which had not been planned to verify the hypothesis, have indicated a protective effect on colorectal cancer. The studies on pancreatic cancer have given inconsistent results. All studies showed no correlation between coffee consumption and breast, gastric, and upper digestive tract cancers. The marginal increase in relative risk found in a few studies on ovarian cancer has been attributed on pharmacological arguments to bias from unknown sources or chance.

Cardiovascular Disease

The epidemiological evidence for a direct link between coffee/caffeine consumption and increased risk of cardiovascular disease is inconclusive. This may be explained by the different consumption modes between and within the populations studied or by atherogenic behaviors positively associated with coffee consumption, such as smoking and high dietary fat and cholesterol intakes. Moderate coffee consumption is not likely to be a significant risk factor for cardiovascular disease.

Serum cholesterol levels There is clinical evidence that high coffee consumption is associated with reversible increased levels of total low-density lipoprotein (LDL) and very-low-density lipoprotein (VLDL) cholesterol, while high-density-lipoprotein (HDL) cholesterol levels remain unchanged. Cafestol (and kahweol) esters, substances specific to coffee present in the lipid fraction, have been identified as the responsible factor. Lipid concentration depends on the brewing method, and is relatively important in the Nordic-style boiled coffee brew, explaining why the first data clearly indicating a link between coffee consumption and cholesterol levels originated in northern Norway, where consumption of boiled coffee was high. Lipid content is negligible in both regular and decaffeinated filter and instant coffees.

Plasma homocysteine levels Elevated homocysteine levels have been suggested, but not yet confirmed, as a risk factor for cardiovascular disease. A link between consumption of high quantities of unfiltered coffee has been shown to increase plasma homocysteine levels by up to 10%. It has not yet been checked if cafestol esters could also in this case be the responsible factor.

Blood Pressure

Abstention from ingestion of caffeine for a period of several weeks may reduce both systolic and diastolic mean blood pressure by 1–4 mmHg. High single doses of caffeine produce after an abstinence of at least 12 h a 5–10% increase of both systolic and, particularly, diastolic blood pressure for 1–3 h. Tolerance develops rapidly with continued consumption and blood pressure stabilizes slightly upwards in a few days. (*See* **Hypertension**: Physiology.)

Cardiac arrhythmias The question of whether caffeine consumption increases the frequency or severity of ventricular arrhythmias is still partially unresolved as there may be particularly sensitive patients: the consensus is now that a 'moderate' coffee consumption, corresponding to 200–500 mg of caffeine per day, is unlikely to increase the frequency of arrhythmias both in healthy subjects and in patients suffering from tachycardia or fibrillation.

Other Health Questions Associated with Coffee Consumption

Coffee and the gastrointestinal function According to a complete review of all the available data '. . . there is no evidence for an association of coffee ingestion with peptic ulcer disease or functional dyspepsia.' The increase in heartburn due to a reduction of the lower esophagal sphincter pressure and consequent gastric acid reflux, associated with both regular and decaffeinated coffee consumption by sensitive individuals, has not yet been explained, and no clear link has been shown with the presence of specific substances in the coffee brew. An increase of the distal colon motility by about 30% of subjects within minutes after coffee consumption has been described in one study. In another study, consumption of two cups of coffee after an overnight fast produced no change in mouth-to-cecum transit time in healthy subjects. Coffee consumption produces contractions of the gallbladder, as is also the case with other foodstuffs. The size of the population suffering gastrointestinal discomfort after drinking coffee, the symptoms themselves, and the mechanism remain unclear, with the possible exception of dyspeptic complaints of patients suffering from duodenal ulcer.

See also: **Caffeine**; **Calcium**: Properties and Determination; **Cancer**: Diet in Cancer Prevention; **Coffee**: Physiological Effects; **Dietary Fiber**: Physiological Effects; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Hypertension**: Physiology; **Magnesium**; **Mycotoxins**: Occurrence and Determination; **Niacin**: Physiology; **Phytic Acid**: Nutritional Impact; **Polycyclic Aromatic Hydrocarbons**; **Potassium**: Properties and Determination; **Zinc**: Physiology

Further Reading

Boekema PJ (1999) Coffee and gastrointestinal function: Facts and fiction – A review. *Scandinavian Journal of Gastroentherology Suppl*. 230.

Bucheli P and Taniwaki MH (2002) Review – Research on the origin, and on the impact of post-harvest handling and manufacturing on the presence of ochratoxin A in coffee. *Food Additives Contaminants* 19: 655–665.

Clifford MN (1999) Review – Chlorogenic acids and other cinnamates – nature, occurrence and dietary burden. *Journal of the Science of Food Agriculture* 79: 362–372.

Committee on Diet and Health, National Research Council (1989) *Diet and Health*, pp. 465–471.Washington, DC: National Academy Press.

Martin PR, De Paulis T and Lovinger DM (2001) Non-caffeine di-cinnamoylquinide constituents of roasted coffee inhibit the human adenosine transporter. *Proceedings of the 19th ASIC Conference*, CD-rom.

Nehlig A (1999) Are we dependent upon coffee and caffeine? A review on human and animal data. *Neuroscience and Biobehavioral Reviews* 23: 563–576.

Schilter B, Holzhäuser D and Cavin C (2001) Health benefits of coffee. *Proceedings of the 19th ASIC Conference*, CD-rom.

Viani R (2000) *Coffee: Ullmann's Encyclopedia of Industrial Chemistry*, CD-rom.

Walker R (1997) Quality and safety of coffee. *Proceedings of the 17th ASIC Conference*, pp. 51–60.

Cognac *See* **Brandy and Cognac**: Armagnac, Brandy, and Cognac and their Manufacture; Chemical Composition and Analysis of Cognac

Cole Crops *See* **Vegetables of Temperate Climates**: Commercial and Dietary Importance; Cabbage and Related Vegetables; Leaf Vegetables; Oriental Brassicas; Carrot, Parsnip, and Beetroot; Swede, Turnip, and Radish; Miscellaneous Root Crops; Stem and Other Vegetables

COLLOIDS AND EMULSIONS

M M Robins and P J Wilde, Institute of Food Research, Norwich, UK

Introduction

In the strictest sense, colloids are dispersions of two or more immiscible materials, where the characteristic length of the inclusions (particles) is between 1 nm and 1 μm. However, in foods, the term 'colloids' generally includes dispersions containing larger particles, and foams, where the bubble size may approach millimeters. Food dispersions thus encompass suspensions such as sugar fondant (solid sugar particles in a saturated sugar solution), emulsions such as milk, cream, sauces, and spreads, and foams such as are found in bread, beer, and icecream. Icecream is an archetypal food dispersion, containing dispersed particles of solid (sugar and ice), liquid (fat), and gas (air). The key property that these foods have in common is the presence of particles, and the consequent high interfacial area between the particles and the continuous phase. The properties of food colloids are dominated by the particles and their interactions, which in turn are controlled by the properties of the interfacial region.

Although other food-relevant systems may contain microscopic particles, such as suspensions of yeast used in brewing, this article is confined to inanimate colloids whose composition and properties are formed naturally, as in milk, or deliberately engineered to manufacture a particular food, such as a dressing.

Figure 1 shows examples of the structure of the common forms of food colloid. Suspensions are usually formed by crystallization processes, which are outside the scope of this article (*See* **Crystallization**: Basic Principles). Similarly, foams have characteristic behavior and properties that are covered more fully in a companion article. Here we focus on emulsions, which are the most common type of food colloid.

Emulsions are liquid-based dispersions containing droplets of oil or water dispersed in water or oil respectively. In order to maintain the droplet structure, the oil–water interface is coated with a layer of emulsifier. In oil-in-water (O–W) emulsions the stability is highly dependent on the emulsifier. However, in water-in-oil (W–O) systems (such as butter) the continuous phase is partially solid, which enhances stability and gives the food a firm texture. Crystallization of the dispersed phase may also be used to generate texture, but in an indirect way, as partial crystallization of the oil droplets in cream and icecream enhances their whipping properties.

Below we outline the composition, formation, and characterization of food emulsions and refer briefly to their safety and nutritional aspects.

Formulation

Food emulsions can be simple O–W or W–O systems, or multiple emulsions where the dispersed phase is already an emulsion (W–O–W, O–W–O). They often contain a complex mixture of ingredients including proteins, polysaccharides, emulsifiers, water, and oil. Each of these ingredients contributes to the overall properties of the emulsion, as detailed in the following section.

Aqueous Phase

The role of the aqueous phase is chiefly as a solvent to the water-soluble components (salts, emulsifiers, proteins, polysaccharides, etc.), as either the continuous phase in O–W emulsions or as the dispersed phase in W–O emulsions. Water molecules possess strong orientation-dependent attractive hydrogen bonds, which result in a high interfacial tension. This makes breaking the dispersed phase into fine droplets difficult. Thus, the interfacial tension forces need to be reduced by the presence of emulsifiers. The pH, ionic strength, and emulsifier concentration influence the formation and physical characteristics of the emulsion, by altering the size of droplets and the interactions between them. The viscosity of the aqueous phase can be altered by the addition of hydrocolloids (gelling or thickening agents, usually polysaccharides such as starch or xanthan gum), which can also affect emulsion formation, rheology, and stability. This is utilized extensively in low-fat O–W emulsion-based food products such as dressings, yogurts, and desserts.

Oil Phase

The oil phase in food emulsions, whether from animal or plant sources, is mainly in the form of triglycerides. The oil also contains levels of di- and monoglycerides, polar lipids, and free fatty acids. The latter three lipids tend to be surface-active, more water-soluble, and sometimes used as food emulsifiers. The three most common fatty-acid chains are the saturated palmitic (C16) and stearic (C18) and the unsaturated oleic (C18:1) and linoleic (C18:2). Plant oils tend to

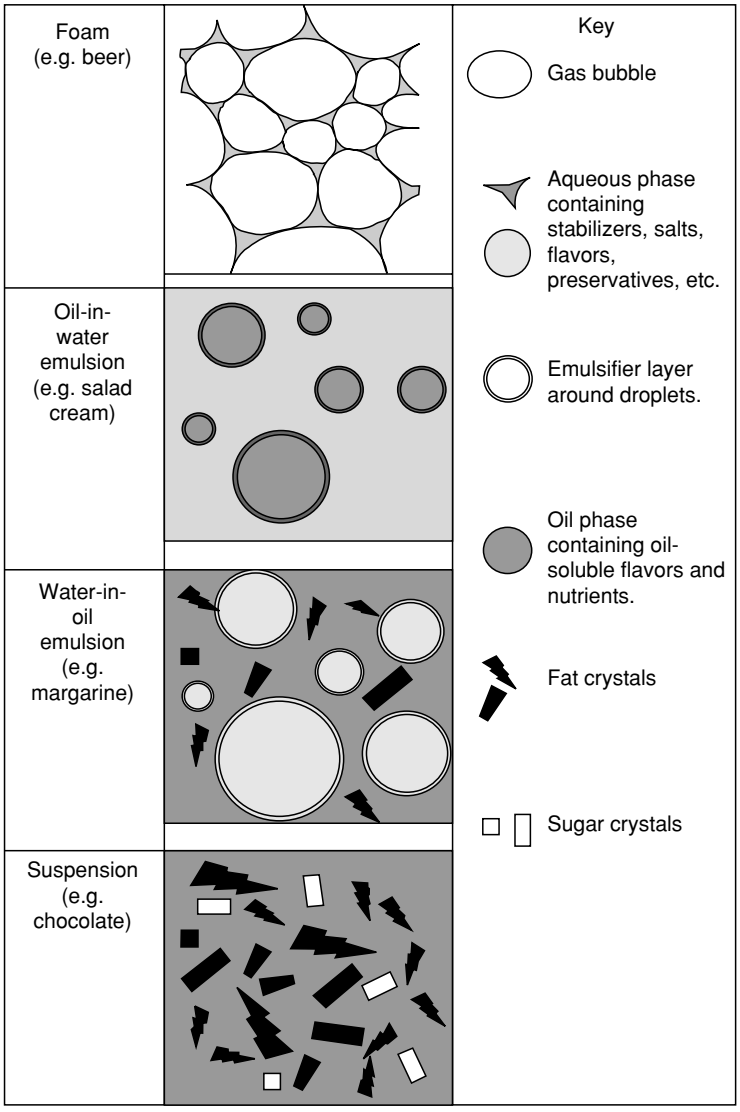

Figure 1 Types of food colloid and their structures.

contain more unsaturated fats than those of animal origin (with the exceptions of coconut and palm oils), and therefore tend to be liquid at room temperature. Fat is an essential nutrient, and the oil phase imparts a characteristic mouth feel to products such as cream, butter, and cheese. The oil phase is also used as a solvent for the more lipophilic emulsifiers, oil-soluble nutrients (e.g., oil-soluble vitamins and essential fatty acids) and some flavor and aroma compounds.

An important property of the oil phase is the solidification or crystallization of the fat. Solid fat in an emulsion has different rheological and textural properties to liquid oil, particularly in W–O emulsions. In addition, fat crystals can form structures that change the texture of fat-continuous products such as chocolate, butter, and cheese. Crystallization of the fat in the dispersed phase of ice cream and whipped cream allows the oil droplets to form solid networks that enhance structure and texture. The viscosity of the dispersed oil phase is also important for homogenization, as droplets made up of high-viscosity oil are less deformable and therefore more difficult to homogenize into smaller droplets.

Emulsifiers

The role of emulsifiers is to form an adsorbed layer around the emulsion droplets. This adsorbed layer lowers the interfacial tension, which aids emulsification, and stabilizes the droplets against flocculation and coalescence. Emulsifiers need to be surface-active, so that they will naturally assemble at an interface. They therefore need to be amphiphilic (i.e., possess both hydrophilic and hydrophobic parts). Emulsifiers can be either low-molecular-weight

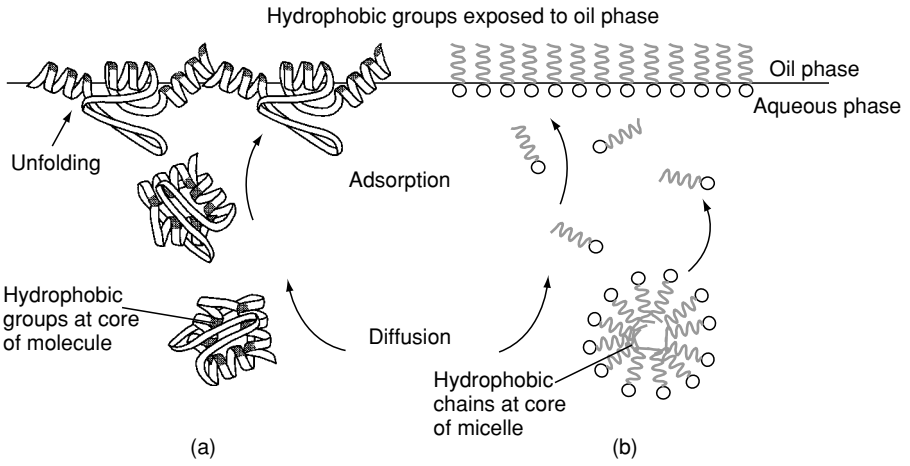

Figure 2 Emulsifier structure and adsorption of (a) proteins/polymers and (b) low-molecular-weight emulsifiers/surfactants. Hydrophobic groups are shaded in gray. Proteins form a folded structure in solution, keeping the hydrophobic amino acids away from aqueous solution. Upon adsorption, they unfold to expose these groups to the oil phase. Similarly, surfactants and emulsifiers form micelles, holding the hydrophobic chains in the core of the structure, which then dissociate and adsorb to the interface.

emulsifiers or macromolecular polymers, as shown in Figure 2. The low-molecular-weight emulsifiers have a small hydrophilic head group and one or more hydrocarbon chains; they include mono- and diglycerides, lecithins, polysorbates, and sugar esters. The most common polymeric emulsifiers in food systems are proteins, mainly from milk, but other sources include soya, meat, and fish, depending on the application. Other polymeric emulsifiers include modified cellulose and block copolymers, although block copolymers are not approved for food use.

Stabilizers

Although emulsifiers help stabilize emulsions, the term 'stabilizer' in the context of emulsions is specifically aimed at polymers which are added to the aqueous continuous phase to improve the stability of O–W emulsions. These polymers are often in the form of gums such as xanthan, guar, carageenan, etc. Addition of the polymers in sufficient concentrations to form a network results in the effective encapsulation of the emulsion droplets by the polymer. This immobilization prevents flocculation of the droplets and reduces creaming. Lower concentrations of polymer can induce flocculation by the depletion mechanism. Depending on the conditions, this flocculation can create a solid network of emulsion particles with enhanced viscoelastic properties (e.g., ice cream and whipped cream). Under different conditions, flocculation by polymers can induce rapid creaming or sedimentation.

Other Ingredients

From a purely physical standpoint, we tend to ignore components in an emulsion that do not contribute to its physical characteristics. However, in food emulsions the physical properties are not the only issue: flavor, sensory, and safety characteristics are also important. To this end, food emulsions contain salts, sugars, flavors, colors, and preservatives. However, these components highlight issues in colloids and emulsions that would not normally be of concern in other food products. Salt is perhaps the most important component as this can affect the solubility of functional ingredients such as proteins and polysaccharides. Specific interactions between, for example, calcium and caseins cause precipitation of the proteins. Salt also affects the electrostatic colloidal stability by screening the electrostatic repulsion forces between particles. Sugars are usually inert as far the physical properties of emulsions are concerned, but they can lower the dielectric constant of the continuous phase, which affects colloidal interactions, and certain sugars, such as trehalose, are known to protect proteins against denaturation during heating and drying. Flavors, colors, and preservatives are usually present at relatively low levels, and generally do not influence the physical characteristics. However, some components are volatile and partitioning can often occur, from one phase to the other. It is important to understand these processes, for example for efficient use of preservatives, so that they remain in the phase that allows them to achieve the desired function.

Formation

Formation is clearly a critical stage in the life of an emulsion. Many of the physical parameters of the emulsion are established here. If the emulsion is to remain stable for significant periods, then the emulsion droplets need to be small enough to slow down

the creaming and separation processes (typically < 1 μm diameter), which will be discussed later. The main factors which determine the 'quality' of the newly formed emulsion which will be discussed here are: (1) the adsorption of emulsifiers to the droplet surface; (2) the homogenization process; and (3) stability during formation.

Adsorption

Emulsion formation requires the dispersion of one phase into small droplets. This results in a massive increase in interfacial area between the dispersed and continuous phase. Classically, interfacial tension is defined as the work required to increase the surface area by a unit amount. The interfacial tension between the two phases is therefore extremely important. Interfacial tension forces act on droplets to minimize surface area, hence, liquid droplets form spheres in the absence of other forces. Lowering the interfacial tension means that less energy is required to increase the interfacial area by a unit value. This is achieved through the use of emulsifiers, which adsorb at the interface between the two phases and lower the interfacial tension by disrupting the hydrogen bonds between the water molecules at the interface. Emulsifiers can be simple lipid-like molecules or complex macromolecules such as proteins; however, they are all amphiphilic so that they naturally align themselves at the interface, as shown in Figure 2. The interfacial tension is determined by the structure of the emulsifier and the interfacial concentration, therefore the most effective emulsifiers are those whose structure allows the formation of a dense, close-packed interfacial layer. The dynamics are also very important because the homogenization process is rapid, so it is the dynamic interfacial tension over time scales < 1 s that is most important, not the equilibrium value. Rapid adsorption of the emulsifier to lower the interfacial tension will thus result in more effective emulsification of the dispersed phase during the homogenization process.

Homogenization

Homogenization is the process by which the dispersed phase is broken up into small droplets (*See* **Agitation**). Normally, rapid mixing of the ingredients is used to create a coarse premix. This is sufficient to break up the dispersed phase into large droplets, and allow adsorption of the emulsifiers prior to final homogenization. Figure 3 shows the three main methods of homogenization. Mechanical methods such as rotor–stator systems induce high shear fields to break up droplets. High-pressure homogenizers are now very common, and simply force the premix through a narrow orifice or valve at high pressures (typically 10–100 MPa or 1500–15 000 lb in^{-2}). Forcing the emulsion through a valve at high pressure creates turbulence and very high shear forces, thus breaking up the droplets. However, shear forces are not the most efficient way of breaking up droplets, as the shear field tends to lose energy rotating the droplet. Forcing the droplets through a simple orifice induces elongational flow forces that break up droplets more efficiently. The droplet size produced is, within certain boundaries, generally inversely proportional to the energy input during homogenization. More accurately, it is related to the energy dissipated into droplet break-up. This is affected by the interfacial tension, and viscosities of the dispersed and continuous phases. As continuous-phase viscosity increases,

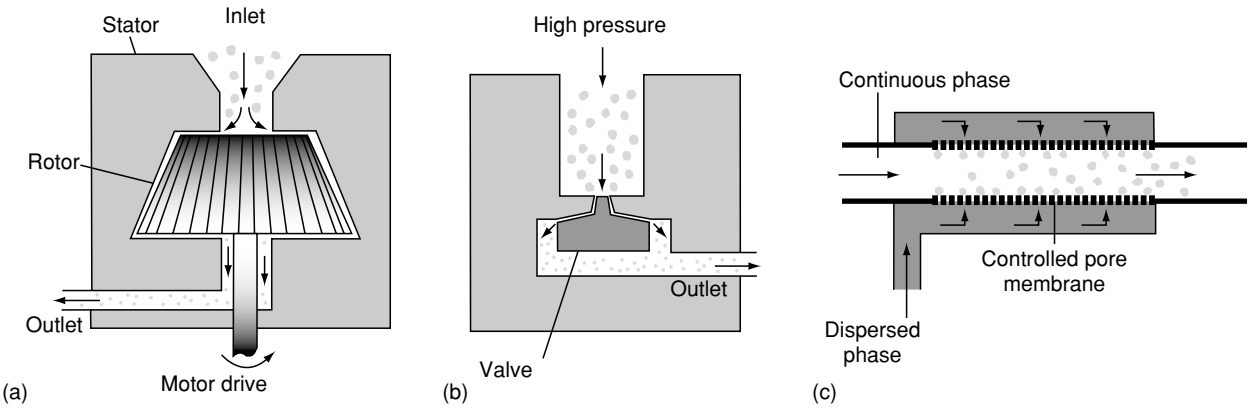

Figure 3 Basic principles of homogenization methods. (a) Rotor–stator device or colloid mill. The rotor, incorporating grooves or blades, spins at high speed, generating high shear fields to break up droplets. (b) High-pressure homogenizers force the droplets through narrow valves or orifices to break up droplets through high shear or elongational forces. (c) Porous membranes are used to extrude the dispersed phase into the flowing continuous phase.

droplet size will decrease, yet as dispersed-phase viscosity increases, the droplets become more difficult to break and droplet size consequently increases. The most energy-efficient process is the microporous approach, where the dispersed phase, or premix, is forced through a porous substrate into the continuous phase. The droplet size produced is dependent on the interfacial tension and the pore sizes. These microporous methods are still being developed and have not yet been widely utilized in the food industry, due to the low throughput.

Stability during Formation

Coalescence or fusing of droplets is a major instability process in emulsions, which will be discussed later. During formation, droplets are exposed to aggressive, turbulent conditions designed to break up droplets. These conditions also serve to increase the number and intensity of collisions between droplets. This increases the probability of recoalescence, particularly during the critical period when the adsorption process is not complete. Therefore, emulsifiers that adsorb slowly are particularly prone to this problem, and if the homogenization method is not optimized, then energy is wasted through recoalescence after breakup. This problem can be reduced through the control of the hydrodynamic conditions immediately after homogenization. This can be achieved by gently dispersing the droplets into a larger volume by either hydrodynamic methods or inert filler particles. This increases the interdroplet distance and reduces the probability of coalescence. However, in a commercial process it may not be convenient to form dilute emulsions and other strategies may be needed to produce acceptable emulsions.

In summary, by controlling the emulsifier type, concentration, and homogenization conditions, it is possible to control the droplet size distribution. It is the size and physical properties of the droplets which are the principal factors controlling the behavior of the whole emulsion, such as the rheology, stability, texture, and other quality characteristics of food emulsions, as described in the next section.

Characterization and Properties of Colloids and Emulsions

Particle/Droplet Size

The size of the dispersed particles or droplets is a key factor in the properties and stability of the dispersion. For example, for a product to appear turbid, the constituent particles must be able to scatter visible light, which means that their diameter needs to be at least 0.4 μm. However, very large particles tend to be less stable to the processes of flocculation (*See* **Flocculation**) and creaming, or impart a gritty mouth feel to the product. The particles in colloidal foods are generally very polydisperse, with a factor of 10 or more between the sizes of the largest and smallest particles present. It is thus very important to generate the optimum particle size distribution for each product, and to insure that it does not change during storage.

Typically, O–W food emulsions contain oil droplets of diameter in the range 0.5–2 μm, and W–O products such as spreads contain water droplets up to 20 μm. Solid dispersed particles, such as sugar crystals in chocolate, have diameters in the range 0.1–50 μm, although grittiness is detectable when the particles exceed about 20 μm.

Measurement of the particle size in colloidal foods or emulsions depends on the type of product and the size range of interest. Most foods can be examined microscopically, using optical methods for larger particles (> 3 μm) and electron microscopy in the submicron range. Microscopy is ideal for qualitative assessments of particle size and shape (*See* **Microscopy**: Light Microscopy and Histochemical Methods; Scanning Electron Microscopy; Transmission Electron Microscopy; Image Analysis), but quantitative measurement of a size distribution is less straightforward. Other direct methods such as sieving are convenient only for solid particles at the larger end of colloidal sizes. Indirect particle-sizing methods include optical and electrical techniques. The most common technique for droplet size determination is laser diffraction, based on the principle that the angle by which light is scattered from a particle is (inversely) related to its size. Measuring the intensity of light at a range of scattering angles enables the size distribution of the scattering particles (droplets) to be estimated. Another popular method for characterizing particle size distribution is the electric sensing-zone technique. In this method the particles are passed singly through a small gap between electrodes, and the volume of the particle is estimated from the change in capacitance during transit.

The light-scattering and electrical capacitance techniques both require very low particle concentrations (typically < 0.01% v/v), and recently there have been developments in methods that can characterize droplet size in concentrated emulsions. The principal techniques are acoustic, since ultrasound propagates readily through concentrated dispersions and detailed analysis of the frequency-dependent attenuation spectrum of ultrasound enables estimation of the particle sizes present. A similar technique is electroacoustics, where (charged) particles in a concentrated dispersion are made to oscillate in an electric field, which

generates sound waves whose characteristics are related to the particle charge and size.

Stability

Emulsions are inherently unstable and, given sufficient time, will eventually separate into bulk layers of oil and water. However, temporary stability can be achieved, and foods are engineered to remain comparatively stable for the required shelf-life. The mechanisms by which emulsions change in structure with time are shown in Figure 4. The figure also shows schematic drawings of the structural changes in the emulsions due to each type of instability. It is clear that creaming and flocculation involve the physical rearrangement of droplets into a layer or clumps (aggregates) respectively, whereas coalescence and disproportionation exhibit changes in the primary droplet size. Since emulsions are not thermodynamically stable systems, there are slow changes taking place from the start, and eventually the droplets grow to such an extent that the emulsion breaks into bulk layers of the component phases.

The interaction forces between droplets are crucial for the stability of the emulsion. If the droplets are highly charged, and there is a low ion concentration in the continuous phase, the droplets experience repulsive forces when they approach each other, and these forces keep them separated as individual

Instability type	Schematic mechanism	Factors promoting instability
Creaming Droplets move under gravity to the top (oil-in-water) or base (water-in-oil) of the container	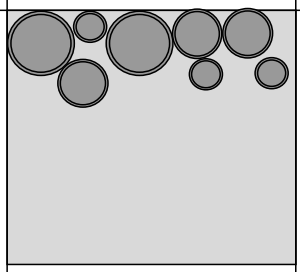	Large density difference between droplets and continuous phase; large droplets; low continuous-phase viscosity
Flocculation Droplets clump together into aggregates, but retain original droplet size	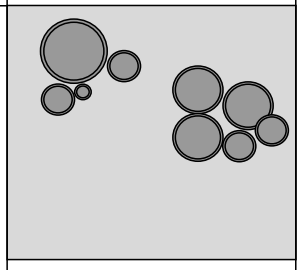	Attractive interactions between droplets; low charge and/or high ionic strength; presence of adsorbing polymer to bridge between droplets; presence of nonadsorbing polymers to induce depletion flocculation
Coalescence Droplets clump together and then merge to form larger droplets	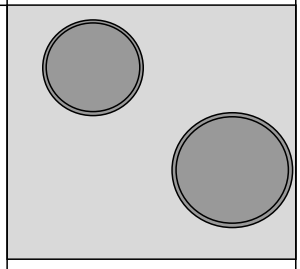	Flocculation or creaming to bring droplets together; an interfacial layer that ruptures easily
Disproportion-ation Swelling of larger droplets by molecular transfer of oil from smaller droplets, which eventually disappear	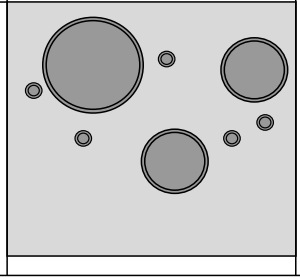	Polydispersity in droplet size; significant solubility of disperse phase (oil) in continuous phase; excess emulsifier; high interfacial tension

Figure 4 Emulsion instability processes.

particles. If an emulsion is stable to flocculation on a certain time scale, coalescence is also prevented because the droplets are repelled before the interfaces get near enough to merge. Certain additives used in food, e.g., polysaccharide 'stabilizers', may cause flocculation of the droplets in order to impart a particular texture, and then the interfacial composition is important in preventing coalescence of the adjoining droplets. The interfacial properties that are important in preventing coalescence of droplets in close contact are steric hindrance, interfacial viscoelasticity (of polymers and proteins), and interfacial mobility (of emulsifiers and surfactants).

Rheology

The rheological properties of emulsions are important functional properties, both during processing and in the final product where the texture is an important component of mouth feel. The response of the emulsion to mechanical stress may mimic simple liquids, as in dilute emulsions such as milk. At the other extreme, they may be designed to impart a solid-like texture, as in mayonnaise.

The rheological properties are controlled primarily by the concentration and size of the droplets, and by the composition of the continuous phase. (*See* **Rheology of Liquids**). Emulsions of low droplet concentration, where the droplets are unflocculated and the continuous phase is a simple liquid, will be of low, constant viscosity at a given temperature (Newtonian behavior). As the droplet concentration increases, and/or flocculation occurs, the emulsion becomes progressively more non-Newtonian. Its response to flow is now dependent on the magnitude of the applied stress, and the time scales of application. These characteristics are deliberately designed into many products, so the food rests on a plate without spreading, but is capable of flowing when dispensed. The apparent lack of response to a low stress (such as gravity) is sometimes attributed to the presence of a 'yield' stress, implying that up to a certain stress level the emulsion is solid-like and flows only when the applied stress exceeds a critical value.

Flavor

Controlling the flavor of foods is described elsewhere, but in emulsions the droplet composition and structure are additional variables. In many traditional foods, flavor components are present in both the oil and aqueous phases, and the full flavor is only released by droplet breakdown in the mouth. The droplet size and interfacial composition (emulsifiers) are also important factors.

Safety

Microbiological food safety relies on either sterilization or the existence of chemical conditions hostile to microbial growth. In an emulsion the foodborne bacteria can only grow in the aqueous phase, but the oil phase can affect the chemical environment for growth. This is because many antimicrobial agents, e.g., organic acids, are able to partition into the oil phase, thus reducing their preservative action in the aqueous phase. Food manufacturers are increasingly taking emulsion structure into account when formulating preservative systems.

Nutrition

The dietary impact of emulsions is the high fat content. Reducing fat intake is the principal target in the fight to reduce chronic heart disease in the developed world. Therefore one approach is to reduce the fat content in food emulsions, but retain the texture and flavor of the original product. A simple way is to increase the viscosity of the product with thickeners; however, the rheological behavior and hence perceived texture is very different. A better way is to produce dispersed droplets that behave like emulsion droplets, but with a lower fat content, by using gels, coacervates, or multiple emulsions (oil droplets containing water droplets). In margarines and spreads, large volumes of water can be dispersed into the fat continuous phase through the use of appropriate emulsifiers and processing. Another approach is to use fat replacers, which are not broken down and digested in the gut; however, use is limited due to other physiological side-effects.

Another important nutritional factor is the absorption of oil-soluble nutrients. Oil-soluble vitamins and antioxidants are crucial for health. They are thought to protect against heart disease and cancer. Their successful absorption is dependent on the processing of the raw materials, emulsion structure, and interfacial composition. Colloid science is now an important part of trying to understand the mechanisms involved.

See also: **Aerated Foods**; **Crystallization**: Basic Principles; **Emulsifiers**: Organic Emulsifiers; Phosphates as Meat Emulsion Stabilizers; Uses in Processed Foods; **Fats**: Production of Animal Fats; Uses in the Food Industry; **Flocculation**; **Homogenization**; **Margarine**: Types and Properties; **Milk**: Liquid Milk for the Consumer; Processing of Liquid Milk; Physical and Chemical Properties; **Rheology of Liquids**; **Stabilizers**: Types and Function

Further Reading

Becher P (ed.) (1983) *Encyclopedia of Emulsion Technology*. New York: Dekker.

Binks BP (ed.) (1988) *Modern Aspects of Emulsion Science*. Cambridge: Royal Society of Chemistry.

Dickinson E (1992) *An Introduction to Food Colloids*. Oxford: Oxford University Press.

Dickinson E (1998) Structure, stability and rheology of flocculated emulsions. *Current Opinion in Colloid and Interface Science* 3: 633–638.

Dickinson E and Stainsby G (1982) *Colloids in Food*. London: Applied Science Publishers.

Friberg S (ed.) (1976) *Food Emulsions*. New York: Marcel Dekker.

Robins MM (2000) Particle size analysis in food. In: Meyers RA (ed) *Encyclopedia of Analytical Chemistry*, pp. 4130–4153. Chichester: John Wiley.

Walstra P (1993) Principles of emulsion formation. *Chemical Engineering Science* 48: 333–349.

Wilde P J (2000) Interfaces: their role in foam and emulsion behaviour. *Current Opinion in Colloid and Interface Science* 5: 176–181.

Wilde PJ (2000) Physically modified proteins. In Doxastakis G and Kiosseoglou V (eds) *Novel Macromolecules in Food Systems*, pp. 161–180. Amsterdam: Elsevier.

COLON

Contents

Structure and Function

R J Levin, The University of Sheffield, Sheffield, UK
G I Sandle, University of Leeds, Leeds, UK

Structure

The adult human colon is about 100 cm long *in situ* and is divided anatomically into the ascending, transverse, descending, and sigmoid colon (**Figure 1**). Longitudinal muscle bands (taeniae coli) in its wall cause haustrae – characteristic puckerings and sacculations. Between these, the wall is thrown into crescent-shaped semilunar folds. The lumen is lined by a mucosa of simple columnar epithelium (colonocytes) containing numerous goblet cells. These cells secrete a mucus assumed to act primarily as a lubricant and protectant, allowing the solid and semisolid material to pass easily along the organ without damaging the epithelium. As the mucus contains secretory immunoglobulins (IgA) it also helps to protect the mucosa from luminal pathogens.

Unlike the small bowel, the colon is devoid of villi but has numerous straight tubular glands or crypts (sometimes called pits) (**Figure 2**) with openings on to the near-flat epithelial surface. These crypts are lined with cells similar to the surface colonocytes. At the crypt base, undifferentiated cells form a stem cell population for all the other epithelial cells of the large bowel. These divide by mitosis to replace those shed from the surface. The crypts also contain a few enteroendocrine cells. The lamina propria, containing plasma cells, macrophages, lymphoid aggregates, and mast cells, underlies the mucosa and extends down to the muscularis mucosae, a thin layer of inner circular

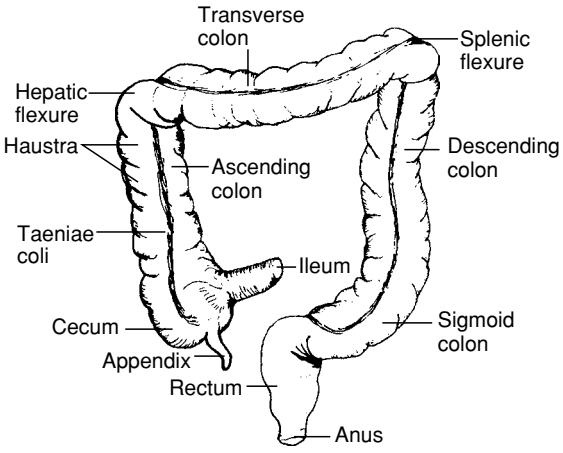

Figure 1 Diagram showing the anatomical regions of the human large intestine. Reproduced from Colon: Structure and Function, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

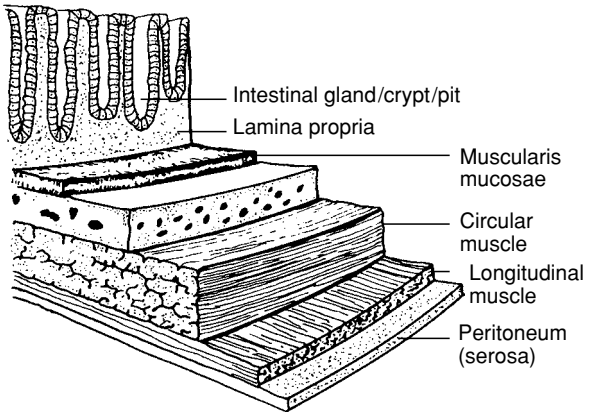

Figure 2 Schematic diagram showing the various major tissue structures of the human large intestine. Reproduced from Colon: Structure and Function, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

and outer longitudinal smooth muscle fibers. Beneath this is the submucous coat containing areolar tissue and the blood and lymph vessels serving the epithelium. The smooth muscle investing the organ consists of an inner circular and the outer longitudinal layer, the bulk of which is arranged in three longitudinal bundles (taeniae).

Nervous Innervation

The autonomic supply to the colon has three divisions: the extrinsic (1) sympathetic and (2) parasympathetic nerves link up with (3) the enteric supply in the bowel wall. The latter functions as an integrating system and consists of two ganglionated plexi. One is called the myenteric plexus, positioned between the circular and longitudinal smooth-muscle layers, and the other the submucosal plexus, lying in the submucosa under the muscularis mucosae. More recently, a third plexus, called the mucosal plexus, has been identified. The mucosal and submucosal plexi are involved in controlling epithelial function but the myenteric plexus interconnects to coordinate motility with absorptive functions by local and extrinsic reflex pathways. Colonic movements include segmentation contractions and peristaltic waves similar to those in the small intestine.

Luminal conditions are probably sensed by afferent nerve endings in the mucosa and the afferent information processed at the enteric plexus, cord, or brain level. The efferent output is then fed back via the extrinsic nerves to the muscle and epithelium.

Function

The colon has three major functions: (1) absorption; (2) secretion; and (3) bacterial fermentation.

Absorption

Salt and water While the jejunum and ileum are the major sites for the absorption of the salt and water ingested into and secreted by the upper alimentary tract, it is the task of the colon to absorb the remaining fluid. Many authors describe this as the 'salvage' function of the colon. Its efficiency can be judged by the estimate that daily, approximately 1000–2000 ml of the fluid enters the colon at the ileocecal valve, while less than 100 ml is lost in the feces. The colon is adapted to extract water and salts from the colonic contents to create initially semisolid and finally solid contents that become the feces. First, it has a much slower rate of luminal transit than is observed in the small intestine, allowing its contents a greater contact time with the mucosa. Second, the colonocytes can generate very large osmotic forces to extract the fluid from the lumen. Third, the tight junctions between the colonocytes are much less permeable to ions and water than those of the small intestine. Fluid absorptive capacity is determined primarily by sodium ion (Na^+) and chloride ion (Cl^-) absorption. The colonocytes actively transfer Na^+ into the convoluted lateral spaces between the cells and water follows osmotically, mainly through the tight junctions. (*See* **Sodium**: Physiology.)

Segmental differences The mammalian colon is not a single organ with a uniform distribution of electrolyte transport along its length. It is well known that different species exhibit segmental differences in their ion transport and also that the different segments of the colon, in the same animal, display differences in ion transport. The most studied animal colons have been those of the rabbit and rat. Although there are obvious similarities in the colonocyte functions in these two species with human colonocytes, neither matches exactly that of the human colon. *In vivo* studies in the human colon indicate that Na^+ and water are primarily absorbed in the ascending and transverse colon, with relatively little or no net absorption in the rectum. The human descending colon (like that of the rabbit) possesses colonocytes with apical membrane channels sensitive to amiloride which abolishes the electrogenic Na^+ absorption (and short-circuit current or net electrogenic ion transfer *in vitro*). In the ascending colon, however, the Na^+ absorption is mainly via an amiloride-insensitive or neutral transfer mechanism(s).

Sodium ion absorption mechanisms Active Na^+ absorption is accomplished by two mechanisms: (1) electrogenic Na^+ transfer, and (2) neutral NaCl transfer.

Electrogenic Na^+ transfer The Na^+ ion moves into the colonocyte across its apical membrane through channels (that can be blocked by amiloride) under the influence of a favorable electrochemical gradient as the inside of the colonocyte is negative to the lumen and has a low Na^+ concentration. It is then pumped out of the cell across the basolateral membrane by the agency of Na^+, K^+-ATPase (the sodium pump or Na^+, K^+-activated adenosine triphosphatase), creating an electric current (rheogenic transfer) and a positive potential difference at the serosa. The ATPase, and thus current and Na^+ transfer, can be inhibited by the cardiac glycoside, ouabain. Mineralocorticoids, especially aldosterone, enhance the entry of Na^+ across the apical membrane by the electrogenic route and thus also increase the absorption of water.

Neutral NaCl transfer This may represent coupled $Na^+ - K^+ - 2Cl^-$ cotransport, coupled $Na^+ - Cl^-$ cotransport, or the coupling of dual ion exchanges such as Na^+ for H^+ and Cl^- for HCO_3^- or OH^-.

Chloride ion absorption mechanisms Chloride ions can be absorbed passively or actively.

Passive absorption This is via an electrical potential difference (pd)-dependent process as the serosal side of the intestine is positive with respect to the lumen.

Active absorption This occurs either as neutral NaCl absorption (Na^+-dependent process) or as neutral $Cl^- - HCO_3^-$ exchange (Na^+-independent process). It has been estimated that in the human colon *in vivo* the $Cl^- - HCO_3^-$ exchange accounts for about 25% of overall Cl^- absorption, the other 75% being via the favorable electrical pd mechanism via the paracellular route.

Potassium ion absorption mechanisms Active K^+ absorption occurs in the distal colon of rat and rabbit and has been shown to be an electroneutral process which operates independently of Na^+ and Cl^-. In rabbit, this K^+ absorptive process seems to reflect a K^+–H^+ exchange which can be unmasked by the presence of ouabain or the removal of Na^+. Recent studies in rat distal colon have identified K^+-ATPase as the mediator of this apical K^+–H^+ exchange, although it is unclear whether a similar mechanism operates in human colonocytes. There is emerging evidence that K^+ movement across the basolateral membrane reflects K^+ recycling via K^+ channels, which are beginning to be characterized in biophysical and molecular terms. (*See* **Potassium**: Physiology.)

Calcium and magnesium absorption While considerable amounts of these ions are present in the lumen of the colon, most are unavailable for absorption as they are bound to insoluble macromolecules. Moreover, the total removal of the human colon (colectomy) does not upset the balance for these ions. There is evidence from rat experiments that the colon has a Ca^{2+} active transport system sensitive to enhancement by vitamin D and to low calcium intake. In the case of magnesium, other experiments with rats have indicated that the ion is absorbed passively, dependent on the electrochemical gradient and the solvent drag of fluid absorption. (*See* **Cholecalciferol**: Physiology; **Magnesium**.)

Secretion

The colon secretes the anions Cl^- and HCO_3^- and the cations K^+ and H^+.

Bicarbonate secretion In the human, approximately 70 mmol HCO_3^- leaves the ileum to enter the colon every day and although the organ secretes bicarbonate, only a few mmol appear in the feces. This is because the secreted HCO_3^- is titrated by the H^+ ions manufactured by the bacterial population of the colon. In diarrheal states, when there is an enhanced secretion of bicarbonate, stools can contain an increased amount of bicarbonate.

The mechanisms and control of bicarbonate movements are complex. Other ions move in association with, or in exchange for, bicarbonate. Bicarbonate can come from the plasma, passing between and through cells, and can also come from the intracellular hydration of CO_2, Hydrogen ions, produced from cellular metabolism, can be secreted into the lumen to influence the production and neutralization of bicarbonate ions. Net bicarbonate secretion is accomplished by a combination of bicarbonate and hydrogen ion transport as the blood, cell cytoplasm, and luminal contents contain bicarbonate and CO_2. A primary source of HCO_3^- and H^+ ions appears to be from the catalyzed hydrolysis of CO_2, as Na^+ and Cl^- absorption is modulated by ambient CO_2. The colonocytes are rich in carbonic anhydrase which can supply bicarbonate and H^+ ions at high catalyzed rates. The enzyme may also be directly involved in ion transfer. Inhibition of carbonic anhydrase reduces colonic absorption.

Chloride ion secretion The colon actively secretes Cl^- ions. The intracellular Na^+ and K^+ of the colonocyte are regulated by the basolateral Na^+–K^+ transport ATPase pumping out Na^+ in exchange for K^+. The K^+ can diffuse out across the membrane by a K^+ channel. This mechanism keeps the intracellular

Na^+ low and the K^+ high, and produces a negative intracellular pd. The uptake of Cl^- ions is by the coupled Na^+–$K^+ - 2Cl^-$ cotransport carrier in the basolateral membrane, and is sensitive to loop diuretics such as furosemide (frusemide) and bumetanide. Chloride ions leave the cell via apical Cl^- channels, under the influence of their intracellular–extracellular concentration gradient and the negative intracellular pd. These Cl^- channels are normally closed and are activated to open by increases in the intracellular calcium ion (Ca^{2+}) and/or cyclic adenosine monophosphate (cAMP) levels brought about by a variety of neurotransmitters, secretagogues, hormones, and paracrine agents. Because the efflux of the Cl^- ions depolarizes the apical membrane, the movement of Cl^- would gradually cease. To maintain the driving force for the Cl^- exit, an increase in the K^+-conductance of the basolateral membrane occurs shortly (20 s) after the increase in apical membrane Cl^- conductance. The exit of K^+ ions across the basolateral membrane induces a hyperpolarization and, because of electrical coupling between the basolateral and apical membranes, the hyperpolarization maintains the pd across the apical membrane. Thus the stimulation of Cl^- secretion in colonocytes (as in many other Cl^--secreting cells) requires the coordination of four independent systems, the apical Cl^- channels, the basolateral Na^+–$K^+ - 2Cl^-$ cotransporter, K^+ channels, and Na^+, K^+-ATPase. Observations in animals have indicated that secreting cells are located predominantly in the crypts, although surface colonocytes are also capable of secreting Cl^- ions.

Potassium ion secretion Although it is known that the colon absorbs and secretes K^+, the cellular sites have not yet been defined. Surface cells of rabbit colon absorb and secrete K^+, but it is not known whether such processes are also found in the cells of the crypts. In rats and rabbits fed high-potassium diets or given aldosterone, K^+ secretion is accomplished first by the active uptake of K^+ at the colonocyte's basolateral membrane by ouabain-sensitive Na^+-K^+–ATPase and Na^+-K^+–$2Cl^-$ cotransport. Its exit across the apical membrane is then by barium-sensitive K^+ channels that appear to be under the control of cAMP and Ca^{2+}, since raising their cAMP and or Ca^{2+} levels in the colonocytes increases the secretion of K^+. Adrenergic agonists also increase the secretion of K^+, which in rabbit colon is mediated by β_1-receptors. The mechanism involves both increases in the basolateral pump activity and in the conductance of the apical membrane to K^+. In diarrhea associated with colonic secretion (e.g., colitis) K^+ loss in the stools can be appreciable.

Short-chain fatty acids (SCFAs) Surprisingly, the major anion of the colonic contents comes from SCFAs (acetate, butyrate, propionate, etc.). These originate predominantly from the bacterial metabolism of carbohydrate passing into the colon from the ileum. In the human colon, the daily production of SCFAs is estimated to be greater than 300 mmol, while the fecal excretion is only 10 mmol. This indicates that 97% of the SCFAs are removed by the colon. Their absorption is rapid and they enhance the absorption of Na^+. (*See* **Essential Fatty Acids**.)

The SCFAs are weak acids that can exist in solution as two species, the undissociated molecule (acid form HA) and the dissociated anion (A^-). The balance of the species depends on the environmental pH and the p*K* values of the acid under question. In the colon, the luminal pH is normally approximately 7 and the SCFA p*K* approximately 4.8, so that the predominant species is the ionized or anionic form. The pathways of absorption of SCFAs across the colon are complex because several mechanisms can operate. These include the following: (1) the generation of luminal CO_2, either from bacteria or from colonocytes, to provide H^+ ions which combine with the anionic form of SCFA (A^-) in an acid microclimate, creating the undissocated acid form (HA) which could diffuse across the apical membrane; (2) luminal H^+ ions from the colonocytes exchange with Na^+, forming the HA; (3) the ionic form, A^-, could diffuse across the tight junctions between the cells; (4) anion exchange mechanisms (e.g., HCO_3^-–A^-) allow the movement of the anionic form across the apical membrane.

In the human colon it has been proposed that approximately 40% of SCFA absorption can be accounted for by the Na^+–H^+ exchange. As SCFA absorption is also associated with a rise in luminal HCO_3^-, the SCFA–HCO_3^- exchange mechanism is suggested to occur. In theory, changes in luminal pH should affect SCFA movements and alterations in acid–base balance should affect their absorption. However, the mucus and unstirred layers at the apical pole of the colonocytes act as a buffer and maintain the surface acid microclimate in the face of considerable changes in the luminal pH.

SCFAs increase the absorption of NaCl partly through the fatty acids acting as metabolic substrates for the colonocytes and partly as a result of the SCFAs directly influencing the transport mechanisms of the apical membranes. Paradoxical results have also been found in animal colons used *in vitro*, such as SCFAs inhibiting Cl^- absorption and in other cases actual secretion of SCFAs by colonocytes. Other studies have shown that SCFAs also cause electrogenic secretion of Cl^- by colonocytes which can be recorded both *in vivo* and *in vitro*. As the SCFAs have been

claimed to be a fundamentally important metabolic fuel for the colonocytes, their reduction or near absence in human conditions such as famine (starvation) and malnourishment (marasmus and kwashiorkor) has been suggested as a major cause of colonic dysfunction and – its consequence – diarrhea. Other studies have shown that colonic and rectal secretory function in starved or malnourished rats can be enhanced, which would again predispose to diarrhea. There are many unanswered questions and unsolved problems in relation to SCFA and colonic ion and fluid transfer. (*See* **Kwashiorkor**; **Marasmus**.)

Bacterial Fermentation

Composition of colonic microflora The bacteria of the human gastrointestinal tract form a complex ecosystem; over 400 species have been identified in the feces of a single subject. In healthy humans, the upper gastrointestinal tract is sparsely populated by bacteria as most are killed by gastric acidity (except acid-loving forms such as *Streptococcus, Staphylococcus*, and *Lactobacillus*) and the passage of intestinal contents is rapid. The bacterial flora of the small intestine is between 10^3 and 10^4 colony-forming units (CFU) per ml of content (mainly Gram-positive anaerobes); this increases to 10^6–10^7 CFU ml^{-1} in the distal ileum and Gram-negative bacteria now become dominant. After the ileocecal sphincter, the bacterial concentrations increase enormously, reaching between 10^{11} and 10^{12} CFU ml^{-1} of fecal material, which can amount to approximately 1.5 kg wet weight of bacteria. Viable bacteria are said to make up one-third of the fecal dry weight. The major species are *Bacteroides, Clostridium, Eubacterium, Peptococcus, Bifidobacterium, Streptococcus* and *Fusobacterium*. (*See* ***Clostridium***: Occurrence of *Clostridium perfringens*; Occurrence of *Clostridium botulinum*; **Microflora of the Intestine**: Role and Effects; ***Staphylococcus***: Properties and Occurrence.)

Metabolic capacity of colonic bacteria The metabolic capacity of the colonic bacteria is diverse. Luminal contents are potential substrates for bacterial enzymic transformation by hydrolysis (examples are estradiol glucuronide, cyclamate), dehydroxylation (bile acids, amino acids), or reduction (unsaturated fatty acids, food dyes, benzaldehydes). The bacteria can also synthesize vitamins such as vitamin K, biotin, and vitamin B_{12}. Antibiotic treatment of humans is known to lower the plasma levels of the first two, suggesting that the bacteria may play a role in supplying some of the daily requirements for these vitamins. (*See* **Biotin**: Physiology; **Cobalamins**: Physiology.)

Carbohydrates are fermented by the bacterial enzymes to SCFA. The major carbohydrate sources are plant cell wall polysaccharides (cellulose, pectins, hemicellulose), starch, intestinal mucus, and mono- and disaccharides. Depending on the amount of fiber in the diet, between 20 and 70 g of carbohydrate is estimated to be fermented per day in the human colon. The SCFAs produced by the fermentation process are absorbed by the colonocytes, for which they represent an important metabolic source and act as a promoter of NaCl and thus fluid absorption. (*See* **Carbohydrates**: Digestion, Absorption, and Metabolism; **Dietary Fiber**: Physiological Effects.)

Another important feature of the metabolic capacity of the colonic bacteria is the potential for creating candidate carcinogens and mutagens (e.g., nitrosamines) that may play a role in carcinogenesis, especially of the colon. In the USA this is the most frequently diagnosed cancer of internal organs. There is some evidence that the western diet (high in beef, fat, and protein, and low in fiber) is a possible cause of the much higher incidence of colon cancer than is found among Africans, Asians, and South Americans. (*See* **Cancer**: Diet in Cancer Prevention; **Nitrosamines**.)

Migrants show a cancer incidence that approximates the prevailing rate of their residence rather than their birthplace. There is, however, another side to the bacterial coin of the colon – the inactivation of carcinogenic compounds by bacteria such as *Escherichia coli, Lactobacillus* spp., and *Bacteroides* spp.

Urea, ammonia, and nitrogen Some 6–9 g of urea (approximately 20% of the daily urea synthesis) is broken down in the human intestine, mainly by the urease of the colonic bacteria. As only 0.4 g of this urea enters the colon from the ileum, and there is no urea and hardly any of its breakdown metabolite ammonia in the feces, most of the urea apparently has to enter the colon by another pathway in order to be metabolized. Unfortunately, the urea permeability of the cleansed human colon appears to be very low and excludes its transmucosal diffusion as a major route. There is evidence that the low permeability is caused by removal of the colon contents which normally increase the permeability of the mucosa to urea; thus the entry path of urea into the untouched colon is probably transmucosal. The catabolism of the urea load produces 200–300 mmol of ammonia and 100–150 mmol of bicarbonate per day. Practically all the ammonia (99%) is absorbed across the mucosa by nonionic diffusion for at colonic pH (6–7) ammonia is largely present as the ammonium ion (NH_4^+). The absorbed ammonium is transported by the portal system to the liver, where it enters the nitrogen pool and is synthesized to urea or amino acids. This metabolism of urea is of little importance to human

nutrition but the enterohepatic circulation of ammonia is important in the genesis of hepatic encephalopathy. In hepatic disease, the liver can fail to detoxify the ammonia and it then enters the general circulation to produce neuropsychiatric disturbances and even coma due to its cerebrotoxic action on the brain. (*See* **Liver**: Nutritional Management of Liver and Biliary Disorders.)

The origin of ammonia in the large intestine is mainly from the breakdown of urea, as described above, but it can also be produced by the bacteria from nonurea sources such as amine and amide nitrogen of proteins, peptides, and amino acids and from plasma glutamine. The colonic bacteria not only produce ammonia but can also utilize it as a nitrogen source for their own synthesis of amino acids and proteins. The oral administration of poorly digested carbohydrates is known to increase the fecal loss of nitrogen. The probable explanation for this is that the nondigestible carbohydrates are metabolized by the bacteria in the colon as a preferential energy source and stimulate the production of new bacteria which are then lost in the feces.

Effect of diet on microflora While there have been numerous studies both in animals and in humans on the effects of diet on the balance of the colonic microflora, the data are conflicting with regard to the ability of the diet to alter specific components of the adult bacterial population. For example, different studies on individuals fed on omnivorous and then herbivorous diets have reported no change, small increases, or decreases in fecal counts of *Bacteroides*. Subjects in the UK and the USA eating the western diet had more *Bacteroides* and fewer enterococci and other aerobics than subjects eating a vegetarian diet in Uganda, India, and Japan. Overall, it appears that eating a meat-poor, high-carbohydrate diet increases the fecal counts of aerobic bacteria, and decreases a number of anaerobic bacteria. (*See* **Vegetarian Diets**.)

Effect of dietary intake on colon contents Since the production of colonic gas (see below) and SCFA is so dependent upon the bacterial fermentation of unabsorbed carbohydrate, much attention has been paid to the proportion of dietary carbohydrate that remains undigested, enters the large intestine and is then broken down and 'salvaged.' Starch, rather than dietary fiber, is the major substrate. In healthy adults it is estimated that usually less than 5% of ingested starch is unabsorbed and enters the colon, although this may be increased to 12% by food processing.

Colon gas production The major components of intestinal gas are nitrogen, oxygen, carbon dioxide, hydrogen, and methane. Gases such as hydrogen sulfide, volatile amines, organic acids, mercaptans, and indoles, collectively amount to less than 1% of the total but as they account for the characteristic unpleasant smell of the flatus they are socially important. The colon is the only source for the production of hydrogen, which is used by the bacteria, expelled in the flatus or absorbed and excreted in the breath. Dietary substrates are the major source for its production as it is low in the fasting subject. Fermentable carbohydrates such as raffinose and stachyose (oligosaccharides found in beans), lactulose, and lactose (in subjects with hypolactasia) all increase the production of hydrogen. Measures of breath hydrogen can be used as an index of its colonic production and also as an indication of when food reaches the colon.

Colonic Function and Diarrhea

It is clear from the various accounts of colonic function that if its absorptive function is impaired or its secretory activity enhanced or both occur, then diarrhea results. A large number of factors can alter the basic net absorptive 'tone' of the colon to a net secretory one and thus induce diarrhea. Examples include bacterial toxins (cholera, *E. coli* STa (heat-stable enterotoxin) and LT (heat-labile enterotoxin)), mediators of inflammatory processes (eicosanoids, histamine, kinins, substance P), neurotransmitters (acetylcholine, vasoactive intestinal peptide (VIP), serotonin), hormones (neurotensin, glucagon, cholecystokinin) and specific endogenous luminal contents such as bile acids. Recently, studies on rats and mice to elucidate the enigma of the terminal diarrhea suffered by human victims of famine and severe malnutrition have shown that starvation induces hypersecretory responses of the enterocytes and colonocytes to neurotransmitters and agents that activate secretion. The mechanism of the hypersecretion is not yet understood but one locus is probably at the apical membrane after the release of the second messengers that activate the Cl^- secretion.

See also: **Biotin**: Physiology; **Cancer**: Diet in Cancer Prevention; **Carbohydrates**: Digestion, Absorption, and Metabolism; **Cholecalciferol**: Physiology; ***Clostridium***: Occurrence of *Clostridium perfringens*; Occurrence of *Clostridium botulinum*; **Cobalamins**: Physiology; **Dietary Fiber**: Physiological Effects; **Essential Fatty Acids**; **Kwashiorkor**; **Marasmus**; **Microflora of the Intestine**: Role and Effects; **Sodium**: Physiology; ***Staphylococcus***: Properties and Occurrence; **Vegetarian Diets**

Further Reading

Binder HJ and Sandle GI (1994) Electrolyte absorption and secretion in the mammalian colon. In: Johnson LR (ed.)

Physiology of the Gastrointestinal Tract, 3rd edn. pp. 2133–2171. New York: Raven Press.
Goldin BR, Lichtenstein AH and Gorbach SL (1988) The roles of the intestinal flora. In: Shils ME and Young VR (eds) *Modern Nutrition in Health and Disease*, pp. 500–515. Philadelphia: Lea and Febiger.
Kasper H and Goebell H (eds) (1982) *Colon and Nutrition*. Falk symposium no. 32. Lancaster: MTP Press.
Lebenthal E and Duffey ME (eds) (1990) *Textbook of Secretory Diarrhea*. New York: Raven Press.
Lomax RB, Warhurst G and Sandle GI (1996) Characteristics of two basolateral potassium channel populations in human colonic crypts. *Gut* 38: 243–247.
McBurney MI (1991) Passage of starch into colons of humans: quantitation and implications. *Canadian Journal of Physiology and Pharmacology* 69: 130–136.
Roediger WEW (1990) The starved colon – diminished mucosal nutrition, diminished absorption, and colitis. *Diseases of the Colon and Rectum* 33: 858–862.
Sandle GI, McNicholas CM and Lomax RB (1994) Potassium channels in colonic crypts. *Lancet* 343: 23–25.
Wrong OM, Edmonds CJ and Chadwick VS (1981) *The Large Intestine – Its Role in Mammalian Nutrition and Homeostasis*. Lancaster: MTP Press.

Diseases and Disorders

D N Moskovitz and Y-I Kim, University of Toronto, Toronto, Ontario, Canada

Diet and Nutrition in Inflammatory Bowel Disease

The term 'inflammatory bowel disease' (IBD) refers to primarily two disorders, ulcerative colitis (UC), and Crohn's disease (CD). Both diseases are chronic, relapsing disorders of unknown etiology. CD and UC, while different disease entities, share certain demographic, epidemiologic, and therapeutic features.

Crohn's Disease

CD is a chronic inflammatory disorder that can affect the small intestine and/or large intestine. It occurs throughout the world, with a prevalence of 10–100 cases per 100 000 people. The disorder occurs most frequently among people of European origin, is three to eight times more common among Jews than among non-Jews, and is more common among whites than nonwhites. Although the disorder can begin at any age, its onset most often occurs between 15 and 30 years of age. There appears to be a familial aggregation of patients with CD such that 20–30% of patients with CD have a family history of IBD.

Clinical presentation Diarrhea, abdominal pain, weight loss, and fever are the typical clinical manifestations for most patients with CD (**Table 1**). Diarrhea may have multiple causes, including excessive fluid secretion and impaired fluid absorption by the inflamed small or large bowel, bile salt malabsorption due to an inflamed or resected terminal ileum, and steatorrhea related to extensive ileal disease or ileal resection with loss of bile salts. Bacterial overgrowth from small bowel strictures, enterocolonic fistulae, and extensive jejunal disease may also lead to steatorrhea. Crampy abdominal pain is a common manifestation of CD. The transmural nature of the inflammatory process leads to fibrotic strictures that often lead to repeated episodes of the small bowel or, less commonly, colonic obstruction. Gross bleeding is much less frequent than in ulcerative colitis (except for some patients with Crohn's colitis). Transmural inflammation is also associated with the development of sinus tracts that can lead to serosal penetration and bowel-wall perforation. This complication may be associated with an acute presentation of localized peritonitis with fever, abdominal pain, and tenderness, and often a palpable mass on physical examination. Penetration of the bowel wall often presents not as an acute abdomen but as an indolent process related to fistulization. The actual clinical manifestation of the fistula depends on the area of involvement adjacent to the diseased bowel segment. Symptoms and signs related to perianal disease occur in up to one-third of patients with CD. These include perianal pain and drainage from large skin tags, anal fissures, perirectal abscesses, and anorectal fistulae. Weight loss and fever are the primary systemic symptoms in CD. There are many extraintestinal manifestations of CD (**Table 2**).

Pathology Inflammation, which may or may not be accompanied by noncaseating granulomas, extends through all layers of the gut wall. The inflammatory

Table 1 Clinical differences between CD and UC

Symptom	*CD*	*UC*
Diarrhea	+	++
Abdominal pain	++	+
Weight loss	++	–
Fever	+	+
Fistula	+	–
Stricture	+	–
Bleeding	+	++

process is often discontinuous, with normal bowel separating portions of diseased bowel. The key pathological feature of CD is an inflammatory process that extends through all layers of the bowel wall. Microscopic examination reveals (1) hyperplasia of perilymphatic histiocytes, (2) diffuse granulomatous infiltration, (3) discrete noncaseating granulomas in the submucosa and lamina propria, (4) edema and lymphatic dilation of all layers of the gut, and (5) monocytic infiltration within lymph nodules and Peyer's patches on the serosal surface of the bowel. Patients with CD can be divided into those with small bowel disease alone (30%), those with both small and large bowel involvement (50%), and those with disease involving only the colon (20%). When CD involves the small bowel, 80% of the time, the terminal ileum is involved. When the colon is involved in CD (Crohn's colitis), all segments of the colon can be affected. Although absence of rectal disease is more characteristic of CD than of ulcerative colitis, the rectum is involved in half those cases of CD in which colonic disease exists.

Etiology The etiological agent responsible for IBD has not yet been determined. One hypothesis explaining the developing of CD begins when the gastrointestinal immune system becomes exposed to a mucosal antigen, perhaps even an antigen normally present within the lumen. On this occasion, the antigen does not evoke the typical antigen-specific suppressor T-cell activity, mucosal unresponsiveness. Rather, because of an antigen-specific immunoregulatory defect, it evokes helper T-cell activity and sets in play an ongoing immune response. This immune response as an epiphenomenon leads eventually to the development of self-antigens and the appearance of autoantibodies. Subsequently, in an attempt to downregulate the antigen-specific response, antigen-nonspecific suppressor T-cells appear. Initially, these antigen-nonspecific suppressor T-cells may prevent disease progression; however, they are gradually depleted, leaving the unregulated antigen-specific helper T-cell activity to predominate. This unregulated antigen-specific immune response leads to the production of lymphokines, which stimulate migration of inflammatory and cytolytic cells to the region. Through this process, the microscopic and gross morphological changes of IBD are manifest.

Table 2 Extraintestinal manifestations of IBD

	CD	*UC*
Conjunctivitis	+++	+
Iritis, uveitis, episcleritis	+++	+
Ankylosing spondylitis	+	+++
Arthritis	+++	+
Erythema nodosum	+++	+
Pyoderma gangrenosum	+	+++
Aphthous ulcers of the mouth	+++	+
Cancer	+	+++
Toxic megacolon	+	+++
Fissure, fistula	+++	+
Pericholangitis	+	+
Primary sclerosing cholangitis	+	+++
Gallstones	+++	–
Bile duct carcinoma	+	+
Chronic active hepatitis	+	+
Amyloidosis	+	–
Nephrolithiasis	+++	–

Treatment The treatment of IBD has evolved over time (Table 3). Historically, therapeutic approaches were based on current beliefs of the disease etiology. The treatment of inflammatory bowel disease during the early 1900s included vaccination, dairy diets, astringents, antiseptics, opium, silver nitrate, and rectal instillations of boracic acid. Early treatment often consisted of rectal insufflation of hydrogen peroxide, irradiation of the right lower quadrant of the abdomen, liver extract, and injections of peptides. In the early 1900s, food was implicated in the cause of inflammatory bowel disease. As a result, dietary programs emerged as treatments. During the 1930–50s, clinical observations led to the use of psychotherapy and psychoanalysis. While such early treatment was a reflection of a lack of understanding of disease etiology, as theories of the pathogenesis of IBD evolved, so did the therapeutic modalities.

Medications have remained the mainstay of treatment. It was not until the 1940–50s that sulfonamides began to be utilized. Sulfasalazine, the first sulfonamide, consists of a 5-aminosalicylic acid (5-ASA) attached to sulfapyridine (an antibiotic). It was soon discovered that the active therapeutic moiety was the 5-ASA component. As a result, the 5-ASA group of drugs consisting of asacol, pentasa, and mesalamine was created, each drug differing in their type and place of delivery to the gastrointestinal

Table 3 Medications used in the management of IBD

Medication	*CD*	*UC*
5-ASA	R	R
Asacol, pentasa, mesalamine, osalazine	M	M
Antibiotics		
Metronidazole, ciprofloxacin	R	
Steroids		
Prednisone, budesonide	R	R
Immunomodulators	R	R
Azathioprine, 6-mp, cyclosporine, infliximib	M	M

M, used for maintenance; R, induces remission.

tract. The 1970–80s saw the emergence of antibiotics in the treatment of IBD. Studies have shown that metronidazole is effective in both inducing and maintaining remission in patients with IBD. Moreover, the use of ciprofloxacin has proven effective in inducing and maintaining remission in patients with CD. The discovery of adrenocorticotropic hormone and, subsequently, the glucocorticoids cortisone and prednisone in the 1950s provided a great impact on the treatment of IBD. As a scientific approach to the study of inflammatory bowel disease emerges, so do therapeutic interventions targeted at changes in the intestinal tract. As the study of the immune system evolved in the late 1800s, new techniques for the suppression of the immune response began to be developed. In the mid-1960s, 6-mercaptopurine, an immunosuppressant, was first utilized in the treatment of IBD. Decades later, the use of immunosuppressants cyclosporine and methotrexate became more prevalent.

Nutritional issues Malnutrition is common in patients with CD. It has been reported that up to 75% of hospitalized patients with CD are malnourished or may present with weight loss, low albumin, anemia, negative nitrogen balance, or vitamin D deficiency. Even in the absence of active disease, 20% of patients with CD were observed to be more than 10% below the desirable body weight. The mechanisms contributing to chronic malnutrition in IBD are multifactorial and include decreased oral intake, increased gastrointestinal (GI) losses, malabsorption, increased nutritional requirements, and drug–nutrient interactions. The nutritional problems of patients with IBD can be divided into those that concern macronutrients and those that concern micronutrients.

Anemia is a common finding in IBD, and it may be related to iron, folate, or vitamin B_{12} deficiency. Iron deficiency resulting from blood loss is seen in about 80% of patients with UC and in about 40% of patients with CD, resulting from poor intake, blood loss from the GI tract, and malabsorption. Macrocytic anemia caused by vitamin B_{12} malabsorption results from involvement of the distal ileum in patients with CD or from folate malabsorption and altered metabolism caused by drug–nutrient interactions and decreased intake of fresh vegetables in both CD and UC. Although rare, deficiency of water-soluble vitamins other than vitamin B_{12} and folate has been reported in patients with IBD, resulting in scurvy (vitamin C), Wernicke encephalopathy (thiamin), photophobia with dermatological changes (riboflavin), and pellagra (niacin). Fat-soluble vitamin deficiency results from malabsorption that may occur because of extensive mucosal inflammation or bacterial overgrowth.

Table 4 Nutritional deficiencies in IBD

Anemia – iron, folate, vitamin B_{12}
Water-soluble vitamins – vitamin C, thiamin, riboflavin, niacin
Fat-soluble vitamins – A, D, E, K
Zinc
Magnesium
Calcium

Vitamin K may be deficient in patients with malabsorption of dietary vitamin K and due to a lack of synthesis by gut bacteria in those who have received antibiotics. Zinc deficiency has been reported in about 40% of patients with CD and is particularly prevalent in patients with severe diarrhea or enteric fistulas. Magnesium malabsorption is relatively common in patients with IBD, especially in patients with extensive small-bowel disease, in those with magnesuria induced by corticosteroids, and in others who have binding of magnesium by malabsorbed fat. Metabolic bone disease in patients with IBD is well recognized, with prevalence estimates of bone disease ranging from 5 to 78%. Comparisons between patients with CD and UC reveal that osteoporosis may be more prevalent in CD. Thus, risk factors for development of osteoporosis include IBD, especially CD, and treatment with steroids (Table 4).

Nutritional management Nutritional support has been a focus of considerable interest in the management of patients with IBD. A dietitian should be consulted to prescribe suitable protein and energy intakes and to modify the diet to account for food intolerances or allergies.

A proper diet must correct calorie and micronutrient deficits caused by self-imposed or iatrogenic dietary restrictions. One must meet dietary goals that are adequate for the nutritional needs of the patient. Lactase deficiency is reported to occur in 5–15% of whites of North European descent. In one series of CD patients, 34% failed to absorb physiological doses of lactose. In most patients, elimination of dietary lactose decreases abdominal cramps and diarrhea, but objective proof of lactase deficiency should be obtained. Prolonged adherence to a milk free diet may contribute to negative calcium balance. Most patients with lactose malabsorption will require only diminution of milk or icecream intake.

While food intolerance has been shown in patients with IBD, it is not a universal problem. There is a significant correlation between incidence of CD and dietary changes. However, there is no role for elimination diets in the global management of IBD. Theories suggest that different diets affect CD by allowing the patients to absorb more food and thus not suffer

from malnutrition. Constant feeding will allow greater absorption of food, allowing for greater capability of fighting disease. The role of various diets, however, remains controversial. It is possible that certain fats or proteins exacerbate disease. There has been extensive research investigating the effects of different fat compositions and the course of disease. The main focus has been on the type of fat. Researchers have shown that the consumption of a diet with a high ratio of polyunsaturated to saturated fatty acid improves the intestinal function in CD patients. Specifically, the ω-3 fatty acids have been proven to be effective in CD management. Disease activity falls as the composition of long-chain triglycerides falls in the diet. For patients with UC, butyrate (a fatty acid with four carbon atoms) has been shown to induce remission. ω-3 fatty acids have also been shown to have antiinflammatory properties that induce remission in patients with ulcerative colitis. Whether nutrition can control inflammation in the same way as medication remains a controversial issue. It has been claimed that if a patient takes nothing by mouth, this will induce remission through bowel rest. Resting the bowel in this way may cause malnutrition and further harm.

In patients with narrowing of the lumen, the rationale for restricting dietary fiber is apparent. Fiber intake can be reduced by avoiding coarse wholegrain breads and cereals, nuts, and most fruits and vegetables. Fruit and vegetables may not pass through strictures and may cause a bolus obstruction behind a stricture. Patients with symptoms suggestive of mild or partial bowel obstruction should avoid the intake of raw fruits and vegetables.

Patients with CD involving the small bowel and patients with extensive small-bowel resection often have fat malabsorption. Fatty acids and their derivatives often exert a cathartic effect upon the colonic mucosa. There is also a direct correlation between loss of fat in stools and loss of calcium, magnesium, and possibly zinc. Patients with steatorrhea experience symptomatic benefit from decreasing their intake of dietary fat. A reduction in fat to 70–80 g often suffices to alleviate diarrhea. The calories lost by removal of calorie-dense fat from the diet must be replaced by more easily absorbed calorie sources such as sugars and starches.

The target protein–energy intake should be 126–146 kJ kg^{-1} day^{-1}, with 1.5–1.7 g per kilogram of protein per day (Table 5). Fluid and electrolyte deficiencies are best treated with an oral rehydration solution. Patients should receive a solution consisting of glucose 90 mmol l^{-1}, sodium chloride 45 mmol l^{-1}, sodium citrate 45 mmol l^{-1}, and potassium chloride 20 mmol l^{-1}. Iron deficiency is best treated with iron supplements. An appropriate dose is 300 mg of ferrous sulfate or gluconate one to three times per day. It is sometimes necessary to administer iron by intravenous infusion or intramuscular injection, as there is evidence that iron in the colon increases oxidative stress and may exacerbate inflammation. Magnesium deficiency is treated with oral supplements including magnesium heptogluconate (Magnesium-Rougier) or magnesium pyroglutamate (Mag 2). The total dose of elemental magnesium required to ensure normal serum magnesium varies between 5 and 20 mmol day^{-1}. Zinc deficiency as a result of losses through the stool is common in IBD. Zinc supplements should be considered in patients passing significant amounts of stool. Zinc deficiency can be treated by the administration of zinc gluconate, 20–40 mg day^{-1}. Patients with CD should take folic acid, 1 mg day^{-1}. Folic acid nutrition is important in light of recent data that folate supplementation may provide protection against both sporadic and IBD-associated colorectal cancer. Supplementation with thiamin, riboflavin, pyridoxine, niacin, and ascorbate should be recommended using a standard vitamin preparation. Calcium supplements of 1000–1500 mg of elemental calcium per day is appropriate, while vitamin D, 2000–4000 IU day^{-1} is often required.

Table 5 Target energy and nutrient intake in CD

Energy	126–146 kJ kg^{-1} day^{-1}
Protein	1.5–1.7 g kg^{-1} day^{-1}
Iron	Iron supplement 300 mg day^{-1}
Magnesium	Magnesium supplement 5–20 mmol day^{-1}
Zinc	Zinc gluconate 20–40 mg day^{-1}
Folic acid	1 mg day^{-1}
Calcium	Elemental calcium 1000–1500 mg day^{-1}
Vitamin D	2000–4000 IU day^{-1}

Eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) are polyunsaturated fatty acids derived from fish oil and marine mammals. Fish oil has been observed to have antiinflammatory effects. Fish oil modulates inflammatory mediators that are considered to be operative in CD. Fish oils have been shown to reduce the production of leukotriene B_4 and thromboxane A_2, inhibit the synthesis of cytokines, and scavenge free radicals. Clinical studies have shown a modest antiinflammatory effect of fish oils in psoriasis, atopic dermatitis, and rheumatoid arthritis. One study included 78 patients with CD in remission who were randomly assigned to 4.5 g of a fish oil preparation or to placebo capsules containing the same amount of medium-chain triglycerides. After one year, the number of patients still in remission was significantly higher in the fish oil group (59 vs. 26%). Another study from Spain found that a diet rich in cold fish meat significantly reduced the rate of relapse by 38% at 2 years follow-up. The difficulty in

evaluating the efficacy of trials is in the high drop-out rate of patients due to poor compliance with the fish-oil diet. The advantage in using fish oil is that there is no significant morbidity associated with fish oils. The side-effects, unpleasant taste and odor, are considered insignificant. A summary of trials suggests that fish-oil preparations can be effective for the maintenance of CD in remission. Further studies are necessary to evaluate the effects of fish oils in conjunction with currently available maintenance therapy for CD.

Patients with more severe symptoms unresponsive to medical therapy often require intensive nutritional support. For the patient requiring hospitalization, complete bowel rest, defined as the complete absence of oral intake, is often employed. However, such management causes malnutrition unless alternative forms of energy are given.

Elemental diets contain predigested forms of macronutrients, no lactose or fiber, and vitamins, minerals, and electrolytes sufficient to meet the recommended dietary allowances for healthy people. The diets are administered through flexible tubes placed into the stomach or duodenum. It is not certain how an elemental diet works. It may be that giving a patient a balanced nutritious diet improves overall well-being. Others have suggested that an elemental diet has a direct effect on the mucosa of the bowel. These feeds are effective because they provide minimal digestion, the reduction of fecal bulk, and a reduction in gastric and pancreatic secretions. Therapeutic value may also be related to induced changes in epithelial growth kinetics. Almost 20 randomized controlled trials have evaluated the efficacy of enteral feeding. All trials evaluated the short-term efficacy on the CD activity index. The trials compared either diet therapy versus standard pharmacotherapy or diet therapy where one formula was compared to another. The trials used small number of subjects, had high withdrawal rates, and contained a heterogeneous patient population. Initial controlled trials showed that feeding patients elemental diets was as effective as prednisone in inducing disease remission. Later studies suggested that remission was more likely to occur with the use of prednisone. In summary, nutritional support is likely more effective than placebo in inducing remission. Diets low in long-chain triglycerides may be as effective as prednisone.

The logic of minimizing trauma and irritation of inflamed bowel appears sound. Peripheral intravenous therapy during total bowel rest can maintain adequate fluid and electrolyte balance and ensure maintenance of vitamin and trace elements. The goals of total parenteral nutrition (TPN) therapy include weight gain, positive nitrogen balance, and remission of symptoms. Complications of TPN include pneumothorax during line insertion, catheter-related infections, cholestasis, and steatosis. TPN has been shown to be effective in patients with active disease, patients with fistula, patients with growth retardation, and patients with short bowel. Trials reported regarding TPN are small, with different types and severity of disease. The experience of the TPN team is highly variable. Several studies show that TPN can induce remission in acute CD. TPN with bowel rest achieved remission in 59 and 79% of patients in two randomized controlled trials. It has been shown that TPN may delay the time to surgery in some patients. There was early enthusiasm regarding the use of TPN in treating fistulous disease. TPN can improve the nutritional status and result in decreased fistula output, but better trials with a longer follow-up are required before TPN can be used as a first-line therapy for fistulous disease. TPN can reverse growth arrest and has therefore been increasingly used as the principal determinant of restored growth. Finally, parenteral therapy is important in the management of patients with short bowel. In short periods, it allows for gut adaptation to occur. Long-term use is required in patients whose functioning small bowel remains completely or partially inadequate for nutritional maintenance. In summary, data suggesting that TPN induces disease remission are indirect and inconclusive but should be used to give nutritional support to patients when other means of support are not possible.

Ulcerative Colitis

UC is an inflammatory disease of unknown etiology affecting the colonic mucosa from the rectum to the cecum. It is a chronic disease characterized by rectal bleeding and diarrhea. Histological features of the disease may be seen in other inflammatory states of the colon, such as those caused by bacteria or parasites. The diagnosis of ulcerative colitis, therefore, rests on the discovery of a combination of clinical and pathological criteria, investigation of the extent and distribution of lesions, and exclusion of other forms of inflammatory colitis caused by infectious agents (*Entamoeba histolytica, Clostridium difficile, Campylobacter*, *Escherichia coli*, and *Shigella*). UC is an inflammatory state confined to the mucosa. The colonic tissue displays small microabscesses, called crypt abscesses, which involve the crypts of Lieberkühn. Polymorphonuclear cells accumulate in the crypt abscesses, and frank necrosis of the surrounding crypt epithelium occurs; thus, the polymorphonuclear infiltrates extend into the colonic epithelium. These microabscesses in the crypts are not visible to the naked eye; however, several crypt abscesses may

coalesce to produce a shallow ulceration visible on the mucosal surface. Occasionally, lateral extension of crypt abscesses may undermine the mucosa on three sides, and the resulting hanging fragment of mucosa will appear endoscopically and radiographically as a 'pseudopolyp.' Following this mucosal destruction, highly vascular granulation tissue develops in denuded areas, resulting in friability and bleeding. The two most prominent symptoms of UC – diarrhea and rectal bleeding – are related both to the extensive mucosal damage that renders the colon less capable of absorbing electrolytes and water, and to the highly friable vascular granulation tissue, which bleeds readily.

Etiology It was once thought that UC was caused by food allergy. The food most likely to cause the reaction was cows' milk. Other foods implicated in the etiology of ulcerative colitis include eggs, wheat, tomatoes, oranges, and potatoes. Circulating antibodies to milk proteins are present in some normal subjects but are more common and in higher titers in patients with UC. Sensitivity to disaccharides, including lactose, is noted in patients with UC.

Nutritional management Five studies to date have investigated the effect of fish oil supplementation on chronic active UC. The studies, however, are small and of short duration, have varying doses of ω-3 fatty acids, have significant rates of drop out, and are confounded by the use of other medications. It can be concluded, however, that fish oil appears to have only a modest treatment effect on active UC and does not seem to be beneficial in maintaining UC in remission.

Short-chain fatty acids (SCFA) have been proposed as a topical treatment for active distal UC. SCFA are organic acids produced by anaerobic fermentation of undigested carbohydrates within the colonic lumen. Acetate, propionate, and butyrate, the two-, three-, and four-carbon SCFA, respectively, account for 90–95% of SCFA in the colon, with isobutyrate, valerate, iosovalerate, and caproate comprising the rest. Once inside the colonocyte, SCFA are an important energy source for the cell. Butyrate is the preferred SCFA to meet colonic energy requirements, acetate is second, and propionate is the least metabolized of the three. The SCFA could account for up to 80% of the energy requirements of the colon and for 5–10% of total body energy requirements. Colonic biopsy specimens from patients with UC have impaired utilization of butyrate as measured by carbon dioxide production. It has been suggested that UC may be the result of an energy-deficiency state of the colonic epithelium. The existence of a deficiency of SCFA production has not been consistently shown to exist in UC. It has therefore been postulated that the problem may lie in the uptake or oxidation of the SCFA by the colonocytes. The rationale behind using SCFA in UC is that supraphysiologic luminal SCFA concentrations may be able to overcome the partial metabolic defect of the colonic mucosa to oxidize SCFA. SCFA stimulate colonic cell proliferation, provide a more effective barrier between mucosa and the intraluminal contents, dilate the resistance arteries of the colon, and increase blood flow and mucosal oxygen uptake. Four uncontrolled studies investigating the effect of SCFA on distal active UC have been published. In all four trials, 50–78% of the patients had clinical improvement, some with complete remission, during the study period ranging from 4 to 6 weeks. Disease activity indices and endoscopic appearance all improved. Five randomized trials published to date also support the role of SCFA in the management of UC. These trials to date are limited by the small number of subjects, varying combinations and concentrations of SCFA, and differences in study design and patient populations. However, these trials collectively suggest that SCFA irrigation can effectively treat refractory distal UC that has failed to respond to standard topical, systemic, or combination therapies with a pooled clinical response rate of 70.6%. As a first-line treatment of mild to moderate distal UC, SCFA is associated with a modest clinical response. SCFA is as effective as topical corticosteroid and 5-ASA in inducing clinical, endoscopic, and histologic improvement in mild to moderate distal ulcerative colitis. SCFA is free of significant side-effects and toxicity, and is well tolerated. SCFA is associated with significant cost-savings compared with standard typical corticosteroids or 5-ASA. Unresolved issues regarding SCFA include their role in maintaining remission of distal UC once active disease is brought under control. Furthermore, the role of SCFA in preventing colon cancer has yet to be delineated. Finally, the most effective concentrations and compositions of SCFA have not yet been determined.

The impetus for using SCFA in the treatment of distal ulcerative colitis was the finding that topical SCFA could effectively treat colonic inflammation in diversion colitis. Diversion colitis is an inflammatory process that occurs in segments of the colon after surgical diversion of the fecal stream. Several studies have documented that endoscopic abnormalities and histologic changes occur in the distal colonic segment of most patients after intestinal diversion. Pathologic examination shows lymphoglandular complexes expanding the submucosa with increased lymphocytes and plasma cells. Cryptitis with abscesses, patchy neutrophil infiltration in the lamina propria, and superficial erosions overlying lymphoid follicles

are common findings. The disease is characterized by bleeding from inflamed colonic mucosa, tenesmus, mucus discharge, and abdominal pain. The majority of patients remain asymptomatic. Diversion colitis is likely caused by a deficiency of SCFA, which serve as luminal nutrients for colonocytes. The differential diagnosis for diversion colitis includes acute self-limited colitis, antibiotic-associated colitis, and preexisting UC or CD. The definitive treatment of diversion colitis is restoration of intestinal continuity. Surveillance for cancer in the diverted segment is not necessary, provided that the underlying condition for which the operation was performed has no malignant potential. Randomized controlled studies have not confirmed a beneficial role of SCFA in the treatment of diversion colitis, a peculiar finding. These studies were hampered by the small sample size and short duration of treatment. In those with preexisting IBD, SCFA combined with antiinflammatory drugs may be effective.

A low-residue diet is recommended during acute phases of colitis, the idea being that insoluble particles irritate the bowel and make diarrhea worse. Recently, however, the rarity of UC in developing countries, together with the ability of dietary fibre to affect colonic function and its bacterial content, has suggested that a low intake of fiber might be a factor in causing UC. A high-fiber diet can be prescribed for patients during quiescent phases of disease.

Initial studies suggested a reduced postoperative morbidity and mortality in patients with UC receiving TPN. However, parenteral nutrition is not an effective primary therapy for UC. Retrospective studies have failed to provide evidence for the use of TPN in inducing remission of UC. Moreover, five prospective studies investigating the use of TPN in inducing remission in UC failed to show any benefit for TPN versus control groups. In all five studies, there was a mean 37% initial response rate and only a 12% sustained response rate. TPN appears to have little influence on refractory patients with active disease, and rarely can TPN avert colectomy. TPN can be useful to prepare the malnourished patient for surgery as well as providing postoperative nutritional support. There have been relatively few studies looking at the use of elemental diets in UC. It is generally considered that enteral nutrition is not effective in achieving remission in active ulcerative colitis.

The role of diet and nutrition in patients with IBD remains controversial. By applying an orderly and comprehensive approach to management, many of the complications of IBD can be contained or prevented. While the role of diet continues to evolve, its efficacy in the management of IBD is undisputed.

Diet and Nutrition in Irritable Bowel Syndrome

Irritable bowel syndrome (IBS) is defined as a functional bowel disorder in which abdominal pain is associated with defecation or a change in bowel habit, with features of disordered defecation and distention. Criteria for IBS have been formalized in the Rome Criteria (Table 6). IBS is a disorder that can be diagnosed positively on the basis of a series of symptom criteria and limited evaluation to exclude organic disease. The incidence of IBS has been estimated to be 1% per year. Prevalence data estimate a range of 2.9–20%. Patients have been divided into subgroups based on the predominant symptom. Symptom subgroups include constipation-predominant IBS, diarrhea-predominant IBS, and IBS with alternating bowel movements. The prevalence of IBS is lower in the elderly and higher in female patients. Only 10–25% of patients with IBS seek medical care. The economic impact in the USA is estimated at $25 billion annually. IBS accounts for 2.4–3.5 million physician visits in the USA annually, making it the most common diagnosis in gastroenterologists' practice (12% of primary care visits, 28% of all gastroenterologist's patients).

Pathophysiology

There is no single physiological mechanism responsible for symptoms of IBS. IBS is considered a biopsychosocial disorder resulting from a combination of psychosocial factors, altered motility and transit, and increased sensitivity of the intestine or colon (Table 7). It has been hypothesized that altered peripheral functioning of visceral afferents and the central processing of afferent information are important in the altered somatovisceral sensation and motor dysfunction in patients with IBS. It has been postulated that persistent neuroimmune interactions after

Table 6 Criteria for the diagnosis of IBS

Manning criteria	Pain relieved by defecation
	More frequent stools at the onset of pain
	Looser stools at the onset of pain
	Visible abdominal distention
	Passage of mucus
	Sensation of incomplete evacuation
Rome II criteria	At least 12 weeks or more, which need not be consecutive in the previous 12 months of abdominal pain or discomfort with two of three features: relief with defecation, onset associated with a change in the frequency of stool, onset associated with a change in the appearance of stool

Table 7 Pathophysiology of IBS

Biopsychosocial disorder
Abnormal motility
Heightened visceral perception
Psychologic distress
Intraluminal factors – lactose, bile acids, short-chain fatty acids, food allergen
Postinfectious alteration of gut function

infectious gastroenteritis resulting in continuing sensorimotor dysfunction might be a cause of IBS. Infectious diarrhea precedes the onset of IBS symptoms in 7–30% of patients.

In some patients, carbohydrate intolerance may contribute to the symptoms of IBS. Data suggest that food allergens may play a role in IBS (Figure 1). Symptoms often improve with dietary exclusion. Stress and emotions affect gastrointestinal function and cause symptoms in patients with IBS. Psychologic symptoms that are more common in patients with IBS include somatization, anxiety, hostility, phobia, and paranoia.

Diagnosis

Many patients' symptoms fluctuate over time. Regardless, IBS is a 'safe' diagnosis. Patients with IBS have a benign course without any risk of developing organic disease. The diagnosis of IBS first involves a careful assessment of the patient's symptoms. Manning or Rome criteria can be used to identify patients with IBS. A thorough physical examination and a limited series of initial investigation are needed to exclude organic structural, metabolic, or infectious diseases. Investigations include hematology and chemistry tests, erythrocyte sedimentation rate, stool examination for occult blood, ova, and parasites, and possible flexible sigmoidoscopy.

Food Intolerance

Data from dietary elimination and food challenge studies support the role of diet in the pathogenesis of a subgroup of IBS patients. The term 'food intolerance' encompasses nonimmunologically mediated adverse reactions to food, which resolve following dietary elimination and are reproduced by food challenge. Food allergy or hypersensitivity is used to describe conditions in which an immunological mechanism may be demonstrable. Food aversion is a psychological avoidance response. A prevalence for food intolerance of 5% has been reported in the general population. Food hypersensitivity is a common perception amongst IBS patients with 20–65% attributing their symptoms to adverse food reactions. The effect of dietary exclusion followed by food challenge

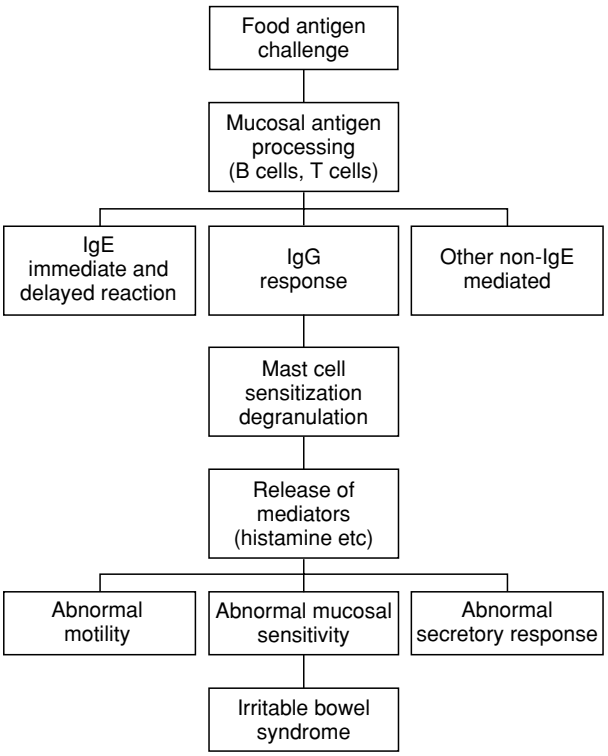

Figure 1 Pathogenesis of food hypersensitivity in IBS.

has been investigated in several trials. Food intolerances have been recognized in 6–58% of cases, with milk, wheat, and eggs being the most commonly implicated foods. In the largest study, the effect of an exclusion diet was evaluated in 200 IBS patients. Symptomatic improvement was reported in 91/189 (48%) patients and was maintained for a mean of 14.7 months. Following dietary challenge, 73/91 (80.2%) of responders identified one or more food intolerances. The investigators conclude that a portion of IBS patients may benefit from dietary manipulation.

The bowel mucosa acts as a physical barrier to a variety of intraluminal dietary and microbial antigens. Under normal physiological conditions intact food antigens can penetrate the mucosal barrier via transcellular or paracellular routes. The mechanism by which the mucosal immune system maintains a state of immunological tolerance is not clear. Studies demonstrating a positive response to elimination diets support the role of a food hypersensitivity reaction in IBS. An immune-mediated food hypersensitivity response involving IgE and IgG antibodies has been postulated as the underlying mechanism in patients with IBS.

While there is no clear evidence to suggest that IBS is an infective condition, up to 30% of patients develop symptoms after an episode of gastroenteritis.

Infective and inflammatory conditions of the bowel cause an increased mucosal permeability, thereby exposing the immune system to an increased load of dietary and microbial antigens. The production of a variety of proinflammatory and immunomodulatory cytokines may serve to prime the mucosal and submucosal immune system establishing a hypersensitivity response. Altered bowel flora has been reported in patients with IBS. Restoration of bowel flora whether by dietary manipulation or through the use of probiotics has been suggested as a treatment for IBS.

Treatment

The dietary treatment of patients with IBS has centered around bran supplementation, manipulating the dietary fiber content of the diet and the identification of food intolerance. There are currently no evidence-based guidelines on how dietitians should treat patients with IBS. An in-depth assessment of a patient's dietary intake is essential prior to any therapeutic dietetic intervention. A history regarding the onset of symptoms in relation to any dietary changes is often useful. Dietary intervention needs to be individualized with respect to the patient's symptoms. Of the eight studies evaluated, five identified food intolerance as a major contributor to symptoms. There are no randomized control trials in this area primarily due to the impracticality of constructing a trial design. In all trials, the methods used to assess the effect of intolerance to suspect foods on symptoms varied. In summary, seven studies mentioned dairy products, six coffee, and five wheat. Others cite eggs, corn, potatoes, onions, fruits, and vegetables. In the majority of the studies, the patients with diarrhea responded favorably to an exclusion diet compared to other subtypes of IBS. The recommendation that elimination diets control the symptoms of IBS is based on poorly designed and incomplete trials. The absence of any objective symptom assessment in the trials raises questions about their validity.

The mainstay of dietary therapy for IBS has centered around the manipulation of dietary fiber. Rees *et al*. conducted a critical review of clinical trials examining the effect of dietary fiber on symptoms in patients with IBS. In the review, it was shown that six out of eight investigations detected no significant difference in most of the symptoms of IBS between fiber and placebo. In the eight trials, the amount of supplement varied, the form in which the fiber was administered differed, and the type of fiber supplement was different. The literature does not support any beneficial effects of increasing insoluble nonstarch polysaccharides in IBS patients. Increasing insoluble wheat products in the diet may be of some benefit in patients with predominantly symptoms of constipation. Many patients complain of bloating with higher doses of fiber. Bran is reported to be no better than placebo in relief of overall IBS symptoms and may be worse than a normal diet for symptoms of IBS caused by intraluminal distention. Fiber may induce bloating by increasing residue loading and bacterial fermentation without accelerating the onward movement of the increased residue. Nonetheless, there appears to be significant improvement in constipation if sufficient quantities of fiber (20–30 g day^{-1}) are consumed. As a result, it is common practice to start with a low dose, increasing gradually, and abandoning high levels of supplementation (> 30 g day^{-1}) if patients experience worsening of symptoms. Therefore, it can be concluded that fiber may have a role in treating constipation with a minimal role in the relief of abdominal pain and diarrhea.

The role of carbohydrate and sorbitol is not routinely considered in the management of IBS. Trials cannot confirm a true malabsorption of carbohydrates in patients with IBS. Nonetheless, studies have shown that patients with IBS develop symptoms after being exposed to a sorbitol diet, when compared to controls. One-third of those with IBS reported lactose intolerance on a subjective basis. The percentage of lactose maldigesters in patients with IBS is the same as in healthy subjects, but the number of subjects reporting lactose intolerance is higher (60% compared to 27%). There is a strong relationship between subjective lactose intolerance and IBS. The inconclusive findings of, and association of, lactose intolerance and IBS can be explained by poorly designed studies. Lactose intolerance has been associated with GI symptoms, and whether this is due to lactase deficiency or increased sensitivity cannot be answered.

In summary, the goal of dietary manipulation in patients with IBS is to help patients control their symptoms. Dietary assessment is essential for all patients with determination of their present dietary intake. Any unusual or abnormal eating practices need to be assessed in relation to the patient's symptoms. It is important to obtain a detailed history of the onset of symptoms and their relation to the individual's eating pattern. Exclusion diets should be tried only when patients complain of multiple food intolerance, and single food avoidance has not helped control symptoms. Patients taking large quantities of sorbitol should be discouraged of such practice, especially if their symptoms are predominantly pain and diarrhea. A milk-free or lactose-free diet should be tried in those patients in whom dairy products are associated with symptoms. Patients continuing on this diet should take calcium supplements. A trial of

a wheat-free diet may be helpful. If a patient's predominant symptom is constipation, an assessment of fluid intake should be undertaken. Regular meal patterns should be encouraged in all patients. Finally, an assessment of the type and quantity of nonstarch polysaccharides consumed should be made. The addition of bran and insoluble fibers should be discouraged, unless the individual feels this is of direct benefit in symptom control. More emphasis should be placed on increasing the proportion of foods containing a higher concentration of soluble nonstarch polysaccharides.

Diet and Nutrition in Diverticular Disease

Diverticular disease is a term encompassing diverticulosis and diverticulitis. Diverticulosis occurs in at least one person in two over the age of 50 years. The prevalence of diverticular disease is age-dependent, increasing from less than 5% at age 40, to 30% by age 60, to 65% by age 85. A male preponderance was noted in early series, but more recent studies have suggested either an equal distribution or a female preponderance. There are geographic variations in both the prevalence and pattern of diverticulosis. Westernized nations have prevalence rates of 5–45%, depending upon the method of diagnosis and age of the population. Diverticular disease in these countries is predominantly left-sided. The findings are markedly different in Africa and Asia, where the prevalence is less than 0.2%, and diverticulosis is usually right-sided. Diverticulosis or diverticular disease of the colon is due to pseudodiverticula in that the wall of the diverticulum is not a full-thickness colonic wall, but rather outpouchings of colonic mucosa through points of weakness in the colonic wall where the blood vessels penetrate the muscularis propria. These diverticula are prone to infection or 'diverticulitis' presumably because they trap feces with bacteria. Among all patients with diverticulosis, 70% remain asymptomatic, 15–25% develop diverticulitis, and 5–15% develop some form of diverticular bleeding. Diverticulitis represents micro- or macroscopic perforation of a diverticulum. The primary process is thought to be erosion of the diverticular wall by increased intraluminal pressure or inspissated food particles; inflammation and focal necrosis ensue, resulting in perforation. If the infection spreads beyond the confines of the diverticula in the colonic wall, an abscess is formed. Patients present with increasing left lower quadrant pain and fever, often with constipation and lower abdominal obstructive symptoms such as bloating and distention. Some patients with severe obstructive symptoms may actually describe nausea or vomiting. This can occur with or without abscess formation. The diagnosis of acute diverticulitis is often made on the basis of the history and the physical examination. On physical examination, the patient often has localized tenderness in the left lower quadrant and, with severe infection and an abscess, may have rebound tenderness in the left lower quadrant. A palpable mass is often identifiable where the sigmoid colon (the most common site of diverticulitis) is infected. Computed tomographïc (CT) scanning has become the optimal method of investigation in patients suspected of having acute diverticulitis, being employed for diagnosis, assessment of severity, therapeutic intervention, and quantification of resolution of the disease. CT scan may be helpful in outlining the colon and identifying an abscess, and is preferable to barium enema for diagnosis in patients with acute illness. After resolution of an episode of acute diverticulitis, the colon requires full evaluation by colonoscopy, barium enema, or both to establish the extent of disease and to rule out coexistent lesions, such as polyps or carcinoma.

Etiology

It has been speculated that low dietary fiber predisposes to the development of diverticular disease. In one study, Burkitt and Painter demonstrated that individuals in the UK eating a Western diet low in fiber had colonic transit times of 80 h and a mean stool weight of 110 g day^{-1}. In comparison, Ugandans eating very high fiber diets had transit times of 34 h and greater stool weights (> 450 g day^{-1}). The longer transit times and smaller stool volumes were felt to contribute to the development of diverticular disease through the increase in intraluminal pressures that predispose to diverticular herniation. The etiology of diverticular disease is unknown. The leading theory suggests that altered colonic motility plays a major role in the development of diverticula. Higher resting, postprandial and neostigmine stimulated pressures in diverticular patients suggest that a delay in transport with augmentation of water reabsorption could cause excessively high pressures forcing mucosa to herniate.

A recent report that evaluated a cohort of over 47 000 men provided strong evidence for the role of dietary fiber. After adjustment for age, energy-adjusted total fat intake, and physical activity, total dietary fiber intake was noted to be inversely associated with the risk of symptomatic diverticular disease. The relative risk was 0.58 for men in the highest quintile compared to those in the lowest quintile for fiber intake. The observation that diverticular disease is less common in vegetarians than nonvegetarians is also compatible with a role for dietary fiber, since vegetables and fruits are important sources of fiber.

Other dietary factors that might contribute to the pathogenesis of diverticular disease have been examined. There is no substantially increased risk associated with smoking, caffeine, or alcohol. However, an association has been noted between obesity in men under 40 years and acute diverticulitis. This finding is compatible with observations that the risk of symptomatic diverticular disease is particularly increased (relative risk 2.35–3.32, 95% confidence interval) by a diet characterized by a high intake of total fat or red meat and a low intake of dietary fiber.

Treatment

The treatment of diverticular disease can be divided into prevention of diverticulosis, uncomplicated diverticulosis, and complicated diverticulosis (Table 8).

The mainstay of treatment for preventing diverticulosis is a diet high in fruit and vegetable fiber. This suggestion is based on observations that low-fiber diets are associated with colonic diverticulosis. Recommendations are also based on results from the Health Professionals Follow-up Study, which showed an inverse association between insoluble dietary fiber intake and the risk of developing symptomatic diverticular disease. In particular, the greatest reduced risk was in those consuming a diet high in fruit and vegetable fiber, with an average consumption of 32 g day^{-1}.

The management of uncomplicated diverticulosis involves a diet high in fruit and vegetable fiber. The majority of patients with diverticular disease will remain asymptomatic. Therefore, this group of patients consists of the largest treatment group. There are no data to support any therapeutic recommendations in this group of patients. There have been multiple uncontrolled studies demonstrating the effect of fiber in patients with diverticulosis, but the lack of a placebo group complicates the results. Other well-conducted studies have reported conflicting results regarding the use of fiber. Despite conflicting data, it appears safe to recommend a diet high in fiber for patients with uncomplicated diverticular disease.

Complicated diverticulosis includes diverticulitis and hemorrhage (Table 9). Patients with mild diverticulitis can be treated as outpatients with broad-spectrum oral antibiotics with activity against anaerobes and Gram-negative rods. Patients should be on the alert for symptoms of increasing pain, fever, or inability to tolerate oral foods. Such patients are treated with a clear liquid diet. Patients admitted to hospital with more severe diverticulitis should be placed on bowel rest with clear liquids or nothing by mouth. Intravenous fluid therapy is required. In addition, intravenous antibiotics should be initiated to target colonic anaerobes and Gram-negative rods. Improvement is expected within 2–4 days, at which point the diet can be advanced. In patients hospitalized with acute diverticulitis, 15–30% will require surgery. After resolution of acute diverticulitis, the likelihood of recurrence and the role of elective surgical resection are important to prevent further attacks. The risk of recurrence after an acute attack ranges from 7 to 62%. Recurrent attacks are less likely to respond to medical and diet therapy and often require emergent surgery. Therefore, elective surgery is often recommended after two attacks of acute diverticulitis.

Table 8 Management of patients with diverticulosis

Prevention	High-fiber diet
Uncomplicated	High-fiber diet
	No antispasmodics
	No antibiotics
Complicated	Broad-spectrum antibiotcs
	Bowel rest
	Surgical consult

Table 9 Complications of diverticulosis

Fistula
Obstruction
Hemorrhage
Perforation
Peritonitis
Abscess

Diet and Nutrition in Infectious Colitis

Infectious or noninfectious causes may be responsible for acute diarrhea, and in selected patients, both can occur simultaneously. Noninfectious causes of diarrhea include drugs, food allergies, primary gastrointestinal diseases such as inflammatory bowel disease, and other disease states such as thyrotoxicosis and the carcinoid syndrome (Table 10). A variety of infectious diseases cause acute diarrhea. Diarrheal diseases represent the second most common cause of death worldwide and the leading cause of childhood death. Diarrhea results in 300–400 deaths per year in children in the USA, approximately 200 000 hospitalizations, 1.5 million outpatient visits, and more than one billion dollars in direct medical costs. Acute diarrheal disease is generally defined as three loose stools (or two loose stools with abdominal symptoms) or more than 250 g of stool per day for up to 7 days. When diarrhea lasts for 14 days, it can be considered persistent; the term chronic generally refers to diarrhea that lasts for at least a month. It is useful to categorize infectious diarrheal diseases by the portion

of the intestine that they are prone to infect since the presenting symptoms vary by region. The initial evaluation of patients with acute diarrhea should include a history of the duration of symptoms and the frequency and characteristics of the stool. The clinical and epidemiologic history is central to patient medical evaluation and management. Diarrhea may be categorized as mild (no change in normal activities), moderate (forced change in activities), or severe (disability generally with confinement to bed). There should be an attempt to elicit the severity of illness, evidence of dehydration and extracellular volume contraction (e.g., with orthostatic vital signs), and character of stool pattern. Fever and peritoneal signs may be clues to infection with an invasive enteric pathogen (*Shigella*, *Salmonella*, or *Campylobacter*). Diagnostic clues may be provided by questioning about factors that might expose a patient to potential pathogens such as residence, occupational exposure, recent and remote travel, pets, and hobbies. A food history may also provide clues to a diagnosis. Consumption of unpasteurized dairy products, raw or undercooked meat or fish, or organic vitamin preparations may suggest certain pathogens. In addition, the timing of symptoms with regard to exposure to suspected offending food can be important clues to the diagnosis. Symptoms that begin within 6 h suggest ingestion of a preformed toxin of *Staphylococcus aureus* or *Bacillus cereus* (Table 11). Symptoms that begin at 8–14 h suggest infection with *Clostridium perfringens*. Symptoms that begin at more than 14 h suggest infection with viral agents, particularly if vomiting is the most prominent feature, or bacterial contamination of food with enterotoxigenic or enterohemorrhagic *E. coli*. It is also important to ask about recent antibiotic use (as a clue to the presence of *Clostridium difficile* infection). In the absence of specific clues in the history, a stool culture should be obtained in patients with the signs of more severe illness. Gross examination of the stool for blood, pus, and mucus is the single most important laboratory test to begin the evaluation of an acute diarrheal illness. This should be followed by microscopic examination of the stool for inflammatory cells. A stool specimen should be submitted for culture if the stool is inflammatory (i.e., with white cells and/or mucus). Specific media, methods, or stains may be required to isolate or identify organisms of interest. A routine stool culture will identify *Salmonella*, *Campylobacter*, *Shigella*, *Aeromonas*, and most strains of *Yersinia*. A stool culture that is positive for one of these pathogens in a patient with acute diarrheal symptoms can be interpreted as a true positive, although antibiotic therapy is not required for all of these organisms. Other organisms that should be considered in selected situations include enterohemorrhagic *Escherichia coli*, viruses, vibrios, *Giardia*, *Cryptosporidia*, and *Cyclospora*.

Table 10 Noninfectious causes of diarrhea

Drugs
Thyroid disease
Carcinoid syndrome
Lactose deficiency
Short gut
Gastrinoma
Tumor
Cystic fibrosis
Schwanoma
IBD
Celiac disease
IBS

Treatment

The management of patients with acute diarrhea begins with general measures such as hydration and alteration of diet. Antibiotic therapy is not required in most cases, since the illness is usually self-limited. Nevertheless, empiric and specific antibiotic therapy can be considered in certain situations.

For most cases of acute diarrhea, the most important form of therapy consists of fluid combined with

Table 11 Infectious causes of diarrhea

Agent	*Clinical features*	*Syndrome*
Staphylococcus aureus, Bacillus cereus (preformed toxin)	Nausea, vomiting, watery diarrhea	Gastroenteritis
Enteric pathogen	Abdominal pain, cramps, voluminous stools	Acute watery diarrhea
Shigella, Campylobacter, Salmonella, Escherichia coli, Yersinia, Aeromonas spp., noncholera vibrios, *Chlamydia trachomatis*	Small volume stools, urgency, tenesmus, dysentery	Colitis, proctitis
Parasite (*Giardia, Cryptosporidium, Cyclospora, Microsporidium*), bacterial overgrowth	Depends upon location of disease	Persistent, chronic diarrhea

electrolytes. Particular attention should be given to the immunocompromised and the elderly. Solutions containing sodium in the range of 45–75 mEq l^{-1} are required. For dehydrated patients, aggressive fluid therapy is required. Oral fluids with Na 60–90 mEq l^{-1}, K 20 mEq l^{-1}, Cl 80 mEq l^{-1}, citrate 30 mEq l^{-1}, and glucose 20 g l^{-1} are often recommended. In non-dehydrated healthy persons with acute diarrhea, fruit juices with saltine crackers supplemented with broths and soups can meet the fluid and salt needs in most cases. In all cases, calories should be provided to facilitate enterocyte renewal. Historically, boiled starches/cereals with salt combined with crackers, bananas, yogurt, soup and boiled vegetables have been used successfully. Diet can return to normal when stools are formed. Starchy foods have the advantage of containing a small percentage of simple proteins that are easily hydrolyzed and well absorbed. Once rehydration is complete and food has been reintroduced, the oral electrolyte solution is continued to replace ongoing losses from stool and for maintenance. The rehydration fluid can be safely continued if other foods or fluids are introduced.

Oral rehydration solutions were first developed in the USA in the early 1950s following the realization that, in many small-bowel diarrheal illnesses, intestinal glucose absorption via sodium–glucose cotransport remains intact. Thus, in diarrheal disease caused by any organism that turns on small-bowel secretory processes, the intestine remains able to absorb water if glucose and salt are also present to assist in the transport of water from the intestinal lumen (Figure 2). From this point of view, it would seem desirable not to restrict food in acute diarrheal illnesses. Carefully controlled trials of various regimens and solutions for oral rehydration have been conducted in nearly every country in the world. The American Academy of Pediatrics Committee on Nutrition endorses the use of oral fluid therapy and posttreatment feeding.

A composition of oral rehydration solution (per liter of water) has been recommended by the World Health Organization (Table 12).

Homemade solutions of sugar and salt can also be used for oral rehydration. An appropriate mixture is half a teaspoon of salt (3.5 g) and 8 teaspoons of sugar (40 g) in 1 liter of water. These homemade solutions lack sufficient amounts of potassium. Moreover, potentially harmful errors in mixing have been reported. The electrolyte concentrations of fluids used for sweat replacement (e.g., Gatorade) are not equivalent to oral rehydration solutions, although they may be sufficient for the otherwise healthy patient with diarrhea who is not dehydrated. Diluted fruit juices and flavored soft drinks along with saltine crackers and broths or soups may also meet the fluid and salt needs in these less severely ill individuals. If available, racecadotril, an enkephalinase inhibitor, may be an effective adjunct to oral rehydration solutions. In one study, it reduced the output and duration of watery diarrhea in a study of 135 Peruvian boys, aged 3–35 months.

Several problems with glucose-based oral rehydration solutions have limited their use. While this therapy effectively replaces the fluids lost in the stool, it does not decrease the stool volume. The continued diarrhea often leads to self-imposed cessation of therapy by the user. Glucose-based solutions prepared with an electrolyte concentration higher than that of the recommended solutions as a result of the addition of too much solute can produce an increase in diarrhea and hypernatremia.

In summary, oral rehydration has become the mainstay for management of acute diarrheal illness caused by infection. Glucose electrolyte solutions should be used for rehydration (75–90 mmol of sodium per liter) and maintenance (40–60 mmol of sodium per liter). The ratio of carbohydrate to sodium should not exceed 2:1. Oral rehydration solutions may be used to treat mild, moderate, and severe dehydration. Mixing of dry ingredients and water at home is acceptable if guidelines are strictly followed. Feeding

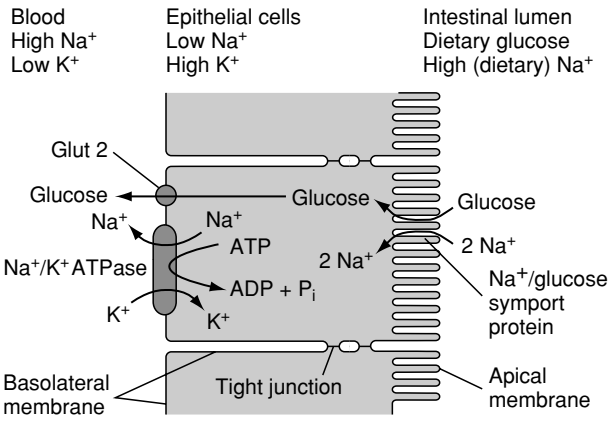

Figure 2 Sodium–glucose transporter.

Table 12 Composition of oral rehydration solution

Ideal	3.5 g of sodium chloride
	2.9 g of trisodium citrate or 2.5 g of sodium bicarbonate
	1.5 g of potassium chloride
	20 g of glucose or 40 g of sucrose
Homemade	1.5 teaspoons (3.5 g) of salt
	8 teaspoons (40 g) of sugar
	1 l of water

should be reintroduced within the first 24 h of the onset of diarrhea.

See also: **Antioxidants**: Natural Antioxidants; **Body Composition**; **Calcium**: Properties and Determination; **Carbohydrates**: Classification and Properties; **Colon**: Structure and Function; **Dietary Fiber**: Properties and Sources; **Enteral Nutrition**; **Fish Oils**: Production; **Folic Acid**: Properties and Determination; **Infection, Fever, and Nutrition**; **Inflammatory Bowel Disease**; **Iron**: Properties and Determination; **Malnutrition**: The Problem of Malnutrition; **Parenteral Nutrition**

Further Reading

Camilleri M (2001) Management of the irritable bowel syndrome. *Gastroenterology* 120: 652–668.

Drossman DA (2000) *The Functional Gastrointestinal Disorders*. McLean, UA, USA: Allen Press.

DuPont HL (1997) Guidelines on acute infectious diarrhea in adults. *The Practice Parameters Committee of the American College of Gastroenterology* 92(11): 1962.

Feldman M, Sleisenger MH and Fordtran JS (eds) (1998) *Gastrointestinal and Liver Diseases*, 6th edn. Philadelphia, PA: W.B. Saunders.

Greenberg GR (1992) Nutritional support in inflammatory bowel disease: Current status and future directions. *Scandinavian Journal of Gastroenterology* 27(supplement 192): 117–122.

Guerrant RL, Gilder TV and Steiner T (2001) IDSA Practice guidelines for the management of infectious diarrhea. *Clinical Infectious Diseases* 32: 331.

Hawkins C (1990) Ulcerative colitis: dietary therapy. In: *Inflammatory Bowel Diseases*. Edinburgh, UK: Churchill-Livingstone.

Jeejeebhoy KN (2002) Clinical nutrition. Management of nutritional problems of patients with Crohn's disease. *Canadian Medical Association Journal* 166(7): 913–918.

Kohler L, Sauerland S and Neugebauer E (1999) Diagnosis and treatment of diverticular disease: Results of a consensus development conference. The Scientific Committee of the European Association for Endoscopic Surgery. *Surgical Endoscopy* 13: 430.

Parks TG (1975) Natural history of diverticular disease of the colon. *Clinical Gastroenterology* 4: 53.

Practice Parameters for the Treatment of Sigmoid Diverticulitis. The Standards Task Force. The American Society of Colon and Rectal Surgeons. *Diseases of the Colon and Rectum* 43: 289.

Ramaswamy K (2001) Infectious diarrhea in children. *Gastroenterology Clinics of North America* 30(3): 611–624.

Rosenberg IH (1990) Diet and nutritional therapy in Crohn's disease. In: *Inflammatory Bowel Diseases*. Edinburgh, UK: Churchill-Livingstone.

Stollman NH and Reskin JB (1999) Diagnosis and management of diverticular disease of the colon in adults. Ad Hoc Practice Parameters Committee of the American College of Gastroenterology. *American Journal of Gastroenterology* 94: 3110.

Cancer of the Colon

S-W Choi and J B Mason, Tufts University, Boston, MA, USA

Background

Colorectal cancer is the second most common cancer in terms of incidence and mortality for both men and women in most of the developed countries of the world. Molecular-genetic studies indicate that colon carcinogenesis is due to a time-dependent accumulation of aberrations in tumor suppressor genes and oncogenes in the colonic epithelial cell, leading to abnormal expression of critical, cancer-related proteins. These aberrations, the most well characterized of which are mutations, offer a molecular explanation for the generally accepted sequence of colon carcinogenesis: normal mucosa, hyperproliferative epithelium, adenoma, and carcinoma.

The majority of colorectal cancers (65–85%) are described as 'sporadic' and have no apparent underlying genetic predisposition, whereas it is estimated that at least 5% of colorectal cancers are due to inherited mutations in major genes ('hereditary' colon cancer), and up to an additional 30% may be due to minor susceptibility genes ('familial' colorectal cancer). Hereditary nonpolyposis colorectal cancer and familial adenomatous polyposis are the two major varieties of hereditary colorectal cancer syndromes.

Migrant studies and other epidemiologic studies indicate that the incidence of sporadic colon cancer is strongly associated with environmental determinants, in particular diet. A diet rich in foods from plant sources (vegetables, fruits, and whole grains) is associated with decreased risk, whereas diets high in animal fat and red meats are associated with increased risk for colorectal cancer. Making appropriate changes in dietary habits and/or supplementing the diet with particular nutrients that are thought to be protective is therefore an attractive approach to preventing this cancer, although it should be emphasized that the prevention of colorectal cancer by this means probably requires long-term changes

over a decade or more in order to effect a meaningful response.

Epidemiology

Rates of colorectal cancer vary considerably with geography. The disease is common in the USA, Australia, New Zealand, Scandinavia, and Western Europe, and is relatively uncommon in Asia, Africa, and South America. Incidence rates vary approximately 20-fold around the world, with the highest rates seen in the developed world and the lowest in India. The international differences, migration data, and recent rapid changes in incidence rates in Italy, Japan, urban China, and male Polynesians in Hawaii show that colon cancer is highly sensitive to changes in the environment. Among immigrants and their descendants, incidence rates rapidly reach those of the host country, sometimes within the migrating generation. The 20-fold international difference may be explained, in large part, by dietary and other environmental differences; indeed, although incidence rates in Japan have been low even until quite recently, the highest rates in the world are now seen among Hawaiian Japanese.

However, colorectal cancer has long been known to occur more frequently in certain families, and there are several rare genetic syndromes that convey a markedly elevated risk. For example, familial adenomatous polyposis, an autosomal dominant hereditary disease, has 90–100% penetrance, with essentially all affected individuals developing colon adenomas, and later colon cancer. Colorectal cancer arising in this setting is thought to have nearly a 100% genetic basis. Colorectal cancer is thus causally related to both genes and environment.

Fat

In Western diets, much of the fat is derived from animal products and may constitute 40–45% of total caloric intake. The relationship between dietary fats and cancer has long been recognized and has received considerable attention as a possible risk factor in the causation of colon cancer. A recent ecological study observed that mortality data of colorectal cancer for 22 European countries, the USA, and Canada were correlated with consumption of animal fat. Increased concentrations of bile acids and fatty acids in the intestine, resulting from a high-fat diet, is suggested as an explanation for the epidemiological association between high intake and high rates of colorectal cancer. Bile acids and fatty acids are considered to be major irritants of the colonic mucosa and are thought to link the high-risk Western-style diet to colon cancer. However, the link between animal fat and colorectal cancer is not entirely consistent. The majority of case-control studies of diet and cancer have shown no significant associations. Furthermore, prospective and intervention studies of fat in studies of the precursor adenoma have yielded conflicting results. The risk of developing an adenoma was reduced by a low-fat diet in a prospective study of 45 000 US health professionals, but not in an intervention study of 400 adenoma patients in Australia.

Not all dietary fats are detrimental in this regard. Intake of vegetable fat is usually shown not to be associated with risk of the cancer. Also, European epidemiologic studies indicate an inverse relationship between fish consumption and colorectal cancer and an inverse correlation with fish and fish oil consumption when expressed as a proportion of total or animal fat. These effects were seen in populations with a high fat intake, indicating that the type of fat in the diet is critical in determining cancer risk. Consumption of fish products rich in ω-3 polyunsaturated fatty acids (PUFA), such as eicosapentaenoic and docosahexaenoic acid, is associated with a low incidence of colorectal cancer and adenoma. Moreover, PUFAs have been shown to normalize altered proliferative patterns of the colonic mucosa in human subjects at high risk for colon cancer. These results emphasize that the enhancing effect of colon carcinogenesis by dietary fat depends on the type of fat as well as the fatty acid composition of the fat.

Meat

Most case-control studies show an increased risk for development of colorectal cancer in those individuals consuming high amounts of meat, especially red meat (this includes lamb, pork, and beef). A high intake of processed meat is also associated with colorectal cancer risk in two cohort studies. Possible mechanisms underlying this epidemiological association include the heterocyclic amines (which are carcinogenic and mutagenic) that are formed when meat is cooked. The possibility that meat alters nitrogen metabolism within the lumen of the intestine and enhances the production of endogenous promoters and carcinogens within the colon is also an attractive mechanism. In a review of several large prospective cohort studies, colon cancer risk was associated with red-meat consumption but not with total or animal fat, which suggests that the risk associated with red-meat consumption is independent of its fat content.

In contrast, consumption of fish or chicken is not associated with risk and might even reduce the occurrence of colorectal cancer. A cohort study in Iowa

women and European prospective studies did not found any significant association between overall meat intake and colon cancer, and two prospective studies show that consumption of white meat or fish is not associated with risk and might even reduce the occurrence of colorectal cancer. A study of Seventh-day Adventists – a predominantly vegetarian population – reported that meat intake was not associated with risk of colorectal cancer.

The data suggest an independent association between red-meat consumption and the risk of colorectal cancer; neither the total protein nor the total fat content of the meat seems to be responsible for this affect.

Calories

Case-control studies suggest that excess calories enhance the risk of colonic carcinogenesis, although it is very difficult to distinguish between energy intake and the intake of dietary fat. In an animal study, carcinogen-induced colon tumors were inhibited by calorie restriction, even though the calorie-restricted rats ingested twice as much fat as control rats. In other animal studies, the incidence and multiplicity of colon tumors were significantly inhibited in animals fed diets containing 20–30% fewer calories than controls.

Obesity appears to be a risk factor for colorectal cancer independent of dietary fat intake. Positive calorie balance and the resulting accumulation of body fat during adult life increase the risk of colon cancer. Even a moderate degree of overweight (body mass index $(\mathrm{kg\,m^{-2}}) > 26$) is associated with increased rates of colorectal cancer. Therefore, maintaining weight in a desirable range appears to be protective.

Vegetables and Fruits

Among the most consistent data that are available from the observational epidemiologic literature are those suggesting that a higher intake of plant foods lowers the risk of cancers at almost every site. The overwhelming majority of descriptive, case-control, and cohort epidemiologic studies have suggested an inverse relationship between consumption of vegetables and fruits and colorectal cancer risk.

There are many biologically plausible reasons why consumption of vegetables and fruits might reduce the likelihood of cancer. Numerous compounds in vegetables and fruits that might exert anticarcinogenic effects have been identified, such as carotenoids, vitamin C and E, folate, flavonoids, phenols, isothiocyanates, and fiber. Each of these phytochemicals and bioactive compounds has been shown to exert some anticarcinogenic activity in laboratory models of cancer.

Vegetables are observed more consistently to convey a protective effect against colorectal cancer than fruit in studies from many countries. This may be due to a difference in composition between vegetables and fruits. An average serving of fruit is substantially higher in sugar, calories, and vitamin C, but lower in carotenoids, vitamin B_6, and folate relative to vegetables. Conversely, raw vegetables tend to be lower in fiber (especially the soluble variety) than both cooked vegetables and fruits. Alternatively, a high vegetable intake, more than a high fruit intake, may merely reflect a generally healthier diet, that is, one relatively low in energy, fat, and sugar.

Although the vast majority of case-control and cohort studies have observed a protective effect of vegetables and fruits on colorectal carcinogenesis, the Iowa Women's Health Study, a large prospective cohort study, observed that total intake of neither vegetables nor fruits reduced the relative risk of colorectal cancer. Similar results were obtained when each vegetable or fruit item was independently analyzed, except garlic. The data from the meta-analyses of colorectal and stomach cancer suggest that a high intake of raw and cooked garlic may be associated with a protective effect against stomach and colorectal cancers. Garlic contains glutathione-*S*-transferase, which is thought to be involved in the detoxification of several potential carcinogens. Individuals consuming more than one garlic clove per week had a 32% reduction in the risk of colorectal cancer compared to those not eating any garlic.

Fiber

The fiber hypothesis is an old hypothesis that originated from an epidemiologic study in which African blacks consuming high-fiber, low-fat diets were observed to have a lower mortality from colorectal cancer than African whites consuming a low-fiber, high-fat diet. The majority of subsequent observational epidemiologic and case-control studies support a protective effect of fiber-rich diets. The effect is thought to be mediated by the shorter intestinal transit time that accompanies a high-fiber diet and the resultant decrease in duration of exposure of the colonic mucosa to potential carcinogens, as well as an increase in stool bulk, which would dilute the colonic contents, including any potential carcinogens.

Fiber is found in vegetables, fruits, whole grains, seeds, nuts, and legumes and consists of nonpolysaccharides (such as lignin) and nonstarch polysaccharides (including cellulose, hemicellulose, pectin, gums, and mucilages). Earlier data suggested that a

protective effect was largely related to less fermentable and more insoluble fibers, especially wheat bran. However, a recent prospective study found a reduced risk of distal colon adenoma with increasing intake of fiber from fruit but not cereals or vegetables. This result suggests that soluble fiber is more protective.

Studies in humans have not yet provided a convincing answer as to whether or not fiber is truly protective. In the Health Professionals Follow-up Study, a large prospective cohort trial, dietary fiber was inversely associated with a risk of developing colorectal adenomas in men, and all sources of fiber were associated with a decreased risk of adenoma. However, data from the Nurses' Health Study, a parallel study in women by the same investigators, failed to indicate any protective effect of dietary fiber on the development of colorectal cancer. A subsequent analysis of the latter cohort, with 10 additional years of follow-up and more thorough adjustment for confounding factors, also showed no protective effect in women with regard to adenomas or cancer. These large prospective epidemiologic studies raise the question as to whether the protective effect of dietary fiber against colorectal cancer may be a gender-specific phenomenon.

A few human intervention trials have demonstrated that wheat-bran supplementation favorably alters a number of biomarkers related to the risk of colorectal cancer, including fecal mutagenecity, fecal secondary bile acids, and rectal cell proliferation. However, two recent well-performed trials which utilized the most widely accepted intermediary biomarker of colon cancer, development of adenomatous polyps, have both yielded negative results: neither a daily dietary supplement containing 13.5 g of wheat bran fiber nor a low-fat, high-fiber diet over 3–4 years reduced the incidence of recurrent adenomas. The results of these two recent studies strongly suggest that dietary supplementation of fiber is not an effective chemopreventive strategy for the prevention of colorectal adenoma recurrence. Nevertheless, the results of these two trials should not be construed as definitive evidence that fiber is ineffective: the follow-up in these trials was relatively brief, and the research question that was addressed (adenoma prevention) is related, but not identical, to the question of cancer prevention.

Antioxidant Nutrients

The antioxidant micronutrients, including vitamin A, carotenoids, vitamin C, vitamin E, selenium, zinc, copper, and manganese, are part of the body's defense against free radicals and reactive oxygen species. These antioxidants are thought to possibly convey protective effects against a number of chronic degenerative diseases through a host of different mechanisms. In some animal studies, vitamins A, C, and E have been shown to have a direct inhibitory effect on tumor growth and damage to both cellular membranes and DNA. However, the epidemiologic literature for the protective effect of antioxidant vitamins in colorectal cancer has little in the way of a consensus. Although early clinical intervention trials suggested a possible preventive role for vitamin E, two recent prospective randomized trials with a more definitive endpoint (adenoma recurrence) have shown no benefit.

Selenium is an essential trace mineral found in cereal grains and seafood. Epidemiologic studies indicate an increased incidence of colorectal cancer in humans in geographic regions where selenium levels are low in the soil and a lower mortality rate of colorectal cancer in areas with high soil selenium levels. The selenium content of foods is determined by the soil in which the food is grown, so assessment of individual dietary intakes using a food content database is of limited value. Results from one study indicate that patients with low plasma selenium levels are at four times the risk for colorectal adenomas. Selenium may alter carcinogen metabolism, affect immune function, protect against oxidant stress, inhibit cell proliferation, and impair tumor metabolism. The most compelling data to data are from a randomized, placebo-controlled intervention trial, where daily supplementation of 200 μg of selenium reduced the incidence of colorectal cancer by 60%.

Folate

Findings from animal and epidemiologic studies indicate that decreased dietary intake of folate or reduced serum or RBC folate levels increases the incidence of colorectal cancer and adenomas. This relationship is underscored in patients with ulcerative colitis, where folate deficiency is commonly induced by increased requirements related to the high turnover of the colonic epithelium, reduced oral intake of nutritious diets, and reduced intestinal absorption resulting from competitive inhibition of folate absorption due to sulfasalazine. Two case-control studies found that folate intake may protect against neoplasia in ulcerative colitis, and the overwhelming consensus of over 15 epidemiologic studies in the general population suggests folate is equally protective in sporadic colorectal cancer. Results from a recent follow-up of the Nurses' Health Study confirmed that colorectal cancer risk was reduced with increased folate intake, as had previously been shown with adenoma recurrence. Interestingly, the study suggested that more

than 10 years of increased intake were necessary to observe a protective effect. Folate deficiency also has been considered as an important factor in alcohol-related enhancement of rectal carcinogenesis because alcohol alters normal folate metabolism in a variety of ways. Although several large multicenter trials are underway to definitively determine the chemopreventive role of folate supplementation in colorectal carcinogenesis, dietary folate intake appears to be inversely associated with the risk of developing colorectal adenomas and cancer.

The mechanism(s) by which folate depletion enhances colorectal carcinogenesis have not been clearly defined, although several candidate mechanisms have been proposed. Folate is an essential factor in a number of critical cellular metabolic pathways that involve the transfer of one-carbon groups, including the *de novo* biosynthesis of deoxyribonucleotides and many biological methylation processes. Therefore, folate deficiency could contribute to DNA damage, instability of the chromatin, impairment of DNA repair, or aberrant patterns of genomic, and/or gene-specific DNA methylation, all of which are phenomena thought to enhance carcinogenesis.

Recently, the relationship of a common polymorphism (C677T) of the methylenetetrahydrofolate reductase (MTHFR) gene (homozygous in 10–15% of the US population) and the risk of colorectal cancer has been examined. MTHFR catalyzes the irreversible conversion of 5,10 methylenetetrahydrofolate to 5-methyltetrahydrofolate. Men with this homozygous mutation have half the risk of colorectal cancer than do homozygous wild-type or heterozygous genotypes. Among men with adequate folate levels, a threefold decrease in risk was observed, but protection associated with the mutation was largely absent in men with a low systemic folate status in the Physician's Health Study. It has been suggested that the cancer protective effect of the MTHFR mutation is related to the increased availability of 5,10 methyleneTHF, and therefore increased ease of nucleotide synthesis. One might speculate that, under low but not high folate conditions, the availability of 5-methyltrahydrofolate for biological methylation constitutes a more critical determinant of whether the cell is pushed down the pathway towards neoplasia.

Calcium, Vitamin D, and Dairy Foods

Many epidemiologic studies suggest an inverse relationship between the risk of colorectal cancer and dietary calcium and/or vitamin D intake. However, the results of epidemiologic research have been inconsistent, even regarding adenomas.

Human clinical trials have been rather promising, although the magnitude of the protective effects have been rather modest. A recent blinded study demonstrated that the administration of low-fat dairy foods, which can double the calcium intake, significantly improved indices of proliferation and differentiation markers in the colonic mucosa. Furthermore, calcium carbonate supplementation (3 g (1200 mg of elemental calcium) daily) provided a significant reduction ($\sim$ 15%) in the risk of recurrent colorectal adenomas as well as reduced numbers of adenomas in a randomized, double-blind trial recently reported by the Calcium Polyp Prevention Study Group. The effect of calcium supplementation was independent of initial dietary fat and dietary calcium intake.

The underlying mechanisms are not clear, but the principal hypothesis is that calcium may precipitate fatty acids and bile acids that are potentially toxic to the colorectal epithelium. Both calcium administration and dairy food administration diminish concentrations of bile acids and fatty acids in the aqueous fraction of the feces, and this is accompanied by a significant decrease in *in vitro* cytotoxicity. Animal studies suggest that supplemental calcium can retard the hyperproliferative stimulus of dietary fat, bile salts, and fatty acids.

Iron

Iron, a necessary nutrient, has been associated with increased cancer risk in humans. *In vitro* and *in vivo* data suggest that iron may be capable of mutagenic effects mediated through free-radical generation or tumor promotion through nutritional mechanisms. As a transition metal, iron possesses loosely bound electrons that are capable of participating in lipid peroxidation reactions. Such reactions are thought to lead to DNA damage and, in some cases, neoplasia.

Prospectively gathered data from a large study suggest that the risk of colon cancer increases in parallel with iron intake throughout the colon for both men and women. A nested case control study similarly showed an increasing trend of risk of cancer of the proximal colon with increasing total iron intake. However, increased body iron stores were not related to the development of colorectal cancer. Even though increased iron intake is related to increased red-meat consumption, the putative relationship between iron intake and colorectal cancer risk remains rather tentative at this point that requires further confirmation.

Heterocyclic Amines

Heterocyclic amines (HCAs) are the carcinogenic and mutagenic chemicals formed from cooking muscle

meats such as beef, pork, fowl, and fish. HCAs form when amino acids and creatine react at high cooking temperatures and are formed in greater quantities when meats are overcooked or blackened. Seventeen different HCAs resulting from the cooking of muscle meats have been identified, which may pose human cancer risk.

HCAs develop a powerful carcinogenic potential when activated by cytochrome P450 1A2-mediated oxidation of the amino group, which is followed by acetylation or sulfation to form direct-acting reactive mutagens that attack key elements in DNA. When fed to rats, several of these HCAs can induce cancer of the mammary glands, colon, and pancreas. Reduction of HCAs in food can be accomplished by avoiding high-temperature cooking, addition of high-antioxidant-containing foods, or addition of soy protein. Nevertheless, there is no definitive evidence that the degree to which meat is cooked is a determinant of colorectal cancer risk.

Alcohol

A large number of epidemiologic studies have examined the association between the risk of colorectal cancer and alcohol consumption. The results are not entirely consistent. Although there are studies showing either increased risk or no association with alcohol intake, there are essentially no studies that show a reduced risk with higher intake.

Alcohol, particularly beer, consumption has been linked most convincingly with rectal cancer. However, there is no strong evidence to suggest that one source of alcohol (e.g., beer) is more associated with risk than alcohol as a whole. Studies of colorectal adenomas, which are colorectal cancer precursors, also have demonstrated an effect for alcohol. The mechanism for an increased risk with alcohol is not known for sure, but the carcinogenicity of acetaldehyde as well as folate and methionine depletion are plausible explanations. Although alcohol is not a direct-acting carcinogen, beer, wines and spirits contain at least 1200 different compounds, such as aldehydes, phenols, and amines, many of which are thought to be related to carcinogenesis.

Tea

Tea is grown in about 30 countries and, next to water, is the most widely consumed beverage in the world. Black tea is consumed primarily in Western countries and in some Asian countries, whereas green tea is consumed primarily in China, Japan, India, and a few countries in North Africa and the Middle East. Most of the studies showing the preventive effects of tea have been conducted with green tea; only a few studies have assessed the usefulness of black tea. Green tea contains polyphenolic compounds, which include flavanoids, flavandiols, flavonoids, and phenolic acids. In black teas, the major polyphenols are theaflavin and thearubigin.

(−)-Epigallocatechin-3-gallate (EGCG) is the primary component of green tea, accounting for 40% of the total polyphenolic mixture. Numerous *in vitro*, human, and animal studies have identified these antioxidant polyphenols, in particular EGCG, as cancer-chemopreventive agents. EGCG inhibits cancer cell growth and represses the catalytic activities of several P450 enzymes such as P450 1A and 2B1. Repression of phase I enzymes involved in cancer initiation, and enhancement of enzymes that play a role in carcinogens detoxification, are thought to guard against carcinogenesis.

Animal studies show that tea consumption protects against colon cancer induced by chemical carcinogens. A cross-sectional study showed a 30% lower risk of adenomatous polyps of the sigmoid colon in persons consuming five or more cups of green tea per day. However, human studies on tea drinking and colorectal cancer are still inconclusive.

Soy

Epidemiologic data suggest that populations that regularly consume soy foods have a lower incidence of colon cancer than populations that do not. Soy is abundant in isoflavones such as genistein and daidzein, which, because of their structural similarity to estrogen and/or their biological activity, are considered to be phytoestrogens. Several other cancer-related properties have also been reported for soy-derived genistein, such as antioxidant activity, an antipromotional effect, inhibition of tyrosine kinase, inhibition of cell-cycle progression and growth and inhibition of angiogenesis. Studies in laboratory animals have further shown that a soy-rich diet inhibits chemically induced carcinogenesis. However, animal studies of the effects of single phytochemicals from soy on intestinal cancer have yielded conflicting results. The relationship between soy intake and the risk of colonic cancer still needs further investigation.

Dietary Recommendations

Diet is important in preventing colorectal cancer. With our current knowledge regarding the dietary risk factors of colorectal cancer, the following suggestions are offered: (1) avoid overweight (BMI > 26), (2) reduce total fat intake to less than 25–30% of total

Table 1 Dietary components that may prevent colorectal cancer

Food or nutrient	*Proposed mechanism*	*Common diet source*	*Estimated effective dose*
Calcium[a]	Reduces bile acid/fatty-acid-induced irritation in colon as well as proliferation	Dairy products, dried beans, kale, broccoli	1200 mg day^{-1}
Vitamin D	Reduces cell proliferation and DNA synthesis, modulates signal transduction pathway, induces apoptosis, promotes repair of membranes and DNA	Fortified milk, salmon, mackerel, fortified cereals, egg yolks	400 IU day^{-1}
Selenium[a]	Reduces risk of excessive oxidative DNA damage and cell toxicity, induces apoptosis, reduces proliferation, restores immune function	Meats, fish, grains, brazil nuts	200 μg day^{-1}
ω-3 fatty acids	Inhibit arachidonic acid metabolism and oxidative reactions, reduce proliferation	Tuna, salmon, sardines, mackerel, and lake trout	3–4 g day^{-1}
Folate[a]	Corrects imbalance of DNA synthesis and DNA methylation, enhances DNA repair	Fortified cereals, citrus fruits and juices, asparagus, spinach, baked beans	400 μg day^{-1}
Fiber	Increases stool bulk, binds with bile acids and other potential carcinogens, reduces fecal pH and anaerobic fecal flora, ferments to short-chain fatty acids	Whole grains, legumes, wheat bran, nuts, barley, beans	20–30 g day^{-1}
Vegetables and fruits[a]	Block the formation of carcinogens, induce a multiplicity of solubilizing and activating enzymes, suppress DNA and protein synthesis		5 more servings day^{-1}
Tea	Antioxidant activity, represses phase I enzymes involved in cancer initiation, and enhances enzymes involved in carcinogen detoxification	Green tea	5 more cups day^{-1}
Soy	Antioxidant activity, antipromotional activity, inhibits cell-cycle progression and growth, and inhibits angiogenesis	Soy bean, tofu, soy milk	

[a]Existing data are particularly compelling.

Table 2 Dietary components that may enhance the risk of colorectal cancer

Food or nutrient	*Proposed mechanism*	*Common diet source*
Animal fat, calories, and red meat[a]	Increasing DNA damage and/or promoting carcinogenesis	Beef, lamb, pork
Heterocyclic amines	Direct-acting reactive mutagens that attack key elements in DNA	Meats cooked at very high temperature (fry, broil, barbecue)
Alcohol	Folate and methionine depletion and carcinogenicity of acetaldehyde, phenol, and amines	Beer, wine, spirit
Iron	DNA damage by lipid peroxidation reactions and mutagenic effects through free-radical generation	Red meat

[a]Existing data are particularly compelling.

calories and saturated fat to less than 10% of total calories, (3) eat five or more servings of fresh vegetables and fruits daily, (4) eat other foods from plant sources, such as breads, cereals, grain products, rice, pasta, or beans several times each day, (5) choose foods low in fat, (6) replace red meat with chicken, fish, nuts, and legumes, (7) eat processed meat in moderation, (8) eat low-fat dairy products, (9) if you drink, keep alcohol consumption moderate (< 2 drinks per day), and (10) consider taking a multiple vitamin containing folic acid, particularly if alcohol is consumed daily. These recommendations are based on the guidelines by the American Cancer Society and currently accepted dietary modulations for the prevention of colorectal cancer (Tables 1 and 2).

Daily supplementation with fiber (20–30 g day^{-1}), calcium (1200 mg day^{-1}), folate (400 μg day^{-1}), fish oils, and selenium (200 μg day^{-1}) have not been proven by prospective randomized control trials to prevent cancer but have considerable potential (Table 1).

See also: **Alcohol**: Properties and Determination; **Amines**; **Calcium**: Properties and Determination;

Cholecalciferol: Properties and Determination; **Energy**: Intake and Energy Requirements; **Epidemiology**; **Fats**: Requirements; **Iron**: Properties and Determination; **Meat**: Nutritional Value; **Soy (Soya) Beans**: Dietary Importance; **Tea**: Chemistry

Further Reading

Alberts DS, Martinez ME, Roe DJ *et al*. (2000) Lack of effect of a high-fiber cereal supplement on the recurrence of colorectal adenomas. *New England Journal of Medicine* 342: 1156–1162.

Baron JA, Beach M, Mandel JS *et al*. (1999) Calcium supplements for the prevention of colorectal adenomas. *New England Journal Medicine* 340: 101–107.

Choi SW and Mason JB (2000) Nutritional chemoprevention of colorectal cancer. *Clinical Perspectives in Gastroenterology* 3: 259–266.

Franceschi S, Parpinel M, Vecchia CL *et al*. (1998) Role of different types of vegetables and fruit in the prevention of cancer of the colon, rectum, and breast. *Epidemiology* 9: 338–341.

Lipkin M, Reddy B, Newmark H and Lamprecht SA (1999) Dietary factors in human colorectal cancer. *Annual Review of Nutrition* 19: 545–586.

Nelson RL, Persky V and Turyk M (1999) Determination of factors responsible for the declining incidence of colorectal cancer. *Diseases of the Colon and Rectum* 42: 741–752.

Schatzkin A, Lanza E, Corle D *et al*. (2000) Lack of effect of a low-fat, high-fiber diet on the recurrence of colorectal adenomas. *New England Journal of Medicine* 342: 1149–1155.

Slattery ML, Edwards SL, Boucher KM, Anderson K and Caan BJ (1999) Lifestyle and colon cancer: An assessment of factors associated with risk. *American Journal of Epidemiology* 150: 869–877.

Potter JD (1999) Colorectal cancer: Molecules and populations. *Journal of the National Cancer Institute* 91: 916–932.

COLORANTS (COLOURANTS)

Contents

Properties and Determination of Natural Pigments

A Lea, The University, White Knights, Reading, UK
B Henry, Phytone Ltd, Burton-on-Trent, UK

Background

Natural colors in foodstuffs may arise in basically three ways. They may be present originally in the plant or animal tissue from which the food is derived (e.g. chlorophyl in green vegetables), they may be generated during processing (e.g., brown Maillard products in fried onions), or they may be added deliberately (e.g., β-carotene in orange squash). This article considers the major pigments in these three groups. (*See* **Browning**: Nonenzymatic; **Carotenoids**: Occurrence, Properties and Determination; **Chlorophyl**.)

Naturally Occurring Pigments

The range of compounds responsible for the natural colors of foodstuffs is surprisingly limited. Plants are the primary sources, yielding anthocyanins, carotenoids, and chlorophyls as the major groupings, with limited contributions from compounds such as betanins and curcumin.

Anthocyanins

The anthocyanins are responsible for most of the red, purple, and blue colors in plant foodstuffs. The skins of red apples, plums, and grapes are rich in anthocyanins, as are strawberries, rhubarb stalks, and red cabbage leaves. Chemically, most anthocyanins are based on the six common benzopyrilium aglycones (anthocyanidins) shown in **Figure 1**. These aglycones are not stable, and the 250 or so anthocyanins which are known in nature reflect the fact that the aglycone is stabilized by substitution with different sugars at different positions. Substitution at the 3 position is

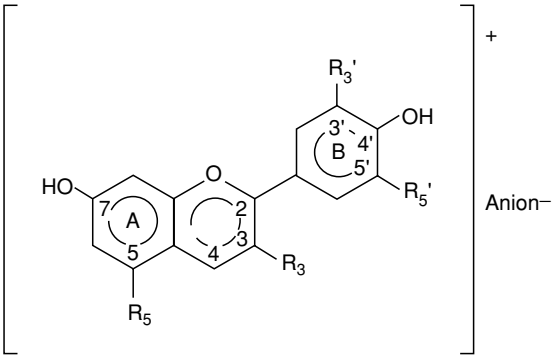

Anthocyanidins: R_3 = OH, R_5 = OH

		R_3'	R_5'
Pelargonidin	(Pg)	H	H
Cyanidin	(Cy)	OH	H
Peonidin	(Pn)	OCH_3	H
Delphinidin	(Dp)	OH	OH
Petunidin	(Pt)	OCH_3	OH
Malvidin	(Mv)	OCH_3	OCH_3

Anthocyanins: Pg. Cy. Pn. Dp. Pt. Mv with
R_3 = *O*-Sugar or *O*-acylated sugar
R_5 = OH or *O*-glucose

Figure 1 The common anthocyanins and anthocyanidins. Reproduced from Colours: Properties and Determination of Natural Pigments, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

most widespread, but 5 and 7 substitution is also common. The substituent sugars may be simple (e.g., glucose, galactose) or complex and unusual (e.g., rutinose, sambubiose). The sugars themselves are sometimes acylated with other complex organic acids – grapes, for instance, contain significant quantities of the caffeic and *p*-coumaric acid conjugates of malvidin 3-glucoside. The polar nature of anthocyanins means that they are soluble only in water or polar organic solvents – they are not associated with lipid in the plant cells and they are not soluble in oils or in organic solvents such as chloroform or hexane.

The color chemistry of anthocyanins is complex. The benzopyrilium structure can exist in four major equilibrium forms depending upon pH (Figure 2). The flavylium salt is the most stable and intensely colored, predominating at low pH. Hence the well-known observation that red cabbage pigment is intense and stable when cooked in the presence of acid, but rapidly turns blue and degrades at a pH near or above neutrality. The blue color is due to the generation of the quinoidal base (which is stable under nitrogen, even at pH 10), which rapidly degrades in the presence of oxygen.

The perceived color of anthocyanins is also highly dependent on copigmentation effects. In the concentrated environment of intact plant cells, the proximity of pectin, metal ions, organic acids, and other polyphenols act to stabilize and to augment anthocyanin color to a remarkable degree. Hence the skins of red grapes seem intensely dark blue or black but, when the pigment is extracted by ethanol during wine-making, the native anthocyanin is revealed as bright red. The visible absorbance maximum of isolated anthocyanins is somewhat dependent on structure, moving from 506 nm (orange) for pelargonidin derivatives to 534 nm (purple-red) for malvidin, delphinidin, and petunidin. A very few anthocyanins are naturally blue, even in the extracted state, due to internal copigmentation with their own sugar-acylating groups.

Even at low pH, the anthocyanins are not stable indefinitely, and they tend to polymerize with other polyphenols or reactive carbonyls. After only a few months' aging, therefore, very little monomeric anthocyanin remains in most red wines and the specific E_{520} drops dramatically. The overall visual chromophore is only slightly affected by polymerization, however, giving a color shift from purple to red. In these circumstances, tristimulus colorimetry is a far more accurate measure of perceived color than is a spot measurement at 520 nm.

Carotenoids

The carotenoids are a range of lipid-associated pigments widely present in higher plants and algae. They are responsible for many of the orange-yellow colors of fruits and vegetables such as red peppers, tomatoes, carrots, and bananas. They are also responsible for the yellow skin ground color of apples such as Golden Delicious. The structure of carotenoids is more diverse than that of anthocyanins and does not fall into such well-defined groups. The basic unit is a C_{40} hydrocarbon structure and those consisting of carbon and hydrogen only are collectively called carotenes, whilst those containing oxygen are called xanthophylls. Some carotenoids are acylic (e.g., lycopene), though the majority contain a six-membered ring at either one end or both ends (e.g., β-carotene). Apocarotenoids are those compounds where an end group has been removed from the normal C_{40} structure.

A few carotenoids appear in animal flesh, such as salmon and shrimp, which acquire them via their diet. The dark blue of the live lobster is a carotenoid–protein complex which is destroyed when the animal is boiled and only the pink carotenoid then remains. Butter, cheese, and egg yolks contain carotenoids which derive ultimately from the vegetable diet of

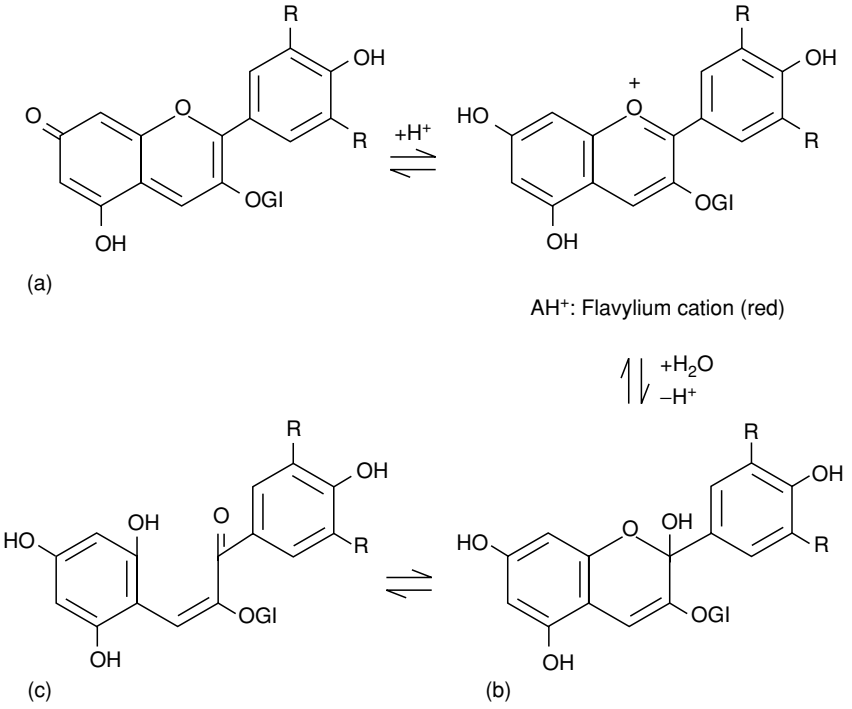

Figure 2 Equilibrium forms of anthocyanins. (a) Quinoidal base (blue); (b) carbinol pseudobase (colorless); (c) chalcone (colorless). Reproduced from Colours: Properties and Determination of Natural Pigments, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

the animals producing them. In animals, many carotenoids act as vitamin A precursors, so have additional value in addition to their properties as pigments. (*See* **Retinol**: Physiology.)

Although some 500 or so carotenoids are known in nature, very few of them are water-soluble – those that are have achieved a disproportionate importance as food colors, as described later. Their general water-insolubility means that they tend to be more stable in processed foods than do anthocyanins, since they remain attached to cellular fragments. Thus, the carotenoid pigments of canned tomatoes are more stable than the anthocyanins of canned strawberries. Nevertheless, carotenoids are still prone to degradation once freed from their cellular matrix, and are particularly susceptible to light and oxygen.

Chlorophyls and Heme Pigments

Chlorophyls are amongst the most noticeable of all plant pigments due to their central photosynthetic role. Like carotenoids, they are water-insoluble and are associated with the lipid portion of plant cells in specialized organelles (the chloroplasts). The bright green of native chlorophyl in plants is determined by the presence of a central magnesium atom in a large porphyrin ring. When green-leaf vegetables are cooked under normal (slightly acid) conditions, this magnesium is easily lost to form the dull green pheophytins. Although this process can be prevented by the use of alkali during cooking and the green color thereby retained, this is no longer a common practice since it leads to oxidation and loss of vitamin C. The green color can also be retained and stabilized if the central magnesium atom is replaced by copper. (*See* **Chlorophyl**.)

The bright red heme pigments, characteristic of animal blood and therefore of fresh meat, contain similar porphyrin ring structures, but the central atom is iron which shifts the chromophore into the red region. Like chlorophyl, the heme pigments are soon degraded when meat is cooked in the normal way. If preserved in the presence of nitrite, however, as in traditional curing, the nitrosohemoglobin stabilizes the pigment and a red color is retained. (*See* **Curing**.)

Other Plant Pigments

Although most orders of higher plants can synthesize anthocyanins to provide red and purple shades, the Centrospermae lack this ability and synthesize instead a group of nitrogenous pigments known as the betalains. These occur in both red forms (betanins) and yellow forms (vulgaxanthins) (Figure 3). The betalains are water-soluble and are generally stable

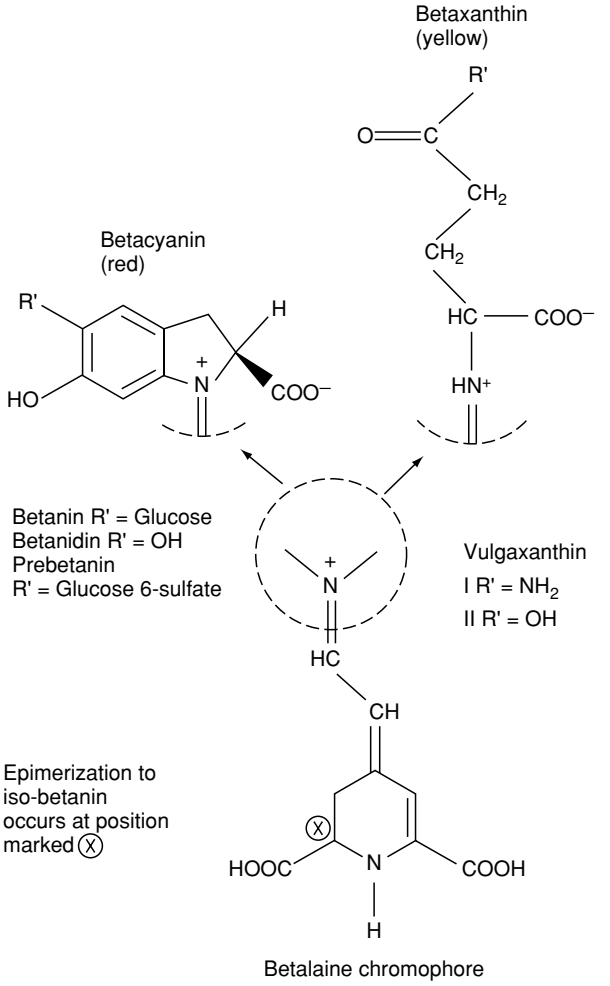

Figure 3 The structure of betalains. Reproduced from Colours: Properties and Determination of Natural Pigments, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

at moderately acid pH, although heat, light, and oxygen will hasten their decomposition. Although of restricted natural distribution, the importance of betalains, particularly the red betanins, is enhanced by their presence in a number of important food crops such as amaranth, chard, and beetroot.

The yellow tubers of turmeric (*Curcuma longa*) are well known as both a spice and a coloring material for food when dried and ground into a powder. The pigment, known as curcumin, has the extended conjugated structure shown in Figure 4. This is subject to keto/enol equilibrium and the pigment converts to an unstable orange form above pH 7.

Pigments Generated During Processing

Pigments generated during processing or cooking are amongst the least well defined in terms of chemical

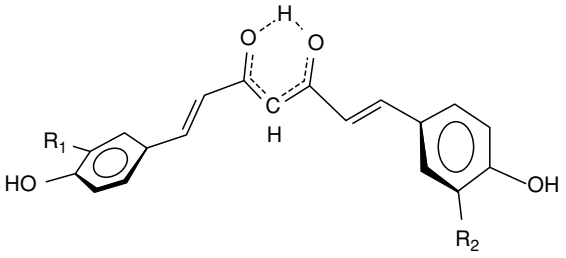

Figure 4 The structure of curcumin. R = H or OCH_3. Reproduced from Colours: Properties and Determination of Natural Pigments, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

structure. They fall into two broad groups – those produced by enzymatic browning (principally from natural phenolics and polyphenoloxidase) and those produced by (nonenzymatic) Maillard reactions between reducing sugars and amino compounds.

Enzymatic Browning

Colors of this sort are generated by the action of polyphenoloxidase or peroxidase enzymes on natural phenolic substrates. The development of color in a freshly cut apple is a typical example. The principal substrates involved are the catechins or procyanidins, although in older literature the phenolic acids are often mentioned. It is now known that phenolic acids do not themselves brown, although they may play a part in coupled oxidation systems which allow the procyanidins to generate the color. The pigments derived from procyanidins have no specific λ_{max} in the visible region of the spectrum, although those derived from simpler phenolics such as catechins may initially display yellow-orange chromophores about 400 nm before further polymerization occurs.

In some food systems, such as apple juice or white wine, only a limited amount of color development is desirable before the product becomes unacceptably dark. In these cases, color formation may be inhibited by antioxidants such as sulfite or by a drop in pH which reduces enzyme activity. In other foods, such as black tea or cocoa, the development of color during processing is critical to the product and is also associated with flavor development processes. The tea system has been well studied and is known to generate two types of phenolic pigment from catechins. These are the theaflavins, which are bright orange-red, and the thearubigins, which are dull brown. Most oxidized phenolic pigments tend to become browner and duller with time, particularly if further nonspecific oxidation takes place. (*See* **Antioxidants**: Synthetic Antioxidants.)

Nonenzymatic (Maillard) Browning

This form of color development is usually associated with heat and is a type of caramelization, in which the temperature for degradation of a reducing sugar is lowered by the presence of catalytic amino groups. The sugar degrades to give aldehydic functions which polymerize to give visible chromophores, intensifying after further complex formation with amino groups. Typical desirable examples include the browning of fried onions, the roasting of meat, or the manufacture of toffee confectionery.

Almost inevitably, Maillard browning is associated with flavor development. The pigments generated are even less well defined than those from the phenolic browning reaction. Although temperatures above 80 °C will generate Maillard pigments rapidly in suitable systems, the reactions will also proceed slowly at room temperature if the water activity is low enough. Thus, apple juice concentrate held at 20 °C for several weeks shows discernible color development from Maillard reactions between fructose and amino acids (in addition to any enzymatic browning which may have taken place earlier). Similarly, tinned condensed milk or dried milk powder develops color in storage due to reaction between lactose and the lysine ε-amino groups in milk protein. Such slow Maillard reactions are generally considered undesirable, in contrast to the fast reactions which are under control during cooking.

Natural Pigments Deliberately Added to Foodstuffs

Most of the natural pigments added to food are derived from one of the categories described above. A list of those permitted in the European Union is given in Tables 1 and 2.

Table 1 Natural and nature-identical colorants (and those of natural origin) covered by European Union regulations

Colorant	*E number*
Reds	
Anthocyanins	E163
Beetroot red	E162
Cochineal, carmines	E120
Orange-Yellow	
Annatto	E160(b)
β-apo-8′-carotenal	E160(e)
Canthaxanthin	E161(g)
Mixed carotenes and β-carotene	E160(a)
Curcumin	E100
Ethyl ester of β-apo-8′-carotenoic acid	E160(f)
Lutein	E161(b)
Lycopene	E160(d)
Paprika extract	E160(c)
Riboflavin, riboflavin-5-phosphate	E101
Greens	
Chlorophyls and chlorophyllins	E140
Copper complexes of chlorophyls and chlorophyllins	E141
Brown	
Plain caramel	E150(a)
Caustic sulfite caramel	E150(b)
Ammonia caramel	E150(c)
Sulfite ammonia caramel	E150(d)
Black	
Vegetable carbon	E153

Table 2 Inorganic colorants covered by European Union regulations

Colorant	*E number*
Calcium carbonate	E170
Titanium dioxide	E171
Iron oxides and hydroxides	E172
Aluminum	E173
Silver	E174
Gold	E175

Blues and Reds

Most anthocyanin extracts are derived from grape skins as a byproduct of wine or grape-juice manufacture. Some extracts are also available from black carrot, elderberry, red cabbage, purple corn, and blackcurrant. A particularly interesting extract is that from red cabbage which contains highly copigmented anthocyanins of much bluer shade than most fruit-derived anthocyanins. The extracts are generally made using acidified water or alcohol, sometimes in the presence of sulfite (which is later removed). The extracts are then concentrated to a liquid or dried to a powder. None of the extracts are pure compounds but consist of a variety of anthocyanin structures complete with polymeric degradation products.

Anthocyanin extracts are well suited to aqueous systems but they cannot be directly used in fatty foods or in foods of pH much greater than 4.5. They are susceptible to copigmentation and complexation effects – for instance, the presence of tin, iron, or aluminum (as in canned fruits) can cause noticeable blueing.

Native (monomeric) anthocyanins are reversibly bleached by sulfite but are irreversibly degraded by ascorbic acid. Polymeric or copigmented anthocyanins are generally more resistant to degradation and are therefore particularly suited to use in soft drinks. Other phenolic compounds such as procyanidin tannins will tend to cause coprecipitation or haze formation. The crude extracts of grape skin which

are high in tannins may also react with food proteins such as gelatin. Anthocyanins, particularly those derived from grapes or black carrot, have found considerable application in providing a natural coloring for soft drinks where they are stable at low pH and moderately resistant to light and pasteurization.

Betanin extracts are prepared from beetroot juice which may be fermented to remove sugar and then dried to a powder. Although generally stable between pH 3 and 7, they are adversely affected by heat, light, oxygen, high water activity, and sulfite. However they have found considerable application in frozen and short-shelf-life dairy products such as icecream and yogurt, and in dry-mix desserts.

The traditional red color cochineal is derived from the dried South American insect (*Coccus cacti*) and consists of a polyhydroxy anthroquinone pigment known as carminic acid. Its aluminum lake, known as carmine, is produced as an insoluble powder which can be converted to a water-soluble product by dissolution into dilute alkali. It is very stable to heat, light, and oxygen and finds application in a wide range of foodstuffs, including meat and dairy products, sugar confectionery, and certain soft drinks.

Orange and Yellow

The carotenoids are generally oil-soluble and therefore well suited for coloring fatty foods such as margarine and dairy spreads. Natural sources include paprika, lutein from *Tagetes erecta* (Aztec marigold), carotenes from palm oil, and lycopene from tomato. Natural carotenes may be incorporated into an oil-in-water emulsion to facilitate their dissolution into soft drinks. A commonly used extract is annatto, which is obtained from the seeds of a tropical shrub, *Bixa orellana*. This consists mainly of bixin, the fat-soluble methyl ester of a carotenoid carboxylic acid known as norbixin, which is a particularly intense colorant providing distinctive colors at dose levels as low as 10 p.p.m. It can be rendered completely water-soluble by alkaline hydrolysis to yield the sodium or potassium salts. These readily complex with protein and are therefore ideally suited as colorants for smoked fish, cheese, and other dairy products. Norbixin can precipitate out at low pH and thus an annatto preparation formulated to color cheese cannot be used to color a soft drink where special-application forms of norbixin are required. Bixin, however, is not affected by pH.

Another traditional water-soluble carotenoid is crocin, an ester of the dicarboxylic acid crocetin, which is the pigment from saffron (the stigmas of *Crocus sativus*). This pigment is also extracted from the seeds of certain oriental *Gardenia* spp.

Synthetic but 'nature-identical' carotenoids are also available as food colorants, principally β-carotene and β-apocarotenal. These are generally used in oil-based systems but may also be made water-dispersible by physical means (emulsification and encapsulation), and thus are commonly used for coloring cloudy-orange soft drinks. Canthaxanthin may be incorporated into the diet of farmed salmon to replace the carotenoids which wild fish would naturally consume, thereby insuring a pink flesh. Synthetic carotenoids and natural paprika and lutein extracts are also used in poultry feed to insure either that egg yolks are of the required color or that the skin of the chicken is of a suitable shade of yellow.

Curcumin, a deodorized turmeric extract, also provides a useful yellow pigment, although its application is limited by its sensitivity to light and its pH dependence. New application forms, however, now allow for its use in toppings and fillings for flour confectionery as well as its more common use in dairy products, desserts, sugar confectionery, soups, and sauces. It has found considerable application, together with annatto, in providing the yellow shade required for vanilla icecream and certain cookies (biscuits) and cakes.

Greens

Chlorophyl extracts are prepared using organic solvents from sources such as grass and lucerne leaves. The green extracts may be used to color fatty foods directly, but are more usually converted into their copper complexes to give a brighter green, followed by preparation of water-soluble sodium or potassium salts. In this form they are used to color sugar confectionery, frozen desserts, and dairy products.

Browns and Blacks

These colors are principally obtained by the use of synthetic caramels and carbon blacks. Four types of synthetic caramel are recognized, which differ in their solubilities and isoelectric points. This affects their stability towards such factors as protein binding (in beers) and alcohol-solubility (in spirits). The appropriate caramel must therefore be chosen for the product in hand. Class IV caramels E150(d) are used throughout the world to color cola beverages. Recently, 'natural' caramels based on malt extracts have also become available. All these colors are based on Maillard browning reactions. Brown pigments based on enzymatic oxidation have not generally been used for addition to food systems, principally

on account of their low tinctorial power and instability to further oxidation.

Carbon black is an insoluble pigment derived from the controlled combustion of plant material and is usually converted into a liquid suspension by milling into a carrier such as sugar syrup or fat. In this form it is used in the manufacture of black sugar confectionary.

Analysis of Natural Colors

Analysis of natural colors as individual pigments is really only possible using chromatographic techniques, although the use of tristimulus colorimetry is valuable in defining and maintaining the exact visual specification of a color in a foodstuff. Extraction of pigments requires specific techniques, e.g., acidified alcohol/water mixtures for anthocyanins, ethyl acetate or dichloromethane for carotenoids. Most analyses are nowadays carried out by high-performance liquid chromatography (HPLC) on reverse-phase columns. For anthocyanins, water/alcohol gradients at pH 1.5 are suitable. For carotenoids and chlorophyls, mixtures of dichloromethane, acetonitrile, and tetrahydrofuran are often used. Detection must be in the visible spectrum – diode array detectors can be helpful in confirming peak identity. Isolation of standards is often a major difficulty since few are commercially available.

In processed food the added pigments may be degraded or oxidized during processing, or may become so tightly bound to the food matrix that their total recovery is impossible. For these reasons the quantitation of added anthocyanins or carotenoids to a foodstuff is only approximate at best, especially if these pigments have themselves been added as naturally occurring mixtures without a defined composition. If simple synthetic β-carotene has been added to a food, it can be extracted by saponification under nitrogen followed by ether extraction and HPLC, and then quantified with reference to an external standard. However, the qualitative identification of pigment origin can be a powerful tool. For instance, fruit desserts colored with grapeskin or with black carrot extracts can be readily identified from each other, and likewise sauces colored with paprika extracts or β-carotene can be easily distinguished. Caramels and phenolic oxidation products cannot be analyzed as single chemical entities due to their heterogeneous nature, although the presence of byproducts such as hydroxymethylfurfural may be useful for the indirect assessment of caramel addition. New techniques such as capillary zone electrophoresis show promise for the analysis of charged materials such as the class IV E150(d) caramels.

See also: **Antioxidants**: Synthetic Antioxidants; **Browning**: Nonenzymatic; **Caramel**: Properties and Analysis; **Carotenoids**: Occurrence, Properties, and Determination; **Chlorophyl**; **Chromatography**: High-performance Liquid Chromatography; **Curing**; **Retinol**: Physiology

Further Reading

Blake CJ (1982) *Determination of Natural Colours in Foods*. *Scientific and Technical Survey* no. 130. Leatherhead: Leatherhead Food Research Association.

Coulson J (1980) Naturally occurring colouring matters for foods. In: Walford J (ed.) *Developments in Food Colours*, vol. 1, pp. 189–218. London: Elsevier.

Henry BS (1996) Natural food colours. In: Hendry GAF and Houghton JD (eds) *Natural Food Colorants*, 2nd edn, pp. 40–79. London: Blackie.

Knewstubb CJ and Henry BS (1988) *Natural Colours – A Challenge and an Opportunity*, pp. 179–186. London: Institute of Food Science and Technology.

Lea AGH (1988) HPLC of natural pigments in foodstuffs. In: Macrae R (ed.) *HPLC in Food Analysis*, 2nd edn, pp. 277–333. London: Academic Press.

Markakis P (1982) *Anthocyanins as Food Colours*. London: Academic Press.

Taylor AJ (1984) Natural colours in food. In: Walford J (ed.) *Developments in Food Colours*, vol. 2, pp. 159–206. London: Elsevier.

Timberlake CF and Henry BS (1986) Plant pigments as natural food colours. *Endeavour* 10: 31–36.

Properties and Determinants of Synthetics Pigments

M J Scotter, Central Science Laboratory, Sand Hutton, York, UK

Introduction

This article will review the use of synthetic coloring materials in foodstuffs, structural classification, chemical stability, interactions with other food additives, usage in foods, and their qualitative and quantitative analysis.

Classification

Synthetic food colors classification can be simplified by grouping into the following chemical classes:

Table 1 Synthetic food colors currently permitted in the European Union and USA

Name	*E number*[a]	*FD&C classification*[b]	*Color index no.*[c]	*Structure type*	*Color shade*
Tartrazine	E 102	Yellow no. 5	19140	Azopyrazolone	Yellow
Quinoline yellow	E 104	None	47005	Quinoline	Greenish yellow
Sunset yellow FCF	E 110	Yellow no. 6	15985	Monoazo	Orange yellow
Carmoisine	E 122	None	14720	Monoazo	Bluish red
Amaranth	E 123	None	16185	Monoazo	Red
Ponceau 4R	E 124	None	16255	Monoazo	Orange red
Erythrosine	E 127	Red no. 3	45430	Xanthene	Bluish pink
Red 2G	E 128	None	18050	Monoazo	Bluish red
Allura red AC	E 129	Red no. 40	16035	Monoazo	Yellowish red
Patent blue V	E 131	None	42051	Triarylmethane	Violet blue
Indigo carmine	E 132	Blue no. 2	73015	Indigoid	Deep blue
Brilliant blue FCF	E 133	Blue no. 1	42090	Triarylmethane	Greenish blue
Green S	E 142	None	44090	Triarylmethane	Bluish green
Black PN	E 151	None	28440	Disazo	Bluish black
Brown FK	E 154	None		Azo[d]	Orange brown
Brown HT	E 155	None	20285	Disazo	Dark brown
Lithol rubine BK[e]	E 180	None	15850	Monoazo	Bluish red
Citrus red[f]		Red no. 2	12156	Monoazo	Scarlet red
Fast green FCF		Green no. 3	42053	Triarylmethane	Bluish green
Orange B[g]		Orange B	19235	Azopyrazolone	Orange

[a]As given in European parliament and Council Directive 94/326/EC on colors for use in foodstuffs.
[b]As published by the US Food and Drug Administration.
[c]As published by The Colour Index, Society of Colourists and Dyers, Bradford, UK.
[d]Mixture of six main components; only used for coloring kippers.
[e]For coloring of cheese rind only.
[f]For coloring of orange skins only.
[g]For coloring of casings or surfaces of frankfurters and sausages only.
Note: Yellow 2G is no longer permitted in the European Union.

- Azo (monoazo, disazo, and trisazo)
- Azo-pyrazolone
- Triarylmethane
- Xanthene
- Quinoline
- Indigoid

Table 1 lists the food colors currently permitted in the UK, European Union (EU), and USA and gives their structural classes.

Azo Food Colors

Azo dyes contain one or more chromophoric azo groups which are usually associated with aromatic systems containing salt-forming substituents generally in the *meta* or *para* position to the azo group. Azo dyes span a wide range of colors: one example is sunset yellow FCF (Figure 1a).

Azo-Pyrazolone Colors

Azo dyes which also contain a pyrazolone group exist essentially as keto-hydrazine tautomeric systems. A well-known example of an azo-pyrazolone dye is tartrazine (Figure 1b).

Triarylmethane Food Colors

Triarylmethane colors are characterized by the presence of a chromophoric system containing a central carbon atom attached to three aromatic moeties with amino, substituted amino and hydroxyl groups substituted in the *para* position, which act as auxochromes. One example of a triarylmethane food dye is green S (Figure 1c).

Xanthene Food Colors

Xanthene dyes are characterized by a chromophoric system comprising essentially a dibenzo-1,4-pyran heterocyclic ring system with amino or hydroxyl groups in the *meta* position with respect to the oxygen bridge. Erythrosine is the only example of a xanthene dye currently permitted for food use in the EU and the USA (Figure 1d).

Quinoline Food Colors

Quinoline yellow (Figure 1e) is the only example of a quinoline dye currently permitted for food use in the EU. The chromophoric system is based on the 2-(2-quinolyl)-1,3-indandione (or quinophthalone) heterocyclic ring system.

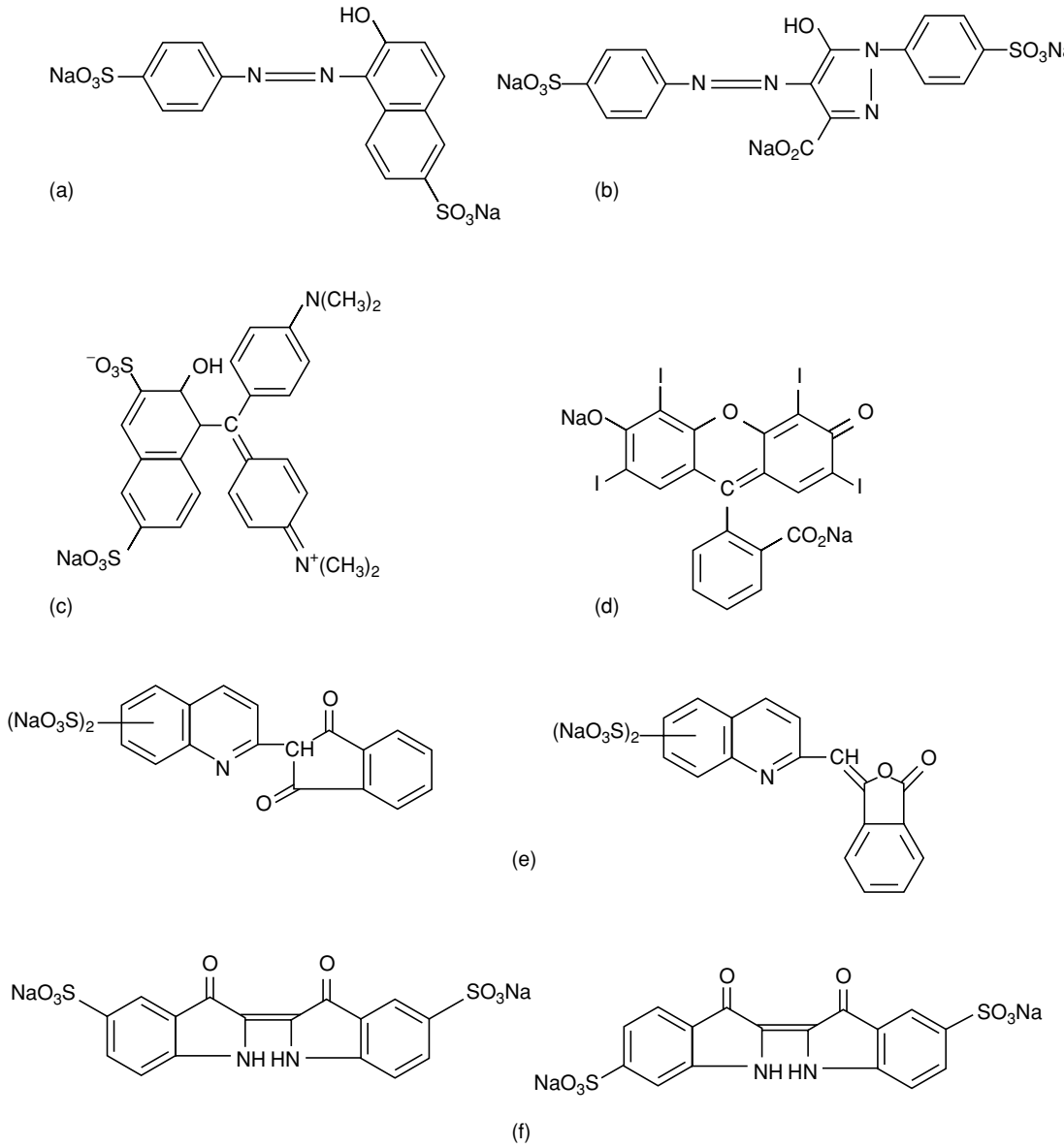

Figure 1 (a) Chemical structure of azo dye sunset yellow FCF. (b) Chemical structure of azo-pyrazolone dye tartrazine. (c) Chemical structure of triarylmethane dye green S. (d) Chemical structure of xanthene dye erythrosine. (e) Chemical structure of quinoline dye quinoline yellow, showing the two main coloring components. (f) Chemical structure of indigoid dye indigo carmine, showing the two main coloring components. Reproduced from Colours: Properties and Determination of Synthetic Pigments, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds) 1993, Academic Press.

Indigoid Food Colors

The only example of an indigoid dye currently permitted for food use in the EU and the USA is indigo carmine (Figure 1f), a blue-violet sulfonated analog of indigo, a naturally occurring dye which exists as a resonance equilibrium between two hybrid structures.

Pigments and Lakes

Pigments are dyestuffs which are generally insoluble in aqueous and organic solvents and have to be dispersed into foodstuffs to effect coloration. Precipitation of water-soluble dyes on to an inert substrate such as alumina forms water-insoluble pigments known as lakes:

$3(\text{Dyestuff-}SO_3^-) + Al^{3+} \rightarrow (\text{Dyestuff-}SO_3^-)_3Al^{3+}$ precipitated and extended on to $Al_2O_3.3H_2O$.

No synthetic liposoluble dyestuffs are permitted in the EU for the coloring of food. Synthetic nature-identical dyestuffs are generally not considered as synthetic food dyes.

Chemistry and Stability

All of the synthetic coloring materials permitted for food use in the EU and USA (except lithol rubine BK and citrus red, which have restricted use), apart from the lake colors, are soluble in water to a greater or lesser extent and insoluble in fats. The degree of water-solubility is determined by the number and relative position of salt-forming groups present in the dye molecule. The most common of these is the sulfonic acid group $-SO_3H$ and the less common carboxylic acid group $-CO_2H$, which form water-soluble anionic dyes. Cationic dyes contain basic groups such as amino $-NH_2$ or substituted amino $-NH.CH_3$, $-N(CH_3)_2$.

Most food dyes are soluble in certain nonaqueous hydrophilic solvents such as glycerine, propylene glycol, and sorbitol, and this allows the preparation of solutions and pastes for use in certain food products. The triarylmethanes and erythrosine are appreciably soluble in the lower alcohols ethanol and isopropyl alcohol. Turbidity or precipitation of color may be experienced upon interaction with hard water.

Coloring materials exhibit excellent stability when stored under cool, dry, and dark conditions. Many factors can and do contribute to colorant stability, such as heat, light, pH, redox systems, other food ingredients (especially preservatives), and trace metals.

Photodegradation

Light is capable of inducing photochemical changes in all dyestuffs, eventually leading to total decolorization. Resistance to photochemical degradation is termed light fastness. Heat, other various agents, and food ingredients are known to accelerate the photodegradation of dyestuffs, whereas others prove to have a stabilizing effect. True azo dyes can undergo three principal types of photochemical reaction; *cis-trans* photoisomerism, photoreduction, and photo-oxidation.

Thermal Degradation

Heat can cause losses of color during food processing and cooking. Coloring materials are added to products at the latter stages of and at the lowest possible temperatures during food processing when further heating is unlikely to take place. For all dyes, processing at very high temperatures will lead to an inevitable loss of color or change in shade due to carbonization.

Acids, Alkalis, and Redox Systems

Not all colors can be used over all pH values and some coloring materials, such as erythrosine, may precipitate from solution at acid pH, whereas others such as indigo carmine will fade rapidly. Color lakes often exhibit amphoteric properties, with both acids and alkalis tending to solubilize the inorganic substrate and thus releasing the free colorant (i.e., color 'bleed').

The majority of permitted food colors exhibit instability when used in combination with oxidizing and reducing agents. Since color depends on the existence of a conjugated unsaturated system within the dye molecule, any substance which modifies this system (e.g., oxidizing or reducing agents such as hydrogen, sugars, acids, and salts) will affect the color.

Metals

All dyes, those with the azo group in particular, will exhibit accelerated fading under both acid and alkaline conditions in the presence of metals including zinc, tin, aluminum, iron, and copper, especially at higher temperatures. This is mostly due to the reducing effect of liberated hydrogen. Dyes will often react with the metal in food cans at a rate proportional to their concentration.

Interaction with Other Food Additives

Preservatives

Canned products containing added color may degrade in the presence of tartaric and citric acids, which may react with the metal of the container to liberate hydrogen. The stability of nine red colors in comminuted meat products in the absence and presence of nitrite has shown that most of the dyes are destroyed to some extent but with nitrite more of the color survives. Subsidiary dye components and colorless fluorescent products are formed as a result of heat processing and in certain cases additional products are observed in the presence of nitrite. Nitrite can also cause rapid detinning to produce Sn^{2+}, a strong reducing agent. Sulfur dioxide is known to cause rapid decolorization of dye solutions. The interaction between sulfite and carmoisine results in the formation of a hydrazo compound via hydrolysis.

Ascorbic Acid

Azo dyes are known to degrade under accelerated conditions in the presence of ascorbic acid at pH 7 but are more resistant to degradation at pH 3. However, ascorbic acid at a concentration of 50 mg l^{-1} can affect all synthetic coloring materials at pH 3 after 3 days' storage at 20 °C.

Sugars

Reducing sugars such as glucose and fructose can reduce azo dyes in aqueous solution. Amaranth in particular may degrade when incorporated into reducing sugar-containing foods which are baked. The presence of baking soda can promote the degradation of the dye markedly in the presence of glucose.

Model Food Systems

During accelerated storage in model soft drink systems, tartrazine degrades very little. Amaranth and sunset yellow FCF are also stable to most additives, with the exception of ascorbic acid and sodium metabisulfite. In the latter case the degradation products appear to be higher-sulfonated analogs of the parent dyes. Prolonged storage under these conditions can therefore cause irreversible degradation leading to colorless products. The formation of specific amines from the degradation of amaranth has been used to estimate the amounts of dye added to soft drinks. Red dyes have been shown to degrade both before and after storage in fish paste products. Ponceau 4R may be reduced to a yellow by hydrogen sulfide or sulfur compounds liberated during processing and storage of certain foodstuffs, and erythrosine may lose iodine to produce fluorescein when incorporated in canned cherries and stored in unlacquered cans.

Color Usage in Processed Foods

Certain food processes have been associated with the use of food coloring matter for some time, principally to:

- reinforce colors already present in foods to meet consumer expectations
- insure uniformity of colour in batch productions
- restore the original appearance of certain foods when color has been, or will be, diminished during processing or storage
- give color to otherwise colorless foods such as candies (sweets), instant desserts, and ice lollies

Table 2 gives a brief summary of color usage in foods in the UK in 1987.

Dye Purity

Most of the dyes used for coloring foods comprise several colored components as well as the main dye. These are collectively known as subsidiary colors. The manufacture of a dye from its starting materials usually involves a number of synthetic stages and transformations such as reduction, amination, sulfonation, diazotization, condensation, and oxidation. The side-reaction products and the precursors of the dyes themselves are collectively known as 'intermediates' and in food dyes these are often sulfonated compounds. Table 3 lists those most commonly found in food colors. Coloring material specifications in general therefore contain criteria for limitations on subsidiary dyes and intermediates as well as certain unsulfonated or free aromatic amines. Separate criteria are prescribed for inorganic impurities such as transition metals, heavy metals, and certain salts.

Inorganic Impurities

The most commonly found inorganic impurities in food colors are sodium chloride and sodium sulfate. Small amounts of phosphate, acetate, carbonate, and iodide may also be present. The criteria for purity with respect to inorganic matter are somewhat different for lake colors.

Organic Impurities

The most common organic impurities present in synthetic coloring materials are small amounts of reaction intermediates. There may be various unsulfonated aromatic compounds as well as the sulfonated analogs present in the finished colorants owing to impurities in the starting materials. Triarylmethane dyes are prepared by condensation reactions during which an uncolored leuco base is formed as an intermediate. The leuco base is then oxidized to the fully conjugated colored dyestuff using oxidizing agents such as lead dioxide, manganese dioxide, or dichromate, which might then be present as low-level inorganic contaminants in the finished dye.

Analysis of Synthetic Food Colors

The major components of synthetic water-soluble food color formulations are active dye (including subsidiary dyes), inorganic salts, and moisture. Other constituents may be permitted diluents or extenders which may be added for standardization purposes or to facilitate the incorporation of the colorant for certain applications.

Table 2 Guide to synthetic food coloring usage in the UK[a]

Commodity class	*Food types*	*Stability requirements*[b]	*Additional comments*	*Permitted colors*[c]	*Typical levels of application (mg kg^{-1} or mg l^{-1})*
Soft drinks and other nonalcoholic beverages	Ready-to-drink cordials, vending machine concentrates, instant teas	LF, AC, PR, FL, TM	Must not accelerate corrosion of metal containers	TZ, SY, P4R, CA, BB, GS, AR, BV, IC, QY, BHT	10–100 Proportionally higher in concentrates
Alcoholic beverages	Beers, ciders, fortified and aromatized wines, spirituous beverages	SD	Limited use	TZ, SY, AM, P4R, CA, BB, GS, AR, BV, IC, QY, BHT, BPN	Up to 200
Confectionery	Boiled sweets, toffees, caramels, gums, jellies, pastilles, licorice, chewing gum Decorations and coatings	TM, SD, FL, AC	Added as late as possible during production; colors often bright/intense	TZ, SY, P4R, CA, BB, GS, AR, BV, IC, QY, BHT, BPN	50–300
Fine bakery wares	Biscuits, wafers, cakes, baking ingredients	TM	Raising agents may be present	TZ, SY, P4R, CA, BB, GS, AR, BV, IC, QY, BHT, BPN	Up to 200
Soups		TM, AC	Requires consistent staining of product relative to carrying liquors	TZ, SY, P4R, CA, BB, GS, AR, BV, IC, QY, BHT, BPN	Up to 50
Meat products	Breakfast sausage Burger meat Luncheon meat	TM, PR	Must show stability in and affinity for protein	AR, R2G AR, R2G AR	25, 20 -, 20 25, -
Meat and fish analogs	Based on vegetable proteins	TM, PR	Must show stability in and affinity for protein, Lakes often used	TZ, SY, P4R, CA, BB, GS, AR, BV, IC, QY, BHT, BPN	Up to 100
Fish and fish products; shellfish	Smoked fish Salmon substitutes Surimi Fish roe Fish paste and crustacean paste Precooked crustaceans Kippers	TM, PR	Must show stability in brine and affinity for protein; surface coloring often used	TZ, SY, P4R, CA, BB, GS, AR, BV, IC, QY, BHT, BPN; AM only in fish roe[e] BFK	100 500 500 300 100 250 20
Cheese	Flavored processed cheese Edible cheese rind Edible cheese casing	AC, PR	Color must not migrate from rind/casing into cheese	TZ, SY, P4R, CA, BB, GS, AR, BV, IC, QY, BHT, BPN (LBK cheese rind only)	100 QS QS
Special dietary foods	Solid food supplements Liquid food supplements Complete formulae	Various	Most have special requirements	TZ, SY, P4R, CA, BB, GS, AR, BV, IC, QY, BHT, BPN	Up to 300 Up to 100 Up to 50
Preserves	Jam, jelly, and marmalade Preserves of red fruits Candied fruits	TM, AC	Must be stable to high temperatures found in jam-making and migration	GS, P4R, SY, QY	Up to 100
Desserts, including flavored milk products	Blancmanges, custards, mousses, dry mixes, sauces Edible ices, icecream	TM, LF, GE	Lakes often used but must not show speckiness in product	TZ, SY, P4R, CA, BB, GS, AR, BV, IC, QY, BHT, BPN	Up to 150

Continued

Table 2 Continued

Commodity class	*Food types*	*Stability requirements*	*Additional comments*	*Permitted colors*	*Typical levels of application (mg kg^{-1} or mg l^{-1})*
Commodity class	*Food types*	*Stability*	*requirements*[b]	*Additional comments*	*Permitted colors*[c]
Typical levels of application (mg kg^{-1} or mg l^{-1})					
Snack foods: dry, savoury potato, cereal- or starch-based	Extruded or expanded savoury products	LF, TM	Products often surface-treated	TZ, SY, P4R, CA, BB, GS, AR, BV, IC, QY, BHT, BPN	Up to 200
	Other savoury products and nuts				Up to 100
Sauces, seasonings	Pickles, relishes, chutney, curry powder, tandoori	TM, AC, LF	Most have special requirements; dye mixtures often used	TZ, SY, P4R, CA, BB, GS, AR, BV, IC, QY, BHT, BPN	Up to 500
	Mustard				Up to 300
Other	Cocktail and candied cherries in syrup and cocktails		Cherry color must not leach	ERY	150–200
	Processed mushy and garden peas (canned)	TM, AC	Requires consistent coloring	TZ, GS, BB	Up to 100

[a]Based on Ministry of Agriculture, Fisheres, and Food (1987) *UK Survey of Colour Usage in Food, Food Surveillance* paper no. 19. London: HMSO; and European parliament and Council Directive 94/326/EC on colors for use in foodstuffs.
[b]LF, light-fastness; AC, acids; PR, preservatives; FL, flavorings; SD, sulfur dioxide; GE, gelling agents and emulsifiers; TM, temperature.
[c]TZ, tartrazine; SY, sunset yellow FCF; P4R, ponceau 4R; AM, amaranth; CA, carmoisine; BB, brilliant blue FCF; GS, green S; BHT, brown HT; IC, indigo carmine; BPN, black PN; ERY, erythrosine; R2G, red 2G; BFK, brown FK; BV, patent blue V; QY, quinoline yellow; AR, allura red; LBK, lithorubine K; QS, *quantum satis* (no maximum level specified).

Dye Content

Azo, triarylmethane, and indigoid dyes are readily reducible by tin (II) chloride, which forms the basis of the titrimetric assay method for dye content. Erythrosine (xanthene type) and quinoline yellow (quinophthalone type) are not reducible and have to be assayed by other means. However, erythrosine is the only commonly used food dye which is insoluble in dilute acid and which can consequently be assayed gravimetrically.

Simple spectrophotometric methods are commonly used for the assay of food dyes. Measurement in the ultraviolet/visible range on instruments with scanning and recording facilities is common. The use of high-performance liquid chromatography (HPLC) for the determination of dyestuff purity is also widespread.

Inorganic Salts

The classical method for the determination of chloride is by the precipitation of chloride as its silver salt. Sulfate may also be determined gravimetrically as the barium salt. Electrometric procedures may also be used for specific ion determination. Chloride, sulfate, and other ionic species may also be determined simultaneously using ion chromatography.

Moisture

Standard procedures for moisture determination are loss on drying, or by nonaqueous titrimetric techniques such as the Karl–Fischer procedure.

Lakes

Alumina is one of the major components of food lakes but is rarely determined. Food lakes are generally soluble in hot dilute ammonia solution and all alumina lakes, with the exception of erythrosine, dissolve readily in hot dilute hydrochloric acid. The resultant solutions give the identifiable reactions of aluminum with base and with alizarin. Qualitative identification of the parent dyes and quantitative analysis of the major and minor components of lakes can be carried out using similarly prescribed procedures as for the water-soluble analoges. Certain lake colors such as erythrosine may prove difficult to analyze and may require special procedures for analysis.

Table 3 Intermediate compounds commonly found in synthetic food colors

Intermediate name	*Trivial name*	*Occurrence*[a]
Aminobenzene	Aniline	R2G
1-Aminobenzenesulfonic acid	Sulfanilic acid	BPN,SY,TZ,Y 2G,BFK
4-Aminonaphthalene-1-sulfonic acid	Naphthionic acid	AM,CA,P4R,B HT
6-hydroxynaphthalene-2-sulfonic acid	Schaeffers acid	AM,P4R,SY,A R,GS
3-hydroxynaphthalene-2,7-disulfonic acid	R-Acid	AM,GS,P4R,S Y
7-Hydroxynaphthalene-1,3-disulfonic acid	G-acid	AM,P4R
7-Hydroxynaphthalene-1,3,6-trisulfonic acid		AM,P4R
4-Acetamido-5-hydroxynaphthalene-1,7-disulfonic acid	Acetyl K-Acid	BPN
8-Aminonaphthalene-2-sulfonic acid	1,7 Cleves acid	BPN
4-Amino-5-hydroxynaphthalene-1,7-disulfonic acid	K-Acid	BPN
4-Hydroxynathalene-1-sulfonic acid	N & W Acid	CA
5-Amino-4-hydroxynaphthalene-2,7-disulfonic acid	H-Acid	R2G
5-Acetamido-4-hydroxynaphthalene-2,7-disulfonic acid	Acetyl H-Acid	R2G
4,4∋-Diazoaminodi(benzenesulfonic acid)	Triazene	SY,TZ,Y2G
6,6∋-Oxydi(naphthalene-2-sulfonic acid)	DONS	SY
Tetrahydrosuccinic acid	Dioxytartaric acid	TZ
4-Hydrazinobenzenesulfonic acid		TZ
5-Oxo-1-(4-sulfophenyl)-2-pyrazoline-3-carboxyllic acid	SPCZ	TZ
2,5-Dichloro-4-(3-methyl-5-oxo-2-pyrazolin-1-yl) benzenesulfonic acid	CSPMZ	Y2G
Fluorescein		ERY
2,4,6-Triiodoresorcinol		ERY
2-(2,4-dihydroxy-3,5-diiodobenzoyl)benzoic acid		ERY
1-H-Indole-2,3-dione (and analogous sulfonic acids)	Isatin	IC
5-Sulfoanthranilic acid		IC
1-H-Indole-2,3-dioxo-1-H-indole-5-sulfonic acid	Monosulfonated indigo	IC
2-, 3-, and 4-Formylbenzenesulfonic acids		BB
N-Ethyl-N-(3-sulfobenzyl)sulfanilic acid	ESBSA	BB
m-Phenylenediamine		BFK
4-methyl-*m*-phenylenediamine		BFK
4,4∋-Bis(dimethylamino)benzhydrol alcohol		GS
4,4∋-Bis(dimethylamino)benzophenone		GS
N,N∋-Diethylaniline		BV
m-Hydroxybenzaldehyde		BV
2,4-Dihydroxybenzylalcohol		BHT
5-Amino-4-hydroxy-2-toluenesulfonic acid		AR
p-Hydroxybenzaldehyde-*o*-sulfonic acid		FG
α-(N-Ethylanilino)-*m*-toluenesulfonic acid		FG

[a]BPN, Black, PN; SY, Sunset yellow, FCF; TZ, Tartrazine; Y2G, Yellow 2G; BFK, Brown FK; AM, Amaranth; CA, Carmoisine; P4R, Ponceau 4R; BHT, Brown HT; R2G, red 2G; Ery, Erythrosine; IC, Indigo carmine; BB, brilliant blue FCF; GS, Green S; BV, patent blue V; AR, Allura red AC; FG, Fast green FCF.

Minor Components

Water-insoluble matter is calculated on the basis of 100% dye content and is usually determined by gravimetric procedures. Atomic absorption spectrophotometry has largely superseded the use of wet methods for trace heavy metals, mainly because of the speed with which analyses can be carried out and the high levels of accuracy and precision which can be attained. Other emergent techniques which have been used with limited application because of their high costs are inductively coupled plasma spectroscopy (ICPS), X-ray fluorescence (XRF, for mercury), and neutron activation analysis (NAA).

Organic Impurities

Ether-extractable matter content is usually determined by Soxhlet extraction and is applicable to all dyes except erythrosine.

Primary aromatic amines are usually determined by diazotization and coupling (to N-1-naphthylethylenediamine, NED) of an appropriate extract, followed by spectrophotometric measurement or

HPLC. HPLC has superseded classical column chromatographic procedures as the most widely used technique for the separation and quantitation of intermediate species. Both gas and thin-layer chromatography have also been used for the determination of certain intermediate compounds. The leuco base content of triarylmethane dyes is determined by carrying out further oxidation of the parent dyestuff and measuring the subsequent increase in absorbance.

Analysis of Foodstuffs for Synthetic Coloring Materials

Essentially food colors must first be extracted from the food matrix, purified to remove potentially interfering coextractives, and concentrated prior to identification and quantitation. Some form of sample pretreatment is often required, such as defatting of meat products or dilution of sugars and gums in confectionery products, before the extraction can proceed.

Leaching

Leaching may be used to remove colorants from the surface of foodstuffs such as sausages, and also from food-packaging materials. In the simplest application, the sample is soaked in an appropriate (usually alkaline) solvent which is then filtered or centrifuged to clarify the colorant solution. Further clean-up is performed as necessary. Newer techniques such as supercritical fluid extraction (SFE) may prove to be useful for the extraction of colorants, intermediates, and interaction products from foods.

Solvent-Solvent Extraction and Ion-Pair Techniques

These are widely used and effective methods of colorant isolation. Simple immiscible solvent pairs may be used where one solvent acts as a carrier for a dye-complexing reagent or soluble ion-exchange resin. The higher alcohols, particularly 1- and 2-butanol, are the most useful solvents for this technique. Amberlite LA-2, a liquid anion-exchange resin dissolved in butanol (or hexane), has been widely used as a dye-extraction medium for foodstuffs.

Quaternary ammonium compounds such as cetylcyclohexyl-dimethylammonium bromide (biocidan) and cetyltrimethyl-ammonium bromide (cetrimide) have been used for the extraction of synthetic dyes from food. More recently, reagents such as tetra-*n*-alkylammonium halides and cetylpyridinium chloride have also been employed for the rapid extraction of anionic dyes, as hydrophobic ion-pair complexes, from food using organic solvents.

Enzymatic Digestion

Pretreatment of a food sample by enzymatic digestion may be used prior to extraction of the colorants in order to release those colorants which may be highly bound or associated with the food matrix. Enzyme–substrate combinations that may be selected include papain (for protein digestion), lipase (lipids), phospholipase (phospholipid), amyloglucosidase (starch), pectinase (pectin), and cellulase (cellulose). Optimization of the pH and temperature conditions is necessary in each case.

Adsorption Techniques

Adsorption, generally referred to as solid-phase extraction (SPE), techniques have been developed for the isolation of food colorants utilizing a variety of adsorption materials such as wool, powdered leather, cellulose, alumina, and polyamide powder. More recently, semimicro adsorption cartridges containing reverse-phase bonded silica materials have found widespread use. Adsorption is achieved by either adding adsorbent directly to a pH-adjusted sample solution or by passage of the sample solution through a column packed with adsorbent. The absorbent is freed of other sample matrix components by washing with appropriate solvents and the colorants selectively desorbed using a different solvent, similar to those used in HPLC.

Dialysis Techniques

Dialysis has had limited application in the extraction and concentration of synthetic dyes from foods but has been used for the isolation and concentration of certain dyes from soft drinks and sugar-rich foods such as jellies and boiled sweets. These techniques are usually automated and employ SPE cartridges followed by HPLC.

Qualitative and Quantitative Analysis of Food Extracts

Spectrophotometry

Numerous techniques are available for the spectrophotometric analysis of colorants: measurements at ultraviolet and visible wavelengths are the easiest to perform. Beer's law can simply be applied to extracts containing single colors, whereas extracts containing two or more colors can be problematic. If the identities of the coloring components are known, their concentrations can be determined providing there is no interaction between them. The distinguishing features of the spectra obtained for single colors may be significantly affected by careful adjustment of the pH

of the solution with acid or alkali, characterized by shifts in absorption wavelength maxima and intensities. Simultaneous spectrophotometric determination of mixtures of food colorants without prior chemical separation has been applied to mixtures of up to five dyes. The data obtained are processed by chemometric approaches which utilize normal absorbance spectra as well as first- and second-order derivative spectra.

Other spectroscopic techniques, such as infrared, Raman (e.g., surface-enhanced scattering techniques), and nuclear magnetic resonance, have been used for the analysis of food colors but do not lend themselves to routine application.

Mass Spectrometry

Because of their inherent lack of volatility, direct analysis of sulfonated azo (and other) food dyes by mass spectrometry (MS) is very difficult. Several reaction schemes may be used to obtain various volatile neutral derivatives to facilitate analysis. Alternative ionization techniques other than electron impact (EI) and chemical ionization (CI) may be used for the analysis of food dyes but so far have had limited application. The most useful of these are fast-atom bombardment (FAB), field desorption (FD), and secondary-ion MS (SIMS). Liquid chromatography–mass spectroscopy techniques are beginning to find use as 'benchtop' instruments become more widespread. Atmospheric pressure chemical ionization (APCI) in positive and negative ion modes as well as pneumatically assisted electrospray have been used to determine mono- and disulfonated azo dyes.

Polarography

Differential pulsed polarography and differential pulse adsorptive stripping voltammetry may be used to estimate dye concentrations in food matrices. The addition of gelatin has been found to be advantageous in the partial identification and determination of food colors due to its pronounced affects on measured peak currents.

Electrophoresis

Of the various electrophoretic techniques available, paper electrophoresis has been the most widely used for the analysis of food colors, utilizing a range of different buffer systems and applied potentials. Cellulose acetate and polyacrylamide gel have been used for the electrophoretic separation of azo and triarylmethane dyes.

Capillary zone electrophoresis (CZE) is an analytical technique that has proven to be applicable to the analysis of food dyes, especially when coupled to photodiode array detectors. The separation of synthetic dyes is influenced by buffer composition, pH, and additives such as cyclodextrins. The related technique micellular electrokinetic capillary chromatography (MEKC) has also found application in the analysis of synthetic food dyes. Mixtures of food dyes may be separated and determined simultaneously. Buffered mobile phases modified with acetonitrile or methanol and containing other additives are often used. Both techniques may be automated readily.

Gas Chromatography

Gas chromatography cannot be used for the direct analysis of food dyes owing to their inherent lack of volatility. It is however a useful technique for the analysis of volatile derivatives and certain intermediate compounds.

Paper Chromatography

Paper chromatographic techniques have been widely used in the past for the identification of food colors but have nowadays been largely superseded by thin-layer, column, and HPLC techniques. Paper chromatography maintains popularity in some laboratories because of its relatively low cost and ease of use, and many suitable solvent systems are available.

Thin-Layer Chromatography

The use of thin-layer chromatographic systems for the separation of food dyes is fairly widespread, but is gradually being superseded by HPLC. Silica gel is the most commonly reported adsorbent used, though alumina, microcrystalline cellulose, and high-performance reverse-phase bonded silicas have widespread use. High-performance thin-layer chromatography with densitometric detection has been used for the determination of dyes in alcoholic and nonalcoholic beverages.

High-Performance Liquid Chromatography

HPLC has become the major analytical technique for the determination of synthetic coloring materials in foodstuffs. The most widely used separation modes are ion-exchange and reverse-phase. Spectrophotometric detection is applied in the visible wavelength range for dyes and subsidiary colors or in the ultraviolet range for intermediates and other organic impurities.

Ion-Exchange HPLC

Dyes, subsidiary colors, intermediates, and impurities have all been characterized using ion exchange,

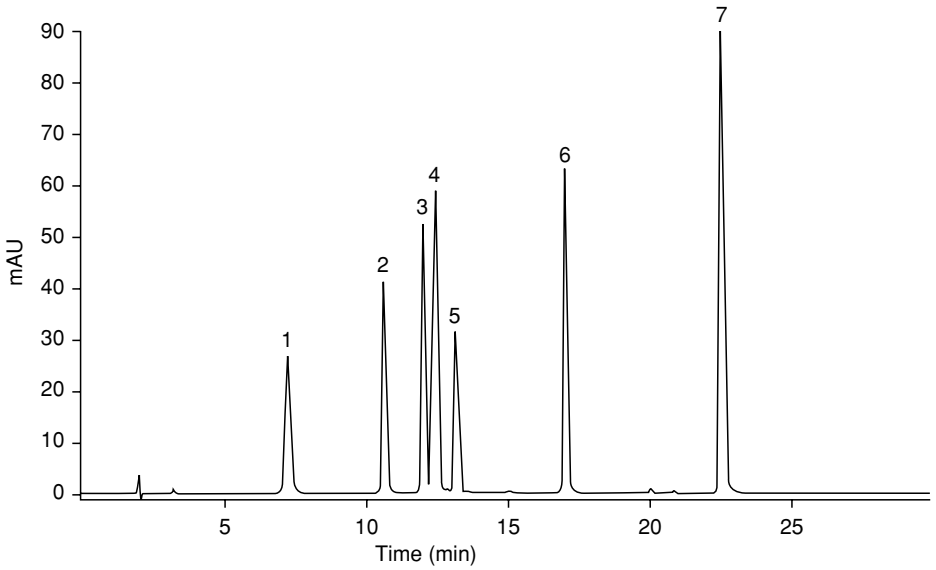

Figure 2 High-performance liquid chromatographic (HPLC) chromatogram of red dye mixture separated by reverse-phase ion-pair gradient elution. Conditions: solvent A = 0.005 mol l^{-1} tetrabutylammonium bromide (TBAB) in methanol, solvent B = 0.005 mol l^{-1} TBAB in 0.01 mol l^{-1} KH_2PO_4. Gradient elution profile: 50:50 to 100:0 (A:B) over 30 min linear, 10 min hold; solvent flow rate = 0.4 ml min^{-1}; column: ODS 20 cm; detection wavelength: 520 nm. Peak identification: 1, amaranth; 2, allura red AC; 3,red 10B – nonpermitted analog of 4, Red 2 G; 5, ponceau 4R; 6, carmoisine; 7, erythrosine. Reproduced from Colours: Properties and Determination of Synthetic Pigments, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds) 1993, Academic Press.

mostly on strong anion exchange (SAX) columns using buffered mobile phases such as borate and perchlorate mixtures and gradient elution. Azo dyes and sulfonated intermediates can be separated using weak anion exchange (WAX) columns with citric acid mobile phase at pH 2.8. Because of poor reproducibility (notably with the use of gradient elution), the requirement for aggressive buffer systems and its consequently relatively short column life, ion exchange has been largely superseded by reverse-phase HPLC for the analysis of synthetic water-soluble dyestuffs.

Reverse-Phase HPLC

Synthetic dyes require buffered eluants to achieve optimum pH conditions for desired separations on reverse-phase HPLC, usually with organic mobile phase modifiers methanol or acetonitrile. The column materials used are short-chain (C_2), octyl (C_8), and octadecyl (C_{18}) alkyl-bonded silicas, though other bonded phases such as amino ($-NH_2$) and cyano ($-CN$) have also been employed. Considerable interest has been shown in the use of macroporous and highly cross-linked polystyrene-divinylbenzene copolymer column packing materials. The mobile phase may alter the affinity of ionic species for the stationary phase by both ion suppression and ion-pairing mechanisms.

Reverse-phase-HPLC with ion suppression has been used to separate and identify various dyes, eluting with phosphate buffers at various pH values and methanol. Ammonium acetate has recently been successfully used in this mode of reverse-phase HPLC.

Ion-pair HPLC is perhaps the most widely used chromatographic technique for the analysis of dyes, subsidiary colors, and uncombined intermediates in foodstuffs. Many published methods are available. Most favored are ion-pair reagents derived from quaternary ammonium compounds such as cetyltrimethylammonium bromide (cetrimide) and tetrabutylammonium phosphate. Again, buffered mobile phases modified with organic solvents are generally used.

Both isocratic and gradient systems can be employed to separate dye mixtures; the latter is often preferred for the separation of complex mixtures. Modern high-performance and computer-aided instrumentation permits the use of many powerful techniques which may be readily applied to aid in the analysis of dyestuffs. These include:

- automated sample handling and processing
- automated chromatographic methods development
- high-sensitivity multiple-wavelength and diode-array detectors

- postrun qualitative and quantitative analysis
- liquid chromatography–mass spectrometry interfacing

Figure 2 shows a typical chromatogram of the separation of red food colors by gradient ion-pair HPLC. It is possible to separate many of the UK-permitted and several nonpermitted dyes on a single column system by employing a combination of ion-pair gradient elution and selective wavelength detection.

Diode Array Detection

Diode array detectors have the ability to record the entire spectral range of an eluting dye component during an analysis. Absorbance data may be collected simultaneously from 190 nm to as high as 800 nm and is achieved in real time. The diode array detector allows the precise determination of absorption maxima of adequately separated solutes and facilitates positive peak identification and peak purity analysis by rapid spectral scanning. Chromatographic peaks may be identified by reference to spectral libraries of previously run reference compounds. Multiple wavelength monitoring with absorbance rationing may be used to characterize and identify multicomponent dye mixtures in single chromatographic runs. Chemometric techniques similar to those used in direct spectrophotometry have been used to characterize and quantify dye mixtures.

These and other analogous methods and techniques are also used for the detection and quantification of nonpermitted dyes in foods.

See also: **Adulteration of Foods**: Detection; **Amaranth**; **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; Supercritical Fluid Chromatography; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; Properties and Determinants of Synthetic Pigments; **Food Additives**: Safety; **Food Labeling (Labelling)**: Applications; **Food Safety**; **Mass Spectrometry**: Applications; **Spectroscopy**: Visible Spectroscopy and Colorimetry

Further Reading

Damant A, Reynolds S and Macrae R (1989) The structural identification of a secondary dye produced from the reaction between sunset yellow and sodium metabisulphite. *Food Additives and Contaminants* 6: 273–282.

King RD (ed.) (1980) The determination of food colours. In: *Developments in Food Analysis Techniques* 2, pp. 79–106. Barking: Applied Science Publishers.

Knowles ME, Gilbert J and McWeeny DJ (1974) Stability of red food colours in the presence of nitrite in canned pork luncheon meat. *Journal of the Science of Food and Agriculture* 25: 1239–1248.

Marmion DM (1984) *Handbook of US Colorants for Foods, Drugs, and Cosmetics*, 2nd edn. New York: John Wiley.

Marovatsanga L and Macrea R (1987) The determination of added azo dye in soft drinks via its reduction products. *Food Chemistry* 24: 83–98.

Peters AT and Freeman HS (eds) (1995) *Analytical Chemistry of Synthetic Colorants. Advances in Colour Chemistry Series*, vol. 2. London: Blackie Academic and Professional.

Reynolds SL, Scotter MJ and Wood R (1988) Determination of synthetic colouring matter in foodstuffs – collaborative trial. *Journal of the Association of Public Analysts* 26: 7–25.

Trace Materials (Colours) Committee (1963) *Analyst* 88: 864.

Venkataraman K (ed.) (1977) *The Analytical Chemistry of Synthetic Dyes*. New York: John Wiley.

Walford J (ed.) (1980) *Developments in Food Colours 1*. Barking: Applied Science Publishers.

Walford J (ed.) (1984) *Developments in Food Colours 2*. Barking: Elsevier Applied Science Publishers.

Colorimetry *See* **Spectroscopy**: Overview; Infrared and Raman; Near-infrared; Fluorescence; Atomic Emission and Absorption; Nuclear Magnetic Resonance; Visible Spectroscopy and Colorimetry

Common Agricultural Policy *See* **European Union**: European Food Law Harmonization

COMMUNITY NUTRITION

J Cade, University of Leeds, Leeds, UK

Introduction

Community nutrition incorporates the study of nutrition and the promotion of good health through food and nutrient intake in populations. This article will consider aspects of community nutrition relating to dietary goals and recommendations for populations; methods of assessing diet in population groups; and promoting healthy eating at the community level.

Community nutrition (public health nutrition) requires a population approach. The community rather than the individual is the focus of interest. This area of nutrition focuses on the promotion of good health and the primary prevention of diet-related illness. The emphasis is on maintenance of health in the whole population, although it will also include working with high-risk groups and other subgroups within the population. Community nutrition includes nutritional surveillance; epidemiological studies of diet; and also the development, implementation, and evaluation of dietary recommendations and goals. A community may be any group of individuals, for example, the population of a town or country, or the residents of an old people's home.

Advice to the Community

Dietary advice to the community has changed over the years, depending on the nutrition-related diseases of importance at the time and our understanding of how they are caused. In the 1930s, for example, the primary concern in Europe and North America was the elimination of deficiency diseases. The concept of a balanced diet was developed to try to provide the minimum requirements of protein, vitamins, and minerals. By the late 1950s, research suggested that some chronic diseases could be related to overnutrition and dieting to lose weight became popular. For example, it was recommended that bread and potato intake should be restricted to help avoid overweight. However, by the 1980s, wholemeal bread and jacket potatoes were being promoted as good sources of dietary fiber.

Recommended dietary allowances (RDAs) for populations were first set in the 1930s and have been revised at regular intervals. The current UK recommendations were published in 1991, and are known as dietary reference values (DRV). They include the reference nutrient intake (RNI) level which is the amount of a nutrient (excluding energy) which is sufficient for almost any individual. The RNI is 2 standard deviations higher than the estimated average requirement and as such is higher than most people need. RNI must be interpreted carefully when used in nutrition education and is not intended to be used by individual members of the public as a guide.

Dietary goals or guidelines, on the other hand, aim to reduce the chances of developing chronic degenerative diseases and are based on data from animal experiments, metabolic studies, clinical trials, and epidemiological research. They give targets for the population to aim at for some future time. The first set of dietary goals was published in Sweden in 1968. The UK has published its own goals. Nutrition goals were included in the Health of the Nation White Paper (1992). Further goals were included in the Department of Health's report in 1994 from the Committee on Medical Aspects of Food Policy (COMA) on diet and cardiovascular disease. The World Health Organization (WHO) has also published population nutrient goals in 1990. These are recommended for use in all parts of the world and have been expressed in absolute terms rather than increases or decreases in existing nutrient intake. The desirable change will then vary with the population. For example, in some developing countries the goal for fat intake (lower limit) suggests the need to increase average intakes slightly. Conversely, for most industrialized countries a reduced fat intake is desirable. Table 1 summarizes the COMA, Health of the Nation, and WHO nutrient goals.

The recommendations are based on the best available evidence and may need to be changed in future. For example, in the future they may need to address optimal nutritional status.

Assessing Adequate Nutritional Status

In order to advise the community we need to know what the community is already eating. The nutritional status of the community can be defined as the presence or absence of diet-related diseases and is related to the health and well-being of the community. There is no simple or single way to measure it so that information has to be gathered from many sources.

Food and Nutrient Supply

National level Information on national food availability is collected annually by many governments in the form of food balance sheets. The Food and Agriculture Organization (FAO) collects and publishes

Table 1 Comparison of dietary goals from the World Health Organization (WHO), Committee on Medical Aspects of Food Policy (COMA), and Department of Health, Health of the Nation

	WHO		*COMA*	*Health of the Nation goals by 2005*	*Dietary reference value*[a]
	Lower	*Upper*			
Total energy (MJ)	[b]		Maintain desirable body weight	↓ Obesity –	
Males				to no more than 6% men	10.6[c]
Females				to no more than 8% women	8.1
Total fat	15	30	↓ to about 35%	↓by 12%, to no more than 35%	33 (35)[d]
Saturated	0	10	↓to no more than 10%	↓ by 35%, to no more than 11%	10 (11)[d]
Polyunsaturated	3	7	*n*-6 PUFA: no further ↑ *n*-3 PUFA: ↑ from 0.1 to 0.2 g day^{-1}		6 (6.5)[d]
Cholesterol (mg day^{-1})	0	300	No increase		
Total carbohydrate (CHO)	55	75	50		47 (50)[d]
Complex CHO	50	70	Increase		37 (39)[de]
Free sugars	0	10	Increase from fruit and vegetables		10[f](11)[d]
Dietary fiber (g day^{-1})	27	40			18[g]
Protein	10	15			
Salt (g day^{-1})		6	6		1.6
Alcohol					

Units are percent total energy unless otherwise stated.
[a]Values are population averages.
[b]Energy intake to allow normal growth, pregnancy, lactation, work, activities and to maintain appropriate body reserves. Body mass index (kg m^{-2}) in adults = 20–22.
[c]Estimated average requirement for men and women, aged 19–49 years, physical activity level 1.4.
[d]Figures in brackets represent percentage of food energy.
[e]Intrinsic and milk sugars and starch.
[f]Nonstarch polysaccharides.
[g]Nonmilk extrinsic sugars.
PUFA, polyunsaturated fatty acids.

these statistics. It lists the total quantity of different foods available for consumption in a country during a specified time (usually one year). Data are limited since there is no information available about subgroups of the population by region, ethnicity, age, or socioeconomic level. Its accuracy is variable, especially for areas with subsistence farming.

Food balance sheets can be used to assess available energy and nutrient intakes per capita. Comparing changes over years can indicate trends towards or away from national food security.

Household level The National Food Survey carried out by the UK Ministry of Agriculture, Fisheries and Food (MAFF) assesses trends in food supplies and, by inference, the food eaten. Nutrient intake can be calculated from these data. It records food purchased rather than food eaten and does not distinguish between the intakes of men, women, and children, nor does it include alcohol, candies (sweets), and soft drinks.

As shown in Table 2, the National Food Survey is particularly useful for assessing trends in food and nutrient intake. In general, the food supply was consistent with the amounts recommended to cover the needs of most of the population.

Itemized till receipts have also been shown to be a good reflector of energy and fat intake in the household in families which purchase the majority of their food at supermarkets.

Individual level Assessing diet at the individual level is usually done by survey. These use several different methods such as a weighed intake, dietary recall or record, or a food frequency questionnaire. Attempts have been made to validate the resulting nutrient levels by measuring biological markers. Surveys may look at a representative sample of the population or a particular at-risk group. The UK government has commissioned a series of regular surveys known as the National Diet and Nutrition Surveys. The first one of British adults was published in 1990. It aimed to recruit a nationally representative sample of adults aged 16–64 living in private households in the UK to inform food and health policy development and evaluation. Subsequent surveys have studied preschool children, the elderly, and young people. A second national survey of adults is currently underway.

Anthropometric Measurements

Weight and height can be used to assess the level of malnutrition or obesity by comparison with reference data. Height can give evidence of past chronic malnutrition, whereas weight is more useful in assessing recent nutritional experience. Overweight or obesity

Table 2 Nutrient intake from the UK National Food Survey, expressed as a percentage of recommended intakes current at the time of the survey

	1958[a]	*1968*[b]	*1978*[b]	*1988*[c]	*1998*[d]
Energy (kcal)	104	108	94	91	93
Protein (g)	100	127	121	123	147
Calcium (mg)	107	191	181	159	119
Iron (mg)	115	122	100	102	99
Thiamin (mg)	126	133	125	153	98
Riboflavin (mg)	108	129	138	123	149
Vitamin C (mg)	222	181	188	213	173

Recommended intakes:
[a]British Medical Association (1950) *Report of the Committee on Nutrition*. London: BMA.
[b]Department of Health and Social Security (1969) *Recommended Intakes of Nutrients for the United Kingdom*. London: HMSO.
[c]Department of Health and Social Security (1979) *Recommended Intakes of Nutrients for the United Kingdom*. London: HMSO.
[d]Department of Health (1991) *Dietary Reference Values for Food Energy and Nutrients for the United Kingdom: Report of the Panel on Dietary Reference Values, Committee on Medical Aspects of Food Policy. Report on Health and Social Subjects 41*. London: HMSO. Nutrient values from Ministry of Agriculture, Fisheries and Food National Food Survey for years stated.

is most practically assessed using the body mass index (weight in kg/height in m^2).

Health Statistics

Birth statistics such as neonatal, perinatal, or infant mortality rates provide indirect information about the nutritional status of a community, particularly its disadvantaged groups. Death rates for diseases which have a nutritional cause can also indicate nutritional status in the community.

Qualitative Data

Useful information about community food and nutrient patterns can be obtained using qualitative methodologies, such as informal interviews or focus group discussions. If carefully carried out, these methods can supplement data collected in the quantitative methods described above and provide an indepth understanding of issues such as motivations, food choices, and particular food habits.

Success Level of General Advice

In order to change a population's diet, DRVs and goals must have scientific credibility, political and technical support, and be recognized as being necessary and acceptable to the consumer. It may take years to achieve the desirable change.

Is the population meeting the DRVs for the UK population? The results of three National Diet and Nutrition Studies (NDNS) in the British population showing the percentage who met the UK dietary goals are presented in Table 3. Most subjects ate more total fat, saturated fat, and refined sugars and less carbohydrate and fiber than recommended. A comparison with a subgroup of 15 000 women from the UK Women's Cohort Study shows that in this health-conscious group more people were able to achieve the dietary goals than in the earlier NDNS study of adult women.

People who consume diets which meet the nutrient goals are more likely to eat cereals, wholemeal and brown bread, skimmed or semiskimmed milk, polyunsaturated margarine, fruit, vegetables including potatoes, low-fat meat, and nonfried fish. They also eat less white bread, butter, margarine, whole milk, high-fat cheese, eggs, fatty meat, and fried fish than those who do not meet the goals. To achieve a particularly healthy diet independent predictive factors have been found to be spending more money on food, being a vegetarian, having a higher energy intake, and a lower body mass index, and being older. Extra costs of the food may make the cost of a diet which meets the dietary goals too expensive for the elderly, unemployed, and low-paid. It is worth noting, however, that it is possible to consume a diet which meets the dietary goals which is substantially cheaper than the average cost of a diet which does not meet the goals.

The National Food Survey can monitor progress made by a population towards meeting the goals. It has been analyzed for a 50-year period from 1940 to 1992. There has been a decline in the percentage of food energy from carbohydrate and an increase in the proportion from fat (Figure 1). There have been substantial changes in the types and quantities of foods consumed over this period. Bread consumption has fallen over this period from around 250 g day^{-1} in 1940 to 110 g day^{-1} in 1992 (Figure 2). Fresh potato consumption has also declined, but fruit and other vegetable consumption has increased from about 200 g day^{-1} to 300 g day^{-1} (Figure 3).

Community versus Individual Advice

Public health examines risk factors for disease in the population as a whole and then designs prevention strategies to reduce them. Meanwhile, the clinician does the same for the individual patient.

Dietary goals for the population are based on identifying population intakes to maintain health. Health is defined as a low rate of diet-related diseases. In assessing whether a population is meeting the dietary goal it is the entire range of nutrient intake which matters. If the intake is normally distributed within the population, then it can be summarized by the average intake of the population and its standard error.

Dietary goals require a population approach to dietary change leading to a change in the average intake. This will result in some individuals consuming more and some less than the stated goal.

Table 3 Percentage of population meeting dietary goals

Dietary goal	*NDNS UK adults (1986–87)*[a]		*NDNS 1½–4½ years (1992–93)*[b]		*NDNS 65 years + free-living (1994–95)*		*UK Women's Cohort (1995)*[d] *Health-conscious women*
	Men	*Women*	*Boys*	*Girls*	*Men*	*Women*	
Total fat $\leq$ 35% food energy	12	15	42[e]	42[e]	46	41	
Saturated fat $\leq$ 15% food energy	11	12	38[f]	38[f]	59	48	43[g]
Total carbohydrate > 50% energy	9	12	60[h]	57[h]	31	30	54[i]
Refined sugars[k] < 10% food energy			12	13	30	36	
Fiber > 25 g day^{-1}	45	16			3[l]	1[l]	32[j]

Note that the goals are not meant to refer to children – the preschool age children survey has been included for comparison.
[a]Gregory J, Foster K, Tyler H and Wiseman M. (1990) *The Dietary and Nutritional Survey of British Adults*. London: HMSO.
[b]Gregory JR, Collins DL, Davies PSW, Hughes JM and Clarke PC (1995) *National Diet and Nutrition Survey: Children aged 1½ to 4½ years*. London: HMSO.
[c]Finch S, Doyle W, Lowe C *et al*. (1998) *National Diet and Nutrition Survey: People Aged 65 Years and Over*, vol. 1. *Report of the Diet and Nutrition Survey*. London: Stationery Office.
[d]Cade J, Upmeier H, Calvert C and Greenwood D (1999) Costs of a healthy diet: analysis from the UK Women's Cohort Study. Public Health Nutrition 2: 505–512.
[e]Less than 35% food energy from fat.
[f]Less than 15% food energy from saturated fat.
[g]0–10% total energy from saturated fatty acids.
[h]Greater than or equal to 50% food energy from carbohydrate.
[i]50–70% total carbohydrate.
[j]27–40 g fiber.
[k]Nonmilk extrinsic sugars.
[l]$\geq$ 20 g day^{-1} fiber.
NDNS, National Diet and Nutrition Studies.

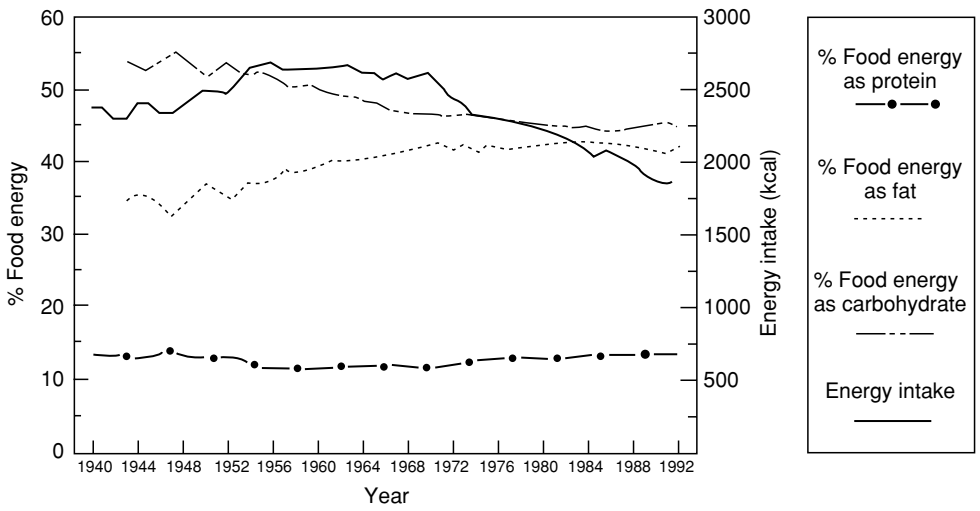

Figure 1 Trends in proportion of daily energy intake in the British diet derived from carbohydrate, fat, and protein (National Food Survey 1940–1992). From Department of Health (1994) *Nutritional Aspects of Cardiovascular Disease. Report on Health and Social Subjects 46*. London: HMSO with permission.

Methods of Giving Advice to the Community

Much can be done to change a population's diet. Diets are chosen by individuals; government should not enforce recommendations concerning diet and health by restrictive legislation. Legislation concerning the production and sale of food does however affect a nation's diet. An alternative approach to change is to increase people's knowledge and awareness of food and its relationship to health.

Health Promotion

Traditionally, nutrition education campaigns have involved only one or two sections of the community such as schools or the media. Success, if it was measured at all, was limited. A more integrated approach is to involve as many groups of the community

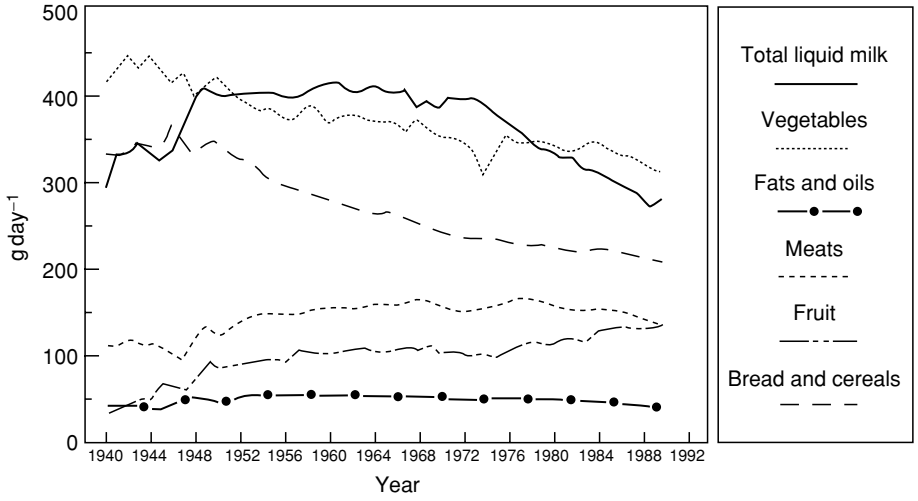

Figure 2 Consumption of selected foods. From Department of Health (1994) *Nutritional Aspects of Cardiovascular Disease. Report on Health and Social Subjects 46*. London: HMSO with permission.

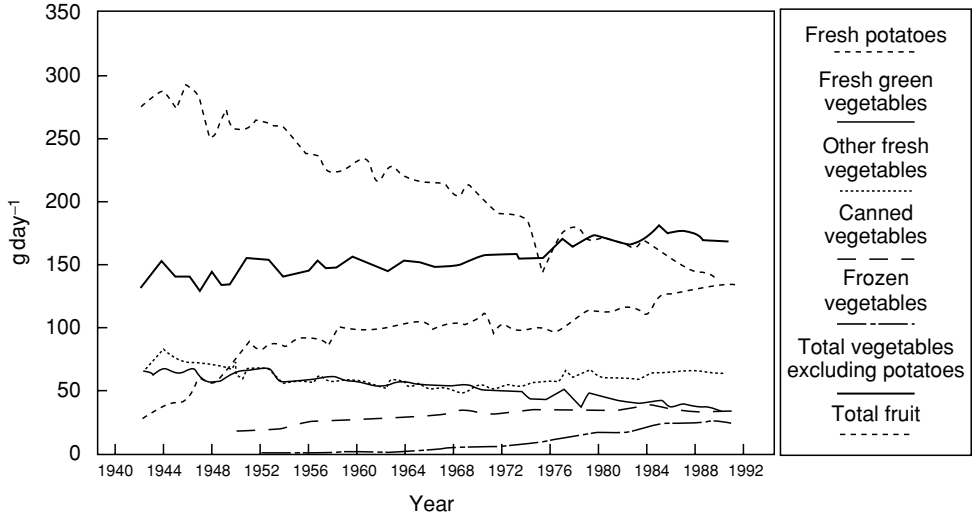

Figure 3 Trends in consumption of vegetables and fruit. From Department of Health (1994) *Nutritional Aspects of Cardiovascular Disease. Report on Health and Social Subjects 46*. London: HMSO with permission.

as possible and to use a variety of educational approaches. For example, Heart-Beat Wales is a health promotion program which involves public education, lay groups, schools, factories, retailers, and others. Look After Your Heart is a multidisciplinary government-backed campaign to reduce risk factors for coronary heart disease. Evaluating the success of these campaigns in terms of reduction of deaths from heart disease is difficult. With regard to lowering fat intake, general recommendations to the public without individual dietary counseling are not very effective. Despite educational campaigns to lower fat intake, the level of saturated fat intake in Europe remains considerably higher than desired.

Knowledge of the health risks of a food does not inevitably lead to a change in its consumption. Why people eat what they do is determined by a number of factors; the health effects of a food may only have a minor influence. Understanding these factors involves the application of psychological models such as the Theory of Planned Behavior or the Stages of Change model. These models explore issues such as beliefs about food and motivations for changing intakes. Health education can change consumers' purchasing and eating habits. For example, the national promotion of low-fat milk immediately increased its consumption by 20%. Leaflets are often used, but are ineffective with mass distribution.

The media can have an important influence on choice of foods consumed. This has been demonstrated recently by the concerns over new-variant Creutzfelt–Jakob syndrome from infected beef and unknown risks associated with genetically modified foods. Media reports have raised awareness of these issues and people have altered their food intake patterns, although there is no evidence to suggest that these changes have been detrimental to overall nutritional status.

Health Professionals

General practitioners are encouraged to provide health promotion, but have little training in nutrition. Patients are rarely referred to a dietitian; the practice nurse often provides most of the advice. Dietary advice provided by untrained members of the primary health care team can be misleading. Community dietitians can provide group-based training and education. The UK Nutrition Society is now recognizing a group of nutritionists with experience in community nutrition and awarding an accreditation in public health nutrition.

Most district health authorities have their own food and health policies which are often limited to measures within their own services. Potential contributors to these policies are groups within the community such as local shopkeepers, industry, and trade unions. Health action zones have been set up to address inequalities in health in the community. Some have incorporated nutrition and access to food as priorities.

Food Labeling

Food should be appropriately and comprehensibly labeled to aid consumer choice. As yet, there is no standard format for labels. The format for nutrition labeling of foods and food claims is regulated by Codex Alimentarius, or the food code. This is a global reference point and deals with creating food standards. Nutrition labeling is often difficult to understand. Supplying the consumer with nutrition information through the internet or in-store computer terminals as well as highlighting the most important information may solve the problem of demands for both more and understandable information.

Web-based Resources

Knowledge of nutrition and health can be acquired from the internet. Recent figures suggest that around 16 million people have access to the internet in the UK. This figure is increasing. It is important for sites to be rated in terms of content, credibility, and usability. A number of sites are on the fringe of conventional nutrition and may be promoting information which is unhelpful or confusing. NHS Direct has been set up to provide telephone or web-based health advice and includes a nutrition component.

Reaching Vulnerable Groups

Nutrition advice should take account of the special needs of particular subgroups, who may be vulnerable to malnutrition due to their physiological requirements, their cultural traditions, or their economic problems.

Babies and Young Children

The health professions have a role in promoting breast-feeding and the preparation of culturally acceptable and appropriate weaning foods. In the UK, in order to meet the current dietary goals the diet is likely to increase in bulk and so reduce energy density. It is important to maintain adequate energy intakes by not changing babies on to low-fat milks too early. Children in lone-parent families tend to have lower carotene and vitamin C intakes than other children. Growth monitoring, particularly in developing countries, can be used to detect growth faltering. Remedial action can then be taken.

Adolescents

On the whole this group is adequately nourished. There is some concern, however, at the low levels of iron intake in teenage girls. Adolescents on slimming diets may be at risk of low intakes of some micronutrients. Appropriate nutrition education at school, backed up with the provision of healthy school meals, may help to encourage suitable nutritional habits in this group.

Pregnancy and Lactation

Most pregnant and lactating women in the UK consume an adequate diet. This is, however, a time when women come into contact with health care services and may be open to nutrition education. Young girls and other women at risk of having a low-birth-weight baby should have particular attention paid to their dietary intake. Periconceptual folate supplements are recommended to prevent neural tube defects.

Elderly

The current dietary goals are appropriate for the elderly. Energy intakes tend to decrease with increasing age and there is a need for the elderly to consume diets with an increased nutrient density. Health professionals and other carers need to be aware of the potential for the elderly to be consuming an inadequate diet. The NDNS for people aged 65 and over found that intakes of vitamin D, Mg, K, and Cu

were low. A substantial proportion of people living in institutions had low biochemical status indices for vitamin C, Fe, and folate.

Low-income Groups

Families on a low income and those living in temporary accommodation may find it difficult to eat a diet which meets the recommendations. There is particular concern for people living in socially disadvantaged areas since there is a greater risk of birth abnormalities and a higher incidence of diseases such as coronary heart disease and cancer in the adult population.

Ethnic Groups

Many aspects of different ethnic dietary patterns will promote health, such as high fruit or vegetable intakes. Some ethnic groups, however, also consume diets which are characterized by a high fat intake and certain groups have a higher prevalence of obesity, diabetes, and coronary heart disease. In some communities women and children may be at risk of vitamin D deficiency. Culturally acceptable nutrition education needs to be in a form that can be understood by the whole community. Personal intervention by a trained member of the local community may give the best results.

See also: **Adolescents**; **Dietary Reference Values**; **Elderly**: Nutritional Status; **Epidemiology**; **Ethnic Foods**; **Food Labeling (Labelling)**: Applications; **Infants**: Nutritional Requirements; **Lactation**: Human Milk: Composition and Nutritional Value; **Pregnancy**: Metabolic Adaptations and Nutritional Requirements

Further Reading

Committee on Medical Aspects of Food Policy (1994) *Nutritional Aspects of Cardiovascular Disease. Report on Health and Social Subjects* 46. London: HMSO.

Department of Health (1991) *Dietary Reference Values for Food Energy and Nutrients for the United Kingdom: Report of the Panel on Dietary Reference Values, Committee on Medical Aspects of Food Policy. Report on Health and Social Subjects* 41. London: HMSO.

Department of Health (1992) *The Health of the Nation: A Strategy for Health in England*. London: HMSO.

Finch S, Doyle W, Lowe C *et al.* (1998) *National Diet and Nutrition Survey: People aged 65 years and over*, vol. 1. *Report of the Diet and Nutrition Survey.* London: Stationery Office.

Gregory J, Foster K, Tyler H and Wiseman M (1990) *The Dietary and Nutritional Survey of British Adults*. London: HMSO.

Gregory JR, Collins DL, Davies PSW, Hughes JM and Clarke PC (1995) *National Diet and Nutrition Survey: Children aged 1½ to 4½ years*. London: HMSO.

Ministry of Agriculture, Fisheries and Food (1999) *National Food Survey 1998*. London: Stationery Office.

World Health Organization (1990) *Diet, Nutrition and the Prevention of Chronic Diseases. Technical Report Series 797*. Geneva: World Health Organization.

CONDENSED MILK

H J Hess, Nestec Ltd, Vevey, Switzerland

Background

Industrial production for both sweetened and unsweetened condensed milk started around the middle of the nineteenth century. However at the beginning of the nineteenth century, food scientists in Europe, mainly in France and England, and in the USA had been working on the possibility of preserving milk as a concentrated liquid. Both products have their industrial roots in the USA.

The first sweetened condensed milk in hermetically sealed cans was manufactured and sold in 1856 by Gail Borden in the USA. The business grew rapidly and spread to Europe by 1866 where a sweetened condensed milk factory was set up in Cham, Switzerland, by Charles A. Page, a US consul assigned to the country. The rapid success of this Swiss-based company within Europe led eventually to an expansion of its manufacturing facilities in the USA. This organization, registered as the Anglo-Swiss Condensed Milk Company, sold its US interest in 1902 to a company established and registered at the end of the nineteenth century as the Borden Condensed Milk Company. In 1904, the remaining European interests of the Anglo-Swiss Company merged with Henry Nestlé, of Vevey, Switzerland, who also manufactured this product. The new company was called the Nestlé-Anglo-Swiss Condensed Milk Company.

Until the early 1880s, unsweetened condensed milk was also produced and was sold open in the market

due to the lack of knowledge and success of long-life preservation at that time.

The basic process for preservation of unsweetened condensed milk by heat sterilization was conceived by John B. Meyenberg in 1882, a Swiss citizen, and an employee of the Anglo-Swiss Condensed Milk Company. The idea of preserving milk without the addition of sugar was made possible by his invention of a revolving sterilizer working with steam under pressure. Lacking sufficient support from his company to continue his work, he migrated, in 1884, to the USA and also obtained a patent for his invention in that country. In 1885, Mr Meyenberg was cofounder of the Helvetia Milk Condensing Company in the State of Illinois, and during the same year, he achieved the first successful manufacture of unsweetened condensed milk. The name of this product was changed to evaporated milk for a clearer distinction from sweetened condensed milk, a situation that prevails today.

The initial phase of commercial production of both products was hampered by various problems and quality defects. Only at the beginning of the twentieth century were constant quality standards reached.

Definition of Products

Sweetened Condensed Milk

Sweetened condensed milk is made by the addition of sugar to whole milk and the removal of water from the milk to about one-half of its original volume. The product is canned or packaged in other containers without sterilization, with the sugar acting as a preservative. International standards prescribe:

- a minimum milk fat content of 8%;
- a minimum milk solids content of 28%.

The minimum sugar content is often not specified precisely but should be sufficient to avoid spoilage.

Permitted stabilizers are usually specified as sodium, potassium, and calcium salts of:

- hydrochloric acid;
- citric acid;
- carbonic acid;
- orthophosphoric acid;
- polyphosphoric acid.

The name of the product may be:

- sweetened condensed milk;
- sweetened condensed whole milk; or
- sweetened full-cream condensed milk.

Legislation in some countries requires a somewhat higher milk solids and fat content, usually 9% milk fat and 31% total milk solids. However, there are also provisions for skimmed sweetened condensed milk with a milk solids content up to 24% and low-fat compositions, in principle, of 4% fat and 24% total milk solids.

Frequently, this product is fortified by the addition of vitamins, mainly A, D_3, and B_1. (*See* **Food Fortification**.)

Evaporated Milk

Evaporated milk is made by removal or evaporation of water from milk but without the addition of sugar or any other preservative material. The canned product is heat-sterilized at 118–122 °C for several minutes. The product also may be packed in any other sterilizable container.

International standards prescribe:

- a minimum milk fat content of 7.5%;
- a minimum milk solids content of 25.0%.

Permitted stabilizers are usually specified as sodium, potassium, and calcium salts of:

- hydrochloric acid;
- citric acid;
- carbonic acid;
- orthophosphoric acid;
- polyphosphoric acid.

In addition, some legislation permits the addition of carrageenan up to 150 p.p.m.

The main name of the product is:

- evaporated milk;
- evaporated full-cream milk; or
- unsweetened condensed full-cream milk.

Legislation in some countries requires a somewhat higher fat and milk solids content, up to 9 and 31%, respectively. However, there are also provisions for either skimmed or low-fat evaporated milks:

- skimmed evaporated milk with a minimum of 20% milk solids;
- low-fat evaporated milks, 4 or even 2% milk fat and a milk solids content of 20–24%.

Fortification with vitamins of either or both A or D_3 is common practice.

Recombined Milk Products

Sweetened condensed milk and evaporated milk are often recombined in countries outside the traditional dairy belt. (*See* **Recombined and Filled Milks**.)

For recombining, imported skim milk powder and anhydrous milk fat are used to make up the milk

components. In recent years, the production of filled milks has become increasingly important. In this case, the milk fat is replaced by locally available and cheaper vegetable fats.

Storage and Packing

An adequate keeping quality of up to 12 months can be achieved for both sweetened condensed milk and evaporated milk, but this largely depends on the storage conditions. At a storage temperature above 25 °C, the aging process and related physical, as well as organoleptic, defects appear more rapidly. For recombined products, the normal shelf-life stability, without major deviations from the original aspect, is reduced by about half of a fresh-milk product. (*See* **Storage Stability**: Parameters Affecting Storage Stability.)

The most commonly used packing is the tin plate can, which offers absolute protection against light, is basically crash proof, and has advantages for handling and storage. Recycling is also an important factor. As alternative packing, glass containers or paper/plastic laminates can be used; however, depending on the packing material, a considerable shelf-life reduction should be taken into account. For paper/plastic laminates, aseptic filling techniques are used, especially for evaporated milk. (*See* **Canning**: Principles.)

Nutritional Considerations

Owing to the concentrated form, both sweetened condensed milk and evaporated milk have an increased compositional analysis compared with milk (Table 1). For sweetened condensed milk, the figure for carbohydrate includes the sucrose content necessary for product conservation; the carbohydrate in evaporated milk and unconcentrated milk is lactose only. The rather high energy value of sweetened condensed milk is due mainly to the amount of carbohydrates present, whereas the protein and fat contents compare more or less with evaporated milk. (*See* **Milk**: Dietary Importance.)

Production and Usage

World production amounts to over 4.5 million tonnes. About one-third is produced as sweetened condensed milk and two-thirds as evaporated milk. As recently as 35–40 years ago, infant feeding was still a major application for either sweetened or unsweetened concentrated milk. Owing to the nature of the products, somewhat distinct usages are common practice.

Sweetened Condensed Milk

Owing to the high sucrose content of > 40% and its viscous consistency, this product is frequently used as a jam-like bread spread.

An early, but no longer popular, use of sweetened condensed milk was to dilute it with water and consume it as a drink. Presently, coffee or tea whitening and sweetening is the major use of sweetened condensed milk. In many countries, it is also used in combination with cocoa or other milk modifiers in the preparation of homemade drinks.

A growing application is in sweet dessert preparations like icecreams, cakes, cookies, etc. Specific to the Latin American continent is its use for Dulce de Leche, basically a caramelized sweetened condensed milk obtained by boiling the can in water for 2–3 h. Different applications vary from country to country and from one geographical region to another, according to traditional consumption habits.

Evaporated Milk

A large area of application is coffee and tea whitening and the preparation of milk-based beverages. The other main field for general usage is in the culinary sector to enhance the taste and texture of mashed potatoes, pasta, quiches, soups, and a wide variety of savory and sweet recipes. The main uses can also vary specifically from one country to another.

Manufacturing Principles

The raw material used for both products is usually cows' milk, although in certain regions, it may be a mixture of cows' and buffalos' milk. (*See* **Buffalo**: Milk.)

Sweetened Condensed Milk

A flow diagram for the manufacture of sweetened condensed milk is shown in Figure 1.

Table 1 Typical comparative figures for standard concentrated products and milk (all values in grams per 100 g)

Product	*Protein*	*Fat*	*Carbohydrates*	*Minerals (ash)*	*Calcium*	*Phosphorus*
Sweetened condensed milk	7.8	8.0	55.2	1.8	0.28	0.23
Evaporated milk	6.5	7.5	9.8	1.4	0.24	0.19
Milk	3.2	3.5	4.6	0.7	0.12	0.09

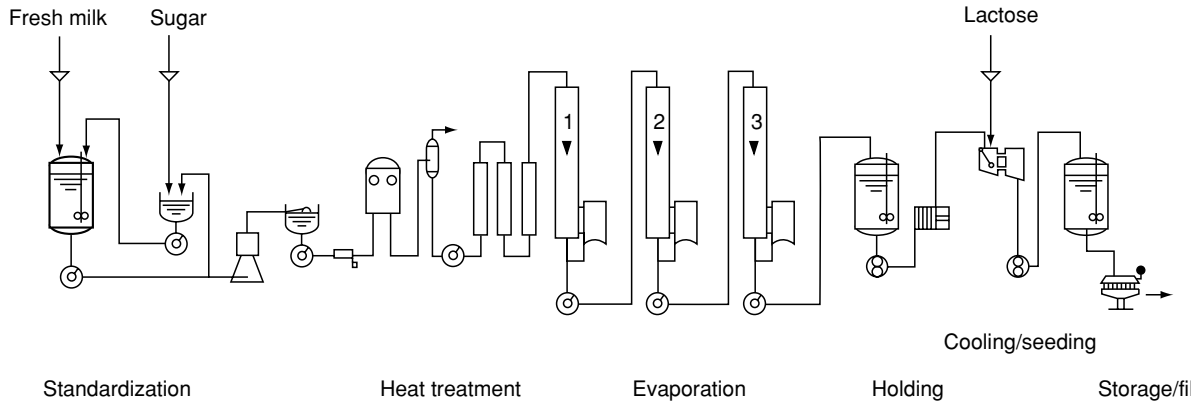

Figure 1 Basic flow diagram for the manufacture of sweetened condensed milk (continuous process). Reproduced from Condensed Milk, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Raw milk is collected and selected according to the usual quality criteria. The milk is analyzed for fat and solids-nonfat (SNF) content and the ratio between both components is adjusted according to the fat/SNF ratio of the final product to be manufactured by the addition of either cream or skim milk.

If the final product should have a composition of 8.0% fat and 21% SNF, the raw milk should be adjusted to the same fat/SNF ratio (see Table 2).

Sugar is dissolved in the cold milk, in principle by recirculation through a dissolving vat. The amount of sugar is determined by the quantity of milk prepared for a batch standardization, its fat content, and the compositional requirement of the sucrose content of the final product, e.g.,

$$\frac{10\,000 \text{ kg of milk} \times 3.25\% \text{ fat}}{100} \times \text{sugar factor}\left(\frac{\text{sugar content of finished product}}{\text{Fat}}\right).$$

After standardization, the milk/sugar mixture undergoes a heat treatment, which consists of a temperature/time combination usually above 100 °C and holding for seconds or minutes. This heat treatment eliminates nearly all the bacteriological flora of the milk, but the choice of temperature/time combination greatly influences final product characteristics such as the consistency and shelf-life stability. (*See* **Heat Treatment:** Chemical and Microbiological Changes.)

The viscosity of the product is an important factor. It is rather difficult to indicate precise heat treatment conditions, since they depend on the equipment used and the characteristics of the raw material milk which can produce different behavior through regional, climatic, and seasonal influences. Constant observation of viscosity and age thickening is necessary to counteract any negative effect. In general, lower heat treatments favor increased viscosities, and higher, intensive heating tends to give lower viscosities. The viscosity should be sufficient to avoid phase separation during storage, and age thickening should be moderate in order to keep the product pourable during its storage life.

Table 2 Example adjustment of the fat and SNF content of raw milk to achieve a final product composition of 8.0% fat and 21% SNF

	Final product	*Ratio*	*Standardized milk*	*Ratio*
Fat (%)	8.0	0.381	3.25	0.381
SNF (%)	21.0		8.52	

Concentration after heat treatment takes place under vacuum and usually in a multiple-effect evaporator.

Sweetened condensed milk is not a sterilized or sterile product; its conservation is assured by the bacteriostatic effect of the high sucrose concentration. The sugar-in-water ratio in the concentrated product must be above 61%. However, osmophilic organisms can develop in this medium, so it is important that the equipment after heat treatment and up to filling is of a hygienic design and is operated under the best hygienic conditions to avoid reinfection. Strict controls for equipment cleaning and sterilization are essential.

After evaporation, the concentration must be as close as possible to the final product's total solids content; an adjustment by, for example, water should be omitted because of the risk of contamination.

After evaporation, the product is cooled to normally 20–25 °C, and at the final concentration, part of the lactose contained in the product is

oversaturated. To avoid autocrystallization, fine-milled and pasteurized dry lactose crystals are added to the concentrate to initiate instant and controlled crystallization. During storage, product stirring should be maintained for several hours to finalize the crystallization process.

Filling, usually into metallic tin plate cans, is the last delicate operation. Cans and lids must be sterilized, and this is often achieved by passing them through a flaming installation. Air in the filling area must be filtered and of an excellent bacteriological quality; the air space in the can must be as low as possible to restrict mold growth. It is not necessary to perform an aseptic operation, but excellent hygienic conditions must prevail.

Evaporated Milk

A flow diagram for the manufacture of evaporated milk is given in Figure 2.

In principle, this product is poststerilized, which is a conventional process, whereby after filling and closing operations, the product is sterilized in the final container (e.g., tin plate cans, glass bottles, etc.).

For stability reasons, the product contains a small percentage of stabilizing salts (either one or a combination of the earlier-mentioned salts). Where legislation allows, the hydrocolloid, carrageenan, may be added. Stabilizing salts are indispensable as manufacturing and technological aids for adjusting or regulating the heat stability of the product during the poststerilization process. Carrageenan influences phase stability (creaming and protein sedimentation) during the product's storage life.

During the last 20 years, aseptically filled evaporated milks, mainly in soft packs, have been developed and commercialized. To realize aseptic filling, the product has to be heat-treated in-line for final sterilization. Ultrahigh temperature (UHT)-type installations are used for this purpose. (*See* **Heat Treatment: Ultra-high Temperature (UHT) Treatments.**)

The processing steps for milk standardization, heat treatment, and concentration/evaporation are similar or identical to those described for sweetened condensed milk, with the exception of the addition of sugar.

After concentration, evaporated milk is homogenized, cooled for intermediate storage, and sterilized after filling.

Heat treatment/heat stability The ability of concentrated milk to withstand high-temperature sterilization is essential. This is commonly described as heat stability and refers to the resistance of milk concentrate to coagulation during sterilization in containers.

Milk has a natural heat stability, which is influenced mainly by compositional factors, like mineral salt content, protein content, degree of acidity, etc. The natural heat stability also varies according to season and lactation period. In order to obtain milk sterilizable at different levels of concentrations, this natural heat stability must be improved and adapted. In practice, this is achieved by temperature/time combination of preliminary heat treatments on the milk before concentration and the addition of certain mineral salts. Precise indications are rather difficult, owing to the various influencing factors. The optimal heat stability has to be found by an empirical approach. In principle, more intensive heat treatments increase the heat stability, but above optimum conditions, a reverse effect is attained.

It must be understood also that processing operations, such as concentration and homogenization,

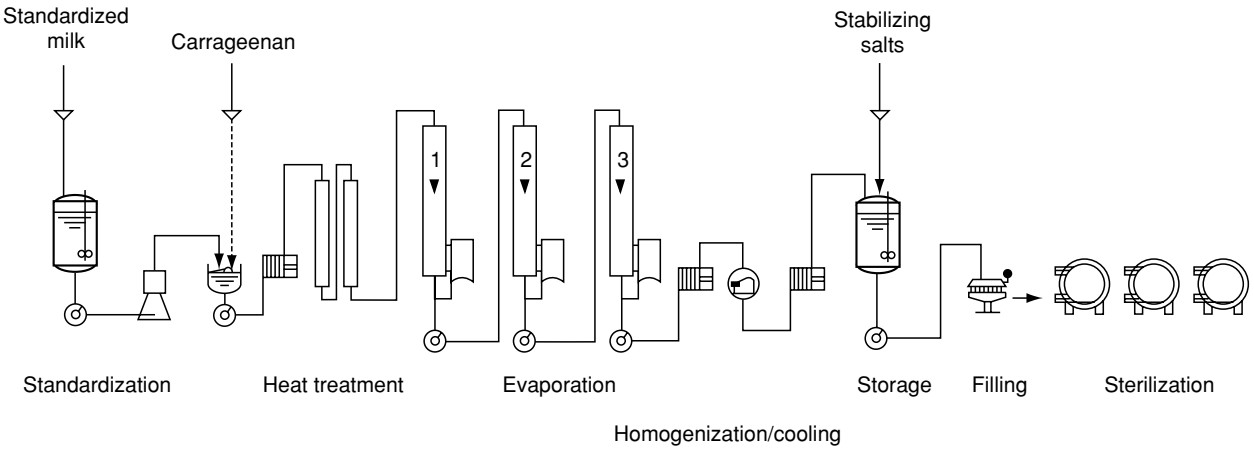

Figure 2 Basic flow diagram for the manufacture of evaporated milk. Reproduced from Condensed Milk, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

have a destabilizing effect on the concentrate. The various operations must be balanced so as to ensure a product of optimum quality.

Concentration After preheating, the milk is evaporated under vacuum. It is of utmost importance that the evaporator works under optimal hygienic conditions, despite the fact that the product is sterilized afterwards. In general, falling film evaporators are commonly used today.

Homogenization In order to obtain a satisfactory homogenization effect, the homogenizer and, in particular, the homogenizing valves must be kept constantly in the best mechanical conditions.

The homogenization temperature should preferably be around 65 °C.

A homogenizing pressure of 200–250 bar is usually applied, and the best results are obtained by a two-valve system in series. In principle, the second valve is adjusted to 20–25% of the total pressure. Excessive pressure has a destabilizing effect, which is irreversible. Following homogenization, the product is cooled for intermediate storage.

Pilot sterilization It is advantageous to fill some cans and to sterilize them under the given conditions in order to verify the heat stability of the concentrate. Corrective actions are further possible by addition of stabilizing salt and adaptation of the sterilizing conditions.

Sterilization After filling into containers (cans), the product is sterilized. The purpose of sterilization is to obtain physical and bacteriological stability. According to practical application, one speaks of commercial sterility.

The effect of sterility is expressed by the *F* value as an integral function of the lethality of the microorganisms present, related to *C. botulinum*, at the specific destruction temperature of 121 °C. Usually, an *F* value of 4 is considered satisfactory.

Today, modern instruments are available for measuring the time/temperature profile within any type of sterilizer.

Usually, rotary and continuous sterilizers are used.

Major Product Defects

Bacteriological Problems

For evaporated milk, the sterilization processes usually applied are proven and sufficiently safe to guarantee product safety and commercial sterility. If problems arise, they are mainly traced back to an insufficient can integrity. Faulty soldering, welding of can bodies, or closure seams are the weakest points. However, the presence of excessive numbers of thermophilic spores can cause severe spoilage. Appropriate line and finished product controls must be put in place. (*See* **Spoilage**: Bacterial Spoilage.)

Sweetened condensed milk is less vulnerable to spoilage because of its high sugar content, but it is not protected against osmophilic organisms. Plant and manufacturing hygiene are the keys to success.

Physical Instability

Instability may be of various individual defects, some of which may be interrelated. Separation problems can, and mainly are, a result of inadequate product viscosities. Certain texture problems may be related to nonoptimal heat treatments, but for sweetened condensed milk, it may also be a result of coarse lactose crystallization.

Age thickening is a constant problem for sweetened condensed milk. The increase of viscosity over the storage time is always present and, from a manufacturing point of view, can be corrected only by an optimal heat treatment of the milk prior to concentration. Considerable practical experience is required to keep this phenomenon under control.

Age gelation of evaporated milk is a similar phenomenon to age thickening for sweetened condensed milk but often appears only and suddenly after long storage times. The product can remain with normal physical properties up to 10 months or more and can then thicken within a few weeks. This gelation is strictly a storage defect and should not be confused with thickening or coagulation during the sterilization process or with coagulation resulting from microbiological activity. The main factors influencing age gelation are, as above, insufficient and nonoptimal preheating conditions of the milk before evaporation, marginal sterilization conditions, and an insufficient stabilizing salt level.

Inferior raw milk quality and final product storage conditions can also play an important role within the complex context of age gelation or thickening for either evaporated or sweetened condensed milk.

Products Manufactured by Recombining

As already mentioned, traditional dairy products like sweetened and unsweetened condensed milk are made by recombination using skim milk powder and anhydrous milk fat for the dairy components. The following is a brief description of these processes. (*See* **Recombined and Filled Milks**.)

When recombining sweetened condensed milk (see Figure 3), according to a process using only flash

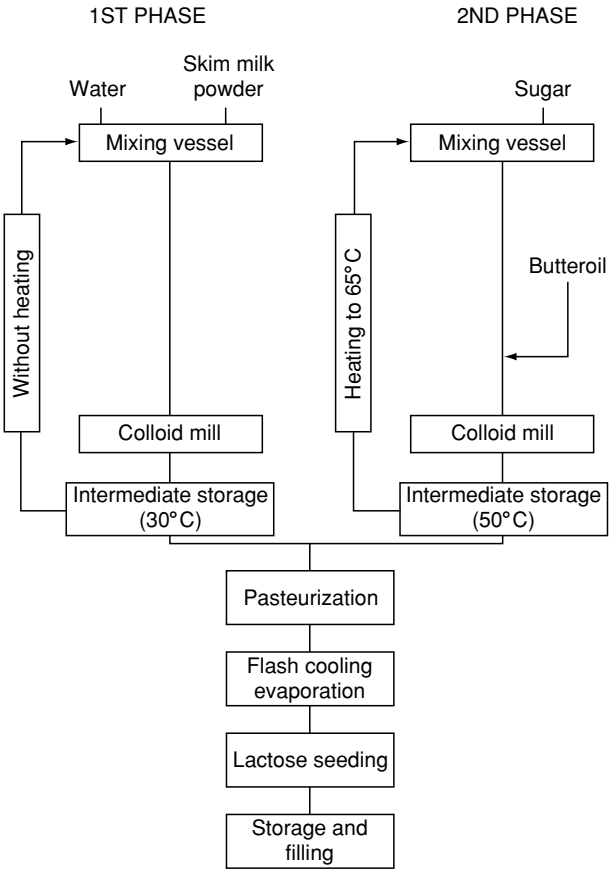

Figure 3 Flow chart for the recombination of sweetened condensed milk. Reproduced from Condensed Milk, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

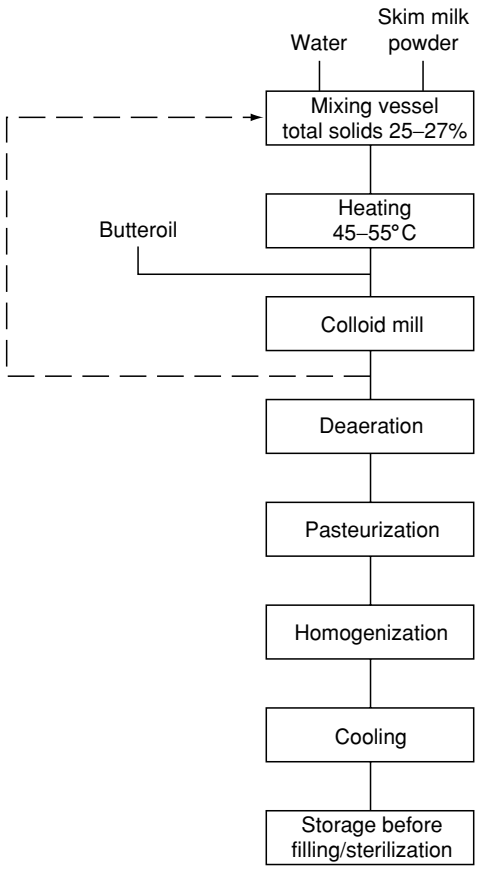

Figure 4 A flow chart for the recombination of evaporated milk. Reproduced from Condensed Milk, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

cooling evaporation, the skim milk solids have to be dissolved at a concentration of 40%, where some formation of lumps is practically unavoidable. However, these lumps are easily dispersible by mechanical force, as obtained through a colloid mill. A recirculation system has a beneficial effect on skim milk powder hydration.

In principle, dissolving sugar in a low-temperature medium does not pose a major problem but is somewhat time-consuming. Therefore, the circulated solution is heated after powder dosage in order to make the sugar dissolve rapidly.

A slightly simplified process for recombining can be used for evaporated milk (see **Figure 4**).

Dissolving skim milk powder at a low concentration generally poses less problems. Skim milk powder is certainly the most critical raw material and, depending on its characteristics, directly affects the physical properties of the final recombined product. This influence depends to a large extent on the heat treatment given to the liquid skim milk before drying. Seasonal and regional influences may also play a role.

Since milk proteins are sensitive to heat, the extent of their denaturation reflects the heat treatment applied and is used for classifying skim milk powders. The latter have been classified by the American Dairy Products Institute into three groups according to the level of undenatured whey protein nitrogen present in the powder after manufacture. In general terms, this is expressed as the whey protein nitrogen index (WPN index) (see **Table 3**). Whereas the high-heat-type powders are, in principle, the most appropriate for sterilized products, the low-heat range powders are usually more suitable for nonsterilized recombined products such as sweetened condensed milk.

Anhydrous milk fat is the principal source of fat used in recombined dairy products. It contributes significantly to the taste and milky character of the product. This is a clear indication for the quality requirements for this raw material. In certain cases,

Table 3 WPN index (values are in milligrams of nitrogen per gram of powder)

Classification	*WPN index*
High heat	≤ 1.5
Medium heat	1.51–5.99
Low heat	≥ 6.0

vegetable oils are used as butterfat substitutes, primarily for economic reasons. Such products are internationally recognized as 'filled milks.' The vegetable oils used in filled milks should be double-refined and deodorized and have a low peroxide value.

Water is a basic ingredient of all recombined dairy products. In general, a good drinking water quality is sufficient and acceptable. However, as water taste, odor, and possibly also color may influence the final product, special monitoring of water quality is important.

Buttermilk powder can be and is used in recombined products. From the technological point of view, buttermilk powder is an emulsifying aid for the fat, since it contains a relatively large amount of the phospholipids that are lost in the separation process of skim milk and anhydrous milk fat. For most applications, only sweet buttermilk is suitable; buttermilk powder obtained from acidified cream for butter production is not suitable. (*See* **Emulsifiers**: Uses in Processed Foods.)

As the natural vitamin content of fresh milk is usually reduced slightly during processing of the raw materials for recombining, vitamins A, D, and B are usually added. For filled milks, further addition of vitamin E is carried out and is recommended for nutritional purposes.

Various processing aids, mainly phosphates but also other emulsifiers or stabilizers, are added to achieve specific product characteristics and consistencies.

All additives must be in accordance with legal prescriptions or comply with the Food and Agriculture Organization/World Health Organization standard (Code of Principles). (*See* **Legislation**: Additives.)

See also: **Buffalo**: Milk; **Canning**: Principles; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; Chemical and Microbiological Changes; **Legislation**: Additives; **Milk**: Dietary Importance; **Recombined and Filled Milks**; **Spoilage**: Bacterial Spoilage; **Storage Stability**: Parameters Affecting Storage Stability

Further Reading

Codex Alimentarium (1984) Rome: Food and Agriculture Organization of the United Nations/World Health Organization.

Hunziker OF (1935) *Condensed Milk and Milk*, 5th edn. La Grange, IL: Hunziker.

Pizzoferrato L, Manzi P, Vivanti V *et al.* (1998) Mailland reaction in milk-based foods: nutritional consequences. *Journal of Food Protection* 61: 235–239.

Ramonaityte DT (2001) Copper, zinc, tin and lead in canned evaporated milk, produced in Lithuania: the initial content and its change at storage. *Food Additives and Contaminants*. 18(1): 31–37.

Webb BH, Johnson AH and Alford JA (1974) *Fundamentals of Dairy Chemistry*. Westport, CT: AVI.

Confectionery *See* **Sweets and Candies**: Sugar Confectionery

CONJUGATED LINOLEIC ACID

C E Fernie, Scottish Crop Research Institute, Invergowrie, Dundee, UK

What is Conjugated Linoleic Acid (CLA)?

Conjugated Linoleic Acid (CLA) is the collective name given to a group of 18:2 free fatty acid positional and geometric isomers, containing two conjugated double bonds (**Figure 1**). Most fatty acids commonly occurring in nature contain methylene-interrupted double bonds, e.g., *cis*-9, *cis*-12 18:2 (linoleic acid). However, conjugated dienes have been known for more than 50 years. They are seen as the products of both autooxidation and of partial hydrogenation of polyunsaturated fatty acids, as well as being minor components of dairy and ruminant lipids. There is increasing interest in CLA, as in animal trials, it has shown potentially beneficial effects on a variety of medical conditions such as cancer and type 2 diabetes, as well as affecting body composition parameters such as bone composition and lean body mass. As a result, this is an area of

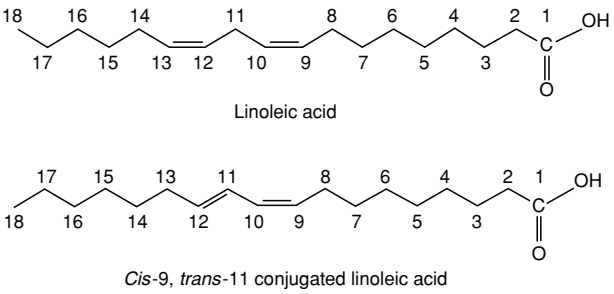

Figure 1 Linoleic acid is a *methylene-interrupted* fatty acid, i.e., it has a CH_2 group between the two double bonds, at position 11. CLA is a *conjugated* fatty acid as there are no groups between the double bonds in the carbon chain.

continuing growth in research in a wide variety of scientific fields. This review will attempt to give an overview of the current state of knowledge in this developing area of interest.

Occurrence in Natural Products

The most common source of natural CLA in the human diet is from the meat and milk products of ruminant animals. CLA isomers are present in these materials at levels of approximately 0.5% of total fat on average, although seasonal variations are observed. The predominant isomer is the *cis*-9, *trans*-11 acid, at levels of 80–90% of the total CLA, with minor amounts of a wide number of other isomers present, ranging from the 6,8 through to the 13,15 positional isomers, in almost all of the different geometric configurations possible (i.e., *cis*, *cis*; *cis*, *trans/trans*, *cis* and *trans*, *trans*), particularly in milk fat and cheese. Biohydrogenation of linoleic acid is thought to be the main mechanism of formation of natural CLA. The process involves the action of the ruminant bacterium *Butyrivibrio fibrisolvens* on dietary linoleic acid (via linoleic acid isomerase) to give CLA as an intermediate in the formation of *trans*-11 18:1 (vaccenic acid) and subsequently 18:0 (stearic acid). As levels of CLA produced by this mechanism do not seem to be high enough to account for the amounts of CLA in ruminant tissues, a second mechanism has been proposed, involving the action of the Δ-9 desaturase enzyme on vaccenic acid within adipose and mammary tissues to give the *cis*-9, *trans*-11 CLA isomer (Figure 2).

The variety of conjugated isomers in ruminant milk and tissue is thought to be due to a combination of double bond migration, and the action of specific *cis*, *trans* isomerases in the rumen.

Altering the feeding regime of the animals can vary the levels of CLA in ruminant fats. Pasture-fed animals have higher levels of CLA than those given

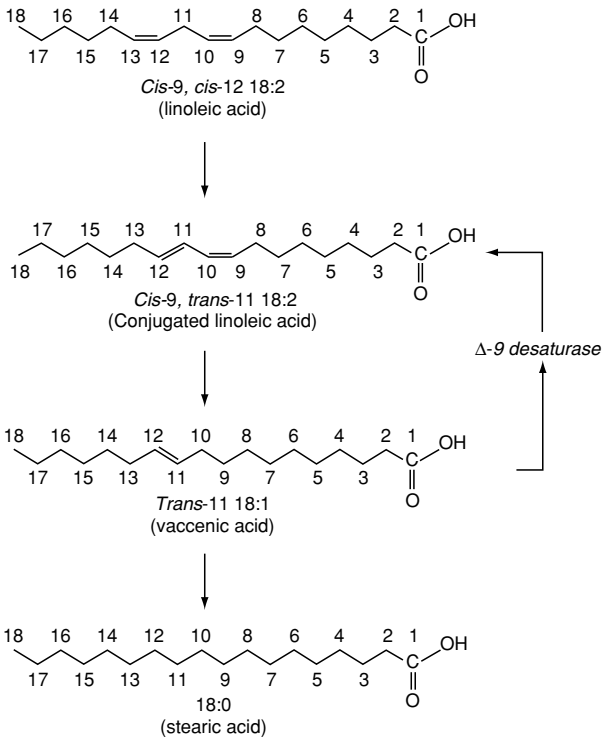

Figure 2 Biohydrogenation of linoleic acid to produce CLA, both as an intermediate towards stearic acid and by desaturation of vaccenic acid.

commercial feed. However, feed enriched with unsaturated fatty acids can also increase the CLA levels in milk and meat. There can be wide variations in CLA production between animals of different breeds, and at different stages of lactation. This may relate to the Δ-9 desaturase level of the individual animals.

CLA has been reported in fish, and its occurrence in farmed fish is probably due to the CLA content of vegetable oil in fish food. However, there is concern that CLA can be misidentified in fish tissues, as it elutes on some gas chromatographic columns in the same region as the naturally occurring 18:4 isomers. This may have been the case in early reports of low levels of CLA in trout, shrimp, and mussels, as currently no mechanisms are known to account for CLA formation in marine foods.

CLA is also present in small amounts, as a result of the hydrogenation process, in margarines, cooking oils and other partially hydrogenated fats.

Commercial Production

Mixtures of CLA isomers are available commercially, for use in scientific studies and as reference standards, and also as dietary supplements. These products, which are usually in the free fatty acid form, are typically produced by the alkaline isomerisation of a

linoleate-rich vegetable oil, such as sunflower or safflower oil. The CLA produced from this process in aqueous solution generally contains four main CLA isomers: *trans*-8, *cis*-10; *cis*-9, *trans*-11; *trans*-10, *cis*-12, and *cis*-11, *trans*-13. Under harsh reaction conditions, these isomers can be present in approximately equal proportions. It is possible, however, to manipulate the reaction temperature and time to give a mixture that is greater than 90% of the *cis*-9, *trans*-11 and *trans*-10, *cis*-12 isomers, in approximately equal proportions. This is more desirable, particularly for supplement usage. By using propylene glycol instead of water, with an appropriate basic catalyst (e.g., potassium hydroxide), the amounts of the minor isomers can be reduced to around 1%, but small amounts of isomerization will always occur.

The processes described above all give mixtures of CLA isomers. It is also possible to obtain CLA enriched in one particular isomer. Enzymes derived from fungi such as *Candida rugosa* and *Geotrichum candidum* show a preference for the *cis*-9, *trans*-11 isomer, and, by preferentially esterifying this CLA isomer, give an unesterified fraction that is rich in the *trans*-10, *cis*-12 isomer. This has been demonstrated on a laboratory scale, but has yet to be optimized for commercial production. Alternatively, a series of selective crystallisations can be used to fractionate the *cis*-9, *trans*-11 and *trans*-10, *cis*-12 isomers, to give material of high isomeric purity. In the future, the production of CLA is likely to expand to include production of CLA-containing triacylglycerols, as this is the form of lipid most generally present in foodstuffs.

Analysis of CLA

The primary concern in analyzing CLA is the complete separation and quantitative identification of the individual isomers in the sample. The form of the sample will also determine the approach to analysis. Samples are generally either in the free fatty acid form, or present as esterified CLA as a component in a biological matrix, from which the CLA needs to be extracted first. A number of analytical techniques have proven useful. ^{13}C NMR allows identification and quantification of all the positional and geometrical isomers in a single analysis, but has the drawback of requiring a comparatively large amount of sample, so it is not helpful for analyzing biological extracts. Gas chromatography (GC), gas chromatography–mass spectrometry (GC-MS) and high-performance liquid chromatography (HPLC) can all give large amounts of information regarding sample composition, but for these methods, the sample must be extracted and also derivatized prior to analysis. A combination of the above techniques is often required to give a comprehensive analysis.

As CLA is prone to isomerization, care must be taken in the extraction method chosen to obtain CLA from tissue samples. A mild alkaline hydrolysis can be used, for instance, to give the CLA as free fatty acid, which can then be derivatized for analysis using one of the methods described below.

Prior to GC analysis, fatty acids must be converted to derivatives (usually methyl esters). Although there are standard methods used to derivatize methylene-interrupted fatty acids, it has become apparent over the last few years that some of these methods, in particular those using an acidic catalyst, are unsuitable for use with conjugated free fatty acids. Under vigorous conditions, they lead to isomerization of the sample and artefact formation. Though base-catalyzed methods do not normally esterify methylene-interrupted free fatty acids, esterification of free CLA can be achieved, but some of the methods used have also shown artefact formation on reaction with CLA. Phase transfer methods using methyl iodide in a basic buffer solution have been utilized on a small scale with consistent results. Despite the drawbacks, acidic methods can give good quantitative results providing that the reaction is carried out under mild conditions with fresh reagent. Esterified CLA can be converted to the methyl esters by the standard method of transesterification with sodium methoxide in methanol solution. The most useful GC columns for resolving complex mixtures of CLA isomers are longer, more polar columns such as the 100-m Cp-Sil 88 or the 120-m BPX-70 capillary columns. Neither of these columns will completely resolve a complex mixture of CLA isomers, although they will give good results on most commercial samples. For this reason, complex mixtures of biological CLA should also be analyzed by either HPLC, or GC-MS, or both.

HPLC has proved especially useful in the analysis of conjugated dienes, such as CLA. A UV detector is usually employed as conjugated dienes show a strong absorbance in the region of 230–235 nm. Silver-ion HPLC, employing more than one column, can be used to separate all the positional and geometric isomers in a commercial CLA mixture, as methyl esters. Reversed-phase HPLC has been used successfully to concentrate the CLA methyl esters of biological samples for further analysis.

For GC-MS, a different type of derivative, usually containing nitrogen, is required to give good-quality spectra of CLA. Both dimethyloxazoline (DMOX) derivatives and 4-methyl-1,2,4-triazoline-3,5-dione (MTAD) adducts are commonly used and can be used to locate the conjugated double bond positions in both CLA and its metabolites.

Biological Activity

As mentioned above, CLA has exhibited biological effects in a number of major medical conditions. However, many of the CLA mixtures used for early trials were poorly characterized. As a result, there is little conclusive information as to which CLA isomers are the most active and under which circumstances. It is only recently that individual pure *cis*-9, *trans*-11 and *trans*-10, *cis*-12 isomers have become available for clinical trials. Consequently, the discussions below assume that studies were carried out with commercial isomer mixtures of free fatty acids, although where a particular isomer has demonstrated a reproducible effect this is indicated.

Anticarcinogenicity

The interest in CLA as a biologically active material, rather than as a minor lipid component or an oxidation product, began in the early 1980s, when researchers at the University of Wisconsin reported the discovery of an extract from grilled minced beef, which was found to inhibit mutagenesis. Since then, many studies have been performed, both *in vivo* using a number of different animals, and *in vitro* with a variety of animal and human cell lines. Although other lipids, such as fish oil, have been shown to have anticarcinogenic effects, CLA appears to be potent at much lower concentrations (e.g., from 0.1% w/w in rat tumors). A number of studies have concluded that the *cis*-9, *trans*-11 isomer is the active CLA isomer against cancer, but this remains to be proven.

In vivo, dietary supplementation with CLA decreased the number and size of mammary tumors in rats and mice, independent of the amount and types of other fat in the diet. However, no effect was seen in trials using tumor cells from hormone nonresponsive cell lines such as the WAZ-2T (-SA) and MDA-MB-231 cell lines. Further studies are required in this field, as this may be relevant to elucidating the mechanism by which CLA acts on tumors. In tests on skin carcinogenesis in mice, topically applied CLA was found to be effective in reducing the size and number of skin tumors, although dietary supplementation with CLA only affected the number and not the size of the tumors. Fewer data have been published on the effect of CLA on prostate cancer and on forestomach and colon cancers, with conflicting data being obtained in different mouse and rat models.

Much of the *in vitro* work on carcinogenesis has been done on the estrogen-responsive MCF-7 human breast cancer cell line. In the majority of these studies, CLA has been shown to reduce cell growth in a dose-dependent manner. Again, however, estrogen-unresponsive cell lines failed to show tumor inhibition on treatment with CLA. Limited numbers of studies on a variety of other human cancer cell lines have also shown tumor inhibition in colorectal, liver, prostate, and lung cells.

A number of studies have been conducted to try and elucidate the mechanism by which CLA acts on cancer cells. Initial suggestions that CLA acts as an antioxidant have recently been contradicted by a number of trials. CLA can block the formation of carcinogen DNA adducts in rats in certain organs. Another potential mode of action would be by affecting the production of eicosanoids implicated in tumorigenesis. CLA is incorporated into both neutral lipids and phospholipids, at varying levels, depending on the tissue type, and can be metabolized to give 18:3, 20:3, and 20:4 fatty acids containing conjugated double bonds. Dietary supplementation with CLA has also been shown to reduce the synthesis of the eicosanoid prostaglandin E_2 in mouse epidermis. However, the direct method of action has still to be determined.

There appears to be a direct link between levels of CLA in breast tissue during maturation and the limiting of lateral branching of the terminal end bud cells of the developing breast, leading to a decreased risk of breast cancer in later life. Again, the mechanism for this effect has yet to be shown. Interestingly, a number of studies examining human dietary CLA intake and breast cancer risk world-wide have failed to show any consistent positive or negative effect of dietary CLA on cancer development. This may be due to limitations of the accuracy of the dietary surveys conducted.

Body Composition

In animal studies, CLA reduced body fat and increased lean body mass in growing animals, for a number of species, including mice, rats, and pigs. Other physiological factors connected with these effects in different studies were a reduction in lipoprotein lipase activity, an increase in the enzyme carnitine palmitoyl transferase (associated with β-oxidation of lipids), and reduction in adipocyte size (but not cell number). Reduced body fat and increased body protein and water have been directly associated with the *trans*-10, *cis*-12 isomer. Potential mechanisms for these effects include alteration of membrane structure, effects on eicosanoid and cytokine production, and changes in peroxisome proliferator-activated receptor (PPAR) activity.

Feeding of CLA also leads to a loss of monounsaturated fatty acids, which is an effect seen in treatment to reduce obesity. It has recently been demonstrated that this effect is due to repression of the expression of the stearoyl-coenzyme A desaturase gene in fat cells by the

trans-10, *cis*-12 isomer, leading to smaller lipid droplets containing lower levels of monounsaturated fatty acids. This is not seen with the *cis*-9, *trans*-11 isomer.

A few studies of the influence of CLA on human body composition have been conducted. The amounts of CLA fed were variable (between 1.7 and 7 g per day), and a number of the subjects studied were either athletes or body builders. In general, although there was little effect on body weight, in most of the subjects, body fat was either reduced or stayed constant, and lean body mass was either constant or increased. However, these variations were in most cases not significant. The existing data are not consistent enough to show a dose-dependent response, and the effects of exercise should be taken into consideration. Also, studies in humans were conducted on adults, whereas animal data were collected from growing subjects. It has been speculated that feeding CLA affects plasma leptin levels and thus acts as a control on food intake. However, recent data in humans showed that, although supplementation decreased leptin levels in the short term, after approximately 7 weeks levels returned to baseline values, and no alterations in either appetite or body composition were observed. Further studies are required with individual isomers, in particular the *trans*-10, *cis*-12 isomer to assess fully the effects and the potential to treat obesity.

Immunity

In mice, dietary CLA is incorporated into liver lipids by displacing linoleic and arachidonic acids, thus modifying the *n*-6 fatty acid composition of the phospholipids. Subsequent experiments in rats have shown that this phospholipid modification affects eicosanoid production. For instance, feeding 1% CLA for 2 weeks reduced serum prostaglandin E_2 levels by approximately 50%. CLA can also lower the levels of some leukotrienes (e.g., LTB_4) in rats. This may then affect cytokine production, but further studies are required in this area. It has effects on immunoglobulin production, specifically increasing levels of IgA and IgG, which are both associated with antiallergic effects.

Few data are currently available from human studies, but a recent trial on women showed no changes in immune function during the course of the study. This may be due to a dose effect, as the levels of CLA per weight in animal models tend to be much greater than the levels fed in the human studies.

Atherosclerosis

Experiments on atherosclerosis have been conducted on both rabbits and hamsters. Feeding CLA at 0.1–0.5% of the diet in rabbits for 22 weeks showed a reduction in plasma triacylglycerol, total cholesterol, and LDL-cholesterol levels. There was also an effect on plaque formation, with a reduction in the CLA fed animals. In hamsters, CLA feeding at similar levels was found to reduce total cholesterol and fatty streak area, although linoleic acid feeding showed similar effects. An isomer effect was seen in the hamster model, with only the *trans*-10, *cis*-12 isomer decreasing triacylglycerol and cholesterol. Contradictory data have been obtained in C57BL/6 mice, in which an increase in aortic fatty streaks was reported.

Bone Metabolism

Lipids play an important part in bone metabolism as they are known to influence both bone modeling and remodeling. Both CLA and *n*-3 fatty acids modulate the production of prostaglandin E_2 in chicks and rats, leading to increased bone formation and decreased resorption. CLA is thought to act on prostaglandin E_2 production by influencing the cyclooxygenase enzymes, which produce prostaglandin E_2 from arachidonic acid. This may occur through the production of competitive 20:4 conjugated CLA metabolites. As also seen in other studies, CLA reduces the level of LTB_4, which is a strong bone resorption factor. CLA may influence cartilage functions in growing animals. As yet, few experiments have been conducted with CLA, and further data are required to elucidate its mechanism of action.

Diabetes

In type 2 diabetes, treatment aims to reduce insulin resistance and improve glucose uptake. In *in vitro* tests, chemicals such as thiazolidinediones show these effects by directly affecting PPARγ expression. CLA, in particular the *cis*-9, *trans*-11 isomer, has also been shown to bind to, and activate, PPARγ, as have other unsaturated fatty acids. In Zuker diabetic fatty (*fa/fa*; ZDF) rats, a diet containing 1.5% CLA produced normal glycemic responses to tolerance tests, as well as reduced hyperinsulinemia.

Enhanced CLA Levels and Food Production

In animals, there are two different reasons to try to increase the levels of CLA in tissue and/or milk. Firstly, CLA has effects on body composition, leading to, for instance, increased lean meat and lowered levels of fat in pigs and cows. This is becoming more desirable as public concern increases due to suggested links between increased dietary levels of fat and conditions such as heart disease and obesity. The second reason is because of the potentially beneficial effects of CLA itself in a number of diseases. Although

current dietary levels of CLA have been shown to have little or no direct effect on various conditions, enhanced levels in food are hoped to lead to a protective effect against conditions such as cancer.

Most studies that aim to increase CLA levels have been conducted with cows. Feeding of mixes containing unsaturated lipids such as vegetable or marine oils leads to increases in CLA levels of up to fivefold. Increasing CLA, however, generally causes a concurrent increase in the level of other *trans* fatty acids, in particular the *trans*-10 octadecenoic acid, which has negative effects in both cows and humans. The levels of this fatty acid have been suppressed by feeding a buffer.

Feeding CLA to chicks has also shown an effect on their immune response, such that less tissue wasting occurs on immune stimulation, thus reducing the need for antibiotic treatment and improving meat quality.

CLA is present in human milk, but not infant formula, and it has been suggested that this is a factor in the lower levels of food allergies, immune deficiency, and diabetes in children that are breast-fed. This, however, has not been tested in sufficiently large data sets.

Enhancing CLA in dietary components is only one way of increasing CLA intake. Other methods include supplementation and production of a functional food. CLA supplements are already available in the USA and Europe, and currently are mainly targeted at the body building population and at slimmers. These products are either liquid or encapsulated CLA, of a composition similar to that of the commercial products described above.

A functional food is defined as a food product that has been enhanced to give specific, beneficial, biological effects. Examples include margarines containing stanol esters, which have demonstrated a cholesterol-lowering effect in humans. These products are designed to be available on supermarket shelves to the average consumer. CLA would be an ideal candidate for incorporation into a food item such as a dairy product, if it could be shown to have specific biological effects in humans at reasonable (i.e., non-toxic) levels.

Safety Issues

There are few data available upon the toxicological effects of CLA. Most of the information comes from animal models, and the data are contradictory. Although few effects have been seen in rats and hamsters, experiments conducted with a variety of mouse models have shown increases in liver weight, transaminase levels, and steatosis. Insulin resistance has also been reported. These effects may be relevant purely to the mouse model. In humans, in a study on overweight or obese volunteers undergoing CLA supplementation, at levels of 3.4 g of CLA per day, blood lipids, liver factors, and other hematological end points, as well as physical parameters such as blood pressure, were examined, but no differences were found between the supplemented and control groups.

This is an area requiring urgent review, particularly with the prospect of CLA incorporation into a functional food and considering the lack of data on specific CLA isomers.

Summary

Conjugated linoleic acid shows potential as a treatment for a variety of human ailments, as well as a means towards obtaining meat products with a fat-to-lean-meat profile closer to that which we currently regard as healthy. However, the data available so far come mostly from animal studies and show conflicting results in some cases. A lack of consistent analytical methodology has further complicated the situation. There is a need for clinical trials in humans with specific CLA isomers, in order to elucidate the mechanisms by which these compounds show their effects, as well as to ascertain the appropriate dosage levels, with no significant side-effects.

See also: **Analysis of Food**; **Fats**: Production of Animal Fats; **Fatty Acids**: Properties; Metabolism; Dietary Importance; *Trans*-fatty Acids: Health Effects; **Food Fortification**; **Functional Foods**; **Immunology of Food**; **Obesity**: Fat Distribution; Treatment; **Spectroscopy**: Nuclear Magnetic Resonance; **Triglycerides**: Structures and Properties; **Vegetable Oils**: Oil Production and Processing; Dietary Importance

Further Reading

Banni S and Martin J-C (1998) *Trans Fats in Human Nutrition*. In: Sébédio JL and Christie WW (eds) Conjugated Linoleic Acid and Metabolites, from Trans Fatty Acids in Human Nutrition. Dundee, UK: The Oily Press.

Christie WW, Sébédio JL and Juaneda P (2001) A practical guide to the analysis of conjugated linoleic acid. *Inform* 12: 147–152.

Food Research Institute, University of Wisconsin-Madison, http://www.wisc.edu/fri/clarefs.htm

Gnädig S, Rickert R, Sébédio JL and Steinhart H (2001) Conjugated linoleic acid (CLA): physiological effects and production. *European Journal of Lipid Science and Technology* 103: 56–61.

Jahreis G, Kraft J, Tischendorf F, Schöne F and von Loeffelholz C (2000) Conjugated linoleic acids: physiological effects in animal and man with special regard to body

composition. *European Journal of Lipid Science and Technology* 102: 695–703.
Sæbø A (2001) Commercial production of conjugated linoleic acid (CLA). *Lipid Technology Newsletter* 7: 9–13.
Yurawecz MP, Mossoba MM, Kramer JKG, Pariza MW and Nelson GJ (eds) (1999) *Advances in Conjugated Linoleic Acid Research*, vol. 1. Champaign, IL: AOCS Press.

CONSUMER PROTECTION LEGISLATION IN THE UK

J Carton, Manchester Metropolitan University, Manchester, UK

Background

We are all consumers throughout our lives. Every member of society is a consumer in one form or another, whether they consume goods, services, or intangibles such as gas or electricity. For the purposes of this article, it is assumed that a consumer is someone who buys or uses such goods or services for private use or consumption. What we consume may depend upon the amount of disposable income that we possess and what stage we have reached in our lives. Factors such as availability and price will also play an influential role. Food is a particularly important consumable, since we all need it to survive.

Protecting the consumer is usually something which political parties agree upon as being a good thing to do, but how they go about achieving it differs. There are also various ways of trying to ensure that consumers are protected, such as education, advice, and industry self-regulation. The emphasis here will be on *legislation* as a means of protecting consumers and on *food* as a major consumer product. In addition, only the main pieces of legislation will be addressed. Discussion will also be mainly confined to the laws which apply in England and Wales, although other countries throughout the world will have similar consumer protective measures.

Consumer protection legislation has only really developed in the UK since the 1960s, and three government-initiated reports have assisted with this development. They are the Final Report of the Moloney Committee on Consumer Protection (Cmnd 1781, July 1962), the Crowther Committee (Cmnd 4596, March 1971), and the White paper, modern markets: confident consumers (July 1999).

The Legal System

Both criminal and civil law play a role in protecting the consumer, and much of the current legislation emanates from the UK's membership of the EU. Sometimes, the aim is not necessarily to protect consumers: it could be to ensure fair trading in the business environment with the protection of consumers as a side-effect.

Criminal law aims to protect the public at large and creates criminal offenses for which penalties, such as fines and even imprisonment, can be invoked by the courts. Action is normally taken by public bodies such as local authority trading standards or environmental health departments. Civil law, however, seeks to settle disputes between two private parties, providing a remedy for someone who has been injured or harmed in some way.

The nature of much legislation in this field is strict liability, which means that it need not be shown that the perpetrator knew or intended to commit an offense or breach a provision, only that they did so. This is illustrated in criminal law by the case of *Parker v Alder (1889) 1 QB 20* in which the seller was found liable for the adulterated milk he was selling even though it was satisfactory when it left his control. In civil law, the case of *Frost v Aylesbury Dairy Co (1905) 1KB 608*, in which milk infected with typhoid germs was sold, demonstrates the strict liability nature of the legislation. It is unlikely that either seller was at fault in any way but both were nevertheless held to be in breach of the relevant legal provisions.

Criminal Legislation

The Food Safety Act 1990

This is the main piece of criminal legislation which specifically regulates food. It is concerned with both the safety of food and hygiene of premises as well as the quality, description, labeling, and composition of food.

There are four main offense sections within the Act which between them cover the areas mentioned above. In addition, there are a great many sets of regulations, orders and codes of practice made under the Food Safety Act which provide detailed requirements to supplement its provisions.

Food Not of the Nature, Substance or Quality Demanded by the Purchaser

Selling food which is not that which was demanded by the purchaser is an offense contrary to section 14(1) of the Act. In fact, there are three separate offenses, selling food which is not, first, of the nature demanded, second, of the substance demanded, and, third, of the quality demanded. Examples of cases prosecuted serve to illustrate what these phrases mean, although it is generally accepted that there is some overlap between the three.

It is generally accepted that 'nature' would refer to foods not being of the variety requested by the purchaser, e.g., where cod is requested, and haddock is supplied as a substitute. In *Shearer v Rowe (1985) 149 JP 698*, butchers were convicted of selling 'minced beef' and 'minced steak,' which were not of the nature demanded since they contained 10% pork and 10% lamb. Prosecutions relating to 'substance' are often taken when a food contains a foreign body or mold as in *Southworth v Whitewell Dairies Ltd (1958) 122 JP 322*, in which a bottle of infants' school milk contained a small sliver of glass which the child drinking it was able to suck up the straw being used to drink the milk. Allegations of food not being of the 'quality' are generally made when there is a failure to meet standards set by regulations or when there is a shortfall in what an ordinary consumer would expect. In *TW Lawrence & Sons Ltd v Burleigh (1981) 146 JP 134*, a regular consumer made a telephone order for mince and, when it was delivered, was shocked at the appearance of the product, in that she felt that there was too much visible white fat. The mince was found to contain 30.8% fat, and the court was satisfied that this was in excess of what the purchaser expected.

Food Which Does Not Meet the 'Food Safety Requirements'

Section 8 of the Act deals with the safety of food. It is an offense, contrary to section 8(1), to sell food which fails to comply with food-safety requirements. In fact, the offense is wider than that in section 14, since it is not restricted to a sale. The offense is committed if such food is offered, exposed, advertised, or in possession for sale [s.8(1)(A)], or has been deposited with, or consigned to, someone else in preparation for sale. This thereby ensures that the whole supply and distribution chain is regulated.

Consumers are further protected in that there is a presumption in section 3 that food is for human consumption. Therefore, if the trader wishes to argue that the food was not ready for consumers, e.g., because it had been rejected or had not yet been checked, then the onus is on him prove so.

There are three ways in which food may fail to comply with food-safety requirements. First, the food may be 'unfit for human consumption' (s.8(2)(b)), e.g., because it is putrid or decomposed. This is illustrated by the case of *Chibnall's Bakeries v Cope Brown (1956) Crim LR 263*, in which a used, dirty bandage was found in a loaf of bread. Second, the food may fail to comply with the food-safety requirements if 'it is so contaminated that it would not be reasonable to expect it to be used for human consumption in that state' (s.8(2)(c)). Food could be contaminated in a number of ways, e.g., by mold, heavy metals, mites, foreign bodies, radioactive substances, or unauthorized additives. In *Kwik Save Group plc v Blaenau Gwent Borough Council (1995), unreported (CO/2246/95)*, a packet of crumpets was found to breach this provision when the crumpets contained mold spots 5 hours after being purchased. Third, the food may have been rendered injurious to health (s.8(2)(a)) as in *Cullen v McNair (1908) 72 JP 376*, in which the addition of a particular preservative to cream rendered the product injurious to the health of a substantial portion of the community, namely invalids and children. To create the offense, the food must have been rendered injurious to health by one of the operations detailed in section 7(1) with the intention that it would be sold for human consumption. The operations are '(a) adding any article or substance to the food; (b) using any article or substance as an ingredient in the preparation of the food; (c) abstracting any constituent from the food; and (d) subjecting the food to any other process or treatment.'

Food Rendered Injurious to Health

As well as supplying, etc. such food, the actual act of rendering the food injurious to health in one of these ways is an offense contrary to section 7(1). This is one of the few consumer protection offenses which does not require the perpetrator to be acting in the course of a trade or business. Generally, this provision is not used as a consumer protection measure but potentially could be used to deal with those who deliberately sabotage food products.

There are many sets of regulations made under the Food Safety Act 1990 which also help to ensure that consumers receive safe food. The Food Hygiene Regulations 1995 regulate the cleanliness of food premises and the use of practices within the food industry. There are specific regulations for the meat, dairy, and egg industries. In addition, there are requirements for the registration and or licensing of certain premises and a prohibition or restriction on the use of certain processes, ingredients or additives.

Food Falsely Described or Presented

Consumers are not only concerned about their food being safe but also require it to be accurately described and presented. Section 15 of the Food Safety Act creates a number of criminal offenses to cover this area. Labels and adverts which depict or describe food must not be false or mislead as to the nature, substance, or quality of the food (see above), otherwise an offense may be committed contrary to section 15(1), (label) or 15(2), (advert). There have been many cases alleging false descriptions of food, although most of them have been taken under other legislation, e.g., *Henderson and Turnbull v Adair (1939) JC 83* in which a product was falsely described as 'Scots Whisky' when it was a blend of 33% Scotch 'pot-still' and 67% Irish 'patent still' spirits. An example of a misleading label is illustrated by the case of *Van den Berghs & Jurgens Ltd v Burleigh (1987) unreported*. The products in question were cream substitutes 'Elmlea Whipping' and 'Elmlea Single.' The use of cream type cartons, the words 'single' and 'whipping,' the coloring of the words, and the rural scene on the carton were considered to be misleading. It was felt that an ordinary consumer would think that that the products were real cream. This is an example of labels, which, when examined in their entirety, were considered to be misleading even though all the information on them was factually correct. The cartons were clearly labeled 'the real alternative to cream.' There is also the possibility that the food can describe itself, and to that end, it is an offence to sell, etc. 'any food the presentation of which is likely to mislead as to the nature or substance or quality of the food' (s.15(3)).

Food Standards Regulations

Again, there are various sets of regulations requiring specific labeling and even the composition of certain foods. Consumers can only make informed purchasing decisions if they know what is in the product. What has to be on a food label depends on what the food is, how it is packed (prepacked or loose), and where it is sold (retail, wholesale, or from a catering establishment). The Food Labelling Regulations 1996, which implement EC Directive 79/112 (now consolidated as Council Directive No. 2000/13/EC), require most prepacked foods, by virtue of regulation 4, to be marked with the general labeling requirement (r.5), i.e., the name of the food, its ingredients, (including the percentage quantity of certain ones) 'best before' or 'use by' date plus the name and address of the manufacturer. In addition the origin of the food and instructions for use must also be indicated if the consumer would be misled without them. Consumers are also interested in knowing whether foods contain any irradiated or genetically modified ingredients, and so there are requirements relating to such processes.

Some regulations lay down compositional standards for certain products as well as additional or alternative labeling requirements, e.g., the Meat Products and Spreadable Fish Products Regulations 1984 require declarations of minimum amounts of meat and fish (regulations 7 and 8) and also declarations of water content on certain sliced products (regulation 9). Regulation 4 and Schedule 4 prescribe minimum meat contents for products such as burgers (80%), pork sausages (65%), cooked potato and meat pies and fish pate.

The Trade Descriptions Act 1968

This piece of legislation can be used as an alternative to section 15(1) or (2) of the Food Safety Act 1990 to deal with falsely described food. It also covers oral descriptions which section 15 does not. The offenses are contained in section 1, and whereas the Food Safety Act is obviously restricted to food, this Act covers 'goods' generally, including food. The offenses within section 1 are for actually applying a false trade description (defined in s.2) to goods (s.1(1)(a)) and for supplying and offering to supply goods to which a false trade description has been applied. Section 14 also contains offences of recklessly or knowingly making false statements in relation to accommodation, services or facilities which could include references to food services such as hotel meals.

The Agriculture Act 1970

Since our food products are derived from crops and animals, it also necessary to control animal feeding stuffs and fertilizers so that what ends up in the human food chain is nutritious and not harmful in any way. There is a great deal of legislation in this area and two of the most relevant pieces are the Feeding Stuffs Regulations 2000 and the Fertiliser Regulations 1991.

The General Product Safety Regulations 1994/SI 2328

These Regulations cover the safety of consumer products generally and are supplemented by Part II of the Consumer Protection Act 1987 and various specific sets of Regulations governing particular products. The General Product Safety Regulations implement those parts of Council Directive 92/59/EEC on general product safety not already covered by the Consumer Protection Act.

The main requirement is that producers (defined in r. 2(1)), should only place safe products on the market. This is known as the 'general safety

requirement' and is laid down in r.7. A breach of this provision may amount to an offence contrary to Regulation 12. Product is also defined in r.2(1) and is confined to consumer products. Whether or not a product is safe depends upon its presenting no or only minimum risk (r.2(1)). The labeling, presentation, packaging, and instructions, as well as its actual design and construction can all be taken into account when trying to assess whether a product is safe. There are various standards available against which safety can be measured as detailed in r.10. These include specific UK laws, European Standards, British Standards, Industry Codes of Practice, and the current practises within a particular industry. As well as these guidelines, the 'safety which consumers may reasonably expect' should also be borne in mind. It is unlikely that these regulations would be used to take a prosecution in relation to a food product, although food clearly falls within the definition of a product in r.2(1). The Food Safety Act 1990 will almost inevitably be a more appropriate act to use, and in any event, it is arguable that the Food Safety Act, or more specifically Council Directive 89/397/EEC which implements it, falls within the exemptions detailed in regulation 3 of the General Product Regulations 1994: 'These Regulations do not apply to . . . .(c) any product where there are specific provisions in rules of Community law governing all aspects of the safety of the product.' If it could be argued that there is some aspect of the safety of food which is not covered by the Directive then presumably the General Product Safety Regulations could be used to deal with that aspect.

The Weights and Measures Act 1985

Legislation in this area covers goods, equipment, marking, and units. In addition, there are specific provisions regulating packers and importers and the average quantity system which is used for most prepacked goods throughout the European Community (Part V, Weights and Measures Act 1985 and the Weights and Measures (Packaged Goods) Regulations 1986). Consumers are protected by the creation of offenses in relation to supplying short-weight products, including food, and using suitable and accurate weighing and measuring equipment. Certain foods are required to be sold in prescribed quantities, e.g., bread in multiples of 400 g, so that consumers can make meaningful comparisons between producers (Weights and Measures (Miscellaneous Foods) Order 1988). Also, certain units have to be used in the sale of food (Units of Measurement Regulations 1986, 1994, and 1995). This is well illustrated by the case of *Thoburn and others v Sunderland (2002) 3 WLR 247*, in which a prosecution was mounted against Mr Thorburn who had continually refused to sell his fruit and vegetables in metric quantities.

The Prices Act 1974

Consumers need full and accurate price information to be able to make informed purchase decisions and compare prices between traders. The legislation in force goes some way towards ensuring that this happens.

Regulations made under this act require price indications to be given, and the Consumer Protection Act 1987 creates an offense of giving a misleading price indication (Part III, s.20). The Price Indications (Food and Drink on Premises) Order 1979 requires written price indications to be given in premises supplying food and drink for consumption on those premises. The Price Marking Order 1999 requires written indications of unit prices, e.g., price per 100 g, as well as the selling price, to allow for comparisons to be made by consumers. Finally, in the area of pricing, the Price indications (Method of Payment) Regulations 1991 require it to be made clear to consumers if there is a charge made for different methods of payment, e.g., a surcharge for using a credit card.

Loosely connected to the area of pricing is the issue of buying on credit. This may be relevant, for example, if consumers buy food hampers on credit terms or pay for a restaurant meal using a credit card. This area is covered by the Consumer Credit Act 1974 and various detailed regulations. Provisions are both criminal and civil in nature requiring certain information in advertisements (Consumer Credit (Advertisements) Regulations 1989) and also in the documentation containing the agreement (s.61 and the Consumer Credit (Agreements) Regulations 1983). There are also important provisions imposing liability on the creditor as well as the supplier in certain circumstances (sections 56 and 75). This could be particularly relevant in relation to some of the civil provisions discussed below.

The Trade Marks Act 1994

This act implements Directive 89/104/EEC and allows traders to register their trade mark to prevent it being used by anyone else. This is very important in the food industry where brand loyalty is a big issue. Any discrete sign capable of being represented graphically and which is distinguishable may be registered as a trade mark (s.1(1)). This may be the shape of a container, or a sound, or a smell. Company names can also be trade marks, e.g., Coca-Cola and Kelloggs. Although the provisions are mainly about traders protecting their brands and their products, it could be argued that there is a consumer protection angle in that consumers come to expect a certain quality and

safety from a particular trader, which they may not receive from someone using the trade mark without the authorization of the holder. The act contains both criminal and civil provisions, and in addition there is a civil tort of 'passing off' which 'is concerned with the protection of marks that identify the goods as emanating from a particular trade source and yet have not necessarily been registered as trade marks.' An example is the case of *J Bollinger v Costa Brava Wine Co Ltd (1961) RPC 116*, in which the latter company was prevented from calling sparkling wine (now known as Cava) 'Spanish Champagne' as it did not originate from the Champagne region of France (Edenborough M (1995) *Intellectual Property Law*, Cavendish Publishing Ltd.)

Reasonable Precautions and Due Diligence

The vast majority of the offenses described above are strict liability. This could be considered necessary to protect the consumer thoroughly, particularly when it comes to issues such as safety. However, it does seem a little unreasonable to attach criminal responsibility to someone when they have done all they can to try and avoid committing the offenses in the first place. Thus, to counteract the strict liability nature of the offenses, most of them will be subject to a reasonable precautions and due diligence defense (s.21(1) Food Safety Act 1990). The exact wording and nature of the defense provision will vary between the acts, but broadly speaking, they are based on the trader being able to demonstrate that they have set up a system (reasonable precautions) and ensuring that the system works (due diligence). There is much case law in this area illustrating what the courts will accept as satisfactory, although it must be remembered that this is a question of fact, and every case will turn on its own circumstances. However, generally speaking there are a number of principles which can be extracted from the cases as guidelines for the setting up and execution of a suitable system. This will include areas such as staff training (*Knowsley MBC v Cowan (1991) 156 JP 45*), sampling (*P & M Supplies (Essex) Ltd v Devon (1991) MR October 20*), documented checks (*Ellis v Price & Company Bakeries Ltd (1968) MR Aug 183*) and specific warranties from suppliers (*Amos v Melcon (Frozen Foods) Ltd (1985) 149 JP 712*). It is important to carry out some form of risk assessment, identifying where problems are likely to occur and putting in place systems to deal with them. This is known as Hazard Analysis of Critical Control Points and forms the basis of any worthwhile quality-assurance program.

Any system must be preventative, all-encompassing, open to review and amendment, and under the direction of the controlling minds of the company. Although, obviously, the resources of the company will be taken into account, what is necessary for a large retail or manufacturing organization is not necessarily appropriate for a small company (*Garrett v Boots Chemists Ltd (1980) (unreported)*).

Civil Provisions

Civil law deals with situations between affected parties, as opposed to issues affecting the public at large. It is concerned more with remedies, usually monetary, for damage or injury rather than punishment. The tort of 'passing off' has already been discussed, and there are several more areas which would be relevant for food. First, the law of negligence is also a tort and is one of the few areas which is fault-based. The person making a claim is not restricted to the purchaser. Second, the law of contract governs, amongst other things, the terms which may be implied into a contract. The nature of a contract is such that usually only the purchaser will have a remedy. Third, the area of product liability aims to make the producer of a product liable for damage it causes if it is deemed to be a defective product. As with negligence, this is not restricted to purchasers.

Contract Law

The Sale of Goods Act 1979 is the principal piece of legislation which implies terms into contracts. There are in fact several pieces of legislation which imply similar terms depending on the type of contract formed. The references given are for the Sale of Goods Act 1979. The common terms which are implied into consumer contracts are that the goods, including food, will be fit for their purpose (s.14(3)) (*Frost v Aylesbury Dairy Co* mentioned earlier), will be of satisfactory quality (s.14(2)), will conform to their description (s.13(1)) or sample (s.15(2)), and can be sold (s.12(2)). A further consumer protection measure is contained in the Unfair Contract Terms Act 1977, which prohibits the exclusion of these implied terms in consumer contracts by virtue of section 6(2).

Contracts for consumer services must be carried out with reasonable care and skill, within a reasonable time and for a reasonable price. These three terms are implied into consumer contracts by virtue of sections 13, 14, and 15 of the Supply of Goods and Services Act 1982.

Law of Negligence

This can be best explained by the case of *Donoghue v Stevenson (1932) AC 562*, which revolved around a decomposing snail found in a bottle of beer, not by the purchaser but by a friend for whom it had been

bought. There was therefore no contractual liability by the retailer to the ultimate consumer. The principle of law established, which can make the producer liable to the ultimate consumer, whether or not the purchaser, can be found in the well-quoted passage from Lord Atkin's speech:

> A manufacturer of products, which he sells in such a form as to show that he intends them to reach the ultimate consumer in the form in which they left him with no reasonable possibility of intermediate examination, and with the knowledge that the absence of reasonable care in the preparation or putting up of the products will result in an injury to the consumer's life or property, owes a duty to the consumer to take that reasonable care.

The consumer of course has the burden of proving the lack of care and also that the damage suffered was a result of the lack of care.

Product Liability

If goods are defective, they may now come under the remit of Part I of the Consumer Protection Act 1987. This makes a producer (s.1(2)) liable for the damage caused by a defective product (s.2(1)). Damage to the actual item cannot be claimed, although damage to other consumer items is recoverable, as is personal injury or loss, providing they are in excess of £275 (s.5). This legislation can be used to take action in relation to food-poisoning incidents, without the need to prove negligence and where the ultimate consumer was not the purchaser and therefore may not have contractual rights.

Conclusion

It can be seen that there is a significant amount of legislation which deals with consumer protection issues. Only the main areas have been indicated, and there are many more, e.g., the Misrepresentation Act 1967 and the Fair Trading Act 1973 Furthermore, none of the areas mentioned have been covered in any depth, and to appreciate the full application of the law and the way in which it is interpreted, it is necessary to read the actual legislation and cases pertaining to it as well as some texts which comment on it. It should be emphasized that prosecutions under the criminal provisions are often seen as a last resort. Officials are keen to prevent offenses being committed in the first place and therefore offer advice and guidance to traders to help them understand their obligations to consumers. Similarly, many potential civil problems can be sorted out by negotiation, mediation, and arbitration without the need to sue a trader. It is important to note that the achievement of consumer protection aims can often be met, in both criminal and civil areas, without recourse to the courts.

See also: **European Union**: European Food Law Harmonization; **Food Labeling (Labelling)**: Applications; **Food Safety**; **Hazard Analysis Critical Control Point**; **Irradiation of Foods**: Legal and Consumer Aspects; **Legislation**: History; International Standards; Additives; Contaminants and Adulterants; Codex; **Quality Assurance and Quality Control**; **Risk Assessment**; **Weighing**

Further Reading

Butterworth's Trading, Consumer and Food Safety Service. London: Butterworths.

Harvey BW and Parry DL (2000) *The Law of Consumer Protection and Fair Trading*. London: Butterworths.

Howells GG and Weatherill S (1995) *Consumer Protection Law*. Aldershot, UK: Dartmouth.

Lasok KPE (2001) *Law and Institutions of the EEC*. London: Butterworths.

Miller CJ, Harvey B and Parry DL (1998) *Consumer and Trading Law: Text, cases and Materials*. Oxford: Oxford University Press.

Oughton D and Lowry J (1997) *Consumer Law*. London: Blackstone Press.

Scott C and Black J (2000) *Cranston's Consumer's and the Law*. London: Butterworths.

Thompson K (1996) *The Law of Food and Drink*. Crayford, UK: Shaw & Sons.

Contact Plate Freezing *See* **Freezing**: Principles; Operations; Blast and Plate Freezing; Cryogenic Freezing; Storage of Frozen Foods; Structural and Flavor (Flavour) Changes; Nutritional Value of Frozen Foods

CONTAMINATION OF FOOD

E Todd, Michigan State University, East Lansing, MI, USA

Introduction

Contaminants in foods can be grouped according to their origin and nature. Essentially, these are microbiological (bacteria, viruses, parasites), extraneous matter (biological, chemical, physical), natural toxins (seafood toxins, mycotoxins), other chemical compounds (pesticides, toxic metals, lubricants, veterinary drug residues, undesirable fermentation products, radionuclides), packaging materials, and poisons introduced through tampering. Most of the agents found in food are natural contaminants from environmental sources, but some are deliberate additives, including those used for tampering. Cooking can increase or decrease the toxicity of some of these contaminants but has little effect on most chemicals. Consumers have specific concerns about food contaminants but often lack the means to make appropriate judgments on what is high risk and what is not. Additives were at one time a major concern, but today, microbiological issues are the greatest, followed by pesticide and animal drug residues and antimicrobial drug resistance.

Microorganisms

Foodborne Disease Bacteria

Bacteria that cause foodborne disease occur worldwide, the most common being *Salmonella, Staphylococcus aureus, Clostridium perfringens*, and *Bacillus cereus*. These bacteria can multiply rapidly in moist, warm, protein-rich foods, such as meat, poultry, fish, shellfish, milk, eggs, and most food after it has been processed. Infectious organisms such as *Salmonella* and *C. perfingens* can multiply in the digestive tract and cause illness by invasion of the cell lining, production of toxins, or both. Other organisms produce enterotoxins, e.g., *S. aureus* and *B. cereus*, or neurotoxins, e.g., *C. botulinum*, in the food during their growth and metabolism. The staphylococcal and *Bacillus* enterotoxins are heat-resistant, as are the spores of *Clostridium* and *Bacillus* species. Spores, therefore, can survive cooking and germinate into vegetative cells which can grow in the food if it is left at room temperature. Three other pathogens, *Listeria monocytogenes, Vibrio vulnificus* and *Escherichia coli* O157:H7, have caused severe illnesses leading to deaths. *Listeria* is widespread in the environment and is associated with processed food that has become contaminated. Most illnesses have resulted from consumption of contaminated smoked fish, delicatessen meat products, milk and soft cheese. *Escherichia coli* O157:H7 has been responsible for meat (mainly hamburger)-associated outbreaks, sometimes causing hemorrhagic colitis and hemolytic uremic syndrome. Cattle appear to be a major source of this organism. However, illnesses have resulted from contaminated sprouts, lettuce, unpasteurized apple juice, unpasteurized milk, yogurt, cheese, and water. In the last decade, *Salmonella enteritidis* has been the source of many outbreaks involving egg products through transovarian infection from hen to egg. In the last few years, *S. typhimurium* DT 104 and other antibiotic-resistant salmonellae have also recently become a concern. *Vibrio* spp. naturally occur in seawater, and outbreaks have been associated with molluscan shellfish, especially when the ambient temperatures have been high enough in the water to allow rapid growth of the organisms within the molluscs. More foodborne illnesses are caused by bacteria than any other group of agents. There are several specific qualitative detection methods for most pathogens, but their detection in food is usually more difficult than in clinical specimens. Quantitative methods exist for only some pathogens. Refer to individual organisms; Food Poisoning, Tracing Origins and Testing; Food Poisoning, Statistics. (*See* **Gums**: Food Uses; Dietary Importance.)

Viruses and Prions

Viruses are difficult to detect in food, and historically, few viral foodborne diseases have been confirmed, but these agents are probably major causes of foodborne and waterborne disease from epidemiological investigations. Hepatitis A virus and Norwalk-like agents (within the small round structured virus group) are typical foodborne enteric viruses. Enteric viruses typically arise from infected food handlers or sewage. Shellfish can accumulate viral particles if they are present in the water and cause infections following their ingestion. Although the foot-and-mouth disease virus is not a human pathogen, it can be transmitted through food to animals, as is believed to have occurred in the large UK outbreak in 2001. (*See* **Viruses**.) Another area of concern is the transmissible spongiform encephalopathies caused by

prions, notably the bovine form, which has been responsible for over 100 deaths in the UK. Since there is no control of the disease, eradication of infected animals, including sheep with scrapie, and removal of bovine brain and spinal tissues from the food chain are currently being pursued in Europe.

Parasites

A limited number of parasites are foodborne (waterborne or person-to-person spread is more frequent), and illnesses are usually mild. *Trichinella* is found in garbage-fed pork and wild carnivores, including bears and walruses, which are eaten by hunters. *Taenia saginata* and *T. solium* are widespread in beef or pigs, respectively, and up to one-third of human infections of *Toxoplasma gondii* arise from contaminated meat. Since *Cryptosporidium* can cause infections in humans and animals, and the parasite can be spread through sewage and cattle feces into water systems. It is resistant to standard chlorination procedures and needs to be filtered out during water treatment. Over half of the tested public water supplies that use surface water have been found to have small amounts of *Cryptosporidium* in the potable water supply. Produce sprayed or washed with contaminated water or ice made with it are potential vehicles for this agent. People at risk for severe cryptosporidiosis include people with weakened immune systems, including AIDS patients and those who take drugs that suppress the immune system. *Cyclospora* is another coccidian parasite associated with consumption of water, but also produce such as raspberries, basil, and lettuce. Other parasites, such as *Anisakis, Pseudoterranova*, and *Diphyllobothrium* are relatively frequent in fish but rarely invade gastrointestinal tissues to cause serious illnesses. All parasites can be destroyed by thorough cooking of raw meat or fish, or freezing for several weeks. (*See* **Parasites**: Occurrence and Detection; Illness and Treatment.)

Natural Toxicants

Seafood Toxins

Seafood may contain natural toxins that are of concern in many parts of the world. These are paralytic shellfish poison (saxitoxins), diarrhetic shellfish poison (okadaic acid and related compounds), neurotoxic shellfish poison (brevetoxin), amnesic shellfish poison (domoic acid), pufferfish poison (tetrodotoxin), ciguatera poison (ciguatoxin and other toxins), and scombroid poison (histamine). All but the last are accumulated in shellfish or fish through the food chain from dinoflagellates, diatoms, or bacteria. All these organisms are phytoplanktonic or benthic forms of life, naturally occurring in specific areas of the marine environment. Scombroid poison is produced through the decarboxylation of the histidine to histamine in certain fish, notably tuna, mackerel, mahi mahi, and marlin, through bacterial spoilage. All these seafood toxins are resistant to normal cooking practices and are not detectable by organoleptic means. Shellfish in potentially toxic areas can be controlled by preventing harvesting until the causative dinoflagellate and diatom blooms have diminished. However, because of their sporadic occurrence and difficulty in detection, it is more difficult to prevent toxic fish from reaching the consumer. Except for histamine, there are few chemical analytical tests for regulatory purposes, and for most toxins bioassays are used. (*See* **Shellfish**: Contamination and Spoilage of Molluscs and Crustaceans.)

Mycotoxins

Mycotoxins are widely produced on moist crops by a variety of fungi. Poultry and livestock have been poisoned, with both acute and chronic toxic effects. However, acute poisonings, including deaths, have rarely been documented for human beings; examples of potential problems include ergotism from *Claviceps* growth on rye, alimentary toxic aleukia from *Fusarium* on cereal grains, and aflatoxicosis from *Aspergillus* contamination of maize and other grains. Less conclusive are links to chronic illness from aflatoxins in peanuts, other nuts and grain crops, citreoviridin in moldy rice, fumonisins, ochratoxin, and zearalenone in grain, trichothecenes (vomitoxin) from winter wheat, and patulin from apple juice, although epidemiological data would support a link between liver cancer and aflatoxin in parts of China. Milk and meat from animals feeding on moldy crops may contain mycotoxins and are potential sources of carcinogens and teratogens. There are chemical and analytical methods for most mold toxins, and some crops are monitored and advisory levels set for specific toxins. There is also an indication that mild gastroenteritis can arise from beer and soft drinks in containers contaminated with adventitious molds penetrating through small holes, e.g., punctured cans and faulty bottle tops. These have not been thoroughly investigated. (*See* **Mycotoxins**: Occurrence and Determination.)

Plant Toxins

Several naturally occurring secondary metabolites from plants can cause illness, e.g., neurolathryism from α-aminoproprionic acid in *Lathyrus* species (wild peas), cyanide poisoning from cyanogenic glycosides in almonds, cherries, etc., solanine poisoning from α-solanine in potatoes. Novel foods used from

plant species that have not been used for food before need to be evaluated carefully to avoid the potential for adverse effects. (*See* **Plant Antinutritional Factors**: Detoxification.)

Extraneous Matter

Extraneous matter is foreign material entering a food during its production, storage or distribution, and can be (1) biological matter, such as insect parts, rodent excreta, animal hair, mites, nematodes, and mold, (2) chemical compounds, such as oil or tar, or (3) physical materials, such as glass, metal, and stones. These are undesirable from an esthetic point of view, but injuries and gastroenteritis after their consumption have also been reported. In addition, psychosomatic illnesses have been documented from people observing disgusting objects in their food. Manufacturers receive more complaints about extraneous matter than all other food-related problems combined, and usually install metal or glass detectors on plant lines to minimize these types of contamination.

Other Chemical Compounds

Environmental contaminants of food-safety concern to both scientists and consumers include trace elements, organometallic compounds, and organic substances such as halogenated hydrocarbon pesticides. These compounds have several characteristics in common: (1) they tend to be stable and persistent in the environment; (2) they tend to accumulate in the food chain; and (3) they can be biotransformed with increased toxicity. Analytical chemistry has made us aware of the presence of agents in foods that have the potential to cause chronic illness, but the adverse effects generally cannot be directly demonstrated in human populations. However, the causes of most cancers are not known, and the number of cases of stomach, colon, bladder, and kidney cancers that might be expected to reflect exposure to environmental chemicals are not increasing. Testicular cancer and brain cancer, however, may be on the rise but are not necessarily related to chemical contaminants in foods. Some chemicals, including pesticides, have been shown in animal studies to disrupt the endocrine system and affect reproductive development in the fetus; however, it is not known whether consumption of food containing small amounts of chemical contaminants has any effect on the human population.

Pesticides

Pesticides include insecticides, herbicides, fungicides, rodenticides, insect repellants, weed killers, antimicrobials, and algaecides, that are designed to prevent, destroy, repel, or reduce pests (animal, plant, and microbial). Some are selective, impacting only target organisms, whereas others have a broad-range toxicity. These may be applied in solid, liquid, aerosol, or gaseous form. Spraying in one area may result not only in contamination locally but, depending on weather conditions and patterns, contamination up to thousands of kilometers away because particles can be carried into the upper atmosphere. Some residues may remain in both fresh produce and processed foods. Metabolically stable lipophilic pesticides tend to accumulate in fat deposits in exposed farm animals, and this can lead to biomagnification in animals higher up the food chain, including humans, but today, with better control of the use of chemicals in the environment, this is a diminishing concern in developed countries.

Disorders of the nervous system, liver, and kidney and various types of cancer typically have a long latency period. The critical parameters which determine whether or not a pesticide will cause any adverse effect are (1) the intrinsic biochemical properties of the chemical, (2) the level of exposure to the chemical, and (3) the duration of exposure to the chemical. The severity of illness ranges over a wide spectrum of effects, depending on the duration and extent of exposure and on the susceptibility of the person exposed. Chronic pesticide exposures in the parts per billion and trillion range through the diet are thought to pose at most a very low risk. However, there are many organochlorine, organophosphate, carbamate, pyrethrin, and arsenical pesticides that alter the immune system in wildlife and laboratory animals. More recently, more selective chemical and biological pesticides have been developed, e.g., spores of *Bacillus thuringiensis*; these are not known to have any adverse affects against humans or domestic animals. (*See* **Pesticides and Herbicides**: Toxicology.)

Organochlorine Pesticides

Organochlorine insecticides were widely used, and many persist in the environment today. Dichlorodiphenyltrichloroethane (DDT) and cyclodienes, such as aldrin, endrin, and heptachlor, have been used worldwide for over four decades against insects of public and veterinary health significance. Because of their stability, metabolites of these and other organochlorine compounds have now reached all parts of the world's surface. The most toxic insecticides, such as endrin, have also caused acute human illness from contaminated food products, e.g., tortillas, and even deaths from consumption of flour in contact with the insecticide during transportation in Pakistan. (*See* **Pesticides and Herbicides**: Types of Pesticide.)

Although many countries have banned the use of DDT, it is still of value in controlling diseases such as malaria. Milk from cattle reared in areas where DDT was used in agriculture can be heavily contaminated (e.g., over 500 p.p.m.). High levels can also be found in mild products, e.g., ghee, butter, cheese, and infant formula. Hexachlorocyclohexanes, e.g., lindane, have also been identified in milk after feed prepared from sprayed crops had been given to cattle, or insecticidal ointment had been applied directly to their skin. Some organochlorine compounds originally unconnected to agricultural use have accumulated in foods because of their long-term persistence, e.g., fish containing mirex, an ant poison and also a fire retardant in household materials. Dioxin, a byproduct of defoliant Agent Orange, waste incineration, and industrial processes is found widely in the environment, although usually at low concentrations. As a result of an industrial accident 25 years ago in Italy, the long-term impact of dioxin is being examined; one effect is a reduction in the number of male births. In 1999 in Belgium, chickens and other food-producing animals ingesting dioxin-contaminated feed cost the Belgian economy more than $750 million. Motor oil containing the dioxin entered the food chain after being mixed with vegetable oils used in the feed manufacture.

Many organochlorine pesticides have been linked to hormone disruption and reproductive problems in aquatic invertebrates, fish, birds, and mammals. They degrade slowly and, being fat-soluble, accumulate in the food chain, eventually ending up in body fat. The long-term effects of organochlorine compounds on human beings are not yet known.

Organophosphates

Organophosphates are irreversible cholinesterase inhibitors and can cause changes in the immune system. Some are highly toxic and are responsible for more accidental pesticide poisonings than any other class of pesticides. Symptoms may be reversed with atropine treatment. They are relatively inexpensive and control a variety of insects. Examples are parathion, malathion, and diazinon. They are widely used on food crops, in residential and commercial buildings and for ornamental plants and lawn care, which means that many people may be regularly exposed to them. Although they do not persist in the environment, residues have been found in fruits and vegetables. There is no known health risk from consumption of this kind of uncooked food containing low residues over a long period of time.

Carbamates

Carbamates are insecticides, fungicides, herbicides, and nematocides. These act in a similar manner to the organophosphates for insecticides and nematocides by inhibiting acetylcholinesterase, except that the action is more rapidly reversible. The toxicity is generally low, with the exception of a few compounds, particularly aldicarb. They are considered to be nonresidual and generally break down more rapidly than chlorinated hydrocarbons. Aldicarb has been absorbed from the soil by plants with leaves and fruit accumulating the chemical. Watermelons grown on previously contaminated soil, and hydroponically grown cucumbers treated with aldicarb have caused anticholinesterase symptoms in several hundred persons in outbreaks in North America. Symptoms of acute carbamate poisoning are the same as for organophosphate poisoning; abdominal pain, nausea and/or vomiting, diarrhea, muscular weakness, twitching, blurred vision, watery eyes, pinpoint pupils, excess salivation, sweating, seizures, disorientation, and excitation.

Pyrethrins and other Botanicals

Botanicals are substances that are naturally produced by plants that are poisonous to pests that prey on those plants. Pyrethrins are derived from plants and, together with their synthetic analogs, are used as insecticides. Pyrethroids affect the nervous system and are neurotoxic. Insects become disoriented, and body functions fail, but pyrethroids do not affect mammals because they do not penetrate mammalian skin, and they have esterases than can detoxify the chemical. Although these are generally safe for mammals, they can be toxic to aquatic life. Strychnine, neem, and rotenone are examples of other botanicals that can be used against pests.

Other Pesticides

Alkyl mercury salts, used as fungicides on seed grain, have caused human illness either through direct consumption of the grain or through meat from pigs eating the contaminated grain but are no longer used in most countries. Other chemicals considered too dangerous today and banned in many countries are the thallium salts formerly used as rodenticides. Bromide fumigants, such as methylbromide, are used against insects, nematodes, viruses, weeds, and fungi and are applied directly to the soil. Crops planted on these soils, particularly lettuce, celery, tomatoes, cucumbers, and strawberries, may accumulate bromine. When dibromoethane was found in flour and cake mixes arising from fumigation of wheat during storage, its use was banned because of its mutagenic, carcinogenic, and teratogenic properties. Other fumigants are methylene chloride, carbon tetrachloride, chloropicrin, phosphine, and ethylene dibromide

(banned). (*See* **Fumigants**.) Alar, a plant-growth regulator, penetrates fruits, cannot be removed by washing or peeling, and can be converted to products that may be carcinogenic in cooked fruit or products. Controversy that arose regarding its potential carcinogenicity led to its removal from use by the manufacturer.

Pesticide Contaminants

In addition to pesticides themselves, certain impurities in formulations can be toxic. Acute delayed toxic effects, including weight loss and pneumonia, have developed from exposure to trimethylphosphorothioate present in malathion sprays. The potential carcinogen nitrosodimethylamine has been found as a contaminant in certain pesticides. Improved pesticide production practices have since reduced the level of these toxic impurities. (*See* **Nitrosamines**.)

Herbicides

After 50 years of use, 2,4-D of the phenoxy family of herbicides, is the most widely used herbicide worldwide. Its major uses in agriculture are on wheat and small grains, sorghum, corn, rice, sugar cane, soybeans, low-till rangeland, and pasture. It is also used on rights of way, roadsides, noncrop areas, forestry, lawns, turfs, and aquatic weeds. There are moves to restrict its applications, even though no scientifically documented health risks, either acute or chronic, exist from approved uses. Atrazine, a triazine herbicide used to control broadleaf and grassy weeds in corn, sorghum, soybeans, sugarcane, pineapple, Christmas trees, and other crops, also has a low toxicity, but animal studies indicate its potential for endocrine disruption and carcinogenicity. In contrast to most herbicides, paraquat, a bipyridyl contact herbicide, is highly toxic through dermal exposure and ingestion, and must be handled with care. Some weeds are developing herbicide resistance, and either new compounds or new approaches may be required for use of herbicides in the future.

Summary

Several pesticides are considered potentially carcinogenic or teratogenic based on various tests including animal studies. However, human data are limited (except for arsenic and perhaps for the use of the defoliant Agent Orange in Vietnam), and there is no epidemiological evidence that any pesticide is a human mutagen, carcinogen, or teratogen at normal exposure levels. Future work needs to concentrate on mechanisms of pesticide toxicity and metabolism, reevaluation of acceptable limits, and their control in the environment. The use of pesticides will not be eliminated in the future, but their use may be reduced, and safer compounds will replace the more toxic and persistent compounds; unfortunately, DDT and other organochlorines will remain in the environment for many years.

Toxic Metals

The sources of toxic metals are atmospheric contamination, sewage sludge, trace contaminants in animal feed, fertilizers, or pesticides. Contamination may also occur during processing, home cooking, and storage in metal containers. Heat does not reduce the toxicity of these metals substantially.

Cadmium

Cadmium is naturally present at a very low level in the environment, but large concentrations are produced as waste byproducts of electroplating, rust proofing, color tints for paint, plastic stabilizers, and batteries. Effluents also arise from textile, porcelain, and enamel manufacturing. Crops grown in sludge containing cadmium can accumulate the metal up to 13 times the soil concentration. When cadmium salts are released into an aquatic environment, crustaceans and fish accumulate quantities large enough to be hazardous if they are a part of the regular diet. (*See* **Cadmium**: Toxicology.)

Lead

Lead present in rocks and soil may be accumulated in plants and transmitted to humans through the food chain. However, most lead comes from atmospheric aerosols and dust particles through industrial processes and automobile emissions, although lead is less used as a petroleum additive today. Urban areas therefore contain more lead particles than rural regions. These fall on soil and are absorbed by crops. Modern canning uses lacquer, side-seam welds, and seamless two-piece cans to replace lead solders. Acidic foods or beverages stored in ceramic vessels made with glazed lead salts or galvanized pans have accumulated lead. In addition, the alimentary tract can receive lead from potable water transported by lead piping or stored in lead-lined tanks, and through the compulsive nibbling by children of paint chips containing lead, with up to 550 μg per day being accumulated. Lead is a cumulative poison, and chronic exposure may lead to damage to the brain, liver, peripheral and central nervous system, and reproductive, renal, and immune systems. (*See* **Lead**: Toxicology.)

Arsenic

Arsenic is widely distributed throughout the earth and reaches food sources from water through the dissolution of minerals and ores, from industrial effluents, and from atmospheric deposition; concentrations in ground water in some areas are sometimes elevated as a result of erosion from natural sources. The average daily intake of inorganic arsenic in water is estimated to be similar to that from food. Arsenic is used commercially in the production of pharmaceuticals, metals, herbicides, wood preservatives, and sheep dips. Arsenic compounds have occurred in seafood, eggs, and cheese. Chronic poisoning from arsenic is unusual unless it occurs naturally in the water supply. However, inorganic arsenic is a documented human carcinogen, and a relatively high incidence of skin and possibly other cancers that increase with dose and age has been observed in populations ingesting water containing high concentrations of arsenic.

Mercury

Mercury is used in electrodes, the manufacture of plastics, slimecides in the pulp and paper industry, paints, and fungicides. In the environment, the toxic form is methyl mercury, which is a neurotoxin as well as a teratogen. The mercury content of crops may increase with the use of sewage sludge. It is found in fish and shellfish living in water contaminated from industrial wastes and in large lakes artificially created for hydroelectric power from drowned, decaying vegetation. Acid rain also helps to mobilize industrial aerosols into water supplies. People with a regular diet of fish containing high levels of organic mercury, or shellfish containing inorganic mercury, are at greatest risk. (*See* **Mercury**: Toxicology.)

Tin, Copper, and Antimony

Although tin can reach aquatic environments through natural sources, industrial pollution and antifouling paint on boat hulls, there is little evidence that it can cause chronic human illness. However, acid foods in tin-lined cans are able to dissolve enough metal to cause acute gastroenteritis. Both copper and antimony become toxic when they are in contact with acidic liquids. Copper pipes used in soft-drink vending machines, and antimony-lined bowls or cups have been implicated in acute illness. The speed of metal solubilization depends on the acidity of the liquids, temperature and time of exposure.

Fluoride

The beneficial effects of fluoride in drinking water on teeth have been demonstrated, but fluoride can be toxic, and intakes of more than 10 mg per day are not recommended. The total consumption of fluoride from both water and food prepared using water has to be considered in determining intake levels.

Lubricants

Polychlorinated biphenyls and polychlorinated naphthalenes were produced as lubricants and should not be brought into contact with food or feed, but cattle may ingest them by licking the oils from farm machinery or animal feed containing leaked oil. Affected animals develop a debilitating condition, and human ingestion of milk is considered hazardous. Because of their extreme persistence in the environment, PCBs are not only banned from future use but are being removed from aging equipment, e.g., transformers, and are destroyed by incineration. Polybrominated biphenyls have also been accidently fed to livestock with resultant human exposure through milk and meat.

Veterinary Drug Residues

Antibiotics are added to reduce disease and improve the growth of farm animals and aquaculture fish; these are of less concern for their chemical effect and more for their ability to increase antimicrobial resistance in strains that might subsequently infect humans. There is evidence that antimicrobial resistance is increasing worldwide but particularly in developing countries. Steroids and growth hormones are used to increase productivity, e.g., bovine somatotrophin for milk, synchronize breeding cycles, and act as tranquillizers. The human effect of consuming foods with these drugs is still being debated, but many countries refuse to accept products derived from animals given these drugs. Also, some drug residues may be allergenic, e.g., penicillin. The withdrawal times of such drugs are critical to keep the residues in food as low as possible.

Undesirable Fermentation Products

The carcinogen ethyl carbamate is a natural fermentation product in the production of wine, either through reaction of ammonia with the antimicrobial agent diethyl pyrocarbonate or through ethanol with carbamyl phosphate. Limits of 30 μg per kilogram of wine and 400 μg per kilogram of brandy have been set to reduce the level of carbamates for human exposure. There have been over 1500 cases and 20 deaths owing to eosinophilic myalgia syndrome in the USA from tryptophan dietary supplements contaminated with derivatives of L-tryptophan. These contaminants

were thought to have arisen during the fermentation process when a genetically modified strain of *Bacillus amyloliquefaciens* was used to produce the tryptophan. However, some purification steps were left out of the process at the same time, and no satisfactory explanation for the contamination has been made.

Radionuclides

Radionuclides are normally present in the environment at low levels from naturally radioactive rock but can also arise from atmospheric fallout from nuclear tests and gaseous emissions and spills from accidents at nuclear reactors. Food is one possible means of transmitting radionuclides to the human body, with a danger for long-term carcinogenic, teratogenic, or mutagenic effects. Higher doses may also affect reproductive capability and shorten the lifespan. Radium-226 and uranium-238 are naturally occurring radioactive elements that may be at high levels in certain locations because of the underlying rock structure. Cereals are the food items that contribute the most radium-226 to our diet. Fallout from tests or accidents persists longer in arctic zones, and persons eating an exclusive meat diet, such as Lapps and the Inuit, are at highest risk from lead-210 and polonium-210. Milk and meat supply most of the radionuclides, such as strontium-89, strontium-90, iodine-131, and cesium-137, to our bodies. The direct effects of the Chernobyl nuclear reactor explosion in Ukraine in the immediate area of the reactor were extreme, but worldwide contamination was not as high as was originally feared and had no lasting impact on the food supply. (*See* **Radioactivity in Food**.)

Packaging and Storage

Packaging can exist in various forms, from crates to wraps. Wood preservatives, such as phenols or copper arsenates, and insecticides against termites, used in crates, can contaminate food during shipping through direct contact or vapors. Wood shavings used for packing can also transfer these chemicals to foods, or to animals if the shavings are used for bedding. In addition, food stored in rooms with walls covered with insecticidal paint, or containing pest strips from which pesticides evaporate over a long period of time, may become exposed to the chemical.

Paper, cardboard, plastic wrap, or aluminum foil may not keep out lindane vapors if not properly wrapped. Phosphoric acid esters as flame retardants in paper, adhesives, and plastic wrappers, e.g., bread bags and margarine or butter wraps, can reach food. Diffusion through these is dependent on the time of contact and type of food wrapped (lipid-rich or moist foods absorb better than other foods). Hermetically sealed containers protect food from all chemical and microbial contaminants if the process is carried out correctly.

Tampering Problems

For centuries, food has been adulterated for fraudulent purposes or deliberately contaminated for injuring or killing specific persons. Recently, commercially produced food has been tampered with for a variety of reasons, including product extortion to obtain money from a company or life insurance firm, attempts to ruin a company by a dissatisfied customer or disgruntled employee, attention in the media for a political cause, or as a joke to observe the reaction of those affected. In dioxin contamination of feed, there may be some situations where this is deliberately done to use waste oils mixed with vegetable oils. Tamper-resistant and tamper-evident packaging has reduced, but not eliminated, these problems. Nowadays the possibility of contamination of the food chain, water supplies or even retail food is a real threat of bio terrorism. (*See* **Adulteration of Foods**: History and Occurrence; Detection.)

See also: **Adulteration of Foods**: History and Occurrence; Detection; **Food Poisoning**: Tracing Origins and Testing; Statistics; **Fumigants**; **Lead**: Toxicology; **Mercury**: Toxicology; **Mycotoxins**: Occurrence and Determination; **Nitrosamines**; **Parasites**: Occurrence and Detection; Illness and Treatment; **Pesticides and Herbicides**: Types of Pesticide; Toxicology; **Radioactivity in Food**; **Viruses**

Further Reading

Colburn T, vom Saal F and Short P (eds) (1998) Endocrine-disrupting chemicals: neural, endocrine and behavioral effects. *Toxicology and Industrial Health* 14: 1–360.

Cooper RL and Kavelock RJ (1997) Endocrine disrupters and reproductive development: a weight-of-evidence overview. *Journal of Endocrinology* 152: 159–166.

Crawford LM and Franco DA (eds) (1994) *Animal Drugs and Human Health*. Lancaster, PA: Techomic.

de Vries J (ed.) (1996) *Food Safety and Toxicology*. Boca Raton, FL: CRC Press.

Doyle MP, Beuchat LR and Montville TJ (eds) (1997) *Food Microbiology: Fundamentals and Frontiers*. Washington, DC: ASM Press.

Farber JM and Todd ECD (eds) (2000) *Safe Handling of Foods*. New York: Marcel Dekker.

Fong WG, Moye HA, Seiber JN and Toth JP (eds) (1999) *Pesticide Residues in Food: Methods, Techniques and Regulations*. New York: John Wiley.

Harvey PW, Rush KC and Cockburn A (eds) (1999) *Endocrine and Hormonal Toxicity*. New York: John Wiley.

Ioannides C (ed.) (1998) *Nutrition and Chemical Toxicity*. Chichester, UK: John Wiley.

Jackson LS, Knize MG and Morgan JN (eds) (1999) *Impact of Processing on Food Safety*. New York: Kluwer Academic/Plenum.

Moffat CF and Whittle KJ (eds) (1999) *Environmental Contaminants in Food*. Sheffield, UK/Boca Raton, FL: Sheffield Academic Press/CRC Press.

van der Heijden K, Younes M, Fishbein L and Miller S (eds) (1999) *International Food Safety Handbook*. New York: Marcel Dekker.

Waston D (ed.) (1993) *Safety of Chemicals in Food: Chemical Contaminants*. Chichester, UK: Ellis Horwood.

Waston D (ed.) (1998) *Natural Toxicants in Food*. Sheffield, UK/Boca Raton, FL: Sheffield Academic Press/CRC Press.

CONTRACEPTIVES: NUTRITIONAL ASPECTS

F Ford, University of Sheffield, Section of Reproductive Medicine, Sheffield, UK

Background

In 1998, the Royal College of General Practitioners (RCGP) celebrated the 30th anniversary of its unique and world-renowned study into the safety and effects of the oral contraceptive (OC) pill on women's health. The study recruited 1400 GPs and 47 000 volunteer women throughout the UK, half of whom were using OCs, the other half of whom had never done so. It has become one of the largest detailed investigations into the pill worldwide and has highlighted many important and significant links and risk factors between the oral contraceptive and a variety of medical conditions.

Diet, Oral Contraceptives, and Cardiovascular Disease

The RCGP study was among the first to show that the risk of cardiovascular disease among pill users is concentrated mainly in users who smoke, especially in older women. The study was also the first to demonstrate a link between the risk of hypertension or arterial disease and the progestogen content of the pill. Most excess cardiovascular disease among users of OCs is due to thrombosis (not atherosclerosis); studies indicate that the lower the OC estrogen dose, the lower this risk. OCs containing the third-generation progestins desogestrel and gestodene have been associated with greater risks of venous thromboembolism than are associated with older progestins, although there is some controversy surrounding these findings.

OC users have lower dietary carotenoid intakes than nonusers. Since OC users smoke and drink more than nonusers, and both factors are associated with lower carotenoid intakes, OC users form a potential high-risk group for disease.

Conflicting reports regarding the possible effect of OCs on folate status prompted researchers to evaluate the relationship between dietary folate intake and the concentration of folate in serum and erythrocytes among users and nonusers of OCs. The authors concluded that the use of OCs produces significantly lower serum folate levels during the first week of the menstrual cycle in spite of adequate folate intake.

Recent epidemiologic studies have shown an increased mortality from cardiovascular diseases in people with higher serum copper levels. Even though a higher serum copper concentration in women using OCs is well known, there is still uncertainty about the influence of newer progestin compounds in OCs on serum copper concentration. This issue is of particular interest in the light of recent findings of an increased risk of venous thromboembolism in users of OCs containing newer progestins like desogestrel compared to users of other OCs. While an elevated serum copper concentration was found in users of all types of OCs, the elevation was more pronounced among women taking OCs with antiandrogen effective progestins such as third-generation OCs containing desogestrel. Further investigation is required to shed light on the possible role of high serum copper concentration in increasing the cardiovascular or thrombotic risk of women using OCs.

Oral Contraceptives and Weight Gain

Although weight gain is frequently reported, as a side-effect of OC use, there is little empirical evidence that such weight gain actually occurs. Published studies on the effects of OCs have shown that long-term use is not associated with increases in weight. Despite these findings, it is a common perception among women that oral and other hormonal contraceptives cause weight gain. In 1995, more than one million American women used Depo-Provera, and more than 500 000 used Norplant. Clearly, it is important to determine whether the use of these drugs can be expected to promote weight gain in women.

Preliminary reports suggested that users' perceptions of weight gain may be justified for some newer contraceptive drugs. Implanted (Norplant) and injected (Depo-Provera) forms of progestin have been found to lead to increases in appetite and body weight.

The mechanisms by which contraceptive hormones may affect body weight are not known. Numerous studies have shown that energy intake and expenditure are altered across phases of the menstrual cycle. Few researchers have examined how these cyclic changes are affected when a contraceptive drug suppresses ovulation. A placebo-controlled experiment showed that, although the phase of the menstrual cycle affected both energy intake and resting energy expenditure, suppression of the cycle with depot medroxyprogesterone acetate did not cause short-term changes in energy intake or expenditure, or cause weight gain, in young healthy women. Further research is needed to determine whether similar effects occur with different contraceptive hormones, over longer periods of time, and in other populations, such as overweight or obese women.

See also: **Carotenoids**: Occurrence, Properties, and Determination; **Copper**: Properties and Determination; **Coronary Heart Disease**: Etiology and Risk Factor; **Energy**: Intake and Energy Requirements; **Folic Acid**: Properties and Determination; **Obesity**: Etiology and Diagnosis

Further Reading

Carranza-Lira S and Bueno Fontal JP (2000) Modification of corporal weight, body fat distribution, blood lipids and glucose levels in oral contraceptive users. *Medical Science Monitor* 6(5): 941–944.

Eck LH, Bennett AG, Egan BM *et al.* (1997) Differences in macronutrient selections in users and nonusers of an oral contraceptive. *American Journal of Clinical Nutrition* 65(2): 419–424.

McLean JA, Barr SI and Prior JC (2001) Cognitive dietary restraint is associated with higher urinary cortisol excretion in healthy premenopausal women. *American Journal of Clinical Nutrition* 73(1): 7–12.

Masse PG, Van den Berg H, Livingstone MM, Duguay C and Beaulieu G (1998) Nutritional and psychological status of young women after a short-term use of a triphasic contraceptive steroid preparation. *International Journal for Vitamin & Nutrition Research* 68(3): 203–207.

Pelkman CL, Chow M, Heinbach RA and Rolls BJ (2001) Short-term effects of a progestational contraceptive drug on food intake, resting energy expenditure, and body weight in young women. *American Journal of Clinical Nutrition* 73(1): 19–26.

CONTROLLED-ATMOSPHERE STORAGE

Contents

Applications for Bulk Storage of Foodstuffs

J Saxena, Banasthali Vidyapith, Rajasthan, India
H Gourama, Pennsylvania State University, Reading, PA, USA
S L Cuppett and L B Bullerman, University of Nebraska-Lincoln, Lincoln, NE, USA

Background

Planting and harvesting of fruits and vegetables are determined by many factors including season of the year, the cultivars, climatological conditions, and the type of soil. In addition, since the majority of fruit and vegetable consumers are often located far from growing areas, most crops must be stored then transported to marketplaces if demand for them is to be satisfied throughout the year and throughout the world. Because fruits and vegetables are essential sources of vitamins, minerals, fiber, and other nutrients, consumers are demanding a larger and more consistent supply of these products. Controlled atmosphere (CA) storage of fruits and vegetables not only increases the shelf-life of these products and maintains good quality but also allows for a consistent year-round supply in the marketplace. CA storage has been shown to lengthen the shelf-life of fruits and vegetables by controlling their levels of transpiration, respiration, and diseases.

Control of Undesirable Plant Processes

The quality of fruits and vegetables after harvest depends upon whether they were harvested at the right maturity stage, and if the metabolic activities after harvest are controlled. Some of the physiological activities of fruits and vegetables such as respiration and transpiration continue after harvest. However, other metabolic activities of these plant products such as photosynthesis and water absorption through the roots stop at harvest. All of these changes create metabolic stresses on fruits and vegetables that shorten the storage life of these crops and reduce the nutritional and organoleptic quality. Optimal storage conditions after harvest can control the loss of quality of these products by maintaining metabolic homeostasis and by minimizing processes like transpiration, respiration, sprouting, and elongation of existing structures, seed germination, and greening.

Control of Transpiration

The loss of moisture after harvest can lead to many changes in the texture and structure of plant products such as shriveling, shrinkage, and product loss. Thus, minimizing the amount of water loss during the post-harvest period is critical for maintaining good quality and freshness of these products. Transpiration rates depend not only on the type of plant product but also on the humidity and temperature of the external environment. Reduction of the rate of transpiration in these food products can be achieved by maintaining low temperatures, high relative humidities, and small vapor pressure differences during storage.

Control of Respiration

Respiration in plant products is defined as a physiological process by which plant tissues exchange gases with the surrounding environment. The principal outcome of this physiological process is the metabolic transformation of carbohydrates such as starch and oxygen to carbon dioxide and water through glycolysis, the Krebs cycle, and the electron transport chain. Respiration, if allowed to occur to any extent, produces heat, thus affecting the temperature of the product. Generally it is known that fruits and vegetables with high respiration rates such as broccoli, sweet corn, mushroom, and asparagus are perishable, whereas products with low respiration rates are stable during storage. The respiration rate of fruits and vegetables is influenced by oxygen level, temperature, carbon dioxide, ethylene, and physical stress. Control of temperature by refrigeration is critical in reducing the rate of respiration. However, many products, such as melons, peppers, cucumbers, etc., cannot be stored for a long time under refrigeration because of their susceptibility to chill injury. Additional control measures such as appropriate packaging and transportation, modified atmosphere (MA) storage, and CA storage of fruits and vegetables should be implemented.

Pioneering work has led to the development of CA storage, an effective process that can extend the shelf-life of fruits and vegetables by controlling the gases within the storage environment. CA and MA are basically processes that change the normal composition of the air (78% N_2, 21% O_2, 0.03% CO_2) surrounding fruits and vegetables. CA storage is usually used in conjunction with refrigeration. The first studies regarding fruit preservation in a gaseous medium were carried out in an atmosphere consisting of more carbon dioxide and less oxygen than is normally present in the air. The benefits of CA and MA include slowing down the respiration rate, sensitivity to ethylene, and, consequently, the ripening process, inhibition of microbial growth, and control of insects. The exact gas composition will depend on the type of product, temperature, and length of treatment. CA and MA have been proven to be beneficial for vegetables such as cauliflower, asparagus, brussels sprouts, celery, mushrooms, cabbage, lettuce, peppers, broccoli, and other vegetables. Among vegetables, artichokes may be held in 2–3% carbon dioxide and 2–3% oxygen, asparagus in 5–10% carbon dioxide and atmospheric oxygen (air), broccoli in 5–10% carbon dioxide and 1–2% oxygen, cabbage in 3–6% carbon dioxide and 2–3% oxygen, okra in 3–5% oxygen and 0% carbon dioxide at 8–12 °C, radishes in 1–2% oxygen, 0% carbon dioxide at 0–5 °C, mature tomatoes in 0–3% carbon dioxide and 3–5% oxygen at 12–20 °C, and partially ripe tomatoes in 0–5% carbon dioxide and 3–5% oxygen at 8–12 °C. Generally, the relative humidity for the CA of vegetables varies from 90 to 98%. CA storage was also found to reduce chilling injuries to cucumbers.

CA storage has also been proven to be good for maintaining the quality and extending the shelf-life of fruits. CA storage has been shown to improve the chilling sensitivity, retarding the ripening and senescence and controlling microbial and insect attacks. One of the principal commercial fruits with which CA storage has been used successfully is apples. Generally, apples are stored in 1–3% oxygen and 1–5% carbon dioxide at 0–5 °C, depending on the variety. These apples are typically harvested before or at the beginning of the increase in the respiration rate. During CA storage, apples can go through many chemical and physiological disorders, such as development of bitter pit, corky lesions, and superficial

scald. CA storage of Golden Delicious apples at low temperatures and low oxygen atmospheres produces the highest amounts of volatile aroma compounds such as branched-chain esters. Low-oxygen atmospheres have also been found to be beneficial for fruit firmness, acidity, and color. It has been reported that there are many advantages in storing apples in CA for a long period of time (up to 3 years) over normal air storage. These advantages include maintenance of firmer fruits, good skin color, higher soluble solids, and low pH. It has also been found that the development of flesh browning after short exposure of 'Fuji' apples to 20-kPa carbon dioxide is a good prediction of CA-induced flesh browning. CA storage has also been shown to extend the shelf-life of green bananas by storing them in 2–5% oxygen and 2–5% carbon dioxide at an optimal temperature of 13 °C. It has been reported that equal amounts of oxygen and carbon dioxide (5%) at 11.7 °C are suitable for Gros Michel bananas held for 20 days. Lacatan and Dwarf Cavendish bananas have been effectively stored for 3 weeks using CA storage conditions of 6–8% carbon dioxide and 2% oxygen at 15–15.5 °C. A mixture of 5–10% oxygen and 0–5% carbon dioxide is generally effective for oranges, 5–10% oxygen and 0–10% carbon dioxide for lemons, 1–2% oxygen and 3–5% carbon dioxide for peaches, 1–2% oxygen and 0–5% carbon dioxide for plums and prunes, 5–10% oxygen, 15–20% carbon dioxide for strawberries, and 3–5% oxygen and 5–10% carbon dioxide for mangos. The temperature range for CA storage of fruits varies from 0 to 5 °C, and for tropical fruits, the temperatures are higher, from 5 to 15 °C.

Control of Ripening

Prior to senescence, fruits go through a ripening stage, which consists of biochemical reactions that transform fruits into sweet, soft products with good aromas. The color also changes from green to yellow, orange, and/or red. Fruits can be divided into categories, based on their ripening physiology. Some fruits, such as apples, tomatoes and bananas, are referred to as climacteric, whereas other fruits, such as oranges, grapes and pineapples, are nonclimacteric. Kidd and West were the first to establish the relationship between decreasing oxygen and raising carbon dioxide and the onset of climacteric changes in apples and pears. These two types of fruits are distinguished by their response to the gas ethylene. Climacteric fruits, when exposed to ethylene, exhibit a respiration surge/climax, which is coincident with their ripening, whereas nonclimacteric fruits do not have a respiration climax and show no ripening changes. Also, during ripening, climacteric fruits produce and release ethylene into the atmosphere. Ripening in climacteric fruits hastens their transition into senescence and, therefore, shortens their storage life. Thus, control of the level of ethylene in the environment is very important for extending the storage life of climacteric fruits. (*See* **Ripening of Fruit**.)

Ethylene production in climacteric fruits has an important physiological role in seed germination, leaf abscission, ripening of fruits and vegetables, and the yellowing of leaves. Methionine is the primary precursor of ethylene biosynthesis. Methionine in the presence of ATP transforms into *S*-adenosylmethionine, which then transforms into aminocylpropane carboxylate (ACC) and ethylene. ACC synthase and ACC oxidase are key regulatory enzymes involved in the biosynthesis of ethylene. Ethylene's biosynthesis and physiological function depend on the oxygen concentration. Apples stored at 3% oxygen generate up to three times less ethylene than the apples stored in the air. Golden Delicious apples stored at low oxygen and at low temperature receive the highest sensory score after 5 months of storage. Maintenance of ethylene concentration of less than $1\,\mu l\,l^{-1}$ in CA storage delays fruit softening and other ripening changes, including the development of disorders, such as superficial scald and bitter pit in apples, both of which are also markedly influenced by preharvest factors and maturity. It is clear that the potential for disorder development, which would be readily manifest in air or refrigerated storage, can be modified significantly by storage in conditions that induce significantly reduced ethylene production rates and/or prevent build-up of ethylene in the storage atmosphere.

High concentrations of carbon dioxide suppress the production of ethylene. A number of investigations have shown that treatment of fruits with elevated levels of carbon dioxide leads to an accumulation of ACC. The inhibition of ethylene formation may be due to the inactivation of ACC oxidase, the enzyme that catalyzes the reaction of ACC to ethylene. However, small concentrations of carbon dioxide (0.1% or lower) have been found to be stimulatory to the production of ethylene. Reducing oxygen levels from 2 to 1%, with 1% carbon dioxide at 1 °C, leads to a decrease in the respiration rate of Red Delicious apples by one-third. Storing apricots at low concentrations of oxygen and high carbon dioxide levels (12–15%) for 30 days or less improves the quality of Perfection and Rival apricots.

Carbon monoxide has also been used in CA storage. Carbon monoxide is an air pollutant that is hazardous to humans and can affect vegetables detrimentally before and after harvest, because of its ethylene-mimicking capabilities. In a study, carbon

monoxide added to air, or 2% oxygen, reduced the respiration rate of head lettuce during storage at 2–5 °C. Carbon monoxide can stimulate carbon dioxide and ethylene production by climacteric fruits but has little effect on respiration rates on nonclimacteric fruits, such as strawberries. However, it has been reported that a combination of 4% oxygen, 2% carbon dioxide, and 5% carbon monoxide may be optimal for delaying maturation and ripening, maintaining good quality and retarding decay during storage of mature green tomatoes at 12.7 °C. It has also been found that packaging precut vegetables, i.e., lettuce, cabbage, celery, broccoli, cauliflower, parsley, and green onions, in 1.5 mil (a mil is a full unit used for thickness) polyethylene bags and modifying the atmosphere within these packages to include 25–50% oxygen and 3–10% carbon monoxide retard brown discoloration. Carbon monoxide has also been shown to prevent browning of mushrooms by inhibiting polyphenol oxidase activity.

Many investigations have shown that preserving foods by atmosphere modification has more advantages than disadvantages, depending on the commodity, variety, atmospheric composition, and temperature. The advantages of preserving fruits and vegetables are as follows:

- reduction of low-temperature diseases, such as chill injury during refrigeration;
- control of insect and microbial attacks;
- better preservation of organoleptic characteristics such as taste and aroma;
- retardation of ripening; and
- the period of preservation is significantly longer than when fruits and vegetables are stored in an uncontrolled atmosphere.

Potential problems with these processes, especially MA storage, include the following: storage under low levels of oxygen may lead to the development of off-flavors owing to anaerobic respiration; certain types of defects, such as brown stains on lettuce and blackheart in potatoes, can be initiated at low levels of oxygen; and at high levels of carbon dioxide, the ripening of some fruits can be irregular.

Control of Microorganisms

Fruits and vegetables contain water and other solutes, which are utilized by microorganisms deposited on their surfaces during the growing season. Most of the microorganisms are opportunistic, invading the tissue as it matures, particularly if the crops are stored under conditions of high temperatures and humidity. Although fungi and bacteria cause substantial losses in vegetables and soft fruits, most losses are caused by molds, since the inherent acid conditions of the product delay bacterial growth. (*See* **Spoilage**: Bacterial Spoilage; Molds in Spoilage.)

It is difficult to determine the full extent of postharvest losses caused by microorganisms; however, conservative estimates place US losses at around 24% of the harvested crops of fruits and vegetables. Postharvest losses have been estimated to be 50% of the harvested crop, and much of this is due to rot caused by microbes, especially fungi. Fruits and vegetables are frequently washed with recirculating water after harvest. This water is often contaminated with soft-rot bacteria and fungi, because of the recirculation, and these can contaminate fresh commodities. Therefore, an effective concentration of a broad-spectrum antimicrobial agent should be maintained in water used for handling fruits and vegetables in order to kill potential pathogens and spoilage microorganisms. Chlorine and sodium orthophenylphenol (SOPP) have been used for this purpose. Bananas are generally surface-contaminated with several fungi, i.e., *Cephalosporium, Gloeosporium, Fusarium*, and *Verticillium* species, which may cause crown rot during marketing. These fungi can be killed by exposure for 1 min to 2 ppm chlorine in water.

Care must also be taken in properly maintaining the storage environment of fruits as a means of controlling microbial losses. Rooms used for ripening bananas, degreening citrus fruits, and refrigerated storage can become heavily infested with spores of pathogenic fungi. These rooms may be disinfected easily and economically by atomizing a water solution containing 1% formaldehyde or 5% sodium hypochlorite into the atmosphere and closing the room for a few hours. Gaseous nitrogen trichloride and sulfur dioxide have also been used to disinfect lemon and grape storage rooms, respectively. Brief heat treatment has been shown to be most successful in eradicating latent or incipient fungal infections in several fruits. Hot water is the best heat-transfer medium because of the availability, heat capacity, and lack of residues on the fruits. Hot water has been advocated for control of *Penicillium* and *Diplodia* species on oranges, *Colletotrichum* species on papaya and mango, and crown rot of banana.

Irradiation has also been investigated as an alternative to chemical means to control insect and fungal contaminants. Irradiation has also been shown to extend the shelf-life of fruits and vegetables by delaying the ripening process. The use of gamma irradiation has shown promising results in controlling specific infections in apples. Ionizing radiation doses for up to 100 krad has been approved for the treatment of produce. Infection may also be prevented by the use of fungicides before harvest or before storage.

Many weak acids, such as benzoic, sorbic, propionic, acetic, nitrous or sulfuric acid, are used to control molds in foods and stored products. (*See* **Fungicides**.)

Biological control of postharvest diseases is an emerging area that has has been widely investigated, owing to increased concern over the safety of these chemicals and to development of resistance of number of plant pathogens to these chemicals. A number of biological controls have been tried, including antagonistic microorganisms, induced resistance, and the use of natural antimicrobial substances. *Trichoderma viride* has been applied to strawberry plants to partially control gray mold caused by *Botrytis cinerea* on strawberry fruits after harvest. Brown rot on peaches has been controlled by *Bacillus subtilis*. Many investigators are trying to enhance the natural defense mechanisms of fruits and vegetables such as the reinforcement of plant cell wall, phytoalexins, PR proteins, and constitutive inhibitors such as monoterpene aldehyde. Other potential nonchemical control mechanisms include the use of ultraviolet light, heat treatment, calcium chloride, and chitosan.

A storage atmosphere that has been modified by reducing the oxygen level and increasing the carbon dioxide level can influence the development of postharvest diseases either by direct inhibition of the pathogen or by altering the resistance of the host. It has been found that 30–40% carbon dioxide in the storage atmosphere inhibits the development of several pathogenic fungi on temperate zone fruits. Tropical fruits may be less tolerant to high levels of carbon dioxide. The following are recommendations for some tropical fruits: avocado in 2–5% oxygen and 3–10% carbon dioxide, banana in 2.5% oxygen and 2.5% carbon dioxide, mango in 3–5% oxygen and 5–10% carbon dioxide, papaya in 2–5% oxygen and 5–8% carbon dioxide, and pineapple in 2–5% oxygen and 5–10% carbon dioxide.

The most commonly used gas combination is 5% carbon dioxide and 3% oxygen, a mixture that provides good storage conditions, along with low temperatures (0–4.4 °C). Dry ice can be used to increase the carbon dioxide content to 10–45% during transit or short-time storage. Such treatment is beneficial for sweet cherries, strawberries, raspberries, and several other products. Among vapor treatments, sulfuric fumigation is a standard practice for control of decay of grapes in storage, but it is injurious to most other fruits and vegetables. Biphenyl pads have been used to control decay in packages of some citrus fruits. (*See* **Fumigants**.)

Another potential disinfectant for fruits and vegetables is ozone. This has been extensively investigated over the last few years. Ozone dissolved in water has been shown to be effective in sanitizing the surface of fruits and vegetables. Sodium orthophenylphenate is frequently used as an antiseptic wash or in a wax on several crops including citrus fruits, pears, sweet cherries, and sweet potatoes.

Control of Insect Pests

Insect pests, along with microbial diseases and spoilage, are responsible for much storage loss that causes extensive economic losses. Various control methods are available for these pests, including hygiene and cultural control, physical control, physical barrier, and chemical and biological control. Treatment by hot water or hot air has been used to control insects and increase the shelf-life of fruits and vegetables. Postharvest has also been found to affect the senescence of fruits by influencing the rate of respiration and ethylene synthesis. (*See* **Insect Pests**: Insects and Related Pests; Problems Caused by Insects and Mites.)

Use of Fumigants

Typical fumigation procedures expose the fruit, at ambient or near ambient temperatures, for several hours to methyl bromide or ethylene dibromide at concentrations of 45 g to a few kilograms per 1000 m^3 of air space in air-tight chambers. The fumigants are highly toxic, and their residues can pose an additional problem in the use of chemical fumigants. For the safety of operators and consumers, applications of fumigants should follow the standards set by the proper regulatory agencies.

Use of Gas Control

Strict control of storage atmospheres has many advantages such as delay of maturation, ripening, senescence, insects, and microorganisms. Treatment of European red mite on Delicious and Spartan apples with close to 100% carbon dioxide or nitrogen at 21–24 °C provides a 100% kill in 2 days. A mixture of 60% carbon dioxide with air provides no mite mortality, even after 7 days of exposure. This indicates that the biocidal effect resulted from low oxygen. The effects of carbon dioxide are also more or less the same on a number of other insects of economic importance. The codling moth is also sensitive to high carbon dioxide and low oxygen levels. The possible phytotoxicity caused by such an atmosphere indicates a potential for using high levels of carbon dioxide on some apple cultivars. Laboratory tests have shown the effects of high carbon dioxide (60%) and low-oxygen (0.5%) atmospheres, applied at 25 °C and at relative humidities of 60 and 95%, as reducing the amount of time required for codling moth mortality.

Design of Storage Areas

Fruits and vegetables can be stored in a CA most effectively in special chambers. These chambers are equipped with a cooling system, devices for creating the necessary gas composition, and devices for controlling and regulating relative humidity. A number of factors should be considered carefully before the construction of any storage facility. These factors include: site selection, integrity of the structure, insulation, vapor barriers, an appropriate refrigeration system, adequate temperature- and gas-monitoring system, and good storage facility layout.

Volume Planning Decisions

The requirements that should be kept in mind before planning for storage chambers are the level of produce loading that can be expected, the distribution of temperature and gases in the mass of produce, and the geometry and parameters of the chambers. If the chambers have very high loads, it will be difficult to obtain a gas mixture of the required composition to maintain CA. It has been shown that produce must be loaded in a continuous stack with a minimum gap between them and the walls. For instance, in the UK, the standard distance between boxes in the stack is 0.5 cm. When loading containers in the chambers, the gap between them is 2.5 cm; in the case of boxes on racks, the gap between them is 5 cm. In all cases, the gap between the stack of produce and the enclosure wall is 2.5 cm. At the same time, it must be considered that these chambers should be economical. Studies in the former USSR have shown that from the viewpoint of minimum costs of enclosures, chambers with a height up to 5.4 m are considered economical. In other countries, this value is 4.0–4.8 m (UK) and 5.6–6.6 m (USA, Italy, France).

Modern CA fruit storage areas consist of individual chambers with load capacities of 50–150 tonnes, or more. Volume planning designs regarding these chambers are generally made on the basis of various considerations, which are summarized in Table 1.

Table 1 Considerations for chamber designs

Chamber design	*Measurements*
Volume	4–5 m^3 $tonne^{-1}$
Size	Vary with the size of boxes
Net height	4.8–5.4 m
Gap	1–2 cm for boxes; 5–10 cm for trays
Minimum capacity	30–50 tonnes
Dimensions of chambers	6 × 6 m, 6 × 12 m, and 12 × 12 m; with a net height of 4.8–5.4 m

Hermetic Sealing and Gas Insulation

To maintain the desired gas mixtures, chambers should be airtight, and because of the cost of airtight insulation of partition walls, these chambers should not be very high. One of the most common materials used for sealing is foamed-in-place polyurethane, which should be protected by a fire-resistant thermal barrier. Another important factor is the selection of the gas-impermeable material used for the construction of the chambers. The most reliable gas insulation materials are galvanized iron sheets, aluminum foil covered with asphalt on one or both sides, vinyl-based resins, polyester and epoxy resins, reinforced fiber glass, and asphalt matriced with rubber. Besides these, plastic sheets, gas-impermeable varnish, paints, etc. have also been used. Galvanized iron sheets and aluminum foil covered with asphalt are the materials that are used most widely in Italy, the USA, and the UK. These barriers should prevent the build-up of moisture within walls and ceiling.

The most difficult places to design in chambers with a CA are the joints between walls, ceiling and floors, and inlets and outlets of pipelines in the building. One of the techniques to maintain the seal between walls and floors is the use of continuous monolithic foamed-in-place polyurethane. Doors lined with galvanized iron sheets on both the inner and outer sides have been used. Wood doors that are sealed with epoxy coatings have also been used. Urethane foam is usually used to insulate doors. The cost of these insulated doors is about 11–14% of the total cost of construction of the storage facility. Other considerations in the design of CA storage facilities include the installation of an observation window, pressure relief system, and continuous leakage testing. Research is in progress to make cold storages airtight and more flexible. A plastic tent has been developed for the CA storage of small quantities of fresh commodities. These CA tents are airtight and can be set up easily, allowing flexibility in the use of this storage space.

See also: **Fumigants**; **Fungicides**; **Insect Pests**: Insects and Related Pests; Problems Caused by Insects and Mites; **Ripening of Fruit**; **Spoilage**: Bacterial Spoilage; Molds in Spoilage

Further Reading

Calderon M and Barkai-Golan R (eds) (1990) *Food Preservation by Modified Atmospheres*. Boca Raton, FL: CRC Press.

Eckert JW (1975) Postharvest physiology. Part 1. General principles. In: Pantastico EB (ed.) *Postharvest Physiology, Handling and Utilization of Tropical and*

Subtropical Fruits and Vegetables, pp. 393–394. Westport, CT: AVI.
Eckert JW (1978) Pathological diseases of fresh fruits and vegetables. In Hutlin HO and Milner M (eds) *Postharvest Biology and Biotechnology*, pp. 161–209. Westport, CT: Food and Nutrition Press.
Forbes-Smith M (1999) Induced resistance for the biological control of postharvest diseases of fruits and vegetables. *Food Australia* 51: 382–385.
Kader AA (1978) Effects of adding CO to controlled atmospheres. In: Weichman J (ed.) *Vegetables and Postharvest Physiology*, pp. 277–284. New York: Marcel Dekker.
Kidd F and West C (1936) *The Gas-storage of English-grown Conference Pears. Report of the Food Investigation Board for 1935*, p. 85. London: Her Majesty's Stationary Office.
Leyte JC and Forney CF (1999) Controlled atmospheres tents for storing fresh commodities in conventional refrigerated rooms. *Hort Technology* 9: 672–675.
Metlitskii EG, Sal'kova EG, Volkind NL, Bondarev VI and Yanyuk VYA (1972) *Controlled Atmosphere Storage of Fruits*, p. 150. Moscow: E'konomika.
Nilsson T (2000) Postharvest handling and storage of vegetables. In: Shewfelt RL and Bruckner B (eds) *Fruits & Vegetable Quality, An Integrated View*, pp. 96–123. Lancaster, PA: Technomics.
Soderstorm EL and Brandl DG (1985) Controlled atmospheres to reduce post harvest insect damage to horticultural crops. *Proceedings of the 4th National Controlled Atmosphere Research Council*, pp. 207–212.
Yahia EM (1998) Modified and controlled atmospheres for tropical fruits. *Horticultural Reviews* 22: 123–183.

Effects on Fruit and Vegetables

M V Rama and P Narasimham, Formerly of Central Food Technological Research Institute, Mysore, India

Introduction

Controlled-atmosphere storage (CAS) is one of the most important innovations in fruit and vegetable storage systems as the composition of gas in the storage affects their storage life. The CAS technology involves reduction of oxygen (O_2) and increasing carbon dioxide (CO_2) as compared to the ambient atmosphere. Sometimes it also involves removal of ethylene and addition of carbon monoxide. CAS implies continuous monitoring and precise adjustment of these gases within the storage container, to a predetermined level. On the other hand, in modified-atmosphere storage (MAS), gas composition is not actively controlled; gas atmosphere within the package is created by the respiratory activity of the fresh produce. Several types of whole and minimally processed fruits and vegetables have been successfully stored in modified-atmosphere packages. Kidd and West originally established the scientific basis of CAS in the early twentieth century. It has been shown to be effective in extending the postharvest life and quality of a wide range of fruits and vegetables.

Increasing commercial use of this technology over recent years in both the short and long term has been in response to market demand for the supply of all types of fresh fruits and vegetables at all times. This article reviews the uses of pre-CAS treatments, physiological changes related to ripening and respiration, sensory and storage quality, and poststorage life of fresh fruits and vegetables stored in controlled atmosphere.

Gaseous Atmosphere in Commercial Use

Although the original idea was to replace refrigeration by CAS, observations show that CAS is most effective in combination with refrigeration, providing additional benefits. Responses to modification in the atmosphere vary considerably among species, organ types, developmental stage, crop temperature, growing conditions, and the degree of ripeness before harvest and include both undesirable and beneficial physiological responses. Recommended controlled-atmosphere conditions during storage for various fruits (**Table 1**) and vegetables (**Table 2**) have been given. Except for apple and pear, for most other fruits and vegetables the typical controlled-atmosphere conditions suggested have no, little, or only some commercial use.

Most fruits and vegetables tolerate as low as 2% O_2 and as high as 5–10% CO_2. The commercial adoption of CAS for a given kind of produce depends on the balance between cost and benefits, because control of gas composition is expensive. **Table 3** gives the general effects of low O_2 and high CO_2 concentrations on fruits and vegetables.

In general, the lower the O_2 and the higher the CO_2 levels, the better will be the controlled-atmosphere effects such as retardation of ripening and senescence. However, the extent to which O_2 level can be reduced and the CO_2 level can be elevated is limited due to the shift from aerobic to anaerobic metabolism, which leads to off-flavors. The incorrect application of CAS may have detrimental effects on the crop being stored. Sprouting and rooting problems may be triggered under certain CAS conditions, as in carrot, onion, and potato. Apple and pear are the main fruit

Table 1 Recommended controlled-atmosphere or modified-atmosphere conditions during storage and transport of selected fruits

Produce	*Temperature (°C)*	*O_2 (%)*	*CO_2 (%)*
Apple[a]	0–2	2–3	1–5
Asian pear[a]	2	3	1–5
Avocado[a]	10–13	2–5	3–10
Banana[a]	12–16	2–5	2–5
Black berry[b]	0–5	5–10	10–19
Blue berry[b]	0–5	5–10	15–20
Cherry[a]	0–5	3–10	10–15
Fig[b]	0–5	5–10	15–20
Grape[a]	0–5	2–5	1–3
Kiwi fruit[a]	0–5	5–10	5–10
Lemon[b]	10–15	5–10	0–10
Lime[b]	10–15	5–10	0–10
Mango[a]	10–15	5	5
Nectarine[b]	0–5	1–2	3–5
Orange[b]	5–10	5–10	0–5
Papaya[a]	10–15	2–5	5–8
Peach[a]	0–5	1–2	1–2
Persimmon[a]	0–5	3–5	5–8
Pine apple[b]	5–10	2–5	5–10
Plum[b]	0–5	1–2	0–5
Strawberry[b]	0–5	5–10	15–20
Tomato[a]	8–12	1–5	0

[a]Climacteric fruits; [b]nonclimacteric fruits.
Adapted from Kader AA (1992) *Post Harvest Technology of Horticultural Crops*, 2nd edn. Oakland, CA, USA: Division of Agriculture and Natural Resources and Bishop D (1996) Controlled atmosphere storage. In: Dellino CV (ed.) *Cold and Chilled Storage Technology*. London: Blackie.

Table 2 Recommended controlled-atmosphere or modified-atmosphere conditions during storage and transport of selected vegetables

Produce	*Temperature (°C)*	*O_2 (%)*	*CO_2 (%)*
Beetroot	0	10	3
Broccoli	0–5	1–2	5–10
Brussels sprout	0–5	1–2	5–7
Cabbage	0–5	2–3	3–6
Capsicum	8–12	3–5	0
Cauliflower	0–5	2–5	2–5
Chilli	8–12	3–5	0
Cucumber	8–12	3–5	0
Celery	0.5	1–4	0–5
Lettuce	0–5	2–5	0
Okra	8–10	3–5	0
Onion	0	3	5
Parsely	0–5	8–10	8–10
Spinach	0–5	21	10–20
Sweet potato	0–5	2–4	5–10

Adapted from Kader AA (1992) *Post Harvest Technology of Horticultural Crops*, 2nd edn. Oakland, CA, USA: Division of Agriculture and Natural Resources and Bishop D (1996) Controlled atmosphere storage. In: Dellino CV (ed.) *Cold and Chilled Storage Technology*. London: Blackie.

types stored commercially under CAS at present. Studies on other fruits and vegetables have proved promising for its wider application.

Precontrolled Atmosphere Storage Treatment

Subjecting fruits and vegetables to various types of preCAS treatments can enhance the effectiveness of CAS. Cooling kiwi fruit to $-0.5\,^{\circ}C$ immediately after harvest reduces the incidence of rot caused by *Botrytis cinerea* and increases the shelf-life during CAS. Curing crops like potatoes, yam, sweet potato, garlic, and onion at high temperature before they are stored in controlled atmosphere is beneficial to enhance their storage life. A prestorage heat treatment effectively controlled 'telescoping' of cut green onions. Several types of pre-CAS treatments have been suggested, especially to improve the quality of apple varieties. Keeping Granny Smith apples at a higher temperature of $46\,^{\circ}C$ for 12 h before CAS resulted in firmer fruits with higher soluble solids-to-acid ratio and lower incidence of superficial scald than fruits which were not preheat-treated. Hypobaric ventilation prevents the accumulation of scald-related volatiles in the epicuticular wax of apples, thus preventing scald development.

Exposing apple, pear, and avocado fruits to high levels of CO_2 shock treatment prior to CAS is shown to be beneficial; such a treatment to golden delicious apples resulted in better green color of the skin and firmer flesh, with no change in titratable acidity. It also reduced the incidence of physiological disorders such as internal browning and mesocarp discoloration in pears and avocado respectively. An initial low O_2 stress (ILOS) reduces superficial scald incidence during CAS of apple varieties like Law Rome and Granny Smith. Scald development is related to the accumulation of 6-methyl-5-hepten-2-one (MHO) in fruits. ILOS inhibits both productions of alpha-farnesene and its breakdown product, MHO. ILOS is suggested to be a potential alternative for chemical scald inhibitors.

Therefore, besides the application of controlled-atmosphere technology to preserve the quality of certain fresh fruits and vegetables, a cold and high temperature, chemicals, low O_2, or high CO_2 pretreatment for a short duration would give additional beneficial effects in extending the storage life. Given below are some of the significant effects of CAS on fruits and vegetables.

Effect on Fruit Ripening

Studies have shown that controlled atmosphere has its most beneficial effects on climacteric fruits at the preclimacteric stage by prolonging this period. The effect is less marked in climacteric fruits at the ripening stage. The climacteric fruits are much

Table 3 General effects of low oxygen and high carbon dioxide concentrations on fruits and vegetables

Effects of low oxygen	*Effects of high carbon dioxide*
Reduced respiration rate and substrate oxidation	Decreased synthetic reactions in climacteric fruit
Delayed ripening of climacteric fruit and breakdown of chlorophyl	Delaying the initiation of ripening and inhibition of chlorophyl breakdown
Prolonged storage life	Inhibition of some enzymatic reactions
Reduced rate of ethylene production and degradation of soluble pectins	Altered metabolism of some organic acids; production of off-flavors
Formation of undesirable flavor and odors with altered texture	Induction of physiological disorders and reduced fungal growth on the crop
Development of physiological disorder	Inhibition of the effect of ethylene
	Retention of tenderness and decreased discoloration levels
	Changes in sugar content and effects on sprouting, as in potatoes

sensitive to ethylene (C_2H_4), hence its removal is particularly recommended for long-term storage. CAS inhibits the production of C_2H_4 by the fruit, which is due to the absence of sufficient oxygen. In Jonathan and Elstar apples, increased CO_2 concentration reduced the activity of ethylene; similar reduced ethylene activity is reported in kiwi fruits and mission figs. The activity of several oxygen-requiring enzymes related to fruit ripening is reduced in CAS. In bananas, although low oxygen levels (1–15%) in store induce ripening, the degreening process of the peel is inhibited. The fruit in which the ripening process is initiated is slowed down in a low-oxygen atmosphere. Holding mature green tomatoes in low O_2 atmosphere delayed the decrease in cell wall galactosyl residues and increased soluble galactose levels, thus causing delayed loss of fruit firmness; this change was associated with increased development of red color, respiration, and ethylene production. So it is suggested that cell wall-related ripening processes are either ethylene-dependent or require only low baseline ethylene levels to occur.

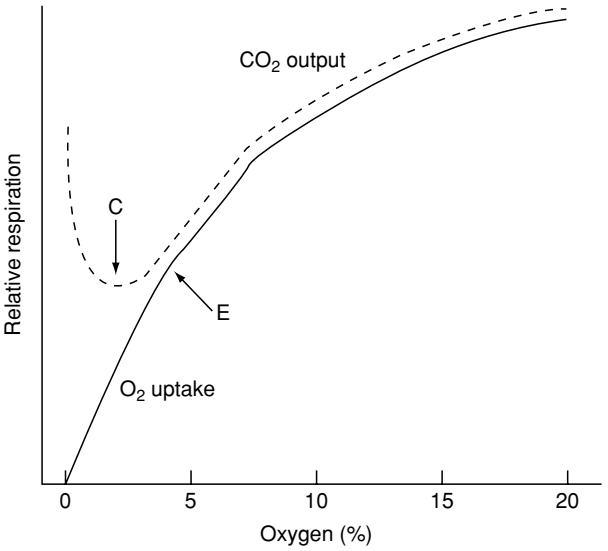

Figure 1 Schematic diagram showing the effect of O_2 concentration on the uptake of O_2 and evolution of CO_2 in fruits and vegetables. E, Extinction point (the lowest O_2 concentration at which aerobic respiration stops); C, critical O_2 concentration (the concentration of O_2 at which CO_2 produced is lowest).

Effect on Respiration

Effect of Low O_2

For some fruits and vegetables, depending on the temperature, the respiration rate is lowered during storage in 3% O_2. When the O_2 concentration of the storage atmosphere is reduced below 8% in particular, the respiration rate decreases as a function of the O_2 concentration (Figure 1). Anaerobic respiration is initiated in the tissue as the O_2 concentration drops to the extinction point. Anaerobic respiration becomes predominant, with a further decrease in the concentration of O_2, and the CO_2 output starts to increase after it reaches its lowest value at the critical O_2 concentration. Further reduction in O_2 concentration below the critical point brings a shift in the respiratory metabolism; pyruvic acid, the end product of glycolysis, is no longer oxidized through Kreb's cycle, but is converted into acetaldehyde, CO_2, and finally ethanol. This causes tissue breakdown and development of off-flavor. Thus the minimum concentration of O_2 necessary to continue aerobic respiration and prevent fermentation will be optimum for the storage atmosphere. This concentration is usually 2–5%, and depends on the kind of produce, temperature, and the duration of storage. Carrot shreds under low O_2 atmosphere exhibited an increase in respiratory quotient (RQ), an accumulation of fructose 1,6-biphosphate, coupled with an increase in the activity of phosphofructokinase (PFK). This suggests the involvement of PFK in the regulation of glycolysis under low O_2 atmosphere.

The decrease in O_2 uptake under reduced levels of O_2 in the atmosphere is due to the decrease in the

activity of oxidases like polyphenol oxidase, ascorbic acid oxidase, and glycolic acid oxidase, and not due to the reduction in the activity of cytochrome oxidase. The storage atmosphere must contain less than 2% O_2 for cytochrome oxidase to be affected; on the other hand, the affinity for other oxidases to O_2 is 20 times more than that of cytochrome oxidase. The increased activity of enzymes like pyruvate decarboxylase and alcohol dehydrogenase is not responsible for the initiation of the fermentation process; these enzymes are mostly active even before exposure to low O_2 levels.

The beneficial effects of low-O_2 atmosphere are also related to the biosynthesis and action of C_2H_4 under the established controlled atmosphere, as it is directly related to ripening and senescence processes. C_2H_4 is produced by the oxidation of 1-aminocyclopropane-1-carboxylic acid (ACC) by ACC oxidase (Eqn (**1**)):

$$\text{1-aminocyclopropane-l-carboxylic acid} + \frac{1}{2}O_2 \rightarrow C_2H_4 + CO_2 + HCN + H_2O \quad (1)$$

The concentration of O_2 that gives half the maximum C_2H_4 production rate ranges between 5% and 7% in various tissues, so that 2–5% atmosphere, usually used in CAS, directly interferes with C_2H_4 synthesis. The action of C_2H_4 on plant tissues at 3% O_2 is only about 50% of that in the air.

Effect of High CO_2

High CO_2 atmosphere affects several processes, including respiration rate, ethylene production, ripening and senescence, and acidity of the produce.

Respiration rate The respiration response to high levels of CO_2 depends upon the crop, cultivar, and the stage of development. Under similar CAS conditions the suppression of respiration was more in golden delicious apples than in Cox's orange pippin apples. In a study on the effect of elevated levels of CO_2 on respiration and ethylene production, changes in respiration seldom coincided with changes in C_2H_4 production. Respiration rate decreased in ripening banana, pink tomato, and pickling cucumber, while it increased up to 30% in potatoes and carrots, and had no effect on the respiration in guava, orange, and onion bulbs. Ethylene evolution was substantially reduced at all levels of CO_2 in guavas and tomatoes, while it was accelerated in bananas, carrots, cucumbers, onions and potatoes, which is possibly due to an early injury response.

Studies in pears have shown that high CO_2 atmosphere affects the glycolytic intermediates and enzymes; inhibition in the activity of phosphofructokinase, a key enzyme of glycolysis, resulted in the accumulation of fructose-6-phosphate and decrease in fructose-1,6-diphosphate, in contrast to the effect of low O_2. Accumulation of succinic acid, an intermediate of Kreb's cycle, has been observed in several crops, such as apples, pears, lettuce, podded pea, and activity of succinic dehydrogenase in peas under high CO_2 levels. The oxidation of several intermediates of the Kreb's cycle is suppressed, as revealed in studies of isolated mitochondria from apples. Formation of ethanol and aldehyde is common in many crops, under high CO_2 atmosphere, despite the presence of sufficient O_2; this phenomenon is called CO_2 zymasis.

Ethylene production A reduction in the O_2 uptake in climacteric fruits such as banana and avocado, without affecting the respiration activity before climacteric, has been observed under high CO_2 atmosphere; also the onset of climacteric rise, which is triggered by endogenous ethylene, is delayed. These results suggest that a high CO_2 atmosphere delays the burst of ethylene (C_2H_4). In most vegetables and nonclimacteric fruits like citrus, grapes, and Japanese pears, O_2 uptake is not affected under high CO_2 atmosphere. However a depression in the O_2 uptake occurs in ripening climacteric fruits and broccoli, accompanied by inhibition of C_2H_4 synthesis. On the other hand, 10% or more CO_2 in the storage atmosphere, in crops like lettuce, cucumber, and lemon, which are susceptible to high CO_2 level, resulted in the elevation of O_2 uptake with induction of C_2H_4 synthesis. These observations indicate that the respiratory response of crops to high CO_2 levels might be mediated mainly by the effect of CO_2 on the synthesis or action, or both, of C_2H_4. In other words, retardation of ripening or senescence and the associated biochemical and physiological changes are caused by concentrations on action or synthesis of C_2H_4 rather than on any direct effect on the respiratory metabolism. Tomatoes at breaker stage under CAS showed a reduced natural ethylene production rate and delayed increase in galactosidase activity. Such effects remained unchanged by the addition of exogenous ethylene. These results suggest that ethylene-mediated responses are impaired by low O_2 and high CO_2 and that exogenous ethylene and low O_2/high CO_2 concentrations are acting antagonistically.

Results of a study on ethylene biosynthesis in Jonathan apple fruits immediately after removal from CAS conditions (0–2% CO_2 with O_2 of 3% or 15% concentrations at 0 ° C) included inhibition of C_2H_4 production, decrease in the concentration of ACC, increased accumulation of ACC under low O_2, but

only in the absence of CO_2, and decrease in the activity of C_2H_4 forming enzyme at 20% CO_2 at both the O_2 concentrations. By these results, inhibition of ethylene production has been attributed to the inhibitory effect of CO_2 on ACC synthase activity.

Similar results are reported on other apple cultivars like Barnack and Beauty. In kiwi fruits, high CO_2 reduced ethylene production and ACC oxidase activity. Added to this, several responses of crops to ethylene such as fruit ripening, abscission, growth inhibition, floral senescence, and the induction of some enzymes, are delayed or prevented in high CO_2 atmosphere.

A comprehensive study of the effect of CO_2 in reversing the action of C_2H_4 in growth of pea segments has suggested that CO_2 competes with C_2H_4 for binding site, and the relative affinity of the site for 15% CO_2 is equivalent to that for 1 p.p.m. C_2H_4. However, the calculation of the binding site with radioactive C_2H_4 suggests that the inhibition of C_2H_4 binding is indirect or due to secondary effects such as pH changes.

Acidity and pH CO_2 dissolves easily in plant tissues, forming carbonic acid, which dissociates into bicarbonate and hydrogen ions. Hence the general assumption is that tissues exposed to high CO_2 levels would have low pH. However, the reported measurements of pH changes in several plant tissues stored in high CO_2 atmosphere have been contradictory. Intact lettuce tissues exposed to 15% CO_2 for 6 days showed a drop of about 0.4–0.1 pH in the cytoplasm and vacuoles respectively. Juice extracted from a crop exposed to high CO_2 has shown increased pH. As alterations in the pH of plant cells might affect the metabolic activities, the change in pH due to CO_2 could be treated as both beneficial and harmful effects of high CO_2 condition for the harvested crops.

Effects on Sensory Quality

Flavor and volatile compounds

Factors like volatile compounds, organic acids, carbohydrates, and phenolics affect the flavor of fruits and vegetables. Flavor determines the eating quality of the produce. Controlled-atmosphere-stored fruits are known to retain good flavor longer than fruits just cold or air-stored, as in kiwi fruits, Russian pears, and tomatoes. For golden delicious apples, CAS has proved best in terms of flavor maintenance. The controlled-atmosphere-stored white cabbages had a fresh peppery flavor and rated higher in flavor quality than corresponding air-stored samples. On the other hand, the flavor quality was inferior in limes and apricots stored in controlled atmosphere. The low O_2 of 3 kPa is reported to be enough to induce reduction in pear aroma compounds; the decadienote moiety appeared to have been more affected than methyl or ethyl moieties. The accumulation of amino acids and reduction in fatty acids and aroma compounds in Fuji apples stored under low O_2 atmosphere indicates that a change in the aroma production involves altering fatty acid and amino acid metabolisms.

Controlled atmosphere will affect the production of characteristic volatile compounds, which give flavor and aroma to fruit and vegetables. The production of volatile compounds is not only suppressed during CAS, as in several apple varieties, but the fruits fail to synthesize normal amounts even after removal from the CAS. The highest volatile emission from Granny Smith apples was after 5 months of CAS and it was highest under low-oxygen (LO) and ultra-low-oxygen (ULO) atmospheres. LO and ULO form valuable technologies for maintaining the sensorial quality of apple.

Development of Off-Flavor

The unfavorable concentrations of O_2 and CO_2 in the storage atmosphere, especially for a long time, result in the development of off-flavors and off-tastes in fruits and vegetables and are generally attributed to shift from aerobic to anaerobic respiration. In intact lettuce heads exposed to 20% CO_2, the development of off-flavor is related to the production of ethanol and acetaldehyde. More than 15% CO_2 induces odors in broccoli and cauliflower. In fruits like Granny Smith, yellow Newton apples, twentieth-century pear, and Angellena plums and peaches, ULO of 0.02–0.25% resulted in the development of alcoholic of-flavor. Sweet potatoes stored in 2% O_2 develop off-flavor with reduced sweet-potato flavor. Controlled-atmosphere-stored strawberries do not develop visible injury: the juice develops off-flavors, which are correlated to the accumulation of ethanol and ethyl acetate and acetaldehyde.

In contrast, avocados stored in air had higher ethanol and acetaldehyde content than those stored in CAS. One of the findings on the biogenesis of off-odor in broccoli under LO atmosphere suggests that broccoli has the enzymic capacity to convert free sulfur-containing amino acids into volatile sulfur compounds, which contribute to the objectionable odor. The precursors of such volatiles, free sulfur-containing amino acids, are synthesized, preferably in broccoli that is stored under LO atmosphere.

Ascorbic Acid and Acidity

The concentration of ascorbic acid in fruits and vegetables can be affected by storage conditions and is

important from the nutritional point of view. The loss of ascorbic acid is hastened in tomatoes when stored in high CO_2 atmosphere. Similar effects are reported in strawberries and blackberries, where the loss of ascorbic acid increases with increase in CO_2. In Chinese cabbage loss of ascorbic acid took place at 30–40% CO_2 levels. Though in general higher CO_2 (> 5%) content has the tendency to reduce ascorbic acid levels, it depends on the type of produce, storage temperature, and the concentration of CO_2. For example, the degradation of ascorbic acid in parsley leaves is retarded with storage in 4.2% O_2 and 1% CO_2 compared with air storage, but it is accelerated with storage in 4.2% O_2 and 6% CO_2. A LO atmosphere is favorable to the retention of ascorbic acid because of the low affinity of ascorbic acid oxidase to O_2. Pineapple stored in atmosphere containing less than 5.4% O_2 retained higher ascorbic acid content, with no effect on total soluble solids (Figure 2). The storage temperature during CAS affects the ascorbic acid content; CAS treatment at 12 °C resulted in lower ascorbic acid retention, while there was no effect at 1 °C in okra. Pears showed a decline in ascorbic acid concentration at enhanced levels of CO_2.

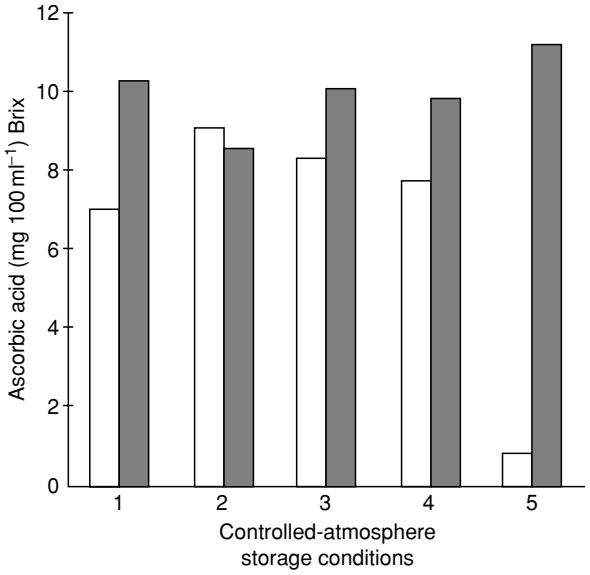

Figure 2 Effect of various controlled-atmosphere storage conditions on ascorbic acid (open columns) and total soluble solids (filled columns) of pineapple stored at 8 °C for 3 weeks and for 5 days at 20 °C. 1, 1.3% O_2 + 0% CO_2; 2, 1.4% O_2 + 11.2% CO_2; 3, 2.2% O_2 + 0% CO_2; 4, 2.3% O_2 + 11.2% CO_2; 5, 5.4% O_2 + 0% CO_2. Adapted from Haruenkit R and Thompson AK (1996) Effect of O_2 and CO_2 levels on the internal browning and composition of pineapples Smooth Cayenne. In: *Proceedings of the International Conference on Tropical Fruits*, pp. 343–350. Kuala Lumpur, Malaysia.

The organic acids, especially malic acid, of apples are retained better if favorable controlled-atmosphere conditions are reached quickly. However, if the CO_2 level reaches 3%, loss of acidity is sometimes stimulated in these fruits. A significant amount of acidity is lost in Valencia oranges stored in controlled atmosphere. In tomatoes the titratable acidity increased during the first 20 days of CAS at 13 °C and it tended to decrease during further storage (Figure 3). Vegetables such as lettuce, spinach, and broccoli always contain higher amounts of titratable acid after storage in air than after storage in controlled atmosphere.

Texture and Color

Due to delay in ripening under CAS, normally a desirable texture is maintained for several fruit types like apple, pears, peaches, apricots, and tomatoes. Retention of firmness is one of the most important advantages of the commercial CAS of apples. Though both high CO_2 and low O_2 contribute to the retention of flesh firmness, the effect of high CO_2 is more prominent than that of low O_2. Induction of enzyme polygalacturonase, which degrades pectic substances, is prevented when mature green tomatoes are stored in 5% CO_2 and 5% O_2 for up to 8 weeks at 12.5 °C: on removal of tomatoes from such an atmosphere to air, fruit softening is resumed due to the synthesis of polygalacturonase. In kiwi fruit, retention of fruit firmness is better in CAS, but presence of even low concentrations of C_2H_4 would counteract this positive effect. The texture of most other fruits and

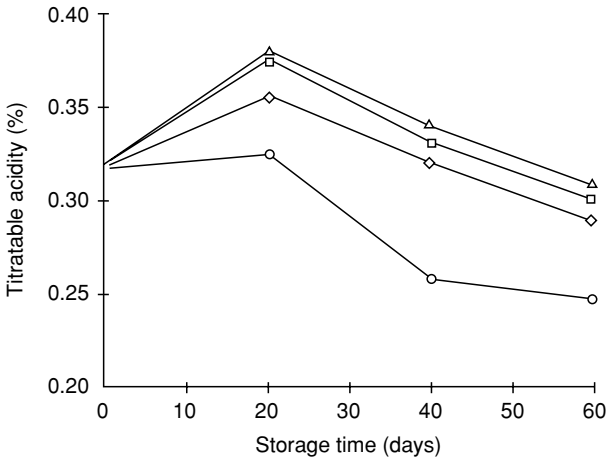

Figure 3 Changes in the titratable acidity of mature green tomatoes stored under different controlled-atmosphere storage conditions for 60 days at 13 °C (all treatments with 5.5% O_2). Circles, control; squares, 3.2% CO_2; triangles, 6.4% CO_2; diamonds, 9.1% CO_2. Adapted from Batu A (1995) *Controlled Atmosphere Storage of Tomatoes*. PhD thesis. Cranfield University, UK.

vegetables is little influenced by LO conditions. The positive effects of high CO_2 conditions reported include delay in the softening rate of strawberries and bush-type berries, reduced splitting in bitter gourd, and the retardation of the toughening of asparagus spears, cauliflower, and snap beans, and retention of bright yellow color and firmness in star fruits.

The reduced rate of chlorophyl breakdown under controlled-atmosphere conditions improves the retention of green color of the produce. High CO_2 concentration in particular preserves the chlorophyl content. Such an effect is obvious in green vegetables like spinach, asparagus spears, broccoli, snap beans, green limes, and cabbage. Some of the winter white cabbage cultivars retained better green color and crisp texture when stored at 0–2 °C under controlled atmosphere, consisting of 5–6% CO_2 and 3% O_2. The degradation of chlorophyl and the synthesis of pigment such as lycopene, carotene, and xanthophyl are suppressed when tomatoes are stored in 5% O_2 and a 5% CO_2 mixture; biosynthesis of anthocyanin is slowed down in plums under controlled-atmosphere conditions. Bitter gourd fruits stored for 3 weeks in controlled atmosphere (2–5% O_2 with 3.5–5% CO_2) had greater retention of green color. In spinach stored under low O_2 atmosphere (0.8%), the color and chlorophyl content were not affected. The extinction point of spinach is found to be > 0.4% but < 0.2% O_2, indicating that spinach can also be stored under LO atmosphere without quality loss due to anoxia. Bell peppers stored under LO showed slower rate of decrease in green color.

Controlled Atmosphere and Storage Disorders

Diseases and Pests

The process of ripening and senescence in climacteric fruits renders them susceptible to infections by pathogens. The delayed ripening and senescence in such fruits by storing them in controlled atmosphere allows them to retain greater resistance to postharvest diseases; thus CAS can reduce disease development in crops by improving their physiological conditions. When exposure time is short, some horticultural crops might be tolerant of more extreme controlled-atmosphere conditions such as <1% O_2 and > 20% CO_2 than are normally used. Such ULO and high CO_2 levels are fungistatic, directly suppressing fungal growth and spore germination. High CO_2 retards fungal decay incidence on cherry, blackberry, raspberry, strawberry, fig, nectarine, and grape. The reduction in disease levels in tomatoes is chiefly due to LO levels with a little additional effect from increased CO_2 levels. LO and high CO_2 levels, at temperatures less than 10 °C, reduce the infection of potatoes by *Erwinia caratovora*. On the other hand, anaerobic conditions created around freshly harvested or uncured potatoes during storage or transport prevent the formation of wound periderm and the oxidation of phenolic compounds, both of which are involved in the defense mechanism against bacterial infections; the depletion of O_2 rather than increase in CO_2 in the atmosphere would increase the roting by both *Erwinia* and *Clostridium*.

The postharvest pathogen of apple fruits *Penicillium expansum* can be better controlled in controlled-atmosphere conditions; its biocontrol agent *Candida sake* has the better ability to colonize under controlled-atmosphere conditions, on the apple fruit surface and in wound tissues, thus protecting apples completely from damage by the growth of *P. expansum*. The incidence of rots caused by *Botrytis cinerea* in white cabbages was reduced to a greater extent under storage atmosphere containing 5–6% CO_2 and 3% O_2; it also reduced or eliminated the incidence of pepper-spotting, a type of nonmicrobial leaf necrotic disorder common to cabbages.

Extreme controlled-atmosphere conditions are insecticidal too. For example, CAS is commonly used in Mexico to reduce the incidence of insect infestation of Manila mango fruits; these fruits can tolerate controlled-atmosphere treatment at temperatures < 44 °C. Fig moth (*Cadra cautella*) could be controlled when dried figs were stored under controlled-atmosphere conditions (1% O_2 with 10–15% CO_2). A short-term insecticidal controlled atmosphere (ICA) containing 40% CO_2 and 0.25% O_2 is shown to be a promising alternative to methyl bromide fumigation for Fuyu persimmons without affecting the marketability of fruits. A CO_2 concentration of 50% is shown to control insects in other tropical fruits like avocado, papaya, and mango.

Physiological Disorders

It is necessary to achieve and maintain precise gas mixtures during CAS, which otherwise may cause irrevocable damages to fruits and vegetables in the form of physiological disorders and injuries (Table 4). The preharvest factors like maturity at harvest, storage season and locality, and size of the fruit are also the factors determining the incidence of CAS disorders. The flesh breakdown injury and internal browning in pear fruits under CAS are associated with the production of acetaldehyde, ethanol, and ethyl acetate and stimulation of free radicals respectively, in the stored fruits. ULO of 0.5% induced a reddish discoloration in Brussels sprouts and an

Table 4 Threshold level of O_2 and CO_2 required for causing injury to some fruits and vegetables and typical injury symptoms

Crop and cultivar	*CO_2 injury level*	*CO_2 injury symptoms*	*O_2 injury level*	*O_2 injury symptoms*
Apple, cox's orange pippin	> 1%	Core browning	< 1%	Alcoholic taste
Apple, golden delicious	> 5%	CO_2 injury	< 1%	Alcoholic taint
Apple, Elstar	> 2%	CO_2 injury	< 2%	Core flush
Apple, Jonathan	> 5%	Flesh browning	< 1%	Alcoholic taint, core browning
Apricot	> 5%	Loss of flavor, flesh browning	< 1%	Off-flavor
Avocado	> 15%	Skin browning, off-flavor	< 1%	Internal flesh breakdown
Broccoli	> 15%	Off-odor	< 0.5%	Off-odor
Cabbage	> 10%	Discoloration of inner leaves	< 2%	Off-flavor
Cucumber	> 5%	Increased softening	< 1%	Off-odors
Lettuce	> 2%	Brown stain	< 1%	Breakdown at center
Kiwi fruit	> 7%	Internal breakdown	< 1%	Off-flavor
Mango	> 10%	Softening, off-flavor	< 2%	Grayish flesh, skin discoloration
Pepper, bell	> 5%	Internal browning and softening	< 2%	Off-odors and breakdown
Strawberry	> 25%	Off-flavor, brown discoloration	< 2%	Off-flavor
Tomato, mature green	> 2%	Discoloration, uneven ripening	< 2%	Off-flavor

Adapted from Thompson AK (1998) *Controlled Atmosphere Storage of Fruits and Vegetables*. New York: CAB International.

extremely bitter flavor in nongreen portions of the sprouts.

The browning of core tissue that is triggered in conference pears is related to decreased ascorbic acid concentration under enhanced CO_2 levels. Pears that show browning produce ethane, which is said to be probably a result of membrane peroxidation.

Some recent tests have shown that apples, when stored under controlled-atmosphere conditions early after harvest, are more susceptible to CO_2 injury like internal browning, flesh softening, and core browning. Incidence of Braebun browning disorder (BBD), a type of CO_2 injury in late-harvested Braebun apples, increases with decrease in the level of O_2 in the storage atmosphere. Small-sized Fuji apples were less susceptible to CAS disorder than large fruits. Delaying storing apple fruits in controlled atmosphere or CO_2 accumulation is recommended as it may reduce the incidence and severity of all the above-mentioned physiological disorders.

The appearance of CO_2 injury symptoms is shown to be a function of concentration, exposure time, and temperature, while injury under low O_2 levels is due to anaerobic respiration which results in the accumulation of toxic byproducts like alcohol and acetaldehyde. On the other hand, certain types of physiological disorders can be curtailed under CAS. For example, a LO of > 5% reduces the brown discoloration of beans which usually develops during harvesting and handling. High CO_2 is reported to inhibit browning in snap beans at the site of mechanical injury and skin browning in lichee fruit.

A hypothesis to explain the mode of action of LO and high CO_2 on the incidence of physiological disorder, such as browning, has been proposed based on a recent study on the effects of CAS conditions on adenosine triphosphate (ATP) and adenosine diphosphate (ADP) levels and the concentrations of NAD(H) and NADP(H) in apple varieties stored. The reduced ATP level and increase in the accumulation of fermentative products might indirectly change the mitochondrial membrane permeability and the electrochemical potential. Alterations in the membrane permeability leads to leakage of ions, protons, acids, and phenolics from vacuoles, causing acidification and oxidation of phenols in the cytoplasm and finally the browning of tissues.

Residual Effects of CAS

Storing fruits and vegetables in CAS can affect their subsequent shelf-life. For example, bell pepper fruits exhibited suppressed CO_2 production and O_2 consumption for at least 24 h after transfer to air. The residual respiratory inhibition in bell peppers is found to be due to a reduction in the mitochondrial oxidative capacity and not due to changes in the ultrastructure of mitochondria. Black raspberry (cv. Bristol) showed greatest weight loss after removal from controlled atmosphere. Several potato cultivars showed complete sprout inhibition, low weight loss, and healthy skin post CAS. With regard to root quality during long-term storage and poststorage processing potential, CAS of LO and high CO_2 has been shown to be an effective method for Lance Asian bell roots. In controlled-atmosphere-stored cabbages, storage and trimming weight losses were lower; recoveries of trimmed cabbages suitable for marketing or processing were higher – up to 22% than in those stored in air.

Though the present literature is ambivalent as regards the quality and rate of deterioration of fruit and vegetables after they have been removed from

CAS, many reports indicate that the storage life is adequate for marketing, and sometimes the marketable life is even better than harvested produce.

Storage of Minimally Processed Fruits and Vegetables

The preference by consumers for vegetables in a 'ready-to-cook' and 'ready-to-eat' convenient form has resulted in the supply of fruits and vegetables in minimally processed form (peeled, sliced, diced, shredded, and topped vegetables) to the market. Such minimally processed (MP) produce has been successfully stored in modified-atmosphere packages (MAP) with little or no loss in their sensory, nutritional, and market quality. MAP using polypropylene bags with microperforations has been shown to be an effective method for storing MP fruits and vegetables. LO atmosphere is used to control spoilage microorganisms on cut spinach leaves for at least 7 days. A high level of O_2 inhibits undesirable fermentation reactions, delays browning during processing, and inhibits microbial growth during the storage of MP fruits and vegetables in sealed packs. MAP has been used commercially to extend the shelf-life of minimally processed mango and other salad vegetables and leaves.

See also: **Acids**: Properties and Determination; Natural Acids and Acidulants; **Apples**; **Ascorbic Acid**: Properties and Determination; **Avocados**; **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Flavor (Flavour) Compounds**: Structures and Characteristics; Production Methods; **Pears**; **pH – Principles and Measurement**; **Ripening of Fruit**; **Sensory Evaluation**: Texture; Taste

Further Reading

Batu A (1995) *Controlled Atmosphere Storage of Tomatoes*. PhD thesis. Cranfield University, UK.

Bishop D (1996) Controlled atmosphere storage. In: Dellino CJV (ed.) *Cold and Chilled Storage Technology*. London: Blackie.

Geeson JD and Browne KM (1980) Controlled atmosphere storage of winter white cabbage. *Annals of Applied Biology* 95: 267–272.

Haruenkit R and Thompson AK (1996) Effect of O_2 and CO_2 levels on the internal browning and composition of pineapples Smooth Cayenne. In: *Proceedings of the International Conference on Tropical Fruits*, pp. 343–350. Kuala Lumpur, Malaysia.

Kader AA (1986) Biochemical and physiological basis for effects of controlled and modified atmospheres on fruits and vegetables. *Food Technology* 40: 99–104.

Kader AA (1992) *Post Harvest Technology of Horticultural Crops*, 2nd edn. Oakland, CA, USA: Division of Agriculture and Natural Resources, University of California.

Luiz A, Xuetong F and James M (2000) Delaying establishment of controlled atmosphere or CO_2 exposure reduces 'Fuji' apple CO_2 injury without excessive fruit quality loss. *Postharvest Biology and Technology* 20: 221–229.

Lund BM (1971) Bacterial spoilage of vegetables and certain fruits. *Journal of Applied Bacteriology* 34(1): 9–20.

Moleyar V and Narasimham P (1994) Modified atmosphere packaging of vegetables: an appraisal (review). *Journal of Food Science and Technology* 31(1): 267–278.

Salunkhe DK and Kadam SS (ed.) (1998) *Handbook of Vegetable Science and Technology – Production, Composition, Storage and Processing*. New York: Marcel Dekker.

Saquet AA, Streif J and Bangerth F (2000) Changes in ATP, ADP and Pyridine nucleotide levels related to the incidence of physiological disorders in 'Conference' pears and 'Jonagold' apples during controlled atmosphere storage. *Journal of Horticultural Science & Biotechnology* 75(2): 243–249.

Smock RM (1979) Controlled atmosphere storage of fruits. *Horticultural Review* 1: 301–336.

Thompson AK (1998) *Controlled Atmosphere Storage of Fruits and Vegetables*. New York: CAB International.

Weichmann J (1986) The effect of controlled atmosphere storage on the sensory and nutritional quality of fruits and vegetables. *Horticultural Review* 8: 101–127.

CONVENIENCE FOODS

J E Tillotson, Tufts University, Medford, MA, USA

History

Today, convenience foods supply much of the food requirements of the world's population in the industrial countries. Their prominence in the modern diet is the result of the urbanization of society during the nineteenth and twentieth centuries. The universal consumption of convenience foods has arisen from the demands of modern lifestyles in the industrial countries requiring, for the majority of people, that their daily nourishment be supplied by commercial enterprises rather than by their own food-producing endeavors.

Ever-growing, global urban populations, increasingly without direct access, ability, desire, or time for growing and preparing food, but with rising financial means, have increasingly wanted the convenience, availability, and taste of commercially processed foods. This market demand has motivated the growth of a global network of food-processing and distribution industries, supplying all manner of convenience foods that, by 1999, had an estimated world retail sale value of US$4000 billion. Today, at any given time, tens of thousands of various types of convenience foods are offered for sale to consumers throughout the world. All of them offer consumers the advantages of food that is readily available, convenient, inexpensive, and often of surprisingly pleasing organoleptic quality, requiring only minimal or no preparation before consumption.

Convenience foods can be looked at in several ways. These foods can be narrowly classified by the type of processing technology employed: canning, freezing, dehydration, chilling, chemical preserving, etc., or by the type of food: frozen and canned vegetables; cake mixes and bakery products; soups, sauces, and condiments; processed meats and fish; chilled and frozen dairy-based products; ready-to-eat and shelf-stable dishes; plus many other types.

Convenience foods can also be further interpreted in a broader sense, beyond the specific production technology or product type, as a vehicle of explaining the dynamics and development of our food supply in recent times, and as an aid in envisioning the future nature of people's nourishment. History teaches us that what constitutes convenience food has changed a number of times to meet the changing food needs of society.

As it has been observed that, 'an army travels on its belly,' so it is equally true that, 'modern society survives on convenience foods.' Today, convenience foods are not simply add-on, optional products, but have now become central to supplying society's nourishment. How do we explain this growth of commercialization of our food supply?

While the availability today of the plethora of convenience foods is the result of brilliant technological innovation and scientific discovery, their current prominence in the diets of the world's industrialized countries cannot be fully explained solely by the application of food technology to the food supply. Food technology has been more of an enabling factor than a motivating force. Other types of technology, including transportation and communications, have also been significant in assisting the market growth of convenience foods.

Above and beyond the diverse technical forces involved, the phenomenon of convenience foods is best understood in the broader context of the social, economic, and political forces that are also responsible for their wide acceptance. During the last century, these diverse forces have shaped and motivated the central role that convenience foods have come to play today in the modern diet. The following interprets the economic, social, technical and political dynamics that have led to the central role of convenience foods in the modern diet. Processing technology descriptions as well as detailed accounts of specific types of convenience food are treated in other sections of this encyclopedia.

Early Convenience Foods

In the mid-nineteenth century, a food-supply revolution began that has continued to the present. This revolution would result in fundamental changes in the forms of food that people in the industrial countries would depend on for their daily nourishment. While food had been preserved as early as 4000 BC by methods such as drying, salting, smoking, and pickling, the nineteenth century urbanization brought the need for a better organized, comprehensive, and dependable system of food processing and distribution.

In this era, the industrial revolution was beginning; the new factories and the accompanying commerce of the period required large new labor forces in the industrial cities of Europe and America. Many

people left rural life in agriculture to work and live in the new industrial centers. In new urban jobs, these workers required food that was both readily available and affordable. A new food-supply system sprang up to meet the new urban population's needs. This initiated the trend that resulted over time in an ever-increasing percentage of the population depending on fewer and fewer others to supply their food. These social and economic changes set in motion the industrialization of the food supply, and with it, the development of convenience foods.

As the industrial era of the nineteenth century was enabled by the application of multiple technologies, the industrialization of the food supply was similarly based on the simultaneous application of multiple technologies. The movement of vast numbers to the new large cities, removed from their own participation in agriculture, created a food-logistics problem that the growth of new railroad systems in Europe and North America would help to alleviate.

Supplying the ever-increasing demand in the new urban centers required transporting food from ever-larger geographical growing areas. Because of the economics involved in the procuring and handling of raw agricultural commodities (e.g., wheat, cattle, vegetables, sugar, milk), the basic processing of these commodities became attractive in new, dedicated processing plants in the growing and producing areas, often located far from the growing cities. The railroads were essential in transporting these processed basic food ingredients to the cities.

The urban populations were able to buy staples (e.g., flour, white sugar, meats, and dairy products) in the newly appearing local grocery stores. Urban workers used their industrial wages to pay for the basic food ingredients required in preparing their daily meals. Their wages also paid for the convenience of letting others grow, process, transport, and sell, as well as for the availability of these basic food ingredients. This established the economic model that became the foundation of convenience foods to this day, whereby money was exchanged not only for the basic agricultural commodity, but also for the labor and other services associated with the food or ingredient offered.

Historically, the majority of early preserved foods used primitive preserving methods (e.g., drying, salting, etc.) that had been developed over centuries of trial-and-error, and were designed solely for the purpose of storing present food surplus of agricultural products against future needs. The quality of these primitive 'convenience foods' was often questionable. During the mid-1800s, the developing new commercial food supply system produced the first products that would conform to the modern definition of convenience foods. With the commercial introduction in the nineteenth century of canning – the heat preservation of food in glass or metal containers – convenience foods assumed new identities, quality and importance. Consumers of this era would become ready buyers of these first mass-produced convenience foods.

In England, Brian Dookin established one of the first industrial canning factories. By 1812, he had started his canning factory and 6 years later was producing cured beef, boiled beef, carrots, mutton-and-vegetable stew, and soups of various types in tinned iron containers. However, due to their high cost, only travelers, explorers, and the military used these initial canned foods.

By 1868, the earlier handmade cans were replaced by machine-cut, assembly-line-produced containers, with a concomitant reduction in production cost and improved quality. In the late 1870s, new methods of processing under steam pressure reduced processing times and made large-scale production possible. By 1910, a single can-making machine could produce up to 35 000 containers a day. These technological advancements, combined with the use of integrated food-handling equipment to peel, shell, clean, slice, chop, and cook various foods, made canned products inexpensive to produce and, as a result, affordable and acceptable as convenience foods by millions of people. The production model developed at this time that yielded low unit costs and reliable quality would be the basis of the manufacturing strategy for convenience foods in the future.

In 1910, American processors, employing more than 68 000 factory workers, were producing over 3 billion cans of food. The large capital investments necessary for a modern cannery of the period favored the formation of large processing companies, many of American origin (e.g., *Heinz, Campbell's*). The large-scale production of these new convenience foods required large and expanded markets for their sale. Aided by the transportation and communication revolution occurring at this time, the United States market was developing from regional to national scope, which, fueled by the country's economic growth and rising incomes, offered the necessary commercial outlets for fast-increasing, canned-food production.

Food processors at this time came to realize that their future market success depended on more than technological, logistical, and manufacturing abilities. In order to build competitive and sustainable distribution, businesses would require an understanding of the factors influencing the demand side of their businesses. They realized that attracting and retaining consumers were critical to their future commercial success.

Beginning in the 1860s, progressive processors started to commit the time, energy, and resources necessary to acquire a better understanding of the consumers' needs. Leading food canners used this information to craft attractive national reputations for their companies and their products. Their marketing response was to build consumer-recognized and trusted brands through effective packaging, advertising and promotion. After the 1880s, rapidly growing food companies began branding their products for national distribution. As the scope, scale, and reputation of their nationally distributed brands expanded, shoppers now, unlike in the past where they depended on a local retailer for buying advice, came to know, trust, and purchase national brands for the quality they embodied. This brand marketing innovation would become the future business model for the success of convenience foods to this day.

Initial development of convenience foods during this period was not limited to canned foods. Other types of convenience foods produced with different technologies and packaged in other types of containers became nationally available during this period through then-proliferating neighborhood stores, beginning to organize into grocery chains. Most famous and enduring among these were convenience entries in the (then) new category of breakfast cereals. Early leaders in packaged cereals were *Kellogg's* 'Corn Flakes' and *Post's* 'Grape-Nuts,' which became most successful in encouraging Americans to substitute these products for their former breakfast choices.

By the 1880s, several rapidly growing companies were packaging and branding a wide range of other food products – *Procter & Gamble, Pillsbury, Nabisco, Borden's* – that are still prominent brands. The immediate and sustained market success of these new-to-consumers forms of convenience foods predicted the major role that these foods and their promotion would play in shaping the future diets in the industrial countries. The consumer attraction and durability of these leading brands also presaged the importance of product innovation, advertising and promotion, and wide distribution to which the market success of convenience foods owed so much.

In addition to the technological, social, and economic factors responsible for the early successful development of convenience foods, political factors also would be significant. Increasingly during the nineteenth century, the safety and quality of various processed foodstuffs were a growing concern with the public, and not without justification. Progressive canners, such as Henry J. Heinz and others, came to realize that because of the sharp practices of some processors, consumers were often suspicious of processed foods: this concern had the potential to adversely affect the industry's future. Government-imposed standards were necessary to prevent mislabeling and food adulteration, occurring all too often in this era.

During the period of the 1880s up to the successful passage of the US Pure Food and Drug Act in 1906, various state and federal legislators, government and academic food experts, civic and professional organizations (American Medical Association, National Consumers' League, General Federation of Women's clubs), and a number of food processors lobbied for government oversight. The passage of this Act and the resulting government agency oversight of the industry gave the public the necessary faith in processed foods to further enhance their wide acceptance. Earlier similar food laws had been passed and implemented in other industrial countries with parallel results. The first British Food and Drugs Act was passed in 1860 and then extensively revised and strengthened in 1872.

Packaged and Branded Convenience Foods

In the period between the last quarter of the nineteenth and the first quarter of the twentieth century, commercially produced convenience foods became established in the diets of the industrial countries. The nascent food-processing industries of that era, producing these early convenience foods, were still learning the skills of economically mass-producing products. Of equal importance, the companies were learning how to brand, distribute and effectively market these early convenience foods on a more national scale.

Consumers not only accepted, but also had an apparent great desire and need for, more and more of these convenient products that fitted their new urban lifestyles so well. The food-processing industries responded aggressively to this growing consumer demand during the early part of the century. The social, technical, economic, and political foundations had been established for the next phase – the 1930s through 1980s – for the rapid growth and proliferation as well as improvement of convenience foods with their 'built-in-chef service.' These second-generation packaged foods would be an integral part of this new era with its lifestyle that would become known as 'the golden age of consumer products.'

New advances in convenience foods would depend on a greater scientific understanding of food processing and handling. While canning had been practiced

since the eighteenth century, it was not until the end of the nineteenth century that canning was established on a firm scientific basis by research at the Massachusetts Institute of Technology. While urban populations were becoming accustomed to early mass-processed convenience foods, the scientific foundation for the myriad advanced convenience foods, which would be introduced in the 1930s and later, was being established in the research laboratories of government institutes, universities and food companies.

A notable achievement of this new science-based food innovation era was the first market introduction in the 1920s of packaged frozen foods by a company founded by Clarence Birdseye, generally considered to be the initial patent-holder and first commercial developer of frozen foods. Frozen foods offered stored processed foods with a markedly improved texture, color, and taste over equivalent canned foods. The availability of frozen foods, with their improved eating quality versus canned foods presaged a trend that would continue during the nineteenth century: a newly developed processing technology, yielding an improved convenience food, would replace products manufactured with an older technology. Consumer markets would respond favorably to the 'new and improved' marketing claims of these new consumer products.

The first part of the twentieth century would prove to be a very productive era for food science and technology. Notable commercial convenience-food successes resulting from new knowledge would be: *Crisco*, hydrogenated vegetable shortening by *Procter & Gamble* in 1911; strained baby food in individual glass jars by *Gerber* in 1927; *Wonder Bread*, packaged, sliced white bread in 1930 by *Continental Baking*; frozen June peas and frozen spinach by *Birds Eye* in 1930; *Ritz* Crackers by *National Biscuit Co.* (*Nabisco*) in 1933, which, in 3 years, would become the world's largest selling crackers; soluble (instant) coffee, first produced by Nestlé in 1938 as the *Nescafé* brand, using countercurrent roast-coffee water extraction, followed by spray drying, and later improved by *General Foods* in 1964 by a freeze-drying technology, marketed as *Maxim* freeze-dried instant coffee; *Minute Maid*, frozen orange juice concentrate by the *Minute Maid Corp.* in 1947; packaged, prepared cake-mixes by *General Mills* and *Pillsbury* in 1948; Frozen TV dinners by *C.A. Swanson's & Sons* in 1953, using flash-freezing technology; and many other famous, convenience foods consumed to this day.

Through the period of the 1950s to the 1980s, there seemed to be no end to the growth of these new branded, packaged foods. This was especially true for the leading brands in the various supermarket categories. The consumer food industry, with the help of sophisticated input suppliers – packaging companies, process equipment manufacturers, and numerous suppliers of food ingredients, flavorings, and food additives – formed industrial clusters that served as sources of product innovation. This resulted in 'new products' introductions entering American supermarkets in the thousands of items per year. New methods of consumer research vastly improved the commerical development process by identifying potential consumer needs and desires prior to the actual physical development efforts. Nevertheless, new product failure rates were always high, but the great profitability of successful market winners justified the cost of both continuous product development efforts and the high advertising and promotion expenses of introducing new products.

These second-generation convenience foods were a perfect fit for the era, and they captured the contemporary consumer tastes and lifestyles. Consumers – having survived the Great Depression of the 1930s and the Second World War – were eager for the benefits of the new consumer society that followed, providing new products often previously unavailable. These new, and improved, consumer products fitted their rising incomes and new social aspirations. New home appliances, such as refrigerators with frozen food compartments and microwave ovens, as well as second family cars and new television sets were all necessary stimuli to the growth of the new branded, packaged convenience foods. The large volumes of these new mass-marketed foods – required for supporting their relative low cost – spurred the transformation of food retailing, giving rise to new large, shopper-friendly, self-serve supermarkets offering and selling thousands of new, attractive food items.

Housewives, often living in new suburban surroundings, driving second family cars and now having the time, desire, income, and shopping interest became avid purchasers of the new convenience foods. The increasing prevalence of then-new TVs in their homes gave marketers of convenience foods an economically effective means of introducing and marketing their products. As a result, the packaged convenience food industry experienced an unparalleled era of great prosperity. While there were safety questions raised concerning the increasing use of food additives, chemicals and new packaging in convenience foods, new laws addressed the majority of these safety concerns but did not restrict their increasing use in food manufacturing as well as the market growth of new convenience foods.

Ready-prepared Convenience Foods

In the middle of the golden era of consumer products, while mass-marketed, supermarket-sold, branded, packaged convenience foods were setting new consumption records, a new type of convenience food was rushing into the market place: ready-prepared foods that offered consumers the ultimate in utility and convenience, food requiring no preparation, complete and ready for consumption. Foods, both as hand-held snacks or full meals that can be purchased in fast-food outlets and restaurants, as well as in supermarkets and stores of all other kinds, and eaten at the site of purchase, in the home, or on the move. While the convenience of ready-prepared food has, in a sense, been available in many forms since the time of the Roman Empire and their cookshops, the rapid and extensive growth plus the commercial origin of these modern ready-prepared foods was without parallel.

By the 1990s, this new wave of convenience ready-prepared foods would largely blunt the phenomenal growth experienced by packaged, branded convenience foods in the US supermarkets since the 1940s. By the early 1990s, total consumer retail spending on foods of all types at all sales outlets had reached approximately US$700 billion annually equally divided by shoppers between traditional processed foods and new ready-to-eat foods. These new convenience foods became available everywhere: in restaurants of all types, catering, and institutional outlets (schools, universities, hospitals, government, and industrial cafeterias). Even supermarkets and stores of many kinds joined the new trend to offer also take-out, ready-prepared foods. By late 1990s, applying the successful marketing and branding techniques developed earlier by the packaged food industry, the annual growth of ready-prepared foods was outdistancing traditional foods sales by approximately four-to-one (4% versus 1–2%) annually. In other industrial countries, a similar trend was occurring, but not yet to the same extent as in North America.

The ever-changing concept of convenience foods was taking on still another identity, driven by a new set of social, economic, and technical factors. Socio-economic factors, in particular, were largely responsible for the success of these new ready-prepared convenience foods. Americans were experiencing unprecedented economic growth, affluence was molding their expectations, and ready-prepared foods became 'affordable luxuries' for all. A major factor in the changing economy was increasing participation by women in the work force; the number of women working outside the home doubled from fewer than one in three in the 1950s to nearly two in three in the 1990s. Overworked, time-constrained consumers, often in dual-income or single-parent households, not having the time, energy, or desire to cook, moved from the traditional never-changing meal schedules of the past to unstructured and fast-paced feeding schedules. Furthering this trend was the number of married households with children in the USA, which declined from 40% in 1970 to 24% in 2000.

Ready-prepared convenience foods were ideally suited, in both form and price, to the new eating patterns. Americans, both young and older, experiencing new lifestyles, increasingly prefer the easy availability, immediate service, and fresher taste of ready-prepared foods, as opposed to highly processed, packaged foods requiring supermarket purchase, tiring transfer to the home, and then preparation and clean-up. Food also became more than just daily nourishment, but often entertainment, too, as people were experiencing new tastes, foods of new ethnic varieties, and new types of restaurants. By the 1990s, Americans appeared to be eating convenience foods continuously and everywhere; concurrently, their away-from-home eating places were changing, as well as the types of food they were eating.

Between 1970 and 1990, the number of full-service restaurants per 100 000 people fell by one-quarter, and the number of bars and luncheonettes by half. New eating spots – fast-food outlets – were taking their place; during the same period, the number of fast-food outlets doubled. The majority of these were part of huge national and even international chains that been started in the 1950s and 1960s (*McDonald's, Burger King, Kentucky Fried Chicken*, now known as *KFC*). These fast-food chains, with their hamburgers, French fries, fried chicken, and soft drinks, combined with new moderate-priced, family-style restaurants (*Denny's, Red Lobster*), pizza shops (*Domino's, Pizza Hut*), and coffee outlets (*Dunkin' Donuts, Starbucks*) and assorted other types, would come to dominate America's new convenience food industry, with an estimated 40 million people per day eating at, or meals from, fast-food outlets in the 1990s. By 1998, the 20 largest fast-food chains would have 79 922 outlets, with annual sales of US$56 billion, accounting for 23% of total US restaurant sales, often located side-by-side, convenient by automobile, in or by major shopping malls and highways.

In order to compete, supermarkets would start to offer ready-prepared items, as did many other types of food shops. Traditional food companies also started to offer chilled, ready-prepared convenience foods. The same economic and social trends that favored the growth of the fast-food outlets also favored an equally strong growth of all types of institutional foodservice. These new convenience foods

with the often pervasive use of their brand names, have led to brand-name, fast-food outlets now prevalent in food courts of shopping malls and as mini-units in corporate, institutional, and governmental food settings.

As the nature of convenience foods has changed, so have the technologies to deliver the new ready-prepared consumer items. The traditional food industry practices the technologies of product formulation, processing, and packaging designed for product stability, often at ambient temperatures or deep-frozen, and months-long storage times, plus potential distribution over long distances. Ready-prepared convenience food technology has other objectives: freshness, short shelf-life, and capability of being prepared on demand for waiting customers in minimum time, by nonskilled labor. Often the fast-food outlet is merely the final stage in a vast, highly complex system of mass production that gathers the necessary agricultural products and ingredients, partially prepares and processes them – often in central commissaries – for final and speedy, minimal preparation at consumer store level. This new segment of the food industry has industrialized the final preparation of food that occurred in the past in the home kitchen.

The nature of ready-prepared food production, often with short-shelf-life products, requires organizations with great skill handling timely logistics and exacting, day-to-day market-demand forecasting. As a result of the significant differences in the necessary business skills required for ready-prepared foods as compared with traditional processed foods, the leading marketers of these new convenience foods are predominantly business organizations quite different from the traditional food processors.

While the major fast-food chains have acquired such market dominance and apparent invincibility in supplying modern consumers' diet at the start of the twenty-first century, political forces are building that could prove to be problems for their further growth. Concern about the narrow dietary composition and high fat content of fast-food menus is growing, and questions are being raised by medical professionals and government health policymakers, especially given their high rate of consumption by the young.

Summary

Over the last 150 years, populations of the industrial countries, as well as those of the recently developing countries, increasingly favor convenience foods as diet mainstays. Convenience foods cleverly meet the nourishment needs of modern life. World-wide shoppers want the advantages these foods offer: taste, quality, variety, economy, availability, nutrition, safety, and – above all – their labor-saving functions. Today, these foods offer the advantage of minimum, or even nil, preparation effort, and they are available everywhere at attractive prices. Their consumer prices are being determined more by the 'added-value' services offered than by the cost of ingredients used in their production. Convenience foods have changed the diet for many, while enhancing our modern, quick-paced lifestyle. They again confirm that 'food is never just something to eat.' (From *Much Depends on Dinner*, by Margaret Visser, 1986).

New food science and technology, combined with associated packaging and processing technologies as well as advances in communication, distribution, and marketing, have enabled the development of a series of convenience foods and added to ever-increasing consumer satisfaction. However, basic to their phenomenal growth are the changing social, economic, and political factors responsible for their wide everyday consumption. As the nature of these environmental factors has changed over the past century, so has the basic identity of convenience foods. As we have seen in the three major past transformations of convenience foods – early, brand and packaged, and now ready-prepared forms – outlined above, so it is likely that the nature of convenience foods will change again in reaction to new social, technical, economic, and political forces.

See also: **Canning**: Principles; Quality Changes During Canning; **Cereals**: Breakfast Cereals; **Freezing**: Principles; **Packaging**: Packaging of Solids

Further Reading

Baas E, van Battum S, Voor bergen M and Zwanenberg A (1999a) *The World Food Markets*. Utrecht: Rabobank International.

Baas E, van Battum S, Voor bergen M and Zwanenberg A (1999b) *The Fight for Stomach Share*. Utrecht: Rabobank International.

Beardsaworth A and Keil T (1997) *Sociology on the Menu*. New York: Routledge.

Bowley R (2001) *Carried Away: The Invention of Modern Shopping*. New York: Columbia University Press.

Emerson R (1979) *Fast Food: The Endless Shakeout*. New York: Lebhar-Friedman.

Hughes D (ed.) (1994) *Building with Tradition*. Wye, UK, Wye College Press.

Kahn B and McAlister L (1997) *Grocery Revolution: the New Focus on the Consumer*. Reading, MA: Addison-Wesley.

Koehn N (2001) *Brand New*. Boston, MA: Harvard Business School Press.

Levenstein H (1988) *Revolution At The Table*. New York: Oxford University Press.

OECD (1998) *The Future of Food*. Paris: Organisation for Economic Co-operation and Development.
Pillsbury R (1998) *No Foreign Food: The American Diet in Time and Place*. Boulder, CO: Westview Press.
Putnam R (2000) *Bowling Alone: The Collapse and Revival of the American Community*, New York, Simon & Schuster.
Schlosser E (2001) *Fast Food Nation*. New York: Houghton Mifflin Company.
Senauer B, Asp E and Kinsey J (1991) *Food Trends and the Changing Consumer*. St. Paul, MN: Egan Press.
Tannahill R (1988) *Food in History*. New York: Three Rivers Press.
Tillotson J (2000) Global supermarkets. *Nutrition Today* 35.

Cookies *See* **Biscuits, Cookies, and Crackers**: Nature of the Products; Methods of Manufacture; Chemistry of Biscuit Making; Wafers

COOKING

Contents

Domestic Techniques

A J Rosenthal, Oxford Brookes University, Headington, Oxford, UK

Introduction

As with all forms of cooking, domestic cooking is intended to improve the palatability of the food, making it more appetizing. Unlike industrial food preparation and catering or food service, domestic cooking is carried out in the end user's home, by people who may not have any technical knowledge of what is happening from an engineering or biochemical point of view. By definition, cooking raises the temperature of the food. This results in a number of simultaneous and interrelated processes which influence the flavor, texture, appearance, nutrient content, and safety of the food. The different techniques of domestic cooking reflect the way in which the temperature of the food is raised (**Figure 1**).

Clearly, there are two basic ways in which energy can be applied to a foodstuff, resulting in a rise in temperature. The traditional route is by contact with a heated medium which causes heat to flow to the surface of the food and then on to the center by conduction. An alternative route is to apply electromagnetic radiation. Of the two types of electromagnetic radiation commonly used in domestic cooking, infrared, radiant heating (grilling) employs short-wavelength radiation which is only able to penetrate a couple of millimetres below the surface of the food. The inner regions of the food heat by conduction. Microwaves have longer wavelengths which are able to penetrate deep into foods, generating heat *in situ*.

In many cases energy transfer during cooking is not by a single mechanism. For example, ovens absorb and emit infrared energy and baked food heat by a combination of convection and radiation. Similarly, barbecues emit infrared radiation as well as generating hot combustion gases, which flow around the food, heating it by convection.

General Considerations

Surface Heat Transfer

The crucial difference between the different techniques of heating by contact with a heated material is what happens at the surface of the food.

When a solid food is placed in a hot fluid, there is a stagnant layer of fluid around the food. This boundary layer acts as an insulating barrier which slows the flow of heat from the fluid to the food.

Air is a good insulator (thermal conductivity, $0.024\,W\,m^{-1}\,K^{-1}$); therefore a boundary layer consisting of air greatly slows the flow of heat from an oven to the food (**Figure 2a**). One way of reducing the thickness of this boundary layer is to agitate the heating medium. In the case of fan-assisted ovens, the air is forced to circulate, thereby impinging on the

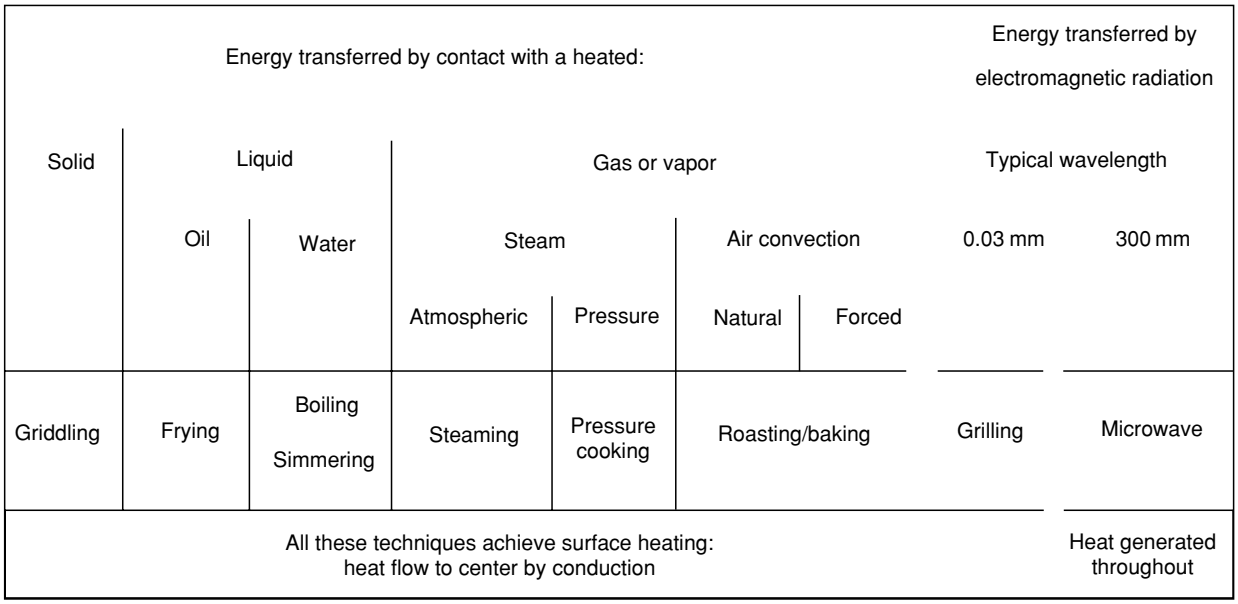

Figure 1 Classification of domestic cooking techniques.

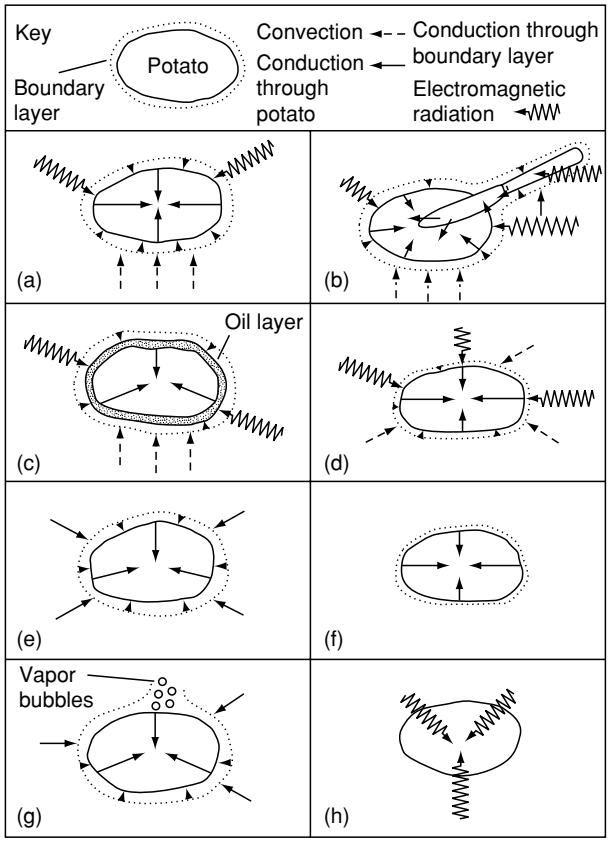

Figure 2 Eight ways to cook a potato: (a) bake; (b) bake with a good conductor (e.g. knife) inserted; (c) roast; (d) forced convection (fan-assisted oven); (e) boil; (f) steam or pressure-cook; (g) fry (chipped); (h) microwave.

surface of the food and physically reducing the thickness of the boundary layer (Figure 2d). The surface heat transfer coefficient in a fan-assisted oven is about 10 times greater than in a conventional oven.

Heat travels by conduction through static boundary layers. Heating media which are better conductors of heat than air will increase the rate of heat flow to the surface of the food. Water has a thermal conductivity of $0.573\,W\,m^{-1}\,K^{-1}$, and the rates of convective heat transfer during boiling and simmering are therefore considerably faster than in ovens at the same temperature (Figure 2e). However, the maximum temperature attainable with water is limited by its boiling point. Alternative liquid heating media, such as oil, have thermal conductivities in the same region and can operate at temperatures around 180 °C. Hence rates of heat flow to the surface in frying are considerably greater than in boiling. This is also partly attributable to physical agitation of the boundary layer caused by the generation of water vapor at the surface of the food in contact with the hot oil (Figure 2g).

When saturated steam is used as the heating medium, heat flow to the surface of the food is accompanied by simultaneous condensing of the steam. The dramatic volume change during condensation results in fresh steam flowing to occupy the void, thereby maintaining a virtually negligible boundary layer (Figure 2). However, if the steam is not saturated, then the noncondensing air accumulates at the surface of the food, forming an insulating layer. In fact, as little as 6% air in steam reduces the surface heat transfer by 90%.

Rate of Heating

During heating the rate at which heat flows into any food depends on the shape and size of the food, the temperature difference between the food and the heating medium, as well as the rate of heat transfer between the food and the heating medium.

Shape and size influence the rate of heating. Large surface areas obviously allow more heat to flow than small surfaces. Shapes with more surface-to-bulk ratio such as spherical foods will heat quicker than other shapes such as slabs. Consequently, if we consider how fast it takes heat to penetrate to the center of two 'spherical' potatoes of different size we see that it takes longer for heat to penetrate to the centre of a large potato compared with a small one. There are two reasons for this: the large potato has a greater distance for the heat to flow to reach the centre and the ratio of the heating surface area to mass is smaller for a large potato.

The driving force in any heating operation is the difference in temperature between the food and the heating medium. The greater the difference, the higher the rate of heat transfer. In the case of domestic cooking, foods heat by what is referred to as unsteady-state heat transfer. That is to say that the temperature difference reduces as the temperature of the food rises and hence the driving force lessens as the process ensues.

The physical properties of the food and the heating system contribute to the heating rate. Factors such as how easily the heat is able to flow through the insulating surface boundary layer and then on through the bulk of the food by conduction towards the center are influential.

Engineers who study heating processes have methods of predicting rates of heating based on the physical properties of materials and heating systems. In order to illustrate the factors which influence the rate of heating, Figure 3 has been produced. As a measure of the heating rate, the time taken for a food of fixed shape and size to reach a particular temperature has been used. In this figure, 2-cm-diameter spherical potatoes starting at 5 °C have been used. The identical potatoes are placed into different heating media at designated temperatures and the time taken for the centre temperature to reach 80 °C has been predicted. The temperature of 80 °C has been chosen as a realistic center cooking temperature for many meats and vegetables, and one which will certainly achieve gelatinization of starch in a food like a potato.

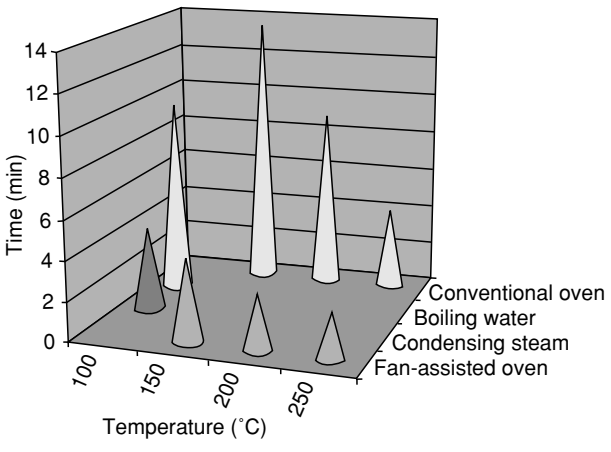

Figure 3 Predicted time for the center of 2-cm-diameter potatoes to reach 80 °C.

Clearly, short cooking times can be achieved in a number of ways. Working at higher temperature achieves more rapid cooking, as does selecting a cooking process with a high surface heat transfer coefficient. Hence, working with a fan-assisted oven is faster at cooking than conventional ovens because they reduce the thickness of the insulating surface boundary layer around the food.

Cooking by Electromagnetic Radiation

Infrared

As with the other traditional forms of cooking, grilling (broiling) causes the surface of the food to heat, and the center then heats by conduction. Absorption of infrared radiation at the surface is proportional to the difference of the fourth power of the temperature of the radiation source and the surface of the food. However, various foods absorb infrared radiation to differing extents depending on their absorptivity (e.g., black body, 1; perfect reflector, 0; water, 0.96).

Conventional grilling and barbecuing are restricted by the fact that radiant energy is supplied from one direction; cooking of all sides is only achieved by frequently turning the food, e.g., on a spit.

In practice, infrared heating is involved in many cooking techniques, such as baking and roasting. In these cases the oven walls heat up by convective heat transfer and then emit infrared radiation.

Microwaves

Microwaves penetrate deep into most foods; their sinusoidal electromagnetic waveform causes dipoles (e.g., water) to oscillate, which in turn causes heat (Figure 2h). As microwaves pass through a food material their energy is lost; although heating does occur within the food, it is greatest at the surface. The actual temperature achieved within the food depends on several factors, including the frequency of the radiation, the strength of the field, the dielectric loss factor of the food, the free moisture present, and the shape of the food.

Quality Changes during Cooking

General Considerations

Mrs Beeton tells us that there are only two basic cooking methods: wet and dry. Wet cooking includes boiling, simmering, steaming, pressure cooking, stewing, and poaching. During wet cooking the surface temperature of the food does not exceed the boiling point of water (usually 100 °C, but up to about 120 °C in domestic pressure cookers); consequently, the surface remains moist. Dry cooking includes frying (deep and shallow), baking, roasting (in an oven with natural or added fat), grilling (broiling), barbecuing, and griddling. The surface temperature during dry cooking may well exceed 100 °C, leading to evaporation of moisture and a dry or crisp surface.

Whatever is being cooked, it is worth bearing in mind that most foods are biological in origin and many are still respiring up to the time of cooking. The rise in temperature during cooking can result in a severe disruption of structural integrity and the termination of normal metabolic activity.

We eat a diverse range of foods, each with its own criteria of quality. Since the changes that occur during cooking are dependent on both the temperature achieved in the product and the time at which the food is held at each temperature, the quality of the product is constantly changing. Most of the quality changes that take place during domestic cooking can be described by first-order Arrhenius kinetics (i.e., after a finite activation energy is reached, the rate of change is proportional to a rate constant, which is itself exponentially affected by temperature). The activation energy and rate constant vary for different quality factors; moreover, any particular factor may behave differently in each food (e.g., thiamin is destroyed more rapidly during cooking of rainbow trout than in a buffered solution with the same temperature–time profile). This means that if a food is cooked on the basis of one quality characteristic (e.g., color in meat, or texture in cakes) and if the time–temperature treatment is varied from an established procedure, then the other quality parameters may be suboptimal. For example, microwaved meat may not develop the same flavor as conventionally roasted meats.

Textural Changes

No single mechanism is responsible for the texture of foods, yet changes in protein structure and solubilization of polysaccharides (and some proteins) are the two primary phenomena involved.

Heat causes proteins to unwind their tertiary structure. Heating for long periods or high temperatures frequently results in irreversible changes in tertiary structure, termed denaturation, which causes changes in functional properties. If the protein is highly charged, the uncoiled amino acid chains tend to repel each other and the protein increases its affinity for water, leading to an enhanced solubility. (*See* **Protein**: Interactions and Reactions Involved in Food Processing; Functional Properties.)

If the protein molecules are reasonably close to their isoelectric point, the amino acid chains tend to attract each other with hydrophobic interactions and hydrogen bonds, forming a network of chains. This association of molecules results in a reduction in the amount of water which is associated with protein, and hence the following consequences: a loss of protein solubility; precipitation of the protein from solution, giving rise to a solid structure (such a change is responsible for the thermal setting or gelation); loss of water-holding capacity accompanied by an aqueous exudate from the product (often observed when cooking meats); shrinkage of the product as both the above occur; and increase in the opacity of the food (e.g., egg white).

Some proteins undergo reversible thermal transitions when heated. For example, collagen (and its partially hydrolyzed derivative, gelatin) is solubilized when heated. Owing to its prevalence in connective tissue, this has important implications in tenderness of cooked meat and is partly responsible for the softening action of long-duration stewing.

Solubilization is also a process which affects polysaccharide material between the cell walls of plants. Dissolution of these pectins results in softening of the plant tissues. Addition of sodium bicarbonate enhances solubilization of pectin by producing the sodium salt and displacing the calcium ions which are naturally chelated in the structure. (*See* **Carbohydrates**: Interactions with Other Food Components.)

Another mechanism commonly involved in providing rigidity in plant tissues is turgor. Denaturation of the proteins present in the cell membranes causes a termination of osmoregulation and a subsequent softening of the tissues. Deliberate protein denaturation is the aim of blanching, a brief exposure to temperatures around 85 °C, with the intention of destroying flavor-modifying enzymes prior to home freezing.

Starch is present in many plant foods. Heating starch with an ample supply of water leads to gelatinization. The granules swell as they absorb water and the crystalline regions are disrupted. This has both a softening and a thickening effect, which is why gelatinized starch provides the viscosity of many sauces. (*See* **Starch**: Structure, Properties, and Determination.)

The light, open texture of baked flour products is due to the presence of gas or air cells in the prebaked dough or batter. During heating the gases expand; this expansion is aided by the generation of additional gas from chemical leavening agents (e.g., sodium bicarbonate) and from an increase in the vapor pressure of the water present. Finally, the liquid matrix is heat-set by denaturation of proteins and starch gelatinization. In the case of popcorn the impervious grain coat acts as the wall of pressure vessel. As the grain is heated the water vapor pressure increases until the wall ruptures. The vapor then flashes off, opening out the endosperm which then heat-sets.

A crisp outer surface is often associated with dry cooking. The temperature in excess of 100 °C results in moisture being lost from the surface.

Color and Flavor Changes

With the exception of added colors sometimes used in domestic cooking (e.g., saffron or cochineal), two types of color change result from domestic cooking: modification of natural pigments present in the raw materials, and browning reactions. Most natural colors tend to be relatively unstable when cooked. In addition to these chemical reactions, in wet cooking loss of natural color may result from leaching of water-soluble pigments into the cooking medium.

Metalloporphyrin colors are found in foods of both animal and plant origin. Myoglobin, the pigment found in muscle, undergoes color change from red-purple to brown when heated; this is a result of certain amino acids in the globin protein constituent of the molecule becoming coordinated with the central iron atom. Chlorophylls are also prone to lose their color when heated in acid conditions. The change from bright green to a greyish green is the result of demineralization of the central magnesium atom from the molecule. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments.)

Sensitivity to pH is exhibited by other natural pigments. Anthocyanins, for example, change from red (in acid conditions), through colorless (when neutral) to blue (when alkaline). This can result in color changes when heat disrupts cells and allows these water-soluble pigments to mix with the cooking water.

Browning reactions tend to occur at low levels of available water. They are therefore common in dry cooking where the surface layers may dry out. Caramelization occurs when concentrated sugars are heated, particularly in the presence of acids or alkalis. Caramelization is of importance in both sugar and flour confectionery. The most famous of these reactions is Maillard browning, which involves the reaction between carbonyl groups (as found in reducing sugars) and amino groups. The reaction gives rise to a variety of brown compounds and characteristic aromas. Simple mixtures of one amino acid and one reducing sugar have been shown to produce distinctive aromas, reminiscent of particular foods (e.g., glucose and cysteine heated at 180 °C for 30 s smells of puffed wheat, yet after 3 min it smells of overroasted meat). (*See* **Browning**: Nonenzymatic.)

Generally speaking, volatile components are lost during cooking, and this can result in a loss of the uncooked flavor. Since domestic cooking is carried out by the end user, loss of volatiles is not completely wasted and may act as an appetizer.

Sustained heating of dry surfaces can result in carbonization, usually accompanied by the generation of smoke. In such circumstances the surface browns, then blackens as it burns; such products are usually regarded as spoiled.

Nutritional Changes

Although the act of cooking involves raising the temperature of the food, the term is frequently undifferentiated from other food preparation procedures, many of which commonly precede cooking. Peeling and trimming are two such operations which can lead to substantial loss of available nutrients, by cutting unsightly portions off the food and throwing them away!

Nutrient loss during cooking is attributable to two basic routes: thermally induced chemical reactions, and leaching of nutrients into the cooking medium.

Many nutrients are thermally unstable and when heated their concentration falls exponentially with time. Obviously, different nutrients have their own rates of destruction. The most sensitive vitamins are ascorbic acid (vitamin C) and folic acid, both of which can be completely destroyed by domestic cooking. Of the essential amino acids, lysine is the least stable to heat and up to 40% may be lost by domestic cooking practices. In general, rapid cooking methods, using high temperature for short times, or microwaves cause less nutrient destruction than long-duration, low-temperature cooking methods such as stewing. (*See* **Amino Acids**: Properties and Occurrence; **Ascorbic Acid**: Properties and Determination; **Folic Acid**: Properties and Determination.) High-temperature short time processes such as frying can result in similar levels of vitamin C, as in raw potatoes.

In addition to thermal decomposition, nutrients can be lost by reacting with each other. For example, proteins will participate in Maillard reactions, particularly when ϵ-amino groups are present.

Sodium bicarbonate is sometimes added to vegetables for its softening effect. Unfortunately, its

addition leads to the destruction of vitamin C, as well as chemically modifying proteins, lowering their biological value.

Cooking processes are not entirely detrimental from a nutritional point of view. For example, frying can result in an uptake of oil, resulting in an increased energy density in foods.

Safety Aspects

A common misconception is that domestic cooking is performed purely to improve the gastronomic experience at the detriment of the food's biological value. In fact, domestic cooking can make safe products which would otherwise contain harmful or toxic components.

Harmful components in food arise in the form of naturally present toxins (e.g., cyanides in kidney beans and cassava), or compounds that interfere with digestion, effectively making the food less nutritious (e.g., trypsin inhibitors found in many legumes). The presence of pathogenic microorganisms may lead to infections, while other microorganisms produce toxins, both of which result in food poisoning. (*See* **Hemagglutinins (Haemagglutinins)**; **Plant Antinutritional Factors**: Detoxification; **Trypsin Inhibitors**.)

Cooking is effective at destroying or removing many of these harmful components. However, some toxins are relatively heat-stable (e.g., the toxin produced by *Staphylococcus aureus*) and may not be destroyed by the temperatures and times incurred during domestic cooking. It is of course likely that untrained domestic cooks will be oblivious to the risks involved in eating particular foods, or to the neutralizing effect of cooking. A further risk arises from cross-contamination of cooked food by raw ingredients, resulting in a potential for fresh growth of pathogenic microorganisms.

See also: **Amino Acids**: Properties and Occurrence; **Ascorbic Acid**: Properties and Determination; **Browning**: Nonenzymatic; **Carbohydrates**: Interactions with Other Food Components; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Folic Acid**: Properties and Determination; **Hemagglutinins (Haemagglutinins)**; **Plant Antinutritional Factors**: Detoxification; **Protein**: Functional Properties; Interactions and Reactions Involved in Food Processing; **Starch**: Structure, Properties, and Determination; **Trypsin Inhibitors**

Further Reading

McGee H (1991) *On Food and Cooking: The Science and Lore of the Kitchen*, 2nd edn. London: Harper Collins.

Domestic Use of Microwave Ovens

T Ohlsson, SIK – The Swedish Institute for Food and Biotechnology, Gothenburg, Sweden

Microwave Ovens

Towards the end of the 1970s, sales of domestic microwave ovens started to grow rapidly in Japan, a couple of years later in the USA, and at the end of the 1980s also in western Europe. The domestic microwave oven industry has now grown into a world-wide multibillion dollar business, with about 24 million ovens being sold annually. On many markets, the microwave oven is the home appliance with the largest sales volume. It is estimated that there are about 225 million microwave ovens in the homes in Japan, the USA and western Europe. User studies have shown that it is predominantly the affluent, 'middle-aged parents with kids' families that own microwave ovens.

This population of microwave oven owners represents a large potential market for food products that can be microwave-heated. The food industry has quickly responded to these market possibilities. The number of new product introductions for the microwave food segment has grown rapidly on many markets; e.g. in the USA during the middle and later part of the 1980s.

Microwave and microwaveable food today constitute an important segment of the food market in Japan, the USA and, increasingly, western Europe. The microwave food segment comprises not only prepared food items, such as meals and side dishes, but also products specially made for microwave heating, the most striking example being the microwave popcorn bag, developed by Golden Valley Microwave Foods in the early 1980s. It is today a billion-dollar market, alone. The presence of microwave ovens in the homes affects the preparation or cooking instruction for more or less all packed food products, even those not traditionally associated with microwave heating.

The growing number of microwave ovens and microwave foods has also greatly influenced the food packaging industry. Requirements of microwave compatibility have greatly changed the way in which the food industry selects food packaging. For prepared foods, microwave-transparent plastic and paperboard packaging has grown at the expense of aluminum packaging. A vast number of packaging designs, specific to microwave heating, have been developed.

The most striking example is the so-called susceptor package, in which an extremely thin metallized film on a PET film is heated by the microwaves to temperatures sufficient to crisp or brown pizza or pie crusts, or to pop popcorn.

The food technology differences between microwave and conventional heating are manifested by differences in development of food flavor and aroma and of food texture and appearance. Thus, modified recipes and special 'microwave-adapted' food ingredients have been developed by the industry to overcome at least in part some of the quality problems of microwave-heated foods.

This very substantial development of foods, packaging, and ingredients for microwave heating has given an increased general knowledge about microwave heating in the industry. Much of this knowledge is empirical, but a clear trend towards more scientifically based R&D has been seen in later years.

Heating Mechanism

What are Microwaves and How Do They Heat Foods?

Microwaves are electromagnetic waves in the frequency range from 300 to 300 000 MHz, corresponding to wavelengths from 1 m to 1 mm. For domestic food applications we are, however, limited to the ISM (Industrial, Scientific, Medical) band of 2450 $\pm$ 50 MHz.

The heating of foods by microwave energy is accomplished both by the absorption of microwave energy by rotation of the dipolar water molecules and by translation of the ionic components of the food. This energy is converted into heat. Thus, both the water content and the dissolved ion content (often salt) are dominating factors in the microwave heating of foods. When the dipolar water molecule is subjected to a microwave field that rapidly changes direction, the dipole tries to align itself with the field direction. There is a time lag, as some response time is required for the water molecule to overcome the inertia and the intermolecular forces in the water. The electric field thus provides energy for the water molecule to rotate into alignment. The energy is then lost to the random thermal motion of the water and equivalent to a temperature rise.

The energy-transfer mechanism will be efficient only if the time between the changes of direction of the electric field is so short that the dipolar molecule aggregates can barely follow the changes. If the time is long (frequency low), the alignment will be good and the energy transfer low. If the time is short (frequency high), the aggregates will not move much between field reversals, and the energy-transfer rate will again be low. Since the statistical number of water molecules that are bound together by hydrogen bonding will decrease with increasing temperature, so the inertia and energy release will be reduced. As the frequency in dielectric heating equipment will be maintained constant and below the energy transfer efficiency optimum, the efficiency will decrease with increasing temperature.

Hydrated ions, such as sodium and chloride from table salt, try to move in the direction of the electrical field, and an electric resistance heating effect is accomplished. The ions are surrounded by water molecules and will, in the movement, transfer energy randomly to the water molecules. These are more mobile at higher temperature and not so tightly bound to ions. The ions can then move more freely and absorb and dissipate more energy. The conductive heating due to dissolved ions increases with increasing temperature.

When a microwave impinges on a food surface, some of it will be reflected, and some will be transmitted and penetrate into the food and gradually be absorbed, while generating heat. Materials can be classified into reflecting, absorbing, and transparent, according to their interaction with the dielectric field. Reflecting materials are mostly metals, where the electric field creates surface currents that only penetrate a few microns into the material. Transparent materials, however, do not absorb the energy to any significant extent. Glass and most plastic materials are typical examples of this. Materials that will absorb microwave energy, according to the heating mechanism explained above, are those containing polar constituents, predominantly water.

The penetration depth (as calculated from dielectric data) for water and three food materials is shown in Figure 1 as a function of temperature. It can be seen that microwave penetration in water rapidly increases with temperature. For beef, the penetration depth is very high in the frozen state, rapidly decreasing during defrosting and leveling out with further temperature increases, whereas the penetration depth continues to decrease with rising temperature for salty ham.

Frozen foods show much lower dielectric properties as compared with thawed foods. Most of the water in frozen foods is present as ice crystals inside the food, and pure ice has very low dielectric properties. However, approximately 10% of the water remains unfrozen as a strong salt solution in the food, thus explaining why frozen foods absorb microwave energy. There is a marked jump in dielectric loss in the melting region, which partly explains the tendency towards so-called run-away heating in microwave thawing of foods, where already thawed parts

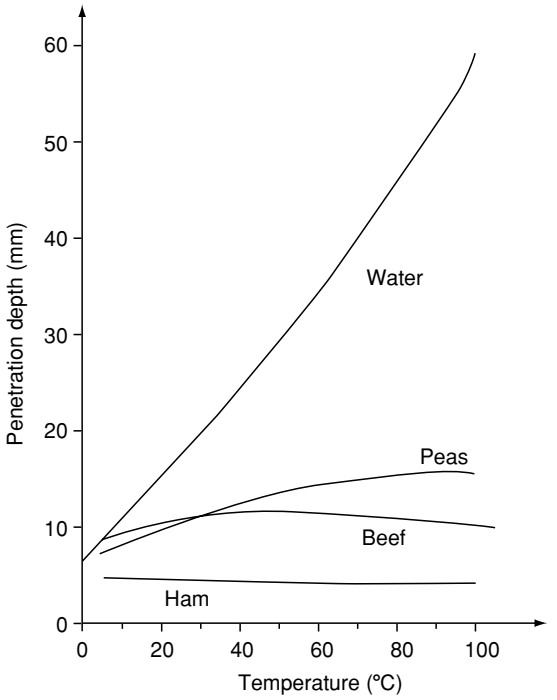

Figure 1 Penetration depth of foods at 2450 MHz.

absorb most of the available microwave energy because of their higher loss factor.

For nonfrozen foods, dielectric loss tends to decrease with increasing temperature, more pronounced so, the higher the water content and the lower the ionic content. This will tend to slow down runaway heating in thawing such foods. Salty foods, however, demonstrate a pronounced increase in dielectric loss with increasing temperature due to conductive losses. Since most foods contain some ionic material, the reduction in dipolar absorption with increasing temperature tends to be partly balanced by increasing conductive losses.

The actual temperature rise in microwave heating depends not only on the energy released in the food but also on its thermal properties, as apparent from the following energy balance:

$$Pt = mc_p\Delta T \text{ or } Pt = m\Delta H,$$

where P is power in watts; t is time in seconds; c_p is specific heat; m is mass in grams; ΔT is temperature increase in degrees Celsius; and ΔH is the change in enthalpy corresponding to the temperature rise.This means that low microwave absorption can still lead to a considerable temperature rise in foods of low specific heat, such as fats and oils.

The release of microwave energy in foods is largely dependent on both the microwave oven field and the microwave penetration pattern in the particular food material. The field distribution is a function of both the type of microwave oven and the manner of introducing the microwaves into it, as well as the type, shape, and distribution of the food inside.

In a rectangular slab of food of even thickness, the power level will gradually decrease inwards to an insignificant level if the slab thickness is large, in comparison with the calculated penetration depth.

For layered materials of different dielectric properties, microwaves will be reflected and refracted also at the interface between these materials, for example between meat and an outer layer of fat. Depending on the thickness and dielectric properties of the layers, standing wave patterns may develop, with maxima and minima. The result can be that the fat layer will be overheated, in spite of its lower dielectric loss factor. A contributing factor will then be the much lower specific heat of the fat material.

In a slab with sharp corners and edges protruding into the microwave field, energy concentrations will occur there, causing selective heating, especially at the corners. A sharp edge or corner will act as an antenna and attract more energy than surrounding areas.

When microwaves impinge on to a food surface at different angles, part of the energy is reflected, part refracted. Refracted energy will be partially absorbed. Most of the remaining energy will be reflected back at the other food surface and so on. Depending on the food geometry, the result can be focusing of energy to certain areas, which may be part of the explanation for concentration heating effects.

For spherical or cylindrical shapes, the result can be a concentration of energy to the center of the food, depending both on the food diameter and on its dielectric properties. This central heating effect occurs for diameters approximately one to three times the penetration depth in the material. For cylinders, concentration effects occur when the electrical field is parallel to the cylinder axis. The effect is stronger for foods with high values for its dielectric properties. At 2450 MHz, central heating usually happens at diameters between 25 and 55 mm, a common food size.

Heating Effects

Microbial Effects

The basis for the application of microwaves to food is the thermal heating effects caused by a combination of dipole rotation and electrical resistance heating. Over the years, there has been considerable discussion about the possibility of additional, nonthermal or *per se*, effects, from the exposure of the food to microwave fields.

In what ways, if any, will the effects on microorganisms of microwave heating differ from that of conventional heating methods?

The problems of insufficient microbiological inactivation in microwave reheating of chilled foods were highlighted by a study on survival of *Listeria* in ready meals heated according to recommended cooking instructions in the UK.

The study gave rise to much concern about the microbiological safety of foods heated in microwave ovens. The requirements that all parts of the food must reach 70 °C have been made clear to all parties involved, as a result of the UK microwave oven safety concerns.

It had already been claimed in the early 1960s by some researchers that microwave heating may have special, nonthermal effects on microorganisms, permitting sterility to be reached at considerably lower temperatures and with shorter treatment times than by other means of heat processing.

A number of comprehensive studies on microorganisms have been reported, which demonstrate the absence of any other effects than purely thermal.

The conclusion reached in 1989 by an IFT expert panel on Food Safety and Nutrition was that it is generally recognized that microwave interactions are solely due to thermal effects.

In 1989, results were published on exposure of milk, claiming the formation of unacceptable levels of D-proline and *cis*-hydroxyproline, involving a risk of brain damage in infants. However, the experimental conditions were subject to grave doubts, and efforts by other research groups to duplicate his results have been futile, even under the very excessive microwave heating conditions used.

Nutritional Effects

There are several good reviews available on the nutritional effects of microwave exposure. The general conclusion is that no significant nutritional differences exist between foods prepared conventionally or by microwave heating.

As regards the nutritional quality after comparable degrees of conventional and microwave heating, it was concluded that protein quality is retained in microwave heating of animal foods and improved in vegetables. Since no crust is formed in microwave cooking, the availability of amino acids, especially of lysine, is high. The content of fat in fatty details of meat and pork may decrease. Minerals and vitamins such as potassium and vitamins B_1 and C, are better retained in microwave heating as the cooking is done in a minimum amount of water – provided excessive heating is avoided.

Sensory Effects

Microwave-heated foods demonstrate more or less pronounced differences in appearance, flavor, and texture from conventionally heated foods. This is a result of the differences in temperature, moisture, and pressure gradients already discussed, and the related changes in structure and binding properties. Absence of surface color can be avoided only by combining microwave heating with other surface heating by hot air or by the use of susceptors.

In conventional cooking, a combination of enzymatic and chemical reactions is given time to develop. These reactions cause textural changes such as softening, firming, and tenderizing, and the formation and/or breakdown of flavor substances to give the combined flavor that is traditionally regarded as the normal or optimal. The absence of surface drying and Maillard browning reactions in microwave heating will also mean the absence of flavor notes that are very important to meat and bread.

The Microwave Oven

It is estimated that there are more than 225 million microwave ovens in homes in the industrialized world today. Penetration into households has reached above 100% in the USA, Japan, and Australia and above 80% in the UK and the Nordic countries in Europe. There are hundreds of models on the market in a wide range of design, control features, sophistication, and price. The basic oven shape and components are illustrated in Figure 2.

Alternating current from the mains is transformed to around 4 kV and rectified to direct current in the power pack. This is fed to the magnetron tube, in which the electric energy is transformed into microwave energy. This, in turn, is transferred through a waveguide to the rectangular oven space, the so-called microwave cavity. The cavity is a closed metal box, designed to prevent any external leakage of microwave energy. It is fitted with a door with microwave seals and often has a microwave tight metal screen for visibility of the oven interior during heating. The food to be heated is positioned on a bottom plate of microwave transparent material or on a rotating plate, or turntable.

Differences can be considerable between the multitude of oven makes and models on the market today, both in generated microwave power and in field distribution. Commonly, microwave output power lies between 500 and 1000 W for domestic ovens. Control features range from a simple timer and a start–stop button to weight and temperature sensors and microchip-based programing. The

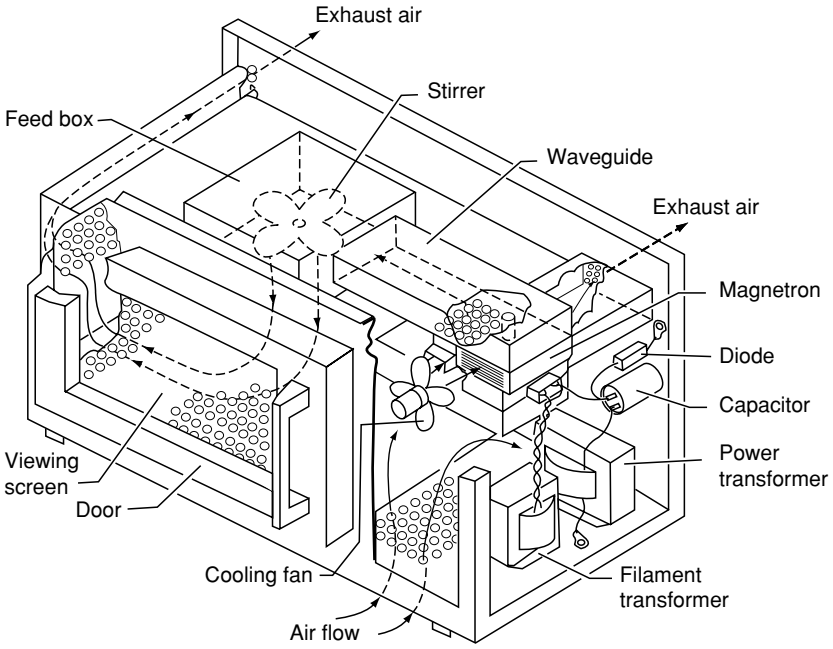

Figure 2 The microwave oven.

dimensions of domestic ovens are normally $30 \times 30 \times 25$ cm, dimensions that are of comparable magnitude to the wavelength in air at 2450 MHz (12 cm), which is the only frequency in use today for domestic microwave ovens. There is a growing trend to combine microwaves with a conventional heat source, such as hot air convection heating. Similar combination ovens already enjoy a substantial share of the domestic oven market on the European continent.

Energy is fed into the cavity by means of some kind of waveguide and antenna arrangement or opening slot in the roof or upper side wall of the cavity. The arrangement is designed to distribute the energy efficiently and evenly inside the cavity to enable rapid and uniform heating of a wide range of food loads.

In the oven space, a limited number of standing wave patterns can be generated as a result of multiple reflections at the metal cavity walls, usually not more than three to five different patterns. This means that the field is not entirely evenly distributed in the three dimensions, but demonstrates a pattern of maxima and minima. At the metal surfaces, the field strength is always zero. The field pattern is continuously modified by using a mode stirrer, a rotating propeller-like device, that will help a number of the possible modes to exist in the oven for certain time intervals. Alternatively, field differences can be compensated for by moving the food through the field on a rotating turntable at the bottom of the oven space. Most ovens use a combination of direct microwave radiation from the in-feed region and a multimode standing wave pattern in the oven.

In addition, ovens are often designed to contain devices that to some extent direct microwaves into the central area of the oven shelf, where foods are placed for heating. These so-called confined modes between the food and the oven bottom have been shown recently to be of great importance for developing ovens with more uniform heating patterns. The major reason for this is a promotion of bottom heating through promotion of the propagation of waves with an exceptionally long horizontal wavelength.

When food material is introduced into the oven cavity, the food composition, geometry of the food and package, and the positioning of the food in the oven will affect the field distribution compared with that in an empty oven.

Computer simulations of the microwave field and food temperature distribution have been used to evaluate the influence of the large number of product and oven factors since 1971, when a one-dimensional temperature profile simulation was presented. In recent years, thanks to the vast improvements in computer calculation power, numerical simulations of the three-dimensional microwave field pattern are possible, even on PCs. Both the FEM (Finite Element Method) and the FDTD (Finite Diference Time Domain) method is used. The FDTD method offers advantages of lower computer requirements and the simplicity of understanding the method and of 'building' models. However, the uniform grid (element) size

gives unnecessarily large matrices. The simulations give primarily the normalized electric field distribution. The translation to temperature patterns will also require modeling of the power dissipation of the oven.

Computer simulations will be an important field for research in the future, for improving the understanding of the complex interaction of the variables influencing the microwave field during heating, as well as for the optimization of oven design and microwave foods.

The rate of heating for a single food component in a microwave oven is primarily a function of the available microwave power, but the sample weight and the dielectric properties of the sample will to some extent affect this available power. Normally, available power will diminish with decreasing sample weight. Another factor that influences the available oven power is the line voltage.

The recognized standard procedure today for determining oven power output is the International Electrotechnical Commission (IEC) standard, which is based on heating a 1000 g water load, from 10 to 20 °C under precisely controlled conditions. The IEC has also issued an internationally recognized standard performance test for microwave ovens.

Microwave Foods and Packaging

Taking into account that microwave ovens are designed to give as even a field distribution as possible in the loaded oven with minimal corner and edge overheating, then food composition, geometry, and positioning as well as packaging will constitute the main remaining factors that determine the heating performance and resulting food quality.

For slab-shaped foods, the product thickness should be even, and preferably limited to less than 2.5 times its microwave penetration depth. Rounded edges and corners will limit preferential corner and edge heating tendencies. The temperature variations over large, flat surfaces of food, in terms of hot and cold spots, are normally larger than the in-depth temperature variation, and may have a greater influence on the overall heating results.

When heating food components of different dielectric and thermal properties together, such as in meals on a tray or a plate with different compartments, care must be taken in selecting the shape and relative arrangement of the components so as to optimize the heating uniformity. In this context, advantage can be taken of the central focusing tendencies of rounded (cylindrical or semispherical) geometry for thick food samples of limited penetration depth. Food components with a high dielectric loss, such as meat stew, can be partially shielded by components of lower loss (such as mashed potato) to even out the temperature distribution. Material with a low dielectric loss is less susceptible to corner and edge overheating, etc. By adjusting the thickness of the material with a low dielectric loss, the formation of variable standing wave patterns inside the food can be kept to a minimum.

In microwave heating, packaging is an integral part of the food product. Packaging composition, shape, and incorporated active functions all contribute to the heating result, in direct interaction with the composition, properties, geometry, and positioning of the food inside.

The packaging material should permit temperatures of at least 120 °C without affecting the handling stability or food protective properties. For low-moisture foods high in oil or sugar, the temperature stability may have to be extended up to 150 °C, because these foods will easily overheat in a microwave field. To permit reheating also in a conventional oven (dual ovenable), a temperature resistance of up to 200 °C is required.

The main packaging alternatives are:

- polymer-coated paperboard;
- mono- or multilayer polymers, laminated or coextruded;
- aluminum foil or foil- and polymer-laminated board;
- glass.

Of the polymer materials in use, PP (polypropylene) will meet the minimum temperature stability requirement, whereas APET (polyester) can be used up to 200–250 °C.

The ideal food package shape is round or oval with vertical sides and rounded edges to limit sharp corners and edges to a minimum. Rectangular containers may be required for practical reasons but should have rounded corners and edges to reduce preferential heating tendencies.

Metals reflect microwaves completely, and food areas close to a metal surface will not be microwave-heated at all. Also, the presence of metal in the package or other metal objects in the microwave oven will markedly influence the microwave field distribution.

Previously, recommendations have been to avoid metal trays or foil laminated packages in microwave heating, because they were likely to cause damage to the magnetron, due to excessive reflection of energy. Since modern magnetrons are considerably less sensitive, aluminum containers, with transparent covers may now be used safely, provided they are not positioned so close ($\leq$ 2 mm) to the oven wall that electrical discharges or sparks can occur.

The metal changes the mode pattern in the oven and thus influences the heating result. This may even be improved compared with heating in a transparent package. Because microwaves can now penetrate only from the top and not from the sides, corner and edge overheating is eliminated, even to the extent that these parts may remain cooler than the central parts of the food.

The product thickness will be limited to about half the penetration depth of the food material, to permit sufficient heating of the bottom layer. The property of metal to completely reflect microwaves can be utilized in the design of packages, which minimize heating differences between food components of different physical properties, shape or thickness. A pattern of aluminum foil can be applied on the package to screen areas that would otherwise overheat, but permit full penetration into areas that would otherwise remain cooler.

The use of susceptors was briefly mentioned earlier in this article. Susceptors, or receptors as they have also been called, are extremely thin films of metal, vacuum-deposited on PET film with paperboard backing. At a metal film thickness of only 25–50 nm, part of the impinging microwave energy is transmitted, part reflected, and a substantial part absorbed generating heat and leading to high temperature in the metal film. When such a susceptor film is applied a few millimeters from the food surface, the generated heat will transfer rapidly to the food and heat the food surface. The surface temperature reached will be sufficient to reverse the temperature and moisture profile in the food and to crisp and brown the food surface, like in conventional grilling. Susceptor temperatures will usually not exceed 115–130 °C when in close contact with the food, but may otherwise reach temperatures up to and above 250 °C. At such high temperatures, there will be some risk of migration of substances into the food from the plastic film and paperboard backing. Therefore, susceptors are being developed that have a built-in temperature-limiting function.

There has been much concern for migration of hazardous chemical components from susceptor packaging material. Extensive research in the USA, UK, Germany, The Netherlands, and Sweden have demonstrated, however, that microwave heating itself does not increase migration rates.

Future Outlook

The market for *domestic microwave ovens* will continue to grow world-wide, even though a near-saturation point has already been reached in a few countries, like the USA and Japan, where all households already have one or more ovens.

The *heating performance* and consistency of microwave ovens will improve gradually as a result of the ongoing international standardization of performance and output power testing methods. At the same time, the growth in fundamental knowledge on field distribution and the interactions between oven design, microwave field, and food will lead to improved oven performance.

Packaging development continues in materials, shape, shielding, susceptors, and other forms of active packaging, in combination with a product development that increasingly takes into account dielectric and thermal properties of foods and their interaction with the microwave field.

There will also be important improvements in oven components, such as solid-state components, which will substantially lower both oven weight, size, and production costs. In Europe, at least, the proportion of combination ovens, combining microwaves with heat by convection or radiation, will continue to grow.

See also: **Browning**: Nonenzymatic; **Packaging**: Packaging of Liquids; Packaging of Solids

Further Reading

Buffler CR (1993) Microwave cooking and processing. In: *Engineering Fundamentals for the Food Scientist*. New York: AVI, Van Nostrand Reinhold.

Cross GA and Fung DYC (1982) The effect of microwaves on nutrient value of foods. *CRC Critical Reviews in Food Science and Nutrition* 16: 355–381.

Jauchem J (1991) Alleged health effects of electromagnetic fields: Misconceptions in the scientific literature. *Journal of Microwave Power and Electromagnetic Energy* 26(3): 189–195.

Mudgett R (1985) Dielectric properties of foods. In: Decareau R (ed.) *Microwaves in the Food Processing Industry*. London: Academic Press, pp. 15–37.

Mudgett RE (1989) Microwave food processing. A scientific status summary by the IFT's expert panel on food safety and nutrition. *Food Technology* 43: 117–126.

Ohlsson T (1983) Fundamentals of microwave cooking. *Microwave World* 4(2): 4–9.

Risman PO (1993) Microwave food hazards. *Microwave World* 14: 21–23.

Sacharow S and Schiffmann RF (1992) *Microwave Packaging*. Leatherhead, UK: Pira International.

Sundberg M (1998) Analysis of industrial microwave ovens. Doctorate thesis, Department of Microwave Technology, Chalmers University of Technology and SIK – The Swedish Institute for Food and Biotechnology, Gothenburg.

Walker J (1987) The secret of a microwave oven's rapid cooking is disclosed. *Scientific American* 258(2): 98–102.

COPPER

Contents
Properties and Determination
Physiology

Properties and Determination

R Barberá, R Farré and M J Lagarda, University of Valencia, Burjassot, Spain

Physical and Chemical Properties

Copper is a tough, soft, ductile, reddish metal, with an atomic number of 29, a relative atomic mass of 63.54 amu, an electronic configuration $1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^{10}\ 4s^1$ and a relative density of 8.95. The melting and boiling points are 1083 and 2595 °C, respectively.

Copper is the 25th most abundant element in the earth's crust, comprising about 0.01%, and has two natural isotopes, ^{63}Cu and ^{65}Cu, with isotopic abundances of 68.94 and 31.06%, respectively. In total, 11 isotopes are known. Copper is found in its native state and also as a component of sulfides, oxides, and carbonates. Copper is used in alloys such as brasses and is completely miscible with gold (Au). It can be only superficially oxidized in air, sometimes acquiring a green coating of hydroxo carbonate and hydroxo sulfate.

The principal states of copper oxidation are +1 and +2. However, Cu(I) is not stable in solution, since the disproportionation reaction has an equilibrium constant of $1.1 \times 10^6\ 1\,mol^{-1}$. The complex formation can stabilize Cu(I) with the complex formed with cyanide presenting a high stability. Most Cu(I) compounds are fairly easily oxidized to Cu(II) compounds, but further oxidation to Cu(III) is more difficult.

The most common oxidation state, especially in aqueous solution, is +2. In diluted acid solution, Cu(II) precipitates as the sulfide (solubility product $10^{-35}\ mol^2\,l^{-2}$) with hydrogen sulfide or ammonium sulfide. Copper dissolves readily in HNO_3 and H_2SO_4 in the presence of O_2, and it is soluble in NH_4OH, ammonium carbonate, or KCN solution. Most Cu(II) salts dissolve readily in water.

Cu(I) compounds are diamagnetic and, except where color results from the anion or charge-transfer band, colorless. Cu(II) complexes and compounds are blue or green, except for red or brown species. The blue or green colors are due to the presence of an absorption band in the 600–900-nm region of the spectrum.

Determination in Foods and Beverages

The copper content in foods and beverages is in the order of micrograms per gram ($\mu g\,g^{-1}$).

Atomic absorption spectrometry (AAS) is probably the most commonly used technique of the wide range available for the analysis of copper in foods and beverages. When the contents are sufficiently high, flame atomic absorption spectrometry (FAAS) is used, and when they are low, electrothermal atomic absorption spectrometry (ETAAS) is employed. Whenever possible, FAAS is the technique of choice, since it is less time consuming and less sensitive to interference (background absorption).

The widespread distribution of copper is a source of problems of contamination, and as a result, special care must be taken when dealing with samples containing very low levels, as occurs with some foods, bearing in mind that the lower the concentration, the higher the risk of contamination. Wherever possible, plastic instruments should be used to prepare biological samples whose copper contents are low. In order to prevent contamination of samples, all the glass- and plastic-ware used must be thoroughly washed with diluted nitric acid. Blanks should be used to check the absence of contamination. To ensure the quality of the results, certified reference materials of a similar nature to the material being analyzed should be used. This also allows any suspicion of contamination to be eliminated.

Sample Digestion

Samples of most types of food and beverage require a prior preparation in order to convert the sample into solution before analysis. The choice of the procedure depends on the element to be determined, the complexity of the matrix, and the analytical technique to be applied.

This preparation can be omitted in some samples, especially in beverages, such as tap water or wine, where, in the latter, a simple dilution of the sample can be sufficient. Nevertheless, in most food and beverage samples, the destruction of the organic matter is necessary to a greater or lesser degree, depending on the technique of analysis to be applied.

The sample preparation is a critical step in trace-element analysis. The classic standard methods for organic-matter destruction are: wet digestion by conductive heating (hot plate, digestion block) or dry ashing (muffle furnace) at a defined temperature. Wet-digestion methods are generally more rapid than dry digestion methods. The disadvantages are that only fairly small samples can be used and also that the digests normally have to be heavily diluted before the analysis, resulting in poor detection limits. These techniques usually use mineral acids, oxidants, and ashing aids. Wet digestions can be carried out with a concentrated acid or mixture of acids, in open or closed systems. Since 1975, microwave-assisted sample digestion has been increasingly used for sample preparation in trace-element determination in food samples.

The advantages of microwave digestion over conventional sample preparation techniques are: a reduction in the time needed for the sample preparation and also of the amounts of acids used for digestion, and the possibility of controlling pressure/temperature, automation, and flexibility in sample preparation. Two types of microwave digestion systems are now available for wet digestions: closed-vessel microwave and open-vessel focused microwave.

Each of these techniques offers specific advantages. Closed systems can be used to prepare samples under controlled digestion conditions with a minimal volume of acid and a reduced risk of contamination. Nitric, hydrochloric, and hydrofluoric acids and their mixtures are the most widely used with closed digestion vessels. Nitric acid is usually added due to its antioxidant properties, and hydrofluoric acid is used when siliceous matter is present. The addition of hydrogen peroxide to nitric acid is also recommended.

Automated focused open-vessel microwave systems operate at ambient pressure and offer flexibility in digestion conditions, with vessels and reflux columns available in borosilicate glass, Teflon perfluoralkoxy, and quartz. These systems can be used to prepare samples sequentially with minimal attended operation, since digestion reagents are added automatically in programmed volumes and dispensing rates. To minimize analyte losses, a ramped, multistage heating program can be used in addition to reflux columns with the digestion vessels. When compared with the closed vessel system, the open-vessel system can process large samples, which leads to lower detection limits. High-pressure and high-temperature conditions are not required to obtain a complete digestion when an instrumental technique such as ETAAS is used for trace-element analysis.

Precautions should be adopted in all wet-digestion procedures when strong acids and strong oxidizing agents are used. The formation of explosive nitro compounds has been reported when fat-rich samples have been digested with nitric acid. Whichever wet digestion procedure is applied, the digests are left to cool and diluted with water.

Dry digestion or ashing procedures are also widely used to destroy organic matter. The main problems with these procedures are the possible losses of analyte by volatilization or by retention on crucible walls during dry ashing. Although several studies strongly indicate that ashing at temperatures up to 450 °C is free of losses by volatilization or retention, it is recommended that the temperature be increased gradually up to 400–500 °C and maintained until white ashes are obtained. Dry ashing is time-consuming, usually taking a day or more, although it requires very little attention from the analyst. The resulting ash can be dissolved in a small amount of solvent, and contamination through reagents is scarce.

To eliminate potential losses and/or to speed up the ashing procedure, ashing aids and matrix modifiers can be used. This, however, always increases the risk of contamination and worsens detection limits. Ashes are usually dissolved in diluted HCl or HNO_3 and it is sometimes necessary to heat them with concentrated HNO_3 and $HClO_4$ until no more fumes are produced, in order to complete the destruction of the organic matter. In some cases, organic-matter destruction can be replaced by acid extractions of the food product. Hydrochloric acid, nitric acid, or a mixture is used as the extractant, and the use of an ultrasonic bath facilitates the extraction.

Some analytical techniques that are useful for copper determination, such as ETAAS, do not require organic-matter destruction in some kinds of food. Nevertheless, even when this is not the case, a complete digestion, minimizing residual carbon content, is not required. The suppression of the digestion step will reduce the time needed for the analysis and the risk of losses and contaminations. Organic-matter destruction and dilution or dissolution of the residue can be replaced by:

- direct injection of the sample (water) or after the appropriate dilution (wine);
- dissolution of the sample in an adequate solvent, e.g., butter in the mixture composed of butylamin/water/tetrahydrofuran (48/12/40);

- addition of a chemical modifier $Mg(NO_3)$ and an emulsifying agent Triton X-100 (0.2% w/v), e.g., to whole milk, nonfat milk, and whey milk;
- preparation of a slurry of finely pulverized sample (slurry) in acidic diluents and subsequently treated by sonication.

The advantages of the slurries are a decrease in the time needed for the analysis and a reduced risk of external contamination due to reduced sample manipulation. Using slurries, copper detection limits lower than $0.2\,\mu g\,g^{-1}$ can be achieved in biological samples.

An alternative to slurry sampling and microwave-assisted acid digestion is ultrasound-assisted extraction. The main advantage of this over slurry sampling is a doubling of the tube lifetime, most likely because the matrix is not introduced into the atomizer, thus avoiding the buildup of carbonaceous residues or silicates on the graphite platform. In comparison with microwave-assisted digestion, it gives a twofold better precision and better detection limits.

Determination

Molecular Absorption Spectroscopy

A large number of reagents have been proposed for the spectrophotometric determination of copper, but only a few, such as pyridil-azo derivatives chloroazoantipyrine, 4-(2-pyridylazo) resorcinol, cuprizone (oxalic acid bis(cyclohexylidene) hydrazine) and diphenylcarbazone, diethyldithiocarbamate are of any real practical use. Cuprizone, one of the reagents used, reacts with Cu(II) to produce a blue, water-soluble complex with a molar absorptivity of $1.6 \times 10^4\,l\,mol^{-1}\,cm^{-1}$ at 600 nm.

Visible spectrophotometric methods have been applied mainly for determining the copper content of water, a matrix that does not require prior digestion. In general, these methods should only be considered as an alternative when atomic absorption or plasma emission techniques are not available. The main limitation of molecular absorption spectroscopy methods is their lack of sensitivity to determine copper at the low content levels of foods and beverages.

A preconcentration using pretreated silica columns or a chelation solvent extraction prior to applying a visible spectrophotometric method will yield sensitivities similar to those obtained with AAS measurements. A concrete example of precolumn derivatization uses a preconcentration column of C_2-bonded silica and pretreated with ammonium 1-pyrrolidinecarbodithioate – hexadecyltrimethylammonium bromide(I) ion-pair in mobile phase buffer (50 mM Na acetate–20 mM $NaNO_3$ adjusted to pH 6.0 with dilute HCl) (4:1). The metal–dithiocarbamate complexes are then backflushed on to an analytical column Spherisorb ODS2 (5 μm). Elution with acetonitrile–buffer (17:8) containing I (15 mM, respectively) and detection at 430 nm (Cu). The detection limits are $0.5\,\mu g\,l^{-1}$, and the results are comparable to those obtained by a conventional method involving chelation, solvent extraction, and AAS.

As an example of postcolumn derivatization, there is an ion chromatographic method that uses derivatization with 2-(2-benzoxazolylazo)-1-naphthol (α-BOAN) detection at 565 nm in a nonionic surfactant solution of Brij35 used for the separation and determination of copper(II). The detection limit is 16 μg of Cu per liter. This method has been used with good results in the separation and determination of Cu(II) in certified reference materials (rice flour and bovine liver).

Moreover, some of the reactions between copper and different organic reagents, such as cuprizone, diethyldithiocarbamate, or 1,5-diphenylcarbazide, have been adapted to flow injection analysis (FIA). This analytical technique has become very popular and has been used to determine the copper content of water and different biological matrixes (cow milk, milk-based infant formulas) or their digests (plants). Detection limits of copper ranging from 0.02 to $0.3\,\mu g\,ml^{-1}$ have been reported in water and about $5\ \mu g\,l^{-1}$ in plant digests after liquid–liquid extraction. These values can be improved with purification or separation after organic-matter destruction and prior to determination.

Fluorescence

Several fluorescent reagents can be used to determine copper in food extracts and digests, the most frequently reported being: 5-(2-carboxybenzenazo) rhodanine and 5-(4-chloro-2-carboxyphenylazo) rhodanine. The intensity of fluorescence is measured at 400–408 nm (excitation at 310 nm), with detection limits ranging from 0.01 to 0.05 μg of Cu(II) per milliliter.

Atomic Absorption Spectrometry

FAAS FAAS is the technique of choice for measuring copper in food-sample digests. The most sensitive line is 324.8 nm using a lean air acetylene flame, the measurement is almost independent of the stoichiometry of the flame, and the interferences are scarce, though large amounts of some transition elements in the presence of mineral acids can depress the response; e.g., $10\,000\,\mu g\,ml^{-1}$ depresses 10 μg of copper per milliliter by 10%. At 324.8 nm, a detection limit of $0.0015\,\mu g\ ml^{-1}$, which is sufficient for many

routine determinations, can be achieved under optimal conditions.

Flame emission can be used at a wavelength of 324.8 nm with a lean nitrous oxide acetylene flame. To improve the sensitivity of FAAS when trace amounts of copper have to be measured, separation and preconcentration/enrichment procedures based on adsorption, chelation, or precipitation processes can be used.

ETAAS ETAAS is recommended when copper contents are very low or the sample volume is very small. The 324.8 nm line is used. Copper is easily atomized from the wall or the platform of a graphite tube, and only minor interferences have been reported, as a result of which no chemical modifier is usually necessary in the ashing step. Temperatures of 800 and 2000 °C are used for ashing and atomization, respectively. The rate-limiting step for the atomization of copper is the reduction of Cu_2O by graphite that precedes atomization.

Chloride interferes in the determination of copper by ETAAS, so chloride should be removed from the sample before electrothermal atomization, a small excess of an oxyacid (sulfuric or nitric) added before the drying or ashing steps, or the method of standard additions should be used for quantification. To suppress chloride interferences, NH_4NO_3 has been used together with pyrolytically coated tubes, which are more sensitive but less robust than uncoated tubes. A 1% ascorbic acid solution has been found to reduce the interference of seawater upon the determination of copper.

Background corrections can be made by using a deuterium lamp or a Zeeman correction, although copper is one of the few elements to suffer from sensitivity loss when the Zeeman correction is used. The 324.7-nm line has an anomalous Zeeman pattern, and several isotopes split the line. The 327.4-nm line may be preferable and should be chosen. The characteristic mass for copper at the 327.4-nm line with Zeeman background correction is about half that with deuterium background correction. The reproducibility of the results with Zeeman background correction is improved when a transverse-heated graphite furnace is used instead of a longitudinal heated one.

One of the advantages in using ETAAS in trace-element analysis is that a complete digestion, minimizing residual carbon content, is not required. This elimination or reduction of sample digestion reduces the time needed for the analysis. Among the proposed methods are direct injection of the sample, dissolution of the sample in an adequate solvent, the use of a chemical modifier together with an emulsifying agent, and the use of slurries.

One of the facts to take into account in the direct analysis of a solid sample by ETAAS graphite furnace, is the sample weight. The use of too little or too much sample may lead to erroneous results, even if the analyte content lies within the linear dynamic range. Masses smaller than 0.3 mg or larger than 1.1 mg lead to overestimated and underestimated results, respectively.

Inductively coupled plasma-atomic emission spectrometry Inductively coupled plasma-atomic emission spectrometry (ICP-AES) provides, under optimal conditions, an approximate limit of detection of $1\,\mu g\,l^{-1}$. Although the cost of ICP-AES instrumentation has fallen considerably, AAS still remains the most widely available atomic technique for most laboratories.

ICP-AES can be used to measure copper ($\lambda = 324.7$ nm) in diluted biological fluids. In the case of water, a prior preconcentration by chelate formation and solvent extraction, or by adsorption on a column may be necessary. Detection limits of $0.07\,\mu g\,l^{-1}$ can be achieved when an ultrasonic nebulizer is used.

In the case of foods, samples have to be digested and preconcentrated or purified, or, in the case of solid samples such as cereals, samples are ground and mixed with graphite powder as buffer in a ratio of 2:3, after which an aliquot of the mixture is transferred to the powder sampler for introduction into the sealed ICP-arc system. The limit of detection is $0.02\,\mu g\,g^{-1}$.

The destruction of the matrix with one or more acids (e.g., HCl, $HClO_4$, and H_2SO_4) increases the possibility of spectral overlap. Therefore, whenever possible, only HNO_3 should be used in sample preparation, because H, N, and O do not add any additional spectral interferences to those already existing in a pure water spectrum.

Inductively coupled plasma-mass spectrometry Inductively coupled plasma-mass spectrometry (ICP-MS) is a powerful technique for trace multielement and isotopic analysis, because of its high sensitivity and ability to determine the isotope composition of a sample using less cumbersome pretreatment procedures than other mass spectrometry techniques.

However, ICP-MS is prone to interferences caused by sample matrix components. The interferences may be caused by polyatomic ions having the same nominal mass as the analyte, signal suppression or enhancement due to nonspectroscopic matrix effects and blockage of the nebulizer and sampler. Therefore, in order to achieve accurate and reliable results, matrix separation is needed when the matrix elements

in the prepared solution interfere with the determination. A detection limit of 4.8 $\mu g\,l^{-1}$ can be obtained.

Chemiluminiscence

FIA systems have been developed to determine micro-amounts of copper in well, rain, lake, and waste waters by using chemiluminiscence reactions. The systems luminol–CN^-–Cu(II) and luminol–$KMnO_4$–Cu(II) permit detection limits ranging from 10 $ng\,l^{-1}$ to 8.5 $\mu g\,l^{-1}$ to be achieved.

Potentiometric Methods

Electroanalytical techniques can be considered as an alternative to AAS or plasma techniques for measuring elements at trace levels. The main advantages of these techniques are an excellent sensitivity and the possibility of the simultaneous determination of different elements. Potentiometric methods also permit the speciation of the analyte, an important point from a nutritional and toxicological point of view.

The application of potentiometric techniques to the determination of an element content in foods requires complete organic-matter destruction. Once the sample has been converted into a solution, a sufficient amount of a base electrolyte should be added, and the oxygen has to be removed, since metals deposited on the electrode in the form of an amalgam or a film are usually very sensitive to oxidation, and oxygen interferes in the determination through its double-wave reduction.

The high sensitivity of the electrochemical methods requires the use of high-purity reagents and the adoption of measures to prevent sample contamination. Copper can be determined in food digests by differential pulse polarography with a stationary Hg-drop or a stick Au-base Hg film working electrode, a Pt counter-electrode and an Ag–AgCl reference electrode. The limits of detection of Cu are 0.45 μM, and the corresponding peak potentials are -0.71 V/-0.74 V/-0.65 V, depending on the supporting electrolyte used.

Anodic stripping voltammetry (ASV) is probably one of most suitable methods for determining copper in biological matrixes. In ASV, copper is selectively deposited out of the sample matrix on to a working electrode at a suitable working potential. After a selected period of deposition, the analytes collected on the working electrode are stripped back into the solution phase, and the oxidation currents are measured.

Different electrodes are available:

- Working electrode: hanging mercury drop electrode or a Hg film deposited on a 3-mm-diameter glassy-carbon rotating electrode;
- Reference electrodes Ag–AgCl or saturated calomel electrode (SCE); and
- Auxiliary/counter-electrode electrode: platinum (Pt) electrode.

The reference and auxiliary electrodes are connected with the sample solution via a liquid bridge filled with a solution of suprapure potassium chloride (KCl). The supporting electrolytes are: 1 M KCl, 2 M NH_4Ac/HAc, 1 M NH_3/NH_4Cl, 0.075 M $HClO_4$.

To determine copper by ASV, copper can be electrodeposited on a hanging/stationary mercury drop working electrode at -1.1 and -1.2 V vs. Ag–AgCl or SCE (with a platinum auxiliary electrode) for times ranging from 100 s to 4 min. The deposited metal can be determined by differential pulse ASV at $+0.04$ V. Detection limits at the $ng\,l^{-1}$ or $ng\,g^{-1}$ level have been reported.

The use of a selective electrode allows copper to be measured over a wide range of concentrations with an approximate practical detection limit of 10^{-8} $mol\,l^{-1}$. Copper ions in aqueous solution can be determined with an ion-selective membrane electrode containing a mixture of copper and silver sulfides. A supporting electrolyte, such as 2.5 M $NaNO_3$ is needed. The electrode response is linear from 1 μM to 1 mM Cu. The presence of Ag, Hg, Fe, and strong reducing agents (hydrazine 20–200 $\mu g\,l^{-1}$ does not affect the potential but reduces the life of the electrode.

Neutron Activation Analysis

Neutron activation analysis (NAA) is a nondestructive method that has been used for copper determination in biological material and bioavailability studies. The values obtained applying NAA correlate well with those obtained by AAS or ICP. NAA is not in common use, being restricted to specialized laboratories.

X-ray Fluorescence

X-ray fluorescence (XRF) techniques allow the determination of copper in very small samples (microsamples) of biological matrixes (e.g., blood serum, a hair, a cell). Samples have to be dried, milled, homogenized and compacted into disks. The destruction of the organic matter is not required. Copper content can be determined by direct XRF using a wavelength dispersive XRF spectrophotometer and a synthetic standard prepared by incorporating the analyte into cellulose. The detection limit is 5 μg of Cu per gram. There is a good agreement between the results obtained by XRF and NAA.

Speciation

ICP-AES and ICP-MS are element-specific detectors for high-performance liquid chromatography (HPLC) that have been used for copper speciation. With this aim, they can be coupled to size-exclusion chromatography for simultaneous monitoring of molecular weight fractions and their associated elements. HPLC coupled with ICP-AES has been used in copper speciation in foods such as soy bean flour extracts. Using size-exclusion chromatography coupled to ICP-AES in human milk, the distribution of copper between the different protein and non-protein fractions can be estimated.

Conclusions

AAS techniques are still the most widely used techniques to determine the copper content of foods and beverages. The most difficult and risky (because of contamination or analyte losses) step in copper determination in foods is sample preparation prior to the element measurement, a fact that is reflected in the numerous sample treatments found in the literature. ETAAS offers the advantage of not requiring complete organic-matter destruction, thus permitting the direct introduction of a sample, slurry, or acid extract.

However, ICP-AES offers the advantage of simultaneous element determination in a sample, as do potentiometric techniques, which also permit the differentiation between oxidation states, although, in the latter, complete organic-matter destruction is required. The use of appropriate reference materials is essential to check analytical accuracy. Currently, NAA remains useful as a reference method. The hyphenated techniques based on the coupling of a separation technique such as HPLC with an element detection will contribute to copper speciation in foods and beverages.

See also: **Mass Spectrometry**: Principles and Instrumentation; Applications; **Spectroscopy**: Fluorescence; Atomic Emission and Absorption; Visible Spectroscopy and Colorimetry

Further Reading

Belarra MA, Crespo C, Martínez Garbayo MP and Castillo JR (1997) Direct determination of metals in solid samples by graphite-furnace atomic absorption spectrometry: Does sample mass influence the analytical results? *Spectrochimica Acta* 52B: 1855–1860.

Bermejo-Barrera P, Domínguez-Gonzalez R and Bermejo-Barrera A (1997) Direct copper determination in whole milk, non-fat milk and whey milk by electrothermal atomic absorption spectrometry. *Fresenius Journal of Analytical Chemistry* 357: 457–461.

Boumans PWJM (ed.) (1987) *Inductively Coupled Plasma Emission Spectroscopy. Part II: Applications and Fundamentals*. New York: John Wiley.

Cao QE, Zhao Y, Cheng X, Hu Z and Xu Q (1999) Sequential fluorescent determination of copper (II) and cobalt (II) in food samples by flow injection analysis. *Food Chemistry* 65: 405–409.

Capelo JL, Filgueiras AV, Lavilla I and Bendicho C (1999) Solid–liquid extraction of copper from slurried samples using high intensity probe sonication for electrothermal atomic absorption spectrometry. *Talanta* 50: 905–911.

Cotton A and Wilkinson G (1989) *Advanced Inorganic Chemistry*, 5th edn. pp. 755–757. NewYork: John Wiley.

Dracheva LV (1990) Ionometric copper determination in aqueous media with use of a copper-selective electrode. *Gigiena i Sanitariya* 8: 81–83.

Gawalko EJ, Nowicki TW, Babb J, Tkachuk R and Wu SL (1997) Comparison of closed-vessel and focused open-vessel microwave dissolution for determination of cadmium, copper, lead, and selenium in wheat, wheat products, corn bran, and rice flour by transverse-heated graphite furnace atomic absorption spectrometry. *Journal of AOAC International* 80: 379–387.

Giglio L, Bocca AP and Caroli S (2000) Determination of the total content and binding pattern of elements in human milk by high performance liquid chromatography-inductively coupled plasma atomic emission spectrometry. *Talanta* 53: 295–303.

Hernández Córdoba M (1995) Copper. In: Townshend A (ed.) *Encyclopedia of Analytical Science*, vol. 2, pp. 852–858. London: Academic Press.

Hwang TJ and Jiang SJ (1996) Determination of copper, cadmium, and lead in biological samples by isotope dilution inductively coupled plasma mass spectrometry after on-line pre-treatment by anodic stripping voltammetry. *Journal Analytical Atomic Spectrometry* 11: 353–357.

Lavilla I, Filgueiras AV and Bendicho C (1999) Comparison of digestion methods for determination of trace minor metals in plant samples. *Journal of Agricultural and Food Chemistry* 47: 5072–5077.

Li GH, Zhang WG and Meng Z (1993) Determination of micro-amounts of copper in waste water by flow-injection analysis using the checking effect of the chemiluminescence reaction. *Lihua-Jianyan-Huaxue-Fenxi* 29: 290–291.

Lima EC, Barbosa F, Krug FJ, Silva MM and Vale MGR (2000) Comparison of ultrasound-assisted extraction slurry sampling and microwave-assisted digestion for cadmium, copper and lead determination in biological and sediment samples by electrothermal atomic absorption spectrometry. *Journal of Analytical Atomic Spectrometry* 15: 995–1000.

Price WJ (1983) *Spectrochemical Analysis by Atomic Absorption*. New York: John Wiley.

Slavin W (1984) *Graphite Furnace AAS: A Source Book*. Norwalk, USA: Perkin Elmer.

Physiology

M A Johnson, The University of Georgia, Athens, GA, USA

Introduction

Copper was identified as an essential nutrient early in the twentieth century. Today, research focuses on the molecular biology of copper metabolism and the role of copper in the heart, blood vessels, nervous system, immune function, formation of red blood cells, and bone health. A major challenge is the lack of a copper protein in blood that can reliably assess the copper status of humans. Although copper has been added to infant formulas for decades, copper is rarely added to enriched or fortified cereals consumed by older children and adults. An easily absorbed form of copper, such as copper sulfate, is added to only a few multiple vitamin/mineral supplements.

Body Content

Adult humans contain about 50–120 mg of copper, with the highest concentrations found in the liver, brain, heart, skeleton, hair, and nails (Table 1). Because of its size, muscle contains the largest percentage of body copper. Copper in liver, kidney, and intestine is strongly related to copper intake.

Table 1 Copper in human fluids, excretions, and tissues

Fluids and excretions	*Level*
Serum	0.5–1.5 mg l^{-1}
Sweat	20–100 μg l^{-1}
Urine	4–66 μg l^{-1}
	5–50 μg day^{-1}
Feces	0.5–2.5 mg day^{-1}
	20–100 mg kg^{-1} dry weight
Menstruation	0.5 mg per period
Hair	10–50 mg kg^{-1} dry weight

Tissues	*Concentration (mg kg^{-1} wet weight)*	*Estimated % of total body copper*
Bone marrow	5–7	15
Liver	5–7	8
Brain	5–7	8
Bone	2–3	19
Skin	2–3	15
Lung, heart, kidney	2–3	< 5
Muscle	< 2	25
Blood	< 2	5
Spleen	< 2	< 1

Reproduced from Johnson MA (1998) Copper: physiology, dietary sources and requirements. In: *Encyclopedia of Human Nutrition*. Academic Press, with permission.

Serum is the most widely used indicator of copper status, but it is not strongly related to the body stores of copper and becomes low only when the stores of copper are nearly exhausted. Similarly, copper in hair is not strongly related to the body levels of copper. Although sweat contains copper, it is difficult to collect and the relationship between body stores and sweat levels of copper has not been established. Sweat is not considered a major excretory route, unless sweat loss is extremely high. Losses of copper in the feces are proportional to intake, while urinary excretion is less than 5% of intake. Unlike for iron, menstrual losses do not deplete the body of copper.

Absorption, Transport, Storage, Excretion

Copper absorption in humans is usually estimated from copper intake minus copper excretion in the urine and feces. However, more accurate estimates of copper absorption and excretion are made by using a stable isotope of copper.

The primary site of copper absorption in humans and other animals is in the intestine in the duodenum. When fed a typical diet, human adults generally absorb about 50–75% of the copper ingested from food. The percentage absorption decreases as the amount of copper ingested increases. In adults, advanced age and gender do not influence copper absorption.

At the lower levels of copper intake, absorption is probably regulated by a saturable active transport mechanism, while at high levels of copper intake passive diffusion plays a role. Metallothionein is a metal-binding protein found in the intestine and other tissues that may also regulate copper absorption. Copper bound to metallothionein in the mucosal cell will be lost when these cells slough off the intestine. As will be discussed in more detail, copper absorption is negatively influenced by numerous dietary factors, including zinc, iron, molybdenum, ascorbic acid, fructose, and sucrose.

Copper is transported in the circulation bound primarily to albumin and to a lesser extent to transcuprein and low-molecular-weight compounds. Newly absorbed copper is transported to the liver and rapidly incorporated into ceruloplasmin. Ceruloplasmin is the transport protein from the liver to other organs. Some liver copper will also be incorporated into metallothionein, especially when copper intake is high. Metallothionein may help detoxify copper when copper exposure is high.

Gestation is a very important period of life for storage of copper. At birth the human liver has 5–10 times the adult concentration of copper. Liver copper gradually decreases during the first year of life as the infant uses liver copper for other functions in the

body. Premature infants are at high risk for copper deficiency unless they receive breast milk or copper-fortified formula.

Copper is lost from the body primarily through the bile and the feces, and retention of copper in the body is controlled primarily through regulation of copper absorption at the intestinal cells and excretion through the bile. Very little of the copper excreted into the bile is reabsorbed. Copper loss from the urine, skin, nails, and hair is very small, and, unlike some other minerals, the kidney does not regulate copper homeostasis.

Functions and Essentiality

Most of what we know about copper comes from livestock or from experimental animals fed diets deficient in copper. Some information is available from infants fed diets low in copper, healthy adults fed low amounts of copper, and adults given total parenteral nutrition without copper. Copper clearly is required for the immune system, the nervous system, the cardiovascular system, skeletal health, for iron metabolism, and for formation of red blood cells. However, the roles of copper in other functions such as the regulation of cholesterol, glucose, and blood clotting are not well understood. Some disorders of the vascular and nervous systems induced by copper deficiency during prenatal development in animals cannot be reversed by copper supplementation after birth. Thus, copper must be provided during pregnancy. Copper's role in cardiovascular and skeletal health also has important implications for major public health problems seen in industrialized nations, namely heart disease and osteoporosis.

Copper functions in several enzymes and proteins in the human body. Table 2 lists copper enzymes found in mammals. Cytochrome c oxidase is essential

Table 2 Human enzymes and proteins that contain copper

Enzyme or protein	*Function*
Cytochrome c oxidase	Mitochondrial enzyme involved in the electron transport chain; reduces oxygen to water and allows formation of adenosine triphosphate; activity is highest in the heart and also high in the brain, liver, and kidney
Ceruloplasmin (ferroxidase I)	Glycoprotein with 6–7 copper atoms; 4 copper atoms involved in oxidation/reduction reactions; role of other copper atoms not fully known; scavenges free radicals; oxidizes some aromatic amines and phenols; catalyzes oxidation of ferrous iron to ferric iron; assists with iron transport from storage to sites of hemoglobin synthesis; about 60% of plasma copper bound to ceruloplasmin; primarily extracellular; activity will be low during severe copper restriction
Ferroxidase II	Catalyzes oxidation of iron; no other functions known; in human plasma is only about 5% of ferroxidase activity
Monoamine oxidase	Inactivates catecholamines; reacts with serotonin, norepinephrine (noradrenaline), tyramine, and dopamine; activity inhibited by some antidepressant medications
Diamine oxidase	Inactivates histamine and polyamines; highest activity in small intestine; also high activity in kidney and placenta
Lysyl oxidase	Acts on lysine and hydroxylysine found in immature collagen and elastin; important for integrity of skeletal and vascular tissue; use of estrogen increases activity
Dopamine beta-hydroxylase	Catalyzes conversion of dopamine to norepinephrine, a neurotransmitter; contains 2–8 copper atoms; important in brain and adrenal glands
Copper, zinc superoxide dismutase	Contains 2 copper atoms; primarily in cytosol; protects against oxidative damage by converting superoxide ion to hydrogen peroxide; erythrocyte levels are somewhat responsive to changes in copper intake
Extracellular superoxide dismutase	Protects against oxidative damage by scavenging superoxide ion radicals and converting them to hydrogen peroxide; small amounts in plasma; larger amounts in lungs, thyroid, and uterus
Tyrosinase	Involved in melanin synthesis; deficiency of this enzyme in skin leads to albinism; catalyzes conversion of tyrosine to dopamine and oxidation of dopamine to dopaquinone. Present in eye and skin and forms color in hair, skin, and eyes
Metallothionein	Cysteine-rich protein that binds zinc, cadmium, and copper; important for sequestering metal ions and preventing toxicity
Albumin	Binds and transports copper in plasma and interstitial fluids; about 10–15% of copper in plasma is bound to albumin
Transcuprein	Binds copper in human plasma; may transport copper
Blood-clotting factors V and VIII	Role in clotting and thrombogenesis not well understood; part of structure homologous with ceruloplasmin

Reproduced from Johnson MA (1998) Copper: physiology, dietary sources and requirements. In: *Encyclopedia of Human Nutrition*. Academic Press, with permission.

for energy production in mitochondria. Ferroxidases such as ceruloplasmin (ferroxidase I) and ferroxidase II are copper enzymes. Many amine and diamine oxidases contain copper. Lysyl oxidase catalyzes the cross-linking of collagen and elastin; thus, copper deficiency causes bone and vascular abnormalities in many species. Superoxide dismutases are antioxidants. Copper deficiency in animals with dark hair can cause dark hair to lighten because of decreased activity of tyrosinase. Copper-binding proteins include metallothionein, albumin, transcuprein, and blood-clotting factors V and VIII. Copper also binds to low-molecular-weight ligands such as amino acids and small peptides.

Assessment of Copper Status

Severe copper deficiency is verified by low serum or plasma copper or ceruloplasmin, low red cell superoxide dismutase activity, anemia, and neutropenia. A major problem with using low ceruloplasmin as an index of copper status in people with chronic diseases such as cardiovascular disease, cancer, or inflammatory diseases is that ceruloplasmin synthesis is elevated during disease states and hence reflects disease state rather than copper status. Therefore, patients at risk for copper deficiency, such as those receiving total parenteral nutrition without copper, should have their serum copper routinely assessed.

Marginal copper deficiency is very difficult to detect and limits our ability to identify the role of copper in human health. Many copper proteins are easily measured in the blood, but their concentrations do not change very much when copper intake is increased or decreased. Studies at the US Department of Agriculture Human Nutrition Research Centers have identified the dietary intakes associated with decreases in copper proteins and other indices. Elderly women fed 0.57 mg copper per day for 105 days had low red cell superoxide dismutase activity, low platelet cytochrome c oxidase activity, low red cell glutathione peroxidase activity, and elevated plasma factor VII, but had no changes in plasma copper or ceruloplasmin. In contrast, young men fed 0.38 mg copper day^{-1} for 42 days had decreased plasma copper and ceruloplasmin concentration and activity. Thus, it appears that copper intakes must be below 0.6 mg copper day^{-1} for copper proteins to be decreased. Also, the magnitude of these decreases is not large enough to be used for routine assessment of healthy adults.

Ceruloplasmin and serum and plasma copper are increased by cigarette smoking, oral contraceptives, estrogen status, pregnancy, infectious disease, hematological diseases, diabetes, coronary and cardiovascular diseases, uremia, cancer, inflammation, and after surgery. Ceruloplasmin is an acute-phase reactant which accounts for its rise during numerous disorders and therefore limits the use of ceruloplasmin as a copper status marker, especially during illness.

Copper Deficiency in Infants

In infants, severe copper deficiency can be diagnosed by assessing plasma or serum copper, ceruloplasmin, and neutrophils. Sometimes copper deficiency is mistaken for iron deficiency because both mineral deficiencies are associated with anemia and low hemoglobin concentrations. However, only the anemia of copper deficiency responds to iron supplements.

Infants are more susceptible to severe copper deficiency than are children and adults. Risk factors for copper deficiency were identified in 52 copper-deficient infants (Table 3). Copper accumulates in the liver late during gestation. These liver stores are used in the first few months of life, because milk contains little copper. The highest risk for copper deficiency is in infants who are malnourished, premature, or of low birth weight. When these physiological conditions are combined with feeding practices such as cows' milk or total parenteral nutrition, the risk for copper deficiency can increase. Breast-feeding helps protect against copper deficiency because the copper in human milk is well utilized by infants.

Copper deficiency is rare among infants in developed countries because infant formulas are supplemented with copper. Also, it is recommended that copper be added to total parenteral nutrition solutions for infants as well as adults. However, where malnutrition is prevalent, copper deficiency may continue to be a

Table 3 Copper deficiency in infants

	% of cases
Predisposing factors	
< 18 months of age	96
Low birth weight (< 2500 g)	40
Fed exclusively or predominantly cows' milk	54
Received total parenteral nutrition	23
Formula-fed	12
Antecedent malnutrition	25
Physical features	
Hypotonia (loss of muscle tone)	43
Hypopigmentation	29
Psychomotor retardation	21
Skin rash	21
Dilated veins	15
Hepatosplenomegaly (complication of anemia)	15
Bone fractures	11

Median age at presentation was 8.3 months for full-term infants and 3.0 months for low-birth-weight infants. Reproduced from Johnson MA (1998) Copper: physiology, dietary sources and requirements. In: *Encyclopedia of Human Nutrition*. Academic Press, with permission.

problem. For example, clinicians in Chile reported that copper deficiency was more common than zinc deficiency and that supplements of copper improved the growth of 11 infants with malnutrition complicated by copper deficiency. It is believed that the cows' milk-based diet used during recovery from malnutrition in many developing countries does not provide enough copper for optimal growth and development.

Nutrient Interactions Influencing Copper Bioavailability

The bioavailability of copper is decreased by high intakes of several nutrients, including zinc, iron, molybdenum, ascorbic acid, sucrose, and fructose. It is well documented that high intakes of zinc interfere with copper absorption in humans and animals. The mechanism appears to involve the induction by zinc of metallothionein synthesis in the intestine and copper becomes 'trapped' in this protein and is not available for transport into the circulation. Zinc probably has little influence on copper status when the zinc to copper ratio is 15:1 or less. However, 50 mg of zinc day^{-1} has decreased some indices of blood copper status in humans. It is recommended that healthy adults limit their supplemental intake of zinc to less than 20 mg day^{-1}. The current recommended dietary allowance in the USA for zinc is 8 mg for women and 11 mg for men.

Although studies in animals have shown that high iron intakes interfere with copper status, there are not enough studies to determine if supplements of iron impair the copper status of people. Interactions between copper and molybdenum are well documented in ruminant animals, but little information is available for humans. High intakes of ascorbic acid induce copper deficiency in many species, but there is very little information in humans. In one study, 600 mg ascorbic acid day^{-1} did not decrease copper absorption, but in another study, 1500 mg ascorbic acid day^{-1} decreased the activity of ceruloplasmin. Supplemental intakes of ascorbic acid in excess of 1000 mg day^{-1} are common in North America, but the long-term impact of excess ascorbic acid on copper status is unknown.

Rats fed copper-deficient diets that contain 50% or more by weight fructose or sucrose develop more severe copper deficiency than when they are fed starch as the source of dietary carbohydrate. It is believed that in other species, including humans, the type of carbohydrate may not markedly influence the copper status of people. For example, in one human study, high intakes of fructose caused an increase in copper absorption but a decrease in red blood cell copper, zinc superoxide dismutase.

Combinations of copper antagonists interfere with copper status in animals. For example, high concentrations of sucrose exacerbate the effects of iron, zinc, and ascorbic acid on copper deficiency in rats. These type of relationships have not been evaluated in humans.

Genetic Diseases

Menkes disease, Wilson disease, and aceruloplasminemia are genetic disorders involving the regulation of copper and/or ceruloplasmin (Table 4). These

Table 4 Genetic disorders of copper metabolism

	Menkes disease	*Wilson disease*	*Aceruloplasminemia*
Genetics	Xq13.3 X-linked 1 in 300 000	13q14.3 Autosomal recessive, 1 in 30 000	3q 23–35 (ceruloplasmin), Autosomal recessive Frequency unknown
Clinical	Onset before birth, gray-matter degeneration, abnormal hair, hypothermia, hypopigmentation, tortuous arteries, early death (< 3 years)	Onset late in childhood, abnormal basal ganglia, liver disease, Kayser–Fleischer rings in eyes, psychiatric disorders	Adult onset, dementia, dystonia, diabetes, retinal degeneration
Laboratory	Low serum copper and ceruloplasmin, low liver copper	Low serum copper and ceruloplasmin, elevated liver copper	Low serum copper and ceruloplasmin, normal liver copper, elevated liver iron
Defect	Impaired copper transport in intestine and placenta	Impaired biliary excretion of copper	Impaired export of iron from reticuloendotheilial cells
Treatment	No effective treatment	Chelation very effective to remove body copper; zinc supplements help prevent further copper accumulation	Unknown
Animal models	Mottled mouse	LEC rat	None

Reproduced from Johnson MA (1998) Copper: physiology, dietary sources and requirements. In: *Encyclopedia of Human Nutrition*. Academic Press, with permission.

disorders are advancing the understanding of the cellular and molecular regulation of copper homeostasis. All three disorders are characterized by low serum copper and low serum ceruloplasmin; however, Menkes disease results in copper deficiency, Wilson disease leads to copper overload, and aceruloplasminemia causes iron overload in certain tissues.

Menkes disease is an X-linked disorder that causes mental deterioration, hypothermia, failure to thrive, connective tissue disorders, and death in early childhood. The advanced neuronal degeneration at birth suggests that the disease begins *in utero* and that copper is needed for fetal development of the nervous system. Copper uptake into cells is normal but copper-dependent enzyme activity is deficient. Placental and intestinal transport of copper is decreased. Neither oral supplements nor injections of copper will reverse this copper-deficient state. There may also be a milder type of Menkes disease that was formerly known as X-linked cutix laxa. Most patients reach adulthood, but have impaired physical and mental function.

Wilson disease is an autosomal recessive disorder that usually presents in adolescence with liver cirrhosis and neuronal degeneration. Copper accumulation in the liver and basal ganglia appears to account for tissue damage but the mechanism of toxicity is unknown. Copper chelation therapy is beneficial if the disease is diagnosed early. Supplements of zinc help limit copper absorption and subsequent accumulation. However, the neurologic symptoms are often irreversible and liver disease may be quite advanced at the time of diagnosis. The metabolic defect appears to be decreased biliary copper excretion and impaired incorporation of copper into newly synthesized ceruloplasmin.

Aceruloplasminemia is an autosomal recessive disorder of iron metabolism that was discovered in some people who were thought to have Wilson disease but did not follow the strict clinical criteria for the disease and did not have mutations in the Wilson P-type adenosine triphosphatase (ATPase). Their symptoms were somewhat similar to the dementia of Parkinson disease, combined with retinal degeneration and elevated liver iron, but normal liver copper. Basal ganglia had high iron concentrations as shown during magnetic resonance imaging of the brain. Mutations in the ceruloplasmin gene resulting in truncations of the open reading frame were found in these patients. Thus, it appears that abnormalities in the ceruloplasmin gene and its protein products cause disordered iron homeostasis.

The genes responsible for Menkes and Wilson disease do not involve ceruloplasmin. Rather, the genetic disorder appears to involve a transmembrane protein with homology to cation-transport P-type ATPases. There is 65% homology between the Wilson and Menkes P-type ATPases. Both contain an area proposed to bind copper. The Wilson gene appears to be expressed predominantly in the liver while the Menkes gene is expressed in most tissues other than the liver. The phenotypic differences in Wilson and Menkes disease are probably due to the tissue-specific expression of the genes. Evolutionary conservation of these types of proteins in bacteria, yeast, plants, and mammals indicates that a unique protein structure necessary for copper transport is preserved throughout the animal and plant kingdom.

Requirements and Recommended Intakes

The lack of a biochemical index of marginal copper status has made it very difficult to determine copper requirements in humans. When recommended dietary allowances in the USA were revised in 1989, there were insufficient data to make a firm recommendation for copper. However, in 2001 Dietary Reference Intakes in the USA were estabilished (Table 5). For adults, the Recommended Dietary Allowance for adults is $0.9\,mg\,day^{-1}$ and the upper level is $10\,mg\,day^{-1}$. Liver damage was the endpoint used to estimate the upper level. The European Commission has also established population references for several nutrients, including copper.

Dietary Sources

The amount of copper in various foods varies by about 100-fold and depends on where a given food was produced and how it was processed (Table 6).

Table 5 Dietary reference intakes for copper

Age (years)	*Recommended dietary allowance (mg day^{-1})*	*Upper intake level (mg day^{-1})*
0–0.5	0.20[a]	–
> 0.5–1	0.22[a]	–
1–3	0.34	1
4–8	0.44	3
9–13	0.70	5
14–18	0.89	8
19–50	0.90	10
≥ 51	0.90	10
Pregnancy		
14–18	1.0	8
19–50	1.0	10
Lactation		
14–18	1.3	8
19–50	1.3	10

[a]Adequate intake.

From: Food and Nutrition Board (2001) Copper. In: Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Mangnese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc, pp. 224–257. National Academy of Sciences. Washington DC: National Academy Press.

Table 6 Copper in foods

Food	*Copper content (mg kg^{-1})*
Dairy	
Cheese	0.4–0.8
Chocolate milk	0.2–0.3
Cottage cheese	0.1–0.2
Cows' milk (skim, 2%, whole, buttermilk)	0.02–0.08
Human milk	0.02–0.08
Yogurt	0.01–0.09
Eggs (fried, scrambled, or soft-boiled)	0.4–0.8
Fish and shellfish (cooked)	
Cod, fish sticks, haddock, salmon, sardines, tuna	0.3–0.8
Shrimp	2–3
Oysters	0.3–16
Fruits	
Apples (red, with peel, raw)	0.1–0.4
Apple juice (canned)	0.02–0.2
Banana (raw)	1–2
Grapes (raw, purple, or green)	0.4–1.4
Grape juice (canned)	0.01–0.13
Orange (raw, all varieties)	0.1–1
Orange juice	0.1–0.3
Dried fruits (currants, dates, figs, prunes, or raisins)	1–5
Grains and cereals (cooked or processed)	
Barley, pearl	0.4
Bread	
White (loaf, rolls, biscuits)	1–2.6
Whole wheat	2–3
Rye	1.6–2.8
Corn (fresh, frozen, cream-style, grits)	0.1–0.4
Flour	
Whole grain	2–8
White	1–3
Macaroni (cooked)	0.6–1
Oatmeal (cooked)	0.3–1.2
Wheat bran	10–20
Wheat cereals (shredded, bran flakes, puffed wheat, raisin bran)	4.5–5.5
Legumes	
Beans (boiled or baked; cowpeas, kidney, lima, navy or pinto)	1–4
Beans (dry; cowpeas, kidney or lima)	5–10
Peanuts (fresh, roasted or butter)	3–10
Meats (cooked: beef, pork and poultry)	
Muscle meat	0.7–1.4
Liver	20–180
Nuts and seeds (almonds, brazil nuts, hazelnuts, pecans, pistachios, sesame seeds, sunflower seeds, or walnuts)	8–18
Vegetables	
Broccoli (fresh, frozen, or boiled)	0.1–0.9
Cabbage (fresh, boiled, or coleslaw)	0.1–0.2
Carrots (raw or boiled)	0.5–1
Cauliflower (fresh, frozen, or boiled)	0.2–1
Onions (raw or cooked)	0.2–1
Peas (green, canned, cooked)	1–1.5
Potatoes	
Baked (with peel)	0.6–1.8
Boiled (without peel)	0.3–1
Spinach (canned, fresh, frozen, or boiled)	0.6–1.2
Sweet potatoes (boiled or baked)	1.6–2.0
Tomatoes (raw, canned or juice)	0.3–1.4

Reproduced from Johnson MA (1998) Copper: physiology, dietary sources and requirements. In: *Encyclopedia of Human Nutrition*. Academic Press, with permission.

Variation in copper content may be the result of soil conditions, type of fertilizer or other agricultural chemicals, weather, time of harvesting, and processing. The amount of copper in a typical serving can be approximated by using portion sizes of 250 g for fluid milk and juices, 120 g for small servings of fish or meat, 100 g for fruits, vegetables, and legumes and 30 g for a slice of bread, small bowl of cereal, or slice of cheese. Copper concentrations are highest in legumes, nuts, and shellfish, and lowest in dairy foods such as cows' milk. Removal of the germ and bran during refining of flour decreases the copper content of grain foods. Copper in drinking water is quite variable and depends on the natural mineral content, the pH of the water, and the use of copper water pipes. Acidic water that flows through copper pipes will pick up some copper. Estimates of the contribution of water to copper intake range from less than 10% to over 50% of the recommended intakes.

In the USA, copper is added to some, but not all, multiple-vitamin/mineral supplements. Cupric oxide is one of the most common sources of copper used in multiple-vitamin/mineral supplements; however, its solubility and bioavailability are probably very poor for humans (but quite good for ruminants). It is recommended that copper sulfate be used in these multinutrient preparations. Most infant formulas are supplemented with copper in the form of copper sulfate. Copper is not routinely added to ready-to-eat breakfast cereals intended for older children and adults.

Copper Toxicity

Tolerance to high intakes of copper varies considerably among species. For example, rats tolerate high amounts of copper, while sheep are intolerant of high intakes of copper. Soluble salts of copper (carbonates, sulfates, and chlorides) are more toxic than the less soluble forms (nitrates, acetates, and oxides). High intake of zinc protects again copper toxicity in many species.

Acute toxicity from copper is rare in humans and usually occurs from accidental or intentional ingestion of copper salts. Acute symptoms include salivation, nausea, vomiting, epigastric pain, and diarrhea. Ingestion of about 100 g or more of copper sulfate can produce hemolytic anemia, liver failure, renal failure, shock, coma, or death.

Data are quite limited to make a scientifically based safe upper limit of copper intake for humans.

Table 7 Regulations regarding total intakes of copper and copper content of drinking water

Agency	*Upper limit*	*Rationale*
Water content		
World Health Organization (WHO)	2.0 mg l^{-1}	Attempt to keep contribution of water to no more than 10% of total copper intake
European Commission (EC)	None	Suggested guide of 0.1 mg l^{-1} at water treatment plant, and 3 mg l^{-1} after water has stood in pipes for 12 h
US Environmental Protection Agency (EPA)	1.3 mg l^{-1}	Based on a 'lowest observed adverse-effect level' of 5.3 mg in humans (nurses, study), consumption of 2 l water daily, and addition of safety factors
	1.0 mg l^{-1}	Nonenforcible concentration of 1 mg l^{-1} based on taste considerations
Total intake		
Food and Agricultural Organization/World Health Organization (FAO)/WHO)	0.5 mg kg^{-1}	Based on no-observed-effect level of 5 mg kg^{-1} in dogs
US National Academy of Science	10 mg day^{-1}	
European Commission	10 mg day^{-1}	Risk of nausea and vomiting increases when this dose is exceeded

Reproduced from Johnson MA (1998) Copper: physiology, dietary sources and requirements. In: *Encyclopedia of Human Nutrition*. Academic Press, with permission.

The copper intake of nurses was roughly estimated at between 5.3 and 32 mg copper, and the dose of 5.3 mg copper day^{-1} has been adopted by some authorities as the 'lowest observed no-effect level'. Other authorities have suggested 0.21–0.5 mg kg^{-1}, or up to 30 mg of copper in a 60-kg person, as the maximum tolerable daily total intake for copper. Proposals for the maximal allowable amount of copper in drinking water are in the range of 1–2 mg l^{-1}, but based on copper poisoning in children, the safe amount for children may be as low as 0.2 mg l^{-1}. This wide range of estimates is attributed to insufficient data, lack of understanding by toxicologists and regulatory authorities that copper is an essential nutrient, lack of awareness that humans are exposed to copper from both foods and beverages, and concern about copper toxicosis syndromes such as childhood cirrhosis. Regulations for copper intakes and copper in drinking water are summarized in Table 7.

Indian childhood cirrhosis was a common killer until the early 1990s when it became preventable and treatable. The disease generally occurs between the ages of 6 months and 5 years, and is characterized by increased hepatic, urinary, and serum copper concentrations. Symptoms include nonspecific complaints such as abdominal distension, irregular fever, excessive crying, and altered appetite. Jaundice is usually a late feature. If left untreated, the illness progresses rapidly within a few months and is associated with liver failure and possibly death. The source of the dietary copper appears to be milk boiled in copper or brass (an alloy with about 70% copper and 30% zinc) vessels that have been used extensively in India, particularly by traditional Hindu families. Some investigators also believe that a genetic disorder enhances susceptibility to this toxicity syndrome associated with excessive dietary exposure to copper. Public health campaigns in India designed to discontinue the use of copper-containing vessels markedly reduced the incidence of this disease.

Contamination of drinking from copper plumbing has emerged as a public health concern. Copper-associated liver diseases in infants and children, similar to the condition seen in India, have now appeared in parts of Europe. The source of copper appears to be drinking water. Treatment for this liver disease involved D-pencillamine (a copper chelator) and in a few cases a liver transplant was conducted. In an investigation in Wisconsin (USA), one water system involved had an alkaline pH of 7.2–7.4 that would not be expected to be corrosive. However, phosphate was added to sequester iron, and this may have prevented calcium and magensium from depositing on the inner surface of the copper pipes and providing a protective coating. Thus, copper was more easily leached into the water from the pipes. Homes built or remodeled within the past 10 years had the highest concentrations of copper in the water. In a second location, symptoms began after installation of new copper pipes in a mobile-home park and the water was naturally corrosive and very low in mineral content. In both areas, the symptoms were nausea, diarrhea, abdominal cramps, and headaches. The copper concentration of the water associated with toxicity was up to 3.6 mg l^{-1}. Flushing the faucet for 1 min decreased the copper content to below 0.25 mg l^{-1}. The authors suggested that owners of homes with new copper pipes should be informed about the risks of copper-contaminated drinking water and have their water tested. They also recommended that copper intoxication be included in the differential diagnosis for those presenting with chronic or intermittent

gastrointestinal upsets, particularly when these symptoms can be linked with a change in the water supply. Infants may be particularly sensitive to high intakes of copper and some investigators have recommended that drinking water contain no more than $0.2\,mg\,l^{-1}$.

See also: **Copper**: Properties and Determination; **Dietary Reference Values**

Further Reading

Failla ML, Johnson MA and Prohaska JR (2001) Copper. In: Bowman BA and Russell RM (eds) *Present Knowledge in Nutrition*, 8th edn., pp. 373–383. Washington, DC: ILSI Press.

Food and Nutrition Board (2001) Copper. In: Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc, pp. 224–257. National Academy of Sciences. Washington DC: National Academy Press.

Hardorn H-B (1999) Copper-associated liver diseases in children. Introduction. *European Journal of Medical Research* 4: 212–213.

Harris ZL and Gitlin JD (1996) Genetic and molecular basis for copper toxicity. *American Journal of Clinical Nutrition* 63: 836S–841S.

Harris ED (1997) Copper. Chapter 8. In: O'Dell BL and Sunde RA (eds) *Handbook of Nutritionally Essential Mineral Elements*. pp. 231–273. New York: Marcel Dekker.

Johnson MA (1998) Copper: physiology, dietary sources and requirements. In: *Encylopedia of Human Nutrition*, pp. 442–450. London: Academic Press.

Johnson MA, Smith MM and Edmonds J (1998) Copper, iron, zinc and manganese in dietary supplements, infant formulas, and ready-to-eat breakfast cereals. *American Journal of Clinical Nutrition* 67: 1035S–1042S.

Lönnerdal B (1998) Copper nutrition during infancy and childhood. *American Journal of Clinical Nutrition* 67 (suppl.): 1046S–1053S.

Olivares M and Uauy R (1996) Limits of metabolic tolerance to copper and biological basis for present recommendations and regulations. *American Journal of Clinical Nutrition* 63: 846S–852S.

Pandit A and Ghave S (1996) Present interpretation of the role of copper in Indian childhood cirrhosis. *American Journal of Clinical Nutrition* 63: 830S–835S.

Scheinberg IH and Sternlieb I (1996) Wilson disease and idiopathic copper toxicosis. *American Journal of Clinical Nutrition* 63: 842S–845S.

Turnlund JR (1999) Copper. Chapter 12. In: Shils ME, Olson JA, Shike M and Ross AC (eds) *Modern Nutrition in Health and Disease*, 9th edn, pp. 241–252. Baltimore, MD: Williams & Wilkins.

CORONARY HEART DISEASE

Contents

Etiology and Risk Factor

F G Huffman, S Koutoubi and S Nath, Florida International University, Miami, FL, USA

Background

Our understanding of the biology and pathophysiology of cardiovascular medicine and pathogenesis of coronary heart disease (CHD), a multifactorial disease with a complex pathophysiology, continues to expand as new horizons in epidemiological, experimental, clinical, and genetic research are required. Even so, CHD remains the leading cause of death, accounting for more than 7 million of all cases of mortality, and is responsible for chronic disability and much of the healthcare costs worldwide. The World Health Organization predicts that CHD will prevail as the primary cause of morbidity and mortality worldwide until the year 2020.

No single risk factor consummates the 'cause and effect' relationship. That the presence of any one isolated risk factor cannot predict which individual will develop CHD. To date, data from myriad

Table 1 Etiology of coronary heart disease (CHD)

Nonmodifiable risk factors
Age and gender (men $\geq$ 45 years; women $\geq$ 55 years)
Family history of premature CHD (CHD in male first-degree relative < 55 years; CHD in female first-degree relative < 65 years)
Modifiable risk factors
Cigarette smoking
Atherogenic diet
Physical inactivity
High blood pressure ($\geq$ 140/90 mmHg)
Obesity
Diabetes mellitus
High blood total cholesterol ($\geq 200\,mg\,dl^{-1}$)
High blood low-density lipoprotein cholesterol ($\geq 130\,mg\,dl^{-1}$)
Low blood high-density lipoprotein cholesterol ($< 40\,mg\,dl^{-1}$)
High blood triglycerides ($\geq 150\,mg\,dl^{-1}$)
Thrombogenic factors
Inflammation or infection
Cytomegalovirus
Chlamydia pneumoniae
Helicobacter pylori
Herpes simplex
C-reactive protein
Serum amyloid A protein
Others
High blood homocysteine
Angiotensin II

research have documented several hundred risk factors that are statistically associated with the development of CHD. Experts in cardiovascular research often agree to classify these etiological factors as major and minor risk factors or modifiable and nonmodifiable risk factors (Table 1). The majority of the CHD epidemic in developed countries are associated with the presence of smoking, high blood pressure, high blood cholesterol, and impaired glucose tolerance. Most often, these common risk factors tend to cluster in individuals who eventually develop CHD events.

Atherosclerosis, hardening of the arterial wall, commences in childhood and, after decade-long cellular, molecular, and biochemical changes, culminates in a heart attack, a well-known indication of CHD. In 1973, the 'response-to-injury' hypothesis of atherosclerosis was first outlined, underlining that the atherosclerosis results from an arterial response to chronic injury. Commonly chronic endothelial cell injury is initiated by tobacco toxins (smoking), high blood pressure, diabetes, oxidized low-density-lipoprotein (LDL) cholesterol, as well as free radicals resulting from oxidation of homocysteine (Hcy), and infection (e.g., chlamydia). In response to an injury to endothelial cells, there is a downregulation of protective mechanisms and upregulation of several factors leading to atherosclerosis. For example, at the site of endothelial injury, there is downregulation of the potent endogenous vasodilator nitric oxide (NO), tissue plasminogen activator (tPA), and prostacycline, leading to a hyperprothrombic environment characterized by upregulation of platelet aggregation and leukocyte adhesion. At the initial stage, circulating monocytes are attracted to the vascular intima, the battleground of the atherosclerotic process. It is worth mentioning that vascular cell adhesion molecule, usually expressed on the surface of endothelial cells due to an inflammatory process, always precedes the binding of monocytes. Monocytes then squeeze through the tight junctions of the endothelial cells, enter the subendothelial layer of the intima and are transformed to phagocytic state macrophage. Macrophages engulf oxidized LDL cholesterol and form the foam cells, the integral part of the 'fatty streak.' At one stage, smooth muscles cells from the tunica media migrate to the intima and release an extracellular matrix. Monocytes, lymphocytes, platelets, endothelial cells, and smooth muscle cells liberate numerous factors, such as platelet-derived growth factors (α, β), fibroblast growth factor, insulin-like growth factor-1, interleukins (IL)-1, -2, -6, and -8, tissue necrosis factor (TNF), which influence further cellular migration and proliferation. The size of the fatty streak increases with the accumulation of different types of cells and lipids. Eventually, calcium is deposited and forms an atherosclerotic plaque (Figures 1 and 2).

The following subsections detail each risk factor as it relates to CHD.

Age, Gender, and Menopausal Status

The absolute risk of CHD increases with age, as there is a progressive accumulation of coronary atherosclerosis with aging. Coronary heart disease is most common after the age of 65 years, irrespective of gender. Men however have a greater risk of CHD than women; with men developing atherosclerotic lesions 20 years earlier than women. Behavioral and physiological differences between men and women such as prolonged exposure to smoking, dislipidemia and hormonal differences may partly explain the gender difference. Among premenopausal women, CHD is rare (except among smokers and oral contraceptive users). However, the CHD risk is precipitous among postmenopausal women, approaching that for men.

Family History

Based on genetic data, twin studies, and phenotypes (measurable traits) linked by inheritance, a positive family history of CHD is one of the most potent etiological influences on CHD independent of other major risk factors such as smoking, high blood

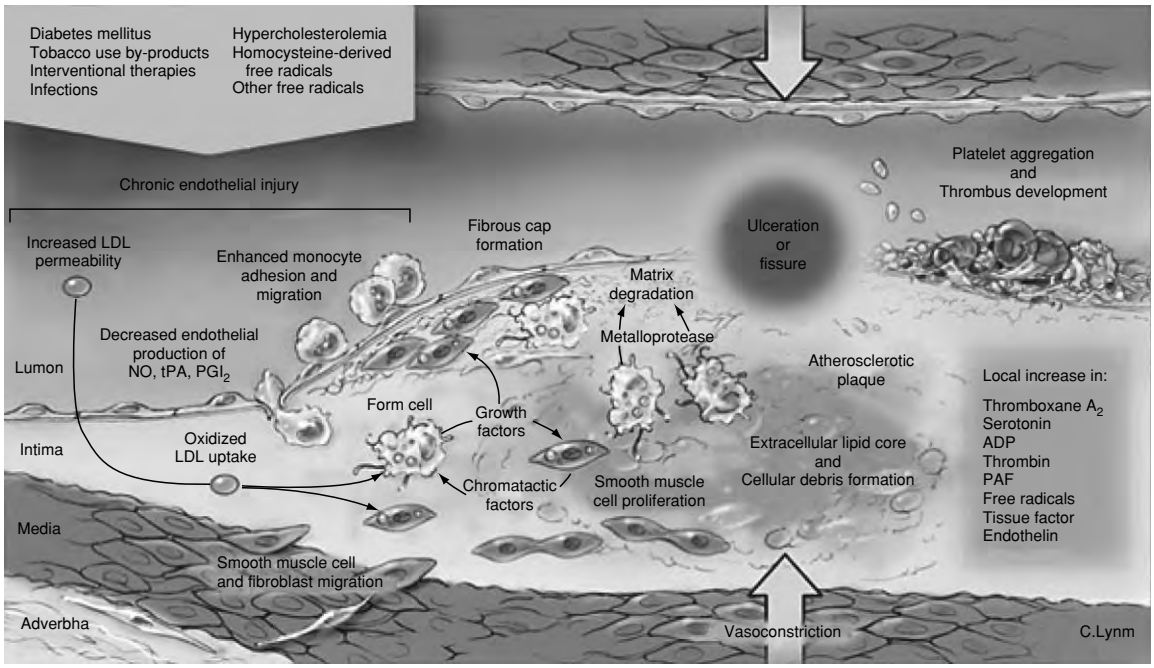

Figure 1 (see color plate 39) Mechanisms in the initiation and progression of atherosclerosis. Endothelial injury results from a variety of factors such as smoking, high blood cholesterol and homocysteine, and ulceration of atherosclerotic plaques, etc. (see text). At the site of endothelial injury, there is a downregulation of endothelial-derived nitric oxide (NO), tissue plasminogen activator (tPA), and prostacyclin (PGI_2) with increased adhesion of platelets and leukocytes, increased permeability to lipoproteins, and vasoconstriction. ADP, adenosine diphosphate; LDL, low-density lipoprotein; PAF, platelet activating factor. From Lefkowitz RJ and Willerson JT (2001) Prospects for cardiovascular research. *Journal of the American Medical Association* 285: 581–587 with permission.

pressure, diabetes, and high LDL cholesterol. For example, a single gene mutation (genetic variation) in lipid metabolism is well documented in the development of CHD. Plasma lipoprotein(a), Lp(a), is a new risk factor for CHD. In normal population, the variability of the *apo(a)* gene accounts for 90% of the variability of plasma Lp(a). African Americans have a higher levels of Lp(a) than white Americans. French Canadians also have a higher level than expected due to relative genetic homogeneity, also known as 'founder effects.' Familial combined hyperlipidemia, another well-known lipid disorder, may cause CHD before the age of 60 years, with a prevalence of about 2%.

The risk of CHD increases in an individual with a history of CHD in a first-degree relative (parent, brother, sister, son, or daughter). The family history is considered positive if clinical CHD or sudden death can be documented in first-degree male relatives before the age of 55 or in first-degree female relatives before the age of 65.

Smoking

Cigarette smoking is one of the most important risk factors for the development and progression of CHD. It has been estimated that smoking is responsible for about 30% of CHD-related deaths. Coronary heart disease events and deaths are attributable to cigarette smoking in both a dose- and duration-dependent fashion. Coronary heart disease risk is two to four times higher in smokers than in non-smokers. Inhaled cigarette smoke contains thousands of substances, many of which are toxic to cells. For example, the extensively studied chemical nicotine appears to be involved in endothelial cell death, diminished endothelial cell mitosis rate, and inhibition of NO synthesis. Detrimental effects of smoking on the cardiovascular system include an acute increase in vascular resistance and blood pressure, impaired oxygen transport, constriction of coronary vessel walls, acceleration of LDL cholesterol oxidation, decreased HDL cholesterol, increased platelet aggregation, and increased fibrinogen production. Smoking also acts synergistically with other etiological factors for CHD, thereby substantially increasing the risk. Passive smoking or environmental exposure also carries a substantial risk in the development of CHD. Like first-hand smoke, environmental smoke contains many toxins, such as carbon monoxide, benzopyrene, and hundreds of other chemicals. Many of these substances are toxic to arterial walls and may cause endothelial dysfunction and substantially reduce coronary microcirculation, an early process of atherosclerosis.

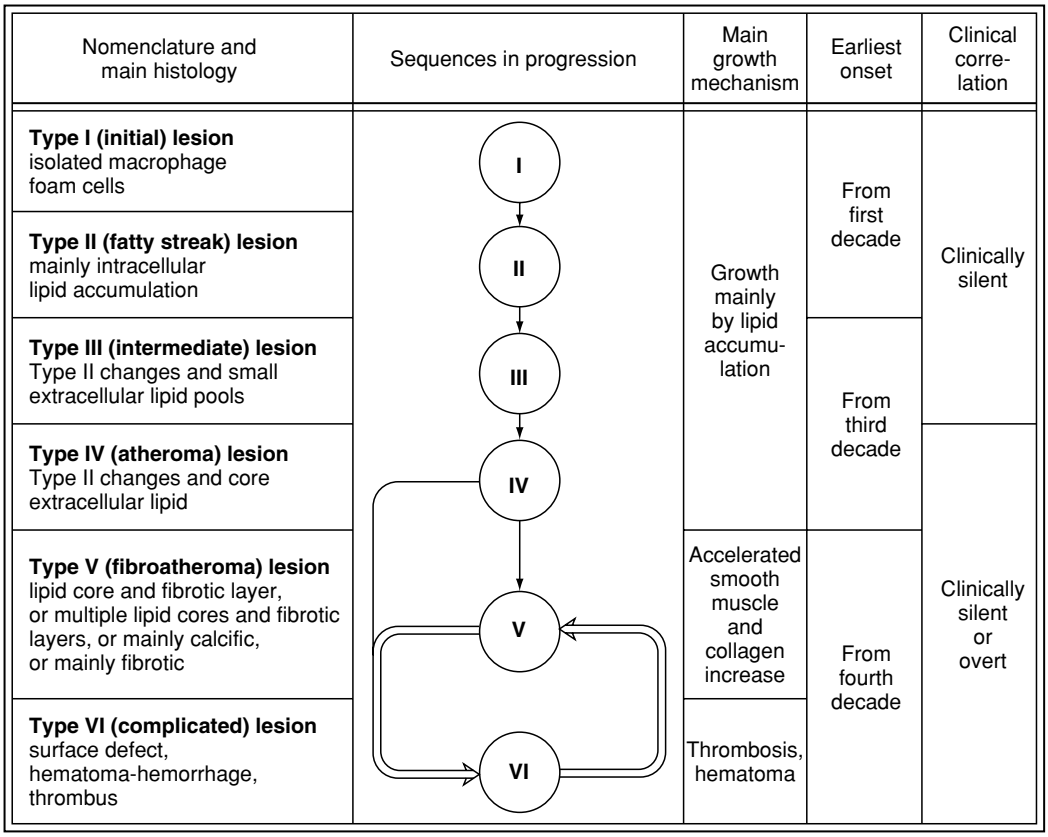

Figure 2 Flow diagram of the progression of atherogenesis. Roman numerals indicate histological characteristics of lesions. Loop I to loop IV illustrates the morphological changes in the lesion due to lipid accumulation. The loop between V and VI illustrates how lesions increase in thickness when thrombotic deposits form on their surface. From Stary HC *et al*. (1995) A definition of advanced types of atherosclerotic lesions and a histological classification of atherosclerosis. *Arteriosclerosis, Thrombosis, and Vascular Biology* 15: 1512–1531 with permission.

Psychosocial Risk Factors

There is mounting evidence that exposure to environmental stressors and psychological reactions to stress, an individual's personality prototype (type A), lack of social support, depression, and low socioeconomic condition are associated with increased CHD risk. Type A behavior is of special interest since it is associated with a two- to threefold higher risk for acute myocardial infarction, as documented by substantial prospective and case-control studies. A comprehensive explanation of the effects of psychosocial factors on CHD risk factors (e.g., high blood pressure) has not yet been established. Psychosocial factors may act in different ways. For example, they may be related to risky behaviors or lifestyle factors such as smoking, poor diet, physical inactivity, or excessive alcohol consumption, which in turn influence the development of CHD. However, the essential role of the psychosocial factors in developing CHD event is mediated by a neuroendocrine mechanism that alters catecholamine levels. Excessive adrenergic (sympathetic) activity may cause platelet aggregation, leading to occlusive arterial thrombus formation. It is also hypothesized that psychosocial stress may also increase circulating IL-6, which in turn acts on the hypothalamic–pituitary–adrenal axis, releasing many hormones that eventually result in abdominal obesity, insulin resistance, and various lipid disorders, the risk factors for CHD.

Atherogenic Diet

Many decades of research have established that diet plays a key role in atherosclerosis and subsequent events such as myocardial infarction. The rates of CHD are strongly associated with atherogenic diets, which influence other biological factors such as high blood pressure, LDL cholesterol, HDL cholesterol, thrombogenic factors, and obesity. The diet also affects CHD in other ways: high-calorie, high-saturated-fat (especially 12–16 carbon atoms) diets, combined with physical inactivity or a sedentary lifestyle, may lead to obesity, diabetes, and high blood pressure. Experimental models have proved that dietary saturated fat may cause atherosclerotic changes even in the absence of other classical etiological

factors for CHD. In contrast, substitution of saturated fat with isocaloric unsaturated fatty acids and complex carbohydrates, especially soluble fiber, lowers LDL cholesterol. Partial hydrogenation of polyunsaturated fatty acids (e.g., margarine) increases LDL cholesterol. Dietary cholesterol has a small effect on elevating LDL cholesterol. A diet high in protein can substantially limit carbohydrate intakes. Proteins from animal sources are generally higher in fat, saturated fat, and cholesterol and are considered atherogenic. Excessive salt intake is associated with high blood pressure in susceptible persons. Dietary supplementation of β-carotene (20 mg per day), a precursor of vitamin A, may increase cardiovascular mortality by 12–26%, especially among male smokers. Many epidemiological studies have shown a J- or U-shaped relationship between alcohol consumption and all-cause mortality. Excessive alcohol consumption is associated with increased CHD, possibly because of an increase in blood pressure and cardiac arrhythmias. In contrast, light to moderate alcohol consumption is cardioprotective, since it increases HDL cholesterol and apolipoprotein A1, and has favorable effects on markers of inflammation (e.g., C-reactive protein and IL-6).

Physical Inactivity

Prospective epidemiological studies have shown that both physical inactivity and a sedentary lifestyle are independent etiological factors and predictive of increased CHD risk. It is, however, difficult to quantify the association between the amount of physical activity and CHD. Results from meta-analyses have concluded that physical activity may lower the rate of CHD events by 50%. The positive benefits of physical activity on CHD outcomes may be partially mediated by lowering high blood pressure, decreasing triglycerides and LDL cholesterol, increasing HDL cholesterol, and increasing insulin sensitivity by lowering blood glucose levels.

High Blood Pressure

Although high blood pressure or hypertension is defined as a systolic blood pressure of 140 mmHg or greater and a diastolic blood pressure of 90 mmHg or greater, this definition is arbitrary. Several large-scale epidemiological studies have consistently demonstrated a positive, strong, continuous, graded, and independent relationship between high blood pressure and the etiology of CHD. The prevalence of high blood pressure is approximately 50% in the persons aged 65 years or older in many countries. High blood pressure promotes the atherogenesis process by several mechanisms: increased endothelin production, increased leukocyte adherence, impaired endothelium-derived relaxant factor production, increased endothelium-derived contractile substances, and increased vascular permeability to lipoproteins. It has been shown that in essential hypertension, endothelial damage inhibits NO production, and there is production of endothelin a potent vasoconstrictor from the activated endothelium. In addition, high blood pressure may also cause structural remodeling of vascular walls (cell proliferation) and accumulation of smooth muscle cells in the tunica intima, which in turn initiates a series of cascades leading to atherogenesis.

Obesity

Obesity, particularly central or abdominal obesity, is considered one of the prime etiological risk factors for CHD, since the adverse effects of obesity on CHD risk factors (blood pressure, lipid profiles, and glucose tolerance) are numerous. Central obesity is defined as a waist-to-hip ratio of > 102 cm in men and > 88 cm in women. Among the several adverse effects of obesity in mediating CHD, the most profound effect is high blood pressure. It has been estimated that more than 75% of high blood pressure is directly attributed to obesity. It is well documented in medical literature that blood pressure increases with weight gain and decreases with weight loss. Another significant negative effect of obesity is on lipid metabolism; obesity increases triglycerides and LDL cholesterol, and decreases HDL cholesterol. Although it is difficult to establish the exact contribution of obesity to glucose intolerance or insulin resistance, the role of excess body fat, particularly abdominal fat, is positively correlated with insulin resistance.

Diabetes Mellitus

Individuals with either type 1 (insulin-dependent diabetes mellitus) and type 2 (noninsulin-dependent diabetes mellitus) are at higher risk for CHD than nondiabetic individuals. Atherosclerosis is responsible for more than 75% of all mortality among diabetics. Silent myocardial ischemia commonly occurs in patients with diabetes. Data from longitudinal epidemiological studies suggest that diabetes or hyperglycemia (random plasma glucose concentration $\geq 200\ \mathrm{mg\,dl^{-1}}$ or fasting plasma glucose concentration $\geq 125\ \mathrm{mg\,dl^{-1}}$) is an independent and well-established etiological factor for CHD. Diabetes, especially type 2 diabetes, is a strong CHD risk factor in women which increases CHD-related mortality by 400%.

The relationship between diabetes and CHD risk is complex, poorly understood, and yet to be elucidated. Diabetes alters endothelial cell function by contributing several mechanisms such as smooth muscle cell proliferation and extracellular matrix production, elevation of endothelin and angiotensin II, enhance platelet adhesion, and decrease in tPA. All these changes

result in the progression of atherosclerosis. In the case of type 2 diabetes, the hyperinsulinemic state exerts mitogenic effects on the smooth muscle cells, leading to an alteration of endothelial structure. In addition, in the pathogenesis of atherosclerosis process, hyperglycemia and other CHD risk factors such as smoking, high blood pressure, hyperlipidemia, obesity, physical inactivity, and insulin resistance are interrelated and intertwined and continue to act independently in patients with diabetes. The majority of patients with type 2 diabetes have insulin resistance, which indeed predisposes then to CHD. There are other factors such as abdominal obesity, elevated triglycerides and LDL cholesterol, low HDL cholesterol, high blood pressure, and alterations in coagulation mechanisms that tend to cluster among type 2 diabetics, collectively known as metabolic syndrome. Elevated plasminogen activator inhibitor-1 (PAI-1) concentration, a novel risk factor of CHD, is another potentially important thrombogenic factor in this syndrome.

Blood Lipids

The essential roles of blood cholesterol and triglycerides in the etiology of CHD are well established. There is a continuous and graded risk between total blood cholesterol and CHD in population-based studies. However, on an individual basis, total cholesterol may be used as a risk factor in determining CHD. In fact, most of the lipid-related risk is explained by the LDL cholesterol levels, and the positive association between high LDL cholesterol ($\geq 160\,\mathrm{mg\,dl^{-1}}$) and the risk of CHD is applicable to individuals with or without established CHD, in both men and women. In epidemiological studies, LDL cholesterol increases as total cholesterol increases. Low-density lipoproteins increase with age, weight gain, and atherogenic diets. Genetic factors such as the heterozygous form of familial hypercholesterolemia further increase LDL cholesterol in the general population. The oxidized form of LDL is the main culprit in atherogenesis. Scavenging macrophages engulf oxidized LDL particles, resulting in the formation of lipid-laden foam cells. Subsequently, in combination with smooth muscle proliferation, the lipid-laden foam cells are transformed to a fatty streak, as mentioned earlier. Oxidized LDL cholesterol also inhibits relaxation factors and promotes contraction factors leading to modification of vascular tone, which eventually leads to vasospasm and thrombus formation.

A low level of HDL cholesterol is a strong independent etiological factor for CHD. Low HDL cholesterol is defined as a level of $<40\,\mathrm{mg\,dl^{-1}}$ (Table 2). High-density lipoprotein cholesterol levels are negatively influenced by cigarette smoking, overweight/obesity, physical inactivity, and type 2 diabetes. It has been hypothesized that cholesterol efflux from atherosclerotic lesions is promoted by HDL cholesterol via a receptor-mediated mechanism.

A meta-analysis of prospective studies has indicated that elevated triglycerides, or hypertriglyceridemia ($\geq 200\,\mathrm{mg\,dl^{-1}}$), is an independent risk factor for CHD after controlling for LDL and HDL cholesterol. Several factors contribute to the development of high triglyceride levels such as cigarette smoking, obesity and overweight, physical inactivity, type 2 diabetes, excess alcohol consumption, and genetic disorders (e.g., familial hypertriglyceridemia). The proposed potential mechanisms by which elevated triglycerides contribute to CHD include increased thrombogenicity, increased chylomicron and very low-density lipoprotein (VLDL) remnant particles, increased LDL and decreased HDL cholesterol, and altered metabolic processes involved in triglyceride transport.

Lp(a) is a cholesterol ester- and apolipoprotein B-containing particle with the presence of glycoprotein apo(a) that can promote atherogenesis by enhancing oxidation of LDL cholesterol, promoting cholesterol deposition in tunica intima, and interfering with fibrinolysis. The level of Lp(a) appears to be genetically determined.

Thrombogenic Risk Factors and Homeostasis

Both case-control and prospective studies have shown that thrombogenic factors (fibrinogen, blood clotting factor VII, tPA and, PAI-1) independently and

Table 2 Classification of total, LDL, HDL cholesterol, and triglycerides ($\mathrm{mg\,dl^{-1}}$)

Lipid fraction	*Classification*
Total cholesterol	
< 200	Low
200–239	Borderline high
≥ 240	High
LDL cholesterol	
< 100	Optimal
100–129	Near optimal/above optimal
130–159	Borderline high
160–189	High
≥ 190	Very high
HDL cholesterol	
< 40	Low
≥ 60	High
Triglycerides	
< 150	Optimal
150–199	Borderline high
200–499	High
≥ 500	Very high

HDL, high-density lipoprotein; LDL, low-density lipoprotein.
From (2001) *Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) Executive Summary*. NIH Publication No. 01–3670. Bethesda, MD: National Heart, Lung, and Blood Institute.

consistently increase the risk of CHD. Coronary thrombosis is generally present in most cases of acute myocardial infarction. Increased platelet counts and platelet aggregation are also associated with recurrent CHD. PAI-1 activity or antigen inhibits fibrinolysis, and increased concentrations of fibrinogen promote platelet aggregation, which in turn induces blood clotting. Several factors elevate fibrinogen, notably increasing age, smoking, obesity and diabetes, use of oral contraceptive pills, and genetic polymorphism. Blood clotting factor VII plays a pivotal role in coagulation initiation and may also be predictive of myocardial infarction. Plasma concentrations of factor VII again are influenced by both environmental and genetic polymorphism. PAl-1 regulates the rate of tPA synthesis by the endothelial cells. Several studies have shown that the levels of PAl-1 and tPA mass concentrations are predictive of initial and subsequent myocardial infarction. Elevated plasma fibrinogen, a consistent procoagulant risk factor for CHD, has several mechanisms; notably, it increases fibrin formation, platelet aggregation, and plasma viscosity. Other markers such as the D-dimer, plasmin-α-antiplasmin complex, are indicators of activation of the fibrinolytic system.

Inflammation and Infectious Agents

There is a considerable body of evidence indicating that inflammation and infection are responsible in the pathogenesis of CHD as potential etiological factors. C-reactive protein (CRP) is of special interest. Both experimental and clinical studies have documented that elevation of plasma CRP, a marker of low-grade inflammation, can predict an increased risk of CHD event in healthy persons independent of other risk factors. Although the definitive mechanism of actions of CRP that triggers low-grade inflammatory response in atherosclerosis is still under further investigation, possible indirect roles have been as follows: (1) CRP may act as a procoagulant, (2) it binds avidly with neutrophils, (3) it facilitates uptake of oxidized LDL cholesterol by macrophages, (4) it induces complement activation, (5) it is chemotactic for monocytes, and (6) it impairs endothelial functions and enhances tissue injury. Other low-grade markers of inflammation, notably amyloid A protein, albumin, IL-6, TNF-α, transforming growth factor β1, and leukocytes may also be involved in the atherogenesis process. In the development of coronary event, the proposed mechanisms of IL-6 include alterations in insulin sensitivity, increased hepatic release of fibrinogen, procoagulant effects on platelets, and increased release of adhesion molecules. TNF-α inhibits lipoprotein lipase, stimulates lipolysis, and influences endothelial function leading to a procoagulant state. Recently, chronic infections by several microorganisms, especially herpes simplex viruses, cytomegalovirus, *Chlamydia pneumoniae*, and *Helicobacter pylori*, have attracted interest in the potential roles in the pathogenesis of atherosclerosis leading to CHD.

Homocysteine

There is substantial evidence supporting elevated plasma levels of Hcy (> 15 μmol l^{-1}) as an independent etiological factor for premature atherosclerotic lesions. Hcy is a metabolite of methionine as a product of many *S*-adenosylmethionine-dependent transmethylation reactions that require several enzymes (cystathionine β-synthase, γ-cystathionase, betaine-homocysteine methyltransferase, methionine synthase, 5,10-methyltetrahydeofolate reductase) and vitamins (folic acid, B_{12}, B_6). Enzymatic defects or vitamin deficiencies cause Hcy accumulation, which causes (1) direct toxic effects to the endothelium, (2) increased adhesiveness of platelets, (3) oxidation of LDL cholesterol, (4) proliferation of smooth muscle, (5) inhibition of nitric oxide formation, and (6) increased thromboxane formation. Meta-analyses document that as much as 10% of the CHD risk may be attributable to elevated Hcy, and each 5 μmol l^{-1} increment of Hcy is associated with a 60–80% increase in CHD. Homocystinuria is a rare autosomal recessive genetic disorder that is associated with a 10-fold increase of blood Hcy levels.

Emerging Risk Factors

Accumulating evidence suggests that the renin–angiotensin II system plays an integral role in the multiple processes critically implicated in the pathogenesis of atherosclerosis. The inflammatory response of angio-

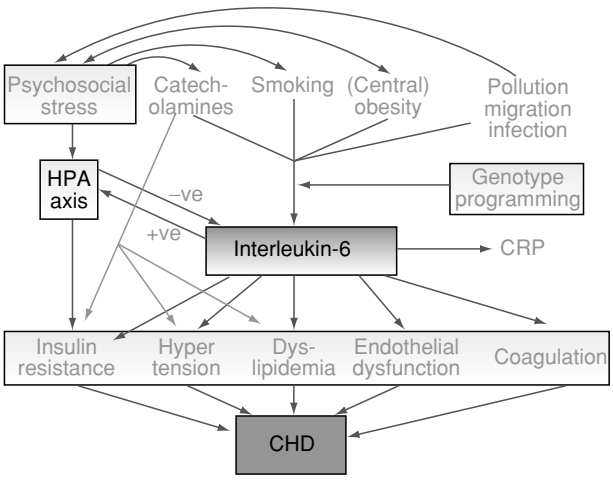

Figure 3 Possible mechanisms of action of IL-6 in the pathogenesis of CHD. From *Atherosclerosis* 148(2): 209–214 with permission.

tensin II on endothelium is mediated by the release of platelet-derived growth factor, insulin-like growth factor, IL-1, IL-6, and IL-8. Angiotensin II may also alter the structural integrity of endothelium, antagonize NO synthesis, increase the growth and migration of smooth muscle cells and fibroblasts, and promote LDL cholesterol oxidation and deposition in the intimal layer. IL-6 may also initiate the atherosclerotic process by changing metabolic, endothelial, and coagulant mechanisms. The possible mechanisms of action of IL-6 in CHD pathogenesis are illustrated in Figure 3.

The multifactorial etiology of CHD and the contribution of these factors, individually or in combination, to the risk of developing a CHD event in future is of prime importance in terms of both primary and secondary prevention.

See also: **Atherosclerosis**; **Cholesterol**: Role of Cholesterol in Heart Disease; **Coronary Heart Disease**: Antioxidant Status; Intervention Studies; Prevention; **Diabetes Mellitus**: Secondary Complications; **Exercise**: Metabolic Requirements; **Hypertension**: Physiology; Hypertension and Diet; **Obesity**: Etiology and Diagnosis; **Smoking, Diet, and Health**; **Vitamin B_6**: Properties and Determination

Further Reading

Boyle E Jr., Lille S, Allaire *et al.* (1997) Atherosclerosis. *Annals of Thoracic Surgery* 64: S47–S56.

Folsom A (1999) 'New' risk factors for atherosclerotic diseases. *Experimental Gerontology* 34: 483–490.

Fuster V, Gotto A, Libby P *et al.* (1996) Pathogenesis of coronary disease: the biological role of risk factors. *Journal of the American College of Cardiology* 27: 964–976.

Hajjar D and Nicholson A (1995) Atherosclerosis. An understanding of the cellular and molecular basis of the disease promises new approaches for its treatment in the near future. *American Scientist* 83: 460–467.

Pahor M, Elam M, Garriaon R *et al.* (1999) Emerging noninvasive biochemical measures to predict cardiovascular risk. *Journal of the American Medical Association* 285: 581–587.

Pasternak R, Gundy S, Levy D and Thompson P (1996) Spectrum of risk factors for coronary heart disease. *Journal of the American College of Cardiology* 27: 978–990.

Refsum H, Ueland P, Nygard O and Vollset SE (1998) Homocysteine and cardiovascular disease. *Annual Reviews of Medicine* 49: 31–62.

Ross R and Glomset J (1973) Atherosclerosis and arterial smooth muscle cell. *Science* 180: 1332–1339.

Schmidt-Ott K, Kagiyama S and Philips M (2000) The multiple actions of angiontensin II in atherosclerosis. *Regulatory Peptides* 93: 65–77.

(1997) *The Sixth Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure*. NIH Publication No. 98–4080. Bethesda, MD: National Heart, Lung, and Blood Institute.

(2001) *Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) Executive Summary.* NIH Publication NO. 01–3670. Bethesda, MD: National Heart, Lung, and Blood Institute.

Van Lente F (2000) Markers of inflammation as predictors in cardiovascular disease. *Clinica Chimica Acta* 293: 31–52.

Wood D, Backer G, Faergeman O *et al.* (1998) Prevention of coronary heart disease in clinical practice: Recommendations of the Second Joint Task Force of European and other Societies on Coronary Prevention. *Atherosclerosis* 140: 199–270.

Antioxidant Status

D S P Abdalla, Universidade de São Paulo, Cidade Universitátia, São Paulo, Brazil

Introduction

Coronary heart disease (CHD) is still the major component of total cardiovascular disease (CVD) mortality. The possibility that antioxidant substances might safely prevent oxidative stress and antagonize the atherosclerotic process has raised great interest in scientific and public communities. Epidemiological studies have demonstrated that those who consume higher amounts of fruits and vegetables have lower rates of CVD events. A possible explanation for this protective effect is the content of antioxidant micronutrients of fruits and vegetables. This section briefly summarizes the basic concepts behind the hypothesis that antioxidants may be helpful to prevent or retard the progression of CVD.

Oxidative Stress in Atherosclerosis

Atherosclerosis is a multifactorial disease of the arterial wall that has been considered to be the primary cause of most cardiovascular diseases, including CHD. The so-called oxidative theory of atherosclerosis has received great attention in the past 15 years, and a large body of experimental evidence has been accumulated. Free radicals may directly damage arterial endothelium, promote thrombosis, and interfere with normal vasomotor regulation. Oxidation-related events have been claimed to be involved in several steps possibly related to atherosclerosis: (1) the activation of endothelial cells to express adhesion molecules; (2) recruitment of monocytes in the subendothelial space and their differentiation into macrophages; (3) formation of foam cells;

(4) differentiation, proliferation, and migration of smooth muscle cells; (5) activation of metalloproteinases; and (6) alterations of vasorelaxation. Oxidative stress, particularly oxidation of low-density lipoprotein (LDL), has been considered to be a crucial event in atherogenesis. Free radicals derived from arterial wall cells can induce oxidative damage to LDL protein and lipids, which significantly increases its atherogenicity. Oxidatively modified LDL (oxLDL) can start the chronic inflammatory reactions of atherosclerosis. OxLDL elicits local vascular cells to produce monocyte chemotactic protein 1 and granulocyte and macrophage colony-stimulating factors promoting monocyte recruitment and differentiation to macrophages in the subendothelial area of arterial walls. As oxLDL is more negatively charged, it is recognized and internalized by the scavenger receptor pathway on macrophages, which is not regulated by negative feedback mechanisms, resulting in a massive uptake of oxLDL leading to foam cell formation. Furthermore, oxLDL is also cytotoxic to vascular cells, thus causing the release of lipids and lysosomal enzymes into the extracellular space of the arterial wall, perpetuating the inflammatory stimulus and enhancing the progression of the atherosclerotic lesion (Figure 1). OxLDL also seems to be involved in the immune response. Ox-LDL and its immune complexes have been found in atherosclerotic lesions. In addition, plasma of rabbits and humans contains autoantibodies that react with several forms of oxLDL. Moreover, a positive correlation between progression of carotid atherosclerosis and the titer of antibodies against oxLDL has been found. OxLDL can also impair endothelial function by inhibiting the release of endothelium-derived nitric oxide ($^{\bullet}$NO). Disturbances of G protein-dependent stimulation of $^{\bullet}$NO release, as well as direct inactivation of $^{\bullet}$NO, may be involved in this effect, resulting in enhanced vasospasm and platelet and monocyte adhesion to endothelium. OxLDL increases endothelial cell tissue factor and decreases protein C activation and may thus increase procoagulation at the endothelial surface. The oxidative hypothesis of atherosclerosis has been validated by a series of *in vivo* studies performed in animal models where the initial development of the atherosclerotic disease was effectively inhibited by supplementation of selected antioxidants.

Diet and Antioxidant Status

Diet has a profound effect on antioxidant status. The overall antioxidant effect of diet is determined by its components and other factors, such as:

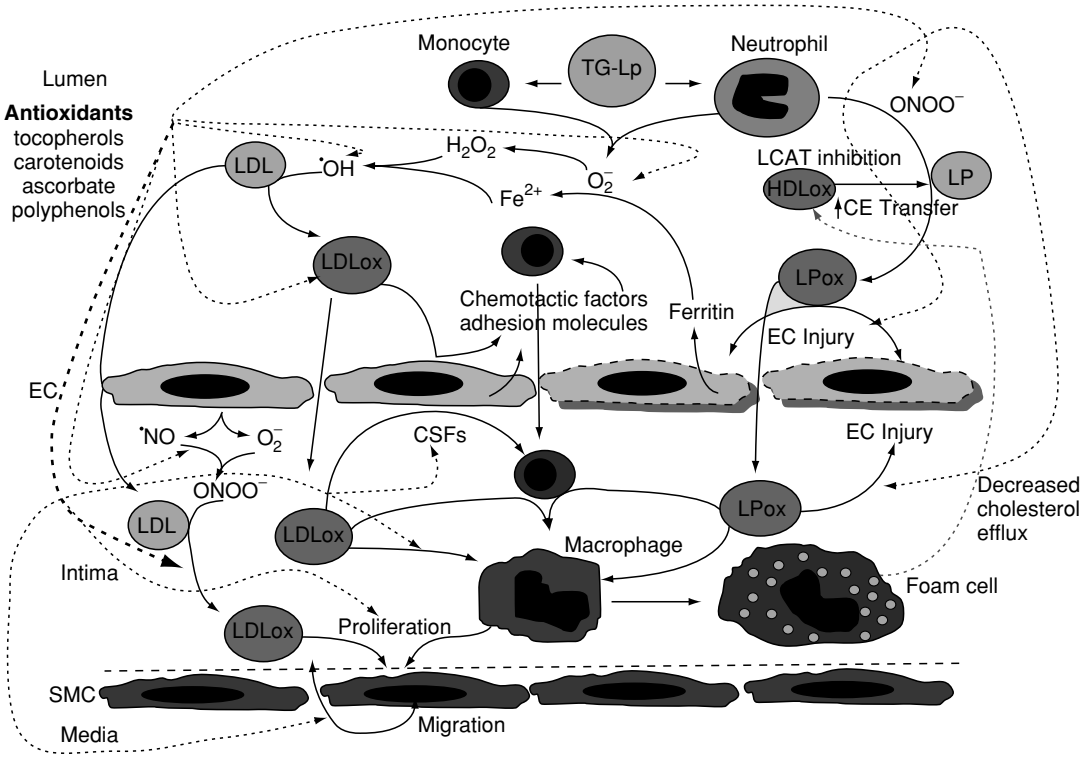

Figure 1 (see color plate 40) Possible effects of antioxidants in the atherogenic process. CE, cholesteryl ester; CSFs, colony-stimulating factors; HDLox, oxidized HDL; LCAT, lecithin: cholesterol acyl transferase; LDLox, oxidized LDL; LPox, oxidized lipoproteins; TG-Lp, triacylglicerol-rich lipoproteins; EC, endothelial cells; SMC, smooth muscle cells.

(1) antioxidant nutrient components; (2) absorption and bioavailability; (3) food processing and storage; (4) food additives and nutritional supplements; (5) chiral form and other chemical characteristics of nutrients. A dietary antioxidant is a substance that significantly decreases the adverse effects of reactive oxygen species, reactive nitrogen species, or both on normal physiological functions in humans. The diet contains antioxidants recognized as essential nutrients or nonnutrients, and such nutrient antioxidants include: vitamin E (tocopherols and tocotrienols), vitamin A, ascorbic acid, and nutrients essential for the normal function of endogenous enzymatic antioxidant systems, such as Cu, Mn, Zn (cofactors of superoxide dismutase), Se (cofactor of glutathione peroxidase) and Fe (cofactor of catalase and peroxidases). The diet also contains nonnutrient antioxidants, including a large number of phytochemicals.

Dietary Antioxidants

Mechanisms of Antioxidant Action and Relation with CHD Pathophysiology

Vitamin E (tocopherols and tocotrienols) Vitamin E is the generic term used to describe at least eight naturally occurring compounds that exhibit the biological activity of α-tocopherol. It includes α-, β-, γ-, and δ-tocopherol and α-, β-, γ-, and δ-tocotrienol. The naturally occurring *RRR* stereoisomer and the synthetic *all-rac*-α-tocopherol are equally well absorbed, yet levels of α-tocopherol in the blood and tissues increase significantly more than *all-rac*. In the liver, the α-tocopherol transfer protein preferentially selects *RRR*-α-tocopherol over γ- and other tocopherols for incorporating into very-low-density lipoprotein (VLDL), which then enters into systemic circulation. Tocotrienols appear in the blood and tissues at significantly lower levels than tocopherols, even when ingested at equivalent or higher amounts.

The antioxidant activity of the tocopherols and tocotrienols is mainly due to their ability to donate their phenolic hydrogens to lipid free radicals. α-Tocopherol acts as a chain-breaking antioxidant by donating its phenolic hydrogen to the chain-propagating lipid peroxyl radical ($LOO^{\bullet}$) in membranes and lipoproteins and replacing the latter with the less reactive α-tocopheroxyl radical (α-$TO^{\bullet}$). Alternatively, α-tocopherol may react directly with the initiating radical to prevent $LOO^{\bullet}$ formation. $LOO^{\bullet}$ may also be eliminated via a radical–radical reaction with α-$TO^{\bullet}$. Ascorbate, ubiquinone 10 and other co-antioxidants, donates electrons to the α-$TO^{\bullet}$ reducing it back to α-tocopherol. Both, α- and γ-tocopherol, the principal forms of vitamin E in the diets of Europe and the USA, respectively, protect against peroxynitrite-induced lipid peroxidation. This process is inhibited more effectively by γ-tocopherol than α-tocopherol. These results suggest that γ-tocopherol can act as a trap for membrane-soluble electrophilic nitrogen oxides, forming stable carbon-centered adducts through its nucleophilic 5-position, which is blocked in α-tocopherol. Besides the reactions with free radicals, tocopherols also react with singlet oxygen either by physical quenching or by chemical reactions. In physical quenching, excited state singlet oxygen ($O_2{}^1\Delta_g$) is deactivated to ground-state triplet oxygen ($O_2{}^3\Sigma_g$) through a charge-transfer mechanism. The tocopherols also react chemically with singlet oxygen and become destroyed after the reaction.

LDL particles may contain approximately 12 mol of vitamin E per mol of LDL. The oxidative resistance of LDL is increased in vitamin E-supplemented individuals, especially at intakes $\geq$ 269–336 mg of α-tocopherol equivalents, and there are strong correlations between oxidative resistance and LDL-α-tocopherol concentrations in these subjects. In vitamin E-supplemented subjects, the rate of oxidation was significantly decreased at 269 and 537 mg of α-tocopherol equivalents per day, leading to the conclusion that vitamin E was the most important variable that determined oxidative resistance of LDL. However, this relationship becomes weaker in unsupplemented subjects. In contrast, some studies on LDL oxidation have suggested that α-tocopherol accelerates the peroxidation of LDL lipids under mild free-radical generation and in the absence of co-antioxidants, such as ascorbate and coenzyme Q_{10}.

The bioactivity of $^{\bullet}NO$ is particularly sensitive to oxidative stress. The superoxide radical combines readily with $^{\bullet}NO$ in a diffusion-limited reaction to form peroxynitrite. In atherosclerosis, hypercholesterolemia, hypertension, and diabetes, the superoxide flux in the arterial wall is increased, leading to reduced $^{\bullet}NO$ bioavailability/activity. α-Tocopherol localizes mainly in lipoproteins and membranes, where it serves to scavenge lipid peroxyl radicals. Despite this action, atherosclerosis is characterized by lipid peroxidation within the arterial wall, even in the presence of α-tocopherol. Experimental studies have shown that arteries deficient in α-tocopherol showed a dose-dependent impairment of $^{\bullet}NO$-mediated arterial relaxation upon exposure to oxLDL. In contrast, vessels containing a high amount of α-tocopherol were resistant to this effect of oxLDL. This effect may be accounted for by inhibition of protein kinase C-α phosphorylation by α-tocopherol, which will prevent protein kinase C stimulation by oxLDL. One hallmark of atherosclerotic plaque is

proliferarion of vascular smooth muscle. Considerable *in-vitro* data indicate that α-tocopherol inhibits the proliferation of smooth muscle cells. This effect was also associated with the inhibitory effect of α-tocopherol on protein kinase C. Thrombus formation within the vessel lumen, an important event in pathophysiology of cardiovascular disease, is often precipitated by the adhesion and aggregation of platelets to a ruptured atherosclerotic plaque. It has been reported that platelet α-tocopherol content is an important determinant in platelet responsiveness toward protein kinase C-dependent stimuli leading to inhibition of platelet aggregation. Moreover, platelet α-tocopherol status appears to be important in regulating $^{\bullet}NO$ and superoxide radical production, which could also contribute to inhibit platelet aggregation. Recent data indicating that patients treated with vitamin E showed a high risk for hemorrhagic stroke could be related to its modulation on platelet function.

The antioxidant efficacy of tocotrienols in membranes is higher than that of tocopherols, in spite of their uptake and biodistribution after oral ingestion being lower than that of α-tocopherol. Cell-culture studies indicate that tocotrienols inhibit cholesterol synthesis by directly regulating the expression of 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR), through a posttranscriptional process involving accelerated degradation of the reductase protein. The molecular mechanism for this suppression by tocotrienols was ascribed to their side-chain's unique ability to increase cellular farnesol, a mevalonate-derived product, which signals the proteolytic degradation of HMGR. Some studies in humans have shown that tocotrienol supplements decreased LDL-cholesterol, apolipoprotein B (apoB), lipoprotein (a) (Lp(a)), thromboxane B_2, and platelet factor 4, suggesting hypolipidemic and antithrombotic effects. However, supplementation studies with α-tocotrienyl acetate, which is hydrolyzed, absorbed, and detectable in human plasma, did not lower cholesterol in hypercholesterolemic subjects, although it was potent in decreasing LDL oxidizability. *In-vitro* studies with HepG2 cells suggest that α-tocotrienol is effective at levels of $10\,\mu mol\,l^{-1}$. Interestingly, γ- and δ-tocotrienol, which lack the 5-methyl substituents present in α-tocotrienol, show a higher HMGR suppression. This structure–activity relationship indicates that in addition to the requirements of the prenyl side-chain for HMGR supression, changes in the methyl substitution on the chromanol ring may also lead to a divergent effect on HMGR activity. The conflicting results obtained in human studies might also be related to differences in the effective concentrations in cells necessary to inhibit HMGR. Moreover, it has recently been found that humans do not respond uniformily to the cholesterol-lowering action of tocotrienols, particularly when cholesterol and alcohol intakes are not controlled. Tocotrienols, as well as other HMGR inhibitors, lower apoB levels partly by upregulating LDL receptors in the liver. This facilitates the liver uptake of circulating apoB-LDL. In addition, tocotrienol has been shown to increase the intracellular proteolytic degradation of apoB and alter the assembly process of VLDL. Thus, it seems that the ability of tocotrienol to reduce apoB levels in blood plasma depends on both the clearance rate of LDL and the production of VLDL. A novel tocotrienol fraction from specially processed rice bran oil enriched with didesmethyl-tocotrienol (with no methyl group on the chromanol ring) has been shown to decrease plasma Lp(a). Additionally, long-term prevention studies in humans are needed to validate this effect, which may have important implications in the prevention of atherosclerosis and thromboembolism.

Carotenoids Carotenoids are pigments found only in plants and microorganisms. Most carotenoids in the diet are provided by deeply pigmented vegetables and fruits. Nearly 600 of these compounds have been identified in nature. Less than 10% of the carotenoids can be metabolized to retinol and act as vitamin A precursors. The predominant carotenoids found in blood plasma (i.e., about 90%) are β-carotene, lycopene, lutein, β-cryptoxanthin and α-carotene. The structure of carotenoids is a key determinant of their physical properties, chemical reactivity, and biological functions. The unique chemical features of each carotenoid, such as size, shape, hydrophobicity, and polarity, determine its ability to be incorporated into the molecular microenvironment and its biological function. Thus, when carotenoids are consumed by humans from the diet, these structural properties influence the absorption, bioavailability, and biodistribution, and may affect their actions at the subcellular level and on biochemical pathways. Carotenoids exist in different geometric forms (*cis*- and *trans*-isomers), which can be interconverted by light, thermal energy, or chemical reactions. After passive absorption by the enterocyte, unmetabolized carotenoids are incorporated into chylomicra followed by the uptake of remnants by liver. *Cis*-β-carotene appears in blood and tissues at significantly lower concentrations than the corresponding *trans* form, even when ingested at equivalent or higher amounts. After absorption, the *cis* form of β-carotene is converted to the *trans* form. This conversion seems to increase the bioavailability of the *trans* form at the expense of the *cis* form. The more nonpolar carotenoids (e.g., β-carotene, α-carotene, lycopene) are predominatly within LDL. The more polar

carotenoids (e.g., lutein) are associated with high-density lipoprotein (HDL). Approximately four carotenoid molecules are associated with each VLDL and one with each LDL particle, whereas for every 1000 HDL particles, only one contains carotenoids. An inverse relation between body mass and plasma concentrations of carotenoids, irrespective of dietary carotenoid intake, has been found in both epidemiological and clinical studies. High concentrations of carotenoids are found in tissues that are rich in LDL receptors, such as the corpus luteum, adrenal gland, and testes.

Carotenoids are efficient quenchers of singlet oxygen and directly scavenge free radicals. In comparison, vitamin A is a poor antioxidant. Lycopene exhibits a better antioxidant capability than β-carotene and lutein. The *in-vivo* antioxidant actions of carotenoids is mainly due to inhibition of lipid peroxidation. Moreover, cell-culture studies and very small clinical trials suggest that lycopene may inhibit macrophage cholesterol synthesis, secondary to the inhibition of HMGR, and increases macrophage LDL receptors. It is worth noting that lipid-lowering pharmacological agents can reduce serum concentrations of lycopene, via a reduction in lipoprotein particle size and competition between the drug and carotenoids for incorporation into VLDL. The biological link between high levels of carotenoids and a reduced risk of cardiovascular diseases has been suggested to be the antioxidant effect of carotenoids in LDL. Moreover, carotenoids have been detected in lipid-rich atherosclerotic plaques, although concentrations are very low. The serum concentrations of β-carotene achieved with supplementation are much higher than those achieved with the consumption of a carotenoid-rich diet. This fact may have contributed to the adverse effects observed in some intervention trials (ATBC, Table 1). Thus, a diet intervention strategy is advantageous considering that: (1) it is risk-free because, until now, no adverse effects have been associated with consuming more carotenoids from food; and (2) benefits may be gained, even if other biologically active compounds, other than carotenoids, are the truly active agents or act in synergism with carotenoids.

Ascorbic acid Ascorbic acid (vitamin C) is an essential micronutrient required for normal metabolic functions in the organism. Like other primates, humans lost the ability to synthesize ascorbic acid, as a result of a mutation in the gene coding for L-gulonolactone oxidase, required for the biosynthesis of ascorbic acid via the glucuronic acid pathway. A lack of ascorbic acid in the diet causes the disease scurvy, which can be prevented with 10 mg per day, an amount easily obtained by consumption of fruit and vegetables. Tissue saturation in healthy men occurs at ascorbic acid intakes of approximately 0.1 g per day. Thus, further increases over this intake may have minimal or no additional effect on tissue ascorbic acid concentration, and hence disease risk. Ascorbic acid is a cofactor for several enzymes involved in the biosynthesis of neurotransmitters, collagen, and carnitine. Ascorbic acid has also been implicated in the catabolism of cholesterol to bile acids by the enzyme cholesterol 7α-monooxygenase and in the steroid metabolism in the adrenals. The role of ascorbic acid in these metabolic pathways is basically to reduce the central metal ion of the mono- and dioxygenases, acting as a cosubstrate in these reactions. Ascorbic acid is an important water-soluble antioxidant in biological fluids. Ascorbic acid scavenges reactive oxygen and nitrogen species, such as superoxide, hydroxyl, nitroxide and aqueous peroxyl radicals, singlet oxygen, ozone, peroxynitrite, nitrogen dioxide, and hypochlorous acid. Two major properties of ascorbic acid make it a strong antioxidant: (1) the low one-electron reduction potentials of both ascorbate and its one-electron oxidation product, the ascorbyl radical, which allows both forms to react with and reduce basically all physiologically relevant oxidants; (2) stability and low reactivity of the ascorbyl radical. The latter readily dismutates to form ascorbate and dehydroascorbic acid, or is reduced back to ascorbate by an NADH-dependent semidehydroascorbate reductase. Ascorbic acid can also act as a co-antioxidant by regenerating α-tocopherol from α-TO$^{\bullet}$, although *in vivo*, this interaction is not clear.

The effect of ascorbic acid on hypercholesterolemia has been investigated in numerous studies, although results are still controversial. In one supplementation study, consumption of 1.0 g of ascorbic acid per day for 4 weeks resulted in a reduction in total cholesterol, whereas in another study, supplementation with (0.060–6.0) g per day for 2 weeks had no effect. The positive effect of ascorbate may be related to its role as a cofactor for cholesterol 7α-monooxygenase, or, its modulating effect on HMGR. Several observational studies have found a significant association between elevated plasma ascorbic acid and increased concentrations of HDL-cholesterol and reduced concentrations of LDL-cholesterol. In relation to thrombosis, two studies found an inverse association between serum ascorbate concentrations and coagulation factors, as well as a positive association between low serum ascorbate and elevated coagulation activation markers. However, these effects were not confirmed by further studies. *In-vitro* studies have shown that physiologic concentrations of ascorbic acid increase PGE_1 (protaglandin E_1) and PGI_1

Table 1 Randomized trials of antioxidants in CHD

Study	Participants	Follow-up (years)	Antioxidants	Results
Primary prevention trials				
ATBC	29 000 male smokers (50–60 years)	5–8	VE (50 mg per day) β-Carotene (20 mg per day)	No effect on coronary mortality (increased incidence of lung cancer)
PHS	22 071 healthy males (40–84 years)	12	β-Carotene (50 mg every other day)	Supplementation confers neither benefit nor harm with respect to CHD or cancer
CARET	18 314 men and women heavy smokers/asbestos workers	4	β-Carotene (30 mg per day) Retinol (25 000 IU per day)	Trend towards excess cardiovascular deaths (relative risk = 1,26; 95% CI = 0.99–1.61)
Women's Health Study	40 000 healthy US health professionals (40–84 years)	Ongoing	VE (600 mg every other day) β-Carotene (50 mg every other day)	Cancer and cardiovascular endpoints
Su. Vi.MAX	12 735 healthy men and women	Ongoing	VC (120mg per day) VE (30 mg per day) β-Carotene (6 mg per day) Selenium (100 μg per day) Zinc (20 mg per day)	Premature deaths, cancers, cardiovascular diseases end points

ATBC, α-tocopherol β Carotene Cancer Intervention Trial; CARET, β-carotene and Retinol Efficacy Trial; CI, confidence interval; PHS, Physicians' Health Study; Su. Vi.MAX, Supplementation en Vitamines et Mineraux Study; VC, vitamin C; VE, vitamin E.

(prostacyclin) production, reducing platelet aggregation and thrombus formation. Low concentrations of ascorbate have also been associated with increased concentration of plasminogen activator inhibitor 1, a protein that inhibits fibrinolysis. High doses of ascorbate, administered either orally or by intra-arterial infusion, have shown beneficial effects on vasodilation. Four studies investigated vasodilation in patients with CVD and found increases of 45–220% in vasodilation after administration of ascorbate (1.0–2.0 g oral or 0.025 g min^{-1} infusion). A 100% reversal of epicardial artery vasoconstriction was observed in coronary spastic angina patients infused with 0.010 g of ascorbate per minute. In the studies reporting an inverse association between plasma ascorbate and angina pectoris and CHD, the association was reduced after adjusting for smoking, suggesting that smokers may need additional ascorbic acid intakes. The possible mechanisms to explain the positive effect of ascorbate on vasodilation are related to its antioxidant activity and are suggested as the following: (1) ascorbate may spare $^{\bullet}NO$ by scavenging superoxide radicals or preventing the formation of oxLDL; (2) maintenance of intracellular concentrations of glutathione by a sparing effect or regeneration of thiols from thyil radicals, which may enhance the bioavailability of $^{\bullet}NO$ or increase the stabilization of $^{\bullet}NO$ through the formation of *S*-nitrosothiols.

Phytochemicals (polyphenols) The secondary metabolism of plants (e.g., shikimate and acetate pathways) generates products that are grouped under the term phytochemicals or polyphenols. The term polyphenols includes several classes of compounds that share a common structure; among them, flavonoids constitute the most important group, including more than 5000 compounds already identified. Subclasses of flavonoids are flavonols, flavones, catechins, flavanones, anthocyanidins, and isoflavonoids. Phenolics are derivatives of benzene with one or more hydroxyl groups associated with the aromatic rings. Several *in-vitro* studies have shown that flavonoids can modulate a variety of mammalian enzyme systems. Some of these enzymes are involved in important pathways that regulate cell division and proliferation, platelet aggregation, and inflammatory and immune responses. Soybean isoflavones, in particular genistein and daidzein, have a weak proestrogenic or antiestrogenic activity, which enable them to interact with estrogen receptors and decrease serum cholesterol concentrations. *In vitro* studies

have shown that genistein, a specific inhibitor for tyrosine kinases, prevents the development of atheroma by inhibiting cell adhesion and proliferation, by inhibiting LDL oxidation and by altering growth factor activity. In addition, soy isoflavones improve coronary vascular reactivity in female nonhuman primates, as well as, the endothelium-dependent flow-mediated vasodilation and arterial compliance in postmenopausal women. The phenolic compounds found in tea (e.g., quercetin, kaempferol, myrecitin, and epigallocatechin gallate) may be protective against CHD. Some epidemiological studies have correlated a high tea flavonoid intake with a lower incidence of CHD (Table 2). Since the 'French paradox' was revealed, it has been demonstrated that red wine, as well as dark grapes, contains many phenolic antioxidant compounds. In particular, *trans*-resveratrol has been shown *in vitro* to act as: (1) a strong antioxidant, (2) a vasorelaxant, (3) a phytoestrogen, and (4) an inhibitor of platelet aggregation, cyclooxygenase-2, and polymorphonuclear leukocytes activity. However, human studies are still needed to confirm the effects of *trans*-resveratrol in the protection against CHD. The antioxidant properties of olive oil phenolics, particularly oleuropein and hydroxytyrosol, have also been demonstrated in cellular and animals models. Among biological activities demonstrated by olive oil phenolics, the following can be mentioned: inhibition of platelet aggregation and human neutrophil respiratory burst; reduced formation of thromboxane B_2 and leukotriene B_4 by human leukocytes; and increased production of $^{\bullet}NO$ by mouse macrophages. Cocoa (*Theobroma cacao* L.) also contains appreciable amounts of phenolic substances, and *in-vitro* and *ex-vivo* studies indicate that they are endowed with certain biological activities such as an antioxidant capacity and immunoregulatory effects.

Table 2 Prospective studies on flavonoid intake and CHD

Population	*Age (years)*	*Follow-up (years)*	*Relative risk[a] (95% confidence interval)*
Cross-cultural study			
Seven-countries Study 12 763 men with CHD	40–59	25	$r = -0.5$; $P = 0.01$
Cohort studies			
Zutphen Elderly Study (The Netherlands) 552 men with stroke	50–69	15	0.27 (0.11–0.70)
Caerphilly Study (UK) 1 900 men with CHD	49–59	14	1.6 (0.9–2.9)
Finish Study 5 133 men + women with CHD	30–69	20	Men: 0.67 (0.44–1.00) Women: 0.73 (0.41–0.32)

[a]Relative risk of highest versus lowest flavonoid intake group, adjusted for age, diet and other risk factors for CHD.

Epidemiological Studies on Antioxidants and Coronary Heart Disease

In the last two decades, observational studies around the world have suggested an inverse association between the consumption of fresh fruits and vegetables, important sources of antioxidants, and the risk of

Table 3 Case-control studies relating antioxidants and CVD

Study	*Comparison groups*	*Comments*
Scotland, Finland, and Southern Italy men	Plasma VE, VC, β-carotene were determined in angina patients and healthy controls	The relative risk of angina between the lowest and highest quintiles of lipid-standardized VE levels was 2.98 (95% CI = 1.07–6.70); after adjustment for cigarette smoking, the relative risks for VC and β-carotene were, respectively, 1.63 (0.76–3.49) and 1.41 (0.63–3.13); antioxidants did not reflect regional CVD mortality rates
EURAMIC	Adipose tissue levels of α-tocopherol and β-carotene were obtained in men from 10 European countries with first acute MI and healthy controls	The multivariate odds ratio in the lowest quintile β-carotene level compared to the highest was 1.78 (95% CI = 1.17–2.71); the associations were strongest among current and former smokers. α-Tocopherol levels were not associated with an apparent increased risk, with an odds ratio of MI in the lowest compared to the highest category of 0.83 (95% CI = 0.57–1.21)

MI, myocardial infarction; VC, vitamin C; VE, vitamin E.

CHD. Several case-control studies (Table 3) and prospective cohort studies (Tables 2 and 4), comparing rates of CHD with different plasma levels or intake of antioxidants, have reported a significantly reduced CVD incidence and mortality in subjects or populations presenting a high intake or plasma levels of vitamin E, vitamin C, and β-carotene, while others have not supported these associations. However from the results of these studies, it was not possible to establish a causal relationship and several limitations. Some of the restrictions to the case-control studies are sample size, biases in cases and control selection, quality of information on exposures, and recall bias. Moreover, blood levels of antioxidants are assessed with a single measurement some time after the index event, thus precluding any distinctions between short- and long-term exposures. In addition, storage of samples for long periods may decrease the concentrations of antioxidants. The prospective cohort studies are also limited by several factors: (1) subjects taking vitamin supplements may also have a healthier lifestyle than nontakers, such as practice of physical activities, equilibrated diet, and no smoking; (2) lack of precision and reliability of nutrient intake

Table 4 Prospective cohort studies of dietary intake and CVD

Study	*Participants*	*Follow-up (years)*	*Comparison groups*	*Results (relative risk; 95% CI)*
Nurses' Health Survey (NHS)	121 000 female nurses (USA) (34–59 years)	8	Upper vs. lower quintiles of dietary and supplemental vitamin E, vitamin C or β-carotene intake	β-Carotene (0.78) Vitamin E (0.66) Vitamin C (0.80; nonsignificant after controlling for vitamin E intake)
Health Professional Follow-up Study (HPFS)	39 910 male health professionals (USA) (40–75 years)	4	Upper vs. lower quintiles of dietary and supplemental vitamin E, vitamin C or β-carotene intake	β-Carotene (0.71) Vitamin E (0.60) Vitamin C (1.29)
First National Health and Nutrition Examination Survey (NHANES-1)	11 349 adults (USA) (25–74 years)	10	Upper vs. lower dietary and supplemental vitamin C intake	Vitamin C (0.66) Intake of other vitamin supplements was not taken into account
Iowa Women Health Study	34 486 postmenopausal women (USA) (55–69 years)	7	Upper vs. lower quintiles of dietary and supplemental vitamin E, vitamin A, retinol, and carotenoid intake	Vitamin E (0.96) No benefit for Vitamin A, retinol, and carotenoids
	2 748 men (Finland) (30–69 years)	14	Upper vs. lower tertiles of dietary and supplemental vitamins or β-carotene intake	Vitamins (0.66) β-Carotene (1.02)
	2 385 women (Finland) (30–69 years)	14	Upper vs. lower tertiles of dietary and supplemental vitamins or β-carotene intake	Vitamins (0.35) β-Carotene (0.62)
	1 899 men with hyperlipidemia (USA) (40–59 years)	13	Upper vs. lower quartiles of serum β-carotene intake	β-Carotene (0.64)

Table 5 Randomized trials of antioxidants in CHD

Study	*Participants*	*Follow-up (years)*	*Antioxidants*	*Results*
Secondary prevention trials				
CHAOS	2002 men and women with angiographically proven CHD	510 days (3–981)	VE (400 or 800 mg per day)	75% reduction in nonfatal MI; nonsignificant increase in cardiovascular death
HOPE	9 541 patients with CHD, stroke or pad	4.5	VE (400 mg per day)	No effect on death from CVD + MI + stroke
GISSI	11 324 men and women with CHD	3.5	VE (300 mg per day)	4.7% (NS) reduction of total mortality + nonfatal MI + CVA; 35% reduction of sudden death
WACS	8 000 women with CHD	Ongoing	VE (400 mg per day) VC (1 g per day) β-Carotene (20 mg per day)	MI, stroke, coronary revascularization, and death from CVD end points
HPFS	20 000 patients with previous angina, stroke, claudication, or diabetes	Ongoing	VE (600 mg per day) VC (250 mg per day) β-Carotene (20 mg per day)	Incidence of coronary mortality and all-cause mortality end points

CVA, cerebrovascular accidents; GISSI, Gruppo Italiano per lo Studio della Sopravvivenza nell'infarto; HOPE, Heart Outcome Prevention Evaluation; HPFS, Heart Protection Study; MI, myocardial infarction; VC, vitamin C; VE, vitamin E: NS, nonsignificant; WACS, Women's Antioxidant Cardiovascular Study.

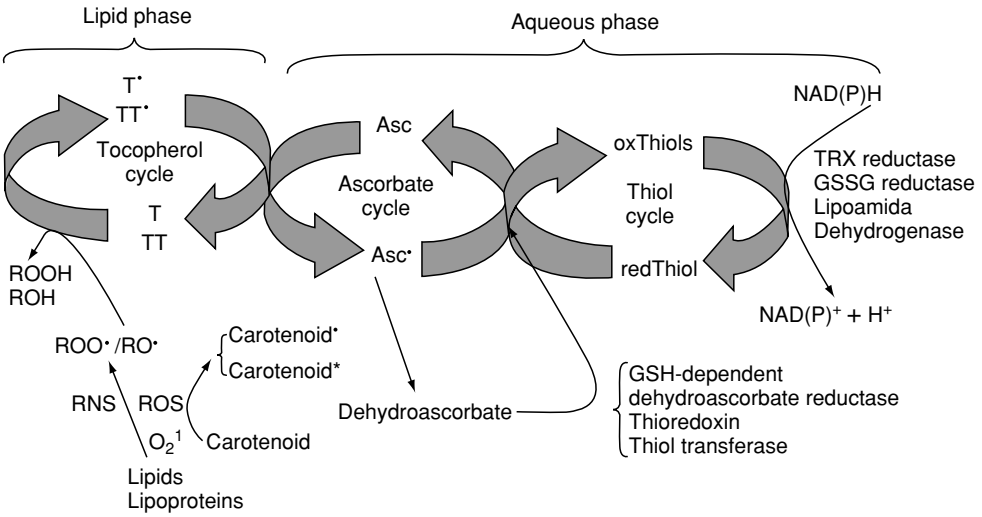

Figure 2 Integrated action of antioxidants. Asc, ascorbate; Asc•, ascorbyl radical; GSSG, oxidized glutathione; carotenoid•, carotenoid-derived free radical; carotenoid*, excited-state carotenoids; oxThiols, oxidized thiols (glutathione, lipoic acid, thioredoxin); RO•, alcoxyl; ROH, lipid-hydroxyl derived; ROO•, peroxyl radical; ROOH, lipid hydroperoxide; ROS, reactive oxygen species; RNS, reactive nitrogen species; T, tocopherols; T•, tocopheroxyl radical; TT, tocotrienols; TT•, tocotrienoxyl radical; TRX, thioredoxin.

ascertained by food-frequency questionnaires; and (3) confounding factors that were neither measured nor controlled for in multivariate analysis.

The randomized intervention trials (Tables 1 and 5) have been considered as the golden standard to test the hypothesis that antioxidant micronutrients effectively have a causal relationship with the decrease of CHD morbidity/mortality. Until now, the results of these trials have raised the possibility that some benefits from observational epidemiology may have been overestimated. β-Carotene and vitamin C appear to confer no overall benefit in the primary prevention of CVD among well-nourished subjects. However, whether risk can be reduced among those with disease or lower baseline levels of these micronutrients remains unclear. For vitamin E, the results are still controversial; while the results of the Cambridge Heart Antioxidant Study (CHAOS) show a

lower incidence of nonfatal infarction and major cardiovascular events in supplemented subjects, HOPE results indicate no benefit on chosen endpoints. There are also criticisms of the randomized trials. Supplementation of a diet with individual antioxidants could not exert optimal effects because the concerted mechanism of action of the antioxidant system (Figure 1) could require the availability of adequate and harmonized levels of several antioxidant components in the diverse microenvironments in the organism. Furthermore, it may be possible that the antioxidants chosen (vitamin C, vitamin E, and β-carotene) could only be markers for the intake of other, as yet unidentified, substance(s) with the true biological effect. The rationale behind the approach of antioxidant supplementation in order to prevent the development and progression of CVD has not yet been firmly established. More basic research is required to ascertain the appropriate antioxidant(s) and doses, oxidation biomarkers, intermediate and final end points, among other factors, before planning further large intervention trials. Thus, until the results of these studies produce the final answers, the consumption of fruits and vegetables and a healthy life style would be advisable.

See also: **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; Role of Antioxidant Nutrients in Defense Systems; **Ascorbic Acid**: Properties and Determination; Physiology; **Carotenoids**: Occurrence, Properties, and Determination; Physiology; **Soy (Soya) Beans**: Dietary Importance; **Tocopherols**: Properties and Determination; Physiology

Further Reading

Antony MS (2000) Soy and cardiovascular disease: cholesterol lowering and beyond. *Journal of Nutrition* 130: 662S–663S.

Carr AC and Frei B (1999) Toward a new recommended dietary allowance for vitamin C based on antioxidant and health effects in humans. *American Journal of Clinical Nutrition* 69: 1006–1007.

Chisolm GM and Steinberg D (2000) The oxidative modification hypothesis of atherogenesis: an overview. *Free Radical Biology and Medicine* 28: 1815–1826.

Clinton SK (1998) Lycopene: Chemistry, biology and implications for human health and disease. *Nutrition Reviews* 56: 35–51.

Diplock AT, Charteux JL, Crozier-Willi G *et al.* (1998) Functional food science and defence against reactive oxidative species. *British Journal of Nutrition* 80: S77–S112.

Gaziano JM (2000) Dietary antioxidants and cardiovascular disease. *Vitamins and Hormones* 58: 299–320.

Hollman PCH and Katan MB (1999) Health effects and bioavailability of dietary flavonols. *Free Radical Research* 31: S75–S80.

Kamal-Eldin A and Appelqvist LA (1996) The chemistry and antioxidant properties of tocopherols and tocotrienols. *Lipids* 31: 671–701.

Keaney JF, Simon DI and Freedman JE (1999) Vitamin E and vascular homeostasis: implications for atherosclerosis. *FASEB Journal* 13: 965–976.

Krinsky NI (1998) The antioxidant and biological properties of the carotenoids. *Annals New York Academy of Sciences* 854: 443–447.

Marchioli R (1999) Antioxidant vitamins and prevention of cardiovascular disease: laboratory, epidemiological and clinical trial data. *Pharmacological Research* 40: 227–238.

Packer L, Weber SU and Rimbach G (2001) Molecular aspects of α-tocotrienol antioxidant action and cell signalling. *Journal of Nutrition* 131: 369S–373S.

Theriault A, Chao JT, Wang Q *et al.* (1999) Tocotrienol: a review of its therapeutic properties. *Clinical Biochemistry* 32: 309–319.

Visioli F, Borsani L and Galli C (2000) Diet and prevention of coronary heart disease: the potential role of phytochemicals. *Cardiovascular Research* 47: 419–425.

Intervention Studies

P C Elwood, Llandough Hospital, Penarth, South Glamorgan, UK

Definition and Aims

In the section on epidemiology a number of research strategies are discussed. These fall into two obvious groups: observation and intervention. In general, observation studies seek to ascertain more about the natural course of a disease process, whereas intervention studies test the effect of a possible therapeutic or preventive measure. (*See* **Epidemiology.**)

There is a further division of intervention studies, and this division is of particular importance in nutrition research. First, there are trials in which a foodstuff or a nutrient is given to subjects, and the effect upon a mechanism or a metabolic process, or the effect on a biochemical or hematologic variant is observed. Measuring the change in blood lipids following changes in dietary fat intake, or measuring the blood pressure response to coffee prepared in different ways, illustrates this kind of study. In these, the numbers of subjects can be relatively small, because a measurement of the outcome variable (serum cholesterol, blood pressure, or platelet aggregation, etc.) is obtained for every subject. Furthermore, such trials can usually be relatively short-term, and often it is acceptable to use a cross-over design, in which each

patient has a period on the intervention diet or supplement being evaluated, and a period on a normal diet without the intervention, thus serving as his or her own control. This last can be a most efficient design, provided the order of the intervention and control periods is random for each subject, and provided an adequate washout time is allowed between the two periods.

In the second kind of intervention trial, the effect of a dietary change, or a nutritional supplement, on the incidence of a disease is studied. This might be the reduction in the mortality of patients who have had a heart attack and who are advised to eat fatty fish, or the effect of a magnesium supplement given to patients judged to be at risk of a heart attack or stroke. Trials such as these are very different in that the outcome is a disease event, and this is likely to occur in only a very small proportion of persons in the trial. As a consequence, the numbers involved have to be very large and/or the period of observation very prolonged.

Both kinds of trials involve intervention and the level of confidence that one can put on the results is, in general, much higher than in studies from which the evidence is purely observational. Most of the design features are the same in the two kinds of intervention trial, but the problems which arise in those which aim to change the progress or reduce the incidence of a disease are substantial and so most of the following discussion will focus on the second.

Design

The only design for intervention studies which is now acceptable, fulfills the basic criteria for randomized controlled trial (RCT). These criteria include the following:

1. The allocation of subjects to the group which will receive the intervention, or to the control group which will not, must be random. Only allocation at random is likely to insure that there are no systematic differences or bias between the groups, which might account for some of the effects found and which might wrongly be attributed to the intervention. In fact, the comparability between the groups should later be checked by comparing the distributions of age, social class, and perhaps smoking, etc. in the two groups to give some assurance that by chance some systematic difference or bias has not occurred.
2. The advice, handling, and monitoring of the two subgroups of subjects – those receiving the intervention and those not – must be identical in every respect other than the intervention being tested. As far as possible the subjects in the trial should be 'blind' with regard to whether or not they are receiving the intervention under test. While this last is easy in a drug trial by the use of placebo tablets, it is unlikely to be possible in a trial of a foodstuff.
3. The outcome, whether this is a change in a biochemical or other variant or a difference in disease incidence or disease progression, must be assessed by an observer who is blind as to whether or not the subject has or has not been receiving the intervention.

The setting-up and conduct of an RCT is a major undertaking, and this is particularly true of a dietary trial. It may be appropriate therefore to consider the use of a factorial design. In this, two or more interventions are separately randomized. Equal numbers of subjects therefore receive each possible combination of the interventions, with some receiving all and some none. This design is acceptable provided that there is no reason to expect strong interactions between the different interventions. It is highly efficient as the main effects are all tested simultaneously without loss of power. For example, a trial by Burr and colleagues evaluated the effects of three dietary changes on mortality after a heart attack: a reduction in saturated fat, an increase in fatty fish, and an increase in cereal fiber consumption. Advice on each of these dietary changes was given to half of the total patient group, selected for each intervention at random.

The compliance of subjects with the intervention being tested is of great importance and should be taken into account when drawing conclusions from a trial. In fact, it is important to monitor both the degree of compliance by those who have been given advice about the dietary factor being tested and the extent to which the control subjects may spontaneously change their consumption of that factor. Ideally, compliance should be monitored in an objective way; for example, the estimation of plasma eicosapentaenoic acid will indicate how much fatty fish has been consumed. If an appropriate objective measure is not available, then at intervals the subjects (both those advised and the controls) can be asked to keep a diary of the food they consume.

Every reasonable effort should be made to achieve and maintain a high level of compliance. At the same time, whether or not a subject complies with the intervention, and whether or not a control subject spontaneously adopts the measure under test, all the subjects must be retained within their originally randomized group, and their results analyzed accordingly. This leads to what is known as an 'intention to treat' analysis. Omission of patients from one or other group because of poor compliance or other reason can unbalance the comparability obtained by

the original randomization, and is highly likely to introduce bias and invalidate any conclusions drawn from the results of the trial.

In planning a trial, careful attention should always be given to the likely power of a study, that is, the likelihood that the planned trial will be large enough, and the outcome measures sensitive enough to detect a statistically significant beneficial effect, should one occur. A number of things affect the power of a study. The most obvious are the number of outcome events and the likely difference in this between the intervention and the control group. Clearly, this last will not be known, otherwise the trial would be unnecessary. An estimate usually has to be made. If the intervention involves the consumption of a foodstuff, such as fish, or fruit and vegetables, the reproducibility of measurement of consumption should be assessed, and any changes in the consumption of other foods or nutrients monitored.

All these estimates should be realistic. It is all too easy to overestimate the likely effect of advice. For example, dietary intakes are relatively fixed, particularly in older subjects, and advice to increase the consumption of, say, fruit and vegetables, or to reduce fat intakes, has relatively little effect.

Primary and Secondary Trials in Coronary Heart Disease

The terms 'primary' and 'secondary' are used to distinguish trials which test a preventive measure in patients who have not already had a clinical episode of a disease (primary) from those based on patients who have already had a myocardial infarct or other clinical event (secondary trials). A distinction of this kind is most often made in coronary heart disease (CHD).

In fact, the concept of primary and secondary trials is not helpful. The terms relate to past history, and not to future risk. In relation to risk of a future vascular event, subjects simply form a continuum, ranging from subjects at exceedingly low risk, such as young, healthy subjects, to patients who have recently had a vascular event. Ranged between these are subjects who smoke, who have raised blood pressure, raised cholesterol, etc., whose risk of a vascular event is somewhere between those of subjects at the two extremes. In the evaluation of the risk of any individual, the occurrence of a past vascular event is only one of a number of risk factors which determine the likelihood of a future event. Figure 1 displays the continuum of risk.

Ideally, all possible preventive measures should be tested in an RCT based upon younger healthy subjects. The numbers required for a trial in such individuals are however formidable. Further difficulties arise in persuading healthy subjects to comply with the intervention to a sufficient degree and for a sufficient time.

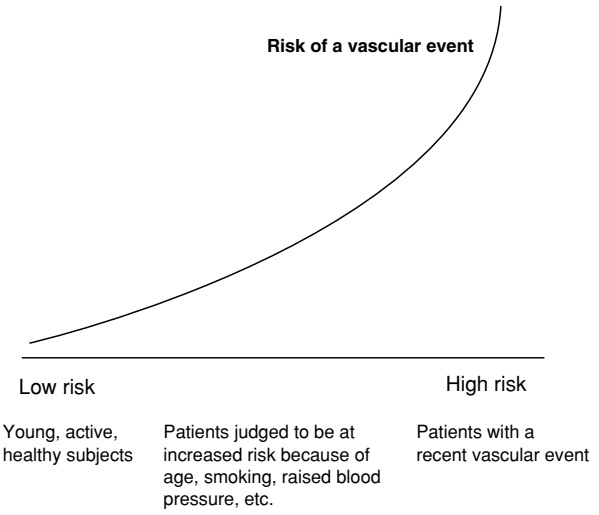

Figure 1 Low verus high risk of a vascular events

Intervention trials are therefore often conducted in subjects who have evidence of an increased risk, be it raised levels of risk factors, or some evidence of vascular disease, such as angina pectoris. Trials can even be based upon patients who have already had a CHD event, such as a myocardial infarct. All such patients know that they are at an increased risk of a further event, and they are likely to be more easily persuaded to comply with a dietary change for long periods. It can be argued that the underlying disease of survivors of a myocardial infarction is at an advanced stage and may not be reversible within the time span of a trial of a reasonable duration. There may well be some truth in this, particularly if the aim of the intervention is to reduce atherosclerosis.

Nevertheless, a myocardial or a cerebral infarct results when a thrombus develops on top of an atheromatous plaque on the wall of a blood vessel. Dietary factors are undoubtedly relevant to the risk of thrombosis. For example, the head of an intravascular thrombus is usually a mass of aggregated platelets, and alcohol reduces platelet activity. A myocardial infarct can precipitate ventricular fibrillation and death, and there is evidence that fish oil reduces the electrical stability of the myocardium. Free radicals from an ischemic lesion can damage the myocardium, and antioxidants from fruit and vegetables may reduce this damage to the heart muscle. There are therefore mechanisms in vascular disease, other than atherosclerotic disease in the vessel wall, which are affected by dietary items and which are appropriately tested in dietary intervention RCTs.

Hypotheses Tested by Dietary Intervention Trials

It would be a mistake to summarize the present situation on dietary trials in CHD, as results are constantly accumulating and the picture is changing. Rather, intervention trials in a few areas will be described briefly, largely in order to illustrate methodological aspects.

Dietary Fat Reduction

Undoubtedly, the relationship of saturated fat to fat reduction has been more investigated in the prevention of CHD than any other potential prophylactic. The early trials focused largely on the reduction of total fat in the diet; interest then focused on the P:S ratio (the ratio of polyunsaturated to saturated fats) and on monounsaturated fats. Almost all metaanalyses or overviews of dietary RCTs have failed to detect any significant reduction in mortality from dietary changes aimed at reducing cholesterol, though several overviews have reported a significant reduction of around 20% in the incidence of nonfatal myocardial infarctions. (*See* **Cholesterol**: Role of Cholesterol in Heart Disease; **Fats**: Requirements.)

An issue worth noting because of its wider relevance to the drawing of conclusions from trials is the degree to which successful trials are cited in preference to negative trials. A report of an overview of 22 RCTs in which the effect of a reduction in dietary fat intake had been examined went on to examine how often these trials had been cited in the general literature. It was found that the few trials in which a reduction in CHD had been observed had been cited six times more frequently than the trials which had failed to find such an effect. This positive bias in the reporting of trials is serious bias in any overview of the literature, whether this is based upon a formal metaanalysis, or simply an impression gained from reading the literature.

The relevance of cholesterol level to vascular disease has been very considerably reinforced by the results of recent RCTs of statin drugs. These drugs substantially reduce cholesterol level, and at the same time are associated with a reduction in vascular disease incidence and mortality. The reduction in disease appears however to commence virtually immediately, and this has led to the suggestion that some of the benefit from these drugs may arise because of effects that have been described on platelet function, on hemostasis and on immunological mechanisms.

Within the area of lipids and vascular disease, one trial stands out. This tested the effect of the so-called 'Mediterranean diet,' a dietary pattern modeled on the dietary pattern in Crete and other Mediterranean communities. The initial finding was a 70% reduction in all-cause mortality together with comparable reductions in nonfatal events, and these results have been sustained in a 4-year follow-up, reported in 1999. The diet advised in this trial was rich in fruit and vegetables, but the group responsible for the study considered that its most important feature was its richness in α-linolenic acid. In the trial this last was achieved by distributing a special margarine containing rapeseed oil to half the subjects. At present only one RCT of a Mediterranean type of diet has been reported. Clearly there is an urgent need for further trials.

Fatty Fish

The hypothesis that fatty fish, fish oil, or eicosapentaenoic acid might reduce the risk of CHD emerged from surveys of Inuit in Greenland in the 1930s. Later, historical studies showed that changes in fish consumption during World War II had been accompanied in The Netherlands and in Norway by decreases in heart disease mortality. Laboratory and animal studies then examined the effects of fish oil on various risk factors and certain mechanisms involved in CHD. This was the background to the setting-up of a major RCT of fish and mortality. A total of 2033 men were randomized factorially to three interventions, one of which was the consumption of fatty fish on at least 2 days each week. The consumption of fish was associated with a reduction in mortality of about 30% and a saving of lives of about 36 per 1000 men advised. Because these estimates of benefit are derived from the results of a single trial they carry a considerable uncertainty. Further trials are required, followed, hopefully, by a metaanalysis. It is notable, however, that this trial gave no evidence that nonfatal infarctions were reduced by the fish consumption. Indeed, there were rather more nonfatal infarctions in the men advised (4.8%) than in the other men (3.2%). It is therefore possible that the effect of fish is to reduce the risk of death after infarction rather than to reduce the incidence of infarction. Evidence consistent with this hypothesis comes from studies of the electrical stability of the myocardium in animals, following experimentally induced infarction.

Fruit and Vegetables

It is believed that the oxidation of lipids is an important step in the development of atherosclerosis. Oxidized low-density lipoprotein cholesterol is potentially highly atherogenic, while antioxidants prevent its oxidation and reduce the risk of other oxidative damage, thus retarding the development of atherosclerosis. There is therefore growing interest in the protective effect of both supplements of the

antioxidant vitamin supplements, including vitamins E, C, and β-carotene, and in fruit and vegetables which carry these and other antioxidants.

Numerous observational studies based on communities and on individuals give supporting evidence of benefit. In fact, the evidence is virtually totally consistent that subjects and groups with high consumptions of fruit and/or vegetables have a reduced incidence of both digestive cancers and of CHD. Despite their consistency, however, these observations are seriously open to confounding by related social and other dietary factors. Hence there is no substitute for RCTs.

A number of RCTs have been conducted to test the effects of supplements of vitamins E, C, and/or β-carotene. A recent overview of these has concluded that 'there is no clear indication for the use of these supplements . . . and the suggestion of potential harm associated with β-carotene in two clinical trials (indicates) that β-carotene in the form of additional vitamin supplements should be avoided.'

On the other hand, RCTs of fruit and vegetables necessitate increasing consumption of these foods by half the subjects studied, and even in a so-called secondary preventive trial the dietary increase has to be either large or prolonged, or both. This makes such trials very difficult. The diets of older subjects tend to be relatively fixed, and it is exceedingly difficult to persuade subjects – even patients who have already had a clinical CHD event – to increase their consumption of any food item. In our experience this is especially true for fruit and vegetables. (*See* **Antioxidants:** Natural Antioxidants.)

Dietary Fiber

For a very long time there has been interest in dietary fiber. At first this focused upon 'roughage,' or unabsorbed dietary material, and measures such as fecal bulk and bowel transit time were commonly used. Interest then focused on cereal fiber, the diseases of possible relevance being bowel cancer and CHD. Numerous observational studies have been reported, but one of the uncertainties in all of these is the effect of confounding factors. In fact, confounding is a major source of possible confusion in all dietary studies that are dependent upon observation. The diet of an individual is determined by a host of factors, including lifestyle, education, social standing and income level, educational and psychological factors. Furthermore, the consumption of any one food or diet factor is interrelated with a host of other foods, and cannot realistically be considered in isolation. For example, an increase in the consumption of wholemeal bread will probably lead to an increase in spreading fats; an increase in fish consumption will probably be balanced by a reduction in meat. It will be almost impossible to take account of all these confounding and interrelated factors in a study of disease incidence in relation to a single food item or nutrient. The final elucidation of relationships between diet and disease must therefore be ultimately dependent upon intervention studies, in which control over confounding is achieved by randomization, and different levels of dietary intake are achieved experimentally.

There has been only one RCT of dietary fiber and CHD. This was conducted by Burr and his colleagues in south Wales. Advice was given to 1017 men who had already had a myocardial infarct. This led to a mean intake of cereal fiber of 19 g per day, compared with 9 g in 1016 controls given no advice about fiber. After 2 years there was no significant difference in the deaths in those given the advice about fiber (12.1%) compared to those not so advised (9.9%).

Interest in fiber has however shifted again, and attention is now being focused on soluble fiber. This, too, is confounded and interrelated with numerous other food items and nondietary items and its importance will only be sorted out by RCTs.

Interpretation of the Results of RCTs

Relative and Absolute Risk Reduction

There are two important aspects of trials that should be understood. First, the results of a controlled trial can be expressed as 'relative' protection, or the 'relative' reduction in the outcome events, or they can be expressed in terms of the 'absolute' effect of the intervention. This can be very easily illustrated.

In the Diet and Reinfarction Trial (DART) 1015 post-myocardial infarction patients were given advice to eat fatty fish on at least 2 days per week. A group of 1018 other patients, selected at random, were given no such advice. During the next 2 years 127 (12.5%) of the men advised to take fish had a CHD event, while 149 (14.6%) given no such advice had an event. The difference between these incident figures implies a relative reduction in the risk of an event of 30%. On the other hand, the absolute reduction was 2% over the 2 years or one IHD event was prevented per year in every 100 patients advised to eat fatty fish.

Both these figures are correct, but each expresses rather different concepts. Thus the relative reduction, or relative benefit, is the proportionate extent to which the risk is reduced in a group in which there has been an intervention, compared to a control group with no such intervention. The absolute reduction, or absolute benefit, is the number of events prevented. Another way of expressing this last is the number needed to treat (NNT) to prevent one disease event. Clearly, for fatty fish, this is 100 patients for 1 year.

While figures of absolute benefit may at first sight appear disappointing, they carry a wealth of information. Thus, in the above example, one can say that, of every 100 patients advised to eat fatty fish, a CHD event will be prevented in one or two each year, but about six will still die from CHD each year. About 92 will derive no benefit because, judging from the experience of the subjects in the control group, they would have survived anyway. No worry! They will have enjoyed the fatty fish and will probably have saved money on meat!

Intervention with drugs has much the same protective effect (for example, antihypertensives achieve a relative reduction in CHD event of about 25%, that is, closely similar to the 30% achieved by fish consumption in DART). However the patients who are given drugs, and who would have survived anyway, will have been exposed to the undesirable side-effects of the drugs, and their therapy will have generated considerable costs. None of this is meant to imply that dietary intervention is a substitute for medication. Raised CHD risk factors should be treated appropriately, but considerations should also be given to dietary intervention. These are not in competition.

Overviews or Metaanalyses

In the nature of things, certainty is unobtainable and even the largest and most carefully conducted RCT can, by chance, give a wrong answer. Firm conclusions should only be based on evidence drawn from a number of trials. The play of chance can never be ruled out, whatever the level of statistical significance, and consistency in a number of studies is likely to be a better guide to truth than the results of any single study. Further to this, the results of any single trial cannot be extrapolated to the general population with certainty. Conclusions from a trial are limited by issues such as the selection of patients tested, the diagnostic criteria used, the way the intervention was achieved, and the possibility that the intervention may have affected other factors of relevance to the disease. Together, these uncertainties can make it inappropriate and perhaps misleading to generalize the results from any one study to patients in general and make them the basis of clinical recommendations.

Hence the importance of overviews of studies, or metaanalyses of data from a number of different studies. This approach is particularly appropriate with RCTs, but it is important that data from every relevant RCT are included.

In fact, even the most meticulous search of the literature followed by a formal overview of all the evidence from relevant RCTs still leaves uncertainties. There is publication bias, and reports which report an effect are generally more likely to be accepted by a journal than those which show no effect. But then this limitation affects judgments based simply on cumulative reading of the literature, however carefully done. Nevertheless, the approach in research that is likely to get closest to the truth is clearly drawing conclusions from all the available acceptable evidence. Cochrane, an epidemiologist, saw this and he commented in a Rock Carling lecture, later published as *Effectiveness and Efficiency*: 'It is surely a great criticism of our profession that we have not organised a critical summary, by speciality or sub-speciality, adapted periodically, of all relevant clinical trials.'

The Cochrane Collaboration

This urging by Cochrane led to the setting-up of the Cochrane Centre in Oxford, and then to the Cochrane Collaboration, a worldwide effort coordinated in 11 Cochrane Centres. The collaboration aims to identify every RCT ever done in medical care, to group these by clinical objective, and make them readily available as databases. These then constitute the basis for statistical overviews, or metaanalyses, which draw on all the acceptable trial evidence relevant to each clinical procedure.

Final Considerations

RCTs of dietary factors are immensely difficult and before one is mounted consideration should be given to a number of general considerations relevant to intervention in cardiovascular disease. In summary, there are unlikely to be any easy answers!

First, compliance with a dietary intervention is likely to be poor. Trials of the effect of dietary advice on cholesterol levels have generally shown very little effect, and to some extent – and probably a very large extent – this is because of the difficulties most people find in making a significant, let alone a large, reduction in dietary saturated fats. Likewise, attempts to increase the intake of dietary fiber or fruit and vegetable intake have achieved relatively little change. Results such as these have far-reaching implications for the acceptability of prophylactic strategies in symptomless subjects, or in the general population.

Second, effectiveness is likely to be low. Geoffrey Rose defined a paradox – population preventive strategies, even those with large overall potential benefit, are likely to confer little benefit on the individual – and he went on to describe the overall benefit/risk balance in strategies involving the general population as 'worrisome.' An overview by Davey-Smith and Ebrahim of nine major population-based RCTs of the effect of advice on diet, smoking, and blood

pressure led to the conclusion: 'The pooled effect of intervention on multiple risk-factors is insignificant.'

Third, the costs of intervention are likely to be high. The cost of an intervention strategy that really matters is not the cost of the change of diet, the cost of nutritional supplements of medication in a single subject. Rather, the cost that matters is that of treating a sufficient number of subjects, for a sufficient time to save one life, or prevent one disease event. While this approach has been well developed with drug therapy, the most relevant costs of a dietary intervention are those that arise from the need for subjects to accept a change in their diet, and for most of us, a possible small change in the long-term risk of a disease is far from the front of our minds when preparing a meal.

See also: **Antioxidants**: Natural Antioxidants; Role of Antioxidant Nutrients in Defense Systems; **Cholesterol**: Role of Cholesterol in Heart Disease; **Epidemiology**; **Fats**: Requirements

Further Reading

Alderson M (1983) *An Introduction to Epidemiology*, 2nd edn. London: Macmillan.

Bradford-Hill A (1971) *Principles of Medical Statistics*, 9th edn. London: Lancet.

Burr ML (1983) Epidemiology for nutritionists. 3. The design of studies. *Human Nutrition: Applied Nutrition* 37A: 339–347.

Burr ML, Fehily AM, Gilbert JF *et al.* (1989) Effects of changes in fat, fish, and fibre intakes on death and myocardial reinfarction: Diet and Reinfarction Trial (DART). *Lancet* ii: 757–761.

Cochrane AL (1971) *Effectiveness and Efficiency: Random Reflections on Health Services*. London: Royal Society of Medicine Press.

Davey-Smith G and Ebrahim S (1996) Health Promotion in Older People for the Prevention of Coronary Disease and Stroke. *Health Promotion Effectiveness Review: Summary Bulletin 1*. London: Health Education Authority.

Margetts BM and Nelson M (1977) *Design Concepts in Nutritional Epidemiology*, 2nd edn. Oxford: OUP.

Muldoon MF, Manuck SB and Matthews KA (1990) Lowering cholesterol concentrations and mortality: a quantitative review of primary prevention trials. *British Medical Journal* 301: 309–313.

Peto R, Pike MC, Armitage P *et al.* (1976) Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. II. Analysis and examples. *British Journal of Cancer* 34: 585–612; 35: 1–39.

Pocock SJ (1983) *Clinical Trials: A Practical Approach*. New York: Wiley.

Ravenov U (1992) Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. *British Medical Journal* 305: 15.

Prevention

F G Huffman and S Nath, Florida International University, Miami, FL, USA

Background

There has been a steady decline of coronary heart disease (CHD)-related mortality in most industrialized nations over the past few years. Despite this fact, CHD remains the leading cause of morbidity and mortality in adults of middle years and older (after age 65), leading to a loss of productivity and substantial healthcare costs in the industrialized countries. The World Health Organization estimates that by the year 2010, CHD would be the leading cause of death in the developing countries as well. Therefore, prevention (both primary and secondary) of CHD is extremely important from a public health point of view.

The preceding section in this chapter highlights the etiology of CHD that includes both modifiable and nonmodifiable risk factors. Among all CHD risk factors, cigarette smoking, unhealthy diet, overweight or obesity, physical inactivity, high blood lipids (high total cholesterol and elevated low-density lipoprotein cholesterol), high blood pressure or hypertension, and high blood glucose (also metabolic syndrome) are preventable or manageable. Several prophylactic therapies and screening of high-risk individuals also offer potential benefits for the prevention of CHD (**Table 1**).

Primary prevention of CHD is generally defined as the modification of risk factors or prevention of the development of these risk factors, with the aim of delaying or preventing a first coronary event (angina pectoris, myocardial infarction, or sudden death), offers the greatest opportunity for reducing the burden of CHD from a personal, economical, and global aspect. The prime objective of primary prevention is to reduce both short-term (< 10 years) and long-term (> 10 years) risk. However, secondary prevention includes therapy or measures to minimize recurrent coronary event (angina pectoris, history of myocardial infarction, history of bypass graft or angioplasty) and decrease mortality and morbidity in patients with established CHD, which is not cost-effective in most cases. The potential benefits of primary prevention exceed those of secondary prevention, since $\sim 30\%$ of patients with first myocardial infarction survive less than 3 months. The first CHD event may progress to prolong permanent morbidity. Both primary and secondary preventions share many common characteristics (**Figure 1**). Therefore, the elements of both

Table 1 Prevention of coronary heart disease

Lifestyle modification	*Goals*	
Smoking cessation	Encourage patient to stop smoking completely	
Healthy diet	25 g of fiber, > 400 g of fruits and vegetables; < 30% fat, < 10% saturated fat, poly- and monounsaturated fat 10% each, < 300 mg of cholesterol of daily calories; salt intake 6 g or less per day	
Weight control	Maintain a healthy BMI (18.9–24.9)	
Physical activity	30–60 min aerobic activity four to six times a week	
Risk factors	*Goal*	*Recommended drugs*
Blood lipid		Statins
LDL cholesterol	< 130 mg dl^{-1}	
Total cholesterol	< 200 mg dl^{-1}	
Blood pressure		Low-dose thiazide
Systolic	130–139 mmHg	Diuretics
Diastolic	85–89 mmHg	β-blockers
		ACE inhibitors
		Calcium antagonist
Prophylactic therapy	*Purpose*	
Aspirin (75–325 mg per day)	For secondary prevention and maybe for primary prevention	
ACE inhibitors	For secondary prevention	
Statins	For both primary and secondary prevention	
Folic acid	For primary prevention	

BMI, body mass index; LDL, low-density lipoprotein; ACE, angiotensin-converting enzyme.

primary and secondary preventions are addressed simultaneously. Combining both primary and secondary prevention, the overall objective of CHD prevention is to reduce the risk of initial or subsequent CHD events, thereby reducing premature CHD-related morbidity and mortality, and prolonging survival.

Primary and Secondary Preventive Measures for CHD

Smoking

Cigarette smoking is a convincing risk factor for CHD, and cessation of smoking (including all forms of tobacco) rapidly and significantly decreases the risk for the development of CHD and CHD-related mortality in a very cost-effective way. Smoking cessation is a complex, multifactorial, and difficult process. Simple brief counseling with a clear, firm, positive, encouraging, and sympathetic attitude is the first step in the management of smoking cessation. Persistent smokers may require extended counseling, more than one session preferably, by a skilled counselor or psychologist. Nicotine chewing gum or transdermal nicotine replacement patches with a progressively decreasing dosage can be helpful for heavily addictive persons as an adjunct to counseling.

Family support is very important in smoking cessation. Encouragement from a family or spouse is of great help. 'Smoke-free' environments at home, restaurants, work sites, public transport vehicles, and other public premises can greatly improve the atmosphere for smoking cessation attempts by smokers.

Diet

Diet is an important determinant for the development of CHD by affecting the obesity, blood pressure, and blood lipids. Therefore, it is crucial to maintain a 'healthy diet,' which constitutes (1) consumption of six or more servings of whole-grain products with ≥ 25 g of fiber, especially soluble fiber; (2) consumption of at least five servings of fruits and vegetables (> 400 g per day); (3) limiting consumption of saturated fat < 10% (preferably < 7%) of total daily calories, and cholesterol < 300 mg per day (preferably < 200 mg per day); (4) consumption of polyunsaturated fatty acids up to 10% of total daily calories; (5) limiting salt intake to 6 g per day, the equivalent of 100 mmol of sodium per day; and (6) limiting alcohol consumption to one or two standard drinks for men per day and not more than one drink for women (a pregnant woman should not drink) per day to prevent or delay the development of CHD risk.

Overweight and Obesity

Patients who are overweight (body mass index (BMI = kg m^{-2}) > 25 kg m^{-2}) or obese (BMI > 30 kg m^{-2}), or with excess truncal or abdominal fat (waist circumference > 102 cm in men and > 88 cm in women), are at increased risk for CHD, because of the adverse effects of fat on blood lipid levels, blood

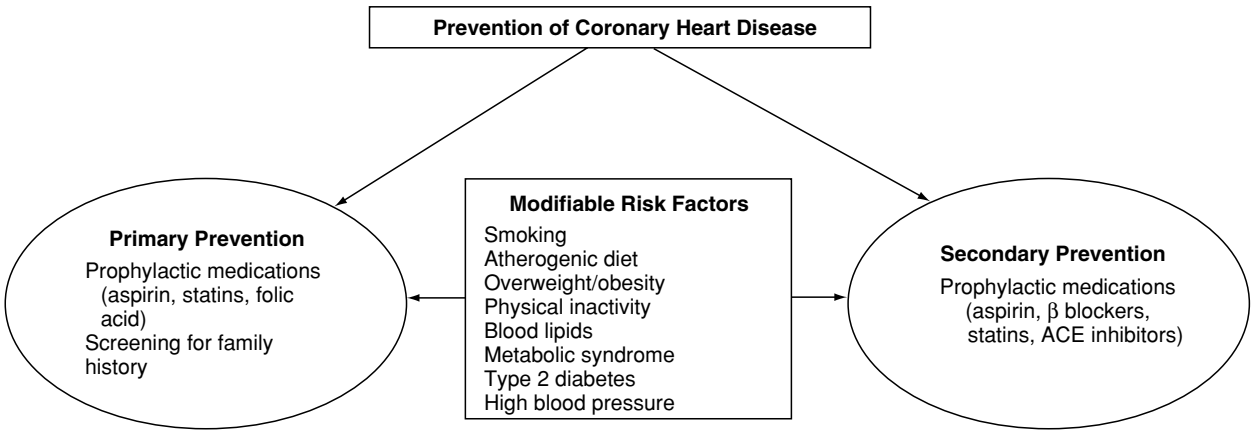

Figure 1 Shared preventable measures in primary and secondary prevention of coronary heart disease. ACE, angiotensin-converting enzyme.

pressure, and glucose tolerance. Prospective epidemiological studies have shown a J-shaped association between BMI and all-cause mortality among Western populations. High-risk populations require professional advice on weight reduction.

Successful weight reduction is not easy to manage, often requiring long-term professional support and encouragement, good motivation of the person, a calorie-restricted lipid-lowering diet, and a suitable physical activity program. Realistic weight-reduction goals should be based on short-term results and agreed between the physician and the patient. It is recommended that a weight reduction of 0.5–1 kg per week is a suitable rate until the target weight goal is achieved. The National Heart, Lung, and Blood Institute, USA has recommended that a person with BMI 27–35 should decrease their caloric intake by 300–500 kcal per day, and a person with BMI > 35 should decrease their caloric intake by 500–1000 kcal per day from their diet to achieve a 10% weight loss in 6 months.

There has been considerable concern about the usage of drugs and their possible side-effects for weight reduction. For example, sibutramine, a sympathomimetic, reduces food intake but may cause hypertension and tachycardia. Lipase inhibitor orlistat (Xenical) has significant effect on weight reduction but may cause severe gastrointestinal symptoms. Weight loss is usually observed within 2 weeks of initiation of therapy. However, researchers agree that a reduction in caloric intake combined with regular physical activity may be the most effective and safe method in reducing excess body fat.

Physical Activity

Prospective epidemiological studies have documented that sedentary lifestyle is a strong predictor of CHD risk, irrespective of other risk factors. Therefore, a regular, suitable exercise program is recommended for a sedentary person, since it has favorable effects on body weight, blood pressure, blood lipids, insulin sensitivity, and glucose tolerance. Epidemiological studies have demonstrated strong evidence that aerobic exercise reduces the risk of CHD. Brisk walking, jogging, cycling, mowing, swimming, aerobic dancing, skipping, and other recreational sports that require movement are good examples of aerobic exercise.

It is now recommended that everyone must perform at least 30–60 min of physical exercise four to six times a week. Generally, an exercise program includes a 5–10-min warm-up by stretching and other gentle activity followed by 20–30 min endurance or aerobic phase, and finally 5–10 min of a cool-down phase. Physical exercise prescription (intensity and duration) for patients with clinically established CHD must be based on comprehensive clinical evaluation and results of exercise testing. The Second Joint Task Force of European and other Societies on Coronary Prevention recommends that the intensity and duration of activity should initially be set low and gradually increased according to the limits imposed by exercise-induced symptoms. Patients recovering from myocardial infarction and other cardiac events should also be given advice about a suitable, gradually increasing exercise program. In these patients, exercise programs must be supervised, and ECG monitoring should be used in the early stages.

Blood Lipids

CHD risk is best evaluated by measuring plasma concentrations of total cholesterol, low-density lipoprotein (LDL) cholesterol, and high-density lipoprotein (HDL) cholesterol. Results from animal experiments, epidemiological, and other studies consistently demonstrate an independent, strong,

positive, and graded relationship between high total cholesterol and elevated LDL cholesterol and the risk of CHD development. Recently, several clinical trials consistently showed that LDL-lowering therapy reduces the total mortality, CHD-related mortality, major coronary events, and coronary artery procedures in persons with established CHD and persons with CHD risk equivalent (i.e., diabetes). Based on the results of prospective epidemiological and clinical trials with both angiographic and clinical end points, an LDL cholesterol level of $< 100\,mg\,dl^{-1}$ is set as the goal of therapy in secondary prevention. Two basic components of LDL-lowering therapy are 'therapeutic life changes' (Table 2) and drug therapy.

Therapeutic Life Changes (TLC) The Third Report of the National Cholesterol Education Program (NCEP) Expert Panel, USA, 2001, recommends a multifactoral life style approach that includes: (1) reduction of saturated fat intakes (< 7% of total calories) and cholesterol (< 200 mg per day), (2) consumption of plant sterols (2 g per day) and soluble fiber (10–25 g per day), (3) weight reduction, and (4) increased physical activity.

Drug Therapy In order to attain the goal of LDL cholesterol levels of $< 100\,mg\,dl^{-1}$, many persons with CHD and CHD risk equivalents will need LDL cholesterol-lowering drug therapy in addition to TLC. Low-density lipoprotein cholesterol should be measured on admission or within 24 h in order to decide the need for drug treatment. Currently, four major classes of drugs are widely used to lower LDL cholesterol effectively: hepatic hydroxymethylglutaryl coenzyme A (HMG CoA) reductase inhibitors (statins), bile acid-binding resins, fibric acid derivatives, and nicotinic acid. Recently, many comprehensive clinical trials with statins have robustly documented that this class of drugs is the most effective in lowering LDL cholesterol from angiographical and clinical end points. Statins also have a better safety profile compared with other classes of drugs. Therefore, at present, the statins are the first line of drugs used to lower total and LDL cholesterol levels. In some special conditions, like familial hypercholesterolemia, a combination of drugs (statin and resin or statin, resin, nicotinic acid) may be used.

Blood Pressure

A large number of epidemiological studies have demonstrated that elevated arterial blood pressure is a risk factor for CHD. The goal for blood pressure reduction is to maintain systolic blood pressure (SBP) < 140 mmHg and diastolic blood pressure (DBP) < 90 mmHg. The goal can be achieved by nonpharmacological treatment (lifestyle modification) alone or in combination with pharmacological therapy.

Table 2 Components of therapeutic life changes[a]

Weight reduction	Balance energy intake and expenditure to maintain a healthy weight (BMI 18.5–24.9)
Increased physical activity	30–60 min aerobic exercise most days of the week
Healthy diet	
Saturated fat	< 7% of total calories
Polyunsaturated fat	10% of total calories
Monounsaturated fat	20% of total calories
Total fat	25–35% of total calories
Carbohydrate	50–60% of total calories
Fiber	20–30 g per day
Protein	~ 15% of total calories
Cholesterol	< 200 mg per day

[a]Adapted from the Third Report of the National Cholesterol Education Program (NCEP), USA, 2001.
BMI, body mass index.

Lifestyle modifications Several randomized clinical trials have evaluated the effectiveness of lifestyle interventions on managing mildly elevated blood pressure in a cost-effective way. The essential features of lifestyle modifications are to: (1) maintain a healthy weight or reduction of excess body weight based on BMI; (2) achieve a moderate level of physical fitness by 30–45 min of aerobic physical activity in most days of the week; (3) reduce salt (sodium) intake moderately (100 mmol per day) and not to exceed total salt intake of 6 g per day (however, results from the meta-analysis of randomized clinical trials presently do not recommend a restriction of salt intake for the normotensive population); (4) increase the intakes of potassium (60–90 mmol per day) preferably from fruits and vegetables; and (5) limit alcohol consumption to no more than two standard drinks for men and one for women per day. Table 3 demonstrates the elements of lifestyle modification or nonpharmacological treatment for elevated blood pressure.

Pharmacological treatment Several large-scale, multicenter, randomized, controlled trials have credibly demonstrated that antihypertensive medications lower CHD-related morbidity and mortality and all-cause mortality. The decision to commence pharmacological treatment relies on the degree of blood pressure elevation, presence of target organ damage (e.g., heart disease, nephropathy, peripheral arterial disease, and retinopathy), presence of clinical cardiovascular disease and other risk factors, or conditions such as diabetes. A general agreement, however, is that an individual with a 'high normal BP' (130–139/85–89 mmHg) and with target organ disease and/or diabetes should be treated with

Table 3 Lifestyle modifications for high blood pressure

Lose weight if overweight or obese
Increase aerobic physical activity 30–45 min for most days of the week
Stop smoking completely
Limit sodium intake to 6 g or less per day
Maintain adequate intake of dietary potassium
Limit alcohol intake to one or two drinks per day

antihypertensive medication after a trial of lifestyle modification. Drug treatment begins with monotherapy, low doses followed by slow titration to achieve 24-h control with once-daily medication depending on the patient's age, needs, and response, as well as economic condition. The most commonly used antihypertensive agents: thiazide diuretics in low dosage, β-blockers (especially if angina is present), angiotensin-converting enzymes (ACE) inhibitors, calcium antagonists, α-blockers, and angiotensin II receptor blockers. Most experts prefer low-dose thiazide diuretics or β-blockers as a first line of antihypertensive agents, unless contraindicated. Other preferred choices for initial therapy are: β-blockers for hypertensives with uncomplicated myocardial infarction; ACE inhibitors for individuals with CHD and decreased ejection fraction, with left ventricular function, diabetes with proteinuria; and thiazide, β-blockers, or calcium antagonist (calcium channel blocker) for isolated systolic hypertension. Isolated systolic hypertension is a CHD risk factor, especially for the elderly, and should be managed properly.

The antihypertensive agents are not free of adverse effects. However, the use of combinations of low doses of two agents from different classes have been shown to potentiate the effect of other agents without incurring any adverse effects.

Blood Glucose

Data from large numbers of epidemiological and (including the Framingham Study) pathological studies document hyperglycemia as an independent risk factor for CHD. Persons with both type 1 and type 2 diabetes and diabetic women are at increased risk for developing CHD than nondiabetics and diabetic men, respectively. The underlying mechanisms for this remain to be answered. In addition to hyperglycemia, diabetics frequently have other classical risk factors for CHD – hypertension, lack of physical activity, low levels of HDL cholesterol, and hypertriglyceridemia– that intensify the risk of developing CHD events.

The effects of glycemic control among diabetics and alteration in CHD risk have not been appraised conclusively. Results from randomized controlled trials show that prudent insulin therapy retards the occurrence of microvascular events and a nonsignificant reduction of macrovascular events in patients with type 1 diabetes. To date, no clinical or epidemiological evidence of the role of blood glucose control in reducing CHD events significantly in patients with type 2 diabetes has been documented. However, there is general agreement that control of blood glucose (hemoglobin $A_{1c} < 7.0\%$) among type 2 diabetics may minimize CHD complications. In order to achieve an acceptable level of hemoglobin A_{1c}, type 2 diabetics often need professional dietary advice (reducing total calorie and fat intake), reduction of overweight or obesity (5–9 kg irrespective of starting weight), and increased level of physical activity as a first line of management. Frequent aerobic exercise (at least 30 min most days of the week) also facilitates blood glucose control and weight control. Oral hypoglycemic agents come as a second line of therapy when a desirable hemoglobin A_{1c} cannot be achieved with lifestyle modification alone.

Table 4 Guidelines to reduce coronary heart disease in patients with diabetes mellitus

Risk factors	*Recommendations*
Smoking	Complete cessation
Physical activity	At least 30 min three or four times per week
Weight reduction	5–9 kg irrespective of starting weight
Hemoglobin A_{1c}	Target: $< 7\%$
Fasting blood glucose	Target: 80–120 mg dl^{-1}
LDL cholesterol	Target: < 100 mg dl^{-1}
Calorie restriction	250–500 kcal less than average daily intake
Blood pressure	Target: $< 130/85$ mmHg

Control of blood glucose also maintains normal plasma lipid levels in patients with type 1 diabetes without nephropathy. Diabetic dyslipidemia (elevated triglycerides, low HDL cholesterol), which is more pronounced among type 2 diabetics, may be corrected by normalizing blood glucose using oral hypoglycemic agents. The primary goal of LDL cholesterol < 100 mg dl^{-1} can be achieved by moderate caloric restriction (250–500 cal less than the average daily intake), $\leq 30\%$ fat, $< 7\%$ saturated fat, and < 300 mg of cholesterol per day. Cholesterol-lowering therapy is recommended if dietary modification fails. The primary goal for blood pressure in diabetics is to maintain blood pressure $< 130/85$ mmHg. Moderate weight loss has also been shown to reduce dyslipidemia and hypertension among diabetics. Smoking cessation is obligatory. Guidelines to reduce CHD for diabetics are shown in Table 4.

Metabolic Syndrome

There is mounting recognition that many persons have a collection of multiple major risk factors and

emerging risk factors for CHD; a cluster of these conditions is known as a metabolic syndrome characterized by abdominal obesity, 'lipid triad' (elevated serum triglycerides, low HDL cholesterol, and high LDL cholesterol) hypertension, insulin resistance, and other thrombogenic risk factors. Primary prevention is mainly focused on reducing body weight through diet and by increasing physical activity. Dietary modification and physical activity also effectively normalize dyslipidemia.

Prophylactic Therapies

To prevent the aforementioned classical risk factors and other thrombogenic factors that play pivotal roles in the CHD events, a great deal of emphasis is placed using several prophylactic drug therapies. Likewise, increased consumption of polyunsaturated fatty acids (ω-3 and ω-6) may be used as antithrombic agents.

Aspirin Compelling evidence from randomized studies insinuates that a low dose of aspirin or acetylsalicylic acid (between 75 and 325 mg) daily may considerably reduce subsequent myocardial infarction and all-cause mortality among patients with postmyocardial infarction. With regard to primary prevention, the use of aspirin and the balance between benefits and hazards is not lucid. However, a recent meta-analysis has suggested that aspirin may be used as a primary prevention agent but requires formal accurate estimation of absolute CHD event risk. A maintenance dose of 75 mg of aspirin is recommended as a secondary prevention for all postmyocardial infarctive patients if not contraindicated.

β-blockers Results from the meta-analysis indicate that β-blockers substantially decrease all-cause mortality in patients following myocardial infarction. Therefore, like aspirin therapy, β-blockers should be reckoned for patients following myocardial infarction if not contraindicated.

ACE inhibitors ACE inhibitors significantly reduce all-cause mortality and the risk of progression of heart failure among patients following myocardial infarction and with chronic left ventricular systolic dysfunction.

Statins Although TLC is considered as the basis of prevention to lower the high total cholesterol and elevated LDL cholesterol, recently, several trials with statins have documented unambiguously that statins may be used as a method for primary prevention without any substantial side-effects. Recent pharmacoeconomic analyses indicate that statin therapy may be highly cost-effective in secondary prevention and moderately so in primary prevention. However, the cost-effectiveness of statins as a primary prevention needs to be evaluated from a global perspective.

Estrogen Abundant observational studies strongly and consistently attest that postmenopausal estrogen replacement therapy (ERT) may protect against CHD. However, recent, multicenter, randomized, placebo-controlled secondary prevention trials have negated this finding and documented that ERT may increase the risk of venous thromboembolism in postmenopausal women with CHD. Therefore, ERT is not recommended for secondary prevention of CHD for the postmenopausal women. The use of ERT may be effective as a primary prevention of CHD. None the less, ERT may be appropriate for postmenopausal women to continue when they are already receiving this treatment.

Folic acid A high level of total plasma homocysteine ($> 15\ \mu mol\ l^{-1}$) is considered as a notable risk factor for CHD based on epidemiological studies. High levels of homocysteine may be acquired, because of a deficiency of folic acid, or congenital, because of genetic mutation of a thermolabile variant of the methylenetetrahydrofolate reductase gene. Supplementation with folic acid (0.5–5.0 mg) reliably and effectively reverses the hyperhomocysteinemia.

Screening

First-degree relatives (parents, siblings, and descendants and other close relatives) of the patient suspected of having familial hypercholesterolemia, or other inherited genetic lipid disorder, or individual with premature CHD (men < 55 years and women < 65 years) should be regularly screened for risk factors as they are at increased risk for CHD.

Conclusion

Prevention of coronary heart disease is a global public health challenge. Given the unique and wide array of the determinants of health (access to quality healthcare, health policies and interventions, physical and social environments, and individual behavior and biology) in the different geographic areas of the world, the orderliness of preventive care for CHD will vary from one country to another. Therefore, it would not be pragmatic to define the aforementioned CHD prevention as a sole model; rather, the universal standards of prevention should be ensued and exercised. The implementation of these recommendations and emerging scientific knowledge in view of CHD prevention should be incorporated in a person's lifestyle as well as in a healthcare provider's professional care. In the near future, the Human Genome Project

will also be able to explicate the genetic factors that predispose to CHD.

See also: **Coronary Heart Disease**: Etiology and Risk Factor; **Dietary Requirements of Adults**; **Exercise**: Metabolic Requirements; **Fats**: Digestion, Absorption, and Transport; **Glucose**: Maintenance of Blood Glucose Level; **Obesity**: Etiology and Diagnosis; **Slimming**: Metabolic Conseqences of Slimming Diets and Weight Maintenance; **Smoking, Diet, and Health**

Further Reading

Assmann G, Carmen R, Cullen P *et al*. for the International Task Force for the Prevention of Coronary Heart Disease (1999) Coronary heart disease: reducing the risk. A worldwide view. *Circulation* 100: 1930–1938.

Assmann G, Cullen P, Jossa F, Lewis B and Mancini M for the International Task Force for the Prevention of Coronary Heart Disease (1999) Coronary heart disease: reducing the risk. The scientific background to primary and secondary prevention of coronary heart disease. A worldwide view. *Arteriosclerosis, Thrombosis and Vascular Biology* 19: 1819–1824.

Barbour M (2000) Hormone replacement therapy should not be used as secondary prevention of coronary heat disease. *Pharmacotherapy* 20: 1021–1027.

Gotto A (1999) Lipid-lowering therapy for the primary prevention of coronary heart disease. *Journal of the American College of Cardiology* 33: 2078–2082.

Grundy S, Balady G, Criqui M *et al*. (1997) Guide to primary prevention of cardiovascular diseases. A statement for healthcare professionals from the Task Force on risk reduction. *Circulation* 95: 2329–2331.

Grundy S, Balady G, Criqui M *et al*. (1998) Primary prevention of coronary heart disease: guidance from Framingham. A statement for healthcare professionals from the AHA Task Force on risk reduction. *Circulation* 97: 1876–1887.

Joint British recommendations on prevention of coronary heart disease in clinical practice: summary (2000) *British Medical Journal* 320: 705–708.

Krauss R, Eckel R, Howard B *et al*. (2000) AHA Dietary Guidelines. Revision 2000: A statement for healthcare professionals from the Nutrition Committee of the American Heart Association. *Circulation* 108: 2284–2299.

Liesbson P and Amsterdam E (1999) Prevention of coronary heart disease. Part I. Primary prevention. *Disease-a-month* 45: 500–571.

Liesbson P and Amsterdam E (2000) Prevention of coronary heart disease. Part II. Secondary prevention, detection of subclinical disease, and emerging risk factors. *Disease-a-month* 46: 4–123.

National Heart, Lung, and Blood Institute (1997) The Sixth Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. NIH Publication No. 98-4080. Bethesda, MD: National Heart, Lung, and Blood Institute.

National Heart, Lung, and Blood Institute (2001) Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) Executive Summary. NIH Publication No. 01-3670. Bethesda, MD: National Heart, Lung, and Blood Institute.

National Heart, Lung, and Blood Institute in cooperation with The National Institute of Diabetes and Digestive and Kidney Diseases (1998) Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults. The Evidence Report. NIH Publication No. 98-4083. Bethesda, MD: National Heart, Lung, and Blood Institute.

Ramsay L, Williams B, Johnston G *et al*. (1999) British Hypertension Society guidelines for hypertension management 1999: summary. *British Medical Journal* 319: 630–635.

WHO (1996) Hypertension control. Report of a WHO Expert Committee. *World Health Organization Technical Report Series* 862: 1–83.

Wood D, Backer G, Faergeman O *et al*. (1998) Prevention of coronary heart disease in clinical practice: Recommendations of the Second Joint Task Force of European and other Societies on coronary prevention. *Atherosclerosis* 140: 199–270.

CORROSION CHEMISTRY

P Marsal, Thionville, France

Background

For most people, the word 'corrosion' carries the image of rust specific to iron-containing metals, surmising that only iron is subject to this phenomenon. In fact, corrosion is the general term for the damage and destruction of most man-made metallic materials. This destructive force of corrosion has always existed but has grown in importance only in the modern era due to the advancement of civilization and technology.

Corrosion may be considered as the deterioration of metals and alloys made by humans, which tend to return to their original state; thus, 'rust' is the natural state of iron, as steel is obtained from iron oxides

found in the earth's crust. The same can be applied to many metals (such as aluminum, zinc, copper, and so on) which are unstable.

Thus, corrosion is an unavoidable problem, and the most important aim is to reduce its occurrence as much as possible. Annual damage due to corrosion in the world is very important. Every domain of industry is affected, including the food industry.

How Corrosion Starts

Prior to corrosion, a disparity or a heterogeneity can always be found, either physical or chemical. The most conspicuous physical disparity is obtained by the contact of two different metals (e.g., tinplate, galvanized steel) present in the same solution (e.g., salt water). Thereby, an electrochemical cell is formed.

Chemical heterogeneity may be represented by differences of ion concentration near identical electrodes or by the Evans cell, which results from differential aeration. The case of a salt water drop on a ferrous material is well known (see **Figure 4**). The anodic site (i.e., the area of corrosion) is always at the center of the drop where there is an oxygen deficit relative to the peripheral zone where aeration (and thus oxygen) is more extensive than in the center of the drop.

Microscopic disparities can also be found; there is no practical industrial process which yields metals with an absolutely homogenous surface, in the strict physical sense: a touch of the metal surface with the hand or simple machining with any tool may be sufficient to damage the surface. A cold-worked zone becomes less 'noble' that is more easily corroded than the surrounding zones as the atoms found in stressed crystals tend more easily to leave the metallic crystal lattice. The various chemical treatments sustained by the metallic material are other sources of heterogeneity.

Electrochemical Basis of Corrosion

A study of corrosion phenomena at a molecular level shows an electron exchange between donating and accepting sites. This is a reduction–oxidation or 'redox' reaction: the oxidizing part (which acts as a cathode) is reduced through electron capture, while electrons are donated by the reducing material (which acts as an anode), which becomes oxidized (**Figure 1**).

Some metal atoms lose electrons to form ions much more easily than others, depending on how strongly the metal's positive ion attracts electrons, and this depends on the nature of the metal. Reactive metals, such as sodium, are those that ionize very easily, whereas unreactive metals, such as platinum, are those that ionize only with difficulty.

- sodium atoms ionize very easily;
- zinc atoms ionize easily;
- iron atoms ionize fairly easily;
- copper atoms ionize with some difficulty;
- gold atoms ionize with great difficulty;
- platinum atoms ionize hardly at all.

Let us take an example: a bar of iron coupled to a bar of copper. A bar of iron builds up a greater concentration of electrons than a bar of copper when dipped in water. In the iron, the electrons are crowded more closely together and so repel each other more strongly. That is, the 'electron pressure' in the iron is greater than in the copper.

The electric potential is more negative (lower) for iron than for copper. (**Figure 2**). What can get the electrons flowing? Something is needed which consumes excess electrons and prevents them from accumulating in the copper bar. Hydrogen ions can do the job. Positive hydrogen ions, short of one electron, are strongly attracted to electrons and combine with them to form neutral hydrogen atoms. In acid water, they are plentiful. At the iron bar, ions continue to leave the surface and pass into the water, and the iron bar starts to corrode. The electrons released by the ionized iron atoms flow through the connecting wire to the copper bar and combine with the hydrogen ions to form gaseous hydrogen. The copper bar cannot corrode because it is protected by an excess of electrons coming from the corrosion of iron. The cathodic protection of iron (in the form of steel) by sacrificial anode (soluble anode) is usually obtained through a coating; galvanized steel is an important industrial example (**Figure 3**). Galvanized steel (zinc plating) on steel dustbins, iron rails, etc. is a very effective form of sacrificial anode protection. It works even when large areas of zinc have gone. As long as some zinc is left in contact with iron, the zinc corodes preferentially. Mildly acidic rain water is usually the electrolyte. The protection life through a metallic coating is roughly proportional to its thickness, as it dissolves progressively to protect base metal. Rust formation, too, is driven by electron exchange between an electron donor site and an electron acceptor site (in this case, oxygen; see **Figure 4**).

Moreover, electron exchange may take place in plain metal when some heterogeneities (physical or

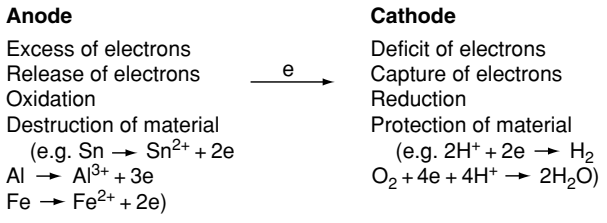

Figure 1 Corrosion as an electrochemical process.

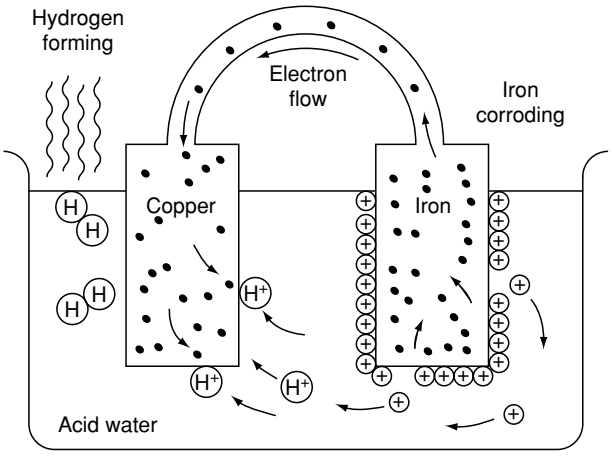

Figure 2 Single electrochemical cell. Reproduced from Corrosion Chemistry, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

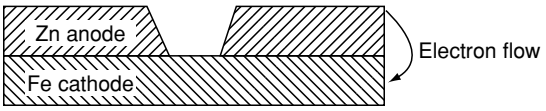

Figure 3 Galvanized steel. Reproduced from Corrosion Chemistry, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

chemical) are present in the metal structure. Microcells form and may cause in pitting (Figure 5).

Parameters Driving Electron Transfer

The most important factors governing electron transfer are potential and easiness.

The potential factor is governed by the electron affinity difference between the two transferring elements. The greater the difference, the faster the corrosion may develop. The potential factor is often called the dissolution potential. Consider, for example, a metal M in a solution containing M^+ ions. The potential of the metal is given by the Nernst formula

$$E = E_0 + \frac{RT}{nF}\log_e(a_{M^+}), \qquad (1)$$

where E_0 is a constant factor, independent of the solution concentration, a_{M^+} is the activity of the M^+ ions in the solution (activity is proportional to concentration), R is the gas constant ($8.31\,J\,K^{-1}\,mol^{-1}$), n is the number of moles of electrons driven by the cell potential E, F is the Faraday (96 500 coulombs), and T is the absolute temperature. The Nernst formula is commonly used as

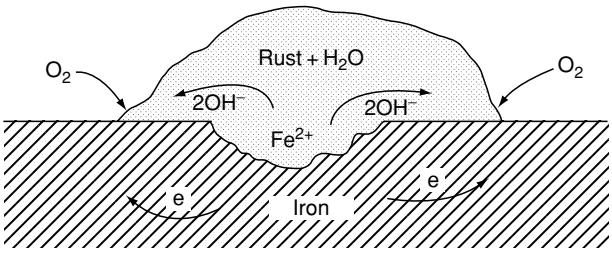

Figure 4 Formation of rust through the formation of a differential aeration (Evans) cell. Reproduced from Corrosion Chemistry, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

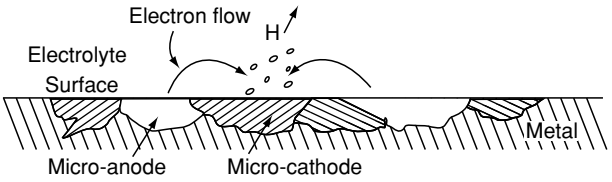

Figure 5 Microelectrochemical cell. Reproduced from Corrosion Chemistry, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

$$E = E_0 + \frac{0.06}{n}\log_{10}(a_{M^+}) \qquad (2)$$

for convenience. From measurements of differences in potential between a metal and a reference electrode, a series of dissolution voltages can be produced (Table 1). Elements that have a greater tendency than hydrogen to lose electrons are described as electropositive, while those that gain electrons are called electronegative, e.g., sodium is more electropositive than aluminum. From this series, it can be seen which of two metals, if placed in contact with each other, will become corroded. For example, in this series, zinc has a much greater dissolution voltage than iron; thus, if a piece of iron is coated with zinc and placed in water, the iron will not rust if the zinc coating is scratched, because zinc loses electrons in preference to iron. This is the principle behind the galvanizing of steel.

The easiness factor reflects any factors affecting the reaction arising from the conditions of the electrolytic media where the electron transfer takes place. If the electron donor (the anodic site) does not easily let electrons loose, the corrosion is said to be donor controlled, and there is an anodic overvoltage. In contrast, the corrosion is controlled by the electron acceptor if it cannot freely accept the electrons; a cathodic overpotential then exists.

Any corrosion cell can be studied through intensity voltage curves or polarization plots. In the laboratory,

special devices quantitatively measure the ease of electron transfer from an anodic site to a cathodic state. Evans diagrams are mostly favored as they are very simple; three examples are shown in Figure 6.

The ideal situation would be a corrosion battery with the lowest possible activity as a result of strong anodic and cathodic overvoltages.

The role of corrosion inhibitors is to increase the overvoltage on the anode, on the cathode, or on both electrodes simultaneously. They polarize the cell, thus reducing its electron flow, often through the formation of an insoluble compound on the surface of one electrode. The corrosion inhibitor behavior (especially the anodic inhibitors) is sometimes unpredictable as most of the compounds used, depending on their concentration and other environmental factors, may have a role as either an accelerator or inhibitor of corrosion. A corrosion accelerator is the reverse of an inhibitor. Every electroactive compound in the electrolyte which is able either to complex the positive ions coming from the anode or to use the electrons arriving at the cathode, either directly or through nascent hydrogen formation, must be considered as a corrosion accelerator.

Table 1 The electrochemical series

Metal	*Electrode reaction*	*Standard electrode potential (V)*
(Active end)		
Sodium	$Na \rightarrow Na^{+} + e$	− 2.712
Magnesium	$Mg \rightarrow Mg^{2+} + 2e$	− 2.34
Beryllium	$Be \rightarrow Be^{2+} + 2e$	− 1.70
Aluminum	$Al \rightarrow Al^{3+} + 3e$	− 1.67
Manganese	$Mn \rightarrow Mn^{2+} + 2e$	− 1.05
Zinc	$Zn \rightarrow Zn^{2+} + 2e$	− 0.762
Chromium	$Cr \rightarrow Cr^{3+} + 3e$	− 0.71
Iron	$Fe \rightarrow Fe^{3+} + 3e$	− 0.44
Cadmium	$Cd \rightarrow Cd^{2+} + 2e$	− 0.402
Cobalt	$Co \rightarrow Co^{2+} + 2e$	− 0.277
Nickel	$Ni \rightarrow Ni^{2+} + 2e$	− 0.250
Tin	$Sn \rightarrow Sn^{2+} + 2e$	− 0.136
Lead	$Pb \rightarrow Pb^{2+} + 2e$	− 0.126
Hydrogen	$H_2 \rightarrow 2H^{+} + 2e$	0.000 (reference)
Copper	$Cu \rightarrow Cu^{2+} + 2e$	+ 0.345
	$Cu \rightarrow Cu^{+} + e$	+ 0.522
Silver	$Ag \rightarrow Ag^{+} + e$	+ 0.800
Platinum	$Pt \rightarrow Pt^{2+} + 2e$	+ 1.2
Gold	$Au \rightarrow Au^{3+} + 3e$	+ 1.42
(Noble end)		

Appearance of Corrosion

Macroscopically, corrosion can appear in the following ways (Figure 7):

- Generalized or uniform corrosion, which appears at the same speed on the whole surface of the metal.
- Pitting corrosion, which appears on localized sites, e.g., at the interface between the metal and an inclusion, on the breaks of a passivation film or coatings. Current density is high in the vicinity of the defect as the anode is small, while the cathode is very large.
- Crevice corrosion, which appears in cracks mainly on stainless steels and under metallic coatings.
- Intergranular corrosion, which appears only on intergranular joints.
- Intragranular corrosion: arborescent corrosion cracks which appear in metallic crystals. Stress often initiates this type of corrosion.

Corrosion Reactions in the Food Industry

Most foodstuffs are in aqueous solution, they are thus conductors and, to some extent, active electrolytes.

Metals used in the food industry are found chiefly in two separate areas:

- metallic cans used for preservation of foodstuffs:

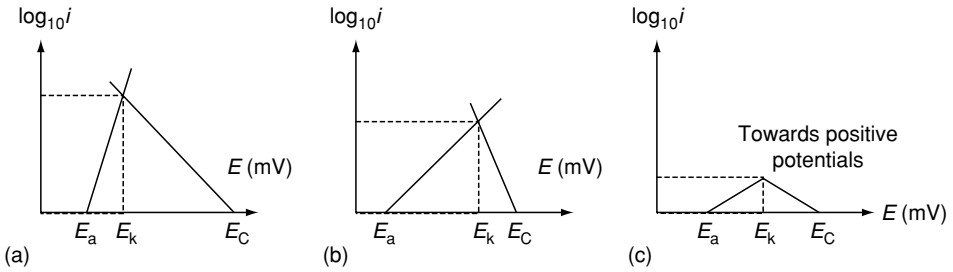

Figure 6 (a) Low anodic overvoltage, high cathodic overvoltage; the galvanic couple is under cathodic control. (b) High anodic overvoltage, low cathodic overvoltage; the galvanic couple is under anodic control. (c) High anodic and cathodic overvoltage; the galvanic couple is under mixed control. $\log_{10} i$, $\log_{10}$ of the corrosion current (or electrons), E_a, from the equilibrium potential: anodic polarization or overvoltage on the electrode supplying the electrons; E_k, mixed and common potential of both; E_c, from the equilibrium potential: cathodic polarization or overvoltage on the electrode capturing the electrons. Reproduced from Corrosion Chemistry, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

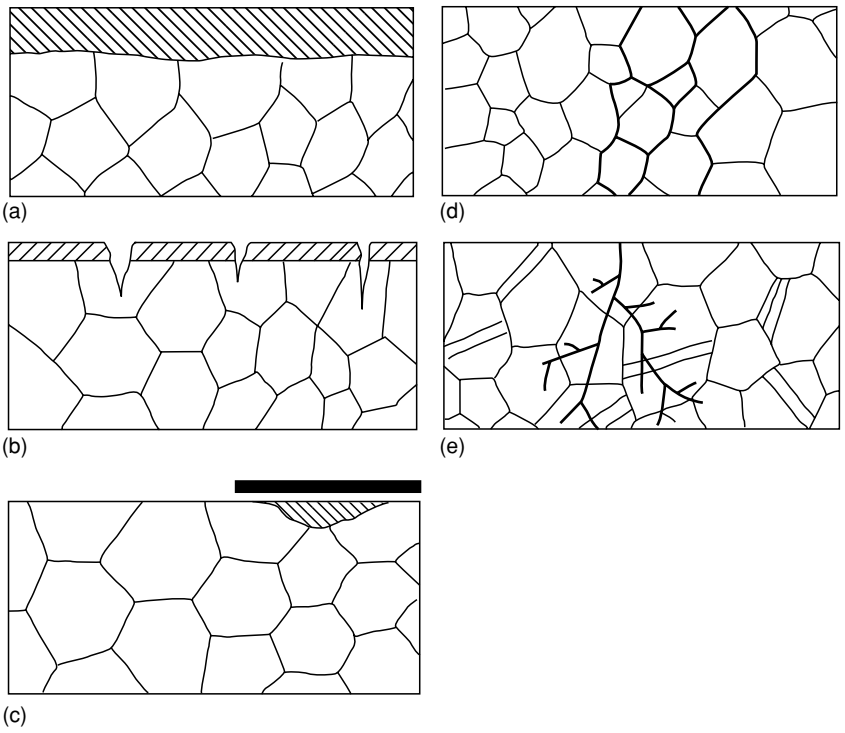

Figure 7 (a) Generalized or uniform corrosion of the metal. (b) Pitting corrosion. (c) Crevice corrosion. (d) Intergranular corrosion. (e) Intragranular corrosion. Reproduced from Macrae R, Robinson RK and Sadler MJ (eds) (1993) Corrosion Chemistry, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Academic Press.

- processing equipment and storage vessels – aluminum alloys and stainless steels are most widely used for this application.

Corrosion of Metal Cans

Worldwide, over 200 billion cans are used every year to preserve and protect a very wide range of foods. Preservation of foodstuffs in closed cans by heat was proposed by the Frenchman Nicolas Appert at the beginning of the nineteenth century. A few years later, the first metallic cans were produced in England from tinplate, a material made from the tin/iron couple used since the Middle Ages for kitchenware. The preserved food industry developed most rapidly in the USA.

The shelf-life of cans should be several years, that is if the rate of corrosion is kept as low as possible. The metal canning of foodstuffs makes use of two base materials:

- Steel coated on both sides with a layer of tin of varying but uniform thickness (0.4 – 1.6 μm). For some 20 years, tin has been replaced, from time to time, by a layer of metallic, oxidized chromium not thicker than 0.015 μm. This composite is usually called 'tin-free steel.'
- Aluminum-based alloys, which have only been used extensively since World War II.

Tinplate is a very asymmetric material; the behavior of the tin/iron cell is shown schematically in Figure 8. Tin in the presence of most foodstuffs behaves as a sacrificial anode, and steel is thus cathodically protected. The anodic behavior of tin is due to the formation of many complexes with organic acids and phenolic compounds (tannins) found in fruits and vegetables. The uncomplexed fraction of tin in solution or in the foodstuff is very low. Following the Nernst equation, tin is more electronegative than it appears from the electrochemical series. (*See* **Tin**.)

Tinplate producers sell about 15 million tonnes per year, two-thirds of which is used for foodstuff canning. Tinplate is not a single product: many types can be manufactured, depending on the quality of the steel, the thickness of the tin coating and sometimes the quality of the lacquer.

Pure tin exhibits good resistance to the acid medium of foodstuffs without oxidizing agents. This corrosion is considered as normal. It even has a beneficial effect on the retention of several organoleptic qualities of foodstuffs. The most important corrosion accelerators (that is, electron acceptors) which may be found in foods are oxygen (air), sulfur dioxide (preservative), sulfur (pesticide), nitrates (water, fruit, vegetables) and trimethylamine oxide (fish). Farming techniques

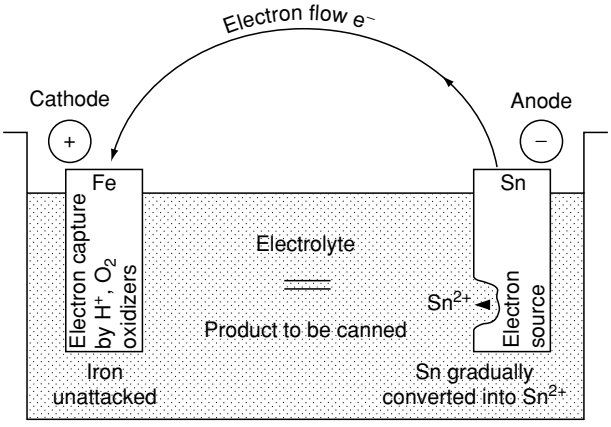

Figure 8 Schematic diagram of a tin/iron cell. Reproduced from Corrosion Chemistry, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

are always changing, and have an influence on the chemical composition of foodstuffs. Corrosion problems have multiplied during the last 20 years as a result of, for example, fast detinning by nitrates, which is even faster as the pH is lowered, and the preferential dissolution of iron due to the presence of pesticide residues such as dithiocarbamates. Contents of up to 40 mg per kilogram of foodstuff, which have no adverse effect on the consumer's health may, in some cases, lead to significant dissolution of tin through corrosion. Similarly, dithiocarbamate residues (a few milligrams per kilogram) are sufficient to invert the tin/iron cell and induce corrosion of the steel. When these additional corrosion risks appear, organic coatings inside the can are needed. For this purpose, lacquers (macromolecular compounds or high polymers) are used. They are insoluble and inert in aqueous media. On account of the nature of the foodstuffs and the various mechanical stresses these lacquers have to withstand, several different types are used. These are characterized either by their barrier behavior or by their degree of flexibility: oleoresinous, organosols, epoxyphenolics, epoxyesters, epoxyurea, and so on are commonly used. These organic films are deposited either on the flat sheet or after can manufacture. In both cases, they are cured thermally in an oven, usually at 200 °C for 10–15 min. Their thickness ranges between 5 and 15 μm, depending on the requirements for preservation of the can and/or the foodstuff. They must also withstand sterilization temperatures (125–130 °C).

Tin-free steels cannot be used alone as they are much more sensitive to acid corrosion, and an organic protective coating is thus essential.

Aluminum is widely used in the food industry, either as a canning material or in various equipment for preparation, storage, and transport of foodstuffs. For metallic canning, it is used much less than steel-based materials: only 2.5 million tonnes per year. In the USA, about 2 million tonnes per year are used due to the large market for beverage cans (beer, soft drinks). In foodstuff cans, aluminum is always used as an alloy (with magnesium or manganese), the exact composition depending on the required mechanical properties needed: for aerosol cans, 99.5% aluminum is used, and for collapsible tubes 99.7%. Aluminum is also used in composite packaging materials associated with polymer films (polyethylene, polypropylene, polyester, Pet, etc.). Theoretically, aluminum is a very passive metal as it can be easily covered with an alumina film (hydrated aluminum oxide). But being an amphoteric metal, it is still highly sensitive to corrosion in acid media (aluminum salts) or alkaline media (aluminates). Pure aluminum is subject to all the corrosion forms mentioned above (uniform, pitting, stress, inter- and intracrystal, galvanic). For this reason, aluminum is often chemically or electrochemically passivated (anodization or chroming) in sulfuric media. The surface treatments greatly improve the adhesion of organic coatings (or lacquers) always used for internal protection of metallic cans to achieve chemical inertness over several years. Epoxyphenolic, vinyl organosols or polyesters are usually used, in one or two layers, as the situation requires. Corrosion of aluminum cans is very rare with such protection. (*See* **Aluminum (Aluminium)**: Properties and Determination; Toxicology.)

Whether the can is in tinplate or aluminum, beer can only be stored in totally inert cans as it is extremely sensitive to contamination by iron or aluminum, resulting in cloudiness and taints.

Low density, high thermal conductivity, and low sensitivity to atmospheric corrosion are the three main specific advantages of aluminum over steel for cans.

Equipment Corrosion

Many aluminum alloys are still used in tanks and containers of all kinds (jars, cans, trolleys, tables, and so on). The meat, fish, milk, cheese, pastry, and confectionery industries all use such equipment. However, in the modern food industry, stainless steels are now more commonly used on account of their good mechanical properties and robustness to frequent cleaning. Similarly, glass and porcelain are now more commonly being used.

There are two main families of stainless steel which possess ferritic and austenitic structures. The first is mostly chromium and iron, while the second contains additional nickel. In both families, some molybdenum may be added to improve corrosion resistance. To prevent corrosion of welded zones, it is necessary

to reduce the carbon content of austenitic steels or to stabilize them by adding titanium and/or niobium. Corrosion resistance and also cost increase with the amounts of other elements used in alloys. It is thus essential to evaluate, as far as possible, the risks and the corrosion resistance properties of the steels used, to eliminate all technical problems, yet at the same time avoiding high costs of production and maintenance. Stainless steels can withstand different types of localized corrosion: intergranular corrosion, pitting, and cracking. Chloride solutions at high temperatures can be highly aggressive: any cracks in the passive layers may lead to rapid corrosion as they become anodic sites with regard to the remainder of the surface. Depassivation may have a mechanical origin (abrasion, wear), and the metal is then permanently depassivated or 'active.' It must be kept in mind that stainless steels have extensive but not universal corrosion resistance. Every use of stainless steel in corrosive media must be treated as a particular problem. The choice of the material must be made in conjunction with a metallurgist, the equipment builder, and the user in order to consider all aspects of the problem (both economic and technical).

Some food industries experience more corrosion problems than others, such as pork butchers and salt meat producers as a result of chlorine- and salt-containing vapor in plants (from brines).

For a long time, the dairy industry large quantities of austenitic stainless steel (18% chromium, 10% nickel) for tubes, heat exchangers, tanks, centrifugation bowls, etc. New varieties of ferritic steels, containing 17% chromium, a low percentage of carbon and some titanium, are now being successfully introduced for boilers.

Corrosion by Cleaning and Disinfecting Products

To obtain regular production of good-quality foodstuffs it is necessary to clean, disinfect, and descale surfaces in contact with food. The frequency of such treatments varies from one industry to another: two to four times a day in the milk industry to once a year in the sugar industry. These treatments require the use of chemical products, which may themselves exert a considerable corrosive action, generating microcavities which will, as the damage increases, be increasingly difficult to clean. (*See* **Cleaning Procedures in the Factory**: Types of Detergent; **Sanitization**.)

With regard to alkaline cleaning agents, stainless steels withstand corrosion well, while aluminium alloys are very sensitive, although corrosion may be reduced by the addition of silicates.

Corrosion of stainless steels in acidic media varies with the particular acid, pH, and oxidizing potential of the solution. In nitric acid solution, austenitic stainless steels face no generalized corrosion, being naturally autopassive. In some cases, slight pitting of the whole surface can be seen. However, aluminum alloys do not exhibit passivation in nitric acid, but here too, it is possible to reduce corrosion speed by adding organic acids such as malic or citric acids. Cleaning solutions based on sulfuric acid always contain corrosion inhibitors for stainless steels.

Chlorination is frequently used by many food industries, owing to its simultaneous strong and cheap disinfecting and bactericidal properties. Its effect is still not fully understood, but it is known that chlorine is effective between pH 7 and 9. As sodium hypochlorite, its effect is due to nascent oxygen, which is very effective against microbial germs. The problem with the use of chlorine is to determine the level of addition required. Resistance to chlorine varies widely between microbial species which are to be killed. In terms of corrosion, chlorine is a strong oxidizing agent. For metallic cans, chlorination of 0.5–2 mg of free chlorine per liter is sufficient to prevent their recontamination through suction via seams in the cans due to the increasing vacuum level developing inside. Higher chlorine levels may induce corrosion phenomena (detinning and rust on tinplate, pitting on aluminum cans).

With the very high chlorine concentrations (300–1500 $mg\,l^{-1}$) needed for thorough disinfection (in dairies, for example), the risk of pitting and crevice formation on stainless steel is higher as the temperature and contact durations increase. The lower the pH value, the higher the risk. Sometimes, corrosion inhibitors are also needed to improve the chemical inertness of stainless steels.

Iodine containing compounds are considered as having no action on stainless steels but should not be used for cleaning and disinfecting aluminum and aluminum alloys.

Solutions of peracetic acid (300 $mg\,l^{-1}$) made from acetic acid, hydrogen peroxide, and water, which have very good bactericidal properties, may be used at room temperature, for short durations (about 20 min) on austenitic steels and aluminum alloys.

Bacterial corrosion, although uncommon, may appear in some food industries, e.g., in buried tubing. Every material, even metals, may be attacked by microorganisms adhering to surfaces and, through their bioactivity, leading to the, accumulation of acids and dissolved gases. For example, we may quote ferrobacteria and sulfate-reducing bacteria. Ferrobacteria, acting on the anodic site, take their energy from the oxidation of ferrous ions to ferric ions, thus initiating the rapid formation of rust as they continuously modify the equilibrium by simultaneous anodic

and cathodic depolarization. Sulfate-reducing bacteria use hydrogen and induce cathodic depolarization: jelly-like vesicles appear, which are living bacterial colonies.

Ways to Prevent Corrosion

Some common ways of preventing corrosion of metals have already been mentioned:

- Organic coatings with inert macromolecular polymers used for cans and for steel-based equipment in food industry plants.
- When using unprotected metals (aluminum alloys and stainless steels), the following must be considered:
  – avoid as far as possible of the joining two metals;
  – choose the best suitable material;
  – modify aggressive media composition with inhibitors;
  – use cathodic protection by coupling a sacrificial anode metal to the material to be protected. The use of a metal as a sacrificial anode does not suppress corrosion and theoretically should not be used for materials in contact with food (unless the corroding metal is specifically authorized for contact with food (e.g., tinplate).

Regulations upon Materials in Contact with Foodstuffs

Corrosion may cause damage during food processing through steam or cooling fluid conveying tubes (perforations or breaks). Equipment failure as a result of corrosion may disrupt the processing of sensitive foodstuffs. If these are to be processed as soon as possible, significant losses may result.

In addition, certain foodstuffs may pick up certain metals. Although some metals may have no toxic effects, they may alter the organoleptic characteristics of the food. For example, beer can turn cloudy in the presence of even small amounts of iron. Also, iron has a deleterious influence on the color of some fruits and vegetables that are rich in phenolic compounds (tannic compounds). Gray or black complexes may form and change the foodstuff hue. For example, white coffee prepared with milk containing very small amounts of iron can take on a grayish appearance.

Sometimes, the presence of metals may have beneficial effects. For example, tin picked up by a foodstuff from the can may preserve the color of clear fruits and vegetables (such as mushrooms, pears, asparagus, and pineapple), owing to the reducing effect of stannous ions (Sn^{2+}).

Firm regulations are always based on a definite listing: any material or object that is not precisely authorized is forbidden. Some local regulations stipulate limits for global or specific migration levels. These vary from one country to another. On a worldwide basis, the *Codex Alimentarius* is the reference. (*See* **Legislation:** Codex.)

Traces of incompletely rinsed cleaning agents may pollute the foodstuff, although this operation is mandatory. The choice of corrosion inhibitors to be added to the cleaning products must always be done with reference to the list of compounds authorized for cleaning.

See also: **Aluminum (Aluminium)**: Properties and Determination; Toxicology; **Cleaning Procedures in the Factory**: Types of Detergent; **Legislation**: Codex; **Sanitization**; **Tin**

Further Reading

Bosich JF (1970) *Corrosion Prevention for Practising Engineers*. New York: Barnes & Noble.

Brun S (1983) Les matériaux en contact avec les aliments. Colloque CNERNA, Paris les 9, 10, 11 février 1983, publié sous la direction de Suzanne Brun. In: *Colloque 'Les Matériaux em Contact avec les Aliments'*. Paris: Technique et Documentation Lavoisier.

Centre Français de la Corrosion (CEFRACOR) (1980) *Corrosion dans les Industries Alimentaires. Symposium, Rennes (France), 23–25 September 1980*. Paris: CEFRACOR.

Crosby NT (1981) *Food Packaging Materials. Aspects of Analysis and Migration of Contaminants*. London: Applied Science.

International Tin Research Institute (1975) *Tin versus Corrosion. Publication No. 150*. Greenford, UK: International Tin Research Institute.

International Tin Research Institute (1980) *Guide to Tinplate. Publication No. 622*. Greenford, UK: International Tin Research Institute.

Landoft D (1993) *Chimie et Corrosion des métauv Traité des* matériaux. vol. 12, Paris: Technique et Documentation Lavoisier.

Marsal P (1965–1985) [Various publications on the subject, particularly on the corrosion of metal packaging including (1981–1982) *On some Corrosion Factors in Canned Foods*; (1977) *Matching Tinplate Cans to their Contents* (1979) *Influence of Nitrates upon Tinplate Corrosion* (1985–1986) *Practical Use of Organic Coatings in the Protection and Decoration of Metal Containers*.] Thionville: French Tinplate Research Centre.

Marsh KS and Brody AL (eds) (1997) *The Wiley Encyclopedia of Packaging Technology*, 2nd edn. Chichester, UK: John Wiley.

Reilly C (1981) *Metal Contamination of Food*. London: Applied Science.

Schweitzer PA (ed.) (1985) *Corrosion and Corrosion Protection Handbook*, 2nd edn. Paris: Technique et Documentation Lavoisier.

Tait WS (1994) *An Introduction to Electrochemical Corrosion Testing for Practicing Engineers and Scientists*. Racine, WI: PairODocs.
Vargel C (ed.) (1979) *Le Comportement de l'Aluminium et de ses Alliages*. Paris: Dunod Technique.
Vargel C (1999) *Corrosion de l'aluminium*. Paris: Dunod Edition.
Vhlig HH (1970) *Corrosion and Protection*. Paris: Dunod Edition.

Crab *See* **Shellfish**: Characteristics of Crustacea; Commercially Important Crustacea; Characteristics of Molluscs; Commercially Important Molluscs; Contamination and Spoilage of Molluscs and Crustaceans; Aquaculture of Commercially Important Molluscs and Crustaceans

Crackers *See* **Biscuits, Cookies, and Crackers**: Nature of the Products; Methods of Manufacture; Chemistry of Biscuit Making; Wafers

Cranberries *See* **Fruits of Temperate Climates**: Commercial and Dietary Importance; Fruits of the Ericacae; Factors Affecting Quality; Improvement and Maintenance of Fruit Germplasm

Crayfish *See* **Shellfish**: Characteristics of Crustacea; Commercially Important Crustacea; Characteristics of Molluscs; Commercially Important Molluscs; Contamination and Spoilage of Molluscs and Crustaceans; Aquaculture of Commercially Important Molluscs and Crustaceans

CREAM

Contents

Types of Cream

C Towler, P A E Cant and K R Palfreyman, Fonterra Research Centre, Palmerston North, New Zealand

Range of Products and Compositional Data

A variety of creams, with different compositions, functions, and types of packaging, can be obtained for consumption (**Figure 1**). Cream consists of emulsified globules of fat in a skim milk serum. A typical particle size distribution in cream is shown in **Figure 2**. Particles less than 1 μm in diameter represent mainly casein micelles; the fat globules are mostly in the 1–10 μm range. The fat provides flavor, and the emulsion form gives characteristic textural attributes and functional properties. The legal classification of products is generally based on fat content. As each country has different laws regarding the composition of different creams, this is not covered here, but **Table 1** gives typical values for the fat contents of different cream types. Other regulations may cover the heat treatment that might be applied to the cream – untreated, pasteurized, sterilized or ultrahigh temperature (UHT)-treated – as well as the limits on

Figure 1 Consumer creams.

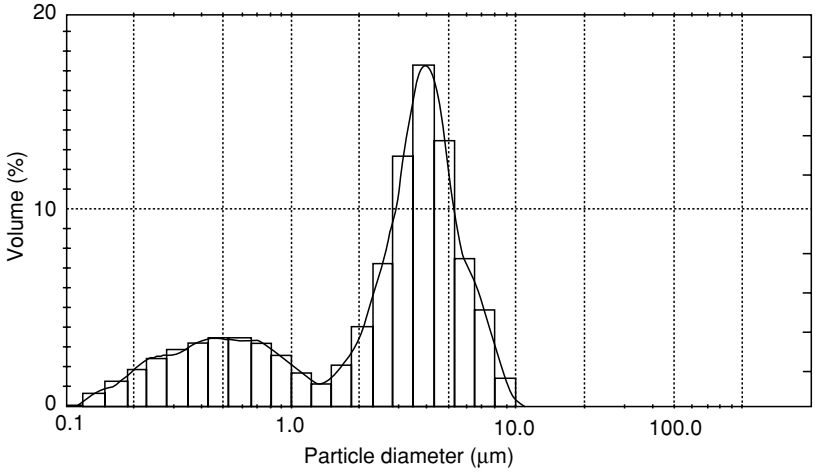

Figure 2 Particle size distribution in cream.

Table 1 Typical fat content of different types of cream

Cream type	*Fat content (%, w/w)*
Clotted	≥ 55
Double	≥ 45
Whipping	≥ 35
Aerosol whipping	≥ 32
In-can-sterilized	≥ 23
Single	≥ 18
Half	12–18

the permitted additives and the normal regulations relating to labelling. Several countries do not allow the sale of unpasteurized cream. Low-fat creams are used as pouring creams for desserts or for addition to coffee or tea. Creams with a higher fat content may be used as whipping creams. Permitted additives vary from country to country. Typical additives are sugar, stabilizers, emulsifiers, and stabilizing salts.

As the fat content increases, the viscosity of the cream increases, and creams with a fat content greater than 60% can be used as spreads. However, the fat content is not the only determining factor in consistency, and spreadable creams can be made with a

lower fat content by reducing the fat globule diameters (homogenization) and adding thickeners. Although cream is generally defined by its fat content, the suspending serum is also important. This serum consists largely of water (approximately 91%) containing lactose (approximately 5%), protein (approximately 2.8% casein and 0.8% whey protein) and other minor constituents such as minerals (0.7%) and vitamins. The levels of fat and milk solids-not-fat (MSNF) in milk are influenced by breed, nutrition of the cow, and lactational or seasonal factors. Another very important component of cream is the membrane that surrounds the fat globules, and this has been the subject of many scientific studies. The major components of this membrane are proteins (41%), phospholipids (27%), neutral glycerides (14%), water (13%), cerebrosides (3%), and cholesterol (2%). Many of the properties of cream are influenced by the membrane and its surface-active components as they affect the stability of the globules and their tendency to agglomerate. Vitamins, minerals, and enzymes are important minor components of the fat and the membrane.

Production and Packaging of Cream

Separation and Standardization

Cream is produced from whole milk by 'separation,' which relies on the density difference between the fat and the aqueous serum. The fat globules will rise in milk according to Stokes' law:

$$v_g = (d^2(\rho_f - \rho_l)g)/18\eta,$$

where v_g = velocity of globule (m s^{-1}); d = diameter of globule (m); ρ_f = density of globule (kg m^{-3}); ρ_l = density of serum (kg m^{-3}); g = acceleration due to gravity (m s^{-2}); η = viscosity of serum (kg m^{-1} s^{-1}). Note that v_g is negative as the equation represents a velocity of settling: $\rho_f < \rho_l$.

The rate of separation can be increased by applying a centrifugal force field, and this provides the basis of the milk separator:

$$v_g = (d^2(\rho_f - \rho_l)r\omega^2)/18\eta,$$

where r = radial distance of the globule from the axis of rotation (m); ω = angular velocity (rad s^{-1}).

The continuous separation of the fat-rich fraction (cream) and the serum (skim milk) is achieved through a stack of rotating disks into which the milk is distributed (Figure 3). Each gap between the disks acts as a zone of separation. Separation takes place in the gaps between the disks, where the denser aqueous phase moves outwards at a greater velocity than the fat globules and is channeled *via* the underside of the disks to the outside and the skim milk outlet. The fat globules concentrate toward the axis of the spinning disks and are channeled out via the upper surface of the disks to the cream outlet. The position of the rising channels is important in maximizing the separation efficiency, and their position on the disks should be in relation to the flows of the two products.

The rotational energy of the streams can be converted to hydrostatic pressure by paring disks (centripetal pumps) and used to pump the products away. Efficiency of separation is measured by the fat content in the skim milk, usually in the form of very small fat

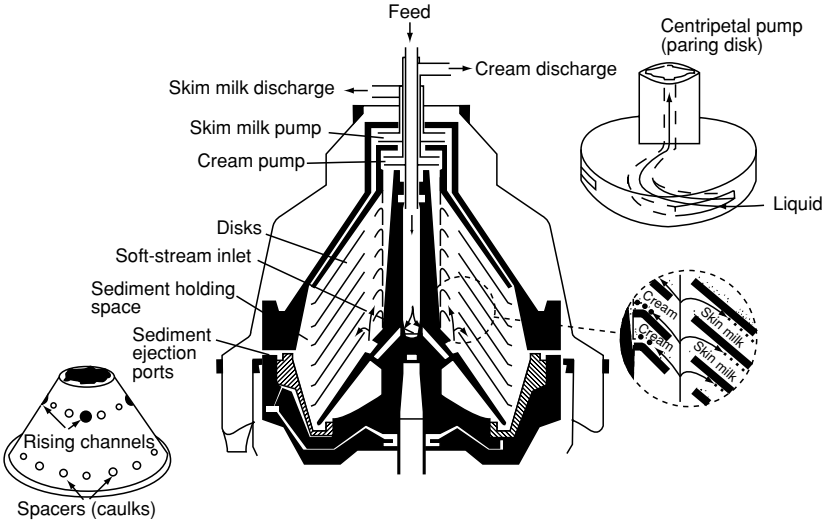

Figure 3 Elements of a milk separator. From Lehmann HR and Zettier K-H (1994) *Separators for the Dairy Industry, Technical Scientific Documentation No. 7*, 4th revised edn. Oelde, Germany: Westfalia Separator AG, with permission.

globules (less than 1 μm). Separation efficiency is influenced by the geometry of the disks (angle, diameter, etc.), their distance apart, and the rotational speed of the bowl. However, there are practical limits to these parameters, and little improvement is made even with very high rotational speeds. A commercial separator would normally work at between 4500 and 6000 rev min^{-1}, depending on size, to give cream with a fat content of approximately 40% and skim milk with a fat content of 0.06%. Incorporation of air markedly reduces the separation efficiency. Hermetic or hydrohermetic sealing systems are used to minimize air incorporation. Milk is fed into a hermetic separator through a hollow drive spindle. The cream and the skim milk are removed by rotating pumps in their respective chambers. The milk attains the rotational speed of the bowl more gradually in the hermetic separator when compared with the paring disk separator where the milk attains the rotational speed of the bowl almost instantaeously. The absence of air and the gentle action of the hermetic separator result in less free fat in the cream than with a paring disk separator.

The fat content of the cream is controlled by the relative flows of the outlet streams. If the flow of cream is restricted, the fat content will increase, but the separation efficiency will decrease if the fat content becomes very high. The separation temperature also influences the separation efficiency due to the effects on the cream viscosity and the relative densities of the fat and the serum. However, higher temperatures may disrupt the membrane, resulting in more free fat in the cream. Phospholipids migrate from the membrane into the serum as the temperature increases; this affects the functional properties of the cream, notably whipping, which may be adversely affected. Optimum separation of milk is normally achieved at 52–55 °C, but some separators are designed to separate milk at around 5 °C, the temperature at which it reaches the processing facility. Such separators have wider gaps between the disks than conventional separators to cope with the more viscous cream. Cold milk separators are always of fully hermetic design. The separation efficiency is not as good as with conventional separators, and there is a greater chance of damaging the fat globule membrane with a high proportion of solid fat in the cream.

Cream is normally standardized continuously by automatic control of the various streams. The fat content of the cream can be monitored through rapid instrumental analysis or by in-line density measurement, and the resultant signals may be used by process logic controllers (PLCs) to adjust the back pressures in the product lines automatically, in order to control the fat contents.

Pasteurization, Sterilization, and Packaging

Cream may be batch-pasteurized at a temperature of 63–65 °C and a holding time of approximately 30 min. Continuous pasteurization in a plate heat exchanger and holding tube is more common. Although 72 °C for 15 s is the legal minimum for milk pasteurization in most countries, the US Food and Drug Administration (FDA) recommends that equivalent heat treatment can be achieved for creams with fat contents >10% by increasing the pasteurization temperature by 3 °C at any given holding time. This recommendation has been incorporated into the New Zealand dairy industry pasteurization standard for cream, with the proviso that the minimum holding time at temperatures greater than 84.7 °C is set at 1 s to give an adequate safety margin. The time/temperature relationship for cream pasteurization at temperatures below 84.7 °C can be expressed as:

$$\log t = (14885/(T + 276.1)) - 41.97,$$

where t = minimum holding time (s); T = minimum temperature (°C).

Temperatures above 80 °C can shorten the shelf-life of the cream, possibly through the activation of bacterial spores. It is important that the cream is handled with care during any processing; positive pumps are recommended to avoid disruption of the fat globule membrane and the release of free fat. Pasteurized cream can be packed in glass, cartons, or plastic pots. Cartons require a water-tight barrier between the cardboard and the cream, and polyethylene is almost universally used, having replaced wax. Pots can be made of polystyrene or polypropylene, the latter being more popular. Covers can be aluminum foil/plastic laminates that are heat-sealed on to the top of the containers with a snap-on cover, or they can be screw-on plastic tops, normally with a tamper-proof plastic ring.

Alternatively, the cream can be in-can-sterilized for extended shelf-life. The cream, which must be of low bacterial count and low acidity, is standardized to a fat content close to 23%, which represents the legal minimum in many countries. Such cream is often called 'reduced cream.' The cream is preheated and homogenized. Homogenization breaks the large fat globules into smaller globules to reduce the tendency for the fat globules to rise (cream), at the same time creating a more stable fat globule membrane as protein is adsorbed from the serum to cover the new fat/serum interface. To obtain a final product without graininess, it may be necessary to add stabilizing salts, such as trisodium citrate, a sodium phosphate, or sodium carbonate, which increase the availability of serum casein due to calcium sequestration and/or

an increase in pH. Other stabilizers, such as carrageenan or sodium alginate, may also be added to enhance the stability and inhibit the separation and syneresis during storage of the processed cream. The cream is filled into lacquered cans that are sealed and sterilized in a suitable retort or continuous sterilizer. Conditions for sterilization are normally 115–120 °C for 10–20 min. Holding times depend on the size of the can and whether agitation is applied during retorting. The high viscosity of the cream reduces heat transfer, and the volume of a can is restricted (<300 ml) as the prolonged heat treatment necessary for larger volumes would increase Maillard browning reactions between protein and lactose, and adversely affect the color and flavor. The heat treatment does have an adverse effect on the stability of the cream, and some agglomeration of fat globules does occur (Figure 4).

Cream may be UHT-sterilized (≥140 °C for ≥2 s) and aseptically packed. The aseptic packaging material includes laminated cartons with plastic (e.g., polyethylene) on the inside, an intermediate aluminum foil barrier, and cardboard on the outside. The cartons may be preformed or formed continuously from a roll of laminate. The material must be first sterilized with hydrogen peroxide and the residual sterilant removed by draining and heating before filling. Pack sizes range from 100 ml to 1 l. Plastic pots may also be used, thermoformed continuously from a sheet before aseptic filling in a laminar filtered airflow and final heat sealing of an aluminum foil/plastic laminate on top of the pot. Pack sizes range from about 10 ml (single serve coffee cream) to 1 l. Lacquered metal cans are used for UHT aerosol cream.

Cream is packed in bulk in plastic pouches or plastic/foil laminate pouches contained in cartons or returnable plastic crates. UHT cream may be packed in bulk using special aseptic filling systems. Volumes up to 1000 l can be packaged, but bulk packaged cream would normally be retailed in packs of 5–25 l.

Cream is susceptible to deterioration because of microbial, enzymatic, and physicochemical changes. Pasteurized cream has a limited shelf-life of a few days and must be kept chilled (< 5 °C). Although microbial deterioration is the major hazard, lipases in the cream release fatty acids, which can result in rancid flavors. The fat globules will tend to rise to the surface and agglomerate unless the cream is homogenized. Homogenization is advantageous for low fat creams to increase the viscosity and inhibit creaming. Storage at a low temperature retards the 'plugging' of nonhomogenized cream as creaming is retarded due to the higher viscosity, and the lower proportion of liquid fat in the globules minimizes agglomeration. Cream in transparent containers should be kept in the dark, as light will cause photo-induced oxidative changes in the milkfat, with consequent deterioration in flavor. The shelf-life of sterilized cream is limited by chemical and physical changes. Spore-forming organisms may be a problem if not killed by preheating. In-can-sterilized cream has a long shelf-life at ambient temperature, although Maillard reactions, which are initiated in the sterilization process, and serum separation (syneresis) may occur during storage. The greatest problem with UHT cream is physical separation of the fat with subsequent agglomeration, although this can be minimized by homogenization. The addition of sodium caseinate or gum stabilizers, such as carrageenan or sodium alginate, will also inhibit the agglomeration of fat globules. Psychrotrophic bacteria in the milk may release heat-resistant proteolytic enzymes, which will result in coagulation and bitter flavors. The shelf-life of UHT cream is limited to some extent by

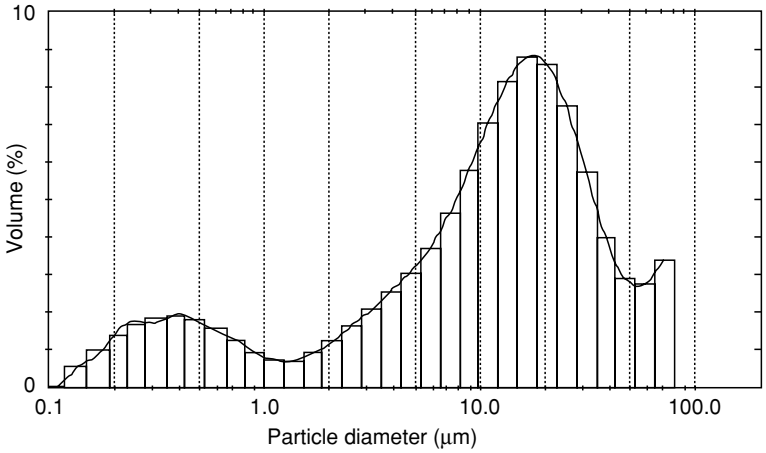

Figure 4 Particle size distribution in an in-can-sterilized cream.

its packaging. Plastics are permeable to air, and consequent oxidation adversely affects the flavor. A layer of aluminum foil in a laminate will prevent oxygen entry, but the flavor will still deteriorate due to non-oxidative reactions.

Freezing of Cream

Freezing of cream will provide protection against microbial deterioration. However, unless the freezing is very rapid, large ice crystals will disrupt the emulsion, resulting in separation on thawing. Bulk frozen cream is used for manufacturing purposes when phase separation is not important or a homogenization step is included. Sometimes, cream is frozen at times of the season when production is high. This cream can be used subsequently for butter production when stocks of fresh cream are insufficient to meet market demands. Bulk frozen cream may be used for the production of cream soups or ice cream.

If cream is frozen rapidly, small ice crystals that do not disrupt the emulsion are formed. One commercial process includes the use of a frozen novelty machine that forms thin rods of frozen cream. Another commercial process uses two continuous stainless steel belts that form a continuous film of frozen cream that breaks into flakes. Alternatively, cryogenic freezing with liquid nitrogen may be used. Frozen cream must be stored at less than −18 °C for a long shelf-life; temperature cycling must be avoided as this will result in the formation of large ice crystals that will damage the fat globule membranes.

Uses of Cream

Cultured (Sour) Cream

Cream can be cultured with suitable organisms that metabolize lactose to lactic acid and also provide other flavoring compounds. The reduction of pH coagulates the protein, which thickens the product and gives a somewhat extended shelf-life when kept under refrigeration. Pasteurized single cream (approximately 18% fat) is normally used as the starting material for culturing. Crème fraîche is a higher fat product that is lightly fermented. Sour cream is used as an ingredient in many savoury foods (e.g., stroganoff, vegetable dishes, dressings, snack dips, etc.).

Coffee Cream

Cream is a favorite 'whitening' agent for coffee as it provides a pleasant texture and flavor to complement the coffee flavor. Long-life coffee creamers can be produced by sterilization (UHT) or by drying. UHT coffee creamers are normally low-fat (10–15%) creams that are given an extensive homogenization

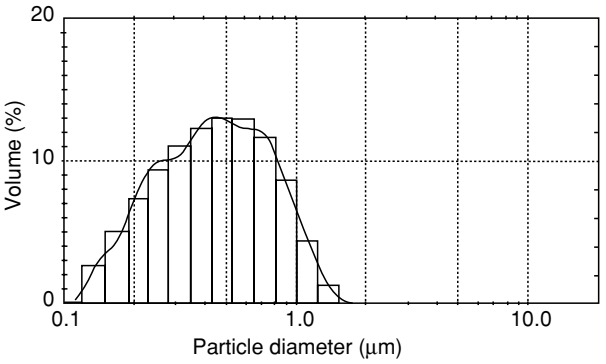

Figure 5 Particle size distribution in a UHT coffee cream.

(Figure 5). The particle size distribution shows a cream with all fat globules less than 2 μm in diameter. The heat and acidity of coffee provide a hostile environment for the cream, and destabilization of the emulsion can result in 'feathering,' with curdled cream forming a characteristic pattern on the surface of the coffee. The presence of calcium in the cream enhances the incidence of feathering, and the presence of calcium-sequestering agents, such as trisodium citrate or sodium phosphates, helps to alleviate this. Feathering may develop with time of storage of the coffee cream as calcium migrates from the serum to the fat globule membrane. Extra protein in the form of sodium caseinate will help to stabilize the cream against feathering. UHT coffee creams are most often packaged in single-serve form-fill-seal containers of about 15 ml.

Cream can be dried to form a powdered coffee creamer. Products available on the market are normally recombined products based on vegetable fat (coffee whiteners).

Whipped Cream

Mechanical agitation of cream will introduce air into the cream as dispersed bubbles. The protein in the cream provides stability for the formation of a foam. The fat globules concentrate at the air/serum interface, where the surface tension is believed to disrupt the fat globule membrane. As agitation continues, the air bubbles become smaller, and the globules interact to form a stable network (Figure 6). The interaction depends on several factors. The fat content has to be sufficiently high (approximately >28%) to give the necessary density of globules at the interface. The membrane around the fat must not be too stable, as some mechanical breakdown is necessary to develop the interaction and 'welding' of the globules. Homogenization has an inhibitory effect on whipping. The temperature and the aging of the cream are important. If the temperature is above 10 °C, excessive

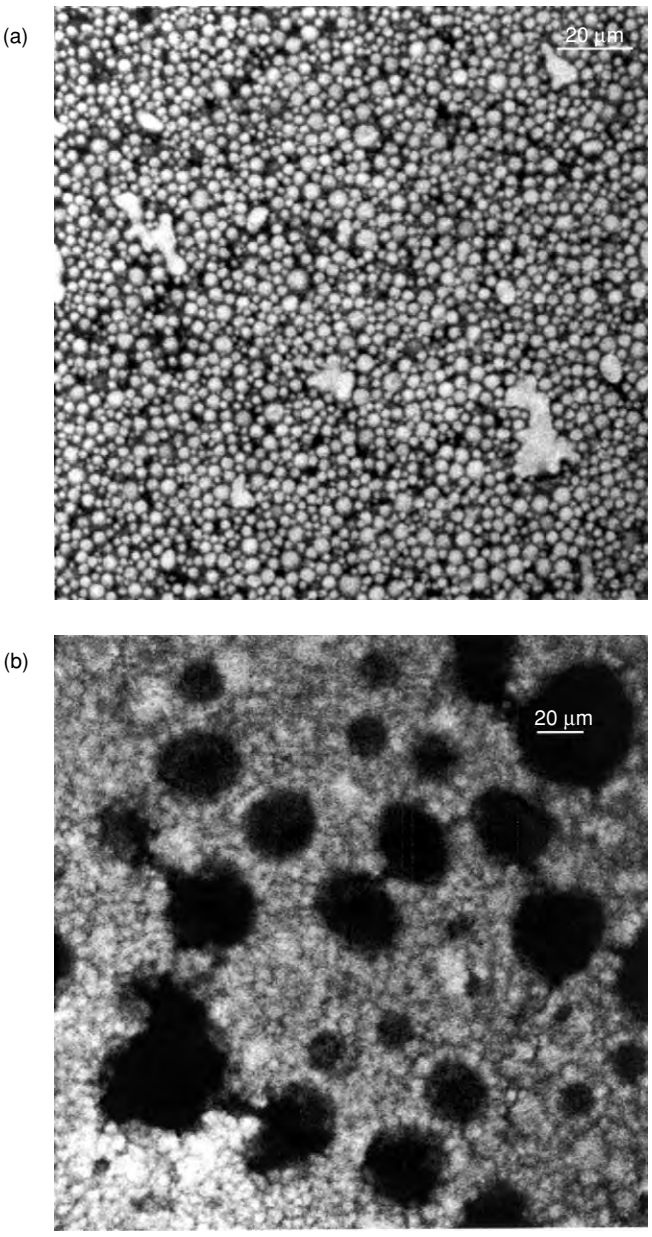

Figure 6 Confocal images of (a) unwhipped cream and (b) whipped cream (dark objects are air cells).

liquid fat will weaken the structure and act as a foam depressant. Cream should be stored refrigerated for at least 12 h before whipping to optimize fat crystallization and membrane destabilization. Long-life whipping cream can be produced by UHT treatment. Some homogenization is necessary to prevent 'creaming' and consequent superficial agglomeration of the cream. The whipping time of UHT cream is thus generally longer than that of normal non-homogenized cream. To give stability during storage, addition of gum stabilizers, such as carrageenan, helps to prevent the association and agglomeration of the fat globules. Whipping properties can be improved by the addition of emulsifiers. Emulsifiers can provide extra surface activity and increase overrun (percentage increase in the volume of the liquid cream when converted to whipped cream). Some emulsifiers (e.g., polysorbates) will increase fat globule interactions during whipping and consequently reduce the whip time and give a stiffer whip. The shelf-life of UHT whipping cream is considerably extended by refrigerated storage. Temperatures $> 25\,^{\circ}C$ and temperature cycling are particularly deleterious to the cream properties. Aerosol cream is packed under pressure with nitrous oxide gas. When the dispenser is activated, the cream is forced out of

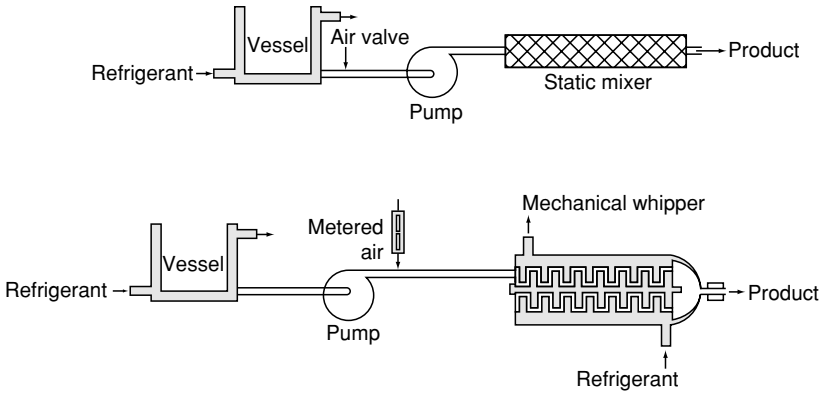

Figure 7 Continuous cream whippers.

the can by gas pressure, and the foam structure is then formed by the expanding gas that was dissolved in the cream. The percentage increase in volume (overrun) of aerosol cream (400–500%) is much greater than that of normal whipped cream (approximately 100%). As a result, the stability of the foamed aerosol cream is not as great as that of normal mechanically whipped cream, but the interaction of the fat globules is still important in providing stability.

The major industrial use of cream is as whipped cream in baking and confectionery to decorate cakes and desserts. Whipped cream can be produced commercially in large batch mixers or in continuous machines (Figure 7).

Control of cream whipping is very important as 'overwhipping' will result in the complete breakdown of the emulsion with 'churning' of the fat and release of free serum. 'Underwhipping' will give a soft foam without adequate structure to hold its shape. As whipped cream has a better freeze–thaw stability than the unwhipped product, it can be incorporated successfully in frozen confections. The addition of stabilizer, such as carrageenan or sodium alginate, will reduce the tendency of the whipped product to synerese.

Cream Liqueurs

Cream liqueurs are popular beverages that use the preservative powers of alcohol for a long shelf-life. Typically, the liqueurs contain 15% fat from the cream, 19% sucrose, and 14% alcohol. Higher-proof products contain 10% fat and 19% alcohol. Emulsification and emulsion stability are provided by sodium caseinate with a fat:caseinate ratio of 1:5, i.e., 3% caseinate in standard liqueurs and 2% caseinate in high-proof liqueurs. Other emulsifiers may be added (e.g., glycerol monostearate (GMS) at 0.34%) to give extra emulsion stability. The most common form of destabilization was found to be calcium-induced aggregation of fat globules, so sodium citrate was introduced into the formula at a level of approximately 0.16% to sequester free calcium in the serum of the liqueur. Some care is required in the use of the citrate as calcium citrate crystals may precipitate from the liqueur if used in excess. The liqueurs are manufactured by dispersing the sodium caseinate and sodium citrate in water, adding the cream and GMS to give a cream base. Ethanol and dissolved sucrose are then added to form a preemulsion that is homogenized to give the final product. The mixing and homogenization are conducted at approximately 55 °C. The ethanol can come from a variety of spirits, and flavors are added to differentiate individual brands of product. Extensive homogenization (e.g., twice at 17.5 MPa and twice at 4 MPa) assists in yielding a stable emulsion. Lack of sufficient homogenization can lead to formation of a 'neck plug' of aggregated fat during storage.

Other Uses of Cream

Cream is the raw material for the production of butter and anhydrous milkfat (AMF). In New Zealand, where cows are pasture-fed, cream for buttermaking is routinely steam-stripped and vacuum-treated during pasteurization. This reduces volatile flavor components from the pasture to desirable levels in the final products.

The Vacreator® (Figure 8) has long been the equipment of choice for cream treatment in New Zealand, but is now being superseded by the Flavourtech® spinning cone column (Figure 9). In both the Flavourtech® and the Vacreator®, steam is injected counter to the flow of cream, and then the vapors and liquid cream are separated under vacuum. The Flavourtech® has a series of fixed and spinning cones that expose the cream to the steam over a large surface area for more effective stripping of the volatile components.

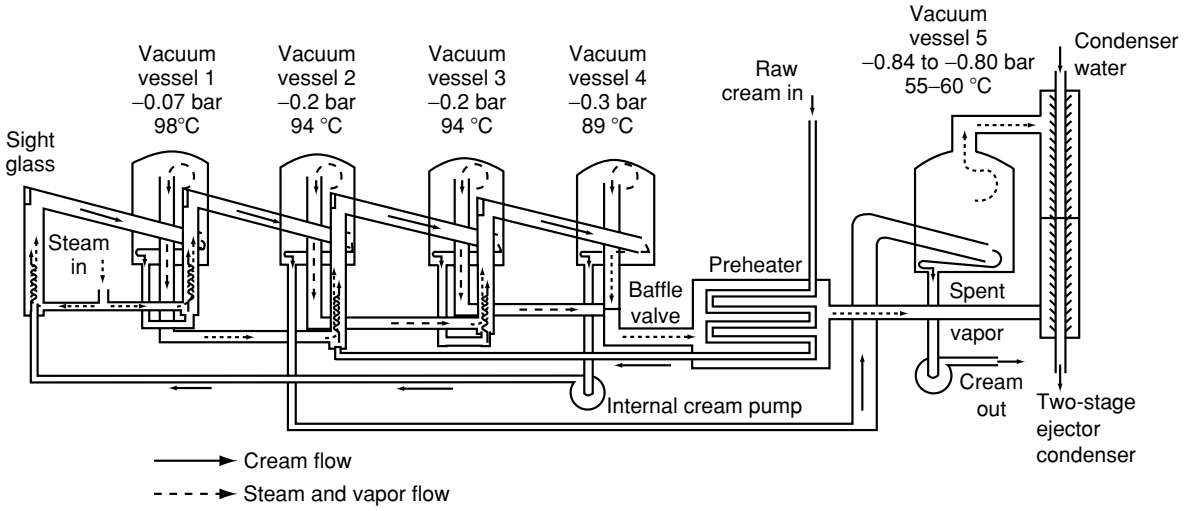

Figure 8 Vacreator®.

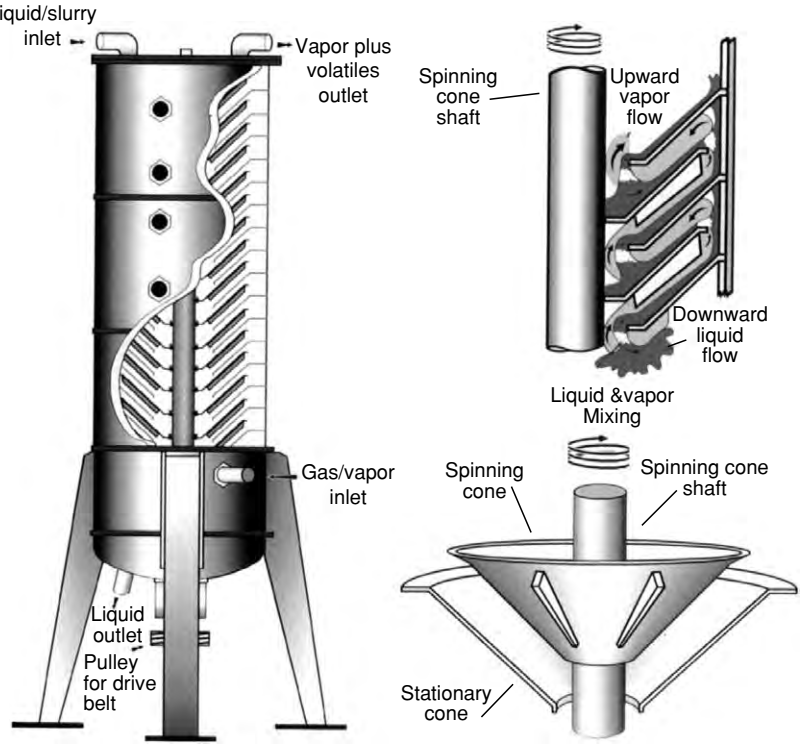

Figure 9 Flavourtech® spinning cone column. From Flavourtech Pty. Ltd., Sydney, Australia, with permission.

Cream can be used as a source of milkfat in foods. It contains more flavor components than butter or AMF. Fresh cream is generally regarded as being the optimum milkfat source for best-quality icecream.

Cream can be dried to produce high-fat powders that may be used as ingredients in foods and may be particularly useful for dry blends (e.g., cake mixes, soup mixes, or icecream powders). The spray drying of cream demands formulations to suit the end use and special driers are required to handle the high-fat product. Cream powders are not produced in large quantities because vegetable-fat-based products are much cheaper and easier to produce. Cream-based powders have the advantage of a superior flavor, but protection against lipid oxidation is required to give the product adequate shelf-life.

See also: **Butter**: The Product and its Manufacture; **Colloids and Emulsions**; **Freezing**: Cryogenic Freezing; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; **Ice Cream**: Methods of Manufacture; Properties and Analysis; **Lactic Acid Bacteria**; **Liqueurs**: Cream Liqueurs; **Oxidation of Food Components**; **Packaging**: Packaging of Liquids; **Pasteurization**: Principles; Pasteurization of Liquid Products; **Separation and Clarification**; **Stabilizers**: Applications; **Sterilization of Foods**; **Storage Stability**: Mechanisms of Degradation

Further Reading

Anderson M and Brooker BE (1988) Dairy foams. In: Dickinson E and Stainsby G (eds) *Advances in Food Emulsions and Foams*, pp. 221–255. London: Elsevier Applied Science.

Bird J (1991) Cream separation. *Journal of the Society of Dairy Technology* 44: 61–63.

Buchheim W and Dejmek P (1997) Milk and dairy-type emulsions. In: Friberg SE and Larsson K (eds) *Food Emulsions*, 3rd edn. pp. 235–278. New York: Marcel Dekker.

International Dairy Federation (1992) *Monograph on the Pasteurization of Cream. Bulletin of the International Dairy Federation No. 271*. Brussels: International Dairy Federation.

International Dairy Federation (1996) UHT Cream. *Bulletin of the International Dairy Federation* No. 315, pp. 3–34. Brussels: International Dairy Federation.

Keenan TW and Dylewski DP (1995) Intracellular origin of milk lipid globules and the nature and structure of the milk lipid globule membrane. In: Fox PF (ed.) *Advanced Dairy Chemistry, Lipids*, 2nd edn, vol. 2, pp. 89–130. London: Chapman & Hall.

Lehmann HR and Zettier K-H (1994) *Separators for the Dairy Industry, Technical Scientific Documentation No. 7*, 4th revised edn. Oelde, Germany: Westfalia Separator AG.

Lynch AG and Mulvihill DM (1997) Effect of sodium caseiante on the stability of cream liqueurs. *International Journal of Dairy Technology* 50: 1–7.

Muir DD (1989) Cream liqueurs. *Journal of the Society of Dairy Technology* 42: 31.

Mulder H and Walstra P (1974) *The Milk Fat Globule. Emulsion Science as Applied to Milk Products and Comparable Foods*. Wallingford, UK: CAB International/ Wageningen, The Netherlands: PUDOC.

New Zealand Dairy Board (1999) *NZCP 7: Pasteurisation Equipment and Instrumentation Code of Practice*. Wellington: New Zealand Dairy Board.

Rothwell J (ed.) (1989) *Cream Processing Manual*, 2nd edn. Huntingdon, UK: The Society of Dairy Technology.

Towler C (1994) Developments in cream separation and processing. In: Robinson R K (ed.) *Modern Dairy Technology, Advances in Milk Processing*, 2nd edn. vol. 1, pp. 61–105. London: Chapman & Hall.

Walstra P (1995) Physical chemistry of milk fat globules. In: Fox PF (ed.), *Advanced Dairy Chemistry, Lipids*, 2nd edn. vol. 2, pp. 131–178. London: Chapman & Hall.

Clotted Cream

S Jolley, Otaihanga, Paraparaumu, New Zealand

History

Clotted, or scalded, cream is made principally in the West Country of England, in the counties of Devon, Cornwall, and Somerset. The process dates back a couple of centuries to a time when these counties were a long and difficult journey from the main consumers of milk in the cities, well before the invention of the mechanical cream separator. Clotted cream was made originally on farms rearing and fattening cattle, where it was found that skimmed milk was more easily digested by young calves than rich full-cream milk, and if the remaining cream was heated, it kept longer and had a pleasant flavor.

The predominant cattle breeds in the area at the time, Jersey, Guernsey, Devon, and South Devon, produced milk of a high fat percentage, with large fat globules that rose to the surface in a thick, firm cream layer. Milk was left to stand in shallow setting pans overnight, for up to 24 h, and heated slowly the next day for 2–3 h until the surface became wrinkled but unbroken, and bubbles appeared just below the surface. It is at this stage that the characteristic flavor developed. Originally, the cream was heated on a tripod over the fire in an open hearth and later on a kitchen range.

The pan was then moved to a cool place allowing the cream to cool slowly and form the deep yellow crust while the body of the cream thickened. The cream was then skimmed off using flat perforated skimmers.

The advent of small hand-driven mechanical separators for farm use simplified the whole process. Milk straight from the cow could be run through the separator, the warm milk at about 35 °C being the right temperature for efficient separation, and the cream scalded immediately. The separated cream was often floated on a small amount of skim for scalding on the stove, to avoid scorching the bottom of the cream layer.

At a later stage, scalding was more likely to take place in a piece of equipment like a *bain marie*. The setting pan containing the cream fitted into an aperture on the top of a water bath, allowing the base and sides of the pan to sit in the hot water circulating underneath.

The Product Today – Craft or Science

Clotted cream is still made on many farms in the West Country and on a commercial scale in several creameries, but little is made elsewhere. The area is a

well-known tourist destination, with Devonshire and Cornish clotted cream teas a specialty, famous throughout the world. One or two small manufacturers in New Zealand, Australia, the USA, and Canada cater for their local market.

Although there remain many Channel Island and Devon cattle in the West Country, much of the milk is now produced by Friesians and other breeds of cow the milk of which is whiter, with a lower fat percentage and smaller fat globules. This factor has an effect on both color and texture of the final product, as do the quality of pasture fed and the use of concentrated feeds. For those producing clotted cream all year round, there is the additional complication of seasonal differences in fat composition, requiring adjustments in the time/temperature combinations used in manufacture to achieve a consistent final product.

Clotted cream is a seasonal product, time-consuming and labor-intensive to make, so commercial enterprises are usually privately owned dairies making clotted cream on a small scale and having the capacity to increase production during the season. It is also a luxury product and will never have a very large market share, so expensive research time cannot be justified to improve manufacturing methods. Any developments in manufacture are largely individual to suit the milk supply of the operation, the experience of the manufacturer, and, ultimately, their customers. No two farm or factory operations are the same, and details of the process will vary in each case.

The traditional product has a sweet, nutty flavor, with a golden crust of grainy texture and a body of cream that is thick, smooth, and deep cream in color. Related products are made commercially for the confectionery trade. Apart from its use in the traditional Devonshire or Cornish cream tea, when it is served with scones or splits (light bread rolls) and strawberry jam, clotted cream can be used on raw fruit, particularly strawberries and raspberries; hot and cold desserts; in savory situations such as on quiches and with baked potatoes; and in many dishes such as sauces and curries, where it adds a rich, full flavor all its own. Although limited quantities are sent from the West Country to cities and catering establishments throughout the country, most is sold locally.

Methods of Manufacture

Farmhouse Production

Farmers who sell clotted cream as a lucrative sideline, usually separate the milk to a fat content of 55–60% as it leaves the parlor, using a small-scale centrifugal separator, either electrically or mechanically operated. The cream is poured into large stainless steel trays, usually rectangular in shape, for the heating process. If the cream is for sale to retail outlets, it can be poured directly into retail containers, provided that they are manufactured to withstand the heating process.

Heating is carried out by one of two methods. The following times and temperatures are only guides. Each operator will have to judge by the results until the most suitable combination is achieved.

1. Forced-air convection oven: The trays or retail containers are placed in the oven at 90 °C for 50–60 min. Initial cooling stages can be carried out in the oven by turning off the heat but allowing the fan to continue to run for up to an hour. Containers are then transferred to a cooling room or refrigerator to finish cooling, taking care not to shake the cream and disturb the crust. Some operators move the cream to a cold room or refrigerator immediately after heating, but if the cream is cooled too rapidly, the crust will collapse.
2. Scalding in a *bain marie*-type apparatus circulating hot water: The scalder usually consists of a shallow open-topped water container on a free-standing frame, in which hot water at 90–95 °C is circulated. The stainless steel top can be designed to take different sized trays or tubs for retailing. The trays containing the cream, or racks holding pots, are set into the covering top of the apparatus, so that hot water is in contact with the sides and bases of the containers. The cream is heated to 77–88 °C for 45–50 min. After heating, the trays are transferred to a cool room for cooling.

In either case, the essentials of the process are:

- cream should not be more than 50 mm deep in the container to ensure even heating;
- air must be able to circulate freely above the containers to allow evaporation to form the crust;
- the temperature/time combination must be high enough to kill milk-souring organisms and prevent off-flavors;
- the heating temperature and time must be sufficient to develop the flavor and texture, and form the crust;
- the cream should not be disturbed during processing to allow the crust to form;
- postheating contamination must be avoided;
- cooling must be done slowly to achieve flavor development, even thickening, and avoid the crust collapsing, particularly in retail containers.

The criteria on which to judge the results are:

- a rich, golden-colored crust that is wrinkled but not broken, with a grainy texture when cold;

- the body of cream that is smooth, thick, i.e. of spreadable consistency, but not solid, a deep creamy color, and with no separation of liquid at the bottom of the container;
- a full, rich, creamy flavor that is nutty, sweet, and lightly caramelized;
- a legal minimum fat percentage of 55%.

Commercial Manufacture

Commercial-scale manufacture has adopted and developed the traditional methods by employing modern technology. Milk destined for clotted cream is separated using a centrifugal separator at a temperature ranging from 35 to 40 °C, to a fat content of 55–60%. This fat percentage may have to be adjusted during the season as the composition of the butterfat varies, in order to achieve a consistent viscosity in the final product. Separation at this high fat percentage can result in the loss of fat in the skim if a higher separation temperature is used, which will drastically affect the yield of product. The yield will also be affected by evaporation from the cream surface during scalding. Evaporation is necessary for the formation of a good crust, but some operators may try to restrict evaporation and loss of yield at the expense of the crust.

Most commercial operations pasteurize the cream, many using plate-heat exchangers, but small dairies may still use batch pasteurizers. From the point of view of product safety, this step would seem to be unnecessary as the cream is heated to, and held at, a higher temperature than that required for pasteurization during the scalding or heating process. Also, the shear to which the cream is subjected could affect the setting time and thickness of the finished cream. Some operators compensate for this by adding a small percentage of homogenized cream prior to heating.

Commercial heating of the cream can be carried out in hot air using a convection oven, or by an extension of the hot-water scalding process. Some operators use a mixture of both methods, depending on the container and end result required. Either method can be automated to a limited extent.

Forced air convection ovens (see Figure 1) can be obtained in a range of sizes from bench-top models to large free-standing models, their capacities depending on the size of containers in the batch of cream being processed. They are convenient and safe, as they do not involve the use of very hot water, and steam is not emitted into the room. It is essentially a batch process, although it would be interesting to see if a bakery oven with a moving conveyor would work as well.

Ovens are convenient for heating cream in retail containers. Cream is poured into tubs or trays after separation and placed on to racks or wire mesh shelves. These can be either slotted directly into the oven or fitted into a large mobile unit that can be pushed into a freestanding oven. Heating times can vary from 65 °C for 40–45 min for the 21-g packs for airline catering, to 90 °C for 1 h for 227-g (8-oz.) tubs. The depth of crust formed on the airline portions is limited by their size, and there is a loss of flavor when using the lower heating temperature.

Figure 1 Free-standing hot-air convection oven showing overhead fans and a mobile rack on the left-hand side, which has shelves for retail containers.

A typical water scalding process (see Figure 2) involves running the cream into containers ranging from 113-g (4-oz.) tubs to large trays for catering purposes holding 4.53 kg. The containers fit into trays that are put into the scalding apparatus, as described for the farmhouse method. Water at 95 °C is circulated for up to 70 min, after which time the trays are lifted out for cooling (See Figure 3).

Some operators say they stir the cream for the first 10 min to ensure even heating, and variations in scalder design include gabled covers above the trays to avoid heat loss.

One operator has adopted a continuous method of scalding in which the cream is filled into pans holding 1.5–2.25 kg of cream. The pans are passed through

Figure 2 Large *bain marie*-style scalder, designed with a gabled cover to avoid heat loss.

Figure 3 Commercial operator transferring a tray of cream in retail tubs from the scalder to the cooling room.

the heated water bath on bars that move at a given speed to ensure the required scalding time, usually 1–1.5 h.

A recently formed company, which commenced a commercial clotted cream enterprise in 1998, produces three clotted cream products, using a combination of hot-water scalding, a convection oven and batch heating in a closed, jacketed vat. The company processes 600–750 kg daily, depending on the season and makes about 15 tonnes in the 2-week pre-Christmas period.

The company's traditional clotted cream is produced by hot-water scalding in four batches daily. Container sizes include 113-g, 227-g, 453-g and 906-g tubs, and 1.81-kg and 4.53-kg trays for catering establishments. A convection oven is used for small 21-g airline catering portions, because trays holding large numbers can be put into the oven on racks, making them easier to handle. The industrial-grade product is made in a batch pasteurizer using quite a different process.

Thickened Cream for Confectionery

Confectionery products, such as clotted cream fudge, toffee, and icecream, require a product that can be described as clotted cream but do not need the crust, the flavor, or the texture of true clotted cream.

Cream at 55% fat is homogenized at 1315–1754 gsc, heated to 65 °C for 45 min, then cooled to 30–35 °C when it has a thick blancmange-like consistency. This is poured into block molds and frozen. It is kept frozen until required by the confectionery manufacturer. The confectionery demand is increasing, and this outlet helps to even out seasonal requirements for the traditional product.

Cooling and Packing

To produce a good traditional clotted cream, whichever heating method is used, slow cooling is an important factor in achieving a firm even texture and the typical grainy crust. If the cream is cooled too rapidly by putting it into a cold store, at a temperature down to 4 °C, the cream loses its full flavor, and the texture becomes hard.

Containers are transferred from the heating area to a closed room at 12–14 °C, where the air is still, or preferably fitted with fans creating positive pressure, to avoid contamination at this vulnerable stage. The room can be fogged, i.e., the air sprayed, using peracetic acid at 25 p.p.m. to control molds and yeasts. It takes 4–5 h at this temperature for the fat to cool sufficiently to avoid crust collapse and loss of flavor. The containers can then be moved to a cold store maintained at 4 °C for a further 12–24 h before lidding and subsequent storage. Some commercial operations lid the retail pots after 5 h of cooling but run the risk of condensation forming inside the pots.

Steps should be taken to avoid contamination during transportation of cream to and between cooling areas, particularly from dust, the main source of mold and yeast infection. Many operators transfer the trays of cream on trolleys, which must be cleaned daily along with the rest of the equipment.

Farming enterprises and commercial operators who scald cream in large trays, either sell it in the returnable trays directly to catering establishments or pack it at a later stage into retail containers. Farmers may ladle the desired amount of cream into cartons or jars for individual customers, or pack it into lined tins for posting to regular consumers.

One commercial enterprise cools the cream briefly before filling into containers using filling machines while the cream is still liquid. Care must be taken with mechanical handling of the cream at this stage, from the point of view of damaging the texture and postheating contamination. If this practice is followed, the cream must be cooled further before the pots can be sealed. Cream processed in retail containers is best cooled for 24 h or overnight before lidding, but some operators seal the containers much sooner than this.

Lids may be of the snap-on variety, which can be applied by hand in small enterprises, with a hand-held applicator, or automatically in a lidding machine. The individual portions for airlines have a foil lid that is heat-sealed in place. This method would be used for larger retail tubs also.

Shelf-life and Problems

Raw cream of good bacteriological quality and hygienic production methods should give a traditional product with a shelf-life of at least 14 days. The most common problem is the appearance of molds on the surface, caused by postheating contamination during transfers in the cooling and packing operations. An entry air lock to avoid direct contact with outside air or an area with positive air pressure where these parts of the process can be carried out will do much to alleviate this problem.

The production of clotted cream is governed in the UK by Cream Regulations. Intending manufacturers should contact the Ministry of Agriculture, Fisheries, and Food or their local authority to find out the relevant details for their operation. If the product is labeled 'clotted cream,' it must contain no ingredients

other than cream, including flavoring or other added ingredients, whether or not that ingredient is a constituent of milk. Nisin is the only permitted preservative in clotted cream, but it is not used widely. If it is used, it must be declared on the label. (*See* **Cream**: Types of Cream; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; **Nisin**.)

Acknowledgment

Thanks to Torridge Vale Creamery, Torrington, Devon, who produce *Definitely Devon* clotted cream, for supplying information about their commercial process for use in this entry.

See also: **Cream**: Types of Cream; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; **Nisin**

Further Reading

MAFF Publications (1971) *Leaflet 438 Farmhouse Production of Clotted Cream*. Alnwick, UK: Lion House.

Martin C (1999) *Clotted Cream*. Redruth, UK: Tor Mark Press.

Rothwell J (ed.) (1989) *Cream Processing Manual*, 2nd edn. Huntington, UK: The Society of Dairy Technology.

Street L and Singer A (1975) *The Backyard Dairy Book*. Dorchester, UK: Prism Press.

Thear K (ed.) (1978) *The Home Dairying Book*. Saffron Walden, UK: Broad Leys.

Cream Liqueurs *See* **Liqueurs**: Composition; Cream Liqueurs

Crustacea *See* **Shellfish**: Characteristics of Crustacea; Commercially Important Crustacea; Characteristics of Molluscs; Commercially Important Molluscs; Contamination and Spoilage of Molluscs and Crustaceans; Aquaculture of Commercially Important Molluscs and Crustaceans

Cryogens *See* **Freezing**: Principles; Operations; Blast and Plate Freezing; Cryogenic Freezing; Storage of Frozen Foods; Structural and Flavor (Flavour) Changes; Nutritional Value of Frozen Foods

CRYSTALLIZATION

Basic Principles

L Yu and S M Reutzel-Edens, Lilly Research Laboratories, Indianapolis, IN, USA

Background

Crystallization is the physical transformation (*phase transition*) of a liquid, solution, or gas to a crystal, which is a solid with an ordered internal arrangement of molecules, ions, or atoms. Many substances of scientific, technological, and commercial importance are crystalline, ranging from large-tonnage commodity materials to high-value specialty chemicals. Crystalline substances of importance to food sciences and nutrition include sugars, sugar alcohols, salts, fats, fatty acids, artificial sweeteners, etc. Crystallization is a means to isolate chemical substances in the solid form for long-term storage and downstream processing. As a purification technique, crystallization relies on the stringent structural requirement for crystal formation to exclude impurities. Crystallization performed under different conditions can yield crystals of different sizes and morphologies, thus providing a way of modifying particles to desired specifications. The growth of large single crystals is essential to the determination of the structures of molecules and crystals by crystallography and the fabrication of optical and electronic devices. The science of crystallization has been developed on the foundation of thermodynamics, kinetics, fluid dynamics, crystal structures, and interfacial sciences.

This article is divided into three sections. The first section, Principles of Crystallization, is concerned with the thermodynamic conditions of crystallization and the kinetics of crystal nucleation and growth. The second section, Molecular Aspects of Crystallization, discusses how the knowledge of molecular and crystal

structures provides mechanistic insights into crystallization and how elemental steps of crystallization can be manipulated at the molecular level to gain control over competing crystallization pathways. The third section, Techniques of Crystallization, describes the various ways in which crystallization can be performed.

Principles of Crystallization

Crystallization usually occurs through a *nucleation and growth mechanism*. Nucleation is the formation of stable molecular aggregates (*nuclei*) capable of growing into macroscopic crystals. Crystal growth is the actual development of the nuclei into visible dimensions. Both nucleation and growth require a thermodynamic driving force. In the case of solution crystallization, the thermodynamic driving force is *supersaturation*. Supersaturation is generated when the concentration of a solute exceeds its *equilibrium solubility*, the solute concentration in a solution that is in equilibrium with solute crystals. Supersaturation can be expressed as the difference between the concentration of a saturated solution c_{ss} and the equilibrium solubility c_{eq}^* or as the ratio between the two: $\Delta c = c_{ss} - c_{eq}^*$ or $S = c_{ss}/c_{eq}^*$. The thermodynamic driving force for crystallization from a melt is generated by lowering the temperature below the crystal melting point, or *undercooling*. For crystallization to occur from a gas, the vapor pressure of the crystallizing component must exceed its equilibrium vapor pressure, i.e., the vapor pressure attained when the gas is in equilibrium with the crystals.

Although any nonzero supersaturation can in principle cause crystallization from a solution, crystallization usually does not occur unless the supersaturation exceeds a certain threshold, or *metastable limit*. A diagram illustrating this phenomenon is shown in Figure 1. The solid line depicts the equilibrium solubility of a solute as a function of temperature. The region below the solid line is *undersaturated*, where crystallization is thermodynamically impossible. The region above the solid line is supersaturated, where crystallization is thermodynamically possible, and the solution is metastable. The metastable limit is indicated by the dashed line, and the region between the solid and dashed lines is called the *metastable zone*. A supersaturated solution free of crystal seeds can remain in the metastable zone indefinitely. For a solute to crystallize spontaneously, the solution concentration must be raised not only above the equilibrium solubility (solid line), but also above the metastable limit (dashed line) into the so-called *labile zone*. The width of the metastable zone depends on the nature of the solute and conditions of crystallization (e.g., stirring, solvent, temperature, pressure, the presence of impurities, and the surface characteristics of the crystallization vessel).

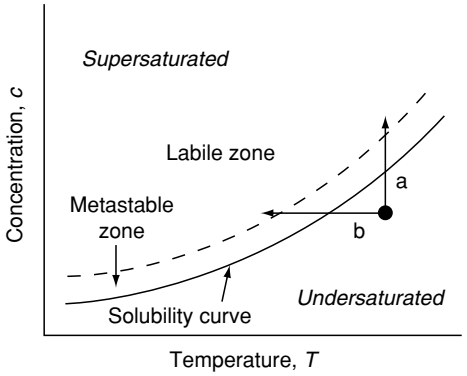

Figure 1 Equilibrium solubility, supersaturation, and metastable limit. An undersaturated solution can be induced to crystallize by solvent evaporation (path a), temperature change (path b), or both.

Crystal nucleation may be classified as primary or secondary. *Primary nucleation* refers to the formation of crystal nuclei from a solution that contained no preexisting crystals. Primary nucleation occurs through both homogeneous and heterogeneous mechanisms. *Homogeneous nucleation* is the formation of nuclei within a homogeneous fluid, and *heterogeneous nucleation* is initiated by contact with foreign particles and surfaces. The effectiveness of surfaces or interfaces as templates for nucleation frequently makes the heterogeneous mechanism the dominant mechanism when particulate contaminants are present. *Secondary nucleation* refers to the generation of crystal nuclei from preexisting crystals, such as those introduced through seeding.

The classical theory of homogeneous nucleation treats crystal nucleation as a thermally activated process whose activation energy is the free energy change (ΔG) for forming a microscopic crystal within a liquid. ΔG is further divided into a surface term (ΔG_s) and a volume term (ΔG_v). ΔG_s is the surface free energy change for creating a crystal–liquid interface, which is thermodynamically unfavorable ($\Delta G_s > 0$) owing to the generation of surface tension. ΔG_v is the volume free energy change for the liquid-to-crystal transformation, which is thermodynamically favorable ($\Delta G_v < 0$) owing to the greater thermodynamic stability of the crystalline product. Thus, the volume and surface terms represent the driving force and the obstacle of nucleation, respectively. The interplay of the two terms dictates whether or not a crystal nucleus is viable, i.e., capable of growing into larger sizes. Since both terms increase in magnitude with the crystal size r, but at different rates (ΔG_s proportional to the surface area and ΔG_v proportional to the volume of the

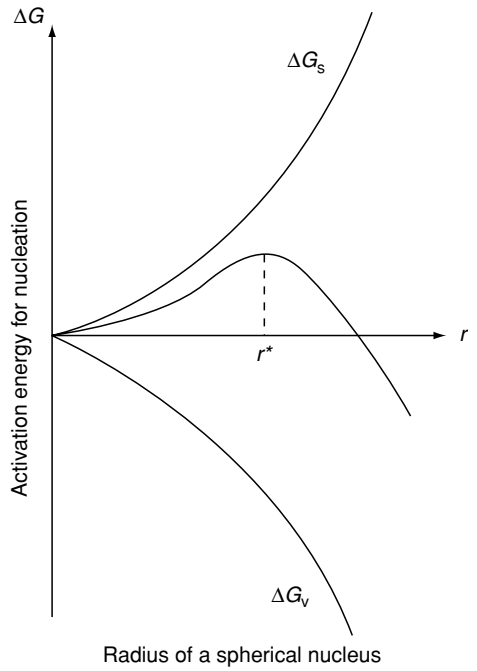

Figure 2 Classical theory of homogeneous nucleation. The free-energy change of crystal nucleation, ΔG, is divided into a surface term ΔG_s, representing the free energy increase for creating a crystal–liquid interface, and a volume term, ΔG_v, representing the free energy decrease for the liquid-to-crystal phase transition. The different dependence of ΔG_s and ΔG_v on the crystal size r leads to a maximum of ΔG at the critical nucleus size r^*. Nuclei dissolve if they are smaller than r^* and grow if they are larger than r^*.

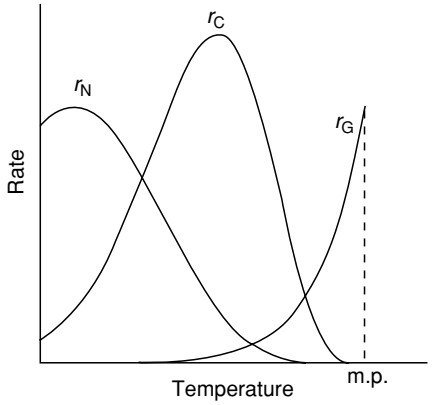

Figure 3 Typical temperature dependence of crystallization rate from a single-component melt. r_G, crystal growth rate; r_N, crystal nucleation rate; r_c, overall crystallization rate. The fastest crystallization rate is achieved at an intermediate temperature between the preferred temperatures for nucleation and growth.

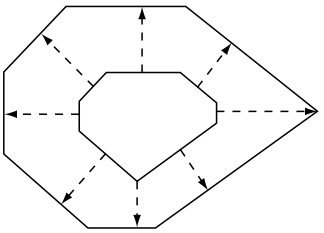

Figure 4 Illustration of the growth of a single crystal. Slower crystal growth rates lead to larger crystal faces.

crystal), ΔG achieves a maximum at some crystal size r^*, termed the *critical nucleus size* (see Figure 2). The critical nucleus size is so named because crystals smaller than r^* will dissolve spontaneously, whereas crystals larger than r^* will grow spontaneously.

The process of crystal growth requires the transfer of mass to the growth interface and the dissipation of heat generated by crystallization. The rate of crystal growth therefore depends on the rates of mass and heat transfer, which in turn depend on the viscosity of the medium and the degree of agitation. Impurities strongly absorbed to crystal surfaces can significantly retard crystal growth and modify crystal morphologies.

Crystallization rate depends on temperature because temperature affects both the molecular motions required for nucleation and growth and the degree of supersaturation. If crystallization occurs from a single-component melt, supersaturation becomes irrelevant, and a characteristic temperature dependence often emerges (Figure 3). The crystal growth rate, r_G, usually increases with temperature until the arrival of the crystal melting point, above which no crystallization is possible. Contingent upon the formation of molecular clusters, nucleation requires a low temperature, but not too low so that molecular mobility becomes limited. Thus, the nucleation rate, r_N, plotted against the temperature typically shows a maximum below the melting point. The combination of the two temperature effects leads to the fastest crystallization rate, r_C, achieved at an optimal temperature that is intermediate between the preferred temperatures for nucleation and growth.

Crystal growth from solution does not proceed at the same rate in all directions. As crystals grow, distinct morphologies or habits emerge as a result of face-specific growth rates. Slower growth rates translate to larger crystal faces and *vice versa*, as shown in Figure 4. The face-specific growth rate depends on both the internal structure of the crystal and growth conditions (e.g., supersaturation and impurities), enabling crystals of the same internal structure to develop different external habits.

A crystallization process may produce crystals of different sizes. The control of particle size distribution relies on the control of nucleation and growth, which can be achieved by adjusting supersaturation, seeding, agitation, resident time, etc. Crystals allowed to remain in contact with a saturated solution can undergo *ripening*, which causes small crystals to

dissolve and large crystals to grow. The driving force for ripening is the slightly higher solubility of smaller crystals owing to their higher surface energies.

Polymorphism is the ability of a molecule to crystallize in multiple crystal forms that differ in molecular packing and/or conformation. A related phenomenon, *pseudopolymorphism*, arises if a molecule can cocrystallize with solvent molecules, forming a *solvate*. A solvate may be stoichiometric or nonstoichiometric, depending on whether the amount of solvent is fixed or variable. Since polymorphs and solvates have different physical properties (e.g., solubility and crystal morphology), more stringent control may be necessary for their crystallization. In crystallizing polymorphic compounds, unstable polymorphs sometimes crystallize first and then transform into stable forms. This observation is summarized by Ostwald's step rule, which states that the first form to crystallize is the most soluble (the least stable) form, which then transforms to the next most soluble form through a process of dissolution and recrystallization, and so on. This rule, which has not gone unchallenged, is applicable to crystallization initiated at sufficiently high supersaturations that the least stable (most soluble) polymorph is driven to crystallize. Obtaining metastable polymorphs frequently requires rapid crystallization and immediate harvest because of potential conversion to more stable polymorphs.

Molecular Aspects of Crystallization

Crystallization at the most basic level is the assembly of molecules; thus, an understanding of the molecular events of crystallization promises the ultimate control of crystallization. Certain crystallization processes have been identified as the assembly of molecular aggregates, rather than single molecules. These molecular aggregates, called *prenucleation aggregates* or *growth units*, are often in the form of dimers, trimers, or high multimers of molecules. The growth units of some carboxylic acids, for example, are found to be cyclic dimers held together by hydrogen bonds. Thus, in the molecular theory of crystallization, a multistep process is envisioned: the association of molecules into prenucleation aggregates in solution, the assembly of prenucleation aggregates into crystal nuclei, and the growth of crystal nuclei into macroscopic crystals (Figure 5).

The identification and manipulation of prenucleation aggregates underlie molecular strategies for crystallization control that go beyond traditional approaches based on the nucleation-growth model. Examples of such strategies include the use of 'tailor-made additives' for polymorph control, morphology modification, and inhibition of unwanted crystallization. Chemical principles of molecular structures and interactions are used to guide the selection of such additives. For example, molecular details of a dominant growth surface can be analyzed to determine suitable impurities that may preferentially bind to and inhibit the growth of the surface, thereby altering the crystal habit. The introduction of impurities with molecular conformations resembling those present in one polymorph ('conformational mimicry') has been used to prevent its crystallization, leading to the crystallization of another polymorph.

Crystal engineering is the design of crystal structures with desired properties (e.g., nonlinear optical activity arising from noncentrosymmetric structures) from molecular structures. The same chemical principles guiding the molecular control of crystallization are employed in such engineering of crystals, as are reliable structural patterns of intermolecular interactions in crystals (e.g., close packing, hydrogen bonding and organometallic coordination chemistry).

The separation of mirror-related molecules that are not superimposable with each other (*enantiomers*) is called *chiral resolution* and is an important application of crystallization. With some molecules, chiral resolution occurs spontaneously; that is, a solution containing opposite enantiomers crystallizes to yield a physical mixture (*conglomerate*) of crystals containing pure enantiomers. A conglomerate can be separated manually into crystals of pure enantiomers, as Louis Pasteur famously demonstrated with sodium ammonium tartrate. With seeding, one enantiomer can be induced to crystallize from a self-resolving solution. This process, known as *kinetic resolution*, requires that the system does not come

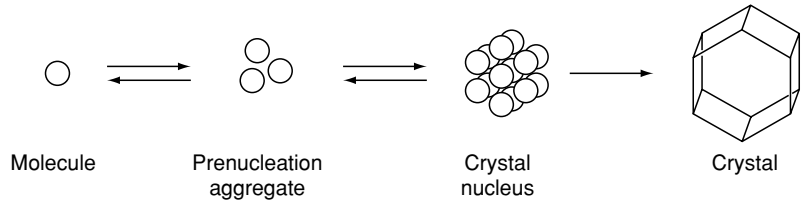

Figure 5 Molecular model of crystallization.

into equilibrium to precipitate both enantiomers. For most chiral molecules, spontaneous resolution is impossible, and opposite enantiomers crystallize together to form a so-called *racemic compound*. In these cases, chiral resolution usually requires the addition of a resolving agent, which is a chiral molecule that forms a crystal with one enantiomer more readily than with the other. Preferential crystallization of one enantiomer may also be accomplished with a chiral solvent, in which the solubility of the opposite enantiomers is different.

Techniques of Crystallization

Crystallization techniques differ in the way in which supersaturation is generated and relieved, and nucleation is initiated. The importance of selecting an appropriate technique lies in the fact that the same crystallization performed under different conditions can yield crystals of different properties (size, morphology, chemical purity, polymorphic form, etc.). The design of a cost-effective crystallization process requires many types of data, including mass and heat transfer and equilibrium solubility as a function of solvent, temperature and pH.

To induce crystallization from a solution, the necessary supersaturation can be generated in several ways, including solvent evaporation, temperature change, antisolvent addition, and chemical reaction. In Figure 1, solvent evaporation at a constant temperature corresponds to a vertical line from the undersaturated to the supersaturated region. Solvent removal at controlled temperatures can be performed either with or without vacuum. Slow evaporation of solvent is frequently used to obtain large, high-quality single crystals.

Crystallization induced by temperature change (*temperature gradient*) takes advantage of the temperature dependence of solubility. In Figure 1, the generation of supersaturation by temperature change without solvent removal corresponds to a horizontal line from the undersaturated to the supersaturated region. If solubility increases with temperature, a high-temperature solution can be cooled to generate supersaturation. In general, slower cooling rates lead to nucleation at higher temperatures, producing fewer and larger crystals. To effect additional control, it is customary to seed the solution so that crystallization is initiated at the desired temperatures. Once crystallization begins, the suspension can be cooled to improve the product yield (assuming that solubility increases with temperature).

Crystallization by *antisolvent addition* (also called *drowning out*) depends on the solvent dependence of solubility. With this technique, a solution to be crystallized is mixed with an antisolvent, which is miscible with the initial solvent and in which the solute is less soluble. The addition of antisolvent lowers the solubility and generates supersaturation. The onset of crystallization is signaled by turbidity, after which precipitation usually follows.

Vapor diffusion is a common technique for growing high-quality single crystals. With this technique, an antisolvent is diffused into a solution through the vapor phase and slowly generates supersaturation. An analogous technique, *liquid diffusion*, is carried out by slow addition of a lower-density antisolvent, allowing the initial solution to float to the top. As the antisolvent gradually diffuses into the initial solution, crystals may grow at the interface.

Crystallization may directly follow chemical reactions that produce low-solubility products. For example, an acidic or basic solute may be crystallized as a less soluble salt with an appropriate counter ion. Acid–base titration is a recrystallization technique used to purify free acids and bases, whereby a soluble salt is first generated at a high or low pH, and then the solution is neutralized to the starting pH to precipitate the solids.

The condensation of vapors produced by *sublimation* provides a way to grow crystals from the gas phase. With this technique, a solid is placed near a condensation surface and heated to produce sufficient vapor pressure. The condensation surface supplies a substrate for the formation of crystals.

Freeze-drying, also called *lyophilization*, is a drying process in which the solvent (typically water) is first frozen and then removed by sublimation in vacuum. Owing to the low temperature used, freeze-drying is frequently a gentle drying technique, especially suitable for thermally labile substances (e.g., proteins and peptides). The low processing temperature often prevents solute crystallization in freeze-drying; nonetheless, many solutes crystallize predictably (e.g., glycine, mannitol and NaCl). (*See* **Freeze-drying**: The Basic Process; Structural and Flavor (Flavour) Changes.)

Spray-drying is a one-step drying process that converts a solution to a powder. In a spray-drier, the feed solution is first atomized into a spray. The spray is then dried through contact with heated air streams. The dried powder is usually separated from the air stream via a cyclone separator and collected at the base of the drying chamber. Spray-dried particles are often roughly spherical and have a narrow size distribution. (*See* **Drying**: Spray Drying.)

Crystallization with the aid of *supercritical fluids* takes advantage of the dissolving power of a substance (e.g., CO_2) existing in the supercritical state, a state of matter created by high pressure and

exhibiting gas- and liquid-like properties. A supercritical fluid rapidly returns to the gaseous state upon reduction of pressure (e.g., by expansion through a nozzle). When this happens, solutes dissolved in a supercritical fluid will precipitate quickly, often yielding fine, evenly sized, and spherical particles. It is also possible to use a supercritical fluid to rapidly extract the solvent from a solution, causing the solute to precipitate.

Spherical crystallization is a technique of forming dense and spherical agglomerates of crystals. This technique is based on adding to a crystallizing medium a small amount of an immiscible liquid (bridging agent) that preferentially wets the developing fine crystals and encourages them to compact into spherical agglomerates.

Epitaxy refers to the oriented growth of one crystal on the surface of another crystalline substance, with the crystallization promoted by the matching periodicity of the two crystal lattices. The lattice match lowers the activation energy for nucleation and growth. Structural studies by crystallography and microscopy of substrate-promoted crystallization have identified epitaxial match as the underlying cause for oriented growth of crystalline overlayers. Epitaxy is also invoked as an explanation for nucleation induced by foreign particles.

See also: **Freeze-drying**: The Basic Process; Structural and Flavor (Flavour) Changes

Further Reading

Davey R and Garside J (2000) *From Molecules to Crystallizers. An Introduction to Crystallization*. Oxford: Oxford University Press.

Garti N and Sato K (eds) (1988) *Crystallization and Polymorphism of Fats and Fatty Acids*. New York: Marcel Dekker.

Jacques J, Collet A and Wilen SH (1991) *Enantiomers, Racemates, and Resolutions*. Malabar, FL: Krieger.

Mullin JW (2001) *Crystallization*, 4th edn. Oxford: Butterworth-Heinemann.

Myerson AS (1993) *Handbook of Industrial Crystallization*. Stoneham, MA: Butterworth.

Cucurbits *See* **Vegetables of Tropical Climates**: Commercial and Dietary Importance; Root Crops of Uplands; Root Crops of Lowlands; Edible Aroids

Cultured Milk Products *See* **Fermented Milks**: Types of Fermented Milks; Dietary Importance; **Yogurt**: The Product and its Manufacture; Yogurt-based Products; Dietary Importance

CURING

M Shimokomaki and E Youssef Youssef, Londrina State University, Londrina, PR, Brazil
N N Terra, Santa Maria Federal University, Santa Maria, RS, Brazil

Introduction

The use of salt and curing is undoubtly the technological procedure which is most commonly employed to transform meat into a meat product of quality. This practice has been applied before Roman times in order to preserve meat products. Curing methods consist of the application of salt (sodium chloride), sodium nitrate, and nitrite, and reducing agents. There is no curing processing without salt and probably nitrate was discovered as an impurity of salt. Sodium nitrite is the most important component since it plays many roles in meat curing. It is the chemical agent which develops the typical color cured meat products through the formation of nitric oxide compound from the reaction of nitrite-myoglobin. In contrast, it helps to develop the typical flavor and increases the stability for oxidative reactions. Most importantly, nitrite inhibits the development of several undesirable pathogenic microorganisms, including *Clostridium botulinum*. These

multiple functions of nitrite make it difficult to find an ideal nitrite substitute; nitrite gives rise to nitrosamine compounds, which are known to be potent carcinogenic agents. However it has recently been reported that on average only one third of daily-ingested nitrite originates in cured meat products. This article will discuss the events that occur in curing processing.

Raw Material

Care should be taken in relation to the raw material. The meat system characteristically presents a weak acid, pH 5.5–6.0, after completion of rigor mortis. This final pH value is important for curing reactions because under these conditions nitrite forms nitrous acid which also participates in curing reactions. This pH control is very relevant, for instance, during cooked ham processing; the pH of the semimembranosus muscle should be above 5.8 and refrigerated pork meat pieces should not be frozen because freezing would create colour defects. The meat temperature should not be too cold as the full series of color reactions cannot take place and the final cured color formation is impaired.

Ingredients

The ingredients applied for curing processing are salt, phosphates, nitrite, sugar, reductants, binders and extenders, and antioxidants.

Salt

Salt (NaCl) is the main ingredient in quantity and has two major functions. It solubilizes myofibril proteins, helping to stabilize meat emulsion: the concentration of 6–8% solution is most effective. In contrast, salt, being a dehydrating agent, would alter osmotic pressure and thus inhibit bacterial growth and subsequent spoilage. Years ago, it was common to use a high concentration of salt and presently this concentration is around 2–3%, therefore it is necessary to store derived meat products under refrigeration. Under higher salt concentrations with subsequent lower water activity values, these products are able to resist bacterial spoilage, for instance, copa, salami, charqui meats in South America, and biltong in South Africa; these are known as intermediate-moisture meat products. Their shelf-life can be extended, even at room temperature, for several months. NaCl is also an important factor to enhance the flavor of meat products. This is a beneficial factor and is why it is included in the preparation of these products but can cause problems by increasing consumers' high blood pressure. However the amount of sodium chloride can be reduced by mixing with potassium chloride, up to 50% concentration.

Sodium Nitrite

Nitrite in meat products has the following functions: (1) it stabilizes the color; (2) it contributes to the flavor of cured meat; (3) it inhibits the development of spoilage microorganisms, in particular avoiding the appearance of botulism; and (4) it delays the development of rancidity. Nitrate is used as a source of nitrite formed after the action of microorganisms and it has not been used so frequently, since it has been substituted by nitrite in meat processing. There are a few exceptions to this: country cured hams and Lebanon bologna.

The color formation in cured meat is the result of a series of complex chemical reactions involving sodium or potassium nitrite (NO_2). This complex series of chemical reactions can be summarized as follows:

$$2KNO_3 > 2KNO_2$$

$$2KNO_2 > 2HNO_2$$

$$2HNO_2 > H_2O + N_2O_3$$

$$N_2O_3 > NO + NO_3$$

The nitric oxide formed combines with myoglobin heme, a constituent of meat pigment, and this colored complex pink color (hemocrome) becomes stable after it has been heated. The other reaction for nitrite is a direct interaction with myoglobin (Fe^{2+}); as this is a strong oxidizing compound, it can produce brown oxidized metmyoglobin (Fe^{3+}). The Fe^{3+} ions formed have the capacity to be active promoters of oxidative reactions and there is a need to convert them to inactive and stable Fe^{2+} forms. This can be achieved by the action of reducing agents added as ascorbate or erithorbate. In the final cured meat, it has been found that residual nitrite found reduced to 10–20% of the original amount added. Only about 5–10% of added nitrite reacts with myoglobin, 1–5% combines with lipid components, and 1–5% forms volatile components.

Phosphates

Sodium phosphates are used in meat processing in order to increase water-binding capacity because this helps to solubilize myofibril proteins. In these functions, only alkaline phosphates are efficient for this role because they raise the pH value, becoming effective water binders. Also they bind heavy metal, having in the end a preservative role to inhibit the development of microbial growth. Normally, phosphates are used in most pumped meats, i.e., ham, bacon, roast

beef, and pastrami, which give the advantage of reducing the cookout, improvement of sliceability, retention of flavor, and juiciness. There is a limit on the quantity of phosphates used of a maximum of around 3000 mg kg^{-1}, because above that level would give a metallic or soapy taste to the product.

Sugar

The addition of sugar to the curing process is primarily for flavor. The presence of sugar counterbalances the harsh taste promoted by a high concentration of salt. In case of fermented products, the addition of sugar is necessary as an energy source for microorganisms to perform fermentation. When heat treatment is used during processing, a browning reaction results in a burning flavor.

Reductants

The isomers ascorbate and erythorbate are commonly used in modern curing practice. Because of their reducing role, their function is primarily as a curing accelerator reducing nitrite to nitric oxide and reducing the meat pigment, thus giving stability to meat product color. By stimulating nitric oxide formation they suppress nitrosamine formation and therefore become an important ingredient, in particular for bacon.

Binders and Extenders

These are used to increase the binding properties of a meat mixture, giving an economic advantage to the producers. Thus, they improve emulsion stability, give better yields and better sliceability, and are permitted in cooked sausage at 3.5%. These compounds are nonmeat proteins like soy and its hydrolyzed products, several starches and of course hydrocolloids like carrageenan.

Antioxidants

There is a group of compounds which function to prevent rancidity. Butylated hydroxytoluene (BHT), butylated hydroxyanisole (BHA), tertiary butyl hydroquinone (TBHQ), and propyl gallate are the major antioxidants in use. Recently the application of natural compounds has become popular and the practice of including for example, vitamin E, in the dietary ration for daily animal consumption has been shown to extend the meat product shelf-life greatly.

Processing

In practice, there are at least three methods of curing: (1) dry curing; (2) wet (pickle) curing; (3) and combined dry and wet curing.

Dry curing

This old-time procedure is the easiest method of curing since all that is needed is to rub the salt alone or in combination with nitrite and/or nitrate and sugars into the surface of the meat. This technique is applied without any water, thus the curing ingredients are dissolved in the moisture of the original samples and then they diffuse. Also, fermented dry-cured sausages are prepared by mixing minced meat and ingredients and stuffing them into natural pork or lamb casings or artificial casings made from reconstituted collagen. Depending on the type of product, they are placed in natural or air-conditioned drying chambers at a specific temperature and time, hence natural microorganisms develop, constituting the ripening of the product. In some cases, starter cultures are added to accelerate the fermentation process and concomitantly sugars are transformed into lactic acid, lowering the pH value. This final pH is similar to the meat isoelectric point lowering the water-holding capacity that would further help the dryness of the product. The advantage of this methodology lies in the higher shelf-life because of the intermediate water activity values and firmness and the product presents more flavor. The disadvantages are economical since more labour is required and there is a waste of processing laboratory space; the process is time-consuming, and the final product has a harsh salty flavor. The list of cured products derived from the application of this technique includes ham (Spanish serrano, Italian parma, French bayonne, American country-style, German westphalia), South American charqui meats, and Italian copa.

Wet curing

Historically this method originally involved the immersion of meat pieces into cold brine containing dissolved curing salts. Although the diffusion of these ingredients was quicker than rubbing dry salt into the meat, it was in fact a time-consuming technology. A technique to speed up curing is injecting pickle into the entire ham through artery pumping. The needle is usually inserted in the femoral artery using the same ingredient as for dry curing but the brine is dissolved in water to make a pickle. Thus, the need to speed up even more this curing technology brought about the multiple-needle stitch automatic injection that is currently used to prepare ham and bacon, followed by tumbling to accelerate brine diffusion and give uniformity of the curing mixture. To finish this process, these products are exposed to heat and/or smoke.

Finally, for the wet curing of minced meat, the curing mixture is added to the meat and fat to prepare the meat emulsion. In order to prevent excessive heat,

ice is included to cool down the process. This helps emulsion stabilization by avoiding protein denaturation and fat fusion point. Thus the product is stuffed in cellulose, plastic, or reconstituted collagen casings, followed by heat treatment, and can be smoked. This technique is applied to process frankfurters, Bologna sausages, and mortadella. However, the disadvantages are again the poor utilization of space and the products have a milder flavor in comparison to dry curing, although less labor is required.

Combined Dry/Wet Curing

Some cured products are processed by applying a combination of both techniques. This is the case for South American charqui meat processing. The conditions for charqui meat preparation are to incorporate the curing ingredients into the meat to achieve uniform distribution, and thus to accelerate curing and to stabilize the color. Figure 1 shows a flow diagram for charqui processing. Essentially it is prepared by using the whole sides of muscles, immersing them in concentrated brine of approximately 25 °B containing 3–4% sodium or potassium nitrite for hours, allowing the brine to diffuse through the meat. The curing procedure can be considerably shortened by injection of brine into the meat to achieve uniform distribution. In a more modern technique, multiple injections are applied automatically, and brine is injected simultaneously. Thereafter samples are submitted to dry salting on a concrete floor on which the meat pieces are stacked into piles separated from each other by layers of coarse sea salt (approximately 1 mm thick). After about 8 h, the meat is restacked and the uppermost meat pieces are repositioned at the bottom of the new piles. The maneuver is repeated every 24 h (three to five times), and, after washing to remove excess salt from the meat surface, samples are subjected to the drying stage directly in the sun on wooden rails. At night, meat pieces are collected and piled on the concrete floor and covered with a tarpaulin. Finally, samples are packed under vacuum in polyethylene bags of 1–5 kg and commercialized.

Functional properties of cured products

The changes promoted by curing processing in relation to raw materials are relevant and consequently color, flavor, texture, and nutritive values are described below.

Color formation

Fresh meat color is primarily determined by the oxidation state of iron and the radical attached to the heme group of the myoglobin pigment that also presents the globular protein portion, globin, in its structure. Myoglobin can be found in three forms: (1) desoxymyoglobin (Mb) or reduced myoglobin, which is a purple red color, with the iron state of Fe^{2+}, (2) oxymyoglobin (MbO_2), oxygenated myoglobin, which is a bright red color with Fe^{2+}; and (3) metmyoglobin (MMb), oxidized myoglobin, brown color with Fe^{3+}. During the curing process, because nitrite is a oxidizing agent for Mb, it converts Mb and MbO_2 to MMb. Thus, nitric oxide reacts with MMb to form nytrosylmetmyoglobin (brown with Fe^{3+}) that is reduced to nitric oxide myoglobin, nitrosylmyoglobin (MbNO), which is the desired bright pink color, although unstable, with Fe^{2+}. Finally, MbNO when heated is denatured to form nitrosylhemochrome (stable pink color). This is the typical cured-meat color with Fe^{2+}, as shown in Figure 2.

Flavor

Cured meat flavor is a complex issue and the chemical reactions responsible for its formation are not well understood. This is understandable since it is believed that there are nearly 1000 volatiles compounds in the fresh meat. The presence of sodium nitrite decreases not only rancidity but also the warmed-over flavor present in cured meat. The reaction of the heme group with nitric oxide giving a typical color is also connected to the cured meat flavor.

Texture

Texture is a subject which is seldom discussed with reference to cured meat. However, some insights have already been reported with whole meat such as charqui meats. Before consumption charqui meats are immersed in several changes of water in order to remove salt. Because of the meat buffering properties, about 3–4% is the final amount of remaining salt. These meat products are tougher than control raw samples. There is five to six times the increase in texture in charqui meat in comparison to raw material and two to three times in relation to desalted charqui, as measured by a texturemeter. A similar pattern is observed after cooking these samples. The water content plays a significant role in the texture of charqui. The moisture content determined in raw material is 76% and the measured shear force is 17 N; charqui contains 46.3% moisture and 96.5 N and desalted charqui 59.0% moisture and 46.5 N shear force. Cooked samples have a moisture value of 52.0, 31.7, and 44.8%, in control, charqui, and desalted charqui, respectively and shear force equivalent to 52, 300 and 86 N for control, charqui, and desalted charqui, respectively. There is an inverse relationship between the amount of water and

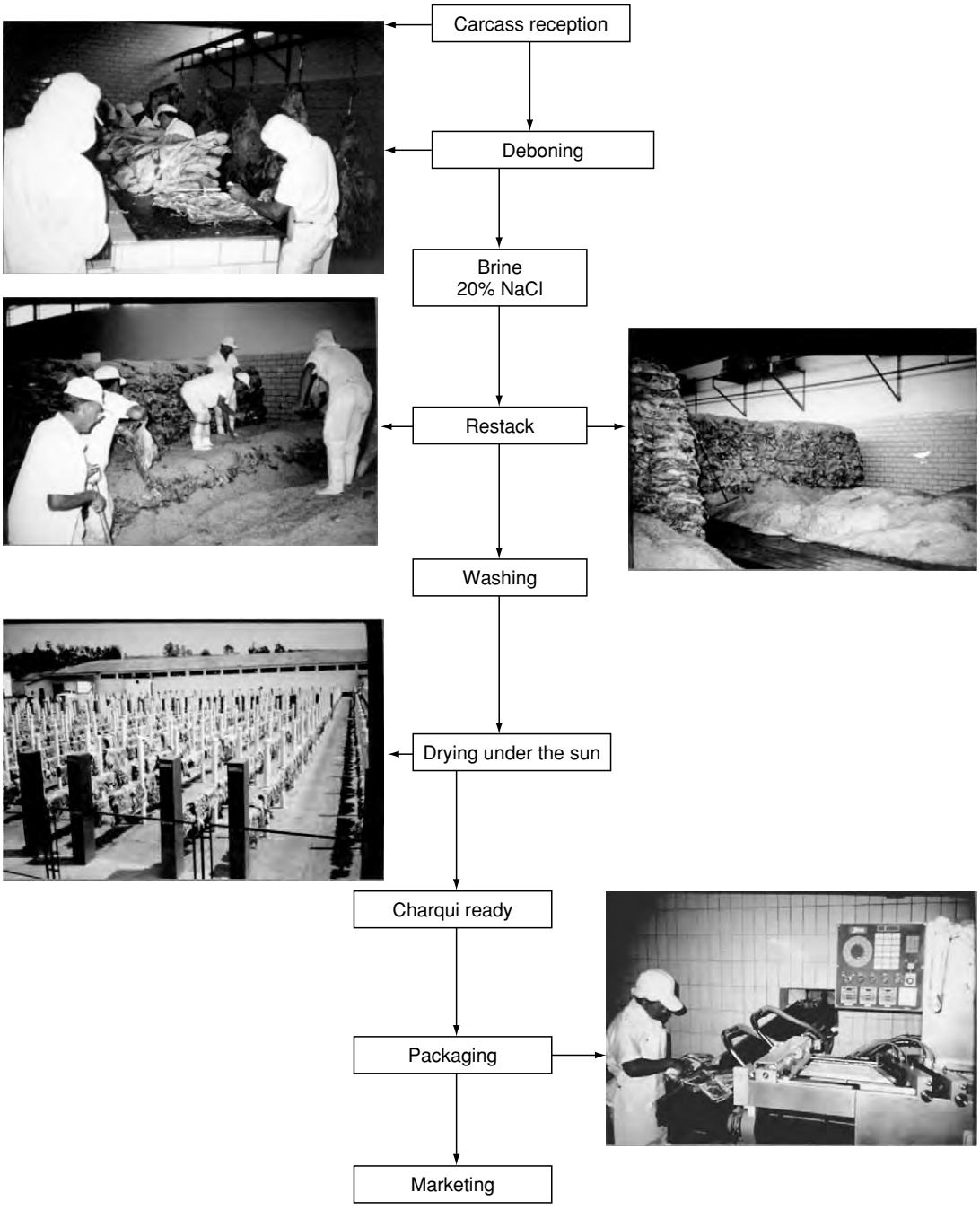

Figure 1 Flow diagram showing charqui meat processing with pictures of the steps of deboning, stacking, drying under the sun, and packaging. Photos courtesy of Industrias Allyson, Santana da Parnaíba, Brazil.

charqui texture, and the shear force increases exponentially with water loss.

Nutritive Value Evaluation

Cured meat products have the disadvantages of the presence of nitrosamine that is formed during curing. This compound is a potent carcinogenic component and has been detected in bacon after frying. On the other hand, because of the high biological value of protein, they present a high nutritive value. This is the case for intermediate-moisture meat such as charqui meats. Recent experiments with cooked samples have demonstrated high protein efficiency ratio of 2.07 and net protein utilization of 87.7, which is close to those of casein: 2.33 and 94.6, respectively.

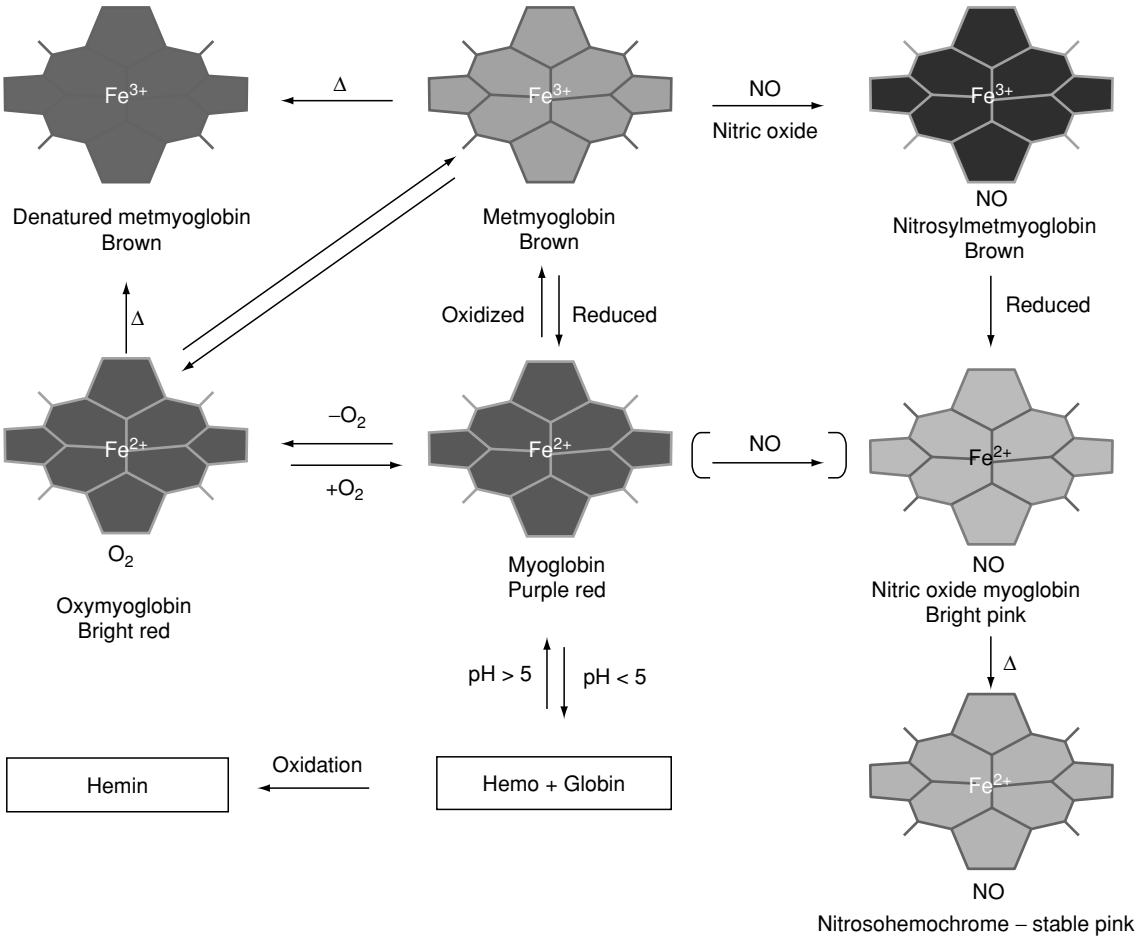

Figure 2 (see color plate 41) Changes in the pigment heme during curing and respective color and iron state.

Fermentation

Fermentation is an old technology applied to meat products that has a preservative activity and also gives rise to specific products with characteristic flavors. Initially, the type of microorganism was naturally selected; today there is an enormous technological advance with the application of starter cultures. In this way, a quicker and more uniform product can be obtained. In general, starter can be be used in two ways: maturation and surface. *Lactobacillus, Pediococcus, Micrococcus*, and *Staphylococcus* normally represent maturation starters. *Lactobacillus* and *Micrococcus* are characterized by the production of lactic acid which brings down the pH from 5.7–6.2 to as low as 5.3, with some processes dropping to pH 4.6: this is important for the reduction of nitrite to nitrous acid. The presence of *Micrococcus* and *Staphylococcus* is also important because they are capable of reducing nitrate to nitrite. In general, fermented meat products can be divided into two major groups: products using whole-meat pieces such as country ham products and charqui meats; and products using small chopped meat (the most representative quantitatively), as fermented sausage with various types of salami.

Microbiological Safety

It is known that curing processing inhibits the development of deteriorating bacteria. It follows the application of the so-called hurdle technology theory, that involves the application of sequential hurdles. Thus, nitrite would inhibit several kinds of bacteria, but Gram-positive bacteria, i.e., *Lactobacillus* and Micrococcaceae, are more resistant and play a relevant role in fermentation. Therefore nitrite has the important function of selecting microflora in particular in fermented sausages. The most important role, though, is to inhibit the development of *Clostridium botulinum*. Salt quantity and the resultant water activity inhibit the bacterial outgrowth, including the enterogenic *Staphylococcus aureus*.

Public Health Concern for Cured Meats

Nitrosamine as a carcinogenic agent is a worldwide concern, therefore its quantity is limited for curing. The amount of 120 p.p.m. is permitted to process bacon in USA, although 40–80 p.p.m. of residual nitrite level is sufficient to inhibit *C. botulinum* spore outgrowth. In Europe an estimated maximum level of 150 p.p.m. for sodium nitrite is permitted for cured meat. This quantity varies depending on the type of products and also on the processing technique. Sodium nitrate is not allowed in bacon and is only permitted to be used in dry-cured meat whilst nitrite is applied in the wet-curing processes. In South America, a level of 200 p.p.m. is allowed for cured meat products and 50 p.p.m. for cured charqui meats.

See also: **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; **Fermented Foods**: Fermented Meat Products; **Nitrosamines**; **Phosphorus**: Properties and Determination

Further Reading

Campbell-Platt G (1995) Fermented meats – a world perspective. In: Campbell-Platt G and Cook PE (eds) *Fermented Meats*, pp. 39–52. London: Chapman & Hall.

Cassens RG, Greaser ML, Ito T and Lee M (1979) Reactions of nitrite in meat. *Food Technology* 17: 46–62, 93.

Chang SF, Huang TC and Pearson AM (1996) Control of the dehydration process in production of intermediate-moisture meat products: a review. *Advance in Food Nutrition and Research* 39: 71–161.

Flores J and Toldrá F (1993) Curing. In: Macrae R, Robinson RK and Sadler MJ (eds) *Encyclopaedia of Food Science, Food Technology and Nutrition*, pp. 1277–1282. London: Academic Press.

Leistner L (1995) Stable and safe fermented sausages worldwide. In: Campbell-Platt G and Cook PE (eds) *Fermented Meats*, pp. 160–175. London: Chapman & Hall.

Sebranek JG and Cassens RG (1973) Nitrosamines: a review. *Journal of Milk and Food Technology* 36: 76–91.

CURRANTS AND GOOSEBERRIES

N Heiberg, The Norwegian Crop Research Institute, Leikanger, Norway
F Maage, Agricultural University of Norway, Aas, Norway

Background

The domestication of currants and gooseberries has taken place within the last 400 or 500 years. Horticulturally, these fruits are not major crops, but they are widely grown in northern European regions. The crop is mainly used in the production of juices, jams, and jellies. Blackcurrant has strong color and aromatic taste, and is an excellent source of ascorbic acid. Long before vitamins were known, people used blackcurrants for medical purposes, mostly as a hot drink against the common cold.

Global Distribution

The genus *Ribes* of the Saxifragaceae family, consists of about 150 species of currant and gooseberry, mainly distributed in the northern temperate regions of Europe and North America. Of the edible types, the main commercially grown species are blackcurrant (*Ribes nigrum* L.) red- and whitecurrants (*R. rubrum* L., *R. sativum* Syme, and *R. petraeum* Wulf.), and the gooseberry (*R. grossularia* L.).

Commercial Importance

The majority of currants and gooseberries are produced in Europe (**Table 1**) with Germany, Russia, and

Table 1 World production of *Ribes* fruit ($\times 10^6$ kg)

	1988	*1998*
Austria	29	19
France	9	10
Germany	173	136
Netherlands	1	2
Scandinavia	24	24
UK	21	10
Western Europe	263	208
Czechoslovakia	35	
Czech Republic		20
Hungary	18	12
Poland	165	175
Slovakia		4
Eastern Europe	218	211
USSR	100	
Russian Federation		204
Ukraine		18
New Zealand	2	4
World	583	650

From Food and Agriculture Organization (FAO) production year books.

Poland as the main producers. Blackcurrants account for more than half of the total world production; this species is dominant in the UK, Scandinavia, and New Zealand, and is gaining importance in other countries. About a quarter of *Ribes* production is gooseberry, but the production of this fruit is decreasing in most countries. Germany, the UK, Poland, Czech Republic, and Russia are the major gooseberry-producing countries. Countries with a high proportion of redcurrant compared with blackcurrants are Germany, Belgium, The Netherlands, and Austria. Currants and gooseberries are at present of little commercial significance in North America, as they are alternative hosts of white blister rust (*Cronartium ribicola* Fisch.), but there is increasing interest in *Ribes* fruits in the USA, and some legislative controls on their cultivation are changing.

Machine harvesting is prevalent in industrial production of blackcurrants. The machines may be used for redcurrants as well but work less satisfactorily for gooseberries.

Besides commercial production, currants and gooseberry are important home garden crops in many countries. They tolerate both low summer and winter temperatures and therefore may be grown far to the north, where few other fruits can be cultivated.

Blackcurrant Cultivars

Old cultivars such as Roodknop and Øjebyn are still grown in Europe, but the 'Ben' cultivars from the UK have been commercially successful in many countries and their hectarage is increasing. Ben Lomond has been dominant so far, but in most countries, there is a swing towards newer cultivars like Ben Alder and Ben Tirran. The main cultivar in New Zealand has been Magnus, but north European cultivars such as Ben Ard and Ben Rua are becoming more widespread there. In the expanding Polish hectarage, Titania, Roodknop, Ben Lomond, and Ben Nevis are popular besides the traditionally grown Øjebyn. Russia has a large range of cultivars adapted to the climatic conditions in the different regions. Breeders are working on blackcurrants in several countries, and new cultivars are released yearly.

Red- and Whitecurrant Cultivars

The major red cultivars grown in Europe are Jonkheer van Tets, Red Dutch, Rondom, and Stanza. Red Dutch is winterhardy and is widely grown in the most northern areas. For the fresh fruit and dessert markets, Red Lake and Jonkheer van Tets have been popular because of their good eating quality. However, more recent cultivars from Holland, Rovada, Rolan, Rosetta, and Rotet, and Junifer from France, are gaining in popularity. White Dutch and White Versailles are most widespread among the whitecurrant cultivars. The whitecurrants lack red pigments and are in fact a color form of redcurrants.

Table 2 Fruit characteristics of currants and gooseberry

Characteristics	*Blackcurrant*	*Redcurrant*	*Gooseberry*
Number of fruits per strig	5–10	6–14	1–3
Fruit weight (g)	0.7–1.4	0.4–0.9	1.5–15.0
Number of seeds per fruit	30–50	5–11	10–45
Weight per seed (mg)	1–2	6–8	4–6

From Maage F (unpublished).

Gooseberry Cultivars

In the nineteenth century, gooseberry was a popular fruit in home gardens in Europe, especially in the UK, and amateur breeders raised hundreds of cultivars. Selection was mainly for fruit size. However, the appearance of American gooseberry mildew (*Sphaerotheca morsüvae*) in 1905 soon drastically reduced the acreages of gooseberry, as the large-fruited types all proved more or less susceptible.

Mildew-resistant cultivars have been released, but so far, they have been inferior to the old types in terms of both fruit size and quality. One that is gaining popularity is Invicta from the UK. Among the old, nonresistant cultivars, the green-fruited cultivar Whitesmith and the red-fruited Whinham's Industry are still grown. New cultivars such as Greenfinch and Pax from the UK, and Rixanta, Rokula, Reflamba, and Rolanda from Germany have been widely tested.

Other *Ribes* Cultivars

Artificial hybridization between *R. nigrum*, *R. divaricatum*, and *R. grossularia* has given rise to the new species *Ribes* × *nidigrolaria*. The fruit quality is somewhere between blackcurrant and gooseberry. Josta, the first cultivar to be released from Germany, is planted mostly in home gardens.

Morphology and Anatomy of the Fruit

The fruits of currants and gooseberry are true berries with the seeds enclosed in a fleshy pericarp. The berries are born in clusters, with every single fruit adjoined to the main strig by a short stem. The fruits ripen in order along the strig, the fruit closest to the branch first and the terminal last. The fruit closest to the branch is largest, and the fruit in the end of the strig is smallest. Gooseberry fruits develop singly or in small clusters with two or three fruits, the distal fruit often being smaller than the others. The fruits on the plant are harvested at the same time, often mechanically. At harvest, the stem follow the fruit, and in the

Table 3 Chemical composition (per kilogram of fresh fruit) of currants and gooseberry

Nutrient	*Blackcurrant*	*Redcurrant*	*Gooseberry*
Water (g)	815	845	880
Carbohydrates (g)	128	96	78
Protein(g)	13	12	8
Fat (g)	2	2	2
Fiber (g)	43	39	22
Pectin (g)	8	7	5
Total sugar (g)	85	57	56
Energy (kJ)	2600	2050	1650
Energy (kcal)	621	490	395
Sodium (mg)	17	14	15
Potassium (mg)	3100	2300	1600
Manganese (mg)	3	2	10
Magnesium (mg)	190	140	115
Calcium (mg)	550	380	250
Iron (mg)	13	9	6
Phosphorus (mg)	480	330	250
Copper (mg)	1	1	2
Zinc (mg)	3	2	2
Ash (mg)	7200	6400	4800
Ascorbic acid (mg)	1600	650	440
Thiamin (mg)	0.5	0.4	0.4
Riboflavin (mg)	0.4	0.3	0.2
Pyridoxine (mg)	1.2	0.5	
Nictinic acid (mg)	2.8	2.5	2.5
Pantothenic acid (mg)	4.0	6.0	2.3
β-Carotene (mg)	1.2	0.6	1.5
Vitamin B_6	0.8	0.5	0.2
Folic acid (mg)	0.2	0.1	0.2

From Green A (1971) Soft fruits. In: Hulme AC (ed.) *Biochemistry of Fruits and their Products*, vol. 2, pp. 375–410. London: Academic Press; Herrmann K (1997a) Inhaltsstoffe der Johannisbeeren. *Die industrielle Obst- und Gemüseverwertung* 82(1): 14–20; Herrmann K (1997b) Inhaltsstoffe der Stachelbeeren. *Die industrielle Obst- und Gemüseverwertung* 82(2): 34–36; Kuusi T (1970) Uber die chemische Zusammensetzung und Kennzahlen einiger finnischer einheimischer Beeren. *Flüssiges Obst*. 37: 188–190, 253–263; Souci SW and Bosch H (1978) *Lebensmittel-Tabellen für die Nährwertberechnung*. Stuttgart: Wissenschaftliche Verlagsgesellschaft.

distal end, the rest of the flower is found. Table 2 lists the range of variation concerning fruit characteristics. Gooseberry has the largest fruits with a great variation between cultivars, redcurrant being the smallest. Redcurrants have the lowest number of seeds and the largest seeds of the berries. The skin of some gooseberry cultivars is hairy, whereas the skin of currants is always hairless.

The pigments of the blackcurrant fruit are located in the skin, the flesh remaining green. The pigment of redcurrants and gooseberries, however, is also present to some extent in the fruit flesh. Unlike currants, the cultivars of gooseberry cover the whole range of fruit colors, from dark to light red, through various shades of green, yellow, and almost white. Blackcurrants have a dark purple color, whereas redcurrants are pure red. The white currants are lacking anthocyanins and have a light yellow–greenish color.

Chemical and Nutritional Composition

The energy in fruits of currants and gooseberries comes mainly from carbohydrates (Table 3). Only a small part comes from protein and fat. Fructose and glucose are the main sugars, with roughly equal amounts of each (Table 4). Sucrose is present, but to a lesser extent. Gooseberries contain small amounts of sorbitol, whereas only traces of sorbitol are present in currants. A characteristic feature of the fruits is the high content of acids. Citric acid is predominant in currants, whereas citric acid and malic acid are present in almost equal quantities in gooseberries.

Ascorbic acid (vitamin C) has probably received more attention than any other constituent of blackcurrants. Less than 50 g of the fruit meets the

Table 4 Some organic characteristics in fruit juice (per kilogram of fresh juice) of *Ribes*: mean values and variation

Characteristic	*Blackcurrant*	*Redcurrant*	*Gooseberry*
Soluble solids (g)	160 (120–200)	105 (80–145)	128 (80–170)
Glucose (g)	35 (30–45)	26 (7–35)	38 (16–52)
Fructose (g)	38 (34–46)	26 (18–31)	42 (21–52)
Sucrose (g)	13 (8–20)	4 (1–6)	12 (2–28)
Total sugar (g)	86 (74–105)	57 (32–68)	92 (48–112)
Titratable acids[a] (g)	39 (22–56)	24 (17–36)	23 (12–32)
Citric acid (g)	42 (32–55)	25 (15–38)	12 (8–14)
Malic acid (g)	6 (3–10)	5 (1–18)	14 (6–21)
pH	2.87 (2.68–3.18)	3.04 (2.86–3.20)	3.00 (2.79–3.06)

[a]Acid content measured using equivalent weight of citric acid.

From Dietrich H, Krüger E, Patz C-D and Schöpplein E (2000) Characterisierung von Schwartzen Johannisbeersorten im Hinblick auf die Saft- und Nektarherstellung. *Obst-, Gemüse- und Kartoffelverarbeitung* 85(2): 86–98; Heiberg N, Måge F and Haffner K (1992) Chemical composition of ten blackcurrant (*Ribes nigrum* L.) cultivars. *Acta Agriculturae Scandinavica, Section B, Soil and Plant Sci*. 42: 251–254; Herrmann K (1997) Inhaltsstoffe der Johannisbeeren. *Die industrielle Obst- und Gemüseverwertung* 82(1): 14–20; Herrmann K (1997) Inhaltsstoffe der Stachelbeeren. *Die industrielle Obst- und Gemüseverwertung* 82(2): 34–36; Måge F (1993) Vegetative, generative and quality characteristics of four blackcurrant (*Ribes nigrum* L.) cultivars. *Norwegian Journal of Agricultural Sciences* 7: 327–332.

recommended daily requirement of ascorbic acid. However, considerable variation exists between cultivars. Other vitamins are also present in the fruits, and the content of minerals is also worth mentioning, especially the high content of potassium.

Attention is now being focused on naturally occurring antioxidants. Besides ascorbic acids, currants, especially blackcurrants, contain many substances that have an antioxidant activity. Blackcurrants possess a high flavonoid content compared with other soft fruits. The most important group of the flavonoids is the anthocyanins. The anthocyanins are the dominant pigments, and blackcurrants contain from 1250 to 2000 mg per kilogram of fresh weight. The anthocyanins of blackcurrant are mainly cyanidin and delphinidin 3-glucosides and 3-rutinosides. In redcurrant, six different cyanidin glycosides have been isolated. Most red European gooseberry cultivars contain only cyanidin 3-glucoside and 3-rutoniside. Another important group of flavonoids is the flavonols, of which glycosides of kaemferol, quercetin and myricitin are present in currants and gooseberries.

The characteristic aromatic compounds of blackcurrant are present in the whole plant. Several studies have been carried out to identify the aromatic components of blackcurrants, and a wide range of volatile components have been identified, but the exact chemical nature of the specific aroma in blackcurrants is unknown.

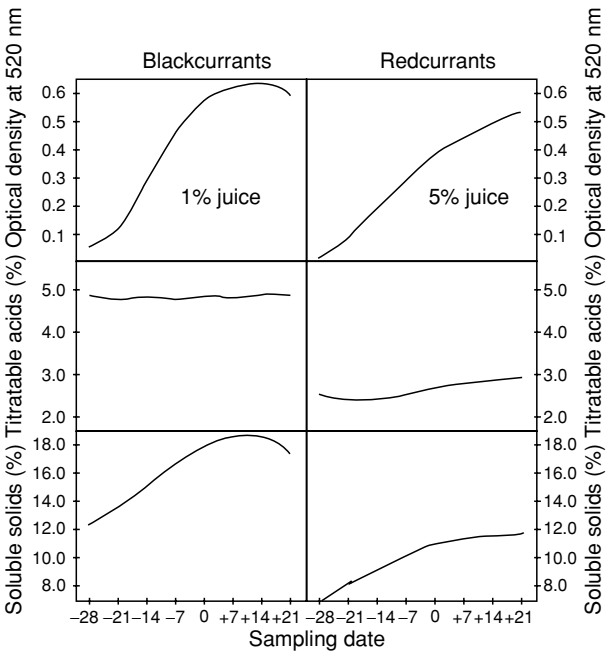

Figure 1 Changes in fruit quality components in black- and redcurrants from 28 days before to 21 days after normal harvest time (titratable acid expressed in terms of citric acid). From Måge F and Heiberg N (unpublished).

The degree of ripening has a great influence on the quality characteristics. The color intensity, content of dry matter, soluble solids, and sugars increase with ripening, whereas the viscosity and content of ascorbic acids decrease. The content of titratable acids reaches its peak about 2 weeks before harvest and changes little during ripening. The development of some quality factors is shown in Figure 1.

The seeds of currants and gooseberry contain about 20% fat, of which 5–20% is γ-linolenic acid, with the highest content occurring in blackcurrant.

Handling and Storage

The strong flavor and the high acidity of currants, especially blackcurrants, make them less attractive for fresh consumption. However, in Europe, there is a market for fresh fruits, especially red currants, which are used for desserts and decoration purposes. With a combination of out-of-season production and storage under controlled-atmosphere conditions, growers are able to market fresh fruits most of the year. Gooseberries have a milder flavor, and a considerable part of the crop is used for dessert purposes. Gooseberries are sold both as unripe, green fruits and as ripe fruits.

For fresh consumption, the *Ribes* fruits are still picked by hand, because harvesting machines reduce the postharvest quality. The currants are picked on strigs, whereas gooseberries are sold as single fruits. The berries should be picked in a dry condition, as they spoil quickly if gathered and packed when wet.

Compared with many other soft fruits, currants and gooseberries keep reasonably well, but without cooling, the fruits decay rapidly after harvest. The weight loss may reach 2–3% within 24 h without cooling. The fruits are usually cooled to 0–5 °C immediately after harvest. Forced-air cooling is recommended. At low temperatures, the blackcurrants have a shelf-life of 2–6 days and redcurrants 2–4 weeks, dependent on maturity stage. Picking the berries slightly unripe increases the shelf-life. The fruits should be kept cooled during transportation and marketing. To increase the shelf-life, the fruits must be stored in a controlled atmosphere, in which it is possible to store blackcurrants for 3–4 weeks, gooseberries for 6–8 weeks, and redcurrants for 8–14 weeks.

Industrial Uses

The most important products of blackcurrant are different types of juices and syrups. Most of the UK crop is used to produce the proprietary brand name drink 'Ribena.' Owing to the special aromatic taste, and the high content of ascorbic acid, blackcurrant products are popular in northern Europe and are

commonly regarded as healthy preparations. The fruits are also used for jams and jellies, and are suitable for flavoring other foods, such as yogurt, icecream, and other diary products, and in bakery products such as pies.

Red currants are mainly grown for juice and jelly processing, often mixed with other fruits with lower acidity, and are used as a component in mixed juice, where they add a stronger character to the product.

Gooseberry is mainly used for jam and canned products and, to some extent, for juice production. All three fruits are used for berry wine production, alone or in combinations with other raw materials. Blackcurrants are also used for sweet liqueurs, such as 'Crème de cassis' in France.

Blackcurrant is rich in polyphenolic color products, and the γ-linolenic fatty acid produced from blackcurrant seeds is used as a health preparation. Manufacture of the oil is based on press cake from blackcurrant juice processing. In blackcurrant, other parts of the plants are also of commercial interest. The leaves may be used for tea, and bud extracts have been used as a flavor component in other foods and as an ingredient in some fragrances.

See also: **Ascorbic Acid**: Properties and Determination; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Tannins and Polyphenols**

Further Reading

Agar T, Streif J and Bangerth F (1991) Changes in some quality characteristics of red and blackcurrants stored under CA and high CO_2 conditions. *Gartenbauwissenschaft* 56(4): 141–148.

Batzer U and Helm HU (1999) Lagerung von Beerenobst. *Erwerbsobstbau* 41: 51–55.

Brennan R (1996) Currants and gooseberries. In: Janick J and Moore JN (eds) *Fruit Breeding Vine and Small Fruit*, vol. 2, pp. 191–297. New York: John Wiley.

Dietrich H, Krüger E, Patz C-D and Schöpplein E (2000) Characterisierung von Schwartzen Johannisbeersorten im Hinblick auf die Saft- und Nektarherstellung. *Obst-, Gemüse- und Kartoffelverarbeitung* 85(2): 86–98.

Green A (1971) Soft fruits. In: Hulme AC (ed.) *Biochemistry of Fruits and their Products*, vol. 2, pp. 375–410. London: Academic Press.

Heiberg N (1986) Quality components of blackcurrant (*Ribes nigrum* L.) fruits as related to degree of ripeness and cultivar. *Meldinger fra Norges landbrukshøgskole* 65(24): 25

Heiberg N and Måge F (1991) Evaluation of ten blackcurrant cultivars. *Norwegian Journal of Agricultural Sciences* 5: 277–282.

Heiberg N, Måge F and Haffner K (1992) Chemical composition of ten blackcurrant (*Ribes nigrum* L.) cultivars. *Acta Agriculturae Scandinavica, Section B, Soil and Plant Science* 42: 251–254.

Herrmann K (1997a) Inhaltsstoffe der Johannisbeeren. *Die industrielle Obst- und Gemüseverwertung* 82(1): 14–20.

Herrmann K (1997b) Inhaltsstoffe der Stachelbeeren. *Die industrielle Obst- und Gemüseverwertung* 82(2): 34–36.

Kuusi T (1970) Uber die chemische Zusammensetzung und Kennzahlen einiger finnischer einheimischer Beeren. *Flüssiges Obst* 37: 188–190, 253–263.

Måge F (1993) Vegetative, generative and quality characteristics of four blackcurrant (*Ribes nigrum* L.) cultivars. *Norwegian Journal of Agricultural Sciences* 7: 327–332.

CYCLAMATES

J F Lawrence, Sir F.G. Banting Research Centre, Ottawa, Ontario, Canada

Introduction

'Cyclamate' is the term given to the artificial sweetener cyclamic acid (cyclohexylsulfamic acid) and its calcium or sodium salts (see **Figure 1**). Cyclamate was discovered in 1937 at the University of Illinois, following the accidental contamination of a cigarette with a derivative of cyclohexylamine. In 1940, DuPont obtained a patent for its production, and in 1950, it was available to consumers. Consumption of cyclamates increased steadily from that time up to about 1969, when it was banned in the USA and other countries due to safety concerns related to its potential carcinogenicity. However, recent studies have contradicted the earlier work, and cyclamate continues to be used in many countries (see section on 'Metabolism and Safety'). Cyclamate is not metabolized by the human body, so it contributes no energy to the diet and is considered a nonnutritive sweetener.

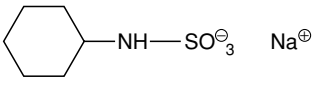

Figure 1 Structure of cyclamate (sodium salt).

Sweetness

In comparison with other intense sweeteners, cyclamate is perhaps the least sweet, being only 30–80 times as sweet as sucrose in actual food uses, depending upon concentration, pH, flavoring agents and other ingredients that may constitute a part of the food product. Aspartame, by comparison, is about 200 times sweeter than sucrose, whereas saccharin is about 300 times sweeter. In high concentrations, cyclamate has an unpleasant aftertaste. However, at low concentrations, it has some bitterness-masking ability that makes it attractive for use in pharmaceutical products. During the 1960s, cyclamate–saccharin combinations became very popular. At ratios of 10:1 (cyclamate:saccharin) on a weight basis, the combination resulted in a pleasant sweetness, minimizing the somewhat disagreeable aftertastes of both sweeteners individually. The synergistic effect of saccharin in combination with cyclamate increased the sweetening power of the mixture compared with the individual sweeteners.

Production, Physical, and Chemical Properties

Cyclamate and its calcium or sodium salts are white crystalline powders with intensely sweet tastes. Cyclamate has a melting point of about 169–170 °C. It is soluble in water up to a concentration of about 1 g in 7.5 ml, whereas the calcium and sodium salts are slightly more soluble, to the extent of about 1 g in 4 ml of water. Cyclamate is rather acidic (the pH of a 10% aqueous solution being 0.8–1.6), whereas similar solutions of the calcium and sodium salts are neutral (pH 5.5–7.5). Cyclamate is relatively heat-stable, microbiologically inert, and nonhygroscopic.

Cyclamate is synthesized by sulfonation of cyclohexylamine with chlorosulfonic acid in chloroform followed by treatment with barium hydroxide and sulfuric acid. Variations of this reaction scheme followed by treatment with sodium or calcium hydroxides yield the salts of cyclamate. Commercial production of cyclamate reached a peak in the USA in 1968, when an estimated 7400 tonnes were

Table 1 Cyclamate use levels in diet food products

Diet food	*Cyclamate level*
Table top sweeteners	25 mg g^{-1}
Milk beverages	0.8 mg ml^{-1}
Beverages and beverage bases	4 mg ml^{-1} prepared drink
Gelatin, puddings, filling	27 mg ml^{-1}
Salad dressings	1.6 mg ml^{-1}
Jellies, jams, preserves	30 mg ml^{-1}
Sweet sauces, toppings, syrups	30 mg ml^{-1}
Chewing gum	20 mg per stick
Hard confectionery	5 mg g^{-1}
Baked goods, baking mixes	2.6 mg g^{-1}

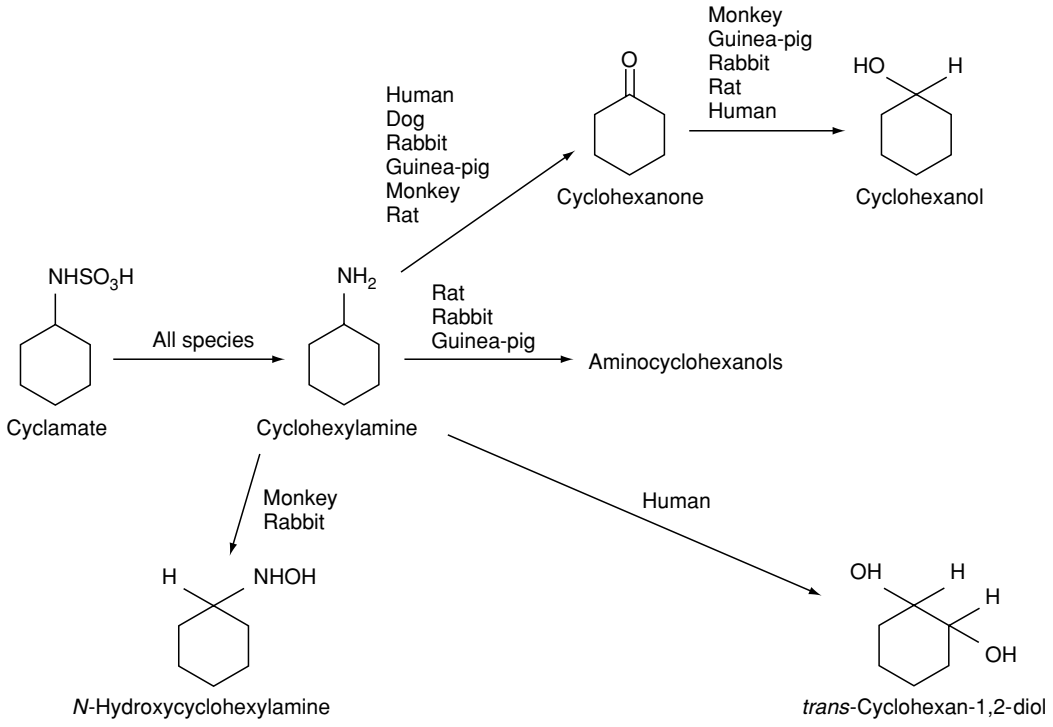

Figure 2 Schematic of the metabolic routes for cyclamate in several mammalian species.

produced. Commercial production dropped dramatically after bans on its use in foods came into effect in many countries around 1970.

Food Uses

Cyclamate is not permitted as a food additive in a number of developed countries, including Canada and the USA. Prior to 1970, cyclamate was used in a variety of diet food products. In 1969, 70% of the usage of cyclamate was in beverages, whereas 15% was in dietetic foods and 15% in table top products. Table 1 lists some of its past applications and approximate concentration levels. Like saccharin, cyclamate is rather stable to heat and moisture, and thus it was suited to sweetening a wide variety of food products. It found wide use as a sweetener in combination with saccharin, as mentioned above.

Metabolism and Safety

Figure 2 shows a schematic of the overall metabolism of cyclamate in a number of mammalian species. In humans, orally administered cyclamate is rapidly excreted via both the urine and feces. More than 98% of the dose is excreted within 1–2 days. Studies with rats and dogs have shown that small amounts of cyclamate can be distributed to all tissues (except the brain), although accumulation is negligible.

Conversion of cyclamate to cyclohexylamine in human subjects was first reported by Japanese researchers in 1966. Very small amounts of cyclohexylamine (0.7% of the administered dose) were found in the urine. This finding stimulated a large number of feeding studies in humans and animals. It appears that Enterococci organisms in the intestine convert cyclamate to cyclohexylamine. Cyclohexylamine may be converted to other metabolites as shown in Figure 2. The metabolism of cyclamate to cyclohexylamine and other products is dependent on the individual concerned. In one study involving 1000 human subjects, 10–30% of the subjects converted ingested cyclamate to cyclohexylamine mostly in the range of < 0.1–8% of the administered dose. However, some individuals converted up to 60% of the cyclamate.

Cyclamate was banned from use in many countries because the results of some rat-feeding studies showed the substance to cause bladder tumors. Many scientists questioned the results of the original studies, and this led to many follow-up studies in a variety of animal species. Of note are the results of a 22-year feeding study in monkeys carried out by the National Institutes of Health in the USA that showed that neither cyclamate (nor saccharin) demonstrated any evidence of carcinogenic effects. The prevailing opinion today is that neither cyclamate nor cyclohexylamine is likely to be carcinogenic to humans, especially at the levels recommended for diet foods (see Table 1). The main reason now that cyclamate is not permitted is that during the follow-up feeding studies to evaluate its carcinogenicity, it was observed that cyclohexylamine, the main metabolite of cyclamate, caused irreversible testicular atrophy in rats. Further research is required to study this effect in detail. The Food and Agriculture Organization/World Health Organization has determined an acceptable daily intake (ADI) value (the maximum amount that could be consumed daily for a lifetime without appreciable risk) for cyclamate to be 11 mg per kilogram of body weight. However, in the UK, a temporary maximum ADI of only 1.5 mg per kilogram of body weight has been established until the results of further research are known.

Analysis

Methods for the determination of cyclamate in foods are not as simple or as straightforward as for those sweeteners such as acesulfam-K, aspartame, and saccharin, which strongly absorb ultraviolet light. The latter can be measured by direct means (usually high-performance liquid chromatography) without recourse to chemical derivatization.

The most common analytical methods for cyclamate involve chemical conversion to cyclohexylamine followed by determination of the amine by gas chromatography with flame ionization detection. Several methods employing high-performance liquid chromatography have been evaluated, which offer potential for cyclamate determination without the need to convert to cyclohexylamine or other product. These employ direct conductivity detection, indirect ultraviolet absorption detection, or postcolumn ion-pair extraction detection.

See also: **Sweeteners**: Intensive; Others

Further Reading

Franta R and Beck B (1986) Alternatives to cane and beet sugar. *Food Technology* 40: 116–128.

IARC (1980) *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans, Some Non-nutritive Sweetening Agents*, vol. 22: Lyon: International Agency for Research on Cancer.

Kretchmer N and Hollenbeck CB (1991) *Sugars and Sweeteners*. Boca Raton, FL: CRC Press.

O'Brien-Nabors L and Gelardi RC (1991) *Alternative Sweeteners*. New York: Marcel Dekker.

CYSTIC FIBROSIS

R W Shepherd, Washington University School of Medicine, St Louis, MO, USA

Introduction

Cystic fibrosis (CF), a common autosomal recessive hereditary disease in Caucasians (1/2200 live births) is characterized by recurrent pulmonary infections, pancreatic insufficiency with maldigestion, malabsorption, and abnormal levels of sweat electrolytes. The abnormal gene locus is localized on the long arm of chromosome 7, most commonly at delta F 508 (in approximately 70% of cases). Biochemical studies have suggested that the defect occurs in the regulation of ion (mainly chloride) transport across epithelial cell membranes, resulting in abnormal exocrine function. This in turn results in a wide range of clinical effects (Figure 1). Recent advances in therapy have greatly improved life expectancy, such that most patients now survive into adulthood. This has resulted in an increasing population of patients requiring long-term medical care best managed through multidisciplinary clinics.

The adverse effects of CF on nutrition and growth have long been recognized and, with improving life span, the assurance of normal nutrition and growth has become an increasingly important aspect of management. Nutritional growth retardation continues to affect a substantial number of patients and is a major factor adversely affecting survival. Optimal nutritional management is, therefore, crucial in terms of growth, quality of life, and perhaps long-term outcome. This review considers the nature and causes of nutritional growth retardation in CF, explores the consequences and range of nutritional deficiencies documented in this disease, and describes current best practice with respect to nutritional therapy.

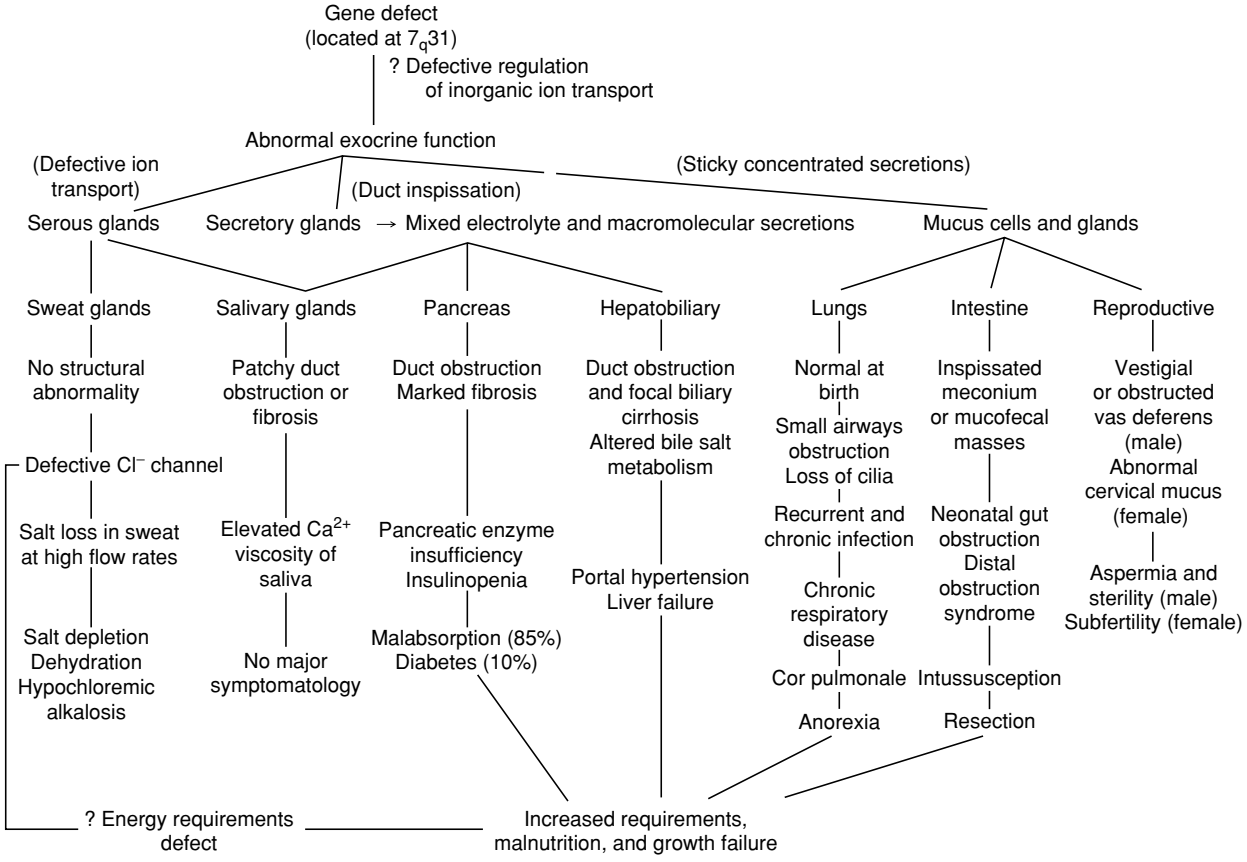

Figure 1 Pathogenesis of cystic fibrosis. Reproduced from Cystic Fibrosis. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

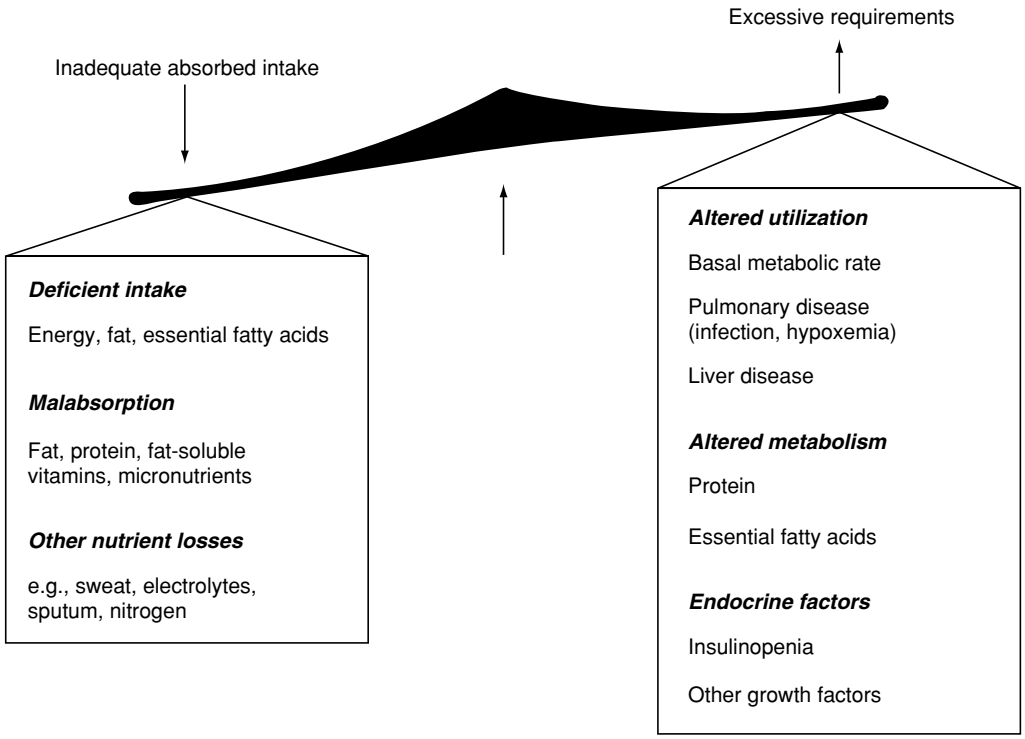

Figure 2 Factors affecting energy balance in cystic fibrosis. Reproduced from Cystic Fibrosis. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

The Nature of Malnutrition in CF

A wide range of nutritional deficits causing many deleterious effects have been described in CF. The heights and weights of many CF patients are markedly skewed towards the lower centile bands, particularly with increasing age. Body composition studies have suggested that underweight CF children have a deficit in body cell mass and body fat, with a reduction in mass, muscle mass, and an excess in extracellular water compared to controls. Whole-body protein turnover studies have suggested that many malnourished patients with severe lung disease are catabolic with a reduction in whole-body protein turnover synthesis occurring during pulmonary exacerbations. The deficit of the body cell mass observed in CF from total body potassium counting can be present from the first few weeks of life, as judged by studies of newborn CF infants diagnosed by neonatal screening. Prior to neonatal screening, which has allowed the earlier introduction of preventive therapy, overt hypoproteinemia was a common presenting symptom. Specific deficiencies of essential fatty acids, fat-soluble vitamins, some water-soluble vitamins, and specific micronutrients, including zinc, iron, and selenium, are well described. Another important consideration, particularly in tropical climates, is continuing specific loss of sodium chloride resulting from the sweat gland defect. Some patients develop CF-related diabetes or CF-related hepatobiliary disease, which can both further contribute to nutritional failure.

The range of nutritional abnormalities occurring in CF seems likely to be caused by a combination of inadequate absorbed intake, nutrient losses, and increased requirements (Figure 2). Although many textbooks describe voracious appetites in CF children, in reality, objective measurements often show an inadequate energy supply compared to recommended intakes, and such patients with apparent excessive appetites may often be attempting to compensate inadequately for requirements and excessive losses. A poor appetite may be due to malnutrition *per se*, but also appears to occur where there is poor pulmonary function and during active pulmonary infection. This latter problem may have a cumulative effect on nutritional status over years as lung disease progresses. (*See* **Energy:** Energy Expenditure and Energy Balance.)

Nutrient losses occur mainly from malabsorption, but in some cases excessive nitrogen loss in sputum, and excessive sodium losses from the sweat gland abnormality (Figure 1) are also important. Fat and protein malabsorption are present in approximately 85% of patients because of pancreatic exocrine deficiency. The pancreatic abnormality involves both water-electrolyte (mainly bicarbonate) secretion as well as lipase, trypsin, and amylase deficiency.

Reduced bicarbonate secretion is universal, but a functional level of digestive enzyme capacity is retained in 10–15% of patients termed pancreatic-sufficient. These individuals may develop recurring pancreatitis and some may eventually become deficient. In untreated cases where enzyme outputs are less than 10% of normal, steatorrhea, azotorrhea, starch intolerance, and malabsorption of fat-soluble vitamins, some minerals, and vitamin B_{12} occur.

There has been a growing interest in the nature of energy and protein metabolism in CF since the recognition of an apparent maladaptation to undernutrition, and the suggestion that there may be an energy-requiring basic cellular defect in this disease. There is good evidence that resting energy expenditure is elevated in CF irrespective of age and gender but is genotype-specific. The energy cost of activity is increased in those with more severe lung disease. These factors play an incremental role in elevating total energy expenditure and the high energy requirement observed in CF patients (up to 150% recommended dietary intake or RDI).

The Consequences of Malnutrition in CF

The major effects include nutritional growth retardation, delayed puberty, adverse effects on pulmonary disease, particularly lung growth and, ultimately, a poorer prognosis. Patients dying of CF have marked nutritional failure in the 1–2 years prior to their death. Malnutrition *per se* can compromise absorptive and immune function, and all of the specific deficiencies mentioned above can have a wide range of secondary adverse effects.

A close relationship exists between malnutrition and pulmonary disease in CF. Pulmonary disease apparently adversely affects energy requirements and whole-body protein metabolism. As previously mentioned, CF patients who have chronic but stable pulmonary disease tend to have excessive protein catabolism and elevated total energy expenditure, while those malnourished patients studied during acute pulmonary exacerbations have a significantly reduced level of whole-body protein turnover. Thus both acute and chronic pulmonary disease appear to have a major effect on energy utilization. There is also evidence that malabsorption and malnutrition *per se* adversely affect the course of pulmonary disease. An improved respiratory prognosis has been observed in those patients with pancreatic sufficiency, and a number of specific effects of malnutrition on the respiratory system have been documented. Malnutrition can affect the central respiratory drive mechanism, the respiratory muscles, and the growth of the lungs. Long-term studies of nutritional rehabilitation of malnourished patients have suggested that improved nutrition may favorably affect the course of pulmonary disease.

Nutritional Management in CF

Treatment of CF patients has become increasingly specialized and is most satisfactorily performed at major referral clinics where a comprehensive and intensive management program is available. It would seem likely that early diagnosis by neonatal screening and the institution of an aggressive therapy program, before irreversible lung damage and chronic nutritional deficits have occurred, may prolong survival. Studies of neonatal screening programs to date have indicated lower morbidity and improved nutrition in those children diagnosed by neonatal screening compared with a clinically diagnosed group, and one study from Denmark has indicated that there may also be a lower mortality. The adverse effects of chronic pulmonary disease on nutritional status and, conversely, the adverse effects of malnutrition on pulmonary status, particularly lung growth, require an early aggressive approach to both aspects of the disease. Pulmonary therapy is beyond the scope of this discussion, but its importance in the prevention of malnutrition in CF cannot be overstated. Optimal nutritional therapy involves the restoration of energy balance by regular nutritional surveillance, and the maintenance of an adequate absorbed protein-energy intake with the prevention and management of specific deficiencies (**Table 1**). Certain phases of the disease, such as those occurring in infancy, during and after pulmonary exacerbations, and during puberty, may be especially important for

Table 1 Nutrient requirements in cystic fibrosis

Nutrient	*Requirement*
Energy	120–150% RDA
Protein	120% RDA
Essential fatty acids	5% of total energy (kJ)
Vitamin A	5000–10 000 IU per day (water-miscible)
Vitamin E	100–300 IU per day (water-miscible)
Vitamin K	5 mg twice weekly
Vitamin D	800 IU per day in temperate zones
B vitamins	200% RDA
Vitamin C	200% RDA
Iron and zinc	120% RDA
Pancreatic supplements (as pH-sensitive microspheres)	Infants: 2000 units of lipase per 120 ml of formula Older children: 500–4000 units of lipase per g dietary fat Maximum dose 10 000 units kg^{-1} day^{-1}

RDA, recommended dietary allowance.

optimal nutritional therapy. Overtly malnourished patients can benefit from long-term enteral nutrition delivered via either nocturnal nasogastric or gastrostomy feeding.

Surveillance

An important part of the routine management program is the regular evaluation of protein-energy balance. This involves assessment of intake, absorbed energy, possible nutrient losses, and the recording of anthropometric data. An example of a computerized surveillance program at the Royal Children's Hospital, Brisbane, is given in Figure 3, showing the benefits of longitudinal evaluation. Certain noninvasive research techniques are deserving of wider application as markers of changing nutritional status in CF. Although many of these patients are within the normal centile bands for weight, measurements of total body potassium (which provide a much more sensitive indicator of growth of the body cell mass) suggest many patients have cell mass deficits. Body impedance, which can be measured by a simple bedside technique, correlates well with body potassium in CF and may serve as a reliable clinical tool for the surveillance of fat-free mass. The advent of measures of total energy expenditure in free-living humans should also help to establish an evaluation of true nutritional requirements in this disease.

Dietary Therapy

Dietary counseling, and dietary and pancreatic enzyme supplements remain mainstays in the preventive management of nutritional problems in cystic fibrosis. To achieve an adequate absorbed protein-energy intake (Figure 2), daily intakes >130% of RDI are usually necessary to compensate for stool losses and increased requirements. In general, it is usually possible to achieve normal growth and nutrition in CF children with dietary counseling, pancreatic enzymes, and high-energy supplements. Although in the past it was standard practice in many CF clinics to prescribe low-fat, high-carbohydrate diets, low-fat diets offer no benefit to CF patients, impair total energy intake, and contribute to essential fatty acid deficiency. Indeed, mortality was higher in clinics prescribing low-fat diets, favoring the prescription of a moderate- to high-fat diet. Since fiber has been shown to inhibit pancreatic enzymes *in vivo* and *in vitro*, fiber-enriched diets should be avoided. Patient tolerance of normal fat intakes of 40% of total energy is satisfactory, provided that appropriate enzyme preparations are used. Palatable and

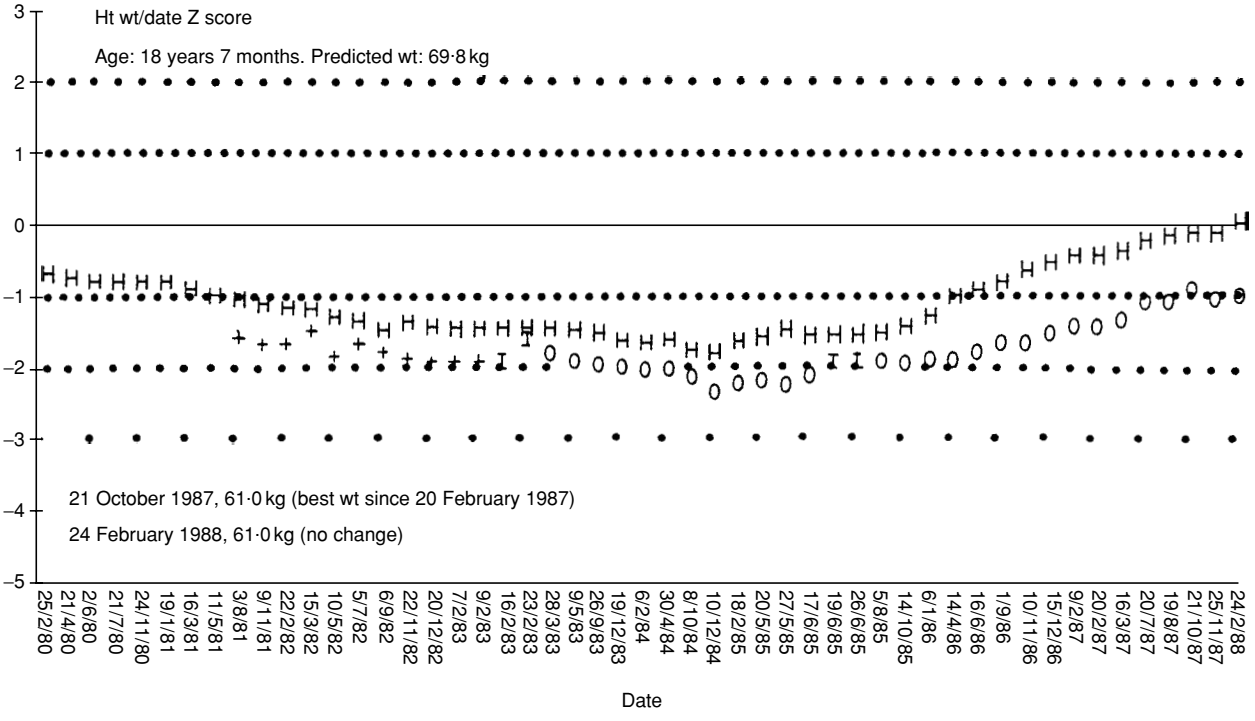

Figure 3 Serial height and weight standard deviation scores from the computerized clinic data records of an adolescent with cystic fibrosis from the age of 10–18 years. Note the gradual decline in both height (H) and weight (I, weight as an inpatient; 0 or +, weight as an outpatient), until about the age of 14 years, when long-term enteral nutritional supplements were instituted by nocturnal intragastric feeding. By the age of 18 there was a gradual improvement in these parameters. Reproduced from Cystic Fibrosis. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

nutritious high-protein energy foods are encouraged and special milk drinks with added energy sources, such as chocolate-flavored protein-enriched powders, glucose polymers, and other foods, represent a good way of increasing total energy intake.

A full range of vitamin and mineral supplements is necessary to prevent vitamin and mineral deficiency. In particular, fat-soluble vitamins presented in a water-miscible form are required in all patients with steatorrhea, administered in a dosage twice the RDI. As mentioned previously, where oral intake with supplements fails to achieve satisfactory height and weight gain, or if the patient is overtly malnourished, enteral supplements can be used. A metaanalysis of 18 studies of enteral supplementation have demonstrated significant benefits in terms of weight gain, growth, body protein accretion and stabilization, and, perhaps, improvement of pulmonary function. High-energy defined semielemental formulae are most suitable for use in the nutritional rehabilitation of malnourished CF patients, and are generally well tolerated. As illustrated in Figure 3, surveillance of weight gain and growth provides a reasonable clinical indicator for nutritional intervention and the assessment of benefits derived.

Pancreatic Enzyme Replacement Therapy

Pancreatic enzyme replacement therapy (PERT) needs to be tailored to individual requirements and is required in about 85% of CF patients, according to the amount of food (particularly fat) eaten and the degree of malabsorption. It is appropriate to assess quantitatively the degree of malabsorption and the ability of the therapy to correct this. Some measures to improve the efficacy of PERT have been a recent area of study in CF, as conventional preparations rarely fully optimize absorption. The objectives of PERT include correction of maldigestion, elimination of symptoms and signs of malabsorption, and sustaining normal nutrition.

A bewildering number of pancreatic enzyme preparations are commercially available in the form of tablets, capsules, enteric-coated microspheres, granules, and powders. Some products have not been designed specifically for the treatment of exocrine pancreatic insufficiency, but for patients with unspecified abdominal pain. Some contain bile salts; in general these should be avoided in patients with pancreatic insufficiency because high concentrations of bile salts may aggravate diarrhea. Preparations that are protected against peptic acid inactivation are preferable to unprotected preparations: unprotected ingested enzymes are degraded to a significant degree as a result of increased gastric acidity, and lowered duodenal pH secondary to depressed bicarbonate secretion for the pancreas. The lipase content is the most important determinant of the effectiveness of these products. Between 80% and 90% of unprotected ingested lipase and trypsin is inactivated in the stomach. It was thought that PERT was complication-free, but excessive doses have been associated with renal calculi, and prolonged excessive dosages with serious fibrosing colonopathy. Guidelines for the safe correct use of PERT have been published. In the UK, the Committee of Safety of Medicines suggests doses of no more than 10 000 u lipase kg^{-1} day^{-1}.

It has been calculated that, in an adult, assuming there is no inactivation of enzymes in the stomach, approximately 30 000 IU of lipase is required to be delivered to the duodenum with an average meal containing about 6 g of fat for normal digestion. In children approximately 500–4000 IU is required per gram of fat ingested. The distribution of fat content of different meals and the mixing enzymes with food in the duodenum seems to be important. Granulated preparations or microspheres are better than tablets in that they enable higher enzyme activities to reach the duodenum and mix with the food. The majority of enzyme preparations contain between 5000 and 20 000 IU of lipase. It can thus be calculated that, for the average adult, if some defence against gastric inactivation is provided and there is an adequate enzyme–meal mix, between 2 and 10 capsules are required per meal to control steatorrhea. Studies in children comparing conventional tablet enzyme therapy to enteric-coated microspheres of pancrelipase have shown that significantly less steatorrhea and azotorrhea occur with the use of the microsphere preparation. In addition, there is an improvement in compliance because significantly fewer capsules are required. Irrespective of the preparation used, there is good evidence that the enzymes need to be delivered appropriately and dispersed evenly throughout a meal, in order to achieve maximum exposure of food to the ingested enzymes.

The success of enzyme therapy is assessed by bowel symptoms, quantitative absorptive tests, such as fecal fat analysis, and the maintenance of normal nutrition. In children, assessment of growth is also essential. Azotorrhea is more frequently abolished by pancreatic enzyme supplements than is steatorrhea, possibly because trypsin secretion is better preserved than lipase secretion in pancreatic insufficiency, and because trypsin is not inactivated by acid but only by pepsin. Poor response to pancreatic enzyme preparations may result from poor compliance, inappropriate timing of administration, the presence of another condition causing steatorrhea (e.g., bacterial overgrowth), or the use of an unprotected, acid-sensitive enzyme preparation.

Adjuncts to Pancreatic Enzyme Therapy

Adjunctive treatment to neutralize or inhibit gastric acid and help to protect pancreatic enzymes against acid inactivation may be useful in selected cases. The use of antacids with pancreatic supplementation to neutralize gastric acid has achieved variable results, probably because some antacids form calcium soaps and precipitate glycine-conjugated bile salts. Histamine H_2-receptor antagonists are of theoretical value because many causes of pancreatic insufficiency (particularly CF) are associated with gastric hypersecretion. If there is adequate reduction of gastric activity, the addition of cimetidine does appear to be effective when used with enteric-coated, microsphere, pH-sensitive preparations. However, these enzyme preparations do not disintegrate and release their contents until encountering an alkaline pH, and despite acid suppression therapy, in some cases, complete correction of steatorrhea has not been possible. Recently, a new approach to adjuvant therapy has been suggested with the use of misoprostol, a synthetic methylated prostaglandin E_1 analog, which decreases secretion of gastric acid and increases duodenal bicarbonate secretion. This therapy may have inherent advantages over histamine H_2-receptor antagonists as adjuvant therapy for pancreatic insufficiency because of the latter effect.

Conclusions

Nutritional problems are common in CF and have an increasingly important role in management. Malnutrition is a major factor influencing survival, and may even influence the course of lung disease as there appears to be a close bilateral relationship between the progression of pulmonary disease and undernutrition, although the factors responsible for such a relationship are incompletely defined. Although it is widely recognized that factors such as nutrient losses from malabsorption, a variable but often inadequate intake, and chronic lung disease contribute to malnutrition, recent studies have also suggested that the primary defect may require energy. A wide range of nutritional deficits occurs in CF; many patients have deficits of body fat and body cell mass, and appear to be in a state of chronic catabolic stress. Whole-body protein turnover would also appear to be abnormal in many CF patients, particularly during pulmonary exacerbations. Specific deficiencies of essential fatty acids, fat-soluble vitamins, and some micronutrients also occur and may compromise pulmonary function and immunity. Optimal nutritional management in CF includes the provision of adequate protein and extra energy as fat and carbohydrate, and an appropriate dosage and administration routine with food, of pancreatic enzyme supplements in a form which minimizes acid–peptic inactivation. Newer adjuncts to pancreatic enzyme therapy, such as prostaglandin analogs, show promise. Food supplements should provide extra energy, essential fatty acids, fat-soluble vitamins in a water emulsion, salt supplements in hot weather, and water-soluble vitamins, and micronutrients. Routine surveillance of nutritional status and absorbed energy should be performed at regular intervals. Deviations from normal weight gain and growth may require nutritional intervention. Recent studies by a number of groups confirm that long-term supplementation can achieve sustained catch-up weight gain and growth, and provide support for the view that reversible nutritional factors may have an important influence on the course of pulmonary disease in those CF patients who have deteriorating lung function, and who are unable to sustain normal growth and nutrition by the oral route. Obviously, the overall goal should be to prevent this situation and provide the CF patient with as normal a lifestyle as possible, with management aimed at preventing progressive pulmonary disease, maintaining normal growth and nutrition, and preventing complications.

See also: **Energy**: Energy Expenditure and Energy Balance; **Immunology of Food**; **Metabolic Rate**

Further Reading

Anthony H, Collins CE and Davidson G *et al.* (1999) Pancreatic enzyme replacement therapy in cystic fibrosis. *Journal of Paediatric Child Health* 35: 125–129.

Durie PR (1993) Gastrointestinal and hepatic complication and their management. *International Seminars in Gastroenterology and Nutrition* 2: 3–9.

Hill CM (1998) *Practical Guidelines for Cystic Fibrosis Care*. London: Churchill Livingstone.

Shepherd RW and Cleghorn G (1990) *Nutritional and Gastrointestinal Problems in Cystic Fibrosis*. Boca Raton, Florida: CRC Press.

Stapleton D, Tunnecliffe L, McGuiness D, Sheriff J and Sly P (1998) Development of a nutrition education behaviour intervention in children with CF. *Australian Journal of Nutrition and Dietetics* 55: 130–137.

D

DAHI

C D Khedkar, College of Dairy Technology, Maharashtra, India
G D Khedkar, Dr Babasaheb Ambedkar Marathwada University, Aurangabad (Maharashtra), India
S D Kalyankar, Marathwada Agricultural University, Parbhani (Maharashtra), India
D N Bajad and A R Sarode, College of Dairy Technology, Maharastra, India

Background

Indian curd, known as 'dahi,' is a very well-known fermented milk product consumed by large sections of the population throughout the country, either as a part of the daily diet or as a refreshing beverage, e.g., lassi. Dahi is also used for the production of Shrikhand: some whey is removed from dahi, and the concentrated curd is mixed with an equal weight of sucrose. Dahi has been a very popular fermented dairy product in the Indian subcontinent, which includes India, Pakistan, Bangladesh, Nepal, and Sri Lanka. The name of this product has been found in the ancient Hindu scriptures, and its medicinal value has also been well documented. Dahi may be consumed directly either sweetened or salted and spiced. It is also consumed with other foods such as rice and wheat loaf. Dahi has assumed a special place in the daily diet of Indian people, who prefer to have dahi once or twice a day at morning or evening meals. In 1999–2000, the production of dahi was estimated to be about 7% of total milk production in India. Since conversion of milk into dahi is an important intermediatary step in manufacture of indigenous fat-rich dairy products like butter and ghee, it can be said that over 40% of the total milk production in India is converted into dahi.

Definition

According to the Bureau of Indian Standards, dahi is a product obtained by lactic fermentation of cow or buffalo milk or mixed milk through the action of single or mixed strains of lactic acid bacteria. This definition does not include milk coagulated by the addition of acids or milk coagulating enzymes. As per the Prevention of Food Adulteration Act (1988), dahi or curd is a product obtained from pasteurized or boiled milk by souring natural or otherwise, by harmless lactic acid or other bacterial culture. Dahi may contain added cane sugar. It should have the same percentage of fat and solids-not-fat as the milk from which it is prepared. Where dahi or curd, other than skimmed milk dahi, is sold or offered for sale without any indication of the class of milk, the standards prescribed for dahi prepared from buffalo milk should apply. This is similar to yogurt made from boiled milk after inoculation with a mixed starter culture which consists of dahi left over from the previous lot. However, it is less acidic than yogurt.

Starter Cultures for Dahi

In the production of dahi, inoculation with a small quantity of the desired fermenting flora, named the starter culture, (since the culture initiates or starts fermentation). The most important organisms in starter cultures are the homofermentative and heterofermentative lactic acid bacteria. Dahi starter cultures can be classified into two groups, according to the characteristics of the microorganisms: mesophilic and thermophilic. Mesophilic starters and cultures are used at a temperature of 16–30 °C and contain mesophilic organisms (Table 1). The most typical representatives are the homofermentative *Lactococcus lactis* ssp. *lactis* and *Lc. lactis* ssp. *cremoris*, which are similar, although the former is able to produce free amino acids, which are believed to stimulate the growth of *Lc. lactis* ssp. *cremoris* strains. Another important species in mesophilic starters is the homofermentative citrate$^+$ *Lc. lactis* ssp. *lactis*, which is able to produce lactic acid and in the presence of lactose to ferment citric acid into a number of other compounds such as diacetyl and carbon dioxide. However, at a pH below 5, carbon dioxide and a number of compounds similar to diacetyl, but non-aromatic, are produced. The heterofermentative *Leuconostoc mesenteroides* ssp. *dextranicum* and *Ln.*

Table 1 Designation of dahi based on the type of starter culture

Designation	*Cultures used*	*Remarks*
Sweet dahi	*Lactococcus lactis* ssp. *lactis, Lactococcus lactis* ssp. *cremoris*,	Single or in combination with or without aroma producers like *Leuconostoc mesenteroides* ssp. *dextranicum, Leuconostoc mesenteroides* ssp. *plantarum*
Sour dahi	*Lactobacillus delbrueckii* ssp. *bulgaricus* and *Streptococcus thermophilus*	Single or in combination with or without aroma producers

mesenteroides ssp. *plantarum* are able to ferment citric acid into carbon dioxide, diacetyl, and several other products. Unfortunately, the *Leuconostoc* species may reduce diacetyl into nonaromatic compounds. The intensity of reduction varies between the various representatives of the strains. Homofermentative thermophilic starter organisms like *Lactobacillus delbrueckii* ssp. *bulgaricus* and *Streptococcus thermophilus* are used to obtain mildly sour dahi. These starters play an important role in producing a good-quality dahi with a firm and uniform consistency, sweet aroma, and clean, acidic taste.

Classification of Dahi

Broadly speaking, dahi may be classified into two types:

1. dahi for churning into desi (indigenous) butter;
2. dahi for direct consumption.

Dahi for direct consumption is further classified into the following subtypes:

1. whole-milk and skim-milk dahi;
2. sweet (or mildly sour), sour, and sweetened dahi.

Technology of Sweet and Sour Dahi Manufacture

The steps involved in the manufacture of sweet (or mildly sour) and sour dahi are shown in Figure 1. The manufacture is as follows:

Preliminary Treatment of Milk

The preparation of the basic mix involves selection of milk, preheating, filtration/clarification, and standardization of milk. The different methods for standardizing the milk indicate the possibilities that exist for either increasing or reducing the various milk constituents, especially the fat content. However, the choice of any particular method is primarily governed by the following factors:

- cost and availability of the raw materials;
- scale of production.

The composition of dahi varies considerably. The composition of dahi made from buffalo milk is detailed in Table 2. Commercial dahi tends to contain around 14% milk solids; the use of ultrafiltration to concentrate the solids in cows' milk and skim

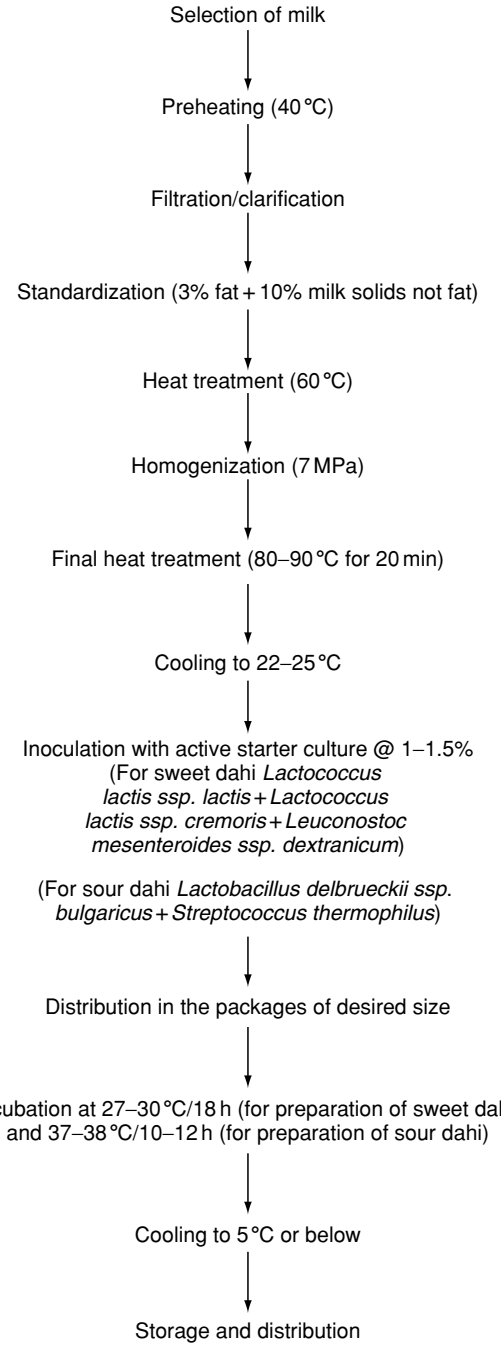

Figure 1 Flow diagram for preparation of dahi.

Table 2 Composition of buffalo milk dahi

Constituents	*Level (%)*
Moisture	85–88
Fat	5–8
Protein	3.2–3.4
Lactose	4.6–5.2
Ash	0.7–0.75
Lactic acid	0.5–1.0
Calcium	0.12–0.14
Phosphorus	0.09–0.11

milk is being considered as a feasible alternative. The importance of total solids stems from the improved consistency imparted to the dahi coagulum, an improvement that is carried further by the homogenization stage by employing a pressure of 17 MPa. Homogenization of milk improves the texture of the coagulum and also reduces the likelihood of whey separation. Homogenization is followed by a final heat treatment of 80–90 °C for a prolonged period of 20 min. The effect of this drastic heat treatment can be summarized as follows:

1. The bacterial load in the milk is reduced, and hence the starter culture has less competition from the contaminating organisms.
2. The whey proteins (albumins and globulins) are denatured and interact with the casein molecules, which form a three-dimensional network on acidification. The network traps the whey proteins, and the dahi coagulum subsequently produced is rendered more viscous.
3. There is a reduction in the amount of oxygen in the milk, and as the normal dahi cultures are microaerophilic, the lowered oxygen tension encourages their growth.
4. Some limited damage to the milk proteins may occur during heating, and the breakdown products can stimulate starter activity.
5. The heat treatment imparts a cooked flavor for misti dahi.

Fermentation of Milk

The acidification of milk during the manufacture of dahi is an important biochemical process which must be carried out under controlled conditions in special incubators and/or fermentation tanks. Fermentation tanks are used as incubators only, and they are usually insulated in order to maintain the appropriate temperature. The processing of the milk and the cooling of the dahi are carried out in other equipment in the production line. Cabinets are also used for this purpose. These comprise a small insulated room which is divided into compartments, and most incubators of this type are multipurpose chambers capable of circulating hot or cold air. Hot air is circulated during the fermentation period, followed by cold air during the cooling stage. Sometimes, these cabinets are used as incubators only, and the dahi can be cooled in refrigerated cold storage. In the case of mildly sour (sweet) dahi manufacture, the temperature is maintained at 27–30 °C, while for sour dahi, it is 37–38 °C.

Cooling of Dahi

After the incubation period, dahi is cooled in order to control the level of lactic acid in the product. The rate of cooling can affect the structure of the coagulum. Thus, very rapid cooling can lead to whey separation, if the protein filaments are concentrated too rapidly, which in turn affects their hydrophilic properties.

Microbiological Quality of Dahi

An examination of the microbiological quality of a product is usually concerned with two aspects, namely: protection of the consumers from exposure to any health hazard and ensuring that the material does not suffer microbiological deterioration during its anticipated shelf-life. Both these aims are of importance to the consumer and producer alike. Thus, over and above any moral responsibility that a company accepts for the integrity of its product, the financial losses that follow a public health incident, or even a high level of consumer complaints, provides a very strong incentive for a manufacturer to give quality assurance a high priority. The type of hazard that may be encountered depends on the nature of the product, and in general terms, dahi can be regarded as 'hygienically safe.' The reason for this confidence stems from the level of acidity present (around 0.7–1.0% lactic acid), for this situation, potential pathogens such as *Salmonella* spp. will be largely inactive. Similarly, 'coliforms' will be unable to survive the low pH encountered, and this inhibition is reinforced by the production of antibiotic substances by the dahi starter organisms. Microbiological standards for dahi are given in Table 3. Typically, the shelf-life of dahi is 2–4 days at 4 °C.

Table 3 Indian Standard Institute specifications for dahi

Characteristics	*Requirement*	
	Sweet dahi	*Sour dahi*
Acidity, lactic (% wt) (max.)	0.70	1.0
Yeast and mold count g^{-1} (max.)	100	100
Coliform count g^{-1} (max.)	10	10
Phosphatase test	Negative	Positive

From Indian Standards Institution (1973) *ISI Document No. 7035*. New Delhi: Manak Bhavan.

Technology of Sweetened Dahi (Misti Dahi) Manufacture

Misti dahi (syn. sweetened dahi, red dahi, payodhi) is a traditional sweetened fermented milk product of the Eastern region of India. It is prepared on a cottage scale by sweetmeat makers everyday to meet local demands. Traditionally, milk with added cane sugar is continuously heated in an open pan at a simmering temperature (68–70 °C) for 6–7 h to concentrate and develop the intense cooked flavor and brown color, increase the viscosity, and induce other physicochemical changes. After cooling to about 40 °C, the mix is inoculated with a commercial misti dahi starter culture, which is generally a mixed lactic culture like LF-40 or *Lb. delbrueckii* ssp. *bulgaricus*, *St. thermophilus*, and some aroma-producing organisms, as in mildly sour and sour dahi. Wide variations in total solids content (27–43%), milk-solids-not fat (11–16%) and sucrose (13–19%) in market samples of misti dahi sold in Calcutta city have been reported. Many flavor defects such as fruity, alcoholic, highly acidic, flat, etc., and textural defects like gassiness, weak body, wheying off, and a thick crust on the top surface have been observed in most of the market samples. In view of the increasing nationwide demand for this product and the growing interest in organized dairy plants for large-scale manufacture, the technology has been developed to suit industrial-scale production. The manufacturing technology for the industrial production of misti dahi is outlined in Figure 2. Fresh buffalo milk is standardized to 3.5% fat and 9% milk-solids-not fat, heated to 65 °C in a plate heat exchanger, and homogenized at a pressure of 56 kg cm^{-2} (one stage). Milk is concentrated to 1.44-fold in a vacuum evaporator. After adding cane sugar (7–15%), the milk is heated at 85 °C for 10 min to generate a cooked flavor. The mix is water-cooled to 40 °C before inoculation with the mixed culture, LF-40. In some cases, sugar, caramel, jaggery, and artificial colors are added to impart a brown color. The inoculated mix is aseptically distributed into presterilized polystyrene containers of the desired size and mechanically transferred to an incubation chamber at 40 °C. During curd formation, the milk must remain stationary. In all postfermentation activities, the gel should be subjected to minimal external influences. Diacetyl is the predominant flavor component, but microquantities of acetoin, acetic acid, and carbonyls are also present. Off-flavors in this product can be traced back to poor-quality raw milk, use of contaminated cultures, a high incubation temperature, and contaminated packages.

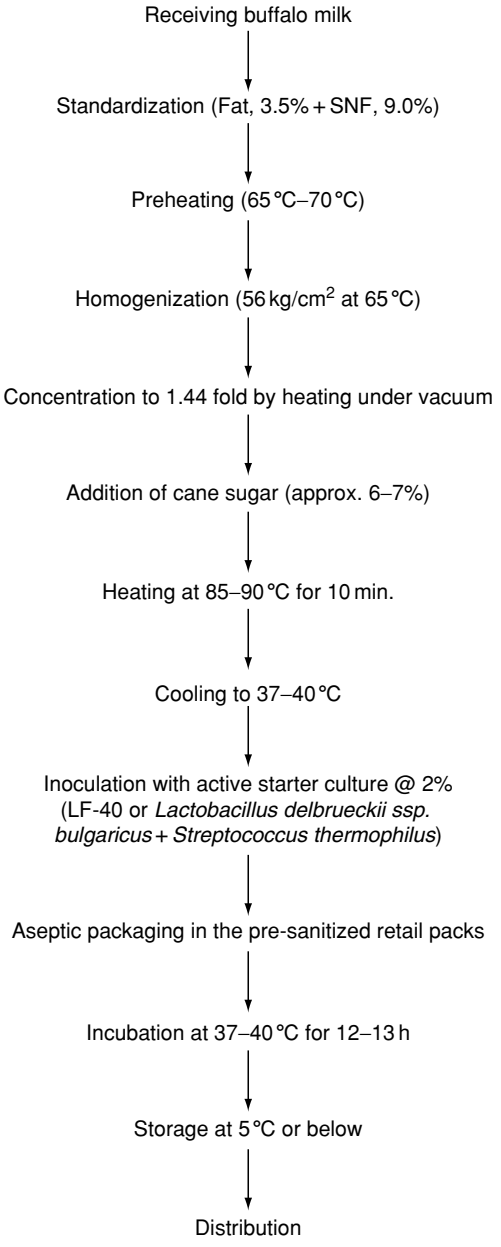

Figure 2 Flow diagram for the preparation of misti dahi (sweetened dahi). The temperature of homogenization is different for both types.

Food and Nutritive Value and Therapeutic Benefits of Dahi

Indian literature has described in detail the importance of dahi in human nutrition. Scientific investigations have proved that dahi is digested and assimilated more easily by virtue of its buffering action and supplies more nutrients than milk. Tricalcium phosphate is converted into monocalcium phosphate, and some insoluble minerals are rendered more soluble due to lactic acid production. Many researchers have found an increase in the levels of

Table 4 Mineral and vitamin content (per 100 g) of dahi

Constituents	*Milk*	*Dahi*
Mineral matter (g)	0.8	0.8
Calcium (mg)	149.0	149.0
Phosphorus (mg)	96.0	93.0
Vitamin A (IU)	118.0	102.0
Thiamin (μg)	55.0	49.0
Riboflavin (μg)	167.0	157.0
Nicotinic acid (μg)	96.0	86.0
Biotin (μg)	29.0	3.2
Pantothenic acid (μg)	202.0	183.0
Folic acid (μg)	161.0	178.0
Vitamin B_{12} (μg)	0.15	
Ascorbic acid (mg)	1.4	1.3

riboflavin (vitamin B_2) and thiamin (vitamin B_1) in dahi.

Converting milk into dahi also tends to stabilize the vitamin C potency more than fresh milk, which is vitally important for the growth and repair of tissues in bones, teeth, and blood vessels, health of gums and teeth, overall body growth, and resistance to infection. Lactose-fermenting bacteria have been found to synthesize vitamin B, present only in small quantities in milk. The mineral and vitamin contents in dahi are listed in Table 4.

Dahi obtained from cows' milk has been found to cure dysentery, diarrhea, thirst, coughs and colds, indigestion, headache, weakness, and many other diseases and ailments. It improves the digestibility of food and allays the fear of diseases of any nature by checking the undesirable putrefactive fermentations in the alimentary canal. The putrefying bacteria inhabiting the human intestines are harmful, because they poison the food with the products of their metabolism. Since the bacterial flora of the intestines live in an alkaline medium, the best thing would be to create an acid medium in the lower intestine. For this reason, the introduction of regular use of dahi is advocated. Its palatability and pleasant flavor have an appetizing effect. Those who do not relish milk like dahi. It is beneficial to youngsters and pregnant and lactating females. It is a good dietary item for all and deserve a high priority in our daily menu for healthful living.

Industrial methods have been developed during the last decade to mechanize the production, packaging, and storage of dahi, and it is now produced regularly on a factory scale. The need for more promising mechanization and upgrading of production technologies, packaging, preservation, and distribution of dahi has now been recognized. Attempts to develop any large-scale process for dahi-based sweets must aim at using the existing processes in dairy plants rather than relying on a new plant design and fabrication, which carry limitations of available engineering skill, high costs, and time loss. In view of the growing market demands for dahi and its derivatives, there is a great scope for modernizing the existing technology for industrial production.

See also: **Fermented Milks**: Dietary Importance

Further Reading

De S (1977) *Outlines of Dairy Technology*. Oxford: Oxford University Press.

Ghosh J and Rajorhia GS (1987) Chemical, microbiological and sensory properties of Misti dahi sold in Calcutta. *Asian Journal of Dairy Research* 6: 11–15.

Ghosh J and Rajorhia GS (1990) Technology for production of misti dahi – a traditional fermented milk product. *Indian Journal of Dairy Science* 43: 239–244.

Indian Standards Institution (1973) *ISI Document No. 7035*. New Delhi: Manak Bhavan.

Laxminarayana H, Nambudripad VKN, Laxmi NV, Anantaramiah SN and Sreenivasamurty V (1952) Composition of dahi. *Indian Journal of Veterinary Science and Animal Husbandry* 22: 13–16.

Yadav JS, Grover S and Batish VK (1993) *A Comprehensive Dairy Microbiology*. New Delhi: Metropolitan.

DAIRY PRODUCTS – NUTRITIONAL CONTRIBUTION

J Buttriss, British Nutrition Foundation, London, UK

Introduction

Dairy products are traditional dietary items in many parts of the world, in particular in regions such as northern Europe where the cooler climate is especially suited to dairying. The history of milk as a food has been documented over the centuries and examples of early dairying are depicted in Egyptian friezes such as that from the sarcophagus of Queen Kawit from Der-al-Bahri, between Luxor and Karnack, dating back 4000 years. There is an even earlier Mesopotamian frieze from the temple of Nin-khasarg, near Ur, which is thought to be 1000 years older.

The popularity of milk as a staple food over the centuries must partly be due to its versatility. Early humans discovered that it could be churned to make butter and fermented with bacterial cultures to produce cheese and yogurt, all of which were methods of preserving some or all of the nutrients in milk for consumption at a later date. (*See* **Cheeses**: Dietary Importance; **Milk**: Dietary Importance; **Yogurt**: Dietary Importance.)

This article summarizes the nutritional contribution made by milk and milk products.

Variety

Today, the range of dairy products on the market is immense. In most countries, a range of milks with differing fat contents is available. For example, in the UK consumers can choose between Channel Islands milk, with 5.1 g per 100 g fat, through whole milk (3.9 g per 100 g) to semiskimmed (1.6 g per 100 g fat) and skimmed milk, which has virtually no fat. Similarly, a wide range of cheeses exists with varying fat contents: at one end of the spectrum are soft fresh cheeses, made with skimmed milk and, at the other, hard cheeses such as cheddar. Also available is cheese made with nonanimal rennet, suitable for vegetarians. In the UK alone, almost 200 different cheeses are produced, and cheese is particularly popular in countries such as France, where an even greater variety is available.

Fermented milk products such as yogurt, smetana, and kefir have always been popular in Middle Eastern countries, but their popularity, particularly that of yogurt, is dramatically increasing in the UK. Again, a wide range of yogurts exists, from the very-low-fat to the creamier, whole-milk or Greek-style product, including set yogurts, stirred yogurts, fruit yogurts, frozen yogurts and fromage frais. (*See* **Fermented Milks**: Types of Fermented Milks; **Yogurt**: The Product and its Manufacture.)

Traditional products such as cream and butter are still in demand, in spite of their relatively high fat content, and are being joined by other 'luxury' products, such as real dairy icecreams, fresh cream desserts, and luxury mousses. To meet the demand for a spread with a buttery flavor, spreads have been developed which incorporate butterfat for taste but often have a lower fat and calorie content. (*See* **Butter**: The Product and its Manufacture; **Cream**: Types of Cream; **Ice Cream**: Methods of Manufacture.)

A number of products also exist to which nutrients have been added, such as the calcium-enriched milks, and yogurts fortified with additional vitamins.

Nutritional Value

Most dairy products offer a wide range of nutrients and this is reflected in the contribution made by milk and milk products to dietary intake in many countries. For example, in the UK, milk and milk products provide over half the calcium in the average diet and make a significant contribution to the needs of other nutrients (Table 1).

Nutritional Significance of Milk and Milk Products

In countries where dairying has traditionally been a strong industry, milk and milk product consumption tends to be widespread and makes a significant contribution to nutrient needs. For example, the average daily intake of milk in the UK in 1998 was a little under half a pint per person (285 ml). Table 2 shows the contribution this quantity of whole milk makes to the nutrient and energy needs of a 4-year-old girl and a man. Skimmed or semiskimmed milk would make smaller contributions to intake of energy and fat-soluble vitamins.

In Scandinavian countries, where dairying is also traditional, liquid milk consumption is typically higher than in the UK. However, in warmer climates, in particular the Indian subcontinent, Africa and

Table 1 Contribution of milk and milk products to daily nutrient intake in the UK

	Whole milk		*Semiskimmed and skimmed milk*		*Cheese*		*All milk and milk products, including cheese*	
	Amount	*% of total intake*	*Amount*	*% of total intake*	*Amount*	*% of total intake*	*Amount*	*% of total intake*
Energy (kJ)	276	3.6	314	4.1	226	2.9	974	12.6
Total fat (g)	4.0	5.2	2.3	3.0	4.5	5.9	12.7	16.7
Saturated fatty acids (g)	2.6	8.6	1.6	5.3	2.8	9.3	8.1	26.9
Polyunsaturated fatty acids (g)	0.1	0.7	0.1	0.7	0.2	1.5	0.5	3.7
Total sugars (g)	4.4	4.3	7.9	7.7	0.1	0.1	16.2	15.7
Calcium (mg)	116	14.1	198	24.1	91	11.1	452	55.1
Iron (mg)	0.1	1.0	0.1	1.0	–	–	0.2	2.0
Sodium (mg)	42	1.7	70	2.8	104	4.1	240	9.4
Vitamin C (mg)	1.7	2.7	2.8	4.4	–	–	5.0	7.9
Folate (μg)	7	2.9	9	3.7	5	2.0	24	9.8
Vitamin A (μg) retinol eq.	32	4.1	34	4.3	49	6.2	141	17.9
Vitamin D (μg)							0.2	6.3

Reproduced from Ministry of Agriculture, Fisheries and Food (1999) *National Food Survey 1998*. London: Stationery Office, with permission.

Table 2 Contribution of 285 ml (half a pint) of whole milk to nutrient needs

	Percentage of UK reference nutrient intake (RNI)	
	Girls (4–6 years)	*Adult men (19–50 years)*
Energy[a]	12.5	7.6
Fat[a]	19.0	11.5
Saturates[a]	37.3	22.7
Monounsaturates[a]	14.3	8.6
Polyunsaturates[a]	2.7	1.7
Carbohydrate[a]	7.0	4.2
Protein	47.5	16.8
Nonstarch polysaccharide	–	–
Vitamin A	32.8	23.4
Thiamin	16.4	11.5
Riboflavin	62.5	38.5
Nicotinic acid equivalents	22.1	14.3
Vitamin B_6	19.4	12.5
Folic acid	17.5	8.8
Vitamin B_{12}	143.8	76.7
Vitamin C	10.0	7.5
Calcium	74.8	48.1
Sodium	23.9	10.4
Chloride	26.6	11.7
Copper	Trace	Trace
Iodine	44.0	31.4
Iron	2.4	1.7
Magnesium	26.7	10.7
Phosphorus	77.0	49.0
Potassium	37.3	16.4
Selenium	15.0	4.0
Zinc	17.7	12.1

[a]There is no RNI for these components of food. For energy, the estimated average requirement has been given; for fat and carbohydrate, the desirable average has been given.
Figures for milk composition from Holland B, Unwin ID, Buss DH (1989) *Milk Products and Eggs*. Fourth supplement of McCance and Widdowson's *The Composition of Foods*, 4th edn. London: Royal Society of Chemistry and Ministry of Agriculture, Fisheries and Food.
Figures for RNIs from Department of Health (1991).

South America, climatic conditions lend themselves less readily to cows' needs, so that the development of dairying and, hence, milk-drinking habits has been far more patchy. Exceptions exist, such as the Masai in Eastern Africa, whose culture is dominated by the cow. Similarly, in parts of India, cows have religious significance but have not until recently been intensively managed for their milk. Table 3 presents estimates of milk consumption in different countries.

As it is rich in calcium needed for skeletal development and maintenance, milk has traditionally been seen as an important food during childhood, pregnancy, and lactation, when calcium requirements are particularly high. This view is still supported today by a recent UK government report. A pint of milk (0.568 l) provides about 700 mg of calcium. (*See* **Calcium**: Properties and Determination; **Children**: Nutritional Requirements.)

Claims of Adverse and Beneficial Effects to Health

Allergy or Intolerance to Cows' Milk

True milk allergy is typically seen in infants. Although estimates vary, it is thought that at its peak incidence in 1–3-year-olds, cows' milk allergy affects about 1% of infants. Most children outgrow this reaction to the protein in milk by their fifth birthday and in many instances it is short-lived, being brought on by a gastrointestinal upset.

Intolerance to the sugar in milk – lactose – occurs in those people who lack or who have low levels of the enzyme lactase, which digests lactose to its components, glucose and galactose. This enzyme is invariably present in the intestines of infants to enable them to

digest their mother's milk, but levels fall off in some adults, particularly those who come from areas of the world where milk consumption is not traditional. Hence, in the UK, lactose intolerance is rare among Caucasians but is experienced by some ethnic groups of Indian, African, or Chinese extraction. In many cases, sufferers have low levels of the enzyme and can thus tolerate small quantities of milk and larger quantities of foods such as cheese or yogurt. In the case of cheese, the lactose has largely been lost with the whey in the cheese-making process. In the case of yogurt, it is suggested that the β-galactosidase enzyme, present in the bacterial culture used to make yogurt, may be able to digest some of the lactose. (*See* **Food Intolerance**: Lactose Intolerance.)

Table 3 Estimates of liquid milk consumption in a selection of countries in 1998

	Liquid milk consumption (kg per person per year) 1998
Australia	103
Canada	90.7
Chile	30.3
Czech Republic	57.0
Denmark	93.8
Finland	143.5
France	75.5
Germany	64.4
Iceland	153.6
Ireland	143.5
Israel	59.2
Italy	85.3
Japan	40.2 (1997 figure)
Netherlands	54.9
Norway	119.3
South Africa	25.1
Spain	94.3
Sweden	116.4
UK	117.9
USA	98.9

Reproduced from International Dairy Federation (1999) The world dairy situation 1998. *Bulletin of the IDF* 339: 49, with permission.

Health Benefits of Yogurt

A number of health benefits of yogurt have been claimed, including an ability of certain types of bacteria used in the manufacture of yogurt to colonize the gut and help fend off harmful bacteria.

Coronary Heart Disease

One of the risk factors for coronary heart disease (CHD) is a raised plasma cholesterol level. A number of factors, including genetic make-up, influence an individual's plasma cholesterol level. One such factor is thought to be intake of saturated fatty acids, particularly in genetically susceptible people who are less able to remove low-density lipoproteins (LDLs) from their blood stream.

Many governments have issued guidelines on diet in relation to heart disease. In the UK, guidelines of energy and nutrient intake include recommendations on total fat intake and on fatty acid intake. The guidelines on saturated fatty acids (saturates) have been set with a view to reducing average plasma cholesterol

Table 4 Contribution of sources of fat and saturates in the household food supply

	Fat		*Saturated fatty acids*	
	g day^{-1}	*%*	*g day^{-1}*	*%*
Whole milk	4.0	5.2	2.6	8.6
Reduced-fat milks	2.3	3.0	1.6	5.3
Cheese	4.5	5.9	2.8	9.3
Total milk, cream, and cheese	12.7	16.7	8.1	26.9
Carcass meat and poultry	6.9	9.1	2.6	8.6
Bacon, ham, meat products	9.8	12.7	3.7	12.3
Total meat	16.7	21.9	6.3	20.9
Total fish	1.3	1.7	0.3	1.0
Eggs	1.4	1.8	0.4	1.3
Butter	4.5	5.9	3.0	10.0
Margarine	2.9	3.8	0.8	2.7
Other fats and spreads	13.7	17.8	2.7	9.0
Total fats and spreads	21.1	27.7	6.5	21.6
Total vegetables	4.9	6.4	1.5	5.0
Total fruit	1.3	1.7	0.3	1.0
Cakes, biscuits and pastries	6.8	8.9	3.3	11.0
Bread, flour, breakfast cereals	1.6	2.1	0.3	1.0
Total cereals	13.1	17.2	5.1	16.9
Other foods and beverages	3.7	4.9	1.6	5.3
Total all foods and drinks	76.9	100	30.1	100

Reproduced from Ministry of Agriculture, Fisheries and Food (1999) National Food Survey 1998. London: Stationery Office, with permission.

concentration by 0.4 mmol l^{-1}, thus bringing it down to a level associated with a substantially lower incidence of CHD than witnessed in the UK at present.

In 1999, saturates provided an average of 14.9% of energy intake. The Department of Health report advises that average consumption is reduced to 10% of total energy (11% of food energy). This advice does not apply to children under the age of 2 years.

Intake of specific foods has not been linked directly with CHD risk. However, a reduction in foods contributing significant amounts of fat (Table 4), especially saturated fatty acids, to the diet has been recommended for those people with a high fat intake. Such advice calls for a reduction in foods such as whole milk and its products, fatty meats, fried foods, spreading and cooking fats, and baked goods such as biscuits, cakes, and pies.

A wide range of lower-fat milks and milk products is now available, so that those people concerned about their fat intake can change to a lower-fat option rather than cutting down on dairy products and losing a valuable source of calcium and other essential nutrients.

Dairy Products and Dental Health

The frequency of sugar consumption is positively related to the incidence of dental caries. Although lactose (milk sugar) is moderately cariogenic if consumed in isolation, when it is consumed as part of milk any cariogenicity is counteracted by protective factors in milk. Consequently, milk without added sugars can be considered to be virtually noncariogenic. (*See* **Dental Disease**: Etiology of Dental Caries.)

Cheese is now recognized to help protect against dental decay. Dental caries results from acid production at the tooth surface, following fermentation of sugar by the bacteria present in dental plaque. It has been known for some time that eating a small amount of cheese at the end of a meal can counteract this effect; however, a recent study indicated that cheese incorporated within a meal confers similar benefits, in so far as it reduces the fall in plaque pH recognized to be associated with decay.

A number of possible mechanisms have been offered to explain this beneficial effect of cheese. Eating cheese causes an increased flow of saliva, which is slightly alkaline and acts as a buffer to the acid. Also, cheese favors remineralization of any damage since it is rich in calcium and its consistency brings calcium and phosphorus into contact with the tooth surface.

See also: **Butter**: The Product and its Manufacture; **Cheeses**: Quarg and Fromage Frais; Dietary Importance; **Cream**: Types of Cream; **Dental Disease**: Etiology of Dental Caries; **Fermented Milks**: Types of Fermented Milks; **Food Intolerance**: Lactose Intolerance; **Ice Cream**: Methods of Manufacture; **Milk**: Dietary Importance; **Yogurt**: The Product and its Manufacture; Dietary Importance

Further Reading

Ashwell M (ed.) (1997) *Diet and Heart Disease: A Round Table of Risk Factors*, 2nd edn. London: Chapman & Hall.

British Dietetic Association (1987) *Children's Diets and Change*. A report of the Child Health and Nutrition Working Party. Birmingham: British Dietetic Association.

Buttriss JL (ed.) (2001) *Adverse Reactions To Food. British Nutrition Foundation Task Force Reporter.* Oxford: Blackwell Science.

Department of Health (1991) *Dietary Reference Values for Food Energy and Nutrients in the UK*. Committee on Medical Aspects of Food Policy. Report on Health and Social Subjects 41. London: HMSO.

Department of Health (1994) *Nutritional Aspects of Cardiovascular Disease*. Report on Health and Social Subjects 46. London: Stationery Office.

Department of Health (1998) *Nutrition and Bone Health; with Particular Reference to Calcium and Vitamin D*. Report on Health and Social Subjects 49. London: Stationery Office.

DHSS (1989) *Dietary Sugars and Human Disease*. Committee on Medical Aspects of Food Policy. Report on Health and Social Subjects 27. London: HMSO.

International Dairy Federation (1999) The world dairy situation 1998. *Bulletin of the IDF* 339: 49.

Ministry of Agriculture, Fisheries and Food (1999) *National Food Survey 1998*. London: Stationery Office.

Moynihan P (2000) Food and factors that protect against dental caries. BNF Nutrition Bulletin 25(2) (in press).

Renner E (1983) *Milk and Dairy Products in Human Nutrition*. Munich: Volkswirtschaftlicher Verlag.

Renner E (1989) *Micronutrients in Milk and Milk-Based Food Products*. London: Elsevier Applied Science.

Silva MF A *et al*. (1986) Effect of cheese on experimental caries in human subjects. *Caries Research* 20: 263–269.

The Dairy Council (2000) *Milk Matters: A Guide to the UK Dairy Industry and its Products*. London: The Dairy Council.

DATE PALMS

A A A Kwaasi, King Faisal Specialist Hospital & Research Centre, Riyadh, Saudi Arabia

Background

The date palm (*Phoenix dactylifera* **L.**) has been known for as long as recorded history. Apart from its importance as a food source for man and animals, all parts of the plant have their use in medicine, pharmacognosy, chemistry, and religion, and in the fishing, horticulture, and construction industries. Like all economic plants, dates are also afflicted by a number of diseases from microbes, insects, and rodents. Dates are one of the few foods that have a high potassium content and, at a certain stage of development, have a low sucrose content. As knowledge in agriculture, medicine, and industry develops, so does the number of new agents that are found to be harmful or useful to man. In the last two decades, there has been a dramatic increase in the number of foods that have been found to be allergenic, and accidental consumption of these foods by susceptible individuals has led to fatalities. Date fruit and pollen are no exception. Results of recent research have revealed that both date pollen and fruit peptides from certain date cultivars can elicit allergic responses in susceptible individuals and share a number of cross-reactive epitopes with some well-known allergens. The beneficial properties of the date are such that its cultivation should be expanded.

History

The date palm has provided inhabitants of the hot dry regions of the world with a highly desirable food and has therefore been thoroughly exploited. The earliest archeological finds dated 5000–6000 BC were from Iran, Egypt, and Pakistan. Remnants of dates have been found in excavations of Neolithic sites dating back 7000–8000 years and in the tombs of Egyptian Pharaohs. The earliest cultivated finds dated around 4000 BC were from Eridu, Lower Mesopotamia in the Bronze Age. Dates are mentioned in Akkadian and Sumerian cuneiform sources (*c.* 2500 BC) and later. The date palm has historical and religious fame in Judaism, Christianity, and Islam. Date palm fronds are used on Palm Sunday and are thought to have been used in Jesus' triumphant entry into Jerusalem (Synoptic Gospels, New Testament), and the book of Psalms states that 'the righteous shall flourish like a palm tree.' Date fruits are eaten by most Christians during Christmas and are used in breaking the fast during the Islamic Holy Month of Ramadan. It is mentioned in the Qur'ān (Maryam 19:25–26) in relation to Jesus' birth, and there are a total of 26 references to dates in this holy book.

Origin and Distribution

The palm is presumed to have originated in the Persian Gulf area and in Western India, but its exact origin is difficult to establish. Apart from remnants of some wild dates still present in areas around Jordan and the Iraq–Iran border, all dates known today are fully domesticated. The date was naturalized in Arabia and spread through Northern Africa, Spain, the drier parts of the Nile valley, and around the Euphrates River. From Spain, it was introduced to America, where it is now grown commercially in the warm valleys of Southern California and Arizona. Dates are cultivated throughout the desert regions of the world and grow well between latitudes of 15 and 35 ° N, from the Canaries and Morocco in the west, to India in the east. It is recorded from all the inner and littoral parts of North Africa, the southern parts of the Balkan Peninsula, and Asia Minor, from Syria, Palestine, Transjordanian, Iraq, Arabia Peninsula, Iran and Baluchistan regions. Dates are grown on family compounds and along streets and highways, and are seen in every conceivable space in the Middle East. The date is cultivated as an ornamental plant in South Europe, where it seldom matures into fruit, except in southern Italy and Spain, and is cultivated in parts of Queensland, Australia.

Botany

The date palm belongs to the Class: Monocotyledoneae, Subclass: Areciae and is a member of the Family Arecaceae (formerly Palmaceae). It is an aborescent, erect, unbranched stem with large petioles that are sheathed at their bases. The tree is characterized by numerous offshoots that are produced at the base of the trunk of young palms. The average mature tree is about 20 m tall, but some may be as high as 30 m. A young, actively bearing date palm, showing offshoots and fruits at the end of the 'khalaal' stage is shown in Figure 1.

The trunk of the palm is covered with persistent grayish leaf bases. It is surmounted by a handsome array of pinnately divided long leaves and needle-sharp fronds. Between 10 and 20 new leaves are

Figure 1 (see color plate 42) (a) Young actively bearing date palm with offshoots or suckers. (b) Bunches of fruit from the same tree photographed at the end of the 'khalaal' stage.

produced annually. Date palm leaves are subtended at their bases by a cylindrical sheath of reticulate mass of tough, fibrous material, together forming a tight protective envelope for the terminal bud. Even though the phyllotaxy of date palm is not always clear, it forms the basis for the identification of some cultivars.

The palm produces many long rope-like adventitious roots that arise from the spherical base of the stem, often traversing long distances in search of water and nutrients. In Egypt and Algeria, mycorrhizal association of the root tips with arbuscular fungi of the genera *Glomus*, *Autochthon*, *Aoufous*, and *Sclerocystis* is common, but *Glomus mosseae* has been found to be most efficient in helping the date palm maintain water balance.

Male and female flowers are borne on different plants (dioecious), but all date plants look alike. Both male and female inflorescences are branched and are completely enclosed during development by sheathing fibrous spathes. The female inflorescence has pendulous branches, and florets (small flowers) are spread out along their entire length. A female flower with branches of male inflorescence inserted and tied, and a male inflorescence with abundant pollen are shown in **Figure 2a** and **2b**, respectively.

A typical female flower (gynoecium) has a small cup-shaped calyx tube formed from three outer perianth segments. It has three inner perianths, which are small and circular, and are closely appressed to the three carpels of the superior apocarpous ovary, each of which is surmounted by a short, hooked stigma. Only one ovary develops into a seed after fertilization. The floral formula of the female flower is: $K_3C_3A_0G_3$.

The male palms produce between 30 and 50 inflorescences yearly. Their spathes are narrower and shorter than the female flowers and are covered with a rust-colored powdery coating (*See* **Figure 2b**). The inflorescence consists of a short, stiff peduncle with numerous copious branches and white, sweet scented flowers along their length. The male flower has six stamens bearing long flattened anthers with undulate margins and has the floral formula: $K_3C_3A_6G_0$. The center of the androecium bears a (vestigial) small rudimentary ovary. Pollen production is prolific, and typical pollen is light, small, and pale yellow in color, and measures about 8 by 25 μm. Date pollen is distinct and only slightly resembles the pollen of *Sabal palmetto* and *Ginkgo biloba* (see scanning electromicrograph of pollen grains in **Figure 3**).

Figure 2 (see color plate 43) (a) Female flower with branches of male inflorescence inserted and tied to the female. (b) Opened male inflorescence with abundant pollen.

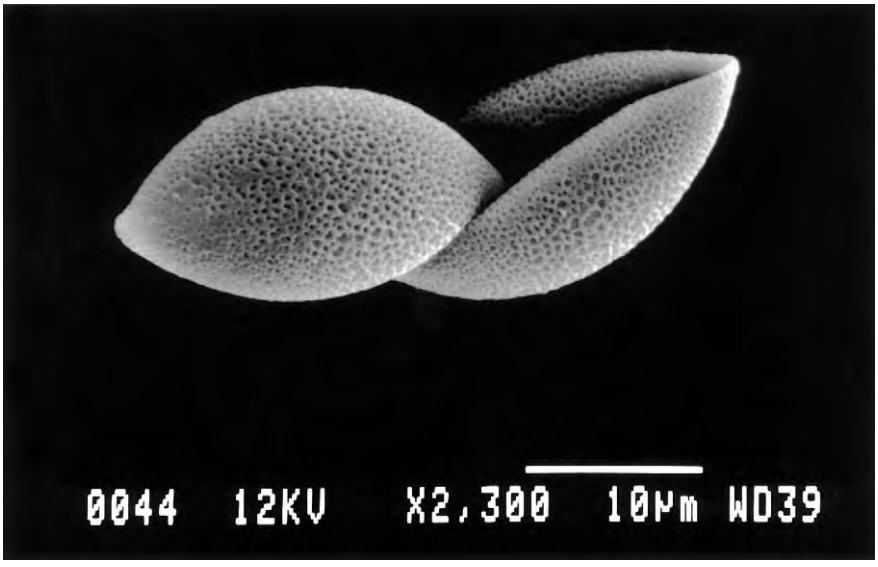

Figure 3 Scanning electromicrograph of two pollen grains.

Genetic Diversity

The heterozygous nature of the palm, its dioecious habit and the long period of cultivation, has led to the establishment of thousands of different date cultivars. In Riyadh (Saudi Arabia) alone, about 100 different cultivars are sold in the local open markets during harvest. A photograph of 18 of these cultivars is

shown in Figure 4. Results of enzymatic and genetic studies have been carried out on the fruits of several cultivars in Algeria and Iraq, and other studies of isoenzymes in leaves demonstrate that there is a high percentage of polymorphic loci, strong heterozygosity, and genetic diversity among cultivars. Extensive date breeding continues, and reports from the African Centre of Diversity indicate that date cultivars able to withstand such characteristics as alkaline soil, drought, frost, heat, heavy soil, high pH, salt, sand, slope, smog, and water logging have been bred. Furthermore cultivars with such fruit qualities as attractiveness, resistance to spoilage, heavy yield, late or early maturity, richness of flavor, smoothness of skin, and coloration are available in the germplasm repository of USDA, California, USA.

Many of the native palms in the Arabian Peninsula and North Africa have originally grown from either volunteer or intentionally nursed seedlings, but in modern date culture, many different cultivars are propagated vegetatively from offshoots or suckers (*See* Figure 1a). There are hundreds of established varieties or cultivars, and it is estimated that there are 350 cultivars of soft dates in the Basra region of Iraq alone.

Dates can be categorized into three different groups depending on the type of fruits they produce: soft, semisoft, and dry. Soft dates have a soft fleshy consistency and contain about 60% sugar in the form of glucose and fructose, and this concentration is sufficient to prevent undesired fermentation. They are used fresh or preserved in the form of 'agwa,' a pressed mixture of dates. Semisoft dates have a firmer flesh and contain much more sugar than soft dates. Dry dates contain much more sugar than soft and semisoft dates (65 to $\geq$70%), and they ripen and dry to a hard consistency while still on the palm tree.

Ecology

The environmental demands of date palms are summed up in the old Arab adage that 'the date palm needs its feet in running water and its head in the fire of the sky.' Date palms enjoy high temperatures and can withstand very low temperatures for short periods only. Temperatures below $-8\,^{\circ}C$ are harmful. They can tolerate annual precipitation of between 31 and 403 mm and grow well in regions of low humidity and where rainfall is low or nonexistent. Flowering is favored by long, hot, and

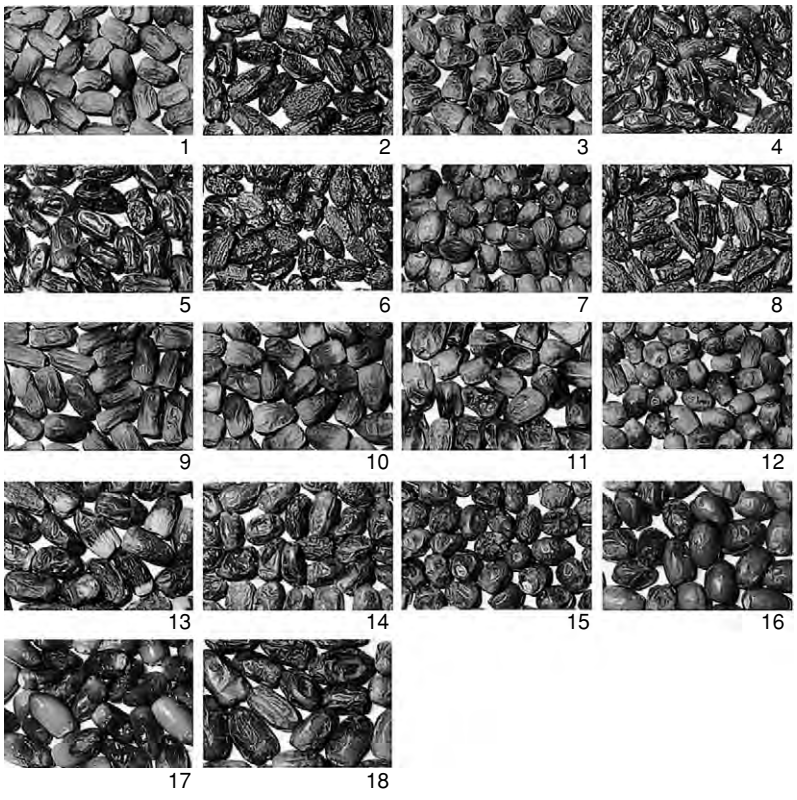

Figure 4 (see color plate 44) Eighteen of the most popular date cultivars sold in Riyadh, Saudi Arabia. Their names are as follows: 1 = Assela, 2 = Labana, 3 = Sabbaka, 4 = Rothana, 5 = Manifi, 6 = Naboot-Ali, 7 = Sefri, 8 = Mabroom, 9 = Roshidia, 10 = Sokary, 11 = Khodri, 12 = Safawi, 13 = Halaw/Helwa, 14 = Hanini, 15 = Barhi, 16 = Juffair, 17 = Qattar, 18 = Seeleg.

completely dry periods, and an average temperature of 30 °C is considered optimal for proper ripening. The plants can tolerate a wide variety of soil types and a fair amount of alkalinity (pH range of 5–8.2) with the only proviso that the soil be well aerated. To bear well, the roots must be in a stratum with less than 1% of alkali salts. Recently, salt-tolerant date cultivars have been bred, and cultivation in halophytic (salty) soils close to the sea has been made possible. Most date palms are grown under carefully controlled irrigations or in oases. Date trees must be pruned to allow new leaves to grow, to prevent rodent and insect infestation, and to facilitate pollination, fruit setting, and harvesting.

Propagation

Dates can be generally grown from seeds or vegetatively propagated from offshoots, but in recent years, organogenesis has been researched, and the use of plantlets from tissue culture has become popular. Although impractical for the poor peasant farmer, it has the advantage that several plantlets can be produced in a very short time from a small amount of meristematic tissue.

Growth of dates from offshoots is, however, the preferred option for the native grower and has almost completely replaced the old fashioned growth from seeds. Adult date palms in their early years produce many offshoots, and these offshoots are removed from the parent plant and transplanted about 9.2 m apart. The young transplanted tree needs protective wrapping or shelter against cold, heat, and wind, and should be watered daily during the first week and once weekly thereafter until the roots have fully developed. In Saudi Arabia, it is common practice for whole mature trees to be dug out and transplanted as a 'quick fix.' Growth from seeds is still used in a few areas for pollen production. Date plantations should be irrigated and the soil fertilized. Intercropping of date trees with citrus and other fruit plants is common practice, and alfalfa (*Medicago sativa*) is often grown.

Pollination, Fruit Formation, and Development

It is estimated that one male plant is sufficient to pollinate 50–100 female date trees, and it is common practice to plant 2.4 male plants per hectare. In Saudi Arabia, dates usually flower between February and April, and dates are pollinated during these months. Date pollen can be stored at various temperatures for differing lengths of time without much loss of viability or ability to fertilize female flowers. Dates should be naturally wind-pollinated, but there is a small contribution from visiting insects and birds. However, the native form of pollination is to insert one to three branches of male inflorescence in the fully exposed female inflorescence (*See* **Figure 2a**). In order to effect efficient fertilization and to maximize fruit yield, this method has largely been replaced by hand pollination, in which pollen is placed in porous sacs or dusters and spread as dust on female flowers or by mechanical spraying. A temperature of 35 °C is optimal for pollen germination. After successful pollination and fertilization, one of the three ovules of the ovary develops into a one-seeded drupe with a fleshy inner tissue and an outer shiny chitinous skin.

The fertilized rudimentary fruit is referred to in Arabic as 'hababauk.' Dates take about 200 days to reach maturity. After pollination, the fruit passes through four distinct phases, known in Arabic as 'kimri,' 'khalaal,' 'rutab,' and 'tamr' stages. The kimri stage is the time from which the rudimentary seed is formed until it is about 17–19 weeks old. The fruit increases rapidly in size and weight, and in its reducing sugar, acid, and moisture content. This ends the time of botanical maturity, and the fruit turns yellow or red, depending on the cultivar. The khalaal stage is the second stage of development (from 19 to 25 weeks). Weight gain is slow, but sucrose content increases, whereas the moisture content decreases. Tannins start to precipitate, leading to a reduction in astringency, making some varieties palatable and commercially mature at this stage. At the rutab stage, which sets in from around 26 to 28 weeks, there is further moisture loss, and in some varieties, sucrose turns into inverted sugar, and there is browning and skin softening. The moisture content decreases to about 35%, and this is the time during which fresh soft dates are sold. The final stage, tamr, is used only to refer to dates that are left on the palm trees to undergo further ripening, and some dates become self-preserving at this stage.

World Date Production

It is estimated that there are about 90 million date palms in the world. One United Nations Food and Agricultural Organization statistical report names 31 countries as the main date-producing countries. Out of these, only 21 countries export any appreciable amount of their produce. However, 72 countries throughout the world import between 11 and over 500 000 tonnes of dates annually. In the year 2000, the world date production totalled 5.21 million tonnes. Date production figures for each of the 20 leading date-producing countries (except India) are shown in **Figure 5**.

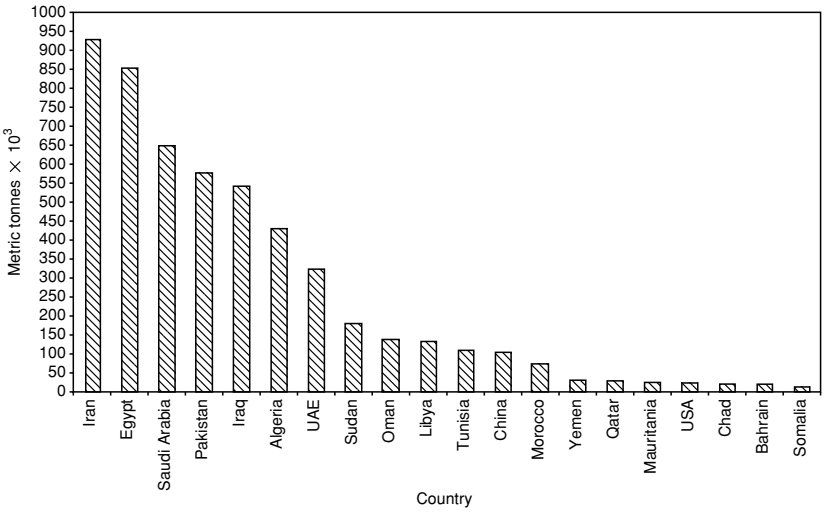

Figure 5 World date-production figures (shown as histograms) for the 20 date-producing countries (figures for India not provided).

Pests, Diseases, and Losses

Dates are afflicted by a number of pests and diseases, which result in substantial losses. Even though new diseases keep emerging, the most important and widespread of these is Bayoud disease, a fungal wilt disease of date trees caused by the fungus *Fusarium oxysporium* f. sp. *albenidis*. Plants infected by this and other wilt fungi make way for opportunistic microbes and pests to cause more damage. Some newly emerging diseases are fungal rhizosis (*Ceratocystis radicicola*) in South Africa and *Serenomyces* in California, and two new pathogens, *Phomopsis phoenicicola* and *Fusarium equiseti*, have been reported from Iraq. The most recently reported diseases (December 2000) are the phytoplasma 'white tip die-back' in Northern Sudan and the 'slow decline' disease in North Africa.

Dates can also be attacked by a range of pests: the scale insect is important in the Dangola District of Northern Sudan, and termite damage is common. Pakistani dates are mainly attacked by borers (*Coleoptera* spp.) and the red palm weevil, *Rhynchosporus ferrugineus*. Recently, this weevil has spread to Israel, Jordan, the Palestinian Authority, and Saudi Arabia. An isolated report from Oman implicates the Indian palm weevil (*Pseudophilus testaceus*) as the cause of some date tree infestations.

Ants and locusts are troublesome, and mice and other rodents feed on leaves, tender roots, and fruits. Damage to fruit through direct infestation by fungi results in the production of phytoalexins, and aflatoxin contamination causing severe fungitoxicity during packaging or marketing has been reported in the cultivars Lulu and Naghal. A list of other disease causing agents, pests, and some of the resultant symptoms is given in Table 1.

Table 1 Diseases of the date palm

Agent	*Symptom/disease*
Alternaria spp.	Leaf spot
Alternaria citri	Brown spot of fruit
Alternaria stemphylioides	Fruit rot
Aspergillus niger	Calyx end rot
Auerswaldia palmicola	Leaf infection
Catenularia fuligenea	Fruit rot
Ceratostomellara dicicola	Root rot
Colletotrichum gloesporoides	Leaf infection
Diplodia phoenicum	Leaf stalk rot, fruit rot, and shoot blight
Endoconidiophora paradoxa	Black scorch and heart bud rot
Fusarium spp.	Inflorescence blight and fruit rot
Graphiola phoenicis	Leaf spot and false smut
Meliola furcata	Black mildew
Meloidogyne spp.	Root knot nematode
Omphalia pigmentata	Decline disease
Omphalia tralucida	Decline disease
Penicillium roseum	Fruit rot
Pestalotia spp.	Leaf spot
Phomopsis phoenicola	Fruit rot
Phymatotrichum omnivorum	Saprophytic contaminant on roots
Pleospora herbarum	Fruit rot
Poria spp.	Wood rot

Harvesting, Packaging, and Storage

Under optimum conditions, dates start fruiting in 3–5 years of planting an offshoot and reach full maturity in 12 years. Dates are usually harvested from July to September, and the estimated yield of dates per bunch is about 5–15 kg. After harvest, dates are packed or transported loose to the market, packing, or processing plants. The majority of techniques that are applied by commercial growers to prevent infestation and spoilage are fumigation, heat treatment, chilled storage, and irradiation. The most

popular fumigant is hydrogen phosphide. Methyl bromide was also commonly used, but, owing to its toxicity to man and animals, it has been banned in most parts of the world. Heat treatment is an effective method for disease and spoilage control, because a 100% mortality can be achieved at all stages of the life cycle of some insects. Refrigeration at 4 °C is used to prolong the storage life of dates, making it possible to keep dates of 20% moisture content for a year without any changes in appearance and taste. Analyses have confirmed that glucose, fructose, sucrose, and protein contents are the same in irradiated as in nonirradiated dates. Combined treatments seem to be the best choice for overall pest and microbial control.

Quality Control Procedures in Date Production

In organized date-handling companies, dates are sorted and cleaned to achieve homogeneity. Different date cultivars ripen and mature at different times and are therefore picked at different times. Dates are often delivered to the processing or packing plants as a mixture, where they are sorted, cleaned, and graded into cull dates, first-, second- and third-grade dates. Cull dates are used for animal feed, whereas first- and second-class dates are packaged, and third-grade dates are used for commercial date products. After sorting and grading, dates are dry-cleaned by moving them over damp towelling, on mechanical shakers, or on rotating cylinders. They are then air-dried and, depending on their use, are sent for further processing, such as dehydration, hydration, glazing, coating, and pitting.

General Uses

There is an old Arab adage that 'there are as many uses for dates as there are days in the year'. The date palm is primarily cultivated for its fruit, which is eaten fresh, dry, or variously processed. It is a high-energy food, containing mainly sugar. In areas where dates are a staple food, the variety of date products includes date honey, date sugar, and date wine, which are made from fresh date fruit juice. A date drink called 'nabidh' is prepared by soaking mashed up dates in water for one night. The traditional use of date syrup in some Islamic countries used to be mainly on a specific occasion: it would be poured on fermented dough (asseda) and eaten on the occasion of the Prophet's birthday. Early records indicate that date syrup has been an overland export material between the Middle East and China for many years, and it is widely used on breads, in softening and preserving dates in jars, and in home-made confectionery. In recent years, date syrup has been used as a base for beverages and bakery products, and in icecreams. Even though attempts to use it in the preparation of soft drinks have proved difficult, a carbonated date syrup-based soft drink of Swiss origin (Tamra) is marketed in Saudi Arabia. Dates are customarily cooked and packaged as candies for local consumption or for export.

Date palm flour is made from the pith of the tree, young leaves are cooked as vegetables, and the terminal bud and palm heart are eaten as salad. Oil is made from the seeds (pits or kernel), which is mainly used for making soap. It was believed in ancient Egypt that when a woman of childbearing age ate two date seeds, she would be childless for several years. The kernel can be ground and used as animal feed. Crystallized inverted sugar from dates is sold in certain date-producing countries, and the seeds can be strung and used as beads. In medieval times, the palm was believed to prevent sunstroke, avert lightning, cure fevers, and drive away mice and fleas. Several other uses have been recorded in folk medicine, and it has been observed that there is a very low incidence of cancers in the Bedouins, attributable to their heavy dependence on dates as a staple food.

Mats, hats, and baskets are woven from date leaves, and the fibers are used for ropes and thread and for rigging boats. In traditional Arab villages, the leaves are used for roofing, and the midribs are used for fencing. The leaves are used as manure, and shredded palm leaves serve as a good substrate for the growth of tomatoes and cucumber. The sap and inflorescence (mostly from senile plants) after fermentation are a good source of palm wine or toddy (alcoholic content 5%), which is either drunk in its natural state or made into liquor. Arrak is an example of an alcoholic beverage distilled from dates. The wood from the trunk is used as firewood, timber for doors, beams, camel saddles, roofing rafters, cords, ropes, baskets, crates, and furniture.

Chemistry of the Different Parts of the Date Palm

Dates, like many plants, are a source of different chemicals. One unexpected compound in date tissue is serotonin, with levels of up to 850 μg per 100 g. Date leaves contain coumarin, luteolin-7-glucoside, luteolin-7-rutinoside, and glycosylapigenin. The main minerals in the leaves and petioles are nitrogen, phosphorus, potassium, and ash.

Date pollen mainly contains cholesterol, rutin, carotenoids, and estrone, and is known to exert

gonadotrophic activity in rats. Pollen mixed with honey is eaten as an aphrodisiac. The main hemicellulose in date pollen is made up of arabinose (46%), galactose (25%), xylose (18%), rhamnose (9%), and 2% uronic acid.

Dry date fruit pulp has an average moisture content of 22.5% with an energy yield of 1151 kJ. It contains 1.9–2% protein, 0.45% lipids, 73.5% carbohydrates, and 7.5% fiber. Its mineral content, in decreasing order, is: potassium > phosphorus > magnesium > calcium > sodium > selenium > iron, but it also contains trace amounts of zinc, copper, and magnesium. Workers in some parts of the world have reported similar chemical contents, but in India, a small amount of arsenic has been detected. Other constituents in the pulp are leucanthocyanins, pipecolic acid, tannins as well as baikiain, and the sarocarp has been shown to contain cholesterol, campesterol, stigmasterol, β-sitosterol, and isofucosterol. Date fruit pulp contains some vitamins: vitamin A, folate, and niacin are in appreciable quantities, but riboflavin, palmitic acid, vitamins B_6, and E, and thiamine are present in trace quantities. (*See* **Vitamins**: Overview.)

The lipid content of date pulp can be as low as 0.2%. It is higher in monounsaturated 16:1 and 18:1 fatty acids than the polyunsaturated fatty acids of 18:2 and 18:3 species. The low protein content of dates is reflected in its poor amino acid content: apart from glutamic acid, aspartic acid, proline, and glycine, which are present at 213, 126, 106, and 95 mg per 100 g, respectively, of date pulp, it contains only trace amounts of most of the other essential amino acids. Some workers have reported high amounts of arginine, but generally, tryptophan, isoleucine and lysine are absent.

Date pits (seeds or kernel) make up between 6 and 12% of the total weight of date fruit and have a moisture content of 5–10%. Other constituents of pits are 7% proteins, 10% oils, 10–20% crude fiber, 55–65% carbohydrates and 1–2% ash. Date pits are relatively high in fatty acids, oleic/linoleic acids, (mostly c. 44–52%) and contain appreciable amounts of tannic acid, 17–24% lauric acid, about 10% each of myristic acid, palmitic acids, and linoleic acid, and trace amounts of stearate, capric, and caprylic acids. At the tamr stage, the pit contains cholesterol, estrones, polysaccharide A (with D-mannose and D-galactose at a ratio of 10:1), polysaccharide B, and xylose. (*See* **Carbohydrates**: Classification and Properties; **Fatty Acids**: Properties.)

Of the essential amino acids, glutamic acid, aspartic acid, and arginine account for about half of the total proteins in date pits, whereas tryptophan is the most limiting, followed by isoleucine and lysine. The oil extracted from date pits has some uses and has the following physico-chemical properties. It has a specific gravity of 0.9207 (at 15 °C), a refractive index of 1.4580 (at 40 °C), an iodine value of 50–55, and a saponification value of 205–210. Date pits also contain a few minerals. These are, in decreasing order: potassium > phosphorus > magnesium > calcium > sodium. The only important microelements detectable in date seeds are iron, manganese, zinc, and copper. From a general human nutrition point of view, date pits are unimportant but have one asset, a high digestibility. The only known human food is the Bedouin dish called 'canua,' which is made from roasted date seeds. However, date seeds are an invaluable source as a main or supplementary diet for animals. It has been suggested that date seed meal can replace barley in chick rations. Ground date seeds can replace 20–75% of ruminant rations, and 75.5% of a wheat bran–barley mixture can be used for feeding carp. Traditionally, date pits are used in feeding pigs, sheep, and camels, for fattening cattle, and for blending in chicken feed. (*See* **Amino Acids**: Properties and Occurrence.)

Commercially Available Date-Derived Products

The only time that dates are used for derived products is when there is a temporary surplus, when dates are damaged, or when dates are undersize or unattractive. Date products can be classified under: sweets, preserves, condiments, breakfast foods, and desserts. Dates are used to a larger extent in the bakery and confectionery industry than they are credited for, probably due to poor or cryptic labeling of such products. It is estimated that over 500 biscuits, cakes, cookies, cereals, toffees, chocolates, breads, and other products that contain date products or are predominantly made from dates can be found in supermarkets, health food stores, and other outlets around the world. One date merchant alone in Saudi Arabia stocks over 110 pastries, and each of these items contains dates or is decorated with dates or date syrup.

Figure 6 shows a basket, woven from date fiber, filled with a decoration of 110 different confectionery products, dates from 14 different dry date cultivars, a background of semisoft brown dates, date leaves, and bunches of mature dates.

Dates are a common supplement in a number of sauces, chutneys, jams, sauces, preserves, and cereals in supermarkets throughout the world. Major worldwide commercial food manufacturers such as Shields Dessert Products, Sun Country, Quaker, Kellogg's, Pillsbury, Betty Crocker, Bakery Wagon, IBSO, Edfina, HP, Crosse & Blackwells, and Shaws use dates in a number of their products. Cereals containing dates

Figure 6 (see color plate 45) Basket, woven from date fiber and filled with 110 different confectionery products made from dates and date syrup (and dates from 14 different dry date cultivars). The background shows semisoft brown dates, date leaves, and bunches of mature dates.

are mostly combined with oatmeal, raisins, walnuts, sultana, and almonds, and products with these combinations have energy yields of between 116 kcal (485 kJ) and 360 kcal (1527 kJ) per 100 g.

Date juice, presscake, and date syrup are the most common commercially derived date products. The raw juice is the main product, and presscake is the byproduct. Date syrup is obtained after evaporation of date juice. Analysis of date syrup from Iraq and Libya reveals an average of 76% total solids, 69–72% inverted sugar, 0.9–3.2% sucrose, 0.9–1.2% proteins, 0.06–0.17% tannin, 0.25–0.60% total pectin (as impure calcium pectate), 0.46–0.76 citric acid, and 1.6–1.8% ash, and a pH of between 4.14 and 4.66. A number of products like date spread, date syrup, caramel, and liquid sugar are made from date juice concentrates. The fruit pulp is used in semisolid infant foods mixed with whey and skimmed milk, and date liquid sugar is used by some manufacturers as a replacement for sucrose in the white layer in cakes. Some low-calorie sweeteners in soft drinks are prepared from dates.

Fermentation Products

Dates should theoretically be used as a main nutrient for microbial conversion to produce a range of useful fermentation products. Unfortunately, unlike beet or cane sugar, date fruit is not a typical sugar crop because most date cultivars contain high levels of inverted sugar. At the consumable stage, most of the sucrose has been inverted by the enzyme invertase into glucose and fructose. In addition, dates are expensive and have a high tannin content that restricts its use as a cheap fermentation product. (*See* **Fermented Foods:** Origins and Applications.)

Dates have, however, been used in the production of a number of products. Lipids have been successfully produced using the fat-producing fungi *Penicillium lilacinum, P. soppi*, and *Aspergillus nidulans*. *Aspergillus niger* has been used for the production of citric acid from date syrup and molasses, and fodder yeast has been produced from Iraqi dates. Postextraction waste material is used as a cultural medium for *Candida utilis* and *Trichoderma viride*, and date seeds are also used to produce citric acid from *Candida lipolytica*.

Dates were used as a reinforcement and flavoring agent in beer-making in ancient Egypt, and references to date wine can be found in early history. What is not clear in these old records is whether the wine referred to was fermented or not. Dates are generally not suitable for wine-making because they lack natural acidity, flavor, and astringency. However, in Iraq, good-quality light-colored dry wines have been produced from zahdi dates, and a darker sweeter wine is fermented from the date syrup, dibis. Pure alcohol can be produced from the fermented mother liquor by distillation, and a so-called sherry can be obtained from the transitional product between wine and the

distilled pure alcohol. Vinegar (*vin-aigre*, from the French meaning 'sour wine') or acetic acid can also be made from date wines that turn sour.

A number of organic acids can be produced either directly or indirectly from date sugars. Single-cell protein from date products has been the focus of investigations for a number of years, and promising results have indicated that bakers' yeast can be produced on date extracts in the same way as it is traditionally produced on molasses.

Allergenicity of Date Fruit and Pollen

Numerous epidemiological studies have reported a steady and significant increase in the world-wide prevalence of allergic diseases in the last two decades. In Saudi Arabia and most of the Middle Eastern countries, it is common knowledge amongst the inhabitants that date pollen causes asthma and allergic rhinitis, and some of the fruits, on ingestion, can cause itching of the mouth and throat. In spite of these observations, there have been no published references on date palm allergens. Results of pioneering studies carried out at the King Faisal Specialist Hospital and Research Centre in Riyadh, Saudi Arabia on the allergenicity of date pollen and fruits in Saudi Arabia have confirmed that date pollen and fruits of some cultivars are allergenic. In the allergic population studied, 25% were allergic to the pollen and 13% to the fruits. These results have been confirmed by workers in Spain, Israel, Kuwait, and France, and cases of anaphylactic shock have been reported. Six pollen components of ≤ 12, 14, 27–30, 37–40, 57, and 65–67 kDa, three fruit components of ≤ 14.3, 27–33, and 54–58 kDa have been found to be the major date allergens. These allergens have been found to share cross-reactive epitopes with a number of foods that have been implicated in the oral allergy syndrome. These allergenic peptides are novel and are unique to some date pollen and fruits cultivars only. However, amino acid sequencing has indicated that date allergens albeit novel, have similar isoelectric points (p*I*s) to, and share sequence homologies with, known allergenic structures, namely profilins, wound-induced proteins, cross-reactive carbohydrate determinants, pathogenicity-related, acidic class III chitinase, and endochitinase proteins.

General Comments and Future Prospects

Clearly, the date palm is an important plant, and date-derived products used in foods are more common the world over than had been known. It is also clear from the production, export, and import data that there is an imbalance in production, consumption, and export. The need to increase date production to meet increasing demand is indicated. Date fruit cultivars that have a high inverted sugar content can be developed further to yield more fructose at their inverted stage. Since dates have a high potassium content, and some hypertensive patients who take diuretics need potassium supplements in their diet, cultivars that have a high potassium, high fructose, and low glucose content can be bred to help with the dietary management of diabetes mellitus. Even though the palm has been known for many years to be one of the most disease-resistant and hardy plants, it appears from the range of infectious agents listed here that the date palm is under threat from new and old pathogens. More effort should be concentrated on breeding high-yielding, disease-resistant, and hypoallergenic varieties. Phytopathological research needs to be focused on, and all aspects of date production improved and diversified.

See also: **Aflatoxins**; **Allergens**; **Amino Acids**: Properties and Occurrence; **Arsenic**: Properties and Determination; **Dehydration**; **Dietary Fiber**: Properties and Sources; **Fermented Foods**: Origins and Applications; **Hemicelluloses**; **Insect Pests**: Insects and Related Pests; **Irradiation of Foods**: Basic Principles; **Jams and Preserves**: Methods of Manufacture; **Syrups**

Further Reading

Al-Ogaidi HKH, Al-Janabi SM and Al-Shakhli IM (1987) Production of carbonated beverage using date juice. Effect of gelatine, tannin and pH on clarification of date juice (Arabic). *Date Palm Journal* 5: 2.

Al-Rawi N, Narkakis P and Beuer DH (1967) Amino acid composition of Iraqi dates. *Journal of the Science of Food and Agriculture* 18: 1–2.

Barett HC (1973) Date breeding and improvement in North America. *Fruit Varieties Journal* 27: 50–55.

Duke JA (1978) The quest for tolerant germplasm. In: *ASA Special Symposium 32, Crop Tolerance to Suboptimal Land Conditions*, pp. 1–61. Madison, WI: American Society of Agronomy.

Khaliel AS and Abou HAN (1985) Formation of vesicular–arbuscular mycorrhizae in *Phoenix dactylifera*, cultivated in Qasim region, Saudi Arabia. *Pakistan Journal of Botany* 17: 267–270.

Knight R Jr (1980) Origin and world importance of tropical and subtropical fruit crops. In: Nagy S and Shaw PE (eds) *Tropical and Subtropical Fruits*, pp. 1–120. Westport, CT: AVI.

Krueger RR (1995) Mystique of the date palm links old and new worlds. *Diversity* 11: 128–129.

Kwaasi AAA, Harfi HA, Parhar RS *et al.* (2000) Cultivar-specific IgE-epitopes in date (*Phoenix dactylifera L.*) fruit allergy. *International Archive of Allergy Immunology* 123: 137–144.

Kwaasi AAA, Parhar RS, Tipirneni P, Harfi H and Al-Sedairy ST (1993) Major allergens of date palm (*Phoenix dactylifera L.*) pollen. Identification of IgE-binding components by ELISA and immunoblot analysis. *Allergy* 48: 511–518.

Kwaasi AAA, Harfi HA, Parhar RS *et al*. (2002) Cross-reactivities between Date Palm (*Phoenix dactylifera* L.) peptides and foods implicated in the oral allergy syndrome (OAS). *Allergy* 57(6): 508–518.

Riedel HR (1986) Dates, where they come from and how they can be used in the confectionery industry (Ge). *Kakao und Zucker* 38: 12.

Sachde AG, Al-Kiasy AM and Norris RAK (1981) A study of the possibility of producing quality wines from some commercial varieties of Iraqi dates. *Mesopotamia Journal of Agriculture* 16(1): 483.

Salim N (1976) *Fructose Increase from Dates and its Separation. Report PDRC-TB-5/76*. Palm and Dates Research Centre, Baghdad, Iraq. Report PDRC-TB-5/76.

Samarawira I (1983) Date palm (*Phœnix dactylifera*). Potential source of refined sugar. *Economic Botany* 37: 181–186.

Shurafa El- MY, Ahmed HS and Abou-Naji SE (1982) Organic and inorganic constituents of date palm pit (seed) in Libya. *Date Palm Journal* 1(2): 537.

Tisserat BH (1979) Tissue culture of date palm. *Journal of Heredity* 70: 221–222.

Degradation *See* **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage; Fungi in Food – An Overview; Molds in Spoilage; Yeasts in Spoilage; **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability; Shelf-life Testing

DEHYDRATION

S P Allison, University Hospital, Nottingham, UK

Definition

Dehydration strictly means lack of water but has come to be used more colloquially to mean salt and water depletion and, even more loosely, to mean a deficit in body fluid of any kind. It is unfortunate that such loose terminology leads to loose thinking, imprecise diagnosis, and, in some cases, inappropriate treatment. In contrast, the terms 'water depletion,' 'salt and water depletion,' and 'plasma volume deficit,' have more precise diagnostic meaning and imply the appropriate treatment for their correction. Pure water depletion is in fact rare. The most obvious example is diabetes insipidus, caused by a lack of antidiuretic hormone. In contrast, salt and water depletion, or desalination, is extremely common. A prime example is that caused by diarrheal diseases, which kills millions of children throughout the world each year.

Physiology

The loss and gain of fluid by the body to and from its surroundings are normally small compared with the enormous daily flux between the different compartments within the body. Assuming the maximal concentrating ability of healthy kidneys, the volume of urine required to excrete waste products may be as little as 500 ml, although it is safer to allow 1–1.5 l. Insensible loss by evaporation from the lungs may be 500 ml, and from the skin surface 400 ml, although this may be greatly increased in hot climates or during fever, owing to sweating. The amount lost from the gastrointestinal tract as feces is usually as little as 100 ml, but in the presence of diarrheal disease, this may increase enormously. Thus, the minimal daily input of fluid to maintain balance in a healthy adult in a temperate climate is approximately 1500 ml, although 2 l provides a safer margin. In warmer climates this requirement is increased considerably. (*See* **Renal Function and Disorders**: Kidney: Structure and Function.)

The distribution of body fluids in an adult between different compartments is shown in **Figure 1**. Total body water is approximately 60% of the bodyweight (more in small children, less in the obese). The intracellular fluid (ICF) is just over 40% of body weight and extracellular fluid (ECF) just under 20% of bodyweight; 25% of the ECF is within the intravascular space, i.e., plasma; the remaining 75% is termed the 'interstitial fluid' and surrounds the cells. The integrity of these body fluid compartments is maintained by the properties of their separating membranes (e.g.,

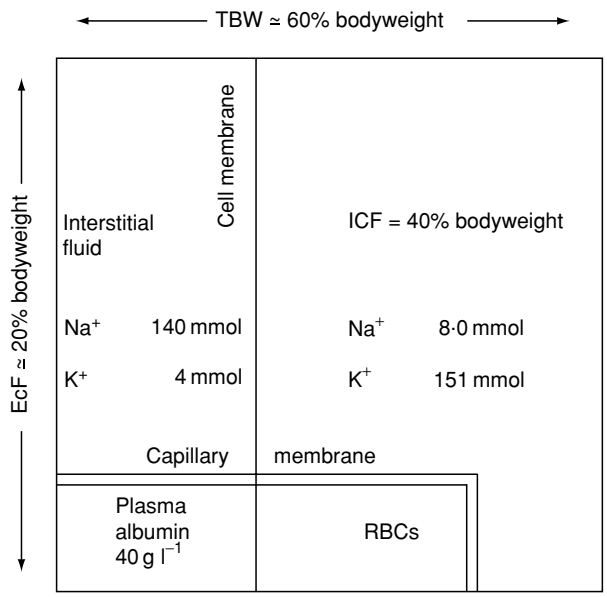

Figure 1 Diagrammatic representation of body fluid compartments and concentrations of sodium ions (Na^+), potassium ions (K^+) and albumin. TBW, total body water; ICF, intracellular fluid; ECF, extracellular fluid; RBCs, red blood cells. Reproduced from Dehydration, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

the cell membrane with its sodium pump, and the capillary membrane and its pore size) and their osmotic or oncotic content (e.g., sodium ions Na^+, for ECF; potassium ions, K^+, for ICF; albumin for plasma). Not only is the external fluid balance important, but also changes in fluid compartments, where shifts may be large. For example, the rate of exchange of albumin between the intravascular and interstitial spaces is normally 10 times its rate of synthesis. Fluid deficits may also have different effects on each compartment. In an average adult, a deficit of 2 l of pure water reduces the total body water evenly by 2 l (and, incidentally, the weight by 2 kg) with a 1400 ml reduction in ICF, a 600-ml reduction in ECF fluid, and a 150-ml reduction in plasma volume. However, a loss of 2 l of salt and water, at a concentration of 140 mmol Na^+ per liter, reduces the total body water by 2 l and the weight by 2 kg but causes a selective loss of 2 l from the ECF, with a consequent reduction in plasma volume of nearly 500 ml. The clinical problems presented by these two situations are of an entirely different order. (*See* **Potassium**: Physiology; **Sodium**: Physiology; **Water**: Physiology.)

To understand the problems of fluid deficit, it is also necessary to consider the large daily volume of secretions, particularly those into the gastrointestinal tract (Table 1). Under normal circumstances, these are almost entirely reabsorbed, with a minimal daily loss of 100 ml in the feces. With gastrointestinal disease, e.g., vomiting, diarrhea or fistula, large volumes of water and electrolytes may be lost to the outside or, in the case of intestinal obstruction or ileus, pooled in the gut and lost to the functional compartments. In an adult, during a normal day, 8000 ml of fluid enter the small bowel; 500 ml as saliva, 1200 ml as gastric fluid, 600 ml as bile, 1200 ml as pancreatic juice, 2000 ml as small bowel secretion, and the remainder as ingested fluid. All but 1.5 l is reabsorbed in the small intestine, and a further 1.4 l is reabsorbed in the colon with much of the remaining sodium chloride. Approximate electrolyte content of gastrointestinal secretions is shown in Table 1.

Table 1 Approximate electrolyte content of gastrointestinal secretions

Secretion	*Sodium (Na^+) ($mmol\,l^{-1}$)*	*Potassium (K^+) ($mmol\,l^{-1}$)*	*Chloride (Cl^-) ($mmol\,l^{-1}$)*
Saliva	44	20	
Gastric	70–120	10	100
Small intestine	110–120	5–10	105
Bile	140	5	100
Pancreas	140	5	75
Diarrhea			
Cholera (adult)	120	15	90
Cholera (child)	100	27	90
Nonspecific (child)	56	25	55

Diagnosis and Monitoring

In common with most diagnoses, that of dehydration depends mainly on the history and examination. The history, combined with a knowledge of the natural history of the presenting condition, provides the main clues to the presence of salt and water depletion, the likely progress of the condition and the necessary treatment.

The physical signs are helpful, but none are pathognomonic. The facies and eyes appear sunken, but this is also seen in cachexia. The mouth and tongue are dry, but this can also be caused by mouth breathing. The skin turgor, assessed by pinching up a fold of skin and observing its recoil, is diminished with salt depletion but also in old age, cachexia, and extreme cold. A marked fall in blood pressure between the lying and sitting or standing positions is also characteristic of salt and water depletion, as is tachycardia. A low output of concentrated urine, associated with a rise in blood urea owing to prerenal failure, is also seen.

Other biochemical indices are less useful in diagnosis. The plasma sodium concentration gives no clue to the total body sodium content, unless water balance is also known. It merely reflects the relative proportion

of sodium and water in the extracellular space. Thus, with pure water lack, the plasma sodium concentration increases. With a mixed deficit of sodium and water, the plasma sodium reflects the relative proportion of the two that are lost.

Progress is first monitored by observing the improvement in the appearance of the patient; guard against fluid overload by watching for edema, a raised jugular venous pressure, or pulmonary edema. A fluid balance chart is valuable, but water balance is best measured by changes in weight, which should be monitored at least daily. Plasma sodium concentration changes can then be interpreted in the light of water balance.

Conditions Associated with Water and Salt Depletion

Diarrhea

Worldwide, diarrhea is by far the most important numerically, causing large numbers of childhood deaths from dehydration. Its severity varies from cholera, in which the stool output may be up to $1\,l\,h^{-1}$ (for electrolyte losses, see Table 1), to other infective conditions in which stool losses are smaller, although potentially life-threatening. Treatment is usually by oral rehydration therapy (see below), although severe cases require intravenous or subcutaneous fluid replacement, where logistically possible. Formulae are shown in Table 2 and are based on the principle that small bowel sodium absorption is linked to that of glucose. The presence of starch, polysaccharide, or glucose in such formulae is vital. Amino acids and peptides also enhance sodium reabsorption.

Vomiting or Nasogastric Aspiration

An examination of Table 1 reveals the likely loss of electrolytes from vomiting or nasogastric aspiration and the logical replacement requirements. Aspirate volumes must be measured and replaced appropriately.

Intestinal Fistula or Short Bowel Disease

Up to 4 l per day may be lost from an upper jejunal or duodenal fistula. Parenteral nutrition is almost mandatory in this situation and must be combined with adequate salt and water replacement guided by measured losses, daily weight changes, plasma electrolyte measurements, and a knowledge of electrolyte concentrations in the fluid lost. Lower bowel fistula or short bowel can be treated by the enteral route, enhancing water and salt absorption using principles similar to those used in oral rehydration therapy. If enteral tube feeding is employed, a feed with sodium content less than 90 mmol causes net sodium secretion into the jejunum, whereas levels greater than this are associated with net absorption. The presence of carbohydrate and protein ensures the additional effect of glucose and amino acids on sodium absorption. The rate of administration may also be important. Too rapid rates may overwhelm the capacity of a shortened bowel to absorb all the fluid presented to it. The total osmolar content should also be considered, although the number of osmoles administered per unit time is more important than concentration.

Metabolic and Renal Disorders

The hyperglycemia of diabetes mellitus causes an osmotic diuresis, leading to large deficits of water, sodium and potassium during acute loss of control, e.g., diabetic ketoacidosis. An osmotic diuresis may also result from excessive urea production owing to excessive protein administration. Hypercalcemia poisons distal tubular function, leading to excessive production of dilute urine. Diabetes insipidus owing to a lack of antidiuretic hormone (ADH) has already been referred to. Alcohol excess dehydrates through inhibition of ADH. Acute or chronic renal damage may be associated with tubular dysfunction, loss of renal concentrating capacity, and water and salt deficit. This may be observed most strikingly in the polyuric phase of recovery from acute renal failure or following relief of obstructive uropathy. Treatment consists of the appropriate administration of water and electrolytes either orally or intravenously.

Excessive Use of Diuretics

Excessive use of diuretics is a common cause of water, sodium, and potassium depletion in the elderly and results from regimens that are too rigid and from lack of patient education. Patients need to know, for example, that if they fall ill from some condition that prevents salt and water intake or leads to excess loss, they should stop their diuretics.

Table 2 Concentration of electrolytes in oral rehydration solutions

	Sodium (Na^+) ($mmol\,l^{-1}$)	*Potassium (K^+) ($mmol\,l^{-1}$)*	*Chloride (Cl^-) ($mmol\,l^{-1}$)*	*Citrate or bicarbonate ($mmol\,l^{-1}$)*	*Glucose[a] ($mmol\,l^{-1}$)*
For mild to moderate diarrhea	35–60	13–25	50	10–20	100–200
For severe diarrhea	90	20	80	10	111

[a]Glucose may be substituted by sucrose, maltose or rice water according to local availability.

Excessive Sweating

Excessive sweating, e.g., in the tropics, merits appropriate replacement. Before adaptation takes place, those going from temperate to hot climates are particularly prone to dehydration and heat stroke. A liberal water intake, combined, if necessary, with additional salt, should be taken prophylactically.

Errors in Diagnosis

Errors in diagnosis are particularly common in already hospitalized patients who are more often the victims of excess fluid administration than of dehydration. This results commonly from a naïve interpretation and reliance on one or more physical signs, e.g., dry mouth, or from a failure to understand the compartmental distribution of body fluids and how these are affected by disease. A proper understanding of simple fluid balance concepts is a prerequisite for managing the many conditions in which dehydration or salt and water depletion are common. Salt and water are also intrinsic components of food and of the regimens used in the nutritional support of the sick.

Oral Rehydration Therapy

Solutions enhance the absorption of water and electrolytes through their substrate content, e.g., glucose, while replacing electrolyte deficit safely, e.g., avoiding hypernatremia in infants. They should be palatable and acceptable to children, readily available and simple to use under all conditions. An alkalizing agent to counter acidosis may be advantageous. Solutions used in the UK for mild to moderate diarrhea contain less sodium and more glucose than the World Health Organization (WHO) formula, which is used for severe diarrhea, e.g., cholera. Dose is according to fluid loss, but a rough guide for adults is 200–400 ml for every loose stool, for children 200 ml per motion, and for infants 1–1.5 times the usual feed. In severe cases, doses may be higher. The WHO policy is to promote its own single oral rehydration solution and to prevent hypernatremia in children by giving extra water between doses. For further details, see *British National Formulary*.

See also: **Potassium**: Physiology; **Renal Function and Disorders**: Kidney: Structure and Function; **Sodium**: Physiology; **Water**: Physiology

Further Reading

Guyton AC (1991) *Textbook of Medical Physiology*. Philadelphia, PA: WB Saunders.

Hladky SB and Rink TJ (1986) Body fluid and kidney physiology. *Physiological Principles in Medicine Series*. London: Edward Arnold.

Turnberg LA (1992) Pathophysiology of secretory diarrhoea. *Hospital Update* 18: 93–104.

Walters G (1983) Disorders of fluid and electrolyte balence. In: Williams DL and Marks V (eds) *Scientific Foundations of Clinical Biochemistry: Biochemistry in Clinical Practice*, pp. 1–23. London: Heinemann.

DENTAL DISEASE

Contents

Structure of Teeth

M E J Curson and M S Duggal, University of Leeds, Leeds, UK

Background

The teeth comprise the hardest structures of the human body and as such, once fully formed, change little throughout life. Their resistance to destruction is such that they play a major role in archeological and forensic studies. Teeth as archeological remains are well known to survive for thousands, if not millions, of years after the death of the animal of which they were a part.

Despite their apparent indestructibility in relation to physical damage, and the ravages of time as skeletal remains, the teeth are susceptible to chemical damage during life by acid dissolution, a process known as dental caries. This disease has occurred in

humans for as long as they have existed. Even in the earliest human skull specimens, the teeth show signs of dental decay, albeit of a limited extent, with only one or two cavities occurring in the mouth of an individual. The susceptibility of the teeth to dental caries has therefore always been present, but the prevalence and incidence of the disease have varied. This will be discussed more fully elsewhere, but an understanding of the dental caries process is based upon a knowledge of the structure of the teeth, which is discussed in this article in overall terms of the teeth and the components that make up a tooth.

Teeth

In all mammals, there are two sets of teeth; in humans, a primary (baby) or deciduous set of 20 teeth is replaced between the ages of 6 and 12 by a secondary or permanent set of up to 32 teeth.

Primary Teeth

The primary teeth are laid down *in utero* and commence calcification at or immediately after birth. The first teeth to erupt are usually the incisors at about 6 months of age, and the final primary teeth to come into the mouth are the second molars between the ages of 2 and 3 years.

The primary teeth have the same basic structure as the permanent teeth, i.e., enamel, dentine, cementum, and pulpal (nerve) tissue. However, the overall morphology of a primary tooth does differ somewhat from the permanent successor. The major difference lies in the crown, particularly of the primary molars, which are more bulbous. The bulbosity is most pronounced on the first primary molar in the lower jaw, or mandible.

The roots of the primary teeth are shorter, smaller, and, in the molar, more curved than the permanent successors. There is therefore less dentine in a primary tooth. As the enamel is also thinner than in permanent teeth, so dental caries progresses through to the nerve faster in primary teeth. Furthermore, as the nerves or pulps of the primary teeth are relatively larger and more prominent, so the risk of abscess, as a result of caries penetrating through to the pulp is that much greater.

Permanent Teeth

In a full set of permanent teeth, there are 32, although not all teeth always develop, so one or several teeth may be missing. The commonest tooth to be absent is the third molar, or wisdom tooth.

The morphology of the permanent teeth is very similar to the primary teeth, but larger and with much bigger roots relative to crown size. The enamel is also much thicker. The dentine comprises the bulk of a permanent tooth and becomes thicker with age. This is because, as teeth age, the wear on the outside of the enamel caused by eating slowly removes the enamel. In wearing away the tooth, the masticated foods produce a response in the nerve, which is slowly reduced as dentine is produced by the odontoblasts. In time, the nerve can shrink to the extent that the pulp chamber almost disappears. Where a coarse, abrasive diet is used, this process of laying down secondary dentine can occur earlier in life and at a much faster rate.

Structure of Teeth

The differences in the structure between primary and permanent teeth are therefore those of size and thickness. The chemical compositions of the enamel, dentine, and cementum of primary and permanent teeth are essentially the same. As the chemistry of their structure affects their subsequent resistance or susceptibility to dental caries, it is appropriate to discuss their structure in detail.

Enamel

Tooth enamel comprises a hard, porcelain-like cap covering the whole of the clinical crown of a primary or permanent tooth. The crown is defined as that part of a tooth that is erupted and is visibly present in the mouth. It is distinct from the anatomic crown. However, with age, as the tooth continues to erupt, the clinical crown becomes the same as the anatomic crown.

The enamel comprises calcium, phosphate, water (as OH) of a form known as hydroxyapatite ($Ca_{10}(PO_4)_6$. 2OH). This basic chemical structure is widespread in the animal kingdom and is the building block of all calcified tissue. Where enamel is unique is that the percentage of inorganic component making up enamel is higher than any other tissue. Analysis of enamel shows that an average of 98% is made up of the inorganic component, largely the hydroxyapatite. (*See* **Calcium**: Physiology.)

Developing enamel, which is of ectodermal origin, is composed of up to 18% calcium and heterogeneous proteins. The majority of the developing enamel matrix consists of amelogenin, which is protein-rich. The remainder of the enamel protein is a high-molecular-weight phosphoprotein, known as enamelin. As the enamel continues to develop, the amelogenin is replaced as the enamel calcifies and then matures. During this process, over 90% of the amelogenin is removed. If this process does not occur properly, e.g., during some metabolic disorders, enamel hypoplasia occurs. At a later stage, if

calcification does not proceed properly during maturation, hypocalcification results. In both instances, the tooth may be more susceptible to dental caries.

Mature enamel consists of enamel crystallites of the aforementioned hydroxyapatite. These crystallites mature rapidly in width but slowly in thickness. The end result is a crystallite of large size.

The crystallinity of enamel varies, depending on its purity and the degree of incorporation of other elements. However, enamel is never a pure substance and always includes other inorganic or organic components during the developmental stage. The main component of interest is fluoride, which may be substituted for part of the hydoxyl ion to form fluorapatite. This is of considerable interest in dentistry, as this change produces an enamel that has a much greater resistance to acid dissolution. The incorporation of fluoride in the enamel will also provide a reservoir of fluoride ion, which plays a role not only in resisting dissolution but more importantly in remineralization.

The crystallinity is affected not only by incorporation of other ions, but also in surface enamel and in physicochemical ionic exchanges. Posteruptively, these changes in enamel chemistry are notably influenced by the uptake and release of fluoride and a reduction in carbonate. The latter is about 2.5% by weight, compared with calcium at 36% and phosphorus at 18%. However, it is felt that the carbonate component, as it is less stable than pure hydroxyapatite, may lead to a disruption of the enamel crystallinity and probably play an important role in enamel maturation. It is felt that as carbonate is more acid-labile than other components, it is lost during early caries or may make a high-carbonate enamel more susceptible to acid dissolution.

Enamel is laid down during development by specialized cells called ameloblasts, and if a higher concentration of fluoride is available in the bloodstream, this trace element becomes incorporated by the ameloblasts in the developing enamel. This mechanism is important when discussing the resistance of enamel to dental caries, but also in the formation of mottled enamel associated with high fluoride intake.

As mentioned above, enamel is laid down by ameloblasts by a process of secretion of an enamel matrix, which then becomes calcified. Further calcium and phosphate salts are laid down as the enamel matures prior to, and soon after, eruption in the mouth. During this period of maturation, the enamel appears to be more susceptible to acid dissolution, demineralization, and the development of carious cavities. The carbonate content of such enamel is high; carbonate is subsequently lost during maturation, but this will be offset by incorporation or the presence of fluoride in the oral environment.

The enamel surface of the outermost layers of 10 μm is of a slightly different histological structure. During the formation of the enamel, the ameloblasts lay down an outer 'surface layer' of different, higher mineral composition. If an increased level of fluoride is incorporated in this layer, that tooth will have a greater resistance to subsequent decay. Even during the process of early demineralization, this 'surface zone' remains intact, probably by a process of continuous remineralization by inorganic ions released from the inner layers of enamel as it dissolves. During the early stages of demineralization (caries), when the inner layers of enamel are dissolved, the outer layer retains its structure but becomes whitish in appearance when seen clinically. This is known as a 'white spot' or 'incipient caries' and is the first clinical sign of decay, but can be reversed by remineralization. The rate at which demineralization occurs will depend not only on the level and frequency of acid attack but also on the resistance of the enamel. This in turn will depend on the degree of mineralization, maturation, and fluoride and carbonate content.

Dentine

The major part of a tooth is dentine. The structure is laid down by cells called odontoblasts, which secrete tubules of dentine from the junction with the enamel to the nerve or pulp of the tooth. The odontoblasts are mesenchymal in origin and the structure they lay down has a lower level of mineralization, as well as smaller crystallites than enamel.

The dentine differs from the enamel in composition, being only 79% inorganic, although this inorganic material is the same hydroxyapatite as in the enamel and resembles bone in composition. The higher organic component consists of collagen type I, making up the walls of the dentine tubules and the dentine process in the centre of the dentinal tubules.

Unlike enamel, dentine has an innervation by means of nerve endings located at the pulpal end of the dentinal tubules. It is for this reason that pain is experienced when dentine is exposed in the mouth. This exposure may occur as part of the carious process when weak or undermined enamel breaks away to form a cavity. Alternatively, when a cavity is drilled by a dentist, in the process of removing decay, the sensitive dentine is cut, also causing pain unless some form of anesthesia is used.

Because of the higher organic content of the dentine, the process of dental decay is somewhat different to that of the enamel. While the caries in enamel is almost entirely a chemical dissolution by acids produced by bacteria, such as *Streptococcus mutans*, a secondary process of proteolytic removal of the collagen occurs in dentinal caries.

Although most cases of dental caries originate with dissolution of the enamel, the structure of the teeth is such that should the gingival tissues or gums recede, the neck of the tooth may be exposed. If this occurs, dentine is quickly exposed, leading to sensitivity of the tooth and, subsequently, dentinal or root caries. As receding gingiva occurs with age, so root caries is associated with the aging dentition and is seen mostly in patients over 50 years of age.

Cementum

The dentine of the root of the tooth is covered by a thin layer of cementum. This structure, also of mesenchymal origin, is again composed of inorganic and organic material but is closer to bone in the proportion of the two components. (*See* **Bone**.)

Cementum provides a mean of bonding the teeth into the bony sockets of the jaws by means of strands of collagen, known as periodontal fibers. The whole root is surrounded by many of these fibers, attaching the tooth to the bone, and providing a system of 'elastic bands,' or a hydraulic system, so that the tooth can be pressed deeper into its socket under biting forces. This mechanism takes up heavy forces on the tooth without them being transferred directly to the bone, which would be very uncomfortable on biting.

The cementum may break down as part of the dental caries process or during gum disease, or be worn away by excessively hard tooth brushing. Because of its lower inorganic content, its resistance to decay or wear is far less than either enamel or dentine, and if exposed, it can wear away quite quickly.

See also: **Bone**; **Calcium**: Physiology

Further Reading

Nikiforuk G (1985) *Understanding Dental Caries: 1. Etiology and Mechanisms. Basic and Clinical Aspects*. Basel: Karger.

Thylstrup A and Fejerskov O (1986) *Textbook of Cariology*, 1st edn. Copenhagen: Munksgaard.

Etiology of Dental Caries

M S Duggal and M E J Curson, University of Leeds, Leeds, UK

Introduction

Dental caries or tooth decay is a pathological process of localized destruction of tooth tissues. It is a form of progressive loss of enamel, dentine, and cementum initiated by microbial activity at the tooth surface. Loss of tooth substances is preceded by a partial dissolution of mineral ahead of the total destruction of the tissue. It is because of this characteristic that dental caries can be differentiated from other destructive processes of the teeth, such as abrasion through mechanical wear and erosion by acid fluids.

Current Concepts of Caries Etiology

There is overwhelming evidence that the initial phase of dental caries involves demineralization of the tooth enamel by localized high concentrations of organic acids. These acids are produced by the fermentation of common dietary carbohydrates by bacteria that accumulate in dental plaque on the teeth. This concept of dental caries etiology is based essentially on the interplay of four principal factors: a susceptible host (the tooth), microflora with an acidogenic potential, a suitable substrate available locally to the pathodontic bacteria (carbohydrate), and a fourth important dimension, time (**Figure 1**). Because of the multiplicity of factors which influence the initiation and progression of dental caries it is often referred to

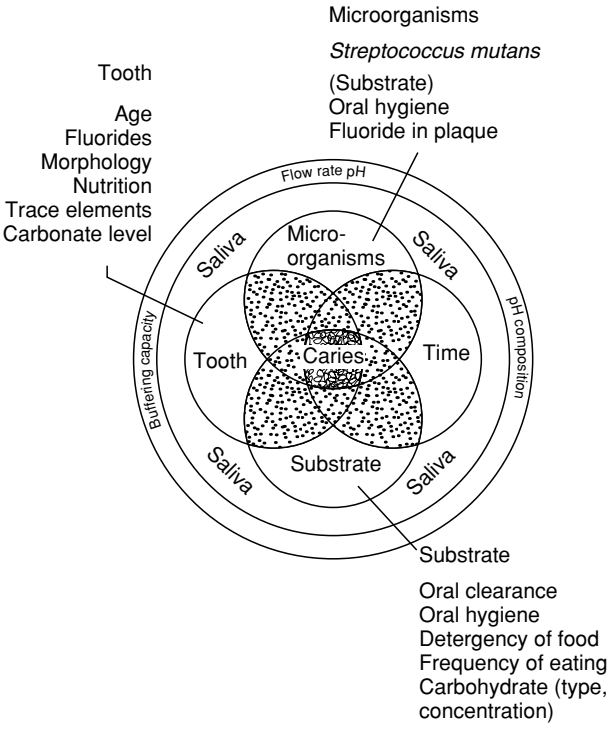

Figure 1 Diagrammatic representation of the parameters involved in the carious process. All factors must be acting concurrently for caries to develop. Dental Disease/Aetiology of Dental Caries, Reproduced from *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

as a disease of a 'multifactorial etiology', an old but appropriate description. In this article the role of microflora and the host susceptibility are discussed. The role of carbohydrates in the diet is discussed later. (*See* **Carbohydrates**: Digestion, Absorption, and Metabolism; Requirements and Dietary Importance.)

The Role of Microorganisms

Considerable difference of opinion exists within the dental research community as to which microorganisms cause dental caries and whether indeed a specific strain of bacteria is responsible in its causation. However, there is evidence that microorganisms, which accumulate in dental masses (dental plaque) on the teeth, cause fermentation of dietary carbohydrates, producing organic acids which cause demineralization of the tooth enamel. This concept is a combination of Miller's acidogenic theory of 1889 and that of dental plaque introduced by Williams in 1897. Subsequent research has supported this original view and the evidence implicating oral microorganisms in the etiology of dental caries is summarized below:

1. Germfree animals do not develop dental caries.
2. Addition of penicillin to the diet of rats prevents caries.
3. Plaque bacteria can demineralize enamel *in vitro* when incubated with dietary carbohydrates.
4. Microorganisms can be found invading carious enamel and dentine.
5. Animals in which diet or dietary carbohydrates are delivered directly into the stomach through a tube do not develop caries.

Role of Dental Plaque

Dental plaque is the name given to the aggregations of bacteria and their products which accumulate on the tooth surface. Plaque collects rapidly in the mouth, although the actual rate of formation varies from one individual to another. When plaque accumulates on the crowns of teeth the natural, smooth, shiny appearance of the enamel is lost and a dull, matt effect is produced. As it builds up, masses of plaque become more readily visible to the naked eye. In direct smears, the early plaque is dominated by cocci and rods, most of which are Gram-positive. In the mature plaque (after about 7 days) the percentage of cocci in the plaque decreases rapidly and filaments and rods constitute about 50% of organisms in plaque. The ecological term for this shift from a predominantly coccal type in the beginning to a predominantly mixed, filamentous flora a few days later is bacterial succession.

There is no doubt that the presence of plaque is required for caries development in humans. The best available evidence suggests that a carious lesion results from the action of bacterial metabolic end products localized to the enamel surface by dental plaque. It is therefore the acidity generated by the microorganisms in the plaque after exposure to carbohydrates that is directly responsible for causing demineralization. This is under the influence of many factors.

Regulators of Plaque Acidity

Various parameters have been studied for their effect on plaque acidity, and hence their effect on caries; these are listed in Table 1. The role of saliva will be discussed later, and carbohydrates as regulators of plaque pH are discussed elsewhere.

Bacterial mass in relation to caries Plaque is required for caries initiation, although the relationship between the plaque quantity and dental caries activity is unclear. Some earlier studies, using a crude methodology, showed a positive correlation between caries increment and oral hygiene, but more controlled investigations failed to establish this finding. Many individuals remain caries-free in spite of an unfavorable oral hygiene. This lack of a clear-cut relationship between amount of plaque and dental caries suggests that plaques vary in their microbial composition and, although present in equal amounts, might vary in their cariogenicity. Impressive evidence indicates that the qualitative nature of the plaque flora determines the metabolism and the potential for caries production. This view has been referred to in the literature as the specific plaque hypothesis.

Bacterial composition of plaque and caries Most published data on gnotobiotic animals have concentrated on the cariogenic potential of streptococci, especially *Streptococcus mutans*. However, several organisms have been found capable of inducing carious lesions when used as monocontaminants in gnotobiotic rats. These include *S. mutans, S. salivarius, S. sanguis, S. milleri, Lactobacillus acidophilus, Peptostreptococcus intermedius, Actinomyces viscosus*, and *A. naeslundii*. However, not all organisms are equally cariogenic. Table 2 shows the acidogenicity and

Table 1 Factors influencing pH of dental plaque

Carbohydrates	*Saliva*	*Plaque*
Quantity of intake	Flow rate	Bacterial mass
Frequency of intake	Buffer capacity	Bacterial composition
Retentiveness	Other factors, e.g., urea, ammonia	Acid production
Modifying effects of other ingredients of foodstuffs		Acid tolerance Acid retention

Table 2 Acidogenicity and acid tolerance of some plaque bacteria

	pH				Initiation of growth
	< 4.0	4.1–4.5	4.6–5.0	> 5.0	
Streptococcus mutans	+	+			5.2
S. sanguis		+	+		5.2–5.6
S. salivarius	+	+			5.2
S. mitis		+	+		5.2–5.6
Lactobacillus	+	+			< 5.0
Actinomyces			+	+	?
Enterococci		+	+	+	?

Adapted from Van Houte J (1980) Bacterial specificity in the etiology of dental caries. *International Dental Journal* 30: 305–325.

acid tolerance of plaque bacteria. In evaluating data shown in Table 2 it should be noted that the final pH represents the pH observed after incubation of pure bacterial cultures in broth with excess carbohydrates. Table 2 also indicates the tolerance of bacterial systems of plaque to acid. Though these data do not reveal the rate at which acid is produced or the pH at which growth ceases, they give some insight into the relative acidogenic, i.e. cariogenic, potential of plaque bacteria.

Bacterial Specificity and Dental Caries

Investigations into the relationship between various plaque organisms and dental caries have mainly focused on streptococci, lactobacilli, and filamentous bacteria.

Streptococci

Studies of *S. mutans* strongly suggest its active involvement in the initiation and progression of dental caries. Members of this species are nonmotile, catalase-negative, Gram-positive cocci. On *Mitis salivarius* agar they grow as highly convex colonies. When cultured with sucrose they form polysaccharides which are insoluble. This property of forming insoluble polysaccharides from sucrose is regarded as an important characteristic contributing to the caries-inducing properties of the species. *S. mutans* can invariably be isolated from incipient lesions as well as lesions that have cavitated on pits and fissures, approximal and smooth surfaces. It can also be isolated with highest frequency from plaque over carious lesions, and the isolation frequency decreases when the samples are taken further away from the lesions.

S. sanguis is also one of the predominant groups of streptococci which colonize the tooth surface. Evidence suggests that caries from this strain is significantly less extensive than that from *S. mutans* and occurs primarily in pits and fissures, whereas *S. mutans* causes smooth surface caries as well.

It would therefore appear that *S. mutans* is the most cariogenic of the family of streptococci. Its cariogenicity is attributed to several important properties:

1. It synthesizes insoluble polysaccharides from sucrose.
2. It colonizes all tooth surfaces.
3. It is a homofermentative producer of lactic acid.
4. It is more acidogenic than other streptococci.

Lactobacilli

Lactobacilli have also been implicated in the etiology of dental caries for many decades. Although a few studies have suggested their involvement in the initiation of caries, they are most often found in large numbers in carious cavities. This and the fact that lactobacilli have a relatively low affinity for the tooth surface suggest that they might have a more important role to play in the progression rather than initiation of dental caries.

Filamentous Bacteria

Filamentous bacteria, especially actinomycetes, have been found to be associated with root surface caries. *Actinomyces viscosus*, an acidogenic bacterium, is almost always isolated from plaques overlying root lesions. The role of *A. viscosus* in the initiation of root lesions is difficult to assess because they are often predominant on the sound root surfaces in subjects experiencing and resisting root caries. More definitive studies are needed to determine the association of *A. viscosus* and root caries.

It must be remembered that most of the microorganisms implicated in dental caries etiology are indigenous organisms and exist in a dynamic relationship with the host. Disease is not a necessary outcome of this association but it may ensue when the balance is greatly disturbed. With *S. mutans* this may occur, for example, when its numbers on teeth are increased owing to a higher or more frequent consumption of carbohydrates.

The Host Factors

Saliva

Saliva significantly influences the carious process. For example, removal of the salivary glands of hamsters greatly enhanced the caries activity when they were fed a high-sucrose diet. Human studies have demonstrated the same results and have shown that if the exposure of teeth to saliva is restricted by blocking the opening of the salivary glands the pH values of plaque, after exposure to carbohydrates, are lower.

Salivary flow and caries rate Humans suffering from xerostomia or decreased secretion of saliva as a consequence of a pathological condition such as sarcoidosis, Sjögren's syndrome, etc., often experience a higher rate of caries. The tooth destruction as a result of desalivation is typically quite rapid. Rampant caries has been demonstrated in patients who have been treated by radiotherapy to the head and neck as irradiation of the salivary glands leads to decreased salivary flow and usually leads to extensive caries in the cervical region of the teeth. These observations have led to the conclusion that saliva is in some way important to maintain the integrity of the tooth substance.

Salivary buffering and dental caries Saliva contains bicarbonate-carbonic acid and phosphate buffer systems. The buffering capacity acquired by saliva by virtue of these ion systems tends to correct pH changes caused by concentration changes of acidic ions produced by fermentation of carbohydrates. A typical reaction of plaque to carbohydrate challenge is an immediate drop in plaque pH followed by a gradual rise to the resting value. However, if plaque and teeth are isolated from the influence of saliva the pH of plaque drops further and remains low for prolonged periods. Thus the buffering capacity of saliva is an important factor in reducing the time the tooth enamel is exposed to the acidogenic challenge.

Other protective factors in saliva Saliva contains a number of factors of glandular origin, including lysozymes and a peroxidase system. Interest has also been focused on the immunological aspect of caries, and specific immunoglobulin A (IgA) antibodies to *S. mutans* have been detected in saliva by immune assays. The concentration of secretory IgA in whole saliva has been reported to be significantly less in subjects with a high caries rate as compared with those with a lower caries experience. Although substantial evidence suggests that the aforementioned factors form an antibacterial system in saliva, the extent to which they contribute to caries resistance of the host is not yet clear.

Tooth

Tooth morphology and dental caries A susceptible host is one of the factors which is required for caries to occur. The morphology of the tooth has long been known to be one of the major determinants of its susceptibility. Clinical observations have suggested that pits and fissures of posterior teeth are highly susceptible to caries. This is thought to result from impaction of food and microorganisms in the fissures which are also difficult to reach with routine oral hygiene aids. An interesting observation is that there is a positive correlation between the depth of the fissures and caries susceptibility.

It has also been observed that certain surfaces of a tooth are more prone to caries than other surfaces. For example, the occlusal surface and the buccal surface of the lower first permanent molar are more likely to develop caries as compared with the lingual, mesial, or distal surfaces, probably because of the presence of the fissures on the occlusal surfaces and the buccal pit on the buccal surface. Similarly, an intraoral variation in susceptibility to caries exists between different teeth. The caries susceptibility by tooth type, in descending order, is as follows: first permanent molar, second permanent molar, second premolar, upper incisor, first premolar, with the lower incisors and canines least likely to develop caries.

Arch morphology and caries Irregularities in the arch form and imbrication of the teeth also favor the development of caries. It has been shown experimentally that enamel in stagnation areas is more likely to develop demineralization, which is manifested as white spot lesions.

Tooth composition and dental caries It is well known that the surface enamel is more resistant to caries attack than the subsurface. This is attributed to a higher concentration of fluoride and other trace metals, such as zinc and lead, which are thought to protect the surface enamel from demineralization. It is because of this that dental caries is often said to be subsurface in origin. Although there is plenty of evidence to support the direct relationship between the fluoride content of the surface layer of enamel and its resistance to caries attack, the relationship of other trace elements is unclear and still being investigated.

See also: **Carbohydrates**: Digestion, Absorption, and Metabolism; Requirements and Dietary Importance; **Immunology of Food**

Further Reading

Newbrun E (1979) *Cariology*, 2nd edn. Baltimore: Williams and Wilkins.

Silverstone LM, Johnson NW, Hardie JM and Williams RAD (1981) *Dental Caries – Aetiology, Pathology and Prevention*. London: Macmillan.

Thylstrup A and Fejerskov O (1986) *Textbook of Cariology*, 1st edn. Copenhagen: Munksgaard.

Van Houte J (1980) Bacterial specificity in the etiology of dental caries. *International Dental Journal* 30: 305–325.

Role of Diet

M S Duggal and M E J Curson, University of Leeds, Leeds, UK

Introduction

The driving force for the development of preventive dentistry and its effective use in patient management has been the expanding understanding of the disease of dental caries itself. As already discussed, dental caries is now recognized as a disease of altered ecology in which the host, oral microflora, and diet interact to present a challenge too strong for the normal defense mechanisms. It is wrong, however, to regard caries as a simple, continuing acid demineralization of the tooth enamel. Although teeth may be frequently exposed to acid environment, caries does not always arise as it is a result of a dynamic interaction of demineralization and remineralization. These are under the influence of a whole range of both cariogenic and protective factors (Figure 1). It is clear from this model that diet forms a part of a large spectrum of etiological factors involved in the caries process. Sugar and other fermentable carbohydrates in the diet constitute an essential aspect of cariogenic potential of food, with other factors, such as the frequency of intake, retentiveness, and buffering potential of foods, also playing an important part. These are in addition to the tooth structure, saliva flow and composition, and the presence of oral bacteria. This article is concerned with the discussion of the dietary considerations related to the etiology of dental caries.

The Role of Diet in the Etiology of Dental Caries

Epidemiological studies have been used to try to establish the relationship of various types of diets and dietary components to caries incidence. Circumstantial evidence linking sucrose consumption and prevalence of dental caries can be readily found in several epidemiological surveys. For example, the prevalence of caries among the native population of Tristan da Cunha, and among Eskimos and Australian Aborigines, was low before western-type food was introduced into their diets. Controlled human studies, mostly on institutionalized subjects, also indicated that sucrose-containing diets were cariogenic.

It is well known that when eating foods containing fermentable carbohydrates the pH of plaque drops, implying localized generation of acid. The drop in pH is the result of fermentation carbohydrates by the plaque bacteria, producing lactic and other organic acids, and is followed by a gradual return to the resting or near-resting pH. This recovery depends on a number of factors, including the buffering potential of plaque and the food, retention, or the rapidity of clearance from the mouth. The whole cycle of pH drop and recovery was first described by Stephan in 1940 and is commonly known as the Stephan curve (Figure 2). (*See* **Carbohydrates**: Digestion, Absorption, and Metabolism; Requirements and Dietary Importance.)

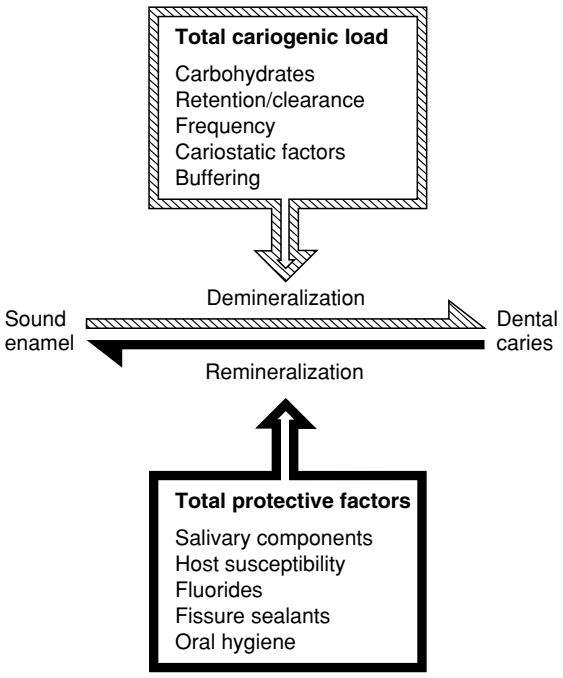

Figure 1 A dynamic model of caries. Reproduced from Dental Disease/Role of Diet, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

It is agreed that foods which produce a pH drop below pH 5.5, known as critical pH, are considered detrimental to teeth. The pH value 5.5 is considered critical because it is thought that enamel demineralization occurs below this point. Some authors regard values below 5.7 as being cariogenic. Foods giving a pH drop between 5.5 and 6.0 are also dubious. This form of ranking is referred to as relative potential cariogenicity. Of greater importance is the relative cariogenic potential of foods compared with known foods of high cariogenicity (sucrose) and low cariogenicity (sorbitol), as agreed at the international conference on cariogenicity of foods at San Antonio in 1985.

In one study, using the plaque-harvesting method, a representative sample of plaque was removed from human mouths before and after consumption of foods, and the pH response of that plaque was

assessed. The response followed a typical Stephan curve and the potential acidogenicity of the foods was compared by measuring the minimum pH recorded and the area enclosed by the curve under the resting pH value. Acidogenicity in a range of snack foods is assessed in Table 1. Interestingly, some foods which are perceived to be 'better for teeth' actually fare quite badly in such a ranking, compared with foods traditionally thought of as 'bad for teeth,' such as chocolate.

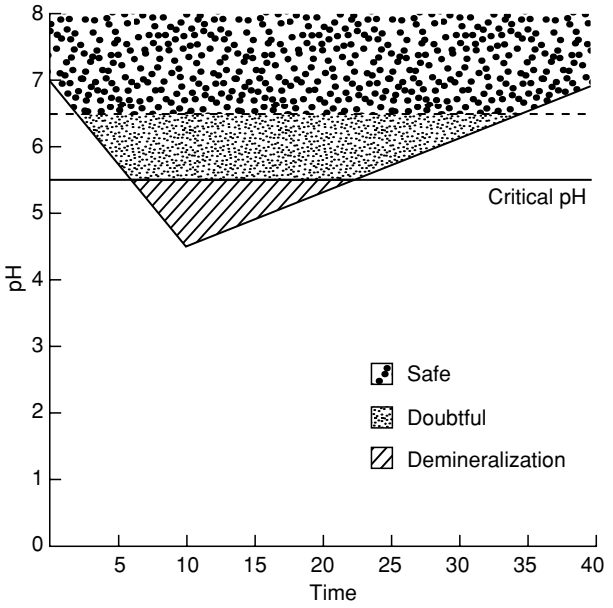

Figure 2 A typical Stephan's curve. Reproduced from Dental Disease/Role of Diet, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

This study bore out a previous study demonstrating that ingestion of chocolate or even apples resulted in a similar pH response. Later work, using the same test procedure, also showed that a wide range of foods containing either sugar or starches, or combinations thereof, are potentially acidogenic and thus possibly cariogenic.

It is important to note that the concentration of fermentable carbohydrate in a food does not affect the pH drop in the mouth, although the period of time taken to return to normal pH levels may be related to concentration. This return to resting pH is as much related to the buffering capacity of the saliva or plaque, buffering capacity of the foods, and the physical properties of the food itself. The retentiveness, and hence clearance rate, of a food is therefore an important determinant of the cariogenic potential of that food. The concentration of a single ingredient of a food has little relation to the cariogenicity; it is the ability of the whole food to promote caries that is important.

Tests for Cariogenicity of Foods

Animal experiments have been used to rank different foods in the order of their cariogenicity. In one such study a selection of human snack foods was ranked by their cariogenic potential index (CPI) by feeding

Table 1 Acidogenic potential of carious snack foods (US) grouped by category and acidogenicity

Group	*Beverages*	*Fruit, etc.*	*Baked goods*	*Sweets*
Least acidogenic	Milk	Peanuts		
	Chocolate milk	Potato chips (crisps)	Bread and butter	
		Apple	Graham crackers	Caramels
				Sugared gum
				Chocolate
				Licorice
				Sugarless gum
	Carbonated beverages	Banana	Cream-filled cakes	
			Sandwich cookies	Orange jellies
	Apple juice	Dates	Doughnut	
		Raisins	Bread and jam	
	Orange juice	Sweetened cereal	Whole-wheat bread	
			Plain sweet biscuits/cakes	
			Apple pie	Rock candy
			Chocolate	Clear mints
			Graham	
			Angel food cake	
Most acidogenic				Sourballs
				Fruit gums
				Fruit
				Lollipops

After Edgar WM (1981) Plaque pH assessments related to food cariogenicity. In: Hefferen JJ (ed.) *Foods, Nutrition and Dental Health*, vol. 1, pp. 137–150. Illinois: Pathotox.

Table 2 Selected rankings of cariogenic potential index (CPI) in rats using human food as snacks

Food tested	*CPI*
Sucrose	1.0
Filled chocolate cookie	1.4
Cereal (14% sucrose)	1.1
Cereal (8% sucrose)	1.0
Cereal (60% sucrose)	0.9
Coated chocolate candy	0.9
Potato chips	0.8
Caramel	0.7
Chocolate bar	0.7
Cereal (2% sucrose)	0.5
Starch	0.5
Sucrose plus 5% Dical	0.4
No meals by mouth	0.0

After Bowen WH, Amsbaugh SM, Monnel-Torrens S *et al*. (1980) A method to assess cariogenic potential of foodstuffs. *Journal of the American Dental Association* 100: 677–681.

laboratory rats via a gastric tube, thus bypassing the mouth. Sucrose was used as a reference food and given a CPI of 1.0. Foods with a score of less than 1.0 were considered less cariogenic than sucrose, while those with a score above 1.0 were thought to be more cariogenic. Interestingly, the concentration of sucrose in a breakfast cereal made little or no difference to the CPI. The results, given in Table 2, show that potato chips (crisps) actually score higher than chocolate bars. It is clear from most of these experiments that any foodstuff containing carbohydrate has the potential to cause significant amounts of acid to be produced at certain sites in the dentition, which can be followed by demineralization of the enamel and subsequent caries. However, it must be remembered that not all occasions of a drop in plaque pH are accompanied by demineralization of the enamel. This has prompted many investigators to question the significance of acid production as a measure of the cariogenicity of the foods. It has also been shown that the total amount of titratable acid produced by the foods does not necessarily parallel the amount of enamel it will dissolve. It is now well accepted that the cariogenic potential of a food is influenced by a number of other factors, including the ability of the foods to remain in the oral cavity and, in some cases, the sequence of food intake. Thus, studies on the relationship between food and dental caries should consider not only foods in themselves but also their relationship to other items of diet with regard to their nature, timing, and order of usage.

Cariostatic Factors in Food

Some components of foods may be cariostatic. Proteins may assist remineralization of enamel or reduce the rate of crystal dissolution. Some fatty acids have been shown to reduce caries in rat studies while phosphates have been shown to have a marked protective effect. Inorganic phosphates have been demonstrated to have a protective influence when added to a cariogenic diet. Organic phosphates such as phytates and glycerophosphates also have a cariostatic action and are thought to reduce the cariogenicity of diets. Although the exact mechanism of action is unknown, studies have indicated that it might be a local modifying influence in the oral cavity, rather than a systemic effect through ingestion. The local effects of phosphates can be attributed to various properties:

1. Phosphates are good buffers; thus they can buffer organic acids produced by plaque flora.
2. Phosphates are known to reduce the rate of dissolution of hydroxyapatite.
3. Phosphates can desorb proteins from the enamel surface; thus they can possibly have a modifying influence on acquired pellicle.

The protection afforded by fluoride is well documented and has led some researchers to refer to dental caries as a fluoride-deficiency disease. These materials are all components of various foods. Furthermore, some cariostatic agents have been isolated from cereals and cocoa and these factors may all influence the level of caries caused. Accordingly the level of fermentable carbohydrate in a food will not be directly related to the degree of caries caused.

Food Retention

Tests have illustrated that, contrary to popular opinion, foods that are perceived to be 'sticky,' such as caramel, tend to clear from the oral cavity faster than many other foods considered cariogenic. As Table 3 shows, after 15 min, white bread was retained in higher quantities in the oral cavity than cake, chocolate, or hard mint. After 30 min, there was more residue from raisins than from caramel. Raisins have been consistently shown to be highly cariogenic.

Beverages, which are perceived to clear quickly from the mouth, actually sustain a low pH level for the same period as a 'sticky' confectionery. In a recent study a range of fruit drinks which were advertised as 'sugar-free' or as 'no added sugar' were assessed on the basis of their ability to reduce the pH of plaque. It was found that the fact that most of these drinks had little natural sugar (sucrose) did not affect their ability to reduce the pH of plaque; when compared with a standard 10% sucrose rinse they were equally acidogenic. This was attributed to the presence of natural sugars – fructose and glucose – in these drinks. These so-called natural sugars are also fermentable by oral

Table 3 Representative figures (mg) for food retention in mouth after eating

Food	*Food retention (mg) after*		
	5 min	*15 min*	*30 min*
Peanuts	4.9	3.3	2.6
Dentyne gum	5.0	3.9	3.1
7-Up	6.3	2.4	2.1
Chocolate milk	7.4	3.8	1.9
Potato chip	12.3	4.9	2.5
White bread	16.1	10.0	3.6
Raisins	16.8	5.7	3.0
Sponge cake	18.8	6.0	4.2
Caramel	19.0	4.2	2.5
Milk chocolate	19.0	6.8	3.0
Cracker (oil-sprayed)	23.8	8.5	3.7
Hard mint	31.9	9.4	2.5
Cracker (plain)	33.6	10.4	3.3
Sandwich cookie	35.0	8.4	4.9

After Bibby BG (1981) Foods and dental caries. In: Hefferren J (ed.) *Foods, Nutrition and Dental Health*, vol. 1, pp. 257–278. Illinois: Pathotox.

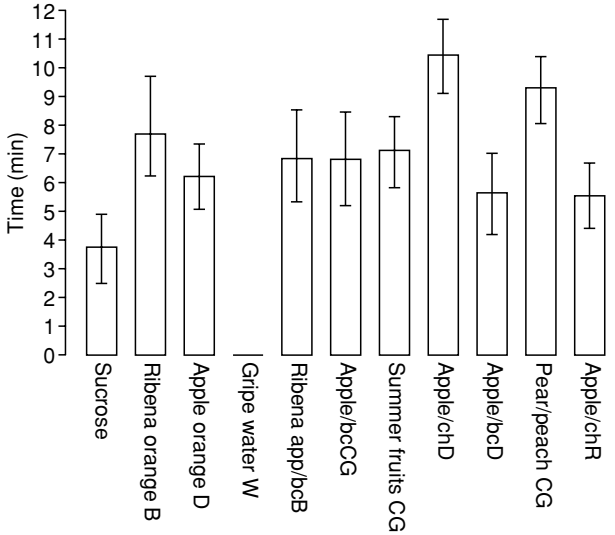

Figure 3 Bar graph showing mean time spent below pH 5.5 for fruit drinks compared with a 10% sucrose rinse. Reproduced from Dental Disease/Role of Diet, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

bacteria and their presence instead of sucrose does not render the drinks any safer for teeth. In fact, when the fruit drinks were compared with 10% sucrose it was obvious that the pH recovery was slower after consumption of fruit drinks and the pH remained below 5.5 for a longer time (Figure 3). This is because of an inherent buffering potential of fruit drinks which gives them an ability to resist any attempts by saliva to buffer the acid. These drinks can therefore be deemed to be more cariogenic than a pure 10% sucrose solution. This research highlighted the fact that the concentration of sugars alone is not the sole determinant of the acidogenicity, and hence cariogenicity, and other factors discussed above are equally important. (*See* **Carbohydrates: Metabolism of Sugars.**)

Eating Pattern and Frequency

On a population level, the average amount of sucrose consumed per capita relates to the average level of caries in the population. However, more detailed studies show the relationship to be less consistent and of low statistical significance. In the classical study often referred to as the Vipeholm study, inmates of a Swedish Medical Institute were fed increased sucrose or other foods in different patterns and caries experience was monitored. Groups of patients receiving high levels of sucrose (up to $330\,g\,day^{-1}$) with their meals experienced minimal increase in caries. But if smaller quantities of sucrose were consumed between meals, very high levels of caries ensued. The relationship was not therefore between quantity of sucrose and caries but rather frequency of intake and caries experience. This relationship, which has been confirmed in human and animal research, sheds light on why population studies do not demonstrate a clear and consistent relationship between sugar consumption and caries. For example, long-term diet–caries studies carried out in the UK showed that statistical correlation of sugar concentration to caries was 0.16 and explained only 4% of the caries variance.

Experience from primitive and developing cultures with little access to sucrose but abundant access to starch is often cited as evidence that sucrose and not starch results in dental caries. This evidence purports to be strengthened by the fact that introduction of western diet (including sucrose) immediately results in development of dental caries. It can be argued that introduction of a western diet is accompanied by increased affluence and an altered eating pattern. Such changes also include differences in the use of cooked starches as much as differences in use of sucrose. Frequency of intake of any food increases dramatically, along with the potential incidence of dental caries. It is also interesting to note that caries has been shown to be associated with a diet consisting of sago starch, as in the Sepik villages in Papua New Guinea.

Sugars and their Role in Caries

In western society, eating frequency has generally increased and snacking has become an accepted aspect of life. This change took place over a long period of the nineteenth and early twentieth centuries

when the incidence of dental caries increased. However, in the past 20 years, while sucrose usage has not changed, dental caries has dramatically decreased. In the case of 5-year-olds, the percentage with tooth decay fell from 73% in 1973 to 48% in 1983.

Reducing the frequency of eating just one of these foods, or reducing the concentration of sugars in a food, is unlikely to have a significant effect on the incidence of caries as nearly all foods contain some fermentable carbohydrate. It has been assumed tacitly both by practicing dentists and by too many dental investigators that the cariogenicity of an individual food is directly proportional to its content of sucrose or other fermentable carbohydrates. There are no quantitative data to support this belief. The effects of high sucrose concentrations in increasing the rate of food clearance of some foods from the mouth and in inhibiting the fermentation process make it seem improbable that high sugar content of itself would be particularly damaging to teeth. The Vipeholm study has been mentioned as evidence for the cariogenicity of sucrose, although investigators have questioned the reliability of a single clinical study from a mental institute. There are a number of contradictory studies that have not been widely recognized. For example, one group of investigators, studying English children, found that they could substantially increase sugar as sucrose in the children's diet without increasing the incidence of caries.

It is the frequency of consumption rather than the amount consumed which is associated positively with dental caries incidence. Considering the already substantial decline in the incidence of dental caries in the west, where frequency of eating has generally increased, and given the fact that dietary manipulation is difficult to achieve, it is reasonable to assume that alternative preventive measures such as the use of fluoride would be unfortunate if public hopes were raised to believe that dietary control alone would solve the problem of dental caries.

See also: **Carbohydrates**: Metabolism of Sugars; Digestion, Absorption, and Metabolism; Requirements and Dietary Importance

Further Reading

Bibby BG (1981) Foods and dental caries. In: Hefferren JJ (ed.) *Foods, Nutrition and Dental Health*, vol. 1, pp. 257–278. Illinois: Pathotox.

Bowen WH, Amsbaugh SM, Monnel-Torrens S *et al.* (1980) A method to assess cariogenic potential of foodstuffs. *Journal of the American Dental Association* 100: 677–681.

Duggal MS and Curzon MEJ (1989) An evaluation of the cariogenic potential of baby and infant fruit drinks. *British Dental Journal* 160: 327–330.

Edgar WM (1981) Plaque pH assessments related to food cariogenicity. In: Hefferren JJ (ed.) *Foods, Nutrition and Dental Health*, vol. 1, pp. 137–150. Illinois: Pathotox.

Edgar WM, Bibby BG, Mundorff S and Rowley J (1975) Acid production in plaques after eating snacks: modifying factors in foods. *Journal of the American Dental Association* 90: 418–425.

Gustafsson BE, Quensel CE, Lanke LS *et al.* (1954) The effect of different levels of carbohydrate intake on caries activity in 436 individuals observed for five years. The Vipeholm Dental Caries Study. *Acta Odontologica Scandinavica* 11: 232–364.

King JD, Mellanby M, Stones HH and Green HN (1955) The effect of sugar supplement on dental caries in children. MRC Special Report Series, no. 288. London: Her Majesty's Stationery Office.

Navia JM *et al.* (1983) Nutrition in oral health and disease. In: Stallard RE (ed.) *A Textbook of Preventive Dentistry*, pp. 90–146. Philadelphia: WB Saunders.

Rugg-Gunn AJ, Edgar WM and Jenkins GN (1978) The effect of eating some British snacks upon the pH of human dental plaque. *British Dental Journal* 145: 95–100.

Stephan RM (1940) Changes in hydrogen ion concentration on tooth surfaces and in carious lesions. *Journal of the American Dental Association* 27: 718–723.

Fluoride in the Prevention of Dental Decay

H Whelton, University College Cork, Cork, Ireland

Introduction

Dental decay or caries is a ubiquitous disease in established market economies and in many developing countries. Dental caries poses a threat to the dentition throughout life. The early caries lesion is known as a 'white spot.' This white spot can do one of three things. It can stay the same, progress to a cavity, or regress to sound enamel. The fate of the white spot is influenced by its exposure to fluoride. Fluoride can help to prevent progression of the lesion and can aid remineralization, in which case the early white-spot lesion can disappear clinically.

Fluoride is widely used in the prevention of dental caries. It is estimated that over one billion people use fluoridated toothpaste worldwide. The second most common vehicle for fluoride is water. Water fluoridation is the controlled addition of fluoride to a public water supply for the purpose of preventing dental caries. In many parts of the world, the level of fluoride in water supplies is adjusted to 1 part per million (ppm). It is estimated that about 317 million people

in 39 countries benefit from artifically fluoridated water. A further 40 million benefit from water supplies which are naturally fluoridated. The fluoridation of toothpastes and public water supplies has been acknowledged as a major public health success in the prevention of dental caries. The addition of fluoride to water supplies and to oral health products has helped to bring about a remarkable decline in dental caries levels since the 1960s. In this paper the occurrence of fluoride in nature, the history of the discovery of its effects on the dentition, its use in the prevention of dental caries, and the risks associated with ingestion of excess fluoride will be considered.

Fluoride in Nature

Fluorine is a naturally occurring element, apparently ubiquitous in nature. Number nine on the periodic table of the elements, it is a halogen gas. It ranks 13th among the elements in order of abundance in the earth's crust. Fluorine is the most electronegative and reactive of all elements. It is a pale yellow, corrosive gas, which reacts with practically all organic and inorganic substances. The element was isolated in 1886 by Ferdinand Frederic Henri Moisson who used an apparatus constructed from platinum. He won the Nobel Prize for chemistry in 1906. Fluorine occurs chiefly in fluorspar (or fluorite, calcium difluoride, CaF_2), cryolite (Na_2AlF_6), and in many other minerals. Ionic fluoride, also referred to as inorganic or free fluoride, is the biologically important form of the mineral.

Fluoride is found in fresh water, sea water, and earth and in foods. Fluoride is present in both groundwater and surface water. The level of fluoride in groundwater varies from less than $0.1\,mg\,l^{-1}$ to more than $25\,mg\,l^{-1}$ according to the geological, chemical, and physical characteristics of the water-supplying area. The fluoride concentration in fresh surface water is generally low, ranging from 0.01 to $0.03\,mg\,l^{-1}$. In sea water, fluoride is found at approximately 1.5 ppm. Marine plants and animals are therefore constantly exposed to large amounts of fluoride. Fluoride in soils is derived primarily from the geologic parent material. Samples of nonindustrially contaminated soils from Germany, Greece, Finland, Japan, Morocco, New Zealand, Sweden, the USA and former USSR have been reported as containing from 30 to 500 ppm fluoride. Soils near fluorite and other types of mineralization may show levels up to more than 5000 ppm.

The normal accumulation of soil fluoride in plants is small. However, a few species of plants are known to accumulate high levels of fluoride, for example tea. The concentration of fluoride in dried tea leaves varies widely (approximately 4–400 ppm) while those of brewed tea range from 1 to 6 ppm depending on the amount of dry tea used, the fluoride concentration of the water, and the brewing time. High levels of fluoride are also found in fish: early reported concentrations range from 0.6 to 2.7 ppm. These samples may have contained bones. With its high affinity for calcium, most absorbed fluoride is deposited in bones therefore fish bones are a source of concentrated fluoride levels. More recent studies which excluded bone fragments from fish reported a range of fluoride values from 0.05 to 0.17 ppm. Other dietary constituents which contain fluoride are fluoridated water and infant formulas and beverages reconstituted with fluoridated water. Most other foods have fluoride concentrations well below 0.05 ppm.

History of Water Fluoridation

In 1901 a letter in the US Public Health Report from JM Eager of the US Public Health Service, stationed in Naples, Italy, reported the occurrence of 'a dental pecularity' known locally as denti di chiae. Denti di chiae were called after Prof Stefano Chiae, a celebrated Neapolitan who first described the appearance of a condition which was later to become known as dental fluorosis.

A similar condition, known as Colarado brown stain, was observed in Colorado Springs, USA, in 1901 by a dentist, Fredrick McKay. The condition manifested itself as a stain of varying intensity from fine white patches to a disfiguring brown mottling. Some 87% of the population in the area were affected. In 1916, McKay, with the help of GV Black, a prominent figure in dentistry at the time, published a thorough description of the condition, describing its appearance as mottled enamel. By the 1920s McKay had suspected drinking water as the source of the problem. In Oakley, Idaho, where mottling was severe, McKay noticed that children living on the outskirts of town who drank from a private spring had no mottling. He advised the town to change to this supply, which they did in 1925. Children born in the town subsequently had no mottling.

In 1931 new methods of spectrographic analysis led to the discovery of fluoride in the water supplies of areas where mottled enamel was endemic. The condition then became known as dental fluorosis.

In 1928 McKay published the observation that, in areas where mottled enamel was found, the prevalence of dental caries appeared lower than would be expected. In 1933 Ainsworth in England showed that dental mottling in Maldon in Essex was associated with a high level of fluoride in its water (4.5–5.5 ppm). He also noted that the prevalence of caries in permanent teeth amongst the children in the area was lower than that for England and Wales as a whole.

The discovery of fluoride (known to act as a protoplasmic poison in some circumstances) as the causative agent in mottling of the teeth was a cause of concern to the US Public Health Service. In response they appointed a dentist, H Trendley Dean, to investigate the condition. Dean devised a six-point scale of severity of dental fluorosis which is still used today. Dean's index facilitates regional and international comparisons of dental fluorosis as well as temporal comparisons. However the index is relatively subjective and the appearance of enamel varies according to lighting conditions, drying of the enamel, and angulation of viewing. Dean's index categorizes the enamel into normal, questionable, very mild, mild, moderate, and severe (Figure 1). Some of the limitations of Dean's index have since been overcome by the development of a standardized photographic method for the recording of dental fluorosis.

Dean combined measurement of the condition of the enamel with analysis of the drinking water throughout the USA. He then set out to identify the water fluoride levels, which represented the best compromise between low caries experience and a level of fluorosis, which could be deemed acceptable. The resulting study became known as the '21 cities' study. Dean recorded the level of dental caries and dental fluorosis in 7257 12–14-year-old children who had been lifetime residents of 21 cities with varying levels of natural fluoride in their water supplies from 0.0 to 2.6 ppm.

The marked inverse association between the fluoride content of the drinking water and the experience of caries in the population is shown in Figure 2. This study showed that, above a concentration of 1 ppm fluoride in the water, very little further reduction in caries was obtained. Dean also illustrated the dose–response relationship between water fluoride levels and dental fluorosis in the '21 cities study.' He found that, where the fluoride concentration exceeded 1.0 ppm, undesirable mottling would be seen in about 10% of the population.

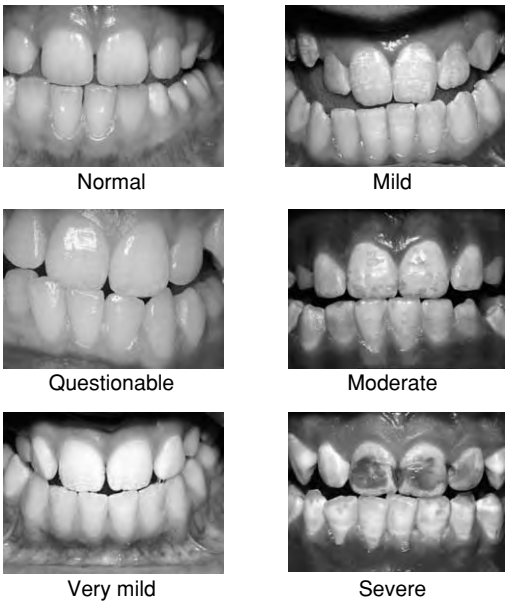

Figure 1 (see color plate 46) Categories of dental fluorosis according to Dean's index.

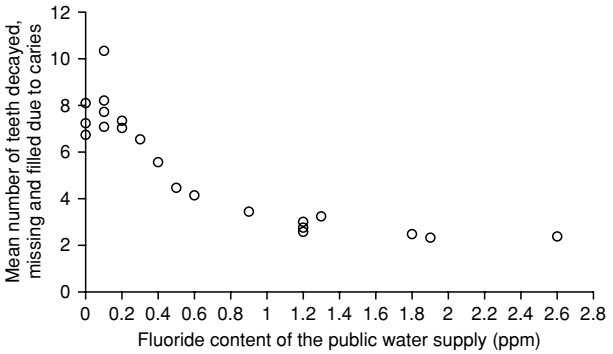

Figure 2 The relation between caries levels observed in 7257 selected 12–14-year-old white school children of 21 cities of four states and the fluoride content (ppm, horizontal scale) of public water supply. Open circles, DMFT (mean number of teeth decayed, missing, and filled due to caries). (Reproduced from F.J. McClure (ed.) (1962) *Fluoride Drinking Waters*, by kind permission of the National Institute of Dental and Craniofacial Research).

In 1945 and 1946 four cities in the USA and Canada began to add fluoride to their water supplies at a concentration of 1 ppm as part of the first controlled field trials of water fluoridation. The fluoridation of water to 1 ppm reduced the number of carious teeth per child by on average 60% relative to children in nonfluoridated cities.

Fluoridation Worldwide

Throughout the world it is estimated that about 317 million people drink artificially fluoridated water and that a further 40 million drink water in which the natural fluoride level is high enough to provide a significant degree of protection against tooth decay. Countries with fluoridation schemes include the USA, Canada, Mexico, Argentina, Ireland, UK, Spain, Australia, New Zealand, Hong Kong, and Singapore.

The USA is the most extensively fluoridated country in the world, with some 135 million people – over half the population – currently receiving artificially fluoridated water and another 10 million receiving naturally fluoridated water. The national health promotion and disease prevention objectives in Healthy People 2010 call for increasing the percentage of

Americans on public water supplies drinking fluoridated water from 62 to 75%.

Within the European Community (EC), the UK, Ireland, and Spain are currently operating fluoridation schemes. In Ireland, fluoridation of public water supplies is mandatory. An EC directive relating to the quality of water intended for human consumption (98/83/EC) sets a maximum admissible concentration (MAC) for fluoride at 1.5 parts of fluoride per million parts of water, irrespective of climate.

Some countries, such as France, Switzerland, and Germany, have implemented salt fluoridation.

Fluoride Ingestion and Absorption

Approximate estimates for daily dietary fluoride intake are shown in Table 1. It is estimated that 75–90% of daily fluoride intake is absorbed. Of that absorbed, the amount excreted varies; it is estimated that on average approximately 50% of absorbed fluoride is excreted, primarily by the kidneys in urine. In the very young more fluoride is retained. Most retained fluoride is deposited in bone because of its strong affinity for calcium.

Assessing fluoride levels in plasma, ductal saliva, or in urine can help monitor recent total fluoride exposure of individuals or populations. Research is ongoing into the use of fingernail clippings as a biomarker for fluoride intake some weeks prior to clipping. Dental fluorosis is a convenient biomarker for fluoride intake; however it is relevant only to the first 6 years of life. Dental fluorosis is a convenient biomarker for fluoride intake, however it is relevant only to the first 6 years of life. Dental hard tissues provide a marker for fluoride exposure during defined developmental periods. Bone provides information on the accumulated body burden of fluoride over a lifetime.

Fluoride Toothpastes

Studies of the effectiveness of fluoride toothpaste in the 1940s and early 1950s failed to demonstrate a caries-preventive effect. It is now believed that the sodium fluoride was not bioavailable due to a reaction with the calcium carbonate or calcium orthophosphate abrasive in the formulation. Since the 1950s studies of 2–3 years' duration have reported that fluoride toothpaste reduces caries experience among children by a median of 15–30%. A dose–response relationship has been demonstrated showing greater caries-preventive effects for higher-dose fluoride toothpaste. The two most commonly used forms of fluoride added to toothpaste are sodium fluoride (NaF) and sodium monofluorophosphate (MFP). There are many different formulations of fluoride toothpaste on the market. The concentration of fluoride in toothpastes varies from 250 to 2800 ppm. The most commonly sold toothpastes contain 1000–1500 ppm. In the European Union the maximum permissible concentration in fluoride toothpaste to be sold over the counter is 1500. Hence toothpaste formulations with greater than 1500 ppm fluoride are available at pharmacies only.

Manufacturers commonly list fluoride levels in toothpaste as a percentage. Table 2 shows the common percentages of NaF and NaMFP found in toothpastes, expressed as ppm in toothpaste.

Fluoride toothpaste contributes to the risk of enamel fluorosis because the swallowing reflex of children aged under 6 years and especially in those aged under 3 years is not well controlled. Early use of fluoride toothpaste by children is associated with the development of dental fluorosis. Ingestion of excessive amounts of toothpaste during the period of enamel formation can cause fluorosis. Enamel formation for all of the anterior teeth and most of the posterior teeth is complete by age 7. The critical window of vulnerability for anterior teeth to develop fluorosis from ingested fluoride has been reported to be 15–24 months for boys and 21–30 months for girls. The amount of toothpaste used accounts for an estimated 60% of the variation in the amount of toothpaste swallowed by children under age 7. It takes 0.75–1.0 g of toothpaste to cover a full head of a child-sized toothbrush. For a 1000 ppm toothpaste, this represents 0.75–1.0 mg fluoride. To minimize toothpaste ingestion up to age 7, manufacturers advise supervision and use of only a pea-sized amount of toothpaste (0.25 g). Many experts also recommend that toothpaste should not be used when brushing children's teeth up to the age of 2 years. The risk of ingesting fluoride toothpaste has led to the development of low-fluoride pediatric toothpastes with 250 or 550 ppm fluoride. These toothpastes have not been shown to be as effective in preventing caries as the

Table 1 Estimated daily fluoride intake from dietary sources

	Fluoridated area	*Nonfluoridated area*
Adults	1.0–3.0 mg	≤1.0 mg
Children	0.5 mg	0.25 mg

Table 2 NaF and NaMFP found in toothpastes, expressed as percentages and as ppm in toothpaste

NaF (%)	*NaMFP (%)*	*ppm*
0.32	1.14	1500
0.22	0.76	1000
0.11	0.38	500

1000 ppm formulation. Over the age of 7 years, when there is little risk of swallowing toothpaste, children who are able to expectorate the toothpaste properly may be allowed to use a greater quantity of toothpaste with up to 1500 ppm fluoride.

Other Fluoridated Products

Many other vehicles for the delivery of fluoride for the prevention of dental caries have been shown to be effective. Under the appropriate circumstances, fluoridated salt (250 ppm) has been shown to be as effective as water fluoridation. Fluoridated domestic salt is available in Switzerland, France, and Germany. Fluoridated milk and sugar have been tested on a pilot basis. Among the other products available are fluoride tablets for daily use, drops, mouth rinses for daily (0.05% NaF, 230 ppm) or fortnightly (0.2% NaF, 910 ppm) use, and products for application by dental professionals, including topically applied gels and varnishes. Applying fluoride gel (2.72% NaF, 12 300 ppm) or other products containing a high concentration of fluoride to the teeth leaves a temporary layer of calcium fluoride-like material on the enamel surface. The fluoride in this material is released when the pH drops in the mouth in response to acid production and is available to promote enamel remineralization. Dental filling materials have also been developed which release fluoride over time; however, fluoride release is limited to a relatively short period after placement. Other products have also been developed for placement in the mouth which would slowly release fluoride over months, if not years. Human clinical trial data on the effectiveness of these products are not yet available. Fortnightly school-based fluoride mouth rinse programs for the prevention of dental caries have been shown to be successful for the duration of the program. However in programs that cease at age 12, the effects have been shown to fade after 4 years.

Mechanism of Action of Fluoride in the Prevention of Dental Caries

Dental enamel is a highly mineralized tissue made up of 96% mineral and 4% organic material and water. The inorganic content of enamel consists of a crystalline calcium phosphate known as hydroxyapatite. The dissolution of these crystals by acid is the start of the caries process. Fortunately, enamel can remineralize as well as demineralize and the enamel surface is in a constant seesaw between demineralization and remineralization. Saliva provides a supersaturated solution of ions and acts as an ion reservoir during this process. When the equilibrium is disturbed, caries can progress.

In the early days scientists thought that the anticaries activity of fluoride was a preeruptive systemic effect as a result of its incorporation into the enamel crystal to form fluorapatite during the time of enamel formation. This crystal is more stable and resistant to acid demineralization than hydroxyapatite. However, investigators have failed to show a consistent correlation between anticaries activity and the specific amounts of fluoride incorporated into the enamel. The theory of preeruptive fluoride incorporation as the sole or principal mechanism of caries prevention has been largely discounted. Another theory was that fluoride inhibits bacterial metabolism; however, the low levels of fluoride found in the mouth after toothbrushing with fluoride-containing dentifrices are ineffective in interfering with processes of growth and metabolism of bacteria. The current theory is that the caries-preventive effect of fluoride is mainly attributed to the effects on demineralization/remineralization at the tooth–oral fluids interface. Levels well under 1 ppm of fluoride in saliva are effective in shifting the balance from demineralization, leading to caries, to remineralization. This is attributed to the fluoride-enhanced precipitation of calcium phosphates, and the formation of hydroxyfluorapatite in the dental tissues.

Benefits and Risks of Fluorides in Contemporary Society

In the 1950s and 1960s dental caries was an overwhelming problem in many countries. For example, in Ireland prior to the introduction of water fluoridation, in the early 1960s 12-year-old children had on average five teeth with decay. In the 1990s 12-year-old children whose domestic water supplies were fluoridated had on average one decayed tooth, and those without water fluoridation had on average two decayed teeth. The dramatic decline in the level of caries both in fluoridated and nonfluoridated areas reflects changes seen in the other established market economies.

Whilst early studies reported up to a 60% reduction in caries with water fluoridation, more recent estimates of effectiveness are lower with an 18–40% difference in caries levels. This apparent reduction in effectiveness is probably due to the widespread use of fluoridated toothpaste in both fluoridated and nonfluoridated communities. In addition, the consumption of foods and drinks manufactured using fluoridated water and consumed in nonfluoridated areas may confer some protection against dental caries to populations in nonfluoridated areas. This is known as the halo effect. Thus, where nonfluoridated communities are exposed to fluoride from sources

Table 3 Objectives and conclusions of the systematic review of public water fluoridation conducted by the NHS Centre for Reviews and Dissemination, University of York, 2000

Objective	*Number of studies included in review*	*Conclusion*
The effects of fluoridation of drinking water supplies on the incidence of caries	26	The best available evidence suggests that fluoridation of drinking water supplies does reduce caries prevalence
The effect over and above that offered by the use of alternative interventions and strategies	9	Water fluoridation has an additional beneficial effect
The effect of water fluoridation on levels of caries across social groups and between geographical locations	15	Inconclusive evidence
Negative effects		
Fluorosis	88	Yes, dose–response relationship with water fluoridation
Bone fracture and bone development problems	29	No association found
Cancer	26	No association found
Other negative effects	33	No clear association found
Differences in the effects of natural and artificial water fluoridation		Not enough evidence to draw conclusions

other than water, the observed effectiveness of water fluoridation is reduced. In Ireland water fluoridation was introduced in the 1960s and the use of fluoridated toothpastes became widespread in the 1970s. Whilst the level of caries has declined, the prevalence of fluorosis amongst Irish children appears to have increased since the 1980s when it was first measured. The alteration in the pattern of caries, the changes in the severity of the disease, and the use of multiple sources of fluoride have prompted a number of reviews of the risks and benefits of fluoridation. In 2000 the University of York presented a systematic review of the evidence on the positive and negative effects of population-wide drinking-water fluoridation strategies to prevent caries. The review summarized the best available and most reliable evidence on the safety and efficacy of water fluoridation. Of all the studies found in the literature search, 214 satisfied the inclusion criteria for consideration in the appraisal of the evidence. Outcomes for the five objectives of the review are shown in Table 3.

The reviewers were critical of the lack of high-quality research on water fluoridation. They recommended: 'the evidence of a benefit of a reduction in caries should be considered together with the increased prevalence of dental fluorosis.'

Those who oppose water fluoridation do so on three main grounds:

1. the effectiveness of water fluoridation in preventing caries in contemporary society
2. the safety of water fluoridation. Concerns around the general health effects of fluoride have been raised. For example, fluoride has a high affinity for calcium and most fluoride, which is absorbed and not excreted, is deposited in bones, hence, the effect of fluoride on bone health has been questioned
3. the ethics of administering fluoride through the public water supply and removal of the right to choose whether one's water supply is fluoridated or not

With regard to effectiveness, the current evidence continues to demonstrate a positive protective effect of water fluoridation against dental caries. Concerning safety, studies have failed to find harmful effects of water fluoridation at 1 ppm. Finally, the ethical debate is a philosophical one and is outside the scope of this paper.

Conclusion

Fluoride has been shown to be effective in the control of dental caries. Its effect is mainly topical at the plaque–enamel interface. The aim of fluoride therapy is to maintain the ambient level of the fluoride ion in the saliva at a constant optimum level. The evidence suggests that the effects of fluoridated water and fluoride toothpaste are additive. However, the risk associated with excessive ingestion and absorption of fluoride is the development of dental fluorosis.

See also: **Fluoride**; **Water Supplies**: Chemical Analysis

Further Reading

Evans RW and Darvell BW (1995) Refining the estimate of the critical period for susceptibility to enamel fluorosis in human maxillary central incisors. *Journal of Public Health Dentistry* 55(4): 238–249.

Fejerskov O, Ekstrand J and Burt BA (1996) *Fluoride in Dentistry*, 2nd edn. Copenhagen: Munksgaard.

Fluorine in web elements. http://www.webelements.com.

Holland TJ, Whelton H, O'Mullane DM and Creedon P (1995) Evaluation of a fortnightly school-based sodium fluoride mouthrinse 4 years following its cessation. *Caries Research* 29(6): 431–434.

Mascarenthas AK and Burt BA (1998) Fluorosis risk from early exposure to fluoride toothpaste. *Community Dental and Oral Epidemiology* 26(4): 241–248.

Murray JJ, Rugg-Gunn AJ and Jenkins AN (1991) *Fluorides in Caries Prevention*, 3rd edn. Oxford: Wright.

NHS CRD (2000) A systematic review of public water fluoridation. Report 18. York: NHS Centre for Reviews and Dissemination, University of York.

Taves DR (1983) Dietary intake of fluoride ashed (total fluoride) v. unashed (inorganic fluoride) analysis of individual foods. *British Journal of Nutrition* 49: 295–301.

ten Cate JM (1999) Current concepts on the theories of the mechanism of action of fluoride. *Acta Odontologica Scandinavica* 57: 325–329.

US Department of Health and Human Services (2000) *Healthy People 2010*. Washington: US Department of Health and Human Services.

US Department of Health and Human Services (2000) *Oral Health in America: A Report of the Surgeon General*. Rockville, MD: US Department of Health and Human Services, National Institute for Dental and Craniofacial Research, National Institute of Health.

US Department of Health and Human Services Center for Disease Control and Prevention (2001) *Morbidity and Mortality Weekly Report* 50: no. RR-14.

Wang NJ, Gropen AM and Ogaard B (1997) Risk factors associated with fluorosis in a non-fluoridated population in Norway. *Community Dental and Oral Epidemiology* 25(6): 396–401.

Whitford GM (1996) *The Metabolism and Toxicity of Fluoride*, 2nd edn. Basel: Karger.

World Health Organization (1994) *Report of a WHO Expert Committee on Oral Health Status and Fluoride Use. Fluorides and Oral Health*. Geneva: WHO.

Detergents *See* **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant; Overall Approach; Modern Systems

Detoxification *See* **Toxins in Food – Naturally Occurring**; **Alkaloids**: Properties and Determination; Toxicology

DEVELOPMENTAL DISABILITIES AND NUTRITIONAL ASPECTS

Contents

Down Syndrome

B Lucas and S Feucht, University of Washington, Seattle, WA, USA

Etiology

Down syndrome (DS) (Trisomy 21) is one of the most common chromosomal abnormalities, affecting one in 600 live births. DS results from the presence of extra genetic material (a third copy) on chromosome 21. It occurs during the prenatal period and affects all races. The incidence of DS increases with maternal age, although 80% of children with DS are born to mothers less than 35 years of age due to increased fertility of younger women. Current knowledge describes three types of genetic changes in chromosome 21 that cause DS:

- nonfamilial trisomy 21, occurring in 95% of individuals;

- unbalanced translocation between chromosome 21 and another chromosome, in 3–4%; of these, approximately 25% result from a familial balanced translocation, and parents may be at increased risk of having another child with DS;
- mosaicism, in 1–2% of individuals with DS, resulting from two present cell lines – one normal and one trisomy 21; these individuals are usually affected less severely than the other types.

Recent research has studied the role of folate and homosysteine metabolism in women who had children with DS. Although no conclusions could be drawn from the small numbers, more work is needed. It is possible that both genotype and nutrition may play a role in the etiology of DS, as it does with neural tube defects.

Features and Characteristics

Individuals with DS demonstrate physical features and other characteristics that define the syndrome and determine the extent of health care, special education, therapies, and vocational planning needed. Developmental delay, hypotonia (reduced muscle tone), and short stature are seen in all individuals with DS. Cognitive levels usually range from mild to moderate mental retardation. See Table 1 for common characteristics.

Overweight and obesity are frequently seen in individuals with DS, resulting from a variety of factors. Other health concerns occurring less frequently, but impacting nutrition, feeding, or physical activity, include congenital heart disease, hearing problems, vision problems, cervical spine abnormalities, seizure disorders, and thyroid disease.

Down syndrome is associated with premature aging, often beginning in the fourth and fifth decades of life. Along with this aging, there is reported to be an increased incidence of Alzheimer's disease, senility, autoimmune diseases, and cataracts. Preliminary studies had documented *in vitro* chronic oxidative stress in persons with DS compared with their control siblings, but the role of this or other biomarkers in accelerated aging is controversial.

Table 1 Characteristics of persons with Down syndrome

Hypotonia
Mental retardation
Short stature
Flat facial profile
Eye features (upward slanting eyelids; skinfolds in corners of eyes)
Dental hypoplasia and irregular tooth placement
Increased risk of overweight and obesity
Accelerated aging process, i.e., Alzheimer-like brain changes, cataracts
Increased incidence of congenital heart disease, thyroid disease, leukemia

Services for individuals with DS are primarily community-based, compared with the past when most of this population would be found in residential institutions. Currently, the diagnosis is made at, or shortly after, birth, and early intervention services and therapies are begun in early infancy. Full inclusion in social, educational, and work environments is supported in most all communities and countries. As adults, many people with DS live in the community (with family, in group or individual homes) and are in vocational programs or gainfully employed.

Growth

Individuals with DS grow more slowly and are shorter in final stature than those without the syndrome of the same age and sex. The growth differences begin in the prenatal period. By puberty, children with DS are typically 1–2 inches (2.5–5 cm) shorter than their peers. The onset of the growth spurt may be delayed 6–12 months in adolescents with DS, and some may never have a growth spurt. Specific medical conditions such as congenital heart disease may also impact growth and weight gain for a child with DS. Since the onset of overweight can begin in late infancy or the preschool period, appropriate prevention strategies by the health care team should begin early.

Specialized Growth Charts

Specialized growth charts have been published for children with DS from birth to 36 months of age and from 2 to 18 years of age. They provide a better picture of length or stature in relation to other children with DS. These DS charts reflect small sample numbers (longitudinal data for weight and length/height for 400 males and 300 females), and include some children with congenital heart disease. The data were collected from 1960 to 1986 on children living at home, and they reflect a tendency to overweight. It is essential to use these specialized charts in combination with standard growth charts, such as the Center for Disease Control Growth Charts: United States 2000. The DS charts do not include head circumference data for younger children, nor do they compare weight/length or height in younger children or body mass index for children over the age of two.

Overweight and Obesity

Increasing rate of weight gain is a risk in persons with DS due to physical characteristics, medical conditions, and environment (see Table 2). With wellness

programs for this population and their families/caregivers, however, obesity can be prevented. A 1992 study of 30 sibling pairs, aged 2–14, found no significant differences between the DS children and their siblings in terms of BMI. The children with DS were less active than their siblings. Although children with DS who grow up with their own families have less overweight and obesity than their institutionalized peers, adults with DS living in the community are more likely to be overweight or obese than those in institutional settings. In 1995, Prasher reported on an assessment of overweight and obesity in 201 adults with DS living in the UK. Using BMI values, 31% of males were overweight, and 48% were obese (BMI > 30). For females, 22% were overweight, and 47% were obese; they also had the more severe forms of obesity. This study found no association between overweight/obesity and the severity of learning disability. Of note was the decreasing fall in BMI with increasing age, but it is unclear if this is an effect of the precocious aging or other related medical conditions.

For children, weight should be closely monitored on standard growth charts to prevent rapid or excessive gain that may not be as clearly seen on the DS charts. Overall, the velocity of growth and weight gain is more important than the actual percentile on either growth chart (see Table 3). For adults, regular health assessment should include weight and BMI monitoring. Adults with DS and their older parents may not have had the benefit of early intervention programs and may not be as enthusiastic about physical fitness and healthy diet plans as younger persons. Regardless of the etiology, the hazards of obesity make prevention and intervention important.

Inclusion in social, educational, and vocational environments brings new dietary challenges. Food choices, from vending machines and fast food outlets, typically favor high-fat, high-energy, nutrient-poor foods and beverages. Nutrition education is required to assist individuals with DS to make healthy food choices for themselves in their communities.

Table 2 Risk factors for overweight and obesity in individuals with Down syndrome

Risk factors	*Effect*
Hypotonia Short stature Low resting metabolic rate Hypothroidism Decreased pulmonary function Cardiac malformations Premature aging Limited instruction in wellness Food used as a reward Limited physical activity	Easily tired Excess energy intake Poor food choices Preference for indoor activities Disinterest in physical exercise

Adapted from Pipes P and Powell J (1996) Preventing obesity in children with special health care needs. *Nutrition Focus Newsletter* 11(6): 1–8.

Table 3 Guidelines for growth monitoring and obesity prevention in children with Down syndrome

Use standard growth charts for assessing weight for length, weight for height, and/or body mass index (BMI)
Down syndrome-specific growth charts may be used with parents to compare their children with their peers for length/height
Height and activity level should be used to determine individual energy requirements (a baseline for children 5–11 years of age is 16.1 kcal (67 kJ) per centimeter of height for males and 14.3 kcal (60 kJ) per centimeter of height for females)
Physical exercise is an important adjunct to an appropriate diet to increase lean body mass and, if needed, decrease fat tissue

Nutrition and Feeding Issues

Individuals with DS require the same nutrients and a varied healthy diet as anyone else. However, the syndrome characteristics may make them more at risk for nutrition and feeding problems. Owing to early development delays, some feeding issues can arise that affect the overall nutrient intake. Some infants may be poor feeders due to hypotonia. As the child grows, delays may be seen in the introduction of textured foods and the transition to feeding themselves table foods. Concerns have also been raised about low energy intake, which may result from efforts to prevent excessive weight gain. This low energy intake may lead subsequently to a low nutrient intake. With adequate monitoring and teamwork with the family, issues around feeding and nutrition can be identified early and appropriate interventions considered. Table 4 provides some general guidance to support positive development of food habits and prevention of excessive weight gain.

Table 4 Nutrition recommendations for children with Down syndrome

Establish early acceptance of a variety of foods
Develop a predictable eating pattern, i.e., three meals a day and snacks, if needed
Eat together as a family as often as possible; do not expect the child to eat alone
To establish good food habits early, limit access to non-nutritious foods
Carry low-energy foods, e.g., vegetable juice, fruit, when away from home
Have water readily available to drink
Avoid the use of food reinforcers in educational settings or for behavior management
Physical activity limits excessive snacking while increasing energy expenditure. (Note: To prevent aspiration, children should not be engaged in physical activity while eating.)

Adapted from Pipes P and Powell J (1996) Preventing obesity in children with special health care needs. *Nutrition Focus Newsletter* 11(6): 1–8.

Energy and Nutrient Intake

Since body mass and rate of growth determine basal energy expenditure, individuals with DS require less energy than their peers. Several studies of 5–11-year-old children, living either in an institution or at home, reported less energy intake than what is recommended for their age groups, or compared with the intake of their typically developing siblings. Obese teenagers with DS consumed 12–13 kcal (50–54 kJ) per cm (between the 10th and 50th percentile of energy intake for typically developing teens).

Over the years, nutrients of concern in children with DS have included calcium, iron, copper, zinc, and vitamins A, C, and E. Some reports have shown the overall intake of vitamins and minerals to be lower in children with DS compared with their controls. This is a concern for those children with a lower weight/height ratio and may be related to food/energy restriction. Earlier vitamin A studies demonstrated a lower intake, decreased absorption, and decreased plasma levels of retinol. How this might relate to abnormal immune response remains speculative.

A recent study by Hopman of 44 children (birth to 4 years) with DS indicated that, while there was delayed introduction of solid foods, the overall nutrient intake was adequate. Iron was the exception, and was low for both the subjects and controls. Energy intake was 27% below the RDA, but when assessed using energy per kilogram of body weight, this group reached the recommended level (102 kcal (427 kJ) per kg). These young children with DS as a group weighed less than their controls and were younger than those previously studied. With concerns about low energy intake and the risk of inadequate nutrient intake, a nutrition referral is appropriate for children with Down syndrome, especially in the early months and years of life.

Feeding Problems

Feeding difficulties and concern about related nutrient intake occur most often during the infant, toddler, and preschool years. If feeding issues continue into older childhood, they may have multiple causes and be more severe. Early feeding problems may include difficulties with coordination of suck and swallow. Most infants with DS are not ready for semisolids by the usual 4–6 months of age. They are usually not able to sit without support due to the hypotonia, and immature oral motor abilities make eating from a spoon difficult, if not unpleasant. If solids are not offered when the child is developmentally ready, however, subsequent feeding problems can occur. Delays in introducing solid foods due to oral motor issues sometimes can lead to a refusal to progress and/or chew more textured foods when the child is developmentally able to do so.

Because of the delayed development, parents often do not have adequate guidance or cues regarding when and how to progress in food textures and self-feeding skills. Feeding problems can also develop from negative experiences and interactions around feeding, i.e., choking or gagging, force-feeding. Infants and children with DS benefit from periodic screening to assess feeding development with subsequent referral to therapists or a feeding team if concerns are present. This screening should begin in infancy.

Supplement Use in Down Syndrome

The use of alternative treatments, such as supplemental vitamins and minerals, is relatively common in individuals with developmental disabilities. Vitamin supplements, with varying formulations and sometimes at high levels, have been promoted for persons with Down syndrome for at least 60 years. Related to the theory of increased oxidative stress, there have been various studies using individual nutrient supplements, i.e., zinc, vitamin A, selenium, and vitamin B_6. The results of these studies were inconsistent, and most had major methodological limitations.

The current popular approach is 'targeted nutrition intervention.' The supplement promoted promises to 'alleviate certain harmful symptoms of Down syndrome (e.g., the susceptibility to infections), and attempt to keep other harmful effects of the syndrome (e.g., mental retardation) from getting worse.' It contains vitamins and minerals for which there is a dietary reference intake (DRI) or a recommended dietary allowance (RDA), amino acids, and other compounds such as coenzyme Q10 and papain, which have no established dietary requirement. When provided as recommended for a child's weight, the supplement does not provide excessive doses of fat-soluble vitamins, which could lead to toxicity. However, some of the water-soluble vitamins are excessively high for infants, i.e. 60 times the dietary reference intake for vitamin B_{12}. Although such high levels are not reported to be deleterious in adults, studies in infants and young children are virtually nonexistent. Therefore, monitoring of these supplements is indicated, as well as education and counseling for families regarding recommended nutrient intakes for infants and children.

Anecdotal reports and testimonials of the benefits of these and similar supplements are widely distributed in lay publications and on the Internet, but no controlled studies have been conducted on these supplements.

When a similar supplement was popular in the 1980s, several double-blind, controlled studies using

the same protocol were conducted with children with DS. There was no difference in the health, development, and cognition between the children receiving the supplements and those who received the placebo. All groups showed improvements in all areas, as would be expected over time.

Health and education professionals working with families who are using such supplements need to be supportive and open, yet make sure that the practice does no harm. They should ask: Is the product safe? Does it work? What is the cost to the family? Are diet, feeding, and growth concerns being addressed at the same time? Is the child still obtaining education and needed therapies?

Other Nutrition-related Issues

Constipation

Constipation is a frequent problem due to hypotonia and decreased motor activity. The colon may retain stool longer leading to loss of water and result in hard, dry stools and long periods between stooling. Typical interventions include increasing the amount of fiber-containing foods and fluids in the diet and participating in appropriate exercise. If feeding problems are present, dietary treatment may be a challenge. Behavior-management techniques are sometimes helpful. Additional medical management may be needed with the use of stool softeners or addition of medical fiber products. Generally, the use of laxatives, suppositories, etc. is not appropriate for children, unless the child's primary care provider is consulted.

Children with DS have an increased incidence of Hirschsprung disease. In this condition, nerve endings are missing in the colon, so stools are not pushed along in the large intestine. This condition is usually identified early in life with subsequent surgical intervention.

Celiac Disease

Celiac disease may occur in 7–16% of children with DS, mostly documented from European sources. Individuals with DS may be predisposed to this condition because of the known increased incidence of autoimmune disorders. Symptoms usually resolve following institution of a gluten-free diet.

Dental Issues

Persons with DS often have smaller jaws and palates. Protruding of the lower jar can lead to poor alignment, which can affect chewing. The tongue also may be increased in size (or the jaw may be too small to support the tongue) and be more furrowed, leading to crevices that retain food and plaque. This can foster bacteria growth and/or bad breath. Dry mouth (xerostomia) is common and is believed to be due to reduced saliva as well as an open mouth and mouth breathing due to the tongue and jaw size.

Tooth eruption is delayed in children with DS until 12 and 18 months of age. There is also a risk for enamel abnormalities that can cause pitting and retention of food and plaque. Gum disease also occurs with greater frequency.

These problems can impact the individual's ability to eat and chew. With the increased risks of dental problems that can impact food intake, early and regular tooth and gum care is essential. For children with DS, the first visit to the dentist is recommended immediately after the first teeth erupt or by the first birthday.

Physical Activity

Individuals with DS are known to be less active than peers and to spend significantly more time indoors. Hypotonic children tire easily and use movement patterns requiring the least expenditure of energy. However, hypotonia tends to improve with age and therapy. The low level of physical fitness is partially explained by findings such as blood pressure, which does not rise regularly with the workload increment. The Health Care Guidelines for Individuals with Down Syndrome recommend establishing regular exercise and recreational programs early in the child's life. Enrollment in an early intervention program, which includes a focus on gross motor activities, is usually recommended within the first months of life.

Many children with DS have few restrictions in physical activity and should be encouraged to be active to increase their function, improve fitness, expend energy for weight management, and have fun. Sports should be limited to noncontact sports that challenge development, coordination, balance, and agility, and are fun to encourage life-long participation. The Special Olympic programs fosters life-long participation in physical activity and has many social rewards.

Approximately 10–40% of individuals with DS have atlantoaxial dislocation (ligaments in the cervical vertebra are more relaxed) resulting from hypotonia. This instability may increase the risk for spinal cord compression and injury. Activities that cause or require significant flexion of the neck, such as high jumping, gymnastics, diving and the butterfly stroke, should be avoided or done under medical supervision.

Promotion of Health and Nutrition

Optimal nutrition and growth, a healthy weight, and overall wellness are positive goals for persons with DS

and their families/caregivers. These are achieved not only by appropriate health care, but also by health and nutrition curricula in educational settings, and wellness programs in vocational and work sites. Key elements in a comprehensive wellness program for individuals with DS throughout the life span include:

- access to accurate and consistent nutrition, wellness, and fitness information and anticipatory guidance;
- interdisciplinary health care teams who develop trusting relationships with families and support access to scientific information and interpretation of that information;
- focus on good food choices and appropriate physical activity (beginning in infancy) to prevent overweight and obesity;
- practical nutrition education (appropriate to cognitive level) related to good food choices, basic food preparation, and physical activity for older children and adults with DS in the community.

See also: **Developmental Disabilities and Nutritional Aspects**: Prader–Willi Syndrome; **Gene Expression and Nutrition**; **Pregnancy**: Maternal Diet, Vitamins, and Neural Tube Defects; **Slimming**: Metabolic Conseqences of Slimming Diets and Weight Maintenance

Further Reading

American Academy of Pediatrics (1994) Health supervision for children with Down syndrome. *Pediatrics* 93: 855–859.

Ani C, Grantham-McGregor S and Muller D (2000) Nutritional supplementation in Down syndrome: theoretical considerations and current status. *Developmental Medicine and Child Neurology* 42: 207–213.

Cohen WI (ed.) (1996) Health Care Guidelines for Individuals with Down Syndrome (Down Syndrome Preventive Medical Check List). *Down Syndrome Quarterly* 1: 2 (available at http://denison.edu/dsq).

Cronk CE, Crocker AC, Pueschel SM *et al*. (1988) Growth charts for children with Down syndrome 1 month to 18 years of age. *Pediatrics* 81: 102. (Note: the birth to 36 month chart for males and females has errors in the weight increments).

Hobbs CA, Sherman SL, Yi P *et al*. (2000) Polymorphisms in genes involved in folate metabolism as maternal risk factors for Down's syndrome. *American Journal of Human Genetics* 67: 623–630.

Hopman E, Cszimadia CG, Bastiani WF *et al*. (1998) Eating habits of young children with Down syndrome in The Netherlands: Adequate nutrient intakes but delayed introduction of solid food. *Journal of the American Dietetic Association* 98: 790–794.

James SJ, Pogribna M, Pogribny IP *et al*. (1999) Abnormal folate metabolism and mutation in the methylenetetrahydrofolate reductase gene may be maternal risk factors for Down syndrome. *American Journal of Clinical Nutrition* 70: 495–501.

Luke A, Sutton M, Schoeller DA and Roizen NJ (1996) Nutrient intake and obesity in prepubescent children with Down syndrome. *Journal of the American Dietetic Association* 96: 1262–1267.

National Down Syndrome Congress. Atlanta, GA (www.ndscenter.org).

Prasher VP (1995) Overweight and obesity amongst Down's syndrome adults. *Journal of Intellectual Disability Research* 39: 437–441.

Sharar T and Bowman ST (1992) Dietary practices, physical activity, and body-mass index in selected population of Down syndrome children and their siblings. *Clinical Pediatrics* 31: 341–344.

Prader–Willi Syndrome

M S N Gellatly, Chelmsford, Essex, UK

Prevalence and Incidence

Prader-Willi Syndrome (PWS) is a complex genetic disorder.

In early 2002, the best estimate of incidence is 1 in 38 000–40 000. Increased awareness of the syndrome and improved diagnostic procedures should not be confused with an increase in prevalence.

PWS occurs in all races and across all ethnic groups.

Characteristics

The documented characteristics of PWS are summarized in **Table 1**. The range and severity of these vary considerably.

Body Composition

People with PWS have an abnormal body composition. Part of the explanation can be attributed to lowered gonadal hormones and their contribution to normal development. The most significant abnormality is a peculiarly high fat to fat-free mass ratio. Bone composition irregularities include lower mineral density, which contributes to the increased risk of osteoporosis, while bone age may be advanced if the child is significantly obese. Hypotonia (poor muscle tone) is primarily caused through central nervous system abnormalities; muscle abnormalities may be observed, such as size variations in type 1 and deficiencies in type 2B fibers.

Table 1 Characteristics of Prader–Willi Syndrome (PWS)

Regulatory
Hypotonia (poor muscle tone)
Hypogonadism (undersized or nonexistent genital organs)
Hypoactivity
Hyperphagia (overeating)
Poor temperature control
Decreased pain threshold and sensitivity
Sleep disturbances
Respiratory disturbances
Rumination
Physical
Distinctive facial characteristics: almond-shaped eyes, high forehead, small downturned mouth, slightly open mouth, prominent nasal bridge
Central obesity
Small hands
Straight ulnar border
Small feet
Dental abnormalities: crowded teeth, poor dental enamel
High palatal arch
Myopia
Strabismus
Cognitive
Learning disabilities
Speech and language delay
Poor articulation
Perseveration (repetitive speech)
Argumentative
Food-seeking
Food-stealing
Compulsions
Stubbornness
Temper outbursts or tantrums
Aggressive behavior to objects and others
Self-injury
Skin-picking
Emotional lability
Other characteristics frequently present in PWS
Obesity
Pale skin
Increased bruising
Reduced motor control
Poor dentition
Reduced salivary flow
Thick, sticky saliva
Absent or irregular menstruation
Undescended testes
Diabetes mellitus
Scoliosis
Osteoporosis
Epilepsy
Narcolepsy
Heart compromise
Cellulitis
Respiratory compromise
Bowel irregularities
Mental health (illness): depression, psychosis, schizophrenia

Modified from Greenswag LR and Alexander RC (1995) *Management of Prader–Willi Syndrome*, 2nd edn. New York: Springer-Verlag; and Waters J (2001) *Babies and Children with Prader–Willi Syndrome*, revised edn. USA: Prader–Willi Syndrome Association.

Endocrine System

Growth hormone and insulin-like growth factors are low in PWS. Gonadotrophins, estrogen, and/or testosterone are usually low in PWS. Increased visceral fat is one cause of glucose intolerance and may lead to diabetes mellitus.

Growth hormone treatment is licenced for children with PWS in many countries, including the UK, USA, and Scandinavia. In addition to aiding linear growth, body composition is altered by increasing the fat-free to fat mass ratio. Early indications from ongoing studies show that adults can also benefit from this treatment.

Neurotransmitters and Peptides

Several of these have been found to be abnormal in PWS. Plasma cholecystokinin (CCK) is raised, yet there is a resistance to the satiety effects. Gamma-aminobutyric acid (GABA) has also been found to be raised in plasma. Leptin levels have been found to be different from those in matched controls in several studies, although the significance of alterations is as yet inconclusive. Increased turnover of serotonin in PWS has been found, and this may contribute to the loss of satiety.

The Syndrome and its Management

The condition is characterized by two distinct phases with multisystem involvement. The hypothalamus is the probable site of major dysfunction as there are numerous alterations to the body's homeostatic mechanisms.

Decreased fetal movement is common. Infants present at birth with severe hypotonia. They are frequently unresponsive to their own needs or to their environment.

Failure to thrive during Phase 1 is common; special feeding measures, e.g., nasogastric tube feeding, are usually necessary, but respiratory support, such as artificial ventilation, is not.

Phase 2 is characterized by hyperphagia (overeating), usually developing between 1 and 6 years of age – although this may be earlier or later – resulting in life-threatening obesity unless energy intake is restricted. The phenomenon of food-seeking and stealing, and surreptitious eating, exhibited in this population is poorly understood by many professionals.

Motor and language development are delayed. Learning difficulties range from very mild to severe.

The placid infant develops into a smiling, friendly child. Most children and adults are loving and

warm; nevertheless, behavior problems are frequently exhibited, e.g., short-lived temper outbursts, stubbornness, manipulative tendencies, including attention-seeking and obsessions, frequently focused on, though not restricted to, food. Sometimes such problems develop into major mental health issues. There appears to be an increased risk of developing recurrent psychoses; these must be distinguished from behavioral phenotypes. Perseveration (repetitive speech) is common; it is unclear if this has an underlying physical or psychiatric basis.

Almost without exception, adults are unable to live completely independent lives.

Prognosis

Gross obesity, leading to premature death, is avoidable and not inevitable. Individual treatment plans and/or programs should be devised and provided in a supportive environment. The management program should include good dietary management at an appropriate energy intake level, to insure weight is kept within reasonable limits.

The most significant development in recent years is the beneficial effect of growth hormone treatment.

Medical Complications

As well as affecting life expectancy, many of these medical complications not only affect health but can precipitate an alteration in living conditions, e.g., to emergency nursing care; crisis management rarely provides the best alternative living environment.

Diabetes Mellitus (DM)

Diabetes mellitus has been frequently reported – almost entirely Type 2 (noninsulin-dependent) – directly linked to obesity. Until weight loss occurs, insulin therapy may need to be prescribed. Blood pressure control, along with no smoking, may be more difficult to achieve in the PWS population.

Osteoporosis

The recorded incidence of osteoporosis appears to be on the increase; this may be due entirely to a longer life span, influenced by lack of hormonal secretions in females, or may simply be more readily recognized as a complication of PWS.

Skin-Picking

Skin-picking can be severe. Infected sores, which were originally minor spots, bites, or operation scars, remaining as open wounds for months or even years, may be related to reduced pain threshold.

Sleep Apnea

Oxygen deprivation leads to sleep disturbances and somnolence; control of weight helps reduce the severity.

Pulmonary Involvement

There are reports that sudden onset of severe chest infection and/or pneumonia can result in death, extremely suddenly, at any age.

Diagnosis

Clinical Diagnosis

Clinical diagnosis criteria were agreed in the early 1990s, and continue to be revised (Table 2).

Table 2 Clinical diagnostic criteria

Major criteria: awarded 1 point each
- Hypogonadism; genital hypoplasia; pubertal deficiency
- Infantile central hypotonia
- Infant feeding problems – failure to thrive
- Rapid weight gain between the age of 1 and 5
- Characteristic facial features
- Developmental delay/mental retardation

Minor criteria: awarded ½ point each
- Small hands and feet for height and age
- Narrow hands with straight ulnar border
- Hypopigmentation
- Myopia, esotropia
- Thick, sticky saliva
- Typical behavior problems
- Skin-picking
- Decreased fetal movement and infantile lethargy
- Short stature compared with family at age 15
- Sleep disturbances/apnea
- Speech articulation defects

Supportive criteria[a]*: no points*
- Temperature control problems
- Unusual skill with jigsaw puzzles
- Scoliosis/kyphosis
- High pain threshold
- Decreased vomiting
- Early adrenarche
- Normal neuromuscular studies
- High pain threshold

Diagnosis should be strongly suspected:
- In those under 3 years of age with five points, three of which must be from major criteria
- In those over 3 years of age with eight points, four of which must be from major criteria[b]

[a]Some clinicians consider more emphasis should be given to some of these.

[b]Smith A, Haan E, Warne G *et al*. (1998) Prader–Willi syndrome: a new study of the Australian Paediatric Surveillance Unit. *Journal of Paediatric Child Health* 34(4): 398–399 require for those:
- under 3 years of age, three major and two minor findings
- 7over 3 years of age, four major and six minor criteria.

Genetic Diagnosis

Clinical diagnosis should be confirmed with genetic studies. PWS is caused by the absence of normally expressed genes on the paternally derived chromosome 15, on the long arm (q11 – q13). The majority with PWS – approximately 70% – have a deletion observed during fluorescence *in situ* hybridization (FISH). Some 29% have maternal uniparental disomy (UPD), confirmed by molecular studies. The remaining 1% have neither deletion nor disomy but an abnormality in the imprinting center in q11–q13. About 5% with deletion or disomy have a translocation or other structural abnormality causing the abnormality in the Prader–Willi critical region (PWCR).

Clinical Differences

Differences are emerging between those with deletion and disomy PWS. However, many early findings are still subjective both in reporting and interpretation. Some of these currently indicate that birth weights seem to be lower in those with deletion, and that a shorter time for tube feeding is needed amongst those with UPD, possibly reflecting a lesser degree of hypotonia.

African-Americans are less likely to have small hands and feet.

Nutritional Management

Early appropriate intervention is essential. Sadly, many individuals still do not receive correct nutritional and dietary management at an early stage.

Whenever a diagnosis is confirmed (or even suspected), dietary treatment should commence.

Poor feeding leading to failure to thrive and hyperphagia leading to rapid onset of weight are the complete antithesis. Yet in individuals with PWS, both are present at different stages of development; nutritional management must therefore encompass these two extremes.

PWS is clearly an eating, and not simply a weight, disorder. It is the most common example of genetic obesity secondary to an impaired satiety response, and must neither be confused nor compared with bulimia nervosa in causality or treatment.

Phase 1

Neonates have a decreased suck, thus breast-feeding is rarely successful. Special methods of feeding, e.g., nasogastric tube feeding, are required almost without exception. Some mothers choose to express breast milk; others find this option impossible to maintain and may therefore need reassurance that infant formula is acceptable.

Transition from tube- to bottle-feeding should not be delayed longer than necessary, but an adequate oral intake is essential before finally removing the tube.

Milk delivery at the right pace is essential; attention should center on three reflexes: lip closure, sucking, and swallowing. In addition to diminished suck, poor head and neck control is common – due to hypotonia – and exacerbates the difficulties encountered during feeding. Advice about positioning the infant is important. Small frequent feeds are advisable.

Infants with PWS rarely wake for feeds and are normally unresponsive to both their own needs and their surroundings.

Alternative Bottles and Teats

Use of a special bottle, e.g. Haberman, Chicco, Mead Johnson Soft, has been found to be invaluable for many parents and infants, although professional opinion as to their suitability is divided. Thus, caution when recommending these is needed. Special teats, supplied with the above bottles, or lamb's or orthodontic types fitting normal feeding bottles may also aid feeding, although it must be noted again that professional opinion as to their advisability is divided.

Provided a safe swallow is present, feeding systems with a soft spout may be suitable.

Cup Feeding

Not only does this method fit with the baby-friendly hospital initiative, but it is recommended for use with sick and low-birth-weight infants by the World Health Organization and UNICEF.

Spoon-Feeding

Sometimes it may be preferable to give liquids from a spoon.

Supplements

It may be necessary to increase the energy density of the feed to insure an adequate energy level is taken. Alterations to feed concentration must always be done under the supervision of a professional, e.g. dietitian or health visitor.

Thickening feeds may help and there are several commercial products available. Again, this should never be done without professional supervision.

Advice relating to vitamin and mineral supplements needs to be given on an individual basis.

Weaning

Introduction of solids may commence around 4 months; parents/carers may need reminding of the

Table 3 Signs of poor feeding

- Poor weight gain in infancy
- Slow intake
- Excessive time spent feeding the child
- Multiple swallows to clear mouth
- Irritability during feeding
- Frequent respiratory infections
- Excessive wet burps
- Excessive drooling
- Family fatigue from special food preparation
- Coughing while swallowing
- Difficulty in moving to new textures so food needs thickening
- Delay in accepting new textures, e.g., more lumps
- Chronic constipation
- Excessive weight gain – due to limited variety of foods because of inability to accept new textures

Adapted from Morris M (1993) Feeding the young child with PWS. *The Gathered View* XVIII(1): 6–7.

risks of commencing solids too early, particularly as infants with PWS often take solids from a spoon better than liquid from a bottle. Introduction of solids not only introduces the infant to new foods and tastes and textures, but assists with mouth muscle development, which is necessary for later speech development. Without good feeding skills it is unlikely that good articulation skills will be present.

Signs of poor feeding within this population are subtle and are summarized in Table 3.

Fluids

Fluid intake may continue to be compromised in toddlers and young children, so particular attention to insure an adequate intake is essential. Thickening of fluids may need to be continued.

Sugars and other concentrated carbohydrate, e.g., honey and sweetened baby drinks should be avoided both to lessen the risk of a 'sweet tooth' or dental caries developing, and to limit energy intake.

Energy Intake

Carers (normally parents) are usually delighted when the infant develops an appetite and takes food easily. At this stage of development, energy intake needs to be adequate but not excessive; comprehensive nutritional assessment by a qualified professional (usually a dietitian) who understands the syndrome is necessary to insure appropriate intakes.

However, it is not uncommon for parents to contribute unwittingly to the failure to thrive – they are so anxious to prevent overweight or obesity, they actually 'starve' the child, who fails to gain weight or achieve expected linear (height) growth. Theoretical needs must be matched by actual intakes. As weight gain can be extremely rapid – there are reports of excess fat deposits preceding the increase in appetite – achieving the right balance is not easy.

Weight should be recorded regularly, at a minimum monthly, but in infants under 12 months, weekly weighing is recommended. Length/height measurements are necessary. Results should be plotted on standard height/weight or body mass index charts – developmental records quickly show any deviation from normal patterns of growth. Linear growth should continue on the same centile, even though this is well below the average; weight should continue along the same centile and not be allowed to become more than two centiles higher than the height one. PWS-specific growth charts have been devloped in some countries.

Energy Requirements

An appreciation of the abnormal body composition is essential; this provides a partial explanation of the need for a lower energy intake. Observations indicate that, while normal children grow on 11–14 kcal (46–59 kj) cm^{-1} height, those with PWS require only 10–11 kcal (42–59 kj) cm^{-1} height. To achieve weight loss, levels of 7–8 kcal (29–33 kj) cm^{-1} height may be needed.

Phase 2

To follow any dietary regime is difficult. Adhering to a strict regime for life is particularly restrictive. For people with PWS not only is an energy-controlled diet necessary, but it is made even more difficult because of the inherent lack of satiety. Teenagers with the syndrome do not exhibit the normal pubertal growth spurt.

Obsessions around food further compromise successful management. Such obsessions with food must never be underestimated and may include:

- needing to know what is for lunch and supper at breakfast time
- creative play revolving around making meals
- collecting recipes and pictures of food
- wanting to watch every food and cookery program on TV

Nevertheless, children can be taught, for example: what is/is not acceptable

- e.g. quantities of food consumed at mealtimes
- what is a 'good' food and what should be avoided because of a high energy density
- normal social behavior, e.g., it is OK to want something to eat late afternoon but not immediately after lunch
- to eat slowly, replacing cutlery on the plate between each mouthful and not gorging

Unsurprisingly, there is no single successful diet. The one that should be followed is that which is acceptable to client and carer, fits in with the family (residential or group home) eating pattern, and provides all macro- and micronutrients at the right energy level to maintain growth and/or weight. The current emphasis amongst professionals for 'healthy eating' regimes at an unspecified energy level is unhelpful to both those with PWS and parents/carers supporting them, because of the satiety problems.

It is possible for older people with PWS to lose weight and return to an acceptable agreed level, provided they are in a suitable environment with understanding and sympathetic support. A body mass index of 30 is an achievable and realistic goal for adults.

As weight reduces, mobility increases; increased self-esteem must not be undervalued.

Fluids

Fluid intake needs to be at the right level. Water softeners, which exchange calcium with sodium salts, must never be used with infants and young children as high sodium intakes would be dangerous. Caution must be observed with artificially sweetened drinks to insure acceptable daily intake levels of such sweeteners are not exceeded.

Water intoxication has been reported (although only as a side-effect or in conjunction with medication), resulting in coma. Thus, statements such as 'you can drink as much as you like' need to be tempered, e.g., up to 1 liters diluted low-energy squash.

Contributory Factors

Successful management of PWS is not restricted to nutritional considerations.

Accessibility of Food

Food-seeking and stealing are peculiar phenomena, hard for the uninitiated to accept, and difficult characteristics to manage. Environmental controls are frequently essential, e.g., locking food cupboards, refrigerators, and freezers. Other measures, which conflict with normally acceptable social behavior, are often needed. These may include restricting access to money (and shops) and insuring no food is accessible, e.g., in a fruit bowl or in another family member's (or resident's) bedroom.

Some individuals will steal and consume any food they find, including some normally not eaten by humans (pica), e.g., cat and dog food, berries and other garden produce, frozen products, discarded food scraps in dustbins.

Research indicates that denial of food-stealing and consumption is related to cognitive learning problems and is not deliberate deception. Strategies for dealing with such events must be known to all concerned and without exception followed – inconsistency complicates management and confuses the individual. A written contract is helpful.

Special Occasions, Holidays, and Treats

People with PWS should be included in social activities, although the importance of planning ahead and providing the necessary supervision should never be underestimated. The preferable approach is to reduce overall energy intake prior to such events rather than afterwards.

People with PWS can have as rough a day as anyone else and may benefit from an occasional extra treat.

Reinforcers or Tokens

People with PWS respond well to programs that allow collection of points towards a predefined reward, or as a reinforcer. Alternatives to foods are preferable, e.g., books, games, toiletries, make-up; some people with PWS respond well to an additional period of attention and time.

Rules and Routine

Consistency in both these is essential – individuals do not respond well to alterations to routine or inconsistency in overall care.

Occasionally rules have to be made which affect others, e.g., eating only at mealtimes. Any management program must insure all involved in care, both direct and indirect – family, friends, teachers, social workers, neighbors, welfare assistants, club organizers – realize the importance of decisions and adhere to them.

Exercise

Parents can assist with improving muscle development from a very early age. The individual program may be devised by a physiotherapist or be adapted from a scheme such as Portage, which is a programme of exercises to assist the development of handicapped babies and young children. All those with PWS, not only children, must be encouraged to participate in physical exercise daily to assist both with improving muscle tone and increasing body metabolism.

The beneficial effects of growth hormone treatment, both increasing muscle mass and improving muscle tome, makes physical activity easier. Increased activity in turn allows a higher energy intake to be consumed.

Education

Many children commence at a mainstream infant school, progressing to a similar junior/primary school. It is acknowledged that the pace of learning falls off from about the age of 8 years, after which time most children, though not all, have difficulties in keeping up with their peers. Few individuals complete education without additional support (welfare assistants or extra individual teaching) in a mainstream setting. A move to a different school, e.g., for those with mild or moderate learning difficulties, may be the better option. Others may require specialist help for other developmental difficulties, e.g., physical handicap or speech delay, and this is best found in specialist schools.

The need to restrict access to food makes adult vocational training difficult but not impossible. Many young adults have false expectations – some of which educators have encouraged, through ignorance rather than intent – which can rarely be met. Failure results and is frequently accompanied by a rapid and unacceptable weight gain. Yet in the right environment people with PWS have achieved success with qualifications in academic, craft, or practical subjects. Many are extremely good with young children and the elderly, provided appropriate supervision is in place.

Behavioral Management

Frequent references have been made to behavior management programs. Each must have individual needs assessed by the appropriate professionals who then devise a sustainable program. The importance of regular reviews, including dietary management, cannot be overstated.

Residential Options and Independence

Adults with the syndrome may give the impression of being eminently capable of both looking after themselves and making appropriate decisions. This frequently leads to inappropriate or inadequate support. Professionals have a duty to plan suitable care in a safe environment. This may be a group home – the Prader–Willi Syndrome Association (USA) led the field in developing PWS-specific homes – a concept gathering momentum in other countries in Europe and Australia. A care village or community, hostel, or sheltered housing may suit some adults better.

Each residential environment must be considered on its own merits as to whether it can handle the multifarious problems of PWS, alongside other residents, and whether it is suitable for the particular individual.

Placements which break down are generally those where proper insight into the complexities of the syndrome are lacking, by frequently dismissing, through disbelief, these management and support issues.

Realistic expectations are essential and no carer should have guilt feelings about restricting access to food.

Support Groups

There are, at the time of writing (early 2002) support groups for parents and professionals in 41 countries. An international body – International Prader–Willi Syndrome Organization (IPWSO) – was constituted in 1991 and each of the 25 constituent countries has both a parent and professional delegate. International conferences are arranged every 3–4 years, when scientists, professionals, carers, parents, and those with PWS are brought together for lectures, seminars, and workshops.

See also: **Diabetes Mellitus**: Etiology; **Growth and Development**; **Obesity**: Etiology and Diagnosis; **Osteoporosis**; **Peptides**

Further Reading

Cassidy SB (1984) Prader–Willi Syndrome. In: *Current Problems in Pediatrics Year Book*, vol. XIV(1). Chicago: Medical Publishers.

Cassidy SB (1997) Prader-Willi Syndrome. *Journal of Medical Genetics* 34: 917–923.

Descheemaeker MJ, Swillen A, Plissart L *et al.* (1994) Genetic counselling. The Prader–Willi Syndrome: a self-supporting programme for Children, Youngsters and Adults. *Genetic Counselling* 5(2): 199–205.

Dykens EM, Goff BJ, Hodapp RM *et al.* (1997) Eating themselves to death: have 'personal rights' gone too far in treating people with Prader–Willi Syndrome? *Journal of Mental Retardation* 4(35): 312–314.

French M (1999) *Prader–Willi Syndrome. Weighing up the Pounds, Policy, Practice and Possibilities*. Colchester, UK: University of Essex.

Gellatly MSN (2001) Prader–Willi Syndrome. In: Thomas B (ed.) *Manual of Dietetic Practice*, 3rd edn, pp. 492–503. Oxford: Blackwell Science.

Greenswag LR and Alexander RC (eds) (1995) *Management of Prader–Willi Syndrome*, 2nd edition. New York: Springer-Verlag.

Hanchett J and Greenswag LR (1998) *Health Care Guidelines for Individuals with Prader–Willi Syndrome*. USA: Prader–Willi Syndrome Association.

Holland AJ (1998) Understanding the eating disorder affecting people with Prader–Willi Syndrome. *Journal of Applied Research in Intellectual Disabilities* 11(3): 192–206.

Holland AJ and Wong J (1999) Genetically determined obesity in Prader–Willi Syndrome: the ethics and legality of treatment. *Journal of Medical Ethics* 25: 230–236.

Holland AJ, Treasure J, Coskeran P and Dallow J (1995) Characteristics of the eating disorder in Prader–Willi Syndrome: implications for treatment. *Journal of Intellectual Disability Research* 39(5): 373–381.

James TN and Brown RI (1992) *Prader–Willi Syndrome: Home, School and Community*. London: Chapman and Hall.

Waters JA (1996) *Handbook for Parents and Carers of Adults with Prader–Willi Syndrome*, revised edn. Derby, UK: Prader–Willi Syndrome Association.

Waters J (2001) *Babies and Children with Prader–Willi Syndrome*, revised edn. Derby, UK: Prader–Willi Syndrome Association.

DEXTRAN

J N BeMiller, Purdue University, West Lafayette, IN, USA

Source, Structure, and Properties

Dextran is a generic term for a family of glucans made by polymerization of the α-D-glucopyranosyl moiety of sucrose in a reaction catalyzed by the enzyme dextransucrase. The common feature is a preponderance of (1 → 6)-linked α-D-glucopyranosyl units.

Several microorganisms produce dextrans with ranges of molecular weights and with structures varying from slightly to highly branched. Commercial dextran is biosynthesized by the nonpathogenic organism *Leuconostoc mesenteroides* NRRL B-512. The basic reaction catalyzed by dextransucrase is n sucrose → (α-D-glucopyranosyl unit)$_n$ + n D-fructose. Branches arise from position O-3 of the glucosyl units. The degree of branching of commercial dextran is about 5%. About 40% of the side chains are single α-D-glucopyranosyl units; about 45% are two units long, and about 15% contain more than 2 units. The average molecular weight of native commercial *Leuconostoc mesenteroides* NRRL B-512 dextran can range from 9 million to 500 million. Dextran of lower molecular weight is produced for clinical applications. Dextran produces relatively low-viscosity solutions, which distinguishes it from other high-molecular-weight polysaccharides. It is a neutral polymer.

Food Uses

Many uses of dextrans in foods have been described and patented, but dextrans are not permitted as foodstuff additives in the USA or Europe. (The generally recognized as safe (GRAS) status was removed by the US Food and Drug Administration, USA, in 1977 because dextran was not being, and had not been, used as an ingredient.) The principal potential uses of dextran in foods appear to be related to its capacity to prevent crystallization, retain moisture, and provide body, but the fact that it is not being used as an additive indicates that no particularly useful functionality that cannot be provided by another hydrocolloid(s) is known. Small amounts of dextran, primarily produced by species of *Leuconostoc* and *Lactobacillus*, are likely to be present in fermented foods that originally contained sucrose.

Dextran formation is detrimental to sugar (sucrose) production, as dextran inhibits crystallization; increases viscosity and, hence, decreases filterability and heat transfer in evaporators, crystallizers, and pans; and results in poor clarification. As a result, steps are taken to minimize dextran formation by contaminating bacteria. Dextranases may be employed to break down any dextran produced. Prevention of crystallization of sugar in foods, beverages, and confections is a potential use.

Metabolism

Dextrans are not broken down by human digestive enzymes. They are degraded by enzymes of bacteria in the large intestine and the released glucose can be absorbed as well as fermented anaerobically.

Medical and Other Uses

Dextran of average molecular weight of 70 000 is used clinically as a plasma volume expander for the treatment of shock and prevention of impending shock. The antithrombic effect of a preparation of average molecular weight of 40 000 provides a prophylactic treatment for venous thrombosis and pulmonary emboli. These and other applications reveal that dextrans of various molecular weights can be safely injected into the blood stream. A complex of ferric hydroxide and dextran is used in the treatment of neonatal anemia in pigs. A principal use is in the

preparation of gels in bead form for size exclusion, ion exchange, and hydrophobic chromatography.

See also: **Anemia (Anaemia)**: Other Nutritional Causes; **Chromatography**: Principles; **Dental Disease**: Etiology of Dental Caries; Role of Diet

Further Reading

DeBelder AN (1992) Dextran. In: Whistler RL and BeMiller JN (eds) *Industrial Gums*, pp. 399–425. New York: Academic Press.

DEXTRINS

J N BeMiller, Purdue University, West Lafayette, IN, USA

Background

Dextrin is a generic term applied to a variety of products obtained by heating a starch in the presence of small amounts of moisture and an acid. Dextrins can be made from any starch and are generally classified as white dextrins, yellow (or canary) dextrins, and British gums. Each is more water-soluble and produces less viscous solutions or dispersions than its parent starch. Each is produced by combinations of slight depolymerization (hydrolysis) and transglycosylation (molecular rearrangement). Transglycosylation produces more highly branched structures and forms glycosidic linkages not found in native starches. Most dextrins are used as adhesives for paper products. Only white dextrins and only small amounts of them are used in prepared foods. White dextrins are prepared by heating a dried, acidified starch.

Other hydrolytic breakdown products from starch are characterized by their dextrose equivalency (DE), which is the percent of reducing power compared to anhydrous D-glucose (dextrose). The DE value is inversely related to molecular weight, i.e., the degree of polymerization (DP), and is, therefore, an indicator of the degree of hydrolysis. (The DE value of anhydrous D-glucose is 100. The DE value of native starch is 0.) When all products of starch hydrolysis were made by treatment of starch with acid, the properties of a product were largely a function of DE. Now with a variety of treatments and processes being used, two products with equivalent DE values (average degree of polymerization) can have rather different properties because of different distributions of molecular sizes.

Maltodextrins are those products having DE values of less than 20, generally DE 5–19. Syrup solids are those products of starch hydrolysis with DE values of more than 20 that are available as dry powders; in other words, they are dried low-DE glucose syrups.

While dextrins are little used in foods; maltodextrins and syrup solids are used extensively. Both are produced from starch by hydrolysis only, i.e., without molecular rearrangement, and are of lower average molecular weight than either dextrins or acid-thinned (thin-boiling) starches, the latter being slightly depolymerized starches that remain in granular form. The primary difference between thin-boiling starches, maltodextrins, and syrups/syrup solids is the degree of depolymerization. The primary difference between dextrins and thin-boiling starches is the method of preparation.

Production

Dextrins are prepared by heating a starch moistened with dilute hydrochloric acid or heating a moist starch in the presence of gaseous hydrogen chloride until a cold-water-soluble product is formed.

Maltodextrins and syrup solids are prepared in basically the same way as are starch-based glucose syrups, except that the process is stopped at an earlier stage to keep the DE value low. Depolymerization of a starch can be effected with either an acid or an enzyme(s) or by combination treatments. These processes are referred to as acid conversions, enzyme conversions, and combination conversions, respectively. Conversions to produce maltodextrins are usually done in a continuous process. In a typical process, a starch slurry (30–40% dry solids, 17–22 Bé) at pH 6.0–6.5 is pasted, generally in a jet cooker ($\sim$105 °C), followed by an atmospheric flash. Conversion (liquefaction) is accomplished with a thermostable bacterial α-amylase at 95–100 °C. Depolymerization is stopped after 60–120 min by enzyme inactivation, and the pH is adjusted. The solution is filtered, treated with carbon, concentrated, and spray-dried to give a maltodextrin or syrup solid preparation. In

Table 1 Typical analysis of a maltodextrin product

	%
Carbohydrate	94.5
Moisture	5
Protein	< 0.1
Fat	< 0.1
Crude fiber	< 0.1
Sodium	< 0.1
Calcium	< 0.03
Phosphorus	< 0.03
Magnesium	< 0.01
Potassium	< 0.005
Sulfite	< 0.001
Iron	< 0.00025
Zinc	< 0.0001

a lower-temperature batch process, steam is introduced into a starch slurry for pasting, which is accomplished at 80–90 °C. Otherwise the processes are identical, except that a debranching enzyme may be used in the batch process. (*See* **Enzymes**: Functions and Characteristics.)

For an acid conversion, a starch is acidified with a food-grade acid, generally to a pH value of about 2. Hydrochloric acid is preferred because of a preference for sodium chloride in the final product. The slurry is passed through a jet cooker to paste the starch and then through a converter to depolymerize it. Alternatively, the starch can be pasted, the pH adjusted after pasting, and the conversion done in a batch process. The degree of hydrolysis, i.e., the DE value, is controlled by a combination of time, temperature, and acid concentration. The process is terminated via cooling and neutralization. The resulting liquor is clarified, treated with carbon, concentrated, and spray-dried.

A variety of maltodextrins can be produced. The variables in the process are the starch source, the means of conversion (including the specificities of enzymes used), and the extent of breakdown (i.e., the DE value of the product). In general, DE values range from 5 to 19; the amount of D-glucose (dextrose) ranges from 0.5 to 3%; and the moisture level is 4–6% (Table 1). As much as 98% of the components can be of DP 3 or higher.

Properties

Dextrins are cold-water-soluble. Upon drying of a dextrin solution, clear films are obtained. Maltodextrins are generally even more soluble, although special crystalline and less soluble types are produced. Depending on the type, 15–60% solutions can generally be made easily at room temperature. Syrup solids are slightly sweet, hygroscopic, semicrystalline, or powdery amorphous products. Maltodextrins are less hygroscopic because of their lower monosaccharide content. They are available as spray-dried powders and in agglomerated and other forms, all of which are essentially white, free-flowing powders. They have the ability to absorb flavor oils and other nonaqueous liquids. Their flowability, compressibility, and low hygroscopicity make them excellent excipients. They provide good to excellent solution clarity. They provide moderate to very low solution viscosity, and low to extremely low browning. They are generally bland, and generally resist caking. They form protective films with oxygen-barrier properties and provide binding, surface sheen, and high solids content without affecting freezing points. Characteristics and properties that increase as DE increases are bulk density, hygroscopicity, ability to participate in Maillard-type browning reactions, solubility, clarity of solutions, osmolality, freezing-point depression, sweetness, and particle size. Characteristics and properties that decrease as DE increases are abilities to form a film, build viscosity, serve as a binder, and provide body.

Typical properties of typical maltodextrins with average DE values of about 5 are formation of clear solutions at concentrations up to 15% at room temperature, production of significant viscosity at 20–40% concentration, very minimal hygroscopicity and sweetness, and excellent film-forming characteristics. Typical properties of typical maltodextrins with average DE values of about 10 are formation of clear solutions at concentrations up to 30% at room temperature, very low hygroscopicity, and low sweetness and ability to effect browning. Typical properties of typical maltodextrins with average DE values of about 15 are formation of clear solutions at concentrations up to 50% at room temperature, slight sweetness, and low hygroscopicity and ability to effect browning. (*See* **Browning**: Nonenzymatic.)

Especially prepared, low-DE, more crystalline maltodextrins produce a fatty mouthful. They are most often made from potato or tapioca starch; oat flour and bran is another source; rice flour has also been used for their preparation. These products, when hydrated, form soft, spoonable gels with a creamy texture. They generally have DE values less than 3.

Food Uses

Dextrins are usually used to produce protective coatings such as those applied to panned confections. Maltodextrins are used in a wide spectrum of food products, including baby foods, bakery products, beverages, breakfast cereals, canned fruit

products, condiments, confections, dairy products, desserts, dietetic and nutritional beverages, frozen meals, infant formulas, jams, jellies, margarines, meat products, dry mixes of all kinds, peanut butter, pet foods, preserves, sauces, snack foods, soups, and syrups. Applications are divided approximately as follows: as carriers for spray-dried and extruded flavors (20–25% of total market), as bulking agents for flavoring materials and high-intensity sweeteners (20–25%), as fat sparers and replacers (20–25%), in nutritional beverages (15–20%), and in confections and coatings (5–15%). Properties that make them useful in these applications are low hygroscopicity (prevention of caking) while being easily dispersible and readily soluble, generally bland taste, encapsulation of essential oils and other flavors (ability to act as a carrier), film formation (ability to coat), binding action (adhesiveness), provision of lubricity, control of expansion of extruded products, provision of bulking, and texture creation. They prevent sugar crystallization and improve chewiness in soft confections, increase shelf-life and maintain moisture levels in hard confections, and speed up the panning process. They are effective binders and excipients for direct compression confections (and pharmaceutical tablets). They control the freezing point and ice crystal growth in frozen dairy products. They control sugar crystallization and sweetness and add solids to bakery fillings, frostings, and glazes. They provide a chewy texture and extend the shelf-life of fruit leathers and granola bars. Maltodextrins are generally excellent spray-drying aids and are used for agglomeration of food gums. Partially crystalline maltodextrins are effective fat sparers and replacers in low-fat or nonfat dairy products. All maltodextrins are digestible, having an energy value of about $3.8\,\mathrm{kcal\,g^{-1}}$. (*See* **Cereals**: Breakfast Cereals; **Dietetics**; **Fat Substitutes**: Use of Fat-replaced Foods in Reducing Fat and Energy Intake; **Infant Foods**: Milk Formulas; **Infants**: Nutritional Requirements; **Low-fat Foods**: Types and Manufacture; Low-fat Spreads.)

Syrup solids are generally used alone or in combination with maltodextrins when one or more of the following characteristics are desired: browning, freezing-point depression, greater solids content, increased solubility, some sweetness, increased clarity.

Digestion and Metabolism

All products made by simple depolymerization of a starch (food dextrins, thin-boiling starches, maltodextrins, dried syrups, syrups) are nutritive saccharides. They are generally recognized as being safe and nontoxic. Human digestive enzymes (pancreatic α-amylases and maltooligosaccharidases of the epithelium of the small intestine) convert each product into D-glucose, which is absorbed, raising blood sugar levels. The D-glucose is then metabolized, so each of these products provides approximately $3.8\,\mathrm{kcal\,g^{-1}}$. Maltodextrins and syrup solids are the carbohydrates of choice for sports drinks and liquid diet formulations because of the low osmolality of their solutions. These properties and their easy digestibility also make them suitable for infant formulas. (*See* **Carbohydrates**: Digestion, Absorption, and Metabolism.)

See also: **Browning**: Nonenzymatic; **Carbohydrates**: Digestion, Absorption, and Metabolism; **Cereals**: Breakfast Cereals; **Dietetics**; **Enzymes**: Functions and Characteristics; **Fat Substitutes**: Use of Fat-replaced Foods in Reducing Fat and Energy Intake; **Infant Foods**: Milk Formulas; **Infants**: Nutritional Requirements; **Low-fat Foods**: Types and Manufacture; Low-fat Spreads; **Starch**: Structure, Properties, and Determination

Further Reading

Auh JH, Lee HG, Kim JW, Kim JC, Yoon HS and Park KH (1999) Highly concentrated branched oligosaccharides as cryoprotectant for surimi. *Journal of Food Science* 64: 418–422.

Chronakis IS (1998) On the molecular characteristics, compositional properties, and structural-functional mechanisms of maltodextrins: a review. *Critical Reviews in Food Science* 38: 599–637.

Schenck FW (1996) Solid starch hydrolysates. *Cereal Foods World* 41: 388–390.

Schenck FW (1996) Solid starch hydrolysates (monograph). *Starch/Stärke* 48: 188–190.

Schenck FW and Hebeda RE (eds) (1991) *Starch Hydrolysis Products: Worldwide Technology, Production and Application*. New York: VCH Publishers.

Van Beynum GMA and Joels JA (eds) (1985) *Starch Conversion Technology*. New York: Marcel Dekker.

Wang Y-J and Wang L (2000) Structures of commercial maltodextrins from corn, potato, and rice starches. *Starch/Stärke* 52: 296–304.

Dextrose *See* **Glucose**: Properties and Analysis; Function and Metabolism; Maintenance of Blood Glucose Level; Glucose Tolerance and the Glycemic (Glycaemic) Index

DIABETES MELLITUS

Contents

Etiology

J Wright, University of Surrey, Guildford, Surrey, UK

Definition and Classification

Diabetes mellitus is a complex disorder of metabolism (fat and protein as well as carbohydrate) caused by either a deficiency of, or defective action of, insulin. It is not a single disease entity, but rather a syndrome of variable clinical presentation and etiology (**Table 1**). (The terms insulin-dependent diabetes mellitus (IDDM) and noninsulin-dependent diabetes mellitus (NIDDM) should now be abandoned in favor of type 1 and type 2 diabetes.)

Basic Features of Diabetes

Type 1 (insulin-dependent) and type 2 (noninsulin-dependent) diabetes account for the great majority of patients. They differ in terms of their clinical presentation, etiology, and basic metabolic defect. Type 1 diabetes is characterized by absolute insulin deficiency owing to B-cell destruction, and type 2 diabetes by relative insulin deficiency due to a combination of dysfunctional insulin secretion and resistance to insulin.

Successful treatment of diabetes with diet, oral hypoglycemic drugs, or insulin enables the majority of patients to lead long and near-normal lives. Increasingly, it is the long-term complications of the disease which are the major challenge in the treatment of diabetes. The most important of these are the microvascular complications (retinopathy and nephropathy), the macrovascular complications (atherosclerosis, coronary artery, and peripheral vascular disease), and neuropathy. These complications are uncommon in type 1 diabetes of less than 10 years' duration but are not uncommon at presentation in type 2 diabetes, probably because of the long period of asymptomatic hyperglycemia which usually precedes the clinical diagnosis. In the majority of patients it is relatively easy to achieve good or at least reasonable health in the short term. However, with increasingly convincing evidence that the incidence of long-term complications can be reduced by maintaining good glycemic and metabolic control, this has become the major emphasis in the clinical monitoring and management of diabetes.

The major differences between types 1 and 2 diabetes are shown in Table 2.

Table 1 Etiological classification of diabetes

Type 1: insulin-dependent diabetes mellitus (IDDM)[a]
Autoimmune
Idiopathic
Type 2: noninsulin-dependent diabetes mellitus (NIDDM)[a]
Other specific types
Genetic defects of B-cell function
Genetic defects in insulin action
Diseases of the exocrine pancreas
Endocrinopathies
Drug- or chemical-induced
Infections
Uncommon forms of immune-mediated diabetes
Other genetic syndromes sometimes associated with diabetes
Gestational diabetes

[a]The terms IDDM and NIDDM should now be abandoned in favor of type 1 and type 2 diabetes.
Based on American Diabetic Association Expert Committee Report (1997).

Table 2 Major features of types 1 and 2 diabetes

	Type 1 diabetes	*Type 2 diabetes*
Age at onset (years)	Usually < 30	Usually > 40
Weight	Weight loss	Normal or overweight
Rate of onset	Rapid	Slow
Ketosis	Usual	Absent
Complications at presentation	Rare	Common
HLA association	Usual	Absent
Islet cell antibodies	Usual	Absent

HLA, human leukocyte antigen.
Adapted from Home PD, Johnston DG and Alberti KGMM (1989) Diabetes mellitus. In: Hall R and Besser G (eds) *Fundamentals of Clinical Endocrinology*, pp. 318–361. Edinburgh: Churchill Livingstone.

Clinical diabetes represents a continuous spectrum from the young, underweight, ketotic, and sometimes comatose patient on the one hand to the elderly, obese and frequently asymptomatic patient on the other. The majority of patients with types 1 and 2 diabetes conform to the patterns in Table 2 but there are frequent exceptions; it is increasingly recognized, for example, that the elderly sometimes develop typical type 1 diabetes. There is more uniformity in young patients, almost all of whom require insulin. An exception is the syndrome of maturity-onset diabetes of the young (MODY), a group of rare inherited disorders which usually present before the age of 25 years, and which, in the first instance at least, do not usually require treatment with insulin.

Diagnosis

Although both insulin deficiency and insulin resistance result in abnormalities of fat, protein, and carbohydrate metabolism, it is upon the changes in blood glucose levels that both diagnosis and monitoring are primarily based.

Until recently, the diagnosis of diabetes was based upon the criteria set by the World Health Organization (WHO), using a standard oral glucose tolerance test (OGTT). In the vast majority of patients, a GTT is unnecessary, and the diagnosis is made on the basis of clinical history, and the random and/or fasting blood glucose level, which are usually unequivocal. (The exception is pregnancy, during which screening for gestational diabetes using a GTT is common practice.) However, strict criteria and a standard diagnostic procedure are important, particularly for epidemiological purposes. The interpretation of a GTT is based on the glucose level in the fasting state and 2 h after a 75-g glucose load (or equivalent); neither the 1-h glucose level nor measurements of urine glucose or plasma insulin are of diagnostic use. The glucose load should be dissolved in 250–300 ml of water and drunk within 5 min after an overnight fast; dietary carbohydrate should not have been restricted during the 3 days prior to the test.

The WHO criteria have recently been revised, and the new criteria are shown in Table 3. The major reason for the changes is the fact that the 1-h postload plasma glucose level of 11.1 mmol l^{-1} (the most important criterion for the diagnosis of diabetes) correlates better with a fasting level of 7 than 6 mmol l^{-1}. A new category of 'impaired fasting glucose' (fasting plasma glucose 6.1–7.0 mmol l^{-1}) has also been proposed because it has been shown that prevalence of diabetic complications (retinopathy) starts to increase significantly in people with a fasting plasma glucose above 6 mmol l^{-1}. (*See* **Glucose**: Maintenance of Blood Glucose Level; Glucose Tolerance and the Glycemic (Glycaemic) Index.)

Impaired glucose tolerance (IGT) is not a separate etiologic category. It is associated with an increased risk of developing diabetes (10–25% within 5 years) and, importantly, the risk of developing the syndrome of obesity, hypertension, and hyperlipidemia as well as the macrovascular complications of diabetes. Individuals with IGT should therefore be monitored carefully, and treated with dietary restriction and weight loss where appropriate. (*See* **Hyperlipidemia (Hyperlipidaemia)**; **Hypertension**: Nutrition in the Diabetic Hypertensive.)

Table 3 Diagnostic values for the oral glucose tolerance test (WHO, 1999)

	Venous plasma glucose (mmol l^{-1})
Diabetes mellitus	
Random, or	≥ 11.1
Fasting, or	≥ 7.0 (whole blood ≥ 6.1)
2 h after glucose load	≥ 11.1
Impaired glucose tolerance (IGT)	
Fasting, and	< 7.0
2 h after glucose load	≥ 7.8 but ≤ 11.1
Impaired fasting glycemia (IFG)	
Fasting	≥ 6.1 but < 7.0

Epidemiology

The current and predicted prevalence of diabetes worldwide is shown in Table 4. As can be seen, it is anticipated that the prevalence of diabetes will (approximately) double by the year 2010, with the major increase occurring in 'developing' countries. Around 98% of the increase is accounted for by type 2 diabetes.

Type 1 Diabetes

There are dramatic differences in the prevalence of type 1 diabetes, both racially and geographically. The

Table 4 Estimated prevalence of diabetes (millions)

Worldwide 1997	124 m[a]
2010	221 m
West Asia 1995	3.6 m
2010	11.4 m
South/central Asia 1995	28.8 m
2010	57.5 m
South-east Asia 1995	8.6 m
2010	19.5 m
East Asia 1995	21.7 m
2010	44 m

[a] ~2% population; 3.5 m type 1, 120 m type 2.

From Amos AF, McCarty DJ and Zimmet P (1997) The rising global burden of diabetes: estimates and projections to the year 2010. *Diabetic Medicine* 14: S7–S15.

annual incidence rate in young people below the age of 20 years varies from around 30 per 100 000 per year in Finland and Scandinavia to less than 1 per 100 000 per year in oriental populations such as the Koreans and the Japanese. Rates are generally much higher in northern than southern Europe (with the exception of the island of Sardinia, which has an incidence similar to Scandinavia), and higher in whites than blacks from the same geographical area. Racial differences in the prevalence of type 1 diabetes may relate to differences in the frequency of susceptibility genes in the population.

The peak age of onset of type 1 diabetes is between the ages of 11 and 13 years. Thereafter, the incidence rate is low during middle life but there is a further peak in the elderly, in whom the presenting features of the disease may be similar to those seen in children. The etiology in this group may be different, but they frequently have islet-cell antibodies, indicating an autoimmune process. In addition to age-related differences in incidence, there is also a seasonal variation, seen only in older children, with a maximum during the winter months, possibly related to seasonal variations in viral illnesses involved in the etiology of the disease. Of particular concern is the fact that the incidence in children appears to be increasing steadily in some countries, including the UK. The reasons for this are unclear.

Type 2 Diabetes

There are also marked differences in the prevalence and incidence of type 2 diabetes. The lowest rates are found in underdeveloped rural communities in Africa and Asia. The prevalence in most European countries is between 1% and 3%, and it is estimated that there may be a further similar number of cases which are undiagnosed. Unlike type 1 diabetes, surveys in the USA show that the prevalence of type 2 diabetes is higher among blacks than among whites.

The anticipated explosion in the prevalence of type 2 diabetes is already evident in certain ethnic groups who have been subjected to rapid 'westernization.' The most extensively studied are the Pima Indians in Arizona and the Micronesian population of the Pacific island of Nauru where inheritance appears to be autosomal dominant, and the prevalence in adults may be as high as 80%. Recently, high rates have also been found in a number of migrant populations, particularly in those from the Indian subcontinent to the west. It has been suggested that these peoples possess a 'thrifty genotype' which confers the advantage of metabolic efficiency when food is in short supply but which leads to obesity, insulin resistance, and type 2 diabetes when food becomes plentiful.

Complications and Mortality

The pattern of complications varies between type 1 diabetes and type 2 diabetes. In both, the major cause of death is from cardiovascular disease, but in type 1 diabetes, microvascular complications (nephropathy and retinopathy) remain important, with a significant increase in deaths from renal disease and its sequelae. Similar microvascular complications occur in type 2 diabetes, but the increased mortality is largely attributable to deaths from cardiovascular disease. These differences probably reflect fundamental differences in the pathophysiology of the two conditions.

Causes

The underlying lesions in type 1 and 2 diabetes are different and result from different pathogenic processes. In the majority of cases of type 1 diabetes there is loss of insulin secretion due to autoimmune destruction of B cells, while in type 2 diabetes there is usually minimal loss of B cells but evidence of dysfunctional insulin secretion and peripheral insulin resistance. In both cases the disease process is complex and almost certainly involves genetic, environmental, and possibly dietary factors.

Type 1 Diabetes

Genetic factors About 60% of the genetic predisposition to autoimmune type 1 diabetes is accounted for by genes of the major histocompatibility complex (MHC), particularly the human leukocyte antigen (HLA) genes which code for a series of cell-surface glycoproteins, class I and class II HLA antigens. Class I proteins are expressed on the surface of most nucleated cells and are involved in the presentation of processed antigens to cytotoxic T lymphocytes. Class II proteins are expressed on B lymphocytes, macrophages, and activated T cells, and are involved in the presentation of antigens to T-helper lymphocytes. Class II proteins comprise α and β chains which are encoded for by genes within the DP, DQ, and DR regions of chromosome 6.

The major association with type 1 diabetes is with possession of the HLAs DR3 and DR4 (**Table 5**).

Table 5 Human leukocyte antigen (HLA) associations and the relative risk of type 1 diabetes

HLA-DR antigen	*Relative risk of type 1 diabetes*	*95% confidence limits*
DR2	0.1	0.05–0.3
DR3	5.0	2.9–8.8
DR4	6.8	3.8–12.1
DR3 and DR4	14.3	6.3–32.4

Data from the Barts-Windsor Study (1983).

However, these are too frequent in the general population to be solely responsible for susceptibility to diabetes, and the reason that these haplotypes appear to confer susceptibility is that they are in linkage disequilibrium (i.e., transmitted together) with other true susceptibility alleles, particularly within the DQ region. One particular DQ polymorphism determines the presence of an aspartate residue at position 57 in the DQ8 β chain. The presence of an aspartate residue in this position changes the configuration of an antigen-presenting cleft in the DQ protein; aspartate confers a more efficient antigen-presenting configuration, and thus enhances antigen recognition and the subsequent immune response. In addition to conferring susceptibility, the possession of certain HLA alleles may confer protection (e.g., DQ6, DQ18).

While a considerable amount is known about the MHC/HLA genes, very little is known about the genes which may account for the remaining 40% of genetic susceptibility. There is some evidence that the insulin gene may be a susceptibility locus, but conclusive evidence is lacking.

Environmental factors Twin studies have shown that the concordance for type 1 diabetes between genetically identical individuals may be less than 40%, indicating that other environmental factors are also involved in triggering the autoimmune process. The most likely of these is viral infection, and several candidate viruses have been identified. The best documented is congenital rubella, with type 1 diabetes developing in 10–20% of affected individuals. Others include mumps, and coxsackie B, cytomegalovirus, and Epstein–Barr virus. There may be several mechanisms by which viral infections may cause B-cell damage. These include direct B-cell infection and damage, induction of antigenic determinants in the B-cell membrane, either as a result of incorporation of viral proteins or exposure of autoantigens, and molecular mimicry (immunological cross-reactivity between viral antigens and B-cell components).

Other possible environmental triggers include toxic and/or nutritional factors. Nitrosamines in smoked mutton have been implicated in the pathogenesis of childhood diabetes in Iceland. Currently there is considerable concern that bovine serum albumin (BSA) in cows' milk may be able to initiate an autoimmune response because of cross-reactivity between BSA and a B-cell component; this issue is clearly of major importance but the connection remains unproven.

Immune factors It is likely that type 1 diabetes develops as the result of an environmental injury in genetically predisposed individuals, the combination of which initiates an autoimmune process which results in insulitis and B-cell death. The precise details of this process are unclear but it is probable that an initial event such as a viral infection results in B-cell damage and exposure of antigens which are usually hidden from the immune system. This triggers a macrophage response and subsequent release of cytokines such as tumor necrosis factor and γ-interferon which are capable of inducing aberrant expression of class 2 antigens on B cells, thus enabling B cells to act as antigen-presenting cells, and to present surface autoantigens directly to T cells. Neither A nor D cells appear to be capable of being induced to express class 2 antigens and are therefore spared.

The result of the autoimmune process is the production of autoantibodies which can be detected in the circulation of almost all newly diagnosed cases of type 1 diabetes. Antibodies may be directed against either cytoplasmic antigens (islet-cell antibodies, ICAS) or against surface antigens (islet-cell surface antibodies, ICSAS), and even against insulin itself (insulin autoantibodies, IAAS). They may be complement-fixing and therefore cytotoxic. Candidate cytoplasmic autoantigens include the enzyme glutamate dehydrogenase (GAD), while candidate cell-surface autoantigens include the glucose transporter GLUT-2.

Autoantibodies may be present in the circulation for months or even years before the clinical presentation of the disease, during which time insulin secretion may decline gradually. There may then be a final event which precipitates the development of clinical diabetes, either a further viral infection or reexposure to the original triggering infection, thus explaining the seasonal peaks in the incidence of new cases of type 1 diabetes.

Idiopathic type 1 diabetes In a small proportion of patients, there are no genetic or autoimmune markers, and the underlying cause (or possibly, causes) of diabetes in this group is unknown. This has been termed idiopathic type 1 diabetes. It has been proposed that type 1 diabetes should be subdivided into type 1A (autoimmune) and type 1B (idiopathic) diabetes, but this classification is not yet generally accepted.

Type 2 Diabetes

The major pathophysiological components of type 2 diabetes are dysfunctional insulin secretion and insulin resistance. Indeed, in the majority of patients with type 2 diabetes, the condition is part of a wider syndrome – the 'insulin resistance' or 'metabolic' syndrome (previously known as syndrome X or Reaven's syndrome), the features of which are shown in Table 6. The definition, nature, and assessment of insulin resistance are described in the following

Table 6 Features of the insulin resistance ('metabolic') syndrome

Impaired insulin-stimulated glucose uptake
Hyperinsulinemia
Central male-pattern obesity
Glucose intolerance or type 2 diabetes
Dyslipidemia*
• Increased plasma triglycerides
• Decreased HDL-cholesterol
• Increased small dense LDL particles
Hypertension
Hyperuricemia (gout)
Hypercoagulable state (increased PAI-1 and factor VII)
Polycystic ovary syndrome
Microalbuminuria

*The 'Atherogenic Lipoprotein Phenotype'.
HDL, high-density lipoprotein; LDL, low-density lipoprotein; PAI-1; plasminogen activator inhibitor-1.

Table 7 Other specific types of diabetes

Genetic defects of B-cell function
MODY1: HNF4α (chromosome 20)
MODY2: glucokinase (chromosome 7)
MODY3: HNF1α (chromosome 12)
MODY4: IPF-1 (chromosome 13)
Mitochondrial DNA 3243 mutation
Genetic defects in insulin action
Type A insulin resistance; leprachaunism
Lipoatrophic diabetes
Diseases of the exocrine pancreas
Pancreatitis, neoplasia
Trauma/pancreatectomy, etc.
Endocrinopathies
Cushing's syndrome, acromegaly
Glucagonoma, pheochromocytoma
Drug- or chemical-induced
Nicotinic acid, glucocorticoids, thyroid hormone
α and β adrenergic agonists
Thiazide diuretics, phenytoin, ciclosporin
Ethanol
Infections
Congenital rubella, cytomegalovirus, etc.
Uncommon forms of immune-mediated diabetes
Insulin and insulin receptor antibodies
'Stiff man' syndrome
Other genetic syndromes sometimes associated with diabetes
Down's syndrome, Friedreich's ataxia
Huntington's chorea, Klinefelter's syndrome,
Lawrence–Moon–Biedel syndrome, myotonic dystrophy, etc.

MODY, maturity-onset diabetes of the young.
Adapted from American Diabetes Association Expert Report (1997).

chapter (**Diabetes Mellitus**: Chemical Pathology). Insulin resistance arises from the interplay of a series of genetic factors on the one hand, together with social and environmental factors on the other.

Genetic factors Genetic factors are clearly much stronger in type 2 than in type 1 diabetes. In most cases it is possible to obtain a family history, and concordance is greater than 90% for type 2 diabetes in monozygotic twins compared with around 40% for type 1 diabetes. In high-risk populations, the risk of diabetes in offspring when both parents have type 2 diabetes may be greater than 50%, and as high as 80% in Pima Indians and Nauru islanders when parents develop the disease before the age of 45 years. Despite these figures, much less is known about the genetic associations of type 2 than those of type 1 diabetes. A number of genetic defects have been identified which are responsible for rare diabetic syndromes (**Table 7**) but in the great majority of cases, inheritance is almost certainly polygenic, with environmental factors playing a major role. There is no evidence for association with MHC or HLA markers, and only limited evidence for the involvement of a number of other candidate genes including the insulin gene itself, and insulin receptor and glucose transporter genes.

Social and environmental factors The major non-genetic factors which predispose to the development of insulin resistance are obesity, physical inactivity, nutritional factors, and smoking.

Increasing obesity is associated with deterioration in glucose tolerance, and the majority of patients with type 2 diabetes are overweight: around 75% have a body mass index (BMI) of over 25 $kg\,m^{-2}$. Of course, the fact that not all obese people develop diabetes is evidence that other factors (particularly genetic predisposition) are important. If obesity has a role in the cause of type 2 diabetes, it is probably in modulating the speed of progression to symptomatic disease, possibly by contributing to the development of insulin resistance and the earlier onset of B-cell exhaustion. The risk of developing type 2 diabetes is related to not only the degree but also the distribution of obesity, with the greatest risk seen with male-pattern central obesity. (The precise determinants of fat distribution are uncertain but it too is at least in part genetically determined.) Central adipose tissue is more metabolically active with higher lipolytic activity than peripheral fat, and it is thus a richer source of nonesterified fatty acids (NEFA). Furthermore, the venous drainage of visceral adipose tissue in particular goes directly to the liver. High levels of NEFA stimulate hepatic gluconeogenesis, impair peripheral (muscle) glucose uptake, and inhibit hepatic degradation of insulin, all of which predispose to the development of hyperglycemia and insulin resistance.

Although obesity is an important risk factor for diabetes, both insulin resistance and the hyperinsulinemia which results from it are predictive of diabetes independently of body weight.

Physical exercise influences both glucose transport systems (GLUT 4) and a number of enzyme systems (e.g., glycogen synthase) in skeletal muscle which have a marked effect on insulin sensitivity. The contribution of smoking to the development of insulin resistance is minor but significant.

As far as diet is concerned, the relative roles of fat and carbohydrate in the pathogenesis of insulin resistance and diabetes is unclear and, indeed, controversial. In many countries where the prevalence of diabetes is low, dietary fat intake is also low and dietary carbohydrate tends to be unrefined, with a low glycemic index and a high proportion of soluble and insoluble fiber, all of which mitigate against the development of diabetes. However, overall energy intake tends to be low in this situation and the prevalence of obesity is low. The situation in 'developed' countries is somewhat different. There is general agreement that diets high in saturated fat predispose to the development of insulin resistance, and as far as overall macronutrient balance in the diet is concerned, high-carbohydrate diets appear to increase rather than decrease insulin resistance when compared to diets high in mono- or polyunsaturated fat.

Thrifty phenotype or thrifty genotype? The reason that the incidence of insulin resistance and type 2 diabetes is so high in 'developing' societies has been ascribed to the possession of a 'thrifty' gene which, in an environment with an unstable, poor food supply, confers a survival advantage due to metabolic efficiency. When individuals with this genotype are exposed to a copious western diet, metabolic efficiency predisposes to weight gain, insulin resistance, and diabetes.

An alternative, or perhaps complementary hypothesis (the thrifty phenotype) derives from an increasing amount of evidence which indicates that nutrition *in utero* and in early infancy influences insulin sensitivity in later life; babies who are small at birth and at 1 year of age tend to be more obese in later life, with insulin resistance and a high risk of type 2 diabetes, hypertension, and coronary heart disease. These two hypotheses are clearly not mutually exclusive.

Other Specific Types of Diabetes

In addition to the common patterns of type 1 and type 2 diabetes described above, IGT or diabetes may occur in association with a large number of other conditions, as shown in Table 7. The majority of these conditions are very uncommon or, indeed, rare. In the majority, the precise mechanism of the carbohydrate intolerance is unclear but in some (the various forms of maturity-onset diabetes of the young (MODY), for example), identification of the underlying biochemical defect has provided useful insights into the mechanisms that may be involved in the common polygenic form of type 2 diabetes.

Screening and Prevention

Type 1 Diabetes

Different arguments can be advanced for screening for the different types of diabetes. Screening for susceptible individuals or for preclinical type 1 diabetes might be justified as part of a prevention program if means were available to avert or delay for a significant time the onset of clinical disease. Once glucose intolerance starts to develop, the progression to clinical disease is usually rapid. Measurement of blood or urine glucose is therefore useless, and screening would, of necessity, be limited to testing for susceptible genotypes or the presence of autoantibodies in children or young people with a family history of the disease. Sporadic cases would, of course, remain undetected, and since a high proportion of patients do not have a family history, this would clearly limit the effectiveness of screening. Any such program would be extremely costly and at present there is insufficient evidence that early intervention might be effective. Measures such as immunosuppressive therapy with ciclosporin or treatment with nicotinamide have been shown either to induce limited remission or to slow the rate of deterioration of B-cell function. A further possible approach is to immunize high-risk individuals against viruses capable of precipitating diabetes but this, too, remains an unproven strategy.

Ultimately, preventive measures are likely to involve gene therapy, but until both the technology and ethical issues of such treatment have been resolved, extensive screening for type 1 diabetes is probably not justified.

Type 2 diabetes

Below the age of 45 years, the prevalence of type 2 diabetes in the population is over 3%, and screening on a large scale is probably not cost-effective. However, the nature of type 2 diabetes (frequently asymptomatic), the high prevalence of diabetes (predominantly type 2 diabetes) in older age groups (**Table 8**) and the high prevalence of complications, both at presentation and in the early years after diagnosis, provide some justification for the establishment of screening programs in the older population. Problems arise in deciding when to rescreen subjects who were initially negative, since the pick-up rate will clearly be lower on retesting. In addition to the elderly, screening may be justified in those ethnic

Table 8 Age-related prevalence of diabetes in white Caucasians

Age group (years)	*Prevalence of diabetes mellitus*
0–9	0.5
10–19	1.8
20–29	4.0
30–39	4.6
40–49	7.5
50–59	13.8
60–69	21.3
70–79	37.1
80+	35.6

Modified from Singh BM, Rutter JD and Fitzgerald MG (1988) The natural history of non-insulin-dependent diabetes mellitus. *Clinical Endocrinology and Metabolism* 2: 343–358.

groups in whom there is known to be a high incidence of type 2 diabetes.

To some extent, however, the question of establishing specific screening programs is becoming redundant as testing for diabetes forms part of a more general health screen which general practitioners are encouraged to provide.

Screening can be based upon testing for glycosuria, measuring fasting or random blood glucose levels or upon measurements of glycated hemoglobin (HbA_1). Of these, blood glucose measurements are more precise and probably to be preferred, provided that they are interpreted in relation to the time of meals.

See also: **Diabetes Mellitus**: Chemical Pathology; **Glucose**: Maintenance of Blood Glucose Level; Glucose Tolerance and the Glycemic (Glycaemic) Index; **Hyperlipidemia (Hyperlipidaemia)**; **Hypertension**: Nutrition in the Diabetic Hypertensive; **Immunology of Food**

Further Reading

Amos AF, McCarty DJ and Zimmet P (1997) The rising global burden of diabetes: estimates and projections to the year 2010. *Diabetic Medicine* 14: S7–S15.

Daly ME, Vale C, Walker M, Alberti KGMM and Mathers JC (1997) Dietary carbohydrate and insulin sensitivity: a review of the evidence and clinical implications. *American Journal of Clinical Nutrition* 66: 1072–1085.

Expert Committee (1997) Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* 20: 1183–1197.

Harris MI, Hadden WC, Knowler WC and Bennett PH (1987) Prevalence of diabetes and impaired glucose tolerance and plasma glucose levels in US population aged 20–74 yr. *Diabetes* 36: 523–534.

Home PD, Johnston DG and Alberti KGMM (1989) Diabetes mellitus. In: Hall R and Besser G (eds) *Fundamentals of Clinical Endocrinology*, pp. 318–361. Edinburgh: Churchill Livingstone.

Mahler RJ and Adler ML (1999) Type 2 diabetes mellitus: update on diagnosis, pathophysiology and treatment. *Journal of Clinical Endocrinology and Metabolism* 84: 1165–1171.

Pickup JC and Williams G (eds) (1997) *Textbook of Diabetes*, (2nd edn). Oxford: Blackwell.

Reaven GM (1988) Role of insulin resistance in human disease. *Diabetes* 37: 1595–1607.

Singh BM, Rutter JD and Fitzgerald MG (1988) The natural history of non-insulin-dependent diabetes mellitus. *Clinical Endocrinology and Metabolism* 2: 343–358.

Wolf E, Spencer KM and Cudworth AG (1983) The genetic susceptibility to type 1 (insulin-dependent) diabetes: an analysis of the HLA-DR association. *Diabetologia* 24: 224–230.

Yki-Jarvinen (1997) MODY genes and mutations in hepatocyte nuclear factors. *Lancet* 349: 516–517.

Chemical Pathology

J Wright, University of Surrey, Guildford, Surrey, UK

Morbid Anatomy of the Pancreas

The Normal Pancreas

Insulin is secreted from the B cells of the islets of Langerhans. There are approximately one million islets distributed throughout the pancreas. Individual cells can be identified using immunocytochemistry, whereby cells are stained using antibodies directed at their secretory products. The B cell is the dominant cell type of the islet but there are a number of others, as shown in **Table 1**. There is a complex functional and anatomical relationship between the cells of the islet and their products. A and D cells are situated on the periphery of the islet with a central core of B cells. It is not entirely clear whether arterial capillary blood reaching the B cells has first passed via both A and D cells, transporting glucagon and somatostatin to the B cells where they act as paracrine factors, respectively enhancing and inhibiting the insulin response to stimuli, or whether capillary blood first supplies the B cell

Table 1 Cell types and secretory products of the islet of Langerhans

Cell type	*Secretory product*
A	Glucagon
B	Insulin
D	Somatostatin
PP	Pancreatic polypeptide

with insulin, then exerts a paracrine effect on glucagon and somatostatin secretion. The function of pancreatic polypeptide is not known.

In addition to insulin, B cells also secrete amylin (also known as islet amyloid polypeptide), a 37-amino-acid peptide which is similar to calcitonin gene-related peptide (CGRP, ~50% sequence homology). It is found in secretory granules and appears to be cosecreted with insulin. Initial studies suggested that amylin may induce insulin resistance in skeletal muscle. Whether altered amylin secretion contributes to the metabolic abnormalities of diabetes is as yet unclear. In type 1 diabetes, amylin production is lost as a result of B-cell damage. In type 2 diabetes, amylin secretion is diminished and the peptide accumulates as amyloid fibrils which are found between the B cells and the adjacent capillaries; there is some evidence that this may impair insulin secretion and play a role in the pathogenesis of type 2 diabetes.

Type 1 Diabetes (IDDM)

In the majority of cases, type 1 diabetes is caused by autoimmune destruction of B cells, probably triggered by a viral infection in genetically susceptible individuals (*See* **Diabetes Mellitus**: Etiology). The first visible stages in this process involve the appearance of macrophages within the islets and the establishment of a typical chronic inflammatory reaction with lymphocytic infiltration; this is termed 'insulitis.' Loss of B cells is a gradual process which may first be detectable several years before the development of clinical diabetes. In some patients, residual B-cell function persists for many years after the onset of diabetes, with persistence of insulitis in remaining B cells. Continuing insulin secretion (as assessed by a positive C-peptide response to glucagon stimulation) is characteristically associated with easier glycemic control.

In established type 1 diabetes, there is almost total loss of B cells. As a result, the islets appear atrophic, although the number and distribution of A, D, and PP cells remain almost normal.

Type 2 Diabetes (NIDDM)

Unlike type 1 diabetes, the gross morphology of the islet is normal (with the exception of islet amyloid accumulation). The extent of B-cell loss is variable and usually only minimal, although there is occasionally a loss of up to 50% of B cells; the number of A and D cells remains unaffected.

The Metabolic Effects of Insulin Insufficiency

The major stimulus to insulin secretion is an increase in blood glucose concentration. A number of other physiological factors also stimulate insulin secretion (**Table 2**), in particular the hormones of the enteroinsular axis (glucose-dependent insulinotrophic polypeptide and glucagon-like polypeptide-1) which augment the insulin response to oral carbohydrate and fat, but only in the presence of a raised blood glucose level. The cephalic phase of insulin secretion, which occurs in anticipation of eating, is mediated by the vagus nerve.

Insulin is thus the hormone of the fed state with low circulating levels in the postabsorptive state, and its metabolic actions (**Table 3**) are primarily concerned with the coordinated disposal, metabolism, and storage of the major nutrients. The major effects are upon carbohydrate metabolism, with control of blood glucose level being the most obvious of these. This depends on a balance between output from the liver and peripheral uptake, utilization and storage of glucose; this balance is largely dependent on those tissues which require insulin for intracellular transport of glucose, particularly skeletal muscle. (*See* **Glucose**: Maintenance of Blood Glucose Level.)

Since diabetes is characterized by the absence of effective insulin activity, it is possible to view the major metabolic features of diabetes as a reversal of the actions in Table 3. This can be due to absolute insulin deficiency (type 1 diabetes) or to a combination of dysfunctional insulin secretion plus insulin resistance (type 2 diabetes).

Counterregulatory hormones Insulin is the only major anabolic hormone in the body (growth hormone also has anabolic properties but these are only manifest in the presence of insulin). The action of insulin is opposed by a number of other counterregulatory hormones, including glucagon, cortisol, catecholamines, vasopressin, and growth hormone. The action of these hormones is important in maintaining fuel supplies (fatty acids and ketones as well as glucose) in the fasted state. These actions are unopposed in the absence of insulin, and in the extreme situation of diabetic ketoacidosis they are probably responsible for many of the metabolic abnormalities. However, the importance of counterregulatory hormones in the pathogenesis of diabetes is uncertain; the ratio of

Table 2 Major physiological stimuli to insulin secretion

Metabolites	Glucose
	Amino acids
	Ketone bodies
Gastrointestinal hormones	GIP
	GLP-1
	Glucagon
Neural	Vagal stimulation

GIP, gastric inhibitory polypeptide; GLP-1, glucagon-like polypeptide-1.

Table 3 The metabolic actions of insulin

Metabolism	*Action*	*Site (tissue) of action*
Carbohydrate metabolism	Stimulates glucose transport,	M,F
	glucose phosphorylation,	M,F,L
	glycolysis, and	M,F,L
	glycogen synthesis	M,L
	Inhibits gluconeogenesis and	L,K
	glycogen breakdown	M,L
Fat metabolism	Stimulates fatty acid synthesis and	L,F
	triglyceride synthesis	L,F
	Inhibits lipolysis	L,F
Protein metabolism	Stimulates amino acid transport and	L,M
	protein synthesis	L,M
	Inhibits protein breakdown	L,M

M, muscle; F, fat; L, liver; K, kidney.

hepatic insulin to glucagon may be an important metabolic determinant but it is interesting that diabetes which occurs as a result of tumors secreting an excess of any one of the counterregulatory hormones is usually mild, indicating the relatively greater importance of insulin. (*See* **Hormones**: Adrenal Hormones; Pituitary Hormones.)

Insulin Resistance

The concept of insulin resistance has been challenged on the grounds that it is not a discrete entity which is always associated with pathology. This is undoubtedly true, and insulin sensitivity (the 'reciprocal' of insulin resistance) is clearly a continuum: in a high proportion of the nondiabetic population, the level of insulin sensitivity can be shown to be similar to that in people with type 2 diabetes. In the majority, normal carbohydrate tolerance is maintained by hyperinsulinemia, but in a proportion, insulin secretion diminishes over time, leading to the development of impaired glucose tolerance and, ultimately, diabetes. The reasons for the decline in insulin secretion are complex: 'glucose toxicity' (the process by which sustained hyperglycemia results in impaired insulin secretion) may play a role, but genetic factors are clearly particularly important.

As shown in Table 4, a number of tests have been used to measure insulin sensitivity. The euglycemic hyperinsulinemic clamp remains the gold-standard method but is both labor-intensive and costly. Mathematical modeling of the fasting insulin-to-glucose ratio (homeostasis model assessment: HOMA), the frequently sampled intravenous glucose tolerance test and the short insulin tolerance test are reasonable alternatives. In simple terms, a high insulin-to-glucose ratio apparently indicates insulin resistance, but there may be difficulties in interpreting these findings. When the high blood glucose level is taken into account, the concomitant insulin level may, in fact, be relatively low, although the absolute rise in insulin level with blood glucose may be similar to that seen in nondiabetics, suggesting a change in the 'set' of insulin level.

Table 4 Tests used to measure insulin sensitivity

Fasting insulin-to-glucose ratio (IGR)
Mathematical modeling of IGR (homeostasis model assessment, HOMA)
Frequently sampled intravenous glucose tolerance test (FSIVGTT)
Short insulin tolerance test (ITT)
Euglycemic hyperinsulinemic clamp

Adapted from Hermans MP, Levy JC, Morris RJ and Turner RC (1999) Comparison of insulin sensitivity tests across a range of glucose tolerance from normal to diabetes. *Diabetologia* 42: 678–687.

A further problem which complicates the interpretation of insulin levels in type 2 diabetes lies in the lack of specificity of immunoassays for insulin, nearly all of which cross-react with either proinsulin, or insulin/proinsulin intermediates ('split proinsulin'), or both. Using a range of highly specific two-point immunoassays for insulin, it has been shown that the level of circulating proinsulin and insulin/proinsulin intermediates (all of which are physiologically much less active than insulin itself) is high in type 2 diabetes, presumably resulting from dysfunctional processing and packaging of insulin by the B cell, and that the level of true, bioactive insulin may, in fact, be low rather than high. To an extent, this challenges the concept of 'hyperinsulinemia' in type 2 diabetes.

The insulin resistance of type 2 diabetes affects both carbohydrate and fat metabolism, and is demonstrable in liver, skeletal muscle, and adipose tissue. Hepatic insulin resistance results in unsuppressed glucose output which is the major cause of fasting hyperglycemia, while resistance to insulin-mediated glucose uptake in skeletal muscle is the major determinant of postprandial hyperglycemia.

The adipocyte is more sensitive to the action of insulin than either muscle or liver, but the dose–response curve is shifted to the right in type 2

diabetes, and adipose tissue lipolysis (particularly central adipose tissue) is increased in the fasting state and incompletely suppressed postprandially. As a result, plasma free (non esterified) fatty acid levels are characteristically elevated, but, in type 2 diabetes, since insulin is always present, lipolysis is never completely uninhibited. As a result, ketosis is not a feature of type 2 diabetes. In contrast, in untreated type 1 diabetes resulting from absolute insulin deficiency, not only is lipolysis uninhibited, but reesterification of free fatty acids to form triglycerides is impaired because of the reduction in supplies of α-glycerophosphate from glycolysis. In the liver, glycolysis is not insulin-dependent and triglyceride synthesis is unimpaired and may even be increased because of an increased supply of free fatty acids. (*See* **Fatty Acids: Metabolism.**)

Hyperglycemia and Glycosuria

Blood glucose concentration is normally carefully controlled and maintained between fairly narrow limits as a result of a balance between hepatic output and peripheral uptake of glucose. Even following carbohydrate feeding, around 60% of absorbed glucose is extracted during the first pass through the liver under the influence of insulin, such that the postprandial rise in blood glucose is normally modest.

Glucose is freely filtered by the glomerulus and reabsorbed by an active transport system in the proximal tubule. Under normal circumstances, the urine contains virtually no glucose as the capacity of the renal tubules to reabsorb glucose exceeds the filtered load. However, as the blood glucose rises, there comes a point at which the filtered load of glucose in the renal tubule exceeds the absorptive capacity, and glucose appears in the urine; this is the renal threshold for glucose. The normal renal threshold is between 6 and 7 mmol l^{-1}. It is lower in those individuals with idiopathic renal glycosuria and during pregnancy. The presence of glucose in the urine exerts an osmotic effect in the distal and collecting tubules of the kidney, countering physiological mechanisms for urine concentration, resulting in an increase in urine volume (polyuria). The resultant loss of water from the body (plus the direct effect of hyperglycemia on osmoreceptors in the hypothalamus) results in polydipsia (excess thirst) and, if fluid intake is inadequate, dehydration.

Clinical Features and Symptoms

Type 1 Diabetes

The classical features of type 1 diabetes are weight loss, polydipsia, and polyuria. This is the direct result of the osmotic diuresis which results from hyperglycemia; the weight loss results from dehydration plus loss of both fat and muscle due to the catabolism which occurs in the absence of insulin. In addition to these specific symptoms, there is frequently a history of increasingly severe, nonspecific symptoms such as tiredness, lack of energy, muscle cramps, and general ill health, often over a period of some months, which may only be recognized retrospectively after the development of full-blown diabetes. In children particularly, the onset may be more rapid, with ketoacidosis supervening before the diagnosis is established.

Type 2 Diabetes

In contrast to type 1, the symptoms of type 2 diabetes are very insidious, frequently very mild, and often absent altogether; up to 50% of type 2 diabetes in the general population may be undiagnosed. Most patients are overweight. Indeed, older patients presenting with weight loss are likely to have type 1 diabetes and to require treatment with insulin from the outset. Polyuria and polydipsia may be noticed, and a history of urinary tract infection and vulval candidiasis can frequently be elicited. Ketoacidosis does not occur but occasional patients present with hyperosmolar, nonketotic coma.

Complications such as neuropathy, arterial disease, and retinopathy are frequently present at diagnosis and may be a presenting feature. In many patients, however, the condition remains asymptomatic and is only detected on routine screening.

Coma in Diabetes

Three types of coma occur in diabetes:

- Hypoglycemic coma.
- Hyperglycemic coma, either ketoacidotic coma (diabetic ketoacidosis: DKA), or hyperosmolar, nonketotic coma.

Hypoglycemic Coma

Hypoglycemia occurs predominantly in patients taking insulin but also in patients taking sulfonylureas, particularly long-acting preparations such as chlorpropamide and glibenclamide. The factors which commonly predispose to hypoglycemia, either alone or in combination, are:

- too much insulin or sulfonylurea
- too little food or delayed meal
- too much physical exercise.

In addition to simple overdosing with insulin, variations in absorption from different injection sites, or

even inadvertent intravenous injection, can also precipitate hypoglycemia. Similarly, exercise, as well as lowering blood glucose in its own right, may provoke hypoglycemia by increasing the rate of absorption of insulin from an injection site. A further important cause of hypoglycemia is alcohol which, by inhibiting gluconeogenesis, potentiates the action of insulin and delays recovery from hypoglycemia.

The onset of hypoglycemic symptoms is usually rapid. The pattern of symptoms varies from patient to patient but is usually constant for any individual. The symptoms are those of neuroglycopenia (incoordination, impaired intellectual function, confusion, inability to concentrate, blurring of vision) and of sympathoadrenal activity (sweating, anxiety, tremor, hunger, palpitations). Unless treated by immediate ingestion of carbohydrate (preferably sucrose or glucose), the patient may lose consciousness. Untreated, the resultant coma may last for several hours, but spontaneous recovery owing to counterregulatory hormone (epinephrine (adrenaline), glucagon, cortisol, and vasopressin) and sympathetic stimulation of gluconeogenesis is the rule. Treatment of the comatose patient consists of glucose gel smeared inside the mouth, intramuscular glucagon (1 mg) or intravenous glucose. It is important to avoid giving excess intravenous glucose, particularly in young children in whom the osmotic effects of resultant hyperglycemia can result in cerebral damage. The dose of intravenous glucose in an adult should not normally exceed 50 ml of a 20% solution. Prolonged coma is occasionally seen in association with massive (sometimes intentional) insulin overdosage, with alcohol and with sulfonylurea-induced hypoglycemia, and may require prolonged intravenous glucose infusion.

Particular care should be taken to identify nocturnal hypoglycemia. This frequently occurs during sleep without waking the patient, and therefore may not give rise to typical symptoms. It commonly causes restlessness or sweating at night, vivid dreams or nightmares, and morning headaches or 'hangover.' It is often associated with paradoxically high and/or rising morning blood glucose levels. This phenomenon (the Somogyi effect) is probably caused by a combination of declining insulin levels and a marked counterregulatory (particularly growth hormone) response.

A distressing experience of some patients is that of hypoglycemic unawareness, which can be both alarming and potentially dangerous. This is known to occur in over 20% of patients after 20 years of diabetes and is frequently associated with autonomic neuropathy and impaired counterregulatory response to hypoglycemia. It is also encountered in patients taking nonselective β-blocking drugs. More recently, it has been reported by a number of patients who underwent a change in treatment from animal to human insulin. The reasons for this phenomenon (if, indeed, it is real) are unclear. Human insulin given by subcutaneous injection produces a slightly faster fall in blood glucose than animal insulin but the pattern of both glycemic and counterregulatory response is otherwise virtually identical. Double-blind studies have generally failed to reproduce an association between human insulin and hypoglycemic unawareness even in patients who have reported the condition. Part of the explanation for the loss or, at least, change in symptoms may be attributable to the fact that, when patients were transferred to human insulin, the opportunity was taken to review and, frequently, to encourage stricter glycemic control, which is known to be associated with a reduction in hypoglycemic awareness.

Hypoglycemia is unpleasant and distressing for patients and their families. It also causes swings in blood glucose control because of both the physiological response and overcompensation by the patient. It should be avoided as far as possible, but not at the expense of abandoning attempts to obtain good control.

Diabetic Ketoacidosis

DKA is the most serious metabolic emergency associated with type 1 diabetes. It is the largest single cause of death in young diabetics and, although uncommon in older patients, when it does occur, it has a very high mortality in this age group. The fundamental cause is either absolute insulin deficiency or, less commonly, relative deficiency associated with an acute physiological stress, in association with increased secretion of counterregulatory catabolic hormones. The resultant metabolic effects consist of the following:

1. Hyperglycemia, owing to grossly impaired peripheral glucose uptake and utilization, plus continuing, uninhibited gluconeogenesis, leading to increased polyuria and severe fluid and electrolyte loss.
2. Ketoacidosis, owing to accelerated lipolysis and ketogenesis, with production of excess acetyl coenzyme A, which is partially oxidized to the ketoacids acetoacetic and β-hydroxybutyric acid. Along with acetone derived from acetoacetate, these 'ketone bodies' accumulate in larger amounts than can be metabolized, and are excreted unchanged in urine and on the breath. Both acetoacetic and β-hydroxybutyric acids are weak acids, but in the amounts produced lead to a metabolic acidosis with a concomitant rise in hydrogen ion concentration (fall in pH).

Precipitating causes include infection (in about one-third of cases), inappropriate insulin treatment (too often due to the erroneous advice to reduce insulin dosage when unwell and not eating) and myocardial infarction. DKA is occasionally the presenting feature in new cases of type 1 diabetes. In around 25% of cases no precipitating factors can be identified.

DKA develops relatively slowly (compared with hypoglycemic coma), usually over a period of 24–36 h. Symptoms include an increase in thirst and polyuria, general ill health, nausea, vomiting, drowsiness, and eventually coma. Clinically, the patient is often febrile, with evidence of dehydration (thin, rapid pulse, postural hypotension). Hyperventilation (owing to the metabolic acidosis) is an important feature, and there is frequently a strong smell of acetone on the breath. Treatment should be instituted as a matter of urgency as soon as the diagnosis is confirmed by the finding of hyperglycemia and ketonuria without waiting for the results of other laboratory tests (blood gases and acid–base status; plasma sodium, potassium, and creatinine; full blood count; urine, sputum, and blood cultures). Treatment consists of replacement of fluid and electrolyte deficits which, in established DKA in adults, are in the order of 5 l of fluid and 500 mmol each of sodium, chloride, and potassium. It is important to remember that plasma potassium levels may not reflect this, and may even be paradoxically high due to the effects of acidosis. Although there may be a similar theoretical deficit of bicarbonate, replacement is unnecessary and even counterproductive; small amounts only should be given as required to raise the blood pH above 7.0. Insulin is given in relatively small doses (5–10 units per hour initially) either by intravenous infusion or intramuscular injection. Intravenous potassium should be given either immediately or as soon as plasma potassium levels start to fall, with regular laboratory monitoring of plasma levels.

Nonketotic Hyperosmolar Coma

Nonketotic hyperosmolar coma only occurs in type 2 diabetes, sometimes as a presenting feature, and consists of gross hyperglycemia and dehydration but without ketosis and significant acidosis. The mechanism of this is uncertain. The condition develops more insidiously than DKA, allowing time for a greater degree of dehydration to develop. Treatment consists of fluid replacement and low-dose insulin therapy. The prognosis is poor, with a significant mortality especially in older patients.

See also: **Diabetes Mellitus**: Etiology; **Fatty Acids**: Metabolism; **Glucose**: Maintenance of Blood Glucose Level; **Hormones**: Adrenal Hormones; Pituitary Hormones; **Renal Function and Disorders**: Kidney: Structure and Function

Further Reading

Berger W and Keller U (1992) Treatment of diabetic ketoacidosis and non-ketotic hyperosmolar coma. *Clinical Endocrinology and Metabolism* 6: 1–22.

De Fronzo RA, Ferrannini E and Simonson DC (1989) Fasting hyperglycaemia in non-insulin dependent diabetes. *Metabolism* 38: 387–395.

Hermans MP, Levy JC, Morris RJ and Turner RC (1999) Comparison of insulin sensitivity tests across a range of glucose tolerance from normal to diabetes. *Diabetologia* 42: 678–687.

Pickup JC and Williams G (eds) (1997) *Textbook of Diabetes*, 2nd edn. Oxford: Blackwell.

Reaven GM (1988) Role of insulin resistance in human disease. *Diabetes* 37: 1595–1607.

Young AA (1997) Amylin's physiology and its role in diabetes. *Current Opinion in Endocrinology and Diabetes* 4: 282–290.

Treatment and Management

J Wright, University of Surrey, Guildford, Surrey, UK

Aims of Treatment

The main objectives in the treatment of diabetes can be summarized as follows:

1. alleviation of symptoms
2. maintenance of good general health
3. prevention of long-term complications.

It is relatively easy to achieve the first of these objectives and, in the majority of patients, reasonable or good general health can be maintained in the short or medium term without necessarily achieving good diabetic control. However, there is now convincing evidence that good control of both blood glucose and blood pressure influences the development of microvascular complications (nephropathy and retinopathy) in both type 1 and type 2 diabetes although, as far as macrovascular disease is concerned, the situation is more complicated. Coronary heart disease (CHD) is the major cause of death in long-standing type 1 diabetes, and in type 2 diabetes even of relatively short duration. Evidence suggests that glycemic control is not closely related to the risk of atherosclerosis and other risk factors such as smoking, hypertension, dyslipidemia (raised triglycerides and

low high-density-lipoprotein (HDL)-cholesterol), physical inactivity, male-pattern obesity, and ethnic origin appear to be more important determinants. The presence of microalbuminuria (which is probably a marker of generalized arterial endothelial damage) is a particularly strong marker of risk from ischemic heart disease. (*See* **Atherosclerosis**; **Coronary Heart Disease**: Intervention Studies.)

Thus it is clear that, if a major objective of treatment is to prevent long-term or associated complications, as much attention must be paid to these other risk factors as to control of blood glucose.

Choice of Treatment Regimen

Diet, oral hypoglycemic agents, and insulin remain the cornerstones of treatment of diabetes. Dietary advice is required by all patients, but the selection of other treatment measures is largely based on individual clinical assessment along with measures of glycemic control. The criteria used in the selection of treatment are summarized in Table 1.

With the exception of pregnancy, in which strict criteria are generally agreed, the classification of hyperglycemia into mild, moderate, or severe is personal to both the physician and the individual patient. A common dilemma is the overweight patient with uncontrolled hyperglycemia in whom treatment with insulin, while improving glycemic control, frequently results in increasing weight gain and little in the way of symptomatic improvement. (*See* **Glucose**: Maintenance of Blood Glucose Level; **Pregnancy**: Nutrition in Diabetic Pregnancy.)

Table 1 Treatment options in diabetes

Treatment option	*Criteria*
Diet alone	Moderate hyperglycemia only Obesity
Oral hypoglycemic drugs[a]	
Sulfonylureas	Moderate to severe hyperglycemia Near-normal body weight
Biguanides	Moderate to severe hyperglycemia Obesity Insulin resistance
Insulin[a]	Severe hyperglycemia Ketosis Unintentional weight loss Nonspecific symptoms (general ill health, tiredness) Specific symptoms (painful neuropathy, amyotrophy) Pregnancy Intercurrent illness, surgery, myocardial infarction

[a]Plus diet.

Diabetic Diet

Although the importance of diet in the treatment of diabetes has been recognized for centuries, fashions have changed considerably, even in recent years. Prior to the introduction of insulin treatment, a number of desperate dietary measures were tried, including semistarvation (leading to such gross reduction in body fat that severe ketoacidosis could not occur). With insulin, it became possible to liberalize energy consumption but, with continuing carbohydrate restriction, diets were inevitably high in fat. The evolution in dietary policy since 1921 is illustrated in Table 2. After 1950, in response to the perception that a high-carbohydrate diet does not have a deleterious effect on diabetic control, and that high-fat diets may contribute to the risk of CHD, there was a marked swing away from carbohydrate restriction, and the promotion of diets high in carbohydrate (particularly complex carbohydrate and fiber) and lower in fat than previously. More recently, a number of studies have shown that high-carbohydrate diets may in fact increase insulin resistance and adversely affect the lipoprotein profile, while diets high in monounsaturated fatty acid (MUFA) result in an improved lipoprotein profile (lower triacylglycerols and higher HDL-cholesterol) without loss of glycemic control. (High-polyunsaturated fat diets have a similar effect but are not recommended because of concerns that they result in increased lipid peroxidation.) An opposing view is that high-carbohydrate diets in which most of the carbohydrate has a low glycemic index achieve at least similar metabolic control. However, many people find such diets difficult to manage and unpalatable, and, given the low level of compliance with high-carbohydrate diets, there are both practical and evidential reasons for recommending an increase in the (monounsaturated) fat content of the diet. As yet, however, a number of national diabetic associations have not modified their guidelines, and the one recommendation with respect to fat which remains universally accepted is that saturated

Table 2 Historical perspective of dietary recommendations for diabetes

	Distribution of calories (% dietary energy)		
Year	*Carbohydrate*	*Protein*	*Fat*
Pre-1921	Starvation diets		
1921	20	10	70
1950	40	20	40
1971	45	20	35
1986	Up to 60	12–20	< 30
1994	[a]	10– 20	[a,b]

[a]Based on individual assessment and goals.
[b]Less than 10% calories from saturated fat.

fat should be restricted to less than 10% of dietary energy. (*See* **Glucose**: Glucose Tolerance and the Glycemic (Glycaemic) Index.)

There are no great differences between the general principles of dietary treatment in insulin-dependent diabetes mellitus (IDDM) and those in noninsulin-dependent diabetes mellitus (NIDDM), although there is frequently more emphasis on weight loss in NIDDM, while meal distribution may be more important in IDDM. In patients treated with insulin, the importance of timing of meals and of taking snacks between meals in order to prevent hypoglycemia should be stressed. Overall energy intake should be determined on an individual basis, and the diet designed to achieve and maintain desired body weight. (*See* **Energy**: Energy Expenditure and Energy Balance.)

The place of fish oils (high in long-chain *n*-3 fatty acids) in the diabetic diet is unclear. There has been concern that fish oils may have an adverse effect on glycemic control, but there is a growing consensus that this potential adverse effect is offset by a greater beneficial effect of fish oils on lipid profile and cardiovascular disease risk. Cholesterol intake should be less than 300 mg day^{-1}. (*See* **Fats**: Requirements; **Fish Oils**: Dietary Importance.)

High salt intake should be discouraged, particularly in hypertensive patients. At the same time, a relatively high potassium intake (particularly from vegetable sources) should be encouraged. (*See* **Potassium**: Physiology; **Sodium**: Physiology.)

In patients with early nephropathy, protein intake should not exceed 12% of total dietary energy. (*See* **Protein**: Requirements.)

Alcohol

Guidelines are similar to those for the general adult population, i.e., a maximum of 30 g (3 units) per day for men, and 20 g (2 units) per day for women. Patients taking insulin need to be especially careful since alcohol, by inhibiting gluconeogenesis, can predispose to severe hypoglycemia, especially at night. (*See* **Alcohol**: Alcohol Consumption.)

Special Foods

In general, special 'diabetic' foods are not recommended. Artificial sweeteners are acceptable in moderation.

Individualization

The most important of all the recommendations is that diets should be specifically tailored for individual patients. This is critical in certain groups, such as growing children, pregnant women, and the elderly, in whom restrictive diets may be inappropriate.

Oral Hypoglycemic Agents

Oral hypoglycemic agents rely on residual insulin secretion for their action. Their use is largely restricted to the management of type 2 diabetes. The number of classes of oral agents has increased in recent years, and further developments are anticipated in the near future. Those currently available are summarized in Table 3.

Sulfonylureas

The sulfonylurea group comprises a large number of closely related compounds (**Table 4**). The important clinical differences between individual drugs are related to differences in half-lives, in route of metabolism and excretion, and in receptor binding (second-generation drugs bind more avidly to the specific sulfonylurea receptor on pancreatic B cells and are effective at lower doses).

The major effect of the sulfonylureas is to stimulate insulin secretion, an effect which appears to be maintained in the long term. The glucose-lowering potency of sulfonylureas is directly related to the starting blood glucose level – the higher the fasting blood glucose, the greater the decrease achieved by sulfonylurea treatment.

Table 3 Oral antidiabetic agents

Class of drug	*Example(s)*	*Mode(s) of action*
Sulfonylureas	Chlorpropamide Tolbutamide Glyburide (glibenclamide) Gliclazide	Enhance insulin secretion
Biguanides	Metformin	Increase peripheral insulin sensitivity Inhibit gluconeogenesis
α-Glucosidase inhibitor	Acarbose	Impairs starch digestion and glucose absorption
Thiazolidenediones	Troglitazone Pioglitazone Rosiglitazone	Increase adipocyte differentiation and net triglyceride deposition; reduce plasma nonesterified fatty acids; increase insulin sensitivity
Non S-U insulin secretagogue	Repaglinide	Enhances insulin secretion

Table 4 Examples of sulfonylurea drugs

Drug	*Half-life (h)*	*Duration of action (h)*	*Daily dose (mg)*	*Doses per day*
First-generation				
Chlorpropamide	35	60	100–375	1
Acetohexamide	4–6	12–18	250–1500	2
Tolbutamide	3–8	6–12	500–2500	2–3
Tolazamide	4–7	12–24	100–1000	1–2
Second-generation				
Glyburide (glibenclamide)	5	12–24	2.5–15	1–2
Glipizide	1–5	12–24	2.5–40	1–2
Gliclazide	12	10–15	40–320	1–3
Gliquidone	4	6–8	15–60	1–3

The fact that the sulfonylureas increase insulin secretion means that, if anything, they tend to encourage weight gain. They are therefore most suitable for lean patients and should be used sparingly in the obese. Weight gain may be less marked with those drugs which have a shorter half-life; if given before meals, these have been claimed to enhance the insulin response to food without inducing sustained hyperinsulinemia.

The major adverse effect of the sulfonylureas is hypoglycemia which may be both profound and prolonged. This is seen particularly with those drugs which have a longer half-life or where this is prolonged by hepatic or renal dysfunction. Chlorpropamide, for example, should be used with caution in the elderly and should be avoided in patients with renal impairment. Other adverse effects are uncommon and include nausea, dizziness, blood dyscrasias, cholestatic jaundice, skin rashes, and water retention (particularly with chlorpropamide, which can be used in the treatment of diabetes insipidus). Flushing with alcohol (probably owing to inhibition of acetaldehyde metabolism) is common in patients taking chlorpropamide and can usually be improved by transfer to an alternative sulfonylurea preparation.

Biguanides

The two drugs in this group are phenformin and metformin. Phenformin has now been withdrawn in most countries because of its association with severe and occasionally fatal lactic acidosis.

Metformin reduces both fasting and postprandial blood glucose levels in diabetics but rarely causes hypoglycemia. The precise mechanism of its action is uncertain but the major effect appears to be an increase in glucose uptake and enhanced glycogen secretion in skeletal muscle, resulting in an increase in insulin sensitivity. Metformin also inhibits gluconeogenesis and thus reduces hepatic glucose output. (This results in a small increase in blood lactate concentration but less than that produced by phenformin and rarely sufficient to cause significant lactic acidosis.) Unlike the sulfonylureas, metformin does not cause weight gain; in fact, it often helps with weight loss, and is therefore most suitable in overweight patients. The drug is not metabolized but is excreted unchanged via the kidney and must therefore be used with caution in patients with renal impairment. The major side-effects of metformin are dose-related and include gastrointestinal discomfort and diarrhea. To minimize gastointestinal side-effects, treatment with metformin should begin with a low dose (500 mg daily) building to a maximum of 1000 mg three times per day as tolerated.

Thiazolidenediones

The thiazolidenediones are a new class of drug recently introduced for the treatment of type 2 diabetes. They bind to a novel nuclear receptor, the peroxisome proliferator activated receptor (PPAR-γ), which initiates effects on fat metabolism, including adipocyte differentiation and lipoprotein lipase (LPL) gene expression. The major effect of the thiazolidenediones is probably on the adipocyte, and their ability to increase muscle insulin sensitivity probably results from a reduction in circulating nonesterified fatty acid (NEFA) levels. Their effect on adipocyte differentiation and triglyceride deposition probably explains their tendency to encourage weight gain. It is unclear whether they also have direct effects on hepatic carbohydrate metabolism, such as induction of glucose transporter expression and inhibition of gluconeogenesis.

Troglitazone was the first of the thiazolidenediones to become available for clinical use but was withdrawn after only a short time in the UK because of concerns over liver damage; it remains in use in many other countries, including the USA. A number of other thiazolidenediones have been introduced recently, these include pioglitazone, ciglitazone, and rosiglitazone. Experience indicates that the glucose-lowering effect of the thiazolidenediones is less than that of sulfonylureas, although they may have a greater effect on lipid parameters. Like metformin, they can also be used in combination with insulin in type 2 diabetes.

The starting and maximal daily doses of the thiazolidenediones are 400 and 600 mg for troglitazone, 15 and 45 mg for pioglitazone, and 2 and 8 mg for pioglitazone.

Prandial Glucose Regulators

A new class of insulin secretagogs has recently been introduced, the so-called prandial glucose regulators. Those currently available are repaglinide and

nateglinide. These are chemically dissimilar to the sulfonylureas but may act via the sulfonylurea receptor on the B cell. These drugs are rapidly absorbed and have a short plasma half-life (< 1 h). Given before meals they produce a rapid release of insulin which helps control 'prandial' glucose levels. This may prove valuable since there is increasing concern that postprandial peaks in blood glucose may be an important risk factor for the development of microvascular complications.

The starting doses of repaglinide and nateglinide are respectively 0.5 mg and 60 mg three times daily with meals, rising to maximum total doses of 4 mg and 180 mg three times daily respectively.

Acarbose

Acarbose is an α-glucosidase inhibitor which competitively inhibits the action of brush-border enzymes in the digestion of oligo- and disccharides, thereby preventing the release and absorption of glucose from dietary carbohydrate. It has only a moderate glucose-lowering potency which is most marked in the postprandial phase. The drug is safe but gastrointestinal side-effects (abdominal discomfort, diarrhea, and severe flatulence) are common and frequently limit the acceptability of the drug to patients. The maximal tolerable dose is 200–300 mg day^{-1}.

Combination Therapy

Oral hypoglycemic agents may be used either alone or in combination with other agents. The combination of sulfonylurea and metformin was shown to be effective in the UK Prospective Diabetes Study. Metformin and insulin are frequently used in combination in overweight patients with type 2 diabetes. Thiazolidenediones may also be used in combination with either a sulfonylurea or metformin; they have also been used successfully in combination with insulin but this practice is not yet universal. Sulfonylureas have also been used in combination with insulin (for example with isophane insulin used overnight) but this has not gained widespread acceptance.

Insulin

Prior to the isolation of insulin in 1921, juvenile-onset diabetes was an untreatable and rapidly fatal condition. With well-managed insulin treatment, such patients can now expect to live a full and virtually normal life. Until the 1970s, insulin for therapeutic use was extracted from either beef or pork pancreas. The resulting bovine and porcine insulins differ from human insulin by two and one amino acids respectively. The majority of insulin is now bioengineered using recombinant deoxyribonucleic acid (DNA) technology. The proinsulin thus produced is cleaved to give insulin which is identical in its amino acid sequence to human insulin. Regular, soluble neutral insulin is relatively short-acting and a number of techniques have been used to produce preparations with longer duration of action. The most successful of these are the complexing of insulin with either zinc or a protein (protamine) to produce zinc and isophane insulins respectively. Within the range of zinc insulins, further variations in duration of activity are achieved by varying the pH and, therefore, the particle size to form either amorphous (microcrystalline, medium-acting semilente) or crystalline (long-acting ultralente) preparations. Lente insulin is a mixture of these two in the ratio 30% semilente to 70% ultralente; the resultant mixture is intermediate-acting. Isophane insulin (also known as neutral protamine Hagedorn, or NPH insulin) is of standard formulation and has a similar, intermediate duration of action to lente insulin.

A range of premixed insulins, containing from 10% soluble and 90% isophane to 50% of each, is available. Such premixtures are not possible with zinc insulins owing to the loss of the short-acting, soluble component which results from complexing with any excess zinc present.

Several new insulins are currently being developed and introduced into clinical practice. In these, the insulin molecule is modified in order to alter either the duration of action or receptor specificity. The first of these to be introduced (Lispro and Aspart) are very fast-acting and can be injected immediately before a meal (in contrast to soluble insulin which should be injected 15–20 min before eating). A new, modified insulin (Glargine) has been introduced recently. This provides stable 'background' plasma insulin levels over a 24-hour period, and has been used successfully in combination with very fast-acting soluble insulin analogs. It is anticipated that insulins with altered receptor specificity with, for example, differential activity on hepatocyte or skeletal muscle metabolism, will have particular advantages in different patterns of diabetes.

A number of patients who were treated initially with either pork or beef insulin and subsequently transferred to human insulin reported a marked loss of hypoglycemic awareness and, largely for such patients, some animal insulins continue to be manufactured. In total, a bewildering range of preparations is now available, but can be simplified, as shown in Table 5.

Choice of Insulin Regime

The decision to start insulin treatment is made on clinical grounds (**Table 1**). In younger patients with

Table 5 Major available insulin preparations

Formulation	Time of action (h)[a] Onset	Peak	Duration	Other names
Crystalline[b]	0.5–1	2–4	8	Regular, soluble, neutral Humalog
Zinc				
Amorphous	1–2	4–6	8–12	Semilen
Crystalline	2–4	14–20	24–36	Ultralente
Mixed[b]	2–4	6–10	12–24	Lente
Protamine[b]	2–4	6–10	12–24	Isophane
Premixed[c]	Dependent on ratios			

[a]Approximate times following subcutaneous injection.
[b]Available mixtures contain crystalline and isophane insulin in the ratios 1:9, 2:8, 3:7, 4:6, and 5:5.
[c]Animal (pork or beef) preparation still available.

Table 6 Examples of insulin regimes

Regime	Formulation
Once daily	Lente, ultralente
Twice daily	Soluble plus lente
	Soluble plus isophane (premixed or variable)
Multiple injections (basal bolus)	Soluble or Humalog two or three times daily before meals, plus either ultralente or isophane at bedtime

evidence of ketosis, the decision is usually unambiguous. Patients with type 2 diabetes frequently come to treatment with insulin but the decision to change from oral agents is often difficult. Unintentional weight loss is an important criterion, as are nonspecific symptoms such as tiredness and lack of energy which usually improve on insulin. However, in overweight but otherwise asymptomatic patients with uncontrolled hyperglycemia, large doses of insulin are often required to achieve improved glycemic control, and this is often at the expense of further weight gain and little in the way of symptomatic benefit. Age is certainly not an absolute criterion and, indeed, many newly diagnosed elderly patients need insulin treatment from the outset. Unfortunately, there are no objective biochemical criteria (apart from glucose and ketones) on which to base treatment decisions. In particular, measurement of endogenous insulin secretion (e.g. C-peptide response to glucagon stimulation), while of interest for research purposes, should not be used in the decision to treat with insulin or not.

Insulin is usually given by subcutaneous injection. Examples of injection regimes in current use are shown in Table 6. A once-daily regime may be used in the elderly but for the majority of patients, good control can only be achieved by using at least twice-daily injections. A number of pen injector devices are now available for the administration of soluble, isophane, and soluble/isophane mixtures. An alternative to intermittent insulin injections is the technique of continuous subcutaneous insulin infusion (CSII) whereby insulin is infused continuously (with boosts at meal times) via an indwelling subcutaneous cannula from a pump which is worn or carried permanently by the patient. Impeccable glycemic control can be achieved with pump treatment and although enthusiasm for it waned at one time, recent technical advances have made it a far more acceptable form of insulin delivery, and it is now widely used in a number of countries. Of the other alternative methods of insulin delivery (oral administration, intraperitoneal injection, nasal spray, implantable programmed infusion devices, etc.), some have been abandoned while others are still under evaluation. Other techniques, such as pancreatic transplantation, hold promise, but the prospects for a practical alternative to insulin injections in the near future seem remote.

Other Measures

Treatment of Hypertension

It is increasingly recognized that control of hypertension is essential in diabetes, and it has been shown to reduce the progression of both microvascular and macrovascular complications in both type 1 and type 2 diabetes. Early studies which indicated that treatment with angiotensin-converting enzyme (ACE)-inhibitors reduced the progression of nephropathy (as judged by microalbuminuria) even in normotensive patients suggested that these drugs may have a particular benefit in diabetes. However, in the UK Prospective Diabetes Study, there was no difference in outcome between patients treated with an ACE-inhibitor (captopril) or a β-blocker. Long-term studies comparing the effects of newer ACE-inhibitors, angiotensin receptor antagonists, α-blockers, and calcium channel antagonists are required to decide the optimum drug(s) for use in diabetes.

Exercise

Lack of physical activity is a major factor in the increased incidence of diabetes in the developing world. Exercise has been shown not only to have a beneficial effect on metabolic control in people with established diabetes but also, in retrospective and prospective studies, to reduce the risk of developing type 2 diabetes.

Patient Education

Perhaps more than in any other disease, the person with diabetes needs to participate in treatment and to take responsibility for the day-to-day management of

the condition. This cannot be successfully accomplished without a clear understanding of treatment objectives and the means of achieving them.

A successful educational program needs to be carefully planned and coordinated, with opportunities for individual as well as group sessions, and a commitment to long-term reinforcement. The contribution of the diabetes nurse specialist, who is usually the ideal person to undertake much of the responsibility for patient education, has proved invaluable. The traditional hospital diabetic clinic is by no means the best environment for teaching and in many areas, dedicated diabetic centers have been established for this purpose. A vast array of literature and audiovisual aids is available both from support organizations, such as national diabetic associations, and from commercial firms.

An essential part of any educational program is long-term evaluation. Despite the apparently obvious need for education, the results, when formally assessed, have not been uniformly encouraging. A reduction in hospital admissions can be achieved but this might simply reflect an improvement in care and support services in general. Furthermore, the effects may not be sustained in the long term. As yet, there is no conclusive evidence that improvements in knowledge lead to better metabolic control; poor compliance is usually attributable to factors other than simple ignorance. Successful self-treatment requires, in many people, fundamental changes in attitude and behavior. It is these benefits rather than the acquisition of knowledge that educational programs can provide.

Assessment of Diabetic Control

Glycemic Control

Regular monitoring of diabetes is essential in order to make short-term adjustments to treatment (particularly insulin dosage) and in order to assess control in the longer term. Short-term assessment is achieved by monitoring glucose in urine or blood, and long-term assessment by measurement of glycated (glycosylated) proteins. Monitoring diabetes by means of urine tests was for a long time the only practical method available, but this must now be considered crude and often misleading. The renal threshold for glucose varies widely in diabetes but is frequently well above target levels for blood glucose; the urine may therefore be free of glucose when the blood glucose is higher than desired. At the same time, urine tests are of no value in detecting hypoglycemia. Persistent, heavy glycosuria is probably an indication of poor control (except in those with a very low renal threshold) but more subtle interpretation of tests is usually impossible.

Occasional, single measurements of blood glucose are also of very limited value in monitoring control since there are wide variations in blood glucose throughout the day in both type 1 and type 2 diabetes. Once a common feature of diabetic clinics, 'one-off' blood glucose measurements have been abandoned in many centers. Regular measurement of blood glucose by patients (home monitoring) has become commonplace and is essential in managing changes in insulin treatment, and in making day-to-day adjustments in insulin dosage (especially with multiple injection regimes), particularly during periods of metabolic instability such as intercurrent illness or pregnancy. A number of test 'sticks' are available (most of which employ dry-reagent chemistry based upon glucose oxidase) and these can be read visually or using one of the many meters designed for the purpose.

Measurement of glycated proteins has become the standard method of monitoring medium-term glycemic control. These compounds result from the nonenzymatic linkage of glucose to proteins (in the case of hemoglobin, to the N-terminal valine residue of the β-chain). The initial aldimine linkage undergoes an Amadori rearrangement to form a stable ketoamine product. The rate of glycation depends on the prevailing glucose concentration, and the extent of glycation of any protein is therefore a function of the average glucose concentration during its lifetime. Glycated hemoglobin (HbA_{1C}) is most commonly used because of its availability and because the half-life of the red blood cell (around 8 weeks) is a convenient period of time over which to assess control. Laboratories should now measure the more specific HbA_{1C} fraction. The level of HbA_{1C} in nondiabetics is around 4.5–5.5%. The results of the DCCT study suggest that 7% should be the target level in well-controlled diabetes, although higher levels may be acceptable, particularly in older patients. Low values may be found in patients who are experiencing frequent hypoglycemia. Spuriously low values are found in conditions associated with reduced red-cell survival (e.g., hemolytic anemias), while high values are found in renal failure (owing to the presence of carbamylated hemoglobin).

In situations such as pregnancy, the half-life of hemoglobin is too long for HbA_{1C} to be useful in monitoring glycemic control as closely as is required. Glycated albumin (half-life of around 19 days) would be ideal for this purpose but there are currently no specific methods for this. As an alternative, measurement of glycated serum proteins or fructosamine (mostly albumin and, therefore, of similar half-life) can be used.

Nonglycemic Control

In addition to measures of glycemic control, it is clearly essential to monitor a large number of other biochemical and clinical parameters on a regular basis in an attempt to anticipate and, if possible, prevent the development of complications. Control of blood pressure is particularly important as this has been shown to influence the progression of both nephropathy and retinopathy. Monitoring is best done as part of a regular (usually annual) formal review (Table 7).

Prognosis

Despite the advances which have been made over the last 70 years, life expectancy is still reduced in both type 1 and type 2 diabetes. This is largely attributable to arterial disease (especially coronary artery and peripheral vascular disease), which occurs at an earlier age than in the nondiabetic population, and to those complications which are peculiar to diabetes, i.e., microvascular disease (particularly nephropathy), and both ketoacidosis and hyperosmolar coma. In type 1 diabetes, life expectancy is reduced by about 25%, with a peak in mortality occurring between 15 and 25 years after diagnosis.

There is a common misconception that the situation is less serious in type 2 diabetes, but this is clearly not so. Overall life expectancy is reduced by between 5 and 10 years, largely owing to the very high incidence of coronary and cerebrovascular disease. Nephropathy is relatively uncommon in type 2 diabetes, but visual loss resulting from maculopathy is as common as in type 1 diabetes and frequently occurs early in the course of the disease. It is important to remember that type 2 diabetes may be only a single aspect of a syndrome which comprises, in addition to carbohydrate intolerance, hypertension, dyslipidemia, obesity, and marked insulin resistance. The diabetes itself may be 'mild,' but patients should not be dismissed lightly in view of the high mortality associated with the syndrome. (*See* **Lipoproteins**.)

Table 7 Parameters to be reviewed at annual follow-up

Biochemical	
Blood	HbA_{1C}, creatinine, potassium
	Fasting lipid profile (cholesterol, HDL-C, triglycerides)
Urine	Albumin (microalbuminuria)
Hypoglycemia	Occurrence, severity, adequacy of warning symptoms
Anthropometry	Weight, body mass index, height c
Clinical	
Cardiovascular	Blood pressure, peripheral pulses
Eyes	Visual acuity, fundoscopy, retinal photograph
Neurological	Vibration threshold, tendon reflexes, peripheral sensation
Podiatry	Feet inspection
Treatment	Diet
	Insulin or oral agents
	Injection sites
	Other drugs
Social factors	Smoking, alcohol consumption
	Impotence
	Exercise

HbA_{1C}, glycated hemoglobin; HDL-C, high-density-lipoprotein cholesterol.

See also: **Alcohol**: Alcohol Consumption; **Atherosclerosis**; **Coronary Heart Disease**: Intervention Studies; **Energy**: Energy Expenditure and Energy Balance; **Fats**: Requirements; **Fish Oils**: Dietary Importance; **Glucose**: Maintenance of Blood Glucose Level; Glucose Tolerance and the Glycemic (Glycaemic) Index; **Hypoglycemia (Hypoglycaemia)**; **Lipoproteins**; **Potassium**: Physiology; **Pregnancy**: Nutrition in Diabetic Pregnancy; **Protein**: Requirements; **Sodium**: Physiology

Further Reading

American Diabetes Association (1998) Diabetes mellitus and exercise. *Diabetes Care* 21 (Suppl. 1): S40–S44.

American Diabetes Association (1998) Nutrition recommendations and principles for people with diabetes mellitus. *Diabetes Care* 21 (suppl. 1): S32–S35.

Daly ME, Vale C, Walker M, Alberti KGMM and Mathers JC (1997) Dietary carbohydrates and insulin sensitivity: a review of the evidence and clinical implications. *American Journal of Clinical Nutrition* 66: 1072–1085.

DCCT Research Group (1993) The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes. *New England Journal of Medicine* 329: 977–986.

DeFronzo RA (1999) Pharmacologic therapy for type 2 diabetes. *Annals of Internal Medicine* 131: 1281–1303.

Garg A (1998) High-MUFA diets for patients with DM: a meta-analysis. *American Journal of Clinical Nutrition* 67: 577S–582S.

Pickup JC and Williams G (eds) (1997) *Textbook of Diabetes*, 2nd edn. Oxford: Blackwell.

UKPDS Group (1998) Intensive blood glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes. *Lancet* 352: 837–853.

UKPDS Group (1998) Effect of intensive blood glucose control with metformin in overweight patients with type 2 diabetes. *Lancet* 352: 854–865.

UKPDS Group (1998) Efficacy of atenolol and captopril in reducing risk of cardiovascular disease in hypertensive patients with type 2 diabetes. *British Medical Journal* 317: 713–719.

Problems in Treatment

B Thomas, Dorking, Surrey, UK

Pregnancy

Pregnancy in Women with Pre-existing Diabetes

It is only relatively recently that women with diabetes have been able to look forward with any degree of confidence to the safe arrival of a healthy child. Even 20 years ago, the combination of diabetes and pregnancy was associated with a greatly increased risk of pre- or perinatal death, miscarriage, congenital malformations, and even maternal death. The incidence of these has now fallen dramatically as a result of better diabetes care and monitoring techniques during pregnancy, although the risks remain slightly higher than for the nondiabetic population.

Strict control of glycemia, from the time of conception onwards, is now recognized as vital for a successful outcome to pregnancy. Maternal hyperglycemia is strongly associated with the development of congenital malformations during embryogenesis in the very early stages of pregnancy. Since such damage may occur even before pregnancy is suspected, preconceptional measures to optimize diabetic control are important. The effectiveness of the current diabetes care plan will need to be reviewed and any necessary changes in treatment implemented. Insulin regimens are usually changed to more frequent injections of shorter-acting insulin to provide a greater degree of control, and type 2 patients normally treated with oral hypoglycemic agents may be changed to treatment with insulin. All patients will need to monitor their blood glucose levels more frequently (several times per day) in order to maintain tight glycemic control.

Dietary intake will also need to be reviewed in terms of its suitability for both diabetes and pregnancy. Changes to the insulin regimen usually necessitate dietary adjustment to insure there is a balance between dietary intake and hypoglycemic medication. A stable meal pattern, usually comprised of small frequent meals and snacks, is essential. Starchy carbohydrate foods (such as bread, rice, pasta, potatoes, and breakfast cereals) should form the basis of main meals and people should be encouraged to accompany them with foods rich in micronutrients (such as fruit, vegetables, low-fat dairy products, lean meat, fish, and pulses). Newcomers to insulin treatment will require guidance on how to recognize, avert, and treat hypoglycemia.

Overweight people who are actively trying to slim should be discouraged from doing so while trying to conceive, and only attempt weight reduction before pregnancy is attempted or after it is over.

Other more general nutritional aspects of pregnancy such as the need for folate supplementation will also need to be discussed.

Once pregnancy commences, regular follow-up is necessary to insure that tight glycemic control is maintained and that the nutritional needs of pregnancy are also being met. If nausea and vomiting are a problem during the early weeks, frequent small snacks of easily assimilated sources of carbohydrate such as sandwiches, crispbread, biscuits, or soft drinks may be necessary to prevent hypoglycemia.

As pregnancy progresses, weight gain should be closely monitored so that emerging problems can be detected at an early stage. Rapid or excessive weight gain is undesirable because it will impair glycemic control and increases the risk of fetal problems. Active weight loss is rarely appropriate but weight can be stabilized, or the rate of gain lessened, by encouraging people to alter their food choice in favor of more fruit, vegetables, and fiber-rich carbohydrate, and fewer fat-rich foods and snacks. Such measures decrease the energy density of the diet without impairing its nutrient content.

Since the physiological stress of pregnancy can sometimes exacerbate or precipitate complications of diabetes such as retinopathy and nephropathy, the monitoring process needs to include these aspects.

Following delivery, breast-feeding should be encouraged. Once established, women may need considerably more carbohydrate (e.g., an additional 40–50 g day^{-1}) to meet the high energy costs of lactation. Nocturnal hypoglycemia is a particular risk while the diabetic mother is still breast-feeding the infant during the night, and extra carbohydrate at bedtime in the form of a sandwich or other starchy carbohydrate is usually necessary.

Gestational Diabetes

Gestational diabetes is a symptomless form of impaired glucose tolerance which appears for the first time during pregnancy and, in the vast majority of cases, resolves after delivery. The condition is usually detected via routine urine screening at the first antenatal visit and confirmed by a fasting blood glucose above 6 mmol l^{-1}.

The significance of a diagnosis of gestational diabetes is a matter of current debate. Pregnancy always worsens glucose tolerance to some extent and it is difficult to determine at which point physiological change becomes pathologically significant. While there is an association between gestational diabetes and macrosomia (infants of high birth weight), with its associated risks to mother and child as a result of a

more difficult delivery, it is now less certain that this is a consequence of the hyperglycemia *per se*. Maternal obesity may be the more relevant factor since this tends to worsen the glucose intolerance of pregnancy (thus taking some women into the 'gestational diabetes' range) and is also the main factor associated with macrosomia in the normal obstetric population.

The benefits of intervention have also been questioned for a number of reasons:

- By the time gestational diabetes is identified, the pregnancy will have advanced beyond the stage of embryogenesis when any risk to the fetus in terms of hyperglycemia-induced malformations can be prevented.
- Since the condition is symptomless, there is no need for symptom relief.
- While gestational diabetes may in some cases progress to frank diabetes necessitating insulin treatment, this cannot be predicted or prevented during the pregnancy itself.
- Weight loss in those who are obese is the only intervention which can improve glucose tolerance and lessen the risk of macrosomia, but active slimming is always contraindicated during pregnancy itself because of the risk of fetal undernutrition. At most, dietary measures to prevent excessive weight gain can be instituted (via adjustments in food choice which preserve high nutrient quality while lessening energy density) and these should be encouraged in all pregnant women, not just those with gestational diabetes.
- A formal diagnosis of gestational diabetes has a number of implications in terms of patient anxiety, acquisition of disease status, and perhaps unnecessary obstetric interventions such as Caesareian section. If little can be done about the condition during pregnancy itself, the justification for making such a diagnosis requires careful consideration.

It now seems more likely that the main significance of impaired glucose intolerance in pregnancy is that it is a strong marker for the development of maternal diabetes in later life, particularly if that person remains or becomes obese. The time for intervention is therefore postpregnancy when women should be encouraged to make diet and lifestyle changes which will help them achieve and maintain a healthy body weight. Such people should be considered 'at risk' for development of diabetes and regularly screened for signs of its development.

Diabetes in Children

The diagnosis of diabetes in a child is a traumatic event for the entire family, and parents usually take far longer than the child to adapt to the situation. Insulin injections, diet, and hypoglycemia are the major worries and a high level of support and reassurance from professional advisers will be required in the early stages.

Fortunately, modern management of diabetes means that the dietary constraints are no longer as disruptive to family life as used to be the case. Children with diabetes are no longer prescribed a diet restricted in carbohydrate, virtually devoid of sugar, controlled by means of complicated exchange lists, and necessitating different meals to those of the rest of the family. Instead they are encouraged to consume a diet based on healthy eating principles, regulated via appropriate meal pattern and food choice, and which is also suitable for the entire family.

The aims of dietary management in diabetic children can be summarized as follows:

1. To achieve good glycemic control without an unacceptable frequency of hypoglycemia.
2. To insure the diet is sufficient to meet the demands for growth.
3. To establish eating habits which will provide protection against complications in later life.
4. To achieve these in a way which causes the minimal amount of psychological and social disruption for the child (and family).

Specialist pediatric dietetic care is essential in order to achieve these aims.

Diet and Glycemia

Most children with diabetes will have the type 1 form of the disease and require insulin. It is therefore essential that there is a balance between dietary carbohydrate consumption and insulin action in order to maintain good blood glucose control and prevent extreme glycemic swings. Most insulin-treated patients experience a mismatch in this balance from time to time and suffer the unpleasant consequences of hypoglycaemia (low blood sugar) but this is particularly likely to occur in diabetic children because both dietary habits and activity levels tend to be more erratic. Children's food likes and dislikes may change rapidly, appetite may be highly variable and, particularly in younger age groups, food refusal may be used as a way of seeking attention or asserting independence. Physical activity levels are also less predictable in children; school breaktime may be spent sitting chatting to friends or running round the playground; at home a child may decide to ride a bike or just watch TV. It is therefore much more difficult to maintain a balance between food intake and the dual hypoglycemic effects of insulin and exercise.

In order to minimize the problem, the child needs to consume fairly constant amounts of starchy carbohydrate at frequent intervals. Dietetic expertise is needed to determine the timing and composition of daily meals and snacks which are most compatible with both the family's eating habits and the child's medication, and this pattern should form the framework of food intake. Parents and carers then need to be taught how dietary adjustments may need to be made to this pattern in the light of blood glucose monitoring results, changes in the child's food intake (e.g., due to illness), activity level, and other circumstances.

Hypoglycemia will still almost inevitably occur from time to time and everyone who has responsibility of care for the child will need to be aware of this and know how to deal with it. This may include considerable numbers of people – not just parents, carers, and school or preschool staff, but other adults such as grandparents, parents of the child's friends, leaders of youth organizations, ballet teachers, babysitters, and others.

Diet and growth

In this age group, the dietary needs for growth must not be overlooked. Dietary energy content is particularly important. In the days when diabetes was controlled by highly regulated and restrictive diets, energy intake was often inadequate and growth stunting a fairly common occurrence.

Energy needs of children vary widely and cannot be assessed accurately from tables of average needs. The best guide to individual dietary energy needs is habitual dietary energy intake. If a child is a healthy weight and growing normally, the usual dietary energy intake should remain unaltered. The underweight child will require dietary measures to boost the energy content of the diet, usually by increasing its energy density rather than the total amount of food consumed. The overweight child will require some curtailment of energy intake, usually by decreasing dietary energy density, in order to achieve weight stabilization.

It is also important that the energy and nutrient content of diabetic children's diets are reviewed at regular intervals. It is sometimes forgotten that dietary needs will change as the child grows. An energy intake which is appropriate at the time of diagnosis may be totally inadequate a year later.

Diet and the Prevention of Complications

Tight glycemic control is imperative for the long-term prevention of complications and this is primarily achieved by maintaining a day-to-day balance between carbohydrate intake and insulin activity. In addition to this, the overall composition of the diet also needs to be one which provides protection against the increased risk of developing cardiovascular disease in later life.

As with diabetic adults, starchy carbohydrate should comprise the greatest proportion of dietary energy intake. Suitable food choices which are popular with children include bread, breakfast cereals, pasta, rice, and potatoes. Wholemeal or fiber-rich varieties can be encouraged for their general health benefits, although have no particular advantages in terms of diabetic control. Cereal fiber intake should not be excessively high in very young children or those with small appetites because such foods are bulky and can compromise the intake of other foods needed to meet energy and nutrient needs. Conversely, high-fiber foods can be encouraged in the diets of children with healthy appetites who are becoming overweight.

There is no need for the diabetic child to eat a sugar-free diet. Sugar-rich foods or drinks should not normally be consumed as isolated snacks, although may at times be necessary to prevent or alleviate hypoglycemia. Candy (sweets) and chocolate can, if desired, be included in the diet as planned carbohydrate top-ups prior to exercise. Modest amounts of sugar-containing foods can be consumed as part of main meals. There are no benefits from substituting fructose or sorbitol for sucrose, nor from the use of specialist foods such as 'diabetic' chocolate; all these products are unnecessary and may cause diarrhea in children. However sugar-free 'diet' drinks sweetened with artificial sweeteners such as saccharine, aspartame, and acesulfame-K are useful alternatives to sugar-containing soft drinks.

Excessive fat intake, especially of saturated fat, should be avoided and healthy eating measures to achieve this are suitable for the entire family. Visible meat fat and poultry skin should be discarded, minimum amounts of fats or oils used in cooking or added to foods, and fat-rich foods and snacks such as pastry, pies, crisps, biscuits, and fried foods regarded as occasional treats rather than regular dietary features. However other fat restriction measures should not be overzealous in very young children or energy and nutrient intake may be compromised. For this reason, reduced-fat milks and dairy products are not suitable for children below the age of 2 years.

Protein intake should be sufficient but also not excessive. Small portions of lean meat, fish fingers, cheese, eggs, and pulse foods such as baked beans are good choices for children. Convenience foods such as burgers, sausages, and other meat products are popular but also have a high fat content so should only be provided occasionally and, where possible, reduced-fat varieties should be chosen.

Minimizing Social and Psychological Disruption

Children hate to feel different from their peers, and while the presence of diabetes inevitably imposes some constraints, every effort should be made to keep these to a minimum. Dietary management therefore has to find the best compromise between maintaining good control while keeping life as normal as possible.

The key to achieving this is individualized rather than standardized guidance. Dietary changes necessary to maintain good diabetic control and long-term health should be made by modifying existing family meal patterns and food choices as much as possible, taking account of individual food preferences, appetite, activity level, and lifestyle, rather than by imposing dietary rules which are alien to the child and disruptive to the family.

Advice should be as positive as possible, with diabetes presented as a condition which presents challenges to be overcome rather than one which precludes normal life. Diabetic children have to learn to understand that because they take insulin, they have to be more careful than other children in terms of what they eat, and when they eat. However diabetes will not stop them from eating foods which other children eat or from doing the activities which others do; it is simply that some advance planning may be needed to incorporate these things into daily life. Most children accept and adapt to this with surprising speed.

Diabetes in Teenagers

A number of factors, particularly dietary ones, can affect diabetic control during the teenage years. An increasingly independent lifestyle means that eating habits often change markedly, with trends towards consumption of snacks and fast foods rather than formal meals, a greater proportion of food consumed outside the home away from adult supervision, and alteration in food choice which may either be widened as new foods are encountered, or sometimes narrowed by the constraints of vegetarianism or slimming.

Energy and nutrient requirements will also change during these years. Sudden growth spurts, especially in boys, can result in a dramatic increase in the need for dietary energy and carbohydrate. The body's demands for micronutrients such as calcium and iron are also high. Activity levels may also alter, and in either direction; some teenagers may take up physically demanding new sports such as squash or jogging while others may give up formal sports altogether and spend increasing amounts of time sitting at a computer.

It is also likely that alcohol will be consumed and this is of major significance for people on insulin. Alcohol potentiates the effect of insulin and, if consumed on an empty stomach, can cause severe, potentially fatal, hypoglycemia. It is therefore imperative that alcohol is not consumed in excessive quantities and always in conjunction with sufficient carbohydrate. This is perhaps the most important dietary message to get across to diabetic teenagers. They do need to be aware of the risks and should be offered written guidance on safe drinking limits, how to use the alcohol unit labeling system, and how much additional carbohydrate is needed when alcohol is consumed. They should also know that many of the beers and lagers popular with this age group have an extremely high alcohol content.

Most diabetic teenagers will require modifications to their diet and insulin regimen in order to cope with the changes associated with adolescence. At the same time, many will tend to become less receptive to the advice of health professionals and begin to resent what they perceive as perpetual nagging about diet, control, and the risk of complications. Some will bend the rules as much as they can. With greater maturity, a more responsible attitude will return and the young adult will become more receptive to health guidance. In the meantime, health professionals should aim to keep the channels of communication open and encourage their clients to regard them not as people telling them 'what to do' but 'how to do it.'

Diabetes in Older Adults

The risk of developing type 2 diabetes increases with age, and increasing longevity now means an increasing prevalence of diabetes in people over the age of 65 years. In the past there has been a tendency to regard diabetes in this age group as a 'mild' form of diabetes and its management confined to mere symptom relief. It is now known that this is not the case; the risk of complications such as blindness, gangrene, renal failure, stroke, and coronary heart disease is greatly increased if the condition is not diagnosed promptly and managed effectively.

In this age group, as in any other, tight glycemic control is important. In most patients who develop type 2 diabetes in later life, this can be achieved by a combination of dietary measures and oral hypoglycemic drugs. However, if these prove inadequate for optimum control, some type 2 patients will be managed on insulin.

Although the objectives of diabetes care are no different in older adults, some specific aspects of care need to be borne in mind:

- Treatment with insulin or sulfonylurea hypoglycemic drugs always poses a risk of hypoglycemia

and in older adults this may have serious consequences in terms of falls and fractures. If patients are frail or unsteady on their feet, a balance has to be found between keeping preprandial blood glucose levels low enough to minimize complications but high enough to prevent the hazards from hypoglycemia.

- The physical effects of aging in older adults tend to impair self-management of diabetes. Poor eyesight may make it difficult to draw up the correct dose of insulin or read a blood glucose monitor; short-term memory loss may mean that a dose of an oral hypoglycemic drug may be forgotten or perhaps taken twice. Physical infirmity may make it difficult to shop, cook, or eat. Psychosocial factors such as loneliness, depression, or bereavement may result in loss of interest in food and self-care.
- Older adults are also more likely to have other diseases associated with aging such as stroke, cancer, or dementia as well as diabetes. The effects of these illnesses, and of the drugs or other therapies used to treat them, on appetite, mobility, or memory can have a profound impact on food intake and hence the maintenance of diabetic control.

Recent surveys have highlighted many inadequacies in the care of older adults with diabetes, both those living in the community and people in residential homes. It is now recognized that better and more individualized care and support could do much to reduce the unnecessary morbidity and mortality associated with the disease, and improve the quality of life in the later years.

Diabetes in People from Ethnic Minority Groups

The prevalence of diabetes is particularly high among some ethnic minority groups. People of South Asian origin living in the UK are four times more likely to develop the disease than the indigenous Caucasian population. In African-Caribbean people, the risk of diabetes is twice as high. Both these groups are also more likely to die from long-term complications of diabetes such as cardiovascular disease and renal disease.

Diagnosis and management of the disease are often hampered by the fact that access to, or uptake of, health care services by this sector of the population is often poor and language barriers can further hamper communication. In order to provide effective care, health professionals need to be sensitive to the cultural beliefs and customs of each individual and understand the practical implications of these, particularly in terms of dietary management. Dietetic expertise is vital to insure that the guidance given is both appropriate and culturally acceptable.

See also: **Adolescents**; **Alcohol**: Alcohol Consumption; **Children**: Nutritional Requirements; Nutritional Problems; **Pregnancy**: Nutrition in Diabetic Pregnancy

Further Reading

Ales KL and Santini DL (1989) Should all pregnant women be screened for gestational glucose intolerance? *Lancet* i: 1187–1191.

Benbow SJ, Walsh A and Gill GV (1997) Diabetes in institutionalised elderly people: a forgotten population?. *British Medical Journal* 314: 1868–1869.

British Diabetic Association, Nutrition Sub-Committee (1989) Dietary recommendations for children and adolescents with diabetes. *Diabetic Medicine* 6: 537–547.

British Diabetic Association, Nutrition Sub-Committee (1993) Dietary recommendations for children and adolescents: an implementation paper. *Diabetic Medicine* 10: 874–885.

British Diabetic Association (1997) *Recommendations for the Management of Diabetes in Primary Care*. London: BDA.

British Diabetic Association (1999a) *Guidelines of Practice for Residents with Diabetes in Care Homes*. London: BDA.

British Diabetic Association (1999b) *Recommendations for the Structure of Specialist Diabetes Care*. London: BDA.

Dornhorst A, Bailey PC, Anyaoku V, Elkeles RS, Johnston DB and Geard RW (1990) Abnormalities of glucose tolerance following gestational diabetes. *Quarterly Journal of Medicine* 284: 1219–1228.

Goldenburg RL and Tamura TT (1996)) Pre-pregnancy weight and pregnancy outcome. *Journal of the American Medical Association* 275: 1127–1128.

Haines LC and Swift PGF (1997) Report of the 1994 British Paediatric Association/British Diabetic Association survey of services for children with diabetes: changing patterns of care. *Diabetic Medicine* 145: 693–697.

ISPAD/IDF/WHO (International Society for Paediatric and Adolescent Diabetes; International Diabetes Federation, World Health Organization) (2000) *Consensus Guidelines for the Management of Insulin Dependent (Type 1) Diabetes Mellitus in Childhood and Adolescence*. Zeist, The Netherlands: Medical Forum International.

Jardine Brown C, Dawson A, Dodds R *et al*. (1996) Saint Vincent and improving diabetes care. Report of the pregnancy and neonatal care group. *Diabetic Medicine* 13(suppl. 4): S43–S53.

Jarrett RJ (1997) Should we screen for gestational diabetes? 'The concept of gestational diabetes was popularised before considerations of evidence based medicine came on the scene'. *British Medical Journal* 315: 736–737.

Walkinshaw SA (1999). Dietary regulation for 'gestational diabetes' (Cochrane review). In: *The Cochrane Library*, issue 2. Oxford: Update Software.

Secondary Complications

B Thomas, Dorking, Surrey, UK

Types of Complications

As the life expectancy of people with diabetes has increased, the long-term consequences of the disease have become more apparent. The treatment objectives of diabetes are therefore not confined to the short-term relief of its symptoms but also aim to prevent or minimize its long-term effects.

Complications of diabetes can be divided into two main categories:

1. Arterial disease of large or medium-sized blood vessels (macrovascular disease)
2. Damage to the microvascular circulation of the retina and kidneys and to peripheral nerves (microvascular disease).

There are several differences in the pathogenesis of these two types of complications. Microvascular complications are unique to diabetes, and their development is associated with prolonged exposure to hyperglycemia. In contrast, arterial disease affects both diabetic and nondiabetic populations alike, although the prevalence of the disease is greatly increased in people with diabetes. However, there are some links between diabetic complications: retinopathy and nephropathy often develop together, and patients with diabetic nephropathy have a greatly increased risk of cardiovascular mortality.

The recent findings from two major prospective studies – the Diabetes Control and Complications Trial and the UK Prospective Diabetes Study – have greatly clarified the role of different risk factors in the development of diabetic complications.

Macrovascular Disease

Macrovascular disease results from atheromatous changes in large and medium-sized blood vessels that may become partly or totally blocked. The resulting impairment can affect the cardiovascular system (causing angina and coronary thrombosis), the cerebrovascular system (causing stroke), and the peripheral blood supply to the lower limbs (causing claudication and tissue damage).

The susceptibility to cardiovascular disease is of particular concern. People with diabetes are at least twice as likely to develop heart disease as the nondiabetic population. Furthermore, the disease tends to strike at a younger age and in both sexes; diabetic women appear to lose much of their premenopausal protection against heart disease and are at a much greater risk than their nondiabetic counterparts. People with diabetes also have a two- to threefold increase in the risk of stroke.

The reason why diabetes enhances the atherogenic process has not been fully explained, but it is undoubtedly due to a combination of factors. Hyperglycemia almost certainly plays a role, probably as a result of the metabolic disturbances associated with it, in particular, disordered lipid metabolism. Abnormal blood lipid profiles are a common feature of both types of diabetes, although the pattern of dyslipidemia is usually different. Type 1 patients tend to have elevated levels of total and low-density lipoprotein (LDL) cholesterol and reduced high-density lipoprotein (HDL) cholesterol, a combination known to enhance atherogenic risk. Type 2 patients commonly have elevated very-low-density lipoprotein (VLDL) triglyceride levels, particularly if they are obese, together with increases in the highly atherogenic small dense LDL subclass. All these lipid profiles are exacerbated by poor glycemic control.

Thrombogenic factors, such as the fibrinogen level or degree of platelet stickiness, may also be adversely affected by diabetes, leading to an increased likelihood of thrombosis and blood vessel occlusion.

The presence or absence of other cardiovascular risk factors also plays a key role. Hypertension is particularly strongly associated with cardiovascular risk in diabetic subjects, and long-term tight blood pressure control in hypertensive Type 2 patients has been shown to result in a significant fall in the incidence of coronary heart disease and stroke. Other known contributors such as cigarette smoking, obesity, and physical inactivity are also particularly detrimental to people at such high risk of cardiovascular disease. Prevention of diabetic macrovascular complications therefore requires multiple risk factor reduction from all these influences, in addition to the maintenance of good diabetic control.

Diet plays an important part in cardiovascular disease prevention. The modern diabetic diet, based on healthy eating principles, confers a much greater atherogenic protection than its low-carbohydrate, high-fat predecessor. Like the rest of the population, people with diabetes are advised to obtain at least 50% of their dietary energy from starchy carbohydrate foods such as bread, cereals and potatoes. Carbohydrate food choices with a low glycemic index (such as oat-based breakfast cereals, pasta, and rice) should be encouraged in order to minimize postprandial glycemia. Dietary fat intake should be reduced to 30–35% of dietary energy, with particular emphasis on avoidance of saturated fat. Use of spreading and cooking fats and oils should be kept

to a minimum and should be mainly derived from monounsaturated sources; excessive intake of polyunsaturated fat is no longer encouraged because of the risk of lipid peroxidation. Adequate consumption of fruit and vegetables (at least five servings a day) is a particularly important dietary measure, since there is growing evidence that these foods confer significant antioxidant protection. In addition, the soluble fiber content of fruit and vegetables, together with that found in oats and pulses, may have beneficial effects on both glycemic control and blood lipids. Foods with a high content of cereal fiber (e.g., wholemeal bread) are encouraged for their high satiety value and gastrointestinal benefits but are no longer thought to offer any particular advantage in terms of glycemic control.

High sodium intakes may exacerbate or contribute to hypertension, and, along with the rest of the population, people with diabetes should be encouraged to reduce their sodium intake by about one-third (from the typical average of 9 g per day to 6 g per day). This can be achieved quite easily by not adding salt to foods, avoiding highly salted foods and making sure the diet is not over-reliant on processed manufactured foods (the major contributors to the sodium content of the UK diet) such as canned foods, ready meals, or convenience foods.

For those who are overweight, weight loss is an over-riding priority. Even modest weight loss (5–10 kg) will have marked benefits in terms of glycemic control, blood pressure, and lipid profile. In Type 2 diabetic patients, weight loss is the treatment outcome most closely related to improved life expectancy. In general, a low-energy diet relatively high in starchy carbohydrate, fiber, and low in fat is the most suitable for weight loss because of its high satiety value. Increasing the daily level of physical activity is also important.

Microvascular Disease

Nephropathy

Diabetic nephropathy results from damage to the microvascular circulation of the kidney. It is characterized by persistent proteinuria, rising blood pressure, and a progressive decline in renal function, which eventually may result in renal failure. The risk of premature death from either uraemia or cardiovascular disease is greatly increased.

Nephropathy affects about 30% of people with Type 1 diabetes, typically appearing about 10–20 years after diagnosis. In those with Type 2 diabetes, it tends to be appear after a shorter time scale, and its incidence varies with ethnic origin, ranging from about 25% in Europeans to as much as 50% in people of African-Caribbean and Asian origin.

Prior to the development of overt diabetic nephropathy, damage to the glomerular capillaries results in leakage of albumin into the urine. These losses are small and symptomless but appear to be an important predictor of the subsequent development of clinical nephropathy. Identification of patients with microalbuminuria (urinary albumin excretion in the range of 30–300 mg per day) is now regarded as an important way of ensuring that preventive measures are applied at an early stage to those at greatest risk.

It is now established beyond doubt that diabetic nephropathy is related to both hyperglycemia and hypertension. Prevention therefore requires tight control of each of these parameters.

In the early stages of the disease, strict glycemic control and effective antihypertensive treatment can reduce microalbuminuria and restore the glomerular filtration rate (GFR) to the normal range. There is evidence that the progression to clinical nephropathy can be slowed and associated mortality reduced if these measures are started early enough.

The role of protein restriction in the prevention of nephropathy remains equivocal. Dietary protein intake is known to influence renal function, and a low-protein diet can retard the progression of chronic renal failure. There is some evidence that a diet that includes moderate dietary protein restriction (to about 45 g per day or 0.6–0.7 g per kilogram of ideal body weight) may also reduce the rate of decline in GFR and reduce urinary protein excretion in some patients, particularly if started before renal impairment becomes too advanced. Further research is required to determine precisely which dietary components are important, at what stage protein restriction should be introduced, and at what level. Current recommendations in the USA are that diabetic patients with overt nephropathy should not exceed a protein intake of 0.8–0.9 g per kilogram of body weight per day. Once GFR begins to fall, a further restriction to 0.6 g per kilogram per day may be advisable.

What does seem clear is that all diabetic patients should be discouraged from consuming excessive protein intakes. In the now outmoded era of carbohydrate restriction, people were often encouraged to satisfy their hunger with protein-rich foods such as meat, cheese, and eggs. As well as being unwise in respect of atherogenesis, such a measure may well have been detrimental to renal function.

Excessive protein intake can be avoided by insuring that foods such as meat, poultry, fish, and cheese are not the largest component of a meal, and that a greater proportion of food intake comprises starchy

foods and vegetables (as recommended in *The Balance of Good Health* healthy eating plate model). Such a measure also will help decrease the consumption of saturated fat and hence improve the lipid profile. This may be particularly important in patients with diabetic nephropathy, most of whom have a disordered lipid pattern and eventually die from cardiovascular disease. Other dietary aspects that also minimize this risk, such as weight reduction and salt restriction, are also high priorities in these patients.

Retinopathy

Diabetic retinopathy is the most common cause of blindness in the UK. Damage to retinal capillaries causes microinfarcts, hemorrhages, and retinal deposits, which, if unchecked, can result in visual impairment. The risk of cataracts is also increased.

Retinal changes usually remain mild and asymptomatic for many years, so-called 'background retinopathy.' However, in time, 'proliferative retinopathy' may develop; new blood vessels appear, and progressive damage eventually leads to retinal detachment or vitreous hemorrhage and loss of sight.

The development of background retinopathy is closely associated with the duration and degree of exposure to hyperglycemia. After 20 years of diabetes, most Type 1 diabetic patients and well over half of those with Type 2 will have some degree of retinopathy. However, not all will develop the sight-threatening proliferative changes, the peak incidence of which occurs after about 15 years of diabetes and remains constant at about 3% per annum thereafter. It is unclear which factors trigger this effect. Poor control is a contributory factor (although, paradoxically, a sudden improvement in diabetic control often leads to an abrupt deterioration in retinopathy for a period).

Prospective studies have now established that long-term tight glycemic control is of primary importance in minimizing the development of retinopathy. For this reason, it is also vital that Type 2 diabetes (which can remain undetected for considerable periods of time in older adults) is diagnosed at any early stage; about 20% of these patients already have significant retinopathy at the time of diagnosis as a result of prolonged untreated hyperglycemia.

If retinopathy does occur, much can now be done to halt its progression. Laser photocoagulation can be used to prevent new retinal blood vessels from proliferating and threatening sight, and new surgical techniques may be able to improve vision or prevent further deterioration. However, these damage-limitation measures need to applied at the earliest possible stage in order to protect sight. Since retinopathy is usually asymptomatic until serious damage has occurred, regular screening of the retina is important. All diabetic patients should have annual ophthalmological examinations.

Neuropathy

Diabetes may result in damage to nerves, causing disordered motor, sensory, and autonomic function. Much less is known about the incidence and prevalence of this complication, because its effects are so diverse and difficult to classify.

About 10–20% of insulin-dependent patients may have evidence of peripheral neuropathy. Peripheral nerve damage is characterized by 'pins and needles,' numbness, or pain in the lower limbs, and is particularly likely to contribute to problems in the diabetic foot. Loss of sensitivity to pain can result in unfelt injuries, which, if associated with poor control or peripheral vascular disease, may become infected, leading to ulcers and, possibly, gangrene.

Symptomatic autonomic neuropathy is more rare, but when it does occur, it may cause severe problems, including postural hypotension, incontinence, and impotence. Neuropathy affecting the gastrointestinal tract can cause diarrhea or constipation. Gastric neuropathy may reduce gastric secretion and delay gastric emptying, causing unpleasant sensations of bloating and resulting in anorexia, or even nausea and vomiting. These problems can seriously affect dietary management. As with other microvascular complications, hyperglycemia promotes neuropathy, and its development is related to the duration of diabetes. There is some evidence that improved glycemic control may benefit nerve function.

Prevention of Diabetic Complications

The factors with the greatest influence on the development of each type of diabetic complication, and hence of most relevance to its prevention, have been outlined above. However, since all patients are potentially at risk from all complications, preventative measures have to be global in nature so that protection from all types is maximized.

Treatment Objectives to Prevent Complications

Central to the prevention of every complication is tight glycemic control. Maintaining preprandial blood glucose level between 4 and 7 $mmol\,l^{-1}$ and glycosylated hemoglobin (HbA_{1C}) at 7.0% or below are the most important treatment objectives.

Correction of hypertension and maintenance of blood pressure at 140/80 mmHg or below may be equally critical, particularly in respect of nephropathy and cardiovascular disease.

Correction of disordered blood lipids is important. Sometimes, this will necessitate drug therapy but often can be achieved by more effective management of obesity or by tightening glycemic control.

People also need to minimize the influence of other known cardiovascular risk factors, particularly those from cigarette smoking, obesity, physical inactivity, and an atherogenic diet.

These treatment objectives must be pursued with equal vigor in both Type 1 and Type 2 patients. In the past, the latter form of the disease has often been regarded as a 'mild' form of diabetes, but it is now clear that it is anything but mild in its effects.

Role of Diet in the Prevention of Complications

Effective dietary management is crucial if these objectives are to be achieved. Diet is central to the maintenance of tight glycemic control, by ensuring there is a balance between glucose entering the blood following the ingestion of food and the amount of insulin (either exogenous or endogenous) available to deal with it. This is primarily achieved by the maintenance of a stable day-to-day pattern of eating in terms of meal timing and carbohydrate consumption, to which minor adjustments are made as needed to meet fluctuations in activity level or other circumstances.

Considerable dietetic skill is needed to help people make the necessary adjustments to their normal eating habits so that carbohydrate intake is compatible with hypoglycemic medication, and also in a way that is acceptable to the individual and hence sustainable in the long term. Patients also have to be taught how to use their blood glucose monitoring results to assess whether the balance between diet and glycemia is appropriate, whether further dietary adjustments need to be made, and, if so, what form these should take.

The dietitian also has to insure that the diet is balanced in terms of energy and nutrient content and of optimal composition for the prevention of cardiovascular disease. The healthy eating guidelines detailed in the UK national food guide (*The Balance of Good Health*) are a good basis for achieving this. This model recommends that starchy cereal foods should be the main part of each meal, that at least five portions of fruit/vegetables should be consumed every day, and that the daily diet should also contain two or three servings of low-fat dairy products and up to two modest servings of meat or alternatives. The consumption of fat-rich and sugar-rich foods should be avoided as much as possible. If these guidelines are adopted, the resultant dietary composition will meet all the diabetic dietary objectives, i.e., contain a relatively high proportion of energy from starchy carbohydrate, a low proportion of energy from saturated fat, a high content of antioxidants and soluble fiber, a lower-than-average sodium content, and not excessive protein content.

In those who are overweight, effective weight management is imperative, because obesity has detrimental effects on glycemia, blood pressure, and blood lipids, and hence impacts on all types of complications. Achieving permanent weight reduction is a notoriously difficult task, but the potential health benefits of even modest weight loss justify the efforts involved. This requires far more than just providing dietary guidance; behavioral strategies and considerable support measures will be necessary to effect permanent lifestyle change

Along with other aspects of diabetes care, dietary management has changed considerably in recent years. Treatment is much more patient-centered, with the individual being encouraged to take greater responsibility for their own self-care in partnership with health professionals. The role of the dietitian is no longer to dispense standard advice to all-comers, often in the form of complicated exchange lists or restrictive diet sheets, but to assess the unique dietary and clinical circumstances of each patient, identify the nutritional priorities and help that individual make appropriate dietary changes. Dietetic expertise is an essential component of the skill mix of the diabetes care team.

Quality of Care Necessary to Prevent Complications

Prevention of diabetic complications depends, firstly, on achieving the best possible control of the diabetes and, secondly, on early diagnosis of any problems that do arise. Diabetes care that consists solely of brief routine visits to an overworked GP or busy outpatient clinic may be sufficient to maintain short-term control of diabetes, but will not provide the quality of care necessary to prevent minor problems becoming major and possibly life-threatening.

Effective diabetes management requires multidisciplinary care from a team that incorporates the skills of a number of health professionals such as dietitians, ophthalmologists, and chiropodists as well as those of clinicians and nurses. Regular monitoring and follow-up are vital; all diabetic patients should have an annual clinical and dietetic review to assess the effectiveness of current treatment, identify any emerging problems, and implement any necessary action. Care provision in the primary and secondary sectors needs to be closely integrated so that all patients have access to expert services as a matter of routine.

There is evidence that the care of many people with diabetes has been substandard, particularly that of Type 2 patients, whose susceptibility to complications has often been underestimated. It is also clear that

better care could significantly reduce much of the morbidity and premature mortality associated with the disease. In the UK, high priority is currently being given to implementing national standards of diabetes care in order to reduce the high economic and human cost of diabetic complications.

See also: **Coronary Heart Disease**: Etiology and Risk Factor; **Diabetes Mellitus**: Treatment and Management; **Hyperlipidemia (Hyperlipidaemia)**; **Hypertension**: Physiology; **Obesity**: Etiology and Diagnosis

Further Reading

American Diabetes Association (2001) Nutrition recommendations and principles for people with diabetes mellitus. *Diabetes Care* 24 (supplement 1).

American Diabetes Association (2002) Evidence-based nutrition principles and recommendations for the treatment of diabetes and related complications. *Diabetes Care* 25(1): 148–212.

British Diabetic Association, Nutrition Sub-Committee (1992a) Dietary recommendations for people with diabetes: an update for the 1990s. *Diabetic Medicine* 9: 189–202.

Diabetes Control and Complications Trial Research Group (1995a) The relationship of glycemic exposure (HbA_1) to the risk of development and progression of retinopathy in the Diabetes Control and Complications Trial. *Diabetes* 44: 968–983.

Diabetes Control and Complications Trial Research Group (1995b) Effect of intensive diabetes management on macrovascular events and risk factors in the Diabetes Control and Complications Trial. *American Journal of Cardiology* 75: 894–903.

European Association for the Study of Diabetes, Diabetes and Nutrition Study Group (2000) Recommendations for the nutritional management of patients with diabetes mellitus. *European Journal of Clinical Nutrition* 54: 353–355.

Ha TK and Lean MEJ (1998) Technical review: Recommendations for the nutritional management of patients with diabetes mellitus. *European Journal of Clinical Nutrition* 52: 467–481.

United Kingdom Prospective Diabetes Study Group (1998a) UK Prospective Diabetes Study 33: Intensive blood glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes. *Lancet* 352: 837–853.

United Kingdom Prospective Diabetes Study Group (1998b) UK Prospective Diabetes Study 38: Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes. *British Medical Journal* 317: 703–713.

Waugh NR and Robertson AM (1999) Protein restriction in diabetic renal disease (Cochrane Review). In: *The Cochrane Library*, Issue 4. Oxford: Update Software.

Yudkin JS, Blauth C, Drury P *et al*. (1996) Prevention and management of cardiovascular disease in patients with diabetes mellitus: An evidence base. *Diabetic Medicine* 13(supplement 4): S101–S121.

Dialysis *See* **Membrane Techniques**: Principles of Reverse Osmosis; Applications of Reverse Osmosis; Principles of Ultrafiltration; Applications of Ultrafiltration

DIARRHEAL (DIARRHOEAL) DISEASES

W Karnsakul, Mahidol University, Bangkok, Thailand, **C Lifschitz**, Baylor College of Medicine, Houston, TX, USA

Background

Diarrhea is one of the most common health care problems in both developing and developed countries. Infectious diarrhea has been noted to constitute 10% of US pediatric outpatient visits; its estimated costs total nearly $1 billion annually. World Health Organization (WHO) statistics indicate that diarrhea is the second leading cause of death in the pediatric population, causing about 3 million deaths per year. Although a dramatic decrease in mortality was achieved through the introduction of oral rehydration therapy program in 1979, infectious diarrhea is still a major problem; a causative microorganism is identified in only about half of all such cases. The epidemiology of the microorganisms has not changed significantly over time.

Diarrhea is defined as an excessive loss of fluid and electrolytes in the stool. In newborns and infants, stool volume greater than $15\,g\,kg^{-1}\,day^{-1}$ is considered diarrhea. By approximately 3 years of age, the stool quantity reaches adult output, and from then on, an amount greater than 250 g per day is

considered diarrhea. Diarrhea can be categorized as acute or chronic, according to its duration; chronic diarrhea lasts more than 2 weeks. The etiology of diarrhea can be traced to both noninfectious and infectious agents such as viruses, bacteria, and parasites. Noninfectious etiologies include food allergies, malabsorption, inflammatory bowel disease, drugs, and endocrinopathies.

Pathogenesis of Diarrhea

Most dietary carbohydrates are ingested as di or polysaccharide and are hydrolyzed into monosaccharides prior to absorption. Carbohydrate in the bowel lumen attracts water, and under normal circumstances, water is subsequently absorbed. Whenever there is an excess of luminal carbohydrate, as in the case of malabsorption, net absorption of water is reversed, and diarrhea occurs. Osmotic and secretory diarrhea are differentiated, based on the pathogenesis (see etiology in Tables 1 and 2). Osmotic diarrhea is related to the presence of unabsorbed substances, such as lactose; nonabsorbable sugars, such as lactulose; or laxatives, such as magnesium. These osmotic solutes draw more water through the tight junctions into the gut lumen. Electrolyte absorption continues and is not influenced by intraluminal osmotic solutes. In osmotic diarrhea, the stool fluid has little sodium or potassium; osmolality is high; and the ion gap, calculated by the subtraction of stool osmolality by the doubled value of the sum of $Na + K$, exceeds $100\,mOsm\,kg^{-1}$. This type of diarrhea commonly ceases when fasting. In secretory diarrhea, typically, the amount of stool is voluminous, even when fasting. Stool osmolality is normal, and the ion gap is less than $100\,mOsm\,kg^{-1}$. The most common cause of secretory diarrhea is bacterial enteritis, such as *Escherichia coli* or *Vibrio cholerae*. The increase in net fluid and electrolyte secretion occurs when the pathogens adhere to, or invade, the enterocytes and produce toxins that enhance intracellular second messengers (cyclic AMP, cGMP, and/or Ca/protein kinases). These mediators activate protein kinase, which inhibit NaCl-coupled influx, and increase Cl efflux to the ion channels.

Table 1 Etiology of osmotic diarrhea

Impaired intraluminal digestion
Exocrine pancreatic insufficiency (cystic fibrosis, Shwachman–Diamond syndrome, etc.)
Decrease in bile acid pool (congenital absence of bile acid synthesis, ileal resection, Crohn's disease, etc.)
Congenital lipase deficiency
Congenital lactase deficiency
Adult-type hypolactasia
Sucrase-isomaltase deficiency
Congenital trehalase deficiency
Inflammatory villous atrophy (celiac disease, protein milk allergy, postgastroenteritis)
Malabsorption
Carbohydrate malabsorption (congenital glucose–galactose malabsorption, fructose malabsorption)
Fat malabsorption (abetalipoproteinemia, hypobetalipoproteinemia, etc.)
Laxative abuse
Increased colonic fermentation
Excessive sorbitol
Lactulose abuse
Increased consumption of simple carbohydrate (juice)

Table 2 Etiology of secretory diarrhea

Infection (bacterial toxins)
Enteroaggregative and enterotoxigeneic *E. coli*, cholera
Intestinal obstruction
Partial bowel obstruction, Hirschsprung's disease
Congenital ion transport defects
Congenital chloride transport defect, congenital sodium transport defect
Neuroendocrine tumors
Gastrinoma, VIPoma, somatostatinoma, mastocytosis, carcinoid tumor, medullary carcinoma of the thyroid
Ileal bile acid malabsorption
Nonosmotic laxative abuse
Idiopathic villous atrophy
Congenital microvillous atrophy
Autoimmune enteropathy or vasculitis

Acute Diarrhea

Acute diarrhea is generally associated with an enteric infection. Viral gastroenteritis primarily affects the proximal small intestine. It presents with watery diarrhea and vomiting, leading to dehydration. Enteritis that affects the small intestine, such as that caused by *Vibrio cholerae*, enterotoxigenic and enteroaggregative *Escherichia coli*, *Klebsiella pneumoniae*, *Giardia lamblia*, and *Cryptosporidium*, produces nonbloody, watery diarrhea. Infection with *Salmonella*, *Shigella*, *Yersinia*, *Campylobacter*, enteroinvasive *E. coli*, and amoeba manifests with fever, abdominal cramps with tenesmus, and bloody mucoid stool. *Clostridium difficile*-associated diarrhea usually follows the use of broad-spectrum antibiotics. Usually, the course of the infectious diarrhea is self-limited and lasts no longer than 14 days.

Infectious Diarrhea

Viral Diarrhea

Rotavirus infection is the most common cause of dehydration in early infancy, particularly between 6 months and 2 years of age. The pattern of viral shedding spreads from shortly before, to possibly 2 weeks after, onset of illness. The infection is most common

during the winter in temperate climates. The viruses spread predominately via the fecal–oral route. Symptoms of rotavirus infection begin with fever, nasal congestion (30–50% of children), and emesis, followed by watery diarrhea without blood or mucus, which can last up to 5–7 days. Vomiting and fever typically disappear within 48 h of onset. Severe dehydration and death can occur in children with underlying malnutrition and short bowel syndrome. In tropical countries or endemic areas, rotaviruses are a cause of traveler's diarrhea. Severe and protracted diarrhea is seen in children with T-cell immunodeficiency and severe combined immunodeficiency (SCID), but rotavirus is not a common cause of diarrhea in children with human immunodeficiency infection. Asymptomatic infants, older children, and adults may shed the virus in their stools.

Diagnosis of rotavirus diarrhea is considered in young children with acute onset of vomiting and watery diarrhea. Viral particles can be detected in the stools by electron microscopy. A simple and inexpensive tool with enzyme-linked immunosorbent assay (ELISA) or latex agglutination can detect viral antigen in the stools.

Management of rotavirus diarrhea focuses on rehydration and contact isolation. Mild to moderate dehydration requires oral rehydration solution therapy. Intravenous fluid therapy is used in children with severe dehydration, recurrent emesis, and failed oral rehydration. Antibiotics and antidiarrheal drugs should not be given. Oral administration of human milk or human immune globulin-containing protective rotavirus antibody is indicated in low-birth-weight infants as passive prophylaxis and in children with immunodeficiencies who develop protracted rotavirus diarrhea.

Astroviruses are the second most important group of viral agents that cause diarrhea in young children. Transmission tends to occur from person to person through the fecal–oral route. Occasionally, contamination with water or shellfish has been reported. Most such infections have occurred in children younger than 4 years of age. Diagnosis is made with direct visualization by electron microscopy, but may not be as sensitive as for rotavirus. ELISAs are available in reference and research laboratories.

Enteric adenovirus diarrhea, which occurs year-round, mostly affects children younger than 2 years of age. Caliciviruses (or Norwalk viruses) and calicilike viruses are a common cause of water- and food-borne outbreaks of acute nonbacterial gastroenteritis, and are particularly associated with ingestion of shellfish and salads contaminated by infected kitchen-workers during food preparation. Transmission occurs from person to person via the fecal–oral route. Viral shedding lasts 5–7 days after the onset of illness but can extend up to 2 weeks, and may persist 4 days after the cessation of symptoms. The illness is indistinguishable from other viral gastroenteritis in children. Adults tend to have an abrupt onset of symptoms similar to those of staphylococcal food poisoning. Several other viruses cause gastroenteritis in children, including picornaviruses, coronaviruses, toroviruses, parvoviruses, parvo-like viruses, and unclassified small round viruses. Some of these viruses are occasionally responsible for outbreaks.

Bacterial Diarrhea

Bacterial infection in the gastrointestinal tract has a variety of manifestations, such as acute food poisoning, nonbloody, watery diarrhea; and dysentery. Antibiotic therapy is recommended in infections by certain bacterial pathogens, as shown in Table 3.

Salmonella Salmonellae are Gram-negative bacilli that are classified into groups A–E. *S. typhi* (group D)

Table 3 Recommended antimicrobial treatment for bacterial diarrhea

Diseases caused by pathogenic bacteria	*Antibiotic treatment*
Cholera	1. Tetracycline 50 mg kg^{-1} day^{-1}; maximum 2 g day^{-1} in four divided doses for 3 days (drug of choice for *V. cholera* 01 and 0139 Bengal): use of tetracycline under the age of 8 years may outweigh the side-effect of developing tooth coloring in resistant strain to other antibiotics 2. Doxycycline 6 mg kg^{-1}, maximum 300 mg, as a single dose 3. Trimethoprim-sulfamethoxazole (TMP-SMX) 8 mg kg^{-1} day^{-1} of trimethoprim if the strain is resistant to tetracycline
Shigellosis	Ampicillin, TMP-SMX, for 5 days
Enteropathogenic *E. coli* diarrhea	TMP-SMX For infants with mild diarrhea, nonabsorbable agents, such as neomycin and gentamicin, given three or four times a day for 5 days
Enteroinvasive *E. coli* diarrhea	TMP-SMX
Yersinia enteritis	Aminoglycosides, cefotaxime, tetracycline (< 8 years old), chloramphenicol, TMP-SMX
Salmonellosis or typhoid fever	Ampicillin, amoxycillin, cefotaxime, ceftriaxone, chloramphenicol, TMP-SMX
Pseudomembranous colitis (*C. difficile*)	Metronidazole 30 mg kg^{-1} day^{-1} in four divided doses for 7–10 days

and many other *Salmonella* serotypes cause bacteremia and typhoid fever, typically with gradual onset of fever, malaise, headache, and abdominal tenderness. Nontyphoidal salmonellosis can present with gastroenteritis ranging from a small volume of stools, to profuse bloody diarrhea and to severely watery stools. Most *Salmonella* infections are sporadic, but transmission occurs from person to person; via contaminated water and food of animal origin (e.g., poultry, red meat, eggs, and unpasteurized milk); and by contact with infected reptiles, such as pet turtles and iguanas. *Salmonella* infection frequently occurs in the extreme ages (younger than 5 and older than 70 years of age), and peaks early in the first year of life. Diagnosis is made by isolation of salmonellae from stool, urine, or blood specimen. Antimicrobial therapy is not recommended in uncomplicated cases of gastroenteritis. Infants less than 3 months of age, and children with complicated and invasive disease, malignancy, hemoglobinopathies, HIV infection, immunosuppressive state, or severe colitis, should be treated with antibiotics. A small number of persons infected with *Salmonella* can develop Reiter's syndrome.

Shigella Shigellae are Gram-negative, aerobic, nonmotile bacteria. Shigellae are easily transmitted from person to person and by the fecal–oral route. Ingestion of 10 shigellae can cause dysentery in adults. The bacteria can survive in water for up to 6 months. The infection tends to occur in children younger than 4 years of age. However, newborn infants can have a subclinical infection, and carriers are commonly found in developing countries. Day-care centers are sources of outbreaks in the developed world. Shigellosis often occurs during the rainy season. *Shigella flexneri* is the most common group in developing countries. The rectosigmoid and distal colon are more affected than the proximal part, which leads to bloody mucoid stools; however, some children present with high fever and watery diarrhea in the first 48 h, followed by abdominal cramps, tenesmus, and a small volume of blood and mucus in the stools. Complications with hyponatremia and hypoglycemia are commonly found in shigellosis. Lethargy and febrile seizures can precede diarrhea. Toxic megacolon, intestinal perforation, hemolytic uremic syndrome, pneumonia, and malnutrition can lead to death. Extraintestinal manifestations rarely occur in shigellosis. Sepsis and disseminated intravascular coagulation are infrequent complications but have high mortality rates. Shigellae can be isolated by a common stool culture, followed by biochemical and serologic tests to identify subgroups and serotypes.

Campylobacter *Campylobacter* is a group of spiral-shaped, motile, flagellated, Gram-negative bacilli that can be transmitted through food and water contaminated with material from infected animals or humans. *Campylobacter* is the most common bacterial cause of diarrheal illness in the USA; the majority of these cases stem from cross-contamination or consumption of raw or undercooked poultry. Most such human illness is caused by one species, *C. jejuni*. The bacteria adhere to the intestinal epithelium and produce enterotoxins, leading to secretory diarrhea. They can penetrate the cells and cause cellular damage and cell death, with subsequent bloody stools. The incubation period can last from 1 to 7 days. Symptoms start with nausea, vomiting, and abdominal pain with fever and myalgia, followed by watery diarrhea or bloody stools. Abdominal pain frequently resembles appendicitis in children older than 2 years of age. The diarrhea can last as long as 2 weeks. Extraintestinal manifestations associated with *C. jejuni* include Guillain–Barré syndrome. Direct examination of stools may demonstrate spiral-shaped organisms and fecal leukocytes. *Campylobacter* is microaerophillic, requiring special culture media and conditions. Correction of fluid and electrolyte imbalance is the sole therapy in this infection. The role of antibiotics is still controversial in complicated cases.

Clostridium difficile Pseudomembranous colitis is associated with a Gram-positive, spore-forming anaerobe that grows when the normal colonic flora is suppressed as a consequence of the use of broad-spectrum antibiotics. This organism produces toxins, particularly toxin A. Usually, it manifests with watery, nonbloody diarrhea and abdominal cramps. Occasionally, bloody mucoid stools may develop. Pseudomembranes may be seen on sigmoidoscopy. A fulminant colitis and toxic megacolon may develop, requiring surgical intervention. *C. difficile* can be found in the stools of young infants without causing any symptoms. Diagnosis is made by detection of toxin A in the stool. It has been seen, however, that in certain infections, toxin A is negative, but toxin B is positive; therefore, stools should be tested for this toxin as well.

***Aeromonas* species** *Aeromonas* species are Gram-negative, oxidase-positive bacilli that may be found in about 2% of children with diarrhea. The diarrhea is usually watery and self-limited; however, bloody stools and persistent diarrhea may occur.

Escherichia coli (E. coli) *E. coli* is part of the normal bacterial flora of the human gut. A few strains can cause gastroenteritis.

Enteropathogenic Escherichia coli *(EPEC)* EPEC consists of 12 serogroups. The EPEC strains do not

produce enterotoxins, but certain strains make Shiga-like cytotoxin. Children younger than 2 years of age are usually affected by EPEC and present with acute or chronic diarrhea, especially in developing countries. Diarrhea can be severe and, if untreated, can last up to 2 weeks. Vomiting and fever occur in 60% of children. Diagnosis can be made with commercial *E. coli* typing sera. Treatment consists of fluid replacement and correction of electrolyte imbalance and antibiotic therapy. Trimethoprim-sulfamethoxazole (TMP-SMX) is administered orally. Intravenous aminoglycoside is added if resistance to TMP-SMX is known. Fluoroquinones are not approved for use in children.

Enterotoxigenic Escherichia coli *(ETEC)* ETEC first adheres to the small intestinal mucosa, and then releases both heat-labile and heat-stable enterotoxins. ETEC is a common cause of nonbloody, watery diarrhea in children, and of traveler's diarrhea in developing countries. Adults in endemic areas do not tend to acquire this infection, probably because of the development of protective immunity. Children can present with rapid dehydration. Mortality is higher among malnourished patients. Patients rarely have fever, and stools do not contain blood or fecal leukocytes. The definitive diagnosis is made by isolation of *E. coli*, which produces enterotoxin and contains a gene encoding for enterotoxin production.

Enteroinvasive Escherichia coli *(EIEC)* EIEC is one of the causes of dysentery in developing countries. Occasionally, there are foodborne outbreaks in the developed world. The Shiga-like EIEC produces a clinical manifestation similar to shigellosis. These strains have the same biochemical characteristics and gene encoding for tissue invasion as *Shigella*. These bacteria invade colonic epithelium and cause ulcers that lead to blood and mucus in stools.

Enterohemorrhagic Escherichia coli *(EHEC)* EHEC infection has a unique presentation, with grossly bloody diarrhea, which is associated with *E. coli* serotype O157:H7. Most of these outbreaks have been related to consumption of contaminated beef. After ingestion of the contaminated food, the organisms adhere to the epithelium and produce cytotoxins, called Shiga-like toxins or verotoxins, which inhibit protein synthesis and result in cellular destruction. Children usually present with watery diarrhea, low-grade or no fever, and severe abdominal cramps. The diarrhea may become bloody. *E. coli* O157:H7 infection may cause hemolytic uremic syndrome in children, and thrombocytopenic purpura in adults. Treatment with antibiotics is not recommended, as some data suggest that antibiotic treatment will increase the incidence of hemolytic uremic syndrome.

Enteroaggregative Escherichia coli *(EAEC)* The EAEC organisms produce one or more enterotoxins that cause injury to the intestinal mucosa. The course may be acute, or it may be chronic, especially in infants. The diarrhea tends to be watery; however, bloody diarrhea may occur occasionally.

Yersinia enterocolitica *Yersinia enterocolitica* can cause inflammation in the ileum and colon and mesenteric adenitis, which mimics appendicitis. The infection occurs most frequently in children between 5 and 15 years of age. Clinical presentation is similar to other acute gastroenteritis. Bloody mucoid stools are found in some patients. This is a self-limiting infection, which, in uncomplicated cases, does not require antibiotics.

Vibrio cholerae Infection with *Vibrio cholerae* is manifested as an afebrile, painless, high-output, watery diarrhea that causes rapid dehydration and electrolyte imbalance, leading to hypovolemic shock within 4–12 h if no treatment is implemented. Stools are colorless with some mucus, and resemble rice water. Transmission occurs via consumption of raw or undercooked shellfish and dried fish. The incubation period is usually 1–3 days. Direct contact has not been described. Drugs of choice include tetracycline and doxycycline. TMP-SMX, erythromycin, or furazolidone may be considered in resistant strains.

Parasitic Infection

Except for giardiasis and cryptosporidiosis, intestinal parasitic infection is an uncommon problem in the developed world. The common mode of transmission is the fecal–oral route and ingestion of contaminated water or food. Infection with *Giardia lamblia* can be asymptomatic or present with nonbloody, watery diarrhea and steatorrhea. Normal hosts with *Cryptosporidium* and *Isospora belli* may present with nonbloody, self-limited diarrhea, whereas persistent diarrhea and malnutrition occur in immunocompromised hosts. The clinical spectrum of *Entamoeba histolytica* infection ranges from mild diarrhea to fulminant rectocolitis, and subsequently with frequent bloody mucoid stools. Roundworm infestation may cause watery diarrhea in tropical and subtropical regions Table 4.

Management Dehydration is the major complication in individuals with diarrhea who need fluid and electrolyte assessment. Rehydration and correction of concurrent fluid loss and electrolyte imbalance are

Table 4 Summary of parasitic infection and management

Disease and/or agents	*Endemic areas*	*Modes of infection*	*Diagnostic tests*	*Manifestations*	*Therapy*
Giardia lamblia	World-wide	Ingestion of contaminated water with feces containing cysts	Microscopic examination, *Giardia* antigen	Acute or recurrent abdominal pain, flatulence, anorexia, failure to thrive	Metronidazole, albedazole, furazolidone
Cryptosporidium parvum	World-wide	Fecal–oral route, ingesting water contaminated with oocysts	Microscopic examination, antigen by EIA	Watery diarrhea, abdominal cramps, vomiting, fever	Human serum immunoglobulin in immunocompromised host
Microsporidia	World-wide	Fecal–oral route, ingesting water contaminated with spores	Microscopic examination	Watery diarrhea in immunocompromised host	Albendazole, metronidazole, atovaquone, but recurrence of diarrhea after therapy is stopped
isospora belli	Tropics, subtropics	Fecal–oral route, ingesting water contaminated with oocysts	Microscopic examination	Protracted, foul-smelling, watery diarrhea, fever, vomiting in immunocompromised hosts	TMP-SMX, pyrimethamine
Cyclospora cayetanensis	World-wide	Fecal–oral route, ingesting water contaminated with oocysts	Microscopic examination	Protracted, watery diarrhea, fever, vomiting in immunocompromised hosts	TMP-SMX for 7 days
Entamoeba histolytica	World-wide	Fecal–oral route, ingesting water contaminated with cysts	Microscopic examination, PCR, isoenzyme analysis, antigen detection, serum antibody test	Watery diarrhea, then bloody mucoid stools, fever, liver abscess	Metronidazole, luminal amebicide with iodoquinol, paromomycin, or diloxaanide furoate
Dientamoeba fragilis	World-wide	Fecal–oral route, ingesting water contaminated with protozoa	Microscopic examination	Intermittent diarrhea, abdominal pain, anorexia	Iodoquinol, paromomycin, or tetracyclin
Balantidium coli	World-wide	Fecal–oral route, ingesting water contaminated with cysts	Microscopic examination, scraping lesions during sigmoidoscopy for histology is more sensitive	Recurrent bloody or mucoid diarrhea	Tetracyclin, or iodoquinol and metronidazole as alternatives
Blastocystis hominis	World-wide	Believed to be fecal–oral route, ingesting water contaminated with cysts	Microscopic examination	Associated with symptoms of bloating, abdominal pain, nausea, mild to moderate diarrhea	Metronidazole or iodoquinol
Trichuris trichiura	Tropics, subtropics	Eating embryonated eggs, from soils or food	Microscopic examination	Protracted diarrhea, rectal prolapse	Mebendazole 100 mg twice a day for 3 days
Strongyloides stercoralis	Tropics, subtropics	Penetration of the skin by infective larvae either from contact with infected soil or autoinfection	Microscopic examination, eosinophilia	Rash, cutaneous larva currens, mucoid voluminous stools, malabsorption, steatorrhea, pneumonitis	Thiabendazole 25 mg kg^{-1} per dose twice a day for 2 days
Trichinella spiralis	World-wide	Eating meat containing encysted larvae	Eosinophilia, bentonite flocculating test, muscle biopsy	Diarrhea, muscle tenderness, subungual petechial hemorrhages	*Trichinella spiralis*

Continued

Table 4 Continued

Disease and/or agents	*Endemic areas*	*Modes of infection*	*Diagnostic tests*	*Manifestations*	Therapy
Shistosoma mansoni, S. japonicum	*S. mansoni*: Africa, Brazil, Suriname, Venezuela,	Skin penetration	Microscopic examination	Pruritus, diarrhea, bloody stool abdominal pain, vomiting, peptic ulcer, portal vein hypertension	Single dose of praziquantel 40 mg kg^{-1} for *S. mansoni*, 60 mg kg^{-1} day^{-1} divided into two or three doses for 5 days
S. japonicum	Caribbean: Far east, South-east Asia				
Capillaria philippinensis	Philippines, Thailand	Eating uncooked infected fish	Microscopic examination	Protracted diarrhea, protein-losing enteropathy, ascites	Mebendazole, albendazole

Table 5 Composition of commonly used ORS for rehydration therapy in children

Solutions	*Sodium* ($mmol\,l^{-1}$)	*Potassium* ($mmol\,l^{-1}$)	*Chloride* ($mmol\,l^{-1}$)	*Base* ($mmol\,l^{-1}$)	*Glucose* ($mmol\,l^{-1}$)	*Osmolality*
WHO solution	90	20	80	30	111	310
ESPGHAN solution	75	20	65	30	139	310
Pedialyte (Ross Laboratories)	45	20	35	30	139	250
Resol Wyeth Ayerst	50	20	50	34	111	
Ricelyte	50	20	50	34	Oligosaccharides	
Infalyte Mead Johnson	50	25			Oligosaccharides	200
Gatorade	23.5	< 1	17	3	45 (glucose, sucrose, fructose mix)	330
Colas	1.6	< 1		13.4	50–150 (glucose, fructose mix)	550–750
Apple juice	5	32			120	730
Orange juice	< 1	50		50	120	
Chicken broth	250	8		0	0	500

Gatorade, colas, and juices are demonstrated for comparison, not advised for rehydration.

the mainstay of the therapy. Oral rehydration solution (ORS) therapy was successfully developed for this purpose in children who can tolerate enteral intake (Table 5). The principle is that glucose and amino acids are transported across the apical membrane of the enterocyte by contransporters. Unlike apical sodium–hydrogen exchange, nutrient–sodium cotransport is not affected by increased intracellular cyclic AMP levels, particularly in bacterial toxin-induced diarrhea. Inadequate potassium replacement can cause muscle weakness, paralytic ileus, and cardiac arrhythmia. Current oral rehydration therapy is safe and effective in children with mild to moderate dehydration and electrolyte imbalance. In severe dehydration, when the patient's vomiting is uncontrollable, or whenever there are associated conditions that interfere with oral intake, intravenous fluid rehydration may be required.

Feeding Feeding should be resumed as early as possible. Breastfeeding during diarrhea results in a shorter recovery period and improved nutritional status compared with fasting. Early and rapid reintroduction of a normal diet is recommended to maintain nutritional status, especially in malnourished children. Multiple studies have demonstrated that most eutrophic children do not need to have lactose eliminated from the diet, have the formula diluted, or undergo slow regrading of formula. Liquids that contain a high amount of simple sugars, such as fruit juices, should be avoided, because of their high osmolality. Although lactose malabsorption is rare in well-nourished infants, in some cases, lactose intolerance may delay recovery, especially in children with malnutrition, severe dehydration, or recent episodes of gastroenteritis. In these cases, a lactose-free formula may be beneficial. The addition of age-appropriate food has been shown to be well tolerated by infants and children with diarrhea.

Medications Antimicrobial agents are imperative in those who have infectious diarrhea, as described above. Agents such as loperamide and diphenoxylate, which reduce gut motility, do not alter fluid and electrolyte net balance, and are not recommended. Bismuth subsalicylate may decrease intestinal secretion, but bismuth and salicylate toxicity are worrisome. Clay adsorbents alter the appearance of the

stool without changing the water and electrolyte balance. Probiotics have been shown to shorten the duration of rotaviral diarrhea and prevent antibiotic-associated diarrhea. Zinc supplementation in malnourished children may reduce the duration and severity of diarrhea. Vitamin A and other micronutrients have yet to be proven efficacious for the treatment of diarrheal patients. Hyperimmune bovine antirotavirus colostrums may be given in selected cases to immunocompromised children.

Prevention Individuals who travel to high-risk areas or developing countries should avoid inappropriately prepared drinks, raw fruit and vegetables, and certain foods. Routine antibiotic use is not generally recommended, because of the risks of drug reaction, antibiotic-related colitis, and an increase in the incidence of resistant bacteria. Vaccines are being developed for better immunogenicity and fewer side-effects.

Chronic, Persistent, or Protracted Diarrhea

When the course of diarrhea continues over 14 days, it is considered to have become chronic. Chronic diarrhea can be caused by pathogenic organisms such as viruses, bacteria, or parasites acting on an individual who usually has risk factors such as malnutrition, associated illness, impaired immune status, intestinal mucosal injury, and/or bacterial overgrowth.

Chronic diarrhea may be congenital or acquired.

Congenital Microvillus inclusion disease (congenital microvillus atrophy), although rare, appears to be the most common cause of congenital intractable watery diarrhea. Diagnosis is based on intestinal biopsy demonstrating villous atrophy, crypt hypoplasia, and, on electron microscopy, microvillus inclusion in the enterocytes. Infants with this disease require total parenteral nutrition.

Tufting enteropathy presents in the first weeks of life with intractable watery diarrhea, but its onset is somewhat later than in microvillus inclusion disease. The remarkable finding is that the majority of the epithelial surface contains focal epithelial tufts (teardrop-shaped groups of closely packed enterocytes with apical rounding of the plasma membrane).

Congenital glucose–galactose malabsorption is a rare disorder with a defect of the glucose–sodium cotransporter system, leading to an increase in osmotic load in the intestinal lumen. Affected individuals, however, are capable of absorbing fructose, which becomes the source of dietary carbohydrate.

Congenital chloride diarrhea is a rare autosomal recessive condition. Maternal polyhydramnios is common. At birth, the infant develops rapid dehydration with marked abdominal distention. Stool chloride is $120\,\mathrm{mmol\,l^{-1}}$, and there is metabolic alkalosis. In congenital sodium diarrhea, clinical manifestations are similar to those of congenital chloride diarrhea, except for the fact that stool sodium is as high as $145\,\mathrm{mEq\,l^{-1}}$, and stool pH is alkaline as opposed to what is found in congenital chloride diarrhea.

Primary bile acid malabsorption is another extremely rare transport defect in the distal part of the ileum. Intractable diarrhea begins early after birth, followed by failure to thrive and fat malabsorption.

Hypobetalipoproteinemia is clinically indistinguishable from abetalipoproteinemia (Bassen–Kornzweig syndrome) and manifests itself with steatorrhea, failure to thrive, hypolipidemia, and acanthocytosis, which are present since infancy. The defect is an incapacity to form micelles, an important mechanism for fat transport into the enterocytes. Chylomicron retention disease may manifest similarly with steatorrhea and failure to thrive. Children with intestinal lymphangiectasia can suffer from steartorrhea, protein-losing enteropathy, and lymphopenia.

Disaccharidase enzyme deficiency can be diagnosed with a breath hydrogen test and confirmed by intestinal biopsy demonstrating normal histology and almost absent disaccharidase activity. Congenital sucrase–isomaltase deficiency becomes evident when sucrose or starch is introduced to the diet. Congenital lactase deficiency is a rare condition. Congenital glucoamylase deficiency has been reported in children with chronic diarrhea, which worsened with an oral starch challenge.

Impaired fat digestion tends to cause frequent bowel movements. Stools are foul-smelling, bulky, greasy, and float on the surface of the toilet water. This is one of the most common causes of failure to thrive in children. A sweat chloride test confirms the diagnosis of cystic fibrosis in some of these children. However, exocrine pancreatic insufficiency may occur in other diseases as well.

Acquired Chronic nonspecific diarrhea, or toddler's diarrhea, is seen in children between 6 and 36 months of age. Except for the history of watery stools, the children are healthy, and experience normal growth and development without abdominal pain. Diarrhea has been associated with initiation of broad-spectrum antibiotics, which change the normal intestinal flora, excessive fluid intake, and fruit juice (particularly apple juice) consumption.

Secondary lactase deficiency can be seen after acute gastroenteritis, particularly in malnourished individuals and those with recurrent episodes of diarrhea in the recent past. Acquired or late-onset lactase deficiency may begin in childhood or preadolescence, mainly in the nonCaucasian population, but it occurs

in many Caucasians as well. Reduction of lactase production is determined by genetic background and may lead to lactose intolerance, causing abdominal pain and/or diarrhea.

An abnormal intestinal villous surface prevents the absorption of digested nutrients, and leads to osmotic diarrhea and steatorrhea. Hypersensitivity to cows' milk or soy protein may cause partial villous atrophy in young infants. Frequent bloody mucoid stool, dehydration, and failure to thrive are not uncommon features. Total villous atrophy is found in celiac disease in infants who consume a diet containing gluten. Intestinal biopsy is the gold standard to make the diagnosis. Partial villous atrophy may be associated with intestinal infection such as *Giardia lamblia*, particularly in individuals with secretory IgA deficiency. Chronic diarrhea occurs often in individuals with congenital hypogammaglobulinemia, combined immunodeficiencies, and acquired immunodeficiency, at times associated with opportunistic infections. Crohn's disease and ulcerative colitis may present with chronic watery or bloody diarrhea. Crohn's disease presents systemic symptoms more often than ulcerative colitis. Enteroenteric and enterocolonic fistula may complicate cases of Crohn's disease with malabsorption and bacterial overgrowth. Infectious diarrhea can mimic exacerbation of inflammatory bowel disease.

Autoimmune enteropathy is a poorly understood syndrome with malabsorption and chronic diarrhea, which usually develops after 6 months of age and does not respond to a gluten-free diet. The presence of specific antienterocyte antibodies in serum can confirm this disorder.

Children with altered intestinal motility, such as partial obstruction or pseudoobstruction, may present with chronic diarrhea. Bacterial overgrowth also leads to diarrhea.

Certain tumors such as gastrinoma, VIPoma, ganglioneuroma, and ganglioneuroblastoma produce neurotransmitters that can cause secretory diarrhea. Tropical sprue can be seen among children who live in certain Caribbean countries, northern South America, Africa, and parts of Asia. Patients develop a diffuse lesion in the small intestine, which leads to malabsorption. Symptoms usually begin with fever, malaise, and watery diarrhea, which then subside. Subsequently, malabsorption and intermittent diarrhea follow and cause malnutrition. Small bowel biopsy demonstrates villus atrophy, an increase in crypt length, and chronic inflammatory cells in the lamina propria. Symptoms respond well to nonabsorbable sulfonamides or tetracycline for 3–4 weeks, folic acid and vitamin B_{12} therapy. When the investigation is thorough, but the etiology is still unknown or the resultant data are inappropriate, one should look for a condition of self-induced diarrhea. The use of stool laxatives can mimic chronic osmotic diarrhea. Determination of stool electrolytes, magnesium, and sulfate can be helpful in making the diagnosis. Creatinine may be detected in the stool in cases when urine is added to the stool by the patient or a caretaker to simulate diarrhea. A low stool osmolality suggests addition of water to the stool.

Management Treatment should be approached according to the pathophysiology of the diarrhea. The main goal is to prevent dehydration, malnutrition, its complications, and associated illnesses. Infection should be investigated and aggressively treated with appropriate antimicrobial agents. Empirical therapy for chronic diarrhea is considered when no specific treatment is available, based on the etiology of chronic diarrhea. Empirical trials of antimicrobial agents may be judiciously used in patients who have a high possibility of a condition leading to bacterial, viral, or protozoal infection. Bile acid-binding agents such as cholestyramine may decrease diarrheal output in children with primary bile acid malabsorption or high bile acid output-induced diarrhea. Reducing the intestinal transit time is helpful to allow the enterocytes to reabsorb fluid and electrolyte. Therefore, opiates are considered as antidiarrheal agents in those patients with noninfectious chronic diarrhea. Octreotide, a somatostatin analog used in tumor-induced secretory diarrhea, has been shown to decrease secretions. Parenteral nutrition should be the last resort for those who are unable to meet nutritional requirements with the enteral approach.

See also: ***Aeromonas***; ***Campylobacter***: Properties and Occurrence; Campylobacteriosis; ***Clostridium***: Botulism; ***Escherichia coli***: Food Poisoning; ***Salmonella***: Salmonellosis; ***Shigella***; **Vibrios**: *Vibrio cholerae*; **Viruses**

Further Reading

American Academy of Pediatrics (2000) *Red Book 2000. Summaries of Infectious Diseases*, 25th edn, pp. 161–642.

Bass DM (1999) Rotavirus and other agents of viral gastroenteritis. In: Behrman RE, Kliegman RM, Arvin AR and Nelson WE (eds) *Nelson Textbook of Pediatrics*, 15th edn., pp. 914–916. Palo Alto, CA: Stanford University.

Klish WJ (1999) Chronic nonspecific diarrhea of childhood. In: McMillan JA, DeAngelis CF, Feigin RD and Warshaw JB (eds) *Oski's Pediatrics. Principles and Practice*, 3rd edn., pp. 1639–1640. Philadelphia, PA: Lippincott Williams & Wilkins.

Lentze MJ (2000) Persistent diarrhea. In: Walker AW, Durie PR, Hamilton JR, Walker-Smith JA and Watkins JB (eds) *Pediatric Gastrointestinal Disease, Pathophysiology,*

Diagnosis, Management, 3rd edn., pp. 39–45. Hamilton, Canada: B. C. Decker.

Lew EA, Poles MA and Dieterich DT (1997) Diarrheal diseases associated with HIV infection. *Gastroenterology Clinics of North America* 26: 259–283.

Mehta DI, Lebethal E and Blecker U (1996) Chronic diarrhea: cause, presentation, and management. Indian *Journal of Pediatrics* 63: 459–472.

Nappert G, Barrios JM, Zello GA and Naylor JM (2000) Oral rehydration solution therapy in the management of children with rotavirus diarrhea. *Nutrition Reviews* 58: 80–87.

Panosian CB (1988) Parasitic diarrhea. *Infectious Disease Clinics of North America* 3: 685–703.

Pickering LK and Snyder JD (1999) Gastroenteritis. In: Behrman RE, Kliegman RM, Arvin AM and Nelson WE (eds) *Nelson Textbook of Pediatrics*, 15th edn., pp. 721–724. Philadelphia, PA: W. B. Saunders.

Rudolph JA and Cohen MB (1999) New causes and treatments for infectious diarrhea in children. *Current Gastroenterology Reports* 1: 238–244.

Schiller LR (1999) Secretory diarrhea. *Current Gastroenterology Reports* 1: 389–397.

Schmitz J (2000) Maldigestion and malabsorption. In: Walker WA, Durie PR, Hamilton JR, Walker-Smith JA and Watkins JB (eds) *Pediatric Gastrointestinal Disease, Pathophysiology, Diagnosis, Management*, 3rd edn., pp. 46–70. Hamilton, Canada: B. C. Decker.

Taylor DN, Connor BA and Shlim DR (1999) Chronic diarrhea in the returned traveler. *Medical Clinics of North America* 83: 1033–1075.

DIETARY FIBER

Contents

Properties and Sources

R Mongeau, Ottawa, Aylmer (Quèbec), Canada
S P J Brooks, Ottawa, Canada

Background

Although dietary fiber has been known for more than 2000 years under various terms (e.g., bran, roughage), the term 'dietary fiber' first appeared in 1953 and referred to hemicelluloses, cellulose, and lignin. The term 'fiber' is somewhat misleading since only a fraction (cellulose) of dietary fiber is fibrillar in nature. To correct this misnomer, other terms (e.g., plantix) have been proposed, but despite these efforts, the term 'dietary fiber' has survived.

This section deals with several different aspects of dietary fiber, including its determination and many of its physiological effects. Most of this information is relatively new and is the result of progress in research brought about in the last 30 years. The first part of this article will focus on the nature and composition of dietary fiber, its properties, and examples of sources of dietary fiber. This is followed by a definition of dietary fiber. Although this should be a relatively straightforward description, there are many different viewpoints concerning the nature and physiological effects that dietary fiber should have. These will be discussed in relationship to the physiological effects of dietary fiber.

Chemical Structure

The composition and structure of dietary fiber differ from plant to plant. It is also a function of the portion of the plant that is edible and the stage of maturation and is largely composed of the cell wall (structural) components that give the plant physical stability. As such, it is made of highly interlinked sugar-based and phenolic-based polymers (hemicelluloses, pectic substances, phenolics, glycoproteins, and proteoglycans) in a matrix of amorphous structure with some enmeshed cellulose microfibrils. The cell-wall components are intimately linked together

through various linkages including protein–sugar bonds.

Polysaccharide Fiber Components

Dietary fiber comprises carbohydrate and noncarbohydrate polymers; most of these are structural components. The carbohydrate polymers are often described as nonstarch polysaccharides (NSP) to differentiate them from the (relatively) easily digested starch components of food. Figure 1 shows the proportion of the monomer constituents of fiber polysaccharides from three food categories (11 cereals, 13 fruits and 11 vegetables) and from canned mushrooms, peanuts, and baked white beans. As can be seen, arabinose is a major NSP component of cereals, peanuts, and beans. Xylose is present in large amounts only in cereals. Pectic substances (uronic acids) represent a substantial part of fruit, vegetable, peanut, and bean NSP. Mushroom NSP is formed mostly of glucose residues. These results were based on the analysis of composites made of up to 20 individual foods purchased over a 30-month period using the Englyst gas–liquid chromatography method. A more recent version of this method would have given a higher proportion of xylose relative to glucose.

The different monomeric residues of Figure 1 are the basic units that form the major NSP macromolecular constituents shown in the upper part of Table 1. Cellulose and β-glucans are generally unbranched polymers containing mostly glucose residues. The β-glucans have mixed β-1,3- and β-1,4-inter residue linkages and are a constituent of dietary fiber from barley and oats. The glucose residues of cellulose are joined by β-1,4-linkages. In addition to their resistance to the enzymes of the human digestive tract, these linkages allow greater inter- and intrapolymer hydrogen bonding than do the α-1,4-linkages of starch. This allows the formation of tightly condensed crystalline regions within the cellulose microfibrils. Cellulose includes structurally amorphous regions where the presence of other sugars can be found. The final cellulose structure is a function of the glucose and nonglucose content: the degree of cellulose crystallization is inversely proportional to the nonglucose residue content. Cellulose is a minor fraction of total cereal dietary fiber but is a major fraction of dietary fiber in other foods. The proportion of cellulose in the dietary fiber of a food is not reflected by the glucose content of the NSP fraction, since glucose is also a major component of other fractions such as β-glucans and hemicelluloses.

Xylose is a part of the primary chain of hemicelluloses and gums, and the secondary chain of gums and pectin. Arabinose is in the primary chain of gums and the secondary chain of hemicelluloses and pectin. Hemicelluloses constitute a major fraction of dietary fiber; this fraction is largely made of arabinose and xylose in cereals and of xylose and glucose in fruits and vegetables (Table 1). Pectic substances are prevalent in citrus fruits. They are also present in vegetables, legumes, and cereals in small amounts (Table 1). Other polysaccharide constituents are mainly represented by galactose, or mannose in bananas. Nonstructural NSPs represent a small portion of dietary fiber. The monomer NSP composition is of limited usefulness in predicting the properties and functions of dietary fiber. This is because the structural architecture of dietary fiber plays a large role in many of its physiological properties, and this structure cannot be predicted by its composition. Indeed, two dietary fibers with a similar component profile (e.g., wheat

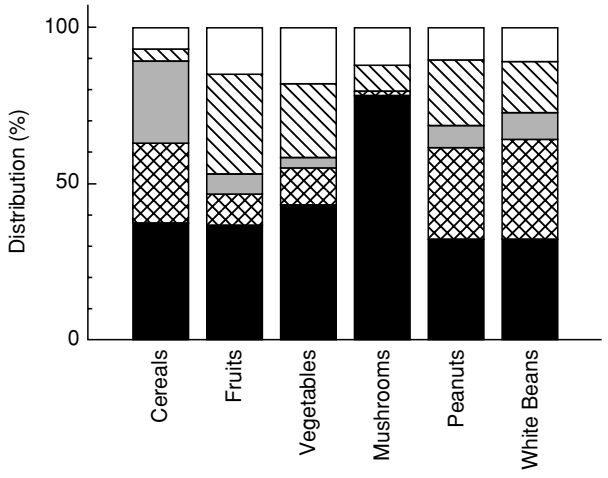

Figure 1 Distribution of the main polysaccharide constituents of dietary fiber from some foods and food categories. The values are means of $n = 11$ cereals, $n = 13$ fruits, and $n = 11$ vegetables. Other values are for $n = 1$. ■: glucose; ⊠: arabinose; ▩: xylose; ▨: uronic acid; □: other sugars.

Table 1 Main macromolecular constituents of dietary fiber

Constituent	*Fruits and vegetables*	*Cereals*	*Legumes*
Polysaccharides			
Hemicelluloses			
Xyloglucans	X		X
Glucuronoxylans	X		
Arabinoxylans		X	
Glucuronoarabinoxylans		X	
Galactomannans			X
Cellulose	X	X	X
β-D-Glucans		X	
Pectic substances (pectin)	X	X	X
Others			
Lignin		X	
Phenolic esters	X	X	
Protein		X	
Glycoproteins	X		X

bran and corn bran) can have clearly different structures, even under optic microscopy, and can show different behaviors in the gastrointestinal tract.

Nonsaccharide Fiber Components

Minor components of dietary fiber that are not polysaccharides and, hence, not NSPs also play an important structural role. Table 1 shows the distribution of the nonsaccharidic polymers that are intimately associated with dietary fiber structure. These polymers include glycoproteins (fruits, vegetables, and legumes), protein (cereals), phenolic esters (cereals), and lignin (lignified tissues of fruits, vegetables, and cereals). Lignin is of special interest because of its role in slowing down the fermentation of dietary fiber. It is a complex group of phenyl-propane polymers formed by the condensation of aromatic alcohols and is especially important in conferring structural stability. As such, highly lignified tissues are found in the stems of plants such as trees and bushes as well as in the stalks of cereals. Lignification occurs at the expense of water and pectin as cells differentiate and mature. This increases cell-wall rigidity in critical areas. Since edible plant tissues are consumed when relatively immature, their cells are largely undifferentiated and, largely, unlignified. Wheat bran and the small seeds covering strawberries are examples of lignified tissues. In most fruits and vegetables, lignin represents only a small part of dietary fiber, although it can be as high as 4% of dry matter in mature pears.

Cutin, waxes, or suberin may also be present in some tissues. These compounds are mixtures of lipids, proteins, and carbohydrates that form the waterproof covering and cuticle on the outer cell wall of plants. They are especially resistant to digestion and fermentation. As such, they impair the digestibility of other cell-wall components and are usually found in the feces.

Structure

Models of primary cell walls of plants were often obtained from nonfood material such as wood and tobacco. Because the cell-wall material forms a large part of these plants, the models provide a wealth of information on the interactions between cell wall components such as carbohydrates and lignin as well as protein and carbohydrates. More recent reports have provided some structural information on the cell walls of plant foods. These reports have shown that, in carrots, protein–protein or protein–polysaccharide linkages may contribute to the formation of a rigid, inextensible cell wall. Other data have shown that protein may form up to 10% of cell walls in immature plants and appears to be important for structural cohesion. The major part of the flesh of fruits and vegetables contains undifferentiated types of cell wall that have significant amounts of highly branched hemicellulosic types of polysaccharides. The cellulose content is usually low, and cellulose is laid down in a more oriented arrangement in the matrix. This may reflect stresses on the plant. The cell walls are often rich in pectic substances and contain approximately 10% protein, as described above. These cell walls are thin and elastic with cell form maintained by osmotic pressure.

In contrast, the outer layers of many seeds and nuts contain thick lignified cell walls that are difficult to break. These serve as mechanical protection to the seed within and as a vapor barrier to prevent desiccation.

Physical Properties

Soluble and Insoluble Dietary Fiber

It is possible to separate total dietary fiber into soluble and insoluble components based on solubility in water, but quantifying the amount of soluble fiber remains problematic. This is because the conditions utilized to measure the proportion of soluble fiber differ among laboratories (e.g., time, temperature, pH, and type of buffer; (see Definition section). It should be noted that the distinction between soluble and insoluble fibers is somewhat artificial, since it is difficult to predict the actual fiber solubility in the gastrointestinal tract.

For many years, it was thought that soluble fiber was not commonly found in foods. Reports on pectin, guar, legume fiber, and oat bran appeared in the early 1960s, and currently, it is estimated that about a third of the daily total dietary fiber intake is soluble fiber. Pectic polyuronides are the major soluble dietary fiber components of vegetables and fruits, but the exact composition varies from plant to plant. When measuring dietary fiber, one must also take into account the method of food preparation because the soluble fiber content of a fiber source is influenced by the method of food preparation. For example, some polyuronides are susceptible to depolymerization at the high temperatures used to cook the foods. Similarly, some insoluble fiber is susceptible to depolymerization at higher temperatures and may be found in the soluble fiber fraction.

In addition to the distinction between soluble and insoluble fiber, viscosity may also be an important fiber characteristic. This is because viscous fibers modestly lower blood cholesterol and triacylglyceride concentrations as well as attenuate the postprandial glucose response. It is thought that the increased viscosity of the intestinal lumen reduces the rate of diffusion of glucose, cholesterol, and triacylglycerides

to their respective receptors along the intestinal wall reducing absorption. The results of a recent meta-analysis using several different clinical studies showed a statistically significant lowering of plasma triacylglycerides and cholesterol in subjects fed for at least 14 days with a diet supplemented with oats, psyllium, pectin or guar gum, but the effect was modest: a decrease of approximately 1 mg per deciliter per gram of soluble fiber was observed. If we consider that an individual with moderately higher blood cholesterol has a total serum cholesterol value of approximately 300 mg dl^{-1}, the change brought about from consumption of viscous, soluble fibers is on the order of 0.3%. Considering that oat bran contains about 20% fiber by weight and only half of this is soluble, one can calculate that 10 g of oat bran per day is needed for each 0.3% change in total cholesterol.

Insoluble dietary fiber is, by definition, the fraction of total dietary fiber (TDF) that is not soluble in hot buffer solution. However, the measured proportion of insoluble fiber can vary depending on the dietary fiber methodology. Insoluble dietary fiber consists primarily of hemicellulose, cellulose, and lignin (*See* **Dietary Fiber**: Determination). It is often incorrectly believed that insoluble fiber is not fermented to any great extent by the bacteria present in the large bowel. However, actual measurements of fiber fermentability have shown that a significant proportion of insoluble dietary fiber is fermented. This is true even for wheat bran, which has a relatively lower insoluble fiber fermentability of between 30 and 40% (approximately 90% of the wheat bran TDF is insoluble dietary fiber). The fermentability of oat bran is even greater – up to 80% of the insoluble fiber (approximately half of the TDF) is fermented in the human large intestine. The degree of fermentability is a property unique to each dietary fiber and depends largely on the nature and the structural arrangement of the fiber components, and also on other physical characteristics such as particle size. For example, the structure of coarse wheat bran favors a high water-holding capacity (see below) and a slow rate of fermentation. This permits the microtrapping of slowly released fermentation gasses, increasing fecal bulk and stimulating defecation.

Water-holding Capacity

Dietary fiber holds water by adsorption and absorption. Some water is also retained outside the fiber matrix (free water). The particle size, chemical composition, and structure of dietary fiber influence the water-holding capacity. The *in vitro* water-holding capacity cannot be used to predict the impact of highly fermentable fiber on colonic function.

Table 2 Effect of grinding on the properties of wheat bran insoluble fiber[a]

Mesh sieve	*Aperture*[b] *(μm)*	*MPS (μm)*[c]	*WHC*[d]	*Glycocholate binding*[e]
As is		800	9.5 ± 0.1	26.5 ± 1.5
20	840	420	8.1 ± 0.1	24.9 ± 1.2
40	420	280	6.6 ± 0.2	23.8 ± 0.8
60	250	180	5.8 ± 0.1	22.1 ± 1.0
80	175	160	5.6 ± 0.1	21.5 ± 0.3

[a]Residue after neutral detergent and porcine pancreatic α-amylase treatments.
[b]Aperture size of sieve used with a Wiley mill, intermediate model.
[c]MPS, geometric mean particle size.
[d]WHC, water-holding capacity, grams of water per gram of insoluble fiber, using the centrifugation method: mean ± SEM of three measurements.
[e]Grams of glycocholate bound per 0.2 g of insoluble fiber; mean ± SEM of four measurements.

Although lettuce fiber isolates retain much water, the water-holding capacity expressed in grams of water held by fiber per gram of edible portion of food has been found to be higher for wheat bran fiber (4.5 g) and carrot fiber (2.1 g) than for that of lettuce and various other foods (0.3–1.3 g). The water-holding capacity of fiber from the brans of ready-to-eat breakfast cereals is positively correlated ($r \geq 0.85$, $P < 0.05$) with its mean particle size (MPS): the water-holding capacity of 160-μm MPS wheat bran fiber is 59% of that of 800-μm MPS wheat bran fiber (Table 2). A large portion of the water held by wheat bran fiber appears to be free water.

The water-holding capacity has a significant effect on fecal output and stool hardness, as discussed above. These factors are important physiological effects of wheat bran and contribute significantly to its laxative effects.

Bile Salt Binding and Serum Cholesterol Reduction

There is considerable interest in the cholesterol-reducing properties of dietary fiber. This is because the influence of serum cholesterol (low-density lipoprotein (LDL) and high-density lipoprotein (HDL) cholesterol, in particular) on cardiovascular disease is considerable. Three different mechanisms have been proposed to account for the serum cholesterol-reducing properties of dietary fiber. As discussed above, a certain soluble viscous fiber may modestly lower cholesterol by increasing luminal viscosity and preventing cholesterol (re)absorption. Acetic and propionic acids, produced by the fermentation of dietary fibers in the gut, are thought to inhibit liver cholesterol synthesis at the metabolic level. Dietary fibers also have the ability to bind cholesterol directly: deconjugated bile salts are bound to pectic substances by hydrogen bonding, and lignin appears to bind bile salts through hydrophobic interactions. The actual

in vivo mechanism is unknown, and it is likely that different fibers operate through different combinations of all three mechanisms to influence blood cholesterol levels.

In vitro evidence suggests that the physical form of the fiber may also play an important role in binding bile salts. In commonly ingested ready-to-eat cereals, glycocholate ($r = 0.90$, $P < 0.001$) and taurocholate ($r = 0.86$, $P < 0.05$) binding have been positively correlated with the MPS of the neutral detergent fiber (NDF; see Determination). Reducing the particle size of wheat bran NDF from 800 to 160 μm reduces its glycocholate binding by 19% (Table 2). The bile salt binding capacity of fiber isolates, although method-dependent, may reflect events in the terminal small intestine. Purified fiber fractions may be poor (cellulose) or strong (lignin) binders. Some rice brans appear to have a high capacity to bind bile salts.

Human cholesterol-lowering studies have linked the consumption of dietary fiber with lower total plasma cholesterol and increased LDL/HDL ratios in addition to reductions in serum triglycerides. Studies such as these have suggested that cardiovascular risk may be reduced by consuming particular types of foods high in dietary fiber, most notably whole cereal grains. However, it should be noted that coronary heart disease (CHD) is a complex disease and is affected by many factors. For example, the reduction in CHD observed in a recent study was larger than would be expected from the beneficial effects of soluble dietary fiber on serum cholesterol levels only. Thus, reductions in CHD are likely to be the result of many factors, one of which is total dietary fiber intake.

Cation-exchange Capacity

Cation exchange is partly dependent on the presence of uronic acid in the nonesterified form. The preparation of the fiber material may decrease the number of nonesterified carboxyl groups and the apparent cation exchange capacity. Wheat bran contains little uronic acid, and its cation exchange capacity is mainly due to diffusion within the dietary fiber network. The cation exchange capacity of fiber from different sources is difficult to compare, unless it is expressed per edible portion. According to the affinity of minerals for carboxylic acid groups, cabbage and coarse wheat bran show a high cation-exchange capacity compared with pectin.

There has been some concern that the ability of dietary fiber to bind minerals (as measured *in vitro*) may lead to mineral deficiencies in individuals consuming high-fiber diets. In North American diets, calculations show that mineral intakes far exceed the potential binding capacity of dietary fiber so that no need for concern arises. In addition, minerals that are bound to fibers or that are trapped in the (as yet, undigested) cell-wall matrix may not be absorbed in the small intestine but could be partially released and absorbed in the colon when the fiber is degraded by bacteria. Absorption of minerals in the colon has been suggested as a mechanism for accounting for increased mineral absorption in rats fed fructooligosaccharides. This may explain why cation adsorption has not been consistently related to mineral bioavailability. However, wheat bran fibers can permanently bind heavy metal ions to decrease their toxicity.

Viscosity and Gelling Properties

As mentioned above, certain soluble fibers, such as oat β-D-glucans, are viscous when dissolved in water, while others, such as pectins, show gelling properties. These fibers influence gastric emptying and absorption rates in the small intestine. Direct measurement of the viscosity produced by concentrations of fiber likely to be used in diets has shown little effect of low methoxy pectin, a slight increase with wheat bran, and a significant increase with high methoxy pectin. Some of the nonstructural dietary fiber components can also increase viscosity.

Microbial Degradation (Fermentability)

Most dietary fiber remains undegraded until it reaches the large intestine, where the extent of fermentation depends on the source as well as several other factors, including the physical structure of the fiber, the presence of specific components in the fiber matrix, the nitrogen source, bacterial adaptation, and transit time. It is generally accepted that soluble fiber is almost completely fermented in the large intestine, but unfortunately, it is also generally believed that insoluble fiber is not fermented. Actually, both insoluble and soluble fiber are extensively fermented. On average, 70–80% of the total dietary fiber from mixed diets (e.g., fruits, vegetables, legumes, and most cereals) is degraded by colonic bacteria. Since insoluble dietary fiber is the major dietary fiber fraction, representing approximately two-thirds of total dietary fiber in Western diets, this means that insoluble fiber is highly fermented. Indeed, as indicated above, 70–80% of the insoluble fiber from oat bran and up to 40% of the insoluble dietary fiber from wheat bran are fermented.

There is some indication that fermentability depends on particle size, with small particle sizes being more readily fermented, but this has not been consistently reported, and the effect appears modest. In rats fed purified diets containing 15% hard red wheat bran, NDF fermentability was largely unchanged (34–36%) when the MPS varied from 1275

to 394 μm. Similar trends were observed when American Association of Cereal Chemists soft and hard wheat brans were fed at various MPS. In these experiments, 37.3–37.6% and 30.0–32.8% of the NDF was fermented, respectively.

Interaction between Structure and Physiology

Food Processing

The physicochemical properties of dietary fiber *per se* are dependent on the chemical composition and the structural characteristics of the fiber. However, as noted above, knowing the former does not allow one to predict the latter. When considering the potential *in vivo* effect of dietary fiber, one must also take food-processing practices into consideration. Many different types of food processing practices are currently employed. Foods can be boiled, canned, frozen, blanched, parboiled, extruded, adiabatically extruded, and milled. All processes can have an effect on dietary fiber content, as measured by current methodologies. For example, heat treatment can break glycosidic linkages in dietary fiber polysaccharides. This can solubilize some of the insoluble fiber as well as reduce the TDF content if the polymers are broken down into small molecular mass fragments. Heat treatment can also reduce the overall length of the polysaccharide chains to lower viscosity and the water-holding capacity of the fiber. The pentosans in dietary fiber can react with amino acids such as lysine (Maillard reaction) and form new polymers that do not behave physiologically like dietary fiber. The boiling of vegetables increases the apparent TDF content, but this seems to be due to the loss of nonfiber components to the water.

It is often difficult to predict before hand the effect of a food-processing method on TDF and the distribution between insoluble and soluble fractions. For example, extrusion cooking of some cereal brans apparently increases its water-holding capacity *in vitro*, although this does not translate into a change in fecal water holding capacity *in vivo*. Extrusion cooking of wheat bran causes small decreases in TDF and insoluble dietary fiber but also gives rise to small increases in soluble dietary fiber. Processing of rice (abrasion or parboiling) removes the seed coat and the majority of the dietary fiber. Some collapsing of the fiber structure also occurs during food preparation and/or mastication due to the reduction of particle size, and the fiber structure of sources such as spinach may be modified by cooking.

Milling is known to significantly affect TDF fermentation in the human colon. Cereal grains have a tough outer seed coat that would be completely resistant to digestion if the cereals were not milled prior to consumption. After milling, the thin-walled dietary fiber structures from the endosperm are partly disrupted and, thus, easily broken during fermentation. Milling, therefore, makes the endosperm accessible to the digestive enzymes in the upper digestive tract but also makes the dietary fiber more accessible to colonic bacteria. However, milling can also negatively affect the physiological properties of the dietary fiber in grains. While coarse grinding and adequate mastication appear to be sufficient for digestion of cereals, fine grinding disrupts much of the cell-wall architecture, and further milling can even interfere with its determination by gravimetric methods. Table 3 shows the effect of milling on the dry bulk volume of wheat bran, as measured in a volumetric cylinder. The volume of the coarse bran was reduced from 5 ml g^{-1} to less than 2 ml after grinding in a M5 Wiley mill using a 0.5-mm screen. Fiber from different sources ground under the same conditions may behave differently during milling, so that their resulting particle sizes may be different.

Table 3 Effect of grinding on dry bulk of wheat brans (ml g^{-1}): duplicate measurements

Sieve aperture [a]	*Bran A* [b]	*Bran B*	*Bran C*	*Bran D*
As is	3.1	3.4	5.1	3.5
2.0 mm (10 mesh)	2.7	2.4	2.8	2.1
1.0 mm (18 mesh)	2.2	2.1	2.4	2.0
0.5 mm (35 mesh)	1.8	1.8	1.9	1.7

[a]Using a MS Wiley mill. Mesh sieve equivalence is indicated in parentheses.
[b]Bran A was American Association of Cereal Chemists-certified soft white wheat bran; brans B–D were soft or hard wheat brans from other sources.

As discussed above, the particle size of unfermented wheat bran fiber may be related to colonic function and fecal characteristics in humans and laboratory animals. Finely ground cellulose or even wheat bran, for example, may have a constipating effect, whereas coarsely ground wheat bran prevents constipation. The influence of particle size on fecal bulking has been measured directly in rats. When rats were fed diets containing 12% hard wheat bran, the fecal wet density was 0.796 ± 0.010 (mean $\pm$ SEM, $n = 12$) in the coarse bran group (geometric MPS: 850 μm) but increased to 0.888 ± 0.013 in the fine bran group (MPS = 308 μm). This increased density was not due to a change in fecal weight but was due to a decrease in fecal volume, paralleling the effect demonstrated for bran in Table 3. In ready-to-eat breakfast cereals, the fiber MPS has been found to vary between 350 and 2000 μm. Increased particle size normalizes gastrointestinal transit time, delays gastric

emptying, and may also affect the extent of insoluble fiber fermentation. Although dietary fiber has many positive effects on the digestive tract, excessive milling of dietary fiber to produce small particles in the range of 7–70 μm may have negative effects. Smaller particles do not exhibit the transit time normalization associated with larger bran particles, and the persorption of a small amount of very fine fiber particles may occur in some regions of the intestinal wall. Contrary to persorbed starch granules, persorbed fiber particles are not degraded in the blood and may impose an increased workload on the kidneys. Data on persorption are scarce, and much work is needed on this subject.

Studies with apples, apple purée, and apple juice have highlighted the link between the physical integrity of the cell wall and the beneficial effects of unprocessed dietary fiber, in this case, a lowering of the glycemic index of a food (the rise in blood glucose that accompanies ingestion and digestion of a carbohydrate-containing food). Only the unprocessed apple had a reduced glycemic index; the apple purée, containing the physically degraded dietary fiber was similar to apple juice in response. This demonstrates the importance of the physical encapsulation of available macronutrients within the fiber matrix. This effect has also been observed with rolled oats where higher glycemic indices were observed with more finely cut oats. Studies such as these show that the integrity of the cell wall is important in bringing about beneficial physiological effects, even in the presence of viscous, soluble fiber types (β-glucans in the cell wall; *see* section Bile Salt Binding and Serum Cholesterol Reduction).

In vitro Measurements

In vitro properties of fiber isolates have been used to estimate their potential *in vivo* effect. However, *in vitro* data should be used with caution. The measurements are influenced by the methods used to follow digestion *in vitro* as well as by the methods used to prepare the fiber. Other confounding influences are also present: the fiber residue obtained *in vitro* may have been altered by heat or chemicals or may still contain residual digestible materials. For these reasons, *in vivo* conditions generally result in a more extensive digestion than that measured *in vitro*.

Digestive Events

An important consideration when examining the physiological effect of dietary fiber is the fact that dietary fiber is concentrated and denuded only in the terminal small intestine. This means that the dietary fiber that is partially associated with digestible materials in the jejunum and duodenum may have different physical properties from those of an entirely denuded fiber matrix at the terminal ileum. In the large intestine, most of the soluble fiber and a large but variable portion of insoluble dietary fiber are degraded by the colonic microflora. As already discussed, fermentation plays an important role in mediating many of the physiological effects associated with dietary fiber.

Definition (Based on Function and Structure)

In the early 1970s, it became evident that cellulose represents only a small part of the total dietary fiber content of foods. Prior to this, there was a tendency to equate dietary fiber with cellulose because that is what was measured using the old crude fiber method. While this helped reinforce the use of the term dietary fiber (cellulose has a fibrillar structure), cellulose represents only a small part of the total dietary fiber content of foods. Thus, at that time, new methodologies were emerging, and it was clear that there was a need for a dietary fiber definition that was linked to the actual physiochemical composition of dietary fiber.

It is in this context that the dietary fiber hypothesis appeared, a few years after Cleave's hypothesis in 1956, which related refined foods and Western diseases. The dietary fiber hypothesis thus appeared as a refocusing of Cleave's original ideas, reversing the original emphasis to become: a diet high in unrefined edible plant material is protective against Western diseases of the bowel. The dietary fiber hypothesis is important in understanding the 'dietary fiber concept,' since it links botanical structure to digestive processes and health outcomes – mainly noninfective colonic diseases such as constipation, diverticular disease, irritable colon, ulcerative colitis, appendicitis, hemorrhoids, polyps, and cancer of the large bowel. It is this link to these and other health outcomes that has characterized much of the work in the dietary fiber field over the last 30 years. The first dietary fiber definition in 1972 sought to describe the plant structures and components that were responsible for mediating these health benefits: "that portion of food which is derived from the cellular walls of plants and is digested very poorly by human beings... Fibre is composed largely of cellulose, ... hemicelluloses, pentosans, pectin and lignin."

There are many different definitions of dietary fiber. However, they all include certain specific and critical aspects. First and foremost, a definition must include the fact that dietary fiber corresponds mostly to the structural material in the edible part of plant foods in the usual human diet. This has been consistently

reflected throughout decades by the inclusion of the terms 'plant cell wall' when referring to dietary fiber (Table 4). The term 'dietary fiber' itself is something of a misnomer, since dietary fiber is not fibrillar (except for cellulose). To reflect this fact, many other terms have been proposed over the years, including unavailable carbohydrate and plantix. None of these terms has succeeded in replacing 'dietary fiber.'

Another important aspect of any dietary fiber definition is that it be resistant to human digestive secretions (including enzymes) and not absorbed in the small intestine. The observation that dietary fiber increased fecal weight initially gave the impression that dietary fiber was resistant to both human digestive secretions and to bacterial digestion in the lower gut (fermentation). While it is true that dietary fiber is not digested and absorbed in the small intestine, a large proportion of dietary fiber is fermented in the colon. The extent of this fermentation is a function of several different factors (see Microbial Degradation (Fermentability) Section), meaning that the extent of fermentation cannot be use as a criterion to define dietary fiber.

Using 'resistance to endogenous digestive secretions' alone as a criterion may be confusing. For example, some dietary fiber may be partly degraded by acids in the stomach, while some small intestine-digestible material can reach the colon where bacterial fermentation occurs. In addition, a definition based on the resistance to digestibility by endogenous digestive secretions could be problematic if the material is not specified as coming from the edible part of plant foods. Indeed, such a definition could be erroneously interpreted as meaning that dietary fiber is any material resistant to *in vivo* digestion or, even as any material measured as dietary fiber by a specific dietary fiber method. Among the materials to be excluded from the dietary fiber definition are 'ruminant fiber' and the products formed during processing and/or cooking. There have been requests to include some other plant components as well as manufactured polymers, and this is a subject of much debate. Among the criteria that could be used to accept or reject a material as dietary fiber, only 'edible plant cell wall' is specific enough to allow an absolute definition. In addition, it links current dietary fiber

Table 4 Use of the term 'plant cell wall' in defining dietary fiber over time

Year	*Text*	*Reference*
1885	The 'cell membranes' of Rubner	Trowell HC (1975) In: Burkitt DP and Trowell HC (eds) *Refined carbohydrate foods and disease*, p. 43. Academic Press, London.
1929	The 'skeletal framework of the plant'	McCance RA and Lawrence RD (1929) The carbohydrate content of foods. Medical Research Council special report series No. 135, pp. 1–173. HM Stationary Office, London.
1981	The 'remnants of plant cells'	AOAC Fiber Consensus (1981) Association of Official Analytical Chemists 95th Annual Meeting, October 1981, Washington, DC.
1985	'Derived from plant cell walls and not digested by human alimentary enzymes'	Trowell H, Burkitt DP and Heaton K (eds) (1985) *Dietary fibre, fibre-depleted foods and disease*, Academic Press, London.
1990	The 'structural plant cell wall composed of polysaccharides and lignin'	Eastwood M (1990) Function of dietary fibre in the large intestine. In: Southgate DAT, Waldron K, Johnson IT and Fenwick GR (eds) *Dietary Fibre: Chemical and Biological Aspects*, pp. 211–219. Cambridge: Royal Society of Chemistry.
1995	'The ''cell wall material of plant foods'' responsible for the physical form and texture of unprocessed plant foods'	Heaton KW (1995) The dietary fibre concept – time for a re-evaluation? In: Sørensen A, Bach Knudsen KE, Englyst HN, Gudmand-Høyer E and Nyman M (eds) *Metabolic and Physiological Aspects of Dietary Fibre in Foods: Recent Progress in the Analysis of Dietary Fibre*, *COST 92*, p. 15. Luxembourg: Commission of the European Commission.
1995	The 'plant cell walls' are the common characteristic of plant foods that constitute a high-fiber diet associated with beneficial effects.	Cummings JH, Hudson GJ, Quigley ME and Englyst HN (1995) The classification and measurement of dietary carbohydrates. In: Sørensen A, Bach Knudsen KE, Englyst HN, Gudmand-Høyer E and Nyman M (eds) *Metabolic and physiological aspects of dietary fibre in food. Recent progress in the analysis of dietary fibre in food*, *COST 92*, pp. 17–36. Luxembourg: European Commission.
1999	'the cell walls of edible plant tissues in the traditional human diet'	Mongeau R, Scott FW and Brassard R (1999) Definition and analysis of dietary fiber. In: Cho SS, Prosky L and Dreher M (eds) *Complex Carbohydrates in Foods*, pp. 305–316. New York & Basel: Marcel Dekker.

definitions to the original 'dietary fiber concept,' which recognized that a diet high in minimally processed edible plant foods provided many health benefits. Confusion often arises because of the overlapping characteristics and physiological effects of the different 'candidate' materials with those of a dietary fiber 'reference.' Examples of 'candidates' are polydextrose (a randomly bonded synthetic glucose polymer), inulin and fructooligosaccharides (naturally occurring glucose-fructose oligomers and polymers), and resistant starch (formed by food-processing methods and largely undigested in the small intestine). This overlapping means that it becomes virtually impossible to distinguish between fiber and nonfiber material based on resistance to digestion alone. Debates such as this started shortly after dietary fiber became popular in the 1970s. They are still ongoing, as demonstrated by the publication of a recent international survey of 147 professionals. While a complete discussion of the merits of either side is beyond the scope of this chapter, it is important to remember the origins of the dietary fiber definition and the dietary fiber hypothesis that firmly linked fiber to physiological outcome.

The definition of dietary fiber was first centered on the 'original dietary fiber concept' that linked the consumption of 'plant cell walls' in unrefined, traditional foods to positive physiological outcomes. Table 5 shows two definitions that were proposed after the turn of the millennium. These two definitions have taken the original ideas of the dietary fiber hypothesis into account as well as included a reference to some specific physiological effects in humans.

While the newer definitions of dietary fiber specifically mention the fact that dietary fiber has physiological benefits, the list is rather specific and deals primarily with effects observed with viscous fibers (lowering post-prandial serum glucose and lowering serum cholesterol) or effects brought about from partial fermentation and water holding capacity (laxation). These effects are well supported by clinical experiments with laboratory animals and with humans. However, the original dietary fiber concept related dietary fiber intake to the diseases of Western man, including heart disease, cancer and obesity. As mentioned above, the link between viscous dietary fiber and heart disease is supported in the literature through its effect on serum cholesterol and triglycerides but the link with other Western diseases has been much harder to demonstrate. It has been well established that eating patterns that favor foods high in dietary fiber and low in red meat consumption significantly reduce the risk of cancer, heart disease and obesity. However, it has been relatively difficult to isolate this effect to dietary fiber consumption *per se*. Part of the problem comes from the almost exclusive reliance on epidemiological studies to demonstrate any association. There are several reasons why epidemiological studies are generally favored. First, these studies provide data on free-living humans rather than in-bred laboratory animals. Second, these studies follow the long-term effect of dietary fiber consumption on health outcomes. Third, the sample size can vary from a few hundred (case control studies) to many thousands of individuals (prospective cohort studies). This gives the studies good statistical power and lends weight to the conclusions. However, epidemiological studies also suffer from a number of shortcomings including: problems with assessing food intake for the past 2–3 years (case control studies), or with monitoring food intake 2–3 times over a period of 5 years and using this to predict disease outcome 15 years later (prospective cohort studies), problems that arise from the use of incomplete dietary tables to calculate dietary fiber intake, and problems arising from the use of inappropriate methods for correcting for energy intake.

The most striking conclusion from epidemiological studies is a failure to provide support for an inverse correlation between dietary fiber intake and colon cancer. This concept has long been entrenched in the literature largely because of studies with laboratory animals that have provided much data in favor of

Table 5 Recent proposed definitions of dietary fiber

Year	Text	Reference
2001	'. . . the edible parts of plant or analogous carbohydrates that are resistant to digestion and absorption in the human small intestine . . . promotes beneficial physiological effects including laxation, and/or blood cholesterol attenuation, and/or glucose attenuation.'	American Association of Cereal Chemists (2001) Report of the Dietary Fiber Definition Committee to the Board of Directors of the American Association of Cereal Chemists. 2001. The definition of Dietary Fibre. *Cereal Foods World* 46: 112–129.
2002	'Dietary Fiber nondigestible carbohydrates and lignin that are intrinsic and intact in plants. Functional Fiber consists of isolated, nondigestible carbohydrates that have beneficial physiological effects in humans.'	National Academy of Sciences (2002) *Food and Nutrition Board. 2002. Dietary Reference Intakes. Proposed Definition of Dietary Fiber*. National Academy Press, Washington DC.

a protective effect of dietary fiber. The animal studies are supported by plausible physiological and biochemical mechanisms that can account for the observed results. However, the epidemiological data is inconsistent in its support for an inverse relationship between colon cancer and dietary fiber intake and clinical trials that examined the effect of diet on the re-appearance of colon polyps (colon cell growths that have a 1/20 chance of forming a cancerous tumor) found no protective effect of dietary fiber. Even though the link between adenomatous polyps and cancer is not straightforward, these studies show that more research is needed to elucidate the relationship between the known physiological benefits of dietary fiber and cancer.

Sources of Dietary Fiber

The total dietary fiber content of a food depends on many factors, including (1) the plant variety, (2) the stage of maturity when harvested, (3) the plant growing conditions, and (4) the method of food preparation. Although not all the methodological issues have been completely resolved with respect to measuring dietary fiber, there is now general consensus on the total dietary fiber content for a majority of foods.

The names of 49 common dietary fiber sources are provided in Table 6. The foods are classified according to their TDF content expressed on an 'as is' basis. Since some methodological variability is inevitable, the letters 'E,' 'M,' and 'P' in the right-hand column designate which methods would give TDF values in the range shown. Other methods not included in the comparison are likely to agree with these values. The letter 'E' refers to NSP as measured by the Englyst GLC procedure and includes a lignin value measured separately with potassium permanganate. The letter 'M' refers to the Mongeau rapid method, and the letter 'P' refers to the Prosky method. The last two methods are gravimetric and include lignin in the final determination. Table 6 shows a considerable agreement between methodologies: 33 of 49 foods are designated 'EMP,' meaning that all three methods give similar values. In addition, three foods are designated 'MP' and 13 foods as 'M' only because sample size limitations precluded measurement with other methods. The TDF values are based on winter and summer collections of up to 20 individual foods over 30 months in different locations.

Most fruits and vegetables contain low amounts of dietary fiber (1.0–2.2% TDF) on an 'as is' basis. Legumes contain intermediate amounts of dietary fiber (around 4% TDF), and cereals contain between 1–2.2% TDF (corn kernels) and 15.5–15.8% TDF (oat bran). The dietary fiber content of foods is greatly influenced by the moisture content. Thus, cereal TDF values are high because these foods contain little moisture (up to 10% at most), whereas fruits and vegetables contain 80–90% moisture on an 'as is' basis. The values of Table 6 are reported as a percentage of weight, but when calculating dietary fiber intake, it is important to factor in the serving size (30 g for breakfast cereals, 50 g for lettuce and raisins, and 125 g for vegetables and fruits). For example, 1 serving of Bran flakes will provide approximately 4 g of dietary fiber, 1 serving of lettuce will provide 0.35 g of TDF, 1 serving of raisins will provide about 1.8 g TDF, and 1 serving of corn will provide about 2 g of TDF.

Table 6 Sources of dietary fiber

TDF[a]	*Foods*	*Methods*[b]
0.1–0.5	White rice, cooked	M
	Orange juice from concentrate	M
0.6–0.9	Lettuce	M
	Cucumbers	EMP
	Pineapple, melons (cantaloupe)	EMP
1.0–2.2	White bead	EMP
	Corn kernel (canned), potatoes (boiled), green beans (boiled), cabbage (raw), celery (raw), onion (cooked), green peppers (raw)	EMP
	Cauliflower (cooked), tomatoes	M
	Apple, bananas, orange, pears, strawberries, cherries, grapefruit, plums	EMP
	Apple pie, blueberry pie, pasta (cooked)	M
2.3–3.0	Rye bread	M
	Mushroom (E = 2.0)	MP
	Beets (canned), rutabaga (raw), carrots (raw), broccoli (raw)	EMP
	Blueberries (E = 1.9)	MP
3.1–4.3	Brown rice	EMP
	Raisins (seedless)	M
4.4–6.4	Bran muffins, wheat cereal, wholemeal bread	EMP
	White beans (baked)	EMP
	Kidney beans (baked), peanut butter	M
6.5–7.3	Peanuts	EMP
	Green peas (boiled)	M
8.0–9.3	Oatmeal cereal	EMP
9.4–10.3	Shredded wheat	EMP
12.3–13.1	Bran flakes	EMP
15.5–15.8	Oat bran	EMP
40.6–40.9	AACC hard wheat bran (E = 36.4)	MP

[a]TDF, total dietary fiber range in grams per 100 g on an 'as is' basis.
[b]Methods: E, Englyst GLC method plus potassium permanganate lignin; M, Mongeau rapid gravimetric method; P, Prosky gravimetric method. From Mongeau R and Brassard R (1989) A comparison of three methods for analyzing dietary fiber in 38 foods. *Journal of Food Composition and Analysis* 2: 189–199; Mongeau R, Brassard R and Verdier P (1989) Measurement of dietary fiber in a total diet study. *Journal of Food Composition and Analysis* 2: 317–326.

Glossary

- Digestive secretions: enzymes and other secretions produced by the digestive organs (e.g., salivary glands and exocrine pancreas, stomach and small intestinal mucosa, and liver); this excludes enzymes produced by intestinal bacteria.
- Digestibility: degradation of ingested food components by endogenous digestive secretions in the upper gastrointestinal tract.
- Fermentability: degradation of undigested food components (e.g., dietary fiber) by the intestinal microflora, mostly in the large intestine.
- Pectic substances: mixtures of acidic and neutral polysaccharides that can be extracted with water from plant tissues and are rich in galacturonic acid; together with hemicelluloses, they form the cell wall matrix in which are enmeshed cellulose fibers.
- Persorption: the paracellular passage of small amounts of solid food components from the lumen of the intestine into the lymphatic and blood circulation, this being without structural changes.

See also: **Bile**; **Carbohydrates**: Classification and Properties; **Cellulose**; **Dietary Fiber**: Determination; Physiological Effects; Effects of Fiber on Absorption; Bran; Energy Value

Further Reading

Anderson JW, Allgood LD, Lawrence A *et al*. (2000) Cholesterol-lowering effects of psyllium intake adjunctive to diet therapy in men and women with hypercholesterolemia: meta-analysis of 8 controlled trials. *American Journal of Clinical Nutrition* 71: 472–479.

Baintner K (ed.) (1986) *Intestinal Absorption of Macromolecules and Immune Transmission from Mother to Young*. Boca Raton, FL: CRC Press.

Bolton RP, Heaton KW and Burroughs LF (1981) The role of dietary fibre in satiety, glucose, and insulin: studies with fruit and fruit juice. *American Journal of Clinical Nutrition* 34: 211–217.

Brown L, Rosner B, Willett W and Sacks FM (1999) Cholesterol-lowering effects of dietary fiber: a meta-analysis. *American Journal of Clinical Nutrition* 69: 30–42.

Carpita NC (1990) The chemical structure of the cell walls of higher plants. In: Krichevsky D, Bonfield C and Anderson JW (eds) *Dietary Fiber: Chemistry, Physiology, and Health Effects*, pp. 15–30. New York: Plenum Press.

Cummings JH (1986) The effect of dietary fiber on fecal weight and composition. In: Spiller (ed.) *Handbook of Dietary Fiber in Human Nutrition*, pp. 211–280. Boca Raton, FL: CRC Press.

Dreher ML (ed.) (1987) *Handbook of Dietary Fiber: An Applied Approach*. New York: Marcel Dekker.

Eastwood M (1990) Function of dietary fibre in the large intestine. In: Southgate DAT, Waldron K, Johnson IT and Fenwick GR (eds) *Dietary Fibre: Chemical and Biological Aspects*, pp. 211–219. Cambridge: Royal Society of Chemistry.

Heaton KW (1995) The dietary fibre concept – time for a re-evaluation? In: *Recent Progress in the Analysis of Dietary Fibre, Metabolic and Physiological Aspects of Dietary Fibre in Foods, COST 92*, p. 15. Luxembourg: Commission of the European Communities.

Lee SC and Prosky L (1995) International survey on dietary fiber: definition, analysis, and reference materials. *Journal of AOAC International* 78: 22–36.

Mongeau R and Brassard R (1982) Insoluble dietary fiber from breakfast cereals and brans: bile salt binding and water-holding capacity in relation to particle size. *Cereal Chemistry* 59: 413–417.

Mongeau R, Scott FW and Brassard R (1999) Definition and analysis of dietary fiber. In: Cho SS, Prosky L and Dreher M (eds) *Complex Carbohydrates in Foods*, pp. 305–316. New York & Basel: Marcel Dekker.

Selvendran RR and Robertson JA (1990) The chemistry of dietary fibre: an holistic view of the cell wall matrix. In: Southgate DAT, Waldron K, Johnson IT and Fenwick GA (eds) *Dietary Fibre: Chemical and Biological Aspects*, pp. 27–43. Cambridge: Royal Society of Chemistry.

Spiller GA (1986) Definition of dietary fiber. In: Spiller GA (ed.) *Handbook of Dietary Fiber in Human Nutrition*, pp. 15–18. Boca Raton, FL: CRC Press.

Trowell H (1972) Ischemic heart disease and dietary fiber. *American Journal of Clinical Nutrition* 25: 926–932.

Wolk A, Manson JE, Stampfer MJ *et al*. (1999) Long-term intake of dietary fibre and decreased risk of coronary heart disease among women. *Journal of the American Medical Association* 281: 1988–2004.

Determination

R Mongeau, Ottawa, Aylmer (Quèbec), Canada
S P J Brooks, Ottawa, Canada

Introduction

This chapter presents a brief overview and comparison of commonly utilized methods for quantifying dietary fiber in food. There are, fundamentally, only two types of methods. Both rely on the removal of nonfiber components by a combination of physical, chemical, and enzymatic procedures prior to measurement of dietary fiber. The gravimetric methods measure dietary fiber by weighing the residue and the gas-liquid chromatographic (GLC) methods measure dietary fiber by hydrolyzing a treated sample with acid to sugar monomers prior to quantification

by GLC. Although, on the surface, both types of methods appear similar, there are fundamental practical differences that have come about because of variations in the individual method steps and differences in the choice of reagents and enzymes. These practical differences have significant consequences since they determine the final composition of dietary fiber as measured by each method.

Procedures have been described for isolating and analyzing dietary fiber in specific foods as well as analyzing carbohydrate polymers that are not digested in the small intestine and are not identified as dietary fiber by the current Association of Official Analytical Chemists (AOAC) methods. The present article, however, will focus on practical methods that are applicable to the routine analysis of a wide range of foods. Most total dietary fiber (TDF) methods also allow for the quantification of different fiber fractions such as soluble fiber, insoluble fiber, hemicelluloses, cellulose, lignin, etc. In this article, only the methods that measure TDF are reviewed. This is partly because of space limitations and partly because the individual fiber components have not yet been positively associated with human physiological effects. For example, soluble and insoluble fiber can be readily measured but these designations have little physiological meaning since the digestive properties and physiological benefits associated with some soluble fiber can be found in some insoluble fibers and the physiological benefits associated with insoluble fibers are also found with some soluble fibers. In addition, some processing techniques can also partly depolymerize insoluble fiber that may then be found in the soluble fiber fraction.

The early work of Van Soest focused on measuring ruminant fiber (mainly hemicellulose, cellulose, and lignin) but these procedures could not be used to measure TDF because they would have underestimated the TDF content of human foods by missing a significant proportion of the soluble dietary fiber. All the modern methods incorporate steps to measure soluble fiber. This is important since it can represent a significant proportion of the TDF of a food. The chemical complexity and physical variability of the dietary fiber fraction in various foodstuffs as well as the various food-processing techniques have contributed to the difficulties encountered in devising adequate dietary fiber methods. Many of the present-day TDF procedures are based on the pioneering work of Southgate.

Ideal reference standards do not exist for dietary fiber analysis. Purified fiber components and unpurified standardized 'reference samples' are both used as standards of analysis, but the former group rarely represents the complex dietary fiber matrix found in foods and the latter group remains variable in composition, structure, and properties.

Historical Perspectives on Dietary Fiber Values

Several different measures of dietary fiber and dietary fiber fractions have been reported for foods over the last 150 years. It is important to put these values into some sort of context to understand the differences between these measurements and the implications for human dietary fiber research.

The crude fiber method was standardized in the late 19th century in Germany. It was initially developed to analyze ruminant feeds and silage. It is a highly empirical method that uses sequential extractions with hot dilute acid and alkali. The crude fiber residue contains no soluble fiber and retains only about 15% of the total hemicelluloses, 50–100% cellulose, and 10–50% lignin. The variability in the method and the variable losses of various dietary fiber fractions mean that crude fiber values are of little value in human nutrition. (*See* **Cellulose**; **Hemicelluloses**; **Lignin**.)

Acid detergent fiber (ADF) represented the first method that reproducibly maintained specific dietary fiber fractions. ADF is obtained by refluxing in $1\,\mathrm{N\ mol\,l^{-1}}$ sulfuric acid containing cetylmethylammonium bromide detergent followed by filtering and washing. This fraction contains mostly cellulose and lignin.

Neutral detergent fiber (NDF) represents the insoluble fraction of TDF. It is obtained by refluxing in a solution of sodium lauryl sulfate followed by filtration and washing. NDF includes most of the cell wall material but excludes the soluble β-glucans and soluble hemicelluloses. It includes the ADF fraction as well as the insoluble hemicelluloses. It thus underestimates TDF because it does not measure soluble dietary fiber.

In some older tables, TDF values are reported as the sum of NDF and pectin (estimated TDF) but, as shown in Figure 1, this value is not always accurate. Figure 1 also presents the relationship between insoluble dietary fiber (essentially NDF but using porcine pancreatic α-amylase to remove starch; see below) and TDF (as determined by the rapid HPB (Health Protection Branch) method; see below) determined for the same foods. The insoluble dietary fiber underestimates the TDF value by approximately 40% but this is not constant. The differences between TDF and insoluble dietary fiber represent actual differences measured on the same food sample but the differences between the estimated TDF and the actual TDF values were performed at different times with

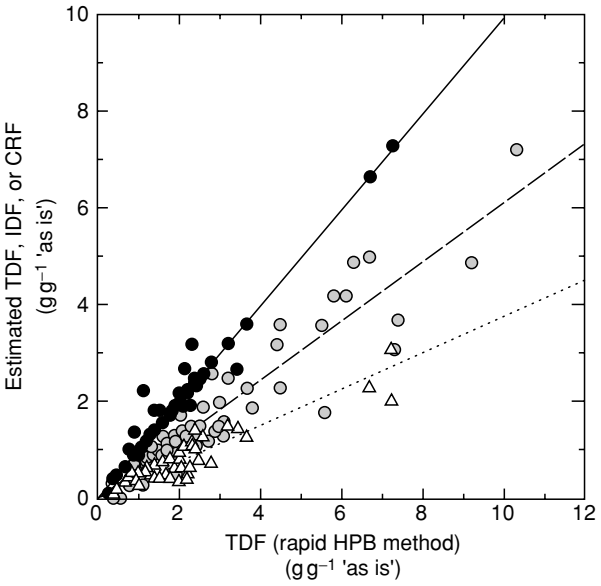

Figure 1 Relationship between crude fiber (CRF; white triangles), insoluble fiber (IDF; gray circles), estimated total dietary fiber (TDF: total of pectin + neutral detergent fiber (NDF), black circles), and TDF values for different types of foods. The TDF and CRF values are from the Canadian Nutrient File (Health Canada 1997) and Mongeau *et al*. (1989). TDF and IDF values were determined by the rapid HPB method (AOAC 992.16, see text). The solid line represents a 1:1 correspondence. The dashed line (IDF) and the dotted line (CRF) are regression lines with the intercept fixed at the origin. Insoluble dietary fiber is essentially NDF but includes a treatment with porcine pancreatic α-amylase. The NDF + pectin values are ultimately from the USDA Nutrient Database (1999).

different procedures and using different food samples so that procedural variations may account for some of the differences. For example, it is known that some previously determined NDF values were measured after incubation at lower (suboptimal) temperatures that may have given higher and more variable values due to incomplete starch digestion.

In addition to the relationships described above, Figure 1 shows the relationship between crude fiber (CRF) and TDF (as determined by the rapid HPB method) for foods from the Canadian Nutrient File. Like the insoluble dietary fiber values, the CRF values significantly underestimate the TDF values; however, in this case, the variability is largely due to the CRF method. Figure 1 shows that no consistent relationship between TDF and CRF or insoluble dietary fiber (and thus NDF). Thus, it is not possible to predict TDF values from either CRF or insoluble dietary fiber: a complete TDF measure is required.

Gravimetric Methods

Gravimetric methods use enzyme and/or chemical treatments to remove material that is digestible in the small intestine. Ideally, these treatments should mimic the digestive processes to arrive at a sample that is representative of the material entering the large intestine of humans. Because dietary fiber represents the structural components of plant cell wall material (*See* **Dietary Fiber**: Properties and Sources), the weight of the residue is corrected for the presence of noncombustible materials (such as minerals) and corrections may also be applied for components that are incompletely digested by the procedure (such as protein). The remaining weight is considered to be dietary fiber. In order to measure the remaining material gravimetrically, it must be filtered through glass crucibles of defined porosities. It is, therefore, important when preparing the sample to insure that it is not too finely ground. The use of a Wiley mill avoids these problems and minimizes particle loss during filtration. Many different gravimetric methods have been published with many variations in reagents, temperatures and incubation times. However, relatively few of these methods have been approved by the Association of Official Analytical Chemists (AOAC) after rigorous testing and interlaboratory verification. Three AOAC approved methods will be discussed in detail in this section: the rapid Health Protection Branch method (rapid HPB method), the Prosky method, and the Lee method. Before presenting these methods, we will first discuss the detergent system.

The Detergent System

When introduced in 1963, the detergent system represented a great improvement over the CRF method and it is still in use today for measuring insoluble dietary fiber. It has been widely used to measure silage and ruminant feed fiber. The system includes the NDF and the ADF procedures, which use hot detergent to remove digestible material and soluble fiber. The detergent system discriminates among the insoluble fibers – hemicelluloses, cellulose, and lignin. A chelating agent in the neutral detergent solution permits the solubilization of insoluble pectin. When NDF and ADF procedures are performed sequentially, the difference in weight between the NDF and ADF extracts is an estimation of the insoluble hemicellulose fraction. The ADF residue contains lignin, cellulose, and (if present) cutin. Lignin can be measured by gently oxidizing the ADF residue with potassium permanganate. A 72% sulfuric acid treatment hydrolyzes cellulose leaving cutin, phlobaphenes, and Maillard products.

NDF and Porcine α-Amylase

Since water-insoluble dietary fibers represent a major fraction of total dietary fiber, the NDF procedure represents a simple method that can form part of an

overall procedure for measuring total dietary fiber in human foods. Nonfiber materials such as starch, most protein, and any fiber-like artifacts remaining in the NDF residue can be efficiently removed by rapid treatment with α-amylase from porcine pancreas (A-3176, Sigma Chemicals Co., St Louis, MO); the rapid treatment consists of 5-min and 60-min incubations at 55 °C and is at least as effective as the 18-h treatment at 37 °C originally proposed by Schaller. The source of the enzyme is specified since pancreatin and other preparations of porcine pancreas α-amylase are much less efficient. Unpurified porcine α-amylase has unique amylolytic and proteolytic activities when applied after the neutral detergent extraction step, making it an ideal enzyme to use for measuring insoluble dietary fiber. (*See* **Enzymes**: Uses in Analysis; **Protein**: Determination and Characterization; **Starch**: Structure, Properties, and Determination.)

The heat-stable α-amylase from *Bacillus subtilis* has also been used to digest starch after the neutral detergent extraction step, but several potential problems arise with this enzyme. It may contain fiber-digesting enzymes that could lead to an underestimation of the cellulose and hemicellulose fractions. On the other hand, it does not completely remove the nonfiber material (such as starch), which will lead to an overestimation of the lignin and possibly the hemicellulose fractions. The final value, therefore, represents a compromise between these two competing factors. Table 1 shows that, for shredded wheat, the NDF residue obtained after treatment with *B. subtilis* α-amylase contained 30% more material than that obtained after treatment with porcine pancreas α-amylase. This residue apparently contained three times more lignin and four times more protein. Both the NDF residue and the apparent lignin content were also higher in samples of bran flakes. Other problems are apparent with the *B. subtilis* α-amylase. For example, toasting wholewheat bread increased the *B. subtilis* α-amylase-measured NDF and lignin values by two- and fivefold, respectively (Table 1), but no such differences were noted with the porcine pancreas α-amylase. The efficacy of the porcine pancreas α-amylase treatment has been verified in various foods and food products and confirmed by other researchers.

In addition to its demonstrated capabilities in avoiding artifactual increases in the estimation of various dietary fiber fractions, porcine pancreas α-amylase enzyme also facilitates rapid filtration in subsequent steps (starch and protein tend to clog the crucibles). The neutral detergent–α-amylase combination is advantageous for research analyses since the neutral detergent solution discriminates well between plant and bacterial cell walls. This is important because bacterial cell walls interfere with the fiber analysis of fecal samples.

Other modifications from the original Van Soest NDF method include the deletion of decalin, 2-ethoxy-ethanol, and sulfite (to maximize the recovery of lignin).

Rapid HPB Method

The rapid HPB method described here is the AOAC No. 992.16 method. In order to obtain a measure of the TDF in a sample, one needs only to combine the insoluble dietary fiber from the NDF + porcine pancreas α-amylase treatment with a specifically tailored method for measuring soluble fiber. In the rapid HPB method, the soluble fiber procedure was devised to maximize the recovery of soluble fiber while excluding protein and other precipitable, nonfiber materials. This latter characteristic is important if one wishes to avoid time-consuming protein determinations that are part of other procedures. The soluble fiber procedure includes a gelatinization step in acetate buffer at pH 4.5 (121 °C) followed by a treatment with heat-stable α-amylase (100 °C) to remove any starch. Starch gelatinization is important to insure complete starch hydrolysis. The soluble material is separated

Table 1 Effect of two sources of α-amylase on the apparent neutral detergent fiber (NDF) and lignin content of breakfast cereals and bread (g per 100 g, dry weight basis)

	Porcine pancreas α-amylase[a]			Bacillus subtilis *α-amylase*		
	NDF	*Lignin*	*Protein*[b]	*NDF*	*Lignin*	*Protein*[b]
Shredded wheat	10.2	0.8	0.4	13.4[c]	2.7[c]	1.7[c]
Bran flakes	10.1	0.9		15.8	2.4	
Wholewheat bread[d]	5.2	0.3		5.9	0.4	
Wholewheat bread, toasted[d]	5.4	0.4		10.1	2.2	

[a]α-Amylase from porcine pancreas (Sigma Chemical Co., catalogue no. A-3176).
[b]Protein measured in NDF residue (N × 6.25).
[c]From Van Soest PJ (1978) Fiber analysis tables. *American Journal of Clinical Nutrition* 31: S284.
[d]From Mongeau R, Brassard R (1980) Rapid digestion of starch and artifact fibre in the measurement of neutral detergent fibre of cereal products. *Getreide Mehl und Brot* 34: 125–127.

from insoluble fiber by filtration and is treated at 60 °C with amyloglucosidase (to finish starch digestion by hydrolyzing branched residues), and then with protease (to hydrolyze the peptide bonds of protein). The soluble fiber is then precipitated in 80% ethanol at room temperature; the sugar and amino acid residues remain soluble in ethanol.

Acetate buffer at pH 4.5 favors the exclusion of protein and appears to prevent the degradation of fiber components that may occur at high temperatures. This does not occur as readily at the relatively higher pH values used by other TDF procedures (e.g., phosphate buffer at pH 7). The heat-processed reference standards extracted from apple and carrot, and used in some collaborative studies, may show lower TDF values due to the loss of pectin during analysis. The differences are less apparent when expressed on a 'per g fresh weight' basis. Without a method for accurately measuring all cell wall components it is difficult to provide an absolute compositional measure of the constituents of dietary fiber as measured by different methods. However, the use of differential procedures to obtain fiber fractions (ADF, NDF), coupled with sulfuric acid digestion, permanganate oxidation and total nitrogen analysis, allows an estimate of the relative contribution of several plant components to the final TDF value.

Figure 2 shows a comparison between several methods for measuring dietary fiber as well as the relative amounts of the components retained by each method. The rapid HPB method retains lignin (a structural polyphenol), the cell wall-associated protein, and a small amount of starch. The significant amount of starch retained by the Prosky method (see below) is readily apparent when TDF values are measured in legumes. Table 2 shows a comparison of four different methods for measuring TDF in green peas. The starch in the residue was estimated by subtracting the values from the total nonstarch polysaccharides (NSPs) + lignin measured by the Englyst procedure. The table also shows the effect of adding porcine pancreas α-amylase.

The method is considered 'rapid' when compared to other fiber methods, particularly when the Fibertec E (for soluble fiber) and the Fibertec I (for insoluble fiber) are used. For example, the maximum number of duplicate determinations using the Prosky TDF method is 20 per week, while 44 duplicate TDF determinations per week are possible using the rapid HPB method. This evaluation takes into account the time required for preparing the reagents and performing the analysis, but not the time required for preparing the samples (freeze-drying, grinding).

Both separate and sequential rapid HPB methods are available and Table 3 presents a comparison between them. In the sequential method, the insoluble residue generated during the soluble fiber determination is treated with neutral detergent and porcine α-amylase yielding the insoluble fiber residue. Thus, a single sample generates both insoluble and soluble fiber measurements to give the TDF value. The separate method requires two samples for analysis. Using a

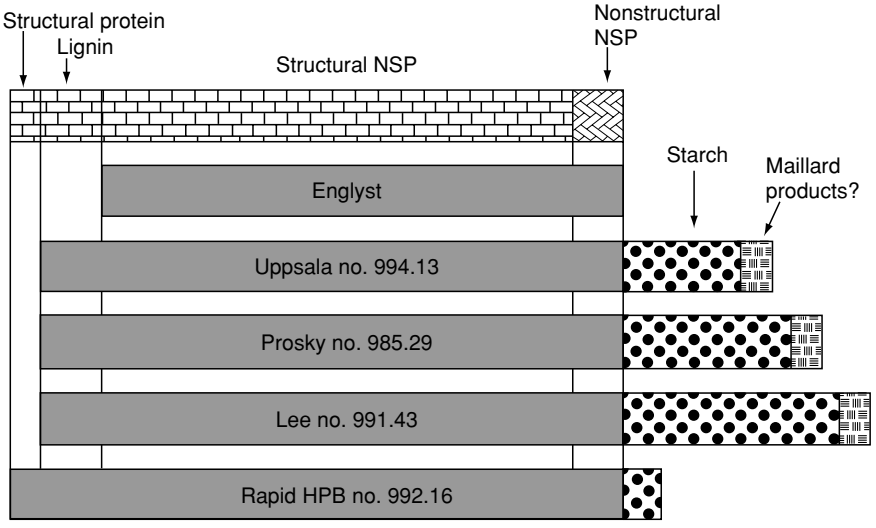

Figure 2 Material measured as dietary fiber by unmodified methods for all food categories. The surface area of the box is proportional to its total dietary fiber (TDF) value. It shows whether the method includes lignin and structural fiber protein or if it excludes starch and Maillard products. The cell wall material corresponds to the structural nonstarch polysaccharides (NSP), lignin, and structural protein. The larger amounts of starch retained by the Theander, Prosky, and Lee methods are most evident when measuring TDF in legumes. AOAC numbers for each procedure are indicated. Adapted from Mongeau R (1995). HPB, Health Protection Branch, Health Canada.

Table 2 Total dietary fiber (TDF) in boiled dried green peas as measured by four different methods with or without added porcine pancreas α-amylase

Method	*TDF on a 'per g dry weight basis' (%)*	*Estimated starch content (%)*
Rapid HPB method, separate	9.9 ± 0.1	2.7
Prosky TDF method	19.0 ± 0.2	11.8
Prosky TDF method + porcine pancreas α-amylase	12.3 ± 0.3	5.1
Lee TDF method	18.8 ± 0.3	11.6
Lee TDF method + porcine pancreas α-amylase	12.3 ± 0.3	5.1
Englyst method (NSP + lignin)	7.2 ± 0.2	
Englyst method (omit DMSO)	10.4 ± 0.1	3.2

NSP, nonstarch polysaccharides; DMSO, dimethyl sulfoxide.
Starch content estimated by subtracting the NSP + lignin value from the TDF estimate. Table adapted from Mongeau R and Brassard R (1994) Comparison and assessment of the difference in total dietary fiber in cooked dried legumes as determined by five methods. *Journal of AOAC International* 77: 1197–1202.

Table 3 Comparison of total dietary fiber (TDF) values by gas liquid chromatography (GLC: nonstarch polysaccharides NSPs + lignin) or gravimetric methods (on a g 100 g^{-1} dry weight basis)

Food	*NSP*[a]	*NSP*[b]	*Permanganate lignin*	*Gravimetric TDF values*[c]		
				Prosky	*Rapid separate*[d]	*Rapid sequential*
Apple	9.9–14.7	13.4	0.8	14.6	14.1	14.6
Beans, green	27.8–32.2	30.6	2.1	33.7	31.3	29.1
Bread, white wheat	2.6–3.3	3.0	0.6	3.5	3.0	3.9
Carrot	22.8–28.9	23.5	1.8	25.4	23.4	23.4
Corn kernels, canned	6.7–8.9	5.9	1.0	7.7	7.8	7.6
Cornflakes	0.7–2.4	1.1	< 0.1	2.5	1.1	2.6
Flour, white wheat	3.0–3.7	2.6	< 0.1	3.0	3.9	3.6
Oats, rolled	7.1–9.5	8.2	0.9	10.5	11.0	9.6
Peas, canned	20.4–22.3	–	–	–	23.3	20.9

[a]Range of values reported by Anderson and Bridges (1988), Englyst and Cummings (1989), Englyst *et al*. (1995) and Marlett (1992): see Further Reading.
[b]NSP as determined by the Englyst GLC method.
[c]TDF as determined by the Prosky method and the rapid separate and sequential methods.
[d]The separate method required analyzing two separate samples to obtain the TDF value.

single sample to measure TDF gives a more accurate and reliable measurement because it eliminates potential cross-contamination that could give rise to falsely high TDF readings. For example, if the insoluble fiber fraction contains some small amount of soluble fiber, it will be counted twice in the final value (once in the insoluble fraction and once in the soluble fraction). The far right column of Table 3 shows that the TDF values are comparable when measured by either procedure and demonstrates the variability in the procedure.

In a comparison of 28 foods from various categories, the rapid sequential HPB method agreed well with the Prosky TDF method (slope = 0.99, intercept = −0.11, $r^2 = 0.985$) and with the Englyst GLC method (slope = 1.03, intercept = 0.15, $r^2 = 0.989$). For comparison with the Englyst method, it was necessary to measure lignin in each food and add its weight to the total NSP (obtained from the Englyst analysis). This gave a TDF value that was comparable to that measured gravimetrically. The regression parameters indicate that the rapid HPB method and the Prosky TDF method retain similar material in the final fraction (Figure 2). However, the negative intercept shows that the Prosky TDF method gives a higher value than the rapid HPB method when measuring foods containing low amounts of fiber. (*See* **Carbohydrates**: Classification and Properties; **Chromatography**: Gas Chromatography.)

Prosky, Lee, and Asp Methods

The Prosky TDF method has also been referred to as the AOAC method since 1985. The Prosky, Lee, and Asp methods are similar but differ in the sources and types of enzymes as well as in the type and pH value of the buffers. The TDF values from all three methods are comparable (see Table 2 for a comparison between the Prosky and Lee methods for green peas). In the Prosky and Lee TDF methods, duplicate samples are treated with α-amylase, protease, and amyloglucosidase to remove digestible material. Four volumes of 95% ethanol is then added and the precipitate filtered through celite C-211 (Fisher Scientific, Mississauga, Ontario, Canada). The residue

Table 4 Polysaccharide and unidentified material in the Prosky total dietary fiber (TDF) method for selected food groups

	Prosky TDF method residue (g 100 g^{-1} dry matter)		
Food	*Starch*	*Total polysaccharides*	*Unidentified material*
Bread	1.0	5.1	0.8
Other cereals	0.3	3.8	1.0
Green vegetables	1.5	25.6	4.6
Other vegetables	1.8	10.2	7.5
Canned vegetables	2.3	12.9	4.0
Fresh fruit	0.2	8.6	3.5
Fruit products	0.1	2.5	1.5
Nuts	0.4	6.7	2.5

Adapted from Englyst *et al*. (1995): see Further Reading. Values represent TDF measurements after correction for ash and 'protein' (N × 6.25). Part of the unidentified material represents lignin but the values are too large to be accounted for as lignin only.

contains a substantial amount of nitrogen requiring a total nitrogen analysis (N) and subsequent back-calculation to correct for protein content (N content × 6.25). One of the duplicates is also analyzed for ash as in other gravimetric analyses and this, too, is subtracted from the residue weight. It is proposed that the use of this protein conversion factor leads to overestimation of dietary fiber (Table 4). The additional protein determination step prevents this method from being considered rapid. As pointed out above and in Tables 3 and 4, the final insoluble dietary fiber residue can retain significant amounts of starch. In a simplified version of the Prosky method, the protease step is skipped since protein is measured in the final residue. This means that a larger amount of protein is subtracted as compared with the original Prosky procedure and this may contribute to some variability. The nature of the celite used in different laboratories has also been considered as a source of variability. Other modifications have also been proposed to improve the precision of the Prosky method. The method comparisons in Tables 2 and 3 were confined to the results obtained with the original version of the Prosky method.

Complete Polysaccharidic Components (GLC Methods)

Southgate published the original TDF method that required the hydrolysis of fiber into its individual polysaccharidic components. In Southgate's method, TDF represented NSPs plus lignin but the method also provided values for hemicellulose and cellulose content. The method used 85% (v/v) alcohol to precipitate soluble starch residues followed by gelatinization and starch digestion with an appropriate enzyme. The original enzyme used by Southgate (takadiastase) had both amylolytic and proteolytic activities. However, in the 1970s, this enzyme became unavailable and other enzymes were used as replacements. It is apparent that some of these replacement enzymes are often less effective (Tables 2 and 3) but the α-amylase from porcine pancreas appears to be adequate. The original Southgate method measured the NSPs colorimetrically but now more specific methods are available such as GLC (*See* **Spectroscopy**: Visible Spectroscopy and Colorimetry.)

The more modern chemical methods use a combination of enzyme and/or chemical treatments to remove digestible material. The residue is then hydrolyzed by acid prior to measurement of the sugar content. The acid treatment needs to be severe enough to assure complete hydrolysis but mild enough to prevent monomer degradation. The neutral fiber polysaccharidic constituents are measured chemically, usually by GLC after derivitization to a measurable form (e.g., by forming alditol acetates). Chemical derivitization is not required when high-performance liquid chromatography (HPLC) is used to measure the residue concentration. Acidic sugars (uronic acids) are measured colorimetrically. The total NSP content is obtained by appropriate calculations and TDF is calculated as the sum of NSPs + lignin. The main GLC methods have been published by Anderson, Englyst, and Theander and their coworkers. Marlett uses a modified Uppsala method. The GLC methods are different in several aspects, including the starch gelatinization step, enzyme treatments, and conditions of acid hydrolysis. This explains why the methods may be in agreement for some foods, but not for others. (*See* **Chromatography**: High-performance Liquid Chromatography.)

Englyst Method

Like all GLC methods, the Englyst method measures NSPs (neutral sugars + uronic acids). However, unlike other methods, the Englyst method is completely applicable to processed foods because it makes use of dimethyl sulfoxide (DMSO) during the gelatinization step to disperse the starch completely. This is followed by treatment with a heat-stable α-amylase and by a mixture of pullulanase-pancreatin

in acetate buffer at pH 5.2. This treatment completely removes all the starch, leaving only the nonstarch polypeptides (Figure 2). The total insoluble and soluble NSPs are then precipitated in 80% ethanol and collected by centrifugation prior to hydrolysis with sulfuric acid. Neutral sugar residues are measured by GLC as alditol acetates and uronic acids are measured colorimetrically. According to Englyst and Cummings, the method can be performed in approximately the same amount of time as the Prosky TDF method. However, the GLC instrument must be equipped with an autosampler for the comparison to be valid. A rapid version of the Englyst method also exists with the sugar components measured colorimetrically but this does not represent a substantial reduction in assay time when compared to the GLC method with an autosampler.

The Englyst method is the only dietary fiber method that does not include lignin. The role of noncarbohydrate components (which include lignin) in *in vivo* fiber digestion may be significant (*See* **Dietary Fiber**: Properties and Sources). There is some debate over whether some of the starch and structural protein should be considered as dietary fiber. (*See* **Dietary Fiber**: Properties and Sources). These components would have to be measured by other methods if they were included in the TDF value.

Uppsala Method

In the Uppsala method, there is a provision for an initial extraction in 80% ethanol if the sample contains very large amounts of mono-, di-, and oligosaccharides. This extraction appears important in reducing any coprecipitation of nonpolymeric sugars and fiber. Starch is digested in acetate buffer at pH 5.0 by incubating with a thermostable α-amylase and then with amyloglucosidase. DMSO is not used and starch may be less efficiently dispersed and, therefore, less efficiently removed (Figure 2). After reprecipitating in ethanol and centrifuging, the residue is hydrolyzed with sulfuric acid. As in the Englyst method, neutral polysaccharides are analyzed as alditol acetates by GLC and uronic acids by decarboxylation with special equipment or by colorimetry.

Unlike the Englyst method, Klason lignin is measured gravimetrically. Potential errors can arise when lignin is measured independently and added back to NSPs. For example, the lignin fraction may include other, nondigested fiber material (such as cellulose) which would then overestimate insoluble fiber. This was apparently the reason behind an apparent lignin value of 0.9–3.5% (dry weight basis) measured in apples, peas, or a food composite. This value is suspect because low permanganate lignin values (obtained after treatment of the insoluble fiber residue with porcine pancreas α-amylase and subsequent workup to obtain an ADF residue) were obtained which were independent of any food-processing method (similar to the values in Table 1). Marlett has discussed the necessity of measuring lignin and the problems related to its measurement.

Anderson Method

As in the Englyst method, the Anderson GLC method uses DMSO to disperse starch prior to hydrolysis. The starch is then digested with porcine pancreas α-amylase in acetate buffer at pH 5.2. After precipitation with four volumes of ethanol, the residue is extracted with water at 100 °C to obtain soluble and insoluble fiber fractions. Following centrifugation, the soluble fiber from the supernatant and the insoluble fiber from the pellet are hydrolyzed with sulfuric acid. Klason lignin is measured as the loss in weight of the nonhydrolyzed insoluble residue after ashing.

Comparison of GLC and Gravimetric Methods for Measuring TDF

As already indicated, there is general agreement between the rapid HPB method and the Prosky and Englyst methods. The regression coefficients were 0.985 and 0.989 for the Prosky and Englyst methods, respectively. Other comparisons have also shown good agreement between values. For example, in a comparison of 25 different foods, TDF values as measured by the rapid HPB method agreed with those reported by Anderson and Bridges (modified Englyst method) and with those reported by Marlett (modified Uppsala method). However, more direct comparisons are needed to determine the potential problems with each method.

Table 3 presents a comparison of the TDF values expressed on a 'per g dry weight' basis to magnify differences which may be partly hidden when the values are expressed on a 'per g fresh weight' basis. In order to compare TDF values, the permanganate lignin values need to be added to the NSP values. Three different levels of variability are apparent from the data shown in Table 3. The range of NSP values obtained from three different laboratories (using three slightly different methods and different food samples) illustrates, in part, the intermethod variability. In addition, there is some interlaboratory variability, as shown by a comparison of four laboratories as well as an inherent variability in the samples (including stage of maturation of the product; (*See* **Dietary Fiber:** Properties and Sources). Direct comparisons are only possible for the results in the five right-hand columns. These samples were composites based on winter and summer collections of up to 20 individual

foods over 30 months in different locations. The results show that, when lignin is included in the final TDF value, there is good agreement between GLC and gravimetric methods for the indicated foods.

Nutritional Labeling and Analysis

It is difficult to decide on a single method for measuring the dietary fiber content of all foods. The challenge is to find a single method that works adequately for all food groups and is independent of sample processing. Several candidates are possible but to a large extent, the final decision will reflect the definition of dietary fiber. (*See* **Dietary Fiber:** Properties and Sources). The dietary fiber definition must not be driven by available methodologies but must be arrived at independently of methodological considerations. As shown in Figure 2, it is possible to make a case for including structural protein, lignin, structural, and nonstructural NSPs, as well as some starch. Each method includes variable amounts of all these components. In addition to these complications, it is also possible to separate TDF into insoluble and soluble components. Once again, each method uses different procedures to measure these fractions so that the measured content of insoluble and soluble fiber will differ from laboratory to laboratory. Consumption of these different fiber fractions has not yet been related to definite physiological effects, although there appears to be a role for viscous soluble dietary fiber in reducing the risk factors for cardiovascular disease. The role of starch and other nonstarch polymers not traditionally associated with dietary fiber has yet to be determined as well. Therefore, the following discussion deals only with practical considerations in choosing a TDF method for food labeling.

Equipment

A method requiring simple, inexpensive equipment would be the most widely applicable because it would be more acceptable to nations unwilling or unable to spend considerable sums of money to monitor labeling. In general, the gravimetric methods require simpler equipment and the rapid HPB method is less costly than other gravimetric procedures because it does not require elemental nitrogen analysis. The Englyst colorimetric method can be completed with a simple spectrophotometer, although this method is less robust than the GLC method. Except for the Englyst method, all methods require a muffle furnace for ashing the samples. Although the gravimetric systems require simpler equipment, a hot-water supply is needed and analyses are greatly aided by purchasing semiautomated filtering/refluxing equipment that would increase the cost. However, these pieces of equipment are usually less expensive to maintain than GLCs.

Operator Time

Because it does not include a protein determination, the rapid HPB method requires the least amount of operator time per duplicate determination. Methods that require sugar derivitization and analysis by GLC tend to be longer than the gravimetric methods, especially when lignin measurement is included. The time required for these methods is greatly reduced if the GLC system includes an autosampler. Semiautomatic equipment is available for the gravimetric determinations but filtering time can be slow, especially for samples that contain a considerable amount of starch.

Lignin

The Englyst method is the only one that does not include lignin in the final TDF value. Most researchers consider lignin to be an important dietary fiber constituent (*See* **Dietary Fiber:** Properties and Sources) and, consequently, most methods include a measure of the lignin content. In the GLC methods, the adoption of the Klason method (residue after sulfuric acid digestion) for measuring lignin means that other nonfiber material such as Maillard products may be included in the lignin fraction. This could lead to overestimation of the lignin content and, consequently, overestimation of the insoluble dietary fiber and TDF values. Lignin may be more accurately measured by a combination of NDF + porcine pancreas α-amylase and ADF treatments prior to permanganate oxidation. This would, however, increase the time required for analysis as well as increase the amount of sample needed per assay.

Independence from Food Processing

Different methods include different amounts of starch in the final TDF estimate (Figure 2). The only method that does not include any starch is the Englyst method because it completely solubilizes the starch with DMSO prior to enzymatic hydrolysis. There are several different types of starch, some of which arise from food preparation such as retrograde starch and so the inclusion of starch in a method could be problematic. An example of the potential for problems comes from measuring the TDF values in cooked and uncooked rice and potatoes (high-starch foods) using the Prosky method. The Prosky method gave higher TDF values in the cooked foods. Problems were also noted with roasted peanuts where nonfiber material was included with TDF as well as toasted wholewheat bread (**Table 1**). Lower cooking temperatures (i.e., 95 °C compared to 100–120 °C)

artificially increased the TDF content of white navy beans when the Prosky or Lee methods were used. Some of the problems may have been related to the choice of enzyme used to degrade starch, as was observed for toasted wholewheat bread (**Table 1**).

In general, the Englyst method is independent of the effects of food processing because it uses DMSO to aid in removing starch and nonpolymeric sugars. This helps exclude food preparation artifacts, such as retrograde starch, as well as Maillard products from the final analysis but it also means that lignin is not included. It also means that lignin is excluded from the final analysis. Other methods that rely on GLC identification of NSPs include Klason lignin but this analysis has other problems (see above). Of the gravimetric methods, the rapid HPB method appears to be independent from food processing and, at the same time, includes lignin in the final value.

Precision

Method variability pooled over all foods has been measured in a US Department of Agriculture study using 25 duplicate food samples. Different laboratories were asked to measure TDF in the samples without knowing their composition or source. The coefficient of variation for a single observation was 3.0% with the rapid HPB method, 4.7% with the Prosky TDF method, and 8.4% for the Uppsala method. This represents an estimation of the intralaboratory precision for selected unprocessed and processed foods. More interlaboratory work comparing precision of several methods is needed.

Remaining Problems

The general agreement between the different methods for measuring fiber is encouraging because it opens the door to universal acceptance of a dietary fiber method. However, many problems remain. First, an alcohol precipitation step is used to separate fiber material from nonfiber material. This step usually follows starch digestion and is designed to precipitate sugar polymers. It is possible that some of the fiber components are not completely precipitated and that some sugars released by starch digestion are coprecipitated. This may vary with the method and may be the source of some of the intermethod variability. In addition, the selectivity of this step in separating sugar polymers from nonpolymeric compounds has been questioned. In particular, polymers such as oligofructose (actually a mixture of oligomers and polymers) do not precipitate during the subsequent centrifugation. The 'dietary fiber' nature of these polymers is currently being debated. (*See* **Dietary Fiber**: Properties and Sources.)

Second, structural fiber protein may represent up to 10% of dietary fiber in immature plant tissues. It is impossible, at the present time, to quantify this protein (although the rapid HPB gravimetric method presumably includes it). Except for the rapid HPB method, this fraction is excluded from the reported TDF values.

Third, the colorimetric method for determination of uronic acid values (included in most GLC methods) needs to be reexamined. This is because the uronic acid content of vegetables was shown to be higher when measured by a decarboxylation method.

Fourthly, there are a small number of foods where significant method disagreement occurs. For example, for soya bean fiber isolate the NSP value using the Englyst GLC method is lower than the TDF value measured gravimetrically. The reasons for these discrepancies need to be examined in order to determine the exact TDF composition measured by each method. These comparisons will greatly improve fiber methodology.

Finally, although both acetone and ether have been used to remove fat from the samples prior to analysis, their efficacy is questionable and only the defatting procedure of Bligh and Dyer is efficient enough to defat all food samples completely. Caution is warranted when substituting defatting procedures since TDF methods have been validated with their respective defatting methods and substitutions may produce unexpected results. For example, the degree of lignin removal with the Bligh and Dyer procedure is unknown. There is some question as to the necessity for complete defatting since the methods were validated with a wide variety of food samples, all with different fat contents. It may be necessary to reexamine the importance of fat removal prior to analysis in samples containing high amounts of fat.

Conclusion

Comparable TDF values can be obtained for many foods using any of the methodologies discussed above. Some foods are more problematic and require a careful choice of methodology. Significant causes of error have been identified, including the potential precipitation of incompletely digested starch and the measurement of Klason lignin, particularly in processed foods. The inclusion of Maillard products in Klason lignin may significantly increase the final TDF value. The following treatment is recommended: NDF, porcine pancreas α-amylase, ADF, and permanganate treatment. This procedure gives permanganate lignin values that are independent of food processing and, when added to NSPs determined by GLC, provides a detailed and known measurement of dietary

fiber. The rapid HPB method, including the neutral detergent and porcine α-amylase treatments, has several advantages and would serve as a good method for food labeling.

See also: **Carbohydrates**: Classification and Properties; **Cellulose**; **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **Dietary Fiber**: Properties and Sources; **Enzymes**: Uses in Analysis; **Hemicelluloses**; **Lignin**; **Protein**: Determination and Characterization; **Spectroscopy**: Visible Spectroscopy and Colorimetry; **Starch**: Structure, Properties, and Determination

Further Reading

Anderson JW and Bridges SR (1988) Dietary fiber content of selected foods. *American Journal of Clinical Nutrition* 47: 440–447.

Armstrong A (1989) US Department of Agriculture Collaborative Study 1988. *Total Dietary Fibre. HPB Laboratory (Canadian) Results*. Statistical Report. Food Statistics and Operational Planning. Health and Welfare, Canada.

Bach Knudsen KE and Mongeau R (1995) Analytical methods for the determination of components conventionally considered as dietary fibre. In: Sørensen A, Bach Knudsen KE, Englyst HN, Gudmand-Høyer and Nyman M (eds), *Metabolic and Physiological Aspects of Dietary Fibre in Food. Recent Progress in the Analysis of Dietary Fibre COST 92*, pp. 205–206. Luxembourg: Commission of the European Communitites.

Bligh EG and Dyer WJ (1959) A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry* 37: 911–917.

Cho S, Devries J and Prosky L (1997) *Dietary Fiber Analysis and Applications*. Gaithersburg, MD: AOAC International.

Dreher ML (1987) *Handbook of Dietary Fiber. An Applied Approach*. New York: Marcel Dekker.

Englyst H and Cummings J (1989) Dietary fibre and starch: definition, classification and measurement. In: Leeds AR (ed.) *Dietary Fibre Perspectives. Reviews and Bibliography*, pp. 3–26. London: John Libbey.

Englyst HN, Quigley ME, Englyst KN, Bravo L and Hudson GJ (1995) *Dietary Fibre. Measurement by the Englyst NSP Procedure. Measurement by the AOAC Procedure. Explanation of Differences*. Report of a Study Commissioned by MAFF, UK. Medical Research Council & University of Cambridge.

Jenkins DA, Kendall CWC and Vuksan V (2000) Viscous fibers, health claims, and strategies to reduce cardiovascular disease risk. *American Journal of Clinical Nutrition* 71: 401–402.

Marlett JA (1990) Analysis of dietary fiber in human foods. In: Kritchevsky D, Bonfield C, Anderson JW (eds) *Dietary Fiber – Chemistry, Physiology and Health Effects*, pp. 31–48. New York: Plenum Press.

Marlett JA (1992) Content and composition of dietary fiber in 117 frequently consumed foods. *Journal of the American Dietetic Association* 92: 175–186.

Mongeau R (1995) Experiences with different official analytical methods in Canada. In: Sørensen A, Bach Knudsen KE, Englyst HN, Gudmand-Høyer E and Nyman M (eds) *Metabolic and physiological aspects of dietary fibre in food. Recent progress in the analysis of dietary fibre, COST 92*, pp. 143–151. Luxembourg: European Commission.

Mongeau R and Brassard R (1989) A comparison of three methods for analyzing dietary fiber in 38 foods. *Journal of Food Composition and Analysis* 2: 189–199.

Mongeau R and Brassard R (1993) Enzymatic-gravimetric determination in foods of dietary fibre as sum of insoluble and soluble fibre fractions: summary of collaborative study. *Journal of AOAC International* 76: 923–925.

Mongeau R and Brassard R (1995) Importance of cooking temperature and pancreatic amylase in determination of dietary fiber in dried legumes. *Journal of AOAC International* 78: 1444–1449.

Mongeau R, Brassard R and Verdier P (1989) Measurement of dietary fiber in a total diet study. *Journal of Food Composition and Analysis* 2: 317–326.

Mongeau R and Brooks SPJ (2001) Chemistry and analysis of lignin. In: Cho SS and Dreher ML (eds) *Handbook of Dietary Fiber*, pp. 321–373. New York, Basel: Marcel Dekker.

Theander O (1981) A review of the different analytical methods and remaining problems. In: James WPT, Theander O (eds) *The Analysis of Dietary Fiber in Food*. New York: Marcel Dekker.

Theander O, Åman P, Westerlund E, Andersson R and Pettersson D (1995) Total dietary fiber determined as neutral sugar residues, uronic acid residues, and Klason lignin (the Uppsala method): collaborative study. *Journal of AOAC International* 78: 1030–1044.

US Department of Agriculture Research Service (1999) *USDA Nutrient Database for Standard Reference. Composition of Foods: Raw, Processed, Prepared*. Release 12.

US Department of Agriculture, Agriculture Research Service (1999) *USDA Nutrient Database for Standard Reference, Release 13*. Nutrient Data Laboratory Home Page, www.nal.usda.gov/fnic/foodcomp.

Physiological Effects

I T Johnson, Institute of Food Research, Norwich, UK

Introduction

The term ‘crude fiber’ was coined in the nineteenth century to describe the material that remained after rigorous nonenzymic hydrolysis of animal feedstuffs.

The possible importance of a similar undigestible residue in human foods was proposed in the early 1970s by the physician and epidemiologist Hugh Trowell, but he also recognized that the crude fiber figures then available for foods had little physiological significance and were of no value in the context of human diets. He therefore used the term 'dietary fiber' to describe the 'remnants of plant cell walls resistant to hydrolysis (digestion) by the alimentary enzymes of man.' Later, he and others redefined it more precisely as 'The sum of lignin and the plant polysaccharides that are not digested by the endogenous secretions of the mammalian digestive tract.' Although some controversy still exists as to the precise definition of dietary fiber, there is no doubt that the term embraces a large and complex mixture of polysaccharides that share few common properties other than resistance to digestion by the endogenous enzymes of the human alimentary tract. The presence of undigested cell wall fragments and dispersed polysaccharides can alter physiological processes throughout the gut, but the effects of different fiber components depend upon their varied physical and chemical properties during digestion, and upon their susceptibility to degradation by bacterial enzymes in the colon. It is important to realize that a single analytical value for the fiber content of a food is a poor guide to its physiological effects. This article will review the main mechanisms of action of resistant polysaccharides in the alimentary tract and their implications for human health.

Fiber in the Alimentary Tract

Food is conveyed through the alimentary tract by rhythmic muscular activity. Digestive enzymes are released into the lumen at intervals to facilitate the breakdown of macromolecules into fragments that can then be absorbed and utilized by the body. The earliest stages of digestion begin in the mouth, where food particles are reduced in size, lubricated with saliva, which also contains the digestive enzyme salivary amylase, and prepared for swallowing. In the stomach, any large particles that remain are broken down by rigorous muscular activity in the presence of hydrochloric acid and proteolytic enzymes. The need to disrupt and disperse intractable food particles and cell walls appears to delay the digestive process significantly. Since cell wall polysaccharides help to determine the texture of many plant foods, and hard foods tend to be chewed more thoroughly than soft foods, the presence of dietary fiber in unrefined foods may begin to regulate digestion at a very early stage. Apples, for example, tend to be consumed more slowly than apple juice, and the absorption of sugar from apples is also slower. Similarly, the rate at which the starch is digested and absorbed from cubes of cooked potato has been shown to be much slower when they are swallowed whole than when they are chewed normally. The nutritional significance of these effects is not entirely clear, but they may limit the rate at which glucose from some foods enters the circulation.

Small Intestine

The small intestine is the largest of the digestive organs, with the greatest surface area, as befits its role as the main site of nutrient absorption. The semiliquid products of gastric digestion are released at intervals into the small intestine and propelled away from the gastroduodenal junction by peristatic movements at about $1\,\mathrm{cm\,min^{-1}}$. The hydrolysis of proteins, triglycerides, and starch continues within the duodenum and upper jejunum, under the influence of pancreatic enzymes. The final stages of digestion occur at the mucosal surface, where the products are absorbed into the circulation, along with water and electrolytes, via the specialized epithelial cells of the intestinal villi. Muscular activity in the small intestinal wall, together with rhythmic contractions of the villi, insures that the partially digested chyme is well stirred. In adults, the first fermentable residues from a meal containing complex carbohydrates enter the colon approximately 4.5 h after ingestion. When a solution containing indigestible sugar is swallowed without food, it reaches the colon about 1.5 h earlier than when the same material is added to a solid meal containing dietary fiber. The presence of solid food residues slows transit, probably by delaying gastric emptying and perhaps also by increasing the viscosity of the chyme so that it tends to resist the peristaltic flow. Soluble polysaccharides such as guar gum, pectin and β-glucan from oats increase the mouth-to-cecum transit time still further.

Carbohydrate Metabolism

In his early work, Trowell proposed that dietary fiber was a major factor in the prevention of the metabolic condition insulin-dependent diabetes. Indeed, he argued that historically, this condition was unknown in western Europe prior to the introduction of mechanized flour milling, because until then, the consumption of unrefined carbohydrate foods had favored slow absorption of glucose, which placed less strain upon the ability of the pancreas to maintain glucose homeostasis. Whatever its merits, this hypothesis has encouraged a great deal of research on the effects of cell wall polysaccharides on the digestion and absorption of carbohydrates, which in turn has shed light on

the potential role of dietary fiber in the control of diabetes. The 'glycemic index' is essentially a quantitative expression of the quantity of glucose appearing in the bloodstream after ingestion of a carbohydrate-rich food. To calculate the index, fasted subjects are given a test meal of the experimental food containing a standardized quantity of carbohydrate. The change in concentration of glucose in the blood is then measured over a period of time. The ratio of the area under the blood-glucose curve in response to the test meal to that produced by an equal quantity of a standard reference food is then calculated and expressed as a percentage. When glucose is used as the standard, complex starchy foods often have glycemic indices lower than 100%.

The physical resistance of plant cell walls during their passage through the gut varies considerably from one food to another, and any cell walls that remain intact in the small intestine will impede the access of pancreatic amylase to starch. Even when enzymes and their substrates do come into contact, the presence of cell wall polysaccharides may slow the diffusion of hydrolytic products through the partially digested matrix in the gut lumen. Pulses tend to give particularly low glycemic index values, probably because legume seeds have relatively thick cell walls, which resist destruction during processing and cooking. This effect of fiber on carbohydrate metabolism cannot be predicted from simple analytical values for total fiber because it reflects the structure, as opposed to the absolute quantity, of cell wall polysaccharides within the food.

Many studies on postprandial glycemia have been conducted using isolated fiber supplements added to glucose test-meals or to low-fiber sources of starch. They demonstrate that, contrary to Trowell's original hypothesis, wheat bran and other insoluble cell wall materials have little effect on glucose metabolism. However certain polysaccharides, such as guar gum, pectin, and oat beta-glucan, which form viscous solutions in the stomach and small intestine, do slow the absorption of glucose. Highly viscous food components may delay gastric emptying and inhibit the dispersion of the digesta along the small intestine, but the primary mechanism of action appears to be suppression of convective stirring in the fluid layer adjacent to the mucosal surface. The rapid uptake of monosaccharides by the epithelial cells tends to reduce the concentration of glucose in this boundary layer, so that absorption from the gut lumen becomes rate-limited by the relatively slow process of diffusion. The overall effect is to delay the assimilation of glucose and hence suppress the glycemic response to glucose or starchy foods in both healthy volunteers and diabetics. A similar mechanism probably inhibits the reabsorption of cholesterol and bile salts in the distal ileum, and this may account for the ability of some viscous polysaccharides to reduce plasma cholesterol levels in humans. (*See* **Carbohydrates**: Metabolism of Sugars.)

Mineral Metabolism

Polysaccharides and phenolic components of cell walls often contain polar groups that can interact with ionized solutes in the gastrointestinal contents. Such molecular binding effects do occur *in vitro* and might, in principle, restrict the availability of nutrients and other substances for absorption in the small intestine. Iron, zinc, and calcium are absorbed relatively poorly from the human diet, and there is a long tradition that high intakes of dietary fiber have an adverse effect on human mineral nutrition. Studies with subjects who, for clinical reasons, have had the colon removed and the small bowel brought to the abdominal surface (ileostomists) suggest that although neutral polysaccharides do not bind minerals, charged polysaccharides such as pectin can displace cations into the colon. However, fermentation of cell walls probably releases iron and calcium into the lumen, whence they may be salvaged by colonic absorption, and there is little objective evidence that dietary fiber *per se* has much of an adverse effect on mineral metabolism. Indeed, highly fermentable oligo- and polysaccharides have been reported to promote mineral absorption in some studies, presumably by stimulating active transfer in the colon.

The effect of phytate (myoinositol hexaphosphate) is generally considered to be more significant than that of fiber. Phytate is often present in close association with cell wall polysaccharides in unprocessed legume seeds, oats, and other cereals. Phytate does exert a potent binding effect on minerals and has been shown to reduce the availability of magnesium, zinc, and calcium for absorption in trials with human volunteers. Phytate levels in foods can be reduced by the activity of endogenous phytase, by hydrolysis with exogenous enzymes, or by fermentation. Dephytinized products may be of benefit to individuals at risk of suboptimal mineral status. (*See* **Phytic Acid**: Nutritional Impact.)

Large Intestine

The large intestine, which is located distal to the small bowel, salvages energy from food residues and from endogenous proteins and carbohydrates that have escaped digestion and absorption. Its other main role is to form and store feces. Fecal microorganisms degrade undigested starch, and many of the polysaccharides that comprise dietary fiber, to yield the

volatile fatty acids butyrate, propionate, and acetate. Whereas the gut of ruminants is adapted to provide a watery, nutrient-rich environment for bacterial fermentation in the foregut, the colon is the main site of permanent bacterial colonization in the human alimentary tract. The proximal colon contains about 200 g of dilute fecal material, portions of which are transferred at regular intervals from the right colon into the transverse and distal segments for partial dehydration and storage. The pattern of motility in the large intestine is similar in principle to that of the rest of the alimentary tract, but the rate of transit is slower. In healthy individuals, stools are passed with a frequency varying from once or twice a day to once every 2–3 days.

The absorption and metabolism of short-chain fatty acids derived from carbohydrate fermentation provide an important route for the recovery of energy from undigested polysaccharides. Butyrate functions as a source of energy for the colonic mucosal cells, whereas propionate and acetate are absorbed and metabolized systemically. The other major breakdown products of carbohydrate fermentation are hydrogen, methane, and carbon dioxide, which together comprise flatus gas. Excess gas production is said to cause distention and pain in some individuals, especially those attempting to increase their fiber consumption, but this is probably caused more by fermentation of oligosaccharides such as stachyose and verbascose, found for example in legume seeds, rather than the cell wall polysaccharides themselves.

Fecal Bulk

The ability of dietary fiber to prevent cancer and various degenerative diseases of the alimentary tract was proposed by Denis Burkitt, who based his hypothesis largely on the concept of fecal bulk. His field observations in Africa, where cancer and other chronic bowel diseases were rare, suggested that populations consuming traditional rural diets rich in vegetables and cereal foods produced bulkier, more frequent stools than persons living in the industrialized West. Burkitt argued that consumption of highly processed cereals in industrialized societies led to chronic constipation, and that this caused prolonged high pressures both within the colonic lumen, and also within the lower abdomen as a result of straining to pass hard stools. This in turn was thought to increase the risk of various diseases of muscular degeneration including varicose veins, hemorrhoids, hiatus hernia, and colonic diverticuli. Furthermore, infrequent defecation was thought to cause prolonged exposure of the colonic epithelial cells to mutagenic chemicals that could initiate cancer. Although Burkitt's overall hypothesis for the beneficial effects of fecal bulk is undoubtedly an oversimplification, it has never been comprehensively refuted.

It is certainly true that the consumption of dietary fiber is a major determinant of both fecal bulk and bowel habit, but the magnitude of the effect depends upon the type of fiber consumed. Soluble cell wall polysaccharides such as pectin are readily fermented by the microflora, whereas lignified tissues such as wheat bran tend to remain at least partially intact in the feces. Both classes of dietary fiber can contribute to fecal bulk but by different mechanisms. The increment in stool mass caused by wheat bran depends to some extent on particle size, but in healthy Western populations, it has been shown that for every 1 g of wheat bran consumed per day, the output of stool is increased by between 3 and 5 g. Other sources of dietary fiber also favor water retention. For example isphagula, which is a mucilaginous material derived from *Psyllium*, is used pharmaceutically as a bulk laxative. Soluble polysaccharides such as guar and oat β-glucan are readily fermented by anaerobic bacteria, but solubility is no guarantee of fermentability, as is illustrated by modified cellulose gums such asmeythylcellulose, which is highly resistant to degradation in the human gut. Fermentation reduces the mass and water-holding capacity of soluble polysaccharides considerably, but the bacterial cells derived from them do make some contribution to total fecal output. Thus, although all forms of dietary fiber exert some laxative effect, the important differences in their properties and behavior make it hazardous to infer the biological effects of any diet from a single analytical measurement of total fiber content.

Fecal Chemistry

Apart from increasing the availability of energy from dietary fiber, the major effect of bacterial fermentation is to regulate the physical and chemical properties of the intraluminal environment. It is now generally accepted that most colorectal carcinomas develop progressively from precancerous lesions called adenomatous polyps. The gradual transition from a normal crypt via a precancerous lesion to a malignant tumor is associated with a progressive loss of differentiation, deregulation of cell growth, and an accumulation of mutations in genes associated with the control of cell proliferation and death. These somatic mutations are assumed to be caused primarily by mutagenic chemicals in the feces, much as mutagens in tobacco smoke induce lung cancer. Human fecal water contains a number of different mutagens that have been shown to be carcinogenic in animals. These include heterocyclic aromatic amines created during the cooking of meat at high temperatures, and

N-nitrosamines derived from dietary proteins. It has not been established conclusively that these substances cause colorectal carcinogenesis in humans, but there is strong circumstantial evidence that they do, and it would seem prudent to adopt nutritional strategies to reduce their concentration in the colon. The fecal bulking hypothesis provides the principal rationale for the current Dietary Reference Values for nonstarch polysaccharides in the UK, which recommend that adults should consume an average of 18 g of fiber per day. Epidemiological studies provide only weak evidence for a protective effect of dietary fiber against bowel cancer *within* the populations of industrialized Western countries such as the USA, perhaps because the overall range of fiber intakes is relatively low. However, there is stronger evidence for a protective effect of fiber intake and fecal weight against bowel cancer when different populations with a broader range of fiber intakes are compared.

The realization that colorectal carcinogenesis is a prolonged multistage process involving changes to a complex array of genes raises new questions about interactions between the colonic epithelial cells, dietary fiber, and other constituents of the fecal stream. Bile acids, which are produced by the liver and secreted into the gut lumen, where they facilitate the digestion and absorption of fat, are known to increase the rate at which colonic mucosal cells divide. It has been proposed that this effect causes bile salts to act as tumor promoters in the human colon, and thereby helps to accelerate the carcinogenic process initiated by fecal mutagens. High-fat diets tend to stimulate the release of bile acids and increase their concentration in the feces, and this may explain at least partially the fact that high fat consumption is a risk factor for bowel cancer in human populations. However, increased fecal bulk will tend to dilute the concentration of bile salts and reduce their residence time. The nonfermentable particulate components of plant cell walls may also provide a finely dispersed solid phase on to which bile acids can bind, so that their concentration in the aqueous phase of the feces is reduced still further. (*See* **Cancer**: Diet in Cancer Prevention.)

Cellular Physiology

Fermentation of carbohydrates lowers the pH of the feces and increases the concentration of short-chain fatty acids in contact with the mucosal cells. There is increasing evidence that these changes modify the intraluminal environment in such a way as to regulate the birth and death of the mucosal cells. Programmed cell death (apoptosis) is a mechanism whereby damaged cells are removed selectively from a tissue in an orderly manner that does not cause inflammation or general tissue desruction. This may well provide a defense against the survival of cells carrying tumorigenic mutations. Apart from the role of butyrate as an essential source of energy for the colonic mucosa, it also suppresses proliferation, increases differentiation, and induces apoptosis in many types of tumor cell grown *in vitro*. Tumor cell lines established from adenomatous polyps and carcinomas from human colon undergo increased apoptosis in the presence of butyrate, at concentrations close to those that occur *in vivo*. Moreover, cells from fully developed carcinomas have been shown to be less responsive to butyrate than those from early lesions, which perhaps implies that abnormal cells that survive and evolve into malignant tumors also acquire an adaptive resistance to butyrate. Unabsorbed starch and fermentable components of cereal fiber may provide a source of increased butyrate, which helps to suppress the proliferation of tumor cells and increase the likelihood of their deletion from the tissue by apoptosis. If so, much may depend on the rate of fermentation. Sugars and small oligosaccharides probably disappear too rapidly after entering the colon to provide a supply of butyrate to the distal colorectum, whereas lignified plant cell walls and some types of resistant starch that are slowly fermented may deliver butyrate efficiently to more distal regions of the colonic mucosa.

Adverse Physiological Effects

There are few well-authenticated adverse effects of dietary fiber. Some cases of pediatric malnutrition caused by very high fiber intakes have been reported, but this problem appears to be confined to grossly unbalanced diets. Similarly, the few cases of gastrointestinal obstruction in the literature appear to stem from very abnormal intakes of wheat bran or from swallowing nonhydrated supplements of viscous polysaccharides intended for consumption as a drink. There is some evidence that patients with a history of gastrointestinal surgery are at increased risk of obstruction due to abnormally high intakes of cereal bran. (*See* **Inflammatory Bowel Disease**.)

Sources of Fiber

It should now be obvious that although all plant foods provide some form of dietary fiber, not all components of fiber have the same physiological effects. In the UK, about 47% of total fibre intake is obtained from bread of various types and from breakfast cereals. The level of fiber in bread depends upon the extraction rate, which is the proportion of the original grain used to make flour. White flour has an extraction rate of about 70% and contains about 3%

nonstarch polysaccharides (NSP), whereas 'wholemeal' (100% extraction) flour contains about 10% NSP. Wheat bran, which contains about 40% NSP, can be added to the diet as a supplement, but it is relatively unpalatable. One of the lasting effects of the dietary fiber hypothesis has been to increase the variety and palatability of high extraction cereal products in the market place. Such products tend to contain primarily insoluble, poorly fermentable polysaccharides, and their consumption is a very effective way of increasing fiber intake to increase stool bulk. Oats, rye and barley contain higher quantities of soluble fiber than wheat. Oat bran is an important source of β-glucan, a soluble viscous polysaccharide that has been shown to slow down glucose absorption and reduce plasma low-density lipoprotein cholesterol levels in humans. A further 45% of total fiber intake in the UK comes from fruits and vegetables. Typically, the levels of NSP in fruits and vegetables are between 1 and 5% of fresh weight, and the polysaccharides are mostly soluble pectins and arabinoglactans, which are readily fermentable in the large bowel. For reasons described earlier, soluble fiber from vegetables and fruit does not have the bulk laxative effect of cereal bran, but it does provide fermentable carbohydrate for fermentation in the colon. Moreover, there is strong evidence that fruit and vegetables protect against cancer through other mechanisms, including the provision of biologically active phytochemicals. (*See* **Dietary Fiber**: Properties and Sources.)

See also: **Cancer**: Diet in Cancer Prevention; **Carbohydrates**: Metabolism of Sugars; **Dietary Fiber**: Properties and Sources; **Inflammatory Bowel Disease**; **Phytic Acid**: Nutritional Impact

Further Reading

Burkitt DP and Trowell HC (eds) (1975) *Refined Carbohydrate Foods: Some Implications of Dietary Fibre*. London: Academic Press.

Bell S, Goldman VM, Bistrian BR *et al*. (1999) Effect of beta-glucan from oats and yeast on serum lipids. *Critical Reviews in Food Science and Nutrition* 39: 189–202.

Johnson IT and Southgate DAT (1994) *Dietary Fibre and Related Substances*. London: Chapman & Hall.

Kushi LH, Meyer KA and Jacobs DR Jr (1999) Cereals, legumes and chronic disease risk reduction: Evidence from epidemiologic studies. *American Journal of Clinical Nutrition* 70: 451S–458S.

Southgate DAT (1992) *Determination of Food Carbohydrates*, 2nd edn. London: Elsevier Applied Science.

Williams GM, Williams CL and Weisburger JH (1999) Diet and cancer prevention: The fibre first diet. *Toxicological Science* 52: 72–86.

Effects of Fiber on Absorption

L Prosky, Prosky Associates, Rockville, MD, USA

Background

Although dietary fiber is not in itself a nutrient, in many countries, it is considered among the six groups of major nutrients along with proteins, carbohydrates, fats, vitamins, and minerals. Its required listing on the nutrient food label in most countries that have labeling procedures begs the question: what is the function of dietary fiber? Among these concerns is: how does the speeding up of the passage of food through the gastrointestinal tract and increase in the stool weight caused by the ingestion of diets high in dietary fiber affect the absorption of nutrients and other substances in the gastrointestinal tract?

Effects on Mineral Absorption

In vitro studies have shown that particular fibers bind minerals such as iron and zinc, suggesting interference of fibers on mineral absorption. Iron deficiency, which is one of the most prevalent deficiency disorders in the world, is present in Western cultures where there is no lack of food. There would also be concern for many of the other minerals for which there is only a marginal dietary intake. Some reports have implicated bran as a very strong iron absorption inhibitor, the action of which could be reversed with ascorbic acid or dietary meat (**Table 1**). High intake of phytate-rich, fiber-rich foods could inhibit zinc absorption, but again, this could be reversed by ingestion of animal protein. This would seem to indicate that there are many other factors involved in the absorption of minerals. In addition, several reports have revealed a tendency for reduced absorptions and retentions to occur for calcium, magnesium, phosphorus, and copper. It should be noted that no reliable method exists for measuring the bioavailability of an ingested mineral that separates this exogenous intake and its associated biological response from the amount of endogenous mineral in the body being reabsorbed and reutilized. Further, it has remained a challenge to detect a biological response sensitive enough to measure a change in the body's storage or use of a mineral that reflects the true amount of a specific dietary mineral absorbed or utilized. These problems are further compounded when the interactions among minerals are considered. The drawback to many of the studies was the relatively short period of observation and the often excessively high

Table 1 Effect of various dietary fibers on the absorption of five essential minerals

Dietary fiber	*Iron*	*Calcium*	*Zinc*	*Phosphorus*	*Magnesium*
Bran	↓↓				
Ispagula	↓				
Psyllium	↓		ne		
Pectin	ne	ne			ne
Methoxylate pectin	ne				
Guar gum	ne				
Cellulose	ne	↓	ne		↓, ne
Beet pulp		ne			
Brown bread		↓*		↓[a], ne	↓[a], ne
Unpolished rice		↓[a]			↓[a]
Unleavened wholemeal bread		↓		ne	
Leavened wholemeal bread		↓		↓	↓
Wheat fiber		↓			
Cellulose in apple compote		↓		ne	↓
Hemicellulose					↓
Fruits, vegetables		↓, ne		ne	↓, ne
Partially hydrolyzed guar gum (Sunfiber®)	↑	↑			↑
Pectin	ne	ne	ne	ne	ne
Cellulose	ne	ne	ne	ne	ne
Inulin, oligofructose[b]	↑	↑	↑	[c]	↑
Galactooligosaccharides[b]		↑			↑
Resistant maltodextrin (Fibersol-2®)	ne	↑			ne
Polydextrose[b]		ne			

↓, decreased absorption; ↓[a], decreased absorption improved with time; ne, no effect; ↑, increased absorption; [b], at present time is not measured as fiber; [c], prevents phosphorus loss from bone not observed

dosage of the particular experimental component. The intakes of dietary fiber foods often bear little relationship to human dietary situations (as much as 60 g per day of bran supplements in some studies). It also has been demonstrated that there is an adaptation to decreased mineral intake. Ingesting 50–60 g of dietary fiber derived from whole maize meal, brown bread, and beans with a mean daily intake of 275 mg of calcium resulted in calcium absorption varying from 61 to 83%. This is far higher than the usual 20–30% absorption that occurs with Western subjects ingesting much lower dietary fiber diets. The results showed that the subjects were in balance with regard to calcium, magnesium, and iron, and that the drawbacks of habitually high dietary fiber intakes and low calcium and other mineral salt intakes of Third World people may be of less significance to health than might be expected. While the nutritional policies of Western societies should not do anything to restrict, at this time, the intake of dietary fiber-containing food, it is apparent that as the world population increases, there will be a greater reliance on plant foods. This is because it is far more economical to use land to produce plant products than to produce meat products. More work will obviously have to be done to establish the mineral absorption effects and the factors that affect them (i.e., phytate, etc.), but for now, the benefits of ingesting dietary fiber far outweigh any negative effects of decreased mineral absorption. Also included in Table 1 are the mineral absorption effects of four food components; the methods for these food components were approved recently by the Association of Official Analytical Chemists (AOAC). These food ingredients are not measured as dietary fiber by the standard AOAC methods for dietary fiber but are carbohydrate polymers of DP-3 and higher and are resistant to hydrolysis and absorption in the human small intestine. Updated definitions of dietary fiber include these components (*See* **Minerals – Dietary Importance**.)

Effects on Vitamin Absorption

There is a paucity of information on the availability of vitamins in relation to the dietary fiber content of the diet. Further, what is available is mainly from animal studies, and the extrapolation of its effects to humans may be difficult.

In previous studies, methyl cellulose or pectin did not affect the utilization of vitamin A in rats, but a combination of the two did appear to decrease liver vitamin A stores. Lignin given with a test meal did not affect serum vitamin A of human subjects, but carotene, which is bound to plant cell walls, is less available than pure or added carotene. (*See* **Carotenoids**: Physiology; **Retinol**: Physiology.)

Dietary methyl cellulose had no effect on the growth of thiamine-deficient or normal rats. Riboflavin and niacin may be less available from wholemeal bread than from enriched white bread. Several fiber sources (course bran, fine bran, cellulose, and cabbage) improved absorption of a load dose of riboflavin in humans. Nicotinic acid, which is bound to fiber in cereal brans and is not available unless released by alkaline hydrolysis or treatment with limewater, was rendered effective in promoting weight gain in animals, made deficient by feeding maize with bound vitamin. The binding was probably due to cellulose or hemicelluloses. (*See* **Niacin**: Physiology.)

A lower bioavailability of vitamin B_6 from cereal than from nonfat dry milk was reported for rats. Vitamin B_6 was available to a lesser degree from soybeans than from beef, and the addition of wheat bran to a diet increased vitamin B_6 availability in human subjects. Fibers digestible by intestinal bacteria (xylan, guar gum, and pectin) intensified the symptoms of B_{12} depletion in rats as compared with the nondigestible fibers (cellulose, lignin, alginate, and wheat bran). (*See* **Vitamin B_6**: Physiology.)

In two additional studies, fiber (in the form of white bread, wholemeal bread, or cellulose) was found to have no effect on urinary folate excretion, but in another study, fiber (maize, rice, and bread) decreased the elevation in serum folate as compared with the vitamin alone. (*See* **Folic Acid**: Physiology.)

From these inadequate data, it appears that in general, the availability of unbound vitamins is not affected by the presence of dietary fiber, but the situation with the bound vitamins will have to be explored further.

Effects on Lipid Absorption

Dietary fiber can influence lipidemia and atherosclerosis by affecting the absorption and metabolism of the lipid components in the blood. Soluble fibers, such as pectin, do affect the lipid uptake and glycemia. Oat bran, which contains β-glucans (soluble fibers), will lower the level of serum cholesterol. It is now established that, whereas insoluble dietary fibers (wheat bran, cellulose) do not influence serum lipids, soluble dietary fibers (guar, pectin, psyllium) exert a hypocholesterolemic effect (Table 2). The mechanism by which this hypocholesterolemic effect is exerted involves increased viscosity of the stomach and small intestine contents, and may also influence bile acid metabolism.

Cereal fibers and fractions (oat, rice, and barley) contribute significantly in reducing total cholesterol in a variety of animal species and in hypercholesterolemic subjects. Long-term population studies (Framingham, LRC, and Helsinki) have demonstrated that for each 1% reduction in total cholesterol, there is a two- to fourfold decrease in atherosclerosis in otherwise healthy individuals. A 6-year prospective study of male health professionals diagnosed free of cardiovascular disease showed that cereal fiber was found to be most strongly associated with a reduced risk of myocardial infarction. Adding oat and barley, in addition to other soluble dietary fiber foods, proved efficacious in lowering cholesterol even further. The data generated by some 60 feeding studies with psyllium convinced the US Food and Drug administration to approve a health claim for psyllium, which allows foods and supplements with at least 1.7 g of soluble dietary fiber from psyllium to claim 'soluble fiber from psyllium per serving, as part of a diet low in saturated fat and cholesterol, may reduce the risk of heart disease.' Although the effects of cereal fiber are less pronounced in normo-cholesterolemic individuals, these changes would have a significant impact on reducing cardiovascular risk for the whole population. Table 3 shows the effects of feeding various dietary fibers on serum and liver cholesterol and serum triglycerides in a rat model. Of particular interest is cellulose, which has no effect on cholesterol and triglycerides, and wheat bran, which has little effect on serum and liver cholesterol but plays a role in reducing serum triglycerides.

To date, the FDA has approved health claims for three substances, oats, psyllium and soy, with regard to their cholesterol-lowering ability in combination with a diet that is low in saturated fat and cholesterol. This reinforces the abundant data available showing the cholesterol-lowering properties of specific dietary fibers. A study was done to compare a soluble fiber, soy-enhanced diet with a diet containing low-fat dairy foods but low in soluble dietary fiber on the level of serum lipids. The test diet reduced the total cholesterol:high-density lipoprotein (HDL)-cholesterol ratio and reduced the risk of coronary heart

Table 2 Influence of dietary fibers on human serum total and low-density lipoprotein (LDL)-cholesterol (percentage of control)

Dietary fiber	*Total cholesterol*	*LDL cholesterol*
Wheat bran	102	97
Corn bran	97	98
Cellulose	104	108[a]
Soy hulls	107	104
Oat bran	87[a]	86[a]
Guar gum	87[a]	84[a]
Locust bean gum	86[a]	86[a]
Gum karaya	90[a]	90[a]
Legumes	87[a]	95

[a]Significant effects.

Table 3 Effect of feeding various dietary fibers on serum and liver cholesterol and serum triglycerides in rat models

Dietary fiber	*Number of studies*	*Serum cholesterol (percentage change)*	*Liver cholesterol (percentage change)*	*Serum triglycerides (percentage change)*
Psyllium	7	−32	−52	−9.6
Oat gum	11	−22	−45	−9.6
Guar gum	3	−23	−43	na
Pectin	20	−17	−35	+3.2
Oat bran	8	−12	−28	−13.4
Soy fiber	6	−11	−17	−10.2
Corn bran	2	−6	−11	−43.9
Cellulose	26	0	0	0
Wheat bran	5	−1	+6	−38

na, not available.

disease (CHD). (*See* **Cholesterol**: Factors Determining Blood Cholesterol Levels.)

In a study of the relationship of dietary fiber and serum lipids among true vegans, lacto-ovo-vegetarians, nonvegetarians, and the general public, only the true vegans showed a significant lowering of cholesterol. The only difference in the dietary fiber intakes among the four groups was a higher intake of pectin in the true vegans.

The digestion of dietary fat begins in the stomach, where it is coarsely emulsified and partially hydrolyzed. No further emulsification occurs in the duodenum, but subsequently, under the action of pancreatic lipase, phospholipids and cholesterol esters are cleaved enzymatically. The activity of pancreatic lipase *in vitro* can be reduced by several of the soluble dietary fibers, including pectins, oat bran, psyllium, and sugar beet fiber, with part of the loss being due to the binding of the fiber to the enzymes. The inhibitory effects of insoluble dietary fibers, wheat bran, and wheat germ can also produce this reduction of pancreatic lipase. Because the mechanism by which soluble dietary fibers alter the breakdown of fat is not well understood, some investigations of the physicochemical aspect of the process were studied. It is well known that emulsification is a key step in fat digestion, because gastric and pancreatic lipases act at the fat globule–water interface, and that the size of the fat globule is a parameter governing the activity of gastric and pancreatic lipases. Could dietary fiber intefere with the normal process of fat emulsification? The two parameters studied were viscosity and electric charge. Five sources of fiber with different properties were studied, namely gum arabic (not viscous), two differently charged pectins, and three uncharged guar gums (of varying viscosity). Under gastric conditions (pH 5.4), the degree of emulsification was not affected by any of the fibers, but the viscous fibers did increase the size of the emulsified droplets, and the effect was more pronounced as the concentration of these fibers increased. Only viscous fibers significantly increased droplet size and reduced droplet surface area. Overall, the droplet size was positively correlated, and the droplet surface area was negatively correlated with the concentration of the medium viscosity guar gum in the range 0–20 mPa.s. The high-viscosity guar gum significantly reduced (by 32%) the triglyceride lipolysis of such emulsions by human gastric lipase, compared with control and low- and medium-viscosity fibers (Table 4).

In duodenal conditions (pH 7.5 and addition of bile), the quantity of emulsified lipids was reduced, and raising the concentration of viscous fibers increased the size of the droplets. The extent of the emulsification, the droplet size, and, consequently, the droplet surface area were strongly correlated with the concentration of the medium-viscosity guar gum in the range 0–4 mPa.s. (and also 0–20 mPa.s.). The high- and medium-viscosity guar gums and one type of pectin significantly reduced the extent of triglyceride lipolysis catalyzed by pancreatic lipase only. Overall, the extent of triglyceride lipolysis was negatively correlated with the viscosity of the duodenal medium. These findings indicate a mechanism by which soluble, highly viscous fibers can alter lipid assimilation by lowering the emulsification of dietary lipids in the stomach and duodenum.

This hypothesis was tested *in vivo* in rats. A coarsely emulsified lipid mixture with and without 0.3% high-viscosity guar gum was intragastrically intubated. After 30 min of digestion in the stomach, the median droplet diameter of the emulsion in the presence of guar gum was about two fold larger. The specific surface area displayed by the emulsion was about half that of the controls. In the presence of the guar gum, the extent of triglyceride lipolysis was about threefold lower in the stomach content and about twofold lower in the duodenum content. Thus, dietary fibers that raise viscosity in the stomach and duodenum contents reduce the extent of emulsification and subsequent triglyceride lipolysis.

Table 4 Influence of dietary fiber charge and viscosity on droplet size, droplet area, and triglyceride lipolysis under gastric and duodenal conditions

Dietary fiber	*Droplet size*	*Droplet surface area*	*Triglyceride lipolysis*
Gastric conditions (pH 5.4)			
Gum arabic (not viscous)	−	−	−
Pectin (low viscosity, charged)	+	↓	−
Pectin (low viscosity, different charge)	+	↓	−
Guar gum (low viscosity, uncharged)	+	↓↓	−
Guar gum (medium viscosity, uncharged)	++	↓↓↓	−
Guar gum (high viscosity, uncharged)	+++	↓↓↓↓	↓
Duodenal condition (pH 7.5 and addition of bile)			
Gum arabic (not viscous)	−	−	−
Pectin (low viscosity, charged)	+	↓	−
Pectin (low viscosity, different charge)	+	↓	−
Guar gum (low viscosity, uncharged)	+	↓↓	−
Guar gum (medium viscosity, uncharged)	++	↓↓	↓
Guar gum (high viscosity, uncharged)	+++	↓↓↓	↓

−, no effect; +, increase effect; ↓, decreases surface area; ↓, decreases lipolysis.

The lipid products of the digestion process (i.e.,mainly monoglycerides, free fatty acids, lysophospholipids, and free cholesterol) are absorbed by the small intestine, where they are rapidly used as substrates for the *de-novo* synthesis of triglycerides, phospholipids, and cholesterol esters. They are then packaged and secreted as chylomicrons into the lymph and finally into the blood.

Other mechanisms can contribute to the absorption of lipids. The binding or entrapment of lipolytic products as well as bile salts in the presence of viscous fibers such as oat bran, guar gum, and pectin can counteract the building of vesicles and mixed micelles in the aqueous phase of the small intestine contents. The limiting step in lipid absorption is the transport across the unstirred water layer associated with the enterocyte brush-border membrane. Highly viscous fibers have been observed to increase the thickness of the unstirred layer, reducing the rate of diffusion and absorption of cholesterol and free fatty acids. These effects have not been observed in the presence of slightly viscous soluble fibers (hydrolyzed guar gum, or chitosan) or insoluble fibers such as cellulose or some hemicelluloses. Changes induced in the intestinal lumen by feeding viscous fibers have been shown to result in reduced recovery of fatty acids and cholesterol in the lymph.

Augmented ileal excretions of lipids have been repeatedly demonstrated in ileostimized human subjects after ingesting meals containing citrus pectin, alginate, oat bran, or β-glucan, and barley fiber. Some investigators have also reported increased fecal or ileal lipid excretions after their subjects' diets were enriched with wheat bran or wheat germ. The ileal or fecal excretion of cholesterol and/or bile acids is markedly increased after addition of various kinds of dietary fibers are added to the test diets. Soluble and viscous fibers such as β-glucans, pectin, or psyllium are particularly efficient in increasing cholesterol excretion,

A principal feature of the postprandial state is a striking increase in triglyceridemia, owing to the secretion of chylomicrons (which transport absorbed dietary lipids and cholesterol) by the small intestine into the circulatory system. Studies with a variety of diets, either supplemented or not supplemented, has shown that different kinds of dietary fibers (wheat fibers, legume fibers, guar gums, pectins, oat bran, β-glucans, psyllium, and mixed fibers, especially those rich in soluble fibers) can reduce the postprandial rise in triglyceridemia. This results from reduced bioavailability of fatty acids and cholesterol induced by the viscous dietary fibers, making the small intestine mucosa less efficient in lipid synthesis and in secreting chylomicrons into the lymph and finally into the circulatory system.

Partially hydrolyzed guar gum lowers total cholesterol, triglycerides and LDL with no change in the HDL in human subjects. Hemicelluloses from soft winter wheat, hard red spring wheat, corn bran, soy hulls, apples, and carrots lowers triglyceride levels, but only the hemicellulose from soft winter wheat had no effect on serum cholesterol. This indicates that the effects may be dependent of chemical interactions with other nutrients in the foods.

Several of the carbohydrate dietary components that are not measured as dietary fiber in any of the accepted methods, but which are not digested or absorbed in the small intestine, appear to have the physiological effects that we have come to identify as

Table 5 Effects of food constituents on lipid parameters

Product	*TC*	*TR*	*LDL*	*HDL*	*BF*
Inulin and oligofructose	↓[a]	↓[a]			
Polydextrose	↓[a]	↓[a]			
Galactooligosaccharides	↓				
Resistant maltodextrin (Fibersol-2®)	↓	↓	–	ne	↓

TC, total cholesterol; TR, triglycerides; LDL, low-density lipoprotein; HDL, high-density lipoprotein; ↓[a], lowers in hypercholestorolemics; BF, body fat; ne, no effect.

fiber effects and have received much attention recently. They have the following properties with regard to lipid absorption (Table 5).

Effects on Carbohydrate Absorption

The rate of delivery of glucose by the gut to the portal vein and the peripheral tissues is influenced by many factors other than the amount of α-linked glucose polymers in foods. The amylose:amylopectin ratio in starchy foods influenced postprandial glycemia, because amylopectin had a much faster rate of hydrolysis by pancreatic α-amylase. Food that was not chewed resulted in a much flatter glycemic response than food that was chewed well before swallowing. Similar responses could be found when food was altered through milling of cereals, indicating the importance of structure and particle size affecting the α-amylase access to the starch substrate and the effects of surface area. Food constituents such as dietary fiber, which form viscous solutions, may slow the rate of starch digestion and hence glucose absorption. The mechanism of the effects of a viscous solution might happen because of (1) delayed gastric emptying; (2) reduced hydrolytic activity; (3) poor mixing of the intestinal contents with the secreted enzymes; and/or (4) increased thickness of the unstirred water later, slowing down the diffusion of glucose.

The concept of a glycemic index (GI) was introduced in 1981 as a measure of the rise in postprandial glucose of various food carbohydrates. Low-GI foods were found to be slowly or incompletely digested in the small intestine, or found to contain fructose, which is only partially converted to glucose. On the basis of this determination, foods such as unripe bananas (rich in amylose) and over-ripe bananas (rich in fructose) had a lower GI than fresh white bread, which is the standard against which the GI of other foods is measured.

Starch in freshly cooked potatoes is readily hydrolyzed by α-amylase and results upon feeding in a high glycemic and insulinemic response. Repeated cooking and cooling of these potatoes increases the amount of starch resistant to the α-amylase of the small intestine. The result of feeding freshly cooked potatoes and potatoes that had been cooled and cooked several times was a higher postprandial peak for plasma glucose and serum insulin in the former group. (*See* **Starch**: Functional Properties.)

In subsequent years, interest has been shifted to the effects of viscous fibers on glucose absorption by the small intestine and insulin release by the pancreas. It has been shown that insoluble fiber sources such as wheat bran had little effect on glucose tolerance and postprandial insulin response. The effects of viscous fibers were striking and depended on the viscosity of the fiber. Oat β-glucan clearly showed this effect, which was similar to sipping glucose over 3 h, rather than ingesting it as a bolus in 5 min. The nibbling versus gorging effect has been shown to be of benefit in diabetes.

The chemically resistant starches have no effect on the glycemic response of the available starch. These starches are chemically resistant and require solubilization with potassium hydroxide prior to analysis. Entrapped starch, which may be physically trapped in, or by, the fiber reacts as a slow-release carbohydrate. This starch simply needs to be released from its bound physical matrix and made available by good mastication or fine grinding of the food ingredient.

Fructooligosaccharides have no major effects on the glycemic response of the accompanying carbohydrate and have no effect on the glycemia *per se*.

Dried legumes peas, beans and lentils, pumpernikel rye bread, and bulgur or cracked wheat have been shown to reduce the rate at which these foods are digested, and therefore improve the glucose tolerance. The structure of the food may also influence the rate at which the foods are digested.

Paradoxically, insoluble dietary cereal fiber have been shown to offer protection from diabetes in several large cohort studies. Low-GI diets or diets with a low glycemic load (dietary GI × dietary carbohydrates) were also found to be related to the development of Type 2 diabetes over a 6-year period. It is not easy to explain why cereal fiber has an effect in reducing Type 2 diabetes, unless it is through associated nutrients such as magnesium or antioxidants. While the effects of wheat bran on glycemia are not striking, it is possible that the phenolics in the bran may be useful antioxidants of importance in the prevention of diabetes.

The early promise shown by viscous soluble dietary fibers in the treatment of diabetes has suffered because of the unavailability of palatable soluble dietary fiber diets. Studies using the insoluble dietary fibers from cereal have not been followed up successfully in more recent studies.

Table 6 Effects of food constituents on postprandial carbohydrate parameters

Product	*Blood glucose*	*Insulin*	*Glucagon*
Partially hydrolyzed guar gum (Sunfiber®)	↓	↓	
Inulin and oligofructose	ne	ne	ne
Polydextrose	↓	ne	
Galactooligosaccharides	–	–	–
Resistant maltodextrin (Fibersol-2®)	↓	↓	↓

ne, no effect; –, no experimental evidence.

Table 6 lists the effects of five food constituents on postprandial carbohydrate parameters. Only the partially hydrolyzed guar gum and resistant maltodextrin measure as dietary fiber by the AOAC method.

Effect on Protein Absorption

The effects of dietary fiber on protein absorption have mostly been ignored. Studies have shown that the increased intake of dietary fiber results in increased nitrogen loss. The form in which the nitrogen occurs in the feces is of considerable importance, if one is trying to deduce the effects of a dietary component on digestion.

Investigations in which the fecal nitrogen was fractionated imply that most of the nitrogen is in the bacterial mass or metabolic products, such as bacterial products, unabsorbed intestinal secretions, or mucosal cell debris. The proportion derived from unabsorbed dietary components represents a loss to the body, which is probably an effect of absorption of protein. There is also evidence that some bacterial nitrogen is derived from urea secreted in the large bowel and some fixation of nitrogen occurs in the large bowel. These uncertainties emphasize the limited value of fecal data for deducing what are essentially intestinal events. The probable solution to this problem would be to carry out ileostomy studies. To date, ileostomy studies centered on protein absorption have not been carried out.

See also: **Carbohydrates**: Digestion, Absorption, and Metabolism; **Dietary Fiber**: Physiological Effects; **Fats**: Digestion, Absorption, and Transport; **Minerals – Dietary Importance**; **Protein**: Digestion and Absorption of Protein and Nitrogen Balance; **Vitamins**: Overview

Further Reading

Cho SS, Prosky L and Dreher M (eds) (1999) *Complex Carbohydrates in Foods*. New York: Marcel Dekker.

Kritchevsky D and Bonfield C (eds) (1995) *Dietary Fiber in Health and Disease*. St. Paul, MN: Eagan Press.

Kritchevsky D and Bonfield C (eds) (1997) *Dietary Fiber in Health and Disease*. New York: Plenum Press.

Lairon D (2001) Dietary fibres and dietary lipids. In: McCleary BV and Prosky L (eds) *Advanced Dietary Fibre Technology*. London: Blackwell Science.

McCleary BV and Prosky L (eds) (2001) *Advanced Dietary Fibre Technology*. London: Blackwell Science.

Marabou Food and Fibre (1976) Symposium held at Marabou, Sundbyberg, Sweden Supplement Nr 14 till Naringforskning, argang 20, 1976. Stockholm: Caslon Press.

Trowell H, Burkitt D and Heaton K (eds) (1985) *Dietary Fibre, Fibre-depleted Foods and Disease*. London: Academic Press.

Vahouny GV and Kritchevsky D (eds) (1982) *Dietary Fiber in Health and Disease*. New York: Plenum Press.

Vahouny GV and Kritchevsky D (eds) (1986) *Dietary Fiber Basic and Clinical Aspects*. New York: Plenum Press.

Bran

A Zitterman, University of Bonn, Bonn, Germany

Definition and Characteristics of Bran

Bran is the outer coating or shell on grain that is removed while processing white flour. Bran consists of the pericarp, the seed coat, and the aleurone layer. The pericarp itself consists of three layers of differentiated types of cells (**Figure 1**). Wheat, oat, rice, and rye are common sources of bran. The bran content of wheat is approximately 15% of the whole grain. In rice, where the beard is coalesced with the corn, the bran content is only 8%. Typical commercial oat bran products contain large amounts of adhering starchy endosperm, because the endosperm tissues in the oat grain do not separate cleanly and efficiently from the outer layers. Bran has very little taste.

Chemical Composition of Bran

Bran is not a chemically defined substance. The composition of bran depends upon many different factors: species and variety, e.g., grain, kernel seize, shape, maturity, size of germ, and thickness of the outermost layer, length of time and condition of grain storage, system of grain conditioning before milling, the milling system itself, and, above all, the type of flour produced. The lower the extraction rate of the flour, the more endosperm passes to the bran.

Bran is an excellent source of dietary fiber (**Table 1**). However, it is important to recognize that bran is not dietary fiber as such. Bran is also a good source of

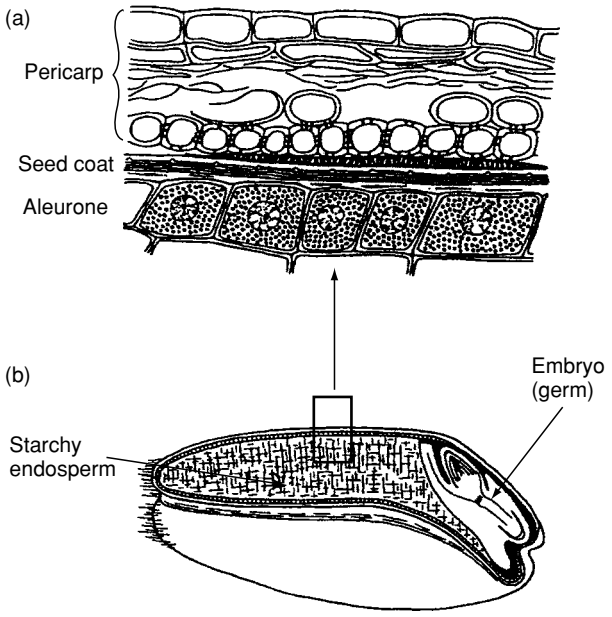

Figure 1 (a) Longitudinal section of the bran layers of a wheat grain showing the different cell types. (b) Longitudinal section through a wheat grain: the rectangle indicates the location of the section shown in (a). From Ferguson LR and Harris PF (1999) Protection against cancer by wheat bran: role of dietary fibre and phytochemicals. *European Journal of Cancer Prevention* 8: 17–25, with permission.

Table 1 Chemical composition of wheat and rice bran

Grams per 100 g dry weight	*Wheat bran*	*Rice bran*
Protein	13–18	12–20
Fat	3–6	3–22
Ash	6–7	9–13
Total dietary fiber	35–58	24–29
Soluble dietary fiber	2–5	2–4
Insoluble dietary fiber	32–53	20–24.5
Cellulose	6–12	6–12.8
Hemicellulose	19–31	8.7–17
Lignin	2–8	3–4

certain nutrients. It is rich in protein, which is primarily located in the aleurone layer. Rice bran is a particularly good source of fat (Table 1), mainly oleic acid (40%), linoleic acid (34%), and palmitic acid (17%). Moreover, approximately 70% of the mineral content and between a third (thiamine) and two-thirds (pyridoxine) of the B vitamins of cereal grains are located in the bran fraction. Bran also contains nonessential ingredients like phytic acid and ligans, substances for which the effects on human physiologic functions have not yet been clearly established. (*See* **Fatty Acids**: Properties; **Vitamins**: Overview.)

Dietary Fiber

The high amount of insoluble dietary fiber in cereal brans (Table 1) comes primarily from plant cell walls. The cell walls consist of heterogeneous mixtures of cellulose, noncellulose polysaccharides, and lignin cross-linked to varying degrees to form a complex matrix. Lignocellulose materials and insoluble arabinoxylanes (hemicellulose) constitute most of the mature cell wall of the plant materials such as cereal brans. In addition, cell walls of the seed coat contain cutin, a three-dimensional polyester, which has similar properties to lignin. Moreover, ferulic acid, the predominant phenolic acid in wheat bran, is bound to the cell wall. (*See* **Phenolic Compounds**.)

Cereal brans can also contain varying amounts of *soluble* arabinoxylanes and β-glucans, but no pectins. β-Glucan is a soluble polysaccharide that is formed exclusively by glucose residues. In contrast to the insoluble cellulose, β-glucans are formed not only by β-1,4 but also by β-1,3 glycosidic bonds. The ratio of β-1,3 and β-1,4 glycosidic bonds is approximately 3:7. Normally, blocks of two to 10 β-1,4-glycosidic glucose residues are interrupted by β-1,3-glycosidic bonds. High amounts of β-glucans are found in the aleurone layer and in the adhering endosperm of some cereal brans. Similar to pectin, the physiologic effects are related to the viscosity of β-glucans. (*See* **Carbohydrates**: Classification and Properties.)

Like wheat bran (Table 1), barley bran contains up to 55% of dietary fibers, of which normally 90% are insoluble. Barley bran contains approximately 5% of β-glucans. Special barley subspecies like 'Prowashonupana' contain up to 15% β-glucans. The dietary fiber content of rye bran is 28–36% and is thus lower than that in wheat bran. However, the amount of soluble arabinoxylanes in rye bran (15–25%) is relatively high compared with wheat bran (1–1.5%). In oat bran, the total amount of dietary fiber is 15–18% with about 50–60% insoluble and 40–50% of soluble dietary fibers. Oat bran normally contains 5–8% β-glucans.

The rice bran produced under normal milling conditions is unpalatable. When brown rice is milled to white rice, the oil in the rice bran comes into contact with a potent lipase also present in the bran, resulting in rapid degradation of the oil to glycerol and free fatty acids. However, if the bran is subject to a short-term high-temperature heat treatment immediately after milling, the lipase activity is destroyed, and a stabilized bran is produced. In contrast to other cereals, the rice grain has the ability to enrich silicium. Therefore, the industry standard for rice bran is a maximum of 0.1% silicium dioxide.

Phytic Acid

Phytic acid, the hexaphosphate of myo-inositol, is an abundant component of bran, where it occurs as a magnesium–calcium salt. The phytic acid content of

different types of bran can vary widely. Rice bran contains approximately 10% phytic acid, whereas wheat bran and rye bran contain only 5%. The phytic acid content of different cereals depends on the botanic subfamily to which the grain belongs. Wheat and rye are members of the Hordeae subfamily, whereas rice is a member of the Oryzeae. The phytic acid content of oat bran (subfamily: Aveneae) is approximately 6%.

Phytic acid can strongly chelate multivalent metal ions, especially the positively charged ions zinc, calcium, and iron. The chelates are named phytates. Such chemical bonds with multivalent cations can be formed intra- and intermolecularly by one or several phosphate groups of the pyhtic acid molecule (Figure 2). These complexes are insoluble.

The distribution of phytic acid in different morphologic layers of seeds can vary significantly. In cereal grains from the Hordeae subfamily, phytic acid is normally concentrated in the aleurone cells and, to a lesser degree, in the germ. The aleurone layer of wheat grain, for example, contains 85% of the phytic acid, the embryo 13%, and the endosperm 2%. In contrast, 80% of the phytic acid content of rice and maize is concentrated in the pericarp and embryo, respectively.

Inositol-6-phosphate and inositol-5-phosphate show the strongest inhibiting effect on mineral bioavailability, whereas the effect of inositol-4-phosphate and inositol-3-phosphate, two products of inositol-6-phosphate hydrolysis, is markedly lower. Phytic acid is hydrolyzed by the enzyme phytase. This enzyme was first described in rice bran. In wheat bran, the highest phytase activity is present in the aleurone layer, where more than 40% of the phytase activity is located. Approximately 34% is present in the endosperm, 15% in the scutellum, and 3% in the germ. In rye bran, the phytase activity is 10–20 times higher than in the rye flour. The total amount of cereal phytase activity varies depending on the food origin: phytase activity is high in wheat and rye grains, medium in barley, and low to very low in oat and maize grain. (*See* **Phytic Acid**: Properties and Determination; Nutritional Impact.)

Lignans

Lignans are phenolic dimers possessing a 2,3-dibenzylbutane structure. Such compounds are known to exist as minor constituents of many plants, where they form the building blocks for the formation of lignin in the plant cell wall. The compounds occur mainly in the glycosidic form. In rye, lignans are predominantly present in the bran fraction (Table 2). The glycosides are converted by fermentation in the proximal colon to mammalian lignans. The two major mammalian lignans, enterodiol and enterolactone, are the products of colonic bacterial metabolism of the plant lignans secoisolariciresinol and matairesinol.

Table 2 Lignan content of whole grains and bran[a]

	Secoisolariciresinol (μg per 100 g dry weight)	*Matairesinol (μg per 100 g dry weight)*
Wheat bran (whole grain)	33	3
Wheat bran	110	0
Rye meal	47	65
Rye bran	132	167
Oat meal	13	0
Oat bran	24	155
Barley (whole grain)	58	0
Barley bran	63	0

[a]Adapted from Mazur and Adlercreutz. *Pure and Applied Chemistry* 70: 1759–1776.

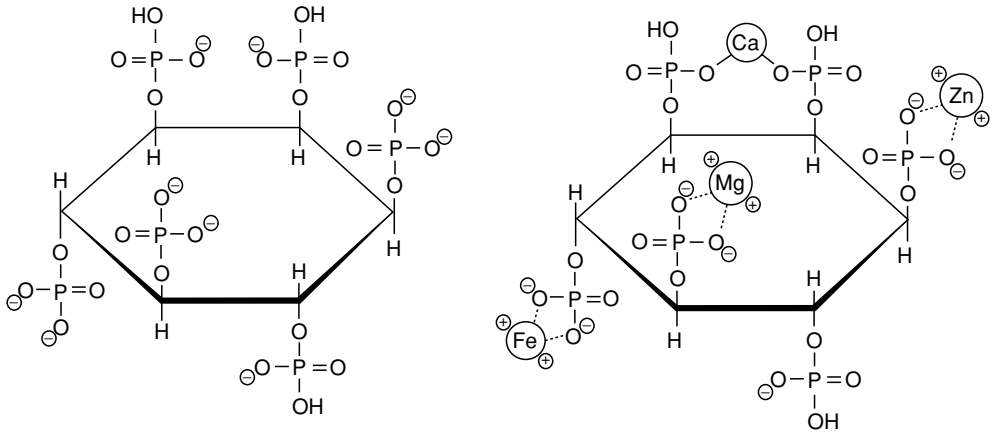

Figure 2 Structure of phytic acid and a phytate chelate at neutral pH. From Ingelmann H-J, Rimbach G and Pallauf J (1993) Phytinsäure – ein antinutritiver Faktor? *Ernhrungs–Umschau* 40: 400–404, with permission.

Physiologic Effects of Bran

Stool weight

Stool weight is significantly increased by adding sources of insoluble fiber to the diet. The physiological effect of dietary fiber is largely determined by the size of the particles. One gram of fine wheat bran increases the stool weight by 3.1 g, and 1 g of coarse bran increases the stool weight by approximately 6.2 g. Barley bran is more effective in increasing stool weight than fine wheat bran but less effective than coarse wheat bran. Comparable amounts of fiber provided by wheat and oat bran have the same effect on daily stool output, even though > 90% of wheat bran fiber and only 50–60% of oat bran fiber are insoluble. Wheat bran increases the fecal concentration of sugars and mass of plant material more than oat bran, whereas oat bran increases the fecal bacterial mass more. Oat bran increases the stool weight by providing rapidly fermented soluble fiber in the proximal colon for bacterial growth, which is sustained until excretion by fermentation of the insoluble fiber. Bacteria and lipids are major contributors to the increase in stool weight with oat bran consumption, whereas ingested plant fiber is responsible for much of the increase in stool weight with wheat-bran consumption.

Constipated subjects should start increasing dietary fiber by 2–4 g, e.g., with 1–2 tablespoons of bran. Plenty of liquid should be taken with bran, as it may cause irritations in the digestive tract. Intake can be increased to three times a day with a meal. Bran should be mixed with food in order to absorb water and increase stool volume. Coarse bran holds water better than fine bran. Depending on the severity of constipation, improvement may take a few days to several months.

Sterol Metabolism

Bile acids Soluble dietary fiber can adsorb bile acids. Accordingly, oat bran and rye bran increase fecal excretion of total bile acids, whereas wheat bran has no or only modest effects. Moreover, oat bran and rye bran influence the metabolism of bile acids. Rye bran increases the daily excretion and percentage of conjugated primary bile acids (cholic acid, chenodesoxycholic acid) and lowers the percentage of free secondary bile acids (desoxycholic acid, lithocholic acid). Rye bran thus decreases the ratio of secondary bile acids to primary bile acids. Rye bran, but not wheat bran or barley bran, reduces the ratio of lithocholic acid to desoxycholic acid. Oat bran enlarges the cholic acid pool, increases the relative proportion of chenodeoxycholate in bile, and reduces the relative proportion of desoxycholic acid, whereas cholate remains unchanged.

Blood cholesterol levels Since the first study performed in 1963 in humans with fiber-rich rolled oats, several studies have reported a beneficial effect of oat bran or oat meal-enriched diets on plasma cholesterol levels in healthy subjects or mildly hyperlipidemic subjects. The daily intakes ranged from 17 to 100 g oat bran, and the duration was 2–12 weeks. Oat bran and oat gum (80% β-glucan) has been served in drinks and instant whips, or baked in bread and muffins. The majority of controlled studies performed in hyperlipidemic subjects showed a decrease in plasma total cholesterol, with most decreases in the range of −4 to −10%. It is noteworthy that decreases in plasma cholesterol are related to decreases in LDL cholesterol, whereas while HDL cholesterol does not generally noticeably change. Consequently, oat bran causes a modest reduction in blood cholesterol level. Larger reductions are seen in subjects who have initially higher blood cholesterol levels ($> 5.9\,\text{mmol}\,\text{l}^{-1}$ or $229\,\text{mg}\,\text{dl}^{-1}$). Studies have shown clearly that soluble β-glucan in oat fiber has a cholesterol-lowering effect in animals and humans. About 3 g of β-glucan per day causes a 10% drop in LDL cholesterol. In bread and muffins, the baking process can decrease the solubility and/or molecular weight of β-glucans, resulting in a weak serum cholesterol-lowering effect. Approximately 100 g of oat bran should be ingested daily to have any effect on plasma cholesterol levels. Oat β-glucan also improves glucose and insulin regulation. The mechanisms of action of viscous polysaccharides like β-glucans in the human gastrointrestinal tract are not completely understood. Increased luminal viscosity is believed to cause delayed gastric emptying and reduce the rate of absorption of nutrients in the small intestine, thereby attenuating the postprandial glucose and insulin response to an oral carbohydrate load. Viscous gums alter bile acid- and cholesterol metabolism, reduce and alter the site of lipid absorption, and undergo fermentation in the colon to short-chain fatty acids. All of these effects might contribute to a reduction in serum cholesterol levels in humans. Full-fat rice bran leads to a similar reduction in serum LDL cholesterol compared with oat bran. Rice bran contains low amounts of soluble fiber, but rice bran oil has a hypolipidemic effect. Essentially all human studies performed since 1969 have shown no effects of wheat bran on fasting blood total cholesterol as well as LDL cholesterol in healthy or hyperlipidemic subjects following a chronic intake of 14–77 g of wheat bran per day. Moreover, in rats, wheat bran essentially does not change fasting blood lipid parameters.

(*See* **Cholesterol**: Properties and Determination; Absorption, Function, and Metabolism.)

Female sex hormones A daily intake of 15 g of wheat bran can significantly reduce circulating levels of estradiol and estrone in premenopausal women by approximately 15–20%. The mechanism by which wheat bran modulates serum estrogen levels most likely involves their enterohepatic circulation. Estrogens are conjugated in the liver to form glucoronides and sulfoglucoronides, and 20–50% of estrogen metabolites are excreted in the bile as these polar biologically inactive substances. Approximately 80% of the estrogens in the gut are reabsorbed, a process that requires deconjugation by bacterial enzymes. Wheat bran intake may lower fecal β-glucoronidase activity. An additional mechanism may be the binding of unconjugated estrogens to fiber in the gut, thus impeding their reabsorption. Oat bran causes reductions in fecal β-glucuronidase activity to a lesser extent than wheat bran. Maize bran has no significant effect on this enzyme.

Vitamin D A high intake of wheat bran can also influence the metabolism of vitamin D, another steroid substance: an additional intake of 20 g of dietary fiber per day from wheat bran can reduce the plasma half-life of tritium-labeled 25-hydroxyvitamin D by approximately 30%. A dietary fiber-induced decrease in vitamin D status may diminish active absorption of dietary calcium. The occurrence of rickets and osteomalacia in Asians living in the UK has been partially associated with a high intake of fiber-rich chapatties (unleavened bread). However, such an adverse effect may only be of clinical importance in subjects with marginal or insufficient vitamin D status and low dietary calcium intake. (*See* **Vitamins**: Overview.)

Effects of Bran on Mineral Metabolism

It has been demonstrated in several human studies that wheat bran and high-fiber diets decrease intestinal Ca absorption. Wheat bran alters the usual inverse relationship between calcium load and fractional absorption in humans. This effect is the result of the high Ca-binding capacity of wheat bran. At pH 8.0, 1 g of wheat bran can bind approximately 0.45 mmol of calcium. This amount is slightly higher than the calcium binding of rye bran (0.38 mmol of Ca per gram of bran) and slightly lower than the calcium binding of rice bran ($0.49\,\text{mmol}\,\text{g}^{-1}$), whereas soy bran only binds 0.19 mmol of calcium per gram at pH 8.0. In wheat bran, phytic acid and dietary fiber are both present in high quantities. Theoretically, each mole of phytic acid can maximally bind 6 moles of calcium. Phytic acid can form insoluble calcium salts resulting in a low bioavailability of calcium, but wheat bran can bind more calcium *in vitro* than expected from the molar binding ratio of bran phytate:calcium. *In vitro* calcium binding to wheat bran is linear over a wide range. Thus, the amount of calcium binding by wheat bran cannot be explained by the content of phytic acid alone. Other bran constituents may influence calcium bioavailability. Although neutral cellulose has little affinity to cations, lignin strongly binds calcium. The lignin in wheat bran is indigestible during the gastrointestinal transit. However, as the result of an adaptation process, wheat bran intake is also associated with a reduced renal calcium excretion. Consequently, a high bran intake does not necessarily increase the risk of a negative calcium balance.

Intakes of high (> 15 g) amounts of bran (wheat, oat, barley) can reduce zinc absorption and decrease iron availability. Measures to reduce the phytate content of cereals like malting and soaking (activation of endogenous phytase activity) can significantly improve zinc and iron absorption. Nevertheless, in diets that provide sufficient amounts of trace elements, decreased absorption does not necessarily influence the trace element balance and status of human subjects. A reduced endogenous fecal excretion can compensate for the lower absorption. However, if zinc and iron intakes are marginal or inadequate, the additional consumption of a fiber concentrate may contribute to the development of trace element deficiencies.

In animal diets, supplemental phytases of microbial and cereal sources improve dietary phytate phosphorus utilization. Such a measure is environmentally advantageous in order to replace the content of inorganic phosphorus in those diets.

Beneficial Health Effects of Bran and its Components

In animals, wheat bran causes a significant reduction in incidence of mammary cancer. A suggested protective effect of wheat bran on breast cancer in humans may be due, at least in part, to the effect on circulating estrogen concentrations and their bioavailability. Moreover, the lignans especially present in the aleurone layer of some cereals (Table 2) may exert antiestrogenic effects in humans.

Wheat bran and its component phytic acid have both been shown to decrease early markers of colon carcinogenesis like certain indices of cell proliferation and certain aberrant cryt foci parameters. Wheat bran, dephytinized wheat bran, and phytic acid induce cell apoptosis, increase cell differentiation,

and favorably affect colon morphology, thus indicating that phytic acid as well as other components of bran may be preventive against colon cancer. A lower concentration of the co-carcinogenic secondary bile acids and a high bacterial production of butyrate may contribute to a normal metabolism and proliferation of the colon epithelial cells after bran intake.

Phytic acid inhibits *in vitro* and *in vivo* the enzymatic degradation of starch and reduces the glycemic index owing to complexation of the starch molecule, amylase, and calcium ions, which are essential for the amylase activity. The reduced rate of glucose absorption and reduced postprandial increase in blood glucose may both be beneficial in prophylaxis and treatment of diabetes mellitus. Protective effects of phytic acid against lead toxicity have been demonstrated in the presence of calcium ions. Acute toxic effects of 1 g of lead per kilogram of diet can be reduced by 20 g of calcium phytate per kilogram.

Commercial Products

Both fine wheat bran and coarse wheat bran are available for the consumer. One tablespoon of bran contains approximately 2 g of dietary fiber. Bran may be used alone as a cereal, mixed with other cereals, sprinkled on cereals or included in bran muffins, hamburgers, etc. Bran is legally regarded as a food. Moisture should not exceed 14%, and the amount of starch in wheat bran should not exceed 15% of the dry weight. Since especially the bran fraction of wheat can accumulate relatively high amounts of cadmium, the cadmium content should not exceed 50–60 $\mu g\,kg^{-1}$. Wheat bran tablets are also obtainable – one tablet equals approximately 2 g of dietary fiber. Moreover, breakfast cereals with a bran content up to 85% are in the market (such as Kellogs All-Bran Plus™). These products contain 11 g of dietary fiber per 40-g portion. Several baked goods are enriched with wheat bran to increase the fiber content. Enrichment should at least result in a doubling of the dietary fiber content, and a good source of dietary fiber should contain 2 g per portion. Although coarse wheat bran increases the stool weight more efficiently than fine wheat bran, the latter is often used to enrich baked goods, as it is easier to process and is preferred by the consumer.

Pure oat bran flakes are also commercially available. One portion (25 g) of oat bran contains approximately 2.3 g of insoluble and 1.95 g of soluble dietary fiber. Moreover, oat bran-enriched breakfast cereals and oat flakes containing 80% oat bran are obtainable, and oat bran can be used as a fat replacement in sausages like Frankfurters (6% oat bran, 30% water).

Protex™ (Food Engineering International, Inc.) is a specially processed rice bran derivative that may be used in baked goods, breakfast cereals, macaroni and noodle products, and milk-like beverages.

New Potential Sources

Psyllium is a new source of dietary fiber. It is made from ground husks of psyllium seeds and has the ability to hold water and form bulk. Psyllium contains up to 85% soluble fiber. Psyllium holds water better than bran, so smaller amounts are needed to be effective. Moreover, psyllium effectively lowers serum LDL cholesterol and blood glucose and insulin levels. In 1998, the US Food and Drug Administration ruled that psyllium, like oat bran, may reduce the risk of heart disease. Psyllium has several advantages over oat bran. It is more versatile, since it is a far more concentrated form of fiber. The US Food and Drug Administration allows the addition of psyllium to breakfast cereals, and FDA officials expect more psyllium-containing foods, from waffles to biscuits, to hit the market.

Soy bran is a dietary fiber concentrate that is obtained during the production of soy protein, and the insoluble portion is separated by centrifugation. The concentrate is defatted and contains the soy hulls. The dietary fiber content is approximately 67–80% (Table 3). Soy bran is sold in competition with other dietary fiber concentrates. It is effective in increasing stool weight (Table 2), and 25 g of soy fiber significantly reduces total serum cholesterol levels, especially in patients with blood cholesterol concentrations above 7.7 $mmol\,l^{-1}$ (300 $mg\,dl^{-1}$).

Several other food residues with a high dietary fiber content are obtained during food processing (Table 4). They can be used by food technologists, e.g., for the fiber enrichment of bread (beer draff) or for the processing of fiber and β-carotene-rich fruit juices (carrot husks).

Wheat fiber (fiber content 97%) is a concentrate of cellulose and hemicellulose that is free of phytic acid

Table 3 Effect of bran and other fiber-rich foods on stool weight

	$\text{Efficiency index} = \frac{\text{increase in stool weight (g)}}{\text{fiber intake (g)}}$
Wheat bran, fine[a]	3.1
Wheat bran, coarse[a]	6.2
Bread, whole wheat grain, fine[a]	5.6
Bread, whole wheat grain, coarse[a]	8.2
Barley fiber concentrate[a]	4.2
Oat bran[a]	4.54
Soy bran[b]	1.5–2.3
Carrots, raw[a]	3.6

[a]Daily intake of 14 g of fiber; [b]daily intake of 30–65 g of fiber.

Table 4 Different fiber concentrates of food residues

Concentrate	*Dietary fiber content (g per 100 g dry weight)*
Beer draff	60–82
Maize bran	42–79
Soy bran	67–80
Pea seeds	83–90
Sugar-beet fiber	67–72
Citrus fiber, orange fiber	50–60
Cacao seeds	55
Carrot husks	60–68

and gluten. This insoluble wheat fiber can be used for the production of common foods and also for the preparation of dietetic foods (e.g., for patients with celiac disease).

The drying of apples results in a dietary fiber concentrate that is rich in soluble fiber (pectins). The combination of wheat and apple fiber influences the ratio of insoluble to soluble dietary fiber in processed foods.

See also: **Bioavailability of Nutrients**; **Cereals**: Contribution to the Diet; **Dietary Fiber**: Properties and Sources; Effects of Fiber on Absorption; **Phytic Acid**: Nutritional Impact; Nutritional Impact

Further Reading

Bartnik M and Jakubczyk T (1989) Chemical composition and nutritive value of wheat bran. In: Bourne GH (ed.) *Nutritional Value of Cereal Products, Beans and Starches. World Review of Nutrition and Dietetics* 60: 92–131.

Batchelor AJ and Compston JE (1983) Reduced plasma half-life of radio-labelled 25-hydroxyvitamin D_3 in subjects receiving a high-fiber diet. *British Journal of Nutrition* 49: 213–216.

Chen HL, Haack VS, Janecky CW, Vollendorf NW and Marlett JA (1998) Mechanisms by which wheat bran and oat bran increase stool weight in humans. *American Journal of Clinical Nutrition* 68: 711–719.

Ferguson LR and Harris PF (1999) Protection against cancer by wheat bran: role of dietary fibre and phytochemicals. *European Journal of Cancer Prevention* 8: 17–25.

Lairon D (1996) Dietary fibres: effects on lipid metabolism and mechanisms of action. *European Journal of Clinical Nutrition* 50: 125–133.

Rose DP, Goldman M, Connolly JM and Strong LE (1991) High-fiber diet reduces serum estrogen concentrations in premenopausal women. *American Journal of Clinical Nutrition* 54: 520–525.

Saunders RM (1990) The properties of rice bran as a foodstuff. *Cereal Foods World* 35: 632–636.

Wood P (ed.) (1993) *Oat Bran*. St. Paul, MN: American Association of Cereal Chemists.

Energy Value

S Brooks, 3W Banting Research Centre, Ottawa, Canada
R Mongeau, Ottawa, Aylmer (Quèbec), Quebec, Canada

Caloric Value in Human Nutrition

All food contains energy, which is needed by the body to fuel metabolic processes. This energy is obtained though the oxidation of macronutrients (carbohydrate, fat, and protein) and is partly captured as chemical energy in compounds such as adenosine triphosphate (ATP). These 'high-energy' compounds are used to drive many processes, including the biosynthesis of cellular components, the maintenance of ion concentrations across membranes, the initial phases of catabolic pathways, cell replication, mechanical function, etc. The large amount of energy released by the hydrolysis of ATP (approximately $7\,\text{kcal}\,\text{mol}^{-1}$ under cellular conditions) is captured by coupling its hydrolysis to other chemical reactions to drive them to completion.

The energy captured as ATP equivalents is one end of a long chain of digestive and metabolic processes that starts as the gross energy of the macronutrients in ingested food (**Figure 1**). Significant losses in energy occur during the digestive process. These can be identified and quantified as losses due to incomplete digestion and absorption (loss in fecal energy to give apparent digestible energy; DE_{app}), losses to fermentation gases and urinary nitrogen energy (to give apparent metabolizable energy), as well as losses associated with the digestive process itself and with the absorption and storage of food (to give net energy). Calculating the metabolizable energy from a food requires the collection of all excreta (including methane and hydrogen gasses). For substances that are fermented, extra information is also required as discussed later in this chapter. Net energy values are more difficult to estimate and require some assumptions about digestive processes (see section on Factorial Models of Fermentation, eqn (19)).

Metabolizable energy values are the commonly reported food energy values found on product labels. These were initially determined in the 1900s by a German scientist who was a pioneer in this area. To honor his name, these factors are commonly referred to as 'Atwater factors.' Both general and specific Atwater factors exist. The specific values were measured in individual foods. The general values (4 for carbohydrate, 9 for fat, 4 for protein, and 7 for alcohol) were calculated from the specific Atwater factors using

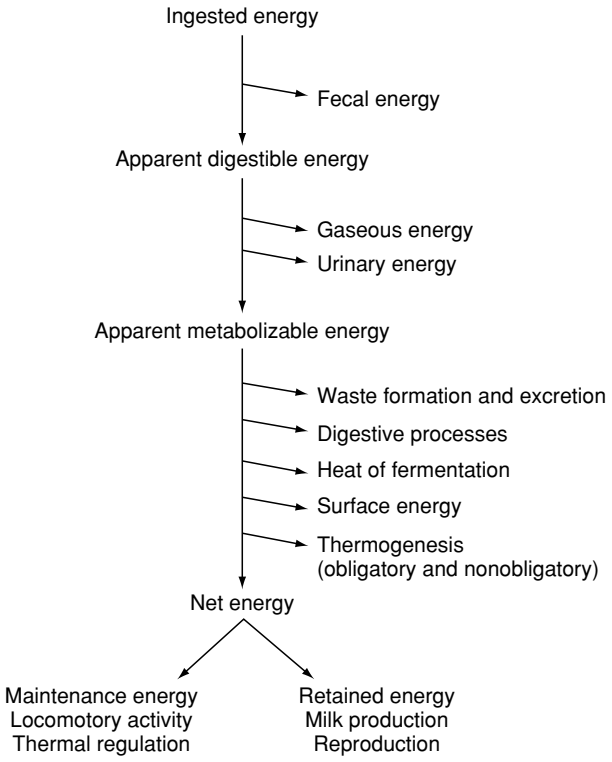

Figure 1 Diagrammatic representation of the energy losses associated with various stages of digestion and the resulting determined energy values.

average intakes of mixed diets. The more modern food energy values, as contained in tables such as the US Department of Agriculture Handbook #8, represent specific Atwater factors as modified by Merrill and Watt. Although we refer to these factors as metabolizable energy values for macronutrients, the values for carbohydrate and fat are, in reality, DE_{app} values. This is because urinary and gaseous energy losses were considered as insignificant for these macronutrients. In the case of the protein Atwater factors, urinary energy losses were taken into account, but gaseous energy losses were, again, assumed to be negligible. The Atwater factors for fat, protein, and carbohydrate represent the total energy available for metabolism, and this heat should be equivalent to the energy measured with a whole-body calorimeter.

Digestible sugars, fats, and protein are readily absorbed in the small intestine by physiological and biochemical processes that capture the maximum energy from these food components. These macronutrients are virtually completely absorbed by the small intestine with little energy finding its way into the large intestine. In fact, with diets high in digestible sugars, fats, and protein, most of the energy entering the large intestine comes from small intestine secretions and from the sloughing off of dead intestinal cells. It is the relative simplicity of the digestive process coupled with the high degree of absorption that allows the comparatively easy metabolic energy calculation for these macronutrients.

Not all food components are readily digested and absorbed in the small intestine. Nonstarch long-chain polysaccharides such as those associated with plant cell walls are not digested by human enzymatic secretions. Many other carbohydrates are not well absorbed by the small intestine. These pass into the large intestine where they are anaerobically fermented by the bacterial population that resides there. For sake of clarity, the terms 'undigested' and 'resistant' will be used here to refer to the degradation processes occurring in the upper gastrointestinal tract (mostly by host digestive enzymes), and the term 'fermentation' and related words will be used to refer to the bacterial degradation processes occurring in the lower gastrointestinal tract. The energy losses associated with fermentation are not as readily measured, and so calculation of the metabolizable energy for fermentable food components is not as straightforward as it is for sugars, fats and protein. The present chapter outlines different methods for obtaining the metabolizable energy of food components not digested and absorbed in the small intestine.

A short note about the methods used to determine macronutrient content of foods is in order. In some countries like the USA and Canada, the carbohydrate content is determined by subtracting the fat, protein, water, and ash content of the food from the total weight (difference method). This measurement includes both 'available carbohydrate' (i.e., carbohydrate that can be digested in the small intestine) as well as dietary fiber. Other countries, such as Great Britain and Australia, measure 'available carbohydrate' directly using chromatographic methods. The carbohydrate values from this latter method do not include dietary fiber. Specific Atwater factors used in the USA and Canada take into account some (but not all) of the energy losses associated with dietary fiber because they rely on the apparent digestibility of carbohydrate to calculate food energy. This approach overestimates energy availability from mixed diets. In Europe, the general Atwater factor for carbohydrate ($4\,kcal\,g^{-1}$, $16.7\,kJ$) is used to calculate carbohydrate energy after subtracting the dietary fiber content. This method underestimates the energy of mixed diets. When 'available carbohydrate' is measured directly, one must also measure dietary fiber and use a dietary fiber-specific energy value to calculate the energy contribution from dietary fiber.

Digestible Energy and Metabolizable Energy: Calculations

Digestible Energy

As noted above, the digestible energy of a substance is the difference between the ingested energy and the energy excreted in the feces. This is, in principle, a simple definition but gives rise to two different digestible energy values depending on how the measurement is made: the apparent digestible energy and the partial digestible energy. We begin by defining IM_s as the ingested mass of a measured substance FM_s as the amount of the substance of interest. This substance can represent any dietary material but for the purposes of this article, we will assume that it escapes digestion in the small intestine and reaches the large intestine where it may (or may not) be fermented. If we further define recovered in the feces, the apparent digestibility (D_{app}) can be calculated as:

$$D_{app} = (IM_s - FM_s)/IM_s. \quad (1)$$

The apparent digestible energy value (DE_{app}) is simply the apparent digestibility multiplied by the heat of combustion (ΔH_c) of the material:

$$DE_{app} = D_{app} \times \Delta H_c. \quad (2)$$

The DE_{app} takes into account only the energy loss associated with incomplete digestion of a single food component and not any potential interaction of the component with macronutrients. If the food component is fermented, part of its energy value will be captured and utilized by the host, part of the remaining energy will appear in the feces, and part will be lost to other processes (discussed below). Thus, the DE_{app} overestimates the true digestible energy.

Partial digestible energy values (DE_{part}) describe the influence of a dietary substance on the energy digestibility of the whole diet. As such, they represent a more complete picture of the energy intake and excretion associated with ingestion of a particular food component. Calculating DE_{part} values requires measurement of overall energy balance. The first step is to define the energy digestibility (D_{energy}) of the overall diet as:

$$D_{energy} = (IE_T - FE)/IE_T, \quad (3)$$

where IE_T is the total ingested energy, and FE is the fecal energy. Partial digestible energy values are determined by adding a varying amount of a food component (as a supplement) to an unchanging basal diet. The energy digestibilities of the unsupplemented and supplemented diets are then determined (eqn (3)) and are plotted as a function of the fraction of ingested energy that is derived from the supplement (IE_S/IE_T). If the supplement has a lower digestibility than that of the overall diet, the graph should be a straight line with a negative slope – the energy digestibility of the overall diet will decrease with increasing amounts of the supplement. The partial indigestibility of the supplement (S_D) is given by:

$$S_D = \Delta D_{energy}/\Delta(IE_S/IE_T) + (1 - D_{energy,0}), \quad (4)$$

where $D_{energy,0}$ is the energy digestibility of the basal (unsupplemented) diet. The DE_{part} is obtained by multiplying the partial digestibility (1 − the partial indigestibility) of the supplement by its heat of combustion.

$$DE_{part} = \Delta H_c \times (1 - S_D). \quad (5)$$

Partial and apparent digestibilities can differ by a considerable amount. In general, $DE_{part} < DE_{app}$ because DE_{part} includes energy losses that are not accounted for by DE_{app} measurements. This is illustrated by an experiment that compared the DE_{app} and DE_{part} values for the fiber component of hard red wheat bran (bran is approximately 40% fiber). Measurement of the diet and fecal fiber content as well as the diet and fecal energy content gave a $DE_{app} = 1.82 \pm 0.17\,kcal\,g^{-1}$ ($7.6\,kJ\,g^{-1}$) and a $DE_{part} = 0.45 \pm 0.10\,kcal\,g^{-1}$ ($1.9\,kJ\,g^{-1}$). The DE_{app} value shows that, for every gram of wheat bran fiber ingested, $4.15 - 1.82 = 2.33$ kcal (9.7 kJ) of energy was lost to feces in the form of fiber. The DE_{part} value shows that $1.82 - 0.45 = 1.37$ kcal (5.8 kJ) of extra energy was lost to the feces in addition to that lost from fecal fiber excretion. Thus, only $0.45/1.82 \times 100\% = 25\%$ of the energy of the fermented wheat fiber was potentially retained by the rats (see below for other calculations on the metabolizable energy value of wheat fiber). These calculations show the importance of determining the total energy losses in a digestion experiment involving fermentable substances.

Metabolizable Energy

Like digestible energy values, it is possible to define two different metabolizable energy values. Livesey (1993) calls the Atwater metabolizable energy values 'apparent metabolizable energy' values (ME_{app}) because they do not take into account all the energy losses normally associated with metabolizable energy measurements:

$$ME = IE - (FE + UE + GE) \quad (6)$$

where UE is the urinary energy and GE is the gaseous energy. If a strict parallel between digestible energy and metabolizable energy is to be made, one could also define ME_{app} as $DE_{app} - (UE + GE)$. In essence,

this is the definition used by Atwater with some assumptions. For fat and carbohydrate, it was assumed that GE and UE were negligible. For protein, it was assumed that GE was negligible but UE was not.

Partial metabolizable energy values (ME_{part}) can also be calculated in a fashion analogous to DE_{part} values. In this case, we define the total energy loss as:

$$\sum E_{losses} = FE + GE + UE. \quad (7)$$

Once again, we define the energy metabolizability of the total diet as:

$$M_{energy} = (IE_T - \sum E_{losses})/IE_T \quad (8)$$

and use this value to define a partial energy 'non-metabolizability' (S_M).

$$S_M = \Delta M_{energy}/\Delta(IE_s/IE_T) + (1 - M_{energy,0}), \quad (9)$$

where $M_{energy,0}$ is the energy metabolizability of the basal (unsupplemented) diet. The ME_{part} is then given by:

$$ME_{part} = \Delta H_c \times (1 - S_M). \quad (10)$$

Equations that minimize the error structure for experimentally determining DE_{part} and ME_{part} are also available.

Like DE_{part} values, ME_{part} values describe the influence of a part of the diet on the energy metabolizability of the whole diet. For protein and fat, it is thought that $ME_{part} > ME_{app}$ because fecal excretion of bacterial nitrogen and fat will decrease the ME_{app} value. Note that energy losses to bacterial fat and nitrogen are not simply due to the loss of macronutrient mass. Energy is required for the *de novo* bacterial synthesis of fat- and protein-containing macromolecular structures, which are then excreted in the feces. This energy is also lost. For carbohydrates, it is thought that ME_{part} and ME_{app} are equivalent since bacteria do not contain appreciable amounts of carbohydrate. For carbohydrates that escape digestion in the small intestine and are partially or completely fermented, like those in dietary fiber, $ME_{app} > ME_{part}$ because of the extra energy losses that occur during fermentation. This energy loss is the subject of this chapter.

Metabolizable or Net Energy?

The discrepancies between ME_{app} and ME_{part} energy values mean that a more accurate system for measuring food metabolizable energy must be employed for diets containing appreciable amounts of undigested and fermented carbohydrates as the greatest differences occur with these food components. Many methods can be used to obtain the ME values of food components that are not digested by the small intestine. Whole-body calorimetric measurements (in a chamber where an individual can be placed for experimentation) can be used to determine the energy value directly, estimates of the metabolizability can be obtained from mixed diets, body energy measurements can be made in laboratory animals and related back to the metabolizable energy of the food, and theoretical calculations can provide a general model for estimating metabolizable energy.

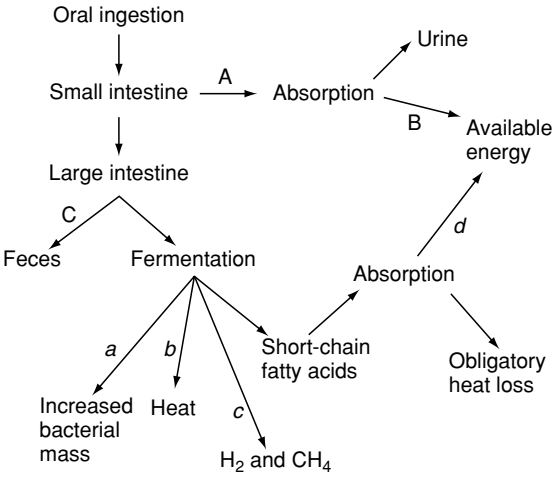

Figure 2 Scheme showing various stages of digestion of a macronutrient that is partially digested in the small intestine and partially fermented in the large intestine. The letters represent fractions of the initial amount that proceed down various pathways (see text).

Large differences exist between the energy losses associated with the digestion of macronutrients (fat, carbohydrate, and protein) and undigested and fermented carbohydrates. Figure 2 shows an example of the various processes that occur in the gastrointestinal tract during the consumption of a partially digested and partially fermented food. Because the final energy value is net of all major losses of energy, many researchers have called it a net metabolizable energy (NME). This should not be confused with the net energy values used for livestock, which represent different energy measurements. Also, it is not the net energy of Figure 1, which incorporates many more energy losses. The NME is a more practical measure of food energy for undigested and fermented substrates because it takes into account energy losses that do not occur for fat, carbohydrate, and protein. Use of this value would allow comparisons between the energy values of different food components.

Mathematical Models for Undigested and Fermented Compounds: Sugar Alcohols

It is technically difficult and expensive to measure all stages of the fermentation process and determine the

energy losses associated with each step. This is true of human whole-body calorimetric measurements since analysis requires a complete energy balance profile for each subject (see below). A more simplistic approach involves the construction of a mathematical model of the digestive events in the large intestine. This approach has most commonly been applied to the energy value of sugar alcohols but can also be extended to other partially or completely fermentable carbohydrates.

Common dietary sugars (such as glucose, galactose, and fructose) have a ketone or aldehyde group that allows cyclization to a hexose or pentose ring structure. Reduction of this carbonyl group to an alcohol produces a sugar alcohol that cannot cyclize. The resulting sugar alcohols are not well absorbed in the small intestine so that a significant percentage can pass intact into the large intestine. Isomalt (Palatinit®), an equimolar mixture of two disaccharide alcohols (α-D-glucopyranosyl-1,1-D-mannitol and α-D-glucopyranosyl-1,6-D-sorbitol), is an example of a sugar alcohol that is virtually unabsorbed in the small intestine. It passes intact into the large intestine, where it is rapidly and completely fermented by the bacterial population that resides there. As such, it represents a model for a theoretical dietary fiber that is completely fermented in the large intestine and can be used as a basis for constructing a mathematical model.

A good mathematical model seeks to define all losses that occur during fermentation of ingested material and assign reasonable factors to these losses. Figure 2 presents the scheme for such a model. Fermentable substances provide carbon units for bacterial growth and reproduction and serve as metabolizable substrates to meet the colonic bacterial population's energy requirements for maintenance and growth. Four major products result from anaerobic colonic fermentation: increased bacterial mass (factor *a*, Figure 2), heat energy loss due to fermentation (factor *b*, Figure 2), methane and hydrogen gas (factor *c*, Figure 2), and short-chain fatty acids (SCFAs). Humans can only use SCFAs as an energy source. Thus, the metabolic energy yield of the sugar alcohol has been reduced by the losses to bacterial mass, as well as the heat and gas produced during fermentation. It has also been reported that SCFAs may produce significantly less ATP per gram of absorbed SCFA than glucose, so an additional energy loss to this difference in metabolic efficiency should also be included (factor *d*, Figure 2).

The fermentation model of Figure 2 can be expressed as an equation that incorporates the four major processes by which energy is lost during fermentation. The equation describing the scheme of Figure 2, as it applies to sugar alcohols, was first developed for the Nutrition Council of Holland and has been recognized by several authors as an alternative method for calculating the energy value of sugar alcohols. For partly fermentable substances, it can be written as:

$$\mathrm{NME} = \Delta H_c \times [(A \times B) + \alpha \times (1 - A - C) - \beta], \tag{11}$$

where, ΔH_c is the heat of combustion of the substance in question, A represents the fraction of the ingested substance absorbed in the small intestine, B represents the fraction of absorbed substance that is metabolized, and C represents the fraction of the ingested substance that is excreted intact in feces (Figure 2). The factor α represents the proportion of energy derived from fermentation of the substance. It can be calculated from eqn (12):

$$\alpha = (1 - a - b - c) \times d. \tag{12}$$

The factor β was not present in the original equation and represents the extra energy lost due to the presence of the fermentable carbohydrate. This can come about through binding to macronutrients, changes in osmotic balance, alterations in transit time, changes in intestinal viscosity, inhibition of digestive enzymes, inhibition of macronutrient uptake, or increased sloughing of intestinal cells. As discussed above (see section on DE_{part} and ME_{part}), measuring the appearance of an undigested food component in the feces is not sufficient to account for all of the energy losses that can occur when a diet is supplemented with a fermentable carbohydrate. In a practical sense, it is impossible to differentiate between any 'extra' energy lost to the feces (β) and the energy lost due to increased bacterial mass that results from fermentation (factor *a*). The solution is to set $\beta = 0$ in eqn (11) and measure total fecal output.

Using the mathematical model of eqns (11) and (12) requires a knowledge of factors A, B, and C, which are obtained from experimental evidence. For sugar alcohols, values for factor A have been obtained from ileostomy patients or by direct sampling using multiple lumen tubes. Factor B (the amount of sugar alcohol metabolized) has been estimated indirectly by measuring urinary sugar alcohol content (the amount of sugar alcohol not metabolized).

In the original factorial model paper, the Nutrition Council of Holland held α to be constant for all sugar alcohols and estimated it at 0.5. This value included fermentation energy losses resulting from increased microbial mass excreted in the feces ($a \approx 0.2$), loss of H_2 and CH_4 ($c = 0.03$–0.08), heat produced by microbes ($b = 0.02$–0.05), and higher heat loss during utilization of short-chain fatty acids as compared

Table 1 Potential error associated with oxidation of 50 g of a carbohydrate versus fermentation and oxidation of the released short-chain fatty acids[a]

Substrate	*RQ*	*Oxidation substrates (mol)*[b]	*O_2 consumed (mol)*	*H^{eq}*[c]	*kcal per 50 g*
50 g of sucrose	1.00	0.146 CHO	1.753	5.012	250.6
50 g of lactitol	0.943	0.409 (CHO) + 0.099 (fat)	1.113 (CHO) + 0.269 (fat)	4.981 (CHO) + 4.682 (fat)	153.1
50 g of lactitol	0.943	0.508 SCFA	1.382	4.663	144.5

[a]The table illustrates the potential error involved in measuring the apparent metabolic energy value of a completely fermentable supplement (lactitol) by indirect calorimetry. This was done by calculating the apparent metabolic energy value using two different methods. It was assumed that lactitol was fermented to a mixture of SCFAs followed by complete oxidation to CO_2 and water plus fermentation gases. In an indirect calorimeter, oxidation of SCFAs can be mistaken as oxidation of fat and carbohydrate in a 0.195:0.808 ratio. Oxidation of these substrates would give 153.1 kcal per 50 g. Oxidation of the SCFA directly would give 144.5 kcal per 50 g of original lactitol. Thus, an error of approximately 6% can occur when indirect calorimetry is used to measure the metabolic energy value of lactitol. The yield of SCFAs from lactitol and the proportion of acetate, propionate, and butyrate (115.1:35.38:24.5 per 100 mol of glucose) represent weighted averages of literature values for ruminants. As such, these calculations approximate the actual error associated with indirect calorimetry, which varies from individual to individual.

[b]Equivalent mol of sucrose or SCFAs obtained from 50 g of substrate.

[c]H^{eq} values from Livesey G and Elia M (1988) Estimation of energy expenditure, net carbohydrate utilization, and net oxidation and synthesis by indirect calorimetry: evaluation of errors with special reference to the detailed composition of fuels. *American Journal of Clinical Nutrition* 47: 608–628. The H^{eq} value for glucose was used for carbohydrate and the H^{eq} value for the fat mixture of an omnivore diet were used for fat. The H^{eq} value for SCFAs represents a weighted average.

with glucose ($d = 0.80$–0.85). Summing these factors gives a value of 0.40–0.53. The value of 0.5 was adopted because the Nutrition Council of Holland did not want to suggest that the value was known with high precision (they could have used 0.46–0.47 as the average) and because they wanted to include some energy lost to increased cell proliferation in the large intestine caused by increased fermentation. Thus, the slightly higher value of 0.5 was chosen as an estimate of the total losses during fermentation. In a later review of the value of α, it was determined that the individual factors that contribute to α should give a value closer to 0.6. However, in the same review, the author noted that the factorial procedure overestimates energy values to the extent that it takes no account of osmotic effects of the sugar alcohol. Thus, the results of energy values calculated by the Dutch method are probably fairly accurate. Many researchers agree that the value of 0.5 represents a reasonable estimate of the fermentation process.

For completely fermentable substrates that are not absorbed by the small intestine, like sugar alcohols, the MNE value is $0.5 \times \Delta H_c$ because $A = 0$ and $C = 0$. As indicated above, the value of β is set to 0, and any 'extra' energy lost due to ingestion is incorporated into the factor α, which has been estimated at 0.5. Thus, for the sugar alcohol isomalt (which is completely fermented and not absorbed by the small intestine), the energy value is approximately 1.9 kcal g^{-1} (8 kJ g^{-1}, approximately half the heat of combustion). For completely fermentable dietary fibers, the $0.5 \times \Delta H_c$ rule of thumb represents an upper limit of the NME since dietary fiber is known to bind fats and cholesterol as well as inhibit carbohydrate absorption. Thus, dietary fiber may interact with other macronutrients to decrease their metabolizable energy values. This will give a lower NME value for dietary fiber since there will be a higher excretion of fecal energy.

When all the energy losses are included, the energy value of dietary fiber can be small compared with the ingested value. For example, measurements with hard red spring wheat fiber give NME values of approximately 0.3 kcal g^{-1}, even though the total ingested energy is approximately 4 kcal g^{-1}. In the case of hard red spring wheat bran, much of the energy is lost to excretion; wheat bran dietary fiber is incompletely fermented so that approximately 60% is lost directly to the feces. In addition, a significant loss of energy comes from dietary fiber interactions with macronutrients plus other losses due to the presence of wheat bran fiber (β). Finally, energy is lost during the fermentation process itself. The energy value of hard red spring wheat fiber is not typical of other dietary fibers, as shall be shown in the following sections (Table 1).

Whole-body Calorimetric Measurements: Lactitol as a Model

Indirect calorimetry (a type of whole body calorimetry where all gas consumption and gas exhalation are continuously monitored) can be used to determine energy expenditure through the use of equations relating total O_2 consumption, CO_2 expiration, and urinary nitrogen excretion to energy utilization. One such equation is:

$$\text{Heat production} = 16.175\,O_2 + 5.021\,CO_2 - 5.987\ \text{urea N} - 4.5\,H_2. \quad (13)$$

In addition to total energy production, indirect calorimetry provides information on macronutrient utilization through the nonprotein respiratory quotient

(RQ). The RQ is the ratio of CO_2 expired to O_2 consumed during a defined period. The theoretical RQ values are known for fat, carbohydrate, and protein oxidation, making it possible to determine total fat, carbohydrate, and protein oxidation. When protein oxidation is estimated from nitrogen excretion, fat and carbohydrate oxidation can be calculated as follows:

$$\text{Nonprotein RQ} = 1.00 \times \text{carbohydrate oxidation} + 0.707 \times \text{fat oxidation}. \quad (14)$$

However, caution is needed when estimating fuel oxidation from nonprotein RQ values, since the RQ itself can vary considerably depending on the fat and protein mixture being oxidized. In addition to the problem of obtaining an accurate picture of fat and carbohydrate oxidation from nonprotein RQ values, it is also not possible to distinguish between the oxidation of carbohydrate + fat and the oxidation of short-chain fatty acids produced by carbohydrate fermentation. When measuring the apparent metabolizable energy value of a macronutrient that represents a small proportion of the total diet, there is a high potential for error when macronutrient oxidation profiles are measured. However, theoretical calculations show that there is much less error associated with measuring energy loss.

An example of the use of indirect calorimetry for measuring the metabolic energy of completely fermentable substances is the determination of the net energy value for lactitol, a disaccharide sugar alcohol (4-*O*-(β-D-galactopyranosyl)-D-sorbitol). The total energy balance was measured during the experimental period in the calorimeter. The retained energy of each subject was then calculated as:

$$\text{RE (retained energy)} = \text{IE}_\text{T} - (\text{fecal E} + \text{urine E} + \text{gaseous E} + \text{expended E}). \quad (15)$$

The energy needed to maintain energy balance (metabolizable energy for maintenance; ME_m) could then be determined as:

$$\text{ME}_\text{m} = \text{metabolizable energy intake} + \text{negative RE} - 1.11 \times \text{positive RE}. \quad (16)$$

The factor 1.11 represents the cost of depositing body energy and was measured in an earlier experiment. The difference in ME_m on a diet supplemented with sucrose versus one supplemented with lactitol is a measure of the energy lost from lactitol, and can be expressed as the change in ME_m adjusted for the metabolic weight of the individuals ($\text{weight}^{0.75}$):

$$\Delta\text{ME}_\text{m} = (\text{ME}_\text{m}/\text{weight}^{0.75})_\text{lactitol} - (\text{ME}_\text{m}/\text{weight}^{0.75})_\text{sucrose}, \quad (17)$$

and the NME can then be calculated as:

$$\text{NME} = \text{ME}_\text{part} - \Delta\text{ME}_\text{m}/(\text{IS}/\text{weight}^{0.75})_\text{lactitol}, \quad (18)$$

where ME_part is obtained from eqn (10), and IS is the ingested mass of the supplement (lactitol). The ME_m values are divided by the metabolic weight to account for differences in metabolic mass. Using this procedure, a value of 2.4 kcal g^{-1} (10 kJ g^{-1}) for lactitol was obtained. Note that some measure of the energy expenditure is required in this experiment (eqn (15)). The authors relied on diaries and ankle actometers and corrected to equal actometer counts. They acknowledged that this was not very satisfactory; there is a potential 15% error in the final net energy value for lactitol.

Metabolizable Energy Value of Dietary Fiber in Mixed Diets

Up to this point, this article has focused on undigested and fermented food components as a general group of substances. This section will deal specifically with dietary fiber, a partly fermented food component that passes through the small intestine unacted upon by human digestive enzymes. Dietary fiber is a mixture of nonstarch polysaccharides and lignin that are structural components of plant cell walls.

As indicated above, it is not appropriate to use metabolizable energy values for dietary fiber since many more energy losses are associated with fiber digestion as compared with the digestion of fat, carbohydrate, and protein. Equations have been developed to take account of these extra losses and provide a method to calculate the NME value of a fermentable dietary component when the fecal energy losses are known. These equations show that approximately one half of the potential energy is available to the host when undigested carbohydrates are completely fermented. This value is widely accepted.

Knowledge of the energy yield from the fermented portion of dietary fiber is not sufficient in and of itself to allow the calculation of dietary fiber NME since a significant proportion of the 'fermentable' dietary components are not fermented but appear in the feces. In addition, some 'extra' energy (factor β, eqn (13)) may be lost during digestion because of the intimate physical interaction between dietary fiber and the food matrix (*See* Dietary Fiber: Properties and Sources). The aim of these studies is to determine a general NME value for dietary fiber that can be used to calculate the metabolizable energy of mixed human diets. To overcome these problems, two solutions have been employed. The first solution involved determining individual digestibility factors for

specific foods and using these to calculate metabolizable energy intake. This method has the advantage that macronutrient digestibility is measured directly for specific foods and multiplied by the specific heats of combustion when available. In this method, the calculated digestible energies of the diets are close to the measured values for mixed diets since the final value reflects the sum of the individual components. This method has some problems. In the system of Merrill and Watt, the digestibility of carbohydrates was determined by difference, so accurate values for carbohydrate digestibility are lacking. This also means that a digestibility value for the fermentable food component portion of the carbohydrate fraction (as determined by difference) is lacking. It is, therefore, not possible to calculate the metabolic energy of a food of known composition since values for undigested and fermented components are not available. This limits the usefulness of this system to foods that have been measured. In addition to this problem, the NME value is appropriate for undigested and fermented food components, whereas the digestible energy value is not. Thus, this system does not allow any correction for energy lost when food components are resistant to digestion but fermented in the lower gut.

The second solution is to use an average digestibility value determined by examining typical mixed diets. This can be calculated by two separate methods: using a factorial approach or using an empirical model. Factorial models relate total protein, fat, carbohydrate, and dietary fiber intake to digestible energy through an equation of the type:

$$\begin{aligned}\text{ME (kca)} = {} & 4.0 \times \text{protein (g)} + 9.0 \times \text{fat (g)} + 3.75 \\ & \times \text{carbohydrate (g)} + \text{NME}_{\text{DF}} \\ & \times \text{dietary fiber (g)}. \end{aligned} \quad (19)$$

In the development of eqn (19), carbohydrate was determined directly (not determined by difference) and is expressed as monomeric units. Dietary fiber must also be measured directly through an approved, reproducible method that is not affected by food processing (*See* Dietary Fiber: Determination). The factors 4 (protein) and 9 (fat) are the general Atwater factors derived from mixed diets. The value of 3.75 for carbohydrate is slightly lower than the general Atwater factor of 4. This reflects the fact that the carbohydrate is expressed as the total monomeric weight.

The value of NME_{DF} (the NME for dietary fiber) comes from an analysis of the dietary fiber digestibility of mixed diets and the loss of energy to feces in mixed Western diets. This value is determined as follows. Firstly, the DE_{part} value for dietary fiber is obtained by comparing the digestible energies of mixed diets to the metabolizable energy calculated by eqn (19). The apparent digestible and metabolizable energies for carbohydrate and fat are identical since the macronutrient digestibilities were accounted for, and urine and gaseous losses of these components were negligible. Protein losses are corrected for urine nitrogen and protein digestibilities. The digestible energy for dietary fibre (DE dietary fiber) was estimated by:

$$\begin{aligned}\text{DE dietary fiber} &= \Delta H_{\text{c}} \times D_{\text{app}} \times (1 - a) \\ &= 4.1\ \text{kcal g}^{-1} \times 0.7 \times 0.7 \\ &= 2.0\ \text{kcal g}^{-1} (8.4\ \text{kJ g}^{-1}). \end{aligned} \quad (20)$$

The apparent digestibility of dietary fiber is around 0.7. The value of factor a in eqn (20) needs to be defined for dietary fiber. For sugar alcohols, the value is approximately 0.2 (see above). A value of 0.3 has been adopted for dietary fibers in humans. This value reflects the fact that the majority of energy loss estimates lie between 0.2 and 0.4 kcal per kcal of fermented energy. As already mentioned, a value of 0.2 was obtained when measuring lactitol energy utilization in human subjects. It has also been suggested that between 20 and 30% of the energy in undigested and fermented carbohydrates is utilized by bacteria in humans. A value of 0.2 kcal per kcal of fermented material has been measured in rats fed an α-amylase-resistant pea starch and a value of 0.4 kcal per kcal of fermented material has been measured in rats fed guar gum. In human diets with high intakes of dietary fiber, an increased fecal energy output of 0.3 and 0.39 kcal per kcal of energy fermented, respectively, has been measured. Overall, these data suggest that about 0.3 kcal of energy is lost to the feces per kcal of energy fermented. Thus, $1.0 - 0.3 = 0.7\ \text{kcal kcal}^{-1}$ is retained.

As indicated above, the digestible energy value does not give an accurate picture of the energy losses that occur during fermentation of undigested food components. Correction factors must be included for energy losses to the heat of fermentation (factor b), losses to fermentation gases (factor c), and the relative efficiency of utilization of short-chain fatty acids (factor d). Correcting the digestible energy for these values gives:

$$\begin{aligned}\text{NME}_{\text{DF}} = \Delta H_{\text{c}} \times D_{\text{app}} \times 0.54 &= 4.1\ \text{kcal g}^{-1} \\ \times 0.7 \times 0.54 &= 1.5\ \text{kcal g}^{-1}. \end{aligned} \quad (21)$$

Eqn (21) shows that the NME value for dietary fiber is approximately $1.5\ \text{kcal g}^{-1}$. Eqn (21) applies strictly to dietary fiber because the ΔH_{c} value and digestibilities were determined for dietary fiber. It is useful for

calculating the metabolic energy values of mixed diets because the digestibilities were determined for mixed diets. It is comparable, therefore, with the general Atwater factors of 4, 9, and 4 for carbohydrate, fat, and protein.

Empirical models represent another method for determining total energy intake. Empirical models are, in general, superior to the factorial models for a number of reasons, the most important being that the empirical models are much more robust to errors in carbohydrate, fat, and protein determinations as well as to the use of general Atwater factors. An example of an empirical equation is as follows:

$$\begin{aligned}\text{ME} = {} & 0.96 \times \text{total energy (kcal)} - 2.5\,\text{kcal g}^{-1} \\ & \times \text{dietary fiber (g)} - 12\,\text{kcal g}^{-1} \\ & \times \text{ingested nitrogen (g)}. \end{aligned} \quad (22)$$

Eqn (22) relates the metabolizable energy to total energy, protein, and dietary fiber only. The factor 0.96 represents the metabolizability of mixed diets in the absence of dietary fiber. The value of 2.5 kcal g^{-1} represents the NME energy loss associated with ingestion of dietary fiber. This value is, therefore the difference between the heat of combustion of dietary fiber and its NME_{DF} value (determined above):

$$\begin{aligned}2.5\,\text{kcal g}^{-1} = {} & \Delta H_{\text{c}} - \text{NME}_{\text{DF}} = 4.0\,\text{kcal g}^{-1} \\ & - 1.5\,\text{kcal g}^{-1}. \end{aligned} \quad (23)$$

A version of Eqn (22) was tested for bias and accuracy of prediction by comparison with 43 human diets containing varied intakes and sources of dietary fiber. The equation showed no bias over a wide range of dietary fibre intake (2–93 g daily). By necessity, this test was performed by comparing DE_{part} values because ME_{part} values cannot be measured directly for fermentable food components. This analysis provides further evidence for a general NME for dietary fiber of 1.5 kcal g^{-1}.

Overall Conclusion

Eating food components that are not digestible but that are fermentable lowers the amount of metabolizable and digestible energy available from food. This occurs because of two different processes. Fermentable food components, like dietary fiber, are not 100% fermentable, and so the NME for these components is lower than that of protein, carbohydrate, and fat. In addition, fermentation itself is associated with many different energy losses, including losses to increased bacterial mass, the heat of fermentation, fermentation gasses, as well as a lower efficiency of host utilization of short-chain fatty acids. Because the digestion of fermentable food components is so different from that of macronutrients, special equations are needed to determine the net metabolizable energy yield.

It is possible to utilize data from several sources to obtain an estimate of the relative losses that occur during fermentative digestion and back-correct to obtain an unbiased estimate of the energy value of dietary fiber. Taking into account all these energy losses, a NME value for dietary fiber of 1.5 kcal g^{-1} (6.3 kJ g^{-1}) has been calculated. This appears to accurately predict energy losses associated with mixed Western diets when analyzed using either factorial or empirical models. Thus, the NME of 1.5 kcal g^{-1} (6.3 kJ g^{-1}) can be used in a fashion similar to the general Atwater factors of 4, 9, and 4 for protein, fat, and carbohydrate to predict the metabolizable energy of mixed Western diets. Individually measured NME values for fiber in whole foods may differ from this value since dietary fiber is fermented to different extents and is intimately associated with the food matrix, which may limit macronutrient digestibility. Thus, individual measurements are needed for direct comparison and may be required for specific types of foods.

See also: **Dietary Fiber**: Properties and Sources; Determination

Further Reading

Bär A (1990) Factorial calculation model for the estimation of the physiological caloric value of polyols. In: Hosoya N (ed.) *Proceedings of the International Symposium on Caloric Evaluation of Carbohydrates*, pp. 209–257. Tokyo: Research Foundation for Sugar Metabolism.

Bernier JJ and Pascal G (1990) Valeur énergétique des polyols (sucres-alcools). *Médicine et Nutrition* XXVI: 221–238.

British Nutrition Foundation (1990) *Complex Carbohydrates in Foods. The Report of the British Nutrition Foundation's Task Force*, pp. 56–66. London: Chapman & Hall.

Calloway DH and Kretsch MJ (1978) Protein and energy utilization in men given a rural Guatemalan diet and egg formulas with and without added oat bran. *American Journal of Clinical Nutrition* 31: 1118–1126.

Cummings JH (1983) Fermentation in the human large intestine: evidence and implications for health. *Lancet* 1: 1206–1208.

FASEB (1994) *The Evaluation of the Energy of Certain Sugar Alcohols used as Food Ingredients*. Bethesda, MD: Life Sciences Research Office, Federation of American Societies for Experimental Biology (FASEB).

Goranzon H, Forsum E and Thilen M (1983) Calculation and determination of metabolizable energy in mixed diets to humans. *American Journal of Clinical Nutrition* 38: 954–963.

Livesey G (1989) Procedure for calculating the digestible and metabolizable energy values of food components making a small contribution to dietary intake. *Journal of the Science of Food and Agriculture* 48: 475–481.

Livesey G (1990) Energy values of unavailable carbohydrate and diets: an inquiry and analysis. *American Journal of Clinical Nutrition* 51: 617–637.

Livesey G (1991) Calculating the energy values of foods: Towards new empirical formulae based on diets with varied intakes of unavailable complex carbohydrates. *European Journal of Clinical Nutrition* 45: 1–12.

Livesey G (1992) The energy values of dietary fibre and sugar alcohols for man. *Nutrition Research Reviews* 5: 61–84.

Livesey G (1993) Comments on the methods used to determine the energy values of carbohydrates: dietary fibre, sugar alcohols and other bulking agents. *International Journal of Food Sciences and Nutrition* 44: 221–241.

Livesey G (1995) Metabolizable energy of macronutrients. *American Journal of Clinical Nutrition* 62(supplement): 1135S–1142S.

Livesey G and Elia M (1988) Estimation of energy expenditure, net carbohydrate utilization, and net oxidation and synthesis by indirect calorimetry: evaluation of errors with special reference to the detailed composition of fuels. *American Journal of Clinical Nutrition* 47: 608–628.

Merrill AL and Watt BK (1973) *Energy Value of Foods. Basis and Derivation. Agriculture Handbook #74*, Washington, DC: US Department of Agriculture and US Department of Health and Human Services.

Mongeau R and Brassard R (1993) Enzymatic–gravimetric determination in foods of dietary fibre as sum of insoluble and soluble fibre fractions: summary of collaborative study. *Journal of the AOAC International* 76: 923–925.

Mongeau R, Brooks SPJ, Lampi BJ and Brassard R (1997) Hard wheat bran and hard wheat bran fiber energy values measured in rats after 6 and 16 weeks. In: Kritchevsky D and Bonfield C (eds) *Dietary Fiber in Health and Disease*, pp. 267–289. New York: Plenum Press.

Nutrition Council of Holland (1987) *The Energy Value of Sugar Alcohols*. Recommendations of the Committee on Polyalcohols, The Hague, Netherlands.

van Es AJH, de Groot L and Vogt JE (1986) Energy balances of eight volunteers fed on diets supplemented with either lactitol or saccharose. *British Journal of Nutrition* 56: 545–554.

Dietary Goals *See* **Community Nutrition**; **Nutrition Policies in WHO European Member States**

DIETARY REFERENCE VALUES

J Pollard, University of Leeds, Leeds, UK

Background to the Dietary Recommendations in the UK

The need to quantify aspects of the nations' diet has occupied scientists for many years. When one considers that an adult eats about one tonne of food each year, it is easy to see why nutrition has become such a public health issue. The food we eat should provide all the nutrients we need, yet in some countries, we have an excess of food and in others insufficient. In fact, it is only relatively recently in evolutionary terms that we have the variety of food in this country available today. For example, as recently as the 1930s, it was estimated that up to one-third of households in Britain could not afford an adequate diet. It is in this context that dietary recommendations were first developed, to ensure that people were able to consume an adequate diet. During the war years, recommendations were developed with the objective of providing enough food to enable a work force to achieve and maintain its productivity. However, as the science of nutrition has developed, and the publics' awareness and understanding has increased, the objectives have changed again towards one of maintaining health. The recommended dietary intakes (RDIs) were published in 1969 and were defined as 'the amount sufficient, or more than sufficient, for the nutritional needs of practically all healthy persons in a population.' These recommendations were subsequently replaced by the recommended daily amounts (RDAs). These were published in 1979, and were defined as 'the average amount of the nutrient, which should be provided per head in a group of people if the needs of practically all members of the group are to be met.' Both of these definitions attempt to make it clear that the standards are for use in population estimates only.

In other words, consumption of nutrients at the level set would be more than adequate for most people. The problem with both these sets of recommendations was that they were misused. People using the standards did not understand the proposed use, how they had been derived, or how accurate they were intended to be. People widely interpreted these standards as minimum amounts of nutrients that everyone had to consume, rather than amounts that would prove adequate for the majority. With this, then, comes the interpretation that more is better, leading to potential overdosing of some vitamins and minerals. Most of the confusion in using the standards lies in the difference between the terms 'recommendations' and 'requirements.' People who did not understand the previous standards used them to assess requirements. A requirement is the amount of a particular nutrient that an individual must consume in order to avoid deficiency. Deficiency may be defined by clinical, physiological, biochemical, or sociological criteria, and each of these may give different values at an individual level for nutrient intake needs. Requirements will differ from individual to individual. What the standards actually give us are recommendations, that is the amount of a nutrient judged to be the appropriate amount to maintain health in virtually all individuals in that group.

In 1987, the Committee on Medical Aspects of Food Policy (COMA) convened a Panel to review the 1979 RDAs. The result of this was the new recommendations, published in 1991, and entitled Dietary Reference Values for Food Energy and Nutrients for the United Kingdom. These new recommendations attempted to clarify the confusion in the usage of the previous recommendations and covered more nutrients, particularly the macronutrients, than the previous reports. The Report also took the objective of maintaining health one step further to try and provide an 'optimal diet'. An optimal diet can be defined as one which 'maximises health and longevity and therefore, prevents nutrient deficiencies, reduces risks of diet-related chronic diseases, and is composed of foods that are safe and palatable'.

Dietary Reference Values – Definition and Derivation

The term 'dietary reference values' (DRVs) refers to a number of different parameters that describe the distribution of the requirements for a particular nutrient found within a population. When one considers that people differ from one another in the amounts of nutrients they need, it would be impossible to represent these differences with just one value, as the previous standards attempted to do. Therefore, when COMA designed the DRVs, they came up with a range of values for each nutrient to represent the differing needs of the population.

The Panel used different criteria for the assessment of nutritional adequacy for individual nutrients, depending on what information was available and on the physiological effects of different nutrients. Evidence came from observational studies of healthy populations, clinical studies of deficiency symptoms, balance studies, or functional tests of a nutrient. The Panel identified a key functional parameter for each nutrient that was considered to be essential for healthy life. That functional parameter was then used to measure the range of needs for that nutrient within a community. Functional parameters used for each nutrient are shown in Table 1. The DRVs are derived from a statistical association describing the variation around the average seen when making measurements of a large number of people. As there is little evidence on the distribution of requirements in a population, the Panel made the assumption that they are normally distributed in almost all cases. Hence, the average requirement for each nutrient was calculated, and the upper and lower limits of its distribution could be estimated. The DRVs are based on this distribution, as shown in Figure 1.

The average amount of a nutrient required is called the estimated average requirement (EAR). The range of normal consumption is taken as the EAR plus or minus two notional standard deviations. The EAR plus two standard deviations is called the reference nutrient intake (RNI) and is defined as "an amount of a nutrient that is enough for almost every individual, even someone who has high needs." This is the equivalent of the old RDAs. The EAR minus two standard deviations is the lower reference nutrient intake (LRNI) and is defined as "the amount of a nutrient that is enough for only the small number of people with low needs." Statistically, we know that the RNI indicates the amount of a nutrient sufficient to meet the needs of up to 97.5% of the population. Similarly, we known that only about 2.5% of the population will have needs lower than the LRNI. So, the requirements of about 95% of the population will lie between the LRNI and the RNI. It should be pointed out that the EAR is not a recommended intake for an individual. It is the estimated average requirement of a nutrient throughout a large group of people: some will need more than this, and some will need less.

For some nutrients, a reference value called the 'safe intake' was assigned. This term was used if there was not enough information to ascertain requirements. This intake level was one that was judged by the COMA Panel to be adequate for almost everyone's needs but not so large as to cause undesirable effects.

Table 1 Functional parameters used by the COMA panel to assess DRVs

Nutrient	*Functional parameter*
Energy	Energy expenditure
Essential fatty acids	Prevention of essential fatty acid deficiency
Fat	Changes from current intakes that would be expected to result in specific changes in physiological and/or health outcomes, particularly in relation to coronary heart disease
Nonstarch Polysaccharides	Healthy bowel function and stool weight
Nonmilk extrinsic sugars	Effect on dental caries
Protein	Nitrogen balance
Vitamin A	Adequate body pool size, based on the amount of vitamin A in the liver
Vitamin B_1 (thiamin)	Urinary thiamin output, thiamin loading, glucose loading, and transketolase activation
Vitamin B_2 (riboflavin)	Measurement of urinary and red cell riboflavin; the determination of erythrocyte glutathione reductase activation coefficient – a measure of riboflavin tissue saturation and long-term riboflavin status
Niacin	Observational requirements to prevent or cure pellagra
Vitamin B_6	Estimates of body pool size and blood vitamin B_6 levels in depletion and repletion of adults on controlled diets
Vitamin B_{12}	Serum vitamin B_{12} levels
Folate	Serum levels, red cell folate, and liver reserves
Vitamin C	Plasma ascorbate levels and prevention of scurvy
Calcium	Growth, calcium retention, urinary excretion, and absorption
Magnesium	Magnesium balance
Phosphorus	1 mmol P: 1 mmol Ca
Sodium	Sodium balance and balance of risks and benefits that might practically be expected to occur, given the prevailing sociocultural environment
Potassium	Amount needed for growth plus losses (fecal, urinary, via skin and hair)
Iron	Estimated gut absorption and losses
Zinc	Systemic needs plus losses (fecal, urinary, skin, hair, semen, and menstruation) and absorption
Copper	Balance studies
Selenium	Glutathione peroxidase activity
Iodine	Prevention of deficiency symptoms of goiter

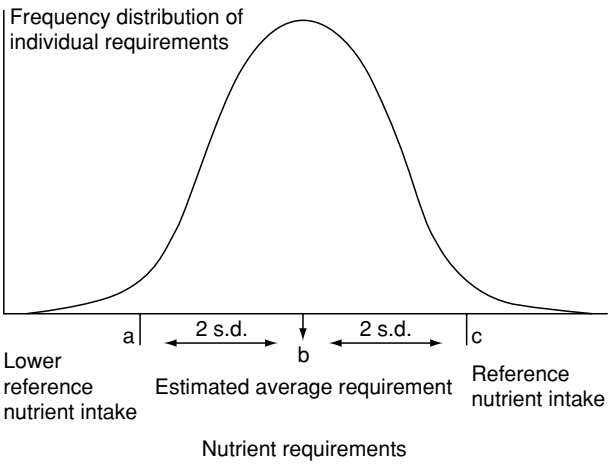

Figure 1 Dietary reference values – definition. From Department of Health (1991) *Report on Health and Social Subjects 41. Dietary Reference Values for Food Energy and Nutrients for the United Kingdom*. London: HMSO, with permission.

Energy, Protein, Vitamins, and Minerals

The RNI is designed to be adequate to meet the needs of most individuals in a population, even those with high needs. If a person with low needs consumed the RNI for most nutrients, this excess would not be harmful, but this is not the case with energy. If someone with low energy needs consumed excess, this would lead to an undesirable weight gain. Therefore, the Panel set energy standards at the average need, knowing that some people will need more and some will need less.

Safe intakes were set for pantothenic acid, biotin, vitamin E, vitamin K, molybdenum, manganese, chromium, and fluoride, as insufficient evidence was available for DRVs.

Fats, Carbohydrate and Nonstarch Polysaccharides (NSP)

As there are no specific deficiency signs or symptoms for fats and carbohydrates, and no absolute requirement for any of these individually, population averages have been derived from the likely effects that changes in their average consumption could be expected to have on certain health outcomes. These are given in the Report as percentages of the EAR for energy that they should, on average, provide. No ranges around the EAR are given.

The DRVs for NSP are based on an estimated desirable average intake, and the expected range of individual intakes around that figure.

Dietary Reference Values – Uses, Limitations, and Assumptions

The DRVs set out to reduce confusion over use, and in order to do this, they deliberately used the term 'reference' in the title to try and convey the message that these were not individual recommendations but a point of reference only. The main proposed uses of the DRVs were for food planners, health policy personnel, and the food industry, although many other people may wish to refer to the reference values.

Assessing the Diet of Individuals

DRVs, as with previous standards, are not designed for assessment of individual diets. It is not possible to know an individuals' requirement for a particular nutrient without measuring the functional parameter for that nutrient. However, the DRVs, unlike the previous standards, can tell us something about the likely adequacy of an individuals nutrient intake. If a person's consumption of a particular nutrient is somewhere between the LRNI and the RNI, it would not be possible for us to say whether or not that persons diet is adequate, as we do not know whether that person has a low, medium or high requirement. However, what we do know is that if an individual's consumption is above the RNI, they are likely to be meeting their needs, but the closer consumption gets to the LRNI, the more likely it is that that person is not meeting their needs. If a person is routinely consuming amounts of a nutrient less than the LRNI, they are likely to be deficient, but functional tests would be required to confirm this. DRVs can only give us an indication of nutritional adequacy in an individual, and great care needs to be taken when using them in this manner.

Assessing the Diet of Groups of People

When assessing the adequacy of diet in groups of people, the RNI should be used where possible to ensure that the risk of any deficiency in the group is small. When assessing the adequacy of the diet of a group of people, some assumptions have to be made. These are that the distribution of intakes in the group is identical to the distribution of requirements and that the people in the group with the highest requirements are generally those with the highest intakes, also that the measured diet is typical of habitual intake. Because of these assumptions, we are unable to say with certainty that intakes below the LRNI are inadequate, and, as with assessing the diets of individuals, measurement of functional parameters is required to identify deficiencies.

Planning Food Supplies

When planning food supplies, it is necessary to ensure that everyone (even those people with high requirements) obtains enough of every nutrient. Therefore, the RNI should be used as a reference point. Even though some people will be receiving in excess of their requirements, the risk of deficiency will be minimal.

Nutrition Labeling

The nutrition information appearing on a food label is used by individuals, and it would be inappropriate therefore to use the RNI as a reference level. If the RNI appeared on food packaging, many people might try and consume that level of the nutrient, even though they might have much lower requirements. Although this may not be harmful, it may be expensive and perhaps wasteful. Therefore, the EAR is used as a reference in nutrition labeling, and the amount of a nutrient in a product is expressed as a percentage of the EAR. However, this still needs some interpretation and is potentially misleading, as some people will still need more than the EAR and some less.

Identifying 'At Risk' Subgroups

By using the LRNI as a reference value, 'at risk' groups in the population may be identified. If people are routinely consuming less than the LRNI, they are likely to be 'at risk' of deficiency. These people then can be examined and their nutritional status assessed using the appropriate functional parameter.

There are numerous assumptions that have to be made, and therefore limitations to using DRVs. When deriving the DRVs, the COMA Panel emphasized that there was insufficient information to be able to set the RNI, EAR, and LRNI for most nutrients with any great confidence. The Panel stressed that some of the functional parameters used were far from ideal and emphasized the need for further research.

The DRVs only apply to healthy people and may not be appropriate for use with some clinical conditions or those with special requirements. DRVs are also not appropriate for use in planning nutritional requirements for intravenous therapy. It also needs to be taken into consideration that the requirement for a particular nutrient is dependent on the composition of the diet as a whole, as the presence or absence of other dietary constituents can affect the absorption and utilization of some nutrients. The DRVs for one particular nutrient therefore assume that DRVs are also being met for all other nutrients and energy.

Ultimately, the value of the DRVs when used as a frame of reference will also be affected by the quality of the intake data against which they are to be compared. The DRVs indicate how the amounts of

different nutrients should average out over time. It is not necessary for an individual to be consuming these amounts of every nutrient at every meal, or even on a daily basis. Therefore, when collecting dietary data for comparison with DRVs, the data should reflect long-term habitual intakes, whether in the individual or at the community level.

In conclusion, the DRVs are a major advance in the use of reference nutrient levels in the science of human nutrition. They build on the old methods whilst expanding the uses and relevance of recommendations. It is still necessary for people to be aware of the derivation and limitations of the DRVs, but they do go some way towards reducing the potential misuse and abuse of these values.

See also: **Carbohydrates**: Requirements and Dietary Importance; **Dietary Requirements of Adults**; **Energy**: Intake and Energy Requirements; **Fats**: Requirements; **Food Labeling (Labelling)**: Applications; **Millets**; **Protein**: Requirements; **Vitamins**: Overview

Further Reading

British Nutrition Foundation Briefing Paper 19 (1992) *Dietary Reference Values – What are They and How Should They be Used?* BNF: London

Buttriss J (2000) Nutrient requirements and optimisation of intakes. *British Medical Journal* 56(1): 18–33.

Department of Health. Report on Health and Social Subjects 41 (1991) *Dietary Reference Values for Food Energy and Nutrients for the United Kingdom*. London: HMSO.

Jackson AA (2001) Human protein requirement: policy issues. *Proceedings of the Nutrition Society* 60: 7–11.

Waterlow JC (1979) Uses of recommended intakes. The purpose of dietary recommendations. *Food Policy* 4: 107–114.

DIETARY REQUIREMENTS OF ADULTS

J Dwyer, Jean Mayer US Department of Agriculture, Human Nutrition Research Center on Aging at Tufts University, Boston, MA, USA

History and Current Status

Nutrients are substances that are not synthesized in the body in sufficient amounts to perform critical functions such as growth, reproduction, or health maintenance and therefore they are required from the diet.

Although humans have known since ancient times that they must eat to live, only recently has it been possible to specify the chemicals and amounts involved with certainty. The rise of this knowledge has led to the growth of nutrition as a science. In the nineteenth and early twentieth centuries, the concepts of energy metabolism and essential nutrients were formulated, and techniques for quantifying nutrient requirements were developed. Experimental and observational work determined the amounts of nutrients required to prevent frank dietary deficiency disease. During the middle and later twentieth century, attention turned to nutrient functions associated with chronic degenerative diseases and additional roles for nutrients. This chapter describes the recent work on dietary requirements as they have been formulated by the Health Canada/Food and Nutrition Board, National Academy of Sciences in the USA in the dietary reference intakes. These deliberations and parallel efforts by expert groups of the World Health Organization/Food and Agricultural Organization of the United Nations and many national or regional governments occupied most of the 1990s. They will conclude early in the twenty-first century.

The landmark Dietary Reference Intake project arose because much new work on nutrient needs, the balance and interactions between nutrients, and the development of chronic degenerative diseases had become available by the early 1990s. New techniques for evaluating the biological functions of nutrients were present. In addition, many nonnutrients in food with biological activities of health significance were being described. The need for upgrading data on the validity and reliability of existing dietary standards was evident. Moreover, new statistical techniques for adjusting nutrient intake data for assessment and planning purposes had recently become available, and these needed to be integrated into existing concepts. Finally, it had become apparent that, in the past, some uses of dietary requirements and nutrient standards had been inappropriate, and new guidance was needed.

Estimates of nutrient requirements are now well delineated for healthy adults, but they are still not so for special groups, such as very premature infants, the very old, and the ill. This remains a challenge for the future.

What Nutrients are Required?

All human beings require the energy-providing nutrients (amino acids, fatty acids and fat, glucose, and other carbohydrates), vitamins, minerals, and water. The specific essential nutrients include organic compounds such as nine essential amino acids (leucine, isoleucine, histidine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine). Cystine, which can replace some of the methionine requirement, and tyrosine, which can replace some of the phenylalanine requirement, are also sometimes considered to be essential. Taurine is thought to be essential in premature infants. There are two essential fatty acids (linoleic and arachadonic), four fat-soluble vitamins (A, D, E, and K), and 10 water-soluble vitamins (niacin, thiamin, riboflavin, folate, vitamin B_6, vitamin B_{12}, biotin, pantothenic acid, ascorbic acid, and choline). A source of glucose is also required, but this can be synthesized from glycerol and glycogenic of the amino acids if they are present in sufficient amounts. Among the inorganic substances, three minerals (calcium, phosphorus, and magnesium), nine trace mineral elements (iron, iodine, selenium, copper, zinc, manganese, fluoride, chromium, and molybdenum), three electrolytes (sodium, chlorine, and potassium), and water are required.

There are many other substances that are naturally present in foods and are known to be required in the diets of other species. For example, several ultratrace elements are essential in experimental animals and may also be essential in humans (arsenic, nickel, silicon, boron, cadmium, lead, lithium, tin, vanadium, and cobalt), but at present, evidence is lacking that they are. Dietary requirements, if any, for many other compounds, the so-called candidate nutrients, are also uncertain, and they are currently being explored. They include carnitine, myoinositol, cholesterol, the nonprovitamin A caratenoids, and dietary fibers. In addition to these substances that appear to be essential for some higher animals but not for humans, scores of other organic phytochemical and zoochemical compounds are present in foods that have various health effects, and some of them eventually may prove to be essential. These include growth factors and coenzymes such as dietary nucleotides, coenzyme Q, lipoic acid, and many others.

How Much is Required?

The amount of a nutrient required depends on the health criterion or indicator of adequacy that is used. For most nutrients, multiple possible criteria exist that are based on various health parameters. Levels of nutrients that are adequate to meet one of these criteria, such as prevention of night blindness, may not be so for reaching another, such as repletion of liver stores of vitamin A. The types of evidence and the criteria used to establish the quantitative requirements for a specific compound vary from nutrient to nutrient. They also vary from one age or physiological stage to another.

The amounts of nutrients that are needed, and sometimes the types of the essential nutrients that must be provided, differ depending on age and physiological state. For example, taurine is an amino acid that may be essential in premature infants but is not so in older infants or adults. Also, the forms in which nutrients occur in foods, fortified foods, and supplements affect their bioavailabiliy and thus their requirements.

How are Nutrient Requirements Expressed?

The quantitative recommendations for nutrient intakes used as reference values or standards for planning and evaluating the nutrient intakes of healthy people are collectively referred to as the dietary reference intakes, or DRIs. The DRIs include the estimated average requirement for nutrients (EAR) and three other reference values: the recommended dietary allowances, the adequate intake, and the tolerable upper level.

The process for establishing estimated average requirements and other dietary reference intakes used in the USA and Canada is summarized below. It is described in detail in monographs on groups of nutrients published by the Standing Committee on the Scientific Evaluation of Dietary Reference Intakes of the Food and Nutrition Board, Institute of Medicine, National Academy of Sciences. They replace both the 1989 Recommended Dietary Allowances, the single reference values used in the USA, and previous reference data used by Canada in the past.

The DRI paradigm has now been used to develop current estimates of nutrient requirements for vitamins and minerals. Requirements for energy, macronutrients, water, and other compounds of possible health significance are now being considered, but as of this writing, no paradigm has been published for these dietary constituents.

Estimated Average Requirement

The estimated average requirement (EAR) is the amount of a nutrient that is estimated to meet the requirement for a specific criterion of adequacy of half of the healthy individuals of a specific age, sex, and life-stage. In setting the EAR, the evidence for

each possible criterion is considered, and the reason for selecting the criterion that is finally chosen is justified. The amount of the nutrient necessary to meet the appropriate criterion of adequacy varies from one individual to the next, but the data are usually distributed normally or can be transformed to achieve a normal distribution. The EAR is not useful as an estimate of nutrient adequacy in individuals, because it is a mean requirement for a group, and the variation around this number is considerable. At the EAR, 50% of the individuals in a group are below their requirement, and 50% are above it. Thus, a person whose usual intake is at the EAR has a 50% risk of an inadequate intake during the reporting period. An individual with an intake between the RDA and the EAR would have a risk of inadequacy between 50 and 2–3%. An individual with a usual intake below the EAR would have a risk of inadequacy between 50 and 100%. This is because the EAR is derived from a group estimate. The precise amount of a nutrient that will be adequate for any given individual is therefore unknown. It can be stated only in terms of probabilities, and thus it is rarely used in clinical practice. For healthy individuals whose usual nutrient intakes are accurately described, the EAR can be used to assess the approximate probability of inadequacy, although the range of error in the estimate is considerable.

Recommended Dietary Allowances

The recommended dietary allowance (RDA) is the average daily dietary intake level that suffices to meet the nutrient requirements of nearly all (97–98%) healthy persons of a specific sex, age, life stage, or physiological condition (such as pregnancy or lactation). The RDA is a nutrient intake goal for planning the diets of individuals.

To ensure that the needs of any given individual are met, a measure of variability around the EAR, usually the standard deviation or, if data are insufficient, an estimated coefficient of variation (CV) of 10 or 15%, is assumed. Then, an amount that covers this variation is calculated. The RDA is therefore defined statistically as two standard deviations above the EAR, assuming that the distribution of nutrient requirements is normal, as it may be for most nutrients that have been studied. That is, the RDA is the EAR + 2 SD. If the SD is not available, using a CV of 10%, the RDA is set at 1.2 times the EAR. If the CV is assumed to be 15%, the RDA is set at 1.3 times the EAR.

The risk, but not the certainty, of inadequacy increases as intakes fall further and further below the RDA. However, the RDA is an overly generous criterion for evaluating nutrient adequacy. By definition, the RDA exceeds the actual requirements of all but about 2–3% of the population. Therefore, many individuals who are below the RDA may still be getting enough of the nutrient in question to be above their requirement level.

The 1989 RDA consisted of a single reference point, with a focus on achieving sufficiency only, for individuals. However, the reference was often mistakenly applied as a standard for the evaluation of population nutrient intakes. In fact, the RDA is not appropriate for assessing the diets of individuals, or for assessing or planning the diets of groups.

Adequate Intake

For some nutrients, it is not possible to set RDA because data on the requirement (EAR) for the function of health significance that is judged to be the most important are not available. However, enough information is available to make some helpful statements about healthful nutrient intake levels. An adequate intake (AI) is set when experts believe that sufficient data on the requirement using the criterion that has been chosen to determine the requirement are still preliminary. More research is needed before an EAR and an RDA for that particular criterion or function can be determined. The AI is also an appropriate goal for the nutrient intake of individuals.

The AI is a value, based on observed or experimentally determined approximations of nutrient intakes by groups of healthy people, that is used when an EAR and an RDA cannot be determined. For infants, the AI is always based on the mean intakes of groups of healthy infants. For adults, in some instances, the AI is set as the mean of diets of healthy individuals in some reference groups. For other nutrients, the criteria are less precisely determined but always chosen to be generous enough for good health. In the DRIs that have been established to date, AIs rather than RDAs are proposed for all nutrients for infants up to age 1, and for calcium, vitamin D, fluoride, pantothenic acid, biotin, choline, vitamin K, chromium, and manganese) for persons of all ages. The AI is *not* a requirement, and it should not be interpreted as such. Rather, it is an average or median intake in a group of healthy people, all of whom are assumed to be meeting their nutrient requirements.

Tolerable Upper Intake Levels of Nutrients (UL)

Until recently, dietary requirements for nutrients were thought to consist solely of assuring adequacy. Now,

it is recognized that avoiding excess is also important. Intakes of nutrients that are far in excess of the RDA can disturb body functions and cause acute, progressive, or permanent disability. Some examples include fluorosis, hypervitaminosis A, and hypervitaminosis D. The DRIs now also address risks of possible excess of inadequate nutrient intakes. The tolerable upper intake level (UL) of a nutrient is the highest level of chronic, usual daily nutrient intake that is likely to pose no risk of adverse health effects to almost everyone in the population. Below the UL, individuals should be able to biologically tolerate this amount of the nutrient. Excesses of each nutrient may have many possible adverse effects. The UL is set for the effects that experts believe are most likely to be prejudicial for health and for which sufficient data are available. The UL is set using qualitative and quantitative evidence and inference-related judgments of experts in a logical process that assesses the associations between nutrient ingestion and the probability of adverse effects. The UL is set by determining the levels at which no observed adverse effects are noted, or the lowest level of intake associated with observed adverse effects. An uncertainty factor is then applied to insure that even very sensitive persons would not experience adverse effects at the UL dose chosen. For many nutrients, data on the adverse effects of large amounts of nutrients are simply unavailable, or they are so limited that a UL cannot be determined. The lack of a UL does *not* mean that the risk of adverse effects is nonexistent from high intakes; it simply means that data are not yet available, and the verdict is unknown. Therefore, caution is warranted in consuming high amounts of nutrients. The UL is *not* intended to be a recommended level of intake. There is no established benefit for healthy individuals of consuming nutrient levels above the RDA or AI. The amounts of individual foods that most people eat rarely reach levels that are likely to exceed the UL. In contrast, nutrient supplements provide more concentrated amounts of nutrients per dose, and thus the potential risk of excess is somewhat greater with them.

Current Estimates of Nutrient Requirements for Vitamins and Minerals

The EAR, RDA, AI, and UL all have specific uses in dietary assessment, for making nutritional recommendations and for dietary planning for individuals. The EARs for selected nutrients and age groups are provided in Table 1. Note that, for all of these nutrients, the estimated average requirements are based on oral intakes of nutrients. The appropriate use of the reference values is described in detail in the various publications of the Food and Nutrition Board.

Water

For adults under usual conditions, 1–1.5 ml of water per kcal energy expenditure is sufficient to allow for normal variations in physical activity levels, sweating, and solute loads of the diet. If renal function is normal, and solute intakes are adequate, the kidneys adjust to increased intakes of water by excreting the excess. If water intakes are inadequate to cover obligatory urine outputs, hydration status can be compromised. Obligatory water losses include 50–100 ml per day in the feces, 500–1000 ml by evaporation or exhalation, and, depending on the renal solute load, 1000 ml or more in the urine. If external losses increase, intakes must increase accordingly to avoid dehydration.

The stage of life and health influence water requirements. Pregnancy increases water needs by perhaps 30 ml per day, but during lactation, milk production increases requirements to at least 1 ml for each milliliter of milk produced or to 1 l per day or more, depending on milk production. In addition, water deficits may arise if disease causes increased losses, damages the kidneys, requires the use of diuretics, or, as in coma, makes it impossible for the individual to communicate their thirst. Fever increases water losses by approximately 200 ml per degree Celsius, and diarrhea losses vary but may be as great as 5 l per day in severe diarrhea. Vomiting and heavy sweating, especially in hot, dry climates and at altitude, also increase water losses substantially.

Energy

Recommendations on energy and the macronutrients are expected from the DRI process in 2002. Energy requirements or inputs must match energy outputs for weight to remain stable. The major sources of energy output are resting energy expenditure (REE) and physical activity. Minor sources include the energy cost of metabolizing food (thermic effect of food or specific dynamic action) and shivering thermogenesis (e.g., cold-induced thermogenesis). Because energy needs are so rarely measured directly, they are estimated in adults by assuming that if an individual is in energy balance and their body composition is not changing, they are in energy balance. To estimate the energy needs of an individual whose weight is stable, the formulas for estimating resting metabolism devised by Owen and colleagues are useful. First, an estimate of REE is obtained to provide a rough approximation of energy needs at rest in health: Male $\text{REE} = 900 + 10$ (weight in kg) or female $\text{REE} = 700 + 7$ (weight in kg). The result is then adjusted for physical activity level by multiplying it by 1.2 for the very sedentary, 1.4 for moderately active,

Table 1 Dietary reference intakes (DRIs): Estimated average requirements[a]

Life stage group	*Vitamin A (μg per day)*[b]	*Vitamin C (mg per day)*	*Vitamin E (mg per day)*[c]	*Thiamin (mg per day)*	*Riboflavin (mg per day)*	*Niacin (mg per day)*[d]	*Vitamin B_6 (mg per day)*	*Folate (μg per day)*[e]	*Vitamin B_{12} (μg per day)*	*Copper (μg per day)*	*Iodine (μg per day)*	*Iron (mg per day)*	*Magnesium (mg per day)*	*Molybdenum (μg per day)*	*Phosphorus (mg per day)*	*Selenuim (μg per day)*	*Zinc (mg per day)*
Infants																	
7–12 months												6.9					2.5
Children																	
1–3 years	210	13	5	0.4	0.4	5	0.4	120	0.7	260	65	3.0	65	13	380	17	2.5
4–8 years	275	22	6	0.5	0.5	6	0.5	160	1.0	340	65	4.1	110	17	405	23	4.0
Males																	
9–13 years	445	39	9	0.7	0.8	9	0.8	250	1.5	540	73	5.9	200	26	1055	35	7.0
14–18 years	630	63	12	1.0	1.1	12	1.1	330	2.0	685	95	7.7	340	33	1055	45	8.5
19–30 years	625	75	12	1.0	1.1	12	1.1	320	2.0	700	95	6	330	34	580	45	9.4
31–50 years	625	75	12	1.0	1.1	12	1.1	320	2.0	700	95	6	350	34	580	45	9.4
51–70 years	625	75	12	1.0	1.1	12	1.4	320	2.0	700	95	6	350	34	580	45	9.4
> 70 years	625	75	12	1.0	1.1	12	1.4	320	2.0	700	95	6	350	34	580	45	9.4
Females																	
9–13 years	420	39	9	0.7	0.8	9	0.8	250	1.5	540	73	5.7	200	26	1055	35	7.0
14–18 years	485	56	12	0.9	0.9	11	1.0	330	2.0	685	95	7.9	300	33	1055	45	7.3
19–30 years	500	60	12	0.9	0.9	11	1.1	320	2.0	700	95	8.1	255	34	580	45	6.8
31–50 years	500	60	12	0.9	0.9	11	1.1	320	2.0	700	95	8.1	265	34	580	45	6.8
51–70 years	500	60	12	0.9	0.9	11	1.3	320	2.0	700	95	5	265	34	580	45	6.8
> 70 years	500	60	12	0.9	0.9	11	1.3	320	2.0	700	95	5	265	34	580	45	6.8
Pregnancy																	
≤ 18 years	530	66	12	1.2	1.2	14	1.6	520	2.2	785	160	23	335	40	1055	49	10.5
19–30 years	550	70	12	1.2	1.2	14	1.6	520	2.2	800	160	22	290	40	580	49	9.5
31–50 years	550	70	12	1.2	1.2	14	1.6	520	2.2	800	160	22	300	40	580	49	9.5
Lactation																	
≤ 18 years	880	96	16	1.2	1.3	13	1.7	450	2.4	985	209	7	300	35	1055	59	11.6
19–30 years	900	100	16	1.2	1.3	13	1.7	450	2.4	1,000	209	6.5	255	36	580	59	10.4
31–50 years	900	100	16	1.2	1.3	13	1.7	450	2.4	1,000	209	6.5	265	36	580	59	10.4

[a]This table presents estimated average requirements (EARs), which serve two purposes: for assessing adequacy of population intakes and as the basis for calculating recommended dietary allowances (RDAs) for individuals for those nutrients. EARs have not been established for vitamin D, vitamin K, pantothenic acid, biotin, choline, calcium, chromium, fluoride, manganese, or other nutrients not yet evaluated via the DRI process.

[b]As retinol activity equivalents (RAEs). 1 RAE = 1 μg of retinol, 12 μg of θ-carotene, 24 μg of l-carotene, or 24 μg θ-cryptoxanthin. The RAE for dietary provitamin A carotenoids is twofold greater than retinol equivalents (RE), whereas the RAE for preformed vitamin A is the same as RE.

[c]As α-tocopherol. α-Tocopherol includes *RRR*-α-tocopherol, the only form of α-tocopherol that occurs naturally in foods, and the 2*R*-stereoisomeric forms of α-tocopherol (*RRR*-, *RSR*-, *RRS*-, and *RSS*-α-tocopherol) that occur in fortified foods and supplements. It does not include the 2*S*-stereoisomeric forms of α-tocopherol (*SRR*-, *SSR*-, *SRS*-, and *SSS*-α-tocopherol), also found in fortified foods and supplements.

[d]As niacin equivalents. 1 mg of niacin = 60 mg of tryptophan.

[e]As dietary folate equivalents (DFE). 1 DFE = 1 μg of food folate = 0.6 μg of folic acid from fortified food or as a supplement consumed with food = 0.5 μg of a supplement taken on an empty stomach.

or 1.8 for very active individuals. More precise estimates of energy expended in physical activity may be used if they are available. Taken together, the product of REE times the activity factor provides an estimate of total caloric needs in a state of energy balance.

Energy needs decrease with declines in lean body mass, the major determinant of resting energy expenditure in health, and with physical activity. Because both actively metabolizing tissue and physical activity usually decline with age, energy needs in older persons, especially those over 70, are less than those of younger persons; in contrast, energy needs rise with pregnancy, because of the demands of fetal growth, in lactation, because of the energy lost required for milk production, and in growth and rehabilitation.

Macronutrients: Protein, Carbohydrate (Including Fiber), and Fat

Requirements for the macronutrients have been reviewed most recently in publications of the Food and Agricultural Organization/World Health Organization (FAO/WHO), and the update by the DRI Committee is expected shortly.

Factors Affecting Nutrient Requirements

Nutrient requirements are potentially affected by age, sex, rate of growth, pregnancy, lactation, and physical activity levels. They are also influenced by route of administration (e.g., by mouth or vein), composition of the diet with respect to nutrients and other constituents that may affect absorption, utilization or excretion, coexisting diseases, and drugs. The specific factors vary, and every nutrient must be assessed individually.

Assessing Dietary Adequacy Using Information on Nutrient Requirements

Dietary assessment determines whether intake is adequate while not being excessive. The first prerequisite for doing this is to obtain a representative estimate of usual dietary intake. Procedures for assessment of individuals and groups vary; a report on assessing dietary intakes is available. The Food and Nutrition Board is also completing a report on the uses of the DRI in planning for individuals and groups.

See also: **Carbohydrates**: Requirements and Dietary Importance; **Dietary Fiber**: Properties and Sources; **Dietary Reference Values**; **Energy**: Intake and Energy Requirements; **Fats**: Requirements; **Protein**: Requirements; **Water**: Physiology

Further Reading

Beaton GH (1999) Recommended dietary intakes: Individuals and populations. In: Shils ME, Olson JA, Shike M and Ross AC (eds) *Modern Nutrition in Health and Disease*, 9th edn., pp. 1705–1726. Baltimore, MD: Williams & Wilkins.

De Onis M, Clugston GA and Underwood B (1998) Primary malnutrition. In: Sadler MJ, Strain JJ and Caballero B (eds) *Encyclopedia of Human Nutrition*, pp. 1246–1254. San Diego, CA: Academic Press.

Dwyer JT (1999) Dietary assessment. In: Shils ME, Olson JA, Shike M and Ross AC (eds) *Modern Nutrition in Health and Disease*, 9th edn., pp. 937–962. Baltimore, MD: Williams & Wilkins.

Food and Nutrition Board, Commission on Life Sciences, National Research Council (1989) *Recommended Dietary Allowances*, 10th edn. Washington, DC: National Academy Press.

Harper AE (1999) Defining the essentiality of nutrients. In: Shils ME, Olson JA, Shike M and Ross AC (eds) *Modern Nutrition in Health and Disease*, 9th edn. Baltimore, MD: Williams & Wilkins.

Owen OE, Holup JL, D'Alessio DA *et al*. (1987) A reappraisal of the caloric requirements of men. *American Journal of Clinical Nutrition* 46: 875–885.

Owen OE, Kavle E, Owen RS *et al*. (1986) A reappraisal of caloric requirements in healthy women. *American Journal of Clinical Nutrition* 44: 1–19.

Panel on Dietary Antioxidants and Related Compound, Subcommittees on Upper Reference Levels of Nutrients and Interpretation and Uses of Dietary Reference Intakes, and the Standing Committee on the Scientific Evaluation of Dietary Reference Intakes (2000) *Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids*. Washington, DC: National Academy Press.

Panel on Micronutrients, Subcommittee on Upper Reference Levels of Nutrients and Interpretation and Uses of Dietary Reference Intakes, and Standing Committee on the Scientific Evaluation of Dietary Reference Intakes (2001) *Dietary Reference Intakes for Vitamins A, K, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Vanadium, and Zinc*. Washington, DC: National Academy Press.

Standing Committee on the Scientific Evaluation of Dietary Reference Intakes and its Panel on Folate, Other B Vitamins and Choline, Food and Nutrition Board, Institute of Medicine (1998) *Dietary Reference Intakes for Thiamin, Ribroflavin, Niacin, Vitamin B_6, Folate, Vitamin B_{12}, Pantothenic Acid, Biotin, and Choline*. Washington, DC: National Academy Press.

Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (1997) *Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D and Fluoride*. Washington, DC: National Academy Press.

Subcommittees on Interpretation and Uses of Dietary Reference Intakes and Upper Reference Levels of Nutrients, and the Standing Committee on the Scientific Evaluation of Dietary Reference Intakes (2000) *Dietary Reference Intakes: Applications in Dietary Assessment*. Washington, DC: National Academy Press.

DIETARY SURVEYS

Contents

Measurement of Food Intake

M Nelson, King's College London, London, UK

Background

Measuring food consumption is a complex and difficult task. It may be carried out for a variety of reasons. For example:

- to determine the adequacy of food supply available to an entire country;
- to determine the adequacy of nutrient intake in a selected group of individuals;
- to establish levels of food self-sufficiency in the development of agricultural policy;
- to examine the relationship between diet and health in groups or individuals;
- as part of fundamental research on nutrient metabolism;

and for many other reasons.

It is fair to say that there is *no* method available that provides an entirely accurate measure of food consumption or nutrient intake, at whatever level the measurement is being made. It is impossible at the national or regional level to measure all food production completely, and to take into account all food and production losses and wastage; at the household level, people alter their purchasing or food-use habits while being observed; and at the individual level, people change their eating habits during a survey, or fail to describe their eating habits accurately. Awareness of the likely sources of error, however, allows for a sensible interpretation of results from studies of dietary assessment.

The type of measurement made, whether at the national, regional, household, or individual level, will be dictated by the purpose of the study being undertaken. This article identifies the points in the food chain at which measurements of food consumption can be usefully made, and highlights the strengths and weaknesses of the techniques available. It also considers the validity and reliability of these techniques when trying to relate diet and disease.

Methods for Measuring Food Consumption

Figure 1 shows the points in the food chain at which food consumption is commonly measured. Point I is at the level of domestic food production. In virtually all countries, a Department or Ministry of Agriculture requires food producers to report the amount of food produced. This information is useful in establishing levels of food self-sufficiency, but cannot be used alone to estimate the average consumption in the population.

Measurements at Point II are used by the Food and Agricultural Organization (FAO) of the United Nations to establish food-balance sheets. These measurements reflect domestic production, plus food imports and food taken from stocks, minus food exports, food moved into stocks, and food used for animal feed or non-food purposes (e.g., sugar used in the brewing industry). They can be used to show the average amount of food available per person in the population.

The next point of food measurement is the household (Point III). Typically, the weights and costs (or imputed value) of all food acquisitions (purchases, gifts, food from gardens and allotments, payments in kind, etc.) are recorded over a set period ranging from 1 week to 1 year. In addition, an inventory of the larder contents may be carried out at the beginning and end of the survey period in order to determine changes in food stocks, and allowances made for estimated food wastage, to yield an estimate of the amount of food available for consumption in the household (Point IV). Alternatively, respondents may be interviewed and asked to recall their purchases and food use over the previous week or two weeks, usually having been forewarned and asked to keep receipts and other informal records of food use and acquisition such as packaging and labels. Household surveys may include records of food purchased and eaten away from home, and household members' consumption of items not directly under the purview of the respondent, such as alcoholic beverages, sweets and soft drinks.

The final point of measurement (Point V) is at the level of the individual. Briefly, there are two

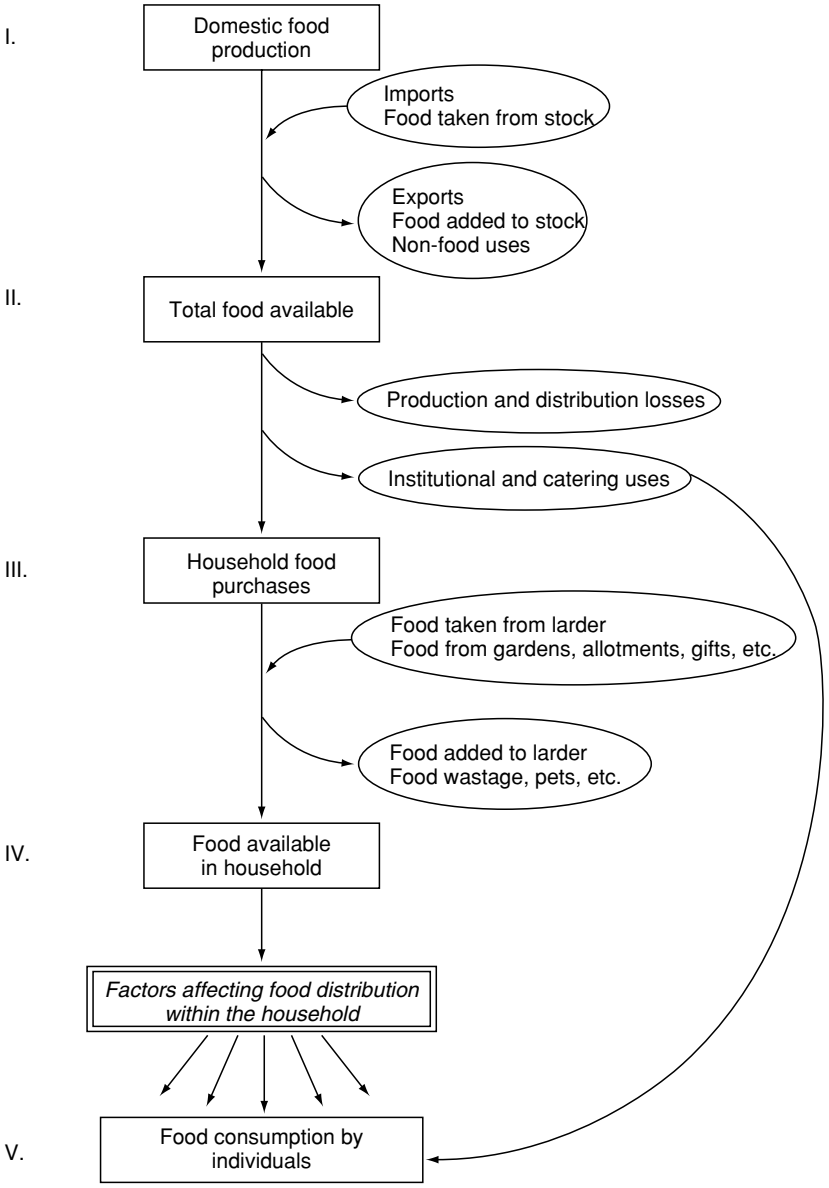

Figure 1 Points in the food supply chain at which it is convenient to measure food availability or consumption. Reproduced from Garrow J, James WPT and Ralph A (eds) *Human Nutrition and Dietetics*, 10th edn. London: Churchill Livingstone, wtih permission.

approaches, *prospective* and *retrospective*. Prospective methods include:

- *Duplicate diet*. The subject collects an exact duplicate or fixed aliquot (e.g., 10% by weight) of all food and drink actually consumed. The aliquot technique is especially well suited to households or institutions. A composite diet may be assembled in which representative local foods are combined into a single sample in proportion to the foods consumed (assessed using other techniques such as household food purchases or weighed inventories). The sample is then analyzed chemically for its nutrient content.
- *Weighed inventory or weighed record*. The subject records a detailed description and measured weight of all food and drink consumed.
- *Household measures*. The subject keeps a record of all food and drink consumed. Amounts are recorded in units of common household measures such as cups, spoons, bowls, etc.
- *Food check list*. The subject ticks a box on a preprinted list of food and drink to indicate each time a named item is consumed. Foods or drinks that do not appear on the list are recorded separately.

In general, the longer the period over which information on diet is collected, the greater will be the

Table 1 Appropriate uses of dietary survey methods

	Dietary survey method							
	Food balance sheets	*Household surveys*	*Surveys of individuals*					
			Prospective			*Retrospective*		
			Duplicate diet[a]	*Weighed inventory*	*Household measures*	*Diet history*	*24-h recall*	*Questionnaires*
Level of dietary measurement								
National	+++	+++	+[b]	+[b]	+[b]	+[b]	+[b]	+[b]
Regional	–	+++	+[b]	+[b]	+[b]	+[b]	+[b]	+[b]
Institution/group	–	+++	+[b]	++[b]	++[b]	++[b]	++[b]	++[b]
Household	–	++[c]	++[d]	++[d]	++[d,e]	++[d]	++[d]	++[d]
Individual	–	+[f]	+++[g]	+++[g]	+++[g]	+++[g]	+++[g]	+++[g]
Type of study								
Epidemiological	+++[h]	+++[h]	+[i]	+++[i]	+++[i]	++[i]	++[i]	+++[i]
Clinical	–	–	++	+++	+++	+++	++[j]	+++
Metabolic	–	–	+++	+[k]	–	–	–	–

[a]Includes other techniques of direct analysis (see text).
[b]Requires sample representative of population, institution or group, or analysis weighted to reflect balance of subgroups.
[c]Requires larder inventory. Short-term measures (e.g., 1 week) may not reflect usual diet in individual households.
[d]The need for data from all household members may distort usual household food consumption patterns.
[e]Semiweighed method (Nelson M and Nettleton PA (1980) Dietary survey methods. I. A semi-weighted technique for measuring dietary intake within families. *Journal of Human Nutrition* 34: 325–348.) reduces total respondent burden within household.
[f]Requires complex mathematical modeling of within-household food and nutrient distribution (Chesher A (1997) Diet revealed?: Semi-parametric estimation of nutrient intake age relationships. *Journal of the Royal Statistical Society, Series A – Statistics in Society* 160(3): 389–420.).
[g]Important to screen out individuals whose responses may not be valid (see text).
[h]Appropriate for ecological studies (see Hiller JE and McMichael AJ (1997) Ecological studies. In: Margetts BM and Nelson M (eds) *Design Concepts in Nutritional Epidemiology*, 2nd edn. Oxford: Oxford University Press.).
[i]See Margetts B and Nelson M (1997) *Design Concepts in Nutritional Epidemiology*, 2nd edn. Oxford: Oxford University. For a detailed discussion of use of dietary survey methods in epidemiological studies.
[j]Requires repeat 24-h recalls for valid classification of subjects according to levels of intake (see text).
[k]Useful only if range of foods is of limited variation in composition, allowing reliable use of food composition tables.
+++, very suitable; ++, moderately suitable; +, limited application; –, not suitable.

precision in the estimate of an individual's food or nutrient intake. However, the intake of some nutrients (like vitamin C) varies more from day to day than the intake of other nutrients (sugar, for example). The greater the day-to-day variation in intake, the longer the period of recording that will be needed to achieve a given level of precision.

Retrospective methods include:

- *Diet history*. The subject is asked by a trained interviewer about usual consumption habits in an extended interview lasting 1–2 h.
- *24-h recall*. A trained interviewer asks quantitatively about consumption of food and drink over the previous 24 h.
- *Food-frequency questionnaires (FFQ)*. These are completed either by the respondent or with the help of an interviewer. FFQs typically ask about the usual frequency and quantity of consumption of a wide range of food and drink. Alternatively, they may focus on the consumption of a limited range of items that are the main sources of intake of a particular nutrient. Some FFQs ask about frequency alone without assessing portion size.

Table 1 summarizes the principal applications of these different techniques. Food-balance sheets are appropriate for estimating national consumption levels, and are therefore of use in epidemiological studies comparing diets between countries. They cannot be used for assessing consumption at the regional, household, or individual level. Household surveys lend themselves to a wide variety of uses. If, for example, a nationally representative sample of households is used, it will be possible to estimate consumption levels nationally as well as in regional, socio-economic, and other subgroups within the population. Some household surveys include larder inventories at the beginning and end of the survey period, which allow estimates of consumption at the level of single households. Techniques are available to estimate the distribution of food and nutrient to individuals of similar age and sex within groups of households, and household surveys may therefore have application in some epidemiological surveys where a knowledge of the *pattern* of consumption to individuals, but not individual intakes *per se*, is required.

Duplicate diet techniques are most suitable for metabolic studies of individuals, where are a precise knowledge of nutrient intake is required. The

Table 2 Main attributes of the different dietary survey methods for measuring food consumption or nutrient intake in human populations

Type of survey	*Strengths*	*Weaknesses*[a]
Food-balance sheets	Determine national food consumption	National only
	Compare diets between countries	May overestimate food availability, especially in more affluent countries
Household budget surveys	Relatively cheap to obtain	Home food only
	Subgroupanalysis possible (by region, income, household composition, etc.)	Usually excludes meals eaten away from home, sweets, soft drinks, alcohol
	Useful where individual data hard to obtain	No individual data
	Time trends if collected annually	May be expenditure only, so weights of foods must be imputed
Without larder inventory	Lower respondent burden	Bias towards overpurchasing, especially in low income and elderly households
		Food acquisition, not consumption
With larder inventory	Consumption data	Distorts normal food purchasing pattern
		Higher respondent burden
Prospective surveys of individuals	Current diet	Labor-intensive
	Direct observation of diet	Requires literacy and numeracy skills
	Vary duration of survey according to needs	High motivation needed
		Distortion and/or omission due to 'health' issues
		Overweight subjects more likely to underreport true consumption levels
Duplicate diet	Direct analysis of nutrient content of diet (no food tables needed)	Very expensive
		Intensive supervision required
		Likely to distort usual diet
Weighed inventory	Widely used, facilitates comparisons between studies or groups	Loss of precision of estimate of nutrient intake compared with duplicate diet
Household measures	No scales needed	Loss of precision compared with weighed inventory
Retrospective surveys of individuals	Inexpensive	Bias: memory, conceptualization, interviewer
	Quick	Daily variation not usually assessed
	Low subject motivation needed	Depends on regular eating habits
	Good cooperation	
	Current or past diet	
Diet history	Assesses 'usual' diet	Tends to overestimate nutrient intake
24-h recall	Very quick	Tends to underestimate nutrient intake
	Repeat observations to obtain daily variation	Does not reflect individual intake unless repeated
Questionnaires	Suitable for very large number of subjects	Requires validation
	Can be posted	Literacy and numeracy skills needed if self-completed
	Focus on particular nutrients	

[a]With the exception of the duplicate diet technique, all methods rely on food composition tables for estimating nutrient intake.

composite diet technique is particularly suitable for estimating the nutrient content of foods eaten by a group of people where food composition tables (q.v.) are likely to be inadequate. The weighed inventory is suitable for assessment of individual food consumption or nutrient intake (using food-composition tables), and thus lends itself to a wide variety of uses, in that accurate estimates of dietary intake for groups of people can be built up from individual data. Where a limited range of foods of accurately known composition is being consumed, the weighed inventory may also be used in metabolic studies. Because the household-measures technique, like the weighed inventory, focuses on the individual as the unit of assessment, it can also be used in a wide variety of studies, but with some loss of precision in comparison with weighed inventories.

The retrospective techniques, because they rely on memory, are regarded as being less objective than the prospective techniques, but they may free the respondent of factors likely to distort the recording of diet (see Table 2). The diet history, particularly, may be used to assess the 'usual' diet of an individual more effectively than other techniques, in that seasonal variations in diet can be taken into account more readily. A single 24-h recall from one individual cannot be used to assess that individual's diet because of day-to-day variations in intake. Single 24-h recalls collected from a large number of subjects, however, may be useful for estimating the intake of a group.

Repeat 24-h recalls collected from an individual can provide an accurate assessment of that individual's diet. Questionnaires are of particular value in epidemiological studies, in that they can provide information economically from a large number of subjects.

Strengths and Weaknesses

Table 2 summarizes the main strengths and weaknesses of the methods available. When measuring diet, it is important that the method chosen is accurate enough to enable any hypothesis to be tested adequately. It is likely that many studies that fail to show a relationship between diet and disease do so, not because no relationship exists, but because the techniques used to assess diet were not appropriate, and measurement error (under-reporting actual intake, for example) was not taken into account. Equally important, if the aim of a study is to relate diet and health in groups of people, for example by comparing diet and health status between countries or regions, there is little point in spending resources on expensive and labor-intensive surveys that characterize the diets of individuals.

It must be decided at the beginning of a study whether measurements of food consumption, or nutrient intake, or both are required. The required precision (standard error) of the measurements will determine the number of subjects and the number of measurements per subject required, or the number of regions or countries to be included in a regression or correlation analysis. If ranking of subjects or groups will suffice, then a technique that has a systematic bias or nondifferential misclassification[1] may be acceptable, but if absolute measures of consumption or intake are required, techniques with a greater validity will be needed. Most prospective methods (apart from duplicate or composite diet techniques) and all retrospective methods rely on food composition tables (q.v.) for estimating nutrient intake, and the weaknesses associated with their use should be considered in any evaluation of findings. Measurements made will be affected by the time of year (summer/winter, wet season/dry season, holy days, festival days), the time of week (weekdays/weekends), and even the time of day (interview responses may differ before and after meals). Also, the precision and representativeness of the findings will be influenced by whether it is present or past diet that is being measured. Measurement of past diet is more difficult but may be important in relation to understanding the dietary origin of chronic diseases such as heart disease or some types of cancer.

In studies of individuals, different techniques require particular skills to complete records or recalls of diet satisfactorily. For example, very young or very old subjects may be unable to recall diet accurately; people with poor literacy and numeracy skills will be unable to keep adequate records.

Food-Balance Sheets (see 'Food Supply' database at http://apps.fao.org/)

These are eminently suitable for comparing diets between countries, and are readily available from the FAO, thereby reducing the cost of such studies. They do, however, show significant bias, in that the estimated availability of foods and nutrients may be considerably greater than the actual levels of consumption. In the UK, for example, food-balance sheet estimates of energy intake are 25% greater than the estimated energy requirement in the population (although some allowance needs to be made for waste and production losses). More important, this bias is probably not constant between countries. More affluent countries are more likely to overestimate consumption than poorer countries. This may be due in part to more efficient recording of food production in affluent countries, and in part to the presence in poorer countries of many small holdings and subsistence farms whose production is underrecorded.

Household Budget Surveys

Household budget surveys of nationally representative samples are regularly conducted by many governments. The results provide a ready-made database for analysis of food trends across time, and between regions or subgroups within the population. Where it is difficult to obtain individual data (such as in cultures where families eat communally from large serving vessels), household surveys may provide a valuable way of characterizing diet.

Their principal disadvantage is that foods eaten away from home are rarely recorded. These usually include a high proportion of sweets, soft drinks, alcoholic beverages, and snack foods. Also, individual intakes cannot be estimated, which may be a particular disadvantage in epidemiological studies, where it is important to relate disease outcome in individuals to diet.

Some household surveys are conducted without a larder inventory, so as to reduce the respondent burden. The assumption is that over a sufficient number of households (say 20), the differences within individual households between purchases and actual consumption will balance out, giving an overall

[1] In nondifferential misclassification, measurement errors that are present affect all subjects equally or proportionately. In differential misclassification, errors in reporting are greater in some groups (for example, overweight or elderly people) than in others. Consequently, comparison between groups is difficult to interpret. Differential measurement errors may undermine attempts to establish diet–disease relationships.

estimate of consumption close to the true level. In practice, it has been observed that surveys of food purchases in developed countries tend, on average, to stimulate purchasing in excess of actual consumption by as much as 15%. This bias is greater in households with respondents over 60 years of age and in respondents from lower income households. As lower income group consumption levels are generally lower, comparisons of the food purchases or estimated nutrient intakes are likely to underestimate the true differences between income groups except where very major differences occur (in fruit consumption, for example). Comparison of food or nutrient *profiles* (per cent of income spent on certain foodstuffs, or nutrient intakes per MJ) may be more reliable. This problem is less pronounced in surveys where the variety of food purchases is limited (as in many developing countries), or where purchases are reported retrospectively rather than prospectively. Reported household food purchases may be supplemented by records from individual household members of consumption of foods eaten outside the home, but are often subject to errors of omission.

Household surveys conducted with larder inventories give a more precise estimate of levels of food consumption and nutrient intake, but they tend to distort usual purchasing patterns when respondents' attention may be drawn to the larder contents. They also have a higher respondent burden. Some household budget surveys include expenditure only. It is then necessary to impute purchase quantities from expenditure. These are likely to be less reliable than surveys that record amounts of food purchased as well as expenditure.

Prospective Surveys of Individuals

These have the advantage of direct observation of current diet. They can be varied in duration to improved the accuracy of classification of individuals' nutrient intakes, i.e., the greater the number of days of information recorded, the more likely it is to reflect the true intake of the individual. For example, carbohydrate intake may be estimated within a few days, whereas energy intake may require a week, and fatty acid intake 2 weeks. Where a large number of days is needed, problems of respondent fatigue can be overcome by having a series of shorter periods (four 4-day periods, for example).

The principal disadvantage of prospective techniques is the level of literacy and numeracy skills required to complete the surveys. This can be overcome to some extent by having an interviewer do the recording and measuring, but this is very invasive for the subject and expensive in terms of staff time. Prospective methods may also encourage distortion of diet and avoidance of foods regarded as unhealthy, particularly sweets and alcoholic beverages, or there may be simple omission of items owing to poor recording techniques. A particular problem that has come to light with the advent of the (very accurate) doubly labelled water method for estimating energy expenditure is that overweight subjects are more likely to underrecord their intake than normal weight subjects. This has important implications when trying to establish links between diet and disease where being overweight may itself be a risk factor (heart disease, for example).

The *duplicate diet technique* is appropriate mainly for metabolic studies, as it is very expensive and time-consuming. Its use in free-living populations requires exceptionally close supervision to ensure that the food samples collected are truly representative of the foods that have been eaten. It has been used in epidemiological studies where a single day's duplicate has been collected from a number of subjects in order to characterize the group diet rather than individual diets. The composite diet technique is easier to use for assessing group diet characteristics by chemical analysis. It is very good for estimating average levels of additives or contaminants present in the diet. It has the disadvantage of relying on food intake or purchasing records or dietary recall.

The *weighed inventory* has been used very widely, and for this reason, there are many values in the literature against which new data can be compared. It is less demanding of both respondent and interviewer than the duplicate diet technique, but because it relies on food composition tables, there is a loss of precision in estimating nutrient intake. The household measures technique is useful where subjects are unused to using scales, or resources are limited, and requires a lower level of numeracy than the weighed inventory, but there is a loss of validity owing to a variation in portion sizes not reflected in their descriptions.

Retrospective Surveys of Individuals

In comparison with prospective techniques, retrospective techniques are relatively inexpensive, demanding less time of both respondent and interviewer. They therefore tend to have slightly higher cooperation rates. For foods that are consumed only occasionally (like liver) or about which there may be some stigma (like sweets or alcoholic beverages) in some countries, and for assessing the intake of nutrients (such as carotene or vitamin B_{12}) that would require many days of recording, there are advantages in using retrospective techniques that ask about the usual frequency of consumption of foods. Retrospective techniques also have the unique advantage of allowing assessment of past diet, although, in

practice, estimates of past diet may be heavily influenced by present diet. Most recalls of diet in the distant past (2 or more years prior to the interview) correlate as strongly with current diet as with the past diet assessed at the time.

Results from retrospective methods, like those from prospective methods, are subject to differential misclassification, but for different reasons. Memory and conceptualization skills (the ability to describe quantitatively what was actually eaten) differ in an unsystematic way between individuals. The very young (under 12) and the very old (over 70) are less likely to be able to recall their diets correctly. The number of food items recalled has been shown to correlate with the estimated energy intake, independent of the true level of intake. Apart from the use of repeat 24-h recalls, retrospective methods provide little or no information on day-to-day variation in diet. Such measures of variation can be useful in determining the likely accuracy of ranking of individuals according to nutrient intake. For the diet history and questionnaires, especially, subjects who do not have regular eating habits will have difficulty in describing the 'usual' frequency of consumption or portion size.

The *diet history* is especially valuable for assessing the 'usual' diet, but the nature of the interview may encourage overreporting, and the results from diet history assessments usually exceed those from weighed inventories, which, in normal-weight individuals, are probably a good reflection of 'usual' intake.

The *24-h recall* is relatively quick, demanding little of respondents' time, but cannot be used to characterize individuals' diets unless it is repeated over a number of days. In general, because of the absence of a 'training effect,' respondents are more likely to omit items than in methods that extend over several days. Estimates of nutrient intake using 24-h recalls therefore tend to be low in comparison with other techniques (whether prospective or retrospective). A new 'triple pass' 24-h recall technique (in which respondents are given two further opportunities in the same interview to recall other items consumed) has resulted in higher estimates of intake.

Questionnaires are suitable for obtaining information from very large numbers of subjects. Their administration can be more readily standardized than other types of dietary assessment (for example, administered by computer or sent by post for self-completion by respondents). Where the hypothesis to be tested relates to specific foods (e.g., vegetables) or nutrients (e.g., calcium), they can be of limited length pertaining to relatively few foods. They can thus have the advantage of brevity while still being able to rank subjects with a known degree of validity. Because of the uncertainty regarding the successful completion and validity of assessment of newly designed questionnaires, it is essential that they are properly evaluated against a dietary survey method in which the errors of assessment can be estimated. If they are to be self-completed after being posted to subjects, literacy and numeracy skills will be needed, and differential misclassification may affect interpretation.

Use of Computers

Computers are used universally to calculate the nutrient content of diets for which no direct chemical analysis is made, but they can also be used directly in the dietary assessment of individuals. Data-entry programs such as UNIDAP (Unilever) have been tailored to allow interviewers to enter patterned food consumption data, thus effectively providing the facility for an interactive diet history. Branching interactive programmes (such as EPICSOFT) allow for 24-h recall data to be entered directly at interview. Short, structured questionnaires (e.g., CALQUEST for measuring habitual calcium intake) have also been devised for interactive use. In Integrated Dietary Analysis (IDA), respondents use a hand-held computer linked with digital electronic weighing scales to collect weighed inventory data without the need for written recording or subsequent computer entry of data. Scales have also been linked to computers with special keypads with symbols instead of letters (FRED and FREDA), allowing respondents to use pictures to indicate the food being consumed; again, weights and codes are recorded automatically. A tape recorder has also been linked to scales (PETRA), allowing subjects to describe foods orally rather than in writing, a weight being recorded automatically on the tape in association with each item described. With the exception of interactive recalls, the use of computers for data collection has been disappointing because of high costs or laborious transcription of data.

Use of Biological Markers

Given the errors that are inherent in any reporting of diet, it is important to have objective measures that reflect usual food consumption or nutrient intake. Biochemical measurements of nutrients or related metabolites in blood or urine have long been used to identify nutrient deficiencies or irregularities. Markers of diet that span the whole range of intake are less common. Collection and storage of appropriate sample material may be a problem in itself (e.g., ensuring completeness of collection of 24-h urine samples). Biological markers do, however, provide important tools for evaluation of the validity of

dietary assessment and for characterization of nutritional status in population studies. For a fuller discussion of the use of biological markers, readers are referred to Margetts and Nelson (1997).

See also: **Dietary Surveys**: Surveys of National Food Intake; Surveys of Food Intakes in Groups and Individuals; **Food and Agriculture Organization of the United Nations**

Further Reading

Bingham SA (1987) The dietary assessment of individuals: methods, accuracy, new techniques and recommendations. *Nutrition Abstracts and Reviews* 57: 705–742.

Cameron ME and Van Staveren WA (1988) *Manual on Methodology for Food Consumption Surveys*. Oxford: Oxford University Press.

Margetts B and Nelson M (1997) *Design Concepts in Nutritional Epidemiology*, 2nd edn. Oxford: Oxford University Press.

Nelson M (2000) Methods and validity of dietary assessment. In: Garrow J, James WPT and Ralph A (eds) *Human Nutrition and Dietetics*, 10th edn. London: Churchill Livingstone.

Nelson M, Morris JA, Black AE and Cole TJ (1989) Between- and within-subject variation in nutrient intake from infancy to old age: estimating the number of days required to rank dietary intakes with desired precision. *American Journal of Clinical Nutrition* 50: 155–67.

Surveys of National Food Intake

J A T Pennington, Division of Nutrition Research Coordination, Bethesda, MD, USA

Purpose/Objectives of National Food Intake Surveys

National food intake (consumption) surveys are usually undertaken by government agencies responsible for the health and/or food supplies of a population. Such agencies may include departments or ministries of health, departments of agriculture, or food administrations.

National food intake surveys may be carried out to assess and monitor the dietary status of the population. Other objectives for food intake surveys are to determine the adequacy and safety of the national food supply, to determine the effects of dietary patterns on nutrient intake and health status, to assess changes in dietary patterns and their affects on nutrient intake, and to identify needs for changes in national food policies or regulations.

The dietary status of the population can be monitored by comparing food and nutrient intakes with recommended intake standards set by government, academic, or private agencies. For example, the average number of servings of fruits, vegetables, dairy products, grain products, and meats or meat substitutes may be compared with the recommended standards, and the average daily intake of nutrients (e.g., protein, fat, carbohydrate, vitamins, and minerals) may be compared with recommended dietary allowances for these nutrients. The dietary status of a nation can also be compared with measures of disease incidence to determine potential relationships between nutrient intakes and dietary deficiency diseases or chronic diseases such as obesity, heart disease, diabetes, stroke, alcoholism, and various types of cancer.

Results from food intake surveys help determine the adequacy of the national food supply and identify populations without adequate food that are underserved by current food relief efforts. Assessments of food and nutrient intake help determine the extent of physical, mental, and emotional problems relating to food insecurity (i.e., inadequate food intake). Dietary surveys help determine which population subgroups as defined by age, sex, race, ethnicity, income, urbanization, or education are at risk for energy and nutrient deficiencies. Food intake surveys should include questions about the use of public food assistance, food banks, and community feeding programs because the responses to these questions may assist government agencies in making decisions about the need for continued or increased support of such programs. Information from the survey may help determine the effectiveness of public health policies regarding food fortification, food stamps, and feeding programs for children, the elderly, institutionalized persons, and those below the poverty level. Survey results may assist governments in revising current food-related policies or developing new policies.

Results from food intake surveys also help determine the safety of the food supply. In addition to providing estimates of nutrient intake, the information from national food intake surveys may help provide dietary exposure estimates for pesticide residues, food contaminants, and food toxins. Concentration data on the levels of pesticide residues, contaminants, and toxins in foods may be merged with the food consumption data from the surveys to provide such exposure estimates. Information on dietary exposure to potentially harmful constituents in foods may be useful for modifying or revising policies concerning agricultural and food manufacturing practices.

Results from food intake surveys help determine the effects of dietary patterns on nutrient intake and on health status. Families perpetuate the dietary patterns learned from their parents and other relatives, and families within communities and larger geographic areas (e.g., countries) tend to have similar dietary patterns. These learned dietary patterns may be modified or changed with exposure to the foods and eating patterns of other cultures via travel, immigration, migration, or the presence of visitors. Media food advertising and the offering of new and different foods from grocery stores and restaurants may also affect dietary patterns. Dietary surveys should include questions about cultural and ethnic dietary patterns so that the relationships of cultural and ethnic factors to food and nutrient intake can be identified. This may allow for the identification of relationships between culturally or ethnically defined dietary patterns on dietary and health status.

Food intake surveys are useful for following changes in dietary patterns, their effects on nutrient intake, and potential implications for health. For this purpose, the surveys are most useful if they are done routinely (e.g., yearly or biennially) using the same or similar methodology so that changes and trends in food and nutrient intake can be identified. Changes in food intake may reflect the impact of the availability of ready-to-eat products from grocery stores, fast food restaurants, and other food vendors. Changes in agricultural and manufacturing practices may result in changes in the composition of foods (e.g., the type of fats and oils in baked products and the type of sweeteners in soft drinks and desserts) so that nutrient intake changes even without changes in food patterns. It is important to have ongoing monitoring of food and nutrient intake as well as monitoring of disease incidence so that cause and effect might be investigated.

Food intake surveys are useful for identifying issues that might require changes in policies or regulations concerning food safety, food labeling, food fortification, food additives, agriculture practices, or manufacturer practices. Ultimately, the food policies and regulations developed by government agencies should serve to maintain and improve the health of its citizens. Food intake surveys may be done in conjunction with health examination surveys that include physical exams, health histories, clinical measures, and/or the collection of biological samples (urine, blood, or other tissue) that are analyzed to assess nutritional or health status. When surveys of food intake and health measures are done concurrently on the same population, there are opportunities to relate food and nutrient intake to health status measures such as blood pressure, body weight, and blood levels of cholesterol, triglycerides, and various nutrients. Changes in the health status of people within a country could be the result of many factors (diet being one of them). Dietary status data derived from food intake surveys could be an important factor when assessing the health of a population. Other factors (physical activity patterns, environmental health, economics, epidemics, etc.) are also very important with regard to the health of a population. Dietary surveys provide an important piece of information that will help governments in making policy and regulatory decisions concerning foods and agriculture.

Sample Population

A food intake survey should include survey participants that are representative of the population in terms of the demographic features of interest to the investigators. Usually, the demographic factors include sex, age, region, urbanization, income, education, race, and ethnicity. There may be other important considerations with respect to population sampling for a national survey. For example, a government agency may wish to oversample at-risk populations or minority populations for whom few dietary data are currently available.

Obtaining a representative sample population for a food intake survey requires the expertise of statisticians who are knowledgeable of the population and its dynamics. The statisticians can assess the currently available population databases (e.g., census database, voters registration database) and determine whether they might be useful to develop a sample population for the dietary intake survey.

The agency responsible for the food intake survey must consult with the statisticians and survey designers regarding important facts such as:

- If individuals in institutions (hospitals, prisons, rest homes, retirement communities, convents, monasteries, etc), individuals without permanent resident status (immigrants, aliens), and individuals without homes are to be included in the survey. Survey participants generally need to have a home address, a telephone or other means of contact so they can be requested to participate in a survey.
- Which demographic features (age, sex, race/ethnic group, income, region, urbanization, education) are of importance to the survey results and how many individuals with the demographic features must be in the survey so that representative dietary intakes can be calculated for them.
- If minority or at-risk groups should be oversampled if they will not be present in sufficient numbers in the overall sample.

- If the survey should be designed to correct for sample bias (e.g., nonresponse rate, oversampling). One method to correct for sample bias is to develop weighting factors for each individual in the survey that pertains to his/her demographic characteristics.
- How to obtain dietary information for those not able to respond for themselves (e.g., young children, ill, elderly, at work, traveling, illiterate, etc.).
- How to maximize survey participation of the selected participants, e.g., by sending a letter of intent, following up with a phone call, minimizing respondent burden, and providing remuneration (if possible) for the subjects' time and effort.
- How to recontact people if necessary (e.g., to clarify information) without violating privacy issues.

Methods Used for National Food Intake

Information about the foods consumed by a nation can be obtained using a variety of methods such as food disappearance (per capita intake), food balance sheets, household inventories, 24-h dietary recalls with a professional interviewer or computer assistance, food diaries (records) kept by survey participants, weighed food intakes, diet histories, duplicate portions, and food frequency questionnaires. There are various advantages and disadvantages of each method with regard to costs, respondent burden, and accuracy and reliability of the information obtained. Costs include the staff time, field survey transport and housing, materials, data processing, and participant renumeration. Respondent burden refers to the time and effort required by the survey participant to provide information. (*See* **Dietary Surveys**: Measurement of Food Intake.)

The preferred methods for national food intake surveys are 24-h dietary recalls or food diaries, and it is most useful if there are two or more days of dietary information for each survey participant. The methods of food disappearance, food balance sheets, and household inventories are less accurate, tending to overestimate food intake, and do not provide specific information based on the age and sex of participants. Weighed food intakes, diet histories, and duplicate portions take much interviewer and participant time and are expensive (i.e., the cost per person is high for the large sample population needed for national data). The weighed intakes and duplicate portion technique also tend to alter usual eating patterns, usually by decreasing total food intake. Food frequency questionnaires may provide useful information but are generally not quantitative and are not specific enough because they include a limited number of foods (as low as 50).

Although the 24-h dietary recalls and food diaries are likely the methods of choice for national surveys, they too have disadvantages. The 24-h dietary recall requires accurate memory and honesty of the survey participant as well as knowledge of food names, descriptions, and serving sizes. Food diaries have the effect of altering usual food intake and are also subject to the honesty of the survey participant. The act of writing down consumed foods in a diary tends to decrease total food intake and may result in 'healthier' food choices that are not typical or usual in countries where obesity prevalence is high. The main problem with dietary recalls and diaries appears to be underestimation of total food intake. Recent studies, using the doubly labeled water technique, have estimated that dietary recalls and diaries may be underestimated by about 20% of energy. It is not known if the underestimation applies equally to all foods or if there is selective underestimation as for example with alcoholic beverages, snack foods, or desserts.

The information derived from dietary surveys is particularly useful to identify potential health problems related to food and nutrient intake. The potential problems can be investigated more fully by collecting and evaluating information obtained from clinical and biochemical measures of nutritional status. Because of the difficulty of obtaining accurate, valid, and reliably dietary information from survey participants, dietary data are usually not considered sufficient for making changes in nutrition policy, i.e., the dietary results need to be verified with clinical and biochemical results. Otherwise the benefits of policy changes (which can be expensive and burdensome on the food industry) may not be forthcoming.

Considerations with Dietary Recalls and Records

The food intake survey should be designed with considerations for the number of days of recalls or records, the days of the week and seasons of the year, memory cues and probes, the interviewing situation and interviewing techniques, the food composition database, default food codes, dietary supplements, cultural sensitivity of the dietary tool, specific food use questions, and the use of biochemical markers to assess the validity of the dietary method.

1. For most nutrients, as many as 7–14 days of dietary information may be necessary to characterize usual intake for an individual. However, for population distributions of nutrient intake, 2 days of food recalls or diaries are usually sufficient. Collecting and processing additional days (beyond

two) of dietary information result in increased financial costs as well as reduced response rates as individuals tire of participating.

2. Because people tend to eat differently on work days and days off work, it is important to factor days of the week into the sample design for food recalls and diaries. Eating patterns are affected by the seasonality of some foods (e.g., watermelon, strawberries, plums, grapes, corn, spring lamb, oysters, and crabs). The seasonality of outside temperatures may affect food preferences and eating patterns. People often prefer hot and hearty soups and meals on cold days and salads or lighter meals on hot days. Feasting and fasting patterns around religious and national holidays are often different than other days. It may be best to avoid collecting data for such days or to factor these days into the sampling design so that the overall results will be representative.
3. Cues and probes provided by the dietary interviewer or the computer-assisted interview help survey participants remember the details about the foods they eat. Questions about snacks and about added ingredients (sugar and cream in coffee, catsup, salad dressing, sauces) are important to obtain complete food recalls and diaries. The use of food models and food pictures also helps to clearly identify foods and to improve the accuracy of serving size estimates.
4. The manner in which food-related questions are asked affects responses. Certain characteristics of the interviewing situation may affect responses about foods that might be depicted as 'good foods' or 'bad foods' by the survey participants. For example, the presence of other family members during the dietary interview may lead to underestimation of the intake of alcoholic beverages, snacks, and dessert items (especially for parents and teenagers). An interview in a health setting such as a clinic may elicit responses with underestimated amounts of fat and overestimated amounts of foods with higher nutrient content (e.g., fruits and vegetables). Interviewers should attempt individual interviews, if possible, refrain from either positive of negative feedback about food choices, and encourage subjects to accurately report all foods. Interviewers can emphasize the confidentiality of the recall or diary and indicate that only grouped results will be reported unless sent in advance via mail to the participants.
5. Dietary interviews have traditionally been done as in-person (one-on-one or in family interviews), but are now also done by computer and by telephone. The computer-assisted interview helps standardize the questions and probes because some potential human errors are removed. Computers do not get tired, sick, forgetful, or irritable, and computers do not require training, quality control checks, or intrainterviewer compatibility checks. Telephone interviews are less expensive than in-person interviews as they do not require an examination site or the travel costs of survey participants. However, telephone interviews have the disadvantage of not permitting the use of food pictures or models of serving portions.
6. To derive accurate and reliable information from the dietary interview, the food consumption data must be merged with a food composition database that contains accurate and relevant data on the nutrient content of the foods consumed by the population. The database should contain both the generic and brand name products typically eaten. To ensure accuracy, the nutrient information on brand name products in the database should be obtained from the manufacturers or from a qualified laboratory. The foods in the database should contain the current levels of nutrient fortification for products usually fortified such as wheat flour, white rice, cornmeal, ready-to-eat breakfast cereals, milk, salt, etc. The database should include foods from fast food and other restaurants as well as manufactured and homemade mixed dishes, ethnic foods, and desserts.
7. Survey participants often have difficulty identifying foods, remembering (or knowing) the ingredients in mixed dishes, and properly identifying serving portions. Food models, pictures, and computer graphics may help participants identify foods and serving sizes. However, default food codes will be necessary when detailed information cannot be elicited from the participants. Default food codes are needed for mixed dishes, especially homemade and restaurant where exact recipes and ingredients are not known. The database can carry mixed dishes with traditional ingredients (from cookbooks) and calculated nutrient values as defaults. Defaults may also be necessary for sandwiches, soups, pizza, pie, cake, cereal, salad, and many other general food categories where more detail cannot be elicited.
8. Information about the intake of dietary supplements is important because supplements can be very potent sources of nutrients and far outweigh the concentration of nutrients obtained from foods. However, it is usually best to maintain information about supplement intake separately from food intake so that survey results can be presented both ways (i.e., as nutrient intakes excluding supplements and intakes with supplements for those individuals who consume them).

Of primary interest is the type of supplement taken (the nutrients included and the brand name), the potency for each nutrient, the frequency of use of the supplement, and the duration of use. There are thousands of supplement products available both over-the-counter and by prescription. Information about supplement intake is difficult to capture and code. A particular complexity in assessing supplement use is that many individuals take supplements inconsistently or in patterns that are difficult to characterize (e.g., when they do not feel good, on certain days of the week, until the bottle is empty, when they feel tired). If possible, the survey participants should be asked to bring their supplement bottles to the interview so that the interviewer can account for the types and amounts taken. If supplement intake data are to be collected during the survey, a supplement database will be needed to convert the supplement intake data into nutrient intake data.

9. The dietary tool developed for a food intake survey may not work equally well for all groups within a population. The method may not retrieve the same level and quality of information from minority populations or population subgroups that differ by culture, urbanization, income, education, or living situation. (In fact, men and women may respond differently to the same tool, and children may respond differently than teenagers or adults.) The food intake tools should be culturally competent, or there should be different tools for different cultures. Infants, children, adolescents, and elderly persons may require surrogate reports or confirmation of dietary intakes from parents or caretakers. People who are institutionalized (i.e., in hospitals, rest homes, retirement homes, prisons, or psychiatric institutions) may require surrogate dietary intake reports from their caretakers.
10. It may be necessary to include specific food use questions in the survey to meet the identified objectives of the agency conducting the survey. For example, there may be questions about the type of cooking and table fat typically used (butter, margarine, lard, suet, shortening), the use of discretionary salt during cooking and at the table, whether the salt is iodized or not, the use of a microwave oven, typical methods of food preparation (frying, baking, grilling), the use of artificial sweeteners, the use of foods with fat or sugar replacers, the use of medical foods, if fat is trimmed from meat cuts, if skin is removed from poultry, or the use of low calorie products (e.g., salad dressings, soft drinks).
11. Dietary surveys are much less expensive than physical exams with clinical and biochemical measurements to assess the nutritional status of a population. However, as noted, the results of dietary surveys are not completely accurate. It would help to validate (or invalidate) the results of dietary surveys by using a short-term biochemical marker, if resources permit. Short-term biochemical markers that have been used to attempt to validate the previous day's nutrient intake include urinary nitrogen, potassium, sodium, and carotenoids for dietary protein, potassium, sodium, and carotene as well as the doubly labeled water technique for energy intake.

Summarizing Data and Data Availability

The information collected during dietary surveys must be stored, reviewed, aggregated, and evaluated with special attention to weighting factors and to the sample number when data are reported by demographic characteristics. Nutrient intakes may be reported as means with standard deviations, medians, and percentile distributions. The data derived from a national food intake survey are generally valid for the population as a whole and for population subgroups for which the sample size is sufficient. Government agencies should make the results of their food intake surveys widely available through publications and on the internet to assist other government agencies that are attempting to undertake food intake surveys for the first time or that wish to improve upon the methods used in their previous surveys. Government agencies should summarize the results of their food intake surveys and explain and justify their conclusions regarding the dietary status of their population. They may also attempt to relate the dietary status of the population or population subgroups to the reported incidence of diet-related diseases in the population or in population subgroups.

See also: **Dietary Surveys**: Measurement of Food Intake

Further Reading

Burk MC and Pao EM (1976) *Methodology for Large-scale Surveys of Household and Individual Diets*. Home Economics Research Report No. 40. Washington, DC: Agricultural Research Service, USDA.

Dwyer J (1999) Dietary Assessment. In: Shils ME, Olson JA, Shike M and Ross AC (eds) *Modern Nutrition in Health and Disease*, 9th edn. Baltimore, MD: Williams & Wilkins.

Food Balance Sheets, Food and Agriculture Organization, United Nations, http://apps.fao.org/page/collections?subset=nutrition.

National Health and Nutrition Examination Survey, Centers for Disease Control and Prevention, United States Department of Health and Human Services, http://www.cdc.gov/nchs/nhanes.htm.

Thompson FE, Byers T and Kohlmeier L (1994) Dietary Assessment Resource Manual. *Journal of Nutrition* 124 (Supplement II): 2245S–2317S.

Willett WC and Sampson L (1997) Dietary Assessment Methods. Proceedings of a symposium held in Boston MA, January 22–24, 1995. *American Journal of Clinical Nutrition* 65(Supplement): 1097S–1368S.

Surveys of Food Intakes in Groups and Individuals

G H Beaton, GHB Consulting, Willowdale, Ontario, Canada

Background

Studies of food intake by individuals or groups are usually undertaken for a narrowly defined purpose. This must influence both the choice of dietary methodology and approach to analysis and interpretation. In the case of individuals, intake assessments are undertaken most often in connection with dietary counseling, monitoring compliance, and, sometimes, as part of a risk-assessment program. Since counseling usually relates to foods, not nutrients and even more often to patterns of food use, qualitative assessment of food use may be all that is needed. When compliance to a strict metabolic regimen is to be assessed, quantitative methodologies and computation of nutrient intakes are likely necessary. The methodological demands are very different, depending upon the intended application. The potential purposes for group surveys are more diverse. The intake data collection might be a part of a national health and nutrition survey; here, primary interest often spans both nutrient intake and food usage, but the important focus is on the distribution of usual intakes among individuals in the target population. Quantitative estimates of intake are needed. More often, group surveys may be undertaken either as part of the evaluation of a program or in the assessment of possible need for an intervention in a subpopulation. Here, the precise measures of interest will depend on the defined evaluation, but usually, quantitative assessments are needed. A different demand emerges when data collection is a part of a specific epidemiological study. The primary objective may be to rank individuals along a continuum of intakes with less concern about absolute levels of intake. Depending on the nutrient(s) or food(s) of interest in the planned epidemiological study, a semiquantitative methodology may suffice and may be more cost-effective. While it is possible to design simplified assessment tools for nutrients such as vitamin A or calcium, caution must be taken to avoid using the data collected for other purposes. The more broadly the data are to be used, the greater demand there is for quantitative methodology. A recent report by the European Food Consumption Survey Method group concluded that repeated 24-h recalls were the preferred method for planned health and nutrition monitoring. This is consistent with earlier decisions relating to the design of the two large US national surveys.

The impact of error terms in dietary data, and the availability of statistical methodology to diminish their impact in analyses, is vastly different between the assessment of individuals and groups.

Group Studies

Epidemiological Studies

Although it is clear that epidemiologists can and do use quantitative food intake data, it is also clear that many such studies collect data with food-frequency or diet-history techniques, most often classified as qualitative or semiquantitative methods. Large epidemiological studies differ in purpose and design from national surveys. They are often designed to examine groups defined by specific characteristics of interest to the investigator rather than representing the population as a whole. They may be concerned with a particular aspect of diet and hence can optimize data collection for this variable. Rather than being primarily interested in the distribution of intakes in a national survey, the epidemiology study usually wants to test the association between the intake variable and a defined outcome. This places emphasis on the ranking of intakes of individuals rather than the distributional characteristics.

Many epidemiologists select food-frequency or diet-history methodologies, sometimes selectively abbreviated, rather than the quantitative methods recommended for national surveys. There are at least two apparent reasons for this choice, one theoretical and one logistical: (1) the frequency questionnaire may be assumed to circumvent the problems of day-to-day variation in intake by asking about usual eating practices (see discussion of this under individuals, below); (2) a frequency questionnaire is more easily administered than a record, one day recall or diet history. Frequency questionnaires may be designed for self-reporting and may be distributed to participants by

mail. Food-frequency data are generally believed to provide less precise quantitative information than expected with daily recalls or records of foods consumed. There may be a trade-off between errors involved and feasibility in the selection of methodologies.

It is almost certain that, in common with all other dietary methodologies, there is an error term attached to the data collected with food-frequency methods. This may be the equivalent of the 'random error' in records and recalls, although potentially of smaller magnitude. However, since the investigator has only a single estimate for each individual, there is no simple approach estimate this error. Approaches have been developed and are still evolving to use highly standardized 24-h recall intake data in the calibration of food-frequency data across multiple studies.

Drawing on the theory of errors in measurement, numerous statistical approaches that offer the possibility of adjusting analyses, or at least assessing the impact of random error on analytical outcome, have been proposed. These require the estimation of the error term in the dietary data. Therein lies a key limitation of the food-frequency instruments; the error term is essentially inestimable.

For large epidemiological studies, involving geographically dispersed subjects, logistical feasibility becomes an overwhelming consideration. The food-frequency method often carries overwhelming logistical advantages for epidemiological studies – advantages that far outweigh the potential or real disadvantages mentioned above.

Evaluation and Monitoring of Interventions

When the objective of a special-purpose survey is to evaluate or monitor an intervention program, there could be interest in a specific dimension of food intake, allowing adoption of specifically designed and targeted food-intake methodology. Often, the investigator may wish to estimate intakes of many foods, patterns of food use, or multiple nutrient intakes. If so, the preferred methodology may be repeated one-day recalls or records. Depending upon planned analyses, the investigator may be more interested in group behavior than in individual response. This too will affect the choice of methodology.

General Comment and Conclusion about Group Assessments

When the research question to be posed concerns the 'population risk' or prevalence of apparently inadequate intakes, the past 20 years have seen major forward strides. Now available are statistical approaches and software that can estimate the distribution of usual intakes with a single day of data on all subjects and a sample of replicates, or preferably, two independent days of data for all subjects (Hoffman *et al.*, 2002; Institute of Medicine, 2000). It has also been demonstrated that the prevalence of apparently inadequate intakes can be approximated by estimating the proportion of individual with usual intakes below the median nutrient requirement for the specific class of person (Institute of Medicine, 2000). Two major difficulties have been resolved. A remaining serious problem is that of alleged underreporting bias (see discussion below). Although many group assessments are undertaken for very specific purposes that allow adoption of specialized and targeted methodologies, there is danger in *post-hoc* analyses of the data base for other purposes.

Studies of Food Intake in Individuals

Although the unit of observation in groups is the individual, the unit of interest is the group. In this section, we address the situation when the unit of observation and of interest are both the individual. While this may seem 'academic,' it will be seen that there are extremely important differences between collecting and analyzing data for individuals as part of a group and for individuals as individuals.

The worst-case scenario arises when the goal of investigation is to evaluate the apparent adequacy of dietary intake for the particular individual (without assessment of the health condition of the individual). This involves an estimation of the individual's *usual* intake of one or more nutrients and comparison of this with their requirement of the nutrient(s). In this situation, three very serious issues must be addressed.

The most obvious problem is that associated with day-to-day variation in reported intake. For most nutrients, the within-person coefficient of variation (CV) falls between 15 and 30%. Vitamin A, vitamin B_{12}, dietary cholesterol, and fatty acid ratios have a notoriously higher day-to-day variation. Consider the very conservative estimate of a CV of 15%. For a single day's intake, this translates to mean that the individual's true usual intake lies within $\pm$ 30% of the estimate (95% confidence interval assuming a normal distribution). With an average of seven independent days of data, the confidence interval might fall to about $\pm$ 11–12%, and with 14 days of data, the interval might be $\pm$ 8%. It has been suggested that for dietary cholesterol, 4 weeks or more of dietary data might be needed to bring the confidence bounds to $\pm$ 10%. To make matters worse, it is generally assumed that the day-to-day variation in intake is random and that its magnitude across individuals is comparable. Both of these assumptions have been shown to be wrong. The magnitude of day-to-day variation is a characteristic of the individual. The

confidence bounds on an estimate of intake for a fixed number of days vary between individuals. Furthermore, a part of the variation is not random. It has been recognized that day of the week and other social patterns exert effects on food intakes of individuals. In addition, many individuals exhibit long-term patterns within their day-to-day variation. The reported energy intakes of one such individual collected across 365 consecutive days) are portrayed in Figure 1. One implication of these longer time trends is that to sample days correctly, one must define the time dimension of 'usual' and sample across that time frame rather than just worrying about the number of days needed for a defined confidence bound. It is seldom possible to collect many days of data for an individual and rarely possible to sample time correctly. Instead, it must be accepted that there are residual errors, often very substantial, in the feasible estimates of the usual intake of the particular individual. It is not clear whether diet-history methods can overcome these problems, and no easy way to address this question has been forthcoming. Compare this with the availability of statistical methods to estimate the distribution of usual intakes with only 2 days of data in the examination of large groups.

A second issue in the assessment of nutrient intake of a particular individual is that, for nutrients such as iron and zinc, we must recognize the potential effect of absorption enhancers and inhibitors concurrently consumed *in the same meal*. With folic acid, it is now suggested that folic acid taken as supplements or added to foods in fortification is absorbed more efficiently than the folate indigenous to foods. While our understanding of effects such as these is improving and algorithms to adjust estimated intake (e.g., to folate equivalents) are appearing, it must be accepted that we face appreciable errors from this source, irrespective of the choice of food-intake methodology. If these become essentially random errors, in population analyses the effects are addressed along with other random errors, but not in examination of the particular individual.

The third issue, this time relating to data analysis and interpretation, arises in the comparison of intake with requirement. Individuals have specific requirements for nutrients, but these differ among seemingly similar individuals categorized by age, gender, body size and composition, physical activity, etc. At best, we can describe the estimated distribution of requirements among individuals like the specific subject. There can never be a way of classifying the intake of a subject as 'adequate' or 'inadequate' unless their usual intake lies well above or well below the range of requirements. If the usual intake falls within the range, in theory one can assign a probability of adequacy. Any probability statement is based on the cumulative distribution of requirements. This is illustrated, for protein (FAO/WHO/UNU 1985 requirement estimates) in Figure 2. If the estimated usual intake of protein were estimated to be $0.1\,g\,kg^{-1}\,day^{-1}$, it would almost certainly be inadequate. Very few people would be expected to have a requirement that low. If the intake were $1.5\,g\,kg^{-1}\,day^{-1}$, one could assert confidently that intake is almost certainly adequate. Very few individuals would be expected to have a requirement that high. If the intake were $0.6\,g\,kg^{-1}\,day^{-1}$ (the median requirement), there

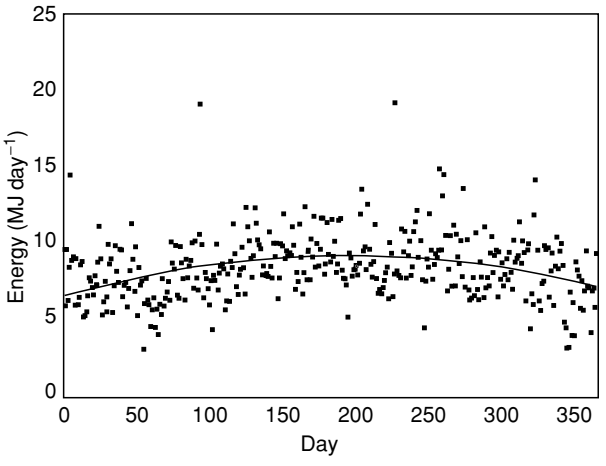

Figure 1 Long-term pattern within the day-to-day variation of energy intake in a single subject who reported intake for 365 consecutive days. From Beaton and Tarasuk (1991), with permission. Data from the Beltsville One Year Study of Dietary Intake.

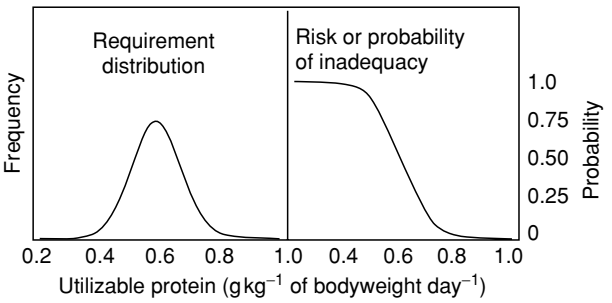

Figure 2 Probability approach to assessment of observed intake. The estimated distribution of protein requirements of young adult men is portrayed on the left (based on FAO/WHO/UNU (1985) *Energy and Protein Requirements: Report of a Joint FAO/WHO/UNU Expert Consultation. World Health Organization Technical Report Series 724*. Geneva: World Health Organization). The probability or risk curve, portraying the likelihood that a particular level of usual intake is inadequate to meet the actual requirement of a randomly selected individual, is shown on the right. Reproduced from Dietary Surveys: Surveys of Food Intakes in Groups and Individuals, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

would be a 50:50 chance that intake is inadequate: half of the individuals would be expected to have a higher requirement, and half would be expected to have a lower requirement. The approach can be formalized by the development of a risk curve, as portrayed in Figure 2.

The description of the probability approach described above assumes that the individual's usual intake is known without error. This is an incorrect assumption, as indicated above. Indeed, the error associated with the estimated usual intake may equal, or exceed, the estimated variability of requirement. This issue was addressed, without satisfying resolution, by the Institute of Medicine in 2000.

The final position we must accept is that, in spite of a sound theory based on probability, in fact we do not have, as yet, any practical tools that allow valid assessment of apparent adequacy of intake in a particular individual. We are deceiving ourselves if we think such assessment can be made, except in the situation of estimated usual intakes greatly higher or lower than the range of requirements (technically, at least two standard deviations of estimated usual intake above or below the tails of the requirement distribution).

What may be very hard to understand is that, while these issues may seem insurmountable in investigating of a particular individual, they become much less important in groups and, more particularly, population assessments. Many of the errors discussed above behave like random error terms. Some cancel out, but all can be addressed using currently available statistical tools. The existence of bias in reporting food intake (systematic under- or overreporting) cannot be corrected by any presently accepted procedure and remains an important issue for both individual and group assessment. The probability assessment of observed intake has established merit when applied to large groups and can even be greatly simplified in many situations.

Summary Comment – Individual Assessment

From the above, it follows logically that if evaluation of nutrient intake in the particular individual is very problematic, one must ask whether or not it is worth attempting. A much simpler dietary methodology can suffice to provide general descriptions of current food-use patterns, even though these may be more qualitative. For many counseling applications, or for monitoring compliance with a prescribed dietary regime, these may suffice and are almost certainly more acceptable to the client. The professional must judge what level of error is acceptable for the purpose at hand. If a high degree of precision is needed in such assessment, the other approaches not involving dietary data should be considered.

Bias in Reporting and Recording Food Intake or Computing Nutrient Intake

Bias in estimated nutrient intake can result from intentional or unintentional (erroneous memory) reporting by the individual; from errors associated with the instrument and/or interviewer involved in recoding the reported intake; or from food coding and the food-composition database. A specific example of the last of these might be the estimation of selenium intake without identification and coding of wheat sources in a country that imports wheat from more than one region. Instrument and interviewer errors have been documented in a number of studies. These sources of error are being addressed through a periodic comparison of chemically measured and calculated composition of composite meals and by careful training of interviewers together with a greater reliance on new computerized interview techniques that incorporate probes and quality checks. The most difficult source of bias to detect and control is that arising from the subject themselves, be it intentional under- or overreporting, or the consequence of imperfect memory. Interview probes may help avoid the latter.

The most widely described bias in estimated intake is with reference to energy. A few early studies demonstrated that obese subjects fed what they reported consuming showed weight changes very different from those seen when intake was being reported (inference was underreporting). When algorithms for predicting basal metabolic rate and total energy needs became available, these were applied to large group databases, and the discrepancies between reported intake and computed need in apparently weight-stable people were cited as evidence of bias in reporting. With the introduction of doubly labeled water techniques for indirect estimation of total energy expenditure, we have seen a new wave of reported bias in estimated energy intakes. Information pertaining to nutrients is sparse, since we seldom have an accepted gold standard for independent estimation of intake.

There can be no question that bias in estimating nutrient intake exists. The theoretical impact is clear. Consider, for example, the impact of underestimating protein intake by 10%. If true intake is well above the range of requirements, the assessed risk of inadequacy would remain 0 and present no problem. If true intake were perhaps $0.8\,g\,kg^{-1}\,day^{-1}$, the assessed probability of inadequacy could increase from

2–3% to about 30%. In groups, the bias likely varies from person to person. For assessments of the type just portrayed, there remains a problem. For epidemiological studies in large groups, 'random' bias does not necessarily hide true associations between intake and some health outcome; it increases 'noise' in the data and decreases statistical power. Conversely, while the presence of an association may be detected, the description of that association involving actual levels of intake is affected. If the bias moves with the outcome of interest (e.g., alleged underreporting of intake by overweight subjects), extremely serious confounding of analyses can result.

To date, no generally accepted method of 'adjusting' nutrient intakes for the probable bias has been developed. Many approaches have been suggested, but each presents serious limitations, at least for some applications. Two general approaches in varying forms have appeared and are mentioned here to illustrate the type of issues that arise. One approach would predict energy need and then adjust nutrient data by a factor (estimated energy requirement)/(estimated energy intake). The second would also predict energy requirement, establish limits based on expected errors in that estimate, then reject subjects with estimated intakes outside those limits. When these approaches are applied to nutrients, they both assume that misreporting of nutrients is exactly proportional to energy (i.e., the nutrient:energy ratio is valid). We lack sound evidence for this assumption. When outliers are excluded, one is likely to generate yet another bias – erroneous reporting that moves a subject into the acceptable range will be missed. Further, if rejection rates are substantial, one must question whether the data continue to represent the target population. A final issue is the simple observation that, at present, we are unlikely to be able to predict the energy requirement of a particular free-living subject with any better precision than we can estimate their energy intake. One must question approaches that attempt to address measurement error by using an erroneous reference.

Bias in estimated nutrient intake is a continuing problem without, at present, a satisfactory solution. Analysts and users of nutritional analyses must be cognizant and cautious in their interpretations.

Conclusion

There are serious limitations associated with the estimation of food and nutrient intakes in individuals. As one moves to larger groups, some of these limitations diminish in importance. Certainly, existing limitations have not prevented major advances in establishing associations between diet and disease. However, at the level of the particular individual, the recognized limitations call into question the merit of attempting detailed dietary assessments. Limitations of data collected in studies of individuals or groups must be weighed up in terms of the defined purpose of the study and planned analytical methodology. When there is a mismatch between purpose and methodology (of collection and analysis), quality declines, and limitations on interpretation are likely to increase.

See also: **Bioavailability of Nutrients**; **Dietary Reference Values**; **Dietary Surveys**: Surveys of National Food Intake; Surveys of Food Intakes in Groups and Individuals; **Energy**: Energy Expenditure and Energy Balance; **Epidemiology**

Further Reading

Beaton G (1999) Recommended dietary intakes: Individuals and populations. In: Shils ME, Olson JA, Shike M and Ross AC (eds) *Modern Nutrition in Health and Disease*, 9th edn., pp. 1706–1726. Baltimore, MD: Williams & Wilkins.

Biró G, Hulshof KFAM, Oveson L and Amorim Cruz JA (2002) Selection of methodology to assess food intake. *European Journal of Clinical Nutrition* 56 (supplement 2): S25–S32.

Hoffman S, Boeing H, Dufour A *et al*. (2002) Estimating the distribution of usual dietary intake by short-term measurements. *European Journal of Clinical Nutrition* 56 (supplement 2): S53–S62.

FAO/WHO/UNU (1985) *Energy and Protein Requirements: Report of a Joint FAO/WHO/UNU Expert Consultation. World Health Organization Technical Report Series 724*. Geneva: World Health Organization.

Institute of Medicine (2000) *Dietary Reference Intakes: Applications in Dietary Assessment*. Washington, DC: National Academy Press.

Liu K (1989) Consideration of and compensation for intraindividual variability in nutrient intakes. In: Kohlmeier L and Helsing E (eds) *Epidemiology, Nutrition and Health*, pp. 87–98. London: Smith–Gordon–Nishimura.

Mertz W and Kelsay JL (1984) Rationale and design of the one-year dietary intake study. *American Journal of Clinical Nutrition* 40: 1323–1326.

Tarasuk V and Beaton GH (1991) The nature and individuality of within-subject variation in energy intake. *American Journal of Clinical Nutrition* 54: 464–470.

Tarasuk V and Beaton GH (1992) Statistical estimation of dietary parameters: implications of patterns in within-subject variation – a case study of sampling designs. *American Journal of Clinical Nutrition* 55: 22–27.

Willet W (1990) *Nutritional Epidemiology. Monographs in Epidemiology and Biostatistics*, vol. 15. New York: Oxford University Press.

Volatier JL, Turrini A and Welton D (2002) Some statistical aspects of food intake assessment. *European Journal of Clinical Nutrition* 56 (supplement 2): S46–S52.

DIETETICS

P A Judd, University of Central Lancashire, Preston, Lancashire, UK

Introduction

This article briefly relates the history of the development of dietetics and the role of the dietitian and describes the current roles of dietitians in some general and specialist areas of practice. The increasingly important roles of the dietitian in research and education are also discussed.

Definition and History of Dietetics

Dietetics is defined as 'the application of the science of nutrition to the human being in health and disease.' However, the term 'dietitian,' used to describe a practitioner of dietetics, was in use long before the science of nutrition had become an accepted discipline. Although Florence Nightingale has been described as the first dietitian because of her insistence, during the Crimean war, that the sick men in her charge would not recover if not well fed, the first use of the title of dietitian was recorded in 1899 in the USA. The dietitian was then described as a person working in a hospital who provided nutritious meals to patients and the earliest dietitians were primarily home economists. The role of the dietitian expanded greatly during the 20th century and is no longer seen simply as the provision of food. The dietitian is seen as the expert in the planning and evaluation of nutritional care for patients requiring therapeutic dietary regimens as well as for the population in general and this is reflected in the breadth of knowledge and skills required for qualification and registration as a dietitian.

The profession of dietetics is a relatively young one, first formalized in the USA in 1917 when the American Dietetic Association (ADA) was founded. In the UK the first dietitians were nurses and the first dietetic department was started in the Edinburgh Royal Infirmary in 1924. Several of the pioneers of British dietetics were able to visit the USA in the 1920s to study the work of established dietetic departments there and after several abortive attempts the British Dietetic Association (BDA) was established in 1936. Other countries followed and in 2000 there were more than 30 dietetic associations throughout the world registered with the International Committee of Dietetic Associations (ICDA).

Dietetics Qualifications

It is now accepted that the practice of dietetics requires a wide range of knowledge and skills. As the scope of dietetic practice has expanded over the years, training programs have continually adapted to insure that the practitioner has the current knowledge and skills required. The education and training of a dietitian usually comprises a degree program (either BSc or MSc), based in a university, including or followed by a period of practical training based in a recognized hospital dietetic department.

Preregistration programs include coverage of basic and applied sciences (chemistry, biochemistry, physiology, nutrition, microbiology) as well as social sciences (psychology and sociology). In addition, as dietetics is concerned with feeding people, a knowledge of the food habits of populations together with detailed knowledge of food composition and food preparation is essential. To this basic foundation is added knowledge of medicine, pathology, and the therapeutic uses of dietary treatment. Increasingly the development of skills required to communicate with all types of people is emphasized, whether counseling individuals or teaching groups.

During the practical training or internship the student dietitian learns to apply the theory learned at university with individuals or groups of people. The training covers all aspects of dietetic practice and the students spend time in different settings, including community care and, in some countries, in large-scale catering establishments. In order to become a registered practitioner students must demonstrate that they have both good theoretical knowledge and are competent practically.

In the USA, UK, Canada, Australia, New Zealand, South Africa, The Netherlands, and many other countries, training programs are regulated by bodies external to the educational establishments and successful completion of such a regulated training allows registration as a dietitian. In the USA regulation of courses and training programs is carried out by the ADA. In the UK, the former Dietitian's Board of the Council for Professions Supplementary to Medicine (CPSM) has been replaced by the Health Professions Council, set up to regulate the 12 professions previously overseen by the Boards of the CPSM. In the UK only state-registered dietitians may be employed in the National Health Service. The registration body, in each case,

produces a statement of conduct which describes the role and responsibilities of the registered dietitian and failure to work within this statement of conduct may result in disciplinary action and removal from the register. Registration in one country does not automatically mean that a dietitian can work elsewhere in the world as levels of education and training are not always comparable from country to country. Each registering body will therefore consider applications from dietitians from other countries and suggest further training if appropriate.

Continuing education and demonstration of continuing competence to practice is increasingly being seen as vital in this rapidly changing profession; in the USA there has long been a requirement to demonstrate continuing education and continuing registration is dependent on this. In the UK it is not yet a registration requirement to demonstrate continued competence to practice, although this was considered as part of the recent review of the Act of parliament defining Professions Supplementary to Medicine and is likely to become mandatory now that the Health Professions Council has succeeded the CPSM. However, the BDA, the professional association for dietitians, has recently introduced a Diploma in Advanced Dietetic Practice, which has as part of its remit the recognition of continuing education.

The Role of the Dietitian

The first dietitians (with the exception of those concerned mainly with food service provision) worked mainly in hospitals. Clinical dietetics and the acute hospital service still claim a large proportion of the graduates from dietetics training but other areas of work are increasingly becoming more important. In the UK changes in the emphasis of health care, particularly the change of emphasis from acute (hospital) care to care in the primary health care setting, has resulted in a marked increase in the number of 'community-based' dietitians and the Community Nutrition Group of the BDA has been the fastest-growing specialist group in recent years.

Dietitians also have many other roles outside the health services. In industry they may work as advisors to food companies, wholesale and retail suppliers of food, and companies producing specialized dietary products. They also work with government agencies, e.g., in dietary surveys of the population, in evaluation of intervention programs, and advising on the practical application of policy. In addition dietitians are increasingly working independently as consultants, e.g., in private practice, journalism, and sports nutrition. Whatever aspect of work a dietitian chooses, one of the primary roles will be that of an educator, whether in assisting individuals to understand and apply a therapeutic regimen, teaching doctors, nurses, or other health professionals about nutrition and dietetics to help them carry out their functions, teaching groups of people about aspects of preventive nutrition, or writing an article for the scientific or lay press. The ability to communicate is therefore central to every dietitian's role.

The Dietitian's Role in Food Service

In the USA and countries which follow the US model, hospital dietitians work in either administrative or clinical (therapeutic) areas. Administrative dietitians manage the provision of food services for all patients and staff. They are responsible for food production and quality control in the delivery of the hospital meal service as well as insuring their nutritional adequacy. They are also often responsible for budgeting and staffing of the dietary departments and usually relate to other administrators and managers, having little or no direct contact with patients or medical staff. The clinical dietitian is the person who has direct contact with patients and the medical and paramedical staff involved in their care.

In the UK the practice of dietetics has evolved along a different route and dietitians do not usually have overall responsibility for food service. However, there is usually close liaison between dietitians and the catering manager to insure the provision of nutritionally sound selective menus. The dietetic manager will also be consulted on matters of policy such as the implications of changes in food preparation systems or the introduction of healthy eating policies. It is acknowledged now that many patients (especially elderly people) are malnourished when they enter hospital and that this often becomes worse during their stay. Dietitians are therefore very much involved in attempts to insure that every patient has a nutritionally adequate diet.

Clinical or Therapeutic Dietetics

The term 'therapeutic dietetics' is used to describe the work of the dietitian in his or her direct dealings with patients who require special diets, for various reasons, as outlined in Tables 1 and 2. The role of the clinical dietitian has broadened in the last decade, both with respect to the range of conditions which are encountered and the setting in which the work is done. In the past the role of the therapeutic dietitian was to calculate, teach, and facilitate compliance to a range of dietary regimens prescribed by medical or surgical practitioners for specific disorders. These functions are still important and, indeed, are

Table 1 Examples of the use of therapeutic diets to alleviate symptoms of disease

Symptoms	*Condition*	*Therapeutic diet*
Dysphagia	Disease of mouth or esophagus	Semisolid diet, with thickeners
Weight loss	Trauma, surgery, carcinoma, severe burns	High-energy
Constipation	Diverticular disease, irritable bowel syndrome.	High-fiber
Diarrhea	Pancreatic insufficiency	Low-fat
	Lactose intolerance	Lactose-free
Polyuria, polydipsia	Diabetes	Controlled energy; high-complex-carbohydrate and low-fat

Table 2 Examples of therapeutic diets used to correct disturbances in physiological or metabolic function

Condition	*Cause*	*Therapeutic diet*
Uremia	Renal failure	Moderate-protein, high-energy, low-phosphate
Edema	Liver or renal diseases	Salt restriction
Malabsorption	Pancreatic disease	Low-fat plus MCT
	Cows' milk protein intolerance	Exclude milk and dairy products
Eczema	Food allergy	Remove identified foods, e.g., cows' milk, eggs
Villous atrophy	Celiac disease	Gluten-free
High blood phenylalanine	Phenylketonuria	Controlled phenylalanine intake plus amino acid supplements

extending as improvements in medicine enable patients who would previously have died in childhood to survive to adulthood with continuing needs for nutritional care. For example, patients with cystic fibrosis require nutritional support as their lungs deteriorate and women with hyperphenylalaninemia need dietary advice in order to help them achieve successful pregnancies.

In addition to these roles, however, dietitians are increasingly involved in the assessment and support of patients not traditionally seen as requiring a therapeutic diet. Planning and implementing feeding for patients who are nutritionally compromised as a result of trauma, surgery, or chronic illnesses such as cancer or acquired immunodeficiency syndrome (AIDS) has become an important part of the workload of the dietitian working in both acute and community settings.

Clinical dietitians are increasingly involved in decisions about the appropriateness of particular dietary regimens or the type of nutritional support required and in many institutions are responsible for prescribing the patient's diet in consultation with the physician or surgeon, who recognizes the dietitian as the expert. In acute care a dietitian is an important member of the nutrition support team, working alongside nurses, pharmacists, and medical staff to advise on the feeding of all patients who need nutritional support. (However, in the UK at least, the medical practitioner is ultimately responsible for the patient and any therapeutic regimen, including dietary treatment.)

As new advances occur in clinical practice, the dietitian must be aware of these, and many dietitians are now involved both in research into the development of new treatments and in evaluating current practice. There has been a rapid increase in the numbers of dietetic practitioners with research qualifications and dietitians are increasingly initiating their own projects, working alongside experienced investigators rather than acting as technicians in other people's projects, as sometimes happened in the past. These advances are contributing in some part to a reevaluation of dietetic practice and result in a need for dietitians to justify the evidence-base for therapies and to set priorities in their work. As dietetics has developed it has become more specialized and this has resulted in the formation of practice groups within the dietetic associations. These groups have become a forum where advances in practice in particular areas can be shared and are often a focus for continuing education, both formally (by running courses) and informally. Some of the more specialized roles in which dietitians are working are described below.

Pediatric Dietetics

The pediatric dietitian has a unique role in that he or she has to combine the metabolic requirements of the disease process or condition with the normal requirements for growth and development. With the advances in early diagnosis of many complex metabolic conditions, children may require complicated diets which are very different from the rest of their family and peers, which need constant modification as the child grows and which may be lifelong. The dietitian is responsible for modifying the diet as necessary to take account of the patient's metabolic

requirements, any feeding difficulties, whether mechanical or physiological, and the patient's food preferences and dislikes as he or she grows. The dietitian is an essential part of the support system for children with inborn errors of metabolism such as phenylketonuria and cystic fibrosis, conditions such as renal or heart disease, food allergies, diabetes, and many others, being able to tailor the diet to the patient's specific needs and having access to the information about special foods and products of which the carer might be unaware or have difficulty in locating.

The pediatric dietitian also has an important role as an educator, often teaching the child's parents initially and later the child him-/herself how to cope with the constraints of a special diet, both at home and in school. As is the case with adult patients, the dietitian will often be able to put the child and family in contact with support groups where newly diagnosed patients or parents will be helped by others with first-hand experience of the disease and its treatment.

Renal Dietetics

Renal dietitians are usually attached to specialized renal units and are an integral part of the team involved in the treatment of people suffering from varying degrees of renal impairment, whether acute or chronic. In the USA there is a legal requirement related to funding of patient care which states that a qualified dietitian must be part of the professional team which develops long- and short-term care plans for renal patients. The dietitian, together with the nephrologist, has responsibility for nutritional assessment, the diet prescription, and for monitoring responses to treatment. In addition, he or she must be able to devise appropriate individualized dietary plans, taking into account any other ongoing disease processes or conditions (e.g., diabetes mellitus) and will teach the patient and family how to manipulate the diet. The dietitian will also monitor dietary compliance. A thorough knowledge of physiology and the pathological processes involved in the various kidney diseases and an ability to interpret the patient's biochemical data are therefore essential for the renal dietitian.

Some disorders may resolve with treatment but others may become chronic and result in permanent kidney failure. In progressive renal disease, patients may initially be managed using diet and drugs alone but as the kidneys fail they will require replacement therapy such as peritoneal dialysis, hemodialysis, or transplantation. Each of these stages requires different dietary treatment. The dietitian will deal with a variety of patients with different types of disease and at different stages of progression, with different needs with respect to diet and will also have to teach the patient how to cope with changes in diet which follow as he or she changes from one treatment to another.

Nutritional Support

Nutritional support of patients who are unable to feed themselves adequately by the normal oral route is now an important area practice. The dietitian's involvement here ranges from simply advising on and providing supplements for the patient who cannot eat enough to designing and advising on complete parenteral nutrition regimens for the unconscious patient in intensive care. Between these two extremes will be patients who need enteral feeds to provide complete or supplementary nutrition for a variety of reasons and for periods varying from a few days to a lifetime.

Patients requiring nutritional support may be acutely ill or may require long-term feeding, sometimes at home. The dietitian must be able to assess the nutritional requirements of each individual and design appropriate feeding regimens under all circumstances. Many patients will be sent home whilst still being tube-fed, either enterally or parenterally, and the community dietitian as well as the dietitian in the acute hospital will both be closely involved in the patient's care and in monitoring progress.

Diabetes Care and Education

Traditionally the dietitian always had an important role in the treatment of patients with diabetes mellitus, and the radical changes in dietary treatment over the last 15–20 years have emphasized this role. Since the early 1980s there has been a move away from diets low in carbohydrate and the basis for the treatment of all people with diabetes mellitus, whether young and insulin-dependent, or older and treated by diet alone or diet and hypoglycemic drugs, is now the supply of an appropriate amount of energy as a low-fat diet with at least 50% of energy from foods rich in complex carbohydrates and nonstarch polysaccharides. In addition, the recognition that similar amounts of carbohydrate from different foods have different effects on blood glucose levels has led to less stress on absolute intakes of carbohydrate and a move towards a qualitative approach to the diet.

Many established diabetics found the change in dietary treatment confusing and have needed help in switching to the new regimen. Newly diagnosed patients also need help in learning to manipulate and control this lifelong disorder. Many dietitians have also been involved in the research which has underpinned the progression of dietary treatment and are now evaluating its effects.

HIV and AIDS

The dietetic care of patients with human immunodeficiency virus (HIV) and AIDS has become increasingly important in the last decade. Dietitians may work with people who are HIV-positive to help them optimize their nutritional status and resist the opportunist infections which eventually cause death in the patient with frank AIDS. Once the person develops AIDS, the dietitian's role becomes both therapeutic and palliative, devising and implementing with the patient regimens to satisfy nutritional requirements when, for example, disease of the gastrointestinal tract results in multiple malabsorption or cancer results in weight loss and anorexia. The advent of new multidrug regimens which require careful planning of meals to match their absorption characteristics and which have side-effects affecting nutritional status has made the dietitian's role even more important.

Other Areas of Specialization

There are many other areas of clinical dietetics in which individuals may specialize, including oncology, liver disease, gastroenterology, eating disorders, gerontology, and care of the mentally ill or mentally handicapped. Many of the activities in these areas and those described in the sections above are not confined to the hospital but require input from the dietitian in the community.

Dietetics in the Community

Community dietitians fulfill a variety of roles which may range from working mainly in clinical dietetics in the community setting, e.g., advising a variety of patients in a general practitioner's clinic, to that of public health nutritionist advising the local health or social services on aspects of food policy. In recent years, due to the changing emphasis in health care, the numbers of community dietitians in the UK have increased markedly. The main increase has been in supplying clinical (therapeutic) care to the primary health care setting and about 21% of UK community dietitians surveyed in 1997 stated that this was their only role. About half of the respondents to the survey were involved in health promotion activity as well as clinical work but only 9% worked solely in health promotion.

Many of the dietitians working in health promotion achieve their objectives by educating other professional groups, such as doctors, nurses, health visitors, and midwives, who will then pass the specific knowledge on to the individual or groups of patients. Prevention of diseases which may be diet-related has recently become a much more important issue, and dietitians are working with schools, health education departments and industry to try to educate the public towards consuming a healthier diet. Dietitians also work as advisors in government departments and are therefore involved in planning nutrition policies for the country as a whole.

Dietitians in Research and Education

The advances in all areas of nutritional knowledge, both in terms of achieving optimal health and prevention of disease as well as in therapeutic nutrition, have led to an increasing number of dietitians working in research. The combination of nutritional and medical knowledge and the ability to translate these into terms of foods eaten means that the dietitian has a unique role to play. Dietitians often approach research from a deductive perspective in order to understand or solve difficulties observed in practice, and the results of this research will often have practical significance and can be incorporated into treatments. Evaluation of practice can also be considered as part of this deductive process and is essential in the current health care climate when increasing reliance is put on measuring effectiveness and the use of evidence-based medicine. In addition, dietitians are increasingly involved in basic experimental and analytical scientific research which is essential for nutrition and dietetics to advance in both clinical and nonclinical areas. Involvement in research has led to registration for higher degrees and the number of dietitians with masters' degrees or doctorates is now considerable in countries such as the USA, UK, Canada, and Australia.

Research is also seen as an important part of the role of those dietitians employed in universities and colleges to teach dietetic and other students. In the UK and other countries, there is a requirement that each dietetic training course has registered dietitians on the staff and in many cases these people also work in the National Health Service in order to keep up to date with current practice. Dietitians are also involved in the education of many other professional groups, including nurses, midwives, and pharmacists. Recently, advances have been made in convincing those in charge of medical education at undergraduate and postgraduate level of the importance of nutrition in medical education and this is also seen as an important area where dietitians should be involved.

See also: **Children**: Nutritional Requirements; Nutritional Problems; **Community Nutrition**; **Diabetes Mellitus**: Treatment and Management; Problems in Treatment; **HIV Disease and Nutrition**; **Renal Function and Disorders**: Nutritional Management of Renal Disorders

Further Reading

American Dietetic Association (1993) *The American Dietetic Association Research Agenda for Dietetics: Conference Proceedings*. Chicago, IL: American Dietetic Association.

American Dietetic Association (1994) Position of the American Dietetic Association: nutrition: an essential component of medical education. *Journal of the American Dietetic Association* 94: 555–557.

Bateman EC (1986) *A History of the British Dietetic Association. The Second Twenty-five Years 1936–1986*. Sunderland, UK: Edward Thompson.

Chernoff R (1983) The dietitian in the hospital setting. In: Schneider HA, Anderson CE and Coursin DB (eds) *Nutritional Support in Medical Practice*, pp. 70–78.Philadelphia: Harper & Rowe.

Department of Health (1994) *Targeting Practice: The Contribution of State Registered Dietitians*. Health of the Nation. London: HMSO.

Department of Health (1994) *Nutrition Core Curriculum for Nutrition Education of Health Professionals*. London: HMSO.

Fox C (1999) *Community Dietetics: Supporting the Future*. Community Nutrition Group of the British Dietetic Association. Birmingham, UK: BDA.

Judd PA (1996) Educating the dietitians of the future. *Journal of Human Nutrition and Diet* 9: 333–338.

Judd PA (1997) Pre-registration training for dietitians – report of the Dietitians Board/BDA working group on pre-registration training. *Journal of Human Nutrition and Diet* 10: 157–162.

Disinfectants *See* **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant; Overall Approach; Modern Systems

Distilled Beverage Spirits *See* **Vodka**; **Gin**: The Product and its Manufacture; Composition and Analysis; **Whisky, Whiskey, and Bourbon**: Products and Manufacture; Composition and Analysis of Whisky; **Rum**; **Vermouth**

Distribution *See* **Transport Logistics of Food**

Down Syndrome *See* **Developmental Disabilities and Nutritional Aspects**: Down Syndrome; Prader–Willi Syndrome

DRESSINGS AND MAYONNAISE

Contents

The Products and Their Manufacture

S C Yang and L S Lal, Providence University, Taiwan, Republic of China

Introduction

Salad dressings are liquid or semiliquid products used to flavor, moisturize, and enrich salads. In addition, they may be used as spreads for sandwiches or as ingredients in a variety of cold and hot recipes.

A two-phase emulsion may consist of oil droplets in water phase (i.e. oil-in-water emulsion) or water droplets in oil phase (water-in-oil emulsion), as shown in Figure 1. Traditional salad dressings are oil-in-water emulsions. Droplets of oil, the dispersed phase, are suspended in a continuous aqueous acidic phase such as vinegar or lemon juice. Other ingredients are added to modify the flavor or texture. Normally oil and water do not combine well, but separate into layers. However, the two phases must remain evenly dispersed in an acceptable salad dressing. (*See* **Colloids and Emulsions**.)

Salad dressings can be divided into four categories on the basis of their formulations and stability:

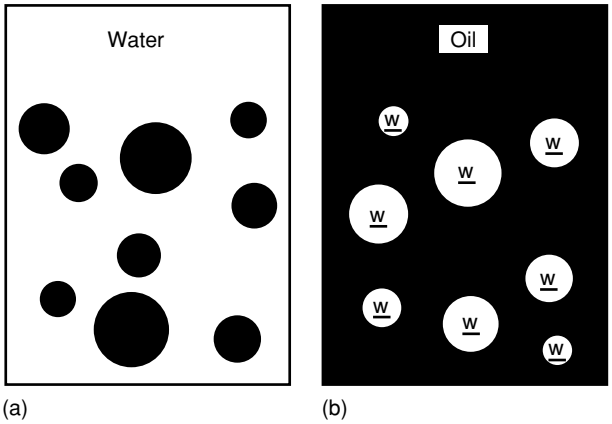

Figure 1 A two-phase emulsion may consist of (a) oil droplets in water (o/w) or (b) water droplets in oil (w/o).

1. oil and vinegar dressings (most are temporary emulsions);
2. emulsified dressings (most are permanent emulsions);
3. cooked salad dressings (most are emulsions containing a starch thickener);
4. low-fat/diet products (most are emulsions containing hydrocolloids for viscosity and texture characteristics).

Each type of salad dressing must conform to certain specifications; in the USA each has a standard of identity established by the Food and Drug Administration (FDA).

Typical Examples and Their Uses

Oil and Vinegar Dressings

Basic French dressing and variations such as mustard French, herbed French, Italian, piquante, chiffonade, and avocado dressings are examples of temporary oil and vinegar emulsions. The ingredients include vinegar, oil, and various dry ingredients. In these products the liquid phase separates after mixing, although dry ingredients, such as mustard and paprika, collect at the oil–water interface to give limited stability to the emulsion. These dressings should be mixed thoroughly immediately before they are used. The harder the mixture is beaten or shaken, the longer it takes for it to separate again.

Emulsified Dressings

The classic example of an emulsified dressing is mayonnaise, a permanent oil-in-water emulsion. Mayonnaise is made from a vegetable oil, vinegar and/or lemon juice, and whole egg or egg yolk. Additional seasonings such as mustard, paprika, and salt are usually added. By legal definition in many countries (e.g., the USA), commercial mayonnaise must contain either egg yolk or whole egg and must have an oil content of not less than 65% by weight. Whole egg or egg yolk functions as an effective emulsifier and creates a permanent emulsion. The rather high percentage of oil contributes significantly to the viscosity of the dressing. The viscous nature of permanent

emulsions makes it difficult for the oil drops to come into direct contact and coalesce. (*See* **Emulsifiers**: Organic Emulsifiers.)

Emulsified French dressing is similar to basic French dressing, except that egg yolk is added to prevent the oil and vinegar from separating. Its preparation is similar to that of mayonnaise. Mayonnaise is sometimes used by itself as a salad dressing but more often serves as the base for a variety of other dressings. Thousand Island, Lousis, Russian, Chantilly, and blue cheese dressings are all mayonnaise-based. These dressings are generally thick and creamy, and many of them contain sour cream.

Cooked Salad Dressings

Cooked salad dressings resemble mayonnaise in appearance but differ in some of the ingredients and in the method of preparation. The basic ingredients are vegetable oil, vinegar or lemon juice, whole egg or egg yolk, milk or water, and a cooked starch paste. Homemade cooked salad dressings are usually thickened with starch and/or egg. Very little, if any, oil is used in these dressings. However, a commercial salad dressing, labeled as such, must meet the standard of identity of the FDA and contain by weight no less than 30% vegetable oil and no less than 4% liquid egg yolk or the equivalent. Although these dressings are often very similar in appearance to mayonnaise, they cannot be labeled mayonnaise because of the lower oil and egg yolk content, plus the fact that they may contain cooked starch. The reduced oil content has the advantage of making this product lower in calories than mayonnaise.

Low-Fat/Diet Products

The recent change in dietary consciousness has affected the type of food consumers select at supermarkets and in restaurants. In 1985, the size of reduced-fat or reduced-calorie sauces and dressings grew faster than the total market. In 1986, this category of food grew at an annual rate of 6% but reduced-calorie products achieved 23% growth. The driving forces were the increased interest in nutritious foods and dietary fiber, less consumption of cholesterol, and a marked increase in vegetarianism.

By moving low-calorie/reduced-fat out of the diet section of supermarkets and repositioning them with regular, standard products, producers improved the image of these products. The low-calorie/reduced-fat products are now recognized by consumers not as diet foods but as nutritious, lower-calorie counterparts to the standard products, with good sensory properties.

Reduced-fat dressings and mayonnaise require processing similar to traditional products and a one-third reduction of oil content. The oils are the same as for the full-fat products. Generally, soya bean oil is used. Gums, particularly xanthum and alginates, are incorporated to stabilize the formula, and water replaces the fat removed. These products rely totally on hydrocolloids for viscosity and texture characteristics.

Food technologists have developed ingredients that can be used to replace or lower the fat content of food products. One such ingredient is Simplesse, a physically modified combination of milk and egg-white proteins processed by an internationally patented heating and blending procedure called microparticulation. Simplesse possesses fat-like characteristics while providing significantly fewer calories than fat for such items as mayonnaise and salad dressings. Other recently developed fat substitutes include Olestra, a sucrose polyester, and Veri-Lo, a creamy white emulsion containing milk fat. All substances must undergo review and gain official approval before they can be used. (*See* **Fats**: Fat Replacers.)

Reduced-sodium mayonnaise and salad dressings are also of interest to health-conscious consumers. The market has grown steadily since 1980, and is expected to make up 23% of the total sauces and dressings market by 1995. Reduced-sodium mayonnaise is made with salted egg yolk containing potassium chloride rather than sodium chloride. Frozen salted egg yolk made with potassium chloride is less viscous than that containing sodium chloride. However, viscosity and stability of mayonnaise made from 10% salted (potassium chloride) egg yolk are acceptable. (*See* **Hypertension**: Hypertension and Diet; **Sodium**: Physiology.)

Continued growth in the area of low-calorie, reduced-fat and reduced-sodium mayonnaise and salad dressings is anticipated. The aging population, combined with increased nutritional awareness, will create a market for such modified products.

Formulations

Oil and Vinegar Dressings

French dressing is prepared from edible vegetable oil and the following acidifying and flavoring ingredients:

1. salt;
2. nutritive carbohydrate sweeteners, such as sucrose, dextrose, corn syrup, glucose syrup, or honey;
3. mustard, paprika or other spices, spice oils or spice extracts;
4. any suitable, harmless food seasoning or flavoring (other than imitation);
5. tomato paste, tomato purée, catsup, sherry wine;
6. any vinegar or vinegar diluted with water to an acidity of not less than 2.5% by weight calculated

as acetic acid, or any such vinegar or diluted vinegar mixed with an additional optional acidifying ingredient such as citric acid, weighing no more than 25% of the weight of the acids of the vinegar or diluted vinegar calculated as acetic acid (lemon juice is another acidifying ingredient which may be used in place of vinegar).

French dressing contains not less than 35% by weight of vegetable oil. The ratio of oil to vinegar in basic French dressing is 3:1. Some chefs prefer a 2:1 ratio. Less oil makes the dressing more tart, while more oil makes it milder in flavor and oilier. For emulsified French dressing, eggs or other emulsifiers (vegetable gum, propylene glycol esters of alginic acid, methyl cellulose, pectin, xanthan gum, etc.) may be added, the total content being no more than 0.75% by weight. Eggs used for French dressing can be fresh, frozen, or dried.

Formulae for commercially prepared dressings and mayonnaise vary according to the specific plant and equipment employed. Also, taste preference is different from region to region. Herein, typical formulae for certain commercial products are presented.

A typical commercial French dressing formula may contain:

Water	12.0%
Vinegar (cider) 5%	22.0%
Sugar	8.8%
Salt	3.5%
Worcestershire sauce	3.5%
Paprika	4.5%
Mustard, garlic, onion	4.7%
White pepper	0.3%
Vegetable oil	40.0%
Gum tragacanth	0.4%
Dried egg-yolk solids	0.3%

In this formula, the total solids of commercial liquid egg yolk is about 44%. Dried egg yolk may be substituted for commercial liquid egg yolk on a total solids basis. French dressing may be mixed and packed in an inert atmosphere in which air is replaced totally or in part by carbon dioxide or nitrogen.

Emulsified Dressings

Mayonnaise and mayonnaise dressing are emulsified semisolid food products prepared from edible vegetable oil, egg yolk-containing ingredients and the following acidifying and flavoring ingredients:

1. salt;
2. nutritive carbohydrate sweeteners, such as sucrose, dextrose, corn syrup, glucose syrup, or honey;
3. mustard, paprika or other spices, spice oils or spice extracts, except turmeric or saffron, and no spice oil or spice extract which imparts to the mayonnaise a color simulating that imparted by egg yolk;
4. any suitable, harmless food seasoning or flavoring (other than imitations), provided it does not impart to the mayonnaise a color simulating that imparted by egg yolk;
5. monosodium glutamate;
6. acidifying ingredients, which may be any vinegar of not less than 2.5% acetic acid, or frozen, canned, concentrated or dried lemon or lime juice with a total acidity of no less than 2.5% calculated as acetic acid; citric or malic acid may be used in a proportion not to exceed 2.5% of the acid of the vinegar calculated as acetic acid.

A typical commercial mayonnaise formula may contain:

Vegetable oil	78.5%
Vinegar, white (10%)	3.8%
Water	11.6%
Sugar	1.8%
Salt	1.2%
Spices (mustard, onion, etc.)	0.3%
Dried egg yolk solids	2.8%

This formula provides a mayonnaise with medium viscosity. For a heavier viscosity, vegetable oil may be increased to 80.5% and water decreased to 9.6%. Mayonnaise may be mixed and packed in an inert atmosphere in which air is replaced totally or in part by carbon dioxide or nitrogen. Dried egg-yolk solids may be substituted for liquid egg yolk on a total solids basis.

Cooked Salad Dressings

Salad dressing is the emulsified semisolid product prepared from ingredients such as those of mayonnaise plus cooked (or partially cooked) starch paste prepared with a food starch, tapioca, wheat or rye flour, or any two or more of these. Salad dressing under the FDA standards of identity resembles mayonnaise in that it is an emulsion of oil in vinegar with egg as an emulsifier. It differs from mayonnaise in that it also contains starch paste as a thickener. Salad dressing may be flavored with one or more of the following ingredients:

1. salt;
2. nutritive carbohydrate sweeteners;
3. mustard, paprika or other spices, spice oils or any suitable harmless food seasoning or flavoring ingredient providing they do not impart a color simulating that imparted by egg yolk;
4. monosodium glutamate.

Salad dressing contains not less than 30% by weight of vegetable oil and no less egg yolk-containing ingredient than is equivalent to 4% by weight of the yolk solids of liquid egg yolk. Salad dressing may also contain not more than 0.75% by weight of any of several polysaccharide gums or their cellulose derivatives, methylcellulose and sodium carboxymethylcellulose; gums are not permitted in mayonnaise. The acidifying ingredients may be any vinegar or frozen, canned concentrated or dried lemon juice. Citric or malic acid may be used in a proportion not to exceed 25% of the total acid of the vinegar calculated as acetic acid. A typical commercial salad dressing formula may contain:

Vegetable oil	30.0%
Starch paste mixture	68.2%
Dried egg-yolk solids	1.8%

Dried egg-yolk solids may be substituted for liquid egg yolk on a total solids basis.

The starch paste is made by cooking and cooling a combination of ingredients in the following proportions:

Vinegar (10%)	17.2%
Salt	4.0%
Starches	7.7%
Sugar	15.0%
Spices (mustard, celery, etc.)	1.1%
Water	55.0%

The starch may be a food starch, modified food starch, tapioca, wheat, rye flour, or any two or more of these. Water may be added in the preparation of the paste. Salad dressing may be mixed and packed in an inert atmosphere in which air is replaced in total or in part by carbon dioxide or nitrogen.

Low-Fat/Diet Products

The Nutrition Labeling and Educational Act (NLEA) of 1990 standardized the nutrient content claims on food labels in the USA. Additionally, the NLEA standardized the serving sizes for each product. The US regulations for foods are published in Code of Food Regulations 21 CFR Part 101 each year. The three claims of particular interest here are 'reduced,' 'low,' and 'free' as applied to calories, sodium, fat, saturated fat, and cholesterol. In the USA, the term 'reduced' refers to a nutritionally altered product that contains at least 25% less of a nutrient or 25% fewer calories than a reference food. The term 'low' refers to a reference amount (or 50 g of food, if the reference amount is small) that contains ≤ 40 calories, ≤ 140 mg sodium, ≤ 3 g fat, ≤ 1 g saturated fat, and $\leq 15\%$ of calories from saturated fat, or ≤ 20 mg of cholesterol. The term 'free' refers to a serving and the reference amount contains no or physiologically inconsequential amount, specifically, < 5 cal, < 5 mg sodium, < 0.5 g fat, < 0.5 g saturated fat and < 0.5 g *trans* fatty acids, < 2 mg cholesterol, or < 0.5 g sugars. Standards may vary by country.

Various fat-replacers, including starch-derived fat mimetics, fiber-based fat mimetics, microparticulated proteins, hydrocolloid gums, and synthetic fat mimetics, have been successfully applied to reduced-fat, low-fat, and fat-free salad dressings. Formulation of a low-fat salad dressing containing 5% oil may contain:

Microcrystalline cellulose	3.2%
Vegetable oil	4.3%
Xanthan gum	0.3%
Modified starch	3.6%
Sugar	5.8%
Egg yolk	4.0%
Vinegar	4.5%
Lemon juice	0.5%
Salt	1.8%
Water	72.0%

The reduced-fat, low-fat, and fat-free salad dressing formulations are generally designed to impart a full-fat organoleptic profile, including texture, appearance, mouth feel, and body.

Major Ingredients in the Formulae

Vegetable oil Oil is the major ingredient contributing to the viscosity and body of mayonnaise. The rigidity of the emulsion depends partly on the size of the oil droplets and partly on how tightly they are packed. The more oil that is dispersed in the emulsion, the stiffer it will be. The choice of oil is governed by quality and local preference. It should be bland. Strongly flavored oils can make excellent dressings but are not appropriate with every food. Today, soya bean, cottonseed, sunflower, and corn oils are used commercially. They may be used singly or in combination. Deodorized, winterized oil should be used with dressings that are to be refrigerated. (*See* **Soy (Soya) Beans**: The Crop.)

Vinegar The functions of vinegar are as a preservative and a flavoring. Acid is the main preservative against microbial spoilage in mayonnaise. The industry makes use primarily of two types of vinegar, namely cider vinegar and distilled vinegar. The ideal vinegar for salad dressing and mayonnaise is one which is mild in odor and taste. Vinegar strength is measured by 'grain', 100 g equalling 10% acetic acid. Most salad vinegars range from 3 to 5% and should be standardized by dilution with water before measuring out for a recipe. (*See* **Vinegar.**)

Cider, malt and wine vinegars are more costly than distilled vinegar, but have unique flavors which contribute character to mayonnaise. Distilled vinegar is used when a neutral flavor is desired. Lemon juice may be used in place of, or in addition to, vinegar when its flavor is desired.

Egg Egg yolk is the emulsifying agent most often used in the home for preparing emulsions requiring stability. The fact that egg yolk is itself an emulsion explains, at least in part, why it is such an effective agent. (*See* **Stabilizers**: Types and Function.)

Eggs for mayonnaise may be fresh, frozen, or dried. Fresh broken eggs usually make a weak-bodied mayonnaise, although the product may stiffen somewhat during storage. Upon freezing and storing raw egg yolk below $-6\,^{\circ}C$, its viscosity increases and gelation occurs. This irreversible gelation of egg yolk makes it useless for mayonnaise production. The addition of 10% salt (sodium chloride) or 10% sugar will partially inhibit the gelation. Mayonnaise made from frozen salted egg yolk is thick and creamy. Dried yolk disperses readily in the aqueous phase of mayonnaise, and results in a thicker product than that obtained from frozen egg when prepared with the same total egg solids content.

Egg yolk is the primary source of color in mayonnaise. The intensity and tone of the yellow color of the yolk determine the color of the finished product, because no other coloring matter is permitted. In mayonnaise formulation, sequestrants, including but not limited to calcium disodium ethylenediaminetetra-acetate (EDTA) and/or disodium EDTA, may be used to preserve color.

Pasteurization of egg products in the USA became mandatory on 1 June 1966 in order to eliminate problems with salmonella. Eggs used for dressings and mayonnaise must meet this requirement.

Biopolymer Proteins and polysaccharides, such as soy protein isolate, whey protein concentrate, xanthan gum, sodium alginate, propylene glycol alginate, locust bean gum, guar gum, gum arabic, gum acacia, starch, modified starch and microcrystalline cellulose, are the two most important types of biopolymers used as ingredients in mayonnaise and salad dressings to enhance the texture, replace part or all of the fat, and enhance the stability of emulsions.

Proteins can form a film around the surface of fat droplets through three stages: (1) native protein molecules diffuse to the interface; (2) protein penetrates the interface; and (3) molecules rearrange to achieve minimum energy. Such actions not only reduce the interfacial tension, but also form a physical barrier to reduce the possibilities of flocculation and coalescence of fat droplets. Gums and/or starches are added to dressings possibly due to their thickening and/or surface-active effect. Gums and/or starches may improve the emulsion stability of the dressings by making the continuous phase more viscous, reducing the possibilities of fat droplets encountering each other, and creating a yield stress which increases emulsion stability by minimizing or eliminating creaming of the fat droplets under gravity force.

Mustard Mustard flour is probably used in more dressing formulae than any other spice. It is used chiefly for flavor but contributes slightly to emulsification. Oil of mustard can be used in place of mustard flour without affecting the emulsion strength. Oil of mustard must be obtained from mustard seed. Synthetic allyl isothiocyanate is not permitted. (*See* **Mustard and Condiment Products**.)

Procedures for Manufacture

Mayonnaise is prepared commercially by both batch and continuous methods, and there are many variations with respect to order and rate of ingredient additions. All equipment should be fabricated from stainless steel since vinegar will corrode ordinary steel and aluminum. Two-stage mixing is commonly employed with high-speed turbine blades in the first stage followed by more severe shearing of the oil into fine droplets in the second stage. The second-stage mixer may possess close-clearance whirling teeth as in a colloid mill.

The Dixie–Charlotte system is probably the one most widely used for the production of mayonnaise and salad dressings. The system consists of two Dixie mixers and a Charlotte colloid mill connected by appropriate piping, valves, and a rotary positive displacement pump. It is arranged so that one mixer feeds premix to the mill while another batch of premix is being prepared in the other mixer. The Dixie mixer is a deep circular tank fitted with three turbine mixers mounted side-by-side on a horizontal shaft close to the tank bottom. The shaft is turned by a variable-speed motor. This mixer can prepare mayonnaise requiring no further processing. The emulsion will be similar to one made with a Hobart mixer. The creamy texture of modern mayonnaise is obtained by pumping this relatively soft, coarse emulsion through the Charlotte mill. This equipment is available in a number of sizes with capacities ranging from 60 to 760 l per batch and yielding 230–4540 l of finished mayonnaise per hour. The system is arranged on a continuous-batch basis and can be modified to be operated as a truly continuous system.

Preparation of mayonnaise premix in the Dixie mixer is started by adding mayonnaise from a previous run to a level which will reach the mixer shaft. This gives the turbine blades a heavy material to work against while shearing the oil into fine droplets. Egg and dry ingredients are then added to the mixer with low-speed agitation. Oil and vinegar are then pumped or fed by gravity from supply tanks into the mixer. Agitation speed is usually increased periodically as the level of the premix rises. If the oil is added too slowly the premix will be thick and only partially emulsified. The subsequent passage through the colloid mill causes an overemulsification, resulting in a less viscous and stable mayonnaise. If the oil is added too quickly, in relation to agitator speed, the premix will develop an oily and curdled appearance; and, if the rate is not reduced, the emulsion will reverse to the water-in-oil type.

The premix is pumped through the Charlotte mill as soon as it is made. Timing should be such that the alternate premix tank should have just finished emptying. This is critical with many mayonnaise formulations as the premix can gel in the dixie mixer. The longer the premix is held in the mix tank prior to milling, the softer the final product will be.

Since salad dressings are similar to mayonnaise except for the lower oil content and the added starch paste, the ingredients function as in mayonnaise, and principles with respect to mixing and emulsion stability are similar. Special care must be given to starch cooking to develop the desired degree of thickening.

Raw starch is not water-soluble, but can be dispersed in cold water. It requires cooking to form a thickened suspension or paste. The sugar, salt, and vinegar are also part of the paste. The process calls for cooking and cooling the paste before combining it with the egg and oil emulsion. In batch processing, some of the paste is usually added to the egg before any oil is incorporated. This weakens the egg. The oil is thereby prevented from forming too tight an emulsion. The oil is not whipped in as vigorously as with mayonnaise. When the balance of the starch paste is mixed in after the oil is added, the entire mass must be fairly soft so that it can be pumped. The colloid mill yields a salad dressing with high viscosity and smooth texture. If the premix is too stiff, it will break down in the mill.

A batch-type starch paste cooker is essentially a jacketed vertical stainless-steel tank equipped with single- or double-acting agitators having side-scraping blades. A similar vessel is used for cooling the paste in which cold water or brine solution is used as the cooling medium. The most modern method for processing starch paste is a continuous one employing the concept of energy regeneration resulting from the use of a plate heat exchanger.

No matter what process is employed, starch paste should be used as soon as possible after preparation. The longer it stands after cooling, the more readily it will break down on being worked into the dressing. This is especially true with high-amylose starches. Cooking starch in vinegar solution tends to degrade or hydrolyze the starch to some extent. Modified starches are usually more resistant to breakdown than unmodified types. Where feasible, it is best to cook the starch with a minimum amount of vinegar in the water solution. (*See* **Starch**: Modified Starches.)

The continuous method employed in the manufacture of mayonnaise and salad dressing provides continuous automatic control so that one worker can process up to 75701 of mayonnaise, salad dressing, and pourable dressings per hour. One operator can monitor the entire system. Manual operation is almost nonexistent now.

Formulae are monitored from a control center. In the case of salad dressings, the system includes four probe-controlled surge tanks: one for oil, one for starch, one for egg, and one for the brine additive. Each of the surge tanks is probed at a high and low level so that, when a low level is reached, a signal is sent to a pump, and the tank will be resupplied with oil, starch, or egg until it reaches the high point. This maintains a constant supply of ingredients. Should any one of the tanks become empty, the entire system halts. Each controlled surge tank leads to a metering pump. Each pump is monitored from the control center so that it delivers the desired volume within 0.5%.

By setting the predetermined speed of each pump, the egg, brine, oil, and starch are metered accurately at the proper flow rate ratio to the inline mixer or preemulsifier where the ingredients are thoroughly mixed and a preemulsion is created. The preemulsion is then pumped with a sanitary pump to a colloid mill, where an extremely fine globule emulsion is created, and then on to the filler. Using this system, up to 25 different varieties of dressing can be made, all within the accuracy of 0.5% in the formulation.

See also: **Colloids and Emulsions**; **Emulsifiers**: Organic Emulsifiers; **Fats**: Fat Replacers; **Hypertension**: Hypertension and Diet; **Mustard and Condiment Products**; **Soy (Soya) Beans**: The Crop; **Stabilizers**: Types and Function; **Starch**: Modified Starches; **Vinegar**

Further Reading

Anonymous (1975) *Instructions for the Use of Dried Egg Yolk in Mayonanise, Salad Dressing and French Dressing*. New York: Henningsen Foods.

Ford LD, Borwankar R, Martin RW, Jr. and Holcomb DN (1997) Dressing and sauces. In: Friberg SE and Larsson K

(eds) *Food Emulsions*, 3rd edn, pp. 361–412. New York: Marcel Dekker.
Friberg SE (1997) Emulsion stability. In: Friberg SE and Larsson K (eds) *Food Emulsions*, 3rd edn, pp. 1–55. New York: Marcel Dekker.
Gorman JM and Ball HR, Jr. (1986) Quality control and product specifications. In: Stadelman WJ and Cotterill OJ (eds) *Egg Science and Technology*, 3rd edn, pp. 325–343.Westport: AVI.
Humphreys WM (1996) Fiber-based fat mimetics: microcrystalline cellulose. In: Roller S and Jones SA (eds) *Handbook of Fat Replacers*, pp. 131–144. New York: CRC Press.
Lopez A (ed.) (1981) Mayonnaise and salad dressing products. In: *Processing Procedures for Canned Food Products, A Complete Course in Canning* 11th edn, book II, pp. 380–395. Baltimore: Canning Trade.
McClements DJ (ed.) (1999) Emulsion ingredients. In: *Food Emulsions. Principles, Practice, and Techniques*. pp. 83–125. Boca Raton, Florida: CRC Press.
McWilliams M (ed.) (1985) Salads and salad dressings. In: *Food Fundamentals*, 4th edn, pp. 99–119. New York: Wiley.
Orthoefer FT (1988) Reduced-fat margarines and dressings. *Journal of the American Oil Chemists' Society* 65: 574–578.
Roller S (1996) Starch-derived fat mimetics: maltodextrin. In: Roller S and Jones SA (eds) *Handbook of Fat Replacers*, pp. 99–129. New York: CRC Press.
Tressler DK and Sultan WJ (eds) (1975) Standards of identity for mayonnaise and *French* dressing. In: *Food Products Formulary*, vol. 2, pp. 394–395. Westport: AVI.

Chemistry of the Products

S C Yang and L S Lal, Providence University, Taiwan, Republic of China

Introduction

Food emulsions, such as mayonnaise and salad dressings, are two-phase systems of immiscible liquids with limited stability. One phase is in the form of finely divided droplets of diameters generally larger than 0.1 μm. This dispersed, internal, or discontinuous phase is suspended in the continuous or external phase. Emulsion systems can be divided into two categories: (1) those consisting of droplets of oil dispersed throughout an aqueous medium, which are usually referred to as oil-in-water emulsions; (2) those in which droplets of water are dispersed throughout an oil or fat medium, which are termed water-in-oil emulsions. Most food emulsions are of the oil-in-water type. Emulsion properties generally depend on the nature of the continuous phase and the proportion of this phase to the dispersed phase. (*See* **Colloids and Emulsions**.)

Standards of identity, in the USA, for mayonnaise define the product as a semisolid emulsion made of egg yolk, edible vegetable oil, and acetic or citric acid. It may also contain salt, spices or spices oils, natural sweeteners and various natural flavoring ingredients. The oil content must be not less than 65% by weight, and the product must contain at least 2.5% acetic acid by weight. Citric acid in the form of lemon or lime juice may replace the acetic acid at a minimum level of 2.5%. The egg yolk may be in liquid, frozen, and/or dried forms. This ingredient provides emulsifying properties and gives the mayonnaise a pale yellow color, which must not be intensified by any other ingredient. Salad dressings are also oil-in-water emulsions, but the concentration of oil is only about 30–45% as compared to 65–82% in mayonnaise. In addition to yolk, stabilizers are permitted in this product. Commercial products of reduced-fat mayonnaise and salad dressings usually contain less than half fat content as compared to regular product. Fat-free products should contain no or physiologically inconsequential amount of fat (< 0.5 g fat) per serving. All reduced/low-fat or fat-free products generally rely heavily on hydrocolloids for viscosity and texture characteristics. **Table 1** shows the composition of selected types of dressings and mayonnaise. Many variations in formulations exist within each type of product. (*See* **Emulsifiers**: Organic Emulsifiers; Uses in Processed Foods; **Stabilizers**: Types and Function; Applications.)

Physical Structure and Rheology

Reduction of interfacial tension is probably the first step in the formation of an emulsion. Surface-active agents form a film around the oil globules and prevent their coalescence. In surrounding the oil droplets, the lipophilic (oil-loving) portion orients itself in the aqueous phase, forming a shell around the droplets of the dispersed phase. By collecting itself at the interface, the emulsifier prevents the dispersed particles from coalescing and separating out, thereby increasing the emulsion's stability.

The microstructure of mayonnaise, as revealed by the scanning electron microscope, contains fat globules of various sizes (**Figure 1a**). Both the dispersed fat and continuous phase are evident. The continuous phase binds tightly to the fat globules (**Figure 1b**). Several very small fat globules can be seen adhering to the large globules. Under the high magnification of

Table 1 Composition of selected types of dressings and mayonnaise

Nutrient	*Amount per 100 g*							
	Blue and Roquefort (regular)	*Cooked (homemade)*	*French (regular)*	*Italian (regular)*	*Russian (regular)*	*Mayonnaise (soybean)*	*Mayonnaise (low-cal)*	*French (low-fat)*
kcal	504	156	430	467	494	717	136	10
Moisture (%)	32.3	69.2	38.1	38.4	34.5	15.3	80.7	95.2
Protein (g)	4.8	4.2	0.6	0.7	1.6	1.1	1.1	0.4
Fat (g)	52.3	9.5	41.0	48.3	50.8	79.4	12.7	0.2
Carbohydrate (g)	7.4	14.9	17.5	10.2	10.4	2.7	4.8	1.8
Fiber (g)	0.1	0	0.8	0.2	0.3	0	0.5	0.3
Vitamins								
A (IU)	210.0	411.0			690.0	280.0	220.0	
C (mg)	2.00	0.60			6.00			
Thiamin (mg)	0.01	0.06			0.05	0	0.01	
Riboflavin (mg)	0.10	0.15			0.05	0	0.03	
Niacin (mg)	0.10	0.25			0.60			
Folic acid (μg)						3.00		
Minerals								
Calcium (mg)	81.0	84.0	11.0	10.0	19.0	18.0	18.0	11.0
Phosphorus (mg)	74.0	87.0	14.0	5.0	37.0	28.0	28.0	14.0
Sodium (mg)	1094.0	734.0	1370.0	787.0	868.0	568.4	118.0	787.0
Magnesium (mg)			10.0			2.0		10.0
Potassium (mg)	37.0	121.0	79.0	15.0	157.0	34.0	9.0	79.0
Iron (mg)	0.20	0.5	0.40	0.20	0.60	0.50	0.2	0.4
Zinc (mg)		0.11	0.08	0.11	0.43	0.16		0.18

Data pooled from Ensminger AH, Ensminger ME, Konlande JE and Robson JRK (1994) Food composition. In: *Food and Nutrition Encyclopedia*, 2nd edn, vol. 1. Boca Raton, FL: CRC Press.

transmission electron microscopy, part of the interfacial film is visible surrounding the lipid droplets of diluted mayonnaise (Figure 2). The electron-dense material at the interface is the emulsifying material from egg yolk. This material, which acts as a surface-active agent, consists of complexes of egg yolk protein fractions such as lipovitellin, livetin, and lipovitellenin. (*See* **Eggs**: Structure and Composition.)

In addition to functioning as an emulsifying agent, egg yolk contributes viscosity to mayonnaise. Increasing the amount of egg yolk from 5 to 12% increases the thickness of the product. Viscous emulsions tend to be more stable. However, amounts of yolk in excess of 12% are not associated with a change in consistency of the mayonnaise. It is essential that there is a sufficient amount of aqueous phase (usually vinegar) to surround oil droplets as they become smaller and expose more surface during emulsification. However, an excess of aqueous phase is not conducive to a good emulsion.

Since salad dressing contains less oil than mayonnaise, there are fewer lipid droplets, and a higher concentration of amorphous material between droplets. The amorphous material is assumed to be cooked starch paste, a stabilizing ingredient added to salad dressing but not to mayonnaise. Starch paste also reduces the rate of lipid droplet coalescence in salad dressing. (*See* **Starch**: Modified Starches.)

Sensory attributes of dressing emulsions, such as creaminess, thickness, smoothness, spreadability, and pourability, are related to their rheological properties. The shelf-life of dressing emulsions also depends on their rheological characteristics. Viscosities of dressing emulsions depend on the shear rate applied. Pseudoplastic flow is the most common type. It manifests itself as a decrease in the apparent viscosity as the shear rate is increased and is therefore often referred to as shear thinning. Viscosities of dressing emulsions also depend on the shear time. Thixotropic behavior is the most common type, in which the apparent viscosity decreases with time when subjected to a constant shear rate. Dressing emulsions that exhibit this type of behavior often contain droplets aggregated by weak forces. Shearing of the material causes the aggregated droplets to be progressively deformed and disrupted, which decreases the resistance to flow and therefore causes a reduction in viscosity over time. For example, pourable salad dressings, such as French varieties, Thousand Island, and Creamy Italian, are primarily viscous in their flow behavior, but exhibit varying degrees of thixotropy and often a measurable yield stress as well.

Some dressing emulsions are not purely viscous or elastic, but instead partly viscous and partly elastic. For example, mayonnaise and spoonable salad dressings show viscoelastic rheological behavior and also possess a yield stress. The viscoelastic behavior

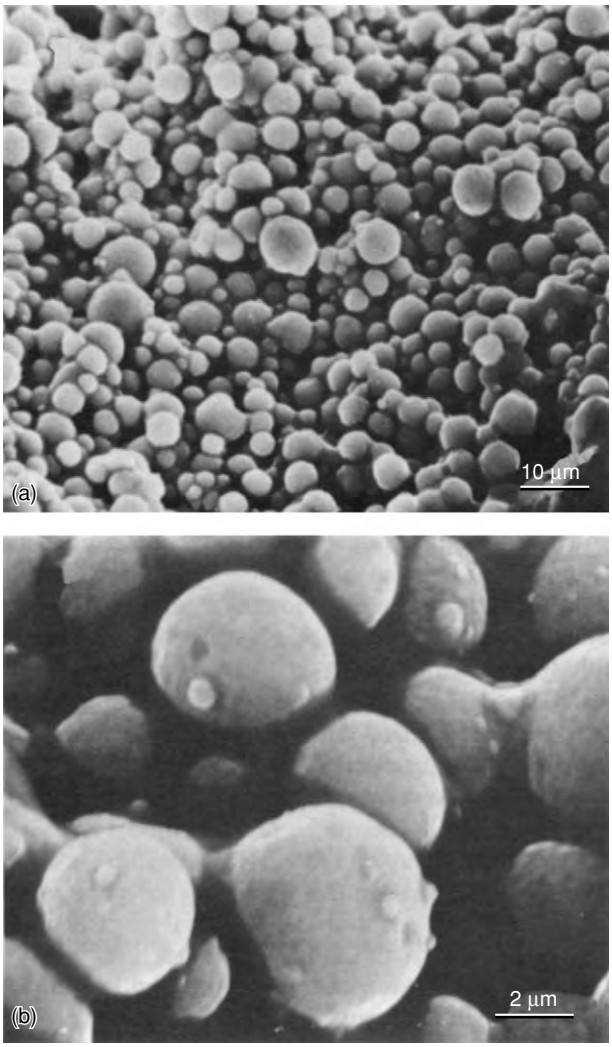

Figure 1 (a) The scanning electron microstructure of mayonnaise reveals fat globules of various sizes. (b) Higher magnification of part (a) showing the continuous proteinaceous phase bound tightly to the fat globules. Reproduced from Dressings and Mayonnaise: Chemistry of their products, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

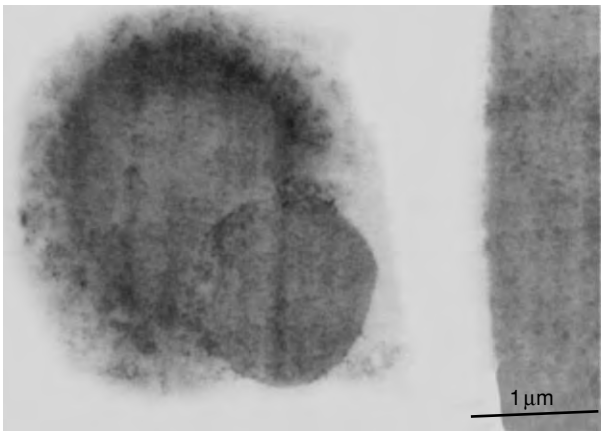

Figure 2 The transmission electron micrograph of mayonnaise. The interfacial film is visible surrounding the lipid droplets. The electron-dense material at the interface is the emulsifying material from egg yolk. Courtesy of *Scanning Electron Microscopy*, Inc., Illinois.

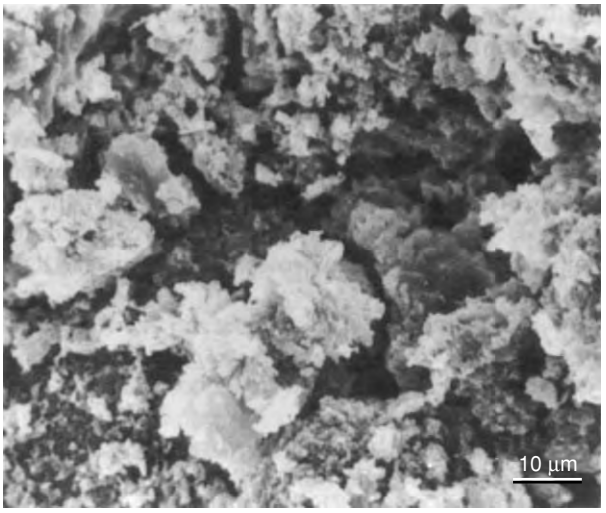

Figure 3 The broken structure of the lower layer of mayonnaise resulting from freezing and thawing. Reproduced from Dressings and Mayonnaise: Chemistry of their Products, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

could be characterized through creep tests or dynamic tests. Creep test consists of the application of a constant stress and ensuing study of the time-dependent behavior of the strain. Dynamic experiment consists of the application of a sinusoidal stress and the resulting sinusoidal strain is measured or vice versa. The amplitude of the applied stress or strain used in this type of test is usually so small that the material is in the linear viscoelastic region, and the rest state of the material can therefore be successfully studied.

The rheological properties of dressing emulsions are a function of the continuous-phase viscosity, the phase volume fraction of the dispersed phase, the droplet size, and the nature of the colloidal interactions. In general, the viscosity of an emulsion increases as the continuous-phase viscosity increases. Typically, gums and stabilizers that have non-Newtonian rheology are added to pourable salad dressings to impart the non-Newtonian character to emulsions and increase the viscosity of the continuous phase and hence the viscosity of emulsions. The emulsion viscosity generally increases with the phase volume fraction of the dispersed phase in a nonlinear manner, and decreases with increasing droplet size. Rheological properties of an emulsion also depend on the relative magnitude of the attractive and repulsive

interactions between the droplets. Viscosities of emulsions can be increased by adding biopolymers to increase the depletion attraction or bridging flocculation, by altering the pH or ionic strength to reduce electrostatic repulsion, and by heating protein-stabilized emulsion to increase hydrophobic attraction. Aggregation may increase the emulsion viscosity. The aggregation state of an emulsion is clearly dependent on the shear rate, and for thixotropic (time-dependent) systems, on shear history.

Factors Affecting Physical Stability

Emulsion Stability

The term 'emulsion stability' refers to the ability of an emulsion to resist changes in its properties over time. The more stable the emulsion, the more slowly its properties change. All emulsions are thermodynamically unstable and, given enough time, undergo phase separation. However, in a kinetic sense, the dressing emulsions can be made stable over shelf-life, maintaining appearance, texture, and flavor that are desirable to the consumer. The stability of dressing emulsions may range from seconds for a product like separating Italian dressing, to years for a product like mayonnaise. Semisolid dressings, such as mayonnaise and spoonable salad dressings, are very viscous emulsions. Due to the high oil content of the mayonnaise emulsion, the oil droplets are forced close together. This close proximity precludes concern for flocculation, and emulsion stability must instead be concerned with preventing coalescence, which can be delayed by employing tough, pliable membranes that are formed from the protein and polar lipid components of egg yolk around the oil droplets. Pourable dressings are much less viscous than semisolid dressings, and are susceptible to creaming. Since the oil droplets are spaced apart from each other, flocculation that could lead to coalescence is also of concern. For low-fat and fat-free dressings, emulsion stability is normally not a major problem due to the low amount of oils. Special considerations have to be given to impart a full-fat flavor and texture for these products.

The emulsion stability is influenced by a number of factors, including the egg yolk, the emulsifying effect of mustard, method of mixing, and storage conditions. The stability of mayonnaise may be evaluated by simply storing samples at room temperature until visible separation occurs. Centrifugation of emulsions hastens the formation of separate layers and is also a useful test of stability. The stability index for mayonnaise is calculated as the percentage of the sample remaining emulsified after being centrifuged at 5000 rpm for 30 min and then allowed to stand for a period of time.

Egg yolk

Egg yolk is itself an emulsion, and it functions as an efficient emulsifying agent. Three types of egg yolk are currently used in dressings and mayonnaise: liquid egg yolk, frozen salted egg yolk, and dried egg yolk.

Liquid egg yolk The emulsifying ability of egg yolk is affected by the age and genetic factors of the bird from which the eggs were obtained. Decreased emulsifying ability is associated with eggs from older birds. Also, low social dominant strains of the White Leghorn breed were found to produce eggs with greater emulsifying ability than eggs from high social dominant strains. The emulsification capacity of liquid yolk decreases as dilution with albumen increases. This is due to the lower solids content and to interactions between albumen proteins and yolk fractions.

Pasteurization does not significantly affect the emulsification capacity of commercial fresh yolk containing 48–49% solids. Heating albumen-free yolk to 61 °C causes no significant change in emulsion stability, but emulsification capacity is significantly increased by heating the yolk to 63 °C. Pasteurization temperatures above 63 °C cause a considerable increase in egg yolk viscosity. However, salted (10% sodium chloride) liquid egg yolk can be pasteurized at 62–78 °C for 5 min without damage to emulsifying properties. In the manufacture of oil-in-water salad dressings, the addition of egg yolk and salt results not only in an increase of emulsion viscosity, but also in a reduction of oil droplet size. (*See* **Pasteurization**: Principles.)

Frozen salted egg yolk Viscosity increases and gelation occurs in raw egg yolk subjected to freezing and storage below −6 °C. The addition of salt reduces gelation and inhibits microbial growth during thawing. Frozen yolk containing 10% sodium chloride is the most common form used in the dressings and mayonnaise industry. Frozen salted egg yolk produces mayonnaise that is stiffer than that made with fresh yolk. Freezing at −29 °C and storage at −23 °C for 1–4 months is detrimental to both pasteurized and unpasteurized salted egg yolk. Acidification in combination with pasteurization, followed by freezing and storage, also damages the emulsification ability of salted egg yolk.

Dried egg yolk The advantages of using dried egg yolk are threefold:

1. less storage space and transportation cost than liquid or frozen egg yolk;
2. not susceptible to bacterial growth during storage;

3. good uniformity, easy to handle in a sanitary manner.

Compared to frozen and to freeze-dried egg yolk, spray-dried egg yolk produces the least stable emulsions. This detrimental effect of spray-drying on yolk is due to the rapid increase in extractability of the free lipids.

Mustard

Mustard has a slight stabilizing effect on emulsions. This effect depends not only on the chemical and physical properties of mustard, but also on the method of incorporating mustard into the mayonnaise. In general, mustard collects at the oil–water interface and tends to prevent droplets of oil from coalescing. (*See* **Mustard and Condiment Products**.)

Method of Mixing

Introduction of oil beneath the emulsion surface provides mayonnaise with improved stability, consistency, and homogeneity. Using this method, the best mayonnaise can be produced when a small amount of oil is first added to the emulsifying agent, followed by the addition of acid, and then the addition of the remainder of the oil.

The appearance of the dispersed and continuous phases in laboratory and commercially made mayonnaises is different. This is due to the difference in the degree of agitation. A colloidal mill is used to prepare the emulsion of commercial products, whereas a catering-type food mixer is used in the laboratory. The colloidal mill results in uniformity in the size and shape of fat globules and a continuous phase with no separation from the fat globules. In mayonnaise prepared in a food mixer, the fat globules are irregular in both size and shape, and some globules seem to join together to form a continuous mass. Other mixing factors that affect consistency of mayonnaise include the amount and composition of the aqueous phase added during the first stage of mixing, the duration of beating and resting periods, and the addition of vinegar at various stages in the process.

Storage Conditions

Mayonnaise and salad dressing become increasingly unstable when stored at 55 °C for 3 days. The rate of droplet coalescence in salad dressing may not be as high as in mayonnaise due to the stabilizing effect of the starch ingredients.

Freezing usually causes the dispersed phase of emulsions to coalesce. Thus, freezing and thawing of mayonnaise result in collapse of the emulsion structure. The emulsion may separate into two layers, the upper layer of oil and the lower layer probably containing water, proteins, sugar, salt, and mustard (Figure 3). The damage to emulsions that are broken by freezing is related to the influence of freezing on the emulsifying agent. (*See* **Freeze-drying**: Structural and Flavor (Flavour) Changes.)

Salad dressing may be stable to frozen storage when the oil does not crystallize, or crystallizes slowly. Other conditions favoring stability include use of a thickening agent such as waxy rice flour and freshly salted unfrozen egg yolk, at a level of 5.3–8%, along with a decrease in oil. At storage temperatures above −18 °C, this combination was stable for 6 months. Other factors affecting emulsion stability include:

1. fermentation of the egg yolk with pancreatic phospholipase A_2;
2. acetylation and succinylation of egg yolk lipoproteins;
3. addition of 0.05% sodium 2-lactylate to egg yolk.

Storage and Shelf-Life

Most dressings and mayonnaise held at room temperature have a shelf-life of approximately 7 months. These products are preserved against microbial spoilage by their acid and sodium chloride content, but they are very sensitive to oxidative deterioration in flavor and should be refrigerated after their containers are opened. (*See* **Storage Stability**: Mechanisms of Degradation.)

The acid and sodium chloride content protects against spoilage by microorganisms such as *Salmonella* and *Staphylococcus* species. Some lactic acid bacteria and yeast may survive at the low pH (pH 4.0) of mayonnaise, but they are destroyed by pasteurization at 60–63 °C for 3–5 min. Care should be taken when mixing dressing with mayonnaise to avoid growth of microorganisms which might occur at the higher pH of the mixture. (*See* **Lactic Acid Bacteria**; **Spoilage**: Yeasts in Spoilage; ***Staphylococcus***: Properties and Occurrence.)

Oxidative rancidity is one of the major problems in relation to the use of vegetable oils. Even a hint of a rancid flavor can ruin an entire batch of salad dressing or mayonnaise. A thin film of oil, such as might be left on containers through inadequate washing, can become rancid very quickly. Thus, all dressing containers must be cleaned thoroughly, and a fresh batch should never be poured into a container containing older dressing.

Time, temperature, light, air, surface exposure, moisture, nitrogenous organic material, and traces of metals are factors responsible for the rancidity of dressings and mayonnaise. In salad dressing and mayonnaise products, the oil is subjected simultaneously to most or all of these adverse conditions:

- The emulsification process increases the exposed surface area of the oil.
- On average, mayonnaise contains 10–12% air by volume, although an inert gas such as nitrogen may replace some of the air.
- Moisture is present.
- Nitrogenous organic material is dispersed in films surrounding the oil globules.
- The products are packed in glass jars that are exposed to light.
- In some production areas traces of metals may be dissolved out of the equipment by the vinegar, although the use of stainless-steel equipment minimizes this.
- The temperature at which salad dressing and mayonnaise are kept may be as high as 38 °C or higher.
- It may be 3–6 months before these products are consumed.

For these reasons, the oil used by the salad dressing industry must be of high quality. Commercial salad dressings are presently made with both unhydrogenated and hydrogenated soya bean oil. Whether or not hydrogenation is necessary to prepare oxidatively stable salad dressing is a controversial issue in the industry. Hydrogenation of soya bean oil with copper and nickel catalysts effectively increases the storage stability of salad dressing at 21 °C but not at 32 °C. The use of butylated hydroxytoluene (BHT) as an antioxidant in the oil, and ethylenediaminetetraacetic acid (EDTA) as a metal scavenger in the starch base, as well as nitrogen packaging, are effective in prolonging the storage stability of salad dressing made with unhydrogenated soya bean oil. Therefore, these additives or nitrogen packaging may provide economic substitutes for hydrogenation of soya bean oil used in salad dressing. (*See* **Antioxidants:** Synthetic Antioxidants; **Soy (Soya) Beans:** The Crop; **Vegetable Oils:** Oil Production and Processing.)

Separation of mayonnaise may occur as a result of long storage, warm temperature, freezing, or considerable shaking or agitation during shipping. This problem is controlled commercially by fine division of the oil droplets, use of an effective stabilizer, cool storage temperatures (0–5 °C), and protection from air and light during storage.

See also: **Antioxidants**: Synthetic Antioxidants; **Colloids and Emulsions**; **Eggs**: Structure and Composition; **Emulsifiers**: Organic Emulsifiers; Uses in Processed Foods; **Freeze-drying**: Structural and Flavor (Flavour) Changes; **Lactic Acid Bacteria**; **Mustard and Condiment Products**; **Pasteurization**: Principles; **Stabilizers**: Types and Function; Applications; ***Staphylococcus***: Properties and Occurrence; **Starch**: Modified Starches; **Storage Stability**: Mechanisms of Degradation; **Vegetable Oils**: Oil Production and Processing

Further Reading

Dickinson E (1992) *Introduction to Food Colloids*. Oxford: Oxford University Press.

Hanson HL and Lorraine RF (1961) Salad dressings stable to frozen storage. *Food Technology* 15: 256–262.

Harrison LJ and Cunningham FE (1985) Factors influencing the quality of mayonnaise: a review. *Journal of Food Quality* 8: 1–20.

McClements DJ (1999) Emulsion rheology. In: *Food Emulsions. Principles, Practice, and Techniques*. Boca Raton, Florida: CRC Press.

Warner K, Frankel EN, Snyder JM and Porter WL (1986) Storage stability of soybean oil-based salad dressings: effects of antioxidants and hydrogenation. *Journal of Food Science* 51: 703–708.

Yang SC (1988) Effect of types of salt on characteristics of 10% salted egg yolk in mayonnaise. *Food Science (Taiwan)* 15: 304–313.

Dried Foods *See* **Drying**: Theory of Air-drying; Drying Using Natural Radiation; Fluidized-bed Drying; Spray Drying; Dielectric and Osmotic Drying; Physical and Structural Changes; Chemical Changes; Hygiene; Equipment Used in Drying Foods

DRUG–NUTRIENT INTERACTIONS

B Caballero and K G Conner, Johns Hopkins University, Baltimore, MD, USA

Background

Understanding the interactions between dietary constituents and pharmacological compounds is essential to monitor drug therapy correctly as well as to assess the potential nutritional impact of medications. Most therapeutic agents exhibit some form of interaction that ultimately affects the nutritional status of the host, by altering absorption or utilization of nutrients. Frequently, these changes are not readily identified or may be obscured by the underlying disease.

The interactions between therapeutic agents and nutrients are part of the large number of interactions occurring between nutritional and nonnutritional constituents of the human diet. These include all substances added to the food chain – incidentally or deliberately – during harvesting, processing, packaging, distribution, and preparation of foods. Some examples of this broad category would include pesticides, food additives, antibiotics, hormones, and environmental toxins.

Drug–nutrient interactions operate in both ways: drugs can have a significant impact on nutrient absorption and utilization. Also, the nutritional status of the host certainly affects the ability of drugs to be absorbed and transported, and to exert an effect at the target tissues.

Drug–nutrient interactions can be broadly classified in two categories: direct physicochemical interaction and physiological or functional interaction. Drug–nutrient interactions can also be classified according to the site of their occurrence: within the food matrix, in the gastrointestinal tract, or during transport, metabolism, and excretion. The mechanisms and sites of drug–nutrient interactions are listed in Table 1.

Physicochemical Interactions

These usually involve some form of molecular interaction between the drug and a nutrient, and occur primarily during digestion and absorption. The usual consequence of this interaction is a reduction in the bioavailability of the drug and/or the nutrient. A well-known example of this mechanism is the binding of metals by the antibiotic tetracycline.

Functional Interactions

Functional Interactions at the Gastrointestinal Tract

These acquire particular significance because alterations in gastrointestinal (GI) function are likely to affect the digestion and absorption both of the drug causing it and of a number of nutrients. The most common GI functional effects altering those functions are as follows.

Changes in GI motility A reduction in transit time may lead to a decreased absorption. A large number of drugs have an effect on gut motility, whether this is their primary therapeutic effect or not. Conversely, food composition affects motility as well. Dietary fiber not only increases motility but also may trap other nutrients and drugs and reduce their availability for absorption.

Table 1 Mechanisms and sites of drug–nutrient interactions

Site	*Mechanism*	*Effect*
Food matrix	Binding and chelation	Decreases bioavailability
Gastrointestinal (GI) tract	Changes in GI motility	Increase in transit time reduces absorption
	Binding and chelation	Decreases bioavailability
	Change in bile acid concentration	Reduces absorption of fat-soluble nutrients
	Gastric pH	Affects absorption of iron, vitamin B_{12}, and others
Circulation	Albumin concentration	Decreases transport of bound substances
	Competitors for albumin binding	Displaces albumin-bound nutrients (fatty acids, tryptophan, etc.)
Target tissues	Antagonistic effects	May increase requirements for antagonized nutrient
	Enzyme activities	Reduces concentration of enzyme product
Excretion	Renal function	Increased excretion may lower nutrient levels, increase requirements
	Sequestration	Same as above

Changes in gastric acid output A reduced chloride production with a subsequent increase in gastric pH retards gastric emptying and may alter the balance between ionized and nonionized forms of therapeutic agents.

Reduction in the concentration of bile acids This affects the absorption of most fat-soluble compounds. Lower bile acid concentration may result from increased binding and excretion or decreased production. For example, the antibiotic neomycin binds to bile acids and increases their fecal excretion, thus reducing their luminal concentration, and in so doing decreasing the absorption of fat-soluble vitamins. This interaction, like many others, can be used therapeutically to reduce bile acid turnover in certain liver diseases, and to lower cholesterol levels by reducing their reabsorption.

Alterations in the GI microflora This may affect the availability of nutrients produced by the normal gut flora, such as vitamin B_{12}. Since many drugs are susceptible to bacterial metabolism, changes in the gut flora may also affect drug bioavailability. In certain cases, drug cleavage by intestinal microorganisms is an expected and necessary step for adequate drug action. For example, the antiinflammatory agent 5-aminosalicylic acid is given as its precursor sulfasalazine, which is converted to the active compound by colonic bacteria. An altered colonic flora affects the production of the active compound. Drugs can also inhibit nutrient absorption by directly affecting protein synthesis in the enterocyte. Since most transport systems require active protein synthesis and turnover, such inhibition would result in a decreased rate of nutrient absorption. Furthermore, certain drugs undergo initial metabolism in the enterocyte, before reaching the bloodstream. Alterations in protein synthesis at the enterocyte, or an impaired turnover of the intestinal epithelia, also affect this process.

Interactions Affecting Transport, Metabolism and Excretion

Functional Synergism or Antagonism

The biological actions of nutrients and drugs can be synergistic or antagonistic, occur in different time periods after exposure, and affect a variety of target tissues. Some of the most common mechanisms are as follows.

Alterations in drug transport Drugs circulate in the bloodstream as free compounds or bound to other blood constituents, usually proteins. Drugs vary greatly in their propensity to bind to circulating proteins, virtually covering the entire spectrum from zero to 100%. For a given drug, the bound fraction tends to be relatively constant under physiological conditions, but responds to changes in pH, electrolyte balance, and the presence of competing molecules. The major transport protein in plasma is albumin, and its concentration and the presence of other compounds with affinity for albumin binding affect the amount of drug that ultimately will be transported by this protein.

Increase in nutrient catabolism Certain drugs stimulate detoxifying systems such as the cytochrome P-450 pathway. Activation of this system may result in an increased catabolism of certain nutrients. In other cases, drugs have a direct action on nutrient catabolism, as in the case of anticonvulsant drugs, which stimulate vitamin D catabolism in the liver.

Biological antagonism This occurs when a drug and nutrient have opposite biological actions, as is the case, for example, of vitamin K and salicylates in the coagulation process.

Increased nutrient losses Many drugs enhance directly or indirectly the urinary excretion of nutrients. Examples include the increased urinary losses of electrolytes caused by aminoglycoside antibiotics, and the increase in urinary ascorbic acid excretion induced by barbiturates.

Host-related Functional Interactions

Nutrients and nutritional status can conversely also affect drug action and disposition. Perhaps the most significant host-related factor affecting drug disposition is protein synthesis. An altered protein synthesis, usually resulting from deficient dietary protein intake or severe diseases, will affect absorption, transport, metabolism, and excretion, all protein-dependent processes. The role of plasma albumin in drug transport was discussed above, and is certainly affected by impaired albumin synthesis and/or sequestration in the extravascular space, as seen in protein–energy malnutrition. It should be noted, however, that malnutrition affects many aspects of drug metabolism, not all in the same direction. For example, drug delivery may be reduced by impaired albumin concentration, but drug concentration in the bloodstream may be increased, owing to impaired clearance, which is also affected by malnutrition.

The plasma amino acid profile may affect the efficacy of drug entry into the central nervous system. At the blood–brain barrier, certain drugs are transported

Table 2 Major drug–nutrient interactions of clinical relevance

Drug	*Class*	*Food/nutrient*	*Effect/mechanism*
Acarbose	Antidiabetic	Food	Delays carbohydrate breakdown: delays glucose absorption
		Iron	Decreased iron absorption
Acetaminophen	Analgesic	Food	May delay extended release; high-pectin food delays absorption
		Alcohol	Increased risk of hepatotoxicity
Acetohexadmide	Antidiabetic	Glucose	Hypoglycemia
		Alcohol	Flushing, headache, nausea, vomiting, sweating, tachycardia
		Sodium	Hyponatremia, SIADH
Acyclovir	Antiviral	Food	No effect; may take with meals
Aluminum hydroxide	Antacid	Thiamin	Affects bioavailability, because of pH
		Iron	Decreased iron absorption
		Phosphorus	Inhibits phosphorus absorption
		Vitamin A	Inhibits vitamin A absorption
Amikacin	Antibiotic	Calcium, potassium, magnesium	Causes renal wasting of these nutrients
Amoxicillin	Antibiotic	Food	Decreased absorption by delayed gastric emptying
Amphotericin B	Antifungal	Potassium, magnesium	Causes renal wasting of potassium and magnesium
Ampicillin	Antibiotic	Food	Decreased absorption by delayed gastric emptying
		Potassium	High doses increase urinary potassium losses
Antipyrine		Green vegetables, beef protein	Decreased absorption
Aspirin	Analgesic	Food	Decreased rate of absorption
		Folic acid	Increased excretion of folate
		Amino acids	Decreased intestinal absorption of amino acids, increased urinary excretion of tryptophan
		Iron	Chronic high dose 3–4 g day^{-1}, iron deficiency possible
		Alcohol	Gastric irritation, leading to possible gastric bleeding
		Curry powder, licorice, teas, raisins, paprika	Potential salicylate accumulation
		Ascorbic acid, vitamin C, fresh fruits, high vitamin C	Increased urinary excretion; decreased concentration in serum and platelets
Astemizole	Antihistamine	Food	Decreased bioavailability
Atenolol	Antihypertensive	Food	Delayed absorption
Atovaquone	Antibiotic	Food	Bioavailability increased, especially high-fat foods
Atropine	Anticholinergic	Iron	Delayed absorption
Azithromycin	Antibiotic	Food	Decreased rate and delayed absorption
Bacampicillin	Antibiotic (penicillins)	Food	Decreased absorption
Barbiturates	Anticonvulsant	Alcohol	Enhanced CNS depression
		Calcium, vitamin D	Increased vitamin D requirements, owing to increased metabolism
		Cyanocobalamin	Increased bone resorption
		Folic acid	Decreased serum levels, leading to megaloblastic anemia
		Serum lipids	Decreased CSF folate and erythrocyte concentration; may increase cholesterol, HDL triacylglycerols
Benzodiazepines	Anticonvulsant	Nutrient	Enhanced CNS depression
Clonazedam		Calcium	Increased vitamin D requirements, secondary to increased metabolism
Clorazepate dipotassium		Vitamin D	Increased bone resorption
Lopazepam		Cyanocobalamin	Decreased serum levels, leading to megaloblastic anemia
Oxazepam		Folic acid	Decreased CSF folate and erythrocyte concentration
		Serum lipids	May increase cholesterol, HDL triacylglycerols
Buprenorphone HCL	Analgesic Narcotic Agonist–antagonist	Alcohol	Enhanced CNS depression

Continued

Table 2 Continued

Drug	*Class*	*Food/nutrient*	*Effect/mechanism*
Butorphanol tartate	Analgesic Narcotic Agonist–antagonist	Alcohol	Enhanced CNS depression
Calcium carbonate	Antacid	Iron	Decreased iron absorption
		Fats	May cause steatorrhea
Captopril	Antihypertensive angiotensin-converting enzyme inhibitor	Food	Reduced absorption
Carbamazpine	Anticonvulsant	Sodium	SIADH
		Food	Enhanced absorption, increased bile production
Carbenicillin iandanyl sodium	Antibiotic	Food	Decreased rate of absorption
Cephalosporins	Antibiotic	General alcohol	Flushing, headache, nausea, vomiting, tachycardia
Cefachlor		Food	Decreased rate of absorption
Cefadroxil		Food	No effect (may take food)
Cefamandole		Vitamin K	Decreased vitamin K hypoprothrombinemia
Cefixime		Food	Decreased rate of absorption
Cefoperazone		Vitamin K	Decreased vitamin K hypoprothrombinemia
Cefotetan		Vitamin K	Decreased vitamin K hypoprothrombinemia
Cefpodoxime proxetil		Food	Bioavailability increased with food
Ceftibuten		Food	Decreased absorption
Cefuroxim axetil		Food	Bioavailability increased with food
Cephalexin		Food	Absorption reduced for suspension; delayed for capsules
Cephradine		Food	Rate of absorption delayed
Cetrizine	Antihistamine	Food	Delays time to serum peak; no effect on overall absorption
Chlorambucil	Antineoplastic	Food	Reduced absorption
Chloramphenicol	Antibiotic	Iron	Increased serum level iron; increased total iron-binding capacity
		Folic acid	Antagonist to physiological action; increased requirements of folic acid
		Vitamin B_{12}	Increased requirements of vitamin B_{12} can cause peripheral neuropathy
Chloroquine	Antimalarial	Food	Increased bioavailability
Chlorothiazide	Diuretic	Food	Increased drug absorption owing to delayed gastric emptying
Chlorpromazine	Antiemetic	Food	Decreased absorption owing to delayed gastric emptying
Chlorpropamide	Antidiabetic	Glucose	Decreased blood glucose concentration
		Sodium	Hyponatremia, SIADH
		Alcohol	Flushing, headache, nausea, vomiting, tachycardia
Cholchicine	Antigout	Cyanocobalamin	Decreased absorption of cyanocobalamin
Cimetidine	Histamine 2 antagonist	Food	Delays absorption
Ciprofloxacin	Antibiotic (quinolone)	Caffeine	Decreased rate of absorption
		Food	Cipro decreased elimination of caffeine
		Calcium	Calcium can bind quinolones
		Mineral supplement	Absorption of divalent cations and trivalent cations decreased by binding to quinolones
Clarithromycin	Antibiotic (macrolide)	Food	Decreased onset of absorption; no change in total amount absorbed
Clonazepam	Anticonvulsants (benzodiazepine)	Nutrient	Enhanced CNS depression
Clorazepate dipotassium		Calcium	Increased vitamin D requirements, owing to increased metabolism
		Vitamin D	Increased bone resorption
		Cyanocobalamin	Decreased serum levels, leading to megaloblastic anemia
		Folic acid	Decreased CSF folate and erythrocyte concentration
		Serum lipis	May increase cholesterol, HDL triacyclglycerols
Cloxacillin	Antibiotic (penicillin)	Food	Decreased rate of absorption

Continued

Table 2 Continued

Drug	*Class*	*Food/nutrient*	*Effect/mechanism*
Codeine	Narcotic agonist, analgesic	Alcohol	Enhanced CNS effect
		Glucose	Can cause hyperglycemia
Corticosteroids	Steroid	Calcium, phosphorus, vitamin D	Decreased absorption of Ca, P; increased urinary excretion; chronic high dose can cause osteomalacia
Dexamethazone Hydrocortisone Methylprednisolone (Prednisone, Prednisolone)			
Corticosteroids	Steroid	Nitrogen	Increased urinary nitrogen losses
		Zinc	Increased urinary excretion and decreased serum levels
		Glucose	Impairs glucose tolerance; increases plasma levels
		Triacylglycerols, cholesterol	Increased serum levels
Co-trimaxazole	Antibiotic	Potassium	Decreased excretion hyperkalemia
		Sodium	Increased excretion hyponatremia
		Folic acid	Potential for folate deficiency
Cyclosporine	Antirejection	Milk, fat, pineapple juice	Increased absorption
Demeclocycline	Antibiotic	Food, calcium, iron	Decreased absorption of dairy products and divalent and trivalent cations
Diazepam	Anticonvulsant	Food	Increased absorption with high-fat meals and delayed gastric emptying
Clonazedam Clorazepate dipotassium Lopazepam Oxazepam			
Dicumarol	Anticoagulant	Food	Increased absorption with high-fat meals and delayed gastric emptying
Didanosine Tab	Antiviral	Food	Decreased rate and extent of absorption
Oral suspension		Fruit juice or acid liquid	Didanosine unstable in acid pH
Digoxin	Cardiac	Food	Delay absorption; adsorbent to high-fiber, high-pectin foods
Dirithromycin	Antibiotic (macrocide)	Food	Slightly increased absorption
Divalproex	Anticonvulsant	Food	Decreased rate of absorption; extent of absorption not affected
Doxycycline	Antibiotic	Food	Decreased absorption of food and milk
Erythromycin	Antibiotic (macrocide)	Food	Increased absorption by delayed gastric emptying
Erythromycin stearate		Food	Reduced absorption by delayed gastric emptying
Ethionamide	Antituberculosis	Pyridoxins	Reports of peripheral neuritis and paresthesia
Etodolac	Nonsteroidal antiinflammatory drug (NSAID)	Food (milk)	Decreased total bioavailability of tolmetin; decreased absorption of ibuprofen
		Sodium	Hyponatremia (indomethacin/ketorolac)
		Potassium	Hyperkalemia (indomethacin/ketorolac)
		Food	Increased rate of absorption
Felbamate	Anticonvulsant	Glucose	Hypoglycemia
		Magnesium	Hypomagnesemia
		Phosphorus	Hypophosphatemia
		Potassium	Hypokalemia
		Sodium	Hyponatremia
Fenoprofen	NSAID	Food (milk)	Decreased total bioavailability of tolmetin
Fenoprofen calcium		Sodium	Hyponatremia (indomethacin/ketorolac)
		Potassium	Hyperkalemia (indomethacin/ketorolac)
		Food (milk)	Decreased bioavailability of tolmetin; decreased absorption of ibuprofen
		Food	Increased rate of absorption
Fluconazole	Antifungal	Potassium	Hypokalemia
Flucytosine	Antifungal	Food	Decreased rate of absorption; no change in extent of absorption

Continued

Table 2 Continued

Drug	*Class*	*Food/nutrient*	*Effect/mechanism*
Foscarnet	Antiviral	Calcium	Hypocalcemia; drug chelates; divalent metal ions
		Magnesium	Hypomagnesemia
		Phosphorus	Hypo- and hyperphosphatemia
		Potassium	Hypokalemia
Furazolidone	Antiinfective	Tyramine-rich foods (avocados, canned figs, aged cheese, cola beverages, coffee, chocolate, wines, soy sauces, fermented meats, yeast preparation, yogurts)	Prolonged large doses result in increased risk for hypertensive crisis
		Alcohol	Flushing, headache, nausea, vomiting, sweating, tachycardia
Furosemide	Diuretic	Food	Delayed absorption
Ganciclovir	Antiviral	Food	Increased area under curve plasma concentration
Glipizide	Antidiabetic	Food	Delayed absorption
		Alcohol	Flushing, headache, nausea, vomiting, sweating, tachycardia
		Sodium	Hyponatremia, SIADH
Griseofluvin	Antifungal	Alcohol	Can increase alcohol effect, flushing, tachycardia
		High-fat food	Increased drug absorption rate
Hydralazine	Diuretic	Food	Increased absorption
Hydrochlorothiazide	Diuretic	Food	Increased absorption by delayed gastric emptying
Ibuprofen	NSAID	Food (milk)	Decreased total bioavailability of tolmetin; decreased absorption of ibuprofen
		Sodium	Hyponatremia (indomethacin/ketorolac)
		Potassium	Hyperkalemia (indomethacin/ketorolac)
		Food	Increased rate of absorption
Indinavir	Antiviral	Food	Decreased absorption of high-calorie, high-fat, and protein
		Grapefruit juice	Decreased area under curve concentration
Indomethicin	NSAID	Food (milk)	Decreased total bioavailability of tolmetin; decreased absorption of ibuprofen
		Sodium	Hyponatremia (indomethacin/ketorolac)
		Potassium	Hyperkalemia (indomethacin/ketorolac)
		Food	Increased rate of absorption
Iron	Mineral	Ascorbic acid	Increased absorption
		Amino acids	Increased absorption
		Calcium phosphate	Decreased absorption
		Zinc	Inhibits absorption
		Vitamin A	Vitamin A deficiency inhibits iron utilization and accelerates development of anemia
		Tea/coffee	Decreased absorption owing to formation of iron tannate
		Vegetable polyphenols	Binds and insolubilizes iron
Isoniazid	Antituberculosis	Food	Decreased intestinal absorption
		Pyridoxine	Decreased metabolism, antagonism
		Food and histamine, tuna, liver, aubergine, parmesan cheese, tomato, spinach, tyramine-containing foods	Headache, redness, itching of eyes and face, chills, diarrhea, palpitation
			Potential hypertensive crisis owing to monoamine oxidase inhibitor activity

Continued

Table 2 Continued

Drug	*Class*	*Food/nutrient*	*Effect/mechanism*
Itraconazole	Antifungal	Food	Increased absorption, increased triacylglycerols
		Potassium	Hypokalemia
Ketoconazole	Antifungal	Alcohol	Flushing, headache, nausea, vomiting, sweating, tachycardia
Labetalol	Antihypertensive	Food	Increased absorption
Lamivudone (3TC)	Antiviral	Food	Decreased rate of absorption
Levodopa	Anti-Parkinson	Food	Decreased absorption; with high-protein meals amino acids compete for absorption
Lithium	Antimanic	Low-sodium diet	Increased lithium concentrations
		High-sodium diet	Increased lithium clearance
		Food	Increased absorption
Lomefloxacin	Antibiotic (quinolone)	Food	Decreased rate and extent of absorption
Loracarbef	Antibiotic	Food	Increased rate of absorption
Lovastatin	Antihyperlipidemia	Food	Increased absorption
Mebendazole	Anthelmintic	Food	Increased absorption
Meclofenamate	NSAID	Alcohol	Additive CNS effects; increased prothrombin time
		Food	Decreased bioavailability
Melphalan	Antineoplastic	Food	Reduced absorption
Mercaptopurine	Antineoplastic	Food	Reduced absorption
Methacycline	Antibiotic	Food, calcium, iron	Decreased absorption of dairy products, cereals, divalent and trivalent cations
Methenamine mandelate	Urinary antiinfective	Milk products, citrus fruits	Excessive amounts inhibits drug conversion
Methosuximide	Anticonvulsant	Alcohol, calcium	Additive CNS effects; hypocalcemia
Methotreate	Antineoplastic	Food	Increased absorption
Methyldopa	Antihypertensive	Vitamin B_{12}, folate	In high doses methyldopa can increase vitamin B_{12} and folate losses
		Food	High-protein meals compete for absorption
Metoprolol	Antihypertensive	Food	Increased absorption
Metronidazole	Antibiotic	Alcohol	Flushing, headache, nausea, vomiting, sweating, tachycardia
		Food	Decreased peak serum concentration but total amount of drug absorbed not affected
Minocycline	Antibiotic	Food, calcium	Decreased absorption
Nafcillin	Antibiotic	Food	Decreased absorption; decreased serum levels owing to altered gastric pH
		Potassium	High doses can cause hypokalemia owing to increased urinary losses
Nifedepine	Antihypertensive calcium channel blocker	Grapefruit juice	Increased serum level of nifedepine flavonoids inhibits cytochrome P-450
		Food	Decreased bioavailability, formulation dependent
Nitrofurantoin	Antibiotic	Food	Increased absorption by delayed gastric emptying
NSAIDs	NSAID	Food (milk)	Decreased bioavailability of tolmetin
Diclofenal Etodolac		Sodium	Hyponatremia (indomethacin/ketorolac)
Fenoprofen Ca Ibuprofen		Potassium	Hyperkalemia (indomethacin/ketorolac)
Ketoprofen Ketorolac Naproxen Oxaproxin Piroxican Sulindac		Food	Decreased absorption of ibuprofen
Tolmetin NA		Food	Increased rate of absorption
Norfloxacin	Antibiotic (quinolone)	Food, dairy products	Decreased rate of absorption
		Multivitamin and mineral supplements	Decreased absorption owing to formation of divalent/trivalent cation complexes with quinolones
Oflaxacine	Antibiotic (quinolone)	Dairy products and mineral supplements	Decreased absorption by polyvalent cations

Continued

Table 2 Continued

Drug	*Class*	*Food/nutrient*	*Effect/mechanism*
Ondansetron	Antiemetic	Food	Increased extent of absorption
		Potassium	Hypokalemia
Oral contraceptives		Ascorbic acid	Decreased ascorbic acid concentration in plasma, platelets, leucocytes
		Vitamin C, folic acid	Decrease in their serum levels
		Vitamin B_{12}	Impairs tryptophan metabolism
		Amino acids, vitamin A, vitamin E, copper	Increase in their serum levels
Oxacillin	Antibiotic	Food	Decreased absorption and decreased serum concentration
		Fats	Oxacillin can cause steatorrhea
Paromomycin	Amoebicide	Food	Increased absorption by delayed gastric emptying
		Vitamins A, D, E, K	Malabsorption of fat-soluble vitamins owing to hypocholesterolemia
Penicillamine	Antidote (chelating agent)	Food	Decreased absorption
		Iron/zinc	Decreased absorption 30–70% of increased zinc absorption; decreased penicillamine absorption
Penicillin G & VK	Antibiotic	Food	Decreased absorption by delayed gastric emptying
		Glucose	Hyperglycemia
Pentamidine	Antibiotic	Calcium, magnesium	Hypomagmanesemia, hypocalemia
		Potassium	Hyperkalemia due to nephrotoxicity
Phenacemide	Anticonvulsant	Fresh fruits and vitamin C	Increased urinary excretion of phenacemide
Phenobarbital	Anticonvulsant (see *Barbiturates*)	Food	Decreased absorption owing to protein binding
		Protein	Low-protein diet increases duration of action of phenobarbitol
		Vitamin D, calcium	Decreased serum vitamin D by cytochrome P-450 hypocalcemia
		Fresh fruits and vitamin C	Increased urinary excretion of phenobarbitol
Phensuximide	Anticonvulsant (succinimides)	Calcium, vitamin D	Decreased serum vitamin D by P-450 cytochrome hypocalcemia
		Vitamin B_{12}, folic acid	Decreased absorption and serum levels of folates; inhibits vitamin B_{12} transport
		Copper	Increased serum levels
Phenytoin	Anticonvulsant (hydantoins)	Fresh fruits and vitamin C	Increased urinary excretion
		Vitamin D, calcium	Decreased serum vitamin D by cytochrome P-450 hypocalcemia
		Enteral feeds	Decreased absorption
		Food	Increased absorption by delayed gastric emptying
Piroxicam		Food	Delayed absorption
Praziquantel	Anthelmintic	Food	Decreased rate and extent of absorption
Primidone	Anticonvulsant	Fresh fruits and vitamin C	Increased urinary excretion of primidone
		Protein	Low-protein diet increases duration of action of primidone
Propranolol	Antihypertensive	High-protein foods	Increased absorption
Propantheline		Food	Decreased absorption
Proxyphene	Analgesic	Food	Increased absorption by delayed gastric emptying
Pyrimethamine	Antimalarial	Folic acid	Decreased serum folate concentrations
Quinidine	Antiarrhythmic	Food	Delayed absorption owing to protein binding
Riboflavin	Vitamin	Food	Increased absorption by delayed gastric emptying
Rifampin	Antibiotic	Food	Decreased absorption
Ritonavir	Antiviral	Vitamins	Can cause vitamin deficiency
		Potassium	Hyper- and hypokalemia
		Cholesterol	Hypercholesterolemia
		Triacylglycerols	Hypertriacylglycerolemia
		Food	Delayed absorption
Oral solution capsules		Food	Increased extent of absorption

Continued

Table 2 Continued

Drug	*Class*	*Food/nutrient*	*Effect/mechanism*
Salicylates	Analgesic	Iron	Long-term chronic use decreases serum iron
Magnesium salicylate			
Choline salicylate		Vitamin C	Decreases concentration in serum and platelets
Sodium salicylate		Amino acids	Decreases their intestinal absorption and increases urinary secretion
Saquinavir mesylate	Antiviral	Food	Increased absorption of high-calorie, high-fat foods
		Calcium	Hypercalemia
		Glucose	Hyper- and hypoglycemia
		Phosphorus	Changes in serum phosphorus
		Potassium	Hyper- and hypokalemia
Spironolactone	Diuretic	Food	Increased absorption by delayed gastric emptying
Sulfonamides	Antibiotic	Food	Delayed with no effect on extent of absorption
Sulfdiazine			
Sulfisoxazole		Folic acid	Decreased intestinal synthesis, absorption and serum levels
Sulfamethoxazole			
Tetracycline	Antibiotic	Food	Decreased absorption
		Minerals	Inhibits absorption of iron, calcium, zinc and magnesium; chelation by polyvalent cations
		Fats	Decreases their absorption
		Vitamin K	Decreases bioavailability
		Vitamin C	Increase urinary losses; decreases plasma and leucocytes and levels
Theophylline	Broncodilator	Charbroiled beef	Increased metabolism of theophylline
		High-fat meals	Increased absorption dependent on formulation
Tolazamide	Antidiabetic	Sodium	Hyponatremia and SIADH
Tolbutamide			
Trimethoprim	Antibiotic	Folic acid	Decreased serum folate levels
Valproic acid	Anticonvulsant	Milk, food, carbonated drinks	Delayed absorption but no effect on extent of absorption
Divalproex sodium			
Valproate sodium			
Oral solution			
Warfarin	Anticoagulant	Alcohol	Inhibits warfarin metabolism
		Vitamin K	Beef liver, pork liver, green tea, leafy green vegetables high in vitamin K inhibit anticoagulant effect
		Vitamin E	Can increase warfarin response
Zalcitabine	Antiviral	Food	Decreased rate and extent of absorption
Zidovudine	Antiviral	Food	Decreased rate of absorption

CSF, cerebrospinal fluid; NSAID, nonsteroidal antiinflammatory drug; SIADH, Syndrome of Inappropriate Antidiuretic Hormone excretion.

into the brain by the same transport system as the large neutral amino acids, and thus must compete with them for use of the carrier binding sites. Diet composition, by affecting the postprandial amino acid profile, may change significantly the clinical efficacy of drugs such as L-dopa, used in the treatment of Parkinson's disease.

Body composition is also a relevant determinant of drug disposition and action. Although most drug dosages are calculated by total body weight, most drugs only act in the fat-free body mass. Thus, at a given body weight, individuals with more body fat will tend to receive a higher effective dose than those with lower body fat. The size of the body fat mass is also important for drugs that are stored in adipose tissue.

Major Drug–Nutrient Interactions of Clinical Relevance

Table 2 provides information on the major drug–nutrient interactions of clinical relevance. The list reflects well-known interactions of drugs that have been in the market for some time. The US Food and Drug Administration (FDA) maintains an on-line database of recently reported interactions and those of new drugs. The database can be assessed at www.fda.gov.

See also: **Amino Acids**: Metabolism; **Malnutrition**: The Problem of Malnutrition; Malnutrition in Developed Countries

Further Reading

Caballero B (1988) Nutritional implications of dietary interactions: a review. *Food Nutrition Bulletin* 10: 9–20.

Knapp HR (1996) Nutrient–drug interactions. In: Ziegler FF and Filer LJ (eds) *Present Knowledge in Nutrition*, pp. 540–546. Washington, DC: ILSI Press.

Neuvonen P (1989) Clinical significance of food–drug interactions. *Medical Journal of Australia* 150: 36–40.

Roberts J (1988) Age and diet effects on drug action. *Pharmacology and Therapeutics* 37: 111–149.

Roe DA (1986) Drug food and drug nutrient interactions. *Journal of Environmental Pathology, Toxicology and Oncology* 5: 115–135.

Roe DA (1994) Diet, nutrition and drug reactions. In: Shils ME, Olson JA and Shike M (eds) *Modern Nutrition in Health and Disease*, 8th edn, pp. 1399–1416. Philadelphia, PA: Lea & Febiger.

Wurtman RJ, Caballero B and Salzman E (1988) Facilitation of DOPA-induced dyskinesias by dietary carbohydrates. *New England Journal of Medicine* 390: 1288–1289.

Drugs *See* **Antibiotics and Drugs**: Uses in Food Production; Residue Determination

DRYING

Contents

Theory of Air-drying

J G Brennan, The University of Reading, Reading, UK

Background

Drying, or dehydration, is the oldest method of food preservation practiced by man, and today it is a major contributor to the convenience food market. The main objective of drying food is to prolong its shelf-life beyond that of the fresh material, while at the same time minimizing undesirable changes in appearance, texture, flavor, and nutritional value during the drying process. In addition to the preservation of the product, drying can bring about savings in storage and transport of foods owing to the reduction in weight and bulk that occurs during most methods of drying.

Water in Foods

Water is an important contributor to the sensory quality of foods and a loss or gain of moisture from a product can adversely affect its acceptability. However, a certain amount of moisture in foods enables undesirable changes to occur as a result of the activity of microorganisms and enzymes or through nonenzymic chemical reactions. (*See* **Drying:** Chemical Changes.) Thus, removing moisture from a food, or making it less available, is one way of extending its shelf-life.

The *moisture content* of a wet material may be expressed on a *wet-weight basis* (wwb), i.e., the mass of water per unit mass of the wet material, or on a

dry-weight basis (dwb), i.e., the mass of water per unit mass of dry solids. The latter way of expressing moisture content is usually used in drying formulae and calculations; see section Theory of Air Drying.

However, when the influence of moisture on the stability of foods is being studied, it is not only the total moisture content that is of importance but also how much of that moisture is available to support the activity of microorganisms and enzymes and enable chemical changes to occur. It is now widely accepted that a fraction of the total moisture in food is strongly bound to individual sites on the solid components and is not available to support these activities. An additional amount of moisture is less firmly bound but is still not readily available as a solvent for soluble components of the food. Such 'bound' moisture exerts a vapor pressure less than the vapor pressure of pure water at the prevailing temperature. A property known as water activity (a_w) is used to represent the availability of moisture in foods thus:

$$a_w = \frac{p_v}{p_w} \qquad (1)$$

where p_v is the water vapor pressure exerted by a solution or wet solid, and p_w is the vapor pressure of pure water at the same temperature. This expression also represents the relative humidity of an air–water vapor mixture ($\times 100$ to give the percentage relative humidity). A graph of moisture content as a function of water activity, with which it is in equilibrium, is known as a *sorption isotherm*. Such plots may be constructed by adsorption or desorption, which, in the case of many foods, results in two different curves, exhibiting hysterisis (Figure 1). Food isotherms may be considered to be made up of three regions. In region A, water is strongly bound to specific sites on the solid components, is not available as a solvent, and so does not support, microbiological, enzymic or chemical activity. The amount of moisture held in this way is known as the monolayer or monomolecular moisture content. This type of moisture usually exits below a_w 0.35, corresponding to a moisture content range 0.05–0.11 (dwb). Above region A, water is less strongly bound to the solids but still does not exert its full vapor pressure. In some publications, region B (Figure 1), is said to contain moisture that is bound in multilayers to the solid, and in region C, structural and solution effects lower the water vapor pressure. However, it is more likely that all these effects have an influence over the whole isotherm. At moisture contents of 1.0 (dwb) and above, most foods exert vapor pressures equal to that of pure water. Temperature affects the sorption behavior of foods according to the expression:

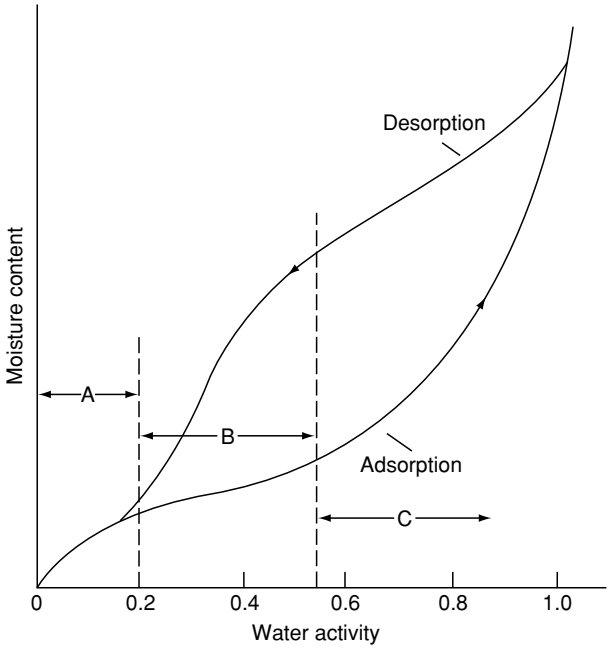

Figure 1 Adsorption and desorption isotherms showing hysteresis. From Brennan JG, Butters JR, Cowell ND and Lilly AEV (1990) *Food Engineering Operations*, 3rd edn. London: Elsevier Applied Science, with permission.

$$\frac{d(\ln a_w)}{d\left(\frac{1}{T}\right)} = -\frac{Q_s}{R}, \qquad (2)$$

where T is the absolute temperature, Q_s is the heat of adsorption, and R is the gas constant. The amount of adsorbed water at any given value of a_w decreases as the temperature increases.

Numerous mathematical models have been proposed to represent sorption isotherms of foods. One of the earliest was the Brunauer–Emmet–Teller isotherm:

$$\frac{a_w}{(1-a_w)W} = \frac{1}{W_m C} + \frac{a_w(C-1)}{W_m C}, \qquad (3)$$

where W is the total moisture content (dwb), W_m is the monolayer moisture content (dwb), and C is constant. This model is said to apply over the a_w range 0.05–0.45. An extended version of eqn (3), which takes into account multilayer adsorption and is said to be applicable up to a_w 0.90, is the Guggenheim–Anderson–DeBoer model:

$$\frac{W}{W_m} = \frac{Cka_w}{(1-ka_w)(1-ka_w+Cka_w)}, \qquad (4)$$

where C is known as the 'Guggenheim' constant, and k is a factor correcting properties of the multilayer molecules with respect to the bulk liquid.

A knowledge of the sorption characteristics of foods is useful in the prediction of drying times and energy requirements for drying processes. When

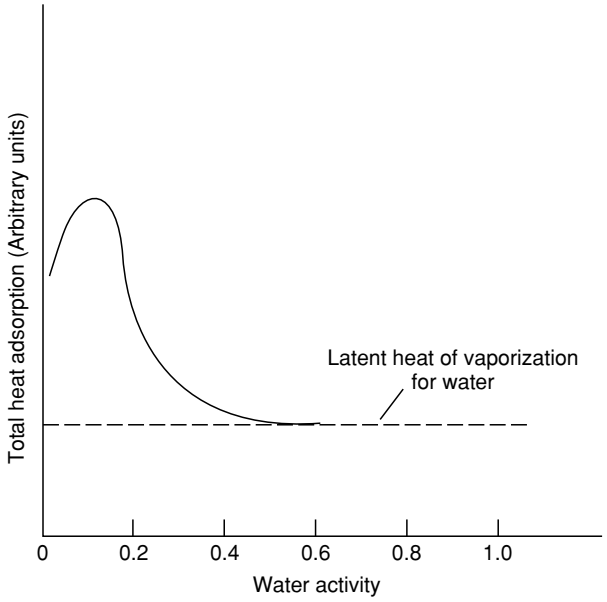

Figure 2 Dependence of the total heat of adsorption (latent heat plus heat of adsorption) on water activity. Adapted from Karel M (1974) Fundamentals of drying processes. In: Spicer A (ed.) *Advances in Pre-concentration and Dehydration of Foods*, pp. 45–94. London: Applied Science.

drying to very low moisture contents, extra energy is required to overcome the bonding of the water to the solids (Figure 2). The influence of a_w on the stability and shelf-life of foods is covered in **Drying:** Chemical Changes. The most common method of drying foods is air drying.

Theory of Air Drying

When a wet material is placed in a current of heated air, the air performs two main functions: it supplies the sensible and latent heat to bring about the evaporation of the moisture, and it acts as a carrier to move the water vapor formed away from the drying surface, thus enabling further evaporation to take place. Consider a wet solid composed of pure water and inert solids, in the form of a thin slab, which is placed in a current of heated air. If it is assumed that drying takes place from one large face only, that the temperature, humidity, and velocity of the air remain constant, and that all the heat is transferred from the air by convection, the drying curve should have the familiar shape shown in Figure 3. It comprises a number of stages or periods.

Period A–B is an *equilibration period* during which equilibrium is attained between the wet solid and the drying air. This period usually takes up a relatively small part of the drying curve.

Period B–C is known as the *constant rate period*, as the rate of drying remains constant during the period.

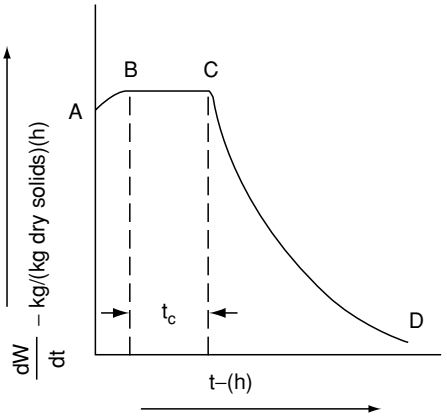

Figure 3 Model drying curve for a slab-shaped wet solid in air at constant temperature, humidity, and velocity. Adapted from Brennan JG, Butters JR, Cowell ND and Lilly AEV (1990) *Food Engineering Operations*, 3rd edn. London: Elsevier Applied Science, with permission.

Throughout this period, the surface of the solid is fully saturated with water. The movement of water from within the solid is sufficient to maintain saturated conditions at the surface. A state of equilibrium exists at the surface as the heat transfer to the surface from the air balances the rate of evaporation from the surface. The temperature at the surface of the solid remains constant at a value that corresponds to the wet-bulb temperature of the air. The driving force that causes water to evaporate into the air is the difference in water vapor pressure at the surface from that in the main air stream. The rate of drying, i.e., rate of change of weight $(-dw/dt)$, may be represented by the expression:

$$\frac{dw}{dt} = -K_g A(p_s - p_a), \qquad (5)$$

where K_g is the mass transfer coefficient, A is the drying area, p_s is the water vapor pressure at the surface of the solid, and p_a is the water vapor pressure in the main stream of the air. Eqn (5) may also be written as:

$$\frac{dw}{dt} = -K'_g A(H_s - H_a), \qquad (6)$$

where K'_g is the mass transfer coefficient, H_s is the absolute humidity at the surface of the solid, and H_a is the absolute humidity in the main air stream.

The rate of heat transfer from the air to the drying surface (dQ/dt) may be represented by the expression:

$$\frac{dQ}{dt} = h_c A(\vartheta_a - \vartheta_s), \qquad (7)$$

where h_c is the convection heat transfer coefficient, ϑ_a is the dry-bulb temperature of the air, and ϑ_s is the wet-bulb temperature of the air (temperature at

the drying surface). If sensible heat changes are neglected, the equilibrium between heat and mass transfer at the drying surface may be represented by:

$$\frac{dw}{dt}L_s = -\frac{dQ}{dt}, \tag{8}$$

where L_s is the latent heat at ϑ_s. Therefore, the rate of drying may be represented by the expression:

$$\frac{dw}{dt} = -\frac{h_c A}{L_s}\ (\vartheta_a - \vartheta_s). \tag{9}$$

The drying rate may be expressed in terms of the rate of change of moisture content ($-dW/dt$):

$$\frac{dW}{dt} = -\frac{h_c A'}{L_s}\ (\vartheta_a - \vartheta_s), \tag{10}$$

where W is the moisture content (dwb), and A' is the drying surface associated with the unit mass of dry solids. If the thickness of the slab is l, and the bulk density of the solids is ρ_s, the rate of drying may be expressed as:

$$\frac{dW}{dt} = -\frac{h_c}{\rho_s L_s l}(\vartheta_a - \vartheta_s), \tag{11}$$

If W_0 is the moisture content (dwb) of the wet material at the beginning of the constant rate period, and W_c is the moisture content (dwb) at the end of that period, the duration of the constant rate period (t_c) may be expressed as:

$$t_c = \frac{(W_0 - W_c)\rho_s L_s l}{h_c(\vartheta_a - \vartheta_s)}. \tag{12}$$

From the equations above, it can be seen that the higher the temperature and the lower the humidity of the air, the higher the rate of drying during the constant rate period. The velocity of the air also influences the drying rate during this period by affecting the heat transfer coefficient thus:

$$h_c = \frac{aG^n}{D_c^m}, \tag{13}$$

where G is the mass velocity of the air, D_c is the characteristic dimension of the system, and a, n, and m are constants. D_c has been represented by the equivalent diameter of the flow channel (cross-sectional area $\times$ 4%perimeter) or the length of the drying surface parallel to the direction of flow of the air. For most tray-drying calculations, the equivalent diameter is used. Values of n reported in the literature range from 0.35 to 0.80. Where no specific values are available, 0.80 is usually used in calculations. For situations where the air does not flow parallel to the drying surface, alternative expressions for h_c are available. When some of the heat is transferred to the drying surface by conduction or radiation, an overall heat transfer coefficient taking these sources of heat into account must be used.

Many solid foods do not exhibit a constant rate period. Others do, but it is usually quite short. Some recent work suggests that, if changes in dimensions of the solid pieces during drying are taken into account, more foods would exhibit a constant rate period.

As drying proceeds, a point is reached, C in Figure 3,when the movement of moisture to the surface is no longer fast enough to maintain the surface in a saturated condition. Beyond this point, the drying rate starts to fall and continues to do so throughout the rest of the drying cycle. Point C is known as the critical point and the moisture content at that point, W_c, the critical moisture content. The characteristics of the material being dried and the drying conditions influence the value of W_c. Period C–D (Figure 1), is known as the *falling rate period*. Many publications have reported the existence of two or more falling rate periods (see below).

During the falling rate period, the external conditions, in particular the velocity of the air, have less influence on the drying rate. This is determined by the factors that affect the movement of water from the interior of the solid to the surface. Numerous mathematical models have been proposed to represent the falling rate period of drying. Many are based on the mechanism whereby the moisture is assumed to move to the surface. Several mechanisms have been proposed, but the mechanism that has been most widely accepted is that in which the moisture moves by diffusion as a result of the concentration gradient between the surface and the interior of the wet solid. This type of movement may be represented by Fick's second law:

$$\frac{dW}{dt} = D\frac{d^2W}{dl^2}, \tag{14}$$

where W is the moisture content (dwb), t is the time, l is the distance, and D is the liquid diffusivity. One well-known solution to this equation, in an abbreviated form, for a slab-shaped solid, drying from one large face only is:

$$\frac{W - W_e}{W_c - W_e} = \frac{8}{\pi^2}\left[\exp\left\{-Dt\left(\frac{\pi}{2l}\right)^2\right\}\right] \tag{15}$$

or

$$t = -\frac{4l^2}{\pi^2 D}\left\{\ln\left(\frac{W - W_e}{W_c - W_e}\right) - \ln\left(\frac{8}{\pi^2}\right)\right\}, \tag{16}$$

where W is the average moisture content (dwb) at time t, W_e is the equilibrium moisture content (dwb), W_c is the moisture content at the start of the falling rate period (dwb), and l is the depth of the slab.

This expression holds for values of $(W-W_e)/W_c-W_e)$ less than 0.60.

Eqn (16) has been widely used to calculate drying times during the falling rate period. However, when it is used, it is assumed that D remains constant throughout the falling rate period. It is widely accepted now that the value of D changes with changes in moisture, i.e., it decreases as drying proceeds. Many authors have found that the falling rate period could be divided into two or more stages, with each stage having its own value of D.

D is dependent on temperature. The relationship most widely used to describe this dependence is of the form:

$$D_e = D_0 \exp\left(-\frac{Q}{RT}\right), \qquad (17)$$

where D_e is the average effective diffusivity, D_0 is the temperature-independent constant, Q is the energy of activation for diffusion, R is the gas constant, and T is the absolute temperature.

Many empirical models, derived from experimental data, have been used to represent falling-rate drying. One of the earliest and most widely used is:

$$\frac{dW}{dt} = -K_c\,(W - W_e), \qquad (18)$$

where dW/dt is the rate of drying (rate of change of moisture content) at a moisture content W (dwb), and K_c is the drying constant (or mass transfer coefficient). K_c is related to temperature thus:

$$K_c = a \exp\left(\frac{-b}{T}\right), \qquad (19)$$

where a and b are empirical constants.

Integrating eqn (18) gives:

$$\frac{W - W_e}{W_c - W_e} = \exp(-K_c t) \qquad (20)$$

or

$$t = -\frac{1}{K_c}\ln\left(\frac{W - W_e}{W_c - W_e}\right). \qquad (21)$$

Note the similarity between eqns (12) and (17). Although widely used, eqn (15) often does not apply to the whole falling-rate curve. The introduction of an empirical exponent, n, can widen its applicability thus:

$$\frac{W - W_e}{W_c - W_e} = \exp(-K_c t^n). \qquad (22)$$

Many other empirical models have been applied to describe falling-rate drying. Most are only applicable under conditions similar to those under which the model was developed. Many are specific to a particular food or related materials.

Neither the diffusion model nor the empirical model, outlined above, takes account of other changes that may occur during drying. Shrinkage may lead to changes in dimensions, density, and porosity of the material undergoing drying. These changes may influence the pattern of drying. Changes in the thermal properties of the material may influence heat transfer and, in turn, the pattern of drying. Many more complex models have been proposed that attempt to take account of these and other changes. Treatment of such models is outside the scope of this chapter.

Spray drying is widely used to dry *liquids and slurries*. The feed is converted into a fine mist or spray, which is brought into contact with heated air in a large drying chamber. Very rapid drying takes place, and a powder is formed, usually within 20 s. Because of the very large surface area created, evaporative cooling maintains the surface of the droplets close to the wet-bulb temperature of the air throughout most of the drying. If it is assumed that drying takes place under constant rate conditions, the time, t, for a spherical droplet to dry from an initial moisture content, W_0, to a final moisture content, W_f (dwb), may be represented by the expression:

$$t = \frac{r^2 \rho_l L_s (W_0 - W_f)}{3h_c(\vartheta_a - \vartheta_s)(1 + W_0)}, \qquad (23)$$

where r is the radius of the droplet, and ρ_l is the density of the liquid.

In practice, not all the moisture is removed under constant rate conditions. More detailed treatments of drying of droplets during spray drying are available in the literature.

See also: **Drying**: Chemical Changes; Equipment Used in Drying Foods

Further Reading

Brennan JG (1994) *Food Dehydration – A Dictionary and Guide*. London: Butterworth-Heinemann.

Brennan JG, Butters JR, Cowell ND and Lilly AEV (1990) *Food Engineering Operations*, 3rd edn. London: Elsevier Applied Science.

Karel M (1974) Fundamentals of drying processes. In: Spicer A (ed.) *Advances in Pre-concentration and Dehydration of Foods*, pp. 45–94. London: Applied Science.

Masters K (1991) *Spray Drying Handbook*, 5th edn. New York: Longman Scientific and Technical.

Mujumdar AS (1995) *Handbook of Industrial Drying*, 2nd edn. New York: Marcel Dekker.

Wang N and Brennan JG (1995) A mathematical model of simultaneous heat and moisture transfer during drying of potato. *Journal of Food Engineering* 24: 47–60.

Drying Using Natural Radiation

M. Mathlouthi, Université de Reims Champagne-Ardenne, Reims Cedex, France

Background

Drying of certain crops and foodstuffs is a very ancient and relatively easy technique for food preservation. Sun-drying has been used by early civilizations and is still used in many tropical and subtropical countries where solar energy remains a low-cost heat energy. However, the traditional method, which involves spreading the product in a thin layer on the ground or on a platform and exposing it directly to the sun has its drawbacks. The quality of the dried product is reduced owing to color deterioration by direct exposure to UV radiation, as well as damage to unprotected products by insects, rodents, and other animals, and microbial spoilage, especially as the relative humidity increases. Improvements in sun-drying have led to solar drying. This technique uses sun as a heat source, which increases the drying capacity of air by heating it and reducing its relative humidity in indirect solar dryers. Recent advances have produced cheaper solar dryers, suitable for the drying requirements of the crop, as well as an improved quality of the dried product. Whatever the drying system, using either direct exposure to the sun or circulation (natural or forced) of heated air in solar dryers, the driving force for dehydration remains the gradient of water vapor pressure between the product and the surrounding air. This explains why natural ventilation is convenient in dry areas like arid zones of Africa or the Artic, whereas the hot and humid air of certain tropical regions is inefficient for drying crops. Such climatic conditions necessitate the use of adapted solar dryers.

To take advangae of solar energy, collectors are used, generally made from low-cost packaging films readily available in developing countries where solar dryers have been introduced with the aim of reducing crop losses. Solar-energy-based dryers are generally low-temperature drying systems. Equilibrium in relative humidity between the air and the product is generally reached at a constant ventilation rate and an average temperature around 50 °C. This kind of drying concerns crops that are dried in bulk in their storage devices. Fruits with a high sugar and acid content, like raisins, figs, apricots, or dates, are particularly well suited to natural drying outdoors. Sometimes, a pretreatment is needed if the dehydration rate and storage stability of the product are to be improved.

Solar Drying Systems

Three types of application of solar energy are used to dry crops:

- Direct – absorption of solar radiation where natural convection of the surrounding air is the origin of mass and heat transfer;
- Indirect – application of solar radiation, which is used to preheat air in a solar collector before blowing it on the crop tray and releasing it through a chimney;
- Mixed-type drying systems, which use both direct solar heat and natural ventilation of the drying chamber with air heated in a solar air heater.

There is another type of classification that considers solar drying systems as passive or active, depending on whether they rely completely on solar radiation and natural air flow, or only partly on solar energy, and also on other heating systems, as well as on motorized fans. All active solar dryers use forced convection.

Direct Solar Radiation Drying

Traditional open-to-sun drying This involves spreading the crop, or the food on the ground, or placing it on horizontal shelves exposed to sun, air, and natural ventilation. From time to time, the crop is stirred to reduce the drying time. This type of sun drying is common in developing countries. Although this method seems simple and cheap, very often, crop losses result from fungal and insect infestation, or unexpected rain. The quality of the dried product is sometimes poor, such as the loss of riboflavin in couscous dried under UV sun radiation as compared with industrial hot-air-dried products. Even though large amounts of fruits and vegetables are still sun-dried, the development of indirect solar dryers seems the solution for the future if it is necessary to decrease the drying time, improve the quality, and minimize crop loss.

Direct solar-energy absorption dryer In this type of dryer, the solar radiation is used directly in the dryer, which is, in this case, a cabinet dryer equipped with a fan and sometimes an additional electric heat source (Figure 1), or a greenhouse tunnel covered with a glass or a polyethylene plastic cover about 150 μm in thickness and, in some cases, with UV and far-infrared protection. In the greenhouse-type dryer (Figure 2), the solar-heated air is circulated through the product by a fan. A metallic structure made of arches connected by longitudinal tubes and supported on vertical lateral tubes is used. This structure is covered with a polyethylene film of the same type as

that used in cultivation greenhouses. Recycling of air in a transparent plastic tube is possible. Fans at the air inlet and exit allow an increased ventilation rate.

Storage-type dryers have roofs that act as a solar collector. In other cases, both the roof and the wall act as collectors. The wall is usually black painted concrete covered with glass and is both a collector and a means of heat storage.

Indirect solar dryer In this type of dryer (also called a distributed-type air-circulation solar dryer), the crop is placed on shelves in an opaque drying chamber. The drying air is heated in a separate solar collector and distributed by either natural or forced (fan) circulation. In this dryer, the crop is not exposed to sun, thus preventing vitamin loss and other UV-sensitive reactions like lipid oxidation. The air temperature is higher than in direct sun-dryers, which reduces the drying time. A typical solar dryer is shown in Figure 3. Clearly, the equipment and maintenance costs of this type of drier are more than that of the direct sun dryer. Efficiency is increased if forced-air convection is installed. The optimization of indirect solar dryers also depends on the quality of the solar energy heating the collectors.

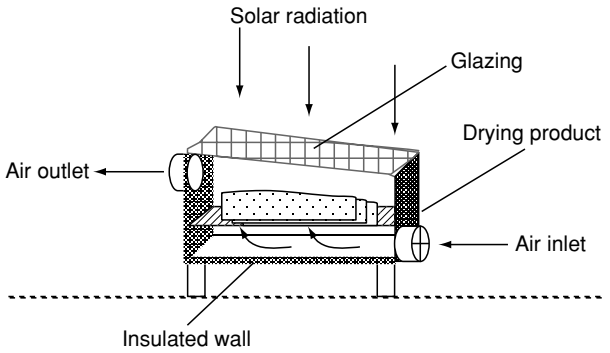

Figure 1 Cabinet solar dryer.

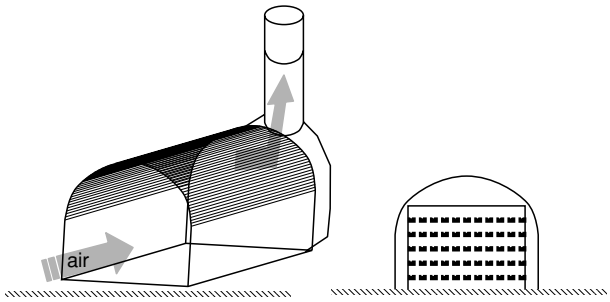

Figure 2 Greenhouse-type solar dryer.

Mixed solar dryer This type of dryer incorporates both direct and indirect solar-energy drying. The dryer consists of a preheated air chamber, a drying chamber and a chimney. However, the drying chamber in this case is glazed so that the product also receives direct solar energy as well as a flow of drying air (Figure 4). Depending on the countries and product, the heat absorber may be made of low-cost recovered byproducts, as in rice drying, where a layer of burnt rice husk covered with a transparent plastic sheet on an inclined bamboo network is used as the air heater's absorber. Direct absorption of solar heat in the drying chamber takes place through a transparent plastic film, which also protects the crop against rain and other degradation agents. In the multi-stacked dryer, different products can be dried at the same time on a series of shelves. Also, the drying chamber has a column exposed on one side to the sun through glazing and, on the opposite side, a wooden access door and vents, thus obviating the need for a chimney.

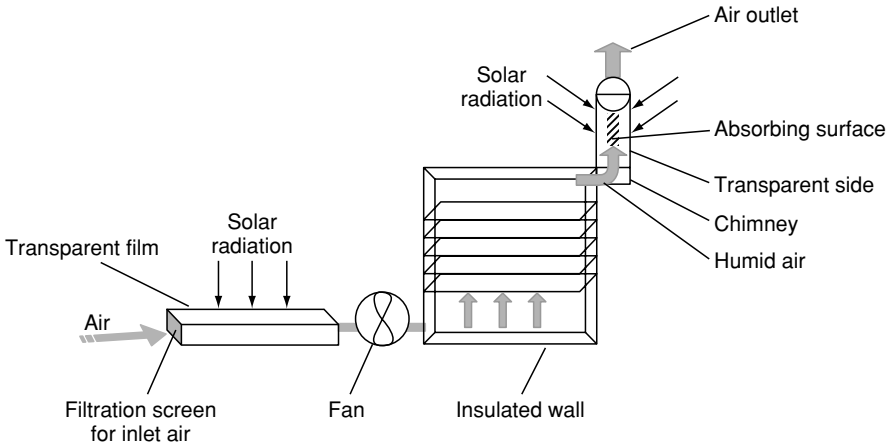

Figure 3 Indirect solar dryer consisting of a collector, drying chamber, and chimney.

Solar Heat Collectors

Solar collectors take advantage of the thermal energy of solar radiation. There are two types: the bare flat-type collector and the covered-plate heat absorber. The surface of the plate is usually painted black; this absorbs sun radiation and transmits it in the form of heat to the drying air. The air flow through the collector is by natural convection or by means of a fan. The efficiency of a collector can be determined by measuring the inlet and outlet temperatures, the surface of the collector and flow rate of the air, and the solar energy available.

Bare Flat-plate Collectors

These collectors consist of a simple air duct with the upper side acting as a heat absorber and the lower surface acting as an insulator (see Figure 5). Important losses of thermal energy are observed at the upper surface, which limits the use of such collectors to low-temperature drying like crop drying in storage silos. In this type of collector a flat plate of aluminum can be used as a heat absorber. The heat loss is reduced by placing the air duct in a box filled with glass wool insulation.

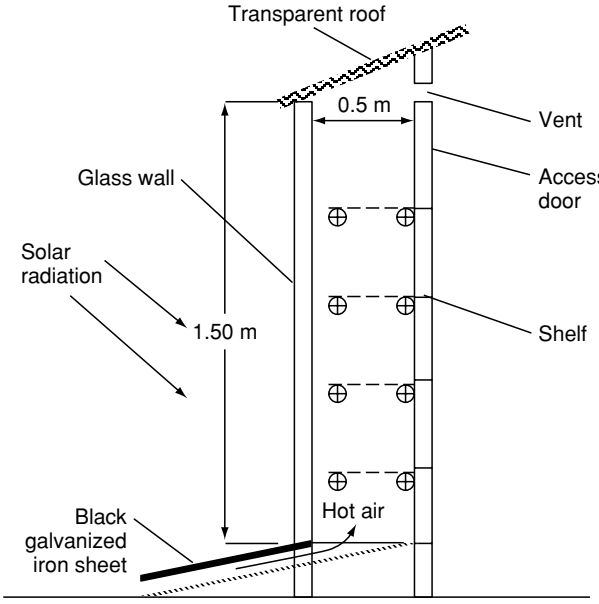

Figure 4 Mixed solar-energy dryer.

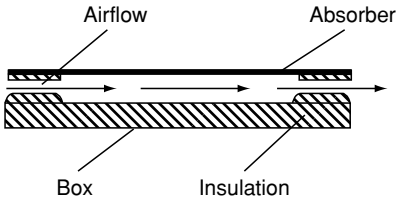

Figure 5 Flat-plate collector.

Covered-plate Heat Collector

To minimize heat losses from the upper side of the collector, glass, plexiglass, or transparent plastic is used as a cover material, placed parallel to the metallic absorber plate. Covered-plate solar-energy air heaters are much more efficient than bare-plate collectors. The air to be heated can be sucked between the absorber and the transparent cover, or sometimes underneath the absorber to prevent dust contamination at the surface of the absorber. In this case, the air circulates between the lower face of the absorber plate and the insulation layer, the upper surface of which is in contact with a static layer of air underneath the glazing. This type of back-pass air heater (Figure 6) is more efficient than one where air flows between the absorber and the cover.

Efficiency of Solar-heat Collectors

Efficiency is defined as the ratio of thermal energy used to heat the drying air over the total solar energy absorbed by the collector in the same time period. The heat used to warm air in the collector is expressed by:

$$Q_a = m_a C_{pa}(T' - T),$$

where Q_a is the heat transmitted to air in the collector ($J\,s^{-1}$), m_a is the mass flow rate of air through the collector ($kg\,s^{-1}$), C_{pa} is the specific heat capacity of air ($J\,kg^{-1}\,K^{-1}$), T' is the collector outlet temperature, and T is the inlet temperature (generally ambient).

If A is the collector area (m^2) and I the flux of incident solar radiation ($W\,m^{-2}$), the total incident energy of the exposed surface of collector is $Q_i = AI$, and the efficiency ratio (η) can be expressed as:

$$\eta = \frac{Q_a}{Q_i} = m_a C_{pa} \frac{(T' - T)}{AI}.$$

In forced-convection air dryers, m_a is easy to measure, as well as the inlet and outlet temperatures. For natural convection, the flow rate of air is difficult to

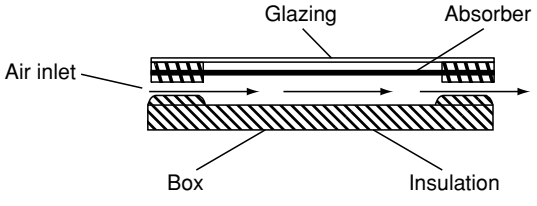

Figure 6 Back-pass air heater.

determine precisely, hence the efficiency ratio is measured. Incident solar radiation absorbed by collector depends on the properties of the material used for the design of the collector like absorbance, transmittance, and reflectance. It is also a function of the position of the sun relative to the collector, the solar altitude angle, and the sunrise and sunset of the collector orientation. A value of 1 kW per square meter of exposed surface may be considered as an optimum.

Application of Solar Drying to Selected Crops

Solar drying is an effective mode of dehydration especially for certain crops (cereal grains, grapes, figs, sweet pepper, thyme) in countries with good sunshine during the harvest period. Many developing countries have sufficient insolation for sun or solar drying, but the dryer needs to be of the right design to minimize crop losses resulting from contamination by insects, molds, and poor storage conditions. In developing countries, greenhouses can be used as solar dryers during the warm period of the year, thus reducing the cost of investment of setting up a new dryer. Solar collectors may be made from low-cost polyethylene film traditionally used to cover greenhouses. In all cases, covered heat collectors and drying chambers are preferable to open-air sun dryers to prevent crop spoilage.

Solar Drying of Grapes

Traditional drying of grapes is still used in many countries and involves spreading or hanging the product on galvanized wires (with a coverage of about 15–20 kg m^{-2}) on the racks. A typical dryer contains six to 10 racks each about 50 m long, 2 m tall, and 1.5 m wide. Racks are placed on open high land to allow free air flow. The period of drying varies from 8 to 20 days, depending on the grape variety, treatment, and climate conditions. The water content in the grapes is reduced from about 3 kg of water per kilogram of dry matter in fresh product to about 0.16 kg of water per kilogram of dry matter in the dried product. Solar dryers using either direct or indirect solar energy have higher drying rates than sun dryers. These rates are improved when grapes are pretreated. Pretreatment solutions contain 4–5% K_2CO_3 and 1–2% dipping oil. Changing the surface tension of grape by dipping for a few minutes in dipping solution increases water diffusion through the waxy cuticle and decreases the drying time.

The quality parameters of solar-dried grapes, especially color, moisture content, and acidity, are more uniform than those of the sun-dried product. Storage stability is improved with SO_2 gas pretreatment and conditioning in a modified atmosphere (1% O_2; 20% CO_2) packages.

Solar Drying of Other Crops

For obvious economic reasons, if a solar dryer is preferred to an open-air sun dryer, it should work to full output. Therefore, it has to be a multipurpose dryer equipped for both low-temperature (below 50 °C) and high-temperature (above 50 °C) drying, eventually using the calories produced by the combustion of agricultural waste as part of the total thermal energy required. Tunnel greenhouse dryers are very efficient during sunny days and can be used with an auxillary combustion heating system to ensure a continuous operation. This type of solar dryer is used for sweet pepper and garlic in Argentina. In terms of mold contamination, the quality of the product is better on sunny days than on cloudy, humid days. A complementary heating source is needed to improve drying during humid periods and to increase the working time. Products like thyme can be dried at about 50 °C using a wire basket solar dryer. Sun-dried thyme herb is suitable for packaging or grinding as a powder. In terms of the essential oils extracted after drying, sun-dried thyme is of a better quality than that of oven-dried thyme, but the percentage of oleoresin is the same for both drying methods.

Quality of Solar-dried Products

Sun drying usually takes too long, and even solar drying lasts several days. Such a long process favors degradation reactions, especially when the water activity is relatively high, above 0.60. Moreover, the storage conditions should be controlled so that the temperature is maintained below 10 °C for grain and nut storage, and the relative humidity below 40%. This requires cooling and dehydration of air. However, if precautions (e.g., the use of clean crops) are taken to minimize spoilage, high-quality sun-dried products can be produced. Other precautions include the use of covered plate absorbers and the use of clean drying air. The first stages of dehydration (first and second day) should be carried out on sunny and dry days. The most dangerous contamination is fungal damage, which can yield mycotoxins. Among the mycotoxins, such as ochratoxin, citrinin, and ergosterol found in grains and nuts, aflatoxins are the most important as regards toxicity, occurrence, and economic consequences, especially in developing countries.

Other degradation reactions such as browning, lipid oxidation or vitamin loss may be minimized by using adequate protection in the dryer. For example,

it is preferable to use low temperatures as well as the UV radiation filtration using colored glazing.

Conclusion

Solar drying has proved to be technically and economically valuable for several crops. It is, however, necessary to develop large-scale dryers that may be used throughout the year for different products to make them attractive to the farmers.

Protection against UV radiation, dust, insects, mold, and other sources of contamination, as well as temperature and relative humidity control, are needed to improve the quality of the product. Likewise, the storage conditions of solar-dried products should be tailored to the crop and the packaging optimized if a long shelf-life is required.

See also: **Drying**: Theory of Air-drying; Physical and Structural Changes; Equipment Used in Drying Foods

Further Reading

Condori M, Echazu R and Saravia L (2001) Solar drying of sweet pepper and garlic using the tunnel greenhouse driers. *Renewable Energy* 22: 447–460.

Ekechukwu OV and Norton B (1999) Review of solar-energy drying systems II: an overview of solar drying technology. *Energy Conversion & Management* 40: 615–655.

Exell RHB (1980) A basic design theory for simple solar dryes. *Renewable Energy Review Journal* 1: 1–14.

Pangavhane DR and Sawhney RL (2002) Review of research and development work on solar dryers for grape drying. *Energy Conversion & Management* 43: 45–61.

Fluidized-bed Drying

J G Brennan, The University of Reading, Reading, UK

Principles of Fluidization

If air is made to flow through a bed of solid particles supported by a perforated plate or grid, frictional drag will cause a pressure drop to develop across the bed. The drag, and hence the pressure drop, will increase as the air velocity increases. When the frictional force overcomes the weight of the particles they move apart, and the bed expands and begins to behave like a fluid. In this state it is known as a 'fluidized bed.' The air velocity at which fluidization first occurs is known as the 'incipient (or minimum) fluidization velocity.' This velocity can be calculated, for spherical particles, using the relationship:

$$v_f = \frac{(\rho_s - \rho)g}{\mu} \frac{d^2\varepsilon^3}{180(1-\varepsilon)}$$

where v_f (m s^{-1}) is the minimum fluidization velocity, ρ_s (kg m^{-3}) is the density of the solid particles, ρ (kg m^{-3}) is the density of the air, g (m s^{-2}) is acceleration due to gravity, μ (N s m^{-2}) is the viscosity of the air, d (m) is the diameter of the particles, and ε is the voidage of the bed (the fraction of the bed that is void). As the air velocity increases above the minimum value, bubbles of air pass up through the bed. These bubbles are small when they leave the perforated plate. As they pass up through the bed they coalesce to form larger bubbles. The movement of these bubbles causes mixing of the solid particles in the bed. As the air velocity increases, mixing becomes more vigorous. At some stage particles will leave the surface of the bed and fall back on to it. At a still higher velocity the particles become entrained in the air and are carried away in the exhaust air stream. The velocity at which this first occurs is known as the 'entrainment velocity.' This velocity may be calculated, for spherical particles, by means of the relationship:

$$v_c = \sqrt{\left(\frac{4d(\rho_s - \rho)}{3C_d\rho}\right)}$$

where v_c (m s^{-1}) is the entrainment velocity and C_d is the drag coefficient. C_d has a value of 0.44 when the Reynolds number of the air is in the range 500–200 000. When heated air is used to fluidize a bed of particles containing moisture, dehydration of the particles will occur. This is the principle of the fluidized-bed drier.

Fluidized-bed Driers

Food particles suitable for fluidized-bed drying should be in the size range 20 μm to 10 mm, uniform in size and shape, and not very sticky. They should not break up easily when subjected to impacts and abrasion in the bed. Some food materials, e.g., cereal grains, fluidize readily over a wide range of moisture contents. Other foods will only fluidize at low moisture content. In the latter cases the wet product may be fluidized by dispersing it in a bed of already dried material.

Fluidized-bed drying chambers equipped with a mechanical agitator to help disperse wet or sticky materials are available. The agitator is located near

Table 1 Typical gas velocities for fluididizing different-sized particles

Average particle size (μm)	*Velocity ($m\,s^{-1}$)*
100–300	0.2–0.4
300–800	0.4–0.8
800–2000	0.8–1.2
2000–5000	1.2–3.0

From Hovmand CV (1995) Fluidized bed drying. In: Mujumdar AS (ed.) *Handbook of Industrial Drying*, 2nd edn, pp. 195–248. New York: Marcel Dekker with permission.

the feed inlet. It disintegrates the feed and disperses the pieces into already dry material, thus facilitating fluidization. Some semisolid foods may be converted into a form suitable for fluidized-bed drying by mixing them with some dried material, and extruding and cutting the mixture into suitably sized pieces.

The air velocity should be as high as possible to promote rapid drying. The upper limit is often determined by the loss of fine product, created in the bed, with the exhaust air. Some loss of fines may be tolerated in order to achieve high drying rates. The fine particles may be recovered from the outgoing air by means of cyclones or fabric filters. Typical air velocities for fludizing particles with densities in the range 1000–2000 $kg\,m^{-3}$ are presented in Table 1.

For most food applications the support for the bed is a simple perforated plate. For very fine particles more complex designs incorporating jets or bubble caps, similar to those in distillation plates, may be used. Fluidized-bed driers usually operate at atmospheric pressure. However, by suitable location of fans they may be run at pressures above or below atmospheric pressure. Heater plates may be incorporated into the drying chamber to improve heat transfer. These are not common in food applications. The drying conditions can be accurately controlled in batch-operated fluidized beds. High rates of drying and uniform moisture content in the product can be achieved. However, batch units are only used for small-scale operations.

There are two types of continuous fluidized-bed drier which differ in the pattern of flow of the particles and air. The plug-flow type (Figure 1a) has a relatively long narrow bed. The feed is introduced at the wet end and moves to the dry end in a plug-flow pattern where the dry product is discharged over an adjustable weir. The distribution of the residence time of the particles is relatively narrow so that uniform drying is accomplished. However, its application is limited to feed particles which are readily fluidizable. In the well-mixed, or backed-mixed, fluidized bed (Figure 1b) the ratio of the length of the bed to its width is unity or below. The particles are vigorously mixed and the bed will have a relatively uniform

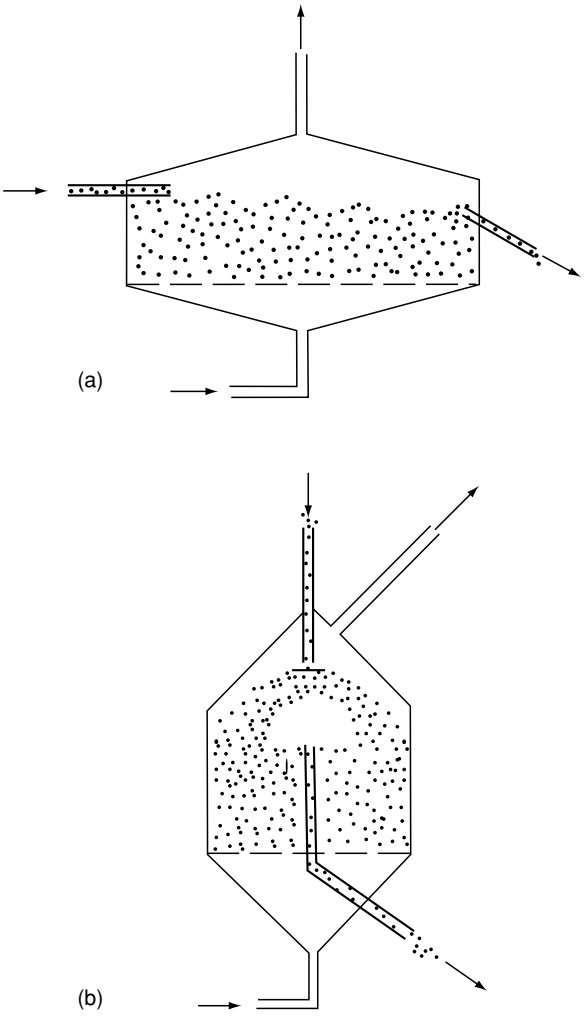

Figure 1 Representations of (a) a continuous plug-flow fluidized-bed drier and (b) a continuous well-mixed fluidized-bed drier. Adapted from Mujumdar AS (ed.) (1995) *Handbook of Industrial Drying*, 2nd edn, by courtesy of Marcel Dekker: New York.

temperature. The wet feed is introduced into a bed of relatively dry particles and is quickly dispersed. Thus, it can handle feeds that are difficult to fluidize. On the other hand, there is a broader distribution of particle residence time, as compared with the plug-flow design, and the product moisture content may be less uniform. The well-mixed bed is best at removing surface moisture, whereas the plug-flow type removes moisture from within the particles. The latter type is most widely used for food applications. Figure 2 shows a continuous plug-flow fluidized-bed drier equipped with a cyclone for the recovery of fines.

Vibrated Fluidized-bed Driers

Fluidized-bed driers may be fitted with vibrating bases, to help move the product horizontally from

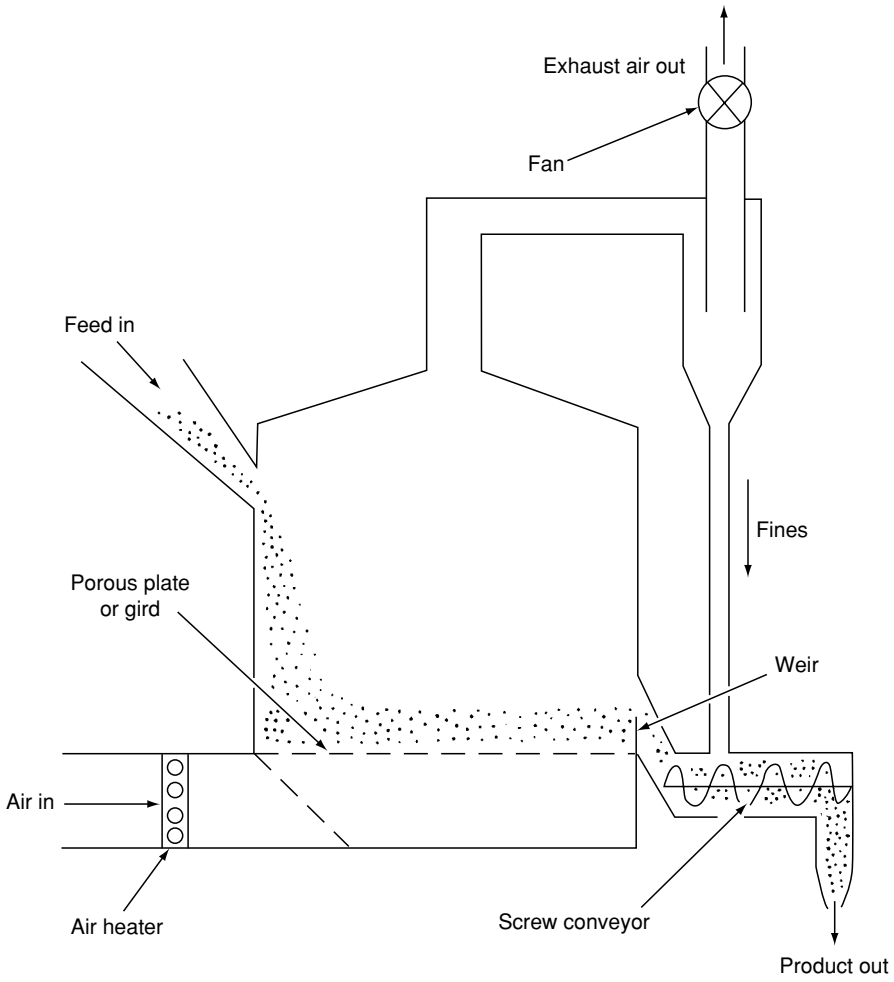

Figure 2 A continuous plug-flow fluidized-bed drier equipped with a cyclone for the recovery of fines. From Brennan JG, Butters JR, Cowell ND and Lilly AEV (1990) *Food Engineering Operations*, 3rd edn. London: Elsevier Applied Science with permission.

the wet end to the dry end. This motion is gentle and suitable for handling fragile materials and agglomerates susceptible to mechanical damage. A vibrating bed is capable of handling particles less uniform in size as compared with the stationary-type bed. A vibrating bed is also better at handling sticky materials. The air velocity required to keep the particles moving is usually much less in a vibrated bed than in a stationary one. Bases vibrate in the frequency range 5–25 Hz and a half amplitude of a few millimeters. Such beds are also known as 'vibro-fluidizers.'

Multistage Fluidized-bed Drying

A number of fluidized-bed driers may be used in series. This enables better control of the drying conditions at different stages in the cycle and can yield a product with a more uniform moisture content as compared with a single-stage drier. The use of a multistage system may also lead to energy savings. One possible combination is a well-mixed bed, to remove surface moisture, followed by a plug-flow bed to complete the drying. Alternatively, a number of plug-flow beds may be used in series. The latter arrangement is mainly used for the food applications. Up to six stages may be used. Vibrating beds are usually employed. The particles fall under gravity from one stage to the next. Such an arrangement is also known as a 'cascade system.'

Applications

Single-stage and multistage fluidized-bed drying systems are used for a wide range of food materials. These include peas, sliced beans, diced carrots, onions, potato granules, meat cubes, grains, flour, salt, and sugar. The air temperature is selected to suit each product. Vibrating fluidized beds are used as after driers and coolers following spray drying. The

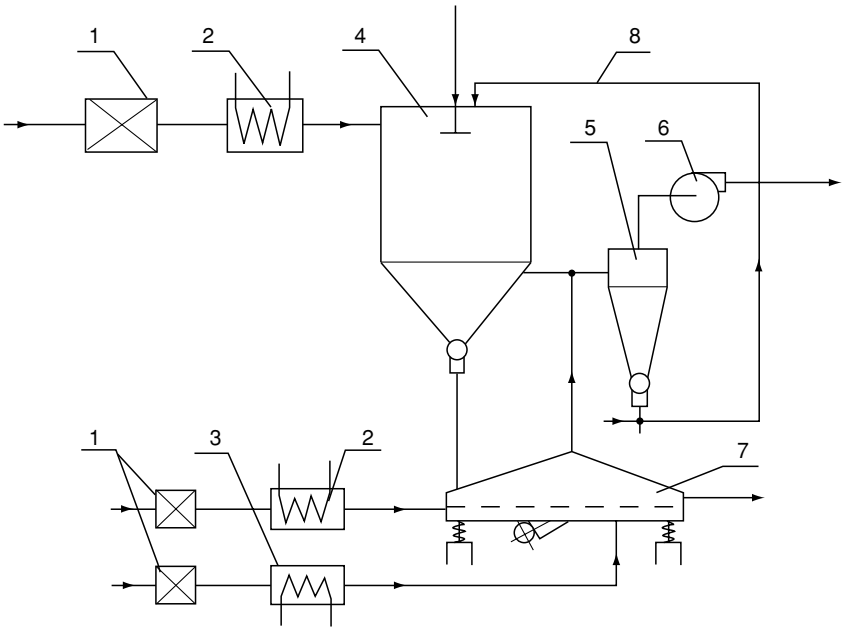

Figure 3 A two-stage spray-drying process for milk: 1, air filter; 2, heater; 3, cooler; 4, spray drier; 5, cyclone; 6, exhaust fan; 7, fluidized-bed drier; 8, return line for powder. From Mujumdar AS (ed.) (1995) *Handbook of Industrial Drying*, 2nd edn, by courtesy of Marcel Dekker: New York.

powder from the spray drier has a moisture content of 5–7%. This is reduced to 3–4% on a vibrofluidizer using dry air at a relatively low temperature. Such two-stage drying is said to facilitate better control over the quality of the product and lead to savings in energy as compared with straight-through spray drying. Some agglomeration of the powder particles may occur in the fluidized bed which can improve the reconstitution characteristics of the product. Such a system is shown in Figure 3. The product may be cooled in the same vibrofluidizer or in a second unit in series with it. Cooling the product facilitates handling and packaging. Spray driers are available with fluidized beds built into the main drying chamber. These are known as integrated fluidized-bed spray driers.

Spouted-bed Drier

One particular design of fluidized-bed drier is the spouted bed. In this type of drier the drying air enters at high velocity through a nozzle located centrally in the conical base of the chamber (Figure 4). This creates a spout of fast-moving particles up through the center of the bed. These particles return to the base of the chamber as a slow-moving annular bed surrounding the spout. A proportion of the inlet air flares out into this annular bed and flows upwards, countercurrent to the downward-moving particles, at a relatively low velocity. In the spout, high rates of heat and mass transfer occur, resulting in high initial rates of drying. As a result of evaporative cooling the temperature of the particles in the spout is kept relatively low. The later stages of drying take place in the annular bed under relatively mild conditions. This type of drier is suitable for drying relatively large particles, greater than 5 mm in diameter. It can cope with heat-sensitive materials. During start-up a high

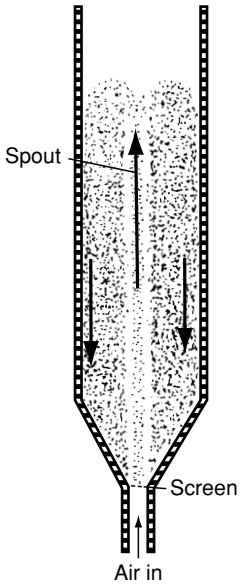

Figure 4 The principle of the spouted-bed drier. From Brennan JG (1994) *Food Dehydration – A Dictionary and Guide*. Oxford: Butterworth-Heinemann with permission.

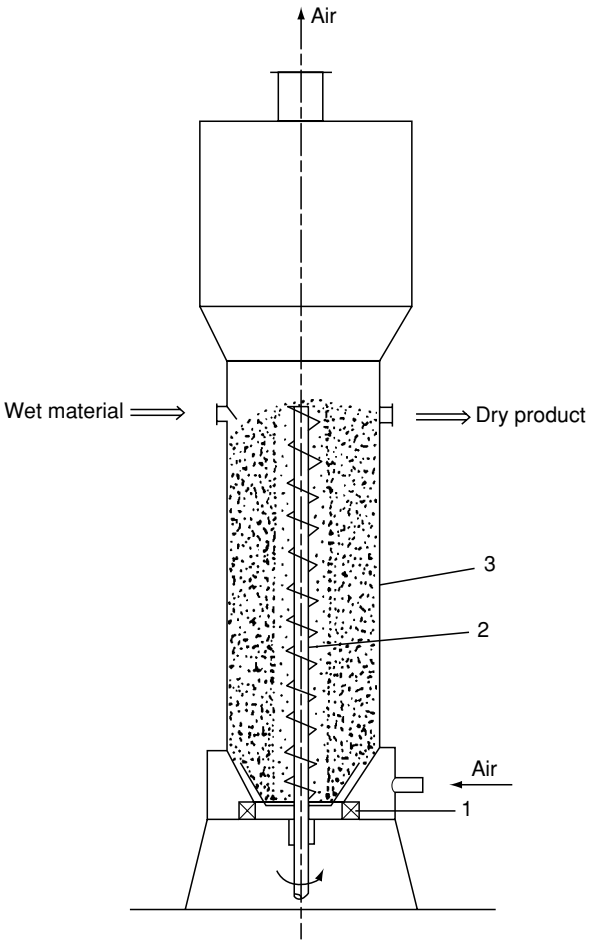

Figure 5 Spouted-bed drier with tangential air inlet and central conveyor screw: 1, tangential air inlet; 2, conveyor screw; 3, drier body. From Mujumdar AS (ed.) (1995) *Handbook of Industrial Drying*, 2nd edn, by courtesy of Marcel Dekker: New York.

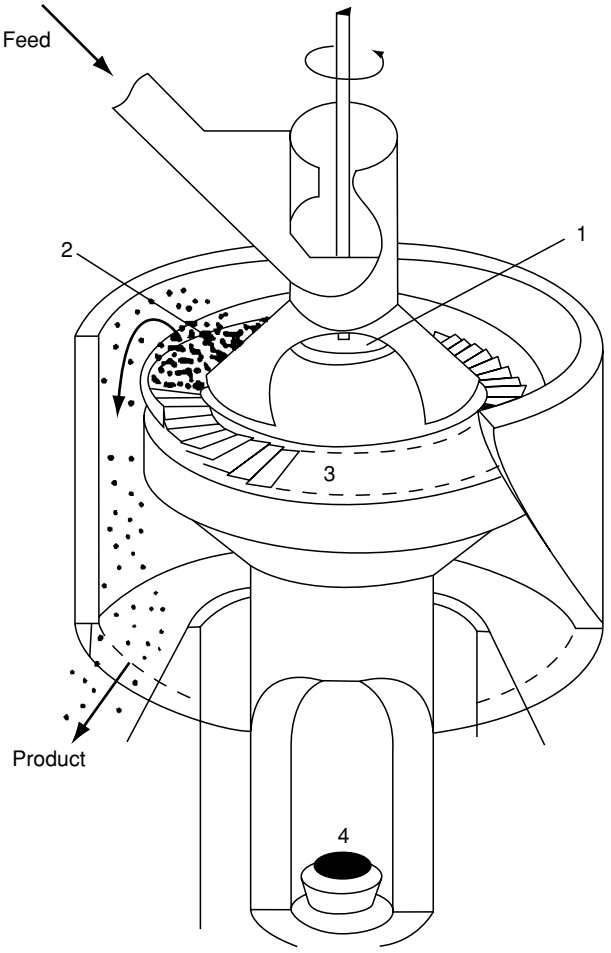

Figure 6 The toroidal-bed drier: 1, rotating disk distributor to deliver raw material evenly into the processing chamber; 2, rotating bed of particles; 3, fixed blades with hot air passing through at high velocity; 4, burner assembly. From Brennan JG (1994) *Food Dehydration – A Dictionary and Guide*. Oxford: Butterworth-Heinemann with permission.

pressure drop has to be developed across the bed to form the spout. Once this is formed the pressure drop decreases to a value less than that which develops across a conventional fluidized bed. A spouted-bed drier featuring tangential air inlet has also been reported (Figure 5). The air enters through slits at the bottom of the chamber. An open screw conveyer is located in the spout to control the upward movement of the particles, independently of the air velocity. This enables smaller particles to be dried as compared with a conventional spouted-bed drier. A high early air pressure drop is not required to initiate spouting. Wheat, corn, and diced carrot are among the foods which have been successfully dried in spouted-bed driers.

Toroidal-bed Drier

This is another novel design of fluidized-bed drier (Figure 6). A high-velocity stream of heated air enters the base of the process chamber through blades or louvers which impart a rotary motion to the air. This creates a compact rotating bed of particles, which may vary in depth from a few millimeters to 50 mm. High rates of heat and mass transfer are attainable in this bed, resulting in high initial rates of drying. This drier can accommodate a wide range of particle sizes and shapes and can be operated on a batch or continuous basis. Not much information on the performance of this equipment for drying foods has been published. However, preliminary studies, in which this author has participated, indicate that it has considerable potential for drying small particulate foods and liquid foods on a similar principle to the spouted bed above. It can be used for roasting products such as coffee beans and nuts, expanding pasta pellets, and possibly cooking or sterilizing other particulate. In a current study it is being used to produce puffed potato cubes.

Centrifugal Fluidized-bed Drier

This consists of a cylindrical chamber rotating about a horizontal axis (Figure 7). The cylinder wall is perforated and a high-velocity (up to 15 m s^{-1}) stream of air flows through these perforations and across the drying chamber. The feed enters at one end of the chamber and the dried product leaves over a weir at

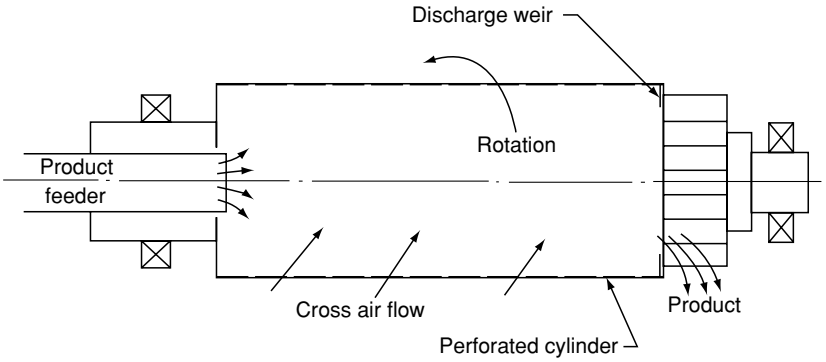

Figure 7 Centrifugal fluidized-bed. From Mujumdar AS (ed.) (1995) *Handbook of Industrial Drying*, 2nd edn, by courtesy of Marcel Dekker: New York.

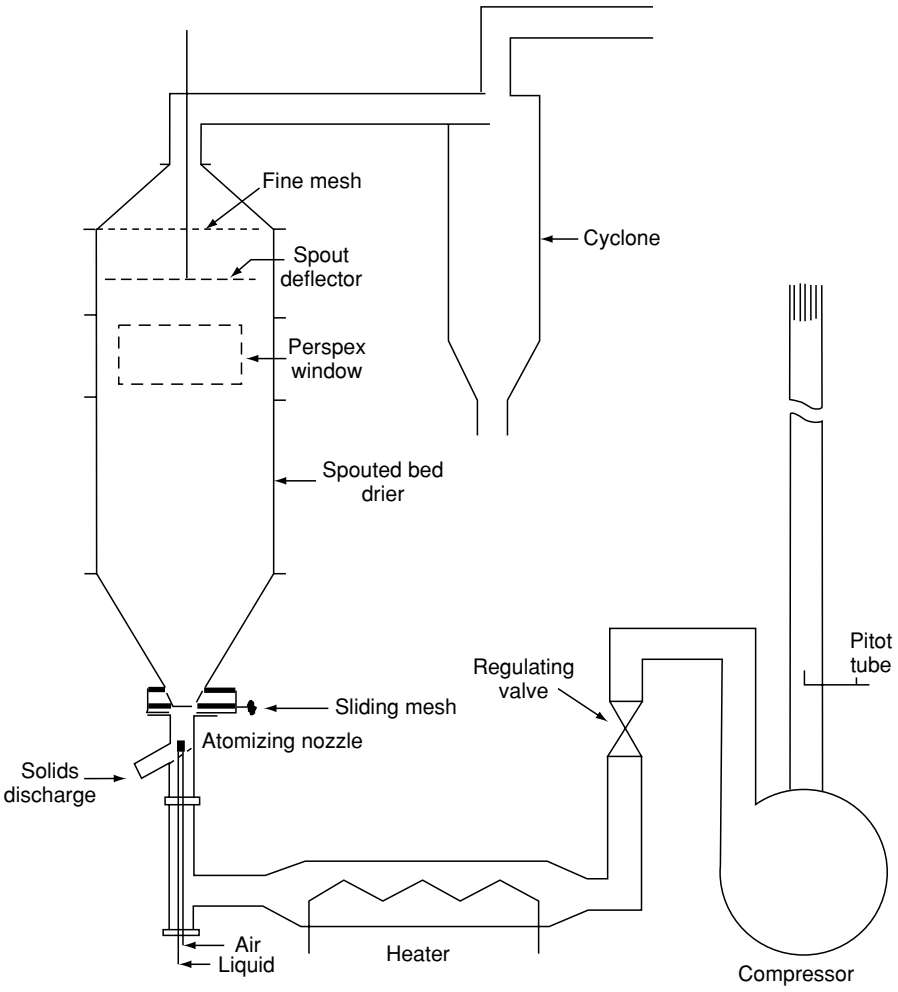

Figure 8 Experimental spouted-bed drier for liquid foods. From Ochoa-Martinez AE, Brennan JG and Niranjan K (1993) Spouted bed dryer for liquid foods. *Food Control* 4: 41–45 with permission.

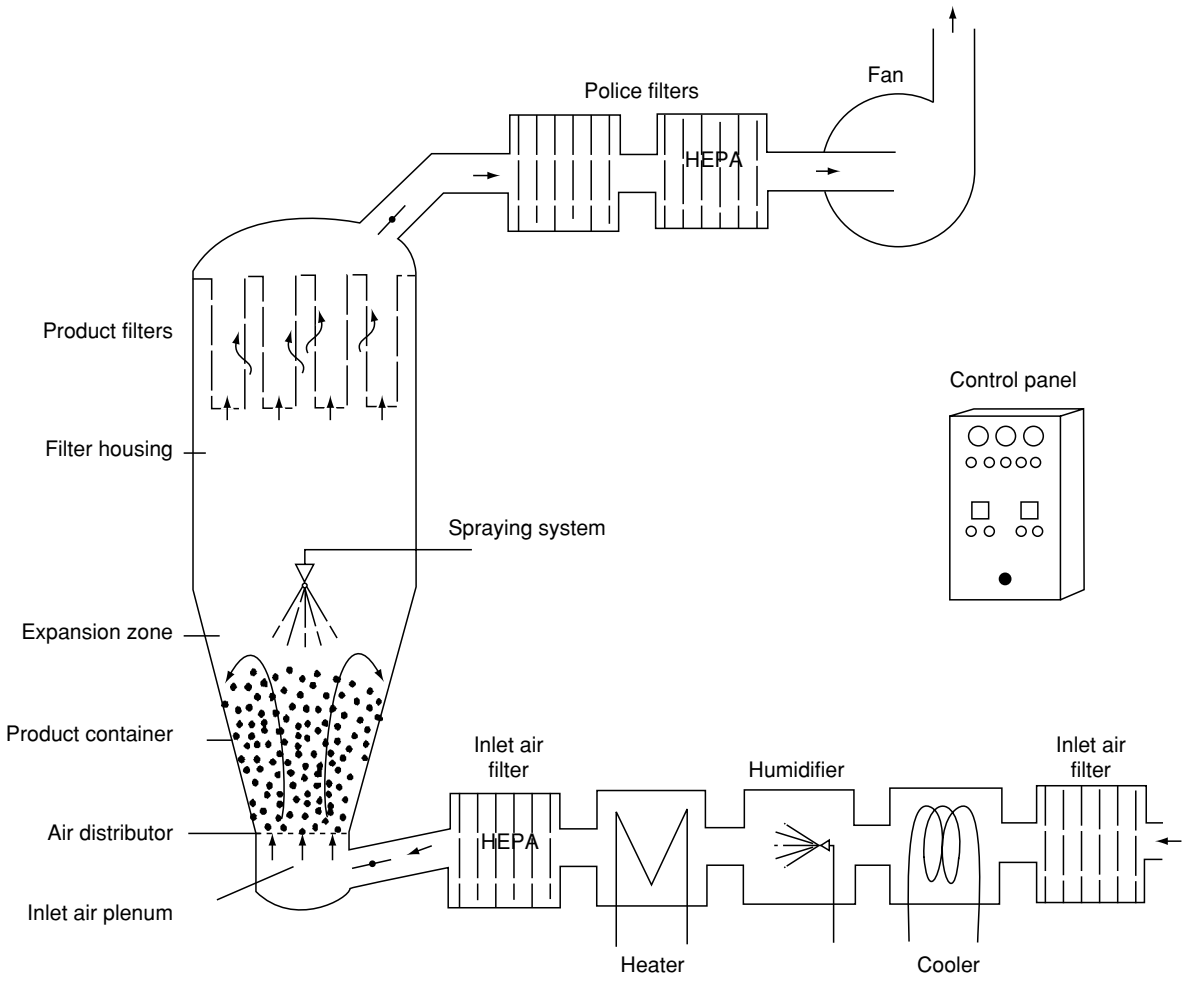

Figure 9 Batch fluidized-bed agglomerator. Courtesy of Niro Aeromatic AG.

the other end. Some 10–20% of the chamber volume is occupied by the food particles. These are made to fluidize during part of each revolution and form a stationary bed during the rest of the cycle. This drier has been used to dry sliced and shredded vegetables which are difficult to dry in conventional fluidized-bed driers and for puffing vegetable pieces, including rice grains, to produce a quick-cooking product.

Novel Applications for Fluidized and Spouted Beds

Fluidized beds of inert particles have been used to dry solid and liquid food materials. A whirling fluidized bed of glass beads has been used to dry diced vegetables and sticky products such as cooked rice. Mushrooms, carrot, beef, and shrimp have been dried at low temperature in a fluidized bed of activated alumina. Paste and liquid foods have been dried in spouted beds of inert particles. Blood has been dried in this way. In a recent study, model liquid foods were sprayed on to polypropylene beads in an experimental spouted bed drier (Figure 8). Drying occurred by a combination of convected heat from the air and conducted heat from the beads. The performance of this drier was compared with that of a laboratory spray drier. Early results indicated that the drying process in the spouted bed was rather more severe than in the spray drier. However, the evaporative capacity of the spouted bed was some six times that of the spray drier, which was a comparable size.

Fluidized beds may be used to produce agglomerated or granulated products. A liquid is sprayed on to a fluidized bed of particles, coating them. Collisions between particles cause them to adhere together to form agglomerates. When the agglomerates reach the desired size they are dried to a stable moisture content by raising the temperature of the fluidizing air. Both batch and continuous agglomerating systems are available (Figure 9).

A relatively new application for fluidized beds is the puffing of fruit and vegetable pieces. The pieces

are predried in a conventional hot-air drier such as a cabinet or tunnel, to form an impervious layer on their surface. They are then introduced into a fluidized bed, operating at a relatively high temperature. The water evaporates internally within the pieces, causing them to expand and producing an open porous structure. The puffed product is then dried down to a stable moisture content in a cabinet or tunnel drier. This has the effect of reducing the overall drying time and produces a product which reconstitutes rapidly. Alternatively, the dry product may be used as a snack food, low in fat, compared with fried products.

Drying Difficult Materials

To contain hazardous materials or unpleasant odors semiclosed-cycle fluidized-bed systems are available. In such systems a large proportion of the exhaust air is recycled. The air from the cyclones goes to a condenser/scrubber which removes the water from it as condensate. Only a small amount of air is released to the atmosphere; the rest is recycled through the heater and drier. The air vented from the system may be burnt off if necessary. Similar systems are available for spray driers. Application of these specialized types of driers to food dehydration is limited at the present time. However, as regulations relating to pollution of the environment become more stringent, the use of such systems may increase.

See also: **Drying**: Theory of Air-drying; Physical and Structural Changes; Chemical Changes

Further Reading

Brennan JG (1994) *Food Dehydration – A Dictionary and Guide*. Oxford: Butterworth-Heinemann.

Brennan JG Butters JR, Cowell ND and Lilly AEV (1990) *Food Engineering Operations*, 3rd edn. London: Elsevier Applied Science.

Filkova I and Mujumdar SS (1995) Industrial spray drying systems. In: Mujumdar AS (ed.) *Handbook of Industrial Drying*, 2nd ed, pp. 263–307. New York: Marcel Dekker.

Hovmand S (1995) Fluidized bed drying. In: Mujumdar AS (ed.) *Handbook of Industrial Drying*, 2nd edn, pp. 195–248. New York: Marcel Dekker.

Ochoa-Martinez LA, Brennan JG and Niranjan K (1993) Spouted bed dryer for liquid foods. *Food Control* 4: 41–45.

Pallai E, Szentmarjay T and Mujumdar AS (1995) Spouted bed drying. In: Mujumdar AS (ed.) *Handbook of Industrial Drying*, 2nd edn, pp. 425–488. New York: Marcel Dekker.

Varnalis A, Brennan JG and MacDougall DB (1998) Puffing of foodstuffs – a review. *European Food and Drink Review* winter 23–26, 28–31.

Spray Drying

J G Brennan, University of Reading, Reading, UK

Principles

This is the most commonly used method for drying food liquids and some slurries. The feed is converted into a fine mist or spray which is then mixed with heated air. Very rapid drying takes place, converting the liquid droplets into powder particles. Because of the small size of the particles – diameters usually in the range 10–200 μm – a very large surface area is available for drying. Also the distance that moisture has to migrate within the particles to the drying surface is relatively small. Thus, short drying times, 1–20 s, are a feature of this method of drying. There is also a significant evaporative cooling effect so that the surface temperature of the droplets does not rise much above the wet bulb temperature of the drying air until drying nears completion. Provided the particles, once dry, are swiftly removed from the drying chamber, heat damage to the product can be limited.

The main features of a single-stage spray-drying plant are shown in **Figure 1**. Inlet fan A draws air in through a filter B and pushes it through the air heater C into the drying chamber D. By means of pump E the feed material is pumped from feed tank F to the spray-forming device G. It is converted into spray which mixes with the drying air in the drying chamber D. The bulk of the dry powder thus formed is removed from the drying chamber via valve H and is carried pneumatically to a storage bin through duct I. The air is removed from the drying chamber via ducting J through one or more air/powder separators K and is exhausted to the atmosphere by outlet fan L. The fine powder recovered by the separators K may be added to the main product stream via valve M or recycled into the wet zone of the drying chamber via duct N.

Spray-drying Equipment

Air heating may be indirect using steam, fuel oil, or gas. Direct heating would obviously improve the thermal efficiency of a drier and in recent years natural gas has been used directly to heat the incoming air. There is some concern over the possible contamination of the dried products with nitrate and nitrite compounds and *N*-nitrosodimethylamine. The introduction of low NO_x (nitrogen oxide) burners has reduced this problem. However, work is still in progress to establish the level of contamination, if any, of food products dried by directly heated air.

The formation of a spray of the feed liquid within a specified size range is an essential feature of spray drying. Too wide a droplet size range can lead to nonuniform drying and/or build-up of powder on the inner wall of the chamber. Droplet size can influence important properties of the dried product such as its bulk density and reconstitution characteristics.

Spray-forming devices, also known as atomizers, are of three general types: (1) centrifugal or wheel atomizers; (2) pressure nozzles or jets; and (3) two-fluid or pneumatic nozzles.

A centrifugal atomizer consists of a disk or bowl on the end of a rotating shaft. The liquid feed is introduced on to the disk near its center of rotation. It accelerates to the peripheral velocity of the disk and is spun off, in the form of a spray, into the drying chamber. There are numerous different designs of spinning disk, some of which are depicted in Figure 2.

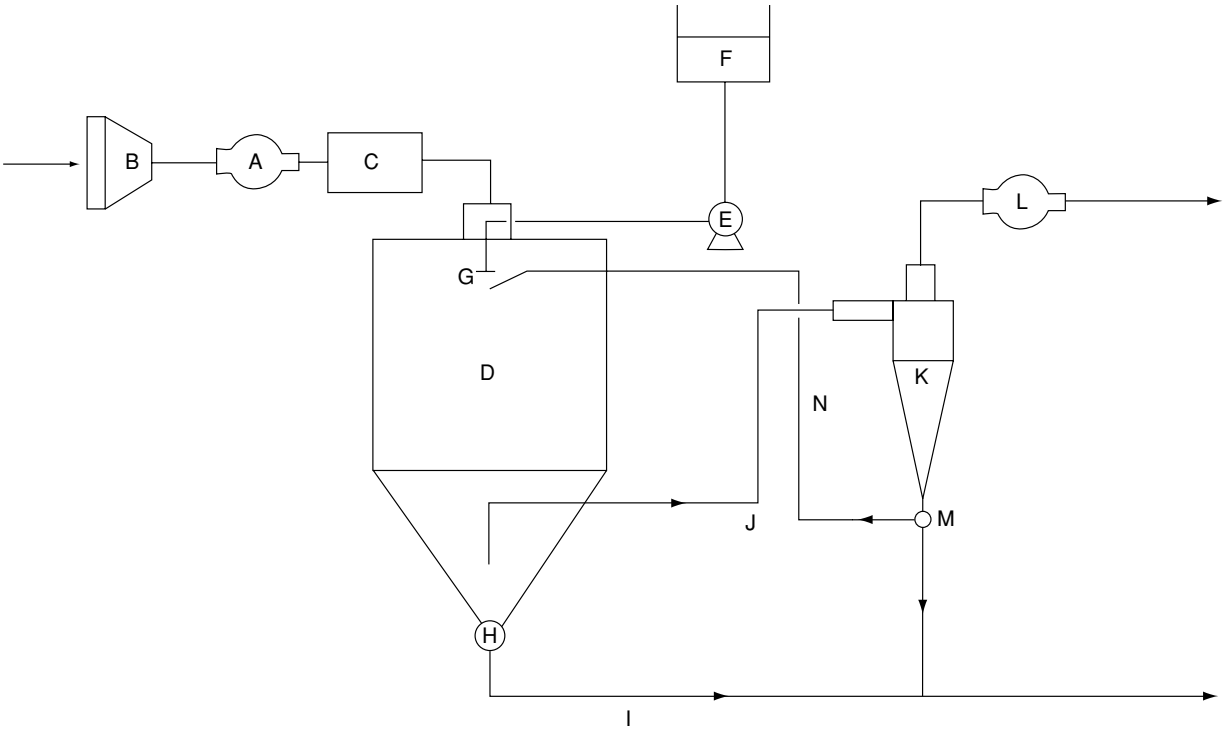

Figure 1 A typical single-stage spray drier, featuring recycling of fines. A, air inlet fan; B, filter; C, air heater; D, drying chamber; E, feed pump; F, feed tank; G, atomizer; H, rotary valve; I, product conveyor; J, duct for outlet air; K, cyclone, L, air outlet fan; M, rotary valve; N, duct for returning fines to chamber. From Brennan JG (1994) *Food Dehydration – A Dictionary and Guide*, Oxford: Butterworth-Heinemann with permission.

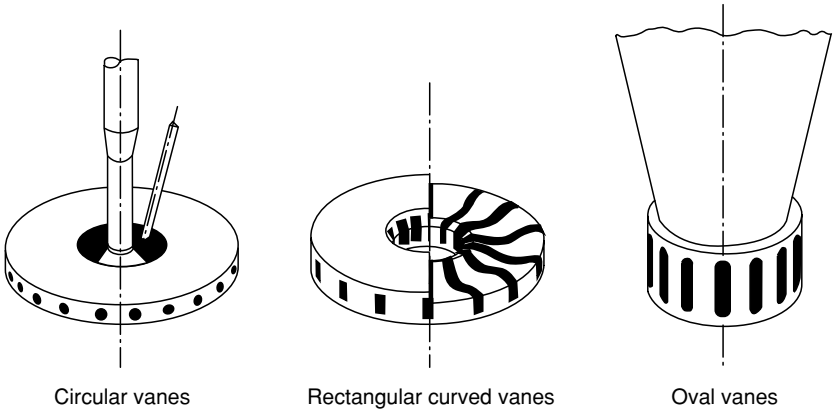

Figure 2 Different designs of centrifugal atomizers. Adapted from Mujumdar AS (1995) *Handbook of Industrial Drying*, 2nd edn. New York: Marcel Dekker with permission.

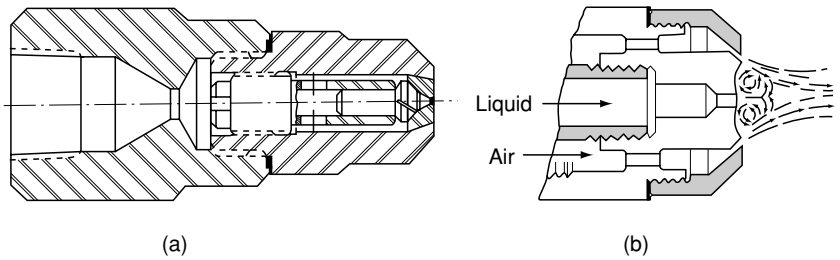

Figure 3 (a) Pressure nozzle; (b) two-fluid or pneumatic nozzle. Adapted from Brennan JG, Butters JR, Cowell ND, Lilly AEV (1990) *Food Engineering Operations*, 3rd edn. London: Elsevier Applied Science with permission.

Disk diameters range from 50 to 300 mm, they rotate at speeds of 50 000–10 000 revolutions per minute, respectively, and produce droplets with mean diameters in the range 1–600 μm. Centrifugal atomizers can produce uniform sprays within a wide range of droplet sizes. The size distribution of the droplets can be altered by changing the speed of rotation. Such atomizers are not subject to blocking or abrasion due to insoluble solid particles in the feed and can handle viscous materials at low pumping pressures.

A pressure nozzle (Figure 3a) features a small orifice, with a diameter in the range 0.4–4 mm, through which the feed is pumped at high pressure, in the range 5–50 MN m^{-2}. A grooved insert, located behind the orifice, imparts a spinning motion to the liquid, producing a hollow cone of spray. Pressure nozzles can produce quite uniform droplets with mean diameters ranging from 10 to 800 μm, which usually results in dried particles with hollow centers. The orifices in such nozzles may be abraded or blocked by insoluble solid particles in the feed. Very high pumping pressures are required to atomize viscous liquids.

In a two-fluid or pneumatic nozzle (Figure 3b) the energy in a high-velocity stream of gas is used to atomize the feed. The gas exits from the nozzle through an annular opening while the liquid feed exits through a circular opening concentric with the annular one. A Venturi effect is created and the liquid is converted into a spray. Such nozzles are subject to blocking and abrasion by solid particles in the feed. The size of the droplets can be changed by varying the gas–liquid ratio. They have relatively small capacities and can produce a spray with a broad spectrum of droplet sizes, especially with high-viscosity feeds.

A recent development involves the use of sonic energy to atomize highly viscous materials. A sonic resonance cup is located in front of a two-fluid nozzle. This creates a field of high-frequency sound waves that breaks up the liquid into droplets. This system is not yet being used commercially.

In the drying chamber of a spray drier the spray of feed contacts the heated air and drying takes place. There must be sufficient dwell time in the chamber for the particles to reach a suitably low moisture content but, once that is attained, the dry particles must be removed quickly from the chamber to avoid unnecessary heat damage. There are many designs of drying chamber available. These may be classified into three groups based on the direction of flow of the feed relative to the air, i.e., concurrent, countercurrent, and mixed flow. Countercurrent chambers are seldom used for food applications because of the risk of heat damaging the product. Concurrent chambers usually consist of a cylindrical body with a conical base. When drying large droplets of a heat-sensitive feed, in particular if the dried particles are tacky and tend to adhere to the wall of the chamber, a tall cylindrical body is used (Figure 4a). The feed is introduced at the top of the chamber, usually through a nozzle atomizer, and follows a fairly straight-line flow path to the bottom of the chamber. The wall of the conical section of the chamber may be cooled to facilitate removal of the powder.

Other concurrent chambers feature a shorter cylindrical body (Figure 4b). The air enters at the top of the chamber tangentially and follows a downward, spiral flow path. The liquid feed enters at the top of the chamber through a centrifugal atomizer. This design of drying chamber is widely used for drying foods which are not very heat-sensitive or tacky. A mixed-flow chamber features both concurrent and countercurrent flow patterns (Figure 4c). Such chambers find limited application in the food industry. The risk of heat damaging the product is greater than in concurrent chambers. Many other types of chamber are available, including flat-bottomed cylindrical and horizontal box-like designs.

In industrial-sized spray driers the bulk of the dried powder is removed directly from the chamber through a rotary valve or vibrating device. Some food powders, mainly those with high sugar or fat contents, become sticky when hot and adhere to the wall of the chamber. Various devices are available to assist in the removal of such products. Hammers may be made to strike the outside wall of the chamber at intervals

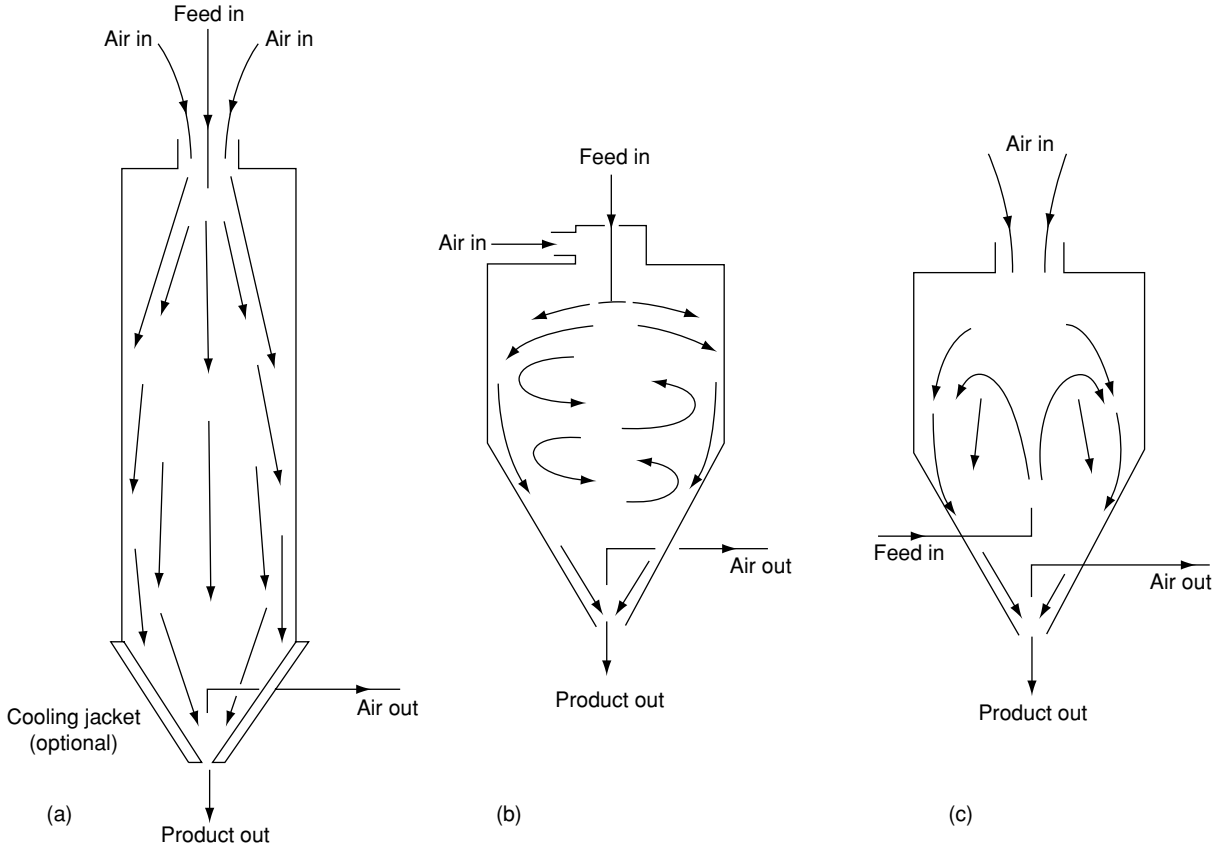

Figure 4 Different designs of spray-drying chambers: (a) concurrent with straight-line flow path; (b) concurrent with spiral flow path; (c) mixed flow.

to loosen deposits. Brushes, chains, or air brooms may sweep the inner surface of the chamber. Reducing the temperature of the chamber wall may help. This can be achieved by removing some of the insulation from the wall or by drawing cool air in through a jacket covering some or all of the chamber wall (Figure 4a). This warm air may then be introduced into the main air supply to the heater, thus recovering some of the heat lost from the chamber. Fine powder entrained in the exhaust air from the chamber may be recovered with the use of dry cyclone separators, filters, or wet scrubbing devices, singly or in various combinations. If some of the feed is used to wash the outgoing air in a scrubber and is then recycled as feed into the drier, savings in energy can result.

Semiclosed or closed spray-drying systems are available in which some or all of the drying gas is recycled. A scrubber-condenser is used to remove the water vapor from the gas. These systems may be used for drying materials with strong, unattractive odors or where gases other than air are used as the drying medium. Aseptic spray-drying systems are also available in which all the air entering the system is presterilized, usually by filtration, and the product is discharged into a sterile packing zone. Such systems are mainly used in the pharmaceutical industry but may find applications in food dehydration as hygienic requirements become more stringent.

Safety Considerations

There are fire and explosion hazards associated with spray drying of liquid foods. A dust explosion may occur when a fine combustible powder is dispersed in air and ignited. For an explosion to occur there must be a sufficient concentration of powder in the air and sufficient oxygen present to support combustion. Such conditions may exist in spray driers under normal operating conditions. Self-ignition may occur in food powders. Oxidative reactions are exothermic and the generation of heat by such reactions can be very fast at the temperatures encountered in spray drying. In thick layers of powder, heat generated in this way cannot escape quickly enough and the temperature rises rapidly, causing ignition. The critical thickness for ignition depends on the temperature and the powder characteristics. Obviously, to minimize the occurrence of self-ignition, the build-up

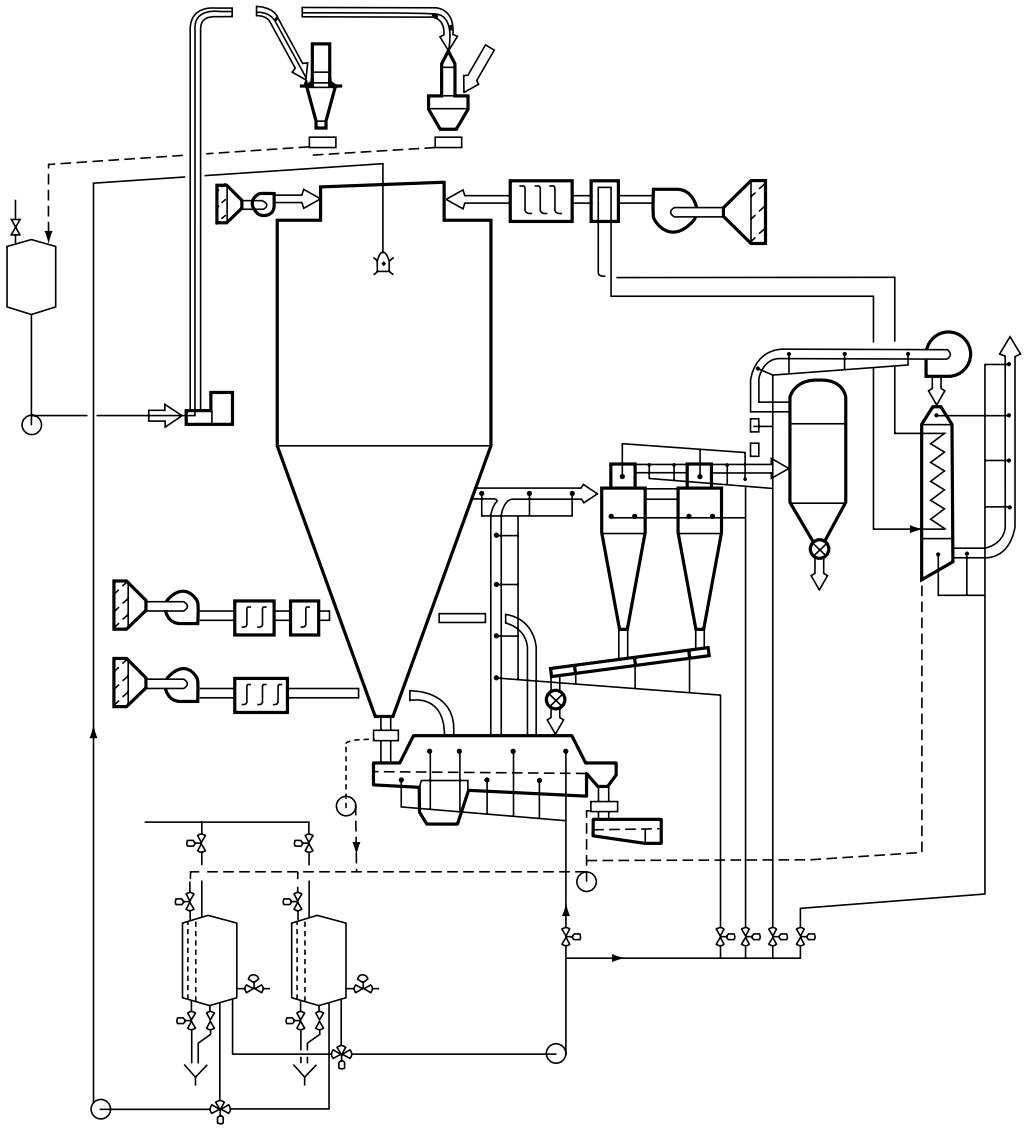

Figure 5 Cleaning-in-place (CIP) system for spray drier layout. From Masters K (1991) *Spray Drying Handbook*, 5th edn. New York: Longmans with permission.

of layers of powders in spray-drying systems should be avoided. Regular and efficient cleaning of the plant (see below) and the use of devices, such as those mentioned above, to remove powder deposits can reduce the hazard. There may be other causes of powder ignition such as heat generated by friction or electrical sparks. In spray driers featuring direct flame heating, very hot particles entering the drier in the heated air may ignite the powder. Good design and maintenance of equipment should minimize the chances of ignition occurring by these means. The fitting of weaker doors or panels, which are designed to fail at relatively low pressures, relieving the explosion, is the most common precaution taken to limit damage due to dust explosions in spray driers. Other measures such as containment or suppression of such explosions are not usually feasible in high-capacity driers.

Efficient cleaning of spray drier installations is essential. This may be done manually using high-pressure water jets. Rotating nozzle heads may be strategically located throughout the system so that all parts of the equipment are contacted by the cleaning fluid. Most modern plants are equipped for cleaning-in-place (CIP). Such a system is shown in Figure 5.

Energy Considerations

Spray drying is an energy-intensive method of removing moisture. In a straight-through spray drier, up to

6000 kJ of energy may be required to evaporate 1 kg of water. By comparison, in a six-stage multiple-effect evaporator with mechanical recompression, as little as 200 kJ may do the same task. Therefore, to conserve energy, liquid foods should be concentrated to as high a solids content as possible prior to spray drying. The upper limit of solids concentration is usually determined by the viscosity of the concentrate which, if too high, may result in poor atomization. The use of direct flame heating of the air to the spray drier is more energy-efficient than indirect methods (see above). The higher the air inlet temperature and the lower the air outlet temperature, the more thermally efficient the drying operation. The upper limit on air inlet temperature is usually determined by the heat sensitivity of the feed. For many foods 200 °C is this upper limit. Some can tolerate 250 °C. It is now possible to go as high as 300 °C for some foods when using modern, multistage spray-drying systems (see below). A low outlet temperature may lead to too high a moisture content in the product. Few driers operate well below 80 °C. Up to 50% of the exhaust air from the drying chamber may be recycled under certain circumstances and can lead to up to a 20% saving in energy. However, this must be balanced against a reduction in evaporative capacity and is really only feasible when high air outlet temperatures are used. Thus, it is not suitable if the product is heat-sensitive. Recovery of heat from the exhaust air leaving the drying chamber offers another opportunity to conserve energy. This may be achieved using air/air heat exchangers so that the incoming air is prewarmed with heat recovered from the exhaust air. There have been considerable improvements in the design of such heat exchangers in recent years. Alternatively, heat recovered from the exhaust air may be used to prewarm the liquid feed to the drier. This can be achieved by the use of wet scrubbers to wash the fines from the exhaust air (see above).

Finishing Operations

In the case of some powders, better control over product quality may be achieved by removing it from the main spray-drying chamber at a moisture content higher than the desired final value and finishing the drying in another type of drier. Vibrating fluidized-bed driers are most commonly used for this purpose. In this secondary drier some agglomeration of fine powder particles may occur, reducing the amount of very fine particles in the end product. If the fine powder recovered from the cyclones is recycled into the wet zone of the drying chamber, this will further reduce the amount of very fine particles in the product (Figure 1). Such a reduction usually facilitates the handling of the dry powder and improves its reconstitution characteristics. The use of such secondary driers can also lead to savings in energy as compared with single-stage drying. The powder leaving the secondary drier may be cooled in another vibrating fluidized bed.

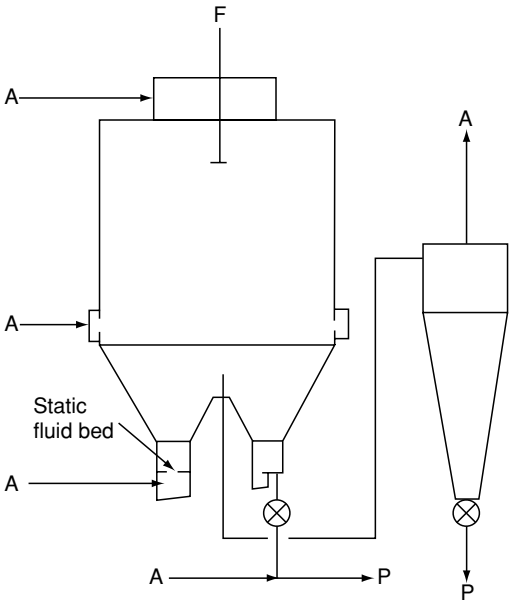

Figure 6 A two-stage layout with integrated static fluid bed in drying chamber: A, air flow; F, feed; P, dried product. From Masters K (1991) *Spray Drying Handbook*, 5th edn. New York: Longmans with permission.

Spray driers are also available which permit multistage drying in one unit. One design features a built-in fluidized bed (Figure 6). This bed is in the shape of a ring at the bottom of the chamber. Heated air is introduced at the top of the chamber and also at the bottom, fluidizing the particles in the annular bed. Drying takes place partly in the main chamber and partly in the fluidized bed. A further supply of air is introduced tangentially into the chamber. This sweeps the wall of the chamber, reducing powder build-up and improving air/powder separation. In the Filtermat drier (Figure 7) part of the drying takes place in the main chamber and the semidried powder falls on to a perforated, moving belt where it is dried to the required moisture content in two further stages, with a short holding period between, by a through flow of warm air.

Spray-dried products containing very fine particles may be difficult to reconstitute. When added to hot coffee, for example, dried milk powder may form clumps which float on the surface of the liquid and require vigorous and prolonged stirring to disperse. The most common solution to this problem is to

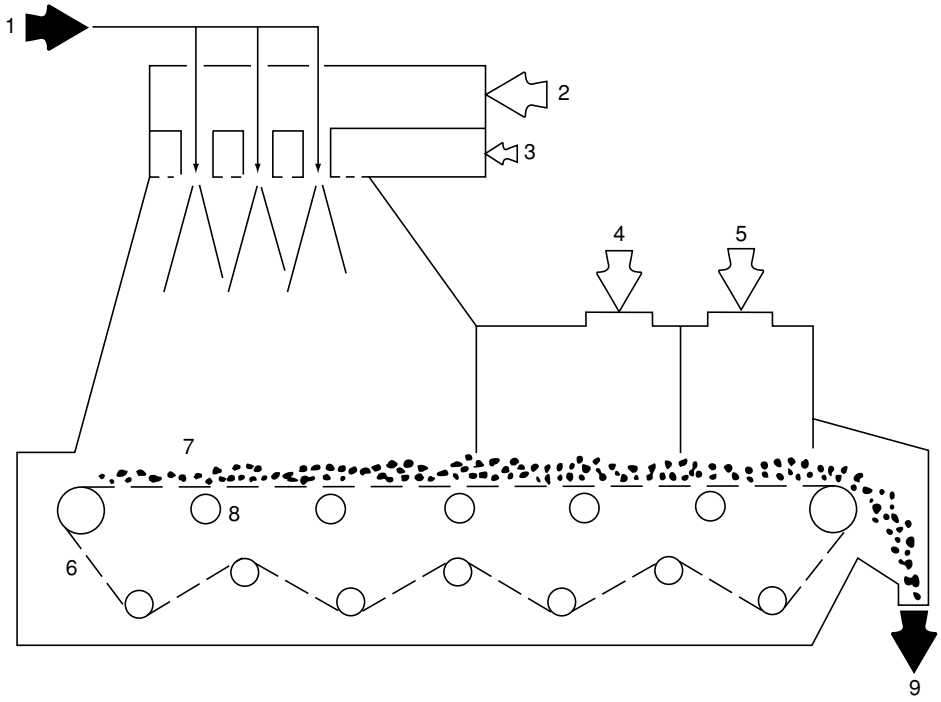

Figure 7 Filtermat spray dryer (DEC): 1, feed to nozzles; 2, primary drying air; 3, ceiling cooling air; 4, secondary drying air; 5, cooling air; 6, textile mesh belt conveyor; 7, powder layer; 8, exhaust air to cyclone; 9, product outlet. From Mujumdar AS (1995) *Handbook of Industrial Drying*, 2nd edn. New York: Marcel Dekker with permission.

increase the size of the powder particles. This may be achieved by adjusting the drying conditions, by recycling of fines, or by agglomeration in a secondary drier (see above). When these measures are inadequate, instantizing may be effected by rewetting the surface of the particles with steam, warm humid air, or a fine mist of liquid, promoting interparticle collisions by swirling the wetted powder in a vortex and redrying the agglomerated powder, often in a fluidized-bed drier.

Control of Spray Driers

Most modern spray driers are equipped with automatic control systems. The main parameter which is controlled is the outlet air temperature. In the case of driers equipped with rotary atomizers, the outlet air temperature is controlled by adjusting the rate at which the feed enters the chamber. The air inlet temperature is separately controlled by adjusting the fuel supply. When nozzle atomizers are in use, the feed rate needs to be constant, and the outlet air temperature is controlled by adjusting the combustion rate of the fuel. In both cases, an additional safety device must be fitted to avoid a rapid rise in outlet temperature should there be an unexpected drop in the feed rate. In addition to the above controls, the starting-up, shutting-down, and cleaning of the plant may be controlled by timing devices.

Spray-Dried Food Products

A very wide range of spray-dried food products is produced worldwide. A summary of typical operating conditions and some properties of a selection of dried powders is presented in Table 1.

Dairy Products

Skim milk is spray dried on a very large scale: some plants handle more than 1 million liters of milk per day. It is not a difficult material to spray dry with no tendency to adhere to the chamber wall and not being very hygroscopic. Many designs of drying chamber are used, featuring both centrifugal and nozzle atomizers. Skim milk powder produced in a simple single-stage drier contains a lot of fine particles, is dusty, and is difficult to handle and reconstitute. These properties can be improved by recycling fines, promoting agglomeration in secondary driers or integral fluidized beds or by rewetting.

Whole milk is rather more difficult to dry. It tends to be rather sticky when hot and can build up a deposit on the chamber wall. Hammers or other devices (see above) to assist in the removal of such deposits can be used. Handling and reconstitution characteristics can be improved by agglomeration. In addition, the dried particles may be coated with

Table 1 Typical drying conditions and properties of some spray-dried foods

Product	*Total solid content of feed (%)*[a]	*Air inlet temperature (°C)*	*Type of atomizer*	*Type of chamber*	*Air outlet temperature (°C)*	*Product moisture content (%)*[a]	*Particle size of powder (μm)*	*Bulk density of powder (kg m^{-3})*	*Notes*
Skim milk (ordinary)	40–50	180–230	Various	Concurrent	85–100	3.5–4.0	30–50	600–700	
Skim milk (agglomerated[b])	40–50	180–230	Various	Concurrent	85–90	3.5–4.0	150–200	450–600	Bulk density may be as low as 250 kg m^{-3} if rewetted
Whole milk (ordinary)	40–50	180–200	Various	Concurrent	75–95	2.5–3.0	30–50	550–650	
Whole milk (agglomerated)[b]	45–50	175–200	Various	Concurrent	90–100	2.5–3.0	150–200	450–550	
Whey (ordinary)	45–50	180–200	Centrifugal	Concurrent	90–100	3.0–4.0	30–40	600–700	
Whey (precrystallized and agglomerated)	40–60	180–190	Centrifugal	Concurrent	80–90	2.0–3.0	100–150	550–650	Particle size may be > 1000 μm and bulk density < 500 kg m^{-3} if a second crystallization stage follows spray drying
Buttermilk	45–50	175–190	Various	Concurrent	75–90	3.5–4.5	30–50	750–830	
Eggs (yolk and whole)	25–48	150–200	Centrifugal	Concurrent	50–80	2.0–3.0		450–550	
Egg (white)	10–12	145–200	Centrifugal	Concurrent	50–80	2.0–3.0			
Instant coffee	35–40	250–300	Nozzle	Concurrent	110–115	2.0–3.0	150–300	180–300	
Instant tea	35–40	200–250	Nozzle	Tall chamber	90–100	2.0–2.5			
Tomatoes	25–40	140–150	Centrifugal	Concurrent with cooled wall	70–80	3.0–3.5			

[a]Wet weight basis.
[b]Agglomeration in fluidized bed.

lecithin to improve their wettability when reconstituted in water or aqueous solutions.

Whey is also spray-dried. A number of drier designs, usually featuring nozzle atomizers, may be used. To reduce the hygroscopicity and improve the handling characteristics of whey powder, the lactose is crystallized into the stable α-monohydrate form, prior to drying.

Buttermilk is also spray-dried. Driers similar to those used for skim milk are used.

Icecream mixes have high sugar contents, up to 30%, which cause problems during drying. One solution is to keep back some of the sugar and add it to the dried mix. Two-stage drying yields a good noncaking powder. (*See* **Ice Cream:** Methods of Manufacture.)

Cheese is ground and mixed with water prior to drying. Some deposition on chamber walls may occur. This can be overcome using devices to assist in powder removal (see above) and/or two-stage drying. Strong odors can escape to the atmosphere during cheese drying. The exhaust air may need to be deodorized or a semiclosed drying system used (see above).

Butter and high-fat powder (60–80% fat) may be spray-dried. Precautions must be taken to avoid breakdown of the emulsion structure and the product must be cooled after drying.

Sodium caseinate after manufacture is relatively viscous and is fed to the drier at a lower than usual solids content (20–25%). Centrifugal atomizers and two-stage drying are usually used.

Eggs

Whole egg, egg yolk and egg white may be spray-dried. The feeds are pasteurized prior to drying. Glucose, if present in the dry powder, can give rise to the development of discoloration and off-odor during storage. Glucose may be removed, prior to drying, by fermentation using bacteria, yeasts, or enzymes. This process is mainly applied to egg white. Whole egg and egg yolk are fed to the drier at 25–27% and 45–48% solids contents, respectively. Centrifugal atomizers are usually used. Egg white feed usually contains 10–12% solids, but this may be increased by reverse osmosis.

Beverages

Instant coffee is another product spray-dried on a large scale. Ground, roasted coffee beans are extracted with hot water in semicontinuous or continuous extractors. The extract contains 20–25% solids which may be fed directly to the spray drier or preconcentrated up to 60% solids. Driers featuring tall cylindrical bodies (Figure 4a) are used, and nozzle atomizers. Air and spray follow a streamlined path down the tower. The conical base of the drier may be jacketed and cool air passed through this to keep the wall temperature low. Fines may be recycled to the drying chamber. Coffee powder may be agglomerated by rewetting. Instant coffee may also be freeze-dried. Coffee substitute made from mixtures of chicory, barley, rye, wheat, and coffee may be dried in a similar manner to the pure coffee. (*See* **Coffee:** Instant.)

Instant tea may also be produced by spray drying. Dried leaves are extracted with hot water to give an extract containing 3–5% solids. This extract is concentrated by vacuum evaporator. The volatile aroma compounds which flash off in the early stages of evaporation are condensed and added back to the concentrate prior to drying. Driers similar to those used for coffee extract are used but at lower inlet air temperatures. (*See* **Tea:** Types, Production, and Trade.)

Fruits and Vegetables

Whole fruits and fruit juices may be spray-dried. However, with the exception of tomatoes, additives must be used to facilitate drying.

Tomatoes are prepared for drying by the hot-break or cold-break processes. The former method is favored as the reconstituted powder made in this way has less of a tendency to settle than that made by the cold-break process. The pulped tomatoes are comminuted and sometimes homogenized and then concentrated to 26–48% solids by vacuum evaporation. The drier most commonly used has a relatively short cylindrical body and conical base. The whole of the drying chamber may be jacketed and cool air drawn through the jacket to cool the chamber wall. A centrifugal atomizer is used. The powder leaving the chamber is cooled on a vibrating fluidized bed.

Citrus and other fruits such as bananas, apples, pears, strawberries, raspberries, peaches, and apricots need to have fillers added to permit spray drying. Without such additives the wall of the drying chamber will become coated with sticky powder and the product will be tacky and hygroscopic, with a strong tendency to cake. The most common additive used is liquid glucose with a dextrose equivalent in the range 15–20. This may be added in amounts up to 75% of the fruit, on a solids basis. Skim milk powder has also been used as an additive. To insure good atomization the fruit must be comminuted thoroughly. Both centrifugal and nozzle atomizers and various designs of drying chamber, some featuring cooled walls and aftercoolers, have been used.

Whole vegetables and vegetable juices may also be spray-dried. These include carrots, potatoes, cabbage, peas, and asparagus. These are easier to dry than fruits. Again, efficient comminution is necessary and additives may be required.

Other Spray-dried Products

Beverage whiteners These are oil-in-water emulsions whose ingredients include vegetable fat, sodium caseinate, corn syrup, and emulsifying agents. They are spray-dried, usually using pressure nozzles, down to a moisture content of 2–5%. Often two-stage drying is used.

Vegetable proteins These may be extracted from soya beans, peanuts, potatoes, and other sources. They are spray-dried using concurrent chambers and pressure jet atomizers. Some dried products may be tacky and hygroscopic and require cooling during handling and packaging.

Single-cell protein (SCP) SCP is produced from methane, methanol, ethanol, and η-paraffins by fermentation, concentrated by centrifugation and spray-dried. The fine powder is often agglomerated. Relatively high inlet and outlet temperatures are used. There is a danger of dust explosion and semiclosed, self-inertizing spray-drying systems may be used.

Meat Purées of beef and chicken may be spray-dried for use in soups and sauces. Concurrent drying chambers and centrifugal atomizers are usually used. Blood from slaughterhouses is spray-dried to produce a range of blood, albumin, and plasma powders. Relatively low drying temperatures are used.

Fish byproducts A range of products, including soluble fish powder, fish meal, fish flour, and fish protein, produced by hydrolysis of fish flesh, is produced by spray drying. The odor released during drying can be a problem.

Enzymes A number of enzyme powders are produced by spray drying. To minimize inactivation, inorganic salts may be added and relatively low drying temperatures used. In some cases aseptic spray drying may be carried out.

Vitamins Vitamins from various sources may be spray-dried. Vitamins B and C are spray-dried directly. Vitamins A and D are usually microencapsulated (*See* **Vitamins**: Determination.)

See also: **Agglomeration**; **Ascorbic Acid**: Properties and Determination; Physiology; **Cholecalciferol**: Properties and Determination; Physiology; **Coffee**: Instant; **Drying**: Theory of Air-drying; Fluidized-bed Drying; Physical and Structural Changes; Chemical Changes; **Eggs**: Use in the Food Industry; **Fish**: Miscellaneous Fish Products; **Fish Meal**; **Ice Cream**: Methods of Manufacture; **Instrumentation and Process Control**; **Powdered Milk**: Milk Powders in the Marketplace; Characteristics of Milk Powders; **Retinol**: Properties and Determination; Physiology; **Single-cell Protein**: Algae; Yeasts and Bacteria; **Soy (Soya) Beans**: Processing for the Food Industry; **Tea**: Types, Production, and Trade; Processing; **Vitamins**: Determination; **Whey and Whey Powders**: Production and Uses

Further Reading

Barbosa-Canovas GV and Vega-Mercado H (1996) *Dehydration of Foods*. New York: Chapman and Hall.

Brennan JG (1994) *Food Dehydration – A Dictionary and Guide*. Oxford: Butterworth-Heinemann.

Brennan JG, Butters JR, Cowell ND and Lilly AEV (1990) *Food Engineering Operations*, 3rd edn. London: Elsevier Applied Science.

Masters K (1991) *Spray Drying Handbook*, 5th edn. New York: Longmans.

Mujumdar AS (ed.) (1995) *Handbook of Industrial Drying*, 2nd edn. New York: Marcel Dekker.

Van Arsdel WB, Copley MJ and Morgan AI Jr. (1973) *Food Dehydration, Drying Methods and Phenomena*, vol. 1 *Principles*, vol. 2 *Practices and Applications*. Westport, CT: AVI.

Dielectric and Osmotic Drying

J G Brennan, The University of Reading, Reading, UK

Application of Dielectric Energy to Drying

Principles

Here the term dielectric heating is used to represent the radiofrequency (RF) and microwave (MW) bands of the electromagnetic spectrum. RF heating is in the frequency range 1–200 MHz and MW from 300 MHz to 300 GHz. By international agreement, certain frequencies have been allocated for industrial use, in order to avoid interference with telecommunications. These are RF 13.56 and 27.12 MHz and MW 2450 and a band within the range 896–915 MHz. The advantages of dielectric heat for drying, cooking, thawing, and melting of foods are that heat generation is rapid and occurs throughout the body of the food material. An additional advantage for drying is that water is heated more quickly than the other components of the food. In conventional methods of drying, heat is transferred to the surface of the food by convection, conduction, or radiation. It has then to transfer throughout the body of the food by conduction. Some penetration does occur in the case of radiant heat but this is limited and often difficult to

control. The penetration of electromagnetic waves in the dielectric range into materials such as foods depends on frequency and the characteristics of the material. The lower the frequency, hence the longer the wavelength, the greater the penetration. Penetration may be quantified by half-power depth, i.e., the thickness of the material, which reduces the wave energy to 50% of the incident level. The half-power depth for water is about 12 mm at a frequency of 2450 MHz and 75 mm at 100 MHz. Penetration into foods is less than into water.

In the case of dielectric heating, the heat is transferred into the body of the material rather than just to the surface. When a wet material is exposed to high-frequency electromagnetic radiation molecules, which are dipolar, such as water, are stressed by the alternating magnetic field and this results in the generation of heat. This is the main mechanism which causes heating on the application of MW energy. At the lower RF frequencies, ionic conduction plays a part. Ions present in the material are accelerated by the electric fields, and this leads to the generation of heat. Since most foods contain ions in the aqueous phase, RF heating can be effectively applied to them. The mechanism of drying using dielectric heating is different from that which operates when convected or conducted heat is used. The rapid generation of heat within the material causes rapid evaporation of water. This creates a total pressure gradient, which results in the rapid movement of liquid water and water vapor to the surface. Thus, shorter drying times and lower product temperatures are attainable compared to more conventional methods of heating.

Overheating of the surface can be avoided. Drying is more uniform as thermal and concentration gradients are less. There is also less movement of solutes within the material. Since the water absorbs most of the heat there is more efficient use of energy compared to drying by the application of convected or conducted heat. However, too high a heating rate can lead to scorching or burning of the material. If the water vapor becomes entrapped within the material, high pressures can develop, which can lead to the rupture of solid pieces.

Equipment

A basic RF (platen) applicator consists of two metal plates between which the food is located or conveyed. Often the conveyor belt may represent one of the electrodes. The plates are at different electrical voltages. The whole system is housed in a metal chamber to prevent leakage of radiation. In a stray field applicator a thin layer of material passes over electrodes of alternating polarity. In the staggered applicator rods extend beyond the plates to form staggered throughfield arrays between which the layer of material passes. This last type can cope with thicker layers Figure 1.

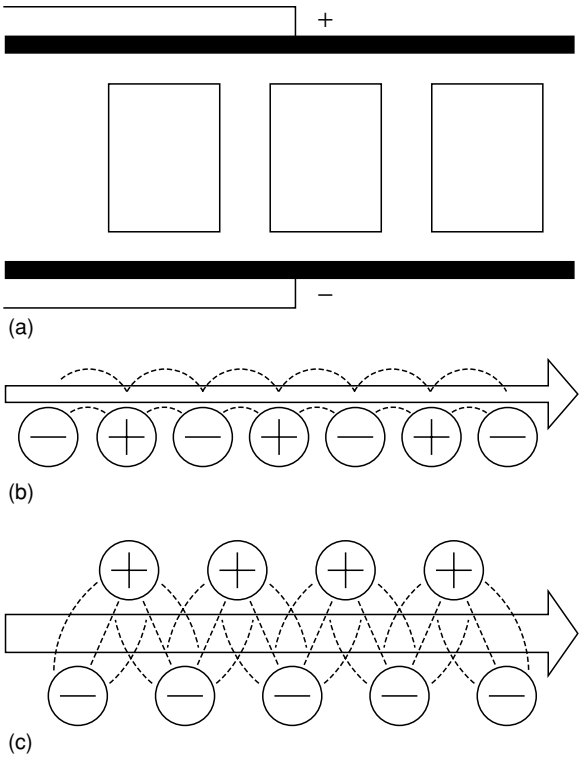

Figure 1 Electrode configurations for dielectric heating sysÙtems. (a) Platen type for bulky objects. (b) Stray field type for thin layers. (c) Staggered type for thicker layers. Adapted from Schiffmann RF (1995) Microwave and dielectric drying. In: Mujumdar AS (ed.) *Handbook of Industrial Drying*, 2nd edn, vol. 1, pp. 345–372. New York: Marcel Dekker.

In a basic batch MW applicator, MWs, generated by a magnetron or klystron, are directed into a metal chamber via a waveguide or coaxial cable. The food is placed in the chamber. To improve the uniformity of heating the beam of MWs may be disturbed by a mode stirrer, resembling a slowly turning fan, which causes reflective scattering of the waves. Alternatively, the food may be rotated in the chamber. In continuous systems, the food is conveyed through a chamber, which has a number of wave injection ports in its sides, top, and bottom.

Applications

In drying of foods, dielectric heating is usually combined with conventional heating. It may be used to preheat the feed to a hot-air drier. By raising the temperature of the feed quickly and causing moisture to move to the surface, it can decrease the overall drying time. Dielectric heating may be applied part-way through the drying cycle, when the food enters the falling rate period. (*See* **Drying**: Theory of Air-drying.)

This can boost the rate of drying. If dielectric heating is applied near the end of hot-air drying it can also shorten the drying time significantly and hence increase the throughout of the drier. It is more usual to use dielectric heating in the later stages of drying.

One of the major applications of RF heating is in the postbaking of biscuits. The objectives in baking biscuits are to produce a product of the right size, shape, color, and moisture content. In a conventional oven, reducing the moisture content to the desired level can take up a large part of the total baking time. The application of RF heating can shorten the baking time. The oven is set to produce biscuits of the right size, shape, and color, but the RF heating is used to remove the remaining moisture, without any further change in these properties of the biscuit. The capacity of an oven can be increased by more than 50% by the use of RF heating. Postbaking by RF heating has also been applied to breakfast cereals and cereal-based baby foods.

MW heating is used, in combination with heated air of high humidity, to dry pasta products. The humid air prevents cracking of the product. The drying time is reduced from 8 h, using heated air only, to 1 h. MW heating is used in drying onions. The moisture content is reduced from 80% to 10% by means of heated air and then down to 5% using MWs. This combined process results in more uniform drying and a saving in energy of up to 30%, compared to hot-air drying. Potato chips (French fries) and crisps are finish-dried after frying to prevent darkening, if the sugar content is high.

MW heating has been used in vacuum drying. (*See* **Drying**: Equipment Used in Drying Foods.) Pasta is dried under vacuum using MW heating. This results in much shorter drying times, compared to the use of conducted or radiant heat. There are reports of other dried products being produced in this way, including fruit juice, tea and enzyme powders, mushrooms, asparagus, and soya beans.

MW heating in freeze drying (*See* **Freeze-drying:** The Basic Process; Structural and Flavor (Flavour) Changes), has been widely studied. The loss factors of ice and liquid water are much higher than that of dry tissue. When a frozen food is exposed to MWs, the ice will absorb energy much faster than the solid. This should be an advantage in freeze drying. However, ionization of the rarefied gases can occur because of the very low pressures used. This can result in plasma discharge and overheating of the food. The use of higher-frequency MWs, 2450 MHz, can reduce this problem. The loss factor of liquid water is much higher than that of ice. If some melting of the ice occurs, the water formed will absorb energy quickly and vaporize. This could cause solid food particles to explode because of the build-up of pressure internally. Good control of the freeze-drying process should prevent this happening. The use of MW heating in freeze drying is technically feasible. However, it has not been used on a large scale as yet.

Osmotic Drying

Principles

This term is applied to the removal of water from fresh foods, mostly fruits and vegetables, by immersing pieces of the food in a solution with a higher osmotic pressure, and hence a lower water activity, than the food. This solution is sometimes referred to in the literature as the hypertonic solution. Water will pass from the food into the solution under the influence of the osmotic pressure gradient. In this process, the walls of the cells in the food act as semipermeable membranes. However, the membranes are not completely selective. Some soluble natural substances, such as organic acids, sugars, salts, and vitamins, may be lost with the water, while solutes from the solution may penetrate into the food. This gain of solutes by the food can contribute to the reduction in its water activity, but may affect the taste of the product.

Osmotic Agents

The solutes used to make up the hypertonic solution must be highly osmotically active, nontoxic, edible, with an acceptable taste and flavor. In the case of fruits, sugars are the solutes used to make up the hypertonic solution. Sucrose is the most widely used sugar, but glucose, fructose, glucose/fructose and glucose/polysaccharide mixtures, high-fructose syrups, and lactose have been used experimentally, with varying degrees of success. The addition of small amounts of sodium chloride (0.5–2.0%) to the sugar solution can enhance the rate of water loss. Other low-molecular-weight substances, such as malic acid, lactic acid, and even hydrochloric acid, can have a similar effect. There is some evidence that the molecular size of the osmotic solute can affect its penetration into the food. The smaller the size of the molecule, the greater the depth and extent of the penetration. For vegetables, sodium chloride is the solute most widely used. Glycerol and starch syrup have also been investigated.

Processing Conditions

In osmotic drying, the rate of water loss is high initially but reduces significantly after 60–120 min. However, it can take days for the process to reach equilibrium. A typical processing time to reduce the weight of the food by 50% is 4–6 h. Some researchers

have reported that, while water loss falls off rapidly in the first 2 h, the gain of solutes from the hypertonic solution continues for longer. The initial concentration of solute in the solution can affect the rate of water loss. Sugar concentrations in the range 40–70% are used. In general, the higher the initial concentration of the sugar, the greater the rate and extent of drying. Sometimes, the upper limit in concentration may be dictated by the gain of solids, which can affect the taste of the product, i.e., making it too sweet. In the case of vegetables, sodium chloride concentrations in the range 5–20% have been used. Again, at the high concentrations, the taste of the product may be adversely affected.

The temperature at which the osmosis takes place can affect the rate of water loss. Generally, increasing the temperature will increase this rate. However, there will be an upper limit on temperature above which the food may be adversely affected. Cell walls may be damaged resulting in excess loss of soluble materials, such as vitamins, from the cells. Discoloration of food may also occur. This upper limit on temperature will be different for different foods. Temperatures above 70 °C are seldom used, and for some heat-sensitive fruits, osmosis may be carried out at room temperature, i.e., c. 20 °C.

The size and shape of the food pieces can also affect the osmotic process. The smaller the pieces, the greater the surface area through which water loss can occur and hence the higher the rate of water loss. However, the smaller the pieces, the more cell damage will occur during slicing. This can result in excessive loss of soluble matter. The shape and size of the pieces may be dictated by the form of the end product, e.g., some fruits such as plums may be processed whole or in halves. Apples and peaches may be in slices or segments. Vegetables may be in the form of slices or cubes. The literature suggests that the upper limit on thickness is 10 mm.

Movement of the osmotic solution over the surface of the food pieces would be expected to increase the rate of water loss. However, vigorous mixing of the pieces is seldom desirable because it causes damage to them. The food pieces may be immersed motionless in the solution. To increase the rate, the solution may be recirculated through the tank by means of a pump. The pieces may be suspended in a vibrating basket in the solution. Another option is to pump the solution through a bed of the pieces contained within a tall vessel.

The weight ratio of food to solution can affect the process, in terms of rate of water loss and solute gain. In general, the lower this ratio, the higher the water loss and the solute gain. Ratios of 1:4 to 1:5 are commonly used.

Blanching, either in water or osmotic solution, will prevent browning. In the case of large fruit pieces, blanching can sometimes shorten the osmotic process, due to relaxation of the structural bonds in the fruit. For some small fruit pieces and for vegetable pieces, blanching can adversely affect the osmotic process, reducing water loss and increasing solids gain. However, this effect is different for different foods.

Recycling of Hypertonic Solution

Following the osmotic process, the isotonic solution will be diluted with the water removed from the food. In order to reuse this solution, its concentration has to be increased to that of the original solution. This can be done by adding more solute to the diluted solution. However, this will result in a surplus of solution which has to be used for some other purpose. An alternative approach is to reconcentrate the solution by vacuum evaporation. There will be a limit to the number of times the solution can be recycled as some darkening can occur. It is usually necessary to filter the solution free of insoluble solids before evaporation. It is necessary to monitor the microbiological quality of the solution which is being recycled. There can be a build-up of microorganisms, mainly yeasts, in sugar solutions, if the processing temperature is low. Pasteurization of the solution can overcome this problem.

Since there is no change of phase during osmotic drying, the energy requirements should be less that in conventional drying methods. Heat is required to dissolve the solute in preparing the hypertonic solution, to maintain the temperature of the solution and to evaporate water from the solution, if it is being recycled. Additional energy is necessary to mix and recirculate the solution. It has been reported that the energy consumption during osmotic drying of fruits is 100–2400 $kJ\,kg^{-1}$ of water removed, depending on the processing temperature and how the surplus solution is treated, compared to 5000 $kJ\,kg^{-1}$ when drying in heated air.

Further Processing of Osmotically Dried Foods

The amount of water removed in osmotic drying, typically about 50% of the fresh weight of the food, does not result in a shelf-stable product. The water activity attainable in fruits and vegetables is usually in the range 0.90–0.95. Consequently, osmosed products have to be processed further to obtain a product with an extended shelf-life.

Air drying Osmotically dried fruits and vegetables may be stabilized by further drying in heated air. The

rate of drying of such products is usually lower than that of fresh material and relatively low air temperatures may have to be used to avoid caramelization and case hardening. The sensory properties of such products are different to those produced by air-drying alone. They have a good color and flavor, with reduced acidity, and a soft chewy texture. Shrinkage is usually less than in air-dried products.

Vacuum drying Osmosed products may be stabilized by vacuum drying. The quality of such products can be superior to those completed by air drying but the process, which is known as osmovac, is more expensive.

Freeze drying Predrying by osmosis can shorten the drying time, and hence reduce, the cost of freeze drying. The end products are less brittle and less prone to browning than those freeze-dried without the osmotic step.

Freezing Freezing of fruits and vegetables is an energy-intensive operation. By partly drying food before it is frozen, freezing costs can be reduced. This process is known as dehydrofreezing. The partial drying prior to freezing can be achieved by osmosis. When thawed, such products can have a better texture, with less drip, than those frozen from fresh. Unlike conventionally frozen foods, they require rehydration after thawing.

Solar drying It has been reported that osmotic drying before solar drying can reduce the load on the solar driers and enhance product quality. Solar energy may be used to reconcentrate the solution before it is recycled.

The range of osmotically dried products available commercially is limited to date. Some good-quality fruit products are on the market. These are osmotically dried, followed by air or vacuum drying. However, there is considerable interest in this process worldwide.

See also: **Drying**: Theory of Air-drying; **Freeze-drying**: The Basic Process; Structural and Flavor (Flavour) Changes; **Drying**: Equipment Used in Drying Foods

Further Reading

Brennan JG (1994) *Food Dehydration – A Dictionary and Guide*. London: Butterworth-Heinemann.

Jones PL (1993) Alternative methods of industrial drying. In: *Encyclopaedia of Food Science, Food Technology and Nutrition*, vol. 3, pp. 1476–1480. London: Academic Press.

Lewicki PP and Das Gupta DK (1995) Osmotic dehydration of fruits and vegetables. In: Mujumdar AS (ed.) *Handbook of Industrial Drying*, 2nd edn, vol. 1, pp. 691–713. New York: Marcel Dekker.

Schiffmann RF (1995) Microwave and dielectric drying. In: Mujumdar AS (ed.) *Handbook of Industrial Drying*, 2nd edn, vol. 1, pp. 345–372. New York: Marcel Dekker.

Physical and Structural Changes

W F A Horner, University of Hull, Hull, UK

Introduction

The concentration of solutes as drying proceeds has consequences on the nature of foodstuffs: some of these changes are reversible and some are irreversible. The manner of drying, its conditions, and extent determine this reversibility and, therefore, how closely the characteristics of the rehydrated food resemble the material from which it originated.

Physical Changes

Changes in color and texture, accompanied, in solid structured materials, by shrinkage and deformation, are expected outcomes of drying. As solutes concentrate in the milieu, osmotic gradients cause water molecules to move out of their close association with hydrophilic macromolecules. The latter tend to aggregate: some, like starch and cellulose, to such an extent that they became almost crystalline. Such close aggregations of molecules are not easily dissociated on rehydration. This helps to explain the frequent observation, illustrated in **Figure 1**, that many foods with the same water content have a higher water activity-(a_w) (which usually means greater susceptibility to microbiological spoilage) when rehydrating than when dehydrating. Although dehydration may also cause changes in the structure of macromolecules, e.g., the denaturation of proteins, experimental evidence has shown that many such denatured structures hold more hydration water than undenatured structures. This is because the denatured form often exposes hydrophilic groups not exposed in the native form. (*See* **Water Activity**: Effect on Food Stability.)

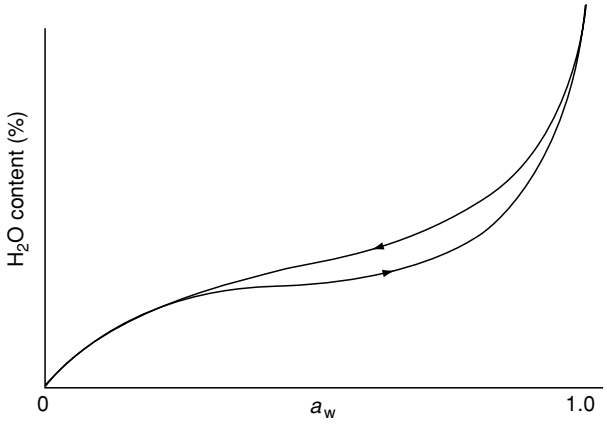

Figure 1 Typical food sorption isotherm showing hysteresis. Reproduced from Drying: Physical and Structural Changes, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

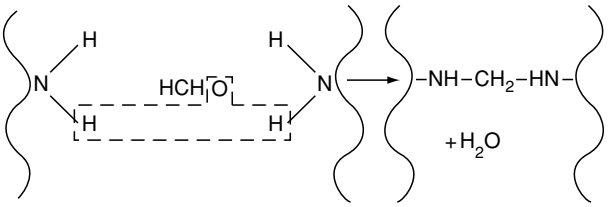

Figure 2 Formation of a methylene bridge between two protein molecules. Reproduced from Drying: Physical and Structural Changes, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

It seems likely, therefore, that the major textural changes in food, characterized by collapse of cell walls and membranes and the change from the gel-like to the fibrous, are due more to the purely physical moving together of macromolecules than to denaturation. Such aggregation may subsequently favor the cross-linking between macromolecules via hydrogen bonds and sulfur bridges, further restricting the extent to which rehydration can occur. If meat, for example, is overdried, down to the a_w 0.15–0.40 region, it reaches maximum toughness and elasticity. Moreover, upon rehydration, muscle tissue dried in air to this extent has lost its natural succulence. To chew such material is to chew inert wetted fibers which very quickly yield their absorbed water to become an unpleasantly dry ball of fibers in the mouth.

Texture Changes and Rehydratability

The modulus of elasticity in vegetable tissue decreases rapidly as, in drying, cells lose their turgor and the plant wilts. In this way, the trend in texture for vegetable tissue appears opposite to that in flesh foods as a_w decreases. It seems, therefore, that the textural properties of foods, such as crispness, hardness, and toughness, on drying, depend on their cellular nature. Pretreatment with a glycerol solution, however, helps to protect vegetable cell walls from disruption and fracture during freeze and air-drying and allows the tissue to return to its original shape on rehydration. Untreated tissue remains partly shriveled on rehydration.

In dehydrated animal tissues, aggregation of the actomyosin fraction causes the texture to become tough and woody so that the flesh rehydrates slowly and incompletely. The formation of methylene (-CH_2-) linkages between myosin molecules was thought to be responsible for the gradual toughening of fish muscle texture during frozen storage. The plausible initiator of such cross-linking was formaldehyde released when trimethylamine oxide, naturally present in marine and some fresh-water fish species, is enzymatically broken down to dimethylamine (eqn (1)). Formaldehyde from this reaction could react with certain protein end-of-side-chain groups like the amino group and possibly form methylene bridges between protein molecules (**Figure 2**). Analytical evidence suggests, however, that such methylene bridges between protein chains occur rarely, if ever. Spoilage product molecules, however, being much smaller than the structural protein molecules, are much more mobile and tend to react together, excluding water and causing irreversible texture changes. The muscle protein myosin, if very fresh, can dry like a gel and can be easily rehydrated. Myosin which is less fresh or is dried under harsh conditions tends to dry as difficult-to-rehydrate fibers.

$$N(CH_3)_3O \rightarrow HN(CH_3)_2 + HCHO \qquad (1)$$

Freezing is a form of dehydration and as such promotes the macromolecule aggregation which hinders the rehydration process. Nevertheless, it is used to improve rehydratability in dried vegetables. Ice crystal growth during freezing forces tissues apart and leaves pores in the material upon drying. Thus, being more porous, the dried food imbibes water more quickly upon rehydration. Water-holding capacity, however, is not improved. The use of cryoprotectants in freezing reduces the tendency of macromolecules to aggregate by holding more water molecules around them. (*See* **Freezing**: Structural and Flavor (Flavour) Changes.)

With the exception of freeze-drying, dehydration processes cause shrinkage which is greatest during the initial stages. If the rate is too fast, shrinking

capillaries bearing water to the food surface soon become blocked by solids carried in the flow. The surface becomes dry and impermeable – a condition known as 'case hardening' – halting the drying process whilst the interior is still wet. Flesh foods dried thus decompose from the inside but, with some vegetable foods, a high initial drying rate is advantageous because the final volume is fixed early in the drying process. As drying proceeds further, the tissues split and rupture internally, forming an open microstructure. The product, in this case, has a low bulk density and good rehydration characteristics. Such materials subjected to low initial drying rates shrink inwards progressively with process time, yielding a high bulk density product which is difficult to rehydrate. (*See* **Freeze-drying**: The Basic Process.)

Liquids like milk and hydrocolloid solutions are frequently spray- or roller-dried. The extremely short contact time (microseconds for droplets of 10 μm diameter) required to dry milk to a powder in a spray drier causes very little damage to components, loss of nutrients, or destruction of microorganisms. However, the powder produced, composed as it would be of particles which are mainly very small, does not easily or quickly reconstitute because surface tension effects on such tiny particles prevent efficient dispersal. To improve reconstitutibility, the fine powder is often rehumidified, causing particles to stick together, and redried as larger more easily dispersible agglomerates. This process is referred to as 'instantization'.

Roller-dried liquids and slurries are exposed to high temperatures for a longer period than spray-dried ones; indeed, the extra contact time may contribute to 'cooking' the paste as well as drying it. Instant or pregelatinized starch powder, which is used as the main functional component of instant pudding mixes, is manufactured in this way.

Unlike the corresponding spray-dried version, the roller-dried product tends to be coarse and more dispersible. Thermal damage to the nutrients and sensory characteristics of such products may be minimized by carrying out the roller-drying process in a vacuum chamber.

Flavor and Color Retention

In spray drying, where there is a huge temperature differential between feed material and drying medium, the consequent large loss of volatiles close to the point of liquid introduction to the chamber may be reduced by the addition of an emulsified oil phase to the feed which alters the droplet size distribution and the pattern of flavor loss. Flavor oils themselves, moreover, can be emulsified with suitable hydrocolloids dispersed in water (continuous phase) so that, on spray drying, the oil droplets are protected from evaporation by the hydrocolloid 'skin,' which instantly encloses them. This microencapsulation process yields almost odorless flavor powders which release their volatiles slowly – an advantage when baking a cake, for example – or quickly when the microcapsules are ruptured, as in 'scratch-and-sniff' perfume samplers.

The retention of the flavor components of a food material during the drying process in unaltered form is a major positive quality attribute for some dried products. Consumers are prepared to pay a premium, for example, on instant coffee expensively produced by freeze-drying which, in spite of losses due to the high vacuum, retains a greater proportion of the highly volatile characteristic flavors than does spray drying. (*See* **Coffee**: Instant.)

The loss of volatiles can be very serious in products which depend significantly upon their aroma for acceptability. Even for milk there is evidence that the loss, during drying, of volatiles such as methyl sulfides and the lower fatty acids may be responsible for the lack of 'fresh milk' flavor in the dried product. (*See* **Sensory Evaluation**: Aroma.)

Heat-induced reactions such as enzymatic and nonenzymatic browning may lead to significant changes in flavor as well as color. Such changes are not always undesirable; indeed, they may be an essential product attribute as in the case of black tea compared to green tea production. Enzymatic changes in dried foods are dealt with in greater detail in the following section, but control can be exercised by selection of the variety and stage of maturity of the material to be processed. In this way it may be arranged that the starting material is as deficient as possible in the substrate or enzyme involved in browning. Blanching vegetables before dehydration and excluding oxygen from the product, respectively, destroy the enzyme and prevent the oxidation stage of the sequence leading to browning. Sulfur dioxide acts both as an enzyme poison and reducing agent and is therefore frequently used to preserve the color of dried fruits and vegetables. Because it is volatile, it is easily driven off as the food is reconstituted and cooked. Normally, sulfur dioxide is applied to vegetable matter prior to drying through dipping them in a solution of sodium metabisulfite for a period which yields a product conforming to the maximum permitted limits for that chemical in that specific material. (*See* **Browning**: Nonenzymatic; **Preservation of Food**.)

Selection of the raw material can also play a significant role in maintaining the desired color quality of dried foods subject to nonenzymatic browning. Potato crisps, if made from stored potatoes, where

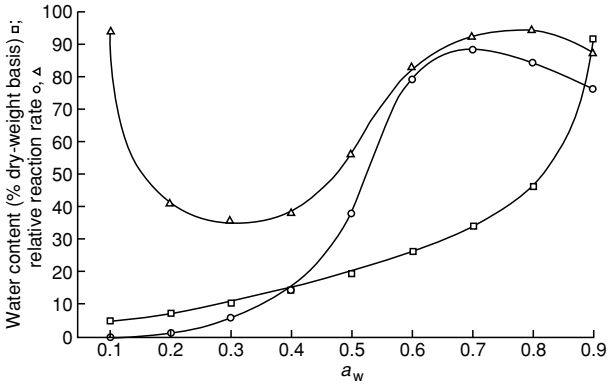

Figure 3 Dried product storage stability (susceptibility to lipid oxidation and Maillard browning) related to its water activity. Squares, sorption isotherm; triangles, lipid oxygenation; circles, Maillard browning. Modified from Labuza TP, Tannenbaum SR and Karel M (1970) Water content and stability of low moisture and intermediate moisture foods. *Food Technology* 24: 543–550.

the starch has begun to break down to sugars prior to germination, become too dark before water content has been reduced by the frying process sufficiently to give the required characteristic crisp texture.

The Maillard nonenzymatic browning reaction sequence is initiated when the carbonyl group of reducing sugar reacts with the amine group of a protein peptide or amino acid. The dire effects of this on the visual quality of spray-dried egg have required that one of the reactants, glucose, be removed by treatment with glucose oxidase, before the drying process.

Fish which has spoiled considerably before drying contains substantially more of the mobile Maillard reactants than fresh fish and so becomes an unsightly brown color upon drying. Figure 3 shows how the rate of browning is affected by the water activity of the product. (*See* **Oxidation of Food Components**.)

Microstructure of Dried Products

Dried foods may be particulate (e.g., potato powder), glass/rubber-like (e.g., many sugar confections and pasta), fibrous, or a combination of these.

Particulate dry foods may range from simple crystalline chemicals, like salt and sugar, to amorphous conglomerates of highly complex substances like milk powder.

In many instances, dried foods are eaten without rehydration (e.g., dried fruits and sugar confections) and their chewy, tough, or brittle texture contributes to the pleasure of eating them. The size range of simple crystalline substances like sugars is critical in determining the smoothness of a fondant and the brittleness/crunchiness of a biscuit.

The manner in which lipid material is distributed in relation to the structural matrix (carbohydrate or protein) and interstitial components affects what is referred to as the 'shortness' in texture of the product. The lipid interrupts the potential continuity of such matrices which would otherwise lead to toughness and elasticity. This 'shortening' facility may be used to isolate relatively small units of glutenacious material so that no continuum is formed, in the baking of shortbread biscuits, for example, or to isolate layers of glutenacious matrices from one another, as in cream crackers and puff pastry.

Removal of water from foods inevitably leads to the precipitation of solutes and the aggregation of insoluble structural components – with consequent effects on texture and rehydratability. Some of the soluble and colloidal components of the food starting material, gelatinized starch, for example, form a glassy continuum within the food material as it dries. This phenomenon is a highly significant feature in the cook–extrusion of starchy materials. Indeed, the viscous glassy continuum formed in the extruder barrel momentarily holds the water, as it suddenly turns to steam the instant the material emerges from the exit die, and allows the extrudate to expand and quickly solidify as the steam escapes. The hygroscopicity and continuity of such glassy components favor very rapid rehydration – an unwelcome property with respect to long-term storage in humid environments for many cook-extruded products destined for the snack-food market. The 'melt-in-the-mouth' characteristic imparted by the products' aeration and hygroscopicity is, however, its strongest sensory advantage – although with some such products rehydration in the mouth is so rapid that the product tends to cling to the palate and the teeth. Packaging for these hygroscopic, aerated extruded products must provide an efficient barrier against the ingress of water vapor and – if, as is usual, the product contains a significant proportion of fat – oxygen. The porosity of dried foods greatly assists the process of oxidative rancidity which is also promoted (Figure 3) by the removal of the coating of water molecules around reactive sites at very low a_w. In many dehydration processes, an attempt is made to maximize dried product porosity (as in puff, foam-mat, and freeze-drying) to improve dehydratability. Less controlled drying processes often yield shrunken or case-hardened products which, due to the capillary constriction, are very difficult to rehydrate. Porosity and rapid hydratability are not always desirable properties in dried foods. In Indonesia, fish may be dried, smoked, and pressed alternately until a dry horny product (ikan kaju) is produced. This product, which is grated over soups as a condiment, resists rehydration from the humid

surroundings for a year or more. (*See* **Extrusion Cooking**: Chemical and Nutritional Changes; **Sensory Evaluation**: Texture.)

Effects of Drying on the Microflora of Foodstuffs

Dehydration is a method of food processing which relies for its effectiveness on the prevention of microbial proliferation by making the food environment too water-retentive for microorganisms to abstract it for their metabolic requirements (Table 1).

Lowering the a_w of many fresh foods from 1.00 to inhibit the growth of only some of the bacteria requires that around 90% of the water content of the raw material be removed unless drying is accompanied by the infusion of solutes as in salting.

The lowering of a_w from 0.95 to 0.60 (enough to suspend growth of all microorganisms) requires only a relatively small reduction in water content but, as drying merely suspends microbial proliferation, rather than destroying the organisms, outgrowth quickly recurs in dried foods as they absorb small amounts of water from a humid environment. Certain species of xerophilic yeasts and osmophilic molds are able to proliferate on material which is still apparently dry. The mold *Wallemia sebi*, associated with rotting wood, has often caused problems of surface spotting, 'dun' spoilage, in dried salt-cured fish. Halophilic bacteria, like *Halinobacterium salinaria*, can cause more severe damage, proteolysis, and putrefaction in salt-cured fish during the 'water reduction/salt-infusion' process if the ambient processing temperature is not kept below 10 °C. The characteristic coloration of the colonies of such organisms, which eventually pervade the entire product surfaces, gives rise to the description 'pink' spoilage. The microbial flora so far described cause spoilage of dried foods at intermediate water contents whether during desorption, in processing, or resorption during subsequent storage in too humid ambients. There are, however, pathogenic species which can affect the safety of dried food products, particularly if atmospheric moisture condenses upon the surfaces.

Certain molds produce mycotoxins whilst growing upon such surfaces. The most dangerous of these are the highly carcinogenic aflatoxins produced by the black mold *Aspergillus flavus*. Of the pathogenic bacteria, *Staphylococcus aureus* is the most tolerant towards low water availability, being capable of growth down to around a_w 0.8, although it is doubtful whether it would produce its exotoxin at these levels. Perhaps of greatest concern, though, is the ability of much of a food's contaminating microflora to survive and remain dormant in the dried food and then proliferate as the reconstituted food is held. Cases of *Salmonella* infection stemming from the use of dried infant formula foods contaminated with such organisms have emphasized this. (*See* **Mycotoxins**: Occurrence and Determination; ***Staphylococcus***: Properties and Occurrence; **Spoilage**: Bacterial Spoilage; Molds in Spoilage; Yeasts in Spoilage.)

It seems to be generally assumed that dried foods once rehydrated are consumed more or less immediately. Since this is not always the case, and because the heat exposure involved in most drying processes is insufficient to destroy pathogens the starting material may contain, careful selection of raw materials and hygienic processing conditions may not be sufficient to guarantee the safety of the product under all possible consumer usage circumstances. Heat or irradiation pasteurization of the starting material may have

Table 1 The water activity (a_w) of foods and microbial inhibition

a_w	*Food examples*	*Microorganisms inhibited*
1.00	Most fresh, high-water-content foods	None
0.95	40% sucrose or 7% salt (NaCl) solutions, breadcrumb, cooked sausage, canned ham, sponge cake, immature cheese	Gram-negative rods like *Escherichia coli*, spores of Bacillaceae
0.91	55% sucrose or 12% salt solutions, Parma ham, mature cheese, reduced-sugar jams	Most cocci and lactobacilli vegetative cells of Bacillaceae
0.88	65% sucrose or 15% salt solutions, salami sausage, Parmesan cheese, fishmeal	Most yeasts
0.82	Wheat flour, dry cereal grains and pulses, fruit cake, intermediate-moisture pet foods, traditional fruit conserves, dry smoked sausages, orange juice concentrates	Most molds and *Staphylococcus aureus*
0.75	26% salt (i.e., saturated NaCl) solution, honey fondants, rolled oats, rich fruit cakes, glacé cherries, salt-cured cod (prior to final drying), marzipan	Most halophilic bacteria
0.65	Dried dates, prunes, and figs, toffees and caramels, soft brown sugar, stockfish, fishmeal dried to 5% water content	Xerophilic molds
0.60	Licorice, fruit gums, salt-cured cod (after final drying)	Osmophilic yeasts

to be considered as an essential part of the process. For example, spray-dried egg should be produced from pasteurized liquid egg. Alternatively, the dry product itself may be treated. Some dried herbs and spices are irradiated to destroy a large proportion of their contaminating flora, especially potential pathogens. (*See* **Irradiation of Foods**: Basic Principles; **Pasteurization**: Principles.)

See also: **Browning**: Nonenzymatic; **Coffee**: Instant; **Extrusion Cooking**: Chemical and Nutritional Changes; **Freeze-drying**: The Basic Process; **Freezing**: Structural and Flavor (Flavour) Changes; **Irradiation of Foods**: Basic Principles; **Mycotoxins**: Occurrence and Determination; **Oxidation of Food Components**; **Pasteurization**: Principles; **Preservation of Food**; **Sensory Evaluation**: Texture; Aroma; **Spoilage**: Bacterial Spoilage; **Water Activity**: Effect on Food Stability

Further Reading

Christian JHB (1963) Water activity and growth of microorganisms. In: Hawthorn J and Leitch JM (eds) *Recent Advances in Food Science*, vol. 3. London: Butterworths.

Davies R, Birch GG and Parker KJ (1976) *Intermediate Moisture Foods*. London: Applied Science.

Duckworth RB (1975) *Water Relations of Food*. London: Academic Press.

Gilbert SG (1986) New concepts on water activity and storage stability. In: Charalambous G (ed.) *The Shelf Life of Foods and Beverages*. Amsterdam: Elsevier.

Hamm R (1963) The water imbibing power of foods. In: Hawthorn J and Leitch JM (eds) *Recent Advances in Food Science*, vol. 3. London: Butterworths.

Horner WFA (1992) Preservation of fish by curing (drying, salting and smoking). In: Hall GM (ed.) *Fish Processing Technology*. Glasgow: Blackie.

Johnston K (1985) Water-sorbent-solute relationships in dry food model systems. Paper given at a Royal Society of Chemistry Discussion Conference, *Water Activity: A Credible Measure of Technological Performance and Physiological Viability*. Cambridge: Girton College.

Labuza TP, Tannenbaum SR and Karel M (1970) Water content and stability of low moisture and intermediate moisture foods. *Food Technology* 24: 543–550.

Mathlouthi M (1986) *Food Packaging and Preservation: Theory and Practice*. London: Elsevier.

Mossel DAA and Westerdijk J (1949) The physiology of microbial spoilage of foods. *Antonie van Leeuwenhoek* 15: 190–202.

Scott WJ (1957) Water relations of food spoilage microorganisms. In: Mark EM and Stewart GF (eds) *Advances in Food Research*, vol. 7. London: Academic Press.

Van den Berg IC and Bruin S (1981) Water activity and its estimation in food systems: theoretical aspects. In: Rockland LB and Stewart GF (eds) *Water Activity: Influences on Food Quality*. London: Academic Press.

Chemical Changes

N Shilton, University College Dublin, Dublin, Ireland

Background

Drying of food products can be carried out by using a variety of different processes, ranging from solar drying, air-drying to the more expensive freeze drying. All of these processes influence the properties of the resulting product in some way, not least in the way that components of the food will undergo chemical reactions during the drying process. Usually, these reactions will be detrimental to product quality, but this is not always the case. For example, drying usually results in a reduction in volatiles due to vaporization, but in certain cases, the volatiles can increase as a result of drying; for example, it has been found that drying of basil leaves in a vacuum-microwave dryer actually increases the amount of volatiles present in samples when compared with fresh samples, as a result of chemical reactions.

There are many different methods of drying food materials. The main drying processes that have been identified as being studied with respect to chemical changes are solar drying, air convective drying, microwave drying, and freeze-drying. The effects of these drying processes on the constituents of food products will be examined in the context of the chemical reactions occurring during the drying process. The chemical changes occurring during the drying process will be examined, and not those that occur during storage.

Reaction Kinetics/effect of Water Activity

As water is removed from the product being dried, the water activity of the product decreases. As the water activity decreases, so too will the reaction rates of some chemical reactions, such as oxidation of free fatty acids. However, for a number of chemical reactions, such as Maillard browning and anthocyanin degradation, the reaction rate increases in the water activity range 0.4–0.8, before decreasing again, whereas other reactions such as autoxidation undergo an exponential type increase in reaction rate as the water activity is reduced to below 0.2. This can be related to an increase in the concentration of reactants as the moisture content falls, which can lead to an increase in reaction rates.

An examination of the chemical kinetics of drying shows that as the temperature of the product increases, so will the rate constant of the decomposition of food components, so it is recommended that foods

be dried at low temperatures or under vacuum. The chemical changes described in this section will show the types of chemical reactions that can occur if the drying process is not controlled carefully.

Effect on Proteins

The main effect of a drying process on proteins is that of denaturation. This is not a chemical reaction in itself (in the initial stages) and usually involves a change in the tertiary or secondary structure of the protein. However, severe denaturation will involve a chemical reaction in the formation or dissociation of disulfide bonds and coagulation of the proteins. The drying temperature will play an important part in determining the changes undergone by proteins during drying. The lower the drying temperature, the less thermal, and so potential chemical damage will be done to the proteins. This has the effect of extending the drying time of the product. An extension in drying time will also increase the risk of degradative reactions as the water activity passes through the critical zones described earlier.

During the spray drying of skim milk, milk proteins can undergo lactolyzation. High drying and outlet temperatures can cause extensive lactolyzation, but reducing the inlet and outlet temperatures that the degree of lactolyzation can minimize this reaction.

Amino Acids

Specific amino acids can be lost as a result of chemical reactions. For example, losses of lysine are higher with drum-dried milk than with spray-dried milk powder. This is most probably as a result of the direct contact with the hot drum, which causes thermal cross-linking between the disulfide bonds in the lysine. One of the most important reactions involving amino acids is the Maillard reaction; this is described in the section on Maillard browning.

Effect on Carbohydrates

For starchy products, such as potatoes, if the drying temperature is too high, there is a risk of gelatinization of the starch granules within the cells. This is often not desired. Increasing the drying temperature above 65 °C has also been found to increase the rate of glucose and fructose degradation during the drying of products such as onions. One of the most important reactions involving carbohydrates is the Maillard reaction; this is described in the section on Maillard browning. This reaction can also modify the resulting aroma profile of the dried product.

During the drying of prunes, three phases involving chemical changes to constituent carbohydrates can be identified; these correspond to the decreasing moisture content. As the moisture content is initially reduced, there is a hydrolysis of sucrose to give fructose and glucose, and at intermediate moisture contents, there is a rapid loss of this fructose and glucose as a result of the Maillard reaction. As the moisture content is reduced further, caramelization of the sugars will occur.

Maillard Browning

Nonenzymatic browning is the major chemical reaction that can take place during the drying of all types of foods. The most prevalent form of nonenzymatic browning that can occur is that of Maillard browning. This is a complex series of reactions that is initiated by the reaction of amino acids and a reducing sugar, in the so-called Amadori rearrangement, that leads, *via* a series of intermediate compounds, to give melanoidins, insoluble compounds with a characteristic brown color. The reaction is very heat-sensitive, and so during drying, the use of higher temperatures, in high-temperature short-time processing, can cause an increase in the reaction rate. Formation of the intermediate compounds takes some time, so it is possible for a dried product to be produced that has no visible signs of browning, but has in fact undergone substantial chemical reactions. This could have an impact on the nutritive quality of the product. The degree of Maillard browning can be controlled by the use of sulfite treatment, through the use of either a metabisulfite solution or sulfur dioxide.

Of potential concern is the possible production of carcinogenic compounds during the Maillard reaction, either as intermediate compounds or as compounds during the final color development. Dry heating also produces aminocarbolines, and the formation of these seems to be favored by free-radical reactions and a low water activity. However, the Maillard reaction can also produce antioxidative components, and these have been found to prevent oxidation in products such as oven-dried peanuts.

It has been found that harsher, higher drying temperatures will cause an increase in Maillard browning, but this can be controlled by monitoring the free amino acid levels in the product and adjusting the drying temperature accordingly. In pasta, a combination of high temperature and a high water activity during the initial stages of drying will contribute to cause nonenzymatic browning by the Maillard reaction. With pasta dried at high and low temperatures (50 and 110 °C), the higher processing temperature causes a large increase in furosine over the lower processing temperature. Furosine is an important precursor in the Maillard reaction.

The Maillard reaction can also result in the degradation of amino acids by the so-called Strecker degradation. This reaction produces aldehydes, ammonia, and carbon dioxide. The aldehydes contribute to the development of aroma during the Maillard reaction, with individual amino acids producing different aldehydes and so different aromas.

Effect on Fats

Oxidation of fats and lipids is most likely more of a problem during long-term storage of dried products. However, high-fat products or those containing unsaturated fatty acids are more likely to undergo chemical reactions during the actual drying process. Dried fish products are an example of a product type that is particularly susceptible to oxidative deterioration of fats and lipids. As oxidation and chemical reactions proceed, the relative proportions of saturated and monounsaturated fatty acids will increase, and the proportions of polyunsaturated fatty acids will decrease. Impingement drying, a high-temperature short-time process, of pressed fish cake can help minimize the oxidation of polyunsaturated fatty acids. However, the formation of free fatty acids during drying can also be a problem, and these will contribute to overall product degradation during subsequent storage. Polyunsaturated fat and carotenoid oxidation can be a problem in products such as dried egg powder. Low spray-drying temperatures are recommended in combination with the use of antioxidants.

Effect on Nutrients

Reactions involving minor nutrients during drying are of particular concern in considering the chemical changes during drying. For example, dried baby foods and infant formulae are products where nutrient destruction could be potentially harmful to the health of the consumer. The type of drying process chosen can have a major impact on the potential for nutrient degradation; for example, very few chemical changes occur during the spray drying of infant formulae, whereas roller drying can cause degradative reactions, such as nonenzymatic browning, by the Maillard reaction in the final product. In some cases, the dried product may have improved nutrient availability over more traditional processes; for example, it has been found that the bioavailability of nutrients in powdered formula is better than that in sterilized liquid formula. This is most likely because the powdered formula had a less severe heat treatment during drying than the sterilization step in the production of the liquid formula.

Effect on Vitamins

The water-soluble vitamins are sensitive to chemical degradation as a result of drying; for example, loss of vitamin C is dependent on the presence of heavy metals such as copper and iron, light, water activity, and the temperature of drying. Losses can typically run between 10 and 50%. Loss of thiamin during food processing has been extensively researched, and losses of up to 89% have been reported with certain combinations of moisture content, drying temperature, and drying time. In general, losses of water-soluble vitamins as a result of conventional drying processes are less than 20%, and for vegetables, these losses are generally less than 5%, except for vitamin C.

The type of drying process again has a major influence on vitamin retention. Freeze-drying preserves the vitamin C content more efficiently than processes such as oven drying or solar drying. This can be attributed to the high drying temperatures of the oven dryer and the slow drying rate of the solar drying process, and many researchers have reported a relationship between increased drying temperature and increased vitamin C degradation. Degradation of vitamin C has been found to follow first-order reaction kinetics in different products, such as guava and potatoes. In general, vitamin C is better preserved when rapid drying rates are used rather than with slower drying rates.

Losses of fat-soluble vitamins can occur as a result of oxidation and, again, can be linked with the drying temperature. Losses of vitamins A, D, and E are often reported to be negligible; the vitamin most prone to loss would appear to be β-carotene, with other studies reporting significant losses of carotenoids and tocopherols. It is often difficult to compare losses of vitamins, as retention will be dependent on a number of factors, such as the food material, type and extent of pretreatments, and drying conditions.

Effect on Dietary Fiber

Dietary fiber has also been shown to be modified during processing. Thermal processing of wheat bran, including by drum drying, causes an increase in soluble dietary fiber and a decrease in insoluble dietary fiber. The total dietary fiber content is unaffected.

Color Degradation

Color is often degraded during drying, especially during the drying of fruits and vegetables. Again, this can be correlated most often with drying temperature; for example, a more rapid drying process improves the retention of color as a result of reduced degradation of carotenoids and chlorophyls in many

vegetable products. Degradation of carotenoids is most often a result of oxidation, particularly as carotenoids have a high degree of unsaturation in their molecular structure.

Pretreatments can also have an effect on the carotenoid content of the final dried product; for example, blanching of carrots in water causes more carotenoid degradation during drying than steam blanching. However, pretreatments is beneficial, as unblanched controls will suffer the greatest degree of carotenoid losses. The use of additives such as citric acid, salt, and sodium metabisulfite can help retain carotenoid during convection drying. Treatment of carrots has been shown to have a pronounced effect on improving the carotenoid retention in carrots. Chlorophyl has been shown to be fairly stable at low moisture contents. However, degradation does occur at low pHs, when it is converted into pheophytin.

Again, the type of drying process used will affect the overall color retention. For example, in a comparison between hot-air drying and microwave drying, it was found that hot-air drying at 60 °C had the least effect on color degradation of kiwi fruit, whereas microwave drying caused the highest degree of browning.

See also: **Browning**: Nonenzymatic; **Carotenoids**: Physiology; **Chlorophyl**; **Drying**: Physical and Structural Changes; **Infant Foods**: Milk Formulas; **Powdered Milk**: Characteristics of Milk Powders

Further Reading

Borquez R, Wolf W, Koller WD and Spiess WEL (1999) Impinging jet drying of pressed fish cake. *Journal of Food Engineering* 40(1/2): 113–120.

Damodaran S (1996) Amino acids, peptides and proteins. In: Fennema O (ed.) *Food Chemistry*, 3rd edn. pp. 321–429. New York: Marcel Dekker.

Hwang JK, Kim CT, Cho SJ and Kim CJ (1995) Effects of various thermal treatments on physicochemical properties of wheat bran. *Korean Journal of Food Science and Technology* 27(3): 394–403.

Jayaraman KS and Das Gupta DK (1995) Drying of fruits and vegetables. In: Mujumdar AS (ed.) *Handbook of Industrial Drying*, pp. 643–713. New York: Marcel Dekker.

Sarria B, Lopez-Fandino R and Pilar-Vaquero M (2001) Does processing of a powder or in-bottle-sterilized liquid infant formula affect calcium bioavailability? *Nutrition* 17(4): 326–331.

Sokhansanj S and Jayas DS (1995) Drying of foodstuffs. In: Mujumdar AS (ed.) *Handbook of Industrial Drying*, pp. 589–625. New York: Marcel Dekker.

Wilford LG, Saberez H and Price WE (1997) Kinetics of carbohydrate change during dehydration of d'Agen prunes. *Food Chemistry* 59(1): 149–155.

Hygiene

M N Ramesh and A Chakkaravarthi, Central Food Technological Research Institute, Mysore, India

Introduction

Food hygiene is a subject of wide scope. It aims to study methods for the production, preparation, and presentation of food, which is safe, and shelf-stable. It covers the proper handling of every variety of foodstuff and drink, and the equipment and apparatus used in their preparation, service, and consumption. Good hygiene is a necessity in all food industries. The consumer has the right to demand both good quality and good shelf-life. The safe and hygienic processing of food products is becoming even more important as public awareness of food safety continues to increase. Hygienic design of any food-processing equipment is essential if a microbiologically safe product is to be produced. Food should be nutritious and attractive. It must be visibly clean and free from noxious materials which may lead to food poisoning or gastroenteritis. Foods heavily contaminated with food-poisoning germs may be normal in appearance, odor, and flavor. Microorganisms may be introduced directly from infected raw materials, from workers, or the environment during the preparation of foods. Poisonous substances may be produced by the growth of bacteria and molds in food.

The food-processing plant is made up of four basic components:

1. factory layout, structure, fittings, and perimeter
2. equipment
3. personnel
4. handling of foodstuffs from raw material to the final finished product

Plant for manufacture of food products should be laid out, preferably in straight lines of product flow. Final or intermediate processes must not be in a position where there could be cross-contamination by raw materials. Appropriate walls should physically separate the preparation areas for raw materials and dried materials.

The risk of contamination of food products with relevant microorganisms during open processing increases with the opportunity to grow in poorly designed driers as well as with the concentration of the microorganisms in the environment. The risk of airborne infection in open handling is much higher. This means that, in open drying, environmental

conditions in addition to appropriate drier design have an important influence on hygienic operations.

This is often overlooked, in spite of the well-known fact that microorganisms may be carried long distances in aerosols, which are easily produced during cleaning of yards, floors, and the outsides of machinery, not to mention water treatment. Another point of importance in open food processing is the traffic of people in process areas. As between the raw material area and the process area, staff, facilities, and even air should be kept strictly separated.

In a closed process the function of the equipment, the cleaning efficiency, and the control of process parameters are most important for hygienic operation. There is a tendency to trust automation rather than common sense for cleaning. Cleaning must be considered as a process, and the product of cleaning should be clean equipment. Inadequate cleaning may be due to mechanical malfunction, the wrong type or concentration of detergent, the wrong time and temperature during cleaning, or wrong water quality with regard to detergents used, or to several of these in combination. Inadequate cleaning should be discovered on process control and never be allowed to cause loss of product through poor hygienic quality. Inadequate cleaning, however, has many other economic consequences, beside the contamination risk. Poor product shelf-life, malfunction and corrosion of equipment, overconsumption of detergents and water, expensive waste water treatment, and energy losses are just a few examples.

Hygienic design of drier should be based on optimum combination of mechanical, process, microbial, and cost considerations. It is important that food product and process are designed by competent food scientists and technologists capable of understanding the individual importance and interrelationships of the physicochemical parameters. The water activity (a_w) requirement of the various microbiological agents of food spoilage and food poisoning should be understood and, particularly, the limiting a_w for known pathogens should be known.

Processing Condition

During conventional drying, hot air is used as a heating medium. Air carries many microorganisms. Concentrations of 100–10 000 microorganisms per cubic meter are quite normal. The concentration differs according to season and location. Therefore, depending on the food produced or processed, control of airborne microorganisms may be necessary.

Generally dry air with low dust content and higher temperature has a low microbial level. The organisms, which can be predominantly present in air, are spores of *Bacillus* spp., *Clostridium* spp., molds and some Gram-positive (e.g., *Micrococcus* spp., and *Sarcina* spp.) species as well as yeasts. If the surroundings contain a source of pathogens (e.g., animal farms or sewage treatment plant), different types of bacteria, including pathogens and viruses (including bacteriophages), can be transmitted via the air. Ducts may collect water and allow the growth of microorganisms, which are subsequently carried by the flow of air. It is important therefore to insure that air ducts are fully self-draining or that other measures are taken to prevent the accumulation of condensate.

The air used for drying must itself be of good quality; thus removal of moisture, oil, and particles from air is essential. Treatments such as inactivation or removal of microorganisms present in the air can be applied. Several methods may be used to reduce the number of microorganisms in the air. These include physical treatments and chemical agents or a combination of both. However, one of the most used methods of producing sterile air with higher assurance of sterility is filtration. Filtration is the removal of particles, including microorganisms, from the air. Air filtration has greatest particle potential of all the separation methods. Air sterilization is the physical removal of microorganisms from the air by filters of appropriate retention efficiency. Depth filters are made of cellulose, glass fiber mixtures with resin, or acrylic binders. The fibrous sheet filters have a low resistance to air flow and a large surface area. Such a filter is required to provide air with an extremely low microbiological load to aseptic areas.

Principles of Hygienic Design

The most important aim in the design of driers is to avoid pockets and crevices. A pocket is relatively large whereas a crevice is narrow and deep. Normally, a pocket is nothing to be afraid of: it is more important to avoid crevices, which can cause trouble, because the detergent solution is not able to rinse them clean. Sometimes hygienic design requirement may conflict with functional design requirements. In such cases, an acceptable compromise can be found. Where this is not possible hygienic design should be preferred and any deviation from the functionality of the equipment has to be accepted, to insure the microbiological safety of foods. Hygienic design should be based on an appropriate combination of mechanical, process, and microbiological requirements. Complying with hygienic requirements may increase the life expectancy of the equipment, and reduce maintenance needs, especially when cleaning-in-place (CIP) is built into the system, subsequently leading to lower manufacturing costs.

Equipment Design

Hygienic design of a very high standard is essential in the modern plant to minimize down time for cleaning and sterilizing. A drier used for drying food should not be of such a nature that the food is in danger of becoming unfit for human consumption, after drying. The planning of a dehydration plant is just as important as the design of the drier. A good drier can be dangerous if the planning is bad. Mechanization and automation have made great progress in the cleaning of the drier – both batch and continuous, previously cleaned manually. Earlier, the machines were opened every day, and if certain parts were not especially hygienic they were given an extra cleaning. But modern driers employ CIP with computer controls. This will not affect the cleaning efficiency of individual components, but the computer helps to optimize the total cleaning result.

To produce microbiologically safe dehydrated foods, a hygienically designed drier is needed. The drier itself can be a main source of bacterial contamination of foods, if the design and fabrication do not conform to high levels of hygienic standards. Hygiene in the drier is necessary to prevent food poisoning and to reduce food spoilage. The main factors affecting this are biological, but others, such as chemical and physical factors, also contribute. A well-designed drier is no safeguard against bacterial infection and other contamination unless it is accompanied by sound instructions for its installation, operation, cleaning, and maintenance.

To insure that it can be freed from relevant microorganisms, the drier must be properly cleaned before the process begins. Equipment must be hygienically designed. Where crevices and/or dead areas cannot be avoided for functional reasons, equipment may have to be dismantled for cleaning. It is easy to decontaminate a clean surface. The drier should be subjected to inspection and treatment to control vermin and infestation. Care must be taken to prevent the contamination of foods with rodenticide and pesticide residues.

Five principles on which a minimum standard could be established are:

1. All surfaces in contact with food should be inert to the food under the conditions of use, and constituents from their surfaces must not migrate into the food or be absorbed by the food in quantities which could endanger health.
2. All surfaces in contact with food must be smooth and nonporous so that tiny particles of food, bacteria, or insect eggs are not caught in microscopic surface crevices and become difficult to dislodge, thus becoming a potential source of contamination.
3. The drier should be so designed that all surfaces in contact with food can be readily disassembled for manual cleaning. In accordance with current practice, CIP techniques are preferred, but where this is not practicable the design must allow for the ready dismantling of the relevant parts for manual cleaning.
4. The drier must be so designed as to protect the contents from external contamination, including accidental ingress of foreign matter, for example, bolts, nuts, washers, and gaskets.
5. The exterior or nonproduct contact surfaces should be arranged so to prevent harboring of soils, bacteria, or pests in and on the equipment itself as well as in its contact with other equipment, and as far as possible with floors, walls, and hanging supports.

Three basic factors that are important for hygienic drier design are:

1. actual design, fabrication and finish
2. materials of construction
3. cleaning system adopted for the maintenance of the drier

The drier should be designed to prevent the entry of foreign materials, and the development of 'out-of-sight' dead spots, especially within the operation chamber and associated pipe work. Bolts and clips should be attached externally to maintain all internal product contact surfaces smooth and easily cleaned. For operator safety, all equipment should be made safe by electrical isolation during dismantling and cleaning. During the design of continuous driers, metallic belts are used to convey the material from inlet to outlet through the drying chamber. Such driers should be designed so that the whole of the belt is accessible for cleaning. The belts should be made endless, preferably by welding the two ends, instead of riveting, which may be a potential spot for microbial contamination. The transmission systems used to drive the conveyor should be kept out of the chamber and sealed to avoid any contamination from the driving system, as shown in **Figure 1**.

The hygienic design approach for the food industry differs from conventional design, as the hygienic requirements have to be integrated with conventional design to achieve good overall design results. The first design requirement is ease of cleaning and maintaining to insure that the drier will perform as expected and to prevent any contamination during the drying process. Under favorable conditions microorganisms grow very rapidly. Consequently, dead legs, gaps, and crevices, where microorganisms can be trapped and multiply, must be avoided. When support structures

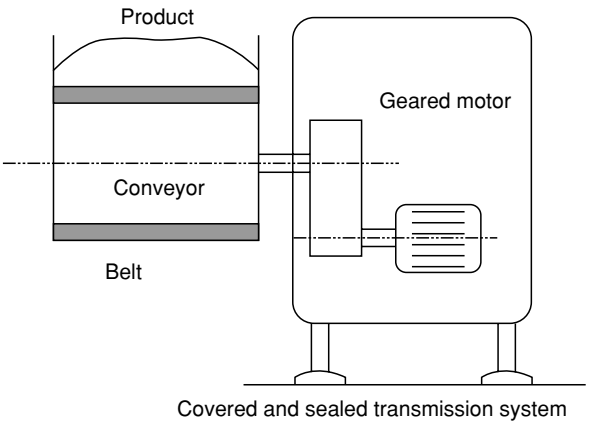

Figure 1 Transmission system for conveyor driers. Reproduced from EHEDG Guideline No. 3 (1994) Hygienic design of equipment for open processing. *Trends in Food Science Technology* 6: 305–310, with permission.

are attached to the floor or to walls, they must be provided with legs of suitable height to provide a minimum clearance for cleaning and inspection.

Fabrication of the drier as per design is an important consideration. The surface finish of the product contact parts should have an acceptable roughness value (*Ra*) of 0.8 μm when measured following ANSI/ASME B.46.1 methodology. Noncontact parts must be smooth enough to insure ease of cleaning. While welding joints, welds should be continuous and free of imperfections. Postwelding operations like grinding and polishing are mandatory.

Choice of Materials

Materials used for construction must be nontoxic, mechanically stable, inert, resistant to the product being dried (at all stages of drying), and also to cleaning and antimicrobial agents at the full range of operating temperatures, The material should also be smooth and with unpainted food contact surfaces. Care is needed to insure that food materials do not attack the metal; for example, acid reacts with mild steel, and salt with aluminum. Equally, the construction materials must not attack the foodstuff, for example, vegetable oil with copper. Appropriate services for cleaning operations, such as hot and cold water, drainage, compressed air, and vacuum lines should be adjacent for cleaning operations. It is easy to choose materials for machines used in the food industry, as those parts in contact with the food product or which are used as protection for the machines should be stainless steel. The other parts come into the category of normal machine construction.

Material selection is generally done separately for:

1. material not in contact, like frames, transmission, etc.
2. materials in direct contact with feed material, like trays, conveyors, feeding hoppers, scooping arrangement, and feeding chutes
3. materials for components like seals, gaskets and bearings, blowers, and ducting

Stainless steel is the universally accepted material for contact parts. Normally only two grades of stainless steel are used: AISI 304 and AISI 316 – a molybdenum alloy stainless-steel that is acid-proof. Both are weldable and can be machined by normal machine tools. AISI 304 is often used externally, as it is only needed to protect the machine from atmosphere, water, and sometimes spilled liquids. A choice must be made between AISI 304 and AISI 316 for parts which are in contact with the product. AISI 304 tolerates the most common cleaning solutions and is suitable for most purposes.

Cleaning Systems

Production operations require careful management on a clean-as-you-go basis to maintain good housekeeping, even during production at peak periods. A structured approach is required to the formal methods and the frequency of cleaning of all plant and equipment, including instructions for reassembly. Automatic cleaning, or CIP, should be introduced at more and more points. For example, a conveyor is used in continuous driers which can be cleaned automatically after use, with the help of cleaning nozzles. A great deal of manual work can be saved, as this type of conveyor is common in the food industry. The automatic wash system has been designed to achieve the advanced cleaning required in today's food-processing environment. The high-pressure wash system is designed to be housed over a conveyor belt unit and employed in the periodic cleaning-down and drying of the food-carrying belts at preset times. Absolute hygiene is required for modern food-processing plants and can be sized and fitted to incorporate any part of the food-manufacturing process line, continually achieving 100% cleaning.

Certain driers often have to be completely dismantled for cleaning, but the dismantling could be made simpler in many cases and involve fewer parts, with quick-releasing systems which are easy to remove. The spaces which are then to be cleaned should be open and incline outwards so that all product material and cleaning liquids can easily drain from the machine. Much has already been done in this field, but the food industry could encourage the manufacturers of the machines to make their

machines easier to clean – and eventually automatically cleanable.

During CIP, various flushing, cleaning, and sanitizing fluids are brought into contact with the wetted parts of the plant. Cleaning may be done on a once-through basis, or some of the cleaning agents may be recirculated to reduce consumption of water, chemicals, and energy. Cleaning is achieved by the physical action of high-velocity-flow jet sprays, agitation, and the chemical action of cleaning agents enhanced by heat. While mechanical forces are necessary to remove soil and to insure adequate penetration of cleaning solutions to all areas, most of the cleaning action is provided by chemicals – surfactants, acids, alkalis, and sanitizers.

A CIP system consists of piping for distribution and return of cleaning agents, tanks and reservoirs for cleaning solutions, heat exchangers, spray heads, flow management devices (supply-and-return pumps, valves, sensors and gages, recording devices) and a programmable control unit, as well as other items. Although a CIP system requires a significant initial capital investment, CIP is economical relative to manual cleaning in most large-scale processes. CIP greatly reduces labor. CIP methods often consume less water and chemicals relative to manual cleaning, especially if cleaning solutions are recycled. A properly designed, validated, and operated CIP system insures consistent and reproducible cleaning. Because there is little or no dismantling of equipment, the down time is reduced and more time is available for productive use of machinery. Consistent cleaning eliminates product contamination and associated rejection of batches. As another important consideration, the CIP technology enhances safety by eliminating or minimizing operator contact with hazardous cleaning agents, bioactive products, and potentially pathogenic biohazard agents. Other unsafe situations are avoided because worker entry into equipment is not required.

A safety lock-out mechanism must prevent leakage of detergent/disinfectant and foodstuffs during dehydration of food. Modern industrial detergents and disinfectants are often strongly acid or alkaline, and if consumed in error in sufficient quantity, could give rise to chemical poisoning. All CIP systems should be supplemented periodically by the more conventional strip-down methods of plant cleaning, with all couplings open and gaskets removed. Chemical disinfectants should be changed at intervals to prevent the development of microbial resistance.

In roller drying, used for manufacturing baby food products, hygiene hazards will occur owing to partly dried materials from the ends of the drier drum leaving the drum and entering the main dry-product collecting stream. This is more noticeable when 'outboard' dam plates bear on the ends of the drum and feed rolls, rather than inboard. To overcome this a reciprocating doctor knife has been used to insure the discharge of a continuous dried-product sheet, in combination with short end and face scrape which remove the partly dried material before it reaches the doctor knife and the dried-product stream. The applicator rolls are manufactured from SS316 stainless steel, and the main drum is hard-chrome-plated to insure that hygienic operating conditions are maintained.

1. Where slurries are fed to the drier, steps should be taken to prevent survival of microorganisms by increasing the overall heat treatment so that, apart from the heat exposure on the drum itself, the slurries are preheated before transferring to the applicator rolls. For thick slurries containing starch, this can also be achieved by using a scraped-surface heat exchanger or a steam injector.
2. Other slurries for roller drying, which are at temperatures optimum for microbial growth, are made up with ice at the milling stage to keep the slurries at low temperatures in the holding tanks before processing.

Dried foods are not always sterile. Some microorganisms and particularly bacterial spores will survive the dehydration process. There will be no growth until the addition of water for reconstitution and consumption. Spices, dehydrated soups, baby foods, and other convenience-dried foods may all contain spores, usually in small numbers. For example, the addition of spices to meat will introduce spores and add to the possibility of *Bacillus cereus* and *Clostridium perfringens* food poisoning.

The principles of good hygiene practice in dehydrated food factories are basically the same as in any other food-handling environment. The scale of food manufacture, and the unique processes developed to produce shelf-stable preserved foods for storage, distribution, and retail sale provide greater opportunities for food-poisoning agents to become hazardous. The most important factor for consideration is that attention to the cleanliness of plant surfaces alone is not enough; proper design and control of the complete process are essential to insure the absence of food-poisoning agents.

Food Hygiene in Dehydrated Food Manufacture

Dehydration reduces the moisture content to levels below which microorganisms cannot grow. Dehydrated foods, although not sterile, may be stored

Table 1 Relationship between water activity and growth of organisms

Water activity	*Foods*	*Predominant microorganisms*
1.00–0.95	Most fresh foods, e.g., meat, fish, poultry, fruits, and vegetables	*Clostridium botulinum, Salmonella* Most normal bacteria
0.90–0.95	Many cured-meat products	*Staphylococcus aureus* (anaerobic)
0.85–0.90		Most normal yeasts, *S. aureus* (aerobic)
0.80–0.85	Very high sugar or salt content	Most normal molds
0.75–0.80		Halophilic bacteria
0.60–0.75	Dried foods	Osmophilic yeasts and molds

Reproduced from Alur MD and Venugopal V (1999) Dried foods. In: *Encyclopaedia of Food Microbiology*, pp. 530–537. London: Academic Press, with permission.

indefinitely because of their low water content. The bacteriological state of a product depends on the degree of contamination before dehydration. The growth and metabolism of microorganisms require water in an available form.

Dehydration is a method of food preservation that arrests the activity of foodborne microorganisms and retards chemical changes in the food. Microorganisms require water in order to grow; however the amount of moisture that permits growth varies with different kinds of microbes. The term water activity is used to designate the amount of moisture available in foods. Drying decreases the a_w of a food below the limits required for microbial growth. The a_w of any food is in the range 0–1.0. The relationship between a_w and growth of organisms is shown in Table 1.

Water activity in the range 0.995–0.980 is best for most bacteria and spoilage is encouraged in meat, fruit, and vegetables. At a_w 0.98–0.93 spoilage by Gram-negative bacteria gives way to spoilage by certain Gram-positive organisms; below 0.93–0.85 micrococci, yeasts, and molds can grow. Below 0.85–0.60 certain fungi and yeasts predominate in spoilage and below 0.60 there is no growth. The function of preservatives, such as salt and sugar, is to reduce the a_w; heat may also remove water. The reduced a_w is important to prevent outgrowth of spores.

It is very important to prevent the entrance into the drying chamber of organisms which are capable of producing toxins. This applies to the toxigenic forms of *C. botulinum* and certain strains of *Staphylococcus aureus*. Dehydrated foods should also be free from certain intestinal microorganisms that are likely to be pathogenic to humans, e.g., members of the genera *Salmonella* and *Shigella*. The bacterial counts of dried foods should be low to avoid development of undesirable flavors during the period of reconstitution. Microorganisms respond to reduced a_w levels. The effect of reduced a_w is characterized by an extension of the lag phase, a suppression of the log phase, and a reduction in the total number of viable microorganisms. The minimal a_w levels for growth and toxin production of a number of foodborne pathogens are shown in Table 2. Hence, the drying process should be extended to reach the particular a_w to eliminate growth of such microorganisms. Heat applied during a drying process causes a reduction in the total number of microorganisms, but the effectiveness varies with the kinds and numbers of organisms originally present.

Table 2 Effect of water activity (a_w) on growth and toxin production by some foodborne pathogens

Pathogenic microorganisms	*Minimum* a_w	
	Growth	*Toxin production*
Aspergillus flavus	0.82	0.78
Staphylococcous aureus	0.84	0.86
Clostridium botulinum type A and B	0.92	0.82
Listeria monocytogenes	0.92	
Bacillus cereus	0.93	
Vibrio parahaemolyticus	0.95	
Salmonella spp.	0.96	
Clostridium perfringens	0.96	

Reproduced from Alur MD and Venugopal V (1999) Dried foods. In: *Encyclopaedia of Food Microbiology*, pp. 530–537. London: Academic Press, with permission.

Usually, yeasts and most bacteria are destroyed but spores of bacteria and molds survive, together with a few vegetative cells of heat-resistant bacteria. More organisms are killed by freezing than by dehydration during the freeze-drying process. If the drying process and storage conditions are adequate, there will be no growth of microorganisms in the dried food. During storage, there is a slow decrease in numbers of organisms. The microorganisms that are resistant to drying will survive. In particular, resistant spores of bacteria, molds, and some of the micrococci survive. In conventional drying of vegetables, the microbial flora will be modified during drying. If the equipment is not clean, an increase in counts may be anticipated. Temperatures up to 90 °C are used during the first stage of drying. The rapid moisture loss from foods during this period induces a cooling effect, and thus the temperature of the product will be 40–50 °C. Hence, reduction in microbial counts is small. In the second stage of drying the product temperatures are

higher (60–70 °C), which may kill the yeasts and many other bacteria.

Mycotoxin formation is more sensitive to a reduction in a_w than is the growth of the mold forming it. To prevent the formation of bacterial toxins or the development of organisms pathogenic to humans, the food product should be dried at temperatures where growth is unlikely to occur. At temperatures of 50 °C, it is probably the minimum. Where some heating below this temperature is unavoidable because of loss of quality in the product, the period of dehydration should not exceed 4 h. Bacterial growth generally does not occur in foods containing less than 15% water. Foods at a_w levels below 0.78–0.80 are not hazardous. They also indicate that mold species are generally more osmotolerant than bacterial species, both in their ability to grow and produce toxins at low a_w levels. *Staphylococcus aureus*, the bacterial species capable of growth at the lowest a_w, has been exploited in some food regulations. Foods can be protected against microbial deterioration by keeping their a_w below a certain value. The approximate lower limits of a_w at which bacteria, yeast, and molds can grow are 0.91, 0.88, and 0.80, respectively. Most food-poisoning bacteria, including C. *botulinum*, *Salmonella*, and *Shigella*, are unable to grow or produce toxins below a_w of 0.95 in foods. Some aerobic sporeforming bacteria such as *Bacillus cereus* can tolerate much lower a_w (0.93) and the most xerotolerant species, like *B. subtilis*, are able to develop at a_w 0.90. *Staphylococcus aureus* is the most xerotolerant and toxin-producing bacterium. The limit of growth in laboratory media is 0.86, but enterotoxin A, which is frequently encountered in foods, is not produced below a_w of 0.88 at 25 °C. Such a difference between a_w levels permitting growth and those allowing the production of toxins is also frequent for mycotoxins produced by toxigenic molds.

As a guide to the storage stability of dried foods, the 'alarm water content' has been suggested. This is the water content which should not be exceeded if mold growth is to be avoided. Although drying stops growth, it does not kill all bacteria in the foods. The 'alarm water' contents for some foods are given in Table 3. These values are equivalent to an a_w of 0.7, at which point microbiological stability can be expected.

Table 3 The 'alarm water' content of some foods

Food product	*Water (%)*
Dehydrated fruits	18–25
Starch	18
Dehydrated vegetables	14–20
Pulses, milk powder, dehydrated meat	15
Rice, wheat flour	13–15
Dehydrated whole eggs	10–11
Whole milk powder	8

Reproduced from Alur MD and Venugopal V (1999) Dried foods. In: *Encyclopaedia of Food Microbiology*, pp. 530–537. London: Academic Press, with permission.

Several organizations are specialized in producing guidelines and standards for food-processing plants. Organizations in the USA and Europe attempt to harmonize hygienic design standards and guidelines, facilitating international trade. Specifically, the USA-based 3-A organization (3-A Sanitary Standards Committees, St Paul, Minnesota, USA), NSF International (NSF International, Ann Arbor, Michigan, USA), and the European Hygienic Equipment Design Group (EHEDG) (EHEDG Foundation, Brussels, Belgium) have established cooperation by exchanging staff and by participation in joint working groups.

See also: **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant; Overall Approach; Modern Systems; **Factory Construction**: Environmental Considerations; Materials for Internal Surfaces; **Food Poisoning**: Classification; **Meat**: Hygiene; **Metals Used in the Food Industry**; **Plant Design**: Basic Principles; Designing for Hygienic Operation; **Spoilage**: Bacterial Spoilage; **Water Activity**: Principles and Measurement; Effect on Food Stability

Further Reading

Adams D and Agarwal D (1988) Clean-in-place system design. *Biopharm* 2(6): 48–57.

Alur MD and Venugopal V (1999) Dried foods. In: *Encyclopedia of Food Microbiology*, pp. 530–537. London: Academic Press.

Chisti Y (1999) Modern systems of plant cleaning. In: *Encyclopedia of Food Microbiology*, pp. 1806–1815. London: Academic Press.

Curiel GJ, Van Eijk HMJ and Lelieveld HLM (1999) Risk control of airborne contamination. In: *Encyclopedia of Food Microbiology*, pp. 1816–1822. London: Academic Press.

EHEDG Guideline no. 1 (1993) Hygienic equipment design criteria. *Trends in Food Science Technology* 4: 225–229.

EHEDG Guideline no. 2 (1993) Welding stainless steel to meet hygienic requirements. *Trends in Food Science Technology* 4: 306–310.

EHEDG Guidelines no. 3 (1994) Hygienic design of equipment for open processing. *Trends in Food Science and Technology* 6: 305–310.

Gurakan GC and Bozoglu TF (1999) Designing for hygienic operation. In: *Encyclopedia of Food Microbiology*, pp. 1790–1794. London: Academic Press.

Hayes PR (1985) *Food Microbiology and Hygiene*. London: Van Nostrand.

Hobbs BC and Roberts D (1987) Food poisoning and food hygiene. In: Javitt R (ed.) *Hygienic Design and Operation of Food Plant*, Connecticut: AVI.

Hyde JM (1985) New developments in CIP practices. *Chemical Engineering Progress* 81(1): 39–41.
International Association of Milk, Food and Environmental Sanitarians. *3-A Sanitary Standards*. US Public Health Service Dairy Industry Committee.
International Commission on Microbiological Specifications for Foods (1996) *Microorganisms in Foods 5: characteristics of Microbial Pathogens*. Blackie Academic & Professional.
ISO/DIS14159. (1998) *Safety of Machinery. Hygiene Requirements for the Design of Machinery.*
Mostert MA and Lelieveld HLM (1999) Overall approach to hygienic processing. In: *Encyclopedia of Food Microbiology*, pp. 1802–1805. London: Academic Press.
Ramesh A (1995) CFTRI Annual Conference – CAC'95, pp. 49–54.
Robson JN and Barnes G (1980) *Hygienic Design and Operation of Food Plant*, pp. 121–124. Connecticut: AVI.
Stewart JC and Seiberling DA (1996) Clean in place. *Chemical Engineering* 102(1): 72–79.

Equipment Used in Drying Foods

J G Brennan, The University of Reading, Reading, UK

Drying in Heated Air

A wide range of equipment is used in air-drying foods, including the following.

Cabinet or Tray Drier

This is a batch drier consisting of an insulated cabinet fitted with a fan, an air heater, and baffles to direct the flow of the heated air between trays of food or through perforated trays and layers of the food material. Dampers control the intake of fresh air and the outflow of the exhaust air. These driers vary in size from small bench units, which house a few trays and are used to carry out laboratory studies, to large industrial installations which may take up to 25 large trays. In the case of industrial driers, the trays are usually loaded on to trolleys which are wheeled in and out of the cabinet. For large throughputs, a number of cabinets may be used in parallel with a staggered loading sequence. Such plants can handle several thousand kilograms of fresh vegetables in a 24-h period. Cabinet driers are used to dry sliced and diced vegetables and fruits, cubed or minced meat, and many other types of food.

Tunnel Drier

This type of drier can handle large quantities of vegetables and fruits on a continuous or semicontinuous basis. It consists of an insulated tunnel, up to 25 m long and 2 m × 2 m cross-section. Prepared food is spread in even layers on trays which are stacked on trolleys. When a trolley of fresh food is introduced into one end of the tunnel a trolley of dried material exits the other end. Heated air flows mainly between the trays, although some throughflow may occur. In a concurrent tunnel both the air and the fresh food enter the tunnel at the same end, the 'wet end,' and travel concurrently along the tunnel to exit at the 'dry end' (**Figure 1a**). This design of tunnel features high initial rates of drying. As the air moves along the tunnel its temperature falls and it becomes more humid. Consequently, the drying conditions at the dry end are poor and it may be difficult to attain low moisture contents. In the countercurrent tunnel (**Figure 1b**), the air is introduced at the dry end and travels countercurrent to the feed exiting at the wet end. The initial drying rate is lower than that in the concurrent design, but at the dry end the fresh, hot air meets the almost dry food. Hence, low moisture contents can be attained but there is a danger of heat damaging the food. Two tunnels may be used in series. The food first passes through a relatively short concurrent tunnel and then through a longer countercurrent tunnel. Thus, high initial drying rates can be achieved and, by using relatively cool but dry air in the countercurrent tunnel, low moisture contents can be attained without heat damaging the food. These features can be included in a single tunnel (**Figure 1c**), known as a center exhaust tunnel. In a cross-flow tunnel (**Figure 1d**), the air travels at right angles to the direction of flow of the feed. This facilitates uniform drying.

Conveyor or Belt Drier

In this type of drier the food is carried through a drying tunnel on a moving belt made up of hinged, perforated metal plates or wire mesh. The conveyor is 2–3 m wide and up to 50 m long. The drying air is usually directed through the belt and layer of food, which may be 75–150 mm deep. Usually the air travels upwards in the early stages of drying and downwards in the later stages. Such throughflow driers may only be used for solid particles that form a porous bed. The throughput per unit floor area is much less than that of a conventional tunnel drier. Using two or more conveyors in series, with solids being redistributed on to the following belt, can increase the throughput and result in more uniform drying, compared to a single-stage drier.

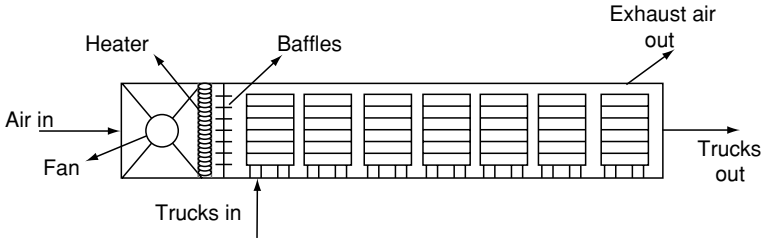

(a)

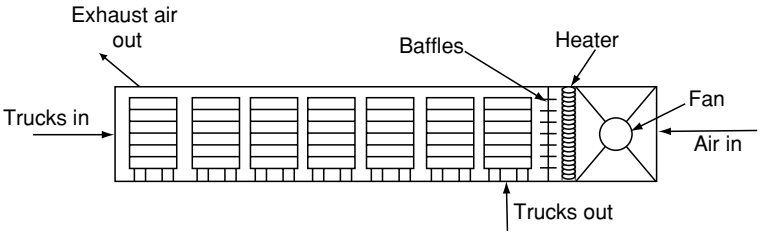

(b)

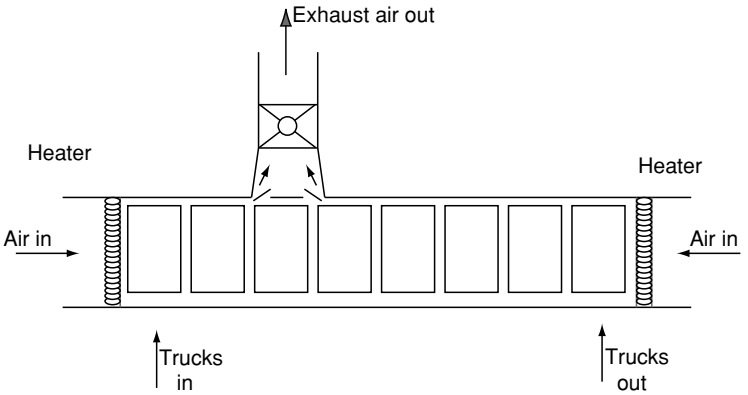

(c)

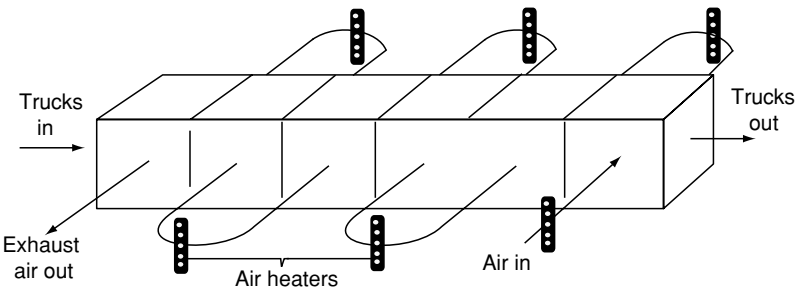

(d)

Figure 1 Tunnel driers: (a) concurrent; (b) countercurrent; (c) center exhaust, (d) cross-flow. Adapted from Brennan JG, Butters JR, Cowell ND and Lilly AEV (1990) *Food Engineering Operations*, 3rd edn. London: Elsevier Applied Science.

Nevertheless, it is a relatively expensive method of air-drying. Such a drier may be used to part-dry a food, e.g., diced vegetables down to 10–15% moisture. The final drying is carried out in less expensive equipment, such as a bin drier.

Bin Drier

This is a deep-bed, throughflow drier which is often used to finish off products that have been partly dried in tunnel or conveyor driers. It consists of a vessel fitted with a perforated base. Partly dried particles are placed in the bin to a height of up to 2 m. Dry air at relatively low temperature passes up through the bed of particles, at a relatively low velocity. Drying takes place over a period of up to 36 h. Moisture migration between particles can lead to a uniform moisture content in the product. Large bins are fitted with their own fan and heater. Smaller portable bins may be filled with partly dried particles direct form the first drier then moved and connected to a common air supply.

Fluidized-bed drier

This type of drier is being used increasingly for drying of foods. (*See* **Drying**: Fluidized-bed Drying.)

Pneumatic or Flash Drier

In this type of drier, food particles are fed into a fast-moving stream of heated air. They are carried in the air stream through ducting which is of sufficient length to give the required drying time. The dry particles are recovered from the air stream by a cyclone separator or filter. The ducting may be arranged vertically, known as air-lift driers, or horizontally. Relatively high air temperatures are used to reduce the length of the ducting. Expansion chambers may be incorporated into the ducting to facilitate control over the air temperature. Two or more of these driers may be used in series, with fresh air being introduced to each stage. To reduce the length of the ducting a closed-loop or ring drier may be used. The food particles circulate a number of times around the loop until they reach the required moisture content. They are then recovered from the air stream by a cyclone. There is a continuous supply of fresh air and feed to the drier and dried product is discharged continuously. Pneumatic driers are used to dry relatively small particles such as grains and flours and as after-driers on spray drier installations.

Rotary Drier

This consists of a cylindrical shell which rotates at an angle to the horizontal. The feed, consisting of particulate material, is introduced at the raised end of the shell and the product is discharged over a weir at the lower end. The shell is fitted with flights internally. As the shell rotates, the flights lift the material up and then it falls down through a stream of heated air, which may flow concurrently or countercurrently to the movement of the material. Rotary driers are not widely used for food dehydration. They are applied to grain drying and removing surface moisture from sugar and salt crystals.

Spray Drier

This type of hot-air drier is very widely used in the food industry for drying liquids and slurries. (*See* **Drying**: Spray Drying.)

Drying by Direct Contact with a Heated Surface

When a wet material is placed in contact with a heated surface, the necessary sensible and latent heat of evaporation is supplied to the material, mainly by conduction, and drying will take place. The general pattern of drying is similar to that of hot-air drying. There are two stages: an initial constant rate period followed by a falling rate period. However, if drying takes place at atmospheric pressure, the temperature attained by the material will be higher than that reached during hot-air drying, as the evaporative cooling effect will be less. To minimize heat damage, when drying at atmospheric pressure, the material is applied to the hot surface in a relatively thin layer. This results in short drying times, usually in the range 2–30 s. The drum drier is widely used for drying food liquids and pastes at atmospheric pressure. Alternatively, drying can be carried out at a pressure below atmospheric, i.e., vacuum drying, thus reducing the temperature at which drying takes place.

Drum, Film, or Roller Drier

A single-drum drier consists of a hollow metal cylinder rotating about a horizontal axis and heated internally, usually by steam. The feed is applied in a thin layer to the surface of the drum. As the drum rotates, drying takes place. The dried material is scraped off the drum surface by a knife some two-thirds to three-fourths of a revolution from the point of application of the feed. The drum may dip into the trough of feed material (Figure 2a) which boils on to the drum surface. This can lead to heat damage to the feed. Alternatively, two unheated rollers may apply the feed to the drum surface (Figure 2b). This method reduces the possibility of heat damage to the material. In the case of starchy materials, a number of rollers may be used to apply the feed (Figure 2c). The feed

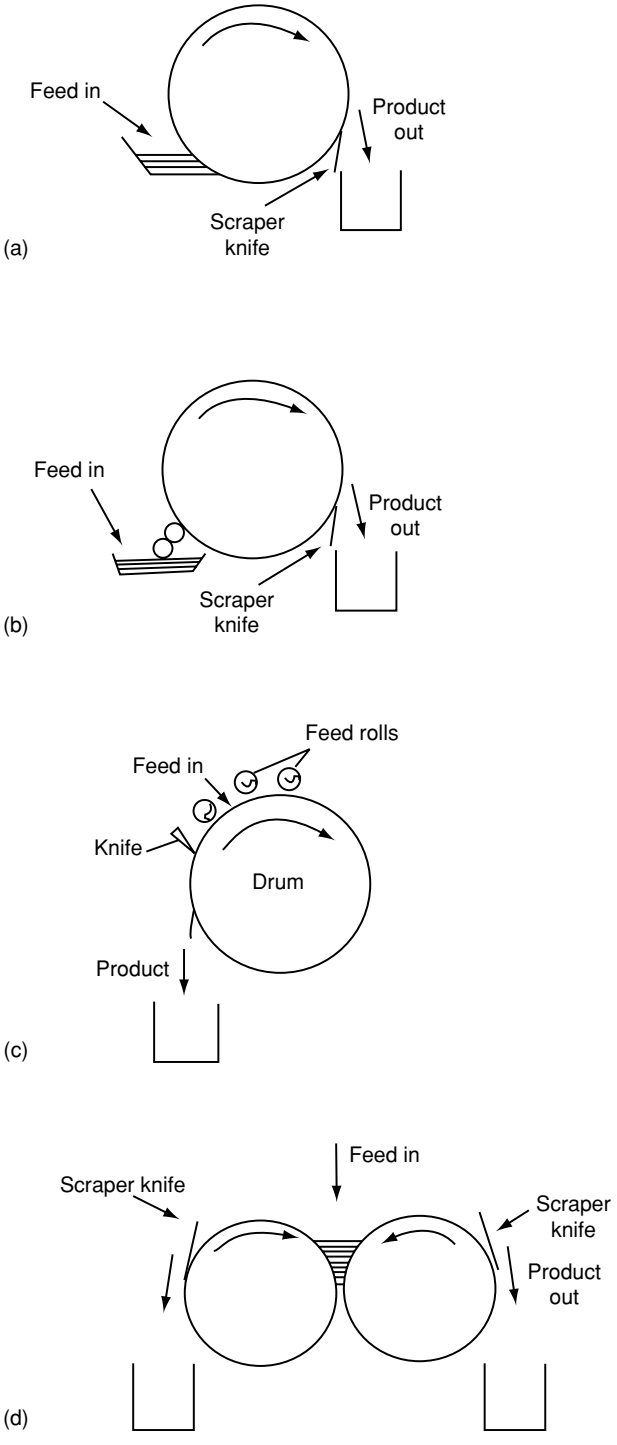

Figure 2 Drum driers – different feeding arrangements: (a) dip feed; (b) unheated roller feed; (c) multiroller feed; (d) double drum with trough feed. Adapted from Brennan JG (1994) *Food Dehydration – A Dictionary* and Guide. London: Butterworth-Heinemann.

is introduced between the first and second roller. The starch gelatinizes and the feed forms a rotating cylinder. Each of the other rollers applies a layer of the gelatinized material to the drum surface. Up to five rollers may be used.

A double-drum drier consists of two cylinders rotating towards each other at the top. The feed enters the trough between the rollers and boils on to both drums. The clearance between the drums may be adjusted to alter the thickness of the layer applied to the drums. This type of drum drier is used for low-viscosity materials that are not very heat-sensitive. The drum diameter ranges from 15 to 150 cm and length from 20 cm to 300 cm. The speed of rotation ranges from 3 to 20 rpm and surface temperature from 110 to 165 °C. They are usually made of chromium-coated cast iron or stainless steel. Potato flakes, instant breakfast cereals, made from an oat flour slurry, and ingredients for dried baby foods are some of the products made by drum drying. For very heat-sensitive materials vacuum-drum driers may be used. The drum is enclosed in a vacuum chamber and the pressure reduced to lower the temperature at which drying takes place. However, these driers have high capital and running costs and are not widely used in the food industry.

Vacuum Shelf or Tray Drier

This type of drier consists of a vacuum tight chamber, usually cylindrical and fitted with hollow shelves. These shelves are heated internally by steam, hot water, or some other thermal fluid which is circulated through them. The shelves typically measure 2.0×1.5 m and the cabinet may contain up to 24 of them. The food is placed on trays which sit on the heated shelves. The chamber is sealed and the pressure inside is reduced to 5–30 kN m^{-2} absolute, by means of a condenser and vacuum pump. The shelves are heated to 35–80 °C. This type of drier can be used for heat-sensitive liquid or solid foods. However, the capital cost, including the condenser and vacuum pump, is relatively high, and so are the maintenance costs. It is only used to dry materials that can bear the high costs of drying. It has been used for fruit juice concentrates, chocolate and malt-based drinks, chocolate crumb, yeast extract, and minced cooked meat for inclusion in ready meals or soup mixes.

Vacuum Band or Belt Drier

In this type of drier a continuous stainless-steel band or a belt made up of wire mesh or hinged metal plates moves clockwise over a heating and cooling drum, inside a vacuum chamber (Figure 3). Additional heat may be supplied by heated platens or radiant heaters. The feed, which must be a liquid or paste, is applied to the band by spraying or by means of a roller. The cooled product is scraped from the band by means of

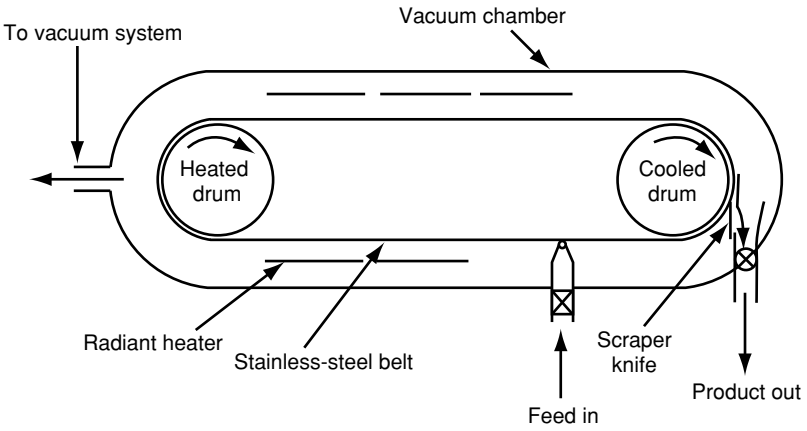

Figure 3 Vacuum band drier. From Brennan JG, Butters JR, Cowell ND and Lilly AEV (1990) *Food Engineering Operations*, 3rd edn. London: Elsevier Applied Science.

a knife and removed from the vacuum chamber via a sealed conveyor or rotary valve. The pressures and temperatures are similar to those used in the vacuum shelf drier. The vacuum band drier is costly to purchase and maintain and so is only used to dry relatively expensive materials, similar to those examples of liquid foods given above.

Drying by the Application of Radiant Heat

Apart from sun or solar drying (*See* **Drying**: Drying Using Natural Radiation), radiant energy is not used in industry to remove water in bulk from food materials. However, it has been used to remove small amounts of water from granular or powder materials.

Infrared Drier

In this type of drier a moving belt or vibrating deck conveyor carries the granular material beneath a bank of infrared heaters. Short-wave lamps are used for very heat-sensitive materials while long-wave bar heaters are used for less sensitive foods. Breadcrumbs, starches, tea, spices, cake mixes, and nuts are among the materials dried in this way.

Drying by the Application of Microwave Energy

(*See* **Drying**: Dielectric and Osmotic Drying.)

Freeze Drying

(*See* **Freeze-drying**: The Basic Process.)

Osmotic Drying

(*See* **Drying**: Dielectric and Osmotic Drying.)

See also: **Drying**: Drying Using Natural Radiation; Fluidized-bed Drying; Spray Drying; Dielectric and Osmotic Drying; **Freeze-drying**: The Basic Process

Further Reading

Brennan JG (1994) *Food Dehydration – A Dictionary and Guide*. London: Butterworth-Heinemann.

Brennan JG, Butters JR, Cowell ND and Lilly AEV (1990) *Food Engineering Operations*, 3rd edn. London: Elsevier Applied Science.

Greensmith M (1998) *Practical Dehydration*, 2nd edn. Cambridge: Woodhead.

Ducks *See* **Poultry**: Chicken; Ducks and Geese; Turkey

E

E-numbers *See* **Food Additives**: Safety

Eating Disorders *See* **Anorexia Nervosa**; **Bulimia Nervosa**

EATING HABITS

H L Meiselman and R Bell, US Army Natick Soldier Center, Natick, MA, USA

Background

Eating habits refer to the wide spectrum of attitudes and behaviors associated with eating. These areas are sometimes comprised in the fields of study referred to as food habits or food choice. Eating habits can be outlined as follows: food attitudes, food acceptance, food selection, food consumption, and food waste. This follows the temporal sequence of eating, beginning with the attitudes that lead to the selection and consumption of food, through to what is not eaten (i.e., wasted). This section on eating habits will follow this temporal sequence and will end with a treatment of meals and meals patterns, since, ultimately, food selection results in the construction of meals that combine all of the influences on eating habits.

Attitudes

Attitudes are related to whether people eat, what they eat, and how much they eat. Attitudes are usually defined as psychological tendencies containing an element of evaluation. The evaluative component is considered to be an important part of what constitutes an attitude. Attitudes are both long-lasting and short-lasting. The long-lasting attitudes are also called traits, which represent the ways in which people differ from one another and tend to be stable over time and over situations. Some of the attitudes that have been considered with respect to eating are these long-term traits. Expectations are the attitudes that products will possess certain sensory attributes at certain intensities or, more generally, the beliefs that products will be liked or disliked to a certain degree. Modern cognitive theories argue that the effective stimulus in any situation is not solely the physical stimulus, but rather is the relationship between the objective stimulus and the preexisting cognitive basis against which the stimulus is judged. This makes the area of food acceptance more complex, because we need to consider the response to the actual food as well as the expectation of that food.

Specific Attitudes and Traits

A number of measures, or scales, have been developed to assess attitudes and traits toward foods. By measuring these attitudes and traits, one can apply them in research studies to learn what factors influence them and, in turn, how they influence eating behaviors. The measures are also valuable for segmenting consumers into subgroups, in which different variables might operate, or in which the same variables might operate differently. These measures come from various fields

such as market research, social psychology, and abnormal psychology. Several important criteria are involved in the assessment of such measures. In addition to the usual criteria of validity and reliability, one must examine the nature and size of the sample on which the measure was developed and validated, whether the measure has been used to predict specific behaviors, and whether the measure has been tested on diverse populations. Below are some of the major scales being used.

The Variety Seeking Tendency Scale (VARSEEK) is a measure of the intrinsic tendency to seek variety in the diet and was developed by van Trijp and Steenkamp in 1992 as a special case of general variety seeking by consumers. The scale has eight items, each of which is rated on a five-point scale (disagree–agree). Individuals who score higher in variety seeking tendency have shown more exploratory tendencies in purchase behavior and more vicarious exploration, and use innovativeness more than those who score lower on the scale. Certain product characteristics, such as the sensory variation between alternatives being offered, may determine whether the intrinsic desire for variety is expressed when making a food selection.

The Food Neophobia Scale (FNS) is a measure of avoidance of novel foods, or foods to which people have not previously been exposed. It consists of 10 items, five positively worded and five negatively worded. Each is rated on a seven-point scale (agree–disagree). This measure has been shown to be a good predictor of an individual's propensity to avoid foods they have not tried previously. In addition, by presenting information that reduces the uncertainty about how a novel food might taste, those individuals with higher levels of food neophobia become more inclined to try that novel food. The Food Neophobia Scale has potential application within the area of new product development of nontraditional food items. Variety-seeking tendency and food neophobia tend to measure the opposites of food seeking and food avoiding; this is reflected in their negative correlation of approximately -0.7.

Dietary restraint refers to the degree to which one tries to control one's eating. It is not a measure of how well one controls eating, but whether one pays attention to eating in an attempt to control it. Dietary restraint has received attention by obesity researchers and others and can be measured by several different scales. Females usually exhibit more dietary restraint than males. The Restrained Eating Scale of the Dutch Eating Behavior Questionnaire consists of 10 items with a temporal response format (seldom, sometimes, often, very often). Stunkard and Messick (1985) also developed a dietary restraint scale as part of their three-factor eating questionnaire, and Herman and Polivy (1975) used a scale comprising a series of statements that allow for the classification of a subject as being either a dieter or a nondieter. Dietary restraint has been used extensively within obesity research, but has broad applicability within eating habits research and with product development aimed at the diet-conscious eater.

It has been shown that measures of variety-seeking tendency, food neophobia, and dietary restraint are stable over long periods of time, and are thus valuable indicators of a group's or individual's tendencies.

Food Preferences

Food preferences are the evaluative attitudes that people express toward foods. Food preferences include the qualitative evaluation of foods, and also how much people like and dislike them. Quantitative food preference measurement has been a part of the field of food habits at least since the 1940s, when the US Army undertook it for menu planning purposes. These studies were summarized by Meiselman in 1988. Of course, nonquantitative food preference measurement has been around since any mother asked their family what they wanted to eat. Food preferences are usually reported either on a hedonic scale ('how much do you like or dislike a food?') or on a preferred frequency scale ('how often do you want to eat a food?'). The nine-point hedonic scale, designed to measure food acceptability, is also used for food preference measurement. (*See* **Food Acceptability**: Affective Methods.)

Food preferences have been analyzed in terms of a number of demographic variables, including race, gender, geography, age, taste physiology, and many disease states. Differences in food preferences have been reported with differences in age, gender, race, and other variables. One of the early theories of obesity argued that obese people had different taste preferences than nonoverweight people. However, this has been shown not to be the case. While gender differences in food preferences have been known for some time, Drewnowski and colleagues have put these differences into a different context, arguing that males prefer salty foods, and females prefer sweet and fatty foods (1997).

One of the main differentiators of food preferences is culture. Elizabeth Rozin has argued that different food preferences are basically rooted in differing flavor principles among cuisines. Flavor principles are the unique flavor components of foods of differing cultures. For example, Greek cuisine is characterized by the combination of olive oil and lemon, whereas southern French cuisine is characterized by

olive oil and herbs. Japanese cuisine is characterized by the use of soy sauce. These traditional flavors help the consumer to identify culturally acceptable foods, and the incorporation of these flavor principles can be used to overcome neophobia to new foods.

Food preference measurement can be linked to menu planning. Beginning in the 1970s, modelers have calculated the contribution of various meal components to overall meal acceptability. This has usually been done with paper and pencil tests, rather than with actual foods and meals. The overall results of these studies show that that the main dish in a meal accounts for the major portion of a meal's acceptability, often 50% or more. Hedderly and Meiselman (1996) used acceptance ratings of actual college meals to conduct these regression analyses, showing that meal components change in their contribution to overall meal acceptance when the style of meal is changed from full plated meals to sandwich meals to pizza meals. Moskowitz and others have also modeled food combinations, boredom in menus, and time preference relationships in menus, trying to predict how much consumers would like an item if they had eaten the item recently or a long time ago.

Food Acceptance

Food acceptance, or the level of liking for a particular food, has been extensively researched and has been covered elsewhere in this volume. (*See* **Food Acceptability**: Affective Methods.) In this section, we will briefly discuss the role of food acceptance in influencing eating habits.

How much someone likes a food should be an important factor in influencing the selection and consumption of food, that is, eating habits. Few people would argue with this rational assumption. However, despite many studies of food acceptance, very few have shown a strong relationship of acceptance with either food choice or food consumption. This is because food consumption and food selection are extremely complex behaviors; complete models to predict food selection and food consumption might need dozens, hundreds, or even thousands of variables. Food acceptance is only one of these variables, although it is probably one of the most powerful. Another reason is that food consumption and food selection are not always rational behaviors. In general, people do not spend a large amount of time thinking about what to eat at every eating occurrence, and many other factors (availability of time and product) may be more instrumental in driving any one set of food selections. So, by itself, acceptance is merely a good measure of how much someone likes a food, and not a predictor of selection or consumption; but in conjunction with measures of other variables, acceptance can be part of a series of factors that operate together to control eating habits.

Food Selection

Food selection, or food choice, is the study of those factors that influence choice. Several fields of research have examined this relationship, including physiology, psychology, economics, and consumer behavior, to name a few. This chapter focuses on food selection as being a function of the interactive combination of the person, the product and the situation in which a food selection is made. Several authors have attempted to model the factors involved in the complex food selection or food choice process.

One difficulty in studying food selection is that it is nearly impossible to control for all factors that might influence choice in a real food choice setting, such as a restaurant or supermarket, or even in the home. Therefore, many studies have tried to explain food selection by using controlled studies in the laboratory. The problem is, once you take away the key elements that might come into play in a real setting, all of which tend to confound the control of a laboratory study, you are no longer examining the very variables that might be the most critical in affecting selection. There is no easy resolution to this controversy, but one approach is to do both laboratory and natural setting research, using the laboratory as a way of generating hypotheses and testing relationships that would identify salient factors influencing choice that could easily translate to real settings, then following up by introducing manipulations of these salient factors in a natural choice environment.

A number of variables have been shown to influence food choice. Habitual behaviors can influence food usage, in certain instances even more so than food acceptability. Familiarity with specific foods, obtained with prior experience, is a strong contributor to food acceptance and food selection, especially for people who are slightly neophobic. Attitudes and traits and expectations also influence acceptability and choice. Persistent negative expectations or stereotypes are important in understanding the critical evaluations of institutional foods and other foods that are regularly criticized such as airline food. In addition, several of the attitudes and traits described earlier (e.g., food neophobia, dietary restraint) also function to drive food choice behavior.

A number of environmental or situational variables have been shown to influence food selection. Increased effort required to obtain food leads to decreases in food selection of that item, and often increased selection of another item to take its place.

The presence of other people leads to increased consumption, probably because meals are longer when eaten with other people present. Food packaging, the information contained on the package or on the food shelf, as well as the placement of the food in relation to other products can influence selection. Finally, the décor, design, and visual and auditory elements of the food selection facility can influence choice.

Food Consumption

No other topic has occupied more research within eating habits than the research devoted to measuring food consumption. This seemingly simple question of how much people eat and why has still not been conclusively settled, leading to numerous books containing plans for healthy nutrition and dieting. Other sections of this book deal with the complexities of food-intake measurement.

Food acceptance or palatability has been shown to be a major determinant of eating habits. People eat larger meals when they are eating food that they like. This has been demonstrated in studies of individuals based on their own food preferences, and in studies in which the food has been manipulated to make it less acceptable. Palatability is a significant factor for meal studies involving both single courses and multicourse meals, as well as for sandwiches and snacks. The palatability effect has been studied in adults as well as neonates and children. Palatability not only affects what you eat, but how you eat. More palatable meals are eaten faster, and people take bigger bites and chew less time with more palatable foods.

Food Waste

Food waste, or the disposal of food, has received much attention in both the sociological and economic literature. In the former area, the study of waste is based on the notion that we are all involved in a world of goods in which we desire to be affiliated with a social structure that has us participating in the process of disposal of food in social and cultural contexts. In economics, the research issue is the cost of food waste in financial terms. In the context of eating habits that we have framed in this chapter, food waste becomes a measurement by which we can infer food consumption, and also a measure of food acceptance.

Food waste is sometimes measured while measuring consumption, because often consumption is measured as the difference between what is provided and what is wasted. The measurement of waste is difficult and time-consuming, because edible food waste must be separated from inedible food waste such as bones. This type of quantitative research has taken place in studies of clinical nutrition and in studies of military ration testing. In the former, subjects are offered foods in a laboratory or clinic. Meal and edible plate waste are analyzed using specialized instruments for estimating the nutrient content of the uneaten foods. Measures include moisture, crude protein, crude lipid, ash, and fiber content; total carbohydrates are then calculated by the difference between what was offered and what remains after analysis. In military ration testing, soldiers in the field are fed a measured and controlled amount of food for a particular meal, set of meals, or days. Other sources of food are controlled. By weighing the food prior to consumption and weighing the leftovers or waste, one can subtract to obtain the weight of the food eaten. Approximations of protein, fat, carbohydrates, micronutrients, or even particular foods, can be made based on the weight of the disposed food.

Meals and Meal Patterns

The patterns of eating in humans are a function of the combination of many, if not all, of the variables mentioned in this chapter. Determining exactly what factors influence meal patterns depends upon how one defines a 'meal' and a 'meal pattern.' This seemingly simple problem is complex because the definition of a meal depends on your perspective. The nutritionist sees meals as combinations of nutrients, the chef sees meals as combinations of courses and dishes, the social scientist sees meals as opportunities for individual expression and for social interaction, and so on. Following this, meal patterns have been defined in several ways, including the macronutrients or energy content in an individual's diet, the amount of food consumed per meal, the interval between meals, and the presence of social interaction.

One of the most interesting parts of the daily meal pattern is the breakfast, the first and smallest meal of the day, following the longest fast of the day. Although there is great variability from person to person in the types of breakfasts chosen, in general, breakfast is different in make-up from the other meals; midday and evening meals are often composed of meat or fish along with starch and vegetables. Breakfast is not similarly composed. Further, breakfast differs greatly from culture to culture. Breakfast ranges from the often-chosen roll and coffee in Southern Europe to the more traditional large cooked breakfasts of England and parts of America. Breakfast is also the meal most often skipped, and the meal most often eaten alone. One can question whether sipping a coffee while walking down the street constitutes a meal and a breakfast.

Breakfast is usually followed by midday and evening meals. Traditionally, many cultures consumed three hot meals per day; now many cultures eat only one hot meal per day. Cultures also vary in which meal of the day is considered the main meal, which is usually heavier and is usually served hot. In some cultures, the main meal is the midday meal, and in others, it is the evening meal. While some cultures have a more narrow definition of what constitutes a main meal (for example, the 'proper meal' of the English consisting of 'meat and two veg'), other cultures have a broader definition of main meals. The development and spread of pizza meals, sandwich meals, and salad meals is gradually changing how we view meals.

De Castro studied meal patterns extensively using a diet-diary technique, and extended this work to a cross-cultural comparison of France, The Netherlands, and the USA. While the study is limited to small numbers of university students, the results portray interesting differences and similarities across cultures. Compared with the French and Americans, who are more similar, the Dutch students consumed more food, had more meals with shorter intervals, ate with more people present, and ate for longer durations at a slower rate. The Americans tended to eat bigger meals as the day proceeded, a behavior that was even more pronounced in the Dutch, whereas the French ate their biggest meal in the early afternoon. However, the researchers concluded that the similarities among cultures are stronger than the differences, suggesting underlying mechanisms for food intake and meals. They also cautioned that differences in methods can lead to differences in results. The spread of the global economy is increasing the need to better describe and understand cultural differences in eating in order to identify opportunities for broader distribution of food products and food service.

There is also interest in continual snacking and meal-taking, often called grazing. A Symposium on Assessing Eating Patterns presented a cross-section of relevant research on meals across many cultures. One study reported snacks as a percentage of total eating in eight European countries: the numbers ranged from 6 to 32% for women and from 7 to 31% for men. Studies in both Greece and Germany showed midmorning, midafternoon, and evening snacking in addition to the regular three-meal pattern. In Germany, snacks accounted for 20% of energy intake, and Swedish data indicated even higher percentages of daily intake in the form of snacks. In a national sample of young adults conducted in the USA, snacking prevalence was shown to have increased significantly from 77 to 84% over the past 25 years. The nutritional contribution of snacks to total daily energy intake went from 20 to 23%, likely due to a 26% increase in energy consumed per snacking occasion and a 14% increase in the number of snacks taken per day. Intake of high-fat dessert snacks has decreased, but the intake of high-fat salty snacks has doubled during the same time period.

A final set of factors that have been studied for their effects on food choices and intake is the increased emphasis on health, diet, and slimness in the past decades, all of which have prompted meal patterns centered on restrained eating. Although restrained eating normally can produce short-term, often cyclic changes in meal patterns among otherwise healthy individuals, there is a danger that the desire for being thin can lead to eating disorders. But the dangers of these restrained eating patterns are not limited to individuals with eating disorders.

The field of human eating habits is very complex and involves many disciplines, only some of which have been covered in this section. The study of human eating habits is expanding rapidly within each discipline and also between disciplines, as the connections between different aspects of human eating habits are being investigated. In addition, the relationship of eating habits to health and fitness is receiving more and more attention, and eventually we will come to understand more fully how eating habits affect general well-being.

See also: **Dietary Surveys**: Measurement of Food Intake; Surveys of Food Intakes in Groups and Individuals; **Food Acceptability**: Affective Methods; Market Research Methods

Further Reading

DeCastro JM, Bellisle F, Feunekes GIJ, Dalix A-M and deGraaf C (1997) Culture and meal patterns: A comparison of the food intake of free-living American, Dutch and French students. *Nutrition Research* 17: 807–829.

Drewnowski A (1997) Taste preferences and food intake. *Annual Review of Nutrition* 17: 237–253.

Furst T, Connors M, Bisogni CA, Sobal J and Flak LW (1996) Food choice: a conceptual model of the process. *Appetite* 26: 247–266.

Hedderley D and Meiselman HL (1996) Modeling meal acceptability in a free choice environment. *Food Quality and Preference* 6: 15–26.

Herman CP and Polivy J (1975) Anxiety, restraint, and eating behavior. *Journal of Abnormal Psychology* 84: 66–72.

Marshall D (ed.) (1995) *Food Choice and the Consumer*. Glasgow: Blackie Academic and Professional.

Meiselman HL (1988) Consumer studies of food habits. In: Piggott JR (ed.) *Sensory Analysis of Foods*, 2nd edn, pp. 267–334. London: Elsevier.

Meiselman HL (ed.) (2000) *Dimensions of The Meal: The Science, Culture, Business and Art of Eating*. Gaithersberg, MD: Aspen.

Meiselman HL and MacFie HJH (eds) (1996) *Food Choice Acceptance and Consumption*. Glasgow: Blackie Academic and Professional.

Oltersdorf U, Schlettwein-Gsell D and Winkler G (1999) Assessing eating patterns – an emerging research topic in nutritional sciences: introduction to the symposium. *Appetite* 32: 1–7 (followed by symposium papers pp. 8–112).

Pliner P and Hobden K (1992) Development of a scale to measure the trait of food neophobia in humans. *Appetite* 19: 105–120.

Shepherd R (1989) Factors influencing food preferences and choice. In: Shepherd R (ed.) *Handbook of the Psychophysiology of Human Eating*, pp. 3–24. Chichester, UK: Wiley.

Stunkard AJ and Messick S (1985) The three-factor eating questionnaire to measure dietary restraint, disinhibition and hunger. *Journal of Psychosomatic Research* 29: 71–83.

Van Strien T, Frijters JER, Bugers GPA and Defares PB (1986) The Dutch eating behavior questionnaire for the assessment of restrained eating, emotional and external eating behavior. *International Journal of Eating Disorders* 5: 293–315.

Van Trijp HCM and Steenkamp J-BEM (1992) Consumers' variety seeking tendency with respect to foods: measurement and managerial implications. *European Review of Agricultural Economics* 19: 181–195.

Zizza C, Siega-Riz AM and Popkin BM (2001) Significant increase in young adults' snacking between 1977–1978 and 1994–1996 represents a cause for concern! *Preventive Medicine* 32: 303–310.

Eels *See* **Fish**: Introduction; Catching and Handling; Fish as Food; Demersal Species of Temperate Climates; Pelagic Species of Temperate Climates; Tuna and Tuna-like Fish of Tropical Climates; Demersal Species of Tropical Climates; Pelagic Species of Tropical Climates; Important Elasmobranch Species; Processing; Miscellaneous Fish Products; Spoilage of Seafood; Dietary Importance of Fish and Shellfish; **Fish Farming**; **Fish Meal**

EFFLUENTS FROM FOOD PROCESSING

Contents

On-Site Processing of Waste

E L Stover and C K Campana, Stover & Associates, Inc., Stillwater, OK, USA

Introduction

Biological waste water treatment is the primary method of preparing food-processing waste water flows for return to the environment. Increasing waste water loads on existing plants and more stringent government discharge requirements have put considerable pressure on the food-processing industry to refine and understand better the design and management of biological waste water treatment processes. Though activated sludge and other biological treatment processes are still frequently operated by general guidelines and 'rules of thumb,' facility design and operation must be guided by consideration of both the physical and biological aspects of waste water treatment. Various modifications and combinations of aerobic and anaerobic biological treatment processes are commonly used in the food-processing industry.

Biological Treatment Processes

Biological treatment processes, whether aerobic or anaerobic, are typically divided into two categories: suspended-growth systems and fixed-film systems. Combined or coupled fixed-film and suspended-growth systems are also commonly used for food-processing waste water treatment. These processes can be batch, semicontinuous, or continuous-flow processes. Suspended-growth systems are more commonly referred to as activated sludge processes, of which several variations and modifications exist. The basic system consists of a large basin into which the contaminated water is introduced, and air or oxygen is introduced by either diffused aeration or mechanical aeration devices. The microorganisms are present in the aeration basin as suspended material. After the microorganisms remove the organic material from the contaminated water they must be separated from the liquid stream. This is normally accomplished by gravity setting. After separating the biomass from the liquid, the biomass increase resulting from synthesis is wasted and the remainder is returned to the aeration tank. Thus, a relatively constant mass of microorganisms is maintained in the system. The performance of the process depends on the recycle of sufficient biomass. If biomass separation and concentration fail, the entire process fails. (*See* **Effluents from Food Processing**: Microbiology of Treatment Processes.)

Another type of suspended-growth system is the aerobic thermophilic treatment process. High-temperature and/or high-strength waste streams as well as waste biological sludges produced during mesophilic biological treatment are excellent candidates for aerobic thermophilic digestion. This process uses natural digestion to stabilize wastes in the 50–70 °C temperature range. A combination of mixing intensity control and oxygen injection rate are utilized to control the reaction rate, oxygen transfer rate, heat loss, and resultant bulk liquid temperature. The air (oxygen) is injected at the bottom of the enclosed reactor vessel, flows up through the bulk liquid, and is collected in the head space of the reactor. Major advantages include significant improvement over mesophilic treatment relative to chemical oxygen demand (COD) removal and volatile solids destruction at significantly lower hydraulic and solids retention times. Total solids reductions of 50–60% have been observed with waste biological sludges. The general requirements to maintain appropriate thermophilic temperatures include sufficient biodegradable organics (COD and/or volatile suspended solids) to provide heat of oxidation, an insulated reactor, and adequate mixing and oxygen transfer efficiency to minimize excessive heat loss.

Fixed-film biological processes differ from suspended-growth systems in that microorganisms attach themselves to a medium that provides an inert support. Biological towers (trickling filters), rotating biological contactors, and anaerobic reactors are the most common forms of fixed-film processes. Biological towers are a modification of the trickling filter process. The medium, which is normally comprised of polyvinylchloride (PVC), polyethylene, polystyrene, or redwood, is stacked into towers which typically reach 4.6–6.1 m (16–20 ft) high. The contaminated water is sprayed across the top and, as it moves downward, air is pulled upward through the tower. Forced air ventilation is also commonly used to insure adequate air flow and oxygen transfer from the gas phase to the bulk liquid. A slime layer of microorganisms forms on the medium and removes the organic contaminants as the water flows over the slime layer.

A rotating biological contactor (RBC) consists of a cylinder arrangement of plastic media, connected by a shaft, set in a basin or trough. The contaminated water passes through the basin where the microorganisms, attached to the media, metabolize the organics present in the water. Approximately 40% of the media surface area is submerged. This allows the slime layer to come into contact with alternately the contaminated water and the air where oxygen is provided to the microorganisms.

The combined or coupled fixed-film and suspended-growth processes offer many advantages for treatment of medium to high-strength food-processing waste waters when compared to use of either technology alone. The coupled process typically requires smaller land areas, and typically will have lower initial construction costs than a process consisting of a suspended-growth system alone. Electrical power costs for coupled systems are typically lower compared to comparable systems using only the suspended-growth process. Process control is usually less demanding because the roughing treatment provided by the fixed-film system helps dampen the organic loading variations on the suspended-growth reactor. When coupled systems are employed for nitrification, much of the organic loading will be removed in the fixed-film system, allowing the suspended-growth system to achieve stable, reliable nitrification.

There are several types of anaerobic reactors that incorporate into the design a relatively high sludge retention time (SRT) to account for the slow cell growth rate associated with anaerobic organisms. Anaerobic filters (AF), upflow anaerobic sludge blankets (UASB), and the Hybrid upflow sludge bed filter

process (Hybrid) are three anaerobic configurations commonly used to achieve and maintain high SRTs.

The AF reactor designs employ certain types of inert support material on which microbial biofilms develop as substrate passes through the reactor. This design enables these processes to operate at very long biomass retention times. Support or carrier material varies in size, shape, density, and placement according to the type of reactor design. The generic term 'anaerobic filter' usually refers to a fixed-film reactor. These processes use large support material such as rocks or plastic media designed for the purpose of forming a static bed in the reactor vessel. Waste flow through such media beds may be upflow, downflow, or horizontal. Other fixed-growth anaerobic systems employ much smaller media with high recycle rates of treated waste water in order to maintain the bed in a constant state of motion within the liquid medium in the reactor. Fluidized-bed reactors use relatively heavy support media such as sand or activated carbon and high recycle rates to keep the bed in a completely fluidized condition. Expanded bed reactors employ lighter material such as diatomaceous earth or ground corncobs and lower recycle rates to maintain the bed in a slightly suspended state. Both systems offer advantages in avoiding clogging, a major problem of the fixed-film anaerobic filters. Expanded beds can also handle suspended solids in the waste feed more readily. Fluidized and expanded beds, with their smaller media, provide much greater specific surface area for biofilm attachment. This means that much higher volumetric loading and conversion rates are possible than with static anaerobic filters, at the cost of additional pumping for recycle and somewhat complex operational requirements.

UASB reactors are the most common high-rate anaerobic treatment process. The process works on the principle of promoting sludge granulation by proper seeding with a granular inoculum of anaerobic bacteria, and adjusting organic loading and upflow feed rates so that biomass is retained as a dense blanket with a clear liquid zone above. Thus, the system acts similar to a biologically active sludge clarifier-thickener. The UASB is operated as a single-pass system, although waste sludge can be settled and returned. The exact biological or physical mechanisms controlling sludge granulation are still poorly understood, even though the UASB process is in successful commercial use. Start-up, however, can be facilitated by inoculating the reactor with well-granulated sludge from an operating system.

The Hybrid process has the advantages of both the upflow AF and the UASB processes while minimizing the disadvantages. The UASB has upward flow of waste water through a dense bed of active granular sludge with good settling properties, and then flows through a less dense blanket of suspended flocs. The waste water leaves the reactor via a solids/liquid/gas separation device at the top of the reactor. This system can retain high biomass concentrations (and high SRTs), but is limited by the settling properties of the granular sludge.

Biological Treatment Process Design

Prior to the last decade, and even today, design engineers have employed 'rule of thumb' techniques for the design of secondary biological waste water treatment processes for food-processing waste waters. Design of many biological systems has been based on empirical criteria derived from past experiences and working plants. Typical design criteria include detention time and biochemical oxygen demand (BOD) loading per unit volume. The need for employment of design models which attempt to describe the functional behavior of biological treatment processes has been recognized. Six of these models for the activated sludge process, which are identified according to the names of the investigators who had major roles in their development, are:

1. Eckenfelder
2. McKinney
3. Lawrence and McCarty
4. Gaudy
5. Weston
6. Stover and Kincannon

The aim of all of these design models is to provide more accurate predictive equations which are in keeping with the underlying metabolic and biological principles governing the waste water purification process. These modern concepts for design utilize various mathematical models which describe relationships governing microbial growth and substrate utilization or waste water purification. The two essential elements of these models are quantitative description of the properties of the growing biomass (the biological kinetic constants) and manipulative operating (engineering) variables. Methods have been developed to obtain the essential biokinetic constants from bench and pilot-scale investigations.

The mechanisms of waste water purification are the same for the suspended-growth activated sludge processes and fixed-film media biological filters and rotating biological contactor type of processes; however, there are physical differences that must be understood and properly evaluated for effective design. The state-of-the-art for design of these fixed-film processes has also been significantly advanced over the past few years. Initial design concepts

considered hydraulic flow rate or waste water organic concentration. More advanced approaches consider soluble substrate removal per unit surface area or contact time based on the assumption of first- or second-order reaction rate kinetics throughout the system. Still other approaches consider the relationship of treatment efficiency versus total organic loading over a wide organic loading range which allows determination of the proper design, considering variability in waste water characteristics as well as effluent quality requirements. Based on these kinetic design approaches, performance data, operating information, and kinetic design data can be obtained from bench and pilot-scale fixed-film systems for design of full-scale systems in a manner similar to activated sludge systems. These secondary biological kinetic design approaches advance the state of the art for consideration of the microbiological and biochemical aspects of biological treatment, as well as consideration of the engineered or operational control variables.

Scale-up Design Considerations

Variations in waste water characteristics, environmental conditions, and biological system performance must be evaluated to design properly a biological waste water treatment system to achieve the allowed effluent criteria reliably. Variability in waste water characteristics (flows and pollutant concentrations) and variability in biological treatment performance are two of the most important considerations during scale-up and design of food-processing waste water treatment facilities. It is extremely important during the experimental planning to develop waste water survey and biological treatability programs for collection of sufficient data to evaluate the variability and statistically analyze the data for design. The waste water characterization requires knowledge of the discharge characteristics due to production schedules and clean-up operations and development of sufficient characterization data for determining the design conditions, for example at the 50 and 90% probable flows and loadings. Consideration of the data variability from the biological treatability study due to the dynamic nature of the heterogeneous microbial populations is more complicated and in most cases has not been properly addressed during design of food-processing waste water treatment facilities.

Modern concepts used for designing biological waste water treatment processes for food-processing waste waters employ mathematical models that describe the relationships governing microbial growth and substrate utilization (aerobic or anaerobic processes). These modeling approaches require the development of biokinetic descriptive constants that can be developed from treatability studies. Stover and Kincannon have recently developed a mathematical model for both fixed-film and suspended-growth aerobic and anaerobic biological systems that eliminates the problem of variability in the biological descriptive substrate utilization constants. The problems with the scatter observed in the development of biokinetic coefficients utilized in other mathematic models have been eliminated by expressing the mass-substrate utilization rate as a function of the mass substrate loading rates. The specific substrate utilization rate (U) is plotted as a function of the food-to-microorganism (F/M) ratio in activated sludge systems and as a function of the mass loading rate in $\mathrm{kg\,day^{-1}}$ $1000\,\mathrm{m^{-2}}$ in fixed-film systems in order to determine the biokinetic constants of $U_{\max}$ and K_B. These constants are then utilized in the mathematical model, and the remaining variability in influent flow rate and influent substrate concentrations can be handled effectively.

Fixed-film Reactor Design

In the Stover–Kincannon model for fixed-film reactors a substrate material balance is conducted around the fixed-film system. The specific parameters of concern for concept design and operational process control for a biotower fixed-film process are shown in Figure 1, and the definition of each of these parameters is presented in Table 1. When considering a fixed-film reactor volume, a mass balance of substrate into and out of that reactor volume can be made as follows:

$$\begin{array}{l}\text{Mass of}\\ \text{substrate}\\ \text{into the}\\ \text{reactor}\end{array} = \begin{array}{l}\text{Mass of}\\ \text{substrate}\\ \text{out of the}\\ \text{reactor}\end{array} + \begin{array}{l}\text{Mass of}\\ \text{substrate}\\ \text{consumed}\\ \text{biologically}\end{array}$$

In the case of the fixed-film reactor, the reactor volume is expressed in terms of thousand cubic meters of media volume. Based on this approach, the mass balance of substrate into and out of the fixed-film reactor volume can be expressed as follows:

$$FS_i = FS_e + \left[\frac{dS}{dt\ A}\right]_G A \qquad (1)$$

where: F = flow rate ($\mathrm{m^3\ s^{-1}}$); S_i = influent substrate concentration ($\mathrm{mg\,l^{-1}}$); S_e = effluent substrate concentration ($\mathrm{mg\,l^{-1}}$); A = surface area of volume ($1000\,\mathrm{m^{-2}}$); and $\left[\frac{dS}{dt\ A}\right]_G$ = specific substrate utilization rate ($\mathrm{kg\ 1000\,m^{-2}\ day^{-1}}$).

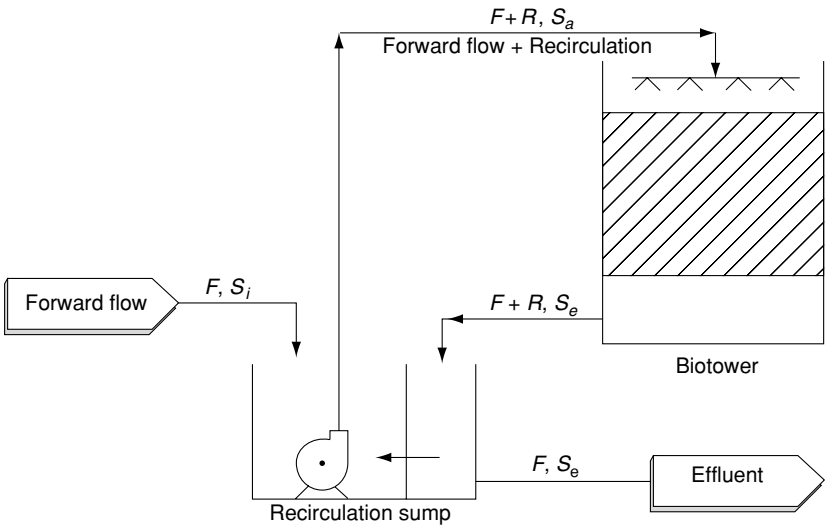

Figure 1 Basic process flow diagram for biotower system with operating parameters.

Table 1 Definition of biotower process parameters

F	Forward flow rate ($m^3\ s^{-1}$)
R	Recirculation rate ($m^3\ s^{-1}$)
$F + R$	Total flow to the biotower ($m^3\ s^{-1}$)
S_i	Substrate concentration in the forward-flow stream ($mg\ l^{-1}$)
S_e	Substrate concentration in the biotower effluent ($mg\ l^{-1}$)
S_a	Substrate concentration in the actual biotower influent ($mg\ l^{-1}$)
d	Biotower diameter (m)
h	Biotower media height (m)
A_v	Specific surface area of media ($m^2\ m^{-3}$)
n	Rotational speed of the distributor ($rev\ min^{-1}$)
N	Number of arms in rotary distributor assembly

The surface area represents the active mass of microorganisms. By plugging the following substrate utilization rate term developed by Stover and Kincannon,

$$\left[\frac{dS}{dt\ A}\right]_G = \frac{U_{max}(FS_i)/A}{K_B + (FS_i)/A} \tag{2}$$

into eq (1), the resulting expression can be solved for the required surface area of media for design, as follows:

$$A = \frac{FS_i}{\dfrac{U_{max}\ S_i}{S_i - S_e} - K_B} \tag{3}$$

where: U_{max} = maximum specific substrate removal rate ($kg/1000\,m^2\ day^{-1}$); K_B = proportionality constant ($kg/1000\,m^2\ day^{-1}$); and $(FS_i)/A$ = specific substrate loading rate ($kg/1000\,m^2\,day^{-1}$).

The expression can also be solved for the effluent substrate concentration that would correspond to a specific design loading and amount of media volume available, as follows:

$$S_e = S_i - \frac{U_{max}\ S_i}{K_B + (FS_i)/A} \tag{4}$$

Eqn (4) can then be used to evaluate system performance at existing facilities and project expected performance for future increased loadings. Eqn (3) can be used to determine design requirements to handle additional loadings.

The biokinetic constants, U_{max} and K_B, can be determined by linearizing eqn (2) or taking the reciprocal as follows:

$$\frac{1}{\left[\dfrac{dS}{dt\ A}\right]_G} = \frac{K_B}{U_{max}} \times \frac{1}{\dfrac{FS_i}{A}} + \frac{1}{U_{max}} \tag{5}$$

where:

$$\left[\frac{ds}{dt\ A}\right]_G = \frac{F(S_i - S_e)}{A} \tag{6}$$

Now, the reciprocal of the specific substrate utilization rate ($F\,(S_i - S_e)/A$) is plotted against the reciprocal of the specific substrate loading rate (FS_i/A). In this linearized form, the y-axis intercept is equal to $1/U_{max}$, and the slope of the line is equal to K_B/U_{max}.

Activated Sludge Reactor Design

In the Stover–Kincannon model for activated sludge reactors, a substrate material balance is conducted around the aeration basin and the secondary clarifier, in a similar manner to that for the fixed-film reactor

model. The specific parameters of concern for concept design and operational process control for activated sludge systems are shown in Figure 2, and the definition of each of these parameters is presented in Table 2. When considering an aeration basin volume, the mass balance of substrate into and out of that reactor volume can be expressed as follows:

$$\begin{matrix}\text{Mass of substrate into the reactor}\end{matrix} = \begin{matrix}\text{Mass of substrate out of the reactor}\end{matrix} + \begin{matrix}\text{Mass of substrate consumed biologically}\end{matrix}$$

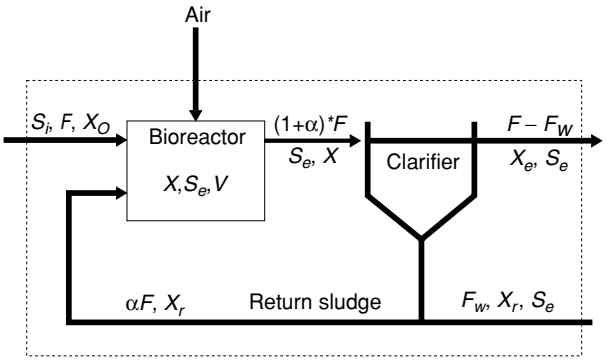

Figure 2 Flow diagram of the activated sludge process showing notation and mass balance envelope.

Table 2 Definition of activated sludge parameters

F	Influent flow rate ($m^3\ s^{-1}$)
S_i	Influent substrate concentration ($mg\ l^{-1}$)
X_o	Influent solids concentration ($mg\ l^{-1}$)
S_e	Effluent substrate concentration ($mg\ l^{-1}$)
X	Biological solids concentration, MLVSS (mixed liquor volatile suspended solids) ($mg\ l^{-1}$)
V	Aeration reactor volume (m^3)
X_e	Effluent suspended solids concentration ($mg\ l^{-1}$)
X_R	Recycle solids concentration ($mg\ l^{-1}$)
F_w	Waste sludge flow rate ($m^3\ s^{-1}$)
α	Recycle ratio
αF	Recycle flow rate ($m^3\ s^{-1}$)
U_{max}	Maximum specific substrate utilization rate ($kg\ kg^{-1}\ day^{-1}$)
K_B	Substrate loading rate at which the rate of substrate utilization is one-half the maximum rate ($kg\ kg^{-1}\ day^{-1}$)
Y_t	True cell yield
K_d	Endogenous decay rate
Y_{obs}	Observed sludge production yield ($kg\ kg^{-1}$ substrate removed)
F/M	Food-to-microorganism ratio ($kg\ kg^{-1}\ day^{-1}$)
U	Specific substrate utilization rate ($kg\ kg^{-1}\ day^{-1}$)

Based on this approach, when considering an activated sludge reactor volume, a mass balance of substrate into and out of the activated sludge reactor volume can be made as follows:

$$FS_i = FS_e + \left[\frac{dS}{dt}\right]_G V \tag{7}$$

where: F = influent flow rate ($m^3\ s^{-1}$); S_i = influent substrate concentration ($mg\ l^{-1}$); S_e = effluent substrate concentration ($mg\ l^{-1}$); V = reactor volume (m^3); and $\left[\frac{dS}{dt}\right]_G$ = specific substrate utilization rate ($kg\ kg^{-1}\ day^{-1}$).

By plugging the following substrate utilization rate term developed by Stover and Kincannon into eqn (7), the resulting expression can be solved for the required aeration basin volume for design, as follows:

$$\left[\frac{dS}{dt}\right]_G = \frac{U_{\max} X(FS_i)/XV}{K_B + FS_i/XV} \tag{8}$$

$$V = \frac{FS_i/X}{\dfrac{U_{\max} S_i}{S_i - S_e} - K_B} \tag{9}$$

where: $U_{\max}$ = maximum specific removal rate ($kg\ kg^{-1}\ day^{-1}$); K_B = proportionality constant, or substrate loading at which the rate of substrate utilization is one-half the maximum rate ($kg\ kg^{-1}\ day^{-1}$); FS_i/XV or F/M = specific substrate loading rate ($kg\ kg^{-1}\ day^{-1}$); and X = biological solids concentration ($mg\ l^{-1}$).

The expression can also be solved for the effluent substrate concentration that would correspond to a specific design loading, mixed liquor sludge concentration, and amount of reactor volume available, as follows:

$$S_e = S_i - \frac{U_{\max} S_i}{K_B + FS_i/XV} \tag{10}$$

Eqn (10) can then be used to evaluate system performance at existing facilities and project expected performance for future increased loadings. Eqn (9) can be used to determine design requirements to handle additional loadings.

The biokinetic constants, $U_{\max}$ and K_B, can be determined by linearizing eqn (8) or taking the reciprocal as follows:

$$\frac{1}{X}\frac{1}{\left[\frac{dS}{dt}\right]_G} = \frac{K_B}{U_{\max}} \times \frac{1}{\frac{FS_i}{XV}} + \frac{1}{U_{\max}} \tag{11}$$

where:

$$\left[\frac{dS}{dt}\right]_G \times \frac{1}{X} = \frac{F(S_i - S_e)}{XV} \tag{12}$$

Now, the reciprocal of the specific substrate utilization rate ($F\,(S_i - S_e)/XV$) is plotted against the reciprocal of the specific substrate loading rate ($FS_i / XV = F/M$). In this linearized form, the *y*-axis intercept is equal to $1/U_{max}$, and the slope of the line is equal to K_B/U_{max}. Once the biokinetic constants have been determined, the required *F/M* ratio at any given loading (F and S_i) can be determined to meet the required effluent quality reliably (S_e). The biokinetic constants, which represent the treatability characteristics of the waste water, can be determined experimentally for any given activated sludge process. These constants can be developed from pilot plant or laboratory-scale studies, as well as, from the full-scale plant data.

Aerobic Thermophilic Process Design Considerations

Evaluations, both bench-scale and pilot-scale operating systems, of aerobic thermophilic reactors have shown that these systems comply with the same types of monomolecular kinetic relationships previously described. Resultant data analyses have shown that the relationships of substrate removal versus substrate applied were applicable to aerobic thermophilic systems. This mathematical model has been used extensively in the design and operation of aerobic thermophilic treatment systems. It is important to note that the kinetics of microbial thermophilic growth exceeds the growth of mesophiles and higher substrate utilization rates increase the amount of substrate that can be effectively treated in a given reactor.

Primary design considerations for aerobic thermophilic biological treatment systems include the following:

- temperature
- aeration and mixing
- matching oxygen transfer with oxygen demand
- sufficient biodegradable organics
- food-to-microorganism ratio
- solids retention time
- kinetics of substrate removal and biomass growth
- heat loss management
- foam control

The thermophilic reactor and mixing and oxygen transfer system are designed for minimal heat loss, maximum mixing intensity, and maximum oxygen transfer capacity in order to optimize and control the reaction temperature. The reactor is designed to minimize heat loss at a variable operating liquid depth by minimizing the cross-sectional surface area to volume ratio. Then with proper reactor material of construction and insulation, the reactor temperature can be controlled by varying the liquid depth, the liquid mixing flow rate, and the air flow rate.

When considering liquid waste water treatment, the following factors are important relative to maintaining autothermal conditions:

- COD concentration/loading
- (F/M) ratio
- oxygen balance (COD balance)
- 12 700–13 200 kJ kg^{-1} O_2

The organics or COD goes to new biomass and carbon dioxide. The primary objective of thermophilic treatment is to convert the majority of the organics to carbon dioxide and heat energy while minimizing the biomass production. Nitrification is also inhibited above about 45 °C, such that excess oxygen requirements are not observed for nitrification. Excess ammonia-nitrogen reacts with the carbon dioxide to produce ammonium bicarbonate alkalinity to assist with reactor pH control.

When considering aerobic thermophilic treatment (stabilization) of waste biological sludges the following equation provides valuable insight into the reaction process:

$$C_5H_7O_2N + 5O_2 \rightarrow 5CO_2 + NH_3 + \text{Energy} + \text{Heat} \qquad (13)$$

By understanding this simple biochemical relationship and the following factors, maintaining autothermal conditions and process control can be related to the following factors:

- oxygen (COD) balance (1.42 K_gO_2/K_g VS (volatile solids) destroyed)
- VS concentration/loading
- VS destruction
- VS balance
- Nitrogen balance
- 18 800–19 200 kJ kg^{-1} VS destroyed

Anaerobic Process Design Considerations

In addition to controlling the loading rate, the mesophilic anaerobic microbes require an optimum temperature (30–35 °C), a pH range of 6.7–7.2, adequate nutrients, and no toxic organic or inorganic compounds. Certain trace inorganic nutrients and vitamins are needed in anaerobic treatment processes. Trace metals such as nickel, iron, cobalt, molybdenum, selenium, and tungsten are stimulatory to methanogens at low concentrations. It is extremely important to monitor and control these parameters during start-up and operations of anaerobic reactors.

In addition, adequate amounts of calcium and magnesium are required to promote the growth of sludge granules which are prevalent in the anaerobic sludge blanket and hybrid processes.

The most effective indicators of performance are pH, volatile acid to alkalinity ratio, and gas (methane) production. The key to successful start-up is to increase the biological solids inventory by gradual increases in the flow and organic loading to the plant. It is critical that the loadings are increased in proportion to the biomass in order to control the F/M ratio. This aspect of both plant start-up and stabilized plant operations can be monitored by the system pH, volatile fatty acids, volatile fatty acid/alkalinity ratio, COD removal efficiency, methane content of the gas, and reactor temperature. Adjustments are required when any of these parameters exceed preset acceptable operating ranges.

Generally, the volatile acid/alkalinity ratio should be maintained at less than 1.0 in order to obtain optimum system performance. When this ratio exceeds 1.0, immediate corrective action(s) should be taken, otherwise the volatile acids will cause a drop in the pH and the methanogen activity will decrease. The methane production is a direct measure of the metabolic activity of the methanogenic bacteria and, as such, has great potential as a diagnostic tool of anaerobic activity. The use of pH alone is not as sensitive an indicator of reactor activity because environmental changes will already have taken place before a pH change is noticed. However, knowledge of the pH is important to good operation of the system and should be maintained at approximately 6.7–7.2 for most applications.

Process Control Requirements

In many cases in the past, biological treatment processes for food industry wastes have been designed and operated by general guidelines or rules of thumb. However, with increasing treatment requirements, it becomes necessary for treatment facilities to be designed and operated more effectively to meet the required effluent quality while continuing to treat varying waste water flows and contaminant loads. Operators must be better trained and educated so that they are better equipped to monitor the biological system and make the proper corrections for changing conditions. The treatment facility to be operated must be designed with the greatest amount of flexibility consistent with the design conditions for that particular food-processing plant. This requires the design engineer to rely on experience and operational feedback from full-scale operations, as well as on parameters and observations obtained from laboratory and pilot plant operations. After transforming this operational experience into a treatment plant design, the design engineer should be responsible for proper training of the operators and for assisting in the facility's start-up.

Changing environmental conditions, especially fluctuations in the waste water characteristics, tend to disrupt the steady-state conditions that the biological treatment facilities are designed to approach. Environmental changes tending to disrupt steady-state conditions (shock loads) which cannot be, or have not been, smoothed by preventive engineering expedients must be accommodated solely by successful biological response or by combined biological and engineering remedial responses. To cope with these situations successfully requires thorough understanding of the process by the plant operational personnel and the incorporation of adequate operating flexibility into the system by the design engineer.

The key to maintaining operational process control and stable operations in biological treatment systems is to provide proper environmental conditions to the biomass or bacteria in the system. The hydraulic flow rate and organic loading rate, along with the variability in these parameters, are two of the most critical parameters relative to maintaining stable operating conditions; of course, pH, temperature, nutrients, dissolved oxygen, and lack of toxic or inhibitory substances are also critical to successful operations. When properly monitoring and controlling these parameters, the key to successful operations of activated sludge systems then becomes matching the number of microorganisms in the system to the organic substrate loading rate to the system, or controlling the F/M ratio of the system.

Operational Process Considerations

Kinetic approaches to successfully applying the required metabolic and biological principles of food-processing waste water treatment for both design and operations have been discussed herein. The kinetic constants $U_{\max}$ and K_B define the biological treatability characteristics of the waste water. These biokinetic constants are normally defined in terms of COD, BOD, and/or total organic carbon (TOC) for food-processing waste waters. They are used for design purposes and operational process control and decision making when properly defined. They provide significant operations value by denoting changing qualitative waste water characteristics, toxicity incidents, and refractory or residual nonbiodegradable organic substances. For high-strength carbohydrate waste waters without toxicity or refractory organics, $U_{\max}$ and K_B are very close to one another in terms of

BOD, COD, and TOC with values typically above 15 kg kg^{-1} day^{-1} in terms of COD. High residual/refractory organics or toxicity will affect both the *y*-axis intercept (U_{max}) and the curvature (K_B) of the specific substrate utilization rate (U) versus the specific substrate loading rate (F/M ratio). Both refractory organics and toxicity reduce U_{max}, and change K_B significantly from U_{max} numerically. Therefore, proper monitoring and use of these kinetic constants can provide an invaluable tool for changing water quality and maintaining stable operational process control of biological systems treating food-processing waste waters.

See also: **Effluents from Food Processing**: Microbiology of Treatment Processes

Further Reading

Allen JD, Sass DW and Stover EL (1986) *Full-scale Operation of an Activated Sludge Plant Treating Corn Wet Milling Waste Water*, pp. 505–510. Proceedings of the 41st Industrial Waste Conference. West Lafayette, IN: Lewis Publishers Inc.

Eckhoff TH and Stover EL (1993) *Anaerobic Treatment of Confectionery Waste Water.* Proceedings of the 1993 Food Industry Environmental Conference. Atlanta, GA: Georgia Tech Research Institute.

Kincannon DF and Stover EL (1982) Design methodology for fixed-film reactors – RBCs and biological towers. *Civil Engineering for Practicing and Design Engineers* 2: 107.

Stover EL (1980) Start-up problems at a plant treating food processing waste water. *Journal of the Water Pollution Control Federation* 52: 249.

Stover EL (1988) *Reliable Design and Operation of Anaerobic Treatment Systems for the Food Processing Industry.* Proceedings of the 1988 Food Processing Waste Conference. Atlanta, GA.

Stover EL (1998) *Anaerobic Treatment Troubleshooting*. Proceedings of the WEFTEC 98 Workshop; Industrial Waste Treatment Operations. Orlando, FL: Georgia Tech Research Institute.

Stover EL (1998) *Industrial Waste Residuals Stabilization by Aerobic Thermophilic Biological Treatment*. Proceedings of the 30th Mid-Atlantic Industrial and Hazardous Waste Conference. Philadelphia, PA: Water Environment Federation.

Stover EL (1999) *Aerobic Autoheated Thermophilic Treatment Process for High Strength Industrial Waste Residuals*. Proceedings of the WFF/Purdue Industrial Wastes Conference. Indianapolis, IN: Water Environment Federation.

Stover EL and Campana CK (1990) *Operational Process Control for High Carbohydrate Filamentous Bulking.* Proceedings of the 1990 Food Industry Environmental Conference and Exhibition. Atlanta, GA: Georgia Tech Research Institute.

Stover EL and Gonzalez R (1988) *Low Temperature Kinetics of Anaerobic Fixed Film Reactors*. Proceedings of the 43rd Industrial Waste Conference. West Lafayette, IN.

Stover EL, Gonzalez RA and King KW (1995) *Biotower Design Concepts for Pretreatment of Food Processing Waste Waters*. Proceedings of the Seventh International Symposium on Agriculture and Food Processing Wastes. Chicago, IL: American Society of Agricultural Engineers.

Microbiology of Treatment Processes

E L Stover and C K Campana, Stover & Associates, Inc., Stillwater, OK, USA

Biological Treatment

During biological treatment, the objective is to remove or reduce the concentration of organic and inorganic compounds by using microorganisms. With biological treatment, the organic material in the waste water is removed by the microorganisms through metabolic processes. The organic compounds may be used by the microorganisms to form new cellular material or to produce energy that is required to sustain the microorganisms.

Heterotrophic microorganisms are the most common group of microorganisms providing the metabolic process for removing organic compounds from contaminated waste water. Heterotrophs use the same substances (organic compounds) as sources of both carbon and energy. A portion of the organic material is oxidized to provide energy, while the remaining portion is used as building blocks for cellular synthesis. Three general methods exist by which heterotrophic microorganisms can obtain energy: fermentation, aerobic respiration, and anaerobic respiration.

In the case of fermentation, the carbon and energy source is broken down by a series of enzyme-mediated reactions, which do not involve an electron transport chain. In aerobic respiration, the carbon and energy source is broken down by a series of enzyme-mediated reactions in which oxygen serves as an external electron acceptor. Anoxic respiration is similar to aerobic respiration, whereby facultative heterotrophic microorganisms use nitrate or nitrite-nitrogen as the final electron acceptor in the absence of oxygen. In anaerobic respiration, the carbon and energy source is broken down by a series of

enzyme-mediated reactions in which carbon dioxide and sulfates serve as the external electron acceptors. These three processes of obtaining energy form the basis for the various biological waste water treatment processes.

Biological treatment processes whether aerobic or anaerobic are typically divided into two categories: suspended growth systems and fixed-film systems. These processes can be batch, semicontinuous, or continuous flow processes. Suspended growth systems are more commonly referred to as activated sludge processes, of which several variations and modifications exists. Fixed-film biological processes differ from suspended growth systems in that microorganisms attach themselves to a medium that provides an inert support. Biological towers (trickling filters) and rotating biological contactors are the most common forms of fixed-film processes. (*See* **Effluents from Food Processing**: On-Site Processing of Waste.)

Activated Sludge

The activated sludge process is the most commonly employed treatment process for the degradation of organics in waste water from the food-processing industry. This article will review activated sludge microbiology, identification techniques of filamentous microorganisms, nonfilamentous microbial problems, higher life forms, and current activated sludge control measures.

Activated Sludge Microbiology

The applicability of biologically treating a particular waste water is a function of the biological degradability of the dissolved organic chemical constituents present in the waste water. The degradation rate of a specific organic compound is a function of the molecular structure of that particular compound, the genera and species (type) of microorganisms utilizing it as a food source, and the time required for the microorganisms to develop the enzymes necessary for substrate utilization.

The basic mechanism for removal of organic materials from waste water by biological treatment processes can be represented by three chemical reactions occurring simultaneously: energy, synthesis, and endogenous respiration. Energy and synthesis are often referred together as oxidative assimilation, and involve the consumption by microorganisms of the organic materials present in the waste water. A portion of this organic material is used as fuel to supply energy for metabolism, while the remaining material provides building components resulting in the formation of new cellular material. Thus, the organic material can be considered as food or substrate consumed by the microorganisms, with the production of carbon dioxide.

In the initial growth phases of a batch-fed biological process when organic matter (food) is present, oxidative assimilation is the primary reaction; however, as the food supply diminishes, the organic matter within the sludge is utilized and results in endogenous respiration as the predominant reaction. In the continuous fed complete-mix activated sludge process, both oxidative assimilation and endogenous respiration occur simultaneously. The degree of oxidative assimilation compared with endogenous respiration is a function of the operating conditions. Higher organic loading rates favor oxidative assimilation, whereas lower organic loading rates favor endogenous respiration.

The reactions of energy, synthesis, and endogenous respiration can be represented by the following simplified equation (eqn (1)):

$$\text{Organic matter} + O_2 \xrightarrow{\text{Microorganisms}} \text{New microorganisms} + CO_2 + H_2O + \text{energy.} \quad (1)$$

This equation is a simplification of the biochemical equation (and of microbial growth); it is qualitative, since no numerical coefficients are included, and significant elements have also been omitted. Nitrogen, phosphorus, and trace quantities of other elements have been eliminated from the left side of the equation. Typically, the heterogeneous population of the active biomass in the activated sludge process consists of 95% bacteria and 5% higher life forms such as rotifers, nematodes, protozoa, etc. The main purpose of the biota in the mixed liquor is to remove (metabolize) the soluble organics in the influent waste water stream and create floc forming bacteria that will settle under gravity in the secondary clarifier, thus leaving a clear supernatant (effluent) that can be discharged.

Situations may arise where not all the bacteria become floc-formers, thereby causing a turbid final effluent. Certain waste water streams can contribute to dispersed bacterial growth, single-celled free-floating bacteria, and the proliferation of filamentous microorganisms. It must be noted that filamentous bacteria are not always the main culprit for bulking or rising sludge problems. In fact, some filamentous organisms can assist in settling by acting as a backbone, whereby floc forming bacteria can grow and attach, therefore, minimizing the potential for shearing of the floc. Also, the presence of some filaments can and do act as a 'catch all' filter for small particles

that do not settle in the final clarifier. Filamentous bulking most often occurs when filaments are present in large amounts, thereby causing sludge compaction and settling to be hindered in the secondary clarifier.

Filamentous Microorganisms

In most activated sludge processes, filamentous microorganisms are routinely observed and actually belong in the microbial population. Their presence in the activated sludge process may actually contribute to a good quality effluent as long as their abundance is held to a minimum. A massive growth of filamentous microorganisms most often leads to bulking problems in the final clarifier, thereby leading to the deterioration of effluent quality. At least 25–30 types of filamentous microorganisms have been found to exist in activated sludge with 10–12 types being the most predominant. Though much effort and research have been expended trying to classify and name all the filamentous microorganisms, the task has yet to be completed. Consequently, unknown filamentous microorganisms (due to a lack of specific genus names) are indicated by a four-digit number. The occurrence of various types of filamentous microorganisms has been correlated with waste water characteristics and operating conditions, as shown in Table 1. This type of information can be used as a diagnostic tool to aid in the identification of potential filamentous microorganisms and the possible reason for proliferation. The growth of filamentous microorganisms is similar to normal bacterial growth except for unicellular cell division. Separation of two new cells does not take place with filamentous organisms, and consequently filaments consist of chains, long or short, of cells. In these cells, normal cell division occurs, but the two cells stay together. However, the presence of a sheath (hollow outer structure) may inhibit cell separation; hence, a large chain of many cells can develop. Usually, the cross-walls between the cells within a filament can be observed when performing a microscopic examination.

Identification of Filamentous Organisms

When bulking of sludge occurs, it can be observed in the final clarifier. Prior to this, it can be microscopically evaluated whether the bulking is due to filamentous microorganisms and, if so, the specific type of filaments. A brief description of different morphological characteristics of filamentous organisms follows. For detailed assessment and identification, the manual *Manual on the Causes and Control of Activated Sludge Bulking and Foaming*, by Jenkins, Richard, and Daigger, is very helpful, or the computer-based rapid Filamentous Bacteria Identification Program 'FILIDENTPro™, Stover & Associates, Inc., which utilizes an enhanced algorithm and search tree that enables the user to identify filamentous organisms easily and reduces the potential for misdiagnosis, is a very good training and identification guide.

The size (length and diameter) and location of filamentous microorganisms vary greatly. Filaments may be long and stout, short and thin, protrude from the floc, be within the floc, be free-floating in the bulk solution, or various combinations of these, may have attached growth adhering to the sheath (Figure 1), and/or have an outer cylinder-shaped clear structure in which some filamentous bacterial cells are wrapped. However, not all filamentous bacteria have a sheath or even attached growth. The shape (straight (Figure 2), bent, coiled, mycelial, smoothly curved) of the filamentous microorganism must be noted, as well as the shape of the individual cells within the microorganism itself. Most often, the cells are rectangular or square, but the need to account for all possible shapes (oval (Figure 3), discoid, round-ended rods (Figure 4), barrel, coccus) is essential to know in order to correctly identify the filamentous microorganism. Other important physical characteristics that distinguish filamentous bacteria from one another are cell inclusions, the

Table 1 Dominant filament types indicative of activated biosolids operation problems

Suggested causative condition	*Indicative filament types*
Low DO	*Sphaerotilus natans, Haliscomenobacter hydrossis*, type 1701
Low *F/M*	*Microthrix parvicella, Haliscomenobacter hydrossis*, types 021N, 0041, 0675, 0092, 0581, 0961, and 0803
Nitrogen and/or phosphorus deficiency	*Sphaerotilus natans, Thiothrix* spp., type 021N; possibly *Haliscomenobacter hydrossis* and types 0041 and 0675
Low pH	Fungi
Septic wastes	*Thiothrix* spp., *Beggiatoa* spp., type 021N

From Richard MG, Jenkins D, Hao O and Shimizu G (1982) *The Isolation and Characterization of Filamentous Microorganisms from Activated Biosolids Bulking*. Report No. 81–2, Sanitary Engineering and Environmental Health Research Laboratory, University of California, Berkeley, with permission.

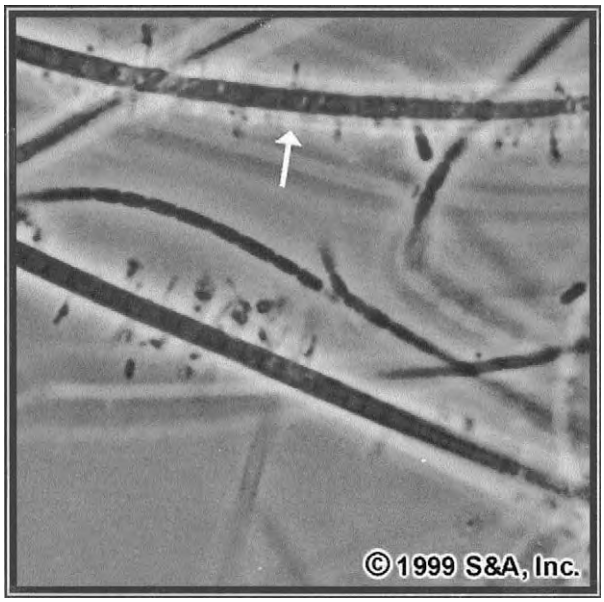

Figure 1 Attached growth of epiphytic bacteria on filamentous organism, Type 0914. Subject to copyright and patent protection of Stover & Associates, Inc., as part of the FIL-IDENTPro™ Filamentous Bacteria Typing Program.

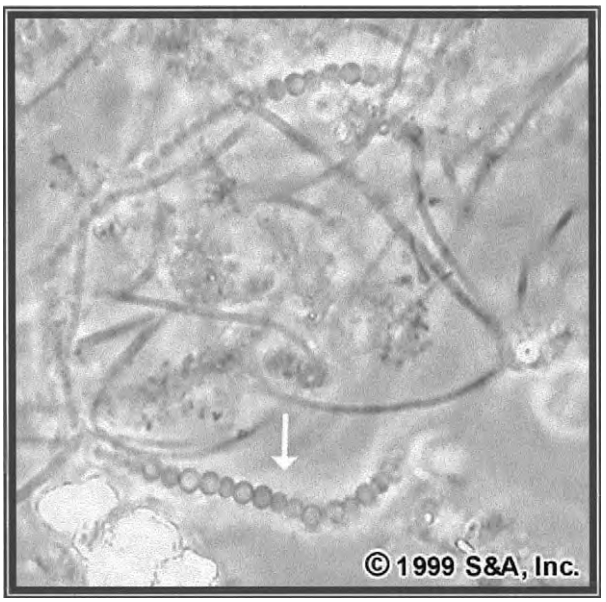

Figure 3 Oval cells resembling a pearl necklace. Filamentous bacteria, Type 1863. Subject to copyright and patent protection of Stover & Associates, Inc., as part of the FIL-IDENTPro™ Filamentous Bacteria Typing Program.

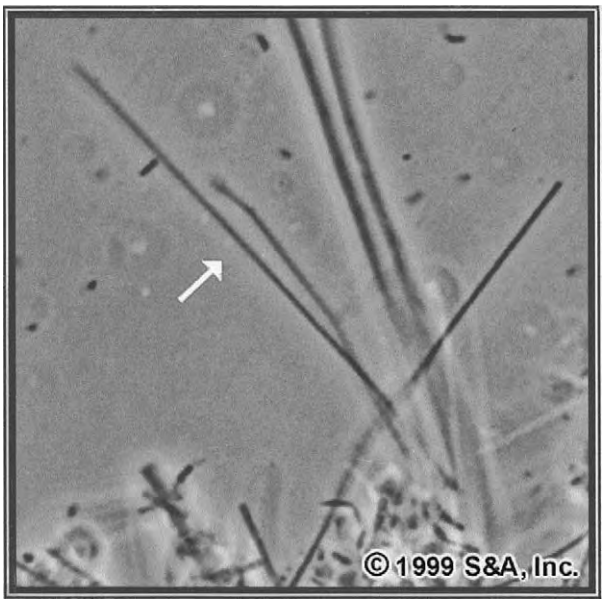

Figure 2 Filamentous bacteria, *Haliscomenobacter hydrossis*, indicating a rigidly straight trichome. Subject to copyright and patent protection of Stover & Associates, Inc., as part of the FIL-IDENTPro™ Filamentous Bacteria Typing Program.

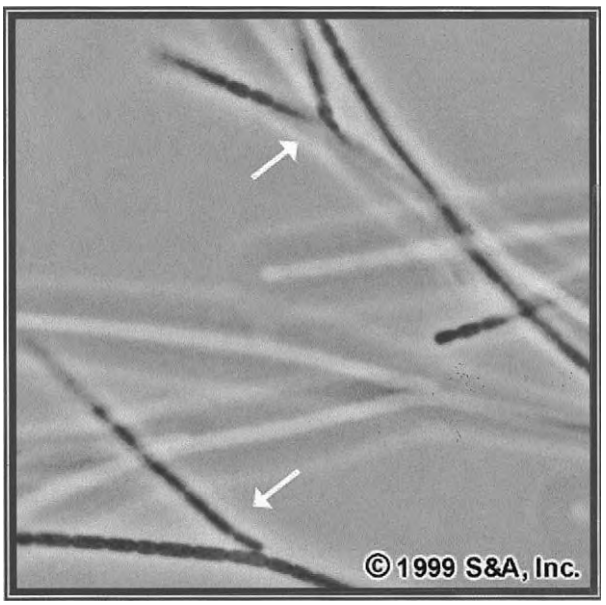

Figure 4 (see color plate 47) Round-ended rods. False trichome branching. Note the appearance of the trichome branches being 'stuck' together. False trichome branching does not have contiguous cytoplasm between the branches. Filamentous bacteria, *Sphaerotilus natans*. Subject to copyright and patent protection of Stover & Associates, Inc., as part of the FIL-IDENTPro™ Filamentous Bacteria Typing Program.

presence or absence of septa, constrictions of the outerwall, motility, rosettes formation, and whether any branching is true (Figure 5) or false (Figure 4). These physical characteristics can be determined microscopically by using a phase-contrast microscope at 1000×.

In addition to identifying the physical morphology, staining of the filamentous bacteria enhances the identification procedure. The Gram Stain and Neisser

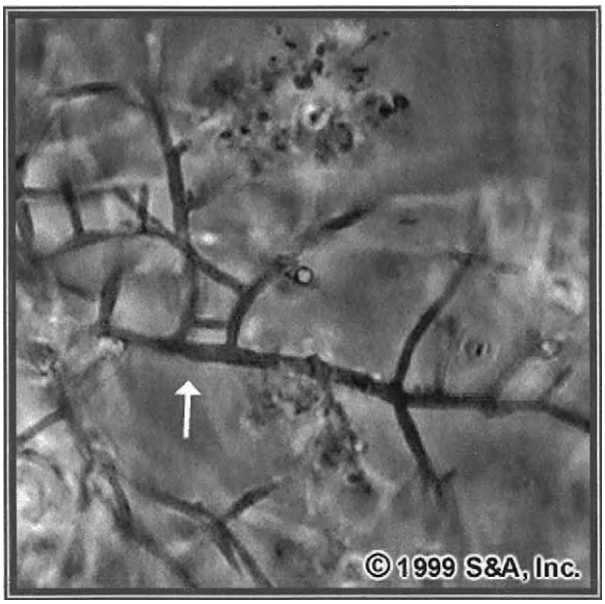

Figure 5 (see color plate 48) True trichome branching. True branching refers to contiguous cytoplasm between the branches. Filamentous bacteria, nocardia form. Subject to copyright and patent protection of Stover & Associates, Inc., as part of the FIL-IDENTPro™ Filamentous Bacteria Typing Program.

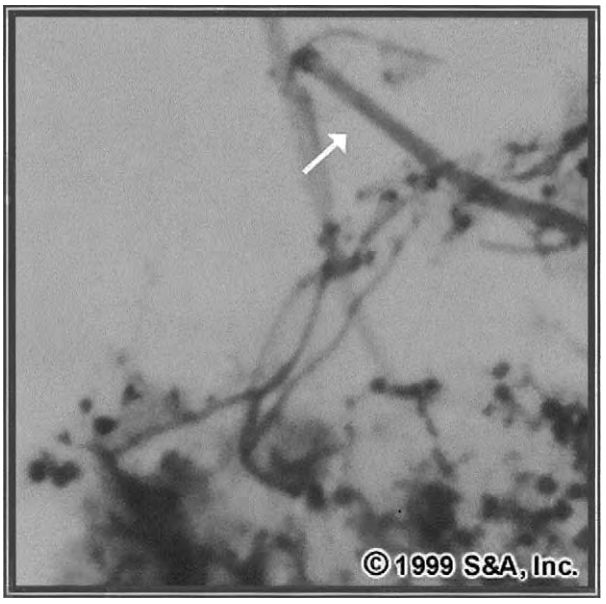

Figure 6 Neisser stain of filamentous bacteria. Blue–violet trichome indicates a positive reaction to the Neisser stain. Subject to copyright and patent protection of Stover & Associates, Inc., as part of the FIL-IDENTPro™ Filamentous Bacteria Typing Program.

Stain are used extensively in the bacterial characterization process. The results are expressed as Gram-positive or -negative and Neisser-positive (Figure 6) or -negative. Other specialized slide preparation tests include:

1. **Sulfur oxidation test** – Ability of the bacterium to store sulfur granules.
2. **Crystal violet stain** – Used for determining whether a filamentous microorganism contains a sheath.
3. **India ink reverse stain** – Indicates the presence or lack of exocellular polymeric material (extracellular polysaccharides).
4. **PHB stain** – Detects intracellular storage of polyhydroxybutyrate.

Microscopic examination of the Gram, Neisser, and PHB stains should be at 1000× magnification with direct illumination (bright field). The India ink reverse stain should be observed at 100× magnification with phase contrast. Crystal violet and the sulfur oxidation test should be viewed at 1000× magnification phase contrast. A general outline is presented in Table 2, indicating important morphological characteristics and stains for the proper identification or typing of filamentous bacteria.

Filamentous bacteria serve as indicator organisms that, if correctly identified, allow for proactive measures to be initiated, thereby reducing the potential

Table 2 Filamentous bacteria identification techniques

Morphology
- **I. Branching**
  1. True
  2. False
- **II. Motility**
- **III. Inclusions (granules)**
  1. Sulfur
  2. Polyhydroxybutyrate
- **IV. Septa**
- **V. Shape of filament**
  1. Straight
  2. Bent
  3. Coiled
- **VI. Attached growth**
- **VII. Constrictions**
- **VIII. Shape of cells**
  1. Spherical or coccus
  2. Rod
  3. Spiral
  4. Oval
- **IX. Sheathed**
- **X. Rosettes**

Stains
- **I. Neisser**
  1. Positive
  2. Negative
- **II. Gram**
  1. Positive
  2. Negative
- **III. Crystal violet**
- **IV. Polyhydroxybutyrate**
- **V. India ink reverse**
- **VI. Sulfur oxidation test**

for process failure. For example, identification of the filament, O21N (Figure 7) can indicate possible nitrogen-deficient operating conditions with high-strength food-processing waste waters.

Once filamentous bacteria become established, they possess a competitive advantage over other organisms relative to substrate utilization and growth. Hence, it is very important to identify the specific filament(s) that cause, or contribute to, bulking at each specific waste-water treatment plant. Early detection of site-specific problematic filamentous microorganisms often leads to heading-off of a bulking sludge condition before serious problems arise.

Nonfilamentous Microbial Problems

Though filamentous microorganisms play an important role in most activated sludge bulking episodes, nonfilamentous microbial growth can also cause negative impacts relative to proper settling and growth of the mixed microbiological culture in the activated sludge process. These include: dispersed growth, toxicity, floating sludge (denitrification) and nutrient deficiency.

1. **Dispersed growth** is generally characterized by single-celled bacteria in which floc development is negated and settling does not occur. This type of growth leaves behind a turbid, murky effluent. Most often, common dispersed growth problems are associated with industrial waste-water treatment facilities where the influent waste stream consists of soluble readily biodegradable organics operating at a high food-to-microorganism (*F/M*) ratio. Thus, the microorganisms are growing in a high log growth-rate phase that does not favor a floc forming settling sludge. Other possible sources of dispersed growth are nutrient (nitrogen and phosphorus)-deficient growth conditions and toxicity such as chronic heavy metals toxicity and associated deflocculation.
2. **Toxicity** or toxic shock loads can create severe operations problems in the activated sludge process. Toxicity implies reduced treatment efficiency and biological upset conditions that stem from the death of a microbe or metabolic impairment. Industrial waste streams often contain compounds, such as cleaning agents and/or surfactants that, in sufficient concentrations, can cause cell inhibition or, even worse, toxicity. Several microscopic bio-indicators can be employed to diagnose these types of problems, as follows:
   - loss or kill of higher life forms and/or protozoa;
   - biomass deflocculation often accompanied by dispersed growth and foaming in the aeration basin;
   - rapid increase in flagellate concentration and activity;
   - loss of biochemical oxygen demand $(BOD)_5$ removal; and
   - proliferation of filamentous organisms upon recovery from the process upset.

   Depending on the severity of the toxic load, the *F/M* ratio can be extremely high initially because of the bacteriological die-off, hence further development of dispersed growth. A spiked dissolved oxygen uptake rate of mixed liquor can be used to diagnose and detect toxicity especially in early stages. Waste-water additions that decrease the oxygen uptake rate by more than 4–5% generally indicate microbiological inhibition.
3. **Floating sludge** in the final clarifier is generally an indication that the aeration system is performing to the extent that the ammonia-nitrogen is being oxidized to nitrite- and nitrate-nitrogen (nitrification). The nitrite- and nitrate-nitrogen under anoxic conditions in the clarifier can be reduced to nitrogen gas and rise to the surface carrying the biological solids with it. This floating sludge condition can be generally corrected by increasing the return sludge rate, increasing the waste sludge rate, thus reducing the sludge inventory in the clarifier, by increasing the *F/M* ratio or, possibly, by adding specific inhibitors of the nitrification process to the system.

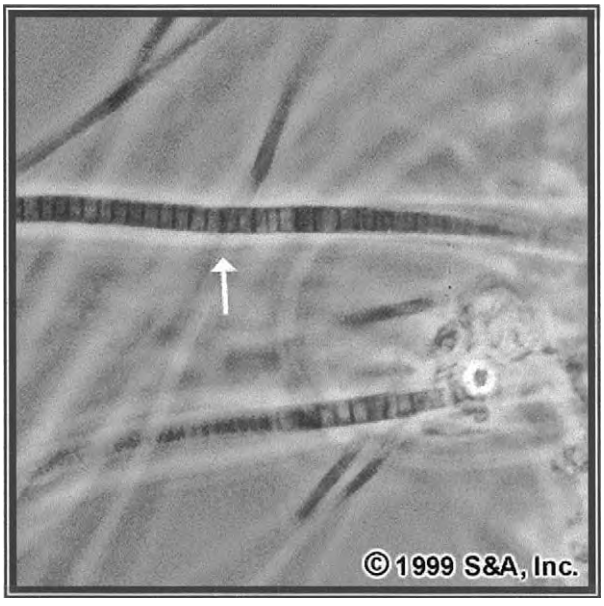

Figure 7 (see color plate 49) Filamentous bacteria type 021N. This slide was photographed at 1000× magnification under phase contrast. Subject to copyright and patent protection of Stover & Associates, Inc., as part of the FIL-IDENTPro™ Filamentous Bacteria Typing Program.

4. **Nutrient deficiency**: Micro- and macronutrient availability is extremely important in maintaining a healthy and active biomass in the aeration basin. The macronutrients, phosphorus and nitrogen, are typically more apt to be supplemented than the micronutrients. The need for the addition of micronutrients is rare because of the small concentrations required and the availability of these nutrients in most influent waste streams. Nutrient deficiency is usually associated with industrial waste-water treatment plants.

   The production of foam and the jelly-like consistency of the activated sludge mixed liquor are indicators of bulking problems associated with nutrient deficiency. The jelly-like (slime) consistency is a byproduct from the bacterial cell when production of necessary cell components such as protein is hampered because nitrogen is not available. The slime, more commonly called exocellular or extracellular polysaccharide, is a metabolic shunt product with surface-active properties that can lead to foaming in the aeration basin. Sludge grown inthis media consists of large globular floc forms that do not settle and compact in the clarifier. This condition can be observed microscopically by using the India ink reverse negative staining technique. Under normal conditions, the ink particles penetrate deeply within the floc, while penetration is blocked when large amounts of exocellular polysaccharides are present, resulting in a whitish 'ghost' appearance to the sludge (Figure 8). Certain filamentous organisms thrive under nutrient deficient conditions. The filaments most often related to nutrient deficiency are types 0041, 021N, 0675, and *Thiothrix* sp. Other filamentous bacteria may be present also; however, these filaments are typically indicators of nitrogen- or phosphorus-limiting conditions.

   The maximum theoretical amount of nitrogen and phosphorus required for biological synthesis is based on a BOD_5:N:P ratio of 100:5:1. This value can be increased or decreased, depending on the sludge retention time (SRT), *F/M* ratio, organic compounds being oxidized, etc. Maintaining soluble nitrogen and phosphorus residuals of at least 1.0 and 0.2 $mg\,l^{-1}$ respectively, generally allows sufficient nutrient availability. The addition of these macronutrients to the aeration basin should closely match the demand; however, a wide variation in the organic loading rate can make this task extremely difficult. Many microbiological problems, both filamentous and nonfilamentous, exist in activated sludge systems due to nutrient-deficiency problems. Care should be taken to insure that all the environmental operating conditions have been properly maintained.

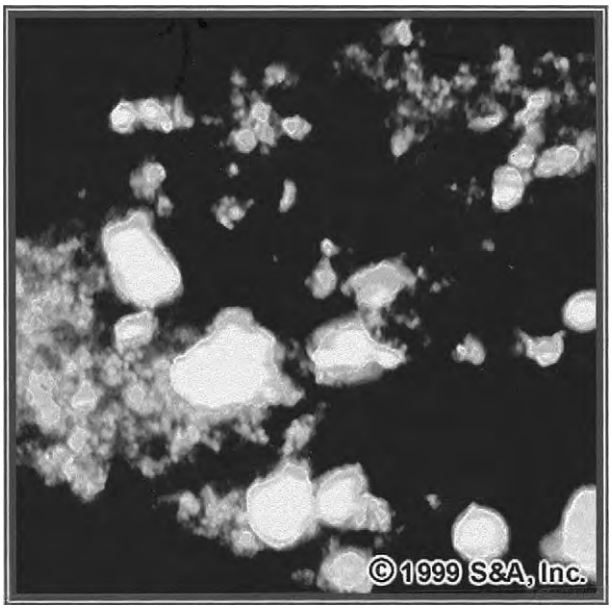

Figure 8 (see color plate 50) India ink staining for extracellular polysaccharide. India ink particles penetrate the flocs almost completely, at most leaving a small clear center. In activated sludge containing large amounts of extracellular material (as shown in picture), large clear areas indicate areas of low cell density. This slide was photographed at 10 × magnification. Subject to copyright and patent protection of Stover & Associates, Inc., as part of the FIL-IDENTPro™ Filamentous Bacteria Typing Program.

Protozoa and Activated Sludge

Effluent quality and plant performance can be related to higher life organisms present in the aeration basin mixed liquor. The biology of the activated sludge aeration basin consists of protozoa, rotifers, annelids, nematodes, and other higher life forms. Approximately 5% of the mixed liquor may be represented by 200 species of protozoa and other higher life forms. The protozoa is generally the most abundant organism in the aeration basin. Protozoa are all single-celled organisms that are hundreds of times larger than bacteria. Therefore, they are easily viewed under a microscope and can be used as an indicator organism of biomass quality.

Locomotion is usually accomplished by external hair-like flagella (tail) or through body movement. Ingested food particles are usually stored inside the vacuoles (food compartments) until enzymes break the food down for absorption through the cell wall. The role of these organisms varies from enhancing microfloral activity and decomposition, thereby aiding oxygen penetration, to contributing to

biomass flocculation. Microscopic examination of the biomass, when performed on a routine basis, such as daily, provides a picture of the condition of the biological population. Normally, microscopic evaluation of the biological solids is made under a 100 × magnification by placing a drop of biosolids on a glass slide. Increased magnification enhances the detail of the observed organisms. Microscopic evaluation of the biomass should be performed daily to monitor for increases or decreases in the various types of higher life growth in the biological population.

The major groups of higher life organisms in activated sludge are: rotifers, free-swimming ciliates, attached (stalked) ciliates, flagellates, amoeba, and higher invertebrates such as nematodes and annelids. The proliferation of certain organisms is dependent upon food availability, toxicity, and operating conditions. The four most common and readily observable indicator forms of higher life and the corresponding environmental conditions for proliferation follow:

- **Flagellated protozoa:** These organisms are normally oval and very small, and can be identified by their long whip-like tails (flagella) that are used for locomotion. The character motion of flagellated protozoa is undulating and relatively slow. If this type of higher life form is predominant, the biological system has a relatively high unstabilized organic content. This could indicate an unusually high BOD_5 loading or a low mixed liquor (biological solids) concentration.
- **Free-swimming ciliated protozoa:** These animals are also oval but are two to five times larger than the flagellated protozoa. The free swimmers can be identified by tiny hair-like cilia over their body. They move very fast in a darting fashion. If free-swimming protozoa are the predominant type of higher form of life, there is probably a moderate to low organic loading level in the system.
- **Stalked ciliated protozoa:** These organisms are much larger than the two previously mentioned and have a variety of shapes. They are equipped with a tail (stem) that they use to attach themselves to solid particles. Some types grow in a colony form that appears very much like a flowering bush; the bodies of the protozoa are located at the end of individual stems that attach to a main stem. Protozoa of this type can be identified by cilia located around the opening for intake of food. The cilia actually circulate water past the food-intake opening. The observer may be able to see small pinpoint particles being consumed by these animals. Stalked ciliated protozoa will predominate in a biological system that has a low organic level of unstabilized BOD of about 10–20 $mg\,l^{-1}$.
- **Rotifers:** These are the largest of the four types of higher life organisms discussed. They have flexible bodies that are also equipped with cilia, used to pull food to the rotifer as well as for locomotion, around the food intake opening. There are many types of rotifers, some of which have forked tails that are used to attach themselves to solid particles while feeding. A biological system in which rotifers predominate normally has a low organic level of unstabilized BOD.

All four forms can be observed in a system at any given time. The observer should attempt to identify which type of higher form predominates and then use the above guide for evaluating plant performance. These higher forms of microscopic organisms are very sensitive to toxic materials, and their presence or absence can help to indicate shock loads. The higher life forms will die before the bacteria are affected, so that routine observations can indicate trouble before it becomes so serious as to kill the bacteria. Also, the protozoa and other animals are strict aerobes, and therefore are indicators of ample dissolved oxygen. The microscopic animals can survive under anaerobic conditions for a few hours, but prolonged deficiencies of dissolved oxygen will be fatal. Low pH conditions will also cause a swift kill of the higher life forms. The protozoa, as well as the rotifers, are predators that continually remove small floc particles and dead microorganisms, and aid in the development of flocs. This predatory action keeps the bacterial population active and contributes to effluent quality by removing nonflocculated bacteria.

Reliable and stable operating conditions can be correlated to the type(s) of higher life organisms present in the system. Low effluent suspended solids concentration and low turbidity are normally achieved with a balance of free and stalked ciliates along with rotifers. A photograph of a rotifer of the type typically observed in activated sludge systems is presented in Figure 9. The establishment of a mixed microbial population in the aeration basin mixed liquor will result in optimum activated sludge performance.

Anaerobic Microbiology

Faculative bacteria are among the largest group of bacteria in nature. These bacteria can function in either an aerobic or anaerobic environment. The most common group of facultative bacteria are the Pseudomonas. Additional common facultative bacteria that have been identified in waste-water

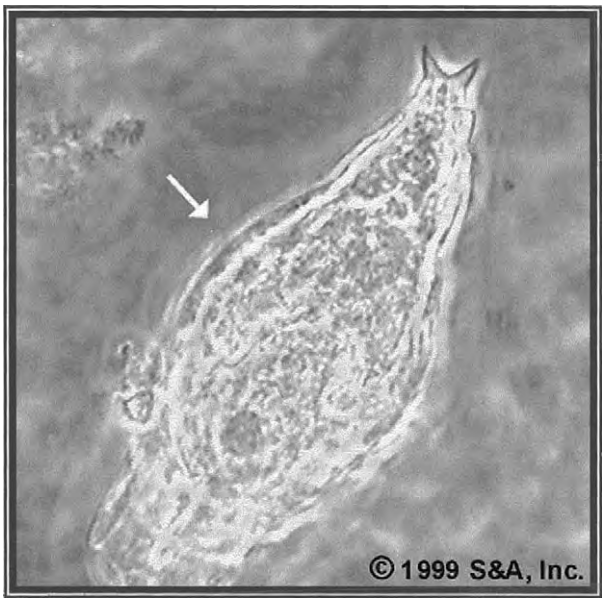

Figure 9 (see color plate 51) Unidentified species of the phylum Rotatoria, commonly called rotifers. This slide was photographed at 100 × magnification. Subject to copyright and patent protection of Stover & Associates, Inc., as part of the FIL-IDENTPro™ Filamentous Bacteria Typing Program.

treatment systems include *Alcaligenes*, *Achromobacter*, *Flavobacterium*, and various enteric bacteria.

The obligate anaerobic bacteria cannot tolerate dissolved oxygen. *Clostridium* is the major group of strict anaerobes. Sulfate-reducing bacteria are also strict anaerobes that belong to *Desulfovibrio*. They can metabolize a wide variety of organic compounds while reducing sulfates to various reduced sulfur intermediates, including hydrogen sulfide. The methane bacteria are also strict anaerobes that require a highly reduced environment for metabolism. They include *Methanobacterium*, *Methanosarcina*, and *Methanococcus*. Certain methane bacteria reduce carbon dioxide with hydrogen to produce methane and water, whereas others metabolize acetate. There is strong competition between the methane bacteria and the sulfate reducers. Even though bacteria are the primary microorganisms in anaerobic environments, protozoa have been observed in some anaerobic treatment systems. These protozoa have been described as flagellated and ciliated protozoa.

Desulfovibrio bacteria can use sulfates as their primary source of electron acceptors. The electron changes produce a series of reduced sulfur compounds, starting with thiosulfates and working through sulfur to sulfides. Energy transfer determines the changes in the sulfates. In the presence of excess organic matter that is readily metabolized by the *Desulfovibrio* bacteria, the reduction reactions go completely to sulfides as shown:

$$\text{Organic matter} + SO_4^{2-} \rightarrow H_2S + CO_2. \quad (2)$$

The hydrogen sulfide is partially soluble. As hydrogen sulfide is produced above its solubility level, it diffuses out of solution into the gases in proportion to its solubility.

At a pressure of 1.0 atm and a temperature of 35 °C, hydrogen sulfide is soluble to a maximum level of 2750 mg l^{-1} at pH 4.0. However, biological systems operate at pH values around 7.0, and as the pH increases above 4.0, hydrogen sulfide forms hydrogen bisulfide, as shown:

$$H_2S \leftrightarrow HS^- + H^+. \quad (3)$$

At pH 7.0, there will be approximately a 50:50 split with a total allowable sulfide concentration of 5765 mg l^{-1} ($H_2S + HS^-$). Formation of bisulfide therefore allows more sulfides to remain in solution. Since hydrogen sulfide is the culprit creating toxicity, toxicity can be reduced by raising the pH above 7.0 to drive the reaction toward bisulfide. Hydrogen sulfide levels below 200 mg l^{-1} should be maintained to eliminate toxicity problems.

The relative concentrations of electron donors (organic matter) and sulfates control the end-product formation. If the sulfate concentration is higher than the organic matter available for metabolism, the sulfate reducers do not have enough electrons to reduce the sulfates completely to sulfides. The overall reduction process can be expected to follow the pattern shown:

$$\text{Organic matter} + SO_4 \rightarrow [SO_3 \rightarrow SO_2 \rightarrow S_2O_3 \rightarrow S_2 \rightarrow HS \rightarrow S] + CO_2. \quad (4)$$

The thiosulfates are readily soluble, whereas, the free sulfur is insoluble.

See also: **Effluents from Food Processing**: On-Site Processing of Waste; **Microbiology**: Classification of Microorganisms

Further Reading

Campana CK and Stover EL (1990) *Filamentous Bulking Sludge Control – a Case Study*. Proceedings of the Sixth International Symposium on Agricultural and Food Processing Wastes, 328. St. Joseph, MI: American Society of Agricultural Engineers.

Eikelboom DH and Van Buijsen HJJ (1981) *Microscopic Sludge Investigation Manual*. Delft, The Netherlands: TNO Research Institute for Environmental Hygiene.

FIL-IDENTPro™ (1999) *Filamentous Bacteria Identification Program*, Stillwater, OK: The Stover Group.

Jenkins D, Richard MG and Daigger GT (1984) *Manual on the Causes and Control of Activated Sludge Bulking and Foaming*. Pretoria: Water Research Commission.

Jenkins D, Richard MG and Daigger GT (1993) *Manual on the Causes and Control of Activated Sludge Bulking and Foaming*, 2nd edn. Chelsea, MI: Lewis.

Richard MG (1989) *Activated Sludge Microbiology*. Alexandria, VA: The Water Pollution Control Federation.

Stover EL (1994) *Control of Opportunistic Filamentous Bacteria in Activated Sludge Treatment of Food Processing Waste Waters*. Environmentally Responsible Food Processing, *AICHE Symposium Series*, vol. 90, p. 42.

Stover EL and Campana CK (1990) *Operational Process Control for High Carbohydrate Filamentous Bulking*. Proceedings of the 1990 Food Industry Environmental Conference and Exhibition. Atlanta, GA: Georgia Tech Research Institute.

Stover EL and Gaden TB (1999) *Filamentous Bacteria Identification*. Proceedings of the WEFTEC® 99 Workshop #110 – Waste Water Microbiology, New Orleans, LA.

USEPA (1987) *Summary Report – The Causes and Control of Activated Sludge Bulking and Foaming*. Cincinnati, OH: US Environmental Protection Agency.

Disposal of Waste Water

E L Stover, Stover & Associates, Inc., Stillwater, OK, USA

T H Eckhoff, M&M/Mars, Inc., Hackettstown, NJ, USA

Introduction

Waste waters produced from food-processing operations vary in general based on the class or type of food processes; however, they share many common characteristics. These common characteristics typically allow biological treatment processes to be used for food-processing waste waters prior to discharge to the environment or prior to recycle and reuse. Water quality standards and criteria have been developed, and are constantly being refined and upgraded for discharge and use requirements. Existing and new environmental laws are being enforced to protect the aquatic environment from effluent discharges. The newer regulations and approaches emphasize watershed-based water quality initiatives and ecological risk assessment. Waste minimization and waste reuse approaches are being emphasized.

Characteristics of Food-Processing Waste Water

The food-processing industry is highly diverse; this diversity is reflected in the enormous variety of food items produced worldwide. In general, the classes of food processors include bakeries, candy manufacturers, meat processors, breweries, specialty convenience-type food processors, baby food manufacturers, fruit and vegetable processors, and dairies. Each type of food-processing facility has waste water problems endemic to the specific processes involved and to the particular product; however, there are common characteristics of nearly all food-processing waste waters. Refer to individual food processes.

The primary waste water pollutants associated with the food-processing industry are food product, raw materials, solvents, detergents, cleaning agents, and disinfectants. These constituents give rise to high biochemical oxygen demand (BOD), chemical oxygen demand (COD), total suspended solids (TSS), total dissolved solids (TDS), nutrients (primarily nitrogen and phosphorus), pH, oil and grease, and color. In general, food-processing waste waters by their nature contain few, if any, toxic constituents. The US Food and Drug Administration (FDA) requirements preclude the use of hazardous or toxic constituents in the preparation and processing of food products. However, these same regulations often require a food processor to clean and disinfect equipment and facilities with compounds that could be problematic during waste water treatment. For example, products that contain quarternary ammonium compounds, which have a strong disinfectant capacity, can pose a problem during waste water treatment if used indiscriminately for production equipment cleaning. For the most part, food-processing waste waters are not considered to be hazardous to human health. A common factor in food-processing waste waters is the enormous volume of effluent generated. Water is used as an ingredient in many foods, and in all parts of the food-processing operation, including product washing, blanching, cooking, cooling, diluting, cleaning, and sanitation. The total volume of water discharged daily may vary greatly, from $2650\,m^3\,day^{-1}$ for a bakery to $2.65 \times 10^6\,m^3\,day^{-1}$ for a cannery of comparable size.

Food-processing industries generally dispose of their waste water by treatment prior to discharge directly to a receiving stream, a publicly owned treatment works (POTW), or land application site. Most food-processing plants utilize a biological treatment system to treat their waste water. These systems include aerobic and anaerobic processes, either independently or in combination. Because of the nontoxic nature of the waste water generated in the food-processing industry, this approach is generally very effective. Some of the disadvantages associated with biological treatment include the generation of biological sludge which must be disposed of, the requirement of a skilled treatment system operator, and the sensitivity of biological systems to climatic change or

process upsets. Prior to discharge, the waste water must be treated to meet certain criteria, depending on the intended route of disposal. These discharge criteria are discussed in the following section.

Water Quality Standards for the Disposal of Waste Water

In the USA, all waste water discharges (including food-processing industries) are regulated at several levels. The ultimate regulatory authority derives from the Federal Water Pollution Control Act (commonly referred to as the Clean Water Act, or CWA) and is administered by the Environmental Protection Agency (EPA), although many states have a delegated authority (in lieu of the EPA) to enforce the various parts of the regulations. This section discusses the basic objectives of the CWA, the administration of water quality standards by states, and the role of toxicity testing in evaluating water quality. Although this discussion presents discharge standards from a US perspective, many other countries have adopted, or will be adopting, similar approaches.

Environmental regulations have progressed in a similar fashion worldwide. Water use and waste water disposal have presented food manufacturers with similar problems in Central and South America, the UK, western Europe, Central Europe, and Asia. In the 1950s to 1970s, economic goals of manufacturers played a more important role in a country's economic development than environmental concerns. As water and waste water problems from all sectors of the economy took on a more national spotlight, environmental laws were passed in an attempt to curb the rising tide of water pollution. Water quality studies of rivers and lakes were begun in the 1970s, and continue today. Many countries have established waste water discharge limitations based on water quality standards. The European Union (EU) is attempting to establish a framework across western and central Europe in environmental legislation to protect rivers, lakes, and groundwater.

More stringent legislation on waste water dischargers, including food processors, has evolved at varying degrees in the many nations of Europe, states of Central and South America, and Asia. The UK and western Europe began enacting environmental legislation in the 1960s and 1970s. Central and Eastern Europe followed in the 1970s and 1980s, gaining on the experience of the USA and western Europe. Many multinational companies now see it as their responsibility and are usually required by local legislation to provide the best water and waste water treatment available to their facilities.

The Clean Water Act

The first comprehensive legislation for water pollution control was the US Water Pollution Control Act of 1948. This law adopted principles of state–federal cooperation in program development but limited federal enforcement authority and federal financial assistance. These principles were continued in the Federal Water Pollution Control Act in 1956 and in the Water Quality Act of 1965. Under the 1965 Act, states were directed to develop water quality standards establishing water quality goals for interstate waters.

These laws were ineffective and the industrial boom in the USA during the 1950s and 1960s brought with it a level of pollution never before seen in the USA. Scenes of dying fish, burning rivers, and thick black smog engulfing metropolitan areas were commonplace on the evening news. In December of 1970, the President of the USA created the US EPA through an executive order in response to these critical environmental problems.

Congress passed the CWA, Public Law 92–500, on 18 October, 1972. Although prior legislation had been enacted to address water pollution, those previous efforts were developed with other goals in mind. For example, the 1899 Rivers and Harbors Act protected navigational interests (commerce), and the 1948 Water Pollution Control Act and the 1956 Federal Water Pollution Control Act only provided limited assistance for state and local governments to address water pollution concerns on their own.

In the Federal Water Pollution Control Act Amendments of 1972, Congress established the National Pollutant Discharge Elimination System (NPDES) whereby each point source discharger to waters of the USA was required to obtain a discharge permit. The CWA required the elimination of the discharge of pollutants into the nation's waters and the achievement of fishable and swimmable water quality levels. The NPDES program represented one of the key components established to accomplish this task. In addition, the 1972 amendments extended the Water Quality Standards Program to intrastate waters and required NPDES permits to be consistent with applicable water quality standards. Thus, the CWA established a combined technology-based and water quality-based approach to water pollution control.

Each state administers the CWA by establishing specific water quality standards for the water bodies in its region. The water quality standards include a narrative limit and a series of specific numerical limits. Each state is required to establish narrative waste

water discharge limits to address Objective 3 of the CWA: 'The discharge of toxic pollutants in toxic amounts is prohibited'. The narrative limits are fully enforceable. Waste water dischargers must therefore comply with both local narrative and numerical limits.

Specific numerical limits are actual physical concentration limits and represent the water quality criteria that most waste water treatment plants were designed to meet. Currently all 50 US states and US territories have established water quality criteria. Specific numerical limits are developed by each state according to the following criteria: the water body-designated use and the water quality criteria necessary to protect that designated use. All US surface waters have been categorized by water body or segment of water body according to specific designated beneficial uses. These designated beneficial uses for each water body are established locally by each state and revised at least every 3 years through periodic reviews. Examples of designated beneficial use classifications include public water supplies, propagation of fish and wildlife, recreational purposes, agricultural, industrial, and navigation.

The exact categories incorporated by each state may differ; however, the basic principles outlined in the CWA for support and propagation of aquatic life and recreation in and on the water are conserved by all states. States often set specific standards for specific uses or establish standards to protect individual species, for example. Special standards are also established for outstanding national resource waters (ONRWs). These are ecologically unique waters with standards set to protect the highest potential water quality.

The water quality criteria for protecting the designated beneficial use for each water body are also established at the state level. The EPA has published criteria recommendations in the *Quality Criteria for Water 1986* (called the *Gold Book*), consisting of scientific information on the effects of specific pollutants on aquatic life or human health. These criteria are used by state regulatory agencies to establish enforceable local discharge limits.

The NPDES Permit

The monitoring and enforcement tool of states and EPA is the NPDES permit. All food-processing industries that discharge waste water directly to surface waters in the USA must have an NPDES permit. To address indirect discharges from industries to POTWs, the EPA, through CWA authority, established the National Pretreatment Program as a component of the NPDES Permitting Program. The National Pretreatment Program requires industrial and commercial dischargers to treat or control pollutants in their waste water prior to discharge to POTWs. Indirect dischargers (to a POTW) must meet specific pretreatment criteria under the authority of the POTW's NPDES permit to insure compliance by the permit holder. Violations of NPDES or pretreatment permit parameters could result in an administrative order (AO), financial penalties, or revocation of the permit, resulting in a cease discharge order and closing the facility. The 5-year NPDES permits contain region and site-specific effluent limits as well as specific effluent monitoring and reporting requirements. The specific numerical limits enforced by the NPDES permit are categorized as conventional parameters and toxicity limits. These toxicity limits are based on whole effluent toxicity testing with aquatic organisms monitoring endpoints such as survival, growth, and reproduction.

Conventional Parameters

Conventional parameters are those pollutants that were initially targeted under the CWA for reduction. These parameters include the following:

- biochemical oxygen demand (BOD)
- chemical oxygen demand (COD)
- oil and grease (O&G)
- total organic carbon (TOC) and total oxygen demand (TOD)
- suspended solids (SS) and volatile solids (VS)
- total dissolved solids (TDS)
- nitrogen forms
- phosphorus
- pH, alkalinity, and acidity
- heavy metals
- specific organic compunds

Although effluent limits on these parameters have reduced the threat of food-processing industry effluents to human health and the environment, more stringent controls are necessary to achieve the goals of the CWA.

Priority Pollutants

In 1987 the CWA was amended to require states to adopt statewide numeric water quality criteria for toxic pollutants (Section 303(c)(2)(B) of the CWA). If numerical criteria were not available, the states were required to adopt criteria based on biological monitoring or assessment methods. A 'hit list' of regulated toxic compounds has been established over a period of years; these compounds are generally referred to as priority pollutants.

The Role of Toxicity Testing in Water Quality Evaluation

The EPA issued a national policy directive in 1984 entitled *Development of Water Quality-Based Permit Limitations for Toxic Pollutants: National Policy.* The EPA policy established control of toxic discharges through the NPDES permit system. Many food-processing NPDES dischargers are now being required to monitor for whole effluent toxicity (WET) to determine if their waste stream is toxic under specific low-flow conditions.

A discharge that repeatedly fails the toxicity test may receive a compliance letter from the state or the EPA which requires the discharger to perform a toxicity reduction evaluation to bring the effluent into compliance. The EPA policy states that biological testing of effluents is an important aspect of the water quality-based approach for controlling toxic pollutants. Effluent toxicity data, in conjunction with other data, can be used to establish control priorities, assess compliance with state water quality standards, and set permit limitations to achieve those standards.

Future Developments in Water Quality Standards

While a major reauthorization of the CWA did not take place in the late 1990s, focused changes on specific issues occurred. The Clean Water Action Plan (CWAP) was released in 1998 and is a broad comprehensive program that emphasizes a watershed management approach to water pollution control, strengthened existing and proposed new water quality standards (WQS) for water bodies, and implemented nonpoint source (NPS) controls for water discharges. The USEPA has a commitment to the CWAP objectives of restoring and protecting water resources on a watershed basis and maintaining a partnership with the States in achieving water resource goals.

In the coming year(s), EPA will continue to implement the Total Maximum Daily Load (TMDL) Program, as well as other watershed based water quality initiatives. The States are required to develop lists of impaired waters. These are waters that do not meet water quality standards, even after point sources of pollution (i.e. NPDES dischargers) have installed the minimum required levels of pollution control equipment. A TMDL specifies the maximum amount of a pollutant that a waterbody can receive and still meet water quality standards, and allocates pollutant loadings among point and nonpoint pollutant sources. The draft rule has elicited much interest and debate from the regulated community and the states due to the probability that the rule will call for increasingly stringent effluent limits on point dischargers, resulting in capital expenditures for larger and more sophisticated waste water treatment plants. Even indirect dischargers to a POTW will be affected as their discharge limits may be lowered to relieve the burden on POTWs.

The water quality standards also contain antidegradation policies which are designed to protect improvements in water quality and existing high-quality waters. This would require new or expanding dischargers to obtain an offset of 1.5 times their proposed discharge before beginning to discharge. These proposed offset requirements are in addition to current Clean Water Act provisions requiring discharge limits to protect water quality standards. Water quality standards will ultimately play an important role in the decision making process for food manufacturers as to where to build new facilities or relocate for consolidation.

Future pollution control strategies can be summarized in two words: waste minimization. With the 1984 congressional amendments to the Resource Conservation Recovery Act (RCRA), also known as the Hazardous and Solid Waste Amendments (HSWAs), the national waste management priority was shifted from treatment and disposal to waste reduction and recycling. The EPA defines waste minimization as the changes that can be made to reduce or recycle waste material that would otherwise have to be treated and disposed. While this effort initially focused on hazardous waste generation and disposal, the effort was soon applied to waste sources in general and waste water treatment.

Waste water reclamation and reuse was a major focus of waste minimization programs and can also provide an alternative water supply. The potential uses for recycled waste water include process water, wash water, rinse water and dilution water. It is evident that waste minimization practices provide opportunities for water conservation and water resources management as well as pollution control.

Prior to appropriate recycling or reuse, the waste water effluent must meet certain criteria for its intended use. As previously discussed, the conventional parameters may often be limiting factors in the water quality necessary for an intended use. Waters high in suspended or dissolved solids, for example, may not be suitable for reuse without additional treatment to remove solids. The list of regulated toxic compounds complicates the issue further as food-processing establishments use more effective sanitizing and cleaning agents to meet other

regulatory requirements. (*See* **Cleaning Procedures in the Factory**: Overall Approach; **Sanitization**.)

Agricultural reuse of waste water has economic advantages. Considerable benefits are realized by agricultural growers using reclaimed effluent as a water source, including a more dependable water supply and reduced water pumping and fertilization costs. Additional treatment by carbon filtration, disinfection, and reverse osmosis may be required for subsequent potable water use. Groundwater recharge with reclaimed waste water is practised in many areas. Zero discharge processes are being designed in many modern food processing facilities, utilizing water reuse and recycling to meet other water needs. Membrane technologies such as microfiltration, ultrafiltration, and reverse osmosis, are used for rinse water recycling and other effluent treatment. (*See* **Water Supplies**: Water Treatment.)

Most industries are resistant to change. As long as waste water treatment and disposal are economically viable and water is readily available, waste minimization will not be integrated into standard plant practice. The EPA is promoting waste minimization by applying increasingly more stringent discharge limits on all industrial and municipal dischargers. In addition to conventional parameters, the EPA has begun to emphasize whole-ecosystem impact of pollutant discharges.

New water quality standards will reflect this emphasis. Three main areas of emphasis for future water quality standards will be sediment toxicity, bioassessment, and ecological risk assessment. Of these three main areas, sediment toxicity and bioassessment standards continue to be defined. Evaluating Standardized and validated sediment toxicity test methods now exist allowing monitoring of sediment quality. Because sediment often serves as a reservoir or depot for toxicant accumulation, monitoring the surface water itself only provides a partial picture of potential ecological hazard. Further, toxicity characterization procedures have been developed for sediments allowing suspect toxicants to be determined in solid-phase samples analogous to procedures developed previously for effluents or surface waters.

Emphasis on whole-ecosystem monitoring and watershed management has prompted the use of ecological bioassessment. The most widely utilized bioassessment practice includes biological surveys which monitor community structure, biological diversity, and the presence or absence of pollution tolerant/sensitive species. Various trophic levels ranging from benthic invertebrates to fish may be included in these surveys.

An emerging trend in ecological water quality monitoring within a given watershed is the establishment of site-specific water quality standards. Although the site-specific water quality monitoring approach has been used primarily with metals to date, the concept may eventually be broadly applicable to a large number of chemicals. This approach essentially compares the toxicity of a specific metal (i.e., Cu, Zn, Pb, Ni, etc.) to traditional biomonitoring test organisms in the laboratory prepared synthetic culture water to?the toxicity of the same metal in the receiving stream. Because many toxicity mitigating and exacerbating factors exist in nature which cannot be accounted for in the laboratory, this approach provides a reasonable approach for determining the potential bioavailability (toxic fraction) of the toxicant in a specific environment. Since no two ecosystems or watersheds are identical, this approach has merit in identifying criteria for each specific ecoregion. In essence, this approach combines the use of laboratory biomonitoring testing with the *in situ* toxicity monitoring.

Several areas have emerged recently in ecological risk assessment. First, the use of sentinel species to provide a general indicator of ecosystem health has become more widely noted within the last several years. One of the best recent examples of the use of a sentinel species in monitoring ecosystem health, is the identification of deformed amphibians in several different locations in North America. Amphibians represent a sensitive class of animals which have raised 'environmental flags' in regions where deformed specimens have been identified. Not only are amphibians indicators of general ecosystem health, but are also potential indicators of risk to the human population. Obviously, the gap between ecological risk suggested by an adverse response in a sentinel species and potential human risk must be bridged.

Although the significance between the presence of endocrine disrupting chemicals (EDC) and their impact in the environment is still being addressed, the second area of intense focus in ecological risk assessment is the identification and monitoring of EDC. Passage in 1996 of the Food Quality Protection Act (FQPA) and Amendments to the Safe Drinking Water Act (SDWA) reflected these concerns and required EPA to develop a screening program using appropriately validated test systems and other scientifically relevant information to determine whether certain substances may have an effect in humans that is similar to an effect produced by naturally occurring estrogen or other such endocrine effects.

Specifically, EPA was required to develop a screening program by August 1998; to implement the program by August 1999; and to report to Congress on the program's progress by August 2000. In 1996, EPA formed the Endocrine Disruptor Screening

and Testing Advisory Committee (EDSTAC), charging the Committee to provide advice on how to design a screening and testing program for EDCs. This report contains the findings of the Committee, and is organized into sections discussing the EDSTAC recommendations on: a Conceptual Framework; Priority Setting; Screening and Testing; and Communications and Outreach. As a result, the EDSTAC recommended that EPA's Endocrine Disruptor Screening and Testing Program (EDSTP) should address both human and ecological (wildlife) effects; examine effects on estrogen, androgen, and thyroid hormone-related processes; evaluate endocrine disrupting properties of both chemical substances and common mixtures. The Committee determined that a tiered approach would be most effective in utilizing reasonably available resources to detect endocrine disrupting chemicals and quantify their effects. The core elements of the approach include initial sorting, priority setting, Tier 1 Screening (T1S), and Tier 2 Testing (T2T). EDSTAC recommended that T1S include three *in vitro* assays, three *in vivo* mammalian assays, and two *in vivo* nonmammalian assays. Specifically, these assays included:

In Vitro

1. Estrogen receptor (ER) binding/transcriptional activation assay,
2. Androgen receptor (AR) binding/transcriptional activation assay, and
3. Steroidogenesis assay with minced testis

In Vivo

1. Rodent 3-day uterotrophic assay (subcutaneous),
2. Rodent 20-day pubertal female assay with thyroid,
3. Rodent 5–7-day hershberger assay, and
4. Frog metamorphosis assay and fish gonadotropic assay

EDSTAC recommended that the T2T battery include a mammalian two-generation reproductive toxicity study, or a less comprehensive test in accordance with the guidelines outlined above, and tests addressing four additional taxonomic groups, including amphibians, fish, and invertebrates as follows:

Mammalian Tests

1. Two-generation mammalian reproductive toxicity study; or
2. A less comprehensive test:
   a. Alternative mammalian reproductive test or
   b. One-generation test

Multigeneration Tests in Other Taxa

1. Avian reproduction (with bobwhite quail and mallard),
2. Fish life cycle (fathead minnow),
3. Mysid life cycle (Americamysis), and
4. Amphibian development and reproduction (*Xenopus*)

See also: **Cleaning Procedures in the Factory**: Overall Approach; **Legislation**: International Standards; **Plant Design**: Basic Principles; Designing for Hygienic Operation; **Sanitization**; **Water Supplies**: Water Treatment

Further Reading

American Society for Testing and Materials (1993) *Standard Guide for Conducting Sediment Toxicity Tests with Freshwater Invertebrates*, E1383–93. Philadelphia, PA: ASTM.

Carawan RE (1989) *Environmental Issues facing Food Processors in the 1990s*. Proceedings of the 1989 Food Processing Waste Conference, pp. 217–237. Atlanta, USA: Georgia Technical Research Institute.

Driscoll TP (1990) *Waste Minimization and Product Recovery in the Crabmeat Processing Industry*. Proceedings of the 1990 Food Industry Environmental Conference, pp. 362–266. Atlanta, GA USA: Georgia Technical Research Institute.

Fort DJ and Stover EL (1997) Development of short-term, whole embryo assays to evaluate detrimental effects on amphibian limb development and metamorphosis using *X. laevis*. In: Dwyer FJ, Doane TR and Hinman ML (eds) pp. 376–390. *Environmental Toxicology and Risk Assessment: Modeling and Risk Assessment*, vol. 6, ASTM STP 1317, Philadelphia, PA: American Society for Testing and Materials.

Fort DJ and Stover EL (1997) Significance of experimental design in evaluating ecological hazards of sediments/soils to amphibian species. In: Dwyer FJ, Doane TR and Hinman ML (eds) *Environmental Toxicology and Risk Assessment: Modeling and Risk Assessment*, pp. 427–442. vol. 6, ASTM STP 1317, Philadelphia, PA: American Society for Testing and Materials.

Kendall RL and Asaday GH (1989) *Environmental Regulations Pertaining to Land Application Systems in the United States*, pp. 1–14. Proceedings of the 1989 Food Processing Waste Conference. Georgia, Atlanta, USA: Georgia Technical Research Institute.

London Environmental Protection Agency (1999) *Introduction to the National Pretreatment Program*. EPA-833-B-98-002.

Lounsbury J (1988) *A Federal Perspective on Waste Minimization*. Proceedings of the 1988 Food Processing Waste Conference, Section 10, pp. 1–6. Georgia, Atlanta, USA: Georgia Technical Research Institute.

Richardson SS (1989) *Waste Audit – a Self Help Approach to Waste Reduction*. Proceedings of the 1989 Food Processing Waste Conference, pp. 48–62. Georgia, Atlanta, USA: Georgia Technical Research Institute.

Richardson SS (1990) *Waste Reduction in Food Processing. A People Management Issue*. Proceedings of the 1990 Food Industry Environmental Conference, pp. 82–89. Georgia, Atlanta, USA: Georgia Technical Research Institute.

Composition and Analysis

M Tarbox, Technology Centre, Kingston Seymour, Clevedon, UK

Introduction

Effluents from food industries show great variability in flow and composition, are substantially organic and biodegradable, and tend to become acidic or start decomposing if their treatment is not immediate. The food industries considered in the present article are:

- meat and poultry
- dairy (processing of milk and milk products)
- breweries and distilleries
- fruit and vegetable processing, fruit juices
- sugar beet
- miscellaneous: egg processing, contact drying of foods, edible oil refining, prepared meals and foods, fish farming; cider, perry and alcoholic beverages

Meat and Poultry

Slaughtering and processing operations release mostly organic materials and the pollutants are classified by their potential effects on receiving water rather than as specific chemical compounds.

Meat and poultry effluents are characterized by high and variable 5-day biochemical oxygen demand (BOD_5), which are influenced by blood content. Improved blood collection is the most effective way of reducing BOD load. Wide differences in discharge volumes and strengths exist. Beef abattoirs produce 1.6–10.2 kg BOD per head and 0.4–3.4 m^3 per head, whilst poultry processors range over 13–45 kg BOD per 1000 birds and 20–32 m^3 per 1000 birds. **Table 1** shows potential pollution loads from modern slaughterhouses.

Table 1 Water volume and pollution per kilogram of carcass

Water volume (m^3) per kg of carcass	*Cattle 4.81*	*Pigs 6.11*	*Poultry 8.11*
Pollution			
COD[a]	32.3	27.3	21.0
BOD_5	13.2	13.2	9.3
FOG	5.2		
Total N	1.6	1.6	
SS	11.8	9.3	4.5

[a]Note that tripe and offal processing contributes more than 50% of the total chemical oxygen demand (COD) pollution.
BOD_5, 5-day biochemical oxygen demand; FOG, fats, oils, and greases.
Data from *Degre'mont Water Treatment Handbook*, 6th edn, vol. 1, 1991. Paris: Lavoisier.

Suspended solids (SS) are both organic and inorganic. Fats, oils, and greases (FOG) float on the surface of water and produce high BOD values.

Nitrogen results from breakdown of proteinaceous materials into amino acids and ammonia.

Bacterial counts are high and include a small percentage of pathogens. Positive results for fecal coliforms simply indicate the possible presence of intestinal pathogens. Coliform counts are in excess of 2×10^6 per 100 ml, which is comparable to crude sewage. Hygiene regulations in food factories show increasing concern to avoid salmonella infection. The risk of animal diseases, for example, anthrax and helminthic infection in some parts of the world, is an important consideration. (*See* **Food Poisoning**: Economic Implications; **Parasites**: Occurrence and Detection; **Zoonoses**.)

Dairy Processing

Milk, a major food commodity, is used for market milk (whole or skimmed, cream), butter, cheeses, anhydrous milk for use in icecreams and frozen desserts, and many other products.

Most (90%) of the pollution load in wastes is derived from dilutions of milk or milk products, but especially from cleaning and rinsing of residues in equipment. They are mainly organic, but large quantities of water are used for cooling or in the condensers of evaporators. Hygiene requirements and continuing trends for improved standards result in extensive use of cleaning compounds and biocides. (*See* **Cleaning Procedures in the Factory**: Overall Approach.)

The high strengths of dairy fluids emphasize the need to prevent wastage and spillage. The organic strengths of these materials are assessed by BOD and chemical oxygen demand (COD). Whole milk (BOD 100 000 mg l^{-1}) is about 250 times stronger

than domestic sewage. The major contributors to the BOD of milk are 4.8% lactose, 3.6% milk fat, 3.5% protein, and lactic acid. BOD is the generally accepted parameter for determining the pollution potential, and is the normal basic design parameter for effluent and sewage treatment plants, despite well-documented disadvantages of the test, e.g., a 5-day test with poor reproducibility ($\pm$ 20%). The COD test is relatively simple, cheap, gives better reproducibility ($\pm$ 5%), is quick (2.5 h), and is used by many dairy plants as a monitor of the effluent strength.

Characteristics of Dairy Wastes

Dairy effluents differ in composition from domestic sewage, often having higher BOD values (up to 2500 $mg\,l^{-1}$), and higher carbon:nitrogen ratios (11:1 compared to 3:1 for sewage). Wide ranges in BOD and COD are possible, but typical BOD averages are 2500 $mg\,l^{-1}$ and for COD 4000 $mg\,l^{-1}$, with BOD:-COD ratios 0.45–0.67.

The use of BOD coefficients assists in controlling processing plants. A BOD coefficient of 1.0 kg BOD per m^3 milk is equivalent to 1% of all the milk solids received, representing loss of the milk and product value and additional costs for effluent treatment. Wastage may represent 0.5–6.0% of the milk received, but an efficient plant has low coefficients (0.5 kg BOD per m^3 milk received; 0.5 m^3 water per m^3 milk received).

Milk fat losses give high BOD values which can be quantified in the standard FOG test using non-aqueous solvent extraction.

Eighty-five percent of the SS in dairy wastes is organic and noted for poor settlability. The values typically range from 400 to 2000 $mg\,l^{-1}$.

The pH value of dairy food waste water is most influenced by either acidic or alkali cleaning compounds. Variations throughout the working day typically range from pH 2 to 11, with daily averages being 4.4–9.4. Temperatures can show wide short-term variations from 11 to 72 °C.

Data from the International Dairy Federation for a number of countries show waste volumes of 0.5–37 m^3 per m^3 milk, indicating for high values the presence of cooling and evaporator waters, even stormwater.

Dairy processing plants show wide fluctuations in rates of flow, strength, and temperature, which are compounded by daily and seasonal fluctuations. In some countries, dairy plants are shut in wintertime. Quantity and strength of wastes are not related to plant size or degree of automation as the main factor is the control of waste (milk). Dairy wastes are free of toxic compounds, other than those that may be introduced by the use of cleaning chemicals and bactericidal compounds.

Whey from casein and cheese-making is a major source of pollution (BOD 34 000 $mg\,l^{-1}$; COD 75 000 $mg\,l^{-1}$). Although whey is often considered a waste product, whey processing methods are utilizing this material for food and animal feeds.

Membrane technology in the dairy and other food industries uses reverse osmosis modules in a range of configurations. Spiral-wound modules are well suited for waste water applications, including whey concentration. (*See* **Membrane Techniques**: Applications of Reverse Osmosis; Principles of Ultrafiltration.)

Breweries and Maltings

Brewing is carried out in most parts of the world. Malting is less widespread in its locations but is a precursor to brewing since malt is essential to the fermentation process.

Malting of barley is the first stage of brewing and whisky distilling, producing steep liquor which is a strong, deeply colored, readily putrescible liquid, prone to excessive frothing. The waste has high BOD, low SS, and lacks nitrogen and phosphorus. The COD:BOD ratio is generally less than 2, with COD up to 4300 $mg\,l^{-1}$ and BOD up to 2300 $mg\,l^{-1}$.

Brewing

In the brewing process, malt and other grains undergo enzymatic changes when insoluble compounds (starches) and high-molecular-weight protein compounds are converted to soluble sugars. After a cooking stage with hops, fermentation using yeast produces alcohol. Yeast is recoverable for reuse or use as food supplements and the draught beer is put into casks. Other beers such as lager, keg, canned, and bottled are aged in a secondary fermentation stage.

Large volumes of strong effluent are produced from most of the processing stages, but they are generally less polluting than maltings liquors.

Historically, spoilt beer was discharged into sewers as a requirement by HM Customs and Excise before allowing duty paid to be reclaimed. This practice continues under strictly controlled procedures.

Brewery wastes comprise 2.4–9.0 m^3 waste water per m^3 beer and 2.3–17.5 kg BOD per m^3 beer with typical BOD 775–1622 $mg\,l^{-1}$ and COD 1220–2944 $mg\,l^{-1}$ and a BOD:COD ratio of 0.67.

Distilleries

In the UK, barley and maize are the main carbohydrates for the production of whisky, gin, and vodka.

The initial processes are similar to brewing. A starch substrate is converted by enzymes into sugars and then fermented. Conversion is unnecessary when using molasses or cane juice. Processes are optimized to maximize the alcohol yield and exhaust the fermentable sugars. BOD of spent mash from cereals averages 12 000–34 000 mg l^{-1}, with pH 3.5–4.0.

Spent mash contains large amounts of organic acids and is liable to rapid anaerobic fermentation because of the high BOD and COD concentrations, producing hydrogen sulfide.

Effluents contain yeast and nonfermentable materials and comprise mainly washing, steaming, cleaning, and cooling waters, with overall average BOD values between 400 and 1600 mg l^{-1} (and up to 10 000 mg l^{-1} if there is no byproduct recovery).

Waste water and pollution load for a cereal distillery are 0.29 m^3 waste water per tonne cereal and 3.4 kg BOD per tonne cereal, respectively, assuming the stillage is used as cattle feed.

Fruit- and Vegetable-Processing Industries

Waste waters from these industries are mainly organic; much of the material is dissolved and highly reactive with bacteria. Special difficulties may arise owing to their high immediate oxygen demand, seasonal load variations, large variations in strength and volume, low pH, lack of nutrients, and the influence of product changes.

Carbohydrates, as simple sugars, such as mono- and disaccharides, promote acidification.

Waste water pH variations are a common problem at fruit and vegetable plants. In many potato-processing plants, most of the pollution load is from caustic or lye peeling, producing high pH values, when starch becomes swollen, partly hydrolyzed and gelatinized in the process and traces of soil are rendered colloidal. Lye peeling of potatoes may create effluent with pH value exceeding 11, BOD 2000 mg l^{-1}, and COD 3000 mg l^{-1}. Modern plants use alternative 'dry lye' methods of peeling which give lower-strength effluents.

Washing and peeling operations for fruit and vegetable canning and freezing produce 15–30 m^3 per tonne of raw product and 8–38 kg COD per tonne of raw product. Blanching creates approximately 5–10 m^3 of waste water per tonne of raw product.

Factories packing fruit and vegetables produce waste waters characterized by soil and plant residues, detergent cleaning compounds, bactericides, storage waxing compounds, and citrus-processing compounds.

Fruit Juices

Effluents are characterized by low pH due to high organic acid content, an imbalance of nutrients dueto low nitrogen and phosphorus, and considerable fluctuations in volume and loads, with the bulk of the pollution load dissolved and not readily settlable.

Typical waste water from apple processing is waste water volume 1.2 m^3 t^{-1}, BOD 1.4 kg t^{-1}, COD 2.0 kg t^{-1}, pH 5.5, BOD 2880 mg l^{-1} and COD 6374 mg l^{-1}.

Sugar Beet Mills

Effluent from sugar beet processing comprises surplus transport water and wash waters. The concentration of organic matter increases from the start of the beet harvesting campaign and reaches a plateau after 1–2 months. Excess transport water, stored in large ponds, may be slightly sugary, containing organic matter such as root hairs, and with a BOD increasing to 3000 mg l^{-1} or more by the end of a campaign.

The total water requirement for processing a tonne of beet is approximately 9–19 m^3, and concentrated effluents have BOD values 4000–5000 mg l^{-1}. Effluent volumes using water recirculation are estimated to be 1.2 m^3 per tonne of beet, 2–3 kg COD per tonne and 200–600 kg SS per tonne. COD values are approximately 1.5 times the BOD value.

Miscellaneous Waste Waters

Egg Processing

High BOD (up to 10 000 mg l^{-1}) and SS owing to product loss and cleaning processes.

Contact Drying of Foods

Evaporation equipment for milk can be used for gum arabic, flavorings, and caramel. The main effect occurs when a product is changed, requiring extensive cleaning out and giving high CODs and, for caramel, high color.

Edible Oil Refining

Sources of waste include leaks and cleaning operations, although refining processes transfer impurities into water, and there is oil–water emulsion. In modern refineries, oil production, refining, and washing stages generate 56–86 kg COD per tonne. Refinery wastes show values of BOD 500–6700 mg l^{-1}, SS 540–5850 mg l^{-1}, and FOG 300–4200 mg l^{-1}. BOD and SS values are directly proportional to FOG concentration.

Prepared Meals and Foods

Factories producing prepared meals may do little processing of raw materials since meat portions are prepared elsewhere, and vegetables are also generally preprocessed. Waste waters result from plant and equipment cleaning, continuously as product mix changes or as spillages occur. Vegetable rinsing and blanching, frying, cooking, and cooling water are other sources.

BOD generation ranges from 9 to 34 kg per 1000 kg of production with waste water concentrations of 600–4000 mg l^{-1}.

Waste from frozen bakery products results from cleaning of equipment and product loss. The major ingredients (flour, eggs, butter, and sugar) are high in BOD, SS, and fat; waste water strength is 2100–4300 mg l^{-1}, equating to 23 BOD kg per 1000 kg of product.

Manufacture of salad dressings, mayonnaise, mustard, and sauces produces high-strength waste waters derived from equipment cleaning, but volumes are relatively small. Typically BOD values are 2700 mg l^{-1}, but overall waste generation is as low as 8 kg per 1000 kg product with waste-water volume 0.3 m^3 per 1000 kg product.

Canneries for soups and baby foods use a variety of vegetables, meat, starch, and fruit ingredients and process raw vegetables on site. Waste-water sources are similar for fruit- and vegetable-processing industries. Wastes vary greatly in volume and strength; typical waste generation is 12 kg per 1000 kg raw product, with BOD value of 560 mg l^{-1}.

Factories making jams, preserves, and jellies produce wastes from cleaning equipment and cooking vessels and are strong owing to dissolved organic materials. Examples are 3–7 kg BOD per 1000 kg finished product with BOD strength 1100–3600 mg l^{-1}.

Confectionery production and bakeries generate high CODs due to sugars and cream, particularly associated with cleaning of utensils and equipment. Pollution loads can be considerably reduced by 'good housekeeping' techniques of brushing and collecting dry materials spilled on floors rather than hosing into drainage channels. A modern bakery using dry cleaning methods and silicone grease may have negligible effluent.

In summary, the prepared foods industries generate typical food-processing wastes with high COD and BOD concentrations, and organic suspended solids. The wastes are comparatively low in nutrients (phosphorus and nitrogen). FOG concentrations are significant where there is frying. Waste volumes and loads vary greatly and the following have been reported for a variety of prepared-food manufacturers: BOD 310–3200 mg l^{-1}; COD 560–7000 mg l^{-1}; SS 200–3700 mg l^{-1}; total P 4–22 mg l^{-1}; N 13–76 mg l^{-1}; FOG 82–2000 mg l^{-1}; with volumes 2.4–85 m^3 per 1000 kg and waste loads 5–26 kg BOD per 1000 kg, 9–53 kg COD per 1000 kg of production, and 1–21 kg SS per 1000 kg of production.

Fish Farming

Fish farms continuously abstract high-quality river or borehole water which is returned to the river contaminated by fish excreta and fish food residues. These large volumes impose considerable pollution loads on to rivers. Pollution is assessed by BOD, SS, and ammonia concentrations and stringent standards apply for Consents to Discharge. Consent is a legal document setting out the terms under which an industrial waste water may be discharged into a public sewer for conveyance, treatment and disposal by a water company in England and Wales (or local authorities in Scotland), or the terms under which the National Rivers Authority (or River Purification Board in Scotland) will allow a water company, local authority, domestic, or industrial effluent to be discharged into a watercourse. Consent values for fish farms may be absolute or differential, typically restricting the maximum increase in BOD, SS, and ammonia between inlet and outlet water to 3 mg l^{-1}, 6 mg l^{-1}, and 0.4 mg l^{-1}, respectively. Similar restrictions on increases apply to color (6 Hazen units), turbidity (3 Formazin turbidity units), and use of chemicals for controlling disease, e.g., Malachite Green (0.1 mg l^{-1}), formaldehyde (1.0 mg l^{-1}), and phenolic compounds (0.005 mg l^{-1}).

Cider, Perry, and Alcoholic Beverages

There is strong seasonal variation in effluent strength, peaking in late autumn with processing of new seasons' apples and pears. Effluents are characterized by low pH (a value of 5, sometimes offset where there is caustic washing of bottles), low SS (200 mg l^{-1}), soluble organic matter giving high BOD (1500–3000 mg l^{-1}), and COD (2500–5000 mg l^{-1}) and negligible nutrients (N, P).

Analysis of Food Effluents

The definitive guidelines for analysis of waters and effluents in the UK are the HMSO Blue Book series entitled *Methods for the Examination of Waters and Associated Materials*. First published in 1976, the series consists of individual books giving the methodology for both manual and automatic techniques for a specific parameter or groups of closely related parameters. Similar standard methods of analysis are used

in North America, Europe, and other developed countries.

The HMSO books are revised only occasionally and there is a tendency for the specialist laboratories to develop their own methods or new techniques and check them against those in the Blue Book. Thus, in 1991, little or no information appears in the Blue Book series on techniques such as capillary column gas chromatography, solid-phase extraction, gas chromatography mass spectrometry, plasma mass spectrometry, plasma emission, and automated discrete analysis. (*See* **Chromatography**: Gas Chromatography; **Mass Spectrometry**: Principles and Instrumentation.)

For routine work, the water companies generally take two types of samples from industrial and food-processing premises. First, spot samples are taken for quality control and compliance with the limits specified in the Consent to Discharge. Second, water companies take composite samples for charging purposes. Where treatment works discharge to rivers, trade wastes are charged on a volumetric and strength basis, the strength being assessed by a comparison of trade sample to domestic sewage using suspended solids and settled COD values. (At coastal sites, if there is no biological treatment, a strength factor based on COD is omitted since charges relate to the treatment given.)

Laboratories frequently observe that trade samples undergo a change in pH from the time of sampling to analysis, by as much as one pH unit. This is particularly noticeable with food wastes, although other factors such as the time lapse and storage temperature are influential.

The COD test is performed on the supernatant of the trade sample which has been settled for 1 h after the pH has been adjusted to pH 7. As a guide, trade samples are invariably stronger than they look. Sometimes settled CODs are higher than shaken values (i.e., unsettled CODs), especially when floating fats or greases are present.

Treatment plants for both sewage and trade wastes have biological capacity designed on a BOD basis. The variability in strength of food wastes requires a range of dilutions to be used to avoid all the oxygen being consumed during the 5-day BOD incubation period. A useful tip is to divide the shaken COD value by 8, the figure being indicative of the dilution factor to be used for the BOD test.

If chlorine is in an effluent, this should be removed before starting the BOD test. If the trade waste has a high or low pH, this should be adjusted to neutrality before starting the BOD test.

As a general guide for food wastes, the BOD value is approximately 50–75% of the COD value, and abnormally low ratios could indicate, for example, the presence of biocides.

The three most important routine analyses are BOD, COD, and SS, details of which are given below.

1. *BOD*. The amount of dissolved oxygen consumed by microbiological action when a sample is incubated for 5 days at 20 °C. Allyl thiourea is added to the dilution water if it is desired to inhibit nitrification in a sample containing nitrifying organisms, so as to obtain a value for carbonaceous BOD.
2. *COD*. The amount of oxygen consumed in a standard arbitrary manner with sulfuric acid and potassium dichromate upon boiling for 2 h.
3. *SS*. The solids retained after filtration through a glass-fiber filter paper (grade C, 70 mm in diameter) followed by washing and drying at 105 °C.

In certain instances more specific analyses may be required for analytes relevant to the particular food-processing operation in question. For example, dissolved or suspended tin may be a significant pollutant from canneries, or heavy-metal catalysts may be lost from hydrogenation facilities. It should be noted that in most of these cases significant pollution only arises from poor manufacturing control, and in these situations pollution also represents an economic loss.

See also: **Chromatography**: Gas Chromatography; **Cleaning Procedures in the Factory**: Overall Approach; **Food Poisoning**: Economic Implications; **Mass Spectrometry**: Principles and Instrumentation; **Membrane Techniques**: Applications of Reverse Osmosis; Principles of Ultrafiltration; **Parasites**: Occurrence and Detection; **Zoonoses**

Further Reading

Barnes D, Forster CF and Hrudey SE (eds) (1984) *Food and Allied Industries Surveys in Industrial Wastewater Treatment*, vol. 1. London: Pitman.

Cybulska G (2000) *Waste management in the food industry.* London: Campden & Chorleywood Food Research Association.

Degrémont Water Treatment Handbook (1991) 6th edn, vol. 1. Paris: Lavoisier.

International Dairy Federation (1981) *Dairy Effluents: Legislation on Water Conservation – Water to Milk Ratios – BOD, COD and TOC Relationships*. Document 139. Brussels: IDF.

Isaac PCG and McFiggans A (1981) Nature and disposal of effluents from malting, brewing and distilling. In: Kirsop BH and Hilton MG (eds) *Food Industry Wastes: Disposal and Recovery*, pp. 144–160. London: Applied Science.

Method for the Examination of Waters and Associated Materials. HMSO, London:

The Measure of Electrical Conductivity and Laboratory Determination of the pH Value of Natural, Treated and Waste Water (1978).
Suspended, Settleable and Total Dissolved Solids in Waters and Effluent (1980).
Chemical Oxygen Demand (Dichromate value) of Polluted and Waste Water (1986), 2nd edn.
5 Day Biochemical Oxygen Demand (BOD5) (1988), 2nd edn (with dissolved oxygen in waters amendments, 1988).
Schmidt CJ, Clements III EV and Farghuar J (1974) Wastewater characterisation for the speciality food industry. *Proceedings of the Fifth National Symposium on Food Processing Wastes*, EPA-660/2-74-058.

EGGS

Contents

The Use of Fresh Eggs

C G Belyavin, Chris Belyavin (Technical) Ltd, Newport, Shropshire, UK

Introduction

Eggs are an important part of the human diet in most countries around the world, and have been for many years. Concerns in recent years about the effects of egg cholesterol on heart disease have remained unproven. However, there has been a tendency for egg consumption per person per year to decrease in many societies over recent years, although the rate of decline has decreased, and in some countries, for example Italy and Austria, the downward trend has been reversed. The decline is partly as a consequence of an increase in awareness about a need for healthy eating and partly because of a change in people's living habits, with more married women working and a move away from a cooked breakfast, a traditional time for eating eggs, towards one based on cereal, toast, and fruit juice.

Egg production worldwide has more than doubled since the early 1960s. In 1961 Europe was the leading egg-producing region, accounting for almost 30% of the world total, closely followed by North and Central America and Asia. Today, Asia now has around 57% of world output, with North America and Europe in second and third places (**Table 1**).

Forecasts for 1999 indicated that production could exceed 46 million tonnes in that year and 47 million tonnes in the year 2000. Although there are some 150 or so countries in the world, the largest 50 poultry industries account for over 90% of world egg production. World trade in eggs is small, at around 1–2% of total production. The USA, The Netherlands, and China are the major exporters of eggs whilst Japan, Hong Kong, and Canada are the major importers.

Against the trend for increasing world egg production is this pattern of decreasing consumption per person per year. In the USA annual consumption per head of population has decreased from 390 eggs in 1952 to 242 in 1996. Figures for Europe are shown in **Table 2**.

In the UK in 1999, nearly 17% of eggs produced went into high-value egg products, showing an upward trend. A comparable figure for the USA would be nearer 30%. The remainder are either sold as shell eggs or are converted to liquid or dried egg as

Table 1 World egg production (million tonnes)

	1994	*1995*	*1996*	*1997*	*1998*
Africa	1.6	1.7	1.6	1.7	1.7
North America	6.0	6.0	6.1	6.3	6.5
South America	2.4	2.4	2.4	2.4	2.4
Asia	19.6	21.1	23.5	25.0	25.7
Europe	9.1	8.9	8.6	8.6	8.6
Oceania	0.2	0.2	0.2	0.2	0.2
World	38.8	40.4	42.6	44.2	45.1

Source: Anon. (1999a) *Poultry International*.

Table 2 European Community egg consumption (number per head per year)

	1994	*1995*	*1996*	*1997*	*1998*
Belgium/Luxenbourg	225	227	227	229	240
Denmark	258	254	225	244	238
Eire	151	156	130	127	122
France	253	264	263	257	252
Greece	218	212	214	214	213
Italy	218	219	221	222	225
Netherlands	173	175	175	176	178
UK	172	167	170	171	166
Germany	219	224	225	227	225
Spain	258	266	250	259	249
Portugal	158	133	145	144	142
Austria	230	230	232	241	240
Finland	168	190	177	167	165
Sweden	204	205	200	200	201
EU-15 (average)	214	217	210	211	211

Source: Anon. (1999c) *Poultry World*.

a means of removing surplus shell eggs from the market. Therefore, the major use of fresh eggs is in their original form sold to consumers in shell as a food.

Storage and Shelf-Life Problems

Shell eggs are a fresh, perishable food and consequently are liable to the same deterioration in quality over a comparatively short period of time that can be associated with any other fresh food product. The problem in the case of eggs is increased by the fact that the shell of the egg is porous and, therefore, the internal contents, which represent the consumable part, are not totally protected from the external environment. This comes about because the origins of the egg were not as a human food but as a source of propagation for the bird and the egg represents a development chamber for the growing embryo. Therefore, the need for gaseous exchange across the shell is inbuilt into the egg via the porous nature of the shell.

Storage Conditions

The conditions under which the fresh egg is stored will influence the rate of deterioration of its quality. However, even the most carefully controlled storage conditions can do no more than slow down the rate of subsequent deterioration. Ideally, eggs should not be stored for more than a few days, although they can remain consumable for up to 21 days after packing.

The rate of deterioration is closely related to the loss of carbon dioxide through the shell and the consequent increase in alkalinity of the albumen. At higher temperatures the rate of carbon dioxide loss is increased and the egg appears stale within a few days. Lower temperatures result in a slight loss and the structure of the egg is maintained at an acceptable level for several weeks.

This tendency towards a deterioration in quality leads to a number of problems with the storage of hen eggs for human consumption. Quality loss manifests itself primarily as the loss of moisture through the shell. There is a breakdown of the egg albumen (white), causing it to spread on breaking the egg, a characteristic not popular with consumers. Essentially it is the thick gel-like albumen which slowly breaks down, resulting in more and more thin, watery albumen. These changes, occurring in the rest of the egg, are associated with a change in proportion of thick to thin albumen.

The egg yolk becomes enlarged, flattened, and mottled and the cushioning effect of the thick albumen is decreased. With advancing deterioration, the yolk becomes detached from any remaining thick albumen. These changes give the egg a stale appearance which is unacceptable.

As the storage temperature is raised the rate of quality deterioration accelerates. At a constant temperature, quality loss is most rapid in the first 2 or 3 days. At high temperatures, the albumen deterioration is substantial and occurs within hours of the egg being laid. Consequently if eggs are to be stored they must be collected frequently and placed in a low-temperature store as soon as possible.

Egg Weight Loss

The porous nature of the egg shell results in egg weight loss under storage conditions. The most important factors influencing egg weight loss are temperature and humidity. As the temperature rises, weight loss becomes increasingly rapid. Relative humidity also influences weight loss and this effect is particularly important at high temperatures. After the first 3 days the weight loss of eggs stored at both a constant temperature and constant relative humidity is almost linear – the same amount is lost each day.

Egg Storage

Egg storage needs to be considered from the point of view of short-, medium-, or long-term storage. Egg quality deterioration and weight loss, as described above, only become excessive when the storage temperature approaches 21 °C. In many countries around the world this temperature is not achieved for prolonged periods of time and is therefore not a problem in the context of short-term egg storage. If higher temperatures are a problem, or longer-term storage is planned, then sophisticated cooling techniques are

needed and the procedure for egg handling is governed by the requirements of the stored eggs. The longer the storage period, the more precise must be the environmental control.

If the eggs are maintained at a temperature of 8–10 °C and a relative humidity of 85%, the quality deterioration and weight loss are slow. Providing the initial egg quality is good and that only intact and naturally clean eggs are stored, they can be held under these permitted controlled conditions for 6–8 weeks and still remain acceptable. Once eggs are removed from such storage conditions to an environment compatible with that found in a supermarket, egg quality deterioration and weight loss are rapid.

Refrigeration is the only means by which eggs can be stored for long periods of time (several months). A temperature of 0 °C and a relative humidity of 85% close to the eggs is essential. Such storage is not encouraged within the European Union (EU) and a regulation states that eggs which have been refrigerated in premises artificially maintained at a temperature less than 5 °C shall not be sold as class A, first-quality eggs.

A further problem associated with egg storage is the persistent risk of condensation on the egg surface which can result in mold growth. If warm, moist air is allowed on to the eggs while in store or as they are being removed from the store, condensation or sweating will occur at certain temperature and relative humidity combinations.

Within the EU eggs that have been washed also cannot be sold as class A first-quality. This regulation applies because of the negative effect on microbiological keeping quality resulting from washing. In some countries, particularly the USA and Australia, egg washing is routinely carried out on all eggs produced for human consumption. Where practiced, egg washing should be carried out to strict standards in terms of water temperature and hygiene.

In some countries eggs are sprayed with an odorless, colorless vegetable oil to prevent loss of moisture and carbon dioxide from the egg during storage. As summer temperatures in several countries can exceed the optimum storage temperature for eggs for prolonged periods of time, this technique helps to prevent rapid deterioration of egg quality but does not improve microbiological keeping quality of eggs kept dry in stores. The oil is generally applied as the eggs move from the washer to the grader or as they move in their boxes along the packing line. The application of the oil is by finely calibrated spray, usually from above, and the oil runs down the egg to give it a complete covering.

Handling and Marketing of Fresh Eggs

In recent years there has been a marked change in attitude and requirements in relation to egg marketing which has significantly influenced the way that eggs are handled over the last 10 years.

Major retail groups (supermarkets) account for a large proportion of the sales of the eggs produced in the UK and, consequently, the requirements laid down by these groups very much influence the policy of the industry as a whole within the UK. It is likely that the market share of such retail groups will continue to increase as consumers respond to quality assurances offered by such groups.

The handling policy now implemented by major egg-marketing companies is one of getting eggs from the point of production to the point of sale as quickly as possible. Eggs have to be treated as a fresh food. A typical time scale now applied would be 48 h at most from production site to retail outlet.

This means that production sites may well be directly linked to, or be situated close to, a packing and quality control center. Where the production facility is not close to a packing center, or eggs have to be held for any period of time (maximum 3 days) prior to packing, they will be held in a short-term store controllable, to provide conditions in line with those described above, to insure the maintenance of optimum quality standards.

Mechanical handling of eggs is now a feature of any modern production and packing plant. Eggs would be transferred from the point of laying via narrow belts to slatted cross-conveyors capable of carrying 150 eggs per meter run of conveyor, moving at a speed of around 4 m min^{-1}. Depending on the production facility, the conveyor would either take the eggs directly into a grading and packing machine or to a farm packing machine which would place the eggs in trays containing either 20 or 30 eggs for subsequent transfer to a plant for grading and final packing. In the latter case the eggs would be stacked on pallets or trollies, each holding around 3200 eggs for easy handling by forklift truck or other means. The eggs would be transferred to the packing plant in temperature-controlled vehicles in most cases (**Figure 1**).

Eggs are a fragile product and are very susceptible to damage. Approximately one in 10 eggs (10%) fails to reach the market place as a grade A, first-quality egg, and the majority of those downgraded eggs are substandard as a result of shell damage or faults. Therefore, correct handling procedures at all stages of the egg collection and handling processes are fundamental to maintaining egg quality.

Figure 1 A temperature-controlled egg transporter. Courtesy of Thames Valley Foods Ltd.

Figure 2 A modern egg-grading and packing machine. Courtesy of Deans Food Group Ltd.

Egg Packing

On arrival at the packing plant the eggs are transferred from the main conveyor, if the production plant is close by, or from the pallets or trollies on to the grading and packing machine (Figure 2). The Dutch and Americans dominate the world market for egg-grading and packing equipment and such machines can grade and pack up to 100 000 eggs an hour. On such machines, egg quality is assessed by a combination of computer-controlled detection of cracks, dirt, and internal blood spots with visual candling involving one or two operators. This process involves shining a light into the whole egg while rotating it to emphasize shell faults and make some of the contents visible to the naked eye. The eggs are then automatically sorted into various sizes by weight prior to packing and selling. Electronics and computerization are playing an increasing role in the operation of a modern egg-grading and packing machine.

The EU operates a four-weight category system ranging from eggs weighing less than 53 g (small) up to eggs in excess of 73 g (extra large). The weight sizes are separated by 10 g bands, 53–63 g being medium and 63–73 g large. The USA operates a six-grade system with a minimum net weight per dozen as follows: jumbo (30 oz.), extra large (27 oz.), large (24 oz.), medium (21 oz.), small (18 oz.), and peewee (15 oz.). In some countries, typically the Middle and Far East, eggs are not graded into sizes prior to sale but are sold by the kilo.

A typical modern UK egg-packing station or plant should be built to food factory specifications and regularly inspected by its customers and DEFRA (Department for Environment, Food and Rural Affairs) egg inspectors, who will check for compliance with EU regulations relating to egg marketing. It would ideally be situated with no production source further than 75 miles (120 km) away, although its market could be several hundred miles away. Eggs can be laid and packed within 24 h, and with the customer within a further 24 h in recognition of the need for a fresh product at consumer level. This requires, and enables, full benefit to be made of egg cooling and the development of a 'cool chain' from producer to retailer. In the UK at least, eggs are not displayed in cool cabinets in retail stores.

A modern egg-packing plant in the UK selling eggs to major retailers could handle 800 combinations of eggs and packaging, taking into account sizes, brands, and pricing. A single plant could be servicing 2100 retail outlets, but only 10 major customers, and packing some 1.6 million eggs a day. The operation, including distribution, could require a staff of 220 working an 8-h shift 5 days a week. Five hours a day would be spent on cleaning operations and the maintenance of hygiene. It should of course be remembered that many much smaller operations exist in any egg production industry around the world and the standards set and achieved may vary considerably.

The management of a large, modern operation needs to coordinate the supply, packing, and distribution and would depend very much on computer technology to do this. Sales information and analysis would be produced and a rolling 5-year plan would be maintained.

Packaging

Egg packaging is an important aspect of handling and selling and represents about 50% of total packing costs. There are many different types of egg packaging, ranging from fiber trays holding 20 or 30 eggs to specific packs made from fiber, plastic, or polystyrene, or even biodegradable materials to comply with concerns of environmentalists. Such packs would generally contain six, 10, 12, or 18 eggs. Niche markets, such as organic and eggs rich in omega-3 fatty acids, have developed for shell eggs as well as those from alternative production systems such as free-range and barn, and the packaging will be used to sell eggs and the egg concept (Figure 3).

Figure 3 Egg-packaging material used for the distribution and marketing of eggs. Courtesy of Thames Valley Foods Ltd.

Egg packs are and will be used to promote an image giving more information on the production system and methods. Value-added shell egg products, such as extra fresh, can only be promoted through the packaging design.

Eggs from Other Species

Traditionally, it is the hen's egg that is most widely used for human consumption and around which modern poultry breeding and egg-marketing techniques have been developed. Eggs from other species, if consumed, tend to be a specialist product sold in exclusive restaurants or stores. A typical example of this would be quail eggs in aspic. Turkey eggs are not generally sold for human consumption and are used more for breeding to produce turkey meat for human consumption. Human consumption of duck, goose, and pheasant eggs is negligible and of no commercial significance today. Eggs from waterfowl are nutritionally richer than those of domestic fowls as their yolks comprise about 36% of the egg's total weight, compared to 30% for hens. The consumption of duck eggs declined significantly because of fears relating to salmonella associated with them.

See also: **Eggs**: Use in the Food Industry; Structure and Composition; Dietary Importance; Microbiology

Further Reading

Anonymous (1983) *Table Egg Storage on the Farm*. Leaflet 563 (revised 1983). London: MAFF Publications.

Anonymous (1988) *A Guide to Using Mechanical Egg Collection Systems*. Springhill, Arley, Coventry: CES Potters.

Anonymous (1999a) *Watt Poultry Statistical Yearbook*. Poultry International. Mount Morris, IL: Watt.

Anonymous (1999b) Diamond's 65 years of eggs handling success. *Poultry World* 153: 25.

Anonymous (1999c) A series of hard knocks. *Poultry World* 153: 13.

Austic RE and Nesheim MC (1990) *Poultry Production*, p. 268. Lea & Febiger, Philadelphia.

Use in the Food Industry

C G Belyavin, Chris Belyavin (Technical) Ltd, Newport, Shropshire, UK

Introduction

The majority of eggs produced in the world are sold in shell for consumers to utilize in the home in traditional ways or for baking. In the UK only 17% of eggs produced go into egg products and in the USA the figure is nearer 30%.

Traditionally, surplus eggs not required for the shell egg market (normally the smaller sizes) and downgraded eggs were turned into egg products. Today, farmers have specific contracts to produce eggs for further processing and any egg that is damaged cannot be used for this purpose. There are essentially three types of egg products: liquid, dried, or frozen, with small markets for peeled, hard-boiled and hard-boiled pickled eggs. The main products are used widely in the bakery and confectionery trades along with other industries. **Figure 1** illustrates the many uses to which processed whole eggs can be put and also shows the way in which whole eggs can be used to produce value-added products for the food industry without going through the process of pasteurization with or without subsequent drying. Value-added products in the USA are often made with pasteurized eggs.

World trade in liquid egg is in excess of 110 000 tonnes traded annually, some 80% of which is traded between EU member countries. Outside of this area, the only significant major market is Japan which purchases around 15 000 tonnes a year. The Netherlands are the number one exporter closely followed by Belgium, between them accounting for about 45% of the total world trade.

Some 28 000 tonnes of dried egg are marketed internationally. Japan and European countries are the major buyers, with the USA being the major seller.

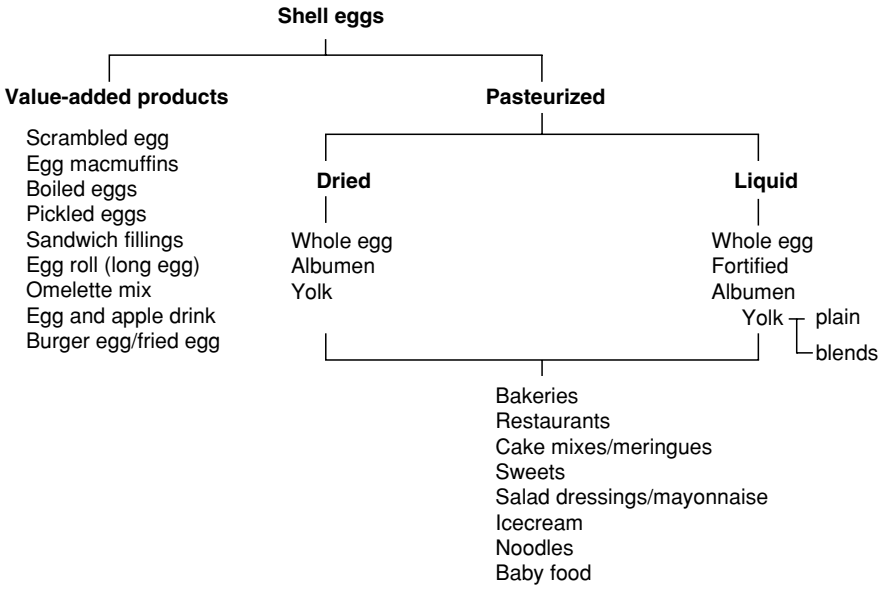

Figure 1 Eggs in the food industry.

When setting up an egg-processing plant it should ideally be located close to its supply of eggs. A constant supply of eggs is important in the efficient operation of the plant. The correct location enables the eggs to arrive at the plant in good condition, thus insuring that the resulting product is of good quality.

Liquid Egg Products

The production of liquid egg products takes place in specialist, large plants specifically designed for the purpose. Such a plant could process in excess of 36 million eggs a year. Mechanical equipment is used to break the eggs automatically and, if necessary, separate the egg yolk from the albumen. Such systems can handle up to 108 000 eggs an hour with electronic detectors to make sure that no yolk ends up in the white at the time of separation. A temperature of 10 °C is optimum for breaking and separating.

Further processing of eggs into liquid product is not a means of disposing of poor-quality eggs. In fact, the increasing concern about food safety and quality has meant that only surplus first-quality eggs and uncracked seconds are used in breaking plants.

At the time of breaking each egg will be visually inspected, either automatically or by an operator, to insure that no undesirable material is passed into the further processing line. Any eggs seen to be unsuitable will be rejected from the system at this stage and disposed of. The contents from the broken eggs are then passed through a sieve to remove any remnants of shell, the chalazae, which are the strands of protein which hold the egg yolk in place, and any other undesirable material.

About 50% of total liquid egg production in the USA is in the form of whole egg (yolk and albumen) and some may be fortified by the addition of extra yolk to increase the ratio of yolk to albumen over and above that found in a normal egg. The balance is divided between production of the separate components, i.e., either albumen or yolk. Liquid yolks may be sold as plain yolk or special blends such as sugared yolk or salted yolk. Such products not only meet the needs of specific customers but the addition of these agents to yolk also helps to maintain the consistency of the yolk, as without these natural additives it tends to separate into a thin liquid surrounding a congealed core if it is frozen. Whole egg tends to be standardized to a solids content of 24.7%. Separate egg white should not contain more than 0.03% yolk on a liquid basis and yolk tends to be standardized to a solids content of 43%.

In the past all these products would have been frozen rapidly in a blast-air freezer and held in this form until used. However, there has been a tendency in large-scale plants to chill the products and handle them as a fresh food, transporting them in refrigerated vehicles to their point of use. In this case, whole egg and yolk liquids are chilled to 4.4 °C and egg albumen to less than 7 °C. In this form the shelf-life of the product is only about 4 days but the costs of canning and freezing are avoided. Freezing is still adopted for small-scale use of liquid egg.

Bakeries as a group are the largest users of whole egg and separate albumen. Mayonnaise

manufacturers and salad dressing makers use large quantities of salted yolks and icecream makers use sugared yolks. Noodle makers and processors of baby foods use plain yolks.

Liquid egg or egg products are pasteurized (Figure 2) during their manufacture to prevent organisms such as salmonella becoming a major health hazard during their subsequent use and particularly after the eggs are removed from storage. Legislation to enforce this was introduced in the UK in 1964 and in the USA in 1966. The pasteurization process in the UK involves subjecting the product to a temperature of 64.4 °C (148 °F) for 2.5 min for whole eggs. Comparable figures for the USA are 60 °C (140 °F) for 3.5 min. Albumen is more susceptible to heat and so a slightly lower temperature is used, whereas the reverse is the case for yolk, where any salmonella is more heat-resistant. The alpha-amylase test is routinely used to test for adequate pasteurization.

Traditionally, liquid egg products were packed in metal cans or transported by large tankers. Today, modern packaging machinery enables the product to be put in large plastic bags contained in rigid outers and each bag can contain 500 kg of liquid egg which is drained from the container by means of a tap inserted in the bottom. The bag is then disposed of and the outer is returned to the manufacturer to be refilled. Product can be put into smaller cartons (Figure 3) which can be supplied to retail outlets for direct sale to consumers or larger cartons for supplying to hospitals and institutions. This is an expensive way of producing liquid egg because of the cost of the packaging. In these cases, liquid whole egg would be used for making scrambled egg and omelettes, for example.

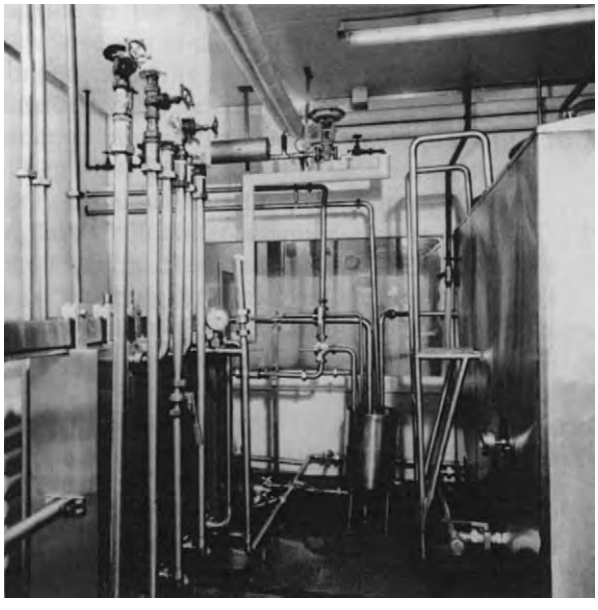

Figure 2 A liquid egg pasteurization plant. Courtesy of Oasters–Fridays (Cranbrook) Ltd.

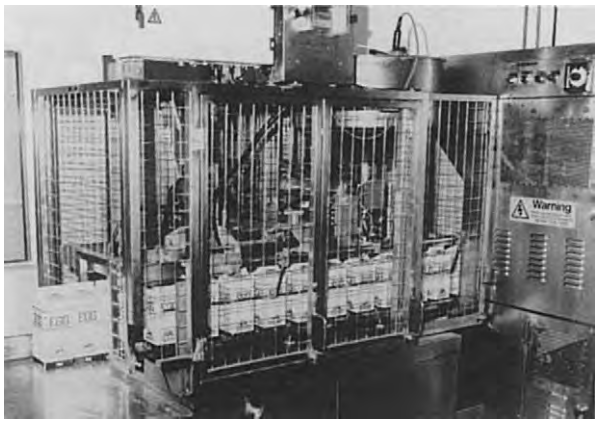

Figure 3 Liquid egg being packed into small cartons for retail sale. Courtesy of Oasters–Fridays (Cranbrook) Ltd.

Over the last 10 years, the production of egg products has been seen as the way in which any decline in egg consumption can be reversed. Consequently, a more diverse range of products is now available, including burger eggs, fried eggs, and meringue mix.

Freezing of Liquid Egg

If liquid egg is going to be frozen, to maintain a quality product, it is important that the freezing process should be rapid and that the frozen product should be stored at a low temperature. Cans of chilled egg should be subjected to an air blast at −30 to −40 °C to insure total freezing within 24 h, producing a uniform product without coagulation of the yolk. Subsequent storage should be at −10 to −20 °C. (*See* **Freezing**: Operations.)

Production of Egg Powder

The production of dried egg products is a further step beyond pasteurization. Spray drying is now the most common method for producing egg powder. However, some egg white is pan-dried and some whole egg is freeze-dried. Dried-egg products have a number of advantages which are listed below.

1. They can be handled and stored with ease and at low cost.
2. They are ready to use immediately with no thawing.
3. They are easy to handle in a hygienic way.

4. They are easy to remove from the container without scraping.
5. No bacterial growth can occur in powder at room temperature providing it is kept dry.
6. There is good uniformity.
7. There is precise control over the amount of water used in formulation.
8. There is no loss when used because dried egg is usually added directly to the batch.
9. No special transfer or storage equipment is needed.
10. Because the moisture content is reduced from around 74% to 2–4% by weight, there is a reduction in weight and volume and a concentration of food value.

A possible disadvantage of powder compared to liquid egg is a loss of 'fresh flavor' and a loss of certain functional properties such as aerating power unless treated with nonreducing sugars prior to heating. This, however, limits its use because of the sweetened nature of the powder.

There are now several different types of dried-egg products available which can be summarized under four major headings: dried egg white, dried plain whole egg and yolk, dried blends of whole egg and yolk with carbohydrates, and special types of dried-egg products.

The growth in egg drying is well illustrated by reference to the USA. Prior to 1941, less than 1% of the eggs produced went to drying plants. Government purchases stimulated drying operations to such an extent that 20 400 tonnes were produced in 1941, 107 000 tonnes in 1942, 118 000 tonnes in 1943, and 146 000 tonnes in 1944. The current annual volume is over 180 000 tonnes. A 30-dozen case of shell eggs will yield about 18 kg of liquid whole egg or 4.5 kg of dried product containing 2.5% moisture.

The handling of shell eggs prior to drying is no different from the handling of eggs that are to be kept liquid. The liquid material is put through a clarifier to remove any bits of shell and is then screened to remove the chalazae and vitelline membranes. Pasteurization is important, not only to control salmonella infection, but also to preheat the liquid so as to insure a low-moisture powder that will not show any scorching. A typical egg-drying plant is shown in Figure 2.

The liquid egg is then pumped under pressure of 17 250–34 500 kPa (2500–5000 lbf in^{-2}) to nozzles through which it is released into a large chamber where it immediately comes into contact with a stream of air that has been heated to temperatures of 120–175 °C. The enormous surface area created by atomization causes instantaneous evaporation of most of the moisture from the egg material and the formation of powder which falls to the floor as a fine powder while the moist air passes on out of the drying chamber at a temperature of 65–71 °C. It is important that the spray drier is designed so that the product stays at relatively low temperatures throughout the drying system and that the dried product is removed from the hot-air stream as rapidly as possible.

Modern driers produce a powder in which the moisture content does not exceed 2%. The temperature of the powder as it leaves the drier may be 65 °C or higher, and this must be reduced quickly to less than 29 °C if the product is to have good keeping qualities. Since the powder is extremely hygroscopic, the temperature must be reduced by contact cooling rather than by exposure to cold air. Two different powder-cooling methods are commonly used: first there is pneumatic conveying with cooling air and second, mixing the powder with carbon dioxide in the form of dry ice. After cooling, the powdered eggs are then packed immediately in sealed containers. Many of them are packed in carbon dioxide to remove the oxygen and lower the pH value in order to improve the keeping quality of the product.

Egg white can be dried in pans rather than through a spray technique. Pan-dried egg white can be in the form of flake or granular albumen or can be milled to a powder form. This form of egg albumen tends to have good whipping properties once reconstituted.

Chief users of dried-egg products are cake mix manufacturers, sweet makers, and manufacturers of meringue powders and, to a lesser extent, mayonnaise makers. Most dried-egg products made for commercial use are packed in polythene-lined boxes or drums. A common pack size would contain 25 kg of product. (*See* **Drying**: Theory of Air-drying.)

Quality of Dried-Egg Products

The quality of dried egg or dried-egg products can be affected by a number of factors. Important are quality of the eggs broken out to make the original product, handling methods, sanitation practices, conditions during processing, pasteurization procedures, drying, and the conditions under which the products are held in storage. Glucose (0.3–0.5%) is removed from liquid egg white prior to drying by fermentation using a yeast or bacterial culture, or by oxidation to gluconic acid using a glucose oxidase–catalase enzyme system. With glucose removed, dried egg whites are completely stable. If glucose is not removed, the product would be unstable because reducing groups from the glucose would combine with amino acids in the protein, leading to a condensation

reaction which would be followed by browning and the development of insoluble proteins. There could also be the development of off-odors and the loss of some functional properties during storage. Plain whole egg and yolk can be pasteurized and dried without removal of the glucose and consequently these are the least stable of the egg products.

Physical properties important in relation to dried-egg products are: bulk density, dispersibility, solubility, and reconstituted viscosity.

Storage of Dried-Egg Products

One of the real advantages of dried-egg products is their ease of storage. Most are relatively stable when stored at room temperature. Dried egg whites can be held under almost any storage conditions for an indefinite period of time. Dried products containing whole egg and yolk should be under refrigeration if held for long periods of time. Some of these are relatively stable at room temperature, particularly those where the natural glucose has been removed prior to drying.

Other Uses of Eggs in the Food Industry

Although liquid egg and dried egg still account for the majority of egg used in the food industry, there are other products which have evolved during the last 10–15 years which are commercially available.

The mass production of hard-boiled eggs has developed in recent years. Normally, white eggs are used in the process rather than brown because the shells remove from white eggs easier than from brown, resulting in less damage to the finished product. In countries where brown egg production predominates, such as the UK, this means that white eggs may have to be imported for the process from countries such as Denmark. However, in recent years flocks of white-egg-laying hens have been established in the UK specifically to supply the processing industry.

Again, only good-quality eggs are used in the process so that the final product is also of good quality. Normally in European Union countries small eggs would be used (under 53 g) for boiling as there is little demand for these sizes of egg in the shell egg market.

The eggs are gradually boiled as they move around a trough arrangement containing high-temperature water, with the objective of them being hard-boiled by the time they have completed a single circuit of the trough. The eggs are then removed, deshelled, and cooled. Once hard-boiled, the eggs can be used for a number of products. They can be pickled in jars in plain vinegar or with spices. They can be used in pies or salads and a major use in this form is in airline meals. Eggs damaged during the boiling or shelling processes would not necessarily be discarded. They can be diverted to the manufacture of sandwich fillings, such as egg mayonnaise, which can be packed into 5-kg tubs for distribution to sandwich manufacturers and the catering trade.

Problems associated with large-scale boiling of eggs, apart from that of deshelling, include optimizing the temperature for both the boiling and cooling processes, effluent discharges, and the prevention of black rings occurring at the junction between the yolk and the albumen. This is associated with the breakdown of sulfur-containing amino acids in the albumen producing hydrogen sulfide which reacts with iron released from the yolk to form iron sulfide.

Hygiene Measures

In all egg-processing plants producing egg products for human consumption a strict code of practice relating to hygiene is essential. Conventional egg products must now meet strict chemical, physical, and functional specifications which include moisture, fat, protein, ash, glucose, reconstituted viscosity, whipping ability, as well as functional performance in the foods in which they are used. Also included are microbiological standards for total plate count, coliform, yeast and mold, *Escherichia coli*, salmonella, coagulase-positive staphylococcus, and *Clostridium perfringens*. (*See* ***Clostridium***: Occurrence of *Clostridium Perfringens*.)

Many plants have a resident food technologist responsible for implementation of a quality control program using an inhouse quality control and microbiology laboratory. Every batch of product produced would be tested for contamination and production plants would be swabbed twice a week to test for the presence of undesirable organisms within such a program.

See also: **Eggs**: The Use of Fresh Eggs; Structure and Composition; Dietary Importance; Microbiology

Further Reading

Anonymous (1998) *Watt Poultry Statistical Yearbook. Poultry International*. Mount Morris, IL: Watt.

Austic RE and Nesheim MC (1990) *Poultry Production*, pp. 280–284. Philadelphia: Lea & Febiger.

Bergquist DH (1981) New development in drying egg products. In: Beuving G, Scheele CW and Simons PCM (eds) *Quality of Eggs*, pp. 7–14. Beekbergen, The Netherlands: Spelderholt Institute for Poultry Research.

Capetillo GO (1981) Egg de-hydration plant. In: Beuving G, Scheele CW and Simons PCM (eds) *Quality of Eggs* pp. 15–23. Beekbergen, The Netherlands: Spelderholt Institute for Poultry Research.

Structure and Composition

W J Stadelman, Purdue University, West Lafayette, IN, USA

Introduction

The term 'egg,' as used in this article, applies to eggs produced by chickens, or other avian species, for human consumption. Material is presented covering the formation of the egg, which helps to explain the structure of the complete egg. The structure is then discussed, including some data on abnormalities that might be found. The third section covers composition of normal eggs and the possibility of modifying composition through changes in management, nutrition, or genetics of the hen.

Formation of the Egg

The egg of the avian species is formed in the ovary and the oviduct. Figure 1 is a schematic diagram of the reproductive system. The ovary of the female chick contains numerous immature yolks, far in excess of the total number of eggs that the hen will produce.

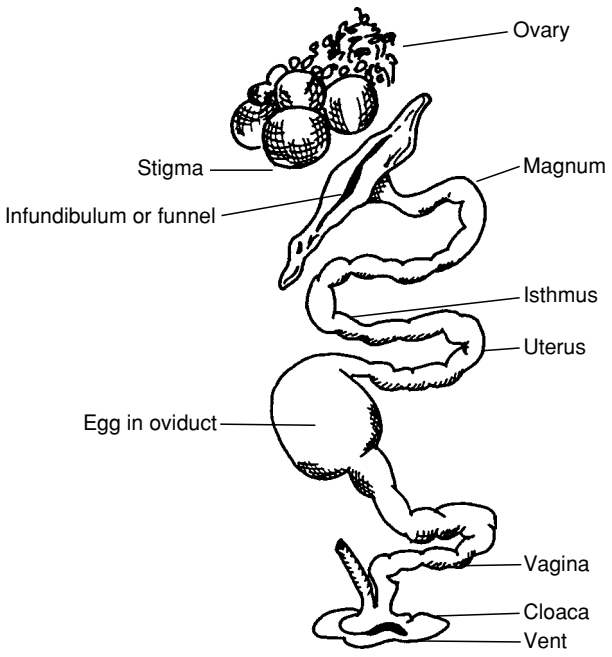

Figure 1 Schematic drawing of the reproductive system of a hen. Reproduced from Eggs: Structure and Composition, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

The immature ovules remain small, about 1–3 mm in diameter, until about 10–12 days prior to ovulation. The yolk increases in size by the deposition of layers of yellow and white yolk material. Growth of the yolk from 6 to 35 mm in diameter occurs during the last 6 days prior to ovulation. The actual color of the yolk material is the result of fat-soluble pigments in the feed. The alternating yellow and white layers are the result of feeding schedules. During the last day of yolk development, a line of smooth muscle develops in the follicle surrounding the yolk. The area is known as the stigma. Just prior to ovulation, the stigma contracts, forcing blood from the surrounding blood vessels. The yolk is then released from the follicle and falls into the infundibulum, or funnel, of the oviduct. Formation of the egg white, shell membranes, and shell follows ovulation. The entire process from ovulation to oviposition (egg laying) takes 24–25 h in the hen.

The oviduct can be divided into five sections based on the portion of the egg produced in that section (Figure 1). These are the infundibulum, magnum, isthmus, uterus, and vagina. Each of these segments expand or shrink as the hen goes into or out of egg production. The weight of the oviduct of a hen in production is about 77 g and when she goes out of production the weight regresses to less than 6 g. When the hen returns to production the oviduct again increases to its previous weight.

The first albumen is deposited around the yolk in the posterior region of the infundibulum. This albumen is rich in mucin fibers. As the developing egg moves through the oviduct it rotates, twisting the mucin fibers to form the chalaza which tends to hold the yolk in the center of the completed egg. The yolk with adhering albumen then moves into the magnum region where more thick albumen is deposited around the forming egg. Most of the protein of the egg white is added to the forming egg in the magnum region of the oviduct. In the isthmus, the inner and outer shell membranes are deposited around the albumen. Through a mechanism called plumping, water is absorbed through the shell membranes and into the albumen, causing the egg to enlarge in size. Plumping occurs in the uterus where the shell is deposited. If a colored egg shell is produced, the pigment, an ooporphoren, is deposited on the surface of the shell in the posterior region of the uterus. Time spent in the vagina is short, but the protein cuticle or bloom is put over the shell in this region. Oviposition follows as the egg moves through the cloaca.

Time spent in each region of the oviduct is about 20 min in the infundibulum, 3 h in the magnum, 1 h in the isthmus, 20 h in the uterus, and a few minutes in the vagina. After the egg is laid, it cools to the environmental temperature from the body temperature of the hen, which is about 42 °C. During the cooling, the liquid shrinks slightly in volume. This reduction in volume causes a separation of the two shell membranes, usually at the large end of the egg, forming the air cell. As moisture and gases are lost from the egg during storage or incubation, the size of the air cell increases.

Structure of the Egg

The avian egg is a complex and highly differentiated reproductive cell. The structure of a cross-section of an egg is shown in Figure 2. Starting with the interior, component parts are yolk, albumen, shell membranes, and the shell.

Structure of the Yolk

The center of the yolk is the latebra, the immature ovule present in the immature pullet. The yolk material is deposited in concentric rings which vary in color depending on the feeding schedule and fat-soluble pigments in the feed of the hen. With normal commercial laying rations the rings vary as light and dark orange-yellow layers. During the yolk material deposition, the latebra neck keeps the germinal disk on the surface of the expanding yolk. The yolk is enclosed by a semipermeable membrane, the vitelline membrane. This membrane is usually about 0.025 mm thick.

Structure of the White

The inner thick or chalaziferous layer surrounds the vitelline membrane. The chalaza consists of mucin fibers which tie into the chalaziferous layer. The next layer of albumen is the inner thin layer, followed by an outer thick layer. The outer section is composed of thin white.

Structure of the Shell Membrane

The shell membrane is composed of two layers (Figure 2). The thickness of the two layers of the membrane varies among different breeds of chickens but is less than 0.02 mm. The membranes are composed of protein fibers that are arranged so as to form a semipermeable membrane. The passage of moisture and gases through the shell membranes is mainly by osmosis. The membrane is one of the primary barriers to bacterial invasion of the egg contents.

Structure of the Shell

The shell is organized as an inner mammillary layer, then a spongy layer, and finally, on the outer surface, the cuticle. The mammillary layer is composed of numerous roughly conical knobs or mammillae, which are oval-to-circular in cross-section. The outer ends of the mammillae are larger than the inner ends, so that spaces between the inner ends allow for a ventilation system which permits passage of gases throughout the shell just outside the shell membrane

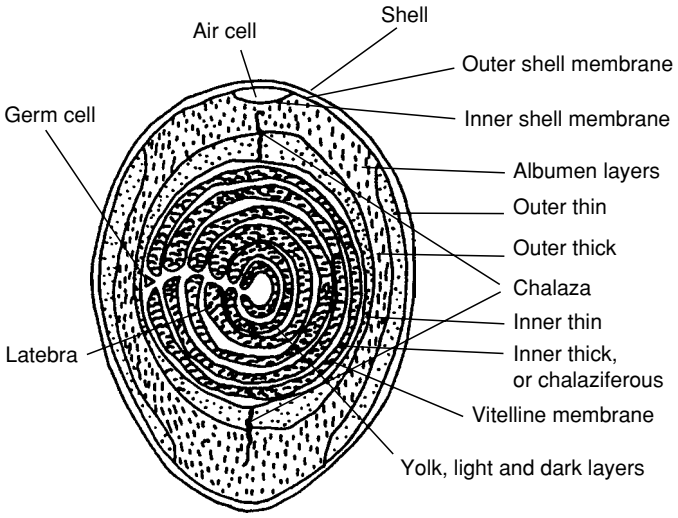

Figure 2 Schematic drawing of a cross-section of an egg. Reproduced from Eggs: Structure and Composition, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

layer. This structure allows gaseous exchange, which is essential for embryo development.

The spongy and mammillary layers are constructed in such a manner that there are numerous oval-to-circular openings through the shell. These are called pores. The normal egg has a greater concentration of pores on the large end, where the air cell is normally located, fewer in the equatorial region of the shell and fewest pores on the pointed end of the shell. The size of pores varies among shells of different hens, whereas eggs of a single hen are quite uniform in number, size, and distribution of the pores.

The outer surface of the shell is covered with a mucin protein which acts as a soluble plug for the pores in the shell. The cuticle is permeable to gas transmission.

Abnormal Eggs

The most common abnormality seen in freshly laid eggs is the double-yolked egg. Such eggs are produced when two yolks are released into the oviduct at about the same time. These eggs are most often produced by pullets just commencing egg production.

Meat spots in eggs are frequent. A meat spot is a bit of tissue from the lining of the oviduct. In white-shelled eggs, the spots are usually white and not noticed unless extremely large. Such large meat spots are removed during candling of market eggs. In brown-shelled eggs, most meat spots are brown and are thus noticed floating in the albumen. Again, large meat spots are removed during the candling operation. A less frequent abnormality is the blood spot. Blood spots are usually removed during the candling operation of shell-egg processing. Blood spots are most frequently found on the surface of the yolk and are the result of a drop of blood being lost from the capillaries of the yolk follicle at the time of ovulation. Blood-stained albumen is also possible if more than a drop of blood is lost during ovulation.

There are also abnormally shaped eggs owing to the malfunction of part of the uterus where the shell is formed. Rough-shelled and thin-shelled eggs are considered to be abnormal and are generally removed from a retail egg pack. Shell-less, thin-shelled, or soft-shelled (egg membrane only) eggs lack proper calcification. Many of these defective eggs are laid prematurely. The incidence of such abnormalities increases in aging hens that have been in production for more than a year.

Composition of the Egg

The avian egg is composed of all of those compounds needed for animal life, as evidenced by the hatching of a normal chick from a fertile egg. As a human food the egg contains all of the nutrients needed by humans with the exception of ascorbic acid (vitamin C).

The egg is naturally divided into the shell, egg white or albumen, and yolk. The shell is about 96% inorganic materials, largely calcium carbonate, with 4% organic material, largely protein, on a dry-weight basis. The egg albumen is primarily water, about 87%. The solids material is largely protein with small amounts of minerals and carbohydrates. The yolk is one of the most concentrated of all biological materials, with only 48% water in fresh egg yolk. The solids of egg yolk are made up of two-thirds lipid material and one-third proteins with small amounts of minerals and carbohydrates. The shell membranes are primarily proteins. Hens' eggs vary in relative amounts of the various parts but a general average is 10% shell, 1% shell membrane, 59% egg albumen, and 30% yolk.

Average proximate chemical composition of the edible portions of eggs is shown in Table 1. Refer to individual nutrients.

Egg Albumen Composition

There are a number of proteins in the albumen. Table 2 lists the identified proteins and gives some of the characteristics of each. The antibacterial nature of several of the proteins constitutes part of the defense against microbial invasions during the incubation of the egg. These defenses work equally well for shell eggs used as human food. These proteins have been separated and characterized. (*See* **Protein**: Chemistry).

Egg Yolk Composition

The egg yolk is a mixture of lipoproteins. An exception is phosvitin, which contains no lipid material and is a phosphoprotein. The protein content of yolk is about 16% and the lipid content varies from 32% to 35%, depending mainly on the strain of bird. The yolk lipid fraction is made up of 66% triglycerides,

Table 1 Ranges in chemical composition of the edible portions of eggs

Ingredient	*Whole egg (%)*	*Egg white (%)*	*Egg yolk (%)*
Water	72.8–75.6	87.9–89.4	45.8–51.2
Protein	12.8–13.4	9.7–10.6	15.7–16.6
Lipid	10.5–11.8	0.03	31.8–35.5
Carbohydrate	0.3–1.0	0.4–0.9	0.2–1.0
Ash	0.8–1.0	0.5–0.6	1.1

From Li-Chan ECY, Powrie WD and Nakai S (1995) The chemistry of eggs and egg products. In: Stadelman WJ and Cotterill OJ (eds) *Egg Science and Technology*, 4th edn. Binghamton, New York: Haworth Press with permission.

Table 2 Proteins in egg albumen

Protein	*Relative amount (%)*	*Characteristics*
Ovalbumin	54	Phosphoglycoprotein
Conalbumin or ovotransferrin	13	Binds metallic ions, iron, copper, and zinc
Ovomucoid	11	Inhibits trypsin
Ovomucin	3.5	Sialoprotein, viscous
Lysozyme or G_1 globulin	3.4	Lyses some bacterial cell walls
G_2 globulin	4.0 ?	
G_3 globulin	4.0 ?	
Ovoinhibitor	1.5	Inhibits serine proteases
Ovoglycoprotein	1.0	Sialoprotein
Ovoflavoprotein	0.8	Binds riboflavin
Ovomacroglobulin	0.5	Strongly antigenic
Cystatin	0.05	Inhibits thiol proteases
Avidin	0.05	Binds biotin

From Li-Chan ECY, Powrie WD and Nakai S (1995) The chemistry of eggs and egg products. In: Stadelman WJ and Cotterill OJ (eds) *Egg Science and Technology*, 4th edn. Binghamton, New York: Haworth Press with permission.

28% phospholipid, 5% cholesterol, and small amounts of other lipids. The major component of the phospholipids is phosphatidylcholine (73%), and other compounds are present in the following amounts: phosphatidylethanolamine, 15%; lysophosphatidylcholine, 6%; sphingomyelin, 2.5%; lysophosphatidylethanolamine, 2%; plasmalogen, 1%; traces of inositol phospholipid. The lipoproteins are classified as high-density lipoprotein (HDL), low-density lipoprotein (LDL), and very-low-density lipoprotein (VLDL). Density is determined by the percentage of lipid material in the molecule. The HDL fraction of the egg yolk consists of phosvitin and the livetins, α-, β-, and χ-livetin. These compounds all have less than 10% lipids. The LDL compounds are α- and β-lipovitellin and contain about 20% lipids. The VLDL lipovitellenin contains about 40% lipid. (*See* **Lipids**: Classification; **Lipoproteins**; **Phospholipids**: Properties and Occurrence; **Triglycerides**: Structures and Properties.)

Cholesterol in the egg is found only in the yolk. The concentration of cholesterol has been found to vary among hens The content of cholesterol per gram of yolk is relatively constant for a hen. For this reason, as the egg yolk size increases, the amount of cholesterol also increases. (*See* **Cholesterol**: Properties and Determination.)

Modifying Egg Composition

Composition of the egg varies slightly with egg size and to a greater degree among eggs from different hens, especially as hens of different strains are compared. The total solids content can be altered significantly by varying the proportion of yolk and white. Genetic selection has apparently made some change in the composition of hens' eggs, as evidenced by differences between the values listed in the US Department of Agriculture *Handbook 8.1*, published in 1973 and revised in 1989. One significant difference was in the saturated fat content of the egg which was found to be reduced by about 0.5%. Part of the difference might be attributed to changes in the diets fed to the laying hens. All nutrients in the egg, with the exception of zinc and choline, may be altered by feeding management or genetics, as reported by one or more research groups.

Modifying Shell Composition

The composition of the shell remains relatively constant, but shell thickness can be altered by reducing the calcium content of the diet. A vitamin D deficiency will also result in thin-shelled eggs. The composition of the shell does not change greatly, but the percentage of shell as part of the total egg may vary from less than 8% to over 12%. Shell membranes have not been found to vary significantly as a result of genetics, management, or nutrition of the hen.

Modifying Albumen Composition

The vitamin content of the albumen can be altered greatly by varying the vitamin content of the hens' diet. The mineral content can also be altered. The protein content remains constant for any hen but varies among hens from 9.7% to 10.6%. Specific amino acid content can be altered by dietary modification, but not sufficiently to affect total protein.

Modifying Yolk Composition

The fatty acid composition of hens' eggs can be altered by feeding fats of various composition to the hen. In addition, the vitamin and mineral content of the yolk can be changed by feeding.

See also: **Cholesterol**: Properties and Determination; **Fats**: Classification; **Lipoproteins**; **Phospholipids**: Properties and Occurrence; **Protein**: Chemistry; **Triglycerides**: Structures and Properties

Further Reading

Li-Chan ECY, Powrie WD and Nakai S (1995) The chemistry of eggs and egg products. In: Stadelman WJ and Cotterill OJ (eds) *Egg Science and Technology*, 4th edn. Binghamton, New York: Haworth Press.

Romanoff AL and Romanoff AJ (1949) *The Avian Egg*. New York: John Wiley.

Stadelman WJ and Cotterill OJ (1995) *Egg Science and Technology*, 4th edn. Binghamton, New York: Haworth Press.

Stadelman WJ and Pratt DE (1989) Factors influencing composition of the hen's egg. *World Poultry Science Journal* 45: 247–266.

USDA (1976) *Composition of Foods: Dairy and Egg Products*. US Department of Agriculture, ARS, Agriculture Handbook 8.1. Washington, DC: US Government Printing Office.

Dietary Importance

W J Stadelman, Purdue University, West Lafayette, IN, USA

Introduction

Eggs are a complete food for a developing embryo of a chick. The only vitamin not found in eggs is vitamin C. Most nutrients in the egg are present in variable amounts depending on breeding, feeding, and management of the hens. The amino acids in the proteins of the egg are relatively uniform among eggs so that egg proteins are often used as the standard against which other protein food sources are evaluated in biological value studies.

Nutrient Distribution in the Egg

The proximate analyses of the edible portion of whole eggs, albumen, and yolk are given in Table 1 of the previous article **Eggs**: Structure and Composition. Values are included for pure yolk and commercially available yolk. The difference is that some albumen is included with the commercial yolk, as evidenced by a change in solids from 51.2% to 44%. The percentages of yolk and albumen vary among eggs from different hens and change with age of the hen, with the percentage of albumen increasing with age of the hen. For rough calculation the egg is composed of about 10% shell, 30% yolk, and 60% albumen.

The overall composition of the egg is influenced by breeding and age of the hens. Nutrition of the hen has little effect on the proximate analysis but can significantly modify the content of some of the nutrients in eggs.

Protein and Amino Acids

The egg white is a protein system consisting of ovomucin fibers in an aqueous solution of globular proteins. The thick and thin layers of egg white differ only in their ovomucin content. The protein fractions of egg white consist of ovalbumin, conalbumin or ovotransferin, ovomucoid, lysozyme, ovomucin, avidin, ovoglobulins, ovoinhibitor, and flavoprotein. The molecular structure and amino acid mapping of several of these proteins have been documented. (*See* **Protein**: Chemistry.)

The absence of lipid materials from the albumen makes possible the formulation of foods with the high protein quality of eggs and a complete absence of fats. The amino acid composition of the several component parts of the egg is listed in **Table 1**. It is apparent that whole egg, egg yolk, and albumen each have a good balance of amino acids. (*See* **Amino Acids**: Properties and Occurrence.)

Table 1 Amino acids in the edible portion of eggs (per 100 g)

			Yolk	
Amino acid (g)	*Whole*[a]	*Albumen*[a]	*Pure*[a]	*Commercial*[b]
Alanine	0.696	0.608	0.861	0.818
Arginine	0.750	0.572	1.000	0.904
Aspartic acid	1.256	1.072	1.639	1.559
Cystine	0.290	0.272	0.301	0.301
Glutamic acid	1.632	1.400	2.126	2.005
Glycine	0.420	0.368	0.518	0.497
Histidine	0.296	0.237	0.434	0.402
Isoleucine	0.682	0.596	0.849	0.790
Leucine	1.068	0.886	1.470	1.364
Methionine	0.390	0.362	0.416	0.416
Phenylalanine	0.664	0.614	0.717	0.696
Proline	0.498	0.410	0.699	0.645
Serine	0.930	0.725	1.432	1.307
Threonine	0.600	0.479	0.892	0.819
Tryptophan	0.152	0.129	0.199	0.191
Tyrosine	0.510	0.410	0.747	0.686
Valine	0.762	0.671	0.934	0.885

[a]Based on data from USDA (1989) *Composition of Foods: Dairy and Egg Products*. USDA Handbook 8.1. Washington, DC: US Department of Agriculture.

[b]Based on data from Cotterill OJ and Glauer JL (1979) Nutrient values for shell, liquid/frozen and dehydrated eggs derived by linear regression analysis and conversion factors. *Poultry Science* 58: 131–134.

There are four major proteins in egg yolk. These are livetins, phosvitin, lipovitellins, and lipovitellenin. The livetins, (α, β, and χ) and phosvitin are classified as high-density lipoproteins; the lipovitellins (α and β) are classed as low-density lipoproteins and lipovitellenin is a very-low-density lipoprotein.

The amino acid balance of egg protein is almost perfect to meet human requirements. The amino acid content of eggs can be modified slightly by feeding and management programs. When hens are molted (loss of feathers), the cystine content of the eggs will be slightly reduced because of the high demand for this amino acid in the growth of new feathers.

Lipid Components

The lipids of the egg are all contained in the yolk in a fresh egg. There may be some migration into the albumen in stale eggs. Nearly all of the yolk lipids are present as lipoproteins. As mentioned above, they are classified as high-density, low-density, and very-low-density lipoproteins. The classification is based on percentage of the lipoprotein that is lipid material. The high-density lipoproteins contain very small amounts of lipid, usually less than 10%. Most human nutritionists consider the high-density lipoproteins to be desirable components of a diet. The low-density lipoproteins, α- and β-lipovitellin, contain about 20% lipids; lipovitellenin, the very-low-density lipoprotein, has 40% or more lipid content. This latter material is considered to be a harmful ingredient in the human diet as its presence tends to increase the serum cholesterol level.

A summary of lipid content is given in Table 2. The saturated fatty acids make up 38% of the total, monounsaturated fatty acids account for 46%, and the polyunsaturated fatty acids comprise the remaining 16%. The amounts of different fatty acids can be dramatically altered depending upon the fatty acids in the diet of the hen. (*See* **Fatty Acids**: Properties.)

A lipid component that has received much attention is cholesterol. According to the value given in Table 2, one large egg with a 17-g yolk has 213 mg of cholesterol. This is now considered to be the value, but a number of studies indicate a wide range of values, apparently the result of breeding, feeding, or management. Age of the hen will also influence the cholesterol content of an egg. This change is due to the increase in size of the yolk in eggs from older hens. The cholesterol level of an egg, on a per-gram-of-yolk basis, is relatively constant for any hen as long as the feed remains the same. The method of analysis for cholesterol in egg yolk can also influence results, as colorimetric methods generally give higher values than high-pressure liquid chromatography. Extensive works have been reported indicating that eggs in the diet of humans had slight or no effect on serum cholesterol levels in the blood.

Table 2 Lipid components of the edible portion of eggs (per 100 g)

		Yolk	
Fatty acids (g)	*Whole*[a]	*Pure*[a]	*Commercial*[b]
Saturated (total)	3.100	9.554	7.663
8:0	0.004	0.012	0.010
10:0	0.004	0.012	0.010
12:0	0.004	0.012	0.010
14:0	0.034	0.102	0.077
16.0	2.226	6.861	5.524
18:0	0.784	2.416	1.915
20:0	0.004	0.012	0.010
Monounsaturated (total)	3.810	11.741	9.444
14:1	0.010	0.030	0.030
16:1	0.298	0.916	0.717
18:1	3.472	10.699	8.623
20:1	0.028	0.084	0.067
22:1	0.004	0.012	0.010
Polyunsaturated (total)	1.364	4.205	3.374
18:2	1.148	3.356	2.842
18:3	0.034	0.102	0.077
20:4	0.142	0.440	0.352
20:5	0.004	0.012	0.010
22:6	0.036	0.114	0.092
Cholesterol (mg)	426.0	1,283.0	1,038
Lecithin (g)	2.30	6.687	5.396
Cephalin (g)	0.46	1.319	1.068

[a]Based on data from USDA (1989) *Composition of Foods: Dairy and Egg Products*. USDA Handbook 8.1. Washington, DC: US Department of Agriculture.
[b]Based on data from Cotterill OJ and Glauert JL (1979) Nutrient values for shell, liquid/frozen and dehydrated eggs derived by linear regression analysis and conversion factors. *Poultry Science* 58: 131–134.

Minerals in Eggs

The egg is an excellent source of many of the trace minerals. The mineral content of egg components is listed in Table 3. The exact mineral content of the egg depends on the mineral content of the hen's diet. The content of all minerals analyzed for could be varied, with the exception of zinc. The minerals of the shell are not considered in the values given in Table 3 as the shell is not a normal item in the human diet. The shell is primarily calcium carbonate and could be a readily available source of calcium. Refer to individual minerals.

Vitamins in Eggs

The vitamins found in the egg vary with the level of vitamins in the layer's feed. The average levels of vitamins are given in Table 4. The fat-soluble

Table 3 Mineral components of the edible portion of the egg (per 100 g)

Mineral (mg)	Whole[a]	Albumen[a]	Yolk	
			Pure[a]	Commercial[b]
Calcium	50.0	5.998	138.552	113.276
Chlorine	174.2	179.64	163.250	163.250
Copper	0.014	0.006	0.024	0.020
Iodine	0.048	0.003	0.133	0.108
Iron	1.440	0.030	3.554	2.861
Magnesium	10.0	11.976	6.024	5.837
Manganese	0.024	0.003	0.072	0.059
Phosphorus	178.0	11.976	487.944	395.080
Potassium	120.0	143.7	96.384	106.022
Sodium	126.0	164.7	42.168	60.008
Sulfur	164.0	167.7	150.600	150.600
Zinc	1.10	0.0	3.132	2.536

[a]Based on data from USDA (1989) *Composition of Foods: Dairy and Egg Products*. USDA Handbook 8.1. Washington, DC: US Department of Agriculture.
[b]Based on data from Cotterill OJ and Glauert JL (1979) Nutrient values for shell, liquid/frozen and dehydrated eggs derived by linear regression analysis and conversion factors. *Poultry Science* 58: 131–134.

Table 4 Vitamin components of the edible portion of eggs (per 100 g)

Vitamin (unit)	Whole[a]	Albumen[a]	Yolk	
			Pure[a]	Commercial[b]
A (IU)	634.0		1946.0	1580.0
D (IU)	49.0		147.6	118.3
E (mg)	1.40		4.217	3.390
B_{12} (μg)	1.00	0.210	3.132	2.512
Biotin (μg)	19.96	7.006	45.662	37.943
Choline (mg)	430.12	1.257	1301.00	1050.09
Folic acid (μg)	46.0	2.994	144.58	120.17
Inositol (mg)	10.78	4.132	23.795	19.891
Nicotinic acid (mg)	0.074	0.093	0.012	0.013
Pantothenic acid (mg)	1.254	0.120	3.807	3.899
Pyridoxine (mg)	0.140	0.003	0.392	0.320
Riboflavin (mg)	0.508	0.452	0.639	0.595
Thiamin (mg)	0.062	0.006	0.169	0.133

[a]Based on data from USDA (1989) *Composition of Foods: Dairy and Egg Products*. USDA Handbook 8.1. Washington, DC: US Department of Agriculture.
[b]Based on data from Cotterill OJ and Glauert JL (1979) Nutrient values for shell, liquid/frozen and dehydrated eggs derived by linear regression analysis and conversion factors. *Poultry Science* 58: 131–134.

vitamins, A, D, and E, are found only in the yolk. Water-soluble vitamins are found in both albumen and yolk components. Refer to individual vitamins.

Designer Eggs

Niche markets for special eggs have resulted in some producers of eggs modifying the nutrient content of the 'normal' egg by enriching the diet of the hen with specific fatty acids, vitamins, or minerals. An example is the 'omega egg.' This egg is obtained by adding a rich source of omega-3 fatty acids (menhaden fish oil, specific algae products, or canola oil). The omega egg producers also increase the vitamin E content of the diet of their hens to get a higher level of both vitamin E and omega-3 fatty acids in the egg. The inclusion of omega-3 fatty acids in the human diet has been reported to reduce the incidence of heart-related problems. Other designer eggs can be produced with increased levels of vitamins, minerals, or other fatty acids. A likely future designer egg is one with increased conjugated linolenic acid.

General Considerations

The exact nutrient composition of eggs is dependent on many factors, including the breed or strain of the hen, age of the hen, diet of the hen, and other management factors.

The per capita consumption of eggs in the USA has declined significantly from a high of about 400 eggs per year in 1945 to a level of about 233 in 1990. The annual consumption of eggs then increased slowly to about 245 in 1998. The decrease in egg consumption has been attributed to three factors: (1) increased buying power, which allows consumers to purchase more expensive substitutes for eggs, such as meats and convenience foods; (2) changing eating patterns, including lighter breakfasts; and (3) concerns about the negative health effects of eating eggs – first the cholesterol issue and more recently the inclusion of *Salmonella enteritidis*, a pathogenic microorganism, in a low percentage of intact shell eggs. In 1990 over 20% of all eggs were converted to egg products prior to sale to the ultimate consumer in the USA. This percentage has increased to over 30% by 1998 as more convenience foods rich in eggs are being offered, as well as the fear of foodborne pathogen.

See also: **Amino Acids**: Properties and Occurrence; **Ascorbic Acid**: Properties and Determination; **Calcium**: Properties and Determination; **Cholecalciferol**: Properties and Determination; **Cobalamins**: Properties and Determination; **Eggs**: Structure and Composition; **Fatty Acids**: Properties; **Fats**: Classification; **Retinol**: Properties and Determination; **Riboflavin**: Properties and Determination; **Thiamin**: Properties and Determination; **Vitamin K**: Properties and Determination; **Vitamins**: Overview; **Vitamin B_6**: Properties and Determination

Further Reading

Cotterill OJ and Geiger GS (1977) Egg product yield trends from shell eggs. *Poultry Science* 56: 1027–1031.

Cotterill OJ and Glauert JL (1979) Nutrient values for shell, liquid/frozen and dehydrated eggs derived by linear regression analysis and conversion factors. *Poultry Science* 58: 131–134.

Ginsberg HN, Karnally W, Siddiqyi M *et al*. (1994) A dose–response study of the effects of dietary cholesterol on fasting and postprandial lipid and lipoprotein metabolism in healthy young men. *Artery Thrombosis* 14: 576–586.

Hu FB, Stampfer MJ, Rimm EB *et al*. (1999) A prospective study of egg consumption and risk of cardiovascular disease in men and women. *Journal of the American Medical Association* 281: 1387–1394.

Li-Chan ECY, Powrie WD and Nakai S (1995) The chemistry of eggs and egg products. In: Stadelman WJ and Cotterill OJ (eds) *Egg Science and Technology*, 4th edn, pp. 105–175. Binghamton, New York: Haworth Publishing.

McCharen C (1988) Nutrition and cholesterol. In: Stadelman WJ, Olson VM, Shemwell GA and Pasch S (eds) *Eggs and Poultry-Meat Processing*, pp. 26–35. Chichester: Ellis Horwood.

Stadelman WJ (1977) Egg proteins. In: Graham HD (ed.) *Food Colloids*, pp. 207–239. Westport, Connecticut: AVI Publishing.

Stadelman WJ and Pratt DE (1989) Factors influencing composition of the hen's egg. *World Poultry Science Journal* 45: 247–266.

USDA (1989) *Composition of Foods: Dairy and Egg Products*. USDA Handbook 8.1. Washington, DC: US Department of Agriculture.

Vorster HH, Bernade AJS, Barnard HC *et al*. (1992) Egg intake does not change plasma lipoprotein and coagulation profiles. *American Journal of Clinical Nutrition* 55: 400–410.

Watkins BA (1995) The nutritive value of the egg. In: Stadelman WJ and Cotterill OJ (eds) *Egg Science and Technology*, 4th edn, pp. 177–194. Binghamton, New York: Haworth Publishing.

Yoshida A, Naito H, Niiyama Y and Suzuki T (1990) *Nutrition: Proteins and Amino Acids*. Tokyo: Japan Scientific Societies Press.

Microbiology

C E Clay, J L Lock, J Dolman, S A Radford and R G Board, University of Bath, Bath, UK

Introduction

The microbiology of eggs must be considered in a biological setting, namely their role in the breeding biology of birds. At the time of lay, an egg contains the nutrients, elements, and water required for embryo development. Natural selection has put in place a defense system that protects the food stored in the yolk from attack by saprophytic bacteria and microfungi. This article deals with the defense system of the eggs of domestic hens, the only species that has been studied in detail. It deals with the events that culminate in the breakdown of the defense system such that the yolk and white become grossly contaminated and, in most instances, markedly changed in appearance. Eggs may be endowed also with defense against pathogenic microorganisms harbored by the parents or present in the nest environment. This does not appear to have been studied in detail and will not be considered further.

The Defense

Physical and chemical systems contribute to the antimicrobial defense of an egg. The former resides mainly in the shell. Both systems occur in the albumen. The yolk appears to contain little, if any, defense against exploitation by saprophytic microorganisms.

The physical defense of the egg shell is most easily understood if its component parts are considered as resistances in series (**Figure 1**). The porous shell of calcite is bounded on its outer surface by a thin layer of glycoprotein, the cuticle (resistance R_1). At the moment of lay the cuticle is immature but it rapidly loses its amorphous state and becomes vesicular

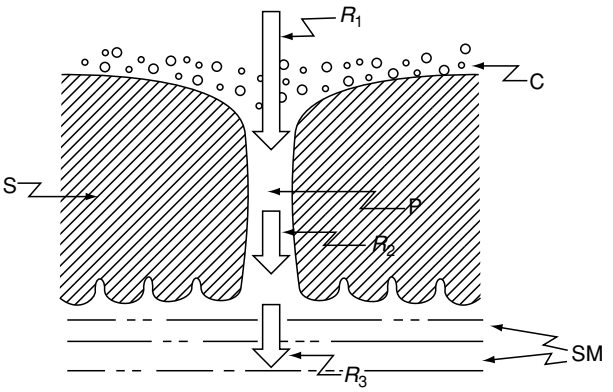

Figure 1 Schematic drawing of hen's egg shell in radial section showing the resistances (R_1, R_2, and R_3) to water movement. C, cuticle; P, pore; S, calcitic shell; SM, shell membrane. For further details, see Board RG and Sparks NHC (1991) Shell structure and formation in avian eggs. In: Deeming DC and Ferguson MWJ (eds) *Egg Incubation*. Cambridge: Cambridge University Press. Reproduced from Eggs: Microbiology, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

in structure with a fissured outer surface. Many of the fissures occur above the outer orifice of pores in the underlying calcite shell. These pores, which in the hen's egg have a straight posthorn morphology, number some $7–17 \times 10^3$ per shell and contribute resistance R_2 (Figure 1). The inner surface of the calcite shell is lined with two membranes. Both are formed from anastamosing fibers having a central core of elastin-like protein and a carbohydrate-rich mantle. These membranes constitute resistance R_3 (Figure 1).

If the germfree contents of an egg at oviposition are to become infected with saprophytic microorganisms, then such organisms have to penetrate the fissured cuticle, traverse the pore canal, and pass through the void spaces in the shell membranes. The agencies that promote translocation of microorganisms across the shell are considered below.

Any microorganism that penetrates the albumen encounters the principal defense system of avian eggs. The albumen contains upwards of six proteins having biological properties with the potential to kill microorganisms (e.g., lysozyme) or impede their growth (Table 1). Of the proteins cited in Table 1, ovotransferrin, a glycoprotein constituting about 12% of the total protein of egg white, plays a cardinal role. It binds two ferric iron atoms to a molecule at two different sites to produce a salmon-pink complex. The metal complexes dissociate in acid (< pH 6.5), but not in alkaline solutions (pH 9–10). Measurements of binding by equilibrium dialysis have shown that the binding constants for iron are very large (about 10^{30}). Changes in the pH of albumen favor iron sequestration by ovotransferrin as well as making egg white an unfavorable medium for microbial growth. At the time of lay, the albumen has a pH of 7.2–7.5 but, owing to the outward diffusion of carbon dioxide, the pH of the albumen drifts to a value of 9.0–9.5 within a few days when eggs are stored at ambient temperatures. The egg white is made an even more unfavorable medium for microbial growth through its minute content of combined nitrogen in a nonprotein form.

From a theoretical view point, lysozyme (a muramidase-*N*-acetylhexosaminidase, EC 3.2.1.17) would appear to be an important component of the antimicrobial defense. It hydrolyzes the β-4,1-glycosidic bonds in peptidoglycans, the exoskeleton of prokaryotic bacteria. Once lysed, this exoskeleton cannot prevent unfettered water uptake by the bacterial cell and the eventual lysis of its cytoplasmic membrane. Although lysozyme accounts for 3.4% of the proteins in the albumen, there is no compelling evidence that it has an important role in the chemical defense of the hen's egg.

Avian embryologists contend that the albumen supports the yolk and young embryo and cushions both from mechanical damage. The support stems from the viscous nature of albumen. Ovomucin is largely responsible for viscosity, and its contribution is augmented by electrostatic attraction between the negative charges of the terminal sialic acids of ovomucin and the positive charges of the lysyl Σ-amino groups of lysozyme. The viscosity of the albumen in a newly laid egg insures that the vulnerable yolk is held at the center of the egg and hence at the furthest distance from infection should microorganisms penetrate the calcareous shell and become lodged in the shell membranes. With storage, the albumen loses its viscosity, and the yolk floats upwards until contact is made with the inner shell membrane. In practice, therefore, the physical defense that the albumen contributes to the antimicrobial defense of eggs is transient. Moreover, there are reasons for believing that loss of viscosity favors bacterial movement in the albumen.

Table 1 Properties of the main proteins of hen albumen

Protein	*Amount in albumen (%)*	M_r	*Characteristics*
Ovotransferrin	12	80 000	Chelation of metal ions, particularly iron
Ovomucoid	11	28 000	Inhibition of trypsin
Lysozyme	3.4	14 600	Hydrolysis of β-1,4-glycosidic bond in peptidoglycans Electrostatic interaction with ovomucin
Ovomucin	3.5	ND[a]	
Ovoinhibitor	1.4	44 000–49 000	Inhibition of several proteases
Ovomacroglobulin	0.5	760 000–900 000	
Ovoglycoprotein	1.0	24 400	
Ovoflavoprotein	0.8	32 000	Chelation of riboflavin
Avidin	0.05	70 000	Chelation of biotin

[a]Value not determined or reported.

For further details, see Tranter HS and Board RG (1982) The antimicrobial defence of avian eggs: biological perspective and chemical bias. *Journal of Applied Biochemistry* 4: 295–338.

Course of Microbial Infection

It is generally accepted that the majority of hens' eggs do not harbor saprophytic bacteria or microfungi at the time of lay. The recent episode of egg-associated cases of food poisoning caused by *Salmonella enteritidis* phage type 4 (PT4) led to the contention that the egg contents may be infected during formation, by organisms either harbored by the ovaries (transovarian infection) or present in the oviduct (oviductal infection). A survey of upwards of 6000 eggs laid by hens known to harbor *S. enteritidis* PT4 demonstrated a low incidence (approximately 0.5%) of contamination of egg contents. Moreover, the majority of these eggs appeared to contain the contaminants in the albumen rather than the yolk. As yet there is insufficient evidence to draw conclusions about the behavior of salmonellae in eggs infected by the transovarian route during storage at chill or ambient temperatures. However, there is abundant information on the course of infection of eggs infected postoviposition with saprophytic bacteria or microfungi. (*See* **Food Poisoning**: Economic Implications; **Spoilage**: Bacterial Spoilage.)

With microfungi, contamination is confined to the shell's surface unless eggs are stored under very humid conditions. Under such conditions, hyphal development gives the shell a 'whiskery' appearance. Hyphae that penetrate the pore canals (Figure 1) spread progressively throughout the egg contents.

The available evidence suggests that free water and some agency to overcome the water resistance of R_1 and R_2 in Figure 1 are prerequisites for bacterial infection of the egg contents. If a warm egg is submerged in a cold suspension of bacteria, the decrease in volume of the yolk and white is greater than that of the shell. The partial negative pressure sucks water across the shell and lodges bacteria in the underlying shell membranes. The water resistance of the cuticle and pores (R_1 and R_2 in Figure 1) can also be overcome by imposing a positive pressure on eggs submerged in water or by suddenly releasing a negative pressure on eggs submerged in water.

Many studies have shown that Gram-positive bacteria, especially those of a coccal morphology, are the dominant contaminants on the surface of the shell. In contrast, the microflora of addled eggs is almost invariably dominated by Gram-negative bacteria (Table 2). The inference that selection of the latter at the expense of the former in the shell membranes has been supported by laboratory studies. (*See* ***Aeromonas***.)

The behavior of contaminants in the shell membranes is determined by storage temperature of eggs and the level of iron contamination of the water in which the contaminants were translocated across the shell. Storage of eggs at 4 °C selects psychrotrophic bacteria, such as *Pseudomonas putida*. At ambient temperatures, psychrotrophic organisms as well as mesophilic ones, such as *Proteus vulgaris*, occur in addled eggs.

Table 2 Types of organisms recovered from contents of eggs

Organisms	*Type of egg*	
	Rotten	*Tainted*
Coli-aerogenes	+	+
Proteus	+	−
Aeromonas	+	−
Pseudomonas	+	+
Alcaligenes	+	+
Achromobacter	+	+
Gram-positive	±	±

For details, see Board RG (1969) The microbiology of the hen's egg. *Advances in Applied Microbiology* 11: 245–281.

When relatively small numbers ($\leq 1.0 \times 10^3$ per ml) of bacteria suspended in water containing less than 1 μg of iron (Fe^{3+}) per ml are inoculated on to the shell membranes of recently laid eggs, very few organisms invade the albumen, and those that do appear to be in a quiescent state. At some stage in storage there is a rapid increase in the level of contamination of the egg contents and, when certain organisms are present, macroscopic changes in the white and yolk become evident (Table 3). The consensus of opinion is that a switch from the quiescent to active growth phase of contaminants in the albumen is the result of (1) chance collision of the contaminants with the yolk, or (2) contact of the yolk with an infected shell membrane. In either case, the yolk appears to short-circuit the chemical defense of the albumen by providing the contaminants with the Fe^{3+} which they require for growth.

If the water used to suspend bacteria contains > 1 μg of Fe^{3+} per ml, then there is a rapid build-up in the level of contamination of the albumen underlying the inoculated shell membrane, and a rapid onset of rotting of the egg contents, if appropriate organisms are present. In this situation, iron present in the water used to suspend the organisms is taken up by the mantles on the fibers of the shell membranes and used for growth by bacteria contained in the membranes. It is the continuous escape of organisms from the membranes that leads to a heavy contamination of the underlying albumen.

It was inferred above that only some of the organisms recovered from the contents of eggs cause macroscopic changes in the yolk or white. It is evident from

Table 3 Role of attributes of contaminants of addled eggs in the rotting process

Attribute	*Action*	*Organisms*
Pigment		
Nonwater-soluble	Discoloration of shell membrane at site of infection, occasionally in the white and on the surface of the yolk	*Cytophaga* *Flavobacterium* *Serratia*
Water-soluble	Discoloration of white	*Pseudomonas aeruginosa*, *P. putida*, and *P. fluorescens*
Proteases	Digestion of white and yolk	*Proteus* *Aeromonas*
Production of hydrogen sulfide	Blackening of the yolk	*Proteus* *Aeromonas*
Lecithinase	Breakdown of yolk emulsion	Coli-aerogenes
Slime production	Increase in viscosity of the albumen	Coli-aerogenes
Odor production	Characteristic odor emitted by infected eggs	*Pseudomonas maltophilia*
None of the above	No macroscopic changes even when eggs harbor 10^8 organisms per ml of albumen	*Alcaligenes* *Salmonella* *Citrobacter*

For further details, see Board RG (1968) Microbiology of the egg: a review. In: Carter TC (ed.) *Egg Quality – A Study of the Hen's Egg*. Edinburgh: Oliver and Boyd.

Table 3 that an organism needs one of four attributes in order to change the appearance of the yolk or white. Microorganisms lacking all of the attributes listed in Table 3 form populations of $>1.0 \times 10^8$ cells per ml of albumen without causing perceptible changes in the appearance of the yolk or white.

See also: ***Aeromonas***; **Food Poisoning**: Economic Implications; **Spoilage**: Bacterial Spoilage

Further Reading

Board RG (1968) Microbiology of the egg: a review. In: Carter TC (ed.) *Egg Quality – A Study of the Hen's Egg*. Edinburgh: Oliver and Boyd.

Board RG (1969) The microbiology of the hen's egg. *Advances in Applied Microbiology* 11: 245–281.

Board RG (1982) Properties of avian eggshells and their adaptive value. *Biological Reviews* 57: 1–28.

Board RG and Sparks NHC (1991) Shell structure and formation in avian eggs. In: Deeming DC and Ferguson MWJ (eds) *Egg Incubation*. Cambridge: Cambridge University Press.

Burley RW and Vadehra DV (1989) *The Avian Egg: Chemistry and Biology*. New York: John Wiley.

Cook N, Ellison J, Kurdziel AS, Simpkins S and Hays JP (2002) A nucleic acid sequence-based amplification method to detect *Salmonella enterica* serotype enteritidis strain PT4 in liquid whole egg. *Journal of Food Protection* 65(7): 1177–1178.

Hall R (2002) Outbreak of gastroenteritis due to *Salmonella typhimurium* phage type I 35a following consumption of raw egg. *Communicable Disease Intelligence* 26(2): 285–287.

Humphrey TJ, Whitehead A, Gawler AHL, Henley A and Rowe B (1991) Numbers of *Salmonella enteritidis* in the contents of naturally contaminated hen's eggs. *Epidemiology and Infection* 106: 489–496.

Metcalfe J, Stock MK and Ingermann RL (1987) *Development of the Avian Embryo*. New York: Alan R Liss.

Palmer PR and Guillette LJ (1991) Oviductal proteins and their influence on embryonic development in birds and reptiles. In: Deeming DC and Ferguson MWJ (eds) *Egg Incubation*. Cambridge: Cambridge University Press.

Tranter HS and Board RG (1982) The antimicrobial defence of avian eggs: biological perspective and chemical basis. *Journal of Applied Biochemistry* 4: 295–338.

Tullett SG (1991) *Avian Incubation*. Oxford: Butterworth-Heinemann.

ELDERLY

Contents

Nutritional Status

A Lehmann, Homerton Hospital, London, UK

Aging Populations

In the last 100 years, there has been a unique marked, continuing aging of the world's populations. Global life expectancy rose from 45 years in 1950 to 64 years in 1990; in the UK, for example, between 1950 and 2000, there was a fivefold increase in the number of people over 85 years old and a 10-fold increase in centenarians (in contrast, male longevity declined in Eastern Europe between 1960 and 1990). The increased life expectancy has been due to a reduction in younger mortality, mainly a reduction in infant mortality, but at present, the fastest growing age group is the elderly. In developing countries, the numbers and proportions of old people are rising with a 300% expected increase in the elderly populations this century (as yet, few demographic texts include epidemic HIV in their published calculations).

In the 1980s, it was hypothesized that increased longevity would be accompanied by 'compression of morbidity' into a shorter period at the end of life, and disability surveys suggest this may be happening very gradually. Nevertheless, the numbers speak very loudly in support of policies aimed at prevention and management of disability in old age.

Aging: Biology

The existence of a postreproductive lifespan in humans is unique among primates. Biological and social pressures are thought to have caused this evolution, e.g., by mother–daughter food-sharing, thus enhancing the daughter's fertility, or elders having knowledge of use at times of occasional disasters such as drought. Natural selection does not impact on individuals past child-bearing age: the aging process has resulted. It is no longer held that a few genes determine our aging rate: rather, the mechanisms are multiple and complex (even though a rare gene mutation may cause premature aging). Aging is likely to occur as a result of an accumulation of random cell and tissue damage, such as somatic mutations, oxidative damage, and the formation of aberrant proteins. The genes may affect somatic and intracellular responses to these random errors, and, of course, lifestyle, toxins, and the environment have profound effects on these processes.

It is important to grasp the consequent increase in variance of almost any parameter measured as people age. There is also a recurring question in old age research: which old people should be called normal? These two factors partially explain the relative delay in understanding the basics of nutrition in old age. It is generally believed that measures that might slow the aging process (as has already occurred to some extent) may also reduce illness and disability rates during old age. Hence the clinical importance of gerontology research.

Nutrition and Longevity. The Story of Dietary Calorie Restriction

It has been known for 60 years that there is one procedure that, if applied to laboratory rodents, reliably slows down the process of aging from a very early age (including prolongation of reproductive period) and in all its diverse manifestations: dietary caloric restriction (DCR). The control groups in some experiments were rodents fed *ad libitum* in captivity. Lengthy longitudinal primate studies are at present under way in the USA to test the potential role of DCR. It is not at present justifiable to extrapolate this finding and apply it to humans, although, presumably, such 'treatment' would reduce the morbidity associated with obesity. The population of Okinawa, Japan consumes fewer calories than other Japanese people, has a higher proportion of centenarians, and has lower overall death rates from both vascular disease and cancer. However, data below will demonstrate that demography and clinical research have most frequently highlighted the importance of undernutrition exacerbating the health of old people: if DCR is shown to have a role, it may be particularly so if started at a young age.

The Concept of Frailty

Frailty has been defined as the loss of a person's capacity to withstand minor environmental stresses. Although 'frailty' is sometimes used to avoid a precise definition of disability, in a more scientific way, it may be useful to help health workers understand what appears to be 'a syndrome of multi-system reduction in physiological capacity compromising an old person's function … and leading to unstable disability.' In frail elderly people, just treating the acute illness is often insufficient: in the clinical context, reduced musculoskeletal function, aerobic capacity, cognitive and neurological function, and nutritional reserve are the features measured to define frailty, because these deficits reduce the capacity for interaction with the environment. In an old patient presenting with an acute illness, all these different aspects of reserve and function need to be optimized if the patient is to have a chance of recovery. The frailty syndrome is sometimes seen as a downward spiral (or the 'dwindles' in the USA), and research confirms the frequently very major role of nutrition in the clinical context.

It is clinically useful to diagnose frailty by measuring the four key capacities, thereby identifying elderly people at risk of unstable disability with a view to reversing these problems if realistic. The suggested methods for measuring the components of frailty for muscle function include grip strength, ability to stand from sitting, aerobic capacity by measuring a 6-min walk in outpatients or a treadmill test, various cognitive tests, and practical functioning tests to assess the mind; and for nutritional state, body mass index (in people unable to stand or markedly kyphosed, demispan or other measures can be used to estimate skeletal size). (*See* **Nutritional Assessment**: Functional Tests.)

Nutrition, Diabetes, and Vascular Risk in Migrant Populations

It has been documented repeatedly that populations of apparently all genetic descents *other* than from temperate climates appeared to change from virtually zero rates of vascular disease associated with obesity, diabetes, hypertension, and hyperlipidemia (and a combination of these – syndrome X) to very high rates and severity of these if they moved from a rural, underdeveloped life to a 'western'-type urban lifestyle. All these conditions are age-associated illnesses and occur at younger ages, even sometimes at reproductive ages in women, and more severely than in European-descent populations. The changes can occur within one generation and are generally believed to be due at least in part to dietary changes in combination with reduced energy expenditure sometimes. This susceptibility to the Western lifestyle is found in genetically extremely diverse populations from Melanesia, the Indian Subcontinent, and Africa, and indigenous Indian people in the Americas, etc. Hunter–gatherers, who were the evolutionary ancestors of Europeans in a temperate climate, may have evolved in response to increased meat-eating and reduced vegetable intake in such a way that does not affect fertility. Increased fat intake has been suggested in the etiology of diabetes, e.g., in the Pima Indians. Several intervention studies support this hypothesis, e.g., if Australian Aboriginal people eat their traditional diet for 5 weeks, it very quickly reverses the diabetes and obesity that they present with. Baschetti has emphasized the role of high-density energy foods in causing these pathologies in susceptible populations and stated that the concentration of sucrose should not exceed $4.18\,\mathrm{MJ\,l^{-1}}$. Dietary potassium (in vegetables and fruit) is reduced by Westernization, and, since it is protective in this context, it has been suggested that potassium be added to drinks. Tropical and subtropical diets have a much larger ratio of vegetables to animal intake compared with those from temperature climates and it is interesting that the traditional Caribbean diet is closest to the present UK government recommended diet.

The high rate of vascular diseases in populations of tropical or subtropical descent is not purely due to these factors. Social factors are very significant, too, poverty being responsible for about 38% of the excess stroke morbidity in Afro-Americans compared with Euro-Americans. The potential health gains from lifestyle modification, however, clearly merit investment in appropriate education and facilities.

Usual Weight Changes During Aging

Both undernutrition and obesity jeopardize the health of old populations all over the world, particularly in urbanized societies. Most data derive from developed countries. Figure 1 shows the usual changes observed cross-sectionally in a representative UK sample.

The average trend of a slight slow weight loss after the age of 60 has been found in numerous surveys (see Table 1): one exception to this was a study in New Mexico on volunteers with no health problems and whom the authors described as socially privileged and unusually health-conscious. (Exclusion of any health problems in a UK community survey in Nottingham, which excluded old people living in institutions, would have meant that 95% of the elderly volunteers would have been excluded.)

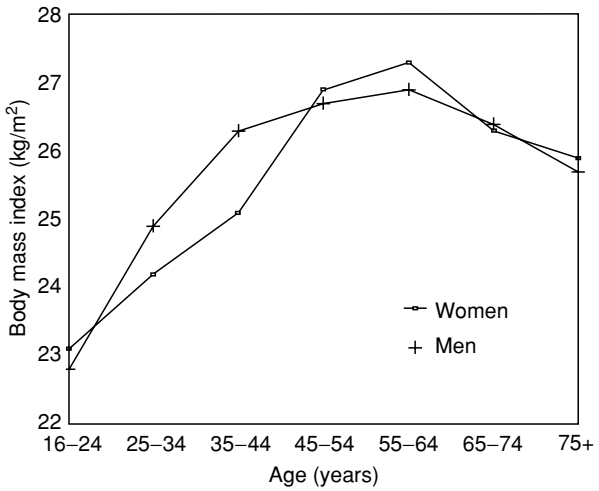

Figure 1 Mean body mass index by age and sex: health survey for England.

Explanations for the slight but distinct tendency for weight loss in old age are awaited. Metabolic changes so far studied during aging would be compatible with a tendency to gain weight. It has been suggested that appetite is reduced during aging by neuroendocrine changes. This 'anorexia of aging' should not be confused with the distinct epidemiology of old patients who notice anorexia and mention it to the health team, and in whom a search for a specific cause has a high yield. Low income, poor dentition, and solitary eating may be important, as may be rising taste or smell thresholds for which there is ample evidence. Since about 95% of the elderly population living in the community in the UK have some illness or limitation, these conditions, the drugs prescribed for them, or the social consequences of them also play a role. Resting metabolic rate is lower in older subjects, even after correcting for the changes in body composition, so this alone would cause weight gain if the diet were to remain constant or adjusting to energy expenditure. One hypothesis that might explain the observed changes and for which there is some limited support in research might be that, as we age, the energy costs of daily activities increase. Several authors have demonstrated that weight stability is associated with the best health in old age. The converse has also been shown: that weight loss is the most ominous predictor rather than thinness itself. People who are observed to be thin in old age include three groups: those who are lean and fit throughout life, smokers who are lean and not so fit, and those most at risk who have lost weight recently. This last group may be perceived as frail.

Body Composition Changes

During aging, there is a tendency for centripetal redistribution of fat, which is variable and includes internal fat deposition. The consequent practical problems for practitioners is the reduced use of limb subcutaneous fat measurements to reflect total body fat. Likewise, in old women especially, this trend plus height loss associated with osteoporosis may produce apparent abdominal protrusion and a very common false label of obesity by doctors assessing on the basis of abdominal examination. Muscle mass decreases during aging, although physical training can to some extent modify this trend. In illness or trauma, dehydration is very common in old people as kidney water conservation ability falls with age, and the threshold to thirst commonly rises with old age.

Gut Changes During Aging

For most purposes, these are negligible in terms of the old person's nutritional status. Probably, the most significant change is oral health, in particular the dental health, which, time and again every year,

Table 1 Longitudinal studies reporting weight change in old age arranged in descending order of magnitude of weight change found (combined group of men and women)

Reference	*Age (years) of subjects*	*Duration of study (years)*	*Number of subjects*	*Weight change (kg per year)*	
				Men	*Women*
Steen (1988)	70–81	11	23	−0.64	−0.55
Stevens *et al*. (1991)	> 65 at entry	25	370		−0.56*
Rissanen *et al*. (1988)	60–69	4–7	294	−0.06	−0.4
	> 70	4–7		−0.36	−0.7
Lehmann *et al*. (1996)	65–100	4	629	−0.21	−0.34
Williamson *et al*. (1990)	65–74	10	1856	−0.11	−0.17
Garry *et al*. (1992)	60–84	9	157	ns	−0.14
Friedlander *et al*. (1977)	65–69	5	1813	ns	
	> 70	5		ns	

ns, not significant.

is shown in many studies to be a relevant factor in explaining undernutrition in the elderly. Esophageal motility may become disordered occasionally during aging, but a much more common finding is in the stomach, where gastric acid production may be reduced. The importance of this in terms of health is particularly the increased risk of gastroenteritis because the acid barrier is reduced. The small bowel and large bowel have a considerable reserve function, and aging *per se* is unlikely to affect this unless of course diseases develop.

What is Optimal BMI in Old Age?

With this question comes that of 'who defines the norm'? Tables most widely available for adults are based on young white volunteers. There are only a few longitudinal data available at present upon which to base BMI advice for elders in any population, and tables are mostly norms using percentiles based on cross-sectional studies, point surveys, and 'healthy' elderly volunteers rather than on health outcomes, unfortunately. As with all BMI descriptive data, the BMI norms have increased, particularly in Westernized societies, first in the USA and now in Europe. The norms established in the 80s were apparently no longer valid in the 1990s. The definition of 'norms' is particularly problematic in old age: in a typical UK community free-living volunteer elder population, only 5% of the over-65s have no medical problem. If all illnesses were excluded in establishing norms, these tables would be based on 5% of the elderly population – who could thus be described as super-fit elders. Compared with our knowledge about younger people, this state of ignorance regarding old people and BMI needs explaining, and may reflect the lack of research investment by public bodies and private companies.

Population Studies

A 1996 review on BMI in old age relied on data from a 1979 survey (Build Study) and had no people aged over 69 (see Table 2). What can be seen from this table, however, is that the optimum BMI in terms of mortality apparently rises very significantly with age. In Finland, a whole population local survey in Tampere of people aged 85 years and older showed a correlation between longer survival and higher BMI (see Figure 2).

A very large US survey of 22 000 subjects also showed a flattening of the U-shaped curve relating BMI to mortality with increasing age cohorts. The obesity risks of vascular disease and cancer do not disappear in old age, but it has been argued that selective survival of those less vulnerable to these risks occurs. The Nottingham Longitudinal Study of Activity and Aging, which measured BMI and health in 629 relatively well but not super-fit elders living in the community at the time of the initial survey, interestingly showed neither an increase in morbidity nor an increase in mortality in the low BMI volunteers initially and at 4-year follow-up. One possible explanation for this apparent anomaly is that whereas most population surveys include significantly disabled or ill elderly subjects, if these people are excluded, thinness has no apparent ominous significance, at least not over 4 years. This fits with the emphasis many authors and clinicians have put on nutrition *reserve* in elders and the importance of regaining this reserve after intercurrent illness or trauma. Nevertheless, the subject cannot be said to be fully understood as yet, and confounders are very likely to be significant and to vary from sample to sample. What the data do suggest, however, is that if old people are fit, well, and mobile, thinness has little ominous significance, but in any studies that include hospital attenders, it appears that thinness is a very powerful predictor of poor outcome.

In terms of disability prevention, the data from the US National Health and Nutrition Examination

Table 2 Effect of age on body mass index associated with the lowest mortality[a]

Age group (years)	*Body mass index (kg m^{-2})*	
	Men	*Women*
20–29	21.4	19.5
30–39	21.6	23.4
40–49	22.9	23.2
50–59	25.8	25.2
60–69	26.6	27.3

[a]Data from Build Study 1979.

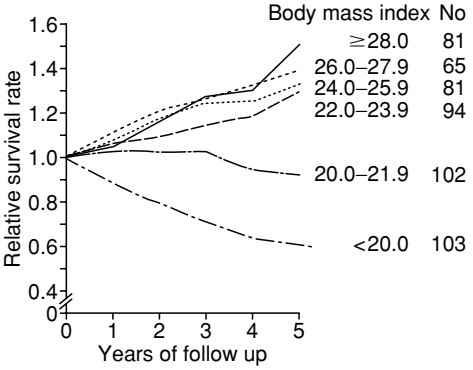

Figure 2 Five-year relative survival curves according to body mass index in subjects aged 85 and over.

Survey (NHANES) in 1994 supported the U-shaped curve hypothesis: *both* extremes of BMI in elderly volunteers apparently independently predicted worse functional status (independence in activities of daily living) (Figure 3). In the old group of women who had lost more than 5% of their weight over 8–16 years, there was an associated doubling in the rate of mobility disability, even after adjustments were made for age, smoking, and education. Weight loss rather than thinness was again the main predictor of poor outcome. In a detailed description of the NHANES follow-up, it was found that the initial BMI at both 15% extremes was an independent predictor of disability in old men and old women, even 11–13 years later, even after correcting for 22 potential confounders, and even after excluding those who died within 2 years of the first survey (in whom, it might have been argued, weight loss was a result of illness).

BMI ranges are therefore a changing phenomenon over time both within and between elderly populations. At present, outcome data are lacking on which to base advice to an elderly person regarding optimum weight, except that overweight is clearly related to joint pains. Survival is reduced in thinner elders overall, but a high waist–hip ratio is probably not advisable in the presence of hypertension, vascular disease, or diabetes.

Table 3 Factors associated with high-risk undernutrition in old age

Environmental causes
Poverty
Social isolation
Depression
Social construction of disability (being housebound)
Individual factors
Poor dentition
Chronic lung disease
History of gastric surgery
Cognitive impairment
Dysphagia
Bereavement and depression
Alcohol excess
Recent unintended weight loss of 3 kg

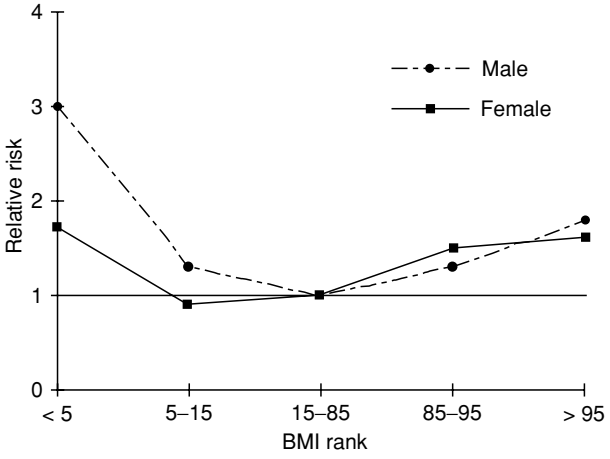

Figure 3 Relationship between body mass index percentile rank and relative risk for functional impairment for men and women. Final logistic regression model after removing from the sample those 307 persons who died within 2 years of the completion of the NHANES-I Epidemiologic Follow-up Study ($n = 2746$).

Hospital Patients

Surveys of nutritional status and health in old age that include disabled, ill, or traumatized elderly patients consistently highlight thinness and recent weight loss as powerful ominous independent predictors of health outcome. Although, with different methods, hospital inpatient surveys yield a wide range of prevalences for undernutrition in old hospital inpatients, in the UK, 40% is accepted by Government Department documents. The commonness of thinness in old inpatients may explain why staff (unlike relatives) usually miss the diagnosis and has prompted the growth of screening tools of varying efficacy. Research suggests that, so far, no questionnaire screening tools to be administered by nurses on general wards are safe, unlike their use by dieticians, and safer alternatives are either anthropometry or simply an eyeball estimation of whether the patient is thin, average, or fat: the nurse is asked to tick the relevant box. (*See* **Nutritional Assessment**: Functional Tests.)

The association and ominous significance of malnutrition with illness and disability are likely to be complex. In some cases, the illness or disability, or their attendant pharmacological and social consequences, may precede and cause negative energy balance. In others, the causal relationship may be reversed. In the clinical context, therefore, it is crucial to tackle both the illness and the malnutrition, if appropriate, and, to avoid recurrence, an attempt should be made to determine the sequence of events in each individual: sometimes, a third factor will have caused both the illness and the malnutrition simultaneously. In a study on very acute illnesses leading to old people's admission, using retrospective dietary histories, it was found that, compared with control groups, these patients had had lower dietary intakes over the preceding month – even though the criterion for inclusion was that the patients were absolutely fine the day before the day of admission. This suggests that

undernutrition may cause or precede very acute illness. Of course, in the clinical context, the reverse is also often found.

Frequently, depression and physical illnesses present simultaneously in the elderly, and in many, a vicious circle is present by the time treatment starts: malnutrition can perpetuate anorexia by increasing small bowel microbial overgrowth, which is even more likely to follow undernutrition in the elderly than in children because of the high prevalence of age-associated achlorhydria (in atrophic gastritis, now known to be due to *Helicobacter pylori* infection) in elderly patients. When depression and anorexia are both very marked and coexist in old age, endoscopy is indicated. In stroke patients, weakness in swallowing is commonly compounded by starvation from the time of stroke, and tube feeding is required to correct this starvation weakness of the bulbar muscles.

Negative energy and nitrogen balance very quickly and severely impair immune function, even in fit elderly volunteers, and even within 9 weeks of starting on an equal-calorie but low-protein diet. In the clinical context, it has been demonstrated that a low lymphocyte count, which can have numerous causes (e.g., inflammation, undernutrition), correlates closely with the risk of readmission of the patient.

What is less well known in elderly care is the importance of careful management of the refeeding syndrome, and work is awaited to see whether this and consequent hypophosphatemia are being missed, as is our clinical impression, in a number of cases of elderly patients in hospital. (*See* **Elderly**: Nutritionally Related Problems; Nutritional Management of Geriatric Patients; **Phosphorus**: Properties and Determination; Physiology.)

In rehabilitation wards for the elderly, research has confirmed very elegantly that poor nutritional status on admission and, in particular, weight loss during the preceding year are the most powerful independent predictors of poor outcome, even after correcting for numerous potential confounders, including 22 that did correlate with outcome. Equally important randomized controlled trials of nutritional supplementation have demonstrated lasting positive outcomes in elderly hip fracture patients and pneumonia patients: functional ability in the long term benefitted from the intervention, and average hospital stay was shortened.

When viewed together, these data explain the primary concern regarding undernutrition in elderly patients and consequent rejection of blanket obesity-avoidance advice intended for other adults, which is sometimes inappropriately followed by old patients.

-Specific nutritional deficiencies in hospital patients
Certain micronutrient deficiencies in old inpatients denote specific pathological conditions. Iron deficiency in old age (in particular postmenopause) in developed countries is due to gut pathology (e.g., ulcer or cancer) in over 90% of cases, according to two London studies. Vitamin D deficiency may be due to anticonvulsant medication, folate deficiency may reflect anticonvulsant medication but, more commonly, inadequate diet due to poor self-care, and this is common in early dementia. It was so common that it was thought to cause dementia, although, at present, the consensus appears to be that early dementia leads to difficulty organizing food if people are living on their own, and the first lab routine test to show this is folate. Folate deficiency is also very commonly associated with alcoholism. (*See* **Elderly**: Nutritionally Related Problems.)

Immobility itself in old age has been shown to be associated with more rapid protein turnover than is seen in fit peers: it was demonstrated long ago that, in the UK, housebound elderly people are in negative nitrogen balance. If protein turnover is more rapid in the elderly during immobility (and endocrine mechanisms are possible), a slight imbalance of intake or expenditure will lead more rapidly to more muscle loss while the body prioritizes visceral protein synthesis. This may explain the downward spiral of frailty sometimes seen in ill, traumatized or disabled patients.

Nutritional Status in Elderly Nursing Home Residents

Undernutrition is very common in this group. Many residents barely eat enough to maintain basal metabolism. The reason for poor nutritional status in advanced dementia may vary, and research has produced a range of answers, including insufficient time spent helping at meal times, inappropriate food consistency, and possibly a change in energy balance control. Drug effects on appetite or bowels and nutrient utilization also need to be considered in each case. Energy expenditure generally drives energy intake in humans – if an old person is less mobile, their appetite will therefore be reduced and micronutrient deficiencies will, and often do, result if the nutrient density of the food is not appropriately increased. Vitamin D deficiency is common in housebound elders without access to sunlight, and sufficient vitamin D consumption would require an unacceptably high intake of fatty fish. Quite why there are such high rates of general undernutrition in this group, however, is unclear, even if the individual micronutrient deficiencies are easier to explain. (*See* **Elderly**: Nutritionally Related

Problems; Nutritional Management of Geriatric Patients.)

In 1995, the Caroline Walker Trust produced multidisciplinary guidelines for optimal nutrition for institutionalized elderly. It is written to be understood by service users and service providers, and covers food provision, dental checks, weight monitoring, etc. plus triggers for referral for medical review (any unexplained 3-kg weight loss). The importance of formulating an individualized plan suitable for each elderly resident is emphasized.

Ethical questions regarding the degree of intervention considered appropriate in terminal illness and in dementia frequently occur in this setting. In several observational studies in nursing homes investigating the management of advanced dementia patients who appear not to be eating enough, tube feeding produced no benefit in terms of comfort and none in terms of survival.

See also: **Elderly**: Nutritionally Related Problems; Nutritional Management of Geriatric Patients; **Nutritional Assessment**: Functional Tests; **Phosphorus**: Properties and Determination; Physiology

Further Reading

Campbell AJ and Buchner DM (1997) Unstable disability and the fluctuations of frailty. *Age and Ageing* 26: 315–318.

Caroline Walker Trust (1995) *Eating Well for Older People. Report of an Expert Working Group. Practical and Nutritional Guidelines for Food in Residential and Nursing Homes and for Communal Meals*. London: Caroline Walker Trust.

Ebrahim S and Kalache A (eds) (1996) *Epidemiology in Old Age*. London: BMJ Publishing Group.

Galanos AN *et al*. (1994) Nutrition and function: is there a relationship between body mass index and the functional capabilities of community dwelling elderly. *Journal of the American Geriatrics Society* 42: 147–168.

Gillick MR (2000) Rethinking the role of tube feeding in patients with advanced dementia. *New England Journal of Medicine* 342: 206–210.

Lehmann A (1996) Nutrition and health – what is new? *Reviews in Clinical Gerontology* 6: 147–168.

Rabeneck L, McCullough LB *et al*. (1997) Ethically justified, clinically comprehensive guidelines for percutaneous endoscopic gastrostomy tube placement. *Lancet* 349: 496–498.

Roth SG *et al*. (1999) Calorie restriction in primates: will it work and how will we know? *Journal of the American Geriatrics Society* 7: 896–903.

Tallis R (ed.) (1998) *Increasing Longevity: Medical, Social and Political Implications*. London: The Royal College of Physicians.

Nutritionally Related Problems

J B Blumberg, Tufts University, Boston, MA, USA

Introduction

Nutrition influences the progressive loss of tissue function with aging and plays an important role in the genesis of several major chronic degenerative diseases. Aging influences nutritional requirements and nutritional status through changes in appetite and in pathways of nutrient absorption, utilization, and excretion (**Table 1**). Further, a variety of socioeconomic and psychological factors as well as the presence and treatment of current disease(s) common to older adults affect dietary patterns and nutritional status. Thus, some of the nutritional problems of the elderly are the consequence of long-term (chronic) interactions between diet and aging, beginning in adolescence and young adulthood, while others are associated with the more immediate (acute) result of old age itself.

Acute nutritionally related problems

Anemia

The most consistently reported hematopoietic change with aging is a higher prevalence of anemia which increases with each successive decade of life. Although anemia is prevalent among healthy elderly, its cause does not generally appear to be due to hemolysis or deficiencies of iron, folate, or vitamin B_{12}. Elderly people with this unexplained anemia have lower leukocyte and neutrophil counts than do nonanemic subjects. Experimental and clinical studies

Table 1

Change in physiologic function	*Change in nutrient requirement*
↓Muscle mass (sarcopenia)	↓Calories
↓Bone density (osteopenia)	↑Vitamins D and K, calcium
↓Immune function (anergy)	↑Vitamins B_6, C, and E, zinc
↓Skin cholecalciferol synthesis	↑Vitamin D
↑Winter parathyroid hormone production	↑Vitamin D
↓Calcium bioavailability	↑Vitamin D
↓Hepatic uptake of retinol	↓Vitamin A
↑Oxidative stress status	↑β-carotene, vitamins C and E
↑Levels of homocysteine	↑Folate, vitamins B_6 and B_{12}
↓Efficiency in pyridoxal utilization	↑Vitamin B_6
↑Gastric pH (gastric atrophy)	↑Vitamin B_{12} and folic acid, calcium, iron, zinc

suggest this anemia of aging is due to a reduction in bone marrow reserve capacity (erythropoietin-dependent progenitor cell proliferation) with decreased hormonal responsiveness to hematologic stress. Chronic inflammation is also a possible mechanism. However, any elderly individual with anemia must be fully evaluated for the possibility of drug-induced or occult gastrointestinal blood loss, bone marrow failure, chronic hemolysis, and nutritional deficiency before attributing an anemia to age alone.

Anorexia and Weight Loss

A decline in calorie intake is a well-recognized concomitant of increasing age. Surveys reveal that almost 20% of older adults consume less than 1000 calories (4.2 MJ) per day; among institutionalized elderly and those whose income is at poverty levels, the prevalence may increase twofold. Protein-calorie malnutrition is a common finding among geriatric patients. Marasmus may occur in older patients because the combination of a small but balanced caloric intake with a contribution from muscle-protein catabolism is sufficient to maintain serum albumin. Thus, weight loss is an ominous sign in the elderly which needs to be taken seriously.

The etiology of anorexia and weight loss in the elderly is multifactorial and includes both physiological and pathological causes. Age-related declines in metabolic rate due to loss of lean body mass (sarcopenia) and decreased expenditure of energy on physical activity contribute to lower calorie intakes in the elderly. Social factors such as poverty, isolation, and requirements for assistance with food shopping and preparation can be a source of impairment to adequate nutrition. Physical factors such as disabilities or immobility and poor oral health can also interfere with conducting activities of daily living associated with nutrition.

Psychiatric disorders, e.g., dementias and depression, as well as states associated with bereavement and isolation affect nutritional status. These conditions are not uncommon among the elderly and are usually associated with anorexia and weight loss. Among institutionalized elderly with low body weight and at risk of malnutrition or B vitamin deficiency, a higher prevalence of dementia and depression is noted. Indifference, memory loss, impaired cognition, and disturbed or erratic behavior associated with these conditions all contribute to reduced food intake. Inadequate nutrition may further exacerbate psychiatric symptoms. Experimental studies suggest an increased activity of the hypothalamic–pituitary–adrenal axis production of corticotropin-releasing factor and/or neuropeptide Y may be involved in the anorexia of some of these psychiatric conditions.

Several diseases common among the elderly affect food intake. Chronic obstructive pulmonary disease increases the effort required to eat. Congestive heart failure is associated with anorexia which is aggravated by the cachexic effect of cardiac glycosides used to treat the condition. Some cancers and all antineoplastic drugs can precipitate anorexia. Esophageal reflux can lead to pain on swallowing and decreased food intake.

Anorexia may also occur as a concomitant of normal aging. Gustatory papillae begin to atrophy in middle age and taste and olfactory thresholds increase with age. Studies in old animals indicate a reduced drive in the endogenous opioid feeding system, particularly the κ-receptor ligand dynorphin, and the increased satiety action of the gastrointestinal hormone cholecystokinin may provide a biochemical basis for the anorexia of aging.

Constipation

Constipation is a common complaint among the elderly, with women reporting this difficulty more frequently than men. The discomfort associated with constipation may affect an older person's food intake. The tone of intestinal smooth muscles and gastrointestinal tract activity diminish with age. However, the lack of dietary fiber, dehydration, and/or immobility appear to be the primary causes of most cases of uncomplicated constipation. A variety of drug treatments common among elderly patients, particularly those with anticholinergic actions, may also cause severe constipation. The requirement for assistance in using the toilet may cause an older adult not to respond to the urge to defecate which can result in a later inability to defecate at will due to desensitization of rectal stretch receptors. Important complications of constipation in the elderly include fecal impaction, abnormally large colon, and intestinal obstruction. Also, straining at stool may create sudden severe circulatory changes in the aged. In most instances, the treatment of simple constipation in the elderly includes advice to increase fiber consumption (and fluid intake) in the diet and/or as supplements (e.g., vegetable mucilloids and modified cellulose products). The pathogenesis of another age-related condition of the colon associated with the absence of fiber in western diets is diverticular disease.

Dehydration

Thirst and water intake are critical to the maintenance of fluid and electrolyte balance since it is only through such intakes that water deficits can be

replenished. Fluid and electrolyte disorders are common in the elderly and abnormalities in salt and water homeostasis are a significant cause of morbidity and mortality in geriatric patients. Older adults respond to dehydration with greater increases in plasma sodium concentration and osmolality than young subjects, yet do not become markedly thirsty and drink enough to dilute their body fluids to predeprivation levels. Thirst in response to heat stress and thermal dehydration is also reduced in the elderly despite a reduced capacity for thermoregulation (heat loss).

The mechanism of the thirst deficit in the elderly is not fully defined. Age-related changes in receptors that detect dehydration of the cell (osmoreceptors) and decreases in plasma volume (baroreceptors) and in the endogenous opioid system have been suggested to play a role in this condition. Changes in renal function with age, such as diminished glomerular filtration rate, urinary concentrating ability, and sodium-conserving capacity, can act to reduce the ability to conserve salt and water and predispose an older person to hypovolemia and dehydration. The decreasing activity of the renin–angiotensin–aldosterone system with age may also contribute to diminished sodium and water homeostasis in the elderly.

Older adults may become dehydrated when exposed to even mild stresses such as water restriction, fever, infection, or diarrhea. When the diminished thirst response in an elderly individual is accompanied by illness or physical incapacity that increases water loss or prevents access to water, dangerous dehydration may follow. In the elderly, dehydration is a common cause of renal impairment and failure and electrolyte disturbances. Dehydration and hyponatremia can also lead to confusional states that further interfere with fluid intake. While the extent of dehydration in free-living elderly has not been established, age-related impairments of thirst sensation place them at risk of dehydration.

Gastric Atrophy

Selective functions of the digestive tract experience some diminution of function with age, although many of these changes do not appear significantly to influence nutrient requirements in the elderly. An increased incidence of gastric acid hyposecretion (hypochlorhydria) resulting from atrophy of the gastric glands of the fundic mucosa of the stomach is noted among about 30% of people over 60 years. This gastric atrophy (type A) is asymptomatic but, when the condition is severe, is associated with pernicious anemia due to a decreased secretion of intrinsic factor. Gastric atrophy is also associated with faster gastric emptying rates and a higher pH with bacterial overgrowth in the proximal small intestine. These latter effects of gastric atrophy may interfere with the bioavailability of several nutrients, including calcium, iron, folic acid, and vitamin B_{12}. In addition to absorbing or utilizing these nutrients, the overgrowth of some aerobic organisms may contribute to the nutritional status of the host via the synthesis and secretion of some nutrients, e.g., vitamin K. The high intraluminal pH of the stomach and proximal small intestine found in gastric atrophy may also prevent the release of some nutrients from dietary fiber or protein and other food complexes. Dietary recommendations for several micronutrients should be adjusted for older adults with gastric atrophy and include the use of appropriate nutrient supplements.

Chronic Nutritionally Related Problems

Cancer

Cancer develops as a result of interactions between genetic and environmental factors. An estimated 90% of all cancer cases are attributed to environmental factors, including tobacco smoking, radiation, industrial pollutants, viruses, and diet. Analyses of epidemiological studies suggest that 35% of cancer deaths may be related to diet. Evidence linking diet and cancer has been derived from experimental *in vitro* and animal studies, metabolic and biochemical studies in humans, and epidemiological investigations. In general, the epidemiological studies have demonstrated that populations migrating from an area with its own pattern of cancer incidence rates acquire rates characteristic of their new location. Further, large changes in incidence rates for many types of cancer within genetically stable populations have been associated with changes in dietary patterns. Although the cumulative data for the relationship between both the causal and protective effects of diet and cancer are compelling, the evidence, particularly with regard to quantitative levels of specific nutrients, remains far from conclusive.

Various studies indicate that a high intake of total fat increases the risk of some types of cancer, notably cancer of the breast and colon but also of the prostate, rectum, and ovaries. Cancer mortality correlates better with total dietary fat intake than with the type of fat. However, several epidemiological studies have demonstrated a strong relationship between saturated fatty acids or animal fat and increased cancer risk. In contrast, animal studies indicate monounsaturated and *n*-6 polyunsaturated fats promote carcinogenesis more effectively, while *n*-3 fatty acids can reduce tumor growth rates. Several mechanisms have been proposed to explain the effects of fat on

cancer, including its alteration of steroid hormones and endocrine changes, immune function, cell membrane fluidity, and (for colon cancer) fecal floral and bile acids. Excess body weight and obesity are also associated with an increased risk of several cancers, including those of the breast, endometrium, ovaries, colon, rectum, prostate, and gallbladder. Other dietary factors, including alcohol and naturally occurring contaminants, e.g., aflatoxins and *N*-nitroso compounds, also pose a potential cancer risk.

Dietary patterns emphasizing foods high in fiber are associated with low rates of certain cancers, especially breast and colon cancer. The protective effect of fiber may be due to a concomitant reduction of total calories and fat and/or by altering gastrointestinal factors such as intraluminal pH, gut flora, and transit time. Fiber-containing foods such as fruits and vegetables are also important sources of β-carotene, vitamins A, C, and E, phenols, and indoles as well as a large number of other phytochemicals with potential cancer-preventive activity. Generous dietary intakes of β-carotene are associated with a reduced incidence of lung, breast, oral mucosa, bladder, and esophageal cancers; however, clinical trials with large supplemental doses of β-carotene resulted in an increased risk of cancer among smokers. Vitamin C has a potentially protective effect against esophageal, stomach, cervical, breast, and lung cancers. Low intakes of vitamin E are strongly correlated with risk of cancer in many, but not all, organs. β-carotene and vitamins C and E may share a related antioxidant, free radical scavenging mechanism against carcinogenesis. Evidence also suggests that calcium and selenium have protective effects against some forms of cancer.

While the evidence associating specific nutrients with cancer risk remains equivocal, dietary guidelines have been promulgated by a number of health organizations which recommend avoiding obesity, reducing total fat to 30% or less of calories, consuming high-fiber foods and a variety of fruits and green and yellow vegetables in the daily diet, and minimizing the intake of alcohol and smoked, salt-cured, and nitrate-cured foods. Evidence suggesting the ability of these dietary patterns or supplementation with putative protective nutrients to reverse existing cancers or lower rates of cancer recurrence is limited and equivocal.

Established cancers as well as antitumor treatments may have a variety of effects on the nutritional status of the affected geriatric patient. Anorexia and weight loss are common and may result in progressive wasting and undernutrition (cancer cachexia). Alterations in protein, carbohydrate, and fat metabolism are associated with many cancers. Malabsorption and protein-losing enteropathy are also a concomitant of various malignancies. Hormonal abnormalities induced by tumors may affect the status of different nutrients. Radiation, surgery, and drug therapies each have consequences which predispose the patient to potentially serious nutrition problems.

Vision Disorders

Aging is associated with extensive postsynthetic modifications to eye lens proteins resulting in aggregation with eventual polymerization and precipitation to form opacities or age-related cataracts. The prevalence of cataract increases with age, reaching about 50% in people 75–85 years old; prevalence rates are higher if early lens changes are included in the calculation. Oxidation of lens proteins is highly correlated with cataract and research studies provide evidence for a close relationship between cataractogenesis and nutrition. An elevated status of dietary antioxidants is strongly correlated with a lower incidence of cataract. Ascorbic acid concentration in the lens can be as much as 30 times that in plasma but decreases in aged and cataractous lens. Vitamin E is associated with protection against oxidative insults and glutathione-dependent antioxidant reactions are compromised in the lens with age and cataracts. The effect of other nutrients, e.g., folic acid, riboflavin, vitamin B_6, taurine, and tryptophan, has been examined but evidence supporting their protective role against cataractogenesis remains limited. Intake of the xanthophyll carotenoids, lutein and zeaxanthin, is inversely correlated with the incidence of cataract and cataract extractions.

Age-related macular degeneration (AMD) is the leading cause of irreversible blindness among older adults. The retina is rich in highly polyunsaturated fatty acids, particularly docosahexaenoic acid, and thus vulnerable to lipid peroxidation. This situation is compounded by exposure to light, high oxygen tension, and high concentrations of retinol. The macula contains a full complement of antioxidant defenses but is unusually rich in lutein and zeaxanthin, which provides its yellow pigment and optical density, and absent in β-carotene and lycopene. Observational studies indicate that generous intakes of lutein and zeaxanthin are associated with a significant reduction in risk of AMD. Zinc plays a role in the metabolic function of several antioxidant enzymes in the chorioretinal complex and limited evidence suggests increased zinc intake may benefit visual acuity in AMD patients.

Coronary Heart Disease

The progressive focal narrowing of coronary arteries by atherosclerosis, beginning as early as age 10, leads

to an inadequate circulation of blood in local areas of heart muscle. These changes can result in angina pectoris, myocardial infarction (heart attack), and sudden death in older adults. These and related disorders are termed coronary heart disease (CHD) and, in western countries, account for more deaths annually than any other group of diseases. The prevalence of CHD is greater among men than women and increases markedly in those over 65 years of age. Medical interventions and changes in lifestyle during the past two decades have reduced the age-corrected death rate from CHD by about 40%; however, without a further reduction in risk factors and/or an improvement in treatments, CHD prevalence and incidence will increase in the future because of the aging of the population. The causes of CHD are multifactorial and include cigarette smoking, high blood pressure, obesity, and high blood cholesterol (hypercholesterolemia).

The relationship between diet, lipoprotein metabolism, atherogenesis, and CHD has been defined by extensive scientific investigations conducted during the last 75 years. This work indicates a strong, continuous, and graded relationship between total blood cholesterol levels, the severity of atherosclerosis, and risk for CHD. Dietary saturated fat and cholesterol raise total blood cholesterol and low-density lipoprotein (LDL) cholesterol levels; LDL is the major atherogenic lipoprotein and accounts for most of the CHD risk associated with elevated plasma total cholesterol. Polyunsaturated and monounsaturated fat lowers total blood cholesterol and LDL cholesterol levels when substituted for saturated fat in the diet. High-density lipoprotein (HDL) is considered to protect against CHD and is associated with female sex, estrogen use, exercise, moderate alcohol consumption, and weight loss.

Diet is considered the cornerstone of therapy to reduce borderline (200–239 $mg\,dl^{-1}$) and high-risk ($\geq 240\,mg\,dl^{-1}$) blood cholesterol levels whether or not drug therapy is eventually added to the regimen. Dietary guidelines to reduce risk of CHD include recommendations of nutrient intake as a percent of total calories of less than 30% from fat (less than 10% from saturated fat, up to 10% from polyunsaturated fat, and 10–15% from monounsaturated fat), 50–60% from carbohydrates, and 10–20% from protein. Total calorie intake should be designed to achieve and maintain desirable weight. In addition, daily cholesterol intake should be less than 300 mg. Individuals at very high risk of CHD should further reduce saturated fat and cholesterol intake. Adherence to these dietary guidelines has been proven to lower blood cholesterol by 10–25%. Elevated plasma homocysteine is an independent risk factor for CHD. Homocysteine concentrations can be reduced by increasing the intake of folic acid, vitamin B_6, and/or vitamin B_{12}; however, prospective studies demonstrating a decreased incidence of CHD with B vitamin interventions have not been completed. Supplementation with vitamin E and/or fish oil may play a role in the primary and secondary prevention of CHD but the results of clinical trials are equivocal.

Diabetes

Type 2 diabetes, formerly called noninsulin-dependent diabetes mellitus, accounts for about 90% of all cases of the disease. The onset of type 2 diabetes is gradual (with no or mild symptoms), usually appearing in midlife among people who are overweight or obese. Although type 2 diabetics display normal pancreatic histology, secrete insulin in response to glucose ingestion, and have normal to high levels of circulating insulin, they are less able to use this insulin (insulin resistance). Type 1 diabetes, formerly called insulin-dependent diabetes mellitus, is characterized by an absolute deficiency of insulin caused by β-cell destruction of the islets of Langerhans in the pancreas. Type 1 diabetes usually begins abruptly in adolescents and young adults of normal body weight. No method has been identified which prevents type 1 diabetes and these patients require insulin administration for survival. Genetic predisposition appears to play an important role in both type 1 and type 2 diabetes. The specific cellular basis of hyperglycemia and related metabolic abnormalities in type 2 diabetes is not yet clear, although three defects appear to be involved: a diminished pancreatic β-cell response to glucose; a reduced synthesis or function of cellular receptors for insulin; and an impaired ability to utilize glucose following binding to cellular receptors. Clinical complications of diabetes include heart disease and stroke, kidney disease, blindness, nontraumatic leg and foot amputations, and death.

Dietary therapy is considered the key to the management of type 2 diabetes. Obesity is strongly associated with the onset and severity of type 2 diabetes with more than 80% of patients being more than 15% in excess of their desirable body weight at the time of diagnosis. Weight loss reduces insulin resistance as well as fasting and postprandial blood glucose levels in overweight diabetics. Dietary recommendations for the management of diabetes include: 50–60% of total calories as carbohydrate (emphasizing complex carbohydrates and water-soluble fiber); decreased protein in those with renal insufficiency; 30% or less of total calories as fat (with reduced intake of saturated fat); and limited alcohol intake (restricted entirely in poorly controlled diabetes). Limited and

equivocal evidence suggests a potential role for vitamins A, C, E, and niacin and the minerals chromium and vanadium in diabetes. Alternative sweeteners, e.g., sorbitol and saccharin, are accepted in diabetes management but their value is unclear. The control of hyperglycemia by diet and medication is associated with reduced risk of some of the long-term complications of diabetes.

Obesity

The determination of obesity in the elderly is problematic since few body weight standards specific to this age group are in use. The prevalence of obesity appears to increase with age. Weight gain normally occurs during the adolescent growth spurt and again during middle age. Basal metabolic rate and energy expended for physical activity decline during adulthood so that fewer calories are required to maintain energy balance; however, many middle-aged and older adults often do not adjust their caloric intake to compensate for their reduced energy requirements and gain weight even though they do not eat more than before. Body weight generally peaks between 35 and 55 years in men and between 55 and 65 years in women; body weight decreases thereafter in both sexes. Obesity aggravates many health conditions found among the elderly including arthritis, cardiovascular diseases and hypertension, and type 2 diabetes.

Osteoporosis

Primary osteoporosis is an age-related disorder characterized by decreased bone mass and increased susceptibility to fractures in the absence of other recognizable causes of bone loss. Postmenopausal osteoporosis (type 1) occurs in women within 15–20 years after menopause and is thought to result from factors related to or exacerbated by estrogen deficiency. Age-related osteoporosis (type 2) occurs in men and women over 75 years of age and may be more directly related to the aging process. Whites are at higher risk of osteoporosis than blacks. Other risk factors include dietary factors, cigarette smoking, physical inactivity, alcohol abuse, being underweight, and family history of osteoporosis. Osteoporosis is more than six times as common among women than men due to their smaller peak bone mass, earlier onset of bone loss, longer life expectancy, and lower calcium intake.

Substantial evidence supports a role for calcium as a protective agent against osteoporosis. High calcium intake during early years contributes to greater peak bone mass and during later years prevents negative calcium balance and reduces the rate of bone loss. Poor calcium nutriture is common among the elderly due to inadequate intake, an age-related decline in the capacity for calcium absorption, and the presence of diseases associated with reduced calcium absorption and/or increased calcium excretion. Vitamin D is essential for the efficient absorption of calcium but vitamin D status is quite low among the elderly due to inadequate intake and age-related impairments in the metabolic conversion of vitamin D to its biologically active form, including lower precursor concentrations of 7-dehydrocholesterol in the skin. Less well-established dietary risk factors for osteoporosis include low intakes of vitamin K or boron and high consumption levels of caffeine, protein, or alcohol. The principal measures employed in the prevention of osteoporosis include estrogen replacement in postmenopausal women, elemental calcium intakes of 1000–1500 mg day^{-1}, vitamin D intakes of 10–20 μg cholecalciferol day^{-1}, and a program of modest weight-bearing exercise. These measures may retard further bone loss in the elderly but will not restore lost bone or reverse the loss in height or collapsed vertebrae characteristic of osteoporosis.

See also: **Anemia (Anaemia)**: Other Nutritional Causes; **Anorexia Nervosa**; **Cancer**: Epidemiology; Diet in Cancer Prevention; **Coronary Heart Disease**: Etiology and Risk Factor; Prevention; **Dehydration**; **Dietary Fiber**: Effects of Fiber on Absorption; **Lipoproteins**; **Obesity**: Etiology and Diagnosis; Epidemiology; **Osteoporosis**

Further Reading

Bendich A and Deckelbaum R (1997) *Preventive Nutrition, The Comprehensive Guide for Health Professionals*. Totowa, NJ: Humana Press.

Chernoff R (1999) *Geriatric Nutrition*. Rockville, MD: Aspen.

Graham I, Refsum H, Rosenberg IH and Ueland PM (1997) *Homocysteine Metabolism: From Basic Science to Clinical Medicine*. Boston: Kluwer Academic.

Harman D, Holliday R and Meydani M (1998) *Towards Prolongation of the Healthy Life Span*. New York, NY: New York Academy of Sciences.

Holliday R (1995) *Understanding Ageing*. New York, NY: Cambridge University Press.

Rosenberg IH (1995) *Nutritional Assessment of Elderly Populations: Measure and Function*. New York, NY: Raven Press.

Rowe JW and Kahn RL (1998) *Successful Aging*. New York: Pantheon Books.

Schneider EL and Rowe JW (1996) *Handbook of the Biology of Aging*. Orlando, FL: Academic Press.

Sies H (1997) *Antioxidants in Disease Mechanisms and Therapy*. San Diego: CA: Academic Press.

Taylor A (1999) *Nutritional and Environmental Influences on the Eye*. Boca Raton, FL: CRC Press.

Nutritional Management of Geriatric Patients

M S J Pathy, Cyncoed, Cardiff, UK
A Bayer, Llandough Hospital, Penarth, UK

Background

At all ages, adequate nutrition is critical for maintaining health and quality of life. In general, dietary protein and energy requirements are substantially less in the over 85s than those aged 65–74 years, and there is an increased need for certain vitamins and minerals. However, individual requirements are influenced greatly by lifestyle and by disease. Maintenance of physical fitness and strength in old age largely depends on continued physical and mental activity, with enough to eat and a diet that is sufficiently varied. Elderly people in low socio-economic groups, those in institutions, and the housebound who are receiving home care are especially vulnerable to undernutrition, through underprovision and inadequate intake. Drugs and disease may also interact with the complex metabolic processes necessary for breaking down and absorption of food. Increasingly, calorie overconsumption is a problem in Western society, with 25% of men and 35% of women in the 65–74 age range being reported as overweight and 10% being severely overweight.

In the UK, the Committee on the Medical Aspects of Food and Nutrition Policy (COMA) laid down general guidelines for dietary requirements of older people together with modifications in disease states. Table 1 summarizes the dietary reference values (DRVs), which replace the earlier recommended dietary allowances (RDA). The reference nutrient intake (RNI) represents the sum of the estimated average requirement (EAR) and two standard deviations.

Energy

A reduction in physical activity and lean body mass in old age lead to a decline in energy requirements by about one-third. However, there is no reduction in basal metabolic rate relative to the metabolically active cell mass. Infection, trauma (accidental or surgical), and the ergonomic impact of disability on daily living tasks all serve to increase energy demands. Dyspnea at rest or exertion (commonly seen in patients with chronic pulmonary obstructive airways disease) and impaired immunological response to infection also require increased energy intake as part of therapeutic management.

COMA recommends that 50% of food energy is derived from carbohydrate – a term that embraces starch and nonstarch polysaccharides. The main source of starch is influenced by ethnicity and culture, with bread, potatoes, breakfast cereals, rice, and pasta predominant in the UK. Distinction is made between intrinsic sugars (milk sugar, and sugar incorporated in the cellular structure of fruit and vegetables) and extrinsic nonmilk sugars (extracted from the fruit stem and roots of plants and added to food, confectionery, and soft drinks). A high extrinsic sugar intake may reduce the dietary quality and in the dentate increases the risk of caries.

Different carbohydrates have different glycemic indexes (Table 2). This is the area of the blood sugar curve produced by a food, compared with the equivalent amount of glucose. Glucose has an arbitrary glycaemic index of 100. Previously, white bread was taken as the standard. Excess high glycemic index carbohydrates are associated with a low high-density-lipoprotein (HDL) cholesterol concentration,

Table 1 Dietary reference values for older men and women (based on COMA recommendations)

Nutrition	*Sex/age (years)*	*Estimated average requirements (EAR)*	*Reference nutrient intake (RNI)*
Energy[a]	M 65–74	9.71 MJ	
	M 75+	8.77 MJ	
	F 65–74	7.96 MJ	
	F 75+	7.61 MJ	
Protein	M 50+		53.3 g
	F 50+		46.5 g
Vitamin A	M		700 μg
	F		600 μg
Vitamin D	M/F		10 μg
Vitamin C	M/F		40 μg
Thiamin	M		1.0 mg
	F		0.8 mg
Riboflavin	M		1.3 mg
	F		1.1 mg
Niacin	M		16 mg
	F		12 mg
Vitamin B_{12}	M/F		No recommendation
Vitamin E	M/F		No recommendation
Phosphorus	M/F		500 mg
Magnesium	M		300 mg
	F		270 mg
Iron	M/F		8.7 mg
Potassium	M/F		350 mg
Zinc	M/F		9 mg
Copper	M/F		1.2 mg
Sodium	M/F		6 g
Fluid	M/F		1.5 l

[a]No RNI for fat/carbohydrates (expressed as EAR due to the close correlation between energy requirement and expenditure).
M, male; F, female.

Table 2 Glycemic index of selected foods (mean values)

Food	*Glycemic index*
Glucose	100
Instant rice	91
Baked potato	85
Cornflakes	84
White bread	70
Wholemeal bread	69
Sucrose (table sugar)	65
White rice (basmati)	58
Boiled potato	56
Banana	55
Baked beans (canned)	48
Orange	43
Pasta (spaghetti)	41
Apple	36
Milk	27
Kidney beans	27
Fructose	23

Derived from 600 entries in Foster-Powell K and Miller J (1995) International Table of Glycemic Index. *American Journal of Clinical Nutrition* 62: 8715–8935.

rapid absorption, and a high postprandial glucose and insulin response, ultimately giving rise to decreased insulin sensitivity. Total cholesterol and low-density-lipoprotein (LDL) cholesterol are not influenced. One community study has confirmed that low-glycemic-index diets are associated with raised HDL cholesterol levels and the longitudinal Nurse Health Study reports that diets with a high glycemic index are associated with an increased risk of coronary heart disease (CHD). Increasing prominence is being given to the role of high-glycemic diets and the increasing prevalence of type 2 diabetes mellitus.

Fat is the most concentrated calorie nutrient. In the USA, it is recommended that 30% or less of calories should be derived from fat, of which no more than 10% should be saturated. The UK Health of the Nation aims to reduce average percentages of food energy derived from fat by at least 12% by 2003 and that derived from saturated fat by 35% by 2005. The chief source of saturated fat is meat, butter, cheese, and cream. Unsaturated fats are derived mainly from plant sources, especially nuts, olive oil, rapeseed oil, some margarines, and fish. Evidence supports the role of saturated fats and trans-fats (hydrogenated polyunsaturates) in atheromatous vascular disease, although infection may be the initial trigger. COMA recommended a reduction of saturated fat and total fat contribution to the diet. However, there is no correlation between polyunsaturated fat intake and LDL levels. Indeed, the unsaturated fat in fish and in olive oil – about 70% monounsaturated – enhances the plasma concentration of the vascular-protective HDL cholesterol. The Nurse Health Study showed that women taking at least one ounce (15 g) of nuts five times or more a week had a greater reduction in risk of CHD than those who took them less frequently. Experimental diets replacing animal fats with nuts lowered LDL cholesterol and raised HDL cholesterol. Walnuts are high in linoleic acids, but most nuts are rich in arginine, a precursor of nitric oxide. A major longitudinal study showed that an increase in polyunsaturated fat intake accompanied by a reduced consumption of trans-fats diminishes the risk of type 2 diabetes mellitus. Studies based on Mediterranean diets (characterized by abundant fruit and vegetables, especially garlic and onions, cereals, beans, nuts, seeds, olive oil, fish and poultry) have reported a reduced risk of CHD and cancer.

Although a reduction in total dietary fat has no affect in reducing blood pressure, increasing the dietary polyunsaturated : saturated ratio or changing to diets rich in linoleic acids produces a moderate blood-pressure reduction. However, one study reported that attempts to enhance compliance with dietary recommendations in the over-70s was only half as successful as in the 60–69-year age group.

Protein

By age 70, skeletal muscle may have lost an average of 40% of its cell mass. Protein synthesis decreases as body cell mass falls with age and is associated with a decline in function. Functional decline is not irreversible and maintenance of physical activity or physical training may increase muscle strength and cardiac output. Protein malnutrition is associated often with profound muscle loss and weakness, and impaired immunological competence. However, this stage is usually compounded by other nutritional deficiencies and by serious disease.

Debate continues about protein requirements in old age. Studies on free-living and institutionalized older people suggest that the daily protein requirement should be about 1.0 g kg^{-1} and about 12–14% total calories. Inflammatory conditions and tissue necrosis (e.g., pressure sores) increase the protein turnover and requirement, and pressure-sore healing rates are influenced greatly by dietary protein intake. Protein malnutrition may follow major surgery or severe acute illness in elderly people and can seriously impair subsequent rehabilitation if left untreated.

Vitamins and Minerals

Vitamin and mineral supplements are necessary to treat well-defined deficiency disorders, but their role is less certain in the absence of obvious clinical

abnormality. Many older people fail to make the reference values for vitamin D, but the synthesis of vitamin D_3 (cholecalciferol) by the action of sunlight (UVL with a wavelength of 290–310 nnm) on the skin contributes some two-thirds of the vitamin D stores. Standard household glass prevents the passage of any therapeutic wavelength of UVL, and housebound and institutionalized patients are particularly susceptible to vitamin D deficiency. Subtle degrees of vitamin D deficiency lead to secondary hypoparathyroidism, increased bone turnover and osteoporotic fractures. Common dietary sources of vitamin D include eggs, cheese, margarine (fortified with vitamins D and A), and oily fish. Randomized controlled trials of dietary supplementation with 17.5 μg (700 IU) and 20 μg (800 IU) of vitamin D and 500 mg of calcium per day show a reduced incidence of hip and other nonvertebral fractures, whereas 10 μg (400 IU) of vitamin D does not reduce the fracture rate. There is considerable evidence for using vitamin D and calcium supplements in the vulnerable elderly, provided overdosage is avoided. Exposure to sunlight, even diffuse sunlight, is also to be encouraged.

Water-soluble vitamin C is readily destroyed during food preparation and storage, and the COMA recommendation of 40 mg does not take this into account. The delay between preparation and delivery to the patient in hospital or nursing home or of meals on wheels to the person at home is often sufficient to substantially reduce the vitamin C content. Meals on wheels are reported to have lost some 90% of the vitamin C content by the time of delivery. Older people eat vegetables more regularly than young adults, but elderly men who live alone have diets poor in vitamin C-rich foods such as fresh green vegetables and fruit. Scurvy is uncommon, but chronic mild deficiency is not infrequent and may play a role in atherogenesis and stroke. The inverse correlation between plasma fibrinogen concentration and serum ascorbate levels suggests that vitamin C supplementation may be cardioprotective.

Vitamin A intake is very variable, but needs tend to decrease with age, as absorption may be increased, whilst plasma clearance is delayed. However, it is often difficult to be certain whether homeostasis is achieved, or whether subclinical vitamin A states are common and unrecognized. Failure of dark adaptation is a well-recognized vitamin A deficiency phenomenon, but dryness of the eyes and perifollicular hyperkeratosis are uncommon deficiency manifestations in the Western World. In some third world countries, vitamin A deficiency is the commonest cause of blindness. The potential benefit of the antioxidant properties of vitamin A and the closely related β-carotene have yet to be established.

Thiamin (vitamin B_1) intake is closely related to energy consumption. Older people generally eat breakfast ('something to eat within 2 hours of rising'), and this commonly includes cereals and fortified cereals containing adequate thiamin. If disease or institutionalization impairs energy consumption, thiamin intake may be deficient. Alcoholism may give rise to confabulation and short-term memory impairment (Korsakoff's psychosis) due to associated thiamine deficiency or, more rarely, Wernicke's encephalopathy – acute confusion, ataxia, and opthalmoplegia. Acute postoperative delirium is not improved by large parenteral doses of the vitamin B group of vitamins. Any associated low plasma thiamine concentration may be an epiphenomenon.

Pyridoxine (vitamin B_6) and folate are important cofactors for the metabolism of homocysteine, and both vitamins independently lower raised homocysteine levels, thought to play a key role in atherogenesis. The Nurse Health Study found that a long-term intake of 3 mg of vitamin B_6 and/or at least 400 μg of folate per day have a favorable impact on CHD. In the Framingham Study, 20% of elderly subjects had folate and vitamin B_6 levels below the recommended intake.

Folic acid in its conjugated form (predominantly polyglutamate) is present in a wide variety of food, especially green leafy vegetables, kidney, and liver. Hydrolysis of the polyglutamate to monoglutamates is necessary for absorption in the small intestine. Folate deficiency may be due to several causes, with inadequate intake the most common, often aggravated by prolonged cooking, which destroys the vitamin. Disease that impairs absorption (e.g., gluten enteropathy or ileal resection) or increases metabolic demand (e.g., vitamin B_{12} deficiency, the leukemias, myelofibrosis, hemolytic anemia, sideroblastic anemia) may result in deficiency. A number of drugs (e.g., trimethoprim, methotrexate, triamterene) inhibit conversion of folate to tetrahydrofolate, which is the main storage and metabolically active form of the vitamin. The anticonvulsants, phenytoin, primidone, and phenobarbitone, and the antibiotic, nitrofurantoin, may also result in folate deficiency, but the underlying mechanism has not been firmly established. Defective DNA synthesis through the purine and pyrimidine pathways following chronic dietary folic acid deficiency or drug-induced activity gives rise to a megaloblastic anemia. Provided that vitamin B_{12} deficiency is not present, treatment with as little as 200 μg daily of oral folic acid is rapidly effective. In practice, pharmacological doses of 1–5 mg daily are used. Folinic acid (Leukovorin), a form of tetrahydrofolate, is preferable if the deficiency is due to folate antagonist drugs. The COMA recommendation of 200 μg per day is adequate

for red cell formation but may not be optimal for homocysteine metabolism and vascular wall integrity.

Dietary deficiency vitamin of B_{12} is rare, except in strict vegans. This is predominantly due to the efficient recirculation of the vitamin and a reported half-life of 480–1360 days. The predominant cause vitamin B_{12} deficiency is loss of intrinsic factor production by the gastric parietal cells, most often due to autoimmune disease giving rise to pernicious anemia. Antibodies against parietal cells and less commonly against intrinsic factor may be found in the gastric juice or serum. The inability to separate vitamin B_{12} from its dietary binding protein may also cause vitamin B_{12} deficiency; lack of gastric acid and pepsin contribute, but an unequivocal explanation has yet to be established. Bacterial overgrowth as in blind loop syndrome competes for available vitamin B_{12} and reduces the amount available from absorption. Gastric surgery by removing some or all intrinsic factor producing cells leads eventually to a low serum vitamin B_{12} concentration.

Because the liver stores large amounts of vitamin B_{12}, evidence of deficiency does not occur for about 5 years, even after total gastrectomy. Ileal disease or surgery often gives rise to vitamin B_{12} deficiency. The gradual but progressive development of anemia and the well-hemoglobinized red cells often results in minimal symptoms until an advanced anemic state. The common symptoms of anemia – fatigue, breathlessness, and poor exercise tolerance, occur. Acute onset of heart failure and angina is not uncommon because of associated CHD and responds well to vitamin B_{12} treatment. Much has been written on the relationship of dementia to vitamin B_{12} deficiency. Although low serum vitamin B_{12} concentrations are not uncommon in demented elderly patients living in the community, there is little evidence of any improvement in cognitive function following B_{12} dosage. When some improvement does occur, it can better be ascribed to concurrent treatment of other comorbid states and correction of impaired total dietary intake and improvement of restricted functional and social activities. However, some neuropsychiatric states are improved by vitamin B_{12} therapy. Neurological symptoms are relatively rare. Symmetrical peripheral neuropathy usually involving the lower limbs is associated with paraesthetic symptoms and evidence of ataxia and posterior column involvement leading to lower-limb weakness and spasticity.

Reduced plasma vitamin B_{12} concentration is an early feature of deficiency, though initially, the blood picture may be normal. Later, macrocytosis and megaloblastic bone marrow changes are characteristic, with low white blood counts and multisegmented polymorphs. When diagnosis is equivocal, the Schilling test is justified. In the presence of anemia, it is useful to measure the reticulocyte response to vitamin B_{12} administration, which reaches its peak on the 4th or 5th day. Treatment of confirmed deficiency with 1000 mcg of intramuscular vitamin B_{12} daily for 2 weeks rapidly replenishes body stores. Subsequently, lifetime intramuscular injections every 3 months is a cost-effective treatment schedule. Absorption of 1–3% of an oral dose of 1 mg of vitamin B_{12} daily generally satisfies the daily requirement of 2–5 μg.

The latest recommended vitamin E requirement is 15 mg per day. Vitamin E deficiency may result from fat malabsorption or defects in the gene for the α-tocopherol transfer protein and may give rise to the insidious onset of peripheral neuropathy. A possible influence of vitamin E (and vitamin C and β-carotene) on excess free radical formation has led the media to speculate on an antiageing effect. Although no benefit has been seen in patients with Parkinson's disease, a placebo-controlled trial of vitamin E 2000 IU daily in people with Alzheimer's disease has suggested a possible slowing of the rate of disease progression. The antiadhesive effect of vitamin E combined with the platelet antiaggregatory effect of aspirin increases the risk of bleeding, and caution is appropriate with supplementary vitamin E in a dose greater than 500 IU per day.

Insufficient iron intake is mainly found in housebound and institutionalized older people. Meat and cereals are the major sources of iron in the diet; fruit and vegetables have much lower proportions. Some flour and cereals are fortified with additional iron. Vitamin C enhances, and phytates or tannins in tea interfere with, iron absorption. Early iron store reduction is demonstrated by low serum ferritin concentrations. As iron deficiency progresses, anemia may develop. Moderate to severe iron-deficiency anemia is mostly due to chronic gastrointestinal bleeding due to drugs (e.g., aspirin, anticoagulants, nonsteroidal antiinflammatory drugs) or disease. Symptoms of iron deficiency may follow the onset of anemia and include fatigue, shortness of breath or exacerbation of concomitant ischemic arterial disease, e.g., angina, congestive heart failure, and intermittent claudication. Treatment of iron deficiency is usually with ferrous sulfate, 200 mg two to three times daily until the hemoglobin and iron stores are replaced. Concomitant administration of ascorbic acid as fruit juice or as a 50-mg vitamin C tablet increases absorption and reduces gastrointestinal side-effects. In the presence of a malabsorption state, iron may have to be given parenterally, but the potential risk of anaphylactic response requires supervision. Dietary advice on foods rich in available iron is an essential part of the long-term strategy, but for

some elderly people, a small dose of oral iron may be a permanent requirement.

Dietary Fiber

Dietary fiber keeps bowel movements regular and prevents constipation when taken with a good fluid intake and adequate exercise. Long-term intake of dietary fiber decreases the risk of CHD and diverticular disease and helps management of diabetes mellitus and hyperlipidemia. Its role in the prevention of colorectal cancer is still under debate. Only a minority of older people consume the recommended daily intake of 25–35 g of dietary fiber. This may be achieved by eating more fruit and vegetables (preferably unpeeled) by substituting whole-grain breads and cereals for white bread and sugary cereals and adding wheat bran and beans to soups, stews, and other dishes. Excess fiber intake should be avoided, as it may impair the absorption of trace minerals.

Clinical Evaluation

Once clinical symptoms and signs appear, nutritional impairment is firmly established. The Mini Nutritional Assessment (MNA; Figure 1) can be undertaken in 15 min and has been used successfully in free-living older people, in residents of nursing homes, and in hospital inpatients. A Danish 5-year longitudinal study showed an increased mortality in subjects with an MNA score of less than 23.5. Subjects in a New Mexico Study with a high MNA score of 27 or greater were classified as having very good health.

Weight is the traditional simple measure of nutritional state and often goes unrecorded in medical and nursing-home case notes. The body mass index (BMI) (weight in kilograms/height, in meters, squared) is not always feasible in disabled and bed-bound persons, but the height can be calculated using the sitting knee height (knee to floor or heel) or the demispan (sternal notch to finger web in the outstretched arm). People aged over 65 years are recommended to have a BMI between 24 and 29. A BMI under 20 indicates undernutrition. Mortality rates increase in the very elderly with low BMIs (under 20).

Body fat distribution changes with age, and in both sexes there is an increase in abdominal fat distribution and an increase of omental fat. In women, a waist/hip ratio over 0.9 is associated with a threefold risk of CHD when compared with a ratio of less than 0.7. Waist circumference is an independent risk factor, and women with a waist circumference of greater than 97 cm (38 inches) have a threefold risk of CHD. Men with a waist circumference of greater than 100 cm (40 inches) have a 2.5- to 4.5-fold increase in one or more cardiovascular risks. Gall bladder disease, cancer, and osteoarthritis are more common in the overweight, and obesity, particularly abdominal obesity, is a significant independent risk factor for insulin resistance and type 2 diabetes.

In older underweight persons, the cornerstone of management is to identify and treat aggressively the cause of weight loss. Poor appetite may improve with treatment of depression or review of drugs, such as digoxin, dopaminergic agents, and anticholinesterases. Treatment of poor oral hygiene, swallowing disorders, or malabsorption syndromes will increase food intake and absorption. Good management of systemic diseases such as cancer, chronic organ failure, infection, and inflammatory disorders, and relevant metabolic diseases such as hyperthyroidism, all require attention to the nutritional state. Attention to social and economic factors and to countering physical disabilities may also be necessary. Highly restrictive 'medically indicated' diets, for example, for lowering cholesterol or managing diabetes, may also lead to unintentional weight loss and nutritional deficiencies, and a more relaxed, well-balanced, and varied diet is likely to be preferable.

Intentional weight loss in obese older people may be associated with functional and health benefits, though some studies indicate that being overweight later in life does not pose a significant health risk. Weight reduction is a partnership between patient and medical advisors. Without this approach, it is difficult to maintain motivation and trust on the part of the patient. Randomized controlled studies firmly establish that low-fat diets produce weight loss. A slowly progressive increase in exercise levels has only a small impact on weight but has a more substantial benefit on function and improves insulin sensitivity.

Nutritional Management in Specific Medical Conditions

Pain and discomfort from inflammation of the gums, mouth, or tongue and missing or illfitting dentures will hinder adequate chewing and decrease food intake. Preventive dental care (from an early age) is the best management, and maintaining active dental health policies is especially important for those in institutions. When maintaining adequate nutrition becomes a problem, high-protein–energy liquids and soft nonacidic foods may be best tolerated, and a mild topical local anesthetic mouthwash before meals may help to relieve pain. Inadequate saliva production arising from infection, anticholinergic drugs, radiation therapy, or Sjogren's syndrome will interfere with taste and swallowing and may be partially

NESTLÉ NUTRITION SERVICES

Mini Nutritional Assessment MNA®

Last name: First name: Sex: Date:

Age: Weight, kg: Height, cm: I.D. Number:

Complete the screen by filling in the boxes with the appropriate numbers.
Add the numbers for the screen. If score is 11 or less, continue with the assessment to gain a Malnutrition Indicator Score.

Screening

A Has food intake declined over the past 3 months due to loss of appetite, digestive problems, chewing or swallowing difficulties?
0 = severe loss of appetite
1 = moderate loss of appetite
2 = no loss of appetite ☐

B Weight loss during the last 3 months
0 = weight loss greater than 3 kg (6.6 lbs)
1 = does not know
2 = weight loss between 1 and 3 kg (2.2 and 6.6 lbs)
3 = no weight loss ☐

C Mobility
0 = bed or chair bound
1 = able to get out of bed/chair but does not go out
2 = goes out ☐

D Has suffered psychological stress or acute disease in the past 3 months
0 = yes 2 = no ☐

E Neuropsychological problems
0 = severe dementia or depression
1 = mild dementia
2 = no psychological problems ☐

F Body Mass Index (BMI) (weight in kg) / (height in m)^2
0 = BMI less than 19
1 = BMI 19 to less than 21
2 = BMI 21 to less than 23
3 = BMI 23 or greater ☐

Screening score (subtotal max. 14 points) ☐☐
12 points or greater Normal – not at risk – no need to complete assessment
11 points or below Possible malnutrition – continue assessment

Assessment

G Lives independently (not in a nursing home or hospital)
0 = no 1 = yes ☐

H Takes more than 3 prescription drugs per day
0 = yes 1 = no ☐

I Pressure sores or skin ulcers
0 = yes 1 = no ☐

J How many full meals does the patient eat daily?
0 = 1 meal
1 = 2 meals
2 = 3 meals ☐

K Selected consumption markers for protein intake
• At least one serving of dairy products (milk, cheese, yogurt) per day? yes ☐ no ☐
• Two or more servings of legumes or eggs per week? yes ☐ no ☐
• Meat, fish or poultry every day yes ☐ no ☐
0.0 = if 0 or 1 yes
0.5 = if 2 yes
1.0 = if 3 yes ☐.☐

L Consumes two or more servings of fruits or vegetables per day?
0 = no 1 = yes ☐

M How much fluid (water, juice, coffee, tea, milk…) is consumed per day?
0.0 = less than 3 cups
0.5 = 3 to 5 cups
1.0 = more than 5 cups ☐.☐

N Mode of feeding
0 = unable to eat without assistance
1 = self-fed with some difficulty
2 = self-fed without any problem ☐

O Self view of nutritional status
0 = views self as being malnourished
1 = is uncertain of nutritional state
2 = views self as having no nutritional problem ☐

P In comparison with other people of the same age, how does the patient consider his/her health status?
0.0 = not as good
0.5 = does not know
1.0 = as good
2.0 = better ☐.☐

Q Mid-arm circumference (MAC) in cm
0.0 = MAC less than 21
0.5 = MAC 21 to 22
1.0 = MAC 22 or greater ☐.☐

R Calf circumference (CC) in cm
0 = CC less than 31 1 = CC 31 or greater ☐

Assessment (max. 16 points) ☐☐.☐
Screening score ☐☐
Total Assessment (max. 30 points) ☐☐.☐

Malnutrition Indicator Score
17 to 23.5 points at risk of malnutrition ☐
Less than 17 points malnourished ☐

Ref.: Guigoz Y, Vellas B and Garry PJ. 1994. Mini Nutritional Assessment: A practical assessment tool for grading the nutritional state of elderly patients. *Facts and Research in Gerontology.* Supplement #2:15-59.
Rubenstein LZ, Harker J, Guigoz Y and Vellas B. Comprehensive Geriatric Assessment (CGA) and the MNA: An Overview of CGA, Nutritional Assessment, and Development of a Shortened Version of the MNA. In: "Mini Nutritional Assessment (MNA):Research and Practice in the Elderly". Vellas B, Garry PJ and Guigoz Y, editors. Nestlé Nutrition Workshop Series. Clinical & Performance Programme, vol.1. Karger, Bâle, in press.

Figure 1 Mini Nutritional Assessment (MNA).

relieved by spraying the mouth with artificial saliva solution. Normal age-related changes in taste result in older people being more sensitive to bitter and sour flavors and less sensitive to sweet and salt. A declining sense of smell means that food aromas that stimulate the appetite must be several times stronger than usual.

Problems with swallowing (dysphagia) may be due to inflammatory, traumatic (e.g., postintubation),

and neoplastic causes in the mouth or esophagus, mechanical or neuromuscular causes, or a pharyngeal diverticulum. Management is primarily directed to resolution of the causative factors, but there will be an intervening period of a need for nutritional maintenance, or, where the underlying cause cannot be treated, long-term nutritional support may be required. Stroke, motor neurone disease, and, less commonly, Parkinson's disease may be associated with severe swallowing problems. Following an acute stroke, swallowing difficulties are generally transient and resolve within a week. However, in 15% of patients, swallowing impairment with risk of aspiration excludes oral nutrition. Dehydration leads to rapid deterioration, and intravenous fluids are required in the first few days. Three options are available for nutritional maintenance until safe swallowing returns. Parenteral nutrition is only practical for a few days and has a very limited role in poststroke management. Fine-bore nasogastric tube feeding is commonly the interim measure for nutritional provision. In the first week or so, emphasis is on energy and protein intake, but a balanced nutrient provision with attention to micronutrients and mineral content becomes essential for longer-term management. Percutaneous endoscopic gastrostomy is well tolerated and is the method of choice for long-term management. The benefits of early PEG feeding have yet to be established, but ongoing randomized trials should provide evidence-based guidance.

In motor neurone disease, the early stages of dysphagia are managed by thickening all fluids and liquidizing and adapting solid food to a thick custard consistency. Unfortunately, food consumption can easily take on an 'eat to live' meaning with the loss of social and pleasurable aspects of eating.

Alzheimer's disease is the commonest cause of dementia in later life. Weight loss is usual, sometimes even early in the disease, and vitamin deficiencies and disturbed eating behavior are common. Where the cognitively impaired individual lives alone and has difficulty in buying or preparing an adequate diet or forgets meals, the cause of the weight loss is readily apparent. In later stages, restlessness and agitation may result in excessive energy expenditure, or aversion to food may develop. Often, however, there is no obvious cause of weight loss. Carers may be able to provide accurate dietary information to enable monitoring of the patient's nutritional status using the MNA. There is current interest in the possible role of inadequate intake of folate and raised homocysteine in the etiology of cognitive impairment. However, it is also likely that vitamin deficiencies may develop as a result dementia. Antioxidant vitamins may have neuroprotective effects and, in the experimental animal, can improve cognitive function. High-saturated-fat diets are reported to be associated with an increased likelihood of cognitive impairment, whereas polyunsaturated fats have a negative influence. Tube feeding of people with advanced dementia has no measurable influence on survival and raises important practical and ethical concerns.

Decline in glucose tolerance and type 2 diabetes mellitus is increasingly common in old age. A raised 2-h postprandial glucose test is associated with increased mortality. Dietary management has changed considerably over recent years. From the early 1990s, reducing dietary fat intake, particularly saturated fat, has been recommended. Currently, a reduction in saturated and trans-fats is advised, but no restriction on monosaturated and polyunsaturated fats. Increasing emphasis is also placed on replacing high-with low-glycemic-index foods. A change in lifestyle is important in both prevention and treatment, and it is inappropriate to consider diabetes in isolation from other risk factors. Hypertension affects at least 50% of older diabetics, and reduced sodium intake enhances the overall management of high blood pressure.

Undernutrition in Institutionalized Patients

Many hospital patients and residents of long-term care facilities cannot feed themselves, and many are malnourished. Since the serving and checking of meals have become a 'nonnursing duty' in the UK, awareness of the patient's nutrient intake may be limited. Meals are often kept for some time in heated wagons before serving, and some nutrients are largely destroyed. Older people are more likely to eat food provided when it is served warm, with portion size individualized, and the eating environment is considered. In one study, in which plated meals were weighed before and after serving, elderly patients were shown often to have inadequate time or ability to cut and eat their meals. When all food was cut up into bite-sized pieces, food consumption improved significantly. Food must also be placed within reach of disabled patients. An uncluttered, well-lit dining room with minimal distractions, providing consistent seating at mealtimes and attractive table settings will enhance eating. Attention to poor hearing enhances enjoyment of mealtime conversation. People with physical disabilities may benefit from feeding aids such as nonskid place mats, weighted cutlery with thick handles, plates with wide, curved lips that help keep food on the plate, and cups with easy-grip handles and special lids for sipping. Cyclical menus should be available, recognizing individual

preferences and offering properly prepared and appetizing meals. Dividing up the daily food ration into lighter, more frequent snacks may be more acceptable than the traditional three meals a day. Special events such as picnics and family dinner parties and special menus for festive occasions such as birthdays will encourage an interest in eating. Patients requiring a softer or liquid diet would benefit by having their food presented creatively.

Water is an important nutrient that is often overlooked. Thirst lessens with age, so older people need to adopt the habit of drinking fluids regularly, or dehydration can occur easily. This is a particular problem in the physically or mentally frail and the institutionalized, who are especially susceptible to the dehydrating effects of fever, vomiting, and diarrhea.

Undernutrition in People at Home

The elderly person living alone is at especial risk of malnutrition. Grief from the loss of a spouse or family and friends, loneliness and depression, greater use or abuse of alcohol, which may substitute for an adequate food intake, lack of sufficient money for food, and disability affecting shopping or preparation of food may all contribute. Poor vision is also an independent determinant of poor eating, leaving people less confident in their ability to cook and less able to get out for groceries. Support from relatives, neighbors, and homecare assistants providing and preparing food deserves recognition. Eating should be a social event, and older people should be encouraged to eat with family and friends whenever possible.

Relevant community services for older people include meals on wheels, which provide hot meals up to 7 days per week to homebound people in their own homes. However, arrangements are often inflexible, and there is limited ability to conform to special dietary needs. Frozen meals delivered weekly and which can then be easily reheated in a microwave oven may be a more appropriate and palatable alternative. Day centers and luncheon clubs have a useful role in supplementing the diet and reinforce the social aspects of feeding. An emergency store of essential food items in case of unexpected illness or bad weather is a useful precautionary measure.

See also: **Aging – Nutritional Aspects**; **Ascorbic Acid**: Properties and Determination; **Carotenoids**: Occurrence, Properties, and Determination; **Cholecalciferol**: Properties and Determination; **Cobalamins**: Properties and Determination; **Dietary Reference Values**; **Dietary Requirements of Adults**; **Elderly**: Nutritional Status; Nutritionally Related Problems; **Energy**: Intake and Energy Requirements; **Folic Acid**: Properties and Determination; **Iron**: Properties and Determination; **Malnutrition**: The Problem of Malnutrition; **Minerals – Dietary Importance**; **Protein**: Requirements

Further Reading

Cummings JH and Bingham SA (1998) Diet and the prevention of cancer. *British Medical Journal* 317: 1636–1640.

Department of Health (1992) *The Nutrition of Elderly People. Great Britain Committee on Medical Aspects of Food Policy Working Group on the Nutrition of Elderly People. Report on Health and Social Subjects. No. 43*. London: Stationery Office.

Ebrahim S and Davey-Smith G (1996) *Health Promotion in Old People for the Prevention of Coronary Heart Disease and Stroke. Health Promotion Effectiveness Reviews*. London: Health Education Authority.

Fiatarone MA, O'Neill EF, Ryan ND *et al*. (1994) Exercise training and nutritional supplements for physical frailty in very elderly people. *New England Journal of Medicine* 330: 1769–1775.

Finch S, Doyle W, Lowe C *et al*. (1998) *National Diet and Nutrition Survey: People Aged 65 years and Older. Volume 1: Report of the Diet and Nutrition Survey*. London: Stationery Office.

Fletcher A and Rake C (1998) *Effectiveness of Interventions to Promote Healthy Eating in Elderly People Living in the Community: A Review. Health Promotion Effectiveness Reviews No. 8*. London: Health Education Authority.

Gariballa SE and Sinclair AJ (1997) Diagnosing undernutrition in elderly people. *Reviews in Clinical Gerontology* 7: 367–371.

Horwitz A, MacFadyen DM, Munro H, Scrimshaw NS, Steen B and Williams TF (1989) *Nutrition in the Elderly*. Oxford: Oxford University Press.

Lilley J and Hunt P (1998) *Opportunities for and Barriers to Managing in Dietary Behaviour in Elderly People*. London: Health Education Authority.

Lipschitz DA (1995) Approaches to the nutritional support of elderly patients. *Clinics in Geriatric Medicine* 11: 715–724.

McGee M and Jensen GL (2000) Nutrition in the elderly. *Journal of Clinical Gastroenterology* 30: 372–380.

Roe L, Hunt P, Bradshaw H and Rayner M (1997) *Review of the Effectiveness of Health Promotion Interventions to Promote Healthy Eating*. London: Health Education Authority.

Seiler WO and Stahelin HB (eds) (1999) *Malnutrition in the Elderly*. Berlin: Springer-Verlag.

Vellas BJ, Garry PJ and Guigoz Y (eds) (1999) *Research and Practice in the Elderly (Nestlé Nutrition Workshop Series. Clinical and Performance Programme,* vol. 1. Basel: Karger.

Walls AW, Steele JG, Sheiham A, Marcenes W and Moynihan PJ (2000) Oral health and nutrition in older people. *Journal of Public Health Dentistry* 60: 304–307.

Electrical Supplies *See* **Power Supplies**: Use of Electricity in Food Technology

ELECTROLYTES

Contents
Analysis
Water–Electrolyte Balance
Acid–Base Balance

Analysis

G R Ahmad, California State University, San Bernardino, CA, USA
D R Ahmad, University of California, Riverside, CA, USA

Electrolytes and their Significance

Electrovalent solids exist as crystal lattices composed of ions bound together by a strong electrostatic force of attraction between positive and negative charges. The solutions of such substances in water conduct electricity and are therefore called electrolytes. Electrolytes are classified as anions or cations, depending upon their direction of movement in an electric field. If an electrolyte bears a positive charge, it will migrate to the cathode (negative pole) and is called a cation. Sodium and potassium are common examples of cations. If the electrolyte bears a negative charge, it will migrate to the anode (positive pole) and is called an anion. Chloride and bicarbonate are examples of anions.

In body fluids, the major electrolytes exist as free ions, while the trace metals primarily occur in combination with proteins. The electrolytes play very important roles in the physiology of living organisms. In the human body, virtually every metabolic process is dependent on, or is affected by, electrolytes. Electrolytes are involved in the maintenance of the body's osmotic pressure, water distribution in the various body fluid compartments, maintenance of proper pH, involvement in oxidation – reduction reactions, acting as cofactors for enzymes, and regulation of neuromuscular irritability or excitability. Thus, abnormal levels of electrolytes could be the cause or consequence of a variety of physiological disorders, which make the determination of eletrolytes one of the most important functions of clinical laboratories. (*See* **Coenzymes; pH – Principles and Measurement.**)

The major electrolytes found in the body are sodium (Na^+), potassium (K^+), calcium (Ca^{2+}), magnesium (Mg^{2+}), chloride (Cl^-), bicarbonate (HCO_3^-), phosphate (HPO_4^{2-}), and sulfate (SO_4^{2-}). In clinical laboratories, the most common electrolytes requested in a medical diagnosis are commonly referred to as the 'electrolyte profile.' This panel includes four electrolytes: HCO_3^-, Cl^-, Na^+ and K^+. (*See* **Calcium**: Physiology; **Magnesium**; **Potassium**: Physiology; **Sodium**: Physiology.)

Specimens Required for Electrolyte Analysis

Electrolytes can be assayed from many body fluids, including serum, heparinized plasma, whole blood, vitreous humor, sweat, urine, gastrointestinal fluid, and aqueous extract from feces. In clinical laboratories, serum and urine are the body fluids from which electrolyte analyses are most frequently performed. These fluids are easily obtained from the patient, and the reference levels (normal) for these fluids are available. When the electrolyte level of body fluid is required, serum is considered the most practical specimen. When electrolytes are to be determined from urine, a timed collection is the preferred method. Timed-collection specimens are needed to allow a comparison of values with reference ranges or for the determination of rates of electrolytes lost from the body. These specimens can be stored at 2–4 °C or can be frozen for delayed analysis. The average serum levels of common electrolytes are as follows:

- bicarbonate (expressed as total carbon dioxide), 22–31 mmol l^{-1};
- chloride, 98–106 mmol l^{-1};

- sodium, 135–145 mmol l^{-1};
- potassium, 3.5–5.0 mmol l^{-1};
- calcium, 8.5–10.2 mg dl^{-1};
- magnesium, 1.8–2.4 mg dl^{-1}.

Analytical Methods

Many methods exist for the assay of electrolytes. These methods include flame emission spectrophotometry (FES), atomic absorption spectrophotometry (AAS), electrochemical analysis with ion-selective electrodes (ISE), redox titration, colorimetry, and fluorescent probes. In the clinical laboratory, the majority of the analyses are performed by FES, ISE, and colorimetry. (*See* **Calcium**: Properties and Determination; **Sodium**: Properties and Determination.)

Flame Emission Spectrophotometry

Flame emission spectrophometry is most commonly used for the analysis of Na^+ and K^+. When metals such as Na^+ and K^+ are sufficiently heated in a flame, they emit light of wavelengths characteristic of the metal. The metal ions absorb heat energy, which causes electrons to be driven out of their normal orbital position. These excited metal atoms are said to be in high-energy states and are very unstable. The extra energy is given off in the form of light. When Na^+ and K^+ are excited, they emit spectra with sharp, bright lines at 589 and 768 nm, respectively. Within certain limits, the amount of light given off by the excited atoms is proportional to the concentration of the metal ions in the solution. To prepare a flame emission spectrophotometer for detecting Na^+ and K^+, standard solutions containing known concentrations (two levels) of these electrolytes are aspirated into the instrument. The digital readout is then set to the known concentrations of the fluid being aspirated. When the flame photometer has been so calibrated, the fluid containing the unknown concentration is aspirated, and the result indicating the concentration is read from the digital display.

Another FES-type system, inductively coupled argon plasma (ICAP), utilizes the principle of atomic emission spectrophotometry. It is similar to FES, except that the plasma has a much higher excitation temperature. This results in lower limits of detection, a wide dynamic range, and virtual freedom from chemical interference. It is possible to use ICAP for the analysis of metals, such as sodium and potassium, as well as many nonmetals, such as sulfur and boron; however, Cl^- cannot be assayed by using ICAP spectrophotometry. Although the principle of ICAP has been used to develop automated instrumentation, at present this system has not been used in clinical laboratories.

Atomic Absorption Spectrophotometry

Atomic absorption spectrophotometry is not commonly employed in clinical laboratories. As AAS is much more sensitive than FES, AAS is used as a reference method for many analytical procedures. The principle on which AAS is based is similar to FES. In AAS, however, the flame serves to dissociate the element from its chemical bonds and place it in a ground state at which it is capable of absorbing light of a wavelength specific for the element. The method involves using hollow cathode lamps containing the metal of interest. The metal in the lamp is subjected to an electric current, causing it to emit light at a characteristic wavelength specific to the element in the lamp. This light passes over a burner to a special detector that measures emitted intensity.

To prepare an AAS for analyzing electrolytes, a standard solution containing a known concentration of the metal of interest is aspirated into the burner; the atoms in this field absorb the light proportional to their concentration in the fluid. The resulting decrease in intensity of the beam of light given off from the hollow cathode lamp is then set on a digital readout to reflect the known concentration. Once the instrument has been so calibrated, the unknown fluid is aspirated, and the electrolyte concentration can be obtained from the digital display. This method can be used in the analysis of Na^+, K^+, Ca^{2+}, and Mg^{2+}.

Ion-selective Electrodes

Potentiometry with ISE is a comparatively recent development in the electrolyte analytical field. It is the measurement of the electrical potential difference between two electrodes, which, in contact with one or more electrolyte solutions, form an electrochemical cell. A high-impedance voltmeter measures the electrical potential difference between the two electrodes. Ion-selective electrodes respond preferentially to a particular ion species in solution. Some potentiometric analytical methods use an ion-selective indicator electrode and a reference electrode. A method developed by Eastman Kodak and used on their Ektachem analyzer uses two identical ISE mounted in a single-use disposable slide. The reference fluid containing ions at fixed concentration is applied to the reference electrode. The sample fluid is simultaneously applied to the indicator electrode. The two solutions are allowed to form a liquid junction in the bridge portion of the slide. This arrangement constitutes a concentration cell.

The specific components of the cell vary according to the ion being tested. Over the physiological range, the relationship between the measured cell potential and the logarithm of the molal concentration of the ion being tested from the fluid is linear. The junction potential is produced between the reference fluid and the sample fluid in the bridge portion of the analysis slide. The magnitude and sign of the junction potential depend on the total ionic concentration and composition of the fluid but are primarily determined by the Cl^- and Na^+ concentrations. The magnitude of this term is minimized by using a reference fluid having ionic concentrations near physiological averages. Applications of this have been developed for the analysis of carbon dioxide (CO_2), Cl^-, Na^+, and K^+. The results are normally reported in molar units.

Other methods employing ISE technology use electrodes of a nondisposable variety. Glass electrodes are made from specially formulated glass with added oxides of various metals. For example, the most practical sodium electrode is made of specialized glass, selective for Na^+. A solid-state electrode employing a homogeneous membrane of silver chloride is used for Cl^- analysis of sweat by direct measurement from the skin surface. A liquid ion-exchange electrode with an ion-selective carrier embedded in a polyvinylchloride matrix is used for chloride analysis of body fluids. A liquid ion-exchange membrane electrode, incorporating the antibiotic valinomycin as the K^+ binder, is the most selective for potassium. A gas electrode is available for total CO_2 (HCO_3^-) analysis. This involves a combination of a glass pH electrode and a reference electrode in contact with a weakly buffered electrolyte solution behind a membrane permeable to CO_2 gas. When the CO_2 gas is forcibly released from the sample, it diffuses through the membrane and reacts with water in a buffer solution, resulting in a change in the hydrogen ion activity of the buffer. The pH electrode senses the change in hydrogen ion activity, which is an indication of the concentration of the CO_2 present in the sample fluid. Ion-selective electrodes have been incorporated into many automated systems used in clinical laboratories for assaying the 'electrolyte profile.'

Redox Titration

Redox titration is characterized by the transfer of electrons from one substance to another with an end point determined calorimetrically or potentiometrically. This method is used for the determination of Ca^{2+}. Calcium is precipitated as calcium oxalate, which is treated with sulfuric acid. The oxalic acid formed is redox-titrated with potassium permanganate (oxidant). This method is time-consuming and is not used in clinical laboratories.

Colorimetry

In clinical laboratories, colorimetry is commonly employed for the analysis of Ca^{2+} and Mg^{2+}. The analysis is based on reacting the metal with an indicator dye to form a colored complex. The density of the resulting complex is related to the concentration of the metal and can be measured spectrophotometrically. A common dye used for assaying total calcium in body fluid is *o*-cresolphthalein complexone. This reagent contains other compounds (8-hydroxyquinoline, urea, and ethanol) which serve to mask interferences and enhance the reaction. Magnesium is commonly determined by complexing Mg^{2+} with a Calmagite reagent, resulting in another chromophore which can also be measured spectrophotometrically. (*See* **Spectroscopy**: Visible Spectroscopy and Colorimetry.)

Fluorescent Probes

The assay of intracellular levels of Ca^{2+} can be conducted in living cells by the use of fluorescent probes. These dyes are introduced into the cell as an ester derivative which is nonpolar and thus freely diffusible across cell membranes. Once within the cell, the compounds are hydrolyzed by nonspecific esterases present in the cytosol. The resulting cation is trapped within the cell owing to its inability to traverse the plasma membrane. Thus, the cation gradually accumulates intracellularly. When excited at a given wavelength, its fluorescence reflects the concentration of free cytosolic calcium.

See also: **Calcium**: Properties and Determination; Physiology; **Coenzymes**; **Magnesium**; **pH – Principles and Measurement**; **Potassium**: Physiology; **Sodium**: Properties and Determination; Physiology; **Spectroscopy**: Visible Spectroscopy and Colorimetry

Further Reading

Anon. (1986) Test methodologies. Potentiometric methodology, section 2. In: *Ektachem Clinical Chemistry Products Manual*. Rochester, NY: Eastman Kodak.

Blick KE and Liles SM (1985) *Principles of Clinical Chemistry*. New York: John Wiley.

Henry JB (1984) *Clinical Diagnosis and Management by Laboratory Methods*, 17th edn. Philadelphia, PA: WB Saunders.

Parker SP (ed.) (1989) *McGraw Hill Dictionary of Scientific and Technical Terms*. New York: McGraw-Hill.

Tietz NW, Pruden EL and Sigaard-Anderson O (1987) Electrolytes, blood gases and acid–base balance. In: Tietz NW (ed.) *Fundamentals of Clinical Chemistry*, 3rd edn. Philadelphia, PA: WB Saunders.

Tsien RY (1989) Fluorescent probes of cell signalling. *Annual Review of Neuroscience* 12: 227–253.

Water–Electrolyte Balance

Y L Cheng and A W Yu, The Chinese University of Hong Kong, Hong Kong, China

Background

The water–electrolyte balance is regulated tightly in the human body for the normal function of cells and organs. The kidneys have a critical role in regulating the composition of water and several important inorganic ions (sodium, potassium, chloride, bicarbonate, hydrogen ion, calcium, and phosphate) in the body and thus homeostasis. This is maintained by an appropriate balance between the excretion of these electrolytes by the kidneys and the daily intake. Hypernatremia and hyponatremia (excess and inadequate plasma sodium), hyperkalemia and hypokalemia (excess and inadequate plasma potassium) are the most common electrolyte disorders and result from an imbalance between intake and excretion of water and these electrolytes. This article is a consideration of the interactions of hormones and the autonomic nervous system on renal tubules in maintaining the water–electrolyte balance.

Water Balance

Water Content in the Body

Water is the most abundant component in the human body. Total body water is about 50% of body weight for normal adult women and about 60% of body weight for normal adult men. Fat has far less water than other body tissues. Consequently, women, obese people and the elderly contain less water in proportion to their body weight because, generally, they have proportionately more body fat.

Body water is distributed in two major compartments that are separated by cell membranes: the intracellular fluid (ICF) and the extracellular fluid (ECF). Approximately two-thirds of the total body water is in the ICF and one-third in the ECF. The distribution of total body water in each compartment is shown in Figure 1.

Water distribution in the ICF and ECF depends on the number of particles (electrolytes and macromolecules) confined in these compartments. The concentration of an individual substance in the two compartments may not be equal because the cell membrane has different permeabilities and channels for different substances. The concentration of a substance in a solution can be expressed as osmolality. Some substances, such as urea and alcohol, freely cross most cell membranes. 'Effective osmolality,' or tonicity, refers to the osmolality that contributes to non-freely permeable substances. Water will move across the cell membrane along the osmotic gradient, i.e., from a compartment with a low effective osmolality to a compartment with a high effective osmolality, until the tonicity is equal on both sides of the membrane. The normal plasma osmolality is about 285–295 mosmoles (mOsm) per kilogram of water.

Water Intake and Water Output

In the steady state, water intake must be equal to water output. For a normal adult, the usual water intake is about 1.5–2.5 l per day (which includes ingested water and water contained in foods), and the endogenous water production by oxidative metabolism is about 0.5 l per day. Water loss is through urine (about 1–2 l per day), 'insensible water loss' by evaporation of water through skin and lungs (about 1 l per day), and stool excretion (about 0.1–0.2 l per day). Urine is the only route by which water excretion

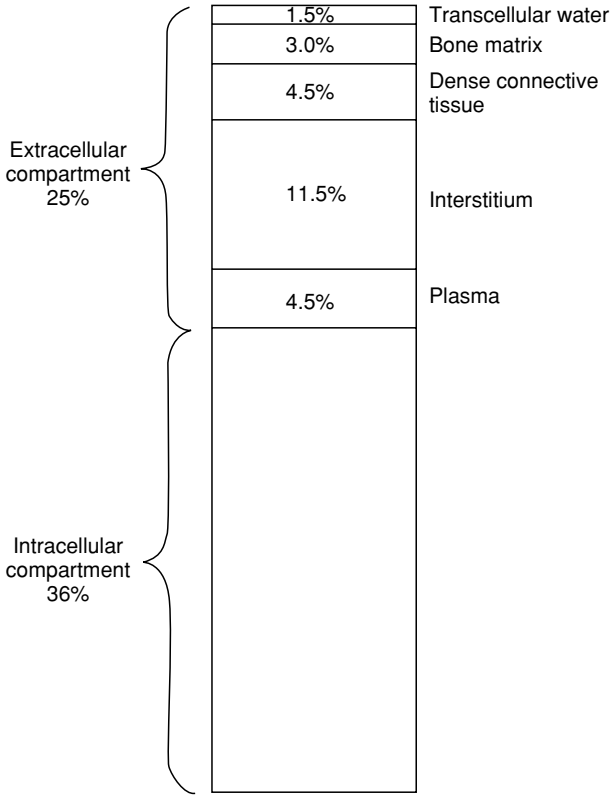

Figure 1 Distribution of body water in typical adult men, expressed as a percentage of body weight.

can be closely regulated. Apart from water intake, urine output also relates directly to the solute excretion, and it cannot be reduced below a certain minimal value required to carry the solute load. This is known as 'obligatory kidney water loss.' In order to produce a maximally concentrated urine (urine osmolality approximately 1200 mOsm kg^{-1}), for instance, a daily solute load of 600 mOsm requires a minimal obligatory urine output of 0.5 l per day (600 mOsm divided by 1200 mOsm kg^{-1}) to excrete the solute.

Regulation of Water Balance

Water balance is the result of the interaction of thirst and arginine vasopressin (AVP), which is also termed antidiuretic hormone, to maintain a stable plasma tonicity. The sensation of thirst promotes water intake, and AVP regulates urinary water excretion. AVP is a nine-amino acid polypeptide that is synthesized in the hypothalamus and stored in the posterior pituitary. Upon stimulation, AVP is released into the blood circulation. It then binds to the receptor (known as AVP receptor 2) on the basolateral membrane, opens up the water channel (known as the aquaporin-2 channel) in the apical membrane of the principal cells of the collecting ducts in the kidneys, and increases the water permeability. The medullary interstitium surrounding the collecting ducts is hypertonic with an osmolality up to 1200 mOsm kg^{-1}. When the water channels are open, water is reabsorbed along the osmotic gradient. In the absence of AVP, the kidneys excrete a large amount of hypoosmotic fluid rapidly. Urine osmolality reflects the kidneys' ability to dilute (removal of water) or concentrate (conservation of water) the urine. In a normal adult, the urine osmolality varies from 80 to 1200 mOsm kg^{-1}.

Thirst is more important than the AVP mechanism. In the absence of AVP secretion (e.g., patient with cranial diabetes insipidus), urinary water loss will increase. If the thirst mechanism is intact, water balance can still be maintained by an increase in water intake. However, patients with a defective thirst mechanism will be predisposed to dehydration, even if the AVP mechanism is normal.

Control of Thirst and AVP Secretion

Both osmotic and nonosmotic factors are involved in the regulation of thirst and AVP release. Amongst these, the osmotic factor (i.e., plasma tonicity) is more important. The factors and their effects are summarized in Tables 1 and 2.

The osmolality sensor (osmoreceptor), the thirst center, and the AVP secretory cells are located in the hypothalamus. The sensor is a group of specialized cells that is remarkably sensitive to the changes of tonicity in the ECF. When there is a water deficit (Figure 2), plasma tonicity will increase. The osmoreceptors will send a signal to both the thirst center and the AVP secretory cells. Intake of water will be increased, and the kidneys will retain water via the action of released AVP.

Conversely, when there is excessive water (Figure 2), plasma tonicity will decrease. Both the thirst center and secretion of AVP will be inhibited to promote net urinary water loss. These responses act as a feedback loop to keep the plasma tonicity and, hence, the body water content within tight limits.

Table 1 Control of thirst

Factors	*Stimulation*	*Inhibition*
Osmotic	↑ Tonicity	↓ Tonicity
Nonosmotic	↓ ECF volume	↑ ECF volume
	Congestive heart failure	Dopamine antagonists
	Hormones/drugs	
	Angiotensin II	
	Catecholamines	
	Parasympathomimetic drugs	

Table 2 Control of arginine vasopressin release

Factors	*Stimulation*	*Inhibition*
Osmotic	↑ Tonicity	↓ Tonicity
Nonosmotic	↓ ECF volume	↑ ECF volume
	↓ Blood pressure	↑ Blood pressure
	Nausea/pain/anxiety	Hormones/drugs
	Congestive heart failure	Dopamine antagonists
	Liver disease	Ethanol
	Hypoglycemia	Norepinephrine
	Adrenal insufficiency	
	Hormones/drugs	
	Angiotensin II	
	General anesthetics	
	Barbiturates	
	Chlorpropamide	
	Clofibrate	
	Cyclophosphamide	
	Nicotine	
	Vincristine	

Kidneys and Water Balance

A nephron is the functional unit of the kidneys. Each kidney contains approximately one million such units. A nephron is composed of a glomerulus and a renal tubule (Figure 3). The renal tubule is subdivided further into the proximal convoluted tubule, the loop

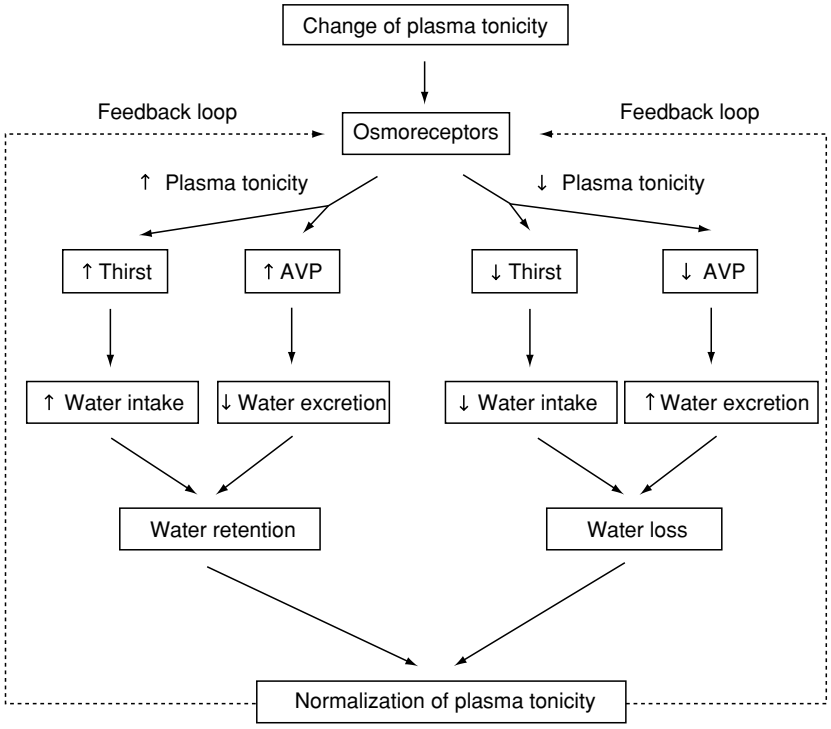

Figure 2 Pathway of water conservation/excretion. AVP, arginine vasopressin.

of Henle, the distal convoluted tubule, and the collecting duct.

The glomeruli filter about 180 l of fluid (which contains water, electrolytes, and glucose) each day. Up to 90–99% of the filtered water is reabsorbed by the renal tubules. It is reabsorbed passively in the proximal convoluted tubules (60–70% of the filtered water) and the descending limbs of the loops of Henle, with down osmotic gradients created by the active transport of sodium and chloride out of the lumina. Water is not permeable in the ascending limb of the loop of Henle and the distal convoluted tubule. The final urine volume is determined by the action of AVP on the collecting ducts.

Sodium Balance

Sodium Content in the Body

For a normal adult, total body sodium content is about 55–60 mmol per kilogram of body weight. The distribution of sodium in different compartments of the body is shown in Figure 4.

Sodium is the major cation in the ECF. Its concentration is about 135–145 mmol l^{-1} and 3 mmol l^{-1} in the ECF and ICF, respectively. The high sodium content in the ECF is maintained by sodium–potassium–adenosine triphosphatase (Na^+–K^+ –ATPase) pumps. The Na^+ –K^+–ATPase pump is located in the cell membrane, and it hydrolyzes a molecule of adenosine triphosphate to release energy to pump three sodium molecules out of the cells (from ICF to ECF) and to transport two potassium molecules into the cells (from ECF to ICF). In normal circumstances, sodium and its accompanying anions (mainly chloride) contribute to more than 90% of the effective osmolality in the ECF. As noted previously, effective osmolality determines the volume in the corresponding compartment. Therefore, sodium is the major determinant of the ECF volume. An excess in total body sodium content leads to expansion of the ECF volume, and a deficit in total body sodium content is associated with contraction of the ECF volume. A stable ECF volume is essential to maintain tissue perfusion, because the plasma volume (one of the compartments in ECF) is directly proportional to the ECF volume.

Sodium Intake and Excretion

A typical Western diet contains approximately 150 mmol of sodium per day. However, social and cultural differences in dietary habits can have a significant effect on sodium intake. Amounts of sodium can be up to 300 mmol per day in some areas in Japan to less than 30 mmol per day in the New Guinea Highlands. In normal health, the amount of sodium

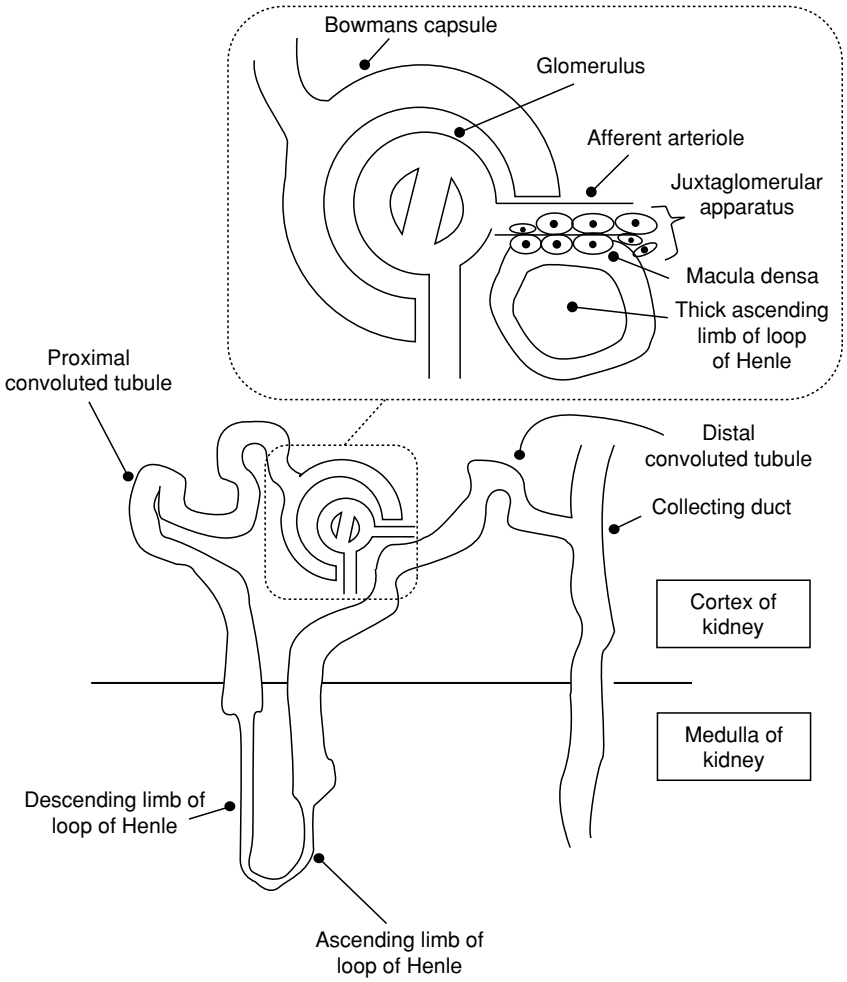

Figure 3 Nephron and juxtaglomerular apparatus.

excretion in the stool (about 5 mmol per day) and sweat (about 10 mmol per day) is small and cannot be regulated. Urine is the only route by which sodium excretion can be precisely regulated to match the sodium intake. Depending on the dietary intake and the status of body sodium content, urinary sodium loss can range from less than 1 mmol in a state of sodium deficit, to more than 500 mmol per day in a state of sodium excess. Thus, the kidneys have an important role in maintaining sodium balance and ensuring a stable ECF volume.

Kidneys and Sodium Balance

In people with normal kidney function, the glomeruli filter about 25 000 mmol sodium each day. This amount is approximately six times the total body sodium. Consequently, control of sodium reabsorption is important in maintaining sodium homeostasis. As shown in Table 3, most of the filtered sodium (60–70%) is reabsorbed at the proximal convoluted tubules. Twenty to twenty-five per cent is reabsorbed at the thick ascending limbs of the loops of Henle, and another 10–15% is reabsorbed at the distal nephrons (distal convoluted tubules and collecting ducts).

Although the distal nephron is responsible for a relatively small amount of sodium recovery, this segment of renal tubule is important for fine-tuning sodium reabsorption.

There are two pathways by which sodium is reabsorbed from the lumen to the renal tubular cells: the paracellular pathway and the transporters at the luminal membrane (Figure 5).

The paracellular pathway refers to channels between renal tubular cells. Sodium is freely permeable in these channels in the proximal convoluted tubule and the thick ascending limb of the loop of Henle, but not in the distal nephron. The transporters at the luminal membrane are different in various segments of the renal tubule (Table 3). There are generally three

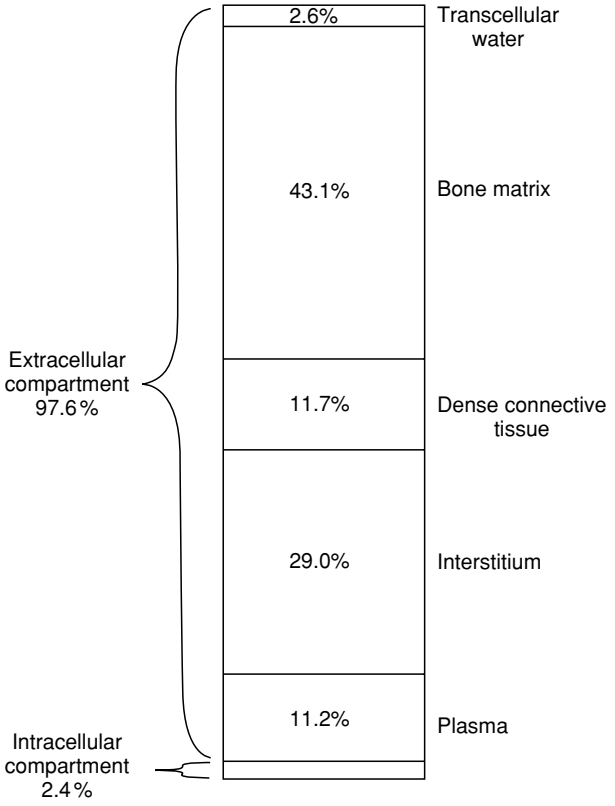

Figure 4 Distribution of body sodium, expressed as a percentage of total body sodium content.

classes of transporters, namely *sodium exchanger*, *sodium cotransporter*, and *specific sodium channel*. The *sodium exchanger* moves sodium into the cell in exchange for another cation (for instance, potassium and hydrogen ion) moving out of the cell. The *sodium cotransporter* couples sodium entry with the entry of other solutes, such as amino acids, glucose, phosphate, potassium, and chloride. The *specific sodium channel* is present in the principal cells of the cortical and medullary collecting duct (the intercalated cell is the other cell type in the collecting duct that is involved in acid or bicarbonate secretion and potassium reabsorption). Under the effect of aldosterone, the channel opens and allows passage of sodium ions only.

The pathways share the same mechanism for sodium transport (Figure 5). There are Na^+–K^+–ATPase pumps that operate across the base of renal tubular cells and also across their lateral walls into the intercellular spaces (also known as the basolateral membrane). The pumps pump out sodium and keep the intracellular sodium at a low level, which then provides a steep electrochemical gradient favoring reabsorption of sodium from the lumen, through the various pathways (paracellular pathway and sodium transporters), into the cells. Having entered the renal tubular cell, the sodium then moves across the basolateral membrane into the peritubular space by the Na^+–K^+–ATPase pumps. The sodium then enters the peritubular capillary and returns back to the ECF.

Regulation of Urinary Sodium Excretion

Numerous mechanisms have been implicated in the regulation of sodium balance and maintenance of a stable ECF. It is unclear whether there are any other mechanisms that have not been discovered yet. The identified mechanisms are overlapped and extensively interrelated. Failure of one of the mechanisms does not result in significant long-term effects. This arrangement is important to maintain stability of the ECF volume. The more constant the ECF volume, the more the body will be protected from changes in the external environment.

The sensing mechanisms include: (1) *mechanoreceptors* in the walls of large blood vessels and cardiac chambers (detecting changes in vascular distension); (2) *chemoreceptors* in the renal tubule (macula densa, responding to an alternation in luminal sodium and chloride ions concentration); (3) *volume receptors* in the central nervous system (sensitive to a change in sodium level in the cerebrospinal fluid); and (4) *peripheral tissue receptors* (monitoring the adequacy of tissue perfusion). The sensors feed back signals to several mediating mechanisms, and the kidneys are then 'informed' by renal nerves, hormones, and physical factors to retain or excrete sodium in the urine.

Factors that affect sodium reabsorption by the renal tubule are listed in Table 3. The major mediating mechanisms are described below.

-Tubuloglomerular feedback The thick ascending limb of the loop of Henle has a cortical segment that closely relates to the afferent arteriole of the parent glomerulus. The cells in this particular segment of the renal tubule constitute the macula densa. The macula densa, the corresponding afferent arteriole, and the nearby extraglomerular mesangium are collectively called the juxtaglomerular apparatus (Figure 3). Changes in the delivery and composition of the fluid flowing past the macula densa elicit changes in glomerular filtration of the same nephron via the interaction of angiotensin II and nitric oxide. For instance, an increase in the delivery of fluid to the macula densa or a low sodium and chloride concentration in the fluid will stimulate the juxtaglomerular apparatus. The renin–angiotensin system is then activated to reduce urinary sodium excretion.

Table 3 Tubular sodium absorption

Renal tubule	*Amount of reabsorbed sodium (%)*	*Sodium reabsorption pathways*		*Factors that stimulate luminal transporter to enhance sodium reabsorption*	*Factors that inhibit luminal transporter to decrease sodium reabsorption*
		Paracellular pathways	*Luminal transporters*		
Proximal convoluted tubule	60–70%	Presence	Sodium–hydrogen ion exchanger	α-Adrenergic agonist[a] Angiotensin II[a] Thyroid hormone[a]	Dopamine[a] Parathyroid hormone[a]
			Sodium–amino acids/glucose/ phosphate cotransporter		
Loop of Henle (thick ascending limb)	20–25%	Presence	Sodium-2 chloride–potassium ion cotransporter	Arginine vasopressin	Loop diuretics Prostaglandin E_2 Cytochrome P450 metabolites of arachidonic acid
Distal convoluted tubule	10%	Absence	Sodium–chloride ion cotransporter	Aldosterone Angiotensin II Sympathetic nerves	Thiazide diuretics
Collecting duct	5%	Absence	Sodium channel	Aldosterone[b] Arginine vasopressin[b]	Potassium-sparing diuretics[b] Atrial natriuretic peptide[b] Bradykinin[b] α_2-Adrenergic agonists[b] Atrial natriuretic peptide[c] Prostaglandin E_2[c]

[a]Factors affecting the sodium–hydrogen ion exchanger.
[b]Factors affecting the sodium channel at the cortical collecting duct.
[c]Factors affecting the sodium channel at the medullary collecting duct.

-Renin–angiotensin system and aldosterone When there is sodium depletion or a reduction in kidney perfusion (for instance, secondary to a low ECF volume), the juxtaglomerular apparatus releases an enzyme, called renin, in the circulation. Renin then acts on angiotensinogen to release angiotensin I. The latter is converted to angiotensin II under the action of the angiotensin-converting enzyme. Angiotensin II stimulates the release of aldosterone, a hormone produced by zona glomerulosa cells of the adrenal cortex.

Angiotensin II reduces urinary sodium excretion by constricting the glomerular arterioles and decreasing the volume of glomerular filtrate, by enhancing sodium reabsorption at the proximal convoluted tubule, and by increasing the responsiveness of tubuloglomerular feedback to a given concentration of sodium and chloride ions at the macula densa. Aldosterone enhances sodium reabsorption by opening up the specific sodium channels at the collecting ducts, and by stimulating the action and synthesis of Na^+–K^+–ATPase pumps at the basolateral membrane of renal tubular cells.

-Sympathetic nervous system The kidneys receive a rich supply of sympathetic nerves. There is evidence that sodium and volume depletion activate renal sympathetic nerves. The kidneys then conserve sodium by constricting the glomerular arterioles, by increasing sodium reabsorption at the proximal convoluted tubules, by increasing renin release from juxtaglomerular apparatus, and by interacting with AVP and atrial natriuretic peptide (ANP).

-Atrial natriuretic peptide Atrial natriuretic peptide is a peptide hormone and is synthesized by atrial myocytes in cardiac atria. When the ECF volume is expanded, the atrial wall distends, and ANP is released. ANP enhances sodium excretion by inhibiting sodium reabsorption at the collecting duct.

Response to Dietary Sodium Intake

Although the relative importance of the different regulatory mechanisms is not well established, many studies suggest that the renin–angiotensin system, aldosterone, and the sympathetic nervous system appear to be more important in conserving sodium

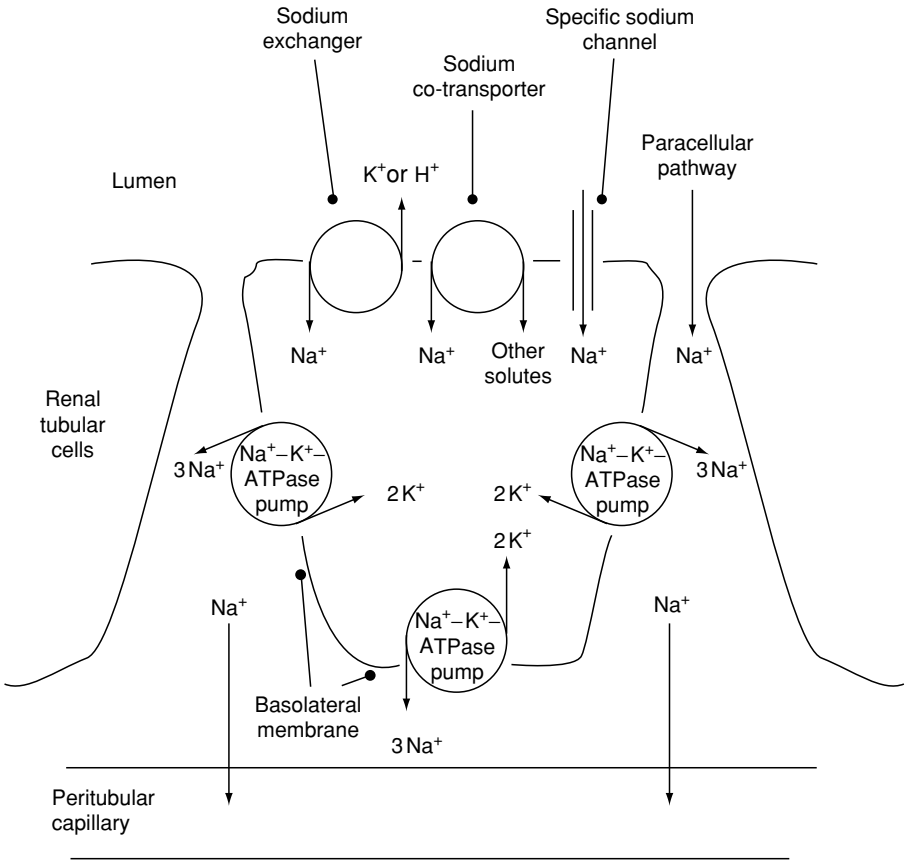

Figure 5 Renal transport of sodium. K^+, potassium ion; H^+, hydrogen ion; Na^+, sodium ion; Na^+–K^+–ATPase pump, sodium–potassium–adenosine triphosphatase pump. The sodium exchanger moves Na^+ into the cell in exchange for another cation (e.g., K^+ and H^+), sodium cotransporter couples sodium entry with the entry of other solutes (e.g., amino acids, glucose, phosphate, K^+ and chloride ion), and the Na^+–K^+–ATPase pump pumps out three Na^+ and transports in two K^+.

for people with a low sodium intake, and the tubuloglomerular feedback seems to be more important in excreting excessive sodium for people with a very high sodium intake.

Potassium Balance

Potassium Content in the Body

The total body potassium content is dependent on sex, age, and, most importantly, muscle mass (which contains 60–75% of total body potassium), and is approximately 40–45, and 50–55 mmol per kilogram of body weight for normal adult women and men, respectively. These values decrease with age and are 20% less in the elderly because of a decrease in muscle mass.

About 90% of the body's potassium is in the ICF. Only 0.4% of the body's potassium is distributed in the vascular space of the ECF (Figure 6).

Potassium is the major cation in the ICF, and the intracellular potassium concentration is approximately 150 mmol l^{-1}. In the ECF, however, the plasma potassium level normally ranges from only 3.5 to 5.0 mmol l^{-1}. The intracellular-to-extracellular potassium concentration ratio is the most important determinant of the resting membrane potential of neuromuscular tissue. Disturbances in this ratio will predispose patients to several clinical disorders, such as cardiac arrhythmia and neuromuscular problems.

Factors that Maintain the Potassium Distribution Between ICF and ECF

Potassium leaks passively from the cell to the ECF space through the ion-selective potassium channels in the cell membrane. A high intracellular potassium concentration is maintained by Na^+–K^+–ATPase pumps that actively transport potassium molecules back into the cells. Factors that alter the activity of Na^+–K^+–ATPase pumps and ion-selective potassium channels affect the transcellular potassium

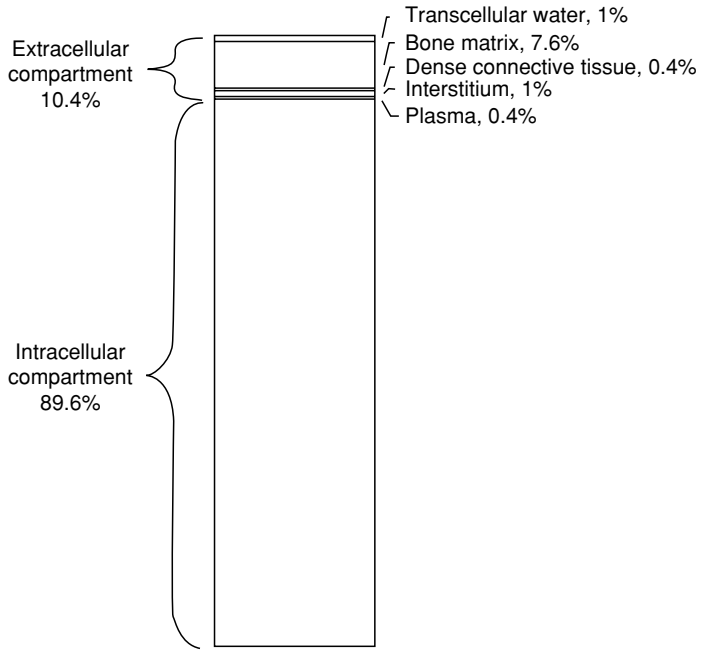

Figure 6 Distribution of body potassium, expressed as a percentage of total body potassium content.

Table 4 Factors that regulate the transcellular potassium distribution

	Movement of potassium
Acid–base status	
Acidosis by mineral acid (e.g., hydrochloric acid)	ICF → ECF
Acidosis by organic acid (e.g., lactic acid, β-hydroxybutyric acid, methylmalonic acid)	No effect
Alkalosis	ECF → ICF
Insulin	ECF → ICF
Glucagon	ICF → ECF
β_2-Adrenergic agonist	ECF → ICF
α-Adrenergic agonist	ICF → ECF
Aldosterone	? ECF → ICF
Hyperosmolality (e.g., mannitol, hyperglycemia)	ICF → ECF
Exercise	ICF → ECF

distribution, resulting in hypo- or hyperkalemia without any change in the total body potassium content. These factors are listed in Table 4.

Potassium Intake and Excretion

Potassium is present in all protein-containing foods, particularly meats. Each gram of dietary protein contains approximately 1 mmol of potassium. Hence, potassium intake is directly related to protein intake. The average Western diet provides 60–100 mmol of potassium each day. In the steady state, about 80–95% of the daily potassium intake is excreted in urine, 5–20% is excreted in the stools (mainly by the colon), and less than 5% is eliminated in sweat. However, in renal failure (with a glomerular filtration rate of less than 20 ml min^{-1}) or diarrhea, potassium loss from the gastrointestinal tract can be significant. Moreover, potassium loss in sweat may be important if perspiration is excessive.

Kidneys and Potassium Balance

About 720 mmol of potassium is filtered through the glomeruli each day. Both potassium reabsorption and secretion occurs in renal tubules. The majority of the filtered potassium (66–75%) is reabsorbed passively at the proximal convoluted tubules and actively (via the sodium-2 chloride–potassium ion cotransporter) in the ascending limbs of the loops of Henle so that only 5–10% of the filtered potassium remains in the early distal convoluted tubules. Fine regulation of potassium excretion occurs at the distal nephron. The collecting duct is the major site of potassium secretion. Under the effect of Na^+–K^+–ATPase pumps at the basolateral membrane of the tubular cell (resulting in a high intracellular potassium concentration) and the effect of sodium absorption via the sodium channels at the luminal membrane (resulting in a negative electrical gradient in the lumen), an electrochemical gradient is set up to promote potassium secretion via the potassium selective channels on the luminal surface of the principal cells of the collecting duct. Factors that modify

Table 5 Factors affecting potassium secretion by the collecting duct

Factors	*Potassium secretion*
Plasma potassium concentration	
↑	↑
↓	↓
Urine flow rate at distal nephron	
↑	↑
↓	↓
Sodium delivered to distal nephron	
↑	↑
↓ (< 30 mM)	↓
Aldosterone	↑
Arginine vasopressin	↑
Poorly absorbed anions in distal nephron	↑
pH	
Acidemia	↓
Alkalemia	↑
Renal ammonia production	↓
Glucocorticoid	↑
Catecholamine	
β_1 adrenergic agonist	↓
α_2 adrenergic agonist	↑
Insulin	↓
Dietary potassium intake	
High	↑
Low	↓
Potassium-sparing diuretic	↓

potassium secretion by the collecting duct are listed in Table 5.

In the medullary collecting duct, potassium reabsorption occurs again. Potassium reabsorption is accompanied by proton secretion that is mediated via hydrogen–potassium ion (H^+–K^+) exchange pumps in the luminal membrane of intercalated cells of the medullary collecting duct. The activity of the H^+–K^+ exchange pump increases with potassium depletion and decreases with potassium loading.

It should be noted that the amount of potassium excreted in the urine depends on the relative degree of potassium secretion and reabsorption at the distal nephron. Moreover, it is clear, therefore, that the net potassium excretion is tightly linked to sodium reabsorption and hydrogen ion secretion in the distal nephron. Generally, renal regulation of sodium and hydrogen ion balance takes priority over regulation of potassium balance.

Response to Dietary Potassium Intake

Urinary excretion of a potassium load occurs over 24–48 h. It is not surprising that other factors, which are rapid and extremely sensitive to a change in potassium intake and output, are involved to maintain the constancy of the ratio of potassium concentration between the ICF and ECF.

High dietary potassium intake triggers several protective mechanisms to increase the cellular uptake of potassium, and to excrete the excessive potassium load by the colon and kidneys. The increase in cellular uptake of potassium is mediated through the release of insulin and catecholamine (Table 4) and probably mediated through an increase in activity of Na^+–K^+–ATPase pumps in skeletal muscle as well. Hyperkalemia stimulates aldosterone secretion. This hormone increases the colonic excretion of potassium directly as well as indirectly through an increase in Na^+–K^+–ATPase pump activity in the basolateral membrane of the colon.

Most of the potassium load is excreted by the kidneys. Similar to the colon, aldosterone increases urinary potassium secretion through an increase in Na^+–K^+–ATPase pump activity, and by increasing the potassium channels of the principal cells of the collecting duct. In addition, a high potassium intake inhibits sodium reabsorption, and hence water reabsorption, at the proximal convoluted tubule and the thick ascending limb of the loop of Henle. Consequently, sodium delivery and urine flow to the distal nephron increase. Moreover, hyperkalemia decreases renal ammonia production, which in turn stimulates aldosterone secretion. All these factors work directly on the distal nephron to increase potassium secretion (Table 5). The adapted kidney can excrete up to 20 times more potassium than the baseline.

Basically, the adaptation response to a low potassium intake is the reverse of that to a high potassium diet. The cellular uptake of potassium decreases by a decrease in the number of Na^+–K^+–ATPase pumps in skeletal muscle but not in other tissues like red blood cells and cardiac muscle. This shifting of potassium extracellularly from skeletal muscle aims to correct an extracellular potassium deficit caused by a low dietary intake. The colon and kidneys also conserve potassium by decreasing the secretion and increasing the reabsorption of potassium, respectively.

See also: **Potassium**: Properties and Determination; Physiology; **Renal Function and Disorders**: Kidney: Structure and Function; **Sodium**: Properties and Determination; Physiology; **Thirst**; **Water**: Physiology

Further Reading

Halperin ML and Goldstein MB (1999) *Fluid, Electrolyte, and Acid–Base Physiology: A Problem-based Approach*. Philadelphia, PA: W B Saunders.

Koeppen BM and Stanton BA (2001) *Renal Physiology*. St. Louis, MO: Mosby.

Kokko JP and Tannen RL (1996) *Fluids and Electrolytes*. Philadelphia, PA: W B Saunders.

Acid–Base Balance

E W Kun and A W Yu, The Chinese University of Hong Kong, Hong Kong, China

Introduction

Virtually all cellular, tissue, and organ systems are sensitive to changes in the hydrogen ion (H^+) concentration in the body. They function best at an extracellular fluid H^+ concentration of 35–45 nmol l^{-1} (each nmol l^{-1} equals 10^{-6} mmol l^{-1}). This concentration is extremely low compared with other ions; sodium (Na^+), for example, has a 3 million times greater concentration. Therefore, H^+ concentration is commonly expressed as pH, a negative logarithm of H^+ concentration. Maintenance of acid–base balance in the body is important because H^+ binds avidly to proteins and changes their shape and function. Many essential metabolic, enzymatic, and transmembrane transport processes can be jeopardized because of alterations of the protein structure and function, resulting in severe organ dysfunction and clinical consequences. With an increase in body H^+ concentration, a patient may experience hypotension, depressed myocardial contractility, and sensorium. With decreased body H^+ concentration, the patient may develop coronary spasm, cardiac arrhythmias, hyperreflexia, muscle spasm, and seizure (Table 1).

In normal conditions, H^+ varies little from the normal value of approximately 40 nmol l^{-1}, even though acids and bases are continually being added to the extracellular fluid. This can be achieved by three homeostatic mechanisms: (1) intracellular and extracellular buffers; (2) changing partial pressure of carbon dioxide in the blood by altering the ventilation rate; and (3) renal H^+ excretion. The basic principles of acid–base physiology and the role of the kidneys and lungs in acid–base balance will be discussed in this article.

General Concepts

An acid is defined as a compound capable of donating an H^+, and a base is a compound that is capable of accepting an H^+. Examples of acids in the body include H_2CO_3, HCl, NH_4^+ and $H_2PO_4^-$, and examples of bases in the body are HCO_3^-, Cl^-, NH_3 and HPO_4^{2-}.

Food and cellular metabolism produces more acid than base. Base is lost in the feces daily. The net effect is the addition of acid to the body fluids. To maintain acid–base balance, acid must be excreted from the body at a rate equivalent to its addition. Acidosis (a disease process that tends to decrease the pH) results if acid addition exceeds excretion. Conversely, alkalosis (a disease process that tends to increase the pH) results if acid excretion exceeds addition. Systemic pH can still be maintained within a normal range at acidosis or alkalosis because of compensatory mechanisms, which are discussed in the section below. If these conditions or processes are left unopposed, the pH will increase or decrease from the normal range, and result in acidemia ($pH < 7.35$) or alkalemia ($pH > 7.45$). Metabolic acidosis results from a primary reduction in plasma bicarbonate concentration and respiratory acidosis results from a primary increase in CO_2 partial pressure (P_{CO_2}). Metabolic alkalosis is a result of a primary decrease in plasma bicarbonate concentration and respiratory alkalosis is a result of a primary decrease in P_{CO_2}. An increase in P_{CO_2} secondary to compensation for metabolic alkalosis is not called respiratory acidosis; a secondary decrease in P_{CO_2} in response to metabolic acidosis is not called respiratory alkalosis.

Acid and Alkali Generation

The major constituents of the human diet are carbohydrates, fats, and proteins. In normal conditions, 15 000 mmol CO_2 is generated from the metabolism of carbohydrates and fats. Although not an acid, this CO_2 will combine with H_2O to form H_2CO_3; therefore, it is termed a volatile acid, reflecting the fact that it has the potential to generate H^+ after hydration with H_2O. An accumulation of this endogenously produced CO_2 will result in respiratory acidosis (Table 2). Fortunately, this large quantity of CO_2 is eliminated by alveolar ventilation, and acid–base balance is maintained.

However, nonvolatile acids (mostly sulfuric acid) are primarily generated from the metabolism of proteins. For example, oxidation of the sulfur-containing amino acids cysteine and methionine yields sulfuric acid (H_2SO_4), whereas metabolism of lysine, arginine, and histidine generates hydrochloric acid (HCl).

Table 1 Clinical manifestations of an acid–base imbalance

Increased acidity (pH < 7.20)	*Increased alkalinity (pH > 7.55)*
Hypotension	Perioral and extremity paresthesias
Depressed myocardial contractility	Muscle spasms
Mental confusion	Hyperreflexia
	Seizures
	Coronary spasm
	Cardiac arrhythmias

Table 2 Generation of acid

Volatile acid – respiratory acidosis
CO_2 production (cellular metabolism)
Nonvolatile acids – metabolic acidosis
Dietary protein and phospholipid metabolism (sulfuric acid, phosphoric acid)
Diabetes mellitus (β-hydroxybutyric acid and acetoacetic acid)
Hypoxia (lactic acid)
Decreased tissue perfusion, e.g., vigorous exercise, heart failure, shock
Decreased renal H^+ excretion (e.g., renal failure, renal tubular acidosis)
Loss of HCO_3^-
Gastrointestinal losses in stool, e.g., diarrhea, external fistulae
Urinary losses of HCO_3^-, e.g., proximal renal tubular acidosis, drug (acetazolamide)

Part of this nonvolatile acid is neutralized by the bicarbonate (HCO_3^-) generated through metabolism of the amino acids, such as aspartate and glutamate, and certain organic anions, such as citrate. The net amount of nonvolatile acid produced depends on dietary intake. Ingestion of a meal with a large amount of meat will result in more acid production, whereas a vegetarian meal can result in less acid generation. In general, approximately 1 mmol kg^{-1} body weight of nonvolatile acid is added to the body each day (50–100 mmol day^{-1} for adults) on a typical diet. The nonvolatile acids are buffered initially, followed by renal excretion of the H^+.

In abnormal clinical conditions, e.g., diabetes mellitus and hypoxia, whereby carbohydrates and fats are incompletely metabolized to CO_2 and H_2O, other nonvolatile acids are produced (Table 2). In diabetes mellitus, when insulin levels are abnormally low, fatty acids are oxidized to yield energy. This leads to the production of acetone and two ketoacids (β-hydroxybutyric and acetoacetic acid). During hypoxia, oxygen delivery to cells is inadequate, glucose is metabolized anaerobically to pyruvate and then to lactate and an H^+ (one per lactate). Therefore, normal individuals performing vigorous exercise, patients with reduced cardiac output (heart failure), and those with hypotension (shock) usually have higher lactic acid levels. The acid produced will combine with HCO_3^- and lower the plasma HCO_3^- concentration producing metabolic acidosis:

$$H^+ + HCO_3^- \rightarrow H_2CO_3$$

With an increased production of acid, the kidneys may increase excretion of H^+ and maintain the acid–base balance. However, in the presence of renal diseases (renal failure or renal tubular acidosis), endogenous acid production or ingestion of acid-producing substances, e.g., aspirin (salicylic acid), methanol, and ethylene glycol, may exceed the capacity of the buffer system and renal excretion of H^+. This will result in metabolic acidosis (Table 2).

The loss of HCO_3^- from the body is also equivalent to the addition of acid and produces metabolic acidosis. This is commonly associated with gastrointestinal or renal loss of bicarbonate (Table 2).

Whereas the stomach is an acid-secreting organ, the gastrointestinal tract distal to the stomach is bicarbonate-secreting. The small bowel has a daily secretory volume of 600–700 ml and this may be markedly increased if the small bowel is diseased. The biliary system secretes about 1 l of fluid per day containing an HCO_3^- concentration of 60 mmol l^{-1}, whereas the pancreas secretes 2 l day^{-1} with an HCO_3^- concentration reaching 120 mmol l^{-1}. Most of the secretions are reabsorbed, but with diarrhea and external drainage of pancreatic, biliary, or small-bowel juice (external fistulae), the loss of HCO_3^--rich fluid results (Table 2).

As mentioned above, alkali can also be produced by the metabolism of certain amino acids but it is neutralized by the larger amount of dietary acid produced. A net alkali excess can only occur when there is exogenous alkali ingestion, primary increase in ventilation with lowering of P_{CO_2} (respiratory alkalosis), and excessive H^+ loss from the stomach or kidneys (metabolic alkalosis). Milk-alkali syndrome is seen in patients with gastric distress who consume large amounts of milk and antacids containing calcium carbonate and sodium bicarbonate. Suppression of parathyroid hormone secretion caused by hypercalcemia from absorbed calcium contributes to failure to excrete the alkali load. Chronically, nephrocalcinosis and renal failure develop, which further reduces the excretion of the absorbed alkali, causing metabolic alkalosis (Table 3).

The Buffering System

One of the major ways in which large changes in H^+ concentration are prevented is by buffering. A 'buffer'

Table 3 Generation of alkali

Volatile alkali – respiratory alkalosis
CO_2 removal (hyperventilation)
Nonvolatile alkali – metabolic alkalosis
Loss of acid
Gastrointestinal losses in gastric fluid, e.g., vomiting, gastric suction
Urinary losses of H^+: diuretics, licorice excess, hyperaldosteronism, Bartter's syndrome, Cushing's syndrome
Exogenous alkali loads
Milk-alkali syndrome
Sodium bicarbonate ingestion

is a pair of substances that can donate or accept an H^+ to moderate changes in H^+ concentration or pH. This is the most immediate mechanism of defense against changes in physiological pH. The major buffers in the body are H_2CO_3-HCO_3^- (pK_a 6.1), $H_2PO_4^-$-HPO_4^{2-} (pK_a 6.8), and H-protein$^-$. The dissociation constant (pK_a) and pH determine whether a buffer pair will bind or release the H^+ in the solution. When the pH is 1.0 unit below the pK_a, H^+ will bind to the buffer pair, whereas at 1.0 unit above the pK_a, H^+ will release.

Bicarbonate/carbon dioxide is the most important extracellular buffer in the body. This buffer system can be described by the following reactions:

$$\underset{\text{(gaseous)}}{CO_2} \leftrightarrow \underset{\text{(aqueous)}}{CO_2 + H_2O} \leftrightarrow H_2CO_3 \leftrightarrow H^+ + HCO_3^- \quad \text{(Carbonic anhydrase)}$$

$$40\,\text{mmHg}\ (5.3\text{kPa}) \qquad 1.2\,\text{mmol}\,l^{-1} \qquad 24\,\text{mmol}\,l^{-1} \tag{1}$$

The hydration/dehydration of CO_2 is the rate-limiting step and is enhanced by the enzyme carbonic anhydrase (CA). When H_2CO_3 is formed, it is ionized to H^+ and HCO_3^- virtually instantaneously. The relationship of pH, HCO_3^-, and P_{CO_2} is described by the Henderson–Hasselbalch equation:

$$\text{pH} = \text{p}K_a + \log\{[HCO_3^-] \div [0.03 \times P_{CO_2}]\}$$

where pK_a = 6.1, HCO_3^- is in mmol$\,l^{-1}$, and P_{CO_2} is in mmHg.

When nonvolatile acid is added to the body fluids, the reaction of eqn (1) is driven to the left. HCO_3^- is consumed during the buffering and plasma HCO_3^- concentration is reduced. In contrast, when acid is removed from the body, the reaction is driven to the right. More HCO_3^- is produced from the dissociation of H_2CO_3 (carbonic acid) and HCO_3^- concentration increases. With a normal plasma HCO_3^- concentration of 24 mmol$\,l^{-1}$ and an extracellular fluid volume of 14 l, this HCO_3^- buffer system can potentially buffer 340 mmol acid. However, for optimal function of this system, regeneration and maintenance of normal HCO_3^- concentration by the lungs and kidneys are required.

In addition, the bones can release buffer compounds such as $NaHCO_3$, $K\,HCO_3$, $CaCO_3$ and $Ca\,HPO_4$ into extracellular fluid when stimulated by a decrease in the plasma HCO_3^- concentration. It has been estimated that as much as 40% of the buffering of acute acid load takes place in the bones. The role of bone buffers is even greater in chronic acidosis, such as that seen in chronic renal failure. One of the consequences is that acid loading directly increases Ca^{2+} release from the bones and urinary Ca^{2+} excretion, promoting stone formation. An inappropriately high-protein diet may increase acid load (a normal diet produces about 70 mmol acid per day) and promote calcium stone formation. In patients with end-stage renal disease, a disorder associated with progressive acid retention caused by impaired urinary acid excretion, there is a gradual reduction in bone calcium stores.

Plasma phosphate and protein also act as extracellular buffers but they are less important quantitatively (plasma phosphate concentration is 1 mmol$\,l^{-1}$ versus 24 mmol$\,l^{-1}$ HCO_3^-).

$$H^+ + HPO_4^{2-} \leftrightarrow H_2PO_4^- \tag{2}$$

$$H^+ + \text{protein}^- \leftrightarrow \text{H-protein} \tag{3}$$

However, intracellularly, phosphate and protein are important buffers besides HCO_3^-. In metabolic acidosis, about 43% of buffering occurs in the extracellular fluid and 57% of buffering is mediated by intracellular buffering. Whereas extracellular H^+ is buffered immediately, buffering by intracellular buffers takes about 2–4 h. This is the time required for H^+ to diffuse into interstitial spaces and enter cells, which occurs more slowly.

Respiratory Control of Acid–Base Balance

Alveolar ventilation provides oxygen for oxidative metabolism and eliminates the CO_2 produced by these metabolic processes in order to maintain P_{CO_2} at about 5.3 kPa (40 mmHg). Approximately 15 000 mmol CO_2 is produced daily and eliminated by the lungs. The stimulation to ventilation is located in the chemoreceptors in the brainstem respiratory center (medulla oblongata) and to a lesser extent in the carotid bodies located near the bifurcation of the carotid arteries. The chemoreceptors are sensitive to changes in the cerebral interstitial pH. With increased CO_2 in the blood, cerebral interstitial pH will decrease and result in the stimulation of ventilation that will return P_{CO_2} to normal. In contrast, interstitial pH will increase with decreased plasma CO_2, resulting in ventilation inhibition (**Figure 1**).

Alveolar ventilation is also affected by metabolic acid–base imbalance. Acidemia caused by metabolic acidosis can increase minute ventilation from the normal of approximately 5 l min^{-1} to greater than 30 l min^{-1} as the arterial pH decreases from 7.40 to 7.00. The central chemoreceptors are relatively insulated by the blood–brain barrier and only CO_2, but not HCO_3^-, rapidly crosses the blood–brain barrier. Hence, the effect on cerebral interstitial pH is slower, and a full ventilation response to nonvolatile acid or a

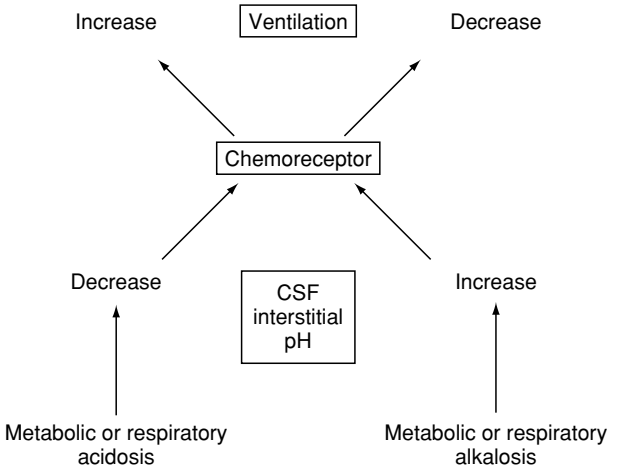

Figure 1 Respiratory control of acid–base balance. CSF, cerebrospinal fluid.

decrease in plasma HCO_3^- may only be achieved in 12–24 h. The P_{CO_2} will decrease by 0.16 kPa for every mmol l^{-1} reduction of HCO_3^- concentration. Conversely, ventilation decreases with metabolic alkalosis with a consequent increase in P_{CO_2} (0.08 kPa per every mmol l^{-1} increase in HCO_3^-), lowering the pH towards normal.

An acid–base imbalance can also be initiated by a change in carbon dioxide tension of body fluids. A primary increase in P_{CO_2} (primary hypercapnia), $P_{CO_2} > 6.0$ kPa, results in respiratory acidosis. This needs to be distinguished from the secondary hypercapnia that results from the correction of metabolic alkalosis. Hypercapnia (respiratory acidosis) could develop from increased CO_2 production, decreased CO_2 excretion, or both (Table 4).

The administration of large carbohydrate loads (greater than 2000 kcal day^{-1}) and parenteral nutrition to critically ill patients may increase CO_2 production. This is rarely the sole cause of hypercapnia. Increased CO_2 decreases cerebral interstitial pH, resulting in an increment in respiratory drive, and the overproduction of CO_2 is usually matched by increased excretion and the generation of respiratory acidosis is prevented. Patients with a marked limitation in respiratory reserve (e.g., chronic obstructive lung disease) and those receiving constant mechanical ventilation may not be able to match the CO_2 production with its removal, resulting in respiratory acidosis (Table 4). Therefore, it is recommended that patients with chronic obstructive airway disease avoid a high-carbohydrate diet to avert further retention of CO_2.

However, excessive alveolar ventilation relative to the prevailing carbon dioxide production results in primary hypocapnia and respiratory alkalosis. This is usually attributed to hyperventilation as a result of increased ventilatory drive, hypoxia, and maladjusted mechanical ventilators (Table 5).

The body responds to hypercapnia and hypocapnia in two phases (acute and chronic). The acute phase takes about 5–10 min to complete. If hyper/hypocapnia persists (chronic phase), adaptation requires a few days to complete.

In acute hypercapnia, CO_2 diffuses into the cells, forming carbonic acid which is buffered by intracellular buffers (such as phosphate or hemoglobulin), resulting in an increase of plasma bicarbonate concentration (Figure 2). On average, plasma bicarbonate concentration increases by 0.8 mmol l^{-1} for each kPa acute increase in P_{CO_2}. Mild increases of Na^+ and K^+ are also observed in addition to the increment in plasma bicarbonate concentration.

Table 4 Causes of primary respiratory acidosis

Increased CO_2 production
High-carbohydrate diet with constant mechanical ventilation
CO_2 insufflation during endoscopic procedures
Decreased CO_2 removal
Pulmonary disease
Asthma
Chronic obstruction airway disease
Pulmonary fibrosis
Mechanical ventilatory defect
Pneumothorax
Hemothorax/hydrothorax
Adult respiratory distress syndrome
Muscular or neuromuscular diseases (e.g., Guillain–Barré syndrome, amyotrophic lateral sclerosis, multiple sclerosis)
Defect in respiratory drive
Brainstem infarct
Sleep apnea
Drugs (opiates and sedatives)
Other
Cardiac arrest
Shock
Severe pulmonary edema
Massive pulmonary embolus

Table 5 Causes of primary respiratory alkalosis

Increased central nervous system respiratory drive
Anxiety
Brain infarction, trauma
Sepsis and fever
Drugs (salicylate, nicotine, doxapram)
Tissue hypoxia
Severe anemia
High altitude
Cyanotic heart disease
Pneumonia
Iatrogenic
Increased mechanical ventilation

If hypercapnia persists, an adaptive increment in plasma bicarbonate concentration is markedly amplified as a result of increasing renal acid excretion, generating new bicarbonate that is returned to the blood. Plasma bicarbonate increases by 3.0 mmol l^{-1} with each kPa chronic increase in P_{CO_2}. The renal response usually takes 1–2 days to develop fully.

In contrast, adaptation to hypocapnia is associated with an immediate decrement in plasma bicarbonate of 1.5 mmol l^{-1} for each kPa acute decrease in P_{CO_2}. With persisting hypocapnia, renal adaptation requires 3–4 days to complete. This results in a larger decrement in plasma bicarbonate (3.8 mmol l^{-1} for each kPa of chronic decrease in P_{CO_2}). The relative changes of P_{CO_2}, HCO_3^-, and pH in various acid–base disorders are summarized in Table 6.

Renal Control of Acid–Base Balance

Despite the efficacy of respiratory and buffering systems in maintaining the acid–base balance, the ventilation rate is limited, and the body buffers eventually deplete if the dietary acid load (50–100 mmol H^+) is not excreted in the urine. The kidneys have the important role of regulating the systemic bicarbonate concentration and defending against acid–base imbalance. The kidneys make appropriate adjustments in the secretion of HCO_3^- and H^+. The renal response requires several days to complete. There are two components in this regulatory process: (1) reabsorption of filtered HCO_3^-, and (2) generation of bicarbonate by net H^+ excretion.

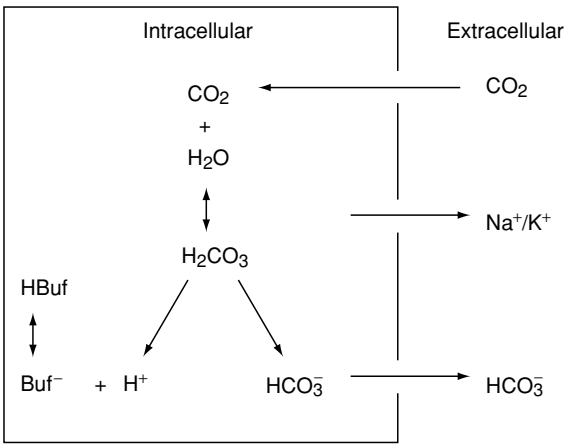

Figure 2 Intracellular buffering response to respiratory acidosis. Buf^-, buffer.

Reabsorption of Filtered Bicarbonate

The glomeruli filter about 4300 mmol bicarbonate per day. About 80% of the filtered bicarbonate is reabsorbed in the proximal tubules, whereas the remainder is reabsorbed in the distal segments. CA has a central role in reabsorption of HCO_3^-. The filtered HCO_3^- generates carbonic acid (H_2CO_3) by combining with the excreted H^+. CA, located in the brush border of the proximal tubular cells, rapidly dissociates H_2CO_3 into water and CO_2. CO_2 then diffuses into the tubular cells where it combines with the OH^- from H_2O, forming HCO_3^- and H^+. H^+ is excreted by Na^+-H^+ antiporters and the HCO_3^- leaves the cells and enters the peritubular capillary blood by Na^+–3 HCO_3^- cotransporter in the basolateral membrane. The tubular fluid pH decreases from 7.40 in the filtrate to about 6.70 by the end of the proximal convoluted tubule (Figure 3).

Renal Tubular H^+ Excretion

The above process reclaims the filtered bicarbonate but does not participate in the excretion of the dietary acid load. The H^+ requires combination with buffers or the formation of ammonium for excretion. Weak acids (titratable acids) filtered from the glomeruli, mostly monobasic phosphate (HPO_4^{2-}), act as important buffers in the tubular fluid (Figure 3). This can account for the excretion of 10–40 mmol H^+ per day. Buffering by phosphate and, to a lesser extent, other buffers such as urate and creatinine is called titratable acidity. In the presence of an acid load, the amount of titratable acidity cannot be easily increased.

Table 6 Changes of pH, P_{CO_2} (kPa) and HCO_3^- (mmol l^{-1}) in simple acid–base

Disorder	*pH*	P_{CO_2}	HCO_3^-	*Expected changes*
Metabolic acidosis	↓	↓	↓	$\Delta P_{CO_2} = 0.16 \times \Delta [HCO_3^-]$
Metabolic alkalosis	↑	↑	↑	$\Delta P_{CO_2} = 0.08 \times \Delta [HCO_3^-]$
Respiratory acidosis	↓	↑	↑	
Acute				$\Delta [HCO_3^-] = 0.8 \times \Delta P_{CO_2}$
Chronic				$\Delta [HCO_3^-] = 3.0 \times \Delta P_{CO_2}$
Respiratory alkalosis	↑	↓	↓	
Acute				$\Delta [HCO_3^-] = 1.5 \times \Delta P_{CO_2}$
Chronic				$\Delta [HCO_3^-] = 3.8 \times \Delta P_{CO_2}$

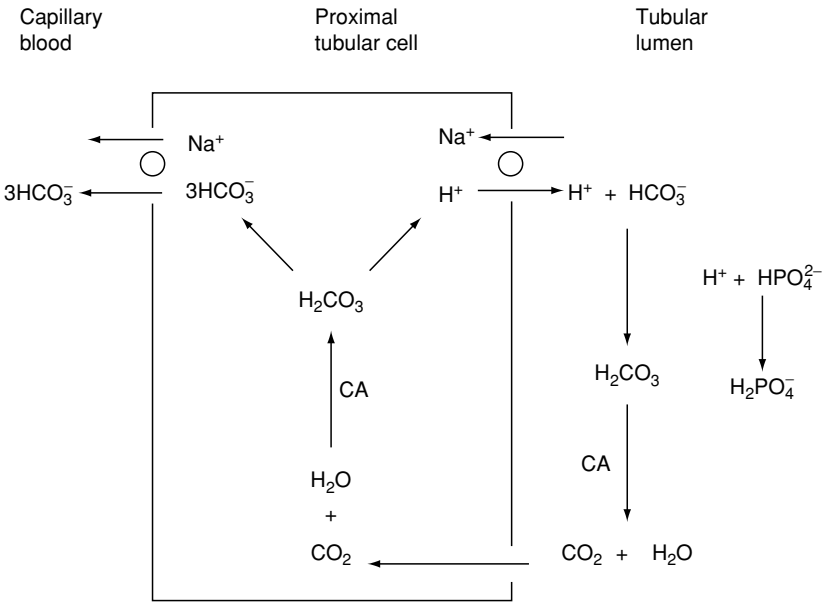

Figure 3 Bicarbonate reabsorption and acid excretion in the proximal tubule. CA, carbonic anhydrase.

Ammonium excretion constitutes the major adaptive response to an acid load.

Ammonium Excretion

The excretion of H^+ as ammonium (NH_4^+) is an important step in acid–base regulation because the rate of NH_4^+ production and excretion can be regulated in response to the physiological needs. NH_4^+ excretion in the urine represents acid excretion and generation of new bicarbonate by the kidneys.

$$H^+ + NH_3 \leftrightarrow NH_4^+$$

The normal rate of NH_4^+ excretion in the urine is about 30–40 mmol day^{-1}. It can be increased to > 300 mmol day^{-1} after a maximum acid load.

NH_4^+ production in the kidneys takes place predominantly in the proximal tubules. Glutamine (primarily from the liver) is the major precursor of ammonia in the kidneys. Glutamine is metabolized in the kidneys to glutamate and NH_4^+.

$$\text{Glutamine} \rightarrow 2NH_4^+ + \text{glutamate}^- \rightarrow 2NH_4^+ + \alpha\text{-ketoglutarate}^{2-}$$

The metabolism of α-ketoglutarate generates two HCO_3^- that are returned to the systemic circulation.

In the proximal tubules, NH_4^+ is secreted into the tubular lumen by Na^+-H^+ antiporter with NH_4^+ substituting the H^+, and the HCO_3^- enters the blood (Figure 4a). Some secreted NH_4^+ is reabsorbed in the ascending limbs of loops of Henle into the medullary interstitium where it exists as both NH_4^+ and NH_3. In the collecting tubules, lipid-soluble ammonia (NH_3) passively diffuses into the lumen and combines with secreted H^+ to form NH_4^+ (Figure 4b). Cationic NH_4^+ is lipid-insoluble and excreted in the urine because back-diffusion into the cells cannot occur.

Regulation of Renal Bicarbonate Reabsorption and Acid Excretion

A variety of factors, such as intracellular pH, effective circulating volume, changes in plasma potassium concentration, as well as several hormones, may affect bicarbonate reabsorption and acid secretion, leading to alkalosis or acidosis (Table 7).

Decreases in intracellular pH, as a result of metabolic acidosis (increased H^+) or respiratory acidosis (increased P_{CO_2}), will increase the availability of H^+ for excretion. At increased acid load or metabolic acidosis, plasma bicarbonate concentration is decreased. A greater concentration gradient allows HCO_3^- to diffuse out of the tubular cells. A decrease in intracellular HCO_3^- will lower intracellular pH and provide a signal to increase HCO_3^- reabsorption and H^+ secretion. In respiratory acidosis or alkalosis, changes in intracellular pH are mediated by the diffusion of lipid-soluble CO_2 into or out of the cells.

The effective volume status can influence bicarbonate reabsorption. Decreases in the effective circulating volume increase bicarbonate reabsorption, whereas increases in the effective circulating volume

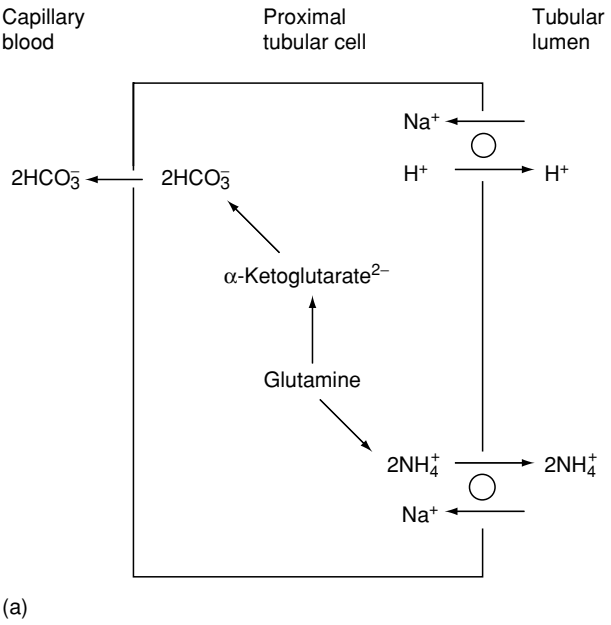

(a)

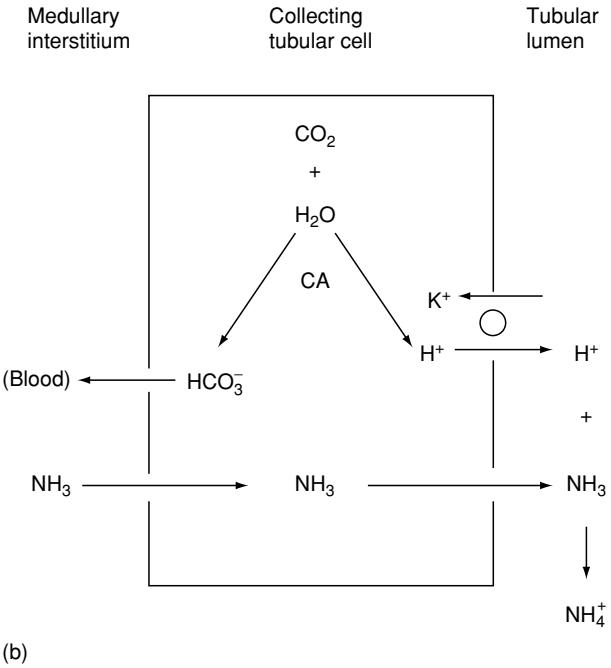

(b)

Figure 4 Acid excretion as ammonium in (a) the proximal tubule; (b) the collecting tubule. CA, carbonic anhydrase.

inhibit proximal bicarbonate reabsorption. This is clinically important in patients with volume-depleted metabolic alkalosis. The inability to excrete excess HCO_3^- prevents resolution of metabolic alkalosis. In this situation, Na^+ reabsorption (together with HCO_3^-) prevents further volume depletion at the expense of systemic pH.

Table 7 Factors affecting renal HCO_3^- reabsorption and/or H^+ excretion

Increase	*Decrease*
Metabolic acidosis	Respiratory alkalosis
Metabolic alkalosis with volume depletion	Volume expansion
Respiratory acidosis	Hyperkalemia
Volume depletion	Parathyroid hormone
Hypokalemia	
Angiotensin II	
Aldosterone	

Table 8 Sequential event in response to an acid or alkali load

	Duration
Extracellular HCO_3^- buffering	Immediate
Respiratory response by changing P_{CO_2}	Minutes to hours
Intracellular and bone buffering	2–4 h
Adjustment in renal HCO_3^- reabsorption and H^+ excretion	24 h to few days

The plasma potassium concentration varies inversely with both HCO_3^- reabsorption and H^+ excretion. With potassium loss, hypokalemia causes a transcellular shift of potassium out of the cells (which contain approximately 98% of total body potassium). To maintain electroneutrality, Na^+ and H^+ move into the cells. Intracellular acidosis results, stimulating HCO_3^- reabsorption and H^+ excretion.

Several hormones also influence bicarbonate reabsorption and H^+ excretion. Angiotensin II stimulates bicarbonate reabsorption in the early proximal tubules by increasing the activity of both the Na^+-H^+ antiporters and the basolateral Na^+-$3HCO_3^-$ cotransporters. Aldosterone, however, acts on distal nephrons and promotes H^+ excretion in the collecting tubules.

The parathyroid hormone inhibits proximal HCO_3^- reabsorption by inhibiting Na^+-H^+ antiporters in luminal membranes and Na^+-$3HCO_3^-$ cotransporters in basolateral membranes. The physiological importance is unknown because the extra HCO_3^- delivered out of the proximal tubules is mostly reabsorbed in the distal tubules. However, patients with primary hyperparathyroidism and hypercalcemia tend to have metabolic acidosis.

Summary

In summary, the pH of body fluids is maintained within a narrow limit by the coordinated function of the buffers, lungs, and the kidneys. Volatile and nonvolatile acids together with other acids and alkali ingested must be excreted to maintain acid–base

balance. The sequential responses to acid or alkali load are depicted in Table 8. The kidney is the ultimate organ to control acid–base balance by adjusting the amount of H^+ excretion. With kidney failure, acid–base balance cannot be maintained despite the presence of other intact buffering systems.

See also: **Acids**: Properties and Determination; Natural Acids and Acidulants; **Amino Acids**: Metabolism; **Diabetes Mellitus**: Secondary Complications; **Glucose**: Function and Metabolism; **pH – Principles and Measurement**; **Phosphorus**: Physiology; **Renal Function and Disorders**: Kidney: Structure and Function

Further Reading

Druml W (2001) Nutritional management of acute renal failure. *American Journal of Kidney Disease* 37 (Suppl 2): S89–S94.

Halperin ML and Goldstein MB (1999) *Fluid, Electrolyte, and Acid–Base Physiology: A Problem-Based Approach*, 3rd edn. Philadelphia: W.B. Saunders.

Henger A, Tutt P, Riesen WF, Hulter HN and Krapf R (2000) Acid–base and endocrine effects of aldosterone and angiotensin II inhibition in metabolic acidosis in human patients. *Journal of Laboratory and Clinical Medicine* 136: 379–389.

Kellum JA (2000) Determinants of blood pH in health and disease. *Critical Care* 4: 6–14.

Khanna A and Kurtzman NA (2001) Metabolic alkalosis. *Respiratory Care* 46: 354–365.

Koeppen BM and Stanton BA (2001) *Renal Physiology*, 3rd edn. St Louis: Mosby.

Kraut JA (2000) Disturbances of acid–base balance and bone disease in endstage renal disease. *Seminars in Dialysis* 13: 261–266.

Kraut JA and Madias NE (2001) Approach to patients with acid–base disorders. *Respiratory Care* 46: 392–403.

Oh MS (2000) New perspectives on acid–base balance. *Seminars in Dialysis* 13: 212–219.

Remer T (2000) Influence of diet on acid–base balance. *Seminars in Dialysis* 13: 221–226.

Rose BD and Post TW (2001) *Clinical Physiology of Acid–Base and Electrolyte Disorders*, 5th edn. New York: McGraw-Hill.

Silver RB, Breton S and Brown D (2000) Potassium depletion increases proton pump (H^+-ATPase) activity in intercalated cells of cortical collecting duct. *American Journal of Physiology and Renal Physiology* 279: F195–F202.

Sugiura S, Inagaki K, Noda Y, Nagai T and Nabeshima T (2000) Acid load during total parenteral nutrition: comparison of hydrochloric acid and acetic acid on plasma acid–base balance. *Nutrition* 16: 260–263.

ELECTROPHORESIS

R J Fritsch, Kraft General Foods Research and Development Inc., Munich, Germany
I Krause, Technical University Munich, Weihenstephan, Germany

Background

Electrophoretic techniques are among the premier methods for the separation and analysis of proteins in foods. This article reviews the basic principles of analytical electrophoresis and the most important separation and detection modes, including the use of immunoreagents, and applications to food analysis/authentication. (*See* **Protein**: Determination and Characterization.)

Principles

Electrophoresis is a general term that describes the migration and separation of charged particles (ions) under the influence of an electric field. An electrophoretic system consists of two electrodes of opposite charge (anode, cathode), connected by a conducting medium called an electrolyte. The separation effect on the ionic particles results from differences in their velocity (v), which is the product of the particle's mobility (m) and the field strength (E):

$$v = mE. \quad (1)$$

The mobility (m) of an ionic particle is determined by particle size, shape, and charge, and the temperature during the separation, and is constant under defined electrophoretic conditions.

Electrophoretic conditions are characterized by the electrical parameters (current, voltage, power), and factors such as ionic strength, pH value, viscosity, pore size, etc., which describe the medium in which the particles are moving.

The removal of heat generated by the passage of electric current is one of the major problems in most forms of electrophoresis. Any temperature difference causes variations in the rates of migration through the

Table 1 Modes of electrophoresis and basic characteristics of systems

Mode	*Characteristics*
Zone electrophoresis	Continuous electrolyte systems, continuous pH and ionic strength, sieving effect possible depending on support medium
Isotachophoresis	Discontinuous electrolyte system, concentrating effect, migration at same velocity
Isoelectric focusing	Continuous electrolyte system, stable and linear pH gradient, no molecular sieving effect

Table 2 Important electrophoretic techniques and types of detection in food analysis

Format	*Electrophoretic technique*	*Type of detection*
In free solution	Capillary zone electrophoresis	On-line electrical
	Capillary isotachophoresis	On-line optical
		On-line thermal
In nonsieving medium	Zone electrophoresis in agarose	Staining for proteins[a]:
	Isoelectric focusing in agarose	Coomassie Blue, Amido Black
		silver
		colloidal gold
In sieving medium	Zone electrophoresis in polyacrylamide:	Localization of specific constituents[a]:
	in homogeneous gel and buffer system	assays for enzyme activity
	in multiphasic gel and buffer system	glyco-, lipo-, phosphoproteins
	with pore size gradients	immunofixation, immunoprinting
	in presence of SDS detergent	
	Isoelectric focusing in polyacrylamide	Blotting on immobilized matrix[a]:
		normal staining methods
		immunoblotting/enzyme labelling

[a]These types of detection are applicable to both the sieving and nonsieving media.

medium, resulting in distortion in the bands of separated molecules. Clearly, it would be ideal if electrophoretic analyses could be carried out at a constant temperature.

The different separation modes and their basic characteristics are summarized in Table 1; a more detailed discussion follows in subsequent sections.

Formats

Generally, all modes of electrophoresis can be carried out either in 'free solution,' in which no anticonvective stabilizers are used (see *capillary zone electrophoresis*), or in a 'support medium,' where anticonvective support matrices suppress the thermally driven convection currents and diffusion in the electrophoretic medium.

In supporting media, the mobility and sharpness of separations can be influenced by additional factors. These include adsorption and ion-exchange effects with the matrix, inhomogeneities within the support matrix, and electroendosmosis. Furthermore, support media offer visualization of separated zones in a conventional format where strips, foils, and slabs can be easily stained, destained, and manipulated in ways not possible for free solutions.

The most commonly practised formats and techniques of electrophoresis in food analysis are listed in Table 2. Gels made from polyacrylamide and agarose are the supporting media of choice today. The use of cellulose acetate strips is popular for routine screening work where the ease of handling, commercial availability of ready-to-use material, and rapidity are relevant attributes. Paper and thin-layer electrophoresis (e.g., on silica gel) are successfully used for the analysis of high-molecular-weight polysaccharides and lipopolysaccharides.

Agarose Gel

Agarose is a highly purified polysaccharide derived from agar, a natural product of red seaweed. Commercially available agarose materials show different, well-characterized levels of electroendosmosis, due to the presence of sulfate and carboxyl groups in agar. Although electrophoresis on agarose gel has been eclipsed by the use of polyacrylamide in the analysis of most proteins and glycoproteins, it remains invaluable in applications where a very large pore size and hence nonrestrictive gel is required, e.g., in:

- immunoelectrophoretic procedures, especially those relying on an immunodiffusion step (immunoprinting and immunofixation);
- the separation of very large molecules with an average hydrodynamic radius above 5–10 nm, such as

antibodies, lipoproteins, some membrane proteins, nucleic acids, and viruses.

Agarose gels, although quite rigid, are much less elastic than polyacrylamide and so are more easily torn and require careful handling.

Polyacrylamide Gel

Favored for producing well-defined, mechanically stable, and chemically inert matrices of variable pore size, polyacrylamide gel (PAG) is probably the most widely used gel medium in electrophoretic food analysis today. The gels are mechanically strong, easy to handle, and completely transparent. Owing to the relatively small pore size, diffusion is limited, so PAG usually provides a better resolution for most proteins (with a molecular weight less than 500 000) than other gel media.

PAG is formed through polymerization of acrylamide monomers with the cross-linking comonomer *N*, *N*′-methylenebisacrylamide, in the presence of free radicals usually provided by the chemical initiator ammonium persulfate or the photochemical initiator riboflavin. The polymerization reaction is regulated through the addition of *N*, *N*, *N*′, *N*′-tetramethylethylenediamine, which provides a source of tertiary amines.

The pore size of the gel can be easily adjusted by controlling both the concentration of acrylamide (% C) and the relative proportion of the cross-linking agent BIS to acrylamide (% T). This results in well-defined molecular sieving properties, providing an additional separation effect dependent on molecular size.

Sample Pretreatments

If sample proteins are insoluble or liable to aggregation or precipitation during an electrophoretic separation, dissociating and unfolding agents such as urea or nonionic detergents (e.g., octylglucoside and Nonidet P-40) are added to the sample or to the gel. Uncharged thiol reducing agents such as β-mercaptoethanol or the odorless compounds dithiothreitol or dithioerythritol are used to protect free thiol groups in proteins against oxidation (loss of enzymatic activity) and disulfide bond formation. Indeed, since these reagents will also cleave disulfide bonds, they can be added together with a dissociating agent such as urea or, preferably, sodium dodecyl sulfate (SDS) to cause the deliberate dissociation of proteins and their aggregates into their constituent subunits. This approach is successfully used in SDS–PAG electrophoresis where samples are run with and without the use of these thiol reducing agents to give an estimate of the degree of protein cross-linking by inter-and/or intramolecular disulfide bonds.

Prior to isoelectric focusing, samples must be desalted, e.g., by ultrafiltration or dialysis.

Gel Electrophoresis

Gel electrophoresis is zone electrophoresis in a chemically inert gel matrix, such as polyacrylamide or agarose. The sample is applied in a small volume as a narrow zone, e.g., in gel slots. As the electric field is applied, each sample component migrates according to its own mobility in a gel medium of constant pH and ionic strength. The separation into 'pure zones' is achieved by maximizing the differential rate of migration, while minimizing zone spreading (dispersion) due to heat convection and diffusion.

At present, there are three different geometrical forms in which gel electrophoresis can be carried out: horizontal slabs, vertical slabs, or vertical cylinders (rods) of gel, often also referred to as tube gels. Slab gels in thinner layers are preferred to thicker slab or tube gels, as they offer faster separation, sharper zones, and more rapid and efficient cooling and subsequent staining. In PAG electrophoresis, there is a clear preference for the horizontal system utilizing ultrathin gel layers polymerized on to carrier foils. The advantages of horizontal versus vertical systems are easier handling, use of premanufactured gels, more efficient cooling, reduced material costs, availability of fully automated systems, and flexibility towards other forms of electrophoresis such as isoelectric focusing.

PAG Electrophoresis

PAG electrophoresis (PAGE) is the most widely used method in the analysis of complex mixtures of proteins. PAGE of proteins can be broadly classified in the following way:

1. Homogeneous systems with only a single separating gel and using continuous buffer media.
2. Multiphasic (discontinuous) systems, where the stacking gel, a nonrestrictive large pore gel, is layered on top of the separating gel, a small-pore gel. Each gel layer is made with different buffers, which may vary in pH, and/or ion mobility, and/or ionic strength (conductivity).

In the stacking gel of a multiphasic system, sample components are separated in the isotachophoresis mode with no molecular sieving effect. After passing the boundary between stacking and separating gel, the sample components are then separated by size and charge in the usual way. The very high resolving power of this method, which is often referred to as

disc PAGE, is due to the formation of very sharp zones produced by the gel and buffer discontinuities. The production of thin starting zones makes disc PAGE very suitable for use with dilute sample solutions.

Gradient PAGE provides a resolution superior to that of a gel of single concentration. Polyacrylamide can be cast in slabs or tubes in which the concentration of acrylamide increases in a continuous manner (e.g., linear or concave) over the length of the gel, thereby producing an increasing sieving effect due to decreasing pore size.

PAGE of food proteins is most efficient in the presence of the anionic detergent SDS. SDS wraps around the polypeptide backbone, breaking up noncovalent protein aggregates (dimers, tetramers, etc.), and cancels differences in intrinsic charge of the proteins. In this way, SDS converts the proteins to rods of negative charges with equal charge densities, i.e., it confers a negative charge to the polypeptide chains in proportion to their length and imparts the same free solution mobility to all proteins, regardless of their identity. Separation of SDS-denatured proteins occurs principally by sieving effects, with lower-molecular-weight proteins migrating more rapidly through the gel. The rate of migration of a protein can be compared to the rate of migration of standard proteins of defined molecular weight, and fairly accurate estimates of protein molecular weight can be obtained. In combination with pore size gradients and discontinuous buffer systems, SDS–PAGE can be an even more powerful and efficient method for protein separation.

Because SDS–PAGE separates molecules according to size, it is a highly favored technique in the preparation of two-dimensional protein maps,a particularly when coupled with a method separating principally according to charge differences in the second dimension.

Isoelectric Focusing

In isoelectric focusing (IEF), amphoteric analytes, such as proteins and macropeptides, are separated according to their isoelectric point (p*I*) values in a stable and linear pH gradient. The gradient increases from a low pH at the anodic end to a high pH at the cathodic end. The samples may be applied at any position, before or after the pH gradient has been formed. When an electric field is applied, the charged sample components move towards the electrode of opposite charge. As each component migrates through the pH gradient and progressively encounters differences in pH, it becomes increasingly less charged. Eventually, it arrives at a pH equal to its p*I*, where its net charge is zero, and its migration stops. Thus, each component is concentrated, or sharply focused, into a narrow zone at its p*I* position, where it remains stationary as long as the electric field is maintained. To counteract the decrease in solubility of proteins approaching pH values close to their p*I*, solubilizing agents such as urea may be included.

The pH gradient is most effectively formed with the use of a complex mixture of synthetic amphoteric molecules (aliphatic polyaminopolycarboxylic acids) known as ampholytes (e.g., ampholines). When an electric field is applied to these ampholytes, they each migrate to their respective p*I*. A relatively stable pH gradient can be formed over the range pH 2–11 and also almost any portion thereof. The development of immobilized pH gradient gels (known as immobilines) has led to a significant improvement. In these gels, amino and carboxy groups (acrylamide derivatives) are covalently attached to the gel-forming polyacrylamide backbone.

IEF is performed in essentially nonsieving media, such as in free solution with a density gradient, PAG of high porosity or in agarose with very low electroendosmosis agarose. Horizontal flat-bed systems with thin or ultrathin layers of PAG are the most suitable. The commercial availability of premanufactured gel slabs has made analytical PAGIEF in both ampholine and immobiline systems even more efficient and accurate. In addition, toxicity problems from handling the neurotoxic acrylamide monomers are avoided. IEF in agarose gels is particularly appropriate for large molecules or when a second-dimension immunodiffusion or immunoelectrophoresis stage is also to be used.

IEF is a powerful technique that can separate proteins differing in p*I* by as little as 0·001. IEF is also a very simple and accurate method for determining the p*I* of a protein.

Isotachophoresis

In isotachophoresis (ITP), sometimes also referred to as steady-state stacking or displacement electrophoresis, separation results from the different mobilities of identically charged ions in a discontinuous electrolyte system with two different solutions, a leading and a terminating electrolyte. The sample is inserted between these two electrolyte solutions. In a specific separation, either cations or anions can be determined, but not both at once. If sample anions are being determined, the leading electrolyte must contain an anion of higher mobility (e.g., chloride) than any of the sample anions of interest, whereas the terminating electrolyte must contain an anion of lower mobility (e.g., glycine) than any of the sample anions of interest. When the electric field is applied, all negatively charged ions begin to move towards the anode, arranging themselves in order of mobility.

Separation of the sample ions takes place between the leading and terminating electrolytes during migration. When the system has reached equilibrium, each individual ionic sample component moves separately as a 'pure' band. Each pure band is stacked between the sample component of next highest and lowest mobility. Thus, in the 'anion-train,' consecutive sample analyte zones form behind the leading ion zone (chloride), which acts as the 'locomotive,' and in front of the terminating ion zone (glycine). The sample zones travel in order of decreasing mobility at the same velocity as the leading electrolyte zone. These differences in mobility cause a stepwise change in the electric field strength from one analyte zone to the next, resulting in an important concentrating effect. Backward or forward diffusion into a preceding or following band is not possible. As a consequence, the concentrating or stacking effect at the boundary between bands produces sharp zones and prevents diffusional broadening.

Analytical-scale ITP is routinely performed in free solution in small-bore capillary tubes and with the use of a constant current at a well-defined pH value, and with on-line detection. The resolved bands emerge as rectangular peaks, not as the common Gaussian-shaped curves in chromatographic separations. Bands can be identified by measuring the electric field strength within the band, by placing two microelectrodes side by side in the capillary channel. Quantitative information is contained in the length of the band. In ITP, electrical detection is most commonly used, although thermal and ultraviolet absorption detection are also applied.

Analysis of Separated Zones

The most important techniques for analysis of electrophoretically separated proteins are summarized in Table 2. Other techniques such as autoradiography and fluorography of radiolabeled analytes so far have not played a significant role in food analysis.

Staining

The most common procedure for analysing proteins separated in gels is staining with a variety of dyes that bind to proteins. Usually, the staining procedure involves the use of a fixative, such as trichloroacetic acid, to precipitate proteins and prevent their diffusion out of the gel. Gels are then soaked in a staining solution, such that the entire gel is uniformly stained. Nonprotein-bound dye must be removed in a destaining rinse. The length of time required for both the staining and destaining steps primarily depends on the thickness of the gel, with the use of ultrathin gel slabs saving a considerable amount of time.

Proteins in gels are most frequently stained with Coomassie Blue or Amido Black dyes, or with photographic amplification systems using silver. The detection limits for Coomassie Blue staining are roughly 100 ng of protein in a band, whereas silver staining is about 100 times more sensitive. Once the gel is stained, it can be photographed or scanned by densitometry for a record of the position and intensity of each band. In general, the accuracy and precision obtainable from this quantitation procedure are not on a par with what is routinely obtained by high-performance liquid chromatography. (*See* **Chromatography**: High-performance Liquid Chromatography.)

Specific staining techniques exist for phosphoproteins, lipoproteins, glycoproteins, and nucleic acids. Enzymes can be localized by assaying for their enzymatic activity, e.g., by converting substrates to sparingly soluble products, which can then be coupled chemically to azo dyes. Agarose gels and cellulose acetate strips are very suitable for direct localization of immunoreagents due to their pore size. Immunofixation and immunoprinting are popular techniques where antibodies and antigens, respectively, migrate towards each other by diffusion (immunodiffusion). Subsequent staining after the removal of unbound antigens and antibodies allows selective detection of specific protein fractions.

Blotting

Blotting refers to the transfer of separated zones of macromolecules, e.g., proteins or nucleic acids, to thin sheets of derivatized paper or adsorptive membrane matrix, such as nitrocellulose or polyvinyldifluoride (PVDF) to which they bind. Immobilized on the surface of a thin matrix, macromolecules are more easily and uniformly accessible to detection reagents and react more sensitively and faster. The common transfer methods are based on simple diffusion (diffusion blotting), solvent flow with/without vacuum (capillary or Southern blotting), or electrophoretic elution (electroblotting). The detection of transferred protein components on blots is carried out with all the procedures described above, specific assays for enzymes included.

Immunoblotting/Enzyme Labeling

Immunological detection of proteins is suitable with all types of immobilizing transfer matrix. After proteins or nucleic acids are transferred, all additional matrix binding sites must be blocked with an excess of nonspecific protein. Then, an antibody specific for a certain class or type of protein is bound and, finally, a second antibody directed against the first antibody. This second antibody can be fluorescence-labeled, radiolabelled, or conjugated to an enzyme for

location by ultraviolet illumination, autoradiography, or enzyme activity, respectively. Almost any enzyme for which a suitable assay method exists could be coupled in theory to an antibody by cross-linking reagents (e.g., carbodiimide, glutaraldehyde, etc.), but alkaline phosphatase and horseradish peroxidase, for which there are several simple and sensitive detection methods, are very popular. The use of these enzyme-labeled antibodies in combination with blotting results in extremely sensitive procedures, which can clearly detect as little as 100 pg of protein on a membrane. (*See* **Enzymes**: Uses in Analysis.)

Immunoelectrophoretic Techniques

Immunoelectrophoresis consists of a combination of an electrophoretic step with the subsequent precipitation of antigen–antibody complexes (immunoprecipitates). The most common procedures in food analysis rely upon the migration of antigenic proteins through or into an antibody-containing gel. Buffers and pH values are usually chosen so that only the antigens migrate, and the antibodies either do not move at all or, at most, migrate only very slowly and thus remain evenly distributed throughout the gel during the whole electrophoresis. The most popular techniques are:

1. Counterimmunoelectrophoresis, sometimes referred to as countercurrent or crossed-over immunoelectrophoresis: in agarose gel of high electroendosmotic flow, antigens electrophoretically move towards uncharged antibodies, which are carried by the electroendosmotic flow in countercurrent, and precipitin arcs are formed.
2. Grabar/Williams method: normal zone electrophoresis in agarose gel is followed by an immunodiffusion stage. Antibody diffusion from troughs in the gel cut parallel to the line of electrophoretically separated components and precipitin arcs are formed.
3. Laurell rocket techniques: antigen proteins are transported electrophoretically through a gel slab containing antibodies at a pH value at which the antibodies remain essentially immobile, resulting in the formation of 'rocket'-shaped precipitate peaks whose height and area are linearly related to the antigen concentration. These techniques allow quantitative analysis of antigens, but are not applicable to complex mixtures.
4. Crossed immunoelectrophoresis. Agarose zone electrophoresis or isoelectric focusing in one dimension is followed by electrophoresis into an antibody-containing gel slab in the second dimension, with the formation of mountain-like precipitation peaks. Qualitative and/or quantitative analysis of peak area is possible.

For many purposes, the formation of an opaque precipitation are in a transparent gel will be relatively clear and perfectly adequate. If desired, the sensitivity of detection can be greatly enhanced by the application of the usual protein-staining techniques (see above) after the removal of nonprecipitated protein (uncomplexed antigens and antibodies). An alternative to staining is the use of radiolabeled, fluorescent, or enzyme-tagged reactants (see Immunoblotting/Enzyme Labeling).

The use of immunoelectrophoresis in food analysis is limited by the availability of specific antibodies, so that cross-reactions with other nontargeted proteins are avoided. (*See* **Immunoassays**: Principles.)

Capillary Zone Electrophoresis

Although conventional gel electrophoresis is extremely powerful and versatile, it is often a slow, labor-intensive technique. Analyses often require several hours, and often several days to complete. But in the last 10 years, rapid instrumental methods have been developed for free solution electrophoresis in capillary tubes with on-line detection. Existing new techniques currently available to the analytical chemist are capillary electrophoresis (CE) in general, and capillary zone electrophoresis (CZE) in particular.

CZE is performed as free solution electrophoresis in capillaries with inside diameters of only 20–200 μm. These small dimensions aid heat dissipation and minimize zone broadening caused by molecular diffusion and convection currents in the electrophoretic medium. As a consequence, relatively high voltages can be used, which allow highly efficient separations within 20 min. Each end of the capillary, filled with electrolyte, is immersed in a separate reservoir containing the same electrolyte and a high-voltage electrode.

CZE is strongly influenced by the degree of inertness of the capillary, especially towards polar interactions with buffer ions and/or analytes. Separations by CZE in inert capillaries (e.g., Teflon) are based only on the different mobilities of charged molecules in an electric field, as in conventional zone electrophoresis. In contrast, CZE in noninert capillaries (e.g., fused silica without a protective coating) uses the separating forces of both electrophoretic migration and electroendosmotic flow.

In these cases, cations in the electrolyte (predominantly protons) associate loosely with the capillary wall, blocking the negative charges of the silanol groups of uncoated fused silica. The hydrated cations are pulled towards the cathode, inducing a bulk fluid flow (the electroendosmotic flow, EEF) in that direction. The EEF can be strong enough to move all species (positive, negative and neutral) in the same direction,

though with different velocities. The EEF may ruin a separation (e.g., by adsorption of the analyte to the wall), but it can also be a crucial feature of the method.

State-of-the-art fully automated CZE instruments introduce samples automatically into the capillaries either by electromigration or by hydrostatic or pneumatic methods. After zone electrophoresis in capillaries, separated sample components are detected in real time as they pass an on-line detector. Ultraviolet absorption and fluorescence detection have proven most useful so far for CZE.

CZE is faster and easier than slab gel electrophoresis. Although CZE is still in its infancy, powerful separation methods have been already developed for amino acids, peptides, nucleic acids, and their fragments. In view of the increasing importance of biotechnology in food manufacturing, CZE is expected to play a major role in analytical food chemistry. (*See* **Amino Acids**: Determination; **Peptides**.)

Application to Food Analysis/ Authentication

Proteins and peptides are the primary targets of electrophoresis in food analysis. Electrophoretic techniques serve to characterize molecular architecture, assess homogeneity, and identify and quantify proteins.

Electrophoresis has allowed food chemists to characterize increases in protein yield, protein–protein interactions, and enzymatic protein degradation. Further studies of compositional parameters and processing treatments have highlighted the importance of these features on the flavor and texture of protein-based foods. This is particularly true in the field of dairy science and technology where the electrophoretic separation and quantification of milk proteins and peptides have always helped new developments. Figure 1 illustrates the simultaneous determination of the major milk proteins by IEF. This efficient procedure can be successfully applied to follow changes in protein distribution in cheeses manufactured by accelerated ripening and/or ultrafiltration procedures.

For studying proteins and larger peptides in complex food systems, the most popular electrophoretic techniques are isoelectric focusing and SDS (disc) electrophoresis in ultrathin, horizontal PAG slabs. Subtle differences in molecular properties such as those due to posttranslational modifications or genetic variation within protein classes can be determined with these methods. Perhaps the ultimate in protein separation is obtained when a two-dimensional electrophoretic approach is used. Truly spectacular resolution can be achieved when one-dimensional isoelectric focusing is combined with two-dimensional SDS gradient PAGE, immunodiffusion, or immunoelectrophoresis. These high-resolution techniques are used to follow proteolytic breakdown and to verify the intact primary structure of food proteins after certain processing treatments.

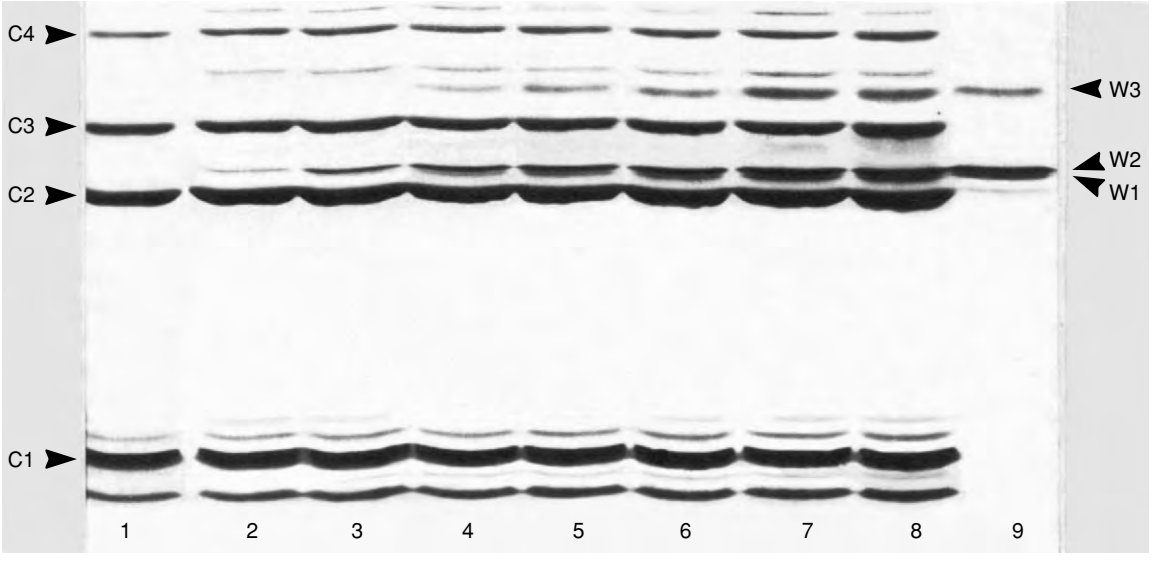

Figure 1 Simultaneous determination of the major milk proteins by isoelectric focusing in ultrathin-layer PAG. C1, α_{S1}-casein B; C2, β-casein A^2; C3, β-casein A^1; C4, α_{S2}-casein; W1, α-lactalbumine; W2, β-lactoglobuline A; W3, β-lactoglobuline B. Samples from left: recombined milk powders with casein: whey protein ratio of 100:0 (1), 98:2 (2), 95:5 (3), 90:10 (4), 80:20 (5), 70:30 (6), 60:40 (7), 50:50 (8), and 0:100 (9). Reproduced from Electrophoresis, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

It is not surprising that the discriminating power and versatility of modern electrophoretic techniques are unsurpassed in the positive identification of specific proteins and peptides. Thus, techniques such as IEF, with or without immunodetection, provide reliable methods in food authentication in order to prevent fraud and adulteration. Examples of current interest include the verification/identification of the species origin of meat, fish, milk and cheese products, the detection of nonmeat and nonmilk proteins in meat and milk products, respectively, the identification of cereal varieties, the classification of pasta into durum wheat and soft wheat types, the characterization of cultivars of crops, fruits and vegetables, and the determination of enzymes in beer and cheese. (*See* **Adulteration of Foods**: Detection.)

Isotachophoresis is the method of choice for the analysis of complex mixtures of organic acids in foods without derivatization but is also applicable to artificial sweeteners and preservatives.

See also: **Adulteration of Foods**: Detection; **Amino Acids**: Determination; **Chromatography**: High-performance Liquid Chromatography; **Enzymes**: Uses in Analysis; **Immunoassays**: Principles; **Peptides**; **Protein**: Determination and Characterization

Further Reading

Andrews AT (1986) *Electrophoresis – Theory, Techniques, and Biochemical and Clinical Applications*, 2nd edn. Oxford: Clarendon Press.

Chrambach A (1985) *The Practice of Quantitative Gel Electrophoresis*. New York: VCH.

Jorgenson JW (1986) Electrophoresis. *Analytical Chemistry* 85: 743A–760A.

Jorgenson JW and Philips M (eds) (1987) *New Directions in Electrophoretic Methods*. Washington, DC: American Chemical Society.

Westermeier R (1990) *Elektrophorese-Praktikum*. Weinheim: VCH.

Elimination Diets *See* **Food Intolerance**: Types; Food Allergies; Milk Allergy; Lactose Intolerance; Elimination Diets

EMERGING FOODBORNE ENTERIC PATHOGENS

D W K Acheson, University of Maryland School of Medicine, Baltimore, MD, USA

Background

Over the last 25 years, there have been an ever-growing number of enteric pathogens linked with foodborne disease (**Table 1**). Following some major foodborne disease outbreaks in the early 1990s, there has been increased focus on understanding the causes and pathogenesis of foodborne disease. As part of this initiative, federal agencies developed improved surveillance systems and invested in gaining a better understanding of how to prevent foodborne disease through safer food processing. During the course of this chapter, the initial section will describe the current major emerging foodborne infections in the context of our new epidemiological knowledge. This is then followed by a more detailed description of some of the principal emerging foodborne infections that are considered to be important. Clearly, different parts of the world have to deal with different enteric pathogens, but the focus of this chapter will be on issues within the USA and other developed countries. It is not the goal of this text to specifically address emerging foodborne pathogens in developing countries that not only have to deal with all the problems seen in developed nations, but also frequently have to deal with other specific pathogens.

Current Trends of Emerging Foodborne Pathogens

In 1996, the Foodborne Diseases Active Surveillance Network (FoodNet) was established. This is a

Table 1 Emerging foodborne pathogens since 1977

Campylobacter jejuni
Cryptosporidium parvum
Cyclospora cayentanesis
Shiga toxin-producing *E. coli*
Listeria monocytogenes
Norwalk-like viruses
Salmonella enteritidis
Salmonella typhimurium DT104
Transmissible spongiform encephalopathy
Vibrio cholerae O139
Vibrio vulnificus
Vibrio parahemolyticus
Yersinia enterocolitica

Table 2 Foodborne pathogens routinely under surveillance as part of FoodNet

Bacteria
Campylobacter spp.
Escherichia coli O157:H7
Listeria monocytogenes
Salmonella spp.
Shigella spp.
Vibrios
Yersinia
Protozoa
Cryptosporidium
Cyclospora

cooperative venture between the Centers for Disease Control, state health departments, the Food Safety and Inspection Service of the US Department of Agriculture, and the Food and Drug Administration. FoodNet is a sentinel network that is designed to produce national estimates of the burden and sources of specific foodborne illnesses. FoodNet currently focuses on obtaining information on nine important and emerging foodborne infection agents (Table 2). FoodNet operates by conducting population-based active surveillance for confirmed cases of the various pathogens shown in Table 2. Since the initiation of FoodNet in 1996, the population under surveillance has grown from five sites and some 14 million persons in 1996 to 29.5 million that were under surveillance during the year 2000. This number now represents over 10% of the US population and involves a number of states including Minnesota, Oregon, California, Connecticut, Georgia, Maryland, New York, and Tennessee.

During the year 2000, there were 12 631 laboratory-confirmed cases of the nine diseases listed in Table 2 at the FoodNet sites. Of these 12 631, 4640 were campylobacteriosis, 4237 were salmonellosis, 2324 were shigellosis, 631 were *E. coli* O157:H7, 484 were cryptosporidiosis, 131 yersiniosis, 101 listeriosis, 61 Vibrios and 22 cyclosporiasis. Since 1996, the levels of the major pathogens have varied amongst the five original sites, and these data are shown in Table 3. As can be seen from Table 3, there have been minor fluctuations but little major change in the nine important pathogens, although for *Campylobacter* and *Salmonella*, the general trend is downward.

One of the key criteria in relation to determining the importance of emerging (and emerged) foodborne pathogens relates to the severity and frequency of disease with which they are associated. In the past, there was a great deal of uncertainty about the morbidity and mortality associated with foodborne disease, but in 1999, a study was published by the Centers for Disease Control that specifically

Table 3 Incidence (per 100 000 population) of diagnosed infections in the original five FoodNet sites

Pathogen	*1996*	*1997*	*1998*	*1999*	*2000*
Campylobacter	23.5	25.2	21.4	17.5	20.1
Cryptosporidium	nr	3.7	2.9	1.8	2.4
Cyclospora	nr	0.4	0.1	0.1	0.1
Escherichia coli O157	2.7	2.3	2.8	2.1	2.9
Listeria	0.5	0.5	0.6	0.5	0.4
Salmonella	14.5	13.6	12.3	13.6	12.0
Shigella	8.9	7.5	8.5	5.0	11.6
Vibrio	0.2	0.3	0.3	0.2	0.3
Yersinia	1.0	0.9	1.0	0.8	0.5

nr, not reported.
Modified from Preliminary FoodNet Data on the Incidence of Foodborne Illness – Selected Sites, United States, 2000 (2001) *MMWR* 50: 241–246.

addressed these issues using FoodNet and other data. The overall conclusion from this study was that there was an estimated 76 million foodborne illnesses per year, resulting in 325 000 hospitalizations and 5000 deaths. An interesting outcome of this work was that known pathogens only accounted for 14 million (18%) of these cases. This obviously begs the question of what is going on in the remaining 80% of cases. Part of the answer lies in our inability to diagnose foodborne disease in all cases. This is due to a mixture of the failure of infected patients to seek medical attention, the failure of physicians to order stool cultures, and the failure of the laboratories to find foodborne pathogens, even when they are present, the latter either due to funding or technological limitations. However, it does also raise the issue of new emerging infectious foodborne agents that either we are unaware of or have not yet been linked with foodborne disease.

Whereas there are over 200 known diseases transmitted via food, including microbes, toxins, chemicals, physical hazards, and prions, there is a relatively short list of the important agents, and many of these could be considered as emerging simply because they were not recognized until recently. Mead *et al*. determined

Table 4 Ranking of the most common microbial causes of foodborne disease with an estimate of the percentage of infections that are transmitted via food, and deaths as a percentage of total foodborne disease-related deaths

Disease or agent	*Estimated total cases*	*Percentage foodborne transmission*	*Percentage of deaths*[a]
Norwalk-like viruses	23 000 000	40	7
Campylobacter spp.	2 500 000	80	5
Giardia lamblia	2 000 000	10	0.1
Salmonella	1 400 000	95	31
Shigella	450 000	20	0.8
Cryptosporidium parvum	300 000	10	0.4
Clostridium spp.	250 000	100	0.4
Toxoplasma gondii	225 000	50	21
Staphylococcal food poisoning	185 000	100	0.1
Shiga toxin-producing *E. coli*	110 000	85	1.4
Yersinia enterocolitica	100 000	90	0.1
Bacillus cereus	27 000	100	0
Cyclospora cayetanensis	16 000	90	0
Listeria monocytogenes	2 500	99	28

[a]Expressed as a percentage of the number of deaths related to foodborne pathogens.
Modified from Mead PS, Slutskev L, Dietz V *et al*. (1999) Food-related illness and death in the United States. *Emerging Infectious Diseases* 5: 607–625.

which are the important foodborne pathogens in terms of frequency of disease, and at the same time determined what proportion of illness due to those pathogens were actually transmitted via food, as opposed to some other means of transmission such as person-to-person or directly from animals. A summary of this information is presented in Table 4 and illustrates the point that some of the most common foodborne infections such as *Campylobacter* and Norwalk-like virus result in a relatively small number of hospitalizations and death compared with other pathogens such as *Listeria monocytogenes*. Another key observation was that five pathogens accounted for more than 90% of the deaths as follows: *Salmonella* – 31%; *Listeria* – 28%; *Toxoplasma* – 21%; Norwalk-like virus – 7%; *Campylobacter* – 5%; *E. coli* O157:H7 – 3%.

Generally, there is a perception amongst many that foodborne disease usually occurs as part of outbreaks (defined as an incident in which two or more persons experience an illness resulting from the ingestion of a common food). In fact, the vast majority of foodborne disease are isolated cases that are not linked to other cases. A recently published summary of documented outbreaks, which occurred between 1993 and 1997, illustrates the important emerging pathogens and foodborne agents that are most frequently linked to outbreaks. During that period, there were 2751 outbreaks investigated by the Centers for Disease Control averaging about 500 outbreaks per year. A total of 86 058 people were sick in connection with the outbreaks, and while only 32% had a known etiology, they accounted for 59% of the 86 058 infections. Table 5 summarizes some of the outbreak data and illustrates the importance of *Salmonella* and *E. coli* as causes of outbreaks and death, but also demonstrates the importance of two preformed toxins, Scrombotoxin and Ciguatoxin, as causes of outbreaks. As already stated, outbreaks account for the minority of foodborne disease. Based on previous work and assuming 76 million cases of foodborne disease per year, during the 5 years from 1993 to 1997, there would have been 380 million episodes of foodborne disease. During this time, only 86 058 cases were linked with outbreaks.

In summary, since 1996 and the inception of FoodNet, we have learnt a great deal about the foodborne agents that are responsible for disease and death in the USA. Many of these were unrecognized 20 or 30 years ago and underscore the need to remain vigilant about new agents and newly emerging pathogens.

Specific Emerging Foodborne Pathogens

In the following sections, a more detailed description of some of the specific emerging foodborne pathogens is given. It is beyond the scope of this text to discuss in detail all the emerging foodborne enteric pathogens, but readers are referred to other specific chapters in this text that deal in more detail with individual microbes.

Campylobacter

Foodborne disease due to *Campylobacter*, which was not recognized until the mid-1970s, is now one of the most frequently diagnosed causes of food-related gastroenteritis. In fact, *Campylobacter* and *Salmonella* probably account for 70% of diagnosed cases of bacterial foodborne disease in the USA.

Table 5 Major causes of foodborne disease outbreaks in the USA between 1993 and 1997, showing the numbers of outbreaks, the numbers of total cases, and the number of deaths associated with a particular etiology

Etiology	*Number of outbreaks (%)*	*Number of cases (%)*	*Number of deaths (%)*
Salmonella	357 (13.0)	32 610 (37.6)	13 (44.8)
Escherichia coli	84 (3.1)	3 260 (3.8)	8 (27.6)
Scrombotoxin	69 (2.5)	297 (0.3)	0
Ciguatoxin	60 (2.2)	205 (0.2)	0
Clostriduim perfringens	57 (2.1)	2 772 (3.2)	0
Shigella	43 (1.6)	1 555 (1.8)	0
Staphylococcus aureus	42 (1.5)	1 413 (1.6)	1 (3.4)
Campylobacter	25 (0.9)	539 (0.6)	1 (3.4)
Hepatitis A	23 (0.8)	729 (0.8)	0
Bacillus cereus	14 (0.5)	691 (0.8)	0
Clostridium botulinum	13 (0.5)	56 (0.1)	1 (3.4)
Norwalk virus	9 (0.3)	1 233 (1.4)	0
Listeria monocytogenes	3 (0.1)	100 (0.1)	2 (6.9)

Modified from Centers for Disease Control (2000) Surveillance for foodborne disease outbreaks – United States, 1993–1997. *MMWR* 49: 1–51.

There are two main species of *Campylobacter* that are responsible for most of the illness seen in humans. *C. jejuni* accounts for the vast majority ($\sim 90\%$), and *C. coli* accounts for the majority of the remainder. Other *Campylobacter* species that have been associated with gastroenteritis in humans include *C. fetus*, *C. upsaliensis*, *C. hyointestinalis*, and *C. lari*. *Campylobacter* species are fragile organisms and tend to die in transport media, so rapid plating of samples is an advantage. Enriched media are usually used that contain a variety of antibiotics to reduce competitive flora. The infectious dose of *Campylobacter* seems to be variable depending on the strain and the host, but low numbers of *Campylobacter* may be all that is needed to cause disease, and in some instances, in human volunteers, less than 100 organisms can cause disease. These organisms are more frequently associated with sporadic disease rather than outbreaks, and person-to-person spread does not appear to be common, although it can occur in a family environment, for example. This sporadic epidemiology suggests that the infectious dose may actually be higher in many situations. *C. jejuni* and *C. coli* are commensals in the intestines of many animals and birds, including domestic pets. The main vehicles for infection are raw meats, especially poultry, milk, and water. Surface water is frequently contaminated with campylobacters, and waterborne outbreaks have occurred.

Campylobacter species are invasive organisms, which means that they have the capacity to invade the epithelial cells lining the intestine. The pathogenesis of *Campylobacter* is dependent on its motility. *In vitro*, nonmotile strains are not capable of invading intestinal epithelial cells. It is assumed that invasion in the intestine is one of the principal pathogenic mechanisms of these bacteria, although the genes involved are not known. The recent publication of the *C. jejuni* genome is, however, helping with the discovery of new genes and determining the function of previously identified genes.

Following infection with campylobacters, there is typically an inflammatory response with marked inflammatory infiltration of the lamina propria resulting in leukocytes being present in the stools. Clinically, symptoms usually occur within 2–3 days following exposure, or up to 7 days after exposure. Nongastrointestinal symptoms such as fever (which may be high), headache, and myalgia may precede the onset of nausea, vomiting, and diarrhea. Diarrhea is usually the predominant symptom. This may be watery or bloody, and is usually associated with severe cramping abdominal pain. Interestingly, the disease is occasionally biphasic with an apparent settling of symptoms after 4–5 days, only to be followed by a recrudescence. A number of complications are associated with *Campylobacter*, including cholecystitis, hepatitis, acute appendicitis, pancreatitis, and focal extraintestinal infections. Longer-term complications include reactive arthritis, Reiter's syndrome, uveitis and Guillain–Barré syndrome; molecular mimicry may be involved in some of these complications. The majority of *Campylobacter* infections are self-limiting and require supportive therapy only, especially in the form of rehydration. Antibiotic therapy is not routinely required, but patients with prolonged or severe symptoms or who have other significant risk factors (AIDS, cirrhosis, diabetes, etc.) should be treated. Erythromycin is a reliable therapy, although there is a general move toward using quinolone antibiotics. Despite the use of erythromycin for many years, the resistance levels have remained low. This is not the case with the quinolones, and there are increasing reports of ciprofloxacin

resistance, which may become more of a problem in view of the expanded use of these antibiotics in agriculture.

Listeria monocytogenes

L. monocytogenes is a Gram-positive motile rod that is one of the most frightening foodborne pathogens because of the high mortality rate associated with infection. Of the seven *Listeria* species, only *L. monocytogenes* is pathogenic for humans. *L. monocytogenes* is a common environmental organism and is frequently present in soil and water, on plants, and in the intestinal tracts of many animals. It has been found in 37 different types of mammals and at least 17 species of birds. Between 1 and 10% of people are carriers of *L. monocytogenes*. This organism is associated with both sporadic disease and outbreaks. Incriminated foods include milk, cheese, raw vegetables, undercooked meat, and foods prepared for instant use such as hot dogs. The infectious dose is not really known: some studies have suggested that it may be very high (up to 10^9 organisms), and others have suggested that it may be as low as several hundred. In practice, the most critical aspect is probably individual susceptibility rather than the infectious dose *per se*. Clinically, it usually begins with nonspecific symptoms such as fever, myalgia and gastrointestinal upset in the form of diarrhea, and nausea. What makes *L. monocytogenes* exceptional as a foodborne pathogen is its very high mortality rate. Of the approximately 1800 cases per year that are estimated to occur in the USA, there are over 400 deaths. This gives a case fatality rate of over 20%. There are certain groups that are especially at risk of developing listeriosis, and these include pregnant women, the elderly, and the immunocompromised. Transplacental transmission in pregnant women is a major concern, although it does not inevitably lead to major consequences. Spontaneous abortion, prematurity, neonatal sepsis, and meningitis are all complications of transplacental transmission. Although *L. monocytogenes* is readily killed by heat and cooking, the fact that it is so ubiquitous makes recontamination a real risk. This then poses a major health problem, because the organism will grow and multiply at standard refrigerator temperatures. Thus, even minor contamination of a product may, after storage, result in high levels of bacteria, even if the product has been adequately refrigerated. Following diagnosis, *L. monocytogenes* is readily treated by penicillins or aminoglycosides. *L. monocytogenes* has also been linked with acute gastroenteritis in individuals with normal immune systems, but it is unclear how frequently this occurs.

Salmonella spp.

Salmonella are one of the most common causes of foodborne illness in humans, and although they are not as a group considered to be emerging foodborne pathogens, some types of *Salmonella* (e.g., *S. enteritidis* in eggs) have definitely emerged as a major problem in recent years. There are many types of *Salmonella*, but they can be divided into two broad categories: those that cause typhoid and those that do not. The typhoidal *Salmonella* such as *S. typhi* and *S. paratyphi* only colonize humans and are usually acquired by the consumption of food or water contaminated with human fecal material. The much broader group of nontyphoidal *Salmonella* are found in the intestines of other mammals and therefore are acquired from the consumption of food or water that has been contaminated with fecal material from a wide variety of animals and poultry.

The number of cases of nontyphoidal *Salmonella* has increased steadily over the last four decades. *S. entertidis* particularly has become a growing problem, especially in hen eggs. In the year 2000 FoodNet data, of the 4237 *Salmonella* that were isolated, 3686 were serotyped, and 23% were *S. typhimurium*, and 15% were *S. enteritidis*. Currently, the estimate is that around 1 in 10 000 eggs are contaminated with *Salmonella*. It is now known that *Salmonella* can penetrate intact eggs lying in fecally contaminated material and also infect eggs transovarially during egg development before the shell is formed. Other than eggs, common sources of nontyphoidal salmonellosis are milk, foods containing raw eggs, meat and poultry, and fresh produce. Essentially, as with many of the other foodborne bacterial infections, *Salmonella* are frequently transmitted through fecal contamination of food because of the large numbers of animals that carry the organism. The infective dose of nontyphoidal *Salmonella* may vary from < 100 to 10^6, depending on the host and on the actual type of *Salmonella*. Irrespective of the type of *Salmonella*, the most critical virulence determinant of these bacteria is their ability to invade the intestinal epithelium, following which they interact with underlying lymphoid tissue.

Clinically, nontyphoidal *Salmonella* typically presents with gastroenteritis 24–48 hours after exposure to the organisms. There is usually nausea, vomiting, abdominal cramps, and diarrhea, which may be watery or occasionally bloody. It is not unusual for there to be associated symptoms of fever, chills, headache, and myalgia. Occasionally there can be long-term consequences following *Salmonella* infection such as reactive arthritis (especially in individuals who are HLA B27-positive), endocarditis, and localized

infections such as osteomyelitis, septic arthritis, and soft tissue infections. Diagnosing infection with *Salmonella* is dependent on culturing the organism, usually from either stool or blood cultures. In the case of nontyphoidal *Salmonella*, it is also worth trying to culture organisms from the incriminated food.

Gastroenteritis from nontyphoidal *Salmonella* is usually self-limiting, and rehydration is the most critical aspect of treatment. Antibiotic therapy is not routinely required for this aspect of *Salmonella* infection and has in some instances been thought to promote chronic carriage. When there is systemic invasion with the bacteria and in cases of enteric fever, antibiotic therapy is important. Third-generation cephalosporins and quinolones are most frequently used, although chloramphenicol has been the mainstay of treatment for typhoid fever for many years and is still used in many developing countries, but chloramphenicol does carry a risk of complications.

Antibiotic resistance has become a major problem with many *Salmonella* serovars. Of particular concern is the recent emergence and spread of *Salmonella* DT104 that carries resistance to multiple antibiotics, including, in some instances, to the fluoroquinolones such as ciprofloxacin. In 2000, *Salmonella* strains that were resistant to Ceftriaxone were also reported, further raising the concern that the emergence of new antibiotic resistance profiles were occurring.

Shiga Toxin-producing *E. coli*

Shiga toxin-producing *E. coli* (STEC) are relative newcomers to the scene of foodborne pathogens. The first STEC to be associated with disease in humans was *E. coli* O157:H7 following two outbreaks of hemorrhagic colitis in 1982. Since then, it has been learned that there are in fact many different serotypes of STEC, and at least 60 different types have been associated with clinical disease. Recent studies have suggested that around 1% of samples submitted to clinical microbiology laboratories in the USA contain STEC, of which around two-thirds are O157:H7, the remainder being non O157. STEC are present in the gastrointestinal tracts of many mammalian species but appear to be especially common in ruminants (cattle, sheep, deer, and goats). Therefore, the main source of STEC in our food supply is bovine products.

Recently, there have been an increasing number of reports associating STEC infection with fresh produce (lettuce, alfalfa sprouts, apple cider) and water. This is thought to be mainly due to contamination with fecal material from cattle pasture. Clinically, STEC cause a variety of diseases ranging from diarrhea, which may or may not be bloody, hemorrhagic colitis, and the hemolytic uremic syndrome (HUS). HUS is a triad of renal failure, thrombocytopenia, and hemolytic anemia. Acutely, HUS has a mortality rate of around 5%, and up to 50% of HUS patients may have some degree of permanent renal insufficiency. The main virulence factor from STEC is the production of one or more bacteriophage-encoded Shiga toxins (Stx), which are of two main types, Stx1 and Stx2. Following ingestion of the bacteria, they colonize portions of the lower intestinal tract and produce the toxins. Stx is then thought to cross the intestinal epithelial cell barrier and damage distant target sites, especially the kidney and brain, by a direct effect on endothelial cells in the microvasculature. The infectious dose of STEC may be very low, in the region of 10–100 organisms in some instances. Symptoms typically develop 2–4 days following ingestion but may occur in as little as 1 day or in as much as 8 days. The diarrhea can be of variable type (bloody or nonbloody) and may contain leukocytes. The type of diarrhea is not a reliable indication of who will go on to develop HUS. The mainstay of treatment for STEC and its major complications is supportive. There is a degree of controversy over the use of antibiotics, and a number of studies have suggested that certain antimicrobials (e.g., trimethoprim-sulfamethoxazole) actually increase the likelihood that a patient will go on to develop serious complications.

Yersinia spp.

Of the three members of the genus *Yersinia*, *Y. enterocolitica* and *Y. pseudotuberculosis* are considered to be foodborne, whereas *Y. pestis* is not. *Yersinia* are not very commonly found as causes of foodborne illness compared with *Salmonella* or *Campylobacter*. However, they are clearly transmitted in food and can cause a significant gastrointestinal illness. The food most frequently associated with yersiniosis is pork. Swine are a major reservoir of these organisms, and although they have been found in many other animals (e.g., sheep, dogs, cats, and cattle), consumption of undercooked pork is a common association. Milk is another frequently reported source, and since *Y. enterocolitica* can survive, and indeed multiply, in milk at 4 °C, small numbers of organisms can become a significant health threat, even if the milk is refrigerated. Symptoms in *Y. enterocolitica* infection can be prolonged, lasting several weeks or even longer. Most infections are, however, self-limiting, although complications may occur such as ulceration and intestinal perforation. A classic long-term complication following yersiniosis is the development of reactive

arthritis. As with other enteric pathogens, this is more likely in patients who are HLA B27-positive.

Viral Foodborne Pathogens

Viral agents are considered to be an increasingly important cause of foodborne illness. A number of different viral agents have been associated with foodborne disease and cause a variety of illnesses varying from a simple gastroenteritis to major systemic upset such as hepatitis. Food and water are vehicles for viruses, but viruses do not reproduce in food, and nor do they produce toxins in food. Some viruses, such as Norwalk, cause large outbreaks, and others seem to be more frequently associated with sporadic disease. Overall, the difficulty in diagnosing viral illness has precluded the development of large amounts of epidemiological data. It is beyond the scope of this chapter to discuss all viral causes of foodborne disease. However, Norwalk-like viruses do deserve a specific mention.

Norwalk virus is a small round-structured virus (SRSV) and was the first virus to be clearly associated with gastroenteritis. Norwalk is a calicivirus, and this group of viruses causes disease worldwide and has been associated with some large outbreaks, often in confined environments such as cruise ships. Outbreaks have been associated with contaminated drinking water, swimming water, consumption of undercooked shellfish, ice, and salads. As with the other enteric viruses, fecal contamination of food or water is usually found to be the ultimate source. The incubation time following exposure is around 48 h, and the clinical illness usually consists of vomiting and diarrhea. The diarrhea is watery without red cells, leukocytes or mucus. The disease is usually self-limiting, settles in 24 h, and requires no specific therapy. Specific diagnosis is difficult. A number of assays are available (but not commercially), including electron microscopy, enzyme immunoassays, and reverse transcriptase–polymerase chain reaction.

Parasites

Two emerging parasitic foodbrone pathogens that cause predominantly watery diarrhea are *Cryptosporidium parvum* and *Cyclospora cayetanensis*. As can be seen from Table 4, *C. parvum* causes a lot of disease, but only 10% of it is considered to be foodborne; in contrast, *C. cayetanensis* causes a lot less disease, but 90% of it is considered to be foodborne, much of which is based on outbreak data. *Cryptosporidium parvum*, for which proven effective therapy does not exist, has gained notoriety for causing persistent chronic diarrhea in immunocompromised patients. *C. parvum* is endemic in cattle and is usually acquired in humans from contaminated water, fresh produce, unpasteurized milk, or person-to-person spread. The incubation period is typically about a week but can be as long as 28 days. *C. parvum* is known to cause large outbreaks, the largest of which was waterborne in Milwaukee, in which around 400 000 individuals became sick. *C. cayetanensis* infection has been associated in the past with consumption of imported berries most likely contaminated with fecally contaminated water. More recently, it has been linked with fresh basil. It can be diagnosed by direct acid-fast microscopy of stools, but it is important to note that most microbiology laboratories will not routinely look for *C. parvum* and/or *C. cayetanensis*, so they may be more common than we think.

Transmissible Spongiform Encepalopathies

Mad cow disease or bovine spongiform encephalopathy (BSE) was first diagnosed in the UK in 1986. BSE is a member of the family of transmissible spongiform encephalopathies (TSEs) that are linked with neurological diseases in a variety of animals, including humans, sheep, elk, and mink (Table 6). Historically, TSEs were recognized in Europe as long ago as the eighteenth century, when a neurological condition known as scrapie was first described in sheep. However, TSEs did not become a public health or food safety concern until the mid-1990s, when the link between BSE and a condition in humans, known as new variant Creutzfeldt–Jakob disease (nvCJD), was made, and it became clear that the disease was due to transmissible agents, that were given the name prions.

Although there are no clear answers as to the origins of BSE, the generally held belief is that BSE originated from feeding cattle rendered protein produced from the carcasses of scrapie-infected sheep or cattle. The rendering process involves processing carcasses by boiling at atmospheric or higher

Table 6 Types of transmissible spongiform encephalopathies (TSEs) in animals and humans

TSEs in animals
Scrapie (sheep and goats)
Chronic wasting disease
Transmissible mink encephalopathy
Bovine spongiform encephalopathy
TSEs in humans
Crutzfeldt–Jakob disease
Kuru
Gerstmann–Straussler–Scheinker syndrome
Fatal familial insomnia

pressures, which results in the generation of an aqueous protein solution under a layer of fat. The fat can be removed and the protein solution further processed into a meat and bone meal product for use as an animal food. Changes were made in the rendering process in the UK around 1980, and it is assumed that this change allowed the survival of the etiologic agent. The recycling of cattle carcasses through the rendering process only served to concentrate the etiologic agent further that eventually led to the epidemic of BSE in the UK in the late 1980s. Thus, the current belief is that the etiologic agent can be delivered orally and is concentrated in certain parts of an infected animal's body, especially central nervous tissues.

In the year 2000, there were 1101 new cases of BSE diagnosed in the UK, so infectious cattle still remain in the UK. BSE has not been limited to the UK, and there have been several hundred cases in other European countries and elsewhere in the world, although some of these are imported rather than native cases. Since 1994, when the first case of nvCJD was recognized in the UK, there has been a steady increase in the numbers of cases. Up to November 2000, there had been 84 cases in the UK, and the trends in incidence since 1994 indicate that the number of cases and deaths per year are on the rise.

Up to the time of this writing, there have been no documented cases of BSE in the USA. The USDA and FDA have taken action to prevent the importation of BSE-contaminated material or animals into the USA. Active surveillance for BSE is underway in cattle in the USA with signs of neurological disease and cattle that are nonambulatory at slaughter. Close to 12 000 brains have been examined so far, with no evidence of BSE. Despite the lack of evidence of BSE in the USA, there is concern that it will appear at some point. The long incubation period and the difficulty in controlling all imports of potentially contaminated products from different parts of the world make it possible that BSE will appear in the USA, at some time in the future. In the general context of food safety, the risks and the likelihood of serious outcomes that are associated with bacterial pathogens such as *Campylobacter*, *Salmonella*, and *E. coli* are significantly greater than the risk of developing nvCJD from BSE. However, nvCJD is frightening, because it is incurable, inevitably leads to death, and there are currently no assays available to test food for the BSE prions, although new tests to determine the presence of nervous tissue in meat products are becoming available.

See also: **Bovine Spongiform Encephalopathy (BSE)**; ***Campylobacter***: Properties and Occurrence; Detection; Campylobacteriosis; ***Escherichia coli***: Occurrence; Detection; Food Poisoning; Occurrence and Epidemiology of Species other than *Escherichia coli*; Food Poisoning by Species other than *Escherichia coli*; **Food Poisoning**: Classification; ***Listeria***: Properties and Occurrence; Detection; Listeriosis; **Parasites**: Occurrence and Detection; Illness and Treatment; ***Salmonella***: Properties and Occurrence; Detection; Salmonellosis; **Viruses**

Further Reading

Allos BM (2001) *Campylobacter jejuni* infections: update on emerging issues and trends. *Clinical Infectious Diseases* 32: 1201–1206.

Centers for Disease Control (2000) Surveillance for foodborne disease outbreaks – United States, 1993–1997. *MMWR* 49: 1–51.

Centers for Disease Control (2001) Diagnosis and management of food borne illness, a primer for physicians. *MMWR* 50: RR-2.

Guerrant RL, Van Gilder T, Steiner TS *et al*. (2001) Practice and guidelines for the management of infectious diarrhea. *Clinical Infectious Diseases* 32: 331–350.

Hohmann EL (2001) Nontyphoidal salmonellosis. *Clinical Infectious Diseases* 32: 263–269.

Kramer JM, Frost JA, Bolton FJ and Wareing DR (2000) Campylobacter contamination of raw meat and poultry at retail sale: identification of multiple types and comparison with isolates from human infection. *Journal of Food Protection* 63: 1654–1659.

Mead PS, Slutsker L, Dietz V *et al*. (1999) Food-related illness and death in the United States. *Emerging Infectious Diseases* 5: 607–625.

Mosier DA and Oberst RD (2000) Cryptosporidiosis. A global challenge. *Annals of the New York Academy of Sciences* 916: 102–111.

Nataro JP and Kaper JB (1998) Diarrheagenic *Escherichia coli*. *Clinical Microbiology Reviews* 11: 142–201.

Paton JC and Paton AW (1998) Pathogenesis and diagnosis of Shiga toxin-producing *Escherichia coli* infections. *Clinical Microbiology Reviews* 11: 450–479.

Prusiner SB (1998) Prions. *Proceedings of the National Academy of Sciences* 95: 13363–13383.

Ranjitham M, Madan M and Chandrasekharan S (1999) *Cyclospora cayetanensis* – an emerging coccidian parasite. *Journal of the Association of Physicians of India* 47: 1198–1199.

Slifko TR, Smith HV and Rose JB (2000) Emerging parasite zoonoses associated with water and food. *International Journal of Parasitology* 30: 1379–1393.

Emission Spectroscopy *See* **Spectroscopy**: Overview; Infrared and Raman; Near-infrared; Fluorescence; Atomic Emission and Absorption; Nuclear Magnetic Resonance; Visible Spectroscopy and Colorimetry

EMULSIFIERS

Contents

Organic Emulsifiers

T Kinyanjui and W E Artz, University of Illinois, Urbana, IL, USA
S Mahungu, Egerton University, Njoro, Kenya

Background

This chapter will narrowly focus on organic emulsifiers, one category in the general class of compounds called surface-active agents, rather than macromolecular stabilizers or any of the other components involved in emulsion stabilization. Some familiar emulsions include those occurring in foods (milk, mayonnaise, etc.), cosmetics (creams and lotions), pharmaceuticals (soluble vitamins and hormone products), and agricultural products (insecticides and herbicides).

The discussion will be limited to food emulsifiers, rather than the use of emulsifiers for nonfood products, such as cosmetics, paints or drug-delivery systems. Emulsifiers are substances that reduce the surface tension at the interface of two normally immiscible phases, allowing them to mix and form an emulsion. Representing one class of a broader group of ingredients called surface active ingredients or surfactants, emulsifiers assert their effects at the interface between oil, water, or air dispersed in a second immiscible fluid. They promote the formation and stabilization of emulsions as well as foams and suspensions. There are several types of emulsions, including oil in water (o/w) water in oil (w/o), oil in water emulsion dispersed in oil (o/w/o) etc.

Emulsions are a special kind of colloidal dispersion: one in which a liquid is dispersed in a continuous liquid phase of a different composition. The dispersed phase is sometimes referred to as the 'internal phase' and the continuous phase as the 'external phase.'

Structure

Food emulsifiers can be categorized (Table 1) on the basis of several characteristics, including origin, either synthetic or natural; potential for ionization, nonionic versus ionic; hydrophilic/lipophilic balance (HLB); and the presence of functional groups.

Lecithin and Lecithin Derivatives

The primary source of lecithin, the only naturally occurring emulsifier used in any significant quantities in the food industry, is soybeans. Soybean oil contains anywhere from 1 to 3% phospholipids in the crude oil. Other sources include corn, sunflower, cottonseed, rapeseed, and eggs. Lecithin is obtained by an aqueous extraction of the oil extracted from soybeans. Phase separation occurs upon hydration of the phospholipids, and the two phases are separated by

Table 1 Some organic emulsifiers

Lecithin and lecithin derivatives
Monoglycerols
Sucrose esters of fatty acids
Hydroxycarboxylic acid and fatty acid esters
Lactylate fatty acid esters
Polyglycerol fatty acid esters
Ethylene or propylene glycol fatty acid esters
Ethoxylated derivatives of monoglycerides
Sorbitan fatty acid esters
Propylene glycol monoesters
Phosphated mono and diglycerols
Sodium stearoyl-2-lactylate
Calcium stearoyl-2-lactylate
Fruit acid esters

centrifugation. The crude extract, after water removal, contains about 35% triglycerols and smaller amounts of nonphospholipid materials. Extraction with acetone is used to produce an oil-free lecithin. The term 'lecithin' has been used to describe both phosphatidylcholine and mixtures of phospholipids. Current recommendations by the International Union of Pure and Applied Chemistry–International Union of Biochemistry suggest the use of 3-*sn*-phosphatidylcholine rather than lecithin to describe 1,2-diacyl-*sn*-glycero-3-phosphatidylcholine. However, a commercial, soybean-derived lecithin preparation contains several different phospholipids, primarily phosphatidylcholine, phosphatidylethanolamine and phosphatidylinositol. The structures are shown in Figure 1. (*See* **Phospholipids**: Properties and Occurrence.)

Commercial lecithin preparations can be either treated or modified chemically to provide a product with altered functional characteristics. Treatment with either hydrogen peroxide or benzoyl peroxide will produce a lighter-colored product. The chemical modification of lecithin by reaction with hydrogen peroxide, plus lactic or acetic acid and water, will produce a hydroxylated product. Hydroxylation occurs at the double bonds, altering lecithin such that its hydrophilic character is increased. The result is a product with improved oil-in-water emulsifying properties relative to unmodified lecithin.

Triglycerols are soluble in acetone, whereas phospholipids are not. Therefore, the greater the percentage of acetone-insoluble material, the greater the phospholipid content in crude lecithin. Because of this, one of the primary criteria for the evaluation of lecithin is the percentage of acetone-insoluble material. Lecithin is also evaluated on the basis of several other parameters including acid value (an indication of free fatty acids), hexane-insoluble matter (an indication of fibrous material), water, peroxide value and metallic impurities. Individual phospholipids in soy lecithin can be quantitated using HPLC.

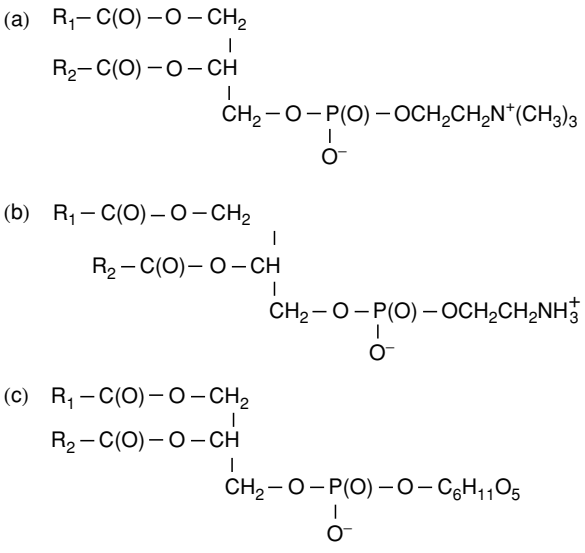

Figure 1 Primary phospholipids reported in commerical lecithin, where R_1 and R_2 are fatty acids: (a) phosphatidylcholine; (b) phosphatidylethanolamine; (c) phosphatidylinositol.

Mono- and Diglycerides

Mono- and diglycerides are the most commonly used food emulsifiers. They consist of esters synthesized via catalytic transesterification of glycerol with triglycerides, with the usual triglyceride source as hydrogenated soybean oil. Mono- and diglycerides are also synthesized directly from glycerol and fatty acids under alkaline conditions. Molecular distillation is used to prepare a purified product containing up to approximately 90% monoglycerol. Monoglycerols can be prepared from the reaction of glycidol (2,3-epoxy-1-propanol) and carboxylic acids with a yield in excess of 90%. Advantages of the process include the synthesis of difficult to produce monoglycerides and a good potential for continuous processing. Mono- and diglycerols have also been obtained from a butterfat fraction by chemical glycerolysis. The enzymatic preparation of mono- and distearin by glycerolysis of ethylstearate and direct esterification of glycerol in the presence of a lipase from *Candida antarctica* has also been reported.

Several tests are used for characterizing commercial sources of mono- and diglycerides, including total monoglycerides, hydroxyl value, iodine value, and the saponification value. With the monoesters, the fatty acid can be attached at either the alpha or beta positions, as with the diglycerides (see Figure 2).

Hydroxycarboxylic and Fatty Acid Esters

To produce an emulsifier with an increased hydrophilic character relative to monoglycerides, small organic acids are esterified to monoglycerols (Figure 3). Some

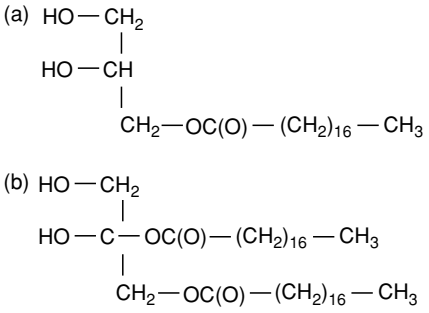

Figure 2 (a) Mono- and (b) diglycerides.

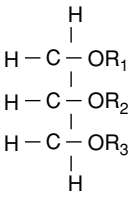

Figure 3 Organic acid ester of monoglycerol, where at least one R is a short-chain organic acid, for example, acetic acid.

of the acids used are acetic, citric, fumaric, lactic, succinic, and tartaric. Succinylated monoglycerols are synthesized from succinic anhydride and distilled. They are used by the baking industry as dough conditioners and crumb softeners. Acetic acid esters of mono- and diglycerides are synthesized from fatty acids plus acetic anhydride or by transesterification. The product is lipid-soluble and water-insoluble. Functions in food include control of fat crystallization and improvement of aeration properties of high-fat foods. They are often added to shortenings or cake mixes. (*See* **Fats:** Classification.)

To synthesize other acid esters, citric acid esters of mono- and diglycerides, glycerol is esterified with a mixture of citric acid and fatty acids. It can also be prepared by the direct esterification of citric acid with glyceryl monooleate. The product is hot water- and lipid-soluble. Functions in food include emulsification, antispattering agent in margarine, improvement of bakery product characteristics, a fat replacement in high-fat foods and a synergist and solubilizer for antioxidants.

Diacetyl tartaric acid esters of monoglycerides (DATEM) are synthesized from diacetyl tartaric acid anhydride and monoglycerides. The emulsification properties of DATEM depend primarily upon the type of fatty acid and the percentage of esterified tartaric acid. Emulsifier quality is based upon results from the analyses for tartaric and acetic acid, acid value, total fatty acids, saponification value, and metallic residues.

Lactic acid esters of mono and diglycerides consist of a mixture of lactic and fatty acid esters of glycerin. The emulsifier is dispersible in hot water. Important qualitative parameters include the percentage of monoglycerides, total lactic acid, acid value, free glycerin, and the amount of water.

Lactylate Fatty Acid Esters

Polymeric lactic acid esters of monoglycerides are also available, commonly known as sodium or calcium stearoyl-2-lactylates (see Figure 4). Typically, there are two lactic acid groups per emulsifier molecule. To produce the emulsifier, a mixture of the fatty acid, polylactic acid, and calcium or sodium carbonate is heated at about 200 °C for about 1 h with agitation in an inert atmosphere. The calcium salt is less dispersible in water than sodium stearoyl-2-lactylate.

Polyglycerol Fatty Acid Esters

Polyglycerol esters (Figure 5) of fatty acid are also used in food products, primarily in baked goods. They consist of mixed partial esters synthesized from the reaction of polymerized glycerol with edible fats. Polyglycerols will vary in degree of glycerol polymerization with an average specified. The source of fatty acids as well as the degree of polymerization can vary, providing a wide range of emulsifiers, from hydrophilic to very lipophilic.

Polyethylene or Propylene Glycol Fatty Acid Esters

Fatty acids can be esterified directly to polyethylene glycol ethers or by enzymatic preparation, which allows better control of the reaction. Propylene glycol fatty acid (C12, C14, C16, C18, and C18:1) monoesters (Figure 6) have been produced by lipase-catalyzed reactions. Propylene glycol monoesters of docosahexaenoic acid and eicosapentaenoic acid that are potentially health-beneficial w/o emulsifiers useful in the food industry have been synthesized by lipase-catalyzed esterification. The HLB of the emulsifier is altered by adjusting the degree of ethoxylation. Fatty acid polyglycol esters are good o/w emulsifiers.

Ethoxylated Derivatives of Monoglycerides

Ethoxylated mono- and diglycerides are produced from the reaction of several moles of ethylene oxide and mono- or diglycerols under pressure. Ethoxylation of monoglycerols results in a product that is much more hydrophilic relative to monoglycerols.

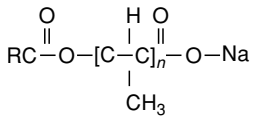

Figure 4 Sodium stearoyl-2-lactylate, where *n* normally averages 2, and R is a fatty acid moiety.

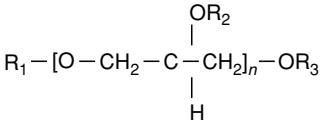

Figure 5 Polyglycerol esters of fatty acids, where R_1, R_2, and R_3 are each a fatty acid and/or a hydrogen, and where the average value of *n* is greater than 1.

The end product of the synthesis is actually a mixture with a distribution range and peak; therefore, many often vary among manufacturers.

Sorbitan Fatty Acid Esters

Polyoxyethylene sorbitan esters are synthesized by the addition, via polymerization, of ethylene oxide to sorbitan fatty acid esters. These nonionic hydrophilic emulsifiers are very effective antistaling agents and, thus, are used in a wide variety of bakery products. These emulsifiers are much more widely known as the polysorbates, e.g., polysorbate 20, 60, and 80 (Figure 7). Polysorbate 20, 60, and 80 utilize lauric, stearate, and oleate, respectively, for the fatty acid portion of the molecule. Polysorbate 60 is a monostearate, whereas polysorbate 65 is a tristearate.

Miscellaneous Derivatives

Fatty acids can be esterified directly to compounds other than glycerol, for example, sugar alcohols like sorbitol, mannitol, and maltitol, and sugars like sucrose, glucose, fructose, lactose, and maltose.

Sorbitol or sorbitan esters are formed from 1,4-anhydro-sorbitol and fatty acids (see Figure 8). Typically, the emulsifier consists of a mixture of stearic and palmitic acid esters of sorbitol and its mono- and dianhydrides. Ethoxylated derivatives can also be prepared by the addition of several moles of ethylene oxide to the sorbitan monoglycerol ester and, depending on the number of moles of ethylene oxide added, have a wide range in HLB.

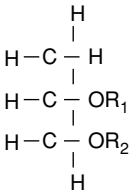

Figure 6 Propylene glycol esters of fatty acids, where R_1 and R_2 represent a fatty acid and/or a hydrogen, and where at least one R represents a fatty acid.

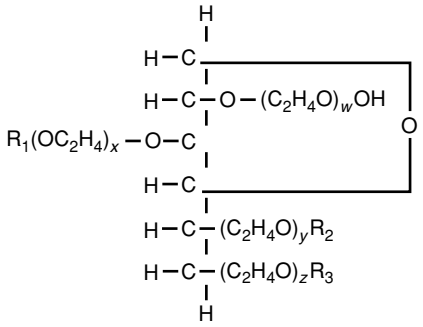

Figure 7 Polysorbates, where $w + x + y + z = 20$ (approx.) and Rs represent a single fatty acid and hydrogens for polysorbate 20, 40, 60, and 80. For polysorbate 65, each R represents a stearic acid moiety. The fatty acids are lauric, palmitic, stearic, and oleic acid for polysorbate 20, 40, 60, and 80, respectively.

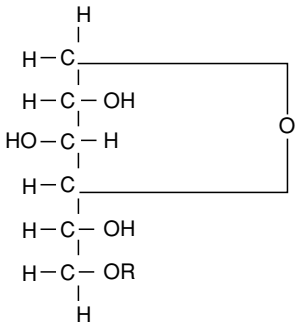

Figure 8 Sorbitan stearate, where R represents a fatty acid moiety, for example, stearic acid, oleic acid, lauric acid, or palmitic acid.

Lactitol (the hydrogenation product of lactose) palmitate is synthesized by direct esterification at a temperature of approximately 160 °C. Workers have reported the enzymatic synthesis of acetylated glucose fatty acid esters. Two immobilized lipases from *Candida antarctica* (SP 382) and *Candida cylindraceae* catalyzed the synthesis of novel acetylated glucose fatty acid esters with glucose pentaacetate and Trisun 80 (80% oleic) vegetable oil or methyl oleate as substrates in organic solvents. The incorporation of oleic acid on to the glucose ranged from 30 to 100%. It was possible to catalyze the synthesis of glucose fatty acid esters with free glucose as the sugar substrate. Other researchers have reported the synthesis of a novel nonionic surfactant, dialkyl glucosylglutamate from δ-gluconolactone, glutamic acid and alkyl alcohols. Sucrose fatty acid esters (Figure 9) can be synthesized using a variety of solvents or by direct esterification. The first description of a practical commercial process for the preparation of sucrose esters of fatty acids was reported in 1956. Enzymatic synthesis of carbohydrate esters of fatty acids has also been reported for the esters of sucrose, glucose, fructose, and sorbitol with oleic and stearic acid and fructofuranose. The reports by researchers indicate the enzymatic preparation of three different 1,6-diacyl fructofuranoses. At low temperatures (5 °C), the synthesis produces quantitative yields of the diesters by simple addition of the original sugar to a solution of the fatty acid in a solvent (acetone), which is accepted by the European Commission (EC) for use in the manufacture of additives. By varying the degree of esterification, the HLB and, hence, the functionality can be controlled. Sucrose

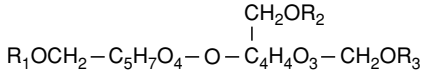

Figure 9 Sucrose fatty acid esters, where at least one of either R_1, R_2, or R_3 represents a fatty acid, and the remainder may represent a fatty acid or a hydrogen; the degree of substitution is 1–3.

monoesters have an HLB value greater than 16 while the triesters have an HLB value less than 1. Monoesters are particularly useful for the stabilization of o/w emulsions, whereas diesters are best for w/o emulsions. With esterification equal to or greater than 5 moles of fatty acid per mole of sucrose, the emulsification properties of sucrose fatty acid esters are lost. But, at that degree of esterification, the sucrose fatty acid polyester can be used as a low-calorie fat replacement since it is neither digestible nor absorbable.

The consistency of both o/w and w/o emulsions can be affected with the addition of ethylene or propylene glycol monostearate. The most common ethylene and propylene glycol esters used as emulsifiers are the monostearate and monopalmitate.

Emulsion Stability

Small droplets and the presence of an interfacial film on the droplets in emulsions make them stable; that is, the suspended droplets do not settle out or float rapidly, and the droplets do not coalesce quickly.

Emulsifiers are generally classified as: anionic emulsifiers, cationic emulsifiers, amphoteric emulsifiers and nonionic emulsifiers. Proteins (caseins) are known to stabilize emulsions, owing to their amphoteric nature. The oil phase viscosity also contributes to the stability of the emulsion stability, but this depends on the adsorption and interfacial behavior of the emulsifiers.

One basis for a theory of emulsion stability concerns the balance between the attractive and repulsive forces of the particle. In other words, attractive (Van der Waals forces or London dispersion forces) would tend toward destabilizing an emulsion, whereas the repulsive forces (electrostatic repulsion between electrical double layers of like sign) would stabilize the emulsion by keeping the droplets separated.

There are several phenomena that can cause emulsion destabilization. Each is affected by the presence of emulsifiers. To understand best the mechanism of emulsion formation and ultimately the forces stabilizing emulsions, the mechanisms of destabilization should be understood.

Mechanisms of Destabilization

Emulsion destabilization can be due to one or all of five possible mechanisms; flocculation, coalescence, sedimentation or creaming, Ostwald ripening, and phase inversion.

-Flocculation The adherence of droplets to form aggregates or clusters and the buildup of these aggregates are referred to as 'flocculation.' It occurs when the attractive forces between the droplets exceeds that of the repulsive forces, without a breakdown in the structural integrity of the interfacial film surrounding the droplets.

-Coalescence When aggregates or flocculates of the dispersed phase combine to form a single, larger drop, the phenomenon is referred to as 'coalescence.' Coalescence is really a reflection of the nature of the interfacial film on the surface of the droplet. A strong, stable film on the surface of the droplet, owing to the addition of the correct concentration of the appropriate emulsifier, will minimize this type of destabilization.

-Ostwald ripening If the two phases forming the emulsion are not totally immiscible, and there are differences in droplet size within the emulsion, larger droplets will form at the expense of smaller droplets owing to a process known as 'Ostwald ripening.' Ostwald ripening is always a factor since variations in initial droplet size always occur in macroemulsions, and both phases are never completely immiscible. The driving force for Ostwald ripening is the difference in chemical potential between droplets of difference sizes. Equilibrium will only exist when all droplets are the same size, which really means a single 'drop' or the presence of two continuous and separate phases.

-Phase inversion The viscosity of an emulsion will increase gradually as more and more of a given phase is added until a critical volume is reached. If more of that same phase is added, exceeding the critical volume, the emulsion will invert, i.e., the discontinuous phase will become the continuous phase.

Mechanisms of Stabilization by Emulsifiers

There are several factors, some of which are dependent on the emulsifiers and stabilizers added, involved in emulsion stabilization. The first is the reduction of interfacial tension by the emulsifiers. Next, is repulsion between droplets due to similar electrical charges on the surface of the droplets. A third is the formation of mesophases or liquid-crystalline phases, which will

provide the most stable configuration for a specified set of conditions. A fourth is the addition of macromolecules or particulate material, which can substantially increase emulsion viscosity and stability. An increase in viscosity of the continuous phase adds to the kinetic stability. However, without a concurrent energy barrier, viscosity will have a small effect on stabilization. Viscosity enhancers increase the stability of the energy barrier. Stable oil-in-water emulsions of argan oil with two different types of mixtures of nonionic emulsifiers have been produced. Three different types of oil (Israeli argan oil, Moroccan argan oil, and soybean oil) have been emulsified with mixtures of Span 80 and Tween 80. The optimum HLB value for argan oil is 11.0. The argan o/w emulsions are stable for more than 5 months at 25 °C. Synergistic effects have been found in enhancing the stability of emulsions prepared with sucrose monostearate. The origin of the oil and the internal content of natural emulsifiers, such as monoglycerides and phospholipids, have a profound influence on its interfacial properties and on the stability of the argan o/w emulsions.

Emulsion stability is also dependent upon the conditions under which the emulsion is formed. This includes not only the constituents of the emulsion but also the emulsifier concentration, emulsion temperature, and physical state (crystalline vs. fluid) of the emulsifier. The order of addition of the constituents is also an important factor. Addition of lecithin to the lipid phase prior to the addition of the aqueous phase can substantially alter the droplet size, liquid crystal formation, and emulsion stability. Another contributing factor is the nature of the internal and continuous phases. Both affect emulsion stability. Two types of emulsions, those prepared with unsaturated emulsifiers and unsaturated oil and those prepared with saturated emulsifiers and saturated oil, are more stable than those prepared with emulsifiers and oil of intermediate or mixed saturation.

Functions of Emulsifiers

Interfacial Tension

The reduction of interfacial tension through the addition of emulsifiers is a key factor in emulsion formation. It allows emulsion formation with considerably less energy input than would be required without the presence of an emulsifier. Once the interfacial film consisting of emulsifier is formed, it acts as an effective barrier to droplet coalescence. It has been found that, in the presence of emulsifiers, the droplet interface may acquire viscoelastic properties, which are important in the prevention of coalescence. A strong interaction between the hydrophilic portion of the emulsifier and the aqueous phase leads to a large reduction in the surface tension of the water; this also affects the type of emulsion formed. A weak interaction between water and the hydrophilic portion of the emulsifier molecule will favor a w/o emulsion, whereas a strong interaction will favor a o/w emulsion.

Electrical Charge

Ionic emulsifiers provide an additional mechanism for emulsion stabilization relative to nonionic emulsifiers, through ion–ion and ion–solvent interactions. In addition, the introduction of charged groups on the surface of the emulsion droplets increases the repulsive forces. Ionic emulsifiers will form an electrically charged double layer in the aqueous solution surrounding each oil droplet.

Liquid Crystal Stabilization

Macroemulsions, although thermodynamically unstable, can attain rather long-term stability, strongly suggesting an intermediate stability level. This was attributed to the formation of a liquid crystalline state by the emulsifier. It has been shown that the presence of a liquid crystalline state reduces the rate of coalescence, even if droplet flocculation occurs. (*See* **Crystallization**: Basic Principles.)

Stabilization by Macromolecules and Finely Divided Solids

Emulsion stability can be increased by the addition of macromolecules like gums and protein. Colloids like xanthan gum, carboxy methyl cellulose and guar gum significantly increase emulsion stability. With both a constant emulsifier and colloid concentration, emulsion stability is enhanced by increased emulsification temperature, increasing the degree of shear, and increase pH, in the range of 3–6. Colloids act by either increasing the viscosity or partitioning into the o/w interface and providing a physical barrier to coalescence.

To evaluate emulsion stability and thereby characterize the potential of an emulsifier, the rate at which the combined destabilization phenomena occur must be determined. These rates can be determined from the changes in the size and distribution of the oil droplets with time. There are several methods available for this determination. Nuclear magnetic resonance is often preferred as a better indication of stability to HLB values. Other methods, used to evaluate the effects of processing on emulsion stability, include centrifugation, turbidity light micoscopy scanning electron microscopy, etc.

Emulsifier Selection

Emulsifier selection is based upon the final product characteristics, emulsion preparation methodology, the amount of emulsifier added, the chemical and physical characteristics of each phase, and the presence of other functional components in the emulsion. Food emulsifiers have a wide range of functions. The most obvious is to assist stabilization and formation of emulsions by the reduction of surface tension at the o/w interface.

Perhaps the most important factor in preparing an emulsion is the selection of the appropriate emulsifier. Several methodologies have been developed to assist in such an endeavor. These include the HLB system of Griffin, the H/L numbers, the water number, the phase inversion temperature, and the emulsion inversion point. Even with the best of methods, selection can be very difficult, except perhaps for the few foods that are relatively straightforward emulsions, like mayonnaise and margarine. Often, one of the best sources of information is the emulsifier manufacturer. Several parameters should be considered during emulsifier selection. These parameters include: (1) approval of the emulsifier by the appropriate government agency, (2) desired functional properties, (3) end-product application, (4) processing parameters, (5) synergistic effect of other ingredients, (6) home preparation, and finally (7) cost.

Obviously, before an emulsifier can be used in a food product, it must be approved by the appropriate regulatory agency. Assuming this criterion is met, the most important considerations would be both the required functional properties of the selected emulsifier and the application. Delineating the required functional properties such as emulsification, starch complexation and crystallization control and the specific end-product application are the two major factors in emulsifier selection. An exact determination of these two parameters should focus attention on a limited number of emulsifiers. The processing methodology and equipment available in the processing facility could further limit the range of emulsifiers that are of potential use. It is at this stage that ingredient supplier(s) could begin to provide helpful assistance.

By far, the most widely used rule for the selection of food emulsifiers is the HLB number. The HLB index is based upon the relative percentage of hydrophilic to lipophilic groups within the emulsifier molecule. The assigned values range from 1 to 20. Lower HLB values indicate a more lipophilic emulsifier, whereas higher values indicate a more hydrophilic emulsifier. Emulsifiers with HLB numbers in the 3–6 range are best for w/o emulsions, whereas emulsifiers with HLB numbers in the range of 8–18 are best for o/w emulsions. Depending upon the application and the types of oils to be emulsified, there is an optimum HLB.

There is an equation that can be used to determine the HLB number for several types of nonionic emulsifiers, particularly the ethoxylated alcohols and the polyhydric fatty acid esters. To determine the HLB for the fatty acid ester type, the following equation can be used:

$$\mathrm{HLB} = 20(1 - S/A),$$

where A is the acid number, and S is the saponification number of the ester.

For the polysorbate type of emulsifier, the HLB value can be determined from the equation:

$$\mathrm{HLB} = (E + P)/5,$$

where E is the weight percentage of oxyethylene, and P is the weight percentage of polyhydric alcohol.

However, there are several factors that reduce the utility of the HLB selection system. One factor is that the HLB system does not work well for ionic emulsifiers, a problem further complicated by the fact that the charge varies with pH. Another factor is that commercial preparations usually contain two or more emulsifiers. These emulsifiers can have a significant synergistic effect that makes it very difficult to apply the HLB system. Another limitation is due to the fact that the HLB system is based upon the molecular structure of the emulsifier and does not take into account the combined oil/aqueous phase/emulsifier system. In addition, once the appropriate emulsifier has been selected, the correct concentration can not be determined from the HLB value. The concentration required is really a function of the droplet size. The smaller the droplet, the greater the surface area and, therefore, the greater the amount of emulsifier that is required for monolayer coverage of each droplet. In spite of these limitations, the HLB is still the most widely used index of emulsifier functionality.

See also: **Colloids and Emulsions**; **Fatty Acids**: Properties; **Phospholipids**: Properties and Occurrence

Further Reading

Artz WE (1990) Emulsifiers. In: Branen AL, Davidson MP and Salminen S (eds) *Food Additives*, pp. 347–393. New York: Marcel Dekker.

Artz WE and Myers MR (1994) Supercritical fluid extraction and chromatography of emulsifiers. *Journal of the American Oil Chemists Society* 72: 219.

Brian V (1999) Dispersion stabilization and destabilization by polymers. In: Dickinson E and Patino RJ (eds) *Food Emulsions and Foams – Interfaces, Interactions and Stability*, p. 20. Cambridge: Royal Society of Chemistry.

Charalambous G and Doxastaxis G (1989) *Food Emulsifiers; Chemistry, Technology Functional Properties and Applications*. New York: Elsevier.

Friberg ES and Larsson K (1997) Emulsion stability. In: Friberg ES and Larsson K (eds) *Food Emulsions*, p. 27–51. New York: Marcel Dekker.

Hasenhuetti GL (1997) Overview of food emulsifiers. In: Hasenhuetti GL and Hartel (eds) *Food Emulsifiers and their Applications*, pp. 1–9. New York: Chapman & Hall.

JEFCA (1994) *Joint FAO/WHO Expert Committee on Food Additives: Summary of Evaluations*. Rome: Food and Agriculture Organization of the United Nations.

Laurier LS (1992) Petroleum emulsions. In: Laurier LS (ed.) *Emulsions – Fundamentals and Applications in the Petroleum Industry*. Washington, DC: American Chemical Society.

Liu KJ and Shaw JF (1995) Synthesis of propylene glycol monoesters of docosahexaenoic acid and eicosapentaenoic acid by lipase-catalysed esterification in organic solvents. *Journal of the American Oil Chemists Society* 72: 1271.

Organization of the Food Chemical Codex (1981) *Food Chemicals Codex*. Washington, DC: National Academy Press.

Rydhag L and Wilton I (1981) The function of phospholipids of soybean lecithin in emulsions. *Journal of the American Oil Chemists Society* 58: 831.

Schuster G and Adams WF (1984) Emulsifiers as additives in bread and fine baked products. *Advances in Cereal Science* 6: 139.

Shimp LA (1985) Process cheese principles. *Food Technology* 39: 63–64.

Yaghmur A, Aserin A and Mizrahi Y (1999) Argan oil-in-water emulsions: preparation and stabilization. *Journal of the American Oil Chemists Society* 76: 15–18.

Yoshida K, Sekine T and Matsuzaki F (1999) Stability of vitamin A in oil-in-water-in-oil-type multiple emulsions. *Journal of the American Oil Chemists Society* 76: 195–200.

Phosphates as Meat Emulsion Stabilizers

L Knipe, Ohio State University, Columbus, OH, USA

Introduction

The emulsifying capacity of phosphates in meat emulsions may take several approaches, depending upon the definition of emulsification. A finely chopped meat mixture is conventionally referred to as a 'meat emulsion.' However, this is now considered to be a misnomer. The so-called meat emulsion consists of solid fat particles dispersed in a mixture of water and many fibrous particles, including connective tissue and muscle fibers. A true emulsion is a stable suspension of two liquids (oil and water) which are not normally soluble in each other. It might be more appropriate to refer to a meat emulsion as a matrix, whose stability is dependent upon the water-holding capacity or binding capacity of the meat proteins in the matrix. However, a finely chopped meat mixture will be referred to as an emulsion in the remainder of this article.

If the proper combination of meat ingredients is combined with the proper processing procedures (e.g., grinding, chopping, emulsifying), a stable emulsion will be prepared which will hold up well during the cooking process. Examples of meat emulsions are bologna sausages or wieners, which have such fine meat particles that they are not distinguishable on the smooth product surface. However, if either the quantity or quality of meat ingredients or the processing methods are inadequate, the meat mixture will be unstable and result in a poor-quality product upon cooking. (*See* **Meat**: Sausages and Comminuted Products.)

If it is assumed that a meat emulsion is not a true emulsion, then phosphates are probably not true emulsifiers, but stabilizers of meat mixtures. There are many factors involved in the stability of meat emulsions which can be affected by the addition of inorganic phosphates. The main effects of phosphates in finely chopped meat systems are on the pH, ionic strength, protein extraction, divalent cation binding, and viscosity. (*See* **Colloids and Emulsions; Stabilizers**: Types and Function; Applications.)

Phosphate Classification and Nomenclature

Inorganic phosphates are classified by the number of phosphorus atoms in the phosphate molecule. The types of phosphates that are important to the meat industry include the orthophosphates, the pyrophosphates, and the straight-chain phosphates.

Orthophosphates contain only one phosphorus atom per molecule. The pyrophosphate molecule consists of two phosphorus atoms linked by a shared oxygen atom. An inorganic phosphate of such structure is referred to as a condensed phosphate. The two pyrophosphates approved by the US Department of Agriculture's Food Safety Inspection Service (USDA FSIS) are sodium acid pyrophosphate and tetrasodium pyrophosphate.

Inorganic phosphates of three or more phosphorus atoms are referred to as polyphosphates. Sodium or potassium tripolyphosphates consist of three linked phosphorus atoms. Sodium hexametaphosphate is a

misnomer for a long straight-chain polyphosphate. The 'meta' designation is correctly given to cyclic polyphosphates. Hexametaphosphate contains an average of 10–15 phosphorus atoms. The name 'sodium hexametaphosphate' has been deleted from USDA regulations and replaced with 'sodium polyphosphate, glassy' (Graham's salt) and 'sodium metaphosphate, insoluble' (Maddrell's salt).

Phosphates and pH

Phosphates affect the pH of both water and meat. The pH values of 1% solutions of the phosphates approved for use in meat products, by the USDA FSIS, appear in Table 1. Whereas meat contains 60–70% moisture, the phosphate effect on the pH of meat, as determined by standard pH measurements, is somewhat less than their effect in water, due to the buffering capacity of meat. Alkaline phosphates increase meat pH in the range 0.1–0.6 units, depending upon the phosphate chosen. The following list indicates the decreasing order of impact on increasing meat pH – pyrophosphates, tripolyphosphates, and hexametaphosphates. Hexametaphosphates are considered to be rather neutral and often have no effect in increasing meat pH. Acid pyrophosphates often, in fact, decrease the pH of meat systems. There is, however, some debate among researchers as to the true importance of meat pH changes, due to the addition of phosphates, on the stability of meat emulsions. (*See* **pH – Principles and Measurement**.)

Water-Holding Capacity of Meat Emulsions

The water-holding capacity of meat could be compared to the action of a sponge and is important to meat processing in that as proteins are able to hold more water they become more soluble. The water-holding capacity of meat is at a minimum at the isoelectric point (pI) of meat proteins. At this point, equal positive and negative charges on the protein molecules result in a maximum number of bonds between peptide chains and a net charge equal to zero. The pI of meat (where water-holding capacity is at a minimum) is in the pH range of 5.0–5.4, which is also the ultimate pH of meat or the pH of meat after it has gone through rigor mortis.

The water-holding capacity of meat is greatly affected by pH. Increasing or decreasing the pH on either side of the pI will result in an increased water-holding capacity by creating a charge imbalance. A charge imbalance is a predominance of either positive or negative charges which will result in a repulsion of the charged protein groups of the same (positive or negative) charge. This repulsion results in increased capacity for water retention and could be compared to the repulsive effect of like charges on two magnets.

An increase in predominance of negative protein charges, due to the addition of phosphates, may also cause better distribution of fat particles in emulsified products. The better fat particle distribution may prevent the clumping of fat particles that occurs during overchopping and which results in an unstable emulsion.

Table 1 pH values for 1% water solutions of US Department of Agriculture-approved phosphates

Inorganic phosphate	*pH*
Tetrasodium or tetrapotassium pyrophosphate	10.2
Sodium or potassium tripolyphosphate	9.8
Disodium or potassium orthophosphate	8.8
Sodium polyphosphate, glassy	7.0
Sodium metaphosphate, insoluble	6.5
Monosodium or potassium orthophosphate	4.4
Sodium acid pyrophosphate	4.2

From Ellinger RH (1972) *Phosphates as Food Ingredients*. Cleveland, OH: CRC Press with permission.

Phosphates and Ionic Strength

Phosphates also affect the ionic strength of meat mixtures. Inorganic phosphates ionize in water to form polyelectrolytes. This ionization causes a masking of positive sites on the meat proteins which further causes an electrostatic repulsion of the proteins. An electrostatic repulsion between meat proteins allows more space between proteins for binding water, which increases the water-holding capacity. This effect is sometimes hard to differentiate from the pH effect on the water-holding capacity of meat. In addition to the repulsion theory, the long-chain phosphates have numerous charged anions along the chains, which allows direct binding of phosphates to water molecules. This is especially true for the long-chain phosphates, such as the hexametaphosphates.

Protein Extraction and Solubilization

During the formation of a meat emulsion, meat proteins are extracted from the fibrous muscle structure and are solubilized to form a solution. This protein extraction is enhanced by the optimal ionic strength and pH of the solution in which the protein is submersed. Once solubilized, with additional chopping or mixing, the liquid proteins are dispersed around the bundles of muscle cells and fat particles. When the mixture is then cooked, the protein solution denatures and coagulates (like egg white) to form a gel around the muscle bundles and fat particles. This gel

serves to stabilize the matrix of the meat mixture or emulsion. (*See* **Protein**: Functional Properties.)

Phosphates have a very valuable role in extracting proteins from muscles. In fact, tetrasodium pyrophosphates are known to have a specific effect on meat proteins in that they are able to extract muscle proteins, and stabilize meat emulsions, to a degree much greater than normally possible considering the effects of tetrasodium pyrophosphates on ionic strength, and pH compared to other phosphates. Part of the specific effect is that pyrophosphates serve to dissociate, or separate, actomyosin into its component parts, actin and myosin. This is very advantageous, as myosin by itself is more beneficial in producing a stable emulsion than actomyosin. Phosphates are also known to extract myosin more readily from fast (white) muscle fibers than slow (red) muscle fibers.

Commercially available phosphates are ranked in the following descending order of impact on extraction of muscle proteins from meat: tetrasodium or potassium pyrophosphates, sodium or potassium tripolyphosphates, sodium hexametaphosphates. There are cases, however, where it has been shown that level of protein extraction in a meat system was not well correlated to the emulsion stability of the final cooked product.

Phosphates and Divalent Cations

Phosphates also bind divalent cations, which are native to meat proteins and which are detrimental to finished processed-meat quality. Divalent cations, such as those of calcium, magnesium, and iron, are often found in untreated or unsoftened water supplies, as well as meat, and are known to decrease the water-holding capacity of meat. This is believed to be due to bridging and binding between several groups of proteins, similar to that which occurs during rigor mortis. Another theory is that divalent cations screen the negatively charged protein groups, minimizing any repulsive effect. However, there is some evidence that increased magnesium concentration enhances the myosin-extracting ability of pyrophosphates.

There is still some debate among researchers as to whether phosphates bind cations which are already bound to the muscle or bind only to free cations in solution within muscle. In either case, the binding of cations in a meat system increases the water-holding capacity. One indication that phosphates are binding divalent cations is in the antioxidant effects of phosphates added to meat products. (*See* **Antioxidants**: Synthetic Antioxidants.)

The long-chain phosphates, such as sodium hexametaphosphates, are most active in chelating calcium ions and short-chain phosphates, such as tetrasodium pyrophosphates, bind magnesium cations most readily.

Viscosity of Meat Emulsions

Phosphates are also important in the physiochemical aspects of meat processing. For example, phosphates reduce the viscosity of meat mixtures. This is important because, in the chopping or mixing steps, meat temperatures will rise due to frictional energy produced as the chopper or emulsifier knives are cutting and moving through the meat mixture. Chopping emulsions containing a soft fat, such as pork fat, to excessively high temperatures will result in an unstable product.

However, when meat is finely chopped to make a stable emulsion-type sausage product, the fat particles need to be reduced to a small enough size so that the extracted meat proteins can coat or entrap the fat. If the fat particles are too large, a coarse, unstable emulsion will result but, if the fat is chopped too much, the fat surface area may be too large or too many fat cells may be broken, also resulting in an unstable product. By reducing the viscosity, a mixture can also be chopped or mixed longer, either to reduce fat particle size or extract more protein to stabilize the mixture further, with less temperature rise. Without phosphates, long chopping or mixing times would result in unstable products.

Phosphates and Sodium Chloride

The functionality of phosphates is greatly affected by the addition of common table salt or sodium chloride to meat emulsions. The effects of combining salt and phosphates in meat emulsions is considered to be synergistic, which means that the measurable effects of the two ingredients combined is greater than the combined effects of the two ingredients added individually. It seems that phosphates exert more effect on the pH and protein solubility and salt exerts more effect on the ionic strength and water-holding capacity. The phosphate effect on muscle proteins appears to diminish at higher salt concentrations ($> 0.6\,\mathrm{mol\,l^{-1}}$ sodium chloride).

See also: **Antioxidants**: Synthetic Antioxidants; **Colloids and Emulsions**; **Meat**: Sausages and Comminuted Products; **pH – Principles and Measurement**; **Protein**: Functional Properties; **Stabilizers**: Types and Function; Applications

Further Reading

Arnold N, Wierbicki E and Deatherage FE (1956) Post mortem changes in the interactions of cations and

proteins of beef and their relation to sex and diethylstilbesterol treatment. *Food Technology* 10: 245.
Ellinger RH (1972) *Phosphates as Food Ingredients*. Cleveland, OH: CRC Press.
Hamm R (1971) Interactions between phosphates and meat proteins. In: DeMan JM and Melnychyn P (eds) *Symposium: Phosphates in Food Processing*. Westport: AVI.
Knipe CL, Olson DG and Rust DE (1985) Effects of sodium hydroxide and selected inorganic phosphates on the characteristics of reduced sodium meat emulsions. *Journal of Food Science* 50: 1017.
Knipe CL, Rust DE and Olson DG (1990) Some physical parameters involved in the addition of inorganic phosphates to reduced-sodium meat emulsions. *Journal of Food Science* 55: 23.
Lyons JW and Siebenthal CD (1966) On the binding of condensed phosphates by proteins. *Biochimica et Biophysica Acta* 120: 174.
Parsons N and Knight P (1990) Origin of variable extraction of myosin from myofibrils treated with salt and pyrophosphates. *Journal of Science of Food and Agriculture* 51: 71.
Shults GW and Wierbicki E (1973) Effects of sodium chloride and condensed phosphates on pH and swelling of chicken muscle. *Journal of Food Science* 38: 991.
Shults GW, Russell DR and Wierbicki E (1972) Effect of condensed phosphates on pH, swelling and water-holding capacity for beef. *Journal of Food Science* 37: 860.
Trout GR and Schmidt GR (1984) Effect of phosphate type and concentration, salt level and method of preparation on binding in restructured beef rolls. *Journal of Food Science* 49: 687.
Whiting RC (1987) Influence of various salts and water soluble compounds on the water and fat exudation and gel strength of meat batters. *Journal of Food Science* 52: 1130.
Wierbicki E, Howker JJ and Shults GW (1976) Effect of salt, phosphates and other curing ingredients on shrinkage of lean pork meat and the quality of smoked processed ham. *Journal of Food Science* 41: 1116.
Xiong YL, Lou X, Wang C, Moody WG and Harmon RJ (2000) Protein extraction from chicken myofibrils irrigated with various polyphosphate and NaCl solutions. *Journal of Food Science* 65(1): 96.

Uses in Processed Foods

T Kinyanjui and W E Artz, University of Illinois, Urbana, IL, USA
S Mahungu, Egerton University, Njoro, Kenya

Emulsions

Natural food emulsions existed, of course, long before food processing. A good example is milk, a natural emulsion/colloid in which fat is stabilized by a milk-fat-globule membrane. The development of technologies for processing oils, such as refining, bleaching, and hydrogenation, has led to the use of food emulsifiers.

The definition of an emulsion has continued to evolve since the 1930s. Becher developed an elaborate definition from several previous authors:

> An emulsion is a heterogeneous system, consisting of at least one immiscible liquid intimately dispersed in another in the form of droplets, whose diameter, in general, exceeds 0.1 μm. Such systems possess a minimal stability, which may be accentuated by such additives as surface-active agents, finely divided solids, etc.

Liquid crystalline phases are significant in emulsion stability. This was reflected in the IUPAC-IUB definition of an emulsion: 'In an emulsion, liquid droplets and/or liquid crystals are dispersed in a liquid.' Sharma and Shah defined both micro- and macroemulsions, differentiating them on the basis of size and stability. Macroemulsions were defined as:

> mixtures of two immiscible liquids, one of them being dispersed in the form of fine droplets with (a) diameter greater than 0.1 μm in the other liquid. Such systems are turbid, milky in color and thermodynamically unstable.

Microemulsions were defined as:

> the clear thermodynamically stable dispersions of two immiscible liquids. The dispersed range consists of small droplets in the range of 100–1000 Å.

Food macroemulsions are unstable systems, even with the addition of emulsifiers. Emulsifiers must be added to increase product stability and attain an acceptable shelf-life. The function of an emulsifier is to mix the oily and aqueous phases of an emulsion, in a homogeneous and stable preparation. The main characteristic of an emulsifier is that its molecules must contain two parts. The first part must have a hydrophilic affinity, and the second part must have a lipophilic affinity. Emulsifiers are generally classified as: anionic emulsifiers, cationic emulsifiers, amphoteric emulsifiers, and nonionic emulsifiers. (*See* **Colloids and Emulsions**.)

Emulsifier selection is based upon final product characteristics, emulsion preparation methodology, the amount of emulsifier added, the chemical and physical characteristics of each phase, and the presence of other functional components in the emulsion. Food emulsifiers have a wide range of functions. The most obvious is to assist in the stabilization and formation of emulsions by reducing the surface tension at the oil–water interface.

By far the most used rule of the selection of a food emulsifier is the HLB value. The HLB value, which is

the 'hydrophile–lipophile balance,' is based on the relative percentage of hydrophilic to lipophilic groups within the emulsifier molecule. Values range from 1 to 20: a lower HLB value indicates a more lipophilic emulsifier, whereas higher HLB value indicates a more hydrophilic emulsifier.

The first step in the formation of a stable emulsion is dispersion of one phase in another. A critical factor in the emulsification process is the formation of a monomolecular layer at the lipid–water interface by the emulsifier. During emulsion formation, there is a large increase in surface area (up to several thousand-fold), which is dependent upon the number and size of the droplets. To form and disperse these droplets, a substantial amount of energy or work must be supplied. Since emulsifiers reduce the surface tension, the addition of emulsifiers reduces the amount of work that must be done to form the emulsion. The most common method of emulsion formation is the application of mechanical energy via vigorous agitation.

Emulsifier Chemistry

Schuster has written a comprehensive treatise on emulsifiers with extensive discussions on the current theories of emulsion formation, emulsifier chemistry, function, and analysis. He included a discussion covering all the emulsifiers currently used in Germany and the USA, plus applications in a wide variety of products.

Applications in Foods

Categories

Food emulsifiers can be categorized (Table 1) on the basis of several characteristics, including: origin, either synthetic or natural; potential for ionization, nonionic versus ionic; HLB; and the presence of functional groups.

Some biopolymer-type emulsifiers, such as whey proteins, will function both as an emulsifier and as a thickener. For underprocessed emulsions, the droplets may not be dispersed finely enough. Overprocessing an emulsion may reduce or eliminate the maximum networking ability of a stabilizing macromolecule colloid (protein or polysaccharide), due to the excessive mechanical or thermal treatment.

An example of an anion-type food emulsifier is sodium stearoyl lactylate. When combined with distilled, saturated monoglycerols (1:1, w/w), vesicles form that are relatively stable in food systems. These work well for such applications as aerating or texturing low-fat foods.

Table 1 Some food emulsifier categories[a]

Emulsifier
Mono- and diglycerols
Propylene glycol monoesters
Lactylated esters
Polyglycerol esters
Sorbitan esters
Ethoxylated esters
Succinated esters
Fruit acid esters
Acetylated mono- and diglycerols
Phosphated mono- and diglycerols
Sucrose esters

[a]Adapted from Zielinski RJ (1997) Synthesis and composition of food grade emulsifiers. In: Hasenhuetti GL and Hartel R (eds) *Food Emulsifiers and their Applications*, pp. 11–38. New York: Chapman and Hall.

When adding an emulsifier–water mesophase as a part of a food formulation, it is important to realize that phase changes may take place due to the effect of the food ingredients (fats, oils, proteins, salts, etc.) on the mesophase, as well as the effect of temperature during pasteurization, sterilization, or freezing of foods.

Functionality

Surface active compounds operate through a hydrophilic head that is attracted to the aqueous phase, and a lipophilic tail that partitions into the oil phase. The surfactant is positioned at the oil–water interface, where it can act to lower the surface or interfacial tension. Lipophilic tails are generally composed of C16 or longer fatty acids. Emulsifiers with shorter fatty acid chains can act as emulsifiers but may hydrolyze to cause soapy or other undesirable off-flavors.

The concentration of added emulsifiers in foods is usually too low to result in any formation of liquid crystalline phases during food processing; nevertheless, some may be formed during processing by the polar lipids present in the food ingredients. This is the case with bakery products such as bread, rolls, and buns based on wheat flour, which contains polar lipids, such as glycolipids and phospholipids. Another example is mayonnaise and salad dressings where the phospholipids from the eggyolk can form liquid crystalline structures. Acetic acid or lactic acid esters of monoglycerols or propylene glycol monostearate are used in whippable emulsions, such as creams, toppings, etc.

Emulsifiers have a wide range of functional properties in addition to the obvious property – stabilization of food emulsions. Table 2 lists the functional properties of food emulsifiers compiled from a variety of sources including product brochures from several emulsifier manufacturers.

Table 2 Functional properties of food emulsifiers[a]

Functions	*Product examples*
Starch complexer	Marcaroni, pasta
Antispattering agent	Margarines
Gloss enhancer	Confectionery coatings
Emulsification, water-in-oil emulsions	Margarine
Emulsification, oil-in-water emulsions	Mayonnaise
Aeration	Whipped toppings
Whippability	Whipped toppings
Inhibition of fat crystallization	Candy
Softening	Candy
Antistaling	Bread
Dough conditioner	Bread dough
Improve loaf volume	Bread
Reduce shortening requirements	Bread
Pan release agent	Yeast-leavened and other dough and batter products
Fat stabilizer	Food oils
Antispattering agent	Margarine and frying oils
Antisticking agent	Caramel candy
Protective coating	Fresh fruits and vegetables
Surfactant	Molasses
Control viscosity	Molten chocolate
Improved solubility	Instant drinks
Starch complexation	Instant potatoes
Humectant	Cake icings
Plasticizer	Cake icings
Defoaming agent	Sugar production
Stabilization of flavor oils	Flavor emulsions
Promotion of 'dryness'	Icecream
Freeze–thaw stability	Whipped toppings
Improve wetting ability	Instant soups
Inhibition of sugar Crystallization	Panned coatings

[a]Adapted from Hasenhuettl GL (1997) Overview of food emulsifiers. In: Hasenhuetti GL and Hartel R (eds) *Food Emulsifiers and their Applications*, pp. 1–9. New York: Chapman and Hall and Artz WE (1990) Emulsifiers. In: Branen AL, Davidson MP and Salminen S (eds) *Food Additives*, pp. 347–393. New York: Marcel Dekker.

Petroswski has compiled a good list of some of the commercial emulsifiers and described their nature, iodine, HLB values, and melting points.

Uses

-Cereal-based products Baked products without emulsifiers have been described as undesirable in terms of texture, taste, and appearance. Current processing, distribution, and storage techniques used for baked products require the use of additives that maintain quality and freshness.

Reviews by Schuster and Adams and Stampfli and Nersten cover research on emulsifiers with respect to baked goods and include excellent discussions on the mechanism of emulsification, the interaction of emulsifiers with starch, and a wide range of applications.

A series of bakery blends have been formulated for several fat-free, low-fat, or reduced-fat bakery applications, including muffins and sweet doughs, cakes and frostings. Emulsifiers function as dough conditioners, giving the product increased resistance to mixing and mechanical abuse, stronger side walls, more uniformity, and improved slicing characteristics. Several different types of bakery products, in addition to bread produced via conventional methods, benefit from added emulsifiers, including white bread produced by a highly mechanized process, buns and rolls, yeast-raised sweet rolls and doughnuts, plus variety breads that contain a significant portion of nonwheat flour components. (*See* **Biscuits, Cookies, and Crackers**: Methods of Manufacture.)

Certain emulsifiers can also function as crumb softeners, another type of dough conditioner. Soft crumb, a characteristic of freshly baked bread, can be retained longer if the appropriate emulsifiers are added. Crumb firming, associated with staling and starch retrogradation, can be delayed for 2–4 days with the addition of emulsifiers. When shortening is added to a wheat dough, it will improve the volume.

Cake volume and crumb texture are both related to the number of air bubbles present in the aerated batter. Commercially manufactured gel phases are available for use as aerating agents in low-fat cakes (Table 3.)

Investigators have attributed the effect of emulsifiers on the delay of crumb firming to the ability of emulsifiers to function as foam stabilizers.

Most investigators have found that lecithin has the same characteristics as other dough conditioners on bakery products: increased water absorption, reduced mixing time, improved machinability, and longer freshness. Lecithin can also improve the wettability of cake mixes. Phospholipids, treated with phospholipase to alter functionality, have been utilized in wheat dough to improve product characteristics. (*See* **Phospholipids**: Properties and Occurrence.)

The wheat proteins, gliadin and glutenin, form a viscous colloidal complex called 'gluten' when mixed into a dough. The properties of the dough are influenced by lipid emulsifiers. Emulsifiers interact with starch to form a complex that slows down the staling process. A blend of whey proteins, mono- and diglycerides, lactic acid fatty acid esters, and lecithin will increase the spreadability, reduce stickiness, increase creaminess, and contribute body to the baked product. (*See* **Wheat**: Grain Structure of Wheat and Wheat-based Products.)

A cake batter is a fat-in-water emulsion, where the aqueous phase contains dissolved sugar and suspended flour particles. In layer cakes, emulsifiers aid in aeration, lubrication, and crumb texture.

Table 3 Influence of emulsifiers on production and quality of baked products[a]

Process	*Influence of emulsifier*	*Advantages*
Mixing	• Improvement of wettability • Stabilization of phases • Better distribution of phases • Interaction of emulsifier/lipid • Interaction of emulsifier/starch surface • Interaction-emulsifier/protein	• Decrease of mixing and speed • Reduced shortening levels • Improvement of mixing • Greater shock tolerance • Improvement of machinability
Fermentation	• Interaction-emulsifier/protein	• Improvement of gas-retaining properties • Shorter fermentation • Greater shock tolerance
Baking	• Emulsifier–starch complex formation	• Improvement of gas-retaining properties • Improved loaf volume • Better texture • Better crumb grain • Better uniformity • Decreased water loss
Storage	• Emulsifier–starch complex formation	• Improvement of crumb softness and longer shelf-life

[a]Modified from Orthoefer FT (1997) Applications of Emulsifiers in Baked Foods. In: Hasenhuetti GL and Hartel R (eds) *Food Emulsifiers and their Applications*, pp. 211–234. New York: Chapman and Hall.

Nonflour lipid additives such as polar lipids (surfactants or emulsifiers) are commonly included in standard baking recipes today. Their effects on dough or cake batter and on the final product include an improvement in texture and an extended shelf-life. The disadvantages of using oil alone in bread include a weak dough, dull crumb, and reduced loaf volume. Emulsifiers can alter the solid fat index and the crystal polymorphism of shortening.

The deleterious effect of wheat bran on bread quality can be offset by the addition of emulsifiers and vital wheat gluten. Lecithin (0.5%) is used in shortening, but it is slightly less effective than 0.5% of ethoxylated monoglycerides, sucrose monopalmitate, or diacetyl tartaric esters for improving whole wheat bread.

A mixture of hydrophilic and lipophilic emulsifiers provides the best dispersibility for bakery products. Diacetyl tartaric acid ester of monoglycerides, a widely used emulsifier in Europe for bread products, appears to act as a dough strengthener and a crumb softener. It can also improve loaf volume and delay staling, and it shows synergistic activity with monoglycerides for some functions.

A family of starch–lipid composites, prepared by passing aqueous mixtures of starch and lipid through a steam jet cooker, offer flavor and texture advantages when combined with food products, particularly ground beef. The jet-cooked dispersions are stable and do not phase-separate, even after prolonged standing. They can be drum-dried and milled to yield dry powders that are not oily to the touch. The lipid component is dispersed within the starch or starch–water matrix as droplets, about 1–10 μm in diameter. The products, trademarked as Fantesk™, are a new type of starch–lipid composite prepared by a continuous steam jet cooking process. (*See* **Bread**: Breadmaking Processes.)

Flat bread products are consumed extensively in the Middle East, North Africa, and the Indian subcontinent. Like other bread products, the flat breads will stale upon storage. The addition of sodium stearoyl-2-lactylate and monoglycerides reduces the degree of firmness upon storage of flat breads. The addition of sucrose esters to chapaties, a flat bread product, allows the substitution of 40% of the whole wheat flour with nonglutenous flour (e.g., corn, sorghum, millet, or soy) with only a slight loss in quality. This would facilitate soy flour fortification of chapaties to improve the nutritional quality of the product.

There is a wide range of baked products, other than bread, with improved characteristics as a result of the addition of emulsifiers. Emulsifiers can be used to prevent, inhibit, or control starch gelation in macaroni, spaghetti, and snack foods, and maintain cooked rice in a dry and nonsticky condition.

Cookie characteristics (volume, top grain, and spreadability) can be improved by the addition of the appropriate emulsifier. The interaction of emulsifiers with starch alters the effect of heat on starch. Puddings, for example, have an increased freeze–thaw stability and increased resistance to retort processing, whereas starch-based sauces have an improved freeze–thaw stability and increased stability during holding at high temperatures.

Sucrose esters of fatty acids are used as nonionic food emulsifiers in nondairy based whipping cream

and dessert toppings, icecream, bread cakes and sponge products, beverage whiteners, and pasta and noodles.

-Dairy products Milk is a complex food emulsion. The emulsion component is composed of fat droplets dispersed in an aqueous phase containing protein.

Ice cream Icecream is both a foam and an emulsion, containing ice crystals and an unfrozen aqueous phase whose freezing point is depressed due to the presence of salts, sugars, and polysaccharide stabilizers.

Many emulsifiers used in icecream have a low HLB value. A number of comparisons of the effects of different emulsifiers in icecream have been made and established that emulsifiers impart desirable structural characteristics to icecream. The emulsifier enables icecream to be extruded into shapes for wrapping, or to retain its shape when it is filled into a cone from a counter freezer. Monolaurate, monooleate, and monostearate are effective in icecream emulsification destabilization, which is also enhanced by the addition of eggyolk solids and butter milk solids due to their surface-active lipid content. The function of emulsifiers in icecream may be summarized as:

1. After homogenization, a minimum amount of emulsifier is needed to prevent coalescence. The emulsion is then stabilized by adsorbed milk proteins.
2. Emulsifiers reduce the amount of protein needed, possibly through the involvement of a mesomorphic phase at the interface.
3. Saturated emulsifiers promote crystallization of the emulsified fat.
4. Emulsifiers promote the destabilization of the emulsion during freezing.

Emulsifiers are particularly useful with formed specialty products, such as icecream sandwiches, slices, and factory filled cones. Emulsifiers coat the milk fat, increasing the ability of the protein film to surround the air cells in the icecream. The results are an improved whippability, smoother texture and body, drier appearance, and slower melting rate. The smoother texture is favored since both ice crystal and air cell sizes are reduced. Two types of emulsifiers are used: (1) mono- and diglycerides and (2) polyoxethylene derivatives of sugar alcohol fatty acid esters.

The polyoxyethylene derivatives of the sorbitan fatty acid esters are particularly effective in imparting stiffness to icecream immediately upon leaving the freezer. The emulsifiers are added not to stabilize the mix, but to impart desirable end-product qualities. Selection of a suitable emulsifier should be based on HLB since optimum product characteristics are dependent on the emulsifier HLB. The optimum emulsifier HLB is 16. Icecream production has been characterized as a deemulsification process due to a combination of agitation, freezing, and the presence of an emulsifier. Emulsifiers give the icecream the dryness and stiffness required by agglomerating the finely dispersed fat globules to properly. (*See* **Dairy Products – Nutritional Contribution**.)

Other dairy products Cream liqueurs represent an unusual type of dairy emulsion that requires long-term stabilization of dispersed fat in an alcoholic environment. Soluble sodium caseinate can be used to stabilize an emulsion that shows finely dispersed fat.

Flack found that mono- and diglycerides, citric acid esters of monoglycerides, lactic acid esters of monoglycerides, and distilled monoglycerides affected the whippability and foam stability of dairy whipping cream that had been pasteurized and homogenized. Lactic acid esters produced a stable foam with a high overrun that was substantially different from a fresh whipped cream, whereas a mixture of citric acid fatty acid esters and distilled monoglycerides produced a foam similar to that of fresh whipped cream. Blends of xanthan gum, carageenan, and galactomannans act as good stabilizers for frozen and chilled dairy products, such as icecream, sherbet, sour cream, sterile whipping cream, and recombined milk.

Processed cheese consists of fat droplets dispersed in a concentrated protein gel network. The emulsion stability in the fat droplets is controlled primarily by adsorbed caseins or hydrolyzed casein fractions. Some manufacturers add mono- and diglycerols as emulsifiers. To control the structure of processed cheese, 'emulsifying' salts, such as polyphosphates, are employed. Although these emulsifying salts are not surface-active, they can play an important role in modifying the activity of the surface-active casein. Cheese yield can be increased with the addition of an emulsifier, such as lecithin. Lecithin was added in quantities of 0.001–0.066% by weight of milk. Studies have reported the use of coconut protein as an emulsion stabilizer in cheese analogs. (*See* **Cheeses: Chemistry of Gel Formation.**)

In concentrated milk products, emulsifiers can aid in emulsion formation and stabilization. Some can increase the heat stability of milks and milk proteins. Usually, low-molecular-weight emulsifiers are utilized, particularly mono- and diglycerides. With certain butter products, phospholipids can be added as antispattering agents to prevent fat spattering during pan and grill frying.

Confectionery The benefits of emulsifiers in both chocolate and sugar confectionery products have been well established. The most common example in confectionery products is the use of lecithin in chocolate to reduce product viscosity and to facilitate handling. Another emulsifier used in chocolate is polyglycerol polyricinoleate, which is used to modify the viscosity of chocolate coatings. (*See* **Sweets and Candies:** Sugar Confectionery.)

Sorbitan monostearate inhibits the migration of fat retarding bloom and extends the shelf-life of chocolate coatings. 'Chocolate bloom' may be defined as the mottled discoloration of confectioners coatings followed by a loss of the gloss. This defect occurs when the coating starts to melt, and some of the melting fat migrates to the surface. The cocoa components that impart the color are left behind inside the coating. Upon cooling, the melted fractions resolidify on the surface to form lighter colored blotches.

Emulsifiers are often used with both semisweet and milk chocolate. Emulsifiers aid in processing by reducing the weep or exudate that occurs with heavy sugar pastes during processing. For example, a marshmallow-based frozen confectionery product is prepared with 0.2–0.8% emulsifier with an HLB between 3 and 9.

Peanut butter contains about 50% peanut oil suspended in peanut fibers. Upon standing, the oil can separate from the peanut fibers, which impairs the product's appearance and palatability. Mono- and diglycerides can be added to emulsify part of the oil during processing, thereby preventing phase separation. Other benefits include an improved stability and spreadability.

Chewing gums contain fats and emulsifiers that soften the gum base and act as carriers for color and flavor. Lecithin is generally used for this purpose. Emulsifiers (sucrose fatty acid esters, sorbitan fatty acid esters, glycerol fatty acid esters, or propylene glycol fatty acid esters) have been used to improve the shelf stability of center-filled chewing gum. The emulsifiers are added to the flavored liquid center at 0.01–0.5% by weight. Some emulsifiers, such as the saturated ethoxylated monoglycerols, can be adsorbed on to starch granules. This property can be used to modify the texture of starch-based sugar confectionery.

Miscellaneous Liquid nondairy creamers (coffee whiteners) are used in the beverage industry. The nondairy whitener modified for use in a hot, acidic environment (coffee) can be prepared by pasteurizing and homogenizing a water-rich lipid emulsion consisting of 6–15% edible fat, approximately 0.6–2% mixed emulsifier (constituting approximately 0.3–1% low-HLB mono- and diglycerides). These whiteners are popular in dry powder form. Evaluation of emulsion stability for liquid coffee whiteners has revealed that if the emulsion is unstable, coagulation can occur. However, with the addition of the correct mixture of emulsifiers, optimum product characteristics are maintained.

To improve the wetting characteristics in instant cocoa drinks, the powder is agglomerated. Lecithin facilitates agglomeration during spray drying. During the spray drying process, the hydrophobic portion of the emulsifier dissolves in the cocoa butter, orienting the hydrophilic portion of the phospholipid towards the surface of the particle. This results in an increased affinity of the cocoa powder for water, thus aiding dispersion and wetting.

Margarine exists as a water–oil emulsion for only a short period of time prior to chilling, during which time, the emulsion is converted to a dispersion of water in a semisolid fat phase. Upon solidification, the product stability is greatly enhanced, since coalescence is essentially eliminated. Emulsifiers fulfill three functions in margarine: (1) assistance in emulsion formation, (2) modification of the crystal structure in the vegetable fat, and (3) antispattering during frying. Typically, a mixture of lecithin or citric acid monoglycerides and monoglycerides is used. A stable emulsion during frying reduces the coalescence of water droplets to form large drops. This facilitates gradual water evaporation, rather than explosive evolution or spattering, during frying. An additional explanation may be that the lecithin sludge formed during frying serves as nuclei for the formation of small vapor bubbles.

To reduce sandiness and 'oiling out' in margarine due to recrystallization, emulsifiers can be added. Sorbitan monostearate and citric acid esters of monoglycerides are effective in preventing recrystallization in tristearin. However, to be effective in margarine, the emulsifier must be very soluble in the oil phase. Other emulsifiers used include polysorbates and polyglyerol esters of fatty acids. To produce a low-calorie imitation dairy product, emulsifiers are usually added to the product. The caloric content can be significantly reduced by replacement of part of the fat with the appropriate emulsifier system. For low-fat margarine containing approximately 40% fat, a mixture of saturated and unsaturated monoglycerides and lecithin has been used successfully. Similarly, emulsifiers are used for partial fat replacement in low-fat, butter-flavored, liquid spreads. The stability of milk fat and water emulsions is dependent on the amount of emulsifier added.

Mayonnaise is an oil–water emulsion containing a high percentage of oil (> 70%). Owing to the high

percentage of fat, coalescence, rather than creaming, is the primary problem. Different protein sources are utilized because of their effectiveness in reducing coalescence.

The manufacturing procedure for mayonnaise is critical since the product contains a very high percentage of oil. Lecithin, contained in the added eggyolk, is usually the only emulsifier added. However, the vegetable oil may contain emulsifiers added to inhibit crystallization, e.g., ethoxylated sorbitan monooleate and monostearate. To inhibit cloud formation in salad dressings and salad oils, emulsifiers can also be added. (*See* **Dressings and Mayonnaise**: The Products and Their Manufacture.)

Emulsifiers can be used to produce emulsions that can deliver flavor, color, and a dispersed lipid phase in a cooked food product. Blending pork volatiles with the appropriate emulsifier, such as the polyglycerol ester of palmitic acid, and warming it will produce a relatively stable, flavored emulsion. The product can be used to replace part of the fat in a pork analog product. The fat level of such a product is 50% that of the regular pork product. In emulsion sausages, the product properties are dependent on a meat–protein matrix, a gel that gives the hot dog, for example, its characteristic sensory properties. The meat proteins are important in forming and stabilizing the emulsion.

See also: **Bread**: Breadmaking Processes; **Cakes**: Methods of Manufacture; **Cheeses**: Processed Cheese; **Dressings and Mayonnaise**: The Products and Their Manufacture; Chemistry of the Products; **Emulsifiers**: Organic Emulsifiers; Phosphates as Meat Emulsion Stabilizers; **Ice Cream**: Methods of Manufacture; **Margarine**: Methods of Manufacture; **Sweets and Candies**: Sugar Confectionery; **Whey and Whey Powders**: Production and Uses

Further Reading

Artz WE (1990) Emulsifiers. In Branen AL, Davidson MP and Salminen S (eds) *Food Additives*, pp. 347–393. New York: Marcel Dekker.

Berger KG (1997) Ice cream. In: Frieberg SE and Larsson K (eds) *Food Emulsions*, 3rd edn, pp. 235–278. New York: Marcel Dekker.

Buchheim W and Dejmek P (1997) Milk and dairy type emulsions. In: Frieberg SE and Larsson K (eds) *Food Emulsions*, 3rd edn, pp. 235–278. New York: Marcel Dekker.

Euston RS (1997) Emulsifiers in dairy products and dairy substitutes. In: Hasenhuetti GL and Hartel R (eds) *Food Emulsifiers and their Applications*, pp. 173–204. New York: Chapman & Hall.

Flack E (1997) Magarine and spreads. In: Hasenhuetti GL and Hartel R (eds) *Food Emulsifiers and their Applications*, pp. 211–234. New York: Chapman & Hall.

Hasenhuetti GL (1997) Overview of food emulsifiers. In: Hasenhuetti GL and Hartel R (eds) *Food Emulsifiers and Their Applications*, pp. 1–9. New York: Chapman & Hall.

John JC (ed.) (1983) *Food Additives; Recent Developments*, pp. 133–185. Park Ridge, NJ: Noyes Data Co.

Kulp K (1989) Emulsifiers for cake systems. In: Charalambous G and Doxastaxis G (eds) *Food Emulsifiers; Chemistry, Technology, Functional Properties, and Applications*, pp. 511–524. New York: Elsevier.

Orthoefer TF (1997) Application of emulsifiers in baked foods. In: Hasenhuetti GL and Hartel R (eds) *Food Emulsifiers and their Applications*, pp. 211–234. New York: Chapman & Hall.

Petrowski GE (1976) Food-grade emulsifiers Part II. *Food Technology*, 30(7): 36–40.

Pszzola DE (1994) Blends reduce fat in bakery products. *Food Technology* 48: 168–170.

Schuster G and Adams WF (1984) Emulsifiers as additives in bread and fine baked products. *Advances in Cereal Science and Technology* 6: 139.

Stampfli L and Nersten B (1995) Emulsifiers in bread making. *Food Chemistry* 52: 353.

Sworn G (2000) Xanthan gum. In: Phillips GO and Williams PA (eds) *Handbook of Colloids*, pp. 103–115. Cambridge, UK: CRC Press, Woodhead Publishing.

Weyland M (1997) Emulsifiers in confectionery. In: Hasenhuetti GL and Hartel R (eds) *Food Emulsifiers and their Applications*, pp. 235–254. New York: Chapman & Hall.

Emulsions *See* **Colloids and Emulsions**

Endives *See* **Salad Crops**: Leaf-types; Other Types of Salad Crops; Root, Bulb, and Tuber Crops

ENERGY

Contents

Measurement of Food Energy

J C Brown, AFRC Institute of Food Research, Norwich, UK

Background

The accurate assessment of the energy value of foods and diets is essential for the study of energy metabolism and energy balance, and for dealing with problems of malnutrition and obesity. Whereas the body oxidizes the constituents of food in the process of metabolism, the bomb calorimeter determines the energy released as a result of direct oxidation of the food constituents, and so enables measurement of an energy value. The energy value obtained is the *heat of combustion*, which is the maximum potential energy present in the food and is called its gross energy. The gross energy makes no allowance for losses of energy, owing to incomplete digestion or during metabolism.

In nutritional studies, the energy values used are for metabolizable energy, which is the energy available to the body from food. These values are usually calculated using energy conversion factors applied to the fat, protein, and carbohydrate content of a food, and take into account the average energy losses in feces and urine. For greater accuracy, normal losses of energy to feces and urine can be deducted from the gross energy, which is determined by bomb calorimetry, to give a metabolizable energy. In more precise studies the metabolizable energy value is measured, rather than calculated, by determining these energy losses directly.

Principles of Bomb Calorimetry for the Determination of Energy Values of Foods

The principle that allows the determination of energy value using bomb calorimetry is based on the first law of thermodynamics. Implicit in this law is the finding termed 'the law of constant heat summation,' which appeared in a paper published by G.H. Hess in 1838. Hess showed that the heat produced in a given chemical reaction was always the same, regardless of the intermediate steps.

The major sources of energy in foods are the organic constituents fat, carbohydrates, and protein, with smaller contributions from alcohol, organic acids, and traces of other organic materials such as mineral oils. The foods are completely oxidized in the bomb calorimeter, to yield the end products: carbon dioxide, water, oxides of nitrogen and sulfur, and heat. In the process of bomb calorimetry, the sample is rapidly combusted in oxygen at increased pressure, and the heat production is measured. The apparatus is calibrated with benzoic acid, a thermochemical standard, and a heat of combustion is derived for the food.

Occasionally, when an absolute measurement of heat of combustion is required, corrections are applied to the result obtained, e.g., for (1) sulfuric acid production from sulfur in the sample, (2) nitric acid production present in the bomb during combustion, and (3) departures from the standard state, pressure, or temperature of reactants or products. For the purposes of determining the energy values of foods or biological samples, the above corrections are normally sufficiently small to be ignored.

Types of Equipment Used

There are two different types of bomb calorimeter currently in use: the adiabatic (without gain or loss of heat) and the ballistic. The adiabatic bomb is favored for accurate nutritional studies. The ballistic bomb gives more rapid determinations but with a lower accuracy.

Development of the Adiabatic Bomb Calorimeter

The structure and operation of all bomb calorimeters, from the earliest (dated around 1900) to those currently in use, are similar. The bomb chamber in which combustion of the sample takes place is typically made from noncorroding steel and submerged in a water-filled calorimeter vessel. The water is continuously stirred and its temperature recorded, before and

after it rises as a result of heat produced by the combustion of the fuel. With the early designs of bomb calorimeter, it was necessary to determine a cooling correction, since some of the heat generated during the combustion was lost to surrounding air. Assessment of cooling correction involved the recording of accurate temperature observations made at precise time intervals, commencing at least 5 min before firing and continuing until at least 5 min after a constant rate of temperature change is resumed.

Major developments in bomb calorimeter design were concerned with providing heated outer water jackets capable of minizing heat loss from the calorimeter vessel surrounding the bomb. In 1900, Richards had used a closely controlled exothermic reaction within the water jacket to control the temperature. Benedict and Higgins, in 1910, introduced the use of an electrical heating coil to control the temperature of the water jacket, but rapid and close matching of the temperatures of the jacket and calorimeter vessel containing the bomb was still impossible. In 1916, Daniels achieved this by using heating elements, making the whole jacket a conductor. This system was automated by Raymond, Canaway, and Harris, in 1957, to produce an adiabatic bomb calorimeter, which was the basis for current systems. The Berthold–Maher and Maher–Cook bomb calorimeters, produced by Gallenkamp in the early 1950s, had a simple copper water jacket to provide some insulation from the surroundings, but around the late 1950s, Gallenkamp were producing an adiabatic bomb calorimeter.

The current adiabatic bomb, produced by Sanyo Gallenkamp, is shown in Figure 1 and is an improved version of the early ones described above. It is equipped with heaters, which, using a sensitive electrical system, maintain the outer water jacket at the same temperature as that of the water in the calorimeter vessel throughout the period of operation. Employment of the adiabatic concept eliminates cooling from the calorimeter vessel and the need for a cooling correction. This gives the system advantages over previous designs: (1) determinations are more rapid and lead to less operator fatigue, since continuous readings of temperature are not necessary; (2) readings only need to be taken once the equipment has stabilized, before and after the temperature rise resulting from combustion. Further improvement in the control of the temperature of the outer water jacket has been achieved by employing a heated reservoir of water (shown in Figure 1), which injects warm water into that already circulating in the outer water jacket.

Operation of the Adiabatic Bomb Calorimeter

The energy value is determined on a homogeneous dry sample that is compacted into a preweighed crucible. The electrodes are connected by a piece of nickel/chromium fuse wire. To facilitate combustion, a piece of cotton is tied to the wire and is placed in contact with the sample below. The crucible and electrodes are sealed in the bomb chamber, and oxygen is introduced to a pressure of 3030 kPa. The bomb chamber is placed inside the calorimeter vessel, which contains a constant mass of water, and the lid of the calorimeter is closed. The temperature of the water in the calorimeter vessel is recorded, and, once stabilized, the sample is ignited. The heat produced by the combustion is liberated into the bomb, water, and calorimeter vessel. The temperature of the outer water jacket increases in parallel with that of the water in the calorimeter vessel as heating elements are activated via balanced thermistors located in the water jacket and calorimeter vessel. A new steady temperature is achieved in 5–8 min and the final temperature reading taken. The heat capacity (heat

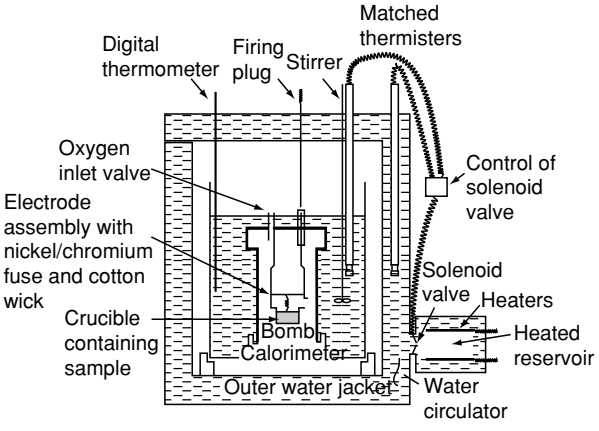

Figure 1 Adiabatic bomb calorimeter. Reproduced from Energy: Measurement of Food Energy, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

released per unit temperature rise) of the apparatus is determined using benzoic acid, and the temperature rise obtained from samples is translated into the energy released. This type of bomb allows corrections for sulfuric acid and nitric acid production to be made, if required, by collection and dual titration of the rinsings from the inside of the bomb chamber. This process is extremely tedious and is unnecessary for most purposes (unless the amount of sulfur present in the fuel leads to a significant quantity of extra heat release when it is converted to sulfuric acid rather than sulfur dioxide). Control of the adiabatic bomb calorimeter by microcomputer is also now possible, but this only controls recording measurements of temperature, and firing. The more labor-intensive parts of the operation – preparation of samples, introduction of oxygen, and cleaning and reloading of the bomb – must still be performed by the operator.

Development of the Ballistic Bomb Calorimeter

In 1959, Fox, Miller, and Payne described an alternative to the adiabatic bomb, giving quicker determinations of energy values. Their design made use of a light bomb casing and eliminated the water-filled calorimeter vessel. On combustion, most of the heat produced is transmitted to the upper parts of the bomb casing and measured using a thermocouple linked to a meter. No attempt is made to measure the equilibrium temperature of the bomb. In the modified version, a thermocouple is attached to the lid and connected to a galvanometer via a series resistance. This bomb, known as the ballistic bomb, was subsequently improved by reducing the thermal capacity of the upper casing.

The current ballistic bomb, shown in Figure 2, is directly descended from that described above. Possible sources of error involved variation in (1) the oxygen pressure, (2) the initial temperature, and (3) the crucible position in the bomb. These were standardized at the optimum level for operation.

Operation of the Ballistic Bomb Calorimeter

The sample, crucible, and electrodes are assembled in the bomb chamber, as in the adiabatic bomb. Oxygen is introduced, and, once stabilized (30 s), the sample is ignited, with maximum deflection on the galvanometer occurring after about 40 s. Deflection of the galvanometer scale is translated into the energy released by calibration with benzoic acid.

Comparison of Adiabatic and Ballistic Bomb Calorimeters

The ballistic bomb has an accuracy of $\pm 1.1\%$ on six determinations, compared with $\pm 0.1\%$ on two determinations with the adiabatic bomb. It was argued that a standard error of less than $\pm 0.5\%$ of the energy value is essential in accurate energy-balance studies. This level of accuracy would not normally be obtained with the ballistic bomb unless about 15 replicates were carried out. It was confirmed that to obtain the same precision on the ballistic bomb as that of the adiabatic, six times as many replicates would have to be carried out. Since the ballistic bomb is only two to three times faster, it is not a

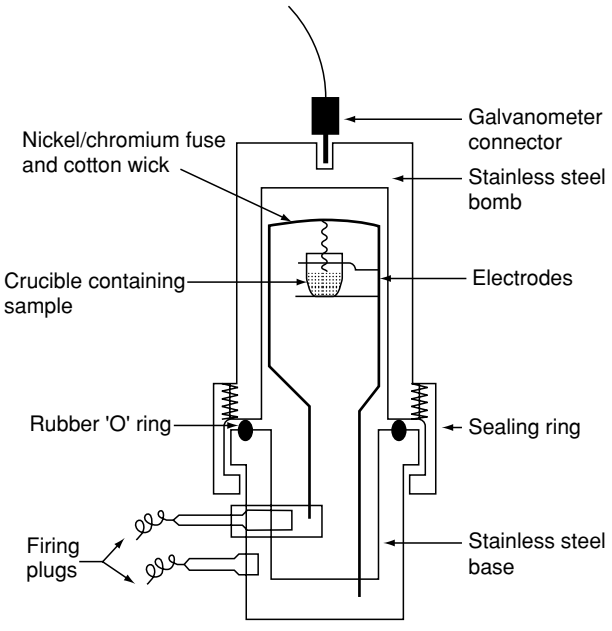

Figure 2 Ballistic bomb calorimeter. Reproduced from Energy: Measurement of Food Energy, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

serious alternative to the adiabatic system for accurate energy balance studies. Table 1 gives average heats of combustion or gross energies of some whole foods and protein, fat, and carbohydrate, determined using bomb calorimetry.

Limitations of Gross Energy Values Obtained for Diets

The energy value determined for a diet using bomb calorimetry is the energy that is released when the fuels are fully oxidized. This complete oxidation does not occur in the body. The energy value obtained for the food (gross energy) takes no account of the digestibility or metabolizability of the energy in the food by the body. Equations (1) and (2) show the relationship that exists between gross, digestible, metabolizable, fecal and urinary energy (*GE*, *DE*, *ME*, *FE*, and *UE*, respectively):

$$GE - FE = DE. \quad (1)$$

The digestibility of the energy of protein, fat, and carbohydrate is reckoned to be about 97%, 95%, and 98%, respectively.

$$DE - UE = ME. \quad (2)$$

This step sometimes involves additional energy losses owing to combustible gases produced during anaerobic fermentation.

The energy in protein is not fully metabolized, some energy-yielding nitrogenous components being present in urine. Atwater determined the heat combustion of food components and found that urine contained $5.23\,kJ\,g^{-1}$ ($1.25\,cal\,g^{-1}$) of absorbable protein, if the subject was in nitrogen equilibrium. (*See* **Protein**: Digestion and Absorption of Protein and Nitrogen Balance.)

The usual purpose of determining the energy value is to obtain a measure of the energy that the body gains from a food, its metabolizable energy value. Metabolizable energy can be measured by determination of gross energy content of diet, feces and urine (eqns (1) and (2)). Collection and analysis of these materials, however, are time-consuming and difficult, especially in humans. The alternative that is most commonly employed, but less accurate, involves the use of appropriate energy conversion factors. These can be used to calculate the metabolizable energy from the composition of the food or diet, or to correct the gross energy obtained by bomb calorimetry for the usual energy losses to feces and urine.

In 1896, Atwater put forward energy conversion factors for calculating the metabolizable energy obtained from protein, fat and carbohydrate of 17, 37 and $17\,kJ\,g^{-1}$ (4, 9, $4\,kcal\,g^{-1}$), respectively, which accounted for losses to feces and urine. These factors vary between foods, and, using the more precise system of Merrill and Watt, it is possible to use the compositional data and energy conversion factors derived for each individual food to calculate the metabolizable energy value. The British system of average factors used for most purposes are $17\,kJ\,g^{-1}$ ($4.0\,kcal\,g^{-1}$) for total protein, $37\,kJ\,g^{-1}$ ($9\,kcal\,g^{-1}$) for total fat, and $16\,kJ\,g^{-1}$ ($3.75\,kcal\,g^{-1}$) for total available carbohydrate (measured as monosaccharides).

An alternative to the use of compositional data for estimating metabolizable energy is to correct the gross energy, determined by bomb calorimetry, for usual energy losses in feces and urine. Miller and Payne assumed 95% digestibility of diets and put forward a system for estimating the metabolizable energy from determined gross energy ($kJ\,g^{-1}$) and

Table 1 Average heats of combustions of whole foods and of protein, fat, and carbohydrate from different types of foods

Material	*Heat of combustion ($kJ\,g^{-1}$, dry-weight basis)*
Food type	
Beef	27.2
Mutton	30.0
Pork	32.8
Poultry, fish	26.4
Eggs	29.7
Butter	37.0
Milk	22.8
Bread	19.0
Rice	18.3
Oats, rolled	20.2
Legumes	18.2
Vegetables, fresh	17.6
Fruit, fresh	18.4
Proteins	
Lean-meat protein	23.6
Milk protein	23.7
Egg protein	24.1
Wheat, gluten	24.9
Legumin	24.2
Fats	
Meat fat	39.8
Lard	40.1
Butter fat	38.8
Cereal oil	39.0
Nut, olive oil	39.7
Carbohydrates	
Glucose	15.6
Cane sugar	16.6
Milk sugar	16.2
Starch	17.5
Cellulose	17.5
Glycogen	17.5
Nonstarch polysaccharide	
Cereal	17.5 (16.7–18.5)
Vegetable	16.8 (16.2–17.9)
Fruit	16.5 (14.9–17.3)

the nitrogen content of the diet (N; g kg^{-1}), which is shown in eqn (3). This equation is appropriate for determination of ME (kg g^{-1}) in Western diets that are low in dietary fiber:

$$ME = 0.95GE - 31.88N. \quad (3)$$

However, 95% digestibility is not always true, especially of diets high in dietary fiber. An alternative system, based on a regression equation that deducts a value of 16.7 kJ g^{-1} (4 kcal g^{-1}) for unavailable carbohydrate (UC), or dietary fiber was proposed. This equation (eqn (4)) was derived using diets of up to 30 g of dietary fiber per day; thus, it has this upper limit.

$$ME = 0.977GE - 27.6N - 16.7UC. \quad (4)$$

A further equation (eqn (5)) takes account of the lowering of digestibility of gross energy diets as the level of dietary fiber (F; %) increases. This equation is used where dietary fiber intake is above 30 g per day, and especially where the diet is high in cereals.

$$ME = (0.95 - F)GE - 31.4N. \quad (5)$$

An equation, based on a regression analysis of many diets, including those of more than 30 g of dietary fiber per day, was derived. This equation (eqn (6)) accounts for the effect of unavailable carbohydrate on the digestible energy content of the diet and can be used in all cases, except with high-cereal diets. Equations (5) and (6) therefore complement each other in terms of application.

$$ME = 0.96GE - 9UC - 30N. \quad (6)$$

The coefficient of individual variation on these equations is around 1% of GE; however, there is uncertainty about how appropriate they are for predicting ME of diets of the elderly, the very young (under 8 years), individuals in the diseased state or those of different races. Since the energy equations are intended for calculating available energy from whole foods, they are most likely to be inappropriate when used for diets including isolates, or novel ingredients, which may fall outside of the scope of the derived equations.

To obtain an accurate value of DE or ME, the heat of combustion of diet, feces, and urine must be determined by bomb calorimetry. Most of the considered variation in metabolizability of a diet or ingredient comes about as a result of a difference in digestibility of energy rather than metabolizability (except for very poorly metabolized materials such as novel sugars that are excreted in the urine). It is therefore sometimes sufficient and more practical in accurate biological studies to make a measurement of DE – which does not require collection and analysis of urine – rather than ME.

See also: **Protein**: Digestion and Absorption of Protein and Nitrogen Balance

Further Reading

Livesey G (1990) Energy values of unavailable carbohydrate and diets: an inquiry and analysis. *American Journal of Clinical Nutrition* 51: 617–637.

McLean JA and Tobin G (1987) *Animal and Human Calorimetry*. Cambridge: Cambridge University Press.

Merril AL and Watt BK (1955) *Energy Value of Foods: Basis and Derivation*. Washington, DC: US Government Printing Office.

Miller DS and Payne PR (1959) A ballistic bomb calorimeter. *British Journal of Nutrition* 13: 501–508.

Intake and Energy Requirements

G R Goldberg, British Nutrition Foundation, London, UK

Introduction

This chapter outlines the progress that has been made in the last 10 years in the evaluation and interpretation of energy intake data, and in the use of energy expenditure data to define energy requirements. In particular, the data presented here illustrate the important contribution of doubly labelled water measurements of total free-living energy expenditure. These have provided points of reference that can be used to assess and define both energy intakes and energy requirements.

Energy Intake

Gross, digestible, and metabolizable energy obtained from food

The energy in food is provided by the four energy-yielding macronutrients, protein, fat carbohydrate and alcohol. The energy obtained from macronutrient oxidation is utilized in a number of ways. It can be used to perform mechanical, electrical, and chemical work; be deposited as new tissue; or exported from the body (e.g., as milk). The gross energy of a food is the energy obtained when it is totally combusted in a bomb calorimeter, and is the maximum amount of energy that can be extracted. However, not all the energy in food is available for use by the body. Not all the macronutrients are digested completely, not all

the products of digestion are absorbed completely, and energy is lost in urine as nitrogenous compounds (including urea, uric acid and creatinine), because the protein that is digested and absorbed is not oxidized completely.

$$\text{Gross energy}_{\text{food}} - \text{gross energy}_{\text{feces}} = \text{digestible energy}$$

$$\text{Digestible energy} - \text{protein energy}_{\text{urine and feces}} = \text{metabolizable energy.}$$

Estimates of Energy Content Using Food Tables

Tables of food composition usually list the energy content of foods not as gross, but as available or metabolizable energy, i.e., allowances have already been made for digestibility, and for the incomplete oxidation of protein by using food specific or mixed diet factors. Differences in the methods used to determine the carbohydrate content of individual foods and inclusion of the nondigestible fractions may lead to an overestimation of the energy content, particularly in fruits and vegetables. However, when calculating the total energy content of a mixed diet, these differences are of little practical significance. The inherent variability of foodstuffs and how they are prepared and cooked are greater sources of error. For example: food tables may not represent the value of foods and drinks as consumed, particularly when dietary intake is recorded in terms of composite dishes and not as separate ingredients; the energy content of foods varies with the proportions of water and fat that they contain; and recipes used by subjects for dishes such as stews, casseroles, cakes, and puddings may differ considerably from the recipe used to calculate the food table value.

Measurements of Energy Intake

-Rationale Dietary assessments form a central component of many areas of clinical and research studies. Various national and international committees make recommendations with respect to the proportions of the dietary energy intake from different food sources that are considered to be best for optimal health. Data from large epidemiological surveys provide vital information for expert committees to consider when formulating allowances and deciding issues of food policy. Energy and nutrient intakes and requirements in different groups of people need to be quantified at an individual level to determine the relationships between nutrition and health. Intakes also need to be quantified at the population level to detect and establish the prevalence of people who are not achieving adequate levels of energy and nutrient intake. Some reported energy intake data from national surveys conducted in the UK and published in the last 10 years are shown in Table 1.

-Methods used to measure energy intake Measurements of energy intake in free-living individuals are obtained by dietary surveys. A record or estimate of all food and drink consumed is collected, and the energy content is then most usually estimated using data from food tables. In some experimental settings, where very accurate measurements are required, duplicate diets comprising foods actually consumed are collected by researchers and subjected to bomb calorimetry. However, under free-living conditions, the effort required of a subject to collect duplicate diets leads to alteration of food patterns and to an underestimate of food consumption.

There are a number of different survey techniques for estimating energy and nutrient intakes. The methods used to collect data about food consumption include weighed intake records, food diaries, collection of duplicate diets, diet history, 24-h recalls, and food-frequency questionnaries. Recalls and questionnaires may also include methods of estimating quantity by reference to models or photographs. The chosen method will depend on a number of factors such as resources available, number of subjects, foods and nutrients of interest, other parameters to be measured, the purpose of study, and subject characteristics such as age, behavioral, and health indices.

Table 1 Mean reported energy intakes of free-living children and adults in the UK

Age range (years)	*Males*		*Females*	
	MJ per day	*kcal per day*	*MJ per day*	*kcal per day*
1.5–2.5	4.39	1045	4.39	1045
2.5–3.5	4.88	1160	4.88	1160
3.5–4.5	5.36	1273	4.98	1183
4–6	6.39	1520	5.87	1397
7–10	7.47	1779	6.72	1600
11–14	8.28	1971	7.03	1672
15–18	9.60	2285	6.82	1624
16–24	10.33	2460	7.14	1700
25–34	10.25	2440	7.02	1670
35–49	10.50	2500	7.27	1730
50–64	9.96	2380	6.76	1610
65–74	8.21	1954	6.07	1445
75–84	7.75	1845	5.88	1400
85+	7.20	1713	5.77	1374

Sources: Gregory J, Foster K, Tyler H, and Wiseman M (1990) *The Dietary and Nutritional Survey of British Adults*. London: The Stationery Office, Gregory JR, Collins DL, Davies PSW, Hughes JM, and Clarke PC (1995) *National Diet and Nutrition Survey: Children Aged 1 to 4 Years. Volume 1: Report of the Diet and Nutrition Survey*. London: The Stationery Office, Gregory J, Lowe S, Bates CJ *et al*. (2000) *National Diet and Nutrition Survey: Young People Aged 4 to 18 Years. Volume 1: Report of the Diet and Nutrition Survey*. London: The Stationery Office, Finch S, Doyle W, Lowe C *et al*. (1998) *National Diet and Nutrition Survey: People Aged 65 Years and Over. Volume 1: Report of the Diet and Nutrition Survey*. London: The Stationery Office, with permission.

-Validity of energy-intake measurements It is often assumed that energy intake is the easiest and most straightforward component of the energy-balance equation to measure accurately. Whilst this may be so under strictly controlled experimental conditions, it is far more problematical in free-living subjects. Estimates of energy intake are subject to many errors, some of which have been well recognized and understood for many years. For example, food table values do not necessarily reflect the actual food as eaten (see above), inadequate descriptions of foods lead to errors in choosing appropriate foods from the food tables, and there may be errors in estimating the quantities of food and drink consumed. However, the magnitude and direction of other errors, and their implications, only began to be appreciated in the last 10 years. Up until then, the mean results for energy intake in studies of diet and health were rarely questioned. It was recognized that during a study, some individuals' reported energy intake would be lower than their habitual intake, but it was assumed that this was balanced by overestimates of intake at the other end of the distribution, thus yielding valid data for a group. Studies were based on the premise that the methods used were accurate and precise and that they provided valid estimates of habitual intake.

-Evaluation of energy-intake measurements The development in the early 1980s of the doubly labeled water (DLW) technique for measuring energy expenditure in free-living people also provided an external validation of measurements of energy intake. On a group basis, energy intake may be presumed to be equal to energy expenditure. Data from DLW measurements have, therefore, not only provided information of energy expenditure and energy requirements (see below) under many physiological, behavioral and experimentally imposed conditions, but also given important insights into the validity and interpretation of energy intake measurements. Of particular concern is that the majority of dietary surveys are significantly biased towards the underestimation of energy intake.

Invalid dietary data have serious implications for the interpretation of results. Apparent underestimates of intake will falsely inflate the prevalence of inadequate nutrient intake, and between-subject differences in qualitative or quantitative reporting creates bias that distorts the real associations between diet and health.

It is not feasible or necessary to carry out concurrent DLW measurements of energy expenditure when the prime interest of a study is the measurement of energy intake. However, the large body of DLW data yielded from many studies conducted over the last 20 years, together with statistical considerations, that take into account the number of subjects, duration of measurements, and variability in energy intake and expenditure, can be used as independent points of reference for evaluating measurements of energy intake.

-Developments in methodology of energy-intake measurements Over the past 10 years, a great deal of research has been undertaken by workers in many countries. The work includes the exploration of the statistical nature of food intakes in order to develop better study designs, and strategies for obtaining reliable and habitual-intake data. Energy intake data are now more closely scrutinized, and critically evaluated and interpreted. The subgroups most likely to provide records of poor validity are also being characterized. This is turn means that psychological assessments (dietary restraint, attitudes to food, health and other topics) as well as the more traditional elements (height, weight, assessment of occupation and leisure activities need to be incorporated into study protocols as a matter of routine. The origins and causes of misreporting due to food- and meal-specific bias, and the socioeconomic, demographic, behavioral, and psychological characteristics of subjects are all the focus of much research.

Energy Requirements

Definition of Energy Requirements

In 1985, the report of the Joint Food and Agriculture Organization, World Health Organization and United Nations University Expert Consultation stated that:

> The energy requirement of an individual is the level of energy intake from food that will balance energy expenditure when the individual has a body size and composition, and level of physical activity, consistent with long-term good health; and that will allow for the maintenance of economically necessary and socially desirable physical activity. In children and pregnant or lactating women, the energy requirement includes the energy needs associated with the deposition of tissues or the secretion of milk at rates consistent with good health.

Rationale for Defining Energy Requirements

Estimates of energy requirements are required as the basis of nutrition policy and dietary advice, to assess the adequacy of diets and evaluate nutritional status, to plan the food supplies of populations, and to plan diets for institutionalized individuals. In many countries, expert committees are charged with estimating the energy requirements of their own populations, usually divided into different age and sex groups.

The average energy requirements currently used in the UK are shown in Table 2.

Determinants of Energy Requirements

All body movements require energy in addition to that required for maintenance. Strenuous activities require more energy than sedentary activities, and the cost of weight-bearing activities increases with body weight. Whilst the gross energy costs of certain activities may be very high during that activity, e.g., during laboring tasks or very athletic recreations, these activities are rarely maintained for long periods of time. The energy costs are diluted by the time spent undertaking sedentary activities, e.g., sitting or standing around, and the average overall daily energy requirement is further reduced by the time spent asleep. The amount of physical activity due to differences in occupation, leisure pursuits, and domestic circumstances determines the overall energy expenditure, and therefore the energy requirements, of otherwise similar individuals.

In addition to the energy required for maintenance and for physical activity, the biochemical synthesis of new tissues for growth, development, repair, and milk production incurs extra energy costs, and therefore the energy requirements, for infants, children, pregnant women and lactating women, and in some clinical conditions. The energy requirements under these circumstances are difficult to quantify, and estimates are often derived from stoichiometric calculations based on rates and composition of weight gain, milk production and tissue repair, and the effects of infection, disease, trauma, and drug therapy on appetite and energy expenditure.

Table 2 Estimated average requirements for energy in the UK

Age range	*Males*		*Females*	
	MJ per day	*kcal per day*	*MJ per day*	*kcal per day*
0–3 months	2.28	545	2.16	515
4–6 months	2.89	690	2.69	645
7–9 months	3.44	825	3.20	765
10–12 months	3.85	920	3.61	865
1–3 years	5.15	1230	4.86	1165
4–6 years	7.16	1715	6.46	1545
7–10 years	8.24	1970	7.28	1740
11–14 years	9.27	2220	7.92	1845
15–18 years	11.51	2755	8.83	2110
19–50 years	10.60	2550	8.10	1940
51–59 years	10.60	2550	8.00	1900
60–64 years	9.93	2380	7.99	1900
65–74 years	9.71	2330	7.96	1900
75+ years	8.77	2100	7.61	1810

Source: Department of Health (1991) *Dietary Reference Values for Food Energy and Nutrients for the United Kingdom*. London: The Stationery Office.

Estimation of Energy Requirements

The 1985 FAO/WHO/UNU Expert Consultation heralded a significant change in the way that recommendations for dietary energy were determined and defined. It was considered that observed energy intakes, even if they were accurate, did not necessarily represent desirable levels for body weight, physical activity, and optimal health. It was concluded that estimates of requirements should 'as far as possible be based on estimates of actual or desirable energy expenditure.' The 1985 Committee also stated that 'Although, in principle, it would be desirable to determine the requirements of children in the same way as for adults, from measurements of energy expenditure, this approach involves many difficulties in practice.' It was concluded that the necessary data were not yet available to base recommendations for energy requirements of infants and children on estimates of energy expenditure, and so the use of information about normally growing infants and children has continued.

-Physical activity levels The 1985 FAO/WHO/UNU report also established the principle of calculating energy requirements for adults (although not pregnant and lactating women) as multiples of measured or predicted basal metabolic rate (BMR). Multiples of BMR are now widely referred to as physical activity levels (PALs). These are a succinct way of expressing the energy requirements of diverse groups of people, and the values provide a means of comparing activity levels between and within individuals. Because total energy expenditure is expressed as a multiple of BMR, most of the interindividual variability due to weight, height, age, and sex is removed.

-Estimates of physical activity levels using factorial calculations The PALs derived by the 1985 Consultation were based on literature values of the energy costs of various occupational and discretionary activities, and assumptions about their duration and intensity. A factorial calculation was then used to determine the theoretical overall 24-h energy requirement. For example, it was estimated that the PAL for a typical rural woman in a developing country was $1.76 \times$ BMR. This was based on the assumptions that within each 24-h period, 8 h were spent asleep ($0.33 \times 1.0 \times$ BMR); 3 h were spent on housework and preparing food ($0.13 \times 2.7 \times$ BMR); 4 h were spent working in fields ($0.17 \times 2.8 \times$ BMR); 2 h were spent on discretionary activities ($0.08 \times 2.5 \times$ BMR); and the remainder of the time (7 h), was spent on sedentary activities ($0.29 \times 1.4 \times$ BMR). A similar calculation indicated that the typical PAL of a man

employed as an office clerk was 1.54 × BMR. The survival requirement, with no allowance for energy needed to earn a living or prepare food, was considered to be 1.27 × BMR; a level of expenditure expected to be found only in totally inactive dependent persons and not compatible with long-term health (see also Table 8). The estimated average daily energy requirements of adults whose work is classified as light, moderate, or heavy, are given in Table 3.

The UK Committee on Dietary Reference Values followed the approach of the FAO/WHO/UNU panel and arrived at PALs considered to be more appropriate for UK populations. In countries like the UK, it is likely that the overall energy expenditure of an individual may be influenced more by recreational activities than by work. PAL values taking both occupational and nonoccupational activity levels into account are shown in Table 4.

-Measurements of physical activity levels from DLW measurements The development in the early 1980s of the DLW technique for measuring total energy expenditure (TEE) in free-living people revolutionized knowledge and assumptions about energy expenditure. The measurement of TEE over relatively long periods is likely, in the majority of people, to be representative of their habitual energy expenditure and, by inference, their energy requirements. TEE measured by DLW, used together with measured or predicted BMR, has allowed PALs to be measured, as well as estimated factorially. Studies conducted over the last 20 years mean that there is now a large dataset of DLW measurements of energy expenditure and PALs measured under many physiological, behavioral, and experimentally imposed conditions. Some of these data are summarized in Tables 5, 6 and 7. These tables illustrate the range of PALs depending on age, and on activity levels that are typical of populations in developed countries.

There are also data that have been obtained from individuals at the extremes of physical activity that can be used to define the upper and lower limits of human energy expenditure and, by inference, the limits of energy requirements and energy intake. Some of these data are summarized in Tables 8 and 9. It should be noted that at the upper level of energy expenditure, it is important to draw a distinction between the maximum achievable level over a limited period of time (e.g., intensive periods of sports) and the maximum sustainable as a long-term way of life given physical fitness and adequate food.

Table 3 Estimated average daily energy requirements of adults whose work is classified as light, moderate, or heavy, expressed as a multiple of BMR

	Light	*Moderate*	*Heavy*
Men	1.55	1.78	2.10
Women	1.56	1.64	1.82

Source: FAO/WHO/UNU (1985) *Report of a Joint Expert Consultation: Energy and Protein Requirements*. Technical Report Series 724. Geneva: WHO.

Table 4 Estimated average daily energy requirements of adults at three levels of occupational and nonoccupational activity, expressed as a multiple of BMR

Nonoccupational activity	*Occupational activity*					
	Light		*Moderate*		*Moderate/heavy*	
	Male	*Female*	*Male*	*Female*	*Male*	*Female*
Nonactive	1.4	1.4	1.6	1.5	1.7	1.5
Moderately active	1.5	1.5	1.7	1.6	1.8	1.6
Very active	1.6	1.6	1.8	1.7	1.9	1.7

Source: Department of Health (1991) *Dietary Reference Values for Food Energy and Nutrients for the United Kingdom*. London: The Stationery Office.

Table 5 Average total energy expenditure and PALs measured by doubly labelled water in different age and sex groups

Age range (years)	*Males*			*Females*		
	MJ per day	*kcal per day*	*PAL*	*MJ per day*	*kcal per day*	*PAL*
1–6	6.1	1452	1.64	5.5	1310	1.57
7–12	9.8	2333	1.74	8.0	1905	1.68
13–17	14.1	3357	1.75	11.4	2714	1.73
18–29	13.8	3286	1.85	10.4	2476	1.70
30–39	14.3	3405	1.77	10.0	2381	1.68
40–64	11.5	2738	1.64	9.8	2333	1.69
65–74	11.0	2619	1.61	8.6	2048	1.62
≥75	9.2	2190	1.54	6.1	1452	1.48

Source: Black AE, Coward WA, Cole TJ and Prentice AM (1996) Human energy expenditure in affluent societies: an analysis of 574 doubly-labelled water measurements. *European Journal of Clinical Nutrition* 50: 72–92.

Table 6 Doubly labelled water measurements of total energy expenditure in 2- to 7-year-old children

Age range (years)	*Males*		*Females*	
	MJ per day	*kcal per day*	*MJ per day*	*kcal per day*
2–3	4.50	1071	4.45	1059
3–4	5.22	1243	4.73	1126
4–5	5.50	1309	4.87	1159
5–6	6.18	1471	5.71	1359
6–7	6.74	1605	6.46	1538

Source: Black AE, Coward WA, Cole TJ and Prentice AM (1996) Human energy expenditure in affluent societies: an analysis of 574 doubly-labelled water measurements. *European Journal of Clinical Nutrition* 50: 72–92.

Table 7 Doubly labelled water measurements of total energy expenditure at very low levels of physical activity

Subjects	*Total energy expenditure*		*PAL*
	MJ per day	*kcal per day*	
Demented elderly females	5.2	1238	1.27
Handicapped male and female adolescents	6.1	1452	1.22
Obese males in calorimeter – no exercise	10.6	2524	1.17
Female athletes in calorimeter – no exercise	7.0	1667	1.21

Source: Black AE, Coward WA, Cole TJ and Prentice AM (1996) Human energy expenditure in affluent societies: an analysis of 574 doubly-labelled water measurements. *European Journal of Clinical Nutrition* 50: 72–92.

Table 8 Doubly labelled water measurements of total energy expenditure at very high levels of physical activity

Subjects	*Total energy expenditure*		*PAL*
	MJ per day	*kcal per day*	
Soldiers on field training on hilly terrain	14.4	3429	1.92
Soldiers on active service in cold conditions	17.8	4238	2.43
Soldiers training in the Arctic	17.8	4238	2.42
Soldiers training in the jungle	19.9	4738	2.63
Soldiers training in the snow	20.6	4905	2.79
Female swimmers tapering training	10.9	2595	1.75
Male swimmers tapering training	16.7	3976	2.08
Female runners in training	11.8	2809	2.03
Female mountaineers (Mt Everest)	12.0	2857	2.00
Male mountaineers (Mt Everest)	14.7	3500	2.44
Female endurance runners in training	12.3	2929	2.25
Female athletes in rigorous training	14.6	3476	2.79
Female Nordic skiers	18.3	4357	2.81
Male Nordic skiers	30.3	7214	3.47
Male Tour de France cyclists	33.7	8024	4.69
Male Arctic explorers	33.1	7881	4.47

Source: Black AE, Coward WA, Cole TJ and Prentice AM (1996) Human energy expenditure in affluent societies: an analysis of 574 doubly-labelled water measurements. *European Journal of Clinical Nutrition* 50: 72–92.

DLW data from men and women people living in rural areas in developing countries have shown that PALs for a very physically active lifestyle are about 2.5.

Low Levels of Energy Expenditure and Energy Requirements in Developed Countries

Many studies conducted in developed counties over the last 20–30 years show secular trends with respect to decreases in the levels of both energy intake and energy expenditure. The latter is due to decreases in the energy expended on physical activity because of labor-saving devices, increased use of mechanized transport for even short journeys, and changes in leisure-time activities. In many developed countries, therefore, energy requirements are falling. Low levels of physical activity are the reason why some expert committees set average energy requirements at a relatively low level. For example, in the UK, because of the inactive lifestyle of much of the population due to their light occupations and nonactive, nonoccupational activity, in the absence of other information, a PAL of only 1.4 is used to estimate energy requirements.

Implications of Low Energy Intakes and Energy Requirements

Current healthy eating guidelines in the UK, USA, and many other countries recommend that people eat a wide variety of foods. As energy requirements fall, so it becomes more difficult for a low energy diet to provide adequate intakes of macronutrients and micronutrients from all the major food groups. Low levels of energy expenditure and physical activity, or more correctly inactivity, have been identified as major risk factors for becoming overweight and obese. This applies to both developed, and increasingly to developing, countries. Low levels of physical activity are also associated with increased morbidity and mortality, and increased risk of diabetes, cardiovascular disease, and osteoporosis. Hence, there are now many health-promotion initiatives that integrate energy and nutrient intake messages with advice about physical activity.

Developments in Recommendations for Energy Requirements

Sections of the 1985 FAO/WHO/UNU report are currently undergoing expert revision by the International Dietary Energy Consultative Group. Data have been reexamined and updated. With respect to energy requirements, the topics under consideration by expert working groups include the equations used to predict BMR, physical activity levels in adults, and the energy requirements of infants, children and adolescents aged 1–18 years, pregnant and lactating women, and older individuals.

Table 9 Summary of physical activity levels (PALs) calculated from doubly labelled water measurements of total energy expenditure

Overall activity	*PAL*
Chair-bound or bed-bound	1.2
Seated work with no option of moving around and little or no strenuous leisure activity	1.4–1.5
Seated work with discretion and requirement to move around but with little or no strenuous leisure activity	1.6–1.7
Standing work (e.g., housewife, shop assistant)	1.8–1.9
Significant amounts of sport or strenuous leisure activity (30–60 min, four or five times per week)	+0.3
Strenuous work or very active leisure	2.0–2.4

Source: Black AE, Coward WA, Cole TJ and Prentice AM (1996) Human energy expenditure in affluent societies: an analysis of 574 doubly-labelled water measurements. *European Journal of Clinical Nutrition* 50: 72–92.

See also: **Carbohydrates**: Classification and Properties; Digestion, Absorption, and Metabolism; Requirements and Dietary Importance; **Children**: Nutritional Requirements; **Community Nutrition**; **Energy**: Measurement of Food Energy; Measurement of Energy Expenditure; Energy Expenditure and Energy Balance; **Food Composition Tables**; **Growth and Development**; **Infants**: Nutritional Requirements; **Metabolic Rate**; **Nutrition Policies in WHO European Member States**; **Pregnancy**: Metabolic Adaptations and Nutritional Requirements; **World Health Organization**

Further Reading

Black AE (1996) Physical activity levels from a meta-analysis of doubly labelled water studies for validating energy intake as measured by dietary assessment. *Nutrition Reviews* 54: 170–174.

Black AE (2000) Critical evaluation of energy intake using the Goldberg cut-off for energy intake: basal metabolic rate. A practical guide to its calculation, use and limitations. *International Journal of Obesity* 24: 1119–1130.

Black AE and Cole TJ (2000) Within- and between-subject variation in energy expenditure measured by the doubly-labelled water technique; implications for validating reported dietary energy intakes. *European Journal of Clinical Nutrition* 54: 386–394.

Black AE, Coward WA, Cole TJ and Prentice AM (1996) Human energy expenditure in affluent societies: an analysis of 574 doubly-labelled water measurements. *European Journal of Clinical Nutrition* 50: 72–92.

Black AE, Goldberg GR, Jebb SA *et al*. (1991) Critical evaluation of energy intake data using fundamental principles of energy physiology. 2. Evaluating the results of dietary surveys. *European Journal of Clinical Nutrition* 45: 583–599.

Black AE, Prentice AM, Goldberg GR *et al*. (1993) Measurements of total energy expenditure provide insights into the validity of dietary measurements of energy intake. *Journal of the American Dietetic Association* 93: 572–579.

Butte NF (1996) Energy requirements of infants. *European Journal of Clinical Nutrition* 50 (supplement 1): S24–S36.

Department of Health (1991) *Dietary Reference Values for Food Energy and Nutrients for the United Kingdom*. London: The Stationery Office.

FAO/WHO/UNU (1985) *Report of a Joint Expert Consultation: Energy and Protein Requirements*. Technical Report Series 724. Geneva: WHO.

Finch S, Doyle W, Lowe C *et al*. (1998) *National Diet and Nutrition Survey: People Aged 65 Years and Over. Volume 1: Report of the Diet and Nutrition Survey*. London: The Stationery Office.

Fogelholm M, Kolset SO, Rasmussen LB, Sjostrom M and Yngve A (2000) Physical activity – an essential part of dietary recommendations? *Scandinavian Journal of Nutrition* 44: 69–70.

Goldberg GR, Black AE, Jebb SA *et al*. (1991) Critical evaluation of energy intake data using fundamental principles of energy physiology. 1. Derivation of cut-off limits to identify under-recording. *European Journal of Clinical Nutrition* 45: 569–581.

Gregory JR, Collins DL, Davies PSW, Hughes JM and Clarke PC (1995) *National Diet and Nutrition Survey: Children Aged 1 to 4 Years. Volume 1: Report of the Diet and Nutrition Survey*. London: The Stationery Office.

Gregory J, Foster K, Tyler H and Wiseman M (1990) *The Dietary and Nutritional Survey of British Adults*. London: The Stationery Office.

Gregory J, Lowe S, Bates CJ *et al*. (2000) *National Diet and Nutrition Survey: Young People Aged 4 to 18 Years. Volume 1: Report of the Diet and Nutrition Survey*. London: The Stationery Office.

Macdiarmid J and Blundell J (1998) Assessing dietary intake: who, what and why of under-reporting. *Nutrition Research Reviews* 11: 231–253.

Prentice AM, Spaaij CJK, Goldberg GR *et al*. (1996) Energy requirements of pregnant and lactating women. *European Journal of Clinical Nutrition* 50 (supplement 1): S82–S111.

Price GM, Paul AA, Cole TJ and Wadsworth MEJ (1997) Characteristics of the low-energy reporters in a longitudinal national dietary survey. *British Journal of Nutrition* 77: 833–851.

Roberts SB (1996) Energy requirements of older individuals. *European Journal of Clinical Nutrition* 50 (supplement 1): S112–S118.

Schoeller DA (1990) How accurate is self-reported dietary energy intake? *Nutrition Reviews* 48: 373–379.

Shetty PS, Henry CJK, Black AE and Prentice AM (1996) Energy requirements of adults: an update on basal metabolic rates (BMRs) and physical activity levels (PALs).

European Journal of Clinical Nutrition 50 (supplement 1): S11–S23.

Starling RD and Poehlman ET (2000) Assessment of energy requirements in elderly populations. *European Journal of Clinical Nutrition* 54 (supplement 3): S104–S111.

Torun B, Davies P, Livingstone M *et al*. (1996) Energy requirements and dietary energy recommendations for children and adolescents 1 to 18 years old. *European Journal of Clinical Nutrition* 50 (supplement 1): S37–S81.

Measurement of Energy Expenditure

P R Murgatroyd, University of Cambridge, Cambridge, UK
W A Coward, MRC Human Nutrition Research, Cambridge, UK

Background

The energy that drives endogenous chemical processes and the performance of external mechanical work is released from the macronutrients fat, carbohydrate, and protein by oxidation and ultimately dissipated as heat. Energy expenditure may thus be measured either directly, by estimating the energy dissipated as heat, or indirectly, by measuring the rates of macronutrient oxidation. The associated measurement techniques are termed direct and indirect calorimetry, respectively. This article describes the ways in which measurements of energy expenditure may be made and discusses the factors influencing their accuracy and precision.

The most accurate and precise primary energy measurement techniques impose constraints on the individual being studied, but these may be turned to advantage in studies that require closely controlled conditions. Where an estimate of free-living energy expenditure is required, secondary indirect techniques, lacking the quality of the primary measurements but relying on them for verification or standardization, must be used. These include the estimation of carbon dioxide production from the elimination rates of the stable isotopes ^{2}H and ^{18}O, and the estimation of energy expenditure from measurements of heart rate or from records of activity.

Direct Calorimetry

Heat is dissipated from the body by *nonevaporative* routes (by radiation, conduction and convection) and *evaporative* routes (through respiratory and perspiratory water loss). These components are estimated, together with the heat generated in external work, by measuring the fluxes of heat from a *calorimeter*, a ventilated room that contains the subject. The nonevaporative heat flux is generally measured by one of two techniques, gradient layer or heat sink calorimetry. Evaporative heat flux measurements are made either by condensing the evaporated water or by measuring the water vapor added to the ventilating air.

Nonevaporative Heat-loss Measurement

-Gradient-layer calorimetry Gradient-layer calorimetry measures the flux of heat through the walls of the calorimeter. The walls are constructed of a thin, rigid material such as glass-reinforced epoxy resin. As heat produced by a subject flows out through the wall, a temperature gradient is established across the wall. This is measured by temperature-sensing layers distributed over the inner and outer wall surfaces. These layers may be resistance thermometers, formed by etching copper films bonded to each side of the wall material, or may be in the form of a thermopile with 'hot' junctions on one side of the wall and 'cold' junctions on the other. The relationship between the mean wall temperature gradient and heat flux is found empirically by calibration with known heat inputs to the calorimeter. To ensure that the heat flux measured arises only from the subject, the rate of change of the heat content of the calorimeter must be very much less than the heat flux from the subject. This is generally achieved by fixing the outside temperature of the wall with a water jacket, the water temperature being controlled to avoid rates of change of water temperature exceeding $0.001\,^{\circ}C\,min^{-1}$.

-Heat-sink calorimetry A heat-sink calorimeter measures the nonevaporative heat lost by a subject in terms of the rate at which heat must be removed from the calorimeter to prevent air-temperature rise and associated heat loss through the calorimeter walls. The heat is removed by recirculating the calorimeter air over a water cooled heat exchanger. Thermal insulation makes the temperature gradient across the walls very responsive to changes in heat production within the calorimeter. This gradient is sensed and used to control the temperature of water entering the heat exchanger so that it removes heat at a rate which minimizes heat losses through the walls and thus equals the heat flux from the subject. The heat flux is expressed as:

$$Q_e = \frac{\Delta t_e}{\Delta t_s} x Q_s$$

where Q_e is the heat extracted from the calorimeter, Δt_e is the rise in temperature of the water flowing through the heat exchanger, Δt_s is the rise in temperature of the same water passing through a standard heater, and Q_s is the heat input to the standard heater.

As with the gradient layer calorimeter, the rate of change of heat content of the calorimeter must be minimized, but due to the higher thermal insulation of the heat sink calorimeter's walls, this can be achieved by controlling the temperature of the air surrounding it to $\pm$ 0.25 °C, avoiding the need for a water jacket.

Evaporative Heat-loss Measurement

-Condensation measurements Ventilating air is conditioned prior to entering the calorimeter. It is first saturated, then cooled to a constant dew point temperature by a water-cooled heat exchanger, then reheated to the calorimeter temperature. Air leaving the calorimeter is passed over an identical heat exchanger to return it to its initial dew point temperature. The heat extracted from the outgoing air is the evaporative heat loss plus a nonevaporative component, which may be deduced either from measurements of ventilation air flow rate and temperature or by measuring the heat added to the air after the ingoing heat exchanger to bring it to the calorimeter temperature. Corrections must be made to the evaporative heat loss measurement for the difference between the latent heats of evaporation at body temperature and the condensation temperature.

-Vapor-phase measurements The rate of evaporative heat loss can be expressed as the product of the rate of vapor production and the latent heat of vaporization. Water-vapor production within the calorimeter is calculated from measures of in- and out-going vapor pressure, total (atmospheric) pressure, and the flow rate of the calorimeter ventilating air. The vapor pressure is best deduced from dewpoint temperature, which can be measured to $\pm$ 0.1 °C using commercial instruments.

Strengths and Limitations of Direct Calorimetry

Measurements of total heat production in direct calorimeters can be of very high quality, certainly yielding results reproducible to within $\pm$ 1% of calibration standards. Response times are also very rapid, particularly for gradient-layer systems using condensation techniques to measure the evaporative heat component. The accurate partition of heat loss into evaporative and nonevaporative components is difficult to demonstrate, as some of the latent heat of vaporization may be supplied by the environment rather than the subject. Direct calorimetry can be something of an engineering *tour de force*, which in part explains the move towards indirect calorimetry in recent years. There are also operational difficulties. Heat dissipated within the calorimeter from sources other than the subject must be accurately measured. These sources, which include radiant heat exchanges through windows, heat dissipated from meals and drinks before they are consumed, and heat lost from excreta, may contribute as much as 15% of the total measured heat. Direct calorimetry still has an important role in combination with indirect calorimetry in the study of heat balance, but this is currently a relatively minor area of research interest. For studies of energy regulation, indirect calorimetry has displaced direct calorimetry because of its overriding advantage of adding to the measurement of energy expenditure the measurement of the oxidation rates of the individual energy macronutrients.

Indirect Calorimetry

Indirect calorimetry measures energy expenditure by estimating the oxidation rates of the energy macronutrients, fat, carbohydrate, and protein, from rates of respiratory exchange of oxygen and carbon dioxide and from the excretion in urine of incompletely oxidized nitrogenous compounds. The respiratory exchanges may be measured in air directly respired through a mask or mouthpiece, in the air flowing through a ventilated canopy over the subject's head or in the air ventilating a small room in which the subject may live for some time. The measurements and calculations required are essentially the same for all these cases.

Gas-exchange-rate Measurements

The variables that must be measured to calculate oxygen and carbon dioxide exchange rates at standard temperature and pressure (STP) under dry conditions are: the ventilation rate of the subject or his enclosure, the concentrations of oxygen and carbon dioxide inspired and expired by the subject or entering and leaving the enclosure, and the moisture content, temperature, and pressure of the ventilating air at the point where its flow rate is measured. It is assumed that the only gaseous variables are oxygen, carbon dioxide, and water vapor, i.e., that nitrogen and the minor components of air are not metabolically active. Rates of gas exchange in directly respired air or in small ventilated canopies are calculated in terms of the product of the ventilation rate and the gas concentration changes generated by the subject as:

$$R_G = F_o \Psi_o f_{N_2o}(f_{Go}/f_{N_2o} - f_{Gi}/f_{N_2i}),$$

where the outgoing ventilation flow rate is measured and:

$$R_G = F_i\ \Psi_i\ f_{N_2i}(f_{Go}/f_{N_2o} - f_{Gi}/f_{N_2i}),$$

where ingoing ventilation flow rate is measured. F is the ventilation rate, Ψ is the correction to STP dry conditions, and f is the fractional concentration of a gas. The subscripts i and o indicate in- and outgoing samples. N_2 refers to nitrogen and the inert components of air, deduced by the difference from O_2 and CO_2 measurements.

When measurements are made in a room, this simple approach yields results that do not immediately reflect the gas exchanges of the subject. An additional term, the product of the calorimeter volume and the rate of change of gas concentration, is needed to account for the rate at which the volume of oxygen or carbon dioxide stored within the room air is changed by the subject. With the addition of this term, and with highly precise determination of oxygen and carbon dioxide concentrations, useful measurements of macronutrient oxidation rates and energy expenditure can be made over periods as short as 30 min in a room whose volume is 150 times greater than that of the subject within it.

Measurement of Macronutrient Oxidation Rates

Protein oxidation is an incomplete process yielding nitrogenous compounds, urea, uric acid, and ammonia, which are excreted in urine. The nitrogen in urine may be measured by several techniques, but the most commonly used procedure is the Kjeldahl method (*See* **Meat**: Analysis). Protein oxidation is calculated as 6.25 × nitrogen excretion rate in grams. The oxygen and carbon dioxide exchanges associated with protein oxidation can be deduced from the coefficients in Table 1. When these exchanges are subtracted from the total respiratory gas exchanges, the 'nonprotein' exchanges remain, and from these, the rates of fat and carbohydrate oxidation are calculated. The precisions achievable for fat and carbohydrate oxidation measurements made over 30 min within a room calorimeter are typically ± 0.4 and ± 0.7 g. (Precision is estimated as ± 1 standard deviation confidence limits from 20 repeated measures of simulated gas exchange and nitrogen analysis.) The accuracies are typically ± 0.2 and ± 0.42 g, respectively, for sedentary activity in an adult subject.

Figure 1 illustrates macronutrient flux measurements made by indirect calorimetry and the oxidation rates of fat and carbohydrate measured during alternating 30-min periods of exercise and rest. Measurements were made within a whole-body calorimeter room with an internal volume of 11 m^3.

Calculation of Energy Expenditure

Once the macronutrient oxidation rates are known, energy expenditure can be calculated from the summation of the products of the macronutrients oxidized and their energy densities (Table 1). The accuracy of energy expenditure measurements achieved in this way can be ± 1% or better. Protein oxidation often accounts for a small (12–15%) and relatively constant proportion of total energy expenditure, and so energy expenditure may be estimated as:

$$EE = 15.82 \times O_2 \text{ consumption} + 5.21 \times CO_2 \text{ production},$$

where EE is in kJ min^{-1}, and gas exchanges are in 1 min^{-1}.

Table 1 Macronutrient oxidation coefficients

	Carbohydrate	*Protein*	*Fat*
Energy equivalent of oxygen (kJ l^{-1})	21.12	19.48	19.61
Respiratory quotient	1.0	0.835	0.71
Energy density (kJ g^{-1})	15.76	18.56	39.33

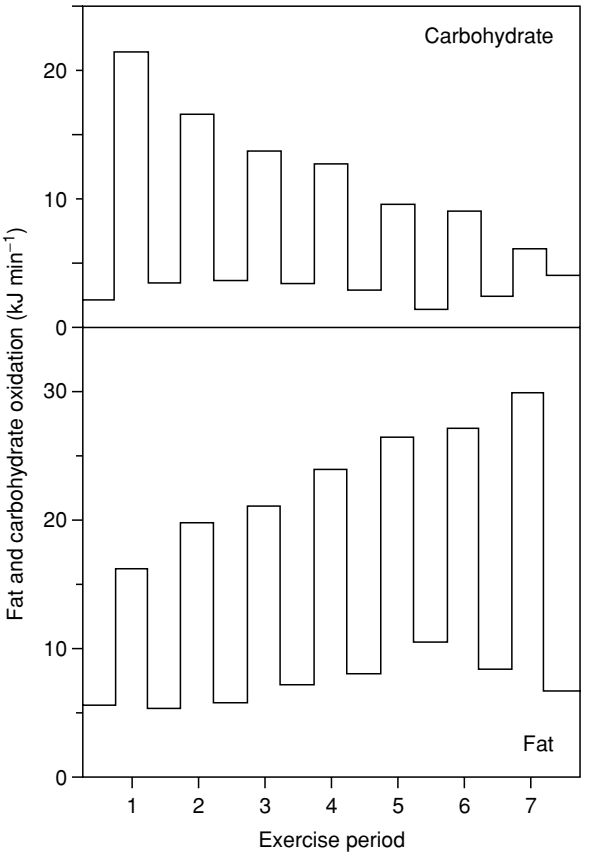

Figure 1 Oxidation rates of fat and carbohydrate measured by whole body indirect calorimetry during alternating 30-min periods of exercise and rest.

The respiratory quotient (RQ) is the ratio of carbon dioxide production to oxygen consumption and reflects the relative contributions of fat, carbohydrate, and protein to the oxidation fuel mixture (Table 1). The energy equivalent of 1 l of oxygen consumed by fat, carbohydrate, or protein oxidation is rather similar (Table 1). By assuming an RQ of 0.85 with 12.5% energy from protein oxidation, energy expenditure can be estimated from oxygen consumption alone as:

$$EE = 20.3 \times O_2 \text{ consumption}.$$

This remains accurate to within about $\pm$ 2.5% over the range of fat to carbohydrate oxidation ratios normally encountered.

Measurement of Energy Expenditure in Free-living Subjects

Doubly Labeled Water

-Principles of the method In the previous section, it was shown that energy expenditure can be estimated from oxygen-consumption measurements. Energy expenditure can also be expressed in terms of measurements of CO_2 production, and this forms the basis of the doubly labeled water method.

The energy equivalence of carbon dioxide, in contrast to that of oxygen, varies significantly with the mixture of macronutrients being oxidized. RQ could range from 0.7, if fat were the source of all energy, to 1.0, if all energy came from carbohydrate (see Table 1). In practice, sustained average RQ values outside the range 0.8–0.9 are rare, and for subjects close to energy balance, RQ may be inferred with greater confidence than this from knowledge of the subject's typical dietary composition. The energy equivalence of CO_2 for a range of RQ values is given in Table 2.

Carbon dioxide production may be measured, and hence energy expenditure estimated, from the difference between the disappearance rates of labeled hydrogen and oxygen in body water. The origin of the difference in isotope disappearance rates is the carbonic anhydrase catalyzed exchange of oxygen in water with oxygen in carbon dioxide (CO_2). Labeled oxygen can thus leave the body as either water or carbon dioxide, while hydrogen can leave only as water.

Table 2 Energy equivalence of carbon dioxide for different values of respiratory quotient

Respiratory quotient	$E_{eq}CO_2$
0.75	26.2
0.8	24.9
0.85	23.8
0.9	22.8
1.0	21.0

The volume of water in a body may be measured by the dilution of isotopes in the body water pool, a technique that has long been used for studies of body composition. In this procedure, small amounts of water labeled with stable isotopes of hydrogen or oxygen (^2H or ^{18}O) are administered, usually orally, and body-water volume calculated from the equation:

$$\text{Isotope dilution space} = \text{isotope given}/\text{isotope concentration found}.$$

This relationship assumes that isotope instantaneously distributes itself in body water in the same way that a marker dye would distribute itself in a flask of water. In practice, because body water is continually being lost and replenished by water intake, the isotope concentration falls in an exponential manner from its initial value. Because labeled oxygen can leave the body as either water or carbon dioxide, while hydrogen can leave only as water, each isotope has an individual logarithmic slope for its disappearance or rate constant (k_H and k_O) as may be seen in Figure 2. Furthermore, the isotope distribution spaces (V_H and V_O) are slightly different, principally because the hydrogen atoms in water exchange rapidly with those in proteins.

The difference between the rates of disappearance of labeled hydrogen and oxygen thus provides a measure of carbon dioxide excretion or production, r_{CO_2}, which may be expressed as:

$$r_{CO_2} = (k_O V_O - k_H V_H)/2 \text{ moles}.$$

The factor of 2 arises because 1 mole of CO_2 is equivalent to 2 moles of water.

To calculate CO_2 production rates, this basic equation is modified to allow for isotope fractionation when losses occur as water vapor or CO_2. There are several ways in which this can be done, but the simplest is to assume that water loss consists of a fractionated component (r_f) and an unfractionated component (r_i), and that r_f is related to r_{CO_2}. For adults in temperate climates, $r_f = 2.1\, r_{CO_2}$, in which case:

$$r_{CO_2} = \frac{k_O\ V_O - k_H\ V_H}{2f_3 + 2.1(f_2 - f_1)},$$

where f_1, f_2, and f_3 are fractionation factors for ^2H and ^{18}O in water vapor and ^{18}O in CO_2, respectively.

-The method in practice The procedure used in the application of this method is that the subject first provides a sample of their body water (e.g., a urine

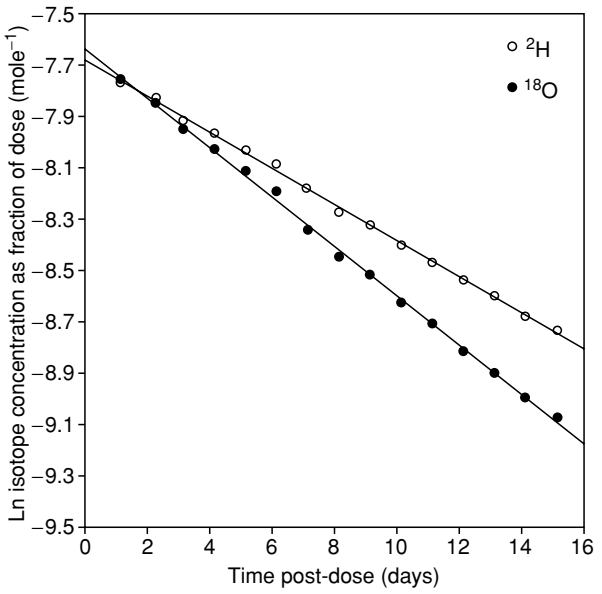

Figure 2 Disappearance of ^{2}H and ^{2}O isotopes from the body water pool in man.

sample) for the measurement of background isotope levels. The isotope dose is then drunk, and samples of urine produced over the next few days are retained for analysis. As few as two samples are required, one near the start of the experiment and one 2–3 biological half-lives of the isotopes later, though precision of the CO_2 production estimate can be improved by taking more samples during this period. The biological half-life of a tracer is the time taken for the isotope concentration to fall to half its original value. In man, this may vary from 3 days for lactating women in tropical countries to 14 days in elderly Western subjects. Average values for the accuracy of the method in comparison to CO_2 production measurements made during whole-body calorimetry studies in man range from -8.7 to $+1.9\%$, and in routine field use, the precision is better than 5%.

So far, the method has been mainly used to try to establish recommendations for energy intake of populations, to test the reliability of food intake measurements, and to assess the significance of energy expenditure in comparison to energy intake as causes of obesity.

Heart Rate

The responsiveness of heart rate to changes in activity level has led to many attempts to relate heart rate to energy expenditure and thereby to estimate free living energy expenditure. Generally, this procedure involves developing an individual relationship between a subject's energy expenditure, measured under laboratory conditions using indirect calorimetry, and simultaneously measured heart rate. The energy expenditure is modulated during this calibration process by varying the subject's activity from rest on a bed, through sitting and standing, to cycle or treadmill work over a range of intensities. The energy expenditure increases little as heart rate is increased by postural changes, but energy expenditure and heart rate both respond proportionally to the increasing intensity of dynamic exercise. A variety of techniques have been employed to describe the relationship between energy expenditure and heart rate. Many use a piece-wise linear approximation and attempt to define a point at which to make the transition from a 'rest' line to an 'exercise' line. A more elegant, though mathematically more complex, approach is to constrain a curve to fit the calibration data taking lines through the resting and exercising data as asymptotes.

The heart-rate method undoubtedly provides excellent qualitative information about energy expenditure and has been shown to be capable of predicting the mean 24-h energy expenditure of eight subjects, measured by whole body calorimetry, to better than 10%, though the accuracy of prediction for individuals varies widely. It is only with the recent development of the doubly labeled water technique that the validity of the extrapolation from laboratory calibration to field energy expenditure can be effectively tested, but heart-rate measurement is likely to remain a useful technique for comparing the relative energy expenditures of study groups when economics preclude the use of doubly labeled water.

Activity Recording

Energy expenditure may be estimated from the records in a diary of activities kept by the subject or an observer. The recordings are used to increment measurements or predictions of resting energy expenditure to provide energy expenditure estimates for each recorded activity, which can be summed to give the total energy expended over a period. As with heart rate, the estimates yield good qualitative results suitable for comparing the relative expenditures of groups.

See also: **Energy**: Energy Expenditure and Energy Balance; **Meat**: Analysis

Further Reading

Black AE, Coward WA, Cole TJ and Prentice AM (1996) Human energy expenditure in affluent societies: an analysis of 574 doubly-labeled water measurements. *European Journal of Clinical Nutrition* 50: 72–92.

Blaxter KL (1989) *Energy Metabolism in Animals and Man*. Cambridge: Cambridge University Press.

Cole TJ and Coward WA (1992) Precision and accuracy of doubly labeled water energy expenditure by multipoint and two-point methods. *American Journal of Physiology* 263: E965–E973.

Coward WA and Cole TJ (1991) The doubly labeled water method for the measurement of energy expenditure in humans: risks and benefits. In: Whitehead RG and Prentice A (eds) *New Techniques in Nutritional Research*, pp. 139–176. San Diego, CA: Academic Press.

Elia M and Livesey G (1992) Energy expenditure and fuel selection on biological systems: the theory and practice of calculations based on indirect calorimetry and tracer methods. In: Simopoulos AP (ed.) *Control of Eating, Energy Expenditure and the Bioenergetics of Obesity*, pp. 68–131. Basel: Karger.

Kalkwarf HJ, Haas DJ, Belko AZ, Roach RC and Roe DA (1989) Accuracy of heart-rate monitoring and activity diaries for estimating energy expenditure. *American Journal of Clinical Nutrition* 49: 37–43.

Kalkwarf HJ, Haas DJ, Belko AZ, Roach RC and Roe DA (1998) Human energy balance: what have we learned from the doubly labeled water method? *American Journal of Clinical Nutrition* 68: 409S–516S.

McLean JA and Tobin G (1987) *Animal and Human Calorimetry*. Cambridge: Cambridge University Press.

Murgatroyd PR, Shetty PS and Prentice AM (1993) Techniques for the measurement of human energy expenditure: a practical guide. *International Journal of Obesity* 17: 549–568.

Prentice AM (1990) *The Doubly-labeled Water Method for Measuring Energy Expenditure*. Vienna: International Atomic Energy Agency.

Energy Expenditure and Energy Balance

N G Norgan, Loughborough University, Loughborough, Leicestershire, UK

Introduction

The origins of energy expenditure lie within the metabolic activities of the body. To maintain an ordered state, as opposed to randomness and chaos, requires energy. To preserve the ionic gradients across cell membranes, and to allow the activity of cells, tissues, and organs as well as their control by enzymes, hormones, and nerves, requires energy. To maintain respiration and circulation, to grow or produce new tissue, to move and to work, all require energy. The energy expenditure results from these activities and functions. However, these are not amenable to individual measurement in the free-ranging human being. Using indirect calorimetry, it is possible to measure the energy expenditure at various levels of exertion – sleeping, lying, sitting, standing, walking, occupational and nonoccupational activities, and under various influences, such as eating, hot and cold environments, disease, pregnancy, and lactation. From the duration and energy cost, the total energy expenditure in each activity under specified conditions can be determined and the total daily energy expenditure (TDEE) calculated. An example is given in **Table 1**. This is a lengthy and laborious procedure and there are doubts about the precision and accuracy of the method. The advent of a new technique based on the disappearance of a dose of doubly labeled water has enabled a measurement of TDEE with some degree of confidence in most cases. Basal metabolism can be measured with relative ease and precision, and subtraction of the basal metabolism from TDEE gives the energy expenditure resulting from physical activity and thermogenesis. Although this model of the components of energy expenditure is less informative, it is felt to be more precise and accurate and so better for investigation of energy balance and imbalance.

Energy balance can be described by the energy balance equation:

$$\text{Energy intake (EI)} - \text{energy expenditure(EE)} = \text{change in energy stores } (\Delta ES)$$

Table 1 Typical time use, energy cost of activity (physical activity ratio: PAR), and energy expenditure in a group of young male office workers of median body weight

Activity	*Time (h)*	*PAR*	*Energy*	
			MJ	*kcal*
In bed	8.00	1.0	2.44	580
Office work				
Sitting	4.50	1.4	1.92	460
Standing	1.00	1.7	0.52	125
Sitting				
Eating	0.75	1.2	0.27	65
Driving	1.00	1.4	0.43	100
Watching TV	3.75	1.2	1.37	325
Miscellaneous	1.50	1.3	0.59	140
Standing				
Personal	1.00	2.0	0.61	145
Washing up	0.25	1.8	0.14	35
DIY	0.50	2.7	0.41	100
Car maintenance	0.50	2.7	0.41	100
Miscellaneous	0.50	2.1	0.32	75
Walking				
Slow pace	0.25	2.8	0.21	50
Normal pace	0.50	3.7	0.56	135
Daily total	24.00	PAL = 1.4	10.20	2435

PAR = (energy expenditure of an activity)/(basal metabolic rate, BMR).
Physical activity level (PAL) = (total daily energy expenditure)/(BMR).

Energy intakes in excess of energy expenditure will result in increased body energy stores and, conversely, when energy expenditure exceeds energy intake the deficit is drawn from the energy stores. As we eat intermittently but expend energy continuously, we regularly alternate between positive energy balance after meals and negative energy balance between or before meals. Energy balance must then be expressed over some time period, usually 24 h.

Positive energy balance is the norm over much of the life cycle as we grow and increase energy stores in childhood and adolescence, in pregnancy, and during adulthood. However, the extent of positive energy balance is quite small in most of us. The energy deposited in growth in the newborn may be one-third of the energy intake but by 6 months it is less than 5%. The fat gain of 10 kg experienced by many between the ages of 20 and 50 years requires only 0.001% of the energy intake.

The existence of energy balance, neither gaining nor losing energy, fat, and weight, is not by itself a sign of good energy nutritional status. The obese and the undernourished are often in a static state of neither gaining nor losing energy. It is only during the dynamic phase that the expected perturbations in energy balance are manifest.

The regulation of energy balance has been a subject of some controversy. Do we eat to meet the requirements for energy expenditure and energy gain, or do we modify expenditure according to the level of intake in order to maintain appropriate energy stores and energy gain? These are questions considered below.

Components of Total Daily Energy Expenditure

Basal Metabolism, Activity, and Thermogenesis

The basal metabolic rate (BMR) is defined as the energy expenditure at complete bodily rest, in a thermoneutral environment, 12–18 h after the last meal. These are conditions which rarely exist or persist in ordinary daily life. They are most commonly met on waking in the morning, when BMR is best measured. BMR amounts to about 4.2 kJ (1 kcal) kg^{-1} body weight h^{-1} in men and 3.8 kJ (0.9 kcal) kg^{-1} h^{-1} in women, i.e., about 4.2 kJ min^{-1}. BMR is not the lowest rate of energy expenditure of the day but its ease of measurement, relative precision, and the size of the existing database of information render it a suitable index for study. The resting metabolic rate (RMR) is measured with the less rigorous exclusion of the effects of previous eating and physical activity and is 3–5% higher than BMR. It shows a circadian (daily) rhythm of amplitude 5–10% with a peak at around 16.00–18.00 h and a trough in the early hours of the morning. Many women show a 5% variation over the menstrual cycle, being lowest 1 week before ovulation and highest 1 week after ovulation.

The energy expenditure at rest is determined to a large extent by the body weight and, in particular, by the lean body mass (*See* **Body Composition**). Men have higher resting energy expenditures than women do. The difference persists when energy expenditure is expressed per kg body weight because men tend to be leaner and each kg has more lean tissue that is more metabolically active than adipose (fat) tissue. Metabolic rate tends to fall with age too as lean body mass declines associated, in part, with a fall in level of activity.

Size is an important determinant of the energy expenditure in activities too, but to this must be added the extent to which the body weight is moved, horizontally and, in particular, vertically. The energy expenditure of walking and running can be estimated reliably from data on body weight, speed, incline, and the nature of the surface, e.g., tarmac, soft sand. Not surprisingly, walkers freely choose stride lengths and step frequencies associated with the lowest net energy cost of travel. Unfortunately, there are few other groups of activities where energy costs can be summarized by simple equations.

Comparisons of the energy expenditure of activities can be made by the physical activity ratio (PAR), i.e., the energy expenditure of an activity divided by BMR. This makes an approximate adjustment for age, sex, body weight and, in some cases, body composition so that values are more easily comparable in the different members of the population. The PAR of sitting quietly is about 1.2; for standing it is 1.7, but for walking at 5 km h^{-1} it is 3–4 and for walking at 5 km h^{-1} up a 25% incline it is 5 or more. The PARs of some everyday activities are shown in **Table 1**. These are values obtained during continuous activity but in everyday life we usually interpose pauses and breaks or a mixture of activities. If a task is hard, e.g., PAR 4–6, as in laboring and construction work, rest pauses are essential if the activity is to be maintained for any length of time. In the fraction of a second it takes to jump upwards, the rate of energy expenditure may be 25–50 times that at complete rest. However, we are only able to maintain activities of PAR 4–5 over the whole of a working day by introducing more rest and recovery pauses into the work pattern. (*See* **Exercise**: Metabolic Requirements).

The third component of energy expenditure has been called thermogenesis. This is an unfortunate term as most of the energy expended (strictly speaking, transduced) appears as heat anyway. Here, thermogenesis refers to the extra heat production

arising from energy expenditure associated with the ingestion and metabolism of food and with the extra heat required to maintain body temperature constant, by shivering or other mechanisms, as opposed to the waste heat which usually has this effect. Each of these components of thermogenesis have been further subdivided, in the case of diet-induced thermogenesis (DIT), into obligatory and facultative or adaptive DIT, while cold-induced thermogenesis has been subdivided into shivering and nonshivering thermogenesis (NST).

Obligatory DIT has several synonyms, e.g., thermic effect of food, heat increment of food. It represents the energy cost of digestion and absorption and the handling of absorbed nutrients, particularly the cost of synthesis. It is some 5–10% of dietary energy but may be higher for single food constituents such as protein. The thermic effect of food is normally included in measurements of energy expenditure and is not assessed separately. Facultative DIT and NST have been examined closely as possible components of energy expenditure that could be altered to allow energy balance to be maintained on widely different levels of energy intake. This is described more fully below. Extremes of climates may affect energy expenditure at rest and volitional physical activity. In everyday life in temperate parts of the world, behavioral and technological adjustments of and to clothing and control of the built environment ameliorate these climatic effects. Over the years, views about whether BMR are lower in the tropics, in particular the South Asia Indians, and raised in indigenous circumpolar populations for their body size and composition have tended to fluctuate more according to the zeal of the investigators rather than the quantity and quality of the evidence.

The TDEE is less variable than the daily energy intake. This is because whereas we may choose to eat a little or a lot, we can never expend no energy, and fatigue can reduce the extent to which we can raise expenditure. With our modern, sedentary lifestyles, resting energy expenditure makes up 70% of TDEE for most of us. Expressed in another way, physical activity levels (PAL = TDEE/BMR) for men and women in light occupations with nonactive leisure time are 1.4–1.5 (Table 1 and Figure 1). Values of up to 2.0 have been recorded in some occupations and 2.5–3.5 in athletes in training and competition, but to reach values of 4 it is necessary to work at the level of Tour de France cyclists. In mammals and birds, the maximum seems to be 7. The existence of a ceiling is a consequence of the need for increased mass of energy-supplying organs with high maintenance and operating costs in very active species matching the high levels of activity.

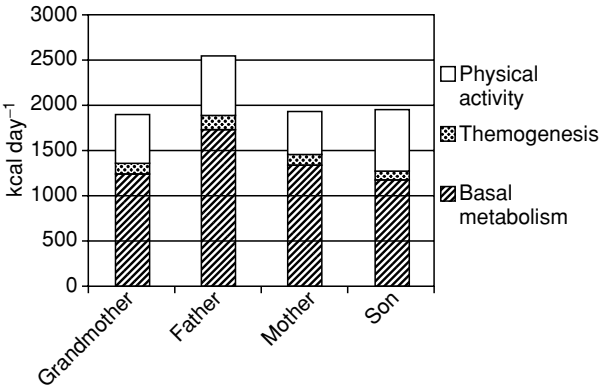

Figure 1 Total daily energy expenditure (TDEE) and its main components in a family: the grandmother is 70 years and weighs 65 kg, the father is 40 years and weighs 75 kg, the mother is 40 years and weighs 62 kg and the 10-year-old son weighs 30 kg. The physical activity levels (TDEE/basal metabolic rate) are 1.52, 1.47, 1.44, and 1.66 respectively.

High TDEE may be achieved by occupations or activities performed at a moderate level for many hours a day or by a shorter period of intensive activity. The physiological characteristics necessary for these two activity patterns are very different. Much higher levels of fitness and work capacity are needed for intense activity without excessive strain. Conversely, fitness improvement and health promotion require activity above a threshold intensity. More time at a lower level may not have the conditioning effect sought. For most of the population, however, exercise such as brisk walking may be of sufficient intensity. Common exercise regimes of three to four times per week for 30–60 min do not alter TDEE significantly. The benefits of exercise lie elsewhere.

Variation and Trends in Energy Expenditure

Individuals differ in their levels of energy expenditure. Much of this may be attributed to their size and composition (the proportions of actively metabolizing lean tissue and the less metabolically active adipose tissue). Men have higher rates than women do because of both size and composition differences. Different patterns of physical activity are the next most likely or important origin of differences in TDEE. Even with the same pattern of activity, individuals may select different rates, unless the activity is paced, as in treadmill walking.

TDEE increases throughout childhood and adolescence. In adulthood, TDEE declines where stable body weight and falling levels of activity are found. More commonly, increases in weight may balance falls in activity such that TDEE may not alter much

between 25–45 years. Thereafter, less child care and career progression, leading to desk-bound work and retirement, may further curtail activity. In addition, BMR falls 2% per decade as lean tissue mass decreases. In very old age, incapacity restricts the ability to be active and PAL levels may be only 1.2. Ill health at other parts of the life cycle may also affect activity and TDEE. In many parts of the world, infection not only causes personal misery but also seriously affects subsistence and occupational activity. Pregnancy has the twin characteristics of an increasing body size plus a customary fall in activity, particularly in the last trimester. Therefore the total extra cost of pregnancy may be less than expected, unless the woman does not reduce activity.

The contribution of genetic factors to differences in energy expenditure is very difficult to investigate and whether any genetic differences are actually expressed in the individual or population depends on a multitude of factors. Results from twin and other studies suggest that variations in RMR, thermic effect of food, level of physical activity, and energy cost of light exercise have a low to moderate genetic component. There are frequent reports in the literature of low BMR in tropical populations, particularly from the Indian subcontinent. Whether these arise from differences in size and composition and nutritional status, or ethnic differences *per se*, is controversial. Some groups have a renowned propensity for fatness, and a hypothesis of a thrifty genotype has been put forward to explain massive overweight and obesity in, for example, Polynesians and the Pima Indians of Arizona, USA, apparently as they altered their traditional lifestyle and took up western habits. The hypothesis suggests that, in response to chronic food shortages over many generations, individuals with more efficient energy utilization survived and bred selectively. Now that high energy intake can be achieved with low energy expenditure, the superior efficiency results in high energy stores. To support the hypothesis, evidence of a lower energy turnover or increased efficiency is required. In Pima Indians some families do have lower RMR than others and these show higher rates of obesity. But, a major problem in human energetics is how to compare the rates of energy expenditure in individuals of different sizes, whether the small, thin individuals in India with apparently low BMR or these large, obese individuals, where low BMR is also being sought. The problem is less difficult for physical activity as the net mechanical efficiency (NME = work done divided by energy expended) is an appropriate index of efficiency. Although NME varies between individuals for reasons that are not entirely clear, there is no consistent pattern of evidence suggesting that the body can alter NME in response to varying planes of nutrition in the short term or over many generations. This attractive hypothesis of a thrifty genotype spread throughout the population appears evolutionarily sound but remains a hypothesis.

Owing to the pervasiveness of modern communication, all societies and populations experienced changes in lifestyle in the 20th century. For many this has meant a shift from subsistence lifestyles, where all material needs were met by 2–3 days' work per week, to long hours of repetitive, often hard, poorly paid labor. In the developed economies, the trend has continued towards increased mechanization, a fall in the length of the working week, and increased holiday entitlements, which have reduced the effort in work, particularly the peak- and high-intensity efforts. The effect of this on TDEE is difficult to gauge. Surprisingly, analyses suggest there are no differences in levels of habitual physical activity, as evidenced by PALs, between the developed and the developing world. This might be explained, in part, by the curtailment of physical activity as a behavioral adaptation to the presence of chronic energy deficiency. It does highlight, however, the pitfalls of taking as self-evident beliefs about energy expenditure and physical activity. There have been few representative national studies of current-day activity and expenditure patterns, let alone studies for earlier periods of the last century. One way round this is to ignore the small effect of the greater energy stores in modern populations and to assume that energy intakes reflect energy expenditure. Again, there are no national representative intake data from earlier years but the evidence suggests that levels of energy expenditure have fallen, albeit only in the last 10–20 years.

Control of Energy Expenditure and Regulation of Energy Balance

There are several types of factors operating at different levels. At the primary level, the setting of regulatory centers in the brain or mechanisms at the cellular level may ultimately determine the level of energy balance. These respond to changes in energy intake and expenditure, at a secondary physiological, biochemical, or neurophysiological level. Traditionally, the regulation of energy balance has been seen as energy intake being controlled to match antecedent levels of energy expenditure. It has been argued recently that regulation of energy balance is less accurate when the energy intake is high in fat, as characterizes many contemporary diets, than for diets with low-fat energy and that this contributes to the modern rise in the prevalence of obesity. In contrast, attention has also been paid to the possibility of the

regulation of energy expenditure to maintain energy balance. In this case, energy expenditure rises when intakes are high and would otherwise lead to an increase in energy stores, and decreases when intakes are low and stores need to be conserved or used efficiently. A consistent effect of high energy intakes or low energy intakes is a change in body size. This tends to buffer the effects of the inequality by altering expenditure in the same direction as intake. However, this is more of an obligatory response in energy expenditure and less the facultative or adaptive response, mentioned earlier. The main mechanism of inefficient energy disposal in overeating has been suggested to be uncoupled oxidative phosphorylation in brown adipose tissue. The most recent suggestion is that the role of the facultative response in the regulation of energy balance may be secondary to its function in regulating the supply of essential nutrients in nutrient-deficient or unbalanced diets.

In undernutrition and during slimming the activity of the sympathetic nervous system, essential for brown adipose tissue thermogenesis, and protein synthesis – an energetically expensive process – may be reduced. However, in humans, by far the majority of the savings in energy expenditure with undernutrition occur owing to the smaller body size and reduced volitional physical activity. The effect of reductions of the metabolic rate of cells is less than 15% of the total reduction. Similarly, the magnitude of DIT of overfeeding at the rates seen in our societies and with our diets is almost certainly low and unlikely to exceed 10% of the extra energy ingested.

A multitude of tertiary-level factors – behavioral, social, and psychological – influence intake and expenditure and, ultimately, energy balance. These include the lifestyle factors referred to earlier, labor-saving devices, availability of private transport, shorter working hours, and inactive leisure. At the same time, the constant availability of food, its variety, and increased palatability and acceptance may easily override control mechanisms evolved for other diet types and nutritional circumstances. Median weights and weights for height or age are increasing in both Europe and North America, illustrating the increased incidence of positive energy balance. Reduced physical activity and increased sedentary behaviors, such as watching television, contribute to lower energy expenditure and are widely regarded as having a role in the increased prevalence of overweight and obesity, but exactly what, it is difficult to uncover. Longitudinal studies in the UK do not consistently show falls in physical activity over the last 10 years, particularly in nonmanual households. There are several dimensions to physical activity, which affect different segments of the population differently.

It has been suggested that to avoid obesity populations should be physically active throughout life at a PAL of 1.7–1.8. As present levels in adults may be nearer 1.55, this would involve the adoption of 20 min vigorous activity per day, such as running or team ball games, or 60 min per day of moderate activity such as brisk walking or cycling. These are clearly difficult to fit into most people's daily lives and a more appropriate strategy would be to perform everyday activities in a more energy-expensive manner, to walk instead of using fuelled transport, to do activities standing instead of sitting, etc. Recently, it has been suggested that fidgeting may play a more important role in energy expenditure than was previously thought. Training studies of the overweight but not obese suggest that increased physical activity plus dieting is associated with weight loss of only some 3 kg over 6 months. Women tend to be less successful than men in reducing body weight and fatness through exercise. It is not known if these differences have physiological or behavioral origins.

One tenet of health and fitness is that body weight should not increase in the adult years. However, data from cross-sectional studies of 7000 male runners, young and old, suggest that in modern society to achieve this requires increasing physical activity with age. Age-neutral body mass index (body weight allowing for height) is not compatible with age-neutral physical activity. Running distances needed to increase 2 km $week^{-1}$ for each increase of 1 year of age. Again, such increases are not easy to adopt and against the trend in the majority of even ardent runners and joggers.

One objection to physical activity as a means of achieving a more favorable energy balance that has been raised is that physical activity and increased energy expenditure cause energy intake to increase with no net effect on energy balance and body weight. The evidence suggests there often is compensation in energy intake but not beyond the amount of energy expended in exercise.

Although weight gain with age is commonplace, the increment in body energy represents only a tiny proportion of the energy throughput. Looked at in this way, regulation and control are still quite accurate.

See also: **Body Composition**; **Exercise**: Metabolic Requirements; **Metabolic Rate**

Further Reading

Ainsworth BE, Haskell WL, Leon AS *et al.* (1993) Compendium of physical activities: classification of energy costs of human physical activities. *Medicine and Science in Sports and Exercise* 25: 71–80.

Blaxter K (1989) *Energy Metabolism in Animals and Man*. Cambridge: Cambridge University Press.

Blaxter K and Waterlow JC (1985) *Nutritional Adaptation in Man*. London: John Libbey.

Bouchard C, Despres J-P and Tremblay A (1991) Genetics of obesity and human energy metabolism. *Proceedings of the Nutrition Society* 50: 139–147.

Department of Health (1991) *Dietary Reference Values for Food Energy and Nutrients for the United Kingdom*. *Report on Health and Social Subjects*, 41. London: HMSO.

Durnin JVGA and Passmore R (1967) *Energy, Work and Leisure*. London: Heinemann.

Ferro-Luzzi A and Martino L (1996) Obesity and physical activity. In: Chadwick DJ and Cardew G (eds) *The Origins and Consequences of Obesity*, pp. 207–266. Chichester: Wiley.

Fox K (1999) Aetiology of obesity X: physical inactivity. In: *Obesity: Report of the British Nutrition Foundation Task Force*, pp. 116–132.Oxford: Blackwell Science.

Garrow JS (1974) *Energy Balance and Obesity in Man*. Amsterdam: North-Holland.

McArdle WD, Katch FI and Katch VL (1996) *Exercise Physiology: Energy, Nutrition and Human Performance*. Baltimore, MD: Williams & Wilkins.

Norgan NG and Ferro-Luzzi A (1996) Human adaptation to energy undernutrition. In: Fregly MJ and Blatteis CM (eds) *Handbook of Physiology Section 4 Environmental Physiology*, pp. 1391–1409. New York: Oxford University Press.

Shetty PS (1999) Adaptation to low energy intakes: the responses and limits to low intakes in infants, children and adults. *European Journal of Clinical Nutrition* 53(suppl. 1): S14–S33.

Stock MJ (1999) Gluttony and thermogenesis revisited. *International Journal of Obesity* 23: 1105–1117.

van Raaij J (1998) Energy. In: Mann J and Truswell SS (eds) *Essentials of Human Nutrition*, pp. 73–90. Oxford: Oxford Medical Publications.

Williams PT (1997) Evidence for the incompatibility of age-neutral overweight and age-neutral physical activity standards from runners. *American Journal of Clinical Nutrition* 65: 1391–1396.

World Health Organization (1985) *Energy and Protein Requirements*. *Report of a Joint FAO/WHO/UNU Expert Consultation*. Geneva: World Health Organization.

Energy Expenditure *See* **Metabolic Rate**

ENERGY METABOLISM

S Rubinstein-Litwak, Clarksville, MD, USA

Introduction

All living organisms need energy to perform the most basic of natural functions, that is, they need energy to survive. Humans consume energy from the moment they are conceived to the moment they cease to exist, and considering that, according to the First Law of Thermodynamics, 'energy is not created or destroyed, just transformed,' where do humans obtain this much needed energy? The answer to this question lies within each cell of each individual organism.

The group of chemical and biological reactions involved in the process of obtaining and spending energy is known as energy metabolism. There are two major sources of free energy: food oxidation and light absorption.

The input of energy needs to be constant and must be converted into a highly accessible form to be used for three major purposes:

- performance of mechanical work in muscle contraction and other cellular movements;
- active transport of molecules and ions;
- synthesis of biomolecules that originate from simple precursors, including macromolecules.

Metabolism

Metabolism is defined as the series of physicochemical reactions that take place in cells with the purpose of synthesizing, converting, utilizing, and catabolizing substrates. The primary goal of metabolism is to form adenosine triphosphate (ATP), nicotinamide adenine dinucleotide (NADH) and building blocks for biosynthesis (i.e., amino acids, fatty acids, and sugars). It is through this constant generation and

consumption of metabolic substrates (i.e., ATP, NADH,) that living organisms transform free energy into channeled energy. Metabolism includes catabolic and anabolic pathways that are almost always distinct. During anabolic or degradative pathways, ATP is utilized, whereas during catabolic or biosynthetic pathways, ATP is generated. Most reactions of the human metabolism are performed in the presence of enzymes, which are protein molecules highly adapted and regulated that catalyze said reactions. Many reactions also require cofactors or coenzymes, most commonly vitamin molecules. The amount of enzymes, the catalytic activity of the enzymes, and the accessibility of substrates present at the time of the reactions regulate the metabolic process. The energy charge, defined as the energy status of the cell based on the relative amounts of ATP, adenosine diphosphate (ADP), and adenosine monophosphate (AMP), plays a central role in metabolic regulations. A low energy charge stimulates ATP-generating pathways, while inhibiting ATP-utilizing pathways. ATP is the only form of energy that can be used directly by the cells; energy released as heat is lost and cannot be utilized.

ATP–ADP Cycle

The ATP–ADP cycle is the basis for energy exchange in biological systems. ATP acts as the principal donor of energy and has a very high turnover. When energy is needed, ATP is hydrolyzed to ADP and orthophosphate (Pi). If more energy is needed, ADP can be further hydrolyzed to AMP and pyrophosphate (PPi). ATP can also be regenerated from ADP during oxidation of fuel molecules or when light is trapped (i.e., photosynthesis) (Figure 1).

Metabolism of Food Molecules

The energy in food becomes available only after the macromolecules contained in it (carbohydrates, fat, or protein) are transformed into less complex structures by undergoing a series of reactions. After adjustment for net absorption and urinary losses, the energy value of these substrates is estimated as 4 kcal g^{-1} for carbohydrates, 4 kcal g^{-1} for proteins and 9 kcal g^{-1} for fat (16.8 kJ g^{-1}, 16.8 kJ g^{-1}, and 37.8 kJ g^{-1}, respectively). Three stages in the generation of energy from food oxidation have been described (Figure 2).

- Stage 1: Large molecules in food are broken down into smaller units. No useful energy is generated in this step.
  - Proteins → Amino acids
  - Polysaccharides → Simple sugars (i.e., glucose)
  - Fats → Glycerol + fatty acids
- Stage 2: The recently formed smaller units are degraded once again, to yield a few simple units that

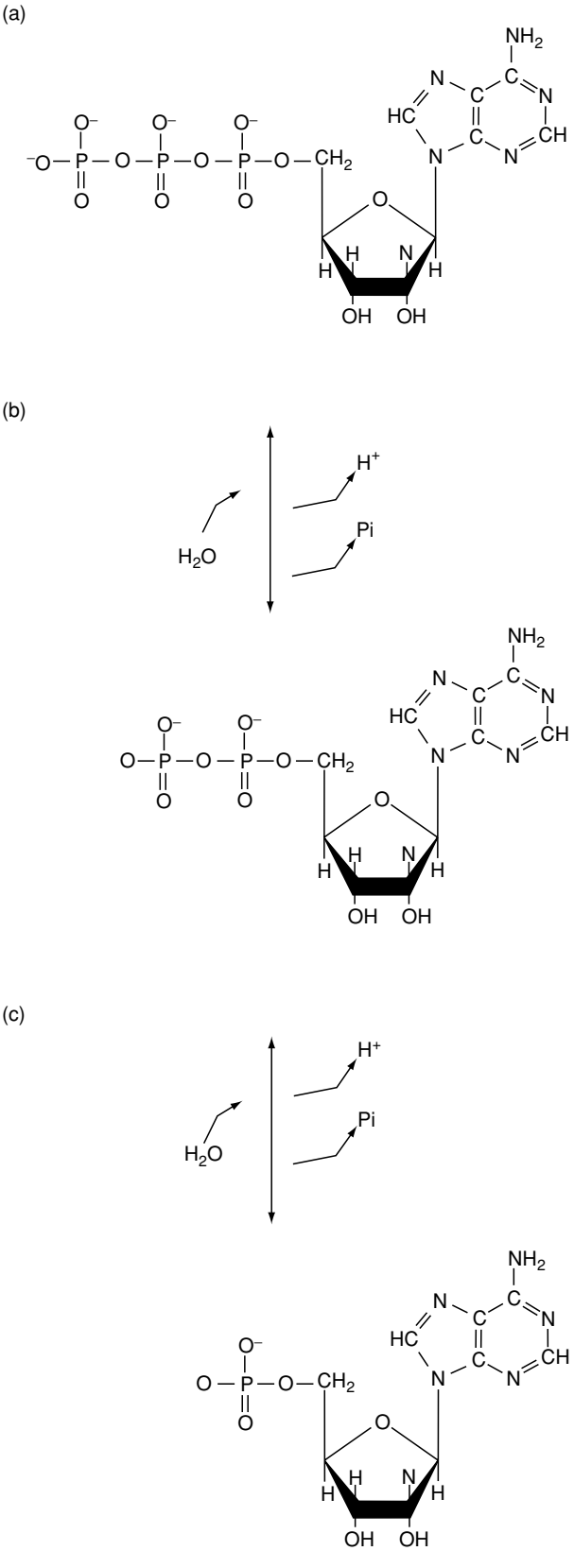

Figure 1 ATP metabolic path. (a) Adenosine triphosphate (ATP); (b) adenosine diphosphate (ADP); (c) adenosine monophosphate (AMP).

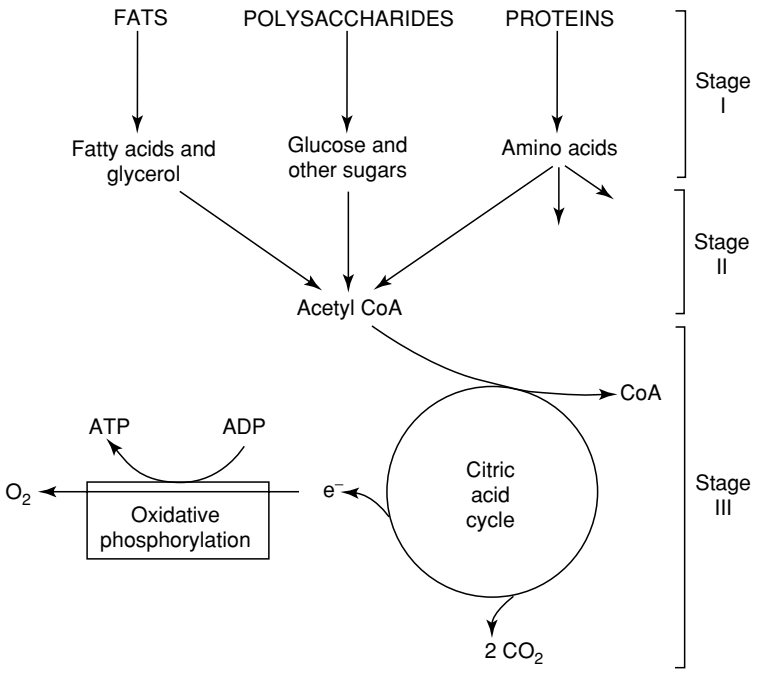

Figure 2 Stages in the extraction of energy from food. From Stryer L (1995) *Biochemistry,* 4th edn. New York: W.H. Freeman, with permission.

play a central role in metabolism. Most molecules are converted into acetyl coenzyme A (acetyl CoA). During this step, some ATP is generated.

- Stage 3: During this step, 90% of the ATP is generated. It includes the citric acid cycle and oxidative phosphorylation.

Carbohydrates

Carbohydrates serve as energy stores, fuels, and metabolic intermediates. When ATP is needed, polysaccharides can be rapidly mobilized to yield glucose, which acts as a prime fuel for the generation of ATP. Phosphorylated sugars are greatly involved in energy production by acting as intermediate substrates in the metabolic pathways. Glucose is the major form of carbohydrate substrate and the preferred energy source for the brain, the renal cortex, and red blood cells.

-Glycolysis Glycolysis includes a series of reactions that convert glucose into pyruvate with a small production of ATP. Reactions occur in the cytosol of the cell. It represents the initial steps that lead to the tricarboxylic acid cycle and the electron transport chain. The glycolytic pathway has a dual role. On one side, it degrades glucose to generate ATP, and on the other, it provides building blocks for the synthesis of cellular components. Glycolysis can be divided into three stages that are presented in Figure 3.

- In the first stage, the substrate is trapped in the cell to form a compound that can be readily cleaved into phosphorylated three-carbon units from where ATP is subsequently extracted. This consists of two phosphorylations and one isomeration to convert glucose into fructose 1,6-bisphosphate. In this step, no energy is extracted, and two molecules of ATP are used.
  - Glucose enters the cell through specific transport proteins. Once in the cell, it is phosphorylated by ATP to form glucose 6-phosphate.
  - Glucose 6-phosphate is isomerated into fructose 6-phosphate
  - Fructose 6-phosphate is phosphorylated by ATP, forming fructose 1,6-bisphosphate.
- In the second stage, some energy is generated. This involves the conversion of fructose 1,6-bisphosphate into 3-phosphoglycerate and a molecule of ATP.
  - Fructose 1,6-bisphosphate is split into glyceraldehyde 3-phosphate (which can be used in the direct pathway) and dihydroxyacetone phosphate. With the help of an isomerase enzyme, *triose phosphate isomerase*, the molecule of dihydroxyacetone phosphate can be converted into a second molecule of glyceraldehyde. The result is the formation of two molecules of glyceraldehyde 3-phosphate from one molecule of fructose 1,6-bisphosphate.
  - Glyceraldehyde 3-phosphate is converted into 1,3-bisphosphoglycerate (1,3-BPG), which is combined with ADP to generate 3-phoshoglycerate

plus ATP. This step represents the first ATP generating reaction in glycolisis.
- In the last stage, 3-phosphoglycerate is converted into pyruvate, and a second molecule of ATP is formed.
- The phosphoryl group in 3-phosphoglycerate is rearranged to form 2-phosphoglycerate.
- 2-Phosphoglycerate is dehydrated to form phosphoenolpyruvate.
- Phosphoenolpyruvate is converted into pyruvate. During this reaction, the phosphoryl group is transferred to ADP to form a molecule of ATP.

The net reaction in the transformation of glucose into pyruvate is:

$$\text{Glucose} + 2\,\text{Pi} + 2\,\text{ADP} + 2\,\text{NAD}^+ \rightarrow$$
$$2\,\text{pyruvate} + 2\,\text{ATP} + 2\,\text{NADH} + 2\,\text{H}^+ + 2\,\text{H}_2\text{O}.$$

Fat

Dietary fat is primarily distributed as triacylglycerols, which are highly concentrated stores of metabolic energy. To be able to use fat as energy, triacylglycerols need to be broken down to fatty acids and glycerol. This is achieved by hydrolysis with the help of a lipase, an enzyme that aids in the degradation of fat molecules. Oxidation of fatty acids takes place in mitochondria. Before entering the mitochondrial membrane, long-chain fatty acids in the outer membrane need to bind to coenzyme A to form acyl CoA. The activated fatty acids are carried across the inner membrane with the help of carnitine. Once in the mitochondria, Acyl CoA goes through a four-step process, called β-oxidation. First, acyl CoA goes through oxidation by flavin adenosine dinucleotide (FAD), then through a hydration, a second oxidation by nicotinamide adenine dinucleotide (NAD), and finally a thiolysis by CoA. The result is the reduction of the carbon chain of the fatty acid by two carbon atoms and the formation of acetyl CoA, NADH and $FADH_2$. Acetyl CoA then enters the tricarboxylic acid cycle by binding to oxaloacetate, while $FADH_2$ and NADH transfer their electrons to molecular oxygen by means of the electron transport chain.

The glycerol moiety that was liberated in the degradation of triacylglycerols can also be further metabolized by entering the glycolitic pathway as a three-carbon compound yielding eventually pyruvate.

Proteins

Proteins need to be degraded to amino acids so the body can use them. Most of the amino acids are used for the synthesis of new proteins, while the remaining amino acids are used for metabolic fuel. Amino acids can be used as the precursors for the synthesis of glucose, ketone bodies, and fatty acids. After the amino group is removed and converted to urea, the carbon chain of ketogenic amino acids is degraded to acetyl CoA or acetoacetyl CoA, to form ketone bodies. At the same time, glucogenic amino acids are degraded to pyruvate or one of the intermediates of the TCA cycle; α-ketoglutarate, succinyl CoA, fumarate, or oxaloacetate to synthesize glucose.

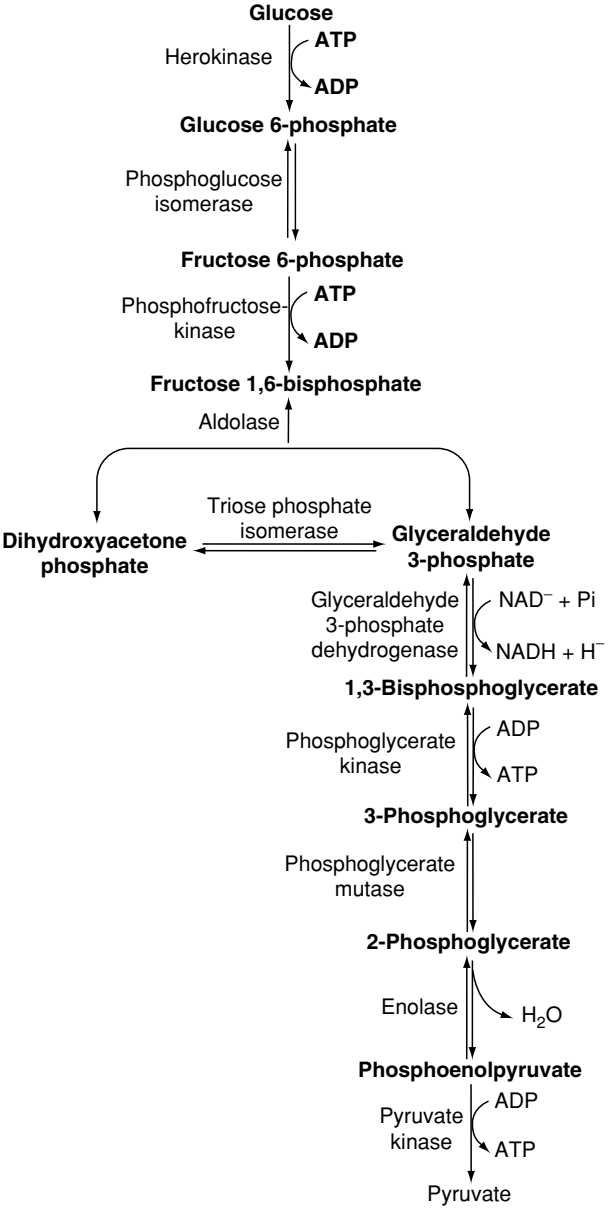

Figure 3 Glycolysis. From Stryer L (1995) *Biochemistry*, 4th edn. New York: W.H. Freeman, with permission.

Tricarboxylic Acid Cycle

The tricarboxylic acid (TCA) cycle is also known as the citric acid cycle or the Krebs cycle. It constitutes a series of reactions that involve the decarboxylation of pyruvate to form acetyl CoA, which in turn is fully oxidized to carbon dioxide (CO_2) and water (H_2O). These reactions take place in the mitochondria. The TCA cycle represents the final common pathway for the oxidation of fuel molecules.

The link between glycolysis and the TCA cycle is the oxidative decarboxylation of pyruvate to form acetyl CoA. Coenzyme A is a heat-stable cofactor that plays a central role within the molecule in metabolism. It can deliver activated acyl groups for degradation and energy generation as well as for biosynthetic purposes.

$$\text{Pyruvate} + \text{CoA} + \text{NAD}^+ \rightarrow \text{Acetyl CoA} + \text{CO}_2 + \text{NADH}.$$

The TCA cycle is graphically represented in Figure 4. The start of the cycle includes the union of oxaloacetate (a four-carbon unit) with the acetyl group of acetyl CoA (a two-carbon unit) to form citrate and CoA. Citrate is then isomerized into isocitrate, which undergoes the first oxidative decarboxylation of the cycle to become oxalosuccinate and then α-ketoglutarate. There then follows a second oxidative decarboxylation that yields a molecule of succinyl CoA. Succinyl CoA becomes attached to guanosine diphosphate (GDP), resulting in the formation of succinate, guanosine triphosphate (GTP) and CoA. This is the only step in the cycle that directly yields a high-energy phosphate bond.

Succinate is oxidized to fumarate, which in turn is hydrated to form malate. The final step in the cycle is the regeneration of oxaloacetate from malate. The electrons that are released during the cycle are transferred to NAD^+ and FAD to be transported in oxidative phosphorylation.

The net reaction of the TCA cycle is:

$$\text{Acetyl CoA} + 3\,\text{NAD}^+ + \text{FAD} + \text{GDP} + \text{Pi} + 2\,\text{H}_2\text{O} \rightarrow 2\,\text{CO}_2 + 3\,\text{NADH} + \text{FADH}_2 + \text{GTP} + 2\text{H}^+ + \text{CoA}$$

Oxidative Phosphorylation

Oxidative phosphorylation is also known as the electron transport chain. It includes the reactions that result in the synthesis of ATP from ADP + Pi. Heat can also be generated when ATP production is uncoupled from the respiratory chain.

Synthesis of ATP

Oxidative phosphorylation constitutes the major source of ATP in aerobic organisms. During the oxidation of fuel molecules, $FADH_2$ and NADH play a central role as the major electron carriers taking the electrons to the ultimate acceptor, oxygen (O_2). The transfer of electrons from NADH and $FADH_2$ to O_2 takes place in the inner membrane of the mitochondria. During this transfer, a number of protons are pumped out of the mitochondrial matrix, forming a proton gradient. Finally, with the help of ATPase, an ATP-synthesizing complex, the protons flow back into the matrix generating the necessary energy to generate ATP.

The electrons are transferred from NADH to O_2 through a chain of three large protein complexes. Each complex is an electron-driven proton pump that contains multiple oxidation–reduction centers. The electron carrying groups that take the electrons from NADH and $FADH_2$ to O_2 are flavins, iron–sulfur clusters, hemes and copper ions (Figure 5).

Complex I, also known as NADH-Q reductase, involves the first process through which the electrons from NADH are transferred. NADH binds to a flavoprotein, flavin mononucleotide (FMN), transferring its two high potential electrons and losing its hydrogen molecules. The result is $FMNH_2$ and NAD^+. From there, the electrons are transferred to coenzyme Q, also known as ubiquinone Q, on their way to complex III. Coenzyme Q is also the entry point for the electrons of $FADH_2$ to the chain. While some energy is lost as heat, ATP is also formed in a coupled reaction by bonding ADP + Pi. Complex III, also known as cytochrome reductase or cytochrome b, is the second proton pump in the electron transport chain. In this step, the electrons from coenzyme Q are transferred to cytochrome b and then to cytochrome c, a water-soluble protein, on their way to complex IV. During the transfer, some protons are once again pumped out of the mitochondrial membrane. A second coupled reaction results in the formation of ATP. Complex IV, also known as cytochrome oxidase or cytochrome a, plays a role in the transfer of the electrons from cytochrome c to O_2. In the final step, the electrons bind to O_2 to form H_2O.

The net reaction from each glucose molecule that enters glycolysis and proceeds through the TCA cycle and oxidative phosphorylation is:

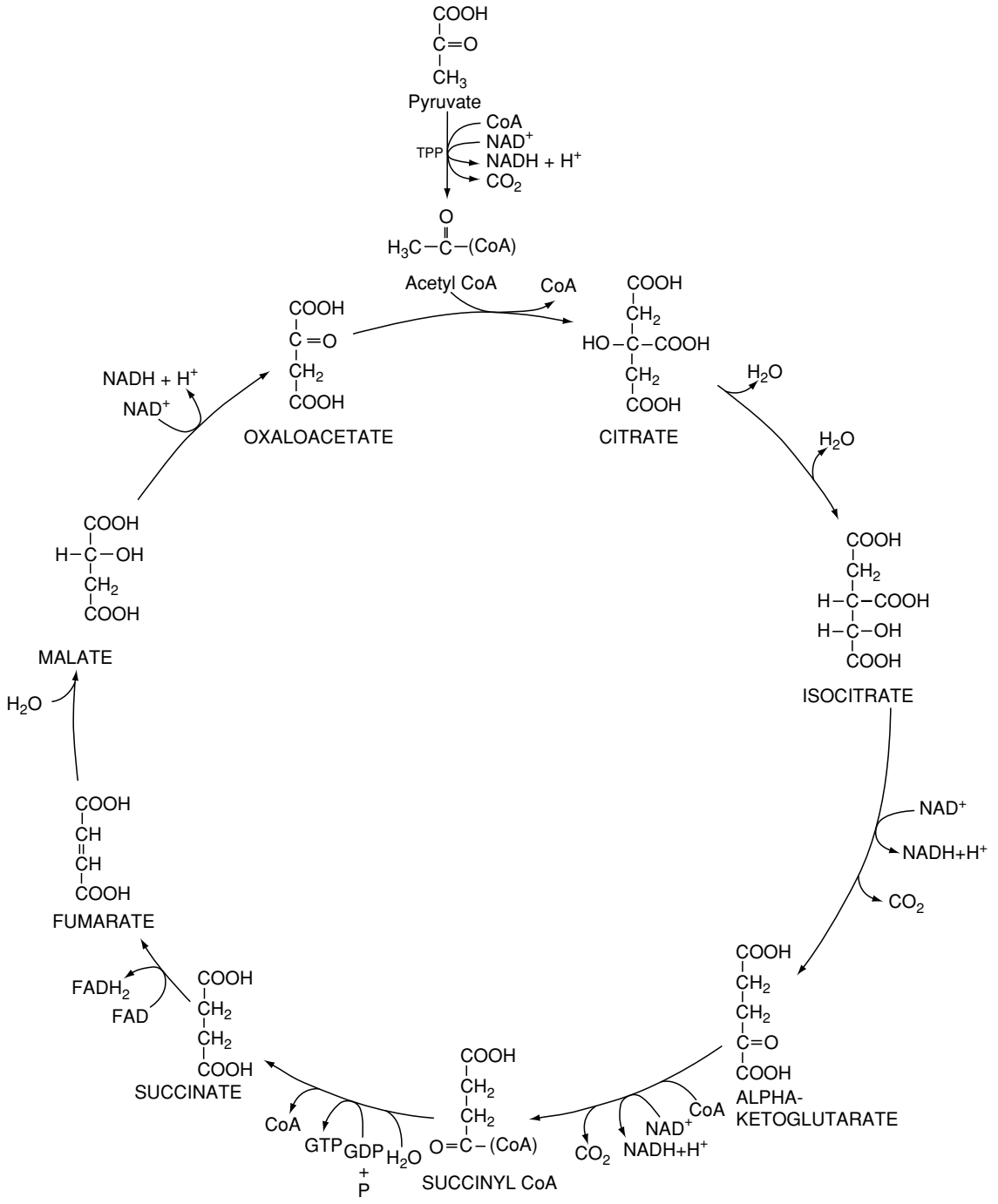

Figure 4 Tricarboxylic acid cycle.

$$\text{Glucose} + 6\,O_2 + 36\,\text{ADP} + 36\,\text{Pi} \rightarrow 6\,CO_2 + 36\,\text{ATP} + 6\,H_2O.$$

Generation of Heat

When there is not enough energy to transform ADP into ATP, mitochondrial electron transport is uncoupled from ATP synthesis by a proton leak. The result is the release of energy as heat. Uncoupling proteins (UCP) are known to be responsible for the leak that sometimes occurs during oxidative phosphorylation. Three uncoupling proteins have been identified. Thermogenin or UCP1 present in brown adipose tissue, UCP2 present in several tissues, and UCP3 mainly present in skeletal muscle. UCP

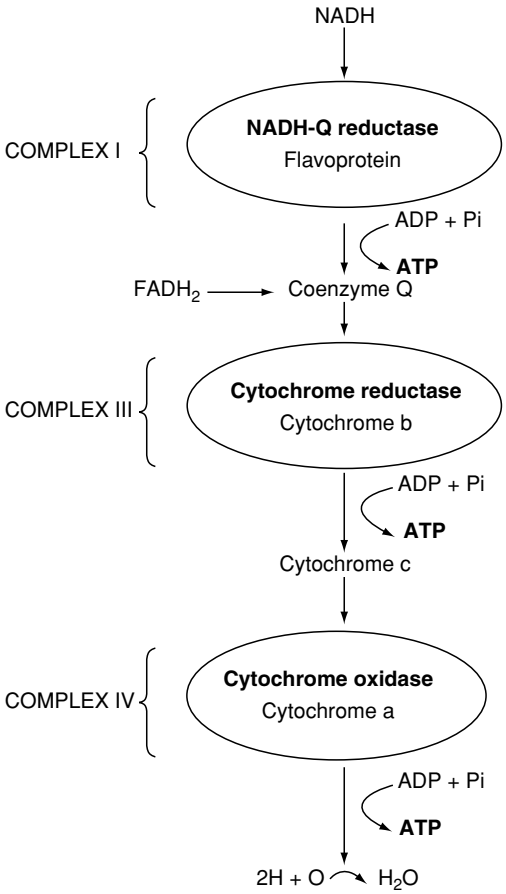

Figure 5 Oxidative phosphorylation.

play a role in energy metabolism by limiting the synthesis of ATP and dissipating energy as heat, thus reducing the efficiency of energy production.

Summary

As described in the previous paragraphs, humans obtain their energy from food. Once in the body, the macromolecules found in food undergo a series of degenerative reactions that yield the necessary substrates to promote the continuation of a process known as energy metabolism. The result of this very precise series of chemical and biological reactions is the transformation of energy in its free form into energy that can be fully utilized by the human body. Carbohydrates are degraded to simple sugars, fat to fatty acids and proteins to amino acids. Eventually, the complete oxidation of these molecules yields acetyl CoA, which acts as a link to the TCA cycle.

In the TCA cycle, acetyl CoA is fully oxidized to CO_2 and H_2O. Throughout the cycle, hydrogen atoms and their electrons are released. The hydrogen atoms are then bound to other compounds, resulting in the production of three molecules of NADH, one $FADH_2$, and one GTP. These new molecules contain the energy originally found in acetyl CoA. For the production of ATP, NADH and $FADH_2$ carry the electrons from the TCA cycle through an electron transport chain known as oxidative phosphorylation. The process ends with the production of ATP and H_2O. An alteration in oxidative phosphorylation can impair the production of ATP and dissipate energy as heat instead. Uncoupling proteins are responsible for this.

See also: **Amino Acids**: Metabolism; **Carbohydrates**: Digestion, Absorption, and Metabolism; Metabolism of Sugars; **Fatty Acids**: Metabolism; **Tricarboxylic Acid Cycle**

Further Reading

Lehninger AL (1975) *Biochemistry*, 2nd edn. New York: Worth Publishers.

Linder MC (1991) *Nutritional Biochemistry and Metabolism*. New York: Appleton & Lange.

Ricquier D and Bouillaud F (2000) The uncoupling protein homologues: UCP1, UCP2, UCP3, StUCP and AtUCP. *Biochemistry Journal* 345: 161–179.

Rubinstein S (1998) Energy metabolism. In: Sadler MJ Strain JJ and Caballero B (eds) *Encyclopedia of Human Nutrition*. pp. 682–687. London: Academic Press.

Schrauwen P, Walder K and Ravussin E (1999) Human uncoupling proteins and obesity. *Obesity Research* 7(1): 97–105.

Stryer L (1995) Metabolic energy: generation and storage. In: *Biochemistry*, 4th edn, pp. 443–652. New York: WH Freeman and Company.

ENTERAL NUTRITION

V A Stallings, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

Enteral Feeding

The nutritional status of hospitalized patients has received much interest since studies in the 1970s and 1980s indicated that up to 50% of surgical and medical pediatric and adult patients have some clinical or biochemical evidence of malnutrition. (*See* **Protein**: Deficiency.)

The four main groups of patients who generally require nutrition support are:

1. Patients with existing moderate-to-severe malnutrition indicated by weight loss in excess of 10%, muscle wasting, peripheral edema and/or albumin less than $30\,g\,l^{-1}$.
2. Patients with mild-to-moderate malnutrition who are at-risk of becoming further nutritionally compromised as a result of their medical condition.
3. Patients with a normal nutritional status but who, as a result of their medical condition, would not be able to commence normal feeding of adequate nutrient intake for a length of time appropriate for age and disease status, thereby potentially leading to a state of malnutrition.
4. Pediatric patients who are unable to maintain a normal pattern of growth and development, or who require nutrition support to promote an increased rate of growth (catch-up growth).

In order to prevent and treat malnutrition, the early identification of the 'at-risk' patient is very important. (*See* **Malnutrition**: The Problem of Malnutrition.)

Providing Enteral Nutrition Support

Oral Feeding

Patients who cannot safely consume solid food may take blenderized food. Patients who are able to take some amount of food but cannot maintain an adequate nutritional status may require oral dietary supplements. There are two main types of dietary supplements:

1. Modular, i.e., single nutrient supplements such as protein, fat, carbohydrates, vitamins, or minerals. There are several formulations of each type of nutrient. For example, protein may be provided as intact proteins, protein hydrolysates, or individual amino acids, and fats as long-chain fats, including essential fatty acids and medium-chain triglycerides.
2. Complete, i.e., balanced nutrient supplements which contain all the nutrients and may be taken as the whole meal, part of a meal, or a snack.

There are a vast number of liquid and semisolid commercial nutritional products that may be used for enteral feeding. These products may be designed for a specific age group (i.e., infant, adult) or specific medical condition (i.e., premature infant, pulmonary failure, renal failure).

Enteral Tube Feeding

Enteral tube feeding is the preferred route of diet administration for all patients with a functioning gastrointestinal tract who, for a variety of reasons, do not take adequate nutrition from food and/or oral dietary supplements. In recent years, research and clinical experience has led to many technical advances in the formulation and presentation of enteral formulas and feeding equipment. As a result, technology-assisted enteral support is now widely used. The most commonly used routes for providing enteral nutrition are via nasogastric and gastrostomy feeding tubes.

-Nasogastric feeding Fine-bore feeding tubes are the tubes of choice as they are narrow, softer, flexible, and more comfortable for the patient than the previously used large-bore nasogastric tubes. The use of fine-bore feeding tubes has reduced many of the common nasal tube complications such as rhinitis, esophagitis, pharyngitis, esophageal strictures, and erosions.

Fine-bore feeding tubes are inserted per nasally with a wire stylet to facilitate passage. The correct position of the tube may be confirmed by aspiration of the gastric contents and testing for acidic pH, or by auscultation of the epigastrium as air is injected into the stomach via the feeding tube. Radiological confirmation of the correct position of the nasogastric tube is required in patients who have an impaired gag reflex, those who are semiconscious or unconscious, and other high-risk patients including infants.

Nasogastric feeding may not be suitable for patients who are at-risk of regurgitation and/or pulmonary aspiration of the feed, and nasoduodenal or nasojejunal feeding tubes should be considered for these patients. Enteral feeding tubes with small weighted tips (3–5 g) are available to facilitate tube passage postpylorically for feeding directly into the

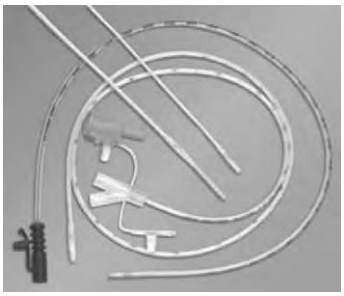

Figure 1 Fine-bore nasoenteral tubes (courtesy of Corpak Medsystems, USA).

duodenum or the jejunum. Figure 1 shows weighted fine-bore nasoenteral tubes (Corpak Medsystems, USA) and two types of unweighted nasogastric feeding tubes. In some patients, endoscopic placement of the feeding tube may be undertaken to ensure rapid and correct placement.

-Gastrostomy feeding Patients who require long-term or permanent enteral feeding should be considered for gastrostomy feeding tube or gastrostomy button (a smaller appliance which is almost flush with the skin) placement. The technique of percutaneous endoscopic gastrostomy (PEG) is a well-established procedure for providing enteral nutrition support. Although the feeding gastrostomy may be placed surgically, PEGs compare favorably, with a lower morbidity and cost; the surgical gastrostomy requires a laparotomy and a general anesthetic, whereas the PEG is performed under local anesthesia using a fiber-optic endoscope. A range of PEG kits are available for infants, children, and adults. A feeding jejunostomy may be considered as the means of providing early postoperative nutrition support in selected patients and may be used for short- or long-term support.

Techniques of Administration of Enteral Tube Feeding

Enteral formulas are usually given from a reservoir (plastic bag) via appropriate tubing to connect to the feeding tube (nasogastric, gastrostomy, jejunostomy). The formula is infused either by gravity or by mechanical peristaltic pump to ensure a selected and steady delivery rate. The enteral feeding administration set should be changed at least every 12 h to minimize the risk of infectious complications. The method of continuous feeding by a pump saves nursing time and ensures that the prescribed amount and rate of formula are provided. Bolus feeding is considered more physiological, but this method of enteral nutrition has a greater incidence of side-effects, such as regurgitation, distention, and diarrhea.

Intermittent enteral tube feeding during the day or the administration of continuous overnight (8–12 h) infusion can be of benefit to patients who may be able to take small quantities of food. This is common practice when patients are being transitioned for total intravenous nutrition support for conditions such as short-bowel syndrome.

Enteral Formulas Tube Feeding

In the past, regular hospital food was blenderized, but these preparations were too viscous for most feeding tubes. Further dilutions lowered the nutrient density, and the resulting blended feed was often nutritionally inadequate. The problems of ensuring nutritional quality while avoiding microbial contamination of these feeds, along with the time taken for their preparation, has led to the increased use of commercially prepared 'ready-to-feed' enteral formulas in most hospital and home care settings.

Types of Enteral Formulas

The four main groups of commercially available enteral diets are polymeric, predigested, disease-specific, and modular. The selection of the most appropriate enteral formulas requires consideration of the patient's age, nutritional status, disease process, and the specific nutritional requirements.

-Polymeric formulas Polymeric formulas are indicated for the majority of patients with normal or near-normal gastrointestinal function. The dietary constituents in polymeric enteral formula generally are as follows: *energy* – carbohydrate as glucose polymers, fat as long-chain triglycerides; *nitrogen* – whole protein; *vitamins*; *minerals*; and *trace elements*. (*See* **Carbohydrates**: Requirements and Dietary Importance; **Energy**: Measurement of Food Energy; **Fats**: Requirements; **Minerals – Dietary Importance**; **Protein**: Requirements; **Vitamins**: Overview.)

Most standard polymeric diets have an energy density of 4.2 kJ (1 kcal) ml^{-1} and a nitrogen concentration of 5–7 g l^{-1}. Some patients may require the energy and nitrogen-dense formulations which contain 8–10 g of nitrogen per liter and possess an energy density of 6.3 kJ (1.5 kcal) ml^{-1} or more. Polymeric diets containing dietary fiber are available and may promote the regulation of bowel function in some enterally fed patients. The use of enteral formulas containing fiber has been suggested in the elderly, those who require long-term nutritional support, and patients with neurological conditions and enteral

inactivity. (*See* **Dietary Fiber**: Physiological Effects; Effects of Fiber on Absorption.)

-Predigested, chemically defined formulas Predigested, chemically defined formulas are also called elemental formulas or diets. These preparations are used in the nutritional management of patients with impaired gastrointestinal function, such as severe impairment of luminal nutrient hydrolysis or substantially reduced functional absorptive capacity (e.g., short-bowel syndrome, exocrine pancreatic insufficiency). Predigested diets contain carbohydrate in the form of glucose polymers, peptides and amino acids as the nitrogen source, fat, usually as a mixture of medium-chain and long-chain triglycerides, and the full complement of vitamins and minerals.

-Disease-specific formulas Specific dietary formulations have been developed for patients with conditions such as liver disease, renal disease, or respiratory disease.

Liver disease Patients with encephalopathy and cirrhosis have abnormal plasma amino acid profiles characterized by raised levels of aromatic amino acids (i.e., tyrosine and phenylalanine) and depressed levels of branched-chain amino acids (i.e., leucine, valine, and isoleucine). These changes in the pattern of amino acid levels have been implicated in the pathogenesis of hepatic encephalopathy. Enteral feeding preparations containing low levels of aromatic amino acids and higher levels of branched-chain amino acids are available for oral or tube feeding: the administration of these diets may be of benefit to patients with encephalopathy associated with acute or chronic liver disease. (*See* **Liver**: Nutritional Management of Liver and Biliary Disorders.)

Renal disease Formulas containing all of the essential amino acids plus histidine, without the nonessential amino acids, have been developed for patients with impaired renal function. It is suggested that these diets reduce urea production and promote the reutilization of nonessential amino acids in the liver by transamination. At present, there is little evidence to support their chronic use to delay progression of chronic renal disease.

Respiratory disease The level of energy from carbohydrate in enteral formulas can result in further increased production of carbon dioxide and respiratory distress in some patients. In order to reduce the production of carbon dioxide, the use of enteral formulas which have a higher fat:carbohydrate ratio has been suggested for patients with respiratory disease. These formulas may facilitate more rapid weaning from ventilatory support.

-Modular diets Modular diets allow specific components of a diet to be altered according to the individual patient requirement. These include products that provide various types of isolated carbohydrate, protein, and fat compounds.

Initiation and Contraindications of Enteral Feeding

Some care teams administer small volumes of diluted enteral feeds during the initial stages of nutrition support in order to minimize any gastrointestinal side-effects. Some studies report that the use of dilute 'starter regimens' does not decrease the frequency of gastrointestinal side-effects but may result in reduced energy and nutrient intake and thus may adversely affect the nitrogen balance. Enteral nutrition is contraindicated in patients with an inaccessible or nonfunctioning gastrointestinal tract (e.g., paralytic ileus, intestinal obstruction).

Complications of Enteral Nutrition

Malposition of enteral feeding tubes may occur, but this can usually be avoided by strict adherence to the protocol for the insertion of feeding tubes, and confirmation of the position of the tube. Unintentional removal of feeding tubes is commonly noted in agitated, disoriented patients. Blockage of tubes can occur unless they are flushed regularly.

Diarrhea is the most commonly reported side-effect, and the factors implicated are high osmotic loads, concomitant antibiotic therapy, and excessively rapid infusion. Problems of lactose intolerance are limited, since the majority of commercial enteral formulas are low in lactose or lactose-free. Tube-feeding-related diarrhea occurs in up to 30% of patient, but its etiology is unclear. Nausea, vomiting, bloating, cramps, and abdominal distension can occur in some patients, especially following bolus feeding or when the feed is infused at a high flow rate. Pulmonary aspiration and regurgitation may be reduced by elevating the head of the patient's bed. Duodenal (transpyloric) feeding may be beneficial in patients at risk for aspiration.

Abnormalities of liver function tests are generally minor and reversible with an enteral feeding regime. Regular monitoring of patients enables early detection of biochemical imbalances, including electrolyte and fluid balance abnormalities. Drug–nutrient interactions can occur with many medications, including theophylline, warfarin, methyldopa, and digoxin.

Monitoring Patients Receiving Enteral Nutrition

Monitoring enterally fed patients is an essential aspect of providing safe and effective nutrition support.

Fluid Balance Charts and Weight

Continuous fluid-balance records enable an accurate review of the patient's actual intake of enteral formula compared to the prescribed regimen, as well as other fluid intake. The prescribed regimen and the actual intake may not be the same, and without records, patients may receive suboptimal (inadequate or excessive) levels of nutrition support. All patients receiving enteral nutrition support should be weighed regularly. The frequency of the weight assessment depends upon the patient's age, nutritional status, and medical condition.

Biochemical and Hematological

Plasma glucose, electrolytes, calcium, phosphate, and magnesium should be monitored regularly, and more frequently in severely malnourished patients. Plasma proteins are useful markers of nutritional status, transferrin and prealbumin in the short-term, and albumin in the long-term. Hemoglobin and the white blood cell and lymphocyte count should be monitored regularly and often. (*See* **Nutritional Assessment**: Biochemical Tests for Vitamins and Minerals.)

Anthropometric and Dynamometric

Measurements of mid-arm circumference and triceps and subscapular skinfolds should be taken regularly and are commonly used for research purposes. (*See* **Nutritional Assessment**: Anthropometry and Clinical Examination.)

Ending Enteral Tube Feeding

As the patient is able to increase oral intake of regular food, the administration of the enteral tube feed is decreased. Oral and enteral tube intake should be monitored using food record charts to document the nutritional quality of the total intake. Enteral feeding should not be discontinued until adequate nutrition can be taken from food and/or oral dietary supplements, and normal hydration and weight maintained.

Nutrition Team

Team management of enterally fed patients reduces associated morbidity and optimizes nutrition support. This interdisciplinary team usually includes a dietitian/nutritionist, nurse, and physician.

See also: **Carbohydrates**: Requirements and Dietary Importance; **Dietary Fiber**: Physiological Effects; Effects of Fiber on Absorption; **Energy**: Measurement of Food Energy; **Fats**: Requirements; **Liver**: Nutritional Management of Liver and Biliary Disorders; **Malnutrition**: The Problem of Malnutrition; **Minerals – Dietary Importance**; **Nutritional Assessment**: Anthropometry and Clinical Examination; Biochemical Tests for Vitamins and Minerals; **Protein**: Requirements; Deficiency; **Vitamins**: Overview

Further Reading

Alpers DH and Klein S (1999) Approach to the patient requiring nutritional supplementation. In: Yamada T, Alpers DH, Laine L, Owyang C and Powell DW (eds) *Textbook of Gastroenterology* 3rd edn, pp. 1081–1107. Philadelphia, PA: Lippincott Williams & Wilkins.

DeLegge MH and Copenhaver (1999) Enteral and parenteral nutrition. In: Brandt LJ and Daum F (eds) *Clinical Practice of Gastroenterology*, 1st edn, vol. 2, pp. 1567–1575. Philadelphia, PA: Churchill Livingstone.

Forchielli ML, Hendricks KM and Lo CW (1999) Home enteral parenteral nutrition. In: Walker WA and Watkins JB (eds) *Nutrition in Pediatrics: Basic Science and Clinical Applications*, 2nd edn, pp. 727–733. Hamilton, Ontario: B.C. Decker.

Mascarenhas MR, Tershakovec AM and Stallings VA (1999) Parenteral and enteral nutrition. In: Wyllie R and Hyams JS (eds) *Pediatric Gastrointestinal Disease*, 2nd edn, pp. 741–757. Philadelphia, PA: W.B. Saunders.

Mascarenhas MR, Kerner JA and Stallings VA (2000) Parenteral and enteral nutrition. In: Walker WA, Durie PR, Hamilton JR, Walker-Smith JA and Watkins JB (eds) *Pediatric Gastrointestinal Disease*, 3rd edn, pp. 1705–1752. Hamilton, Ontario: B.C. Decker.

Shike M (1999) Enteral feeding. In: Shils ME, Olson JA, Shike M and Ross AC (eds) *Modern Nutrition in Health and Disease*, 9th edn, pp. 1643–1656. Philadelphia, PA: Williams & Wilkins.

Smith P, Zeitzer J and Leibowitz AB (1995) Enteral nutrition. In: Haubrich WS, Schaffner F and Berk JE (eds) *Gastroenterology*, 5th edn, pp. 3209–3220. Philadelphia, PA: W.B. Saunders.

Stoker TW and Castle JL (1999) Special diets. In: Walker WA and Watkins JB (eds) *Nutrition in Pediatrics: Basic Science and Clinical Applications*, 2nd edn, pp. 727–733. Hamilton, Ontario: B.C. Decker.

Torosian MH (ed.) (1995) *Nutrition for the Hospitalized Patient: Basic Science and Principles of Practice*. New York: Marcel Dekker.

Enterobacteriaceae *See* ***Escherichia coli***: Occurrence; Detection; Food Poisoning; Occurrence and Epidemiology of Species other than *Escherichia coli*; Food Poisoning by Species other than *Escherichia coli*

Enzymatic Browning *See* **Browning**: Nonenzymatic; Toxicology of Nonenzymatic Browning; Enzymatic – Biochemical Aspects; Enzymatic – Technical Aspects and Assays

Enzyme Immunoassay *See* **Immunoassays**: Principles; Radioimmunoassay and Enzyme Immunoassay

ENZYMES

Contents

Functions and Characteristics

J R Whitaker, University of California, Davis, CA, USA

Introduction

Enzymes are absolutely essential for life. They occur in all animals, plants, and microorganisms. Nearly all chemical reactions in living organisms are a result of enzyme activities. Enzymes cause reactions to occur rapidly at ambient temperatures with a high degree of specificity. As an example, it takes 24–72 h at 100–110 °C and 6 N hydrochloric acid to hydrolyze a protein. Gastrointestinal-tract enzymes accomplish this in 2–4 h at 37 °C. There are thousands of different enzymes in living organisms.

Definition of an Enzyme

An enzyme is a protein with catalytic properties due to its power of specific activation. Some enzymes also require cofactors for activity. The properties and characteristics of enzymes are described below. A few ribonucleic acids, called ribozymes, also act as catalysts for a limited number of hydrolytic reactions, but they do not fit the definition of an enzyme as defined above.

History

Although the effects of enzymes, such as fermentation, digestion, milk clotting and meat tenderization, were known for centuries, the first clear recognition of an enzyme as a unique compound was in 1833, when Payen and Persoz reported that an alcohol precipitate of malt extract contained a thermolabile substance 'diastase' (now known as amylase) because of its ability to separate starch from insoluble envelopes of starch grains. They identified some characteristics of enzymes – precipitation with alcohol, thermolability, and specificity – and designed a naming system (stem of substrate name plus the suffix '-ase') that is used universally today. Schoenbein, in 1855, discovered peroxidase in plants that, in the presence of hydrogen peroxide, causes a brown solution of gum guaiac to turn blue. Later (1856), Schoenbein also discovered polyphenol oxidase in mushrooms, which causes them (and many other plant tissues) to turn brown. Berthelot in 1860

discovered invertase in yeast, which causes an inversion in the optical rotation of a sucrose solution. Several thousand enzymes are now well known.

The relationship between enzymes and living cells was a controversial topic between 1875 and 1900. Pasteur, a microbiologist and enologist, argued that fermentation was inseparable from living cells. Liebig, a chemist who demonstrated the action of pepsin on proteins, held that enzymes are chemical substances, active in the absence of cells. Büchner, in 1897, separated broken yeast cells from liquid, showing that the cell-free extract carried out fermentation, thereby ending the controversy. In 1878, Kühne proposed the name 'enzyme' (Greek for 'in yeast') for the substance transforming one compound to another. (*See* **Fermented Foods**: Origins and Applications.)

Protein Nature of Enzymes

Enzymes are proteins. This was generally accepted in the late 19th and early 20th centuries. However, this idea was challenged in the late 1920s by Willstätter. He purified peroxidase until no protein was detectable, but appreciable activity remained. Therefore, he concluded that enzymes could not be proteins. Sumner, in 1926, crystallized urease from Jack Bean meal and showed it to be a protein. This set off a polemic debate between Willstätter and Sumner that involved many scientists. In 1930, the famous biochemist Haldane indicated that, with the single exception of urease, almost nothing was known about the chemical nature of enzymes. However, soon thereafter, scientists at the Rockefeller Institute in New York, led by the distinguished biochemist, John Northrop, crystallized pepsin, trypsin, chymotrypsin, and carboxypeptidase A, among others. All were proteins. More than 600 enzymes, all proteins, have now been crystallized.

As proteins, enzymes have specific structures. Each has a fixed molecular size and a specific amino acid sequence (primary sequence), as determined by the gene for that enzyme. The primary structure is folded, via α-helices, β-pleated sheets, β and γ bends and random-coil segments, to give secondary structure. Further folding, giving a tertiary structure, results from thermodynamic and kinetic requirements that most of the hydrophobic amino acid residues must be inside the protein, away from water, and most of the hydrophilic amino acid residues must be on the surface. (*See* **Protein**: Chemistry.)

Many enzymes consist of single polypeptide molecules. Others have two or more subunits (identical or different) per molecule, giving rise to a quaternary structure. Some enzymes associate to form macromolecular structure systems to convert substrate to product(s) more efficiently. *Escherichia coli* pyruvate dehydrogenase is a complex of three different enzymes and five cofactors (M_r of 4.44×10^6 Da).

Early on, researchers investigated water-soluble enzymes. Now, much emphasis is on structurally bound enzymes.

Solvation of Enzymes

Water is an important factor in determining native enzyme structure, as well as activity. Hydrophobic amino acid residues are largely inside the enzyme molecule, away from water, and most of the hydrophilic amino acid residues are on the outside. Enzymes contain 30–40% (w/v) of bound surface water (170–220 mol of water per 10 000 gram of protein). The relatively high surface hydrophobicity of a folded polypeptide chain, and its lack of hydration by water molecules, determines whether an enzyme has a quaternary structure or not. In less polar solvents than water, but miscible with water, an enzyme probably would not fold in the same way, and is most likely inactive. Enzymes can be more stable suspended in immiscible organic solvents than in water. (*See* **Water Activity**: Effect on Food Stability.)

Size

Enzymes are large molecules. Ribonuclease ($M_r =$ 13 683) and lysozyme (14 100) are small enzymes, alkaline phosphatase (80 000) and mushroom polyphenol oxidase (128 000) are of intermediate size, and β-galactosidase (520 000) and glutamate dehydrogenase (2 000 000) are large enzymes.

Catalytic Nature of Enzymes

The most unique feature of enzymes is their ability to bind compounds (called substrates) stereospecifically and to convert them to other compounds (called products) very efficiently. Two measures of enzyme efficiency are turnover number and rate enhancement. Example turnover numbers are 10^2–10^7 mol of substrate converted to product per second per mole of enzyme active site for chymotrypsin and catalase, respectively. Some examples of rate enhancement (based on the same temperature and concentration) are: hydrogen peroxide is converted to water and oxygen by catalase 3.5×10^8 times faster than noncatalytically; invertase hydrolyzes sucrose to glucose and fructose 5.6×10^{10} times faster than does 1 M hydrochloric acid; urease hydrolyzes urea to carbon dioxide and ammonia 4.2×10^{11} times faster than does 1 M hydrochloric acid. This remarkable rate enhancement makes life possible at ambient temperatures.

Enzymes increase rates of conversion of substrate to product by lowering the activation energy, E_a. E_a is the minimum energy that a substrate must acquire above the ground-state energy to go to product. In two of the examples above, catalase lowers E_a from 18.0 to 6.4 kcal mol^{-1} (3.5×10^8 rate enhancement) for the oxidoreduction of hydrogen peroxide to products and urease lowers E_a from 24.2 to 8.7 kcal mol^{-1} (4.2×10^{11} rate enhancement) in the hydrolysis of urea.

Enzymes convert substrates to products by well-known chemical mechanisms. Because of their efficiency, it was once considered that they must have magical qualities. We now know that several well-known factors can account for this efficiency. These are: (1) conversion from inter- to intramolecular reaction; (2) proximity and specific orientation of substrate in the active site of the enzyme; (3) catalysis by distortion of bonds; (4) general acid–general base catalysis; and (5) nucleophilic–electrophilic catalysis. These will be explained below.

Active Sites of Enzymes

Reactions occur at active sites of enzymes. The active site is a small area, a cavity or hole on the surface of the enzyme. The active site consists of 10–15 amino acid residues brought together by folding from different parts of the primary structure of the protein. One part of the active site is responsible for stereospecific binding of the substrate in proximity to the (second) transforming part of the active site. In α-chymotrypsin, residues Ser189, Gly216, and Gly226 are responsible for binding the peptide substrate, and residues His57, Ser195, and Asp102 are responsible for catalyzing hydrolysis of the scissile peptide bonds. The amino group of Ile16 and the carboxyl group of Asp194 assist in maintaining the shape of the active site. The binding site recognizes L-Tyr, L-Phe or L-Trp residues of the substrate. Peptide bonds involving other L-amino acids or D-amino acids are not hydrolyzed. There must be very close complementarity, following any induced fit, between the shape of the substrate and the active site for binding to occur. (*See* **Amino Acids**: Properties and Occurrence.)

Proximity and Orientation Effects

Proximity and orientation effects include (1) intra- vs intermolecular catalysis, (2) reaction entropy, and (3) effective concentration of reactive groups. An absolute requirement for catalysis is that the substrate must bind stereospecifically in a correct orientation at the active site to form the enzyme–substrate complex, as

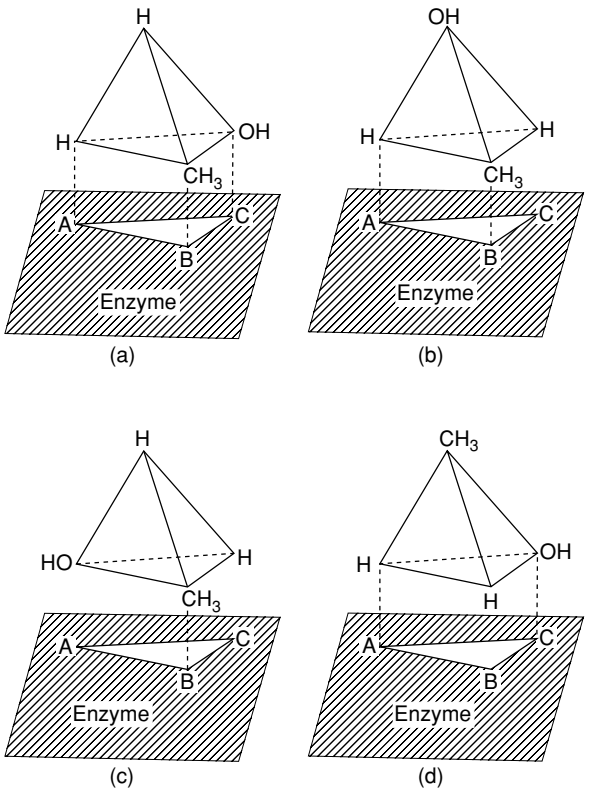

Figure 1 Schematic representation of several possible modes of orientation of ethanol on the surface of alcohol dehydrogenase. A, B, and C are binding points for two Hs, OH, and CH_3 of ethanol on the surface of the enzyme. Only (a) shows the ethanol bound properly with the enzyme surface.

shown in eqn (1), where E is the enzyme, S is the substrate, E·S is the enzyme–substrate complex, E·P is the enzyme–product complex, and P is the product. The obligatory E·S complex converts an intermolecular reaction to an intramolecular reaction. This increases the rate of an enzyme-catalyzed reaction compared with a noncatalyzed reaction by 10^3, 10^{15}, and 10^{22} times for bimolecular, trimolecular, and termolecular reactions, respectively.

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{E{\cdot}S} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{E{\cdot}P} \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} \mathrm{E} + \mathrm{P}. \qquad (1)$$

Binding of the substrate at the active site involves multiple (a minimum of three and up to 12) contacts between the substrate and the enzyme (Figure 1).

As a result of E·S, catalytic groups of the enzyme are 0.1–0.2 nm from the substrate bond to be transformed, usually by general acid–general base or nucleophilic–electrophilic catalysis (Figure 2). This proximity effect is equivalent to about a 10 M concentration of the corresponding intermolecular catalyst (His57 in Figure 2).

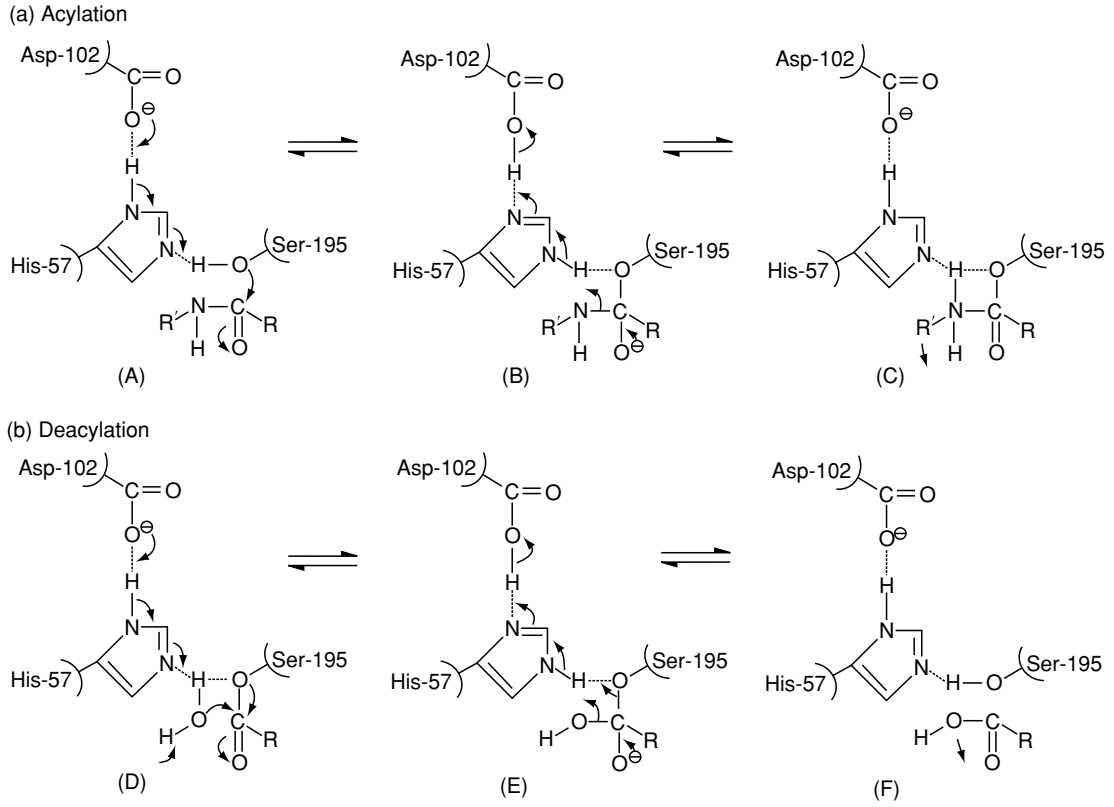

Figure 2 Proposed mechanism for α-chymotrypsin-catalyzed reactions. (a) Acylation. (b) Deacylation.

Table 1 Factors responsible for the rate enhancement of chymotrypsin-catalyzed hydrolysis of *N*-acetyl-L-tryptophanamide relative to the hydroxide ion and imidazole[a]

(1)	Rate constant for hydroxide ion catalysis	3×10^{-4} m^{-1} s^{-1}
(2)	Rate constant for imidazole catalysis	4.8×10^{-10} m^{-1} s^{-1}
(3)	Experimental rate constant for α-chymotrypsin catalysis	4.4×10^{-2} s^{-1}
(4)	Calculated rate constant for α-chymotrypsin catalysis (cumulative):	
	(a) Conversion of imidazole catalysis to intramolecular reaction (10^1)	4.8×10^{-9} s^{-1}
	(b) Nucleophilic catalysis by serine-OH (10^2) (see **Figure 2**)	0.8×10^{-7} s^{-1}
	(c) Proximity and orientation factors (10^3)	4.8×10^{-4} s^{-1}
	(d) General acid catalysis by imidazole (10^2) (see **Figure 2**)	4.8×10^{-2} s^{-1}

[a]At 25 °C; data from Bender ML, Kezdy FJ and Gunter CR (1964) The mechanism of action of proteolytic enzymes. XXXIII. The anatomy of an enzymatic catalysis. α-Chymotrypsin. *Journal of the American Chemical Society* 86: 3714–3721.

Catalysis by Distortion

There is a close complementarity between the structure and shape of a substrate and its binding position in the active site. This led Emil Fischer to the 'lock and key' analogy for the fit between a substrate and an enzyme. Later data established that enzymes permit some (small) latitude in the structure and shape of a substrate, leading to the 'rack mechanism' (Lumry) or the 'induced-fit' concept (Koshland). This results in a decrease in the activation energy needed, with a rate enhancement of 10^2–10^3.

Based on the factors above, an enzyme increases rates of reactions by 10^8–10^{28}, depending on how many substrates are involved in the reaction. Bender *et al.* in 1964 experimentally and theoretically determined rates of hydrolysis of *N*-acetyl-L-tryptophanamide by the three catalysts hydroxide ion, imidazole, and α-chymotrypsin (Table 1). There is remarkable agreement between the rate calculated by theory (Table 1, 4d) and determined experimentally (Table 1, 3).

Effect of Environmental Factors on Rates of Enzyme-catalyzed Reactions

Time, enzyme concentration, substrate concentration, nature of the substrate (including physical state), pH, temperature, solvent, activators, and inhibitors all affect rates of enzyme-catalyzed reactions.

Time

Enzyme-catalyzed reactions are time-dependent. Initially, the concentration of product, $[P]$, formed is linearly related to reaction time. How long the linear relation holds depends on the initial substrate concentration, $[S]_0$, in relation to the Michaelis constant K_m ($[S]_0$ at which $v_0 = 0.5\ V_{max}$; see eqn (2)), enzyme stability, pH, temperature and activators and reversibility of reaction. A tangent drawn to the initial part of the curve gives the initial velocity, v_0 ($= dP/dt$), an invaluable experimental parameter in enzymology.

Enzyme Concentration

For most enzyme-catalyzed reactions, v_0 is directly proportional to the enzyme concentration, $[E]_0$. Doubling $[E]_0$ doubles v_0. This is analytically very convenient since enzyme concentrations in biological systems are based on determination of v_0 under standard conditions. However, there are several cases where direct relations do not hold, as described by Whitaker in 1994.

Substrate Concentration

Because of the obligatory E·S (eqn (1)), the relation between $[S]_0$ and v_0 gives a right-hyperbolic plot (Figure 3). The extent of saturation of an enzyme with a substrate, to form E·S, is dependent on the relation of $[S]_0$ to K_m. At $[S]_0 << K_m$, the reaction rate is first order with respect to $[S]_0$ ($dP/dt = k[E]_0[S]_0/K_m$). At $[S]_0 >> K_m$, the reaction rate is zero order with respect to $[S]_0$ ($dp/dt = k[E]_0$), thereby giving the maximum velocity, V_{max}, under the conditions used. At $[S]_0 = K_m$, $v_0 = 0.5V_{max}$. When $[S]_0 \geq 0.01K_m \leq 100K_m$, the reaction rate is a mixture of first and zero order.

Enzyme-catalyzed systems that follow Figure 3 obey Michaelis–Menten kinetics

$$v_0 = k[\mathrm{E}]_0[\mathrm{S}]_0/(K_m) + [\mathrm{S}]_0). \qquad (2)$$

Substrate activation and inhibition and allosteric behavior cause deviations from this relation.

A plot of $1/v_0$ vs $1/[S]_0$ gives a linear relationship (Lineweaver–Burk equation), permitting V_{max} and K_m to be determined readily (Figure 3 insert).

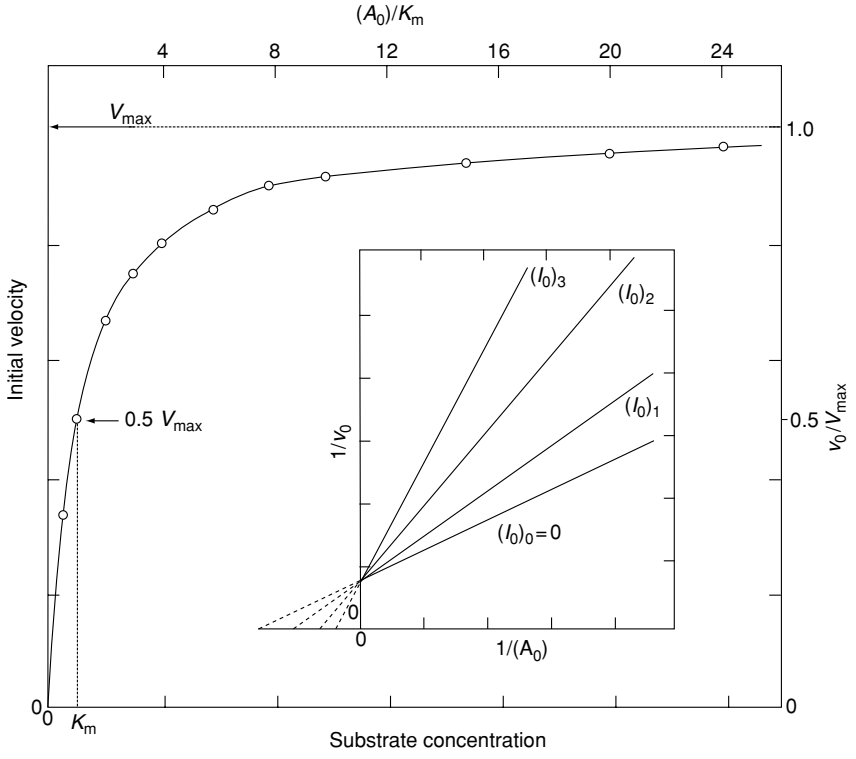

Figure 3 Relation between initial velocity, v_0, and initial substrate concentration, $[S]_0$, for an enzyme-catalyzed reaction. Insert: effect of three different concentrations of a competitive inhibitor on $1/v_0$ versus $1/[S]_0$ plot. $[I_0]_0$, $[I_0]_1$, $[I_0]_2$, and $[I_0]_3$ are 0, 0.5, 1.5, and 3.0 times K_i, respectively.

pH

Enzyme-catalyzed reactions are sensitive to the pH of the reaction, resulting in a pH optimum (maximum v_0). Examples of pH optima are: porcine pepsin, pH 2; horse-radish peroxidase, pH 6; bovine trypsin, pH 8; bovine milk alkaline phosphatase, pH 10; catalase, pH 4–9. The relation between pH and v_0, bell-shaped or sigmoidal, results from the effect of pH on the ionization of essential catalytic or binding groups in the active site of the enzyme and the effect of pH on stability of the enzyme as a protein. (*See* **pH – Principles and Measurement.**)

Temperature

Temperature affects enzyme-catalyzed reactions by influencing v_0, the stability of the enzyme as a protein and various equilibria, including K_m and pK_a of catalytic groups. There is a logarithmic effect of temperature on the rate of reaction (Figure 4, curve A), as predicted by the Arrhenius relation:

$$k = A_e^{-E_a/RT}, \qquad (3)$$

where k is a rate constant, A is a proportionality constant, T is temperature (degrees Kelvin), R is the universal gas constant, and E_a is the Arrhenius activation energy. At higher temperatures, depending on the specific enzyme, enzymes lose activity because of instability (see curve B, Figure 4), resulting in a decrease in activity. The temperature at which the maximum v_0 is obtained is the temperature optimum (apex of curve C, Figure 4) under the conditions used.

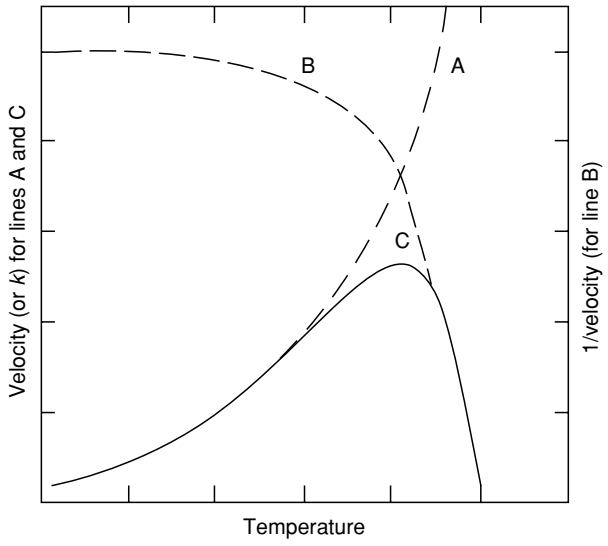

Figure 4 Effect of temperature on v_0 of an enzyme-catalyzed reaction. Curve A shows the effect of temperature on the velocity of substrate conversion to product. Curve B shows the effect of temperature on the stability of the enzyme. Curve C shows the actual effect of temperature on the velocity of substrate conversion to product.

Enzyme Activation

Any compound that increases v_0 when added to an enzyme–substrate system is an activator. These compounds can be cofactors (coenzymes or prosthetic groups), allosteric effectors, conformational effectors, or converters of proenzymes to enzymes (by proteolysis, usually).

Cofactors are small organic or inorganic molecules essential for enzyme activity. Organic cofactors are often built around the B vitamins (pyridoxine, thiamin, riboflavin, niacin, and pantothenic acid), whereas inorganic cofactors include Zn^{2+}, Mn^{2+}, Mo^{5+}, Mo^{6+}, Cu^{2+}, Mg^{2+}, Co^{2+}, Ca^{2+}, Fe^{2+}, Fe^{3+} and Cl^-. For example, more than 150 enzymes require Zn^{2+} as a cofactor. Coenzymes are bound via noncovalent bonds, whereas prosthetic groups generally are covalently bound to the protein. (*See* **Coenzymes.**)

Enzyme Inhibition

Any compound that decreases v_o when added to an enzyme–substrate system is an inhibitor. Inactivation caused by adverse pH, temperature or solvents is not included. Inhibitors occur naturally in many biological materials, may be produced synthetically, or may be metal ions (particularly Hg^{2+}, Pb^{2+}, Ag^+, Cu^{2+}). Enzyme inhibitors are important in pathogenic microorganism control, food storage, and food-quality preservation. Many pharmaceuticals are designed to inhibit specific microbial enzyme systems. Fungicides, herbicides, and insecticides are targeted to specific enzymes. Products of enzyme reactions are often inhibitors, since they resemble substrates. (*See* **Fungicides**; **Pesticides and Herbicides**: Types, Uses, and Determination of Herbicides; **Storage Stability**: Mechanisms of Degradation.)

Naming of Enzymes

Enzymes are named by adding the suffix '-ase' to the stem of the substrate name, e.g. *amylum* (Latin for 'starch')/*amylase*, *maltose/maltase*, *carbohydrate/carbohydrase*, *lipid/lipase*. However, there are several enzymes that act on a carbohydrate, for example, or a single enzyme may act on several substrates. Other trivial names are derived from the source of the

enzyme or the color of the purified enzyme. In 1955, the International Union of Biochemistry appointed members from many countries to a commission on enzymes, with the charge to develop a systematic method of naming enzymes. In 1961, a classification was adopted, based on dividing 'the enzymes into groups on the basis of the type of reaction catalysed and this, together with the name(s) of the substrate(s), provided a basis of naming individual enzymes'. Retention of the trivial name (already in general use) and addition of a four-digit numbering system were also adopted.

All enzymes fit one of six types of chemical reactions. These are (with the first digit of the number and a generalized reaction):

EC 1. *Oxidoreductases*
$AX + B \rightleftharpoons A + BX$, where X is H or e^-; B is oxygen etc.

EC 2. *Transferases*
$AY + B \rightleftharpoons A + BY$, where Y is methyl, acyl, glycosyl, etc.

EC 3. *Hydrolases*
$AB + H_2O \rightleftharpoons AH + BOH$, where water is a substrate.

EC 4. *Lyases*

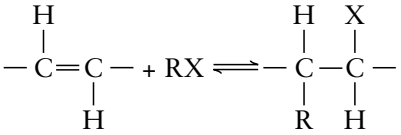

EC 5. *Isomerases*

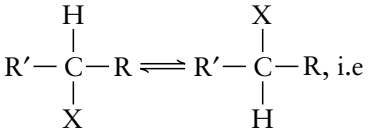

D– $\rightleftharpoons$ L–, cis $\rightleftharpoons$ *trans*, keto $\rightleftharpoons$ enol, etc.

EC 6. *Ligases (synthetases)*
$A + B + AJP \rightleftharpoons AB + ADP + P_i$ (or AMP + PP), where P_i is inorganic phosphate.

The first digit of the EC number indicates the class to which the enzyme belongs, the second digit the subclass, the third digit the subsubclass, and the fourth digit serial number of the enzyme in its subsubclass. For example, the trivial name of the enzyme catalyzing the reaction shown in eqn (4) is alcohol dehydrogenase, the systematic name is alcohol:NAD^+ oxidoreductase, and the number is EC 1.1.1.1. The first digit indicates that it is an oxidoreductase, the second that the donor is a primary alcohol, the third that the acceptor is NAD^+ or $NADP^+$, and the fourth that the enzyme is alcohol dehydrogenase.

$$CH_3CH_2OH + NAD^+ \rightleftharpoons CH_3CHO + NADH + H^+. \quad (4)$$

The naming of an enzyme is complete when the organism, organ, and isozyme (if present) are indicated. For example, one of the two pig pancreas α-amylases is porcine pancreatic α-amylase-1, and the other is porcine pancreatic α-amylase-2. The isozyme number is assigned consecutively, on the basis of the fastest moving isozyme towards the anode by electrophoresis, as '1', and so on.

See also: **Amino Acids**: Properties and Occurrence; **Coenzymes**; **Fermented Foods**: Origins and Applications; **Fungicides**; **Pesticides and Herbicides**: Types, Uses, and Determination of Herbicides; **pH – Principles and Measurement**; **Protein**: Chemistry; **Storage Stability**: Mechanisms of Degradation; **Water Activity**: Effect on Food Stability

Further Reading

Anonymous (1992) *Enzyme Nomenclature, Recommendations of the International Union of Pure and Applied Chemistry and the International Union of Biochemistry*. Orlando, FL: Academic Press.

Bender ML, Kezdy FJ and Gunter CR (1964) The mechanism of action of proteolytic enzymes. XXXIII. The anatomy of an enzymatic catalysis. α-Chymotrypsin. *Journal of the American Chemical Society* 86: 3714–3721.

Boyer PD (ed.) (1970-1978) *The Enzymes*, 3rd edn, vols. 1–18. San Diego, CA: Academic Press.

Colowick SP and Kaplan NO (editors-in-chief) (1955–1999) *Methods in Enzymology*, vols 1–300. San Diego, CA: Academic Press.

Schwimmer S (1981) *Source Book of Food Enzymology*. Westport, CA, USA: AVI.

Whitaker JR (1994) *Principles of Enzymology for the Food Sciences*. New York: Marcel Dekker.

Uses in Food Processing

G B Quaglia and L Gennaro, Istituto Nazionale di Ricerca per gli Alimenti e la Nutrizione, Rome, Italy

Introduction

Enzymes play important roles in various aspects of food processing. In some cases they have made possible the development of new products and, in other cases, the improvement of traditional products. At least 75% of all industrial enzymes are hydrolytic in

action, and they are used for the depolymerization of natural substances. Proteinases remain the dominant type because of their wide use in dairy products processing (as coagulants); carbohydrases, used in baking, brewing, distilling, and especially in starch processing, represent the second largest group. The conventional distribution of the current world sales of enzymes is to assess them by their application sectors (dairy, detergent, starch, textiles, and other; **Figure 1**) This has been the practice for many years but it is important to consider the composition of section listed as other in **Figure 1**, in order to develop a view of future markets. The section covered by 'others' includes the following: alcohol, baking, fats and oils, flavor, fruit, and wine. The distribution of sales for 2005 is expected to be very different, because 'other' sectors will become the single largest part of enzymes sales, up to approximately 47% (**Figure 2**).

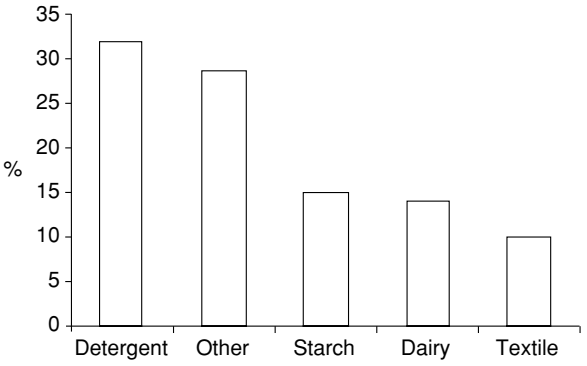

Figure 1 Distribution of enzymes sales. Data from Godfrey T and West SI (1996) Introduction to industrial enzymology. In: Godfrey T and West SI (eds) *Enzymology*, pp. 1–8. London: MacMillan Press.

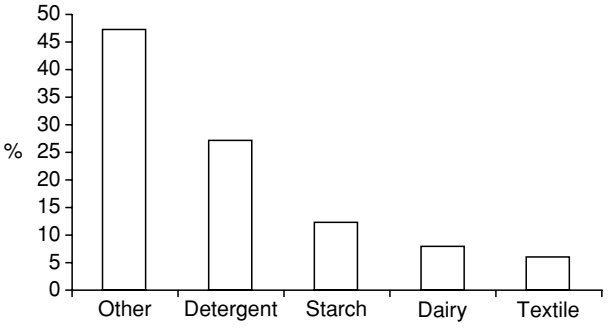

Figure 2 Forecast distribution of enzyme sales. Data from Godfrey T and West SI (1996) Introduction to industrial enzymology. In: Godfrey T and West SI (eds) *Enzymology*, pp. 1–8. London: MacMillan Press.

The advantage of using enzymes in food technology is primarily derived from two intrinsic properties: catalytic rate enhancement and specificity. The ability of enzymes to enhance reaction rates by many orders of magnitude may seem obvious, but it is particularly significant when these rates are achieved under very mild conditions. For most enzymes, this means an aqueous environment at atmospheric pressure, a pH within the range 3–9, and a temperature between 15 and 50 °C. In this way, the formation of chemically induced byproducts is vastly reduced. For example, in starch processing, chemical degradation yielded a product containing up to 6% of residual polysaccharide, as well as colored degradation products, the removal of which required a further purification step; the most recent enzymatic process produced a clean product containing at most 0.1% oligosaccharide. Moreover, the mild processing conditions facilitate the handling of sensitive substrates (such as foodstuffs) where extremes of temperature, pH, or pressure would be undesirable. Enzyme specificity is the other great advantage. Specificity, in fact, not only reduces interference by undesirable substrates but also minimizes the problem of unwanted byproducts, which are always a costly inconvenience in industrial processing. Finally, enzymes can be easily inactivated by thermal or chemical means once the desired degree of conversion has been achieved.

The versatility of enzymes in food processing will be evident by the numerous examples cited in this work. There are a number of ways in which an enzyme preparation can be presented to the market: as an amorphous or microcrystalline powder, in solution, or in some immobilized form. The nature of the preparation will be determined largely by the requirement of every industry. Liquid preparations are often preferred, because volume measurements are simple, but preparing the enzyme in an insoluble form would not only facilitate recovery of enzymes, but also permit reuse, which has obvious economic advantages. For this reason immobilized or bound enzymes have been developed by physically or chemically binding the enzyme to an insoluble support. The main cases of food processing in which immobilized enzymes are used are the conversion of glucose to fructose by glucose isomerase in the production of high-fructose syrups, and the clarification of fruit juices by means of pectic enzymes.

Any organism, whether plant, animal, or microorganism, is a potential source of enzymes. However, the nature of the source dictates the availability, the cost of source material, the ease of recovery, and many other factors (**Table 1**). The majority of enzymes used on an industrial scale are from microbial sources (**Table 2**). One exception was that of rennin, the protease derived from calf stomach, which is used in cheese-making. Even here a microbial substitute,

Mucor rennin, is now being used in a significant proportion of the dairy industry worldwide, and a recombinant chymosin (the main rennet enzyme) will be produced on a large scale. The application of genetic engineering, particularly recombinant DNA technology, is in fact having a major impact on the development of new sources of industrial enzymes for food industry, even if in western countries there is an increasing tendency among consumers not to approve the application of genetically engineered organisms to food and food ingredients.

Selected examples of enzymes used in food processing are mentioned below, and some of these will be discussed in detail:

- Meats are tenderized using the enzymes papain, ficin, and bromelain.
- Invertase is added to the fondant centers of chocolate-covered cherries. This enzyme hydrolyzes some sucrose, thereby decreasing the number of sugar crystals in the fondant, and softening the chocolate center.
- Glucose isomerase is used to convert glucose to the sweeter fructose. This process is carried on commercially with an immobilized enzyme system.
- Pectinases are used commercially to break down pectin molecules in certain fruit juices such as apple juice in order to avoid a hazy appearance; clear, sparkling juice is the final product.
- Rennet is the crude extract of the protease rennin. It is widely used to clot milk in the making of cheese.
- An enzyme system involving glucose oxidase and catalase has been used to remove traces of oxygen in tightly packaged food products. This system is also used to remove small amounts of glucose from white eggs before drying them. Browning may occur during storage if white eggs contain even small quantities of glucose.

Table 1 Sources of major industrial enzymes

Enzyme	*Process*	*Source*
Proteases – alkaline	Detergents	Bacterial
Proteases – acid	Cheese-making	Animal, fungal
Proteases – neutral	Various	Bacterial
Amylases	Starch industry	Bacterial, fungal
Isomerases	High-fructose syrups	Fungal
Pectinases	Wine, beer, condiments	Fungal
Lipases	Various	Bacterial, fungal, animal

From Cowan DA (1994) Industrial enzymes. In: Moses V and Cope RE (eds) *Biotechnology*, pp. 311–340. UK: Harwood Academic Publisher with permission.

Use in Tenderizing Meats

The Structure of Meat

It has been established beyond doubt that the consistency of meat is governed by two principal factors: the quantity and the properties of the muscle tissue collagen and the mechanically contractile state of the muscle. (*See* **Meat**: Structure.)

Collagen fiber consists of triple-helix tropocollagen molecules that are linked by covalent bonds, providing a structure of high mechanical tenacity. In muscle tissue collagen both thermolabile and thermostable bonds have been identified. The thermolabile linkages are prevalent in the collagen of young animals, and are readily broken when the meat is cooked. In older animals the thermostable bonds predominate, with the result that the meat has harder collagen fibers when cooked, i.e., tougher meat. Both the amount of collagen and the type of bond are thus important in determining meat consistency. A further important factor is the state of contraction of the muscle protein.

Once the animals have been slaughtered, chemical bonds form between the thick and the thin filaments

Table 2 Various sources of enzymes

General source	*Specific source*	*Advantage/disadvantages*	*Example*
Animals	Human blood	No longer a viable source. Available in vast quantities	Pancreatic lipase
	Bovine, porcine, and ovine tissues	Enzymes are often of low stability	Rennin
		Purification often complex	Trypsin
Plants	Leaves, fruit	Purification can be a problem	Papain
Microorganism	Bacterial, algal, fungal, and yeast	Large-scale production	Protease
		Technology advanced	Amylase
		Purification often simple	Lipase
		High yields	
		Wide variety of unusual enzymes available	
		Limited range of microorganism approved for use in food industry	

From Cowan DA (1994) Industrial enzymes. In: Moses V and Cope RE (eds) *Biotechnology*, pp. 311–340. UK: Harwood Academic Publisher with permission.

that make up the mobile part of the muscle. It is these bonds that make the muscle hard and reduce its stretching properties. Hardness is most noticeable in highly contracted muscle, where a compact and inelastic structure forms.

In order to make it tender, meat is subjected to 'ripening' ('hanging') once rigor mortis (a stiffness of the muscle resulting from glycolysis) ceases. The process is widely exploited to make the meat not only tender, but also tasty and more digestible. This process takes place when the carcass is cut into sides and quarters. Hanging entails keeping the sides for several days in controlled temperature and humidity chambers. The mode of action of the tenderizing process during this ripening and conditioning is not entirely understood. It is suggested that during this period there is a partial dissolution of the Z-lines, whose function is to anchor the thin filaments (myofibrils) of the muscle tissue. During such a process, dissociation of the thin filaments from the Z-lines leaves the entire muscle structure more stretchable and tender. However, the partial degradation of the Z-lines is due to the action of a specific proteolytic enzyme, which has been given the name of calcium-activated sarcoplasmatic factor (CASF). (*See* **Meat**: Structure.)

Artificial Tenderizing

The normal tenderization process thus seems to involve one or more specific proteases present in the muscle cells; and it is reasonable to expect that the process can be reproduced artificially with the use of proteases. This practice was already in use among the pre-Columbian inhabitants of the Americas. According to Hernan Cortés, the conquistador of Mexico, these people would wrap up pieces of meat in papaya leaves. The papain (a proteolytic enzyme) naturally present in the leaves acted on the structure of the meat, making it tender.

Artifical tenderization is achieved by resort to various sulfhydryl proteases of vegetable origin, such as papain (from the papaya plant), bromelain (from pineapple), and ficin (from the fig tree) or microbial proteases. Factors such as pH, the concentration of the enzyme, and temperature also influence the enzyme action and the degree of tenderization.

Since the enzymes employed in this field are inactive at low temperature (0–10 °C) but increase in activity as the temperature rises (65–80 °C), becoming inactive again at 82 °C, the tenderization process, for all practical purposes, begins and ends with cooking. The problem, then, is to cause the enzyme to act in such a way as to produce a controlled and homogeneous tenderizing, since excessive action during the cooking process would produce a formless lump of tissues.

Tenderization by enzymes is achieved through the action of the enzyme on the connective tissue (collagen and elastin) of the meat, and cannot be compared, in terms of its mechanism, with the natural tenderizing process described earlier.

Enzymes come in powder form, for dusting on the surface of the meat, or in liquid form; in this case the meat is either immersed in solutions of known concentrations or (a more effective method) by injecting the enzyme into the blood stream a short time before slaughtering. This is known as pretenderizing. The method in this case relies on the fact that (1) the animal's vascular system offers an excellent means of distribution among the body tissue; (2) the heart is an efficient pump; (3) blood acts as a diluent and insures uniform distribution of the enzyme; and (4) the enzymes used are safe food ingredients.

The time elapsed between pretenderizing and slaughtering – from 2 to 20 min – allows adequate distribution in the tissues. The amount of solution will vary with the weight, age, and sex of the animal. The enzyme does not act immediately upon introduction into the animal, since the temperature is below optimum, but begins to do so when the meat is cooked.

Apart from the difficulty of achieving homogeneous distribution throughout the tissues, artificial tenderizing with enzymes is further limited by the fact that the enzymes used are of low specificity, in that they act on both connective tissue (collagen and elastin) and on contractile (myofibrillary, muscle structure) protein. These two aspects make it difficult to control the tenderization process.

To overcome these constraints, collagenases are now starting to be used. Crude collagenases are highly specific proteases which act on native or denaturated collagen but not on muscle tissue, so that the dissolution of the actomyosin and Z-line proteins is restricted, while the connective tissue network is weakened. Collagenases are obtained from aerobic bacteria (*Achromobacter* spp.) and *Clostridium* strains. The outlook is promising for their application in this field.

Use in Hydrolyzing Starch

The conversion of native starch, which is the basic constituent of many vegetable products, such as maize, and potatoes, to soluble sugars is the most important application of enzymes in food industry; and in fact the process has almost entirely supplanted that of acid hydrolysis for glucose production. (*See* **Starch**: Structure, Properties, and Determination.)

In the last 30 years, as new enzymes have become available, starch hydrolysis technology has been

transformed. There has been a big move away from acid and today the vast majority of starch hydrolysis is performed using enzymes. Furthermore, in the 1970s, an enzyme technique made possible the production of syrup sweeter than glucose – high-fructose syrup. The production of this syrup has significantly boosted the growth of the starch industry in certain countries.

Depending on the enzymes used, syrups with different compositions and physical properties can be obtained from the starch (Table 3). The dextrose equivalent (DE) value is used as an indication of the degree of hydrolysis of a syrup. The DE value of starch is zero, and that of dextrose is 100. Syrup with DE values from 35 to 43 are still widely produced by acid hydrolysis, despite the drawbacks mentioned here. However, due to the formation of byproducts, it is difficult to produce low- and high-DE syrups of high quality in this way.

There are essentially five groups of enzymes involved in the hydrolysis of starch: the endo-amylases (Enzyme Commission (EC) 3.2.1.1) act primarily on the α-1,4 linkages, and products of reactions are oligosaccharides of varying chain lengths.

The exo-amylases (amyloglucosidases or glucoamylases; EC 3.2.1.3) act on the same substrates as endoamylases, but they are also able slowly to cleave α-1,6 linkages. They act externally on substrate bonds from the nonreducing end and produce low-molecular-weight products.

The debranching enzymes (such as pullulanase, EC 3.2.1.41, and isoamylase, EC 3.2.1.68) act exclusively on the α-1,6 linkages.

A fourth group, the isomerases (EC 5.3.1.5), act on glucose syrups to convert them to fructose syrup.

Finally, cyclodextrin glycosyltransferase (CGTase, EC 2.4.1.19) is an enzyme capable of hydrolysing starch to a series of nonreducing cyclomaltooligosaccharides referred to as cyclodextrins.

There are three basic steps in enzymatic starch conversion to glucose: gelatinization, liquefaction, and saccharification. In the first step, the raw starch is heated to above 40 °C to disrupt the starch granules and expose the polymer to enzymatic action (Figure 3). In its native granular form, starch resists attack from hydrolytic enzymes, so that the application of these (enzyme hydrolysis) calls for acid or enzyme pretreatment in order to bring about depolymerization and consequent reduction in the viscosity typical of gelatinized starch. The starch slurry (pH 6.5) is gelatinized by cooking at high temperature with α-amylase. The gelatinized starch is then liquefied by thermostable α-amylase to a soluble dextrin hydrolysate with a DE of 5–15. After liquefaction, the pH of slurry is adjusted

Table 3 Classification of enzymes involved in the commercial production of starch syrups, maltodextrins, and cyclodextrins

Type	*Common name*	*Source*	*Substrate specificity*	*Optimum*	
				pH	*Temperature (°C)*
Endoamylase	Bacterial amylase	*Acillus subtilis*	α-1,4-Glycosyl	6.0	65–70
		Bacillus licheniformis	α-1,4-Glycosyl	5.0–7.0	90
	Fungal α-amylase	*Aspergillus oryzae*	α-1,4-Glycosyl	4.5	50–60
Exoamylase	Amyloglucosidase	*A. niger*	α-1,4-Glycosyl α-1,6-Glycosyl	4.0–5.0	60
	Bacterial β-amylase	*Bacillus* spp.	α-1,4-Glycosyl	5.0	55–60
		Clostridium spp.	α-1,4-Glycosyl	5.5–6.0	75–85
α-1,6-amylase	Pullulanase	*Klebsiella aerogenes*	α-1,6-Maltotryosyl	5.0	60
	Isoamylase	*Pseudomonas* spp.	α-1,6-Heptasacch	4.0	50–55
Exo-α-amylases	Exomaltotriohydrolase	*Streptomyces griseus*	α-1,4-Glycosyl		
		B. subtilis	α-1,4-Glycosyl		
	Exomaltotetrahydrolase	*B. circulans*	α-1,4-Glycosyl		
		Pseudomonas stutzeri	α-1,4-Glycosyl		
	Exomaltopentahydrolase	*Pseudomonas* spp.	α-1,4-Glycosyl		
	Exomaltoexahydrolase	*B. subtilis*	α-1,4-Glycosyl		
		B. circulans	α-1,4-Glycosyl		
Isomerase	Glucose isomerase	*B. circulans*	Aldo/keto pentose Aldo/keto hexose	8.2	65
Glucanotransferase	Cyclodextrinase	*B. macerans*	α-1,4-Glycosyl	5.0	50
		Thermoanaerobacter	α-1,4-Glycosyl	4.5	90
Endo-/exocellulase	Fungal cellulase	*Trichoderma reesei*	β-1,4-Glycosyl		
Endoprotease	Bacterial protease	*B. licheniformis*		6.5	90
		B. subtilis		6.5	65.70

Data from Guzman-Maldonado H and Paredes-Lopez O (1995) Amylolytic enzymes and products derived from starch: a review. *Critical Reviews in Food Science and Nutrition* 35: 373–403.

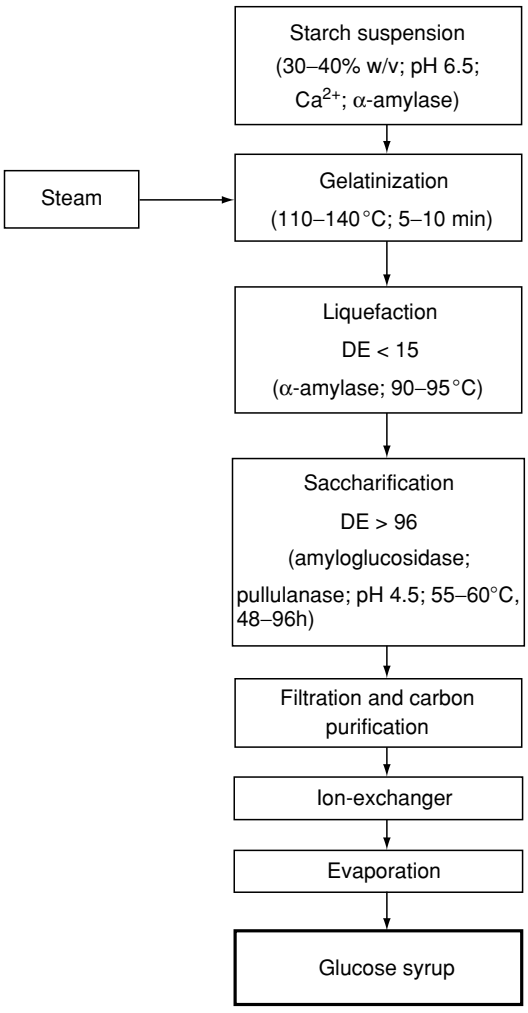

Figure 3 Production of glucose syrups. DE, dextrose equivalent.

to 4.5 and the temperature lowered to 60 °C. Amyloglucosidase is now added and the saccharification reaction is allowed to continue for 48–96 h in a stirred tank until the maximum level of glucose is reached.

Replacing soluble amyloglucosidase with its immobilized form allows for using the same enzyme batch several times in order to shorten the hydrolysis period and to apply a continuous process. However, the use of immobilized amyloglucosidase is less efficient in the production of glucose syrups than the soluble enzyme, giving a maximum DE value of 96% instead of 98%. The reason for this lower maximum DE is the increased formation of isomaltose and the decreased hydrolysis of the high-molecular-weight oligosaccharides.

The carbohydrate composition of typical glucose syrup is as follows:

- 94–98% glucose
- 1–3% maltose
- 0.3–0.5% maltotriose
- 1–2% higher saccharides.

It is possible, in industrial processing, to raise the sweetening power of glucose syrup by converting part of the glucose to fructose by isomerization with immobilized glucose isomerase obtained from a selected strain of *Bacillus coagulans*. This catalyzes the conversion of the glucose syrup in a fructose syrup containing at least 42% fructose.

Finally, it is worth mentioning cyclodextrins, which have been studied for nearly a century but not utilized for food applications until the 1990s because they were too expensive and there was insufficient information to declare them nontoxic. Now, their use is permitted in a variety of food applications.

In a so-called nonsolvent process a reactor is filled with starch and the selected enzyme under sterile conditions (because no chemicals are present to prevent microorganism growth); after the completion of the reaction, heat or acid inactivates the CGTase, and water is evaporated.

Cyclodextrins can stabilize volatile materials, such as flavors and spices, and can also stabilize emulsions of fats and oils, shielding them from oxidation and preventing rancidity. They can be used to debitter citrus juices, and can be widely used in the solubilization of organic compounds. Their market is still limited, because of their cost, but a novel production of cyclodextrins in transgenic potatoes has been developed at lower production costs.

Use in Baking

The use of wheat and other cereals as a source of flour for baking (particularly in the production of bread, probably the major staple food item in the western world) provides a demonstration of a natural enzymatic process which can be artificially modified to optimize product quality. Amylases are a natural constituent of flour. The levels of amylase activity in flour are strongly influenced by the climatic conditions in which the grain was grown. A high amylase content, the result of a moist climate, results in high dextrin production during baking. The resulting dough has low water retention and tends to be 'sticky,' while the loaf is crumbly and dry. A dry environment yields flour with a low amylase content. As a consequence, the dextrin yield is low, fermentative gas production is poor and loaf size is small. (*See* **Bread**: Breadmaking Processes.)

Normal flour, milled from sound wheat, contains significant amounts of β-amylase, but little or no α-amylase. β-Amylase can produce some maltose to aid fermentation without the presence of α-amylase,

but the amount produced is relatively small. The supplementation of dough with α-amylase affects the functional properties of the dough and may determine characteristics that are critical for automated manufacturing processes. Largely for this reason, α-amylase, sometimes in conjunction with protease, is commonly used by the baking industry.

Malted barley has traditionally served as the primary source of added α-amylase. This was the practice years ago when bread formulations were quite lean and contained relatively little sugar. Today, the primary reason for adding amylases to baked products is to improve the processing conditions as well as the overall quality of the baked product. Although barley malt is still used, supplementation with fungal and, to a lesser degree, bacterial amylases has become more common.

Cereal amylases remain active until the temperature reaches about 70 °C. They can cause some additional starch hydrolysis after the starch has gelatinized. Thus, cereal amylases, whether added to the flour (as a supplement) or native (due to sprouting), can produce excessive hydrolysis in high-moisture bread products if too much is present in the formulation.

Bacterial amylases (from *Bacillus* species) are generally the most heat-stable (> 100 °C); some remain active all the way through the bread baking cycle. They must be considered to be most problematic in bread production since they can continue to hydrolyze the starch throughout the baking process. In fact, some bacterial amylases are capable of continuing working even after baking and packaging and while sitting on the grocery shelf. Excessive starch hydrolysis results in poor loaf characteristics, including low loaf volume and dense, gummy crumb texture. The heat stability of bacterial α-amylase is of less concern in products such as cookies and crackers, which have a higher baking temperature and contain little water.

Fungal α-amylases, usually from *Aspergillus oryzae, A. niger, A. awamori*, or species of *Rhizopus*, is normally used to supplement the amylolytic activity in flour. Enzymes from these sources can raise the levels of fermentable monosaccharides and disaccharides of dough from a native level of 0.5% to concentrations that promote yeast growth. The sustained release of glucose and maltose by added fungal and endogenous enzymes provides the nutrients essential for yeast metabolism and gas production during panary fermentation. The *A. oryzae* α-amylase is sometimes favored for baking applications since this fungal enzyme is heat-labile at 60–70 °C and does not survive the baking process. Its thermolability prevents enzymatic action on the gelatinized starch in the finished loaf, which would cause a soft or sticky crumb.

Even proteolytic action on the gluten component is vital to the quality of the product. The protein component makes up a relatively small percentage of the flour, compared to starch, but it is responsible for the formation of gluten, the unique viscoelastic protein network peculiar to wheat. Because gluten protein is primarily responsible for the mixing and sheeting behavior of dough, it is an obvious candidate for enzymatic modification by protease. In particular, supplementation with proteases helps to break down the gluten protein so that the dough is softer and more extensible. Since cereals usually have low proteinase levels, supplementation with fungal (*A. oryzae*) proteinase is routine. The more thermostable bacterial proteases are generally used only in biscuit and pizza dough where the higher cooking temperature rapidly inactivates the enzymes and prevents overproteolysis and stickiness. While most commercially available proteases contain relatively nonspecific endo- and exoproteolytic activity, the units of activity will have been determined on nongluten substrate protein. This can make it difficult to choose the more efficient enzyme to hydrolyze the gluten protein in a particular process. The ability of any protease to hydrolyze the gluten effectively must be determined by experimentation in an appropriate formulation and processing scheme. Because the majority of the effects desired on wheat doughs involve altering the viscosity or rheology of the dough, an endoprotease is more effective than an exoprotease. Furthermore, protease treatment improves crumb characteristics, and the improved flow properties of the dough contribute to better, more uniform product shape.

Enzymes in the group of disulfide reductase/disulfide isomerase are being investigated as possible contributors to controlled and effective gluten modification.

Finally, it was seen that, in bread making, treatments with fungal (or bacterial) enzymes lower the viscosity of bread dough, improving the ease of manipulation by manual workers or machines. Furthermore, favorable effects on taste and crust properties are observed, and the storage characteristic of breads change, giving a product with a softer, more compressible crumb that firms more slowly and keeps longer.

Use in Juice Clarifying

The Fruit Juice Production Process

With certain fruit juices, an initial phase consists of the transformation of plant tissue into a semifluid system made up of cells, fragment of cell walls, and cellular liquid, and a second phase where the juices

are extracted by mechanical separation (pressing, straining, centrifugation).

Once extracted, the juice can be clarified further. Where problems arise in the production process, especially where demand is for limpid (clear) juices, they are for the most part due to the presence of cell wall material, of variable composition, during the ripening process – cellulose, hemicellulose and, at a high degree of esterification, pectin (polymethylgalacturonic acid). In unripe fruit, pectin is present in an insoluble form whereas, on ripening, it is in part broken down into a more soluble form and the fruit becomes soft.

As a result of this partial solubility, some of the pectin passes into the fruit juice, which becomes viscous and hard to separate from the pulp. This explains the low yield from extraction of the juice and of its soluble components, and the fact that juice so extracted is difficult to filter and clarify.

Enzymes in Fruit Juice Clarification

The use of enzymes (Table 4) to overcome the difficulties described is becoming more widespread today with a twofold aim (Figure 4):

- In the treatment of the ground pulp – to raise the juice yield, to improve the extraction rate for certain flavor and color components, and to obtain a partial or total liquefaction of the tissue (for homogenized foods, nectars, juice, and pulp concentrates);
- In the treatment of the juice – to reduce viscosity and facilitate concentration, to clarify in order to obtain a limpid product; and to increase the speed of filtration and thus prevent the formation of precipitates in the juice.

Commercial pectolytic enzymes, commonly referred to as pectinases, are a mixture of pectin esterase (EC 3.1.1.11), which triggers the deesterification of the pectin to pectic acid and methanol; polygalacturonase (EC 3.2.1.15), which hydrolyzes the internal bond of pectic acid; and pectin lyase (EC 4.2.2.10), that splits the glycosidic bonds or either pectate or pectin. These enzymes are chiefly produced from molds of the *Aspergillus* genus. They are used to clarify apple and grape juices and wine. All these preparations contain traces of other enzymes as well, such as cellulase, amylase, protease, and xylanase: when added to the product they bring about the hydrolysis of the soluble pectin and the elimination of its colloidal properties. The process is accompanied by the flocculation and then the removal by filtering or centrifugating of any particles so formed. Traditionally, the industrial utilization of pectic enzymes in fruit juice processing has been conducted in conventional batch reactors using soluble enzymes. Unfortunately, after each cycle of operations the enzymes cannot be recovered for further use, and are inevitably present in the final product, altering organoleptic properties. In this context, the immobilization of pectolytic enzymes has proven to be very advantageous for continuous processing.

Another aspect of using enzymes in fruit juice processing is wine-making; the ideal enzyme preparations for wine-making are different from those for fruit juice processing. In fruit juice processing the enzymes are inactivated very shortly after they have done their job, for example by pasteurization. In wine-making no such heat treatment is imparted. The enzymes therefore maintain their activity over a longer period, and side activities may adversely influence wine quality during storage. During wine-making, enzymatic reactions begin during the ripening of the grapes and continue through the harvest, alcoholic and malo-lactic fermentations, clarification, and even after bottling. The wine-makers still have the power to influence a few of the reactions with the tools of pectinolytic and glycosidic enzymes. These speed up the natural process of wine-making, make the fullest use of facilities and equipment, and improve the quality of wine. Pectic enzymes (from fungi, typically *A. niger, Penicillium notatum* or *Botrytis cinerea*) are used to reduce haze or gelling of grape juice at various stages of the wine-making process. A β-glucanase prepared from a selected strain of *Trichoderma* is now being used in wine-making as an adjuvant for filtering and clarifying purposes of wines made from grapes attacked from the fungus *B. cinerea*. In fact, the *Botrytis* fungus produces β-glucans (polymers of glucose with a high molecular weight), which pass into the wine. These large molecules hinder clarification and rapidly clog filters.

Finally, an improvement of our knowledge of grape composition and of *Saccharomyces cerevisiae* would allow the creation of a 'superyeast strain' by genetic manipulation to depectinize the grape must, to extract and release specific components such as color (especially in the case of red wine) or aromas, or to make the malo-lactic fermentation by heterologous cloning of enzymes from other microorganism.

Use in Cheese Production

Milk Clotting

The entire cheese-making process is based on milk clotting. Clotting can be brought about by natural or induced acidification to pH 4.6 (the isoelectric point for casein) or by enzyme action (Figure 5). The preparation most frequently used for this purpose is rennet, a crude proteolytic extract obtained

Table 4 Classification of pectic enzymes acting on pectins or pectic acids

EC suggested name	*Common name*	*EC number*	*Substrate*	*Action pattern*
Deesterifying enzymes				
Polymethylgalacturonate esterase (PMGE)	Pectinesterase	3.1.1.11	Pectin	Random
Depolymerizing enzymes				
Hydrolases				
Endopolygalacturonase (endo PG)	Polygalacturonase	3.2.1.15	Pectate	Random
Exopolygalacturonase 1 (exo PG1)	Polygalacturonase	3.2.1.67	Pectate	Terminal
Exopolygalacturonase 2 (exo PG2)	Polygalacturonase	3.2.1.82	Pectate	Penultimate bonds
Endopolymethylgalacturonase (endo PMG)	Pectin hydrolase		Pectin	Random
Exopolymethylgalacturonase (exo PMG)	Pectin hydrolase		Pectin	Terminal
Lyases				
Endopolygalacturonate lyase (endo PGL)	Pectate lyase	4.2.2.2	Pectate	Random
Exopolygalacturonate lyase (exo PGL)	Pectate lyase	4.2.2.9	Pectate	Penultimate bonds
Endopolymethylgalacturonate lyase (endo PMGL)	Pectin lyase	4.2.2.10	Pectin	Random
Exopolymethylgalacturonate lyase (exo PMGL)	Pectin lyase		Pectin	Terminal

Pectic enzymes acting on oligogalacturonates have not been included in this table because they are not very abundant and are of little interest for industrial pectin degradation. Enzymes have been classified and named according to the Enzyme Commission (EC) (IUPAC-IUB recommendations). Data from Alkorta I, Garbisu C, Liama MJ and Serra JL (1998) Industrial applications of pectic enzymes: a review. *Process Biochemistry* 33: 21–28.

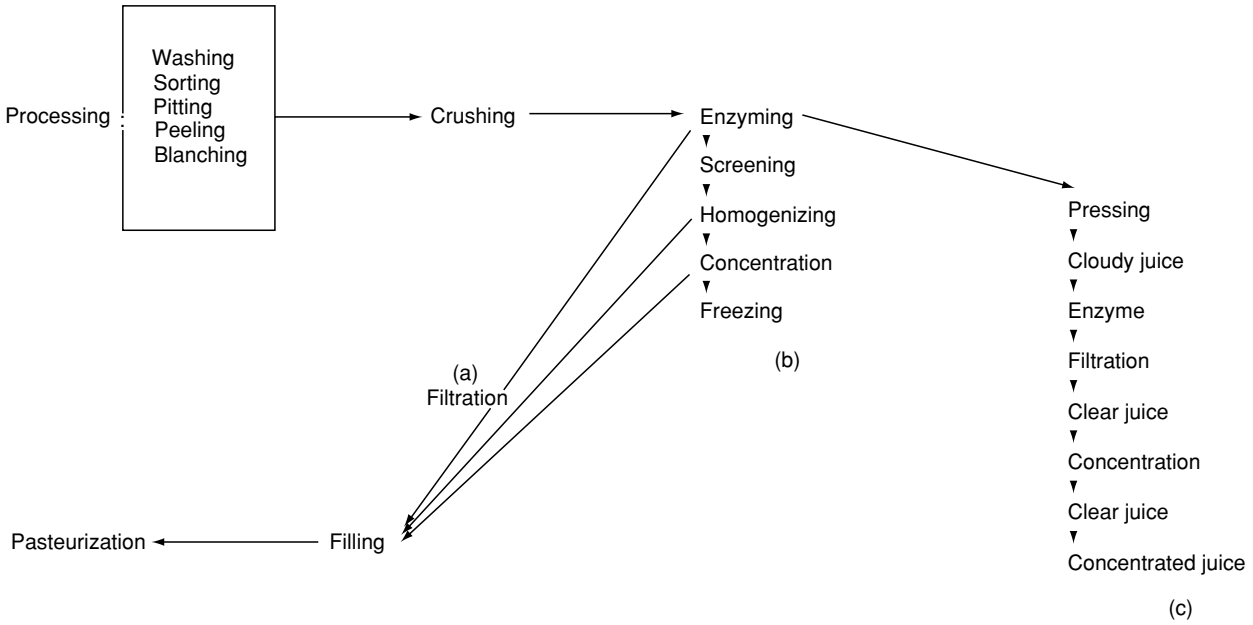

Figure 4 Use of enzymes in the preparation of (a) juices, (b) homogenates, and (c) clear juices.

from the fourth stomachs of calves, piglets, lambs, and kids and containing 85–95% chymosin and 10–15% pepsin.

In milk, primary soluble proteins are the whey proteins α-lactalbumin and β-lactoglobulin. The insoluble proteins found in the colloidal particles include casein micelles. κ-Casein is a calcium-insensitive protein which forms a protective layer around the calcium-sensitive caseins (α_s1-, α_s2-, β-, and γ-), resulting in stable casein micelles.

Chymosin, or rennin (EC 3.4.23.4), triggers the clotting of milk once the equilibrium of charges in the casein micelle has been destabilized. It causes the cleavage of κ-casein at the Phe 105–Met 106 bond, which results in a release of hydrophilic glycopeptide which passes into the whey, while the *para*-κ-casein remains in the micelles, forming the coagulum needed to produce cheese. (*See* **Milk**: Physical and Chemical Properties.)

Milk-Clotting Enzymes

With the marked expansion in world cheese production registered since the 1960s, it has become increasingly difficult to obtain stomachs of young ruminants

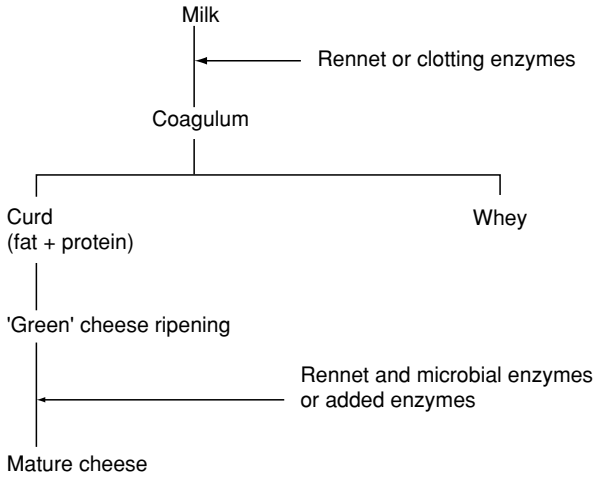

Figure 5 Flow diagram for cheese manufacturing.

from which chymosin can be extracted. In the USA in particular, a general reduction in the number of calves available for slaughter has reduced the supply of rennet, increased the price, and stimulated interest in developing rennet substitutes. Several rennet substitutes have been developed, including bovine rennet from adult cows, fungus proteinase, and other proteolytic enzymes. However, they have a much greater level of nonspecific proteolytic activity, and higher thermostability, which more completely degrade the milk proteins to peptides, leading to a reduction in yield and poor flavor development in some types of cheese. Consequently, there have been numerous attempts to produce chymosin in microorganisms.

Several of these substitute preparations (Table 5) are a mixture of rennet and pepsin but newly developed recombinant DNA techniques have led to their use in cloning the calf chymosin gene in *Escherichia coli*. Recombinant chymosin contains only the protein resulting from the expression of the sequence of one gene and thus contains one variant only. Patents have been taken out for the production of milk-clotting enzymes from *Endotia parasitica, Mucor pusillus* var. Lindt and *Mucor miehei*. Varieties of cheese have been made from recombinant chymosin and evaluated in comparison with cheese produced using the natural enzyme. No significant differences could be detected between them, with regard to the recovery of milk solids, the rate of proteolysis during ripening, or in the characteristics of the final cheese product. Under US law, these enzymes may be used for making any type of cheese.

Cheese Ripening

Although some varieties of cheese are consumed fresh, most varieties that have been subjected to rennet clotting are ripened for periods ranging from 4 weeks to 2 years, the actual duration being in an approximately inverse relation to the humidity content. During the ripening process, a series of chemical and biological changes takes place, in the course of which proteins, lipids, and residual lactose break down into primary and secondary products, such as peptides and fatty acids (lactic, propionic, etc.).

In the right combination, all these components converge to produce the characteristic flavor of the respective cheeses. The cost entailed in the ripening operation has a considerable influence on the cost of the finished product. Accordingly, if it is possible to shorten the ripening time without adversely affecting quality, the capital outlay required can be contained, not to mention the risk of freak retarded fermentation phenomena. (*See* **Fermented Milks**: Types of Fermented Milks; Other Relevant Products.)

Accelerated Ripening Enzymes and Use of Lipases

The ripening time can be reduced by:

- adding, during the manufacturing process, enzymes which are not produced by lactic bacteria;
- introducing additional lactic bacteria (as starters) or extracts of lactic bacteria (as starters) or extracts of lactic bacteria cells;
- using slurry techniques.

Directly incorporating enzymes into the milk undergoing treatment can be costly but has the advantage of insuring their homogeneous distribution. Proteases, peptidases, lipases, and decarboxylases are used for this purpose (Table 6). While the ripening of certain varieties (blue, Parmesan, etc.) is markedly influenced by lipase action, proteolysis is in all types of cheese the primary biochemical event in determining flavor and texture. Neutral proteases of bacterial origin such as neutrase (a metal protease, effective in speeding up the ripening process whether or not inserted into positively charged liposomes) or, again, a serine protease obtained from *Brevibacterium linens* (the main component of surface microflora on cheese), have proved to yield an acceptable product even if elasticity is somewhat sacrificed at times. Moreover, if the crude enzyme containing the protease (but amino peptidase as well) is used, better results can be achieved in terms of the release of peptides and amino acids without any bitter taste developing.

Studies on food and analytical-grade commercial lipase have shown that these have a different specificity in the hydrolysis of short-chain fatty acids and a different specificity in the release of the free fatty acids among them. In fact, depending on the chain-length specificity of a given lipase, its addition to a milk product may enhance the flavor of the cheese,

Table 5 Enzymes in milk clotting

Enzyme	*IUB number enzyme classification (EC)*	*Type and source*[a]
Calf, Piglet, and lamb rennet (chymosin)	3.4.23.4	(A) Stomach
Mammalian rennet		
pepsin A	3.4.23.1	(A) Mammalian stomach
Pepsin B	3.4.23.2	
Gastricsin	3.4.23.3	
Chymosin	3.4.23.4	
Pig (porcine pepsin)	3.4.23.1	(A) Pig stomach
Blends of two or more animal-derived enzymes		(A)
Blends of animal-derived/microbial-derived enzymes		(A) + (M)
Microbial coagulant	3.4.23.6	(M) *Mucor miehei*
Microbial coagulant	3.4.23.6	(M) *Mucor pusillus* Lindt
Microbial coagulant	3.4.23.6	(M) *Endotia parasitica*
Microbial coagulant	3.4.23.6	(M) *Penicillum janthinellum*
Microbial coagulant	3.4.23.6	(M) *Rhizopus chinensis*
Microbial coagulant	3.4.23.6	(M) *Aspergillus niger*

[a]A, animal; M, microbial.
IUB, International Union of Biochemistry; EC, Enzyme Classification.

Table 6 Enzymes in accelerated cheese ripening

Enzyme	*IUB number enzyme classification (EC)*	*Type and source*[a]
Lipase	3.1.1.3	(A) Edible forestomach tissue of calf, kid, or lamb
Microbial lipase	3.1.1.3	(M) *Aspergillus niger*
Microbial lipase	3.1.1.3	(M) *Aspergillus oryzae*
Microbial lipase	3.1.1.3	(M) *Mucor mihei*
Lactase	3.2.1.23	(M) *Saccharomyces lactis* *Kluyveromyces* spp. *Escherichia coli*
Microbial serine proteinase	3.4.21.14	(M) *Aspergillus niger*
Neutral proteinase	3.4.24.4	(M) *Bacillus subtilis*
Neutral proteinase	3.4.24.4	(M) *Aspergillus oryzae*
Proteinase	3.4	(M) *Penicillium camemberti*
Proteinase	3.1.1.1	(A) Pancreatic tissue
		(M) *Micrococcus caseolyticus*
Proteinase	3.4.24.4	(M) *Bacillus subtilis* *Aspergillus niger*

[a]A, animal; M, microbial.
IUB, International Union of Biochemistry; EC, Enzyme Classification.

accelerate the cheese ripening, or assist in the preparation of enzyme-modified cheese (EMC), an important commercial flavor used in the USA for the manufacture of dips, sauces, dressings, and crackers. EMC is produced from cheese curd by the addition of lipases at elevated temperatures, increasing the content of free fatty acids about 10-fold. Depending on the chain-length specificity of the lipases used, there is either predominant liberation of short-chain fatty acid (C4–C6), which results in a sharp, tangy flavor, or of medium- and long-chain fatty acids (>C12), which convey a soapy taste and are readily metabolized to other flavor ingredients. Addition of lipase to pasteurized cows' milk can generate flavors similar to goat, sheep, or raw milk if used with the appropriate microbial consortia.

Currently, lipase technology is so advanced that the production of compounds such as flavors or emulsifiers are regarded as an old application. Commercially available lipases are usually derived from microorganisms, but since the advent of genetic engineering techniques, an increasing number of lipases are being commercially manufactured from recombinant bacteria and yeast. Some industrial applications of microbial lipases are reported in Table 7.

Use in Brewing

Traditionally, beer is produced by mixing crushed barley malt and hot water in a large circular vessel called a mash copper. This process is called 'mashing.' Besides malt, other starchy cereals such as maize, sorghum, rice, and barley, or pure starch itself, are added to the mash. These are known as adjuncts. After mashing, the mash is filtered in a lauter tun. The liquid, known as 'sweet wort,' is then run off to the copper, where it is boiled with hops. The 'hopped wort' is cooled and transferred to the fermenting vessels where yeast is added. After fermentation, the so-called 'green beer' is matured before the final filtration and bottling. (*See* **Beers**: Biochemistry of Fermentation.)

This is a much-simplified account of how beer is made. Brewing can quite happily proceed without any added enzymes, relying entirely on the natural enzymes from the ingredients and the yeast to

produce good-quality beer. From time to time, however, problems occur that cannot be easily identified or rectified. Enzymes can often provide a useful diagnostic tool, giving the immediate remedy as well as suggesting the long-term solution. Table 8 shows some of the problems that can occur during brewing and recommends some of the possible enzyme treatments that can be used to solve the problem. It also serves as a summary of the application of enzymes in brewing, together with Figure 6, which shows where the appropriate enzymes are usually used in the brewing process. Supplementation can in fact occur at various stages during the process. During the initial step of decoction and mash formation, in which probably 80% of the original starch is converted into fermentable sugars (glucose, maltose, maltotriose), bacterial and fungal α-amylases are often added, together with bacterial and fungal glucanases. The latter are important in the removal of β-glucans, β-1,3, and β-1, 4-linked polymers, which are viscous gum-like substances. Failure to hydrolyze the glucans introduces problems in the efficiency of wort filtration and in the later production of hazes. Further amylase and glucanase addition may occur during the filtration of the wort. During subsequent fermentation, particularly for low-calorie beers, where soluble polysaccharide levels need to be minimized, fungal α-amylases and glucoamylases are added.

A typical example of the use of enzymes to resolve a brewing problem is when a haze appears in the bright filtered beer (this includes the enigmatic invisible haze, which cannot be detected by the naked eye, but is recorded as a machine haze). The haze is usually ascribed to three probable causes: proteins, starch/polysaccharides, or nonstarch polysaccharides. It may result from other factors such as the passage of diatomaceous earth, oxalic acid crystals, microorganisms, or filter fibers, which can be easily identified under a microscope or using an enzyme system. In fact, adding β-glucanase, papain, and fungal

Table 7 Industrial application areas for microbial lipases

Industry	*Effect*	*Product*
Dairy	Hydrolysis of milk fat	Flavor agents
	Cheese ripening	Cheese
	Modification of butter fat	Butter
Bakery	Flavor improvement and shelf-life prolongation	Bakery products
Beverages	Improved aroma	Beverages
Food dressing	Quality improvement	Mayonnaise, dressings, and whipped toppings
Health food	Transesterification	Health foods
Meat and fish	Flavor development and fat removal	Meat and fish products
Fat and oil	Transesterification	Cocoa butter, margarine
	Hydrolysis	Fatty acids, glycerol, mono-, and diglycerides

Modified from Godtfredesen SE (1993) Lipases. In: Nagodawithana T and Redd G (eds) *Enzymes in Food Processing*, pp. 205–214. California: Academic Press.

Table 8 A brewer first-aid kit

Location	*Symptom*	*Remedy*
Cereal cooker	Glutinous starch	Add heat-stable bacterial α-amylase
	Retrograded starch	
Mash mixer	Enzyme-deficient malt	Add bacterial α-amylase
	Filtration problems	
	High malt glucans	Add heat-stable fungal β-glucanase
	Poor extract recovery	
	High adjunct brewing	Add neutral proteinase
Fermentation	Starch in beer	Add fungal α-amylase
	Require highly attenuated beers	Add amyloglucosidase or pullulanase and β-amylase
	Rapid diacetyl removal	Add α-acetolactate decarboxylase
Maturation and filtration	Low sweetness	Add amyloglucosidases
	Fuel for secondary fermentation	
	Chill haze	Add papain
Bottling	Poor resistance to staling in bottle	Add immobilized glucose oxidase in the crown liner

Partially reproduced from O'Rourke T (1996) Brewing. In: Godfrey T and West SI (eds) *Enzymology*, pp. 103–131. London: MacMillan Press with permission.

α-amylase or amyloglucosidase to some samples and incubating them at 28–30 °C for 48 h, it can be seen whether any treatment clears the sample. If it works, the problem is identified and an easy solution has been found.

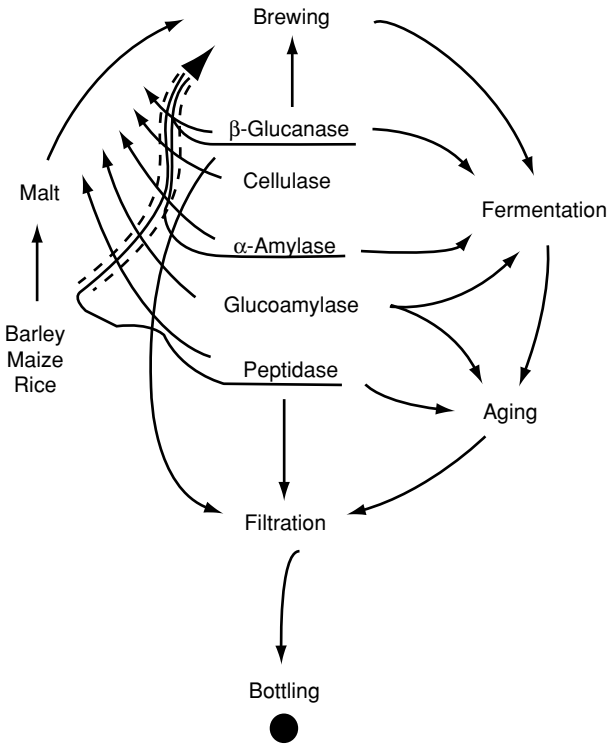

Figure 6 Enzymes used in brewing. Data from Bigelis R (1993) Carbohydrases. In: Nagodawithana T and Redd G (eds) *Enzymes in Food Processing*, pp. 121–147. California: Academic Press.

Use of Transglutaminase in Functional Properties of Food

Transglutaminase (EC 2.3.2.13) catalyzes *in vitro* cross-linking in whey proteins, soya proteins, wheat proteins, beef myosin, casein and crude actomyosin refined from mechanically deboned poultry meat. In recent years, on the basis of the enzyme's reaction to gelatinize various food proteins through the formation of cross-links, this enzyme has been used in attempts to improve the functional properties of food. The modification of food proteins by transglutaminase may lead to textured products, help to protect lysine in food proteins from various chemical reactions, encapsulate lipids and/or lipid-soluble materials, form heat- and water-resistant films, avoid heat treatment for gelation, improve elasticity and water-holding capacity, modify solubility and functional properties, and produce food proteins of higher nutritional value through cross-linking of different proteins containing complementary limiting essential amino acids.

An overview of the application possibilities for microbial transglutaminase in food processing is given in Table 9. In meat and fish processing it is of great interest to maximize the yield of marketable products, particularly by restructuring low-value cuts and trimmings to improve their appearance, flavor, and texture. In vegetables and fruit transglutaminase is used to maintain freshness, by coating them with a membrane containing transglutaminase and proteins. Literature reports also cite a method developed for reducing the allergenicity of some food proteins and/or peptides: α_{s1}-casein was treated

Table 9 Overview of application of microbial transglutaminase in food processing

Source	*Product*	*Effect*
Meat	Hamburger, meatballs, stuffed dumplings	Improved elasticity, texture, taste, and flavor
	Canned meat	Good texture and appearance
	Frozen meat	Improved texture and reduced cost
	Molded meat	Restructuring of meat
Fish	Fish paste	Improved texture and appearance
Krill	Krill paste	Improved texture
Collagens	Shark-fin imitation	Imitation of delicious food
Wheat	Baked foods	Improved texture and high volume
Soya bean	Mapuo doufu	Improved shelf-life
	Fried tofu (aburaage)	Improved texture
	Tofu	Improved shelf-life
Vegetables and fruits	Celery	Food preservation
Casein	Mineral absorption promoters	Improved mineral absorption in intestine
Gelatin	Sweet foods	Low-calorie foods with good texture, firmness, and elasticity
Fat, oil, and proteins	Solid fats	Pork-fat substitute with good taste, texture, and flavor
Plant proteins	Protein powders	Gel formation with good texture and taste
Seasonings	Seasonings	Improve taste and flavor

Partially reproduced from Zhu Y, Rinzern A, Tramper J and Bol J (1995) Microbial transglutaminase – a review of its production and application in food processing. *Applied Microbiology and Biotechnology* 44: 277–282 with permission.

with transglutaminase to manufacture cross-linked casein, which was less allergenic.

In developing countries, many people are still suffering from starvation and efforts are being focused on producing acceptable protein foods from nonanimal protein, to solve the problem of protein deficiency. On the other hand, in addition to their awareness of health problems caused by obesity, people in developed countries are increasingly aware of the environmental burden caused by surplus livestock. When presented with a novel food product, consumers are very sensitive to properties such as flavor, nutritional value, appearance, shelf-life, and palatability. In this respect, protein modification by enzymes, especially by microbial transglutaminase, the mass production of which can be achieved by fermentation from cheap substrates, is one of the most promising alternatives in developing novel protein foods.

Up to now, commercial transglutaminase has been merely obtained from animal tissues. The complicated separation and purification procedure results in an extremely high price for the enzyme. The production of transglutaminase derived from microorganisms was not reported until the late 1980s, and recently a modified downstream process for purifying microbial transglutaminase was described. The microbial process no doubt has an advantage in its independence from regional and climatic condition, in addition to its reasonable cost. But it is still of great interest to improve fermentation and downstream processing to reduce production cost. Modification of strains by genetic engineering is one of the alternatives.

Cleaning

A major use of enzymes as proteases has been in the production of detergents. These are used widely in the food industry, especially for cleaning-in-place applications, where the equipment is sensitive to heat. (*See* **Cleaning Procedures in the Factory**: Types of Detergent.)

Pancreatin (extracted from the pancreatic glands of animals), which contains the protein-degrading enzyme trypsin, was used as early as 1913 in laundry detergents. The detergent, being mainly composed of ordinary washing soda (sodium carbonate), had an excellent softening effect on water and helped to dissolve dirt. The trypsin, however, had only a limited effect.

During the 1960s, there was a significant new development in the detergent industry when a mixture of serine and microbial proteases, which appeared to be less degradable by soaps, alkalis and temperature, replaced trypsin. These enzymes have a serine-histidine active site and produce their effect in the pH range 6.5–10.0 and a temperature range of 30–60 °C.

There was a temporary setback in the early 1970s when it was realized that enzymes could cause allergic reactions among the staff working in the enzyme plant, in detergent factories, or the end users.

The introduction of encapsulated proteases in the 1980s and the withdrawal of the US Federal Trade Commission's restrictions were followed by a resurgence in the demand for proteases in detergents. Proteases remove protein stains and deposits from a wide range of food sources. These organic stains have a tendency to adhere strongly to textile fibers and equipment surfaces. The proteins act as glues, preventing water-borne detergent systems from removing some of the other components of the stain/deposit, such as pigments and particulate material. Nonenzymatic detergents are inefficient in removing proteins and can result in permanent stains due to oxidation and denaturing caused by bleaching and drying.

Proteases hydrolyze proteins and break them down into more soluble polypeptides. Through the combined effect of surfactants and enzymes, stubborn stains/deposits can be removed. Though protein stains can be easily digested by enzymes, oil and fatty stains have always been troublesome to remove. The trend towards lower cleaning temperatures has made the removal of grease spots an even greater problem.

Amylases are also used to remove residues of starchy foods, such as mashed potato, spaghetti, oatmeal porridge, custard, gravy, and chocolate.

See also: **Beers**: Biochemistry of Fermentation; **Biscuits, Cookies, and Crackers**: Methods of Manufacture; **Bread**: Breadmaking Processes; Chemistry of Baking; **Cakes**: Methods of Manufacture; Chemistry of Baking; **Cheeses**: Starter Cultures Employed in Cheese-making; Mold-ripened Cheeses: Stilton and Related Varieties; Surface Mold-ripened Cheese Varieties; **Enzymes**: Functions and Characteristics; Uses in Analysis; **Fermented Milks**: Types of Fermented Milks; Other Relevant Products; **Wines**: Production of Table Wines; Production of Sparkling Wines

Further Reading

Alkorta I, Garbisu C, Liama MJ and Serra JL (1998) Industrial applications of pectic enzymes: a review. *Process Biochemistry* 33: 21–28.

Bigelis R (1993) Carbohydrases. In: Nagodawithana T and Redd G (eds) *Enzymes in Food Processing*, pp. 121–147. California: Academic Press.

Chitpinityol S and Crabbe MJ (1998) Chymosin and aspartic proteinases. *Food Chemistry* 61: 395–418.

Cowan DA (1994) Industrial enzymes. In: Moses V and Cape RE (eds) *Biotechnology*, pp. 311–340. UK: Harwood Academic Publisher.

Godfrey T and West SI (1996) Introduction to industrial enzymology. In: Godfrey T and West SI (eds) *Enzymology*, pp. 1–8. London: MacMillan Press.

Godtfredesen SE (1993) Lipases. In: Nagodawithana T and Redd G (eds) *Enzymes in Food Processing*, pp. 205–214. California: Academic Press.

Goll DE, Robson RM and Stromer MH (1977) Muscle proteins. In: Whitaker JR and Tannenbaum SR (eds) *Food Proteins*, pp. 121–174. Westport: AVI.

Guzman-Maldonado H and Paredes-Lopez O (1995) Amylolytic enzymes and products derived from starch: a review. *Critical Reviews in Food Science and Nutrition* 35: 373–403.

Lanzarini G and Pifferi PG (1990) Enzymes in fruit juice industry. In: Cantarelli C and Lanzarini G (eds) *Biotechnology Applications in Beverage Production*, pp. 189–221. London: Elsevier.

Linko P (1987) Enzymes in the industrial utilization of cereals. In: Kruger JE, Lineback D and Stauffer CE (eds) *Enzymes and their Role in Cereal Technology*, pp. 357–374. St Paul: AACC.

Mathewson PR (2000) Enzymatic activity during bread baking. *Cereal Foods World* 45: 98–101.

O'Rourke T (1996) Brewing. In: Godfrey T and West SI (eds) *Enzymology*, pp. 103–131. London: MacMillan Press.

Schmid RD and Verger R (1998) Lipases: interfacial enzymes with attractive applications. *Angewandthe Chemie International Edition* (in English) 37: 1608–1633.

Whitaker JR (1982) Enzymes of importance in high protein foods. In: Dupuy P (ed.) *Proceedings on the Use of Enzymes in Food Technology*, pp. 329–358. Paris: Lavoisier.

Zhu Y, Rinzem A, Tramper J and Bol J (1995) Microbial transglutaminase – a review of its production and application in food processing. *Applied Microbiology and Biotechnology* 44: 277–282.

Uses in Analysis

J R Whitaker, University of California, Davis, CA, USA

Introduction

Enzymes are ideal for analyses because of their high specificity and sensitivity. Any compound that affects the activity of an enzyme, i.e., substrate, inhibitor, or activator, can be determined rapidly and quantitatively by adding such compounds to a controlled enzyme reaction and comparing the results without the compound (control). Also, measurements of activities of selected enzymes in a biological specimen under standard conditions are useful in medicine for the diagnosis and prognosis of almost all diseases. More recently they have been used for determining specific genetically related diseases. They are also used for assessing nutritional deficiencies, physical damage to organs, suitability for and adequacy of processing and quality of foods, identifying subcellular organelles and membrane structures, and structural configurations of molecules. Enzymes are ideal reporters for measuring antigens to specific antibodies (enzyme immunoassays).

Enzyme usage in analyses is not new. Peroxidase was used for hydrogen peroxide analysis as early as 1845. Enzymatic carbohydrate determination in biological materials was an accepted method by the end of the nineteenth century. But most progress in using enzymes analytically has occurred in the last 20 years as a result of readily available pure enzymes, better understanding of the factors which affect enzyme activity, enzyme kits for specific determinations, easily usable sensitive techniques, and instrumentation matching the specificity and sensitivity of the enzymes. The rapid progress in sequencing genes, including the human genome, is due to two specific types of enzymes: the restriction endonucleases and the polymerases.

Principles and Applications

Compounds to be measured must bind stereospecifically at the active site of an enzyme or another specific location on the enzyme such that the activity of the enzyme is affected (eqn 1).

$$\begin{array}{ccccccccc} E & + & S & \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} & E{\cdot}S & \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} & E{\cdot}P & \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} & E + P \\ \Big\| X & & \Big\| K_d & & \Big\| X & & \Big\| K_d & & \\ E{\cdot}X & + & S & \underset{k'_{-1}}{\overset{k'_1}{\rightleftharpoons}} & E{\cdot}S{\cdot}X & \underset{k'_{-2}}{\overset{k'_2}{\rightleftharpoons}} & E{\cdot}P{\cdot}X & \underset{k'_{-3}}{\overset{k'_3}{\rightleftharpoons}} & E{\cdot}X + P \end{array} \quad (1)$$

E, enzyme; *S*, substrate; *E·S,* enzyme-substrate complex; *E·P,* enzyme-product complex; *P,* product; *X,* inhibitor or activator; *E·X,* enzyme-inhibitor (or activator) complex; *E·S·X,* enzyme-substrate-inhibitor (or activator) complex; *E·P·X,* enzyme-product-inhibitor (or activator) complex; K_d, dissociation constant; k_1, k_2, k_3, k'_1, k'_2 and k'_3 are forward rate constants.

There must be a predictable relationship between the concentration of a compound and its effect on enzymatic activity. Standard curves for compounds and

their products, including enzymes, to be determined are required in order to obtain absolute concentrations. In preparing the standard curve of initial velocity, v_o, versus compound or product concentration, purity of the analyte is essential and exact reproduction of experimental conditions is required. In the total change method (below), strict adherence to experimental conditions (such as rate of reaction) is not required, but it is essential to have pure standard compounds and products and to insure the reaction has gone to completion.

Determination of Substrate Concentration

The concentration of any compound that serves as a substrate, S, for an enzyme can be determined quantitatively by its effect on the initial velocity, v_o, of the reaction, provided K_m and V_{max} are known, as shown in eqn 2, and provided $[S]_o < 20\,K_m$:

$$v_o = V_{max}[S]_o/(K_m + [S]_o) \qquad (2)$$

Best results are obtained at $[S]_o < 5\,K_m$. This is called the rate (or kinetic) assay method.

The concentration of a compound that is a substrate can also be determined by a total change method when all the substrate is converted to product, which is then measured. It is essential to insure that all substrate has been converted to product, which may take several hours at the concentration of enzyme used in rate assays. Therefore, higher concentrations of enzymes are used in total change method assays than in rate assays. The total change method cannot be used for reactions that reach an equilibrium between substrate and product unless the product is trapped, forcing the reaction to completion. An example is alcohol dehydrogenase (alcohol:NAD^+ oxidoreductase, EC 1.1.1.1) acting on ethanol and NAD^+, as shown in eqn 3. Hydroxylamine is used to trap ethanal as the oxime, forcing the reaction to completion (eqn 4). The NADH concentration is measured spectrophotometrically.

$$CH_3CH_2OH + NAD^+ \rightleftharpoons CH_3CHO + NADH + H^+ \qquad (3)$$

$$CH_3CHO + NH_2OH \rightleftharpoons CH_3CHNOH + H_2O \qquad (4)$$

Special care must be taken in determining concentrations of substrates which exist in more than one anomeric form, in equilibrium with each other. An example is the mutarotation of reducing sugars. Glucose at equilibrium exists in the α, β, and open-chain forms in the ratio of 33:66:1, respectively. Glucose oxidase, often used to determine glucose concentrations, has strong preference for the β-D-glucose anomer, versus the α-D-glucose anomer (156:1 relative rate of oxidation). Therefore, both the total change method and the rate assay method may not give the correct concentration, unless a standard curve for glucose is established under the exact same conditions.

Some examples of compounds that can be quantitatively determined because they are enzyme substrates are shown in Table 1. Only a very few examples are given, since, with more than 1000 enzymes with individual specificities, at least that number of compounds can be uniquely determined. (*See* **Fructose**; **Lactose**; **Starch**: Structure, Properties, and Determination.)

Advantages of enzyme analyses for substrates, versus other types, include: (1) the very high specificity, permitting isomers to be distinguished; (2) no need for partial purification; and (3) the ability to work at room temperature, or below, with thermolabile compounds. An example will illustrate this. Reducing sugar methods, such as alkaline copper solutions, ferricyanide or 3,5-dinitrosalicylic acid for glucose, will determine all reducing sugars in the solution, plus other reducing compounds. However, glucose oxidase, which catalyzes the reaction

Table 1 Analytical determination of compounds that are substrates[a]

Compound	*Enzyme used*	*Analysis method*
Ammonia	Glutamate dehydrogenase	S; NADH → NAD^+
Ethanol	Alcohol dehydrogenase	S; NAD^+ → NADH
D-Fructose	Hexokinase	S; $NADP^+$ → NADPH, in coupled assay
D-Glucose	Glucose oxidase	S; H_2O_2 with peroxidase
L-Lactic acid	L-Lactate dehydrogenase	S; NAD^+ → NADH
D-Lactic acid	D-Lactate dehydrogenase	S; NAD^+ → NADH
Lactose	β-Galactosidase	S; H_2O_2 with peroxidase and glucose oxidase
Starch	Amyloglucosidase	S; NADH → NAD^+, in coupled assay through glucose and glucose oxidase
Urea	Urease	E; NH_3 electrode

[a]S, spectrophotometric method; E, electrometric method.

Adapted from Guilbault GG (1984) *Analytical Uses of Immobilized Enzymes*. New York: Marcel Dekker.

shown in eqn 5, will only determine glucose. Furthermore, it is absolutely specific for D-glucose versus L-glucose and oxidizes β-D-glucose 156 times more rapidly than α-D-glucose. The reaction shown in eqn 6 is not catalyzed by glucose oxidase. The rate is slower than that of eqn 5 and therefore cannot be used to determine glucose concentration by the rate method. Therefore, the H_2O_2 concentration is determined by use of peroxidase and a chromagen that gives a colored compound that is measured spectrophotometrically (coupled reactions). The peroxidase and chromagen concentrations must be high enough so that the rate of reaction shown in eqn 5 is rate-determining, not the reaction shown in eqn 7.

$$\beta\text{-D-Glucose} + O_2 \underset{\text{oxidase}}{\overset{\text{Glucose}}{\rightleftharpoons}} \beta\text{-D-Glucone-}\delta\text{-lactone} + H_2O_2 \quad (5)$$

$$\beta\text{-D-Glucone-}\delta\text{-lactone} + H_2O \longrightarrow \beta\text{-D-Gluconic acid} \quad (6)$$

$$H_2O_2 + \text{Chromagen} \xrightarrow{\text{peroxidase}} \text{Colored compound} \quad (7)$$

Determination of Inhibitor Concentration

Any compound that decreases v_o when added to an enzyme–substrate system is an inhibitor. Compounds that are reversible inhibitors can only be determined by rate assay methods, while irreversible inhibitors can be determined by either rate assays or total change assays. When X in eqn 1 binds only to E, so as to decrease enzyme activity (i.e., both $E{\cdot}S$ and $E{\cdot}X$ are formed), it is a competitive inhibitor; k'_1 is zero. Therefore, the concentration of the compound can be determined from the ratio of v'_o/v_o, where v'_o is the initial velocity determined in the presence of a fixed concentration of inhibitor, [I], and v_o is the initial velocity determined in the absence of inhibitor provided K_m, [S], and K_i (= [E][I]$_o$/[EI]; K_d in eqn 1) for the inhibitor are known. This relationship is shown as:

$$v'_o/v_o = (K_m + [\mathbf{S}]_o)/\{(1 + [\mathbf{I}]_o/K_i)/K_m + [\mathbf{S}]_o\} \quad (8)$$

If X binds to E and $E{\cdot}S$ equally well (eqn 1) and k'_1 and k'_2 are zero, the X is a linear noncompetitive inhibitor. [I]$_o$ can be determined from:

$$v'_o/v_o = 1 + [\mathbf{I}]_o/K_i \quad (9)$$

provided K_i for the inhibitor is known.

If X binds only to $E{\cdot}S$ (eqn 1) and k'_2 is zero, then X is a linear uncompetitive inhibitor. [I]$_o$ can be determined from:

$$v'_o/v_o = (K_m + [\mathbf{S}]_o)/\{K_m + [\mathbf{S}]_o(1 + [\mathbf{I}]_o/K_i)\} \quad (10)$$

provided K_m, [S]$_o$, and K_i are known.

If X reacts chemically with E to form an irreversible product, E-X, then the extent of inhibition is linearly related to the concentration of X as long as [X] < [E] and the reaction is allowed to go to completion (total change method). The concentration of X can be calculated, provided [E]$_o$ is known by:

$$v'_o/v_o = ([\mathbf{E}]_o - [\mathbf{E\text{-}X}])/[\mathbf{E}]_o \quad (11)$$

where [E-X] = [X] and [E]$_o$ is concentration of total active enzyme used. Table 2 gives a few examples of compounds that can be determined enzymatically because they are enzyme inhibitors, thereby decreasing v_o when added to an enzyme–substrate system.

Theoretically, any compound that is a substrate can be determined by inhibition kinetics since two substrates (one at a fixed concentration, while the unknown substrate concentration is variable) in the system compete for binding at the active site of the enzyme, based on their relative K_m values. One need only determine v_o for one of the substrates (control), and treat the second substrate as a competitive inhibitor to determine v'_o. (*See* **Aluminum (Aluminium)**: Properties and Determination; **Ascorbic Acid**: Properties and Determination; **Lead**: Properties and Determination; **Mercury**: Properties and Determination.)

Determination of Activator Concentration

Any compound that increases v_o when added to an enzyme-substrate system is an activator. Compounds

Table 2 Analytical determination of compounds that are inhibitors[a]

Compound	*Enzyme used*	*Analysis method*
Pesticides	Cholinesterase	E; H^+; thiocholine
DDT	Carbonic anhydrase	S; H^+ with indicator
Heparin	Ribonuclease	S; degradation of RNA
Ascorbic acid	Catalase	S; uric acid
Ag^+	Alcohol dehydrogenase (yeast)	S; $NAD^+ \rightarrow$ NADH
Al^{3+}	Alkaline phosphatase	S; *p*-nitrophenol
Hg^{2+}	Alcohol dehydrogenase	S; $NAD^+ \rightarrow$ NADH
Pb^{2+}	Peroxidase	F; fluorescent product
F^-	Liver esterase	E; H^+
S^{2-}	Peroxidase	F; fluorescent product

[a]The methods listed for the anions may not be absolutely specific: similar types of ions may interfere.
DDT, dichlorodiphenyltrichloroethane; E, electrometric method; F, fluorescent method; S, spectrophotometric method.
Adapted from Guilbault GG (1984) *Analytical Uses of Immobilized Enzymes*. New York: Marcel Dekker.

that form a rapid equilibrium complex with an enzyme (and/or $E{\cdot}S$) as shown in eqn 1 can only be determined by rate assay methods. This is characteristic of coenzymes, such as NAD^+, $NAD(P)^+$, nucleotide triphosphate coenzymes adenosine triphosphate (ATP), uridine 5′-triphosphate, cytidine 5′-triphosphate) and some metal ions. At a fixed concentration of enzyme and substrate, the concentration of activator, A, can be determined from:

$$v_o^o = v_o^2 - v_o = V_{\max}[\mathbf{A}]/(K_d + [\mathbf{A}]) \qquad (12)$$

provided $K_d\,(=[\mathbf{E}][\mathbf{A}]/[\mathbf{E}{\cdot}\mathbf{A}])$ and $V_{\max}$ are known. v_o^2 is initial velocity in the presence of the activator. In most cases, v_o in the absence of coenzyme is zero, so that $v_o^0 = v_o^2$ is observed; this is the increase in activity in presence of a given concentration of A. (*See* **Coenzymes**.)

Activators that form tight complexes ($K_d < 10^{-9}$ mol l^{-1}) with enzyme, such as prosthetic groups and some metal ions, increase activity of the enzyme in a linear fashion (v_o^o is directly proportional to $[\mathbf{A}]_o$). Therefore, only a total change method is useful.

It is essential to prepare a standard curve of v_o versus $[\mathbf{A}]_o$ in order to convert measured v_o values to absolute concentration. Purity of the standard activator and reproduction of experimental conditions are essential.

Table 3 gives a few examples of compounds that activate enzymes. Several inferences may be drawn from the examples presented. With the exception of O_2, all the compounds listed are cofactors. They are absolutely required for enzymatic activity.

One enzyme may be used to analyze for several compounds. For example, luciferase can determine luciferin (substrate), ATP, Mg^+, and O_2 concentrations, since all affect the activity of the enzyme (eqn 13).

Table 3 Analytical determination of compounds that are activators[a]

Compound	*Enzyme used*	*Analysis method*
ATP	Luciferase	S; NADH → NAD^+, in coupled reaction
FAD	D-Amino acid oxidase	E; O_2 electrode
K^+	Phosphofructokinase	S; activation of enzyme
Mg^{2+}	Luciferase	F; chemiluminescence
Zn^{2+}	Aminopeptidase apoenzyme	S; *p*-nitrophenol
O_2	Luciferase	F; chemiluminescence

[a]A few examples of many. There are alternative methods, depending on instruments, substrates, and enzymes available.
ATP, adenosine triphosphate; FAD, flavin adenine dinucleotide; E, electrometric method; F, fluorescent method; S, spectrophotometric method.
Adapted from Guilbault GG (1984) *Analytical Uses of Immobilized Enzymes*. New York: Marcel Dekker.

$$\text{Luciferin} + O_2 \xrightarrow[\text{ATP, } Mg^{2+}]{\text{Luciferase}} \text{Oxidized luciferin (+ phosphorescence)} \qquad (13)$$

Determination of Enzyme Concentration

Determining enzyme concentration by measuring the initial velocity, v_o, at which a substrate is converted to product seems straightforward. Yet it is fraught with difficulties unless careful attention is given to the design of the experiment, as time, substrate, pH, temperature, activators, and inhibitors all affect v_o for a fixed concentration of enzyme. Therefore, if the enzyme concentration, $[\mathbf{E}]_o$, is to be measured accurately and precisely, all these other parameters must be controlled exactly among experiments. It is important to determine the initial velocity, v_o, so that changes in substrate concentration, stability of the enzyme, approach to equilibrium, and/or inhibition by the product are not factors. The substrate concentration does not need to be much greater than K_m when v_o is measured. Any substrate concentration, when precisely reproduced, will give a linear relationship between v_o and $[\mathbf{E}]_o$, with a few exceptions. The temperature should be below optimum temperature so the enzyme is stable and the pH must be close to the optimum pH for maximum sensitivity.

Determination of absolute concentration of enzyme based on v_o requires a standard preparation of the enzyme with known purity, based on activity. For example, if it is known that a 5 µl aliquot containing 1 µg of 100% active pure enzyme has a v_o of 30 µmol of product formed per minute, while a 5 µl aliquot of an unknown sample has a v_o of 6 µmol of product formed per minute under the same conditions, then the unknown aliquot contains 0.2 µg of the enzyme.

Table 4 gives a few selected examples of analytical determinations of enzyme concentration. (*See* **Spectroscopy**: Fluorescence; Overview.)

Structural Analysis

Because of their high specificity, enzymes are ideal probes of molecular structure. They can distinguish between: D/L, *cis/trans*, α/β, and other stereospecific configurations; nature, location, and sequence of monomeric units in heteropolymers (proteins, carbohydrates, nucleic acids); and secondary, tertiary, and quaternary structures of polymers. They are useful in determining location of enzymes and substrates in organs (histochemistry) and in identifying subcellular organelles (based on enzyme composition), cell walls and membranes. The restriction enzymes and ligases are essential tools in recombinant DNA

Table 4 Analytical determination of enzyme concentration[a]

Enzyme	*Substrate used*	*Analysis method*
Catalase	H_2O_2	E; O_2
β-Galactosidase	*o*-Nitrophenyl-β-D-galactoside	S; *o*-nitrophenol
Invertase	Sucrose	S; use of glucose oxidase/peroxidase coupled assay
Lipase	Tributyrin	E; H^+
Pectin methylesterase	Pectin	E; titrate H^+ formed
Peroxidase	H_2O_2, 4-aminoantipyrine, *p*-hydroxybenzoate	S; oxidized, conjugated 4-aminoantipyrine
Phosphatase, alkaline	*p*-Nitrophenyl phosphate	S; *p*-nitrophenol
Polyphenol oxidase	Pyrocatechol	S; benzoquinone
Xanthine oxidase	Hypoxanthine	S; uric acid

[a]There are several methods available for each enzyme. Availability of instruments and substrates often determine choice of method.
E, electrometric method; F, fluorescent method; S, spectrophotometric method.
Adapted from Guilbault GG (1984) *Analytical Uses of Immobilized Enzymes*. New York: Marcel Dekker.

technology and genetic engineering, being highly specific for nucleotide primary and secondary structures. As probes of structure, unequivocal determination of activity, or lack of activity, is more important than the absolute velocity of product formation (v_o).

Determination of v_o

Determination of v_o for formation of product (dP/dt), disappearance of substrate ($-dS/dt$), change in energy (as heat), or change in pH is required for analytical use of enzymes, as indicated by the preceding discussion. This requires the ability to detect a change during the reaction. Changes in absorbance, fluorescence, phosphorescence, electrical properties including pH, or heat are appropriate monitors. Excellent instrumentation is available for detection of all these changes. The best methods are those where a continuous recording of the change is possible, often with software to convert the measured rate of change directly to concentration of the compound being measured. Data accumulation and analysis are easy when the change is continuously recorded. However, sometimes an aliquot of the reaction must be removed periodically, another reagent added to give a colored derivative (for example) of product or substrate and then the concentration of the derivative determined. Sometimes v_o can be measured only after chromatographic separation of the products. These are lengthy, laborious assays that frequently lead to one-point assays. v_o, based on one-point assays, must be accepted with great caution.

Often, solutions to difficult analyses are possible via coupled enzyme assays. A coupled enzyme assay involves two or more enzymes, in which the product of the last enzyme is measured as shown in eqn 14. The intent is to measure the concentration of A (or E_1). However, the product B is analytically very difficult to determine, while D is very easy to determine:

$$A \xrightarrow{E_1} B \xrightarrow{E_2} C \xrightarrow{E_3} D \qquad (14)$$

A specific example will illustrate this better (eqns 15–17).

$$\text{Sucrose} + H_2O \xrightarrow{\text{Invertase}} \text{Glucose} + \text{Fructose} \qquad (15)$$

$$\text{Glucose} + O_2 + H_2O \xrightarrow[\text{Oxidase}]{\text{Glucose}} \text{D-Gluconic acid} + H_2O_2 \qquad (16)$$

$$H_2O_2 + \text{Chromagenic substrate} \xrightarrow{\text{Peroxidase}} \text{Fluorescent product (or colored)} \qquad (17)$$

Coupled reactions often involve terminal enzymes requiring NAD^+ (NADH) or $NADP^+$ (NADPH), since the extinction coefficient at 340 nm is 6.22×10^3 mol l^{-1} cm^{-1} for NADH (or NADPH) versus NAD^+ (or $NADP^+$), a very large change.

Coupled reactions require enzymes with similar pH optima and that v_o ($= k_2$ [E][S]$_o$/ (K_m + [S]$_o$)) be at least 100 times larger for reactions catalyzed by E_2, E_3, etc., than for E_1. This is usually met by using high concentrations of E_2, E_3, etc. The rationale for this is the concentrations of final product measured (D in eqn 14; hydrogen peroxide in sucrose determination eqns 15–17) must be a stoichiometric reflection of concentration of substrate A (or E_1) being determined. The examples in Tables 1–4 illustrate how frequently coupled reactions are used.

Soluble enzyme systems are not required for determining v_o (and analyte concentration). In fact, present emphasis is on the use of immobilized enzyme systems in columns, electrodes, biosensor chips and antibody–enzyme conjugates (enzyme-linked immunosorbent assay (ELISA), for example). Immobilized enzymes are more stable than nonimmobilized enzymes and may be used repeatedly. Dozens of enzyme electrodes are available (several

commercially) that have response times equivalent to a hydrogen (pH) electrode and can be used repeatedly for several months. Biosensors involving multiple microchips are being developed, which may determine up to eight or more enzymes, substrates, or other compounds simultaneously when a drop of blood, urine, or other biological fluid is added. (*See* **Immunoassays:** Radioimmunoassay and Enzyme Immunoassay.)

See also: **Aluminum (Aluminium)**: Properties and Determination; **Ascorbic Acid**: Properties and Determination; **Coenzymes**; **Fructose**; **Immunoassays**: Radioimmunoassay and Enzyme Immunoassay; **Lactose**; **Lead**: Properties and Determination; **Mercury**: Properties and Determination; **Spectroscopy**: Overview; **Starch**: Structure, Properties, and Determination

Further Reading

Bergmeyer HU, Bergmeyer J and Grassl M (eds) (1983) *Methods of Enzymatic Analysis*, 3rd edn. Weinheim: Verlag Chemie.

Colowick SP and Kaplan NO (eds) (1955–1999) *Methods in Enzymology*, vols 1–300. Orlando: Academic Press.

Guilbault GG (1984) *Analytical Uses of Immobilized Enzymes*. New York: Marcel Dekker.

Whitaker JR (1974) Analytical applications of enzymes. In: Whitaker JR (ed.) *Food Related Enzymes. Advances in Chemistry Series* 136: 31–78.

Whitaker JR (1984) Biological and biochemical assays in food analysis. In: Stewart KK and Whitaker JR (eds) *Modern Methods of Food Analysis*, pp. 187–225. Westport: AVI.

Whitaker JR (1985) Analytical uses of enzymes. In: Gruenwedel DW and Whitaker JR (eds) *Biological Techniques. Food Analysis: Principles and Techniques*, vol. 3, pp. 297–377. New York: Marcel Dekker.

Whitaker JR (1994) *Principles of Enzymology for the Food Sciences*, 2nd edn. New York: Marcel Dekker.

Whitaker JR (1996) Enzymes. In: Fennema OR (ed.) *Food Chemistry*, 3rd edn, pp. 514–518. New York: Marcel Dekker.

EPIDEMIOLOGY

P C Elwood, Llandough Hospital, Penarth, South Glamorgan, UK

Aims of Epidemiology

Epidemiology is the study of health and disease in the community. There are two groups of aims: first, to describe the distribution, the pattern, and the natural history of disease in the general population, and second, to identify factors that may be causal in a disease process, and to evaluate strategies for the control, management, and prevention of a disease.

In nutritional research, epidemiology aims to identify associations between nutrients or food items and the risk of a disease, and then test these associations to verify whether they are likely to be causal. In practice, epidemiology is multifaceted and a number of preliminary considerations will help to put it into context within general medical and biological research.

Epidemiologists rarely generate hypotheses. Most ideas about the causation of disease arise from clinical observations, metabolic studies, or other intensive investigations. The role of the epidemiologist is to take these ideas and test them in representative samples of the general community or, if appropriate, in representative series of patients. In doing this the epidemiologist not only tests the relevance of a factor to the disease, but also seeks to obtain an estimate of the relative importance of the mechanism or factor. Furthermore, in a well-designed comprehensive study, the extent to which the effect of any particular factor of interest is independent of the effects of other factors is established. This generally requires large complex and long-term studies, based on samples of subjects that truly represent the population.

The interest of the epidemiologist is therefore on communities and his or her attention focuses ultimately on lifestyles, on nutrition, and on environmental factors. Thus the recent finding that certain thrombosis-related factors, such as plasma fibrinogen, are very strongly predictive of future ischemic heart disease events leads the epidemiologist on to search for the dietary and lifestyle determinants of fibrinogen. The likely relevance of antioxidants to cancer and other conditions leads the laboratory worker to study free radicals and their cellular and other effects, while the epidemiologist studies the relative protection given by diets rich in natural antioxidants, and the effects of dietary antioxidant supplements. The

hypothesis that the regular consumption of fish is associated with a low incidence of vascular disease leads the biochemist to study the metabolism of fish oils and mechanisms linking fish oils to blood pressure and other cardiovascular risk factors, while the epidemiologist sets up a randomized controlled trial of fish eating and the incidence of vascular disease. The different research approaches – clinical, laboratory, and epidemiological – are complementary and all are needed.

Only the epidemiologist, however, can estimate, from studies based upon representative population samples of individuals, the relative importance of one particular factor to a disease, in comparison with a host of other relevant factors. Thus, magnesium is a key element in many enzymes and is known to be involved in a large number of metabolic and other biological mechanisms. Only the epidemiologist can however estimate the importance of magnesium to health and disease in the population, and the part played by magnesium deficiency, if any, in, say, cardiovascular disease within a community.

The Concept of Risk Factors

Epidemiologists were probably the first to use the term 'risk factor,' applying it to any factor which identifies an individual who is at increased risk of a disease.

The term is however vague and its limitations should be understood. Thus the term is used to cover factors which are inherent in a subject and cannot be altered (such as sex and race, etc.), factors which are likely to be causal and which can be changed (such as dietary fat intake, blood pressure, and plasma fibrinogen) and factors which confound relationships of interest (such as age and social class). Factors in this last group are perhaps the most difficult to identify and deal with. For example, in examining the relationship between a nutrient and a disease, calorie intake can often be a confounding variable, in that it is probably related to the intake of the nutrient of interest (for most nutrients, the more food that is eaten, the greater the intake of any nutrient), and calorie intake, or total energy, is a good surrogate measure for exercise level which, in turn, is strongly related to the incidence of vascular and other diseases.

Another suggested modification of the concept of 'risk factor' is to refer to 'predictors' of a disease (such as cholesterol level, blood pressure, plasma fibrinogen and other biochemical and hematological factors) and 'determinants' of those predictors (such as dietary fat intake, salt intake, smoking habit, and other lifestyle and dietary factors). This last grouping has the great advantage that it acknowledges the difference between factors involved at the level of metabolic processes, and extrinsic factors, upon which preventive strategies which are under the control of subjects themselves, and population-based intervention initiatives, can be focused.

Epidemiological Methods

A variety of strategies are used in epidemiology. Basically these can be divided into prospective and retrospective, and into observational and experimental. Every combination of these is possible: changes over time; comparisons between communities; individual case-control comparisons; cross-sectional surveys; prospective cohort studies; and the randomized controlled trial (RCT) with intervention. Another way in which the various strategies can be classified is by the degree of certainty of the conclusions drawn from them. This last is so important that the various study designs which are described below are listed in the order of the likely certainty of the conclusions that can be drawn from the evidence they yield.

Trends with Time

There have been great changes in the mortality of many diseases, including many cancers and coronary heart disease (CHD), in most countries during the past century and it is possible to look for dietary changes which parallel these changes in disease incidence. Great difficulties arise, however, because information on past disease is uncertain and knowledge of past dietary intakes is very poor. Diagnostic criteria change, and the attributions of certain causes to deaths have changed markedly over the years. In addition, estimates of the nutrient content of food change, as reflected in food composition tables. The MONICA study, set up by the World Health Organization in about 40 countries to study changes over time in CHD, and in relevant risk factors, is probably the most ambitious study of this kind ever mounted.

Even if reliable data on both diets and disease were available, the two cannot be linked with certainty because many other relevant factors may have changed concurrently. Diet is only one aspect of lifestyle and it should never be considered in isolation.

Cross-community Comparisons

There are differences in disease rates in different communities and some of these are large. Relating these to differences in dietary intakes has generated a number of interesting hypotheses. There are difficulties however in that the dietary data available for the different

communities may not be truly representative for those communities, and the disease data may not be truly comparable. The biggest stumbling block is the virtual impossibility of linking differences in disease incidence to dietary differences since many other factors – genetic, behavioral, and environmental – will almost certainly be different and may well be far more important in determining disease risk than dietary differences.

Case-control Studies

The identification of patients with evidence of a disease and the comparison of possible risk factors in these with the same factors in disease-free subjects is a basic research strategy. Difficulties arise, however, because the 'cases' are often patients with advanced disease, who may not be representative of all such patients. Furthermore, the observer cannot always be 'blind' with regard to whether a subject is a 'case' or a 'control,' and so that bias can easily be introduced into the data on risk factors. Further difficulties arise in the selection of suitable control subjects without the disease, and though some of these difficulties can be overcome by matching for factors such as sex, age, and area of residence, uncertainties remain about comparability between the cases and the controls with respect to factors other than those believed to be relevant to the disease.

In nutritional research the case-control strategy has only a very limited place because it necessitates inquiries about dietary intakes before the onset of the disease, and it is very difficult for a patient to separate out this information from dietary changes which have been made since symptoms of the disease commenced. Furthermore, the case-control approach has very little place in diseases with a high case fatality, such as cancer and heart disease, because only a selected subsample of all cases survive to be included in any case-control study and these will certainly not be representative of all patients with the disease. (*See* **Cancer**: Epidemiology.)

Cross-sectional Survey

One of the most common activities in epidemiology and in social science is the conducting of surveys. These can give estimates of the prevalence of a disease, the distribution of dietary intakes, and the levels of risk factors, and associations between these can be examined. For such purposes, it is essential that the population sample selected for the study is truly representative of the community being studied, and that a high response rate is achieved because otherwise representativeness will be compromised. Attention has also to be given to the reproducibility of measurements made. INTERSALT, an international collaborative study coordinated by WHO, was a good example of a cross-sectional study. In this, about 40 research centers obtained data on blood pressure and on salt excretion from 24-hr urine collections (a surrogate for salt intake) in about 200 subjects per center.

Cross-sectional studies have the great advantage that all the data collected are contemporaneous, that is, there is no dependence upon the memories of subjects, nor is there any waiting for disease to develop. On the other hand, surveys have the great disadvantage that causal factors which predated the onset of the disease cannot always be distinguished from a factor which has been affected by the disease itself. Thus, in a cross-sectional survey, differences in dietary intakes between subjects with CHD and other subjects, could equally well be the result of changes in diet made by patients after their heart attack, or the onset of angina, as the cause of the disease process.

One valuable role for cross-sectional surveys is the identification of determinants of risk factors for disease. Thus the associations between dietary and other lifestyle factors and, for example, serum cholesterol, blood pressure, or serum fibrinogen can be examined.

Prospective Studies

The prospective, cohort, or longitudinal study is the classic tool of the epidemiologist. In this, a representative population sample of subjects (a cohort) is examined and then followed forward in time. The association between the levels of the various factors measured at baseline, and the development of the disease, is examined. In other words, the predictive power for subsequent disease of dietary, lifestyle, biochemical, hematological, and other factors is examined. The great advantage of this strategy is that the suspected causal factors are measured before the disease becomes evident, and causal factors can therefore be distinguished from factors which represent effects of the disease.

The identification of factors showing greater than chance associations with subsequent disease is however not the final answer. The degree to which the predictive power of a factor for the disease is independent of other factors must be examined. Thus, on first analysis the dietary intakes of many dietary factors is likely to be found to be predictive of a range of diseases. It must be examined whether or not such relationships are simply a consequence of an age effect (older subjects having an increased risk of most diseases and, probably, a reduced intake of most nutrients) or due to confounding by smoking

(smokers having intakes of many nutrients which are different from nonsmokers, and smokers having an increased risk of a wide range of diseases). In practice it is found that most factors, dietary and other, which are predictive for a disease, are interrelated to a whole network of other factors relevant to that disease. Unscrambling these relationships presents great difficulties, especially when the different factors are estimated with different degrees of precision. Other difficulties arise because of the possibility of overcontrol for confounding, and because errors in the measurement of an independent factor, such as a dietary factor, lead to the underestimation of an association.

Intervention Trials

There are two broad uses of intervention trials in dietary research. First, there is the small, short-term intervention trial that examines a mechanism or tests the association between factors. Thus, the intake of a food item, such as salt, saturated fat, fatty fish, and vitamin C, or the intake of dietary fiber, fruit and vegetables, or dairy produce can be changed and changes observed in biochemical or hematological factors known to be relevant to disease. Possible determinants of platelet function, oxidized lipoproteins, and blood pressure can be thoroughly evaluated using this strategy.

The second, and ultimately far more important use of intervention trials is to test the relevance of a factor to the incidence of a disease. The rules for such RCTs have been well worked out and if an RCT is to yield reliable evidence then there must be no compromise on certain cardinal aspects. These last include random allocation of subjects to the intervention ('treatment') or the 'control' group; the follow-up of all subjects admitted to the trial, whatever their level of compliance, is essential if bias is to be avoided ('intention-to-treat analysis'); and the assessment of outcome events must be made 'blind' with regard to whether the subject was in the treated or the control group. Unfortunately, in dietary RCTs it is rarely possible to comply with a further point which is regarded as important in RCTs, namely 'blindness' of the participants as to whether they are receiving the intervention measure or not. This last is easily achieved in drug trials, but the use of a dummy placebo is seldom possible in dietary trials.

While RCTs give evidence with a high level of certainty, they still have limitations. First, they usually have to be large and long-term. For example, even with the inclusion of thousands of subjects, and prolongation for several years, almost all RCTs so far reported have failed to give conclusive evidence on the relevance of dietary fat intake to survival. Hence there needs to be overviews of all available and relevant evidence, as described below.

Other limitations of dietary RCTs include the difficulty in persuading subjects to make a change in their diet which is of sufficient size, and is maintained for a sufficient time, to be likely to affect the incidence or the progression of a disease. In fact, setting-up a trial with random allocation to an intervention group often leads to 'contamination' of the control group, either through contact between subjects themselves, or through reports in the media. At the very least, dietary and other changes in the two groups, intervention and control, must be carefully monitored. Then it is important to avoid as far as possible changes in dietary factors other than the one of interest, and those changes that do occur should be monitored and recorded. Thus an increase of, say, fish consumption is likely to lead to a decrease in meat intake, and an increase in wholemeal bread may lead to an increase in the use of spreading fats.

Limitations of Approach

Difficulties in epidemiological approaches to research on nutrition and disease include the following:

1. The etiology of most diseases is complex and for very few diseases is there a single cause. Not only are food items likely to interact together, and not only is a change in one dietary item likely to be accompanied by changes in other food items, but other lifestyle factors are likely to be related to the dietary intake of any particular nutrient. Smoking and exercise level are two nonnutritional factors that confound almost every possible nutritional relationship. Social class and education are also relevant to most nutritional factors and to the risk of almost every disease.
2. The difficulties in adequately characterizing the diet of any individual subject are enormous and in most nutritional epidemiological studies this is necessary for thousands of subjects. Methods of estimating dietary intakes are dealt with elsewhere and it has to be accepted that even the most careful and exacting of these give at best only a crude estimate of what is required: that is, the usual or the average dietary intake of subjects over a very long period of time. This is a most important limitation because uncertainties in the measurements of independent variables will reduce the power of a study and true effects may escape detection.
3. Third, the effects of food items or dietary patterns are likely to be very long-term, and in the

study of relationships with disease it is possible that the diet at the time the disease is diagnosed is no longer of any relevance, yet the estimation of dietary intake of an individual in the past – perhaps 10 or 20 years previously – is clearly almost impossible.

4. There are often difficulties in recruiting enough subjects to give power in a study to test a hypothesis adequately. Recruitment can be especially difficult in nutritional studies in which the collection of dietary data, or the making of a change in diet, can be quite disruptive. In estimating the power of a proposed study it is important that a realistic estimate is made of the likely size of the predicted effects. It is all too easy to overestimate this, just as it is easy to overestimate the validity of estimates of dietary intakes. The power of a study to detect a real effect depends ultimately on the size of the effect, and this is usually smaller than one would like, and the size of the effect, or intervention, being studied.
5. In uncommon diseases, and in diseases with a long 'incubation' period, a case-control study may be all that is possible, and this strategy has the greatest limitations. On the other hand, while RCTs yield the most conclusive evidence, they are the most difficult of all the strategies and, as has been indicated, they raise additional difficulties in nutritional research.

Progression in Epidemiological Research

There is a logical sequence in epidemiology, from the observation perhaps of area differences in the prevalence of a disease, or a change over time, to cross-sectional studies that determine prevalence and test associations with dietary or other factors, to intervention trials which examine a preventive strategy focused, for example, on a nutritional factor. A progression such as this will of course be governed to some extent by findings in clinical and in laboratory research and the different approaches are not in any way totally independent.

Such a sequence is seldom seen, and opportunistic studies which explore some aspect are often all that are feasible. However, research on magnesium does illustrate a logical but incomplete sequence. The demonstration of magnesium as an essential element in many enzymes and its involvement in many biological mechanisms led to studies of tissue magnesium in samples taken postmortem. These have led to long-term prospective studies of dietary Mg intakes and the incidence of vascular disease. These led to RCTs of the benefit of Mg infusions given immediately after a myocardial infarction. The sequence will not be completed, and the relevance of dietary Mg to disease will not be apparent until long-term RCTs of an increased Mg intake have been completed and examined in a metaanalysis.

Interpretation of Data

A clear distinction must always be made between the testing of a hypothesis which was defined before a study was set up, and the generation of new hypotheses from data collected for another purpose. This latter, which is often termed 'fishing' or 'dredging,' is permissible only if it is recognized for what it is and if the further associations detected are tested in new *ad hoc* studies. The dangers inherent in generating multiple hypotheses, and in then treating those that achieve an acceptable level of statistical significance as if they were prior hypotheses, are very great in dietary studies because of the number of independent variables.

Firm conclusions should only be based on evidence drawn from a number of studies. The play of chance can never be ruled out, whatever the level of statistical significance, and consistency in a number of studies is a better guide to truth than the results of a single study. Hence the importance of 'overviews' of studies, or metaanalyses of data from a number of different studies. It is most important that data from every relevant study is included in an overview. Unfortunately, certainty on this last is impossible, as there is always a 'publication bias,' that is, reports which confirm a null hypothesis are far less likely to be published than those which show an effect.

There is another aspect of this last point. Not only does one look for consistency in the results of different studies using the same design strategy; one should also look for consistency in the results of studies with different approaches. This is necessary even when there is convincing evidence from, say, a number of RCTs, because an increase in the consumption of one nutrient or one foodstuff is usually accompanied by a change in the consumption of some other nutrients or food items. Confidence is enhanced, however, if prospective studies demonstrate an association between a food item and a disease, if RCTs show a change in the disease incidence, and if these effects are consistent with the results of metabolic studies. Examples of such consistency are rare in dietary research but the benefit of fatty fish in heart disease has been shown in historical studies, in comparisons between communities, in laboratory-type studies of fish oil and vascular risk factors, and in (admittedly only one) intervention RCT.

The Cochrane Collaboration

There have been a number of basic advances in research methodology and foremost among these must be the introduction of the RCT. Another, of perhaps equal importance, has been the development of the overview or metaanalysis. The recognition of the need for these throughout clinical practice led to establishment of the Cochrane Collaboration. This is now an international organization that seeks to identify every acceptable RCT within medical care, makes these available within selected diseases or disease groups as databases, and organizes systematic reviews within different health care interventions.

The Collaboration is an exceedingly valuable research resource. For example, a recent search under the headings 'heart disease' and 'diet' produced, within a minute or two, 328 references to RCTs.

See also: **Aging – Nutritional Aspects**; **Cancer**: Epidemiology; **Food Safety**; **Malnutrition**: The Problem of Malnutrition

Further Reading

Alderson M (1983) *An Introduction to Epidemiology*, 2nd edn. London: Macmillan.

Armstrong BK, While E and Saracci R (1992) *Principles of Exposure Measurement in Epidemiology*. Oxford: Oxford University Press.

Bradford-Hill A (1971) *Principles of Medical Statistics*, 9th edn. London: The Lancet.

Burr ML (1979) Medical memorandum: criteria for planning a research project. *Community Medicine* 1: 157–159.

Burr ML (1985) Progress in food and nutritional science. *Nutritional Epidemiology* 9: 149–183.

Burr ML, Elwood PC, Fehily AM and Sweetnan PM (1983, 1984) Epidemiology for nutritionists: a series of five articles on epidemiological methods. *Human Nutrition: Applied Nutrition* 37A: 259–264, 265–269, 339–347, 419–425; 38A: 215–222, 329–334.

Cochrane AL (1972) Effectiveness and efficiency. Random reflections on health services. The Nuffield Provincial Hospital Trust 1972. London: Royal Society of Medicine Press Ltd.

Last JM (1988) *A Dictionary of Epidemiology*, 2nd edn. Oxford: Oxford University Press.

Margetts BM and Nelson M (1977) *Design Concepts in Nutritional Epidemiology*, 2nd edn. Oxford: Oxford University Press.

Willett W (1998) *Nutrition Epidemiology*. Oxford: Oxford University Press.

ESCHERICHIA COLI

Contents

Occurrence

K S Venkitanarayanan, University of Connecticut, Storrs, CT, USA
M P Doyle, University of Georgia, Griffin, GA, USA

Introduction

Escherichia coli was first described by Dr. Theodore Escherich in 1885. The natural habitat of *E. coli* is the large intestine of warm-blooded animals, including humans. The bacterium is Gram negative, rod-shaped, and can be easily grown on laboratory culture media. However, most strains generally do not survive very well outside of the host's gastrointestinal tract. Therefore, its presence in the environment is considered as evidence of recent contamination with mammalian or avian feces. Because many fecal-borne microorganisms are pathogenic in animals and humans, the presence of *E. coli* in water and foods is indicative of a potential hazard.

Environmentally, *E. coli* is most often associated with fecally contaminated waters. Sources of contamination include animal feces and discharge of untreated sewage. This has become a cause of concern

in recreational areas and beaches, as such areas are often closed when populations of *E. coli* increase. *E. coli*-contaminated water is also of concern to the food industry. Shellfish, through their filter-feeding apparatus, can concentrate in their tissue *E. coli* and other bacteria present in water. Many shellfish are eaten raw or poorly cooked whereby contaminating bacteria are ingested. In addition, the use of fecal-contaminated water for irrigation of crops and the use of animal manure as a fertilizer can cause contamination of vegetables and other foods of plant origin. Not only is *E. coli* an indicator of the possible presence of fecal-borne pathogens, but some strains are themselves pathogenic. Most strains of *E. coli* are harmless commensals and appear to be avirulent in healthy humans. However, a few strains, with well-characterized traits that are known to be associated with pathogenicity, cause infections of the urinary tract, peritoneum, central nervous system, wounds, and gastrointestinal tract. Those of greatest concern in water and foods are the intestinal pathogens, which are classified into five major groups: the enterotoxigenic *E. coli* (ETEC), the enteroinvasive *E. coli* (EIEC), the enteropathogenic *E. coli* (EPEC) the enterohemorrhagic *E. coli* (EHEC), and the enteroaggregative *E. coli* (EAEC).

Entry into the Food Chain

Since *E. coli* is naturally occurring in the intestinal tract of animals, its entry into foods most often occurs by direct or indirect contact with fecal material. Since humans are the principal source of ETEC, EPEC, EIEC, and EAEC, contamination with human feces is the principal factor in transmission of these pathogens. *E. coli* O157:H7, the predominant serotype among the EHEC, can colonize cattle and other ruminants such as sheep and deer. In cattle, *E. coli* O157:H7 localizes primarily in the rumen and colon, and is shed in feces. Studies have revealed that *E. coli* O157:H7 can survive in bovine feces for many months but also retains its ability to produce toxins. Two important routes by which foods may become fecally contaminated with *E. coli* include:

1. Contamination from food handlers. This can be prevented by education of food handlers in proper hygienic practices, e.g., washing hands thoroughly after defecation.
2. Contamination of animal products with feces from the animal. During slaughter and subsequent processing operations, contamination of carcasses with *E. coli* O157:H7 from the digesta or feces of cattle can occur. Hygienic slaughter practices and treatments to clean and disinfect carcasses may be inefficient in completely removing bacteria from carcasses. During milking of dairy cattle, milk can become contaminated with *E. coli*. One study has found that 46% of raw milk samples and 29% of milk products contained *E. coli*.

Water can also serve as a vehicle for contamination of foods by *E. coli*. Water-borne outbreaks of *E. coli* O157:H7 infection have been reported in the USA, Scotland, Canada, and Japan. *E. coli* O157:NM was implicated in a water-borne outbreak in South Africa. At least four water-borne outbreaks of *E. coli* O157:H7 infection in the USA were associated with swimming water in lakes. Since natural waters often contain *E. coli* of animal origin, and may be contaminated with animal feces and untreated human sewage, their use in food harvesting or processing can cause contamination of the food products, e.g.:

1. Direct contamination of seafood. This is especially a problem with filter feeders such as shellfish, also but also has been observed in many other types of seafood.
2. Contamination of plants during irrigation or fertilization. In some areas, raw human sewage is used to fertilize crops intended for human consumption. Fruits and vegetables grown on soil fertilized with cattle manure or irrigated with water contaminated with cattle manure have the potential of being a vehicle of *E. coli* O157:H7. An outbreak of disease caused by *E. coli* O157:H7 occurred after handling potatoes that had been fertilized with cow manure.
3. Contamination of food-processing equipment or food contact surfaces.

Occurrence in Foods

Since raw foods of animal and plant origin are subject to potential fecal contamination, *E. coli* has been isolated from a wide variety of foods. Foods that have been associated with outbreaks of *E. coli* infections include meat and meat products, fish, poultry, milk and dairy products, vegetables, and water. The incidence of *E. coli* in foods varies greatly in different countries. Investigators in India isolated enterotoxigenic *E. coli* from 33–52% of uncooked chicken sausages and 10–43% of uncooked pork sausages. *E. coli* was detected in 4–7% of 597 retail seafood products in San Francisco, USA. A survey of retail meats revealed that *E. coli* O157:H7 was isolated from 1–2% of beef, pork, poultry, and lamb samples in Madison, USA. A study in Mexico revealed that 74 of 75 Mexican cheeses contained *E. coli* over acceptable levels, whereas seven of 25 non-Mexican cheeses contained unacceptable levels. In Egypt, 28% of fresh

soft white cheese samples contained *E. coli*, of which half the isolates were enterotoxigenic. Twenty-nine percent of dairy products tested in the former Yugoslavia contained *E. coli*. *E. coli*, including some pathogenic serotypes, was isolated from 82 of 105 milk samples from various sources in India. However, a survey of commercially available cheese and milk samples for enterotoxigenic *E. coli* in the United States revealed that none of the *E. coli* isolates recovered was enterotoxigenic. A 1997 survey conducted on the occurrence of *E. coli* in ready-to-eat food products in the UK revealed that the bacterium was present in 2.2% of beef products, 2.8% of chicken products, 1.9% of ham products, 1% of pork products, 5% of plant products, and 2.8% of turkey products. A study of retail foods in Seattle, USA, found Shiga toxin-producing *E. coli* in 18% of pork, 12% of chicken, 7% of turkey, 10% of fish, and 5% of shellfish samples. A recent survey of ground beef for *E. coli* O157:H7 conducted by the United States Department of Agriculture found only 59 positive samples out of 7009 specimens, indicating that the prevalence of this pathogen in US ground beef is about 0.8%.

Fate of *E. coli* in Foods

The behaviour of *E. coli* in foods is similar to that of other Gram-negative bacteria. They are generally sensitive to heat and are killed in foods that are adequately cooked. Studies with *E. coli* O157:H7 have revealed the organism to be slightly more heat-sensitive than *Salmonella* spp., with *D*-values in ground beef at 57.2, 60, 62.8, and 64.3 °C of 270, 45, 24, and 9.6 s, respectively. Heating apple juice at 60 °C for 1.6 min reduced the *E. coli* O157:H7 population by 4−$\log_{10}$. *D*-values of *E. coli* in shrimp paste were 5.8 and 0–3 min at 57 and 65 °C, respectively. Hot-water sprays are useful for reducing the bacterial load on animal carcasses. Treating sides of beef with water at 83.5 °C for 20 s can reduce *E. coli* populations 3 $\log_{10}$ without any detrimental effects on meat color. *E. coli* O157:H7 is as sensitive to radiation as most other enteric pathogens. For example, in chicken, the *D*-values of *E. coli* O157:H7 at 5 and −5 °C were 0.27 and 0.42 kGy, respectively. In ground-beef patties at 4 and −16 °C, the *D*-values for *E. coli* O157:H7 were 0.24 and 0.30 kGy, respectively, compared with 0.62 and 0.80 kGy, respectively, for *Salmonella*.

E. coli O157:H7 can grow in a variety of foods, including meat, milk, cantaloupe, and watermelon cubes, and salad vegetables such as lettuce and cucumber. A study on the survival and growth characteristics of *E. coli* O157:H7 in unpasteurized milk revealed that the pathogen survived at 5 °C for 28 days, and increased in populations at 8 and 15 °C, with rapid growth growth observed at the latter temperature. In pasteurized milk samples held at 22 °C, *E. coli* O157:H7 populations increased by approximately 5.0 $\log_{10}$ colony-forming units (CFU) ml^{-1} in 7 days, and then decreased to undetectable levels by day 28. When inoculated in large populations, *E. coli* O157:H7 could survive in mayonnaise (pH 3.6–3.9) for 5–7 weeks at 5 °C and for 1–3 weeks at 20 °C. *E. coli* O157:H7 also can survive for extended periods of time in several acidic foods such as cheese, yogurt, and apple cider. *E. coli* O157:H7 survived in cottage cheese for more than 60 days. The pathogen can survive the fermentation process in yogurt for 5 h at 42 °C, and for 7 days in yogurt held at 4 °C.

Freezing can kill approximately 1 $\log_{10}$ of *E. coli* in foods. Prolonged frozen storage will not completely kill *E. coli* unless the organism goes through several freeze/thaw cycles. However, such conditions would normally destroy the palatability of food. *E. coli* O157:H7 survived for up to 9 months with no decrease in population in ground beef held frozen at −20 °C with no freeze/thaw cycles. *E. coli* is more resistant to sodium chloride and sodium nitrite than *Salmonella*. *E. coli* in brain–heart infusion broth can grow in 4% sodium chloride and 400 μg of sodium nitrite per milliliter at 15 and 35 °C.

Several studies have determined the fate of *E. coli* in a variety of dairy products. Generation times of *E. coli* in skim milk were 28–35 min at 32 °C and 66–109 min at 21 °C. When lactic starter cultures were included in the milk, growth of *E. coli* was inhibited at both temperatures. A study of the fate of *E. coli* during the manufacture of Camembert cheese revealed that *E. coli* grew slowly until the curd was cut and hooped, then increased rapidly. During overnight storage at 10 °C, as the pH of the cheese decreased to 5.0 or below, the viable *E. coli* population decreased sharply. Depending on the strain tested, *E. coli* was not detectable in the cheese after 0–9 weeks of storage at 10 °C. Inoculating *E. coli* directly onto the surface of ripening Camembert cheese resulted in rapid growth and subsequent survival for up to 7 weeks of storage.

Parameters that influence the growth of *E. coli* in foods roughly parallel those of other Gram-negative bacteria. The minimum water activity (a_w) for growth ranges from 0.94 to 0.97. The optimum pH for growth is approximately 7.0, with a minimum and maximum pH for growth of 4.5 and approximately 9.0, respectively. Growth for some strains can occur from 7 to 46 °C, with maximal growth rates occurring at 35–37 °C. *E. coli* O157:H7 grows very poorly or not at all in many media at temperatures at and

above 44 °C. *E. coli* is capable of very rapid growth in high-nutrient foods. Populations can double within 30–40 min under optimal conditions.

Acid Tolerance of *E. coli* O157:H7

E. coli O157:H7 has an unusual acid tolerance, as evidenced by involvement in outbreaks associated with high-acid foods such as apple cider, yogurt, and fermented meat sausage. *E. coli* O157:H7 can survive for extended periods of time in synthetic gastric fluid. Further, exposure of *E. coli* O157:H7 to mild or moderate acidic environments can induce an acid-tolerance response, which enables the pathogen to survive extreme acidic conditions. For example, acid-adapted cells of *E. coli* O157:H7 survived longer in apple cider, fermented sausage, and hydrochloric acid than nonacid-adapted cells. However, *E. coli* O157:H7 is not unusually heat-resistant or salt-tolerant unless cells are preexposed to acid to become acid adapted. Acid-adapted *E. coli* O157:H7 cells also have increased heat tolerance.

The effect of diet and gastrointestinal acidity in cattle on *E. coli* O157:H7 and its acid tolerance has received recent attention from microbiologists. A study published in 1998 evaluated the role of diet (grass vs grain) and the resulting volatile fatty acids (VFA) levels and pH in the rumen and colon on induction of acid tolerance in *E. coli*. The study revealed that the rumen and colon fluids of grain-fed cattle had higher VFA concentrations and lower pH values than those of cattle fed only grass. Further, larger populations of *E. coli* and acid-resistant *E. coli* were found in rumen and colonic fluids of cattle fed a high-grain diet than that in cattle fed grass only. A following study by a different group of investigators evaluated the effect of diet (high grain vs hay-rich diets) on the duration of shedding of *E. coli* O157:H7 by cattle and the acid resistance of the pathogen shed by the animals. However, these investigators observed that hay-fed cattle shed *E. coli* O157:H7 longer than grain-fed animals, and the acid resistance of *E. coli* O157:H7 was similar, irrespective of the diet.

Prevention of *E. coli* Contamination

The most important considerations for preventing *E. coli* contamination are:

1. preventing fecal contamination by use of clean water supplies, avoiding fecal material during slaughter and milking, and practicing hygienic food handling techniques;
2. thorough cooking of all foods, especially foods of animal origin, including seafood, and pasteurization of milk and fruit juices;
3. avoiding cross-contamination of cooked foods with raw foods of animal origin and unsanitized surfaces of food-processing equipment;
4. proper cleaning and sanitizing of food-processing equipment;
5. practicing good personal hygiene;
6. storing cooked or perishable food at 4 °C or lower.

See also: **Cheeses**: Types of Cheese; Surface Mold-ripened Cheese Varieties; **Cleaning Procedures in the Factory**: Overall Approach; ***Escherichia coli***: Food Poisoning; Food Poisoning by Species other than *Escherichia coli*; **Food Poisoning**: Classification; **Food Safety**; **Lactic Acid Bacteria**; **Microbiology**: Classification of Microorganisms; **Microflora of the Intestine**: Role and Effects; **Milk**: Analysis; **Pasteurization**: Pasteurization of Liquid Products; ***Shigella***; **Spoilage**: Bacterial Spoilage; **Starter Cultures**; **Vibrios**: *Vibrio cholerae*

Further Reading

Banwart GJ (1989) *Basic Food Microbiology*, 2nd edn. New York: Van Nostrand.

Diez-Gonzalez F, Callaway TR, Kizoulis MGv and Russel JR (1998) Grain feeding and the dissemination of acid-resistant *Escherichia coli* from cattle. *Science* 281: 1666–1668.

Doyle MP and Padhye VV (1989) In: Doyle MP (ed.) *Foodborne Bacterial Pathogens*, pp. 235–281. New York: Marcel Dekker.

Gross RJ and Rowe B (1985) *Escherichia coli* diarrhea. *Journal of Hygiene, Cambridge* 95: 531–550.

Hovde CJ, Austin PR, Cloud KA, Williams CJ and Hunt CA (1999) Effect of cattle diet on *Escherichia coli* O157:H7 acid resistance. *Applied Environmental Microbiology* 65: 3233–3235.

International Commission on Microbiological Specifications for Foods (1980) *Microbial Ecology of Foods. Food Commodities*, vol. 2. New York: Academic Press.

Jay JM (2000) *Modern Food Microbiology*, 6th edn. Gaithersburg, MD: Aspen.

Kornacki JL and Marth EH (1982) Foodborne illness caused by *Escherichia coli*: a review. *Journal of Food Protection* 45: 1051–1067.

Meng and Doyle MP (1998) Microbiology of Shiga toxin-producing *Escherichia coli* in foods, pp. 92–107. In: Kaper JB and O'Brien AD (eds) *Escherichica coli and Other Shiga Toxin-producing E. coli Strains*. Washington DC: ASM Press.

Meng J, Doyle MP, Zhao T and Zhao S (2001) *Escherichia coli* O157:H7. In: Doyle MP, Beuchat LR and Montville TJ (ed.) *Food Microbiology: Fundamentals and Frontiers*, pp. 193–213. Washington, DC: ASM Press.

Detection

K S Venkitanarayanan, University of Connecticut, Storrs, CT, USA
M P Doyle, University of Georgia, Griffin, GA, USA

Use of *Escherichia coli* as Indicator Organism

Escherichia coli is the predominant member of the facultatively anaerobic portion of the human colonic normal flora. The bacterium's only natural habitat is the large intestine of warm-blooded animals and, since *E. coli*, with some exceptions, generally does not survive well outside of the intestinal tract, its presence in environmental samples, food, or water usually indicates recent fecal contamination or poor sanitation practices in food-processing facilities. The population of *E. coli* in these samples is influenced by the extent of fecal pollution, lack of hygienic practices, and storage conditions. The mere presence of *E. coli* in food or water does not indicate directly that pathogenic microorganisms are in the sample, but it does indicate that there is a heightened risk of the presence of other fecal-borne bacteria and viruses, many of which, such as *Salmonella* spp. or hepatitis A virus, are pathogenic. For this reason, *E. coli* is widely used as an indicator organism to identify food and water samples that may contain unacceptable levels of fecal contamination.

Most testing done to isolate *E. coli* from food samples is undertaken to assess the acceptability of manufacturing practices, including sanitation regimes, used in food-processing facilities. Several methods have been developed to achieve this. The traditional cultural approach is a three-step procedure. The first step estimates the number of coliforms in the sample, the second step estimates the number of fecal coliforms, and the final step isolates *E. coli*. The coliform group is defined as members of the family Enterobacteriaceae that produce acid and gas from lactose at 37 °C. Included in this group are *E. coli* and other species of the family such as *Citrobacter*, *Enterobacter*, and *Klebsiella*. Many of these genera are not primarily of fecal origin, and so the presence of coliforms in a sample does not necessarily indicate fecal contamination. The fecal coliforms group includes those coliforms that produce acid and gas from lactose when grown at temperatures between 44 and 45.5 °C. These organisms are typically of fecal origin. *E. coli* is usually the dominant bacterium of the fecal coliform group. The two most commonly used methods for enumerating coliforms, fecal coliforms, and *E. coli* in foods both utilize the most probable number (MPN) technique of statistically estimating the number of bacteria in the sample. In the method recommended by the American Public Health Association, the Association of Food and Drug Officials of the USA, and the Association of Official Analytical Chemists (AOAC), the food or water sample is decimally diluted, and 1 ml of each dilution is transferred into each of three separate tubes containing 10 ml of lauryl sulfate tryptose (LST) broth with inverted Durham tubes. These tubes are incubated at 35–37 °C for 24 h. From each tube with growth and gas production, a loopful of culture is transferred into a tube containing 10 ml of brilliant green lactose bile (BGLB) broth and an inverted Durham tube. These tubes are incubated at 35–37 °C for 24 h. Gas production is indicative of lactose fermentation by coliform bacteria. The number of positive tubes in each dilution should be recorded and an MPN table consulted to determine the number of coliforms in the sample. Both LST and BGLB tubes should be reincubated for an additional 24 h to detect slow lactose fermenters.

For the quantitation of fecal coliform bacteria, a loopful from each gas-producing LST culture is transferred into a tube containing 10 ml of *E. coli* (EC) broth and an inverted Durham tube. These tubes are incubated at 44.5 °C for 24 h. Gas production is indicative of fecal coliform bacteria. To continue with the isolation of *E. coli*, gas-producing cultures in EC broth are streaked onto plates of eosin-methylene blue (EMB) agar or Endo agar and incubated for 24–48 h at 35–37 °C. Representative colonies are chosen and biochemically tested (Table 1).

Another method of detecting *E. coli*, recommended by the UK Ministry of Health, and more commonly used in European countries, also uses the MPN procedure. Decimal dilutions of food or water samples

Table 1 Biochemical properties of *E. coli*

Test	E. coli	*ETEC*	*EPEC*	*EIEC*	*O157:H7*
Lactose fermentation	+	+	+	v	+
Gas from lactose	+	+	+	−	+
Sorbitol fermentation	+	+	+	v	−
Motility	+	+	+	−	+
Indole	+	+	+	v	+
Methyl red	+	+	+	+	+
Voges–Proskauer	−	−	−	−	−
Citrate	−	−	−	−	−
Lysine decarboxylation	+	+	+	v	+
Ornithine decarboxylation	v	v	v	−	+
β-Glucuronidase	+	+	+	+	−

ETEC, enterotoxigenic *E. coli*; EPEC, enteropathogenic *E. coli*; EIEC, enteroinvasive *E. coli*.
+, > 90%positive; −, < 10%positive; v, 10–90% positive.

are inoculated into tubes of MacConkey broth containing inverted Durham tubes, with triplicate tubes receiving 1 ml of each dilution. These tubes are incubated for 24–48 h at 35–37 °C, and the MPN of coliforms is determined using the number of tubes in each dilution showing gas production. To detect fecal coliforms, a loopful from each tube of MacConkey broth showing gas production after 24 h is transferred into a tube containing 10 ml of BGLB broth and an inverted Durham tube, and into a second tube containing 10 ml of peptone water. These tubes are incubated at 44 °C for 24 h. The indole test is performed on the peptone–water culture. Samples showing gas formation in BGLB broth and indole production are considered positive for fecal coliforms. These cultures can then be streaked onto EMB or Endo agar, and representative colonies are biochemically confirmed as *E. coli*.

Several agar media can be used for isolation of *E. coli*. EMB agar contains eosin Y and methylene blue to inhibit Gram-positive bacteria, and lactose for differentiation of lactose-fermenting colonies. Strong lactose fermentation by *E. coli* will cause formation of a typical green sheen over dark colonies. A medium commonly used for isolation of *E. coli* from clinical specimens is MacConkey agar, which contains crystal violet and bile salts to inhibit Gram-positive organisms. *E. coli* will form bright pink to red colonies, surrounded by zones of precipitated bile salts/neutral red complexes. There has been a great interest in recent years in the use of substrates of β-glucuronidase in the detection of *E. coli*. Studies have indicated that 94–99% of *E. coli* strains produce this enzyme, whereas few other species do. Two substrates that have garnered the greatest attention are 5-bromo-4-chloro-3-indolyl-β-glucuronide (BCIG), which forms a blue-colored compound when hydrolyzed by the enzyme, and methylumbelliferyl glucuronide (MUG), which forms a fluorescent compound upon hydrolysis. Of the two, BCIG may be more desirable because it does not require the use of an ultraviolet light, and the color remains more localized. Many investigators have suggested the addition of one of these compounds to standard media to further differentiate between *E. coli* and other bacterial species. Researchers in one study filtered water samples through membrane filters, incubated the filters on Endo agar, and transferred filters onto nutrient agar containing MUG. They demonstrated that 98% of *E. coli* isolates hydrolyzed MUG to produce fluorescent colonies within 4 h at 35 °C with no false positives. It is indicated by the AOAC that incorporation of MUG in LST broth in the MPN technique may enhance the specificity of the method for detecting *E. coli*. The AOAC is currently studying the use of a hydrophobic grid membrane/β-glucuronidase method for detection of *E. coli*. However, the β-glucuronidase test is not 100% sensitive or specific. *Shigella* spp. typically also produce the enzyme. Several β-glucuronidase-positive *Salmonella* strains have also been identified. Some *E. coli* strains do not produce the enzyme. One study has shown that a mean of 34% and a median of 15% of human fecal isolates of *E. coli* were β-glucuronidase-negative when grown in LST containing MUG. Among the consistently β-glucuronidase-negative *E. coli* strains, there are some pathogenic isolates, including the enterohemorrhagic *E. coli* O157:H7 strains.

Isolation of Pathogenic *E. coli*

Methods used to isolate *E. coli* as an indicator organism from foods have not proved to be efficient for isolating pathogenic strains of *E. coli*. This is largely because pathogenic strains often differ considerably from nonpathogenic *E. coli* in growth patterns. Pathogenic strains frequently show delayed growth at 44 and 45.5 °C, particularly when initially present in low populations. Some pathogenic strains will not produce acid and gas from lactose in LST, BGLB, or EC broths within 48 h. It has also been shown that growth in media containing sodium lauryl sulfate and growth at 44.5 °C can cause a loss of plasmids, known to encode many virulence factors associated with pathogenic *E. coli* strains. One study indicated that up to 95% of *E. coli* cells lost plasmids during selective enrichment cultures. Therefore, the methods commonly used for detection of *E. coli* as an indicator organism should not be used to attempt isolation of pathogenic strains from foods or water.

The US Food and Drug Administration Bacteriological Analytical Manual describes a procedure for isolating pathogenic *E. coli* that overcomes many of the problems associated with the MPN procedures. This method is not quantitative. Its goal is the isolation, but not enumeration, of pathogenic *E. coli* from foods. A 25-g sample of food is transferred into a flask containing 225 ml of brain heart infusion broth and is incubated for 3 h at 35 °C to allow for resuscitation of injured cells. The culture is mixed with 225 ml of double-strength tryptone phosphate broth and incubated at 44 °C for 20 h. Representative colonies are picked and tested for biochemical characteristics.

Isolation of Pathogenic *E. coli* from Clinical Specimens

The standard methods for detection and isolation of *E. coli* from foods are generally not practiced in clinical laboratories. The numbers of viable *E. coli* cells in

clinical specimens are generally much higher than in foods. Clinical specimens are usually streaked directly onto selective-differential media such as MacConkey or EMB agar without any enrichment or selective enrichment cultures. Typical lactose-fermenting colonies are picked and tested with a battery of biochemical tests (Table 1).

Detection of *E. coli* Virulence Factors

The major problem in isolating pathogenic strains of *E. coli* is that most *E. coli* possess no characteristics that have been directly associated with production of disease, such as production of toxins or invasiveness. Therefore, it is assumed that most *E. coli* strains are avirulent in healthy humans. Hence, when attempting to isolate pathogenic strains, isolated colonies of *E. coli* must be assayed for characteristics related to pathogenicity. This is particularly a problem in fecal specimens where large numbers of nonpathogenic *E. coli* are usually present.

The process of screening *E. coli* isolates most often involves determining the serogroup (O antigen) by agglutination in O-specific serum. Polyvalent sera containing antibodies against somatic (O) antigens frequently associated with enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), enteropathogenic *E. coli* (EPEC), and enterohemorrhagic (EHEC) strains are commercially available. This test is easy and quick to perform. However, the disadvantages are that the procedure does not specifically test for disease-producing potential (e.g., not all strains of an ETEC-associated serogroup are ETEC), and serological cross-reactions can occur between serogroups.

Other tests to determine the virulence potential of *E. coli* isolates include assays for toxicity. Among the best known is the ligated rabbit ileal loop test, in which cell suspensions are injected into a ligated segment of rabbit intestine. Heat-labile enterotoxin (LT)-producing bacteria will cause fluid accumulation within the segment. LT can also be detected by vascular permeability in animals and by inducement of morphological changes in tissue culture cell lines Y–I and Chinese hamster ovary (CHO). Culture filtrates of heat-stable enterotoxin (ST)-producing ETEC strains cause fluid accumulation in the intestines of infant mice. Identification of EIEC and *Shigella* spp. may include the Sereny test for invasiveness. This involves applying a suspension of bacteria to the eye of a guinea-pig. If the bacteria are capable of cellular invasion, ulceration and opacity will occur in the eye within 3 days. This test has been replaced to some extent with invasiveness assays in tissue culture cells.

The advent of molecular biological techniques has provided methods to test directly for virulence factors known or believed to be involved in pathogenesis, often without the need for bacterial isolation at all. Factors for which such tests have been developed include toxins and adhesins. Polymerase chain reactions (PCR) targeting heat-labile toxin genes of ETEC and Shiga toxin genes of EHEC have been reported. An *E. coli* species-specific PCR based on the *malB* operon (gene encoding maltose permease) that detects all *E. coli* strains was developed and applied for detecting *E. coli* in water and soft cheese. Another *E. coli*-specific PCR was also developed using primers specific for 16S ribosomal RNA genes of the organism. Recently, a multiplex PCR that can simultaneously detect ETEC, Shiga toxin-producing *E. coli*, and attaching and effacing strains of *E. coli* was also reported. These methods have been widely researched because they allow confirmation of the presence of a particular microorganism faster and more accurately than cultural methods. However, they often require special equipment or reagents that most clinical, public health, and food-microbiology laboratories do not possess. This is particularly true in developing countries, where there is a greater need for *E. coli* detection systems. Another molecular method becoming commonly used for rapid and specific detection of various types of *E. coli* involves the use of genetic probes. Radiolabelled or chemiluminescent-labelled segments of DNA complementary to regions of genes encoding toxins or other virulence factors are hybridized with the sample, and the presence of the target gene is indicated by the amount of bound radioactivity. This procedure is very sensitive; the specificity is dependent on the structure of the probe and the conditions used during hybridization. Note that genetic probes detect the genes for virulence factors, and therefore presumably the organism itself, whereas ELISA detects the virulence factor that is expressed by the organism. Genetic probes have been used to detect LT- and ST-producing bacteria in food, water and stool, adhesins of ETEC, and the invasiveness proteins of EIEC. DNA probes that utilize chemiluminescent substrates such as streptavidin/biotin system eliminate the need for radioactivity, with little reduction in sensitivity.

Enzyme-linked immunosorbent assay (ELISA) is an immunological method used for detection of *E. coli* strains. There are several forms of ELISA, but all are variations on a single theme. The advantages offered by ELISA methods include rapidity of diagnosis with little or no decrease in sensitivity from tissue culture or animal methods. ELISA has been used to detect LT and ST, adhesins of ETEC, and invasiveness-related outer-membrane proteins of EIEC in stool specimens and culture filtrates. While having great utility in detection of *E. coli* virulence factors in clinical

samples, ELISA methods often do not perform as well in food samples, due to interference from food components.

Detection of *E. coli* O157:H7

Isolation of enterohemorrhagic *E. coli* O157:H7 must be approached differently than using the methods for isolating other *E. coli* strains because *E. coli* O157:H7 grows poorly or not at all at 44 °C in many media used to isolate *E. coli*. None of the methods described above can dependably isolate *E. coli* O157:H7. However, *E. coli* O157:H7 has some biochemical differences from most other *E. coli* strains that can be exploited in isolation and identification methods (**Table 1**). *E. coli* O157:H7 ferments sorbitol slowly or not at all, and does not produce functional β-glucuronidase, whereas most of the other *E. coli* strains are positive in both tests. Further, *E. coli* O157 strains do not ferment rhamnose on agar plates, whereas 60% of non-sorbitol-fermenting *E. coli* belonging to other serogroups ferment rhamnose on agar plates. Clinical laboratories that test for *E. coli* O157:H7 usually streak fecal specimens onto plates of sorbitol MacConkey agar (SMAC) (which contains sorbitol instead of lactose). Sorbitol-nonfermentative colonies are chosen and serologically tested for the O157 and H7 antigens. Positive colonies should then be tested for Shiga toxin production. (*See* **Immunoassays**: Principles.)

Several methods for isolation of this organism from foods have been suggested. Media that can be used to test both sorbitol fermentation and β-glucuronidase activity have been developed. For example, Sorbitol MacConkey agar containing MUG is commonly used for differentiating *E. coli* O157:H7 from other *E. coli* strains. In this medium, *E. coli* O157:H7 strains form colorless and nonfluorescent colonies due to their inability to ferment sorbtiol and break down MUG. However, non *E.coli* O157:H7 strains in Sorbitol MacConkey agar containing MUG, form pink colonies that are fluorescent under UV light. The recovery rate of *E. coli* O157:H7 on Sorbitol MacConkey agar can be improved by previous enrichment in trypticase soy broth supplemented with cefixime ($0.5\,mg\,l^{-1}$) and vancomycin ($8\,mg\,l^{-1}$). Another method for improving the selectivity of Sorbitol MacConkey agar for isolating *E. coli* O157:H7 is inclusion of cefixime and potassium tellurite ($2.5\,mg\,l^{-1}$) that permit growth of *E. coli* O157:H7 but inhibit most of the other *E. coli* strains. Another study indicated that addition of novobiocin can increase the selectivity of the medium for isolating *E. coli* O157:H7. Injured or stressed cells of *E. coli* O157:H7 do not grow well on selective media such as Sorbitol MacConkey agar, especially in the presence of selective agents such as novobiocin, cefixime, and potassium tellurite. A study published from the UK reported that delaying the exposure of injured *E. coli* O157:H7 cells to selective agents by inclusion of a recovery step such as resuscitating the cells on a membrane placed on the surface of tryptic soy agar followed by transfer of the membrane onto Sorbitol MacConkey agar increased the detection of stressed cells of the pathogen without sacrificing specificity. Another medium, called Rainbow agar, has been reported to be superior to Sorbitol MacConkey agar in recovering heat-injured cells of *E. coli* O157:H7. Rainbow agar contains chromogenic substrates specific for β-galactosidase and β-glucuronidase, and produces a spectrum of colored colonies ranging from black to gray to red to blue to violet, which differentiates *E. coli* O157:H7, *E. coli* O26:H11, other Shiga toxin-producing *E. coli*, and Shiga toxin-negative *E. coli*. Enterohemolysin agar is a nonselective, differential medium used for detecting enterohemolysin expressed by about 90% of Shiga toxin-producing *E. coli*. (*See* ***Salmonella***: Detection; ***Shigella***.)

DNA probes and polymerase chain reaction (PCR) provide powerful and rapid tools for specific and sensitive detection of *E. coli* O157:H7. Most probes are directed against genes encoding Shiga toxins for detecting *E. coli* O157:H7 and other Shiga toxin-producing *E. coli* (STEC) strains. Shiga toxin production is the only trait shared by all STEC that is not found in other *E. coli* strains. Probes against all Shiga toxins of strains producing human illness, including Shiga toxin 1 (Stx1) and Shiga toxin 2 (Stx2), would be necessary to detect all STEC strains, because some strains produce only one Shiga toxin. Specificity would not be 100%, because probes against Stx1 genes would also be expected to react with Shiga toxin-producing *Shigella* strains. Additionally, it is not known if all STEC are human pathogens. Many studies have shown the capability of such Shiga toxin gene probes to detect STEC strains in colony blots and enrichment broth cultures. The probes have also been successful at detecting the Shiga toxin-producing bacteria in feces and in foods, including ground beef, raw milk, and salads. Further, a variety of PCR assays specific for attaching and effacing (eae), hemolysin, plasmid pO157, and 60-Mda plasmid genes have been developed for detecting *E. coli* O157:H7. Finally, a number of immunoassays based on somatic (O) and flagellar (H) antigens specific for *E. coli* serotype O157:H7 have been reported. An ELISA procedure utilizing a monoclonal antibody (4E8C12) specific for an outermembrane protein of *E. coli* O157:H7 was capable of detecting *E. coli* O157:H7 in foods in less than 20 h.

See also: **Escherichia coli**: Occurrence; Occurrence and Epidemiology of Species other than *Escherichia coli*; **Immunoassays**: Radioimmunoassay and Enzyme Immunoassay

Further Reading

Banwart GJ (1989) *Basic Food Microbiology*, 2nd edn. New York: Van Nostrand Reinhold.

Doyle MP and Padhye VV (1989) In: Doyle MP (ed.) *Foodborne Bacterial Pathogens*, pp. 235–281. New York: Marcel Dekker.

Hill WE (1996) The polymerase chain reaction: applications for the detection of foodborne pathogens. *Critical Reviews in Food Science and Nutrition* 36: 123–173.

Hill WE, Ferreira JL, Payne WL and Jones VM (1985) Probability of recovering pathogenic *Escherichia coli* from foods. *Applied and Environmental Microbiology* 49: 1374–1378.

International Commission on Microbiological Specifications for Foods (1988) *Microorganisms in Foods. 1. Their Significance and Methods of Enumeration*, 2nd edn. Toronto: University of Toronto Press.

Jay JM (2000) *Modern Food Microbiology*, 6th edn. Guithersburg, MD: Aspen.

Mehiman IJ (1984) In: Speck ML (ed.) *Compendium of Methods for the Microbiological Examination of Foods*, 2nd edn, pp. 265–285. Washington, DC: American Public Health Association.

Meng J, Doyle MP, Zhao T and Zhao S (2001) *Escherichia coli* O157:H7 (1997). In Doyle MP, Beuchat LR and Montville TJ (eds) *Food Microbiology: Fundamentals and Frontiers*, 2nd edn. pp. 193–213. Washington, DC:ASM Press.

Nataro JP and Kaper JB (1998) Diarrheagenic *Escherichia coli*. *Clinical Microbiology Reviews* 11: 142–201.

Padhye NV and Doyle MP (1992) *Escherichia coli* O157: H7: Epidemiology, pathogenesis, and methods for detection in food. *Journal of Food Protection* 55: 555–565.

Food Poisoning

K S Venkitanarayanan, University of Connecticut, Storrs, CT, USA
M P Doyle, University of Georgia, Griffin, GA, USA

Introduction

Pathogenic strains of *E. coli* cause more types of infections in humans than any other Gram-negative bacterium. They are a frequent cause of urinary tract infections and meningitis, and also have been associated with wound infections and pneumonia. As early as 1923, *E. coli* was suspected to be associated with infantile diarrhea. The first documented outbreak of foodborne *E. coli* infection in the United States occurred in 1971. The implicated food was imported French Camembert and Brie cheeses. *E. coli* strains that cause gastrointestinal disease have been classified into five groups: the enterotoxigenic *E. coli* (ETEC), the enteroinvasive *E. coli* (EIEC), the enteropathogenic *E. coli* (EPEC), the enterohemorrhagic *E. coli* (EHEC), and the enteroaggregative *E. coli* (EAEC). Each of the five groups of gastrointestinal pathogenic *E. coli* consists of certain serotypes that are commonly associated with that group (**Table 1**).

Enterotoxigenic *E. coli* (ETEC)

ETEC produce at least one of two types of enterotoxins, namely heat-stable enterotoxin (ST) or heat-labile enterotoxin (LT). Originally, ETEC were recognized as the etiological agents of diarrhea in piglets. Subsequently, ETEC were isolated from diarrheal stools of children and adults. ETEC cause a watery diarrhea similar to that produced by *Vibrio cholerae*, although usually not as severe in intensity. Diarrhea is usually accompanied by a low-grade fever, abdominal cramps, malaise and nausea. The mechanism of ETEC infection is similar to that of *Vibrio cholerae*. Following ingestion, ETEC cells using fimbriae attach to epithelial cells lining the ileum. The fimbriae are proteinaceous filaments that specifically recognize carbohydrate receptors on the intestinal cells. Adherence allows the bacteria to resist removal by the peristaltic movement of intestinal contents.

After adhering to the intestinal wall, ETEC produce one or more toxins (i.e., LT and ST) that induce a diarrheal response. LT is a heat-labile protein with a native molecular weight of about 86 000. LT consists of an enzymatically active 28-kDa A subunit and five identical 11.5-kDa subunits B. The A subunit, consisting of A_1 and A_2 peptides linked by a disulfide bond, is responsible for the enzymatic activity of the toxin. The B units are arranged in a ring form, and bind strongly to the ganglioside GM_1 and weakly to the ganglioside GD_1. Following binding to intestinal mucosal cells, LT is endocysted and translocated through the cell by a process involving trans-golgi vesicular transport. The A_1 peptide transfers ADP-ribosyl moiety from NAD to the α-subunit of the GTP-binding protein, G_s, which in turn stimulates adenyl cyclase activity. LT ADP ribosylates the adenylate cyclase regulatory subunit, causing the accumulation of cyclic AMP (cAMP) in the intestinal cell. This leads to an increased secretion by crypt cells and decreased absorption by villous tip cells. There are two distinct classes of LT: LT-I is immunologically related to cholera toxin; LT-II is similar to cholera

Table 1 Characteristics of groups of gastrointestinal *E. coli* pathogens

Group	*Virulence factors*	*Serogroups*
ETEC	Enterotoxins (LT, ST) Fimbriae	O6, O8, O15, O20, O25, O27, O63, O78, O80, O85, O115, O218ac, O139, O148, O153, O167
EIEC	Colonic invasiveness	O28ac, O29, O124, O136, O143, O144, O152, O164, O167
EPEC	Enteroadhesiveness factor (?)	O18ab, O18ac, O26, O44, O86, O111, O114, O119, O125, O126, O127, O128ab, O142, O158
EHEC	Shiga toxins Locus of enterocyte effacement Fimbriae (?)	O26, O111, O113, O121, O145, O157
EAEC	Fimbriae (?)	O3, O4, O6, O7, O17, O44, O51, O68, O73, O75, O77, O78, O85, O111, O127, O142, O162

ETEC, enterotoxigenic *E. coli*; EPEC, enteropathogenic *E. coli*; EIEC, enteroinvasive *E. coli*; EHEC, enterohemorrhagic *E. coli*; EAEC, enteroaggregative *E. coli*.

toxin in function but is not serologically cross-neutralizable. LT-I is produced by *E. coli* strains that are pathogenic for both humans and animals, whereas LT-II is mainly produced by strains of animal origin. Two antigenic variants of LT-II, namely LT-IIa and LT-IIb have been reported. Similar to LT-I, LT-II also increases intracellular cAMP level; however, the receptor for LT-II is GD_1, instead of GM_1.

ST is a heat-stable protein that is much smaller than LT, with a molecular weight of less than 5000. There are two forms of ST, i.e., STa and STb, but only STa is involved in human infection. STb-producing strains are usually of animal origin and do not infect humans. STa consists of 18–19 amino acids with a molecular weight of 2 kDa. Two types of STa have been reported, which include STp (ST porcine or STIa) and STh (ST human or STIb). Each of these two variants contains 13 identical amino acids. STa is produced from its precursor containing 72 amino acids, as a result of cleavage by a signal peptidase. The primary receptor of STa is guanyl cyclase C (GC-C). Stimulation of GC-C occurs once STa binds to it, resulting in the production of increased levels of intracellular cyclic guanosine monophosphate (cGMP). This ultimately results in the inhibition of sodium chloride transport and increased secretion of water from intestinal cells.

Both LT-I and ST are plasmid-encoded, whereas LT-II is chromosomally encoded. The large plasmids, approximately 50–70 MDa in size, that encode LT-I and ST usually encode the fimbriae used for adherence as well. ST genes are located inside a transposon-like element. The presence of these genes on plasmids has contributed to their spread throughout homoconjugative strains.

In developed countries with good hygiene, ETEC is not an important cause of disease. Patients typically have a mild, uncomplicated, watery diarrhea that requires only symptomatic treatment. In these countries, ETEC is most often associated with 'traveler's diarrhea', contracted while travelling in another country with lower hygienic standards. Occasionally, when food or water becomes contaminated, extensive outbreaks of ETEC diarrhea can occur. In 1975, over 2000 staff and visitors at a national park in the USA developed an infection after the park's water supply became contaminated with an ETEC strain that produced both LT and ST. Outbreaks of ETEC have been mainly associated with drinking contaminated water, but examples of foods that have been implicated in outbreaks include crab meat, curried turkey mayonnaise, raw vegetable salad, and Brie and Camembert cheeses. A recent study published by the Centers for Disease Control and Prevention, USA, reported an estimated 79 000 cases of ETEC diarrhea in the United States, with about 70% of the cases foodborne.

In developing countries, ETEC can be a major cause of dehydration and death, particularly among children under 5 years of age. This higher incidence of ETEC infection with higher mortality rates is associated with lower hygienic standards and poor nutrition. In many countries, children are believed to have two to three ETEC infections per year during the first 3 years of life. The disease caused by ETEC in these situations can be severe, and is often indistinguishable from cholera.

Enteroinvasive *E. coli* (EIEC)

EIEC strains are biochemically, genetically, and pathogenically related to *Shigella* spp. Similar to *Shigella* spp., EIEC are nonmotile, lactose negative, and lysine decarboxylase-negative. Further, serologically, EIEC and *Shigella* spp. demonstrate numerous cross-reactions. The mechanisms by which they cause disease are also very similar.

The pathogenesis of EIEC involves cellular invasion and spread, which are mediated by chromosomal and plasmid-borne virulence genes. Like *Shigella* spp., EIEC strains infect the colon by invading epithelial cells, a phenotype mediated by both plasmid and chromosomal loci. Moreover, EIEC elaborate endotoxins, which are believed to play a role in the development of diarrhea. The events that occur in the

pathogenesis of EIEC involve epithelial cell penetration, lysis of the endocytic vacuole, intracellular multiplication, movement through the cytoplasm, and extension into adjacent epithelial cells. Severe infections result in a strong inflammatory reaction leading to ulceration of the mucosa. The resulting symptoms are typical for bacillary dysentery: fever and diarrhea, with some blood and mucus in the stools. Some patients will also demonstrate chills, headache, muscle pain, and abdominal cramps.

The genes mediating invasion are carried on a 140-MDa plasmid. Another plasmid-borne gene (designated as *sen*) in EIEC, which codes for a novel 63-kDa protein (enterotoxin) has been cloned and sequenced. A mutation in the *sen* gene has been reported to cause a significant loss in the enterotoxic activity of the EIEC. Although the role of enterotoxins in EIEC pathogenesis is unclear, toxins are believed to play a critical role in the development of watery diarrhea.

EIEC strains have caused outbreaks of disease world-wide, with most reported outbreaks taking place in countries with high hygienic standards, such as the USA, the UK, France, and Australia. Despite these outbreaks, EIEC is not a common cause of gastrointestinal disease. Outbreaks have resulted from the consumption of contaminated Brie and Camembert cheese and potato salad, but water remains the most commonly implicated vehicle of infection.

Enteropathogenic *E. coli* (EPEC)

The EPEC group has historically been a collection of serogroups associated with infantile and neonatal diarrhea. Until the mechanism by which EPEC cause disease was elucidated, the group was defined principally on the basis of O and H serotypes. However, with a better understanding of the mechanism of EPEC's pathogenicity, this group of pathogens is currently defined on the basis of pathogenesis characteristics. EPEC strains do not produce LT or ST enterotoxins that are associated with ETEC. They are also noninvasive when evaluated by the Sereny test, but some strains are capable of invading tissue culture cells. Some strains do produce Shiga toxin, but these strains have since been reclassified as Shiga toxin-producing *E. coli* (STEC).

EPEC are capable of producing attaching and effacing lesions on host cells. EPEC strains harbor a characteristic large plasmid that confers adherence on the surface of cells. A three-stage model has been proposed in the pathogenesis of EPEC. The first stage involves localized adherence of EPEC to host cells, which is mediated by a type IV fimbria known as bundle-forming pilus (BFP) encoded by fourteen *bfp* genes present on the large plasmid. Studies have shown that mutations in *bfp* genes prevent expression of BFP, resulting in a failure of localized adherence of EPEC to cells. The second stage involves infliction of attaching and effacing lesions on cells, which are mediated by a 35 637-bp pathogenicity island known as the locus of enterocyte effacement (LEE). The attaching and effacing lesions consist of localized destruction of microvilli and the formation of a cup-like pedestal composed of cytoskeletal proteins at the site of EPEC binding to host cells. In the third stage, EPEC attaches intimately to host cells via a receptor followed by cellular alterations in the cytoskeleton. The *eae* gene located in the locus of enterocyte effacement encodes an outermembrane protein intimin, which confers strong attachment of EPEC to cells. Studies of mutations induced in the *eae* gene have revealed that EPEC mutants not only lack the ability to attach intimately to host cells, but also have a diminished ability to cause diarrhea in experimental human infections.

EPEC infections are most commonly associated with infantile diarrhea, which can last for as long as 14 days. Other symptoms can include fever, vomiting, and abdominal pain. When adults are affected by EPEC, symptoms are usually diarrhea, nausea, abdominal cramps, headache, and chills. EPEC infantile diarrhea most commonly occurs in tropical countries with poor hygiene. In Brazil, EPEC is the most common cause of diarrhea in children under 6 months of age. In temperate countries, EPEC are not an important cause of diarrhea, although occasional outbreaks do occur. Foods implicated in outbreaks include pork, meat pies and water.

Enterohemorrhagic *E. coli* (EHEC)

The observation that some strains of *E. coli* produce toxins that kill Vero (African green monkey kidney) tissue culture cells by Shiga toxin provides a basis for the STEC group of *E. coli*. Shiga toxins (also known as Vero toxins) inhibit protein synthesis by cleaving rRNA at a specific nucleotide, in a manner identical to the action of Shiga toxin produced by *Shigella dysenteriae*. Two major Shiga toxins have been characterized. Shiga toxin I (Stx1, VT1) is structurally practically identical to Shiga toxin, and can be neutralized by antiserum raised against Shiga toxin. Shiga toxin II (Stx2, VT2) is structurally similar to Stx1, having at least 74% nucleotide homology with Stx1, but cannot be neutralized by antiserum raised against Shiga toxin or Stx1.

The STEC group as a whole has not been associated with foodborne illnesses. However, a subdivision of

the STEC, known as enterohemorrhagic *E. coli* (EHEC), has been associated with outbreaks and sporadic cases of hemorrhagic colitis following the consumption of contaminated foods. EHEC is a more prevalently recognized cause of illness in developed countries than developing countries. Interestingly, a principal reservoir of EHEC is cattle and possibly other ruminants, whereas humans are the principal reservoir of EIEC and ETEC strains involved in human illness. For this reason, EHEC infections in humans are more often associated with foods of animal origin, or those contaminated with feces of animals than the other types of pathogenic *E. coli*. Additionally, the illnesses caused by EHEC strains are far more likely to lead to death or long-term disabilities than those caused by other *E. coli* gastrointestinal pathogens.

The two major virulence factors in the pathogenesis of EHEC infections include the production of Shiga toxins, and attaching and effacing lesions which are similar to those incurred by EPEC. The Shiga toxins are composed of a single catalytic subunit A, and five receptor-binding B subunits. The B subunits form a pentamer with a center pore through which the carboxyl terminus of the A subunits extends. The B pentamer of Shiga toxins binds to a specific glycolipid receptor known as globotiosyl ceramide (Gb_3), which is present on the cell surface. This is followed by entry of the A subunit into the cell by endocytosis and transport into the Golgi apparatus, and then to the endoplasmic reticulum. The A subunit translocates into the cytoplasm, where it catalyses the N-glycosidation of adenine 4324 of 28S ribosomal RNA, leading to a failure of peptide chain elongation and inhibition of protein synthesis. The block in protein synthesis results in the death of renal endothelial cells and intestinal epithelial cells, or any cells containing Gb_3 receptor. Both Stx1 and Stx2 are encoded by a bacteriophage related to the lambda phage.

Similar to EPEC, EHEC possess the entire LEE, encoded by a 35-kb pathogenicity island. Although the classical attaching and effacing histopathology has been observed in gnotobiotic piglets, infant rabbits, and cultured epithelial cells infected with *E. coli* O157:H7, such lesions have not been reported in clinical biopsy specimens from patients with EHEC infection. The cellular responses leading to attaching and effacing histopathology in EHEC infections include high concentrations of polymerized actin on mucosal lesions, increased levels of IP_3 and intracellular calcium. In addition to Shiga toxins, and attaching and effacing lesions, a plasmid-encoded fimbria has been identified among *E. coli* O157:H7 strains, but its importance in adherence and pathogenesis has not been demonstrated.

Most EHEC isolates belong to the serotype O157:H7, although other serotypes, such as O26:H11, and O111:NM, also have been reported in human illnesses. Hemorrhagic colitis is characterized by a watery diarrhea that progresses to grossly bloody diarrhea, indicative of significant amounts of gastrointestinal bleeding. Severe abdominal pain also results, although fever is usually not present. The illness typically lasts from 2 to 9 days. It is rarely fatal, except among the elderly and young children. However, hemorrhagic colitis and other gastrointestinal infections caused by EHEC strains are often antecedants to more serious conditions, presumably caused by action of Shiga toxins on kidneys and other organs. Hemolytic uremic syndrome (HUS) and thrombotic thrombocytopenic purpura (TTP) occasionally result from a few days to a few weeks after acquiring a gastrointestinal infection with EHEC, especially among the very young and the elderly. Both are manifestations of damage to kidneys, which can lead to kidney failure and death. The clinical features of TTP are similar to HUS, but TTP also involves damage to the central nervous system. Patients often develop blood clots in the brain, frequently leading to death.

EHEC was first implicated as a cause of human illness following two outbreaks of hemorrhagic colitis in the USA in 1982. Forty-seven people became ill after eating contaminated ground-beef sandwiches. Several outbreaks have since been reported from the USA, Canada, and UK. A chain of outbreaks involving more than 40 *E. coli* O157:H7 cases was reported from Scotland in 1996. In the UK, the number of laboratory isolations of EHEC O157 from human infections increased from 76 in 1986 to 1087 in 1997. A massive outbreak of *E. coli* O157:H7 involving more than 9500 cases and nine deaths occurred in Japan in 1996. A study published by the Centers for Disease Control and Prevention, USA, in 1999 reported an estimated 62 500 foodborne cases of *E. coli* O157:H7 annually in the United States. Outbreaks of hemorrhagic colitis have been associated frequently with inadequately cooked ground beef. Outbreaks have also been associated with consumption of a variety of other foods such as salami, roast beef, venison jerky, unpasteurized milk, apple juice, coleslaw, and contaminated municipal water. Moreover, vegetables such as lettuce, alfalfa and radish sprouts, and potatoes have also been associated with outbreaks of *E. coli* O157:H7 outbreaks.

Enteroaggregative *E. coli* (EAEC)

First described in 1987, EAEC are now recognized as a cause of infantile diarrhea, especially in cases that persist for more than 14 days. EAEC are especially a

problem in developing countries; however, an increasing number of cases have been reported in industrialized countries. EAEC are characterized by their ability to adhere to cultured cell monolayers with an aggregative or 'stacked brick' pattern.

The pathogenesis of EAEC has not been fully elucidated; however, a characteristic histopathologic pattern and a number of virulence factors have been identified. A surface fimbrial structure designated as Aggregative Adherence Fimbriae I (AAF/I) that mediates adherence to Hep-2 cells and erythrocyte hemagglutination has been identified in EAEC. A 60-MDa plasmid containing two separate gene clusters encodes AAF/I. A second fimbriae, AAF/II, that is morphologically and genetically distinct from AAF/I has also been reported in EAEC. Although EAEC is reported to produce a 4100-Da homolog of the heat-stable enterotoxin (EAST1), its role *in vivo* in pathogenicity is not clear. EAEC also produces a 108-kDa cytotoxin that incurs destructive lesions in a rat ileal loop, and is believed to be responsible for mucosal damage in patients. Lesions such as shortening of villi, hemorrhagic necrosis of villous tips, edema, and mononuclear infiltration of submucosa are observed in EAEC infection.

Typical clinical manifestations of EAEC infection include watery, mucoid diarrhea, and low-grade fever with no vomiting. In some patients, grossly bloody stools are observed. In studies with volunteers, the mean incubation period of EAEC infection was 14 h, with a range of 7–22 h. The duration of diarrhea is typically longer than 14 days. Although outbreaks of EAEC were commonly reported from developing countries such as Brazil, Chile, Mexico, India, and Bangladesh, some of the recently reported outbreaks occurred in industrialized countries, including Japan and the UK. A massive outbreak of EAEC involving 2697 schoolchildren was reported from Japan in 1993. Four outbreaks of EAEC occurred in the UK in 1994. Although foods have not been implicated as vehicles in any EAEC outbreaks, the role of food and water in the transmission of EAEC merits investigation. Diagnosis of EAEC infection involves isolation of *E. coli* from stool specimens and a demonstration of aggregative adherence pattern in HEp-2 cells. A DNA probe and a PCR assay specific for EAEC have also been reported.

Treatment

In developed countries, medical intervention in cases of gastrointestinal *E. coli* infections is rarely required. The illness typically is of short duration and resolves without any significant incident. Antibiotic therapy is usually not necessary, with the exception of newborn children with severe EPEC infection. In most other cases, treatment with antimicrobial agents does not appear to have a substantial effect on the duration or intensity of disease. A principal concern is to prevent dehydration, especially among infants. Diarrheal episodes in developing countries are generally considerably more severe than in developed countries. Oral rehydration with aqueous solutions containing sodium, potassium, chloride, bicarbonate, and glucose has substantially decreased mortality. The World Health Organization recommends a rehydration solution containing 90 mM sodium, 20 mM potassium, 80 mM chloride, 30 mM bicarbonate and 111 mM glucose.

Infection by *E. coli* O157:H7 is usually self-limited as well, although patients with severely bloody diarrhea, HUS, or TTP are hospitalized. Care must be taken to prevent dehydration and excessive blood loss. Patients who develop HUS or TTP will often require extensive treatment, including dialysis and blood transfusions.

Control and Prevention

Strains of ETEC, EIEC, and EPEC that cause human disease are usually not associated with animals. The principal source of these bacteria is believed to be feces of human carriers. Therefore, the primary step in preventing these infections is avoiding fecal contamination of foods and water. Water used for cleaning of food contact surfaces should be chlorinated to ensure absence of viable *E. coli* cells. Untreated human sewage should not be used to fertilize vegetables or crops intended for human consumption. Food handlers should be educated in proper personal hygiene and in techniques of handling food safely. Foods should be heated thoroughly to kill *E. coli* and stored under refrigeration conditions to prevent bacterial multiplication. *E. coli* O157:H7 and other EHEC serotypes differ from other *E. coli* gastrointestinal pathogens in that the strains that cause human illness are also found in the intestinal tract of domestic animals used in food production. Consumers should be certain that all foods of animal origin are thoroughly cooked or pasteurized before consumption, and care should be taken to prevent contamination of ready-to-eat foods with uncooked material. Food producers and processors should avoid contamination of animal products with feces during slaughter or milking.

Vaccination of cattle is one potential method for reducing carriage of *E. coli* O157:H7. Although a number of studies have attempted to develop a vaccine targeting various colonization factors such as intimin, an effective vaccine against *E. coli* O157:H7 in cattle has not been reported. However,

an alternative strategy involving administration of probiotic bacteria to cattle that reduce colonization of *E. coli* O157:H7 was recently reported. Protection against ETEC infection has been demonstrated in animals immunized with fimbrial antigens used by these organisms to adhere to intestinal epithelium, and in the litters of pigs immunized with these fimbriae while pregnant. Several studies aimed at inducing immunogenicity against ETEC in humans have been reported. A recent study revealed that peroral administration to children of inactivated, whole cells ETEC coupled with the B subunit of cholera toxin was effective in eliciting an IgA response against ETEC colonization factor antigen I.

See also: **Diarrheal (Diarrhoeal) Diseases**; ***Shigella***; **Vibrios**: *Vibrio cholerae*

Further Reading

Bertschinger HU, Neif V and Tschape H (2000) Active oral immunization of suckling piglets to prevent colonization after weaning by enterotoxigenic *Escherichia coli* with fimbriae F18. *Veterinary Microbiology* 71: 255–267.

Donnenberg MS and Nataro JP (2000) In: Cary JW, Linz JE and Bhatnagar D (eds) *Mechanism of Pathogenesis and Toxin Synthesis*, pp. 87–130. Lancaster, PA: Technomic Publishing.

Doyle MP (1990a) Foodborne illness: pathogenic *Escherichia coli*, *Yersinia enterocolitica* and *Vibrio parahaemolyticus*. *Lancet* 336: 1111–1115.

Doyle MP (1990b) In: Cliver DO (ed.) *Foodborne Diseases*, pp. 209–215. San Diego, CA: Academic Press.

Doyle MP (1991) *Escherichia coli* O157:H7 and its significance in foods. *International Journal of Food Microbiology* 12: 289–302.

Doyle MP and Padhye VV (1989) In: Doyle MP (ed.) *Foodborne Bacterial Pathogens*, pp. 235–281. New York: Marcel Dekker.

Gorbach SL (1987) Bacterial diarrhea and its treatment. *Lancet* ii: 1378–1382.

Gross RJ and Rowe B (1985) *Escherichia coli* diarrhea. *Journal of Hygiene* 95: 531–550.

Gyles CL (1998) Vaccines and Shiga toxin-producing *Escherichia coli* O157:H7 in animals, pp. 434–444. In: Kaper JB and O'Brien AD (eds) *Escherichia coli* and other Shiga toxin-producing *E. coli* strains. Washington, DC: ASM Press.

Mead PS, Slutsker L and Dietz V *et al*. (1999) Food-related illness and death in the United States. *Emerging Infectious Diseases* 5: 607–625.

Meng J, Doyle MP, Zhao T and Zhao S (2001) *Escherichia coli* O157: H7 (1997) In: Doyle MP, Beuchat LR and Montville TJ (eds) *Food Microbiology: Fundamentals and Frontiers*, 2nd edn, pp. 193–213. Washington, DC: ASM Press.

Nataro JP and Kaper JB (1998) Diarrheagenic *Escherichia coli*. *Clinical Microbiology Reviews* 11: 142–201.

Nataro JP, Steiner T and Guerrant RL (1998) Enteroaggregative *Escherichia coli*. *Emerging Infectious Diseases* 4: 251–261.

Savarino SJ, Hall ER, Bassily S *et al*. (1999) Oral, inactivated, whole cell enterotoxigenic *Escherichia coli* plus cholera toxin B subunit vaccine: results of the initial evaluation in children. *Journal of Infectious Diseases* 179: 107–114.

Zhao T, Doyle MP, Harmon BG *et al*. (1998) Reduction of carriage of enterohemorrhagic *Escherichia coli* O157:H7 in cattle by inoculated with probiotic bacteria. *Journal of Clinical Microbiology* 36: 641–647.

Occurrence and Epidemiology of Species other than *Escherichia coli*

P Guglielmetti, Università di Siena, Siena, Italy
A Bartoloni, Università di Firenze, Firenze, Italy

Background

The Enterobacteriaceae are a large family of microorganisms, widely distributed in nature, that may have either saprophytic or pathogenic interactions with the host. In this article, some members of the Enterobacteriaceae, other than *Escherichia coli, Salmonella* spp., *Shigella* spp., and *Yersinia enterocolitica*, are reviewed, with particular reference to their occurrence, isolation, and identification procedure and their role in foodborne poisoning. (*See* ***Shigella***; ***Yersinia enterocolitica***: Properties and Occurrence.)

Characteristics and Occurrence

The very large group of Gram-negative bacteria in the Enterobacteriaceae family are among the most important organisms in the field of human health. The bacteria included in this family share properties. They are Gram-negative, nonsporing, straight rods (0.3–1.0 μm by 1.0–6.0 μm), motile by peritrichous flagella, or nonmotile. They grow well on peptone, meat extract, and usually MacConkey's media, both aerobically and anaerobically, often with gas production by fermentation, rather than oxidation. They are catalase-positive and oxidase-negative, reduce nitrate to nitrite, and have a 39–59% guanine and cytosine (G + C) content of deoxyribonucleic acid (DNA).

The family of Enterobacteriaceae includes 26 genera and more than 60 species. Enterobacteriaceae are widely distributed on plants, fruits, vegetables in

the soil, and water, and in the intestine of humans and animals. Some strains occupy a limited ecological niche (e.g., *Salmonella typhi*); others (e.g., *Klebsiella* and *Proteus*) are widely distributed in the environment.

Only organisms in four genera – *Escherichia, Salmonella, Shigella*, and *Yersinia* – have been clearly documented as enteric pathogens. The occurrence and characteristics of genera potentially implicated in intestinal disease, but for which a definitive role has to be confirmed, are discussed here with special reference to food poisoning.

Citrobacter

There are three species in the *Citrobacter* genus that may have some importance. *Citrobacter freundii* is isolated from human and animal feces, as well as from extraintestinal specimens. It is also commonly found in soil water and food. *Citrobacter diversus* is a rare but important agent of neonatal meningitis and infections in compromised hosts. *Citrobacter amalonaticus*, widely distributed in the environment, is found in a variety of human clinical specimens as an opportunistic pathogen.

Edwardsiella

The most frequently identified species in the *Edwardsiella* genus is *E. tarda*, which is rarely isolated but a well-documented cause of extraintestinal infections. It is suspected, but not firmly established, to be an enteric pathogen. Other species in the *Edwardsiella* genus are extremely rare.

Enterobacter

All *Enterobacter* species are found in water, sewage, soil, and vegetables. *Enterobacter cloacae* is the most frequently isolated *Enterobacter* species from humans and animals. Its role as an enteric pathogen has not been demonstrated. However, it may be an opportunistic pathogen and has an increasing importance in hospital-acquired infections (intensive care units, emergency units, urology). *Enterobacter cloacae* is less susceptible to chlorination than *Escherichia coli*. It may also be isolated in meat, hospital environments, and on the skin of man as a commensal.

Enterobacter sakazakii has no importance as an enteric pathogen, but it is a rare cause of neonatal meningitis and sepsis; it occurs in the environment and in food.

Enterobacter agglomerans is occasionally isolated from clinical specimens and has no conclusive importance as an enteric pathogen. *Enterobacter aerogenes*, phenotypically very similar to *E. gergoviae*, has been identified in dairy products and water.

Hafnia

The only species in the *Hafnia* genus is *H. alvei*, previously known as *Enterobacter alvei*. It is frequently found in clinical specimens, especially in feces from healthy individuals. This organism has been reported by some authors as a potential causative agent of intestinal disease. It has also been isolated from water and dairy products.

Moellerella

The genus *Moellerella* has been recently described (1984), and it includes only one species, *M. wisconsensis*, which has been reported by some authors as a potential agent of gastroenteritis. *Moellerella wisconsensis* has been mainly isolated from human feces.

Morganella

Morganella morganii is the only species in the genus *Morganella*. This organism was once assumed to be a cause of diarrhea, but confirmation about its role in intestinal disease has been lacking in recent years. It is now considered an opportunistic pathogen, and has been isolated from blood, sputa, and pus in patients with bacteremias, respiratory tract, and wound infections. *Morganella morganii* is widely distributed in nature and is commonly identified from the intestines of humans, dogs, other mammals, and reptiles.

Proteus

Among the species in *Proteus* genus, *P. mirabilis* and *P. vulgaris* may cause infections in humans. The former is more frequently identified in clinical specimens than *P. vulgaris*, mainly in human urinary tract infections. *Proteus* urinary tract infections are more commonly found in hospitalized and compromised hosts. Approximately 25% of the healthy population are intestinal carriers. The patients may infect themselves by acquiring the environmental strains or with their own flora. *Proteus* urinary tract infections may give rise to bacteremias that are often fatal.

Proteus mirabilis has been suggested as a possible causative agent of outbreaks of gastroenteritis, resulting from the consumption of contaminated food. Its role as a causative agent of acute gastroenteritis is very difficult to assess because of the high carrier rate in healthy individuals.

Providencia

Among the species in the *Providencia genus, P. alcalifaciens* may be isolated from human and animal feces. The role of *P. alcalifaciens* in disease production is not known.

Serratia

Serratia marcescens is an important opportunistic pathogen, is responsible for many nosocomial infections, and is not an intestinal pathogen. Many other *Serratia* species have been described, all of which are widely distributed in the environment (soil, water, plants, intestine of different species).

Klebsiella

Klebsiellae are opportunistic pathogens that may give rise to several extraintestinal infections. An increase in *Klebsiella* species infections has been reported, particularly in those hospitals where the strains isolated were multiple-antibiotic-resistant. The main reservoir of these strains is considered to be the human intestinal tract, whereas hands are the main vehicles for transmission. Enterotoxin-producing strains have been described.

Klebsiellae are widely distributed in nature, occurring in botanical, aquatic, and soil environments. The two species *K. terricola* and *K. planticola* are thought to be pure environmental strains. *Klebsiella pneumoniae* and *K. oxytoca* are normally found in human and animal intestinal tracts, but in lower numbers, if compared with *E. coli*. *Klebsiella oxytoca* may also be isolated from plants and aquatic environment. Klebsiellae are not considered enteric or foodborne pathogens, but their role in determining intestinal disease is not excluded.

Many other genera and species are included in the large family of Enterobacteriaceae, but only those genera that may be important in food contamination and as opportunistic pathogens are described here.

Growth Factors, Isolation, and Identification

In common with all Enterobacteriaceae, the bacteria listed above grow very well on a broad variety of culture media. The isolation depends on the origin of the specimen. In clinical microbiology laboratories, specimens from a normally sterile body site are cultured on blood or chocolate agar. These media are very rich in nutrient elements and allow the growth of many strains. All the Enterobacteriaceae grow well on such media. Specimens that are likely to be a mixture of microorganisms (feces and vomitus) are generally cultured on selective media to enhance the recovery of the Enterobacteriaceae. A preliminary grouping of enteric bacteria and other Gram-negative organisms may be carried out using media such as MacConkey agar or eosin methylene blue agar. The characteristic appearance of the most common organisms described on blood agar or MacConkey agar allows a preliminary identification of these bacteria. (*See* **Food Poisoning**: Tracing Origins and Testing.)

When a small number of Enterobacteriaceae are thought to be present in the specimen, enrichment techniques in liquid media may be of value to isolate them. This is generally the case for the bacteria previously reviewed, which are present in lower numbers when compared with *E. coli*. Enrichment techniques are associated with subcultures in appropriate selective media in which the presence of antibiotics and other substances makes the isolation of the suspected microorganism easier. For example, the selectivity of the media for *Klebsiella* isolation is based upon the high resistence of this genus to carbenicillin, in contrast to other Enterobacteriaceae. Selective media are routinely used in clinical microbiology laboratories to isolate Enterobacteriaceae from feces. Commonly used selective media for Enterobacteriaceae are MacConkey agar, xylose–lysine–deoxycholate agar, or Hektoen enteric agar. Most of these media inhibit the growth of Gram-positive and commensal organisms and differentiate between lactose fermenters and nonlactose fermenters. Special enrichment and selective media are often available according to the species investigated. These kinds of media are very efficient for isolating *Morganella, Serratia, Proteus*, and *Providencia*. For example, when feces are examined for *Proteus*, the rate of isolation is increased from 8.2 to 23.6% for *P. mirabilis*, when primary plating is preceded by enrichment with tetrathionate or selenite broth. Selective or enrichment methods are not available for *Citrobacter, Enterobacter*, and *Hafnia*, and for *Edwardsiella*, little information exists on selective isolation. Although klebsiellae are usually present in the intestinal tract in low numbers, their elevated and mucoid appearance in agar media is very characteristic and easy to identify. However, the isolation of this microorganism from polymicrobic sources can be facilitated by using selective media.

The identification of Enterobacteriaceae at the genus and species level is based on several conventional biochemical tests. The suspected colony is inoculated, from a pure culture in nonselective media, into a series of tubes with different media for each biochemical test; the tests are read after 24 and 48 h. At present, the procedure is simplified by using a series of miniaturized and standardized tests that are available commercially. These identification procedures are utilized in the large majority of clinical microbiology laboratories.

To isolate and identify these bacteria from food, enrichment procedures are normally performed.

Role in Foodborne Poisoning

In general, the microorganisms reviewed are part of the normal intestinal flora. Mainly for this reason, it is very difficult to establish their role in intestinal disease. More than the isolation and identification of a single species in feces of patients with diarrhea and/or in a suspected food, it is the epidemiological analysis of the episode that is important, especially in the case of a foodborne outbreak. Moreover, some of these organisms, mainly *Enterobacter, Proteus, Serratia*, and *Klebsiella*, may be isolated from food without any evidence of a concomitant intestinal pathology in consumers. The high frequency of recovery and count of the *Klebsiella–Enterobacter–Serratia* group indicates that these bacteria are natural flora of some foods, such as vegetables, and are not necessarily contaminants from humans.

In general, proof of their role in causing illness and their mode of transmission is lacking. For this reason, the reports on foodborne poisoning caused by these Enterobacteriaceae are sporadic and lack definitive confirmation.

Nevertheless, some authors have reported the association of these bacteria with human intestinal disease. Furthermore, the evidence of the production of an enterotoxin by some of these bacteria suggests their hypothetical role as agents of intestinal disease.

Reports are available for *Citrobacter, Edwardsiella, Enterobacter, Hafnia, Moellerella, Proteus, Providencia*, and *Klebsiella*.

Citrobacter has been isolated from feces of patients with diarrhea and from apparently normal persons. The same observations were obtained in animals. In adults with diarrhea, the microorganism was more often found than in asymptomatic controls; *Citrobacter* in symptomatic patients was above all isolated in pure culture. Unfortunately, the lack of data concerning serotypes and biological activity of the isolated strains did not allow the formation of a consensus opinion concerning their pathogenicity. *Citrobacter* has also been isolated from feces of patients and from implicated food in a few foodborne disease outbreaks. Patients had mild to serious diarrhea; symptoms were abdominal cramps, fever, chills, headache, tenesmus, vomiting, nausea, dizziness, and fainting. The documented incubation period in one of these outbreaks was 10–15 h, and the duration of the illness was 12 h. Strains potentially implicated in these outbreaks were isolated from corn pudding, spoiled milk, raw milk, macaroni, meat, liver sausage and smoked meat.

After an outbreak in a nursery, *Proteus mirabilis*, serotype 09:H2, was isolated from spaghetti with cheese, and from the feces of symptomatic children who had eaten the implicated food. The role of *Proteus* species has been suspected in other foodborne outbreaks, but the lack of information about other intestinal pathogens, particularly staphylococci and *Clostridium perfringens*, or in some cases the association of well-known agents of gastroenteritis, does not permit a final conclusion.

A well-documented role of *Proteus* spp. and *Klebsiella* spp. in the genesis of foodborne infections is that of histamine fish (scombroid) poisoning. Scombroids, in common with many other Atlantic and Pacific fish, contain high levels of histidine in their flesh. When the fish are inadequately cooled, *Proteus* and *Klebsiella* may multiply rapidly, decarboxylating histidine into histamine and other related products. These substances are responsible for an allergy-type illness, which starts within an hour of ingestion of the implicated fish. (*See* **Fish**: Spoilage of Seafood.)

Organisms belonging to the *Providencia* genus have been isolated more frequently in children and adults with diarrhea, compared with asymptomatic subjects. In a foodborne disease outbreak, *Providencia*, associated with *Proteus mirabilis* and *Providencia rettgeri*, has been isolated from patients with watery diarrhea and from the suspected food (chicken fricassee).

Klebsiellae are considered normal inhabitants of the human intestinal tract. Nevertheless, occasional reports suggest their potential role as agents of gastroenteritis. Several strains of *K. pneumoniae* gave positive responses in rabbit ileal loop, and a heat-stable enterotoxin, similar to the *E. coli* toxin, has been purified and characterized. A foodborne disease outbreak that involved 190 people has been associated with *Klebsiella* contamination of food, but no definitive evidence was produced. *Klebsiella* is a well-established cause of opportunistic infections, particularly in hospitalized patients. Food prepared and supplied by hospital kitchens is at risk of colonization by Enterobacteriaceae. *Klebsiella* was identified in high numbers from 12 foods in a hospital kitchen and has been isolated from vegetables and seeds, sugar, and milk. The most probable origin of these contaminants is the incoming food, but spread of these microorganisms in the hospital environment may also occur by means of colonized subjects. In addition, other Enterobacteriaceae, such as *Enterobacter, Serratia, Citrobacter*, and *Proteus*, are commonly found in vegetables.

Like *K. pneumoniae, Enterobacter cloacae* in the rabbit ileal loop is strongly positive in certain strains, and a heat-stable toxin, resembling the heat-stable enterotoxin of *E. coli*, has been partially purified and described. The presence of such toxins may be

the *in-vitro* evidence of the potential pathogenicity *in vivo*. Nevertheless, there are few reports of foodborne outbreaks associated with *Enterobacter*.

Hafnia alvei has been associated with an outbreak of mild gastroenteritis among hospitalized individuals and staff members. It has been isolated from the feces of patients with diarrhea and from milk that was implicated as the vehicle of the outbreak.

Edwardsiella has been identified in the feces of both patients with diarrhea and asymptomatic subjects. The role of this genus in intestinal disease is unknown.

See also: **Fish**: Spoilage of Seafood; **Food Poisoning**: Tracing Origins and Testing; ***Shigella***; ***Yersinia enterocolitica***: Properties and Occurrence

Further Reading

Balows A (ed.) (1991) *Manual of Clinical Microbiology*, 5th edn. Washington, DC: American Society for Microbiology.

Drasar BS and Barrow PA (1985) *Intestinal Microbiology*. Washington, DC: American Society for Microbiology.

Holt JG and Krieg RN (eds) (1984) *Bergey's Manual of Systematic Bacteriology*, vol. 1. Baltimore, MD: Williams & Wilkins.

Isemberg HD, Schoenknecht FD and von Graevenitz A (1979) *Collection and Processing of Bacteriological Specimens. CUMITECH 9*. Washington, DC: American Society for Microbiology.

Le Minor L and Veron M (1982) *Bacteriologie Medicale*. Paris: Flammarion Médecin-Science.

Food Poisoning by Species other than *Escherichia coli*

P Guglielmetti, Università di Siena, Siena, Italy
A Bartoloni, Università di Firenze, Firenze, Italy

Background

Enterobacteriaceae other than *Escherichia coli*, *Salmonella* spp., *Shigella* spp., and *Yersinia enterocolitica* have rarely been reported as causes of food poisoning. This article analyzes the role, clinical features and characteristics, detection, treatment, and prevention of foodborne disease associated with this large group of Enterobacteriaceae. (*See* ***Shigella***; ***Yersinia enterocolitica***: Properties and Occurrence.)

Role in Foodborne Poisoning

Among the members of the large Enterobacteriaceae family, *E. coli*, *Salmonella* spp., *Shigella* spp., *Yersinia enterocolitica*, and pathogenic strains of *E. coli* are well-established agents of bacterial food poisoning and foodborne infections. Although other genera in this family (*Citrobacter*, *Edwardsiella*, *Enterobacter*, *Hafnia*, *Moellerella*, *Proteus*, and *Klebsiella*) have rarely been reported as the cause of foodborne outbreaks, their definitive role has not been demonstrated.

It is extremely difficult to evaluate these organisms as causative agents of intestinal disease for many reasons. First, these Enterobacteriaceae are part of the normal intestinal flora, and their presence in both symptomatic and asymptomatic individuals rules out their role as causative agents of intestinal disease. Moreover, the higher prevalence of certain species in stool culture in subjects with diarrhea, compared to that in healthy people, could be caused by overgrowing of these nonpathogenic bacteria as a result of changing conditions associated with the diarrheal disease.

Some of these organisms are well-established opportunistic pathogens but only under certain circumstances, such as in hospitalized patients or in compromised hosts. Although the susceptibility of the individuals ingesting the potential pathogenic microorganism may have a very important role, until now, no significant association has been reported between these Enterobacteriaceae and intestinal disease in infants, old people, malnourished people, and immunocompromised patients.

Other conditions specifically linked to the strains may have an importance in transforming these harmless commensals into facultative intestinal pathogens. Such conditions include the virulence or enterotoxigenicity of the ingested strains and the number of the organisms ingested. That certain strains among the *Klebsiella* and *Enterobacter* genera might be responsible for some outbreaks is suggested by reports of the production of heat-labile (LT) and heat-stable (ST) enterotoxins, which are similar to the enterotoxins produced by *E. coli*.

Another problem in interpreting the role played by these Enterobacteriaceae in outbreaks of food poisoning is that they are widely distributed in the environment and are commonly found as commensals and contaminants of food. Under favorable growth conditions, they multiply and may be isolated in large numbers in food implicated in outbreaks.

To strengthen the hypothesis, the suspected strain isolated from the feces of symptomatic individuals might be identified in implicated food to which

outbreaks have been attributed. Its concurrent recovery in vomitus of patients and in feces of subjects who had eaten the implicated food, even if asymptomatic may be important for a better understanding of the outbreaks. (*See* **Food Poisoning**: Tracing Origins and Testing.)

A factor that may cause ambiguous interpretations is that samples of the suspected food are generally contaminated after they were eaten, and the rapidly growing cultures can outgrow and completely mask the causative organism.

Information about the diffusion of these Enterobacteriaceae in nature may be found in the preceding article.

Members of genera *Citrobacter*, *Edwardsiella*, *Enterobacter*, *Hafnia*, *Moellerella*, *Proteus*, and *Klebsiella* have been associated with food poisoning.

Clinical Features and Characteristics

Citrobacter

In a few food-poisoning epidemics, *Citrobacter* has been isolated from symptomatic patients with diarrhea, from food handlers, and, occasionally, from implicated food.

In the cases where the suspected organism has been identified, the clinical picture of the disease varied from mild to intense diarrhea, with abdominal cramps, nausea, vomiting, fever, chills, headache, tenesmus, dizziness, and fainting.

Citrobacter has been identified as a suspected agent of outbreaks in milk, meat, spaghetti, vegetables, and liver sausage.

Edwardsiella

Edwardsiella tarda has been isolated from the feces of both asymptomatic and symptomatic individuals with diarrhea. This organism has also been identified from the feces of food-industry workers, but its role in foodborne disease outbreaks is unknown.

Enterobacter and *Hafnia*

In 1923 and 1932, two food-poisoning epidemics were associated with organisms of the *cloacae–aerogenes* group, but the symptoms and incubation periods were consistent with staphylococcal intoxications.

Hafnia alvei, previously known as *Enterobacter alvei*, has been associated with an outbreak of mild gastroenteritis where milk was suspected as the vehicle. Other contemporary cases have been reported in the same city where the hospital was located, and a common source that served the implicated food has been identified. In this investigation, the analysis of the serotype showed that the strains isolated from the hospital and urban cases were identical.

A fatal case of food poisoning has been reported in which a man became ill after eating a stew. The incubation period of illness was 4 days. *Hafnia alvei* was isolated from the patient but not from the suspected food, which was discarded.

The role of the organisms in the *Enterobacter* and *Hafnia* genera as potential intestinal pathogens is supported by, besides clinical and epidemiological reports, the observation that *Enterobacter cloacae* has given a positive response in the isolated rabbit ileal-loop test. The toxin implicated has been purified, and the analysis has reported its similarity to the heat-stable enterotoxin (ST) produced by *Escherichia coli*.

Moellerella

Moellerella wisconsensis, from the newly described genus *Moellerella*, has been frequently isolated from the feces of symptomatic individuals. Most of the isolates have been from children with diarrhea. Its role in foodborne disease outbreaks is unknown, but it seems to be an intestinal potential pathogen for humans.

Proteus

The presence of strains belonging to the *Proteus* genus is normal in human and animal intestinal tracts. In the past, many cases of food poisoning, in which *Proteus* was suspected as potential causative agent, have been reported before the role of staphylococci and *Clostridium perfringens* in determining acute gastroenteritis was known. Nevertheless, foodborne disease outbreaks have been reported where *P. mirabilis* has been isolated from feces of subjects with diarrhea in the absence of *Salmonella* spp. and staphylococci.

The clinical features included nausea, vomiting, diarrhea, abdominal cramps, collapse, cyanosis, and sleepiness. Fever was reported in only a few cases. The incubation period was 3–5 h after eating the implicated food. In one of these epidemics, *P. vulgaris* has been isolated from inadequately cooked brawn prepared from brine-pickled meat. Another reported outbreak involved a suspected *P. mirabilis* strain that was isolated in large numbers from the vomitus and feces of symptomatic subjects and from a suspected baked ham. The clinical features and incubation periods suggested the hypothesis of staphylococcal intoxication, but the suspected food was not analyzed for staphyloenterotoxin.

Histamine fish poisoning was responsible for 182 among 2745 (6.6 %) foodborne disease outbreaks

of known etiology reported to the Centers for Disease Control (USA) from 1972 to 1986. (*See* **Histamine**.)

In one outbreak, a strain of *Klebsiella pneumoniae*, capable of producing large quantities of histamine, was isolated from tuna. In general, *Proteus* spp. are responsible for histamine food poisoning.

The clinical features of the illness are characterized by symptoms resembling those of a histamine reaction within 5–60 min of eating the implicated fish. Burning of the mouth and throat, headache, dizziness, and flushing are common. Other symptoms in descending order of frequency are abdominal cramps, nausea, vomiting and diarrhea. Urticaria and bronchospasm may also occur. Symptoms of intoxication are self-limiting and usually resolve in a few hours; in general, no therapy is required. Histamine fish poisoning is reported all the year round primarily from coastal areas.

Klebsiella

Klebsiellae are well-established opportunistic pathogens and are not normally considered as causative agents of intestinal disease. Nevertheless, these organisms have been occasionally reported as potential pathogens in acute gastroenteritis. *Klebsiella pneumoniae*, like *Enterobacter cloacae*, causes a positive rabbit ileal-loop response, and a heat-stable enterotoxin similar to that produced by *E. coli* has been purified and characterized.

A possible outbreak of foodborne *Klebsiella* infection has been reported in which 190 people who had eaten at a boarding school developed gastroenteritis. The suspected food was eaten during lunch. All the subjects developed watery diarrhea with violent abdominal pain 10–15 h after eating the implicated meal. Fever was not reported, and all the subjects recovered within 24 h without antibiotic therapy. Klebsiellae were identified from the soup and beef cooked in rice. *Klebsiella* was the predominant organism in stool culture and has been also identified in high numbers from the suspected food. However, these observations cannot establish these organisms as primary causative agents of the outbreak because the food was improperly stored prior to sampling.

Klebsiellae are very important agents of opportunistic infections in hospitalized patients. Food prepared and supplied by hospital kitchens is at risk of colonization by Enterobacteriaceae. Klebsiellae were identified in high numbers from 12 foods in a hospital kitchen and have been isolated from vegetables, seeds, sugar, and milk. The more probable origins of these contaminants are the incoming foods, but spread of these microorganisms in the hospital environment may also occur by means of the colonized subjects. Other Enterobacteriaceae, such as *Enterobacter*, *Serratia*, *Citrobacter*, and *Proteus*, are also commonly found as contaminants in food prepared in hospital kitchens. Hospitalized patients may be colonized by these organisms and may have a higher risk of developing opportunistic infections.

Detection

The detection and identification of Enterobacteriaceae in suspected specimens are not difficult. Appropriate specimens for laboratory diagnosis include feces, vomitus, gastric contents, and suspected and nonsuspected food. Cultures of the food preparation environment and food handlers may also be useful.

The laboratory should be alerted to investigate also the organisms considered part of the normal flora, as well as the common agents of food poisoning. The quantity of the isolate per gram from the incriminated food item is important, but the specimen analyzed may be derived from properly stored food after it was served. The count of the organism in feces of the patients is not significant; it is more important to demonstrate the common source of origin of the isolates. Serotyping, susceptibility to a series of antibiotics, and bacteriocin analysis (as well as other methods) are utilized in confirming that isolates are derived from a common source. Plasmid analysis may also be useful.

Detailed information about isolation and identification procedures is given in the preceding article. (*See* ***Escherichia coli***: Occurrence and Epidemiology of Species other than *Escherichia coli*.)

Treatment

The food poisoning associated with these strains of Enterobacteriaceae is generally self-limiting. Supportive measures are the mainstay of therapy in these cases. Fluid replacement is the major consideration in all of these illnesses.

Antibiotics play no role, either in therapy or in prophylaxis. Antiperistaltic agents are of little benefit in controlling diarrhea and are contraindicated in patients with symptoms that suggest the responsibility of an invasive pathogen (fever, or leucocytes in feces).

The syndrome associated with histamine fish poisoning may be relieved by antihistamines.

Prevention and Control

The mainstay of food-poisoning prevention is the proper selection, preparation, and storage of food. Inadequate refrigeration is the single most frequent factor responsible for foodborne disease outbreaks.

Enterobacteriaceae grow very well in food at temperatures ranging from 4 to 60 °C. Growth may be prevented if cold food is properly refrigerated and if hot food is held at temperatures above 60 °C before serving.

Poor personal hygiene by food handlers frequently contributes to contamination. Education of the public, nurses, physicians, and those working in eating establishments is crucial in preventing food poisoning by these Enterobacteriaceae, as well as by all the other pathogens implicated in intestinal disease. (*See* **Catering:** Nutritional Implications.)

Reporting all the suspected episodes to public health authorities is essential both in investigating an outbreak and in preventing additional cases.

See also: **Catering**: Nutritional Implications; ***Escherichia coli***: Occurrence and Epidemiology of Species other than *Escherichia coli*; **Food Poisoning**: Tracing Origins and Testing; **Histamine**; ***Shigella***; ***Yersinia enterocolitica***: Properties and Occurrence

Further Reading

Balows A (ed.) (1991) *Manual of Clinical Microbiology*, 5th edn. Washington, DC: American Society for Microbiology.

Drasar BS and Barrow PA (1985) *Intestinal Microbiology*. Washington, DC: American Society for Microbiology.

Gorbach SL (ed.) (1986) *Infectious Diarrhea*. Boston, MA: Blackwell Scientific Publications.

Mandell GL, Douglas RG and Bennett JE (eds) (1990) *Principles and Practice of Infectious Diseases*, 3rd edn. New York: Churchill Livingstone.

Rienman H and Bryan FL (eds) *Food-borne Infections and Intoxications*, 2nd edn. New York: Academic Press.

ESSENTIAL FATTY ACIDS

R Shireman, University of Florida, Gainesville, FL, USA

Essentiality of Polyunsaturated Fatty Acids

In the late 1920s, the polyunsaturated fatty acid (PUFA) linoleic acid (18:2*n*-6) was first described as necessary in the diet of rats to promote normal health. Soon, it was apparent that both 18:2*n*-6 (α-linoleic) and 18:2*n*-3 (α-linolenic) are essential in the diet because animals require these for growth and function of all tissues but lack the capacity for inserting double bonds in the *n*-6 and *n*-3 positions. In the 1930s, G. O. Burr and M. N. Burr described nutritional essential fatty acid deficiency (EFAD) in a variety of animals, and A. E. Hansen demonstrated its existence in children. In ensuing years, work in the laboratories of R. T. Holman and S. B. Johnson established normal levels of total serum fatty acids (FA) in humans of different ages, as well as the distributions in various serum lipid classes (e.g., triacylglycerols, phospholipids). These data allowed comparisons of serum fatty acids of the *n*-9, *n*-6, and *n*-3 series in normal and disease states. The single best index of essential fatty acid status was reported first to be the total serum *n*-6 concentration. Later, it was realized that in deficient states, endogenous synthesis of the 20:3*n*-9 from 18:1*n*-9 occurs; thus, the serum level of 20:3*n*-9 provides another index of status. Essential fatty acids (EFAs) may be classified in two ways: (1) those essential in the diet because animals cannot synthesize *n*-6 and *n*-3 FA *de novo* and (2) PUFA with essential functions. As pointed out by A. A. Spector, linoleic acid appears to be essential not so much because of an immediate cellular function, but because it is the precursor of 20:4*n*-6, which has numerous essential functions. Likewise, the importance of dietary 18:3*n*-3 is that it is the precursor of 22:6*n*-3, which is found in membrane phosphatidylethanolamine (PE) and phosphatidylserine (PS) in large proportions in nervous tissue. Animal cells are able to elongate and desaturate dietary 18:2*n*-6 and 18:2*n*-3 to form two long-chain fatty acids, arachidonic (AA, 20:4*n*-6) and docosahexaenoic (DHA, 22:6*n*-3). Since these are essential for cellular growth and function, they are usually listed as EFA. However, 10:5*n*-3 (eicosapentaenoic, EPA) has many effects in cells, but it has no proven essential functions. DHA (22:6*n*-3) has been recognized as especially important for premature and newborn infants. Some experts suggest using the terms 'indispensable' and 'conditionally dispensable'; for example, if 20:4*n*-6 is present in the diet, there is an absence of deficiency symptoms in healthy adults who have little or no dietary 18:2*n*-6. Thus, the 'true' linoleate requirement depends greatly upon the intake of other FA.

Essentiality is based on biochemical evidence and functional measures associated with deficiency. Numerous experiments with nonhuman primates and other animals have shown that functional development of the retina and occipital cortex in the fetus and neonate is abnormal in α-linolenic acid (18:3*n*-3) deficiency because of a subsequent lack of 22:6*n*-3 production. Such studies are difficult to perform because two or more generations of animals usually have to continue on the deficient diet before definitive effects are seen. Biochemical interactions and various conversions of dietary PUFA make it difficult to state an exact requirement for the EFAs. PUFA intakes among adults are highly variable, but the average estimated current intake is shown in Table 1. Major sources of individual PUFAs are shown in Table 2. Most diets contain much more *n*-6 than *n*-3, unless supplemented with products such as fish oil or flaxseed oil.

Essential Fatty Acid Deficiency

Human EFAD of dietary origin is not common in Western countries, and most EFAD research has been conducted on rats. It may occur in patients on long-term total parenteral nutrition (TPN) when EFAs are not included in the TPN formulation. In countries where protein malnutrition occurs, EFAD may occur because of either very low dietary EFA or lack of enzyme synthesis. Originally, it was thought that 18:3*n*-3 could substitute for 18:2*n*-6, but this proved not to be so because the products of the metabolism of these PUFA differ in function. Dietary 18:3*n*-3 does not alleviate the dermatitis of EFAD nor promote growth to the same extent as 18:2*n*-6. The estimated minimal 18:3*n*-3 dietary requirement for children is 0.54% of calories, whereas that for 18:2*n*-6 is about 1.1%. The estimated minimal requirement of 18:3*n*-6 for adults is 1–2 energy percent, with an optimum of 3–6% (approximately 1 tablespoon of oil); the optimum for 18:2*n*-3 is about 2%. The estimated ideal intake range of *n*-3/*n*-6 is 1:2 or 1:3. In both children and adults, the ratio of the two should be considered, because an overabundance of either suppresses some of the desaturation/elongation reactions for the other.

Primary (dietary-related) and secondary etiologies of EFAD in humans are listed in Table 3. There may be other conditions that are secondary to gastrointestinal (GI) pathology or enzyme defects. In cases such as chronic liver disease or hereditary enzyme deficiencies, the EFAD is not due to a lack of 18:2*n*-6 and 18:3*n*-3 in the diet but is due to an inability to metabolize them further to the necessary end products (20:4*n*-6 and 22:6*n*-3). Some experts contend that if 18:2*n*-6 is sufficient in the diet, typical symptoms of EFAD do not occur, but it is accepted by most that *n*-3 FA are needed for normal neuronal development and visual acuity, based on animal studies. The importance of the long-chain 20:4*n*-6 and 22:6*n*-3 in humans is illustrated by the effects of genetic enzyme deficiencies such as Sjogren–Larsson syndrome and acrodermatitis enteropathica; despite normal intakes of 18:2*n*-6, there are severe symptoms of EFAD including mental retardation, ichthyosis, and very abnormal prostaglandin metabolism (Table 4).

In the past, a serum triene/tetraene ratio of 1 was considered a biochemical indication of serious EFAD. An EFA-deficient diet results in insufficient 18:2*n*-6, 20:4*n*-6, and 22:6*n*-3 in tissues, but there is also marked enrichment in eicosatrienoic acid (20:3*n*-9)

Table 1 Estimated average intake of PUFA in Western diets

Polyunsaturated fatty acid	*Intake (g per day)*
18:2*n*-6	22.5
18:3*n*-3	1.2
20:4*n*-6	0.6
20:5*n*-3	0.05
22:6*n*-3	0.08
Total *n*-6/*n*-3 ratio	16.7

Table 2 PUFA major sources and functions

Fatty acid	*Major sources*	*Major functions*
18:2*n*-6 α-Linoleic	Safflower, sunflower, sesame, soybean, and corn oils	Component of epidermal permeability barrier; precursor of 20:4*n*-6
18:3*n*-3 α-Linolenic	Linseed, canola, flax and other seed oils	Precursor of EPA, DHA
20:4*n*-6 Arachidonic	Meats, dairy products	Precursor to bioactive eicosanoids; signal transduction; gene expression
20:5*n*-3 Eicosapentaenoic (EPA)	Marine fish oils	Competitive inhibitor of 20:4*n*-6
22:6*n*-3 Docosahexaenoic (DHA)	Marine fish oils, marine algae	Neuronal membrane structure; gene expression; prevention of apoptosis
18:3*n*-6 γ-Linolenic (GLA)	Evening primrose oil, borage seed oil	Precursor to 20:3*n*-6
20:3*n*-6 Dihomo-γ-linolenic (DHLA)	Milk (small amount)	Competitive inhibitor of 20:4*n*-6

Table 3 Etiology and characteristics of primary and secondary EFAD

Etiology	*Result*	*Effect*
Chronic malnutrition, especially in children	Lack of dietary fat and protein Decreased enzyme synthesis	Very abnormal serum FA patterns Low total serum PUFA
Long-term fat-free TPN	Lack of dietary EFA	Abnormal serum 20:3*n*-9/20*n*-6
Various fat malabsorption conditions	Secondary to pancreatic insufficiency, bowel cancer, or other serious GI diseases	Low total serum PUFA
Sjogren–Larsson syndrome	Genetic enzyme defect in desaturation and elongation	Low serum levels of C20 and 22 PUFA
Acrodermatitis enteropathica	Genetic defect in ε5 and ε6 desaturase activity	Very low levels of 22:6*n*-3
End-stage liver disease	Impairment in desaturase and elongase activities	Very low serum levels of LCPUFA

Table 4 Characteristics of EFAD in humans

Known clinical manifestations
- Dermatitis; dry, scaly skin; impetigo; eczema; generalized erythema
- Coarse, sparse hair
- Increased frequency of stools
- Decreased growth rate
- Cellular hyperproliferation in skin, alimentary tract, and urinary tract
- Possible immune impairment
- Slow wound repair

Suggested manifestations associated primarily with 22:6n-3 deficiency
- Slower early learning behavior (based on animal studies)
- Visual perturbations (based on animal studies)
- Mental retardation in human genetic defects involving desaturases and elongases

Known biochemical aberrations
- Serum 20:3*n*-9/20:4*n*-6 greater than 0.4
- Increased serum 16:1*n*-7, 18:1*n*-9, and 20:3*n*-9; decreased 18:2*n*-6, 20:3*n*-6, and 20:4*n*-6

in blood and liver. Status is now usually measured more accurately as the 20:3*n*-9/20:4*n*-6 ratio in serum phospholipids. A normal value is considered to be about 0.1–0.2. Occasionally, when abnormalities are due to an enzyme deficiency rather than to diet, some other index may be more useful. For example, the normal conversion of 18:2*n*-6 to 20:4*n*-6 results in a ratio of 18:2*n*-6/20:4*n*-6 of about 1.6, but in severe dietary EFAD, it may be 1.2. In achrodermatitis enteropathica, however, very little arachidonic acid is synthesized, and this ratio may be 10 or larger.

Serum FA patterns and symptoms that resemble EFAD occur in a variety of disorders. Marginal or frank EFAD is somewhat common in young cystic fibrosis (CF) patients, but whether this is a consequence of decreased fat absorption or a defect in FA metabolism, or both, is not known. While all serum FA levels are decreased in some CF patients, the fact that fewer desaturation products are present has led to the speculation that there is an enzymatic defect in FA metabolism. Oxidative stress disorders tend to show FA patterns characteristic of EFAD. The reason is unknown. Perhaps reactive oxygen species or products of oxidative damage act to decrease the activity of desaturases. Whereas high amounts of PUFA in the diet are said to increase the requirement for vitamin E, in studies on concomitant EFAD and vitamin E deficiency, measures of autooxidative susceptibilty in rat red blood cells indicated that EFAD actually potentiated the vitamin E deficiency. Patients with malabsorption syndromes or long-term depressed oral intake or TPN may exhaust stores of EFA and exhibit clinical signs of EFAD, especially dermatitis. Also, TPN formulations with very high ratios of *n*-6 to *n*-3 have resulted in symptoms of *n*-3 deficiency, including neuropathy and immune impairment in animals.

Nervous tissues in primates and other animals are rich in *n*-3 PUFAs and their products. There have been numerous reports of learning deficits and other effects on nervous tissue in animals deprived of all *n*-3 PUFAs. If 22:6*n*-3 is not added to the media of retinal photoreceptor cells in culture, they degenerate and die, presumably because of activation of an apoptotic pathway. But reports of effects in humans have been less convincing, largely because long-term controlled dietary experiments cannot be conducted. Episodic numbness, weakness, pain in the legs, and blurring of vision occurred in children on TPN formulas that did not contain *n*-3 PUFA. These effects were reversed after inclusion of *n*-3 FA in the formula. Researchers have noted that some of the physical symptoms of attention-deficit hyperactivity disorder are similar to those of EFAD and that many of the diagnosed children have decreased serum 20:4*n*-6 and 22:6*n*-3 levels. Conversion of 18:2*n*-6 and 18:3*n*-3 requires

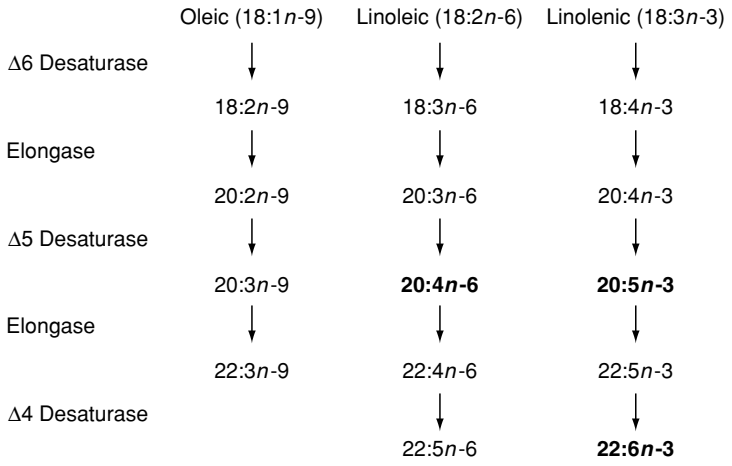

Figure 1 Metabolic pathways for fatty acid elongation and desaturation. Bold type designates major end products of the pathway.

desaturase and elongase enzymes, which may not be fully active in the preterm infant and neonate. Long-chain PUFAs (LCPUFAs), especially 22:6*n*-3, are particularly important for brain and nerve cell development *in utero*, but the fetus does not synthesize the long-chain EFA to any extent. Thus, maternal intake may be more important than has generally been recognized. Many investigators support the inclusion of *n*-3 PUFA, especially 22:6, in formulas for both premature and term infants. There remains some controversy over whether the inclusion is needed, since several studies have shown no detectable benefits for breast-fed infants whose mothers took supplemental dietary 22:6*n*-3. While there is some evidence of the efficacy of *n*-3 long-chain PUFA in increasing early visual maturation in preterm infants, there is no demonstrated long-term benefit. Six randomized trials with supplementation of infant formulas with LCPUFA were reviewed by Simmer, who concluded that there is little evidence to support the hypothesis that supplementation with 20:5*n*-3 and 22:6*n*-3 conferred any benefit on visual or cognitive development or influenced the growth of term infants.

Metabolism of PUFAs

In general, cells try to maintain a balance of FA of different chain lengths and degrees of desaturation to control membrane fluidity. PUFAs can be oxidized for energy, but the functional and structural roles of EFA depend largely on their metabolic products. All 20 carbon FA are referred to as eicosanoids, and 20:3*n*-6, 20:4*n*-6, and 20:5*n*-3 can all be used for prostaglandin synthesis. The metabolism of the *n*-9, *n*-6, and *n*-3 families can occur along the same pathways but normally does not, at least to the same extent (Figure 1). Furthermore, major products of the pathways may differ among cells, depending on enzyme activities. For example, the end product of 20:4*n*-6 in endothelial cells is prostacyclin (PGI_2), whereas the end product in platelets is thromboxane (TXA_2). The balance of LCPUFAs in the cell is regulated largely by induction or suppression of the relevant desaturases and elongases. In EFAD, there is increased desaturation of both 18:0 and 18:1. Negligible amounts of oleic acid (18:1*n*-9) are normally converted via the pathway in Figure 1, but in EFAD, cells try to produce long-chain PUFA needed for structure and functions by conversion of 18:1*n*-9 (oleic acid) to 20:3*n*-9. The ratio of 20:3*n*-9 to 20:4*n*-6 can be used as an index of deficiency.

In the presence of excess PUFA, the Δ9, Δ6, and Δ5 desaturases are suppressed, in part at the level of gene expression. Experiments with cultured endothelial cells have shown that added *n*-3 FA will be mostly elongated to C22 products, thus effectively removing competitors of 20:4*n*-6. But such experiments are lacking in many other types of cells. A. A. Spector has pointed out that even though cells can produce 22:5*n*-6, very little is normally produced possibly because most of the precursor, 20:4*n*-6, leaves the pathway for later prostaglandin synthesis. In *n*-3 PUFA deficiency, more 22:6*n*-6 is produced, but it does not fully replace the function of 22:6*n*-3. EFA desaturation and elongation occurs mainly in the liver and is then delivered to other cells. DHA (22:6*n*-3) is not synthesized in neurons, despite its high concentration there; however, it can be synthesized from 18:3*n*-3 and 20:5*n*-3 in the surrounding astrocytes. It is hypothesized that astrocytes synthesize DHA and release it as a trophic factor for neurons. EFA retroconversion, or shortening of the chain, can occur but its metabolic significance, if any, is not known.

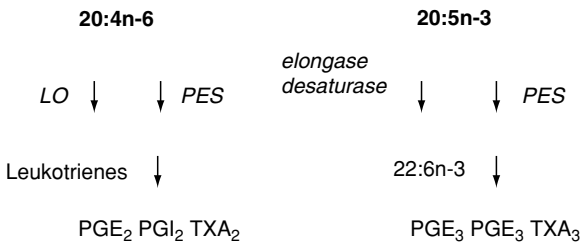

Figure 2 Summary of arachidonic and eicosapentaenoic acid metabolism.

Arachidonic acid undergoes metabolism to the series-2 prostaglandins (i.e., two double bonds) and to the leukotrienes of series-4 (four double bonds), as simplified in Figure 2. The oxidation of eicosaenoic acids (20:4*n*-6, 20:3*n*-6, and 20:5*n*-3) is catalyzed by the enzyme prostaglandin endoperoxide synthetase (PES), which has two separate activities usually referred to as cyclooxygenase (CO) and peroxidase. Figure 2 shows the difference in the products in the pathways that involve the same enzymes. Cyclooxygenase converts 20:3*n*-6 to PGG_1, 20:4*n*-6 to PGG_2, and 20:5*n*-3 to PGG_3. Subsequent conversion to hydroxy analogs occurs next via peroxidase activity. Metabolism of these endoperoxides to characteristic products depends largely on the cell type. For example, in arterial endothelial cells, PGG_2 is largely converted to PGI_2, prostacyclin, which causes vessel vasodilation and also inhibits platelet aggregation. In platelets, however, PGG_2 is converted to TXA_2 which is important to normal blood clotting, since it causes vasoconstriction and platelet aggregation. These metabolites must be continuously synthesized as they have a very short half-life. The various products tend to act in a 'balanced antagonistic' way; while TXA_2 is the most potent proaggregatory agent known, it is normally counteracted by prostacyclin. When the balance is altered, excess thromboxane leads to abnormal platelet adhesion and aggregation and therefore it may be implicated in coronary thrombosis. Although 20:3*n*-6 and 20:5*n*-3 can be used to form similar products that have one or three double bonds, respectively, the series-1 and series-3 prostaglandins do not have the same biological activity as the series-2. In addition, the precursors (18:3*n*-6 and 20:5*n*-3) compete with 20:4*n*-6 for cyclooxygenase action. This is why EPA (20:5*n*-3) in fish oil is said to be beneficial in preventing thrombosis.

Lipoxygenase (LO) catalysis of 20:4*n*-6 forms hydroxy fatty acids known as leukotrienes (e.g., LTA_4, LTB_4) These are compounds that have four double bonds, at least three of which are conjugated double bonds. Other eicosanoids can compete with 20:4*n*-6 for LO activity and thereby reduce the amounts of series-4 leukotrienes formed.

Functions of EFAs and their Metabolites

There are numerous known and some purported functions of PUFA. The major function of both α-linoleic acid (18:2*n*-6) and α-linolenic acid (18:3*n*-3) is as a precursor to other essential long-chain fatty acids, although 18:2*n*-6 itself may be important in membrane and lipoprotein structure and may have some cellular functions. The functions of the metabolites are very different, except that both 20:4*n*-6 and 22:2*n*-3 are components of membrane phospholipids and are involved in signal transduction. Also, both are known to modulate or regulate gene expression for several enzymes. All LCPUFAs suppress transcription of genes encoding several lipogenic enzymes including fatty acid synthetase and S14 protein and induce certain genes encoding peroxisomal and cytochrome P_{450} enzymes. The net result is that at a high PUFA, suppression of lipogenesis and triacylglycerol synthesis occurs. They regulate the expression and/or activity of the acyl desaturases and perhaps some of the enzymes of the glycolytic pathway. Otherwise, the effects of 20:4*n*-6 and 22:2*n*-3 are quite different. The various products of 20:4*n*-6 metabolism also have very diverse, sometimes antagonistic, functions. In addition, the competition of the other 20C*n*-6 and *n*-3 FA for the same enzymes makes it difficult to assign cause-and-effect functions. General functions are shown in Table 5.

Among the lesser known effects of EFA are the effects on lipoprotein metabolism. Normal EFA status is necessary for overall lipoprotein metabolism and especially for efficient high-density lipoprotein processing by the liver. In dietary deficiency with normal liver function, there is decreased incorporation of phosphatidylcholine into bile and decreased chylomicron production. Altered partitioning of available EFA has been noted in animals and increased hepatic conversion of linoleic to arachidonic acid occurs. Also, hepatic Δ6 and Δ5 desaturases are upregulated. More eicosanoid precursors are distributed into hepatic phosphatidylethanolamine than

Table 5 Some general functions of arachidonic acid, docosahexaenoic acid, or their metabolites

Membrane structure	Blood clotting
Signal transduction	Immune responses
Vasoconstriction	Vasodilation
Cell proliferation	Apoptosis
Modulation of gene expression	Lipoprotein synthesis and metabolism

usual. The *n*-3 PUFA decrease serum triacylglycerol through reduced very low density lipoprotein production. The precise biochemical nature of these effects has not been established. Another less known effect of EFAs is that they enhance the effects of vitamin D in increasing calcium absorption and help decrease urinary excretion of calcium. EFA-deficient animals develop severe osteoporosis and ectopic calcification.

The effects of PUFAs on cell proliferation, tumors, and cancers are highly variable, depending on the cell type and proportions of FA. They may be modulators of carcinogenesis with distinct tissue-specific pro- or anticancer effects. LCPUFAs activate the peroxisomal proliferation activator receptor, which may mediate their purported effects on gene expression, cellular growth and differentiation, and apoptosis. Both 18:2*n*-6 and 20:4*n*-6 are said to inhibit mitogen-stimulated proliferation of lymphocytes in culture, but they may have no effect or a stimulatory effect in other cell types. In EFAD in animals, diminished immunity occurs because 20:4*n*-6 metabolites are needed for normal immune responses. However, the *n*-3 PUFAs suppress immune response and may be of benefit in inflammatory and autoimmune disorders. As mentioned above, the prostaglandin and thromboxane derivatives of 20:4*n*-6 are bioactive molecules that act as paracrine hormones and have many well-known, important functions on blood clotting, smooth muscle contractions, and immune responses. When released from phospholipids, 20:4*n*-6 activates protein kinase C and therefore modulates the activation of growth factor receptors such as EGH receptors. It is thought to have other cell-signaling functions as it can inhibit GTPase activated protein.

The *n*-3 PUFAs are said to inhibit endothelial activation. This may occur in at least two ways; 20:5*n*-3 competes with 20:4*n*-6 for conversion to the prostaglandin, which normally stimulates this activity, and 22:6*n*-3 decreases expression of endothelial leukocyte adhesion molecules. Usually, prostacyclin-mediated cytokine stimulation of expression of adhesion molecules is a normal physiologic protective function, but in pathological states, this leads to an undesirable sequence of events that may result in atheroma formation. These *n*-3 PUFAs may help confer a protective effect against atherogenesis. Also, *n*-3 FA promote an antiarrhythmic action, apparently owing to a slight hyperpolarization of heart muscle membranes that prevents 'easy' induction of action potentials.

Leukotriene metabolites have profound effects on white blood cell chemotaxis and phagocytosis, regulation of neutrophil and eosinophil function, smooth muscle contraction and bronchoconstriction in the pulmonary system, and changes in capillary permeability. In neutrophils, 20:4*n*-6 is metabolized primarily by the 5-LO pathway leading to the formation of LTA_4, which is a precursor for a potent chemotactic factor. Thus, a decrease in 20:4*n*-6 concentration can result in an antiinflammatory response. Since 20:5*n*-3 and 20:3*n*-6 compete with 20:4*n*-6 for the LO pathway, the result is similar because less conversion to LTA_4 occurs. Although the prostanoids and leukotrienes formed from 20:4*n*-6 have many extremely important physiologic functions, overproduction of some of these compounds appears to be involved in certain disease processes or states, including thrombosis, myocardial infarction, cardiac arrythmia, arthritis, asthma, inflammatory conditions, and autoimmune disorders. Reduction of tissue levels of 20:4*n*-6, competition for enzymes by other FA, or inhibition of the enzymes by drugs may in some cases ameliorate some of the symptoms associated with the disorder(s). For example, aspirin, ibuprofen, and several other drugs inhibit the activity of cyclooxygenase, and steroidal antiinflammatory compounds such as hydrocortisone inhibit the enzyme phospholipase A_2 that releases 20:4*n*-6 from membrane phospholipids. All of these result in less prostaglandin synthesis. Inhibition of phospholipase A_2 also reduces leukotriene synthesis.

DHA (22:6*n*-3) is the major PUFA in neuronal membranes; it is especially enriched in PS. PS is important structurally and functions in translocation of specific proteins, such as Raf-1 and protein kinase C, to the plasma membrane. Also, it is hypothesized to be involved in cellular apoptosis. Raf-1 translocation is a first step in the transduction of growth factor signaling. Depletion of 22:6*n*-3 leads to a decrease in the synthesis of PS and lower levels of PS in brain cell membranes. PS levels are much higher in nerve tissues than in other tissues, and *n*-3 deficiency results in a dramatic reduction of PS only in nerve cells. Thus, *n*-3 deficiency seems to have profound effects on PS-related signaling events in the nervous system. In animals, a decrease in 22:6*n*-3 may be associated with compensatory increased levels of 22:5*n*-6 leading to: modification of neural membrane composition and functional changes in enzyme activities, suboptimal retinal and brain development, and a poorer performance in learning tasks. It may also be correlated with peroxisomal disorders associated with neuronal deterioration. Brain concentrations of 22:6*n*-3 plateau with 18:3*n*-3 intake of about 0.7% energy, but this is influenced by the 18:2*n*-6 intake. Several studies have shown adverse effects on growth and function of animals with high intakes of 22:6*n*-3.

The amount, type, and balance of dietary FA along with antioxidant nutrients impact the immune system. The amelioration of symptoms of autoimmune diseases in some studies with high doses of fish oils

was attributed to effects of 20:5*n*-3 suppression of T-lymphocyte proliferation, apoptosis of autoreactive lymphocytes, and decreased proinflammatory cytokine production. Long-term effects of such treatment on host immunity require further study. T-cell proliferation was shown to be decreased after administration of 2.4 g per day 18:3*n*-6 (γ-linolenic) in borage seed oil, but not with 18:3*n*-3 or 18:2*n*-6. The effect therefore appears to be a competitive effect of 20:5*n*-3 and 20:3*n*-6 with 20:4*n*-6. Dietary 20:5*n*-3 and 18:3*n*-6 also appear to inhibit arachidonic induced production of thromboxane (TBB_2) by platelets.

Recent Advances

In the past decade, much research has focused on fish oil supplementation, with high levels of 20:5*n*-3 and 22:6*n*-3. The fatty acid content of menhaden oil is shown in Table 6, and possible benefits and concerns of supplementation in Table 7. Three areas have been of special interest: effects on immune responses, coronary heart disease, and behavior. In all of these areas, there have been inconsistencies in the reported results, some of which may be due to different diets and conditions.

Some researchers have suggested a role for EPA as well as DHA in normal nerve cell membrane structure. This is based largely on reports of lower than normal levels of EFA in some patients with schizophrenia and the moderately positive effect observed with EPA supplementation. However, since EPA can undergo further metabolism to DHA, it is unclear whether EPA has anything other than a precursor role in these cells. There are trials in progress on the use of EPA in the treatment of schizophrenia.

Rats with decreased brain levels of 22:6*n*-3 show an impaired performance in learning tasks compared to rats fed 22:6*n*-3 and increasing the amount of 22:6*n*-3 to a 10:1 ratio *n*-6/*n*-3 in infant formula of primates was associated with greater accretion of 22:6*n*-3 in neonatal brain and retina. Patients with moderately severe dementia of thrombotic cardiovascular disorder and Alzheimer's dementia supplemented with fish oil were followed for one year and showed improved scores on a psychometric test as well as better red blood cell (RBC)-deformation scores. It was unclear whether the improvement was a direct effect on neuronal cells or due to improved RBC function. Low-serum 22:6*n*-3, along with the associated decrease in phosphatidylcholine, is now considered to be a significant risk factor for Alzheimer's dementia. Of concern is the reduction in Δ6 desaturase activity that occurs with aging and possible lack of adequate 22:6*n*-3 synthesis, even with adequate intake of 18:3*n*-3. This FA comprises more

Table 6 Composition of menhaden oil

Polyunsaturated fatty acid	*Grams per 100 g of oil*
18:2*n*-6	2.0
18:3*n*-3	1.4
20:4*n*-6	1.1
20:5*n*-3	13.0
22:5*n*-3	5.0
22:6*n*-3	8.5

Table 7 Possible beneficial effects vs. safety concerns related to dietary *n*-3 PUFA

Possible benefits of supplementation in selected individuals
Decrease in serum triacylglycerols
Decrease in myocardial infarctions in persons with diagnosed CHD
Decreased inflammation and relief of certain conditions, as rheumatoid arthritis, psoriasis, lupus and inflammatory bowel syndrome
Decreased cardiac arrhythmias (animal studies, not very definitive in human studies)
Possible safety concerns related to excess intake (based on animal or human studies)
Increased bleeding times and associated risk of hemorrhagic stroke
Increased peroxidation of fatty acids
Inhibition of immune cell function and decreased resistance to infection
No change or increase in serum LDL-cholesterol
Imbalance of PUFAs (very low *n*-6/*n*-3) with resultant behavioral abnormalities
Concerns related to interactions with drugs
Interactions with drugs that affect eicosanoid metabolism and clotting mechanisms
Interactions with drugs that affect eicosanoid metabolism and immune responses
Potentiation of cytotoxicity of some anticancer drugs
Possible cases for 'special' supplementation
Chronic liver disease – may require supplementation of 20:4*n*-6 and 22:6*n*-3
Cystic fibrosis – may require supplementation of 20:4*n*-6 and 22:6*n*-3 in addition to 18:2
Inherited desaturase and/or elongase deficiencies, as in some peroxisomal disorders
Subnormal 20:6*n*-3 serum levels associated with certain types of depression, dementia, attention-deficit disorder, adrenoleukodystrophy, long-chain hydroxyacyl-CoA dehydrogenase deficiency, and dyslexia

than 30% of the structural lipid of neurons and is very enriched in retinal rod outer segments. Subnormal levels are associated with neurological conditions, including depression, attention deficit hyperactivity disorder, and dementia. Dietary intervention to increase plasma 22:6*n*-3 levels has been associated with improvements in visual or neurological deficits in adrenoleukodystrophy, long-chain hydroxyacyl-CoA dehydrogenase deficiency, dyslexia, and Alzheimer's dementia. Intervention trials are now underway to determine the effects of dietary 22:6*n*-3 from algal sources on development of Alzheimer's dementia. This source, which contains much less 20:5*n*-3, will be evaluated for safety and efficacy in hopes of reducing the side-effects of high levels of 20:5*n*-3, including immune suppression and bleeding tendencies.

Diets rich in *n*-3 PUFAs have been associated with suppression of cell-mediated immune response. Although this accounts for the reported beneficial effects of fish oil in arthritis, suppression of immune response is of great concern in some individuals. Fish-oil diets suppress lymphocyte proliferation, monocyte and neutrophil chemotaxis, and tumor necrosis factor, interleukin-2, and interleukin-6 production, and also reduce expression of adhesion molecules on the surface of endothelial cells, monocytes, and lymphocytes. This is the basis of its antiinflammatory effect. LTB_4, which is a potent chemoattractant for monocytes, important in regulation of inflammatory response and cytokine synthesis, declines in the presence of excess 20:5*n*-3. Also, PGE_2, which normally increases during infections and acute inflammation, declines, and the series-3 prostaglandins and series-5 leukotrienes, which have a less potent effect, increase. These effects are attributed to 20:5*n*-3, rather than 22:6*n*-3.

Because of anecdotal and epidemiological evidence suggesting that FO had a positive influence on decreasing the incidence of CHD, in 1993, the US Food and Drug Administration (FDA) conducted a study to determine whether to allow a health claim for *n*-3 fatty acids. At the time, the FDA concluded that findings were inconsistent in the general population, and there was insufficient evidence to support a health claim for *n*-3 FA and CHD. Some suggested an adverse effect. A part of the problem with the early epidemiological and intervention studies is that plasma total cholesterol or low-density lipoprotein cholesterol (LDL-C) was used as the biomarker for reduced risk of CHD. Almost all studies with fish oil showed no significant decline in LDL-C, and some showed an increase. More recently, the FDA again considered whether to allow the health claim, and in this review, different endpoints and clinical measures of the risk of CHD were evaluated. These included the endpoints of fatal or nonfatal myocardial infarction and clinical measures of reduction in fasting and postprandial serum triacylglycerol levels, reduction in platelet aggregation and adhesion, and changes in the composition of serum lipoproteins. In the mid- to late 1990s, all intervention trials of persons with diagnosed CHD supplemented with fish oil or *n*-3 derivatives showed no decrease in LDL-C but decreases in the risk of CHD using the mentioned endpoints. After reviewing the relevant data on beneficial effects on the risk of CHD, adverse effects of any type and the intakes associated with these effects, the FDA found that the use of 20:5*n*-3 and 22:6*n*-3 as dietary supplements is safe, provided the intake does not exceed 3 g per day from all food and supplement sources.

See also: **Fatty Acids**: Properties; Metabolism; Gamma-linolenic Acid; Analysis; Dietary Importance; *Trans*-fatty Acids: Health Effects; **Fish Oils**: Composition and Properties; Dietary Importance

Further Reading

Harbige LS (1998) Dietary *n*-6 and *n*-3 fatty acids in immunity and autoimmune diseases. *Proceedings of the Nutrition Society* 57: 555–562.

Holman RT (1998) The slow discovery of the importance of omega-3 essential fatty acids in human health. *Journal of Nutrition* 128: 4275–4335.

Lagarde M, Spector AA, Galli C, Hamazaki T and Knapp HR (eds) (1999) Fatty acids and lipids from cell biology to human disease; Proceedings of the 3rd Congress of ISSFAL. *Lipids* 34: S1–S350.

Lewis CJ (2000) *Letter Regarding Dietary Supplement Health Claim for Omega-3 Fatty Acids and Coronary Heart Disease*. US Food and Drug Administration. Docket # 91N-0103 http://vm.cfsan.fda.gov/~dms/ds-ltr11.htm

Perkins EG and Visek WJ (eds) (1983) *Dietary Fats and Health*. Champaign, IL: AOCS Press.

Spector AA (1999) Essentiality of fatty acids. *Lipids* 34: 51–53.

www.nal.usda.gov/fnic/foodcomp

ESSENTIAL OILS

Contents

Properties and Uses

J Grassmann and E F Elstner, Technical University of Munich, Munich, Germany

Background

Essential oils have been used in folk medicine for a very long time, but their mode of action in healing was not clear. Recently, these oils have become increasingly important, such as in aromatherapy, which is controversial, but also in traditional medicine and nutrition. For this reason, an increasing number of researchers are becoming interested in the biochemical effects of these oils.

This article gives an overview of the definition, abundance, components, biosynthesis, and biological functions of essential oils in the plant kingdom, but the main aim is to describe the use and biochemical effects as well as the pharmacological properties of these oils. A short chapter deals with the stability and, in this context, with the problem of aging of essential oils.

Essential Oils: Definition and Abundancies

Essential oils (EOs) are complex mixtures of more or less lipophilic substances obtained by distillation or pressing of plants or parts of them. Most of the components are terpenoids (isoprenoids), mostly mono- and sesquiterpenes. Furthermore, this main group is followed by aromatic compounds (phenylpropanoids) derived from the shikimate–chorismate pathway, polyketide derivatives, paraffinic hydrocarbons, alcohols and ketones, fatty acids, and finally sulfur- and nitrogen-containing compounds. The compartments of highest concentration of EOs in the plant are special oil cells, which are eventually converted into oil containers. They are also found intercellularly, with the oil compartments being divided into schizogenic and lysigenic oil containers. Furthermore, milk tubes and plant-surface gland hairs may also contain EOs. The plant families containing EOs include the Apiaceae, Asteraceae, Cupressaceae, Lamiaceae, Myrtaceae, Pinaceae, Poaceae, and Rutaceae.

Isoprenes and monoterpenes play crucial roles in atmospheric chemistry (tropospheric ozone formation) where annual global fluxes between 250 and 450 million tonnes are estimated to arise from the vegetation, mainly from trees. This important function in the atmosphere is due to their fast reaction with reactive oxygen species (ROS) such as the OH radical (see also section on Antioxidative Effects). Terpenoids also seem to play an important ecological role in plant–herbivore interactions, since their biosynthesis is induced by insect feeding, eventually triggered by an elicitor from oral secretion of the feeding insect.

Technical Procedures: Extraction from Plants or Plant Organs

The technical procedures for obtaining EOs from plant material comprise steam distillation, extraction and, as mostly in the case of citrus oil, pressing of peels. These procedures are dealt with in a special chapter of this encyclopedia. (*See* **Essential Oils**: Isolation and Production.)

Physical Properties, Chemical Structures, and Reactivities

Fundamental work on EOs was started as early as 1884, by O. Wallach (1847–1941). EOs are in a liquid state at ambient temperatures, but they are volatile (with boiling temperatures between 50 and 320 °C), thus distinguishing them from mineral or fatty oils. Because of their volatility, EOs have a characteristic odor. They produce a transparent spot, on adsorbent paper, that disappears within 24 h. Most of the EOs are almost colorless, have a high refractive index, and are optically active (see below). They are soluble in most organic solvents, but have a very limited solubility in water. Their density is lower than that of water, with the exception of the oils of sassafras, clove, and cinnamon. An important feature is their property of being steam-distillable, thus

enabling physical separation from other lipophilic plant constituents such as 'fixed oils' or membraneous components. In the following the most important groups of compounds of essential oils will be listed.

Most Important Components

Today, several thousand different components are known, by far the most prominent being the terpenoids, followed by phenylpropanoids and other less important ingredients of plant oil compartments.

-Monoterpenes The monoterpenes (MTs) are widely distributed and sometimes contribute over 90% of the total oil content. They can be divided into regular and irregular monoterpenes and iridoids. Most components of essential oils have regular structures. Irregular members such as chrysanthemic acid and its derivatives are found in EOs of Asteraceae. Other examples of irregular MTs are the artemisanes, santolinanes, and lavandulanes or the artemisiaketon, which is present in *Achillea millefolium*.

Iridoids are another group of the MT family, in which the bicyclic ingredients of the Valerianaceae-like baldrinal, valeranone, or valtrate compounds from the olive tree (oleacin, oleuropein) or from yellow gentian (gentisin, sweroside, gentiopicrin) are the best known.

Since the 'regular' MTs are the most common components in essential oils, this chapter will deal only with this group. They can be subdivided into MT hydrocarbons, MT alcohols and their esters, MT aldehydes, MT ketones, and MT ethers such as cineol or dill ether. A special MT is the endoperoxide ascaridol.

MT hydrocarbons This group comprises several important compounds present in various essential oil compartments of different plant groups and include well-known compounds such as pinene, limonene, and terpinene. These compounds are frequently found as isomers, e.g., α- and γ-terpinene, as optical isomers, for example (+)- and (−)-limonene or both, for example (−)-α-phellandrene and (−)-β-phellandrene or (+)- and (−)-α-pinene together with (−)-β-pinene.

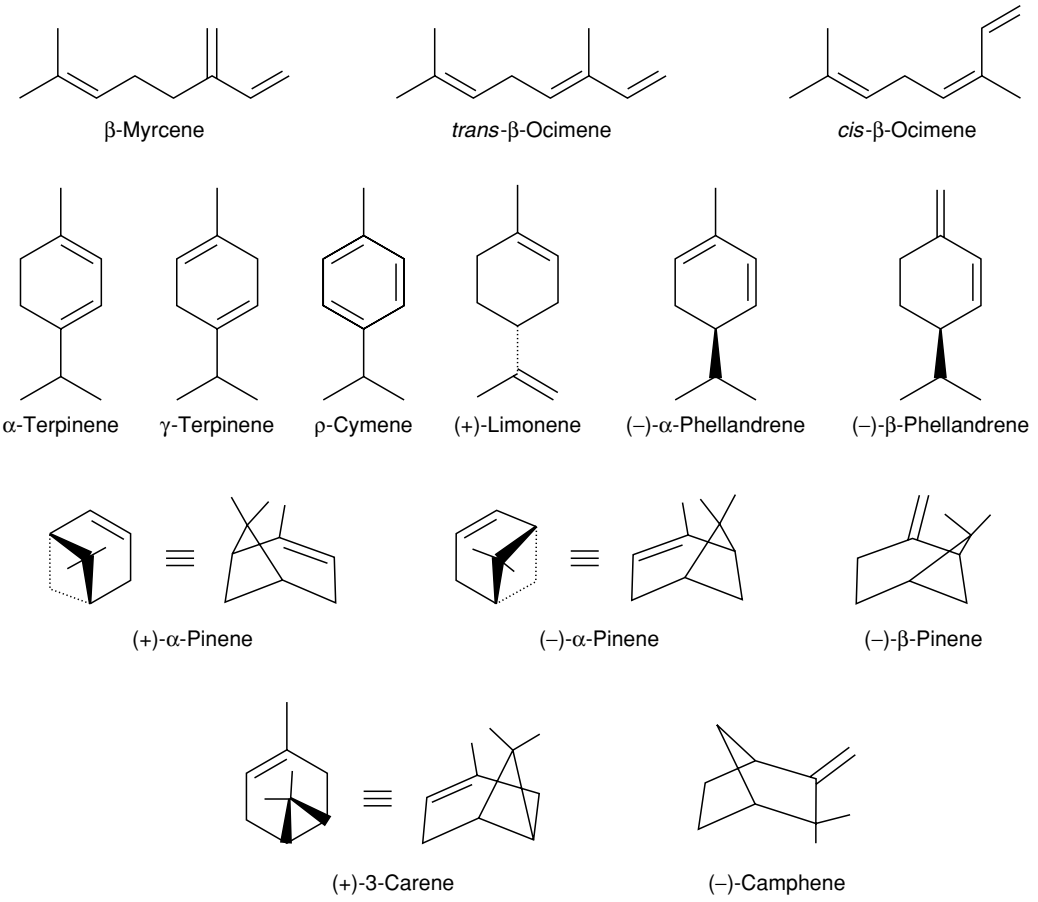

Figure 1 Examples of acyclic and cyclic structures of regular monoterpene hydrocarbons in essential oils.

A selection of important structures in this class is shown in Figure 1.

MT alcohols Important members of this class are the widely distributed compounds geraniol, nerol, linalool, citronellol, terpineol, and borneol. They also comprise both acyclic and cyclic forms. Several important structures are shown in Figure 2.

MT aldehydes and MT ketones Again, both acyclic compounds (geranial, neral, citronellal) and cyclic forms (carvone, menthon, pulegone) are found. A mixture of the isomers geranial and neral is called citral. Bicyclic MT ketones are the well-known (+)- and (−)- camphor, fenchone, and thujone (Figure 3).

MT ethers and endoperoxides Menthofuran, dill ether, cineol, and ascaridol are representatives of this class of compounds and are shown in Figure 4.

-Sesquiterpenes With more than 50 different structural types, sesquiterpenes constitute another important group of compounds within the EOs. They arise from farnesylpyrophosphate, which in turn is formed by head-to-tail condensation of geranylpyrophosphate and isopentenylpyrophosphate, forming a C15 carbon skeleton (see below). They are especially abundant in essential oils of the plant family of the Zingiberaceae and sesquiterpene lactones in the Asteraceae. β-Carophyllen-1, 2-epoxide is found in

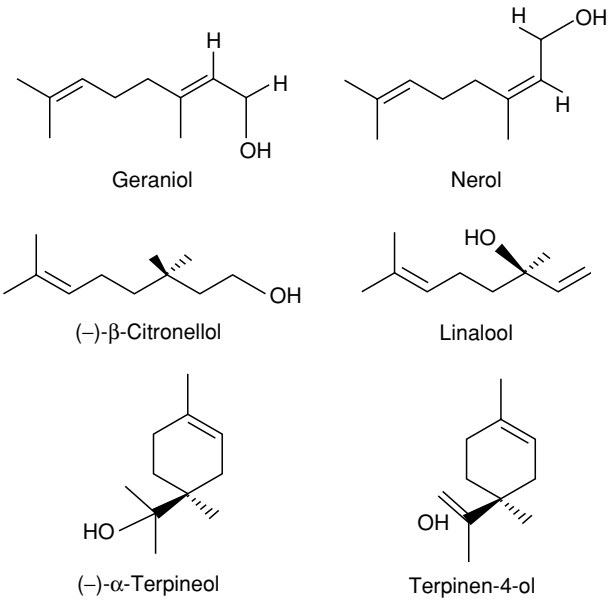

Figure 2 Examples of acyclic and cyclic structures of monoterpene alcohols in essential oils.

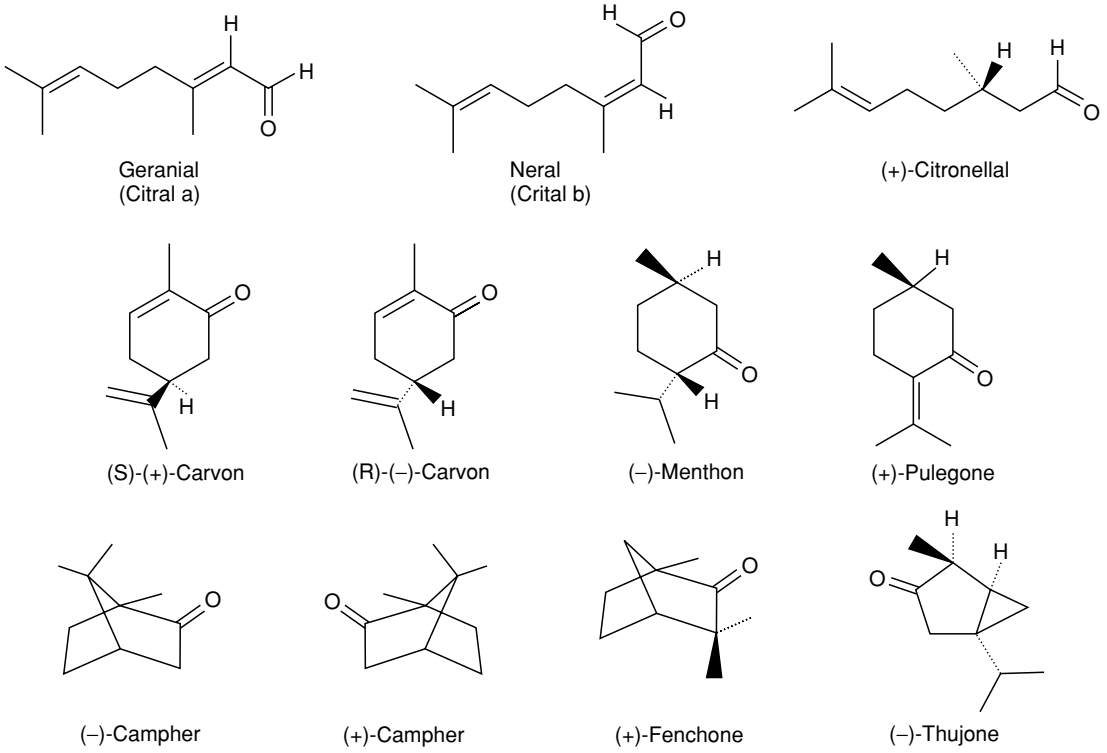

Figure 3 Examples of acyclic and cyclic structures of monoterpene aldehydes and ketones in essential oils.

clove and sage oil, in which it can be used for quality control (Figure 5).

-Aromatic compounds of the phenylpropane group Aromatic compounds (ACs) are minor components in EOs. They essentially comprise phenylpropanoids, propenyl- or allylphenols and coumarins. Typical members from this group, like anethol or estragol, are found in Apiacean oils (fennel, anise) but also in Rutacea. Eugenol and vanillin are two typical ACs of this type (Figure 6).

-Other compounds EOs also may contain degradation products of membraneous unsaturated fatty acids such as 'green odor' components (*cis*- or *trans*-hexenal, hexanal, and hexanol) and several lactones. Furthermore, compounds from terpene degradation, for example C_{13}-norisoprenoids, and sulfur- and nitrogen-containing compounds (e.g., pyridine derivatives in spearmint oils) are found. Several volatile terpenoids are present in plants as glycosides and contribute to the overall, typical fruit flavors.

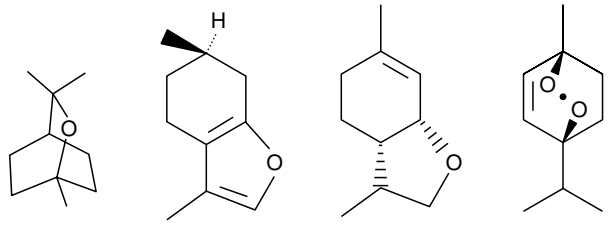

Figure 4 Examples of structures of monoterpene ethers and endoperoxides in essential oils.

Biosynthetic Pathways

The two cell compartments of plant isoprenoid biosynthesis, plastid and cytosol, have two different ways to synthesize terpenoids: the classic mevalonate pathway and the alternative 1-deoxy-D-xylulose-5-phosphate (DOXP) pathway, which occurs in the plastids and leads to, among others carotenoids,

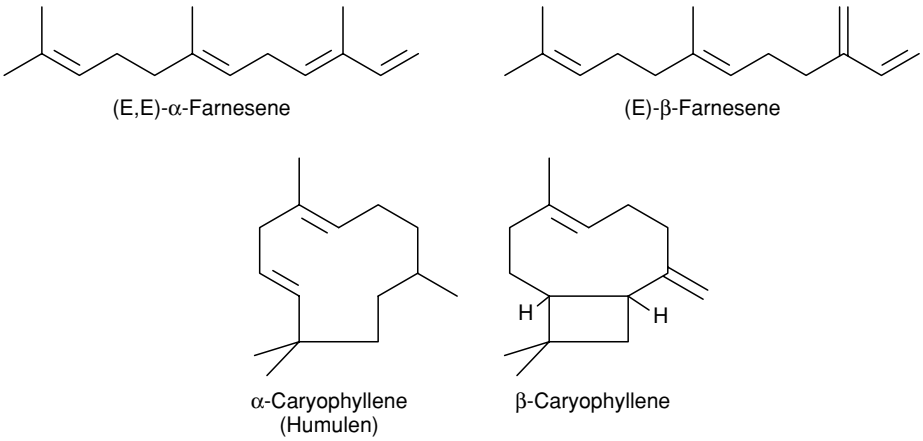

Figure 5 Examples of structures of sesquiterpenes in essential oils.

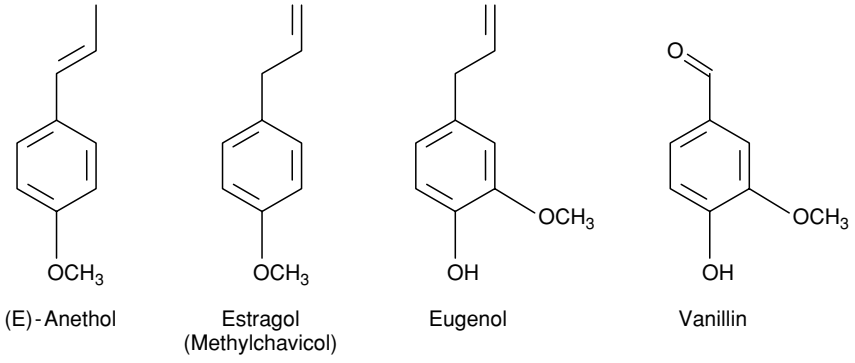

Figure 6 Examples of structures of aromatic compounds of the phenylpropane group.

monoterpenes, and isoprene. In the cytoplasm, the mevalonate pathway takes place and forms sesquiterpenes, for example.

The Mevalonate Pathway

The biosynthesis of monoterpenes in the cytoplasmatic space proceeds via the well-known mevalonic acid pathway, starting from acetyl CoA by various condensations including activation by ATP and reduction by NADH. A key step is the two-step reduction of hydroxymethyl-glutaryl-CoA (HMG CoA) into mevalonate under CoA release by the enzyme HMG CoA reductase. This step has become extremely important, since it represents the main possibility of interaction with several types of inhibitors with this pathway and is used for the treatment of hypercholesterolemia in humans. After the condensation of two C5 bodies (activated isoprenes: isopentyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), with pyrophosphate as a thermodynamically preferred leaving group (splitting into two orthophosphates, E_0' approx. $-35\,\mathrm{kJ\,mol^{-1}}$), the product is geranylpyrophosphate (GPP), the first C10 backbone of the monoterpenes. The formation of acyclic monoterpenes from GPP is obvious, and the generation of monocyclic and bicyclic monoterpenes involves monoterpene cyclases. These enzymes catalyze an isomerization–cyclization mechanism generating the α-terpinyl cation. Loss of a proton leads to the formation of hydrocarbons and addition of water to alcohols. Generation of oxygenated terpenoids also can be carried out by classic oxidases that catalyze allylic hydroxylation, oxidation of the alcohol to an α, β-unsaturated ketone, and possibly reduction of the double bond. Thus, the α-terpinyl cation is the origin for a large number of different monoterpenes.

The precursor of sesquiterpenes, farnesyl pyrophosphate, is the product of the addition of one molecule IPP to GPP. Similar to monoterpene cyclizations, rearrangements, isomerizations, or oxidations lead to a wealth of different structures.

The DOXP Pathway of IPP Biosynthesis

Recently, another pathway, starting not from acetyl-CoA but from pyruvate and glyceraldehyde-3-phosphate (GAP), was established for plastidic-derived terpene derivatives in the middle of the 1990s, especially by the Lichtenthaler group in Germany. This pathway, in contrast to the above mevalonate pathway, is not inhibited by HMG CoA-reductase inhibitors. It includes, after a thiamin-dependent condensation of pyruvate and GAP (CO_2-release), DOXP. An intramolecular C-C-skeleton rearrangement catalyzed by DOXP reducto-isomerase generates 2-*C*-methyl-D-erythrose-phosphate (MEP). In 2000, it was shown that the following steps involve a cytidylyl transfer (catalyzed by MEP cytidylyl-transferase) generating 4-(cytidine 5′-diphospho)-2-*C*-methyl-D-erythritol (CDP-ME), which is phosphorylated by CDP-ME kinase, leading to 2-phospho-4-(cytidine 5′-diphospho)-2-*C*-methyl-D-erythritol. This compound finally undergoes cyclization to form 2-*C*-methyl-D-erythritol-2, 4-cyclodiphosphate (MECDP), a reaction catalyzed by MECDP synthase. MECDP is then converted into DMAPP, but the reactions leading to this product are still unknown.

Biological Functions in the Plant Kingdom

There are good reasons to assume ecological functions for most of EOs such as acting as allelopathic agents, repellants, or attractants in plant–plant or plant–pathogen/herbivore interactions. Observations in the Californian Chapparal and corresponding experimental approaches have made clear that certain evergreen shrubs such as *Salvia leucophylla* or *Artemisia leucophylla* retain a 1–2 m plant-free area in the grassland around them by inhibiting the growth of annual species such as *Avena fatua*, *Bromus rigidus*, *B. mollis*, *B. rubens*, *Festuca megalora*, or *Erodium circutarium*. As reason for this phenomenon volatile terpenes present in the epidermal glands of these shrubs such as cineol, camphor, α- and β-pinene, camphene, and thujone have been identified acting as germination inhibitors. The terpenoids are emitted during hot periods and condense in the soil surface around the shrubs.

Another function is seen in defense and wound sealing in pine tree species where a coordinated induction of the biosynthesis of MTs and resin acids is observed. MTs are thought to function as a solvent and vehicle for resin acids whereby the MTs evaporate from the wound region, and resin acids undergo oxidative polymerization and crystallization in this region, closing the wound and thus preventing access to pathogens.

Furthermore, volatile MTs released in response to herbivore activity or insect feeding are 'designed' as deterrents or antibiotics by plants, but they may act as a guide for predators or host-seeking parasites, as shown for parasitoid wasps, which find their egg host, a herbivorous caterpillar: they inject their eggs into the herbivore and thus propagate their life cycle at the expense of their egg hosts. Plants in extreme environments often have to tolerate extreme temperatures. The direct correlation between isoprene concentration and thermotolerance has led to the speculation that these compounds increase

thermotolerance by interacting with membranes of the plants. Such an interaction is also assumed for physiological and pharmacological functions in man (see below). In horticulture, MTs already may have or receive some practical function, acting as flowering promoters or as nematicides.

Uses and Effects

Use of EOs

The use of EOs is mainly restricted to the perfume and flavoring industries; above all, citrus oils are used in perfumery and to flavor soft drinks or sweets. Recently, aromatherapy has become increasingly important, with the treatment of functional disorders or even diseases by using a pleasant odor, which is often achieved by evaporating EOs. This kind of therapy, however, is controversial, mainly because most studies are questionable, but the use of a pleasant smell can help a person to relax and may help them to feel better. In pharmacy, oils are mostly applied as combined preparations like bath salts, embrocations, and inhalants; another application is external antiseptics. Examples of oils used for pharmaceutical use include eucalyptus oil (and its main component, cineol), peppermint oil, and menthol. Peppermint oil is also often used to alleviate headaches by rubbing the oil into the temples, and its effect may be due to its cooling effect on the skin. In folk medicine, there are various applications for essential oils and their components, as shown in Table 1 (see also section on Pharmacological Properties).

Some terpenoids like pinenes from turpentine, geraniol, (+)-citronellal, or eugenol are also used as a starting material for the synthesis of other substances. The applications described indicate that the most commonly used essential oils are citrus oils like sweet orange, lemon, or grapefruit oil, followed by peppermint and eucalyptus oil.

Biochemical Effects

EOs are mixtures of many compounds belonging to several biochemical classes (see above). Thus, a distinction has to be made between the properties of single isolated compounds and EOs generally, i.e., oil mixtures. To make things even more complicated, although an isolated compound may exhibit toxic properties if tested on its own, it may prove inactive in its 'natural environment' as an oil mixture.

Table 1 Pharmacological properties of some essential oils

Action	*Essential oil*
Expectorant	Anise, fennel, eucalyptus, pine needle
Spasmolytic	Anise, fennel, lavender, caraway, peppermint
Antiinflammatory	Sage, camomile

Uptake, Metabolism, and Elimination in Man

Essential oils and their components are readily absorbed via the respiration tract in the case of inhalation; resorption through skin and uptake in the stomach after oral application are less effective but still possible. After being absorbed, the terpenoids are readily distributed throughout the body.

The terpenoids are metabolized in the liver, e.g., by hydroxylation or glucuronidation, and then eliminated through the kidneys. Small quantities reach the bile or mother's milk unchanged, so, during breastfeeding, the baby may take up terpenoids. Some terpenoids, such as cineole and menthol, are also eliminated unchanged through the respiration tract or skin. Thus, even if they are taken orally, these terpenoids can exert their effects in the bronchial tract, which explains their preferential use in diseases of the respiratory tract, e.g., bronchitis.

There are very few studies on the pharmakokinetics of essential oils or their components, so it is not clear which concentrations they really can reach in blood or plasma. Some studies have examined plasma concentrations of α-pinene, limonene, and/or cineol, and have reported values of 5–20 μM, but mostly, such studies have been carried out with too few participants. Since, as already mentioned, the components of essential oils are very lipophilic, they can be enriched in adipose tissues or membranes and therefore reach much higher concentrations in those environments. If they are distributed in biological membranes, they can exhibit both stabilizing and destabilizing effects.

Pharmacological Properties

EOs exhibit antiseptic, sedative, spasmolytic, and irritating properties. Some oils have local anesthetic and antiinflammatory effects, and recently, anticarcinogenic and serum-cholesterol-lowering activities of terpenoids have been discussed. Some terpenoids are also capable of inhibiting HIV.

The antibiotic and antiseptic properties are against bacteria, fungi, and yeast. The most antiseptic oils are savory, cinnamon, thyme, clove, lavender, and eucalyptus. Thymol, a compound found in thyme oil, for example, is 20 times more antiseptic than phenol. It has been shown that the different enantiomers of limonene and α-pinene, respectively, exhibit different antibacterial and antifungal activities. This result draws attention to the fact that receptors may play a

role in the antiseptic properties of essential oils. Another possible mechanism of antimicrobial activities involves membrane disruption by lipophilic compounds, and it has been shown that reducing the lipophilicity by the introduction of a hydroxyl group reduces the antimicrobial activity of certain diterpenoids.

In vitro, many oils like camomile, clove, mint, and thyme exhibit spasmolytic activity. Some essential-oil drugs like mint or European vervain have a positive effect on gastrointestinal spasms; they are also capable of stimulating gastric secretion and are therefore known as digestives. The oils of fennel, dill, and caraway have positive effects on stomach upset and flatulence.

EOs coming into contact with the skin and mucous membranes may lead to irritations such as reddening and a sensation of hot or cold. These irritating properties are probably due to an increase in capillary blood flow or to an effect of terpenoids on the calcium channels, as shown by the cooling effect of peppermint oil. These effects are also the basis for the use of EOs in ointments, embrocations, and gels.

Eucalyptus oil is thought to increase the motility of the ciliated epithelia in the bronchia, thus explaining the expectorant effect of this oil.

Some terpenoids, such as limonene, geraniol, menthol, and carvone, have been shown to possess anticancer activity, increasing tumor latency and decreasing tumor growth. These effects may result from the ability of the terpenoids to inhibit the activity of HMG-CoA-reductase. This is probably also the reason for the serum-cholesterol-lowering capability of some terpenoids.

The anti-HIV activity can be attributed to different mechanisms; the terpenoid carnosolic acid, for example, has been shown to inhibit HIV-1 protease, whereas the terpenoid nigranoic acid inhibits the reverse transcriptase.

Toxicity

The first point to mention in this section is that one should not confuse essential oil plants and essential oils: even if the plant is harmless, the essential oil derived from it may be toxic. Fortunately, acute toxicity by oral intake is generally very low with an LD50 of 2–5 $g\,kg^{-1}$ or above.

There are more toxic oils such as boldo (LD50 = 0.13 $g\,kg^{-1}$), chenopodium (LD50 = 0.25 $g\,kg^{-1}$), thuja (LD50 = 0.8 $g\,kg^{-1}$), pennyroyal (LD50 = 0.4 $g\,kg^{-1}$), and mustard oil (LD50 = 0.34 $g\,kg^{-1}$) and their components, such as thujone (LD50 = 0.2 $g\,kg^{-1}$), carvone (LD50 = 0.16 $g\,kg^{-1}$), carvacrole (LD50 = 0.8 $g\,kg^{-1}$) and pulegone (LD50 = 0.47 $g\,kg^{-1}$). Thujone is a strong neurotoxin and is found in thuja and sage. Pulegone can be oxidized to menthofuran in the liver and can cause liver damage.

If taken up erroneously in high quantities, intoxication and even death may result. This is especially the case for camphor (LD50 = approx. 1.5 $mg\,kg^{-1}$), which causes epileptic convulsions, and also for EOs from plants such as clove, wintergreen, and parsley.

There is very little knowledge about chronic toxicity, but it seems that the use of essential oils as food flavorings is essentially safe. Especially in aromatherapy, the interactions of products – which may be harmless if used alone – may lead to unexpected side-effects.

There have been several observations concerning skin toxicity (irritations or sensitation after topical applications, e.g., cosmetics or purfumes), including phototoxicity (e.g., photodynamic compounds in the Apiaceae = Umbelliferae family; see below). Allergic reactions caused by EO application are mainly due to the formation of isomers, cyclic compounds, or peroxides during storage under inadequate conditions such as high temperature, light, and access to air and oxygen (see below).

Carcinogenic effects have been observed in rodent models, in which EOs from sweet flag have been found to cause intestinal tumors, and hepatic tumors induced by sassafras, sweet basil, or terragon. Carcinogenic activity is probably due to various allylphenols and alkenylbenzenes (arene structures), and because of the different metabolic pathways, the toxicity has been calculated to be 13 million times higher in rodents than in humans.

Antioxidant Effects

The pathogenesis and symptoms of inflammatory processes are accompanied and/or initiated by the production of ROS. These include, for example, the OH radical, superoxide, hydrogen peroxide, hypochlorite, and peroxynitrite, and are generated, amongst others, by several types of inflammatory cells, e.g., neutrophils and alveolar macrophages, to defend the organism against invading microorganisms. Continuous production of these ROS accompanied by a weak antioxidant defense in the organism may lead to tissue and finally organ damage. Recently, many researchers have regarded antioxidative properties as an important fact in the mode of action of several drugs, food ingredients, or food additives. (*See* **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; Synthetic Antioxidants, Characterization and Analysis; Role of Antioxidant Nutrients in Defense Systems.)

EOs have been shown to have significant antioxidant properties in different biochemical model reactions simulating different pathological situations. It has been shown that several essential oils show a high antioxidant capacity regarding lipid peroxidation, which simulates the situation in biological membranes during oxidative stress. Since lipid peroxidation can cause cell death, EOs may be able to protect these cells. EOs also have a positive effect on reactions of stimulated neutrophilic granulocytes, which appear in great numbers during inflammatory processes. Therefore, the antioxidative capacity of essential oils may be an explanation for their antiinflammatory properties, acting by several different mechanisms on different organizational platforms.

Another pathological situation related to oxidative stress is the oxidation of low-density lipoprotein (LDL), which is regarded as an important factor during atherogenesis. Recently, it has been shown that after incubation of human plasma with components of essential oils, these can be enriched in LDL and protect it from copper-induced oxidation.

EOs and Food

As mentioned previously, citrus oils are used in perfumery and to flavor soft drinks or sweets. The use of spices to flavor foods is an indirect use of essential oils, since often, the characteristic taste of herbs is based on their content of EOs oils. EOs are also antibacterial against some species involved in meat spoilage, so the use of spices in food may not only improve the taste but also prevent spoilage. Essential oils of sweet basil, bay, cinnamon, clove, and thyme, for example, show antimicrobial activity against a wide range of foodborne Gram-positive and -negative bacteria.

Lipids containing unsaturated fatty acids are important food components, but they are very susceptible to autooxidation, which leads to undesirable products. To retard this process, different antioxidants like butylated hydroxytoluene and butylated hydroxyanisole are used. Owing to safety considerations (some of these synthetic antioxidants may be carcinogenic), natural antioxidants are under investigation as to whether they can replace synthetic antioxidants. Since essential oils have been shown to have a high antioxidant capacity (see section on Antioxidant Effects), particularly regarding lipid peroxidation, these substances may be good alternatives to synthetic antioxidants.

Stability of EOs

During storage of EOs photoisomerization, photocyclization, generation of peroxides and free acids, thermoisomerization, transition-metal catalyzed decay, and other reactions may initiate the decomposition of essential oils, which leads to changes in color and consistency. Since this alters the properties of the oil – the generation of peroxides in tea tree oil, for example, is responsible for the irritating activities of this oil – it is important to store essential oils securely under a nitrogen atmosphere, away from light and heat, and perhaps to add antioxidants.

For the determination of peroxides and thus the 'age' of an essential oil, conventional methods such as iodine titration and peroxide determination via photometrical procedures have been described, but the best and most convenient method is based on chemiluminescence by following the light emission of the peroxides formed in the presence of sulfite. This method is much faster and produces much less toxic waste than the traditional iodine titration.

See also: **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; Synthetic Antioxidants, Characterization and Analysis; Role of Antioxidant Nutrients in Defense Systems; **Essential Oils**: Isolation and Production

Further Reading

Baratta MT, Dorman HJD and Deans SG (1998) Chemical composition, antimicrobial and antioxidative activity of laure, sage, rosemary, oregano and coriander essential oils. *Journal of Essential Oil Research* 10: 618–627.

Baratta MT, Dorman HJ, Deans SG *et al*. (1998) Antimicrobial and antioxidant properties of some commercial essential oils. *Flavour and Fragrance Journal* 13: 235–244.

Bruneton J (1999) *Pharmacognosy, Phytochemistry, Medicinal Plants*, 2nd edn. Paris: Lavoisier.

Cowan MM (1999) Plant products as antimicrobial agents. *Clinical Microbiology Reviews* 12: 564–582.

Kreuzwieser J, Schnitzler J-P and Steinbrecher R (1999) Biosynthesis of organic compounds emitted by plants. *Plant Biology* 1: 149–159.

Leung AY and Foster S (1996) *Encyclopedia of Common Natural Ingredients*, 2nd edn. New York: John Wiley.

Lichtenthaler HK (1999) The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* 50: 47–65.

Lis-Balchin M (1997) Essential oils and 'aromatherapy': their modern role in healing. *Journal of the Royal Society of Health* 117(5): 324–329.

Loza-Tavera H (1999) Monoterpenes in essential oils: Biosynthesis and properties. *Advances in Experimental Medicine and Biology* 464: 49–62.

Paré PW and Tumlinson JH (1999) Plant volatiles as a defense against insect herbivores. *Plant Physiology* 121: 325–331.

Isolation and Production

B Sankarikutty and C S Narayanan, Council of Scientific and Industrial Research, Trivandrum, India

Background

Essential oils can be given a simple definition as the predominantly volatile and odorous fraction isolated by some physical process from vegetable materials. The term 'essential oil' was coined by Paracelsus and other alchemists to represent the quintessence or the total odor and flavor of each plant species. In plants, essential oils impart their distinctive and often diagnostic odor. Essential oil from a selected single plant species is made up of organic compounds whose nature and relative proportions are dependent on a number of agricultural factors, viz. environment, climate, soil conditions, time of harvesting and postharvest handling prior to isolation. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Essential oils isolated and identified from the vast number of plant species amount to over 3000, and of these, several hundreds have been produced commercially. The price of any particular commercial essential oil is dependent on the percentage oil yield from the plant species, its availability (production rate), and finally, the most important factor, its demand (application).

Essential oils are isolated from various plant components such as leaves, fruit, bark, root, wood, heartwood, gum, balsam, berries, seeds, flowers, twigs, and buds. The plant components are suitably processed to yield essential oils, which are mostly devoid of cellulose, glycerides, starches, sugars, tannins, salts, and minerals. The yield of essential oils from plants varies widely, and the broad range is 0.05–18.0%. Information on isolation, production, and markets of some selected commercial essential oils are presented in Table 1. The essential oil is present in various plant components in oil sacs and is isolated by comminution and the action of heat, water, and solvents. Distillation, selective solvent extraction, and mechanical expression are the three general basic methods employed for essential oil isolation with improvements or modifications incorporated for each as available. Thus, essential oils include concretes (flower 'concretes' – hexane extractives), absolutes, resinoids, distillates, etc. (*See* **Herbs**: Herbs and Their Uses.)

Distillation Techniques

The most popular physical method for essential oil isolation is distillation. Prior to distillation, plant materials are in most cases dried and are then suitably ground so that the oil sacs are broken and a maximum of surface area is exposed for efficient oil release. Various distillation techniques widely adopted for essential oil isolation are briefly discussed below.

Direct-heating Distillation (Hydrodistillation)

Suitably ground plant materials are placed in a boiler with water completely covering them. As heat is slowly applied, steam alone will be initially formed, and the distillate will be clear. With continued heating, the essential oil starts to distil over with the steam, and the distillate becomes milky white. Distillation is continued until the distillate becomes clear, with no more oil distilling over from the material. Even though the essential oils have relatively high boiling points, codistillation like this brings about a satisfactory recovery of the oil because, in accordance with Dalton's law, a mixture boils when the sum of the vapor pressures of the individual components equals the atmospheric pressure. However, this method is somewhat slow and requires close manual attention and separation of the oil and aqueous phases of the condensate.

Steam Distillation

This method is widely used and is faster than direct distillation. Steam under pressure is passed through the prepared plant material, and the volatile oils are condensed along with the water. Steam can also be bled directly into the boiler. To avoid thermal decomposition of lower-boiling-point constituents, steam pressure is increased only gradually. The distillation temperature with water or steam at atmospheric pressure is usually slightly under 100 °C and can be reduced further by drawing a partial vacuum on the still.

-Hydrodiffusion In this method, the steam enters the still at the top and passes down through the charge. Oil and water vapor condense on coils fixed below the bottom grid within the still. The oil and water are separated in the usual way. The plant design results in considerable energy savings and minimum degradation of the volatile oil. This method is found to be very advantageous in the distillation of seed oils.

Vacuum Distillation

This method is faster than steam distillation. Vacuum distillation is generally adopted to rectify an oil and very rarely to distil an oil directly from the plant material.

Table 1 Isolation, production, and markets of a selection of commercial essential oils

Essential	*Botanical source*	*Plant part used; method of oil isolation*	*Principal producing countries/areas*	*Approximate volume of production (tonnes per year)*	*Major consuming markets*
Amyris	*Amyris balsamifera* L. First	Wood; steam distillation	Haiti	70	France, Germany, USA
Anise	*Pimpinella anisum* L.	Dried fruit; steam distillation	Spain, Russia, Poland	40–50	France
Anise, Star	*Illicium verum* Hook f.	Dried star-shaped fruit; steam distillation	China, Vietnam	60–70	France
Basil (Réunion)	*Ocimum basilicum* L.	Herbaceous tops; steam distillation	Réunion, Comoros, Madagascar, South Africa	10–12	
Basil (sweet)	*Ocimum basilicum* L.	Herbaceous tops; steam distillation	Egypt	~2	USA, Europe, Switzerland, Japan
Bay, Turkish (Laurel)	*Laurus nobilis* L.	Leaves; steam distillation	Turkey	1–2	Germany, Netherlands, USA, UK
Bay, West Indies	*Pimenta racemosa* Mill	Leaves; steam distillation	Dominica	40–45	USA, UK, Switzerland
Bergamot	*Citrus aurantium* L. ssp. *bergamia*	Fruit peel (rind); expression	Italy, Ivory Coast	175–225	USA, France, Netherlands
Caraway	*Carum carvi* L.	Ripe seeds; steam distillation	Western Europe	10–30	North America
Cardamom	*Elettaria cardamomum* Maton	Seeds; steam distillation	India, Sri Lanka, Guatemala	10–20	Western Europe, Scandinavia, USA
Cassia	*Cinnamomum cassia* Blume	Leaves, twigs and inner bark; steam distillation	China	140–150	USA, Japan, Western Europe
Cedarwood, Southern Red, East African	*Juniperus virginiana*	Wood; steam distillation	USA, China, Kenya	700–1 400	USA, Western Europe, Japan
	Juniperus procera	Wood; steam distillation			
Cedarwood, Texas	*Juniperus mexicana* Schiede	Wood; steam distillation			
Cinnamon	*Cinnamomum zeylanicum* Nees	Dried inner bark; steam distillation	Sri Lanka	100	Western Europe, USA, India
Citronella, Ceylon	*Cymbopogon nardus* Rendle	Herb; steam distillation	Sri Lanka	~2 000	USA, Europe, Japan, Mexico
Citronella, Jawa	*Cymbopogon winterianus* Jowitt	Herb; steam distillation	Indonesia, China, Guatemala, India	~2 000	USA, Europe, Japan, Mexico
Clove	*Eugenia caryophyllata* Thunb	Unopened bud; steam distillation	Madagascar, Indonesia, Tanzania, Sri Lanka	2 000	USA, Western Europe, Japan
Coriander	*Coriandrum sativum*	Seeds; steam distillation	Russia	20–40	USA
Davana	*Artemisia pallens* Wall.	Flowers; steam distillation	India	~2	Western Europe, USA
Eucalyptus	*Eucalyptus globulus, E. citriodora, E. dives* 'A' & 'C'	Leaves; steam distillation	China, Portugal, Spain, South Africa, Brazil, Australia	~1 800	Western Europe, USA, Japan
Fennel, bitter	*Foeniculum vulgare* var. *vulgare*	Seed; steam distillation	Germany, Spain, India, China	10–20	USA, Japan
Fennel, sweet	*Foeniculum vulgare* var. *duler*	Seed; steam distillation	Spain	10–20	France
Garlic	*Allium sativum* L.	Bulbs; steam distillation	Egypt	2–3	France, Spain

Continued

Table 1 Continued

Essential	*Botanical source*	*Plant part used; method of oil isolation*	*Principal producing countries/areas*	*Approximate volume of production (tonnes per year)*	*Major consuming markets*
Geranium	*Pelargonium graveolens, P. radula*	Green herbage; steam distillation	China, Egypt, Réunion, Algeria, Morocco	~300	USA, Western Europe, Japan
Ginger	*Zingiber officinale* Roscoe L.	Rhizomes; steam distillation	India, China	~30	USA, Europe, Japan
Jasmine	*Jasminum officinale* L.	Flowers; solvent extraction	Egypt, Morocco	12–15	Western Europe, France, USA
Lavandin	*Lavandula hybrida*	Flowering tops; steam distillation	France, Italy, Spain, Hungary	~750	USA, Europe, Japan
Lavender	*Lavandula officinalis* Chaix	Flowering tops; steam distillation	Eastern Europe, France, Spain	~200	
Lavender, spike	*L. latifolia* (DC) Vill.	Flowering tops; steam distillation	Spain, France	~50	
Lemon	*Citrus limon* L.	Outer peel; expression	USA, Argentina, Italy, Ivory Coast, Brazil	2 000–2 500	Europe, Switzerland, Japan, Russia
Lemon grass	*Cymbopogon flexuosus* Stapf *C. citratus* Stapf	Green herbage; steam distillation	India, Guatemala, China	800–1 300	Western Europe, USA, Russia, Japan
Lime	*Citrus aurantifolia* Swingle, *Citrus latifolia* Tan.	Outer peel; distillation or expression	Mexico, Peru, Jamaica, other West Indian Islands, Western Africa	~900	USA, Europe, Switzerland, Japan
Litsea cubeba	*Litsea cubeba*	Fruit; steam distillation	China	500–600	USA, Western Europe, Japan
Mint					
Corn mint	*Mentha arvensis* L.	Green herbage; steam distillation	China, South America	3 000–4 000	Brazil
Peppermint	*M. piperita* L.	Green herbage; steam distillation	USA	2 000–2 500	Europe, Japan
Spearmint	*M. spicata* L.	Green herbage; steam distillation	USA	1 000	Europe
Neroli	*Citrus aurantium* L. ssp. *amara* L.	Fresh blossoms; steam distillation	France, Italy, Tunisia, Egypt, Comoros	~ 2	USA, Japan, UK, Germany
Nutmeg	*Myristica fragrans* Houtt.	Kernels of fruits; steam distillation	Indonesia, Sri Lanka, Grenada	120–130	USA, Western Europe
Onion	*Allium cepa* L.	Bulbs; steam distillation	France, Egypt	~5	Western Europe
Orange, bitter	*Citrus aurantium* ssp. *amara*	Fruit peel; expression	Jamaica, Dominican Republic, Haiti, Brazil	20–25	USA, Western Europe
Orange, sweet	*Citrus aurantium* L. var. *dulcis*	Fruit peel; expression	Brazil, USA, Central America, West Indies	15 000	USA, Western Europe, Japan
Palma rosa	*Cymbopogon martin* Stapf var. *motia*	Herbaceous tops; steam distillation	India, Indonesia, Brazil	60–70	France, USA, India, Brazil
Patchouli	*Pogostemon cablin* Benth.	Leaves; steam distillation	Indonesia, China	500–550	USA, Europe, Switzerland, Japan, India
Pepper	*Piper nigrum* L.	Berries; steam distillation	India, Indonesia, USA, UK	25–35	USA, Germany, UK
Petitgrain	*Citrus aurantium* ssp. *amara*	Leaves and twigs; steam distillation	Paraguay, Mediterranean countries	280	USA, Europe, Switzerland
Pimento (Allspice)	*Pimenta dioica* L.	Fruits; steam distillation	Jamaica, USA	35–50	USA, Western Europe, Russia

Continued

Table 1 Continued

Essential	*Botanical source*	*Plant part used; method of oil isolation*	*Principal producing countries/areas*	*Approximate volume of production (tonnes per year)*	*Major consuming markets*
Rose	*Rosa damascena* Mill., *Rosa centifolia* L.	Flowers; steam distillation	Bulgaria, Turkey, Morocco	15–20	USA, Europe
Rosewood	*Aniba rosaeodora*	Wood; steam distillation	Brazil, Peru	150	USA, Western Europe, Brazil
Sandalwood	*Santalum album* L.	Wood; steam distillation	India, Indonesia	250–350	USA, Western Europe
Tagetes (marigold)	*Tagetes glandulifera* Schrank	Flowers, stem and leaves; steam distillation	East and South Africa	~1	UK
Thyme	*Thymus vulgaris* L., *Thymus zygis* L. var. *gracilis* Boiss	Herbaceous tops; steam distillation	Spain	20–30	USA, Western Europe
Vetiver	*Vetiveria zizanioides* Stapf	Root; steam distillation	Haiti, Indonesia, Réunion, China	245–265	USA, Western Europe, Japan
Ylang-ylang	*Cananga odorata* forma *genuina*	Flowers; steam distillation	Comoros, Madagascar	120–130	USA, Western Europe, East and South-east Asia

Quality of Distilled Essential Oils

Conditions of distillation must be carefully established and controlled according to the nature of the raw material to ensure *optimum*, not necessarily maximum, yields of quality essential oil. Extended periods of distillation adversely affect the oil quality and also the cost of distillation. The distillation conditions must be established by trial in order to maintain the quality of the oils and also to avoid only batch-to-batch variation.

The yield of oil depends on the efficient separation of the essential oil from the steam/oil vapor condensate. The distillation water also may contain a variable proportion of the essential oil in colloidal suspension or in solution. Its recovery is achieved either by redistillation or by continuously recycling the aqueous distillate through the still, and this system is termed 'cohobation.' When the recovery is incomplete or unsatisfactory by cohobation, the distillate is passed through a packed column of solvent or codistilled with an immiscible solvent such as hexane or pentane, and the solvent is carefully removed by controlled high-vacuum distillation. Some losses of low-boiling components will occur, and these oils must be used with discretion.

A modified frontal liquid chromatographic technique has also recently been employed for recovery of oil from the distillate. Here, the distillate water flows through a column packed with porous poly(tetrafluorethylene), which holds on its surface 20% by volume of a low-boiling water-immiscible solvent. The solvent is renewed as necessary during the process of distillation, and the amount of solvent required is about 1–3% of the water volume. Essential oil is recovered from water distillates by this technique at $500\,l\,h^{-1}$ in a $30\,cm \times 60\,cm$ column.

Rectification

Essential oils are redistilled to improve a particular property, to effect a higher degree of separation and purity, or to concentrate or enrich a particular fraction of the oil. Such redistilled oils are called rectified oils. Traces of water, resinous materials, undesirable flavor notes, etc. may be removed by rectification. The processing methods for concentration or enrichment include, in addition to fractional distillation, solvent extraction, countercurrent extraction, thin-film evaporation and molecular distillation. In some cases, distillation followed by any of the other methods is also employed for the desired fractionation or enrichment of the oil.

Solvent Extraction

Heat-sensitive essential oils can be isolated by extraction using an organic solvent. Solvents used should be of a low boiling point, free of odor and impurities, and inert towards the oil constituents. Pentane or hexane is preferred for flower oils. For spices, in most cases, solvent extracts (volatile and nonvolatile fractions) are more representative of the total flavor than is the steam-distilled oil, which contributes only towards the aroma. To obtain spice essence (oleoresin) extracts, solvents such as ethanol, chlorinated

hydrocarbons, and acetone are employed, solvent selection depending on the nature of the spice and its flavor constituents.

Expression (Pressing)

This method is generally employed for fruit peels rich in oils. In this technique, the selected whole fruits are washed and then crushed between rollers, and oil is then separated from the juice. Also, they are automatically handled in specially designed extractors that recover both juice and oil concurrently. The oil collected is washed with a spray of water and then separated by means of a centrifuge. The method is limited generally to citrus fruits (orange, lemon, and lime). Steam-distilled citrus oils are usually byproducts of the frozen juice concentrate industry. The oil may be recovered by distillation from the juice-and-oil emulsion. The expressed oils and distilled oils differ widely in their flavor characteristics.

Enfluerage

Essential oils of delicate flower petals are prepared by this technique. The method involves the absorption of oil on purified fat on special pressing equipment. The process is repeated until the fat is saturated, and it is then removed and extracted with alcohol and recycled. To reduce the time and cost of the process, hot fat at 40–60 °C is also used for absorption of the oil.

Carbon Dioxide Extraction

High-pressure extraction with supercritical or liquefied gas, e.g., CO_2, is a relatively new technique for obtaining essential oils without thermal or hydrolytic deterioration. Thus, the product obtained by this technique is without artefacts or solvent residues. The use of liquefield CO_2 at 0–10 °C and 8–80 bar results in a product devoid of fixed oils, protein, waxes, chlorophyl and pigments. In the supercritical state, the pressures generally used for extraction of essential oils at 40–80 °C range from 75 to 200 bar.

Considerations of Quality

The composition, and thereby the quality, of the essential oil of a selected plant species depends mainly on the nature of raw material used, the method of isolation employed, and the oil's storage conditions. In distilled oils, differences in oil quality are more likely to arise from variations in distillation time, rate of distillation, efficiency of vapor condensation, method of separation, and postdistillation handling. Irrespective of the method of isolation adopted, to produce a standard quality of essential oil, it is always necessary to ensure that the raw material used is of the correct botanical source and geographical origin and that it is clean and free from contamination and has been correctly prepared for oil isolation. Also, the conditions or the relevant parameters involved in the chosen method must be clearly defined.

See also: **Flavor (Flavour) Compounds**: Structures and Characteristics; **Herbs**: Herbs and Their Uses

Further Reading

Considine DM and Considine GD (1982) *Foods and Food Production Encyclopedia*. New York: Van Nostrand Reinhold.

Dafesera DJ, Tarantilis PA and Polissiou MG (2002) Characterization of essential oils from lamiaceae species by fourier transform Raman spectroscopy. *Journal of Agricultural Food Chemistry* 50(20): 5503–5507.

Essential Oils and Oleoresins: A Study of Selected Producers and Major Markets (1986) Geneva: International Trade Centre UNCTAD/GATT.

Guenther E (1949) *The Essential Oils*, vols I–VI. New York: D. Van Nostrand.

Heath HB (1981) *Source Book of Flavours*. Westport, CT: AVI.

Rogers JA, Jr. (1981) Oils, essential. In: *Kirk-Othmer Encyclopedia of Chemical Technology*, vol. XVI, 3rd edn. New York: Wiley.

Wilson NO, Ivanova MS, Watt RA and Moffat AC (2002) The quantification of citral in lemongrass and lemon oils by near-infrared spectroscopy. *Journal of Pharmacy and Pharmacology* 54(9): 1257–1263.

Ethanol *See* **Alcohol**: Properties and Determination; Metabolism, Beneficial Effects, and Toxicology; Alcohol Consumption

ETHNIC FOODS

J Dwyer and O I Bermudez, Jean Mayer US Department of Agriculture, Human Nutrition Research Center on Aging at Tufts University, Boston, MA, USA

Introduction

Ethnic foods are foods that are regarded as unique to a particular cultural group, race, religion, nation, or heritage. Consumers of ethnic diets often have cultural but also socioeconomic, religious, or regional characteristics that are distinctive. From the cultural standpoint, ethnic foods serve as a familiar link with the past and help those who are accustomed to them maintain ethnic identity. Ethnic foods provide eaters from other culinary traditions with opportunities to experience new cuisines and cultures. They are also important in designing or tailoring culturally relevant nutrition interventions to reach nutritionally vulnerable, hard-to-reach ethnic groups within the community. In addition to being familiar, many ethnic foods have actual or perceived health importance that may heighten their appeal. Finally, ethnic foods provide a platform for developing attractive new foods that appeal to broader segments of the population.

Foods that are regarded as ethnic foods in a culture vary over time and with the degree of acculturation of the group consuming them. Even within major race and ethnic subgroups, many variations in ethnic foods exist. Some foods only bear a passing resemblance to the original cuisine from which they were derived. For example, in the USA, pizza, tacos, chile con carne, and chicken chow mien are very different from the Italian, Mexican, and Chinese dishes that inspired them. Hispanic-Americans with a Mexican-American heritage differ in their food preferences from Puerto Rican Americans, whose eating patterns involve a blend of Spanish, African, Taino Indian, and continental American influences. Similarly, Asian-Americans of Filipino, Japanese, or Chinese origins have very different food choices and traditions. Among the 300 different Native American groups, the staples and other foods that they regard as unique and distinctive to their cultures also vary. In the UK, foods from China, Japan, and the Indian subcontinent are referred to as Asian food whereas in the USA, they refer to Chinese and Japanese cuisines.

Most ethnic cuisines have been influenced by centuries of food exchanges between indigenous groups and immigrants from all over the world. For example, the American cuisine is a multiethnic blend. No single eating pattern has prevailed over all the others, and no ethnic or cultural group has emerged with an unchanged food pattern.

The traditional Mediterranean diet includes liberal amounts of fruits, vegetables, legumes, and grains, high amounts of monounsaturated compared to saturated fats, moderate consumption of alcohol, liberal amounts of fish, and low amounts of meat and milk products. It has been associated with decreased risks of coronary artery disease. Although there is debate about the extent to which such a regimen is exportable, some Mediterranean ethnic foods have become very popular.

Ethnic food preferences and choices are important aspects of the cultural heritage they represent. Some of the many sources available on food-related aspects of culture are summarized in Table 1; these include both the cultural anthropology literature on foodways, cookbooks, and nutrition information materials that describe the foods and their uses.

Origins of Ethnic Foods

Table 2 describes the factors that are thought to be involved in the development of ethnically distinctive cuisines and ethnic foods. Nutritional anthropologists subscribe to various theories on how ethnic food patterns have developed (Table 3). No one theory appears to predominate at present.

Individuals in different ethnic groups have different ways of categorizing foods, which are sometimes at variance with medical views of foods as sources of nutrients that are based on physiological and biochemical characteristics. They may also differ in their attitudes toward medical care and their health behaviors, in their beliefs about the therapeutic and preventive properties of foods, and the appropriate balance between different foods. For example, there are systems in various cultures that categorize foods into hot and cold, or yin or yang – both classification schemes that are based on properties other than nutrient content (Table 3).

Ethnic Differences in Food Classification Systems

Cultures differ not only in the foods that are eaten but also in how foods are viewed and classified by eaters. Some of the relevant factors are listed in Table 4.

Ethnic groups vary in their foodways, social and physical contexts in which food preparation, serving, and eating occur. Food classification systems may

Table 1 Useful sources to consult on ethnic-specific foods, food habits, and cuisines

Scientific journals
Journal of Food Composition and Analysis
Journal of the American Dietetic Association Ethnicity and Disease
Nutrition International
Proceedings of the Nutrition Society
Appetite
Nutrition Today
American Journal of Clinical Nutrition
World wide web
http://www.sallys-place.com/ food/ethnic_cuisine
Oldways http://www.oldwayspt.org/htm/meet/htm
Food-based dietary guidelines and graphics for ethnic eating patterns
Ethnic cookbooks
Eating Well
Flavor and Fortune
Food Heritage Press (Duxbury, MA)
Time-Life series on foods of the world
Nutritional anthropologists
Organizations
American Institute of Wine and Food
Oldways Food Preservation Trust
Society of Ingestive Behavior

Table 2 Factors involved in the development of ethnically distinctive cuisines

- Food availability constraints (geography, climate, range of native plant and unusual species in area)
- Social factors: migrant (both voluntary and involuntary) social structures, social relationships, social status
- Economics
- Philosophical, religious, and moral attitudes
- Individual factors

affect nutritional status because they cause the exclusion of healthful foods or food groups from the diet, encourage consumption of harmful items, or produce a false sense of security among eaters that all of their nutritional needs are met when in fact they are not.

Dietary Assessment

Familiarity with ethnic foods is important for dietary assessment. The foods themselves, their names, composition, portion sizes, and frequency of consumption vary. Accurate assessment of nutrient intakes depends on a knowledge of these factors. Some ethnic foods may be especially rich sources of nutrients or other food constituents, such as phytochemicals. Presently, there is much interest in the possible health effects of the traditional soy foods contained in many Asian diets, which contain relatively high amounts of soy isoflavonoids. These compounds have weak estrogenic effects and have been associated with decreased risks of coronary heart disease in some studies. Other ethnic foods such as some home-brewed alcoholic beverages contain contaminants such as heavy metals that have adverse health effects.

Table 3 Theories on how ethnic food patterns develop

Theory	*Comments*
Environmentalism	Ethnic food traditions are determined primarily by environmental and external influences
Cultural determinism	Humans mold their environments and diets quite independently of their physical surroundings
Cultural ecology	Both culture and environment are important influences
Cultural history	Historical influences affect how foods evolve and develop over time
Functionalism	Foods are used to identify or place people within society

Table 5 provides some examples of the importance of using ethnically appropriate dietary assessment tools. One challenge in evaluating diets that include many ethnic foods is the appropriate choice of methods for dietary assessment. Methods such as semiquantitative food frequency questionnaires are often designed to capture intakes of important nutrients for those who conform to the predominant food culture. These are often inappropriate for subgroups with very different dietary patterns. Subgroup members may call the foods by different names, consume items not included on the food lists, or vary in amount and frequency of consumption. The use of such instruments to assess intakes of individuals with unconventional dietary preferences can lead to highly erroneous conclusions. Subgroup-specific food frequency questionnaires that take account of these factors are sometimes available, and when they are these should be used. For example, well-validated ethnic-specific food frequency questionnaires have been developed for Japanese and Chinese-Americans living in Hawaii, and for elderly Puerto Ricans. However, for many groups, it is not possible to obtain accurate information on intakes without further data collection and validation of new food frequency questionnaires. Some information can be obtained by adding important ethnic foods to existing food lists, but nutrient analyses are still likely to be inaccurate, since consumption of foods within each food group on the otherwise unadjusted questionnaire is apt to be very different among those with ethnic eating patterns than it is among those who consume mainstream diets. Another alternative is to use food

Table 4 How food is classified in different cultures

Food habits	Foods categorized as core, secondary core, or peripheral, depending on their frequency of consumption
Gastronomic	Foods versus nonfood, edible versus nonedible
Religious/philosophical	Sacred versus profane, allowed versus forbidden
Opposing categories or properties	Hot versus cold, yin versus yang, or other schemes for balancing food intakes to prevent or cure disease. The important properties cannot be determined from their chemical composition
Medicinal	Foods used as treatment for certain diseases or psychological problems
Social system	Foods are viewed as symbolic of a culture or belief
Emotional significance	Foods for the sick, foods for celebrations or special needs
Nutritional nutrients	Chemical composition – by macro and micro
Hygienic	Wholesomeness, cleanliness, freedom from environmental contaminants, microorganisms or other toxins

Modified from Kaufman-Kurzrock DL (1989) Cultural aspects of nutrition. *Topical Clinical Nutrition* 4: 1–6.

Table 5 Examples of the importance of ethnically appropriate dietary assessment tools

Kristal AR, Feng Z, Coates RJ *et al*. (1997) Associations of race, ethnicity, education and dietary intervention with the validity and reliability of a food frequency questionnaire: the Women's' Health Trial feasibility study in minority populations. *American Journal of Epidemiology* 146: 856–867

Tucker KL, Bianchi L, Maras J, Bermúdez OI (1998) Adaptation of a food frequency questionnaire to assess diets of Puerto Rican and non-Hispanic adults. *American Journal of Epidemiology* 148: 507–518

Wheeler M, Haider SQ (1979) Buying and food preparation patterns of ghetto blacks and Hispanics in Brooklyn. *Journal of the American Dietary Association* 75: 560–563

Winkelby MA, Albright CL, Howard PB, Lin J, Fortmann SP (1994) Hispanic/white differences in dietary fat intake among low educational level adults and children. *Preventive Medicine* 23: 465–473

records, recalls, or direct observation of food intake. This permits exact descriptions of the amount, type, and use of all foods, including unusual items. Food records also provide information on how ethnic foods are used in everyday diets. They identify differences in food use that may exist between the subgroup and the mainstream culture. Food records also provide the information from which ethnicity-specific dietary probes and semiquantitative food frequency questionnaires can be developed. For those who eat many different foods in small amounts rather than discrete portions of foods, or for those who eat from a common pot, these assessment techniques are more appropriate than food frequency questionnaires.

A second problem in the dietary assessment of ethnic cuisines is that nutrient databases for converting dietary intakes into nutrients for ethnic foods may not be available. In research studies on individuals who use many ethnic foods, nutrient databases that include entries for ethnic favorites are needed to obtain complete information about intakes. Whenever possible, food tables that are specific to the country or ethnic group should be used. Country-specific tables of food composition and nutrient databases are increasingly becoming available and these are the best option. For individuals who eat largely ethnic cuisines in countries where the parent food culture is different, often special food composition tables for the ethnic foods that are most popular in mainstream diets are available. However, the recipes described may differ somewhat from those that members of the ethnic group prepare for themselves. Even for ethnic foods that are included in nutrient databases, food composition information may be incomplete, inappropriate, or lacking. Moreover, nutrient values for foods that are rarely consumed outside of the ethnic group may be absent, especially if the ethnic group is small. This is a problem with many local foods consumed by Alaskan natives, for example.

The dietary assessment task is more difficult when no country or ethnic specific food table is available. Recipes made from locally available ingredients or imported foods may differ in their food composition from recipe values in food tables used in western countries. For example, a typical curry dinner as served in India, Ghana, and the UK varies greatly in its nutrient composition. For this reason, country-specific food consumption tables are needed. The option of adding ethnic foods to the food composition database for the foods in question from country-specific food tables prepared elsewhere is not recommended unless there is good reason to suspect that the foods are similar. Imported ethnic foods and those eaten in the country of origin may differ considerably both in their nutritive value and in other food constituents.

The degree of error that incomplete entries for ethnic foods introduce in nutrient calculations may be large, so it is important to consider these factors in interpreting the resulting information. Food composition tables should separate true null values (analyzed and found to be devoid of the nutrient in question) from simple lack of information (no analytic information available). Before drawing conclusions about the nutrient contributions of such foods, it is important to ascertain how much information is missing.

Some computerized nutrient databases now include food composition values for imported ethnic foods. When such values are not included, the best option is to obtain representative chemical analyses of the major food items that are likely contributors of the nutrients or other substances of interest. This procedure is time-consuming and expensive but it may be necessary if the task is to describe the intakes of individuals who consume large amounts of specific items of unknown chemical composition that are expected to make valuable nutritional contributions.

A third option is to use botanical and zoological classification schemes and expert knowledge of food-processing techniques to develop provisional values for the likely food composition of unknown foods. However, such estimates are usually very crude.

In the past few years, some culture-specific interactive computer programs have become available for some ethnic groups (e.g., Mexican-American, Chinese-American) that greatly simplify dietary assessment tasks. These programs include photographs of various ethnic favorites as well as mainstream foods, and common food names in both English and the native language, and permit queries by typical meal patterns. Expanded nutrient database entries that include foods eaten in both cultures permit accurate calculation of nutrient intakes.

Dietary Planning and Interventions

Table 6 provides some examples from the recent literature of ethnically appropriate nutrition interventions.

The consumption of ethnic foods is at root a cultural phenomenon but it has physiological implications. Sometimes alterations in eating habits are necessary to achieve nutritional ends, but this is a difficult goal to achieve. Food habits are among the oldest and most deeply entrenched aspects of many cultures, and they cannot be easily changed. When they are changed, they often produce unexpected and unwelcome reactions.

There is no simple formula for success, but ethnic-specificity and cultural sensitivity are important ingredients in crafting successful nutrition planning and interventions. The expertise of nutritional anthropologists who specialize in cultural anthropology as it relates to food and foodways is helpful. Some general principles that must be kept in mind are discussed below.

Some foods are more important than others to eaters. Foods in diets constitute a continuum or gradient that consists of the core diet which includes universal, regular staple foods; the secondary core of foods that are used widely but not universally; and peripheral foods, that are used infrequently. Such groupings are useful in planning nutritional interventions since they often suggest opportunities for making acceptable changes as well as possible pitfalls that need to be avoided. Food habits are always difficult to change. The most resistant habits of all involve foods in the core diet, especially those that are also viewed by the eaters as ethnically appropriate and vital to group cohesion.

The language used about foods, and the names of favorite foods themselves must be used in recruitment. Ethnic-specific nutrition information, recommendations, dietary guidance, and education are also vital if interventions are to be appropriate to the dietary patterns and culture of intended recipients. In many countries, ethnicity, socioeconomic status, educational level, geographical location, religious, and ethical views are often confounded. As these other factors vary, so will the appropriateness of the intervention.

The importance of including ethnic foods in determining the ultimate acceptance of nutrition interventions varies within as well as between ethnic groups. In general, those who are the least accultured to the larger mainstream society, those with language barriers, the ill, the geographically isolated, the economically disadvantaged, traditionalists, and the religiously conservative may regard ethnic foods as particularly important. Those who are innovators and desire acceptance in the larger culture may regard ethnic foods as less important. The degree to which new foods are incorporated into eating styles varies depending on contact between cultures, food availability, trade, and social status of the group introducing the food, as well as philosophical, religious, and individual factors. Immigrants differ in the extent to which they wish to remain unassimilated in their foodways and true to their native cuisines. Ethnic borrowing and mingling rather than rigid preservation of unique eating traditions often occur.

Ethnic-specific food-based dietary guidelines and graphics to guide food choices are now available that present nutritionally appropriate food patterns in a culturally acceptable manner for many ethnic groups. Many guidelines are currently available.

Table 6 Examples of the importance of ethnically appropriate nutrition interventions

Coates RJ (1992) Dietary intervention in low income, minority women. In: *President's Cancer Panel: National Cancer Institute.* Bethesda: National Cancer Institute: National Institutes of Health

Coates RJ, Bowen DJ, Kristal AR *et al*. (1999) The Women's' Health Trial feasibility study in minority populations: changes in dietary intakes. *American Journal of Epidemiology* 149: 1104–1112.

Connett JE, Stamler J (1984) Responses of black and white males to the special intervention program of the Multiple Risk Factor Intervention Trial. *American Heart Journal* 108: 839–878

Gorder DD, Dolocek TA, Coleman CG *et al*. (1986) Dietary intake in the MRFIT: nutrient and food group changes over 6 years. *Journal of the American Dietary Association* 86: 744–751

Howard-Pitney B, Winkelby MA, Albright CL, Bruce B, Fortmann SP (1997) The Stanford Nutrition Action Program: a dietary fat intervention for low literacy adults. *American Journal of Public Health* 87: 1971–1976.

Jerome NW (1980) Diet and acculturation: the case of Black-American immigrants. In: Jerome NW, Kandel RF, Pelto GH (eds) *Nutritional Anthropology*, pp. 275–325. Pleasantville, New York: Redgrave

Koehler KM, Harris MB, Davis SM (1989) Core, secondary and peripheral foods in the diets of Hispanic, Navajo, and Jemez Indian children. *Journal of the American Dietary Association* 89: 538–540

Kumanyika SK, Morssink C, Acturs T (1992) Models for dietary and weight change in African-American women: identifying cultural components. *Ethnicity and Disease* 2: 166–175

McNabb W, Quinn M, Kerver J, Cook S, Karrison T (1997) The Pathways church based weight loss program for urban African American women at risk for diabetes. *Diabetes Care* 20: 1518–1523

Nestle M (1995) Mediterranean diets: historical and research overview. *American Journal of Nutrition* 61 (suppl 13): 135–205

Meiselman HEI (2000) *Dimensions of the Meal: The Science Culture, Business and Art of Eating*. Gaithersburg, Maryland: Aspen Publishers

Pareo-Tubbeh SL, Romero LJ, Boumgartner RN, Garey PS, Lindemen RD, Koehler KM (1999) Comparison of energy and nutrient sources of elderly Hispanics and non-Hispanics in New Mexico. *Journal of the American Dietary Association* 99: 572–582

Passin H, Bennett (1943) JW Social process and dietary changes. In: *The Problem of Changing Food Habits*. Council Bulletin no. 1908. Washington, DC: National Research

Rozen E (1989) *Flavor Principle Book*. New York: Penguin Books

Trichopoulou A, Lagions PH (1997) Healthy traditional Mediterranean diets; an expression of history and lifestyle. *Nutrition Review* 55: 383–389

They have the advantage of including ethnic foods as a part of overall diets. These guidelines are helpful as starting points for developing population-based nutrition education materials for healthy persons. Graphics that depict appropriate food choices have the additional advantage of being understandable even to those who may be illiterate.

Special attention must be paid to the ethnic appropriateness of nutritional counseling for those who are ill or in especially vulnerable groups from the physiological standpoint. Therapeutic diet manuals are sometimes available or obtainable that can be helpful in designing diets that provide medical nutrition therapy that is culturally appropriate.

Additional useful techniques for crafting nutrition information, education, and counseling interventions include the use of ethnic and regional terms for foods and recipes. Interventions should be adopted to be culturally appropriate. Food preparation and cooking demonstrations can be used to show how to incorporate recommendations into existing food patterns. Nutrition education and counseling should include attention to ethnic issues. Community social structures in the ethnic group are helpful to make contacts with the target ethnic groups.

Those who develop nutritional interventions that involve food enrichment, fortification, or special supplements must also pay attention to the role of ethnic foods and food patterns in the diet. In many cultures the very target groups for such efforts, who are nutritionally vulnerable because of their age, sex, economic status, or physiological state, are likely to have eating practices that are deeply embedded in their culture. Persons in certain age, sex, or physiological categories are often expected to avoid some foods and to eat others. Nutrition counselors must remember that care of the sick is heavily intertwined with food and health beliefs, social relationships, and acculturation. In summary, nutrition workers must avoid cultural biases and ethnocentricity in dealing with their clients in both dietary assessment and planning.

See also: **Dietary Surveys**: Measurement of Food Intake; Surveys of National Food Intake; Surveys of Food Intakes in Groups and Individuals

Further Reading

Amler RW and Hull HJB (eds) (1987) *Closing the Gap: The Burden of Unnecessary Illness*. New York: Oxford University Press.

Bermudez OI, Falcon LM and Tucker KL (2000) Intake and food sources of macronutrients among older Hispanic adults: association with ethnicity, acculturation, and length of residence in the United States. *Journal of the American Dietary Association* 100: 665–673.

Broussard-Marin L and Hynak-Hankinson MT (1989) Ethnic food: the use of Cajun cuisine as a model. *Journal of American Dietary Association* 89: 1117–1121.

Brown LK and Mussel K (eds) (1984) *Ethnic and Regional Foodways in the US: The Performance of Group Identity*. Knoxville, TN: University of Tennessee Press.

Cassel J (1957) Social and cultural implications of food and food habits. *American Journal of Public Health* 47: 732–740.

Davidson A (1995) *The Oxford Companion to Food*. Oxford: Oxford University Press.

Eliades DC and Suitor CW (1994) *Celebrating Diversity: Approaching Families Through their Food*. Arlington, VA: National Center for Education in Maternal and Child Health.

Gabaccia DR (2000) *We are What we Eat: Ethnic Food and the Making of Americans*. Cambridge, Massachusetts, USA: Harvard University Press.

Grivetti LE (1981) Cultural nutrition: anthropological and geographical themes. *Annual Review of Nutrition* 1: 47–68.

Harwood A (1981) *Ethnicity and Medical Care*. Cambridge, Massachusetts, USA: Harvard University Press.

Jerome NW, Kandel RF and Pelto GH (1980) *Nutritional Anthropology: Contemporary Approaches to Diet and Culture*. Pleasantville, NY: Redgrave.

Kaufman-Kurzrock DL (1989) Cultural aspects of nutrition. *Topics in Clinical Nutrition* 4: 1–6.

Kavanaugh KH and Kennedy PH (1992) *Promoting Cultural Diversity: Strategies for Health Care Professionals*. Newbury Park, CA: Sage Publications.

Keys A (1995) Mediterranean diet and public health: personal reflection. *American Journal of Clinical Nutrition* 61 (suppl): 1321s–1323s.

Kuhnlein HV (1989) Culture and ecology in dietetics and nutrition. *Journal of the American Dietary Association* 89: 1059–1989.

Kumanyika S and Morssink CB (1999) Working effectively in cross-cultural and multicultural settings. In: Owen AL, Splett PL and Owen GM (eds) *Nutrition in the Community: The Art and Science of Delivering Services*, pp. 542–565. Boston: McGraw-Hill.

Lyons GK, Woodruff SI, Candelaria JI, Rupp JW and Elder JP (1996) Development of a protocol to assess dietary intake among Hispanics who have low literary skills in English. *Journal of the American Dietary Association* 96: 1276–1279.

Meiselman H (ed.) (2000) *Dimensions of the Meal: The Science, Culture, Business and Art of Eating*. Gaithersburg, Maryland: Aspen.

Nestle M (1995) Mediterranean diets: historical and research overview. *American Journal of Nutrition* 61 (suppl 13): 135–205.

Passin H and Bennett JW (1943) Social process and dietary changes. In: *The Problem of Changing Food Habits*. Bulletin no. 1908. Washington, DC: National Research Council.

Pelto GH, Pelto PJ and Messer E (1989) *Research Methods in Nutritional Anthropology*. Tokyo: United Nations University.

Rozen E (1989) *Flavor Principle Cookbook*. New York: Penguin Books.

Simoons FJ (1990) *Eat Not This Flesh: Food Avoidances from Prehistory to Present*. Madison, Wisconsin: University of Wisconsin Press.

Trichopoulou A and Lagios PH (1997) Healthy traditional Mediterranean diets: an expression of history and lifestyle. *Nutrition Review* 55: 383–389.

Tucker KL, Bianchi L, Maras J and Bermudez OI (1998) Adaptation of a food frequency questionnaire to assess diets of Puerto Rican and non-Hispanic adults. *American Journal of Epidemiology* 148: 507–518.

Williams RL (ed.) (1969) *Foods of the World*. New York: Time-Life Books.

EUROPEAN UNION

European Food Law Harmonization

P Berry Ottaway, Berry Ottaway & Associates Ltd, Hereford, UK

Introduction

With the signing of the Treaty of Rome on 25 March 1957, the European Economic Community (EEC) came into existence with the founder members being the six signatories to the Treaty: Belgium, France, Germany, Italy, Luxembourg, and The Netherlands. With entries at different intervals over the following 25 years, the Community increased in size to 15 members with the addition of the UK, Spain, Greece, Portugal, Denmark, Sweden, Ireland, Finland, and Austria. The enlargement of the Community (now known as the European Union) is expected to continue with an anticipated 25 member states by 2005.

In the Treaty of Rome, emphasis was placed on consumer protection across the Community. Article 39 states *inter alia* that the objective of the Common

Agricultural Policy is to insure 'that supplies reach consumers at reasonable prices' and Article 86 prohibits undertakings from 'limiting production, markets or technical developments to the prejudice of consumers.'

Right from the inception of the Community it was realized that to insure the freedom of trade in foodstuffs there had to be common rules on labeling, food additives, food composition, and food safety. Initially, the emphasis was on food additives, with the first directive being that on coloring matter authorized for use in foodstuffs for human consumption, which was adopted in 1962. This was followed 2 years later by a directive on preservatives, which was adopted in July 1964. Both these directives gave lists of the permitted additives and controls on the use of the additives for certain applications. During this time, work started on directives to control other groups of food additives such as antioxidants, emulsifiers, and stabilizers.

In 1965 the European Commission published its Recommendation 65/428/EEC of 20 September 1965 to the member states, which required prior communication to the Commission, in draft form, of certain laws, regulations, and administrative instructions. This included the notification of intended new legislation or amendments to existing legislation on foodstuffs.

Harmonization Programs 1969–85

The next major step was the harmonization program which was drawn up in 1969 following the Council Resolution 617/69/EEC of 28 May 1969. This included the drawing-up of a program for the elimination of technical barriers to trade in foodstuffs that resulted from disparities between provisions laid down by law, regulation, or administrative action in the member states. The Resolution also introduced the requirement for mutual recognition of inspections.

The program that emerged from this resolution was, in retrospect, too ambitious. It was divided into priority groups with a time schedule of five stages, 70 subjects were to be covered, and the work was to be completed by 1971. Needless to say, the schedule was by no means complete by this deadline and there was another Council Resolution of 17 December 1973 that included a revized calendar for the elimination of technical barriers to trade in foodstuffs. This required completion of the program by 1978. Again, progress was far slower than anticipated and a further revision was adopted in 1977.

One of the main objectives of the harmonization program during the 1970s was the development of compositional or 'recipe' legislation, covering specific groups of foodstuffs. This 'vertical' legislation, as it came to be known, was considered essential for free movement of foods and a large number of directives were envisaged, each concerned with the composition and control of a product category. By the mid-1970s, over 30 categories of foodstuffs had been identified as requiring recipe or vertical legislation.

The first of these directives was that on cocoa and chocolate products, which is still in force, having been amended numerous times. Although work commenced on over 25 vertical directives during the first half of the 1970s, very little of this proposed legislation had been adopted by 1977. Most of the delays were caused by lack of agreement between the member states as to what was important for the control of the product category. The degree of detail envisaged by the European Commission was found to be incompatible with the current practice in each of the member states.

In 1977, one of the main objectives of the revised program was to review the requirements for many of the vertical directives and to stop work on those that were not considered essential to the facilitation of trade. The program required that the European Commission concentrated on legislation that covered all classes of food such as labeling, prescribed quantities, packing by weight or volume, and purity criteria for the additives already covered by directives. A number of vertical directives that were at an advanced stage were also to be finalized. These included the directive on fruit jams, jellies, marmalades, and chestnut purÕe which was finally adopted in 1979, the directive on coffee extracts and chicory products (1977), and the directive on natural mineral waters (1980).

Of the other proposals for vertical directives, one controlling the levels of erucic acid in fats and oils was adopted in 1976. The directive on the methods of analysis for determining the erucic acid content was eventually adopted in 1980. Whilst these two directives were introduced for reasons of food safety, they have traditionally been regarded as belonging to the vertical legislation.

A directive defining edible caseins and caseinates and specifying the chemical standards and labeling requirements for these ingredients was adopted in 1983 after protracted discussions on the proposals. It took another 2 years before the directive on the methods of analysis was agreed and this was followed a year later in 1986 by a second directive laying down the sampling methods for the chemical analysis.

Except for directives on the methods of analysis of coffee and chicory extracts, the sampling and analysis of preserved milk products, and the manufacture of fruit nectars, the only other vertical directive that was

retained was that on the quality of water intended for human consumption.

By 1980, the original 1969 and 1973 programs had been virtually abandoned and only a handful of the proposed vertical directives survived. These directly related to nine commodities or products, discussed above.

Following the 1977 initiative, work was concentrated on the directive on the labeling, presentation, and advertising of foodstuffs which was adopted in 1979 and on the specific directives required by the directive on materials and articles in contact with food which had been adopted in 1976.

During the early 1980s a number of proposed directives were under discussion but very few were adopted. By 1985 only two-fifths of the original 1969 and 1973 harmonization programs had been implemented. The food industry was finding the cost of compliance extremely high, particularly in the implementation of the requirements of the vertical directives that had been adopted.

A major turning point in the progress towards harmonization and an important influence on the revised programs for the 1980s was the European Court of Justice judgment on the 'Cassis de Dijon' case. This concerned a German prohibition on the import of the blackcurrant liqueur Cassis de Dijon from France as the alcohol content was lower than that permitted by German law. The court ruled that the barrier created by the German regulation was incompatible with the Treaty of Rome, and that quantitative restrictions on imports should be prohibited. This judgment provided a general principle for the trade in foods between member states which was that, provided a foodstuff complies with the regulations in force in the member state in which it is manufactured, it should be permitted on the market of any other member state.

The Cassis de Dijon case is now regarded as being pivotal in the major change of approach by the European Council and Commission to the harmonization program. There was a shift away from the vertical directives and it was determined that future Community legislation on foodstuffs should be limited to provisions that are justified by the need to protect public health, provide consumers with necessary information about foods, provide for necessary public controls, and insure fair trade.

The 1985 Initiative

By the early 1980s it was becoming very clear that, despite revised timetables and changes to strategy, the harmonization program was not working and a completely new approach to food legislation was required. The need for a new approach to harmonization was given an additional impetus in 1985 with the publication of a White Paper from the Commission to the Council entitled *Completing the Internal Market*. The White Paper addressed the fact that there had been a significant imbalance between the 'horizontal' food legislation (e.g., labeling and additives), where considerable progress had been made, and the vertical legislation covering composition, in which there had been relatively little progress.

To overcome these problems it was proposed that a concept of food legislation was required which only contained provisions that could be justified as being essential in the general interest. The main principles to be followed were the protection of public health and the assurance of fair trading. The legislation should also insure that consumers were provided with information and protection in matters other than health and that provisions were made for necessary public controls.

The new approach, initiated by the 1985 White Paper, redefined the method of harmonization and replaced the vertical system with a horizontal one. It also attempted to simplify the procedures to reduce the amount of work involved and to speed up the harmonization process. This new approach was to be based on three general principles. In summary these were:

- that there was the need to distinguish between what was essential and required harmonization and what may be left to the mutual recognition of national legislation and standards;
- that Council Directives based on Article 100 of the Treaty of Rome would be restricted to specifying essential requirements for health and safety which would be obligatory in all member states and compliance with these would allow free movement of products within the Community;
- that, whilst the harmonization of standards would be encouraged, the absence of a European standard could not be used as a barrier to trade. In the absence of a European standard the recognition of national standards would be the guiding principle.

The White Paper, and a Commission communication which followed it, defined a new approach in which the vertical system was to be replaced by a horizontal one. The procedures were also simplified to reduce the work involved and to speed up the harmonization process. The new program was to be centered on five framework directives that were considered crucial to the single market in foodstuffs. These were:

- additives
- food labeling and presentation

- official control of foodstuffs
- foods for particular nutritional uses
- materials and articles in contact with food

The new proposals required the European Council to adopt the framework directives that laid down the general principles and essential requirements for health and safety. The European Commission had the responsibility to introduce measures giving the details and technical specifications that would enable the requirements to be achieved. For example, the framework directive on food additives (89/107/EEC) outlines the general principles for the evaluation and control of food additives based on the criteria of safety and need. From this are derived a number of technical directives giving the lists of permitted additives, their conditions of use, and their purity criteria.

A further, and significant, influence on the need for a change of harmonization policy was the publication of a report in 1988 entitled *The European Challenge 1992 – The Benefits of a Single Market*. The research was funded by the European Commission and the chairman of the steering committee of the research group was Paolo Cecchini, hence the popular description of the paper as the Cecchini Report. This substantial report evaluated all the industry sectors, including those involved with food production and manufacture. It pointed out that the foodstuffs sector was the biggest contributor to jobs and most value-added of all the industries in the European Community, and that it appeared to be well placed to confront international competition in the 1990s. However, it then qualified these statements by saying that appearances may be misleading as the trend towards global consolidation is dominated by companies from the USA. With the exception of Unilever and Nestlé, US companies fill the other eight places of the world's top 10 food companies.

The second, and very crucial, proviso was that the Community market for foodstuffs was segmented by a range of trade barriers that appeared to be on the increase. The harmonization program, which by then had been in operation for almost a quarter of a century, had failed in its primary objectives.

The estimation of the financial cost to the food industry which resulted directly from nontariff barriers was in the range of Ecu500–1000 million annually, not counting the inevitable restrictions in consumer choice. The report also estimated that the savings, which would follow the elimination of these trade obstacles, would represent between 2 and 3% of the value-added component of the food sector, which alone accounted for just over 4% of the Community's gross domestic product. The savings were also estimated to equal up to a 2-year gain in the industry's productivity.

The 1990s

Even with the more simplified and streamlined system introduced with the 1985 initiative, progress was slow and much of the proposed legislation was not adopted at the introduction of the Single Market in January 1993, 35 years after the signing of the Treaty of Rome.

Between the end of 1991 and early 1993, five hygiene directives were adopted. Directive 93/43/EEC covered general food hygiene whilst the other four were for specific groups of foods such as milk and milk products; poultry meat; farmed game bird; and rabbit meats and wild game meats. This plethora of hygiene requirements was soon found to be overlapping and a few years later moves had already started to effect a consolidation into a single piece of legislation.

Other legislation in the early 1990s covered organic food production and control, genetically modified organisms, and the protection of designations of origin and geographical indications.

A communication from the Commission to the Council and Parliament on the *Management of the Mutual Recognition of National Rules after 1992* was issued in December 1993. This document listed 20 barriers to trade in foodstuffs which had been identified by member states during the period 1990–92 and for which many of the Community measures planned in the 1985 White Paper were yet to be adopted or to enter into force.

By the end of 1994 the greater part of the proposed harmonization legislation was in place. Two of the important additives directives, those on colors and sweeteners, were adopted in June 1994, and although the directive on additives other than colors and sweeteners reached common position in December 1994, it was only finally adopted in early 1995.

Legislation still outstanding by the middle of 1995 included the proposed regulation on novel foods and the directive on quantity indication (QUID), both of which achieved common position in June 1995. However, further delays meant that both the novel foods regulation and the QUID directive were only finally adopted in January 1997. By this time draft directives to amend the directive on sweeteners and the directive on additives other than colors and sweeteners had also been circulated.

A number of directives required under the framework directive on foods for particular nutritional uses were still outstanding by mid-1995, as were proposals for the fortification of foods and food supplements. A proposal for a directive on food claims was

withdrawn and a new proposal on unit pricing introduced.

During 1994 the Commission received a total of 35 complaints about obstacles to trade in foodstuffs in contravention of Article 30 *et seq* of the Treaty of Rome. An analysis of these complaints showed that the principal obstacle was the national requirements of prior authorization procedures, whether or not they involved the issuing of certificates, designed to control the composition of the foodstuffs, in particular those containing nutrients such as vitamins and trace minerals and/or additives.

Other examples of complaints listed in the Commission's report for 1994 were the requirement that an accompanying note be produced when sugar is imported into Italy (this was in order to control its use in wine), the ban on the use of sorbic acid in milk-chocolate fillings in France, and the requirement imposed by the Spanish authorities that an official health certificate of the country of origin be produced as evidence that products have been manufactured in accordance with the rules in force in the country of origin and that their consumption presents no health risks.

Post-2000

At the end of 1999 it was apparent that the program was still incomplete. Although some directives and regulations had been under discussion for a number of years, they had still not been adopted. These included a directive on sports nutrition products, one on fortified foods, and a controversial directive on food supplements. There were also a number of outstanding technical directives and regulations, such as those on the control of additives used in flavorings, and further controls on the labeling of foods containing genetically modified organisms.

In December 1999 the European Commission published its White Paper on food safety. This reiterated previous statements that the European Union's (EU's) food policy must be built around high standards of food safety that serve to protect, and promote, the health of the consumer. It also stated that the central role of the Internal Market was to insure that consumers were offered a wide range of safe and high-quality products coming from all member states. This principle should also apply to food imported from countries outside the EU.

The White Paper stated that both EU and member-state food safety systems had been under unprecedented pressure during the food and animal feed emergencies in the second half of the 1990s. These emergencies had exposed weaknesses in the systems both at EU and member-state level and the White Paper assessed the requirement for action by the responsible authorities (the Commission, member states, and European Parliament) to reinforce, improve, and further develop existing food safety systems.

The paper outlined the requirement for a farm-to-table policy that included traceability of animal feed, human foods, and their ingredients.

It was also determined that risk analysis must form the foundation on which the future food safety policy should be based. This was developed into the three components of risk analysis, which are risk assessment, risk management, and risk communication. The paper also stated that, where scientific evidence is insufficient, inconclusive, or uncertain, measures based on the 'precautionary principle' should apply.

A new legal framework of the control of human food and animal feed was laid out in the paper. The inclusion of animal feed was on the premise that the safety of food from animal origin begins with safe animal feed.

The Commission proposed that there should be a comprehensive piece of legislation to recast the different food control requirements already in existence. This would take into account the general principle that all parts of the food production chain should be subject to official controls. An annex to the paper gave an action plan on food safety, which included a number of initiatives relating to food and animal feed safety. This plan covered 84 proposals for regulations, directives, and supporting procedures. The timetable given in the plan required that the majority of the legislation would be adopted by the Commission by the end of 2000, with final adoption by the Council and European Parliament by the first half of 2002. By the end of 2001 many had already fallen behind the timetable.

The rationale for the establishment of a European Food Authority was given in detail. The Food Authority was envisaged as having particular responsibilities for both risk assessment and communication on food safety issues.

From early 2000 onwards two subtle changes were taking place in the European food law. One was the development of legislation not directly related to the original harmonization program but dealing with emerging issues. This included detailed legislation on contaminants in foods such as mycotoxins and dioxins.

The other change that has been taking place is the move from directives to regulations for a wider spectrum of food law. This is an interesting move as it effectively speeds up the legislative process. Directives require member states to modify their internal legislation to incorporate the requirements of the directive

within a given period, which can be quite long. Regulations, however, require the member state to introduce the legislation as the unaltered text of the regulation and the time allowed is relatively short, often a matter of months. As a consequence, regulations allow for a more rapid implementation of the law.

General Food Law

One of the most glaring omissions, which had still not been resolved by the end of the 1990s, was the legal definition of a food and a single document giving the principles and philosophy of European food law. Although the precepts of the legal aspects of European control had been introduced into individual directives and regulations, there was no single official document to cover these aspects. For many years the concept of a directive on general food law had been under discussion, with the objective of producing a piece of legislation to clarify the parameters for the structure and operation of European food law.

It was not until 1997 that the European Commission published its Green Paper on food law, which laid down the common principles underlying the food legislation and established food safety as the primary objective of food law. The Green Paper was followed at the end of 1999 by the Commission's White Paper on Food Safety. Chapter 6 of the White Paper advocated a new legal framework for food safety embodied in legislation covering general food law.

Just a year later, at the end of 2000, the Commission published a proposal for 'a Regulation laying down the general principles and requirements of food law, establishing the European Food Authority, and laying down procedures in matters of food safety.' Article 2 of the proposal gave a definition of food for the first time. In essence, this covered any substance or product, whether processed, partially processed, or unprocessed, intended to be, or reasonably expected to be, ingested by humans. It included drink, chewing gum and any substance, including water, intentionally incorporated into the food during its manufacture, preparation, or treatment. Exclusions from the definition included medicinal products, cosmetics, tobacco and tobacco products, and narcotic or psychotropic substances.

As stated in its title, the proposed regulation also contained the requirement for the establishment of a European Food Authority, which would have the objective of providing scientific advice and scientific and technical support for EU legislation in all fields that have a direct or indirect impact on the safety of human food and animal feed.

The mission for the Authority as given in the proposal was the contribution to a high level of protection of human life and health and in this respect to take account of animal health and welfare, plant health, and the environment. This was to be carried out in the context of the operation of the internal market.

The Authority was to comprise a management board, an executive director and staff, an advisory forum, a scientific committee, and permanent scientific panels. The proposal specified eight scientific panels, which would be required to cover the various aspects covered by the Authority's remit. The panels were to be composed of independent scientific experts.

A further part of the proposed regulation specifically related to food safety and covered the requirement for a rapid alert system, emergencies, and crisis management. This initiative was mainly as a consequence of a number of European food scares during the late 1990s such as bovine spongiform encephalopathy (BSE) in cattle and the Belgian and Dutch dioxin crises.

The proposals also included the requirement for full traceability of food, animal feed, and food-producing animals. The responsibilities of food and animal feed operators are given and the concept of the precautionary principle is laid down in law.

These proposals were finally adopted in January 2002 as Regulation (EC) N° 178/2002.

See also: **Legislation**: History; Additives; Contaminants and Adulterants

Further Reading

Berry Ottaway P (1995) *Harmonisation of European Food Legislation*. FT Management Report. London, UK: Pearson Professional.

Cecchini P, Catinat M and Jaquemin A (1988) *The European Challenge 1992 – The Benefits of a Single Market*. Aldershot, UK: Wildwood House.

Commission recommendation of 20 September 1965 concerning prior communication to the Commission in draft form of certain laws, regulations and administrative instructions. OJ 65* 160 of 29 September 1965, p. 2611. Luxembourg: Office for Official Publications of the European Communities.

Communication from the Commission to the Council and to the European Parliament (1985) *The Completion of the Internal Market: Community Legislation on Foodstuffs*. COM(85)603 Final. Brussels.

Communication from the Commission to the Council and European Parliament (1993) *The Management of the Mutual Recognition of National Rules after 1992*. COM(93)669 Final. Luxembourg: Office for Official Publications of the European Communities.

Council resolution of 28 May 1969 drawing up a programme for the elimination of technical barriers to trade in industrial products. OJ69 C 076 of 17 June 1969, p. 1. Luxembourg: Office for Official Publications of the European Communities.

Council resolution of 17 December 1973 on industrial policy. OJ73 C 117 of 31 December 1973, p. 1. Luxembourg: Office for Official Publications of the European Communities.
European Court of Justice (1979) Case 120/78. European Courts Reports p. 649.
Goodburn K (ed.) (2001) *EU Food Law*. Cambridge, UK: Woodhead/CRS Press.
Green Paper on General Principles of Food Law in the EU (1997) European Commission COM(97)176 Final.
Proposal for a regulation of the European Parliament and Council laying down the general principles and requirements of food law, establishing the European Food Authority and laying down procedures in matters of food safety (2000) European Commission COM(2000)716 Final. Luxembourg: Office for Official Publications of the European Communities.
Report from the Commission to the Council and European Parliament (1995) *The Single Market in 1994*. COM(95)238 Final. Luxembourg. Office for Official Publications of the European Communities.
Treaty establishing the European Economic Community, Rome, 25 March 1957. London: HMSO.
White Paper on Food Safety (2000) European Commission COM(2000)1. Luxembourg: Office for Official Publications of the European Communities.
White Paper from the Commission to the European Council (1985) *Completing the Internal Market*. Luxembourg: Office for Official Publications of the European Communities.

Evaporated Milk *See* **Condensed Milk**

EVAPORATION

Contents

Basic Principles

P H Ferguson, Cornwell Products, Edenbridge, UK

Background

Evaporation is the removal of water, or another solvent, from a solution by the application of heat – primarily the latent heat of vaporization – to the solution. Thus, it can be seen that evaporation involves a heat-transfer process that heats a liquid in order to generate a vapor, which in turn has to be separated and, in nearly all cases, subsequently condensed. The theory of evaporation therefore is based upon the principles of heat transfer and, in particular, their application to the boiling (evaporation) of fluids, and the condensation of vapours. (*See* **Heat Transfer Methods**.)

Considering the process economics aspect, the heat-transfer step is important as its efficiency is the factor having the greatest influence on both the capital and the operating costs. The vapor–liquid separation is also critical as the effectiveness with which it is carried out influences the product losses (by way of vapor entrainment), pollution of the condensate steam byproduct, and potential corrosion problems of the condenser surfaces. The subsequent condensation of the evaporated vapor also requires careful consideration, because not only does it have a bearing on the operation of the evaporator system, but perhaps more importantly, it significantly affects the system's operating economy.

Conventionally, the heat required is provided by low-pressure steam (generated from a range of primary fuels), although in practice, this requirement can be minimized by the reuse of vapors from the evaporation process itself. Less common than these indirect heating techniques are direct heating methods (e.g., solar heating of salt brines).

General Principles of Heat Transfer

Heat can be transferred by conduction, convection, or radiation. The last means of heat transfer is not usually significant in industrial evaporation systems.

Heat Transfer by Conduction

The quantity of heat transferred between two points (see Figure 1) is defined as

$$Q = -KA\frac{T}{X}, \tag{1}$$

where Q is the quantity of heat, K is the thermal conductivity, A is the surface area through which heat is passing, T is the temperature difference between the two points, and X is the distance between two points (the negative sign shows that the heat flow is in the opposite direction to the temperature gradient).

Thermal conductivity is a property of the material through which the heat is passing, and can vary from, for example, 15.93 for stainless steel, 0.62 for water, to 0.024 for air, a difference factor of nearly 700. (Units are $J\,s^{-1}\,m^{-1}\,K^{-1}$.) Interestingly, stainless steel (the most common material of construction for food industry evaporators) has one of the lowest thermal conductivities for metals, whilst water has one of the highest thermal conductivities for liquids.

Another useful parameter to consider is thermal resistance:

$$R = \frac{X}{K}. \tag{2}$$

Where we are considering heat transfer through a composite wall (i.e., several layers of different materials), thermal resistances are cumulative in their effect on heat transfer. Hence, if one layer is a material with a high thermal resistance (e.g., an insulating material or fouling scale), the overall heat transfer will be low, even if other layers have a low resistance (e.g., metals).

In the instance of conduction through the wall of a tube (Figure 2), the surface area across which the heat transfer takes place is the tube wall itself. The equation for determining the heat load transferred is:

$$Q = -K2\pi rl\frac{T}{W}, \tag{3}$$

where r is the tube radius, l is the length of the tube, and W is the wall thickness.

Heat Transfer by Convection

Heat transfer by convection occurs as a result of the movement of fluid on a macroscopic scale, in the form of eddy and circulation currents. This convective movement can take two forms:

1. Natural convection: these currents arise from the heating process itself, because of the differences in density between the heated fluid and the surrounding colder fluid, which causes the former to rise, and the latter to sink to take the place of the former.
2. Forced convection: in this type of convection, the currents are created by an external device, e.g., a circulation pump causing turbulent flow in a pipe.

When heat transfer occurs from a surface into the body of a fluid, natural convection currents are weakest at the surface, which is covered by what is effectively a static film. Consequently, heat transfer across this film can only occur by conduction, and, as mentioned above, thermal conductivity in fluids is low. Hence, the main resistance to heat transfer into fluid in a pipe is this film adjacent to the pipe wall. An increase in the velocity of the fluid moving through the pipe will reduce the thickness of this static film and give rise to an overall increase in the heat transfer into the fluid.

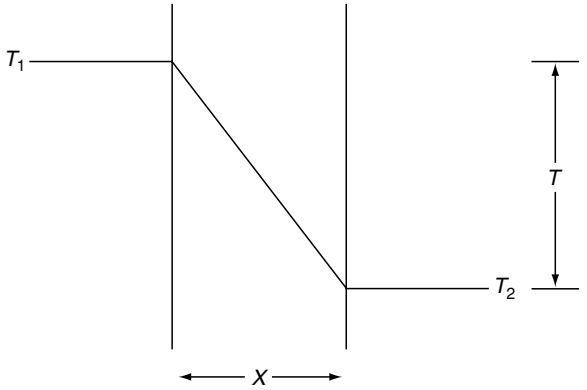

Figure 1 Heat transfer by conduction between two points. Reproduced from Evaporation: Basic Principles, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

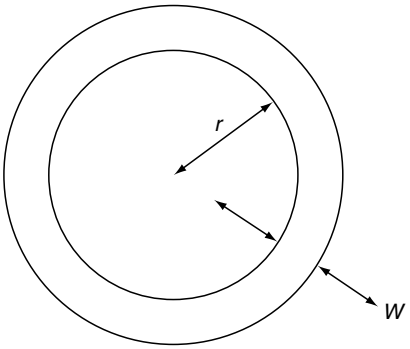

Figure 2 Heat transfer through the wall of a tube. Reproduced from Evaporation: Basic Principles, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

In theory, the heat load transferred across this film is defined as in eqn (1). However, in practice, it is difficult to calculate the film thickness, X, and so the following relationship is used:

$$Q = -hAT, \tag{4}$$

where h is the heat-transfer coefficient.

Thermal resistance is therefore the reciprocal of the heat-transfer coefficient, i.e., $1/h$.

Forced convection is of more importance than natural convection in industrial food processes and equipment, where the fluids are under turbulent flow conditions. It should be noted, however, that evaporators can be classified as either 'natural circulation' or 'forced circulation' systems. In the former case, the eddy and circulation currents described above are greatly enhanced by the currents caused by the rising bubbles generated in the boiling processes.

Whenever possible, streamline flow conditions should be avoided or at least minimized, as the heat transfer coefficients of fluids (in convective heat transfer) are much greater than thermal conductivity factors (in conductive heat transfer). With very viscous fluids (e.g., food pastes and slurries), turbulent flow can only be produced by a high-pressure drop across the heat-transfer device (e.g., by the input of a large quantity of pump energy).

In a tubular heat exchanger (e.g., an evaporator), where perhaps one fluid is flowing inside a pipe and being heated (or cooled) by another outside, the consideration must be of heat transfer both inside and outside the tube. In the latter case, flow can be either lengthwise along the tube (in either direction) or at right angles to the single tube or tube bundle.

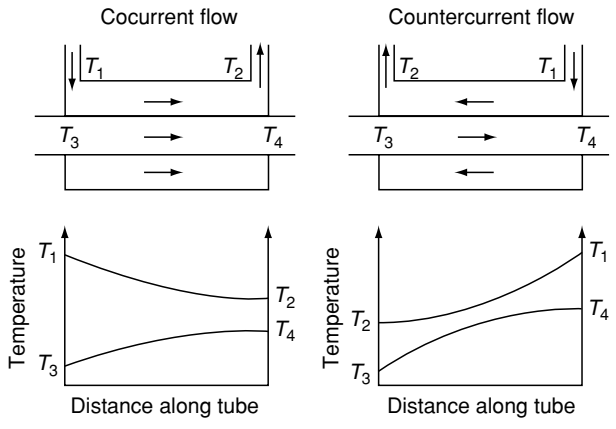

Figure 3 Logarithmic mean temperature difference. Reproduced from Evaporation: Basic Principles *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Also, those fluids passing along the length of the heat exchange tube (either inside or outside) will experience either an increase or decrease in temperature, obviously because of heat transfer. This means that in order to correctly quantify this heat transfer, it is first necessary to define the difference in temperature between the fluids, given that it will not only vary at different points along the length of the tube, but will also vary depending on whether cocurrent or countercurrent flow is being employed. It is necessary to calculate an 'average' value for the temperature difference, and the factor normally employed is the logarithmic mean temperature difference (LMTD):

$$\text{LMTD} = \begin{cases} \dfrac{(T_1 - T_3) - (T_2 - T_4)}{\log_e (T_1 - T_3)/(T_2 - T_4)} & \text{(cocurrent flow)} \quad (4a) \\ \dfrac{(T_2 - T_3) - (T_1 - T_4)}{\log_e (T_2 - T_3)/(T_1 - T_4)} & \text{(countercurrent flow),} \quad (4b) \end{cases}$$

where T is the temperature (see Figure 3 for a definition of subscript numbers).

Heat-transfer Aspects of Evaporation and Boiling

As already stated, evaporation is the removal of water (or other solvent) from a solution by boiling the liquor in a suitable heat exchanger (i.e., the evaporator) and removing the resultant vapor.

In boiling liquids (e.g., inside an evaporator tube), the heat-transfer coefficient (h) depends very much on the temperature difference (T) between the heating surface and the boiling liquid.

Boiling can be classified into three types: interface evaporation, nucleate boiling and film boiling (Figure 4).

Interface Evaporation

Here, relatively few bubbles are formed on the heating surface, and they move to the vapor–liquid interface by natural convection, causing negligible agitation in the liquid. Full boiling is often suppressed owing to the hydrostatic head of the liquid above the heating surface. This phenomenon is particularly likely at the bottom of a vertical evaporator tube.

The heat-transfer coefficient increases slightly with increasing temperature difference.

Nucleate Boiling

Here, a large number of bubbles form on the heating surface, and their escape to the vapor–liquid interface

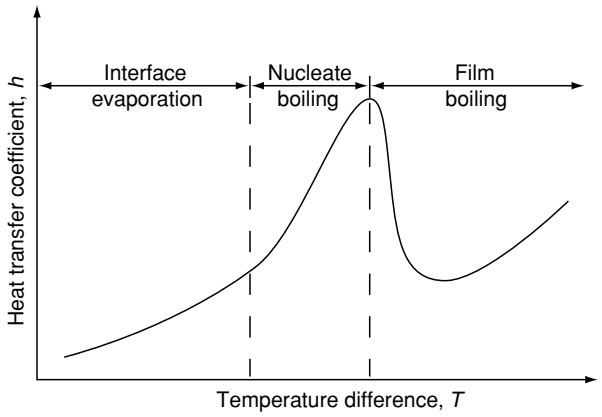

Figure 4 Types of boiling. Reproduced from Evaporation: Basic Principles, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

brings about significant turbulence in the liquid. This in turn means that the heat-transfer coefficients are greater, and rise more rapidly with increasing temperature difference, than in the instance of interface evaporation. The maximum heat-transfer coefficient is reached at the top end of this boiling category.

Film Boiling

Film boiling occurs at higher temperature differences, when the quality of vapor being generated at the heating surface is so great that the bubbles start to coalesce to form a vapor film over the heating surface. This in turn means a large drop in heat-transfer coefficient as the latter is obviously much lower for vapors than for liquids, and therefore the quantity of heat transferred falls in proportion.

The critical temperature difference at which the heat transfer coefficient is at its maximum depends on the physical properties of the boiling fluid and also on the geometry of the heating surface.

With even further increases in temperature difference, the heat-transfer coefficient eventually starts to rise again, as heat transfer by radiation (across the vapor film) starts to become a significant part of the total heat transferred. These very high temperature differences are rarely found in practice, however, and most evaporation systems are designed to operate at, or just below, the temperature difference at which the heat transfer coefficient becomes a maximum.

Boiling-point Rise

In addition to the nature of the boiling regime present, another characteristic of evaporation systems is the boiling-point rise (BPR). The boiling point of a solution (e.g., a liquid foodstuff) is always higher than that of pure water, and the magnitude of this temperature elevation is proportional to the concentration of solids in the solution – hence, as the solids increase during the process of evaporation, the boiling point will increase. By way of example, the BPR of milk is only 0.5 °C at 16% total solids, rising to 2 °C at 50% and 5 °C at 80% total solids.

The consequence of this phenomenon is that the temperature difference between the heating medium (e.g., steam at 110 °C) and the boiling liquid (e.g., milk at 50% total solids; BPR = 2 °C; boiling point = 102 °C at atmospheric pressure) is less (110 °C − 102 °C = 8 °C) than in the case of pure water (110 °C − 100 °C = 10 °C). The 20% drop in the temperature difference, from 10 °C to 8 °C, means a pro rata drop in heat transfer, as it is directly proportional to the temperature difference. The resultant vapor will be at the temperature of the boiling liquid and hence will be superheated by the amount of the BPR (i.e., 102 °C − 100 °C = 2 °C).

In practice, the sensible heat associated with this superheating is often harder to recover than the latent heat itself and effectively is partly or totally lost, causing a loss in energy economy.

Heat-transfer Aspects of Condensation

When a saturated vapor is brought into contact with a cooler surface, and falls below its dew point, vapor will condense on that surface, giving out its latent heat of vaporization.

This condensation can occur in two ways:

1. Dropwise condensation: here, the condensing vapor forms a large number of small individual drops on the surface, which do not wet the surface but, after growing slightly, fall away and hence expose fresh surface for further similar condensation to take place on. This type of condensation can result in very high heat-transfer rates, as heat does not have to flow through a film of liquid by conduction. However, dropwise condensation is rarely found in practice and in fact has only been definitely demonstrated with pure steam, under specialized conditions.
2. Film condensation: in this form of condensation, the condensing vapor forms a continuous film on the cooler surface, and subsequent condensation actually takes place on this film of liquid. If the surface is vertical, the condensate film flows downwards, the flow being predominantly streamline, and hence heat-transfer will be by conduction. It can be assumed that the temperature of the film at the cool surface is equal to that of the surface itself, and the film temperature at the vapor surface is at the same temperature as the vapor.

Another aspect of vaporization to consider is that, in most evaporation systems encountered, the vapors will contain a small quantity of noncondensable gas (e.g., air that has either entered the product or via leaks into the main body of the evaporator, most of which will often be operated under varying levels of vacuum). This mixture of gas and vapor, when cooled in the condenser, will give rise to a layer of condensed vapors only, on the heat exchange surfaces of the condenser. Heat transfer to this surface will, therefore, be not only by the vapor condensing, and hence releasing its latent heat, but also by the transfer of sensible heat from the noncondensing gas. The design of the condenser will need to take this factor into account.

See also: **Heat Transfer Methods**

Further Reading

Coulson JM and Richardson JF (1970) *Chemical Engineering*, vol. 1, 2nd edn. Oxford: Pergamon Press.

Kay JM (1968) *An Introduction to Fluid Mechanics and Heat Transfer*, 2nd edn. Cambridge: Cambridge University Press.

Perry RH and Chilton CH (1973) *Chemical Engineers' Handbook*, 5th edn. New York: McGraw-Hill.

To LC, Tade MO and Kraetzl M (1999) *Robust Nonlinear Control of Industrial Evaporation Systems*, 340 pp. London, UK: World Scientific.

below its boiling point (**Figure 1**). As the outside of the tube is heated by steam, the liquid rises in temperature, and its density falls, causing it to rise. As it passes up the tube, its temperature reaches boiling point, and small vapor bubbles appear in the liquid. This causes a greater drop in the overall density, causing the liquid–vapor mixture to rise further, and as the liquid continues to pick up further heat, additional vapor is produced, and more bubbles form, until the volume of vapor is greater than that of the liquid. Eventually, at the top, only vapor and liquid mist occurs.

It is possible to classify this overall flow regime into five stages:

1. Single-phase flow: at the base of the tube, where the liquid enters – no bubbles present.
2. Two-phase flow: liquid predominates, but contains a swarm of small vapor bubbles.
3. Plug flow: liquid plus large vapor bubbles (formed by coalescence of small bubbles).
4. Annular flow: large vapor bubbles coalesce to occupy central core of tube, and the liquid forms a film on the inner wall of the tube.
5. Mist flow: at the top end of the evaporator – the liquid film breaks up, and the remaining liquid appears as droplets in the flow of vapor out of the top of the tube.

This type of evaporator is also known as a natural-circulation evaporator.

Uses in the Food Industry

P H Ferguson, Cornwell Products, Edenbridge, UK

Background

This article deals with the main types of evaporator used in industry and how to select the correct configuration, and finally, a range of applications in the food, dairy, and related industries are considered.

Types of Evaporator

Primarily, evaporators are characterized according to the manner in which the liquid phase circulates.

Climbing-film Evaporator

This is the traditional form of vertical tube evaporator where liquid to be evaporated enters the evaporator tube at the bottom, at a temperature

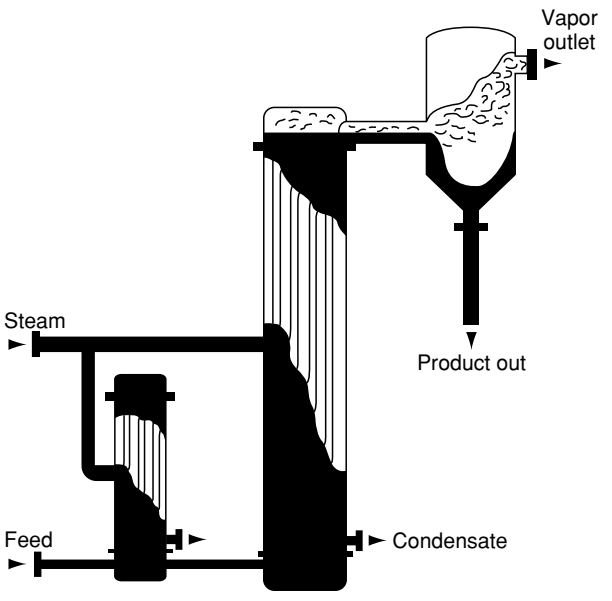

Figure 1 Climbing film evaporator. Reproduced from Evaporation: Uses in the Food Industry, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Forced-circulation Evaporator

This form of evaporation plant (Figure 2) is employed with liquids which, for technical or economic reasons, cannot be concentrated with natural circulation evaporators. These are very viscous fluids, liquors that cause scaling and fouling problems (exhibiting inverse solubility/temperature characteristics), or corrosive liquids. The disadvantages are higher capital and operating (i.e., pumping) costs than for film evaporators.

In this liquid flow configuration, product is pumped at high velocity through a heat exchange section where boiling is prevented by the application of a hydrostatic pressure: this absence of boiling greatly reduces the scaling on heat-exchange surfaces. Boiling subsequently takes place in the separator.

Flash evaporators are effectively special cases of forced-circulation evaporators, where boiling is suppressed artificially by means of an orifice or valve, and the product is flashed through into a separator. Higher temperature rises are therefore possible through the heat exchanger, giving reduced circulation rates and pumping costs. However, this also reduces the suitability of the plant for products containing substances with inverse temperature/solubility characteristics. Plate heat exchangers are often used for flash-evaporation duties.

Forced-circulation evaporator configurations are often used as crystallizers – the alternative being the cooling evaporator. In the evaporative crystallizer, the conditions are approximately isothermal, and supersaturation is achieved as a result of the removal of solvent. Evaporative crystallization must of course be used where the solubility shows little variation with temperature.

Falling-film Evaporator

Compared to the falling-film evaporator (Figure 3), the energy economy of the climbing-film system remains relatively poor. The residence time of the liquid is greater, as are the associated temperatures (and temperature differences), in the climbing-film units, and hence they are rarely found in the food and related industries (e.g., for use with heat-sensitive products). Other advantages often found in the much more common falling-film arrangement are easier control, simpler construction, and greater ease of cleaning.

In a falling-film evaporator, the liquid – as the name implies – enters the top of the tube, and passes down to the bottom, as a film. This latter characteristic is effected primarily by insuring that the liquid, as it enters the top of the evaporator, is superheated so that, as it passes into the top section of the tube, vapor is flashed off, filling the central core of the tube and thereby holding the liquid as a film against the tube wall. Another positive difference, in terms of energy economy, between the falling- and climbing-film evaporators, is that the former type – because its liquid feed is superheated – does not have part of the tube surface employed for preheating purposes (i.e., to bring it to the boiling point). It is more

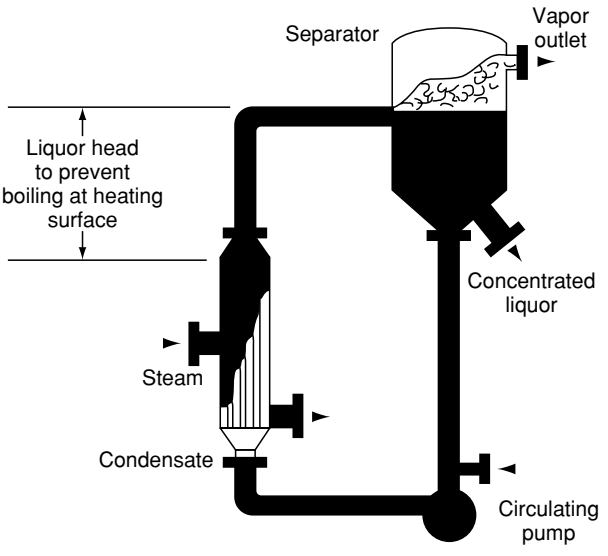

Figure 2 Forced-circulation evaporator. Reproduced from Evaporation: Uses in the Food Industry, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

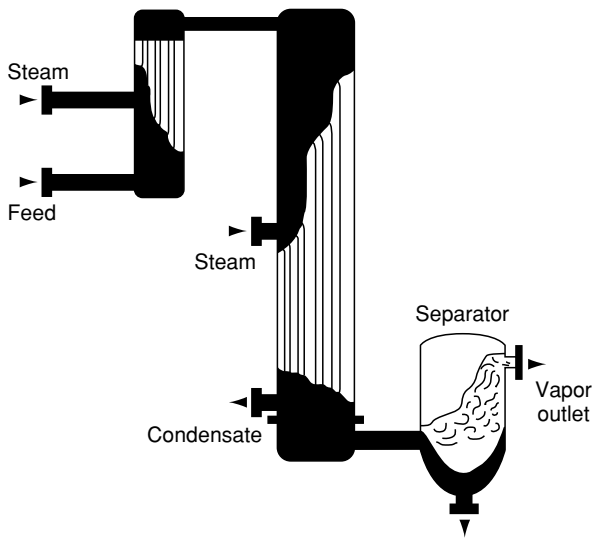

Figure 3 Falling film evaporator. Reproduced from Evaporation: Uses in the Food Industry, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

cost-effective to preheat the liquid in a separate heat exchanger. This can be either a stand-alone system or, more likely, one that is integrated into the overall evaporator plant – for example as a preheating coil located in the steam jacket around the main tube.

Falling-film evaporators in the food industry normally operate under vacuum, thus reducing the product boiling point and minimizing heat damage.

Other Types of Evaporator

The thin-film (or wiped-film) evaporator can be used for very viscous, scaling, or crystallizing materials. The tubular body of this evaporator is fitted with a central rotor, which has blades along its length. These blades almost touch the tube wall (with a clearance of typically about 1 mm) so that, when they rotate during the normal operation of the evaporator, a very thin and well-agitated film of liquid is formed on the surface of the tube wall. In some equipment models, the outer edge of the blade is hinged, so that this edge is forced by centrifugal action on to the wall, and consequently, the thickness of the film so produced is determined by the balance between this centrifugal force and the hydrodynamic forces produced in the liquid film on which the blade rides. This version of the thin-film evaporator is usually known as a swept-surface evaporator. Another configuration of this evaporator group is the scraped-surface evaporator. Here, as the name suggests, the blades (made from a synthetic material or perhaps rubber) actually scrape against the wall. Owing to the relatively high capital cost of this evaporator type, it has relatively limited applications.

Strictly speaking, most driers are forms of evaporators, where the evaporation process is taken to the extreme where effectively all the solvent is removed, leaving the dissolved solids in dry form.

As mentioned at the beginning of this article, evaporator classification is usually based on the way in which the liquor circulates. However, it is also possible to classify evaporation plants according to the mechanical configuration, the most common types being tubular or plate. Although tubular systems are most widespread, plate systems have certain attractions. As the plant height is relatively low, they can operate in process areas where there is limited headroom. Hence, building and installation costs can be modest. It is also often possible to increase their heat-transfer surface at relatively little additional expense. (*See* **Drying**: Theory of Air-drying.)

Evaporator Selection

Once the need for an evaporator for a particular product application has been established, it is necessary for a specification to be prepared, which will be obviously helpful to both purchaser and supplier to identify plant configuration and component selection. In carrying out this exercise, the following points should be borne in mind:

1. Throughput:
   - hourly rate and/or daily rate;
   - length of production run (hours/days).
2. Feed specification:
   - total solids (or equivalent, e.g., degrees Brix);
   - other relevant parameters (e.g., suspended solids, pH, volatile components).
3. Product specification:
   - temperature;
   - total solids;
   - other requirements or characteristics (e.g., vis cosity).
4. Utility requirements:
   - steam, electricity, air, cooling/chilled water (in cluding temperature);
   - existing on-site availability, and proximity to proposed evaporator location.
5. Plant dimensions, building requirements (or outdoor location), foundation loadings, installation aspects (e.g., access for cranage, etc.).
6. Plant access for operation and cleaning (e.g., ladders, platforms, inspection ports) and maintenance (e.g., extra headroom and/or additional space for unhindered removal of individual plant components). Manways should be easy to use (power operated, if necessary).
7. All cleaning requirements ('cleaning in place' (CIP) and also manual cleaning operations – both routine and emergency) must be specified. Additional site facilities identified and procured as necessary.
8. Effluent loadings and associated treatment facilities quantified.
9. Specify individual components (e.g., choice of pumps, electronic controls, etc.). Materials of construction and surface finishes confirmed.
10. Manufacturing and equipment standards to be laid down. Plant testing (e.g., pressure and vacuum). The need for acoustic and thermal insulation should be examined.
11. Control and automation to be specified, including links to adjacent plant, e.g., drying equipment.
12. Obtain recommended spares list, and order as required.
13. Determine pretreatment process which might affect evaporation process. Identify storage/buffer tanks needed, for both feedstock and end product.
14. Identify training requirements, including operator manuals, maintenance schedules, plant

drawings, etc. Plant inspection, delivery, installation and commissioning activities, and arrangements for plant handover.

Applications

A list of established evaporation duties in the food and related biological industries would include the materials listed in Table 1. A few of the applications in Table 1 are considered in more detail in the remainder of this article.

Dairy Products

-Milk From the first recorded use of evaporation, as a means of enhancing the shelf-life of milk, batch vacuum pans were almost universally employed up until the 1920s. However, after stainless steel came to replace copper and aluminum as the material of construction, more sophisticated evaporator configurations (requiring CIP techniques, in turn employing stronger chemicals – now feasible with the new material of construction) became possible. The natural circulation (i.e., rising film) evaporators that replaced the earlier pans were mainly tubular systems and were used up until the early 1950s, when falling film evaporators became more common, and eventually ousted the older styles of evaporator. All three types of plant employed a vacuum, so as to lower the boiling point of the milk, and hence minimize thermal damage (and burn on). Single-stage plants had given way to multieffect units before 1940. All modern-day milk evaporators are multieffect plants, and they almost always employ either thermal (TVR) or mechanical (MVR) vapor recompression. (See below for further details of vapor recompression systems.)

For an example of a typical installation, consider a multiple-effect TVR unit. Prior to the milk entering the first effect, it is usual practice for the milk to be pasteurized. This is generally effected by ensuring that the milk-feed line, which passes through each of the jackets so as to preheat the milk (final effect jacket first, first effect last), goes through a special section in the first effect jacket. Here, the temperature is boosted by extra steam. After regenerative cooling, the milk then enters the first effect or is diverted for additional heat treatment (for example, holding at elevated temperatures, as required for milk to be used for high-heat powder production). The milk then passes through each of the calandria in turn, its concentration obviously increasing (and boiling

Table 1 Materials in the food and related biological industries that require evaporation in their manufacture

Dairy products	*Food products*	*Other biological liquids*
Baby foods	Agar	Amino acids
Buttermilk	Beef stock	Calcium formate
Fat-enriched milk	Beer	Corn steep liquor
Icecream mix	Brewers' yeast	Distillery pot ale
Reconstituted milk	Chestnut extract	Distillery spent wash
Skim milk	Chicken stock	Fish stick liquor
Sweetened condensed milk	Clam broth	Formic acid
Whey	Coffee extract	Gelatine
Whole milk	Dextrose	Gluconic acid
Yogurt milk	Glucose	Glycerine
	Golden syrup	Hydrolyzed protein
Fruit juices	Hopped wort	Itaconic acid
Apple	Invert sugar	Lactone solution
Apricot	Jam	Lignin
Blackcurrant	Jelly	Liquorice water
Cherry	Lactic acid	Methanol extracts
Grape	Malt extract	Molasses effluent
Grapefruit	Meat broth	Monosodium glutamate
Lemon	Meat extract	Pentaerythritol
Lime	Soy sauce	Phenol formaldehyde resin
Mandarin	Sucrose	Tannin extract
Orange	Tea extract	Wheat starch effluent
Passion fruit	Turkey stock	
Pear	Vegetable extract	
Pineapple	Whole egg	
Plum	Yeast cream	
Raspberry	Yeast extract	
Strawberry		
Tangerine		
Tomato		

temperature decreasing) in each subsequent effect. The last effect is often termed the 'finisher' and operates at a higher temperature (and temperature difference) than the penultimate effect, due to an additional steam supply. One or two thermal vapor recompression units are used. The internal design of the finisher is usually configured to handle the higher viscosity associated with the end product.

Alternatively, an MVR system could be employed. Depending on the overall plant capacity, and final product solids, the system would typically comprise a main calandria, where most of the water would be removed, followed by a finisher (possibly TVR), to remove the remainder of the water. (*See* **Milk**: Processing of Liquid Milk.)

-Whey Whey is often characterized by its level of acidity, and hence can be either low-acid 'sweet' whey (from hard or semihard cheese) or high-acid 'acid' whey (from soft cheese, e.g., cottage cheese).

The whey (typical 6% total solids) is normally evaporated after pretreatment to remove fat (and/or cheese fines) up to between 40 and 60% total solids, the choice of the final product solids depending on subsequent processing needs. If the whey concentrate has to be transported, 40% is preferred, to avoid crystallization taking place *en route* and causing unloading difficulties. Alternatively, when the whey is to be spray-dried (on-site), 60% is better. Both TVR and MVR falling-film equipment can be used, the latter configuration usually giving – in this application as well as others – better energy economy. (*See* **Whey and Whey Powders**: Production and Uses.)

Fruit Juices

-Citrus juices High-quality orange juice has been concentrated to between 65 and 75% total solids. Some types of comminuted orange can only be evaporated to about 45%.

Grapefruit juice is typically concentrated to 65% total solids, whilst lemon and lime juices are processed to 45 and 40% total solids, respectively. At these final evaporator concentrations, color quality is maintained intact.

In order to maximize flavor retention, an aroma recovery system should be incorporated into the plant design.

-Apple juice Most apple juice plants concentrate to 72–75% total solids. Aroma recovery should be used to ensure maximum retention of the original fruit flavor characteristics.

-Tomato juice It is possible to use film-type evaporators to preconcentrate tomato juice up to approximately 20% total solids. To obtain higher solids, a swept-surface unit can be considered.

-Other juices Typical concentrations that are possible from commercial evaporation systems are blackcurrant 65–70%, grape 70–75%, pineapple 65%, raspberry and strawberry 70–75%.

Food and Related Products

-Starch industry The main evaporator duties in the corn starch industry are associated with the concentration of corn steep liquor (e.g., for animal feed and antibiotic production medium purposes). Because corn steep liquor is primarily a byproduct, it contains many ingredients, including mainly sugars and proteins, and their derivatives. It shows, therefore, relatively high fouling characteristics, particularly at higher concentrations. Its evaporation requires the use of a plate or tubular falling-film evaporator for the lower concentration stages, and a forced-circulation arrangement for the final effect. Similar byproduct recovery systems, using evaporation technology, are employed in the wheat starch industry. (*See* **Starch**: Sources and Processing.)

-Sugar industry The beet and cane sugar industries use evaporation to concentrate the dilute sugar steam, as extracted from the beet or cane, up to the final concentration prior to crystallization, itself a form of evaporation. Such systems are typically multieffect tubular systems, incorporating vapor recompression. (*See* **Sugar**: Refining of Sugarbeet and Sugarcane.)

The production of sugar syrups (glucose, maltose, fructose) also requires evaporation: the raw syrup (often a product of the acid, alkali, or enzymatic hydrolysis of starch) is typically generated at 20–30% total solids and will need concentration up to 70–80% final solids. The evaporator is almost always a falling-film unit, incorporating either thermal or mechanical vapor recompression. Unlike the byproducts of starch processing (see above), sugar syrups show very low fouling factors on evaporation.

-Gelatin industry The current method is to employ a multiple-effect TVR or single-stage MVR (with or without finisher). The feed material will usually be a dilute gelatin material at around 5% total solids. It will be concentrated to typically 20% total solids, at which point, there is an option of sterilizing the product by live steam injection (at 135–145 °C) and holding the product for up to 8 s. Flashing into a vacuum chamber cools the product down to 100–105 °C, and then the gelatin is further concentrated to a final level of 30% total solids or thereabouts, usually by flash evaporation.

-Lactic acid Food-grade material is usually generated by fermentation, and the resultant acid – after primary removal of impurities – will be at around 10% total solids. This can be concentrated by evaporation to 80–82% total solids. A typical installation, operating at a feed rate of 5 tonnes per hour, consists of three effects (including thermal vapor recompression) and a finisher unit. The boiling temperatures in these three effects are, respectively, 87, 75, and 50 °C, and 50 °C in the finisher. Grade 316 stainless steel is required due to the acidity of the product.

-Beverage industry Most fermented beverage industries that employ distillation to produce a concentrated alcoholic product (e.g., whisky, gin, cognac) generate large volumes of low-alcohol 'waste' streams, i.e., the remaining fraction after the alcohol has been removed by distillation.

The Scotch whisky industry produces two effluent streams: 'pot ale' from malt distilleries and 'spent wash' from grain distilleries. Although these materials closely resemble each other, spent wash is normally more difficult to evaporate to high solids due to a higher level of suspended solids and a greater degree of fouling.

An example of an evaporator to handle, say, 60 tonnes of feed stock (pot ale) per hour at 4% total solids, and concentrate to a final value of 50% total solids, would be configured as follows: a double-effect falling-film MVR system (with turbine fans to recompress the vapor, and optionally a double calandria for the first effect to handle the volume of feed material), supplemented by a forced-circulation finisher. The feed temperature depends on prior handling and storage. The boiling temperature in the two MVR effects would be, respectively, 96 and 91 °C. The finisher would operate at 91 °C also. The main fans would draw around 500–550 kWh, which, together with other motors on the evaporator, would give an overall power consumption of some 12 kWh per tonne of feed stock. TVR systems are still used; however, most new installations incorporate MVR.

Other whisky distilleries (in Ireland and the USA) yield similar byproducts, which can be evaporated in a similar manner.

The neutral spirits industry (gin, vodka, etc.) generates byproducts of a related composition, albeit higher in inorganic salts (e.g., calcium sulfate), which have inverse solubility curves, leading to increased levels of fouling at higher concentrations. Again a forced-circulation layout will be the preferred design for the finisher.

Effluent from the production of cognac requires pretreatment to remove tartaric acid and tartrates before evaporation can be carried out.

Energy Economy

The vapor removed in a single evaporator stage (effect) basically equals the amount of steam fed to that effect. The temperature of the evaporated vapor is, however, lower than the temperature of the heating steam and cannot therefore be used as such to reheat the same stage; instead, the vapor in such a system is led to the condenser, where the remaining heat content is lost. The specific steam consumption (i.e., steam used/vapor removed) of a single-effect evaporator is thus unity.

The specific steam consumption can be halved by the addition of a second effect, which is heated by the vapor from the first effect but is operated at a lower temperature because the heating medium is at a lower temperature.

Addition of further effects thereby reduces (i.e., improves) pro rata the specific steam consumption. Although it would be possible in theory to keep adding effects *ad infinitum* so as to reduce the specific consumption to an almost negligible level, in practice, this possibility is limited for two reasons. First, the increase in capital cost associated with adding extra effects has to be balanced against the reduction in operating cost (i.e., steam); for example, in the dairy industry, it was found that additional effects over and above seven did not pay for themselves in terms of savings in operating costs. Second, between each effect, there is a drop in temperature of perhaps 5 °C (because of the above-mentioned difference in temperature between heating medium and vapor driven off), and coupling this with the fact that, for most food and related products, there is a maximum and minimum evaporation temperature (for reasons of thermal and bacteriological stability, respectively), there is thus another limit to the maximum number of effects permissible in industrial practice.

An alternative method of improving the energy consumption is by the use of vapor recompression. This technique involves taking the product vapor from the effect, compressing it to increase its temperature, and then returning it for reuse in that same effect as the heating medium; the vapor condenses as a result and is subsequently led away for disposal. However, in the process, the specific steam consumption for that effect has been nominally halved. It must be remembered, of course, that additional energy is required to recompress the vapor, but this is small when compared to the energy content of the compressed vapor generated.

The vapor can be compressed either thermally or mechanically: in practice, this means, respectively, either by fresh, high-pressure steam or by a form of mechanical compressor. Where steam is employed, the

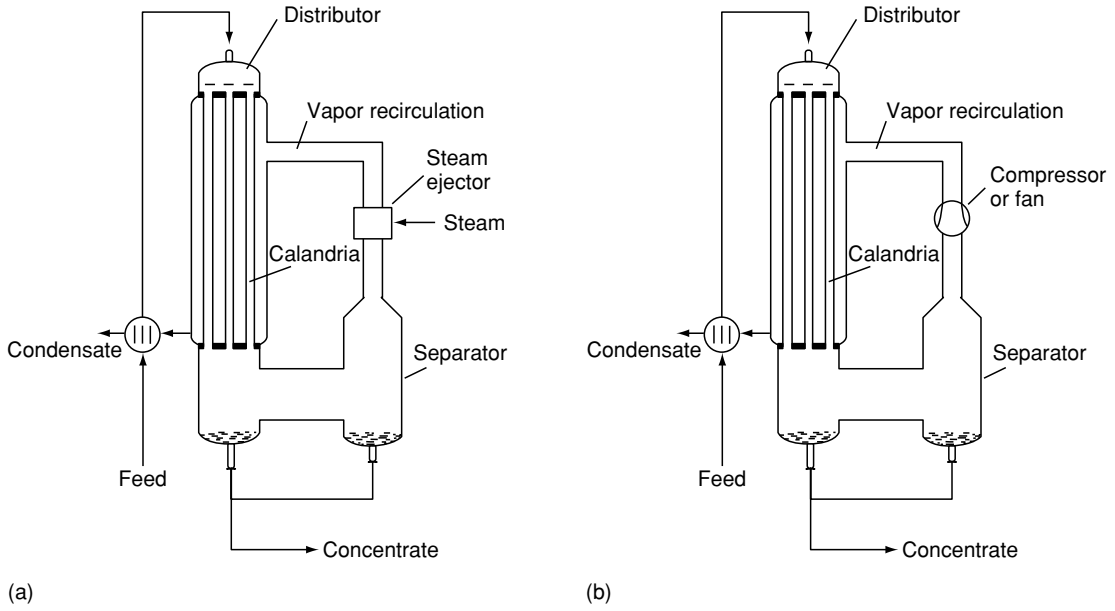

Figure 4 Main components of vapor recompression evaporators: (a) TVR and (b) MVR. Reproduced from Evaporation: Uses in the Food Industry, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

system is known as TVR, whilst mechanical systems are known as MVR. This latter system will usually involve a fan-based compressor rather than a conventional centrifugal-type unit, but individual applications and circumstances will dictate the final choice.

The relatively small additional energy costs required for this vapor recompression are hence in the form of either extra steam or extra electricity costs. The main components of a vapor-recompression evaporator are shown in Figure 4.

As can be seen, once the evaporation process has been brought up to a steady-state operating condition (that is, from start-up), no further addition of primary heating steam is needed (i.e., excluding steam to the steam ejector, for thermal vapor compression). Evaporation is brought about solely by recompressing and 'recycling' vapor separated from the product. In practice, however, heat losses – and often the requirements of (steady-state) control systems – mean that a small but finite quantity of additional steam is required.

Energy Consumption

The power demand of the compressor depends on the vapor quantity and the pressure difference (equating to the temperature difference, T). This means that the power demand is inversely proportional to the heat-transfer surface of the evaporator. When the heat-transfer surface of the evaporator is increased, the T required and the power demand decrease.

With evaporating temperatures between 50 and 70 °C, the power demand of the compressor can be generally calculated using the following equation:

$$P = CMT, \qquad (1)$$

where P is the power demand (kWh^{-1}), M is the vapor mass flow (tonne h^{-1}), C is a constant, which depends mainly on the size of the compressor and is usually between 2.5 and 3.0 (kW $tonne^{-1}$ $°C^{-1}$), and T is the temperature difference across the compressor.

Usually, the T is between 3 and 8 °C, which corresponds to a specific power consumption of between 8 and 20 kW per 1000 kg of water evaporated.

See also: **Drying**: Theory of Air-drying; **Milk**: Processing of Liquid Milk; **Starch**: Sources and Processing; **Sugar**: Refining of Sugarbeet and Sugarcane; **Whey and Whey Powders**: Production and Uses

Further Reading

Coulson JM and Richardson JF (1976) *Chemical Engineering*, vol. 2, 2nd edn. Oxford: Pergamon Press.

Kay JM (1968) *An Introduction to Fluid Mechanics and Heat Transfer*, 2nd edn. Cambridge: Cambridge University Press.

To LC, Tade MD and Kraetzl M (1999) *Robust Nonlinear Control of Industrial Evaporation Systems*, 340 pp. London, UK: World Scientific.

EXERCISE

Contents
Muscle
Metabolic Requirements

Muscle

S C Dennis and T D Noakes, University of Cape Town Medical School, Observatory, South Africa

Introduction

Muscles in the human body can be broadly classified into three main types: smooth, cardiac, and skeletal. Smooth muscles surround many of the body's internal organs and control such functions as the motility of the gastrointestinal tract and the flow of blood through a vascular bed. Cardiac (heart) muscle regulates the rate at which blood is pumped around the body. Skeletal muscles account for around 40% of body weight and are responsible for posture and movement. In this article, attention is focused on the physiology of skeletal muscle.

Skeletal Muscle Organization

Unlike smooth and cardiac muscles, which contract spontaneously and are regulated by hormones and electrical messages from the subconscious regions of the brain, skeletal muscle fiber cells are under voluntary control. With the exception of certain reflexes, skeletal muscle fibers contract only when an electrical signal is sent from the motor cortex of the brain to the muscle via the spinal and motor nerves.

Motor Units

The response of a skeletal muscle to electrical impulses from the brain depends on the motor nerves used to send the signal to contract. When signals to contract are sent via motor nerves that supply as few as 5–15 muscle fibers, tension is adjusted in small increments, and movements are precise. However, when signals to contract are delivered to the muscle by motor nerves that supply several hundred muscle fibers, large increases in tension lead to strong, but less precise, contractions.

Information on the contraction of the skeletal muscle motor units is relayed back to the central nervous system from stretch receptors in the muscle spindles and tension receptors in the Golgi tendon organs. This feedback allows the initial movement to be refined by either reducing motor unit activation or recruiting additional motor units.

Muscle Fiber Types

The small to intermediate motor units in skeletal muscle are generally recruited first during exercise. They are comprised of either red slow-twitch (type I) or red fast-twitch (type IIa) muscle fibers, which are relatively resistant to fatigue. Type I and IIa muscle fibers contain high concentrations of a red oxygen transport protein called myoglobin and obtain most of their energy for contraction from oxidative processes in structures called mitochondria.

In contrast, large motor units are used only when either a very heavy weight has to be lifted or an explosive movement is required. These units are comprised of white fast-twitch (type IIb) muscle fibers, which contain few mitochondria and little myoglobin, and fatigue rapidly. White muscle fibers are designed to power escape reactions and, as will be described later, they generate energy for contraction via an oxygen-independent pathway which maintains maximum force development for around 20 s.

Skeletal Muscle Ultrastructure

When cross-sections of skeletal muscle are viewed under a microscope, type I, IIa, and IIb fibers are found to be arranged in parallel and to be mixed randomly. Ratios of types I:IIa:IIb range from 51:41:8% in untrained persons, to 78:19:3% in elite endurance athletes. (*See* **Cells**.)

Myofibril Arrangement

Within the different muscle fibers, there are several hundred to several thousand smaller parallel fibers known as myofibrils. Surrounding each myofibril is a lattice-like network of tubules, which is part of the sarcoplasmic reticulum. As will be described, the sarcoplasmic reticulum is involved in the initiation and termination of contraction.

Interspersed amongst the myofibrils are columns of small semicircular mitochondria which provide

energy in the form of adenosine triphosphate (ATP). Between the mitochondria are scattered glycogen granules and triglyceride droplets. The muscle triglyceride and glycogen stores are important intracellular sources of fuel during prolonged, low-intensity exercise. (*See* **Glycogen**.)

When viewed laterally at high magnifications, myofibrils are seen to be comprised of a series of units known as sarcomeres. Sarcomeres are separated by structures called Z disks, which pass from myofibril to myofibril. Adjacent to the Z disks are light I bands and in the center of each sarcomere is a dark A band.

Myofilament Arrangement

The I and A bands of the sarcomere are caused by a partial interdigitation of yet smaller parallel fibers, called thin and thick myofilaments. Thin myofilaments emerge from the Z disks to form the I bands, and centrally positioned thick myofilaments create the A band.

Within the A band, around 1500 thick myofilaments are surrounded by hexagonal arrays of twice that number of thin myofilaments. Thick myofilaments are anchored by a lattice network of proteins at the center of the sarcomere. Thin myofilaments are kept in register by their interdigitation with the thick myofilaments and their attachment to the Z disks.

Skeletal Muscle Contractile Proteins

When sarcomeres contract, the thin myofilaments slide between the stationary thick myofilaments and pull the Z disks closer together. Sarcomeres shorten whenever the proteins of the thick and thin myofilaments are allowed to interact.

Myosin

Thick myofilaments are comprised of around 300 spirally arranged protein molecules called myosin. Each myosin molecule is composed of two heavy chains and four light chains. The heavy chains form fibrous, rigid 'coiled coils' in what are known as the tail and neck regions and then split into globular heads, on to which the light chains are attached. Hinges at either end of the neck allow the myosin heads to protrude from the thick myofilaments and rotate freely during the cross-bridge cycling (described later) with the thin myofilaments.

Actin

In thin myofilaments, chains of globular G actin protein molecules are assembled into double helical filaments. On each G actin molecule are sites to which myosin heads will bind and, in resting muscle, these myosin binding sites are blocked by tropomyosin protein molecules, which are similar in structure to the myosin tail.

Tropomyosin–Troponin Complex

At 40-nm intervals along the thin myofilaments, tropomyosin molecules are attached to troponin complexes. In the troponin complexes, there are three proteins: troponin I, troponin T, and troponin C. Troponin I anchors the complex to the thin myofilament; troponin T attaches to the tropomyosin molecules, and troponin C binds calcium. As will be described, an elevated intracellular calcium ion (Ca^{2+}) concentration is the mechanism by which electrical signals lead to muscle contraction.

Skeletal Muscle Electromechanical Coupling

Electrical signals to contract are delivered to the midpoints of muscle cells by branches of the motor nerve. As the motor nerve enters the muscle, it divides into a number of nerve endings that come to lie in troughs in the outer membranes of each of the muscle cells of the motor unit.

Neuromuscular Transmission

Within the nerve endings are about 300 000 small vesicles, which contain the chemical neurotransmitter, acetylcholine. When a wave of electrical excitation arrives at the nerve ending, some Ca^{2+} ions flow into the nerve and cause around 300 vesicles to fuse with the nerve cell membrane. Acetylcholine molecules released from the vesicles then diffuse across a 20–30-nm cleft and bind to receptors on the underlying muscle cell membrane. At the membrane, the binding of acetylcholine opens receptor-operated channels which allow positively charged sodium ions (Na^+) to flow into the electronegative interior of the muscle cell. In the millisecond or so before the acetylcholine is broken down, the entry of Na^+ ions decreases the electrical potential across the muscle cell membrane and triggers a subsequent wave of electrical excitation along the muscle cell.

Electrical Conduction Along Muscle Cells

Electrical excitation arises when the electrical potential across the muscle cell membrane decreases to a threshold at which voltage-gated Na^+ and potassium (K^+) channels open. First, Na^+ channels open and a rapid influx of Na^+ ions causes the inside of the cell to become positive for 1–2 ms. Then, as Na^+ channels close, K^+ channels open and a rapid efflux of K^+ ions

returns the interior of the muscle cell to its resting electronegative state.

With the transient depolarization of the muscle cell membrane at the neuromuscular junction, ions are pulled away from adjacent muscle regions which decreases their membrane electrical potential. This sets off further voltage-activated Na^+ influx. In this manner, waves of depolarization propagate outwards from the center of the motor unit towards the ends of the muscle fibers.

Voltage-dependent Sarcoplasmic Reticulum Ca^{2+} Release

Waves of depolarization on the muscle cell surface descend to the myobrils via transverse T tubules, which are continuous with the surface cell membrane and pass into the muscle cell at the junctions between the thick and thin myofilaments. On either side of the T tubules, the cisternae or 'chests' of the sarcoplasmic reticulum are attached to the T tubules by structures known as foot processes.

These foot processes are comprised of four proteins which open Ca^{2+} channels in sarcoplasmic reticulum cisternae by altering their shape during waves of T tubule membrane depolarization. Under these circumstances, the rapid release of Ca^{2+} from the cisternae of the sarcoplasmic reticulum exceeds the rate at which Ca^{2+} is pumped back into the sarcoplasmic reticulum via the previously described network of longitudinal tubules. As a result, the concentration of Ca^{2+} ions around the myofilaments rises. (*See* **Calcium**: Physiology.)

Ca^{2+} Activation of Thin Myofilaments

As the free Ca^{2+} concentration inside the muscle cell increases from its resting value of around $4\,\mu g\,l^{-1}$ to $40–400\,\mu g\,l^{-1}$, the troponin–tropomyosin protein complexes on the thin actin myofilaments rotate to expose the binding sites for the myosin heads. The mechanism by which this occurs is not precisely known, but one possibility is that the binding of Ca^{2+} to troponin C tugs the tropomyosin into the grooves between the double helical actin myofilaments.

Cross-bridge Cycling

Once thin actin myofilaments are 'activated' by Ca^{2+} ions, they are pulled towards the center of each sarcomere by a process known as cross-bridge cycling. Cross-bridge cycles are formed by repeated attachments and detachments of the myosin heads of the thick myofilaments to the actin of the thin myofilaments.

Energy for cross-bridge cycling comes from the binding of ATP to the myosin head, which causes the myosin head to be cocked like the trigger of a gun at a 'strained' 90° angle to the thick myofilament. At this angle, the bound ATP is cleaved to adenosine diphosphate (ADP) and inorganic phosphate (P_i) and that cleavage creates an actin-binding site on the myosin head. Attachment of the myosin head to actin then leads to a release of P_i which causes the 'strained' 90° myosin head to move to an 'unstrained' 45° angle to the thick myofilament. This movement pulls the actin thin myofilament and the attached Z disk some 12 nm towards the center of the sarcomere. At full excursion, the ADP is then replaced by ATP which causes a rapid detachment from the actin and a return of the myosin head to the 'strained' 90° angle.

Myosin head detachments do not occur in unison. While some myosin heads are hydrolyzing ATP to reestablish actin-binding sites for further cross-bridge cycling, other myosin heads are attached. Thus, while free intracellular Ca^{2+} concentrations remain high, a sarcomere shortens continuously and that continuous shortening, repeated in millions of sarcomeres in series, is what we see as muscle contraction.

Muscle Movement

The slowest step in a muscle contraction depends on the load opposing the movement of the myosin heads to their 45° state. When a muscle shortens against a near-zero load, contraction velocity is limited by the rate at which the myosin heads can hydrolyze ATP to recreate an actin-binding site. However, as the load opposing shortening is increased, control of contraction velocity shifts away from the rate of ATP hydrolysis towards the speed at which the myosin heads can detach from the actin filament.

A useful analogy is a man pulling on a rope tied to an increasingly heavy weight. As the weight becomes heavier, less of his time is spent in changing hands and more is spent in hauling on the rope.

Maximum Shortening Velocity

In unloaded contractions at high velocity, where less than 5% of the myosin heads are attached at any one time, maximum rates of shortening depend almost exclusively on the speed at which the myosin heads can hydrolyze ATP to form actin-binding sites. Since fast-twitch muscle fibers hydrolyze ATP at more than twice the rate of slow-twitch muscle fibers, muscles with a high proportion of fast-twitch motor units are able to contract more rapidly against a minimal load than are muscles with mainly slow-twitch motor units.

Maximum Power Output

A high proportion of fast-twitch muscle fibers is also of advantage for maximum power output. Optimum

power output is achieved by adjusting the number of motor units recruited so that each shortens against one-third of its maximum load. Against such loads, muscles contract at around one-third of their maximum velocity, and about 20% of the myosin heads are attached. Maximum power output is therefore also largely determined by the rate at which myosin heads can hydrolyze ATP to recreate binding sites for actin.

Maximum Load

As the load opposing shortening is further increased, however, rates of ATP hydrolysis become less important. Once a muscle can only restrain a load, all of the myosin heads are attached and the maximum load is more a function of the size of the muscle than of its fiber-type composition.

Energy Supply for Muscle Movement

Irrespective of the mode of muscle contraction, the ATP hydrolyzed by the myosin heads has to be immediately resynthesized to allow the cross-bridges to detach. In working muscles, the half-life of an ATP molecule can decrease to a few seconds.

Phosphocreatine Hydrolysis

With a sudden increase in muscle work, ATP for relaxation and contraction is regenerated from ADP and P_i via a number of mechanisms. The most immediate source of ATP is from a reservoir of chemical energy stored in the form of creatine phosphate. From measurements of muscle creatinine phosphate breakdown to creatine and P_i during severe exercise, it has been calculated that net creatine phosphate hydrolysis can sustain maximal contractile activity for about 4 s.

Oxygen-independent Glycogen Utilization

Following the initial hydrolysis of creatine phosphate, the next major source of energy for ATP resynthesis is from the breakdown of the glucose molecules in muscle glycogen to lactate ions via a metabolic pathway known as oxygen-independent glycolysis. Because it takes time for muscle blood flow to increase at the start of exercise, the oxygen-independent glycolytic pathway is particularly important in the white fast-twitch (type IIb) muscle fibers that power escape reactions. Type IIb muscle fibers generate almost all of their ATP for contraction by an acceleration of glycogen to lactate conversion.

During sprinting, glycogen breakdown is increased by activation of an enzyme called glycogen phosphorylase. Glycogen phosphorylase activity is increased by two rather complex mechanisms. Both occur simultaneously but, for the sake of simplicity, they are described separately.

One mechanism of glycogen phosphorylase activation is via a hormonally controlled conversion of the enzyme from a 'b' form to a more active 'a' form. If a need to escape is anticipated, circulating epinephrine (adrenaline) concentrations rise and stimulate the intramuscular production of cyclic adenosine monophosphate (cAMP). Increases in cAMP concentration cause a rapid conversion of glycogen phosphorylase b to a when free Ca^{2+} concentrations increase for contraction. (*See* **Hormones**: Adrenal Hormones.)

Another mechanism of glycogen phosphorylase activation arises from the sudden increase in ATP hydrolysis at the myosin heads. This ATP is initially resynthesized by a breakdown of creatine phosphate to creatine and P_i and by a condensation of two ADP molecules to form one of ATP and one of adenosine monophosphate (AMP). A reduction in creatine phosphate concentration and a rise in P_i, and AMP concentrations act in concert to stimulate glycogen phosphorylase activity.

At the same time, changes in creatine phosphate, P_i, and AMP concentrations also accelerate the rate of glycolysis by increasing the forward reaction and decreasing the backward reaction of a substrate cycle involving two enzymes known as phosphofructokinase and fructose diphosphatase. Fructose diphosphatase is unique to type IIb fibers and inhibition of its backward reaction allows the forward rate of glycolysis in these fibers to be increased more than 1000-fold.

In response to intense muscle activity, glycogen utilization in type IIb fibers rises from around 0.01 to 11 g min^{-1} kg^{-1} fresh weight. Since the glycogen content of human muscle is roughly 15 g kg^{-1} fresh weight, oxygen-independent glycolysis could in theory provide enough energy for 80 s of maximum activity. In practice, however, the mobilization of glycogen is limited to around 5 g kg^{-1} fresh weight and maximum power output cannot be maintained for much longer than 20 s.

When ATP is resynthesized by glycolysis, rather than by mitochondrial oxidative phosphorylation or creatine phosphate breakdown, the hydrogen ions (H^+) produced by its hydrolysis are not reconsumed and H^+ accumulation may be an important factor in the rapid development of fatigue under these circumstances (Figure 1). When carbohydrates are metabolized at high rates to fuel intense muscle activity, hydrogen ions accumulate from the turnover of glycolytic ATP. Hydrogen ions buffered by P_i^{2-} ions form P_i^- ions, which inhibit P_i^- release from the myosin heads and slow cross-bridge cycling.

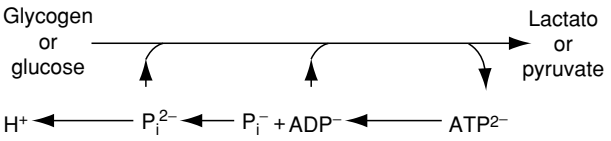

Figure 1 Hydrogen ion production from glycolytic adenosine triphosphate (ATP) turnover. P_i, inorganic phosphate; ADP, adenosine diphosphate. Reproduced from Exercise: Muscle. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

This feedback inhibition of cross-bridge cycling by P_i^- accumulation insures that the muscles' demand for energy cannot exceed the energy supply. If that were to occur, falling ATP concentrations would lead to an irreversible muscle rigor and cell death.

Oxygen-dependent Metabolism

Oxygen-independent metabolism is therefore of little use to the endurance runner. For sustained exercise, the H^+ ions arising from ATP hydrolysis have to be used in its resynthesis and this means that most of the ATP must be derived from mitochondrial oxidative phosphorylation.

The major fuels for mitochondrial oxidative phosphorylation are muscle glycogen, blood glucose, and circulating fatty acids. Muscle glycogen and blood glucose are particularly important fuels during high-intensity exercise. As work rate is increased beyond 70% of an individual's maximum oxygen (O_2) uptake ($V\text{O}_{2\text{max}}$) to 95% of $V\text{O}_{2\text{max}}$, the contribution to energy production from carbohydrate oxidation rises to 100%.

For some reason, fatty acid oxidation can only provide enough ATP for around 50% of maximum power output. Fats are therefore mainly used, together with carbohydrate, in prolonged low-intensity (30–65% of $V\text{O}_{2\text{max}}$) exercise. During such exercise, the contribution to energy production from fat oxidation climbs from 37–39% to 62–67% as rises in blood glucose utilization increasingly fail to compensate for muscle glycogen depletion. (*See* **Fatty Acids:** Metabolism.)

The advantage of muscle fatty acid oxidation in prolonged exercise is that it spares some liver glycogen breakdown for tissues such as the brain which are totally dependent on glucose. Without fat oxidation, it can be calculated that the carbohydrate reserves of the liver would be exhausted within an hour or so of exercise, while the remaining fat cell triglyceride stores would be sufficient for another 2–3 days.

Consequently, muscle fuel utilization is ultimately regulated by the delivery of glucose from the liver into the blood stream. Release of liver glucose is promoted by decreases in circulating glucose concentration and rises in circulating glucagon and epinephrine concentrations. Glucagon and epinephrine binding to the liver cell membrane increase liver glycogen phosphorylase b to a conversion via a mechanism similar to that in muscle fibers. The only differences in the liver are that glucagon binding accelerates intracellular cAMP formation, and epinephrine binding raises free intracellular Ca^{2+} concentrations. (*See* **Glucose:** Function and Metabolism; Maintenance of Blood Glucose Level.)

Glucagon and epinephrine binding also increase the provision of fatty acids to the muscles by activating the cleavage of triglyceride stores to fatty acids and glycerol. In fat cells, both hormones act via a stimulation of cAMP production to convert a less active triglyceride lipase b enzyme to a more active triglyceride lipase a enzyme. Release of fatty acids extends endurance by decreasing blood glucose utilization.

Use of blood glucose and fatty acids by contracting muscles is regulated by the rate of mitochondrial ATP resynthesis. Mitochondrial ATP resynthesis is coupled to a metabolic pathway known as the electron transport chain. In the electron transport chain, H^+ ions and electrons from carbohydrate and fat breakdown are oxidized to water and this oxidation only proceeds when ADP is available for phosphorylation. Thus, until the carbohydrate stores become depleted, muscle metabolism is controlled by the demand for energy rather than the supply of fuel.

See also: **Calcium**: Physiology; **Cells**; **Fatty Acids**: Metabolism; **Glucose**: Function and Metabolism; Maintenance of Blood Glucose Level; **Glycogen**; **Hormones**: Adrenal Hormones

Further Reading

Berman MC, Gevers W and Opie LH (eds) (1985) *Membranes and Muscle*. Oxford: IRL Press at Oxford University Press.

Coggan AR (1991) Plasma glucose metabolism during exercise in humans. *Sports Medicine* 11: 102–124.

Hawley JA, Myburgh KH, Noates TD and Dennis SC (1997) Training Techniques to improve fatigue resistance and enhance endurance performance. *Journal of Sports Science* 15: 325–333.

Jones NL, McCartney N and McComas AJ (eds) (1986) *Human Muscle Power*. Champaign, Illinois: Human Kinetics.

Newsholme EA and Leech AR (1983) *Biochemistry for the Medical Sciences*. Chichester: John Wiley.

Peachy LD, Adrian RH, Geiger SR (eds) (1983) *Handbook of Physiology*, sect. 10. Bethesda: American Physiological Society.

Metabolic Requirements

S C Dennis and T D Noakes, University of Cape Town Medical School, Observatory, South Africa

Anaerobic and Aerobic Metabolism

It has become traditional to group energy metabolism during exercise into that which is independent of oxygen (O_2) supply, so-called anaerobic metabolism, and that which requires the utilization of O_2, so-called aerobic metabolism. Anaerobic metabolism includes net muscle creatine phosphate hydrolysis and conversion of muscle glycogen or blood glucose to lactate. Aerobic metabolism includes the mitochondrial oxidation of the endproducts of the breakdown of muscle glycogen, blood glucose and circulating fatty acid.

Anaerobic (Oxygen-Independent) Metabolism

Anaerobic metabolism is particularly important in short-duration, high-intensity exercise. With extreme exertion, most of the adenosine triphosphate (ATP) for contraction is generated from a net breakdown of creatine phosphate and an acceleration of the conversion of glycogen or glucose to lactate. (*See* **Glycogen**.)

The metabolic pathway leading to lactate formation is called anaerobic glycolysis but, strictly speaking, it should be termed O_2-independent glycolysis. Contrary to popular opinion, blood lactate accumulation during intense exercise is not a 'threshold' response to inadequate O_2 delivery. Rather, muscle lactate production increases as a continuous function of work rate. Blood lactate accumulates only when the rate of lactate efflux from the working muscles into the blood stream exceeds the rate of lactate clearance from the blood by oxidation to carbon dioxide (CO_2) in skeletal and heart muscle, and by conversion to glucose in the liver.

During exercise in which the work rate is increased progressively, there are a number of factors that promote muscle lactate production. One is the hormonal acceleration of muscle glycogen breakdown by the rising concentrations of epinephrine (adrenaline) in the blood stream. Another is the greater recruitment of fast glycolytic (type IIb) muscle fibers at high exercise intensities. A third is the metabolic acidosis arising from the increasing reliance on carbohydrate oxidation at high work rates.

In contrast to the general belief, hydrogen ions (H^+) do not come from lactic acid production (eqn 1).

$$\text{Glucose or glycogen} \longrightarrow 2\,\text{Lactate}^- + 2H^+ \qquad (1)$$

When ATP formation is taken into consideration and the likely electrical charges at intracellular pH are summed, the reactions of the O_2-independent glycolytic pathway do not produce a net gain of H^+ ions (eqn 2).

$$\text{Glucose} + 2\text{MgADP}^- + 2P_i^{2-} \longrightarrow 2\text{Lactate}^- + 2\text{MgATP}^{2-} \qquad (2)$$

where ADP = adenosine diphosphate; P_i = inorganic phosphate; and Mg = magnesium.

Instead, metabolic acidosis during progressive exercise is more a consequence of the increased rate of glycolytic ATP turnover. When ATP is resynthesized by oxidative phosphorylation or by phosphate transfer between creatine phosphate and ADP, the H^+ ions produced by ATP breakdown are utilized in its resynthesis (**Figure 1a, b**).

However, when ATP is resynthesized by glycolysis, the H^+ ions arising from its hydrolysis are not reconsumed (**Figure 1c**). Net proton production therefore occurs irrespective of whether lactate is formed or pyruvate is delivered to the mitochondria for oxidation.

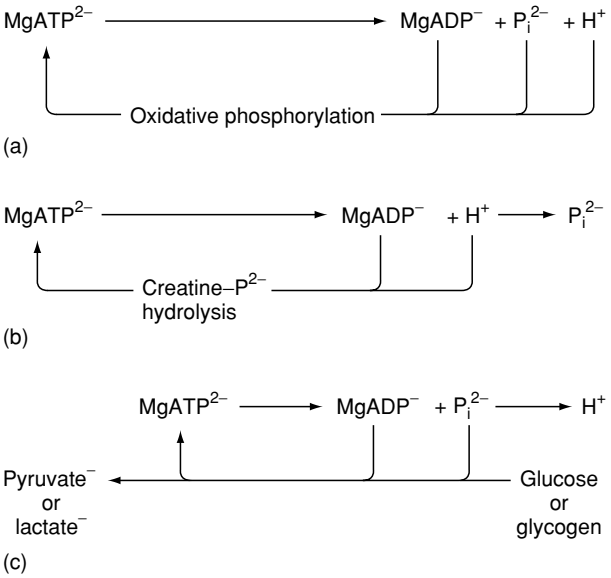

Figure 1 H^+ ion production from glycolytic adenosine triphosphate (ATP) turnover. Unlike when ATP is resynthesized by oxidative phosphorylation or creatine phosphate breakdown, the H^+ ions arising from ATP hydrolysis are not reconsumed when ATP is resynthesized by the conversion of glycogen or glucose to lactate or pyruvate. H^+ ions are therefore produced whenever carbohydrate is metabolized. ADP, adenosine diphosphate. Reproduced from Exercise: Metabolic Requirements, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK, Sadler MJ (eds), 1993, Academic Press.

Since any acceleration of carbohydrate utilization will increase H^+ ion production, lactate formation is more a consequence of, than a cause of, metabolic acidosis. Increases in intracellular H^+ concentration shift the lactate dehydrogenase and lactate permease equilibria towards lactate production and H^+ plus $lactate^-$ coefflux (Figure 2). Thus muscles 'dump' fuel (lactate) to remove H^+ ions into the blood stream whenever carbohydrate utilization is increased to provide energy for exercise of very high intensity.

Once H^+ ions accumulate, however, further acidification by glycolytic ATP turnover is prevented by a slowing of muscle contraction. The H^+ ions combine with the P_i^{2-} ions from creatine phosphate breakdown to form P_i^- ions, which inhibit P_i^- release from myosin heads (Figure 3). In terms of muscle cell survival, this metabolite-induced 'mechanical arrest' during intense, 'anaerobic' muscle activity is an important protective mechanism. If muscles were to become truly 'anaerobic,' energy demand would exceed energy supply and ATP depletion would lead to irreversible muscle rigor and cell death.

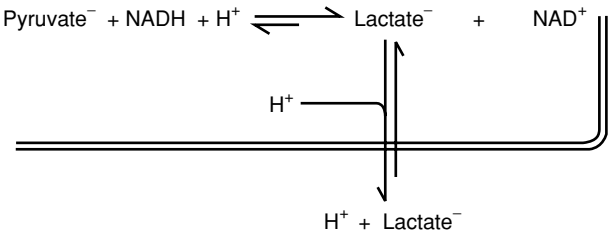

Figure 2 Muscle lactate formation and efflux. Increases in intracellular H^+ ion concentration, arising from rapid carbohydrate oxidation, promote lactate production and efflux from muscle cells. Lactate formation is therefore more a consequence of, rather than a cause of, metabolic acidosis. NADH, reduced nicotinamide adenine dinucleotide; NAD^+, oxidized nicotinamide adenine dinucleotide. Reproduced from Exercise: Metabolic Requirements, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

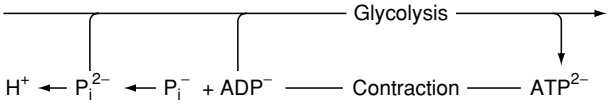

Figure 3 Inhibition of contraction by H^+ ion accumulation. H^+ ion accumulation from glycolytic adenosine triphosphate (ATP) turnover is self-limiting in that H^+ ions $+P_i^{2-}$ ions form P_i^- ions, which slow contraction by inhibiting P_i^- release from the myosin heads. This 'mechanical arrest' is an important protective mechanism to prevent ATP depletion during intense muscle activity. ADP, adenosine diphosphate. Reproduced from Exercise: Metabolic Requirements, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Aerobic Metabolism

Because metabolite-induced mechanical arrest develops during high-intensity exercise in which there are high rates of muscle glycogen and/or blood glucose metabolism, such exercise cannot be sustained indefinitely. For continued physical activity over several hours, the contribution to energy production from carbohydrate utilization must be reduced from nearly 100% to around 60% or 2–3 g min^{-1}. At these rates of carbohydrate oxidation, most of the H^+ ions arising from glycolytic ATP turnover leave the muscle cells in exchange for extracellular sodium ions (Na^+). (*See* **Carbohydrates:** Digestion, Absorption, and Metabolism.)

Since the rate of ATP production from fatty acid utilization is approximately 50% slower than from carbohydrate metabolism, exercise requiring 40% of the energy to come from fat oxidation has to be performed at intensities corresponding to 70–85% of maximum O_2 uptake ($\dot{V}O_{2max}$). Typically, an athlete runs a 42-km marathon (126–240 min) at 85% of $\dot{V}O_{2max}$ and a 100-km ultramarathon (360–540 min) at 70–75% of $\dot{V}O_{2max}$.

During exercise at 70–75% of $\dot{V}O_{2max}$, the average rate of muscle glycogen utilization is around 1·5 g min^{-1}. At this rate of depletion, an athlete's prerace working muscle glycogen content of around (30 g kg^{-1}) × 15 kg or 450 g would be sufficient for approximately 5 h of exercise.

In practice, however, depletion of the carbohydrate stores in the liver may cause premature fatigue. About 135 g of glycogen is stored in the liver after a meal and these stores are sufficient only to sustain exercise at ultramarathon pace (70–75% of $\dot{V}O_{2max}$) for around 3 h. As the initially rapid breakdown of muscle glycogen gradually slows from around 2.1 to 1.0 g min^{-1}, conversion of liver glycogen to blood glucose progressively rises to a maximum rate of approximately 1.1 g min^{-1}. At this point, further increases in blood glucose oxidation can no longer compensate for the decline in muscle glycogen utilization, and work rates must be decreased to allow more energy to come from fat metabolism. Unlike carbohydrate stores, fat reserves could provide energy for many days of exercise.

If an athlete fails to slow down when the demand for blood glucose exceeds liver glucose output, circulating glucose concentrations fall, producing a condition known as hypoglycemia or, in common sporting parlance, the 'bonks.' Because the brain depends on an adequate glucose supply, hypoglycemia leads to a lack of coordination, an inability to concentrate

and, ultimately, collapse. This collapse prevents an individual from dangerously depleting the body's carbohydrate stores. (*See* **Glucose**: Maintenance of Blood Glucose Level.)

For prolonged 70–75% of $V\text{O}_{2\text{max}}$ exercise, an athlete must therefore ingest carbohydrates and spare liver glycogen. Provided that enough carbohydrate is ingested to satisfy the eventual 2 g min^{-1} demand for blood glucose by the muscles, an athlete can continue exercising for some time after muscle glycogen has been completely exhausted.

Despite the maintenance of high rates of carbohydrate oxidation and normal blood glucose concentrations, however, a point is eventually reached where an athlete can no longer sustain an exercise intensity of 70–75% $V\text{O}_{2\text{max}}$. Under these circumstances, there is an increased perception of fatigue. Some believe that this fatigue is due to the slow utilization of plasma branched-chain amino acids by skeletal muscle during prolonged activity. Since the contribution to energy production from branched-chain amino acid metabolism is trivial, it has been suggested that branched-chain amino acid oxidation might provide a communication link between working muscles and the brain.

Over the course of a marathon, plasma branched-chain amino acid concentrations have been shown to decline from 650 to 500 mg l^{-1}. At the same time, the increases in circulating free fatty acid concentrations during exercise were found to raise the free plasma tryptophan amino acid concentration from 1.4 to 2.9 mg l^{-1}, by competing with tryptophan for binding to blood albumin.

Decreases in the circulating branched-chain amino acid concentration and rises in plasma free tryptophan concentration with sustained physical activity promote tryptophan entry into the brain. Passage of tryptophan across the blood–brain barrier is normally limited by its low free concentration and by a competition from branched-chain amino acids for access to the neutral amino acid transporters.

Once in the brain, tryptophan is converted to the neurotransmitter 5-hydroxytryptamine (serotonin) in the serotonergic nerve terminals of the hypothalamus. Serotonin biosynthesis is unusual in that it is exclusively regulated by the rate of entry of tryptophan into the brain.

Since serotonin is involved in the control of sleep, food intake, pain sensitivity, pituitary hormone release and mood, it is tempting to speculate that the elevated serotonin levels in the brain might contribute to the perception of fatigue in prolonged exercise, when muscle fuel supply appears to be adequate. Certainly, excessive serotonin release would account for the improvement in mood and increased drowsiness after long-duration, low-intensity exercise.

The Athlete's Diet and Supplementation

Additional Protein

Whether increased amino acid utilization during and following exercise significantly increases an athlete's daily dietary protein requirement is not known. Despite considerable research on this subject, there are still two schools of thought. Some argue that the 78% safety margin in the 0.8 g per kg of body weight per day recommended dietary allowance (RDA) for protein is sufficient to cover the increased amino acid requirements of both endurance- and strength-trained athletes. Others feel that frequent exercise elevates the daily need for dietary protein. The balance of evidence suggests that an athlete should probably ingest around 1.2 g of protein per kg of body weight per day. (*See* **Protein**: Requirements.)

Daily protein intakes of 1.2 g per kg of body weight, however, do not necessitate protein supplementation. An athlete's larger than normal daily kilojoule ingestion more than adequately covers any increase in his or her protein requirements. Furthermore, there is no evidence that protein intakes in excess of 1.2 g per kg of body weight per day improve sporting performance. Provided that an athlete eats enough to meet his or her daily energy expenditure, protein is not a major fuel for exercise. A significant use of protein for energy occurs only in starvation, when body protein is converted to carbohydrate to maintain an adequate supply of glucose to the brain.

Higher than recommended (1.2 g per kg of body weight per day) protein intakes are also unlikely to help increase muscle size. Muscle hypertrophy is under much more subtle control than the supply of amino acids for protein synthesis.

Additional Vitamins and Minerals

Another question frequently asked by athletes is whether or not they should increase their intake of vitamins and essential minerals. Unless an athlete is malnourished or follows an extreme vegetarian diet, the answer is probably no. Vitamin and mineral deficiencies are extremely uncommon among athletes and there is no evidence to show that extra vitamins and minerals improve performance. Vitamin and mineral supplements are not being promoted to meet the unique nutritional requirements of athletes, but rather to satisfy commercial interests.

However, there is growing interest in the possibility that very heavy training and competition may suppress the immune response and thereby increase an athlete's risk of infection. Since vitamin A and perhaps also vitamins C and E are thought to improve

resistance to infection, an athlete might consider an increased intake of these vitamins during periods of intense training and competition. (*See* **Ascorbic Acid**: Physiology; **Retinol**: Physiology; **Tocopherols**: Physiology.)

Additional Carbohydrate

With heavy training, it is essential that an athlete eats sufficient carbohydrate. Studies of the effects of daily, 2-h, 70% of $V\text{O}_{2\text{max}}$ training runs on leg muscle glycogen contents have shown that a standard 40–45% carbohydrate diet is not sufficient to maintain body carbohydrate stores. For the last 4 or 5 days' intensive training, the subjects' preexercise muscle glycogen contents progressively declined from around 22 to 5–6 g kg^{-1}.

Even a high-carbohydrate (70%) diet failed to replenish completely the subjects' muscle glycogen stores between exercise bouts. With the ingestion of 560 g of carbohydrate per day, muscle glycogen levels still fell from approximately 22 to 17 g kg^{-1}.

Three days before an endurance race, therefore, an athlete should stop training and follow what is known as a 'carbohydrate-loading' regime. During carbohydrate loading, the optimum strategy is to eat 500–600 g of simple carbohydrates for the first 24 h and then consume the same quantity of mainly complex carbohydrates on the following 2 days. By ingesting large amounts of carbohydrate and not exercising, athletes can increase their muscle glycogen content from 22 to 36 g kg^{-1} and their liver glycogen stores from 70 to 90 g kg^{-1}.

As mentioned above, it is also important for an athlete to drink carbohydrate solutions during prolonged exercise to prevent hypoglycemia. Provided that the volumes described below are ingested, it does not matter whether the carbohydrate is glucose, sucrose, maltose or a commercial 7–22-chain-length glucose polymer. In each case, peak rates of exogenous carbohydrate oxidation occur 75–90 min after ingestion and are 0.7–0.9 g min^{-1}. (*See* **Carbohydrates**: Metabolism of Sugars.)

The exception to this pattern is fructose. Because ingested fructose is (1) slowly absorbed from the intestine and (2) converted to glucose in the liver before being oxidized, its maximum rate of use is only 0.4–5 g min^{-1}.

Maintaining Water and Electrolyte Balance

Interestingly, the first evidence that carbohydrate ingestion can enhance performance during prolonged exercise was published in the 1930s but was ignored by athletes, coaches, and scientists until the mid-1980s. Even the idea that fluids should be ingested by athletes during prolonged exercise is of recent origin. Prior to the early 1970s, the prevailing opinion was that any fluid ingested during exercise, including marathon running, adversely affected performance. This fact is frequently overlooked by today's athletes and their advisors who have emphasized the perceived value of fluid replacement during exercise, sometimes to excess.

Factors Influencing Fluid Loss

The important principles underlying fluid and electrolyte balance during exercise are that the rate at which sweat is lost is determined principally by the athlete's rate of energy expenditure (metabolic rate) which is largely determined by his or her mass and speed of movement. Thus sweat rate is highest in activities of short duration and high intensity, such as competitive running over distances of 3–10 km, and is slower in prolonged, lower-intensity exercise, such as marathon and ultramarathon running or ultratriathlon events. Sweat rate also tends to be greater in heavier than in lighter athletes. At the same speed, the heavier cyclist or runner has to expend more energy in moving a larger mass.

When cyclists and runners exercise at the same metabolic rate, the sweat rate of cyclists is lower than that of runners. Cyclists travel faster than runners and therefore lose more heat by convection to the passing air than by sweat evaporation.

Sweat rate is also influenced by the ambient temperature but, in practice, environmental conditions may affect sweat rates less than expected. When forced to exercise in severe heat, athletes usually choose to exercise at a lower exercise intensity. Athletes are well aware that they will fatigue rapidly when exercising in the heat, and a lower metabolic rate reduces their sweat rates.

Contrary to popular opinion, however, 'heat exhaustion' is not caused by excessive body heat accumulation. In very severe environmental conditions, the rate of energy expenditure appears to be regulated specifically to prevent the development of high sweat rates, severe dehydration, and heat injury.

There is also some uncertainty as to whether or not all the fluid lost during exercise must be replaced if serious dehydration is to be prevented. Up to 1.8 l of fluid can be stored in association with the glycogen contained in a carbohydrate-loaded athlete's working muscle and liver. Thus, in activities that largely deplete liver and muscle glycogen contents, a mass loss of up to 1.8 kg can be incurred without the athlete's fluid status being seriously compromised.

Fluid Replacement

In athletes losing more than 1.8 kg of mass, the principal aims of fluid ingestion during exercise are to replace the fluid and electrolytes lost during exercise and to provide carbohydrate in solution to supplement the body's limited carbohydrate stores. Usually, adequate fluid replacement has been considered of greater importance than carbohydrate supplementation, but more modern studies indicate that an adequate carbohydrate supply is at least as important as optimum fluid and electrolyte replacement.

Failure to replace adequately the sweat losses incurred during exercise leads to progressive dehydration. Symptoms of dehydration are: (1) an increase in heart rate with a fall in cardiac stroke volume; (2) a reduction in sweat rate; (3) a decrease in blood flow to the skin; and (4) a progressive rise in body core temperature.

It is not yet clear to what extent the symptoms of dehydration adversely influence athletic performance. A small weight loss resulting from unreplaced sweat losses during exercise might enhance performance by reducing the mass which the athlete must carry. Eventually, however, there must be a crossover point at which the negative cardiovascular and other effects of more severe dehydration outweigh the beneficial effects of weight loss.

Electrolyte Replacement

In considering fluid replacement, it is important to realize that, in health, it is not the fluid content of the body that is regulated but the electrolyte concentration in the extracellular and intracellular compartments. In the 15-l extracellular compartment of a 70-kg man, the major electrolyte is sodium (Na^+) and, in the 30-l intracellular compartment, the main electrolyte is potassium (K^+).

As Na^+ is the principal electrolyte in sweat, the initial site of fluid loss is from the extracellular compartment, which includes the circulating plasma volume. The extent to which dehydration of the extracellular compartment occurs is determined by the amount of Na^+ lost in sweat.

An important consequence of exercise training, especially in the heat, is a reduced sweat Na^+ content. The result is that fluid losses from the extracellular compartment and circulation are decreased in trained, heat-acclimatized athletes. This might explain why highly trained athletes are better able to cope with dehydration during exercise, because more of their fluid loss will come from the intracellular compartment than from the plasma.

Because the concentration of electrolytes regulates the volumes of the extra- and intracellular compartments, it follows that rehydration of these compartments can only occur when their electrolyte losses are corrected. Consequently, there is the need to ingest Na^+ and, perhaps, some K^+ during and after exercise.

Solutions containing sodium chloride (NaCl) are best ingested together with carbohydrate. Carbohydrate digestion leads to Na^+–glucose cotransport across the intestinal brush border and that accelerates the absorption of NaCl and H_2O into the blood stream.

Practical Implications

The present opinion is that an athlete should ingest between 300 and 800 ml of fluid per hour. The high (800 ml) value is for heavy athletes exercising at fast speeds; the low (300 ml) value is for light athletes exercising slowly. Ideally, the fluid should contain NaCl at concentrations of 5–6 g l^{-1} but salt concentrations closer to 2 g l^{-1} are more palatable.

To provide carbohydrate at rates equal to that at which glucose is absorbed into the blood stream and oxidized by the working muscles (approximately 1 g min^{-1}), a mono-, di-, or oligosaccharide should also be added to the solution. For an athlete drinking 600 ml h^{-1} or 10 ml min^{-1}, it can be calculated that the optimum concentration of carbohydrate after 1–2 h of exercise is around 0.1 g ml^{-1}.

See also: **Ascorbic Acid**: Physiology; **Carbohydrates**: Digestion, Absorption, and Metabolism; Metabolism of Sugars; **Glucose**: Maintenance of Blood Glucose Level; **Glycogen**; **Protein**: Requirements; **Retinol**: Physiology; **Tocopherols**: Physiology

Further Reading

Burke LM and Deakin V (eds) (1994) *Clinical Sports Nutrition*. Sydney: McGraw Hill.

Clarke N (1991) *Sports Nutrition Guidebook*. Campaign, Illinois: Human Kinetics.

Coyle EF (1991) Timing and method of increased carbohydrate intake to cope with heavy training, competition and recovery. *Journal of Sports Science* 9: 29–52.

Hawley JA and Burke LM (1998) *Peak Performance: Training and Nutrional Strategies for Sport*. Sydney: Allen and Unwin.

Hawley JA, Dennis SC and Noakes TD (1995) Carbohydrate, fluid and electrolyte replacement during prolonged exercise. In: Keis CK and Driskell JA (eds) *Sports Nutrition: Minerals and Electrolytes*, pp. 233–263. Boca Raton, CRC Press.

Noakes TD (1991) *Lore of Running*. Champaign, Illinois: Human Kinetics.

Noakes TD, Hawley JA and Dennis SC (1995) Fluid and energy replacement during prolonged exercise. In: Tory

SS and Shepherd RJ (eds) *Current Therapy in Sports Medicine*, 3rd edn, pp. 517–520. Philadelphia: Mosby-Year-Book.

Peterson K (1988) *Eat to Compete. A Guide to Sports Nutrition*. Chicago: Year Book Medical Publishers.

EXTRUSION COOKING

Contents

Principles and Practice

R C E Guy, Campden & Chorleywood Food Research Association, Chipping Campden, UK

Extrusion Cooking

Extrusion cooking technology was invented in the 1940s by the Adams Company to manufacture snack foods from maize grits. It is derived from the earlier extrusion technologies used to form pasta at temperatures of 70–80 °C. These processes involved the conveying, mixing, and compressing of moist wheat semolina to form dough and to extrude it as well-defined shapes.

In extrusion cooking, a feedstock of moist powders is also conveyed, mixed, and compressed before shaping at a die (**Figure 1**). However, the dough temperature is raised to much higher levels in the range from 110 to 200 °C on the screws within a barrel. This means that any water present is superheated within the dough but remains in the liquid state. The high temperature also causes changes in the structures and form of natural biopolymers such as starch and soya globulins. Physical bonds can be broken at high temperatures, and the new forms of the polymer can be used to create novel structures for foodstuffs. A particular feature of extrusion cooking is the ability of the process to manipulate these polymers at low moistures and to form doughs or fluid systems from them with the minimum of moisture.

The basis of the technology is a screw system within a tube or barrel, which conveys the dough towards small openings called dies. In the confined space of the barrel, the dough is compressed and heated to high temperatures at high pressures before being extruded through the dies into the atmosphere. An extrusion cooking process has a number of key features, the feeding devices supplying the raw material feedstocks, the design of the screw system and its barrel, the dimensions and number of the dies, and the devices that handle the extrudates.

Extrusion cooking is a continuous process in which the raw material feedstock is metered in at a constant rate, and the machinery maintains a steady-state equilibrium. This is achieved by balancing the forward flow produced by the screws against the pressure at the die. It is a multivariate process depending on all the independent variables of the machinery and the raw materials to produce a range of dependent variables in the machinery and a further range of dependent variables in the extrudates and products. The consequence of varying one independent variable may be to change several of the product characteristics. Therefore, a good systems approach must be used to control such a process.

The range of products manufactured by extrusion cooking can be divided into two groups: animal products and human foods. The animal products themselves include basic feeds for stock animals, marine feeds for freshwater and saltwater fish, shrimps or prawns. They also include large markets for both dry and canned dog and cat foods and treats, and a number of products for other pets and for wild and tame birds. Human foods include the original maize snacks and many new forms including modern versions of prawn crackers and other traditional snacks.

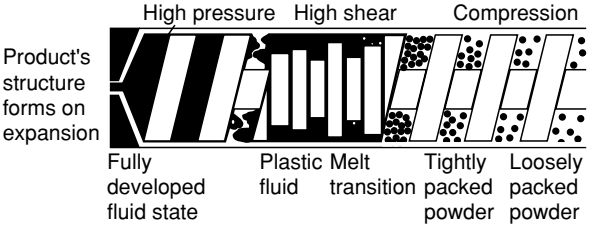

Figure 1 Diagram of the development of raw materials in a twin screw extrusion cooker.

There are many breakfast cereal products formed into balls, hoops, and flakes that have large traditional markets in the USA and UK and growing markets in many areas. Similar expanded pieces of rice and maize are used in chocolate confectionery bars. These bullets of cereal are also used in other products where a crisp, expanded-foam particle is required, such as in granola bars, biscuits, and even icecream.

The range of extruded foods has been extended to coatings, such as breadings. These are used on fried foods, particularly frozen fish and chicken portions. Extruded products have replaced some of the traditional baked crumbs. They are made by expanding cereal extrudates and slicing them into crumbs. This process may be taken a step further in the manufacture of baby foods from extruded feedstocks of cereals and milk powders. For these products, the extrudates are ground to a fine powder, sieved and used to replace the powders manufactured by the more expensive roller drying process.

Principles of Extrusion Cooking: Overall Balance

The extrusion cooking process is used to transform a feedstock of raw materials into a fluid dough and to continue to change the physical nature of this dough until it is forced out of the die (Figure 2). At that stage, the dough must have attained the ideal form to create a structure in the extruded product. The transformation of the native raw materials has an optimum range for most product types. For some products, only small changes are required, whereas for others, a very high degree of change must be induced. For each individual product type, the process must be maintained in a balance between the independent input variables to obtain the ideal conditions within the barrel to achieve this level of transformation.

A good view of an extrusion cooking process is given by the publications of Meuser and van Lengerich. They showed that the relationships between the independent process inputs from the machinery and raw materials created a series of dependent variables within the barrel of the extruder. Most notable of these are residence time, temperature, pressure, moisture level, applied mechanical energy, and mass flow rate. Biopolymers in raw materials, such as starch, were affected by these variables and changed in physical form to dominate the physical features of the fluid dough. As the dough was extruded, the new forms of biopolymers controlled the expansion process influencing the shape, extent of expansion, and the foam texture of the extrudates. The important principle in this section is that the continuous process is in a metastable steady state. Guy and Robert showed that all extrusion cooking processes fluctuate as the balance is maintained between the independent inputs of the process. A fluctuation in an independent input could upset the balance and cause a larger fluctuation in the outputs and therefore in the product quality. It is important, therefore, to minimize any fluctuations in the input variable such as the powder and liquid feed rate, feedstock raw material mix, and quality.

Principles of Extrusion Cooking: the Screw Design

The screws of an extrusion cooker are designed to convey a powder or viscous fluid from the feed port to the die and to force the material through the die (Figure 3). Single screws were used in the early machines leading on from the pasta equipment. More recently, pairs of corotating intermeshing screws have been adapted from their use in the plastics industry to become important machines for the food and feed industries. For conveying materials, the screws are designed with a pitch of 0.5–1 times the screw diameter and a small clearance between the leading edge and the wall. This provides a screw with excellent conveying power for viscous fluid and powders. The effect of a simple conveying screw on the raw materials is minimal, and even with a very small die, the physical transformation of biopolymers is small.

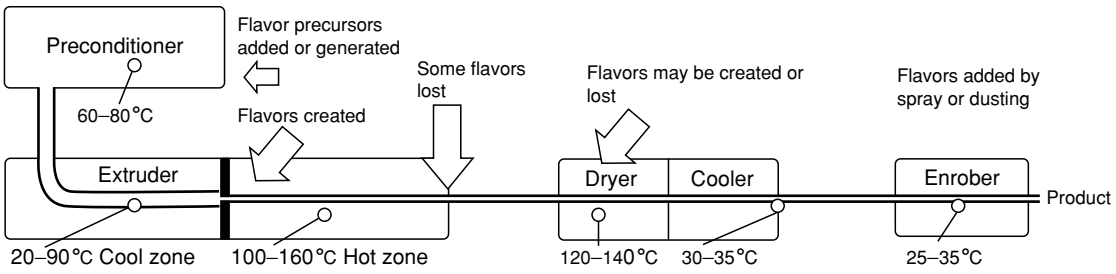

Figure 2 Extrusion cooking equipment in a processing line, with a powder feeder, preconditioner, extruder, dryer and enrober.

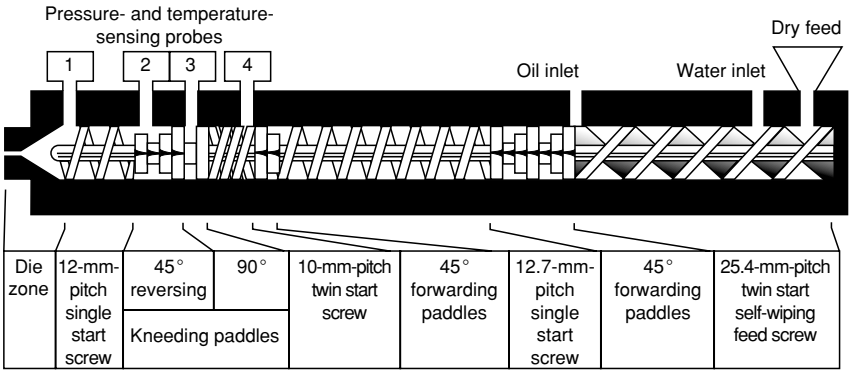

Figure 3 Side view of screw system for a twin screw extruder used at CCFRA showing the functions of the elements.

The second feature of design for a screw system is to apply high levels of mechanical shear to the feedstock being conveyed along its length. In the single screw, this is achieved by reducing the depth of the flights as the material approaches the die and by introducing cuts in the flights to increase back-leakage, thus extending the residence time in the high temperature/high shear zone for the fluid. In the twin screw machines, special sections are added in which the flights are reversed to give a back-pressure pump acting against the main flow. These sections are made weaker by cutting gaps in the flights so that the main pumping screws overcome them. This achieves a steady flow of the fluids though the reversing section under shear.

A special form of reverse section can be constructed from elliptical paddles placed on each shaft in a series. If successive pairs are placed at an angle to the upstream pair, a screw-like structure can be formed. This type of section acts on the fluid as it passes through, kneading the fluid against walls and the other elliptical paddle. An important feature of such a section is that it is set up as a reversing screw to ensure that all the paddles are full of fluid. If they are not full, they have little or no shearing effect on the fluid.

The final role of the screws is to push the fluid melt through the dies in a controlled manner. Normally, the extruder is run with the final sections completely full so that pressure is transmitted from the main conveying screws to the dies. In those cases where reverse sections are used, there may be a pressure drop after the reverse section, and a final section of conveying screw can be used to pump the fluid through the die. In most processes where the extrusion melt fluid has a temperature above 100 °C, the pressure in the die will need to be higher than the water vapor pressure. This insures that the fluid emerges from the die and expands in the air rather than in the die cavity. Normally, the pressure at the die for good control of hot melt extrusions is above 40 bar.

Principles of Extrusion Cooking: the Transformation of Biopolymers

The main feature of extrusion cooking is the transformation of the biopolymers in a raw material feedstock. In 90% of the extruded products manufactured, the active biopolymer is starch, and for the remainder is a protein from oilseeds, such as soya or wheat gluten. Physical transformation is necessary to obtain a more suitable form of the biopolymer for structure creation in the postextrusion expansion processes (Figure 1).

Starch is found in the world's major crops, cereals, such as wheat, rice and maize, and in important tubers such as potato and manioc. In all cases, it is found as partial crystalline granules, 5–80 μm in maximum dimension. The polymers can be released from the granules (Figure 4) by melting their crystalline regions and applying large shearing forces, or by dispersing in an excess of water (>30–35%). For extrusion cooking, low moisture is used, 15–20% in many cases, and the melted starch must be dispersed purely by mechanical shear to obtain a melt fluid of partially or fully dispersed aggregates. The melting temperature for a low moisture feedstock may be in the range of 120–160 °C. Therefore, it is necessary to heat the feedstock to this temperature by the action of the screw and by barrel heating. In low moisture systems with up to 25% moisture, the action of the screws on the powders under compression to a density of $1.4\,g\,ml^{-1}$ will normally achieve the required temperature by frictional heating and the dissipation of mechanical energy from the drive motor.

The second part of the transformation of the starch is more subtle. It has been shown that the major polymer in starch, amylopectin, has a molecular weight range of 10^8–10^9 Da. This size range is too

large for good flow properties in fluids at the concentrations found in the melt fluids. It interferes with the expansion process and reduces the extensibility of the starch films in cell walls. However, the application of high shear inputs to the fluid using reversing screws and high compression ratios can degrade the starch to a much smaller range of $1–5 \times 10^6$ Da. This material is much more extensible and gives a more expanded range of products.

The principle of extrusion is to transform the biopolymers, by heat and shearing forces, to a new form. This may be a completely dispersed form in which the native structures have all been destroyed, and the polymers now form a continuous fluid in water. Such a processed starch would be suitable for forming films that can retain gases during expansion to form fine foams. In examples with proteins, the fluid may be made to texturize during laminar flow through long dies.

The processing of starch may be set up to produce any type of starch structure from the native to the fully dispersed and degraded forms. If the range is divided into a number of stages, as in Figure 4, the type formed at each stage may be associated with a different product in the list discussed earlier.

Principles of Extrusion Cooking: Miscellaneous Reactions and Physical Effects

The extrusion cooker is a chemical reactor that can cause interactions between some of the ingredients such as amino acids and reducing sugars to take part in Maillard reactions and similar degradative and condensation reactions. These reactions produce color and flavor compounds and may also reduce amino acid and vitamin contents. More details of these reactions are given in the chapter on Extrusion Nutrition.

Extrusion cooking also has a unique action that serves to sterilize the materials being processed when their processing temperatures are above 100 °C. The time spent in the extrusion processing zone is short, only 45–120 s in most processes, but the high shear fields and temperatures of 140–160 °C in the shear zone help to reduce the viable cell counts to low levels. Finally, the explosive release of superheated water from within cells ruptures their structure and kills off the residual level of viable cells.

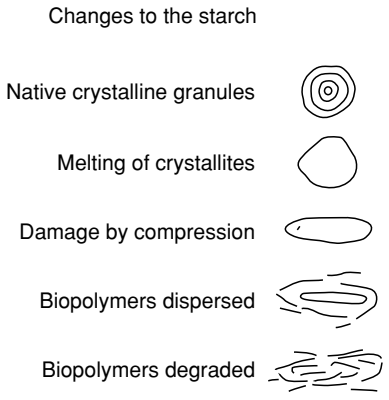

Figure 4 Stages in the transformation of starch in an extruder.

Practice of Extrusion Cooking: Animal Feedstuffs

The basic stock feeds are made using cookers and pelleting presses. This does not gelatinize the starch and merely pasteurizes and compacts the material as loose pellets. The second level of processing is obtained with expanders that add steam to an open screw system feeding pelleting presses. This process also fails to gelatinize more than a small amount of starch, until a die is used, and the equipment becomes a full extrusion cooker.

Feedstuffs with gelatinized and dispersed starch are made for special weaning foods and for all the marine feeds for fish and shrimps. In these products, the pellets are held together by the melted dispersed starch so that a strong structure is formed.

Extruders are also used to improve the nutritional value of oilseed and other protein stocks by denaturing proteins that inhibit digestion such as lectins and trypsin inhibitor in soya beans and dough ball forming elastic glutenins in wheat flours.

Practice of Extrusion Cooking: Pet Foods

Dry pet foods are made from an extruded feedstock of cereals, fortified with protein-rich materials such as dried meat meal, chicken digest, and similar materials. They tend to be fairly dense products with a specific density of 0.35–0.45 g ml^{-1}. The structural material is normally starch, which is supplied by a cereal such as wheat or maize for standard products, and rice for special dietary feeds. There may also be higher levels of oils and fats and minerals to provide a balanced diet for the animals.

Many processes incorporate a preconditioning unit (Figure 2) in which the bulk of the feedstock is conveyed and mixed with oil and steam to produce a hot feedstock at 80 °C for the extruder. This enables a higher throughput of materials to be processed by the extruder.

Processes containing high oil contents must be balanced with materials that absorb oil and prevent

slippage in the machinery. Alternatively, the oil can be sprayed on to the extrudates postextrusion.

Practice of Extrusion Cooking: Snack Foods

The snack foods are made in several forms by either direct expansion or expansion of dense cooked pellets. In direct expansion, the feedstock may be formed from a cereal or potato base mix containing 75–85% starch. The extrusion process is high shear at a low moisture to give highly degraded starch with no granular structure and a large amount of molecular degradation. This allows an expansion to a very low specific density of 0.14–0.20 g ml^{-1}. Directly extruded snacks may be formed as tubes at the die and injected with a cream or jam through the center to manufacture a coextrusion or coextruded snack. The simple products are coated with oil and flavorings to complete the products.

For the pellet process, a similar feedstock is processed at low shear to melt the starch crystalline regions but to retain its granular structure. The cooked melted fluid is cooled to 95–100 °C after cooking, before extruding to the final shape. This soft dense extrudate is cut into a pellet and dried to < 12% moisture. As it cools to ambient temperatures, it assumes a glassy structure and is known as an intermediate pellet or half-product.

A snack is formed by heating the half-product for a few seconds in hot oil or hot air at 190 °C. Recent developments have seen the production of such products in two layers with a half-product. On frying, this forms a large central cavity and is sometimes called a '3-D snack'.

Practice of Extrusion Cooking: Breakfast Cereals

A simple form of breakfast cereal is made for a cereal feedstock with wheat, maize, or rice flours and some small additions of sugar, salt, and proteins. The feedstock is processed at 20–25% moisture to a medium expansion at 150 °C, to give a specific density of 0.22–0.28 g ml^{-1}. The desired products are formed with a coarse cellular texture and thick cell walls, so that they remain crisp in milk for a few minutes. This is achieved by melting the starch granules and degrading about 50% of the granular structures.

The recipes for directly expanded breakfast cereals may be varied to form single cereal varieties such as crisp rice or corn pops, or to form multigrain cereals from blends of wheat, rice, maize, barley, and oats.

A unique form of the expanded breakfast cereal is the high bran product in which wheat bran may be used at levels of 70–75%. In this product, a small amount of starch (15–20%) is used to bind the bran fibers together in a cellular structure. The starch must be melted and completely dispersed to achieve this effect.

The second major form of extruded breakfast cereal is the flaked product, such as the cornflake or multigrain flake. This product is similar in processing requirements to the half-product snack in that it is made from a cereal feedstock at 120–130 °C at low shear. The starch granules are melted, but not dispersed to any significant level. In the process, the cooked cereal feedstock is extruded as a dense fluid at 90 °C and cut into small beads. Each bead forms the basis for a flake when rolled out to a thickness of 0.5–0.7 mm. This process occurs under temperature control at 40–60 °C and is followed by a toasting process at high temperature 240 °C for a few seconds to blister and dry out the flakes. The degree of processing of the starch is very important for the quality of the flakes when they are consumed in milk as a breakfast cereal. If the starch is sheared too much, the flakes take up water too quickly and become limp and flabby in the bowl.

Practice of Extrusion Cooking: Breadings and Coatings

Traditional breadcrumbs are manufactured by making loaves of bread, and shredding and drying their crumb to produce a small range of products. Several extrusion cooking processes have been set up to make similar products from the same materials and a wider range from other ingredients, which cannot be used in traditional bread. The extrusion of a wheat-flour feedstock at 28–35% moisture at temperatures of 110–120 °C at low shear produces expanded extrudates with cellular textures similar to that of breadcrumbs. Starch granules within the structure should be melted to lose their crystalline structure, but only dispersed to a small degree so as to provide material for expansion. The extrudates may be cut while still warm and moist to form crumbs, and then dried to <10% to form the glassy structure required for coating crumbs. The recipe for coating crumbs may be based on wheat flour and modified with some chalk to increase the number of cells. Other cereals such as maize and rice may be used with, or instead, of wheat flour to give a variety of products. Maize adds an attractive yellow color to products, whereas rice give whiter crumbs.

Practice of Extrusion Cooking: Baby Foods

Extrusion cooking has been used to produce expanded baby foods from rice and vegetable flours

such as peas and carrots. These products are similar to snack foods in that the starch must be well dispersed to give good expansion and a light texture. The process must be controlled in terms of temperature to keep the mass temperature as low as possible <130 °C during processing.

The second form of baby foods is more important because it represents a larger amount of production. It was developed to match the roller-dried instant baby food, which is consumed as a porridge, prepared by mixing water or milk with a dry instant powder. The extruded product is made by processing a feedstock of cereals with milk solids under conditions of medium shear input and controlled temperature. The aims are to melt the starch granules, but to retain their structures to a large extent. This gives a high paste viscosity when added to water or milk and produces a more attractive porridge. It is also important to control the temperature of the fluid to reduce any browning reactions that would discolor the products and reduce the nutritional quality of the proteins.

See also: **Cereals**: Breakfast Cereals; **Extrusion Cooking**: Chemical and Nutritional Changes; **Infant Foods**: Weaning Foods; **Snack Foods**: Dietary Importance

Further Reading

Chang YK and Wang S (eds) (1996) *Advances in Extrusion Technology, Aquaculture/Animal Feeds and Foods*. Lancaster, PA: Technomic.

Frame N (ed.) (1994) *The Technology of Extrusion Cooking*. Glasgow, UK: Blackies, Academic & Professional.

Guy RCE (ed.) (2001) *Extrusion Cooking, Principles and Technologies*. Cambridge, UK: Woodhead Publishing.

Kokini JL, Ho C-T and Karwe MV (eds) (1991) *Food Extrusion Science and Technology*. New York: Marcel Dekker.

Mercier C, Linko P and Harper JM (eds) (1990) *Extrusion Cooking*. St. Paul, MN: American Association of Cereal Chemists.

Chemical and Nutritional Changes

I T Johnson, Institute of Food Research, Norwich, UK

Background

The extraordinary versatility of the extrusion cooking process and its variants has led to the use of this technology in a very wide range of applications by the food-processing industry. In developed countries it is widely employed to fabricate snack foods with novel characteristics of flavor and texture, but which usually have little nutritional significance in the diet as a whole. However, the technique can also be applied to the production of more substantial products such as breakfast cereals and crispbreads and even weaning foods, which may contribute a very large proportion of the total nutrient intake of infants. Even in the developing world, where a suitable technical environment exists, extrusion cooking may be applied to the production of staple foods from indigenous crops. This remarkable variety of applications is a direct result of the extreme physical conditions which the process imposes upon raw materials, and the complex transformations which are achieved in the final product. The chemical and physical reactions which underlie these changes can have both positive and negative implications for nutrition and food safety. This issue has attracted increasing attention in recent years.

Changes in Carbohydrates

Cereals and potatoes were the first commodities to be widely processed by extrusion cooking, and they are still the most common. This fact has helped to focus particular attention on the behavior of complex carbohydrates during high-temperature extrusion. Cereal grains and vegetable tubers are storage organs which contain starch, and some free sugars, embedded within a complex cellular matrix. Starch consists of glucose molecules, linked in the case of amylose exclusively by α-1,4 bonds, or, in the case of amylopectin, by a combination of α-1,4 and α-1,6 bonds, which give a branching structure to the molecule. All such links are susceptible to hydrolysis by the digestive enzymes of the human alimentary tract. Cell walls are much more complex structures containing a variety of different polysaccharides, all of which are resistant to digestion. These materials, together with lignin, comprise dietary fiber. (*See* **Carbohydrates**: Interactions with Other Food Components; **Dietary Fiber**: Properties and Sources; **Lignin**; **Starch**: Structure, Properties, and Determination.)

Native, uncooked starch also resists digestion because it is localized within plant cells in the form of semicrystalline granules containing varying proportions of amylose and amylopectin. The polysaccharides are tightly bound together by hydrogen bonding, but the density and orderliness of the bonding vary within the granule matrix. In the presence of moisture, hydrogen bonding in the less ordered regions of the granule is susceptible to

disruption by heat. This allows the entry of water into the granule, which then swells and becomes increasingly susceptible to mobilization and dispersion of its constituent polysaccharides. Where substantial quantities of excess water are present, the granular structure of starch is completely converted into a viscous fluid dispersion. This process is called gelatinization and it is one of the most important functions of processing and cooking because it renders starch readily available for digestion in the small intestine.

During the extrusion cooking process a dough is driven continuously through a die by one or more screws rotating within a high-pressure barrel. As the dough is pressurized its temperature rises. An external source of heat may also be employed, and working temperatures ranging from 100 to 180 °C are commonly achieved. When the dough emerges from the die, the sudden fall in pressure causes rapid expansion and instantaneous generation of steam. Thus, the extrusion cooking process is characteristically one in which high temperature, pressure, and shear forces are imposed upon a starch-rich matrix, which has a relatively low water content, for a short time. All of these conditions are varied at will by the processor, in striving to optimize the texture and other properties of the final product.

There have been numerous empirical investigations of starch behavior during extrusion under different conditions, but no completely adequate theoretical account has been developed. Nevertheless, some useful generalizations are possible. Under low-moisture conditions, starch granules degrade or melt at a temperature higher than that of conventional gelatinization. This process can take place within the high-temperature environment of the extruder barrel. The high pressures and shear stresses of extrusion cooking increase the efficiency of gelatinization so that finished products often possess a completely amorphous and hence readily digestible starch matrix, which nevertheless has a very low water content. Moreover, it is not only starch granules which are degraded under extreme extrusion conditions. Scanning electron micrographs of cereal grains such as maize demonstrate a loss of structural organization within the extruded tissue, and an apparent absence of intact cell walls.

One of the most important attributes of dietary fiber is the structural organization which it confers upon lightly processed plant foods. There is strong evidence that intact cell walls tend to encapsulate gelatinized starch granules in conventionally cooked foods, and thereby slow the access of pancreatic enzymes so that starch digestion is delayed. This leads to a relatively slow rate of glucose absorption and an attenuation of the peaks and troughs of blood glucose which occur after meals. In studies with human volunteers, test meals consisting of breads processed by extrusion cooking have been shown to be more rapidly digested than conventionally cooked breads. The principal aim in the dietary management of diabetes mellitus is the maintenance of blood glucose at levels as close as possible to the normal range, and it has been suggested that diabetics should avoid extruded cereal products and snack foods. However, this principle has not been fully confirmed in patients consuming mixed diets over a long period of time, and there is little or no evidence that such relatively minor differences in carbohydrate availability are of any nutritional significance in healthy subjects.

Apart from the disruption of cell walls, dietary fiber also undergoes chemical changes during extrusion cooking. Under relatively mild conditions the total fiber content of extruded wheat flour remains more or less constant, but there is some increase in the soluble fiber content at the expense of insoluble fiber. Under more severe conditions a moderate increase in total fiber content has been reported. Clearly, this material is not true dietary fiber in the sense of cell wall polysaccharides. It probably represents a combination of Maillard reaction products and amylose-resistant starch complexes. The physiological significance of these substances is unknown, and they may be best regarded as artifacts.

Interaction of Protein and Carbohydrates

Under extreme extrusion conditions the disruption of complex carbohydrates extends to the molecular level. A proportion of the starch may be hydrolyzed to release glucose which then becomes available for further chemical reactions within the extruder. The most important of these is the Maillard reaction which occurs between reducing sugars and the free amino group of lysine, and gives rise to a very complex family of 'browning products.' Such compounds alter the flavor and color of the finished product and are responsible for many of the organoleptic consequences of cooking. However, Maillard reaction products are usually indigestible and, since lysine is an essential amino acid, any such loss of bioavailability has an adverse effect on the nutritional quality of protein. (*See* **Protein**: Interactions and Reactions Involved in Food Processing.)

The extent of lysine loss in extruded products depends upon the extrusion parameters and the composition of the dough. In general the biological availability of lysine is markedly reduced at temperatures in excess of 180 °C, but it is improved by increasing

the moisture content to more than 15% and by minimizing the presence of reducing sugars.

Losses of Micronutrients

There are relatively few foods which can be consumed without some form of processing or cooking. Indeed, as in the case of starch, mechanical and thermal processing is often essential in order to render nutrients available for absorption and utilization by the body. On the other hand a high water content, high temperature, and vigorous mechanical disruption tend to cause losses of vitamins and minerals by leaching or chemical reaction. Thus, a balance is needed to achieve the optimal nutritional value in a manufactured food, and this applies to the use of extrusion technology, as much as to any other form of industrial processing. (*See* **Vitamins**: Overview.)

Water-Soluble Vitamins

The water-soluble vitamins are a diverse group of reactive compounds. In general, a high moisture content and temperature will favor the destruction of such compounds during processing, but of course these effects can be minimized by a short duration. With its relatively low moisture content and short residence times, extrusion cooking has some advantages over other forms of processing, but these tend to be offset if the particular application requires a very high temperature and shear stress. This complex situation is reflected by the diversity of results reported for water-soluble vitamin losses after extrusion.

In general, thiamin is the B-group vitamin which appears to be most vulnerable to destruction during extrusion. On the other hand, riboflavin is retained well and is least susceptible to processing variables. In one set of experiments with extruded maize grits, for example, the recovery of added thiamin was only 46% under the most mild conditions used, and fell still further when temperature and screw speed were increased. In contrast, riboflavin recovery was 92% under the mild conditions and was relatively unresponsive to increases in temperature. However, increases in screw speed and water content did lead to a significant reduction in riboflavin content. In studies with an extruded crispbread formulation it has been shown that losses of thiamin and other B-group vitamins are linearly related to energy input. Similar considerations apply to folic acid, pyridoxin, and niacin, although these vitamins appear to be more stable. (*See* **Niacin**: Properties and Determination; **Vitamin B_6**: Properties and Determination; **Riboflavin**: Properties and Determination; **Thiamin**: Properties and Determination.)

Vitamin C is susceptible to oxidation during processing, cooking, or prolonged storage. Extrusion cooking, like other methods of processing, leads to losses of vitamin C from raw materials such as potato. Supplementation can be used to overcome this problem, but retention is maximized by supplementing after extrusion. At high operating temperatures the stability of vitamin C during extrusion cooking is favored by a low water content. (*See* **Ascorbic Acid**: Properties and Determination.)

Fat-Soluble Vitamins

Vitamin A activity may be destroyed during extrusion cooking, but the losses depend both upon extrusion conditions and the nature of the carotenoids in the commodity. Higher screw speeds have been reported to favor increased retention of β-carotene in maize meal. Vitamin E appears to be particularly susceptible to severe extrusion conditions, and losses as high as 66% have been reported for an extruded product containing wheat germ. There is little information available on the fate of vitamins D and K. (*See* **Retinol**: Properties and Determination; **Tocopherols**: Properties and Determination.)

Minerals

Unlike many other nutrients, suboptimal mineral status occurs relatively frequently in industrialized societies. This is particularly true for iron and calcium, which are essential constituents of red blood cells and bone matrix, respectively. Mineral nutrients are susceptible to aqueous leaching and to reduced bioavailability due to the formation of unabsorbable organic complexes. Extrusion cooking offers potential advantages over other forms of processing because of the generally low moisture content of extruded doughs. However, it has been reported that retention of the phytate in wheat bran, which forms indigestible complexes with zinc, iron, and calcium in the human small intestine, is higher in extruded products. This is probably because the phytase of cereals is destroyed more readily at the high temperatures of extrusion cooking. There is conflicting evidence as to whether this is of any nutritional significance in humans. (*See* **Bioavailability of Nutrients**; **Calcium**: Physiology; **Iron**: Physiology; **Phytic Acid**: Nutritional Impact; **Zinc**: Physiology.)

Investigations using techniques which model the release of bioavailable iron *in vitro* suggest that extrusion may help to solubilize iron in maize. Furthermore, flavoring substances and preservatives added

to extruded snack products in the final stages of production may act as enhancers of iron absorption.

Extrusion of abrasive commodities inevitably leads to destructive wear within the die and other vulnerable surfaces of the extruder. A measurable rise in available iron levels from this source has been observed in some samples of food extruded under experimental conditions. In theory, other metallic contaminants might enter extruded foods, but there is no evidence that the phenomenon constitutes either a benefit or a hazard in any commercially manufactured products.

Destruction of Antinutrients and Toxins

Although the most familiar role of extrusion cooking is in the manufacture of commercial food products in industrial societies, it also lends itself to the production of nutritious and palatable flours from such staple commodities of developing countries as amaranth and sorghum. In these applications the process has been shown to improve protein digestibility, probably by a combination of partial protein denaturation and the destruction of protease inhibitors. Furthermore, the safety of the product may be improved by the destruction of mycotoxins. This is particularly important in the case of peanut meal, where extrusion cooking in the presence of ammonia has been shown to reduce aflatoxin concentrations significantly. (*See* **Amaranth; Sorghum**.)

Microbial cells and spores are also rapidly destroyed at the temperatures achieved during extrusion cooking. The efficiency of sterilization depends upon a combination of temperature and residence time within the extruder. For example, in studies with the spores of the heat-resistant microorganism *Bacillus stearothermophilus* FS 1518 it has been shown that, at 165 °C, destruction of spores was complete after a residence time of 75 s, whereas at 182 °C a residence time of only 45 s was sufficient. One possible application of the technique is in the sterilization of spices, as an alternative to irradiation or treatment with ethylene dioxide. Black pepper has been successfully sterilized by this method. The extruded product is then reground and has been shown to be indistinguishable from conventionally processed pepper. (*See* ***Bacillus***: Occurrence.)

The Sensory Properties of Extruded Foods

From the point of view of the food manufacturer, one of the principal advantages of extrusion cooking lies in its ability to compress a series of manufacturing steps into a single continuous process. The extreme conditions of temperature, pressure, and shear which this requires also allow the texture and shape of the final product to be varied in ways which have been used to create a variety of foods with novel organoleptic properties. However, the physical conditions which favor the creation of expanded products with an attractive texture also encourage the loss of volatile compounds which make an important contribution to flavor and aroma. This is also a major concern in other applications of extrusion cooking, such as the sterilization of spices or cocoa, but in both these examples it has been shown that the extrusion conditions can be manipulated so as to achieve a practical balance of microbiological decontamination and preserved sensory properties. (*See* **Sensory Evaluation**: Sensory Characteristics of Human Foods.)

Formation and Loss of Volatile Flavorings

The extreme physical conditions associated with extrusion cooking encourage a variety of chemical reactions which can modify the sensory properties of the final product. This creates both problems and opportunities for the food scientist. The high temperature generated within the extruder barrel, coupled with the 'flashing' of superheated water as the extrudate emerges from the die, inevitably leads to the loss of volatile compounds which would otherwise contribute to the flavor and aroma. For this reason many products are subsequently flavored by enrobement with mixtures of oil and synthetic flavorings during the final stages of production. To overcome the cost and other limitations of this approach, various attempts have been made to stabilize flavorings added to the feedstock before extrusion. Techniques employed include careful selection of carrier materials, complexation with compounds such as cyclodextrin, or the use of multiwall microcapsules to protect flavorings from thermal damage and volatilization. These approaches have been successful under experimental conditions, but their wide application has been limited by the need to tailor sophisticated thermostable flavorants to suit each particular product and processing strategy. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Nonvolatile Flavoring Compounds

Despite its adverse effects on protein quality, the Maillard reaction is of considerable importance as a source of aroma, flavor, and colored compounds which contribute to the palatability of cooked and processed foods. Through its secondary and tertiary stages, the Maillard reaction gives rise to a large number of products. The volatility of these compounds varies, but many are known to be retained

within the extrudate during extrusion cooking and can contribute to the sensory quality of the final product. The pattern of Maillard compounds which is obtained is sensitive to the primary amino acid and sugar substrates available, and to temperature, water activity, pH, and residence time within the extruder. Since all of these factors can, in theory, be varied during the design of a product, there would appear to be considerable scope for 'designing' the flavor of the product, along with the texture and shape.

Conclusions

There is no evidence that extrusion cooking poses unique problems for the food manufacturer. All forms of food processing alter the nutritional quality of foods. These changes are frequently beneficial, but deleterious effects such as changes in protein quality or vitamin content will often occur. The balance of positive and negative changes has to be judged in relation to the role which the final product will play in the diet of the consumer. For example, a reduction in protein quality, perhaps in a cereal product which has been subjected to extreme extrusion conditions, will need to be taken much more seriously in an infant food intended for use in developing countries than it would in a snack food intended for occasional consumption in the industrialized west.

See also: **Ascorbic Acid**: Properties and Determination; ***Bacillus***: Occurrence; **Bioavailability of Nutrients**; **Carbohydrates**: Interactions with Other Food Components; **Curing**; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Phytic Acid**: Nutritional Impact; **Protein**: Interactions and Reactions Involved in Food Processing; **Riboflavin**: Properties and Determination; **Sensory Evaluation**: Sensory Characteristics of Human Foods; **Sorghum**; **Starch**: Structure, Properties, and Determination; **Thiamin**: Properties and Determination; **Tocopherols**: Properties and Determination; **Vitamins**: Overview; **Zinc**: Physiology

Further Reading

Alonso R, Grant G, Dewey P and Marzo F (2000) Nutritional assessment in vitro and in vivo of raw and extruded peas (*Pisum sativum* L.). *Journal of Agricultural and Food Chemistry* 48: 2286–2290.

Ames JM, Gur RC and Kipping GJ (2001) Effect of pH and temperature on the formation of volatile compounds in cysteine/reducing sugar/starch mixtures during extrusion cooking. *Journal of Agricultural and Food Chemistry* 49: 1885–1894.

Camire ME (1998) Chemical changes during extrusion cooking. Recent advances. *Advances in Experimental Medicine and Biology* 434: 109–121.

Camire ME, Camire A and Krumbar A (1990) Chemical and nutritional changes in foods during extrusion. *Critical Reviews in Nutrition and Food Science* 29: 35–37.

Cazzaniga D, Basilico JC, Gonzalez RJ, Torres RL and de Greef DM (2001) Mycotoxins inactivation by extrusion cooking of corn flour. *Letters in Applied Microbiology* 33: 144–147.

Martin-Cabrejas MA, Jaime L, Karanja C *et al.* (1999) Modifications to physicochemical and nutritional properties of hard-to-cook beans (*Phaseolus vulgaris* L.) by extrusion cooking. *Journal of Agricultural and Food Chemistry* 47: 1174–1182.

Mensa-Wilmot Y, Philips RD and Sefa-Dedeh S (2001) Acceptability of extrusion cooked cereal/legume weaning food supplements to Ghanaian mothers. *International Journal of Food Science and Nutrition* 52: 83–90.

Mercier C, Linko P and Harper JM (eds) (1989) *Extrusion Cooking*. St Paul, MN:American Association of Cereal Chemists.

F

FACTORY CONSTRUCTION

Contents

Environmental Considerations

M O Brolchain, National Standards Authority of Ireland, Dublin, Ireland

General Requirements

The suitability of a site and premises for the manufacture of safe food at an economic cost is the main consideration when deciding on the location and construction of a food factory. It is frequently quite difficult to meet these requirements. A number of factors can put the safety of food at risk. These factors include the external environment of the factory, the internal design and layout, the materials and finishes used in construction, and the security of the site. Also included is the visual impact, both externally and internally, as this can affect the attitudes of the workforce. To manufacture food economically, it is necessary, in addition to choosing the right process and equipment, to have the correct layout and construction. The layout of the process must be such as to minimize labor costs, and the premises must be constructed so that cleaning can be carried out efficiently and effectively. When designing a factory layout, allowance should be made for possible future expansion.

External Environment

Food can be contaminated by airborne material, birds, insects, and rodents. Hence, food factories should be sited in areas free from these possible sources of contamination.

Airborne material includes dust, microorganisms, gases, and particulate matter. Dust in factory air usually results from unpaved roads and yards, and handling of products containing dust in close proximity to the factory. The most common sources of airborne microorganisms are effluent plants, other food processing plants, and decaying vegetation. Discharges of gases and particulate matter into the atmosphere are usually associated with heavy industry but can result from the domestic or industrial use of coal. Discharges of gases can also be associated with industries that use solvents.

Birds, insects, and rodents must be excluded from food factories. They must therefore also be excluded, in as far as is practicable, from the immediate environment around the factory. To achieve this, the environment around the factory should not be attractive to birds, nor should there be favorable locations for the harboring and breeding of insects and rodents. Birds are attracted by trees and a supply of food. Stagnant or slow-moving water and overgrown vegetation provide favorable locations for insects and rodents. (*See* **Insect Pests**: Insects and Related Pests; Problems Caused by Insects and Mites.)

It can be seen from the above that there are a number of environmental factors that can affect the suitability of a particular location as a site for a food factory. These requirements are in addition to the normal requirements when choosing any factory site. A sufficient supply of suitable water and facilities for the disposal of effluent must also be available. (*See* **Effluents from Food Processing**: On-Site Processing of Waste; Disposal of Waste Water.)

Internal Environment

There are many factors that must be taken into account when designing the internal environment of a food factory, if the safety of food is to be assured and

the cost of maintaining a hygienic environment is to be minimized. The main factors are plant layout, plant construction, surface finishes, services, and personnel facilities. These are dealt with, in detail, in the following sections of this article. Badly designed or incorrectly installed processing equipment can have serious implications for food safety. Hence, care must be exercised in choosing and installing equipment. To obtain the ideal internal environment, the factory should be designed around the process and the equipment. (*See* **Plant Design**: Basic Principles; Designing for Hygienic Operation; Process Control and Automation.)

Plant Layout

Figure 1 shows the layout of a food processing factory. The principles used in this layout apply to all food factories. It can be seen from the figure that there is a logical flow of raw material from intake to final dispatch. The raw material preparation, the processing, and the packaging are carried out in separate areas, and there is separate storage for raw materials and for both finished product and packaging. It should be noted that access to the packaging area is via a changing room. This is necessary for access to areas where high-risk food is exposed.

High-risk foods are foods that may be eaten cold or eaten after warming only, and which have the potential to support the growth of food poisoning microorganisms.

It should also be noted that the windows are on the north side of the building. The purpose of north-facing windows is to prevent them from being a source of heat. The choice for the location of service equipment is important, if the installation and operational costs are to be optimized.

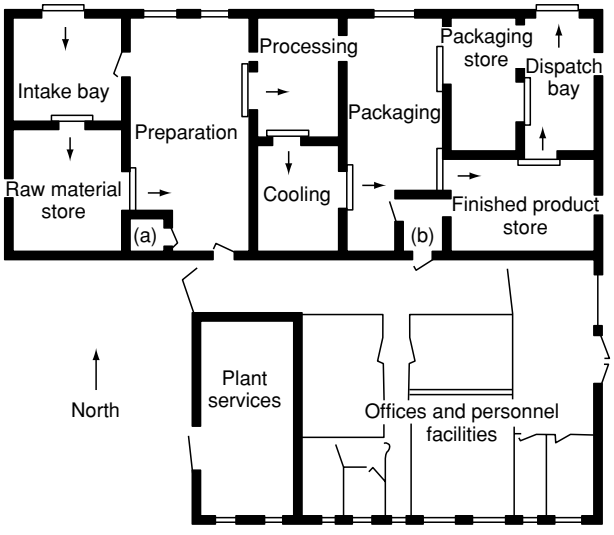

Figure 1 Plan of a food factory. Reproduced from Factory Construction: Environmental Considerations, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Constructional Requirements

As food must be processed in a hygienic environment, the cost of maintaining this environment should be taken into account, when designing and constructing a food factory. The additional capital cost associated with a well-designed factory is more than offset by the savings on cleaning costs.

When designing roof structures, external valleys should be avoided. Even partial blocking of such valleys will lead to the presence of stagnant water, which can be a source of food contamination. For similar reasons, gutters must be designed so that they can be easily cleaned. Drainpipes should be water-trapped at the entrance to storm or other drains. Access to roofs should be external rather than internal, so as to avoid contaminants from roof surfaces being brought into the premises.

Internal open-girder work should be avoided, and ceilings should be smooth and impervious. Ceilings should be accessible for cleaning, and ideally, the surface should be coved where ceilings and walls meet. In many food-production areas, the ceiling must have thermal insulation. This may be necessary either to prevent surface condensation or to reduce the energy costs of maintaining low environmental temperatures.

Walls should also be insulated, to prevent surface condensation. They should be free from rising damp and be protected, where appropriate, from damage by forklift trucks, by use of a guard rail. In modern factories, there is always a coved surface where walls meet the floor. It is also usual for there to be a coved surface where walls join at right angles. The purpose of these coved surfaces is to avoid corners and thus aid cleaning. Panels are frequently used to construct walls, and, if used, they must be adequately sealed. There should be no internal horizontal ledges, such as window sills. Such ledges lead to the accumulation of dust and foreign material. Where ledges are unavoidable, they should be sloped, at a minimum angle of 45 °.

Windows must be accessible for cleaning. In food handling areas, they should be double glazed to avoid condensation. In such areas, transparent polyvinyl chloride should be used instead of glass. Broken glass is a frequent cause of food contamination. Windows should only be used as a source of natural light and should be north-facing. Windows should never be used, in a production area, as a secondary source of ventilation.

External doors must be rodent-proof, which means that the gaps between the door and the frame, and the door and the floor, or riser, must not exceed 6 mm. External personnel doors, with the exception of emergency doors, should not lead directly into a production area. All personnel doors should be fitted with a self-closing device. Other doors, both external and internal, which are in frequent use, should be protected by a plastic strip curtain or an efficient electric air curtain, if they are not self-closing.

When designing floors, allowance must be made for the effect of the use of a forklift truck or the installation of vibrating equipment. In production areas, floors should be constructed so that they have natural slopes to drains. The use of heavy metal grids to cover open drainage channels should be avoided, as they are difficult to keep clean. In high-risk production areas, it is now being recommended that facilities should exist to enable the floor to be flooded with a sterilant solution.

Surface Finishes

When choosing surface finishes in a food factory, due cognisance must be taken of the type of food being handled and of the type of operations to be carried out in each area of the factory. These operations usually include environmental cleaning, and hence the use of detergents and, possibly, steam. There are no materials, therefore, that can be recommended for use as factory finishes, in all areas of a food factory. (*See* **Cleaning Procedures in the Factory**: Overall Approach.)

For floors, there are three broad categories of materials that may be used: concrete, tiles, and polymers. Floors must be impervious, easy to clean, and not slippery, even when wet. The suitability of a given flooring material may, also depend on its resistance to impact, wear and abrasion, and its inertness to chemical or biochemical attack. Ordinary concrete can be used in some areas, but it is subject to attack by strong acids and alkalis. Additives, or special aggregates such as granite chips, can be used to increase surface strength and nonslip characteristics. Ceramic or quarry tiles can be used for factory floors. However, floor tiles are usually only as good as the bedding, jointing, and grouting materials that are used. When laying tiles on floors, rubber latex cements, epoxy resins, and furane cements are often recommended. Polymers can be used on their own, as a surface finish or as additives to cement. Most polymers used in food factories are derivatives of epoxy, polyester and acrylic resins, and rubber latex. Care must be exercised in selecting a suitable polymer and in preparing the surface prior to laying the polymer.

Walls must be impervious, smooth, and easy to clean. To achieve this requirement, cement rendering, paints, and polymers are used. The choice is dependent on the operating environment. Failures in paints are usually attributable to inadequate preparation of the surface, to environmental conditions, or to unsuitability of the paint. Such failures can lead to contamination of food, as a result of flaking. Suitable finishes on walls can also be obtained by using cladding or paneling, provided that they are properly sealed.

Services

In modern food factories, it is usual to have all the service lines, with the exception of drainage, located between the ceiling and the roof. Services are then fed vertically to the various pieces of equipment. This design has many advantages. In particular, it makes the maintenance of hygiene in food handling areas more economical. In some factories, this type of design is impractical. In such cases, the service lines must be designed and installed so that they do not interrupt the smooth finish on walls and ceilings. They must also be designed so that they are accessible for cleaning. The problem with exposed service lines is that dust and food particles form deposits on the top surfaces of the lines. It is usually expensive to maintain such lines in a clean condition. When services are being installed, it is critical that service entry and exit points are adequately sealed, to prevent access by rodents.

As mentioned earlier, windows should not be used for ventilation in food factories. Ventilation must be sufficient to prevent condensation on walls, ceilings, and overhead structures. Airflows within the factory must be designed so as to avoid product contamination. Where there are single air entry and air exit points, airflows must be in the opposite direction to product flows. Air-intake points must be at least 1 m above both internal and external surfaces. They must be fitted, at least, with a fly screen and ideally with dust filters. It is now generally accepted that high-risk processing areas should be kept under positive pressure and that the air supply should be filtered to the required class. Where refrigerated air is required, for food-safety reasons, in production areas, low air velocities are required to reduce the chill factor. To achieve low air velocities, various refrigeration systems may be used.

In most areas of food factories, floor drainage is essential. All floor drains must be fitted with an effective water trap and a suitable grill. The sewerage lines must be adequate in size for effective drainage during washing. When designing factory drainage, allowance must be made for equipment, which requires a drainage outlet. Manholes should

not be present internally within the factory, and, if present, they must be doubly sealed. As food factory sewers have to be cleaned on occasion, manholes must be provided outside the factory to enable this cleaning to take place.

Facilities for Personnel

Adequate facilities must be provided for personnel. Such facilities include a canteen, locker room, toilets, and hand-washing facilities in production areas. The provision of suitable showers is also recommended. It is important that these facilities are well designed and can be maintained in a hygienic condition. This is necessary to maintain the correct hygiene ethos in the workforce. Toilets and canteens must be ventilated separately and must not lead directly into a production area. The doors leading into the lobby and into the toilets should be fitted with a push plate on the outside and a handle on the inside. Ideally, the lobby outside toilets, used by production personnel, should be of sufficient size to enable personnel using the toilets to hang up their protective clothing, prior to entering the toilet. Hand-washing facilities in production areas and in toilets used by production workers must be fitted with taps that are either knee-, foot-, or electronically operated, to prevent recontamination of hands after washing. Hand-drying facilities must be provided, and air driers are not recommended in production areas, as they could lead to the spread of aerosolized bacteria. Paper towel or cabinet roller towels may be used in such areas. Access to high-risk production areas should be restricted. Entry and exit should be via a changing room with hand-washing facilities.

Security

It is now becoming the norm to control access to food factory premises. This is usually achieved by fencing the factory site and having a single combined entry and exit point manned by security personnel. In some cases, this is supplemented by the use of television cameras and monitors for surveillance of the premises and the perimeter fences. Frequently, car parking is only permitted in a parking area external to the fenced site. Car parks can be located within the site, if under the supervision of the security personnel.

The purpose of this level of security in food factories is to reduce the level of pilferage and to prevent entry to the site by unauthorized personnel. The latter is of increasing importance, as a means of preventing either inadvertent or deliberate contamination of the company's products. Various systems can be used for the authorization of entry to the site of visitors, and their identification while on site.

Esthetic Design

In designing food factories, the emphasis has tended to be on functional design. The appearance of the buildings and site from an esthetic point of view has often been ignored. All food factories should be attractive and pleasing to the eye. The importance of a visually pleasing factory and factory site cannot be overemphasized. Such a site helps to create a better ethos in the workforce and better hygiene practices. It also helps to improve the image of the company projected to customers and prospective customers.

In summary, there are many factors that must be taken into account when designing a food factory if the external and internal environments of the factory are to be suitable for the production of safe food at an economical cost. Ignoring any one of these factors can result in the environment being unacceptable for food production.

See also: **Cleaning Procedures in the Factory**: Overall Approach; **Effluents from Food Processing**: On-Site Processing of Waste; Disposal of Waste Water; **Insect Pests**: Insects and Related Pests; Problems Caused by Insects and Mites; **Plant Design**: Basic Principles; Designing for Hygienic Operation; Process Control and Automation

Further Reading

Codex Alimentarius Commission – Joint FAO/WHO Food Standard Programme – *Recommended International Codes of Practice, CAC/RCP 1 to 32*. Rome: Food and Agriculture Organization of the United Nations.

European Commission (1980) Directive 80/778/EEC, *Quality of Water Intended for Human Consumption*. O.J. No. L229 30/8/80 p 11–28.

European Commission (1985) Directive 85/397/EEC, *Intercommunity Trade in Heat Treated Milk*. O.J. No. L226 24/8/85 pp. 13–29.

European Commission (1989) *Directive 89/214/EEC, Inspection of Fresh Meat Establishments*. O.J. No. L87 33/3/89 pp. 1–51.

European Commission (1991) *Directive 91/493/EEC, Conditions for the Production and Placing on the Market of Fishery Products*. O.J. No. L268 24/9/91 pp. 15–34.

Katsuyama AM and Strachan JP (eds) (1980) *Principles of Food Processing Sanitation*. Washington, DC: Food Processors Institute.

MAFF. *Proofing of Buildings Against Rats, Mice and other Pests*. Ministry of Agriculture, Fisheries and Food. London: Her Majesty's Stationery Office.

Sheppard WL, Jr (1981) Nagging floor problems. *Food Engineering International* April: 26–30, 35–36.

Smith DA (1983) Growing problems – mould in food plants. *Food Processing* January: 35–36.

Materials for Internal Surfaces

A C Jolly, Frodsham, Cheshire, UK

Introduction

As a result of the fact that the synthetic resin-based material has been used over the last 20 years in a rather ill-informed way, the rate of exploitation of the technology and properties offered has been slow, held back by the inevitable on-site failures. The resin-makers possessed no experience of on-site conditions (metereological, physical, and quality of labor); the architects, specifiers, and civil engineers usually received no professional education in polymers *per se* (in the UK, for example, 'resins' has nearly always been taken to mean 'epoxies'), and, finally, the contractor's labor force had no real understanding of the specific handling requirements, whilst the formulators of the products and systems were simply commercially keen to sell their wares. The tense of the above verbs is chosen deliberately in view of the fact that, at last, there appears to be a realization by all concerned that successful use of synthetic resin-based products requires a basic understanding of the polymers concerned and the preparational work and handling they require.

Background

Synthetic resins have been with us in a substantial way since the end of World War II. Before this time they had been seen to be having a great influence in the coatings field, as was indeed borne out later with the volume sales of alkyd-based paints, giving quicker tack-free times and a wide range of possible colors. Furthermore, the beginnings of molding materials other than Bakelite were appearing.

World War II itself had, of course, a gigantic catalytic effect upon the rate of development of resins for practical purposes. In the subsequent years they have gradually, and almost surreptitiously, made enormous inroads into the building industry. They have been used not only in paints and varnishes, but also in sealants, roofing, as foamed insulation, in cladding for the exterior and interior of buildings, adhesives, seamless wall and floor finishes (coatings and screeds) and, of course, in the upgrading of concrete (in terms of resistance to wear and carbonation), to name just some of the main examples.

There is absolutely no doubt that synthetic resins possess the properties and characteristics to improve the design, technology, and efficiency of building construction and refurbishment. What has been the greatest drawback is that too many disasters have been associated with their use. This has by no means been attributable to their deficiencies, but usually to other very significant factors.

Amongst these, the major factor is a complete lack of understanding of what synthetic resins actually are, how they should be used, and where they should be used. They have been wrongly specified and installed, and operatives have mishandled their installation.

The Resins and their Uses

Synthetic resins are organic (carbon-based) materials with molecules that can be made to link up with one another into larger units. These basic building blocks are called monomers and the larger groups they form are called polymers. There are many kinds of linking arrangements and many different ways of stimulating the link-up (or polymerization).

Emulsions are usually water-based and consist of resins most frequently based upon styrene, butadiene, or acrylic polymers (thermoplastic materials) which harden when the water or other solvent evaporates, allowing the polymer groups to coalesce.

Alternatives to thermoplastic emulsions are thermosetting polymers (sometimes dissolved in organic solvent) that undergo a permanent chemical change when they cure. The reaction that transforms the resin liquid or paste into a solid tends to continue long after the initial cure is complete. The result is a harder and more chemically impervious material than is the case with the emulsion types referred to above.

In general, and taking the above into consideration, synthetic resins are suitable where seamless surfaces are required to minimize the occurrence of trapped dirt or contamination, and to provide easily cleaned hygienic surfaces (i.e., good British Standard (BS) 4247 rating). Coupled with this facility are the factors of excellent serviceability, safety and, if applied correctly, esthetics. Furthermore, they can act as binders in screeds and provide, as in the case of polyurethanes (PUs), insulating and also roof stabilization foams, to mention just a few applications.

No one resin type will provide satisfaction in all situations encountered. What follows is not a complete statement of the advantages and disadvantages of the various types, but it will give a general idea as to how careful specifiers have to be.

Resin Species

Resins are as varied in their nature as food. Similarly, epoxy is a subdivision of the former, as cheese is of the latter, both being collective terms for a number of connected variants.

Specification of the correct resins requires as much understanding as that required by a diner choosing a menu for a successful meal: no one would order Blue Stilton to quench a thirst. Examples of the variations in the nature of some synthetic resins are as follows:

1. PUs give all-round properties resulting from their hardness, coupled with flexibility; the surface preparation required to avoid their delamination has to be more scrupulous than with, for example, acrylics; PUs possess excellent acid resistance.
2. In general, acrylics are used for hardness and, principally, color retention.
3. Polyesters carried in styrene offer quick cure and good chemical resistance to some chemicals, but produce a strong smell during application and need a temperature of at least 10 °C to cure satisfactorily.
4. Epoxy types (i.e., epoxy–hardener two-pack systems) are versatile in that the resultant properties can be varied by selection of the appropriate hardener; they undoubtedly have excellent resistance to alkalis and can be used in a solvent-free form, but their resistance to marring and certain chemicals is poor. (Marring is the phenomenon of marks being left on a surface through the moving contact of another over it. Epoxy-derived surfaces are prone to this marring effect.)
5. Methacrylates cure quickly and at low temperatures, and exhibit rapid hardening, but during application they give off a pungent smell which dissipates quickly. (They are used, with special care, extensively in Europe in food factories and breweries.) Again, chemical resistance properties vary with relation to specific chemicals.

The methods of curing of the above resins differ widely, from using the moisture in the atmosphere to requiring a peroxide catalyst (when it is known as a free radical mechanism).

Floors

There has long been a search for a floor finish that is as near perfection as possible, and the arguments concerning seamless surfaces and tiles have continued for some years.

In the years since 1945, seamless resin-based floor finishes have developed considerably from small beginnings. Initially they were mainly to be found in the form of clear varnishes or paints based on synthetic resins such as alkyds and phenolics, which gave harder coatings than the traditional types then available.

Using resins as binders for fillers and/or aggregates to produce polymer concretes or screeds (self-leveling or troweled) began to be feasible with the exploitation of the emerging technology associated with epoxy resin chemistry. Since then, other polymers have come to the fore in this applicational area; PUs, unsaturated polyesters, and methacrylates are perhaps the most prominent. The substitution of part of the water component of a cementitious formulation with a thermoplastic emulsion resin, in order to offer a moderately upgraded performance over the unmodified original, yielded an economical product (so-called polymer concrete but, more correctly, polymer-modified concrete).

When there is a new development in any area, there are usually two positive reactions amongst those who are to specify their use. One is to resist the trend, to adhere to the traditional methods or products and look for examples of failures that argue for this attitude. Caution is to be applauded and encouraged, but not if it means totally resisting cost-effective improvements. The opposite situation is perhaps more frightening in that a specifier welcomes a new development with open arms and wishes to use it at the slightest provocation. The disastrous consequences can be heightened by the fact that the first few installations may be successful and the next, failures.

It must be firmly established that no single synthetic resin will offer all of the properties that will cope with the varying in-service requirements presented to the specifier, whether it be in connection with a new floor area or a refurbished one. For example, the solvent resistances are quite different, as are the resistances to specific chemicals. Physical properties associated with impact, abrasion, elongation, and hardness also differ widely. Consideration during application of health and safety, and particularly odor, are also of considerable relevance.

The retardation of the rate of acceptance of all seamless resin-based floors has been caused by highlighting the failures rather than the greater majority of successes. These failures have, in the main, been attributable to the following, and stating them will not only be of historical interest but act as a cautionary and educational exercise:

1. Lack of basic understanding by specifiers of the substance, nature, and properties of synthetic resins and the formulated derivatives.
2. Lack of experience of many self-styled 'specialist' contractors in installing the systems on the one hand, and a lack of basic understanding of this particular medium to cope with the unexpected on site, on the other.
3. Lack of site knowledge by certain of the formulators of the systems, resulting in the systems not being 'workman-proof,' but only 'technologist-friendly'!

Indeed, in many cases the decision as to whether a coating (measured in μm) or a screed (measured in mm) should be laid is made without real understanding and often decided purely on cost – a sure recipe for disaster! A specific procedure for preparing the surface to take the finish is essential.

Coatings

In order simply to keep the level of dusting of concrete down, impregnation or coating with a clear resin is satisfactory. Acrylics, methacrylates, PUs or epoxy types, in unmodified or clear form, give good results. As far as impregnation or priming is concerned, the variable porosity of a concrete surface must first be taken into consideration; in this respect just one application of the resin may not seal the surface as, in areas of high porosity, there will be no resin present on or at the surface. In such cases additional applications have to be made until a satisfactory state is arrived at; if much extra solvent-carried medium has had to be added, the surface must be allowed to vent before proceeding further.

Clear or pigmented (paint) multicoat systems are suitable where the substrate is smooth or of very low profile. Antislip characteristics can be built in by casting fine sand in an intermediate film.

Perusal of trade and academic literature will provide information concerning the respective chemical resistance of the various resin species. Various key facts must be elicited and thence a specification drawn up in an informed way.

Screeds

If a substrate is pitted, potholed, or simply uneven, or has special requirements, a screed is needed. By substituting part of the water in a cementitious formulation with a thermoplastic emulsion, the so-called polymer screeds are constituted. These provide nondusting flat surfaces of varying hardness and certainly upgrade the 'straight' cementitious finish. However, they do not have the degree of chemical and abrasion resistance often required in industry. This is where other synthetic resins come in – epoxies, PUs, acrylics, and methacrylates. These are now well-known, and success is achieved by combining the most advantageous combinations of resin and aggregate (the selection of the latter plays a major part as far as wear and antislip properties are concerned). The main two methods of application are by trowel or, where a formulation where self-level is used, by pouring.

One of the requirements of the food industry, as mentioned before, is a sterile floor; in order to achieve this, the topping must give an excellent BS 4247 rating. One method, of course, is to steam-clean or to use virtually boiling water. Many systems, such as epoxies, do not possess a high enough heat distortion temperature to resist such daily treatment and, as a result, develop hairline cracks (thermal degradation). To some extent this can be overcome by introducing some elasticity, but the solution is to use a medium that does possess a high distortion point.

Another approach is to utilize a PU-modified concrete; this has been done and improvements are still being made (Figure 1).

Table 1 details the characteristics of a typical epoxy-based screed. Choosing from a wide variety of hardeners which, with the epoxy resin itself, form the two-pack system, a wide variation of properties can be achieved.

Table 2 details the characteristics of a typical PU-modified concrete used where high heat distortion temperatures are required, together with excellent physicals.

Table 3 proposes a scheme for the project manager to follow when installing a resin floor.

Finally, one very important matter, often overlooked, is cleaning or aftercare. The chemicals used previously may well not be at all appropriate for the resin-based surface, and it is the duty of the formulator, through the contractor, to identify the best means and detail it in the relevant technical data sheets. (*See* **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant; Overall Approach; Modern Systems.)

Figure 1 Polyurethane-modified concrete laid by Swansback. Photographic Unit of the Boots Company plc.

Table 1 Typical characteristics of an epoxy concrete (filler and/or aggregate and resin)

Characteristic	*Value*	*Units*
Compressive strength	55–110	N mm^{-2}
Compressive modulus	0.5–20	kN mm^{-2}
Flexural strength	25–50	N mm^{-2}
Tensile strength	9–20	N mm^{-2}
Elongation at break	0–15	%
Linear coefficient	25–30 × 10^{6}	per °C
Water absorption (7 days at 25 °C)	0.1	%
Maximum service (temperature under load)	40–80	°C
Rate of development of strength	6–48	h at 20 °C

Table 2 Typical characteristics of a polyurethane-modified concrete

Characteristic	*Ucrete HF*	*Concrete (124)*[c]
Density	2090 kg m^{-3}	2250 kg m^{-3}
Compressive strength	50 N m^{-2}	27 N m^{-3}
Elastic modules (compression)	1200 N m^{-3}	21 000 N m^{-2}
Tensile strength	6 N m^{-2}[a]	2.1 N m^{-2}[b]
Flexural strength	14 N m^{-2}	3.5 N m^{-2}
Coefficient of thermal expansion	1.5 × 10^{5} per °C	1.1 × 10^{5} per °C
Thermal conductivity	1.1 W m^{-1} per °C	1.3 W per m per °C
VICAT softening point	130 °C	
Surface resistivity	3 × 10^{8} Ω	4 × 10^{4} Ω

[a]1800 lbs in^{-2}.
[b]300 lb m^{-2}.
[c](124) means 1 part cement, 2 parts sand and 4 parts aggregate.

Table 3 Suggested scheme for managing a resin-based floor installation

Stage effort	*Item*
60% Evaluation[a]	History of substrate
	Past usage
	Service requirements
	Identification of appropriate system and preparation
	Installation constraints
	Specification settled
	Selection of proficient contractor
20% Installation I (preparation of the surface)	Contractor efficiency
	Achievement of suitable surface state
	Grinding, scrabbling, or shotblasting
	Rectifying deficiencies
	Lack of damp-proof membrane (DPM)
	Contamination
	Unevenness
	Slope
20% Installation II (application of system)	Contractor efficiency
	Scrupulous following of instructions
	Adequate barriers and notices
	Client cooperation
	Employee discipline

[a]Using, in part, a formulated questionnaire.

Walls

Much of the reasoning and statements contained in the Floors section of this article apply to walls. For example, a seamless surface is usually easier to keep sterile than a grout-containing tiled surface, although in many situations tiles are preferable. A good example of the latter is in the corridors of the London Underground system where gouging and graffiti are made very difficult by using specific and glossy tiles; but here sterility is not a significant factor.

In the food industry, as in industry as a whole, the wall finish must resist the chemical and physical attack of the inherent service conditions. For example, it is quite common to steam-clean and, in such cases, the basic resin in the finish is required to (but does not always!) possess a sufficiently high heat distortion temperature to resist the frequent expansions and contractions.

It has been firmly established by psychologists that the color and texture of surroundings have a great effect on the mental attitude of those in the vicinity. This is not surprising when one considers the effect the sun has on all of us when it comes out. Until 20 years ago, there was no pressure to arrange for the decor in the fitting shop to be the same as that in the chairman's office. Although this attitude has changed in essence, the desire could not be sensibly realized owing to the technical shortcomings of available prestigious wall finishes. In other words, the formulation of multicolored and textured systems of classical or upmarket appearance has never been a problem, but the technical inadequacies of the systems did not make them suitable for the more demanding service conditions found in food industry situations.

Multicolored, durable, and resistant coatings have been around for some time, but they have been paint-like and thus of flat profile (Figure 2); in addition, the application of more than one coat was required, thus involving considerable labor costs. They have also been associated during application with the unpleasant smell and fire risk characteristic of solvents. This has meant either causing aggravation amongst those working in the vicinity, or carrying out the particular project only when the building was totally vacated; the suitable time for application was therefore somewhat restricted. Textured systems have often had an upmarket appearance suitable for the chairman's office, but lacked the physical and chemical resistance properties needed for more aggressive service conditions, or they have had the necessary degree of resistance but not the necessary classical and upmarket appearance needed for the more prestigious sites.

Figure 2 A polyurethane-based wall finish by Sempol Surfaces Ltd in association with an epoxy seamless floor in a food factory.

Thus the way has been open for some time for systems of virtually universal suitability, with minimum hazards on application, and the ability to provide textured or smooth finishes. These are now gradually becoming available, especially utilizing the properties of PU-based systems which are no longer handicapped by yellowing on exposure to light.

Within reason, what applies to floors applies to ceilings. Thus, by using resin-based systems, a complete hygienic envelope can be achieved for floor, walls, and ceiling.

Roof Stabilization

Many factories in the food industry were built before World War II and consequently have tile roofs. These can become unstable with the passing years. Puresin-sprayed systems have been developed to secure and stabilize these structures internally without complete rebuilding.

Stability and insulation are achieved through technically advanced closed-cell, rigid polyisocyanurate (or related) foams which are tailor-made for the purpose with British Board of Agrément (BBA) approval. Application is by spray at right angles to the prepared underside of tiles or slates of a roof requiring refurbishment. The latter procedure, coupled with a very fast-curing or setting time, insures that foam does not run out from between the tiles or slates and down the outer surface of the pitched structure. Such foam has fire-retardant qualities, passing class 1 (BS 476 part 7, 1971). Typical characteristics of the system are shown in Table 4.

Table 4 Typical stabilizing foam properties

Property	*Value*
Closed-cell content	90% minimum
Thermal conductivity[a]	Initial 0.017 WMC Aged 0.020 WMC
Water vapor transmission[b]	10.6 μg per Nh[c]
Service temperatures	−50 °C to 70 °C
Density (overall)	42 kg cm^{-3} (approximate)
Tensile strength	117 kPa
Shear strength	77 kPa
Burning characteristics	See text

[a]See comments in text on roof ventilation and insulation.
[b]25 mm thick at 30 °C, 88% relative humidity.
[c]Nh are Newton hours.
All tests results according to BS 4370.
WMC, watts per square metre per Kelvin.

Figure 3 Stabilizing polyurethane Spraybond foam being applied. Courtesy of Liquid Polymers plc.

As with all resin-based systems, the installation must be carried out by trained and experienced contractors who can appraise the site as being suitable, recognize and carry out the vital preparation work, and spray the system correctly. In addition, they should insure that the roof space is adequately ventilated to avoid structural deterioration through condensation (Figure 3).

The foam is sprayed between (not over) the rafters and the purlins and over the battens (i.e., on their

inside face), leaving the outer face uncoated and thus allowing moisture evaporation. A foamed thickness of 25 mm is preferred. Before expanding (foaming), the liquid penetrates cracks and joints and, on curing as a foam, bonds the slates or tiles, thus increasing the stability of the structure chiefly against the force of winds.

In the industrial context, the duration of application will depend on the surface area of the structure which may, of course, be of considerable size. A single day is all a house-sized roof will take.

As already mentioned, a roof thus treated will be sealed against wind, driven rain, and snow, i.e., storm conditions. However, it must be noted that this will not prevent leaks owing to cracked or displaced tiles, slates or faulty flashings, valleys, and gutters. These are avoided by correct, appropriate preparation of the roof before refurbishing with the system.

Being one of the best insulants commercially available, the foam, when of the order of 25 mm thick, achieves a degree of improved insulation that should be approximately 0.8 $W\,m^{-2}$ per °C. Because of this, coupled with good ventilation, the chances of condensation are reduced.

There are other attractions associated with this method of refurbishment. It is carried out at a lower cost than reroofing. The application is easy, being from inside the roof, so that no scaffolding is required. Furthermore, there is minimum disturbance of below-roof activity during the initial installation.

Final Comments

Most problems stem from lack of knowledge of the nature of the synthetic resin involved. Such knowledge must be acquired through the medium of either an informed contractor, fair-minded suppliers who are prepared to say 'you need a system based upon a type *x* resin which we do not make', or an independent resin consultant. The academic institutions can help considerably by including far greater input on the basic subject of synthetic resins in their syllabuses for graduate courses in architecture, civil engineering, and surveying.

An informed, experienced, specialist contractor is essential and can often be identified by the fact that the company has existed for a considerable number of years and can offer visits to past installations or, at least, produce a substantial list of case histories and offer good postinstallation technical service. The basic technical knowledge referred to above is likely to be present if the contractor formulates his or her own systems, and is not purely a 'menu-reader' of material suppliers' literature which cannot hope to cover all of the potentially threatening situations arising on site.

For success, not only should a system based upon the correct resin be selected, but also a good major percentage of the effort put into the project should be in preinstallation activity by way of full appraisal of the site, coupled with the resultant, specifically appropriate substrate preparation. Sadly, all too frequently this is carried out after, for example, a floor finish has failed, and only then do protimeters and the like appear on the scene! A protimeter is an instrument which measures the degree of moisture in a surface (concrete) by means of a probe and the reading on a dial of a small current of electricity which indicates the percentage moisture. Increasing the depth of the hole into which the probe is placed allows moisture levels at different depths to be determined – so quantifying 'rising damp.'

See also: **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant; Overall Approach; Modern Systems

Further Reading

Jolly AC (1965) Polyurethane floor coatings. *Flooring and Finishing News* 10(5): 33, 34, 37, 38.

Jolly AC (1968) Seamless flooring in the can. *Cleaning and Maintenance* (4): 22, 23, 25, 26, 53.

Jolly AC (1979) A hundred years of resins. *Polymer Paint and Colour Journal* 169(4001): 514, 515.

Jolly AC (1982) Seamless floors versus tiles. *Polymer Paint and Colour Journal* 172(4062): 10.

Jolly AC (1985) Resin floor finishes. *Architects' Journal* 181(6): 63–66.

Jolly AC (1987) Educating for change. *Construction Surveyor* 21(64): 9, 10.

Jolly AC (1989a) Elastic, seamless roof finishing. *Architects and Surveyors Institute Journal* 1(1): 29, 30.

Jolly AC (1989b) Seamless resin floors in perspective. *Building, Trade and Industry* 5(4): 28, 29.

Jolly AC (1989c) Stabilising roofs using foam technology. *Dairy Industries International* 54(10): 36, 37.

Jolly AC (1990a) Getting the seamless floor right. *Building and Maintenance* May/June, pp. 11, 12.

Jolly AC (1990b) Are the technological advances available being fully utilised? *Building and Maintenance* Sept/Oct, pp. 7, 8.

Jolly AC (1990c) The dangers of surface knowledge. *Dairy Industries International* 56(11): 25, 27.

Factory Inspection *See* **Hazard Analysis Critical Control Point**

FAMINE, STARVATION, AND FASTING

E M Widdowson[†] **and A Ashworth**, London School of Hygiene and Tropical Medicine, London, UK

Background

Despite man's ingenuity and technological advances, it is a sad reflection that at the start of the twenty-first century, famine and starvation still abound, often with such catastrophic loss of life that it can be measured in millions. Indeed, 'food security' has become an industry, and 'humanitarian assistance' the vocation of a burgeoning number of organizations who do nothing else. It is a damning indictment of our times that when the medical relief organization Médecins Sans Frontières was awarded the Nobel Peace Prize, they were working in over 80 countries.

We shall examine the causes of famine and starvation, both past and present, and some of the physiological changes that have important consequences for those looking after the afflicted. We shall also look at fasting and starvation in the midst of plenty, in particular anorexia nervosa and bulimia.

Causes and Effects of Famine

Famine can be defined as a prolonged scarcity of food causing many in a district or country to die from starvation and/or disease. In communities dependent on subsistence farming, 'hungry seasons' are not uncommon, but coping strategies exist to mitigate the adverse effects and insure survival. For example, nonessential items may be sold and debts called in, and there may be greater use of wild foods. What distinguishes 'famine' from 'hunger' is that in famine, the usual coping strategies are insufficient. Thus famines are characteristically accompanied by extreme measures that may disrupt future livelihoods, for example, the selling of animals, land, and other essential items. In a famine, families are often driven to eat unconventional items, such as rodents, beetles, worms, roots, leaves, bark, grass, candles, and leather.

Famines have occurred from time to time throughout history, and there are many records of them in early writings, such as the Bible. In general, they were caused by crop failure due to droughts, floods or pestilence (e.g., locusts, blight). In Europe in the Middle Ages, they were often accompanied by outbreaks of typhus, cholera, typhoid, dysentery, smallpox, measles, and scarlet fever, and famine led to their spread and severity. In more recent times, misguided political gambits have caused disastrous food shortages, as witnessed by the famines in the former Soviet Union and China, where privately owned small farms were merged into collectives and crops procured by the State to provide the wealth to support industrialization. Some 30 million people are estimated to have starved to death in China during Chairman Mao's Great Leap Forward.

Starvation may be used as a political tactic and knowingly engineered (as in the internment camps of Belsen). Incongruously, countries gripped by famine often continue to export staple food crops, as was the case in the Irish famine of the 1840s and more recently in the famines in Ethiopia and Sudan. Mass movements of populations in search of food or fleeing from conflict situations have become commonplace on our television screens. Tented camps and humanitarian aid may be a refuge from starvation, but thousands of people with weakened immune systems living in close proximity are vulnerable to epidemic diseases, and many die as a result. Table 1 shows some of the twentieth century famines and the estimated number of deaths.

Increasingly, famines are linked with ineffective political structures, armed conflicts, environmental degradation, and ethnic or religious unrest emanating from disregard of civil and human rights. The effects are wide-ranging. The use of land mines, for example, limits access to local markets and hampers food relief efforts in the short term, but also continues to kill, maim, and disrupt agricultural production in the long term, as seen in Mozambique, Angola, and Somalia. Forced conscription of young men for combat adversely affects food production and food security in rural areas, and arms purchases divert foreign exchange from already under-resourced health and development budgets and deter inward economic investment. Evidence of the effects of climate change can be seen in unusually prolonged periods of low rainfall at one extreme and increased

[†]Deceased.

Table 1 Selected twentieth century famines.

Year	*Location*	*Number of deaths*	*Notes*
1899–1900	India	1 million	Drought
1915–1918	Germany	200 000	War
1918–19	Uganda	4 400	
1920–21	North China	500 000	Drought
1921–22	Russia, Ukraine and Volga region	1.3–5 million	Drought
1928–29	China, Shensi, Honan, and Kansu	3 million	
1932–33	Russia	3–10 million	Collectivization
1941–43	Greece	450 000	War
1941–44	Leningrad	1 million	War
1941–42	Poland, Warsaw	43 000	War
1943	Ruanda-Urundi	35 000–50 000	
1943–47	India, Bengal	3 million	Drought and war
1958–62	China	30 million	Collectivization and natural disasters
1967	India, Bihar	Relief operations mitigated the effects of the disaster	Drought
1971–73	Ethiopia	1.5 million	Drought
1988	Sudan	250 000 in Bahr El Ghazal	War
1998	Sudan	60 000	War

[a]Adapted from Alamgir M (1981) An approach towards a theory of famine. In: Robson JRK (ed) *Famine: Its Causes, Effects and Management*. London: Gordon and Breach Science and Watson F (1999) One hundred years of famine – a pause for reflection. *Field Exchange* Issue 8: 20–23.

frequency of hurricanes and floods at the other. These wreak havoc to crops and livelihoods, decrease soil fertility, increase erosion, desertification and salination, and disrupt transport routes. These factors, however, are associated with famine only where populations are chronically poor. Hence, combating poverty is a major concern in famine prevention and mitigation.

Structural and Physiological Consequences of Undernutrition and Starvation

Starvation implies that little or no food is eaten. Undernutrition arises when some food is eaten, but the amount is insufficient to meet physiological and metabolic requirements. In many ways, the effects of starvation and undernutrition are the same, and the differences are those of degree. It goes without saying that a starving or undernourished person loses weight, but the organs and tissues lose weight at different rates. The least essential tissues suffer first; hence, adipose tissue is mobilized and the fat oxidized to provide the energy necessary for the metabolic processes that must go on if life is to be preserved.

Some loss of nitrogen is inevitable despite the increased production of free fatty acids, and this comes, to a large extent, from the proteins of skeletal muscle, leading to shrinkage of the muscle fibers. The breakdown of muscle proteins provides the amino acids alanine and glutamine, whose carbon skeleton can be used to make glucose. The skin also loses protein, both structural and cellular, and becomes thin and easily infected. The loss of its protein, together with the loss of subcutaneous fat, makes the skin hang loosely on the bones.

The internal organs lose weight, the liver sometimes a great deal. In edematous malnutrition, however, the liver may increase in weight by as much as 50% owing to the accumulation of fat, which is caused, at least in part, by a reduced synthesis of lipoproteins that transport triglycerides out of the liver and into plasma. The alimentary tract loses weight and becomes thin, and there are marked changes in the gut mucosa, especially the villi, which become flattened and broader, and have reduced levels of digestive enzymes. The heart and kidneys tend to lose weight in parallel with the body, and their function is impaired. The brain is preferentially protected, although there is a reduced density of neuronal synapses and dendrites.

Healthy, well-nourished individuals operate with considerable 'spare capacity.' Thus, our bodies are able to respond when we run very fast or eat a very large meal. This excess capacity, however, is expensive to maintain, and when food intakes are insufficient, it is reduced to conserve energy. This 'down-sizing' of the body's functional capacity affects every system, organ, and cell, and is termed 'reductive adaptation' (see **Figure 1**). One example of reductive adaptation is the slowing down of the sodium pump in cell membranes. The purpose of this pumping activity is to maintain the correct concentration gradient, and it normally accounts for about one-third

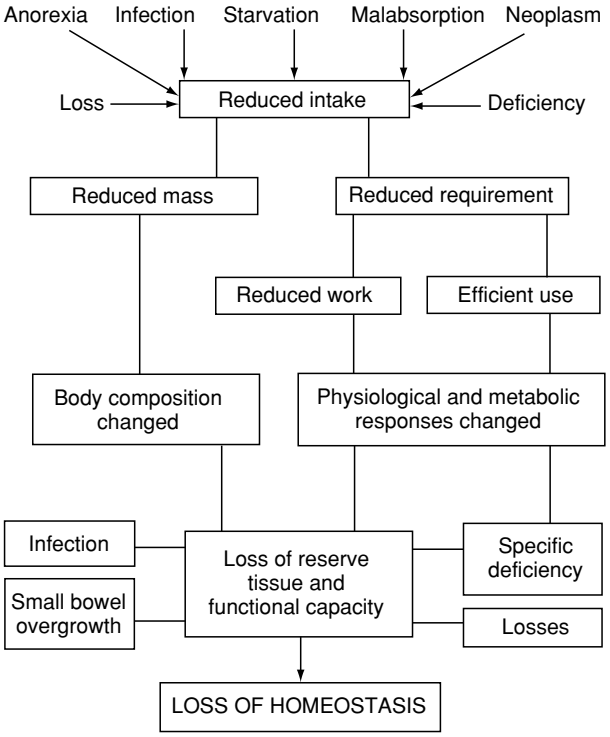

Figure 1 Schema showing the changes that occur in severe malnutrition. When malnutrition is prolonged, the adaptive changes are insufficient to maintain homeostasis. Reproduced with permission of Oxford University Press from Golden MNH (1996) Severe Malnutrition. In: Weatherall DJ, Ledingham JGG and Warrell DA (eds) Oxford Textbook of Medicine, Vol. 1.

of the basal energy expenditure. Although energy-sparing, this slower pump rate is not entirely beneficial, as it allows sodium to accumulate inside cells and potassium to be lost, thus contributing to potassium deficiency. There are several serious consequences of potassium deficiency, including fluid retention and edema, and impaired cardiac function. Other functional changes in starvation include a slower heart rate, reduced ability of the kidneys to excrete excess sodium and excess fluid, which increases the risk of heart failure, and reduced ability to make glucose, increasing the risk of hypoglycemia and hypothermia. Thus, there are both positive and negative consequences of 'reductive adaptation'.

The resting metabolic rate per kilogram bodyweight is not greatly reduced in moderate undernutrition, but in a starving person, it ultimately falls, as does body temperature. In moderate undernutrition, the digestive tract functions normally, provided there is no infection, but in a starving person, diarrhea may become severe. This has been attributed, at least in part, to the colonization of the normally sterile upper intestine with bacteria, which then produce toxins, damaging the gut mucosa and increasing fluid secretion into the gut. The dehydration resulting from this may prove fatal. In addition, some of these bacteria deconjugate bile salts, reducing the body's ability to absorb fat.

The survival time of an adult who takes no food depends upon the availability of clean water. If plenty of water is available, a previously well-nourished adult can survive without any food for about 60 days – the exact time depending partly on the amount of fat in the body. Starving or near-starving infants and young children are in a far worse plight, particularly if they are also short of water, because their requirements for energy and water per kilogram bodyweight are greater than those of adults.

Extreme wasting occurs among hospitalized patients, the causes varying with the nature of the disease or type of trauma. Some conditions, including cancer, markedly depress the appetite leading to semi-starvation as a result of low food intake. Some conditions lead to actual loss of nutrients, for example gastrointestinal diseases and postoperative drainage. Injury and infections trigger an increased demand for nutrients to fuel the immune response, including amino acids for the synthesis of acute phase proteins. These are derived from muscle protein breakdown, causing wasting. Malnutrition in hospital, however, is also often the result of prolonged underfeeding due to inadequate or inappropriate nutritional support. Even in affluent countries, as many as 50% of hospitalized patients are malnourished.

Refeeding after Undernutrition and Starvation

When World War II ended in Europe in May 1945, several groups of doctors and nutritionists went to the liberated camps to supervise the feeding of the starving inmates. It was generally believed that the digestive tract would not be able to tolerate food for some time, and that intravenous feeding would be necessary in the early stages. This proved not to be true. The previously starving men had good appetites and took very large amounts of food (25–42 MJ, 6000–10 000 kcal per day). They gained about 10 kg in 6 weeks, and the response was similar in young and old men. The same is true for severely malnourished children. After a few days of treatment, they become very hungry and will double their food intake and gain weight very rapidly. After a few weeks, when their weight deficit has been corrected and they approach a normal weight for their height, their appetites and food intakes revert back to levels appropriate for well-nourished children of a similar size.

The physiological and metabolic changes that occur in starvation have to be taken into consideration

during refeeding, otherwise death from cardiac failure, infection, hypoglycemia, and hypothermia are likely. Treatment is considered in two main phases: a short stabilization period to treat infections and correct the main metabolic disturbances (e.g., fluid and electrolyte imbalance, and micronutrient deficiencies) and a longer rehabilitation period to replace lost tissues. Many humanitarian aid agencies looking after starving populations have a very organized system of identifying those who are severely malnourished so they can prioritize them for supervised refeeding. Aid agencies often use food preparations that have been specifically formulated for severe malnutrition to provide the correct amounts of energy, protein, potassium, magnesium, and micronutrients. In hospitals and nutrition rehabilitation centers, suitable preparations can be made from milk, sugar, oil, and an electrolyte/mineral mix, and/or modified local foods.

The essential characteristics of successful dietary treatment in the stabilization phase are small, frequent feeds day and night that provide just enough energy and protein to meet maintenance needs. This will halt any further loss of tissue and, together with provision of potassium, magnesium, and micronutrients, will restore normal functioning of the deranged metabolic machinery. In the rehabilitation phase, the aim is to restore the lost tissues. This involves providing large amounts of the nutrients that provide the building blocks for tissue synthesis, and high energy intakes to cover the energy cost. Thus, the aim in this phase of rapid catch-up growth is to encourage children to eat as much as possible and to provide them with diets that are high in energy and protein, whilst continuing to give extra potassium, magnesium, and micronutrients.

Fetal Undernutrition

The question sometimes arises as to whether an infant born small for its gestational age because of undernutrition before birth will remain small. This is difficult to answer because the genetic make-up of human individuals is so diverse, and genes, as well as nutrition, influence the rate of growth. Animal studies with pure-bred strains provide some information. The 'runt' pig, born much smaller than its littermates, is unlikely to catch up to its larger siblings, and rats that are born very immature, when undernourished during the suckling period, take less food after weaning and remain small for the rest of their lives. There is some evidence that this is also true for small newborns in developing countries, especially if their smallness is because they are short in length, rather than thin. It is difficult, however, to separate the effect of fetal undernutrition from the growth-limiting effects of being raised in the same environment that detrimentally affected fetal growth. Interestingly, no long-term effects on adult size were found in those individuals whose mothers suffered 6 months of starvation in the Dutch famine of 1944–45 but then were raised with a good standard of living. There seems little doubt, however, that the earlier undernutrition is imposed, the more likely it is that the effect will be permanent. Barker and colleagues have suggested that retarded growth of the developing fetus is associated with a number of chronic conditions in adulthood, including diabetes, hypertension, and ischemic heart disease.

Role of Fasting in Health and Disease

Some people fast even when they are healthy and there is plenty of food available. The reason may be political: thus, 'hunger strikes' are used by political prisoners and others from time to time to draw attention to their cause. Some fast for religious reasons. Religious fasting is usually of a short duration. Ramadan, for example, held in the ninth month of the Muslim year, involves going without food or drink between sunrise and sunset for the month. This short period of abstinence is likely to do little harm, and in any case, those who might be injured by it (e.g., pregnant women) are exempt from the fast.

A more serious problem is anorexia nervosa. This is a psychiatric eating disorder that tends to run in families and appears to be increasing in prevalence. The sufferers, usually adolescent girls and young women, are obsessed with the idea that they are too fat, and that they must lose weight and remain very thin. They may have been overweight in the past, but not necessarily so. They often come from better-off homes, where there is no shortage of food. Long-standing cases may repress the sensation of hunger, but for others, the sensation persists, and so the term 'anorexia' is misleading. Sometimes, the desire for food overcomes them; they eat a great deal and then deliberately induce vomiting; this condition is called bulimia. Or they may resort to laxatives and diuretics. The effects of insufficient food are the same as those already described, and death from starvation is not unknown. One of the symptoms is often amenorrhea. This is also characteristic of ballet dancers and gymnasts who deliberately eat too little food in order to maintain an acceptable size and shape. (*See* **Anorexia Nervosa**; **Bulimia Nervosa**.)

If the patient with anorexia nervosa can be persuaded to take more food, which is not easy, then physical recovery and restoration of weight take place just as in others who have starved and then been refed.

See also: **Anorexia Nervosa**; **Bulimia Nervosa**

Further Reading

Action Against Hunger (2001) *The Geopolitics of Hunger, 2000–2001: Hunger and Power*. London: Lynne Rienner.

Alamgir M (1981) An approach towards a theory of famine. In: Robson JRK (ed) *Famine: Its Causes, Effects and Management*. London: Gordon and Breach Science.

Barker DJP (1994) *Mothers, Babies and Disease in Later Life*. London: BMJ.

DeRose L, Messer E and Millman S (1998) *Who's Hungry? And How Do We Know?: Food Shortage, Poverty, and Deprivation*. Tokyo: United Nations University Press.

Dreze J, Sen A and Hussain A (eds) (1995) *The Political Economy of Hunger*. Oxford: Clarendon Press.

Garrow JS, James WPT and Ralph A (eds) (2000) *Human Nutrition and Dietetics*, 10th edn. London: Churchill Livingstone.

Golden MHN (1996) Severe malnutrition. In: Weatherall DJ, Ledingham JGG and Warrell DA (eds) *Oxford Textbook of Medicine*, 3rd edn., vol. 1, pp. 1278–1296. Oxford: Oxford University Press.

Keys A, Brozek J, Henschel A, Mickelesen O and Taylor HL (1950) *The Biology of Human Starvation*. Minneapolis, MN: University of Minnesota Press.

Von Braun J, Teklu T and Webb P (1999) *Famine in Africa: Causes, Responses and Prevention*. Baltimore, MD: The Johns Hopkins University Press.

Waterlow JC (1992) *Protein Energy Malnutrition*. London: Edward Arnold.

Watson F (1999) One hundred years of famine – a pause for reflection. *Field Exchange* Issue 8: 20–23.

Widdowson EM (1951) The response to unlimited food. In: *Studies of Undernutrition, Wuppertal 1946–9. MRC Special Report Series No. 275*. London: Her Majesty's Stationery Office.

World Health Organization (1999) *Management of Severe Malnutrition: A Manual for Physicians and other Senior Health Workers*. Geneva: World Health Organization.

FAO *See* **Food and Agriculture Organization of the United Nations**

FAT SUBSTITUTES

Use of Fat-replaced Foods in Reducing Fat and Energy Intake

D L Miller, Central Soya Company Inc., Fort Wayne, IN, USA
L J Cheskin, Johns Hopkins University, Baltimore, MD, USA

Introduction

The availability of 'fat-free,' 'low-fat,' and 'reduced-fat' foods has skyrocketed in recent years. This rise in the number of fat-modified products is due in large part to the development of 'fat replacers' which are substances that replace one or more of the functional qualities of fat in a food without supplying all of calories associated with fat.

The safety of using fat replacers has received much attention, but comparatively few studies have investigated how these products will influence food intake and energy regulation in humans. Until recently, few studies examined the effects of varying the level of fat in foods on energy intake and body composition. Indeed, until the mid-1980s, there was relatively little emphasis on the role of dietary fat in obesity, and the technology for formulating palatable reduced-fat foods was limited. Thus, only recently have studies assessing the effectiveness of such substances in reducing both dietary fat and energy intake been undertaken.

Despite this paucity of published data, some nutrition professionals and the general public alike assume that using fat-replaced products will bring about automatic reductions in fat intake. We know very little, though, about how consumers will use fat-replaced foods. For example, fat-replaced foods may be substituted for higher-fat versions of foods (using low-fat mayonnaise instead of regular mayonnaise), and/or substituted for 'forbidden foods' (eating fat-free potato chips when regular potato chips had been omitted from the diet), or, finally, used as a license to increase intake of other types of foods ('If I use the fat-free salad dressing I can have a piece of cheesecake for dessert'). If fat-replaced foods are used ineffectively, this could potentially result in increases in

energy or fat consumption instead of the intended decreases. Also, it is debated whether the overconsumption of dietary fat in and of itself leads to negative health outcomes, or if it is the resulting increase in overall energy intake due to the overconsumption of dietary fat that contributes to adverse health outcomes. Unfortunately, the message that much of the general public has received is: 'I can eat as much food as I want as long as it is low in fat or fat-free.'

This article summarizes the existing scientific literature regarding the effects of adding low-fat/fat-replaced foods to the diet on fat and energy consumption, and discusses the efficacy of using fat replacers as a means of reducing dietary fat and total energy intake.

Background and Significance

In the developed world and, recently in the developing world, high consumption of dietary fat has been epidemiologically linked to a variety of medical conditions, including obesity, coronary artery disease, and certain types of cancer. Many regard it as the top dietary problem in the USA. Currently, dietary fat comprises nearly 34% of the energy content of the American diet. Most health guidelines recommend that no more than 30% of daily energy be derived from dietary fat in order to reduce the incidence of related morbidity and mortality.

Perhaps the most obvious method of decreasing the percentage of energy from fat is to substitute low-fat foods for high-fat foods. However, it is difficult for many people to restrict their food choices to achieve this end. Laboratory-based experiments suggest that, for most people, high-fat foods are overeaten because they are highly palatable. When a substantial amount of fat is removed from the diet, the resulting menu is often bland and monotonous. Even those whose health is directly threatened (postmyocardial infarction patients, those with type 2 diabetes or hypertension) find it difficult to maintain long-term compliance with a reduced-fat diet. The development of fat-replaced foods has the potential to reduce fat and energy consumption by satisfying the preference for the taste of a high-fat diet while consuming low-fat foods. The recent availability of highly palatable reduced-fat or fat-free foods offers consumers new choices, but because there have been few controlled studies of how these products will be used, questions remain about their efficacy in reducing dietary fat intake.

Several issues to consider when assessing products made with fat replacers are:

1. Why do many people eat more dietary fat than the recommended 30% of total energy, and will fat-replaced foods satisfy people's desire to consume high-fat foods?
2. Will consuming foods made with fat replacers aid in lowering the intake of fat and energy, or will the fat and/or energy reduction be compensated for in subsequent episodes of eating?
3. Will these products be used in place of regularly eaten foods, as a license to eat previously 'forbidden foods,' or to allow for increased consumption of other foods?

Why is Fat Overconsumed?

Palatability

Fats contribute much of the texture, flavor, and odor of a variety of foods. Although not usually consumed in pure form, fat is readily consumed as part of complex foods containing the other macronutrients (protein and carbohydrate) in varying amounts. Thus, it is difficult to isolate a fat-specific preference in foods. Fats contribute a number of sensory properties of food. First is the olfactory perception of volatile fat-soluble molecules that impart the characteristic aroma to many foods. Second, fats endow foods with certain textures (such as softness, oiliness, elasticity, flakiness, viscosity, or smoothness) and mouth feel, (the food's distribution in the oral cavity during chewing and swallowing). These textural and mouth-feel characteristics enhance the richness of food flavor, and positively influence the palatability of the diet.

A number of experiments have studied the nature of diet palatability. Preferences for sweet/fat mixtures such as milkshakes, cake frostings, and ice cream have been identified by sensory panel testing. In one sensory study, the palatability of sweetened, fat-free milk and unsweetened heavy cream were both rated relatively low, but the combination of sugar and fat in sweetened heavy cream was highly appealing. Anecdotally, people who 'crave' sweet desserts often describe themselves as 'carbohydrate cravers;' however, it is fat, not sugar, that provides the majority of the energy in sweet, rich desserts. In addition, not just sweetened, high-fat foods are considered highly palatable. A survey of US military personnel found that the most preferred foods were 'savory' high-fat foods such as steak, French fries, and whole milk. Some of the least preferred foods in this survey were vegetables, skim milk, and diet soda, which are very low in fat. Other surveys of attitudes toward dietary fat also indicate that highly preferred foods often have a high-fat content. In these studies, taste is the primary reason given for the selection of a particular food. Because fat imparts the textural and mouth-feel

characteristics associated with high palatability, high-fat foods are often chosen.

Can Fat Preference be Changed?

The question of whether our preference for high-fat foods is innate or learned is of great importance. If it is learned, it can presumably be more easily altered than if it is innate. It is believed that the preference for sweet taste, for example, is innate. A preference for sweetness was useful in identifying foods that were safe to eat, and may have aided survival for our nomadic ancestors. It follows that an innate preference for fat may have also been adaptive for survival: consumption of a dense, easily stored energy source would provide an energy reserve for periods of scarcity. Unfortunately, if humans did acquire adaptations that favored high fat consumption, these adaptations have become maladaptive in today's society, which is characterized by an abundance of fatty food and opportunities to consume energy above our metabolic needs.

There is some evidence in favor of an innate fat preference. Children generally display preferences for high-fat foods. In experiments with high- and low-fat yogurt shakes, however, it was found that preferences for new foods high in fat content can be learned, and that conditioning of these preferences is the result of the postingestive consequences of consumption. It has been contended that Pavlovian or associative conditioning is central to the acquisition of food preferences. Because high-fat foods are palatable and satisfying, which are positive consequences of consumption, children may learn to like these foods. Such foods are also often used as treats or rewards for children, which may magnify this preference.

Animal studies suggest that, at least in rats, fat may be preferred at an early age. Appetite for fat was measured in 12–15-day-old infant rats. Intake of an oil emulsion solution was nearly as high as intake of a dilute sucrose solution ($0.03\,\mathrm{mol\,l^{-1}}$) or a milk formulation similar to rats' milk. It was concluded that the taste for fat is acquired at an early age and fat is as pleasant to the pups as sweet solutions and mother's milk.

In humans, however, there is no evidence to support the hypothesis that there is an innate, unlearned preference for fat. Indeed, the possibility of an innate fat preference seems unlikely because the form and function of fat are not unitary across food systems. Furthermore, it has been found that there is no relationship between taste preferences for high-fat foods and childhood (< 10 years)-onset obesity, drawing the conclusion that environmental as opposed to genetic factors may be more salient determinants of taste preferences and food choice.

Variability of Fat Preference

Gender differences While it is mere anecdote that 'Jack Sprat would eat no fat and his wife would eat no lean,' surveys have provided epidemiologic evidence that there are, indeed, sex differences in regard to fat preferences. Although both men and women seem to find high-fat foods highly palatable, men often derive the bulk of their dietary fat from meat, especially red meat, while women derive the bulk of their fat intake from margarine, whole milk, shortening, and mayonnaise, and are more likely than men to express preferences for sweet/fat desserts like cake and icecream. Such sex differences exist among obese individuals as well. The favorite foods of obese men and women were surveyed and it was found that obese men cited predominantly fat-protein foods (meat dishes) among their favorites, whereas obese women listed more carbohydrate/fat sources and more sweet-fat foods (doughnuts, cake, cookies, and chocolate).

Obese/lean differences A possible mechanism for obesity is that obese persons may have an enhanced preference for high-fat foods, leading to overconsumption of energy-dense foods. In sensory tests, obese individuals do, indeed, show a preference for higher levels of fat in foods than lean individuals. Several investigators have found that body weight is related to preferences for fat. It was found that obese and formerly obese individuals preferred higher levels of fat in mixtures of dairy products and sugar than did lean individuals. However, in a later study, Drewnowski found that only the subset of obese individuals with a history of large weight fluctuations showed an enhanced fat preference. Even among normal-weight individuals, there is a positive relationship between sensory preferences for fat in a variety of foods and percent body fat. As body fat increases, the percent of energy derived from fat increases. In one 3-year longitudinal study, high weight gain was associated with high fat intake in both men and women. In all, this work suggests that enhanced preferences for fat could be important in the development and maintenance of obesity; however, no controlled, laboratory-based experiment has studied this issue directly. Further research is needed to understand how individual differences in preference for dietary fat influence human food intake and body composition.

Fat Replacers and Fat Preference

There is a theoretical concern about the use of fat replacers, especially products like sucrose polyester (SPE), that mimic the properties of fat so closely, that they may reinforce and maintain the preference for fatty foods. It can be argued that the most

effective strategy for fat reduction is to decrease the preference for high levels of dietary fat. The effects of two different reduced-fat diets have been studied: one allow the use of fat replacers and one which did not, and the effects on the preference for a limited range of high-fat foods were noted. This study found that the group which was given low-fat foods without fat replacers (no fatty sensory qualities) showed a decrease in the preferred level of fat in foods, whereas the group using fat replacers showed no such shift. It was concluded that the preference for fat in foods is governed more by exposure to fatty flavors than by the actual level of fat in foods. It was further suggested that the preference for fat can be lowered, and that the best strategy for lowering fat preference is, therefore, to avoid foods with fat replacers. Because of methodological limitations, however, firm conclusions cannot be drawn from this research. First, hedonic measures were obtained on only four foods. Since fat imparts so many different sensory properties to foods, it appears unlikely that changes in preferences for fat in one type of food will generalize to other types of foods. Second, because subjects consumed products at home, there was little control over the experimental setting. Finally, data were obtained from self-reported diet records, with scant checks for compliance.

A recent study measured participants' sensory ratings to foods with various levels of fat while they were consuming diets that varied in overall fat content: 37, 30, and 26% energy from fat. Participants rated their sensitivity to and liking for fat in several foods at baseline, after each 8-week diet period, and at the end of the entire study. They found no differences in liking of foods as a function of fat level of the overall diet. Thus, exposure to a low-fat diet for an extended period did not diminish fat preference, and exposure to a higher-fat diet did not increase fat preference. In this study, however, the low-fat diets did not use fat-replaced foods exclusively. Thus, although the hypothesis that exposure to fat-replaced foods may maintain a fat preference may have merit, further research is needed to investigate this issue fully.

Energy density Metabolism of dietary fat yields approximately 9 kcal g, more than twice the yield of 4 kcal g for carbohydrate or protein. There is evidence that people tend to eat a consistent weight of food. Given this, the relatively high energy density of fat would be an important factor in its overconsumption. For example, 50 g of potato chips (which typically derive 60% of their energy from fat) provides 269 kcal of energy, while an equal weight of pretzels (which typically derive about 8% of their energy from fat or less) provides only 187 kcal. This effect was tested experimentally in a serving of a milk-based preload that had varying amounts of water added to change the volume (300, 450, and 600 ml), while macronutrient and energy content and palatability were held constant. The resulting drinks differed solely in the density of the provided nutrients and energy. The results showed that the energy density of these milk-drink preloads affected energy intake at lunch 30 min later, such that intake was significantly lower (by 18%) after the high-volume, low-energy-density drink than after the low-volume, high-energy-density drink. Further, this energy deficit was not compensated for at the dinner meal. Thus, consuming more energy-dense, higher-fat foods may easily lead to higher daily energy intakes than eating foods that are low in fat and less energy-dense.

Fat and satiety It has been proposed that fat may be overeaten because it does not satisfy hunger as well as other macronutrients. Ingested fat may differ from ingested carbohydrate and protein in how it affects such factors as stomach distention, stomach emptying, nutrient absorption, hormonal release, and/or oxidation of nutrients. Indeed, fat and carbohydrate have very different postingestive dispositions. In experiments that measured postabsorptive metabolism, dietary fat was shown to be metabolized much more slowly than carbohydrate or protein. Carbohydrate ingestion produces rapid rises in blood glucose, while fat ingestion often depresses blood glucose. The 'glucostatic theory,' which suggests that the sensation of hunger is maintained until blood glucose levels reach adequate levels, would posit that carbohydrates produce more rapid satiety than fats. Opponents of the glucostatic theory argue that there are other, more important factors influencing satiety, including the release of 'satiety hormones.'

One such putative satiety hormone is cholecystokinin (CCK). CCK is particularly released when amino acids and/or some types of fat enter the small intestine. Thus, the ingestion of fat could lead to early satiety despite having less of an effect on blood glucose than carbohydrates. Because of these differing physiological effects, it is difficult to make any definitive statements regarding the satiety value of fat versus carbohydrate. However, data from controlled-laboratory feeding studies suggest that, for some people, carbohydrate may be more satiating. One group reviewed various studies that utilized a preloading model (giving fixed amounts of macronutrients and assessing the effects on subsequent satiety/hunger and food intake) to detect any differences in the satiety value of fat versus carbohydrate. They concluded that carbohydrate seems to have greater

satiety value than fat in individuals with certain subject characteristics (e.g., obese and those concerned with their body weight). Overall, however, there is little evidence that carbohydrate is more satiating than fat, and additional research is needed to determine if overconsumption of fat is related to a physiological insensitivity to the amount of fat in food or other factors.

Types of Fat Replacers

The National Heart, Lung, and Blood Institute of the National Institutes of Health recommends that the food industry accelerate its efforts to develop and market foods reduced in total fat, saturated fat, and cholesterol content. The food industry response has been vigorous. Recent advances in food science and technology have led to the development of reduced-energy and energy-free, fat-like substances. The use of these substances in foodstuffs provides consumers with new options for highly palatable, lower-energy, low-fat products. The American Dietetic Association (ADA) favors the use of these products for consumers seeking to reduce dietary fat intake and perhaps body weight. The ADA's 1991 position paper on fat replacements states: 'the ADA recognizes the innovative development and use of traditional food ingredients and processing methods to reduce or replace fat in foods'.

Fat replacers comprise a diverse range of chemical and processing sources; they possess varied sensory, functional, and physiologic properties. Some of the currently used and newly developed fat replacers are described below. Table 1 provides a summary of the replacers and their characteristics.

Carbohydrate- and Protein-Based Fat Replacers

Several carbohydrate- and protein-based materials are presently in use. These substances generally add bulk, viscosity, or structure to a food by stabilizing water into a gel-like structure. In contrast, microparticulated protein mainly provides a textural continuity (often associated with fat). Because it consists of particles so small that the units are undetectable by oral sensory systems, microparticulated protein gives the illusion of a creamy, smooth texture. Protein-based fat replacers can also provide bulk and viscosity characteristics similar to carbohydrate fat replacers and can supply added functionality to food systems, including opacity and emulsification. These products are partially or fully digestible and provide between 1 and 4 kcal g.

Unfortunately, protein- and carbohydrate-based materials are very limited in their functional usage in general fat replacement. Both types of fat replacers can only be used in a few categories of food (e.g., dairy products, dressings, spreads, cold applications, and some meat products); further, some are not well-suited for baking application, cannot be used to fry foods, and have a limited shelf-life due to high water content. Most of the carbohydrate- and protein-based fat replacers cannot be used for the development of lipid-associated flavors or flavor modification in foods; microparticulated protein-based fat replacers, in contrast, do have a clean flavor base and good flavor-release qualities. Studies indicate that protein- and carbohydrate-based replacers pose little or no toxicological risk when used in the moderate quantities proposed.

Lipid-Based Fat Replacers

One lipid-based fat replacer is currently being marketed (olestra or Olean) and others are under development. Lipid-based fat substitutes have the advantage of having similar functional and sensory properties to the fats they replace. Thus, the potential applications, in contrast to carbohydrate- or protein-based substances, include most or all areas where fat and oils are presently used.

SPE or olestra, its generic name, is a fat-like material compound of hexa-, hepta-, and octa-esters of sucrose and long-chain fatty acids. Its physical properties are nearly identical to those of conventional dietary fats. Because of the molecular bonds in SPE, it cannot be hydrolyzed by pancreatic lipases, and the large molecule, consequently, cannot be absorbed through the intestinal mucosa. Because SPE is not digested or absorbed, it passes through the gastrointestinal tract intact and is excreted in the feces, adding no energy to the diet.

Caprenin, another lipid-based fat substitute, consisting of a glycerol molecule esterified with caprilyc, capric, and behenic acids, was developed for use in chocolate. Unlike SPE, the glycerol-based molecule is hydrolyzable by pancreatic enzymes, but provides only 5 kcal g (instead of the 9 kcal g of triglyceride) due to the poor absorption of behenic acid. Other low-calorie fat replacers include Salatrim (5 kcal g) and Aldo MCT, Captex, and Neobee (8.3 kcal g). Other lipid-based fat replacers include partially and poorly digested organic compounds. A list of such fat replacers can be found in Table 1.

Other Fat Replacers

Polyglycerol esters form an intermediate category between lipid-based and nonlipid-based fat replacers. The functional metabolic properties of polyglycerol esters range between carbohydrate- and lipid-based fat replacers, depending on their level of esterification. Polyglycerols may be partially or wholly

Table 1 Types of fat replacers

	Composition	*Chemical name/process (if any)*	*Trade name/company*	*Applications*
Carbohydrate-based				
Starch-based	Modified starch		Amalean Leanbind N-Lite PureGel Sta-Slim	Cheesecake, pourable salad dressings, sauces, cream fillings, dips, spreads, bakery, dairy products, meat and poultry products, icecream, yogurt
	Maltodextrin		C*Pure CrystaLean Lycadex Maltrin Novelose	Extruded products, bakery products, dressings, sauces, spreads, frozen desserts, toppings, cakes, buttercream and fillings, snacks
	Hydrolyzed flours	Enzymatically and acid-hydrolyzed flour	Quaker Oatrim Rice-gel Rice-trin	Processed meats, bakery, prepared foods, baby foods, sauces, cheese spreads, cream cheese
Fiber-based	Celluloses and hemicelluloses		Avicel Ac-815 Cellulon Ex-cel Fibrex Just Fibre	Mayonnaise, salad dressings, butter substitutes, cheeses, sour cream substitutes, spreads, bakery products
	Fructose polymers	Inulin	Raftiline Fibruline	Icecream, cheese spreads, chocolate, dressings, meat
	Gums	Gellan gum Xanthan gum Acacia gum Guar gum	Kelcogel range Ketrol range Gelamix Rhodigel Spraygum Procol range Supercol F	Margarines, spreads, cakes, cookies, dips, frozen desserts, salad dressings, sauces, gravies, instant products, soups, toppings, canned foods, frozen foods
	Gels	Carrageenan	Aquagel BenGel Liangels	Bakery products, beverages, sauces, confectinery, dairy products, salad dressings, meat products

Protein-based				
	Whey protein		Dairy-Lo Dairylight Nutrilac Simplesse Trailblazer	Frozen desserts, dairy products, baked goods, salad dressings, spreads, sauces, toppings, frosting
	Soy protein		Soy protein Concentrate Supro range	Meat and dairy products, patties, hamburgers, frankfurters, cream cheese
	Milk protein		Complete Milk Protein Lactalbumin Lactomil Miprodan	Yogurt and cultured products, dressings, mayonnaise, sauces, dips
	Microparticulated zein protein (from corn gluten)		Lita	Frozen desserts, mayonnaise, sour cream, salad dressing, whipped topping, milk, yogurt, confectionery and chocolate
Lipid-based				
	Low-calorie fats	Medium-and long-chain triglyceride molecules	Aldo MCT Caprenin Captex Neobee M-5 MCT Oil Salatrim	Chocolate, chocolate coatings, confectionery products, replacement for liquid vegetable oils in products
	Synthetic fats	Sucrose polyester	Olestra	Cooking and frying utilities.
			Colestra	Currently only approved for use in savory snack foods
		Other esters	Alkyl glycoside fatty acid polyesters Carboxy/carboxylate esters Dialkyl dihexadedecyma lonate (fatty acid ester) Polyhydroxyl esters	Baked goods, chocolate flavored products, frozen deserts, margarine, meats, salad dressings, snack foods

digestible, but are generally used in low concentrations relative to the quantity of fat they replace as part of an aqueous emulsion.

Low-Fat Diet Research

Fat replacers are a relatively recent development in the field of studying dietary fat intake. Thus, much of the research regarding dietary fat intake predates the development of many fat replacers. This section reviews research involving manipulations of dietary fat intake with and without fat replacers. Some studies have reduced the fat content of the diet by using foods naturally or intrinsically low in fat (fruits, vegetables, and grains). Others have used specific fat-replaced foods available at the time of study or a combination of naturally low-fat foods and fat-replaced foods. The following studies will serve to illustrate the effects of reducing the fat content of the diet on both energy and fat intake.

Short-Term Dietary Fat Manipulation

A number of studies have investigated the effects of manipulating the fat content in certain meals on energy intake on a short-term basis (≤ 5 days). One such study manipulated the energy and fat content of a midday meal for 5 days by using intrinsically low-fat foods, as well as some commercially available reduced-fat foods. They found that both men and women compensated for energy dilutions in the diet. Interestingly, compensation for excess of energy intake was weaker, especially when the additional energy was derived from dietary fat. Unfortunately, the data for fat intake were obtained via diet records, and such data must be interpreted cautiously due to the potential errors and biases inherent in self-reported measures. However, similar results were obtained in a well-controlled residential laboratory study that manipulated the carbohydrate and fat content of certain meals. Participants (lean males) compensated well and quickly for the caloric dilution in this study, but when the energy level was again raised to baseline levels, participants did not compensate and consequently overate. In a subsequent residential study that studied the affect of altering fat and carbohydrate content of a lunch meal using four conditions (high-fat, high-carbohydrate, low-fat, low-carbohydrate; with 3 days per condition), energy compensation was observed regardless of macronutrient composition (mean daily energy intakes: 2824, 2988, 2700, and 2890 kcal respectively). However, energy intake from dietary fat was fairly constant in all conditions except the high-fat condition, which resulted in significantly greater energy intake than the other three conditions. These studies suggest that decreasing the fat (and thus the energy) content of the diet will result in compensation for energy in subsequent meals or snacks, but not for fat. Furthermore, there is some evidence that when additional fat is included in the diet, it is unlikely that a spontaneous reduction in energy and fat intake will occur to compensate for this surfeit. It is important to note that these findings are dependent on both the magnitude of the manipulation (amount of fat/energy reduction) and the characteristics of the study participants (e.g., lean/obese, restrained (concerned about body weight)/unrestrained (not concerned about body weight)).

Longer-Term Manipulations

The studies discussed above manipulated fat intake within specific meals or over the course of a day. This type of intervention provides useful information about short-term regulation of fat and energy intake when dilutions and surfeits are encountered in specific meals or snacks, but they provide no information about long-term regulation of fat and energy intake or the ability to maintain compliance with a low-fat diet. In contrast, a few naturalistic studies have examined the effects of reducing intake of dietary fat in certain groups over a longer timeframe. These studies utilized various strategies, including nutritional counseling, behavioral therapy, and financial incentives, to aid subjects in reducing their dietary fat consumption. These studies generally found that the intervention groups did consume less dietary fat than control groups. Intervention periods in these studies ranged from 12 to 20 weeks. These studies suggest that compliance with a low-fat diet can be maintained over at least this moderate amount of time. There are few data, however, regarding longer-term compliance with a low-fat diet. One group of researchers followed 303 women participating in the Women's Health Trial for 1 and 2 years. These women either received no specific intervention or intensive instruction in maintaining a low-fat diet by such means as reducing cooking fats, substituting low- or no-fat variants of high-fat foods and increasing traditional low-fat foods in the diet. After 1 year, the intervention group had reduced their dietary fat intake by 45% of their baseline intake and energy intake by 59%. Most of the reduction occurred within the first 6 months of the intervention and was maintained at 1- and 2-year follow-up. Results of naturalistic studies such as these are often used as a basis for advocating low-fat diets and that compliance to low-fat diets can be maintained over time. However, although these studies may have validity, they relied heavily on self-report (diet records or food-frequency questionnaires) as a source of data on fat intake, and should be interpreted cautiously.

Laboratory-based experiments provide more accurate intake measurements and stringent controls over the experimental setting. Three controlled laboratory studies have investigated the effects of high- and low-fat diets on energy intake and body weight over varying periods of time. The first study allowed lean and obese participants to consume a low-energy-density diet comprised of intrinsically low-fat foods (fruits, vegetables, and whole grains) or a high-energy density diet ad libitum for 5 days (0.7 vs 1.5 kcal g). Both obese and nonobese groups significantly reduced their energy consumption on the low-energy diet. Nearly twice as many calories were consumed during the high-energy-density diet compared to the low-energy-density diet (3000 vs 1570 kcal day). Weight change data were not reported.

Two other studies, both conducted at Cornell, are often cited in both scientific and popular literature as proof that ad libitum consumption of low-fat foods can reduce fat intake and produce weight loss. The low-fat diets in these two studies included fat-replaced foods (a limited selection of margarines, salad dressings, and mayonnaises available at the time the studies were conducted) as well as intrinsically low-fat foods. In the first of these studies, all food (meals and snacks) was provided for 24 women divided into two groups: < 101% ideal body weight (IBW) and > 101% IBW. Each participant was fed three diets (in a counterbalanced order): low-fat 15–20%, medium-fat 30–35%, and high-fat 45–50% of energy. Participants could eat the foods ad libitum. Each diet was fed for 14 days with a 7-day washout between test periods. Energy consumption was found to be positively correlated with the level of dietary fat: total daily energy consumed on the low-fat diet averaged 2087, the medium-fat diet 2352, and the high-fat diet 2614 kcal. Over the 2-week intervention periods, the diets did not produce any statistically significant weight changes.

The second Cornell study was similar, but extended the intervention period to 11 weeks. It is unclear whether normal-weight or obese women were studied: although a mean subject height is reported, no mean weight is reported, though it is noted that individuals < 101% of IBW (according to Metropolitan Life standards) were excluded. This study examined two diets: one was comprised of 20–25% energy from fat (low-fat) and the other, 35–40% energy from fat (control). Again, the women were allowed free access to foods. It was found that participants consumed an average of 286 fewer kcal day on the low-fat diet. In contrast to other Cornell study, these participants did lose weight on both diets (low-fat: − 2.54 kg and control: −1.26 kg). There were no reasons given by the authors for the weight loss on the control diet.

A longer-term animal study examined the impact of replacing fat using a carbohydrate-derived fat replacer on three groups of 20 female Sprague-Dawley rats. The rats were chronically fed either a high-fat diet (45% energy from fat) or a low-fat diet (25% energy from fat) containing either no added fat replacer (control diet) or the maltodextrin, Paselli. The three diets differed in texture. Rats were fed the control diet for 10 days, and were then divided into three groups, receiving one of the three diets for 42 days. Preference tests for the three diets were assessed before and after the 42-day period. It was found that all but two of the rats preferred the diet with the fat replacer over the high-fat and control diets, and that food intake was the greatest in the group fed the fat-replacer diet (741 g per 58 days = control, 736 g per 58 days = high-fat, 900 g per 58 days = fat replacer). However, energy intake was greatest in the group fed the high-fat diet (3250 kcal per 58 days = control, 3871 kcal per 58 days = high-fat, 3585 kcal per 58 days = fat mimetic). The higher energy intakes of the high-fat and fat replacer-fed animals resulted in greater weight gain during the experimental period compared with control rats, largely due to an increase in fat mass. These data suggest that the inclusion of highly preferred, fat-replaced foods in the diet may result in a reduction in fat intake, but consumption of such foods is not associated with reductions in energy intake.

Overall, the evidence from the human studies presented here indicates that low-fat diets are potentially useful for reducing both total energy intake and the amount of fat consumed. While these early results have some validity, their interpretation is often oversimplified. It is important to recognize that the methods available to assess the effects of low-fat diets have limitations that may influence outcome. Participants in the three laboratory-based studies reviewed above allowed subjects to consume food ad libitum during the dietary interventions; however, this access was limited to the particular fat-level that was under study. Thus, during a low-fat condition, the subjects could only choose from low-fat foods, while in the high-fat condition, only high-fat choices were available. Although the subjects rated the high- and low-fat diets equally palatable, this does not mean that, given a choice, they would voluntarily eat the low-fat foods, making it difficult to determine if the findings are applicable to free-living individuals. As demonstrated in the animal study described above, even when restricted to a highly palatable low-fat diet, overall energy intake may not be reduced.

Labeling Effects of 'Low fat'

A further limitation of these initial studies is that they did not provide participants with any information (nutritional or otherwise) about the foods that they were eating. Food products labeled with claims such as 'lite,' 'reduced-fat,' and 'fat-free' send powerful messages to the consumer. The results of the low-fat diet studies cited above might be different for individuals who are aware of whether their diet is low or high in fat or who knowingly choose to substitute a low-fat food for a high-fat food outside an experiment. A few studies have examined the impact of such information on subsequent food intake. It was found that intake at a self-selection meal was higher following a yogurt preload labeled 'low-fat' than following equicaloric yogurt preloads labeled 'high-fat.' Another study fed 96 men and women a snack of potato chips made with olestra or triglyceride every afternoon for 2 weeks in a counterbalanced order. Half of the participants were given full nutrition labels about the potato chips and the other half were given bags that simply said 'potato chips.' Interestingly, in the group that received no nutrition information about the chips, all participants ate the olestra and the triglyceride potato chips in similar quantities. In the group that received information, a subset of the participants, those who are very concerned about their body weight and tried to control their food intake (as measured by the Three Factor Eating Questionnaire), consumed significantly more olestra potato chips (60 ± 7 g) than regular potato chips (50 ± 17 g). Other subgroups (obese/normal weight; male/female) did not eat the chips in different quantities. Results from these studies indicate that information about fat (and energy) content can influence intake of the low-fat food itself or of food eaten subsequently. It is apparent that cognitive processes can override or interact with physiological processes in regard to food intake and, therefore, play an important role in energy balance. It may be those who are the most restrictive in their dietary choices who may be at most risk for overconsuming fat-free and reduced-fat foods.

Cognitive processes may also be involved in the phenomenon of food intake being controlled by weight or volume rather than by its energy content. In other words, individuals may control how much they eat by setting a 'standard portion size' of food consumed rather than a 'standard energy amount.' As discussed previously, it is likely that it is the low-energy density of the diets provided in these low-fat diet studies and not a physiological satiety mechanism that accounts, at least in part, for the differences in the amount of energy consumed. In each of these studies, subjects ate similar gram weights of food on both low-fat and high-fat diets. It is likely that subjects are eating a 'standard' weight or volume rather than specifically responding to the covert energy manipulation. Thus, a diet of low energy density consumed in equal amounts to a higher-density diet would provide less energy. Fat-replaced foods may have their greatest impact on fat and energy intake by simply reducing the energy density of the foods we consume.

There is evidence that energy intake may be more a function of the energy density of the diet than the actual fat content. Three studies have found that energy intake is constant despite variations in fat content when the energy density of the diet is held constant. Given this robust effect, a group of researchers investigated the effects of varying the energy density of the diet while holding fat content constant. This manipulation was achieved primarily by altering the water content of foods. The researchers then fed normal-weight women all of their meals in a laboratory setting over three 2-day periods using three levels of energy density in provided entrées (low $= 1.03\,\mathrm{kcal\,g^{-1}}$; medium $= 1.17\,\mathrm{kcal\,g^{-1}}$; and high $= 1.34\,\mathrm{kcal\,g^{-1}}$). The women, like participants in the earlier studies, ate a constant weight of food across the dietary conditions, resulting in a 30% decrease in energy intake in the low-energy-density condition compared to the high-energy density condition; however, there were no corresponding changes in ratings of hunger and fullness. These results provide clear evidence that alteration in energy density of food can affect energy intake. Similar studies need to be done to investigate the effect of using foods made with fat replacers to reduce energy density instead of water.

If fat-replaced foods are consumed in much greater quantities than their high-fat counterparts, however, this net energy reduction will be nullified. The belief that one can consume an unrestricted amount of low- or no-fat food and still expect to bring about desirable dietary change and/or weight loss or maintenance is not supported by the animal or human literature. For example, Sprague-Dawley rats were fed high-fat (41% energy from fat) or fat-free (0% energy from fat) pound cakes in addition to chow for 30 days. The fat-free pound cake was prepared with nonfat milk and egg whites, and used modified cornstarch, xanthan gum, and guar gum to simulate the orosensory properties of fat. Rats fed the high-fat cake and chow consumed more energy ($P < 0.05$) and gained more weight ($P < 0.05$) than did rats fed fat-free cake and chow. The rats fed the fat-free cake, however, overate and gained more weight than did the chow-only controls. The authors concluded that removing

the fat from the cake reduced, but did not eliminate, its obesity-promoting effects, and that low-fat diets have to be consumed in moderation if used for weight control purposes.

To examine the question of whether isolated reduction of fat with no other dietary restriction leads to weight loss in humans, the effects of a low-fat diet consumed ad libitum were compared with a low-fat diet with caloric restriction. It was found that participants on the calorically restricted, low-fat diet lost significantly more weight and reduced fat intake to a greater extent than those on the low-fat diet consumed ad libitum. Hence, advocating that an individual can eat as much as he/she desires and lose weight as long as the food is low in fat content is unwarranted based on these data. For maximum weight control, total energy intake must also be restricted below energy needs.

Another important question is whether varying the fat content, while keeping energy equal, of an energy-restricted diet enhances weight loss. This question was addressed by an animal study. It was found that when rats previously maintained on a high-fat diet (40% of energy) were switched to equicaloric diets (75% of their previous ad libitum intake) that varied in fat content (12, 28, or 45% fat), rats fed the 12% fat diet lost significantly more body fat than the rats fed the 45% fat diet. These results indicate that, at least in rats, an energy-restricted diet produces greater weight loss when it is also low in fat. Although similar data are not available for humans, if weight loss is the goal, it is reasonable to recommend restriction of both dietary fat and overall energy intake to achieve the maximum benefit. Thus, reduced-fat foods will enhance weight loss if an individual who is consuming greater than 30% of energy from fat restricts food choices to those low in fat but not high in energy, while keeping the volume of food consumed constant.

Fat-Replaced Foods and Disease Risk

Although the usefulness of fat-replaced foods for energy reduction and weight loss requires further investigation, it is reasonable that reductions in dietary fat may be helpful in reducing the risk of diseases associated with high fat intake, such as cardiovascular disease and type 2 diabetes. In a recent study, 30 men and women (15 of whom had diabetes) were supplied with five low-fat or fat-free products or their regular fat counterparts and asked to consume them as part of their regular diet in a free-living, free-choice setting for 3 days. No difference was found in usage of the fat-modified foods between people with or without diabetes. Participants consumed the fat-replaced foods in similar quantities to the full-fat counterparts, which resulted in a significant reduction in overall fat (g) intake, but total energy intake was not different between the fat-modified and the full-fat conditions. Cholesterol and saturated fat intake were also significantly lower in the low-fat condition than in the full-fat condition. It has been shown that foods made with fat replacers were less acceptable than foods made with sugar replacers among participants diagnosed with diabetes. Foods made with olestra were less acceptable still than foods made with fat replacers in general. In a recent survey, it was found that knowledge about fat replacers, fat-replaced foods, and general nutrition was poor among individuals with type 2 diabetes. Respondents were much more familiar with sugar replacers than fat replacers. Thus, despite evidence that inclusion of fat-replaced foods may be helpful for people with diagnoses such as type 2 diabetes, knowledge about and acceptance of such foods are low.

Noncaloric, Lipid-Based Fat Substitute Research

Because noncaloric, lipid-based fat substitutes have the greatest potential range of applications as fat replacers, it is important to review the current research on the effects of consuming these materials on food intake and body composition. At this time, only olestra (Olean) is approved for use by the Food and Drug Administration in the USA. Olestra's approval is limited to applications in savory snacks.

Animal studies comparing varying levels of fat and SPE have demonstrated that the higher the percentage of energy from fat in the diet, the greater the fat intake and, in some cases, the body weight. For example, in one study mature female Sprague-Dawley rats were fed five diets that varied in fat content: control diet (21%), low-fat diet (2%), high-fat diet (63%), and two medium-fat diets (30% and 51%) fat. The rats compensated for the low-fat diet compared to the control diet by increasing food intake, resulting in no difference in overall energy intake. However, when the control diet was replaced with the high-fat diet, the rats became obese and hyperinsulinemic. When rats fed the high-fat diet were switched to the medium-fat diets, they also increased food intake to compensate for the energy reduction. These rats were more obese than control rats, but less obese than the rats fed the high-fat diet, suggesting that a modest reduction in dietary fat content can attenuate fat's obesity-promoting effects. In another animal study, the effect of SPE replacement was examined on the body weights of lean and obese Zucker rats by

feeding 8-week-old lean and obese animals either a control diet (15% corn oil) or an SPE diet (5% corn oil and 10% SPE). The obese, control-fed animals gained more weight than the animals fed the SPE diet, while the lean rats given the fat substitute did not have significantly different body weights compared to the lean controls. It was concluded that the obese rats could not defend their higher weights when the fat content of their diet was diluted using SPE.

The first controlled study using sucrose polyester in humans was conducted in 1982. The effect of using SPE in hypocaloric diets for obese persons seeking weight loss was investigated. The effects of SPE and a hypocaloric diet were studied in a counterbalanced, crossover design over two 20-day periods. Participants also had access to non-SPE-modified snacks in the evening, which provided an opportunity for subjects to compensate for the calorie/fat dilution. Participants showed weak compensation for the missing calories, resulting in a significant overall reduction in energy intake ($P < 0.05$) and fat intake ($P < 0.05$). However, a potential confounding factor exists because subjects were actively trying to lose weight in this study, which may have reduced their likelihood of consuming snacks. Thus, this study does not provide an adequate design to assess energy/fat compensation.

In another study the fat content of the diet of five obese men confined to a metabolic ward for 36 days was diluted by replacing the fat in margarine and mayonnaise with SPE. Energy intake was measured during three periods: preintervention baseline (days 1–9), intervention (days 10–28), and postintervention baseline (days 29–36). During the intervention period, the manipulation created a 10% reduction of overall energy of the diet. Participants increased the amount of food consumed during the SPE period relative to the pre- and postintervention baselines so that there was no significant effect of diet on energy intake (baseline = 3924 kcal day^{-1} and SPE=3812 kcal day^{-1}). Participants did not selectively ingest more fat, but rather increased their intake of all three macronutrients to compensate for the energy dilution created subsequent to the addition of the SPE.

Parallel studies in the USA and the UK were the first to study the effect of olestra in lean individuals. Twenty-four healthy, lean males participated in each trial that investigated a SPE manipulation in a breakfast meal (biscuits/scones and margarine). Subjects were fed the three breakfasts (control = 765 kcal, 20 g SPE = 582 kcal, and 36 g SPE = 445 kcal) in a counterbalanced, crossover design followed by self-selected lunch and dinner meals. Subjects also recorded any evening snacks to account for 24-hr energy intake. Despite geographical differences, similar results were obtained in both studies. The lean men compensated well when the breakfast was reduced in energy and fat due to SPE replacement, but there was very little fat-specific compensation. The end result was a significant ($P < 0.0034$ and $P < 0.0066$) reduction in fat consumption in the 20 g and 36 g SPE substitution, respectively, compared to placebo.

Children also appear to compensate well for energy deficits, but not for fat specifically. In one study (Birch *et al*., 1993), the intakes of 29 normally developing 2–5-year olds were measured during 2 days per week over 5 weeks. Approximately 14 g of SPE was substituted for fat (− 125 kcal) in the children's diet over the course of the morning and afternoon. By the end of the first day, the children had compensated for some of the energy dilution and by the end of the second day they had compensated for all but 24 of 125 kcal. Like the adult men, children did not show macronutrient-specific compensation. Instead, the manipulation produced a reduction in the percentage of fat consumed without a reduction in total energy of the diet because the children increased their intake from carbohydrate sources.

In other studies, neither fat nor energy compensation was detected when SPE was incorporated into foods. In one study 47 g of SPE was incorporated into meals and snacks across a day (creating a diet which derived 20% of energy from fat). This manipulation represents a more severe reduction in the amount of fat replaced by SPE than in previous studies. The results were that this level of fat/energy reduction was not compensated for during the day that SPE-replaced foods were provided, and only partial (74%) energy compensation occurred the following day. Incomplete fat/energy compensation was also seen in two studies from the Netherlands. The first incorporated SPE into croissants and found incomplete short-term energy compensation so that daily energy intake was reduced. This group also found that energy was not fully compensated for in a study which used SPE to replace fat in warm meals over a longer term (12 days). However, these studies used control meals that were actually higher in fat content and energy content than the typical Dutch meal, as either triglyceride oil or SPE oil was simply added to dishes. This resulted in a 'surfeit' of fat and energy in the control condition, while the SPE condition matched more closely the fat and energy content of the typical Dutch meal. Given this and other methodological issues, firm conclusions on energy compensation cannot be based on these data.

Overall, therefore, the efficacy of using foods prepared with SPE to reduce fat and/or energy intake is

equivocal. The existing evidence from the relatively few studies that have been published suggests that SPE incorporation into foods may aid in reducing the amount of fat consumed. The evidence is, however, less clear regarding SPE's efficacy in reducing total energy intake. Again, it is important to note that all of these studies used covert manipulations: participants were unaware of the fat manipulation. This may or may not be representative of how such products will be consumed in a free-choice, full-knowledge setting. More studies are, therefore, needed to clarify the effects of fat replacers on energy intake.

There has also been considerable focus on possible negative health effects associated with SPE, especially olestra, most notably gastrointestinal (GI) complaints and decreased absorption of fat-soluble vitamins. Since olestra is nonabsorbable and passes through the gut unchanged, it has been anecdotally linked with GI complaints such as diarrhea, bloating, and cramps. However, one study found no difference in prevalence of GI complaints following a single episode of snacking on chips made with triglyceride or chips made with olestra. Unpublished data from the Olestra Post-Marketing Surveillance Study (OPMSS) found only a limited association with one symptom (bloating) among the highest consumers of olestra. In addition, OPMSS data indicate that GI complaints are rarely attributed to fat replacers such as olestra, while most food-related GI problems are attributed to beans, and spicy and diary foods. Further, another study, also using OPMSS data, found that olestra consumption was not associated with changes in serum carotenoids or serum concentrations of other fat-soluble vitamins; however, serum vitamin K was higher in the highest consumers of olestra. This may be due to olestra being fortified with vitamins A, D, E, and K. Thus, according to current measures, olestra does not present a significant risk of GI upset or reduced absorption of fat-soluble vitamins.

Conclusions

In this article, we have examined the available data regarding whether or not reduced-fat foods, especially those made using fat-replacer technology, are useful in reducing the current trend to overconsume dietary fat and energy in western societies. This question is difficult to answer because it is only since the 1980s that dietary fat consumption has been a focus of nutritional research, and many of the advances in fat-replacer technology are of even more recent vintage. What is clear is that foods that taste good are consumed more readily than those that do not. Thus, it is axiomatic that the availability of low- or no-fat foods that are also highly palatable may aid in compliance with low-fat diets that were previously bland and unsatisfying. However, while fat-replaced foods offer consumers new food choices, considering the conflicting data, it should not be assumed that the use of fat-replaced foods will, indeed, result in significant reductions in fat and energy intake.

The research cited in this article supports the notion that fat-replaced foods may aid in reducing dietary fat intake, but perhaps not overall energy intake. Most studies using traditional low-fat foods and currently available fat-replaced foods have resulted in at least partial compensation for energy reductions, but have not resulted in macronutrient-specific (or fat-specific) compensation. Results from SPE studies are equivocal in respect to energy compensation, with some finding energy compensation, while others do not. Better controlled, laboratory-based human studies are needed to estimate just how useful SPE replacement will prove to be in reducing overall fat and energy intake.

Also still unclear is how consumers will use new and existing fat-replaced foods. Will they be used as a one-to-one substitution for foods previously high in dietary fat, or as a license to overeat other rich foods? It may be that the most important predictors of the successful use of fat-replaced products will be the motivation of the consumer to bring about a reduction in his/her intake of dietary fat. More naturalistic studies exploring the potential usage patterns of fat-replaced products are needed to determine their usefulness in bringing about these desired dietary changes, and additional nutrition education regarding fat-replaced foods is necessary among those with diet-modifiable diseases, such as type 2 diabetes, cardiovascular disease, and hypertension.

Because overall energy intake has been shown to be a critical factor in weight loss and weight maintenance, the use of fat-replaced foods alone should not be expected to produce spontaneous improvements in weight management or obesity. Lasting changes in body weight will still be dependent upon long-term behavioral change that includes not only modification of fat intake, but also reductions of overall energy intake, along with an increase in energy expenditure. Because fat is the most energy-dense macronutrient, substituting low-fat foods can substantially reduce the energy density of the diet, provided these foods are also low in energy. If the energy density of the diet is reduced and the volume of intake remains constant, reductions in total energy intake are likely.

These caveats aside, fat-replaced foods, like other modified foods (e.g., with aspartame or saccharin) could aid motivated individuals to reduce their intake of dietary fat and energy. Thus, fat replacers may

prove a useful tool in reducing fat intake, but, as with most novel approaches, more detailed investigations need to be conducted to examine the efficacy of such products in reducing fat intake. Future studies are needed to resolve issues of fat and energy balance using fat-replaced foods, especially those replaced by SPE, and how such foods can be used to alter the energy density of the diet. Results of such studies will have important implications for how fat-replaced foods should be used in the prevention and treatment of obesity and related conditions such as cardiovascular disease and diabetes. Longitudinal studies, both laboratory-based and naturalistic, need to be conducted to determine the best strategies for long-term compliance with a low-fat, low-energy-density diet.

See also: **Energy**: Intake and Energy Requirements; **Fats**: Fat Replacers; **Obesity**: Etiology and Diagnosis; Fat Distribution; Treatment

Further Reading

Bell EA, Catellanos VH, Pelkman CL, Thorwart ML and Rolls BJ (1998) Energy density of foods affected energy intake in normal-weight women. *Am J Clin Nutr* 67: 412–420.

Birch LL, Johnson SJ, Jones MB and Peters JC (1993) Effects of a nonenergy fat substitute on children's energy and macronutrient intake. *Am J Clin Nutr* 58: 326–333.

Cheskin LJ, Miday R, Zorich N and Filloon T (1998) Gastrointestinal symptoms following consumption of olestra or regular potato chips: A controlled comparison. *JAMA* 279: 150–152.

Drewnowski A and Greenwood MRC (1983) Cream and sugar: human preferences for high-fat foods. *Physiology and Behavior* 30: 629.

Duncan KH, Bacon JA and Weinsier RL (1983) The effects of high and low energy density diets on satiety, energy intake, and eating time of obese and nonobese subjects. *Am J Clin Nutr* 37: 763.

Gatenby SJ, Aaron JI, Morton G and Mela DJ (1993) Nutritional implications of reduced-fat foods in free-living consumers. *Appetite* 21: 178.

Glueck CJ, Hastings MM and Allen C (1982) Sucrose polyester and covert caloric dilution. *Am J Clin Nutr* 35: 1352.

Harris RBS (1994) Factors influencing energy intake of rats fed either a high-fat or a fat mimetic diet. *Int J Obes* 18: 632.

Harris RBS and Jones WK (1991) Physiological response of mature rats to replacement of dietary fat with a fat substitute. *J Nutr* 121: 1109.

Kendall A, Levitsky DA, Strupp BJ and Lissner L (1991) Weight loss on a low-fat diet: consequence of the imprecision of the control of food intake in humans. *Am J Clin Nutr* 53: 1124.

Lissner L, Levitsky DA, Strupp BJ, Kalkwarf HJ and Roe DA (1987) Dietary fat and the regulation of energy intake in human subjects. *Am J Clin Nut* 46: 886.

Mela DJ and Sacchetti DA (1991) Sensory preferences for fats: relationships with diet and body composition. *Am J Clin Nutr* 53: 908.

Miller DL and Cheskin LJ (1999) Acceptance and use of fat-replaced foods in persons with diabetes: Olestra Post-Marketing Surveillance Study. Abstract. Society for the Study of Ingestive Behavior, Annual Meeting.

Miller DL, Castellanos VH, Shide DJ, Peters JC and Rolls BJ (1998) Effect of fat-free potato chips with and without nutrition labels on fat and energy intakes. *Am J Clin Nutr* 68: 282–290.

Rock CL, Thornquist MD, Kristal AR, Patterson RE, Cooper DA, Neuhauser ML, Neumark-Sztainer D and Cheskin L (1999) Demographic, dietary and lifestyle factors differentially explain variability in serum carotenoids and fat soluble vitamins: baseline results from the sentinal site olestra post-marketing surveillance study. *J Nutr* 129: 855–864.

Rolls BJ and Hammer VA. Fat, carbohydrate and the regulation of energy intake. *Am J Clin Nutr*, in press.

Shide DJ and Rolls BJ. Information about the fat content of preloads influences energy intake in healthy females. *J Am Diet Assoc*, (submitted).

FATS

Contents

Production of Animal Fats

T J Wertelecki and R K Bodarski, Wroclaw Agricultural University, Wroclaw, Poland

Introduction

All metabolic processes carried out in animal cells depend on exogenous energy sources. The level of anabolic and catabolic processes in plant cells is lower compared with animal cells, especially with organisms of stable temperature. The chemical composition of food for humans and animals are exogenous sources of energy synthesis, which is in contrast to the plant world, where one of the energy sources is solar radiation – the warranty of photosynthesis.

The main energy sources are carbohydrates and fats and occasionally proteins. The highest values of energy content derive from lipids (energy donors); lipids are also an important form of energy store in animals (acceptor energy). Fat (animal fat) has another more important function in organisms: storing metabolic water (endogenous water); as building timber; creating the structure of muscle; and as protection for other organs – kidneys, liver, intestine, and pancreas. The main function of lipids in organisms is thermostability. Lipids are often connected with lipophil vitamins: A, D, E, and K. Lipids are an important source of essential and important fatty acids. As animal fat is vital it is suggested that is impossible to eliminate fat (animal fat) from the human diet and especially from that of animal.

Evolution of Human and Animal Production

The evolution of the species *Homo sapiens* was connected with the animal world. In the earliest times human beings were hunters, who trapped animals for food – meat (protein and fat) and skin were of the greatest value. From the beginning of hunting, humans began a new era of human nutrition (with animal products on the menu). Later on humans bred animals and so could use animal produce without hunting: animals were domesticated in Asia nearly 10 000 years ago and in Europe 5000 years ago. From this time humans started to carry out breeding programs with animals. Initially breeding animals was uncharted territory but as civilization evolved, it was based on developing sciences. The underlying aim was to increase production value of domestic animals, but at first civilization developed faster than animal production. The principles and aims of breeding focused on one thing only: more animal products – meat, milk, eggs. At that point the quality of the product was not an issue, because there was always a lack of energy and protein.

The revolution in the ideas behind animal breeding started with the development of human food sciences. A wider knowledge of human nutrition created new expectations for animal products, especially as far as quality was concerned. The top priority was animal protein as a source of amino acids (exogenous) for the human diet. The production of animal protein in meat, milk, and eggs increased over time and correlated with this an increase in production of animal fat. This fat was concentrated in tissue, which is impossible to eliminate in an organism. At the same time, opinions on the harmful effect of animal fat on human health were proposed by respected academics, and the lipid theory of coronary heart disease was expounded. This opinion changed the culture of human nutrition towards a decreased use of animal fat and increased use of plant oils. For example, in the USA consumption of animal fat was reduced by about 21%, over a 30-year period, and in developing countries over a similar period (correlated with a marked lack of protein-energy balance in the diet) the consumption of animal fat increased by about 80%! This situation created a negative balance between production and consumption of animal fat. In developing

countries now there is substantial overproduction of animal fat (with regard to the high breeding value of domestic animals) (Table 1).

The development of food sciences and chemical analysis has shown that consumption of animal fat is beneficial for human health: oily fish, modified concentrate and ratio of fatty acids in meat, and eggs. This opinion has suggested the possibility of the changed value of animal fat preferred by human food direction.

The Process of Lipid Synthesis in Animal Organisms

The main function of fat in animal organisms is as an energy-building store. The animal organism accumulates energy in glycogen stored in muscles and liver, but it is only 1% of energy used. The main sources of endogenous energy are accumulated in fatty acids which create fat tissue. The synthesis of endogenous animal fat could be conducted metabolically from carbohydrates, feed fat, and proteins. Part of this endogenous fat is synthesized in a simple way by absorbing and building in organs fat form without transforming it from feed fat. The main precursor of the fat synthesis process is coenzyme A (Co-A), which is created by the chemical conversion of carbohydrates and exogenous lipids.

Fat synthesis is an important chemical process which guarantees the stabilization of the organism. The main organ of lipid synthesis or degradation is the liver. The liver is a many-sided organ which is useful for energy and the process of metabolism. There are between 800 and 2000 mitochondria in a cell of the liver. Lipid synthesis in animal cells is an effect of positive energy balance. Intake delivers a high energy level compared with requirements. This energy is supplied by carbohydrates (starch for monogastric animals, as it can be easily digested, and for ruminants structural carbohydrates such as cellulose or easily digested starch) and secondly by exogenous lipids in the form of fat. Occasionally proteins supply energy for the chemical process in the organism.

Starch is the most important carbohydrate in animal production: it is used for energy and lipid synthesis. For example, 1 g of starch provides 4.15 kcal GE g^{-1} while 1 g of fat provides 9.4 kcal GE g^{-1} (GE, gross energy (calorimetric value)). Excess intake of energy concentrated in carbohydrates can be metabolized to fat in the cells of the liver and stored.

The basic biochemical structure of animal endogenous lipids is the chylomicron, which can be transported with blood and deposited in fatty tissue. Accumulated exogenous fats can be transported from the digestive tract to the blood and fat tissue without transformation in a similar way.

Fat synthesis is conducted in the organism in two ways: (1) liponeogenesis – biosynthesis of different specific fatty acids in liver cells; and (2) lipogenesis – synthesis of triacylglycerols. Lipogenesis influences the quality and time ratio of fat synthesis and deposition in tissue. The synthesis of fatty acids depends on different factors, such as: ACP (acyl carrier protein, a 9 kDa protein), another protein within the SH cysteine group, called the HS enzyme, NADP (nicotinamide adenine dinucleotide phosphate), FMN (flavin mononucleotide) and CoA (coenzyme A) (Figure 1). This process is calling β-reduction.

Synthesis of fatty acids in animal products (fat tissue) may be conducted in an intermediary path from exogenous fatty acids. These fatty acids are transported to reserve tissue or organs either with or without transformation and as phospholipids of membrane cells, or to yolk egg lipids. A simple schema showing transformation and deposition of fat and fatty acids in chicken is given in Figure 2.

In the animal world, for example, in mammalian species, the synthesis of fatty acid could access acids with 18 carbon atoms. In energy balance in an organism easily digested carbohydrates also have an important function. Some energy is gained from undigested carbohydrates by endogenous enzymes fermented by microorganisms from the cecum and colon intestine. In this way animals may gain short-chain fatty acids, which are precursors of energy and fatty acids, especially for monogastric animals.

Milk plays an important role in human diets. Synthesis of milk fat is a specific process. The main components of milk are triacylglycerols. The synthesis of triacylglycerols is conducted in epithelial cells of the mammary gland in two ways: first, from blood lipids (rebuilding blood triacylglycerols to milk form) and second, in *de novo* synthesis. Sixty percent of milk (especially long-chain) fatty acids can be gained from blood fat very-low-density lipoproteins (VLDL) or from chylomicrons containing ingested fatty acids directly from the intestine. The main sources of VLDL and chylomicrons are fat from feed and endogenous tissue fat. Medium- or short-chain fatty acids of milk are synthesized in the epithelial cells of the mammary gland or from blood precursors. The differences in fat synthesis between mono- and polygastric animals are shown in sources of carbon atoms for fatty acids. In monogastric animals the source is glucose while in ruminants it is actetate or beta-hydroxybutyrate (BHBA). In the next conversion fatty acids are esterified at the cytoplasmic surface of the smooth endoplasmic reticulum (SER) of the mammary epithelial cells to hydroxyl groups of

Table 1 Balance of animal fat production and animal fat utylization (1000 metric tons) and consumption of animal fat in food

	Years								
	1969			*1984*			*1999*		
Type of country	*Total production*	*Total domestic utylization (in food, feed and technical industries)*	*Difference between a–b*	*Total production*	*Total domestic utylization (in food, feed and technical industries)*	*Difference between d–e*	*Total production*	*Total domestic utylization (in food, feed and technical industries)*	*Difference between g–h*
	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*
USA	5270	4021	+1249	5843	4054	+1674	6995	5397	+1598
Developed countries	18443	18067	+376	24752	22555	+2197	21647	19395	+2252
Developing countries	3196	3870	−674	4965	7212	−2247	9749	12415	−2666
Type of country	*Total consumption in food (1000 metric tons)*	*Consumption per capita per year (kg)*	*Consumption per capita per day (g)*	*Total consumption in food (1000 metric tons)*	*Consumption per capita per year (kg)*	*Consumption per capita per day (g)*	*Total consumption in food (1000 metric tons)*	*Consumption per capita per year (kg)*	*Consumption per capita per day (g)*
USA	1798	8.6	17.5	1638	6.8	14.5	1872	6.7	14.3
Developed countries	11677	10.9	21.4	14093	11.7	21.4	11088	8.5	15.5
Developing countries	2604	1.0	2.4	4478	1.3	3.0	8400	1.8	4.2

Source of date: FAOSTAT, FAO Statistical Databases, Nutrition Data, Food Balance Sheets, Animal Fats, 1969, 1984, 1999, on-line http://apps.fao.org - files: 01267652.csv (for 1969 year); 01267649.csv (for 1984 year); 01267623.csv (for 1999 year).

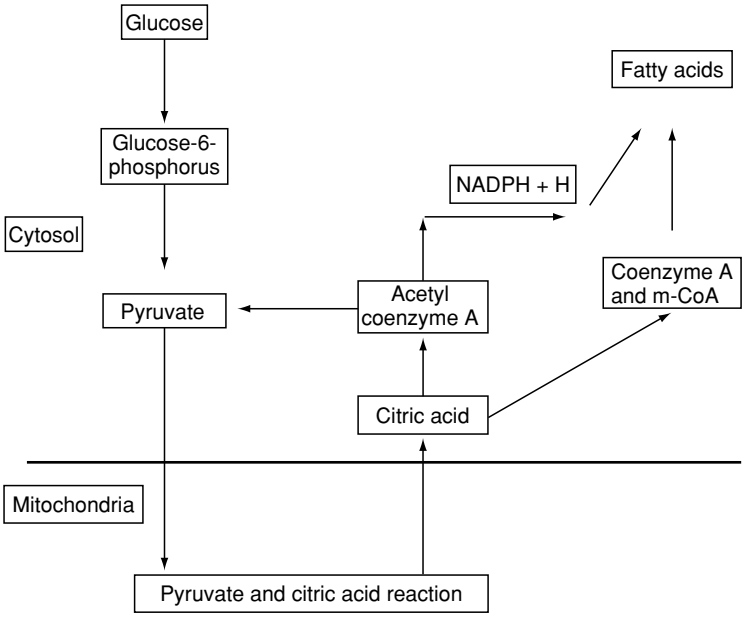

Figure 1 Simple schema of metabolic fat synthesis.

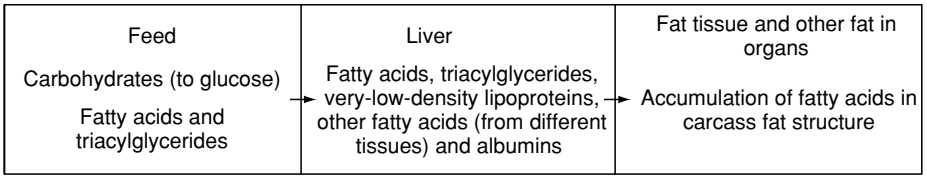

Figure 2 Route of synthesis and deposition of fatty acids and fat in chicken.

the glycerol molecules. In this way triacylglycerols are synthesized; they are agglomerated and create micro-lipid droplets. They are secreted from the SER surface into the cytoplasm, and transported directly to small milk fat globules (<0.5 mm) and unite to form larger droplets. The cytoplasmic lipid droplets may fuse with each other and be secreted as larger milk lipid droplets. The size of fat globules in milk ranges from less than 0.5 to greater than 15 mm. The micro and cytoplasmic lipid droplets are apparently surrounded by a nonbilayer coating of protein and gangliosides (polar lipids); the large lipid droplets are typical membrane-surrounded milk fat globules.

Types of Animal Product, Nutritive Value, Quality and Quantity of Animal Fat

Fat and Fatty Acid Content in Animal Product

Muscles and carcass Different species of animals have different contents of fat in the organism and in their products Table 2.

The average highest content of fat is in pork and mutton. In the organism fat accumulates under skin tissue, around some organs and intestines or between muscles. Compared with other animal products, the content of meat fat ranges from 3 to 55% (depending on animal species).

The characteristic value of the poultry carcass is because of the greatest concentration of fat under skin

Table 2 Nutrient content in animal species and products, including fat

	Water (%)	*Protein (%)*	*Fat (%)*	*Minerals (%)*
Species of domestic animals				
Cattle	54	15	26	4.6
Pig	58	15	24	2.8
Sheep	60	16	20	3.4
Hen	56	21	19	3.2
Mare	60	17	17	4.5
Products	*Fat content (g 100 g^{-1})*			
Pork	24.5			
Beef meat	6.2			
Mutton	8.9–31.6			
Turkey meat	5.7			
Lard	99.5			

tissue and the skin compared with other animals, such as pigs. Poultry meat, and especially chicken, is a valuable product, which contains about 20% of fat (in total with skin). Poultry meat without the skin has about 5% fat content. This meat has a lower level of saturated fatty acids (about 33% of all fat content) and a high level of unsaturated fatty acids – especially polyunsaturated fatty acids (PUFA) – 14%. Duck meat has a higher level of fat than chicken or turkey. In summary, fat content is about 35–45% with skin and under skin tissue, and 25–27% is saturated fatty acids (SFA). Meat from game birds like: grouse, partridge, pheasant, and pigeon has a fat content about 5, 7, 9 and 13% respectively, and one fourth of this is SFA. The composition of fatty acids is in an interesting new product in the human diet – ostrich meat – is given in Table 3.

An important characteristic of fat quality parameters is not only the relation between fat and the weight of the carcass but also the content of fatty acids SFA, monounsaturated fatty acids (MUFA), and PUFA.

Table 4 gives a comparison of the fatty acid content of different animal species. The highest content of PUFA is observed in poultry meat and the worse balance in ratios of fatty acids is in mutton or beef. Probably the genetic and nutrition sciences works could modify in future the profile relations of fatty acids, especially in beef meat.

A higher level of SFA is observed in ruminants than in monogastric animals (for example, SFA in ruminants may be about 50% of total fat content). This profile depends on specific rumen degradation.

In cattle there is more MUFA and PUFA in fat between the muscles compared with under the skin or around the organs. The quantity and quality of this intermuscular fat is a decisive factor in the cooking and taste of this meat (marbling of beef meat). In pig production the highest level of fat is accumulated in fat under skin tissue and in particular areas. In the carcass the deposition of fat is different: there is a low level in ham and high level in the shoulder.

Fish and especially oily fish is an important source of fat in the human diet in some countries. Fish may be classified according to the fat content in the carcass. In the human diet fish are classified as lean (low fat content) such as flatfish, cod, where the concentration of body fat is not greater than 5%, and oily fish, like salmon, trout, and tunny (total fat content is 9–20%). Ocean fish have a genetic tendency to accumulate the highest level of unsaturated fatty acids, in particular PUFA, which makes up about 20% of total fat (Table 5).

PUFA have an important function in human health. PUFA are classified in groups omega-3 and omega-6. A high ratio of omega-6 to omega-3 fatty acids in the human diet has been linked with an increased risk of cancer, cardiovascular disease, allergies, depression, obesity, and autoimmune disorders. Animal fat does not have a beneficial ratio of fatty acids from omega-6 to omega-3 groups. Animal products used in dietary health protection include fish oil, which contains the highest level of fatty acids in the omega-3 group.

One form of fatty acids is conjugated linoleic acid (CLA), which is found in ruminant meat. This fatty acid is healthy for humans. Studies have shown that CLA can change fat to muscle, reduce heart disease, and fight diabetes.

Table 3 Composition of fat in ostrich meat

Component	*Content*
Fat (g 100 g^{-1})	1.3
Cholesterol (mg 100 g^{-1})	57
Fatty acids (in % of total fatty acids)	
Saturated	
16:0	21
18:0	13
Monounsaturated	
16:1	4
18:1	31
Polyunsaturated	
18:2–6	16
18:3–3	6
20:4–6	5
20:5–3	0.4
22:6–6	0.7

Table 4 Fat level and fatty acid composition of muscles of different domestic animals

	Total fat (%)	*% in total fat content*		
		SFA	*MUFA*	*PUFA*
Beef fat	67	43	48	4
Sheep fat	72	50	39	5
Pig fat	71	37	41	15
Chicken fat (meat and skin)	18	33	42	19
Duck fat (meat and skin)	43	27	54	12

SFA, saturated fatty acid; MUFA, monounsaturated fatty acid; PUFA, polyunsaturated fatty acid.

Table 5 Fatty acid content in fish

Product	*Total fat (g 100 g^{-1})*	*20:5 n-3 (g 100 g^{-1})*	*22:6 n-3 (g 100 g^{-1})*
Herring	9.0	0.7	0.9
Mackerel	13.9	0.9	1.6
Salmon	5.4	0.3	0.9
Trout	7.7	0.1	0.5

Milk The highest concentration of fat is observed in sheep milk, in contrast with mares' milk, which has less fat. Cows' milk has a fat level of 3.2–4% (Table 6).

The chemical composition of milk lipids is made up of glycerophospholipids and sphingolipids (Table 7). The physical form of milk fat is a sphere (size 2–10 μm). In one molecule of fatty acid of milk there are 4–20 carbon atoms. In domestic animals the milk fats that predominate are fatty acids: palmitic (C16:0), oleic (C18:1), myristic (C14:0), and to a lesser extent, capric acid (in small ruminants). In unsaturated fatty acids linoleic acid (C18:2) and linolenic acids, have an important position in milk fat, but the ratio between these acids depends on the form and level of animal nutrition.

In human nutrition when raw milk is not consumed, the main function in diets is fulfilled by butter, which is a source of energy and vitamin sources. Butter has a high concentration of short-chain (saturated), fatty acids as well as vitamins A and D.

Eggs In laying hens there are two types of production: (1) eggs for reproduction: fatty acids are particularly important for the hatchability of eggs during incubation, and the quality and health of the chick; and (2) eggs for consumption: fatty acids have an influence on human health.

Eggs for consumption are a good source of lipids and vitamins. Fat in eggs is centered in the yolk, and is $32\,g\ 100\,g^{-1}$ of yolk weight. Important components of Hens' egg yolk are oleic acid ($11\,g\ 100\,g^{-1}$), linoleic acid ($2.61\,g\ \ 100\,g^{-1}$), and arachidonic acid ($0.50\,g\ \ 100\,g^{-1}$) (Table 8). Yolk triacylglycerols are about 60%, and they are the main source of fat energy. Monodiacylglycerols are 1.5–6.7%; the highest level is deposited in high-density lipoprotein

Table 6 Chemical composition of milk and fat content

Species	*Water (%)*	*Protein (%)*	*Fat (%)*	*Lactose (%)*	*Minerals (%)*
Cow	87	3–3.6	3.2–4.4	4.5–4.9	0.7–1.6
Goat	84–87	3.6–4.3	4–7	4.6–5.0	0.85
Sheep (ewe)	80–84	5.2–6.5	4.6–8.3	4.0–4.6	0.95
Mare	91	2	1.1–1.6	5.7–6.4	0.35
Reindeer	68	10.5	17–19	2–2.8	1.5
	% of total fat milk				
Triacylglycerides	98				
Diacylglycerides	0.25–0.48				
Monoacylglycerides	0.02–0.04				
Phospholipids	0.6–1.0				
Cholesterol	0.2–0.4				
Glycolipids	0.006				
Free fatty acids	0.1–0.4				

Table 7 Comparison of important fatty acids in milk from a variety of animals (molar percent)

Fatty acids	*Rabbit*	*Guinea pig*	*Mouse*	*Rat*	*Sow*	*Cow*	*Sheep*	*Goat*
Short-chain								
C4:0	< 1	< 1	< 1	< 1	The sum of C4:0	11	8	8
C6:0	< 1	< 1	< 1	< 1	to C12:0 = 2%	5	5	5
C8:0		< 1	< 1	6		1	4	4
Medium-chain								
C10:0	27	< 1	7	17		3	6	13
C12:0	3	< 1	12	13		3	5	7
C14:0	1	2	16	13	2	10	10	12
Long-chain								
C16:0	8	34	33	28	28	23	22	24
C18:0	2	3	2	3	6	10	10	12
C18:1	9	31	14	13	35	29	22	17
C18:2	10	23	8	7	14	2	4	3
C18:3	1	4	< 1	< 1	< 1	< 1	< 1	< 1

Table 8 Lipids in egg yolks from laying hens (%)

Specification	*Percentage share*
Triacylglycerides	56.1–65.5
Monodiacylglycerides	1.5–6.7
Free fatty acids	0.7
Cholesterol esters	0.1–0.5
Cholesterol	5.2–6.6
Phospholipids	28.3–31.2

(HDL) fraction and a lower level is in the low-density lipoprotein (LDL) plasma yolk fraction.

Modification of Quality and Quantity of Animal Fat

The scientific promoter of lipid metabolism, and in particular the influence of organic feed compounds on lipid synthesis in animals, was the German Oskar Kellner. Kellner ascertained that 1 kg of digested feed components such as starch, fat, or protein produced in cattle a particular level of fat, e.g., 1 kg starch produced 248 g fat, 1 kg protein produced 235 g fat. The results of trials in practice showed that in reality the levels of synthesis in organisms depended on nutrition value, the feed taken, the physiological conditions of the organism, age, digestibility, and the development of the digestive tract. One important point in lipid synthesis is the ratio between quality and quantity of feed consumption. The interaction between different feed components such as starch and cellulose also influences energy metabolism and energy balance.

Lipid synthesis in an organism is conducted differently in monogastric and polygastric animals. The level and system of digestibility and kinds of energy and lipid precursors (exogenous) are important.

Monogastric Animals (Domestic Animals, e.g., Poultry and Pigs)

The highest growth rate, highest protein synthesis and changing fat relation between protein content and fat in tissue are characteristic parameters of slaughtered monogastric animals. When breeding this group there are only two possibilities for changing body composition. One is genetic research and hybridization – there are new genetic lines of animals which have unusually high muscle content. The second way is a nutrition combination, using special balanced diets and special components in feed, which may modify fatty acid content (Figure 3).

Within the feeding system of young slaughtered poultry (chickens) there is a specific procedure relating to quality of feed. The important point is a ratio between energy content in feed and protein during the growth phase. The nutrition system must be correlated with body development in the short life span, which is only 6 weeks. This system allows increased protein synthesis in muscles compared with fat synthesis in the organism and is dependent on energy balance.

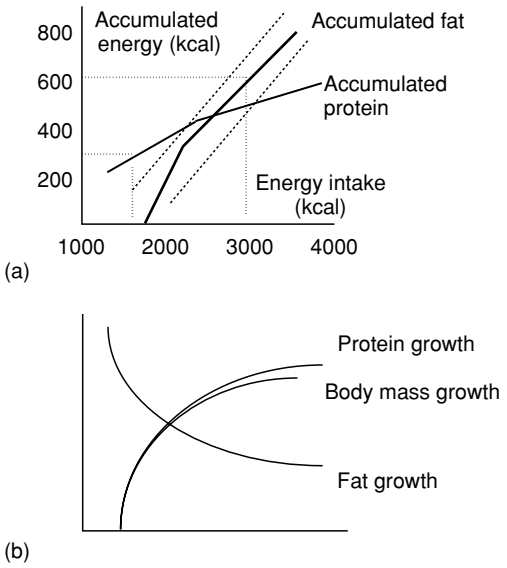

Figure 3 Influence of (a) feed energy intake and (b) proteinon body composition in poultry (chickens).

In chicken or turkey endogenous fat is synthesized from glucose, and the final product of carbohydrate degradation and digestibility – starch. In poultry diets 95% of starch is given by corn, wheat, and barley. The level of energy in poultry feed is balanced by exogenous fat supplementation by oil and tallow. The maximum level of this supplement is about 10% in feed, but is often 3–6%. The balance of energy in poultry is important, because when the balance increases by about 10–25%, fat synthesis is also increased. At this point feed intake is decreased and amino acids and minerals intake are also decreased (only in negative composition mixture).

The influence of systems, contents, and level of nutrition on development and quality poultry products has been determined in scientific research. Change in carcass composition is highest in poultry, especially in relation to fat and protein content. This suggests the possibility of modifying carcass composition by nutritional changes in a short time. One important factor is increased feed protein level and decreased growth rate of endogenous body fat.

A mean increase in protein concentration in the diet ($1\% = 10\,g\,kg^{-1}$) reduces fat content in the carcass of chicken 0.7–1.4% ($7–14\,g\,kg^{-1}$). A similar effect is observed using low-energy components in feed. Research trials have demonstrated that a high level

of rye plus enzyme in the diets decreased endogenous level of fat in the body composition of chickens compared with a corn diet. The specific combination of different energy sources in feed (tallow, soyabean oil, rapeseed oil, fish oil) may influence the quality and composition of endogenous fat in poultry. This correlation between composition of feed fat and carcass fat has been demonstrated in other studies. There was a strong correlation between dietary fatty acid composition and tissue fatty acid composition for all fatty acids except C14:0, C20:0, C20:2*n*-6, and C20:4*n*-6.

Not only energy sources in feed but also protein levels and sources have an effect on abdominal fat. In one study, protein concentrates were used in different combinations, compared with mixtures of a DKA-grower diet containing 11% fishmeal and 35% of a 1:1 mixture of faba bean and full-fat rapeseed and a DKA-grower diet containing 18% soyabean meal, 7% fishmeal, and 4% rapeseed oil. These nutrition combinations influenced differentials in abdominal fat composition: abdominal fat content was 12, 21, and 40 g. Of total fatty acids in abdominal fat, 37.55, 33.96 and 30.43% was saturated, and 62.43, 65.74 and 69.39% was unsaturated. When used in the feeding system in chickens, high-fat yeast (HFY; *Cryptococcus curvatus*) reduced the proportion of oleic acid in breast and thigh muscle in compare to chicks fed mixtures with tallow or without fat supplementation.

The main fat content in poultry is accumulated under skin tissue. The content of fatty acids in this tissue can be modified by different feed additives (Table 9).

The influence of nutritional factors on laying hens, performance and production is important. In reproduction eggs of laying hens, the lower content of linoleic acid or linoleneic acid gives rise to decreased hatchability, lower weight of embryos and chicks, and slower rate of development during incubation. Cholesterol has an important function in this process, too. Modification of consumption eggs is very popular, especially as regards changes of fatty acids and vitamin content. Changing eggs using the nutrition system is called nutritionally enriched eggs (NEE) (Table 10).

During feed modification of egg content, it is important that the cholesterol level is 90% of that required genetically.

In pork production the fatty composition of feed grains has an important function. In pigs fed mixtures with high levels of barley, the lard was white and strong, compared with pigs fed mixtures principally made up of corn, in which case the lard was yellow and clammy.

There are three important principles in pig production:

1. Pigs fed a diet with a lower protein level cannot produce a high level of protein in muscles (in relation to genetics);
2. The highest level of protein in a low-energy diet may increase fat synthesis;
3. All organic components in a low-energy diet may increase fat synthesis.

Table 9 Fatty acid composition in skin influenced fat supplements

	Specification of fatty supplements to mixture			
	Corn oil 6%	*Tallow 6%*	*Corn oil 5% and cod-liver oil 1%*	*Tallow 5% and cod-liver oil 1%*
Fatty acids				
Lauric (C12:0)	0.2	0.1	0.0	0.1
Myristic (C14:0)	0.8	2.9	0.8	3.4
Palmitic (C16:0)	16.1	23.4	15.9	23.3
Oleic (C18:1)	28.8	39.2	28.4	32.6
Linoleic (C18:2)	44.0	13.8	39.0	16.4

Table 10 Comparison of the contents of normal and modified eggs and feed additives which could change the fatty profiles in yolk

Feed additives	*Level of added components to mixtures*	*Nutrient content of NEE compared to ordinary egg: quantity/100 g egg content*		
			Ordinary egg	*NEE*
Fish oil	10–20 kg	Total saturated fatty acids (g)	3.3	2.8
Full-fat ground linseed	100–150 kg	Total unsaturated fatty acids (g)	6.4	6.9
Rapesced oil	5–10 kg	MUFA (g)	4.4	4.4
BHT, BHA, etoxyquin	100–250 g	PUFA (g)	1.2	2.2
Vitamin E	100–200 g	Linoleic acid (g) (n-3)	1.0	1.3
		Omega-6: omega-3 ratio	9.90	1.75
		Unsaturated-to-saturated acids ratio	1.94	2.46
		Cholesterol (mg)	400	350

NEE, nutritionally enriched eggs; BHT, butylated hydroxytoluene; BHA, butylated hydroxyanisole; MUFA, monounsaturated fatty acids; PUFA, polyunsaturated fatty acids.

Ruminants

The possibility of modifying fat quality and content in the body and milk of ruminants can be conducted genetically and nutritionally. In genetic terms cross-breeding special species of cattle with famous domestic lines is a popular route. For example, the Japanese meat-cattle Wagyu has a high intramuscular fat content (shown by marbling of the meat) and a high ratio of MUFA to SFA. A second example of genetic modulation in fatty acid composition in beef is cross-breeding between domestic cattle and bison in the USA, creating a new species which was called beefalo. Beefalo meat has less total fat, less saturated fat, less cholesterol and fewer calories in comparison with domestic breeds of cattle. In milk fat there are large differences between species of cattle, for example, Holstein cattle produce low fat milk (3.2–3.5%), Ayrshire and Brown Swiss give 3.9–4.0% fat and the highest content of fat is in milk produced by Guernsey and Jersey cows (4.6%). Milk fat and meat content can be modified by nutrition. Fattening young cattle on grass fields (extensive model) has decreased fat content in carcass in comparison with the intensive model of feeding mixtures with grass and high level of grains. Feeding milk-cattle a mixed ration with lower effective fiber reduces the fat content in milk. Similarly, supplementing the diet with *trans* fatty acids to inhibit milk fat synthesis decreased milk fat level.

In ruminants improving the profile of fatty acids in the feed does not directly improve the profile of fatty acids in products, whether milk or meat. In normal conditions cows only ingest 2.5–3% of fatty acids in dry matter in their daily feed ration. Fatty acids are delivered by triacylglycerides, galactolipids, and phospholipids by ruminal microbial lipase activity, and are modified to stearic acid. This form is available in the intestine. Unsaturated fatty acids from feed are biohydrogenated in the rumen to the saturated acid form and fat in the ruminant's body (tallow) are highly saturated. The fat of ruminant milk content compared to other fats has a high degree of short-chain fatty acids. The precursors of fatty acids in the rumen are carbohydrates which produce volatile fatty acids: acetate, propionate, and butyrate. Using a fresh grass feed in cows' nutrition could modify the content of healthy fatty acids in meat, giving a decrease in SFA, increased intake of omega-3 fatty acids and CLA. Studies have shown that the ratio between omega-6 and omega-3 fatty acids is decreased from 4:1 to 2.3:1 in the meat of cattle fed grass. A similar effect is observed in milk cows fed grass; the content of CLA was three or five times higher then in ruminants fed grain diets. The final way of modifying fatty acids in milk is to use protected fatty acids before rumen degradation.

See also: **Eggs**: Dietary Importance; **Fats**: Uses in the Food Industry; Digestion, Absorption, and Transport; Requirements; Fat Replacers; Classification; Occurrence; **Fatty Acids**: Properties; Metabolism; Gamma-linolenic Acid; Analysis; Dietary Importance; *Trans*-fatty Acids: Health Effects; **Poultry**: Chicken

Further Reading

Bączkowska H and Ślósarz A (1987) *Żywienie Drobiu*. Warsaw: Państwowe Wydawnictwo Rolnicze i Leśne.

Banaszkiewicz T, Janocha A and Osek M (1995) Wpływ mieszanek DKA z udzialem surowcow krajowych na wyniki poubojowe kurczat brojlerow. *Biuletyn Naukowy Przemyslu Paszowego* 34: 33–41.

Dhiman TR, Anand GR, Satter LD and Pariza MW (1999) Conjugated linoleic acid content of milk from cows fed different diets. *Journal of Dairy Science* 82: 2146–2156.

Dils RR (1983) Milk fat synthesis. In: Mephan TB (ed.) *Biochemistry of Lactation*, pp. 141–157. Amsterdam: Elsevier.

French P, Stanton C, Lawless F *et al*. (2000) Fatty acid composition, including conjugated linoleic acid, of intramuscular fat from steers offered grazed grass, grass silage or concentrate-based diets. *Journal of Animal Science* 78(11): 2849–2855.

Grudniewska B (1994) Hodowla i użytkowanie świń. Olsztyn: Wydawnictwo Akademii Rolniczo-Technicznej.

Horbańczuk O (2001) Nutritive value of ostrich meat. *World Poultry* 17: 42–43.

Jackisch B and Jeroch H (1993) Results of further studies of rye-containing feed mixtures with the enzyme preparation "Nutrizym-feed-additive" in broiler fattening. *Archives of Animal Nutrition* 42: 55–61.

Jamroz D (ed.) (2001) *Żywienie Zwierząt i Paszoznawstwo*. Warsow: Wydawnictwo Naukowe PWN.

Kean MG and Allen P (1998) Effects of production system intensity on performance, carcass composition and meat quality of beef cattle. *Livestock Production Science* 56: 203–214.

Larbier M and Leclercq B (1995) *Żywienie Drobiu*. Warsaw: Państwowe Wydawnictwo Naukowe PWN.

McDonald P, Edwards RA, Greenhalgh JFD and Morgan CA (1995) *Animal Nutrition*, 5th edn. Edinburgh: Addison Wesley Longman.

Miller GJ (1986) Lipids in wild ruminant animals and steers. *Journal of Food Quality* 9: 331–343.

Minakowski W (1990) Biochemia kręgowców. Warsaw: Państwowe Wydawnictwo Naukowe.

Munro HN (1976) Fat content and composition of animal product. National Academy of Sciences. /www.nap. edu/openbook.

Narahari D (2001) Nutritionally enriched eggs. 2001. *Poultry International* 40: 22–30.

Palmquist DL, Beaulieu AD and Barbano DM (1993) Feed and animal factors influencing milk fat composition. *Journal of Dairy Science* 76: 1753–1771.
Scaife JR, Moyo J, Galbraith H, Michie W and Campbell V (1994) Effect of different dietary supplemental fats and oils on the tissue fatty acid composition and growth of female broilers. *British Poultry Science* 35: 107–118.
Simopoulos AP and Robinson J (1999) *The Omega Diet*. New York, NY: Harper Perrenial.
Trziszka T (2001) Jajczarstwo. Wrocław: Wydawnictwo Akademii Rolniczej.
Willham RL (1976) Genetics of fat content in animal products. National Academy of Sciences. /www.nap.edu/openbook.

Uses in the Food Industry

L Boothby, Natland, Kendal, UK

Background

The first oil and fat products to be used by humans, apart from direct ingestion of vegetable oils from plants, were from animal sources. As the world's meat industry developed, animal fats became useful byproducts for baked goods. In the early 1900s, the introduction of hydrogenation, or hardening, revolutionized the usefulness of vegetable oils and enabled the production of fats and blends, with a wide range of physical properties overcoming some of the limitations associated with animal fats. Today's fat technologists also modify fats by fractionation and interesterification or a combination of two or more of the above methods to give a wide selection of tailor-made fats and blends. In more recent years, nutritional considerations have increased the usage of vegetable oils and fats. Similarly, the implication that *trans* acids in the diet increase the risk of coronary heart disease by raising low-density lipoprotein (LDL) cholesterol and reducing high-density lipoprotein (HDL) cholesterol levels has led some manufactures, particularly in yellow fats, to use blends with no hydrogenated oils. *Trans* acids also occur in animal and milk fats, albeit at lower levels than hydrogenated vegetable oils. (*See* **Vegetable Oils**: Dietary Importance.)

Functionality

In addition to being an energy source and fulfilling nutritional requirements, fats are necessary to carry out one or more of the following functions in various applications: (1) incorporate air; (2) provide lubrication; (3) soften texture; (4) aid moisture retention; (5) as a heat exchange medium and flavor enhancer in fried goods.

Bakery Fats

Properties

General-purpose fats or shortenings must have the following properties to perform satisfactorily in cake-making:

1. The shortening must be stable in the β-prime polymorph.
2. There must be a proportion of high melting triglycerols present: 5% is normally sufficient.
3. The ratio of solid to liquid triacylglycerols should be such as to give a plastic solid over the normal range of ambient temperatures.
4. The oxidative stability must be sufficient to meet product shelf-life requirements. (*See* **Oxidation of Food Components**; **Triglycerides**: Structures and Properties.)

Properties (1) and (2) are necessary to give a stable, aerated batter and good texture in the baked product. Property (3) is achieved by using a suitable blend of component oils. The shortening must blend readily with other ingredients during mixing. Plasticity also depends on the temperature history of the fat. During processing, rapid chilling is necessary to form a network of inter-locking small crystals which enmesh the liquid triacylglycerol phase. Slow chilling produces large, irregular crystals, and in products such as ghee a continuous liquid phase results. Storage temperature and ambient temperature in the bakery also affect plasticity. Property (4) is satisfied in most cases by using unmodified liquid oils for the liquid phase. If long-shelf-life products are required, oxidative stability is increased by using lightly hydrogenated oils for the liquid phase, e.g., soyabean oil with iodine value reduced from 130 to 105, or rapeseed oil reduced to iodine value 95 from 115. (*See* **Ghee**.)

General-purpose fats are suitable for pastry manufacture. β-Phase fats, notably lard, also produce excellent pastry products. Lard has a long plastic range but produces cake batters of poor stability and resultant poor cake texture. Its use in shortenings is thus limited to about 20% maximum. However, if costs permit, changing the polymorph to β-prime by interesterification significantly increases the inclusion level to 60–70%.

Component Oils

The physical characteristics of some typical shortening components are given in **Table 1**. Most fat

Table 1 Physical characteristics of some shortening components

Fat	Melting point[a] (°C)	Percentage of solid triacylglycerols at different temperatures (°C)					Stable polymorph
		10	20	30	35	40	
Butterfat	30	47	16	5	1	0	β-prime
Lard	36	49	26	8	5	2	β
Tallow	43	55	41	20	15	7	β-prime
Palm	37	50	22	11	5	2	β-prime
Palm (I)	42	52	35	15	9	6	β-prime
Palm (H)	49	95	92	80	63	42	β-prime
Palm stearin (F)	46	67	45	22	18	13	β-prime
Fish (H)	36	60	40	18	6	0.5	β-prime
Fish (H)	49	90	85	70	50	40	β-prime
Rapeseed (H)	33	75	50	15	4	0	β

[a]BSI slip melting point (BS684, Part I, 1.3). I, interesterified; H, hydrogenated; F, fractionated.

Table 2 Typical general-purpose shortening blends and pulsed nuclear magnetic resonance (pNMR) profiles

Component	Not all vegetable fat (%)		All vegetable fat (%)	
	1	2	1	2
Tallow	65			
Hydrogenated fish oil (30 °C)		80		
Hydrogenated fish oil (49 °C)		20		
Palm oil			25	20
Hydrogenated palm oil (49 °C)			5	10
Hydrogenated rapeseed oil (33 °C)			40	25
Rapeseed oil	35		30	
Hydrogenated rapeseed oil (20 °C)				45
Percentage of solid triacylglycerols (pNMR)				
10 °C	35	45	47	45
20 °C	22	27	20	22
30 °C	12	10	12	10
40 °C	3	2	1	1.5

manufacturers have their own stock of standard hydrogenated or fractionated oils. There is no justification, for example, in a refiner carrying a range of hydrogenated soyabean oils with melting points ranging from 20 °C to 50 °C in steps of 5 °C since most user requirements can be met by blending a limited number of hardstocks. All the figures quoted for solid glyceride levels were determined by pulsed nuclear magnetic resonance (pNMR) using International Union of Pure and Applied Chemistry (IUPAC) method 2.150. Tallow, hydrogenated marine oils, hydrogenated cottonseed oil, hydrogenated palm oil, and palm stearin are all β-prime fats used extensively as hardstocks for shortening blends. (*See* **Fish Oils**: Composition and Properties.)

Shortening Blends

Typical shortening blends are given in Table 2, together with their pNMR profiles. Fat suppliers aim to have a library of formulations which can be used to give identical end products, so that least-cost options are always available. Shortening blends without middle melting triacylglycerols give satisfactory baked goods but there can be some loss of eating quality due to the increased level of high-melting triacylglycerols. Advantages of this type of formulation are an extended plastic range. Most manufacturers increase solid triacylglycerol levels by 2–4% during summer months. Hydrogenated marine oil-based shortenings possess excellent cake-making properties but the increasing all-vegetable trend has seen a decline in this type of shortening.

High-ratio shortenings have similar blends to the general-purpose fats, except for the addition of 2.8–3.2% α-monoacylglycerol, and in some cases a similar level of diacylglycerols. High ratio refers to the higher levels of liquor and sugar to flour that can be used. Resultant cakes are moister, sweeter, and resist staling longer than conventional cakes. Shortenings with 1–2% monoacylglycerol are used to give intermediate ratios of liquor and sugar.

Pourable or fluid shortenings are suspensions of emulsifiers, sometimes with the addition of a small level of high-melting β-phase fat, either in liquid vegetable oil, or in a lightly hydrogenated vegetable oil to give improved shelf-life.

Pumpable shortenings for large bakeries have similar blends to conventional shortenings. They are stored and delivered in jacketed tanks with temperatures maintained between 24 and 28 °C depending on the blend. At the bakery the shortening is delivered from controlled temperature storage by positive displacement pumps and metered into the mixer. (*See* **Bread**: Dough Mixing and Testing Operations.)

Puff Pastry Fats

Puff pastry fats need to be tough and elastic to achieve the multiple alternate layers of fat and dough required to produce good lift. Puff pastry margarines seem to be preferred to fats. To achieve the necessary consistency, relatively high-solid glyceride contents are used together with correct processing. An excellent puff pastry fat can be produced from lard containing 12–15% fully hydrogenated lard. Tallow-based fats are also produced. Vegetable-based fats utilize palm stearin and, hydrogenated palm oil together with other hydrogenated vegetable oils.

Biscuit Fats

Dough Fats

Biscuit dough fats must be β-prime-stable to avoid bloom development during the life of the biscuit. However, certain different criteria are required compared to shortenings:

1. The melting profile must be sharper for good eating quality. There must be generally less than 5% solid triacylglycerols at 35 °C and 0.5% maximum at 40 °C.
2. The oxidative stability needs to be high for a long-shelf-life product, typically 40 h minimum at 100 °C (measured by Rancimat instrument). Thus, liquid oils are normally excluded from blends, except for some specialized 'health' biscuits.

Typical dough fat blends and pNMR profiles are given in Table 3. Historically, the biscuit industry used large quantities of hydrogenated marine oils. Palm or hydrogenated palm oils are normally used to achieve the β-prime polymorph in vegetable blends.

Biscuit fats can be supplied ready for use in boxes. The major producers receive blends in road tankers at temperatures several degrees above the complete melting point, store in suitable tanks, and subsequently chill and texturize as required.

Cream Fats

Biscuit cream fats require medium to high solid triacylglycerol levels at ambient temperatures, together with a rapid melting profile, and, ideally, zero solid triacylglycerols at body temperature. On high-speed production lines the fat must set quickly and give good coherence to the biscuit. Biscuit creams are basically mixtures of icing sugar and fat with or without added colors and flavors. Typical cream fats and their melting profiles are given in Table 4.

The lauric oils – coconut and palm kernel – are ideal starting materials. In addition to rapid melting they are very bland and can be refined to very low color levels. Hydrogenated rapeseed and soyabean oils with high *trans* acid levels can be used singly or in combination with lauric fats. In these blends low-melting saturated acids are partially replaced by *trans* acids – mainly elaidic acid, which has a similar melting point to lauric acid (42 °C).

Low *Trans* Acids Blends

Suitable low *trans* acids blends can be formulated using fractionated or interesterified oils. Examples are given in Table 5 for different applications. To achieve zero to trace levels of *trans* acids it is necessary to use deodorization temperature below 235 °C, particularly if liquid oils are included in the blend; this precludes the use of physically refined oils.

Frying Oils

Frying oils need to be as stable as possible under the harsh conditions used. This is achieved by using oils with the lowest possible level of unsaturation concomitant with good eating quality and very low trace-metal levels. Thus, hydrogenated oils should be posttreated to remove traces of catalyst completely. Stability is also important to achieve a reasonable shelf-life, particularly in potato crisps and snack foods, with their large surface area to weight ratios, and oil contents up to 40%.

Oils are delivered in bulk to the snack-food industry. Suitable oils are lightly hydrogenated soya bean or rapeseed, with linolenic acid levels reduced below 3%, palm olein, and blends of palm oil with vegetable oils. Sunflower is used successfully to produce health-branded products. Frozen chip producers require harder oils to avoid problems during blast-freezing. Suitable oils are palm or hydrogenated palm oil.

Table 3 Examples of biscuit dough fat blends and pulsed nuclear magnetic resonance (pNMR) profiles

Animal/vegetable	*Fat (%)*	*All vegetable*	*Fat (%)*
Hydrogenated fish oil (32 °C)	55	Hydrogenated soya oil (35 °C)	50
Palm oil	35	Palm olein	10
Hydrogenated palm oil (42 °C)	10	Palm oil	40
Percentage of solid triacylglycerols (pNMR)			
20 °C	25		28
30 °C	8		10
35 °C	2.5		4
40 °C	0		0.5

Table 4 Examples of biscuit cream fats and melting profiles

	Fat (%)	*Fat (%)*	*Fat (%)*
Coconut oil	65		
Hydrogenated palm kernel oil (34 °C)	15	100	
Hydrogenated palm oil (42 °C)	20		
Hydrogenated soyabean oil (34 °C)			100
Percentage of solid triacylglycerols (pNMR)			
20 °C	41	74	60
30 °C	4.5	14	20
35 °C	2	2	3
40 °C	0	0	0

pNMR, pulsed nuclear magnetic resonance.

Table 5 Low *trans* acids blends and pulsed nuclear magnetic resonance (pNMR) profiles

Component oils	*General shortening*		*Biscuit dough*		*Biscuit cream*	
	1	*2*	*1*	*2*	*1*	*2*
Palm stearin/coconut 80:20 (I)	15					
Palm (I)	40			85		
Palm	20					
Rapeseed oil	25					
Palm/rapeseed oil 83:17 (I)		100				
Palm/coconut oil 70:30 (I)			100			
Palm kernel oil				15		
Palm kernel stearin (F)					100	
Fully hydrogenated palm/palm kernel 30:70 (I)						100
Percentage of solid triacylglycerols (pNMR)						
20 °C	24	23	29	27	70	57
30 °C	10	10	8.5	11	8	21
35 °C	3	5	2	5	0	4
40 °C	1	1	0	1		0

These products contain 2–5% oil, which is diluted by the end users' frying media.

The catering industry uses palm oil, beef-tallow liquid vegetable oils, and long-life frying oils based on lightly hydrogenated vegetable oils. The latter normally contain a low level of a β-phase fat such as fully hydrogenated rapeseed oil. Chilling and texturizing produce a stable pourable frying oil. An antifoaming agent (methyl polysiloxane) is generally added to long-life frying oils. Antioxidants are not normally added; they give protection prior to use but there is little carry-through to the finished product. Residual tocopherols in vegetable oils give some benefit. (*See* **Antioxidants**: Synthetic Antioxidants; **Tocopherols**: Properties and Determination.)

Processing

Shortenings

Fats and blends are stored above their complete melting points immediately prior to processing. The first stage is normally a water precooler to reduce variations in feed temperature to the main chilling unit. The main chilling unit is a continuously scraped surface heat exchanger, comprising a cylinder with refrigerant on the outside, and inside a rotating shaft with floating scraper blades. The molten fat is pumped under high pressure through the annular space whilst the scraper blades remove chilled fat from the cylinder wall. This step initiates the crystallization before passing to the second unit, sometimes referred to as a 'B' unit or 'puddler' or 'worker.' In the 'B' unit the fat completes most of its crystallization whilst being agitated by a combination of paddles and baffles; most of the heat of crystallization is dissipated.

Absence of a 'B' unit results in nonplastic shortening. Processing plants normally consist of several chilling and 'B' units linked in series. The final stage involves further mechanical working by passing the shortening through an extrusion or texturizing valve immediately prior to filling the container – normally polythene-lined boxes. The whole system generally operates at pressures between 2 and 3 MPa. Depending on the blend, chilling temperatures can vary between 16 °C and 22 °C and filling temperatures between 19 °C and 25 °C.

The final stage after filling is 'tempering' or 'annealing'. Containers are stored in a temperature-controlled environment between 24 °C and 28 °C for at least 48 h. This step allows the shortening to stabilize in the β-prime polymorph and extends the plastic range, so enabling the shortening to be used over a wider temperature range without loss of performance.

Most shortenings have air or, preferably, nitrogen introduced prior to chilling to give a white appearance; the normal level varies between 8% and 15% by volume. The level of gas does not influence final performance and is introduced for cosmetic reasons. Shortenings without gas perform equally well but have a pale yellow, petroleum jelly-like appearance. Colors, such as β-carotene and vitamins, can be added if required. Colored shortenings do not normally include nitrogen.

Pumpable shortenings are processed similarly but after texturizing are stored in special temperature-controlled tanks at tempering temperatures. Further work softening occurs when the shortening is pumped into and out of the delivery vehicle, and again just prior to mixing by the end user. Omission of nitrogen from pumpable shortenings gives more accurate metering into the mixer.

Puff Pastry Fats

Puff pastry fats can be chilled in a similar manner to shortenings or on open chilling drums with a scraper

blade running the length of drum and refrigerant on the inside. Moisture condensation can give problems in the latter process. After chilling, the fat is allowed to crystallize slowly in long resting tubes or in skips. The fat is then plasticized before extruding into blocks. The drum method gives fats of excellent plasticity. No tempering is necessary but low storage temperatures should be avoided.

Other Fats

Confectionery, toffee, icecream, cream, and frying fats are chilled, texturized, and packed as shortenings, but without tempering. (*See* **Ice cream**: Methods of Manufacture.)

See also: **Antioxidants**: Synthetic Antioxidants; **Bread**: Dough Fermentation; **Fish Oils**: Composition and Properties; **Ghee**; **Ice Cream**: Methods of Manufacture; **Oxidation of Food Components**; **Tocopherols**: Properties and Determination; **Triglycerides**: Structures and Properties; **Vegetable Oils**: Types and Properties; Oil Production and Processing; Dietary Importance

Further Reading

Bailey AE (1985) In: Applewhite TH (ed.) *Industrial Oil and Fat Products*, 4th edn, vol. 3. New York: John Wiley.

Baldwin RR, Baldry RP and Johansen RG (1972) Fat systems for bakery products. *Journal of the American Oil Chemists Society* 49: 473–477.

Hoerr CW (1960) Morphology of fats, oils and shortenings. *Journal of the American Oil Chemists Society* 37: 539–546.

Knightly WH (1981) Shortening systems: fats, oils and surface-active agents – present and future. *Cereal Chemistry* 58: 171–174.

Lefebvre J (1983) Finished product formulation. *Journal of the American Oil Chemists Society* 60: 295–300.

Vaisey-Genser M and Ylimaki G (1989) Baking with canola oil products. *Cereal Foods World* 34: 246–258.

Digestion, Absorption, and Transport

T H M Da Costa, Universidade de Brasília, Brasília, DF, Brazil

Background

Fats or lipids are the concentrated form of energy in the diet, providing $9\,\text{kcal}\,\text{g}^{-1}$ ($37\,\text{kJ}\,\text{g}^{-1}$), and have evolved as the most advantageous form of energy storage, owing to their hydrophobic characteristics. Triacylglycerols are the most common of the lipids either in the diet or as lipid tissue stores. They are formed by the esterification of glycerol and fatty acids (1:3 molar ratio). Triacylglycerols vary depending on the three fatty acid residues they contain; simple triacylglycerols contain a single type of fatty acid residue. For example, trioleoylglycerol or triolein (a simple triacylglycerol) has three oleic acid residues, and mixed triacylglycerols contain two or three different types of fatty acid residues. The fatty acid moiety varies greatly in nature, and the ingestion of triacylglycerols containing saturated or unsaturated, long- and/or short-chain fatty acids depends on the source of lipids in the diet.

Fat is an important food energy source in the Western world, ranging from 30 to 40% of daily energy intake, equivalent to the daily consumption of 50–120 g in an adult. Other lipids play less important roles as energy sources under normal physiological circumstances, but serve as structural components of cell membranes, metabolic cofactors and as hormones (phospholipids, sphingolipids, cholesterol, diacylglycerol, and steroids). The diet constitutes the major source of lipids for the body and supplies essential fatty acids and lipid-soluble vitamins that cannot be synthesized endogenously. Lipogenesis, *de novo* synthesis of triacylglycerols from carbohydrate, occurs when there is a high ingestion of the latter.

The first part of this chapter reviews present knowledge on the mechanisms of digestion and absorption of dietary lipids, mainly triacylglycerols, and the second part deals with the complex transport system of lipids in the circulation.

Lipid Digestion and Absorption

Owing to their hydrophobic characteristics, lipids are digested by a dynamic and complex mechanism that facilitates their solubilization and hence availability for enzymatic action. The process can be divided into four main phases: emulsification, hydrolysis of fatty acid ester bonds by lipases, aqueous dispersion of lipolytic products in bile acid micelles, and uptake by the enterocytes (**Figure 1**).

In the process of fat digestion, there is a minor contribution of lingual lipase, which hydrolyzes some long-chain fatty acid triacylglycerols, and gastric lipase, which contributes to the hydrolysis of short- or medium-chain fatty acid triacylglycerols. The main process of lipid digestion takes place in the small intestine. In the duodenum, triacylglycerols aggregate in droplets and are emulsified by the action of the bile salts, monoacylglycerols (originating from intragastric lipolysis), and lysolecithin. The fat

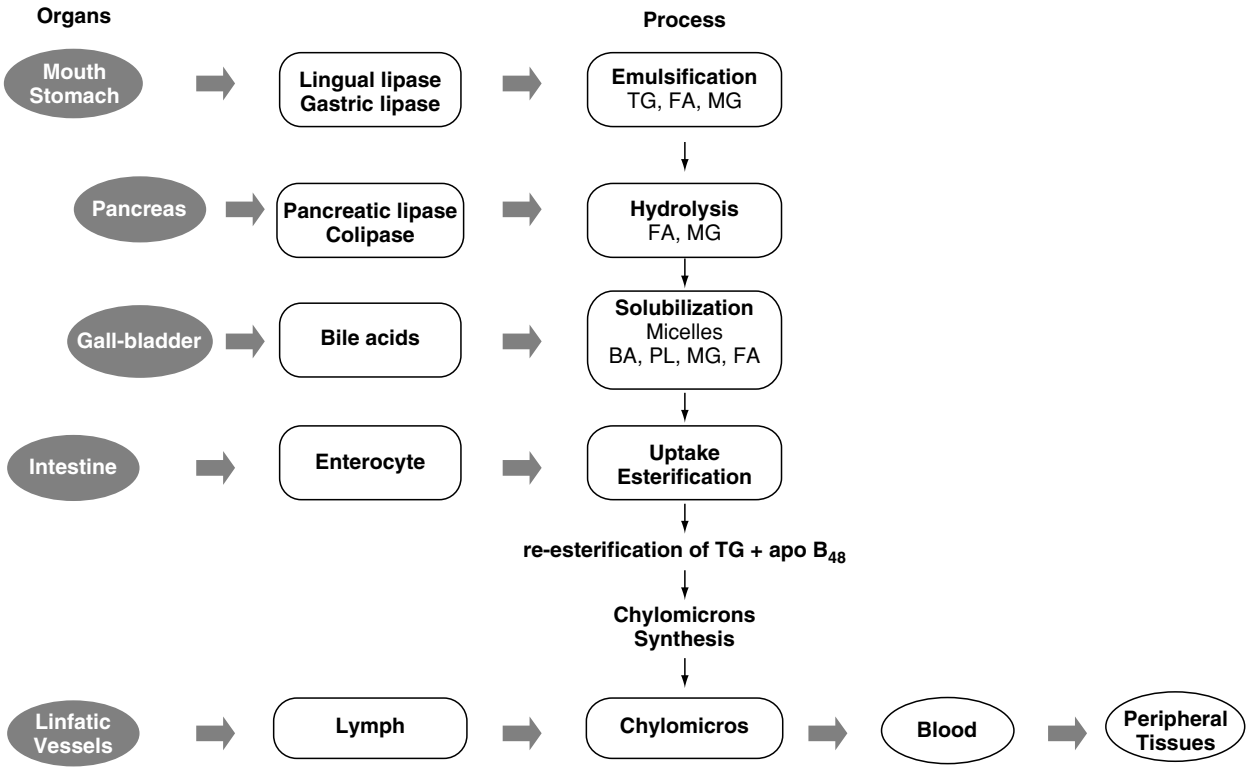

Figure 1 Process of triacylglycerol digestion and absorption showing the organs, enzymes, and processes involved. TG, triacylglycerols; FA, fatty acids; MG, monoacylglycerol; BA, bile acids; PL, phospholipids.

droplets are attacked by the colipase and pancreatic lipase complex. Colipase, a pancreatic protein that protects the lipase from denaturation and anchors it to the lipid–water interphase of the droplet, permits access of the lipase to the inner core of triacylglycerols. In fact, when the lipase binds to the interphase of the lipid droplet, the catalytic rate of the enzyme is increased (surface activation). Pancreatic lipase hydrolyzes ester links in the 1–3 positions of the triacylglycerols releasing fatty acids and 2-monoacylglycerol (Figure 2).

Phospholipids from the diet or secreted in the bile (lecithin) are attacked by pancreatic phospholipase A_2, which cleaves the fatty acid at the 2 position, forming lysophospholipids and lysolecithin, respectively. Lysolecithin aids in the emulsification process. Cholesteryl esters, the main form of cholesterol in the diet, are hydrolyzed to free cholesterol by a pancreatic enzyme cholesterol esterase and aggregate into micelles. Cholesterol esterase also hydrolases retinyl esters. Its action increases the polarity of cholesteryl and retinyl esters, which are highly hydrophobic molecules, allowing their micellar solubilization. Free cholesterol derived from biliary secretion and cell decay is also incorporated into micelles. When the proportion of bile acids and phospholipids reaches a critical micellar concentration, micelles are formed and take up free fatty acids and monoacylglycerols, and the expanded micelle is water-soluble.

In the aqueous interface between the micelles and the epithelial cell surface, known as unstirred water layer, the micelles dissociate, and the concentration of fatty acids, monoacylglycerols, and lysophosphoglycerols increases, and these diffuse freely through the apical enterocyte membrane. Saturated fatty acids with a carbon length of 18 are poorly absorbed in the free state but well absorbed as 2-monoacylglycerols. Very-long-chain fatty acids with high melting points are less efficiently absorbed. Triacylglycerols and diacylglycerols, which are not acted upon by the lipase, do not dissolve well in bile salts and remain in lipid droplets. Unabsorbed fat in feces is usually below 5%. Bile salts that are separated from the micelles remain in the intestinal lumen and are absorbed in the distal ileum by an active transport system or are lost in the feces.

Short-chain (C:2–C:4) and medium-chain-length (C:6–C:14) fatty acids are more hydrophilic and are rapidly absorbed independent of micellar solubilization or re-esterification within the intestinal cells. They are transported via the portal circulation directly to the liver. In the newborn, about half of the

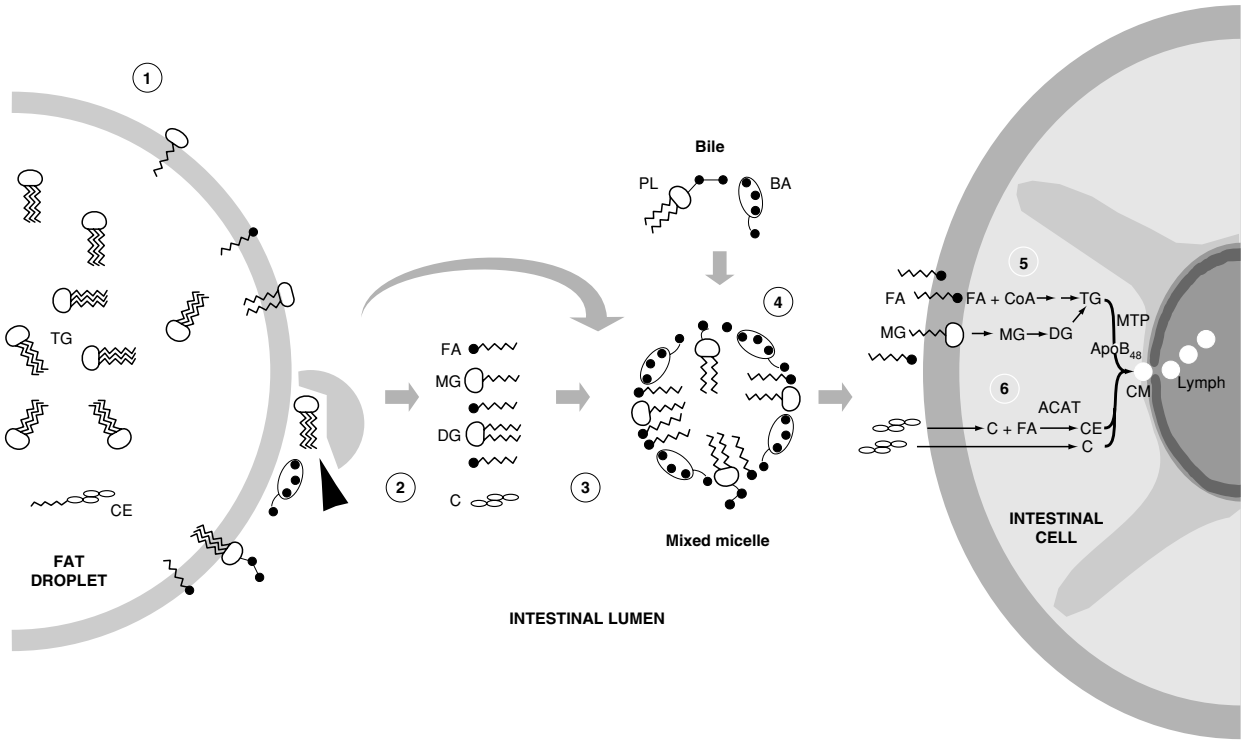

Figure 2 Diagrammatic representation of triacylglycerol and cholesterol digestion and absorption in the gastrointestinal tract. A fat droplet is shown (1) with insoluble triacylglycerols (TG) and cholesteryl ester (CE) in the core. The products of early lipolysis by lingual and gastric lipases: fatty acids (FA), diacylglycerol (DG) and monoacylglycerol (MG) and also dietary phospholipids with their polar heads face the outside of the droplet. Pancreatic lipase (2) acts at the water–oil interphase hydrolyzing TG molecules, with the participation of bile acids (BA) and colipase, which anchors the lipase to its catalytic site. Cholesterol esterase acts on cholesteryl ester forming free cholesterol (C). Lipolytic products (3) interact with bile acids (BA) and phospholipids (PL), facilitating their solubilization into micelles (4). In the unstirred water layer, micelles dissociate, and their content diffuses to the enterocyte. Inside the intestinal cell, fatty acids (FA) are activated and are re-esterified to TG (5). Cholesterol is absorbed and re-esterified with a fatty acid by the action of the enzyme acylCoA:cholesterol acyltransferase (ACAT) originating cholesteryl ester (CE) (6). Chylomicron (CM) synthesis requires the enzyme microsomal triacylglycerol transfer protein (MTP) to assemble the structural protein apo-B_{48} to triacylglycerols. Mature chylomicrons also incorporate cholesteryl esters to the core, and phospholipids, free cholesterol and apoplipoproteins to the surface. Chylomicrons exit the enterocytes by exocytosis to the lymph. Adapted with permission from Ros E (2000) Intestinal absorption of triglyceride and cholestrol. Dietary and pharmacological inhibition to reduce cardiovascular risk. *Atherosclerosis* 151: 357–379.

medium-chain fatty acids from milk can be absorbed in the stomach. In the enterocyte membrane, a fatty acid-binding protein (FABP) facilitates the transport of long-chain fatty acids (carbon chain length > 12) into the cytosol and reduces the potential cytotoxic effect of fatty acid soaps. Cholesterol binds to a sterol carrier protein as it enters the cell. In the enterocyte, cholesterol is re-esterified with a fatty acid by the action of the enzyme acyl-CoA:cholesterol acyltransferase (ACAT). The fatty acids and monoacylglycerols are re-esterified to triacylglycerols within the endoplasmatic reticulum of the enterocyte, and then are combined with phospholipids, cholesterol, cholesteryl esters, and apolipoproteins, forming chylomicrons. Apo-B_{48} is the structural protein of chylomicrons. The process of chylomicron assembly takes place in the Golgi apparatus, and the movement of triacylglycerides from the endoplasmatic reticulum to the Golgi appears to be a rate-limiting step in the process. The union of triacylglycerol to apo-B_{48} requires a microsomal triacylglyceride transfer protein (MTP). Finally, chylomicrons are released by exocytosis to the lymphatic system and enter the blood circulation by the left subclavian vein (Figure 2).

Lipid Transport in the Plasma

Lipids are transported in the hydrophilic blood plasma environment as lipoproteins. Lipoprotein particles consist of a core of triacylglycerol and cholesteryl esters, in varying proportions, surrounded by a more hydrophilic layer of phospholipids, unesterified cholesterol, and protein. In the circulation or as a constitutive component, each lipoprotein particle

contains specific proteins, termed apolipoproteins (Figure 3).

There are at least 12 apolipoproteins described and present in different proportions among the lipoproteins (apo-AI, II and IV, Apo-B_{48} and B_{100}, apo-CI, II and III, apo-D and apo-EII, III and IV). The apolipoproteins, apart from giving stability to the lipoprotein particle, also direct the lipoprotein to its site of metabolism by interactions with specific enzymes and transport proteins on cell membranes. The ratio of lipid to protein, which determines the density of lipoproteins, is important in the classification of the five major lipoproteins: chylomicrons; VLDL, very-low-density lipoprotein; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; and HDL, high-density lipoprotein. As the lipoproteins are delipidated, the diameter reduces, and the density increases. The lipoproteins exist in a dynamic state, and their separation into one of the above classes, according to their flotation characteristics, refers to average values. Chylomicrons, the least dense of all lipoproteins, have the highest proportion of triacylglycerols (85–90%) and transport mainly lipids from exogenous (dietary) origin. As already mentioned, during their assembly in the enterocytes, they acquire apo-B_{48}. Nascent chylomycrons also contain apo A-I, apo A-IV, and, to a lesser extent, apo A-II. Apo A-I and apo A-II enter the circulation in chylomicrons but are rapidly transferred to nascent HDL (Figure 3). Apo A-IV secreted by the enterocytes is likely to mediate the anorectic effect of a lipid meal. It is proposed that apo A-IV acts as endogenous antioxidant and is involved in fat absorption.

VLDL, IDL, and LDL are related particles that transport endogenous (synthesized in the liver) triacylglycerols and cholesterols to peripheral tissues.

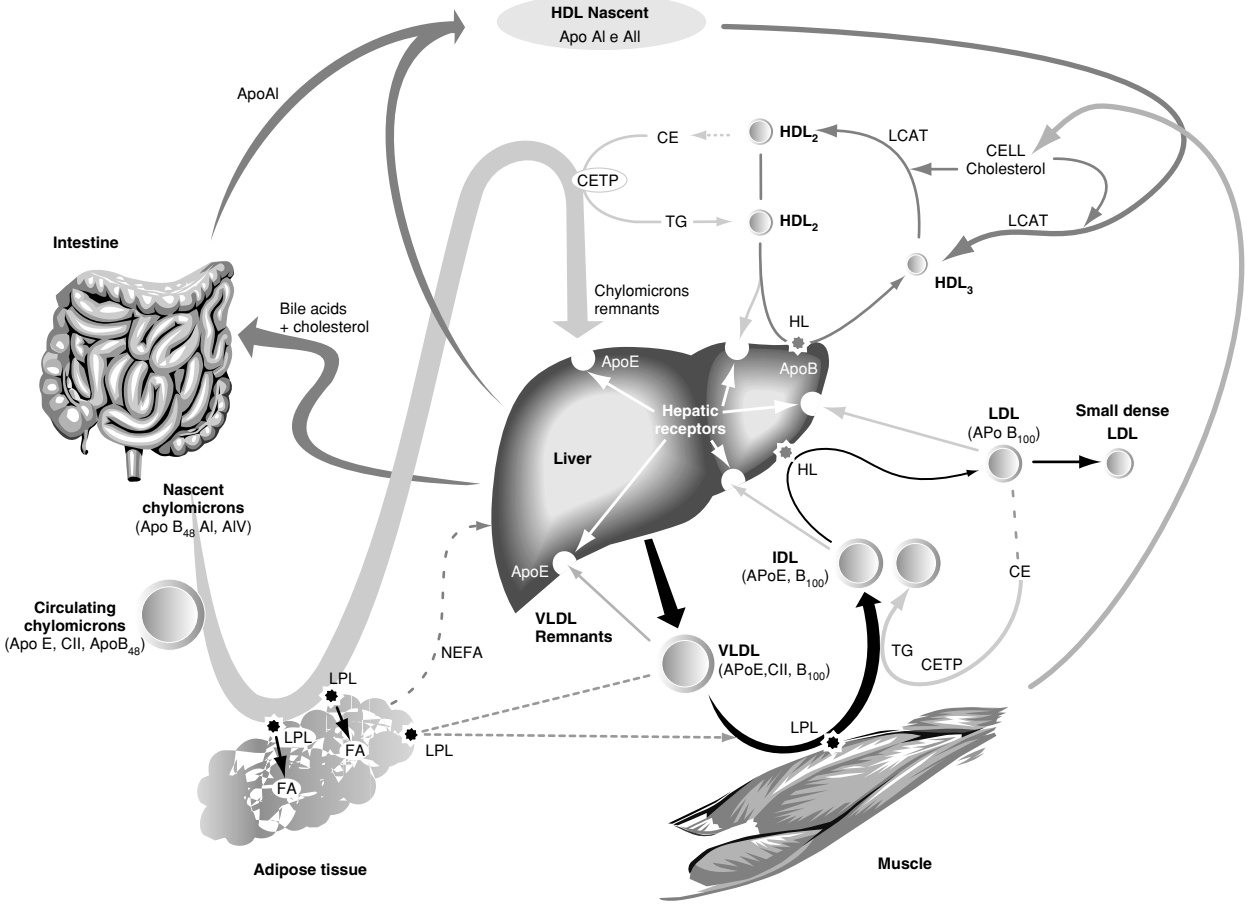

Figure 3 Scheme for the transport of triacylglycerol and cholesterol in plasma. The classes of lipoproteins are shown: chylomicrons and chylomicrons remnants; VLDL, very-low-density lipoprotein; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; HDL, high-density lipoprotein, with its subclasses: nascent-HDL, HDL_2, and HDL_3. The major apolipoproteins associated with each lipoprotein are presented in parenthesis. Enzymes involved in the process of lipoprotein metabolism: LPL, lipoprotein lipase; HL, hepatic lipase. Enzymes related to lipoprotein lipid exchange: LCAT, lecithin:cholesterol acyltransferase; CETP, cholesterol:ester transfer protein. Hepatic receptors are indicated with their ligand for binding to cell-surface Apo-E or Apo-B. For more details, refer to the text. Adapted with permission from Packard and Shepherd (1996).

HDL, which is assembled in the plasma through the action of LCAT (lecithin:cholesterol acyltransferase) from hepatic or enteric nascent HDL, transports cholesterol from peripheral tissues to the liver and functions as an acceptor/donor of apolipoproteins to other lipoproteins (Figure 3). Among the lipoproteins that transport endogenously synthesized lipids, VLDLs are the principal source of triacylglycerols for the tissues (50–55%).

In the postprandial state, circulating chylomicrons interact with HDL and acquire apolipoproteins E and CII. Apolipoprotein CII activates the enzyme lipoprotein lipase in the capillary endothelium of adipose tissue and skeletal muscle. The interaction between the particle and the enzyme causes hydrolysis of the triacylglycerols in the core of the chylomicron. Repeated action of the lipases as the blood circulates removes most of the triacylglycerols, and the resulting particle (chylomicron remnant) is enriched in cholesteryl esters, apo-B_{48}, and apo-E. Apo-CII returns to HDL. In the liver, specific remnant receptors are activated by apo-E, and the remnants are taken into the cell by endocytosis. Lysosomal hydrolysis in the hepatocytes liberates fatty acids, cholesterol, and amino acids. The overall result of the chylomicron transport process is completed as dietary triacylglycerols are delivered to peripheral tissues and cholesterol to the liver.

Lipids within the liver, formed by *de novo* synthesis from carbohydrates (lipogenesis), chylomicron remnants, and nonesterified fatty acids (NEFA) mobilized from the adipose tissue, are incorporated into VLDL particles and secreted into the blood stream during the postabsorptive or fast period. The assembly and secretion of hepatic VLDL is a complex process, and the final particle is composed of a relatively rich core of triacylglycerols with smaller amounts of cholesteryl ester, surrounded by a monolayer of phospholipids and cholesterol enclosed by apo-B_{100} that differs from the apo-B_{48} secreted by the enterocytes. In the hepatocytes, MTP plays a comparable role, as in the enterocytes, in the assembly of triacylglycerides with apo-B_{100}. Secreted VLDL receives apo-CII and apo-E from HDL and is acted upon by the lipoprotein lipase in the same way as the chylomicrons. The resultant lipoprotein of VLDL is an IDL. Some of the IDL are catabolized in the liver through binding to a specific receptor. The IDL particles that remain in the plasma undergo transformation resulting in the conversion to LDL. In this process, all the apolipoproteins are removed, with the exception of apo-B_{100}. LDL is almost entirely composed of cholesteryl esters surrounded by apo-B_{100}. LDL can be further delipided, resulting in small dense LDL that are more atherogenic. Circulating LDL delivers cholesterol to hepatic and extrahepatic tissues, and the process involves the interaction with a specific receptor in the cell membrane (LDL receptors).

In the liver, cholesterol is converted to bile acids or incorporated into the bile and excreted into the intestine to facilitate lipid absorption. Hepatic cholesterol is also distributed to other tissues as it is incorporated into VLDL. Circulating HDL capture plasma unesterified cholesterol formed from cell degradation and convert it to cholesteryl ester through the action of LCAT. The cholesteryl ester moves to the core of the particle. LCAT is activated by apo-AI, the major apolipoprotein present in the HDL (Figure 3). Major subclasses of HDL are designated HDL_2 and HDL_3 based on density. By the action of LCAT, HDL_3 is produced from discoidal HDL and HDL_2 from HDL_3. HDL_2 also exchange lipids with other lipoproteins (chylomicrons and VLDL fractions) through the action of the enzyme cholesterol ester:transfer protein (CEPT). Hepatic lipase acts on IDL and HDL_2-triacylglycerol and a phospholipase on HDL_2-phospholipid, which generates LDL and HDL_3, respectively (Figure 3). These reactions establish a cycle by which cholesterol is removed and returned to peripheral tissues or is disposed of by the liver.

See also: **Cholecalciferol**: Properties and Determination; Physiology; **Cholesterol**: Absorption, Function, and Metabolism; **Coronary Heart Disease**: Etiology and Risk Factor; **Fatty Acids**: Properties; Metabolism; Dietary Importance; **Phospholipids**: Properties and Occurrence; Physiology; **Retinol**: Properties and Determination; Physiology; **Tocopherols**: Properties and Determination; Physiology; **Triglycerides**: Structures and Properties; Characterization and Determination

Further Reading

Applebaum-Bowden D (1995) Lipases and lecithin:cholesterol acyltransferase in the control of lipoprotein metabolism. *Current Opinion in Lipidology* 6: 130–135.

Gibbons GF, Islam K and Pease RJ (2000) Mobilization of triacylglycerol stores. *Biochimica et Biophysica Acta* 1482: 37–57.

Hussain MM, Kedees MH, Singh K, Athar H and Jamali NZ (2001) Signposts in the assembly of chylomicrons. *Frontiers in Bioscience* 6: D320–D331.

Ros E (2000) Intestinal absorption of triglyceride and cholesterol. Dietary and pharmacological inhibition to reduce cardiovascular risk. *Atherosclerosis* 151: 357–379.

Spiller GA (ed.) (1996) *Lipids in Human Nutrition*. New York: CRC Press.

Van Greevenbroek, MMJ and de Bruin TWA (1998) Chylomicron synthesis by intestinal cells *in vitro* and *in vivo*. *Atherosclerosis* 141: S9–S16.

Requirements

N M F Trugo and A G Torres, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

Background

Dietary fat is a term used to designate the major dietary lipids, triacylglycerols. Other dietary lipids are the phospholipids, sterols, waxes, fat-soluble vitamins, carotenoids, and other minor compounds. Fats are required in the human diet to provide energy and essential fatty acids, and to improve the absorption of fat-soluble vitamins, among other functions. They can also enhance food palatability and influence food texture.

Determination of fat requirements as a basis for recommendations is a complex issue owing to the inherent characteristics of fat metabolism and its interaction with energy metabolism. Furthermore, recommendations of desirable ranges of fat intake are influenced by scientific controversies and changing views on both the benefits and risks to human health associated with fat intake. Recommendations may also vary according to the dietary patterns and the prevalence of diet-related noncommunicable diseases in a determined population and in different physiological situations throughout the life-cycle. Most of the data available relating fat intake and health in the general population come from studies in affluent societies and deal with the prevention of chronic diseases, possibly caused by excessive fat intake. An adequate level of fat intake in physiological periods of high and/or special demands has also been of concern. During early development, adequate fat provision will not only influence fetal and infant nutrition but will also have consequences on health in later life.

In this chapter, fat requirements based on current scientific knowledge and the accepted recommendations regarding fat intake for the general adult population, for pregnant and lactating women, and for infants and young children will be specifically addressed.

Fat Requirements for the General Population

Dietary Fat and Energy

Dietary fat intake world-wide is highly variable, ranging from less than 10 to approximately 45% of total energy, being highest among affluent societies. Triacylglycerols present a high digestibility (approximately 95–99%) and the highest energy density among macronutrients (approximately $9\,kcal\,g^{-1}$ or $38\,kJ\,g^{-1}$). Since the metabolic fate of individual fatty acids depends on dietary energy and on energy balance, the intake and requirements for fat and essential fatty acids (EFA) are usually described as the percentage of energy in the diet (en%), rather than total intake (g).

Since the contribution of dietary protein to energy intake shows little variation, low-fat diets are associated with high carbohydrate intakes, and high fat diets with low carbohydrate intakes. Although it is generally assumed that a low fat intake helps to prevent the development of chronic diseases, this may not be necessarily true, since a consequent high intake of carbohydrates can have specific metabolic effects that, in the long term, may lead to chronic diseases, such as type 2 diabetes. Therefore, a compromise should be pursued, regarding an optimal ratio of fat to carbohydrate in the diet. It is currently accepted that dietary fat intake should be moderate: up to 30 en% for sedentary or overweight individuals or up to 35 en% for lean, active individuals, to avoid or minimize the occurrence of chronic diseases associated with high fat intakes. In contrast, fat intake should not be lower than 15 en%, since low fat intakes, common in developing countries especially in rural areas, could limit energy intake and also reduce the absorption of liposoluble vitamins and essential fatty acids.

It has been hypothesized that high levels of dietary fat could lead to obesity. Although it is generally accepted that obesity is the result of a chronic positive energy balance, the homeostatic mechanisms involved in the regulation of energy metabolism discriminate between energy substrates, at both the cellular and whole-body levels. The metabolism of protein and carbohydrates is strictly regulated in the acute and long terms, possibly because of their relatively small storage capacity. Fat, in contrast, has a high storage capacity and is subjected to less efficient homeostatic controls. Additionally, it has been well demonstrated that healthy humans do not synthesize fatty acids *de novo* in appreciable amounts, when eating moderately low and appropriate amounts of fat and carbohydrates, respectively. These evidences lead to the suggestion that high fat intakes *per se* could lead to net weight accretion and obesity. However, an elevated fat consumption alone would not lead to weight gain if not accompanied by a positive energy balance. An elevated consumption of fat is generally associated with high energy intakes, possibly due to a phenomenon known as passive overconsumption or simply due to the elevated energy density of fat. A restriction of dietary fat intake up to the level of 30 en% could help to reduce

energy intake and to improve health. However, fat restriction should be accompanied by total energy restriction and by an exercise plan if weight reduction is to be attained.

Saturated and Unsaturated Fats

The molecular structure of fatty acids will determine their metabolic fate and effects from the cellular level to the whole organism. Fatty acids vary in their chain length, number, position, and geometric configuration of double bonds, which are the main structural differences affecting fatty acid metabolism and, consequently, requirements. Saturated fatty acids are devoid of double bonds, whereas unsaturated fatty acids may have one (monounsaturated) or more (polyunsaturated) double bonds in their carbon chain.

Saturated fatty acids are high in animal fats and palm and coconut oils, and their intake is associated with some of the negative metabolic effects of fats, especially those related to the development of cardiovascular diseases (mainly atherosclerosis and coronary heart disease). The critical links between saturated fat consumption and cardiovascular diseases are total blood cholesterol and low-density lipoprotein (LDL)-cholesterol levels, and thrombosis. Thrombosis is the major determinant of morbidity and mortality from coronary heart disease, and may be caused by elevated saturated fat in the diet. Controlled trials demonstrated that the saturated fat content, as well as the type of saturated fatty acids in the diet, affect serum lipid and lipoprotein levels. The consumption of saturated fatty acids, compared with that of carbohydrates, leads to an increase in total and LDL-bound cholesterol, and it is very likely that high serum cholesterol levels may cause atherosclerosis and coronary heart disease. Therefore, saturated fat intake should be restricted to less than 10 en%. Saturated fatty acids are not equally atherogenic, lauric (12:0), myristic (14:0), and palmitic (16:0) acids seem to elevate LDL-cholesterol levels when compared with carbohydrates, whereas fatty acids with less than 12 carbon atoms (short and medium chain fatty acids) and stearic acid (18:0) seem to have only slight or no effects. These effects are justifiable by the different metabolic fates of these saturated fatty acids. Short-chain fatty acids seem to be oxidized more efficiently than long-chain fatty acids, and stearic acid is probably converted to oleic acid (18:1*n*-9) very rapidly. Although there are differences in the potency of lauric, myristic, and palmitic acids, they are considered the major hypercholesterolemic saturated fatty acids. In case of palmitic acid, its position in the triacylglycerol molecule is also of importance. When bound to the C-2 atom of the glycerol molecule (as in pork lard), it is more hypercholesterolemic than in the C-1 position (such as in palm oil or cocoa butter). A reduction in total saturated fat intake should lead to a decreased consumption of the three aforementioned fatty acids since they are usually the most abundant saturated fatty acids in the diet.

Contrary to saturated fat intake, unsaturated fat intake leads to a decrease in serum cholesterol and LDL-cholesterol levels when substituting for carbohydrates in the diet. The magnitude of this effect is modest when compared with the cholesterol-raising effect of saturated fat. When limiting the intake of saturated fat up to 10 en%, unsaturated fat consumption should increase. The number and position of double bonds in unsaturated fatty acids will influence their effects on serum lipids and lipoproteins levels. The main polyunsaturated fatty acids (PUFA) in the human diet are of the *n*-6 and *n*-3 series, and increasing their intake up to a certain limit should have beneficial effects on human health. However, PUFA are susceptible to peroxidation and generation of reactive oxygen species, which could increase the susceptibility of LDL to oxidation, leading to an increased risk of atherosclerosis, and are also involved in the etiology of certain types of cancer. Therefore, PUFA intake should not be higher than 10 en%.

Although the consumption of monounsaturated fatty acids slightly decreases serum cholesterol and LDL-cholesterol levels, they are generally considered metabolically 'neutral' when compared with polyunsaturated and saturated fatty acids. Therefore, monounsaturated fat intake should account for the remaining of the energy intake contributed by fat, which should correspond to more than 10 en%. However, when the intake of *cis*-monounsaturated fatty acids is elevated (18:1*n*-9/18:2*n*-6 ratio of 10:1), essential fatty acid deficiency may occur.

Essential Fatty Acids

Since animals, in contrast to vegetables, are incapable of synthesizing fatty acids with double bonds at positions *n*-6 or *n*-3, owing to a lack of Δ-12 and Δ-15 desaturase activities, these compounds should be present in the human diet. Linoleic (18:2*n*-6) and α-linolenic (18:3*n*-3), considered essential for humans, are the main fatty acids of the *n*-6 and *n*-3 series in the modern diet. Owing to the high intake of vegetable oils, linoleic acid is present in the diet in higher amounts than α-linolenic acid. PUFA are incorporated into cellular membranes where they present structural functions. In addition, 18:2*n*-6 and 18:3*n*-3 are substrates for the synthesis of long-chain polyunsaturated fatty acids with 20 or 22 carbon atoms (LCPUFA): arachidonic acid (20:4*n*-6;

ARA), dihomo-γ-linolenic acid (20:3*n*-6), eicosapentaenoic acid (20:5*n*-3; EPA), and docosahexaenoic acid (22:6*n*-3; DHA). DHA has an important structural role in the brain and retina, and the LCPUFA with 20 carbon atoms are precursors for the synthesis of eicosanoids, a group of important signaling molecules.

When EFA are not present in the diet in appropriate amounts, deficiency will take place, and various tissues and organs will be affected, owing to EFA ubiquitous structural and regulatory functions. Although its validity has already been questioned, the tissue ratio of 20:3*n*-9 (Mead's acid) to 20:4*n*-6 (triene/tetraene ratio) is still considered a sensitive biochemical marker of essential fatty acid deficiency and should be lower than 0.4 for the human. Mead's acid is the product of elongation and desaturation of oleic acid, which is an important substrate for elongases and desaturases only in the case of deficiency of *n*-6 and *n*-3 fatty acids. Diets with 0.1–0.5 en% of linoleic acid are sufficient to normalize an abnormally high triene/tetraene ratio, but for optimal function of various tissues, much higher intakes of 18:2*n*-6 are necessary. The adequate level of linoleic acid in the diet depends on the criteria used for its determination. It has been recommended that linoleic acid intake should be 1–2 en% to avoid clinical and biochemical signs of EFA deficiency, whereas its intake should be 3–5 en% for the prevention of chronic diseases.

α-Linolenic acid is much less efficient in the treatment of EFA deficiency than linoleic acid, raising the possibility that the former is a conditionally essential nutrient. However, its essentiality for the development and maintenance of normal functions in the retina and brain has now been confirmed, although the adequate level of intake has not yet been strictly determined. It has been estimated that for adult men, the consumption of 2 g day^{-1} of 18:3*n*-3 should provide 75–85% of the 350–400 mg day^{-1} requirement of long-chain *n*-3 fatty acids (20:5*n*-3 and 22:6*n*-3). The intake of approximately 1 en% or 3 g day^{-1} of 18:3*n*-3 and 800 mg of 20:5*n*-3 plus 22:6*n*-3 should provide an adequate supply of *n*-3 fatty acids for healthy adults.

Fatty acids of the *n*-3 and *n*-6 series will compete for the same enzymatic systems: acyl-transferases for incorporation into membranes; elongases and desaturases for LCPUFA synthesis; and lipoxygenase, cycloxygenase and other enzymes for eicosanoid synthesis. Therefore, the balance between tissue levels of *n*-3 and *n*-6 fatty acids will depend on the balance between dietary EFA. Although the optimal ratio of *n*-6/*n*-3 in the diet has been set between 4:1 and 10:1, scientific evidence is lacking to support ratios equal or higher than 5:1. Furthermore, an adequate tissue ratio of *n*-6 to *n*-3 is important to prevent coronary heart disease and stroke, since *n*-6 fatty acids are precursors of eicosanoids that stimulate platelet aggregation and vasoconstriction, in contrast to *n*-3 fatty acids, which are precursors of eicosanoids that stimulate vasodilatation and are much less aggregatory.

Isomeric Fatty Acids

The double bonds in fatty acids may present two geometrical configurations: *cis* (the most common) and *trans*. *Trans* fatty acids are formed during partial hydrogenation of vegetable fats and oils, which are widely used in the food industry for the production of margarine, shortenings, frying fats, among other products. Isomeric fatty acids formed through biohydrogenation in the rumen may be naturally present in meat, milk, and dairy products, which are enriched in *trans*-11-octadecenoic acid (*trans*-11–18:1). The main sources of total *trans* fatty acids in the Western diet are the hydrogenated vegetable fats, which are rich in octadecenoates with double bonds between Δ9 and Δ12.

The consumption of monoenes with a *trans* double bond has long been associated with elevated total and LDL-bound cholesterol in serum, and possibly with reduced levels of high-density lipoprotein cholesterol. In controlled trials in which the *trans* fatty acid intake was increased, an increment in total and LDL-bound cholesterol in serum occurred. It has also been hypothesized that *trans* fatty acids are more potent in eliciting such effects than saturated fatty acids. However, this subject is still a matter of debate, because the dietary levels that would lead to altered serum lipid and lipoprotein levels are much higher than that generally consumed. It is a consensus that *trans* fatty acids intake should be kept low (up to approximately 2 en%), but apparently, there is no need for further limiting when the intake of total fat and of essential and saturated fatty acids is appropriate. However, it should be emphasized that consumers of higher amounts of *trans* fatty acids may present a higher risk of morbidity from cardiovascular diseases.

Other geometrical isomers of unsaturated fatty acids are also present in foodstuffs, the conjugated fatty acids, which are polyunsaturated fatty acids with conjugated double bonds, i.e., double bonds separated by two carbon atoms instead of three. The main dietary sources of conjugated fatty acids are meat and dairy products from ruminants, and the most common are the octadecadienoic or conjugated linoleic acids (CLA). Conjugated linoleic acids have important biological effects that could be beneficial,

even when present in the diet at very low amounts (≤ 1 en%). Effects of CLA range from a potent anticarcinogenic effect to changes in body composition with reduction in the adipose tissue size. Advisable and upper tolerable levels of intake for CLA have not yet been defined, because of the scarce scientific data concerning CLA consumption and its effects in human subjects.

Fat Requirements During Pregnancy and Lactation

Pregnant Women

In the periconceptional period, maternal nutrition is particularly important owing to its influence on the physiological process of fat accretion during pregnancy. Maternal fat stores are critical for placental development and embryonic cell division and differentiation in the first trimester of pregnancy, for the intense fetal growth during the third trimester, and for the production of milk in early lactation. The building up of adequate fat stores requires an increase in energy and fat intakes during pregnancy. The estimated additional energy for pregnant women according to the WHO Report on Energy and Protein Requirements (1985) is 300 kcal day^{-1} for the whole pregnancy, with 30 en% as fat. Dietary energy from fat should not be lower than 20% for women of reproductive age, who might otherwise be exposed to an insufficient supply of EFA and fat-soluble vitamins.

Owing to the demands of EFA for fetal, placental, mammary gland, and uterine growth, and for the increase in maternal blood volume, maternal EFA intake during pregnancy should be increased from 3 to 4.5% of the total energy, and the ratio *n*-6/*n*-3 should be maintained at approximately 5:1.

High consumption of fish and dietary supplementation with *n*-3 LCPUFA are associated with a modest increase in the length of gestation and with reduced incidence of low birth weight and prematurity, suggesting that the dietary consumption of preformed DHA could be advantageous.

Placental transfer of ARA and DHA during pregnancy occurs preferentially in relation to other fatty acids. This can be explained by the presence of a plasma membrane fatty acid-binding protein that presents the highest affinities and binding capacities for both fatty acids, especially DHA. It seems that elongation and desaturation of EFA in human placenta are limited, and the fetus depends on the transfer of ARA and DHA from maternal circulation. There is a substantial increase of ARA and DHA in fetal circulation, and fetal organs such as the liver and brain incorporate large amounts of *n*-3 and *n*-6 LCPUFA into membrane phospholipids by the end of the third trimester. Therefore, the preterm infant will be endowed with a reduced provision of ARA and DHA *in utero*, and will be at a disadvantage concerning fat stores and membrane LCPUFA when compared with term infants.

Prematurity and low birth weight are associated with a high risk of neurodevelopmental disorders and disabilities. An inadequate supply of ARA and DHA during the period of rapid vascular and brain growth could lead to membrane fragility, leakage and breakdown, resulting in ARA peroxidation, vasoconstriction, inflammation, and ischemia, which could explain those disabilities. Fetal exposure to *trans* fatty acids, which are transferred across the placenta, has been related to low birth weight in premature infants.

Lactating Women

The total fat content in human milk is highly variable and is affected by several factors but is secreted in adequate amounts to support infant growth and development in a wide range of maternal nutritional status, unless overt undernutrition is present. The fatty acid composition of the triacylglycerols in milk, however, is dependent on maternal fatty acid intake, body stores, and endogenous synthesis. Gestational age, stage of lactation, nutritional status, and genetic background may also influence the fatty acid composition of human milk.

Saturated and monounsaturated fatty acids are the major fatty acids in human milk, comprising approximately 83% (wt/wt) of the total fatty acids in mature milk. In well-nourished mothers, approximately 11% (wt/wt) of fatty acids in human milk is linoleic acid and 1% α-linolenic, while total *n*-6 and *n*-3 LCPUFA account for 1.2 and 0.6%. Total PUFA accounts for 6% of total energy in milk. PUFA transfer from maternal circulation to milk and the synthesis of LCPUFA from the EFA precursors are a matter of concern, because of their importance for infant growth, neurodevelopment, and visual function. Studies with lactating women using stable isotope methodologies demonstrated that the major part of PUFA in human milk is derived from maternal body stores, which reflect long-term intake, and not from direct dietary transfer. In women on omnivorous diets, about 30 and 12%, respectively, of linoleic acid and ARA secreted in milk were derived from the diet. Endogenous synthesis from dietary linoleic acid accounted for about 3–25% and 1–3% of dihomo-γ-linolenic acid and ARA in milk, respectively.

Assuming an appropriate weight gain during pregnancy, an additional average energy allowance

of 500 kcal day^{-1} is recommended throughout lactation, according to WHO (1985). An additional maternal intake of 1–2% of energy in the form of EFA (3–4 g day^{-1}) is recommended during the first 3 months of lactation, and thereafter a further increase of up to 4% of energy (about 5 g day^{-1}) is required, owing to depletion of maternal fat stores. An inverse correlation between *trans* fatty acids and ARA and DHA in plasma lipids has been reported in infants, suggesting an impairment in LCPUFA synthesis and metabolism. Therefore, maternal intake of *trans* fatty acids is of concern when their intake is excessively high or when EFA intake is low during pregnancy and lactation.

Fat Requirements in Infancy and Early Childhood

Both the amount and quality of dietary fat can affect child growth and development. Breast milk is the best available source of dietary fat and fatty acids for infants, providing between 50 and 60 en% as fat and LCPUFA in adequate amounts for brain development. The human infant should be preferentially breast-fed, but when infant formulas are used, the fatty acid composition should ideally correspond to that of breast milk. Of special concern is the adequate provision of LCPUFA to preterm infants who present an insufficient intrauterine supply of these fatty acids.

LCPUFA are required for adequate brain growth and development. Their requirements are especially high during the last trimester of pregnancy and in the first months of life, when the brain is rapidly developing, and there is a significant deposition of DHA in the central nervous system. The brain gray matter and membranes of the retina present low deposition of EFA but contain high amounts of LCPUFA, which can be incorporated from plasma.

Synthesis of LCPUFA was demonstrated in term infants from the first week of life and even in very-low-birth-weight infants through stable isotope methodologies. However, the contribution of endogenous synthesis to the total plasma LCPUFA pool is small. The rate of ARA synthesis in term infants is low, and it accounts for only about 6% of total ARA renewal in plasma. Therefore, the term infant and especially the preterm infant possibly do not synthesize enough LCPUFA to attain their requirements for brain development during the first few weeks after birth.

Infants fed human milk, compared with infants fed DHA-free formulas, present higher plasma concentrations of LCPUFA and also higher levels of DHA in red blood cell phospholipids and in brain cortex. Clinical trials in premature infants showed beneficial effects of infant formulas enriched in ARA and DHA, such as higher LCPUFA concentrations in plasma, improvement of visual acuity and cognitive abilities, without compromising growth and safety. Several studies with healthy full-term infants indicated that infants fed breast milk or DHA-supplemented formulas presented better visual resolution acuity at 2 months and possibly at 4 months of age than infants fed DHA-free formulas.

A WHO/FAO Expert Committee (1994) based its recommendations for EFA and LCPUFA provision to infants on the composition of breast milk from well-nourished mothers, which provides both parent EFA and their LCPUFA products in apparently adequate amounts to fulfill infants' requirements. The WHO/FAO recommendations described below are similar to those published in position papers of other national and international expert committees.

Total EFA (*n*-6 plus *n*-3) should comprise 5–6 en% (or 0.6–0.8 g kg^{-1} day^{-1}) and LCPUFA 0.8 en%. The total *n*-6/*n*-3 ratio should be in the range of 5:1 to 15:1. To insure tissue proliferation, membrane integrity, and eicosanoid formation, linoleic acid should be supplied as a minimum of 1–4.5 en% or 0.5–0.6 g kg^{-1} day^{-1}, but its intake should be limited to amounts lower than 10–12 en%, since its excess impairs production of the *n*-3 LCPUFA. Arachidonic acid could be considered conditionally essential during early development, owing to its specific neural and vascular functions, the role of its derived eicosanoids in cell regulation, and the interactions between the *n*-3 and *n*-6 fatty acids series. In this regard, 40 mg of ARA should be provided for the term infant, whereas for the premature infant, the amount would be 60–100 mg kg^{-1} day^{-1}, since, in these infants, a low level of ARA is associated with low pre- and postnatal growth.

Total *n*-3 fatty acids should amount to 70–150 mg kg^{-1} day^{-1}, and the intake of DHA should be 20 and 40 mg kg^{-1} day^{-1} for term and preterm infants, respectively. The DHA/ARA ratio should be 1:1 to 1:2, since excess DHA impairs ARA metabolism. Intake of α-linolenic acid should be approximately 1 en% in the absence of intake of EPA and DHA, or 0.5 en% when there is a supply of these LCPUFA, although the bioequivalency of *n*-3 EFA and its LCPUFA products has not yet been determined in infants.

Regarding formulas for preterm babies, a mixture of oleic acid and saturated fatty acids, with a predominance of the former, could provide the required energy from fat, since this mixture would possibly meet the desirable criteria of digestibility and lack of interference with essential fatty acid metabolism.

According to the FAO/WHO (1994) recommendations, during weaning and at least until 2 years of age,

a child's diet should contain 30–40 en% as fat and provide similar levels of EFA to those found in breast milk. Special attention should be given to the complementary foods used in the weaning period in many developing countries, where cereals or tuber-based diets with a low energy density are used, resulting in a drastic drop in energy intake. Low-fat diets with less than 30 en% as fat have been associated with adverse effects on child growth, but this might also have been associated with a low intake of total energy and other nutrients.

The increase of the prevalence and severity of obesity and type 2 diabetes mellitus in the pediatric population is of particular concern, mainly in affluent societies, which reinforces the need for appropriate nutrition, including the quantity and quality of dietary fat, not only in childhood but also in early development.

See also: **Atherosclerosis**; **Coronary Heart Disease**: Etiology and Risk Factor; **Essential Fatty Acids**; **Fats**: Digestion, Absorption, and Transport; **Fatty Acids**: Dietary Importance; *Trans*-fatty Acids: Health Effects; **Infants**: Nutritional Requirements; **Lactation**: Human Milk: Composition and Nutritional Value

Further Reading

Al MDM, van Howelingen AC and Hornstra G (2000) Long-chain polyunsaturated fatty acids, pregnancy and pregnancy outcome. *American Journal of Clinical Nutrition* 71 (supplement): 285S–291S.

Chapkin RS (2000) Reappraisal of the essential fatty acids. In: Chow CK (ed.) *Fatty Acids in Foods and their Health Implications*, 2nd edn, pp. 557–568. New York: Marcel Dekker.

FAO/WHO (1994) *Fats and Oils in Human Nutrition*. Report of a Joint Expert Consultation (19–26 October 1993). Rome: FAO.

Gibson RA and Makrides M (2000) *n*-3 Polyunsaturated fatty acid requirements of term infants. *American Journal of Clinical Nutrition* 71 (supplement): 251S–255S.

Grundy SM (1999) The optimal ratio of fat-to-carbohydrate in the diet. *Annual Reviews of Nutrition* 19: 325–341.

Jones PJH and Kubow S (1999) Lipids, sterols, and their metabolites. In: Shils M, Olson JÁ, Shike M and Ross AC (eds) *Modern Nutrition in Health and Disease*, 9th edn, pp. 67–94. Baltimore, MD: Williams & Wilkins.

Koletzko B (1999) Response to and range of acceptable fat intakes in infants and children. *European Journal of Clinical Nutrition* 53: S78–S83.

Pariza MW, Park Y and Cook ME (2000) Mechanisms of action of conjugated linoleic acid: evidence and speculation. *Proceedings of the Society for Experimental Biology and Medicine* 223: 8–13.

SanGiovanni JP, Berkey CS, Dwyer JT and Colditz GA (2000) Dietary essential fatty acids, long-chain polyunsaturated fatty acids, and visual resolution acuity in healthy fullterm infants: a systematic review. *Early Human Development* 57: 165–188.

Sauerwald TU, Demmelmair H, Fidler N and Koletzko B (2000) Polyunsaturated fatty acid supply with human milk. Physiological aspects and *in vivo* studies of metabolism. *Advances in Experimental Medicine and Biology* 478: 261–270.

Uauy R and Hoffman DR (2000) Essential fat requirements of preterm infants. *American Journal of Clinical Nutrition* 71 (supplement): 245S–250S.

Vellas B, Chumlea WC and Lanzmann-Petithory D (eds) (2001) Diet and cardiovascular diseases. *Journal of Nutrition Health & Aging* 5: 129–204.

Fat Replacers

B G Swanson, Washington State University, Pullman, WA, USA

Fat

Reducing dietary fat is a major, if not the primary, dietary goal for many consumers. With encouragement from health groups and government agencies, the public is increasingly choosing foods and beverages naturally low in fat, as well as new, revolutionary low-fat and nonfat foods and beverages. The development and use of a wide variety of food ingredients, known as fat replacers, are making many low fat or nonfat foods possible.

Fat, protein, and carbohydrates are essential components of the human diet. Fat is the most concentrated source of energy, contributing $38\,kJ\,g^{-1}$; protein and carbohydrates contribute approximately $17\,kJ\,g^{-1}$ to the diet. Some dietary fat is essential to enable the body to function properly. Fat is responsible for transporting fat-soluble vitamins A, D, E, and K and is also a source of essential fatty acids necessary for good health. (*See* **Essential Fatty Acids**.)

Most consumers enjoy the taste, texture, and aroma fat gives to foods. Fat consumption, however, is often considered too great by public health organizations and nutritionists since a high intake of dietary fat is associated with an increased risk for obesity, some types of cancer, and possibly gall bladder disease. Epidemiological, clinical, and animal studies provide strong and consistent evidence for a relationship between saturated fat intake, high blood cholesterol, and increased risk for coronary heart disease.

There are two primary types of fat replacers being investigated: 'energy-free fat substitutes' and 'energy-reduced fat mimetics.' Energy-free fat substitutes are neither digested nor absorbed and therefore contribute no fat, calories, or cholesterol to the diet. Fat substitutes are formulated and synthesized to exhibit physical and cooking properties similar to fats or oils, and are expected to replace some or all of the fat in cooking, as well as in foods.

Fat Substitutes

A group of compounds, called sucrose fatty acid polyesters with the trade names 'olestra 7' or 'olean 7,' produced by Procter and Gamble Co., were approved for use in savory snacks by the United States Food and Drug Administration (FDA) in 1996, amidst great controversy regarding the potential health benefits and risks. Recent consumer and industry surveys suggest that the benefits of fat substitutes far outweight the purported risks, and are being favorably received by many consumers. Sucrose polyester fatty acid mono- and di-esters are FDA-approved emulsifiers at concentrations of less than 1% in food emulsions. (*See* **Emulsifiers**: Organic Emulsifiers.)

Carbohydrate (sucrose) fatty acid polyesters are synthesized from fatty acid methyl esters and sugars or sugar derivatives. Fatty acids available from many triacylglycerol sources and sugars, such as sucrose, glucose, raffinose, stachyose, and trehalose, as well as sugar alcohols are combined, with the sugar as the center and the esterified fatty acids extending away from the sugar like spokes in a wheel. Sugars heavily saturated (esterified) with fatty acids, such as hexa-, hepta-, or octa-fatty acid esters of sucrose, are not hydrolyzed by sterically and hydrophobically hindered enzymes in human digestive tracts, are not absorbed, and are poorly metabolized by colonic microorganisms. The steric and hydrophobic hindrance of lipolytic enzymes is in response to the many fatty acids esterified to sucrose, and the accompanying hydrophobic cloud surrounding the ester bonds on the sucrose molecule. (*See* **Carbohydrates**: Metabolism of Sugars; **Triglycerides**: Structures and Properties.)

Prior to petitioning the FDA for approval of olestra 7, Procter and Gamble evaluated olestra 7 in more than 100 animal and 25 clinical studies involving thousands of individuals with no observations of mutagenic, teratogenic, or developmental responses. Carbohydrate fatty acid polyesters exhibit organoleptic and physical properties similar to conventional vegetable oils or animal fats (triacylglycerols), including necessary heat stability for frying.

Structured lipids or medium-chain triacyglycerols are being prescribed for treatment of patients suffering from malabsorption symptoms, pancreatic insufficiency, bile duct obstruction, steatorrhea, and lipoproteinemia, and to bypass surgery patients and infants requiring feeding formulas. Structured lipids are generally triacylglycerols made up of two short to medium-chain (< 12 carbon) fatty acids combined with a single long chain (> 12 carbon) fatty acid. Structured lipids provide the physical and functional properties of fat while contributing approximately one-half the calories or normal edible oils. 'Caprenin 7' and 'Salatrim 7' are structured lipids approved for use as confectionary fats as well as for use in bakery and filled dairy products in the USA.

Malonate esters, synthesized from malonic acid, hexadecane, and fatty acids, were developed by Frito-Lay, Inc. as a calorie-free substitute for high-temperature frying applications. Feeding studies with rats indicate that only trace concentrations are absorbed, and the liver is the primary organ of distribution and elimination.

Esterified propoxylated glycerol is similar to a natural triglyceride, except that oxypropylene is incorporated between the glycerol and the fatty acids. Short-term animal studies suggest that esterified propoxylated glycerol is safe and resistant to hydrolysis.

Trialkoxytricarballate is also similar to natural triglycerides with tricarballylic acid replacing the glycerol and saturated or unsaturated alcohols replacing the fatty acids. Animal studies indicate that trialkoxytricarballates are not digested. Trialkoxytricarballates may be used to replace vegetable oil in cooking applications as well as in food emulsions.

Polyorganosiloxane compounds are also noncaloric, nonabsorbable liquid oils derived from silica that are chemically inert and not toxic. The polysiloxanes are stable, maintain viscosity over a wide temperature range, are resistant to oxidation, hydrolysis, and degradation, and are similar in solubility characteristics to nonpolar lipids.

Jojoba oil and its derivatives have many applications as constituents of cosmetics and pharmaceuticals. Jojoba margarine and mayonnaise have been shown to exhibit functional properties different than controls, but perform acceptably in most food applications. Unfortunately, jojoba oil is sensitive to hydrolysis by pancreatic lipases. The *in vivo* digestibility of jojoba oil may be as much as 20%.

Other fat-based fat substitutes or replacers include Dur-Lo, a vegetable oil mono- and diglyceride emulsifier manufactured by Durkee Industrial Foods Corporation, which can replace all or part of the shortening content in cake mixes, cookies, icings,

and numerous vegetable oil-based dairy products when emulsified with water. Mono- and di-glycerides are considered generally recognized as safe (GRAS) by the FDA. Emulsion systems using soybean oil or milk fat can reduce fat and calories significantly by replacing fat on a one-to-one basis, while utilizing less fat and contributing fewer calories.

Fat Mimetics

Fat mimetics are compounds that help replace the mouthfeel of fats by contributing a greater viscosity to the liquid phase in the mouth, but cannot substitute for fats on a gram-for-gram basis. Fat mimetics do not have the nonpolar chemical properties of fats and cannot be used for frying applications because of water concentration and heat lability.

Fat mimetics are categorized as 'protein-based,' 'carbohydrate-' or 'starch-based', or 'cellulose-based.' Many of the fat mimetics have been marketed and are available in a wide variety of formulations as salad dressings, frozen desserts, and baked goods.

'Protein-based' fat mimetics marketed under product names Simplesse 7, Trailblazer 7, and Finesse 7 are structurally modified natural proteins derived from milk or egg white. Simplesse is the initial protein-based fat replacer prepared with a patented heating and homogenization process of microparticulation in which the proteins are aggregated into tiny, round particles that create the creaminess and mouthfeel of fat. Simplesse cannot be used in heated foods because the protein will congeal and lose the creamy mouthfeel. Traiblazer and Finesse are similar modified protein texturizers prepared with unique and innovative combinations of heat, acidulation, and homogenization of natural proteins, selected carbohydrates, and water.

Carbohydrate- or starch-based fat mimetics are formulated from GRAS oligo- or poly-saccharides chemically extracted and prepared from hydrolyzed, often modified, corn, potato, or tapioca starch. Many calorie-reducing carbohydrate starch-based fat mimetics are available for use as ingredients in low-calorie or 'lite' foods.

Table 1 Selected applications and functions of fat replacing

Specific application	*Fat replacer*	*General functions*[a]
Baked goods	Lipid-based	Emulsify, provide cohesiveness, tenderize, carry flavor, replace shortening, prevent staling, prevent starch retrogadation, condition dough
	Carbohydrate-based	Retain moisture, retard staling
	Protein-based	Texturize
Frying	Lipid-based	Texturize, provide flavor and crispiness, conduct heat
Salad dressing	Lipid-based	Emulsify, provide mouth feel, hold flavorants
	Carbohydrate-based	Increase viscosity, provide mouth feel, texturize
	Protein-based	Texturize, provide mouth feel
Frozen desserts	Lipid-based	Emulsify, texturize
	Carbohydrate-based	Increase viscosity, texturize, thicken
	Protein-based	Texturize, stabilize
Margarine, shortening, spreads, butter	Lipid-based	Provide spreadability, emulsify, provide flavor and plasticity
	Carbohydrate-based	Provide mouth feel
	Protein-based	Texturize
Confectionery	Lipid-based	Emulsify, texturize
	Carbohydrate-based	Provide mouth feel, texturize
	Protein-based	Provide mouth feel, texturize
Processed meat products	Lipid-based	Emulsify, texturize, provide mouthfeel
	Carbohydrate-based	Increase water-holding capacity, texturize, provide mouth feel
	Protein-based	Texturize, provide mouth feel, water-holding
Dairy products	Lipid-based	Provide flavor, body, mouth feel, and texture; stabilize, increase overrun
	Carbohydrate-based	Increase viscosity, thicken, aid gelling, stabilize
	Protein-based	Stabilize, emulsify
Soups, sauces, gravies	Lipid-based	Provide mouth feel and lubricity
	Carbohydrate-based	Thicken, provide mouthfeel, texturize
	Protein-based	Texturize
Snack products	Lipid-based	Emulsify, provide flavor
	Carbohydrate-based	Texturize, aid formulation
	Protein-based	Texturize

[a]Functions are in addition to fat replacement.

Carbohydrate-based fat mimetics are generally digested and absorbed to some extent, contributing 19 kJ or less per gram of carbohydrate. Carbohydrate-based mimetics taste bland, are soluble in water, somewhat pH- and heat-stable, and, most importantly, contribute a smooth, creamy texture and mouthfeel of fat, as well as the spreadability and appearance when replacing all or part of the fat in natural or emulsified foods. General categories of foods, potential fat replacers, and the functions of the fat replacer in addition to fat replacement are presented in Table 1.

Cellulose-based fat mimetics contribute no calories to the diet in contrast to the carbohydrate- or protein-based fat mimetics. One cellulose-based fat mimetic is Avicel, a blend of microcrystalline cellulose and carboxymethyl cellulose. Cellulose-based fat mimetics contribute a creamy mouthful and little taste to foods.

With increasing evidence that fat consumption and obesity are detrimental risk factors for chronic disease, and that reducing the intake of dietary fats and oils will contribute to improved public health, the demand of health conscious consumers for reduced fat foods will continue. Despite medical opinion and media attention to the benefits of reducing fat consumption, many consumers find it difficult to make long-term changes in their eating habits when taste and enjoyment are compromised by reduction in dietary fat. (*See* **Obesity**: Treatment.)

Unique achievements in food science research and technology are resulting in safe ingredients to replace fats in foods. When limiting fats, calories, and sugars in the diet, consumers need to satisfy their basic essential nutritional needs. No fat replacer is a panacea, and additional reduced fat and reduced calorie foods will neither compensate for poor dietary habits nor replace the need for moderation and good overall nutrition. So, the consensus for the need to reduce dietary fat is offset by the difficulty in identifying alternative foods that encourage consumers to make long-term dietary changes. Low-fat and low-calorie foods and beverages formulated from ingredients that replace hidden and obvious fats and oils can help to maintain a healthy, nutritious diet.

See also: **Bioavailability of Nutrients**; **Carbohydrates**: Digestion, Absorption, and Metabolism; Metabolism of Sugars; **Cholesterol**: Role of Cholesterol in Heart Disease; **Emulsifiers**: Organic Emulsifiers; **Essential Fatty Acids**; **Obesity**: Treatment; **Triglycerides**: Structures and Properties

Further Reading

Akoh CC and Swanson BG (eds) (1994) *Carbohydrate Polyesters as Fat Substitutes*. New York: Marcel Dekker.

Akoh CC (1998) Fat replacers. *Food Technology* 52(3): 47–53.

Best D (1992) Fat substitutes: Finding method in the madness. *Prepared Foods* 161(12): 21–31.

Calorie Control Council (1991) *Fat Replacers*. Atlanta, GA: Calorie Control Council.

Fat Substitutes Update (1990) *Food Technology* 46(3): 92–97.

Hamm DJ (1984) Preparation and evaluation of trialkoxytricarballylate, trialkoxycitrate, trialkoxyglycerylether, jojoba oil and sucrose polyester as low calorie replacements of edible fats and oils. *Journal of Food Science* 49: 419–428.

Haumann BF (1992) Fat replacers. *Inform* 3(12): 1270–1287.

Nabors LO (1992) Fat replacers: Options for controlling fat and calories in the diet. *Food Nutrition News* 64(1): 5–6.

Rudolph MJ, Greenward CG and Flesch RP (1991) Fat replacers: U.S. Markets and Technologies. *Decision Resources* February 12: 14-1–14-8.

Sacchetti M (1970) *Sull'aceto balsamico modenese*. Bologna: Edagricole.

Schlicker SA and Regan C (1990) Innovations in reduced-calorie foods: A review of fat and sugar replacement technologies. *Topics in Clinical Nutrition* 6(1): 50–60.

Stern JS and Hermann-Zaidins MG (1992) Fat replacements: A new strategy for dietary change. *Journal of the American Dietetic Association* 92: 91–93.

Swanson BG (1996) Low-calorie fats and synthetic fat substitutes. In: Roller S and Jones SA (eds) *Handbook of Fat Replacers*, pp. 265–274. London: CRC Press.

Classification

M H Gordon, The University of Reading, Reading, UK

Introduction

Together with proteins, carbohydrates and water, fats represent the major bulk constituents in food and other biological materials. Fats comprise several classes of chemical compounds, termed lipids, that have common solubility properties, i.e., soluble in low-polarity organic solvents such as chloroform, diethyl ether and light petroleum, but sparingly soluble in water. Although these characteristics are valid for many lipids, there is a range of polarity amongst various members of this class and many lipids have a low solubility in nonpolar solvents such as light petroleum. Chloroform is often the best solvent for a wide range of lipids. The term 'fats' may be applied to all lipid-soluble extracts from biological materials,

or a distinction may be made between oils, which are liquid, and fats which are solid. Chemically, fats include many compounds which are fatty acids or esters of fatty acids. However, other components are classified as lipids on the basis of their solubility properties, even though they are not esters of fatty acids. (*See* **Fatty Acids**: Properties.)

Lipids can be classified into three groups: simple lipids, compound lipids, and derived lipids. Simple lipids are esters of fatty acids which produce two classes of compounds on hydrolysis. Compound lipids are esters of fatty acids which produce three or more classes of compounds on hydrolysis. Derived lipids cannot be hydrolyzed to fatty acids.

Simple Lipids

The main simple lipids are triglycerides (also known as triacylglycerols), steryl esters, and wax esters. Hydrolysis of these lipids yields glycerol and fatty acids, sterols and fatty acids, and fatty alcohols plus fatty acids, respectively. The most important of these simple lipids for food scientists are the triglycerides. They are the major components of edible oils and fats, often representing more than 95% of refined oils. Triglycerides are esters of the trihydric alcohol glycerol with three fatty acids (Figure 1). Many of the properties of triglycerides are dependent on the component fatty acids. Thus, the melting point of the triglyceride reflects the melting point of the component fatty acids, with three high-melting-point fatty acids yielding a high-melting triglyceride. Unsaturation in the fatty acids makes the triglyceride susceptible to autoxidation, just as the fatty acid itself would be. (*See* **Triglycerides**: Structures and Properties.)

All triglycerides are susceptible to hydrolysis in the presence of a catalyst. Acids, bases, or enzymes belonging to the hydrolase class, especially lipases, may act as the catalyst for the hydrolysis of triglycerides.

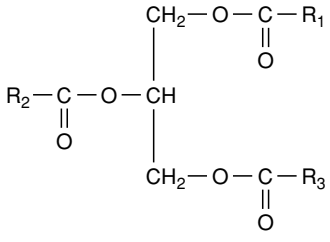

Figure 1 Triglyceride structure. Reproduced from Lipids, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Steryl esters always occur together with sterols in plant, animal, or microbiological tissues. Wax esters may accumulate in considerable amounts in some biological tissues and this class comprises the main constituent of beeswax and jojoba oil.

Compound Lipids

There are many members of this group. Probably the most important compound lipids for food science are the phospholipids.

Phospholipids

Phospholipids are esters of glycerol, fatty acids, phosphoric acid, and other alcohols. The most common phospholipids are phosphatidylcholine, phosphatidylethanolamine, phosphatidylinositol, and phosphatidylserine. These phospholipids share the common features of fatty acids esterified to the 1 and 2 positions of the glycerol backbone with the phosphate group esterified to the 3 position (Figure 2). (*See* **Phospholipids**: Properties and Occurrence.)

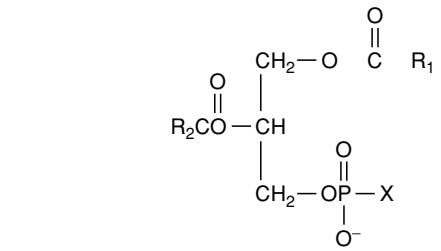

Phosphatidylcholine (Lecithin)	X = —$OCH_2CH_2\overset{+}{N}(CH_3)_3$
Phosphatidylethanolamine (Cephalin)	X = —$OCH_2CH_2\overset{+}{N}H_3$
Phosphatidylserine	X = —$OCH_2CH(\overset{+}{N}H_3)(CO_2^-)$
Phosphatidylinositol	X = 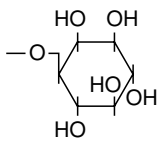
Phosphatidylglycerol	X = —$OCH_2CHOHCH_2OH$

Figure 2 Phospholipid structure. Reproduced from Lipids, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Whereas triglycerides occur in lipid droplets in storage cells of plants, animals, and microorganisms, phospholipids are the major components of cell membranes. They are always present in small amounts in edible oils extracted from biological tissues. However, the bulk of the phospholipids are removed by degumming during the refining of edible oils. Phospholipids differ from triglycerides in being surface-active and they are used in food products as emulsifiers because they migrate to the interface between oil and water, and reduce the inter-facial tension, thereby stabilizing an emulsion. Commercial phospholipid preparations are called lecithin. (*See* **Emulsifiers**: Organic Emulsifiers.)

Glycolipids

Glycolipids yield fatty acids, glycerol, and carbohydrates on hydrolysis. Monogalactosyldiglycerides (Figure 3) are glycolipids which are commonly found in plant leaves and algae. They contain a high proportion of polyunsaturated fatty acids and appear to play a role in photosynthesis. They are also found in the central nervous systems of some animals. Digalactosyldiglycerides often accompany monogalactosyldiglycerides in the chloroplasts of higher plants and algae, but generally at lower concentrations. Mono- and digalactosyldiglycerides are surface-active compounds which are known to play a key role in the baking quality of wheat flour. They assist the incorporation of air into the dough, thereby increasing the loaf volume, and they also have antistaling properties. (*See* **Bread**: Chemistry of Baking.)

Sulfoquinovosyldiglyceride (Figure 4) is one of the most polar plant lipids. It occurs in the leaves of plants, where it appears to play a role in photosynthesis. It is a significant dietary lipid in green leafy vegetables such as spinach.

Cerebrosides and cerebroside sulfates are further examples of glycolipids. These are discussed further under sphingolipids.

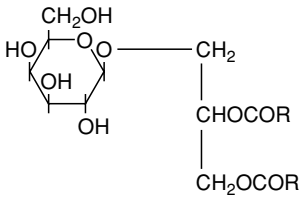

Figure 3 Monogalactosyldiglyceride structure. Reproduced from Lipids, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

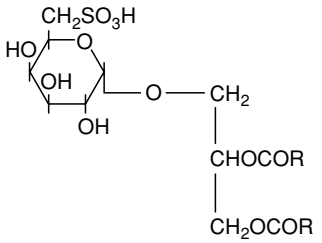

Figure 4 Sulfoquinovosyldiglyceride structure. Reproduced from Lipids, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Sphingolipids

Sphingolipids are important membrane components in both plant and animal cells. They are present in especially large amounts in brain and nerve tissue. Mammalian enzymes include those which can catalyze the hydrolysis of sphingolipids. If one of these enzymes is missing, disorders known as lipidosis may occur as the lipid accumulates in the tissues.

Sphingolipids (Figure 5) are derivatives of the amino alcohol sphingosine, which is found in animal tissues, or phytosphingosine, which is found in plant tissues. Lipids derived from other related bases are included in this group. Sphingosine is one of more than 60 long-chain amino alcohols found in animals, plants, and microorganisms. The bases commonly contain between 12 and 22 carbon atoms in the chain. A ceramide, which is an amide formed from a fatty acid and sphingosine, is the characteristic parent structure of all sphingolipids.

The fatty acid groups in ceramides are long-chain, with up to 26 carbon atoms, and are commonly saturated, monounsaturated, or hydroxy fatty acids. Although free ceramides have been found in small amounts in plant and animal tissues, they are generally present as a component of more complex sphingolipids.

Sphingomyelins are the most abundant sphingolipids in the tissues of higher animals. The 1-hydroxyl group of ceramide is esterified with phosphoric acid esterified with choline or with ethanolamine.

Cerebrosides are glycosphingolipids which contain a carbohydrate linked by a glycosidic linkage at the 1-hydroxyl position of the long-chain base. Cerebrosides are the major sphingolipids of plants. Wheat flour contains cerebrosides from at least four classes of amino alcohol. Plant cerebrosides contain only glucose as their component sugar, while animal cerebrosides contain galactose, glucose, or di-, tri- or tetrasaccharides. (*See* **Wheat**: The Crop; Grain Structure of Wheat and Wheat-based Products.)

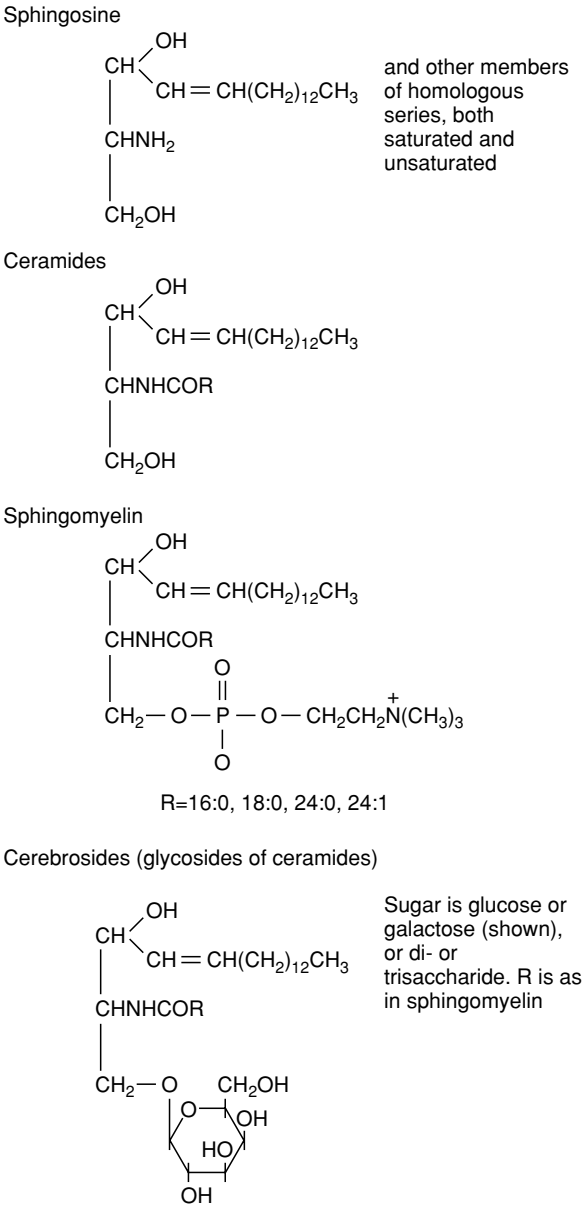

Figure 5 Sphingolipid structure. Reproduced from Lipids, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Lipoproteins

Lipoproteins may also be classified as compound lipids. These are compounds of proteins, fatty acids, and alcohols and may include other classes. Lipids and proteins are largely held together by hydrophobic interactions. Transport lipoproteins and membrane lipoproteins are important biomolecules in human physiology. (*See* **Lipoproteins**.)

Derived Lipids

This group of lipids includes a range of compounds which vary widely in their structures. The main derived lipids include: (1) fatty acids, (2) fat-soluble vitamins and provitamins, (3) alcohols – including sterols – (4) terpenoids; and (5) ethers.

Fatty Acids

Although simple and compound lipids are esters of fatty acids, the free fatty acids only occur in relatively small amounts in food and food raw materials. Edible oils usually contain only small amounts of free fatty acids when isolated from the plant, animal, or microbiological source. The concentration of free fatty acids rises in bruised fruit such as palm fruit if the lipases present are not inactivated rapidly by heating.

However, normal values of free fatty acids are below 5% in most crude edible oils. The free fatty acids present are mixtures similar to those present in the triglycerides. Most free fatty acids are removed from crude edible oils during refining because of the undesirable effects on flavor, and the reduction of the smoke point caused by free fatty acids. A refined oil usually contains less than 0.1% free fatty acids.

Vitamin A

Vitamin A is a C_{20} primary alcohol occurring either as vitamin A_1 or vitamin A_2. The vitamin occurs in animal and marine products such as butter (1 mg per 100 g) and eggs (3.7 mg per 100 g). It is commonly used in the form of acetate or palmitate esters as a food additive. Vitamin A can also be formed *in vivo* from some carotenoids such as β-carotene. (*See* **Retinol**: Properties and Determination.)

Vitamin D

The term 'vitamin D' refers collectively to a group of compounds derived from sterols, which are effective in preventing the development of rickets in children. Ergocalciferol (vitamin D_2) and cholecalciferol (vitamin D_3) are the most common compounds with vitamin D activity. Ergocalciferol is formed by the irradiation of ergosterol with ultraviolet light whereas cholecalciferol is formed in an analogous manner from 7-dehydrocholesterol. (*See* **Cholecalciferol**: Properties and Determination.)

Vitamin E

Eight compounds belonging to the tocopherol and tocotrienol classes have vitamin E activity. Vegetable oils in general contain more vitamin E than animal fats, and vegetable oils rich in polyunsaturated fatty acids tend to be good sources of vitamin E. (*See* **Tocopherols**: Properties and Determination.)

Vitamin K

Compounds exhibiting vitamin K activity possess a 2-methyl-1, 4-naphthoquinone ring with differing chain-length polyprenyl substituents at the 3 position. The dietary requirement for vitamin K by the adult human is extremely low, because the vitamin is synthesized in the intestine by intestinal bacteria. The main dietary sources of vitamin K are green vegetables, which commonly contain > 100 μg per 100 g.

Carotenoids

Carotenoids contribute the yellow, orange, or red colors of many fruits and vegetables, such as carrots, peaches, and tomatoes. They are commonly extracted with vegetable oils, including crude palm oil, which is a rich source containing 500 p.p.m. (*See* **Carotenoids**: Occurrence, Properties, and Determination.)

Sterols

Sterols occur in the membranes of plants, animals, and microorganisms and are termed phytosterols, zoosterols, and mycosterols, respectively. Cholesterol is the main zoosterol, but sterols in plants commonly occur as mixtures with β-sitosterol, campesterol, and stigmasterol representing three of the major phytosterols. These sterols are all Δ^5-sterols (Figure 6), but Δ^7-sterols may also be present in small quantities (Table 1). Molecules with a sterol-type structure that lack an endocyclic double bond are termed sterols. Methyl sterols, also known as triterpenyl alcohol, have an additional methyl group at carbon-4 in the A-ring of the molecule. Methyl sterols and dimethylsterols commonly occur with sterols in plant membranes. Bacteria usually lack sterols in their membranes, but yeasts accumulate considerable quantities of sterols, which may represent up to 10% of the cellular dry weight. Algae produce a wide variety of sterols. (*See* **Cholesterol**: Properties and Determination.)

Terpenes

Terpenes are oligomers or polymers of isoprene, 2-methyl-1, 3-butadiene. Monoterpenoids and sesquiterpenoids contain 10 and 15 carbon atoms respectively, and these components occur in essential oils and are very important flavor compounds. Monoterpenoids include acyclic and cyclic hydrocarbons, alcohols, ketones, or lactones such as myrcene, limonene, and menthol. Sesquiterpenoids include

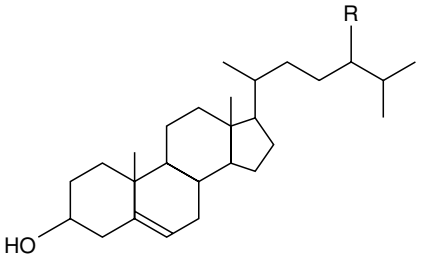

Cholesterol	R = H
β-Sitosterol	R = – C_2H_5
Stigmasterol	R = – C_2H_5 ; Δ^{22}
Campesterol	R = – CH_3
Δ^5-Avenasterol	R = CH – CH_3
Brassicasterol	R = – CH_3 ; Δ^{22}

Figure 6 Sterol structure. Reproduced from Lipids, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 1 Compositions of sterol fractions of 19 vegetable oils

	Total sterols (%)	*Composition (%)*[a]									
		I	*II*	*III*	*IV*	*V*	*VI*	*VII*	*VIII*	*IX*	*X*
Oil, corn	1.2	Tr[b]	Tr	23	6	66	4	1	Tr	–	–
Rice bran	1.8	Tr	Tr	28	15	49	5	1	2	–	–
Wheat germ	2.6	Tr	Tr	22	Tr	67	6	3	2	–	–
Coconut	0.2	1	Tr	8	13	58	14	6	–	–	–
Palm	0.3	1	Tr	14	8	74	2	1	–	–	–
Palm kernel	0.1	3	Tr	9	11	70	6	1	Tr	–	–
Peanut	0.2	Tr	Tr	15	9	64	8	3	1	–	–
Soya bean	0.4	Tr	Tr	20	20	53	3	3	1	–	–
Sunflower	0.4	–	–	8	8	60	4	15	4	–	1
Safflower, linoleic-rich	0.4	–	Tr	13	9	52	1	20	3	2	Tr
Safflower, oleic-rich	0.4	–	–	15	10	52	1	15	5	2	Tr
Olive (France)	0.2	–	–	2	1	91	2	4	Tr	–	–
Olive (Italy)	–	–	–	3	1	84	12	Tr	Tr	–	–
Castor	0.3	Tr	–	10	22	44	21	2	1	–	–
Kapok	0.3	Tr	Tr	9	2	86	2	1	–	–	–
Cottonseed	0.4	Tr	Tr	4	1	93	2	Tr	Tr	–	–
Linseed	0.4	1	Tr	29	9	46	13	2	–	–	–
Rapeseed	0.6	Tr	10	25	Tr	58	2	5	–	–	–
Sesame	0.6	–	–	19	10	62	7	2	–	–	–
Cocoa butter	0.3	2	Tr	9	26	59	3	1	Tr	–	–
Coffee seed	1.8	Tr	Tr	19	20	54	6	1	Tr	–	–

[a]I. cholesterol; II. brassicasterol; III. compesterol, IV. stigmasterol; V. β-sitosterol; VI. Δ^5-avenasterol; VII. Δ^3-stigmasterol; VIII. Δ^5-avenasterol; IX and X. unidentified.

[b]Tr, trace, less than 0.5%.

Data from Itoh T, Tamura T and Matsumoto T (1973) Sterol composition of 19 vegetable oils. *Journal of the American Oil Chemists' Society* 50: 122–125.

bicyclic or tricyclic compounds such as zingiberene, which is the main constituent of the oil of ginger, or caryophyllene, which occurs in the oil of cloves. (*See* **Essential Oils:** Properties and Uses.)

Diterpenoids include the bitter principles, which are commonly bitter and have physiological effects. Triterpenoids and higher terpenoids have no flavor properties. Triterpene alcohols, which occur in small amounts in edible oils, have been found useful in the identification of oils in blends. Carotenoids are the major tetraterpenoids, while chicle, which is used in chewing gum, is the main polyterpenoid used in foods.

Ethers

Ether lipids occur in which one of the hydroxyl groups of a diglyceride or phosphatidyl ester is linked to an alkyl group. The alkyldiacylglycerols are common constituents of some marine oils while vinyl ethers or plasmalogens are found in blood.

See also: **Bread**: Chemistry of Baking; **Carotenoids**: Occurrence, Properties, and Determination; **Cholecalciferol**: Properties and Determination; **Cholesterol**: Properties and Determination; **Emulsifiers**: Organic Emulsifiers; **Essential Oils**: Properties and Uses; **Fatty Acids**: Properties; **Lipoproteins**; **Phospholipids**: Properties and Occurrence; **Retinol**: Properties and Determination; **Tocopherols**: Properties and Determination; **Triglycerides**: Structures and Properties

Further Reading

Abu-Hadeed AM and Kotb AR (1988) A method for the quantitative determination of individual oils in a blend. *Journal of the American Oil Chemists' Society* 65: 1922–1926.

Akoh CC and Min DB (2002) *Food Lipids: Chemistry, Nutrition and Biotechnology*, 2nd edn, New York: Marcel Dekker.

Gunstone FD (1996) *Fatty Acid and Lipid Chemistry*. London: Blackie Academic & Professional.

Gunstone FD, Harwood JL and Padley FB (1994) *The Lipid Handbook*, 2nd edn, London: Chapman and Hall.

Harwood JL and Russell NJ (1984) *Lipids in Plants and Microbes*. London: Allen and Unwin.

Hui YH (1996) *Bailey's Industrial Oil and Fat Products*, 5th edn. New York: Wiley.

Itoh T, Tamura T and Matsumoto T (1973) Sterol composition of 19 vegetable oils. *Journal of the American Oil Chemists' Society* 50: 122–125.

Ratledge C and Wilkinson SG (1988) *Microbial Lipids*, vols 1, 2. London: Academic Press.

Salunkhe DK, Chavan JK, Adsule RN and Kadam SS (1992) *World Oilseeds: Chemistry, Technology, and Utilization*. New York. Van Nostrand Reinhold.

Occurrence

M H Gordon, The University of Reading, Reading, UK

Fats

Fats can be extracted from biological tissues with organic solvents. They comprise complex mixtures of chemical classes, known as lipids, which share the same solubility characteristics. Fats include solid fats and liquid oils. They occur in animals, plants, and microorganisms either as storage lipids, which are potential sources of energy by β-oxidation, or as membrane lipids. Storage lipids mainly comprise triacylglycerols, whereas membrane lipids include phospholipids, sterols, sphingolipids, and glycolipids. (*See* **Phospholipids**: Properties and Occurrence; **Triglycerides**: Structures and Properties.)

Membrane lipids occur in cells in a phospholipid bilayer which has been described as a fluid mosaic. The fatty acid chains of the phospholipids are directed towards the center of the membrane. Globular protein molecules penetrate into either side or extend entirely across the membrane. The membrane is considered to be asymmetric with oligosaccharide chains protruding from glycolipids and glycoproteins on the outer surface of the plasma membrane of eukaryotic cells. Sterols and polar lipids are embedded into the membrane. Phytosterols, including β-sitosterol, campesterol and stigmasterol, occur in plant cell membranes, whereas cholesterol occurs in animal cell membranes.

Fat in Animal Products

Body Fat

The bulk of the meat consumed in the UK is obtained from sheep, cattle, and pigs, with chickens and turkeys being major poultry products. The fat of the adipose tissue of animals usually consists of more than 99% acylglycerols, but fat occurring within the muscles of animals, which are consumed as meat, contains a considerable amount of phospholipids and of unsaponifiable components, such as cholesterol. Palmitic acid, stearic acid and oleic acid are the major fatty acids present in the fat of animals used for meat, and therefore fats derived from animals such as lard and tallow are relatively saturated (Table 1) and consequently semisolid. Fat derived from poultry is more unsaturated. (*See* **Fatty Acids**: Properties.)

Table 1 Major fatty acids in animal fats (% w/w)

Fatty acid	*Lard*	*Beef tallow*	*Bovine milk fat*
4:0			3.3
6:0			1.6
8:0			1.3
10:0			3.0
12:0			3.1
14:0	0.5–2.5	1.4–6.3	9.5
16:0	20–32	20–37	26.3
16:1	1.7–5.0	0.7–8.8	2.3
18:0	5.0–24	6–40	14.6
18:1	35–62	26–50	29.8
18:2	3.0–16	0.5–5.0	2.4
18:3	< 1.5	< 2.5	0.8
20:0	< 1.0	< 0.5	
22:1			

The fatty acid content of depot fat (storage lipid) varies depending on the part of the animal from which the fat is derived. Thus, the oleic acid content of sheep depot fat varies from 19.1 to 35.7%, depending on the anatomical source of the fat. The fatty acid content of depot fats in animals is also dependent partly on the diet since *de novo* synthesis *in vivo* supplies only part of the fatty acids. Unsaturated lipids consumed by ruminants such as cattle and sheep are hydrogenated by rumen microflora before they are incorporated into depot fats. (*See* **Fatty Acids**: Metabolism.)

Milk Fat

Bovine milk contains a complex mixture of lipids, with triacylglycerols representing the major component at 97–98% of the total lipid content. However, diacylglycerols, monoacylglycerols, free fatty acids, sterols, phospholipids, hydrocarbons, and sterol esters are also present. The fat content of milk from cattle reared in developed countries is in the range of 3.4–5.1%. The fatty acid composition of milk fat is extremely complex, with over 400 fatty acids having been identified; the major components are indicated in Table 1. Although the level of linoleic acid (18:2) is normally below 3% due to hydrogenation of dietary linoleic acid in the rumen, the feeding of oil coated with a layer of formaldehyde-treated protein has allowed the linoleic acid content of milk fat to be increased above 15% in experimental animals. The fatty acid composition of milk fat varies with the species of animal. Thus, the oleic acid content of camel milk fat is 38.9% compared with that of horse milk, which is 18.7%, and similar differences are also found in other fatty acids. (*See* **Milk**: Dietary Importance.)

Fat in Vegetable Products

Lipids in plant materials can also be divided into membrane lipids and storage lipids. The composition of membrane lipids varies with membrane function. Cytoplasmic or plasma membranes contain about half the dry mass as lipid. Major components include phospholipids up to 65%, glycolipids up to 20%, sterols up to 5%, and neutral lipids including hydrocarbons, diacylglycerols, and pigments. The main phospholipids are phosphatidylcholine, phosphatidylethanolamine, and phosphatidylinositol. Mitochondrial membranes contain up to 98% of the lipids as phospholipids. Chloroplast membranes are unique in nature, since the major lipid components of their lamellae are glycosylglycerides and not phospholipids.

Triacylglycerols occur in oil bodies, which are surrounded by a monolayer or half-unit membrane. As the oil content of seeds increases, there is a large increase in the number of oil bodies. Lipids are stored in the oil bodies until required as a source of energy by the growing plant. Storage lipids from plant sources are of considerable industrial importance, with world production of refined plant oils exceeding 50 million tonnes per year. The commercial sources of edible plant oils include: oilseeds, with soya bean, rapeseed, sunflower seed, peanuts, and cottonseed being important commercially; tree and bush products, including palm oil, coconut oil, olive oil, cocoa butter, and palm kernel oil; and grains, particularly corn oil. (*See* **Vegetable Oils**: Dietary Importance.)

The oil content of the oil-bearing material (Table 2) is one of the factors that affects the commercial utilization of the crop. However, other factors include yield per hectare and ease of harvesting. In the case of soya bean oil, the oil is a byproduct in meal production, and therefore it is a very important oil commercially despite having a relatively low oil content in the seed.

Table 2 Oil content of a number of oil-bearing vegetable materials

Oil-bearing material	*Oil content (% w/w)*
Coconut	65–68
Babassu	60–65
Sesame seed	50–55
Palm fruit	45–50
Palm kernel	45–50
Peanut	45–50
Rapeseed	40–45
Sunflower seed	35–45
Safflower seed	30–35
Olive	25–30
Cottonseed	18–20
Soya bean	18–20

Crude vegetable oils are extracted from fruits or seeds by pressing or by solvent extraction, usually with hexane, and in some cases, the two processes are applied sequentially. The extracted crude oil contains minor lipids as well as triacylglycerols. Phospholipids, sterols, sterol esters, tocopherols, free fatty acids, diacylglycerols, carotenoids, chlorophylls and hydrocarbons are commonly present. The fatty acid composition of common plant oils is shown in Table 3. There is considerable research by plant breeders into the development of plant varieties with modified fatty acid composition. This has led to the development of low-erucic acid rapeseed oil, which is considered preferable as a food oil to traditional rapeseed oil, which has an erucic acid content of about 45% of the total fatty acid content. Animal experiments suggest that high levels of erucic acid may lead to fatty lesions in heart muscles. High-oleic sunflower oil has also been developed because of nutritional interest in this type of oil.

Fish Oil

Fish can be classified as marine lean fish, marine fatty fish and freshwater fish. Marine lean fish, including cod, haddock, and hake, have a flesh lipid content of 0.1–1% with little seasonal variation. However, the flesh lipid content of marine fatty fish, including sprats, mackerel, and herring, varies from <1 to $>25\%$ in some species, with considerable variation, depending on the season and the diet. For example, the lipid content of sprats varies from a maximum of 17% in autumn to winter and falls to about 6% in spring to summer. Freshwater fish, including rainbow trout, haplochromis, and rock bass, have relatively low fish lipid contents of $<4\%$. (*See* **Fish Oils**: Dietary Importance.)

In the flesh of lean fish such as cod, 65% of the total lipids are phospholipids intimately associated with muscle protein, and 35% are neutral lipids including triacylglycerols and sterols. In the flesh of fatty fish, a large proportion of the lipids are triacylglycerols present as extracellular globules in the muscle. Sprat flesh contains 8–15% triacylglycerols and 0.6–1.9% phospholipids, whereas rainbow trout contain 2.2% triacylglycerols and 0.9% phospholipids. Both species also contain about 1% sterols. The fatty acid contents of the fish lipids of several species of fish are shown in Table 4. The unsaturated fatty acids are commonly mixtures of positional isomers. Fish oils are of considerable interest as nutritional supplements because of the high level of eicosapentaenoic (20:5) and docosahexaenoic acids

Table 3 Fatty acid composition of plant oils (% w/w)

Fatty acid	Coconut	Palm kernel	Sunflower	Peanut	Corn	Olive	Palm	Cotton seed	Soya bean	Rapeseed	Cocoa butter
6:0	0–0.8	< 1.5									
8:0	5–9	3–5									
10:0	6–10	3–7									
12:0	44–52	40–52					< 1.2				
14:0	13–19	14–18	< 0.5	< 0.1	< 1	tr	0.5–5.9	0.5–2.0	< 0.5	0.9–1.2	
16:0	8–11	7–9	3–10	6–15.5	8–19	8.7–18.3	32–59	17–29	7.0–12	4.5–6.0	24.2–27.0
16:0	< 1	< 1	< 1	< 1	< 0.5	0.2–2.4	< 0.6	0.5–1.5	< 0.5		
18:0	1–3	1–3	1–10	1.3–6.5	0.5–4	1.8–4.4	1.5–8.0	1.0–4.0	2.0–5.5	1.5–2.1	32.6–35.4
18:1	5–8	11–19	14–65	36–72	19–50	56.4–82.2	27–52	13–44	19–30	48.3–60.7	33.8–36.9
18:2	< 2.5	0.5–2	20–75	13–45	34–62	4.2–18.9	5.0–14	33–58	48–58	18.8–22.0	2.7–4.0
18:3			< 0.7	< 1	< 2	0.3–1.1	< 1.5	0.1–2.1	4.0–12.1	9.3–10.8	
20:0	< 0.5	< 1	< 1	1–25	< 1	0.2–1.3	< 1	< 0.5	< 1.0	0.6–0.8	
22:0			< 1	1.5–4.8	< 0.5	tr		< 0.5	< 0.5	0.1–0.2	
22:1			< 0.5	< 0.1				< 0.5		0.1–5.1	
24:0				1–2.5	< 0.5			< 0.5		0.2	

tr, trace amount.

Table 4 Fatty acid composition of fish oil (% w/w)

Fatty acid	Cod flesh	Haddock	Mackerel	Sprats	Rainbow trout	Cod liver
14:0	1.4	1.5	8.6	6.0	3.5	2.8
16:0	19.6	20.0	17.6	21.5	13.3	10.7
16:1	3.5	4.0	10.0	5.3	4.8	6.9
18:0	3.8	6.1	2.2	2.4	3.8	3.7
18:1	13.8	14.2	14.8	16.5	18.7	23.9
18:2	0.7	2.2	1.0	1.6	5.5	1.5
18:3	0.1	0.4	0.8	1.3	5.9	0.9
18:4	0.4	0.5	2.0	3.3	2.1	2.6
20:1	3.0	2.6	8.6	7.0		8.8
20:4	2.5	3.3	0.9	0.7	7.2	1.0
20:5	17.0	12.0	9.4	8.1	5.1	6.0
22:1	1.0	0.1	10.2	12.0		5.3
22:5	1.3	2.4	1.2		6.2	1.3
22:6	29.8	24.5	8.7	10.8	21.0	14.3

(22:6) in some oils. These fatty acids are significant because they are believed to play a role in reducing the incidence of coronary thrombosis. Fats from marine mammals have a similar composition to fish oil.

Microbial Oil

There has been considerable interest in microbial lipids in recent years because of the possibility of commercial production of oils and fats from microbial sources. The lipid composition of microorganisms is also useful in taxonomy.

Bacteria

The major lipids in Gram-positive bacteria are phosphatidylglycerol and phosphatidylethanolamine, in which the acyl chains are mainly odd-numbered branched-chain fatty acids. In contrast, Gram-negative bacteria contain phosphatidylethanolamine with even-numbered straight chain and cyclopropane acyl chains. The lipids in Gram-positive bacteria are limited to the cytoplasmic membranes and the mesosomes, whereas in Gram-negative bacteria and mycobacteria, the lipids are interwoven with the external layers of cells, and there is no clear distinction between membrane and cell-wall lipids. The lipid content of many species of bacteria is rather low, although mycobacteria, corynebacteria, and nocardia may contain substantial amounts, e.g., 23% in mycobacteria grown in glycerol as the carbon source.

Algae

Macroalgae such as seaweeds and phytoplankton have low lipid contents, but microalgae grown in the presence of light and in a nitrogen-poor medium accumulate storage lipids rich in triacylglycerols. Oil droplets are clearly visible in the cells. Lipid contents as high as 70% can be achieved in algae grown under optimum conditions for lipid biosynthesis. Some algae may accumulate relatively high concentrations of derived lipids of commercial interest. Thus, *Dunaliella salina* accumulates more than 10% of its dry mass as β-carotene.

Yeasts

Yeasts vary in oil content with low-oil, medium-oil and high-oil types containing < 5, 5–15 and > 15% oil as percentage cell dry weight, respectively. Oleaginous yeasts have been defined as those that produce

>25% oil. The oil content and composition varies between the various yeast subfamilies, and members of a particular genus, but also varies within strains and depends on the growth conditions. Up to 65% of the biomass of oleaginous yeasts, e.g., *Candida curvata* or *Lipomyces* spp., may comprise oil, if the yeast is grown under nitrogen-deficient conditions.

Molds

There are many examples of oleaginous molds that accumulate > 25% oil. Oil contents up to 75% have been described. Molds show a greater diversity of fatty acid composition compared with yeasts, but most species contain oleic acid, palmitic acid, and linoleic acid as the major acids. Higher proportions of polyunsaturated fatty acids occur in molds than in yeasts, and molds are of interest commercially for the production of oil containing polyunsaturated fatty acids of nutritional interest. Thus, *Mucor javanicus* has been used to produce an oil containing γ-linolenic acid, which has been sold as a nutritional supplement in health food shops.

Fat in Processed Foods

As well as occurring naturally in the membranes of all biological tissues, fats are important in the production of processed foods. The major industrial applications of edible oils and fats are in frying oils, margarine, bakery fats, shortenings, mayonnaise, salad dressings, peanut butter, chocolate, and confectionery products. The oil acts as a heat transfer medium for frying, and it has a major effect on the texture and flavor of the processed foods. Fat-soluble vitamins and pigments are also present in the foods. (*See* **Fats**: Uses in the Food Industry.)

Phospholipids in the form of commercial lecithin are another class of lipids added to processed foods. Lecithin is added in baking formulations, cake mixes, chocolate, confectionery products, icecream, instant foods, margarine, nondairy toppings, and shortenings. The lecithin in these products affects the texture by improving the shortening effect of fat, stabilizing emulsions, reducing the viscosity of sugar in fat dispersions, improving the wettability and dispersion of powders in aqueous solutions, and assisting with foam stabilization.

Other lipids are also important as surfactants or emulsifiers in foods. Thus, mixtures of mono- and diglycerides are the most common emulsifiers used in foods to stabilize water-in-oil emulsions. They also have various roles as surfactants, including retarding the staling of bread, improving the behavior of foods during extrusion cooking, improving the palatability and reducing the stickiness of caramels and toffees, and improving the dispersion of coffee whiteners. Other lipid surfactants include acetylated monoglycerides, sucrose esters, sorbitan esters and polyoxyethylene sorbitan esters. (*See* **Emulsifiers**: Organic Emulsifiers.)

See also: **Emulsifiers**: Organic Emulsifiers; **Fats**: Uses in the Food Industry; **Fatty Acids**: Properties; Metabolism; **Fish Oils**: Dietary Importance; **Milk**: Dietary Importance; **Vegetable Oils**: Dietary Importance

Further Reading

Akoh CC and Min DB (2002) *Food Lipids: Chemistry, Nutrition and Biotechnology*, 2nd edn. New York: Marcel Dekker.

Cambie RC (1989) *Fats for the Future*. Chichester, UK: Ellis Horwood.

Gunstone FD, Harwood JL and Padley FB (1994) *The Lipid Handbook*, 2nd edn. London: Chapman & Hall.

Gunstone FD and Norris FA (1983) *Lipids in Food*. Oxford: Pergamon Press.

Harwood JL and Russell NJ (1984) *Lipids in Plants and Microbes*. London: Allen & Unwin.

Hui YH (1996) *Bailey's Industrial Oil and Fat Products*, 5th edn. New York: Wiley.

Kamel BS and Kakuda Y (1994) *Technological Advances in Improved and Alternative Sources of Lipids*. London: Blackie Academic & Professional.

Ratledge C and Wilkinson SG (1988, 1989) *Microbial Lipids*, vols 1 and 2. London: Academic Press.

Salunkhe DK, Chavan JK, Adsule RN and Kadam SS (1992) *World Oilseeds: Chemistry, Technology, and Utilization*. New York: Van Nostrand Reinhold.

Shukla VKS and Gunstone FD (1992) *Oils and Fats in the Nineties*. Lystrup: International Foods Science Centre.

Weiss TJ (1983) *Food Oils and their Uses*, 2nd edn. Westport, CT: Avi.

FATTY ACIDS

Contents

Properties

B J F Hudson[†], University of Reading, Reading, UK

Background

Fatty acids, *per se*, do not occur in fresh foods, but comprise very important structural components of virtually all food lipids. They are to be found in the form of esters of glycerol in triglycerides, as esters of phosphatidic acid and its derivatives in phospholipids, and in similar combined forms in glycolipids, sulfolipids, sphingolipids, cerebrosides, etc. Fatty acids also occur in esterified forms in combination with natural aliphatic alcohols in waxes and with sterols in sterol esters. (*See* **Fats:** Classification; **Phospholipids:** Properties and Occurrence; **Triglycerides:** Structures and Properties.)

Free fatty acids are of industrial importance and have many applications. In food, however, their presence is a sign of rancidity as a consequence of lipolytic activity, and this is undesirable as it causes flavor taints.

Fatty acids are all lipophilic in character but in other respects may vary considerably from one another. Their distinctive chemical, physical, and physiological properties carry through to the properties of the lipids in which they occur. Features of particular importance are their melting, crystallization, and solubility behaviors, their type and degree of unsaturation (and hence chemical reactivity) and their potential for *in vivo* conversion to physiologically significant derivatives.

The most important natural fatty acids are listed in Table 1, but many others, probably hundreds, are known, and yet many others have been synthesized. This emphasizes the extreme complexity, in chemical terms, of a natural fat or oil. If it contains two different fatty acids, six triglycerides are possibly present, and for three fatty acids, 18 triglycerides may be present. A more typical oil, e.g., soya bean oil, with at least six fatty acids will contain well over 100 different triglycerides, in greatly varying amounts.

Nomenclature

For practical purposes, natural fatty acids are straight-chain even-numbered aliphatic carboxylic acids that may be saturated ($CH_3(CH_2)_xCO_2H$) or of a similar structure, but unsaturated, with up to six double bonds. The latter are normally arranged along the chain, separated from each other by methylene groups and with *cis*-conformations. The more important fatty acids (Table 1) have traditional ('trivial') names, since their systematic names are rather clumsy. In modern scientific literature, they are described by a shorthand notation. Thus, arachidonic acid is 20:4*n*-6, meaning that it is a straight-chain fatty acid with 20 carbon atoms and 4 *cis* double bonds, the first starting at the sixth carbon from the *end* of the chain, and therefore the other three at the ninth, twelfth, and fifteenth carbons, respectively. This system of nomenclature is now superseding the earlier one in which the positions of the double bonds were described by reference to the carboxyl end of the chain, counting the -CO_2H as carbon no. 1. Thus, by the old system, arachidonic acid was 20:4c6c9c12c15.

Sources

Referring to Table 1, the first five compounds are crystalline solids melting from around 40 to 70 °C and are the most frequently occurring saturated fatty acids. Palmitic and stearic acids occur universally in food lipids and, when comprising a substantial proportion of the total fatty acids, result in a *fat* rather than an *oil*. Lauric and myristic acids occur in

[†]Deceased.

Table 1 The more important natural food fatty acids

Trivial name	*Empirical formula*	*Shorthand structure*
Lauric acid	$C_{11}H_{23}CO_2H$	12:0
Myristic acid	$C_{13}H_{27}CO_2H$	14:0
Palmitic acid	$C_{15}H_{31}CO_2H$	16:0
Stearic acid	$C_{17}H_{35}CO_2H$	18:0
Arachidic acid	$C_{19}H_{39}CO_2H$	20:0
Palmitoleic acid	$C_{15}H_{29}CO_2H$	16:1 *n*-7
Oleic acid	$C_{17}H_{33}CO_2H$	18:1 *n*-9
Petroselinic acid	$C_{17}H_{33}CO_2H$	18:1 *n*-12
Erucic acid	$C_{21}H_{41}CO_2H$	22:1 *n*-9
Linoleic acid	$C_{17}H_{31}CO_2H$	18:2 *n*-6
α-Linolenic acid	$C_{17}H_{29}CO_2H$	18:3 *n*-3
γ-Linolenic acid	$C_{17}H_{29}CO_2H$	18:3 *n*-6
Arachidonic acid	$C_{19}H_{31}CO_2H$	20:4 *n*-6
Eicosapentaenoic acid[a]	$C_{19}H_{29}CO_2H$	20:5 *n*-3
Docosahexaenoic acid[a]	$C_{21}H_{29}CO_2H$	22:6 *n*-3

[a]In these cases, systematic names are given in the absence of trivial names.

certain tropical oils such as coconut and palm kernel oils. Arachidic acid occurs, albeit in small amounts, in a wide range of tropical vegetable fats. Animal fats, in addition to palmitic and stearic acids, contain small amounts of odd-numbered and branched-chain saturated acids. The lower fatty acids from butyric (4:0) to capric (10:0) are significant in ruminant milk fats. (*See* **Palm Oil**.)

The next four items in Table 1 are monounsaturated fatty acids. Oleic acid is universally important in food lipids, in some cases being as much as 80% of the component fatty acids (e.g., olive oil). Palmitoleic acid is also ubiquitous, though in much smaller amounts. Erucic and petroselinic acids, though important in a few seed oils, are of restricted occurrence, the former in the Cruciferae and the latter in the Umbelliferae. Except for erucic acid, these and all the acids referred to later in this section are liquid at ambient temperatures.

Linoleic acid is of almost universal importance in vegetable oils, but it does not occur to a significant extent in animal products. This diunsaturated fatty acid, the only really important one, is the 'essential fatty acid' on which most nutritional studies have been based and is the key member of the widely advertised 'polyunsaturates.'

The triunsaturated fatty acids, α- and γ-linolenic acids, are found, though very rarely together, in many seed oils, the former being the commoner. α-Linolenic acid is the major fatty acid of leaf and root lipids and many seed oils, especially those of northern latitudes. γ-Linolenic acid, though less widespread, is found in the oil seeds of many plant species, in some algae and fungi and certain animal products.

Of the last three fatty acids named in Table 1, arachidonic acid is of exclusively animal origin (e.g., eggs and certain meat offals) whilst the last two occur in substantial proportions in fatty fish (herring, mackerel) and fish liver oils (cod, halibut). (*See* **Fish Oils**: Dietary Importance; **Offal**: Types of Offal.)

Apart from those listed, and certain isomers and homologs of them, many other fatty acids, though less common, are well known. These may contain conjugated unsaturated systems (e.g., punicic acid), *trans* double bonds (e.g., columbinic acid), triple bonds (e.g., crepenynic acid), homocyclic (e.g., sterculic acid) or heterocyclic (e.g., vernolic acid) rings or substituent functional groups (e.g., ricinoleic acid). Some of these are occasionally consumed in food lipids.

Glycerides

Food scientists and technologists are much more concerned with glycerides than with fatty acids as such. Many, but not all, of the properties of fatty acids are mirrored in those of the nonpolar triglycerides. However, the mono- and diglycerides, in which, respectively, only one or two of the hydroxyls of glycerol are esterified, have both lipophilic and hydrophilic properties. The resulting surface activity enables them to serve as emulsifiers and foam stabilizers in processed foods. (*See* **Emulsifiers**: Organic Emulsifiers; Uses in Processed Foods.)

Mono- and Diglycerides

Monoglycerides containing a fatty acid (RCO_2H) exist in the form of 1-monoglycerides (1) or the isomeric 2-monoglycerides (2). Both forms revert, on storage, to an equilibrium mixture in which the 1-monoglyceride dominates (90%). Monoglycerides, accompanied by diglycerides, are produced by the base-catalyzed interesterification (randomization) of a triglyceride (5) with glycerol. The resulting mixed product is often used without purification, but if high-quality monoglycerides are desired, these can be obtained by molecular distillation of the mixed product. Monoglycerides have important applications in the food-processing industry for the stabilization of both water-in-oil (e.g., margarine) and oil-in-water (e.g., artificial cream) emulsions. (*See* **Margarine**: Methods of Manufacture.)

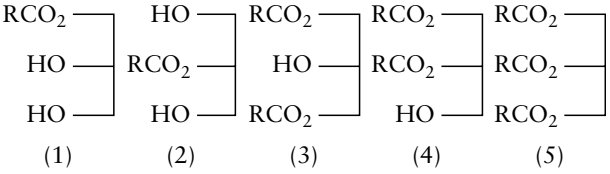

Diglycerides also exist in two forms ((3) and (4)), the 1,3-diglyceride (3) being dominant in the

equilibrium mixture. They are much less polar than monoglycerides and have few food applications.

Triglycerides

These are the dominant components (more than 90%) of all edible oils and fats. Since they contain about four to 10 significant amounts of different fatty acids, they comprise a very large number of individual triglycerides. The fatty acids are, however, not randomly distributed but tend to follow certain rules.

In oils and fats of vegetable origin, the preferred sites for saturated acids are the 1- and 3-positions, whilst unsaturated acids occupy the 2-position for preference. In animal fats (e.g., lard), the reverse is the case, except for the fats of ruminants (e.g., beef and mutton tallow, milk fat), in which the distribution is approximately random.

The positions of substituent saturated and unsaturated fatty acids in the triglyceride molecule are important in determining the hardness, elasticity, and melting characteristics (Table 2) of fats containing them. These physical parameters control the texture and mouth-feel of foods in which they are major components. Such compositional features are particularly important in the case of monounsaturated disaturated triglycerides, which can be represented as the isomers SUS and SSU.

Relatively hard vegetable fats (SUS-type triglycerides, 30–40 °C), such as cocoa butter melt sharply at, or slightly below, body temperature and hence are ideal components in chocolate, cream fillings, and confectionery couvertures. Admixture with fats of the SSU-type causes a fall in melting point and a marked increase in melting range. Such mixed triglycerides may therefore be well adapted to the formulation of margarines and bakery fats. (*See* **Cocoa**: Production, Products, and Use; **Fats**: Uses in the Food Industry.)

The melting and crystallization characteristics of both pure and mixed triglycerides are also very dependent on 'tempering' (close control of holding time and temperature), since they exist in distinct polymorphic forms, the most stable being that with the highest melting point.

Table 2 Melting points (°C) of fatty acids and derived glycerides

Fatty acid		*1-Monoglyceride*	*1,3-Diglyceride*	*Triglyceride*
12:0	44	63	56	46
14:0	54	70	65	57
16:0	63	77	72	65
18:0	70	81	78	73
18:1 *n*-9 (*cis*)	16	35	21	5
18:1 *n*-9 (*trans*)	44	58	55	42
18:2 *n*-6	−6	12	−3	−13

Natural fats can be physically separated into desirable fractions by various industrial processes, viz. by 'dry,' 'wet,' or 'solvent' fractionation, with appropriate careful temperature control. Fats such as edible tallow, lard, palm oil, palm kernel oil, and coconut oil can, in the 'dry' process, be separated by pressing or filtration into a higher-melting crystalline fraction and a liquid oil. In the 'wet' processes, the fat is mixed with an aqueous solution of a surface-active agent, which disperses the crystalline fraction into the aqueous phase. The separation is more efficient than in the dry processes. The most efficient and selective fractionation process makes use of a volatile solvent, usually acetone, to segregate, to a large extent, glycerides according to their degree of unsaturation: SSS, SUS, SUU, and UUU.

Reactions of Triglycerides

In the context of food, very few intentional reactions are permissible, since both the starting material and the reaction product must be of acceptable edible quality. Thus, only two kinds of reactions are of practical importance – hydrogenation and interesterification. To these, however, must be added lipolysis and autoxidation – undesirable spoilage reactions affecting oils and fats under unfavorable storage conditions, and thus causing rancidity.

Hydrogenation The objects of hydrogenation are (1) by increasing the degree of saturation of an oil or fat to raise its melting point and thus extend its applications, and (2) by reducing its degree of unsaturation to reduce its proneness to autoxidation and thereby to enhance its stability. Hydrogenation is always partial and sometimes only marginal. In chemical terms, it involves not only the saturation of double bonds but also *cis–trans* isomerization and migration of some double bonds to neighboring positions in the fatty chains.

Hydrogenation involves the use of hydrogen and a supported metal catalyst, commonly nickel on Kieselguhr, at temperatures between 100 and 200 °C. Conditions are varied according to requirements. Thus, the use of a fresh catalyst and low temperatures minimizes isomerization, whereas an old ('poisoned') catalyst at higher temperatures promotes isomerization and minimizes saturation. *Trans* fatty acids can be formed in high proportions in hydrogenated fats, with consequent rises in melting points (note the contrast in melting points between corresponding *cis* and *trans* 18:1 *n*-9 compounds in Table 2).

Interesterification The same base-catalyzed reaction, as described earlier for the production of monoglycerides, is applicable to the randomization of

individual fats or to the randomization of mixtures of two or more fats or oils. The process has several advantages. For example, lard, which in its natural form solidifies very slowly on cooling, producing large, 'grainy' crystals, after interesterification, rapidly sets on cooling to a compact mass of very small crystals. Edible tallow, with an inconveniently high melting point, after interesterification with a soft oil, such as soya bean oil, yields a soft fat with a very useful melting range. (*See* **Soy (Soya) Beans**: The Crop.)

Lipolysis Triglycerides are hydrolyzed, albeit very slowly, by aqueous media, but much more rapidly by basic catalysts and, more importantly, by endogenous lipolytic enzymes. Oils and fats (e.g., olive oil, palm oil, tallow) from fruit coats or animal tissues are particularly prone to lipolysis if they are allowed to remain too long in contact with bruised tissue before processing. As a result of such handling, major processing losses are suffered as a result of the need to carry out refining operations to remove harsh flavor taints.

Autoxidation Access of atmospheric oxygen to unsaturated fatty acids or glycerides leads to deterioration through oxidation. Saturated and monounsaturated fatty chains oxidize very slowly and do not as a rule cause problems. Diunsaturated chains, as in 18:2 *n*-6, oxidize more rapidly and polyunsaturated chains very rapidly. Oils such as fish oils are so liable to oxidative degradation that they cannot be used as such in food formulation but must first be hydrogenated. (*See* **Oxidation of Food Components**.)

Autoxidation depends on free-radical chain reactions, which involve the interaction of oxygen with free radicals generated at methylene groups adjacent to double bonds, especially at methylene groups between two double bonds. Hydroperoxides are thus formed, and these, being unstable, break down to fission products, including some of low molecular weight. Equation (1) is a summary of a typical reaction sequence.

$$R_1\text{–CH=CH–CH}_2\text{–}R_2\text{–G} \longrightarrow R_1\text{–CH=CH–CH–}R_2\text{–G} \longrightarrow \underset{\displaystyle \underset{\displaystyle |}{\text{OOH}}}{} \qquad R_1\text{–CH=CH–CHO} + \text{HO–}R_2\text{–G} \qquad (1)$$

(R_1 and R_2 are alkyl and alkylene portions of the fatty chain; G is the nonvolatile glyceride residue.)

The aldehydes and other volatile oxidation products have strong, unattractive odors and flavors, and make a predominating contribution to 'rancidity.' Further, peroxides are toxic food components, so autoxidation must be minimized by all possible means. It can be avoided, or at least greatly delayed, by careful and prompt handling and storage of oils and fats, by avoidance of prooxidative situations such as contact with metals, and by the use of various kinds of antioxidants that inhibit steps in the autoxidation process. (*See* **Antioxidants:** Natural Antioxidants; Synthetic Antioxidants.)

See also: **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; **Cocoa**: Production, Products, and Use; **Emulsifiers**: Organic Emulsifiers; Uses in Processed Foods; **Fats**: Classification; Uses in the Food Industry; **Fish Oils**: Dietary Importance; **Margarine**: Methods of Manufacture; **Offal**: Types of Offal; **Oxidation of Food Components**; **Palm Oil**; **Phospholipids**: Properties and Occurrence; **Soy (Soya) Beans**: The Crop; **Triglycerides**: Structures and Properties

Further Reading

Allen JC and Hamilton RJ (1989) *Rancidity in Foods*. London: Elsevier Applied Science.

Chapman D (1965) *The Structure of Lipids by Spectroscopic and X-ray Techniques*. London: Methuen.

Forss DA (1972) Odour and flavour compounds from lipids. In: Holman RT (ed.) *Progress in the Chemistry of Fats and Other Lipids*, vol. 13. Oxford: Pergamon Press.

Gunstone FD (1967) *An Introduction to the Chemistry and Biochemistry of Fatty Acids and their Glycerides*. London: Chapman & Hall.

Gurr MI and James AT (1975) *Lipid Biochemistry: an Introduction*. London: Chapman & Hall.

Hilditch TP and Williams PN (1964) *The Chemical Constitution of Natural Fats*. London: Chapman & Hall.

Hudson BJF (ed.) (1990) *Food Antioxidants*. London: Elsevier Applied Science.

Markley KS (ed.) (1967) *Fatty Acids, their Chemistry, Properties, Production and Uses*. New York: Interscience.

Mostofsky DI, Yehuda S and Salem N (eds) (2001) *Fatty Acids: Physiological and Behavioral Functions*. Totowa, USA: Humana Press.

Patterson HBW (1983) *Hydrogenation of Fats and Oils*. London: Elsevier Applied Science.

Patterson HBW (1989) *Handling and Storage of Oilseeds, Oils, Fats and Meal*. London: Elsevier Applied Science.

Swern D (ed.) (1979) *Baily's Industrial Oil and Fat Products*, 4th edn. New York: Wiley.

Metabolism

P Muñiz, Universidad de Burgos, Burgos, Spain

Fatty Acid Availability

Fatty acids are the principal form of stored energy for most organisms. Almost all tissues can utilize fatty acids for energy, exceptions being the brain, blood cells, skin, and renal medulla. The fatty acids present two advantages compared with other energetic biomolecules: their high energy density and low solubility. Thus, oxidation of tryacylglycerols releases $37\,kJ\,g^{-1}$ compared with the 16–17 $kJ\,g^{-1}$ for sugars, glycogen, and amino acids, and fatty acids can be stored without water.

Most of these fatty acids are stored in adipose tissue, comprising specialized cells, the adipocytes. Mammals can store 5–25% or more of their body weight in the form of lipids, and 90% of these lipids are tryacylglycerols (TGs). The fat storage is not static, so the tryacylglycerols are in a constant state of turnover. The liver, the major center of lipid metabolism, is largely responsible for regulating lipid levels in the body. Among its important functions are: synthesis of TG from fatty acids, synthesis of other lipids such as phospholipids and cholesterol, desaturation of fatty acids, and catabolism of tryacylglycerols for use as energy.

Since most of the tissues contain only small amounts of storage lipids, energy production depends on a continuous supply of fatty acids. The fatty acids are derived primarily from two sources: exogenous TG obtained from diet and endogenous TG stores mainly in the adipose tissue.

Diet Fatty Acids

Tryacylglycerols are the predominant type of lipid ingested by humans, and they are absorbed as fatty acids from the intestinal mucosa (enterocytes) into the lymphatic system. The free fatty acids (FFAs) inside the enterocytes are reesterified into TGs and packaged with proteins and phospholipids to form chylomicrons. Chylomicrons enter the lymphatic system and eventually into the circulatory system at the thoracic duct.

Most of the chylomicrons are removed from the blood into cells by the action of lipoprotein lipase, an enzyme located on endothelial cell lining the capillaries primarily in adipose tissue, skeletal and cardiac muscle, and the liver. Lipoprotein lipase hydrolyzes the TGs into fatty acids and glycerol, and is taken up into these specific tissues. In adipose tissue, the FFAs are primarily reesterified and stored as TGs, in other organs as muscle and liver, and small amounts of TG are stored intracellularly. The primary sites of FFA oxidation are cardiac and skeletal muscle and liver.

Adipose Tissue Fatty Acids: Lipolysis

Fat stored as tryacylglycerols in adipose tissue and mobilized in the form of plasma free fatty acids is the main source of energy in the absence of dietary substrates; up to 150 g of fatty acids can be liberated daily, corresponding approximately to half of the daily caloric needs.

Hydrolysis of adipose tissue TGs occurs via three consecutive reactions and is catalyzed by two enzymes: the triacylglicerol lipase, also called hormone-sensitive lipase (HSL), and the monoglyceride lipase (MGL). The rate-limiting enzyme in adipose tissue lipolysis is HSL. HSL hydrolyzes TGs and diglycerides, whereas the participation of MGL is required for complete hydrolysis of monoglycerides. Then, the three FFA molecules leave the adipose cell and enter the circulation, where they are transported bound to albumin. The remaining glycerol backbone also leaves the cell, because adipose tissue does not contain the enzyme glycerol kinase, and so glycerol is not reutilized for reesterification in adipocytes.

HSL exhibits a preference, although not absolute, for the (*n*-3/*n*-6) fatty acids from TGs, with approximately three- to fourfold higher hydrolysis rates to the 1(3)-ester bonds than to the 2-ester bond. Fatty acids, i.e., oleic and oleoyl coenzyme A (CoA), and monoacyglycerols have been shown as inhibitors to HSL. This feedback inhibition was proposed to prevent the accumulation of fatty acids and free cholesterol in the cells and to ensure that the capacity of MGL to hydrolyze MGs is not exceeded. Recently, adipocyte lipid-binding protein (ALBP) was shown to interact with HSL, raising the possibility that ALBP sequesters fatty acids resulting from HSL-catalyzed lipolysis, and thus preventing the feedback inhibition by lipolytic products.

Regulation of adipocyte lipolysis Fatty acid concentrations are the result of the balance between the rate of appearance (lipolysis) and the rate of clearance. Today, it is generally accepted that lipolysis is controlled mainly by the activity of the sympatic nervous system and by plasma insulin levels. The main regulator factor of the rate of adipose tissue lipolysis is the activity of HSL.

HSL activity is mainly regulated by the circulating concentrations of insulin and epinephrine. These hormones can change during times of stress, such as

starvation, trauma, or physical exercise, or after feeding, resulting in an alteration in lipolytic rates and FFA concentrations. Insulin decreases the activity of hormone-sensitive lipase, which is responsible for the breakdown of adipose TG and concomitant release of the free fatty acids into the circulation. Epinephrine, norepinephrine, and adrenocorticoids, especially adrenocorticotropic hormone, increase the rate of fat mobilization by stimulating HSL activity.

HSL is hormonally regulated via phosphorylation/dephosphorylation. The hormone stimulated by HSL acts through a membrane-bound receptor and activates the membrane-bound enzyme adenylate cyclase, resulting in an increase in cyclic adenosine monophosphate (cAMP) levels in the adipocyte. cAMP directly activates a protein kinase A (PKA), which phosphorylates and activates HSL. Lipolytic agents generally increase the levels of cAMP far above the concentrations required for maximal activation of protein kinase A.

Epinephrine is the primary stimulator of lipolysis working through a β-adrenergic receptor in the plasma membrane of the adipose cell resulting in phosphorylation and activation of HSL. The physiological state can affect the sensitivity to epinephrine. For example, in short-term (3-day) fasting, the lipolytic responsiveness to epinephrine infusion is enhanced, whereas in obesity, the responsiveness to epinephrine is blunted. Catecholamines are able to stimulate lipolysis via three subtypes of β-adrenergic receptors, which are positively coupled to adenyl cyclase by Gs proteins and to inhibit lipolysis via α2-adrenergic receptors, negatively coupled to the enzyme by Gi proteins. In humans, the interplay between α2 and β-adrenergic receptors plays an important role in modulating cAMP levels in adipocytes.

However, the inhibition of HSL catalyzes its dephosphorylation, and so insulin induces a decrease in cAMP levels and a concomitant decrease in PKA activity, and inhibits HSL, thus catalyzing its dephosphorylation. Glucose is the predominant signal for insulin release, so an effect of glucose intake on lipolysis would be expected. Thus, in humans, an increase in glucose concentration in the plasma stimulates insulin release with concomitant increase in glucose uptake and fatty acid reesterification, and a decrease in lipolysis. The inhibitory effect of glucose on lipolysis is mediated entirely by insulin.

Under most physiological circumstances, the rate of lipolysis is largely determined by the balance between the stimulatory effect of epinephrine and the inhibitory effect of insulin. For example, in short-term fasting, there is a marked increase in the rate of lipolysis, which can be attributed to both a decrease in blood glucose (and thus insulin) concentration and an increase in epinephrine concentration and sensitivity to epinephrine. The striking aspect of the regulation of lipolysis is that the factors that are primary regulators (insulin and epinephrine) are not released in response to a signal related to any aspect of fatty acid metabolism. For example, acute changes in fatty acid concentrations by infusion do not affect insulin, catecholamines, or the rate of lipolysis. Therefore, insulin and catecholamines, rather than being primary regulators of lipolysis, may be considered to be regulators of other physiological and metabolic processes that also affect lipolysis. In fact, the rate of resting lipolysis generally provides fatty acids at a rate that is far in excess of the rate required for oxidation. Fatty acids that are released from adipocytes, but not oxidized, are reesterified, thereby completing a substrate cycle.

Cellular Uptake of Fatty Acid

Fatty acids from the diet or produced by lipolysis are transported, bound to proteins in blood, and their cellular uptake is similar. The majority of complex lipid is hydrolyzed by two related lipases, hepatic lipase and lipoprotein lipase, and this hydrolysis enables free fatty acids to enter cells.

The uptake of fatty acids across the plasma membrane takes place either by diffusive processes, involving partitioning of the fatty acid molecule into the lipid bilayer of the plasma membrane, or by the presence of different lipid-binding proteins in the cell membranes, called plasma membrane fatty-acid-binding proteins (FABPpms).

At least five plasma membrane-associated fatty acid binding proteins were proposed to play a role in free fatty acid (FFA) uptake, a plasma membrane fatty-acid-binding protein (FABPpm, 40 kDa), fatty acid translocase (FAT, 88 kDa), fatty acid transport protein (FATP 63 kDa), caveolin, and 56-kDa kidney fatty acid binding protein. Once inside the cell, fatty acids are bound noncovalently to a family of intracellular fatty acid-binding proteins (FABPs) that may play a role in the transfer of FFAs to intracellular destinations where they are used. FABPs comprise a family of 14–15-kDa proteins that bind long-chain fatty acids.

Catabolism of Fatty Acids

Once within the cell and depending on the tissue and its metabolic demand, fatty acids are either converted to tryacylglycerols or membrane phospholipids, or oxidized in the mitochondria for energy production. Some metabolic routes are common to most of the

cells, like β-oxidation, whereas others are tissue-specific for prostanoid synthesis.

Before being directed into storage or membranes or oxidation, fatty acids are first activated to acyl-CoAs.

Fatty Acids Activation and Transport into Mitochondria

A requisite step for fatty acid entering several metabolic pathways is the formation of fatty acyl-CoA-thioesters, catalyzed by fatty acyl-CoA synthetase, with some substypes showing fatty acid chain-length specificity. The fatty acyl-CoA products of these enzymes are also bound to FABP.

Short- and medium-chain fatty acids are transported into the mitochondria matrix as free acids and form the acyl-CoA derivatives over there. However, long-chain fatty acids are normally activated in outer mitochondrial membrane, but this may also occur at the surface of the endoplasmic reticulum.

The enzyme acyl-CoA synthetase, also called acyl-CoA ligase or fatty acid thiokinase, catalyzes the formation of a thiol ester bound between the fatty acid and the thiol group of coenzyme A (eqn (1)).

$$\text{Fatty acid} + \text{ATP} + \text{CoA-SH} \rightarrow \text{Fatty acyl-CoA} + \text{AMP} + 2\ \text{Pi} \qquad (1)$$

The reaction is accompanied by the hydrolysis of ATP to form AMP and pyrophosphate. The pyrophosphate is rapidly hydrolyzed by inorganic pyrophosphatase to two molecules of phosphate, and the synthetase reaction is strongly promoted.

The long-chain acyl-CoA derivatives cannot be transported into the mitochondria matrix directly. They must be converted to acylcarnitine derivatives and then transported through the inner membrane of the mitochondrion. This transport system consists of the enzyme carnitine palmitoyltransferase I (CPTI) localized in the mitochondrial outer membrane, of an integral inner membrane protein the carnitine:acylcarnitine translocase, and of the enzyme carnitine palmitoyltransferase II localized on the matrix side of the inner membrane as shown in Figure 1. The enzyme CPTI catalyzes the formation of the *O*-acylcarnitine, which is transported across the inner membrane by a translocase. Inside the mitochondria, the enzyme carnitine palmitoyltransferase II (CPTII) transfers the fatty acyl group back to CoA and obtain fatty acyl-CoA.

β-Oxidation Reactions

Once inside the mitochondrial matrix, the fatty acyl-CoA is catalyzed by the enzymes in the β-oxidative pathway, resulting in acetyl-CoA production. Energy produced in the β-oxidation of fatty acids and the oxidation of acetyl-CoA in the tricarboxylic acid cycle is used to generate ATP via oxidative phosphorylation. The fatty acid oxidation begins with the oxidation of the carbon that is in 'β' position to the carboxyl carbon, and so this process has come to be known as β-oxidation.

The mitochondrial β-oxidation of saturated fatty acid acids proceeds by repeated cycles of four reactions (flavin adenine dinucleotide (FAD)-linked dehydrogenation, hydration, nicotinamide adenine dinucleotide $(NAD)^+$-linked dehydrogenation, hydration, NAD^+-linked dehydrogenation, and thiolysis) with chain shortening being achieved by the successive cleavage of a C_2 unit as acetyl-CoA at each cycle until the substrate has been completely converted in acetyl-CoA. This is catalyzed by the sequential action of four enzymes families (acyl-CoA dehydrogenase, enoyl-CoA hydratase, L-hydroxyacyl-CoA dehydrogenase, and L-ketoacyl-CoA thiolase) each with a different substrate specificity for short-, medium-, and long-chain acyl-CoAs.

As shown in Figure 2, the first reaction is catalyzed by acyl-CoA dehydrogenase, a flavoprotein that is present as a family of three enzymes that differ in their specificity for the length of acyl-CoAs. Their mechanism involves the oxidation of Cα–Cβ bound

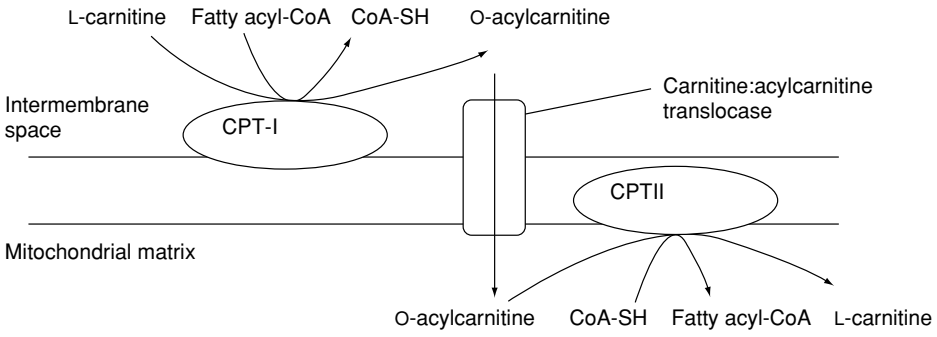

Figure 1 Transport into mitochondria of fatty acid.

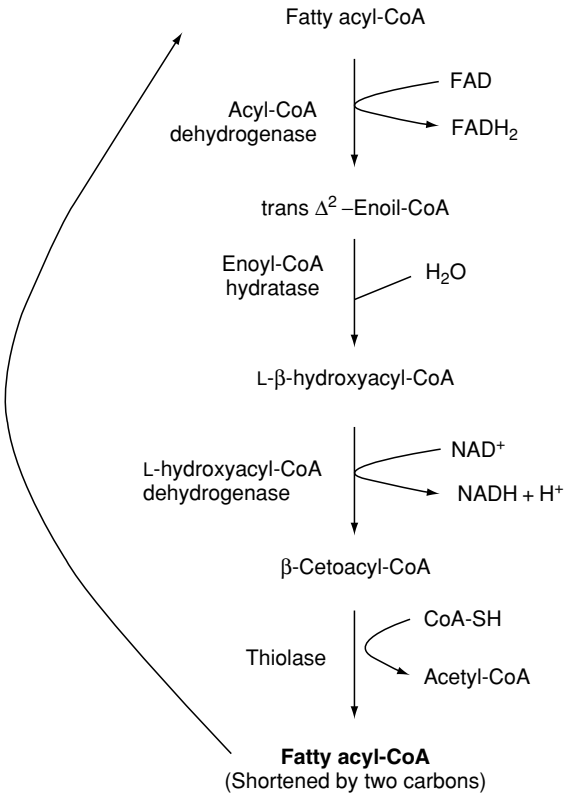

Figure 2 β-Oxidation of saturated fatty acids.

to form a *trans* double bond. During the oxidation of the fatty acid, the FAD is reduced to $FADH_2$, and the electrons are transfered to an electron transfer flavoprotein (ETF). Reduced ETF is reoxidized by a specific oxidoreductase that donates electrons to the electron transport chain at the level of coenzyme Q. The next past is a hydrogenation reaction, catalyzed by the ligase enoyl-CoA hydratase that saturates the new double bond by the addition of a hydroxyl group in the beta position. At least three enoyl-CoA hydratase activities are known. These enzymes present stereospecific and specifically converted *trans*-enoyl-CoA derivates to L-β-hydroxyacyl-CoA. The fourth reaction cleaves the β-keto ester, yielding one molecule of acetyl-CoA and a fatty acyl-CoA shortened in two carbons.

The repetitions of this cycle with the shortened fatty acyl-CoA achieve the complete β-oxidation of fatty acid. At the final step of this cycle, a fatty acid with an even number of carbons yields two molecules of acetyl-CoA.

Fatty acids with an odd number of carbon atoms are rare in mammals but are common in plants and marine organisms. Humans and animals, whose diets include these food sources, metabolize odd-carbon fatty acids via the β-oxidation pathway. The final product of β-oxidation in this case is the 3-carbon propionyl-CoA instead of acetyl-CoA. Three specialized enzymes then carry out the reaction that converts propionyl-CoA to succinyl-CoA, a TCA cycle intermediate that may be oxidized to CO_2.

The oxidation of *polyunsaturated fatty acids* involves the action of two additional mitochondrial enzymes NADPH (nicotinamide adenine dinucleotide phosphate)-linked enoyl-CoA isomerase and $\Delta^2\Delta^4$–enoyl-CoA reductase. They are needed to handle the *cis*-double bonds of naturally occurring fatty acids.

Other Aspects of Fatty Acid Oxidation

Peroxisomes also have a system for the limited β-oxidation of long-chain fatty acids to short-chain acyl-CoA esters and acetyl-CoA. This short-chain product must be transferred to the mitochondria for further breakdown. Peroxisomal β-oxidation is similar to mitochondrial β-oxidation, except that the initial double-bond formation is catalyzed by an FAD-dependent acyl-CoA oxidase.

In *microsomes*, unsaturated fatty acids are oxidized to form eicosanoids and epoxy and hydroxy-fatty acids. Some of these oxidative steps yield regulatory molecules involved in cell signaling. For example, arachidonic acid released from the *sn*-2 position of the membrane phospholipids by the action of phospholipase A_2 provides a substrate for cyclooxygenaxes, lipoxygenase, or monooxygenase.Cyclooxygenase and lipoxygenase products of arachidonic acid (20:4, *n*-6) produce eicosanoids as prostaglandins, thromboxanes, and leukotrienes. These eicosanoids play a critical role coordinating physiological interactions among cells and have an influence in a wide variety of functions, including those of the central nervous system and the contraction of smooth muscles. They also inhibit the mobilization of fatty acids from adipose tissue, have an antiinflamatory effect and modify the blood pressure, aggregation of blood platelets, and cardiac function.

Branched-chain fatty acids are oxidized by α-oxidation pathway. The CoA ester of this metabolic pathway can undergo β-oxidation in the normal way.

Regulation of Fatty Acid Oxidation

Acetyl-CoA formed by mitochondrial oxidation can be oxidized by the citrate cycle. The complete oxidation of fatty acids in peripheral tissues is therefore a major source of metabolic energy and heat in humans. In liver, most of the acetyl-CoA formed by β-oxidation is converted to ketone bodies (acetoacetate and 3-hydroxybutyrate), which are exported as metabolic fuels for extrahepatic tissues. Peroxisomal β-oxidation in the liver is less pronounced than in mitochondria, but it has a key role in metabolizing some very-long-chain fatty acids, although it is

not directly linked to ATP synthesis. The rate of β-oxidation is adjusted to physiological needs. This sequence of events has two potential sites of regulation, (a) transport of fatty acids into the cell, and (b) enzymatic regulation.

Fatty acid transport Long-chain fatty acids can be transferred passively across mammalian cell membranes. A simple diffusion process becomes significant at higher molar ratios of fatty acid to albumin as the concentration of free fatty acid in solution is increased. However, under physiological conditions of low fatty acid to albumin ratios in the circulation, the major fraction of uptake appears to be mediated by a protein component. Therefore, the plasma-membrane-associated and cytoplasmic fatty-acid-binding proteins are involved in cellular fatty acid uptake, transport, and metabolism in tissues. These binding proteins may also act modulate the metabolism of long-chain fatty acids implicated in the regulation of cell growth and other cellular functions.

An acute increase in fatty acid concentration causes a corresponding increase in fatty acid uptake without necessarily increasing fatty acid oxidation. Thus, fatty acid oxidation is more likely to be regulated by factors inside the cell than by limitations in transport.

The FABP exists within both the cytosol and nucleus, which suggests that fatty acids or fatty acid-CoA may be in the nucleus and act as ligands that regulate the activity of specific transcription factors.

Enzymatic regulation When plasma FFAs enter the cells, the first important step in their oxidation is the formation of fatty-acyl-CoA. The enzymatic regulation of CPT-I will then determine whether the fatty-acyl-CoAs will be transferred into the mitochondria for subsequent oxidation.

The reaction catalyzed by CPT-I has been considered to be rate-limiting. CPT-I is subject to regulation at the transcriptional level and to acute control by malonyl-CoA. Thus, the N-terminal domain of CPT-I is essential for malonyl-CoA inhibition.

Malonyl-CoA, which is synthesized by acteyl-CoA carboxylase, is an essential metabolic intermediate in the regulation of fatty acid oxidation. A decrease in malonyl-CoA level results in an increase in CPT-I mediated fatty acid uptake into the mitochondria. As fluctuations in tissue malonyl-CoA content are parallel with changes in acetyl-CoA carboxylase activity, which in turn is under the control of 5′-AMP-activated protein kinase, the CPT-I/malonyl-CoA system is part of a fuel-sensing gauge, turning fatty acid oxidation on and off, depending on the tissue's energy demand.

No allosteric regulation or control of their activities by covalent modification of the other enzymes of the β-oxidation pathway appears to have been described, and it is now recognized that flux control of flux through a pathway is shared between all the enzymes involved, although their individual flux-control coefficients may be very different.

Fatty Acid Biosynthesis. Lipogenesis

Long-chain fatty acids are essential constituents of membrane lipids and are important substrates for energy metabolism of the cell. When total energy intake is in excess of the energy used, fatty acids and triacylglycerol are synthesized.

The synthesis of fatty acids *de novo* is achieved by the sequential condensation of two-carbon units derived initially from acetyl-CoA.

Palmitate, the most abundant acid, is synthesized *de novo* from acetyl-CoA, by acetyl-coA carboxylase (ACC) and fatty acid synthase (FAS) in the cytoplasm.

ACC catalyzes the carboxylation of acetyl-CoA to malonyl-CoA in an ATP-dependent manner. The reaction catalyzed by this enzyme is shown in Figure 3.

Two forms of ACC have been identified. The enzyme that is involved in the synthesis of long-chain fatty acids has a molecular weight of 265 kDa, and the other isoform, with a molecular weight of 275–280 kDa, may be involved in the regulation of mitochondrial oxidation of fatty acids.

FAS plays the central role in *de-novo* lipogenesis in mammals, catalyzing all reaction steps in the conversion of acetyl-CoA and malonyl-CoA to palmitate by enzymatic reactions catalyzed by FAS (Figure 4).

The reaction of FAS involves increasing the length of the acetyl group by C_2 units derived from malonyl-CoA in a stepwise and sequential manner. For instance, in the synthesis of palmitate, there are over 40 steps with at least 30 acyl intermediates. A protein known as acyl carrier protein (ACP) with its 4′-phosphopantetheine prosthetic group was identified as the coenzyme that binds all acyl intermediates as thioester derivatives.

Seven enzymes and reactions are involved in the synthesis of palmitate, as shown in Figure 5. These reactions are essentially the same in all organisms. Acetyl and malonyl are introduced as acyl carrier protein conjugates (reactions 1 and 2). Decarboxylation

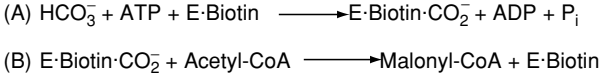

Figure 3 Reactions catalyzed by acetyl-coenzyme A carboxylase.

$$\text{Acetyl-CoA} + 7\ \text{Malonyl CoA} + 14\ \text{NADPH} + 14\ H^+ + 7\ \text{ATP} + H_2O \longrightarrow \text{Palmitic ac.} + 8\ \text{CoA} + 14\ NADP^+ + 7\ \text{ADP} + 7\ P_i$$

Figure 4 Reaction catalyzed by fatty acid synthase.

Enzyme	Reaction
(1) Acetyl transacylase	Acetyl-CoA + ACP-SH ⟶ Acetyl-ACP + CoA-SH
(2) Malonyl transacylase	Malonyl-CoA + ACP-SH ⟶ Malonyl-ACP + CoA-SH
(3) β-Ketoacyl synthase (condensing enzyme)	Acetyl-ACP + synthase-SH ⟷ acetyl-synthase + ACP-SH Acetyl-synthase + Malonyl-ACP ⟶ Acetoacetyl-ACP + CO_2 + synthase-SH
(4) β-Ketoacyl reductase	Acetoacetyl-ACP + NADPH + H^+ ⟷ D-Hydroxybutyryl-ACP + $NADP^+$
(5) β-Hydroxyacyl dehydratase	D-Hydroxybutyryl-ACP ⟷ Crotonyl-ACP + H_2O
(6) Enoyl reductase	Crotonyl-ACP + NADPH + H^+ ⟶ Butyryl-ACP + $NADP^+$
(7) Thioesterase	Palmitoyl-ACP + H_2O ⟶ Palmitate + ACP-SH

Figure 5 Reactions and component enzymes of fatty acid synthase.

drives the β-ketoacyl-ACP synthase and results in the addition of two carbon units to the growing chain (reaction 3). The β-ketoacyl derivative is reduced in three consecutive steps (reactions 4–6) to the saturated acyl derivative, which then acts as a primer for further elongation and reduction cycles to yield ultimately a palmitoyl derivative. The latter is either hydrolyzed to free palmitate (reaction 7), as in animal cells and bacteria, or transferred to CoA-SH to form palmitoyl-CoA, as in yeast, or utilized directly in the triacylglycerol and phospholipid synthesis.

In animal cells, the component enzymes of fatty acid synthesis are covalently linked on a single polypeptide chain. From the N-terminus of FAS to its C-terminus, amino acids encoding β-ketoacyl synthatase, acetyl-CoA transacylase, malonyl-CoA transacylase, dehydratase, enoyl reductase, β-ketoacyl reductase, acyl carrier protein, and thioesterase activities are organized into discrete domains in the order indicated. This is in contrast to bacterial and plant FAS, which are multienzyme complexes consisting of discrete monofunctional enzymes and fungal FAS, which has functional components that are distributed between two nonidentical polypeptides.

The most important feature of animal FAS is its multifunctional character. The animal FAS represents one of the more complex multifunctional polypeptide structures discovered to date because a single polypeptide structure contains all the catalytic components required to direct the 37 reactions that lead to the formation of palmitic acid from acetyl and malonyl-CoA. The functional enzyme is a homodimer of the 250-kDa subunits and contains, in separate domains, the seven different catalytic activities and the site for the prosthetic group, 4′-phosphopantetheine, of the acyl carrier protein (ACP). The multifunctional nature of the subunits and their head-to-tail organization produce a highly efficient enzyme complex capable of carrying out multiple reactions leading to the synthesis of a palmitate molecule from one acteyl and seven malonyl moieties. This multifunctional organization of FAS in vertebrates was probably selected in the course of evolution, because it is more efficient and because the partial activities are more easily regulated in a coordinated fashion. The substrates for each reaction can be channeled to the active center of the subsequent enzymatic step within the same polypeptide, thus providing a high efficiency in long-chain fatty acid synthesis. Moreover, regulating production of the single polypeptide has an effect at all the catalytic steps in a concerted manner.

Additional Elongation

Palmitate is the primary product of the fatty acid synthase, but the cells may synthesize many other fatty acids. Shorter chains can be made easily if the chain is released before it reaches 16 carbons in length. Longer chains are made through special elongation reactions, which occur both in the mitochondria and at the surface of the endoplasmic reticulum. The mitochondrial reactions involve addition and subsequent reduction of acetyl units. The endoplasmic reticulum reactions involve the addition of two-carbon units at the carboxyl end of the chain by means of oxidative decarboxylations involving malonyl-CoA.

Introduction of a Single *Cis* Double Bond

Prokaryotes and eukaryotes are able to introduce an unsaturation (one or more double bond) in a fatty

acid. Double bonds in fatty acids are predominantly of the *cis* (or *Z*) configuration. Many prokaryotes introduce double bonds into fatty acids anaerobically, whereas eukaryotes have adopted an O_2-dependent pathway. So, the advent of an aerobic environment several billion years ago allowed eukaryotes, cyanobacteria, and some bacilli to desaturate the methylene groups of long-chain fatty acids using enzymes called fatty acid desaturases. The transition from anaerobic to aerobic respiration yielded more than an order of magnitude of energy efficiency, leaving a surplus available for processes such as oxidative desaturation. In addition, the ability to regulate membrane fluidity controlling the number of double bonds in fatty acids within the membrane in response to changing temperatures probably conferred a selective advantage to organisms capable of aerobic desaturation.

The anaerobic biosynthesis of unsaturated fatty acids involves dehydration of a hydroxy substrate followed by *trans–cis* isomerization of the double bond. The aerobic reaction can occur anywhere in the fatty acid chain without the need to activate the desired bond toward dehydrogenation. Fatty acid desaturation involves an enzymatic reaction in which a double bond is introduced into an acyl chain, and a molecule of oxygen is completely reduced to water. Two classes of fatty acid desaturase enzymes capable of converting saturated to monounsaturated fatty acids are known, one soluble and the other membrane-bound. Desaturase enzymes need the presence of NADH and oxygen, and two other proteins, cytochrome b_5 and cytochrome b_5 reductase, are involved.

Mammals require polyunsaturated fatty acids but must acquire them in their diet. As such, they are referred to as essential fatty acids (linoleic and linolenic acid). However, mammals can add additional double bonds to unsaturated fatty acids from their diets. Three major groups of dietary fatty acids are oleic acid, linoleic acid, and linolenic acid. These fatty acids serve as precursors for the biosynthesis of a long-chain polyunsaturated fatty acid with carbon lengths of C20/C22. These biosynthetic pathways consist of a series of desaturation and chain elongation steps. Desaturation occurs only toward the carboxylic end of fatty acids in mammalian tissues, and there is no direct crossover among unsaturated fatty acids from one group to another.

Regulatory Control of Fatty Acid Biosynthesis

Variations in the composition and quantity of the diet and rates of *de novo* fatty acid and triacylglycerol synthesis are tightly regulated by nutritional and hormonal status to meet the body's requirement for lipid and energy storage.

The activity of the lipogenic pathway is dependent on nutritional conditions, both in the liver and in the adipose tissue. Consumption of a diet rich in carbohydrates stimulates the lipogenic pathway, whereas starvation or consumption of a diet rich in lipids and poor in carbohydrates decreases its function. Flux through the lipogenic pathway depends on the availability both of lipogenic substrates and of cofactors such as NADPH produced by the pentose phosphate pathway. Regulation of the activity of the lipogenic enzymes involves both short-term and long-term mechanisms.

ACC is the rate-limiting enzyme in the biosynthesis of fatty acids. The ACC activity and the rate of fatty acid synthesis fluctuate rapidly in response to various factors such as hormonal, dietary, developmental, and genetic factors. In the regulation of ACC, short-term changes in the enzymatic activity of ACC are through allosteric regulation, phosphorylation/dephosphorylation events and polymerization/depolymerization mechanisms linked to the presence of specific metabolites as Ca. However, long-term regulation mechanisms, particularly those acting at the gene level of ACC, are controlled at a transcription level and involve changes in the concentrations of regulated enzymes, for example, those observed in animals being fasted and then fed.

The expression of the FAS gene is also highly dependent on the nutritional conditions in liver and adipose tissue. Expression of the gene is barely detectable in starved animals and is stimulated by refeeding with a high-carbohydrate, fat-free diet.

Thus, the amount of carbohydrate in the diet appears to be instrumental in turning on transcription of genes coding for lipogenic enzymes in liver and adipose tissue. Rates of expression of the genes for FAS and ACC are stimulated by glucose in adipose tissue *in vivo* and *in vitro*. Expression of FAS and ACC is stimulated in the liver and of ACC in a pancreatic β-cell line. In each case, regulation involves stimulation of transcription, and to stimulate gene transcription, glucose must be metabolized. To date, the role of insulin, either direct or indirect, is still disputed. Indeed, insulin is reported to restore lipogenic enzyme gene expression in diabetic rats, and glucagon (via cAMP) inhibits expression of the genes for lipogenic enzymes. In adipose tissue, insulin increases indirectly the expression of FAS and ACC by stimulating glucose transport.

Although normal tissues have low levels of fatty acid synthesis, a number of recent studies have demonstrated high levels of FAS expression in a wide variety of human cancers, including carcinoma of the colon, prostate, ovary, endometrium, and breast. The widespread expression of FAS in human

cancer suggests that fatty acid synthesis helps promote tumor growth. This is in marked contrast to its role as an anabolic energy storage pathway in liver and adipose tissue, and fatty acid synthesis is now associated with clinically aggressive tumor behavior and tumor-cell growth and survival, and has become a novel target pathway for chemotherapy development.

See also: **Adipose Tissue**: Structure and Function of White Adipose Tissue; Structure and Function of Brown Adipose Tissue; **Fats**: Digestion, Absorption, and Transport; Requirements; Classification; **Fatty Acids**: Properties; Metabolism; Gamma-linolenic Acid; Analysis; Dietary Importance; **Hormones**: Pancreatic Hormones; Steroid Hormones; **Triglycerides**: Structures and Properties; Characterization and Determination

Further Reading

Chirala SS, Jayakumar A, Gu A and Wakil SJ (2001) Human fatty acid synthase: role of interdomain in the formation of catalytically active synthase dimer. *Proceedings of the National Academy of Sciences, USA* 98: 3104–3108.

Coleman RA, Lewin T and Muoio DM (2000) Physiological and nutritional regulation of enzymes of triacylglycerol synthesis. *Annual Review of Nutrition* 20: 77–103.

Dutta-Roy AK (2000) Cellular uptake of long-chain fatty acids: role of membrane-associated fatty-acid-binding/transport proteins. *Cellular and Molecular Life Science* 57: 1360–1372.

Eaton S, Barlett K and Porfarzman M (1996) Mammalian mitochondrial β-oxidation. *Biochemical Journal* 320: 345–357.

Girard J (1997) Mechanisms by which carbohydrates regulate expression of genes for glycolytic and lipogenic enzymes. *Annual Review of Nutrition* 17: 325–352.

Holm C, Osterlund T, Laurell H and Contreras JA (2000) Molecular mechanisms regulating hormone-sensitive lipase and lipolysis. *Annual Review of Nutrition* 19: 365–393.

Hoppel CL (1976) Carnitine palmitoyltransferase and transport of fatty acids. In: Martonosi A (ed.) *The Enzymes of Biological Membranes*, vol. 2, pp. 119–143. New York: Plenum.

Jump DB (1999) Regulation of gene expression by dietary fat. *Annual Review of Nutrition* 19: 63–90.

Kerner J and Hoppel C (2000) Fatty acids import into mitochondria. *Biochimica et Biophysica Acta* 1486: 1–17.

Kim K (1997) Regulation of mammalian acetyl-coenzyme A carboxylase. *Annual Review of Nutrition* 17: 77–99.

Kuhajda FP (2000) Fatty-acid synthase and human cancer: new perspectives on its role in tumor biology. *Nutrition* 16: 202–208.

Rasmussen BB and Wolfe RR (1999) Regulation of fatty acid oxidation in skeletal muscle. *Annual Review of Nutrition* 19: 463–484.

Shanklin J and Cahoon EB (1998) Desaturation and related modifications of fatty acids. *Annual Review of Plant Molecular Biology* 49: 611–641.

Smith S (1994) The animal fatty acid synthase: one gene, one polypeptide, seven enzymes. *FASEB Journal* 8: 1248–1259.

Sprangers F, Romijn JA, Endert E, Ackermans MT and Sauerwein HP (2001) The role of free fatty acids (FFA) in the regulation of intrahepatic fluxes of glucose and glycogen metabolism during short-term starvation in healthy volunteers. *Clinical Nutrition* 20: 177–179.

Stanley H and Sherratt A (1994) Introduction: the regulation of fatty acid oxidation in cells. *Biochemical Society Transactions* 22: 421–422.

Storch J and Thumser AEA (2000) The fatty acid transport function of fatty acid-binding proteins. *Biochimica et Biophysica Acta* 1486: 28–44.

Subrahmanyamm SC, Jayakumar A, Gu A and Wakil SJ (2001) Human fatty acid synthase: role of interdomain in the formation of catalytically active synthase dimer. *Proceedings of the National Academy of Sciences, USA* 98: 3104–3108.

Sul H and Wang D (1998) Nutritional and hormonal regulation of enzymes in fat synthesis: studies of fatty acid synthase and mitochondrial glycerol-3-phosphate acyltransferase gene transcription. *Annual Review of Nutrition* 18: 331–351.

Towle HC and Kaytor EN (1997) Regulation of the expression of lipogenic enzyme genes by carbohydrate. *Annual Review of Nutrition* 17: 405–433.

Wakil SJ (1989) Fatty acid synthase, a proficient multifunctional enzyme. *Biochemistry* 28: 4523–4530.

Gamma-linolenic Acid

F D Gunstone, Scottish Crop Research Institute, Invergowrie, Dundee, UK

Background

Linolenic acid (GLA) is an important member of the *n*-6 family of polyunsaturated fatty acids, being an intermediate in the bioconversion of linoleic acid to arachidonic acid. Since the formation of GLA is rate-determining in this sequence of changes there are several circumstances when it might be desirable to add it to the diet. GLA is available in seed oils of evening primrose, borage, and blackcurrant, and methods of raising the concentration of this acid in these sources have been described.

Structure

γ-Linolenic acid (GLA) is all-*cis*-6,9,12-octadecatrienoic acid and may also be designated as 18:3 (*n*-6).

This later term indicates an acid with 18 carbon atoms and three double bonds starting on the sixth carbon from the methyl group. It is understood that, in the absence of other indicators, the double bonds have a *cis* configuration and are methylene-interrupted, i.e., the unsaturated centers are separated from each other by one CH_2 group.

$$CH_3CH_2CH=\overset{cis}{C}HCH_2CH=\overset{cis}{C}HCH_2CH=\overset{cis}{C}H(CH_2)_7COOH.$$

There are two important families of polyunsaturated fatty acids: the *n*-6 family based on linoleic acid, and the *n*-3 family, based on linolenic acid. Since animals – including humans – cannot produce linoleic or linolenic acids themselves, these acids are essential parts of the diet, obtainable from vegetable sources or from animals that have already derived these acids from a plant source. Once ingested, the two C18 acids can be metabolized to important C20 and C22 acids through a series of changes involving desaturation and elongation. The same enzymes are required for both families of acids, and there is competition for these. The most important acids in these sequences are arachidonic acid (20:4) in the *n*-6 series, and eicosapentaenoic acid (*n*-3 20:5) and docosahexaenoic acid (22:6) in the *n*-3 series. These acids are important components of phospholipid membranes, and the two C20 acids are also precursors of prostaglandins and other eicosanoids.

GLA is the first intermediate in the metabolic conversion of linoleic acid to arachidonic acid. The change involves introduction of a (third) double bond at the Δ6 position under the catalytic influence of the 6 desaturase enzyme. This step is believed to be the rate-limiting stage in the metabolic pathway. For humans in good health, this presents no problem, but there are some conditions such as aging, smoking, diabetes, high alcohol intake, stress-related hormones, and viral infections that inhibit the rate-determining step and where it is considered wise to supplement the intake of GLA. Nutritional factors can also influence this metabolic pathway.

Table 1 shows the series of changes by which linoleic and linolenic acids are converted to the long-chain polyunsaturated fatty acids.

Occurrence of GLA in Seed Oils and Microorganisms

As an intermediate in the conversion of linoleic acid to arachidonic acid, it is not surprising that GLA is found at low levels in many animal fats. It is present in cow milk fat (~0.1%) and human milk fat (0.35–1.0%), and its presence in the latter has often been cited as an indication of its importance. However, richer sources of this acid are found in the seed oils of some less common plants and are also produced by some microorganisms.

There are three commercial sources of GLA from seed oils: evening primrose (*Oenothera biennis* and *O. lamarckiana*), borage (*Borago officinalis*), and blackcurrant (*Ribes nigrum*). Typical fatty acid profiles for these three oils are given in Table 2. Ucciani found GLA in 164 higher plant families belonging to 10 botanical families.

Evening primrose oil was the first to be developed and used as a source of GLA. It contains 6–14% (generally 9–10%) of this acid and is accompanied by high levels of linoleic acid. Borage (starflower) is a richer source of GLA in terms of oil content of the seed and level of GLA in the oil, but it has less linoleic acid than evening primrose oil and also contains a

Table 1 *n*-6 and *n*-3 families of polyunsaturated fatty acids

n-6	n-3
18:2 (9,12) Linoleic	18:3 (9,12,15) α-Linolenic
↓ a	↓ a
18:3 (6,9,12) γ-Linolenic	18:4 (6,9,12,15) Stearidonic
↓ b	↓ b
20:3 (8,11,14)	20:4 (8,11,14,17)
↓ c	↓ c
20:4 (5,8,11,14) Arachidonic	20:5 (5,8,11,14,17) Eicosapentaenoic
	↓ b
	22:5 (7,10,13,16,19)
	↓ b
	24:5 (9,12,15,18,21)
	↓ a
	24:6 (6,9,12,15,18,21)
	↓ d
	22:6 (4,7,10,13,16,19) Docosahexaenoic

a, 6-desaturase; b, elongase; c, 5-desaturase; d, β-oxidation.

Table 2 Fatty acid composition of evening primrose oil, borage oil, and blackcurrant oil and of oils from *Mortierella isabellina* and *Mucor javanicus*

	epo[a]	*bor*[a,b]	*bck*[a]	*Mortierella*	*Mucor*
Oil (%)	16–26	27–35	22–26		
16:0	7–10	10	6–7	27	22–25
18:0	1.5–3.5	4	1.5	6	5–8
18:1	6–11	18	9–11	44	38–41
18:2	65–80	37	46–49	12	10–12
18:3 (*n*-6)	8–14[c]	22	14–16	8	15–18
18:3 (*n*-3)			13–14		
18:4 (*n*-3)			2.5–3		

[a]epo, evening primrose oil; bor, borage oil; bck, blackcurrant seed oil.
[b]Also 20:1 4%, 22:1 2.5%, 24:1 1%.
[c]Generally 9–10%.

series of *n*-9 higher monoene acids viz. 20:1, 22:1, and 24:1 at a combined level of about 8%. Blackcurrant oil is intermediate between the other two sources in its level of GLA, but it also contains the 18:3 and 18:4 *n*-3 acids.

Molds belonging to the Mucorales family produce GLA as the only 18:3 acid, and commercial processes have been developed in Japan based on *Mortierella isabelina* and in the UK based on *Mucor javanicus*. UK production has been discontinued following a change in ownership of the producer company. These products differ from the seed oils in their higher levels of saturated and monoene acids and their lower level of linoleic acid. This makes the oil more suitable as a source of pure GLA or of upgraded triacylglycerols, since the necessary separations are easier to effect.

The blue–green algae *Spirulina platensis* (18–21% GLA) and *S. maxima* (12% GLA) are also potential sources of GLA. However, this is present in various galactosyldiacylglycerols or phosphatidic acids rather than in triacylglycerols.

The distribution of these fatty acid chains between the *sn*-1, 2, and 3 positions has been examined in a number of ways. Results for GLA are summarized in Table 3.

Because of the high level of linoleic acid in evening primrose oil the major triacylglycerols contain two (or three) linoleic acyl chains: LLL 54.3 mol%, LLG 17.6%, LLO 13.7%, LLP 7.9%, and other 6.5% (L = linoleic, G = γ-linolenic, O = oleic, and P = palmitic acyl group). The triacylglycerol composition of borage oil is more complex, with 10 molecular species each exceeding 4 mol%: LLG 15.4, OLG 12.3, LLL 10.1, LGG 9.8, PLG 9.5, PLL/OOG 8.9, OLL 8.2, SLG/POG 5.0, LLE 4.1, OOL 4.1, and other 12.6 (symbols as above and E = 20:1, S = stearic and). No sterochemistry is implied in these symbols, and PLG, for example, represents the sum of six isomers.

Evening primrose oil grows in many countries in Europe and also in New Zealand, but for commercial reasons, the crop is now grown mainly in China (75–80%) and Eastern Europe. Seed is produced at a level of 1000–1200 kg ha^{-1}. Annual production is estimated to be about 10 000–12 000 tonnes of seed yielding 1300–1500 tonnes of oil. Over 10 years to 1999, the price of the oil fell from $45 to $17.5.

Britain is the main source of borage (starflower), but it is also grown in Holland, Canada, New Zealand, and Poland. Annual production of seed is estimated at 3000–4000 tonnes, which provides about 800 tonnes of oil. Borage seed is $3000–4000 per tonne. The price of the oil fell from $60 to $35 per kilogram over the 10 years to 1999.

Blackcurrant seed oil is a byproduct of the production of blackcurrant juice and jelly, and production is estimated to be perhaps 50–100 tonnes per annum.

These oils are most often supplied in capsules with added tocopherol and ascorbyl palmitate to serve as antioxidants. The products have a shelf-life of about 3 years.

Table 3 Distribution of GLA between the *sn*-1, 2, and 3 positions in evening primrose, borage, and blackcurrant seed oils

	epo[a]	*epo*[b]	*bor*[b]	*bck*[b]
Tag	9.6	9.3	24.8	15.9
sn-1	7.2	3.6	4.0	4.1
sn-2	10.7	10.7	40.4	17.4
sn-3	10.9	13.5	30.1	25.8

For abbreviations, see Table 2.
[a]From Laakso P and Christie WW (1990) *Lipids* 25: 349–353.
[b]From Lawson LD and Hughes BG (1988) *Lipids* 23: 313–317.

Sources with Higher Levels of GLA

It is possible to raise the level of GLA in these oils by selective enzymic reaction, and enhanced evening primrose and borage oils, with twice the normal level of GLA, have been produced and offered for sale. These materials are mainly triacylglycerols but may contain up to 20% of diacylglycerols. Procedures are based on the fact that some enzymes discriminate against acids having a double bond close to the acyl function and GLA with Δ6 unsaturation can be distinguished from the more common unsaturated acids with Δ9 unsaturation (oleic and linoleic). This differentiation is apparent in hydrolysis, esterification, acidolysis, and alcoholysis reactions promoted by appropriate lipases. The following examples are typical.

- Borage oil (22% GLA) is hydrolyzed completely (*Pseudomonas* species, 35 °C, 24 h), and the acids are selectively esterified with lauric alcohol in the presence of *Rhizopus delemar* lipase at 30 ° for 20 h. The latter enzyme discriminates against GLA, and this acid is concentrated in the unesterified acids, which finally contain 70% GLA (74% recovery). If the esterification is repeated, 94% GLA is obtained with 68% recovery (*C. rugosa*, 35 °C, 15 h).
- Selective hydrolysis of borage oil (22% GLA) with *C. rugosa* lipase at 35 °C for 15 h gives glycerol esters with 46% of this acid that can hardly be raised above this value. After removal of free acids, a second hydrolysis raises the concentration to 54% with 76% recovery. On a pilot-plant scale, 7 kg of borage oil gives 1.5 kg of upgraded oil (56%

GLA) after removal of free acid by molecular distillation. Starting with an already upgraded borage oil (10 kg, 45% GLA), high-quality GLA (2.1 kg, 98% pure, 49% recovery) is obtained by hydrolysis and two separate partial esterifications.

There are several reports of plants such as rapeseed/canola genetically modified to produce GLA, but commercial products are not yet available. One example that has been cited has the following fatty acid composition: 16:0 4.2%, 18:0 3.7%, 9–18:1 21.6%, 11–18:1 3.4%, 18:2 26.0%, 6,9,12–18:3 37.0%, 9,12,15–18:3 1.3%, and other 2.8%.

Nutritional and Medical Uses

The major applications for the oils are in the area of healthfood supplements, but markets have also been developed in infant nutrition, pet food, and cosmetics, and the total market for these oils is around $50 million per year.

Available evidence indicates that only around 5–10% of the daily intake of linoleic acid can be converted to GLA and beyond. For a 60-kg adult with a dietary intake of 5–20 g day^{-1} of linoleic acid, the endogenous rate of formation of GLA will be 250–1000 mg day^{-1} or around 4–17 mg kg^{-1} day^{-1}.

Human breast milk contains 100–400 mg l^{-1} of GLA + DGLA (20:3 *n*-6). A 5 kg baby consuming 1 l of milk per day receives 20–80 mg kg^{-1} day^{-1} of these two acids combined. DGLA is converted to two main eicosanoids: PGE_1 and 15-hydroxy DGLA. PGE_1 (prostaglandin E_1) is antiinflammatory, antithrombotic, a vasodilator, and stimulates the formation of cyclic adenosine monophosphate, which inhibits phospholipase A_2, whereas 5-hydroxy DGLA is a potent natural antiinflammatory agent inhibiting the conversion of AA (arachidonic acid) to its 5- and 12-lipoxygenase metabolites. GLA and DGLA are considered to be useful in the treatment of diabetes, atopic eczema, inflammation, stress, cardiovascular disease, and cancer.

Soft gelatin capsules containing evening primrose oil have been sold as nutritional supplements for more than 20 years and of borage oil (also marketed as starflower oil) for more than 10 years. Samples may be blended with other oils (fish, linseed) or have added vitamins, minerals, or herbal extracts. Nutritional supplements probably account for 80% of the total usage of GLA oils.

GLA has been used to treat a wide range of conditions including: atopic eczema, dermatitis and other inflammatory skin conditions, diabetic neuropathy, breast pain, premenstrual syndrome, high blood pressure, and cancer, but not all of these claims have been confirmed.

Doses have not been fully defined, but the following levels have been recommended (levels refer to GLA and not to GLA-containing oils): nutritional purposes (25–50 mg day^{-1}), therapeutic use (100–500 mg day^{-1}), and pharmacological effect (500–2000 mg day^{-1}).

Two products have been licensed as pharmaceuticals Efamast™ for relief of cyclical mastalgia and Epogam™ for treatment of atopic eczema. There are also a range of preterm and term infant formulations on the market (particularly in Europe) containing GLA from various sources. Levels of GLA in infant formula range from 0 to 0.9%. Many skin-care products (creams, lotions, soaps) contain GLA oils at levels between 0.1 and 2.0%.

See also: **Essential Fatty Acids**; **Fatty Acids**: Metabolism; Dietary Importance; **Vegetable Oils**: Types and Properties; Oil Production and Processing; Composition and Analysis; Dietary Importance

Further Reading

Christie WW (1999) The analysis of evening primrose oil. *Industrial Crops and Products* 10: 73–83.

Clough P (2001) Specialty vegetable oils containing gamma linolenic acid and stearidonic acid. In: Gunstone FD (ed.) *Structured and Modified Lipids*. New York: Marcel Dekker.

Gunstone FD (1992) Gamma linolenic acid – occurrence and physical and chemical properties. *Progress in Lipid Research* 31: 145–161.

Horrobin DF (1992) Nutritional and medical importance of gamma linolenic acid. I. *Progress in Lipid Research* 31: 163–194.

Huang Y-S and Mills DE (1996) *γ-Linolenic Acid Metabolism and its Roles in Nutrition and Medicine*. Champaign, IL: AOCS Press.

Huang Y-S and Ziboh A (2001) *Recent Advances in Biotechnology and Clinical Applications of Gamma Linolenic Acid*. Champaign, IL: AOCS Press.

Ucciani E (1995) Sources potentielles d'acide gamma-linolenique: une revue. *Oleagineux Corps gras Lipides* 2: 319–322.

Analysis

E W Hammond, Greenisle-Consulting, Warkton, Kettering, UK

Background

Fatty acids occur in food lipids mainly as esters of glycerol. Fatty acids are represented in food lipids (typical ranges shown in brackets) by

triacylglycerides (> 90%), diacylglycerides and monoacyglycerides (in the range 0.5–4%), other polar lipids such as lecithin (0.5–5%), phytosterol esters (up to 1%), and free fatty acids (up to 1%).

Extraction of Lipids from Foods

The extraction of lipids from foods in order to analyze their fatty acids is often difficult and specialized. For example, foods may be essentially anhydrous (e.g., biscuits, wheat flour) or wet (e.g., fresh meats, fish). Each food type may require very different handling to obtain the total lipids present. The fat in biscuits can be obtained just by suspending the crushed samples in light petroleum. However, to obtain the total lipids from flour requires rigorous extraction with solvents such as *n*-butanol. Samples containing water in any quantity require dehydrating first, which process might change the lipid materials, or extraction with alcohol based solvent mixtures. The references included in the bibliography cover a range of techniques for specific materials.

Determination of Total Free Fatty Acids

Free fatty acids (FFA) occur in refined edible oils at less than 0.1% (w/w) and in crude oils typically from 1 to 15%. FFA at such levels are present because of lipid hydrolysis and not as a normal natural component of vegetable or animal lipids. The high levels sometimes found in crude oils are as a consequence of tissue damage leading to the release of natural lipases. Tropical fruits such as palm are particularly prone to high FFA levels in crude oils. (*See* **Phospholipids**: Determination; **Triglycerides**: Characterization and Determination; **Vegetable Oils**: Composition and Analysis.)

The normal method of determination is titration (AOAC28:029–28:034) with a solution of potassium hydroxide [approximately 0.1 M in 95% (v/v) ethanol]. This procedure is applicable to all oils and fats that are soluble in the solvent mixture 1/1 (v/v) of ethanol 95% (v/v) and diethyl ether. The titrimetric procedure uses an indicator solution which is phenolphthalein [10 g l^{-1} in ethanol 95% (v/v)], for determining the end point. In the case of fats, which give colored solutions, a potentiometric technique should be used to determine the end point. The result of the titration is the average of two duplicate determinations and is expressed as one of the following:

1. **Acid value (AV)**: the number of milligrams of potassium hydroxide required to neutralize 1 g of the fat. This is given by the formula:

$$\text{A.V.} = (56.1PV)/m,$$

where V = number of milliliters of potassium hydroxide; P = exact molarity of potassium hydroxide; and m = mass (g) of the test portion.

2. **FFA%**: the acidity in per cent given by the formula:

$$\text{Acidity} = (PVM_r)/(10m),$$

where V = number of milliliters of potassium hydroxide; P = exact molarity of potassium hydroxide; M_r = relative molecular mass (see Table 1); and m = mass (g) of the test portion.

When determining FFA%, it is conventional to express the value as 'oleic acid.' However, where the type of fat is known, the relative molecular mass should be used for improved accuracy. This can be calculated from the average molecular weight of fatty acids, determined from the analysis of fatty acids by gas chromatography. Otherwise, the values for M_r, as shown in Table 1, should be used.

To maintain precision, the size of sample to be used in the determination will depend on the expected level of FFA and should follow the data in Table 2. (*See* **Chromatography**: Gas Chromatography.)

Method: Total FFA by Titration

1. A quantity of the solvent mixture (1/1 ethanol/diethyl ether) should be neutralized, just prior to use, by dropwise addition of 0.1 M potassium hydroxide solution after adding the phenolphthalein indicator at the rate of 0.5 ml per liter of solvent.
2. The quantity of sample required is determined by reference to Table 2. Duplicate samples are weighed out accordingly, paying attention to accuracy. The weights are recorded. The samples are dissolved in 50–150 ml of the solvent, each.

Table 1 Total free fatty acids

Type of fat	*Expressed as*	M_r
Coconut, palm kernel	Lauric acid	200
Palm oil	Palmitic acid	256
All other oils	Oleic acid	282

Table 2 Sample mass required for determination of FFA

Expected acid value	*Percentage FFA*	*Mass of test portion (g)*	*Accuracy of weighing (g)*
< 1	(< 1)	20	0.05
1–4	(< 2)	10	0.02
4–70	(2–10)	2.0	0.01
20–70	(10–40)	0.5	0.001
> 70	(> 40)	0.1	0.0002

3. While the solution is stirred continuously, titration with 0.1 M potassium hydroxide should proceed to the end point. The end point is a pink color, which should persist for at least 10 s. The volume of titrant is recorded.

Notes:

1. If the FFA is very low (< 0.2%), atmospheric carbon dioxide may interfere significantly. It is useful to replace the air in the titration flask with nitrogen.
2. If the solution becomes cloudy during titration, the volume of neutralized solvent may be increased. Warming should be avoided.
3. If the quantity of hydroxide required for titration exceeds 10 ml, a solution of 0.5 M may be used.

Determination of Individual Free Fatty Acids

Certain situations arise where it may be advantageous to determine a single fatty acid or the distribution of fatty acids in the FFA fraction of a fat or oil. In this case, titration is of little use, since it does not discriminate between the different fatty acids. The quickest method is to use gas chromatography with an added internal standard. In this method, quantitation of both individual and total FFA is obtained and can be measured down to 0.001% with confidence.

It is possible to analyze free fatty acids directly by capillary gas–liquid chromatography (GLC) using special deactivated acidic phases, such as Restek Stabilwax-DA (Stabilwax-DA crossbond fused silica capillary column; Thames Restek UK Ltd., Berkshire, UK). These reduce the tailing effects and nonproportional losses that are a consequence of hydrogen bonding of FFA on the column. Analysis of derivatives is advised to reduce quantitative errors that can result.

Where FFA is to be analyzed directly, it should first be isolated by thin-layer chromatography (TLC) or clean-up column procedures (see below). Prior to this, an internal standard of heptadecanoic acid (C17:0) should be added to the fat at a level consistent with the expected level of FFA present in the fat. (*See* **Chromatography**: Thin-layer Chromatography.)

The FFA, dissolved in toluene (up to 5 mg ml^{-1}, total FFA), is injected (0.5 µl) using the direct on-column technique. The column can be a 30-m length of 0.53-mm ID fused silica, with a 0.25-µm, cross-bonded film suitable for the analysis of underivatized fatty acids (see Figure 1). A 1.0-m length of fully deactivated, blank fused silica should be fitted as a 'retention gap.' Typical conditions for analysis are shown in Figure 1. A full profile, including unsaturated fatty acids, from butyric acid (C4:0) to arachidic (C20:0) acid can be achieved within 30 min.

Where the chain-length distribution of the FFA is wider than C14 to C20, it may be necessary to apply correction factors to the detector response. The flame ionization detector responds to nonoxidized carbon in a linear relationship. However, the carboxyl carbon is oxidized and does not respond; therefore, formic acid gives no response, while acetic shows the response of one carbon, and so on through the series. Since an internal standard is used in this method, it is a simple job to determine the response factors for the FFA of interest relative to this standard.

Analysis of Derivatized FFA

Methods There are often problems in the measurement of FFA by GLC. It is advisable to derivatize for improved accuracy. Again, it will be necessary to use an internal standard of C17:0 and relative response factors. Two approaches may be used: methyl esters of FFA or silyl ethers of FFA.

In the case of derivatives of FFA, one must consider whether they are to be analyzed free (isolated) or in the presence of other lipid classes such as triacylglyceride (TAG) and partial glycerides. If short-chain, volatile FFA is present (e.g., in butterfat), it is best not to have any concentration step. By far the most specific reagent is diazomethane, but care must be exercised in its use, as the reagent is highly toxic and potentially explosive.

The preparation of methyl esters in this way will provide good quantitative data on the FFA portion of the fat only. If silyl ethers are prepared, and the whole fat mixture is analyzed by GLC, total fat information is obtained.

Methyl Esters via Diazomethane An appropriate amount (up to 50 mg) of the isolated FFA is dissolved in diethyl ether (2 ml) containing a few drops of methanol. The diazomethane in diethyl ether solution is prepared, with all manipulations carried out in a fume cupboard. Sufficient diazomethane solution is added to the sample to leave a slight excess of yellow color remaining. The mixture is left for no longer than 5 min; otherwise, artificially high results may occur. Formic acid in methanol (10%) solution is added dropwise to remove the excess reagent. This solution is now ready for analysis without concentration. The type of GLC column used is typically a 30 m × 0.32 mm i.d. fused silica, with a 1.0-m retention gap of fully deactivated fused silica. A polar phase of bonded FFAP at 1.0-µm film thickness is suitable. The sample is injected using the on-column technique and not split injection. The carrier gas may be hydrogen (linear velocity 40 cm s^{-1}) or helium

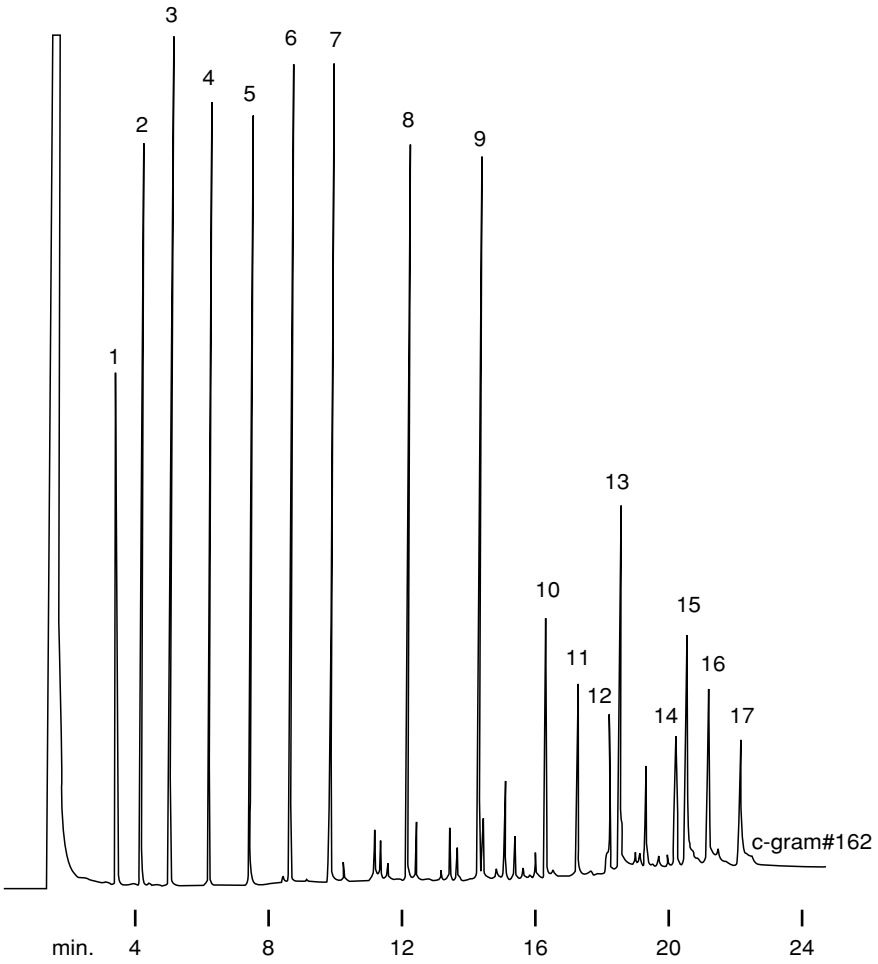

Figure 1 Organic acids and free fatty acids [Stabilwax DA, direct injection; 30 m, 0.53 mm i.d., 0.25 μm Stabilwax-DA (cat. #11025), 0.5 μl direct injection of a 5 mg ml^{-1} standard]. Oven temperature: 100 °C (hold for 2 min) to 250 °C at 8 °C min^{-1}. Injection and detection temperature: 280 °C. Carrier gas: helium. Linear velocity: 40 cm s^{-1} (flow rate: 5.2 cm^3 min^{-1}). FID sensitivity: 8×10^{-11} AFS. 1: C2:0, acetic acid; 2: C3:0, propionic acid; 3: C4:0, butyric acid; 4: C5:0, valeric acid; 5: C6:0, caproic acid; 6: C7:0, enanthic acid; 7: C8:0, caprylic acid; 8: C9:0, capric acid; 9: C10:0, lauric acid; 10: C11:0, myristic acid; 11: C12:0, pentadecanoic acid; 12: C13:0, palmitic acid; 13: C14:0, palmitoleic acid; 14: C15:0, stearic acid; 15: C16:0, oleic acid; 16: C17:0, linoleic acid; 17: C18:1, linolenic acid. Reproduced by permission of Restek Corporation.

(linear velocity 20 cm s^{-1}). The column initial temperature will depend upon the type of sample but typically is 100 °C, held for 5 min. The temperature is then programmed to 200 °C at 5 °C min^{-1}. The elution order and profile will appear very similar to that for FFA above (Figure 1), with the advantage that peak width and retention time will be improved.

***O*-Trimethylsilyl Ethers (*O*TMSi)** The most useful reagents for the derivatization of fats are bis-trimethylsilyl acetamide (BSA) and trimethylsilyl imidazole (TSIM) (Pierce Chemical Co.), with the latter being the stronger reagent, particularly when diacylglycerides (DAG) and monoacylglycerides (MAG) are present. Derivatization is achieved in the neat reagent or in a solution of the lipid in chloroform or tetrahydrofuran (10 mg ml^{-1}). During the procedure, the *anhydrous* solution is warmed at 50 °C for 5 min in a closed vial. In this reaction, all free hydroxyls and carboxyls are derivatized. Normally, 100% volume excess of the reagent over the fat is sufficient, where the level of FFA and partial glycerides is not abnormally high. The minimum level of reagent will be 50–100-fold molar excess over total free hydroxyls and carboxyls. When using chloroform, allowance must be made for the 2% of ethanol stabilizer. The resulting solution is stable for up to 5 h if maintained anhydrous in a closed vial. After this time, it should be discarded. The solution is injected directly 'on column' and *not* via a split technique. For GLC, use a column of fused silica, 0.53 mm i.d. by 10 m in length with a bonded phase of OV1, OV101 (or

equivalent) of film thickness 0.1–0.2 μm. A 1.0-m length of fully deactivated, blank fused silica is fitted as a 'retention gap.' The helium carrier gas is set high at about 30 ml min^{-1} and a start temperature of 50 °C held for 5 min. The temperature is then programed at 5 °C min^{-1} up to 320 °C in a single ramp. The elution order is fatty acids, MAG, DAG, and TAG in order of molecular weight. The method is suitable for neutral fat extracts. In some fats, the level of free sterols is significant, and OTMSi ethers of sterols will appear on the chromatogram.

Fatty Acid Methyl Esters from Triacylglycerides

There are a number of methods used to prepare fatty acid methyl esters (FAME) from TAG to enable their analysis by GC. The boron trifluoride/methanol reagent is widely used (AOAC -28:057). The reagent is toxic and unstable during storage, and has been found to be prone to form artefacts from some oxygenated, cyclic, and some polyunsaturated fatty acids.

Transmethylation techniques are very rapid, but they produce FAME only from glyceride esters and not from any FFA component. It is important, also, that the reagent and fat solutions are anhydrous; otherwise, significant (and even large) amounts of unesterified FFA can result. Of this group, 0.5 M sodium methoxide/methanol is widely used. This reagent is hazardous to prepare and dispose of. Anhydrous potassium hydroxide/methanol (0.2 M) is perhaps a more preferable reagent, being easy to prepare as required, although storage is acceptable provided it is kept dry. It should be noted that if conjugated unsaturated fatty acids are known to be present, they might be unstable in alkaline reagents. These reagents should be stored in bottles with polycarbonate stoppers; glass stoppers will fuse to glass bottles.

For up to 50 mg of sample in 3–5 ml of reagent (with 1.0 ml of toluene added as solubilizer), heated to 50 °C, the reaction time is normally 15 min. At the end of reaction, the sample is allowed to cool, and 5 ml of aqueous acetic acid (5%) is added carefully, followed by up to 10 ml of hexane (dependent upon sample size) such that the ester concentration is about 5 mg ml^{-1} (suitable for direct sampling on to GLC). This mixture needs to be shaken thoroughly and allowed to separate. The lower layer is removed by aspiration and discarded. A 5-ml aliquot of water is added, and the mixture shaken and separated again. Most of the top layer is then transferred to a vial containing about 2 g of anhydrous sodium sulfate, which dehydrates the solution ready for GLC. No concentration step should normally be necessary.

Acid methanolysis reagents methylate most lipid classes including FFA. They require relatively long reaction times but very short processing times, which means that technicians are not tied to one technique for long periods. Since the reagents are acidic, they are not suitable for epoxy fatty acids or other acid-labile fatty acids. Anhydrous hydrochloric acid/methanol (typically 5% or saturated) is suitable and popular. This produces methyl chloride during storage. In addition, while it is easy to use and prepare by bubbling hydrogen chloride gas from a cylinder into anhydrous methanol, this can be hazardous. Alternatively, acetyl chloride (50 ml) can be added to cold (5 °C) anhydrous methanol (450 ml); methyl acetate is formed as a byproduct. However, several workers, when using this latter reagent, have reported yield problems.

Sulfuric acid/methanol with toluene as solubilizer (1:10:20 by volume of sulfuric acid–toluene–methanol) is a convenient reagent. It is easy and safe to prepare, is stable at ambient for long periods, and is easy to use. Sulfuric acid is added to cooled and stirred methanol, followed by the toluene. The solution is stored in a brown stoppered bottle and must be kept dry. Acid-resistant gloves and a face shield should be worn while mixing in the acid. For use, 5 ml of reagent is added to up to 50 mg of sample, which is refluxed for 60 min in a tube. The sample is then diluted with 5 ml of water. A volume of hexane is added (to give an approximate sample concentration of 5 mg ml^{-1}), and the mixture is shaken thoroughly. On separation, the bottom layer is aspirated to waste. A further 5 ml of water is added and the mixture shaken. The top layer is transferred to a vial containing 2 g of anhydrous sodium sulfate. This solution is then ready for analysis by GLC. If short-chain (< C10) fatty acids are present, the toluene may interfere with the GLC analysis. In this case, a reagent made up without the toluene solubilizer should be used, *but note* that an extended reflux (90 min) may be necessary, unless the fat dissolves rapidly.

Chromatographic Methods

Some of the specific points have been covered above and should be considered in combination with this information.

Thin-layer Chromatography (TLC)

FFA is easily separated from other lipid classes on silicic acid TLC plates. This technique is excellent for preparative work. Plates coated with Kieselgel with ('G') or without ('H') calcium sulfate binder can be used. The plates should be activated before use. For speed, individual plates may be placed flat on the

turntable in a microwave oven for 2 min at full power and then cooled. Multiple samples may be spotted in lanes or, for preparative work, the sample may be applied as a continuous streak 1 cm above the base of the plate. The sample area is focused before the main chromatography by developing the plate to just above the sample application line twice in a solvent of diethyl ether. The plate is air-dried between each focus development, and then developed in the main solvent, which, for FFA, is a mixture of diethyl ether/light petroleum 40–60° with added formic acid (proportions 18/82/1 by volume). After development to within 1 cm of the top of the plate, the plate is air-dried and sprayed lightly with a methanolic solution of dichlorofluorescein (0.1%). After drying, the plate is viewed under ultraviolet (UV) light (254–320 nm), which causes most lipid components to fluoresce. Lipids that absorb in the UV (e.g., conjugated unsaturates) appear as violet spots. FFA lies between TAG and DAG with an Rf of approximately 0.6. If there is some confusion over the position of the FFA, a standard can be run in 1-cm lanes at the sides of the plate. Samples that contain FFA with a wide range of chain lengths (C4 to C20, for instance) often give a broad band or even double bands. The longer-chain FFA is less polar than the short-chain FFA and therefore runs slightly higher up the plate. The FFA band should be marked, scraped off into a sinter glass disc filter stick, and eluted off with diethyl ether. Careful concentration yields the dry FFA fraction. If small or volatile fractions are suspected (this is particularly useful for radioactive labeled fractions), they may be stabilized prior to concentration by the addition of a known, small amount of 10% potassium hydroxide in methanol (enough to form the salts). After concentration, the salts are dissolved in a small volume of 1:1 diethyl ether:methanol containing sufficient formic acid to re-form the acids prior to methylation.

High-performance Liquid Chromatography (HPLC)

It is recommended that FFA be analyzed by GLC, but for various reasons, the use of HPLC may be dictated. To overcome the lack of a chromaphore, most workers use UV-absorbing derivatives of FFA. The most popular and successful has been the phenacyl ester. The dansyl piperazide derivative has been applied successfully for fluorescence. (*See* **Chromatography**: High-performance Liquid Chromatography.)

Typical chromatography conditions for phenacyl esters include a 25-cm × 4-mm i.d. column packed with C18 reversed phase material of 5-μm particle size. A solvent of acetonitrile–water (80:20, v/v) is run isocratically at 2.0 ml min^{-1} for the first 30 min. The solvent is then linearly programed to 85:15 (v/v) for another 15 min. There is a considerable overlap of peak compositions, and it is essential that a full set of standard materials is chromatographed to determine retention behavior and times. For instance, myristic acid (C14:0) has an elution time of about 16 min, whereas arachidonic acid (C20:4) elutes around 18 min. Stearic acid (C18:0) has an elution time of about 50 min.

Gas–liquid Chromatography (GLC)

This item has been covered earlier. However, the applications described are designed for capillary GLC. Modern instruments together with capillary or wide bore columns are relatively easy to use and produce high-quality results. However, the methods do not preclude the use of packed columns, although the chromatography of unesterified FFA is less successful. For esterified FFA, packed columns of either 2 or 4 mm i.d. by about 2 m in length should be used, with a carrier gas of nitrogen at 30 ml min^{-1} for a 2-mm column and 60 ml/min for a 4-mm column. For the column packing, a phase of 10% SP2330 on 100/120 mesh Supelcoport should be used, and this should be operated isothermally at 180 °C or programed from 50 to 200 °C at 5 °C min^{-1} after holding for 5 min. The sample solution is then injected directly on-column into a heated injection area set isothermally at 200 °C.

Determination of *Trans* Fatty Acids

Isolated, methylene-interrupted *trans* double bonds show an infrared (IR) band absorbing at about 967 cm^{-1}. This absorption may be used to estimate the *trans* content in edible fats. The AOAC publishes a typical method AOAC 28.052–28.067 for *trans* values (TV). There are some drawbacks with the IR method, which can lead to some inaccuracies. Essentially, when TV is measured on triacylglyceride, the value obtained is some 2 units higher than when the same sample is measured as methyl esters. This problem is worse for samples below 15% *trans*. Also, isomerized or oxidized fats (mainly) exhibit conjugated species that show absorption close to the isolated *trans* bond. This interferes with the correct allocation of baseline and makes low values less easy to define accurately. The following method takes into account these problems and introduces the use of standards and references, although it must be remembered that problems still exist with the determination of absolute TVs.

The method measures the TV on triacylglyceride, but we recommend conversion to methyl esters for values of less than 5% trans. The values are expressed as the percentage of trielaidin compared with a standard curve of trielaidin in tristearin. Confusions in

interpretation exist where there are compounds present that show absorption in the region 970 cm^{-1}. Pure trielaidin and tristearin (99%) are required as a standard and reference, together with a recording double-beam IR spectrometer suitable for quantitation between 1100 and 900 cm^{-1} and a matched pair of cuvettes of path length 1 mm (deviation < 1%) with windows of sodium chloride or potassium bromide. The solvent is carbon disulfide and solutions and measurements are all at 20 °C. The solvent is toxic, and therefore, all manipulations of open solutions should be done in a fume cupboard and appropriate solvent-resistant gloves must be worn. Alternatively, the use of Fourier transform infrared (FTIR) greatly improves the methodology and accuracy.

Method: Determination of *Trans* Value by IR Spectroscopy

1. The spectrometer is set to record in the range 1050–900 cm^{-1}, with a narrow slit and slow recording speed. (*See* **Spectroscopy**: Infrared and Raman.)
2. Solutions are prepared according to Table 3 and made up to exactly 10 ml in volumetric flasks.
3. A cuvette is filled with solution 1 and placed in the reference beam. The absorbances of the other solutions are recorded against this.
4. For each spectral record, a straight baseline is drawn joining the *minima* at 1000 and 925 cm^{-1}. The absorbance of the peak is calculated. If transmission is recorded, the absorbance can be determined by the formula:

$$\text{Absorbance } A = \log_{10}(BD/BC),$$

where B is the zero/∞ chart point, D is the center point of the baseline, and C is the peak apex.
5. The calibration line is constructed (this should be a straight line).
6. If necessary, all samples are liquefied and homogenized. Solutions in carbon disulfide of 200 mg (to the nearest 0.1 mg) in 10-ml volumetric flasks are made up to volume.
7. All sample solutions are read against the same reference solution 1 (tristearin). The baselines are constructed, and the peaks are measured as above.
8. The values (equivalent to trielaidin) for samples are read off the standard line, and the percentage *trans* from the sample weight is calculated.
9. If there is a significant degree of interference to the construction of the baseline, the samples should be converted to methyl esters. This should also be done where the TV is below 5%. In this method, the standard is methyl elaidate, and the reference becomes methyl stearate. All other manipulations are the same.

Table 3 Calibration solutions for IR spectroscopy

Material (mg)[a]	*Solution*				
	1	*2*	*3*	*4*	*5*
Tristearin	200	175	150	100	50
Trielaidin	0	25	50	100	150

[a]Weighed to the nearest 0.1 mg.

See also: **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography; **Fats**: Classification; **Phospholipids**: Determination; **Spectroscopy**: Infrared and Raman; **Triglycerides**: Characterization and Determination; **Vegetable Oils**: Composition and Analysis

Further Reading

Borch RF (1975) *Analytical Chemistry* 47: 2437–2439.
Fales HM, Jaouni TM, Babashak JF *et al*. (1973) *Analytical Chemistry* 45: 2302–2303.
Hamilton RJ and Hamilton S (eds) (1992) *Lipid Analysis – A Practical Approach*. Oxford: IRL Press, Oxford University Press.
Hammond EW (1993) *Chromatography for the Analysis of Lipids*. Boca Raton, FL: CRC Press.
Horwitz W (ed.) (2000) *Official Methods of Analysis of the Association of Official Analytical Chemists*, 17th edn. Washington, DC: AOAC.
Madison BI, Depalma RA and D'Alonzo RP (1982) *Journal of American Oil Chemists' Society* 59: 178–181.
Sheppard AJ and Iverson JL (1975) *Journal of Chromatographic Science* 13: 448.
Torres AG, Trugo NMF and Trugo LC (2002) A mathematical method for the prediction of retention times of fatty acid methyl esters in temperature-programmed capillary gas-chromatography. *Journal of Agricultural and Food Chemistry* 50: 4156–4163.
Wood R and Lee T (1983) *Journal of Chromatography* 254: 237–246.
Yanagisawa I and Yamane M (1985) *Journal of Chromatography* 345: 229–240.

Dietary Importance

C J Field, University of Alberta, Edmonton, Alberta, Canada

Background

In Western countries, an adult eats on average 75–150 g of fat each day, representing 30–45% of the energy in the diet. Triacylglycerols constitute more

than 95% of dietary fat. Each triacylglycerol molecule is composed of three fatty acids esterified to a glycerol backbone, making fatty acids a major constituents of dietary fat. In recent years, it has become clear that fatty acids are more than just a source of energy. Fatty acids, especially long-chain polyunsaturated fatty acids (PUFA), are important regulators of many biological processes important in maintaining health and preventing disease.

Types of Fat in the Diet

Because of the wide range of foods consumed, the human diet contains a great variety of fatty acids. The chain length and degree of saturation are used to classify fatty acids, and these characteristics confer physiological properties. Fatty acids in the diet can therefore be saturated fatty acids (SFA) and contain no double bonds with the maximum number of hydrogen molecules attached, monounsaturated (MUFA) and contain a single double bond or PUFA and contain two or more double bonds. All dietary fats and oils contain a mixture of saturated and unsaturated fatty acids. Individual fatty acids are often referred to by their common names, but are more correctly identified by a systematic nomenclature that classifies them according to their chain length, position and number of double bonds (Table 1). This nomenclature indicates the number of carbon atoms in the acyl chain of the molecule, followed by a colon and then the number of unsaturated bonds. This is followed by the letter '*n*' (or ω) and the number of carbon atoms from the methyl end to the first double bond. Mammalian cells are able to synthesize (from nonfat precursors) saturated and unsaturated fatty acids of the *n*-9 and *n*-7 series but lack the Δ-12 and Δ-15 desaturase enzymes (which is found in most plants) for insertion of a double bond at the *n*-6 or

Table 1 Nomenclature of fatty acids and major sources in the diet

Chemical name	*Common name*	*Fatty acid notation*	*Sources*
Butanoic	Butyric	C4:0	Butter fat
Octanoic	Caprylic	C8:0	Palm kernel oil
Decanoic	Capric	C10:0	*De-novo* synthesis; coconut oil
Dodecanoic	Lauric	C12:0	*De-novo* synthesis; coconut oil
Tetradecanoic	Myrsitic	C14:0	*De-novo* synthesis; milk fat, coconut oil
Hexadecanoic	Palmitic	C16:0	*De-novo* synthesis and most plant and animal fats, including milk, eggs, animal fats, meat, coca butter, palm oil (other vegetable oils contain some but lesser amounts), and fish oils
Octadecanoic	Stearic	C18:0	*De-novo* synthesis and most plant and animal fats, including milk, eggs, animal fats, meat, and cocoa butter
9-Hexadecenoic	Palmitoleic	C16:1*n*-7	Desaturation of palmitic acid, fish oils, animal fats
9-Octadecenoic	Oleic	C18:1*n*-9	Desaturation of stearic acid and all plant and animal fats, including milk; eggs; animal fats; meat; cocoa butter; most vegetable oils, especially olive oil
9,12-Octadecadienoic	Linoleic	C18:2*n*-6	Cannot be synthesized in mammals; some milks, eggs, animal fats, meat; most vegetable oils, especially corn, sunflower, safflower, cottonseed and soybean oils; green leaves
9,12,15-Octadecatrienoic	α-Linolenic	C18:3*n*-3	Cannot be synthesized in mammals; green leaves; some vegetable oils, especially canola, soybean, walnut, mustard and linseed oils
6,9,12-Octadecatrienoic	γ-Linolenic	C18:3*n*-6	Synthesized from linoleic acid, borage, blackcurrant and evening primrose oils
11,14,17-Eicosatrienoic	Mead	C20:3*n*-9	Synthesized from oleic acid; indicator of essential fatty acid deficiency
Eicosanoic	Arachidic	C20:0	*De-novo* synthesis; peanut oil
8,11,14-Eicosatrienoic	Dihomo-γ-linolenic	C20:3*n*-6	Synthesized from γ-linolenic
5,8,11,14-Eicosatetraenoic	Arachidonic	C20:4*n*-6	Synthesized from linoleic acid via γ-linolenic and dihomo-γ-linolenic acids; small amount in meat
5,8,11,14,17-Eicosapentaenoic	Eicosapentaenoic (EPA)	C20:5*n*-3	Synthesized from α-linolenic acid; fish oils
7,10,13,16,19-Docosapentaenoic	Docosapentaenoic	C22:5*n*-3	Synthesized from α-linolenic acid via EPA; fish oils and animal tissues (brain)
4,7,10,13,16,19-Docosahexaenoic	Docosahexaenoic (DHA)	C22:6*n*-3	Synthesized from α-linolenic acid via EPA; fish oils and animal tissues (brain)

n-3 position (Figure 1). As the two types of PUFA, *n*-6 and *n*-3 fatty acids, are essential substrates for many of the regulatory lipids in the body, the precursor fatty acids linoleic (18:2*n*-6) and α-linolenic (18:3*n*-3) must be consumed in the diet (Figure 1). For the most part, the longer chain and more unsaturated *n*-6 and *n*-3 fatty acids such as arachidonic acid (20:4*n*-6), eicosapentaenoic acid (EPA, 20:5*n*-3), and docosahexaenoic acid (DHA, 22:6*n*-3) can be synthesized from either linoleic or linolenic acid (Figure 2).

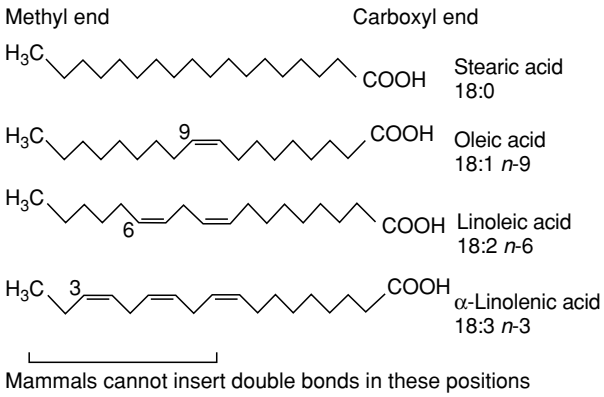

Figure 1 Double-bond location in major dietary fatty acids.

Sources of Fat in the Diet

The quantities and types of fatty acids ingested vary greatly, depending on the dietary source (Table 1). Animal fats contain about 40–60% saturated fatty acids. Some plant oils also contain SFA (i.e., palm oil, palm kernel oil, and coconut oil), which are widely used in processed foods. MUFA are found in animal fats and plant oils, with olive oil being a rich source. Plant oils like corn, soybean, cottonseed, sunflower, and safflower generally have more than 50% of their fatty acids as linoleic acid and are considered excellent sources of PUFA. Poultry and game also contain a small amount of linoleic acid. Linolenic acid, which comprises the majority of most people's *n*-3 fatty acid intake, is found in high amounts in

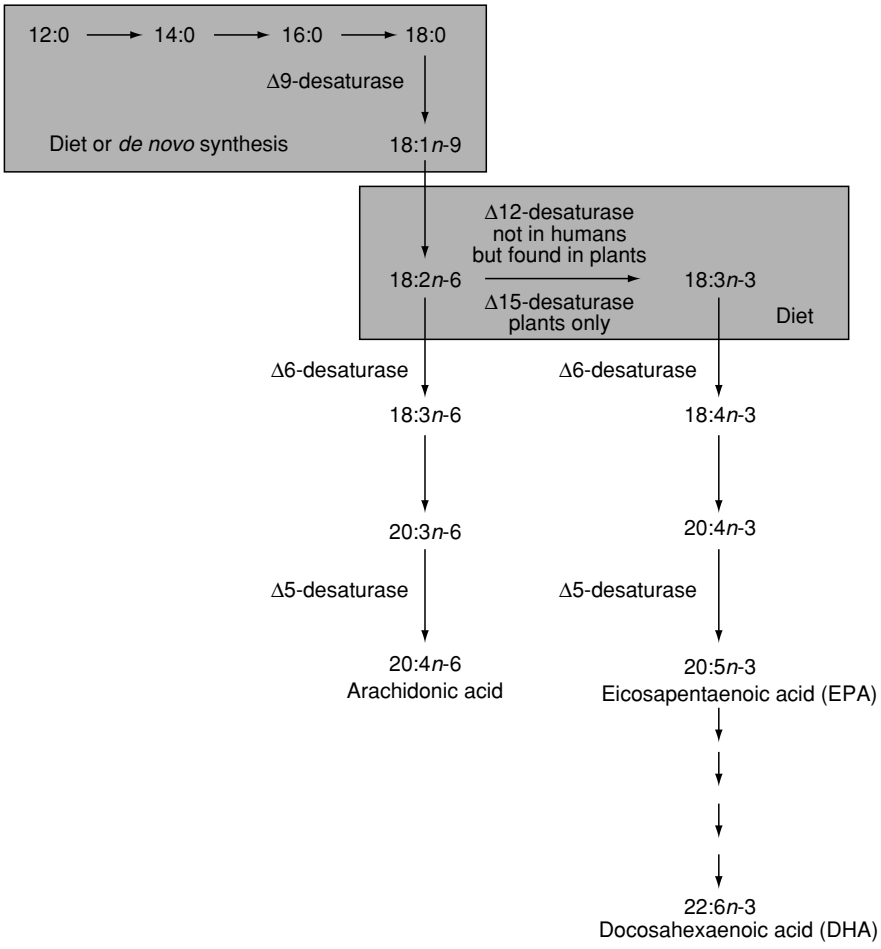

Figure 2 Elongation and desaturation of the major dietary fatty acids.

several plant oils (e.g., canola and flax) and, to a lesser extent, in other plant oils, green leafy vegetables, soybeans, and nuts. The more polyunsaturated long-chain *n*-3 fatty acids, eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), comprise 20–30% of the fatty acids in cold water fish (particularly fatty fish such as herring, mackerel, fresh tuna, sardines, salmon, eel), other marine animals, and in oils extracted from the livers of fish which live in warmer waters (e.g., cod). There are new sources of EPA and DHA, made from single cell organisms, which are available to the food industry to supplement food. Most natural fats contain fatty acids in the *cis* form, but a small number of *trans* fatty acids also occur naturally, principally in meat and dairy products. The majority of the *trans* fatty acids in the diet come from foods made with hydrogenated vegetable oils. Generally, most fruits and vegetables contribute insignificant amounts or no fat to our diets.

Dietary Importance

The dietary importance of fatty acids is listed in Figure 3.

Energy and Essential Fatty Acids

In addition to contributing to the texture, flavor, and aroma of foods, fatty acids are an important concentrated source of energy (9 kcal or 37 $kJ\,g^{-1}$) for most cells in the body. Fat is a particularly important energy source for infants and young children, and restriction is associated with poor growth and delayed development. Although the biological importance is not completely clear, there is evidence suggesting that individual fatty acids are oxidized, metabolized, stored, and released at different rates. Fat slows down the digestion of foods, thus contributing to meal satiety. Dietary fat aids in the absorption of the fat-soluble vitamins. Specific fatty acids have important structural, biochemical, and regulatory functions that are required for optimal tissue function, growth, and repair. Insufficient consumption of the dietary essential fatty acids (linoleic and linolenic acid) results in a deficiency syndrome characterized by immune dysfunction, infections, scaly dermatitis, growth retardation, hair loss, thrombocytopenia, diarrhea, and poor wound healing.

- Energy
- Source of essential fatty acids
- Membrane structure & function
- Gene regulation
- Neurological and retinal structure and function
- Immune function
- Disease prevention & treatment

Figure 3 Dietary importance of fatty acids.

Membrane and Cellular Function

Fatty acids, particularly PUFA, when consumed in the diet, integrate into cell membrane lipids. Although membranes can vary in their lipid content, most cell membranes contain approximately 50% lipid and 50% protein. Phospholipids, the major lipid found in cell membranes, are particularly high in long-chain (18–26-carbon) PUFA. The type of fatty acid incorporated into cell membranes is influenced by dietary fatty acids. It can be assumed that all of the *n*-6 and *n*-3 fatty acids found in cell membranes are derived from dietary sources. Changing the relative amount and type of PUFA in cell membranes has been shown to affect many membrane-related functions such as membrane fluidity, ion channel flow, transporter activity, signal transduction, enzyme activity, hormone binding, cell-receptor action, cell-to-cell communication, release of mediators, and susceptibility to microbial invasion and cancer.

Eicosanoid Synthesis

Linoleic acid and α-linolenic acid are substrates for the synthesis of physiological regulators called eicosanoids. Eicosanoids include prostaglandins, prostacyclins, thromboxanes, and leukotrienes. These 'local hormones' are potent mediators of many biochemical processes and play key roles in the regulation of blood clots, blood pressure, blood lipid levels, immune function, inflammation, pain and fever, and reproduction. The two major pathways for eicosanoid synthesis are via the enzymes cyclooxygenase (produce prostaglandins and thromboxanes) and lipooxygenase (produce leukotrienes, hydroxyeicosatrienoic acids, and lipoxins).

Dietary fat composition influences eicosanoid synthesis by changing the supply of substrates for the synthesis of the longer-chain *n*-3 and *n*-6 PUFA. Consuming large amount of linoleic acid increases the quantity of arachidonic acid in cell membranes. Upon activation, in a variety of cells, arachidonic acid is converted to eicosanoids of the 2 series and leukotrienes of the 4 series. However, dietary α-linolenic acid is converted to EPA in the membrane, and when cells are activated, EPA is converted to eicosanoids of the 3 series and leukotrienes of the 5 series. In general, eicosanoids formed from *n*-3 PUFA oppose, or have weaker effects than, eicosanoids formed from *n*-6 PUFA. In addition to competing as substrates for cyclooxygenase, *n*-3 fatty acids can directly suppress the activity of this enzyme. *n*-3 fatty acids also inhibit Δ-6 and Δ-5 desaturase activity, reducing the synthesis of arachidonic acid in the membrane. Indeed, it has been shown that increasing the proportions of *n*-3 in relation to *n*-6 PUFAs in the diet decreases the

quantity of proinflammatory, vasoconstrictive, platelet aggregatory, and immunosuppressive compounds that are regulated by eicosanoids. Recent investigations have shown that γ-linolenic acid (20:3*n*-6) present in some plants (i.e., evening primrose oil, borage, and flaxseed) may also influence the type of eicosanoid synthesized by a cell.

Gene Regulation

PUFA have been shown to alter the expression of numerous genes involved in the metabolic function of the cell. It has been demonstrated that specific dietary fatty acids can modulate the expression of a variety of genes (and transcription regions on genes) coding for key regulatory proteins in metabolic pathways such as those involved in digestion, lipogenesis, glycolysis, glucose transport, inflammation, and cellular communications. For example, dietary *n*-6 and *n*-3 PUFA suppress hepatic lipogenesis by suppressing transcription of fatty acid synthase. Suppression is greater with DHA, EPA, and arachidonic acid than with linoleic acid or α-linolenic acid, whereas SFA and MUFA have little effect. In the future, fatty acid–gene interactions will likely be translated to dietary prescriptions to maintain health and modulate both the response to injury and the progression of diseases in genetically susceptible individuals.

Neurological Development and Function

DHA and arachidonic acid selectively accumulate during fetal and infant brain development and are found concentrated in the brain and retina. Fetal accretion of these lipids occurs during the last trimester of gestation, and a deficiency can result in visual impairment and brain dysfunction. Although DHA and arachidonic acid can be synthesized from dietary precursors, the efficiency of this conversion in young infants (particularly preterm infants) may not be sufficient to meet their high need. Postnatally, human milk provides high amounts of both arachidonic acid and DHA to the infant, but the amount of these lipids in milk can be influenced by the mixture of fatty acids in the maternal diet. The preterm, and perhaps fullterm, infant not fed mother's milk may be at risk of deficiency if these lipids are not provided. Animal studies have clearly illustrated that provision of arachidonic acid and DHA to young animals improves visual and neurological function. Despite considerable evidence of improved visual function in formula infants provided DHA (and in some studies also arachidonic acid), the dietary requirements for these fatty acids in the infant vary greatly between countries. Recently, animal studies have provided evidence that DHA obtained from plasma is involved both in dopamine and serotonin metabolism in the brain and in the retina for the function of rhodopsin. These findings should guide clinical studies to more sensitive measures of the functional roles of dietary *n*-3 fatty acids. At present, there are promising preliminary results regarding the efficacy of treating various visual disorders and forms of depression with dietary DHA.

Immune Function

There is now considerable scientific evidence, from animal and human studies, that the amount and type of fatty acids in the diet can modulate immune function. A deficiency of either *n*-6 or *n*-3 essential fatty acids is associated with a reduced immune function. Amongst the fatty acids, the *n*-3 fatty acids possess the most potent immunomodulatory activities, and amongst the *n*-3 fatty acids, those from fish oil (EPA and DHA) are more potent than α-linolenic acid. In particular, it has been demonstrated that feeding DHA and EPA (often in a mixture of 'immunoactive' nutrients) decreases the rate of infections and improves outcome (i.e., reduced complications, length of stay, hospital cost) in immunosuppressed patients. These effects are likely mediated through mechanisms that involve indirect (i.e., via eicosanoids, hormone action) and direct (regulation of transcription factors) effects on genes in immune cells.

Many of the detrimental and lethal effects of infection and injury (especially detrimental effects) are mediated by the proinflammatory cytokines (e.g., tumor necrosis factor α (TNF-α), interleukin (IL)-1 and IL-6) and eicosanoids. Thus, a reduction in the synthesis of proinflammatory mediators would be beneficial during critical illness and autoimmune states. In both clinical trials and animal studies, feeding DHA and EPA has been demonstrated to decrease the expression of proinflammatory cytokines, genes, eicosanoids, and adhesion molecules, whereas *n*-6 PUFAs increase the expression of proinflammatory genes. These observations have formed the basis for successful interventions with oral EPA and DHA to treat acute and chronic inflammation, and for disorders that involve an inappropriately activated immune response (i.e., inflammatory bowel disease, rheumatoid arthritis, psoriasis, asthma, multiple sclerosis, systemic lupus erythematosus, and atopic dermatitis).

A number of studies have also reported suppressed cell-mediated immunity, but not generally humoral immunity, in humans or animals fed very high levels of DHA and EPA. One might conclude that consumption of long-chain *n*-3 PUFA might be contraindicated in those with poorly functioning immune systems. However, the very low level of DHA and EPA and the high level of linoleic acid currently in the diet of most people questions the physiological relevance of

n-3 immunosuppression in a healthy individual. To support this, there are many studies reporting benefits to the immune system when EPA/DHA are fed to immunosuppressed individuals, such as those with cancer or recovering from surgery. This suggests that the effect of *n*-3 PUFA on immune function may depend on the composition and content of other fatty acids in the diet.

Disease Prevention and Treatment

Epidemiological studies have concluded that consumption of different amounts and types of fatty acids over a lifetime can alter the risk for many chronic diseases (e.g., heart disease, cancer). There is also scientific evidence to support that dietary fatty acid composition can be used in the prevention/treatment of diseases such as coronary heart disease (CHD) and cancer. Both *n*-6 and *n*-3 PUFA have been shown to be of potential benefit in the treatment of acute respiratory distress syndrome and asthma. Beneficial effects of feeding *n*-3 fatty acids have been shown in the secondary prevention of CHD, hypertension, type 2 diabetes, and, in some patients with renal disease, rheumatoid arthritis, inflammatory bowel diseases, and chronic obstructive pulmonary disease.

Coronary heart disease Individual fatty acids differ in their ability to change blood cholesterol and lipoprotein levels, two major risk factors for CHD.

SFA For the past 30 years, epidemiological and experimental studies have found that the consumption of a high SFA diet increases low-density lipoprotein (LDL)-cholesterol levels, a major risk factor for CHD. The precise mechanism for these effects has not been established, but high intakes of SFA decrease the removal of plasma LDL by LDL receptors, and they may promote endogenous cholesterol synthesis. More recently, studies have found that SFA have varying effects on plasma lipid levels, depending on their chain length. Much of this work has been carried out on animals, and there is considerable controversy as to the effect of stearic, palmitic, myristic, and lauric acids on plasma lipid levels and cholesterol synthesis. It is likely that the deleterious effect of SFA on CHD in epidemiological studies is explained by the hypercholesterolemic effects of myristic, palmitic, and perhaps also lauric acid. There are less data to support a hypercholesterolemic effect of consuming stearic acid. Recently, research has suggested that the cholesterol-raising properties of myristic and palmitic acid may be related to the intake of PUFA, such that, when PUFA falls below a threshold intake (around 5% of energy), the cholesterol-raising properties of these SFA are greatly augmented.

MUFA Generally, studies have found that populations that consume high MUFA diets have a lower risk of CHD. Clinical and animal studies have confirmed that consumption of MUFA (i.e., oleic acid) is associated with lower total- and LDL-cholesterol levels and higher high-density lipoprotein (HDL)-cholesterol levels. Consumption of MUFA has also been reported to increase LDL-receptor activity.

n-*6 PUFA* Epidemiological studies report that populations that consume 10% of their fats as linoleic acid have a lower risk of developing CHD. A lower dietary ratio of PUFA/SFA is associated with higher plasma cholesterol concentrations and a higher population risk of CHD. In clinical studies, it has been shown that consumption of linoleic acid decreases the synthesis of cholesterol in the body and increases LDL-receptor activity, resulting in lower total- and LDL-cholesterol levels. Most countries' dietary recommendations have attempted to promote an increase in the intake of *n*-6 PUFA for its beneficial effects on reducing the risk of CHD. However, some studies have also indicated that a high intake of *n*-6 fatty acids (much higher than the current recommendation of 10% of energy) may reduce HDL-cholesterol levels and shift the physiologic state to one that is prothrombotic and proaggregatory, changes that are associated with an increased risk of CHD.

n-*3 PUFA* Epidemiological studies conclude that the incidences of CHD and atherosclerosis are inversely associated with consumption of *n*-3 PUFA, particularly EPA. The antiatherosclerotic actions of *n*-3 PUFA have been supported in animal and human clinical trials. Experimentally, it has been demonstrated that consumption of *n*-3 fatty acids (α-linolenic, DHA, or EPA) reduced plasma triacylglycerol levels (very-low-density lipoprotein-triacylglycerols), a risk factor for heart disease (especially in women). Evidence for the beneficial effects of consuming *n*-3 PUFA on blood triacylglycerols, blood pressure, cardiac and vascular function, eicosanoids, coagulation, and immunological responses has prompted many health agencies to recommend increasing the intake of these fatty acids to reduce the risk of CHD. There is still, however, a high degree of uncertainty in defining both the type and amount of *n*-3 PUFA that is considered preventative. In a recent metaanalysis of studies examining the relationship between fish consumption and CHD, it was concluded that consumption of 40–60 g of fish each day was associated with a markedly reduced CHD mortality in high-risk populations but not in low-risk populations. In cardiac patients, there are reports of beneficial effects of ingesting oily fish on the incidence of bypass graft

occlusions, forearm vasodilation, arrhythmias, nonfatal stroke, nonfatal myocardial infarction, and cardiac death. There is also increasing evidence for an antiarrhythmic action of *n*-3 PUFA in humans. Most of the studies in high-risk populations have been carried out by feeding fish oils (EPA and DHA), and less is known about the effects of their plant-derived precursor, α-linolenic acid.

Trans-*fatty acids* A high consumption of *trans*-fatty acids is associated with an increase in LDL-cholesterol and a decrease in HDL-cholesterol, two changes that would increase the risk of CHD. *Trans*-fats account for about 2–8% of the calories in our diets, compared with 12–14% for saturated fats. Currently, most nutritionists and researchers recommend that hydrogenated fats containing *trans*-fatty acids be limited or reduced in the diet.

Cancer Human epidemiological and experimental studies have established a positive relationship between high (mainly SFA) intake and cancer. However, high intakes of SFA or *n*-6 PUFA increase the growth of tumors in several animal models, while decreasing the *n*-6/*n*-3 ratio (by increasing the intake of *n*-3 fatty acids) has been shown to inhibit the growth and metastasis of rodent tumors. There is convincing evidence that both DHA and EPA can inhibit the growth of cancer cells (breast and colon) both *in vitro* and in explants. It has been suggested that the *n*-6/*n*-3 ratio of the diet, rather than the absolute intakes of these PUFA, may define cancer risk. Possible mechanisms include both direct effects on tumor growth and indirect effects on anticancer immune defenses. Although many studies support a recommendation to decrease the *n*-6/*n*-3 ratio in the diet, most cancer agencies do not yet feel that there are sufficient human data to make recommendations to the population. Current recommendations to prevent cancer have been aimed at promoting a decrease in dietary fat intake, focusing on reducing the intake of SFA. There are currently no specific fatty acid recommendations for those with, or recovering from, cancer. There are data to support beneficial effects of supplementing individuals with DHA/EPA. These benefits include reducing tumor growth, improving anticancer immune defenses, improving the response to surgery, chemotherapy, and radiation treatments, and decreasing cachexia-mediated weight loss.

Estimated Intakes and Recommendations for Fatty Acids in the Diet

Quantitative dietary guidelines for fats were first issued in the late 1970s and have changed relatively little since that time. Current recommendations for total fat in the diet of adults by most Western countries is less than or equal to 30% of energy. The lower limit of fat intake to meet the energy needs of adults is assumed to be between 10 and 15% of dietary energy, provided that enough carbohydrates are available. This minimum level depends on the fat requirement to meet energy needs, the need for essential fatty acids, and the amount of fat in the diet that is necessary to absorb fat-soluble vitamins (particularly vitamins A and E). Most countries recommend that the intake of SFA should not exceed 10% of the energy intake, PUFA should comprise about 10% of energy and the remainder of the intake of fat be comprised of MUFA. A comparison of dietary fat and fatty acid recommendations between several countries is shown in Table 2.

SFA and MUFA

In the past 30 years, the absolute consumption of SFA in Western diets has declined. For example, in the UK, SFA intake has declined by 40%. Based on the lower risk of some chronic diseases in Mediterranean countries, where the population consumes diets high in MUFA, it is generally recommended that the population replace SFA with MUFA. However, since 1970, the intake of MUFA in the UK has declined by 30%.

Table 2 Dietary fat recommendations from selective countries around the world

Country	*Total fat (percentage of energy)*	*SFA (percentage of energy)*	*MUFA (percentage of energy)*	*PUFA (percentage of energy)*
UK	33	10	None stated	6–10
Canada	< 30	10	None stated	> 3.5
USA	30	10	None stated	None stated
Japan	20–25	6–8	8–10	6–8
New Zealand	30–35	12	20	8
Iceland	25–30	15	None stated	None stated
Malaysia	20–30	10	10	10
Singapore	20–30	10	10	10

Source: Inform (2001) *International News of Fats, Oils and Related Materials* 12(2): 132–140, published by the AOCS Press.

PUFA

For adults, the recommended intake for essential fatty acids is in the range of 3–5% of dietary energy for linoleic acid, and 0.5–1.0% of dietary energy for linolenic acid (about 1–4 g day^{-1}). During the past 60 years, in the UK, the consumption of PUFA has increased by 25%, despite a decrease in total fat intake. This was likely contributed to by the increased consumption of linoleic acid in margarines and various plant oils (sunflower, safflower, cottonseed, soybean, canola). Estimated intakes of PUFA are similar in North America, with the exception that the intake of *n*-3 fatty acids may be lower than in the diet of those in the UK. This increased intake of linoleic acid (approximately 10–14 g day^{-1}) over the past decades has changed the balance between *n*-6 and *n*-3 fatty acids in the diet. Dietary intake studies, in developed countries, estimate that the dietary *n*-6/*n*-3 fatty acid ratio approaches 14:1–20:1. However, most studies suggest that a ratio between 4:1 and 10:1 is needed to achieve the health benefits associated with consumption of *n*-3 fatty acids.

Despite the health benefits of *n*-3 PUFA, the mean daily intake falls far short of any current recommendation. The intake of DHA and EPA in most developed countries is small and related to the intake of fatty fish, and so promoting fish consumption is the most obvious way to increase *n*-3 PUFA intake. In North America, it is recommended that people should have at least two servings of fish per week. On a cautionary note, PUFA are oxidized more readily than SFA and MUFA and may increase the susceptibility of cellular membranes to lipid peroxidation. Indeed, increased free-radical production has been demonstrated in animals fed diets rich in *n*-3 PUFA. Consuming additional antioxidants such as vitamin E can minimize oxidation risk.

Conclusions and Summary

Fat contributes more than a third of the energy in the diet of most people. Fat comprises a number of different fatty acids varying in carbon chain length and degree of saturation. Specific dietary fatty acids have many essential roles in the body in addition to providing an energy-efficient fuel. Changing the types or fatty acids in the diet changes the composition and function of membranes, eicosanoid synthesis, regulates gene expression, controls neurological and retinal development, and modulates immune function in health and disease. There is considerable research demonstrating the importance of *n*-3 PUFA or the *n*-6/*n*-3 PUFA ratio in the prevention and treatment of CHD and cancer. On the basis of current knowledge, an increase in the consumption of long-chain *n*-3 fatty acids seems a prudent recommendation. In summary, dietary fatty intake represents a major environmental factor affecting optimal growth, development, response to injury, and overall health in the population.

See also: **Cancer**: Diet in Cancer Prevention; Diet in Cancer Treatment; **Fats**: Requirements; **Fatty Acids**: Properties; Metabolism; Gamma-linolenic Acid; **Fats**: Classification

Further Reading

Calder PC (1999) Dietary fatty acids and the immune system. *Lipids* 34 (Supplement): S137–S140.

Calder PC and Deckelbaum RJ (1999) Dietary lipids: more than just a source of calories. *Current Opinion in Clinical Nutrition & Metabolic Care* 2: 105–107.

Gibney MJ (1999) Strategies for altering population intakes of fats and fatty acids. *Proceedings of the Nutrition Society* 58: 189–191.

Hamosh M and Salem N, Jr. (1998) Long-chain polyunsaturated fatty acids. *Biology of the Neonate* 74: 106–120.

Innis SM (2000) The role of dietary *n*-6 and *n*-3 fatty acids in the developing brain. *Developmental Neuroscience* 22: 474–480.

Marckmann P and Gronbaek M (1999) Fish consumption and coronary heart disease mortality. A systematic review of prospective cohort studies. *European Journal of Clinical Nutrition* 53: 585–590.

Milner JA and Allison RG (1999) The role of dietary fat in child nutrition and development: summary of an ASNS workshop. American Society for Nutritional Sciences. *Journal of Nutrition* 129: 2094–2105.

Mori TA and Beilin LJ (2001) Long-chain omega 3 fatty acids, blood lipids and cardiovascular risk reduction. *Current Opinion in Lipidology* 12: 11–17.

Roche HM (1999) Unsaturated fatty acids. *Proceedings of the Nutrition Society* 58: 397–401.

Simopoulos AP (1999) Essential fatty acids in health and chronic disease. *American Journal of Clinical Nutrition* 70: 560S–569S.

Uauy R and Valenzuela A (2000) Marine oils: the health benefits of *n*-3 fatty acids. *Nutrition* 16: 680–684.

Trans-fatty Acids: Health Effects

A Aro, National Public Health Institute (KTL), Helsinki, Finland

Effects on Serum Lipoproteins

Early studies on the effects of *trans*-fatty acids on serum total cholesterol levels in human subjects gave conflicting results. Studies in experimental animals

like rats, mice, and pigs did not show any appreciable difference between the effects on serum cholesterol of partially hydrogenated vegetable oils and nonhydrogenated oils. Therefore, the common impression was that *trans*-fatty acids were not hypercholesterolemic or that their effect on serum cholesterol concentration was weak and clinically not important. This impression was not changed until studies focusing attention separately on low-density lipoprotein (LDL) and high-density lipoprotein (HDL) became available in the 1990s.

Since 1990, several dietary intervention trials have indicated that, when compared with oleic acid, *trans*-fatty acids from partially hydrogenated vegetable oils or fish oils increase serum LDL-cholesterol and tend to reduce HDL-cholesterol in a dose-dependent manner (Figure 1). The increase in LDL-cholesterol has been apparent in all studies, but without a clear dose-effect. The HDL-cholesterol concentration has showed little change by small doses of *trans*-fatty acids, but higher amounts, in the order of 6–11% of energy, have clearly reduced HDL-cholesterol concentrations in a dose-dependent manner. In most studies, the LDL-cholesterol-raising effect has been somewhat weaker than that of equal amounts of C12–16 saturated fatty acids, but this may partly depend on short study periods. Corresponding effects on serum apoproteins, an increase in apoprotein B and a decrease in apoprotein A-I concentration, have been found in those studies where apoprotein concentrations have been measured. In some studies, an increase in serum triglycerides has been observed but the effect has been weak and generally statistically nonsignificant.

Partially hydrogenated fish oils that are used in some countries like Iceland and Norway and were previously also consumed in other north-western European countries seem to have a more powerful effect on serum lipoproteins than *trans*-fatty acids of vegetable origin (Figure 2). The isomeric composition of C18:1 *trans*-fatty acids is rather similar in partially hydrogenated vegetable and fish oils, but the various long-chain C20–22 *trans*-isomers that are formed when fish oils are chemically hydrogenated may have an exceptionally strong LDL-cholesterol-raising effect. The effect on lipoproteins of the delta-11 isomer *trans*-vaccenic acid, the main *trans*-isomer of animal fats, has not been studied separately from other fatty acids, because pure preparations have not been available for this purpose. The effects of different C18:1 *trans*-isomers may vary, because the isomers are metabolized by different routes. *Trans*-vaccenic acid is desaturated to conjugated linoleic acid (CLA) by the delta-9-desaturase, whereas the main isomers of partially hydrogenated vegetable and fish oils, the delta-9-(elaidic acid) and delta-10-isomers, cannot be metabolized by this pathway. The effects of CLA on the risk of coronary heart disease (CHD) are poorly known, but some CLA-isomers may have antiatherogenic properties according to experimental studies in animals. Information on the metabolism of *trans*-fatty acid isomers is limited. Generally, the *trans*-isomers seem to be metabolized by the same pathways as the corresponding *cis*-isomers.

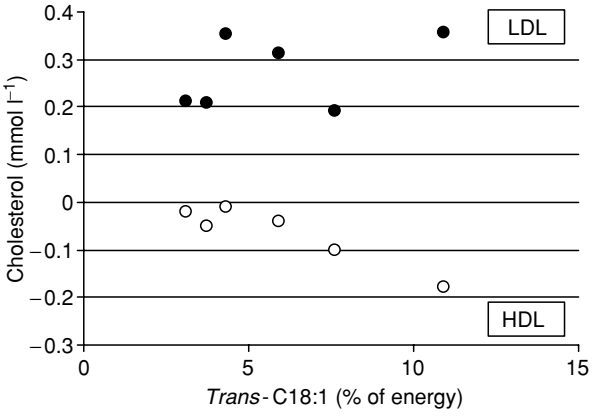

Figure 1 Effects on serum low-density lipoprotein (LDL)-cholesterol (closed symbols) and high-density lipoprotein (HDL)-cholesterol concentrations (open symbols) of *trans*-fatty acids from partially hydrogenated vegetable oils as compared with oleic acid in six controlled dietary trials. Data from Katan MB, Zock PL and Mensink RP (1995) *Trans* fatty acids and their effects on lipoproteins in humans. *Annual Review of Nutrition* 15: 473–493.

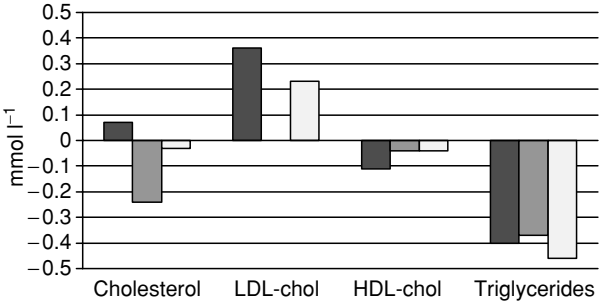

Figure 2 Effects on serum and lipoprotein lipids by 3-week feeding of partially hydrogenated fish oil (dark columns), partially hydrogenated vegetable oil (gray columns), and butter (light columns). The columns indicate changes from a customary Norwegian background diet. LDL-chol, low-density-lipoprotein cholesterol; HDL-chol, high-density lipoprotein cholesterol. Data from Almendingen K, Jordal O, Kierulf P, Sandstad B and Pedersen JI (1995) Effects of partially hydrogenated fish oil, partially hydrogenated soybean oil, and butter on serum lipoproteins and Lp[a] in men. *Journal of Lipid Research* 36: 1370–1384.

In the clinical trials with partially hydrogenated vegetable oils and fish oils, amounts of *trans*-fatty acids (3–10% of energy) have been used that are higher than the habitual consumption in most countries. The *trans*-fatty acids have been derived from specially designed products high in *trans*-fatty acids and not from commonly available foods. Therefore it is difficult to estimate the effects of the amounts that are habitually consumed by people in western countries as part of their diets. Studies on the composition of foods have revealed that processed foods that contain *trans*-fatty acids often contain more *cis*-unsaturated fatty acids than the corresponding foods that are based on saturated animal fats. As an example, icecream that is manufactured from partially hydrogenated vegetable oil contains more *cis*-unsaturated fatty acids than icecream based on milk fat or coconut oil. Thus the effects of moderate amounts of *trans*-fatty acids may be balanced by the neutral or LDL-cholesterol-reducing effects of the concomitant *cis*-unsaturated fatty acids of the same foods.

Trans-Fatty Acids and Risk of Coronary Heart Disease

The observed effects of dietary *trans*-fatty acids on serum lipoproteins, an increase in LDL-cholesterol concentration combined with a reduction in HDL-cholesterol, are expected to increase the risk of CHD. The associations between *trans*-fatty acid intake and risk of CHD have been studied in several case-control studies and cohort studies. Some studies have supported the concept of increased risk of CHD by high consumption of *trans*-fatty acids, but the results have been somewhat controversial.

In most case-control studies the fatty acid composition of adipose tissue or serum has been used as an indicator of *trans*-fatty acid intake. In three studies, C16:1 *trans*-fatty acids that are mainly derived from fats of animal origin were higher in cases than in control subjects, but in other studies the C16:1 *trans*-isomers did not show significant associations with the risk of CHD. In six of the studies, no significant case-control difference was found in C18:1 *trans*-fatty acids. One study showed a direct association between C18:1 *trans*-fatty acids in platelet phospholipids and risk of CHD, and one study showed an inverse association between C18:1 *trans*-fatty acids in adipose tissue and risk of sudden cardiac death. One study from the USA, based on food intake data, showed an increased relative risk of CHD in the highest quintile of *trans*-fatty acid intake but a decreased risk in the third quintile, representing moderate intake. In the largest of the case-control studies, the multicenter EURAMIC study, adipose tissue samples from 671 male survivors of a first acute myocardial infarction and from 717 reference men were studied. No overall case-control difference was found in C18:1 *trans*-fatty acids (Figure 3). The

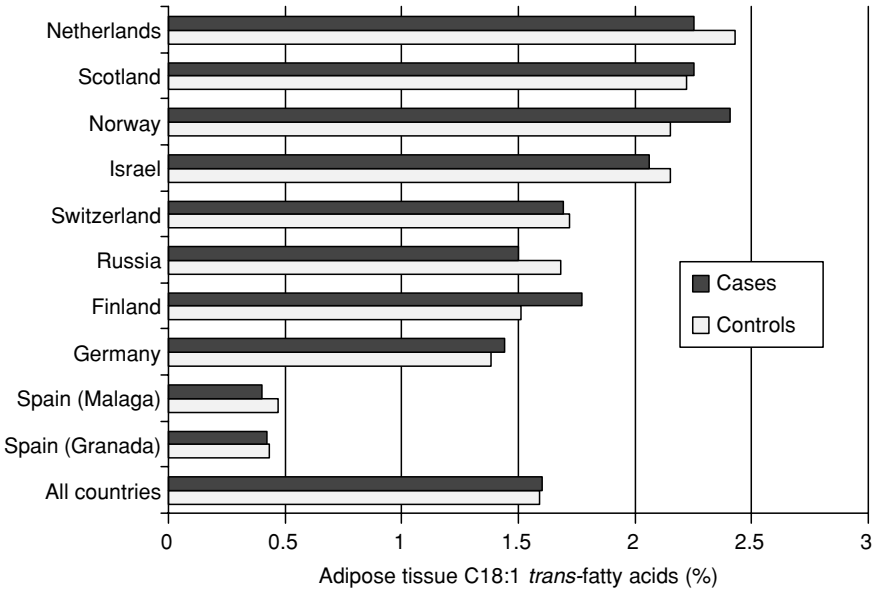

Figure 3 Mean proportions of adipose tissue C18:1 *trans*-fatty acids in male survivors of a first myocardial infarction (cases) and male control subjects without coronary heart disease (controls) in the nine countries of the European multicenter, case-control EURAMIC study conducted in 1991–92. Data from Aro A, Kardinaal AFM, Salminen I *et al*. (1995) Adipose tissue isomeric *trans* fatty acids and risk of myocardial infarction in nine countries: the EURAMIC study. *Lancet* 345: 273–278.

case-control differences differed between the countries. In Norway and Finland, cases with acute myocardial infarction had significantly higher proportions of *trans*-fatty acids in the adipose tissue than control subjects. In Russia and one of the Spanish centers, a significant difference in the other direction was found. The results showed up to sixfold differences between the countries in adipose tissue *trans*-fatty acids, suggesting large differences in the average dietary intakes of *trans*-fatty acids between the nine participating countries.

The differences in dietary intakes were subsequently confirmed by the TRANSFAIR study in 14 European countries. When the two Spanish centers of the EURAMIC study that showed very low proportions of C18:1 *trans*-fatty acids in the adipose tissue were excluded, the results suggested a borderline significant increase in the relative risk of CHD in the two highest quartiles of adipose tissue C18:1 *trans*-fatty acids compared with the lowest quartile. The C18:1 *trans*-fatty acids are the largest and clinically most important cluster of *trans*-fatty acids. The C16:1 *trans*-fatty acids are a minor component mainly derived from ruminant animal fats. Different *cis*/*trans*-isomers of C18:2 are derived in relatively small quantities from both partially hydrogenated vegetable oils and ruminant animal fats. Epidemiological studies do not suggest any major impact on the risk of CHD by these small *trans*-fatty acid clusters.

The discordant results of the case-control studies seem to reflect more analytical problems, problems in interpreting the fatty acid composition of tissues, and difficulties in distinguishing between the causes and effects of the disease in subjects with diagnosed CHD than real effects on the risk of CHD. The differing results may also indicate that *trans*-fatty acids in tissues are markers of some other associated factors that have more impact on CHD risk than the small amounts of *trans*-fatty acids in the diet. Overall, the results seem to agree with the concept that the current average intakes of *trans*-fatty acids are too low to have a significant effect on CHD risk.

Four cohort studies have addressed the association between the intake of *trans*-fatty acids measured either with semiquantitative food-frequency questionnaires or modified dietary history methods. All these studies showed a significantly increased relative risk of CHD in the highest quintile or tertile of *trans*-fatty acid intake compared with the lowest (Table 1). The increased risk was associated with high intakes of C18:1 *trans*-fatty acids only, whereas moderate intakes were not significantly associated with the risk of CHD. In the female and male cohorts from the USA and in the male ATBC study cohort from Finland, the mean intake of *trans*-fatty acids in the highest quintile was between 1.6 and 2.9% of energy intake. In the Dutch Zutphen elderly study, the median intake of the highest tertile was much higher – 6.4% of energy intake. The results of the two US studies suggested that the increased risk of CHD was due to *trans*-fatty acids from partially hydrogenated vegetable oils only, whereas in the Dutch study, no difference between the risk effects

Table 1 Multivariate[a] relative risk (RR) and 95% confidence intervals (CI) of coronary heart disease in cohort studies in relation to *trans*-fatty acid intake

	Intake of trans*-fatty acids*				
	Quintile 1 (low)	*Quintile 2*	*Quintile 3*	*Quintile 4*	*Quintile 5 (high)*
Nurses' Health Study (Hu *et al*., 1997)[b]					
No. of cases 939					
RR (95% CI)	1.0	1.07 (0.86–1.33)	1.10 (0.89–1.37)	1.13 (0.91–1.39)	1.27 (1.03–1.56)
Health Professionals					
Follow-up Study (Ascherio *et al*., 1996)[b]					
No. of cases	112	140	147	154	181
RR (95% CI)	1.0	1.12 (0.86–1.44)	1.12 (0.87–1.44)	1.12 (0.86–1.46)	1.21 (0.93–1.58)
ATBC study (Pietinen *et al*., 1997)[b]					
No. of cases	244	288	268	297	302
RR (95% CI)	1.0	1.10 (0.93–1.31)	0.97 (0.82–1.16)	1.07 (0.90–1.28)	1.14 (0.95–1.35)
	Tertile 1 (low)	*Tertile 2*	*Tertile 3 (high)*		
Zutphen elderly study (Oomen *et al*., 2001)[b]					
No. of cases	24	30	44		
RR (95 CI)	1.0	1.34 (0.76–2.37)	2.00 (1.07–3.75)		

[a]Adjusted for various risk factors of coronary heart disease.

[b]Oomen CM, Ocké MC, Feskens EJM *et al*. (2001) *Trans* fatty acid intake is associated with an increased 10-year risk of coronary heart disease in the Zutphen Elderly Study. *Lancet* 357: 746–751.

of vegetable and animal *trans*-fatty acids was evident. In a fifth analysis based on the cohorts of the Seven Countries Study, initially studied in the late 1950s and early 1960s, the intakes of both saturated fatty acids and C18:1 *trans*-fatty acids showed significant statistical associations with the 25-year CHD mortality. This study was based on analyzed average diets and individual data were not available. In 14 out of the 16 study cohorts, *trans*-fatty acids were mainly derived from animal products. The results thus represent rather the effects on CHD risk of concomitant saturated fatty acids, present in far higher amounts in the main animal sources of *trans*-fatty acids, than the relatively small amounts of *trans*-fatty acids present in these foods.

It has been calculated from the combined results of the cohort studies that a decrease in *trans*-fatty acid intake by 2% of energy decreases the risk of CHD by 25%. This would correspond roughly to the effect of a reduction of 5% of energy in the intake of saturated fatty acids. A reduction of 2% of energy in *trans*-fatty acid intake is seldom possible, because the current average intakes of *trans*-fatty acids are below 2% of energy in most countries, but a reduction of saturated fatty acid intake from 15 to 10% of energy would be feasible in most western countries.

Mechanisms of Action of *Trans*-Fatty Acids

The effects of *trans*-fatty acids on serum lipoproteins clearly have a mechanism that is different from that of saturated C12–16 fatty acids. Both increase the LDL-cholesterol concentration in serum but, in addition, saturated fatty acids increase serum HDL-cholesterol concentration, whereas *trans*-fatty acids decrease the HDL-cholesterol concentration. The effect of *trans*-fatty acids may also appear more slowly, within 3–6 weeks, than that of saturated fatty acids that becomes apparent in 2–3 weeks' time.

An increase of the cholesterol ester transfer protein (CETP) activity by *trans*-fatty acid feeding has been documented in several but not all human intervention studies with *trans*-fatty acids. CETP transfers cholesterol from HDL to very-low-density lipoprotein (VLDL) and LDL. Increased CETP activity could thus explain the decrease in serum HDL-cholesterol concentration combined with an increase in LDL-cholesterol found in intervention studies with *trans*-fatty acids. This could also explain the negative findings in many animal studies, since mice, rats, and pigs do not exhibit CETP activity. The increase in lipoprotein a (Lp(a)) found in most *trans*-fatty acid supplementation studies is another difference between the effects of saturated and *trans*-fatty acids. Lp(a) levels are mainly genetically determined, and the mechanism of the effect of *trans*-fatty acids on LP(a) levels remains unexplained.

In experimental studies on the effects of fatty acids in isolated mouse islets of Langerhans, *trans*-fatty acids increased maximal insulin output compared with the corresponding *cis*-isomers. In human studies, no clear effects on insulin sensitivity by *trans*-fatty acids have been observed.

The effects on hemostatic variables have been measured in two intervention trials. No effects on fibrinogen, factor VII, or tissue plasminogen activator levels, ascribed to *trans*-fatty acids, could be observed. In one of the studies, plasminogen activator inhibitor-I (PAI-I) antigen and activity were increased by the consumption of partially hydrogenated vegetable oil, but this finding was not confirmed by the other study. As far as blood coagulation and fibrinolysis are concerned, there seems to be little difference between the effects of saturated and *trans*-fatty acids.

Trans-fatty acids compete with linoleic acid and alpha-linolenic acid for the metabolizing enzymes. In experimental animals they impair the microsomal desaturation and elongation of the essential fatty acids to their long-chain metabolites arachidonic acid and docosahexaenoic acid. It has been suggested that *trans*-fatty acid intake of the mother during pregnancy and lactation may be harmful for premature and newborn babies, because a high *trans*-fatty acid intake may impair the formation of the essential long-chain polyunsaturated fatty acids in the baby. In adults with ample consumption of polyusaturated fatty acids, this effect of *trans*-fatty acids is probably without practical importance.

Trans-Fatty Acids and Cancer

Experimental Studies

There is much information from animal studies concerning the effects of different fatty acids on experimental cancer, but very few of these yield information on the effects of *trans*-fatty acids. In two studies in mice comparing elaidic acid (*trans*-9 C18:1) as pure fatty acid with esterified oleic acid in triglycerides, a possible tumor-promoting effect of elaidic acid was observed. Subsequent studies comparing *trans*- and *cis*-isomers esterified in triglycerides have shown quite similar effects of *trans*- and *cis*-monoenoic fatty acids on the development of experimental colon cancer and breast cancer.

Epidemiological Studies

It is difficult to assess the effects of dietary *trans*-fatty acids independently of the intakes of the much more

abundant saturated and *cis*-unsaturated fatty acids. Therefore, data on *trans*-fatty acids have not been reported in the majority of the studies. Most studies on the associations between dietary fatty acids and risk of cancer have addressed breast cancer. In the largest of the cohort studies, the Nurses' Health Study, no association between the intake of *trans*-fatty acids and risk of breast cancer was observed. In some case-control studies the fatty acid composition of adipose tissue or serum has been used as a biomarker of intake. In two of these, no association was found between *trans*-fatty acids and the risk of breast cancer. In a third study, *trans*-fatty acids were associated with an increased risk of breast cancer in the lowest tertile of polyunsaturated fatty acids. In one study, a significantly lower proportion of *trans*-11 C18:1 (*trans*-vaccenic acid), mainly derived from cheese, was found in postmenopausal women with breast cancer as compared with matched healthy control subjects, but the *trans*-9 and *trans*-10 isomers in serum showed no association with the risk of breast cancer.

Intake of fatty acids has been associated with the risk of colorectal cancer in some prospective studies, but the contribution of *trans*-fatty acids to the risk has not been assessed in any of the studies. Based on these scanty data, it can be concluded that available evidence does not suggest any appreciable role to the intake of *trans*-fatty acids in relation to the risk of any type of cancer.

Future Aspects

In the past, high intakes of *trans*-fatty acids, exceeding 4% of energy intake, have been recorded in some countries like Norway and the Netherlands, due to the consumption of hydrogenated vegetable and fish oils, and in Canada and the USA, due to common use of partially hydrogenated vegetable oils. As awareness of the undesirable effects on serum lipoproteins of partially hydrogenated vegetable and fish oils has increased, table margarines and other soft spreads containing no or low concentrations of *trans*-fatty acids have become commonly available but industrial shortenings and fats for deep-frying remain high in *trans*-fatty acids. For reasons of texture, e.g., in icecream and leafy pastries, or stability, e.g., in deep-frying and in industrially produced fatty foods with a long shelf-life, *trans*-fatty acids or saturated fatty acids or both will be used in these foods even in the future. In cooking and baking at home, hard fats can be replaced by vegetable oils, but it seems advisable to restrict the consumption of industrially manufactured fatty foods because of their undesirable fatty acid composition. If and when *trans*-fatty acids are removed from foods, as much as possible they should be replaced by *cis*-unsaturated fatty acids. If saturated fatty acids are substituted for *trans*-fatty acids, the benefit in terms of CHD risk is questionable.

The proportion of *trans*-fatty acids derived from ruminant animal products will increase concomitantly with the reduction of *trans*-fatty acids derived from partially hydrogenated vegetable oils in foods. Already in 1995, dairy products and meat contributed more than 50% to total *trans*-fatty acid intake in seven out of the 14 European countries participating in the TRANSFAIR study. It is possible to substitute *trans*-vaccenic acid and CLA for saturated fatty acids in milk and beef fat by modifying the feeding of cows. So far there are no studies indicating whether modifying the composition of milk fat in this way should be considered beneficial or harmful with respect to effects on CHD risk factors.

See also: **Cancer**: Diet in Cancer Prevention; **Cholesterol**: Factors Determining Blood Cholesterol Levels; **Coronary Heart Disease**: Etiology and Risk Factor; **Lipoproteins**

Further Reading

Almendingen K, Jordal O, Kierulf P, Sandstad B and Pedersen JI (1995) Effects of partially hydrogenated fish oil, partially hydrogenated soybean oil, and butter on serum lipoproteins and Lp[a] in men. *Journal of Lipid Research* 36: 1370–1384.

Aro A (1998) Epidemiological studies on *trans* fatty acids and coronary heart disease. In: Sébédio JL and Christie WW (eds) *Trans* Fatty Acids in Human Nutrition, pp. 235–260. Dundee: Oily Press.

Aro A (2001) Complexity of issue of *trans* fatty acids. *Lancet* 357: 732–733.

Aro A, Kardinaal AFM, Salminen I *et al.* (1995) Adipose tissue isomeric *trans* fatty acids and risk of myocardial infarction in nine countries: the EURAMIC study. *Lancet* 345: 273–278.

Ascherio A, Hennekens CH, Buring JE, Master C, Stampfer MJ and Willett WC (1994) Trans-fatty acids intake and risk of myocardial infarction. *Circulation* 89: 94–101.

Hu FB, Stampfer MJ, Manson JE *et al.* (1997) Dietary fat intake and the risk of coronary heart disease in women. *New England Journal of Medicine* 337: 1491–1499.

Ip C and Marshall JR (1996) *Trans* fatty acids and cancer. *Nutrition Review* 54: 138–145.

Katan MB, Zock PL and Mensink RP (1995) *Trans* fatty acids and their effects on lipoproteins in humans. *Annual Review of Nutrition* 15: 473–493.

Kris-Etherton PM (ed) (1995) *Trans* fatty acids and coronary heart disease risk. *American Journal of Clinical Nutrition* 62: 655S–707S.

Lichtenstein AH (1997) *Trans* fatty acids, plasma lipid levels, and risk of developing cardiovascular disease. *Circulation* 95: 2588–2590.

Oomen CM, Ocké MC, Feskens EJM *et al*. (2001) *Trans* fatty acid intake is associated with an increased 10-year risk of coronary heart disease in the Zutphen Elderly Study. *Lancet* 357: 746–751.

Pietinen P, Ascherio A, Korhonen P *et al*. (1997) Intake of fatty acids and risk of coronary heart disease in a cohort of Finnish men. The alpha-tocopherol, beta-carotene cancer prevention study. *American Journal of Epidemiology* 145: 876–887.

Precht D and Molkentin J (1995) *Trans* fatty acids: implications for health, analytical methods, incidence in edible fats and intake. *Die Nahrung* 39: 343–374.

FDA *See* **Food and Drug Administration**

FERMENTED FOODS

Contents

Origins and Applications

B J B Wood, University of Strathclyde, Glasgow, UK

Introduction

Many entries in this *Encyclopedia* will deal, directly or indirectly, with the actions of microorganisms of foodstuffs. Some will be entirely negative, relating to food poisoning, disease transmission, and toxin production. Others will show that changes effected by microbes can be beneficial and even desirable, bringing about changes that enhance the texture, flavor, color, edibility, and hence the value of the foodstuff concerned. This entry discusses the benefits and probable origins of the fermentations. It further attempts to place them in a historical and economic context.

The Beginnings of Food Fermentation

Although we cannot know how or when the practice of fermentation in relation to food began, all available evidence suggests that it is a very ancient part of human development. In fact, fermentation can be seen as a subset of biodeterioration. All things that can serve as foods for humans (or any other animal) are, by their capacity to act as foods, more or less unstable. If they are food for mammals, then they will also be able to support the life of many types of bacteria and molds. In practice, even such apparently stable materials as wood are open to attack by microbes and insects. There is a powerful drive throughout nature to use the energy and nutrition present in organic materials, eventually mineralizing these materials back to carbon dioxide, water, and simple mineral salts.

We see attack by such microorganisms as, primarily, a deterioration, or loss of desirable characteristics. Some such changes, however, turn out to have desirable characteristics. Such changes may produce improvements in flavor, odor, and texture. In addition, there is substantial evidence of improvements in nutritional status and value arising from controlled fermentations. Various studies have shown that even simple fermentations of basic cereal grains with the indigenous (mainly lactic acid bacteria) microflora generate significant increases in the levels of some B-group vitamins. There are strong suggestions that plant proteins, which are somewhat difficult for the monogastric digestive system to handle, can be made more digestible through the actions of proteolytic enzymes produced by bacteria and fungi. Many adults are intolerant to lactose, the principal sugar of milk, but fermentations such as that to produce yogurt reduce or eliminate this problem. Fermentations can

also reduce or remove toxins and other 'antinutritional factors.' Perhaps the most important example of the former is the detoxification of cassava, while an example of the latter is hydrolysis of phytic acid, with removal of sequestering by it of divalent metals. Perhaps even more significant in the present context may be the prospect that fermentation increases the stability of the affected food, increasing its resistance to attack by other organisms. Before the advent of modern methods for food preservation, such as freezing and canning, even a short extension in storage life would be precious. The prolonged storage possible for the main protein in milk (casein) consequent upon conversion to a hard cheese (such as Cheddar) would thus be of immense value. We must also remember that availability evenly spread through the year is a very modern phenomenon. In any given region (particularly in the temperate parts of the world) food supply is very seasonal. The idea that tomatoes, apples and new potatoes would always be on sale would have seemed unreasonable even a couple of decades ago. Now it is taken for granted throughout the industrialized world.

An Example

The existence of small pits (perhaps a meter cube, or even less) dug into the chalk of southern England has long been known, but their purpose remains a matter for controversy. One suggestion has been that they would have been used to store grain. However, *a priori* this seems unlikely. In the ground grain will be exposed to moisture. All seeds will take up available water as a preparation for germination, and will then become more easily attacked by molds than the dry seeds would be. Chalk, being very porous, will readily conduct water. Thus grain storage in pits dug into the surface of chalk would seem to be an inappropriate method for conserving the material. However, experiments showed an interesting sequence of events. Dry grain was packed into pits prepared for the purpose, and the pits were then sealed with a cap of clay or other impervious material. Initially, the grain closest to the clay took up water and was then colonized by a range of fungi and other microorganisms. Eventually this agglomeration of grains and hyphae formed an impervious layer, preventing further ingress of moisture. In addition, respiration by the stored grain and the microbes feasting in the outer layers of the stored material used up the available oxygen, replacing it with carbon dioxide. Thus, at the sacrifice of the outermost layer of the stored material, the bulk of it was preserved for a long time.

This does not correspond with our normal idea of a food fermentation. However, it has illustrations for a number of key points. The pit makers could have had no understanding of the processes at work in their pits. Despite this, the prevailing conditions favored growth of microbes in a beneficial way. There was some loss of stored material, but this was an acceptable price for the storage of the rest in useable condition. All of the foregoing is speculation; we can have no certain knowledge of how the pits were used by their makers. The grain storage hypothesis is nevertheless a valid one, and experiments show that the pits would operate as conservation devices.

Historical Background

Very often our understanding of how a fermentation began is speculation based on our interpretation of the modern process in the light of our knowledge of conditions in the past. Sometimes this is aided by ancient records, pictures such as those from Egyptian tombs, continuing oral traditions, but in many cases such evidence is scanty or entirely absent. Thus there is a great deal of guesswork in our attempts to reconstruct how particular fermentations originated. In most cases this is unimportant, but it can be irritating when we are faced with a product such as kefir. Here the fermentation relies upon structures called grains. Modern analyses show that these grains are highly structured arrangements of yeasts and bacteria upon polysaccharide membranes. These organisms can be separately cultivated in the laboratory, but they do not reform into the characteristic kefir grains when cultured together. How did the grains arise? Was this a single event, or did it happen several times (kefir is widespread across Russia and Eastern Europe)? Clearly, we are unlikely ever to know the answers to such questions. Even in simpler cases, such a bread making or the conversion of grapes to wine or grain to beer, we have no records showing the beginnings of such activities. Often there are myths showing such knowledge as gifts from gods or other superior intelligences. Also foods such as bread and wine frequently have a place in sacramental activities. The giving and consumption of bread and wine, central to the Christian Eucharist, are also found in older traditions, associated with the Roman gods, for example. This infers the great importance of such activities at the change from hunter/gatherer economies to the beginnings of settled agriculture. I suggest that this knowledge was an essential part of that change. The ability to store foods enhanced survival over winter and other periods of hardship.

Converting dry grains and other seeds into something more appetizing than a gruel must have made agriculture more attractive and valuable. Alcohol, despite its dangers, provided (and still provides), in reasonable moderation, a basis for social interaction.

The apparent changes in the nature of consciousness effected by alcohol must have had spiritual significance, and the economic power possessed by those who control its production and distribution remains important to this day. The vital part played by bread in the agrarian economies of Europe is made starkly plain by the draconian laws controlling the activities of medieval bakers, and that of beer by ordinances such as the German beer laws. The trade in fish sauces throughout the Roman Empire affords another example. We tend to associate fish sauces with the Far East, but it is known that several fish sauces have been produced in the Mediterranean region since ancient times. Archeological evidence indicates that these products were traded in substantial amounts throughout the Empire, even as far as Hadrian's wall. Although we cannot know for certain their nature, it seems clear that these were important condiments in the Roman kitchen.

Options for Food Preservation

Before the advent of bottling, canning, and other methods using a combination of heat and an effective seal against recontamination of the heat-treated food by microbes, there were very few safe technologies available. Cold was recognized as valuable in delaying the onset of putrefaction, but was not generally available for most of the year. The richest people had ice-houses, insulated structures where the winter's snow and the ice from pools could be conserved for at least part of the warmer months, then used to store food, as well as providing the means for preparing cold drinks. Access to such ice-stores was only for the very rich, and zero Celsius is of limited value in food preservation. Despite this, preservation of food in ice attracted the interest of the great Sir Francis Bacon, who died in 1626 from a cold caught as a consequence of experiments in preserving fowl with ice. The discovery that ice mixed with salt gave lower temperatures was of very limited value, and must have been an extremely expensive option.

For most people, the main technologies depended upon reducing the water available for microbial growth. Air-drying is still important in preserving many foods in tropical countries, and even in northern latitudes the drying of fish was important. In practice, such drying is normally accompanied by some microbial growth, frequently demonstrated by the odors emanating from the drying food, particularly fish. Salting is a very ancient practice, its importance demonstrated by salt taxes, the Roman payment of salt to soldiers (hence 'salary'), and so forth. Sugar also has importance, principally, although not entirely, in preserving fruits and some vegetables. Their naturally dry state confers particular value on seeds, even though they can be fairly uninteresting foods without considerable processing. On the other hand some potential foods, such as cassava (*Manioc esculenta*) contain such powerful natural preservatives that they are lethal if consumed in their natural state, while others, such as millet, contain unacceptable levels of tannins or other bitter principles.

Thus, the preservative properties associated with some food fermentations must have been important even at the dawn of agriculture, and became much more so as urban civilization developed. Equally, the abilities of fermentations to detoxify things such as cassava are essential if these materials are to become suitable for human consumption. Another preservative agent derived from fermentation, vinegar, has not been mentioned so far. Its value as both a condiment and a preservative has, however been recognized since the most ancient times. It also had importance in medicine and healing. With the rapid transfer of the British population from a mainly rural existence to an increasingly urban one, consequent upon the Industrial Revolution, the enhanced demand for vinegar resulted in development of the 'quick' vinegar process to replace the slow Orleans method. It must also be recognized that fermentation often went together with other methods of preservation, in order to give the desired results. In particular the capacity of salt, even at fairly low concentrations, selectively to enhance growth of certain lactic acid bacteria (LAB) is vital to the success of important vegetable and meat fermentations. The possible significance of essential oils and other factors derived from herbs and other flavorings will be considered below.

Physical and Chemical Properties of Raw Materials for Fermentations

Appropriate fermentation techniques for given products will be strongly influenced by the physical and chemical properties of the raw materials to be used in its preparation. It is therefore useful to attempt a classification of raw materials. What follows will be entirely pragmatic, and is not intended as a formal classification scheme. Instead it will be a grouping by the characteristics most relevant to fermentation procedures. In order to do this, it is first desirable to review the relevant physical and other characteristics.

Of the physical properties, probably the most important will be the availability of water, as this will strongly influence the types of organisms whose growth can be supported on a particular medium, and, in the extreme case, whether any organisms at all will grow on the material in question. Water availability can be limited by simple physical dryness,

i.e., the actual absence of water, and by osmotic pressure creating a physiological drought. Salt and/or sugar in foodstuffs most commonly induce the latter. It should be noted that quite low concentrations of salt, well below that which might be expected to induce osmotic stress, can strongly influence the course of fermentations. The structures of some fruits and vegetables are remarkably well developed to resist attacks by both plant and microbial invaders. The cassava tuber affords a particularly highly developed example of combined physical and chemical defenses. It has both a tough, fibrous exterior, and a capacity to produce toxic levels of hydrogen cyanide when it is damaged. Such materials require to be mechanically damaged to breach the outer layers. This is generally accomplished by chopping, shredding, or slicing, depending on the material and intended product. The outermost layers may also need to be removed, being tough, fibrous, and lacking in nutritive value. Where toxins are present (as in cassava), measures to neutralize them will also be required. In some cases, however, agents developed by the plant as protection may be desired in culinary practice. Examples include the irritants present in onions and the roots of plants such as radishes. While stress has been placed on roots, tubers, and bulbs in the foregoing discussion, arising from their greater numbers of types, certain other structures also perform in similar manners. Perhaps the most obvious example is the cocoa fruit. This is a robust structure with a tough outer layer, which must be breached before the inner pulp can undergo fermentation both to release and chemically modify the seeds that will yield cocoa powder after roasting and grinding.

The chemical composition of the fermentable material is obviously of overriding importance in determining the types of organisms that can grow on it and the outcomes of fermentations which they effect. In some cases there will be ample supplies of (for example) sugars to give desired levels of organic acids, B-group vitamins to support growth of fastidious LAB, and amino acids. In many cases, however, there will be insufficient readily usable material. Then, organisms which can produce and export appropriate hydrolytic enzymes (amylases or proteases, for example) will be at an advantage. They may also prepare the way for later growth of less competent organisms such as yeasts and LAB.

The part played by chemicals in herbs, onions, and garlic can be particularly interesting. We value these materials principally as flavors in foods, but many of them also have long traditions as healing agents and medicinals. There is increasing evidence that these traditions have a secure basis in scientific fact. claims that onions, garlic, and many culinary herbs are powerfully inhibitory against bacteria associated with food poisoning have been reviewed; it is reported that onions, garlic, allspice, and oregano 'killed every bacterial species tested.' These are clearly valuable properties if shown to apply to actual foodstuffs, but an interesting question arises. Many fermented foods use onions, garlic, and various herbs. Some sausages use particularly high levels of garlic (e.g., salami). Sausages are harsh environments even without these inhibitors. Yet bacteria, such as species of *Carnobacterium* and *Micrococcus*, are present in these products, and are thought to be essential for flavor development. How selective is the toxicity of these plant products? There are clearly interesting biochemical challenges here, even if these claims were shown to be fully justified in the field, i.e., in actual food fermentations. It has been pointed out that, for example, salamis, although particularly heavily spiced with garlic, have been associated with serious illness due to *Escherichia coli*. Thus claims of this type need to be assessed very carefully, with purely laboratory studies and robust trials in actual foods being run in partnership with each other, if the effects on food microflora of herbs and spices are to be fully understood.

The foregoing is not intended as a complete discussion of the issues involved here. Rather it is an indication of some of the factors that require consideration when attempting to understand the factors influencing the initial development and subsequent evolution of food fermentations.

A Classification of Raw Materials

Animal Products

Meats This term is taken to include the flesh from mammals, and also that from fishes, birds, and crustaceans. In all cases, the material as obtained postmortem is usually in the form of chunks of tissue, dissected organs (such as muscles, liver), whole organisms (often, but not always eviscerated; small fish are an obvious example). Occasionally the material may be comminuted in the course of its extraction, but relatively large pieces are more usual. Sometimes, particularly in small fish for conversion to fish sauce or for fermentation with rice, the material can be used without further physical modification. More usually it is necessary to comminute the material by mincing or grinding. This is primarily to insure sufficiently intimate mixing of salt, spices, farinaceous materials, and other additives that both promote the desired fermentation and confer desired texture and flavor on the finished product. It also simplifies packing the prepared mixture into appropriate casings.

In certain cases the raw material may be of relatively low intrinsic value, with fermentation being a way of utilizing residues. Typical of this situation is producing oriental fish sauces from trash fish and the viscera, heads, tails, and other unwanted parts remaining after marketable fish has been prepared for sale. For high-value products, such as quality sausages, producers will specify the meat carefully, with prime cuts being essential for the best products, although sausages can represent a way of using lesser-quality meat economically, while still yielding a merchantable product.

Milk Milk is surely the archetype for a fermentation substrate. Its production, particularly by traditional methods, guarantees that it carries a substantial microbial load ab initio. Even in modern parlors using mechanical milking and strict hygiene, the milk still has a significant number of microbes in it. The high water content, presence of readily fermentable carbohydrate (lactose) and nitrogenous materials, B-vitamin content, and the near-neutral pH combine to provide nearly ideal growing conditions for a wide variety of organisms, including potential pathogens. The temperature of freshly drawn milk is also near the optimum for many bacteria. Thus modern processing, for direct consumption, butter production or fermentation, requires rapid cooling of the milk, swift transport to a processing center, and rapid thermal processing (usually pasteurization or ultrahigh-temperature (UHT) treatment) to provide a safe and relatively stable product. Despite impassioned argument about the greater flavor claimed for cheeses made from unpasteurized milk, some developed countries continue to insist on appropriate heat treatment for all milk to be used for making cheeses. On the other hand it must be acknowledged that the bacteria in unpasteurized milk tend to be predominantly LAB, and that the selected cultures employed in modern production originate from isolates originally found in spontaneous fermentations. The relatively anaerobic conditions, which develop in bulk milk, rapid acidification and various factors elaborated by LAB (bacteriocins, hydrogen peroxide) combine to suppress the growth of many pathogens in cheeses. Thus cheese and yogurt production are still valid technologies for societies where modern milk processing is not available to most of the producers.

Plant Materials

The great diversity of plant materials means that any attempt at grouping into classes will be fraught by difficulties caused by overlaps and ambiguities. For example, tomatoes and cucumbers, although obviously fruit, are often placed among the vegetables, while the edible stems of rhubarb are placed alongside fruits. Thus, what follows is entirely arbitrary, reflecting the practical problems in fermenting plant foods.

Leafy material Grass and other forage materials are converted into silage on an enormous scale as an essential part of modern ruminant farming. Although not normally thought of as a food fermentation, i.e., not part of the human diet, this is probably the most important among leaf fermentations in the wider context. Among materials for direct human consumption, the best-known products are kimchi and sauerkraut. The latter is a highly industrialized process in the major producing nations. Undoubtedly it was developed principally as a storage process for conserving excess cabbage through the winter in areas where the severe winter weather would destroy cabbage heads left in the fields, while cut heads would quickly deteriorate in store if held as the unmodified material. The key factors for fermentation are provided for by adding about 2–3% by weight of salt to shredded cabbage heads, then packing tightly into containers where the sap, withdrawn from the plant tissue by osmotic action of the dissolving salt, quickly displaces remaining air from between the plant material. Bacteria naturally present on the plant surfaces drive the lactic fermentation, although there is a tendency towards using selected cultures now. This simple process gave a product that could be conserved without deterioration for many months, even in the absence of refrigeration. Obviously there is no need to continue this process as a means for conserving cabbage in industrialized societies, where advanced storage systems and transportation which can deliver any food from any part of the world without regard to season provide complete freedom from the old domination by the solar cycle. However, this has not reduced demand for the fermented product. On the contrary, those accustomed to it have taken its production with them when they have migrated to and settled in other parts of the world. Consequently its production is, if anything, increasing. The technology is basically the same as for the materials discussed in the next section, with the key difference that they use salt in prepared brine, whereas here dry salt is added to the raw material.

Fruits Cucumbers are perhaps the classical example here, although olives would be cited first in Mediterranean climates. Other examples include tomatoes. Small onions and certain types of radish roots, although not fruits, are processed in the same way. Again the technology is very simple, easily applicable on a village or domestic level, and originally a way of conserving seasonal surpluses of produce against

winter shortages. Essentially the prepared material is immersed in a salt solution (brine) and a spontaneous lactic fermentation develops. In reasonably cool conditions the products are stable during prolonged storage. The observations about the survival of sauerkraut technology after the conservation need had passed apply with equal force here, as do those on industrialization of the processes. Olives are remarkable in the market penetration that the fermented product has achieved outside its traditional consumers.

Another type of conservation of fruits relies on fermenting the released juices to produce stable liquids. This is a very pedestrian way of describing the delights of wines and the products from apple (cider; USA 'hard cider') and pear (perry) juices, but is deliberately done to place them in a specific context. Undoubtedly the capacity of the products to induce intoxication would have attracted attention as soon as they were discovered. Indeed we have ample evidence that birds and mammals are attracted to fermenting fruits, and will manifest clear evidence of drunkenness, and even of addiction to ethanol. In many cases the intoxication became associated with religious experiences among humans. Despite these complicating sociological factors, we can properly see alcoholic fermentation as a preservation strategy, essentially indistinguishable from those presented in the preceding two sections. Sweet fruits have a very short storage life, although strategies have been developed to permit storing some hard fruits (notably apples and pears) for extended periods. At the other extreme, grapes are so easily bruised that picking them for table use demands great care; the yeasts always present on the grape skins quickly ferment the fruit juice at the slightest opportunity. The only practical alternative to fermentation is drying to raisins, currants, and sultanas.

Roots, bulbs, and tubers In general, root crops are stable enough for their storage in simple systems, and even for them to be left in the ground during winter until required. Thus, with minor exceptions such as the radishes referred to above, root crops are not fermented in temperate climates. Bulbs are also generally stable enough for winter storage, so fermentation in brine is probably done more for culinary than for conservation reasons. The New World gave us, among many other valuable crops, three novel starchy crops that have now become so indispensable that we often forget their origins. Of these, one is a cereal (maize) but the other two are root crops, potatoes (*Solanum tuberosum*), and cassava (*Manioc esculenta*). Of these two, I have no knowledge of potatoes being fermented, except in the very limited sense of them being a source of starch for alcohol production. Potatoes can produce various toxins as defenses against predators, but protecting the tubers from exposure to light easily prevents production of the most important of these, and I am not aware of any evidence that fermentations will reduce toxicity of green tubers. On the other hand, fermentation is absolutely essential for making cassava safe for use. Most varieties are extremely toxic, because of the presence of a cyanogenic glycoside. In the intact tuber this, and the enzyme that can hydrolyze it, are strictly segregated, but mechanical damage brings them into contact, so releasing the highly toxic hydrogen cyanide. It is clear that even very primitive peoples from its native habitats had developed techniques for making the tuber comparatively safe to eat. These are not primarily fermentations, depending as they do on the enzymatic action described above to destroy the cyanogenic glycoside. However, the conditions under which the tubers are harvested and prepared for detoxification insure that there is an abundant microbial flora associated with them. LAB, which do not use iron-containing enzymes, are resistant to cyanide, and their growth on the cassava portions produces acidic conditions that both favor the action of the relevant enzyme and help in removal of the cyanide as the unionized gaseous form. Subsequent processing includes drying, reduction to granular form, and heat treatments such as (in the case of products such as the West Africa gari) roasting. Unfortunately, even garification will not reliably remove the last traces of cyanide. Although the residual material is below the acutely toxic concentration, it can, by competing with iodide, produce the thyroid gland malfunction known as goiter.

Seeds The plant products discussed so far have been characterized by high water content. This it true even of the firmest products, such as the root crops. This water content means both that any injury leaves the foodstuff liable to fermentation or other microbial attack, and that even the undamaged item has a restricted storage life because eventually water loss will reduce it to a flaccid and uninviting state. However, some vegetables can be made suitable for storage by deliberate, controlled dehydration. With seeds, the evolutionary development resulted in a very dry, tough, long-lasting product. Seeds are also packed with nutrients placed there by the parent plant to support initial growth of the new plant whose embryo the seed contains. Thus a seed has many characteristics which make it potentially desirable as a human or animal food. From the human viewpoint, however, seeds have some disadvantages as foods. As harvested, they are difficult to eat. If processed in the obvious way, by soaking and/or boiling in water, the

product is apt to be rather bland and unappetizing. Indeed, of the various grains and seeds harvested by humans, only rice is consumed in this way to any extent. Commination by crushing or grinding can effect limited improvements, but the resulting gruels or porridges are still of limited appeal. Much the same is true if the material is formed into a paste and baked or roasted. On the other hand, even the simplest fermentations can effect considerable improvements. For example, a common breakfast dish in Africa is a maize porridge in which the ground maize is left overnight in water, and then boiled next morning, ready for eating. Overnight there is substantial acidification through the action of LAB. Although this is a fairly minor change, the result is a much more appetizing porridge, especially with the help of a little crude sugar. Most such mixtures of flours and water will develop substantial populations of LAB, alone or associated with yeast cells. Microbial gas production will cause some expansion of the paste or dough, and give a tasty and appealing product upon cooking. The most obvious example of this type is of course sourdough bread. It seems reasonable to believe that this is a very ancient product, and archeological evidence is emerging to support this view. A simple improvement would be to add a part of a good ferment to a fresh mixture of flour and water. This process, sometimes called 'back slopping,' remains the basis for much sourdough bread production to the present.

A more dilute mixture of flour and water would give a crude alcoholic drink. The really important development here was the discovery that germinating the seeds greatly increased the alcoholic content, and the general quality of the resulting drink. Beer production in the west further modifies the properties of the drink by drying and heat treatment of the germinated seeds (barley malt). African sorghum beer does not require this added refinement, and it seems reasonable to think that the earliest barley fermentations would have used 'green' (undried) malt. Again the fermentations would have used a mixture of LAB and yeasts. It is clear that from fairly early times residual yeast from beer and wine fermentations was used for bread making. Equally, *kvass* fermentation from rye bread and water uses a portion of sourdough bread ferment to start a new brew going, although once started, it is maintained by 'feeding' the ferment with additions of bread and water.

Rice flour mixed with legume seed flour and water will ferment with leavening provided by carbon dioxide generated by heterofermentative LAB. In parts of India this simple process yields products such as idli and dosa.

Thus, seeds rich in starch are converted to many foods and drinks by a range of fermentations which depend on yeasts and LAB, alone or in associations. Fermentations of protein-rich seeds, particularly the soybean (*Glycine max*) and the ground nut (peanut, monkey nut, *Arachis hypogea*), of tropical and subtropical origin, together with the seeds of tropical leguminous trees, have provided both staple foods (such as tempeh and oncom (ontjom, onjom, etc.)) and flavoring materials (soy sauce, miso, etc.) throughout the tropics and subtropics for many thousands of years. The driving force for such developments was probably the rather bland nature of the starting materials, but we now know that fermentations can increase the digestibility of constituents of the seeds, particularly the proteins, enrich the products with extra B-vitamins, decrease concentrations of some toxic or antinutritional factors, and increase the availability of essential mineral nutrients. These fermentations are frequently multistage, using mold fungi to manufacture a range of digestive enzymes, then yeasts and LAB to complete the process. The origins of these fermentations are lost in time, but we know that soy sauce and miso have existed for at least 3000 years. Legume seeds tend to be rather well supplied with antinutritional factors, and detoxification during the fermentation is a significant benefit of the process. Perhaps the most remarkable example of seed detoxification, however, concerns the castor bean (*Ricinus* spp.). This seed contains a wealth of toxic materials, including ricinoleic acid (a powerful purgative), strongly allergenic proteins, and ricin. Despite this, in parts of Nigeria the seeds are collected, boiled, dehulled, wrapped in leaves, and allowed to ferment, then used as a food flavoring, apparently without any ill effects on its consumers.

Types of Fermentations

Although space constraints forbid a detailed discussion, it is worth noting the range of fermentation procedures developed for dealing with foods. The simplest are probably the liquid fermentations used in making drinks such as beers, wines, and some of the milk products such as yogurt and buttermilk. These are sometimes called suspension cultures because the microbial cells float suspended in the surrounding liquid. Another name is stirred tank reactor (STR) fermentations. Fermentations involving molds usually require free access to the air, both to meet the organism's need for abundant oxygen, and to remove the products of the fermentation, principally carbon dioxide and heat. In one case however (tempeh) the process requires a very limited exposure to oxygen, traditionally achieved by wrapping the substrate in banana or teak leaves, although plastic bags pierced with a pattern of fine holes now often replace these.

Such fermentations, which do not require free liquid water, are described as solid substrate fermentation (SSF). Some lactic fermentations, for example of starchy or farinaceous materials, might also be placed here, as might sausage fermentations. When a solid substrate is in free water, it appears that this is just a type of suspension culture, but there is evidence to suggest that the active microbes attach themselves to the solid material to a considerable extent, thus producing some characteristics of a SSF. The kefir grain provides an interesting problem. Attachment of the cells to the polysaccharide membrane suggests SSF, but analysis of the fermentation shows that there are large numbers of cells suspended in the liquid, quite free of the grains, and that the proportions of bacterial species in suspension are radically different from the proportions within the grains. In Orleans vinegar production the acetification is effected by a layer of bacterial cells growing as a film or skin at the interface between water and air, so exhibiting some characteristics of a SSF. The 'quick' process features a layer of cells on the surface of the wood shavings, irrigated by the recycling liquid, but also supplied with abundant oxygen, thus showing an even closer relationship to SSF. Citric acid production by *Aspergillus* used to require static growth on the surface of a liquid layer, rather like the Orleans vinegar process. Some processes were described using growth on bamboo supports and a liquid recycle like the 'quick' vinegar process, but it is not clear if these were applied on a production scale.

The Future for Traditional Fermentations

Some very ancient fermentations continue to flourish and develop on a very large scale. Soy sauce production almost seems capable of expansion without limit, as does cheese. Yogurt has been a remarkable success story, with market penetration far beyond its traditional homelands. Vegetable fermentations, such as sauerkraut and olives, also seem to grow, although they perhaps do so more slowly than some of the other cases cited. The sourdough bread story is another remarkable one, with the product now on offer at (it seems) every airport in the USA. There does not seem to be any fear for the future of these and a selection of other products.

On the other hand, I feel substantial concern for some of the more modest traditional fermentations found in the developing countries. These have made a significant contribution to the people's nutrition in the past, and the example of tempeh shows how they can substantially improve the nutritional status of the poorest folk. Tempeh itself seems to be robustly addressing the challenge offered by modernizing Indonesian society, but this is an exception. Sorghum beer in South Africa is another example where (in this case with intelligent support from central government) a traditional food item continues to enjoy a place in an evolving society's markets. In many cases, however, modernization tends to displace traditional foods with imported ones perceived as more 'western' by precisely those consumers who are least able to afford the change to imported food fashions. The drift from rural to semiurban living exacerbates the situation. The traditional food fermentations are essentially rural, low-technology processes conducted at an almost domestic level. Hygiene considerations alarm the western observer, but in general the adapted consumer population seems to suffer no harm from these products. There are exceptions of course. The continuation of poisoning outbreaks caused by consumption of tempeh bonkrek in Indonesia, despite the authorities' efforts to ban its production, demonstrate the kind of danger which exists. On balance, however, I feel that loss of traditional knowledge regarding the conduct and management of traditional fermentations will be a serious matter. This knowledge tends to be the preserve of women. Young people are not interested in learning about these matters, and emerging societies tend to place an even lower value on 'women's knowledge' than developing ones do. Courses on food, and fermentation need to place stress on the value of such information. This is perhaps the only way in which we can hope to conserve this unique knowledge pool.

See also: ***Bifidobacteria* in Foods**; **Fermented Foods**: Origins and Applications; Fermented Meat Products; Fermentations of the Far East; Beverages from Sorghum and Millet; Soy (Soya) Sauce; **Fermented Milks**: Types of Fermented Milks; Products from Northern Europe; Other Relevant Products; Dietary Importance; **Leavening Agents**; **Mycotoxins**: Classifications; **Soy (Soya) Beans**: Processing for the Food Industry; **Starter Cultures**; **Yeasts**

Further Reading

Bushell ME (ed.) (1984) *Progress in Industrial Microbiology*: vol. 19; *Modern Applications of Traditional Biotechnologies*. Oxford: Elsevier.

Campbell-Platt G (1987) *Fermented Foods of the World: A Dictionary and Guide*. London: Butterworths.

Campbell-Platt G and Cooke PE (eds) (1995) *Fermented Meats*. Glasgow: Blackie Academic and Professional.

Dirar HA (1993) *Indigenous Fermented Foods of the Sudan*. Oxford: CAB International.

Hesseltine CW and Wang HL (eds) (1986) *Indigenous Fermented Foods of Non-Western Origin: Mycologia Memoirs*, no. 11. New York: New York Botanical Garden.

Huang HT (2000) *Science and Civilisation in China, 6(v): Fermentations and Food Science*. Cambridge: Cambridge University Press.
Lee C-H, Steinkraus KH and Reilly PJA (eds) (1993) *Fish Fermentation Technology*. Tokyo, Japan: United Nations University Press.
Pederson CS (1979) *Microbiology of Food Fermentations*, 2nd edn. Westport, CT: Avi Publishing.
Reddy NR, Person MD and Saluniche DK (eds) (1986) *Legume-Based Fermented Foods*. Boca Raton: CRC Press.
Rose AH (ed.) (1982) *Economic Microbiology*, vol. 7: *Fermented Foods*. London: Academic Press.
Sherman PW and Flaxman SM (2001) Protecting ourselves from food. *American Scientist* 89(2): 142–151.
Steinkraus KH (ed.) (1989) *Industrialization of Indigenous Fermented Foods*. New York: Marcel Dekker.
Steinkraus KH (ed.) (1996) *Handbook of Indigenous Fermented Foods*, 2nd edn. New York: Marcel Dekker.
Sudarmajdji S, Suparmo and Raharjo S (eds) (1998) *Reinventing the Hidden Miracle of Tempe*. Jakarta: Indonesian Tempe Foundation.
Wood BJB (ed.) (1998) *Microbiology of Fermented Foods*, 2nd edn. Glasgow: Blackie.

Fermented Meat Products

F-K Lücke, Fachhochschule, Fulda, Germany

Introduction

Fermented meat products have been subjected to the action of microorganisms or tissue enzymes so that they are modified significantly in a desirable fashion. They are subdivided into fermented sausages (products made from comminuted meat stuffed into casings) and fermented unground meats. This article provides basic information on the manufacture of fermented meat products and on the factors affecting their safety and quality.

Types of Fermented Meats

Fermented Sausages

Fermented sausages are manufactured from comminuted lean and fatty tissue, mixed with salt, spices, sugar, and, in most cases, curing agents (nitrite, nitrate, ascorbate). Particularly in commercial production, starter cultures are also added. The mix is then filled into casings and subjected to a microbial fermentation. Many, but not all varieties of fermented sausages are subsequently aged to dry them and to develop the desired sensory properties. Another criterion for the classification of fermented sausages is the surface treatment (e.g., smoking, mold growth). **Table 1** gives a classification of fermented sausages, along with the definitions of terms used in this article.

An official classification of fermented sausages is important to avoid misleading the consumer and unfair competition between manufacturers. Different parameters are used in different countries; however, they are all more or less related to the degree of drying and the percentage of high-quality raw material (lean muscle tissue) in the formulation. For example, the German Code of Practice (Leitsätze) specifies the minimum content of collagen-free lean meat for different sausage varieties while key parameters in France are the ratio between collagen and protein, the fat content and the percentage of moisture, based on fat-free dry matter.

Fermented Unground Meats

To produce aged hams and related raw cured meats, salt is added by rubbing it on to the meat surface or by immersing the cut into a brine or by injecting brine

Table 1 Classification of fermented sausages

Class	*Final water activity*	*Total ripening time*	*Use of smoke*	*Mold growth on surface*	*Examples*
Dry	< 0.90	> 4 weeks	No	Yes	Traditional Italian salami, French saucisson sec
Dry	< 0.90	> 4 weeks	Yes[a]	Yes	Traditional Hungarian salami
Dry	< 0.90	> 4 weeks	Yes or no	No	German Dauerwurst
Semidry	0.90–0.95	< 4 weeks	No	Yes	Various French and Spanish fermented sausages
Semidry	0.90–0.95	Usually 10–20 days	Yes (with exceptions)	No	Most fermented sausages in Germany, the Netherlands, Scandinavia, USA
Undried	0.94–0.97	< 2 weeks	Yes or no	No	German Teewurst, Spanish sobrasada, Thai nham[b]

[a]During fermentation.
[b]Normally cooked before eating.

into the tissue. Then, cuts are kept at low temperatures (10 °C or below) until the salt is evenly distributed. During subsequent aging at elevated temperatures (15–30 °C), they not only lose some moisture but are also tenderized by tissue proteases. Only such aged products may therefore be termed fermented. They are normally eaten raw, and include raw (pork) ham, but also Italian coppa and some products from beef muscle (e.g., Swiss Bündnerfleisch, Turkish pastirma).

The Manufacture of Fermented Sausages

Choice and Comminution of Raw Material

The microbiological quality of the raw material (lean and fatty tissue) should be controlled by continuously monitoring the temperature during meat handling and transport, by visual inspection of shipments, and by selecting suppliers following good hygienic practice in slaughtering and butchering. For the manufacture of fermented sausages, the pH of the raw material should be 5.8 or lower because higher pH values favor the growth of undesired or hazardous acid-sensitive bacteria. Initial pH values of the sausage mix between 5.8 and 6.0 may still be tolerated if a rapid onset and a sufficient rate and extent of lactic acid formation are ascertained. To obtain products of good sensory quality and shelf-life, the fatty tissue used should be low in polyunsaturated fatty acids and peroxides. To minimize fat deterioration, the material is usually chilled or frozen in order to keep the temperature during comminution below 2 °C, and exposure to oxygen should be minimized during communition and filling.

Formulation and Filling

By removal of oxygen and addition of salt, the normal spoilage flora of aerobically stored fresh meat (mainly pseudomonads) is suppressed. With few exceptions, the target water activity (a_w) of sausage mixes is 0.955–0.965. This is mostly achieved by adding 30–35% fatty tissue (giving a moisture content of about 50% in the mix) and 2.5–3.0% salt. Higher initial a_w values favor growth of Enterobacteriaceae (including *Salmonella*) whereas at a_w below 0.955, acid formation by lactic acid bacteria may be too slow to suppress *Staphylococcus aureus* reliably, a bacterium which still grows well at this a_w.

For most semidry fermented sausages, nitrite is used as curing agent at input levels between 100 and 150 mg $NaNO_2$ kg^{-1}, normally in combination with 300–500 mg sodium ascorbate per kg. Nitrite contributes to the inhibition of Enterobacteriaceae early in fermentation but has little effect on *Staphylococcus aureus* and on lactic acid bacteria. Moreover, by reacting with heme and nonheme iron, nitrite brings about the development of the desired curing color and curing aroma. Dry sausages are often made with nitrate rather than nitrite or with reduced input levels of nitrite. Such products must be fermented at lower temperatures. There is no need whatsoever to add more than 150 mg sodium nitrite kg^{-1}, more than 300 mg potassium nitrate kg^{-1}, or to combine nitrite and nitrate.

Carbohydrates are added to the mix to provide an energy source for lactic acid bacteria. Addition of 0.3% glucose or another rapidly fermentable mono- or disaccharide will enable the lactic acid bacteria to lower the pH from 5.8 to 5.3 or below, which is normally sufficient for microbiological safety. If the initial pH of the mix is higher (5.8–6.0) and/or a lower final pH is desired, the input of rapidly fermentable sugar is increased to 0.5–0.8%. The acidulant glucono-δ-lactone (GdL) is only used for products with a short shelf-life because many strains of lactic acid bacteria ferment this compound to lactic and acetic acid, and acetic acid interferes with reactions leading to and stabilizing desired sensory properties. Also, use of excessive amounts of fermentable sugars, in conjunction with insufficient drying and elevated aging and/or storage temperature, results in high levels of lactic and acetic acids, sometimes even in pore formation and swollen packs due to CO_2 production by heterofermentative lactobacilli. Lactobacilli may also contribute to the spoilage of fermented sausage by forming hydrogen peroxide which attacks heme compounds and polyunsaturated fatty acids. This in turn leads to gray or greenish discolorations and flavor defects, particularly if soft, improperly stored fatty tissue has been used.

In addition to their effects on aroma and flavor, some spices contain antioxidative compounds and may, by means of their manganese content, increase the rate of acid production during sausage fermentation.

During the 1970s and 1980s, it became common practice to use starter cultures in industrial sausage fermentations. Most commercial preparations contain a combination of lactic acid bacteria with nonpathogenic catalase-positive cocci. Such cultures have also been shown to be useful for the manufacture of a wide spectrum of traditional indigenous sausage varieties on a small scale, and also replaced, to a large extent, so-called back-slopping methods where a fermented sausage from the previous batch is used to inoculate the following batch. The disadvantage of the latter method is that desired catalase-positive cocci will tend to become diluted out, and strains with undesirable metabolic properties may be selected.

Suitable starter bacteria must be sufficiently active in the a_w range between 0.93 and 0.96 and at the fermentation temperature chosen (in Europe, 20–25 °C), and must be able to ferment the sugar added. To avoid excessive lag phases, resuscitation times in sausage mix must be short. Obviously, they must not be pathogenic or toxinogenic, and should not form biogenic amines nor metabolites detrimental to the sensory properties of the product.

Use of lactic acid bacteria is important for the safe production of rapidly fermented semidry sausages, whereas at low ripening temperatures their effect is small. Of the species of lactic acid bacteria that are commercially available, *Lactobacillus sakei* strains tend to acidify the sausages most rapidly, and to be most competitive towards acid-sensitive bacteria and other lactobacilli. Acidification rates of *Pediococcus pentosaceus* and *L. curvatus* depend much on the strain used, and may be equal to or higher than those of *L. plantarum* and *L. pentosus*. *P. pentosaceus* is often stated to form a milder aroma. In the temperature range between 20 and 25 °C, acid formation by *P. acidilactici* is too slow to be of much benefit. Suitable starters should thrive in the product while not forming gas, hydrogen peroxide, biogenic amines, or other compounds of sensory or toxicological concern.

Starters containing catalase-positive cocci are of most benefit to products prepared with nitrate, and products to have a long shelf-life. Adding sufficiently high levels (10^6–$10^7\,g^{-1}$) of nonpathogenic *Staphylococcus* or *Kocuria* spp. largely eliminates the need for growing these bacteria in the product by keeping the pH in the permissive range (> 5.3). Strains of *Staphylococcus carnosus, S. xylosus*, and *Kocuria varians* are commercially available. Whether or not a yeast (*Debaryomyces hansenii/Candida famata*) is useful depends very much on the aroma note desired. Surface inoculation with molds is advisable in the manufacture of mold-ripened products. Mold starters must be able to colonize the surface without sporulation and without formation of mycotoxins or antibiotics. Strains of *Penicillium nalgiovense* and *P. chrysogenum* have been selected for sausage fermentation.

Fermentation

Fermentation temperatures between 20 and 25 °C are most common in European-type sausage fermentation. For US-type summer sausage, higher temperatures (up to 41 °C) are used. Rapid onset and rates of acid formation must be ascertained to suppress pathogens such as *Salmonella* and *Staphylococcus aureus* under these conditions. Some traditional products are fermented at lower temperatures for longer times. Sausages expected to have a long shelf-life should be fermented at lower temperatures in order to restrict acid formation and to obtain maximum activity of microorganisms that reduce nitrate and are involved in the formation of aroma and flavor. As pointed out above, fermentation temperature must also be lowered if the initial a_w is outside the range between 0.955 and 0.965, if no nitrite is added, or if the pH is not to drop below 5.3. For genuine Hungarian salami, fermentation temperatures are as low as 12 °C.

Early after stuffing, residual oxygen is consumed by meat enzymes, and oxygenated myoglobin is turned into brownish metmyoglobin. Nitrite accelerates metmyoglobin formation but slows down the former process. Subsequently, acid formation starts, and catalase-positive cocci reduce nitrate to nitrite. Acid favors the reaction of nitrite with metmyoglobin to give rise to pink nitric oxide myoglobin. Residual nitrite is reduced by the sausage microflora. As acid formation continues, the pH drops to about 5.3. At this pH, growth of acid-sensitive bacteria, including catalase-positive cocci, are inhibited, and the water-binding capacity of the mix becomes minimal. Then, fermentation is usually slowed down by adjusting the temperature to about 15 °C and by lowering the relative humidity (RH) in the chamber, thus reducing the a_w of the sausages. The flavor of sausages consumed after little or no aging is dominated by lactic acid and some acetic acid, and by compounds determining the flavor of fresh meat.

Surface Treatment and Aging

After fermentation, most sausages produced in central and northern Europe are smoked. This leads to inhibition of microbial growth on the surface and to a delay of oxidative changes. On unsmoked products common in France and Italy, but also in Spain and Switzerland, a surface bloom develops, mainly consisting of salt-tolerant yeasts (e.g., *Debaryomyces hansenii*) and of molds. This layer also protects the product from oxygen and facilitates drying.

Sausages are usually aged at 12–15 °C. The climate should ensure a slow but steady removal of moisture from the sausages and prevent undesired mold growth on the surface while avoiding uneven drying of the product and concomitant undesired changes in appearance and texture. Hence, the RH in the chamber is gradually lowered to about 75–80%, and air velocity is adjusted to about $0.1\,m\,s^{-1}$. Most sliceable semidry fermented sausages are dried to moisture contents of about 40%, which corresponds to a weight loss of about 18% and an a_w value around 0.93. Dry sausages ($a_w < 0.90$) have 35% moisture or less, corresponding to a weight loss of 25% or more. Obviously, more water must be removed if the formulation includes less fat and more lean meat than common.

During aging, the flavor and aroma characteristics for fermented sausages are formed and stabilized. Lipids are precursors of many unbranched aldehydes, 2-alkanones, and short-chain unbranched fatty acids. The process involves tissue lipases, autoxidation reactions, and the transformation of the reaction products by microorganisms. Use of nitrite results in partial inhibition of oxidative changes of lipids, and in a shift in the reaction products. Tissue proteases split proteins into peptides that are subsequently transformed by microorganisms into amino acids and branched-chain volatile fatty acids. Ethyl esters also contribute to sausage aroma. Sufficient activity of catalase-positive cocci is important for the development and stabilization of the desired sensory properties. To obtain a sausage with full aroma and flavor, it is therefore important to limit rate and extent of acid formation and to age the product for a minimum of several weeks. If undesired mold growth is prevented, oxidative changes in the lipid fraction (rancidity) limit the shelf-life of most semidry or dry sausages.

Safety of Fermented Sausages

Hazards to be controlled during manufacture of fermented sausages are listed in Table 2 and analyzed on the basis of exposure, severity of disease, and epidemiological data, i.e., reports on outbreaks due to the consumption of fermented sausage.

Preventive Measures and Critical Limits to Control Microbiological Hazards

Figure 1 provides a flow diagram of the manufacturing process and includes preventive measures against growth of *Salmonella* and other pathogenic Enterobacteriaceae and *Staphylococcus aureus*.

The critical limits given have been validated by commercial practice. However, they do not cover the whole range of types of fermented sausages. Lowering one hurdle (e.g., addition of less salt or curing agents) must be compensated for by fortifying another hurdle (e.g., by increasing the rate of acid formation). Experience shows that incriminated products have been either fermented at too high temperatures without starters and with nitrate rather than nitrite, or have been eaten after a very short fermentation with little, if any, drying.

In contrast to *Salmonella*, the infective dose of enterohemorrhagic strains of *Escherichia coli* (EHEC) is low. Hence, sausages containing EHEC are a hazard to the consumer, even if growth of EHEC is reliably suppressed during fermentation. The presence of EHEC in meat, especially in beef and lamb, cannot be excluded, and the extent of destruction of EHEC during sausage ripening ranges from less than 1 log (in undried products) to 3–4 log (in smoked dry products of low final pH). Hence, it is crucial to minimize fecal contamination of beef and lamb carcasses during skinning and evisceration. Moreover, beef- or lamb-containing raw sausages with short ripening times should be excluded from the diet of small children (the main risk group for hemorrhagic colitis and its life-threatening sequela disease, hemolytic–uremic syndrome).

Protein degradation by tissue enzymes during prolonged storage of raw meat, in conjunction with growth of psychrotrophic lactic acid bacteria capable of histidine decarboxylation (in particular, certain strains of *Carnobacterium* spp. and of *Lactobacillus curvatus*), appears to be the main risk factor for the formation of biogenic amines. Hence, the formation of these amines can be minimized by proper selection of raw materials and starter cultures.

Table 2 Analysis of microbial hazards during the manufacture of fermented sausages[a]

Agents	*Growth potential during fermentation*	*Prevalence in 25-g samples*	*Minimum cell density for disease*		*Severity of disease*	*Epidemiological evidence*	*Risk priority*
			Normal	*Risk groups*			
Salmonella	Low	0.1–5%	Medium	Low	Low	Yes	High
Enterohemorrhagic *Escherichia coli*	Low	0.001–0.1%	Low	Very low	High	Yes	High
Listeria monocytogenes	Very low	10%	Probably high	Low	High	No	Moderate
Staphylococcus aureus	Low	10%	High (toxin formation in product)		Very low	Yes	Moderate
Yersinia enterocolitica[b]	Low	<1%	Probably high		Low	No	Low
Bacteria forming biogenic amines	Yes[c]	Yes	High (toxin formation in product)		Very low	No	Low
Molds forming mycotoxins	Low	Low	High (toxin formation in product)		Chronic	No	Low[d]

[a]Spore-forming bacteria and *Campylobacter jejuni* are no significant hazard.
[b]Pathogenic serotypes.
[c]Strains of lactic acid bacteria.
[d]Mainly relevant for mold-fermented types.

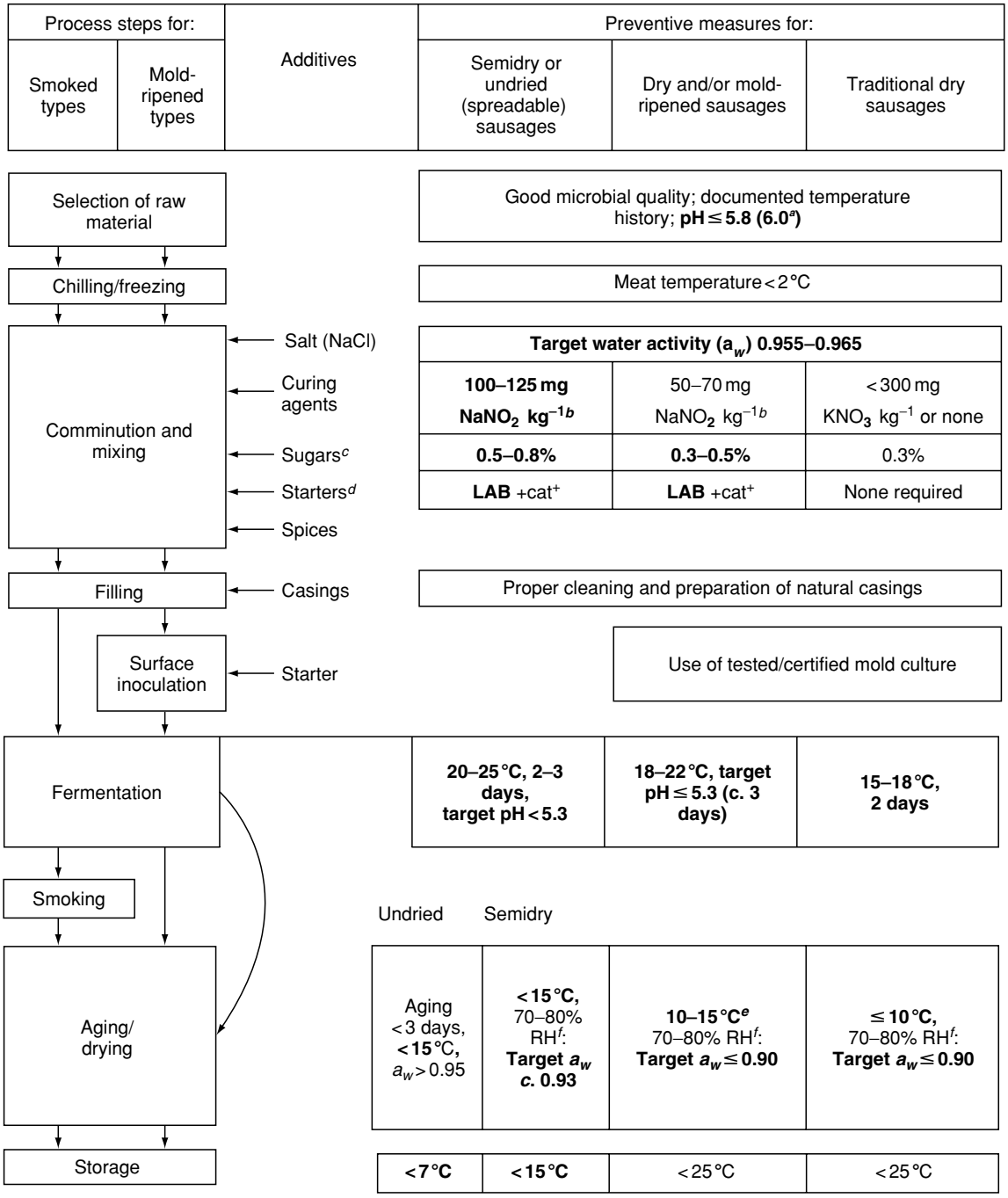

Figure 1 The manufacture of major types of fermented sausages, and measures preventing growth of *Salmonella, Staphylococcus aureus*, and *Listeria monocytogenes*. Critical limits at critical control points are printed in **bold**. [a]Adjustment of sugar addition and fermentation conditions may be necessary. [b]With maximum 500 mg sodium ascorbate kg^{-1}. [c]Rapidly fermentable by the starter used: 0.8% should be used if meat pH is above 5.8. [d]LAB, lactic acid bacteria; cat^+, catalase-positive cocci (*Staphylococcus, Kocuria*). Starters must be active at the fermentation temperature selected. [e]For mold-ripened products, temperature should be lowered to about 10 °C after 2 days of fermentation. [f]RH, relative humidity. Adjustment depends on pH and a_w of sausage. Modified from Lücke FK (2000a) Fermented meats. In: Lund BM, Baird-Parker AC and Gould GW (eds), *The Microbiological Safety and Quality of Food*, pp. 420–444. Gaithersburg: Aspen.

Mold growth on fermented sausages may be inhibited by appropriate control of temperature and RH during aging, storage, and distribution, by excluding oxygen from the packages, and/or by surface treatment with smoke, permitted antifungal agents, or waxes. Products to be mold-ripened should be inoculated with a suitable nontoxic mold strain and dried at temperatures below 15 °C.

Viruses already present in livestock at the time of slaughter normally do not affect humans. However, since these viruses are, as a rule, only slowly inactivated during sausage ripening, fermented meats should, like unprocessed meats, not be exported from areas where viral diseases of meat animals prevail. Viruses pathogenic for humans also retain their infectivity during sausage fermentation, and, as in all cases where food is handled, contamination of the food by human feces must be avoided by proper personal hygiene.

Parasites (protozoa, nematodes, tapeworms) that may occasionally escape detection during inspection of the animals and their meat at the slaughterhouse are inactivated by low a_w values as prevail in dried sausages, and are reliably absent from meats dried to a_w values of 0.90 or below. Parasites are also unlikely to survive the a_w and pH values typical for most semidry sausages. However, in the USA, pork is not generally inspected for the absence of *Trichinella*, and regulations stipulate that semidry and undried fermented sausages must be made from meat which has been stored frozen, or be heated to 58.3 °C (137 °F).

The Safe Manufacture of Fermented Unground Meats

For the manufacture of raw hams and comparable products, it is very important to select cuts of normal pH (≤ 5.8), particularly for large pieces; otherwise, salting proceeds slowly, and the risk of growth of pathogens during the salting process increases unless special precautions are taken.

In contrast to fermented sausages, changes in lipids are of little significance, and the microbiology and sensory characteristics of raw ham are little affected by nitrite. Hence, many aged products with long shelf-life are made with nitrate rather than nitrite, or with no curing agent at all.

The method of salting depends on the type of product and on local tradition. Injection of brine is uncommon in the manufacture of products to be aged and dried. Dry salting proceeds only slowly, but is preferred for most high-quality, expensive products. The meat is salted until sufficient salt has diffused into its core. During this period and the following equilibration time, the temperature must be low (≤ 5 °C in the core of large pieces) in order to prevent growth of any psychrotrophs (including nonproteolytic *Clostridium botulinum*) possibly present in the interior of the meat. Target a_w is 0.96 or below in all parts of the cut. The time required depends on the geometry of the cut, the concentration gradient of salt, and the ratio between meat and brine. There is little microbial activity in the ham under these cool, salty conditions. However, in recycled brines, a population may develop containing halotolerant psychrotrophic Gram-negative bacteria, such as *Halomonas* spp. Using such brines may improve the sensory quality of nitrate-cured hams because *Halomonas*, unlike other bacteria, actively reduces nitrate under the conditions prevailing in concentrated curing brines, and may also form aroma precursors.

After salting and equilibration, excessive salt and moisture are removed from the surface. To prevent growth of *Staphylococcus aureus*, this should be performed at 18 °C or below. Subsequently, the hams may be smoked (e.g., Westphalian ham) or not (e.g., Parma or Serrano ham), and dried to the desired a_w. Once they are microbiologically stable, top-quality hams are further aged at higher temperatures (25–30 °C) to make them very tender and tasty. If the surface is kept moist enough, a surface bloom of catalase-positive cocci, molds, and yeasts may develop and affect the sensory properties of the ham; otherwise, there is little, if any, role of microorganisms in the development of aroma and flavor during aging.

See also: ***Escherichia coli***: Food Poisoning; **Food Poisoning**: Classification; Tracing Origins and Testing; Statistics; Economic Implications; **Hazard Analysis Critical Control Point**; **Lactic Acid Bacteria**; ***Listeria***: Properties and Occurrence; **Meat**: Preservation; Eating Quality; Analysis; Nutritional Value; Hygiene; Extracts; **Mycotoxins**: Occurrence and Determination; **Starter Cultures**; **Zoonoses**

Further Reading

Campbell-Platt G and Cook PE (eds) (1995) *Fermented Meats*. London: Blackie Academic and Professional.

Demeyer DI (1992) Meat fermentation as an integrated process. In: Smulders FJM, Toldrá F, Flores J and Prieto M (eds) *New Technologies for Meat and Meat Products*, pp. 21–36. Nijmegen: Audet Tijdschriften.

Geisen R (1993) Fungal starter cultures for fermented foods: molecular aspects. *Trends in Food Science and Technology* 4: 251–256.

Hammes WP and Knauf HJ (1994) Starters in the processing of meat products. *Meat Science* 36: 155–168.

Incze K (1998) Dry fermented sausages. *Meat Science* 49(suppl. 1): S169–S177.

Jessen B (2000) Meat starter cultures and meat product manufacturing. In: Francis FJ (ed.) *Wiley Encyclopedia*

of Food Science and Technology, pp. 1614–1622. New York: Wiley.
Leistner L and Gorris LGM (1995) Food preservation by hurdle technology. *Trends in Food Science and Technology* 6: 41–46.
Lücke FK (1998) Fermented sausages. In: Wood BJB (ed.) *Microbiology of Fermented Foods*, 2nd edn, pp. 441–483. London: Blackie Academic and Professional.
Lücke FK (2000a) Fermented meats. In: Lund B, Baird-Parker AC and Gould GW (eds) *The Microbiological Quality and Safety of Food*, pp. 420–444. Gaithersburg, MD: Aspen.
Lücke FK (2000b) Utilization of microbes to process and to preserve meats. *Meat Science* 56: 105–115.
Ordóñez JA, Hierro EM, Bruna JM and de la Hoz L (1999) Changes in the components of dry-fermented sausages during ripening. *Critical Reviews in Food Science and Nutrition* 39: 329–367.
Prochaska JF, Ricke SC and Keeton JT (1998) Meat fermentation – research opportunities. *Food Technology* 52(9): 52–57.
Varnam AH and Sutherland JP (1995) *Meat and Meat Products*. London: Chapman & Hall.

Fermentations of the Far East

W R Stanton, Stanbury, Keighley, UK
J D Owens, The University of Reading, Reading, UK

Background

The origin of oriental food fermentations is lost in antiquity, although there is indirect evidence that the earliest 'cultivated plants' were microorganisms, fermentation being part of the hunter–gatherer practices preceding agriculture. As mentioned later, ideas of what is wild and what is cultivated have changed radically in recent years; most so-called virgin tropical forest is in fact carefully managed by the local people, who have taken advantage of their 'human' skills to extract and detoxify products, whilst leaving them in the wild, biochemically protected against the biota. No precise archeological timing can be assigned, because the techniques adopted by humans were acquired by observation of the behavior of other animals. However, in order to understand these Far Eastern foods, it may be helpful to visualize them as occupying two main groups, a Chinese heartland and migrant-spread group and an indigenous products group.

Chinese Heartland and Migrant-Spread Group: Panoriental Fermented Foods

The Chinese heartland and migrant-spread group consists of those foods that are identifiable as originating in China and, to a lesser extent, the Indian subcontinent. They are encountered in the Far East, following the spread – over the last 2000–3000 years – of migrants from these two founts of culture, mainly south- and eastwards from the temperate and monsoonal regions in which they putatively originated (**Figure 1**).

Because these foods became, through the process of transfer, out of context, as far as both climate and substrate were concerned, they underwent subtle changes:

- in the organisms which took part in the fermentations;
- in the duration of the fermentations;
- in the utensils, materials, and conditions for operating the processes.

For example, the Mucorales such as *Rhizopus*, liking warm humid conditions, replaced *Aspergillus* and *Penicillium*. Bamboo, the banana leaf, and the coconut shell replaced pottery.

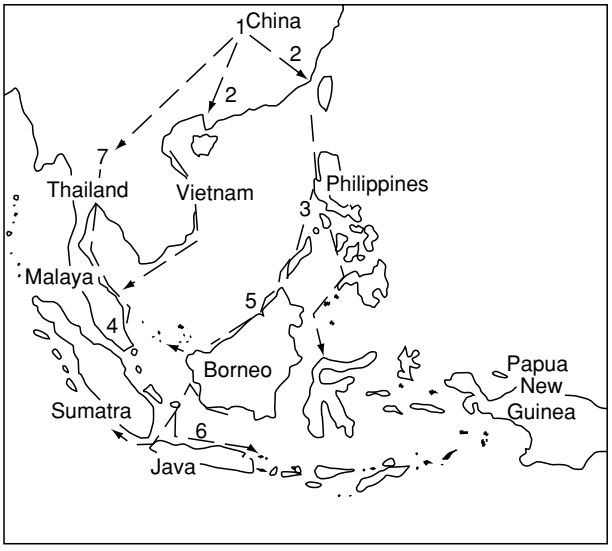

Figure 1 Human migration and the spread of food fermentation techniques in the Far East. 1, Postulated center of origin, 3000–4000 years ago; 2, Land routes within the empire to the 'vassal' kingdoms of Indochina; 3, Sea routes and migratory routes hugging coasts of the South China Sea, or traversing open sea by island hopping; 4, An ancient porterage route across South Malaya; 5, Small victualling and trading ports along the Northwest coast of Borneo; artefacts from this area show strong connections with Chinese trade; 6, Java: a territory rich in traditions of fermented foods; the raw materials, techniques and nomenclature provide evidence for the 'Chinese origin' hypothesis; 7, In addition to the coastal trade, there is a strong tradition of overland migration into Thailand as well as *in-situ* development of richly endowed indigenous cultures, such as the Khmer. After Stanton WR (1985) Food fermentation in the tropics. In: Wood BJB (ed.) *Microbiology of Fermented Foods*, pp. 193–211. London: Elsevier, with permission.

Nevertheless, the links of the tropical foods with their continental cultural roots may be seen in the form of the 'new' foods and the processes used for their manufacture, just as clearly as the links with the cultural and linguistic traits of the Indian subcontinent and China are themselves identifiable.

In-situ (Tropical) Group: Indigenous Products

Much earlier migration initiated the use of the great diversity of plant and animal resources of the equatorial lands of Southeast Asia and their associated freshwater and marine biotopes. This process took place over, putatively, the last 30 000 years, i.e., over the period of the spread of Meso- and Neolithic cultures through the South and Southeast Asian regions. The separation of the two groups of foods is incomplete, but this systematic approach is helpful in understanding the recipes and the current concern over the loss of species diversity, through degradation of habitats with the expansion of the human population, and thus the potential loss of novel raw materials.

Role of Food Fermentation in the Humid Tropics

The function of both groups in the diet has been, and remains, to exploit otherwise toxic or inedible raw materials, to preserve the harvest, to recover wastes and generate attractive, stimulating, and accessory foodstuffs, and to conduct these exercises with economy of expenditure and energy. Figure 2 shows how fermentation fits the scheme of preservation techniques.

Panoriental Fermented Foods

Typical of panoriental fermented products – now used as foods worldwide – are those liquid relishes derived from the soy bean, of which soy sauce is but one of many 'sho', i.e., salt-assisted, processes. The soy bean furnishes the type of substrate for transformation by fermentation, embodying many of the characteristics of a plant raw material requiring processing before it is edible.

Soy sauce is the best known, but there are also solid-state bean fermentations, the *vegetable cheeses*, such as tempe, which have evolved to become 'meat substitutes.' The role of these bean cheeses in diets is more that of balancing, in a cereal diet, vitamin and essential amino acid deficiency. It includes improving digestibility (reducing bean flatulence) and eliminating antinutritional factors, particularly antitryptic factors.

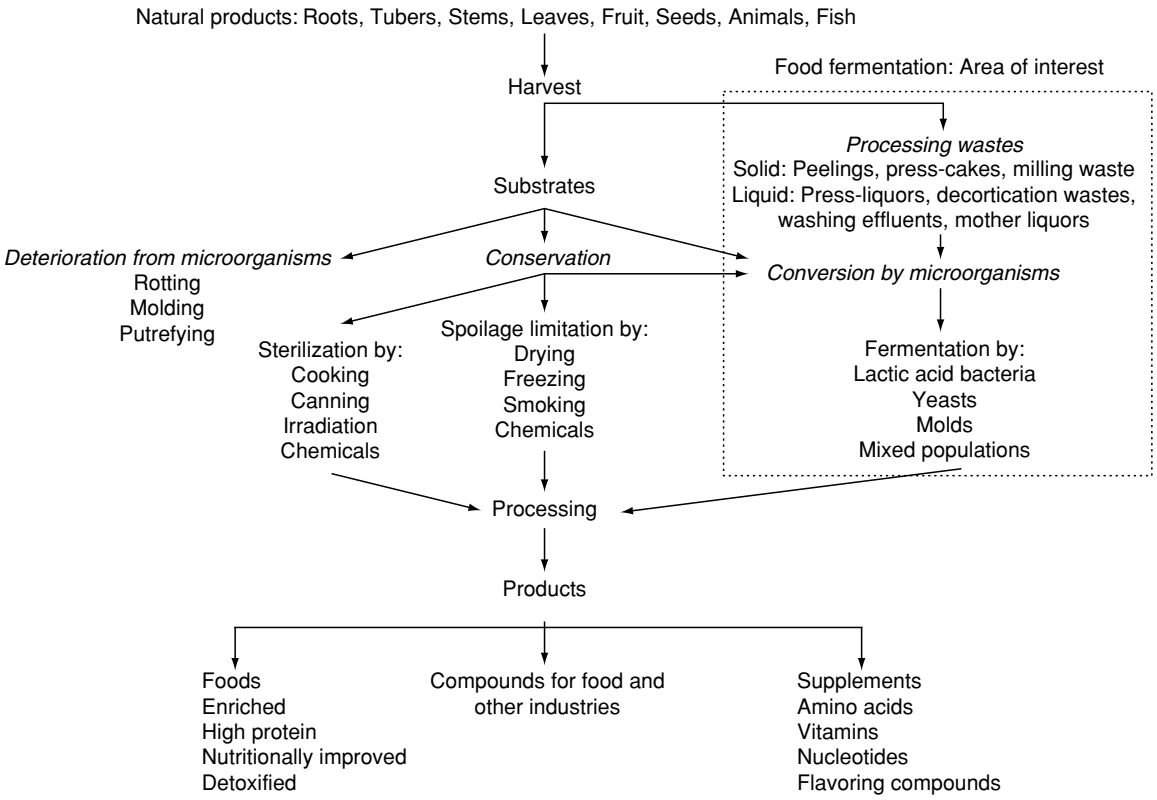

Figure 2 Fermentation in relation to other processes used in the conversion of raw materials to foods. After Stanton WR (1969) Some domesticated lower plants in Southeast Asian food technology. Adapted from Ucko PJ and Dimbleby GW (eds) *The Domestication of Plants and Animals*, pp. 463–469. London: Duckworth.

Tempe

Tempe is a traditional Indonesian solid-substrate fermentation in which soybeans are hydrated and acidified, dehulled, cooked, and then fermented with *Rhizopus* spp. The cotyledons become covered and penetrated by a dense, white, nonsporulating mycelium, which binds them into a compact, sliceable mass. Tempe is cooked by frying, steaming, or boiling, and is consumed as a main ingredient of a balanced meal with rice or starchy tubers and vegetables or as a snack food. It is estimated that 50% of the 200 million Indonesians consume tempe daily as a meat substitute. The high protein content and pleasant, relatively bland taste has led to it occupying a small, but expanding part of the vegetarian market in Japan, the USA, and Europe.

Manufacture Unlike soy sauce manufacture, the fermentation of tempe is a short-duration process. The small-scale manufacture of tempe is carried out as follows:

1. Wash soy beans.
2. Soak for 24 h at tropical ambient temperature ($30 \pm 5\,^{\circ}C$), during which time a natural lactic acid fermentation occurs, lowering the pH of the beans to 4.5–5.3.
3. Dehull by treading or by machine.
4. Wash and float off hulls.
5. Cook the cotyledons by boiling or steaming for 0.5–2 h.
6. Drain well and cool to below 35 °C.
7. Inoculate with *Rhizopus* spp.
8. Pack into perforated polythene bags or wrap in banana leaves.
9. Ferment at ambient temperature ($30 \pm 5\,^{\circ}C$) for about 24 h to yield tempe.

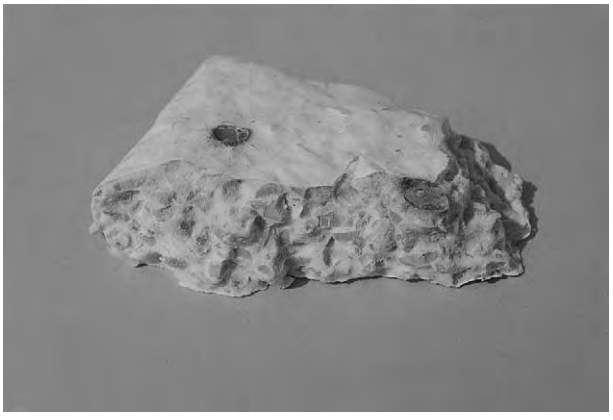

Slide 1 (see color plate 52) Tempe. Growth of a fungus, *Rhizopus oligosporus*, binds cooked, dehulled soybeans into a solid cake. Tempe is a traditional Indonesian meat substitute that also has a small part of the vegetarian market in Japan, the USA, and Europe.

Acidification is important to avoid misfermentation and invasion of undesirable bacteria. Perforated plastic, normally polythene, was first used by Hesseltine – a pioneer of the study of the industrialization of oriental fermented foods – in the late 1960s; it is possibly one of the most elegant examples (in its simplicity and small-scale applicability) of advances in solid-state fermentation technology during the twentieth century. Although primarily a mold fermentation, most commercial tempe also includes a substantial bacterial population. There are suggestions that the bacteria fulfill useful functions, such as the production of vitamin B_{12}, but their presence is not necessary for the production of good tempe. The composition of tempe is given in Table 1. The wide range of values shown in Table 1 is attributable to a diversity of factors, including variation in the quality of the beans and the extent of the use of additives and adulteration. However, under laboratory conditions, the protein content should be near the upper limit of the range of values shown, i.e., 50%, as loss of protein in preparation is generally below 2%. Apart from binding the cotyledons into a firm cake, the main changes in the beans effected by the fermentation are:

1. A net loss of 7–10% in dry matter. Approximately 15–20% of the initial substrate is used by the mold, yielding 50–100 g of dry biomass per kilogram of dry tempe.
2. Hydrolysis of a substantial proportion of the triacylglycerides to free fatty acids and mono- and diacylglycerides. In mature tempe, free fatty acids comprise 15–30% of total crude lipid. Fatty acids serve as a major source of carbon and energy for the mold, and around 25% of the initial lipids present in the cotyledons are utilized during the fermentation.
3. Hydrolysis of about 25% of the initial protein, most of which remains in the tempe as free amino acids. A small part is oxidized, with the release of ammonia and increase in the pH of the cooked cotyledons from the initial 4.5–5.3 to 6.5–7.0 in mature tempe.

Table 1 Composition of tempe

Nutrient	*Content (%, dry-weight basis)*
Protein	40–50
Crude fat	15–30
Carbohydrate	15–30
Crude fiber	5–12
Ash	1.5–3.0

4. The mold does not use the soy bean oligosaccharides, stachyose, raffinose, and sucrose, and the concentrations of these change little during the fermentation. However, a large proportion (70–80%) of the oligosaccharides in the original beans is lost by leaching during the hydration, washing, and cooking of the cotyledons.
5. Isoflavone glycosides are largely hydrolyzed to the aglycones, dadzein and genistein, which, it is suggested, are more available than the isoflavones in cooked soybeans. There is currently considerable interest in possible health-promoting properties of isoflavones.

Other Tempe-type Fermentations

Many of the fermentations make use of waste materials, such as oil extraction residues or other food processing by-products.

Common in western Indonesia is a product called oncom (ontjom). It is made by a tempe-like fermentation, but the process differs in some respects:

1. Peanut (*Arachis hypogaea*) press cake is used in place of the soy bean.
2. The preferred fungus is the orange–red-spored *Neurospora sitophila*.
3. Sporulation of the fungus is encouraged, in contrast to tempe manufacture, where it is avoided.

Oncom also may incorporate cassava press cake or the residue from soy milk preparation, or it can be prepared solely from soy milk residue. Coconut oil extraction residue is another processing material that can be fermented with *Rhizopus* spp., giving a product called tempe bongkrek. However, this may be lethal, if incorrectly prepared, because of the growth and toxin production by *Pseudomonas cocovenenans*.

Slide 2 (see color plate 53) Oncom (ontjom) in Bandung market, Indonesia. Oncom is similar to tempe but uses peanut press cake and the orange-spored fungus, *Neurospora sitophila*. The material in the background is fried tempe.

Natto and Related *Bacillus*-fermented Products

Natto is produced by an exclusively bacterial fermentation of the soy bean using *Bacillus subtilis*. In common with most other Japanese fermented foods, it originated in China, where it is named shi. In contrast to the preparation of tempe, the hulls are left on the steamed beans. In Japan, the manufacture is now highly automated and uses carefully selected pure starter cultures.

The cooked beans are inoculated with starter culture, packaged in small, usually polystyrene foam, containers and fermented at a relatively high temperature (40–42 °C) for 12–16 h. The packages are then cooled and held at 0–5 °C for 1–2 days for maturation. The result is a slime-coated product in which the beans remain visible and separate. The major biochemical change is the hydrolysis of protein to peptides and amino acids. Natto is eaten with boiled rice and is also used as a flavoring agent in other dishes.

Slide 3 (see color plate 54) Natto. A Japanese *Bacillus* proteolytic fermentation of whole soybeans producing a characteristically slimy, strongly flavored product.

Slide 4 (see color plate 55) Thu-nuo, dried disks. A traditional Thai *Bacillus* proteolytic fermentation of soybeans.

Similar *Bacillus*-fermented soy bean products are Thai thua-nuo and kinema, made in eastern Nepal, Darjeeling, Sikkim, and Bhutan, but in these cases, the products are still made by traditional methods. Consequently other *Bacillus* spp., as well as non-bacilli, commonly occur in the fermentations, and the products often develop strong ammoniacal odors. Thua nuo is prepared either as a paste or as dried disks, and kinema is sold in the markets as a paste. Neither is eaten directly as a condiment but they are used to flavor soups and stews, etc.

Other Cereal–Legume Fermentations

Cereal–legume mixtures are commonly used as the substrate for fermentations. Whereas, as explained elsewhere, wheat is added in soy sauce manufacture, rice is added in Japanese miso.

In contrast to the above fungal fermentations, from the Indian subcontinent come a number of rice-flour and bean-flour bacterial fermentations. These include two widely used products, idli and dosa.

South Indian Products The beans most commonly used for the Indian fermentations are Vigna spp., the green and black grams: *V. radiata* (L.) Wilcek (formerly *Phaseolus aureus* Roxb.) and *V. mungo* (L.) Hepper (formerly *Phaseolus mungo* L.). *Vigna* is an Old World genus, and the two grams are closely related to the cowpea or blackeye 'peas' (*V. unguiculata* (L.) Walp., formerly *V. sinensis* (L.) Hassk.). These beans may be confused in reports with beans of the unrelated New World genus *Phaseolus*, beans of which are not used. Any process development of grain legumes should be preceded by careful study of the correct identity of the raw material, as many legumes, unlike cereals, are very poisonous when incorrectly or inadequately processed.

Idli The beans and rice are normally milled separately, and water is added to give a thin gruel, or they may be soaked and wet milled, adding about 1% salt to the final gruel. The bean–rice mixture, in which the proportions vary widely from 10:1 to 1:10, is left covered overnight at 28–32 °C. The temperature is ambient shade temperature, although the duration of the fermentation is not as critical as for the tempe fermentation. The mixture may be stirred from time to time but should be kept anaerobic (microaerophilic) to encourage the growth of the heterofermentative lactobacilli. At the end of the fermentation, the slurry is poured on to an idli steamer surmounted by a perforated plate, in which there are shallow, saucer-shaped depressions some 10 cm in diameter and 2 cm deep. A cotton cloth prevents the slurry oozing through the depressions. The cloth also aids removal of the steamed cakes. Cooking for 15–20 min produces the raised cakes of idli, which are then ready to be consumed.

Dosa Similar fermented mixtures, in which lentils (*Ervum lens*) may be used in place of the gram beans, are prepared and poured on to a hot griddle pregreased with ghee, coconut, or other oil, according to culture.

These products are to be found wherever Indian communities have settled. They vary in flavor and in the added spices, but they are all lactobacillic, rather than yeast fermentations, possessing characteristics in common with sour doughs.

Indigenous Products

In contrast to the panoriental group, the indigenous products are a highly diverse group of foods and beverages, springing from the gathering activities of local people who have inherited, or even retained, recognizable Neolithic or bamboo culture traits and have made full use of their habitats rich in diversity of biota.

Cereal–Legume Fermentations

Cereals, principally rice and maize, and various legumes are involved (predominantly of the genera *Glycine, Vigna*, and *Arachis*). However, tree legumes are also used, as in Africa.

Starches of the Rainforest

Edible carbohydrate-containing plants are in short supply in the rainforests. Many plant species are too loaded with additives to be exploited by fauna other than those with specialized guts, such as leaf monkeys. This shortage limits the population carrying capacity. Those that are 'possible foods' may contain poisonous substances associated with their storage organs, e.g., oxalates. The forest and swamp starch-containing plants include the aroids (Araceae), such as the taros (*Colocasia*), the endangered elephant-foot yams (*Amorphophallus*), and the true yams, *Dioscorea* spp.

Tapé (Tapai)

Cassava (*Manihot esculenta* Cranz.) tubers are the most widely used material for the production of the popular sweet and sour alcoholic tapé, but this product is also made from glutinous rice, maize, or millet.

Manufacture The peeled cassava (tapioca in Southeast Asia) tubers or rice grains are inoculated, after boiling or steaming, by dusting with a microbial

Slide 5 (see color plate 56) Tapé in Bandung market, Indonesia. Cooked cassava roots fermented with fungi to produce a slightly soured, slightly alcoholic, and sweetened cassava.

starter powder, ragi. This is a stabilized mixture of yeasts and molds fermented as a dough and dried as small, flat cakes. Many types of ragi cake can be found in Southeast Asian markets, each type being applicable to a particular substrate. The inoculated cassava tubers are incubated in baskets, lined and covered with banana leaves. The inoculated rice is wrapped in small banana leaf or plastic packets. After 2–3 days at ambient temperatures (25–30 °C), the products soften and develop a sweet/sour slightly alcoholic flavor. The product may be sold for direct consumption, or after air-drying. Alternatively, the fermented mass is squeezed, and the paste and exudate are consumed separately; the latter, after further fermentation, is drunk as an alcoholic liquor.

Wines and Acidic or Alcoholic Beverages

Based on the ragi fermentation are the rice wines, such as tuak, of the peoples of Borneo. Some of the wines found in Southeast Asia are derived from the influence of migrating cultures, particularly those wines derived from rice and the inflorescences of palms, such as the coconut and talipot palm (*Borassus flabelliformis*), a feature of the landscape of Sri Lanka and Indo-China. Others, derived from regionally indigenous plants, are considered below. Reference to these is more often to be found in ethnographic and economic botany texts than in books on food.

Tropical forest people often 'tap' plants for a drink, using xylem and phloem saps. The rattans, for example, are a forest source of water. The evolution of fermentation of saps may have followed almost incidentally from the tapping of sugar-rich phloem saps from emerging palm inflorescences.

Products of the Vitaceae are occasionally found, but the Rosaceae of temperate zones are not represented.

Local fermented beverages are made from the saps of a number of genera of palms, such as *Arenga* and *Nypa* as well as from the talipot palm (*Borassus* spp.). Unlike in West Africa, the oil palm (*Elaeis guineensis*) is not tapped in Southeast Asia.

Other true rainforest trees, furnishing sugary and starchy substrates, are the jack and bread fruits (*Artocarpus*) and the durian (*Durio zibethinus*). The local Indo-Malay name for the fermented product of the durian is tempoyak. *Baccaurea grifithii* is another rainforest tree, producing an aromatic resinous fruit that is fermented to a retsina-like product.

Strong-smelling, biologically active, heterocyclic ring, sulfur-containing compounds (extending the 'garlic range' to many other families) are common in rainforest biochemical systems, the above quoted durian being but one example. Like the durian, they are locally exploited but are increasingly rarely encountered.

Fermented Sugar Cane Juice

Sugar cane (*Saccharum officinarum* L.) is indigenous to Southeast Asia, and the juice is fermented by the indigenous forest people, such as the Semelai of Malaysia. The fermentation is controlled by filtering the raw juice through a filter made of the inner bark of a dipterocarp tree, *Shorea curtisii*. The bark is rich in phenolic compounds, some of which are transferred to the liquor.

Detoxification via Fermentation

The recipes for these true forest products, and the plant species themselves, are in danger of being lost as the forests are cleared. The diversity of these forest products is reflected in the range of biochemicals that are detoxified by fermentation. These include the following: alkaloids and other heterocyclic, nitrogen-containing poisons – the nitriles and cyanides (the cause of optical and peripheral neuropathies), an unusual one being that *of Cycas* spp., or false sago, the seeds of which furnish a source of starch; poisonous amino acids, peptides and proteins; steroids; saponins; anthocyanins and related phenolics and organic acids such as oxalic. In the plant, these secondary compounds function as selective attractants, antibiotics, and feeding deterrents. As has been shown in the examples above, the bioactivity of some of these compounds has been exploited to advantage in the preparation of the fermented foods, but, as with so many aspects of tropical plant study, much remains to be recorded and analyzed.

Fermented Fish Products

Fermented, salted, or dried fish and crustaceae are used ubiquitously in oriental cooking. The products used range from solids and pastes to very salty liquids,

the salt being necessary for conservation, more than for culinary purposes. There is clearly a major subdivision into products in which the process solely involves autolysis by fish-derived enzymes, and those in which carbohydrate fermentations, employing microbial enzymes, also occur.

Salt-aided Fermentation

Outstanding amongst the desiccated products is katsuobishi, a mold-fermented product made from tuna in Japan. It is as hard as Parmesan cheese and is used in a similar way.

Belacan is a softer autolyzed shrimp paste from Indonesia and Malaysia; similar intermediate-moisture fish and shrimp pastes occur throughout Southeast Asia.

The salty fish sauces, such as the patis of the philippines, are autolytic products, relying on the enzymes of the fish gut rather than the associated microflora. These salt-aided autolytic products normally contain 14% or more of salt and are thus only of value as relishes to the main dish.

Carbohydrate-aided Fermentation

In some fish pastes and slurries of Southeast Asia, rice, usually as tapé or koji, also plays a part in the preservative process. The rice provides a carbohydrate source for a lactic acid fermentation and allows the use of less salt in the product.

Acceptance

Unlike soy sauce and, to a lesser extent, tofu, miso, and nata (the jelly-like sweet mentioned later), the acceptance of the fish products outside their home markets is limited, possible exceptions being the fish and shrimp pastes, such as belacan of Indonesia. Furthermore, the fish products mentioned above are not easily amenable to modification. Each depends on a particular local fish species, or group of species, and the taste for them is acquired, just as the predilection to the Scandinavian fermented herrings is very much an acquired taste, not even shared by the rest of Europe.

Nevertheless, in a world of increasing protein shortage and with an increasing proportion of by-catch and trash fish in the harvests, owing in part to over-fishing, as well as fish waste from processing, the opportunity to improve nutrition worldwide afforded by some of these products and processes should not be overlooked.

Pickled Fish

Unlike the above-mentioned products, pickled products rely on the addition of acetic acid and sugar, or acid fruit extracts, and do not play a significant role in the regional food supply.

Fermented Meats

Fermented meats are not important in the diet of the peoples of the Far East, in comparison with the fish products, although they are to be found in Thailand, for example.

Microbes as Food

Finally, in this review, the traditional uses of the microbial cultures themselves as the end product are considered. This is quite different from the development of industrial processes for the production of microbial protein, a large-scale, capital-intensive exercise requiring many more process steps than the traditional processes. The traditional foods are derived from bacteria or from fungal fruiting bodies.

Bacteria The polysaccharide and amorphous cellulose coat formed by *Acetobacter xylinum* is eaten in the Philippines under the general name nata. This is made by encouraging the growth of the aerophilic *Acetobacter* on sugar and ammonium phosphate-enriched acetic-acidified (pH 5) coconut water (nata di coco), pineapple (nata di pina), or other fruit juice.

Manufacture The substrate (a nitrogen- or phosphate-enriched sugared juice) is inoculated with the liquid phase of a previous culture (1:10 of substrate), held in a covered (not airtight) container, preferably undisturbed and in the dark. The liquor is maintained at a temperature below 31 °C, preferably about 28 °C, for approximately 2 weeks. The product is a surface cake of a bacterial embedded cellulosic gel, which acquires a particular flavor according to the substrate used. This gel, which can be up to 3 cm thick and is composed of cells plus polysaccharide coat, must be soaked to remove excess acetic acid before use. It is cubed, marketed fresh in a syrup, or marketed in cans, bottles, or plastic pouches.

Use of waste The bacterial process is a convenient way of making a useful product from a range of fruit-processing wastes, the character of which – unlike many of the fermentations mentioned here – is easily adjustable, appears internationally widely acceptable, and is already exported in limited quantities. However, it is a product that could be produced locally without much hazard and is acceptable to a wide range of palates.

Fungal Fruiting Bodies The fruiting bodies of basidiomycetes (and some ascomycetes) are mentioned

here, as they are ubiquitous foods of Southeast Asia, produced by fermentation and exploiting waste or inedible material. The main differentiation is between those fungi with ephemeral (autolysing) fruiting bodies, such as the paddy straw mushroom (*Volvariella volvacea*), and those with durable (dryable) mushrooms, such as the highly prized shiitake (*Lentinus edodes*).

Future

New Applications of Oriental Fermented Foods

The interest of food scientists in traditional fermentations arises not only from the fact that they furnish novel foods and are challenging from the point of view of the techniques needed to effect industrialization, but also from the fact that the unit processes have a wide industrial application. The enzyme, flavor enhancement, color, texture, gel stabilization, and antioxidant properties have come under scrutiny for the production of shelf-stable, intermediate-moisture foods, flavoring agents such as monosodium glutamate, flavor-enhancing nucleotides, and even coloring matters, e.g., the red–purple of *Monascus purpureus*.

Health Hazards

The products mentioned in this review have undergone millennia (to quote Hesseltine) of testing. Nevertheless, in spite of the advances in food hygiene and the current emphasis on care in avoiding contamination at harvest, safe storage, sanitary processing and marketing, the reported incidence of food poisoning has been rising. There are several reasons for this phenomenon, but, in the present context, the following question arises: how have the people in hot, humid countries circumvented microbial food poisoning in the past?

The answer lies in the fact that the traditional techniques for raw material processing, described here, are energy-economical lessons in living with the microbial world; the appearance of mycotoxins and the various bacterial poisons, mentioned elsewhere in this work, are examples of the failure to learn those lessons.

Conclusion

Some of these oriental fermentations, such as soy sauce and, to a lesser extent, tempe, have already found acceptance in the industrial world. Others have yet to be discovered. A further group will be lost forever if steps are not taken to preserve the diversity of the resource base and record the processes of conversion of these resources to food or, better still, to conserve the environmental conditions so that the original inhabitants may continue to practice their art.

See also: ***Bacillus***: Occurrence; Detection; Food Poisoning; **Enzymes**: Functions and Characteristics; Uses in Food Processing; Uses in Analysis; **Lactic Acid Bacteria**; **Mycotoxins**: Classifications; Occurrence and Determination; Toxicology; **Preservation of Food**; **Probiotics**; **Starter Cultures**; **Toxins in Food – Naturally Occurring**

Further Reading

Bray F (1984) Agriculture. In: Needham J (ed.) *Science and Civilization in China*, vol. 6, part II. Cambridge: Cambridge University Press.

Burkhill HM (1935) *The Economic Products of the Malay Peninsula*. London: Crown Agents.

Campbell-Platt G (1987) *Fermented Foods of the World*. Oxford: Butterworth.

Caplice E and Fitzgerald GF (1999) Food fermentations: role of microorganisms in food production and preservation. *International Journal of Food Microbiology* 50: 131–149.

Cummings JH (1997) *The Large Intestine in Nutrition and Disease*. Brussels: Institut Danone.

Fukushima D (1985) Fermented vegetable protein and related foods of Japan and China. *Food Reviews International* 1: 149–209.

Hachmeister KA and Fung DYC (1993) Tempe: a mold-modified indigenous fermented food made from soybeans and/or cereal grains. *Critical Reviews in Microbiology* 19: 137–188.

Huang KT (2000) Fermentations and food science. In: Needham J (ed.) *Science and Civilization in China*, vol. 6, part V. Cambridge: Cambridge University Press.

Reddy NR, Pierson MD and Salunkhe DK (eds) (1986) *Legume-based Fermented Foods*. Boca Raton, FL: CRC Press.

Stanton WR (1969) Some domesticated lower plants in Southeast Asian food technology. In: Ucko PJ and Dimbleby GW (eds) *The Domestication of Plants and Animals*, pp. 463–469. London: Duckworth.

Stanton WR (1985) Food fermentation in the tropics. In: Wood BJB (ed.) *Microbiology of Fermented Foods*, pp. 193–211. London: Elsevier.

Steinkraus KH (ed.) (1996) *Handbook of Indigenous Fermented Foods*, 2nd edn. New York: Marcel Dekker.

Steinkraus KH (ed.) (1989) *Industrialization of Indigenous Fermented Foods*. New York: Marcel Dekker.

Wood BJB (ed.) (1998) *Microbiology of Fermented Foods*. London: Blackie.

Yokotsuka T (1991) Proteinaceous fermented foods and condiments prepared with koji molds. In: Arora DK, Mukerji KG and Marth EH (eds) *Handbook of Applied Mycology, vol. 3, Food and Feeds*, pp. 329–373. New York: Marcel Dekker.

Beverages from Sorghum and Millet

J R N Taylor, University of Pretoria, Pretoria, South Africa

Introduction

Throughout Africa, the indigenous tropical cereal grains sorghum (*Sorghum bicolor* L. Moench), pearl millet (*Pennisetum glaucum* L. R.Br.), and finger millet (*Eleucine coracana* L. Gaertn.) have been used since prehistoric times to prepare fermented beverages (Figure 1). In general, these traditional beverages are popular as ever, and in southern and eastern Africa their industrial manufacture has become increasingly common. Although the character of the industrially produced beverages is substantially the same as the traditional products, high proportions of maize (corn), (*Zea mays* L.), which is native to the USA, and even maize alone are normally used in their manufacture.

Figure 1 Sorghum and millet growing together in a field in the Northern Province of South Africa. Courtesy of Beryl Fabian, CSIR, South Africa. Reproduced from Fermenting Foods: Beverages from Sorghum and Millet, *Encyclopaedia of Food Microbiology*, Robinson RK, Batt CA and Patel PD (eds), 2000, Academic Press, with permission.

The fermented beverages may be alcoholic, or substantially nonalcoholic. The alcoholic beverages have many names, including sorghum beer (South Africa), opaque beer (Zimbabwe), jwala (Tswana), utywala (Zulu), Chibuku (southern and eastern Africa), and pito (west Africa). Chibuku is a commerical brand name; the word is from the African 'mine' language Fanagalo and can be translated as 'according to the recipe book.' In this article the term 'sorghum beer' is used generically for the southern and eastern African product. Traditional African beers are very different from European lager-type beer, partially on account of the particular properties of the tropical cereal grains. Sorghum beer is characterized by its opacity due to the presence of semisuspended particles of starch, cereal, and yeast. The starch gives it a viscous consistency. The beer is generally pinkish brown in color due to anthocyanidin pigments from the sorghum or millet. It is sour in taste owing to the presence of lactic acid but, unlike lager beer, is not hopped; and it is consumed while actively fermenting. In contrast, pito is clear or somewhat cloudy and is not particularly sour in taste, but like sorghum beer it is consumed while it is actively fermenting. African beers are generally rather lower in alcohol than lager beer. This is related to the considerably lower β-amylase (EC 3.2.1.2) activity of sorghum and millet malts, in comparison to barley malt.

The substantially nonalcoholic fermented beverages have been less well documented but all involve some form of lactic acid fermentation. Products include kunun-zaki (Nigeria), hulu mur (Sudan), and motoho oa mabele (South Africa). These are made from sorghum or pearl millet; the grain may or may not be malted according to local preference. The best-known nonalcoholic beverage is mageu (mahewu or magou) from southern Africa, which is a lactic acid-fermented gruel, widely produced industrially using almost exclusively maize.

Traditional Methods of Making Sorghum Beer

Sorghum beer is intrinsically linked with African culture, traditionally being brewed by the female members of the family for use at ceremonies including weddings, funerals, and other social gatherings.

The basic sorghum beer process involves the following steps: souring (lactic acid fermentation), cooking (starch gelatinization), mashing (thinning and conversion of gelatinized starch to sugars),

straining (grain separation), and alcoholic fermentation. It should be noted that under home brewing conditions it is very difficult to maintain the brews at the optimal temperatures for each of the steps. This leads to an overlap of the processes across different steps, so that two or more brewing steps may take place at the same time.

Malt is the vital ingredient in brewing. In essence, malting involves germinating the grain for several days to produce enzymes (most importantly, amylase enzymes) not present in the grain, then drying and milling it. In the brewing of sorghum beer, the malt amylase enzymes hydrolyze grain starch to sugars. The sugars are fermented to lactic acid by lactobacilli and, during alcoholic fermentation, yeasts convert the sugars to ethanol and carbon dioxide, producing beer.

In traditional home brewing, the grain is malted by first soaking it for about a day, either in water-filled, clay vessels, or in woven grass baskets placed in a stream. After soaking, the steep water is drained off and the damp grain is allowed to germinate for 2–4 days, depending upon the ambient conditions. Germination is conducted in small baskets or in sacks, allowing the circulation of air. At the end of germination, the germinated grain is spread out in a thin layer on the ground and sun-dried. Once dry, the malt is ground by hand between two stones.

To brew beer, a small portion of ground malt is mixed with hot water; either slurried with the water and then cooked, or added to previously heated water. The warm mixture is then covered and allowed to stand overnight to sour. The following day, the soured wort is diluted with additional water and more milled grain is added. This starch-rich adjunct material is either malt, or more commonly whole, coarsely ground sorghum or millet grain. The entire mixture is then boiled (2–6 h), and then allowed to cool and thicken. The following day, a second batch of malt is added to the now cooled brew. The amylase enzymes in the malt liquefy the mash. Depending upon the taste of the brew and the preference of the brewer, it may be recooked, cooled, and more malt added. As the wort cools, wild yeasts in the environment ferment the brew, giving a bubbling, effervescing liquid. A characteristic of traditionally brewed sorghum beers is their fruity flavor, due to fermentation by wild yeasts. The brew is allowed to ferment for a few days When it is considered to be acceptable, it is strained, traditionally by passing it through a braided grass strainer, or today more generally through a perforated steel sheet. The beer is then served in decorated clay drinking vessels. Figure 2 shows a woman serving out the beer with a ladle made from a plant gourd; the latter may also serve as drinking vessel.

Figure 2 Serving traditionally made sorghum beer from clay pots using a gourd ladle. Courtesy of Beryl Fabian, CSIR, South Africa. Reproduced from Fermenting Foods: Beverages from Sorghum and Millet, *Encyclopaedia of Food Microbiology*, Robinson RK, Batt CA and Patel PD (eds), 2000, Academic Press, with permission.

Modern home brewers may make their own malt, or more commonly they obtain it from large commercial maltsters who produce malt of more consistent quality. Over the years, milled maize or brown bread has tended to replace malted or unmalted sorghum or millet as the adjunct. However, the procedures followed are essentially the same as those of ancient times. Figure 3 shows a modern-day sorghum beer home-brewing recipe.

Traditional Methods of Making Nonalcoholic Beverages

The procedures are substantially the same as used for making sorghum beer, but certain steps are ingeniously optimized to result in a substantially nonalcoholic product. Figure 4 shows a scheme for making kunun-zaki. It is notable that the malt is simply mixed (not mashed) with the cooked adjunct. Alternatively, according to local preference, boiling water may be added to the malt. Both processes limit the amount of enzymic hydrolysis of starch to fermentable sugars and hence favor a lactic acid, rather than an alcoholic fermentation. Further, the high initial temperatures in such processes (approx. 50 °C) enable the growth of thermophilic lactic acid bacteria but not yeast. Later, when the temperature falls towards the range suitable for alcoholic fermentation, mesophilic bacteria take over. Kunun-zaki has a pH of below 5.0 and a

titratable acidity around 0.1–0.2%. The predominant microorganisms are the bacteria *Lactobacillus plantarum* and *Bacillus subtilis*, whereas yeast numbers are low. Zaki means sweet in Nigerian Hausa language and the product is often spiced with ginger, cloves, and pepper, then sweetened.

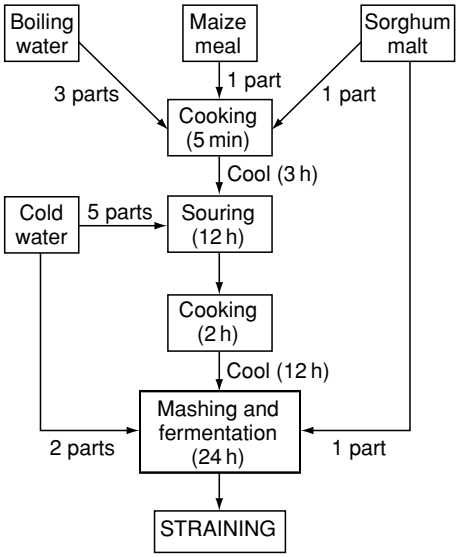

Figure 3 Modern-day home-brewing recipe for sorghum beer. Courtesy of Nola, Randfontein, South Africa. Reproduced from Fermenting Foods: Beverages from Sorghum and Millet, *Encyclopaedia of Food Microbiology*, Robinson RK, Batt CA and Patel PD (eds), 2000, Academic Press, with permission.

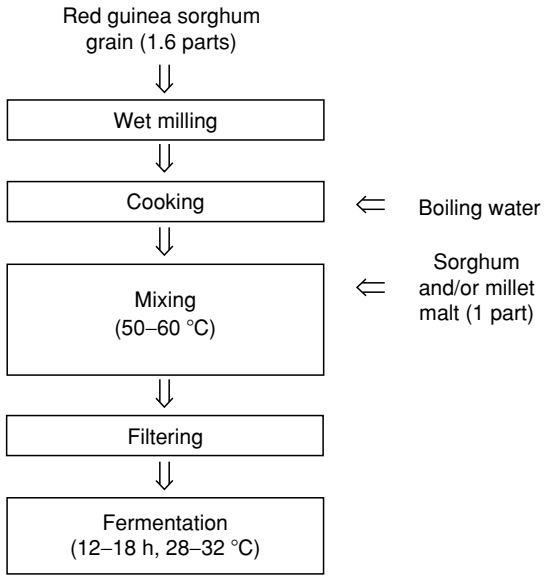

Figure 4 Scheme for preparing kunun-zaki (from a description by Inyang CU, Dabot YA (1997) Storability and potability of pasteurized and sterilized ''kunun-zaki'': a fermented sorghum beverage. *Journal of Food Preservation and Processing* 21: 1–7.

Industrial Sorghum Beer Brewing

Industrial sorghum beer brewing was initiated by municipal governments in South Africa and Zimbabwe in the early part of the 20th century. However, the adaptation of the traditional craft to factory scale was hampered by technical difficulties. In 1954, the Council for Scientific and Industrial Research (CSIR) in South Africa was commissioned to undertake applied research on behalf of the municipal sorghum beer brewers. This technical assistance aided a spectacular increase in scale from home brewing (approximately 50–200-l brews) to factory brewing (approximately 15 000–27 000-l brews), made possible largely by the development and institution of carefully specified raw material standards and the use of specialized equipment. Modern sorghum beer brewing equipment allows rapid heating and cooling, enabling the key brewing steps to be split into separate unit operations and conducted under optimal conditions of temperature (Figure 5).

Malting

The first major problem encountered in scaling-up sorghum beer production related to the quality of the sorghum malt. At that time, virtually nothing was known about the science and technology of sorghum malting. Because sorghum and millet malt contains substantially lower levels of β-amylase than barley malt and amylase activity can be limiting in sorghum beer brewing, the CSIR researchers developed a sensitive method of determining the

Figure 5 Modern industrial sorghum beer brewery. Courtesy of Traditional Beer Investments of Johannesburg, South Africa. Reproduced from Fermenting Foods: Beverages from Sorghum and Millet, *Encyclopaedia of Food Microbiology*, Robinson RK, Batt CA and Patel PD (eds), 2000, Academic Press, with permission.

diastatic power (i.e., the joint level of α- and β-amylase activity) of sorghum malts. The assay remains the most important quality parameter for sorghum malt.

The malting process is physically split into three unit operations: steeping, germination, and drying. Steeping is carried out in steel tanks of several thousand liter capacity. The grain is soaked in water for a period of normally between 6 and 24 h. The water is aerated to assist the grain respiration and may be changed at intervals to allow the grain to respire in air and also to clean it.

As sorghum malt of high and consistent quality is required as an ingredient in industrial brewing, today most is produced in modern pneumatic maltings and is referred to as 'conversion' or 'industrial' malt. The steeped grain is germinated indoors, under conditions of controlled temperature, moisture, and aeration to produce malt to a diastatic power specification of 28–35 sorghum diastatic units (SDU) per gram. The principle of pneumatic malting is that air is blown through the bed of grain by means of a powerful fan. The air aerates the grain, removes carbon dioxide, and maintains it at an even temperature. Another type of malt used in industrial brewing, referred to as 'souring' malt, serves as a source of lactobacilli and as a substrate for lactic acid fermentation. Souring malt is generally produced by germinating sorghum grain outdoors on a concrete floor (floor malting) and as a consequence is generally of lower diastatic power than conversion malt.

The malt is normally dried by blowing heated, dry air through it, at temperatures no greater than 50 °C, until the moisture content of the malt is approximately 10%. Drying at this low temperature helps conserve the amylase activity of the malt. Alternatively the malt can be sun-dried.

Modified Reef-Type Brewing Process

Several different industrial brewing processes are used in southern and eastern Africa, as a result of differences in the taste preferences of the customers, and differing levels of brewing skills and equipment in the breweries. The most common is the Reef-type process (Figure 6), originally developed in the gold-mining region of South Africa known as the Reef, and later modified to include a second cooking and mashing stage.

Souring In South Africa, the lactic acid fermentation (souring) is an essential part of the brewing process. Souring is generally performed by incubating a 10% slurry of 'souring' sorghum malt at 48–50 °C. This favors the growth of thermophilic, homofermentative *Lactobacillus delbreuckii* (*leichmannii*). The sour is normally inoculated with a portion of a previous, successful sour containing high levels of viable lactic acid bacteria – a process of backslopping. Some breweries, however, use specially prepared freeze-dried cultures. The lactic acid sour lowers the pH of the beer to below 4.0, which helps to prevent the complete hydrolysis of starch into sugars, slows the rate of microbial spoilage, and inhibits the growth of pathogenic organisms. The sour also contributes to the beer's characteristic flavor. Breweries in other countries generally simply add industrially produced lactic acid during the brewing process.

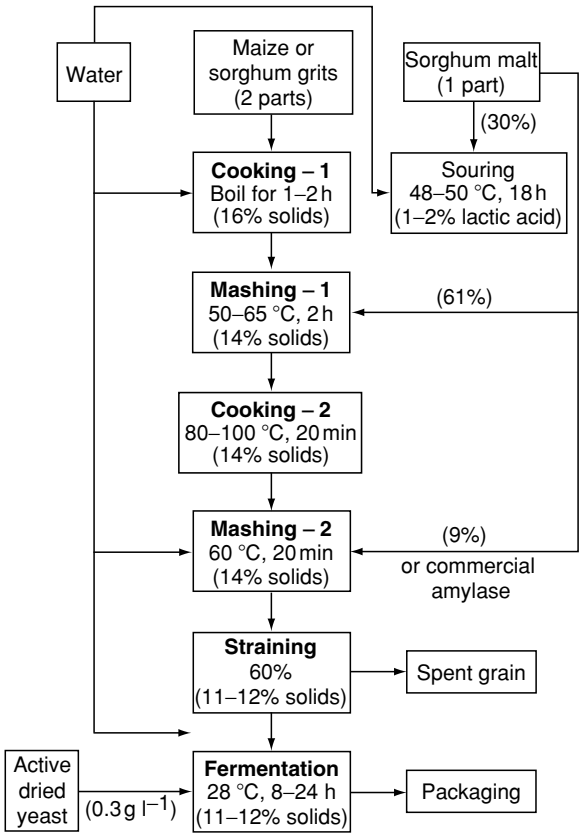

Figure 6 Modified Reef-type sorghum beer brewing process. Reproduced from Fermenting Foods: Beverages from Sorghum and Millet, *Encyclopaedia of Food Microbiology*, Robinson RK, Batt CA and Patel PD (eds), 2000, Academic Press, with permission.

Cooking After the lactic acid fermentation the sour is diluted with water, then maize and/or sorghum meal is added as a starchy adjunct and the mixture is boiled for approximately 1–2 h under atmospheric conditions, or for less time under pressurized conditions, to gelatinize the starch. Industrial brewers in South Africa generally use highly refined maize grits (low-fat endosperm meal) as adjunct. The use of whole, unrefined grain has been largely abandoned owing to the high fat content of the grain, which causes rancidity problems when stored in a milled

form. The high fat content also negatively affects the beer foam. Outside South Africa, the use of less-refined meal, or even whole milled grain is still the norm.

Mashing The mash is cooled to approximately 60 °C with cold water. Conversion malt is then added. As a consequence of the action of the malt α-amylase (EC 3.2.1.1) converting the gelatinized starch to dextrins, the mash is very rapidly liquefied. The β-amylase in the malt hydrolyzes the dextrins to maltose, a fermentable sugar. The low pH of the mash (about 4.0) is below the optimum for sorghum α- and β-amylase activity, which limits starch hydrolysis, preventing the production of too thin a beer and limiting the potential alcohol content of the beer. Some industrial brewers supplement or substitute costly conversion malt with less expensive commercially produced amylase enzymes (α-amylase and/or amyloglucosidase), to facilitate the thinning of the mash and fermentable sugar production. Mashing is carried out for about 2 h at constant temperature (50–65 °C).

Second cooking and mashing The gelatinization temperature of sorghum starch is 68–72 °C. Therefore under the conditions of mashing in the previous step, the sorghum malt starch is not gelatinized. In the modified Reef-type brewing process the mash is recooked (80–100 °C for approximately 20 min). This second cooking stage pasteurizes the mash and also gelatinizes the sorghum malt starch, leaving virtually no ungelatinized starch in the mash. Thereafter, the mash is cooled to 60 °C and a second short mashing period (approximately 20 min) is carried out either with a small amount of conversion malt or commercially produced amylase. The second cooking and mashing steps effectively increase the amount of beer that can be obtained from a given quantity of malt.

Straining Straining removes the coarse cereal particles from the wort, and is normally carried out directly after mashing and prior to fermentation. Industrial brewers strain the mash at elevated temperature using centrifugal decanters and vibrating screens. Before straining there are approximately 14% solids in the mash and after straining approximately 11–12% solids remain in the wort.

Alcoholic fermentation Plate or tubular heat exchangers are used to cool the strained wort to 28 °C. The wort is transferred to stainless-steel vessels and alcoholic fermentation is brought about by pitching with commercially produced yeast (0.3 g l^{-1}). For convenience, active dried yeast is generally used. This yeast is a strain of *Saccharomyces cerevisiae*, specially selected for its ability to ferment sorghum beer rapidly. It is a top-fermenting, nonflocculating type. The pitched wort is then allowed to ferment for 8–24 h. Unlike clear lager beer, sorghum beer is not filtered.

Serving Sorghum beer may be transported by road in bulk (100–500 l) for sale in beer halls as draft beer. Alternatively, the beer is packaged in milk-type cardboard cartons or plastic containers (0.5–25 l) for sale in the retail market. The containers may be returnable and reusable or one-trip, nonreturnable, depending on the market. However, all are characterized by having a small vent in the top to allow the escape of the carbon dioxide produced by fermentation. When ready for consumption, sorghum beer has an alcohol content of approximately 3% by weight. The shelf-life of the product is approximately 5–10 days after straining. Beyond this time, the beer is 'flat' and has settled out, and is frequently spoiled by microbial infection.

Variations in Industrial Brewing Process

In central Africa, the sorghum beer brewing process is much simpler than that described above. Commercial enzymes only are used to convert the adjunct. Whole milled maize is used as the cereal ingredient, although a little sorghum or millet grain may be added for color and flavor. There is no souring step and in some cases no lactic acid addition either. The beer is also often not strained.

Industrial Nonalcoholic Beverage Production

In view of the commercial importance of mageu – some 100 million liters are produced annually – its industrial production will be described, despite the fact that strictly speaking, mageu is an almost exclusively maize beverage, although it can be made from sorghum. Research by scientists of the CSIR and private industry in South Africa in the 1960s led to today's manufacturing process (**Figure 7**). After the maize meal has been cooked to gelatinize the starch, wheat flour (a source of β-amylase to produce maltose) and/or cane sugar is added to promote fermentation. Fermentation is by a commercial thermophilic lactic acid bacteria culture that only produces lactic acid. An interesting development is the widespread use of artificial flavoring in mageu – flavors such as banana, cream, ginger, and guava are popular. The mageu is either pasteurized and/or chemical-preserved to extend its shelf-life and, like sorghum beer,

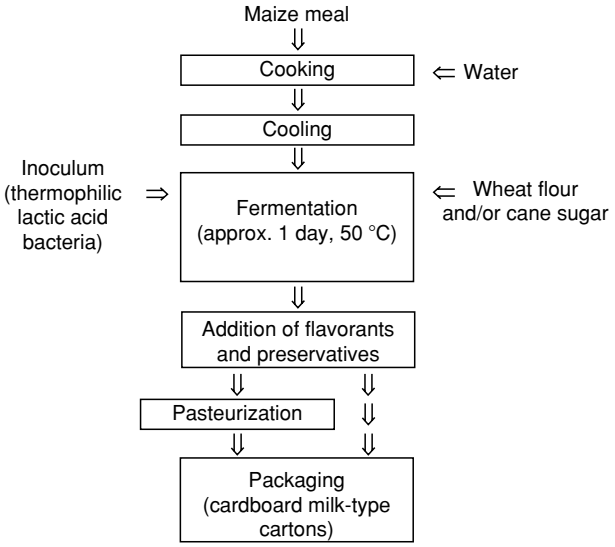

Figure 7 Process for the industrial manufacture of mageu. Reproduced from Fermenting Foods: Beverages from Sorghum and Millet, *Encyclopaedia of Food Microbiology*, Robinson RK, Batt CA and Patel PD (eds), 2000, Academic Press, with permission.

packaged in milk-type cartons. An alternative mageu-making process involves the use of pregelatinized maize meal, produced by extrusion cooking or gun puffing, which obviates the need for cooking equipment. The use of this so-called powder mageu is ideal for preparation in institutions such as mine hostels or hospitals.

Nutritional Aspects

The nutritional value of fermented beverages is influenced to a great extent by two factors, neither of which is associated with fermentation. These are the concentration of cereal solids, and the degree to which the cereal has been refined. The total solids content of commercial mageu is between 5% and 8%, rather lower than the sugar content of many carbonated soft drinks. Higher concentrations are not possible, as the starch would increase the viscosity of the product to that of a porridge. As stated, there is a trend for industrially manufactured fermented beverages to be made with maize endosperm meal rather than partially processed or milled whole grain. Research has shown that brewing sorghum beer with maize grits (the common practice in South Africa) as adjunct substantially reduces its vitamin B and mineral content.

Advantages of Lactic Acid Fermentation

Many nutritional advantages have been claimed for the lactic acid fermentation of cereal beverages and porridges. These include improved starch digestibility, increased protein quantity, improved protein quality and digestibility, increased vitamin levels, decreased antinutritional factors such as tannins and phytate, improved mineral availability, production of bacteriocins and prebiotics, source of probiotic microorganisms, and improved palatability and acceptability. Unfortunately, there are few firm data on the extent of these nutritional improvements. Probably the most unequivocal advantage of lactic acid fermentation is the reduction in pH. Sorghum beer and mageu have a pH of around 3.5. At this low pH, pathogenic bacteria such as *Salmonella typhimurium* are destroyed, heat-resistant bacterial spores do not germinate, and the rate of growth of spoilage bacteria is much reduced. Because many households in Africa still do not have access to potable water and few have refrigerators, these beverages are a valuable safe source of nutrition.

Sorghum Beer

Disregarding the question of whether the consumption of any alcoholic beverage is nutritionally desirable, the nutritive value of sorghum beer is here compared with that of lager beer. There is little evidence, certainly with respect to commercial sorghum beer, to support the often-quoted statement that sorghum beer is more of a food than an alcoholic drink (Table 1). However, sorghum beer has some obvious and not so obvious nutritional advantages. Although the energy content of the two beers is similar, sorghum beer has less ethanol and substantially more complex carbohydrates. The protein content of sorghum beer is much higher, and much of this is yeast protein, which has a better essential amino acid content than cereal protein. The higher thiamin and riboflavin levels are also due to the presence of yeast. The lower nicotinic acid level is due to the use of a high proportion of maize in the grist. The high levels of iron and some other minerals in sorghum beer are the consequence of leaching-out of vessels, because of the low pH of the beer. This can also be disadvantageous. In small-scale brewing where mild steel (as opposed to stainless-steel) vessels are commonly used, excessively high levels of iron have been found in the beer. A benefit of sorghum beer consumption not shown is that, uniquely among beers, it contains measurable quantities of both insoluble and soluble dietary fiber.

Future Directions

As mentioned, these traditional fermented beverages remain popular on the African continent. However, in South Africa there is clear evidence that their industrial production is declining, apparently due to competition from western products such as lager beer and

Table 1 Nutrient composition and contribution to recommended dietary allowance (RDA)[a] of commercial sorghum beer and lager beer

Nutrient (mg l^{-1} unless otherwise stated)	*Sorghum beer*	*Lager beer*	*Percentage contribution to RDA l^{-1}*	
			Sorghum beer	*Lager beer*
Alcohol (% w/w)	2.50	3.59	[a]	
Energy (kJ l^{-1})	1651	1605	15	14
Protein (g l^{-1})	5.40	2.80	10	5
Thiamin	0.24	0.03	17	2
Riboflavin	0.39	0.19	24	12
Nicotinic acid	2.93	4.72	16	26
Cu	0.21	0.09	7	3
Fe	1.90	0.07	19	0.7
Zn	1.40	0.10	9	0.7
Mn	1.60	0.16	40	4
Ca	48	44	6	6
Mg	137	83	39	24
K	312	357	8	9
Na	20	69	1	3
P	245	246	31	31

[a]No RDA.

Reprinted from van Heerden IV. In defence of sorghum beer. *SA Food Review* supplement, Feb/Mar 1988, pp. 83–84 with permission.

carbonated soft drinks. The traditional beverages suffer from a number of technical defects, especially variable quality, short shelf-life and inappropriate packaging. These defects are not insurmountable. For example, bottled, pasteurized sorghum beer has been developed. More importantly, however, traditional beverages have a poor image, especially among affluent younger people. The sorghum beer industry in southern Africa is addressing both these problems by attempting to raise the technical standard and status of its personnel to that in the lager beer industry. The CSIR and the University of Pretoria in South Africa run training courses in sorghum malting and sorghum beer brewing technologies. Several hundred people from all over southern and eastern Africa have attended these course and the sorghum beer technology course is becoming a requirement for brewers in the industry.

Because sorghum and millets are uniquely suited to cultivation in the semiarid tropics, there is considerable interest in Africa and India in using them as a replacement for imported barley in the production of lagers, stouts, and distilled beverages. In 1988 the government of Nigeria banned the importation of barley malt. Today in Nigeria, lager beers are produced using unmalted sorghum grain as the sole cereal source. Commercial amylases and proteases are used to produce sugars and free amino nitrogen from the sorghum grain starch and protein. In addition, research and development is being undertaken worldwide to produce clear lager-type beers using malted sorghum, which would itself provide the necessary hydrolytic enzymes.

See also: **Alcohol**: Metabolism, Beneficial Effects, and Toxicology; Alcohol Consumption; **Beers**: History and Types; Biochemistry of Fermentation; **Lactic Acid Bacteria**; **Maize**; **Malt**: Malt Types and Products; Chemistry of Malting; **Millets**; **Sorghum**; **Starch**: Structure, Properties, and Determination; Sources and Processing; Functional Properties; **Yeasts**

Further Reading

Daiber KH and Taylor JRN (1995) Opaque beers. In: Dendy DAV (ed.) *Sorghum and Millets: Chemistry and Technology*, p. 299. St Paul: American Association of Cereal Chemists.

Haggblade S and Holzapfel WH (1989) Industrialization of Africa's indigenous beer brewing. In: Steinkraus KH (ed.) *Industrialization of Indigenous Fermented Foods*, p. 191. New York: Marcel Dekker.

Hallgren L (1995) Lager beers from sorghum. In: Dendy DAV (ed.) *Sorghum and Millets: Chemistry and Technology*, p. 283. St Paul: American Association of Cereal Chemists.

Holzapfel WH (1989) Industrialization of mageu fermentation in South Africa. In: Steinkraus KH (ed.) *Industrialization of Indigenous Fermented Foods*, p. 285. New York: Marcel Dekker.

Inyang CU and Dabot YA (1997) Storability and potability of pasteurized and sterilized "kunun-zaki": a fermented sorghum beverage. *Journal of Food Preservation and Processing* 21: 1–7.

Murty OS and Kumar KA (1995) Traditional uses of sorghum and millets. In: Dendy DAV (ed.) *Sorghum and Millets: Chemistry and Technology*, p. 185. St Paul: American Association of Cereal Chemists.

Novellie L (1986) Sorghum beer and related fermentations of southern Africa. In: Hesseltine CW and Wang HL

(eds) *Indigenous Fermented Food of Non-western Origin*, p. 219. Berlin: Cramer.
Novellie L and De Schaepdrijver P (1986) Modern developments in traditional African beers. In: Adams MR (ed.) *Progress in Industrial Microbiology*, vol. 23, p. 73. Amsterdam: Elsevier Science.
Palmer GH, Etokakpan OV and Igyor MA (1989) Review: sorghum as brewing material. *MIRCEN Journal 5*: 265–275.
Rooney LW, Kirleis AW and Murty DC (1986) Traditional foods from sorghum: their production, evaluation, and nutritional value. In: Pomeranz Y (ed.) *Advances in Cereal Science and Technology*, vol. 8, p. 317. St Paul: American Association of Cereal Chemists.
South African Bureau of Standards (1970) *Standard Test Method for the Determination of the Diastatic Power of Malts Prepared from Kaffircorn (Sorghum) Including Bird-proof Varieties, and from Millet*. SABS Method 235. Pretoria: South African Bureau of Standards.
South African Bureau of Standards (1979) *Standard Specification for the Production of Mageu*. SABS 1199–1979. Pretoria: South African Bureau of Standards.
Tomkins A, Alnwick D and Haggerty P (1988) Fermented foods for improving child feeding in eastern and southern Africa: a review. In: Alnwick D, Moses S and Schmidt OG (eds) *Improving Young Child Feeding in Eastern and Southern Africa: Household-level Food Technology*, p. 136. Ottawa: International Research Development Centre.
Van Heerden IV (1989) Sorghum beer – a decade of nutritional research. *Proceeding of the Second Scientific and Technical Convention*, p. 293. Johannesburg: The Institute of Brewing Central and Southern African Section.
World Health Organization (1996) *Fermentation: Assessment and Research*. Report of a joint FAO/WHO workshop on fermentation as a household technology to improve food safety. Food Safety Unit, Division of Food and Nutrition. Geneva: WHO.

Soy (Soya) Sauce

M Sasaki and N Nunomura, Kikkoman Corporation, Noda City, Japan

Background

'Shoyu' is the Japanese name for soy sauce. It is made mostly by fermentative methods and is used mainly as an all-purpose seasoning in Japanese cuisine.

In Japan the total production is about 1.04 million kiloliters (kl), against a total population of 125 million. The annual consumption of shoyu per capita is about 8.3 l, of which 3.5 l are consumed in the home and the remaining 4.8 l institutionally and industrially (1999). The consumption volume in Japan is decreasing slightly, but that in American and European area is increasing.

There were 5363 manufacturers in 1957, but by 1999, the number had decreased to 1611. Five large companies together produce about one-half of the total shoyu. Five different types of shoyu are available in Japan: koikuchi, usukuchi, tamari, shiro, and saishikomi.

Koikuchi is made from a mixture of soya beans and wheat kernels in almost equal amounts, and is characterized by a deep reddish brown color and a strong, pleasant aroma. Low-salt shoyu of koikuchi is also enjoyed, and its consumption is growing. It contains half the salt of regular koikuchi shoyu, and it is considered to be healthier.

Usukuchi is characterized by a lighter brown color and milder flavor than koikuchi. It is used mainly for cooking, especially when the original color and flavor of the foodstuff are to be maintained.

Tamari is characterized by a high content of amino acids, and has a dark brown color and distinctive flavor. It is made from soya beans with only a small amount of wheat in the ratio 10:1–2. It is only consumed around the Nagoya region of Japan as a cooking and dipping sauce for several typical Japanese dishes, such as raw fish ('sasimi' in Japanese) and broiled eel ('kabayaki' in Japanese). One of the characteristics of Japanese-type shoyu is the use of almost equal amounts of soya beans and wheat as raw materials. Tamari represents the Chinese type and originates from China. In contrast, shiro is made from wheat with only a small amount of soya beans in the ratio 10:1–2. It is characterized by a very light yellow color, and it has a very low amino acid content and very high sugar content. It is used mainly for cooking and is only consumed around the Nagoya region. Saishikomi is produced from equal amounts of wheat and soya beans in the koji, and raw shoyu is used instead of brine. It is used mainly as a dipping sauce.

Table 1 Production of six kinds of shoyu (Japan Agricultural Standard (1999))

Type	*Amount (kl)*	*Market share (%)*
Koikuchi	686 171	82.9
Usukuchi	115 848	14.0
Tamari	15 233	1.8
Shiro	5 949	0.7
Saishikomi	4 610	0.6
Total production	827 811	100.0

The annual production of the five kinds of shoyu, recognized as being of a quality equal to the Japan Agricultural Standard by the Japan Soy Sauce Inspection Institute in 1999, is shown in Table 1.

As shown in Table 1, about 83% of all shoyu is of the koikuchi type, and thus this type of soy sauce is regarded as the typical Japanese shoyu. The preparation of koikuchi is described below.

History

The origin of shoyu is considered to be from a Chinese food, called 'sho' in Japanese. The first description of sho was found in the book 'Shurai' (in Japanese), which was written in China 3000 years ago. According to this book, sho was made by aging a mixture of mold culture, foxtail millet, dried meat, and a liquor in a bottle for 100 days. The final product was in the form of a mash or 'miso,' which is now one of a number of typical fermented foods in Japan (a semisolid salty food made from soya beans, rice or barley, and salt by fermentation). Soya beans are not described in the book as one of the raw materials of sho, but it is quite possible that soya beans were used, because it is known that soya beans were widely cultured in China 4000 years ago. The first documentation for the clear liquid part of sho is in the Chinese book 'Chin-Min-Yao-Shu' (Saimin-Yojutsu in Japanese) written in the sixth century AD. Sho or hishio was also made from fish and salt at the beginning of Japanese history. Sho made from soya beans is presumed to have been introduced from China into Japan along with other foods by the Chinese Buddhist priest Ganjin (AD 1254). The first record of the name 'shoyu' appeared in 'Ekirinbon-suyoshu' (AD 1595), and it is likely that the basic manufacturing process of today's Japanese-type shoyu was derived by the early seventeenth century. In the Edo era (1603–1867) of Japan, the technology for preparing shoyu developed dramatically and the scale of production increased. However, remarkable improvements in the processing of shoyu have taken place since 1950, with dramatic advances in both biochemistry and technology.

Shoyu was exported to The Netherlands from the port of Nagasaki for the first time in 1668, and, at present, it is exported to 94 countries as a seasoning of world-wide appeal.

Outline of the Manufacturing Process of Koikuchi Shoyu

The process for production of koikuchi consists of five major steps: treatment of raw material, koji making, mash fermentation and aging, pressing and refining, and pasteurization. Among the steps, koji making, mash fermentation, and pasteurization are the most important and difficult processes in shoyu production.

The manufacturing process is shown in Figure 1.

Raw Materials

The amounts of raw materials used for shoyu production in Japan in 1997 are listed in Table 2.

Soya beans Soya beans originated in the East, and the cultivated soya beans, e.g., *Glycine max*, have been derived from *Glycine ussuriensis*, which is a wild species growing in eastern Asia. (*See* **Soy (Soya) Beans: The Crop; Processing for the Food Industry.**)

Whole and defatted soya beans contain approximately 33 and 45% (w/w) protein, respectively. The key component for shoyu production is protein, and about three-quarters of the total nitrogen in shoyu originates from soya beans.

At least 60–70% of the total protein in soya beans is accumulated in particles called 'protein bodies' or aleurone grains in the developing cotyledons. About 90% of the seed protein (about 36% of seed dry weight) is composed of globulins. The globulins are a complex mixture of compounds, and the major difference between the proteins is the molecular size, as separated by ultracentrifugation. The 7S and 11S fractions make up about 70% of the total globulins; the minor constituents are 2S (15%) and 15S (9.1%). The 7S fraction consists of β- and γ-conglycinin (35%), as classified by serological characteristics, whilst the 11S fraction consists of glycinin (41.8% of the total globulins). (*See* **Protein: Chemistry.**) The major components of the soluble carbohydrate fraction in the mature soya bean seed are: sucrose (41.3–67.5%), stachyose (12.1–35.2%) and raffinose (5.2–15.8%). Soya bean oil consists of 94–97% fatty acid glycerides, together with 1.6–2.3% phospholipids. (*See* **Carbohydrates: Classification and Properties; Fatty Acids: Properties; Phospholipids: Properties and Occurrence.**)

Wheat Wheat belongs to the genus *Triticum* of the family Gramineae (the grasses). The origin of the wheat is not entirely certain, but cultured species

Table 2 Raw materials used for shoyu production in Japan (1999)

Raw material	*Amount (tonnes)*
Soya beans, whole	29 587
Soya beans, defatted	154 872
Wheat	154 727
Salt	190 466

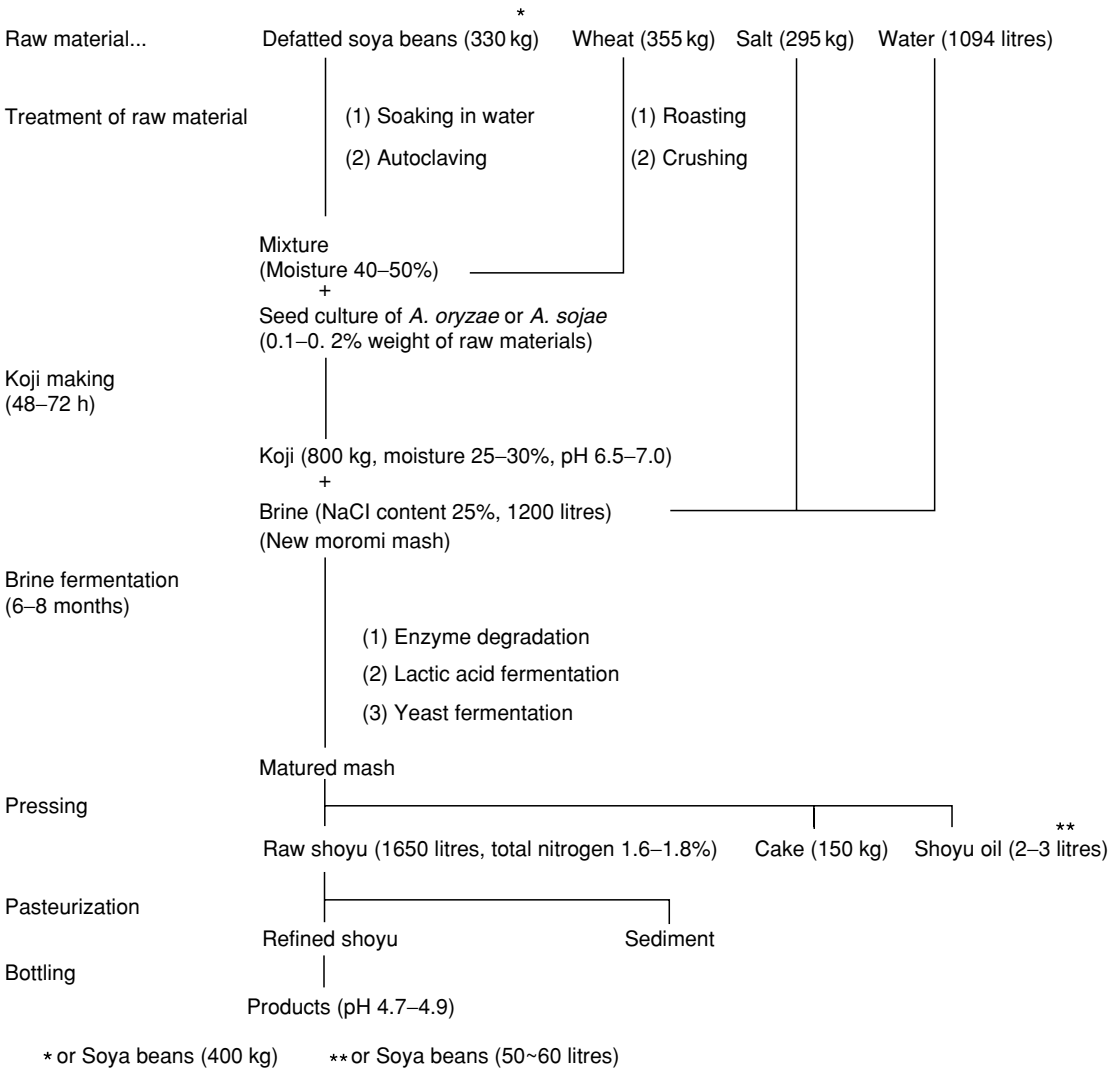

Figure 1 Manufacturing process of Koikuchi shoyu.

originated in prehistory. Cultivation of wheat is believed to have originated in the mountainous regions of Syria and the area formerly part of Palestine of the Middle East. (*See* **Wheat**: The Crop; Grain Structure of Wheat and Wheat-based Products.)

Wheat is commercially divided into seven classes on the basis of botanical species, habit of growth and color: hard red winter wheat, hard red spring wheat, soft red winter wheat, soft red spring wheat, hard white winter wheat, hard white spring wheat, soft white winter wheat, and soft white spring wheat.

Hard wheat is more suitable for making shoyu, because the protein content is higher than in soft wheat. The carbohydrates in wheat include starch, cellulose, and sugar. Starch, which is by far the most abundant constituent of the carbohydrate fraction, is important chiefly as a source of energy for the koji mold in the koji-making process. The proteins of the starchy endosperm are largely the gluten-forming proteins gliadin and glutenin, both of which are insoluble in water and in aqueous salt solutions. About a quarter of the total nitrogen in shoyu originates from the proteins in the wheat. (*See* **Starch**: Structure, Properties, and Determination.)

Salt The salt should contain as little iron and manganese as possible, because oxidative browning of shoyu is synergistically accelerated by the presence of these elements.

Water Water for shoyu should be clean bacteriologically and chemically. For the same reason as with salt, low iron and manganese contents are preferable.

Treatment of Raw Materials

Soya Beans

The protein in raw soya beans is native and cannot be hydrolyzed by the enzymes of the koji mold, and it is necessary to denature the protein to render it susceptible to enzyme action. Conditions for denaturing the soya beans affect digestibility, i.e., the ratio of the total nitrogen of a shoyu to that of the raw materials, and hence denaturing methods have been vigorously investigated, including the development of machines. As a result, protein digestibility has increased from 60 to 90%.

There were many problems to be solved in improving the digestibility. For example, if soya bean proteins remain partly undenatured, shoyu becomes cloudy when it is diluted with water or heated for cooking. This cloudiness diminishes its commercial value. However, over-denatured soya protein loses accessibility for enzyme reactions; as a result, digestibility is low. The methods of cooking can be classified into five groups, according to the stages of historical development from the first generation to the fourth generation, as described below. (*See* **Protein**: Digestion and Absorption of Protein and Nitrogen Balance.)

First generation (traditional method) Before the nineteenth century, soya beans were soaked and then boiled or steamed at atmospheric pressure. At that time, the digestibility was about 60% for whole soya beans.

Second generation Around 1938, the method of cooking was changed to autoclaving under pressure. Soya beans or defatted soya bean grits were sprinkled with hot water and then cooked at a gauge pressure of $0.8\,kg\,cm^{-2}$ for about 1 h. After stopping the steam, the cooked soya beans were kept in the cooker overnight with the valve unopened.

The digestibility was, at best, about 66% for defatted soya beans, because of overdenaturation of the protein.

Third generation The method that characterizes the third generation was invented by Tateno and coworkers (1955) and is called the 'NK method.'

Thoroughly moistened soya beans or defatted soya bean grits were cooked in a rotary cooker under a pressure of $0.8\,kg\,cm^{-2}$ for 45 min, and then immediately cooled to below 40 °C by reducing the inside pressure with a jet condenser. This technique considerably increased the protein digestibility from 66 to 87% for defatted soya beans.

This method contributed greatly to the development of techniques for treating the raw material.

Fourth generation (high-temperature/Short-time (HTST) method) (1970) For convenience, this technique is divided into the following two methods.

Method 1 Thoroughly moistened soya beans are cooked under a higher pressure and a shorter time than with the NK method, using saturated steam, i.e., $2.0–7.0\,kg\,cm^{-2}$ for 15 s to 5 min.

The protein digestibility using this method is 84.24 ($2\,kg\,cm^{-2}$, 5 min) to 89.70% ($6\,kg\,cm^{-2}$, 0.5 min) for defatted soya beans.

Method 2 Soya beans are cooked without water using superheated steam at a gauge pressure of $4–8\,kg\,cm^{-2}$ at 220–287 °C for atleast 15 s.

The protein digestibility is almost the same as that recorded for method 1 using saturated steam, but this method has the advantage of making it possible to store the heat-treated raw materials.

Others The treatment of soya beans with water containing ethanol was investigated, but this method was not used because of a large number of problems.

The extruder method has been used and has been found to be suitable for small quantities of material.

Wheat

Wheat kernels are roasted without adding water. In the past, roasting was carried out in an iron pan, but this method was replaced with a continuous roaster at the beginning of the 1900s. Nowadays, the HTST method, described above, is employed. Control of the process is important. If the wheat is insufficiently roasted, the raw starch cannot be digested by the amylase from the koji mold. However, if the wheat is over-roasted, the protein digestibility decreases.

The roasted wheat kernel is coarsely crushed into flour or fine pieces.

Koji Manufacture

The main purpose of this process is to produce the enzymes needed for hydrolysis of the raw materials. Many nutrients for the yeasts and lactobacilli that multiply in the next stage, and some flavor components that influence the quality of shoyu are produced in this process.

Procedure

Nearly equal amounts of cooked soya beans and roasted, crushed wheat (6:4–4:6) are mixed and inoculated with a pure starter culture of *Aspergillus*

oryzae or *A. sojae* – the so-called koji starter or seed mold. The inoculated mixture is then transferred to the equipment for koji cultivation, and spread on to large perflorated stainless steel trays (3 (wide) × 12 (long) × 0.6 (high) m) to a depth of 30–40 cm, and then incubated in a room at 25–30 °C for 2–3 days. During this period, the temperature, moisture, and aeration are controlled to allow the seed mold to grow on the mixture, and to promote the production of enzymes. A temperature above 35 °C leads to a decrease in enzyme production and, sometimes, to death of the koji mold. The temperature of the koji is controlled by stirring. The first stirring is performed at about 20 h, and the second at about 25 h, after inoculation. The resulting end product (clear yellow to yellowish green in color) is 'koji.'

The method described above is a batch system. In recent years, an advanced system for continuous koji cultivation has been developed and employed industrially.

Koji Molds

The koji molds used for shoyu are *A. oryzae* and *A. sojae*. These species differ not only in their conidial morphology, but also in several physiological characteristics of shoyu-manufacturing importance. Generally, *A. oryzae* is characterized by a high productivity of α-amylase, and *A. sojae* is characterized by a high productivity of protease. *A. oryzae* is used not only for shoyu production, but also for the other Japanese fermented foods, such as 'miso' and 'sake.' However, the use of *A. sojae* is limited to shoyu manufacture. However, about four times as many manufacturers use *A. oryzae* than *A. sojae*.

The breeding of koji molds has been investigated, using mutation, crossing, and protoplast fusion, in an attempt to increase their enzyme productivities. At the present time, several koji molds with a high enzyme productivity of protease and amylase are used to ensure a high productivity of the whole process.

Since the discovery of aflatoxin, the question has arisen as to whether or not koji molds produce this carcinogenic mycotoxin. The results of a large number of investigations have proved that no Japanese industrial strains of koji mold produce this toxin. (*See* **Mycotoxins**: Occurrence and Determination.)

Mash Manufacture and Aging

Procedure

The mixing of koji and brine is called 'shikomi' in Japanese, and the mixture is called the 'moromi' mash. The salt content in the brine is 23–25%, and the ratio of brine to koji is usually 1:1.2–1:1.3 (v/v).

The koji and brine are measured simultaneously, and the mixture is pumped into deep fermentation vessels. Wooden casks of 5–10 kl, or 10–20 kl concrete tanks, were once used as fermentation vessels, but recently, they have been replaced with resin-coated, iron tanks of 50–300 kl or fiber-glass tanks of 30–100 kl. This moromi mash is stirred by air pressure to prevent salt-tolerant lactobacilli and wild yeasts from growing.

At one time, shoyu was manufactured over the four seasons. Koji and the mash were prepared in February or March at a room temperature of 5–15 °C, and the mash was fermented and aged from spring to autumn. The shoyu that was brewed under these conditions contained larger amounts of total nitrogen and amino acids, but only a moderate amount of lactic acid. On organoleptic evaluation, it was considered better than shoyu brewed in the other seasons. Nowadays, in modern shoyu factories, in order to produce an excellent-quality shoyu and reduce the fermentation period, the temperature of the mash is controlled; the control simulates the temperature changes of a mash brewed from spring to autumn. The koji is first mixed with brine at about 0 °C to keep the temperature of the mash below 15 °C for several days. Selected salt-tolerant lactic acid bacteria (*Tetragenococcus halophilus*) are added as a starter, and then the temperature is gradually raised to 28–30 °C after 20–30 days. During this period, a salt-tolerant yeast, such as a pure culture of *Zygosaccharomyces rouxii*, is added. After the vigorous alcoholic fermentation has finished, the temperature is dropped again and kept at around 25 °C over the final 2 months. The addition of *Candida (Torulopsis)* yeasts is recommended to obtain a good 'volatile' flavor in the final product. A period of 6–8 months is necessary to complete fermentation and aging. In the past, fermentation was carried out without the addition of lactic acid bacteria or yeasts. Natural conditions controlled the fermentation by adventitious bacteria and wild yeasts. (*See* **Lactic Acid Bacteria**.)

Chemical Changes in Mash Fermentation

In the early stages of mash fermentation, proteins from the raw materials are hydrolyzed to lower-molecular-weight peptides and free amino acids.

Approximately 20% of the starch from wheat is consumed by the molds during koji fermentation, and the rest is retained for the moromi mash. Most of this remaining starch and the other carbohydrates from wheat are converted to hexoses and pentoses in the moromi mash. A proportion of these sugars is subsequently metabolized into lactic acid, acetic

Table 3 Properties of *Zygosaccharomyces rouxii*, *Candida versatilus*, and *Candida echellsii*

	Z. rouxii		*C. versatilis*		*C. echellsii*	
	Growth	*Optimum*	*Growth*	*Optimum*	*Growth*	*Optimum*
Temperature (° C)						
NaCl 0%	< 35	25–30	< 35	25	< 30	25
NaCl 18%	< 42	25	< 32–35	25–30	< 33–35	25
pH						
NaCl 0%	3.0–7.0	4.5–6.0	3.0–7.0	4.0–5.5	3.0–7.0	5.0
NaCl 18%	3.5–5.5	4.0–5.0	3.0–7.0	4.0–5.0	3.0–7.0	4.0–5.0
NaCl (%, w/v)	< 24–26		< 26		< 26	
Water activitv (a_w)	0.787–0.81		< 0.787		0.787	
	NaCl (%)		*NaCl (%)*		*NaCl (%)*	
Vitamin requirement	*0*	*18*	*0*	*18*	*0*	*18*
Thiamine	N	N, A	E	E	A	E
Calcium pantothenate	E	E	N	N, A	N	A, E
Biotin	E	E	E	E	E	E
Inositol	A	E	N, A	A	N	N, A

E, essential; A, accelerative; N, nonessential.
From Hisaoy (1977) *Microbial Ecology*, No. 4, p. 34. Tokyo: Japan Scientific Societies Press.

acid, and other organic acids by lactobacilli. As a result, the pH drops from an initial value of 6.7–7.0 to 4.7–4.8.

In the next stage, vigorous alcoholic fermentation (from the action of the yeast *Z. rouxii*) occurs, and the remaining sugars are metabolized into ethanol and a number of minor flavor compounds. For example, *Candida versatilis* or *C. echellsii* converts ferulic acid and 4-hydroxycinnamic acid from wheat to 4-ethylguaiacol and 4-ethylphenol, which enhance the quality of shoyu.

As described above, the dominant species of yeasts found in the mash are *Tetragenococcus halophilus*, *Z. rouxii*, *C. versatilis*, and *C. echellsii*. The major properties of these microbes are listed in Tables 3 and 4.

Table 4 Properties of *Tetragenococcus halophilus*

1. Gram-positive *Micrococcus*, diameter 0.6–0.9 μm
2. Halophilic
3. Facultative anaerobe
4. Catalase-negative
5. Indole is not formed
6. Nitrates are not reduced
7. Nonmotile
8. Spores are not formed
9. Growth

	Growth	*Optimum*
Temperature (°C)	20–42	25–30
pH	5.5–9.0	7.0
Water activitv	0.94–0.808	0.99–0.94
NaCl (% w/v)	5–24	5–10

Vitamin requirement: Biotin, vitamin B_2, B_6, nicotinic acid, pantothenic acid
Amino acid requirement: Leucine, isoleucine, valine, glutamic acid, arginine, histidine, tryptophan, phenylalanine

Pressing

The fermented mash is put into a cloth, and the liquid part is squeezed out under hydraulic pressure, which sometimes reaches 100 kg cm^{-2}, for 1–3 days. The difficulty of pressing shoyu mash is due to the viscosity of more than 300 cps.

The liquid part of the mash is called raw shoyu, or 'nama-shoyu' in Japanese. The residue from the pressing is called shoyu cake, or 'shoyu-kasu' in Japanese. This can be used as an additive in animal feed. The final moisture content of shoyu-kasu is less than 25%. At one time, pressing required large amounts of labor compared with the other processes, but today, very efficient and automatic machines are available.

Refining

The raw shoyu is stored in a tank and separated into three layers: (1) sediment at the bottom, (2) clear shoyu in the middle, and (3) an oily layer at the top. The oil layer is called shoyu oil, or 'shoyu-abura' in Japanese, and is removed by decanting. The middle layer, clear shoyu, is heated at 115–120 °C for a few seconds in a heat exchanger in order to kill any vegetative microbial cells, denature enzymes, coagulate proteins, develop the agreeable reddish brown color, and generate aroma.

The clear shoyu is then filtered, bottled, and marketed.

Retail Product

Chemical Composition

The chemical compositions of various types of shoyu are presented in Table 5.

Aroma and Flavor Compounds

All the production processes are related to the formation of flavor compounds, i.e., the heat treatment of raw material, koji culturing by molds, lactic acid fermentation, yeast fermentation, aging of the moromi mash, and pasteurization. Among these processes, yeast fermentation in the moromi mash and pasteurization contribute most to aroma and flavors. (*See* **Sensory Evaluation**: Aroma; Taste.)

The first attempt to identify the flavor compounds of shoyu was made by Tahara in 1887. Since then, many studies on shoyu flavor have been reported, and, to date, over 300 volatiles have been detected. Among them, caramel-like aroma compounds, such as furanones, and phenolic compounds contribute most to the flavor of Japanese shoyu. (*See* **Phenolic Compounds**.)

4-Hydroxy-2(or 5)-ethyl-5(or 2)-methyl-3(2H)-furanone (HEMF) was isolated for the first time from the natural product (1976). It has an intense, sweet aroma and a shoyu-like flavor. As it is present at high levels (100 p.p.m.) in shoyu and has the lowest threshold of detection in water (less than 4×10^{-5} p.p.m.) among the compounds found in shoyu, the aroma value (the concentration of component/the threshold concentration of component in water) is calculated to be $> 5 \times 10^{6}$. Accordingly, it is a significant contributor to shoyu flavor and is believed to be a 'character impact compound' of shoyu.

In the 1990s, HEMF has been found in various foods such as miso (1991), Emmental cheese (1994), lager beer (1996) and coffee (1993), but the levels in these foods are much lower than that in shoyu. The homologs of HEMF, 4-hydroxy-2,5-dimethyl-3(2H)-furanone (HDMF), and 4-hydroxy-5-methyl-3(2H)-furanone (HMMF) have also been isolated from shoyu. The amount of HDMF in shoyu is very low (about 10 p.p.m.), but HMMF is present at high levels (100 p.p.m.). HDMF is found in many foodstuffs, such as pineapple, strawberries, roasted almonds, and beef broth, and in various model systems, such as the degradation of fructose, the pyrolysis of D-glucose, and roasting of alanine and rhamnose. Similarly, HMMF is found in many foodstuffs, such as beef broth, wild raspberries, and guava fruit, and in various model systems, such as the degradation of L-dehydroascorbic acid, and roasting of glycine and xylose. The formation of these compounds by a nonenzymatic reaction has been proved.

HDMF and HMMF in shoyu are probably formed from sugars and amino acids during heating processes, such as the heat treatment of the raw materials and the pasteurization of the liquid part of the mash. However, HEMF is presumed to be biosynthesized through the pentose phosphate cycle by shoyu yeasts. Not only shoyu yeasts but also the other yeasts employed for alcoholic beverages and single-cell protein can change intermediates, such as D-ribulose 5-phosphate, to HEMF. Moreover, it has been proved that D-xylulose 5-phosphate, one of the intermediates in the pentose phosphate cycle, is present in both the enzymatic hydrolysates of isolated soya bean protein and in shoyu moromi mash before the growth of yeasts in shoyu fermentation, and is the precursor for HEMF.

The amount of HEMF produced depends greatly on the concentration of sodium chloride in the shoyu. The highest productions occurred at a sodium chloride concentration of 17–18%.

HEMF has also been reported to show antitumor activity.

Phenolic compounds, such as 4-ethylguaiacol and 4-ethylphenol, have an important relationship with the aroma of shoyu. These compounds are derived from ferulic acid and 4-hydroxycinnamic acid in koji, respectively, by the action of *Candida (Torulopsis)* yeasts. Based on the results of sensory evaluations, the optimum concentration of 4-ethylguaiacol has been found to be 0.8 p.p.m. in shoyu.

Many aldehydes, such as isobutyraldehyde, isovaleraldehyde, and so on, increase during pasteurization, as a result of the aminocarbonyl reaction and Strecker degradation. The aroma derived from pasteurization is called 'fire aroma,' or 'Higa' in Japanese.

The typical quantitative analysis of koikuchi shoyu is shown in Table 6.

Shoyu flavor consists of HEMF and other minor components, and, whilst HEMF gives the fundamental flavor, many minor components such as 4-ethylguaiacol give the product character.

Color

The color of fresh koikuchi-type shoyu is deep reddish brown, and it is very sensitive to oxidation. The color of shoyu packed in bottles or cans is usually stable but darkens rather quickly after opening the seals. During the brewing process, the development of shoyu color derives mainly from nonoxidative browning reactions, but after opening the seals, it is mainly due to oxidative and nonenzymatic browning reactions between amino acids and sugars, which are

Table 5 Typical principal compositions of different kinds of shoyu

	Koikuchi	*Usukuchi*	*Tamari*	*Shiro*	*Saishikomi*
Be Baume degree	21.55	22.24	25.14	20.54	25.08
NaCl (%, w/v)	16.25	18.60	17.50	17.45	14.30
Total nitrogen (%, w/v)	1.60	1.19	2.58	0.53	2.17
Ethanol (%, v/v)	2.60	2.96	3.32	4.19	3.21
Color (Y value)	1.32	22.14	0.08	75.93	1.25[a]
pH	4.80	4.83	4.93	4.62	5.00
Ammonia (mg ml^{-1})	1.29	1.06	2.17	0.59	1.65
Organic acids (mg ml^{-1})					
Pyroglutamic acid	2.62	1.34	5.55	0.79	6.61
Lactic acid	6.76	5.26	11.10	nd	3.45
Acetic acid	1.09	0.94	2.37	0.73	0.93
Formic acid	0.05	nd	0.14	nd	0.04
Citric acid	1.00	0.15	1.99	0.30	3.18
Succinic acid	0.33	0.32	0.41	0.06	0.63
Amino acids (mg ml^{-1})					
o-Phosphoserine	0.72	0.54	1.25	0.18	1.44
Taurine	nd	nd	0.58	nd	nd
L-Aspartic acid	1.85	4.39	7.12	0.53	5.61
L-Threonine	2.81	2.07	4.02	0.51	3.33
L-Serine	4.00	2.95	5.48	0.49	4.88
L-Glutamic acid	12.67	10.51	13.80	6.06	11.99
L-α-Aminoadipic acid	0.41	0.45	0.45	0.10	0.32
Glycine	2.11	1.60	2.84	0.68	2.58
L-Alanine	6.88	2.48	5.80	1.07	6.01
L-Citruline	0.17	0.21	0.66	nd	nd
L-Valine	4.09	2.95	5.73	1.21	5.15
L-Methionine	0.95	0.73	0.85	0.32	0.94
L-Cystathione	0.10	0.10	0.21	0.03	0.20
L-Isoleucine	3.64	2.72	3.64	0.88	3.77
L-Leucine	5.57	4.29	4.40	1.56	5.07
L-Tyrosine	0.48	0.27	0.62	0.37	0.49
L-Phenylalanine	3.40	2.69	3.88	0.68	4.16
β-Alanine	0.53	0.28	0.60	0.14	0.85
L-β-Aminoisobutyric acid	nd	nd	nd	nd	nd
γ-Aminobutyric acid	0.33	0.15	0.52	0.21	0.73
L-plus allo-δ-hydroxylysine	0.38	0.40	0.91	0.08	1.65
L-Ornithine	0.33	0.40	0.94	0.03	0.00
L-Lysine	3.84	3.01	5.77	0.84	4.88
L-Histidine	1.01	0.90	1.00	0.42	1.15
L-Carnosine	0.67	0.54	1.34	0.12	1.34
L-Arginine	3.91	3.18	4.19	1.26	5.71
L-Proline	3.33	1.78	3.88	1.28	3.71
Sugars (mg ml^{-1})					
Sucrose	nd	nd	nd	nd	nd
Ribose	nd	nd	nd	nd	nd
Mannose	0.47	nd	0.66	nd	0.43
Fructose	nd	13.11	0.93	0.74	nd
Arabinose	2.78	6.57	1.31	4.65	5.49
Galactose	3.32	3.15	3.15	1.12	5.55
Xylose	0.46	1.31	0.74	5.37	0.74
Glucose	11.91	13.67	8.16	58.66	7.75
Ions (p.p.m. w/v)					
K	6140	5210	9980	1450	6700
Mg	655	510	1000	154	934
Ca	170	111	163	59	224
Fe	9.2	6.4	10.3	2.3	10.6
PO_4	4590	3750	7110	2040	5970
SO_4	94	90	217	72	103
NO_3	nd	nd	nd	nd	nd

[a]Diluted with water to four times.

nd, not detected.

From Sasaki (unpublished).

Table 6 Typical analysis of flavor constituents in Japanese shoyu

Flavor component	*Concentration (p.p.m.)*
Ethanol	31 501.10
Lactic acid	14 346.57
Glycerol	10 208.95
Acetic acid	2 107.74
4-Hydroxy-5-methyl-3 (2H)-furanone (HMMF)	256.36
2,3-Butanediol	238.59
Isovaleraldehyde	233.10
4-Hydroxy-2 (or 5)-ethyl-5-(or 2)-methyl-3 (2H)-furanone (HEMP)	232.04
Methanol	62.37
Acetol	24.60
Ethyl lactate	24.29
2,6-Dimethoxyphenol	16.21
Ethyl acetate	15.13
Isobutyraldehyde	14.64
Methyl acetate	13.84
Isobutyl alcohol	11.95
Furfuryl alcohol	11.93
Isoamyl alcohol	10.01
Acetoin	9.78
n-Butyl alcohol	8.69
4-Hydroxy-2,5-dimethyl-3 (2H)-furanone (HDMF)	4.83
Acetaldehyde	4.63
2-Phenylethanol	4.28
n-Propyl alcohol	3.96
Acetone	3.88
Methionol	3.65
2-Acetylpyrrole	2.86
4-Ethylguaiacol	2.77
Ethyl formate	2.63
4-Butanolide	2.02
Methional	1.42
4-Ethylphenol	0.34
Dimethyl sulfide	0.04

From Yokotsuka T, Sasaki M, Nunomura N and Asao Y (1980) *Journal of the Brewing Society of Japan* 75: 717.

themselves related to flavor in shoyu. In particular, these oxidative reactions give rise to an inferior organoleptic quality of shoyu, in that the existence of oxygen leads to a decrease in the number of 'good' flavor compounds, such as HEMF and HMMF, and an increase in the number of off-flavors resulting from isobutyric acid, isovaleric acid, and many pyrazine compounds. To preserve flavor and color, it is important to store shoyu at a cool temperature. (*See* **Browning**: Nonenzymatic.)

Safety

The main concern with the safety of Japanese shoyu is the long-term effects of consumption, and whether or not shoyu contains mycotoxins and mutagens.

Mycotoxins As strains of *A. oryzae* or *A. sojae* are taxonomically related to aflatoxin-producing molds, the question has arisen as to whether or not the molds produce aflatoxins. Fortunately, chemical, enzymatic, and molecular biological research has shown that none of the koji molds produce aflatoxins. Moreover, the possible presence of other mycotoxins, e.g., ochratoxins, sterigmatocystin, patulin, cyclopiazonic acid, and penicillic acid, in shoyu has been checked, and results have indicated that, with the exception of very few strains that produce cyclopiazonic acid, koji molds do not produce such toxins.

Long-term effects of shoyu consumption The long-term effects of feeding rats shoyu and its effects on the gastric mucosa have been evaluated. Animals fed shoyu were smaller than the controls, but were healthier, more active, and lived longer. Breast tumors developed in 10 control rats, but none developed in those fed shoyu, and it was concluded that shoyu did not appear to be a carcinogen in rats. Moreover, the acute toxicity of shoyu was explained as being due to its sodium chloride content.

Mutagens After an investigation of mutagens in shoyu using the Ames test, it has been found that shoyu treated with nitrite at a concentration level detected in the human mouth, i.e., less than 50 p.p.m., did not induce any mutagenicity, though mutagenic material was formed in shoyu when 2000 p.p.m. of nitrite was present at pH 3.0. It was concluded that shoyu cannot be an inducer of mutagenicity at the level of human consumption.

South-east Asian Countries

China The annual production of soy sauce in China was estimated to be 500 000 kl in 1993.

The manufacture of soy sauce in Peking and Shanghai nowadays is different to the traditional method and is as follows. The koji is prepared by a usual method on a large scale, culturing *Aspergillus oryzae* with a mixture of steamed soybeans and wheat or wheat bran (6–8:4–2), and then mixed with salt water to make a hard mash, with a moisture content of about 80% and a salt concentration of about 6–7%. This hard and low-salt mash is kept at 45–50 °C for 20–30 days for enzymatic digestion. The digested mash is extracted with hot salt water (90 °C, Be 17–19) and then with hot water. The salt-free residue is suitable for animal and bird feed. There is no alcoholic fermentation of mash or pressing of the mash, as there is in Japanese shoyu manufacture. The yield of soy sauce on a nitrogen basis is 70–80%. About 80% of Chinese soy sauce is made by this process. There are four governmental standard grades of soy sauce in China, and the standard of the highest grade is as

follows: total nitrogen 1.6%, reducing sugar 4%, sodium chloride 19% or more, respectively.

Acid hydrolysis of plant protein for soy sauce manufacture was forbidden by law in China 1979, because the contents of arsenic acid and hydrochloric acid were too high.

Taiwan The annual production of soy sauce in Taiwan was estimated to be 12 000 kl in 1994, which is equivalent to 6 l per capita per year.

About 80% is the so-called tou-yu, and this is made from soybeans and wheat originating from China. There were 433 soy sauce manufacturing plants in Taiwan in 1978, of which the eight largest producers comprised at least 45% of local market. Soy sauce products include genuine fermented soy sauce (25%), chemically hydrolyzed soy sauce (5%) and blended soy sauce (70%). There are three national standard grades of soy sauce in Taiwan, and their total nitrogen percentages were 1.4, 1.1, and 0.8 g per 100 ml, respectively, in 1994.

Five to 10% of Taiwan soy sauce is estimated to be in-yu, which is made only from black soybeans and very much resembles Japanese tamari, made only from yellow soybeans. The black soybean koji is washed with water, before being mixed with salt water to make a mash so as to remove the bitter taste of the product.

Fermented soybean sauce similar to in-yu and tamari is also produced in the southern part of China and seems to be the prototype of the soy sauce prepared only from soybeans. Tamari mash is fermented usually in wooden kegs in Japan, but the soybean sauce in Taiwan, Thailand, Singapore, Malaysia, Indonesia, and the southern part of China is fermented in China jars (approximately 150 l in volume) and in the sunshine.

Korea The annual production of soybean sauce in Korea was 420 000 kl in 1996, which included 180 000 kl of industrially produced soy sauce and 240 000 kl of home-made product. The fermented soy sauce produced industrially in Korea is of the Japanese koikuchi type, but about 70% of its production is estimated to be chemically hydrolyzed soy sauce. The home-made soy sauce is prepared by a traditional method, in which cooked soybeans are mashed and made into small balls, which are dried in the sun during the day and kept warm at night for several days in the winter. Molds grow on the surface of the balls and penetrate the interior of the balls, which are then called 'meju.' Sufficiently dried meju is packed in a rice-straw bag for aging for a few months. The aged meju balls are then placed into brine and kept in large China jars in the sunshine for several months for fermentation. The ripened meju–brine mixture is separated into supernatant liquid and sediment residue. The liquid is soy sauce, and the residue is soybean paste, which is sometimes mixed with red pepper. In recent years, better-quality meju has been prepared industrially, in which cooked soybeans are inoculated with *Aspergillus oryzae* mold in a controlled fermentation room. Housewives take the improved meju home and add it to brine to make soy sauce and paste.

Other countries In Singapore, Malaysia, and Indonesia, two types of soy sauce are manufactured: a light soy sauce (in Cantonese 'sung-show' or in Hokkien 'chiuw-cheng'), and a dark soy sauce mixed with cane molasses (in Cantonese 'low-chow' or in Hokkien 'tau-iu'). In Indonesia, soy sauce is called 'keitjap' or 'kecap,' and similarly, two types of kecap are popular in Sumatora and the latter in Java.

The per capita daily consumption of kecap in Indonesia is 10–15 ml. The kecap mash is washed with salt water twice and the mixed extract is concentrated in the sun. The residue from the extraction is used for tauco or miso of an inferior grade, but the genuine tauco is made from soybeans and wheat flour cultured with *Aspergillus oryzae* mold. The molded-cultured material, or 'koji', is dried in the sun, mixed with salt water, and fermented and concentrated in the sun for 3–7 weeks. Tauco is a slurry containing 10% protein, and is sold in glass bottles, thus differing from Japanese miso, which is a paste containing 25% protein. In the aforementioned countries the chemical acid hydrolysate of soybeans and the semichemical or shinshiki soy sauce are also popular, but the amount of extract produced is not known.

See also: **Browning**: Nonenzymatic; **Carbohydrates**: Classification and Properties; **Fatty Acids**: Properties; **Lactic Acid Bacteria**; **Mycotoxins**: Occurrence and Determination; **Phenolic Compounds**; **Phospholipids**: Properties and Occurrence; **Protein**: Chemistry; Digestion and Absorption of Protein and Nitrogen Balance; **Soy (Soya) Beans**: The Crop; Processing for the Food Industry; **Starch**: Structure, Properties, and Determination; **Wheat**: The Crop; Grain Structure of Wheat and Wheat-based Products

Further Reading

Arora DK, Mukerji KG and Marth EH (eds) (1991) *Handbook of Applied Mycology, Foods and Feeds*, vol. 3. New York: Marcel Dekker.

Chichester CO, Mrak EM and Schweigert BS (eds) (1986) *Advances in Food Research*, vol. 30. London: Academic Press.

Hesseltine CW and Wang Hwal (eds) (1986) *Indigenous Fermented Food of Non-Western Origin*. Berlin: J Cramer.
Perlman D (ed.) (1974) *Applied Microbiology*, vol. 17. London: Academic Press.
Reddy NR, Pierson Merle D and Salunkhe DK (eds) (1986) *Legume-based Fermented Foods*. Boca Raton, FL: CRC Press.
Steinkraus KH (ed.) (1983) *Handbook of Indigenous Fermented Foods*. New York: Marcel Dekker.
Steinkraus KH (ed.) (1989) *Industrialization of Indigenous Fermented Foods*. New York: Marcel Dekker.
Wood BJB (ed.) (1985) *Microbiology of Fermented Foods*, vol. 1. London: Elsevier.

FERMENTED MILKS

Contents

Types of Fermented Milks

H Roginski, The University of Melbourne, Werribee, Victoria, Australia

Background

Milk spontaneously fermented with the adventitious microflora has traditionally been consumed in all parts of the world. Historically, fermented milks were made at the same site where the fresh milk was produced, i.e., on farms or near the herds of dairy animals tended by nomadic tribes. Allowing milk to ferment spontaneously was a simple way of preserving milk solids because most spoilage microorganisms do not grow well at pH 4.5 or lower, characteristic of fermented milks. Fermented milk could be stored for an extended period compared with the raw milk. In warm climates, whey was often partly removed from fermented milks to create an environment in which milk solids could be further preserved.

The sensory and physical attributes of traditional fermented milks reflect the local environment and climatic conditions that facilitate the selection of microorganisms that could dominate the microflora of milk. It is a mixed population of lactococci and lactobacilli, and in some cases, also yeasts and molds, which exist together in a protocooperative relationship. As a result of their spontaneous growth, which normally proceeds in a relay-like fashion, where distinctly different groups of microorganisms dominate the milk environment in succession, various metabolites are formed, and these in turn affect the chemical status and physical appearance of the product. The most important of these metabolites is lactic acid, which causes coagulation of casein micelles at pH 4.6. The bacteria that produce lactic acid can be either mesophilic or thermophilic, homofermentative or heterofermentative.

Lactic acid and other metabolites of the milk microflora have a preservative effect on milk, primarily by lowering the pH of milk and, as we now know, through the production of bacteriocins and peptides, some of which may inhibit spoilage microorganisms.

There is archeological evidence that fermentation of milk has been known for millennia. There is also a reference to a fermented milk product in the Bible (Genesis 18:8). Nowadays, a multitude of fermented milks, sometimes also called 'cultured milks,' are produced commercially world-wide. Most of these products reflect long-standing traditions in different parts of the world. Some of these traditional products have gained world-wide acceptance, such as yogurt, which originated in the Balkans, and is now known and produced in most countries.

Also, a number of products have been developed as a result of progress in clinical and industrial microbiology in the last century. Such fermented milks are manufactured with the use of starter cultures for which various health-promoting effects have been claimed.

According to the definition proposed in 1963 by the International Dairy Federation, 'fermented milks are products prepared from milk, skimmed or not, concentrated or not, with specific cultures; the microflora is kept alive until sale to the consumer and may not contain any pathogenic germ.' Such an understanding of the term 'fermented milks' appears to have been commonly accepted in the dairy world. This term implies a liquid or semiliquid consistency of the product and therefore, in common understanding,

does not apply to cheese. The removal of whey in the cheese manufacturing process seems to constitute the most significant demarcation line between 'fermented milks' and cheeses. However, whey is partly removed during manufacture of a small number of products that are traditionally classified as fermented milks.

The majority of known fermented milks are made from cows' milk, but milk from buffaloes, sheep, goats, horses, camels, yaks, reindeer, and asses is also used for this purpose. Fermented milk was originally obtained by mixing fresh milk with the remainder of the previous batch of the product. Most home-made fermented milks are still prepared in this way. Commercial production of fermented milks is not complicated, with the characteristics of the final product being determined chiefly by the metabolic properties of microorganisms used in manufacture and the treatment of milk before the fermentation. A crude empirical selection of cultures for the manufacture of fermented milks began in antiquity, but it was not until the twentieth century that impressive progress was made in starter culture selection and improvement techniques. Knowledge of the physiology of commercially useful microorganisms is still advancing, and this, as well as applications of genetic engineering to starter organisms, will expand the range of technological manipulation of starter cultures and result in a better control of manufacturing process.

Consumption of fermented milks varies widely between different regions of the world. People of Northern and Western Europe have traditionally consumed, for climatic and historical reasons, relatively large amounts of fermented milks. For example, in Finland, the per capita consumption of all fermented milks in 1995 was 38.5 kg, of which 15.5 kg (40%) was yogurt. Consumption of fermented milks is also high in Bulgaria, Israel and several other countries.

Classification of Fermented Milks

Different criteria can be employed as a basis for classification of fermented milks, such as the region of origin, chemical composition and physical characteristics, temperature and type of fermentation, and types of microorganisms involved in fermentation.

Classification of fermented milks proposed by Kurmann, presented below in a modified form, underlies the importance of the starter culture in determining the characteristics of the product:

1. Fermented milks made with thermophilic bacteria, incubation temperature 30/35–40/45 °C.
   1.1. Lactic acid fermentation, without producing appreciable amounts of gas and alcohol.
   Yogurt and similar traditional fermented milks: yogurt (Bulgaria, Turkey, etc.), dahi (India), laben (Saudi Arabia), mast (Iran), matzoon (Armenia), tarho (Hungary), zabady (Egypt). Diluted yogurt: eyran (Turkey), doogh (Iran). Whey-drained (concentrated) yogurt: labneh, lebneh (Lebanon and other Arab countries), kashk, kaskg (Middle East/Iran), tan, than (Armenia), tulum, torba (Turkey), skyr (Iceland). Dehydrated (dried) yogurt: kurut (Middle East/Turkey), karut (Pakistan), quurut (Afghanistan), khurud (Mongolia).
   1.2. Acid fermentation, without producing appreciable amounts of gas and alcohol, using mainly human intestinal bacteria.
   Single-strain fermentation: acidophilus milk, 'bifidus' milk, yakult.
   Mixed-strain cultures of different formulae: AB type (*Lactobacillus acidophilus* and *Bifidobacterium* spp.); ABT type (*Lb. acidophilus, Bifidobacterium* spp., and *Streptococcus thermophilus*); BAP type (*Bifidobacterium* spp., *Lb. acidophilus*, and *Pediococcus* spp.); and other formulations.
2. Fermented milks made with mesophilic bacteria, incubation temperature 10/15–20/30 °C.
   2.1. Lactic acid fermentation with simultaneous production of slime.
   Scandinavian fermented milks: långfil/tätmjölk (Sweden), viili (Finland), tettemelk (Norway), and similar products.
   2.2. Lactic acid fermentation using butter cultures. Fermented milks prepared with butter cultures.
   Artificial buttermilks: 'cultured buttermilk' (which is a type of cultured milk produced in North America) and similar products.
   2.3. Concentrated fermented milks.
   Commercial products, *e.g.*, ymer and lactofil (Scandinavia).
   Traditional home-made milks, *e.g.*, Kellermilch and Lagermilch in German-speaking areas of Europe.
   2.4. Mixed lactic acid and ethanol fermentation.
   Kumys (North Central Asia), leben, laban (Lebanon, Iraq, and Egypt), and other similar products.
   Kefir (Caucasus) made with kefir grains.
   Artificial preparations, *e.g.*, kefir made without grains.

3. Mixed material plant–milk fermentations
   3.1. Products where plant material is a substrate for fermentation
      Kishk (Middle East), niyogurt (Philippines).
   3.2. Products where plant material is claimed to be a carrier of specific microorganisms and/or enzymes.
      Nordic ropy milks.
4. Various unclassified fermented milks.
5. Buttermilks
   5.1. Conventional cultured buttermilk, a byproduct of cultured butter manufacture.
   5.2. Cultured buttermilk obtained by fermentation of a sweet buttermilk.
   5.3. Yogurt buttermilk obtained by churning yogurt into butter and a liquid byproduct
6. Cultured creams

The major classes of fermented milks and their interrelationships can be graphically presented in the form of a 'tree of fermented milks' (see Figure 1), which shows the diversity of these products, further enriched by various physical forms in which fermented milks can be offered to consumers.

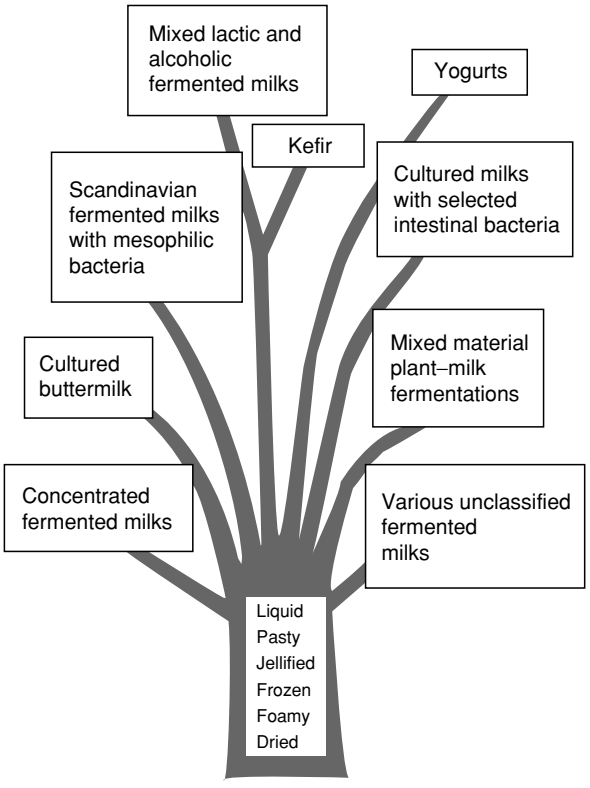

Figure 1 Tree of fermented milk types. From J.A. Kurmann (1984) Aspects of the production of fermented milks. In: *Fermented Milks – Proceedings of an IDF seminar in Avignon*, pp. 16–26. International Dairy Federation Bulletin No. 179. Brussels: International Dairy Federation (1984), with permission.

There are many variations and modifications to these products in different parts of the world, which usually reflect local traditions as well as continuous efforts to develop new products. For example, various types of fermented whey beverages have gained popularity in some European countries.

Taxonomic Location of Starter Cultures

The main thermophilic bacteria used in the production of fermented milks include lactobacilli, *Streptococcus thermophilus* and bifidobacteria. DNA fingerprinting is now commonly used to determine the taxonomic location of these organisms. As a result, important changes have been made to the taxonomic status of certain well-known starter bacteria. For example, organisms previously classified as *Lactobacillus acidophilus* are now divided into the following six species: *Lb. acidophilus*, *Lb. crispatus*, *Lb. gasseri*, *Lb. amylovorus*, *Lb. gallinarum*, and *Lb. johnsonii*. Similarly, the lactobacilli previously classified as *Lactobacillus casei* are now located within either *Lb. rhamnosus* or *Lb. paracasei* ssp. *paracasei*. The extensively studied and well-publicized *Lactobacillus* GG is now a strain of *Lb. rhamnosus*.

The main organisms used in mesophilic starters belong to *Lactococcus lactis* (with subspecies *lactis* and *cremoris*) and *Leuconostoc* spp. Yeasts, belonging mainly to *Saccharomyces* spp., and molds such as *Geotrichum candidum*, are also included in some mesophilic starter cultures.

Manufacturing Technology

Manufacture of fermented milks consists of several basic steps. Most fermented milks can be produced as either 'set' or 'stirred' products, and this feature determines the sequence of major operations (see Figure 2).

Preparation of Milk

Milk for production of fermented milks should comply with the following requirements: low bacterial counts, absence of pathogenic organisms and absence or very low concentrations of various inhibitory substances that may find their way into milk, such as residues of antibiotics from the treatment of mastitis, residues of dairy sanitizers, *etc*.

The first step in milk preparation is usually the clarification of milk, to remove mechanical impurities and somatic cells. This operation is carried out in centrifugal clarifiers/separators, but filtration of raw milk can also be used for this purpose.

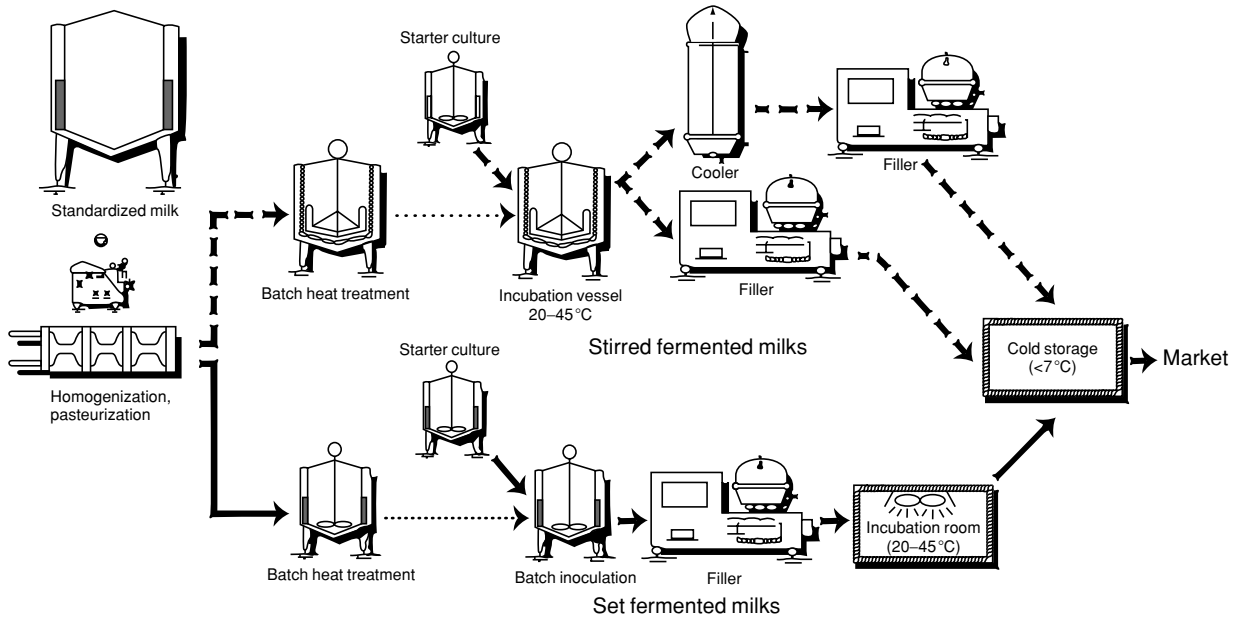

Figure 2 Basic steps in manufacture of fermented milks. From International Dairy Federation (1988) *Fermented Milks – Science and Technology*. International Dairy Federation Bulletin No. 227. Brussels: International Dairy Federation, with permission.

The milk has to be standardized, i.e., its composition altered to meet legal requirements and/or to improve the economy of production and/or to modify the characteristics of the final product. Usually, standardization means adjustment of fat content and sometimes also of nonfat solids (NFS) content. NFS content in milk can be increased by (1) evaporation, which usually removes 10–20% water, (2) addition of skim milk powder, or (3) addition of milk concentrate. Membrane filtration, especially ultrafiltration (UF), is increasingly being used to concentrate milk for production of yogurt and various other fermented milks. The main advantages of using UF retentates in production of fermented milks are: (1) enhanced viscosity; (2) reduced syneresis (even at lower fat content); (3) increased product yield (in manufacture of concentrated fermented milks) due to a partial retention of whey proteins; and, quite often, (4) better taste due to the absence of 'chalkiness' frequently associated with the use of milk powder.

Many fermented milks, including yogurt, can be made from reconstituted or recombined milk. This technique has obvious advantages in locations where it is difficult to obtain raw milk of a consistently high microbiological standard. However, fermented milks made from a totally reconstituted or recombined milk account for only a small fraction of the total production of fermented milks.

The next manufacturing step is homogenization of milk, which, by reducing the diameter of the fat globules, prevents fat separation and has a whitening effect on the milk. Homogenization also improves (1) the stability of fermented milks during storage (by reducing syneresis) and (2) their rheological properties. The usual conditions of homogenization are 10–20 MPa at 50–60 °C. Higher temperatures are sometimes used, e.g., 65 °C in production of filmjölk, a Scandinavian fermented milk.

Homogenization is followed by heat treatment, in order to: (1) pasteurize the product; (2) make the milk a more nutritious growth medium for the starter organisms, through the release of amino acids and other growth factors, reduction of the redox potential and elimination of inhibitory substances; (3) improve physical properties of fermented milks and reduce syneresis; and (4) prevent formation of rancidity through inactivation of lipases. To achieve these objectives, it is common practice to batch-pasteurize milk at 80–90 °C for 20–30 min. Heating of milk to the required temperature in a plate or tubular heat exchanger is followed by pumping into a holding unit where the temperature is maintained for the predetermined period. In some plants, milk is heated in a continuous process in a closed system at 90–95 °C for 5 min or more. Heat exchangers are also used to cool milk to the fermentation temperature. In cases where the only objective of heat treatment is pasteurization of milk, high-temperature, short-time (HTST) pasteurization is usually used.

Another step recommended in manufacture of some fermented milks is the deaeration of milk, which is said to reduce syneresis and improve the consistency of the final product.

Preparation of Starters

Starters for the commercial production of fermented milks can be prepared in several different ways, depending on the type of final product and the type of manufacturing facilities. The majority of fermented milks are now made with defined starters, but for the production of some traditional fermented milks, for example kumys and leben, undefined starters are usually used. In these starters, various lactic acid bacteria and yeasts remain in a protocooperative and possibly also a symbiotic relationship. Traditionally, a starter was a residue from the previous batch that was added to suitably prepared milk. Nowadays, such starters are propagated aseptically in the laboratory to serve as an inoculum for the bulk starter unit.

The majority of fermented milks are made with multiple-strain starters, which are defined mixtures of two or more strains. Apart from enabling manufacturers to make better-quality products, multiple-strain starters can cope better with bacteriophage challenge if suitable bacteriophage-unrelated strains are selected, though this may not always be feasible. A few types of fermented milks are made with defined single-strain starters. Milks fermented with selected intestinal bacteria belong to this category, *e.g.*, yakult, made with *Lactobacillus paracasei* ssp. *paracasei* (formerly known as *Lactobacillus casei* strain Shirota), and acidophilus milk, made with *Lb. acidophilus*.

The importance of the physical characteristics of fermented milks is reflected in the use of microorganisms which synthesize a glycocalyx, a polysaccharide or glycoprotein material which forms 'capsules' around the cells that produced it. Such capsules, sometimes referred to as 'slime' or 'mucus', act as a stabiliser in the product, improving viscosity, disguising syneresis of the milk protein gel, and making the product less susceptible to thixotropic phenomena. Capsule-forming organisms are incorporated into multiple-strain starters used in the manufacture of many types of fermented milks.

In the conventional starter preparation methods, several subcultures in the laboratory are needed to produce, from the initial freeze-dried or liquid culture, the inoculum volume required for the bulk starter unit. This time-consuming step needs skilled personnel to ensure effective quality assurance at all stages of starter preparation. Because of demands on manpower, space and capital, many manufacturers of fermented milks, rather than preparing their own starters, prefer to rely on concentrated commercial starters, which contain up to 10^{11} cells per ml and can be added to the bulk starter unit or directly to the production tank or vat (conventional starters may contain up to 10^{9} cells per ml). Concentrated starters are available in either deep-frozen or freeze-dried form.

Starter cultures selected for the manufacture of fermented milks should not produce appreciable amounts of lactic acid in the product during refrigerated storage. Postfermentation acidification usually leads to product defects such as syneresis of the protein coagulum and the consequent expulsion of whey.

Fermentation

The temperature and time of fermentation can be varied to influence the final product characteristics. The fermentation temperature usually does not differ much from the optimal growth temperature of the starter organisms. It is essential that the temperature remains constant during fermentation. Sometimes, in order to improve the physical properties of the final product, manufacturers choose a lower temperature and a longer time of fermentation to promote the formation of mucoidal capsules by some starter bacteria. For the stirred product, fermentation is carried out in the production tank or vat; the 'set' product is fermented in retail containers.

Cooling

Fermentation is terminated by cooling which, depending on whether the product is stirred or cup-set, can be achieved by different techniques. Cooling of the product, usually to refrigeration temperature, allows the manufacturer to control acidity and initiate the 'cold gelation' of the curd at the appropriate time. The cooling rate is often critical to the quality of the product and is usually determined empirically by each manufacturer. In general, manufacturers aim to lower the temperature of the 'set' product to $< 10\,^{\circ}C$ as quickly as possible, and then continue cooling it at a slower rate to the storage temperature.

In the manufacture of stirred fermented milks, cooling is often carried out in two stages: at first to approximately $20\,^{\circ}C$, at which point the flavor ingredients are added, and the product is packaged into retail containers, and second to the refrigeration temperature. The second stage of cooling is often conducted at a slow rate and may take up to 12 h. This method is said to result in improved rheological characteristics of the product.

Flavoring

Fermented milks are often flavored by the addition of fruit pulp or purée or other flavoring ingredients. Fruit pulp, with a sugar content of approximately 60%, is usually added at the level of about 10%. Pectins present in fruits exert a stabilizing effect on the product, but additional stabilizers may be added

to the product at this stage. These include high-molecular-weight compounds such as gelatin, or various hydrocolloids of plant origin ('gums'), including compounds derived from some seaweeds. For the stirred product, flavoring ingredients are added after or during the cooling stage, whereas the flavoring of the 'set' product is carried out prior to fermentation and filling into retail containers.

Packaging

Fermented milks are packed in containers impermeable to water and odors, insoluble in water, and free from foreign odors. In some applications, impermeability to light, and low permeability to oxygen are also required. Glass, often dyed to reduce the effect of light on the product, was used for a long time as a packaging material for fermented milks, but has now been almost totally replaced by synthetic materials and paper coated with various polymers such as polyethylene. The paper carton, usually formed into a rectangular container, can also be impregnated with resins and/or coated with aluminum or silicone. Synthetic materials ('plastics') are light and can be easily molded into light-weight containers of diverse shapes. The cup is the most common shape of container, but bottles and pouches are also used. Containers for products containing still-viable yeast cells may have a three-layer lid that allows the escape of carbon dioxide generated during storage.

Regulatory Aspects

Fermented milks offered for sale must conform to legal requirements, which, apart from setting microbiological quality standards to safeguard public health, usually define the chemical composition of the product, its shelf-life and, in many cases, the number of viable cells of starter bacteria that must be present in the product at the point of sale. In Japan, for example, a product classified as fermented milk or 'lactic milk drink' must contain at least 10^7 viable cells of lactic acid bacteria and yeasts per milliliter. Standards for fermented milks in some countries include the requirement for the live cells to be present in the product at the end of its shelf-life (declared on the package as the 'use-by' date) but do not specify the minimum numbers. The compositional requirements, usually specified as the minimum content of NFS and fat, differ widely between countries.

The final acidity is often included in the standards as (1) the minimum concentration of lactic acid, (2) the minimum titratable acidity according to an arbitrary scale, e.g., degrees Soxhlet–Henkel (°SH), Dornic (°D), or Thörner (°Th), or (3) the maximum pH of the product.

Products universally known, such as yogurt, are subject to fairly detailed regulations, whereas less well-known products may not have legally binding standards of identity and have to comply with requirements set for other products deemed similar for purposes of control by relevant government authorities.

Health-related Claims

Traditions of folk medicine in many parts of the world have endorsed fermented milks as beneficial not only in the prevention of various health disorders but, sometimes, also as an aid in treatment of these disorders. In addition to traditional products, fermented milks made with 'probiotic' cultures have expanded the range of claimed health benefits. For some of these effects, it would be difficult to separate the effect of starter organisms alone from that of milk fermented with these organisms.

Studies of human intestinal flora are fraught with methodological difficulties. It is not easy to obtain samples of intestinal contents that would accurately reflect quantitative and qualitative relationships in the gut. In addition, the microbial population of the gut is rather stable in healthy individuals and is not affected by the diet. Lactobacilli and bifidobacteria ingested with the diet do not supplant existing intestinal bacteria. However, sufficiently rigorous evidence has now been gathered to suggest that in some disease states, consumption of fermented milks may benefit the host.

In general, health benefits claimed for many types of fermented milks refer to their effect on (1) the microbial ecology of the gut and/or (2) specific ailments unrelated to the gut ecology. Health claims in both these categories include:

- immunomodulatory effects, including enhancement of antitumor activity;
- antimutagenic properties;
- stabilization of gut mucosal barrier, which may lead, among other things, to suppression of allergic reactions;
- prevention and treatment of diarrhea caused by rotavirus and bacterial infections;
- alleviation of radiotherapy-induced diarrhea;
- restoration of normal ecological relationships in the gut after antibiotic treatment;
- alleviation of constipation;
- inhibition of pathogenic organisms;
- alleviation of lactose intolerance;
- lowering of serum cholesterol;
- antihypertensive effects.

It would be reasonable to expect that relationships in the gut can be influenced primarily by the

thermophilic organisms ingested with fermented milks. Beneficial properties of the mesophilic bacteria would be due to their metabolites present in the product before they are consumed, rather than to any lasting effect on the microflora of the gut. In this context, it is interesting to note that an organism known as *Lactococcus lactis* L1A, isolated from samples of fermented milks produced on farms in northern Sweden, was recovered from fecal samples up to 15 days after the consumption of fermented milk containing it had ceased.

During the past few decades, there has been a noticeable revival of popular and scientific interest in 'probiotic' microorganisms and fermented milks. Many of the health benefits claimed for these microorganisms and products still remain to be substantiated by unequivocal experimental and/or epidemiological evidence.

See also: **Cheeses**: Chemistry of Gel Formation; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Functional Foods**; **Homogenization**; **Lactic Acid Bacteria**; **Milk**: Processing of Liquid Milk; **Packaging**: Packaging of Liquids; **Probiotics**; **Recombined and Filled Milks**; **Starter Cultures**

Further Reading

Bottazzi V (1983) Other fermented dairy products. In: Reed G (ed.) *Biotechnology: Food and Feed Production*, vol. 5, pp. 315–363. Verlag Chemie, Weinheim.

International Dairy Federation (1984) *Fermented Milks – Proceedings of an IDF Seminar in Avignon*. International Dairy Federation Bulletin No. 179. Brussels: International Dairy Federation.

International Dairy Federation (1988) *Fermented Milks – Science and Technology*. International Dairy Federation Bulletin No. 227 Brussels: International Dairy Federation.

International Dairy Federation (1996a) *Nutrition Newsletter No. 5*. Brussels: International Dairy Federation.

International Dairy Federation (1996b) *The World Dairy Situation 1996*. International Dairy Federation Bulletin No. 314. Brussels: International Dairy Federation.

Kosikowski FV and Mistry VV (1997) *Cheese and Fermented Milk Foods, Origins and Principles*, 3rd edn, vol. I. Westport, CT: FV Kosikowski, L.L.C.

Kurmann JA, Rasic JLJ and Kroger M (1992) *Encyclopedia of Fermented Fresh Milk Products*. New York: Van Nostrand Reinhold.

Litopoulou-Tzanetaki E and Tzanetakis N (1999) Fermented milks – range of products. In: Robinson RK, Batt CA and Patel PD (eds) *Encyclopedia of Food Microbiology*, pp. 774–784. London: Academic Press.

Marshall VM (1987) Fermented milks and their future trends. I. Microbiological aspects. *Journal of Dairy Research* 54: 559–574.

Nakazawa Y and Hosono A (eds) (1992) *Functions of Fermented Milk – Challenges for the Health Sciences*. London: Elsevier Applied Science.

Robinson RK (ed.) (1991) *Therapeutic Properties of Fermented Milks*. London: Elsevier Applied Science.

Tamime AY and Robinson RK (1988) Fermented milks and their future trends. II. Technological aspects. *Journal of Dairy Research* 55: 281–307.

Tamime AY and Robinson RK (1999) *Yoghurt: Science and Technology*, 2nd edn. Boca Raton, FL: CRC Press.

Wood BJB (ed.) (1992) *The Lactic Acid Bacteria in Health and Disease*. London: Elsevier Applied Science.

Products from Northern Europe

H Roginski, University of Melbourne, Werribee, Victoria, Australia

Introduction

Fermented milks have always played an important role in the nutrition of people inhabiting northern Europe. In Scandinavia in particular, consumption of fermented milks has a long tradition. Historically, fermented milks, often referred to as sour milks, were made and consumed at home. Most of these products have a characteristic thick consistency and ropiness, and could be kept for weeks or even months in a cool room.

The home production of fermented milks is now much less common in Scandinavia, but many types of commercially manufactured products have the same attributes as their homemade precursors. Consumption statistics provide ample evidence of the continuing demand for Nordic fermented milks, although the growing popularity of yogurt has led to some erosion of this traditional market in recent decades.

Interest in Nordic fermented milks has recently been stimulated by reports of prophylactic and therapeutic effects associated with starter flora and its metabolites, which have been observed in experiments on animals.

General Features

Nordic fermented milks constitute a group of products distinctly different from fermented milks made elsewhere primarily due to their unique physical properties, characterized by high viscosity and ropiness. In the cool climate of northern Europe, homemade fermented milks are the result of the spontaneous growth of mesophilic microflora present in raw milk. Low

temperatures seem to be critical for the domination of the milk environment by encapsulated strains of lactococci. Lactobacilli are isolated only rarely from spontaneously fermented milk in Scandinavia. Molds and lactose-fermenting yeasts grow well in acidified milk and are therefore found in some types of Nordic fermented milks, and influence their properties.

Other traditional fermented products popular in Scandinavia include cultured buttermilk and cultured creams, which have a characteristic aroma and flavor, derived primarily from diacetyl and related volatile compounds. Also popular are concentrated milks, made by the removal of whey from the fresh coagulum or, more recently, using membrane-concentrated milk.

Microorganisms Present

Diverse cultures are used commercially in Scandinavia for various types of fermented milk, and have been derived from traditional products. In many of these cultures, the bacteria responsible for ropiness are indispensable. However, the presence of nonropy organisms is also essential in many cultures to insure the formation of metabolites which impart typical organoleptic properties to the product.

Overall, the following groups of microorganisms have been isolated from various types of traditional Nordic fermented milks:

- Mesophilic bacteria: *Lactococcus lactis* sspp. *cremoris, lactis*, and *lactis* biovar *diacetylactis; Leuconostoc mesenteroides* sspp. *cremoris* and *dextranicum*.
- *Lactobacillus helveticus*. It is significant that the optimum growth temperature of a strain isolated from the Norwegian and Finnish version of tätmjölk was 10 °C lower than that reported for other strains of this species. Thermophilic *L. helveticus* has been isolated from the Icelandic fermented milk skyr.
- *Streptococcus thermophilus* and *Lactobacillus delbrueckii* ssp. *bulgaricus*, found in skyr.
- The yeasts *Candida kefir* and its teleomorph *Kluyveromyces marxianus*, as well as *Torulopsis holmii* and its teleomorph *Saccharomyces exiguus*, have been isolated from various types of Nordic fermented milk. The long shelf-life of ropy milk has been attributed in part to the presence of yeasts, which are inhibitory against a number of potential mold contaminants. The inhibitory compound was inactivated by heat.
- The white mold *Geotrichum candidum* is found in viili, tette, and other products. *G. candidum* isolated from a sample of ropy milk and added to a ropy culture (*L. lactis* ssp. *cremoris*), which was previously not associated with this mold, stimulated the growth of the culture in milk. A cell-free filtrate of the medium (sterile whey) on which the mold had grown had a similar effect.

Forms of lactococci classified as 'intermediate' between *Lactococcus lactis* sspp. *cremoris* and *lactis* (on the basis of their physiological characteristics) have also been found in some ropy milks.

The total number of viable bacteria in these products can be very high. For example, up to 10^{10} cfu g^{-1} have been reported in the Finnish ropy milk called pitkäpiimä. This is a higher concentration of live bacteria than that in some starter cultures used in the commercial production of fermented milks.

Traditional Nordic fermented milks owe their characteristic high viscosity and ropiness to the vigorous growth of capsule-forming lactococci, mainly *L. lactis* ssp. *cremoris*. Encapsulated strains of *L. lactis* sspp. *lactis* and *Leuconostoc* species were also isolated from some products. However, strains of these species which clearly do not produce capsules are also found in Nordic fermented milks.

Glycocalyx

Terms like 'slime,' 'capsule,' and 'mucoidal (Muc^+) phenotype' are used interchangeably in the literature, usually without ascertaining whether the material in question is capsule-type or slime-type. 'Capsule' is defined as a compact layer of polysaccharide and 'slime' as a diffuse layer of polysaccharide, both exterior to the cell wall. As a rule, the capsule of lactococci is tightly bound to the cell that produced it. In cases where the physical nature of the material has not been studied, it would be more appropriate to use 'glycocalyx,' a general term which describes the polysaccharide-containing material lying outside the cell wall, although this term is seldom used.

Figure 1 shows an encapsulated strain of *L. lactis* ssp. *cremoris*.

In some places, the capsular material clearly does not follow the chain of lactococci. It is possible, however, that the detachment from the cells occurred during the preparation of the slide. Slime seems to be produced in particular abundance during the late exponential and early stationary phases of growth.

Normally, both encapsulated and nonencapsulated forms of the same species can be isolated from a product. As an example, the types and numbers of bacteria isolated from pitkäpiimä are shown in Table 1.

L. lactis ssp. *cremoris* was the organism most often encountered in this product. It is worth noting that encapsulated *L. lactis* subsp. *lactis* was found only in samples of ropy milk which did not contain slime-forming cells of *L. lactis* subsp. *cremoris*.

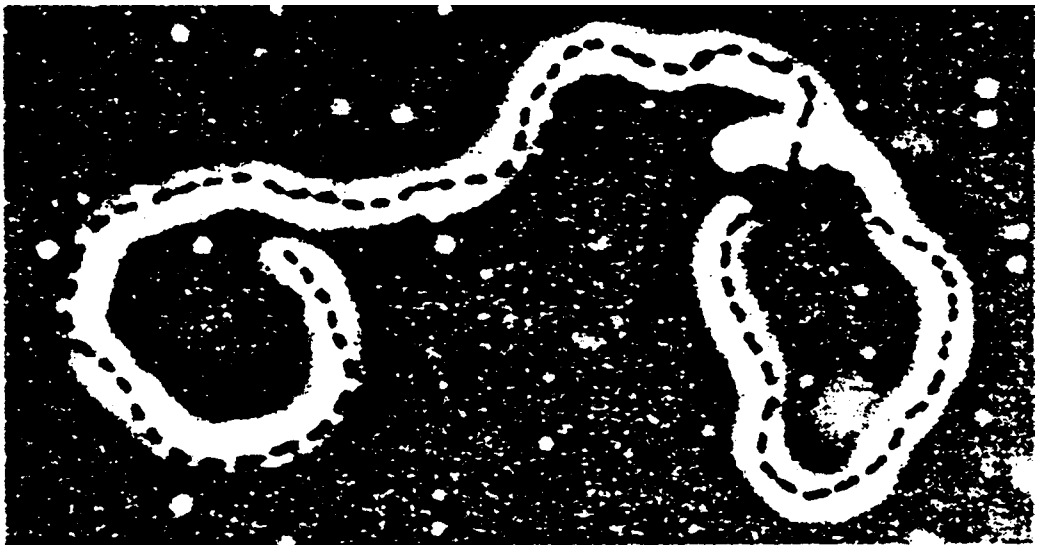

Figure 1 *Lactococcus lactis* ssp. *cremoris*, strain Va, after 6 h growth. Preparation stained with methylene blue, magnification × 1350. From Forsén R (1966) Die Langmilch (Pitkäpilmä). *Meijeritieteellinen Aikakauskirja* 26: 1–76 with permission.

Table 1 Types of bacteria identified in pitkäpiimä; the numbers of slime-forming colonies are shown in parentheses

Ropy milk culture	*Number of characterized colonies*	Leuconostoc *spp.*	L. lactis *subsp. lactis*	L. lactis *subsp.* lactis	L. lactis *subsp.* cremoris	Streptococcus viridans *group*[a]
H_1	29 (4)	9	2 (2)		18 (2)	
H_4	16 (5)	3	6 (1)	2	5 (4)	
H_5	20 (2)	6	3	4 (2)	7	
VA	42 (10)	5 (5)	7 (4)	30 (1)		
Jo	13 (3)		7 (1)		2 (2)	4
In	17 (4)	2	3 (1)	4	8 (3)	
Total	137 (28)	(5)	(9)	(3)	(11)	
Percentage of slime-forming colonies	100%	18%	32%	11%	39%	

[a]*Streptococus viridansä* physiological group includes *S. thermophilus*, but the result reported above does not imply the presence of *S. thermophilus* in pitkäpiimä.
From Forsén R (1966) Die Langmilch (Pitkäpiimä). *Meijeritieteellinen Aikakauskirja* 26: 1–76 with permission.

The growth of slime-forming lactococci isolated from pitkäpiimä depended on the presence of nicotinic acid and riboflavin in the growth medium. It has also been observed that strains of both *L. lactis* sspp. *cremoris* and *lactis* found in pitkäpiimä were more likely to grow at 40 °C than were the strains of these subspecies isolated from commercial starter cultures for fermented milks. The O_2 demand and CO_2 production of slime-forming cultures of *L. lactis* sspp. *cremoris* and *lactis* from pitkäpiimä has been observed, at 19 °C, to be two to three times higher than that of the strains which do not produce slime.

Characteristics of Capsular Material

Studies on ropy fermented milk using scanning electron microscopy showed that the polysaccharide slime material remains associated with the cells which produced it, while also attached to casein micelle clusters, forming a network responsible for the ropy consistency of the product. Other milk proteins may also be involved. In addition, proteins of bacterial origin may be present in slime. However, the results of studies on slime composition are affected, to some extent, by the technique used to isolate and purify the slime.

In one study, the slime produced by a strain of *L. lactis* ssp. *cremoris* isolated from a Swedish ropy milk was found to contain 47% protein, 20% methyl pentoses, approximately 10% hexose-like sugars, and almost 3% sialic acid. The composition of a phosphorus-containing polysaccharide produced by *L. lactis* subsp. *cremoris* isolated from a sample of the Finnish fermented milk viili is shown in Table 2.

Rhamnose, glucose, and galactose were present in this phosphopolysaccharide at the molar ratio of 1:1.45:1.75. The purification procedure (preparative

Table 2 Composition of the slime material and phosphopolysaccharide excreted by *Lactococcus lactis* ssp. *cremoris* SBT 0495

	Content (%)	
	Crude slime material	*Purified polysaccharide*
Total carbohydrate	42.3	78.9 (90.5)[a]
Rhamnose		21.7
Glucose		31.4
Galactose		38.0
Methylpentose	20.6	21.4
Hexosamine	4.8	ND[b]
Sialic acid	0.8	ND
Uronic acid	1.9	ND
Glycerol		ND
Phosphorus	1.8	3.1
Protein	21.2	ND

[a]Values after hydrofluoric acid cleavage.
[b]Not detected.
From Nakajima H, Toyoda S, Toba T *et al*. (1990) A novel phosphopolysaccharide from slime-forming *Lactococcus lactis* subspecies *Cremoris* SBT 0495. *Journal of Dairy Science* 73: 1472–1477 with permission.

sodium dodecyl sulfate–polyacryl amide gel electrophoresis) removed protein, glycerol, hexosamine, sialic acid, and uronic acids from the crude slime material. The phosphate group was attached to the β-D-galactopyranosyl residue. The molecular weight of this phosphopolysaccharide was estimated to be 1.7×10^6 by gel permeation chromatography. The phosphopolysaccharide has a negative charge; it is, therefore, possible that it could form complexes with basic proteins found on the cell surface of lactococci, due to electrostatic interactions.

Determinants of Glycocalyx Formation

Plasmid DNA Observations of the loss of the mucoidal (Muc+) phenotype at higher incubation temperatures led to studies which have demonstrated the involvement of plasmid DNA in the synthesis of glycocalyx by lactococci. The ropy phenotype of strains of *L. lactis* ssp. *cremoris* was found to be associated with a 17-MDa plasmid in Swedish strains and with a 30-MDa plasmid in Finnish strains. This indicates that distinctly different plasmids coding for the ability to produce slime are present in separate populations of starter bacteria. Plasmids on which slime synthesis is encoded can be transferred to strains with no slime-forming ability by a number of currently available techniques, including electroporation or cotransfer involving a bigger plasmid, on which an ability to ferment lactose is encoded.

Proteins of the cell surface Two cell surface proteins of mol. wt 26 kDa and 42 kDa have been shown to be associated with slime formation in lactococci. The synthesis of these proteins is probably plasmid-encoded.

Lipoteichoic acid Experiments demonstrating the presence of lipoteichoic acid (LTA) antigen on the cell surface of *L. lactis* ssp. *cremoris* led to a hypothesis linking LTA with the formation of a capsule or slime. LTA, which was also detected in the growth medium during early stationary phase and during storage at refrigeration temperature, may form associations with proteins and polysaccharides of the slime material. In addition, LTA present on the cell surface inhibits autolytic enzymes of the bacterial cell wall, which has obvious advantages for the shelf-life of the product. LTA also interferes with cell division: 20% of the cells of one particular slime-forming strain were found to have an abnormal morphology due to disturbances in cell division attributed to LTA.

Effect of limiting factors Even in the presence of an energy source, bacterial growth may be affected by some limiting factors, such as an unfavorable physical or chemical environment or the exhaustion of an essential nutrient. It has been suggested that the continuing catabolism of energy substrates can then provide an impetus for the production of polymers, such as the slime material, as either energy reserves or waste products.

Instability of slime production The variable rate at which lactococci produce slime is a well-known phenomenon, especially during incubation at a higher temperature, e.g., 30 °C. In addition, the serial transfer of a ropy strain, even at a lower temperature which normally favors the production of slime material, often leads to the loss of this trait. Sometimes, this unpredictability of slime formation causes difficulties in the commercial manufacture of ropy fermented milks.

Although the loss of the ropiness trait in starter strains has been observed in the commercial production of ropy milks, a similar problem has not been reported in the traditional manufacture of these products, in which a mixture of undefined organisms is used. A mixed population of lactococci and leuconostocs is likely to have effective mechanisms for retention in the culture of the plasmid(s) on which the ability to produce capsular material is encoded, especially if other properties important for the survival of the bacterial population are encoded on the same plasmid(s). The data reported in Table 1 lend support to this concept.

Bacteriophage

Attack by bacteriophages can cause problems in the manufacture of ropy fermented milks. Thirteen phages of different morphology have been found in an extensive study on 90 samples of viili from 20 dairies in Finland. The main morphological type was characterized by an isometric head with a long, noncontractile tail. The host organisms for these phages included strains of *L. lactis* sspp. *cremoris* and *Leuconostoc mesenteroides* ssp. *cremoris*. Some phages differed only in their host specificity.

A phage of nonencapsulated *L. lactis* subsp. *cremoris* isolated from viili was shown to dissolve the capsules of some strains, including a strain of *L. lactis* ssp. *lactis*. The capsules disappeared soon after infection, and maturing phages were seen inside the cells 2 h after infection, appearing as tightly packed bundles. The phage titer declined rapidly after heating for 15 min at 50–70 °C, but particles capable of plaque formation were detected even after heating for 5 min at 100 °C.

Products and Processes

In the domestic production of fermented milks, no heat treatment was applied to milk prior to fermentation, and the milk was inoculated with the residue remaining in the fermentation vessel from the previous batch. In this way, a characteristic flora dominated by lactococci became established in the vessel. Nowadays, in the commercial production of Nordic fermented milks, defined strains of starter microorganisms are used. The types of starter microorganisms have a distinct influence on the properties and shelf-life of these products.

Scandinavian fermented milks fall into six major classes, as shown in Table 3. Not all products in each row are identical to each other; the table is provided as a guide.

Table 3 The major classes of Scandinavian fermented milks[a,b]

Sweden	*Norway*	*Denmark*	*Finland*
Långfil	Tettemelk		Viili
Filmjölk	Kulturmelk	Tykmælk	Talouspiim
Lttfil	Skummet Kulturmelk		Rasvatonpiim
Krnmjölk	Kjernemælk	Kærnemælk	Kirnupiim
Grddfil	Rømme	Crème fraîche	Kermapiim
Lactofil		Ymer	Kokkeli

[a]Each row represents a separate class of products.
[b]There are also many special product names used by commercial manufacturers of fermented milks.

Långfil

Långfil or tätmjölk has a mild, sour taste and a ropy, dough-like consistency produced by encapsulated variants of lactococci. Occasionally, lactobacilli are also found in this product. Traditionally, the leaves of plants such as butterwort (*Pinguicula vulgaris*) and sundew (*Drosera* species, especially *Drosera rotundifolia*) were added to milk before leaving it to sour. They seem to exert a rennet-like action. No relationship has been established between the bacterial flora of butterwort or other plants and the bacterial composition of långfil.

Långfil cultures were preserved for later use by simple techniques, such as dipping a piece of cloth (in Sweden) or birch twigs (tettemelk, in Norway) in the finished product and then allowing them to dry. A new batch was started by immersing the twigs, or placing the dried cloth, in fresh milk.

Viili contains various lactose-fermenting yeasts and *Geotrichum candidum*, in addition to lactococci and leuconostocs. The cream layer is usually covered with the mold. Viili is made from unhomogenized milk and is eaten with a spoon.

Pitkäpiima, which is always ropy, is used as a drink. In domestic production, the cream layer, removed from the pitkäpiima after fermentation, was used to make butter.

Viilipiimä, from which the cream is not removed, also serves as a drink.

The following steps are included in the commercial production of viili:

- standardization of fat content at the minimum level of 3.9%
- heat treatment at 89–90 °C for 15–30 min or 92–96 °C for 4–5 min
- cooling to 18–20 °C
- addition of 4–8% starter culture
- incubation for 18–20 h to reach the pH of 4.6
- cooling to the storage temperature of 8–10 °C.

The final pH, after cooling, is 4.3–4.4 and the shelf-life is between 10 and 15 days.

Cultured Milk

Filmjölk, sometimes also known as kulturmjölk (cultured milk), is a popular Swedish fermented milk characterized by a typical flavor and aroma (derived primarily from diacetyl) and a fairly high viscosity, with a fat content of 3.0%. Filmjölk, which is used as a drink, was developed in the early 1930s. The starter culture contains the acid-producer *L. lactis* ssp. *lactis* and the flavor and aroma producers, *L. lactis* ssp. *lactis*, *L. lactis* ssp. *lactis* biovar *diacetylactis* and

Leuconostoc mesenteroides ssp. *cremoris*. All these strains are propagated together.

In the production of filmjölk:

- standardized milk is often deaerated at 78 °C to prevent or to alleviate defects such as syneresis of the coagulum, granulation, lumpiness, and low viscosity
- the deaerated milk is homogenized at 10–20 MPa and 70 °C, and then heated at 95 °C for 2–6 min
- the milk is cooled to 17–24 °C and inoculated with 1% starter culture
- after 17–24 h, coagulum forms at pH 4.6
- the product is cooled to 8 °C and packaged into retail containers.

Normally, a shelf-life of 10 days at 8 °C is achieved.

Filmjölk with 0.5% fat is called lättfil; the same product is known as skummet kulturmelk in Norway.

Tykmælk, a characteristic Danish fermented milk, is produced commercially in a similar manner to the Swedish filmjölk. It contains 3.5% fat.

Buttermilk

Traditional buttermilk (kärnmjölk, kjernemelk, kærnemælk, kirnupiimä) is made by churning cultured cream which has a relatively low fat content. In this manner, a small amount of cultured butter and a relatively high volume of buttermilk are obtained. A product known as cultured buttermilk is made by the fermentation of skim milk or low-fat milk.

Starter cultures contain, in addition to the bacteria used for filmjölk manufacture, *Lactococcus lactis* ssp. *cremoris*. Slime-producing strains are included in the starter culture for some types of cultured buttermilk made in Finland. The production process involves:

- heat treatment at 85–90 °C/20–30 min or 92–96 °C/4–5 min
- cooling to 20–23 °C
- incubation with 1–4% starter culture
- 15–20 h fermentation
- cooling to 5–10 °C, during which the pH decreases to 4.4–4.45.

The flavor and aroma of cultured buttermilk are very similar to those of filmjölk. The shelf-life varies from 7 to 12 days.

Cultured Creams

Gräddfil, rømme, crème fraîche and kermapiimä are cultured creams made by the fermentation of heat-treated cream with the same cultures that are used in the manufacture of cultured buttermilk. The legal requirements regarding the fat content of cultured creams and production practices vary between Scandinavian countries. For example, in Denmark, production involves:

- the cream being standardized at 9, 18, 38 or 50% fat, and then homogenized
- heat treatment at 90 °C for 5 min
- cooling to the incubation temperature of 20–27 °C
- inoculation with 2% starter culture
- 16–20 h incubation in the tank
- cooling to 5 °C, during which the pH of the final product falls to 4.4.

In two types of Swedish gräddfil, the legally required fat content is 12 and 34.5%; Norwegian rømme contains 20 or 35% fat; and in Finland there are three types of cultured cream, at >12, 35–40 and 42% fat.

Concentrated Fermented Milks

These include ymer, lactofil, kokkeli, and skyr.

Ymer is a Danish fermented milk, which contains at least 11% nonfat milk solids (including 5–6% protein) and 3.5% fat. It is usually produced from the ultrafiltered retentate, which contains about 15% total solids. Production involves:

- blending retentate with cream and homogenization
- heat treatment at 90°C for 5 min and cooling to 20–27 °C
- inoculation with 4% starter culture containing *L. lactis* sspp. *lactis* and *cremoris* and *Leuconostoc mesenteroides* ssp. *cremoris*
- fermentation until the pH reaches 4.5 (after 16–20 h)
- stirring, cooling to 5 °C, and storage for 24 h
- stirring again and packaging.

The product is often cooled in two stages, first to 14 °C and then to 5 °C.

In the traditional process, fermentation of the skim milk is followed by cutting of the coagulum and drainage of the whey at 40 °C. Alternatively, a quarg separator can be used to remove the whey. Pasteurized cream is then blended with the skim product, and this is followed by homogenization, cooling to 12–14 °C and packaging. The shelf-life of ymer is ≈20 days at 5 °C.

Lactofil, produced in Sweden, is similar to ymer. The fat content in lactofil is 5% and the starter culture is similar to those used in the manufacture of filmjölk and karnmjölk.

Kokkeli, traditionally made in eastern Finland, is prepared by warming spontaneously soured milk in an oven and removing the separated whey.

Skyr is manufactured in Iceland from skim milk. Unlike all other traditional Scandinavian fermented

milks, it is fermented by a thermophilic microflora. The milk is then heated to facilitate the syneresis of the protein coagulum. Whey separation is achieved by either straining through linen bags or using quarg separators. Ultrafiltration is now also used to recover whey proteins, which are added to the skyr before packaging. Microorganisms isolated from skyr include thermophilic lactic acid bacteria (*Streptococcus thermophilus, Lactobacillus delbrueckii* ssp. *bulgaricus* and *L. helveticus*) and lactose-fermenting yeasts. Typical skyr contains 17.5% total solids, including >13% protein and 0.4% fat.

Product Composition and Shelf-life

Composition

The total solids and fat content of Nordic fermented milks is prescribed by law. Often, a minimum number of viable microorganisms is prescribed by the food standards.

Starter activity in milk leads to the production of lactic acid, as well as flavor and aroma compounds. For example, *Lactococcus lactis* ssp. *lactis* metabolizes the citrate present in milk to diacetyl, acetoin, 2,3-butylene glycol, and CO_2. Citrate is also metabolized by leuconostocs, mainly to diacetyl and acetoin (at pH $\leq$ 5.5), and acetate (at pH > 5.5).

Differences in vitamin content have been observed between fermented milks and raw milk. For example, viili and cultured buttermilk (containing 1.9% fat) were found to contain more folate than raw milk, by 48 and 65%, respectively. However, their vitamin C content was lower than that of milk, by 58 and 24%, respectively, and their riboflavin was lower, by 12 and 11%, respectively. The concentration of vitamin B_{12} in both products was also lower than in milk, by > 20 %.

Shelf-life

Very high numbers of viable bacteria have been observed in some spontaneously fermented milks. This may be due to the reduced rate of cell autolysis (discussed above).

The shelf-life of traditional ropy fermented milks is exceptionally long, due mainly to the presence of glycocalyx, which improves the rheological behavior of these products.

Features such as the cream layer and the presence of the mold *Geotrichum candidum* on the surface of some products provide added protection from spoilage organisms, thus further extending the shelf-life.

Commercially made products are also noted for their long shelf-life. Even if some preexisting defect reduces the water-holding capacity of the protein matrix, e.g., if proteolysis were allowed to proceed unhindered in raw milk, the presence of glycocalyx would mask the syneresis of the protein coagulum.

Health-related Effects

Considerable progress has been made in demonstrating certain beneficial effects of Nordic fermented milks in experimental animals. However, unequivocal experimental and/or epidemiological evidence still needs to be gathered to substantiate claims of similar effects in humans.

Effect on Immunity

It has been suggested that fermented milks may play an immunomodulating role, for example they stimulate the functions of gut-associated lymphoreticular tissue (GALT), and this effect has been attributed mainly to antigenic structures of the surface of lactococci. In particular, *L. lactis* ssp. *cremoris* isolated from viili has been shown, in studies with human lymphocyte cultures, to stimulate the secretion of immunoglobulins, primarily those of the immunoglobulin M class. In addition, T lymphocytes showed considerable proliferation in response to the same strain.

A substance active as a mitogen of murine B cells was purified from the slime produced by another strain of *L. lactis* ssp. *cremoris*, also isolated from viili. The mitogenic compound was a phosphopolysaccharide, containing rhamnose, glucose, galactose, and phosphorus. Its activity was higher than that of unpurified slime.

A significant induction of the cytotoxicity of peritoneal murine macrophages against sarcoma cells, by *L. lactis* subsp. *cremoris* isolated from viili, has been demonstrated *in vivo*. A single intraperitoneal injection of the freeze-dried cells retarded the growth of ascitic and solid sarcomas in mice. However, the same preparation showed no direct cytotoxic activity against the same sarcoma cells *in vitro*, which suggests that the antitumor effect of this organism is through the enhancement of the cytotoxicity of the macrophages of the host. The exact mechanism of this enhancement is not known, but the slime has been observed to increase glucose consumption *in vitro* by intraperitoneal macrophages.

Freeze-dried preparations of viili, långfil, and ropy yogurt, used as intraperitoneal injections daily for 9 days after the tumor inoculation, also significantly retarded the growth of murine solid sarcomas *in vivo*. The maximum antitumor effect was induced by a dose of 10 mg kg^{-1} of the viili preparation, 50 mg kg^{-1} of the långfil preparation and 100 mg

kg^{-1} of the ropy yogurt preparation. Thus, the effect of both viili and långfil was clearly stronger than that of ropy yogurt. All three preparations significantly enhanced the delayed cutaneous hypersensitivity response to oxazolone, which was depressed in tumor-bearing mice. The antitumor effect of these ropy milks is thought to be mediated by the immune responses associated with host's macrophages and/or T cells.

Antimutagenic Activity

Strains of *L. lactis* ssp. *cremoris* isolated from viili, both ropy and nonropy, reduced the mutagenicity of nitrosated beef extract by 40%, as determined by the Ames test, using *Salmonella typhimurium* as the test organism.

Lowering of Serum Cholesterol

In experiments on rats receiving diets containing viili, nonropy fermented milk, or acidified skim milk, the serum cholesterol level of rats on a viili-containing diet was lowest of the three groups. Their ratio of high-density lipoprotein cholesterol to total cholesterol was the highest of the three groups. The mechanism of this cholesterol-lowering effect is unknown.

Antibacterial Effects

Antagonistic effects of lactic acid bacteria against common pathogens and spoilage bacteria have been well demonstrated *in vitro*. For example, lactococci and their capsular material have been shown to inhibit the growth of *Staphylococcus aureus, Escherichia coli*, and some clostridia.

Future Trends

The future position of Nordic fermented milks in their traditional markets seems assured, despite the popularity that yogurt and probiotic products have gained in the past few decades.

Many types of fermented milks are produced nowadays in Scandinavia and some of them do not clearly fit into the categories outlined above, for example, products which are mixtures of ropy milks with acidophilus milk. With the growing interest in various probiotic cultures, this trend is expected to continue, with more combinations of this kind being offered to consumers.

See also: **Cream**: Types of Cream; **Fermented Milks**: Types of Fermented Milks; **Lactic Acid Bacteria**; **Mycotoxins**: Classifications; **Probiotics**; **Starter Cultures**; **Yeasts**

Further Reading

Alm L and Larsson I (1983) Från forntid till framtid. Den nordiska tätmjölken – en produkt med gamla anor. *Nordisk Mejeriindustri* 10: 396–399.

Bertelsen E (1983) Kulturmjölksprodukter i Norden. *Nordisk Mejeriindustri* 10: 386–390.

Forsén R (1966) Die Langmilch (Pitkäpiimä). *Meijeritieteellinen Aikakauskirja* 26: 1–76.

Forsén R (1989) Characterisation of antigenic surface structures of lactococci and their possible immunobiological effects. In: *Fermented Milks and Health*, workshop proceedings, Arnhem, September 1989. Ede, The Netherlands: NIZO.

Forsén R, Niskasaari K, Tasanen L and Nurmiaho-Lassila E-L (1989) Studies on slimy lactic acid fermentation: detection of lipoteichoic acid containing membrane antigens of *Lactococcus lactis* ssp. *cremoris* strains by crossed immunoelectrophoresis. *Netherlands Milk and Dairy Journal* 43: 383–393.

Gudmundsson B (1987) Skyr. *Scandinavian Dairy Industry* 4: 240–242.

IDF (1988) Fermented milks – science and technology. *International Dairy Federation Bulletin* no. 227, 164 pp.

Kitazawa H, Toba T, Itoh T, Kumano N and Adachi S (1990) Antitumor activity of ropy sour milks in murine solid tumor. *Japanese Journal of Zootechnical Science* 61: 1033–1039.

Kitazawa H, Itoh T and Yamaguchi T (1991) Induction of macrophage cytotoxicity by slime products produced by encapsulated *Lactococcus lactis* ssp. *cremoris*. *Animal Science and Technology* 62: 861–866.

Kitazawa H, Toba T, Itoh T, Kumano N, Adachi ST and Yamaguchi T (1991) Antitumoral activity of slime-forming, encapsulated *Lactococcus lactis* ssp. *cremoris* isolated from Scandinavian ropy sour milk, "viili". *Animal Science and Technology* 62: 277–283.

Kitazawa H, Yamaguchi T, Miura M, Saito T and Itoh T (1993) B-cell mitogen produced by slime-forming, encapsulated *Lactococcus lactis* ssp. *cremoris* isolated from ropy sour milk, viili. *Journal of Dairy Science* 76: 1514–1519.

Kontusaari S and Forsén R (1989) Finnish fermented milk "viili": involvement of two cell surface proteins in production of slime by *Streptococus lactis* ssp. *cremoris*. *Journal of Dairy Science* 71: 3197–3202.

Laukkanen M, Antila P and Antila V (1988) The water-soluble vitamin contents of Finnish liquid milk products. *Meijeritieteellinen Aikakauskirja* 46: 7–24.

Macura D and Townsley PM (1984) Scandinavian ropy milk – identification and characterization of endogenous ropy lactic streptococci and their extracellular excretion. *Journal of Dairy Science* 67: 735–744.

Nakajima H (1995) Characteristics of fermented milk produced by slime-forming *Lactococcus lactis* subsp. *cremoris*. *Snow Brand R&D Reports* 104: 97–169.

Nakajima H, Toyoda S, Toba T *et al*. (1990) A novel phosphopolysaccharide from slime-forming *Lactococcus lactis* subspecies *cremoris* SBT 0495. *Journal of Dairy Science* 73: 1472–1477.

Neve H, Geis A and Teuber M (1988) Plasmid-encoded functions of ropy lactic acid streptococcal strains from Scandinavian fermented milk. *Biochimie* 70: 437–442.

Robinson RK and Tamime AY (1990) Microbiology of fermented milks. In: Robinson RK (ed.) *Dairy Microbiology*, vol. 2: *The Microbiology of Milk Products*, 2nd edition, pp. 291–343. London: Elsevier Applied Science.

Saxelin M-L, Nurmiaho-Lassila E-L, Merilainen VT and Forsén R (1986) Ultrastructure and host specificity of bacteriophages of *Streptococcus cremoris, Streptococcus lactis* subsp. *diacetylactis*, and *Leuconostoc cremoris* from Finnish fermented milk "viili". *Applied and Environmental Microbiology* 52: 771–777.

Sundman V (1953) On the microbiology of Finnish ropy sour milk. *Proceedings of the Thirteenth International Dairy Congress*, vol. 3, pp. 1420–1427. The Hague, The Netherlands: Mouton & Co.

Other Relevant Products

B McKevith and C Shortt, Yakult UK Ltd, Acton, London, UK

Introduction

Fermented milks and other related products are available in many different countries worldwide. Some have a long history of consumption while others have been developed more recently due to scientific advances. The microorganisms used to make these products are varied but often feature lactic acid bacteria (LAB) which during fermentation convert lactose into lactic acid. Many of these products are generally good sources of a number of important nutrients such as protein, calcium, and zinc and some also have other beneficial effects (particularly on the gastrointestinal tract (GIT)). Indeed, in some parts of the world these products have a long history of therapeutic use. Due to increased medical interest early in the twentieth century, many products were developed commercially. Currently one of the major areas for growth is the development of fermented products containing specific probiotic bacteria.

History

The true origin of fermented milks is difficult to establish but it is suggested that they originated in western Asia and were carried to the east where new variants were developed to suit the different climate. Recipes for soured milks containing fruits and spices existed in ancient Greek and Roman times and fermented milks have played a role in many different cultures throughout history, e.g., they are mentioned in the Bible and the sacred books of Hinduism. Storable dairy products, presumably cultured, are also mentioned in the archaic texts from Mesopotamia dating back to before 3000 BC. In the preindustrial world, milk and/or dairy products of some description had a place in the diet of most people who had domesticated animals. A combination of chance contamination, climatic conditions, and serendipity probably played a part in the development of traditional fermented dairy products. Fermentation of milk allowed humans to store and use excess milk and to travel long distances. Today, fermentation is used to produce dairy foods with different flavors, textures and health benefits.

Several different factors influence the preparation of traditional fermented milk products, with agricultural factors heavily affecting the process (**Table 1**). Cultured dairy products can be prepared from the milk of different mammals, with cows', buffalos', goats' and sheep's milk being the most common. In some parts of the world the milk of other animals (such as camel, yak, and reindeer) is used.

Yogurt and fermented milk of some description are probably the most popular fermented dairy products. Variations within countries and even within regions are seen due, in part, to different ethnic groups with distinct cultures and preferences. Some of the traditional fermented milks and related products produced in different parts of the world are outlined in **Table 2**.

Table 1 Factors influencing production of traditional fermented milk products

Factor	*Example*
Agricultural	Current farm practices
	Animal husbandry
	Milk availability
	Species of animals used for milk production
	Species of animals used for other agricultural purposes
	Breed of species for milk production
	Consumption and suitability of milk
Environmental	Weather patterns
	Climate
Cultural	Eating habits of population
	Stage of development of commercial milk processing
	Consumption and suitability of milk

Adapted from FAO (1990) *Animal Production and Health Paper. The Technology of Traditional Milk Products in Developing Countries*. Rome: FAO/WHO.

Table 2 Examples of traditional fermented milk products and other related products

Regions	*Products*
Eastern Europe	
Bulgaria	Kisle mliako, Bulgarian milk, brano mliako, boza
Czech Republic	Kefir, Acidophilus sour milk, biokys
Hungary	Sour cream, yogurt (tarho)
Poland	Kefir, buttermilk, milk ''champagne,'' polkrem (soured cream)
Romania	Oxygala
Middle and Near East	
Greece	Tiaourti
Israel	Leben, zivda
Iran	Dough, Kashk
Iraq	Leben, kushuk
Lebanon	Laban
Turkey	Ayran, torba
Syria	Lebnye, lebenen bezet
Africa	
Algeria	Aoules, tammart
Egypt	Bouza, laban zabady, laban rayeb, laban khad
Kenya	Maziwa iala, busaa
Nigeria	Nono
South Africa	Magou, maas
Sudan	Goubasha
Russia	Kefir, koumiss, prostokvasha, rjazhenka
Scandinavia	
Denmark	Tymaelk ymer
Finland	Piima, kurz, villia
Iceland	Skyr
Norway	Taette (cellar milk), seterroømme
Sweden	Ropy milk
Asia	
China	Targ, airak
India	Dahi, lassi
Mongolia	Koumiss, tarag

Adapted from IDF (1988) Fermented milks and science and technology. *Bulletin of the International Dairy Federation* 227: 4–164 and Kurman JA, Rašić JLJ and Kroger M (1992) *Encyclopedia of Fermented Fresh Milk Products*. New York: Van Nostrand Reinhold.

Fermented Milks and Other Related Products – Definitions

Generally, fermented or cultured dairy products are made from milk (or its components) and modified by incubation with cultures of microorganisms. The 1997 International Dairy Federation (IDF) international draft standard described fermented milks as milk products fermented by the action of specific microorganisms and resulting in a reduction of pH and coagulation. The microorganisms must 'be viable, active and abundant (at least 10^7 CFU g^{-1}) in the product to the date of minimum durability.' The standard further controls the type of microorganisms to be used as being nonpathogenic and nontoxic to produce either a pure fermentation (a lactic acid fermentation) or a mixed fermentation (containing also an alcoholic fermentation). Specific cultures are required to be used for the production of certain defined products, such as cultured buttermilk, yogurt, and Acidophilus milk. There is some disparity among different countries regarding the legal definition of some fermented milk products. Some countries have prescriptive legislation, in line with the above IDF standard, while many others do not.

Some definitions for fermented milks and/or related products imply or state (in legislation or industry standards) that live microorganisms should be present in the final product. This is true in countries such as Japan, the Netherlands, Denmark, and Austria, which stipulate that there must be a high proportion of viable microorganisms in the product (often $\geq 10^7$ ml^{-1}). In contrast, a few countries such as the UK, Germany, and Sweden also recognize products that are heat-treated after fermentation as long as the appropriate consumer labeling is used.

Currently changes are being proposed by the Codex Committee on Milk and Milk Products to the standard for fermented milks. The forthcoming standard will cover all fermented milks (including yogurt and composite products), give descriptions of the different products, essential composition and quality factors, food additives, contaminants, hygiene, and labeling. Issues currently under discussion include at what point the viable organisms should be controlled (e.g., at point of sale to the consumer, when the product leaves the manufacturer, or at the date of minimum durability); the definition, composition, and minimum acidity for mild yogurt; the definition of composite fermented milk products, and the labeling of heat-treated products. The standard still has several hurdles to overcome before it is adopted.

Below are some descriptions for fermented products that are currently used.

Yogurts

Yogurt is traditionally produced from milk and/or cream through the action of two particular strains of LAB – *Lactobacillus bulgaricus* and *Streptococcus thermophilus*. Codex currently defines yogurt as a milk product obtained by fermentation using these two specific strains with or without optional additions (such as milk powder, other LAB, and sugar). In many countries (e.g., Sweden, France, Belgium, Portugal, Spain, USA) legislation exists allowing only *L. bulgaricus* and *S. thermophilus* to be used to produce yogurt. In the UK yogurt can be made using both of these strains or just one of them. In other countries (e.g., Switzerland) strains such as *L. acidophilus* and bifidobacteria may be used in addition to the traditional yogurt strains. It is likely that the

new Codex standard will build on the draft definition currently used for mild yogurt (a product made from *S. thermophilus* and lactobacilli other than *L. bulgaricus*). In Japan no separate legislation exists for yogurt and so it is not obligatory that the two traditional strains are used, either collectively or separately. Similarly, in Finland there are also no compositional regulations for any milk products.

Soft Cheeses

Cottage cheese is a high-moisture, unripened soft cheese made from skimmed milk. Most of the microorganisms used in its manufacture (including species of *Lactococcus lactis*) are destroyed during the process, which involves scalding the curds. After washing, the curds are then mixed with a creamy dressing, which has usually been ripened with a flavor-producing culture. Cream cheese is a soft cheese that is made from cream. A starter culture of *S. lactis* and two species of *L. lactis* is used to produce a rich, full, and mildly acidic flavor and a soft texture. Cream cheese can be made from cream of various fat contents. Crème fraîche is a lightly fermented cream with a rich nutty flavor. Cultured or sour cream is also prepared from cream. After pasteurization and homogenization, it is incubated with a mesophilic starter to produce a thick tangy product. Fromage frais is made from fermented milk and has a mild flavor. It is thicker and more concentrated than yogurt as some water is removed after fermentation. Quark or quarg is a soft, unripened cheese made in a similar way to cottage cheese. Its production often involves heating after coagulation by the starter culture.

Probiotic-Type Products

Probiotic products contain live microbial food ingredients that are beneficial to health. Bacteria of human origin are often used in probiotic products intended for humans. Species of lactobacilli and bifidobacteria are most common but bacteria of nonhuman origin and yeasts such as *Saccharomyces boulardii* are also used. In some countries *Streptococcus thermophilus* and *L. bulgaricus* are considered probiotic strains. Fermented milk drinks and yogurts are popular forms for probiotic foods, although nondairy fermented products also exist, such as fruit juice drinks.

Other Fermented Drinks

The proposed Codex standard will characterize certain fermented milks, including Acidophilus milk, kefir, and kumys, by their specific microorganism(s). However there are a number of other products which are described elsewhere. Acidophilus milk is a product made using the probiotic bacteria *L. acidophilus*, a bacterium from the flora of the normal GIT. It and related products (such as Acidophilus yogurts) are popular in the USA.

Cultured buttermilk was originally made using buttermilk (the byproduct of churning fermented cream into butter) but it has an unpredictable quality and short shelf-life since the whey often separates out and off-flavors develop. Modern equivalents are made from skim or low-fat milk and fermented with acid and flavor starters such as *S. lactis* and/or *S. cremoris* together with *S. paracitrovorous*. Buttermilk is often used as an ingredient in the baking industry.

Laban is a cultured milk drink, usually based on cows' milk and consumed in many Gulf countries. Similar products exist in Egypt, Tunisia, and Morocco. Homemade traditional laban contains a mixture of LAB, molds, and yeasts. Three strains of LAB are used in its commercial preparation. Lassi is a yogurt drink from East India that may be sweetened and flavored with rosewater or fruit. Kefir originated in Russia and is now also popular in Poland, Germany, and Sweden. This traditional product is usually made using cows' or goats' milk and kefir grains which contain a complex mixture of microorganisms including bacteria (such as streptococci and lactobacilli), yeasts, and a polysaccharide of microbial origin. Koumiss (kumiss, kumys, kumyss) also originated in Russia but is based on mares' milk. *L. bulgaricus* and lactose-fermenting yeasts are used in its production. Ropy milks, popular in Scandinavia, use starters that produce slime (mucus), giving a unique viscous quality to the final product.

Frozen Products

There are a few frozen products, like tsutsugi from Mongolia, made from naturally fermented cream, and a number of newer products like Acidophilus icecream in the USA, which contain probiotic strains. Although viability may decrease slightly during freezing, it is still possible to achieve levels of 10^6–10^7 CFU ml^{-1} in frozen products. Frozen yogurt has become popular since the 1970s, especially in the USA. It is made by freezing regular stirred yogurt after the addition of stabilizers, sweeteners, and fruit and flavorings.

Other Related Products

Some fermented products are made using nondairy ingredients such as soya milk, coconut milk, grains, and rice as a base, e.g., boza, a nonalcoholic drink made from cereal (such as wheat, rye, maize, or millet) and widely consumed in Bulgaria and the Balkan peninsula. Acidified milk products (such as sour cream) can be made by direct acidification using organic acids. Nonfermented products also exist,

such as nonfermented Acidophilus milk, available in the USA where *L. acidophilus* is added to freshly pasteurized milk and the product is then kept refrigerated to prevent bacterial growth. In this way the milk acts as a carrier for the microorganism but it does not have the sour taste usually associated with fermented dairy products.

Microbiology

Originally the microorganisms involved in fermentation were used to preserve food of animal and plant origins. As other methods are now available (e.g., refrigeration and pasteurization), microorganisms are used for other reasons – to change the flavor of the raw material (possibly to produce a more acceptable product), to improve the nutritional value of the raw material, and to produce a food which offers beneficial health effects (e.g., probiotic products).

The microorganisms used in the starter culture determine the characteristics of the final product. More specifically, the product's quality and characteristics are dependent on the way the chosen microorganisms utilize the nutrients available in the medium.

Historically, fermented dairy products were made using starter cultures containing a heterogeneous mixture of LAB and yeasts, using one batch to inoculate a subsequent one. This wild fermentation can lead to products of variable quality and taste, due to contamination and the genetic instability of some microorganisms. Although this traditional method is still used in many parts of the world (especially for home production), starter cultures containing known microorganisms have been available commercially since the beginning of the twentieth century. The starter cultures act in a known way, giving greater control over the end product in terms of safety, quality, and taste.

Specific starter cultures are often used in geographical regions, for example, ropy, sour milks containing only streptococci or streptococci and lactobacilli predominate in northern Europe. The type of starter culture can be used to classify fermented milk products in the following manner (Table 3).

LAB are Gram-positive nonsporing cocci, coccobacilli, or rods. Lactobacilli are generally recognized as safe because of their extensive use over the centuries and more recently their role in food technology, with negligible problems. LAB involved in the production of fermented milks and related products are distributed over a large number of bacterial species and include species of *Lactobacillus*, *Streptococcus*, and *Bifidobacteria* genera. Lactobacilli have greater difficulty growing in milk that is neutral or slightly alkaline compared to streptococci. In a mixed culture of these two species, the streptococci may establish growth in neutral or weakly alkaline milk, and from their acid production make the milk suitable for subsequent lactobacilli growth.

Table 3 Classification of fermented milk products by starter culture

Starter culture used	*Example of fermented dairy product*
Thermophilic	
Classic cultures, e.g., *Lactobacillus bulgaricus* and *Streptococcus thermophilus*	Yogurt Acid buttermilk
Mesophilic	
Some species of streptococci, e.g., *Streptococcus lactis*, lactococci and some *Leuconostoc* species	Conventional buttermilk Cream cheese Crème fraîche Cottage cheese Fromage frais Sour cream Ropy milks – Scandinavia
Lactic acid bacteria/yeast	
Mixture of lactic acid bacteria and yeasts, e.g., *Saccharomyces lactis*	Kefir Koumiss
Probiotic	
Some species of lactobacilli, e.g., *Lactobacillus acidophilus, L. casei*, and bifidobacteria, e.g., *Bifidobacteria bifidum*	Bioyogurts Acidophilus milk Yakult

Adapted from National Dairy Council (1997) *Topical Update 8. Nutritional Benefits of Yogurt and Other Fermented Milk Products* and IDF (1988) Fermented milks and science and technology. *Bulletin of IDF* 227: 4–164.

LAB use the lactose in milk as an energy source. During fermentation lactose is cleaved by bacterial enzymes to produce galactose and glucose. The glucose formed is then further fermented to produce pyruvic acid, which is metabolized to lactic acid. There are no traditional products made solely using yeast but some products such as kefir and koumiss use yeasts as well as LAB in their manufacture. The yeasts produce CO_2 via an alcoholic fermentation. This CO_2 may be utilized by the other microorganisms and therefore is not usually present in the final product.

Many products have been developed over the last 100 years due to advancements in microbiology, increased understanding of the fermentation process, and increased knowledge of the impact of nutrition on health. Unlike products which often rely on spontaneous fermentation, newer products use selected strains of bacteria for specific qualities. Some commercialized products may be copies of traditional products, for example, using different microorganisms to produce a product similar to a traditional one.

Probiotics

Most microorganisms in fermented milks and related products do not survive transit through the GIT. The body has a number of different mechanisms (e.g., stomach acid, bile, and mucous secretions of the GIT) to decrease or prevent the movement of microorganisms through the GIT. Only robust strains (usually of human origin) such as those termed probiotics can survive and reach the large intestine in a viable state. Here they can modulate the gut flora, leading to specific health benefits.

The word probiotic is derived from the Greek, meaning for life. The concept was used by von Ferdinand in 1954 to describe gut flora modulation using lactobacilli and bifidobacteria. The definition has evolved and recently probiotics for use in human nutrition were redefined as 'live microbial food ingredients that are beneficial to health.' As such a probiotic may consist of a single strain or mixture of microorganisms, its activity is not limited to the GIT but may affect microbial communities at other sites, such as the urogenitial and respiratory tracts; it may have nonmicroflora-mediated effects, e.g., effects on immune parameters.

A number of microorganisms are used in probiotic products, including strains of lactobacilli (e.g., *L. acidophilus, L. casei*, and *L. plantarum*), bifidobacteria (*B. breve, B. longum*, and *B. bifidum*), other LAB (e.g., *Enterococcus faecium*) and non-LAB species such as strains of *Escherichia coli*.

Although viability has been stressed as an important criterion for probiotic microorganisms, recent studies have shown that nonviable probiotic organisms and their cellular components may be efficacious in certain situations. Research in this area is ongoing but, from a commercial perspective, viability of the microorganisms in probiotic products continues to be a key issue.

Nutritional Aspects

Fermented milks and related products reflect the nutritional composition of the milk or base ingredient from which they are made. Generally those made from milk are good sources of a wide variety of nutrients but their nutritional composition often varies from milk due to changes during fermentation (Table 4).

Cultured products are more digestible compared to unfermented ones. In some communities fermented products are used as weaning foods for this reason. During fermentation acid is produced slowly, allowing small particles to form and increasing the surface area for later human enzyme digestion.

Table 4 Effect of lactic acid bacteria on composition of fermented milk

Before fermentation	*After fermentation*	
	Decreased	*Increased*
Lactose	Lactose	Organic acids Lactic acid Galactose, glucose Polysaccharides
Protein	Protein	Peptides Free amino acids
Fat	Fat	Volatile and long-chain free fatty acids
Vitamins	Vitamins	Vitamins (e.g., folate)

Adapted from Nakazawa Y and Hosono A (1992) *Functions of Fermented Milk Challenges for the Health Sciences*. London: Elsevier.

Fermented milk products have lower lactose levels than milk, normally ranging from 30 to 50%. This is because lactose is hydrolyzed by microbial beta-galactosidase during fermentation to produce galactose and glucose, making fermented products useful for the substantial number of people with a lactase deficiency (absence or decreased production of the enzyme needed to break down lactose). Also live bacteria from fermented products that survive the gastric acid release their microbial beta-galactosidase into the small intestine, aiding lactose digestion.

A larger proportion of free amino acids and peptides will be present in fermented milk products due to microbial digestion, although the overall amino acid concentration will be similar to the base ingredients. This may not be of major importance, as most people have no difficulty digesting protein.

The vitamin content of the final product will depend on the base ingredients used, the starter cultures, and the processing conditions (especially heat treatment). LAB can synthesize vitamins such as folate *in vitro* and *in vivo*. Therefore, there may be increased vitamin levels in some fermented products. It is doubtful, however, whether these vitamins would affect the vitamin status of healthy adults as their bioavailability in the GIT is currently unknown.

The fermentation process does not generally affect the mineral content of the product, although there is some suggestion that bioavailability may be enhanced. Like dairy milk, fermented dairy milk products are excellent sources of a range of minerals such as calcium, phosphorus, and zinc.

Traditional fermented milks and related products can make a considerable contribution to the nutritional status of some groups, particularly in developing countries, where they provide a good source of nutrients (particularly protein) at a low cost.

Other Benefits

Fermented milks and other related products may have other benefits above and beyond their nutritional content. Traditional products (such as laben from Syria) have been used therapeutically for decades for disorders of the stomach, intestine, and liver, and for stimulating the appetite long before the existence of bacteria and other microorganisms was recognized. Both koumiss and kefir are used therapeutically in Russia.

The glucose liberated from the lactose hydrolysis during fermentation is converted into lactic acid, which contributes to the taste of the products. It also helps preserve fermented foods by acidifying the product and preventing or decreasing the growth of pathogenic organisms. Certain probiotics may have a similar effect within the human GIT either by changing the environment (decreasing the pH) or by colonization resistance (preventing other microorganisms from adhering to the GIT), thereby decreasing the incidence and/or duration of intestinal infection. It has also been established that some probiotic bacteria are capable of producing bacteriocins – substances that are toxic to other (usually related) strains of bacteria. Other probiotic components may also have beneficial effects on unwanted microorganisms, for example a supernatant of one probiotic has recently been shown to downregulate *Helicobacter pylori* infection in humans.

A number of health effects have been established for probiotics, including reduction of duration of rotavirus diarrhea, alleviation of the symptoms of lactose intolerance, reduction of harmful intestinal microbial enzyme activities, immune enhancement, and decreased fecal mutagenicity. However, it cannot be assumed that all probiotic strains of even a specific species have the same or even any desirable properties.

In some countries where yogurt has not been part of the normal diet, the consumption of fermented dairy products is generally low. However, their popularity has increased in recent years (**Figures 1** and **2**).

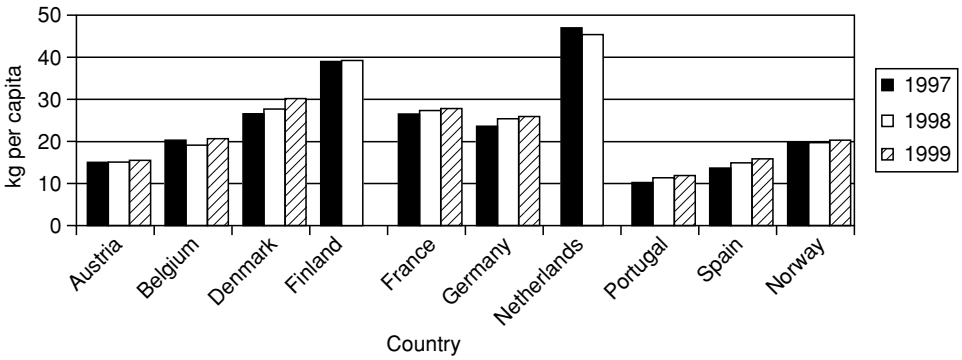

Figure 1 Consumption of milk drinks and fermented products, including yogurt: part 1. Aadapted from IDF (2000) The world dairy situation 2000. *Bulletin of the International Dairy Federation* 335: 8.

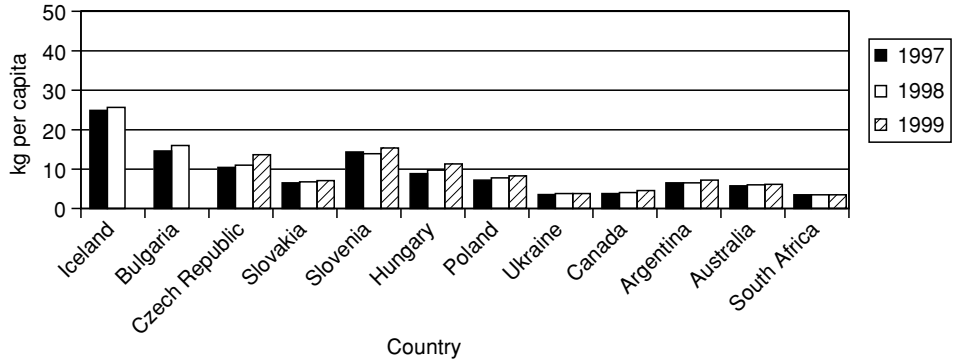

Figure 2 Consumption of milk drinks and fermented products, including yogurt: part 2. Adapted from IDF (2000) The world dairy situation 2000. *Bulletin of the International Dairy Federation* 335: 8.

This may be due to the wide variety of products now available, their simple production technology, their low cost, their long shelf-life, and their healthy image.

The growth area for fermented milks and related products appears to be in the added-value and health sector of the market – probiotic, functional, and fortified products. A number of factors will influence the acceptance and further growth of the market, including the introduction of more specialty products, quality and sensory improvements, development of fermented products using plant material, genetically engineered strains, and continued or increased interest in 'healthy' foods.

See also: ***Acidophilus* Milk**; **Cheeses**: Types of Cheese; **Fermented Foods**: Origins and Applications; **Fermented Milks**: Types of Fermented Milks; **Lactic Acid Bacteria**; **Lactose**; **Microflora of the Intestine**: Role and Effects; **Yogurt**: The Product and its Manufacture; Yogurt-based Products; Dietary Importance

Further Reading

Codex Alimentarius Commission (1975) *Codex Standards for Milk and Milk Products*, vol. 12. Codex Stana 11(a). Rome: FAO/WHO.

Codex Alimentarius Commission (2001) *Report of the Fourth Session of the Codex Committee on Milk and Milk Products*. Rome: FAO/WHO.

Dairy Industry Federation (2000) *Yoghurt Code of Practice for the Composition and Labelling of Yoghurt*. London: Dairy Industry Federation.

FAO (1990) *Animal Production and Health Paper. The Technology of Traditional Milk Products in Developing Countries*. Rome: FAO/WHO.

IDF (1988) Fermented milks and science and technology. *Bulletin of the International Dairy Federation* 227: 4–164.

IDF (1997) Standards for fermented milks. IDF-Doc 316.

IDF (2000) The world dairy situation 2000. *Bulletin of the International Dairy Federation* 335: 8.

Kiple KF and Conee Ornelas K (eds) (2000) *The Cambridge World History of Food*, vol. 1. *Dietary Liquids, Milk and Dairy Products*, pp. 692–702.

Kurman JA, Rașić JLJ and Kroger M (1992) *Encyclopedia of Fermented Fresh Milk Products*. New York: Van Nostrand Reinhold.

Nakazawa Y and Hosono A (1992) *Functions of Fermented Milk Challenges for the Health Sciences*. London: Elsevier Science.

National Dairy Council (1997) *Topical Update 8. Nutritional Benefits of Yogurt and Other Fermented Milk Products*. London: National Dairy Council.

O'Brien J (1999) Sugar profiles of cultured dairy products in the UK. *Journal of Human Nutrition and Dietetics* 12: 242–250.

Dietary Importance

C D Khedkar, College of Dairy Technology, Warud, (Pusad), Maharashtra, India
G D Khedkar, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India
N V Chavan, Adarsha College, Hingoli, Maharashtra, India
S D Kalyankar, Marathwada Agriculture University, Parbhani, Maharashtra, India

Background

During the past 80 years, considerable attention has been directed sporadically on benefits derived from the consumption of milk products containing lactic acid bacteria (LAB). The role of fermented milks (FM) in human nutrition is well documented. Man knew the virtues of these fermented foods, even during the early days of civilization. In earlier days, these foods were produced by natural fermentations with the main objective of preserving milk. In Europe, Asia, and Africa, sour milk was known to be more stable than fresh milk. It preserved the high-quality nutrients present in milk in a relatively more stable form. From prehistoric times, Man learned to use milk as food, and, although the origin of the art of preserving dairy products by lactic acid fermentation is lost in antiquity, biochemical and microbiological knowledge of fermentation is comparatively recent. On the Indian subcontinent, the conversion of milk in every household by souring with the left-overs of the previous day's sour milk has been a common practice ever since the Aryans inhabited the land. In this way, the life and utility of milk nutrients were extended. Today, the practice of preserving milk by fermentation has become a routine household technology, and these products are almost compulsory on this subcontinent. As far as the Western countries are concerned, FMs like kefir, koumiss, leben, skyr, crowdiss, and yogurt produced by lactic fermentation of milk have been consumed by people either as refreshing beverages or as nutritious or therapeutic foods and dietary adjuncts.

The milk of several species, including the cow, buffalo, goat, sheep, mare, camel, and reindeer, has been used for the preparation of FMs. Each type of FM has its own characteristic texture, body, flavor, and composition, depending on the type of milk used, the type of starter cultures (LAB with or without aroma producers and yeasts), and the procedures followed in its preparation. The FMs popularly used in various countries may be classified broadly into four categories: (1) moderately sour types with a

pleasant aroma associated with diacetyl, e.g., cultured milk; (2) sour and very sour types owing to high acid production, e.g., yogurt and dahi; (3) ethanol in addition to lactic acid, e.g., koumiss and kefir; and (4) probiotic fermented milks, e.g., acidophilus milks especially prepared with human strains of *Lactobacillus acidophilus*.

Fermentation with LAB is one of the oldest methods of food processing and food preservation used by mankind. In many areas of the world, the fermentation process is initiated by adventitious indigenous microorganisms or through 'backslopping' with mixed or unknown cultures. Often, this type of fermentation may be slow or unpredictable. Nevertheless, it affords a vital means of preserving highly perishable foods, especially where refrigeration is lacking. In developed countries, fermentation is initiated with pure starter cultures that have predictable performance potentials. Recent scientific and technological advances in starter-culture management and process control have yielded a variety of products with superior chemical, physical, nutritional, and sanitary qualities. A number of FMs, cereals, and beverages are prepared through the action of lactobacilli. Nearly every civilization has consumed FMs. Over the past two decades, the per-capita consumption of FMs has increased more than 35-fold owing to their image as nutritious and healthful foods. Likewise, the per-capita production and consumption of cheeses have seen notable growth in the USA, Europe, and the Indian subcontinent. Thus, FMs have been, and still are, of extreme importance in the nutrition of people throughout the world.

Table 1 Changes in milk constituents resulting from fermentation

Constituents	*Changes*
Milk proteins (3–4.5%)	Coagulated into a smooth curd with casein particles finely dispersed; partially peptonized (0.1–0.7%) and utilized for microbial growth, resulting in the build-up of microbial cell proteins, increase in non-protein nitrogen and release of peptides and amino acids
Lactose (4.5–5%)	Utilized (1–2%) by starter bacteria producing mainly lactic acid (0.6–2%), volatile acids, flavor compounds, and CO_2 may also be formed by heterofermentative types; alcohol and CO_2 may be produced by lactose-fermenting yeasts
Milk fat (3.5–7%)	Fermentation leads to partial digestion of lipids
Mineral matter (0.7–0.8%)	No direct appreciable change
Vitamins	No change in fat-soluble vitamins A, D, and E; slight decrease in B-complex vitamins, depending on the strains used

Source: Khedkar CD and Khedkar GD (1993) Fermented milks: Dietary importance. In: *Encyclopeadia of Food Science, Technology and Nutrition*. Academic Press, London, pp. 1810–1814.

Nutritional Significance of Fermented Milks

Lactobacilli use the nutrients in milk for their own metabolism and growth, and multiply from one to 10 million cells per milliliter. The microorganisms are present in the FM, not only as viable cells, but also as autolyzed cells that give rise to primary and secondary metabolites and enzymes that they may have produced during fermentation and storage.

The nutritional value (NV) of a FM depends on the availability and digestibility of nutritive constituents, and the changes in these constituents resulting from microbial growth and fermentation processes (Tables 1 and 2). Similarly, the NV of fermented food also depends on the nutrient content and the bioavailability of the nutrients. Generally, the energy value of a FM is similar to that of the milk from which it was prepared, yet it is claimed that the FM are more nutritious (the compositions of milk and yogurt are given in Table 2). Fermentation makes some of the nutrients more available for absorption. The interactions between milk nutrients and starters are detailed below.

Table 2 Composition of milk and yogurt (per 100 g)

Nutrients	*Whole milk*	*Whole-milk plain yogurt*	*Low-fat plain yogurt*	*Very-low-fat/low-calorie flavored yogurt*	*Greek yogurt*	*Plain fromage frais*
Energy (kcal)	66	79	56	41	115	113
Protein (g)	3.1	5.7	5.1	4.3	6.4	6.8
Fat (g)	3.9	3.0	0.8	0.2	9.1	7.1
Calcium (mg)	115	200	190	130	150	89
Zinc (mg)	0.4	0.7	0.6	0.4	0.5	0.3
Phosphorus (mg)	92	170	160	110	130	110
Magnesium (mg)	11	19	19	13	12	08

Source: Buttriss J (1997) Nutritional properties of fermented milk products. *International Journal of Dairy Technology* 50 (1):21–27.

Predigestion of Protein and Protein Quality

The proteins in milk are of an excellent quality biologically, and both caseins and whey proteins are well endowed with essential amino acids (EAA). The fact that the protein content of FM, like yogurt, is often elevated by concentration or addition of skim milk solids means that it is an even more attractive source of protein than liquid milk. It is important that the proteins in FMs are totally digestible, because some degree of initial proteolysis is caused by the starter organisms. The extent of this breakdown depends on the strains of bacteria being employed, but, in general, at least some release of amino acids and peptides can be expected during incubation and storage. FM products contain significantly higher levels of free amino acids (FAA) than milk, because the heat treatment and proteolytic activity of the starter organisms. The native milk proteins, which form a hard curd in the stomach, are converted into a soft curd containing finely dispersed casein particles resulting from the action of starter organisms in the FM. The FM proteins are useful to children, old people, and those with stomach ulcers, because the soft curd in FM is easier to digest than the milk.

It has been observed that rats fed with diets containing FM products had a higher feed consumption, greater weight gains, and a more efficient utilization of their feed. Some medical practitioners advocate the use of dahi and buttermilk by children not only for treatment of common intestinal ailments but also for their general nutrition. In tropical countries like India, where milk is exposed to heavy bacterial contamination because of the tropical atmosphere and low standards of hygiene, leading to rapid spoilage, it is better to convert milk into FM in order to preserve its NV.

Predigestion of Milk Fat and Fat Quality

Although much of the FM sold in industrialized countries is produced from skim milk, traditional materials have always contained some 3–4% milk fat. It has been proposed that the digestibility of fat is improved via the action of the bacteria present during the fermentation process, but whether this effect really exists is highly debatable. The overall energy content of FM reflects both the fat content of the milk from which it was made and whether or not ingredients such as cream or sugar have been added. Humans have a double requirement for lipids. Lipids act as a storage fat, which is composed of saturated fatty acids, and serve as a source of energy or as a protection for vital organs. In addition, they play an important role by acting as structural fat, which, together with proteins, forms many of the essential membranes in animal cells, particularly in areas like brain and nerve cells. These are essential to many physiological processes. Some essential nutrients are fat-soluble and are found primarily in foods that contain fat. These nutrients are essential fatty acids (EFA) and fat-soluble vitamins. Shahani and Chandan (1979) cited the fact that the LAB possess lipase activity, as evidenced by lipids in cultured products being partially degraded. Their assays for lipase activity, however, were performed using culture organisms with tributyrin emulsions as a substrate, and they did not prove that the bacterial lipases acted on lipids in the cultured products. A significant increase in free fatty acids over the milk has been observed in FM. The lactic acid fermentation and enzymic degradation of amino acids also produce small quantities of volatile fatty acids.

Predigestion of Lactose

The milk carbohydrate, lactose, causes intestinal problems in a number of persons who are lactose-intolerant. These persons are deficient in the intestinal enzyme, β-galactosidase, and thus must restrict their intake of milk and dairy products. Milk does not contain this enzyme. Lactose intolerance is normally defined as the occurrence of gastrointestinal symptoms after administration of a single test dose, about 50 g of lactose in aqueous solution. In a large proportion of the world's population, β-galactosidase activity is low or absent. This enzyme is present in sufficient quantities in lactic fermented milks. Consequently, the lactose level in FM can be lower than in milk, although this is not always the case, as skimmed milk powder or non-fat milk solids are sometimes added during the manufacture of some of these products; during fermentation, 20–50% of the lactose in milk is hydrolyzed to its component monosaccharides, glucose and galactose, by the starter culture organisms. The production of β-galactosidase increases during the fermentation of yogurt, reaching a maximum after 4 h of incubation. This enzyme may be released from the microbial culture during digestion, thus indicating that it may be present in the intestine of persons consuming FM. Lactose-intolerant subjects have a much lower rise in blood glucose and thus a much lower lactose utilization efficiency. A lower lactose content would presumably help tolerance of the product to those with a reduced ability to digest lactose. But clearly, the explanation is not this simple, because FM with a fairly high lactose content is also better tolerated by such individuals than the equivalent amount of lactose in milk. Lactic

acid is produced as a byproduct of the fermentation process. It acts as a preservative, influencing the physical properties of casein curd by inducing a finer suspension, which improves the digestibility of casein. It also improves the utilization of calcium and other minerals, and inhibits the growth of potentially harmful bacteria.

Effect of Fermentation on Vitamin Content

Milk is a rich source of vitamins, particularly A, D, and some B-complex vitamins. Breed, diet, climate, geographical location, age and stage of lactation and other factors can influence the vitamin content of the milk and, in turn, affect the vitamin content of the FM. There is conflicting evidence about whether there is an increase, decrease, or no change in the level of vitamins during the fermentation of milk. Various manufacturing treatments may affect the content of labile vitamins, whereas there is metabolism of vitamins by the LAB during the log-growth phase and their subsequent synthesis by the same bacteria. An indication of the changes, resulting from heat treatment, fermentation, and storage, is given by the International Dairy Federation (1983). The vitamin content of FM varies with the type of milk used (particularly the fat content of the milk, which influences the amount of vitamin A and other fat-soluble vitamins present), with the strain of bacteria and with fermentation conditions. Vitamin C is heat- and light-labile, but it is more stable in the acidic conditions of FM than in normal milk. The levels of some B-complex vitamins are reduced, owing to the requirement of some of the LAB. Losses of up to 90% of vitamin B_{12} have been reported with specific bacterial strains, but losses are not always this great and can be reduced considerably by the use of a 'supplementary culture' capable of synthesizing significant amounts of vitamin B_{12}. However, several cultures are able to synthesize the folic acid, increasing the level of this vitamin present in the final product; e.g., *Propionibacterium freudenreichii* ssp. *Shermanii*, which is used as a starter culture in some of the cheeses, is known to synthesize vitamin B_{12}.

In the current climate, in which there are concerns about the folate intake of women during the early stages of pregnancy, this aspect may prove to be a nutritional advantage. Increases in vitamin B and P (riboflavenoid) group vitamins have been observed in kefir, as a result of the activity of yeasts and acetic acid bacteria. Fermentation has been reported to increase folic acid in buttermilk, sour cream, yogurt, bifidus milk, kumiss, dahi, lassi, laben, skyr, and kefir. The effect of fermentation on the vitamin content is given in Table 3.

Table 3 Effect of fermentation on vitamin content (μg per 100 g)

Vitamins	*Whole milk*	*Buttermilk*	*Sour cream*	*Yogurt*	*Cottage cheese*
Thiamin	30	34	35	44	21
Riboflavin	170	154	149	214	165
Niacin	100	58	67	114	128
Pantothenic acid	300	275	360	591	215
Vitamin B_6	40	34	16	49	68
Folacin	6	trace	11	11	12
Vitamin B_{12}	0.40	0.22	0.30	0.56	0.63

Department of Health (1991) *Dietary Reference Values for Food, Energy and Nutrients for United Kingdom. Report on Health and Social Subjects*. London: HMSO.

Effect of Fermentation on Mineral Absorption

Most experiments performed recently did not confirm the improved absorption of calcium because of the more acid pH in the intestine following consumption of FM. When the bioavailability of essential minerals and trace elements was investigated in rat experiments using diets based on milk, yogurt, pasteurized yogurt, and a commercial diet, it was observed that the bioavailability (as measured by intestinal absorption, urinary excretion, and bone content) of calcium, phosphorus, magnesium, and zinc from all diets was superior to that from the commercial diet. In experiments with albino rats, the availability of calcium and phosphorus from FM was increased by about 7 and 11%, respectively, compared with milk. These results are explained by the fact that colloidal calcium complexes and lactic acid, both present in FM, enhance calcium absorption.

Bioactive peptides derived from the tryptic hydrolysates of casein, known as caseinophosphopeptides (CPP) possess physicochemical properties that enable chelation of various bi- and trivalent minerals to be carried out, thereby enhancing mineral solubility in the lower small intestine. It has been suggested that moderate and exchangeable binding of calcium to CPP is responsible for the high absorbability of calcium from milk. Results of experiments conducted at the University of Melbourne School of Dental Studies have proved that there is a significant reduction in the tooth decay in laboratory animals fed with CPP.

Nutritional Significance of Single-cell Proteins in Fermented Milks

In recent years, the nutritional importance of microbial cell protein has received considerable attention. When FM is consumed, a large number of bacterial cells (about 10^7 CFU g^{-1}) and the products released from these cells enter the digestive tract. Most of the

living cells are inactivated in the stomach by the action of hydrochloric acid and pepsin. It has been estimated that about 60% of the microbial cell nitrogen content (excluding cell wall material and nucleic acids) can be utilized by the body. The EAA content of microbial cell proteins in FM ranges from 2.0 to 6.5 mg per 100 g of product. The cells of some bacteria were found to be rich in methionine, lysine, and cystine. Studies on amino acid patterns of acid hydrolysates of microbial cell proteins have revealed the presence of all the known amino acids. Cells of mixed cultures contained higher concentrations of many amino acids than those found in cells of individual cultures. There were also considerable species variations in the pattern of amino acids in the microbial cells. When the cells were subjected to proteolytic enzymes, simulating to some extent the conditions during passage in the alimentary canal, significant differences were found between species with regard to the extent of hydrolysis and concentration of EAA released. Cells of *Lactobacillus acidophilus* and of mixed cultures of *Streptococcus salivarius* ssp. *thermophilus* and *Lactobacillus delbrueckii* ssp. *bulgaricus* released the highest concentrations of EAA. The biological activity of many EAA in FM protein is much higher than that of corresponding amino acids in milk. These results indicate that the microbial cells of starter bacteria consumed along with FM are likely to be hydrolyzed partially or fully in the digestive tract, releasing EAA, and some of the starter cultures are more useful than others in this respect.

Biological Value of Fermented Milks

The cumulative effect of the available EAA from milk proteins and from microbial proteins is reflected in the protein quality (biological value) of the total protein content isolated from FM prepared with different cultures. The protein quality of Indian FM, dahi was assessed by microbiological assay using *Streptococcus zymogenes* as the test organism. It was found to be higher (3–30%) than that of milk used for preparing dahi. Samples prepared by using mixed cultures of *Lactococcus lactis* ssp. *lactis*, *Lc. lactis* ssp. *cremoris*, *Str. salivaricus* ssp. *thermophilus*, and *Lb. delbrueckii* ssp. *bulgaricus* gave higher values than those containing other cultures. Incorporation of the culture of *Propinobacterium freudenreichii* ssp. *Shermanii*, along with the starter cultures, produced a further improvement in protein quality. It has also been found that there is increased secretion of digestive enzymes by salivary glands when stimulated by the curd particles, and that the FM proteins are twice as digestible as milk proteins.

Biopeptides in Fermented Milks

Several of the peptides derived from milk proteins represent 'extranutritional' substances, which display significant physiological influence. During the microbial fermentations milk proteins undergo controlled proteolysis under the combined influence of native microflora (comprising several spore-producers) and the starter bacteria. In addition, proteinases produced by the starter organisms could complement the complex enzyme system of starter microorganisms, leading to the accumulation of beneficial peptides. Cheeses and processed cheese products, therefore, offer the potential for developing specialized functional food products displaying prophylactic properties, besides their natural nutritional virtues. This represents a hitherto unexplored area of food research, which has tremendous future potential, from both an academic and commercial perspective.

The folklores regarding the nutritional, healthful, and therapeutic aspects of FM have been amply substantiated through recent biochemical, nutritional, and physiological research. With the growing awareness of beneficial attributes of such products, there has been a tremendous growth of market for probiotic FM products in the Western countries, especially in Japan and developing countries. Prospects for the genetic modification of starter organisms to enhance the NV and probiotic attributes have opened challenging vistas for the new product development and diversification. FM as a potential carrier for natural source of biological peptides is a fascinating area for technological development.

See also: ***Acidophilus* Milk**; **Dahi**; **Fermented Milks**: Dietary Importance; **Yogurt**: The Product and its Manufacture; Yogurt-based Products; Dietary Importance

Further Reading

Buttriss J (1997) Nutritional properties of fermented milk products. *International Journal of Dairy Technology* 50(1): 21–27.

Department of Health (1991) *Dietary Reference Values for Food, Energy and Nutrients for United Kingdom. Report on Health and Social Subjects*. London: HMSO.

International Dairy Federation (1983) *Cultured Foods in Human Nutrition. Doc. 159*, pp. 18–28.

Khedkar CD and Khedkar GD (1993) Fermented milks: Dietary importance. In: *Encyclopeadia of Food Science, Technology and Nutrition*. Academic Press, London, pp. 1810–1814.

Mathur BN (1998) Nutrition and probiotic aspects of cheese and fermented milk products. In: *Advances in Cheese and Fermented Milk Products: A Compendium of Short Term Course Notes*, Karnal, India: National Dairy Research Institute, pp. 210–217.

Shahani KM and Chandan RC (1979) Nutritional and healthful aspects of cultured and culture-containing dairy foods. *Journal of Dairy Science* 62: 1685–1694.

Yamauchi K (1992) Biologically functional proteins of milk and peptides derived from milk proteins. IDF Bulletin 272: 51–58.

Fiber *See* **Dietary Fiber**: Properties and Sources; Determination; Physiological Effects; Effects of Fiber on Absorption; Bran; Energy Value

FIGS

K K Sinha, T.M. Bhagalpur University, Bhagalpur, India

Background

Along with date, vinifera grape, and olive, the fig was an important food crop for the ancient civilization of the eastern Mediterranean region. It has been referred to in many legends and songs of historical and mythological background. It has also been mentioned frequently in the Bible. Theophrastus was also familiar with its several varieties, as is indicated in his 'Enquiry into Plants.'

The edible or common fig, *Ficus carica* L., is a member of the family Moraceae, which includes more than 1000 species distributed throughout the world. Some well-recognized ornamental species of the genus *Ficus* include *Ficus benghalensis* L. (Indian banyan), *F. elastica* Roxb. (Indian rubber fig), *F. pumila* L. (Climbing fig) and *F. religiosa* L. (Peepal tree).

The fig is thought to be a native of southern parts of Arabian peninsula, Italy, the Balkan peninsula, and Russia. It was brought into cultivation probably in the southern parts of the Arabian peninsula by 3000 BC. Later on, it spread into Iraq, Syria, and Turkey, and into all the Mediterranean countries. During the age of exploration by Columbus, the fig was taken into most subtropical areas of the western hemisphere. It is widely cultivated in the tropical, subtropical, and warm temperate areas of the world. The major production is found in the Mediterranean region, from Turkey in the east to Spain and Portugal in the west. It is also grown commercially in parts of the USA and Chile and, to a small extent, in Arabia, Iran, India, China, and Japan.

Greece is now the leading fig-producing country, followed by Algeria, Morocco, Syria, and Italy. Turkey is, however, the largest producer of dried figs.

General Habit

The fig is a small to moderate-sized deciduous tree, 5–8 m high, with a short twisted trunk crowned with irregular branches. Shoots can also develop frequently at the base of the trunk. The bark is pale gray. Leaves are broad, ovate or nearly orbicular, more or less deeply (3–5) lobed, rough above, and pubescent below, long stalked, leaf blade 10–25 cm long, dark green with pronounced venation. The female fig plants have larger, denser, and more spreading crowns of leaves than males. Moreover, leaf fall begins earlier in males.

Fruit Morphology

Fruits are mostly solitary, axillary, green, yellow brown, purplish, or even black, depending on the cultivar and more or less pear-shaped with either a velvety or glabrous skin. Normally, these fruits are of moderate size, but in certain cultivars, these may be up to 6 cm in diameter. The plastid pigments in the fruit skin of fig are chlorophyl *a* and *b*, β-carotene, lutein, violaxanthin, and neoxanthin.

The edible fig is a multiple fruit that is botanically known as syconium. It consists of a fleshy hollow receptacle with a narrow aperture at the tip and numerous small flowers lining the inner surface. The true fruits, which are small achenes, are borne on short stalks on the inside of the syconium (**Figure 1**).

Types and Cultivars

Depending upon the nature of the flowers and the method of pollination, figs can be cronologically distinguished into four different classes:

1. Common fig or Adriatic fig
2. Caprifig
3. Smyrna fig
4. San Pedro fig.

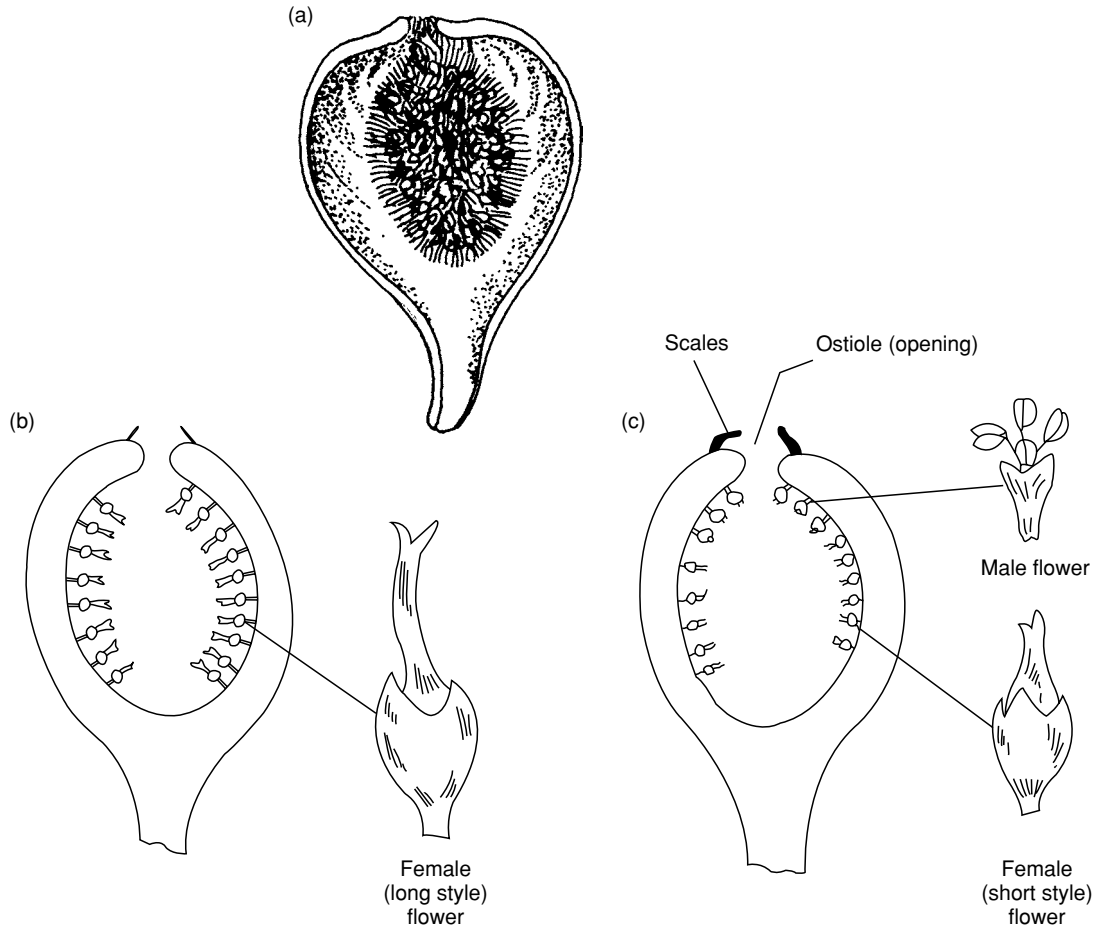

Figure 1 Longitudinal sections of fig fruits. (a) Edible fig fruit. (b) Edible fig fruit (diagrammatic). (c) Capri fig fruit (diagrammatic).

Common Fig

In this type, the individual flowers are pistillate (female), and consequently, the fruits develop without pollination and fertilization (parthenocarpic) and have no seeds. Two crops are produced annually. The first crop (*brebas*) is larger and more juicy, and is usually eaten fresh. The fruits are borne on the old wood. The second crop is produced in the axils of the leaves, and the fruits are used fresh or dried. Some common cultivars of this type are Kadota (Dottato), Mission, Adriatic, Brown Turkey, Celeste, and Conadria.

Caprifig

The caprifigs are wild figs that grow naturally in the Mediterranean region and Western Asia and probably represent the primitive type. Although the fruits do not have any commercial value, they are cultivated in most fig-raising countries, as they are essential for the development of Smyrna fig. The life history of the caprifig is closely connected with that of a small wasp (*Blastophaga psenes*), which brings about cross-pollination.

Caprifigs produce three crops of fruit in a year. The spring crop (Profichi) contains staminate (male) flowers and the so-called gall flowers. These are similar to pistillate flowers but have short-styled ovaries. The fig wasp enters the young figs and lays eggs in the gall flowers. In about 2 months, the new generation of wasps hatches out and emerges from the fig having pollen grains throughout the body surfaces. By this time, the summer crop of figs (Mammoni) are produced, which contain mostly gall flowers. The wasps again enter these figs and deposit eggs in most of them. Although these flowers are pollinated by wasps, the presence of the larvae without eggs may develop fertile seeds. Late in the season, the winter crop of figs (Mamme) is developed and remain on the tree over winter after the wasps visit on them. The larvae mature in April, when a new crop of spring fig is ready to receive the wasps, and the annual cycle is resumed.

Smyrna Fig

No staminate flowers are produced in Smyrna figs, and these are absolutely dependent on cross-pollination from caprifigs. This process is known as 'caprification' and is brought about artificially. Branches of caprifigs of the spring crop are suspended on the Smyrna tree. The wasps coming from Spring crop enter the developing Smyrna figs and affect pollination. The ovaries of Smyrna figs have longer styles that prevent the wasp from depositing eggs at the proper site. The ovules thus develop normally after fertilization. Smyrna figs have a superior nutty flavor owing to the presence of fertile seeds. They are the most important fig of commercial value and are extensively cultivated in Asia Minor, Greece, Algeria, and parts of Portugal and California. Early attempts to cultivate Smyrna figs in California and other places met with continued failure until the spring crop of caprifigs and the fig wasp were introduced. Calimyrna is the most common and widely grown cultivar of Smyrna type. Some cultivars like Krymskii 43 and Bianco Grosso have also been recommended for cultivation in Southern Uzbekistan.

San Pedro Fig

This is an intermediate type of fig that has two crops annually. Like the common fig, the first crop (breba) is completely parthenocarpic and does not require pollination and fertilization of flowers, but the second crop develops only when the flowers are pollinated, as in the case of Smyrna fig. San Pedro, King, and Gentile are some of the common cultivars of this type.

The common fig is the only type that is grown in India. It is considered to be a hybrid between the imported *F. carica* and the indigenous species. A large number of cultivated forms are known in which the fruits vary in shape, size, color of skin, color and flavor of flesh, and period of ripening. Some of the cultivars grown in India are Black Ischia, Brown Turkey, Turkish White, Kabul, and Marseilles. The figs grown in many parts of India are named after the locality and exhibit no special distinction to warrant varietal names. Poona fig is of a medium size, bell-shaped and light purple in color with a rosy flesh. In South India, Poona fig and Marseilles are commonly cultivated, the latter thriving well on the hills above 1500 m. The fruits of Marseilles are medium-sized and pale green on the rind with a whitish sweet flesh.

Fig Cultivation

Soil

The fig tree does not require a very rich soil for cultivation. Alluvial or loamy soil or yellow or reddish brown colour, with a rocky or murum bed, prevents the roots from penetrating deep into the soil, and favors the side-growth of rootlets, which is very desirable. Fig trees also thrive well in clayey soil, but the land must be well drained. Rich black soil is unsuited to fig trees because the plant grows tall and runs to leaf, and the fruit is considerably inferior in both size and taste.

The fig is also one of the salt- and drought-tolerant crops. It can tolerate a fairly high level of sulfate or chloride salts, but it can be injured when even a small amount of sodium carbonate is present in the soil. In highly alkaline soil, the leaves show tip burn, and there is considerable leaf fall. A red loam with a live substratum and about a meter deep is considered best for figs in South India.

Heavy and poorly drained soils usually produce fruits of inferior quality. Soils with a high lime content produce fruits of better quality suitable for drying.

Climate

The fig is a subtropical fruit. In the dormant condition, the mature trees can withstand temperatures as low as -12 to $-9.5\,^{\circ}C$ depending upon the cultivar, but the young trees may have to be protected during winter because they are not very hardy. Low temperatures of about 1.5–4.5 °C at the time when the trees are not dormant are very much damaging. Similarly, early autumn frost or spring frost when the plant starts sprouting causes considerable damage to the tree. In India, the temperature of commercial fig growing areas in Southern and Western parts of the country seldom falls below 4–5 °C, and therefore, there is no damage resulting from low temperatures. In Northern India, where the temperature falls to freezing, the plants remain dormant at that time and normally are not damaged. Figs grow well when the temperature range is 15.5–21 °C.

The climate also greatly influences the characteristics of fig fruits. It affects the shape and size of the fruit, as well as the color of the skin and pulp, quality, and tendency towards parthenocarpy. The best-quality figs are produced in regions with a dry climate, especially at the time of fruit development and maturation. High humidity coupled with a low temperature usually results in fruit splitting and a lower fruit quality. Fruits exposed to warm breezes in hot areas, though slightly sweeter, remain small in size. By contrast, mild temperatures result in large, succulent fruits.

Propagation

The plant is readily propagated by cuttings, although all types of budding and grafting as well as air-layering

are successful. Hardwood cuttings, 20–30 cm long and 0.5–0.7 cm thick, taken from 1–2-year-old shoots in late summer are selected for this purpose. These cuttings are treated with root-promoting substances, stored in packing material like sawdust at room temperature for about 4 weeks, and then planted. Cuttings taken from the base of the shoot and lower part of the crown give better results. In about 2 months, they begin to throw roots and shoots and make a few leaves. If they are properly taken care of, the plants become fit after a year for transplantation, otherwise they may take a further 10–12 months. After a 12–15 months, they are transplanted into orchards. The spacing varies depending upon the size of the tree, variety, and type of culture adopted.

Figs can also be air-layered successfully, and monsoon is the best time for it. One-year-old branches, if layered in June, can be planted in the field in August–September. Top working to change the variety of established trees is sometimes possible and may be done by shield or patch budding or cleft or bark grafting. Micropropagation of figs is also possible.

Pruning and Training

Fig trees are pruned annually and trained to the desired height and shape to keep the plant most productive and to facilitate harvesting and other orchard operations. As the tree normally bears two crops in a year, the first pruning is done on the wood of the previous season and the second on new wood of the current season; the time and the amount of pruning are adjusted according to the type, habit of growth, and bearing capacity of the tree. The main aim of pruning is to induce the growth of fruit bearing wood and so improve the yield of fruits. In addition, pruning increases the fruit weight in early cultivars. Besides pruning, some special methods like notching are also adopted to stimulate the production of laterals on vigorous upright branches.

Irrigation and Manuring

Fig is fairly drought-resistant, and it is seldom irrigated in many fig-growing countries. However, reports from Cairo indicate a greater shoot growth and yield of fruits in irrigated fig plants. In some cases, fruits from irrigated trees have been found to 13–16% heavier. In India, the fig is irrigated only during dry months. However, care should be taken that at the time of ripening of fruits, no irrigation is given, because that may result in the production of insipid fruits. Excessive water in the soil may also cause splitting of fruits. The plant also responds to manuring, but the type and amount of manure vary in different areas.

Fruit Development and Harvesting

The fig tree begins to bear a fair crop from the second or third year after planting, though rooted cuttings and layered plants are known to bear fruits even in the first year of planting. Under favorable conditions, the trees continue to bear fruits for 12–15 years, after which they show a marked decline in yield. Normally, the trees bear two crops in a year, but in some types, even a third crop is obtained.

The figs commonly grown in India are parthenocarpic in nature and do not need any cross-pollination with wild caprifig (caprification), which is very common practice in other countries. However, fruit setting is also inhibited under certain conditions. It has been suggested that parthenocarpy is favored or inhibited in a given type by climatic condition of the growing area. The parthenocarpic fruits develop to a normal size and have a desired sugar content, but as they are completely seedless, the baking industries do not want to use it, because they lack the crunchy quality imparted by fig seeds.

As fresh figs are extremely perishable, fully ripe fruits cannot be transported to distant markets. It is better to harvest the crop when the fruits are slightly immature (ripened). Ripe fruits are picked either from the tree by twisting the neck at the stem end or by cutting it or gathered after they drop on the ground. In the major fig-growing countries, special devices are employed for collecting fruits from the tree, and the pickers are protected against the acrid juice.

Drying

Fresh figs do not keep well even under cold storage (0 °C) for more than a month. Fully matured figs can be either frozen or dried if they have to retain their flavor and color for several months. Details of the process of drying and curing differ in different countries depending largely upon the variations in varieties cultivated.

Dried figs are the most important fruit product of fig. In one method of drying, the fruits are firstly subjected to sulfur fumigation and then dried in an electric drier at a temperature of 70–72 °C. In another method, the fruits are initially soaked in boiling salt water for 30 s, then dried for few hours in the sun and for 8 days in the shade.

Pests, Diseases, and Disorders

The following major pests have been reported from fig trees:

- Stem-boring beetles (*Batocera rufomaculata, B. rubrus*)
- Leaf defoliators (*Adoratus duvauceli*, *A. versutus*)

- Scale insects (*Aspidiotus lataniae*)
- Fig flies (*Lonchaea aristella*).

In addition, dried figs have been found to be damaged by fig moth (*Ephestia cantelia*) and the Indian meal moth (*Plodia interpunctella*). Some of these pests can cause more than 50% losses in yield. The fig leaf roller (*Phycodes mina*) has been reported in Haryana (India), for which excellent control was obtained with trichlorphon (0.05%) and chlorpyrifos (0.03%).

Figs are also susceptible to root-knot nematodes, which can be controlled by using DD-mixture.

Leaf rust is an important disease of figs in all the fig-growing countries of the world that is caused by *Cerotelium fici* (Syn. *Uredo fici*). Rust-affected leaves show small, round, brownish to black eruptive lesions mostly on the lower surface. The affected leaves drop off prematurely and, in severe cases, affect the yield of fruits. This disease can be controlled by dusting with sulfur or spraying with zineb or Bordeaux mixture.

Anthracnose (caused by *Sphaceloma fici-caricae*), a leaf spot disease (caused by *Cylindrocladium scoparium*) and mosaic disease (caused by virus) of fig have also been reported from different parts of India.

Some important diseases of fig fruits are caused by *Botrytis cinerea, Alternaria fici, Aspergillus niger, Fusarium moniliforme*, and *Cladosporium sicophilum*, which can be controlled by using some effective fungicides. Rhizopus rot (caused by *Rhizopus stolonifer*) of fig is also quite common among storage rots and can be controlled by dipping the fruits in a solution of aureofungin and copper sulfate.

Sunburn and fruit splitting are the important disorders of fig plants. Newly planted young trees are mostly affected by sunburn in which the affected parts crack, and the bark peels off. Heavy pruning exposing the trunk and branches is responsible for sunburn. Whitewashing of exposed parts can prevent this malady. Fruit splitting usually occurs if there is a shower during ripening of fruits. Fruit splitting damages the fruit, making it completely unfit for consumption.

Uses

Figs are consumed fresh, dried, preserved, canned, or candied. Fresh figs are delicious and can be used as a dessert or for jam. The great bulk of the crop is consumed as dried fruit. Better grades of dried fig are used for making fancy packs, whereas other grades are used for the production of alcohol and wine. Figs are also roasted like coffee beans and can be used as a substitute of coffee. Other fig products are spiced or pickled figs, fig bread, fig meat, and fig brownies.

Fresh figs generally consist of 84% pulp and 16% skin. The chemical composition varies with type. The average nutritional composition of dried figs (per 100 g) has been recorded as follows:

Energy: 220 cal
Protein: 3.3 g
Carbohydrate: 48.0 g
Fat: 1.5 g
Fiber: 12.4 g
Sodium: 62 mg
Potassium: 970 mg
Vitamin C: 1 mg
Thiamin: 0.08 mg
Riboflavin: 0.1 mg
Nicotinic acid: 0.8 mg
Pantothenic acid: 0.5 mg
Folic acid: 9 μg
Vitamin B_6: 0.26 mg
Calcium: 250 mg
Iron: 4.2 mg
Phosphorus: 89 mg
Magnesium: 80 mg
Copper: 0.3 mg
Zinc: 0.7 mg

Figs owe their food value chiefly to their mineral and sugar contents. The total mineral content is two to four times that of most other fresh foods, and only cheese and a few types of nuts have a higher calcium content. They are richer in iron and copper than nearly all fruits and vegetables and most other dried fruits. The nutritive index of fig in comparison with that of other fruit is reported to be as follows: fig, 11; apple, 9; resin, 8; date, 6; pear, 6.

The fruit, fresh or dried, is valued for its laxative property. It is diuretic, demulcent, emollient, and nutritive, and is used in the form of confection and syrups. Figs are considered useful in the prevention of nutritional anemias. The ash of figs is highly alkaline.

The fig tree yields a latex that is used as an anthelmintic. It is toxic when administered parenterally to animals but has no toxic effect when administered orally. The anthelmintic action has been traced to ficin, a proteolytic enzyme that has the remarkable power of digesting living helminths. This ficin also reacts with protein of the skin, causing dermatitis. Some individuals are adversely affected by such materials and cannot ingest fresh figs without skin eruptions around the mouth. Many field workers in fig orchards suffer from dermatitis on the arms and body if they brush against the limbs and leaves, or as they handle the fresh fruits. Gloves and protective aprons are generally provided for such workers.

Fresh fig latex also shows milk-clotting activity, which may be 30–100 times greater than that of the animal rennet prepared from calf stomach mucosa.

See also: **Date Palms**; **Jams and Preserves**: Methods of Manufacture; Methods of Manufacture

Further Reading

Anonymous (1956) *The Wealth of India. Raw Materials*, vol. IV, pp. 26–32. New Delhi: Council of Scientific and Industrial Research.

Condit IJ (1941) *Fig Culture in California. California Agricultural Extension Service Circular 77*, pp. 1–67. Berkeley, CA: University of California.

Condit IJ (1947) *The Fig. A New Series of Plant Science Books*, vol. XIX, pp. 1–222. Waltham, MA: Chronia Botanica Company.

Condit IJ (1955) *Fig Varieties: A Monograph. Hilgardia 23 (No. 11)*, pp. 323–538. Berkeley, CA: California Agriculture Experimental Station.

Eisen G (1901) *The Fig: Its History, Culture and Curing. USDA, Division of Pomology Bulletin No. 9*, pp. 1–317. Washington, DC: Government Printing House.

Hill AF (1952) *Economic Botany*, pp. 420–422. Bombay: Tata McGraw-Hill.

Nagpal RL (1966) *Fig Cultivation in India. Farm Bulletin No. 42*. New Delhi: Indian Council of Agricultural Research.

Sadhu MK (1996) Fig. In: Bose TK and Mitra SK (eds) *Fruits: Tropical and Subtropical*, pp. 650–666. Calcutta: Naya Prakash.

Samson JA (1980) *Tropical Fruits*. London: Longman.

Simpson BB and Conner-Ogorzaly M (1986) *Economic Botany. Plants in Our World*, pp. 129–130. New York: McGraw-Hill.

Song Q, Yang D, Zhang G and Yang C (2001) Volatiles from *Ficus hispida* and their attractiveness to fig wasps. *Journal of Chemical Ecology* 27(10): 1929–1942.

Young JH (1999) The early regulation of figs in the United States. *Technology* 6: 229–234.

FILTRATION OF LIQUIDS

J Keifer, Formerly of Filtrox-Werk AG, St Gallen, Switzerland

Introduction

Filtration and separation play an important role in the preparation of beverages, especially beer and wine, which, as the products of a fermentation process, require an extensive reduction in microorganisms and haze-forming substances at the end of the production cycle. The aim of this clarification step is to obtain a product which does not contain any harmful microorganisms, is visually bright, and has the required shelf-life.

Selection of the clarification process necessitates a careful analysis of the separation problems involved. Not only is the result of the clarification process itself a criterion, but additional factors also have to be taken into account.

The clarification processes commonly used in the brewing and wine-making industries are kieselguhr filtration, sheet filtration, cross-flow microfiltration (for wine) and dead-end membrane filtration. The first two processes are the dominant filtration technologies.

Kieselguhr Filtration

Kieselguhr or diatomaceous earth filtration in the brewing and wine-making industries is a precoat filtration with the addition of further kieselguhr as body-feed. In brewing, it is always a pressure filtration to maintain the carbonation of carbon dioxide content of the beer; in wine-making, it may be pressure or vacuum filtration.

The application of kieselguhr filtration is restricted to a maximum solids content of about 0.2%. Feeds with higher solids contents always require pretreatment.

Basic Technology of Kieselguhr Filtration

Microorganisms and haze particles are much smaller than the sieve aperture of that part of the filter's structure used as a mechanical support for the kieselguhr cake. Therefore, several precoat layers of coarse kieselguhr are applied to obtain a tight cake which retains the body feed and beer/wine solids. Continuous dosing of body feed prevents a rapid pressure increase by dilution of the solids and, therefore, results in longer filtration runs.

The relationship between filtration pressure, filtrate flux, or specific flow rate (flow per unit area), cake thickness, and cake permeability is obtained from Darcy's law and its derivations.

One derivation of Darcy's law, containing the parameters of cake porosity, specific gravity, and thickness, is:

$$U = \frac{\Delta P}{\mu a \rho s (1 - \sum) L}. \qquad (1)$$

As ρs and Σ are constants for a specific cake, the expression can be simplified to:

$$U = \frac{\Delta P}{\mu a c_k L}, \qquad (2)$$

where U is the filtrate flux (m s^{-1} or m^3 s^{-1} m^{-2}), μ is the filtrate viscosity (N s m^{-2}), ΔP is the differential pressure (N m^{-2}), a is the specific filtration resistance (m kg^{-1}), ρs is the specific gravity of the cake-forming particles (kg m^{-3}), Σ is the filter cake porosity, c_k is the filter cake constant (kg m^{-3}) ($=\rho s(1 - \Sigma)$), and L is the cake depth (m). This expression can be applied to model the different means of operating a kieselguhr filter, that is:

- constant flow, variable pressure;
- constant pressure, variable flow; and
- variable flow, variable pressure.

The first of the above is commonly used in practice for beer and wine filtration.

Kieselguhr Filtration of Beer

A typical beer plant consists of the components shown in Figure 1. The process starts with the kieselguhr filter unit filled with water. Then, after a two-step precoat, as specified in Table 1, the water is pushed out by incoming beer; a portion of the resulting diluted beer is collected and blended back to the unfiltered beer during the production run. (*See* **Beers:** History and Types.)

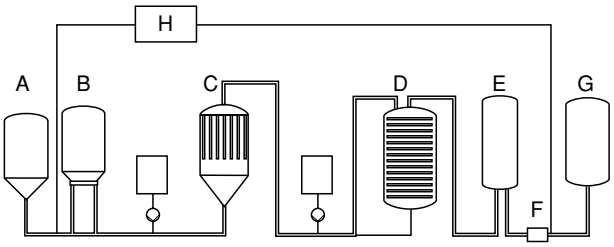

Figure 1 Components of a typical beer filtration plant: A, storage tank for unfiltered beer; B, buffer tank for unfiltered beer; C, kieselguhr filter including dosing unit; D, polyvinylpolypyrrolidone stabilization unit, including dosing unit; E, buffer tank for bright beer; F, carbonation unit; G, bright beer tank; H, cleaning-in-place (CIP) station.

Table 1 Kieselguhr permeabilities for different process steps

Process step	*Permeability*	*Examples*
First precoat	1.2–1.5 Darcy	Manville Hyflo Ceca Dic B Dicalite Speedplus
Second precoat	100–250 m Darcy	Manville Standard Supercel Ceca CBR Dicalite Speedflow
Bodyfeed	30–250 m Darcy	Manville Filtercel Ceca CBL Dicalite 215 Also types from second precoat

The beer in the filter plant is then recirculated for several minutes to stabilize the filtrate haze level. At the end of this step, the beer should reach a clarity of no more than 0.6 European Brewery Convention (EBC) units. At the end of recirculation, the collection of filtered or bright beer can start.

Quality Parameters of the Filtrate

Microbiology Normal beer before filtration contains about 1–10 million yeast cells per milliliter. Kieselguhr filtration can reduce this figure to a maximum of 5 yeast cells per 100 ml – a reduction by a factor of 2×10^7 to 2×10^8. Bacteria – if present – are reduced by a factor of 10^3–10^4.

Haze The total haze is formed by the optical effects resulting from the presence of yeast cells, proteins, and a lot of other high-molecular-weight substances. In brewing, the haze level is measured in EBC units. A level of 0.6 EBC or less is acceptable; less than about 1 EBC is detectable only by accurate measurement of the portion of a light source scattered at a 90° or 25° angle by the suspended particles.

Oxygen pick-up A low dissolved oxygen content of the packaged product is important to maintain its shelf-life, that is, to avoid changes in its appearance or sensory characteristics within the consumption period. Normally, an oxygen pick-up of less than 0.1 mg l^{-1} beer is acceptable.

Process conditions Only typical ranges of values for the process parameters are given below since beer is a natural product which is available in many varieties. Batch sizes, for example, show variations by a factor of about five.

Filtration speed and batch size The filtration speed (or flux) strongly influences batch size; the higher the speed, the smaller the batch, the volume of which is often measured in hectoliters (hl).

The filtration speed cannot be standardized. Normal values – depending on the filter system and the type of beer – vary from 2.5 to 10 hl m^{-2} h^{-1}. The higher filtration speeds are normally obtained by using vessel filter systems, preferably candle filters or horizontal leaf filters. The traditional plate-and-frame filter is operated at lower filtration rates.

The filter is precoated at 7–10 hl m^{-2} h^{-1}.

Batch sizes normally range from 15 to 70 hl m^{-2} filter area. Large batch sizes can be obtained by the

use of enzymes to degrade β-glucan, thus reducing bodyfeed requirements. β-Glucan in its gel form leads to beers of extremely poor filtrability. The filtrability can be measured by the so-called Esser test, a membrane filtration test.

Temperature The filtration temperature is less than 1 °C, ideally −1 °C. Lower temperatures are better, as so-called chill hazes comprising proteins and other high-molecular-weight substances are precipitated and, therefore, removed by the filtration.

Additional treatment Together with kieselguhr filtration, beer is treated sometimes by the addition of silica gel or polyvinylpolypyrrolidone (PVPP) to improve the colloidal or haze stability. The space consumed within the filter by this additional processing aid leads to smaller batch sizes. Therefore, the PVPP treatment is mostly performed in a second, separate filter unit with a regeneration station for the PVPP.

Filter types The traditional filter for kieselguhr filtration, that is, the plate-and-frame filter, is being replaced more and more by modern vessel filters which can be easily automated and require less manual work.

Vessel filters can be classified by the kind of kieselguhr support: candle filters with vertical, cylindrical filter elements, hanging from a perforated plate into the vessel; horizontal leaf filters with centrifugal sludge discharge; and, uncommonly, vertical leaf filters. Candle filters are preferred, because of their excellent filtration characteristics and the absence of any moving parts.

Performance ranges from 30 to 1000 hl h^{-1} for the horizontal leaf filter and from 100 to 1000 hl h^{-1} for the candle filter.

Kieselguhr Filtration of Wine

For wine filtration, the same principles as for beer filtration are valid. The process parameters differ: the filtration speed is higher, ranging from 10 to 20 hl m^{-2} h^{-1}, due to the better filtrability of wine; the temperature is higher, at up to 20 °C; and different grades of kieselguhr are used. (*See* **Wines**: Production of Table Wines; Production of Sparkling Wines.)

Several stages of filtration can be differentiated in wine production: the first filtration after the main fermentation, a second after the use of finings, and a possible third as a prefiltration stage prior to membrane filtration just before bottling. The kieselguhr types vary from coarse, at the first filtration, to very fine, at the last.

The filter types normally used are the plate-and-frame, the horizontal leaf filter and, for smaller units, the vertical leaf filter. Candle filters are normally not used, because of the numerous types of wine and the risk of mixing taking place upon batch change. The horizontal leaf filter can be emptied and filled again with the new variety; this reduces the extent of mixing.

The performance of typical plants ranges from 10 to 600 hl h^{-1}. The plants are much smaller than beer filters and normally are operated manually.

Sheet Filtration

Sheet filtration is normally used as a second filtration for polishing or obtaining a further reduction of microorganisms. In the wine industry, it is also used for prefiltration. For these purposes, a wide range of filter sheets is available, from a coarse prefiltration to a final filtration with bacterial reduction up to a factor of 10^8. Figure 2 shows an electron micrograph of a filter sheet.

In brewing, filter sheet grades vary only over a narrow range; for polishing (reduction of haze level), for obtaining yeast-free beer and for reduction of bacteria. Tighter filter sheets, with bacterial reduction by a factor of 10^6–10^8, are not used in the brewing industry, mainly for economic reasons; the presence of colloids shortens sheet life.

The process conditions for sheet filtration depend on the type of sheet. A survey is presented in Table 2.

The permeability value for filter sheets is given in liters per square meter per minute (1 m^{-2} min^{-1}), measured at 20 °C and 100 kPa with prefiltered water. Filter sheets with permeabilities lower than 30 l m^{-2} min^{-1} are only used as pharmaceutical grades.

The typical life span of a filter sheet depends on its porosity, the quantity and kind of solids to be removed, and the filtration speed. Typical values for beer filtration vary from 50 to 200 hl m^{-2}.

The filter sheets commercially available consist of cellulose, kieselguhr, and, sometimes, perlite. Asbestos-containing sheets are no longer available, due to the well-known problems caused by the use of asbestos.

The disadvantages of filter sheets include so-called drip losses through the exposed edges of the sheets, especially with high-value products, and the high manual input required in handling.

Recent developments of vessel filter systems, incorporating the use of filter sheet material as the filtration medium, eliminate these problems. Filter sheets formed in so-called lenticular cartridges permit a filtration free from drip losses. The handling of such cartridges is very easy; up to 3.6 m^2 are combined in a single unit, as opposed to an absolute maximum of

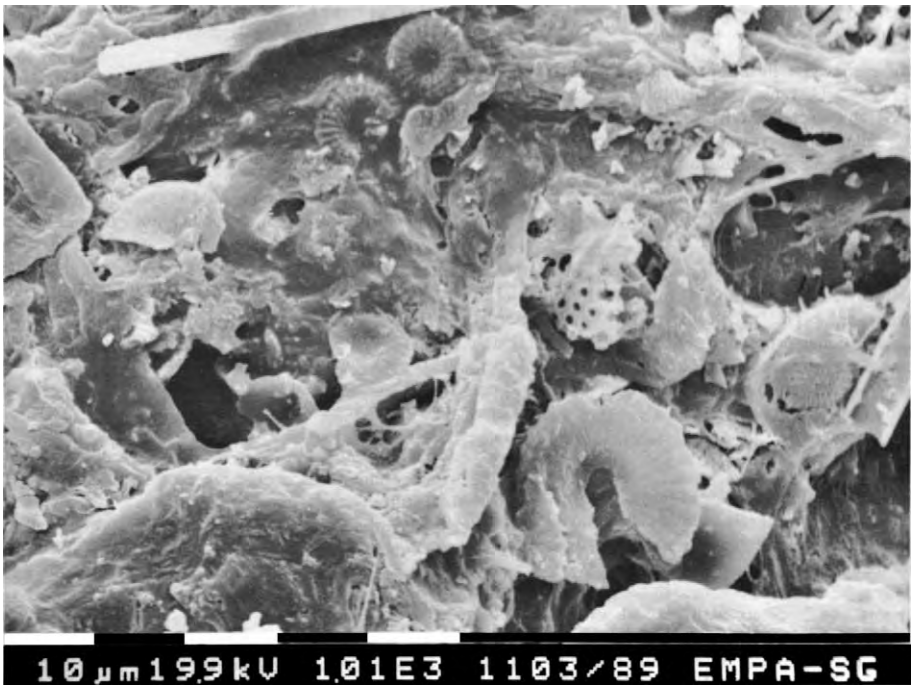

Figure 2 Electron micrograph of a filter sheet.

Table 2 Typical filter sheets

Permeability ($1\,m^{-2}\,min^{-1}$)	*Flux ($hl\,m^{-2}\,h^{-1}$)*	*Use*	*Examples*
300	6	Prefiltration of wine	Seitz K200 Filtrox K30 Carlson XE150
150	2–3	Polishing and filtration of beer	Seitz K150 Filtrox 71H Carlson XE280
130	4	Polishing and filtration of wine	Seitz K150 Filtrox K70 Carlson XE280
80–90	1.5	Removal of yeast from beer	Seitz K100 Filtrox St 110 Carlson XE1200
80–90	3	Removal of yeast from and reduction of bacteria in wine	Seitz K100 Filtrox St 110 Carlson XE1200
30	2	Removal of bacteria from wine	Seitz EK1 Filtrox St 140

$1.4\,m^2$ per individual sheet. Under typical process conditions, it is possible to attain about double the filtration speed of the normal sheet filter.

Final Filtration of Wine

As a last step before bottling, wine can be treated by membrane filtration. The removal of all harmful bacteria requires membranes with a pore size of 0.45 μm. Such a membrane is expensive and therefore requires a very effective prefiltration. This can be achieved by using a lenticular cartridge with its deep-bed filtration medium, or by cartridge filters, constructed from wound cotton, glass, or polypropylene fibers or nonwoven fibrous polymeric media, which may be pleated for optimal filter surface per volume.

Integrity checks of the final membrane are possible.

Cross-Flow Microfiltration

All the techniques described above are based on consumable materials such as kieselguhr, filter sheets, or

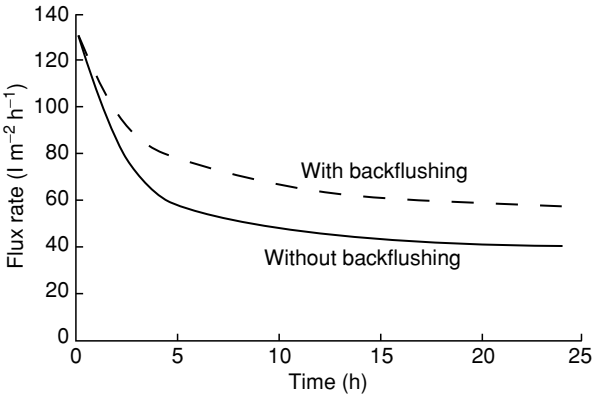

Figure 3 Typical behavior of a cross-flow microfiltration plant used in wine production.

different cartridges. A technique that can combine several filtration stages into a single operation, hence reducing wastage, is cross-flow microfiltration, based on the well-known cross-flow or tangential flow principle. The application of pore sizes of 0.1–1 μm allows an alternative to the systems used until now.

Cross-flow microfiltration systems are based either on capillary or tubular membranes (internal diameter 0.8–4 mm), or flat sheet membranes in conjunction with a special turbulence-promoting mesh. Various polymeric materials are used and ceramic membranes have been tested.

The process conditions should provide a turbulent flow, to prevent concentration polarization and, hence, rapid membrane clogging.

In wine filtration, cross-flow units have found some application (Figure 3). Typical values for the filtrate or permeate flux are 30–80 l m^{-2} h^{-1}, and run times vary from 6 to 30 h. The main problem is the prevention of a rapid flux decay. The reduction in flux can be prevented to an extent by backflushing or other fouling-reducing techniques.

In beer filtration, the cross-flow technique is still not economic due to low process temperatures and the high colloidal fraction of the solids, which results in flux rates of only 20–30 l m^{-2} h^{-1}; and the large volumes of beer involved. Further investigations into the use of membranes will lead to the attainment of better fluxes in the future.

Prospects and Conclusion

The long-term preservation of a food or beverage is a critical requirement of the future consumer. Additionally, chemical or thermal treatments of food are no longer accepted by an increasing number of consumers. For the microbiological stability of beer and wine, filtration in the submicro-meter range is gaining more importance.

Filter media with guaranteed properties, guaranteed solids- or dirt-holding capacities, and microorganism separation characteristics are required as the final step of treatment.

Environmental concern will cause the substitution of waste-producing techniques by others; for example, cross-flow technology and filter media will replace kieselguhr and sheet filtration.

Economic implementation of new techniques will probably arise from the increasing costs of waste disposal, not from any additional advantages in quality, handling or energy consumption. In contrast, the cost savings associated with a reduction in waste will be balanced by a higher energy consumption.

See also: **Beers**: History and Types

Further Reading

Dickenson C (ed.) (1987) *Filters and Filtration Handbook*, 2nd edn. Morden: Trade and Technical Press.

Kiefer J (1990) Kieselgurfiltration – Anwendungsgrundlagen für den Kesselfilter. *Brauwelt* 44(130): 1964–1970.

Matteson MJ and Orr C (eds) (1987) *Filtration – Principles and Practices*, 2nd edn, pp. 133–162. New York: Marcel Dekker.

Rushton A, Ward AS and Holdich RG (2000) *Solid–liquid Filtration and Separation Technology*, 2nd edn. New York: Wiley-VCH.

Scheuermann EA (1989) Klärfiltration mit Filterschichten. *Filtrieren und Separieren* 2(3): 69–76.

FISH

Contents

Introduction

Q Bone, The Marine Laboratory, Plymouth, UK

Kinds of Fish used as Food

An extraordinary diversity of different kinds of fish and fish products are used as food, from jawless lampreys to lungfish (and, at times, even coelacanths), but the most abundant kind of fish, the bony ray-finned teleosts (like carp and cod), make up the great majority of fish used directly as food. The teleosts also make up the greater part of fish used as fish meal, which we eat indirectly after it has been used as animal feed or plant fertilizer, for most fish meal is made from small pelagic teleosts. Some 20 000 teleosts are known presently, but new species are still being described from both marine and freshwaters, and the final total may well exceed the present estimate of 30 000 (with around 12 000 from freshwater). Many of this grand total are naturally too inaccessible, too rare, or otherwise unsuitable for use as food, but most of the different teleost families contain food fish species, and some families like the Clupeidae, Gadidae, and Salmonidae are of special importance. In contrast, the other major fish group, the sharks and rays (elasmobranchs), has only some 800 species, almost all entirely marine. Relatively few are of significant importance as food, though flesh from small sharks (dogfish) is much used in the UK as rock salmon, whilst other elasmobranchs provide liver oils. Apart from the teleosts and elasmobranchs, the only other food fish important for their flesh (and, in particular, their eggs) are the ray-finned sturgeons, one of the small number of surviving Chondrostean fishes.

Elasmobranchs and Teleosts

The general structural features of the two main groups of food fish are shown in Figure 1. The elasmobranch body is covered with hard denticles, often variously ridged or pointed; in the jaws, these form formidable successions of teeth. In contrast, teleosts are covered with bony scales (sometimes lost or minute, as in eels), and these too may be variously sculptured and ridged, although they are usually more or less circular. In both groups, the general body plan is the same, with two sets of paired fins as well as dorsal and ventral unpaired fins. The paired fins are attached to the pectoral and pelvic girdles, the former linked to the central vertebral column. On either side, serial myotomal muscle blocks (the main edible portion of most) are separated by tough connective tissue myosepta, which attach to the vertebral column and to the connective tissue layer underlying the skin.

However, the design of the fins and axial skeleton is different in the two groups: in sharks, the skeleton is cartilaginous (strengthened by granular or prismatic calcifications where necessary); the vertebral column is simpler than in teleosts, and the ribs are small. The unpaired fins are attached to basal cartilages. Shark fins are supported by jointed cartilaginous elements at their bases and by thin elastoidin rods (ceratotrichia) along their length, which stiffen the outer border of the fin. Lacking the bony rays of teleost fins, the supporting elements of shark fins

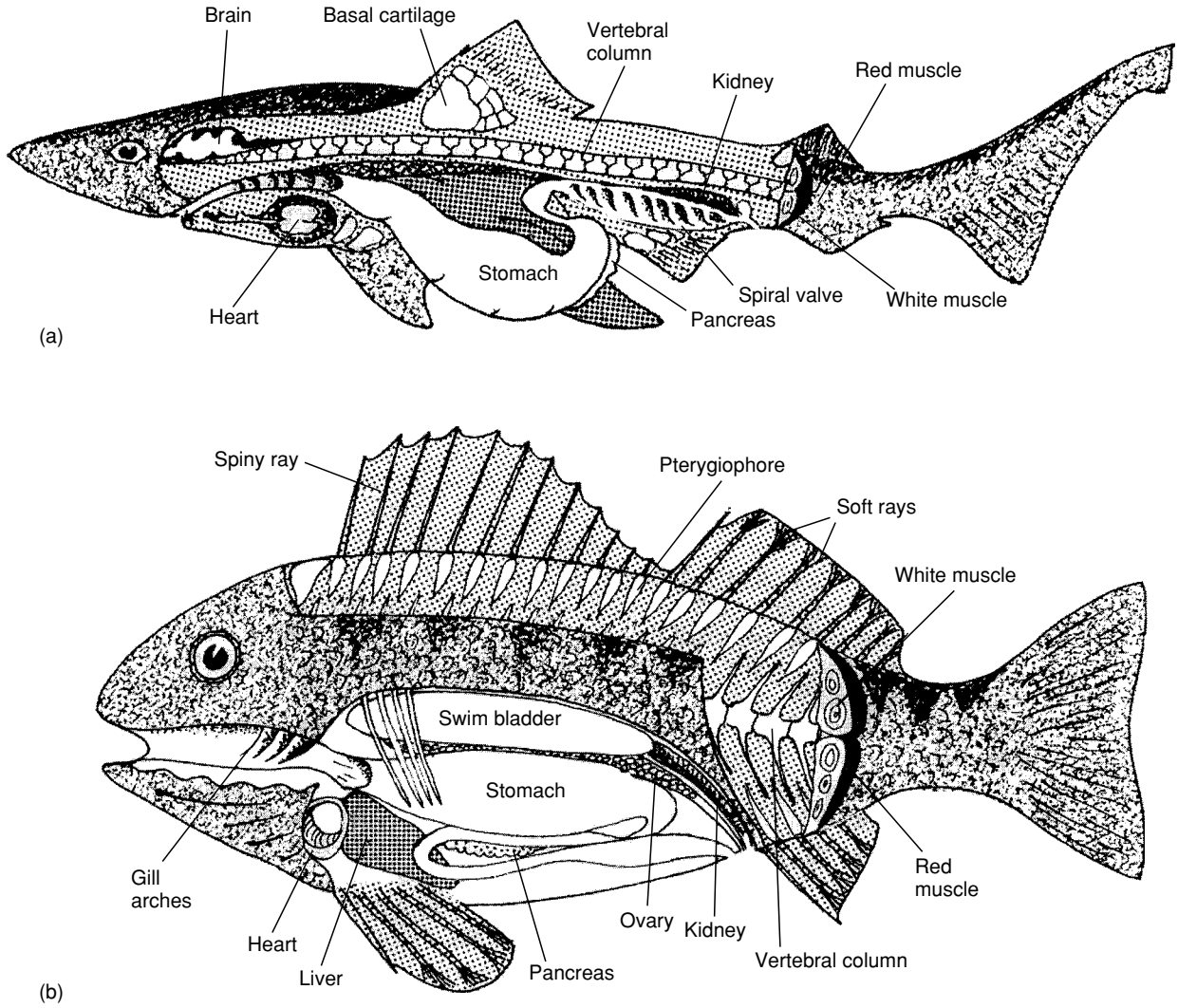

Figure 1 General features of (a) a shark and (b) a teleost fish. Note the absence of a swim bladder and less prominent skeletal elements in the shark, and the presence of a large liver and spiral valve. In both, the myotomal muscle consists of an outer layer of dark flesh and an inner white layer (see text). (a) After Lagler KF, Bardach JE, Miller RR and Passino DRM (1977) *Ichthyology*, 2nd edn, pp. 506. Chichester, UK: Wiley. (b) After Dean B (1895) Fishes living and fossil. An outline of their forms and probable relationships. In: *Columbia University Biology Series*, vol. 3. London and New York: Macmillan and Co.

can be sufficiently softened by cooking to be eaten as an expensive delicacy. They are much less flexible than teleost fins and, unlike the latter cannot be furled and used for delicate fanning and flapping movements. The pectoral fins act in almost all sharks as lifting foils, like aeroplane wings, and they can adjust their angle of attack and be swung back and forth to some extent; the other fins of sharks can only bend stiffly from their bases. Rays differ from sharks because their pectoral fins (which they use for swimming) are greatly enlarged and supported by jointed cartilages to their tips, so that they are much more flexible and can make delicate flapping movements. Naturally, the fins are highly muscular; hence, it is the pectoral muscles rather than the reduced myotomal muscles that are used as food, and they are easily separated from the supporting cartilages when cooked.

In teleosts, the skeleton is much more prominent than in elasmobranchs: the vertebrae have neural and hemal spines; in addition, the unpaired fins are attached to bony, spinous pterygiophores, and there may also be intramuscular bones in the myosepta, as in clupeids such as herring (*Clupea harengus*) and the shads (*Alosa* spp.), which are irritatingly obvious when they are eaten. Teleost fins are very different to those of elasmobranchs, for they are supported by flexible, jointed rays (lepidotrichia) controlled by

muscles at their bases moving individual rays, and hence can be furled or make most delicate movements (e.g., in seahorses or knifefishes).

Internally, the chief differences between teleosts and elasmobranchs are as follows: many teleosts have a gas-filled swim bladder, used for buoyancy (and also in some for sound production and reception, and in some freshwater species for respiration); the gills are of a slightly different design; all elasmobranchs and only a few teleosts have a spiral valve in the intestine, and the reproductive apparatus is different. Only a very few teleosts (e.g., damsel fishes, Pomacentridae) have internal fertilization (universal in elasmobranchs) and, except in bottom-dwelling species, store lipid in the large liver for buoyancy.

A notable physiological difference between the two is that elasmobranchs are close to osmotic equilibrium with seawater by virtue of high concentrations of urea and trimethylamine oxide (TMAO) in the body fluids, whereas in marine teleosts, the body fluids are much more dilute than seawater, and hence they lose water and gain ions across the gills and other permeable surfaces. In fresh water, of course, the reverse occurs. The high urea and TMA0 content of elasmobranchs naturally has important consequences for spoilage.

Fish as Food

Both adult and larval teleosts form the basis of special fisheries as do the eggs of certain species (e.g., sturgeons), but elasmobranch fisheries concern adults only. Even in adults, whilst myotomal muscular tissue is the chief food source, other nonmuscular tissues, such as fins, gonads, and livers, may be important foods (of economic, if not always of much nutritional, importance, e.g., shark fins). In some instances where the bones are eaten (e.g., sardines and whitebait, and, in Japan, the roasted vertebral columns of certain larger species), these are important sources of dietary calcium and phosphorus. Fish are also a valuable source of the vitamins nicotinic acid, riboflavin, and vitamin B_{12}. Refer to individual nutrients.

Above all, however, fish are important as an excellent protein source, for muscular tissue makes up a larger part of the fish body than in terrestrial vertebrates, from 37 to 68% of the total bodyweight. What is more, in all food fish except rays, the locomotor musculature is arranged in serial myotomal blocks along the body so that, at least in larger fish, it is an easy matter to gut the fish and separate the flesh from the skeleton as fillets. The relatively larger muscular mass in fish, and their importance as a protein source, is directly related to their life in water. Because most fish are well streamlined, the major drag force opposing their forward motion is skin friction drag, proportional to the square of forward speed. Thus, the power output required from the musculature is (approximately) proportional to the cube of swimming speed. Even modest swimming speeds require a large mass of locomotor muscle, but only a small weight penalty is incurred by the large amounts of (edible) musculature needed for locomotion, since the water in which fish live is not much less dense than the musculature. Even those fish that do not reduce their density (or achieve neutral buoyancy) by storing gas or fat for static lift only weigh in water around 5–7% of their weight in air.

The locomotor muscle is divided in almost all fish into a lateral red component, used during cruising, and a much larger white or pale yellow mass, which is active during short bursts of higher speed. A similar division is seen in the pectoral muscles of skates and rays. The two types of muscle have a different enzyme spectrum and metabolism, and also differ in lipid content and vascular bed. In yellowfin tuna (*Thunnus albacares*), for example, the red muscle contains twice the amount of lipid than the white muscle, as well as a large amount of myoglobin absent from the white muscle. The functional division into two muscle types of different composition (most simply distinguished by color) is sometimes reflected in the use of the muscles as food. For example, the internalized red muscle mass of the porbeagle shark (*Lamna cornubica*) is regarded as a delicacy when smoked, whereas the larger white portion of the myotomes is cooked and eaten unsmoked.

Raw fish muscle contains around 20% protein (the contractile lattice of actin and myosin and various sarcoplasmic proteins) so that protein forms the major energetic component of the fish, as shown in Table 1.

Table 1 Protein energy as percentage of total energy of some food fish and meat

Food	*Protein energy (%)*
Cod fillet (*Gadus callarias*)	95
Albacore (*Seriola lalandi*)	90
Haddock (*Melanogrammus aeglefinus*)	96
Sole (*Solea vulgaris*)	85
Flounder (*Pleuronectes flesus*)	95
Halibut (*Hippoglossus hippoglossus*)	66
Herring (*Clupea harengus*), an oily fish	41.5
Chicken (white meat)	76
Lean beef	69
Medium-fat beef (cooked)	27.4
Medium-fat pork (cooked)	13.9

From Geiger E and Borgstrom G (1962) Fish protein – nutritive aspects. In: Borgstrom G (ed.) *Fish as Food*, vol. 2, p. 128. London: Academic Press.

In comparison, even lean beef has relatively less protein energy. It is the high protein content of fish flesh that permits the preparation of increasingly important surimi products (*See* **Fish**: Miscellaneous Fish Products). Some fish also offer a useful lipid source, because, as well as the liver oils and lipid in ovaries, there can be much lipid in the muscles themselves, especially in scombroids and clupeids. In mackerel (*Scomber*), adipose cells between the muscle fibers and lipid droplets within the fibers make up 17% of the cross-sectional area of the lateral strip of red muscle. Unsurprisingly, fresh mackerel fillets contain some 16% lipid (protein: 19%), so that lipids make up the greater part of the total energy. As important as content is the nature of the lipid, for fish lipids contain ω-3 polyunsaturated fatty acids, shown in a number of studies to confer significant benefits in the prevention of cardiovascular diseases. Some fish contain lipids that are unsuitable for human consumption; for example, the castor oil fish, *Ruvettus*, gains its name from the purgative qualities of the low-density oil it stores for buoyancy purposes, and since it is abundant in oil sacs within the skeleton, naïve consumers are warned not to suck the bones. (*See* **Fatty acids**: Dietary Importance.)

Both fisheries and aquaculture are crucial for providing protein in many communities, some 5% of all protein eaten per capita in low-income countries. In addition, fish protein fed to animals provides indirectly for further human protein food. It is hardly surprising that fish form an important part of the diet in many countries, and that as well as being hunted with ever more sophisticated equipment, fish are farmed on an increasing scale. The annual per capita consumption ranges from over 20 kg to less than 5 kg, being highest in countries with a coastline facing a productive sea, such as Iceland or Japan, and lowest in such countries as Switzerland and Turkey. The average annual per capita consumption in 1996 was some 13.3 kg (excluding China). The most recent worldwide statistics available (2000) show that fish continue to increase in importance as a food source.

Of these different fish, the overwhelming majority are teleosts, as they are in numbers of species. The total elasmobranch catch is less than 1% of the marine fish caught.

History of Fish as Food

Fish have been used as food from ancient times, and quite a wide variety of fish skeletal remains are known from middens and spoil heaps, including both freshwater and marine species. The presence of marine fish remains in inland Late Old Stone Age refuse heaps (around 40 000 BC) in the Dordogne, France, implies that, even at that date, some means of preservation (possibly drying over smoky fires) must have been used to counter the rapid perishability of fish flesh. Air-drying, smoking, salting, and pickling in brine were probably developed around 4000 BC and permitted the wider availability and use of fish as food. The preservation treatments used varied according to the kind of fish: air-drying and salting were suited to leaner fish, whilst oily fish, such as herring or salmonids, had to be covered in brine to prevent rancidity owing to lipid oxidation. Today, more rapid transport of fresh fish and improved techniques of chilling and freezing are of much greater importance than these other means of preserving fish flesh, although curing in various ways may be important in increasing monetary value, as manifested, for example, in the price difference between fresh herring and kippers. (*See* **Preservation of Food**.)

Preparation

Different storage, preparation, and cooking methods may profoundly alter the value of fish as food, since they greatly influence important constituents of fish flesh, such as lipid and vitamin content. These effects will not be considered here. (*See* **Fish**: Processing.)

Problems in using Fish as Food

Ensuring the safety of seafood (fish and fish products and shellfish) requires government regulation and expense, but several Hazard Analysis Critical Control Point programs are now being put in place to minimize and avoid illness and disease caused by seafood, and consequent difficulties in import and export of seafood products. Two special cases of the problems with fish as food are where some or all of the fish may be dangerously poisonous. The puffer fishes (Tetraodontidae) contain high concentrations of the Na^+-channel blocker, tetrodotoxin (TTX), which is an exceedingly powerful nerve- and muscle-blocking agent. Such fish (a delicacy in Japan) are carefully prepared by licensed chefs to remove the tissues containing the highest concentrations of TTX before consumption in special Fugu restaurants; nevertheless, between 1987 and 1996, there were 32 deaths from eating puffer fish. Connoisseurs apparently require the initial symptoms of low doses of TTX (tingling of the lips) to enjoy the delicacy. The second special case is that of tropical marine fish. Herbivores may accumulate ciguatera toxin, apparently of algal origin, which also passes to carnivores such as barracuda, *Sphyraena*, when they eat the contaminated herbivores. The liver and viscera are particularly liable to be poisonous, and as yet, there is no reliable

test to distinguish fish with ciguatera toxin (apart from feeding portions to animals such as kittens).

Future Trends

Despite overexploitation of many fish stocks and consequent (if belated) control measures, it is clear that fish will continue to play an important role as food, chiefly as a consequence of improved methods of farming both of freshwater and marine species. A better understanding of fish stock management will be important in keeping capture fisheries at around present levels, which are probably not far from the maximum sustainable. In contrast, aquaculture (which provided 28.8 million tonnes in 1997) will increase steadily. In 1996, over a quarter of the world supply of food fish for human consumption came from aquaculture (mainly Chinese carp and other cyprinids). It also seems likely that there will be an increasing amount of fish and fish products entering international trade, particularly from developing countries, where exports almost doubled in value between 1990 and 1995. Future trends here seem likely to show different methods of keeping fish free from spoilage such as by the use of controlled atmosphere packaging using mixtures of N_2, O_2, and CO_2, and a large increase in the use of such products as surimi. An entirely different reason for the future importance of fish as food arises as the result of dietary studies, which have either demonstrated or suggested the value of fish diets in preventing some common diseases in developed countries. For example, fish are generally low in cholesterol and, compared to meat, have a much more complex spectrum of lipids high in polyunsaturated fatty acids. Of particular importance is the high proportion of the ω-3 series of fatty acids in oily fish, contrasting with the ω-6 fatty acids of vegetable origin. Epidemiological studies indicate that significant reductions in coronary heart disease may be obtained by the weekly consumption of only 300 g of such fish. There are also possibilities of modifying the diets of farmed fish to augment the levels of other components valuable for the human diet. (*See* **Fish Farming**.)

See also: **Fatty Acids**: Dietary Importance; **Fish**: Processing; Miscellaneous Fish Products; **Fish Farming**; **Preservation of Food**

Further Reading

Cato JC (1998) *Economic Values associated with Seafood Safety and Implementation of Seafood Hazard Analysis Critical Control Point (HACCP) programmes*. FAO *Fisheries Technical Paper*, Rome: FAO.

Cutting CL (1962) Historical aspects of fish. In: Borgstrom G (ed.) *Fish as Food*, vol. 2, pp. 128. London: Academic Press.

Dean B (1895) Fishes living and fossil. An outline of their forms and probable relationships. In: *Columbia University Biology Series*, vol. 3. London and New York: Macmillan and Co.

Geiger E and Borgstrom G (1962) Fish protein – nutritive aspects. In: Borgstrom G (ed.) *Fish as Food*, vol. 2, pp. 29–114. London: Academic Press.

Hall GM (ed.) (1997) *Fish Processing Technology*, 2nd edn. London: Blackie Academic & Professional.

Halstead BW (1978) *Poisonous and Venomous Marine Animals of the World*, p. 1043. Princeton, NJ: Darwin Press.

Lagler KF, Bardach JE, Miller RR and Passino DRM (1977) *Ichthyology*, 2nd edn, pp. 506. Chichester, UK: Wiley.

Catching and Handling

P Chattopadhyay, Jadavpur University, Calcutta, India

Introduction

The devices used to catch fish are known as gear, i.e., hooks, baits, nets, and rope. Fishing gear of numerous designs is available for use in commercial or sport fishing. Many factors have to be taken into consideration when selecting the method and gear to be used to catch a particular species of fish in a specific area. The choice depends on the species being fished, value of the individual species to the fisherman, the depth of water, and the characteristics of the sea bed. The different species of fish have different habits, movements, and reaction to stimuli. Demersal species are usually found near the sea bed, whereas pelagic species are found anywhere between the sea bed and surface. The movements and habits of each species are controlled by water temperature, salinity, spawning habits, migration, available food resources, and the temperature barriers that exist in any particular location. A few methods of catching fish are discussed below.

Net Methods

Trawling

A fishing method in which a boat is dragging a large net through the water is called trawling. When the net is somewhere in the water column, it is called midwater trawling, but when the net is dragged along the

ocean floor it is known as bottom trawling. Many important species are caught with the bottom trawl. Large 'doors' are used ahead of the net to keep the net down yet opened wide (Figure 1). There are many designs of these doors. The most common are otter doors. Fuel consumption can be reduced by cutting holes into the doors in appropriate places. Trawling therefore requires a strong boat to pull the fishing gear through the water as well as the fish. It disturbs the bottom of the ocean but is still an efficient method of catching large quantities of bottom fish. The doors used for midwater trawls are much lighter than those used for bottom trawling.

If two boats are used, each dragging one side of the net, the same opening of the net could be achieved without using any doors (Figure 2). For this, cooperation between two boats is necessary. Pair trawling is a fuel-efficient method. The ability of smaller fish to escape is controlled by the size of the net mesh. However, as the bag fills up with fish, and the larger fish block the net holes, small fish become trapped deep inside the net and are unable to escape. Regulating the mesh size of the net may be partially effective in minimizing the catch of undersized fish. Recent research suggests that the shape of the net mesh may affect its efficiency: a square mesh may allow for better escape of undersized fish than the traditional diamond mesh.

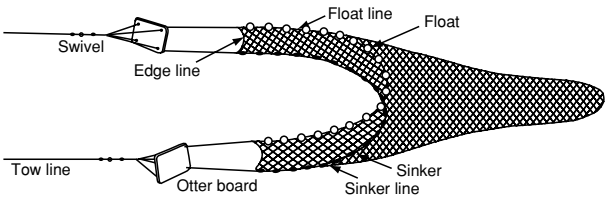

Figure 1 An otter trawl net. Reproduced from Yamaha Motors (1986) *Yamaha Fishery Journal Composite*, p. 134, with permission.

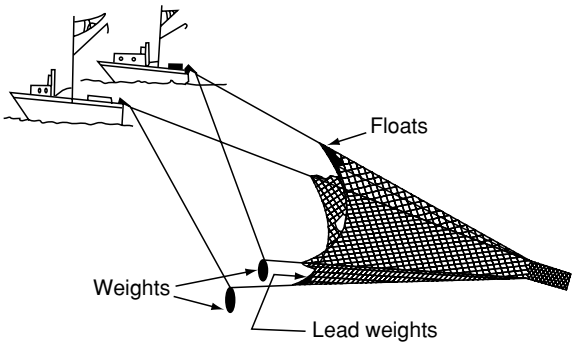

Figure 2 Two-boat midwater trawl.

Table 1 shows some data on the relationship between fish length and weight. Table 2 relates the mesh size of a net to the length of fish retained. The lengths given represent the size at which about 50% of the fish are retained and 50% can still escape through the net. The data can also be expressed for a particular fish (Table 3).

The popularity of trawling has been expanded with the development of factory ships with facilities for on-board freezing.

Seining

Seine nets are often used to catch schooling fish (Figure 3). Pelagic fish, for example, mackerel, are most easily caught in this way. These are upper-water-column fish and swim large distances. Small boats, sometimes two boats, set off from the mother ship and carry a relatively small mesh net which is set around the school of fish. If the net is a 'purse' seine, a string on the bottom is then pulled tight so that the bottom of the net can be closed off. The net is then hauled in until most of the fish are in a small area of water. This is called hardening of the net. The fish are then removed from the water, with hand nets or with

Table 1 Optimizing the catch by regulating mesh size[a]: length–weight data for some common fish

Fish	*Weight (kg)*		
	0.45	*0.9*	*1.36*
	Length (cm)		
Pollock		51	
Cod	36	48	64
Haddock	38	46	66
Whiting	38	51	
Red fish	31	38	51
Mackerel	31	41	
Yellow tail	38	46	
Rainbow trout	36		

[a]Mesh size refers to the main part of the net.
Data courtesy of the National Marine Fisheries Service.

Table 2 Mesh size (cm) versus retention length (cm) for some common fish

	Mesh size (cm)				
	10	*13*	*14*	*16*	*17*
	Retention length (50% retained)				
Cod	34	45	47	49	56
Haddock	33	43	45	47	53
Yellow tail flounder	22	29	31	32	36
Pollock	34	44	46	48	54
Winter flounder	21	28	29	31	35
Dals (plaice)	24	31	33	34	39

Data courtesy of the National Marine Fisheries Service.

Table 3 Relative rates of retention for cod

Length (inches)	*Percent retained*	
	14 cm mesh	*15 cm mesh*
27	99	97
25	97	98
22	78	56
21.6	74	50
20	53	29
19.8	50	27
18	27	11
16	9	3

Data courtesy of the National Marine Fisheries Service.

mechanical unloaders (pumps). Purse seining is used for the capture of species utilized in bulk reduction processes such as fish meal or species of high value. The greatest damage is done to the fish at the time of unloading from the boat because of the postmortem physicological changes that have occured in the fish. Fish in rigor should not be pumped. A modification of the seining method is practiced from the shore. The mother ship becomes a land-side attachment and the net, after setting, is hauled to shore by either people or a vehicle. This method is often used for more valuable species such as blue fish, striped bass (in New York), or for salmon (in Scotland).

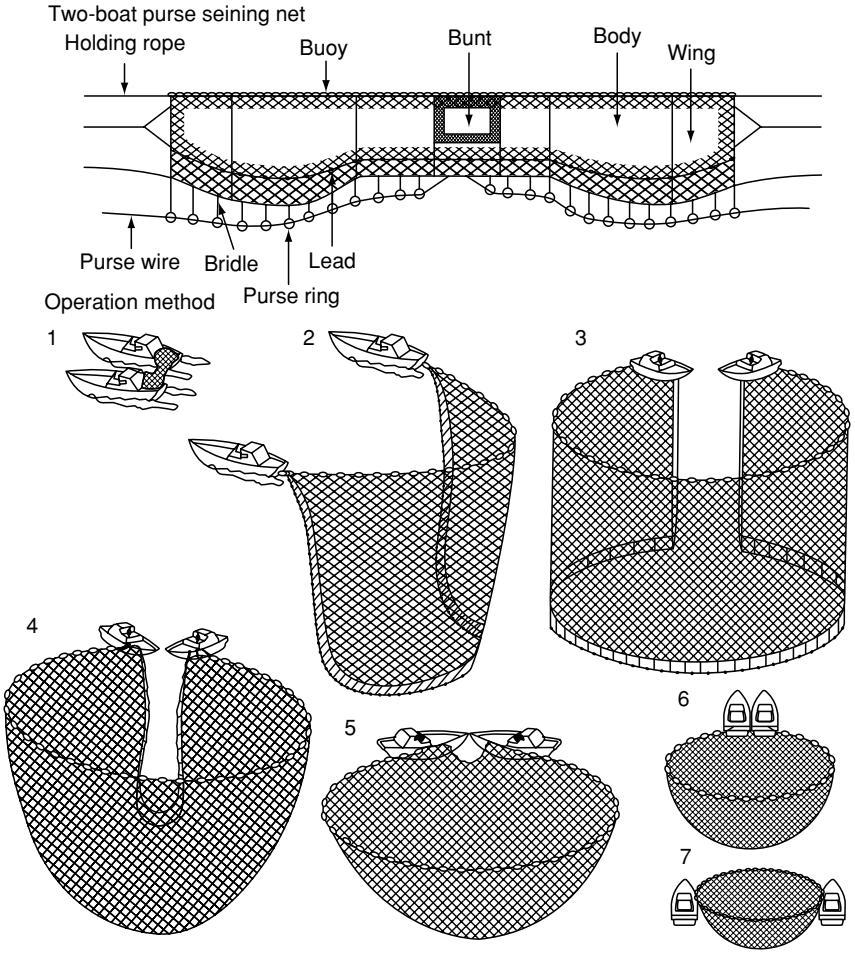

1. Sharing the load of a net and moving forward with their bows linked.
2. Separating from each other and casting a net where a school of fish is located.
3. Surrounding the school.
4. Pursing begins.
5. and 6. Net hauling.
7. The bunt is positioned between two boats and fish is scooped with a big dip net.

Figure 3 Seine nets. Reproduced from Yamaha Motors (1986) *Yamaha Fishery Journal Composite*, p. 178, with permission.

Gillnetting

Gillnetting involves stretching out a net to intercept fish (Figure 4). The net can be set at any level in the water column by manipulating the rope length and by the use of anchors and floats. Placement is often chosen to intercept natural fish migration. The fish do not see the net and simply swim into it. Once their gills go through the net, they can no longer escape, and they drown. The thrashing of fish prior to death leads to both external and internal bruising. If the nets are not well attended, the dead fish will be untreated for too long and will be of poor quality. For this reason, gillnetting is one of the least desirable ways to catch fish from a quality standpoint, even though it is extremely efficient. It is an easy method to use with small boats, for example, for salmon in Alaska. The nets are generally hauled from one side of the boat and immediately sent back to fish on the other side, or a drum is used to wrap the net around as fish are simultaneously removed by hand.

Gillnetting has become controversial because of 'ghost fishing'. With modern net materials like nylon, these nets are difficult for fish to see and may continue to catch fish after the fisherman has lost them.

Cast Nets

Weighted cast nets or variations thereof are usually used by the artisan fisherman. These are simply thrown out to sea from land or a small boat; the weight around the net causes them to sink and trap (catch) the fish.

Herding devices

Another net-based catch method involves the use of traps, and weirs (Figure 5). These devices herd fish into an area where they can then be easily removed. This fixed gear is expensive and time-consuming to build, and also depends on the fish coming to the trap rather than a boat going to the fish. Thus, traps work best when the migration of the desired fish species can be predicted accurately. They often work best in rivers or streams or close to the shore where natural terrain features channel fish movement. The fish can be caught live simply by hand-netting them out of the

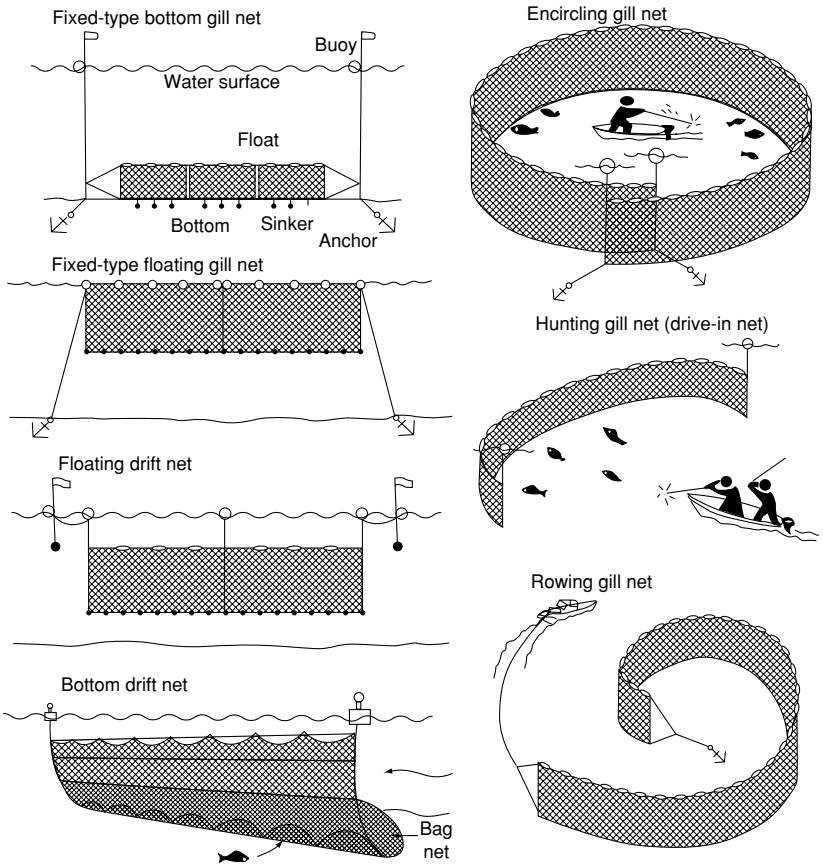

Figure 4 Examples of gill and drift nets. Reproduced from Yamaha Motors (1986) *Yamaha Fishery Journal Composite*, p. 50, with permission.

trap. Otherwise fish can be kept alive in the traps for some time and harvested when needed. Large pound nets are used along the shore to collect and harvest these fish.

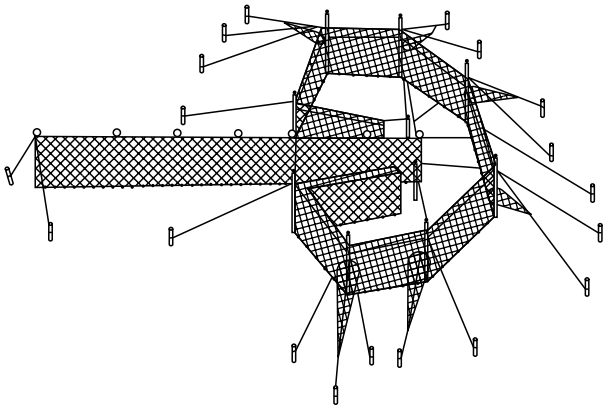

Figure 5 A small fixed net. Reproduced from Yamaha Motors (1986) *Yamaha Fishery Journal Composite*, p. 67, with permission.

Hooking Methods

Longlining

Longlining is simply the use of a long fishing line with lots of hooks, usually evenly placed (e.g. every 1 m) along the line (Figure 6). This method can be used on rougher grounds (bottoms) where a trawler cannot be used. This system has many important benefits over trawling, including the potential for size and species selectivity by the proper choice of hook size and shape, and bait; lower energy cost; and less damage to the ocean bottom.

Placing the bait on to numerous hooks is very labor-intensive. New forms of mechanical baiters have been developed that replace the tremendous effort involved in hand baiting. It is also labor-intensive to remove the bait and fish from the line as it is hauled back on board. Equipment to do this automatically has been developed. If longlining is done properly, the fish can be processed on board one at a time as they are brought on board.

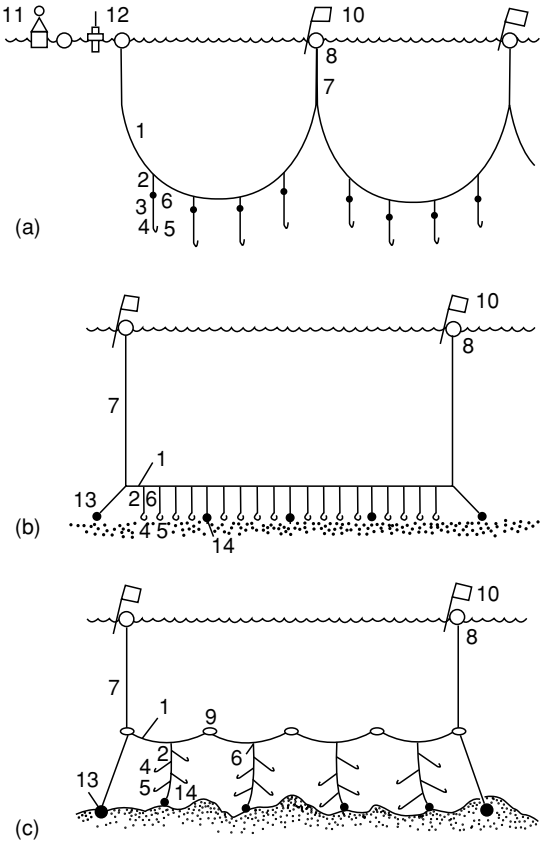

1. Trunk line 2. Branch line 3. Sekiyama 4. Snood 5. Hook
6. Swivel 7. Buoy line 8. Float 9. Antipressure float
10. Marker flag 11. Marker light 12. Radio buoy 13. Anchor

Figure 6 Examples of longlines. (a) Drift longline; (b) bottom longline; (c) vertical line bottom longline. Reproduced from Yamaha Motors (1986) *Yamaha Fishery Journal Composite*, p. 219, with permission.

The hooks used for longlining must be of the appropriate size for the fish that is to be caught. The choices of hook size and bait make this form of fishing very selective (Figures 7 and 8). Since each hook is designed to catch fish above a certain size, this technique is particularly effective in avoiding undersized fish. Hook shape may also affect species selectivity as well as the amount of bait needed (a major expense with longlining). The Mustad Company of Norway has developed the most advanced longline systems appearing on the market.

Trolling

Trolling is usually done close to the water surface. The boat moves through the water with a number of lines each attached to a pole (Figure 9). A number of hooks are attached to each line. The fish are almost always landed live. Trolling allows for selective handling of fish and thus leads to a very high quality. Troll-caught salmon are particularly prized, especially if they are to be smoked subsequently. The absence of either internal or external bruising maximizes the appearance and the recovery (weight) of this extremely expensive product. It must be noted that subsequent careful handling is important to preserve the higher initial quality.

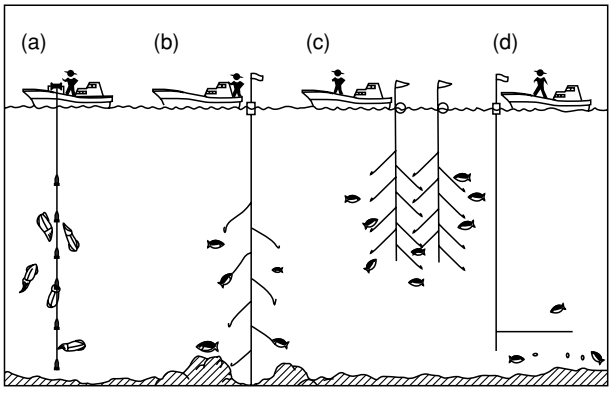

Figure 8 Use of vertical lines. Reproduced from Yamaha Motors (1986) *Yamaha Fishery Journal Composite*, p. 211.

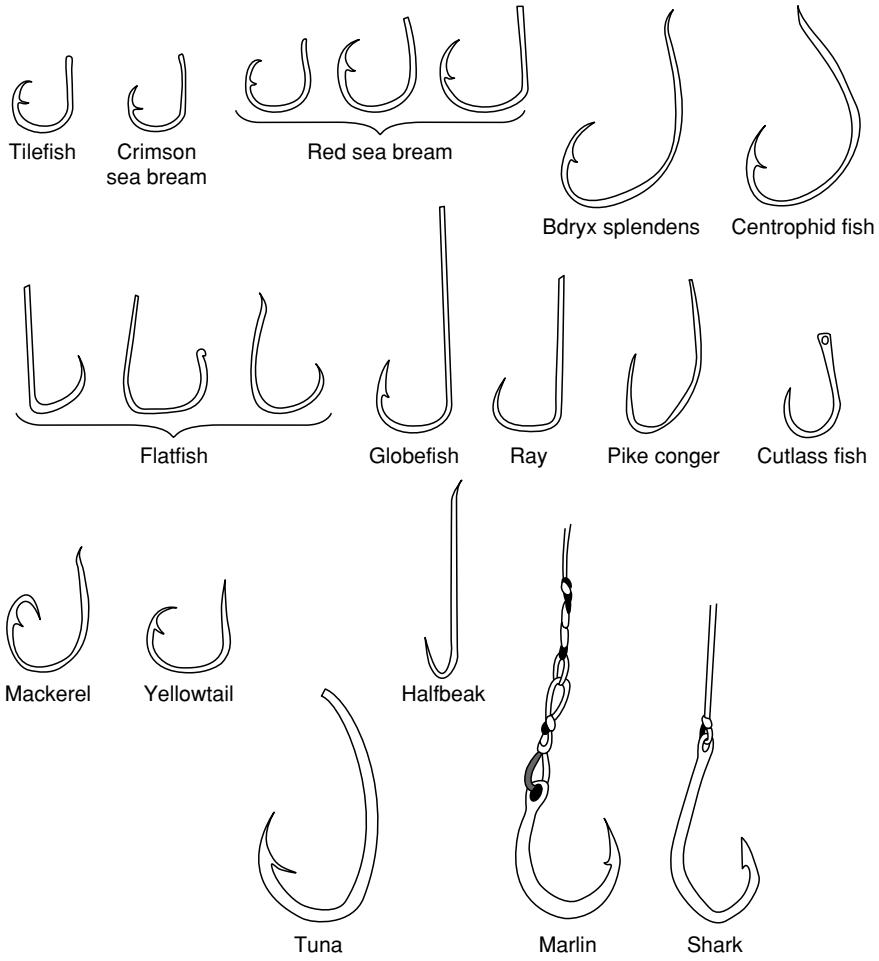

Figure 7 Various fishing hook shapes. Reproduced from Yamaha Motors (1986) *Yamaha Fishery Journal Composite*, p. 220.

Jigging

Jigging was first developed commercially in Japan. When fish are caught by this method, bait and/or special lights (at night) are used to attract the fish to an area (Figure 10). The fish tend to bite on any flashing item including the many hooks present in the water. Some part of the catch may also be snagged on the hooks. The system of powerful lights on the boat coupled with brightly colored jigs and small in-the-water lights is particularly attractive to certain species of squid. Bait is an important cost in all the hooking methods, particularly longlining. Each longline hook uses about 45 g of bait. Therefore, there is an interest to develop appropriate baits, i.e., those that can withstand the physical constrains of the mechanical system and also attract the appropriate fish. Research is being done currently on taste preferences of fish, i.e., flavor attractants. This will help to develop appropriate artificial baits.

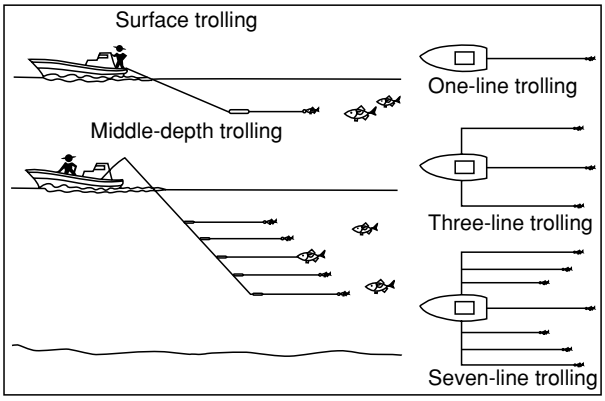

Figure 9 Examples of trolling. Reproduced from Yamaha Motors (1986) *Yamaha Fishery Journal Composite*, p. 211.

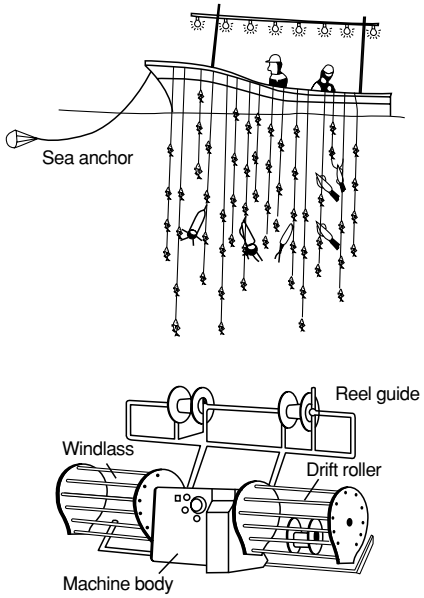

Figure 10 Automatic squid-fishing jigs. Reproduced from Yamaha Motors (1986) *Yamaha Fishery Journal Composite*, p. 24.

Other Techniques

A number of more unusual techniques may also be used. Spears and harpoons may be used for some of the larger fish. Fishermen sometimes simply detonate a bomb in the water and then collect all the dead fish – usually of low quality. A wide variety of other catch methods are used for shellfish. These animals often live below the ocean floor and thus require techniques like surface dredging (Figures 11 and 12) or raking and tonging to recover the product (Figures 13 and 14). A dredge has steel teeth that scrape and stir up the ocean floor. The scallops are then caught in a large steel bag behind the dredge. Tongs are really a pair of forks with a center pivot that are used to pick up clams or oysters from the ocean floor. Classical rod

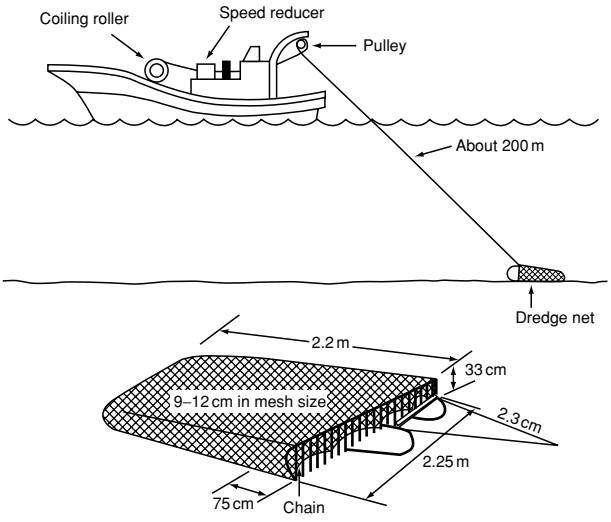

Figure 11 Operation of a dredge. Reproduced from Yamaha Motors (1986) *Yamaha Fishery Journal Composite*, p. 75.

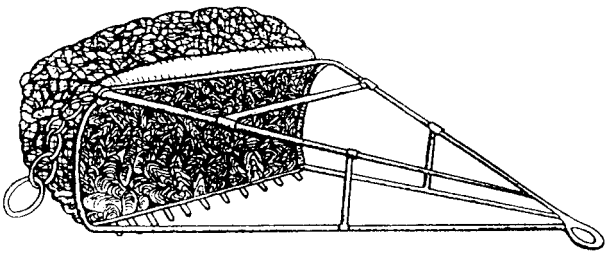

Figure 12 An oyster dredge. Reproduced from Wheaton FW and Lawson TB (1985) *Processing Aquatic Food Products*. New York: John Wiley, p. 95, with permission.

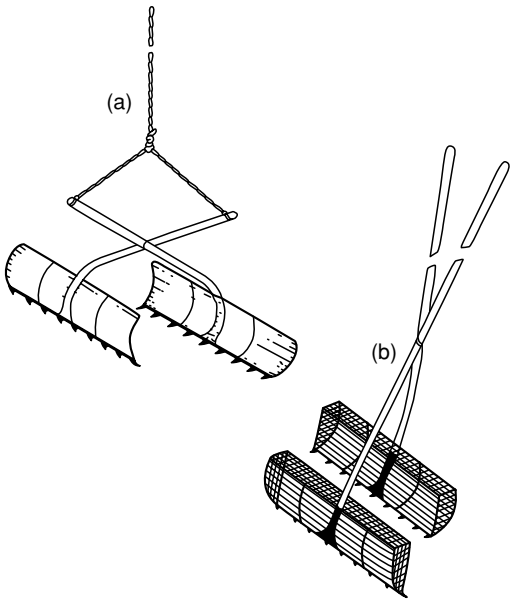

Figure 13 Oyster tongs. (a) Patent tongs; (b) hand tongs. Reproduced from Wheaton FW and Lawson TB (1985) *Processing Aquatic Food Products*. New York: John Wiley, p. 100, with permission.

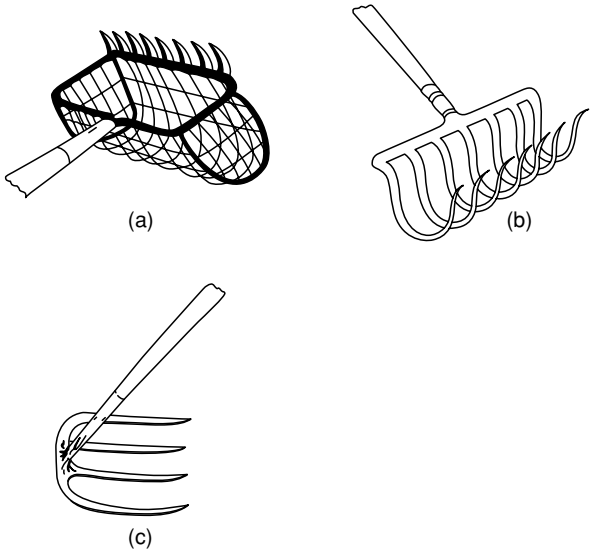

Figure 14 Clam rakes. (a/b) Hard clam rakes. (c) Soft shell clam hoe or fork. Reproduced from Wheaton FW and Lawson TB (1985) *Processing Aquatic Food Products*. New York: John Wiley, p. 101, with permission.

and reel fishing can be used commercially for larger, more valuable fish like tuna and swordfish.

There are many other methods used to catch fish as well as a variety of other types of equipment, such as pots and traps for shellfish (Figure 15). Most of these methods are adopted to the specific fish and local fishing conditions.

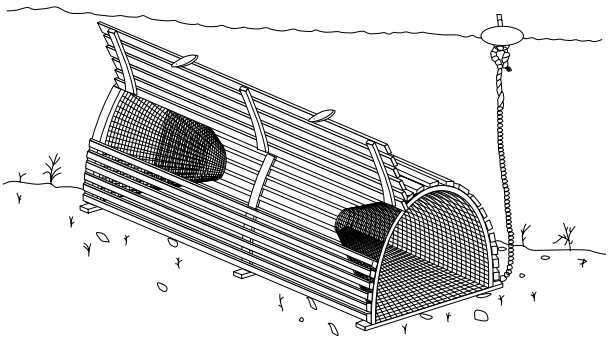

Figure 15 A New England loster pot. Reproduced from Wheaton FW and Lawson TB (1985) *Processing Aquatic Food Products*. New York: John Wiley, p. 85, with permission.

Traps are often used in areas through which fish regularly move or congregate. Traps are set in relatively shallow water adjacent to the land. They may be of many sizes and configurations, but all aim to prevent fish from leaving once they enter; any vessel can set and attend a trap.

Pots are used to capture crustaceans whose principal movement is by legs on the sea bed; pot fishing may be inshore or deep-sea. Pots of different sizes and configurations are set out and attract the species by means of baits, either cut-up fish or other sea creatures, or in a prepared, packaged form. Once the animal enters the pot through a specially designed entrance, it is unable to exit again and is trapped.

Harpooning can be used commercially on large species having high individual value. Work and time involved in searching, capturing, and bringing aboard each fish restricts the operation to valuable species. A retrieving line runs from the pole of the harpoon to the operating vessel; both the hunting and retrieving lines are run out from coils or tubs positioned on the vessel's foredeck.

Handling and Preservation on Ships

Handling of fish on ships needs special care. Fish begin to spoil immediately after death. Since the rate of spoilage is increased by any rise in temperature, the refrigeration of the catch by chilling or freezing is necessary. Spoilage is caused by enzymic, bacterial, and oxidative changes. Enzymes are present in the flesh and stomach of the fish, and the gut wall and the neighboring flesh, which are penetrated and softened by these enzymes after death, are thus easily invaded by spoilage bacteria. For this reason, in trawl, seine, and line fisheries for demersal fish, the fish is gutted and washed immediately after catching. In pelagic fisheries, where small fish are caught in greater numbers and not gutted, the spoilage is

quicker. Rancid odors and flavors are produced when atmospheric oxygen combines with fat in the fish, and temperature is the most important single factor affecting the rate of spoilage.

Chilling

The chilling system should maintain the lowest possible temperature. Considering the problem of quality, length of voyage, labor, and investment, chilling in melting ice is the most common method. An alternative to this is the use of chilled or refrigerated sea water (CSW or RSW), which has the additional advantage of handling the catch. The landing deck or container must be kept clean to prevent contamination of each haul by previous landings, and constant attention to hosing the landing deck area is necessary. Baskets, boots, and other items in contact with the catch should be thoroughly washed before the fish are landed. Fish should be carefully handled on the deck to avoid bruises and cuts; after landing, they are shifted to holding bins and mixed with ice to cool them. Gutting is done manually or by automatic gutting equipment on board. Gutting should remove all the guts and liver, and the belly cavity should be opened sufficiently to make the washing effective. After washing, the gutted fish are stored in melting ice to bleed; this insures that the flesh is free from discoloration. Chilling of fish in ice is done through stowage in bulk, in boxes, in pens, or in shelves.

In bulking, the fish and ice are mixed in intimate contact in order to achieve maximum storage life. It is carried out in a hold divided into pens fitted with removable shelves of wood or metal. Bulking to a depth greater than 40 cm is not recommended and, with some species, the limits are lower; otherwise, there would be excessive weight loss owing to the pressure. With bulking, there are difficulties in unloading the catch owing to the problem of separating the ice and the fish, a process which may involve a large amount of labor. Sometimes, for this reason, substantial water is added and fish are pumped out with a centrifugal pump.

The boxing of fish is usually preferred to bulking because of better cooling, and the elimination of handling of the fish (bruising and injury) at the dock, leading to an improved quality of product at the point of consumption. Boxing involves placing ice and fish in a container (box) of standard dimensions (normally 30 and 65 kg capacity) and stowing the box in the hold. The only real disadvantage is that boxing requires more hold space per unit weight of fish; return of reusable boxes may also be a problem.

In the shelf stowage system, the fish is stowed in single layers, gut cavity down, on a bed of ice. Sometimes a little ice is spread on top. This insures complete bleeding and retards spoilage by chilling, but it is a process which is not widely used on vessels because it is not space-efficient and the fish may become dehydrated as a result of the movement of air over the unprotected fish.

In pen stowage, fish and ice are placed in alternate layers to a depth of 70 cm, a second pen board is placed over the bottom pen, and the same process is repeated. This storage system is widely used for stowing fish in ice. In general, crushed or flaked ice is used in the chilling of fish on board. The storage life of fish in ice can be slightly extended through the use of antibiotics or salt-water ice. Antibiotics such as chlorotetracycline and oxytetracycline in concentrations up to 5 p.p.m. in ice are permitted in some countries to inhibit bacterial action. Separation of salt and water during freezing or melting is a problem with salt-water ice. Immersion of fish in CSW or RSW is used as an alternative to ice for the chilled storage of fish. The sea water may be chilled by the direct addition of ice (CSW), or by mechanical refrigeration with a heat exchanger (RSW). Storage of fish in CSW or RSW helps in unloading the fish from the vessel, as they can be pumped out and into a holding tank.

Freezing

Freezing at sea is necessary if the vessel remains at sea for a period up to 90 days. Storage in ice, or by other means of chilling, is adequate up to 15 days, after which the fish rapidly become inedible. If fish are frozen on board and stored at low temperature, the fishing vessel can remain at the fishing grounds until the hold is full. This improves the economics of fishing, and the general quality of fish landed. Depending on the facilities available on board and the markets, fish may be frozen in the round, gutted, gutted and headed, or in the form of fillets. A number of conventional freezers may be used at sea with little modification. The freezer and its refrigeration system have to conform to national regulations and insurance requirements for a fishing vessel. Many countries do not allow the use of ammonia as a refrigerant in fishing vessels. The design, operation, and type of refrigeration system must take into account the movement of the vessel, vibration, sea-water corrosion, and the extra rough usage under conditions at sea. Fish to be frozen and iced fish are handled in the same way, except that the former are frozen rather than merely cooled. Filleting is the step before freezing, and the fish should be held at chill temperature, below 5 °C, from catching to freezing, in order to retard spoilage and avoid the ill-effects of rigor mortis.

Rigor mortis, or death-stiffening, is the physical change that occurs in all animals, including fish, as a result of a series of complex reactions that go on in

the flesh tissue after death. It is one of the factors which, after a certain period of time, retards the postmortem autolytic and bacterial decomposition of the flesh and its protein, and thus extends the shelf-life. Rigor mortis in fish starts 1–7 h after death and its peak in slaughtered fish, kept in ice, lies between 5 and 22 h after death; the total duration of the rigor is 30–120 h. A prolongation of the rigor mortis period is of great economic importance, and harvesting and on-board handling practices should aim to extend rigor and maintain fish quality. Rigor mortis lasts longer when the fish has exerted less muscular activity prior to death, as it is refrigerated immediately. Immediate slaughtering of the fish after capture also appreciably extends rigor. The onset of rigor mortis is quicker at higher temperatures, and may occur only 10–20 min after death at about 30 °C. It is therefore essential that the fish is chilled quickly to avoid problems arising from rigor mortis during freezing. If possible, it is desirable to freeze fish before rigor mortis begins. Freezing in rigor can damage the fish, and only poor-quality frozen fillets or blocks can be made from such fish. However, while some fish can be frozen postrigor, others become quite unsaleable. In addition, prerigor frozen fillets are sometimes subject to contraction upon thawing; they should be transferred to the cold store immediately on removal from the freezer.

Handling in Port

Containers, boxes, and the portable sections of a fish room are handled by cranes. The discharge of a large vessel stowing its catch in pens and shelves is achieved by filling a basket with fish, attaching it to the end of a rope, hoisting it up and swinging it ashore; this operation may be mechanized using bucket elevators and ship-to-shore conveyors. In the chilled or refrigerated sea-water systems of stowage, pump unloading is convenient and widely used. The fish which are frozen on board are unloaded by standard, endless looped-canvas elevator belts. After landing, the fish should not be exposed to warm ambient temperatures of the auction market. Fish can be sold, after taking random samples from the boxes, containers, or shelves, as soon as the vessel arrives in the port, so that the fish can be unloaded directly on to the waiting transport. Cleanliness, care, and cooling are most important for first-class fish.

It is common to land fish at a market (fish pier) where it is put up for auction. This is also convenient for the ship and secures maximum prices for the fish. However, since these auctions can cause delay in the further distribution of the fish, affecting its quality, there should be several sales during the day and a corresponding distribution system (e.g., special fish trains) to avoid such problems. This will help to insure that the fish reach inland wholesale or retail dealers in time for further sale. Fish should always remain in ice throughout the period in the market, and special market boxes or kits can be used. Fish intended for distribution fresh and uprocessed are usually brought from the market or other landing points to a packing house. They are iced or reiced and packed in nonreturnable or returnable containers with the requisite insulating properties for transport to the inland wholesaler or retailer.

See also: **Fish**: Processing; **Shellfish**: Commercially Important Crustacea; **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage

Further Reading

Borgstrom G (ed.) (1961) *Fish as Food*, vols I and IV. New York: Academic Press.

Burgess GHO, Cutting CI, Lovern JA and Waterman JJ (eds) (1965) *Fish Handling and Processing*. Edinburgh; Her Majesty's Stationery Office.

Connell JJ (ed.) (1980) *Advances in Fish Science and Technology*. Farnham: Fishing News (Books).

FAO (1997) *Freezing in Fisheries*. FAO Fisheries technical paper 167. Rome: Food and Agriculture Organization.

Kristjonssen H (ed.) (1959) *Modern Fishing Gear of the World*. Farnham: Fishing News (Books).

Regenstein JM and Regenstein CE (1991) *Introduction to Fish Technology*. New York: Van Nostrand Reinhold.

Wheaton FW and Lawson TB (1985) *Processing Aquatic Food Products*. New York: John Wiley.

Fish as Food

G Lozano and A Hardisson, University of La Laguna, Tenerife, Canary Islands, Spain

Background

Seafood plays an important role in human nutrition, contributing near 125 000 million tonnes per annum of finfish, shellfish-molluscs, shellfish-crustaceans, and a few other edibles, both from fisheries catch and from aquaculture production. New technologies in fish manipulation, conservation, and manufacture have ensured a constant distribution and high quality of products all year round, and actual aquaculture bloom imply the local appearance of new species on

the market that were not common previously for the local population feeding habits or were even relatively unknown.

Fish (bony fish and elasmobranch fish) represent a large amount of marine-origin meals for direct human consumption and support a wide series of transformation techniques for subproducts or by-catch such as oil, flour, protein, and concentrates production.

Fish and Nutrition

For many centuries, fish has been one of our main foods and at present supplies almost 10% of the animal-origin proteins eaten at mealtimes; this proportion is much higher in developing countries, where fish may be the only source of animal protein. Japan, Norway, Iceland, Spain, and the UK are examples where fish, fishing, and aquaculture play an important role within the population's socioeconomic and feeding activities (Tables 1 and 2).

It is possible to divide fish into two large groups according to the forms of feeding advantage: primary products and secondary or subproducts. The former include fresh, frozen, smoked, salted, and canned fish, and the latter include oil, protein concentrates, and hydrolyzed products.

Two different types of fish are traditionally considered (in this case, exclusively finfish) under the generic denominations of white and blue fish. The first includes those species that have a white or whitish musculature, without the strong presence of darker-colored lateral muscular fascicles, and corresponds to the group of lean and semifat fish (with a fat content no more than 1.5 and 8–10%, respectively), such as cod (*Gadus morhua*), salmon (*Salmo salar*), and hake (*Merluccius merluccius*). The second type is characterized by its darker muscular tissue, the presence of a high degree of vascularization in the lateral muscular fascicles (Vogt's musculature, of fundamental importance on fish swimming in oceanic migratorial species as tunnids), and a high fat content (over 10%), which makes them fatty fish, e.g., the sardine (*Sardina pilchardus*), herring (*Clupea harengus*), mackerel (*Scomber scombrus*), and eel (*Anguilla anguilla*).

However, this distinction between lean, semifat, and fatty fish should be used with some caution, since a large number of species, e.g., the sardine (fatty fish) or hake (semifat fish), can have an enormous range of seasonal variation in their fat content owing to feeding-habits variation and sexual-products elaboration.

The food industry supplies a wide range of different fish types and products (finfish, shellfish-molluscs,

Table 1 World fish catch (1975–1986)

	1975	*1980*	*1984*	*1985*	*1986*
Composition ($\times 10^6$ tonnes)					
Finfish	60.5	63.5	73.4	75.8	81.2
Crustaceans	2.1	3.3	3.3	3.5	3.9
Molluscs	3.8	5.2	6.1	6.3	5.4
Total catch	66.4	72.0	82.8	85.6	91.5
Hunted (%)	93.0		90.0	87.7	
Farmed (%)	7.0		10.0	12.3	
Disposition					
Human consumption (%)	70.4	73.8	71.9	70.7	70.5
Of which, percentage consumed as					
Fresh	38.9	27.3	25.6	25.9	28.3
Frozen	25.3	30.6	34.0	33.6	33.2
Cured	16.7	21.8	20.8	21.5	20.8
Canned	19.1	20.3	19.6	19.0	17.6

Data from Hillcoat JB (1990) Fish farming. In: Somogyi JC and Hötzel D (eds) *Marine Foods. Lebbensmittel aus dem Meer*, pp. 10–21. Basel: Karger with permission.

Table 2 Production and use of world fisheries (millions of tonnes)

Production	*1990*	*1992*	*1994*	*1995*	*1996*	*1997*[a]
Continental fisheries						
Aquaculture	8.17	9.39	12.11	13.86	15.61	17.13
Catch	6.59	6.25	6.91	7.38	7.55	7.70
Total	14.76	15.64	19.02	21.24	23.16	24.83
Marine fisheries						
Aquaculture	4.96	6.13	8.67	10.42	10.78	11.14
Catch	79.29	79.95	85.77	85.62	87.07	86.03
Total	84.25	86.08	94.44	96.04	97.85	97.17
Total aquaculture	13.13	15.52	20.77	24.28	26.38	28.27
Total catch	85.88	86.21	92.68	93.00	94.63	93.73
Total fisheries	99.01	101.73	113.46	117.28	121.01	122.00
Use						
Human consumption	70.82	72.43	79.99	84.69	90.62	92.50
Reduction	28.19	29.29	33.47	30.78	30.39	29.50

[a]Preliminary estimation.

Data from FAO (1999) *El estado mundial de la Pesca y la Acuicultura 1998*. Versión en español. Parte 1, pp. 1–18. Rome: FAO with permission.

shellfish-crustaceans), both catch and aquaculture production, and marine, salty, or continental in origin:

- fresh and cold storage fish;
- frozen fish;
- dried fish;
- salted fish;
- smoked fish;
- conserved fish (canned or tinned fish);
- semiconserved fish (canned or tinned not pasteurized);
- Japanese traditional products (Neriseihin) as surimi and kamaboko;
- krill (euphausiids);
- trepang (dried sea cucumber);
- fish oils;
- fish flour and meal;
- hydrolyzed fish (siloed fish);
- fish protein concentrate.

Cold-temperature preservation can be achieved through a wide range of systems, from refrigeration by simple addition of ice to cryogenic ultrafreezing. Preservation may be increased by storage in modified-atmosphere conditions, ionization and vacuum packaging. Some species require special conditions for freezing such as sardine (superficial freezing: superchilling), tuna (brining solution of calcium chloride), or shrimp and prawn (brining solution of sodium chloride with sugar adding).

Refrigeration can be carried out by three different methods: cold sea water (CSW), refrigerated sea water (RSW) or a combined method of CSW/ice or RSW/ice, with or without the addition of antibiotics to water and ice. Freezing is based on very cold air flow, surface–plate contact and brining solutions (euthectic solutions), although air flow and cryogenics are the best techniques. All of these cold-freezing systems can be adapted for two different purposes: maintenance of a high-quality product along a determined, specific short time, or obtaining an acceptable quality in the long term.

Dried, salted, and smoked fish are short-term edibles in which only water loss and salt concentration ensure preservation, smoked fish being more durable due to semi-cooking of meal and adsorption of some preservative salts from wood (as tanines) during the smoking process.

The differences between packed fish (jarred, canned, tinned, etc.), conserves and semiconserves depend on the heat used first to prevent microbial spoilage (sterilization of cans involves the use of temperatures up to 105 °C for 30 min). Anchovies in olive oil is a good example of a tinned fish semiconserve, as fish fillets are salt-pressed and oil-soaked, but no sterilization is carried out after canning.

Neriseihin products, such as surimi and kamaboko, are a very important world-wide market, as they form the basis for a few new marine preparation foods as cheap sustitutives of more expensive natural meals: 'lobster tail,' 'crab sticks,' and 'young eels' are popular products.

Fish oils, flours, meals, hydrolyzed solutions and protein concentrates may be obtained from specific fisheries (Peruvian anchoveta (*Engraulis ringens*)) or as subproduct or bycatch of other principal catches. Ecuadorian and Peruvian flours, which are used for production of aquaculture feed, show fundamental differences in general characteristics (high quality and low cost) compared with European flours, as the latter product is a result of damaged specimens discharged for human consumption, or specimens of no commercial interest by primary demand, all with a low-quality composition and similar cost to that of the Peruvian and Ecuadorian counterparts.

Other types of seafood (and continental water species) should be mentioned, although their importance is local or anecdotal, for example, sea urchins (genera *Echinus, Paracentrotus, Sphaerechinus*). A few species are traditionally caught and consumed in Spain and France, thus increasing demand and initiating the introduction of specific aquaculture industries, which produce both fresh specimens for direct consumption and manufactured products commercially denominated 'sea urchin caviar,' presented as jarred or tinned semiconserves.

Among all these food sources, the krill fisheries that exploit different species of the genus *Euphausia, Thysanoessa, Meganictiphanes*, and *Nictiphanes*, with a total catch of 160 000 tonnes in 1996, are good examples of a new component of human nutrition and aquaculture feed, although their fundamental role in the marine trophic systems must take total catch quotas under strict controlled management.

General Chemical Composition

From a nutritive point of view, the most interesting chemical components of fish (comprising finfish, cephalopods, shellfish-molluscs and shellfish-crustaceans) are proteins and fats (Tables 3–5 and Figure 1), as well as water, which can constitute 70–80% of the total composition, mineral salts (usually refered to as ash), and vitamins. Carbohydrates and fiber are practically nil in fish.

Proteins and Nonprotein Nitrogen Substances

The value of fish proteins is very high. It is better than meat protein and, although inferior to milk and egg protein, has a stable composition of essential amino acids, with slight deficiencies of methionine and

Table 3 Composition of several seafoods

	Protein (g per 100 g)	*Fat (g per 100 g)*	*Energy (kJ per 100 g)*
Fat fish	18–23	3–23	420–1260
Lean fish	16–29	< 1	294–336
Clam	10.70	0.50	198
Mussel	10.80	1.90	282
Scallop	19	0.90	352
Oyster	10.20	1.40	297
Squid	17	1.30	341
Octopus	10.60	1.00	239
Lobster	15–20	0.5–2.5	294–320
Norway lobster	15.00	0.70	278
Crab	19.50	5.10	518
Prawn	21.00	1.30	400
Sardine	23	13.8	672–1176
Mackerel	23	17.3	756–1302
Tuna	28	8.0	609–1680
Gilthead seabream	17.00	2.70	386
Cod	17.70	0.40	311
Hake	11.80	1.80	266
Turbot	16.10	3.60	425

Adapted from Deslypere JP (1990) Effect of fish consumption compared to intake of fish oil. In: Somogyi JC and Hötzel D (eds) *Marine Foods. Lebbensmittel aus dem Meer*, pp. 10–21. Basel: Karger, with permission.

Table 4 Composition of several transformed seafoods

	Proteins (g per 100 g)	*Fats (g per 100 g)*	*Energy (kJ per 100 g)*
Canned tuna (olive oil)	23	13	874
Smoked salmon	21.50	9.30	710
Canned sardine (olive oil)	22	16.2	977
Fish protein concentrate A	~75	1	1292
Fish protein concentrate B	~75	10	1630
Fish flour	~65	~8	1350

Data from Mataix-Verdú J and Mañas-Almendros M (eds) (1998) *Tablas de Composicion de Alimentos Espaoles*, 3rd edn. Granada, Spain: Editorial de la Universidad de Granada and Lozano (not published) with permission.

Table 5 Composition of some seafish from the Canary Islands (central-eastern Atlantic)

Species	*Proteins (g per 100 g)*	*Fats (g per 100 g)*	*Minerals (g per 100 g)*
Sardine	14.92	12.79	3.68
Round sardinella	16.28	9.13	3.47
Chub mackerel	15.72	6.73	4.41
Blacktail comber	12.67	1.70	6.73
Large-eye dentex	12.11	3.30	7.38
Common pandora	13.34	2.82	4.30
Axillary sea bream	12.50	2.25	5.46
Bogue	15.41	9.39	4.41
Salema	12.69	2.17	3.40
Parrotfish	12.11	1.62	4.30

Data from Lozano G, Ros MM, Quintero ME, Balguerías E and Lozano F (1987) Consideraciones alimentarias en especies de peces de consumo común en las Islas Canarias. *Alimentaria* 180: 47–56.

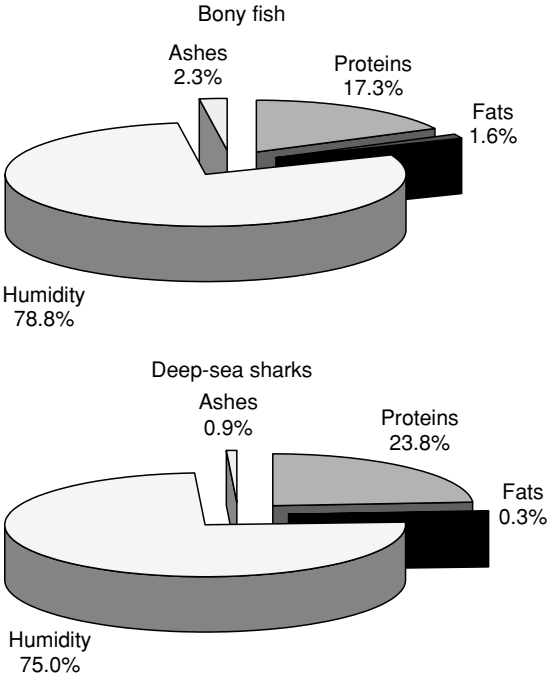

Figure 1 Mean percentage values of proteins, fats, humidity and ashes of several bony fish (top) and elasmobranch fish (deep-sea sharks) (bottom) of normal human consumption in the Canary Islands. Data from Hardisson A, Báez A, Pascual P and Lozano G (1997) Trace metal content in deep-water sharks from the Canarian seawaters. *J. Vet. Pharmacol. Therap*. 20(1): 283–284.

threonine, and an excess of lysine – very important if we take into account the fact that this latter amino acid is lacking in cereals.

The protein content of a species depends on many biological and environmental circumstances (sex, age, sexual maturity, nutritive conditions, etc.); this variation can be rhythmic, periodic, or nonperiodic. Finfish can show a wide range of variation between ~10 and ~30% in wet weight.

The biological coefficient value of protein (B_{vp}) compares retained nitrogen (N_r) with gastrointestinal absorbed nitrogen (N_a):

$$B_{VP} = N_r/N_a \times 100.$$

Nonprotein nitrogenous substances are important in the taste of the fish, with the highest concentrations being found in migrant and elasmobranch fish. The nitrogenous substances are mainly free amino acids, peptides, purine bases, urea and trimethylamine oxide.

Fats

Fats are mainly composed of unsaturated fatty acids, including oleic acid (18:1*n*-9) and high unsaturated fatty acids such as linoleic acid (18:2*n*-6) and linolenic acid (18:3*n*-3), among others. These have an

important role in the control of blood cholesterol levels and in the prevention of cardiovascular problems of arteriosclerotic origin. This is an important feature in the case of blue fish, such as sardine, anchovy (*Engraulis encrasicolus*), herring and bluefin tuna (*Thunnus thynnus*), as well as other species of pelagic fish. As in the case of proteins, fats undergo high seasonal variations in their content, in this case related to trophic and genetic periods and migrations.

Recent studies show increasing interest in deep-sea shark fisheries because of the high levels of unsaturated hydrocarbon compounds in shark liver oils. Squalene ($C_{30}H_{50}$) and the hydrogenated form squalane ($C_{30}H_{62}$) play an important role as lubricants and emolients and as a base for a range of products in the pharmaceutical and cosmetic industries, because they help through-the-skin absorption and avoid water loss.

Carbohydrates

Carbohydrate concentrations are very low (0.1–1%) in fish, and their contribution to the diet is insignificant.

Vitamins

Fish is a rich source of fat-soluble vitamins, such as vitamins A and D. In lean fish, they are found almost exclusively in hepatic oils (the best example being cod liver oil), but in the case of fatty fish, they are also present at high levels in muscular tissues. Vitamins B_1, B_2, and B_{12}, and nicotinamide are also found in significant proportions in the liver and gonads.

Minerals and Heavy Metals

Fish contain many important minerals, mainly located in the viscera, muscles, and bones. Among the most abundant elements are sodium, potassium, calcium, iron, chloride, phosphate, sulfate and iodine. Iodine is very important in the treatment of illnesses caused by the malfunction of the thyroid gland, which can cause delays in growth and/or dwarfism.

In terms of the alimentary structure of fish, the heavy metal content present in the edible parts is an important factor because of the toxic characteristics of these cations and their slow accumulation in the organisms of the trophic chain by bioaccumulation or biological amplification (Table 6). Mercury, lead, tin, cadmium, and vanadium are particularly important heavy metals and are regulated, as far as the maximum total content permitted for consumption, acceptable daily intake, etc., are concerned, by the World Health Organization and national organizations working together on health problems.

Table 6 Mercury levels in some seafish from the Canary Islands (central-eastern Atlantic)

Species	*Mercury level (p.p.m., w/w)*	*Samples*
Sardine	0.025	Fresh
Round sardinella	0.003	Fresh
Chub mackerel	0.003	Fresh
Blacktail comber	0.004	Fresh
Large-eye dentex	0.006	Fresh
Common pandora	0.003	Fresh
Axillary sea bream	0.005	Fresh
Bogue	0.005	Fresh
Salema	0.004	Fresh
Parrotfish	0.003	Fresh
Tuna	0.234	Canned
Albacore	0.270	Canned
Yellowfin tuna	0.005	Canned
Skipjack	0.082	Canned

Data from Lozano G, Ros MM, Quintero ME, Balguerías E and Lozano F (1987) Consideraciones alimentarias en especies de peces de consumo común en las Islas Canarias. *Alimentaria* 180: 47–56.

Digestibility

A significant advantage of fish as a food is its easy digestibility, which results from the special configuration of the muscular structures. These are composed of short muscular segments (myotomes) separated by conjunctive covers (myosepts), which together produce short muscular fibers. Myosepts are more easily transformed to gelatines at 65–70 °C, causing the destruction of muscular structures, and improved access to fibers, than in other animal meats.

Fatty fish are less digestible than lean fish but provide more energy to the consumer. With a mean content of 18–20% protein, fatty fish provide up to 1302 kJ per 100 g, compared with 336 kJ per 100 g in lean fish. Taking into account losses after removal of the head and internal organs (up to 50–60% of the fresh weight), these values are reduced to 357 and 155 kJ per 100 g, respectively.

The digestibility coefficient (D_c) compares absorbed nitrogen (N_a) with ingested nitrogen (N_i) as follows:

$$D_c = N_a/N_i \times 100.$$

D_c values range from 70% (ray) to 100% (cod and hake) in finfish, 75–85% in molluscs, and 70–90% in crustaceans.

Popular Varieties and National Preferences

In general, it is difficult to list the entire range of species or groups of species that comprise the fishing resources of a country, region, or specific geographical area, given the number of variables of differing importance, such as industrial and small-scale fishing, fleet type and

catch methods, the oceanographic and biological characteristics of the environment, migrant and sedentary species, coastal and oceanic areas, insular systems, and national and local gastronomic traditions.

The number of variables increases in the case of nations whose fishing activity includes freshwater species (in the strict sense of the term, or combining marine and freshwater characteristics) and which can boast aquaculture potential in one or many of its varied forms. The situation in such cases is clearly complex when it comes to handling data concerning catches, fishing effort, and production.

Denmark's fishing and production are based on gadoids, pleuronectids, clupeids, and scombrids; Australia's on lobsters, crabs, tuna-like fish and sharks; Uruguay's on squids and hake (*Merluccius* spp.); Mexico's on lobsters, scombrids, anchovy (*Engraulis ringens*) and sardine (*Sardinops* sp.); Japan's on scombrids and lobsters (with, in this case, considerable aquaculture production); and Norway and Greenland's on salmonids and gadoids; USA's on salmonids, cyprinids, anguilids, clupeids, gadoids and scombrids (although one should not overlook the importance of shark-fishing in the deep waters of the Pacific coast); Germany's on salmonids, cyprinids, anguilids, clupeids, gadoids, and scombrids; the UK's on gadoids, merlucids, clupeids, flat fish, and salmonids; and Spain's on, among others, anguilids, cyprinids, salmonids, scombrids, sparids, gadoids, merlucids, flat fish, cephalopods, and crustaceans (Table 7).

Table 7 Marine and freshwater species of fish normally consumed in Western Europe

Species	*Marine*	*Freshwater*
Sturgeon		+[a]
Sardine	+1	
Herring	+1	
Anchovy	+1	
Salmon		+[a]
Trout		+
Northern pike		+
Tench		+
Carp		+
Eel		+[b]
Cod	+2	
Hake	+2	
Mackerel	+1	
Tuna	+1	
Skipjack	+1	
Turbot	+2	
Plaice	+2	
Sole	+2	

[a]Freshwater genetic areas.
[b]Marine genetic areas.
+1, Pelagic fish; +2, demersal fish (ground fish).
Data from Lozano G and Hardisson A (1993) Fish as food. In: Macrae R, Robinson RK and Sadler MJ (eds) *Encyclopaedia of Food Science, Food Technology and Nutrition*, vol. III, pp. 1834–1838.

Preferences for Lean, Semifatty, and Fatty Species

Local preferences for a particular type of fish (lean, semifatty, or fatty) depend on many factors. Some of these are seasonal and biological (catching species on a massive scale as they migrate during a given period, presence near areas of reproduction, disappearance and replacement of species), whereas others have more to do with traditions and culture (consumption of a species during a short period, even though it is present in the fishing zone for all or part of the annual cycle) or social, economic, and political considerations (moving to other fishing grounds, changes in selling price, employment complications in the sector, the imposition of reduced catch quotas, etc.).

In the western Mediterranean countries (Spain, for example), species of white fish (lean and semifatty) are consumed throughout the year, whereas the greatest consumption of blue fish is in the summer, the most important species being bluefin tuna (*Thunnus thynnus*), bonito (*Sarda sarda*), skipjack (*Katsuwonus pelamis*), albacore (*Thunnus alalunga*) and yellowfin tuna (*Thunnus albacares*). In winter, the mainstays are mackerel (*Scomber scombrus*), chub mackerel (*Scomber japonicus*), sardine (*Sardina pilchardus*) and anchovy (*Engraulis encrasicolus*). The same is true, to a large extent, of other neighboring Mediterranean (France and Italy) and Atlantic areas (Portugal).

This pattern is also seen clearly in the Canary Islands (central-eastern Atlantic), where multispecific small-scale fishing is practiced for white fish and crustaceans (demersal species) between October and April, whereas during the period May to September, activity centers almost exclusively on bluefin tuna, skipjack, albacore, bigeye (*Thunnus obesus*) and yellowfin tuna (pelagic species), which are also caught by large Russian, Korean, and Japanese fleets.

Turkey deserves special mention, since, despite its extensive scombrids, clupeids and flat fish resources, it is not a fish-consuming country. In spite of strenous efforts by the government, the traditional cultural situation has remained unchanged. Thus, an extremely important source of food, for both internal consumption and export earnings, is still untapped because Turkey is not a fishing nation.

Usage of Marine versus Freshwater Species

As mentioned earlier, the harnessing and use of marine or freshwater species depends on a wide range of circumstances, most notably, in the case of freshwater fish, the presence of rivers, lagoons, and

lakes in the fishing area. This enables the resource to be extracted and distributed without any great difficulty, given that, for the most part, marketing of this type of fish tends to be limited and is almost never exported.

Nevertheless, both manufactured products (particularly smoked fish) and species, which are produced by means of aquaculture, are opening up new markets for freshwater fish in countries, or regions, where previously they had been rejected or unknown, and where consumption was practically nil.

Manufactured products are derived from eel, sturgeon (*Acipenser sturio*), salmon and trout (*Salmo* spp.), and, in the case of aquaculture species, tilapia (*Tilapia* spp.), tench (*Tinca tinca*), carp (*Cyprinus* spp.), salmon and trout. All of these form part of a sizeable range of widely distributed and economically important species.

The most important characteristic of aquaculture is its capability of producing high-quality edibles, both nutritional and health properties at an almost constant market price, irrespective of seasonal or socioeconomic situations or politics (migrations, weather condition, political restriction to foreign fleets to fishing in national fishing grounds, bans owing to biological reasons, total allowable catches).

Of special significance is the controlled addition of modified diets to species in order to increase their growing rates, condition factor, and certain biological characteristics (e.g. resistance against diseases), thus creating an improved food for human consumption. Components such as folic acid, ascorbic acid, and L-carnitine are being used in new types of foods.

In the EU member states, all production (marine and continental) is regulated under the 'Sectorial Program of Coastal Fisheries and Aquaculture' that assures management, assessment, market procedures, and scientific aid in quinquennial programs. Some species that were not accessible via aquaculture a few years ago are now being consumed in great quantities. For example, turbot (*Scophthalmus maximus*) has seen an extraordinary increment in production to 10 000 tonnes in 1999 (4400 tonnes in Spain alone).

Problems of Supply and Demand

The complex nature of the various sectors of the fishing industry, catches, distribution, processing, and marketing, can be explained by a series of factors. In some cases, a small number of species are fished for a particular type of industry (e.g. canning or smoking factories), or in order to supply a traditional and unchanging market requiring a constant level of quality catches throughout clearly defined periods.

These circumstances may vary considerably as a result of resource modifications produced by natural or artificial causes (variations in the environment, changes in the recruitment rate, overfishing) or by changes in eating habits (varying in demand caused by the introduction of species produced through aquaculture, or caught in areas previously not fished by the fleet). Such species displace or at the very least compete commercially with those that are traditionally consumed. The crises affecting the anchoveta in Peru (only 200 000 tonnes caught in 1985), cod in Greenland (7000 tonnes in 1986), the demise of the European herring, or the Saharan Bank sparids, are prime examples of resource modifications, whereas cases of aquaculture include the introduction of salmon, trout, tench, and eel farming in Spain.

See also: **Fish**: Introduction; Catching and Handling; Fish as Food; Demersal Species of Temperate Climates; Pelagic Species of Temperate Climates; Tuna and Tuna-like Fish of Tropical Climates; Demersal Species of Tropical Climates; Pelagic Species of Tropical Climates; Important Elasmobranch Species; Processing; Dietary Importance of Fish and Shellfish; **Fish Farming**

Further Reading

Bligh EG (1990) *Seafood Science and Technology*. Oxford: Fishing News Books.

Brito G, Díaz C, Hardisson A and Lozano G (1994) Concentrations of Mercury in fresh and salted marine fish from the Canary Islands. *Journal of Food Protection* 57: 246–248.

Connell JJ (1995) *Control of Fish Quality*, 4th edn. Oxford: Fishing News Books.

FAO (1999) *El estado mundial de la Pesca y la Acuicultura 1998. Versión en español. Parte 1*, pp. 1–18. Rome: FAO.

Hardisson A, Báez A, González T, Pascual P and Lozano G (1997) Trace metal content in deep-water sharks from the canarian seawaters. *Journal of Veterinary Pharmacology and Therapeutics* 20: 283–284.

Lozano G, Ros MM, Quintero ME, Balguerías E and Lozano F (1987) Consideraciones alimentarias en especies de peces de consumo común en las Islas Canarias. *Alimentaria* 180: 47–56.

Martin RE and Flick GJ (1990) *The Seafood Industry*. New York: Van Nostrand Reinhold.

Mataix-Verdú J and Mañas-Almendros M (eds) (1998) *Tablas de Composición de Alimentos Españoles*, 3rd edn. Granada, Spain: Editorial de la Universidad de Granada.

Somogyi JC and Hötzel D (eds) (1990) *Marine Foods. Lebensmittel aus dem Meer*. Basel: Karger.

Demersal Species of Temperate Climates

G D Khedkar and B V Jadhao, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India
C D Khedkar, College of Dairy Technology, Warud (Pusad), Maharashtra, India
N V Chavan, Adarsha College, Hingoli, Maharashtra, India

Background

Fishes of many families, shapes, and sizes obtain their food in the near-bottom zone and show morphological and behavioral adaptations for life on or near the seabed. These are the demersal fishes, comprising both benthic and benthopelagic species, the latter usually performing vertical migrations to feed. Demersal fishes occur at all depths and in all near-bottom habitats of the oceans. In this article, emphasis is given on demersal species such as cod, haddock, plaice, and sole, which live on or near the bottom of the sea; the continental shelves along the coasts of Europe, Greenland, Iceland, and North America. Worldwide, about 85% of the total continental shelf area has sandy or muddy substrate. Only about 6% is rocky or gravelly, and the remaining areas are coral reefs or shell beds. However, there are both latitudinal and depth-related patterns. Corals are almost entirely confined to low latitudes where organically enriched muddy areas are also most extensive, particularly near the mouths of major rivers or below highly productive upwelling areas. Sandy, rocky, and gravelly sediments are more common at high latitudes. Regional and local modifications of the distribution and character of soft sediments are common as a result of water currents flushing the shelves. In addition to offering a range of physical habitats to fishes, continental shelves are usually highly productive, and especially in temperate and boreal regions, demersal species of the shelf waters are very abundant and support some of the world's major fisheries. Of the approximately 13 500 marine fish species, 1000–2000 species inhabit continental shelf-water of temperate zones. The majority of these are demersal species.

Taxonomic Diversity, Geographical Patterns, and Assemblages

A wide range of families, including both cartilaginous and bony fishes, have demersal representatives in shelf-waters. There are several rather consistent geographical patterns, however, on both a worldwide and regional scale. The abundance and biomass of demersal fish are considerably higher at high latitudes, i.e., in temperate climates, normally due to the high abundance of a few species, especially gadiform fishes such as hakes (Merluccidae) and cod-like fishes (Gadidae), but also flatfishes (Pleuronectiformes) and rockfishes belonging to Scorpaenidae (Figure 1). Predominant differences between the oceans have evolved, the gadiform fishes being more diverse on temperate Atlantic shelves than on North Pacific shelves, where scorpaenids are very diverse and few gadiforms occur. Narrow shelves tend to have fewer demersal species than wide shelves and shelf seas. The highly productive upwelling areas associated with eastern boundary currents have few demersal species, and most are benthopelagic. Hakes are very well adapted to these environments and live both demersally and pelagically, mainly feeding on other fish.

Adaptations of Demersal Fishes

One adaptation apparent in demersal species is in response to the vertical gradient in light intensity. Many species can adjust their body coloration to their environment almost continuously, while some species like Atlantic cod (*Gadus morhua*) have different persistent color varieties depending on the habitat in which they grow up and live. Most shelf and upper slope fishes are countershaded. They have dark backs to reflect enough of the strong downwelling light to match the dim upward component of the ambient light, with silvery flanks to mirror the background from all visible angles. However, bright sunlight in the clearest oceanic waters penetrates to only 1000 m or so, and in turbid slope waters, penetration is reduced to considerably less than this. Therefore, the great bulk of the ocean is unlit by the sun, resulting in a reduction in countershading for protection, often accompanied by an overall darkening in body colour, with increase in the living depth of the species. Because the red end of the spectrum is most readily absorbed by sea water, darkening of the body at depth can be achieved either by black or by red pigmentation. Most demersal species tend to be blackish, but striking examples of red camouflage are found in the orange roughy, *Hoplostethus atlanticus*, and in *Sebastes* species. The rapidly diminishing intensity of light with depth in the ocean has resulted in the refinement of a range of sensory systems in demersal bottom fishes. Surprisingly, there is no general trend towards a reduction in eye size as occurs in pelagic fishes.

Unlike the pelagic fishes, which are slender and often torpedo-shaped, the typical demersal fishes are

not so well adapted in terms of body shapes or physiology for prolonged fast swimming. However, many shapes ranging from an eel shape to the more classical fish shapes of the cod, sparids, and rockfishes have proven successful (Figure 1). Regulating buoyancy may not be as important to demersal fish as to pelagic organisms. Demersal fish species come in all sizes from small gobies, reaching a length of only a few centimeters, to the Atlantic halibut (*Hippoglossus hippoglossus*) and Greenland shark (*Somniosus*

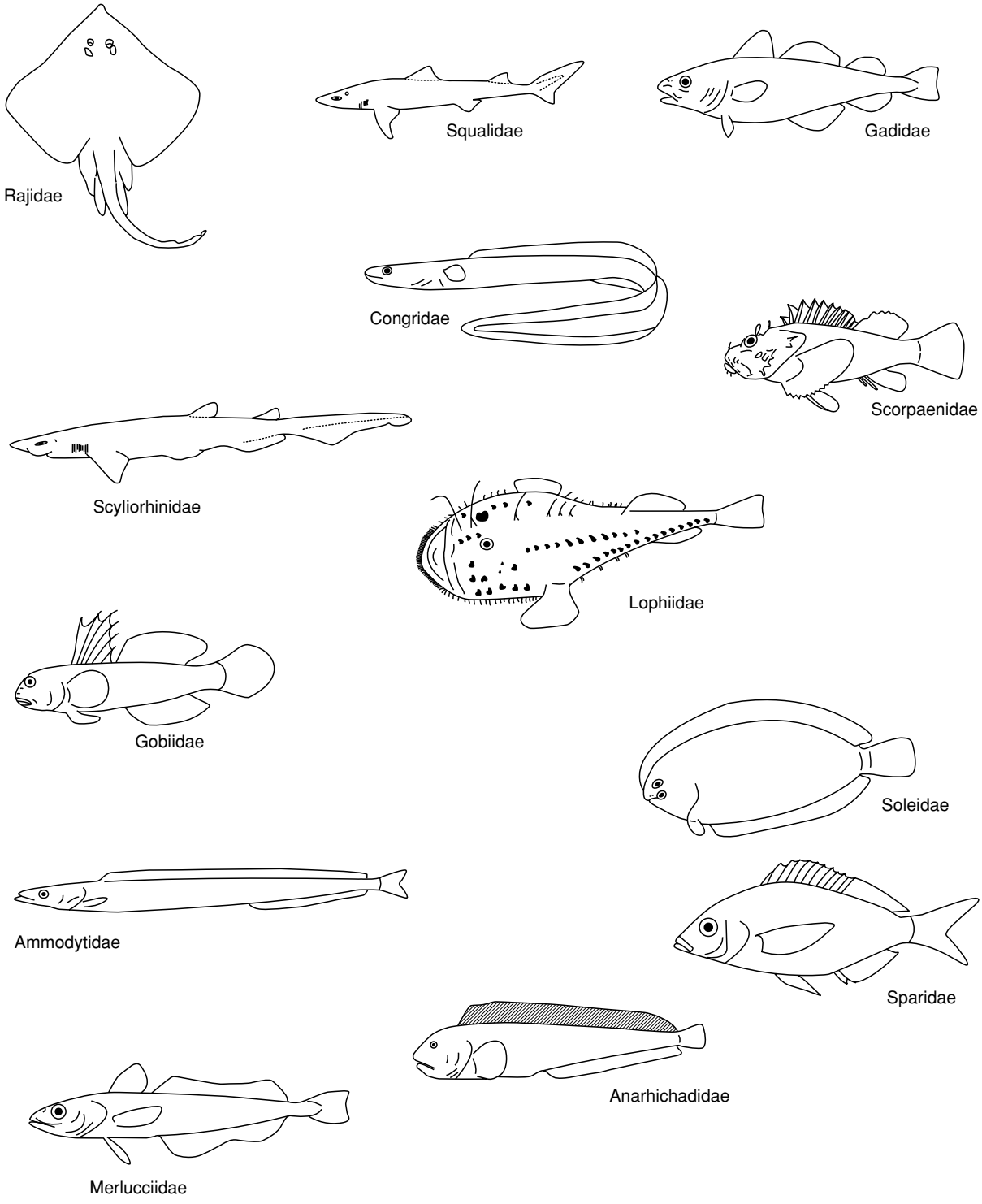

Figure 1 Body shapes of fish from selected demersal families.

microcephalus), which may reach 2 and 6.5 m in length, respectively. Many modes of life are possible in the demersal environment. Some fishes are sluggish and have evolved a typical 'lie-and-wait' strategy, while others are very active, and some almost behave as pelagic species. Flatfishes and several rays spend part of their time buried in the sand and may only emerge for certain times of the day or tidal cycle to feed. Most of the demersal species are more active during the night or at dusk and dawn. Camouflage may be attained in many ways, i.e., through body shape, skin coloration, and modification of skin into the appendages resembling algae or debris. The retention of functional and substantial eyes, together with specialized adaptations, must surely be in response to the widespread phenomenon of the biological light, which is produced by very many organisms in the deep sea, including those associated with the sea floor. In contrast to the pelagic oceanic fishes, relatively few deep-sea demersal fishes have evolved elaborate light-organ systems of their own. For example, there is no need for ventral camouflage in bottom dwellers or bioluminescent lures in a habitat at the sediment–water interface where food ultimately collects and is thus found more easily simply by foraging.

Feeding and Diets of Demersal Fishes

Most of the demersal species of temperate region are carnivores, customarily grouped into three categories, piscivores, benthophages, and zooplanktivores, that predominantly eat fish, benthos, and zooplankton, respectively. Many demersal fishes depend on pelagic production by feeding on vertically migrating nekton and zooplankton entering or inhabiting the near-bottom layer. Piscivores usually have the capacity to deal with live prey that are large compared with their own body size, and so they tend to have large gapes. Some swallow their prey whole, and others tear and bite the prey into smaller pieces. However, the techniques vary widely between species within the demersal fish group. Piscivores generally have several rows of conical backwardly pointing teeth both on the jaws and within the mouth and pharynx. These primarily prevent the prey from escaping and aid swallowing. Benthophagous fishes also have several prey capture methods. Some apply suction after detecting prey on the surface of the sediment or within the sediment, and others bite off the siphons of buried bivalves. Others feed on shelled animals and have strong, muscular jaws. Zooplanktivores often lack dentition on the jaws and rely on suction for ingesting their rather small prey whole. Most species probably pick their prey one at a time rather than filter-feed as clupeids and mackerels do at high prey densities. In most demersal species, there are significant ontogenetic changes in diet and feeding ranges. At metamorphosis, when larval characters are lost, and most demersal fish take up a demersal life style, pronounced diet changes from feeding on planktonic crustaceans to feeding on benthic or benthopelagic prey may occur. As a fish grows in size, larger or more energy-rich prey are needed to satisfy its energy requirements. Many species therefore go through a succession of feeding modes during their lifetime, usually starting as planktivores and ending as piscivores or benthofages. For a given species, pronounced diet shifts tend to occur at rather fixed sizes, but habitat shifts may happen concurrently, and it is sometimes difficult to determine which is more significant. Many species are quite opportunistic, and their diet varies greatly between habitats and seasons, often depending on the food availability.

Major Demersal Species

The major species of demersal species include the Gadiformes or cod-like fish (sometimes called the Anacathini – fish without spines). These include commercially valuable species like the cod (*Gadus morhua*), haddock (*Melanogrammus aeglefinus*), ling (*Molva molva*), and whiting (*Merlangius merlangus*).

Cod

The cod belongs to the family, the Gadidae (genus *Gadus*), represented by at least 19 species in the demersal habitats. They can be distinguished from other genera in the family by having three dorsal and two anal fins. There are 10 species, of which six possess a barbell, a sensory process that hangs down, or may be extended forwards, from the lower lip. In four of these species, the upper jaw is larger than the lower jaw: the cod, *Gadus callarias*, green–yellow with a number of dark spots and a white lateral line; the haddock, *G. aeglefinus*, gray with a prominent black blotch on the shoulder and a black lateral line; the bib or pout, *G. luscus*, very deep in the body and copper-colored, with some dark irregular bands, and the poor cod, *G. minutus*, similar to the last, but smaller and not so deep in shape and brownish yellow with no bands. Then there are the two species that have a barbell but whose lower jaw is larger than the upper jaw: the coal-fish or saithe, *G. virens*, with very dark skin on top, almost black, but a white lateral line and the teeth in the upper jaw all about the same size; and the pollack-whiting, *G. esmarkii*, with the teeth on the outer margin of

the upper jaw stronger than the others; this is a small fish with a beautiful golden color. The geneva that have no barbell include the whiting, *G. merlangus*, with silvery sides and a black spot at the root of the pectoral fins and the pollack, *G. pollachius*, dull green with a lower jaw longer than the upper jaw, the blue whiting, *G. poutassou*, which is like a whiting, but has a very long first anal fin and is smaller; and the silvery cod, *Gadiculus argenteus*, which lives in deep water over the edge of the continental shelf. The cod is undoubtedly the most important demersal food fish of our seas, and not only around the British Isles, but in the Arctic waters of Iceland, northern Norway, and Bear Island.

A large cod may produce as many as four million eggs at one spawning. All members of this family lay floating eggs; their small transparent spheres then drift with the moving waters and develop into the young fry. The spawning period is February to April, and the principal areas are those of the Great Fisher and Ling Banks, and Long Forties and Flamborough grounds. These have been determined by the charting of eggs and fry caught in plankton nets and from the landings of ripe and spawning fish. The fry, which are only 6 mm long when they hatch, spend about 2.5 months as part of the plankton, and are subject to the same hazards of shortage of food or excessive abundance of predators as the young of other species. They feed almost entirely upon the small crustaceans (copepods) of the plankton. During this period, circulating currents keep them within the North Sea, and when they are just over 2 cm long, they seek the bottom. Once on the sea-bed, they are extremely difficult to catch, especially as the nursery regions contain a good deal of rough ground. Still feeding upon crustacea, but now the bottom-living amphipods, isopods, and small crabs, the young cod grow quickly on the typical nursery grounds and begin to be caught by the commercial trawlers when they are 1.5–2 years old and 30 to 36 cm length; in some deeper regions of the North Sea, however, growth is slower, and they may not enter the fishery until they are 3 years old. The age of a cod is generally determined from its scales and not its otolith. Cod do not begin breeding until they are about 5 years old and have reached a length of some 69 cm.

Adult cod feed mainly upon fish, particularly herring, mackerel, small haddock, and sand eels; they also take mollusks, crustaceans, and bristle worms to a lesser extent. There is a marked contrast between the feeding of the cod and that of its close relative, the haddock. This is interesting in view of the well-established rule that two closely allied species cannot remain competing with one another in the same habitat: one, if slightly the more efficient, will certainly eliminate the other. The cod and haddock exist together on the same grounds but have distinctly different food preferences. Whereas the cod feeds mainly on fish, the haddock, except for sand eels, feeds almost entirely upon invertebrates: bristle worms, crustacea, mollusks, and particularly brittlestars, and small sea urchins. The haddock food is limited to rather small slow-moving animals found close to, or burrowing in, the bottom. Cod and hake eat larger and more quickly moving animals, even considering only fish of a comparable size. The nature of the mouths of these three species is significant. The mouths of hake and cod are much larger than those of haddock and practically, if not absolutely, terminal. They are provided with sharp teeth, and all the mouthparts are strong. Haddock, however, have a small mouth positioned somewhat ventrally. Their mouthparts are soft, and the teeth are ordinarily dull. They are, however, provided with better-developed muscular lips. Cod and hake are well equipped for capturing large or fast-moving objects, but haddock, with their smaller, ventrally positioned mouths, are best equipped for picking small animals off or out of the bottom. In addition, the heaviest built portion of the haddock is the anterior part of the body, and this may help them to remain more easily in a forwardly tilted position while feeding.

Haddock

The haddock is perhaps commercially second in importance to the cod among the trawl-caught fish. It is very abundant in the North Sea, particularly in the northern half, and further north to the Faeroes, Iceland, and the Barents Sea, but it does not go as far north towards Spitsbergen as does the cod. On the western side of the Atlantic, the haddock is also abundant on the coastal banks from Cape Cod to the Cabot Strait, and large concentrations are found on the southern hald of Grand Banks of Newfoundland. The main spawning period of haddock, like that of cod, is from February to April. The haddock is smaller than the cod, becomes mature much earlier, when only 2 years old, and is less prolific. A 2-year-old fish of about 25 cm in length will produce some 30 000 eggs, a 3-year-old (30 cm long) some 100 000, and a 6-year-old (only about 40 cm long) some 280 000. Haddock in the trade category 'extra large,' i.e., over 152 cm in length, are rare today; occasionally, however, much larger 'jumbos' are caught. For both cod and haddock, there is quite a marked difference in the growth rates of fish from the northern and southern halves of the North Sea. In the north, a 5-year-old haddock will have an average length of only 32 cm, whereas in the south, a haddock

of the same age will measure 39 cm. The sea-bed to the southern North Sea is warmer than that of the northern North Sea, and this is not simply because of the difference in latitude but because of the greater depth. The deeper water of the north is cut off in summer from the warmer zones by the formation of a discontinuity layer. In the much shallower south, where the water is more disturbed, no marked separation of upper and lower layers takes place, and so the sea-bed is kept warmer, and the fish grow more quickly.

Spawning takes place at various points in the very large area over the northern North Sea and in a much smaller isolated patch just north of the Outer Hebrides. It is still not known to what extent the Hebridean spawners are linked to the main North Sea stock. The positions of spawning within the large area are not fixed; their centers shift from place to place in different years.

Until the small embryo fish begin to produce their characteristic pigment, it is exceedingly difficult to distinguish the eggs of the haddock from those of the cod, and they are often found floating together. They drift and develop, as do those of the cod, but the fry remain in the plankton longer and so do not seek the bottom until they are over 5 cm long. Since the young haddock are larger than the young cod when they first settle down to a bottom-living life, they can be caught more easily.

Whiting

Among the four members of the genus *Gadus* that lack the barbell, the whiting, *G. merlangus*, is the most notable. It forms a considerable part of the catch of North Sea trawlers and seiners and is also abundant in the English Channel and Irish Sea; indeed, it is the most important demersal species for the fishermen of Northern Ireland. The catches in the North Sea have indeed been increasing since 1982, and it has been estimated that the total European catch for human consumption from this area was 140 000 tonnes with, in addition, some 30 000 tonnes of smaller fish taken for fishmeal. The whiting spawns principally in April and May throughout the North Sea but mainly within the 90-m line. In their pelagic phase, after reaching 2 cm in length, the young whiting associate with jellyfish, particularly *Pelagia cyanea*, which become very abundant in June; they shelter under the bell and appear to swim among the tentacles without coming to any harm. When almost 6 cm in length, they go down to the bottom. In their first 2 years, they are found in greatest numbers between the latitudes of 55 and 57° N. Like the hake and cod, the whiting feeds largely upon smaller fish, but at times may also take large quantities of the more active swimming crustaceans, such as shrimps, prawns, mysids, and euphausiids.

See also: **Fish**: Introduction; Catching and Handling

Further Reading

Gerking SD (1994) *Feeding Ecology of Fish*. San Diego, CA: Academic Press.

Gomes MC (1993) *Predictions under Uncertainty. Fish Assemblages and Food Webs on the Grand Banks of Newfoundland*. St John's, Canada: ISER, Memorial University of Newfoundland.

Hoines A, Bergstad OA and Albert OT (1995) The food wed of a coastal spawning ground of the herring (*Clupea harengus L.*) In: Skjoldal HR, Hopkins C, Erikstad KE and Leinaas HP (eds) *Ecology of Fjords and Coastal Waters*. Amsterdam: Elsevier Science.

Jobling M (1995) *Environmental Biology of Fishes*. London: Chapman & Hall.

Laevastu T (1996) *Exploitable Marine Ecosystem: Their Behaviour and Management*. Oxford: Fishing News Books.

Pelagic Species of Temperate Climates

G D Khedkar and B V Jadhao, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India

C D Khedkar, College of Dairy Technology, Warud (Pusad), India

N V Chavan, Adarsha College, Hingoli, Maharashtra, India

Background

The pelagic species found in summer throughout the waters of the European continental shelf are continually on the move in search of food; they feed on small fish such as sprats and sand eels, as well as on krill plankton. These species move between the surface and lower layers of the sea (2–400 m), depending upon the season and topographical features of the sea-bed. As surface-feeders, they have long provided a useful and easily caught source of food. Although the individual species are distinct, the group overall has a number of unifying features that tend to influence its exploitation and utilization by humans.

Mackerel

The scientific name of the mackerel caught in the north Atlantic is *Scomber scombrus*. Other species of *Scomber*, and species of *Scomberomours, Auxis*, and *Restrelliger*, are also described as mackerel in many parts of the world. The name mackerel is used in the UK principally for the species *Scomber scombrus*, but the name can also be used in the UK for any species of *Scomber* offered for retail sale. *Scomber scombrus* is sometimes called Atlantic, northern or Boston mackerel, to distinguish it from *Scomber japonicus*, which is variously called chub, Pacific, or Spanish mackerel. A small mackerel is sometimes called a cock mackerel, a joey or a little boy in England, or a pinner in Scotland.

Distinguishing Features

The Atlantic mackerel has a rounded, elongated body with two widely spaced back fins, the first with 11–13 spiny rays, the second with soft rays; the club mackerel has nine or 10 spines in the first back fin. There are four to six (usually five) finlets between the second back fin and the tail, and between the anal fin and the tail. There is a small keel on each side of the tail stub. The tail is forked. The back of the mackerel is a brilliant greeny blue, and the head is a steely blue–black with a small yellow patch behind the eye. There are 23–33 dark wavy bands across the back of the fish down to the midline. The sides have a silvery or coppery sheen, the belly is silvery white, and there is a broken black line just below the lateral line. The scales are small, and the skin feels velvety. There is a swimbladder in the chub mackerel but not in the Atlantic mackerel.

Life History

At the start of the winter, the mackerel migrate to overwintering areas. These are not well defined, but three are known around UK, one in the North Sea along the Norwegian Deeps, another along the edge of the continental shelf in the western Celtic Sea, and a third close to the south coast of Cornwall. The first two are typical overwintering areas and are similar to those off the east coast of North America; the mackerel are mixed with other species, for example scad in the Celtic Sea and Norway pout in the North Sea. Off Cornwall, the pattern is different. Large shoals of mackerel with few other fish among them move close inshore into relatively shallow water, although separate shoals of pilchard and scad are found in the area. The Cornish stock of winter mackerel was unknown before the mid-1960s, presumably because it was not there, and it is assumed to be an unstable offshoot of the stock in the Celtic Sea. Should the unknown factors that attracted the shoals suddenly change, there is always the possibility that they may withdraw just as suddenly from the area; nowhere else in the north Atlantic are adult mackerel found in a comparable position. Mackerel do not feed while overwintering, but they are not totally inactive. They form large dense shoals near the sea bed during daylight but rise and disperse during darkness. Occasionally, they disperse for no obvious reason and reform hours or even days later many miles from their previous position. Mackerel begin to move away from the overwintering areas in the spring to spawn and to start their feeding migrations. Spawning begins along the edge of the continental shelf in southern Biscay during February. By late March, spawning occurs throughout the Bay of Biscay and the Celtic Sea, and to the west of Ireland, and continues until June, but at all times, the greatest concentrations of eggs are found along the edge of the shelf. The older, larger fish start spawning first and, when spent, start swimming northwards along the edge of the shelf towards Shetland. Smaller, younger fish start spawning later and do not migrate so far. The youngest fish remain in the Celtic Sea throughout the summer. The spawning season in the North Sea is later, from May to August, and is most productive north of the Dogger Bank. Those that spawn south of the Dogger Bank are part of the western overwintering stock that spawns and feeds in the southern North Sea. Once spawning is completed, the North Sea mackerel mix with the older fish from the western stock around Shetland until the winter migration begins. Each female mackerel spawns on average about 300 000 eggs, which, when fertilized, are about 1 mm in diameter. Each contains a large oil globule that helps it to float in the upper water layers until it hatches. The larvae, 2–3 mm long, hatch after 4.5 days at 15 °C, or 8.5 days at 10 °C, and by the end of the first year, the young mackerel are 15–20 cm long. Growth continues to be rapid, and by the end of the first year, the young mackerel are 28–34 cm long. The growth rate is then considerably reduced, and mackerel 35–40 cm long may be anywhere between 5 and 20 years old. While a large fish is usually an old fish, one mackerel 57 cm long was found to be only 6 years old.

Length and Weight

An adult mackerel 30–35 cm long weighs 300–500 g before gutting. A joey mackerel, 20–25 cm long and probably in its second year, weighs 100–200 g. Mackerel weighing more than 1 kg are exceptional.

The Mackerel Fishery

Large, dense shoals may be found near the surface all round the UK in the summer and autumn, but their occurrence is sporadic in the southern North Sea.

The English summer mackerel fishery was traditionally from April to September; this has now been extended from February to October so that, coupled with the huge winter fishery now established in the southwest from October to March, mackerel supplies are available throughout the year. The Scottish mackerel fishery from June to August could also possibly be extended. Mackerel fishing methods have changed dramatically over the past 50 years. Before 1939, mackerel were caught mainly in the spring by drift net, an interim occupation for drifters between the winter and summer herring fisheries. After 1945, mackerel landings were small for some years, and the catch was taken mainly in the southwest during the summer by feathering, that is, small boats using hand lines with colored lures made from feathers or plastics. Once the mackerel shoals were found to remain in Cornish waters throughout the winter, the hand-line catch increased considerably in the late 1960s. As the market for mackerel improved, trawlers began to engage in mackerel fishing, using midwater mackerel in quantity off the west of Scotland in late summer.

Composition of Mackerel

The mackerel is a fatty fish, and the fat and water content vary with season. The fat content of mackerel caught off southwest England is lowest in May after spawning and reaches its peak between October and December after the fish have fed during the summer and autumn. A typical range of fat content throughout the year is 6–23%, and the protein content is 18–20%. The fat content of mackerel caught off the west coast of Scotland during August and September is similar to that of Cornish mackerel during December and January.

Herring

The herring, *Clupea harengus*, is the most plentiful fish found in the north Atlantic. *C. pallasi* is a closely related herring species found in Pacific. So important is it in sea life that it is given pride of place. It is known as a creature very well adapted for living in the upper water layers. Its blue–green back matches the color of the depths when viewed from above, and its silver flanks, seen from the side or below, mirror the varying lights and shades of the surrounding waters.

Distinguishing Features

The body of the herring is deeper than it is thick, and the length of the fish is about five times the greatest depth. The upper part of the body is dark blue–green or steel blue, the snout is blackish blue, and the sides and belly are silvery. The lower jaw protrudes slightly beyond the upper jaw. There is a single short back fin, a short anal fin near the tail, and a deeply forked tail fin. The pelvic fins are behind the start of the back fin, whereas on the sprat (*Sprattus sprattus*), they are in front.

Life History

Nearly all the other sea-fish of commercial importance lay eggs, which float freely in the water, drifting as part of the plankton; the herring is an exception and deposits its spawn in masses on the sea-bed, particularly on a gravel bottom, thickly plastering stones, shells and even the backs of crabs with it. As soon as the eggs are laid, they sink and stick to whatever they touch. In the summer when the fish are feeding, they are perhaps in smaller shoals, but now it seems that these combine together as the spawning time approaches. Some shoals are said to be 13–15 km long and 3–5 km wide. Each herring lays at least 10 000 eggs, and so the quantities of spawn produced by a shoal are enormous. The spawning grounds are an attraction to many bottom-living fish, especially haddock, which gorge themselves on this rich food. The eggs are fertilized in the water by the male herring, which discharge their sperm at the same time as the females lay their eggs. In North America, native Indians have long collected these eggs as a seasonal delicacy, and nowadays, an entire industry is centered on the collection of herring eggs for export, mostly to Japan. The eggs are about 1 mm in diameter and are incubated for 10–30 days depending on the sea temperature; 14–20 days is typical for the North Sea. The newly hatched fry or larvae, with yolk sacs attached, are 6–10 mm long and planktonic, and drift with the current. Since they depend upon nearby supplies of plankton for food, population levels can be directly affected by changes in the availability of the zooplankton. When the larvae reach a length of 40 mm, they begin to develop scales, take on a silvery sheen and move to inshore nursery grounds. Here, they are often caught, together with the young of other fish. The small fish now collect together in shoals, and tend to come in towards the coast into shallow water and sometimes large estuaries like the Thames or the Wash in the UK. In the Thames estuary, enormous shoals of whitebait are fished by so-called stow nets, which are huge conical nets lowered below anchored vessels and streamed out with the ebb and flow of the strong tides. After a sojourn of some 6 months or so in the estuaries and coastal regions, the herrings appear to scatter over the North Sea. They do not join the main shoals until they become sexually mature for the first time, and this, curiously enough, occurs in different individuals from the ages of 3–5 years.

Body Length

Most of the herring landed in the UK are between 230 and 300 mm long; herring caught off Norway and Iceland are often larger, up to 360 mm. Occasionally, herring can reach a length of about 430 mm, but this is exceptional.

Herring Fishery

In Scotland, there are several small fisheries at the beginning of the year. In the Minch, between the Outer Hebrides and the mainland, there is a fishery first for spent fish, then for recovering spents, and later on, in February and March, a stock of early spawning fish comes in. In the Firth of Forth area, there is a fishery for full and filling fish in January, extending to one for spawning fish in March. In Ireland, there are two fisheries for full and filling herring and later for spawners, one at the north and one in the southeast; both are over by the end of February. April and May is a lean period of poor fisheries, mainly for recovering spents. The small fishery at Lowestoft does not always materialize. In Ireland, too, they catch recovering spents in both the north and south in May, and sometimes in the south in April. During June to August, there is the great summer fishery in the northern North Sea for good-quality full and filling fish, the herring having nourishing themselves on the rich crops of zooplankton. In England, the drift-net fishery is confined to the North Shields area, where immature and recovering spents are caught in June, and full and filling fish in July and August.

The ring nets are employed for the herring fishery. A shoal of fish is located either by echo-sounding or by a feeling wire (a wire that gives a tremor when it is hit by the passing fish) and then surrounded by a net cast out as the boat circles around. Herring do not spend all their time near the surface; for a good part of the daytime, they are near the bottom, and a knowledge of this habit has led in recent years to another new and very considerable development: fishing for them with special fast trawls that skim over the bottom and have their head ropes kept very high by extra floats or underwater kites. The Germans were the pioneers in this and, for many years, have operated a large fleet of herring trawlers, working particularly on the Fladen grounds. During July and August, some of the companies in Hull, UK have also adopted this type of fishing in the North Sea. Fishing off the Yorkshire coast is in full swing during September, with landings at Whitby, Scarborough, and Grimsby, but by the end of September, all the drifters are concentrated at Yarmouth and Lowestoft ready for the great October and November fishery for the immense shoals of fine-quality full and filling that flood the East Anglian grounds prior to spawning in the Channel. This is the climax of the year; there is no more important herring fishery elsewhere in the world.

Processing and Preservation of Pelagic Species

Pelagic fishes are fatty and are not normally gutted at sea; they therefore spoil quickly unless they are chilled immediately after catching and kept chilled. They develop off odors after 1–2 days at 10 °C, become soft and spoil rapidly after 3 days, and become putrid after 5–6 days. Stale or noticeably spoiled fish should not be used for processing. In stale fish, the eyes are sunken, cloudy, and discolored red or brown. The skin loses its bloom, and its colors lose their intensity and brilliance, giving a washed-out appearance. The gills are dark red or brown, and a dark blood red mucus oozes from the gill covers. The odor of the gills and body is sour, sweaty, or strongly oily; sometimes a smell of ammonia is also present.

Chilling and Freezing

For freezing, the fish should be chilled immediately after capture and frozen within 24 h. Frozen fish should be properly glazed and kept in cold storage at −30 °C, which will keep it in good condition for at least 6 months. The cold-storage life can be extended up to 1 year if the fish is packed in polyethylene bags topped up with water and frozen in vertical freezers. The block is doubly protected against deterioration in storage by the encasing ice and packaging. Gutting before freezing is not necessary and may be a disadvantage. The fish can be hot-smoked to give a ready-to-eat delicatessen product, and the flesh of the hot-smoked fish can be used as a basis for a pâté. It can also be split or filleted and then kippered in cold smoke in a manner similar to herring. For a product of good eating quality, a fish with a fat content of at least 10% should be used. The smoked product can be whole gutted fish, with or without the head on, or fillets. Smoked whole fish is gutted, or headed, and gutted as required, the gut cavity cleaned, and the black bellywall lining removed. For hot-smoked fillets with skin left on are cut from the whole fish. Brining requires some care. In order to minimize the risk of food-poisoning organisms growing in the finished product, it is recommended that the salt concentration in the water phase of the product be at least 3%. This concentration, expressed in terms of water plus the solid components of the flesh, is less than 3%. To measure the salt concentration in the water phase, it is also necessary to measure the water content of the product. A salt concentration of 3% does

not render hot smoked fish unacceptably salty to the consumer. The rate of uptake of salt during brining depends principally on the size of the fish; large fish take longer to brine than small fish. Thus, to achieve a reproducible uptake of salt, all the fish in a batch should be of roughly the same weight. The brined whole fish can be tied by the tails in pairs and hung over tenter rods for smoking, or hung individually by the tails in keyhole slots cut in metal frames designed for the purpose. Fillets are laid on trays made of nylon-coated or stainless steel wire mesh. Full tenters, speats, or trays are loaded on to trolleys, which, when fully loaded, can be left to drain for 1–2 h or wheeled directly into the kiln without draining. In kilns with more than one trolley, the first loaded trolleys should be positioned at the air-outlet end, and the last loaded trolley at the air-inlet end. The smoking procedure in a mechanical kiln is as follows. With the kiln thermostat first set at 30 °C, the air-inlet half to threequarters open, and the main fan and chimney fan on, the smoke producer is set to produce maximum smoke, the recirculation damper being set to balance the smoke pressure in kiln. This predrying and smoking period takes about 1 h, during which time, surplus moisture dries off the surface of the fish, and the skins begin to set. The thermostats are then reset at 50 °C, the air inlet reduced to a quarter open, and the smoke pressure balanced by resetting the recirculation damper. The temperature and humidity in the kiln begin to rise, and the fish start to cook; high humidity prevents excessive weight loss. Once the temperature reaches 50 °C, it should be kept at that temperature for 30 min. The skin of the fish then becomes firm, and color begins to develop. For the final cooking stage, the order of the trolleys is reversed. The thermostats are reset at 80 °C, the air inlet closed, and the recirculation damper opened only just enough to relieve any back-pressure on the smoke producer. By the time the kiln temperature reaches 80 °C, small fish need a further 40–45 min to complete the process; larger fish need about 75 min. As a rough guide, the whole smoking process should take about 3 h. The fish attain a dark golden brown color, and the skin is dry with a silky sheen. The flesh should be completely cooked; the thickest part of the flesh at the shoulders of a whole fish should be opaque, with no jelly-like flakes.

Canning

These fishes may be canned as steaks, as fillets in a variety of sauces, or as a cold-smoked product. To make canned steaks, the whole fish are headed and gutted, and the belly cavity is cleaned out. The black skin is removed from the belly wall, and traces of blood along the backbone are brushed away. The fish are then cut into steaks of a length to suit the pack. The pieces are packed vertically in the can to give a product that resembles a middle-cut steak of a large fish. After adding 3 g of salt, the can is closed and heat-processed for 90 min at 115 °C. Vegetable oil or tomato sauce can be added to the can before closing. The finished cans should be stored for about 1 month before labeling and distribution; there is no further significant change in flavor or texture during storage after that time.

Salting

Quantities of fish were at one time pickle-cured in barrels for export, and small amounts are still packed in this way. A typical process is as follows. Fresh fish are split along the back and opened with a jerk to break the ribs. For hot-smoked fillets, skin left on is cut from the whole fish. Guts, gills, and any protruding bones are removed, and blood is washed away from the backbone. The fish are then soaked in fresh water for up to 2 h, rinsed, drained, and then dipped in dry salt. The fish are packed in barrels in circular fashion with tails to the center, and salt is sprinkled on each layer. The first two or three layers are packed skin downwards, and successive layers skin upwards. About 1 kg of salt to 3 kg of fatty fish or 4 kg of lean fish is required throughout the pack. When the barrel is full, it is closed, laid on its side, topped up with brine through the bung, and left for 10–12 days, further brine being added, if required, to keep the barrel full.

See also: **Canning**: Principles; **Fish Farming**; **Freezing**: Principles; **Preservation of Food**

Further Reading

Bligh EG (1992) *Seafood Science and Technology*. Oxford: Fishing News Books, Blackwell Scientific.

Burt JR, Hardy R and Whittle KJ (1992) *Pelagic Fish: The Resource and its Exploitation*. Oxford: Fishing News Books, Blackwell Scientific.

Grahame J (1987) *Plankton and Fisheries*. London: Edward Arnold.

Gregoire F (1993) Biological characteristics of Atlantic Mackerel (*Scomber scombrus L.*) sampled along the Canadian coast between 1983 and 1991.

Lockwood SJ (1988) *The Mackerel – Its Biology, Assessment and the Management of a Fishery*. Oxford: Fishing News Books, Blackwell Scientific.

Scott WB and Scott MG (1988) *Atlantic Fishes of Canada. Canadian Bulletin of Fisheries and Aquatic Sciences Number 219*.

SFIA (1989) *Pelagic Species Review – Resources and Markets. Report No. 2002*. Edinburgh: Sea Fish Industry Authority, Market Development Department.

Tuna and Tuna-like Fish of Tropical Climates

G D Khedkar and B V Jadhao, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India
C D Khedkar, College of Dairy Technology, Warud (Pusad), Maharashtra, India
N V Chavan, Adarsha College, Hingoli, Maharashtra, India

Background

Over the last two decades (1981–2001), tuna has become one of the most commercially important fish species, a source of foreign exchange earnings for many developing countries. Initially, larger fishing nations dominated the industry, but with the introductin of the 200-mile (roughly 320-km) exclusive economic zone, an increasing number of developing countries now take advantage of their tuna resources. Important fisheries exist throughout tropical and temperate areas of the world's oceans. Although tuna constitutes less than 3% of the world fish production, it is one of the most valuable fisheries. The southwest Pacific is the main tuna fishing area and accounts for one-third of total catches. The eastern Central Pacific was overfished in the early 1980s, leading to a drop in production in 1984. There was a subsequent fleet movement to the southwest Pacific. As a result, yellowfin resources recovered. Tuna has many uses in the Western nations. A number of tuna utilization possibilities have emerged in the tropics. Internationally, tuna is traded fresh, frozen, or canned. Japan remains virtually the only major market for fresh tuna, much of which is consumed as sashimi. Tuna for the sashimi market is deep frozen to temperatures below −40 °C. The sashimi market is sensitive to quality-size, fat content, and method of handling the fish. Normally, larger fish with a higher fat level are preferred. Most of the frozen tuna internationally traded is eventually canned. Canned tuna is generally presented in the following product forms: solid, chunk, flakes, grated, and smoked. A small quantity of tuna is traded for use as pet foods, oil, and tuna meat. Principal uses of internationally traded tuna are canning, other forms of processing, or special Japanese products. Different parts of the fish are used for different product types. For canning, the firmer loins are used for 'solid pack,' and those less firm or broken are used for 'chunks.' The scrapings, which are not blood, meat, or skin, are collected and sent to the flake or grated lines.

Table 1 lists the characteristics of several commercially important tropical species of tuna.

Characteristics of Major Tropical Species of Tuna

Skipjack (*Katsuwonus pelamis*)

Skipjack is usually 45–80 cm in length and 3–6 kg in weight (Figure 1a). This species is primarily frozen and used as a raw material for canning. In Japan, skipjack is also smoked and dried for katsuobushi production. The landings of this fish have increased steadily over the last 40 years from 25% of world tuna production in 1965 to 32% in 1975, 48% in 1988, and 51% in 1997. Skipjack replaced yellowfin as the most important component of total tuna landings in 1972 and has since consistently accounted for the majority of total landings. Japan has traditionally been the most prominent producer of this fish, but in recent years, the USA, France, and Spain have increased their operations. Several developing countries such as Indonesia and the Philippines have also emerged as important producers. Thailand has emerged as a major market for fresh and frozen skipjack in recent years, importing an estimated 250 000 tonnes in 1988 and surpassing the USA as the world's most important frozen skipjack import market. Major markets of canned skipjack are the USA, UK, and Germany.

Yellowfin Tuna (*Thunnus albacares*)

Yellowfish tuna is usually 70–150 cm in length and 7.2–150 kg in weight (Figure 1b). A major use of this fish is freezing and processing as canned fish for markets in western Europe, the USA, and Japan. Japan also uses yellowfin on the sashimi market. The landings of this fish have increased in recent years, due to new purse seine fisheries in the Western Indian Ocean and the Pacific. The largest importer of frozen yellowfin is Italy. In 1987–88, Italy imported 100 000 tonnes as a raw material for its canneries. Japan and the USA are the second and third most

Table 1 Some commercially important tropical species of tuna

Common name	*Scientific name*	*Size*	*Weight*
Yellowfin	*Thunnus albacares*	160–200 cm	40–130 kg
Bigeye	*Thunnus obesus*	90–180 cm	20–80 kg
Bluefin			
Northern bluefin	*Thunnus thynnus*	160–200 cm	40–130 kg
Southern bluefin	*Thunnus maccoyii*	160–200 cm	40–130 kg
Skipjack	*Katsuwonus pelamis*	48–80 cm	3–6 kg
Albacore	*Thunnus alalunga*	40–90 cm	4–15 kg

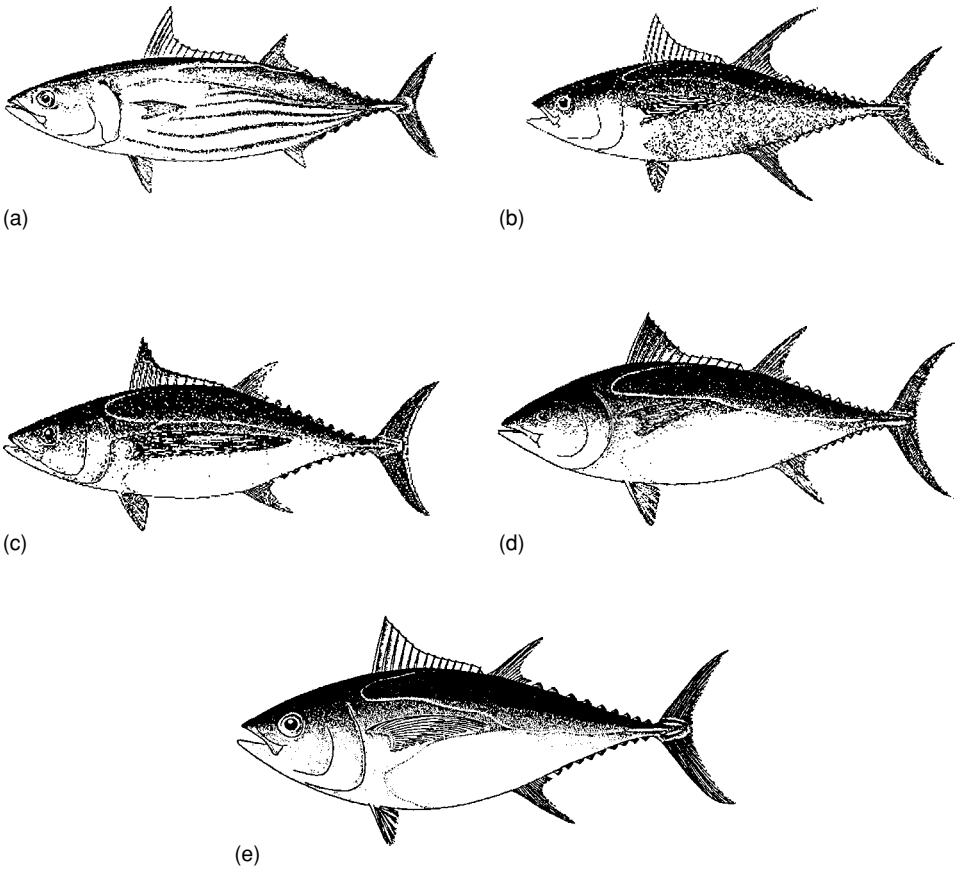

Figure 1 Commercially important species of tuna from tropical climates. (a) Skipjack (*Katsuwonus pelamis*); (b) Yellowfin tuna (*Thunnus albacares*); (c) Albacore (*Thunnus alalunga*); (d) Southern bluefish (*Thunnus maccoyii*); (e) Bigeye tuna (*Thunnus obesus*).

important markets: Japan imported 50 000 tonnes annually in addition to its domestic production during the 1980s, while the USA imported 40 000 tonnes.

Albacore (*Thunnus alalunga*)

Albacore usually weighs 4–15 kg, with a length of 40–90 cm (Figure 1c). This is a cosmopolitan fish, occurring in tropical waters of all oceans. At least two stocks (northern and southern) exist in both the Atlantic and the Pacific oceans, with little or no interchange across the warm equatorial waters. The northwest Pacific is the most important fishing area, followed by the northeast Atlantic and eastern Central Pacific. The landings of this species varied between 180 000 and 225 000 tonnes from 1975 to 1988, but a general decline in annual production is evident. Japan and Taiwan each account for approximately one-third of the world's annual albacore production. Most of the albacore production is in canning in the USA as solid whitemeat tuna. The USA is a major importer, averaging 70 000 tonnes annually. Some albacore is sold as fresh and canned fish, and consumed in local markets in France and Spain.

Southern Bluefin (*Thunnus maccoyii*)

Southern bluefin typically weighs 40–130 kg, with a length of 160–200 cm (Figure 1d). It occurs in southern oceans at a latitude of 39 °S. The fishing areas south of Australia, in the Tasman Sea, and in the Eastern Indian Ocean account for 70% of total annual production. The catch varied between 33 000 and 47 000 tonnes annually during 1975–88. Management measures were introduced by Australia in the 1980s to prevent overexploitation of the species, and production levels have since declined to 19 000 tonnes in 1989. Australia and Japan are the principal producers, consistently accounting for more than 85% of total production worldwide. The fish is generally sold fresh on the sashimi market in Japan, where it may fetch as much as US$35–45 kg^{-1}.

Bigeye tuna (*Thummus obesus*)

Bigeye tuna typically weighs 20–80 kg and is 90–180 cm in length (Figure 1e). This species is

distributed in tropical and subtropical seas but is absent from the Mediterranean. World production of this fish during 1980–90 was approximately 200 000 MT annually. The eastern Central Pacific has traditionally been the most important area, accounting for a third of the total bigeye production. Japan, South Korea, and Taiwan are the only producers with selective fishing gears targeting bigeye, and Japan is the largest consumer of bigeye. Although some of the production is frozen, most is sold as fresh fish and sold on the sashimi market.

Quality Requirements for the Tuna Market

The raw material must be handled carefully so as to prevent bruising and damage to the flesh. Thus, long lining is the best method of catching tuna for export, since damage is minimal. The flesh of fish killed under stress undergoes chemical changes, a condition called 'burning,' which causes the flesh to become mushy and inedible. Speed and accuracy are needed in processing tuna for export. Fish has to be bled, gilled, gutted, and chilled rapidly. Tuna for the sashimi market is graded not only on visual characteristics such as a bright/clear skin, clear and moist eyes, elastic skin, and undamaged smooth abdominal walls but also on the fat content of the flesh. A small flap is cut in the tail region to inspect the color and feel for the oil content.

In most buyer countries, tuna imports are generally covered by food-safety regulations applicable to fresh/frozen fish. The limits for heavy metals vary: in Italy, the upper limit for mercury is 0.7 mg kg^{-1}, and in Japan, it is 0.4 mg kg^{-1}. The permissible limit for cadmium is 0.5 mg kg^{-1} in most buyer countries, while lead levels vary from 0.5 to 2.0 mg kg^{-1}. Internationally, traded tuna is also tested for histamine, a tuna muscle protein that causes vomiting, diarrhea, and skin irritation if ingested at high levels. For most buyer countries, the permitted levels vary from 200 to 500 mg kg^{-1}.

Internationally Traded Tuna Products

Internationally traded tuna products include the following

- chilled, deheaded and gutted, or gilled and gutted;
- frozen, deheaded, and gutted, or gilled and gutted;
- canned, 'solid packs' (a mixture of pieces of tuna);
- canned, 'chunks' (a mixture of pieces of tuna);
- canned, 'flakes' (smaller pieces of tuna);
- oil and meals;
- pet food and animal feed.

The canned products are packed in oil, brine, or sauce. A range of flavoring and seasoning additives are used in canning, including salt, monosodium glutamate, hydrolyzed protein, and various spices. Smoked tuna products are also internationally traded to a small extent.

Sashimi

The Japanese tuna market is distinct. It is centered on raw consumption in a form called sashimi. The sashimi market demands large species such as bluefin, bigeye, and, to a lesser extent, yellowfin. The sashimi market is sensitive to quality-size, fat content, and method of handling the fish. Normally, larger fish with a higher fat level are preferred. The highest-quality northern bluefin is used exclusively for high-class restaurants, while the lower-quality product is used for household sashimi. Southern bluefin is used mostly by the restaurant trade, while bigeye use is split between restaurants and households. Sashimi processing involves progressive cutting of the flesh into cuts and portions illustrated in Figure 2. Different sashimi cuts (Figure 3) from different species have various market values, depending on the fat content. Those with a higher fat content have a lighter color. Prime sashimi is 'otoro' (pink), followed by chutoro (darker pink). Otoro and chutoro are types of toro and have a fat content of about 25%. Inner muscle blocks, more reddish in color with 14% less fat are marketed as akami. Prices vary according to size, season, freshness, moisture content, flesh texture, and taste.

Dried/smoked tuna products, called fushi in Japan, constitute a smaller but important percentage of Japanese tuna products. Katsuobushi refers to skipjack fillets that have been boiled then broiled over charcoal several times, dried, and then inoculated with a mold during processing. Shavings of katsuobushi fillets are marketed as kezuribushi. Arabushi is

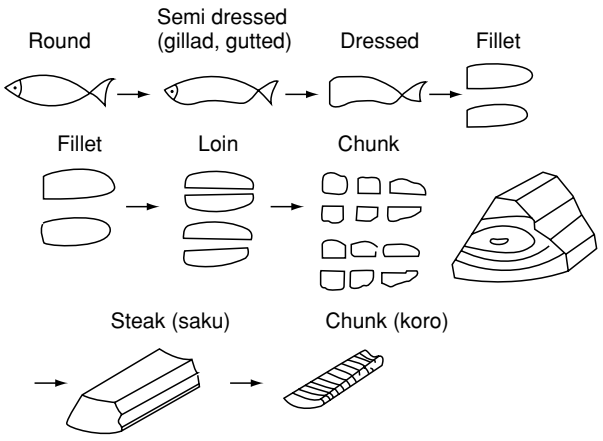

Figure 2 Stages in the preparation of sashimi portions.

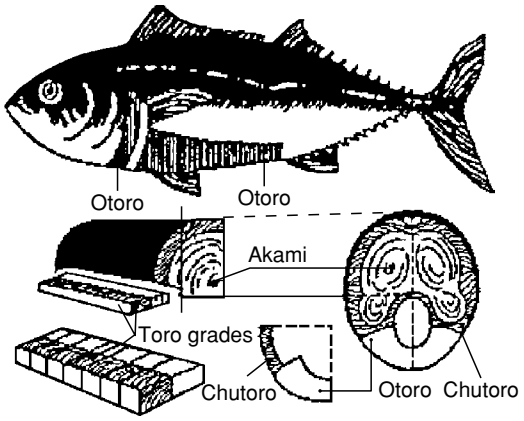

Figure 3 Distribution of cuts in tuna.

produced by drying and smoking skipjack loins, using fish caught by pole and line. Namaribushi is similar to arabushi, but the basic softness of the product is retained by reduction of smoking period. Imports of fishery products are in the form of arabushi, with the final processing including mold inoculation being done in Japan. Other products include tsukudani, a tuna pickle made out of skipjack flesh, boiled and seasoned with a solution of soya sauce and other flavoring materials. Shiokara, a fermented product, is produced from the viscera, stomach, or low-fat meat of skipjack by a process of salt fermentation.

Nutritive Value of Tuna

Tuna is very rich in high-quality protein. It is richer than beef or chicken and is one of the best sources of amino acids (Table 2). Tuna flesh contains substantial quantities of vitamins A, B_{12}, and D. It is also a very rich source of phosphorus, iodine, and fluorine. Studies of tuna have shown that *Katsuwonus pelamis* muscle consists of 69–72.6% moisture, 22.4–25.5% protein, 1.6–2.6% lipid, and 1.5–2.4% ash; a lipid content of up to 4.1% has been reported for some temperate species. Anatomical fractionation has yielded 69.7% meat and a total waste (head viscera) of 30%, a meat–bone separation has yielded 84% mince, and after heat processing and canning, the amino acid content has been found to remain essentially the same except for histamine and the sulfur amino acids. Hence, the protein quality can be said to be unaffected by precooking and retorting of canned tuna. Thiamin shows about 5% retention in the canned product. Nicotinic acid and riboflavin have also shown some reduction, and minerals assayed have shown a low retention. These losses might be due not only to heat destruction but also leaching of vitamins and minerals into the liquid portion of the canned product.

Table 2 Comparative appraisal of nutrients in raw and canned tuna

Nutrients	*Raw tuna*	*Canned tuna*
Essential amino acids (mg per 16 g of N)		
Histidine	6.59	5.07
Isoleucine	4.49	4.47
Leucine	8.24	8.33
Lysine	8.79	8.67
Cystine	1.62	2.72
Phenylalanine	3.62	1.47
Threonine	4.57	3.81
Tryptophan	4.57	1.19
Valine	5.31	5.33
Vitamins and minerals (per 100 g)		
Thiamin (μg)	189	11
Riboflavin (μg)	305	152
Nicotinic acid (mg)	28	19
Sodium (g)	0.27	2.39
Potassium (g)	0.91	0.63
Calcium (mg)	2.94	0.80
Iron (mg)	8.76	5.59
Copper (mg)	2.54	0.63

Source: Johnson AM (1993) Tuna and tuna-like species of tropical climates. In: Macrae R, Robinson RK and Sadler MJ (eds) *Encyclopaedia of Food Science and Nutrition*, pp. 1846–1850. London: Academic Press.

Global Supply Trends for Tropical Tuna

The world catch has steadily increased since the 1980s due to fishing by fleets in Latin America and Asian developing countries, the move of French and Spanish purse seiners to the western Indian Ocean, and increasing exploitation of western Pacific resources. Japan is the largest producer, even though its catch of high-value species has declined. The catches of Spain, France, Indonesia, Mexico, and South Korea have experienced the most growth. Skipjack continues to be the main species caught; 1.2 million tonnes were harvested in 1989. Yellowfin catches increased from 500 000 tonnes in 1980 to 905 000 tonnes in 1989. The major fishing grounds are in the southwest Pacific, accounting for about one-third of the total worldwide catch. Other major areas in order of importance are: eastern Central Pacific, western Indian Ocean, Northern Pacific and eastern Central Atlantic. The total canned tuna production increased more than 50% from 602 000 tonnes in 1980 to 1 099 000 tonnes in 1989. More than 80% of all catches are canned, 16% are consumed as sashimi, and the rest are cured. The leading country in canning remains the USA, which produces around 275 000 tonnes per year. Thailand's canned production grew the most during the 1990s. The share of developing countries in total canned production increased from 16% in 1980 to 44% in 1989. Trade in fresh and frozen tuna reached 1.3 million tonnes in 1989, representing about 75% of

all tuna trade. Increase in the Japanese demand for sashimi and canned yellowfin doubled fresh and frozen imports from 1980 to 1989. Thailand is now the largest importer of frozen tuna as it imports practically all the raw material for its canneries. Similarly, Italy and the Ivory Coast, both dependent on raw material imports, expanded purchases of frozen tuna. The main suppliers of fresh chilled and frozen tuna continue to be South Korea and Taiwan, mainly selling to the Japanese sashimi market. Prepared/preserved tuna exports reached 411 269 tonnes in 1989. Nearly 40% of total canned tuna production entered international trade in 1989.

Global Demand Trends for Tropical Tuna

Japan continues to be the main tuna-consuming country – sashimi consumption has declined over recent years, whereas canned tuna and katsuobushi have expanded. The total per capita supply has increased somewhat from 5.60 kg in 1980 to 6.35 kg in 1988–30% canned, 25% katsuobushi, and the remainder sashimi. Per capita consumption in the USA increased only slightly, from 2.4 kg in 1980 to 2.7 kg in 1988. In Spain and France, per capita consumption of canned tuna almost doubled. In Italy, canned tuna accounts for about 20% of total fish consumption.

Canned tuna is the most interesting product form to watch. Despite dramatic adjustments in the industry during the past decade, canned tuna's share of total consumption has remained at 80%. In Japan, consumption is likely to expand further on the wave of occidental food. By 2005, a per capita consumption of 2.7 kg is predicted. The USA, however, has reached saturation.

Prospects for Developing Countries

Loining operation is one of the prospects for developing countries, especially for those in the Caribbean and South America. In an attempt to contain production costs, US canners operating in Puerto Rico are actively considering options for loining tuna in lower-wage countries. Bumble Bee and Star-Kist are already importing frozen loins from Mexico, Equador, Costa Rica, and Thailand. The preparation of loins for canning is a labor-intensive operation. The cost squeeze for US canners could provide the impetus to perfecting the technology to allow loining to take place away from the cannery location. US canners have the advantage in that the main product form for lightmeat tuna is chunks. Therefore, quality control of imported frozen loins is relatively easy when compared to solid packs. The import duty of frozen tuna loins is 0%, the same as that for frozen whole tuna imports. Developing countries, especially those in the South Pacific should closely watch developments in loining technology, as this might provide a good opportunity to obtain a better yield from their resources.

See also: **Canning**: Principles; **Fish**: Processing; **Fish Farming**

Further Reading

Cushing DH (1982) *Fisheries Biology*. Madison, WI: University of Wisconsin Press.

Cushing DH (1996) *Towards the Science of Recruitment in Fish Populations*. Oldendorf Luhe, Germany: Ecology Institute.

Denman KL (1994) Physical structuring and distribution of size in oceanic foodwebs. In: Giller PS, Hildrew AG and Raffaelli DG (eds) *Aquatic Ecology: Scale, Pattern and Process*. Oxford: Blackwell Scientific.

FAO (1999) *Species Catalogues. No. 125, Vol. 2 Scombrids of the World; No. 125, Vol. 17(2) Clupeid Fishes of the World*. Rome: Food and Agriculture Organization.

Johnson AM (1993) Tuna and tuna-like species of tropical climates. In: Macrae R, Robinson RK and Sadler MJ (eds) *Encyclopaedia of Food Science, Technology and Nutrition*, pp. 1846–1850. London: Academic Press.

Levin SA, Powell TM and Steele JH (eds) (1993) *Patch Dynamics, Lecture Notes in Biomathematics 96*. Berlin: Springer-Verlag.

Powell TM and Okubo A (1994) Turbulence, diffusion and patchiness in the sea. *Philosophical Transactions of the Royal Society of London B* 343: 11–18.

Seuront L, Lagadeuc Y, Schertzer D and Lovejoy S (1999) Universal multifractal analysis as a tool to characterize multiscale intermittent patterns: example of phytoplankton distribution in turbulent coastal waters. *Journal of Plankton Research* 21: 877–922.

Stequert B and Marsac F (1989) *Tropical Tuna Surface Fisheries in the Indian Ocean. FAO Fisheries Technical Paper 282*. Rome: Food and Agriculture Organization.

Demersal Species of Tropical Climates

G D Khedkar and B V Jadhao, Dr Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India
N V Chavan, Adarsha College, Hingoli, Maharashtra, India
C D Khedkar, College of Dairy Technology, Warud (Pusad), Maharashtra, India

Introduction

Demersal fishes have been a major source of protein for people all over the world for thousands of years. Worldwide, about 85% of the total continental shelf area has sandy or muddy substrate. Only about 6% is rocky or gravelly, and the remaining areas are coral reefs or shell beds. However, there are both latitudinal and depth-related patterns. Corals are almost entirely confined to low latitudes where organically enriched muddy areas are also most extensive, particularly near the mouths of major rivers or below highly productive upwelling areas. Sandy, rocky, and gravelly sediments are more common at high latitudes. The tropical waters contain a greater number and diversity of fish species than waters of the higher latitudes. Demersal fish of economic importance are concentrated on the continental shelf and on the continental slope down to about 500 m. They live on or near the bottom of the ocean. A regular landing from a deep-sea trawling session would include catfish, croakers, soles, groupers, shinynose, snappers, mullets, and megalops. World catches of demersal species increased rapidly during the first three-quarters of the 20th century. Since then the catches from some stocks have declined, due to overfishing, while other previously underexploited stocks have increased their yields (Table 1). The limits of biological production have probably been reached in many areas. In order to maintain the fisheries and to protect the ecosystems which support them, careful management is needed.

Migration in Demersal Species of Tropical Climates

Diurnal migrations, either vertically or horizontally, are common in demersal fishes. More fascinating, however, are the long-distance seasonal migrations associated with reproduction, feeding, and sometimes overwintering. Many examples were already well described several decades ago, based on traditional tagging data. But nowadays scientists are employing the application of modern electronic tagging techniques, whereby perfect positional information can be derived after retrieval of the tag. The traditional tagging experiments have provided evidence of migrations across deep-sea areas. Even transatlantic crossings have been recorded for cod. Although such basin-wide migrations may not occur regularly, they certainly show that demersal fishes are quite capable of considerable movements over long distances.

Sensory Systems in Demersal Fishes of Tropical Climates

The shelf environment is generally well-lit and most demersal fishes have well-developed vision. In addition demersal fish may use mechanoreceptors (sense of pressure, motion, and sound), chemosensory (olfaction and gustation) and electrosensory (i.e., ability to detect electric fields) systems. Different systems may play essential roles during different stages of important processes such as prey detection, capture and handling, predator avoidance, or reproduction. In general, mechanoreception by the acousticolateralis system (lateral line organ and ear) tends to be most important for the detection of predators, conspecifics, and other physical disturbances in the environment. Olfaction is very important to benthofages that have to search the substrate for food, whereas planktivores

Table 1 Total world catch of major demersal fish species (1990–2000)

Scientific name	*Common name*	*Catch (tonnes)*
Ammodytes spp.	Sand eels	1 003 343
Ariidae spp.	Sea catfish nei	242 815
Cantherhines spp.	Filefish	234 446
Elasmobranchii	Sharks, skets, rays	337 819
Gadus macrocephalus	Pacific cod	425 467
Gadus morhua	Atlantic cod	2 317 261
Macruronus magellanicus	Patagonian grenadier	230 221
Macruronus novaezelandiae	Blue grenadier	255 421
Melanogrammus aeglefinus	Haddock	273 459
Merluccius capensis	Cape hakes	266 854
Merluccius gayi	South Pacific hake	197 911
Merluccius hubbsi	Argentine hake	526 573
Micromesistius poutassou	Blue whiting	628 918
Nemipterus spp.	Threadfin breams	186 201
Pleurogrammus azonus	Atka mackerel	237 343
Pleuronectiformes	Flatfishes nei	289 551
Pollachius virens	Pollock	385 227
Sciaenidae	Croakers	492 528
Sebastes spp.	Atlantic redfishes	318 383
Theragra chalcogramma	Alaska pollock	3 182 645
Trisopterus esmarkii	Norway pout	299 145

and piscivores tend to rely more heavily on vision. Gustation is most important after capturing prey. Benthofages may have elaborate adaptations of their olfactory system. In addition to a well-developed nose, they may have barbells on the snout (e.g., gadids, zoarcids) and modified and extended first rays of their pectoral fins (e.g., Triglidae) densly packed with olfactory sensory cells. The electrosensory system is particularly well developed in some sharks and rays, and they may primarily use this system for prey location, but also for orientation and navigation.

Life History Characteristics of the Demersal Fishes of Tropical Climates

Life history strategies are best examined in the context of the assemblage or community within which the fish operates. No species lives in isolation, but interacts with and influences the activities of all the other species it regularly encounters. Ecological understanding ultimately grows out of the study of communities, and this perspective is the one most deep-ocean ecologists have adopted. Comparative study of several systems, when possible, can be very insightful. The samples required for this sort of study, however, are not easily come by. Large numbers are required, which, in the best case, should encompass most life stages and time series over seasons, over years, and over space. A single sampling gear rarely suffices, and even a suite of different nets, for example, might also be complemented with cameras and other 'snapshot' devices. For deep-sea fishes, it is hard to find any area where such requirements are met. One that comes somewhat close is in the deep abyssal region of the eastern Atlantic.

Most demersal teleosts (bony fishes) are oviparous and produce free-floating eggs. Mating and spawning happen at the same time, often in mid-water. Some notable exceptions are the viviparous redfish (Sebastes) that release numerous pelagic larvae, and the wolfish (Anarhichadidae) which deposits a cluster of large eggs that is guarded until hatching. Sand eels (Ammodytidae) also have demersal eggs but do not protect the eggs. Most teleosts have no parental care other than making sure that they mate and spawn in an area and habitat that enhances the survival probability of their eggs, larvae, and pelagic juveniles. Most pelagic teleost eggs are small, i.e., a few millimeters in diameter, and fecundity is high, i.e., thousands to millions of eggs per female. Batch spawning is common in the highly fecund species, e.g., cod may spawn 10–13 batches of its very small eggs within a spawning season.

Demersal sharks are either oviparous or viviparous, whereas rays and chimaeroids are oviparous, attaching their few large encapsulated eggs to debris or macroalgae. All species produce rather few young in each batch, i.e., from a couple to a few tens of eggs/juveniles, and lifetime production is very low compared with most teleosts. Upon hatching or birth, the young resembles the adults. Spawning seasons vary in duration, and the general rule is that the more seasonal the production cycle, the more fixed and limited is the spawning season. High-latitude fishes and those living in monsoon or upwelling regions have comparatively short spawning seasons, whereas tropical and subtropical fishes have protracted seasons or year-round reproduction. Larvae hatched from small pelagic eggs have a yolk sac that may provide sufficient nutrition for a few days or weeks, after which they depend on exogenous feeding, usually on small crustacea such as copopod nauplii. The mortality in this perod is very high indeed, and only a minute fraction of the total number of eggs spawned will eventually result in a surviving demersal juvenile. The teleost postlarvae pass through metamorphosis upon reaching a certain size. Characters such as fins and juvenile pigmentaion develop, and sense organs also become fully developed. In flatfishes, the eye migration occurs. In general, the larva is transformed into a fish which is morphologically and behaviorally adapted for demersal life. Associated with this change is the settling on the seabed, at least for parts of the diurnal cycle. Settling areas are often, but not always, separated from the feeding areas of the older fish. This reflects the different habitat and food requirements of juveniles and adults, but may also reduce cannibalism. In the majority of teleost species, the demersal nursery areas are shallower than feeding grounds of older conspecifics. Estuaries and offshore shoals are typical nursery areas, also rocky shores where the substrate and macroalgae offer protection and a variety of prey. Tidal flats and sandy beaches are typical habitats of juvenile flatfish. A gradual ontogenetic shift in depth distribution happens as the juveniles grow larger.

The expected longevity of demersal species of tropical climates varies greatly. A general pattern is that longevity increases with increasing adult size, but there are many exceptions to this rule. Some species, i.e., redfishes, can probably live for at least 30–40 years, but this is unusual for shelf species. Small species such as Norway pout may live for 5–6 years at most, but in exploited areas this seldom happens. In most shelf waters, the fisheries have influenced age distributions to the extent that life expectancy is significantly reduced.

Somatic growth patterns also vary widely among demersal fishes of tropical climates. In fast-growing species living in strongly seasonal environments, the overall growth trajectory may show seasonality either because food supply and feeding rates vary, or because the somatic growth is influenced by reproductive activity/gonad growth. Growth rate usually declines significantly when the fish becomes mature, but never ceases entirely. Attempts have been made to classify species into *r*- and *K*-selected types (*r* and *K* are coefficients expressing rate of reproduction and somatic growth, respectively). The *r*-selected types are those with short life spans but fast growth and rather small adult size. They mature at a small size and young age, and can rapidly take advantage of short and unpredictable favorable environmental conditions by increasing their numbers. The *K*-species, however, tend to invest more energy in somatic growth by growing more slowly and maturing later in life, when they are capable of producing many young. The populations of such species comprise many age groups, and they are not, like the *r*-species, so vulnerable to repeated recruitment failures due to low survival of young over a range of years. There are many species that do not readily fit into these classes, and it is unusual for demersal species of tropical climates to have the extreme *r*- or *K*-strategies seen among epipelagic fishes and demersal deep-sea fishes, respectively.

Species Distribution

The vast extent and inaccessibility of the near-bottom oceanic environment have resulted in extremely patchy sampling and arguably therefore a quite restricted, and no doubt biased, view of overall demersal fish distribution and community composition. Slope areas in certain regions, especially some parts of the North Atlantic, are notable exceptions yet, on the basis of relative areas, the proportion of the abyssal floor of the world ocean that has ever been trawled is so minuscule that any evenly distributed species represented in collections by more than 20 specimens *in toto* is potentially more numerous across the globe than in humanity! The relative uniformity of physical features, such as temperature and salinity at considerable depth, could allow very widespread distribution patterns to occur. The detailed knowledge of the distribution of most of the demersal species of tropical climates, however, remains obscure. Such knowledge is largely derived from taxonomic and rare zoogeographic studies. Rarely does it come from studies of full assemblages where species, ecology, and interrelationships are taken into account. Yet it is these investigations that contribute most to our understanding of the deep demersal environment and the community ecology of fishes there.

Important Species of Demersal Fishes of Tropical Climates

The Clarias

Clarias gariepinus, also known as the African mud catfish, exists in the wild but it is also cultivated in ponds, cages, and pens and is of great commercial importance. This is an omnivorous fish with a preference for a planktonic diet. It also feeds on other types of food items such as insects, insect larvae, pupae, fish, and fish remains. It also has a propensity for being carnivorous. Propagation of these fishes is widely practiced in the tropics. Peak production of eggs by the female and hatching of the fertilized eggs are known to take place during the rainy season. A production mean of about 28 000 eggs and a hatching rate of 86.4% have been recorded during the rainy season. A low propagation yield at other times has been attributed to the resting period in the ovarian cycle of the animal. Induction of spawning has been carried out with 11-desoxycorticosterone acetate or carp pituitary suspension. Some of the processed products made from Clarias include fermented, dried, and hot-smoked products. The nutrient content of *Clarias* spp. is given in Table 2.

The Ariomma

Ariomma bondi and *A. melanum* are the two species. The former is found in 200–400 m depth of water; it is silvery and has a big head with small eyes. *A. melanum* inhabits 200–600 m of water, is brown in color, and has a small head with small eyes. The drift fish (*A. bondi*) is a relatively unexploited deep-sea fish found in tropical West Africa from Senegal to Gabon and in the western Atlantic from Nova Scotia through to the Gulf of Mexico. The length range is about 10–19 cm, with a corresponding average weight of 31.88 g. Size is an important technological property of a fish, of significance to its fishery and subsequent utilization. The nutrient content of *Ariomma* spp. is given in Table 2. The length–weight measurement, as well as the weight composition, has been found to

Table 2 Nutrient content of *Ariomma* and *Clarias* spp.

Nutrient	*Composition*	
	Clarias spp.	*Ariomma spp.*
Moisture	70.5–72.0	75.5–77.5
Protein	18.2–18.6	19.47–20.12
Lipid	6.7–7.1	3.32–3.39
Ash	2.6–2.9	2.67–2.71

meet the technological requirements of canning. In canned *Ariomma*, the edible portion increases from 51.66 to 59.38% when the whole, headless fish with the soft bones and fins are canned, consequently increasing the calcium and phosphorus content of the product. Canned *Ariomma* will be cheaper than canned tuna since it is relatively inexpensive and has a lower processing cost.

Demersal Fisheries

Demersal fisheries use a wide variety of fishing methods to catch fish on or close to the seabed. It is defined by the gear used, type of fishing activity, and varieties of fishes which are caught. A very wide range of fishing gear is used in demersal fisheries, the main ones being bottom trawls of different kinds, which are dragged along the seabed behind a trawler. Other methods include seine nets, trammel nets, gill nets, set nets, baited lines and long lines, temporary or permanent traps, and barriers. Catches from demersal fisheries make up a large proportion of the marine harvest used for human consumption and are the most valuable component of fisheries on continental shelves throughout the world.

Demersal fisheries have been a major source of human nutrition and commerce for thousands of years. Models of papyrus pair trawlers were found in Egyptian graves dating back 3000 years. The intensity of fishing activity throughout the world, including demersal fisheries, has increased rapidly over the past century, with more fishing vessels, greater engine power, better fishing gear, and improved navigational aids. Many demersal fisheries are now overexploited and all are in need of careful assessment and management if they are to provide a sustainable harvest.

The products of demersal fisheries are mainly used for human consumption. The species caught tend to be relatively large and of high value compared with typical pelagic species. Demersal fisheries are also known as ground-fish fisheries.

The five demersal marine fish with the highest average catches over the decade 1990–2000 are Alaska pollock, Atlantic cod, sand eels, blue whiting, and Argentine hake (Table 1). All these species spend a considerable proportion of their time in midwater. Sand eels spend most of their lives on or in the seabed. The vast majority of demersal fisheries take place on the continental shelves, at depths of less than 200 m.

Management of Demersal Fisheries

It is necessary to manage the demersal fisheries for biological, social, and economic purposes. Biological goals used to be set in terms of maximum sustainable yield of a few main species, but a broader and more cautious approach is now being introduced, which includes consideration of the ecosystem within which these species are produced and which takes account of the uncertainty in our assessment of the consequences of our activities. The most important biological goal is to avoid extinction of species. At a global level it is evident that economic goals are not being achieved because the capital and operational costs of marine fisheries are about two times higher than the gross revenue.

There are large numbers of examples of adverse social impacts of changes in fisheries, often caused by the effects of larger, industrial fishing operations on the quality of life and standard of living of small-scale fishing communities. Biological management of demersal fisheries has developed mainly from a single-species 'yield-per-recruit' model of fish stocks. The yield is controlled by adjusting the mortality and size of fish that are caught. In many shared international fisheries, such as those governed by the European Union, the annual allocation of catch quotas is the main instrument of fisheries management. This requires costly annual assessment of many fish stocks, which must be added to the operating costs when looking at the economic balance for a fishery.

The instruments for limiting the size of fish caught are mesh sizes, minimum landing sizes, and various kinds of escape panels in the fishing gear. These instruments can be quite effective, particularly where catches are dominated by a single species.

There are two classes of economic instruments for fisheries management: corrective taxes and subsidies and property rights. The latter demands less detailed information than the former and may lead to greater stakeholder participation in the management process, because it fosters a sense of ownership.

The management of demersal fisheries will always be a complex problem, because the marine ecosystem is complicated and subject to change as the global environment changes. Management will always be based on incomplete information and understanding and imperfect management tools. A critical step towards better management would be to monitor performance in relation to the target objective and provide feedback in order to improve the system.

One of the changes which have taken place over the past few years is the adoption of the precautionary approach. This seeks to evaluate the quality of the evidence, so that a cautious strategy is adopted when the evidence is weak. Whereas in the past such balance of evidence arguments were sometimes applied in order to avoid taking management action unless

the evidence was strong, the presumption now is that in case of doubt it is the fish stocks rather than the short-term interests of the fishing industry which should be protected. This is a very significant change in attitude, which gives some grounds for optimism in the continuing struggle to achieve sustainable fisheries and healthy ecosystems.

See also: **Fish Farming**

Further Reading

Cochrane KL (2000) Reconciling sustainability, economic efficiency and equity in fisheries: the one that got away? *Fish and Fisheries* 1: 3–21.

Cushing DH (1996) *Towards a Science of Recruitment in Fish Populations*. Oldendorf/Luhe, Germany: Ecology Institute.

Gerking SD (1994) *Feeding Ecology of Fish*. San Diego: Academic Press.

Gomes MC (1993) *Predictions under Uncertainty. Fish Assemblages and Food Webs on the Grand Banks of Newfoundland*. St John's, Canada: ISER, Memorial University of Newfoundland.

Gulland JA (1988) *Fish Population Dynamics: The Implications for Management*, 2nd edn. Chichester: John Wiley.

Hoines A, Bergstad OA and Albert OT (1995) The food web of a coastal spawning ground of the herring (*Clupea harengus* L.). In: Skjoldal HR, Hopkins C, Erikstad KE and Leinaas HP (eds) *Ecology of Fjords and Coastal Waters*. Amsterdam: Elsevier Science.

Jobling M (1995) *Environmental Biology of Fishes*. London: Chapman and Hall.

Kurlansky M (1997) *Cod. A Biography of the Fish that Changed the World*. London: Jonathan Cape.

Laevastu T (1996) *Exploitable Marine Ecosystem: Their Behaviour and Management*. Oxford: Fishing News Books.

Pelagic Species of Tropical Climates

G D Khedkar and B V Jadhao, Dr Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India
C D Khedkar, College of Dairy Technology, Warud (Pusad), Maharashtra, India
N V Chavan, Adarsha College, Hingoli , Maharashtra, India

Introduction

Of the total world catch of over 80 million tonnes of fish, roughly 36 million tonnes are accounted for by groups of typically pelagic fish. Of the total fish catch, approximately 25 million tonnes are used for reduction purposes and it is likely that a very high proportion of this comes from pelagic fish landed in bulk. The production of fish meal from oily fish is around 6 million tonnes, with an associated 1.4 million tonnes of oil, implying a conversion ratio of about 3.4:1. Allowing for approximations, these figures imply that about 70% of tropical pelagic fish caught are used for purposes other than direct human consumption. In addition to meal and oil, over 1.7 million tonnes of canned pelagic fish products is produced.

The landmass between the tropics of Capricorn and Cancer (32°28′S and N) is considered as the tropical region. Most of this region, barring some desert areas in Africa and the Indian subcontinent, is considered to be lush with vegetation and animals and is characterized by a climate predominantly determined by annual monsoon rains. From a socio-economic point of view, the tropics, in spite of its expected lushness, has some of the poorest and least developed countries in the world. The great bulk of mariculture in the tropics takes place on land-based ponds, which draw sea water through natural inlet channels, and in shallow bays. Unlike in temperate regions, offshore culture facilities using large floating and/or submersible cages are virtually nonexistent. In particular, the almost exponential growth of the shrimp culture industry in the last 15 years or so, and the adverse environmental influences this development has had on the coastal environment, and the future potential in mariculture development in the tropics as a means of closing the projected gap in supply and demand for aquatic products, especially for the upper end of the market, by year 2015, have created a fresh awareness of tropical mariculture.

Pelagic Environment

Microscopic planktonic organisms of the oceans form the basis of marine food chains and the more complex food webs. A bucketful of tropical sea water may contain several thousand microscopic plants and animals. Microscopic plant life includes all the members of the phytoplankton that provide food for herbivores. The phytoplankton is composed of various types of unicellular algae, which range from a few microns to several hundred microns in size. The major groups of algae include the diatoms, which are the dominant type of phytoplankton. It is essential for diatoms to remain in the surface waters to carry out photosynthetic processes, and therefore these algae have developed a variety of adaptations to prevent sinking. Other common types of phytoplankton include the dinoflagellate algae and coccolithophores.

Since all the oceans are connected with one another, and temperature appears to be the most important factor that limits the distribution of marine fish, there is basic similarity in the families throughout the tropical belt (Table 1). Pelagic fish are those which live at or near the surface of the sea. There are two categories of pelagic marine fish which form important fish resources in the tropics: those that occupy the open ocean, further from the shore and from the surface (to a mean depth of about 150 m), are generally large, swift, carnivorous species, such as tuna, as well as smaller species: the second category comprises fish which occupy the surface or near-surface waters of the continental shelf. These fish are also found in large inland lakes and reservoirs in the tropics, where species such as clupeids or cyprinids are important in commercial and artisan fisheries. These species have been introduced there because of an increasing demand for fish, and to allow better utilization of available water resources in such places (Table 2).

Important Characteristics of Pelagic Species

There are more kinds of surface-living fish in tropical and subtropical waters than those of higher latitudes. The pelagic species are the predominant catches in the tropics with a total exploitable potential of about 15–45×10^6 tonnes. Their common characteristics, including abundance and density, allow efficient mass fishing. Each species has different characteristics, which sometimes fluctuate with seasons. The continuity of fisheries for pelagic fish depends not only on the abundance of the stocks but also on their distribution. Pelagic fish, being more dependent on the body of water they are in than on the type of sea bed, show variable patterns of migration. It is in warmer seas that the larger members of the mackerel family (Scombridae) – the tunnies, bonitoes, and albacores – abound, likewise Coryphaena, the sailors' 'dolphin' and shoals of flying fish (Exocoetidae) and the smaller lantern fish (Myctophidae), upon which these larger fishes prey. The opah or moonfish (Lampris), the larger sunfish (Mola), and the swordfishes (Xiphiidae) are also typical of milder conditions. The northern boundary of these subtropical waters is usually taken to be a line at the sea surface having a mean annual temperature of 12 °C; this isotherm, as such lines are called, skirts the south-western tip of Ireland and the entrance to the English Channel, so that from time to time in summer some of these exciting exotics stray into cooler parts. One of the tropical species which shows a marked variation in chemical

Table 1 Distribution of some pelagic species of commercial importance in the tropics

Species	*Common name*	*West African (Atlantic Ocean)*	*East African (Indian Ocean)*	*Central American (Atlantic Ocean)*	*The Caribbean (Atlantic Ocean)*	*South American (Atlantic Ocean)*	*South American (Pacific Ocean)*	*Indian Ocean*	*Western Central Pacific Ocean*	*Eastern Central Pacific Ocean*
Engraulis spp.	Anchovy	+				+	+			+
Sardina pilchardus	Pilchard						+	+		+
Sardinops spp.	Sardine			+	+	+	+	+		
Sardinella spp.	Sardinella	+		+	+	+	+	+		
Ethamalosa fimbriata	Bonga	+		+	+	+				
Scomber spp.	Mackerel	+		+	+	+	+			
Caranx spp.	Horse mackerel	+	+	+	+	+	+	+	+	+

Reproduced from Johnson AM (1993) Pelagic species of tropical climates. In: *Encyclopaedia of Food Science, Technology and Nutrition*, vol. 3, pp. 1850–1856. London: Academic Press, with permission.

Table 2 Some pelagic species in inland lakes and reservoirs

Water body	*Main pelagic species*	*Yield characteristics*
Thailand (Ubolratana reservoir)	Clupeids	36% of total landing
Mexico	Atherinids	40–50 kg ha^{-1}
Surinam (Bokopondo reservoir)	Characid	Dominant open-water fish
Ghana (Lake Volta)	Clupeid	23% of landing
Lake Tanganyika	Clupeid	22.5 kg ha^{-1}
Zimbabwe	Clupeid	22 kg ha^{-1}
Lake Victoria	Clupeid	23 kg ha^{-1}
Nigeria	Clupeid	3140 t average biomass

Reproduced from Johnson AM (1993) Pelagic species of tropical climates. In: *Encyclopaedia of Food Science, Technology and Nutrition*, vol. 3, pp. 1850–1856. London: Academic Press, with permission.

composition is the West African shad (*Ethamalosa dorsalis*). This fish shows a range in fat content of 2–7% (wet weight) over the year, with a maximum in July. It has also been observed that the oil content of some of these species varies with size. A range of products can be obtained from the pelagic species of tropical region. The limitation in utilization is their small size; nevertheless, these are converted to low-cost products such as meal and oil.

Examples of Pelagic Species of Tropical Climates

Sardines

Sardine is a generic term applied to a number of different kinds of small saltwater fish which are prepared, cooked, and packed in a special way. Sardines are actually canned herring and the Maine sardine is the Atlantic herring, *Clupea herengus* (Figure 1). The edges of the scales on the belly of these fish are rough. These species lay their eggs on the seabed; Atlantic herrings usually lay their eggs at depths of 40–200 m and in some cases intertidally on gravel and small stones and the Pacific herring spawns on seaweed between tidemarks. Both species spawn on narrow strips. They mature at about 3–4 years of age and their annual fecundity amounts to 40 000–100 000 eggs. Throughout life the total fecundity would be about 10 times greater and only two survivors are needed to replace the stock. Hence the annual mortality each year of the eggs, larvae, and juveniles is high; indeed, it is approximately the inverse of the annual fecundity. The natural mortality of the adults is that sustained under predation in the absence of fishing and, as might be expected, it is rather difficult to establish; that of the herring might be about 10–20% of numbers per year.

These fish travel in huge schools and live in the open ocean, ranging from Greenland to North Carolina. The smallest of these herring, 12–17 cm in length, are dressed and canned as sardines, while larger ones are cut into small steaks and packed as 'fish steaks' or filleted and smoked to make kippered snacks.

Figure 1 The Maine sardine (*Clupea herengus herengus*).

Sardinella longiceps is the most important and widely distributed of all the species in the tropical environments. Sardines provide a tasty, low-calorie package. They provide valuable protein, minerals, vitamins, and cholesterol-reducing omega-3 fatty acids. Herring has been canned since the early 1780s when Napoleon recognized that there was a need to preserve food. Even today with the convenience of the refrigerator and the freezer, sales of canned food have been rising yearly. Sardines are also processed without preservatives as they are sterilized by heat in a sealed aluminum can.

The Maine Sardine

The Maine sardine is recognized as the world standard for excellence in sardines. Around the globe there are more than 20 varied species packed as sardines, but it is the tropical sardine that is unsurpassed for quality. This versatile little fish has long been an important resource to the state of Maine, longer in fact than Maine has been a state. Weirs, crude brushwood traps, were used to capture the small fish as they swam in shallow coastal waters. In the nineteenth century, Maine commercial fishermen adopted very similar fishing techniques and set stationary weirs in coves and along coastal areas that are often visited by young herring – the fish come in at night with the flood tide and are left trapped in the weir when the tide goes out.

Today, modern technology has refined the sardine fishing industry to include the use of large boats called seiners that take advantage of sophisticated detection equipment such as depth recorders, echo-sounding devices, and even light airplanes to watch for the flash of silver that marks nearby schools of small herring. The seiners use large nets to encircle schools efficiently, then a drawstring is pulled to close the bottom and trap the fish within. The use of modern-day seines gives fishermen the flexibility to follow the fish offshore rather than wait for them to enter stationary shoreside weirs. To insure that only the freshest product reaches the public, netted fish are quickly dispatched to modern canneries for careful processing.

The Maine sardine industry had its real beginning as a result of the Franco-Prussian War of 1870. At that time, for the US supply of the very popular oil-fried canned French sardine, it became increasingly important that the brand should meet high standards before being released for sale. Every case lost is meticulously inspected and carefully evaluated for strict adherence to 21 quality factors and it is operated under the rigorous Hazard Analysis Critical Control Point program monitored by the Food and Drug Administration. The nutrient content of tropical

sardines is as follows: moisture 65%; fat 14%; crude protein 18%; ash 1.7%; inorganic phosphorus 175 mg; and amino nitrogen 105 mg. These characteristics are important to the canning industry.

The Sprats

The sprats, *Clupea sprattus* (L.), are small clupeoids about 10–13 cm in length. They live for about 5–7 years, but the most abundant age group is the third year. Like other clupeoids, they feed on copepods and other plankton animals. They spawn in spring and summer, but in the northern Adriatic they spawn between December and March. They mature from the end of their second year to the end of their third year and lay between 10 000 and 40 000 eggs each year. The natural mortality is probably high. There are major spawning grounds in the southern North Sea and in southern Norwegian fjords. The eggs, larvae, and juveniles are fully pelagic. Sprats are found in the Baltic, in the North Sea, in the northern Adriatic and off Romania in the Black Sea. Fisheries were established in the Scottish firths, between Bergen and Stavanger in Norway, off Brittanny in France and around the English coast, particularly in the Wash and in the Thames estuary. In the main, these are winter fisheries worked with drift nets and stow nets, i.e., bag nets hung from fixed poles in the tidal streams.

The Mackerel

The common mackerel of tropical climates, *Scomber scombrus*, is the most important pelagic fish. It occurs in large shoals and is caught, as the herring is, in driftnets when it is in its plankton-feeding phase of spring and early summer. Mackerels are larger than herring, reaching as much as 40 cm in length or more. They spawn in the midwater in productive seas where the larvae and juveniles grow up, and they spend the rest of their lives there. They feed on copepods and other plankton animals. They swim quickly and some of the larger ones are predatory. The natural mortality of mackerels is perhaps high, up to 30% of numbers per year. They do not shoal as herring do but there may be small and transient aggregations.

The catch of mackerel has been increasing within the past few years. The catches have reached as much as one million tonnes from two stocks, one in the North Sea and the other to the west of the British Isles. They are found in the Mediterranean but catches have only amounted to about 30 000 tonnes each year. Considerable catches of Atlantic mackerel have been made off the eastern seaboard of the USA. The stock of Atka mackerel in the North-west Pacific yielded annually about 100 000 tonnes. The Pacific mackerel yielded large catches for a period, after which they declined. The horse mackerel (*Trachurus trachurus* L.) has usually been caught in trawls. It is somewhat larger than the common mackerel.

Mackerel meat has higher enzymatic activity than other fish, and the decomposition of substances related to adenosine triphosphate (ATP) is quick. The successful storage of raw mackerel for processing is, therefore, subject to the temperature used. Frozen storage at −18 to −30 °C could give a storage life of about 2–3 months. Storage with crushed ice (0 °C) would keep the fish fresh for about 48 h. Due to the higher variations in its fat content and its high enzymatic activity, mackerel present problems during processing and utilization. The nutrient content of fresh mackerel shows moisture content of 76%, crude protein 18%, fat 4%, carbohydrate 0.7%, and mineral 1.3%.

The Menhaden

The menhaden (*Brevoortia* spp.) comprise a group of species found in the tropical and subtropical west Atlantic. The most important species is the Gulf menhaden. *B. patronus* found in the Gulf of Mexico. Landings of menhaden showed a slow but steady increase up to 1993 (1.3 million tonnes) and have subsequently leveled off at a slightly lower level.

The Capelin

The capelin (*Mallotus villosus*), an arctic fish found north of the polar front, is caught off the Canadian seaboard, around Iceland and Greenland and in the northern parts of the Norwegian and Barents seas. Landings first became prominent in the mid-1960s and increased to a peak of 4 million tonnes in 1977. They have since decreased progressively to only just over a million tonnes in 1987, partly as a result of the collapse of the Barents sea stock.

Preservation of Pelagic Fish Quality on Board and Ashore

Overall, the world fisheries biological potential is about 120 million tonnes and fishery resources around the world are close to maximum catch limits. Many of the recent global changes in landings appear to be more the result of climatic fluctuations of stock size than of fishery development of management processes. There has been a steady increase in landings during the 1990s; 1997 (110 million tonnes) was slightly more than the record catch in 1996 (108 million tonnes). Much of the increase is accounted for by three pelagic shoaling species (Peruvian anchovy, South American sardine, and Japanese sardine) and the semi-demersal Alaska pollock. Pelagic fishing usually gives a single-species catch. When

caught in coastal fishing by smaller boats, mixed catches may demand species sorting as well as size grading. Traditionally, offshore fishing with big vessels has been for fishmeal and oil production, while the coastal fisheries have supplied the fresh-fish market and the processing industry, especially canning.

The increasing demand for high-quality raw material for all types of production has put pressure on fishermen to utilize all available knowledge and organize the processes on board in an attempt to optimize quality, value, and efficiency to satisfy the market. Special grades of fishmeal demand the highest degree of freshness, thereby requiring rapid and efficient chilling for this type of production. The interest in using small pelagics as raw material for animal feed, either as silage or as whole fish frozen in blocks, is also a challenge for the on-board handling and storage of fish. Byproducts are of increasing interest, e.g., roe production and collection of other raw material for biotechnological utilization. The choice of using ice or chilled/refrigerated sea water depends on fish species, the quantity of fish to be handled, and the type and value of product to be produced on board or ashore. The use of ice when boxing pelagic fish for human consumption was long regarded as uneconomical, but the development of methods for automatic addition of ice has made this more attractive. Likewise, improved methods of internal transport and storage using new methods for identifying and labeling each box, which can be monitored and recorded by microprocessor techniques, are important developments.

Filleting of fish on board is a specialized process and increases the product value, but sets certain limits on fishing capacity and often on processing flexibility. The advantage of on-board processing is that freshness can be assured. Also, the question of full utilization of the catch is important; today dumping is too easy and fish are wasted in an effort to store only the most valuable products.

Highly specialized vessels for surimi production based on pelagic fish have been developed but, so far, this venture has not been successful due to many technical problems and limited availability of suitable fish resources.

Chilling

Many of the pelagic resources are found close to the shores, giving small vessels the opportunity to participate in the fisheries. The need for chilling is important, particularly when temperatures are high (> 15 °C), while low temperatures may allow bulk transport if the distance is short. Sorting of fish is necessary ashore due to both damage and autolysis, which increases at high temperatures. Often only a small part can be utilized for consumption. Small vessels can improve handling and increase quality of the catch by using boxes or small insulated containers with ice or CSW. It is important to avoid anaerobic conditions in containers without circulation. Therefore the system of pumping air via perforated pipes into the bottom of the containers can be used. Many such installations have been tried out and proven to be beneficial. To reach a practical solution to benefit the fishermen, quality must be paid for; otherwise new investments will not be made. The Food and Agriculture Organization has been heavily involved in giving advice and is undertaking projects to increase the utilization of small pelagics for human consumption. In some cases when small pelagics are intended for curing or salting and sun drying, the process may start on board. Containers with brine, or chilled brine, can be used, starting the brining on board so that high-quality raw material can be brought back for sun drying. The best results are achieved when small fish are chilled and brined rapidly. Bigger fish (> 25 cm) should be gutted if the brine is not chilled.

Canning

In many countries canning is the most important utilization of small pelagics. The many different products of sprat, sardine, herring, mackerel, pilchard, and tuna are well accepted. This is important in the market due to good storage stability and high nutritional value. Fish for canning must be of the highest quality, and because most small pelagic fish are very delicate, damage due to handling often results in a high degree of postharvest losses. Also, the many old purse seiners operating without chilling make substantial losses, especially in areas with high temperatures such as India, California, Chile, or Peru and South Africa. Autolysis in combination with rough handling results in belly bursting, which is very dependent on storage time and temperature. Reducing storage time and temperature can greatly improve the quality of the raw material. Chilling is of the utmost importance and only a few hours' delay at 15–20 °C will reduce storage ability by several days. This means that the quality of landings of small pelagic fish for canning is dependent on season and fish species, on the chilling rate on board, and on the loading or unloading method. The bigger fish can often withstand rough handling. Disintegration of fish results in lower productivity in the processing plant. In a Mexican canning plant the quantity of sardines nabbed by machine decreased from 160 to 60 kg per person per hour when the proportion of belly burst increased from zero towards 100%. For the small fatty species,

oxidation must also be avoided. For fresh consumption and canning it is not a problem if chilling is done properly and if fish is utilized within 2–4 days, as is usually the case.

See also: **Fish**: Introduction; Catching and Handling; Fish as Food; Pelagic Species of Temperate Climates; **Hazard Analysis Critical Control Point**

Further Reading

Cushing DH (1982) *Fisheries Biology*. Madison: University of Wisconsin Press.

Cushing DH (1996) *Towards the Science of Recruitment in Fish Populations*. Oldendorf Luhe, Germany: Ecology Institute.

Denman KL (1994) Physical structuring and distribution of size in oceanic foodwebs. In: Giller PS, Hildrew AG and Raffaelli DG (eds) *Aquatic Ecology: Scale, Pattern and Process*. Oxford: Blackwell Scientific.

FAO (1999) *Species Catalogues* no. 125, vol. 2. *Scombrids of the World*; no. 125, vol. 17(2) *Clupeid Fishes of the World*. Rome: FAO.

Johnson AM (1993) Pelagic species of tropical climates. In: *Encyclopaedia of Food Science, Technology and Nutrition*, vol. 3, pp. 1850–1856. London: Academic Press.

Levin SA, Powell TM and Steele JH (eds) (1993) *Patch Dynamics Lecture Notes in Biomathematics 96*. Berlin: Springer-Verlag.

Longhurst A (2002–1) Pelagic biogeography. In: *Encyclopaedia of Oceans*. London: Academic Press.

Powell TM and Okubo A (1994) Turbulence, diffusion and patchiness in the sea. *Philosophical Transactions of the Royal Society of London B* 343: 11–18.

Seuront L, Lagadeuc Y, Schertzer D and Lovejoy S (1999) Universal multifractal analysis as a tool to characterize multiscale intermittent patterns: example of phytoplankton distribution in turbulent coastal waters. *Journal of Plankton Research* 21: 877–922.

Important Elasmobranch Species

J P H Wessels, Danhof, South Africa

Introduction

There are economically important fish species in both subclasses of the cartilaginous fish (class Chondrichthyes: **Figure 1**). These two subclasses are Elasmobranchii (previously referred to as Euselachii), which include sharks, skates, and rays, and Holocephali or chimeras. Holocephali differ from Euselachii in respect of the number of gills (they have one gill slit compared with 5–7 in sharks, skates, and rays) and in the fact that their teeth are fused to form plates. The economic use of shark and ray flesh and other portions such as livers is widespread and is not a recent development. Traditional oriental uses for flesh and fins abound: the flesh of many different sharks and rays are used in the western world, Central America and in some South American and African countries and traditionally by the Aborigines of Australia. Like other natural oils, shark oil was used as fuel for lamps, even in lighthouse towers.

Of the total 1994 world catch of fish (just under 100 million t), 730 784 t was reported to consist of sharks, rays, and chimeras. Of this quantity, the chimeras (elephantfish, or St Joseph shark) represented only 5032 t. More than 190 000 t of skates and rays were caught during 1994. Of the species identified, the following sharks were most plentiful in the world catch: *Carcharhinus falciformis*, the silky shark (15 400 t): *Mustelus schmitti*, the smooth houndshark (11 450 t); and *Squalus acanthias*, the spotted spiny dogfish (20 589 t). Catches recorded during 1994 were 1763 t for the basking shark (*Cetorhinus maximus*) and 1049 t for the porbeagle (*Lamna nasus*). The whiptail ray (*Dasyatis akajei*; 4041 t), the cuckoo ray (*Raja naevus*; 2822 t), the thornback skate (*R. clavata*, 1298 t), and the spotted ray (*R. montagui*; 827 t) were dominant catches amongst the rays. Catches recorded for the guitarfish species (*Rhinobatus percellens* and *R. planiceps*) were 1110 and 30 t respectively and for unidentified guitarfish 1559 t.

Table 1 lists the countries that catch and process sharks, rays, and skates. This table shows that fishing activities are not limited to specific geographical areas. **Table 1** also lists the major products of shark, ray, and skate exploitation. Apart from the countries shown, Ivory Coast and Senegal featured as producers of shark products in previous years, while Hong Kong is reported to have produced 3 t of dried salted shark fin during 1989. Shark oil and shark liver oil are recorded as products of Korea and the Maldives, where an estimated 20 t of shark oil and 18 t of shark liver oil were produced in 1989. A further 18 t of shark oil of unknown origin was traded.

Apart from the countries listed, considerable landings of elasmobranch fish were made in France and Brazil (more than 20 000 t in 1991), Argentina and Spain (more than 15 000 t in 1991), Australia, Nigeria, and the former Soviet Union (3100–7600 t in 1991). While Japan is the world's major producer of shark, ray, and skate products, trade statistics reveal that Hong Kong and Singapore are the major centers for trade in shark fins. A large portion of the world's elasmobranch catch is bycatch, i.e., a byproduct of

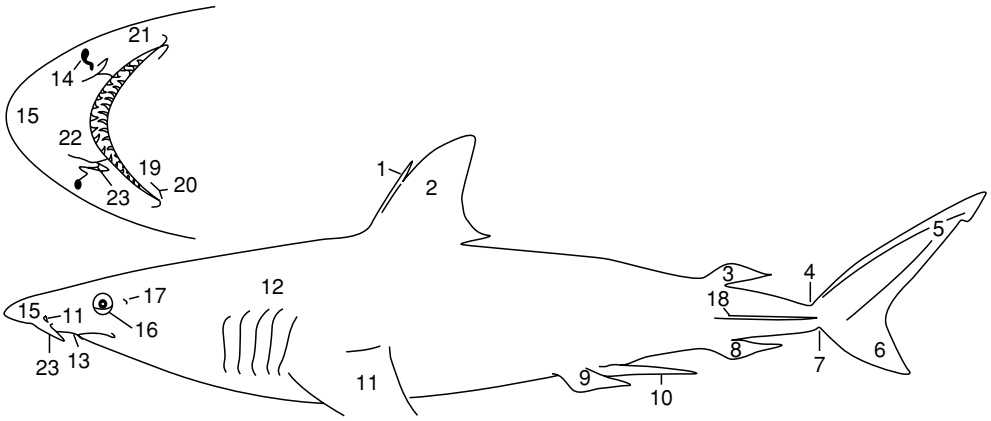

1. Dorsal fin spine.
2. First dorsal fin.
3. Second dorsal fin.
4. Notch at caudal base above.
5. Upper caudal lobe.
6. Lower caudal lobe.
7. Notch at caudal base below.
8. Anal fin.
9. Pelvic fin.
10. Clasper of male.
11. Pectoral fin.
12. Gill-slits.
13. Mouth.
14. Nostril.
15. Snout.
16. Nictitating membrane.
17. Spiracle.
18. Peduncular keel.
19. Lower lip.
20. Lower labial fold.
21. Upper lip.
22. Nasal flap.
23. Nasal cirrus.

Figure 1 Most of the features of a typical chondrichthyan fish. Not all of these features are found in any one fish.

some directed fishery such as a trawling, longline, treknetting, or seining industry. Where elasmobranch fishing is directed, use is made of lines (even handlines) and hooks.

Commercial Products Derived from Elasmobranch Species

Shark and Ray Flesh (Fresh, Chilled, or Frozen)

The smaller species of shark are generally used as sources of fresh, chilled, or frozen meat, while the larger sharks provide fins and hides. In the world context, the spotted spiny dogfish (*Squalus acanthias*), which grows to a length of 1 m and which, unlike most other sharks, is readily skinned, constitutes the major source of shark meat. The relatively small smooth houndstooth sharks (*Mustelus canis, M. manazo*, and *M. mustelus*), the blue shark (*Prionace glauca*, which grows to almost 4 m in length), the porbeagle (*Lamna nasus*, which reaches 3 m in length), and the mako shark (*Isurus glaucus*, of similar potential size as the porbeagle) are also important sources of shark meat. However, in the case of larger specimens of these species, the flesh would be used mainly as an ingredient of fish pastes, or sold as steaks, or included in smoked or minced form of bacalao. In Japan the flesh of the porbeagle, the gummy shark (*Mustelus* spp.), and the whiptail ray (*Dasyatis akajei*) are the only elasmobranch species readily eaten as fresh fish. It is reported that these species do not acquire an ammoniacal taste shortly after being caught.

The elephantfish (*Callorynchus* spp.) is a source of particularly fine flesh of a good white color which, like the flesh of some noncartilaginous fish, is lean. The low-fat content has consumer appeal in countries where there is an awareness of the detrimental effect of a high-fat intake. Low-fat levels of 0.1–0.3% are recorded in the flesh of various shark species. Low muscular fat levels are usually associated with high levels of fat storage in the liver; also in noncartilaginous fish. Certain of the smaller *Carcharhinus* species are of importance in some areas, e.g., the black-tipped shark (*C. limbatus* or *C. johnsoni*) and, in Italy and France, even the large dusky shark (*C. obscurus*) has importance. The tiger shark (*Galeocerdo cuvieri*), which grows to a very large size, is used as a source of shark flesh in several areas of the world. The thresher shark of the Caribbean (*Alopias vulpinus*) is a large shark: it's meat is perceived to be of very good quality and is said to resemble swordfish.

In Japan, it appears that little shark flesh is eaten as fillets or cutlets, but shark flesh has found its way into what might be called surimi-based products. For this purpose certain gel-forming properties are required. Of the major species mentioned above, three appear to have found use in chikuwa and kamaboko-type products: *Prionace glauca, Mustelus manazo*,

Table 1 Products (t) derived from sharks, skates, and rays produced by different countries during 1989

Country	Fresh or chilled fillets	Frozen fillets	Frozen sharks	Sharks dried, salted, or in brine	Sharks, rays, etc. (dried not salted)	Shark fins (dried, not salted)	Shark fins (dried and salted)
Chile			234				
China							563[a]
Colombia				645			
Denmark		1	55				
Faroe Islands			340				
Fiji						14	
Greece			100				
India							190
Indonesia						475	
Ireland			134				
Italy			1500[a]				
Japan			10 015				
Korea						36[a]	
Maldives						13	
Mexico			3224				
New Zealand	19	479	969				
Norway			105				
Pakistan				5250			317
Peru			518	678	1326		
South Africa			660				
Sri Lanka					221		54
Thailand				1000[a]	2500[a]		
UK			1108				
USA	1068	1518					
Uruguay			4			4	
Yemen						60[a]	
Unidentified						1179	

[a]Represents Food and Agriculture Organization (FAO) estimate from sources available to the FAO. Data from FAO (1991b) *Fisheries Statistics Commodities*. FAO Yearbook. FAO Fisheries Series 37, FAO Statistics Series 101. Rome: Food and Agriculture Organization.

and *Isurus glaucus*. Shark flesh is not necessarily used on its own in these products, but rather as a substitute for traditional white fish types. *Lamna nasus* and *Prionace glauca* are said to produce weak (nonelastic) gels, but the flesh of most sharks and rays has good elastic properties. Good gel-forming is associated with a high myosin content, myosin being that portion of the muscular protein that solubilizes in 1.2–1.5 mol l^{-1} sodium chloride solution.

As far as skates, stingrays, and guitarfish are concerned, there is no indication that specific species are more suitable for certain products. Whole gutted skate and skate wings, skin-on or skinned, and skate nobs (chunks of flesh), used fresh and in the frozen form, are world trade commodities. Stingrays are sold alive or fresh in Japan. In as far as rays tend to be more limited in distribution than sharks, there is a tendency for some species of rays to predominate in the catch from specific areas.

The flesh of skates and rays passes through rigor mortis more rapidly than that of cod, for example, and it is not inclined to suffer from the gaping phenomenon that occurs in some fish fillets.

Of 16 shark species studied as a source of flesh, some of the better known species have high fillet yields, in excess of 45% of the original body mass. These include *Squalus acanthias, Mustelus manazo*, and *Carcharinus falciformis*. The yield of the hammerhead shark (*Sphyrna tudes*), at 54.4%, is almost as high as the best, i.e., the blacktip shark (*C. limbatus*).

Canning of Shark Flesh

Some years ago a small shark-canning industry was set up in Murmansk, a port on the Barents sea.

Smoked and Salted and/or Dried Flesh of Sharks and Rays

Flesh of the shark species listed above is smoked and salted and/or dried in various countries. Some of the larger sharks appear to be only used in this manner. The hammerhead sharks (*Sphyrna tudes* and *S. zygaena*), which are not favored as a source of fresh meat, are used for the production of salted and dried meat in the Caribbean and South America. A notable exception is schillerlocken, a smoked product

produced in Germany from the belly flaps of a small shark, the dogfish; in addition, some smoked shark delicatessen products are made in Italy from shark of popular species and size. Shark flesh appears to be consumed primarily in salted form in some South and Central American countries, but it is also used in the fresh or chilled form and for freezing.

The flesh of the ray is hot-smoked, while a fermented product, subsequently salted, is made in Iceland. Rays of the genus *Gymnura* (a butterfly ray) are dried (unsalted) in Indonesia, while the ray is also used in this manner in Gambia, and guitarfish are dried and salted in India. (*See* **Smoked Foods**: Principles; Applications of Smoking.)

Mercury and Ammonia Formation in Shark Flesh

There are two aspects which cannot be ignored in a discussion on the use of the flesh from different cartilaginous fish. The first is the mercury content of the flesh of sharks, which is sometimes raised, and the second is the postmortem formation of ammonia.

Mercury Content

Relatively high contents of mercury, up to about $8\,mg\,kg^{-1}$ of flesh, have been recorded in shark flesh, while values of up to $1\,mg\,kg^{-1}$ are on record for skates, and $0.3\,mg\,kg^{-1}$ for chimeras (*Callorynchus* spp.). This chemical property is shared with other long-living, wide-ranging finfish such as swordfish and marlins. Cartilaginous fish, however, often contain very low levels of mercury. In one particular laboratory the mercury contents of 1003 samples of *Isurus glaucus*, measured as total mercury, ranged from 0.02 to 8.29 mg per 100 g, with a mean of 0.54 mg per 100 g of flesh, while 163 samples of *Galeorhinus galeus* yielded 0.2–3.75 mg per 100 g of flesh (mean of 0.88 mg).

Maximum limits for the mercury content of flesh – operative in the world fishery market at present – lie between 0.5 and $1\,mg\,kg^{-1}$ of flesh. As these animals are generally quite mobile, i.e., not restricted to narrowly defined geographic areas, one expects that the age of the fish would be the major determinant of the level of mercury that will have been taken up and retained by the body. This great mobility is thought to be one of the reasons for the high mercury content of shark flesh; they take in more water than would a slower-moving fish of similar size. Also, for reasons not understood, sharks secrete the heavy metal at a lower rate than they take it up. (*See* **Mercury**: Properties and Determination; Toxicology.)

Ammonia Formation

While ammonia production owing to bacterial action is a common feature of fish flesh, the effect is exaggerated in the flesh of cartilaginous fishes. These fish possess, in addition to the trimethyl amine oxide of fish blood and muscle, a further nitrogen-rich compound, urea (carbonyl amide), which functions in osmoregulation to maintain the fish in balance with its salt-water environment. Bacteria that come in contact with the blood or flesh of the fish after death break down these compounds to release amines and, in the case of urea, ammonia with its characteristic pungent 'fishy' smell; the release of ammonia is a first step in the breakdown process, resulting from bacterial urease activity. (*See* **Amines**.)

Bleeding of sharks is recommended as soon as possible after death, a step which is not practiced in some of the *Squalus acanthias* fisheries – hence the reddish color of the flesh in such cases. Rapid cooling by icing or freezing and strict hygiene are further recommended practices. Urea can also be removed by leaching with water or brine, while in the case of salted, dried flesh the urea content will have been reduced. Even if no bacterial action takes place during the earlier stages of salting which will result in urea breakdown, the leaching effect which accompanies salting results in considerable loss of urea from the flesh.

The urea contents of the flesh differ between and within species. According to one source, *Squalus acanthias* lies at the low end of the range with 1570 mg urea per 100 g of flesh and smooth hammerhead shark (*Sphyrna zygaena*) is at the higher end with 2038 mg per 100 g. Skate fall in the same class as sharks in terms of their urea content, in the range 1–2%. The few values available for St Joseph shark range from 1.4% to 1.8%.

Utilization of Shark Fins for Soup

One set of shark fins usually consists of two pectoral fins, the first dorsal fin (and in large sharks with two dorsal fins, the second) and the lower lobe of the tail fin. These fins are either dried or salted and dried and used in the Orient and by Chinese communities elsewhere to produce shark fin soup. Small sharks do not yield suitable fins, while some larger species do not produce fins of desirable quality. In general, however, the fins of any shark longer than 1.5 m can be used.

Of the sharks of suitable size, only nurse shark (*Ginglymostoma cirratum*) fins and the sawshark (*Pristiphorus mudipiunis*) pectoral fins appear to be of no commercial value.

The preferred shark species are the hammerhead, the mako, and the blue, mentioned above, and the tope or soupfin, *Galeorhinus galeus*. Fins from many of the large and the very large sharks are used, however, including those of the great white shark (*Carcharodon carcharias*), the fibrous meat of which, incidentally, is eaten in part of northern Japan. The very large tiger shark (*Galeocerdo cuvieri*) is also listed as a source of good-quality shark meat. Several *Carcharinus* species provide fins for processing in Pakistan.

Shark Hides for Leathermaking

Shark skins used for leathermaking must be removed from the fish in a very fresh state. Only sharks larger than 1.5 m are skinned for leather, with the exception of the small Pacific Ocean shark species, the brown cat shark (*Apristurus brunneus*), the hide of which is used to prepare the special boroso leather.

The nurse shark (*Ginglymostoma cirratum*) is the species of first choice, except in Japan, where preference is for the skin of the blue shark (*Prionace glauca*) and that of members of the Rhinobatidae family (guitarfish) which have relatively small skin dentiles. These species do not produce suitable hides as far as the shark leather industry of the USA is concerned. In the USA the tiger shark is seen as having a good hide. If properly cut and prepared, however, the hide from any other larger shark species is acceptable. This situation prevails since technical advances in shark hide tanning were made by the Ocean Leather Corporation of Newark in New Jersey, USA. These advances permitted the production of softer leather rather than the rough sandpaper-like skins produced earlier from shark skin.

The Food and Agriculture Organization has drafted a code of practice for the utilization of sharks for food and for skin production.

Sharks' Teeth

Sharks' teeth represent another potential product, similar to leather, in the sense that the end products are consumer goods. There is at present no established market for sharks' teeth, except on a small scale in some holiday resort areas, where mako shark teeth are sought after as tourist items.

Other Products

Other shark products that remain to be discussed represent the result of further extensive processing or are a nutrient source or products with pharmaceutical or other uses.

The Japanese Product Meikotsu

The soft cartilage of shark or skate is diced, boiled, and cooled in water, the remaining muscle and hard cartilage are removed, and the soft cartilage is again boiled and then sun-dried. This product is made in Japan for export to China. Shark cartilage in powder or tablet form is available for treatment of various cancers and autoimmune conditions such as arthritis. While there is strong evidence of dramatic beneficial effects in some instances, the medical fraternity seems to be skeptical.

Extraction of Chondroitin

Chondroitin is extracted from the hard and soft cartilage of the shark. The chondroitins are mucopolysaccharides composed of glucuronic acid and galactosamine units.

Shark corneas have been used to implant on to human eyes, at least in the USA and Australia.

Shark Liver Oil, Vitamins A and D, and Squalene

Shark flesh is virtually free of oil. On the other hand, the livers can contain high but variable levels of oil. The size of the livers varies greatly with the species; larger sharks tend to have larger livers while within species heavier sharks have relatively larger livers. (*See* **Fish Oils**: Dietary Importance.)

The liver constitutes about 17.5% of the mass of the large tiger shark (*Galeocerdo cuvieri*), against 2.9% of that of the soupfin shark, in this instance *Galeorhinus japonicus*. Some deep-water dogfish sampled in Australia contain relatively large livers. For the Portuguese shark (*Centroscymnus coelolepsis*) a mean of 27% of body mass was recorded and for the leafscale gulper shark (*Centrophorus squamosus*) 23%. There is seasonal variation in the relative liver mass and oil contents of the livers, with an upper limit of liver mass in the large sharks about 20% of the body mass.

Marine oils are usually used as raw materials for products such as baking fats and margarines. The oil of the cartilaginous fish is often not made up predominantly of triacylglycerols. The fatty acids in such nontriacylglycerol-rich oils are associated with diacyl glyceryl esters which may contribute considerably to the nonsaponifiable fraction of such oil.

The fatty acid composition of shark oils is not well documented in modern literature. Oil of the head or the liver of the spotted spiny dogfish (*Squalus acanthias*) has a relatively low content of saturated

fatty acids in comparison with most fish oils, i.e., 20% or less, and high levels of C_{20} and C_{22} highly unsaturated fatty acids. Head oil, for example, was reported to contain 10.4–15.5% eicosapentaenoic acid and 17.7–21.8% docosahexaenoic acid, the two fatty acids claimed to be dietary essentials. Amongst the best defined actions of these fatty acids in the human body are reduced blood pressure, thinning of the blood, reduced blood triacylglycerol levels and brain development and visual acuity in very young children. An oil (probably a body oil) from the porbeagle (*Lamna cornubica*) contained an even lower level of fully saturated fatty acids and a higher level of docosahexaenoic acid. The monoenoic fatty acid content of this oil was lower than that of the dogfish head oil (45–47%). Also, the liver of the dogfish is rich in monoenoic acid compared with the oil from the flesh of the same fish. Three elasmobranch species (*Carcharias melanopterus* 31%, *Galeocerdo cuvieri* 40%, and *Pristis cuspidatus* 37%, of the Rajiformes order) are listed as having oils with a high content of saturated fatty acids. (*See* **Fatty Acids**: Dietary Importance.)

Interest in shark oils lies not in the fatty acids of which the oils are composed, but primarily in the fact that some of the oils contain high levels of vitamin A and others high levels of a long-chain hydrocarbon called squalene.

The kitefin (*Dalatius liche*) and hammerhead (*Sphyrna tudes*) sharks with their large livers yield oil with high vitamin A potency, while the liver oil of the relatively small blacktip shark (*Carcharinus limbatus*) also has a high vitamin content. These vitamin contents are higher than those of other somewhat larger species of the same genus: *C. obscurus* (dusky shark), *C. plumbeus* (sandbar shark), *C. falciformis* (silky shark), *C. leucas* (bull or Zambezi shark), *C. brevipinna* (spinner shark) and *C. taurus* (sand shark, Taiwan). Smooth houndstooth shark (*Mustelus manazo*) and mako shark (*Isurus glaucus*) oils are also high in vitamin content, as are oils from the black shark (*Galeus glaucus*) and the soupfin shark (*Galeorhinus japonicus* and *G. galeus*). The vitamin content values are reportedly up to 100 000 IU of vitamin A per gram of dogfish oil and even up to 156 000 IU in oil from the liver of adult *G. galeus*. In general, values are only about 10% of these, with much lower values (< 100 and 180 IU) reported for the basking shark (*Cetorhinus maximus*) and the chimerid *Hydrolagus colliei* (ratfish). Compared with the liver oils of most finfish, the elasmobranch oils contain very low levels of vitamin D: 1 IU g^{-1} for the above-mentioned ratfish to 25 IU g^{-1} for the oil from the liver of the skate, *Raja inornata*. (*See* **Retinol**: Physiology.)

The importance of natural vitamin-rich oils in the diet of humans and farm animals has declined greatly since the last world war as a result of the successful synthesis at competitive costs of synthetic vitamins A and D. The focus of the shark oil market has therefore moved to Japan and some western countries, where squalene is extracted from suitable shark oils. After hydrogenation of the almost fully saturated hydrocarbon (C_{30}) to squalane, this material is used as an ingredient of cosmetics. (*See* **Cholecalciferol**: Physiology.)

Elasmobranch oils have been classified into three groups on the basis of their content of unsaponifiable matter. The first group is similar to the majority of finfish oils, in the sense that they have a low nonsaponifiable material content, i.e., up to 2%. In these oils, cholesterol would be the major nonsaponifiable material. The second class of elasmobranch oils contains 10–35% nonsaponifiable material which, in addition to cholesterol, largely consists of long-chain alcohols naturally present as diacyl glycyl esters. Ratfish oil is in this category. The third group is the high-squalene group. The nonsaponifiable material of the liver oils from the sharks in this group is obviously high.

The basking shark (*Cetorhinus maximus*) is one of the sharks of which the liver oil contains the highest content of squalene, i.e., 45–75% expressed as unsaponifiable matter (another source gives a range of 22.8–55.3%). Deep-sea shark species which have oils with very high unsaponifiable matter contents are *Squalus mitsukurina, Centrophorus squamosus, C. calceus, C. acus, C. atromarginatus, C. lusitanicus*, and *C. niaukang*. The cosmopolitan kitefin or seal shark (*Dalatias licha*) is also said to produce oil with a very high content of unsaponifiable matter.

A small amount of shark oil is traded in view of its use in industry. Special physical properties of such oils render them suitable for use as additives in certain lubricants and polishes.

See also: **Amines**; **Cholecalciferol**: Physiology; **Fatty Acids**: Dietary Importance; **Fish Oils**: Dietary Importance; **Mercury**: Properties and Determination; Toxicology; **Retinol**: Physiology; **Smoked Foods**: Principles; Applications of Smoking

Further Reading

FAO (1991a) *Fishery Statistics: Catches and Landings*. FAO Yearbook. FAO Fisheries Series 36, FAO Statistics Series 98. Rome: Food and Agriculture Organization.

FAO (1991b) *Fisheries Statistics: Commodities*. FAO Yearbook. FAO Fisheries Series 37, FAO Statistics Series 101. Rome: Food and Agriculture Organization.

FAO (1991c) *Draft Code of Practice for the Full Utilization of Sharks*. FAO Rome December 1991. FAO Fisheries Circular 844. Rome: Food and Agriculture Organization.

Kreuzer R and Ahmed R (1978) *Shark Utilization and Marketing*. Rome: Food and Agriculture Organization.

Lane IW and Comac I (1996) *Sharks Still Don't Get Cancer*. New York: Avery.

Smith M and Hiemstra C (eds) (1986) *Smith's Sea Fishes*. Grahamstown, South Africa: JLB Smith Institute of lchthyology.

Processing

M W Moody, Louisiana State University, Baton Rouge, LA, USA

Introduction

From earliest recorded history, humans have used fish and fishery products for sustenance, recreation, and many other positive aspects of developing society. Today, fish – and the associated fish-processing industry – remains an important economic segment of nearly every modern culture. Countries controlling water and coastal resources vehemently protect their interest. So sensitive are the issues of fisheries management and harvesting rights that disputes have been the source of international conflicts, embargoes, and even wars.

There are hundreds of species of fish. Nearly every large body of environmentally sound water, both fresh and salt, is currently producing or has the potential to produce commercially important fishery products. Although most of the fishery resource is harvested from wild stocks, aquaculture is becoming increasingly important as a stable and reliable source of fin fish. Worldwide landings of fin fish approach 122 million metric tons. In 1996, the top fish-producing countries were China, Peru, and Chile. China alone harvested more than 33 million metric tons of fish, more than 25% of world production. Because of the great diversity and distribution of fishery stocks and specific consumer demand for processed-fish products, the international trade of processed fishery products is significant. The estimated annual per-capita world consumption of fish and shellfish (estimated live-weight equivalent) for human food is an estimated 15.2 kg.

This article reviews the basic concepts of basic fin fish processing and associated quality considerations.

Fin fish can be processed into human food, animal food, or into components for industrial applications. Such inherent characteristics as oil content, flesh color, bones, size, and unit cost will usually determine the suitability for the final product.

Onboard Handling

Fin fish are harvested by various capture methods, including nets, trawlers, seines, traps, and hook and line. In many cases, regulations will dictate approved methods if conservation is an issue. The degree that fish are properly handled or mishandled at the time of capture will directly affect the quality of the final processed product. Fish should be quickly and thoroughly chilled and properly stored prior to processing.

Most fish are initially landed on to fishing vessels for temporary storage and transport to shore processing facilities. Depending upon time and distance, the fish may be iced, chilled, or frozen. In some instances, the harvesting vessel is designed and equipped to process fish on board the vessel.

Processing Plant Operations

In its simplest form, fin fish processing may involve only simple evisceration. It may also be complex, requiring specialized and expensive equipment, such as in the value-added production of fish analogs. Fin fish can be processed manually, with automated equipment or, more typically, by using a combination. Automated processing is more suited to processing fish that are uniform in dimension and where high volume is a consideration. Automated fish-processing equipment can be expensive and may require a high degree of maintenance. Manually processed fish, on the other hand, generally provide the highest meat yield, and this technique is used where product appearance is a major consideration.

Although there can be many varied steps and techniques used in fin fish processing, the common steps that may be used include grading or sorting, heading, evisceration, scaling, skinning, cutting, trimming, filleting, and one of many methods of preservation. A processing line is commonly used for fin fish operations. The raw material, usually whole, live, or drawn fish, is introduced into the plant and moves from one processing step to another until that day's delivery has been completed. Each processing step must be synchronized to avoid bottlenecks and delays in processing. Fish harvests can be seasonal or unpredictable, and processing plants must cope with daily shortages or gluts. The objective of fin fish-processing operations is to maximize meat yield and product

shelf-life, size and grade product accurately, and generate the minimum amount of waste. Meat yield is the bottom line and will depend upon the species being processed and the method of processing used.

Fish come in all sizes and shapes. Consequently, the final product can vary greatly in size, shape, and weight. There can be variability within the same species. Markets, however, require product sizing uniformity and consistency. In the last few years, there have been significant developments in sizing and grading technology. Computers, using advanced optic and weighing systems, can sort fish fillets and other meat cuts quickly and accurately. Fin fish-processing operations generate a significant amount of waste in the form of viscera, skin, bones, and unrecoverable meat. This is a growing problem since disposal options are becoming more restrictive. Fish processors are subject to governmental regulations, as is any food manufacturer. Fish nomenclature, fish-processing regulations, and international trade requirements are all challenges to modern fin fish processing. The following discussion considers some of the significant preservation methods and value-added products for processed fish.

Fresh Processing

Fish products processed and offered as fresh have a relatively short shelf-life. Depending on the quality of the fish and the conditions of storage, the shelf-life can be as short as a few days or as long as several weeks. Typical fresh products include whole drawn, headed, fillets, and other meat cuts. There is a direct correlation between the temperature of fresh fish products and the shelf-life. For maximum shelf-life, fish should be stored at temperatures near the freezing point of water (i.e., 0 °C). At this point the flesh will not freeze, but bacterial and enzymatic activities will be minimized. As storage temperatures rise above 0 °C, bacterial and enzymatic activities can increase significantly, dramatically shortening the shelf-life. Where applicable, crushed ice is used to hold fresh fish. The ice maintains the desired temperature of the fish. Water from the melting ice 'washes' microorganisms from the product, thus extending shelf-life. Little or no sophisticated packaging is required for the product. Most fin fish can be processed for fresh markets. (*See* **Storage Stability**: Parameters Affecting Storage Stability.)

Frozen Processing

To reach worldwide markets effectively, frozen fish processing is becoming increasingly important. Not all species of fish are suitable for freezing. Fish with a high fat content may become rancid in a relatively short time. In some cases, there may be undesirable textural changes or discolorations. There are a variety of frozen fish products, such as frozen fish blocks and breaded products, that are not available using any other method of preservation. Modern fish-freezing technology and equipment are economical, efficient, and specific for fish products. Bulk frozen products, such as frozen fish blocks (fillets, mince, or other meat cuts frozen as a single unit), are normally frozen using a direct-contact freezer such as a plate freezer. Individually quick frozen (IQF) products use the efficiency and speed of cryogenic freezers. Other types of freezers used in fish processing include blast and immersion freezing (brine freezing). Regardless of the type of freezing equipment used, there are common factors that have an impact on the product quality. (*See* **Freezing**: Blast and Plate Freezing; Cryogenic Freezing.)

Rate of freezing Rapid freezing produces smaller ice crystals than slow freezing. These smaller ice crystals are responsible for less structural damage to the fish flesh. In addition, rapid freezing minimizes the destructive concentration of cellular constituents. Rapid freezing is encouraged when processors use thin, prechilled packages of fish, and high-capacity, low-temperature freezers. (*See* **Freezing**: Principles.)

Freezer storage temperature There is no significant microbiological activity in frozen fishery products, but enzymatic activity can have a pronounced effect. Most frozen fishery products must be maintained at extremely low temperatures to negate enzyme destruction. Research shows that storage temperatures in the range −23° to −29°C are acceptable for most fishery products. Lower temperatures are even better but, at some point, economic considerations are prohibitive.

Packaging Frozen fish products are packaged to minimize dehydration and oxidation. Desiccation of poorly packaged frozen fish is a serious problem. Over a period of time, a highly undesirable drying condition (freezer burn) may occur. Fish fats and oils are easily oxidized in the presence of air. The resulting rancidity is responsible for undesirable flavors and color. Fish processors depend on commercial packaging materials and wraps to serve as effective barriers to minimize dehydration and oxidation. The fish-freezing industry has made great use of ice glazes for minimizing these same effects, especially in IQF products. Typically, ice glazes are formed by immersing or spraying frozen fish fillets or other fish products in chilled water immediately after emerging from the cryogenic freezing equipment. In this manner, a

coating of ice is formed around the product. The fish-processing industry also uses modified-atmospheric packaging (MAP) and vacuum-packaging to extend shelf-life and maintain quality. However, in the USA the use of such technology may be restricted without approved controls on fresh, unfrozen fish products because of potential growth of anaerobic pathogens such as *Clostridium botulinum*. (*See* Storage of Frozen Foods; **Freezing**: Structural and Flavor (Flavour) Changes.)

Fish Canning

Fish has been successfully canned commercially for more than 100 years. However, with the advent of improvements in freezing technology, transportation, marketing, and storage facilities, the percentage of canned fish products has been declining. One of the major advantages of canned fish is the relatively long shelf-life when stored at ambient temperatures. At the same time, canned fish is notably different from fresh and frozen fish. Canning requires subjecting fish to high temperatures to produce commercial sterility. The resulting product is fully cooked. Salmon, tuna, and herrings (including sardines and anchovies) are species of high commercial value that are traditionally canned. Fish may be delivered to the canning facility fresh (salmon and herrings) or frozen (tuna). Generally, fish are processed for canning by first heading, eviscerating, scaling, and cleaning. Additional processing steps can be quite variable, depending on the species and the desired end product. For example, salmon are packed raw into cans and, although the fins are removed, the skin and bones, which soften and are edible after processing, are considered part of the pack. Tuna and some of the herrings are cooked before packing the cans. Cooking removes excess water from the fish tissue and improves the appearance of the final pack. Bones, skin, and dark-meat portions are removed from tuna before packing into cans. (*See* **Canning**: Principles.)

Canned fish may or may not be packed with additional ingredients, such as oils, water, or sauces. Salmon are not generally packed with any additional media. Tuna may be packed in vegetable oil, water, or broth. Sardines and herrings may be packed in oils, flavors, or a number of sauces such as mustard or tomato. Cans used in the process are of many sizes and shapes. Metal cans are sealed with double-seaming machines that exhaust the head space to produce the proper vacuum. Hermetic sealing is essential if product integrity is to be maintained throughout heat processing, cooling, and storage. Leaks will cause spoilage and jeopardize product safety. Canned fish are low-acid foods and, consequently, will readily support the growth of most microorganisms. *Clostridium botulinum* is a heat-tolerant microorganism capable of growing and producing a highly potent toxin in canned fish that has been underprocessed, or contaminated as a result of container leaks. For this reason, most countries require minimum heat process times and temperatures to destroy all microbiological vegetative cells and spores to commercial sterility. Properly sealed and packed cans are heat-processed using steam under pressure in a retort. The processing time depends on can size, initial temperature of the fish at the time of processing, and the internal temperature of the retort. After processing, cans are quickly cooled in cold chlorinated water, labeled, and packed in shipping containers. Canned fish are subject to deteriorative changes during heat processing and subsequent storage. For example, meat discoloration may occur if the product is overprocessed. Struvite, a harmless glass-like crystal of magnesium compounds, may develop in canned tuna after an extended storage period; the use of appropriate food additives, such as chelating agents, usually minimizes the problem. (*See* **Canning**: Quality Changes During Canning; ***Clostridium***: Occurrence of *Clostridium botulinum*; Botulism.)

Dried, Salted, or Smoked Fish Processing

Curing and drying fish with salt and smoke were among the earliest methods used by humans to preserve fish. Over the years, these products have developed traditional markets and today represent a relatively small but important commercial segment of the processed fish industry. These products depend on low water activity to minimize the growth of microorganisms. Dried salted fish are processed by increasing tissue surface area, usually by splitting the fish, packing in salt to reduce tissue moisture quickly and, finally, air drying to achieve a moisture content low enough to provide product stability. Smoked fish can be processed using a variety of techniques and types of equipment. There are two basic types of smoked fish: cold-smoked and hot-smoked. Cold-smoked fish, as the name implies, is generally processed at lower temperatures than hot-smoked fish over a longer period of time. The salt content may be higher and the texture drier. Hot-smoked fish is succulent and is processed in a significantly shorter period using much higher temperatures. (*See* **Preservation of Food**; **Smoked Foods**: Principles; Applications of Smoking.)

There are several important steps in the processing of most smoked fish. Salting the raw fish before smoking firms the texture, adds a desirable flavor and, in some cases, may serve as a preservative. Soaking fish in a brine solution is the most desirable method of adding salt. This procedure provides

control and uniformity. Several factors directly affect the rate at which salt is absorbed, such as: (1) the strength of the brine solution; (2) the amount of exposed tissue (skin retards salts penetration); (3) the fat content of the fish (the higher the fat content of the fish, the slower the salt penetration); (4) the size of the fish or pieces of fish (the larger the pieces, the slower the penetration); and (5) the amount of agitation and temperature of the brine. After brining, but before smoking, the surface of the fish is air dried to facilitate 'pellicle' formation.

Drying is performed by a skilled individual and usually within the oven that will be used to smoke and cook the fish. A properly formed 'pellicle' will produce an attractive finished product. Smoke particles will cling to its surface, imparting a pleasing golden or bronze color. After drying, the fish is smoked and then cooked to the required internal temperature for a specified period. Cooking time and temperature will vary, depending on the final salt concentration and the use of other food additives. Modern smoking ovens have automatic controls that carefully monitor critical parameters. *C. botulinum* is a microorganism of great concern to smoked fish processors. In some countries, cooking times and temperatures and salt concentrations are established by regulations to eliminate the risk associated with this organism. In addition, the final product must be held under suitable refrigeration to prevent the growth of microorganisms.

Further Processing

Full use of the fish resource is a difficult challenge for the processing industry. There are many species that, because of economics and demand, are not suitable for traditional processing. In addition, fish frames and other filleting wastes resulting from processing operations represent a significant source of recoverable meat. Mechanical flesh-separating equipment can be used to recover meat economically. Most systems work by forcing edible meat through a perforated surface. In this way, the fish flesh is separated from the skin, bones, and other unwanted fish waste materials. The resulting comminuted material is paste-like and is called fish mince. Ideally, the mince should be as white in color as possible, but mince can be dark if it contains blood, organs, or pigments. Mince has many potential uses. Its primary use may be as an extender in other fish and seafood products. Fish mince may also serve as a base material for the manufacture of analogs and other fabricated seafood and meat products. Surimi is made from mince that has been washed in water to remove many of the solubles, including color, to render a highly functional material. Surimi can be processed and formed into various imitation seafood products, such as shrimp, crabmeat, and lobster meat. The production of these imitation seafood products from surimi is highly technical and requires expertise and sophisticated equipment. The process begins with the addition of certain functional ingredients that will give desired properties to the final product, including water, starches, protein, fats, and other ingredients. In addition, various natural and artificial colors and flavors may be blended with the surimi to mimic the desired seafood product. Ingredients including color and flavors used must be in compliance with food regulations and laws. The resulting product is extruded into the desired shape, cooked, and packaged.

Quality Considerations

Fish and fish products are among the most perishable of all foods. Decomposition and spoilage can begin shortly after harvesting, and can be minimized only with extreme care and diligence on the part of the fisherman and processor. The two primary causes of spoilage in fish are microbiological and enzymatic. Bacteria are the primary microorganisms that affect fish quality but molds and yeasts can also be responsible for spoilage of fishery products. Large numbers of microorganisms are found naturally in the gut and body slime of living fish. When fish die as a result of harvesting or when they are butchered during processing, bacteria are introduced into muscle or organ tissues and cause spoilage and decomposition. Under favorable conditions, bacteria can grow and reproduce extremely rapidly. Unchecked, decomposition is quick. Proper temperature controls and sanitation procedures can provide the maximum shelf-life for most seafood products. Fish may harbor a large variety of bacteria, even some pathogens. However, since fish are generally cooked prior to consumption, most do not pose an immediate health hazard. Processing techniques used to process fish require individual evaluations to determine control methods. Please refer to the previous example of *C. botulinum* and the use of vacuum or MAP packaging.

Processors are concerned not only about the number of microorganisms and controlling their numbers, but also about contamination by specific types of microorganisms. Pathogens, even in small numbers, are capable of causing disease or producing toxins. Naturally occurring proteolytic enzymes cause spoilage, called autolysis, by producing undesirable textural and flavor changes. Like microbial spoilage, the undesirable effects of enzymes are minimized by proper handling and temperature control. However, unlike microbial spoilage, enzyme activity may occur during frozen storage.

Several natural toxins may occur in processed fin fish products. Ciguatera and scombroid (histamine) are, perhaps, the most well known. Ciguatera is most commonly associated with reef fish in tropical zones. Fish flesh is made toxic as a result of fish consuming the benthic dinoflagellate *Gambierdiscus toxicus*. Scombroid poisoning is caused by the ingestion of fish containing a high level of histamine. Histamine is produced in fish flesh by enteric bacteria after harvesting. In most cases, this condition can be traced to improper cold storage of the fish. Scombridae (tuna and mackerel) and Scomberesociadae fish are more commonly associated with scombroid poisoning, although other families have been implicated. Other naturally occurring fish quality considerations include paralytic shellfish poisoning and parasites.

Fish Products

Products resulting from fin fish processing are diverse and complex and yet most of the world production can be summarized within several categories (Table 1). One reason for the complexity is product and species nomenclature. The common name of a fish or fish product in one region may not be the same as in another. In addition, there is the large number of fin fish species. A consistent and meaningful nomenclature system of fish and fish products is necessary for international trade. Although there have been attempts by various organizations and governmental agencies to standardize fish nomenclature, this issue remains a major problem not only in international trade but also within commercial domestic trade. Quality and sanitation are other major considerations for fish products. Microorganism count and species under certain conditions may give an indication of the overall quality of the finished product and under what conditions it was processed. These microbiological requirements and standards, however, may vary considerably among countries and nationalities. In general, most countries involved in international trade will consider aerobic plate count, enteric microorganisms, and pathogens as microbiological indicators of food quality and sanitation. Mandatory government fishery inspection programs are becoming increasingly important. In the USA, the Hazard Analysis Critical Control Point (HACCP) concept is being considered as the basis for a national fishery inspection program. HACCP requires processors to evaluate all processing steps and to identify those steps where hazards could occur if the step is not controlled. Control systems, monitors and records are required to insure that the processing of the fish is in compliance with requirements. (*See* **Hazard Analysis Critical Control Point**.)

Table 1 World fish products, 1996

Product	*Metric tons*
Fish, frozen whole and fillets	16 720 040
Fish, oil	1 395 204
Fish, cured (salted, dried, smoked, spiced, marinated, vinegar- or sugar-cured)	4 490 446
Fish, canned and prepared (packed in containers: fermented, sausages, balls, salads, paste, roes livers)	5 447 264

Source: Food and Agriculture Organization of the United Nations (1998) *FAO Yearbook. Fisheries Statistics, Commodities*. Rome: FAO.

See also: **Canning**: Principles; Quality Changes During Canning; ***Clostridium***: Occurrence of *Clostridium botulinum*; Botulism; **Freezing**: Principles; Blast and Plate Freezing; Cryogenic Freezing; Storage of Frozen Foods; Structural and Flavor (Flavour) Changes; **Hazard Analysis Critical Control Point**; **Preservation of Food**; **Smoked Foods**: Principles; Applications of Smoking; **Storage Stability**: Parameters Affecting Storage Stability

Further Reading

Burgess GHO, Cutting CL, Lovern JA and Waterman JJ (1967) *Fish Handling and Processing*. New York: Chemical Publishing.

Food and Agriculture Organization of the United Nations (1998) *FAO Yearbook. Fishery Statistics Commodities*. Rome: FAO.

Martin RE and Flick GJ (eds) (1990) *The Seafood Industry*. New York: Van Nostrand Reinhold.

US Government (1991) *Fisheries of the United States, 1990*. Washington, DC: US Government Printing Office.

Ward DR and Hackney CR (eds) (1991) *Microbiology of Marine Food Products*. New York: Van Nostrand Reinhold.

Wheaton FW and Lawson TB (1985) *Processing Aquatic Food Products*. New York: Wiley.

Miscellaneous Fish Products

T Motohiro, Formerly of Kagoshima University, Kagoshima, Japan

Background

Fish and shellfish have long been consumed by some maritime tribes for their protein source. However, a finding that the highly unsaturated fatty acids in

marine oils have been related to prevention of heart disease is noteworthy in modern Western societies conscious of their health, and the consumption of seafood is increasing significantly. (*See* **Fish Oils**: Dietary Importance; **Shellfish**: Commercially Important Crustacea; Commercially Important Molluscs.)

The majority of fish and shellfish are safe as food, but there are some species that are naturally toxic or poisonous and may cause illness or even death if they are consumed. This article reviews raw-fish dishes of Japan (e.g., *Fugu*), nature and risks from ciguatoxins, caviar and its production, and eels and their uses.

Raw-fish Dishes of Japan

Raw-fish dishes – sliced fish flesh called *sashimi* in Japan – have been served with vinegar for more than 1000 years, but are now more popular eaten with soy sauce and grated Japanese horse radish. Flesh from freshly caught fish is in the majority, but frozen–thawed flesh is often supplied to the dishes. Bluefin tuna (*Thunnus thynnus*), sea bream (*Pagrosomus major*), striped bonito (*Sarda orientalis*), yellowtail (*Seriola quinqueradiata*), and flounder (*Paralichthys olivaceus*) are the species mostly consumable in all seasons, and tiger puffer (*Fugu rubripes*) is considered a prize delicacy in winter. The flesh for the dishes should be of top quality regarding color, flavor, taste, and texture.

Freshly caught bluefin tuna and striped bonito are normally subjected to deep freezing, at a temperature as low as −45 °C, at sea to prevent myoglobin, the dominant pigment in fish muscle, from turning to metmyoglobin, which is a cause of meat discoloration. Thus, the thawed flesh of deep-frozen tuna and bonito is suitable to be served as raw-fish dishes.

More than 100 species of puffer fishes (family Tetrodontidae) are broadly distributed from temperate to tropical waters, but only about 10 species have been consumed as human food by a handful of countries, principally Japan. There are marked differences in toxicity among species, some being virtually nontoxic. Fresh flesh and some parts of the puffer fish may be enjoyed as raw or cooked dishes if properly prepared, but other parts, such as ovary and liver, are extremely toxic (**Table 1**). Only licensees may prepare and serve dishes of the puffer fish for consumers, under conditions specified by the Ministry of Health and Welfare, because the major cause of food-poisoning deaths is ingestion of toxic puffer fishes.

The responsible toxins are tetrodotoxin (TTX) and related compounds (**Figure 1**). The lethal dose of TTX for people is estimated to be 10 000 mouse units, equivalent to 2 mg of TTX. Poisoning as a result of eating puffer fishes is serious, and causes neurological symptoms: tingling in the lips and extremities, paralysis, and death by respiratory arrest and cardiovascular collapse. The cardiovascular effects account for the relatively high lethal rate in puffer-fish intoxications.

Table 1 Toxicity of puffer fishes

Species	*Ovary*	*Testes*	*Liver*	*Skin*	*Intestine*	*Muscle*
Fugu niphobles	A	C	A	B	A	C
F. poecilonotum	A	B	A	B	B	C
F. vermiculare vermiculare	A	D	A	B	B	C
F. pardale	A	C	A	B	B	D
F. vermiculare porphyreum	A	D	A	B	B	D
F. ocellatus obscurum	A	D	B	B	B	D
F. chrysops	B	D	B	B	C	D
F. rubripes rubripes	B	D	B	D	C	D
F. xanthopterum	B	D	B	D	C	D
F. stictonotum	B	D	B	C	D	D
Lagocephalus laevigatus inermis	D	D	B	D	D	D
L. lunaris spadiceus	D	D	D	D	D	D
Liosaccus cutaneus	D	D	D	D	D	D
Canthigaster rivulata	D	N	C	B	C	D
Diodon holacanthus	D	N	D	D	D	D
Chilomycterus affinis	D	N	D	D	D	D
Ostracion immaculatum	D	D	D	D	D	D
Lactoria diaphana	D	D	D	D	D	D
Aracana aculeata	D	D	D	D	D	D

A, strongly toxic, lethal at less than 10 g; B, moderately toxic, not lethal at less than 10 g; C, weakly toxic, not lethal at less than 100 g; D, negative, not lethal at less than 1000 g; N, no data available.
Data from Hashimoto Y (1987) *Marine Toxins and other Bioactive Marine Metabolites*. Tokyo: Japan Scientific Societies Press.

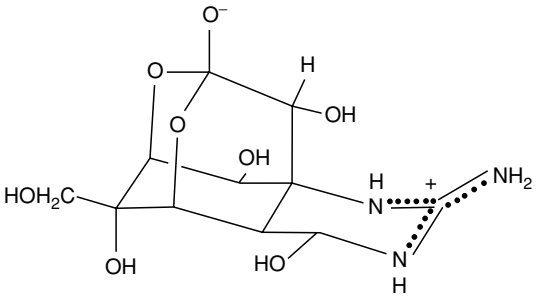

Tetrodotoxin

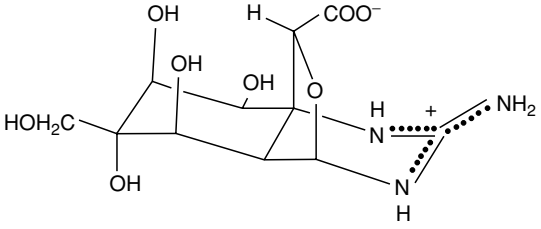

Tetrodonic acid

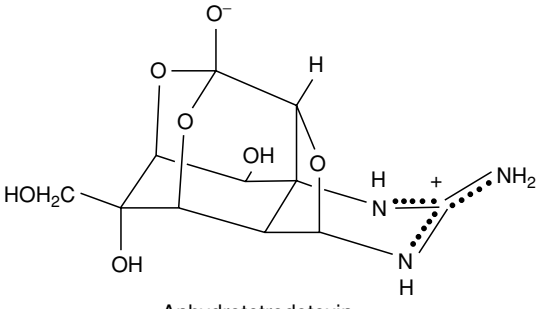

Anhydrotetrodotoxin

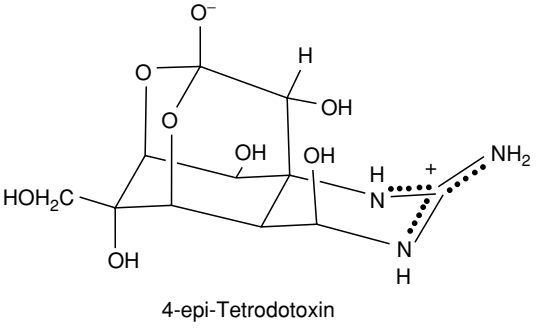

4-epi-Tetrodotoxin

Figure 1 Structures of tetrodotoxin and related compounds. From Nakamura M and Yasumoto T (1985) Tetrodotoxin derivatives in puffer fish. *Toxicology* 23: 271–276, with permission.

TTX has been demonstrated to be produced by intestinal bacteria of TTX-bearing animals, as well as some species of marine bacteria. This fact, along with the distribution of TTX in nature, made it possible to suggest the mechanism involved in the toxification of TTX-bearers (Figure 2).

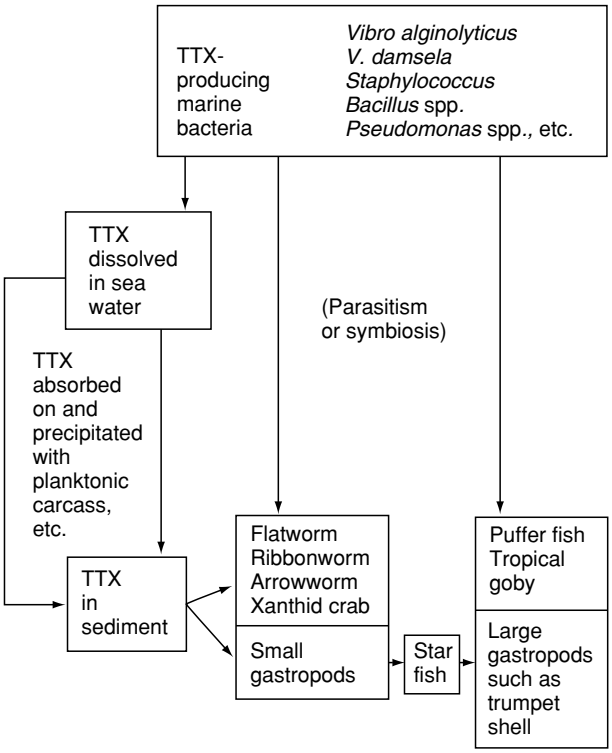

Figure 2 Assumed mechanism of toxification of TTX-bearing animals. From Noguchi T, Arakawa O and Hashimoto K (1989) Tetrodotoxin with special reference to its origin and the mechanisms involved in toxification of puffers. *Journal of the Food Hygiene Society (Japan)* 30: 281–288, with permission.

Nontoxic or slightly toxic puffer fishes have been successfully cultured, and an increase in demand is likely.

Nature and Risks from Ciguatoxins

About 300 species of fish and shellfish, inhabiting shallow waters around coral reefs, have been known to cause ciguatera, a poisoning with unique symptoms. Fish apparently become toxic and lose the toxicity rather more slowly. Although the name ciguatera is derived from the Spanish term for a marine snail, the syndrome is currently recognized as the result of consumption of certain tropical fish. The risk of ciguatera is highest from the consumption of herbivorous reef fish or of the carnivorous fish that feed on them.

It is thought that the principal source of ciguatoxin, the most important toxin, is the benthic dinoflagellate *Gambierdiscus toxicus*, which tends to cluster on the surfaces of certain tropical seaweeds. Because of this growth habit, the benthic dinoflagellate is efficiently accumulated by herbivorous reef fish. Ciguatoxin was isolated from *G. toxicus* obtained directly from

coral reef seaweeds. Cultures of *G. toxicus*, however, have been found to produce primarily a different toxin, maitotoxin, the structure of which is not yet clarified. Other benthic dinoflagellates, with growth habits and distributions similar to those of *G. toxicus*, may also be involved. The known differences between ciguatoxin and brevetoxin lie in the longer duration of the effects of ciguatoxin in human victims. Symptoms are nausea and neurological deficits similar to those of paralytic shellfish poisoning, but more severe and enduring. Victims of ciguatera occasionally complain of relapses, in some cases precipitated by consumption of alcohol, for months or even years after the initial attack. Deaths are rare. The neurological dysfunction includes a reversal of the sensations of hot and cold, called 'dry-ice' sensation. The distribution of ciguateric fish cannot be defined strictly, but distinct zones of toxicity are known by local fishermen. In notorious carnivorous fish, such as barracuda, which are capable of both migrating and accumulating high levels of toxin, it would be difficult to discriminate toxic fish from others.

Although the characterization of ciguatoxin has been a long-standing matter owing to the complexity of the molecule and the difficulty in obtaining adequate material, instrumental analyses with purified material have elucidated the complete structure (Figure 3). Chromatographic evidence indicates that several other related toxins are also involved in ciguatera poisoning.

Caviar and its Production

The eggs of many fish species, and of some other aquatic animals such as sea urchin, are sought as a food item. Caviar is one of the most well-known names, especially in connection with sturgeon (*Acipenser medirostris*) eggs. Caviar and other fish egg products are considered a specialty food and thus tend to bring a high price.

Fish eggs have a substantial food value. On a dry-weight basis, eggs contain 45–48% protein, 11–32% lipid, and 1.3–4.5% ash, respectively. To produce caviar, live mature fish have to be used; from dead sturgeon, it is difficult to remove the eggs as they adhere firmly to the ovary sac. The belly is cut open, from the pectoral fins to the anus. The ovary is then drawn out and processed immediately. The general procedure used in caviar processing begins with washing the ovaries to remove most of the blood and slime. The ovaries are then cut open to extract the eggs. The egg-membrane tissue of the sturgeon is normally hyperplasic and muscular, but it can be softened enzymatically by the action of lysozyme. The eggs are typically salted, using either dry salting or a salt brining. For dry salting, 10% (w/w) of salt is added to the eggs as they are passed through a 33-mm sieve; stirring may ensure faster and more effective penetration of the salt into the eggs. After 3 h of salting, the eggs are drained on a wire net. In the case of salt brining, the eggs are allowed to stand for 1 h in saturated brine, then drained upon a wire net overnight. For further dehydration, the salted eggs are pressed in a cotton bag. The salt-cured eggs are packed in a container, tightly sealed, and stored at 5 °C for aging.

Salted salmon eggs are called red caviar. The quality of red caviar depends on species, maturity, and freshness of the eggs. The eggs should be of spawning

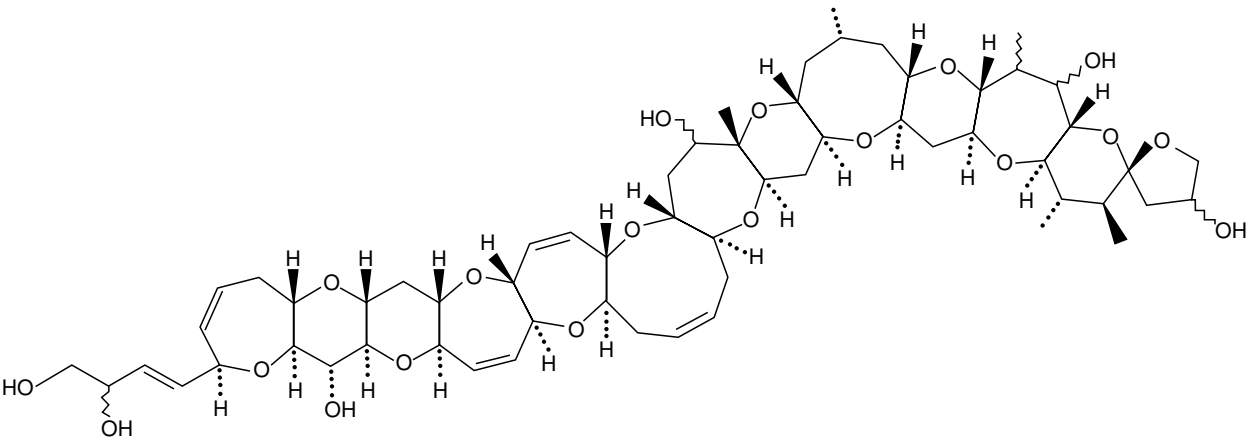

Figure 3 Structure of ciguatoxin. From Murata M, Legrand AM, Ishibashi Y and Yasumoto T (1989) Structure of ciguatoxin and its congener. *Journal of the American Chemistry Society* 111: 8929–8931, with permission.

salmon caught from coastal waters and be extracted within 2 h of catching. Egg-membrane tissue in a mature ovary is naturally hyperplasic and muscular, and results in a low-quality product. It is also inadequate if eggs are used from the salmon late in the fishing season, or immediately before spawning. A high-quality red caviar is attributed to the use of eggs of pink salmon (*Oncorhynchus masou*).

Once separated from the raw salmon, the ovaries are washed thoroughly in fresh water to remove blood vessels and slime, then rubbed through a gauze cotton net to isolate the eggs. An apparatus consisting of gauze cotton net, a sloping wire net, and a receiver is useful for commercial purposes. The eggs that have been separated through the gauze cotton net run down the slope and fall into the receiver. The mass accumulated in the receiver is put into a tub containing saturated salt brine (which has been boiled and cooled prior to use), stirred for 20 min, removed and put into a basket, drained for 24 h, packed in a barrel or a wooden box, and cold-stored for aging.

The oil content of salmon eggs averages 12% on the basis of total weight. Salmon eggs yield a light-colored oil. The oil contains about 45% of highly unsaturated fatty acids with molecules having 20–22 carbon atoms. About one-third of the total fat is phospholipid, probably lecithin. The protein of salmon eggs is of a high quality. It contains all of the eight essential amino acids (isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine) in substantial amounts. Salmon egg protein is relatively high in lysine, methionine, and isoleucine. (*See* **Amino Acids**: Metabolism.)

Eels and their Uses

Eels are difficult to sell in the USA but are highly prized in Europe and several areas of the Far East, including Japan. The major species of food eels are freshwater eels (Anguillidae), pike eels (Muraensocidae) and conger eels (Congridae). Of the 16 species of *Anguilla* eels found worldwide, most are in the waters around Southeast Asia. *Anguilla japonica* and *A. marmorata* are the inhabitants in the waters around Japan, while *A. rostrata* and *A. anguilla* are in the Atlantic. For food eels, the important species are *A. japonica* and *A. anguilla*. *Muraenesox cinereus* is the commercially valuable species of pike eel. Conger eels number about 110 species in the world, but white-spotted conger (*Conger myriaster*) is the only species prized as food. Eels are mostly processed to produce 'spitchcock' with seasoning juice. Eels weighing roughly 200 g are put into a basket and subjected to cut-feed for a couple of days in running water. In producing 'spitchcock,' back-cutting, meat separating, and sizing are the initial processes, performed manually. The spiting and broiling processes are carried out mostly by machines.

In the broiling process, an infrared ray burner is commonly used. The broiling time depends on the flame temperature and the size of the slices of flesh, but generally, 4–5 min is long enough for the first broiling without seasoning juice, then 5–10 min for the second broiling with the juice. A seasoning juice is brushed over the flesh surface during an automated run from the first to the second broiling processes. The broiling process greatly affects the food quality: either insufficient or excessive broiling gives a bitter or burnt flavor to the product. The seasoning juice consists of soy sauce, sugar, and sweet liquor, in a ratio of 1:1:0.5.

Smoked eel is another type, traditionally popular in Europe, and is widely consumed. Preliminaries for processing smoked eels are almost similar to those for 'spitchcock.' The fresh fillets are cured in 5–10% salt brine for about 10 h; salt concentration and curing time are controlled according to the size of the fillet. The cured fillets are then subjected to trimming, washing, draining, and drying. Smoke drying is important in processing smoked eel and is performed at a controlled temperature within 60–80 °C for 5–6 h, using hardwood or sawdust. A gradual increase in temperature, using adequate amounts of sawdust, is required to prevent the flesh from scorching. The half-finished product is taken out from the smoking kiln, any soot is swept off, and the fillets are arranged in a carton box to equalize the water contents.

The 'spitchcock' eels are the traditional food of Japanese people during the hottest period of the summer. They contain (per 100 g of eel) 47.1 g of water, 23.0 g of protein, 24.4 g of lipid, 3.1 g of non-fibrous carbohydrate, 2.4 g of ash, 150 mg of calcium, 300 mg of phosphorus, 0.8 mg of iron, 510 mg of sodium, 300 mg of potassium, 150 μg of retinol, 5000 IU of retinol potency, 0.75 mg of thiamin, 0.74 mg of riboflavin, and 4.1 mg of niacin. Refer to individual nutrients.

Pike eel is a white-flesh fish, which is rich in fat and taste, and consumable as 'spitchcock,' soup, or 'vinegared' products, or, after processing, to *surimi* (ground meat), as the base for a high-quality fish-gel product. However, a number of tiny bones remaining in the flesh are disadvantageous in processing any type of marine food.

See also: **Amino Acids**: Metabolism; **Fish Oils**: Dietary Importance; **Shellfish**: Commercially Important Crustacea; Commercially Important Molluscs

Further Reading

Hall S (1991) Natural toxins. In: Word DR and Hackney CR (eds) *Microbiology of Marine Food Products*, pp. 301–330. New York: Van Nostrand Reinhold.

Motohiro T (1975) An elementary knowledge of frozen food. In: Amano K, Kaneko S, Kameda K *et al*. (eds) *Handbook of Frozen Food*, pp. 39–42. Tokyo: Asakura-Shoten.

Tanikawa E, Motohiro T and Akiba M (1985) *Marine Products in Japan*, 2nd edn. Tokyo: Koseisha Koseikaku.

Spoilage of Seafood

P Dalgaard, Technical University of Denmark, Lyngby, Denmark

Background

The world's annual production in capture fisheries has remained at ~ 90 million tonnes during the last decade, and only aquaculture production in marine and fresh water has allowed for global increases in raw material for fishery products. At the same time, international trade in seafood is increasing, and optimal utilization of available resources and prevention of losses during storage and distribution are more important than ever. Most seafoods have a short and variable shelf-life, and the increasing interest in marketing of less preserved products makes prevention of spoilage an important challenge.

The flesh of most fresh fishes has a high pH of 6–7 *post rigor* and a high content of free amino acids in the muscle extractives. Many marine and some freshwater species contain trimethylamine oxide (TMAO), resulting in a high muscle redox potential, and some species contain high levels of polyunsaturated fatty acids. Furthermore, fish and shellfish often harbor a natural psychrotolerant microflora. These specific properties result in microbial activity and lipid oxidation being responsible for spoilage of most fresh and lightly preserved products. This explains the short shelf-life of these products (**Table 1**). Postharvest losses of food have been estimated to be as high as ~ 25% resulting from microbial activity alone, and this problem is particularly relevant for seafood. To reduce losses resulting from spoilage and to insure that products are safe at the time of consumption, knowledge about product shelf-life under different storage conditions is important. Storage temperature, packaging, hygienic standard of raw material, and gutting are the most important factors influencing the shelf-life of seafoods (**Table 1**).

When new seafoods are developed, or existing products are modified or stored under new conditions, shelf-life must be determined by sensory methods. However, product attributes, to be measured by instrumental methods, are useful when industry and national or international authorities need to evaluate raw material or products. Development of such instrumental methods and techniques to predict shelf-life can rely on information about reactions responsible for product spoilage. The present article concerns these microbial, chemical, and biochemical spoilage reactions and methods to determine and predict shelf-life of seafood. In addition, safety hazards related to seafood consumption are presented.

Sensory Changes and Shelf-life

Sensory properties of seafood, as determined by the senses of sight, smell, taste, and touch change during storage, and a series of phases can be identified in most products. Fresh fishes typically have a species-specific taste and odor that disappear after a few days of chilled storage. Later, off-odors and off-tastes, i.e., off-flavors, which are often ammonia-like, fishy, sour, sulfurous, or yeasty become notable, increasing in intensity and finally resulting in overt spoilage. Sensory methods relying on panels of trained assessors are used to determine the shelf-lives of products based on the degree of sensory changes.

Sensory methods frequently used with seafoods include the Torry scheme, various simplifications of this method, EU schemes for freshness grading, and the quality index method (QIM). The Torry scheme, developed in Scotland in 1948 for fresh fishes, relies on the assessment of raw appearance, raw odor, raw texture, cooked odor, cooked texture, and cooked flavor of samples. Each quality attribute is scaled by five or 10 descriptive terms. Numerous less extensive sensory schemes have been used to determine the shelf-life from the appearance, odor, taste, and texture of raw or cooked products. Often, sensory changes are expressed as a simple overall grading (I, II, or III) and shelf-life determined as the time when the percentage of samples in grade III reaches a defined level. The EU requires numerous fish and shell-fish to be graded by specific EU schemes if marketed within the community. Assessments of appearance, odor of skin, outer slime, eyes, gills, and peritoneum are carried out with whole or gutted fish. Assessments are typically carried out in auctions by trained personnel, and fish are placed into grades Extra (E), A, B or Unfit (C), based on schemes for

Table 1 Shelf-life of selected seafoods

Product	*Storage temperature (°C)*	*Packaging*	*Shelf-life (days)*
Fresh seafood			
White-fleshed from temperate/cold waters	0	Air or VP	12–18
Cod	−2.5	MAP	36
Caught in warm waters	0	Air	18–35
Tilapia	4	MAP	> 25
Large fishes like halibut, tuna, and salmon	0	Air or VP	19–22
Dark-fleshed, small fish			
Low-fat	0	Air	6–9
High-fat	0	Air	4–6
Shellfish, temperate and cold waters	0	Air	3–10
Shellfish, warm waters	0	Air	8–12
Sharks and rays	0	Air	8–10
Frozen and thawed	0	Air	12–18
Thawed cod	2	Air or VP	9–12
Thawed cod	2	MAP	12–>20
Sous vide cooked	3–4	VP	15–>21
Lightly preserved			
Cold-smoked salmon	5	Air	~14
Cold-smoked salmon	5	VP	20–56
Hot smoked cod (lightly salted)	0–1	Air	24–26
Brined roe (lump fish)	5	VP	75–90
Cooked and brined shrimps	5	Air	40–>60
Cooked and brined shrimps	5	MAP	84–>200
Semipreserved and preserved			
Cod (salted)	0	Air	500
Sugar salted herring	5	Brine	> 360
Frozen			
Whitefish, whole or fillet blocks	−18	Glazed	120–240
Fatty fish	−18	Glazed	90–120
Mackerel	−26	VP	> 540

MAP, modified-atmosphere-packaging; VP, vacuum-packaging.
Modified from Dalgaard P (2000) Fresh and lightly preserved seafood. In: Man CMD and Jones AA (eds) *Shelf Life Evaluation of Foods*, 2nd edn. pp. 110–139. Gaithersburg, MD: Aspen.

different groups of species. These freshness categories do not allow the remaining product shelf-life to be determined directly, and the demerit point or QIM can be used to obtain this information. The QIM relies on sequential evaluation of a relatively large number of fish attributes. Demerit or QIM points of a sample, i.e., the sum of the attribute scores, increase linearly with product storage time, and rates of spoilage at a constant storage condition can be calculated easily. QIM schemes allow the remaining shelf-life of whole or gutted fish of several species to be determined. Evaluation of fish fillets remains difficult, and further studies are needed to validate QIM for determination and prediction of the shelf-life of fresh fish above 0–5 °C, i.e., when products are temperature-abused.

Microbiological, Chemical, and Biochemical Changes

Sensory methods are always needed to determine the shelf-life of new or modified seafoods. Unfortunately, sensory methods relying on a panel of assessors are expensive to use and difficult to calibrate. Therefore, it is desirable to supplement or replace sensory evaluations by instrumental analyses. Instrumental methods are important, e.g., to determine whether the characteristics of seafoods comply with specifications as part of purchase agreements or with standards and guidelines established by national or international authorities.

Microbial activity is responsible for spoilage of most lean fresh and lightly preserved seafood. Chemical reactions, primarily oxidation and activity of autolytic enzymes, become important spoilage reactions when microbial activity is reduced by processing or preservation (Table 2). It is noteworthy that the range of product characteristics and storage conditions within which a particular spoilage reaction actually limits product shelf-life remain poorly described for practically all seafoods.

Microbiological Changes

Fish and shellfish are poikilotherms, i.e., their body temperature is similar to that of their surroundings,

Table 2 Seafood spoilage indicators

Spoilage indicator	*Sensory property*	*Spoilage reaction*	*Examples of products where compounds are formed*
Trimethylamine (TMA)	Ammonia-like	Microbial	Most marine fishes and in some fresh water fishes like Nile perch and tilapia that contain TMAO
Ammonia	Ammonia	Microbial and enzymatic	Particularly in cephalopods, crustaceans, and dark-fleshed fishes
Histamine	Pepper-like[a]	Microbial	Tuna, mackerel, mahi-mahi, herring, anchovies and other seafoods with a high level of free histidine, particularly when stored $>5\,^{\circ}C$
Cadaverine, putrescine, and tyramine	Putrid, rotten[a]	Microbial	Most fresh and lightly preserved seafoods
Volatile sulfur compounds	Sulphurous, rotten	Microbial	Aerobically stored fresh fish, temperature-abused fresh and lightly preserved products, but also sporadically in sous vide cooked and chilled fish where very high levels can be observed
Short-chain alcohols and carbonyls (ethanol, propanol, butanol, 2-methylbutanal, 3-methylbutanal, 2,3-butadione)	Solvent-like, malty	Microbial	Most fresh and lightly preserved seafood
Hypoxanthine (Hx)	Bitter taste	Enzymatic and microbial	All fishes, but rates of Hx formation vary markedly between species owing to differences in activity of the autolytic enzymes that generate the substrate precursor inosine. As examples, Hx formation is very rapid in redfish, modest in cod, slow in mackerel, and very slow in swordfish
Dimethylamine (DMA)	Ammonia-like	Enzymatic	TMAO containing seafood and particularly in hake when stored frozen at $>-20\,^{\circ}C$
Free fatty acids (C_{16}–C_{22})	Rancid, oily, bitter and metallic	Enzymatic	Fatty fishes stored at conditions of temperature abuse as well as in both lean and fatty fishes during frozen storage at inappropriately high frozen storage temperatures
Propanal	Solvent-like	Chemical	From oxidation of *n*-3 fatty acids
Pentane, hexanal		Chemical	From oxidation of linoleic acid and useful at indicator for this type of oxidation
Cis-4-Heptanal	Cardboardy, paint-like	Chemical	During frozen storage of cod
2,4-Heptadienal, 2,4,7-Decatrienal	Stale, whale oil flavor	Chemical	Fatty fishes and fish oil containing products

[a]As a result of high flavor detection thresholds, biogenic amines contribute to the spoilage of seafood only when present in extremely high concentrations.

and the often complex microflora on their outer surfaces to some extent reflect the microbiology of the water in which they live. However, only some of the numerous species of microorganisms on live fishes are important in seafood spoilage. In fact, a pattern of microbial growth and activity during storage of seafood is frequently observed; the so-called specific spoilage organism (SSO) concept (Figure 1). On newly processed, fresh, or lightly preserved seafood, the SSOs are usually present in very low concentrations and constitute only a minor part of the total microflora. During storage under particular conditions of temperature, atmospheres, NaCl, available water, preservatives, etc., SSOs, which are often a single microbiological species, grow faster than the remaining microflora, produce metabolites responsible for off-flavors, and finally cause sensory rejection of a product (Figure 1). SSO are typically present at levels of about 10^7 colony-forming units (cfu) g^{-1} when products become sensorily spoiled. In contrast to SSOs, the spoilage microflora, i.e., the microflora found at the time of sensory product rejection, are frequently a mixture of microbial species, some of which can be unimportant in terms of spoilage. The substantial importance of microbial activity in seafood spoilage has been established by comparing off-flavor development in sterile and naturally contaminated blocks of fish muscle. Identification of a SSO relies on a comparison of product spoilage characteristics with those of isolates from the spoilage microflora. Spoilage potential and spoilage activity are evaluated as the qualitative and quantitative ability, respectively, of isolates to produce off-odors and spoilage metabolites. The yield factor for production of a metabolite by a microorganism is a measure of the amount of metabolite produced per cell. Yield factors are most useful to determine whether the levels of cells and metabolites in a seafood correspond to the spoilage activity of an isolate from the spoilage microflora, i.e., if the isolate is likely to be the SSO. Evaluation of proteolytic activity need not be included in SSO identification, because microorganisms metabolize seafood extracts, primarily free amino acids and TMAO, and causes sensory product spoilage before protein degradation is detected.

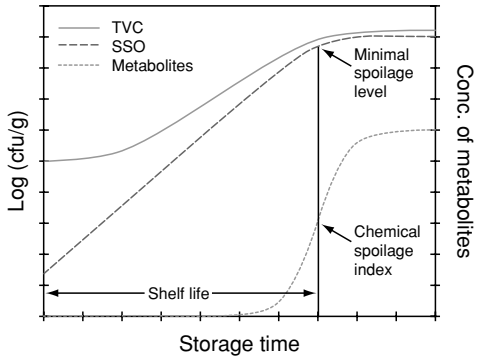

Figure 1 Specific spoilage organism (SSO) concept. Typical changes in total viable counts (TVC), SSO, and metabolites produced by SSO during storage of fresh seafood. Modified from Dalgaard P (2000) Fresh and lightly preserved seafood. In: Man CMD and Jones AA (eds) *Shelf Life Evaluation of Foods*, 2nd edn, pp. 110–139. Gaithersburg, MD: Aspen with permission.

As shown in Figure 1, the level of SSO and their metabolites can be used as indices of product spoilage. Furthermore, product shelf-life can be predicted from the initial level and from the growth rate of an SSO. Finally, it follows from the SSO concept that targeted inhibition of a SSO can be used in the development of new products that have an extended shelf-life.

The SSOs of a number of seafoods have been identified and habitats of the live fishes, product characteristics, and storage conditions clearly determined the microorganisms responsible for product spoilage (Table 3). The specific parameters that allow certain microbial species to become dominant in the spoilage microflora can be determined by studying SSOs in pure cultures. As an example, *Shewaenella putrefaciens*-like bacteria grow vigorously at the pH of most fresh seafoods but are substantially inhibited at pH values below 6. This explains why *Pseudomonas* spp. are more important SSOs in aerobically stored fresh seafoods with a pH of ~ 6 or below. *S. putrefaciens*-like organisms, *P. phosphoreum*, *Aeromonas* spp., species of Enterobacteriaceae and *Vibrio* spp. reduce TMAO to trimethylamine (TMA) in anaerobic respiration. This facilitates their growth in vacuum or modified-atmosphere-packed seafoods and explains the development of TMA in different fresh and lightly preserved packed products. Spoilage of these products, however, can also be caused by lactic acid bacteria, although these are unable to reduce TMAO to TMA. Gram-positive microorganisms are generally more resistant to heat, frozen storage, salt, low water activity, low pH, high carbon dioxide levels, and irradiation than Gram-negative microorganisms. However, existing data for the quantitative effect of environmental parameters on growth kinetics of microorganisms are insufficient to predict precisely the microbial species that become the SSO in a seafood, where combinations of product characteristics or storage conditions have not previously been evaluated in storage trials with naturally contaminated products.

Shewanella putrefaciens-like organisms are primarily important in spoilage of aerobically stored fresh fish (Table 3) but they also may contribute to spoilage

Table 3 Specific spoilage organisms in seafood

Seafoods	*Typical specific spoilage organism*
Fresh and chilled, stored in air	*Shewanella* putrefaciens-like [a,b]
	Pseudomonas spp.[c]
Fresh, chilled and stored in vacuum or modified-atmosphere packaging	*Photobacterium phosphoreum*[b]
	Lactic acid bacteria[c]
	Brochothrix thermosphacta[c]
Fresh and lightly preserved products stored at ambient temperature	*Aeromonas* spp.
	Vibrio spp.
	Enterobacteriaceae
	Enterococcus faecalis
Sous-vide cooked and chill stored	Gram-positive spore formers
Lightly preserved[d] and chill-stored	Lactic acid bacteria[e]
	Enterobacteriaceae[f]
	P.phosphoreum and other marine *Vibrio* spp.
Semipreserved, salt cured and chilled	*Halobacterium* spp., *Halococcus* spp. and osmotolerant molds and yeasts
Fermented and chilled	Molds and lactic acid bacteria

[a]Refer to *Shewanella putrefaciens, Shewanella baltica*, and other closely related H_2S-producing, Gram-negative bacteria.
[b]Typical of fishes from marine and temperate waters.
[c]Typical of fishes from freshwater and fishes from warmer waters.
[d]Include, for example, cold-smoked salmon, cooked and brined shrimps and brined roe products.
[e]Include, for example, *Lactobacillus curvatus* and *Lactobacillus sake*.

of vacuum-packed seafoods if present in high levels when products are packed. *P. phosphoreum* is primarily important in spoilage of packed seafood (Table 3), but it produces TMA in some aerobically stored fresh fish, and it remains unclear as to what extent this contributes to spoilage of the aerobically stored products. High levels of *Aeromonas* spp., with a substantial spoilage potential, have been detected in both chilled and temperature abused freshwater fish stored aerobically or under modified atmosphere packaging. The importance of these organisms in spoilage needs further study, as do the taxonomy and metabolism of the *Vibrio* spp. and species of Enterobacteriaceae occurring in temperature-abused seafoods (Table 3). *Psychrobacter immobilis* is frequently observed in high concentrations in seafood but is not included as a SSO in Table 3, because of a very limited spoilage potential. Recent studies suggest that lipolytic activity of these organisms contributes to the spoilage of fresh sardines, and this deserves further investigation. Finally, it should be stressed that, for numerous seafoods, the SSOs and their spoilage domain, i.e., the range of product characteristics and storage conditions within which a particular SSO causes spoilage, remain to be identified. This is particularly the case for lightly preserved products processed by various emerging preservation technologies and products containing seafood together with other ingredients.

Chemical and Biochemical Changes

Chemical, biochemical, and other changes in seafood are frequently studied with the objective of identifying indices of freshness or spoilage to be measured by instrumental methods in relation to processing or distribution. Single compound quality indices (SCQI) have been studied to find an individual substance for which the concentration increased during storage to a critical level corresponding to sensory product rejection. Owing to the complex and dynamic nature of seafood spoilage, SCQI often have a limited range of applicability with respect to product characteristics and storage conditions. Consequently, sums or ratios of concentrations of several compounds have been used as indices of freshness or spoilage. More recently, multiple compound quality indices (MCQI) have been developed using multivariate statistical methods to select a combination of instrumentally measured responses that correlate with changes in sensory attributes or remaining shelf-lives of products.

Total volatile nitrogen (TVN) and TMA are important indices of spoilage in fresh and lightly preserved seafood. TVN primarily includes ammonia, TMA, and DMA. The European Commission (Council Regulation No. 95/149/EEC) specified that TVN should be used as an index of spoilage when sensory evaluation casts doubt on fish freshness. Critical limits of 25–35 mg-N of TVN per 100 g have been indicated for different fish species. In TMAO-containing seafood, 10–40 mg-N of TMA per 100 g are typically present at the time of sensory spoilage. Nevertheless, *Pseudomonas* spp. can spoil, e.g., fresh Nile Perch without transforming its content of TMAO into TMA. Off-flavors caused by volatile amines depend on pH and possibly other product characteristics. Sugar-salted herring with a pH of <6, for example, may contain as much as ~ 75 mg-N of

TVN per 100 g and ~ 35 mg-N of TMA per 100 g without appearing sensorially spoiled. Even in TMAO-containing seafood, little TMA is produced until the SSO reaches high numbers and TMA can only be used as an index of spoilage and not as an index of freshness (Figure 1). Gadoid fishes and maybe some other species contain the enzyme TMAOase (EC 4.1.2.32) in their muscle tissue. The enzyme transforms TMAO into DMA and formaldehyde. The later causes protein denaturation and textural changes in fish muscle tissue, and the DMA content correlates with the sensory texture ratings of thawed hake. TMAOase is most active at -5 to $-10\,^{\circ}C$, but in hake, substantial amounts of DMA can be formed during frozen storage above $-20\,^{\circ}C$. In contrast with white-fleshed fish, considerable amounts of ammonia can develop in cephalopods, crustaceans, and dark-fleshed fish like herring, owing to their high content of free amino acids. Clearly, the usefulness of TVN, TMA, and DMA as indices of spoilage depends strongly on the fish species and storage conditions.

In chilled fresh and lightly preserved seafood, adenosine triphosphate (ATP) is degraded sequentially into adenosine diphosphate (ADP), adenosine monophosphate (AMP), inosine monophosphate (IMP), inosine (Ino), and hypoxanthine (Hx). IMP is formed by autolytic enzymes, whereas spoilage bacteria contribute to Ino and particularly Hx formation. The K-value $= ([Ino] + [Hx]) \times 100/([ATP] + [ADP] + [AMP] + [IMP] + [Ino] + [Hx])$ and the quantitatively similar K_i-value $= ([Ino] + [Hx]) \times 100/([IMP] + [Ino] + [Hx])$ and G-value $= ([Ino] + [Hx]) \times 100/([AMP] + [IMP] + [Ino])$ are used as measures of fish freshness. In Japan, the critical K-value limit for fish to be consumed raw is 20. These values increase linearly during the first days of chilled storage of most fishes and therefore can be used as indices of freshness. However, in some fish species, e.g., redfish, the K-value reaches maximal levels well before sensory rejection and cannot be used as a general index of spoilage (Table 2).

Production of biogenic amines in seafood varies strongly owing to large variations in the free amino acid precursors between fish species and the variability in the ability of spoilage bacteria to produce agmatine, cadaverine, histamine, putrescine, spermidine, spermine, tyramine, and tryptamine. Consequently, individual biogenic amines are rarely appropriate as SCQI. However, different ratios or sums of concentrations of biogenic amines as well as dedicated MCQI have been suggested and critical limits corresponding to the end of shelf-life determined, e.g., for fresh salmon, herring, lobster tails, mahi-mahi, plaice, rockfish, shrimps, tuna, whiting, and cold-smoked salmon. Biogenic amines are heat-stable and therefore, together with organic acids, appropriate for evaluation of raw material used in canned products. The effect of temperature, NaCl, pH, and atmosphere on microbial formation of biogenic amines in seafood needs further study.

The activity of spoilage microorganisms frequently results in the formation of volatile sulfur compounds, short-chain organic acids, aldehydes, ketones, and alcohols, and some of these metabolites are clear indices of spoilage owing to their very unpleasant odors. However, critical concentrations corresponding to the end of shelf-life remain to be determined and validated.

Degradation of lipids by enzymatic hydrolysis or chemical oxidation can result in rancid off-flavors in seafood. In chilled but particularly in temperature-abused fresh fatty fish, enzymes from tissue, intestinal content, and possibly microorganisms hydrolyze triacylglycerol into free fatty acids, causing the so-called acid rancidity. Free fatty acids are also formed by enzymatic hydrolysis during frozen storage of both lean and fatty fishes. With salmon, frozen storage of raw and cooked samples confirmed enzymatic hydrolysis as being responsible for the formation of free polyunsaturated fatty acid. Aroma-recombination studies with fresh fish showed these acids to cause the same oily, bitter, and metallic taste as observed after frozen storage of salmon. Lipid oxidation limits the shelf-life of several frozen and dried seafoods as well as of fish oil-containing products. The peroxide value (PV) and the thiobarbituric acid reactive substance (TBARS) value are classic measures of primary and secondary oxidation products, respectively. Primary oxidation products have no off-flavor and may be both formed and degraded during storage of seafoods. However, seafoods with PV of 10–20 meq of peroxide per kilogram of lipid and TBARS values of 1–2 μmol of malonaldehyde equivalents per gram of product fat often have rancid off-flavors.

It is laborious to identify microbial metabolites, oxidation products, or other compounds to be used as indices of spoilage. Chromatographic (gas chromatography and high-performance liquid chromatography) and spectroscopic techniques including near infrared and nuclear magnetic resonance can be used to determine various changes in seafoods during storage. Chromatographic and spectroscopic techniques provide large amounts of data, and in order to extract compounds or responses that correlate with sensory changes, multivariate statistical methods, particularly partial least-squares regression, can be applied. Identified combinations of compounds or responses can be used as MCQI, and the approach may facilitate development of methods for rapid detection of

relevant compounds, e.g., by using electronic noises or visual indicators. Future studies are likely to identify various MCQI for determination of spoilage cause by microbial activity, autolytic changes, or chemical reaction in seafood.

Prediction of Shelf-life

Transportation of seafood and fish raw material is increasing globally, and this has resulted in needs for methods to evaluate and predict effects of conditions of storage and distribution on shelf-life. Quality attributes can be measured in newly processed products, but unless changes in these attributes can be predicted as a function of storage conditions, measurements at the time of processing will be of little value to evaluate products as experienced by consumers/buyers. Empirical models for relative rates of spoilage, kinetic models for growth of SSO, and relations between indices of spoilage and remaining product shelf-life or time of storage have been suggested for shelf-life prediction.

Empirical Relative Rates of Spoilage Models

Temperature is the single most important factor influencing the shelf-life of fresh and lightly preserved seafood. Despite the complex and dynamic nature of seafood spoilage, it has been possible to predict the effect of temperature on the shelf-life of various seafoods. The predictions rely on relative rates of spoilage (RRS), i.e., the keeping time at a reference temperature like 0 °C divided by the keeping time at T(°C). The effect of temperature on RRS is similar for various species of fresh fish, but at the same time, RRS may differ substantially between fresh and lightly preserved products. As shown in Figure 2, the apparent activation energy (E_A), for the effect of temperature on shelf-life, can be only $\sim 20\,\text{kJ mol}^{-1}$ ($\sim 5\,\text{kcal mol}^{-1}$) for hot smoked fish and $>100\,\text{kJ mol}^{-1}$ ($>24\,\text{kcal mol}^{-1}$) for cooked and brined shrimps stored in modified-atmosphere packaging. Average RRS values of different fresh seafoods from cold and temperate waters, including superchilled and modified-atmosphere-packed products can be predicted between -3 and $+15$ °C by using eqn. 1, named the square-root spoilage model. For other seafoods, the Arrhenius model, which includes the parameter E_A to describe the apparent activation energy, or a simple exponential model, including the slope parameter, b (eqn. 2), is more appropriate for the effect of temperature on RRS (Figure 2). RRS models allow shelf-life to be predicted at different storage temperatures if only the shelf-life of a product has been determined at a single known temperature (T_{ref}) (see eqn. 2).

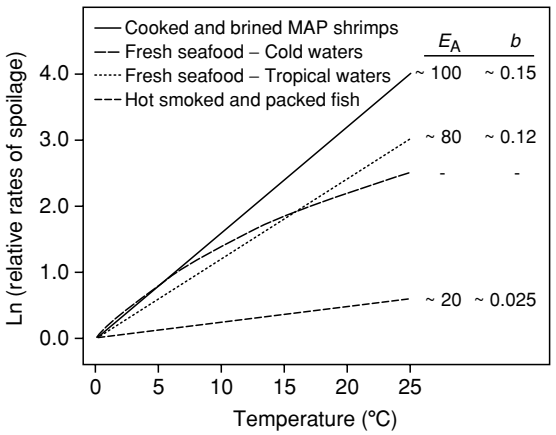

Figure 2 Relative rates of spoilage and corresponding temperature characteristics. Apparent activation energy determined by the Arrhenius equation (E_A) and slope parameter in exponential spoilage model (b). The square root model appropriate for fresh seafood from cold waters is shown in eqn (1). Modified from Dalgaard P (2000) Fresh and lightly preserved seafood. In: Man CMD and Jones AA (eds.) *Shelf Life Evaluation of Foods*, 2nd edn, pp. 110–139. Gaithersburg, MD: Aspen with permission.

$$\sqrt{\text{RRS}} = \sqrt{\frac{\text{shelf-life at } 0°C}{\text{shelf-life at } T\ °C}} = 1 + 0.1 \times T(°\text{C}) \quad (1)$$

$$\text{shelf-life at } T(°\text{C}) = \frac{\text{shelf-life at } T_{ref}(°\text{C})}{\exp\left[b \times (T - T_{ref})\right]} \quad (2)$$

Kinetic Models for Growth of SSO

Mathematical models are available to predict the effect of temperature, atmosphere, water activity, and other factors on the growth of some SSO from seafood as well as on the growth of several pathogenic microorganisms. Nevertheless, only some models have been validated and confirmed to predict shelf-life and growth of pathogenic microorganisms in seafood accurately, *i.e.*, within $\pm 25\%$.

Growth of SSO in aerobically stored seafood is often exponential until the end of the shelf-life is reached (Figure 1) or includes only a very short lag phase. This has allowed the shelf-life of different aerobically stored fishes to be predicted from growth of the SSO, *i.e.*, *Shewanella putrefaciens* and psychrotolerant *Pseudomonas* spp. The square-root model (eqn. 3) is typically used to predict the effect of temperature on maximum specific growth rates (μ_{max}) of spoilage and pathogenic microorganisms in seafood.

$$\sqrt{\mu_{max}} = \alpha(T - T_{min}). \quad (3)$$

More extensive models, including the effect of temperature and carbon dioxide concentrations, are available to predict the growth of *P. phosphoreum*

Table 4 Seafood associated hazards and their prevention

Hazard	*Safety concern*	*Preventive measures*
Histamine	Microbial production of above *c*. 500 p.p.m. in products	Reduce microbial growth and activity by storage of fresh seafood at $< 2\,^{\circ}C$ and lightly preserved products at $< 5\,^{\circ}C$
Biotoxins	Consumption of warm-water fishes with tetrodo- or ciguatoxin or filter feeding shellfish with toxins produced by marine algae	Monitoring and management of fishing areas. Biotoxins persist cooking and preservation of seafood
Viruses	Viral gastoenteritis due to consumption of raw or steamed molluscan shellfish and ready-to-eat seafoods	Monitoring of harvesting areas, depuration of molluscs, cooking seafood prior to consumption
Pathogenic bacteria indigenous in aquatic environment		
Clostridium botulinum, psychrotolerant	Growth and toxin formation in ready-to-eat (hot-smoked and fermented) products from colder regions	Salting ($> 3.5\%$ NaCl) and chilled storage ($< 5\,^{\circ}C$) prevent growth and toxin production in seafood
Vibrio parahaemolyticus, V. vulnificus, V. cholerae, V. algionlyticus and *Plesiomonas shigelloides*	Growth to high levels in raw marine seafoods from warmer waters and in cross-contaminated ready-to-eat products	Avoid consumption of improperly cooked seafood, particularly molluscan shellfish
Pathogenic bacteria from the animal/human reservoir and general environment		
Salmonnella spp. and *Shigella* spp.	Contamination by even low levels	Avoid direct or indirect faecal contamination of products. Bacteria is inactivated by cooking
Clostridium botulinum proteolytic type A and B	Growth and toxin formation in products	Inactivate bacteria and their spores by canning (> 2.4 min at $121\,^{\circ}C$) or prevent growth by chilling ($< 10\,^{\circ}C$), acidification (pH < 4.5) or curing ($> 10\%$ NaCl).
Listeria monocytogenes	Growth in ready-to-eat seafoods	Reduce growth by low storage temperature. Bacteria are inactivated by cooking
Staphylococcus aureus	Growth and toxin formation in products	Prevent growth by chilled storage of products
Campylobacter spp.	Contaminated molluscan shellfish. Bacteria die in seawater	Avoid consumption of raw molluscan shellfish
Parasites	Intake of live nematodes, or trematodes in raw and ready-to-eat seafood	Removal of infected fish or inactivation of parasites by heating ($55\,^{\circ}C$, 1 min), freezing ($-20\,^{\circ}C$, 24 h) or curing

and *B. thermosphacta* as well as the shelf-life of fresh modified-atmosphere-packed seafoods. Kinetic models predict shelf-life from the initial level of SSOs in products, and further studies are needed to determine typical distributions of spoilage organisms in seafoods.

Time–Temperature Integration

Mathematical shelf-life models can be used to evaluate product temperature profiles as recorded during distribution of seafoods. Software that includes both RRS models and kinetic models for growth of SSO is available to predict product shelf-life under constant and fluctuating storage conditions. The seafood spoilage predictor (SSP) (available from http://www.dfu.min.dk/micro/ssp/) is an example of a widely used program. Time–temperature integration (TTI) tags that change color in response to a temperature history are another type of device to evaluate product shelf-life. Storage temperature influences the shelf-life of different fresh and lightly preserved seafoods very differently (Figure 2), and selection of TTI tags with appropriate temperature kinetics is of utmost importance for successful application of this technology. TTI tags with apparent activation energies that correspond to those of most seafoods are available. However, for semi- and lightly preserved seafoods with an extended shelf-life (Table 1), existing tags relying on enzymatic, polymerization, or diffusion reactions may not always be available with appropriate response times.

Spoilage and Safety of Seafood

Growth of spoilage microorganisms in seafood reduces the risk of pathogenic microorganisms reaching high levels before products are spoiled and therefore will not be consumed. Spoilage microorganisms and reactions can be inhibited by classical processing and preservation techniques, *e.g.*, salting, drying, smoking, canning, and freezing (see appropriate articles in the present chapter). In addition, hurdle technology, minimal processing, and targeted inhibition of SSO can be used as mild preservation procedures to extend the shelf-life of seafood. Clearly, inhibition of spoilage reactions should be linked with a carefully evaluated product safety.

The major hazards responsible for seafoodborne diseases are histamine, biotoxins, viruses, bacteria and parasites (Table 4). Seafood is estimated to cause up to 10% of all outbreaks and approximately 3% of all cases of foodborne disease. The safety of seafood products varies considerable and is influenced by a number of factors, including the origin of the fish or shellfish (marine or freshwater; cold or tropical regions; clean or polluted waters), microbiological ecology of the product, handling and processing practices, and preparation before consumption (Table 4).

Histamine fish poisoning is the single most frequent seafoodborne disease ($\sim 30\%$ of all cases). Fortunately, symptoms including facial flushing, vomiting, diarrhea, and headache are relatively mild and of short duration. Molluscan shellfish is the type of seafood most frequently causing seafoodborne diseases ($> 50\%$ of all cases). These shellfish are filter-feeders and concentrate algae, bacteria, viruses, and chemicals from the surrounding water. Despite this situation, molluscan shellfish (with intestines) are often consumed raw and then may cause viral gastroenteritidis (Norwalk-like and hepatitis A), infections by *Vibrio* spp. and various shellfish toxic syndromes resulting from toxins produced by different marine algae (dinoflaggelates). In regions of Asia, raw freshwater fishes are used in various dishes, and as a result of this, a high percentage of the population can be infected with parasites (trematodes). Several other aspects of seafood safety are less closely linked to consumer behavior (Table 4). Because of the increasing international trade with fish raw material and seafoods, the safety of these products is of global importance. Continued efforts to apply hazard analysis critical control point systems in all steps of the seafood chain are essential to improve the safety record of these products at an international level.

See also: **Chilled Storage**: Microbiological Considerations; ***Clostridium***: Botulism; **Fish**: Tuna and Tuna-like Fish of Tropical Climates; **Food Poisoning**: Classification; **Histamine**; ***Listeria***: Listeriosis; **Minimally Processed Foods**; **Preservation of Food**; **Sensory Evaluation**: Sensory Characteristics of Human Foods; **Shellfish**: Contamination and Spoilage of Molluscs and Crustaceans; **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage; **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability; Shelf-life Testing

Further Reading

Ashie INA, Smith JP and Simpson BK (1996) Spoilage and shelf life extension of fresh fish and shellfish. *Critical Reviews in Food Science and Nutrition* 36: 87–121.

Colby J-W, Enriquez-Ibarra LG and Flick GJ (1993) Shelf life of fish and shellfish. In: Charalambous G (ed.) *Shelf Life Studies of Foods and Beverages*, pp. 85–143. Amsterdam: Elsevier.

Dalgaard P (2000) Fresh and lightly preserved seafood. In: Man CMD and Jones AA (eds) *Shelf Life Evaluation of Foods*, 2nd edn. pp. 110–139. Gaithersburg, MD: Aspen.

Feldhusen F (2000) The role of seafood in bacterial foodborne diseases. *Microbes and Infection* 2: 1651–1660.

Gram L and Huss HH (2000) Fresh and processed fish and shellfish. In: Lund BM, Baird-Parker AC and Gould GW (eds) *The Microbiological Safety and Quality of Foods*, pp. 472–506. Gaithersburg, MD: Aspen.

Huss HH (1995) *Quality and Quality Changes in Fresh Fish*. FAO Fisheries Technical Paper. No. 348.

Huss HH, Ababouch L and Gram L (2003) *Assessment and Management of Seafood Safety and Other Quality Aspects. FAO Fisheries Technical Paper.*

Leisner JJ and Gram L (2000) Spoilage of fish. In: Robinson RK, Batt CA and Patel PA (eds) *Encyclopedia of Food Microbiology*, pp. 813–820. San Diego: CA.

Olafsdóttir G, Luten J, Dalgaard P *et al.* (1998) *Methods to Determine the Freshness of Fish in Research and Industry.* Paris: International Institute of Refrigeration.

Dietary Importance of Fish and Shellfish

A Ariño, J A Beltrán and P Roncalés, University of Zaragoza, Zaragoza, Spain

Background

During the 1990s, the world's population consumed ever larger quantities of fish. In discussing food uses of fish, the term 'fish' refers to edible species of finfish, molluscs, and crustaceans coming from the marine or freshwater bodies of the world, either by capture fisheries or by aquaculture. In 1990, direct human consumption of fish amounted to 71.7 million tonnes (live weight equivalent), rising to 93.8 million tonnes in 1997, an increase of almost a third over a period of only seven years. The average availability per capita was 15.9 kg per year during 1996, and fish provided (per day) 28 kcal, 4.4 g of protein, and 1.0 g of fat. There are wide differences among countries in fish consumption, as measured by the average yearly intake per person, ranging from countries with less than 1.0 kg per person to countries with over 100 kg. In recent years, the volume of fishery products marketed fresh has increased; in 1997, about a third of all fish was marketed fresh, compared with only a fifth in 1987.

Edible fish muscle contains 18% protein and 1–2% ash; the percentage of lipids varies from less than 1% to more than 20% in high-fat finfish. While fish has a prominent role as a source of valuable animal proteins in many communities, it seldom has a dietary significance as a source of kilocalories. Overall, fish provides about 16% of the animal proteins in the world and is a source of vitamins, minerals, and essential fatty acids. Food and nutritional professionals and consumers have known for years that fish is a high-protein food that is low in energy, total fat, and saturated fat when compared with other protein-rich animal foods. In addition, a large proportion of the fat in fish is polyunsaturated, the ω-3 (*n*-3) fatty acids.

Because of increased evidence for the cardiovascular benefits of fish (particularly fatty fish), consumption of at least two fish servings per week is recommended to maintain health. The predominant beneficial effects include a reduction in sudden death, decreased risk of arrhythmia, lower plasma triacylglycerol levels, and a reduced blood-clotting tendency.

As fish have become more popular, there have been increasing reports of foodborne diseases attributed to these foods. Foodborne diseases following exposure to fish can result from the food itself (toxic species, allergies), but also bacterial or viral contamination of the fish, naturally occurring seafood toxins, or the presence of additives and chemical residues due to environmental contamination.

General Characteristics of Food Finfish

A very large number of species of finfish are used for food by the world's population as a whole. The dressing percentage of finfish (60–70%) is similar to that of beef, pork, or poultry. The percentage of edible tissue in dressed carcasses of finfish (without head, skin, and viscera) is higher than that of other food animals, because fish contains less bone, adipose tissue, and connective tissue. There are three main categories of finfish widely used as foods. The bony fishes, the Teleosts, provide two compositional groups: white fish (or lean fish) such as cod, haddock, most flat fishes and whiting, and fatty fish, such as herring, sardines, salmon, mackerel, and tuna. The third category includes the cartilaginous Elasmobranch fish, such as dogfish, shark, and skate.

White Fish

The flesh of these fish is very low in fat and consists primarily of muscle and thin layers of connective tissue. The concentrations of most of the B vitamins are similar to those in mammalian lean meats, though fish may contain higher amounts of vitamin B_6 and B_{12}. The mineral levels are also similar, although the very fine bones that are eaten with the fish flesh can raise the calcium content; fish is also a significant source of iodine. These fish accumulate oils in their livers, which are a rich source of vitamin A (retinol), vitamin D and long-chain polyunsaturated fatty acids (PUFA) in their triacylglycerols (TAGs).

Fatty Fish

These fish have fat in their flesh, which is usually much darker than that of white fish, with similar blocks of muscle and connective tissue. The amount of fat is related to the breeding cycle of the fish, and after breeding, the fat content falls considerably. The flesh of the fatty fish is generally richer in the B vitamins than that of white fish, and there are significant amounts of vitamins A and D present. The mineral concentrations are not very different, but fatty fish is a better source of iron. The fat of these fish, or more correctly the oil, is particularly rich in very-long-chain PUFA, especially those of the ω-3 (*n*-3) series such as eicosapentaenoic (EPA) and docosahexaenoic (DHA) acids. These fish accumulate oils in their muscles, belly flap, and skin (subdermal fat).

Cartilaginous Fish

These fish include the sharks and rays, whose flesh is rich in connective tissue and relatively low in fat, although they do accumulate oils in their livers. Compositionally, the concentrations of the vitamins and minerals are very similar to those in white fish. These fish are remarkable in that they have a high urea content, and so protein values based on total nitrogen are overestimated.

General Characteristics of Food Shellfish

The term 'shellfish' includes any aquatic invertebrate, such as a mollusc or crustacean that has a shell or shell-like exoskeleton. Owing to the presence of tough exoskeletons, the edible portion in shellfish (around 40%) is less than in finfish, with the exception of cephalopods, whose dressing percentage is 70–75%. In many communities, these foods are very highly valued gastronomically.

Molluscs

A wide range of molluscs are eaten by man, including bivalves such as mussels, oysters, and scallops, gastropods such as winkles and whelks, and cephalopods such as squids and octopuses. The flesh is muscular with low levels of fat, although the fat is more saturated and richer in cholesterol than that of finfish. The mineral levels in shelfish are usually somewhat higher than in finfish, and the vitamin concentrations are low. The bivalves and gastropods are often eaten whole after boiling and sometimes raw; usually, only the muscular mantles of cephalopods are eaten, after cooking. In some cultures, only selected parts are eaten, for example only the white adductor muscle of the scallop is eaten in North America.

Crustacea

These include a range of species both fresh water such as crayfish, and marine such as crabs, shrimps, prawns, and lobsters. These animals have a segmented body, a chitinous exoskeleton, and paired, jointed limbs. The portions eaten are the muscular parts of the abdomen and the muscles of claws of crabs and lobsters. The flesh is characteristically low in fat and high in minerals, with vitamin levels similar to those found in finfish.

Nutritional Role of Fish and Shellfish

Interest in the health benefits of fish and shellfish began decades ago when researchers noted that certain groups of people – including the Inuit and the Japanese, who rely on fish as a dietary mainstay – have a low rate of ischemic diseases (i.e., heart attack or stroke). Part of the cardiovascular health benefits of fish oils and their *n*-3 fatty acids can be explained by the ability of both EPA and DHA present in fish to reduce serum TAGs and very-low-density lipoproteins (VLDL-cholesterol). A growing body of evidence indicates that foods rich in ω-3 PUFAs confer cardioprotective effects beyond those that can be ascribed to improvements in blood lipoprotein profiles. As a result, current recommendations point out the benefits of eating fish at least twice a week, though even relatively small amounts of fish may decrease the overall risk of death from heart disease. Most experts do not advise routine use of fish oil supplements, and they favor eating fish and shellfish regularly in the context of a healthy diet and a regular pattern of physical activity.

The caloric value of fish is related to fat content and varies with species, size, diet, and season. Fish is generally considered to be a low-calorie food when compared with meats and poultry. Most lean or lower fat species of fish such as cod, hake, flounder, and sole contain less than 100 kcal per 100-g portion, and even fatty fish like mackerel, herring, and salmon contain approximately 250 kcal or less in a 100-g serving. Most crustaceans contain less than 1% fat in the tail muscle because depot fat is stored in the hepatopancreas, which is in the head region. With fish, each consumer can ingest fewer calories to meet their daily protein needs, which makes fish a good choice for diets designed to help you lose or maintain a desirable weight.

Fish Lipids, ω-3 PUFA Content, and Health

In fish, the depot fat is liquid at room temperature (oil) and is seldom even visible to the consumer; an exception is in the belly flaps of salmon steaks.

Many species of finfish and almost all shellfish contain less than 2.5% total fat, and less than 20% of the total number of calories comes from fat. Almost all fish have less than 10% total fat, and even the fattiest fish such as herring, mackerel, and salmon have no more than 20% fat (see Table 1). To obtain a good general idea of the fat content of most finfish species, the color of the flesh should be considered. The leanest species such as cod and flounder have a white or lighter color, and the fattier fish such as salmon, herring, and mackerel have a much darker color.

The triacylglycerol depot fat in edible fish muscle is subject to seasonal variations in all marine and freshwater fish from all over the world. Fat levels tend to be higher during times of the year when fish are feeding heavily (usually during the warmer months) and in older and healthier individual fish. Fat levels tend to be lower during spawning or reproduction. When comparing fat contents between farmed and wild-caught food fish, it should be considered that farmed species have a tendency to show a higher proportion of muscle fats than their wild counterparts. Also, the fatty acid composition in the lipids of farmed fish is dependent on the type of dietary fat used for raising the fish. Cholesterol is independent of fat content and similar in wild and cultivated fishes.

Most protein-rich foods including red meat and poultry as well as fish contain cholesterol. Almost all types of fish and shellfish contain well under 100 mg of cholesterol per 100 g, and many of the leaner types of fish have typically 40–60 mg for 100 g of edible muscle. It is known that most shellfish also contain less than 100 mg of cholesterol per 100-g serving. Shrimp contain somewhat higher amounts of cholesterol, over 150 mg per 100 g, and the squids are the only fish product with a significantly elevated cholesterol content, which averages from 300 mg per 100-g portion. Fish roe, caviar, the internal organs of fish (such as livers), the tomalley of lobsters and the hepatopancreas of crabs can contain high amounts of cholesterol.

ω-3 PUFA in Fish and Shellfish

The polyunsaturated fatty acids of many fish lipids are dominated by two members of the ω-3 family (*n*-3), C20:5 *n*-3 (EPA) and C22:6 *n*-3 (DHA). They are so named because the first of the several double bonds occur three carbon atoms away from the terminal end of the carbon chain.

All fish and shellfish contain some ω-3 but the amount can vary, as the relative concentrations of PUFAs are species-specific (see Table 2). Generally, the fattier fish contain more ω-3 fatty acids than the leaner fish. The amount of ω-3 fatty acids in farm-raised products can also vary greatly, depending on the diet of the fish or shellfish. Many companies now recognize this fact, and provide a source of ω-3 fatty acids in their fish diets. ω-3 fatty acids can be destroyed by heat, air, and light, so the less processing, heat, air exposure, and storage time, the better the ω-3 can be preserved in fish. Freezing and normal cooking cause minimal ω-3 losses, whereas deep frying and conditions leading to oxidation (rancidity) can destroy some ω-3 fatty acids.

The beneficial effects of eating fish for human health have been well documented. Research has shown that EPA and DHA are useful in (1) treatment or prophylaxis of thromboembolic conditions, (2) reducing cardiovascular mortalities and increased survival of patients, (3) decreasing stroke incidence, and (4) preventing fatal heart arrhythmia. DHA could also play an important role for full brain and eye tissue development and functions (learning ability and visual

Table 1 Fat levels in fish and shellfish commonly found in the marketplace

Low (<2.5% fat), less than 20% of total calories from fat	*Medium (2.5–5% fat), between 20 and 35% of total calories from fat*	*High (>5% fat) between 35 and 50% total calories from fat*
Cod, haddock, pollock	Anchovy	Catfish (farmed)
Grouper, Snapper	Bluefish	Eel[a]
Most flatfishes	Bream	Herring[a]
Pike	Carp	Mackerel[a]
Shark, skate	Swordfish	Salmon[a]
Tilapia	Trout	Sardine
Whiting, hake		Tuna (bluefin)
Most molluscs		Whitefish
Most crustaceans		

[a]More than 10% fat.

Table 2 Selected fish and shellfish grouped by their ω-3 fatty acid content

Low-level group (<0.5 g per 100 g)	*Medium-level group (0.5–1 g/100 g)*	*High-level group (>1 g per 100 g)*
Carp	Bass	Anchovy
Catfish	Bluefish	Herring
Cod, haddock, pollock	Halibut	Mackerel (most species)
Grouper	Pike	Sablefish
Most flatfishes	Red Snapper	Salmon (most species)
Perch	Swordfish	Tuna (bluefin)
Snapper	Trout	Whitefish
Tilapia	Whiting	
Most molluscs	Clams	
Most crustaceans	Oysters	

acuity) in late pregnancy or in preterm infants. Possible relationships between ω-3 fatty acids and other disorders such as digestive tract cancers, rheumatoid arthritis, inflammatory reactions, and asthma are also currently being studied (see Table 3).

Fish proteins

Both finfish and shellfish are a highly valuable source of proteins in human nutrition. The protein content of fish flesh, in contrast to that of fat, is highly constant, independent of seasonal variations related to the feeding and reproductive cycles, and shows only small differences among species. Table 4 summarizes the approximate protein content of the various finfish and shellfish groups. Fatty finfish and crustacea have slightly higher than average protein concentrations. Bivalves have the lowest values if the whole body mass is considered, since most of them are usually eaten whole, whereas values are roughly average if specific muscular parts alone are consumed; the latter is the case with the scallop, in which mostly only the adductor muscle is eaten.

The essential amino acid composition is given in Table 5. Fish proteins, with only slight differences among groups, possess a high nutritive value, similar to that of meat proteins and slightly lower than that of egg. It is worth pointing out the elevated supply, relative to meat, of essential amino acids such as lysine, methionine, and threonine. In addition, due in part to the low collagen content, fish proteins are easily digestible, giving rise to a digestibility coefficient of nearly 100.

The recommended dietary allowances (RDA) or dietary reference intakes (DRI) of protein for human male or female adults are within the range of 45–65 g per day. In accordance with this, an intake of 100 g of fish would contribute 15–25% to the total daily protein requirement of healthy adults and some 70% to that of children. A look towards the dietary importance of Mediterranean diet is convenient; one of its characteristics is the high consumption of all kinds of fish, chiefly fat fish. In many Mediterranean countries, fish intake averages over 50 g per day (edible flesh); thus, fish protein contributes over 10% to the total daily protein requirements steadily over the whole year in those countries.

Less known is the fact that the consumption of fish protein, independently of the effect exerted by fish fat, has been related to a decrease in the risk of atherogenic vascular diseases. In fact, it has been demonstrated that diets in which fish is the only source of protein increase the blood levels of HDL in comparative studies with milk or soya proteins.

Nonprotein Nitrogen (NPN) Compounds in Fish

NPN compounds are mostly in the fiber sarcoplasm and include free amino acids, peptides, amines, amine oxides, guanidine compounds, quaternary ammonium molecules, nucleotides, and urea (Table 6). This fraction accounts for a relatively high percentage of total nitrogen in the muscle of some aquatic animals, 10–20% in teleosts, about 20% in crustacea and molluscs, 30–40% – and in special cases up to 50% – in elasmobranchs. In contrast, NPN in land animals usually represents no more than 10% of total nitrogen.

Most marine fish contain trimethylamine oxide (TMAO); this colorless, odorless, and flavorless compound is degraded to trimethylamine (TMA),

Table 3 Summary of beneficial effects of eating fish for the cardiovascular condition and other diseases

Cardiovascular disease (CVD)	*Other diseases*
Protects again heart disease	Protects against age-related maculopathy
Prolongs lives of people after a heart attack	Alleviates rheumatoid arthritis
Protects against sudden cardiac arrest caused by arrhythmia	Protects against certain types of digestive tract cancers
Protects against stroke (thrombosis)	Mitigates inflammation reactions and asthma
Lowers blood lipids such as triacylglycerols and VLDL-cholesterol	
Lowers blood pressure	

Table 4 Protein content of different groups of fish and shellfish

Fish group	*g per 100 g*
White finfish	16–19
Fatty finfish	18–21
Crustacea	18–22
Bivalves	10–12
Cephalopods	16–18

Table 5 Proximate content of essential amino acids in fish and shellfish (g per 100 g of protein)

Fish group	*Ile*	*Leu*	*Lys*	*Met*	*Phe*	*Thr*	*Trp*	*Val*
Finfish	5.3	8.5	9.8	2.9	4.2	4.8	1.1	5.8
Crustacea	4.6	8.6	7.8	2.9	4.0	4.6	1.1	4.8
Molluscs	4.8	7.7	8.0	2.7	4.2	4.6	1.3	6.2

Table 6 Nonprotein nitrogen compounds in several commercially important fish species and mammalian muscle (mg per 100 g wet weight)

Compounds	*Cod*	*Herring*	*Shark species*	*Lobster*	*Mammal*
Total NPN	1200	1200	3000	5500	3500
Total free amino acids:	75	300	100	3000	350
Arginine	< 10	< 10	< 10	750	< 10
Glycine	20	20	20	100–1000	< 10
Glutamic acid	< 10	< 10	< 10	270	36
Histidine	< 1.0	86	< 1.0		< 10
Proline	< 1.0	< 1.0	< 1.0	750	< 1.0
Creatine	400	400	300	0	550
Betaine	0	0	150	100	
Trimethylamine oxide	350	250	500–1000	100	0
Anserine	150	0	0	0	150
Carnosine	0	0	0	0	200
Urea	0	0	2000		35

which gives a 'fishy' odor and causes consumer rejection. This compound is not present in land animals and freshwater species (except Nile perch and tilapia from Lake Victoria). TMAO reductase catalyzes the reaction and is found in several fish species (red muscle of scombroid fish, white and red muscle of gadoids) and in certain microorganisms (*Enterobacteriaceae*).

Creatine is quantitatively the main component of the NPN fraction. This molecule plays an important role in fish muscle metabolism through its phosphorylated form; it is absent in crustacea and molluscs.

Endogenous and microbial proteases yield some free amino acids; taurine (Tau), alanine (Ala), glycine (Gly) and imidazole-containing amino acids seem to be the most frequent. Gly and Tau contribute to the sweet flavor of some crustaceans. Migratory marine species such as tuna, characterized by a high proportion of red muscle, have a high content (about 1%) of free histidine (His). A noticeable amount of this amino acid has also been reported in freshwater carp (*Cyprinus carpio*). The presence of free His is relevant in several fish species because it can be microbiologically decarboxylated to histamine (see Scombroid Fish (Histamine) Poisoning below).

Nucleotides and related compounds generally play an important role as coenzymes. They participate actively in muscle metabolism and supply energy to physiological processes. They have a noticeable participation in flavor; moreover, some of them may be used as freshness indices. Adenosine triphosphate (ATP) is degraded to adenosine diphosphate (ADP), adenosine monophosphate (AMP), inosine monophosphate (IMP), inosine (Ino), and hypoxantine (Hx). This pattern of degradation takes place in finfish, whereas AMP is degraded to adenosine and thereafter to Ino in shellfish. The degradation chain is very fast to IMP or AMP in finfish and shellfish, respectively. IMP degradation to Ino is generally slow, except in scombroids and flat fish. Ino degradation to Hx is slower. IMP is a flavor potentiator, whereas Hx imparts a sour taste, and high levels are toxic. ATP, ADP and AMP decompose quickly to form a build-up of Ino and Hx. As this corresponds well with a decline in freshness, the ratio of the quality of inosine and hypoxanthine to the total quantity of ATP and related substances is called the *K*-value and used as an index representing freshness of fish meat.

Guanosine is an insoluble compound that gives the characteristic brightness of fish eye and skin. It is degraded to guanine, which does not have this property; therefore, brightness decreases until it completely disappears.

The NPN fraction contains other interesting compounds, such as small peptides. Most of them contribute to flavor; besides this, they have a powerful antioxidant activity. Betaines constitute a special group of compounds that contribute to the specific flavor of different aquatic organisms: homarine in lobster, glycine-betaine and butiro-betaine in crustacea, and arsenic-betaine in crustacea. Arsenic-betaine has the property of fixing arsenic into the structure, giving a useful method for studying water contamination with this element.

Fish Vitamins

The vitamin content of fish and shellfish is rich and varied, although somewhat variable. In fact, significant differences are neatly evident among groups, especially regarding fat-soluble vitamins. Furthermore, vitamin content is highly variable within groups, also showing large differences among species as a function of feeding regimes.

The approximate vitamin concentration ranges of the various finfish and shellfish groups are

summarized in Table 7. The RDA or DRI for male or female adults is also given, as well as the percentage supplied by 100 g of fish. Regarding fat-soluble vitamins, vitamin E (tocopherol) is distributed most equally, showing relatively high concentrations in all fish groups, higher than those of meat. However, only a part of the vitamin E content is available as active tocopherol on consumption of fish, since it is oxidized as it protects fatty acids from oxidation. The presence of vitamins A (retinol) and D is closely related to the fat content, and so they are almost absent in most low-fat groups. Appreciable but low concentrations of vitamin A exist in fatty finfish and bivalve molluscs, whereas vitamin D is very abundant in fatty fish. In fact, 100 g of most fatty species supply over 100% of the RDA of this vitamin.

Water-soluble vitamins are well represented in all kinds of fish, with the sole exception of vitamin C (ascorbic acid), which is almost absent in all of them. The concentrations of the rest are highly variable; however, with few exceptions, they constitute a medium-to-good source of such vitamins, comparable with, or even better than, meat. The contents of vitamins B_2 (riboflavin), B_6 (piridoxin), niacin, biotin and B_{12} (cobalamin) are relatively high. Indeed, 100 g of fish can contribute up to 38, 60, 50, 33, and 100%, respectively, to total daily requirements of those vitamins. Fatty fish also provide a higher supply of many of the water-soluble vitamins than white fish and shellfish, namely piridoxin, niacin, pantothenic acid, and cobalamin. Crustacea also possess a relatively higher content of pantothenic acid, whereas bivalve molluscs have a much higher concentration of folate and cobalamin.

A Mediterranean diet, rich in fish – and especially in fatty finfish – contributes steadily over the year to an overall balanced vitamin supply. The last row of Table 7 illustrates this; the supply of vitamins D, B_2, B_6, B_{12}, and niacin of this particular diet is more than 15% of the daily requirements, whereas all other vitamins, except ascorbic acid, contribute to a lesser, but significant, extent.

Fish Minerals

The approximated composition of fish in selected minerals is given in Table 8. The first point to note is the fact that all kinds of finfish and shellfish present a well-balanced content of most minerals, either macro- or oligoelements, with only a few exceptions. Sodium content is low, similar to other muscle and animal origin foods. However, it must be taken into account that sodium is usually added to fish in most cooking practices in the form of common salt; also,

Table 7 Vitamin content of the different groups of fish and shellfish (mg or μg per 100 g)

	A (μg)	*D (μg)*	*E (mg)*	*B_1 (mg)*	*B_2 (mg)*	*B_6 (mg)*	*Niacin (mg)*	*Biotin (μg)*	*Pantothenic acid (mg)*	*Folate (μg)*	*B_{12} (μg)*	*C (mg)*
White finfish	Tr	Tr	0.3–1.0	0.02–0.2	0.05–0.5	0.15–0.5	1.0–5.0	1.0–10	0.1–0.5	5.0–15	1.0–5.0	Tr
Fatty finfish	20–60	5–20	0.2–3.0	0.01–0.1	0.1–0.5	0.2–0.8	3.0–8.0	1.0–10	0.4–1.0	5.0–15	5.0–20	Tr
Crustacea	Tr	Tr	0.5–2.0	0.01–0.1	0.02–0.3	0.1–0.3	0.5–3.0	1.0–10	0.5–1.0	1.0–10	1.0–10	Tr
Molluscs	10–100	Tr	0.5–1.0	0.03–0.1	0.05–0.3	0.05–0.2	0.2–2.0	1.0–10	0.1–0.5	20–50	2.0–30	Tr
Cephalopods	Tr	Tr	0.2–1.0	0.02–0.1	0.05–0.5	0.3–0.1	1.0–5.0	1.0–10	0.5–1.0	10–20	1.0–5.0	Tr
RDA/DRI	1000	5	10	1.2	1.3	1.3	16	30	5.0	400	2.4	60
%RDA per 100 g	0–10	0–100	2–30	1–20	2–38	5–60	1–50	3–33	2–20	0.3–12	40–100	0
% RDA/M.d.[a]	2	50	7	5	15	25	18	5	8	2	100	0

[a]Mediterranean diet.

Table 8 Selected mineral content of the different groups of fish and shellfish (mg per 100 g)

	Na	*K*	*Ca*	*Mg*	*P*	*Fe*	*Zn*	*Mn*	*Cu*	*Se*	*Cr*	*Mo*	*I*
White finfish	50–150	200–500	10–50	15–30	100–300	0.2–0.6	0.2–1.0	0.01–0.05	0.01–0.05	0.02–0.1	0.005–0.02	0.005–0.02	0.01–0.5
Fatty finfish	50–200	200–500	10–200	20–50	200–500	1.0–5.0	0.2–1.0	0.01–0.05	0.01–0.05	0.02–0.1	0.005–0.02	0.005–0.02	0.01–0.5
Crustacea	100–500	100–500	20–200	20–200	100–700	0.2–2.0	1.0–5.0	0.02–0.2	0.1–2.0	0.05–0.1	0.005–0.02	0.01–0.05	0.01–0.2
Molluscs	50–300	100–500	50–200	20–200	100–300	0.5–10	2.0–10	0.02–0.2	0.02–10	0.05–0.1	0.005–0.02	0.01–0.2	0.05–0.5
Cephalopods	100–200	200–300	10–100	20–100	100–300	0.2–1.0	1.0–5.0	0.01–0.1	0.02–0.1	0.02–0.1	0.005–0.02	0.01–0.2	0.01–0.1
RDA/DRI			1000	420	700	10	15		1.5	0.07			0.15
%RDA per 100 g			1–20	4–50	15–100	2–50	1–30		1–100	25–100			8–100
% RDA/M.d.[a]			6	5	30	18	2		2	100			100

[a]Mediterranean diet.

surimi-based and other manufactured foods contain high amounts of added sodium. Potassium and calcium levels are relatively low, too, though the latter are higher in fish than in meat; in addition, small fish bones are frequently eaten with fish flesh, thus increasing calcium intake. Fish is a good source of magnesium and phosphorus, at least as good as meat. These elements are particularly abundant in crustacea, fatty finfish show elevated levels of phosphorus, and bivalve molluscs have high amounts of magnesium.

Fish is a highly valuable source of most oligoelements. Fatty fish provides a notable contribution to iron supply, similar to that of meat, whereas shellfish have higher concentrations of most dietary metals. In particular, crustacea and bivalve molluscs supply zinc, manganese, and copper concentrations well above those of finfish. Worth mentioning is the extraordinary dietary supply of iodine in all kinds of finfish and shellfish; however, this depends on the concentration present in feed, particularly in plankton organisms.

In summary, 100 g of fish afford low levels of sodium and medium-to-high levels of all the remaining dietary minerals. In fact, they can contribute up to 50–100% of total daily requirements of magnesium, phosphorus, iron, copper, selenium, and iodine. A Mediterranean diet, rich in fatty fish and all kinds of shellfish, can lead to an overall balanced mineral supply, which may well reach over 20% of the daily requirements of phosphorus, iron, selenium, and iodine.

Sanitary Aspects of Fish Consumption

Fish and shellfish consumption, besides its nutritional role as a nutrient source, may also cause syndromes and diseases due to infection or intoxication. Table 9 summarizes the most frequent sources of fish food diseases. Some of the etiologic agents are common to other food products (parasites, bacterial pathogens, viruses, and chemicals), either from the environment or from processing. Fish and shellfish contain other more specific toxic compounds like endogenous toxins, toxins from trophic chains, and bacterial toxins.

Table 9 Fish and shellfish toxicity

1. *Endogenous*
   - Puffer fish poisoning
   - Other less important endogenous toxic molecules: blood toxins, certain fish roe
2. *Toxins from trophic chains*
   - Biological origin
     - Ciguatera poisoning
     - Paralytic shellfish poisoning
     - Diarrhetic shellfish poisoning
     - Neurotoxic shellfish poisoning
     - Amnesic shellfish poisoning
   - Environmental contamination: chemicals, heavy metals, radioactivity, etc.
3. *Microbial activity*
   - Bacterial pathogens
   - Virus
   - Scombroid fish (histamine) poisoning
4. *Parasites*
5. *Toxic compounds from fish processing*

Endogenous Toxins

Puffer Fish Poisoning is characterized by a very potent toxin called 'tetrodotoxin' (TTX), which is produced by some members of the family *Tetraodontidae* (pufferfish). The toxin is mainly found in the liver, ovaries, and intestines, whereas muscle tissue is normally free of this. The symptoms are somewhat similar to those described for paralytic shellfish poisoning (PSP), i.e., initial tingling and numbness of lips, tongue, and fingers, leading to paralysis of the extremities and the respiratory system, which may lead to death. The origin of TTX is not clear; it has been linked to certain common marine vibrios (part of the microflora of pufferfish) that produce a form of the toxin. Low concentrations of the toxin (1–4 mg) may constitute a lethal dose for humans; it must be taken into account that internal organs of pufferfish usually contain 40 $\mu g\,g^{-1}$ in intestines to 400 $\mu g\,g^{-1}$ in ovaries.

Toxins from Trophic Chains

Ciguatera Poisoning results from the consumption of fish that have been fed with toxic dinoflagellates, *Gambierdiscus toxicus* and *Prorocentrum concavum*, being the most common. This poisoning is associated to tropical and subtropical areas and it affects specially large fishes from bottom and reef zones (grouper, barracuda, red snapper and sea basses). The symptoms are both gastrointestinal and neurological, the onset time is usually within a few hours and the prognosis for recovery is good if the patient survives the first 24 h; however, the recovery period can be rather prolonged in some cases.

PSP results from the ingestion of bivalve molluscs that have consumed toxinogenic dinoflagellates (microscopic marine plancktonic algae) of the genera *Alexandrium*, *Gymnodinium*, and *Pyrodinium*. The ultimate cause of PSP is a complex of at least 12 toxins – saxitoxins – with neurological effects. The symptoms appear within 0.5–2 h of eating toxic shellfish and can cause death by respiratory paralysis, usually within 12–24 h; patients surviving this period recover completely. The lethal dose for humans is 1–4 mg, and the Food and Drug Administration (FDA) limit is 80 µg per 100 g of shellfish tissue. A single raw mussel can contain up to 24 mg, and cooked mussel, 4 mg.

Diarrhetic Shellfish Poisoning is caused by the ingestion of shellfish containing thermostable toxins (at least seven have been identified, including okadoic acid, two dinophysitoxines and two pectenotoxins) and causes gastrointestinal disorders in many parts of the world. Dinoflagellates of the genera *Dinophysis* and *Aurocentrum* produce the toxins responsible for the illness within a time range of 30 min to a few hours; no fatalities have ever been observed, and patients recover in 3–4 days.

Neurotoxic Shellfish Poisoning is characterized by incoordination, paralysis, and convulsions caused by lipid-soluble toxins called 'brevetoxins'. *Gymnodinium breve* produces the toxins; the maximum permitted level (MPL) is 80 μg of brevetoxin per 100 g of flesh. No human deaths have been recorded.

Amnesic Shellfish Poisoning is caused by a water-soluble toxin (domoic acid) that produces nausea, headache, disorientation, and vomiting, but the most characteristic symptom is the permanent loss of short-term memory. The MPL is 20 p.p.m. domoic acid, and poisoning may cause human death.

Toxins from bacterial activity

Scombroid Fish (Histamine) Poisoning is a chemical intoxication following the ingestion of food containing high levels of histamine. Histamine is formed in some fishery products by enzymatic decarboxylation of histidine during the growth of certain spoilage bacteria that produce the enzyme histidine decarboxylase, such as certain Enterobacteriaceae, a few *Clostridia* and *Lactobacillus* spp. The possible activity of endogenous decarboxylase enzymes on free histidine leading to histamine formation in fish muscle is not well known. The most active histamine producers are *Morganella* (*Proteus*) *morganii*, *Klebsiella pneumoniae*, and *Hafnia alvei*; these microorganisms can be found in most fish, probably as a result of postharvest contamination. The fish species involved are mostly those with a natural high content of free histidine, such as *Scombrida* spp., but also *Clupeida* spp. In fact, the denomination 'scombroid fish poisoning' would be better changed to 'histamine poisoning.' Histamine concentrations of 10 mg per 100 g are considered significant; 20 mg per 100 g is the maximum allowable level in the EU, 50 mg per 100 g is a hazard in the USA Food and Drug Administration (FDA), and 100 mg per 100 g is generally considered toxic. The incubation period between ingestion and the onset of symptoms is rather short, from a few minutes to a few hours. The most common symptoms are: tingling and burning sensations in the mouth (peppery tasting), gastrointestinal complaints, and a rash with itching. The duration of the illness is also rather short, symptoms usually subsiding in a few hours.

Some evidence supports the hypothesis that certain substances in spoiled fish may potentiate the toxicity of histamine. Cadaverine and putrescine enhance the toxicity of histamine probably by inhibiting endogenous enzymes responsible for histamine detoxification.

See also: **Coronary Heart Disease**: Etiology and Risk Factor; **Dietary Reference Values**; **Dietary Requirements of Adults**; **Fats**: Digestion, Absorption, and Transport; **Fatty Acids**: Properties; Metabolism; Gamma-linolenic Acid; **Fish**: Fish as Food; **Fish Farming**; **Food Composition Tables**; **Minerals – Dietary Importance**; **Protein**: Food Sources; Requirements; **Shellfish**: Commercially Important Crustacea; **Vitamins**: Overview

Further Reading

Ackman RG (1995) Composition and nutritive value of fish and shellfish lipids. In: Ruiter A (ed.) *Fish and Fishery Products. Composition, Nutritive Properties and Stability.* Wallingford, UK: CAB International.

Exler J (1987, updated 1992) *Composition of Foods: Finfish and Shellfish Products, Human Nutrition Information Service Agriculture Handbook 8–15*. Washington, DC: US Department of Agriculture.

Food and Drug Administration (1989) *The Fish List, Food and Drug Administration (FDA) Guide to Acceptable Market Names for Food Fish Sold in Interstate Commerce 1988*. Washington, DC: US Government Printing Office.

Haard NF (1995) Composition and nutritive value of fish proteins and other nitrogen compounds. In: Ruiter A (ed.) *Fish and Fishery Products. Composition, Nutritive Properties and Stability.* Wallingford, UK: CAB International.

Holland B, Brown J and Buss DH (1993) Fish and fish products. In: *Supplement to the 5th edition of McCance and Widdowson's The Composition of Foods*. London: The Royal Society of Chemistry and Ministry of Agriculture, Fisheries and Food.

Huss HH (1994) *Assurance of Seafood Quality*, FAO Fisheries Technical Paper 334.

Huss HH (1995) *Quality and Quality Changes in Fresh Fish*. FAO Fisheries Technical Paper 348.

Lands WEM (1988) *Fish and Human Health*. Orlando, FL: Academic Press.

Lovell RT (1989) *Nutrition and Feeding of Fish*. New York: Van Nostrand Reinhold.

National Fisheries Institute at http://www.nfi.org.

Nettleton JA (1993) *Omega-3 Fatty Acids and Health*. New York: Chapman & Hall.

New York Seafood Council at http://www.nyseafood.org.

Southgate DAT (2000) *Meat, Fish, Eggs and Novel Protein*. In: Garrow JS, James WPT and Ralph A (eds) *Human Nutrition and Dietetics* (10th edn). Edinburgh, UK: Churchill Livingstone,

United States Department of Agriculture. Composition of foods, available at http://www.nal.usda.gov/fnic/foodcomp.

FISH FARMING

A G J Tacon, Kaneohe, HI, USA

Background

The term 'fish farming' is synonymous with the term 'fish culture' or, more correctly, 'aquaculture.' According to the Food and Agriculture Organization of the United Nations (FAO):

> the definition of aquaculture is understood to mean the farming of aquatic animals, including fish, mollusks, crustaceans and aquatic plants. Farming implies some form of intervention in the rearing process to enhance production, such as regular stocking, feeding, protection from predators, etc. Farming also implies individual or corporate ownership of the stock being cultivated. For statistical purposes, aquatic organisms which are harvested by an individual or corporate body which has owned them throughout their rearing period contribute to aquaculture while aquatic organisms which are exploitable by the public as a common property resource, with or without appropriate licenses, are the harvest of fisheries. Aquaculture production specifically refers to output from aquaculture activities, which are designated for final harvest for consumption or other purposes such as ornamental purposes.

Fish farming is not new, with the first known monograph on the subject appearing in China in the fifth century BC with the publication of the 'Treatise on Fish Culture' by Fan Li in 473 BC. Moreover, ancient Egyptian pyramid bas-reliefs show fishing scenes and fish raised in ponds, and archeological sites attest to the fact that the Romans also reared fish in ponds or 'piscifactorias.' However, with the exception of China and the Asia and Pacific region, where fish culture has always traditionally played an important role as a provider of much needed food fish for human consumption, it was not until the twentieth century that aquaculture emerged on to the global arena as a major food-production sector and economic activity capable of rivaling the terrestrial animal-livestock farming sector.

Farmed Species

In marked contrast to terrestrial animal livestock production systems, where farming is usually restricted to the production of a handful of different animal species (i.e., pigs, cattle, poultry, sheep) under a relatively narrow range of habitats and farming systems (i.e., grazing, mixed, and industrial farming systems), *aqua*culture is more akin to *agri*culture in the wide diversity of species cultivated and farming systems employed. The 207 different animal and plant species currently cultivated on a global basis are listed in the following sections.

Fishes or Finfish (Pisces) (123 Species Reported in 1998)

Freshwater fishes (73 reported species)

- carps, barbels, and other cyprinids (24 species, including: silver carp (*Hypophthalmichthys molitrix*), grass carp (*Ctenopharyngodon idella*), common carp (*Cyprinus carpio*), bighead carp (*Aristichthys nobilis*), crucian carp (*Carassius carassius*), roho labeo (*Labeo rohita*), catla (*Catla catla*), mrigal carp (*Cirrhinus mrigala*), white amur bream (*Parabramis pekinensis*), mud carp (*Cirrhinus mulitorella*), black carp (*Mylopharyngodon piceus*) Thai silver barb (*Puntius goniono-tus*), Java barb (*Puntius javanicus*), nilem carp (*Osteochilus hasselti*), golden shiner (*Notemigonus crysoleucas*), roach (*Rutilus rutilus*), goldfish (*Carassius auratus*), tench (*Tinca tinca*), hovens carp (*Leptobarbus hoeveni*), pond loach (*Misgurnus anguillicaudatus*), rudd (*Scardinius erythrophthalmus*), bleak (*Alburnus alburnus*), freshwater bream (*Abramis brama*), rhinofishes (*Labeo* spp));
- tilapias and other cichlids (12 species, including: Nile tilapia (*Oreochromis niloticus*), Mozambique tilapia (*Oreochromis mossambica*), blue tilapia (*Oreochromis aureus*), three-spotted tilapia (*O. andersonil*), redbreast tilapia (*Tilapia rendalli*), longfin tilapia (*Oreochromis macrochir*), tilapia (*Oreochromis spilurus*), redbelly tilapia (*Tilapia zillii*), jaguar guapote (*Cichlasoma managuensis*), blackbelt cichlid (*Cichlasoma maculicauda*), yellow-belly bream (*Serranochromis robustus*), velvety cichlids (*Astronotus* spp.);
- miscellaneous freshwater species (37 species, including channel catfish (*Ictalurus punctatus*), mandarin fish (*Siniperca chuatsi*), climbing perch (*Anabas testudineus*), catfish, hybrid (*Clarias gariepinus* × *Clarias macrocephali*), snakeskin gourami (*Trichogaster pectoralis*), piaractus/cachama blanca (*piaractus brachypomus*), pangas catfish (*Pangasius pangasius*), striped snakehead (*Channa striatus*), giant gourami (*Osphronemus goramy*), cachama (*Colossoma macropomum*), kissing gourami (*Helostoma temmincki*), North African catfish (*Clarias gariepinus*), Indonesian snakehead (*Channa micropeltes*), Northern pike (*Esox lucius*),

upside-down catfishes (*Synodontis* spp.), Wels catfish (*Silurus glanis*), black bullhead (*Ictalurus melas*), Asian redtail catfish (*Mystus nemurus*), Snakehead (*Channa argus*), Philippine catfish (*Clarias batrachus*), pike-perch (*Stizostedion lucioperca*), largemouth black bass (*Micropterus salmoides*), marble goby (*Oxyeleotris marmorata*), netted prochilod (*Prochilodus reticulatus*), mudfish (*Clarias anguillaris*), European perch (*Perca fluviatilis*), bagrid catfish (*Chrysichthys nigrodigitatus*), silver perch (*Bidyanus bidyanus*), Pacific fat sleeper (*Dormitator latifrons*), dorada (*Brycon moorei*), barred sorubim (*Pseudoplatystoma fasciatum*), duckbill catfish (*Sorubim lima*), South American catfish (*Rhamdia sapo*), knife fishes (*Notopterus* spp.), African bonytongue (*Heterotis niloticus*), Murray cod (*Maccullochella peelii*), golden perch (*Macquana ambigua*)).

Diadromous fishes (18 Reported Species)

- sturgeons, paddlefishes (two species, including Siberian sturgeon (*Acipenser beeri*), beluga (*Huso huso*));
- river eels (three species, including Japanese eel (*Anguilla japonica*), European eel (*Anguilla anguilla*), Short finned eel (*Anguilla australis*));
- salmons, trouts, smelts (11 species, including Atlantic salmon (*Salmo salar*), rainbow trout (*Oncorhynchus mykiss*), coho salmon (*Oncorhynchus kisutch*), chinook salmon (*Oncorhynchus tshawytscha*), ayu sweetfish (*Plecoglossus altivelis*), sea trout (*Salmo trutta*), Arctic char (*Salvelinus alpinus*), brook trout (*Salvelinus fontinalis*), European whitefish (*Coregonus lavaretus*), chum salmon (*Oncorhynchus keta*), grayling (*Thymallus thymallus*));
- miscellaneous diadromous fishes (two species, including milkfish (*Chanos chanos*), barramundi (giant sea perch; *Lates calcarifer*));

Marine fishes (32 reported species)

- flounders, halibuts, soles (five species, including bastard halibut (*Paralichthys olivaceus*), turbot (*Psetta maxima*), other flatfishes (Pleuronectiformes), common sole (*Solea vulgaris*), Atlantic halibut (*Hippoglossus hippoglossus*));
- cods, hakes, haddocks (one species, including Atlantic cod (*Gadus morhua*));
- redfishes, basses, congers (18 species, including Japanese seabream (*Pagrus major*), gilthead seabream (*Sparus auratus*), European seabass (*Dicentrarchus labrax*), puffers (Tetraodontidae; species not given), blackhead seabream (*Acanthopagrus schlegeli*), mangrove red snapper (*Lutjanus argentimaculatus*), greasy grouper (*Epinephelus tauvina*), sharpsnout seabream (*Diplodus puntazzo*), Russells snapper (*Lutjanus russelli*), goldlined seabream (*Rhabdosargus sarba*), areolate grouper (*Epinephelus areolatus*), goldsilk seabream (*Acanthopagrus berda*), red drum (*Sciaenops ocellatus*), goatfishes (*Upeneus* spp.), white seabream (*Diplodus sargus sargus*), rabbitfish (*Siganus* spp.), meager (*Argyrosomus regius*), cabinza grunt (*Isacia conceptionis*));
- jacks, mullets, sauries (seven species, including Japanese amberjack (*Seriola quinqueradiata*), flathead gray mullet (*Mugil cephalus*), Japanese jack mackerel (*Trachurus japonicus*), cobia (*Rachycentron canadum*), silversides (Atherinidae), fourfinger threadfin (*Eleutheronema tetradactylum*), snubnose pompano (*Trachinotus blochii*));
- tunas, bonitos, billfishes (one species, southern bluefin tuna (*Thunnus maccoyii*)).

Crustaceans (Crustacea) – 26 Species Reported in 1998

Freshwater crustaceans (nine reported species, including giant river prawn (*Macrobrachium rosenbergii*), Chinese river crab (*Eriocheir sinensis*), red swamp crawfish (*Procambarus clarkii*), yabby crayfish (*Cherax destructor*), red claw crayfish (*Cherax quadricarinatus*), marron crayfish (*Cherax tenuimanus*), Signal crayfish (*Pacifastacus leniusculus*), Danube crayfish (*Astacus leptodactylus*), sawtooth caridina (*Caridina denticulata*));

Sea-spiders, crabs: (three reported species, including Indo-Pacific swamp crab (*Scylla serrata*), blue crab (*Callinectes sapidus*), swimcrabs (*Portunus* spp.));

Lobsters, spiny-rock lobsters (two reported species, including longlegged spiny lobster (*Panulirus longipes*), mud spiny lobsters (*Panulirus polyphagus*));

Shrimps, prawns (12 reported species, including giant tiger prawn (*Penaeus monodon*), whiteleg shrimp (*Penaeus vannamei*), fleshy prawn (*Penaeus chinensis*), banana prawn (*Penaeus merguiensis*), blue shrimp (*Penaeus stylirostris*), Indian white prawn (*Penaeus indicus*), kuruma prawn (*Penaeus japonicus*), akiami paste shrimp (*Acetes japonicus*), other natantian decapods (Natantia), endeavor shrimp (*Metapenaeus endeavouri*), common prawn (*Palaemon serratus*), redtail prawn (*Penaeus penicillatus*)).

Molluscs (Mollusca) (43 Species Reported in 1998)

Freshwater molluscs (one reported species, Japanese corbicula (*Corbicula japonica*));

Abalones, winkles, conchs (three reported species, including perlemoen abalone (*Haliotis midae*), periwinkles (*Littorina* spp.), stromboid conch (*Strombus* spp));

Oysters (nine reported species, including Pacific cupped oyster (*Crassostrea gigas*), American cupped oyster (*Crassostrea virginica*), slipper cupped oyster (*Crassostrea iredalei*), European flat oyster (*Ostrea edulis*), Sydney cupped oyster (*Crassostrea commercialis*), mangrove cupped oyster (*Crassostrea rhizophorae*), shell-loving oyster (*Ostrea conchaphila*), Chilean flat oyster (*Ostrea chilensis*), hooded oyster (*Saccostrea cucullata*));

Mussels (10 reported species, including blue mussel (*Mytilus edulis*), Mediterranean mussel (*Mytilus galloprovincialis*), New Zealand mussel (*Perna canaliculus*), green mussel (*Mytilus viridis*), Korean mussel (*Mytilus coruscus*), Chilean mussel (*Mytilus chilensis*), South American rock mussel (*Perna perna*), Australian mussel (*Mytilus planulatus*), choro mussel (*Choromytilus chorus*), *cholga mussel* (*Aulocomya ater*));

Scallops, pectins (five reported species, including yesso scallop (*Pecten yessoensis*), Peruvian calico scallop (*Argopecten purpuratus*), great Atlantic scallop (*Pecten maximus*), queen scallop (*Chlamys opercularis*), Pacific calico scallop (*Argopecten ventricosus*));

Clams, cockles, arkshells (13 reported species, including Japanese carpet shell (*Ruditapes philippinarum*), razor clams (*Solen* spp.), blood cockle (*Anadara granosa*), northern quahog (*Mercenaria mercenaria*), inflated ark (*Scapharca broughtonil*), pullet carpet shell (*Tapes pullastra*), groved carpet shell (*Ruditapes decussatus*), common edible cockle (*Cardium edule*), butter clam (*Saxidomus giganteus*), donax clams (*Donax* spp), Pacific littleneck clam (*Prothaca staminea*), striped venus (*Venus gallina*));

Squids, cuttlefishes, octopuses (two reported species, including octopuses (*Octopus* spp.), common cuttlefish (*Sepia officinalis*)).

Miscellaneous Aquatic Animals (Four Species Reported in 1998)

Frogs and other Amphibians (One Reported Species, frogs (*Rana* spp.));

Turtles (one reported species, soft-shell turtle (*Trionyx sinensis*));

Sea squirts and other tunicates (one reported species, red bait (*Pyura stolonifera*));

Sea urchins and other echinoderms (one reported species, sea urchin (*Strongylocentrotus* spp.)).

Aquatic Plants (classification: Plantae) (11 species reported in 1998)

Brown seaweeds (two reported species, including Japanese kelp (*Laminaria japonica*), wakame (*Undaria pinnatifida*));

Red seaweeds (seven reported seaweeds, including laver (*Porphyra tenera*), eucheuma (*Euchema cottonii*), *gracilaria seaweeds* (*Gracilaria* spp.), elkhorn sea moss (*Kappaphycus alvarezii*), spiny eucheuma (*Euceuma denticulatum*), gellidium seaweeds (*Gellidium* spp.), harpoon seaweeds (*Asparagopsis* spp.));

Green seaweeds: (two reported seaweeds, including green laver (*Monostroma nitidum*), caulerpa seaweeds (*Caulera* spp.)).

Farming Practices and Feeding Methods

As one would expect with the cultivation of over 200 different animal and plant species, farming practices vary widely from region to region, country to country, species to species, and even for the same species. To a large extent, farming practices are determined by the ease of the culture of the target species, the aim of the proposed farming activity (subsistence or cash income), the economic unit value of the cultured species (higher-value species facilitating the use of higher-cost production methods), and the availability of resources to the farmer; (the last including the availability of water, land, seed and/or brood stock, fertilizers, feed, cash, and finance).

Farming practices can be undertaken under a variety of environmental conditions, including freshwater, brackish water, or marine waters. According to the FAO, for statistical aquaculture purposes, freshwater is meant to signify waters with a consistently negligible salinity, brackish water is meant to signify waters in which the salinity is appreciable but not to a constant high level (enclosed coastal and inland water bodies in which the salinity is greater than freshwater but less than marine water are also regarded as brackish), and marine is meant to signify coastal and offshore waters in which the salinity is maximal and not subject to significant daily and seasonal variation.

Current aquaculture farm production methods depend to a large extent upon the target species and the availability of resources to the farmer, and may range from the use of low-input (and consequently low-output) extensive farming systems (equivalent to 'grazing' in terrestrial livestock production) to the use of large-scale semiintensive and intensive industrial farming systems (equivalent to 'mixed' and 'industrial' livestock production systems). In general, higher-value cash crop species such as salmon and shrimp tend to be cultivated alone as a monoculture, whereas lower-value staple food fish species such as the Chinese carps tend to be cultivated together as a polyculture of different species with complementary rearing and feeding habits.

Farm production units can be broadly divided into outdoor and indoor production systems and, in the case of outdoor farming systems, may range from the use of earthen or concrete/plastic lined ponds, artificial tanks or raceways, floating or submerged net cages or fenced pen enclosures, poles or suspended ropes and net bags, rice fields or managed large water bodies (including reservoirs), to the use of indoor tanks and raceways/ponds within closed water recirculated aquaculture production systems.

Feeding methods employed by farmers range widely from species to species and country to country, depending upon the farm production unit employed and availability of fertilizer and feed resources. In the case of extensive pond-based farming systems where animals are usually reared at low stocking densities within large volumes of water, the growth of the cultured species is usually entirely dependent upon the consumption of live food organisms naturally produced within the water body. Such *in-situ* feeding practices are typical of existing farming systems for aquatic plant and filter-feeding mollusc and finfish species. By contrast, within semiintensive and intensive farming systems, fertilizers and feeds are usually externally provided either to augment the production of natural food organisms within the culture system (in the case of fertilizers) or as an exogenous direct source of nutrients in the case of externally provided feeds. Feeds may range from the use of fresh natural food items such as agricultural byproducts and animal food items (such as insects, worms, molluscs, fish, slaughterhouse wastes within small-scale, or backyard fish farming systems) to the use of farm-made or industrially manufactured nutritionally complete aquaculture feeds, the latter usually composed of a combination of different processed plant and animal-feed ingredient sources formulated so as to approximate to the known dietary nutrient requirements of the cultured target species. Such complete diet feeding methods are typical of semiintensive and intensive finfish and crustacean production systems.

Global Aquaculture Production

Aquaculture has been the world's fastest growing agricultural food production sector for over a decade, with the sector growing at an annual percentage rate (APR) of 11% since 1984, as compared with 3.1% for terrestrial farm animal meat production and 0.8% for landings from wild harvested fisheries over the same period. According to the latest statistical information from the FAO, total global aquaculture production in 1998 was reported as 39.4 million tonnes and valued at US$52.5 billion. At a species group level, finfish contributed over half of total aquaculture production by weight in 1998 (20 million tonnes or 50.8%), followed by mollusks (9.1 million tonnes or 23.2%) and aquatic plants (8.5 million tonnes or 21.7%). The growth of the different major specific groups over the past decade has been rapid, with most groups exhibiting double digit growth rates over the period 1984–1998, including finfish (12.3% per year), mollusks (11.5% per year), aquatic plants (7.7% per year), and crustaceans (16.0% per year).

By culture environment, over half (54.1%) of global aquaculture production originated from marine or brackish coastal waters in 1998 (down from 61.3% in 1984), as compared with 45.9% for freshwater aquaculture production (up from 38.7% in 1994), the average growth rate from 1984 to 1998 being highest for freshwater aquaculture production (APR 12.5%), followed by brackishwater (APR 10.2%) and mariculture (APR 9.9%). The main species groups reared in freshwater were finfish (97.9%), crustaceans and finfish in brackish water (56.8 and 35.1%, respectively), and mollusks and aquatic plants in marine waters (46.4 and 44.2%, respectively).

By region, Asia produced over 90.8% of global aquaculture production by weight in 1998 (35.81 million tonnes), with mainland China standing above all other countries with a reported total aquaculture production of 27.1 million tonnes or 68.6% of global aquaculture production in 1998. Including mainland China, all of the world's top 10 aquaculture producing nations were found in Asia in 1998, and included India 2.03 million tonnes Japan 1.29 million tonnes, Philippines 0.95 million tonnes, Indonesia 0.81 million tonnes, South Korea 0.80 million tonnes, Bangladesh 0.58 million tonnes Thailand 0.57 million tonnes, Vietnam 0.54 million tonnes, and North Korea 0.48 million tonnes (these top 10 producing countries representing 89.1% of global aquaculture production by weight; Table 1). The next major regions in terms of production by weight were Europe (4.97% or 1.96 million tonnes: Norway 0.41 million tonnes, Spain 0.31 million tonnes, France 0.27 million tonnes, Italy 0.25 million tonnes, UK 0.14 million tonnes, Netherlands 0.12 million tonnes), followed by South America (1.70% or 0.67 million tonnes: Chile 0.36 million tonnes, Ecuador 0.15 million tonnes, Brazil 0.095 million tonnes, Colombia 0.046 million tonnes), North America (1.66% or 0.65 million tonnes: USA 0.44 million tonnes, Canada 0.090 million tonnes, Mexico 0.041 million tonnes, Cuba 0.038 million tonnes), Africa (0.48% or 0.19 million tonnes: Egypt 0.14 million tonnes, Nigeria 0.020 million tonnes, Madagascar

Table 1 Top producers in 1998: total aquaculture production

Country	Production[a] (tonnes)	Production[b]	Growth APR[c] (1984–98)	Growth[d] (1997–98)	Total value (US$1000)	Unit value (US$kg^{-1})
China	27 071 942	68.6	+16.2	+12.7	25 449 016	0.94
India	2 029 619	73.8	+11.2	+9.0	2 222 789	1.09
Japan	1 290 486	77.1	+0.5	−3.7	4 126 039	3.20
Philippines	954 512	79.5	+5.5	−0.3	639 080	0.67
Indonesia	814 090	81.6	+7.2	+4.7	2 149 508	2.64
South Korea	796 632	83.6	+1.2	−23.4	766 268	0.96
Bangladesh	583 877	85.1	+12.7	+13.9	1 493 670	2.56
Thailand	569 577	86.5	+13.0	+3.1	1 806 795	3.17
Vietnam	537 870	87.9	+12.3	+5.7	1 356 724	2.52
North Korea	481 500	89.1	−2.9	−1.6	302 950	0.63
USA	445 123	90.2	+2.4	+1.5	781 069	1.75
Norway	408 862	91.3	+23.6	+11.3	1 133 580	2.77
Chile	361 430	92.2	+33.3	−3.6	1 001 544	2.77
Spain	313 518	93.0	+1.8	+31.1	282 208	0.90
France	273 920	93.7	+2.7	−4.7	614 156	2.24
Taiwan	255 205	94.3	+0.3	−5.5	818 395	3.21
Italy	249 625	94.9	+7.5	+15.2	471 139	1.89
Ecuador	146 590	95.3	+12.0	+8.3	731 299	4.99
Egypt	139 389	95.7	+13.4	+89.8	327 263	2.35
UK	137 421	96.0	+17.1	+5.9	427 895	3.11
Total world	39 430 834	100.0	+11.0	+9.4	52 458 185	1.33

[a]Total aquaculture production (includes finfish, crustaceans, molluscs, miscellaneous aquatic animals/products, and aquatic plants).
[b]Accumulative total as a percentage of the total world aquaculture production.
[c]Annual percentage growth rate in production by weight between 1984 and 1998.
[d]Percentage change in production by weight between 1997 and 1998.

0.0069 million tonnes, South Africa 0.0052, Zambia 0.0042 million tonnes), and Oceania (0.36% or 0.14 million tonnes: New Zealand 0.094 million tonnes, Australia 0.028 million tonnes).

By economic country grouping, approximately 90.0 and 82.2% of global aquaculture production in 1998 was produced within developing countries (35.49 million tonnes). Moreover, whereas the developing country share of aquaculture production has increased from 72.6% (7.37 million tonnes) of total aquaculture production in 1984 to 90% (35.49 million tonnes) in 1998, the share of production from developed countries has decreased from 27.4% (2.78 million tonnes) in 1984 to 10% (3.93 million tonnes) in 1998. Aquaculture production within developing countries has grown four times faster (12.8% per year since 1984) than within developed countries (2.7% per year since 1984).

In general terms, aquaculture's contribution toward total world fishery landings has increased threefold since 1984, aquaculture production increasing from 10.15 million tonnes or 11.4% of total fishery landings in 1984 to 39.43 million tonnes or 31.1% of total fishery landings in 1998. By continent, aquaculture supplied 45.3% of total fisheries landings in Asia (up from 21.1% in 1984), 10.9% of total landings in Oceania (up from 3.7% in 1984), 10.2% in Europe (up from 6.9% in 1984), 8.0% in North America (up from 4.5% in 1984), 5.7% in South America (up from 0.5% in 1984), and 3.2% in Africa (up from 0.9% in 1984).

Major Farmed Finfish and Crustacean Species

Inland freshwater species currently form the bulk of finfish aquaculture production (17.4 million tonnes or 86.6% of total finfish production), with diadromous and marine finfish species constituting only 9.5 % (1.9 million tonnes) and 3.9% (0.78 million tonnes) of total finfish production (20.06 million tonnes) in 1998. On a species group level, the major cultivated finfish species groups in 1998 can be listed as follows:

- cyprinids (total production 14.14 million tonnes, 23 freshwater species reported in 1998; growth rate over period 1984–98: 13.3% per year);
- tilapia (total production 0.97 million tonnes, 12 freshwater species reported in 1998; growth rate over period 1984–98: 13.4% per year);
- catfish (total production 0.45 million tonnes, 12 freshwater species reported in 1998; growth rate over period 1984–98: 10.2% per year);
- salmonids (total production 1.29 million tonnes, 11 diadromous species reported in 1998; growth rate over period 1984–98: 13.6% per year);

- eels (total production 0.22 million tonnes, three diadromous species reported in 1998; growth rate over period 1984–98: 8.3% per year);
- milkfish (total production 0.37 million tonnes, one diadromous species reported in 1998; growth rate over period 1984–98: 0.3% per year);
- jacks/mullets (total production 0.20 million tonnes, seven marine species reported in 1998; growth rate over period 1984–98: 1.5% per year); and
- redfishes/basses (total production 0.20 million tonnes, 16 marine species reported in 1998; growth rate over period 1984–98: 16.1% per year).

In addition to the above finfish species group, the major crustacean groups produced in 1998 were:

- marine shrimp (total production 1.11 million tonnes, 71.2% of total crustacean production, 11 species reported in 1998, growth rate over period 1984–98: 15.4% per year); and
- freshwater crustaceans (total production 0.32 million tonnes, 20.4% of total crustacean production, nine species reported in 1998; growth rate over period 1984–98: 15.7% per year).

Table 2 summarizes the top twenty cultivated finfish and crustacean species in 1998.

Global Production by Feeding Habit and Nutrient Supply

In terms of feeding habit and nutrient supply, global aquaculture production in 1998 may be broadly divided into six basic categories, reflecting primary trophic behavior utilized by farmers, as follows:

- photosynthetic aquatic plants: 8.57 million tonnes or 21.7% of total aquaculture production, examples including all brown, red, and green seaweeds;
- filter-feeding mollusks: 9.14 million tonnes or 23.2% of total aquaculture production, major cultivated species including oysters, mussels, scallops, clams, and cockles;

Table 2 Top 25 finfish and crustacean aquaculture species in 1998

Species	*Production (tonnes)*	*Production (percentage of world total)*	*Growth (APR, 1984–98)*	*Growth increase (%),1997–98*	*Major producing countries by weight (percentage of total species)*
Silver carp[a]	3 308 419	15.3	11.1	2.5	China 94.7, Bangladesh 2.9
Grass carp[a]	2 894 017	28.7	19.1	6.7	China 97.0
Common carp[a]	2 465 283	40.1	11.3	10.6	China 78.2, Indonesia 7.5, Egypt 2.0
Bighead carp[a]	1 584 289	47.4	11.6	2.0	China 98.9
Osteichthyes[a,b]	1 556 915	54.6	10.7	8.7	China 62.3, India 5.5, Bangladesh 4.3
Crucian carp[a]	1 036 164	59.4	24.6	20.1	China 99.6
Nile Tilapia[a]	793 931	63.1	21.0	7.1	China 66.2, Thailand 12.9, Philippines 7.6
Roho labeo[a]	754 677	66.7	14.1	8.9	India 75.5, Bangladesh 12.6, Myanmar 11.3
Atlantic salmon[c,d]	687 906	69.8	28.1	6.4	Norway 52.4, UK 16.1, Chile 15.6, Canada 7.2
Catla[a]	628 757	72.7	12.9	8.8	India 87.5, Bangladesh 12.2
Giant tiger prawn[e]	577 990	75.4	18.5	8.8	Thailand 41.5, Indonesia 17.9, Vietnam 15.1
Mrigal carp[a]	560 556	78.0	15.3	8.7	India 92.8, Bangladesh 6.8
White amur bream[a]	449 282	80.0	13.1	3.3	China 100
Rainbow trout[c,d]	438 013	82.1	6.8	2.5	Chile 17.1, Italy 10.9, Norway 10.7, France 10.2
Milkfish[c]	369 003	83.8	0.3	0.4	Philippines 44.0, Indonesia 40.1, Taiwan 15.8
Osteichthyes[d,f]	335 620	85.3	28.8	16.3	China 91.4, South Korea 4.4, Japan 2.1
Channel catfish[a]	256 129	86.5	6.8	7.5	USA 99.9
Japanese eel[c,d]	206 773	87.5	8.1	− 4.4	China 78.9, Japan 10.6, Taiwan 8.3
White leg shrimp[e]	191 009	88.4	14.4	13.0	Ecuador 67.8, Mexico 12.4, Panama 4.7, Belize 3.9
Mud carp[a]	160 092	89.1	12.1	6.7	China 99.9
Black carp[a,d]	153 633	89.8	11.7	10.8	China 99.4
Japanese amberjack[d,f]	147 115	90.5	− 0.4	6.2	Japan 99.8
Fleshy prawn[e]	143 932	91.2	6.7	37.8	China 99.4
Giant river prawn[e]	130 313	91.8	21.9	19.5	China 47.5, Bangladesh 36.8, Taiwan 6.2
Chinese river crab[e]	123 249	92.3		22.4	China 99.9

[a]Freshwater finfish.
[b]Miscellaneous freshwater fishes (species not given).
[c]Diadromous finfish.
[d]Carnivorous finfish.
[e]Crustaceans.
[f]Miscellaneous marine fishes (species not given).

- filter-feeding finfishes: 5.52 million tonnes or 14.0% of total aquaculture production, major cultivated species including silver carp, bighead carp, and catla;
- herbivorous/omnivorous finfishes: 12.0 million tonnes or 30.4% of total aquaculture production; major cultivated species including grass carp, common carp, crucian carp, Nile tilapia, rohu, mrigal carp, white amur bream, milkfish, channel catfish, and mud carp;
- benthophagic omnivorous scavenging crustaceans: 1.56 million tonnes or 3.9% of total aquaculture production; major cultivated species including giant tiger prawn, whiteleg shrimp, fleshy prawn, banana prawn, giant river prawn, Chinese river crab, marine crabs, and crayfish; and
- carnivorous finfishes: 2.53 million tonnes or 6.4% of total aquaculture production; major cultivated species including Atlantic salmon, rainbow trout, Japanese eel, Japanese amberjack (yellowtail), black carp, Japanese seabream, coho salmon, Mandarin fish, gilthead seabream, and European seabass.

In terms of nutrient supply, it can be seen from the above trophic classification that nearly half (i.e., 17.71 million tonnes or 44.9%) of global aquaculture production in 1998 was based on the production of marine plants and mollusks within extensively managed farming systems, receiving little or no supplementary nutrient inputs (i.e., in the form of inorganic and/or organic fertilizers). Similarly, over a quarter of total finfish production (i.e., 5.52 million tonnes or 27.5% total finfish production) was based on the production of filter-feeding finfish species, these species usually being reared within extensive/semi-intensively managed farming systems (i.e., within earthen ponds, pen enclosures, rice fields, or small water bodies) as a polyculture of mixed finfish species at low (extensive) to moderate (semiintensive) stocking densities, with finfish growth being dependent upon the filtration of live/suspended food organisms/nutrients from the water body (the production of the latter being augmented or not through the application of supplementary nutrient inputs in the form of fertilizers).

In marked contrast, the production of the remaining finfish (14.53 million tonnes or 72.5% total finfish production) and crustacean species (1.56 million tonnes) is almost entirely based on the provision of exogenously supplied nutrient inputs in the form of either industrially compounded nutritionally complete aquafeeds (or, to a lesser extent, natural food items of high nutrient value such as 'trash fish') or supplementary farm-made aquafeeds (often in combination with fertilization). Interestingly, carnivorous finfish species (2.53 million tonnes) represented 12.6% of total farmed finfish production in 1998, including 72.9% of total finfish production in developed countries (1.35 million tonnes from a total of 1.85 million tonnes) and 6.45% of total finfish production in developing countries (1.17 million tonnes from a total of 18.19 million tonnes).

Food Supply and Concluding Remarks

In terms of per caput availability of 'food fish' from aquaculture (i.e., the production of farmed aquatic finfish and shellfish on a whole live-weight basis, and excluding farmed aquatic plants; 30.86 million tonnes in 1998) this has increased by 261% from 1.45 kg in 1984 to 5.23 kg in 1998, with supply growing at an average rate of 10.4% per year. By contrast, the per caput availability of 'food fish' from capture fisheries (i.e., 62.45 million tonnes, excludes captured fish destined for reduction into fishmeal) has remained static, decreasing from 10.88 kg in 1984 to 10.58 kg in 1998. On the basis of the above data, over 33.1% of global 'food fish' supplies was supplied by aquaculture in 1998. In terms of total farmed meat production, aquaculture currently ranks fourth in terms of global farmed meat production (19.5 million tonnes in 1998; after gutting/shelling), with pig first (88.0 million tonnes), beef and veal second (55.3 million tonnes), and chicken third (52.1 million tonnes). In general, the main factor driving the apparent high demand for staple food fish, and in particular of low-value farmed freshwater food fish species feeding low on the aquatic food chain, within most developing countries is their greater affordability to the poorer segments of the community, including the rural poor, compared with other animal-protein sources.

With the global population now exceeding 6 billion, there is a growing trend toward increased competition between users or stakeholders, including both the agricultural and nonagricultural sectors, for available resources, including land, energy/fossil fuels, water, and nutrient sources. As a consequence of this, there is also a trend within animal husbandry (including aquaculture) toward the intensification of farming systems and increased livestock/aquaculture production per unit area/time, as well as an increasing trend toward the vertical integration of animal and feed production systems so as to reduce risks and minimize production costs. It follows, therefore, that if aquaculture is to sustain its current growth into the third millennium farming systems should become increasingly more efficient in terms of resource use and have little or no adverse impacts on society and the environment.

See also: **Marine Foods**: Production and Uses of Marine Algae; **Shellfish**: Commercially Important Crustacea; Commercially Important Molluscs

Further Reading

Bardach JE, Ryther JH and McLarney WO (1972) *Aquaculture: The Farming and Husbandry of Freshwater and Marine Organisms*. New York: Wiley-Interscience.

FAO (2000) *Yearbook of Fishery Statistics 1998. Vol. 86/2. Aquaculture Production. FAO Statistics Series No. 154 and Fisheries Series No. 56*. Rome: Food and Agriculture Organization of the United Nations.

Li S and Mathias J (1994) *Freshwater Fish Culture in China: Principles and Practice*. Amsterdam: Elsevier.

Tacon AGJ and Forster IP (2000) Global trends and challenges to aquaculture and aquafeed development in the new millennium. In: *International Aquafeed – Directory and Buyers' Guide 2001*, pp. 4–25. Uxbridge, UK: Turret RAI.

FISH MEAL

S M Barlow, International Association of Fishmeal and Oil Manufacturers, St Albans, Hertfordshire, UK

Introduction

Fish meal is a brown powder which normally contains a high level of protein and appreciable quantities of lipid and minerals. Of world fish meal production, about 90% is produced from oily species of fish, such as sardine, anchovy, capelin, and menhaden, and less than 10% from white fish offal (frames), such as cod and haddock. Of the annual world catch of fish, about one-third is used as raw material for fish meal production, this being sustainable catches for which there is no direct outlet for human consumption.

Nature of the Product and its Use

The nutritional values of fish meals differ, mainly depending on the species of fish from which the meals are prepared. In practice, users have classified fish meals into four broad groups, namely herring-type, anchovy/pilchard, menhaden, and white-fish. The main nutritional characteristics of these four broad groups of meals are summarized in **Table 1**.

The protein in fish meal has a high biological value in diets for animals. It is rich in essential amino acids, particularly lysine and the sulfur amino acids, and the presence of fish meal in a complete diet will supplement any deficiencies of the amino acids in vegetable proteins. The amino acid make-up of fish meal made from whole fish is similar when expressed on a protein basis. The lipid material provides a concentrated energy source and it can contribute towards essential fatty acid requirements. Fish meal is a rich source of the major elements, calcium, phosphorus, sodium chloride and magnesium, the trace elements, iron, zinc, and selenium, and the B-group vitamins, especially B_{12} and choline. Unlike vegetable sources of protein, the phosphorus in fish meal is fully available. (*See* **Essential Fatty Acids**, and refer to individual nutrients.)

Although particular proteins might on analysis be found to contain certain levels of amino acids, biological experiments have indicated that these amino acids might not be completely available to the animals ingesting the protein. The digestibilities of the nitrogen and of the critical amino acids of fish meal are approximately 89% for pigs, 85% for chicken, and 94% for the rat. Fish meal is of low degradability in the rumen, while the undegraded or bypass protein is of high digestibility in the small intestine. Meals with a low solubles content may be marketed as specially selected for ruminants. The degradability in the rumen of the protein of such meals is in the range 0.3–0.4.

The main animal species consuming fish meal in 1988 were poultry (60% of world fish meal production), pigs (20%), fish (10%), fur-bearing animals (2.5%), and ruminants (2.5%). The practical uses of fish meal in diets for these species are summarized below.

Poultry

The principal consumers are broilers, followed by turkeys, breeding birds and, to a lesser extent, laying hens. The recommended minimum level of fish meal in a broiler diet is 2%, with the maximum in a broiler finisher diet of 6–8%. The maximum level is normally dictated by the desire not to impart a fishy taint to the flavor of the broiler meat. High-fat fish

Table 1 Principal nutritional values for fish meals (all data on 'as received' basis)

	White-fish meal	*European herring-type fish meals*	*South American anchovy-type fish meals*	*Menhaden fish meal*
Proximate analysis (%)				
Moisture	10	9	9	8
Crude protein	65	70	65	61
Crude fat	5	9	9	10
Crude ash	20	10	16	19
Crude fiber	0	0	0	0
Energy content (MJ kg^{-1})				
Poultry				
ME	12.1	14.0	13.2	12.8
TME	13.8	15.7	14.5	14.1
Pigs				
DE	15.6	18.1	16.9	16.2
ME	13.7	15.9	14.9	14.2
Ruminants				
ME	13.2	16.2	13.2	12.8
TDN	89.5	105.5	91.8	92.7
NE_1	7.6	9.8	7.8	7.6
Amino acids (%)				
Lysine	4.49	5.47	5.07	4.53
Methionine	1.69	2.16	1.95	1.64
Methionine plus cystine	2.29	2.88	2.60	2.25
Tryptophan	0.61	0.83	0.78	0.49
Minerals				
Ash (%)	20.0	10.14	15.4	18.0
Calcium (%)	8.0	1.95	3.95	5.26
Phosphorus (%)	4.8	1.50	2.60	2.98
Sodium (%)	0.77	0.42	0.87	0.34
Magnesium (%)	0.15	0.11	0.25	0.14
Potassium (%)	0.9	1.20	0.65	0.72
Iron (p.p.m.)	300	150	246	438
Zinc (p.p.m.)	100	120	111	151
Selenium (p.p.m.)	1.50	2.78	1.39	2.22
Vitamins (p.p.m.)				
Panthothenic acid	15	30.6	9.3	8.8
Riboflavin	6.5	7.3	2.5	4.8
Nicotinic acid	50	126	95	55
Folic acid	0.5	0.5	0.16	NA
Choline	4396	4396	4396	4396
Vitamin B_{12}	0.07	0.25	0.18	0.06
Biotin	0.08	0.42	0.26	0.26

ME, metabolizable energy; TME, true metabolizable energy; DE, digestible energy; TDN, total digestible nutrients; NE_1, net energy lactation; NA, not available.

meals are more likely to do this than low-fat fish meals (e.g., white-fish meals).

The minimum levels are based on practical experience and experimental results indicating that this level of fish meal produces increases in meat or egg production and better feed utilization compared with diets not containing fish meal. It is impossible to state a typical level of fish meal incorporation in a broiler diet. This depends upon the alternative feed ingredients available and the consumer preference for different characteristics of the broiler flesh. The differences are largely geographically based.

Turkey starter/grower diets would normally have a minimum fish meal level of about 5%, with a maximum of about 8%, even of high-fat fish meals. Layer diets would have, in practice, no minimum or maximum constraints.

Recent interest in the health benefits from the long-chain ω-3 polyunsaturated fatty acids is increasing interest in their residual levels in meat resulting from fish meal feeding. (*See* **Poultry**: Chicken; Ducks and Geese; Turkey.)

Pigs

The minimum level in weaner diets for pigs typically weaned at 3–4 weeks of age through to 20 kg liveweight is about 5% of fish meal, with this level dropping to 2% in grower diets (20–50 kg). Breeding and

lactating pigs also have a minimum fish meal level of about 4% in most diets. Finisher pig diets (> 50 kg) probably have no minimum constraint. However, in circumstances of large availability of supplies of fish meal at relatively low prices compared with vegetable proteins, such as in the main fish meal-producing countries of South America, when it is economically advantageous to include fish meal, a constraint is normally placed upon the maximum level of fish meal to be used in order to ensure that no fishy taint is imparted to the pig flesh. For a high-fat fish meal containing > 10% of fat is is not unusual to see a maximum constraint of 5% in finishing diets. (*See* **Pork**.)

Fish

The principal cultivated fish species using fish meal in their diet are prawns, salmon, eels, trout, and yellow tails, i.e., mainly the marine carnivorous species. These do not use vegetable proteins efficiently, especially if they are fast-growing. Fish meal content of salmon diets is believed to average 45%, with slightly higher inclusion in diets for Atlantic salmon than for Pacific salmon. In prawns a typical fish meal content in the diet would be about 30%; the level of incorporation for trout would be 35% and for eels about 50%. (*See* **Fish Farming**.)

Fur-Bearing Animals

Fish meal is commonly used in mink feeding. The amount of fish meal used varies, but up to 60% of the protein can originate from fish meal provided that the meal is of high quality. Antioxidant-stabilized fish meal, or low-fat fish meal prepared from fresh fish, is preferred.

Ruminants

Fish meal fed at relatively low levels (0.75 kg day^{-1}) to high-yielding dairy cows (over 251 per day) has been shown to increase the milk yield significantly. The effect has been particularly pronounced with high-forage diets, especially with grass silage as the main forage source. Probably of greater economic significance is the fact that dietary fish meal increases the reproductive performance of the dairy cow. For example, conception rates have been shown to increase from 44% to 64%. This increases calving frequency and milk production.

Fish meal also has a role in the diets of sheep, improving productivity and product quality and reducing production costs. It has been found to be particularly important during the last 2–3 weeks of pregnancy for ewes given mainly roughage diets, during the first 5–6 weeks of lactation when body fat in the ewes can help to sustain a large milk production, as a feed supplement in high-energy diets for fattening of early weaned lambs, and for older lambs on high-roughage diets. Finally, it has been found to be useful in manipulating body fat stores, thus reducing excess body fat prior to going to market.

In diets for beef cattle, especially those on high-forage diets, fish meal (200–400 g day^{-1}) increases growth rate. It has been recently shown to provide a means of increasing lean growth at the expense of body fat, which decreases in over-fat finishing cattle on only poor roughage plus fish meal. (*See* **Beef**.)

Source of Raw Material

The main producing and exporting countries of fish meal are shown in Figure 1. South America, Japan, the former Soviet Union and the Scandinavian countries are the main producers. These same countries, with the exception of Japan and the former Soviet Union, are the major exporters.

The principal species of fish processed into meal and oil in these countries are listed in Table 2, together with the name commonly used (sometimes erroneously) in the fish meal trade to describe these types of fish meal.

One of the distinguishing features of the fish meal industry, compared with other sectors of the fishing industry, is the ability to process large volumes of fish in a short time. A company in South America could be processing 250 000 t of fish in a year and, because of the seasonal nature of the fishing, the storage capacity of the plant would be greater than 1000 t of fish.

Because ambient temperatures in the north of Chile and in Peru are relatively high, it is necessary to have the capacity to process the stored fish in 24 h or less. Fish stored at 30 °C will soon turn into a hydrolyzed liquid, resulting in considerable losses of raw material and reducing the yield of the meal. In addition, the quality of the meal produced by such stale fish is poorer than that produced from fresh fish, resulting in poorer performance from livestock fed on the resultant meal. Most manufacturers now monitor the freshness of raw material by measuring its volatile nitrogen content. The lower this is, the fresher the raw material. Special-product fish meals for farmed fish, early weaned pigs, and ruminants are made from fish with less than 90 mg of volatile nitrogen per 100 g of fish.

In colder climates, such as Iceland, it is possible to store the raw fish for much longer periods than in tropical or semitropical climates.

Over the years, a number of chemical agents have been added to the raw fish in order to extend its shelf-life. However, this practice is currently not

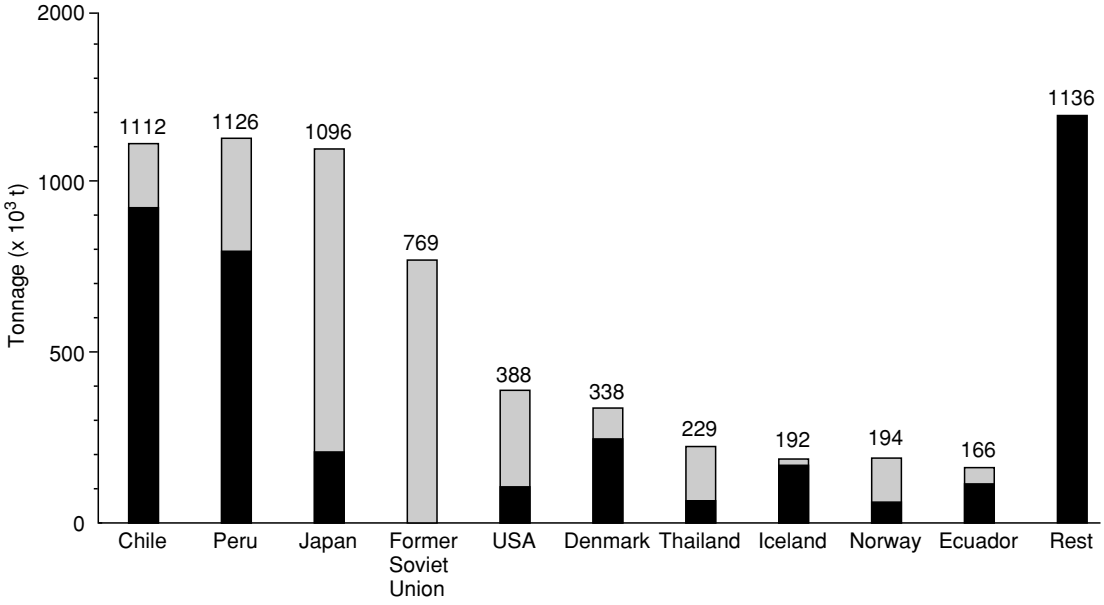

Figure 1 Production (open columns) and export (filled columns) of fish meal, 1988. Reproduced from Fish Meal, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 2 Main species of fish processed into meal

Country	*Principal fish species*	*Common designation used in fish meal trade*
Chile	Anchovy (*Engraulis ringens*)	Anchovy
	Sardine (*Sardinops sagax*)	
	Horse mackerel (*Trachurus murphyi*)	
Peru	Anchovy (*Engraulis ringens*)	Anchovy
	Sardine (*Sardinops sagax*)	
Japan	Sardine (*Sardinops*)	Anchovy
	White-fish offal	White-fish
Denmark	Sand eel (*Ammonodytes* spp.)	Herring
	Norway pout (*Gadus esmarki*)	
	Herring (*Clupea harengus*)	
Norway	Capelin (*Mallotus villosus*)	Herring
	Herring (*Clupea harengus*)	
Iceland	Capelin (*Mallotus villosus*)	Herring
	Herring (*Clupea harengus*)	
	White-fish offal	White-fish
USA	Menhaden (*Brevoortia tyrannus*)	Menhaden

very common and physical means of preservation, such as icing or holding in refrigerated sea water, are more in favor.

Manufacture of Meal

Almost all fish meal is made by cooking, pressing, drying, and grinding fish in machinery designed for that purpose (Figure 2). On some occasions the pressing stage may be omitted since with white-fish raw material there is little oil to be removed.

When fish are cooked, the protein is coagulated, much of the water and oil separates and can be removed by pressing; raw fish, on the other hand, even when subject to very high mechanical pressure, lose very little liquor. A cooker consists of a long cylindrical plant, the outside of which has a steam jacket, and through which the fish are moved by a screw conveyor. Some cookers also inject steam into the cooked material. No drying occurs during cooking. Cooking has an important sterilizing action and assists in insuring freedom from undersirable microorganisms.

After cooking, the fish is pressed in order to remove some of the water and oil. A mixture of oil, water, and some solids is squeezed through the perforations of

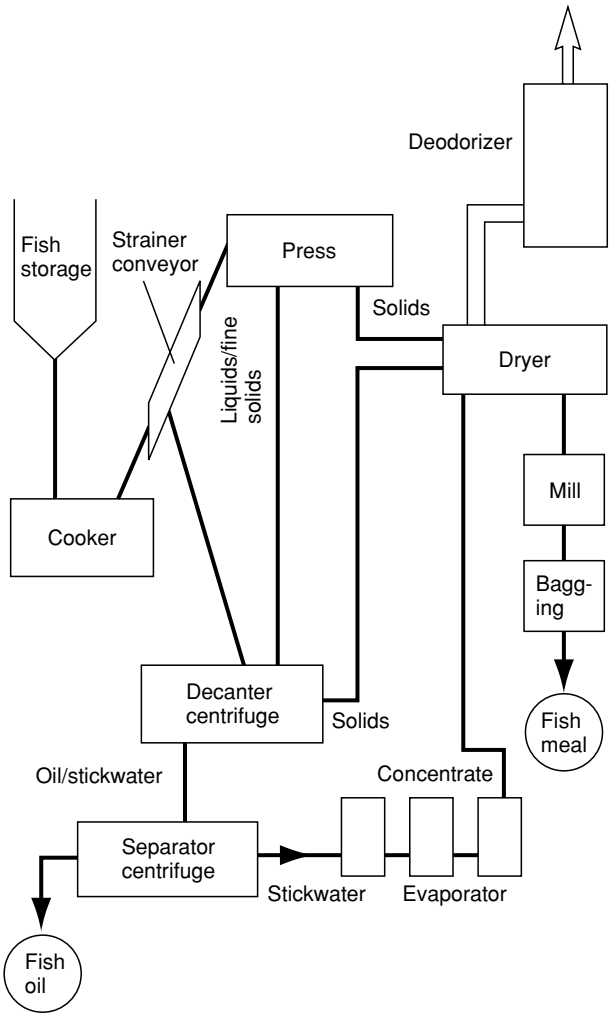

Figure 2 Fish meal processing, shown in diagrammatic form. Reproduced from Fish Meal, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

the press and later these three phases are separated; the solids and the concentrated aqueous phase (stickwater concentrate) are returned to the presscake from the end of the press and the separated oil is polished and stored as a separate product.

The presscake, together with the recovered solids and the stickwater concentrate, is then introduced into a drier to reduce the moisture level down to 10% to make the product stable.

There are two main types of drier – direct and indirect. In the direct drier, very hot air at a temperature of up to 500 °C is passed over the material as it tumbles rapidly in a cylindrical drum. The indirect type of drier consists either of a steam-jacketed cylinder or a cylinder containing steam-heated disks which also tumble the meal. Some recently developed driers operating at lower temperatures use indirectly heated hot air, or vacuum drying. The fish meal produced is sometimes referred to as a low-temperature product, particularly suitable for fish, early-weaned pigs, and mink.

Storage of the Finished Meal

Most fish meal, except white-fish meal, contains a residue of 8–12% of lipid which could not be pressed out of the product economically. This lipid reacts with oxygen and produces heat, which if not controlled properly can result in scorching or even ignition of the fish meal during storage. Thus most of the world's fish meal contains added antioxidant. This is normally ethoxyquin added at a level of about 750–1000 p.p.m. (*See* **Antioxidants**: Synthetic Antioxidants.)

The meal is then cooled through a rotary cooling drum, or, more usually, stored directly in 50-kg bags or in heaps of several hundred tonnes on the ground or in warehouses. The bags are stacked in

such a way as to allow a circulation of air in the vicinity of each bag, so that any heat produced during storage as a result of residual oxidation of the meal is easily removed to the atmosphere. Bulk heaps of meal are turned by means of payloaders from time to time. This again allows any heat produced in the meal to escape. Most large storage areas of meal are equipped with thermocouples which constantly measure the temperature of the meal to insure that there is no risk of overheating.

Provided that the meal has been treated with antioxidant and maintained in a dry condition, and not allowed to overheat, it is possible to store the product for longer than 12 months without any change in nutritional quality. If moisture levels are allowed to rise above 15%, undesirable bacteria and mold growth may occur; this is true for most types of feedstuffs and organic material.

See also: **Antioxidants**: Synthetic Antioxidants; **Beef**; **Essential Fatty Acids**; **Fish Farming**; **Pork**; **Poultry**: Chicken; Ducks and Geese; Turkey

Further Reading

Barlow SM and Windsor ML (1983) Fishery by-products. *CRC Handbook of Nutritional Supplements* VII: 253–272.

Bowman SJ (1990) *Digest of Selected Statistics*. Potters Bar, UK: International Association of Fish Meal Manufacturers.

FAO, Fishery Industries Division (1986) *The Production of Fish Meal and Oil*. FAO Fish. Technical Paper 142 (rev. 1). Rome: Food and Agriculture Organization.

Pike IH (1984) Recent developments in feeding fish meal to farm animals. *The Feed Compounder* September: 10–13.

Pike IH, Andorsdottir G and Mundheim H (1990) *The Role of Fish Meal in Diets for Salmonids*. IAFMM Technical Bulletin no. 24. Potters Bar, UK: International Association of Fish Meal Manufacturers.

Windsor M and Barlow S (eds) (1981) *Introduction to Fishery By-Products*. Farnham, UK: Fishing News Books.

FISH OILS

Contents

Production

G M Pigott and B W Tucker, Sea Resources Engineering, Inc., Kirkland, WA, USA

Production of Fish Oils

The importance of fish oils has increased greatly during recent years as the nutritive value of certain components is being emphasized. Prior to this time, unrefined foul-smelling oxidized fish oils known as 'cod liver oil' were spooned as a vitamin supplement for vitamins A and D. Later, partially refined and hydrogenated fish oil was used for the manufacture of margarine. However, most of the oil was, and is, used for industrial purposes such as components in paints, varnishes, and lubricants, where the quality for human acceptance as a food receives little consideration.

Highly refined fish oils are now in demand for their ω-3 fatty acids, which have been shown to have a profound effect on the cardiovascular and neurological systems. During the last few years, beginning with menhaden oil, fish oil is being approved for human consumption as food supplements and as an ingredient in foods. This creates a demand for better quality control through improved handling of oil in conventional meal plants and in other commercial processes in which fish oil is a by-product. Great improvements have been, and are being, made in maintaining the condition of raw materials and primary products so that fish oil does not have any major oxidation problems prior to extracting and/or refining oil for human and animal consumption.

The term 'fish oil' or 'marine oil' is normally used for the class of lipids that originates from the marine and freshwater bodies of the world, as opposed to those from land plants and animals. Any plant or animal raised in these aquatic environments has the high profile of ω-3 fatty acids not found in most land plants and animals. Hence, by practice, since the large majority of oils from this aquatic environment come from fish, marine oils, fish oils, and aquatic plant and animal oils are used interchangeably throughout the world and are commonly called fish oils.

Fish to be used as a source of oil must be cooled and stored at refrigerator temperatures to minimize enzymatic breakdown of certain components. Unfortunately, this is often not possible in tropical and subtropical areas, and the resulting quality of both the fishmeal and oil is significantly reduced. Not only does spoilage and oxidation occur, but also such degradation results in the formation of compounds, such as sulfur, which are catalyst poisons.

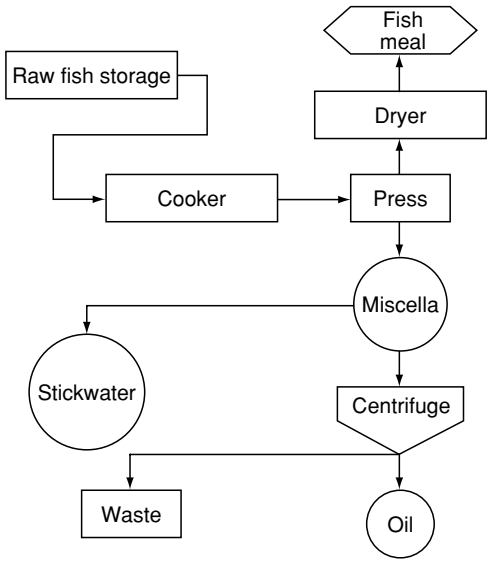

Figure 1 Conventional fish meal process.

Extraction of Fish Oils

Although both extraction and refining technology is advancing rapidly, most of the fish oil being produced in the world is as a byproduct of the conventional fish meal process, a wet rendering technique. Other processes include rendering, hydrolysis, silage production, solvent extraction, critical extraction, and ion exchange. Other new techniques are rapidly being developed to ensure a high-quality product for human and animal food. The superiority of high ω-3 fish oils is creating a demand for these oils to become a common component in engineered and formulated foods consumed by humans.

Conventional Fish Meal Process

Other than improving the machinery and equipment made possible by technical developments, the conventional fish meal process has not had any major procedural changes for the past 50 years. The process, as depicted in Figure 1, involves cooking, pressing to separate the liquid from the solids, and drying the solids. The liquid phase contains water, oil, water solubles, and suspended solids. Oil is removed from the mixture by centrifuging, after which it is pumped into large storage tanks to await bulk shipment to market. The large fish meal operations are designed for processing several hundred tonnes of industrial fish per day. Some of these plants can handle 50–60 tonnes of fish per hour and are located in areas of the world where large schools of fish such as menhaden, herring, sardines, and anchovy are located.

Until recently, a large portion of the world's fish meal products were produced from the large resource of industrial fish in Peru and Chile. However, with the growing market for high-quality fish oils and the decrease in numbers of wild fish in South America, there is a steadily increasing amount of oil being produced in other areas. Often this is related to growing government environmental regulations limiting the amount of processing waste that can be discarded into bodies of water.

Cooking Cooking the raw fish is necessary to coagulate the proteins, thus allowing the separation of oil from the solids. There are several types of cookers used in fish meal production. The most often used is a steam-jacketed cooker where the heat is supplied to the fish from steam passing through a jacket that surrounds the cooker. Many cookers also have a steam-heated augur in hollow flights that agitate the cooking fish as well as supplying additional heat.

Batch cookers are used in some processing plants, but the large volumes of fish passing through a conventional meal plant require a continuous process. This improves the economics, consistency, and quality control of the fish oil.

Pressing and drying Cooked fish solids are separated from the liquid phase by a press, normally a single screw or a multiscrew press. Wet solids emerging from the press are called 'the press cake.' The liquid phase, or press liquor, as it comes from the press contains water, oil, water solubles, and suspended solids. From the press, the press cake is passed through a continuous direct or indirect dryer to remove the moisture. The resulting fishmeal is the

primary product of this process, and oil is considered a byproduct that varies in composition and yield.

Separating oil from the press liquor Oil is separated from the other liquid components by first passing the press liquor over a vibrating screen and then centrifuging. The screen removes the solids that have been carried over during the processing operation. An oil quality-control operation, called 'polishing,' is often carried out whereby steam or hot water is used to wash or strip the oil. This final washing and then centrifuging procedure removes many of the contaminants that contribute to degradation during storage. Still quite warm from the cooking process, the oil is pumped into large holding tanks, where it is held for shipment. The solids recovered during screening of the press liquor are added to the press cake prior to drying. The remaining water phase of the press liquor is often sprayed on to the fish being dried in the dryer. This recovers the dissolved solids and reduces the environmental impact of disposing of them in the wastewater.

With the increasing importance of preventing oxidation of the ω-3 fatty acids used for human foods and health, oil should be cooled before being stored and should be introduced into the tank near the bottom and removed from the top. In fact, the quality of oil from conventional fishmeal plants can be significantly improved if it is immediately cooled after being extracted from the hot cooked fish residue and before being pumped into the storage tanks. In order to keep the free fatty acid content low during storage of oil, sludge from the bottom, which contains bacteria and other microorganisms, and water on the surface should be regularly removed. Also, an antioxidant is usually added to the oils after centrifuging, or during other phases of the processing, to further reduce oxidative degeneration during storage.

Dry Rendering

Dry rendering is carried out in a single cooker–dryer in which the fish are cooked and dried to the final product in one operation. This process is quite popular in meat plants where both bone meal and meat scrap meal is made from the low-fat continuing byproduct materials. However, since the fat or oil is not removed in the dry rendering process, the technique is used only in small fish processing plants that utilize low-fat fish. The normally utilized industrial fish containing over 5% oil would render a meal containing 20% or more oil. This would be a gummy, highly oxidized product of little commercial value.

There have been numerous efforts to modify dry rendering techniques for fishmeal, but the economics of processing have prevented viable commercial application. Certainly, in the future, as fish oil becomes a normal component of human foods, some of the processes abandoned in the past will be modified and utilized.

Hydrolysis

Fish flesh can be hydrolyzed by proteolytic enzymes, either endogenous or added, to produce a high-quality functional protein product. The general scheme for this process is shown in Figure 2. Excess oil is removed by centrifugation, or, with high-fat fish, some of the oil comes to the surface and can be decanted. Since the resulting liquefied product has good emulsion abilities, there is a certain amount of oil left in the finished product. In fact, up to 5% oil remains in the protein hydrolysate due to the formation of a water–oil emulsion. If the hydrolysate is dried to remove water, the oil remains in the dried product.

Depending on the proteolytic enzyme chosen, the process is carried out under either acid or basic conditions. The process of hydrolysis carried out by endogenous enzymes at a low pH (e.g., 2–4) has an advantage in that the resulting product is quite stable from spoilage by microorganisms. However, the reaction can be slow and uncontrollable, depending on the source and portions of fish used as substrate. The hydrolysis by endogenous enzymes varies quite widely, depending on the amount of feed in the

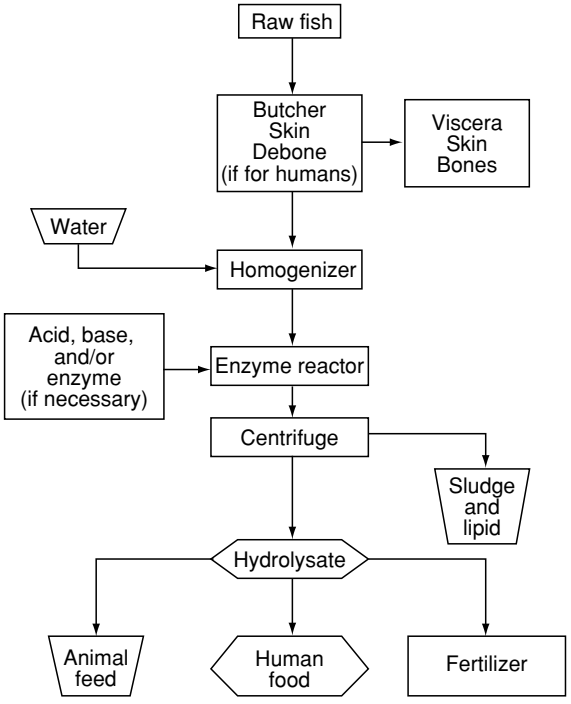

Figure 2 Fish hydrolysate process.

harvested fish stomachs. This is due to the activation of stomach enzymes by the amount and type of food being ingested.

When endogenous enzymes are used for hydrolysis, with an acid addition that maintains the pH at 4 or below, the product is known as 'fish silage.' The crude product resulting from this form of hydrolysis is used for animal feeds or fish fertilizer. As a fertilizer, the product has many advantages since it is a highly digested liquid that can be used as an efficient foliage spray. The resulting hydrolyzed fish product, when properly processed to the end point, has little fish odor and is much more efficient than petrochemical fertilizers, as well as being environmentally friendly.

Hydrolysis can be accomplished at higher pH values, particularly in the basic range, using enzymes extracted from vegetables or manufactured by microbial action. Some of these enzymes are much more efficient than endogenous enzymes in that processes can be more precisely controlled and completed in much less time. However, there is a negative factor in using processes that require near neutral or basic conditions, since the products have a short shelf-life unless stabilized after the hydrolysis is complete.

Solvent Extraction

There has been a considerable amount of work involving the extraction of fish oils by solvents. Alcohol was used in a major development program during the 1960s and 1970s backed by what is now the US National Marine Fisheries Service, National Oceanographic and Atmospheric Administration. This program, an effort to make fish protein concentrate for human consumption, was a dismal failure due to the poor economics of the process. Alcohol, a polar solvent, is not an efficient fat extractor since it preferentially removes water and results in the necessity for carrying out several extractions to ensure recovery of the oil. The cost of removing water by solvent and reclaiming the solvent from the emulsified residue of water, oil, and water solubles is economically unsound.

Extraction with nonpolar solvents is an efficient method of removing oil from animal raw material. A process using ethylene dichloride as the solvent was used commercially for animal feed. This is an efficient and safe process if properly supervised and controlled. However, the problems involving the use of a toxic solvent and the subsequent governmental restrictions prevented extensive use of this solvent for preparing human fish protein concentrate.

The recovery of high-quality oil has proven to be a problem with all solvent extraction processes. The solvents remove many pigments and other components not desired in the oil, and the economics of recovering pure oil from the solvent have not been satisfactorily solved. However, newer techniques are being developed.

Purification and Storage of Oils

Alkaline/Acid Treatment

Purification or refinement of fish oils involves a series of alkaline and acid treatments to remove unwanted products. Alkaline treatment removes the free fatty acids and products of oil decomposition, which produce the strong rancid flavor and odor. It also coagulates the protein to release the oil. The process involves agitating the oil with alkali for various times depending on the temperature and fatty acid content of the oil. Following this treatment, often involving 12 or more hours of agitation and settling, the clear oil is washed with hot water to remove the residuals and the remaining caustic compounds. Pigments must be removed by bleaching, usually by natural or activated clays.

Lipase Hydrolysis

Over the past decade, the use of microbial lipases to produce *n*-3 fatty acid concentrate has been studied. A major advantage of this technology is that there are no extremes of pH or high temperatures, which have an adverse effect on fish oil *n*-3 highly unsaturated fatty acids (eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA)). This technology has tremendous potential for conserving processing energy and increasing the selectivity of specific products. Furthermore, the resulting acylglycerol/ethyl ester mixture can be distilled with much less adverse effect than when fish oil is distilled.

Resin Refinement

A cold process utilizing macroporous cation resins has been developed for improving fish oil chemical, physical, and sensory quality characteristics. Resins include both weak and strong cation and anion resins. Results have been reported indicating that a strong acid cation resin process results in many characteristics that are superior to molecular distillation and thawing fractionation.

Special Products from Fish Oils

Liver Oils

The chemical structure of fish oils gives them certain desirable properties for human consumption and other uses. Certain species of fish with low-fat flesh, such as cod, store most of their oil in their relatively

large livers. This oil was a valuable source of vitamins A and D until the development of less expensive synthetic forms of these vitamins in the late 1940s destroyed the fish liver oil market. Oil from livers is separated by gentle heating with steam. The oil is then washed and centrifuged.

Concentrating Fatty Acids

Various methods have been developed for extracting certain components from fish oils for specific medical or nutritional value. These processing techniques include distillation, crystallization, various chromatographic methods, and supercritical fluid extraction. The pharmaceutical value of fish oil components depends on maintenance of their natural form and structure. For example, it is important to retain the *n*-3 highly unsaturated (long-chain) fatty acids (HUFA) if they are to be valuable in human health. Not only are whole body oils a source of these special fatty acids, but fractionated oils are valuable as a concentrated source of specific desirable fatty acid esters such as those of EPA and DHA.

Molecular distillation Molecular stills can be used for separating oil components based on the molecular weight of the specific component and its vapor pressure. The large difference in molecular weight of vitamins A and D allows good separation of the two components in fish liver oils. Conversely, the small differences between many fatty acid components in oil make it difficult to accomplish effective separation or enrichment of a particular fatty acid.

Chromatographic techniques Fatty acids in fish oils can be fractionally crystallized at low temperatures by mixing them with solvents in which certain fatty acids are more soluble. Separation of components in fish oil can be made by chromatographic techniques in which each component will move through at a different rate in a two-phase system.

Supercritical Separation The potential for commercial scale separation of oil components (e.g., EPA and DHA) was significantly improved by the development of fractionation techniques involving supercritical fluids. Carbon dioxide is the most widely used solvent for food and pharmaceutical products since there are no residual problems involving toxicity, flammability, or environmental pollution. At a pressure of 7384 kPa (1071 psi) and a temperature of 31.1 °C, carbon dioxide becomes a supercritical fluid. As such, it acts as a solvent capable of enriching certain components in an oil. By complexing fish oil with certain chemical compounds (e.g., urea for concentrating EPA and DHA), continuous countercurrent extraction will yield highly enriched products and enriched still raffinate or bottoms. For example, EPA and DHA can be separated from urea-complexed fish oil by this method. In this case, the product is highly enriched EPA concentrated in the raffinate.

See also: **Fatty Acids**: Properties; Dietary Importance; **Fish Oils**: Composition and Properties; Dietary Importance

Further Reading

Ackman RG (1982) Fatty acid composition of fish oils. In: Barlow SM and Stansby ME (eds) *Nutritional Evaluation of Long-chain Fatty Acids in Fish Oil*, pp. 25–88.

Eckert CA, Knutson BL and Debenedetti PG (1996) Supercritical fluids as solvents for chemical and materials processing. *Nature* 383: 313–318.

Liu Y (1995) *Production of Fish Oil from Proteolytic Hydrolysis of Fish Processing By-products*. Thesis, Master of Science, University of Washington, Seattle, WA.

Nicklason P and Pigott GM (1989) Simultaneous heating, forming and extrusion of surimi-based products. *Journal of Food Engineering* 9(3): 210–229.

Pigott GM (1999) *A New Look at Fish Processing Technolog. Symposium for Aquaculture of Central America. Aquaculture and Environment: Together in the New Millenium*. Honduras: Associacion Nacional de Acuicultores de Honduras, San Pedro Sula, pp. 109–115.

Pigott GM and Tucker BW (1990) *Seafood: Effects of Technology on Nutrition*. New York: Marcel Dekker.

Composition and Properties

B W Tucker and G M Pigott, Sea Resources Engineering Inc., Kirkland, WA, USA

Background

Knowledge of the chemistry of fish oils has expanded rapidly during the past few decades, beginning with the development of gas chromatography (GC). More recently, interest in the healthful effects of ω-3 highly unsaturated fatty acids (HUFAs) has spurred identification and characterization of the fatty acid constituents of triglycerides (triacylglycerols).

Composition

Lipids occurring in fish oils are mostly the saponifiable triacylglycerols, with far lesser amounts of di- and monoglycerides. Also present are the unsaponifiable

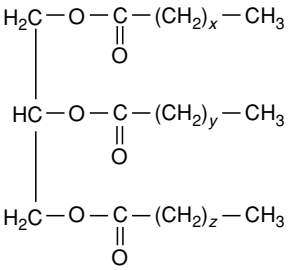

Triglyceride

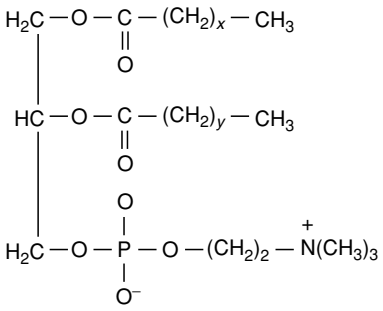

Lecithin

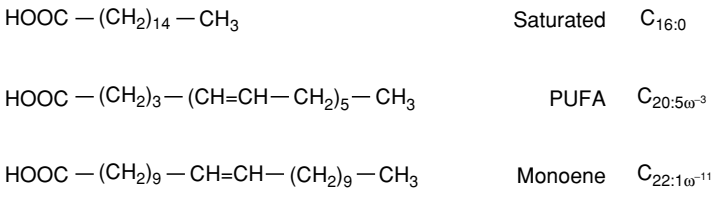

Fatty acids

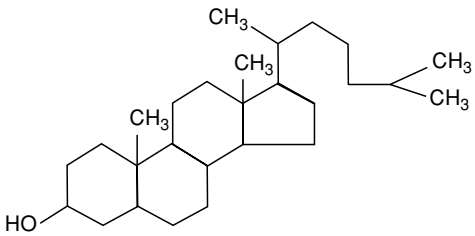

Cholesterol

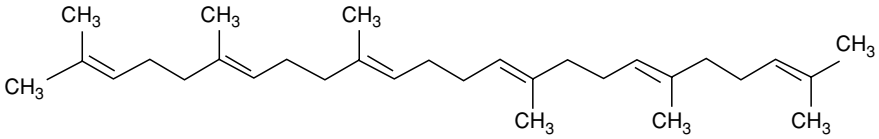

Squalene

Figure 1 Examples of component compounds found in fish oils. Three representative fatty acids (saturated, monounsaturated, and polyunsaturated) typify those found in fish oil triglycerides. Lecithin is the major phospholipid present; cholesterol is the major sterol. Squalene is a hydrocarbon often present at low levels. Reproduced from Fish Oils: Composition and Properties, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

phospholipids, sterols, wax esters, hydrocarbons, and, in certain species, diacyl glyceryl ethers, also known as alkoxydiglycerides. Additionally, minor fat-soluble components such as vitamins, pigments, and contaminants may be present as well as oxidation products. (*See* **Fats:** Classification.)

Fatty Acids

The fatty acid component of the triglycerides and other lipids are saturated, monounsaturated (monoenoic), or polyunsaturated (polyenoic) chains with lengths generally of 12–24 carbon atoms (Figure 1). More than 95% of these chains have even numbers of carbon atoms and are *cis*-oriented. Up to 5% may be odd numbered (Table 1), branched chain, and/or furanoids (F acids). Although 60–80 fatty acids have been identified from a single species, there are four pairs that comprise 80–85% of the fatty acids in fish oil: $C_{14:0}$ and $C_{16:0}$, $C_{16:1}$ and $C_{18:1}$, $C_{20:1}$ and $C_{22:1}$, and $C_{20:5}$ and $C_{22:5}$. The nomenclature specifies first

Table 1 Range of selected fatty acids in commercial fish oils

	Range of total fatty acids (%)		
Fatty acid	*Menhaden oil*	*North American herring oil*	*Mullet oil*
$C_{14:0}$	7.2–12.2	4.6–8.4	4.6–11.5
$C_{16:0}$	15.3–25.6	10.1–18.6	20.2–33.7
$C_{16:1}$	9.3–15.8	6.2–12.0	13.4–29.3
C_{17}	0.2–3.0		0.0–8.2
$C_{18:0}$	2.5–4.1	0.7–2.1	1.8–5.4
$C_{18:1}$	8.3–13.8	9.3–25.2	7.1–13.6
$C_{18:2}$	0.7–2.8	0.1–0.6	0.7–2.7
$C_{18:3}$	0.8–2.3	0.0–1.1	0.3–1.3
$C_{18:4}$	1.7–4.0	1.1–2.8	0.8–4.5
C_{19}			0–2.5
$C_{20:5}$	11.1–16.3	3.9–15.2	4.6–8.1
$C_{22:1}$	0.1–1.4	6.9–30.6	
$C_{22:5}$	1.3–3.8	0.3–1.3	1.7–4.6
$C_{22:6}$	4.6–13.8	2.0–7.8	0.7–3.9

Source: Stansby ME (ed.) (1990) *Fish Oils in Nutrition*. New York: AVI, Van Nostrand Reinhold.

the number of carbon atoms in the chain and next the degree of unsaturation (or number of double bonds) present. (*See* **Fatty Acids**: Properties; Metabolism.)

Compared with other oils or fats, fish oils contain not only a greater variety of fatty acids, but also considerably more long-chain HUFAs. The basic pattern of fish triglycerides usually finds polyenoic fatty acids preferentially bound in the middle (2) position, with monoenes at the ends (1 and 3 positions), although not all species conform to this pattern. Saturated fatty acids comprise 20–35% of the total fatty acids in most fish oils, although herring oil may be up to 50% saturated. The $C_{16:0}$ content is between 10 and 25%, that of $C_{14:0}$ is 3–13%, and that of $C_{18:0}$ is 1–4%. Proportions of fatty acids vary considerably between and within species. Herring oil is characterized by its high percentages of long-chain monoenes, usually about 30%, but often more. Menhaden oil has more $C_{14:0}$ and less $C_{22:6}$ than other fish oils. Monoenoic fatty acids, often present in significant quantities include $C_{16:1}$, $C_{20:1}$, and $C_{22:1}$, the last two being mainly exogenous, originating as dietary components. Cetoleic acid, $C_{22:1\ \omega-11}$, is the major isomer, whereas erucic acid, $C_{22:1\ \omega-9}$, is a relatively minor component. (*See* **Triglycerides**: Structures and Properties.)

More than 90% of the polyunsaturated fatty acids (PUFAs) in fish oils are of the ω-3 configuration: an ethylene bond at the third carbon from the nonpolar or ω end of the molecule. These PUFAs make up 30–40% of the total fatty acids, whereas ω-6 fatty acids constitute less than 10%. In contrast, PUFAs in vegetable oils are almost all ω-6, with less than 5% being ω-3. (*See* **Vegetable Oils**: Dietary Importance.)

Fish oils are characterized by ω-3 HUFAs with five or six ethylenic bonds per acid, and their presence reflects the composition of the diet and other environmental factors, especially water temperature. Fish from colder waters contain larger amounts of HUFAs, which allows more cell membrane flexibility. As related to human health, the most important HUFAs are eicosapentaenoic acid (EPA), $C_{20:5\ \omega-3}$, and docosahexaenoic acid (DHA), $C_{22:6\ \omega-3}$. Although up to 50% of the total fatty acids in certain species may be made up by these fatty acids, the usual concentration found is 17.5–33% with a considerable variation between and within species. Pink salmon contains about 32% HUFAs, whereas the sablefish content is only about 7%.

The influence of environmental factors on HUFA content is illustrated by the difference in DHA levels in menhaden oil. In one study of menhaden, large samplings over 2 years at various times from the Atlantic coast of North America showed a DHA content ranging from 7.3 to 13.1%, whereas the content in menhaden from the US Gulf of Mexico coast was 4.2–8.2%. Another study of herring oil indicated that the EPA contents from Alaskan fish ranged between 11.4 and 15.2%, whereas herring from the east coast of Canada contained just 3.9–8.8%.

Farmed fish (aquaculture) have fatty acid profiles reflecting the fatty acid composition of their feed. Fish oil containing high levels of EPA and DHA is included in feeds used in production of both freshwater and marine species, not only to provide essential ω-3 fatty acids for fish health, but also to generate products nutritionally similar to wild species.

Unreliability of Data

This general discussion of fish oil composition must be qualified. Much available data on fish and fish oil composition are taken from single samplings, small numbers of fish, or samples from specific days. Unfortunately, such data have come to be considered true for all oil from a specific fish species. There exists a difference between species as well as a significant variation in a single species that is influenced mainly by season, but also by catch year, geographical location, spawning cycle, maturity or age of the fish, water temperature, and salinity. The composition of available feed that is determined by the forementioned variables contributes the most significant factor. Even analyses of commercial fish oil from many thousands or millions of fish will vary (Table 1). Statistically reliable data on fatty acid content in fish oil are lacking for nearly all species analyzed. Because of the wide fluctuation in oil content in a single species owing to these variables, data on the average oil content are not meaningful.

As much as a 10% variation in oil content among individual fish of the same species may be found. Fatty acid analyses of commercial fish oils given as ranges of values are more predictive of actual content than data from analysis of oil extracted from a small number of specimens of fish or fish tissues. Additionally, oil as a byproduct of fish meal production will have a different fatty acid content than oil from the same species that has been solvent extracted, perhaps owing to phospholipid fatty acids remaining in the fish meal. HUFAs are particularly susceptible to this phenomenon. (*See* **Fish**: Fish as Food.)

Liver Oils

Fish liver oils contain slightly different proportions of triglyceride and phospholipid fatty acids, as well as minor components, than body oils. The DHA concentration is apt to be higher than in herring or menhaden body oils. Shark liver oils may have amounts of diacylglyceryl ethers approaching those of triglycerides. These glyceryl ethers are also found, to a far lesser degree, in certain shark body oils and a few other species of fish. The fatty acids of these alkoxydiglycerides are predominantly saturated, and monoenoic in the etherlinked portion, and the esterlinked fatty acids are about 25% HUFAs.

Minor Components

Phospholipids, approximately 50% lecithin and 25% cephalin, contribute to the total fatty acid, especially HUFA, content. Phospholipids are associated with cell membranes and, therefore, are scarce in oil produced as a byproduct of meal production. Hydrocarbons are present at very low levels in most fish oils, but the liver oil of several species of shark may be up to 90% squalene (Figure 1) with small amounts of pristine and zamene. Hydrocarbons are usually found in those fish oils containing alkoxydiglycerides. (*See* **Phospholipids**: Properties and Occurrence.)

Sterols, almost all being cholesterol (Figure 1), are always present in crude fish oils, with concentrations between 5 and 8 mg per gram of oil. Processing, however, can remove much free cholesterol and some cholesterol esters.

Wax esters, while present in certain species of fish (e.g., orange roughy), are not found in liver oils or in marketed fillets as the esters occur just under the skin and are removed by deep cutting during fillet preparation. The fat-soluble vitamins A, D and E are always found as minor components, although levels present are subject to the variables previously mentioned. While body oils contain relatively low amounts of A and D, liver oils are considerably higher. Menhaden oil contains 60–150 RE/g of vitamin A and 1.25–2.50 μg/g of vitamin D. Vitamin A in halibut, shark and tuna liver oils is of high potency (up to 210 000 RE/g) while cod liver oil has a lower content (up to 1800 RE/g). Vitamin D in cod liver oil is usually less than 2.5 μg/g while some tuna liver oils reach 1250 μg/g. Vitamin A in halibut, shark and tuna liver oils is of high potency (up to 700 000 IU g^{-1}), while cod liver oil has a lower content (up to 6000 IU g^{-1}). Vitamin D in cod liver oil is usually less than 100 IU g^{-1}, while some tuna liver oils reach 250 000 IU g^{-1}. (*See* **Cholecalciferol**: Physiology; **Retinol**: Physiology; **Tocopherols**: Physiology.)

The vitamin E (tocopherol) content of fish oils is similar to that of vegetable oils. Typical published values range between 40 and 630 $\mu g\,g^{-1}$, with cod liver oil about 560 $\mu g\,g^{-1}$ and herring oil about 100 $\mu g\,g^{-1}$.

Carotenoid pigments such as astaxanthine may be present as a result of dietary components. These pigments give an orange–red tint to the oil. (*See* **Carotenoids**: Physiology.)

Oxidation and hydrolytic products, free fatty acids and amines, peroxides, carbonyls, and volatile compounds will be present to various degrees, reflecting the stability of the oil. Spoilage impurities from fish can contribute objectionable off-odors and off-flavors to their extracted oils. Insoluble matter, such as moisture, dirt, protein, and rust is removed or reduced to $< 5\%$. Soluble impurities include pigments, trace metals, oxidation and decomposition products (e.g., sulfur and phosphorus), mono- and diglycerides and unsaponifiables. Contaminants that occasionally may be found, depending on the source of crude oil, are pesticide residues and chlorinated biphenyls. The processing steps of hydrogenation and deodorization remove or reduce these compounds to nondetectable levels.

Stability

Deterioration of fish oils is associated with oxygen and/or heat. Free fatty acids are hydrolytic products of triglycerides, and levels above 3% are indicative of poor oil quality. Most changes, however, are the result of oxidation of fatty acid portions of triglycerides. The ethylene bonds of fatty acids are very reactive and combine with oxygen and peroxides. Because fish oils are highly unsaturated, they are more vulnerable to exposure to air than vegetable oils, even though the oxidative mechanisms operate in the same manner, albeit at an increased rate. An induction period during which changes are not readily detected is followed by autoxidation involving formation of free radicals and free radical chain propagation (Figure 2). During this time, free radicals react with oxygen to form hydroperoxides. Terminal

Initiation step:

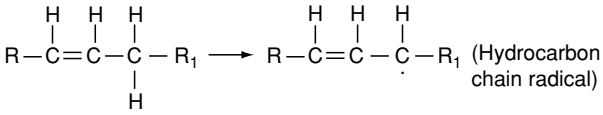

Propagation steps:

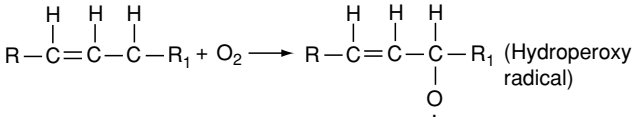

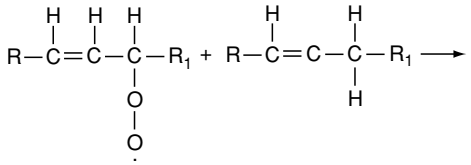

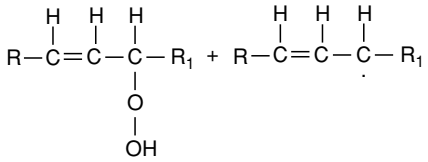

Decomposition of the hydroperoxide forms additional radicals:

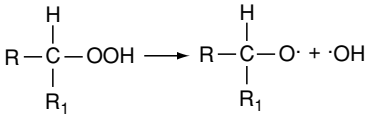

LIPID OXIDATION

AH represents the antioxidant:

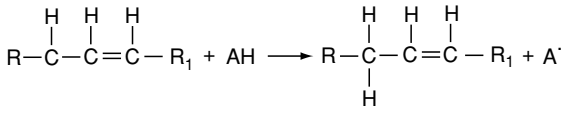

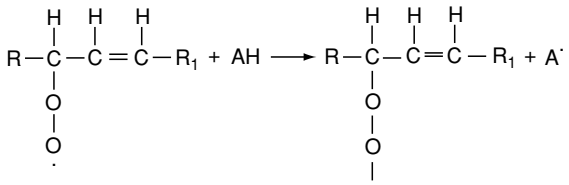

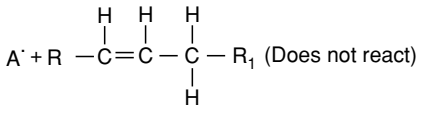

ANTIOXIDANT ACTION

Figure 2 Oxidation and free radical chain propagation leading to rancidity. Antioxidants inhibit and interrupt free radical propagation. Reproduced from Fish Oils: Composition and Properties, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

oxidation products include peroxides, carbonyls such as malonaldehyde, and objectional flavor and odor compounds created by the decomposition of peroxides and oxidation of breakdown products. Plots of storage time versus degree of oxidation give a typical induction curve, indicating an accelerating oxidation rate. The increase in carbonyls and hydroperoxides is roughly equivalent to oxygen uptake. Compared with most other oils, fish oils increase in peroxide value (PV) earlier and the break in the induction curve is less sharp. (*See* **Oxidation of Food Components**.)

Flavors and Odors

'Fishy' flavors and odors occur early in oxidation and will continue if the oil is stored with limited access to air. Even when the PV is low, reversion flavors, characterized as 'beany' or 'grassy,' can be present. Reversion is characterized by a return to preprocessing flavors after storage and occurs before the onset of rancidity. The rancid odors and flavors developing later are typical of all rancid fats and oils. Capillary gas-chromatographic methods for analysis of volatile products have made possible the chemical identification of specific odours. Fish oil stability is best maintained by low-temperature storage and exclusion of oxygen. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics; **Sensory Evaluation**: Aroma.)

Sometimes fish oils are produced from fish that have not been properly refrigerated, and considerable bacterial or enzymatic spoilage may have occurred. Volatile products of spoilage will be partially transferred to the oil. Initially, spoilage products that predominate are nitrogenous, especially trimethylamine. Later, sulfur compounds may contribute to the more obnoxious odors as well as act as catalyst inhibitors.

The presence of prooxidants, such as free fatty acids and other natural impurities, hematin compounds, metal ions, peroxides, heat, air, and light as well as antioxidants and the levels of ω-3 fatty acids, determine the length of the induction period. With ω-3 fatty acids, there is a less distinct break in the induction curve, and antioxidants are less effective during this period. Since chemical tests do not give values that consistently correspond to a specific degree of rancidity, additional sensory evaluation by trained test panels is necessary for satisfactory stability determination. Different degrees of rancidity are readily detectable.

Antioxidants

Antioxidants are substances that, at relatively low concentrations, inhibit the rate of reaction of an oxidizable substrate with oxygen by reacting with a free radical early in the oxidation process. The intermediate compound formed is unable to continue the free radical chain reaction. Synthetic antioxidants – polyphenolic compounds – are added to fish oils to increase induction periods, as they are more effective than the naturally occurring antioxidants that are present in oils of both fish and plants. The tocopherols, eight or more, but mostly α-tocopherol, occur over a range of concentrations in fish oils from 40 to 630 μg per gram of oil. While the vitamin E activity is highest in α-tocopherol, antioxidant activity is less than in some of the other tocopherols. The tocopherols are readily oxidized and are destroyed by heat. Among compounds that exert synergistic antioxidant activity are lecithin, squalene, and metal sequestrants such as ascorbic, citric, and phosphoric acids. (*See* **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants.)

Determination of Fatty Acids and Triglycerides

Gas–liquid chromatography (GLC) is a commonly used method for lipid analysis. Although marine fatty acids are generally derivatized by transesterification, saponification and extraction followed by derivatization to fatty acid methyl esters (FAMEs) may be used. Separation of lipid classes and purification of FAMEs before GC, mass spectroscopy, and Fourier transform infrared spectroscopy, which enable analysis of individual fatty acid derivatives separated by GLC, is followed by computer prediction of identities. High performance liquid chromatography (HPLC) provides higher resolution and greater accuracy, but is more expensive and not universally available. HPLC can identify individual triglycerides (acylglycerides) while mono and diglycerides may be determined, after HPLC separation, with an evaporative light scattering detector. (*See* **Chromatography**: Thin-layer Chromatography; Gas Chromatography; **Fatty Acids**: Analysis; **Mass Spectrometry**: Principles and Instrumentation; **Triglycerides**: Characterization and Determination.)

Chain length can be determined by the linear relation between the logarithm of retention time and the number of carbon atoms. **Tables** of calculations of equivalent chain length (ECL) may be used to identify FAMEs separated by GC.

Further refinement of GC and related technology culminated in a collaborative study to develop an Association of Official Analytical Chemists (AOAC)-approved method to analyze fish oils and fish oil ethyl esters of PUFAs. This method, utilizing wall-coated open tubular GC (WCOT-GC) and flame ionization is especially suitable for analysis of

encapsulated fish oil and concentrated ω-3 components obtained by fractionation from triglycerides.

When such instruments are not available, GC may be used with several other AOAC, American Oil Chemist's Society or International Association of Fish Meal Manufacturers methods such as bromination, hydrogenation, argentation TLC, urea adduction, ozonolysis and/or hydrazine reduction. Bromination distinguishes between saturated and unsaturated FAMEs, as does hydrogenation with reanalysis. Argentation TLC involves separation of FAMEs on silver nitrate-impregnated silica gel-coated TLC plates followed by GC analysis of the fractions. Urea crystallization concentrates branched-chain, cyclic, and unsaturated FAMEs or nonesterified acids. Ozonolysis has long been used to determine the structure of unsaturated fatty acids and esters. The resulting ozonides are then fragmented by chemical reduction or oxidation. Partial hydrazine reduction can locate ethylene bonds.

Determination of Minor Components

Lipids that are present in relatively small amounts or are present in oil from very few species include phospholipids, cholesterol, hydrocarbons, and diacylglyceryl esters. Triglycerides are extracted with diethyl ether, whereas chloroform–methanol extracts phospholipids. Cholesterol is determined by WCOT-GC.

Commercial fish oils are routinely analyzed for free fatty acids, iodine value (IV), moisture, impurities, and color. Free fatty acids are measured by titrating with alkali. The IV measures unsaturation and, therefore, the oxidation potential of the oil. Iodine combines with fatty acids at ethylene bonds and thereby enables a measurement of unsaturation. As oxygen or peroxides react at these bonds, the IV increases. By the Wijs method, oil is combined with potassium iodide, and values are expressed as grams of iodine absorbed per 100 g of sample. Moisture is determined by distillation with an immiscible solvent. The percentage of unsaponifiable matter is gravimetrically determined after several alkali extractions of saponifiables. Color is designated by Gardner numbers after comparison with consecutively numbered standards.

Determination of stability in fish oils involves analyzing for hydrolytic and oxidative degradation products. Chemical tests most often used are to determine free fatty acid, the IV, the thiobarbituric acid (TBA) value, the anisidine value and the gain in weight of the sample during the course of oxidation. Since these tests do not give values that consistently correspond to a specific degree of rancidity, additional sensory evaluation by trained test panels is necessary for satisfactory stability determination.

The hydroperoxide decomposition product, malonaldehyde, reacts with TBA to give a pink color. PV indicates the content of hydroperoxide, an initial chemical oxidation product. The hydroperoxides oxidize potassium iodide and the liberated iodine is titrated with thiosulfate. The anisidine value is related to the level of breakdown products of hydroperoxides present and measures unsaturated aldehydes. Vitamins and contaminants may be detected by several AOAC methods. Heavy metals are best detected with cyclic instrumental neutron activation, although atomic absorption spectrometry is adequate where only a few selected (known) metals are to be determined. Chlorinated biphenyls are best indicated by WCOT-GC. Refined fish oil ready for further processing has minimal identifiable impurities and is, essentially, all triglyceride. (*See* **Contamination of Food; Heavy Metal Toxicology**.)

See also: **Carotenoids**: Physiology; **Cholecalciferol**: Physiology; **Chromatography**: Thin-layer Chromatography; **Fats**: Classification; **Fatty Acids**: Properties; Metabolism; Analysis; **Fish**: Fish as Food; **Mass Spectrometry**: Principles and Instrumentation; **Phospholipids**: Properties and Occurrence; **Retinol**: Physiology; **Tocopherols**: Physiology; **Triglycerides**: Structures and Properties; Characterization and Determination; **Vegetable Oils**: Dietary Importance

Further Reading

Ackman RG (1982) Fatty acid composition of fish oils. In: Barlow SM and Stansby ME (eds) *Nutritional Evaluation of Long-chain Fatty Acids in Fish Oil*, pp. 25–88. London: Academic Press.

AOAC (1990) *Official Methods of Analysis of the AO AC*, 15th edn. Washington, DC: Association of Official Analytical Chemists.

AOCS (1998) *Official Methods and Recommended Practices of the AOCS*, 5th edn. Champagne, IL: American Oil Chemists' Society.

Hamilton RJ and Rice RD (eds) (1995) *Fish Oil: Technology, Nutrition and Marketing*. London: PJ Barnes and Associates.

Nollet LM (ed.) (1996) *Handbook of Food Analysis*, vol. 1. New York: Marcel Dekker.

Pigott GM and Tucker BW (1990) *Seafood: Effects of Technology on Nutrition*. New York: Marcel Dekker.

Stansby ME (ed.) (1967) *Fish Oils*. Westport, CT: AVI.

Stansby ME (1981) Reliability of fatty acid values purporting to represent composition of oil from different species of fish. *Journal of the American Oil Chemist's Society* 58: 13–16.

Stansby ME (ed.) (1990) *Fish Oils in Nutrition*. New York: AVI, Van Nostrand Reinhold.

Windsor M and Barlow S (1981) *Introduction to Fishery By-products*. Farnham: Fishing News Books.

Dietary Importance

R Rice, The Fish Foundation, Tiverton, UK

Fish Oils

The oils from cold-water fish are unique among all of the edible oils and fats. They are the only nutritionally significant source of the long-chain omega-3 (ω-3 or *n*-3) polyunsaturates. The liver oils from fish such as cod and halibut have an added feature in that they are the major natural sources of vitamins A and D.

Changes in eating habits over the last 100 years have resulted in marked changes in fat intake, and in particular in the intake of the polyunsaturates considered essential for human health. Two families of polyunsaturates are considered essential, and they are known as ω-3 (or *n*-3) and ω-6 (or *n*-6). Average intakes of the ω-3 family have fallen considerably, while at the same time, intakes of the ω-6 family have increased dramatically. For reasons that will be explained, this change has placed more emphasis on the need to insure an adequate intake of preformed long-chain ω-3 polyunsaturates from fish. The causes and consequences of this change will be explored in what follows. In particular, attention will focus on the dietary role of the polyunsaturates, on the way in which dietary intake of polyunsaturates has changed over the years, and on the health impact of these dietary changes, in areas such as heart disease, inflammatory disease, skin, lung and kidney diseases, and mental health.

A summary of current suggestions for recommended intakes will complete this overview.

Dietary Polyunsaturates

Essential Fatty Acids

The world is now generally familiar with the concept that certain food components are vital to continued good health. Vitamins, minerals, and the essential amino acids are well studied and documented. What is less well known is that there are also certain types of fat that are essential to good health, fats which cannot be manufactured by the human body, and thus must be supplied preformed in the diet. The concept of 'essential' fats was originated by George and Mildred Burr and colleagues, working in the USA in the late 1920s. They found that rats fed on synthetic diets devoid of any fat developed scaly tails and skin problems that could be corrected by adding some corn oil. Their work started the thinking that there might be some factor or factors in fats that were essential to health. This work, and that of other pioneers such as Ralph Holman and Hugh Sinclair led to the realization that there were two distinct factors, originally termed vitamins F_1 and F_2. It was soon discovered that these factors were not vitamins as such, but nevertheless had significant health impacts if they were excluded from the diet. They were therefore called 'essential fatty acids,' and further work identified the two substances as polyunsaturated fatty acids, linoleic acid (18:2, *w*-6) and α-linolenic acid (18:3, *w*-3). Though this work was of necessity carried out on laboratory animals, such as rats, current nutrition thinking is that humans are equally in need of the essential polyunsaturates. The impact of a lack of dietary polyunsaturates is a long-term impact, with effects taking many weeks or months to become apparent. In such circumstances, it is difficult to see how ethical experiments could be designed to prove the need for the essential polyunsaturates in humans, though some data from human accident victims, and parenteral feeding studies do provide limited direct proof of their essentiality for humans.

Metabolism of Polyunsaturates

The dietary 18 carbon fatty acids linoleic and α-linolenic are converted, by a common enzyme system, into 20 and 22 carbon derivatives (see **Figure 1**). Since the conversion enzymes are the same for both ω-3 and ω-6 polyunsaturates, the phenomenon of competitive inhibition occurs. This means that if there is a predominance of one family over the other, the subordinate family gets crowded out of enzyme sites, and the conversion process favors the dominant family. The result is that an excess of the dietary ω-6 polyunsaturate linoleic acid inhibits conversion of the ω-3 polyunsaturate α-linolenic acid to the 20 and 22 carbon forms eicosapentaenoic acid and docosahexaenoic acid. Fish are the only nutritionally significant source of preformed long-chain ω-3 polyunsaturates, so consumption of fish can in principle compensate for the inhibition of the conversion process. However, in many Western countries, fish consumption is low and falling. The long-chain polyunsaturates (of both families) have most of the true biological functions of the essential polyunsaturates. These can be broadly categorized as:

1. Components of cell membranes, conferring critical properties on the membrane.
2. Precursors of short-lived, powerful biological regulators known collectively as eicosanoids.
3. Components in the regulation of gene-expression.

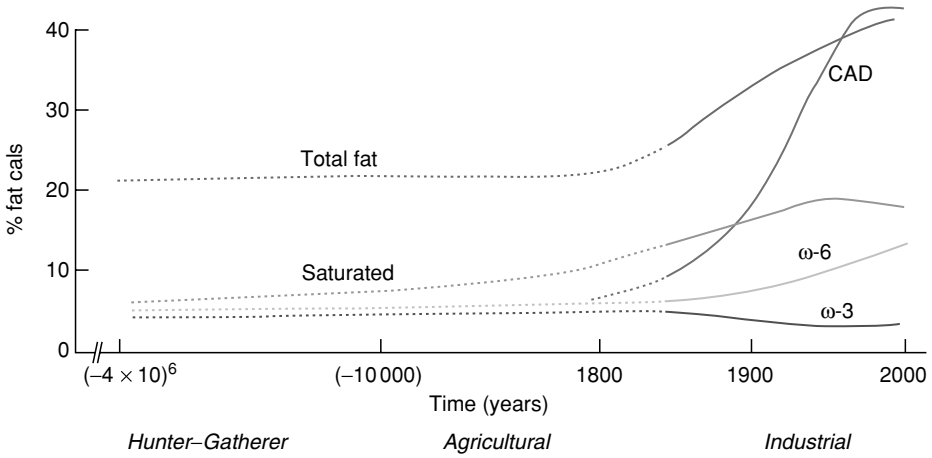

Figure 1 Dietary fat and polyunsaturate changes during man's evolution. CAD, incidence of coronary artery disease. From Leaf A and Weber PA (1987) A new era for science in nutrition. *American Journal of Clinical Nutrition* 45:1053 with permission.

Membrane Composition

All cells in the human body consist of a membrane surrounding the cell contents. These contents vary, depending on whether the cell is a skin cell, a bone cell, or indeed a nerve cell. Whatever job the cell does, the membrane surrounding it will be structurally similar, and an important feature is that it contains significant amounts of polyunsaturates. The exact structure and composition of the membrane in different cells within the body are determined by various factors including genetics, location, function, and diet. Diet affects membrane properties and function, by means of the amount and make-up of the polyunsaturates supplied in the diet and thus available for incorporation. Every cell in the human body has a need for both ω-6 and ω-3 polyunsaturates. Both are needed for optimal health, and if one or the other dominates, health can be compromised, as will be made clear in later sections.

Eicosanoid Precursors

The concept of endocrine hormones (such as insulin) is well established: these are chemical substances secreted into the bloodstream and carried by the circulatory system to any and every cell in the body, and which have specific receptors in target cells that react to the presence of the hormone, bringing about a change.

There is another system of regulators found within the body, known collectively as 'eicosanoids,' named after the 20 carbon fatty acids (Greek eicos=20) from which they are formed. These substances are much more powerful in their actions than the endocrine hormones, so much so that they cannot be centrally produced and distributed, but must be produced locally, in the cells where they are needed, and then destroyed soon afterwards. So far, two families of eicosanoids are well known and understood: the prostaglandins, and the leukotrienes. A third subgroup, known collectively as lipoxins, may also be a part of this wider group, though lipoxins are less well studied or understood at present.

The dietary 18 carbon polyunsaturates are metabolized to 20 carbon forms, which are then available for further metabolism by oxidative enzymes such as cyclooxygenase, or lipoxygenase. The oxidation products of the cyclooxygenase enzyme are available for further metabolism to prostaglandins, whereas the oxidation products of the lipoxygenases are further metabolized to form leukotrienes. Both ω-3 and ω-6 polyunsaturates are capable of being metabolized by these enzyme systems, and the resultant end products, though structurally similar, have functional properties that are slightly but significantly different. The point can be illustrated by reference to the circulatory system.

The clotting power of blood is obviously a valuable self-repair mechanism, but one that can sometimes be triggered when clotting is not needed, and this can lead to formation of a clot, or thrombus, within the intact circulatory system, resulting in a blockage of the flow of blood through a blood vessel. If this happens in an artery supplying blood to heart muscle, a heart attack can be triggered. If it happens in an artery supplying blood to the brain, a stroke can result. The circulatory system has an elaborate system of checks and balances to control this process. The cells that line the interior surfaces of blood vessels are known as endothelial cells, and they are equipped with an enzyme system that makes a substance called prostacyclin from the C20 polyunsaturates contained

within the endothelial cell membrane. Prostacyclin acts as an inhibitor of clot formation, so that if a blood clot comes into contact with endothelial cells, it is dissolved. If there is massive physical damage to the blood vessels, in circumstances where the clotting ability of blood is clearly required to prevent major blood loss, the endothelial tissue being destroyed by the trauma cannot make prostacyclin, and thus clotting is enabled. The proclotting system is based on a prostaglandin called thromboxane, and this is produced in the blood platelets circulating in the bloodstream. Enzymes within the platelet cell convert the C20 polyunsaturates in the platelet membrane to thromboxane. Both ω-3 and ω-6 polyunsaturates can be converted to prostacyclins and thromboxanes. However, and herein lies the significance, the thromboxane produced from conversion of the ω-3 20 carbon polyunsaturate eicosapentaenoic acid (EPA, 20:5, ω-3) is a weak promoter of clotting, whereas its counterpart produced from the ω-6 20 carbon polyunsaturate, arachidonic acid (AA, 20:4, ω-6) is strongly proclotting. The prostacyclin produced from the ω-3 family and the prostacyclin produced from the ω-6 family appear to have comparable levels of anticlotting power, so that the net effect of increasing the level of ω-3 polyunsaturates in cell membranes is to make clotting less likely. Of course, this can be carried to extremes, and if the ω-3 family dominates, clotting can be inhibited to such an extent that uncontrolled bleeding and subsequent blood loss can threaten life. Inuits eating their traditional diets have found themselves in this situation, and there are many reports of Inuits dying from blood loss after wounds and accidents. This imbalance is as much a problem as having too much ω-6, which predisposes to thrombotic diseases such as heart attacks and strokes.

Gene Expression

The way in which the information that is contained within the genetic code is actually put into physical form is called 'gene expression.' Though the mechanisms that are used to achieve this are known, there is still much to learn. At present, there is evidence to show that dietary polyunsaturates do have an influence over this process, but the precise way in which this influence is exercised is not understood.

Time Trends in Dietary Polyunsaturate Intake

In view of the fact that our ability to accurately measure small amounts of polyunsaturates has only existed since the mid-1950s, it is difficult to research the levels of dietary polyunsaturates much before this date. Estimates can be made, knowing the levels of polyunsaturates in certain foods (from modern-day measurements) and knowing what foods would have been available to eat at certain times. Using this approach, experts in the area have estimated that during the times prior to and during the period in which modern man evolved, the balance of the two families of dietary polyunsaturates would have been close to 1:1. Though many thousands of years may have passed since those times, evolution is such a slow process, that in evolutionary terms, the human race is still in the 'Stone Age' stage of physical development. It is therefore reasonable to assume that the biological systems that control human health today will be similar to those that evolved all those eons ago. Those systems evolved on a mix of the two families that was close to 1:1. Today, in the West, we have a diet that supplies around 10 times more of the ω-6 polyunsaturates than the ω-3s. This has come about because ω-6 intakes have increased while ω-3 intakes have declined, and much of this change has taken place within the past 50 years (see **Figure 2**).

Since the mid-1950s, data from the UK show that national intakes of the ω-6 family have more than doubled in response to advertisements promoting the cholesterol-lowering benefits of vegetable oils and margarines (such as sunflower, sesame, corn, etc.) in an attempt to increase protection against heart disease. At the same time, our national intake of ω-3 polyunsaturates has dropped, mainly as a result of reduced fish consumption and a preference for white fish like cod or haddock that have relatively low oil (and so low ω-3) content. A further complicating factor is that our ability to elongate the 18 carbon ω-3 polyunsaturate, α-linolenic acid to the 20 and 22 carbon forms is compromised by the current high intake of the ω-6 polyunsaturate linoleic acid (see **Figure 1**). Thus, not only is the dietary intake of the 20 and 22 carbon ω-3 polyunsaturates reduced, but our ability to make our own from the 18 carbon form is also reduced.

The overall consequence of these long-term dietary changes is that the cells in our bodies are increasingly being primed with the wrong mixture – too much ω-6 and not enough ω-3 – with variable effects depending on which part of the body is being considered.

Health Impact of a Change in Dietary Polyunsaturates

The exact impact of a change in the dietary balance between the two families of polyunsaturates will depend on which area of the body is under consideration. The following sections will examine the impact on the health of heart, brain, skin, lung, and kidney.

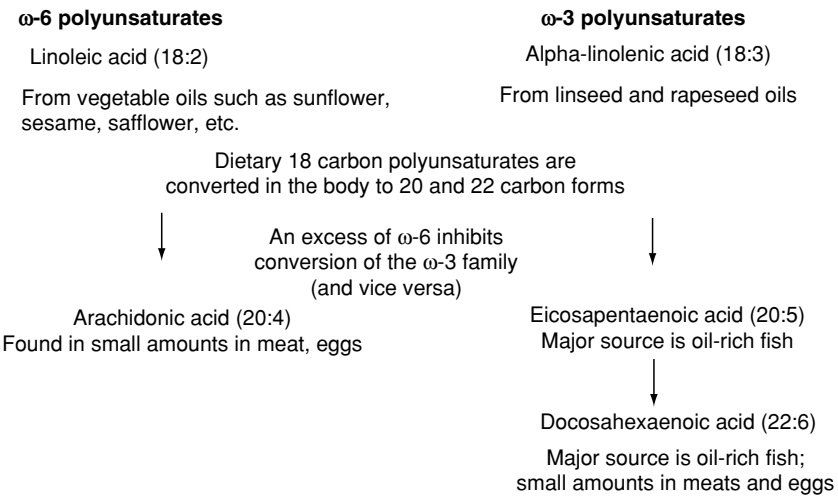

Figure 2 Importance of dietary balance with respect to the polyunsaturates. The figures after the name of a polyunsaturate denote the number of carbon atoms in the molecule, with the figure after the colon indicating the number of double bonds.

Heart Health

Since the mid-1970s, when interest in this field was sparked by observations that native-dwelling Greenland Inuits had high fat/high cholesterol diets yet no heart disease, much research has focused on the role of the ω-3 polyunsaturates in heart disease. As a result of this work, it is now widely acknowledged that the long-chain ω-3 polyunsaturates have a beneficial impact on several established risk factors for heart disease. The blood lipid profile is improved, blood pressure and viscosity are lowered somewhat, blood clotting is made less likely, and the risk of development of abnormally rapid heart rhythm is reduced.

Cholesterol The initial investigations focused on the effect of the long-chain ω-3 polyunsaturates on blood lipids. Early work suggested that the ω-3s might have a cholesterol-lowering effect. The apparent lowering of cholesterol was in fact subsequently shown to be due to a reduction in the level of very-low-density lipoprotein cholesterol (VLDL) particles in individuals with abnormally high levels of such particles. The VLDL particles were cleared from the blood of these patients, as a result of the ability of long-chain ω-3 polyunsaturates to lower serum levels of triacylglycerides (triglycerides), the major component of VLDL particles. This observation resulted in a lowering of total serum cholesterol, as a consequence of the removal of the small amount of cholesterol present in the large amounts of VLDL particles. This led to the erroneous impression that low-density lipoprotein (LDL) cholesterol was being lowered, something that was later shown to be incorrect. An elevated serum LDL level is a major risk factor for heart disease, and this mechanism was for a time thought to reflect the way in which the ω-3s offered their heart protection. The consensus of opinion now is that the long-chain ω-3 polyunsaturates do not have any marked impact on serum LDL cholesterol levels, and that the protective form of cholesterol, high-density lipoprotein (HDL) cholesterol, is elevated by around 10% in most people. The early studies on the ability of the long-chain ω-3 polyunsaturates to reduce blood levels of triacylglycerides were carried out largely in the UK. Though, at the time, an elevated serum triacylglyceride level was not considered to be an independent risk factor for heart disease, that view is now changing, particularly since the inverse relationship between serum triacylglyceride levels and high-density lipoprotein cholesterol levels has been better understood.

The long-chain ω-3s have also been shown to have a moderate but clinically significant blood pressure-lowering effect, especially in subjects with above-normal starting levels. Blood viscosity is also mildly but significantly lowered by the long-chain ω-3s. The impact of the long-chain ω-3 polyunsaturates on the balance of proclotting and anticlotting forces within the blood clotting process also contributes to protection against a heart attack, though this effect is not easy to demonstrate in a clinically meaningful way.

Though a heart attack is triggered by a blood clot formed within the circulation, the real damage occurs when the rhythm control system of the heart causes it to beat very rapidly in an attempt to pump more blood to the brain. This leads to the heart pumping so rapidly that its effectiveness as a pump is lost (ventricular arrhythmia), and death from brain oxygen

starvation follows rapidly. Myocytes are the cells within the heart tissue that generate the tiny electrical impulses that cause heart muscle cells to contract, and the heart to 'beat.' When myocytes are given extra ω-3 polyunsaturates in their make-up, they are much less prone to develop rhythm irregularities when some sort of adverse condition, such as a heart attack, occurs. Work carried out by Harvard Professor Alex Leaf has established that the type and level of polyunsaturate present in myocyte membranes have a major impact on the cell's susceptibility to the development of abnormally rapid rhythms. What this means in terms of human health is not easy to research, because of the practical and ethical problems involved. The nearest approach is the use of laboratory investigations with animals. Work by Dr John Charnock, at the CSIRO in Australia, has shown that giving more of the long-chain 20 and 22 carbon ω-3 polyunsaturates (e.g., EPA and DHA) to rats at high risk of developing cardiac rhythm disturbances completely prevented the appearance of the rhythm irregularities.

Studies with Heart Patients The key question is whether eating more long-chain ω-3 polyunsaturates (either from fish or from fish oil supplements) is useful in helping people at high risk of dying from a heart attack to live longer. Now, there are three separate, independent, controlled studies from different parts of the world demonstrating that this is indeed the case.

The first such study was carried out by a team from the Medical Research Council in the UK. One thousand heart attack survivors were advised to eat more fish, or take more long-chain ω-3 polyunsaturates in the form of fish oil. When they looked at the death rate over the following 2 years, the researchers found that, compared with a similar group of patients advised to eat more dietary fiber, or to eat less fat, the fish group had about 30% fewer deaths. A subsidiary investigation of these patients revealed that the effect was due to the ω-3 polyunsaturates in the fish, rather than some other component.

An Indian study also looked at the impact of recommending more ω-3 polyunsaturates to heart attack victims, and found a striking protective effect of the ω-3 polyunsaturates compared with patients given placebo.

The most recent, and by far the most important, study scientifically was reported from Italy in 1999. There, doctors studied more than 11 000 heart attack victims. Half were given an ω-3 supplement derived from fish oil, and the other half a placebo. After more than 3 years of follow-up, analysis revealed that the total mortality in the patients given ω-3 fatty acids was 20% lower than in those patients not so treated, and the incidence of sudden cardiac death was reduced by 45%.

Though the early studies used relatively large amounts of ω-3 (3000–6000 mg per day), more recent clinical studies have used smaller amounts (1000–2000 mg per day) and shown them to be effective at reducing death rates after a heart attack. These studies clearly show that if heart attack survivors were to be recommended to eat oil-rich fish, or take a fish oil supplement immediately after their attack, their chances of falling victim to a second attack would be greatly reduced.

Brain Health

Brain cells are different in many ways from other body cells. In contrast to most other body cells, they are formed at or around birth, and after the first year or so of life, few further brain cells are formed. The cells do, however, increase in size and form synapses, or connections, with other brain cells. There is also a high proportion of membrane in brain cells compared with content. As in heart cells, the mix of polyunsaturates in the brain cell membrane has a marked impact on how it works and, of course, on brain function.

Studies completed within the past few years have demonstrated that the level and amount of polyunsaturates in the diet can alter the composition of brain membrane polyunsaturates. Other recent studies have shown that adding ω-3 polyunsaturates to the diet can alter the way the brain functions. Increasing the dietary intake of ω-3 polyunsaturates has been shown to:

> reduce severe manic depression; ameliorate the worst effects of schizophrenia; reduce aggression under stress; improve the behavior and reading skills of children with attention deficit hyperactivity disorder (ADHD); and improve problem-solving ability in very young children.

An inadequate intake of ω-3 has also been implicated in postnatal depression and may be associated with dyspraxia in children. The 3–4 point IQ advantage found in breastfed infants compared with bottle-fed infants is also probably related to the natural ω-3 content of breast milk. Only recently have some infant formula milks been fortified with long-chain ω-3s, a trend that seems certain to continue.

Skin

The membranes of skin cells are perhaps some of the most important in the human body since they represent the ultimate barrier between our inner selves and the outside world. The part that polyunsaturates play in skin health has been extensively investigated, and

their role is relatively well understood. The moisture barrier properties of the skin are of key importance, and without a good level and balance of polyunsaturates in skin cell membranes, moisture loss would be severe. The sensitivity of skin cells to problems like psoriasis and dermatitis is related to the level of ω-3 polyunsaturates present. Recent research has suggested that there may be a link between the level and type of polyunsaturate in skin cell membranes and the recent upsurge in prevalence of asthma and allergic conditions.

One of the adverse conditions that skin cells have to face is sunlight. Studies have shown that sensitivity to ultraviolet radiation is reduced in skin cells that have a good level of ω-3, implying some protection against sunburn. The same research suggests that this could ultimately lead to a degree of protection from skin cancer.

Lungs

The cells that line the airways in the lungs are in effect modified skin cells. Lung cells containing higher levels of ω-3 are less prone to damage by cigarette smoke, and probably by other environmental pollutants. When smoke contacts the lungs, an inflammatory reaction is set up leading to congestion in the lungs and ultimately to a condition known as chronic obstructive pulmonary disease (COPD), or emphysema. Higher than average levels of ω-3 in the diet lead to a reduction in COPD, probably because the lungs are less sensitized and do not react in the same hyperresponsive way as lungs with low ω-3 levels. Cigarette smoking is none the less very damaging to health.

Kidneys

The role of the ω-3 polyunsaturates in healthy kidney function is well documented, although the research is taking a long time to penetrate renal units in hospitals. Work carried out over many years by Dr Jim Donadio at the Hormel Institute in the USA has shown consistently that ω-3 supplementation in patients with failing kidneys restores much of their kidney function and often eliminates the need for a transplant. This effect is thought to be due to a reduced inflammatory response following a greater intake of fish oil.

Associated work has also shown that the presence of ω-3 polyunsaturates when administering the immune suppressing drug, cyclosporin A, can reduce some of the drug's damaging side-effects.

Intake Recommendations

Various bodies have attempted to make recommendations on the amounts of the ω-3 polyunsaturates that should be consumed both by adults and, more recently, by infants. Table 1 summarizes the major recommendations made to date. Some of the earlier recommendations are not specific as to whether the recommendations relate to short-chain or long-chain ω-3 polyunsaturates, so have limited value. Generally, as research has led to a greater understanding of the issues involved, the recommendations have become more specific, and discussions are under way to further refine the existing recommendations. The most recent was an *ad-hoc* Expert Workshop held in the USA in April 2000. The make-up of the group, and their detailed recommendations, can be seen by visiting the ISSFAL website at www.issfal.org.uk. A summary of their recommendations can be seen in Table 2 for adults, and Table 3 for infant feeding. A consensus is emerging that intake of long-chain ω-3 polyunsaturates should be around 0.2–0.3% of energy intake, which, for a 2000-kcal intake, means a total long-chain ω-3 intake of about 400–600 mg per day, or 3–4 g per week.

Recommendations on polyunsaturated fatty acid composition of infant feeding formulae have not yet

Table 1 Summary of recommendations for intake of ω-3 polyunsaturates

Source	*Specific ω-6:ω-3 ratio recommended*	*Other specific recommendations*
National Nutrition Council (Norway) (1989)	None	ω-3 0.5%en (1–2 g per day)
Nordic Nutrition Committee (1989)	None	ω-3 0.5%en (1–2 g per day)
NATO Workshop on ω-3/ω-6 (1989)	None	0.8 g per day EPA/DHA (0.27%en)
Scientific Review Committee of Canada (1990)	5:1–6:1	ω-3 at least 0.5%en
British Nutrition Task Force (1992)	6:1	EPA 0.2–0.5%en; DHA 0.5%en
Scientific Committee for Food of the European Community (1993)	4.5:1–6:1.5	ω-3 as 0.5% of total calories
FAO/WHO Expert Committee on Fats & Oils in Human Nutrition (1994)	5:1–10:1	Consider preformed DHA in pregnancy
Committee on Medical Aspects of Food Policy (1994)	None	EPA/DHA 0.1–0.2 g per day (1.5 g per week); at least two portions of fish/week

Table 2 Recommended AI[a] for adults (by an *ad-hoc* Expert Workshop held in USA, April 2000)

Fatty acid	*g per day (2000 kcal)*	*Percentage*
LA, 18:2, ω-6	4.44	2.0
(upper limit)[b]	6.67	3.0
ALA	2.22	1.0
EPA[c] + DHA[d]	0.65	0.3
DHA[e] *minimum*	0.22	0.1
EPA *minimum*	0.22	0.1

[a]If sufficient scientific evidence is not available to calculate an estimated average requirement, a reference intake called an Adequate Intake is used instead of an RDA. The AI is a value based on experimentally derived intake levels or approximations of observed mean nutrient intakes by a group (or groups) of healthy people. The AI for children and adults is expected to meet or exceed the amount needed to maintain a defined nutritional state or criterion of adequacy in essentially all members of a specific healthy population.
[b]Although the recommendation is for AI, the Working Group felt that there is enough scientific evidence to also state an upper limit for LA of 6.67 g per day based on a 2000-kcal diet or of 3.0% of energy.
[c]Elcosapentaenoic acid, C20:5; ω-3.
[d]Docosahexaenoic acid, C22:6; ω-3.
[e]For pregnant and lactating women, ensure 300 mg per day of DHA.

Table 3 Adequate intake (AI)[a] for infant formula/diet (by an *ad-hoc* Expert Workshop held in USA, April 2000)

Fatty acid	*Percentage of fatty acids*
LA, 18:2; ω-6.[b]	10.00
ALA 18:3; ω-3.	1.50
AA, 20:4; ω-6.[c]	0.50
DHA, 22:6; ω-3.[c]	0.35
EPA, 20:5; ω-3.[d]	< 0.10

[a]If sufficient scientific evidence is not available to calculate an estimated average requirement, a reference intake called an Adequate Intake is used instead of an RDA. The AI is a value based on experimentally derived intake levels or approximations of observed mean nutrient intakes by a group (or groups) of healthy people. The AI for children and adults is expected to meet or exceed the amount needed to maintain a defined nutritional state or criterion of adequacy in essentially all members of a specific healthy population.
[b]The Working Group recognized that in countries like Japan, the breast milk content of LA is 6–10% of fatty acids, and the DHA is higher, about 0.6%. The formula/diet composition described here is patterned on infant formula studies in Western countries.
[c]The Working Group endorsed the addition of the principal long-chain polyunsaturates, AA and DHA, to all infant formulas.
[d]EPA is a natural constituent of breast milk but, in amounts more than 0.1% in infant formula, may antagonize AA and interfere with infant growth.

been widely published. The *ad hoc* Expert Group proposed that infant feeds should contain around 0.35% of the fat in the form of the long-chain ω-3 polyunsaturate DHA, together with 0.5% in the form of the long-chain ω-6 polyunsaturate arachidonic acid.

Conclusion

At first sight, the idea that a simple dietary alteration can bring about such significant changes in so many areas of the body is difficult to accept. The key to understanding this is the cell membrane, and the role of polyunsaturated fatty acids in the health and well-being of that membrane. Over the past 50–100 years, the level and amount of the two families of polyunsaturates, the ω-6 family and the ω-3 family, have changed. ω-6 intake has doubled, whereas ω-3 intake has halved. The impact of this change is not immediately obvious; the effects are subtle and take a long time to manifest themselves. During that time-scale, other things have changed also, so that pinpointing the factors responsible for certain effects is difficult. What cannot be denied is that during the past 50–100 years, the West has seen more heart disease, more depression, more aggression, more antisocial behavior, more lung disease, more skin problems, and more kidney ailments. Altering population intakes of dietary polyunsaturates so as to consume more long-chain ω-3 polyunsaturates, and perhaps less of the ω-6s would help to bring about a reversal of these trends.

See also: **Cholesterol**: Role of Cholesterol in Heart Disease; **Coronary Heart Disease**: Etiology and Risk Factor; **Fatty Acids**: Properties; Metabolism; Gamma-linolenic Acid; Analysis; Dietary Importance

Further Reading

Boelsma E, Hendriks HFJ and Roza L (2001) Nutritional skin care: health effects of micronutrients and fatty acids. *American Journal of Clinical Nutrition* 73(5): 853–864.

Das UN (2000) Beneficial effect(s) of *n*-3 fatty acids in cardiovascular diseases: but, why and how? *Prostaglandins Leukotrienes and Essential Fatty Acids* 63(6): 351–362.

Donadio JV, Larson TS, Bergstralh EJ and Grande JP (2001) A randomized trial of high-dose compared with low-dose Omega-3 fatty acids in severe Iga nephropathy. *Journal of the American Society of Nephrology* 12(4): 791–799.

Hamazaki T and Okuyama H (eds) (2001) *Fatty Acids and Lipids – New Findings*. Proceedings of the 4th Congress of the International Society for the Study of Fatty Acids and Lipids. Basel: Karger.

Holm T, Andreassen AK and Aukrust P (2001) Omega-3 fatty acids improve blood pressure control and preserve renal function in hypertensive heart transplant recipients. *European Heart Journal* 22(5): 428–436.

Huang Y-S and Sinclair AJ (eds) (1998) *Lipids in Infant Nutrition*. Champaign, IL: AOCS Press.

Jumpsen J and Clandinin MT (1995) *Brain Development: Relationship to Dietary Lipid and Lipid Metabolism*. Champaign, IL: AOCS Press.

Nettleton JA (1994) *Omega-3 Fatty Acids and Health*. New York: Chapman & Hall.

O'Keefe JH (2000) From Inuit to implementation: Omega-3 fatty acids come of age. *Mayo Clinic Proceedings* 75(6): 607–614.

Price PT (2000) Omega-3 polyunsaturated fatty acid regulation of gene expression. *Current Opinion in Lipidology* 11(1): 3–7.

Rogers PJ (2001) A healthy body, a healthy mind: Long-term impact of diet on mood and cognitive function. *Proceedings of the Nutrition Society* 60(1): 135–143.

SanGiovanni JP (2000) Meta-analysis of dietary essential fatty acids and long-chain polyunsaturated fatty acids as they relate to visual resolution acuity in healthy preterm infants. *Pediatrics* 105(6): 1292–1298.

Simopoulos AP (ed.) (1998) *The Return of ω-3 Fatty Acids to the Food Supply. 1. Land-based Animal Food Products and their Health Effects*. Basel: Karger.

Teitelbaum JE and Walker WA (2001) Review: the role of omega 3 fatty acids in intestinal inflammation. *Journal of Nutritional Biochemistry* 12(1): 21–32.

FLAVOR (FLAVOUR) COMPOUNDS

Contents

Structures and Characteristics

T Cserháti and E Forgács, Chemical Research Center, Budapest, Hungary

Introduction

The quantity and composition of flavor compounds in foods and food products exert a marked influence on consumer acceptance and, consequently, on the commercial value of the products. It has been established many times that one of the main properties employed for the evaluation of product quality is flavor, that is, an adequate flavor composition considerably enhances the marketability of products. A desirable flavor is due to the complex mixture of various volatile and nonvolatile constituents mixed in proper proportions. However, flavor is not only a physical and/or chemical term; it includes the interaction of food and consumer. Flavor sensation is a dynamic process: the volatile or nonvolatile flavor compounds first have to be released from the foodstuff and then transferred to the adequate receptor organs. An exact knowledge of the stability of the flavor compounds in the face of hydrolysis, oxidation, and other environmental and technological conditions is also of paramount importance, because it may help to predict the shelf-life of products and to assess the influence of technological steps on flavor compounds, resulting in more consumer-friendly processing methods. Furthermore, the qualitative determination and identification of flavor compounds may contribute to the establishment of the provenance of the product and to its exact qualification. Unfortunately, traditional analytical methods are generally unsuitable for the separation and accurate determination of particular fractions of flavor compounds. Moreover, they do not contain any useful information on the concentration of the individual flavor compounds and are not suitable for their identification. The high separation capacity of various chromatographic techniques makes them a preferred method for the analysis of flavor compounds in foods. In addition to influencing commercial value, some flavor compounds are believed to have advantageous effect on human well-being as enhancers of appetite and promoters of digestion.

This article overviews the flavor compounds found in various foods and food products, their chemical composition, physicochemical parameters, sensorial characteristics and their separation, quantitation, and identification by various chromatographic techniques.

Chemical and Physical Characteristics of Flavor Compounds

A considerable amount of compounds have been identified in a wide variety of foods producing flavor sensation. The chemical structures of these molecules show high diversity and even flavor compounds with

highly similar chemical structures may have an entirely different sensorial effect. Therefore, the classification of flavor compounds according to their chemical structure (i.e., linear aldehydes, alcohols) is possible but from a practical point of view is irrelevant. Most flavor molecules are hydrocarbons, alcohols, aldehydes, ketones, or esters. The backbone of these compounds generally consists of a linear or branched alkane or alkene chain. Saturated and unsaturated ring structures and terpenoids also occur among the basic structures of flavor compounds. Much effort has been devoted to elucidating the quantitative relationship between organoleptic characteristics and molecular structure or physicochemical parameters of flavor compounds. Because of the complexity of molecular structures and organoleptic properties, these calculations have been of moderate success. Correlations can be observed for only a definite set of structurally similar compounds that have a very low predictive value for flavor compounds with different molecular characteristics. Most of these molecules are relatively small; they are volatile or semivolatile and especially suitable for gas chromatographic (GC) analysis. As flavor compounds generally occur in foods in very low concentrations, their enrichment before the chromatographic separation process is a prerequisite of any successful analysis. However, it has to be borne in mind that the efficacy of various preconcentration methods such as steam distillation, solvent extraction, Soxhlet method and supercritical fluid extraction (SFE) considerably depend on the flavor compound to be extracted and on the conditions of extraction. Organic solvent partition and solid-phase microextraction (SPME) can also be employed for the enrichment of flavor compounds. Furthermore, volatile flavor compounds can be purged from the accompanying matrices carried to a cold trap and then can be cryofocused.

Flavor Compounds in Foods and Food Products

Spices

Spices are important commercial commodities with diverse applications in foods and food products. The accurate measurement of the composition of flavor reveals not only the sensory value of the spice but can also be used as a marker of authenticity. Because of the marked structural diversity of flavor compounds in different spices, a large number of chromatographic methods have been developed and applied for the separation, quantitation, and/or identitification of individual components of flavor in spices. Thus, flavor compounds extracted from cinnamon (inner bark of *Cinnamon zeylanicum* Nees) by solvent-assisted SFE have been analyzed by series coupled capillary GC (cGC). The presence of the structurally similar aldehydes and alcohols as the main flavor components has been verified (Table 1). SPME, SFE followed by cGC, and mass spectrometric (MS) detection have been employed for the analysis of flavor compounds of commercial cassia. The results proved that cassia contains not only linear but also benzene-based flavor components (Table 1). A simple and easy-to-carry-out thin-layer chromatographic (TLC) method has been employed for the separation of vanilla bean extracts using automated multiple development. The occurrence of 5-(hydroxymethyl)-2-furfural, 4-hydroxybenzoic acid, 4-hydroxybenzaldehyde, vanillic acid, and vanillin has been identified. It was stated that the method is suitable for a botanical identification and authenticity test of mercantile vanilla products. High-performance liquid chromatography (HPLC) has also found application in the analysis of flavor compounds extracted from cured vanilla beans (*Vanilla fragrans*). The measurements indicated that the extract contained glucovanillin, *p*-hydroxybenzoic acid, *p*-hydroxybenzaldehyde, vanillic acid, vanillin, and ethyl vanillin. The characteristic

Table 1 Alphabetic list of chemical names of flavor compounds found in cinnamon and commercial cassia

No. of flavor compound	*Chemical name*
Cinnamon (inner bark of *Cinnamon zeylanicum* Nees)	
1	*trans*-Cinnamaldehyde
2	Cinnamyl acetate
3	Cinnamyl alcohol
4	Eugenol
5	2-Hydroxycinnamaldehyde
6	2-Methoxycinnamaldehyde
Commercial cassia	
1	Benzaldehyde
2	δ-Cadinene
3	β-Caryophyllene
4	Cinnamaldehyde
5	Cinnamyl acetate
6	Cinnamic acid
7	Cinnamyl acetate
8	Cinnamyl alcohol
9	*p*-Cymene
10	Coumarin
11	Ethyl cynnamate
12	α-Humulene
13	Linalool
14	Methyl cinnamate
15	α-Phellandrene
16	β-Pinene
17	Safrole
18	Terpinene-4-ol
19	α-Terpineol

odorants of fresh rhizomes of ginger (*Zingiber officinale* Roscoe) have also been studied by aroma extract dilution analysis and multidimensional cGC-MS. The chemical name and odor descriptors of flavor compounds identified in the fresh rhizomes of ginger are listed in Table 2.

The data prove again that compounds with similar molecular structure may have highly different organoleptic properties and the sensory effect of flavor compounds with different structures may be similar. Another study established that the pungent flavor compounds in the extract of ginger consist of 3-, 4-, 5-, 6-, 8-, 10-, 12-, and 14-gingerol, 6-, 8-, and 10-shogaol and zingerone. An online SFE-cGC-MS method was developed and employed for the separation and identification of flavor compounds in the extract of caraway fruits (*Carum carvi* L.). The presence of 23 flavor compounds has been verified, indicating the contribution of a considerable number of different molecules to the organoleptic characteristics (Table 3).

The flavor compounds in turmeric powders have been separated by both HPLC and cGC. The HPLC method quantitated bisdemethoxycurcumin (0.49–0.90 wt%), demethoxycurcumin (0.58–1.10 wt%), and curcumin 0.94–3.18 wt%). The cGC method identified 10 structurally different compounds, suggesting that the analytical method applied exerts a marked influence on the results (Table 3). A combined method (cGC-MS/sniffing port analysis) has been used for the evaluation of commercially dried

Table 2 Alphabetic list of chemical names and odor descriptors of flavor compounds found in the fresh rhizomes of ginger (*Zingiber officinale* Roscoe)

No. of flavor compound	*Chemical name*	*Odor descriptor*
1	Borneol	Dry, camphoraceous
2	Bornyl methyl ester	Earthy, musty
3	Camphene hydrate	Camphoraceous
4	1, 8-Cineole	Camphoraceous
5	Citronellal	Japanese pepper tree-like
6	Citronellyl acetate	Musty, dusty, rosy
7	Decanal	Green, waxy
8	(*E*)-2-Decenal	Fatty, green
9	2, 6-Dimethyl-5-heptenal	Green, fruity, melon-like
10	(*Z*)-3, 7-Dimethyl-3, 6-octadienal	Green
11	(*E*)-3, 7-Dimethyl-3, 6-octadienal	Green
12	(*E*)-2-Dodecenal	Fatty, green
13	2-(2′, 3′-Epoxy-3′-methyl-butyl)-3-methylfuran	Green, earthy, citrus-like
14	Geranial	Citrus-like
15	Geraniol	Floral, rosy
16	Geranyl acetate	Floral, rosy
17	2-Heptanol	Mushroom-like, herbaceous
18	Isoborneol	Musty, dusty
19	Isoeugenol	Spicy, floral
20	Linalool	Floral
21	2-(3′-Methyl-2′-butenyl)-3-methylfuran	Green, minty
22	Neral	Citrus-like, peely
23	Nerol	Floral
24	Nonanal	Floral, waxy, green
25	2-Nonanone	Fruity, fatty
26	(*E*)-2-Octenal	Green, nutty, burdock-like, fatty
27	Octanol	Mushroom-like
28	2-Octyl acetate	Fruity, floral
29	2-Pinen-5-ol	Musty, dusty
30	*cis*-Rose oxide	Floral
31	*trans*-Rose oxide	Floral
32	4-Terpineol	Camphoraceous
33	2-Undecanone	Musty, dusty, green
34	Zingiberenol	Metallic, lemony

Table 3 Alphabetic list of chemical names of flavor compounds found in the extract of caraway fruits and turmeric powder

No. of flavor compound	*Chemical name*
Extract of caraway fruits (Carum carvi *L.*)	
1	(*E*)-Carveol
2	Carvone
3	β-Caryophyllene
4	Decanal
5	(*E*)-Dihydrocarvone
6	(*Z*)-Dihydrocarvone
7	3, 7-Dimethyl-2, 6-octa-dienoate
8	τ-Elemene
9	Geranial
10	2, 4-Hexadienal
11	Limonene
12	Linalool
13	(*E*)-*p*-mentha-2, 8-dien-1-ol
14	(*Z*)-p-mentha-2, 8-dien-1-ol
15	2-Methylbenzaldehyde
16	4-Methylbenzaldehyde
17	β-Myrcene
18	(*E*) -β-Ocimene
19	(*Z*)-β-Ocimene
20	Perillaldehyde
21	α-Pinene
22	Sabinene
23	α-Terpineol
Turmeric powder (capillary gas chromatography)	
1	β-Bisabolene
2	Coumaran
3	Curcumene
4	Curlone
5	Dehydrozingerone
6	2-Hydroxy-5-methyl-aceto-phenone
7	3, 4, 5-Trimethyl-2-cyclo-pentene-1-one
8	ar-Tumerol
9	ar-Tumerone
10	Zingiberene

bell peppers (*Capsicum annum*) after rehydration. The chemical names of volatile components and their odour descriptors in parenthesis are compiled in Table 4.

The data prove that flavor compounds with non-typical odor may also contribute to the formation of the sensorial properties of red pepper. The effect of thermal decomposition of capsaicin on the composition of flavor compounds in the absence and in the presence of oleic acid has also been assessed by cGC-MS. It was found that 9-octadecenamide and *N*-vanillyl-9-octadecenamide are only formed in the presence of oleic acid. Characteristic flavor compounds of onion generally formed enzymatically from odorless precursors when it is cut or crushed. Various cGC-MS methods have also been applied for the identification of these onion volatiles, and the occurrence of unstable thiosulfinates such as (*E*)-ethylidene sulfoxide, (*Z*)-ethylidene sulfoxide, methanesulfenic acid methylthio ester, methanesulfenic acid *n*-proprylthio ester, methanesulfenic acid *n*-propenylthio ester, propanesulfenic acid methylthio ester, propenesulfenic acid methylthio ester, propanesulfenic acid *n*-propylthio ester, propenesulfenic acid *n*-propyl-thio ester, and isomers of zwiebelane has been demonstrated. Similar cGC-MS technique has been used for the separation and quantitative determination of volatile sulfur compounds in garlic. The compounds and their concentration in mg $100\,g^{-1}$ fresh tissue are as follows: allyl methyl sulfide (0.167 ± 0.015); dimethyldisulfide (0.204 ± 0.026); diallyl sulfide (0.122 ± 0.014); allyl methyl disulfide (1.193 ± 0.068); dimethyl trisulfide (0.075 ± 0.011); diallyl disulfide (5.297 ± 0.846); allyl methyl trisulfide (0.562 ± 0.078); 3-inyl-4(H)-1,2-dithiin (4.088 ± 2.266); diallyl trisulfide (1.288 ± 0.183), and 2-vinyl-4(H)-1, 3-dithiin (830.069 ± 1.757). (*See* **Herbs**: Herbs of the Compositae.)

Table 4 Alphabetic list of chemical names and odor descriptors (in parantheses) of volatile flavor compounds found in the dried bell peppers (*Capsicum annuum*) after rehydration. (Odor descriptors have not been determined for each component)

No. of flavor compound	*Chemical name*
1	Acetic acid
2	Benzaldehyde
3	2, 3-Butadione (caramel, butter)
4	Butanal
5	1-Butanol
6	2-Butanone
7	2-Buten-2-one
8	*β*-Cyclocitral
9	Diethoxyethane
10	Dimethyldisulfide
11	Dimethylsulfide
12	Dimethyltrisulfide (rotten, onion, leek)
13	*tert*-Dodecanethiol
14	Heptanal(lemon, orange)
15	*cis*-2-Heptenal
16	2-Heptanone
17	*trans*-3-Heptene-2 -one (mushrooms)
18	2, 3-Hexadione
19	Hexanal (grassy, green)
20	*trans*-2-Hexenal
21	1-Hexanol
22	Limonene
23	2-Methoxy-3-isobutyl-pyrazine (bell pepper)
24	Methyl acetate
25	2-Methylbutanal (chocolate)
26	3-Methylbutanal (chocolate)
27	3-Methyl-1-butanal
28	2-Methylfurane
29	6-Methyl-5-hepten-2-one
30	4-Methyl-2-hexanone
31	5-Methyl-2-hexanone
32	2-Methylpropanal (chocolate)
33	1-Methyl-1H-pyrrole
34	Nonanal
35	*β*-Ocimene (fish, rotten, sickly)
36	Octanal
37	1-Octane-3-ol
38	2, 3-Pentadione
39	Pentanal
40	1-Pentanol
41	2-Pentanone
42	1-Pentene-3-ol
43	1-Pentene-3-one (plastic/chemical)
44	3-Pentene-2-one
45	Propanal

Wine and Cider

Besides the characteristic and attractive color, the qualification of wine and cider markedly depends on the quantity and composition of flavor compounds. Because of their impact on the commercial values of these products, flavor compounds in wine and cider have been intensively investigated. Volatile molecules in wines have generally been analyzed by cGC-MS too. The flavor compounds found by this method in Galician (Spain) white wine are compiled in Table 5 illustrating again the complexity of the composition of flavor compounds. A similar cGC method found a fairly different profile of flavor compounds in the wine of Verdejo (Spain): ethyl lactate + *cis*-hexen-1-ol; acetic acid; propanoic acid; linalool; τ-butyrolactone; butanoic acid; α-terpineol; hexanoic acid; benzyl alcohol; 2-phenylethanol; octanoic acid; ethyl hexadecanoate, and decanoic acid. As volatile components also play an important role in the quality of cider they have been separated and identified by cGC-MS. The flavor compounds and their concentrations in fruity and sharp ciders in $mg\,l^{-1}$ are compiled as follows: 1-butanol (2.24;4.09), amyl alcohols (134.29;170.85), isobutanol (7.73;21.92),

Table 5 Alphabetic list of chemical names of flavor compounds found in Galician wine (Spain)

No. of flavor compound	*Chemical name*
1	Benzyl alcohol
2	Butiric acid + hotrienol
3	2-Butoxyethanol
4	Citronellol
5	Decanoic acid
6	Diendiol
7	Diethyl malate
8	Diethyl succinate
9	3-Ethoxypropanol
10	Ethyl-2-butanoate
11	Ethyl-2-butyrate
12	Ethyl decanoate
13	Ethyl hexanoate
14	Ethyl-3-hydroxybutirate
15	Ethyl lactate
16	Ethyl-octanoate
17	Furfural
18	Geraniol
19	Hexanoic acid
20	*cis*-3-Hexen-1-ol
21	*trans*-3-Hexen-1-ol
22	Hexyl acetate
23	3-Hydroxybutiric acid
24	Isoamyl acetate
25	Isobutanol
26	Isobutyric acid
27	Isovaleric acid
28	Linalool
29	2-Methylbutanol
30	3-Methylbutanol
31	4-Methyl-1-pentanol
32	3-Methylthioethanol
34	Nerol
33	2-Methylthiopropanol
35	Octanoic acid
36	2-Phenylethanol
37	2-Phenylethyl acetate
38	Terpineol

2-phenylethanol (185.24;57.30), benzyl alcohol (0.53;1.54), ethyl lactate (11.86;100.27); diethyl succinate (2.27;2.32), ethyl caprylate (2.22;1.92), ethyl caprate (3.39;2.15), ethyl laurate (0.68;0), ethyl palmitate (0.38;0.66), *i*-amyl acetate (0.28;0.58), 2-phenylethyl acetate (1.21;0), dihydro-2(3H)-furanone (3.50;7.90), acetoin (0;1.92), 2,3(R,R,S,S)-butenediol (37.63;90.89), 2,3-butenediol (meso form) (22.10;75.34), 3-(methylthio)-1-propanol (1.41;1.80), 4-ethylphenol (1.37;3.33), caprylic acid (1.22;3.51), capric acid (1.25;2.20), lauric acid (4.88;0.77), and myristic acid (1.91;0). The data indicate that the marked variation in the concentration of flavor compounds really reflects the difference between the two cider varieties. (*See* **Wines**: Wine Tasting.)

Beer and Hop

Similarly to wine and cider, and for the same reasons, flavor compounds in beer have also been vigorously investigated. It has been emphasized many times and empirically proved that the yeast used for fermentation and the conditions of fermentation play a decisive role in the development of flavor compounds and a considerable number of compounds contribute to the organoleptic aspects of beer quality. The presence of anethole, apiole, asaron, camphor, caryophyllene, carvone cedrol, cineol (eucaliptol), cinnamladehyde, citronellal, cuminaldehyde, cymole, eugenol, farnesol, geraniol, limonene, linalool, menthol, 2-methylbutanal, 3-methylbutanal, methyl ketone, methyl slicylate, myrcene, paradol, α-pinene, thujone, and thymol in beers has been demonstrated. It was further found that the oxidation products of humulene present in hop (five isomers of humulene diepoxy) have a typical herbal/spicy flavor and influence advantageously the organoleptic properties of beer. It has been further verified that nonvolatile compounds such as purine nucleosides and free bases may also contribute to the sensory characteristics of beer. (*See* **Beers**: Chemistry of Brewing.)

Size exclusion chromatography has been employed for the separation and quantitative determination of this class of aroma compounds and the concentration ranges ($mg\,l^{-1}$) found in 18 commercial beers were as follows: guanosine and deoxyguanosine (19–110); adenosine and deoxyadenosine (3–43); xanthin (1–41); guanine (1–11), and adenine (1–17). As the occurrence of 2-furalaldehyde (FA) and 5-hydroxymethyl-2-furaldehyde (HMFA) is considered as an indicator of the deterioration of quality, an HPLC method has been developed and employed for their determination in commercial beers. Concentrations of FA and HMFA ($mol\,l^{-1}$) varied between 1.01–8.41 and 2.62–6.80, respectively. Because of their considerable contribution to the bitter taste and typical taste of beer, the composition and amount of so-called hop bitter acids have been extensively studied. The successful HPLC separation of cohumulone, humulone, adhumulone, colupulone, lupulone, and adlupulone from the SFE extract of hop has been reported. The same aroma compounds has also been separated by microemulsion electrokinetic chromatography in 10 min analysis time.

Coffee and Tea

Flavor characteristics play a decisive role in the evaluation and commercial acceptance of coffee and tea. Because of their outstanding importance many methods have been developed and successfully applied for the separation, identification, and

quantitative determination of flavor compounds in coffee and tea. Thus, aroma components of two varieties of roasted coffee (*Coffea arabica, C. canephora* var. *robusta*) have been analyzed by cGC-olfactometry. The chemical names of flavor compounds and their organoleptic evaluation are listed in Table 6. The composition of the flavor constituents was different in the coffee varieties, suggesting that similar data sets can be employed for the identification of the origin of the product. Nonvolatile aroma precursors present in green coffee have also been separated by high-performance gel filtration chromatography. The data indicated that the concentration of total chlorogenic acid, trigonelline, and caffeine show a marked variation according to the green coffee samples. (*See* **Coffee**: Analysis of Coffee Products; **Tea**: Chemistry.)

The production of tea is a complicated biochemical and biological process. The final quality depends on the chemical composition of the original green leaves, and on the enzymatic processes. A cGC method has been used to study the effect of the concentration of sodium chloride on the efficacy of head-space analysis of volatiles present in heat-treated green tea leaves. The results indicated that NaCl markedly increase the recovery of furfural (from 7.94 to 11.59 μg ml^{-1}) and 5-methylfurfural (1.65 to 2.95 μg ml^{-1}) but did not influence the recovery of benzaldehyde (0.30 μg ml^{-1}) and ethyldimethylpyrazine (0.69 μg ml^{-1}). Another study indicated that the decomposition rate of lipids during tea processing results in the formation of the flavor compounds *n*-hexanal and *trans*-hexanal.

Table 6 Alphabetic list of chemical names and odor descriptors of flavor compounds found in two varieties of roasted coffee (*Coffea arabica, C. canephora* var. *robusta*)

No. of flavor compound	*Chemical name*	*Odor descriptor*
1	Acetaldehyde	Fruity, pungent
2	2, 3-Butanedione	Buttery
3	(*E*)-β-damascene	Honey-like, fruity
4	2, 3-Diethyl-5-methylpyrazine	Earthy
5	Dimethyl trisulphide	Cabbage-like
6	2-Ethyl-3, 5-dimethylpyrazine	Earthy, roasty
7	4-Ethylguaiacol	Phenolic, spicy
8	2-Furfurylthiol	Roasty
9	Guaiacol	Phenolic, burnt
10	2-Isobutyl-3-methoxypyrazine	Green, earthy
11	3-Mercapto-3-methylbutylformiate	Catty, roasty
12	Methanethiol	Putrid, cabbage-like, sulfurous
13	Methional	Boiled potato-like
14	2-Methylbutanal	Malty
15	3-Methylbutanal	Malty
16	2-Methyl-2-butenthiol	Foxy, skunky
17	2-Methyl-3-furanthiol	Boiled-meat-like
18	Methylpropanal	Fruity, malty
19	1-Octen-3-one	Mushroom-like
20	2, 3-Pentanedione	Buttery
21	Propanal	Fruity
22	4-Vinylguaiacol	Phenolic, spicy

Dairy Products

Sensorial properties also determine the type and marketability of dairy products. Because of their paramount importance, flavor compounds have been extensively studied, mainly in cheeses. Cheese flavor is influenced by the biodegradation of milk protein, fat, lactose due to the enzyme systems of milk, rennet, and microorganisms. The composition of volatile flavor compounds in Roncal (Spain) cheese has been investigated by preconcentrating the analytes by the 'purge-and-trap' method and then separating them by cGC-MS. The combined technique sussessfully separated hydrocarbons, alcohols, aldehydes, ketones, acids, esters, and sulfur compounds. The flavor compounds identified are shown in Table 7 and prove again the high complexity of aroma composition. (*See* **Cheeses**: Chemistry and Microbiology of Maturation.)

The flavor compounds of Swiss cheese and Gorgonzola have also been analyzed by simultaneous distillation-solvent extraction-cGC. In contrast to Roncal cheese, the main aroma components of Swiss cheese consisted of (in μg kg^{-1}) 2-heptanone (3776), 2-nonanone (1992), 3-methylbutanal (1292), 2-octanone (1188), 2-undecanone (1178), 2-pentanone (999), 2-tridecanone (873), dimethylsulfide (619), phenyl acetaldehyde (500), and heptanal (429). The aroma composition of Gorgonzola cheese was markedly different, containing 2-nonanone (3212), 2-heptanone (2916), 3-methyl-1-butanol (2534), 2-undecanone (1884), 3-methylbutanal (1120), 2-pentanone (955), 2-tridecanone (890), phenyl acetaldehyde (643), 1-octen-3-ol (601), and methyl propanoate (379).

It has been further demonstrated that the composition of flavor compounds depends not only on the variety of cheese but also on the character of starter culture. In the case of Hispánico cheese it has been demonstrated that the concentration of some flavor compounds considerably depended on the properties of starter culture. The chemical names of flavor compounds identified are listed in Table 8. The stereoselective separation of gamma and delta lactones has been achieved by enantioselective multidimensional cGC. It was found that butter, whipped cream, evaporated milk, full-cream milk, cheddar, parmesan, limburger, emmental and blue cheese contain a considerable amount of C_{10} and C_{12} lactones which

Table 7 Alphabetic list of chemical names of flavor compounds found in Roncal cheese (Spain)

No. of flavor compound	*Chemical name*
1	Acetic acid
2	Acetic acid ethyl ester
3	Butan-2, 3-dione
4	Butanoic acid
5	Butanoic acid prop-2-enyl ester
6	Butan-1-ol
7	Butan-2-ol
8	Butan-2-one
9	Carbon disulfide
10	Decanoic acid ethyl ester
11	Decenal
12	Dimethyl disulfide
13	Dimethyl sulfide
14	Heptane
15	Heptanoic acid ethyl ester
16	Heptan-2-ol
17	Heptan-2-one
18	Hexane
19	Hexanoic acid
20	Hexanoic acid ethyl ester
21	Hexan-2-ol
22	Propan-2-en-1-ol
23	Hexan-2-one
24	Hexan-3-one
25	3-Hydroxybutan-2-one
26	Ethanol
27	Ethylbenzene
28	1-Methoxy-2-propanol
29	3-Methylbutanal
30	3-Methylbutane-1-ol
31	3-Methylbutanoic acid
32	3-Methylbut-3-en-1-ol
33	Methylcyclohexane
34	Methylcyclopentane
35	5-Methylhexan-2-one
36	2-Methylpropanoic acid
37	2-Methylpropan-1-ol
38	Nonanal
39	Nonan-2-one
40	Nonenal
41	Octanoic acid ethyl ester
42	Octen-2-ene
43	Pentane
44	Pentan-2-ol
45	Pentan-2-one
46	Propan-2-one
47	Propanoic acid
48	Propan-1-ol
49	Propan-2-en-1-ol
50	Styrene
51	Undecenal

markedly influence the sensorial properties of the products.

The compounds causing flavor defects in milk powder and milk-based baby formula have been separated and identified by cGC and the involvement of hexanal, heptanal, and nonanal in the development of 'cardboard-like' off-flavor has been established.

Table 8 Alphabetic list of chemical names of flavor compounds found in Hispánico cheese

No. of flavor compound	*Chemical name*
Influenced by the starter culture	
1	2-Butanol
2	2-Butanone
3	2, 3-Butenedion
4	Ethyl butyrate/1-propanol
5	Methyl acetate
6	2-Methyl-2-buten-1-ol
7	3-Methyl-2-pentanone
8	2-Methyl-1-propanol
9	2-Pentanone
10	2-Propanol
11	2-Propanone
12	Toluene
Not influenced by the starter culture	
1	Acetaldehyde
2	Acetonitrile
3	Acetophenone
4	1-Butanol/*m*-xylene
5	Carbon disulfide
6	2, 3-Dimethyl benzofuran
7	Ethanol
8	Ethyl acetate
9	Ethyl benzene
10	Ethyl toluene
11	Heptane
12	1-Heptanol
13	2-Heptanone
14	Hexanal
15	1-Methoxy-3-propanol
16	3-Methyl-1-butanal
17	3-Methyl-1-butanol
18	3-Methyl-3-buten-1-ol
19	Naphtalene
20	2-Nonanone
21	Octane
22	2, 3-Pentanedione
23	1-Pentanol
24	2-Pentanol
25	*o*-xylene
26	*p*-xylene

Meat, Fish, and Oils

A large amount of volatile and nonvolatile flavor compounds has been found in foodstuffs containing meat, fish, and oil as the main ingredient. The chemical character and composition of flavor compounds found in these classes of foodstuffs substantially deviated from those determined in other food varieties. A dynamic head space cGC method has been developed and used for the analysis of flavor compounds in chicken breast meat. The following compounds have been found in chicken breast meat previously heated to 70 °C: H_2S, CO_2O, methanol, propanal, and acetone, 2-propanol and thiobismethane, 2-methylpropanal, 1-propanol, 2,3-butanedione, hexane, 2-butanone, 3-methylbutanal, 2-methylbutanal,

1-penten-3-ol, pentanal, dimethyl disulfide, hexanal, 1-hexanol, 2-heptanone, heptanal, 1-heptanol, benzaldehyde, 7-octen-4-ol, octanal, *d*-limonene, undecane, nonanal, and dodecane. Olfactometry and cGC-MS have been employed for the identification of flavor compounds and their odor descriptors in boiled cod and trout. It was found that acetaldehyde (sweet), 2-methylpropanal (malty), butane-2,3-dione (buttery-like), 3-methylbutanal (malty), 1-octen-3-one (mushroom-like), and (*Z*)-1,5-octadien-3-one (geranium-like) occurred in both fishes. Moreover, boiled cod minces contained methaneethiol (sulfurous), trimethylamine, dimethyl sulfide (cabbage-like), 2-methylbutanal (malty), methional (boiled potato-like), dimethyl trisulfide (cabbage-like, putrid) and dimethyl tetrasulfide (cabbage-like, putrid). (*See* **Fish**: Fish as Food; **Meat**: Analysis.)

In contrast, boiled trout contained propionaldehyde (sweet), pentane-2,3-dione (buttery-like), hexanal (green), (*Z*)-3-hexenal (green), (*Z*)-4-heptenal (biscuit-like), (*Z, Z*)-3, 6-nonadienal (fatty, green), and (*E, Z*)-2, 6-nonadienal (cucumber-like). A similar method has been used for the study of the odor active compounds in cooked mussel (*Mytilus edulis*) and the following compounds have been identified: 2, 3-butenedione (buttery, caramel), 1-propanol (fruity, plastic), dimethyl disulfide (sulfury), hexanal (green), *m*-xylene (plastic), heptanal (citrus fruit, green), (*Z*)-4-heptenal (boiled potato), 1,2,4-trimethylbenzene (plastic), octanal (citrus fruit, orange), (*E*)-2-penten-1-ol (mushroom), (*E*)-2-heptenal (sulfury, grassy), ethylpyrazine (nutty), dimethyl trisulfide (sulfury, green, marine), methional (boiled potato), 2-nonanol (fruity, solvent), (*E, E*)-2,4-octadienal (cucumber), 1-acetylpyrazine (nutty), 2-acetylthiazole (grilled hazelnut), 4-methylthiazol (roasted, meaty), 4-ethylbenzaldehyde (fruity, anisic, minty), 2-acetyl-2-thiazoline (grilled hazelnut), and 2,6-dimethylnaphtalene (grilled, grassy).

Volatile aroma compounds have also been determined in oils using various cGC techniques. The flavor compounds found in extra virgin, regular, extra light and extra mild cooking oils are listed in Table 9. The concentration of flavor constituents of oils remarkably depended on the technological steps applied for the production of oils. The amount of aliphetic aldehydes indicating quality deterioration has also been measured by a cGC method using derivatization by thiazolidine. The presence of C_3, C_4, *i*-C_5, C_6, and C_9 aldehydes in sesame, soybean, rapeseed, cottonseed, corn, olive, and sardine oil, as well as in beef fat, lard, butter, margarine, mayonnaise, cheese, egg yolk, fresh milk, chocolate, and potato chips has been demonstrated.

Table 9 Alphabetic list of chemical names of flavor compounds found in extra virgin, regular, extra light, and extra mild cooking oils

No. of flavor compound	*Chemical name*
1	Acetic acid hexyl ester
2	Butanal
3	2-Cyclohexen-1-ol
4	(*Z*)-2-Decanal
5	Ethyl acetate
6	Heptanal
7	Heptane
8	2-Heptanone
9	(*E*)-2-Heptenal
10	(*Z*)-2-Heptenal
11	Hexanal
12	Hexane
13	(*E*)-2-Hexenal
14	(*Z*)-3-Hexenal
15	(*E*)-3-Hexen-1-ol
16	(*Z*)-3-Hexen-1-ol
17	4-Hexen-1-ol
18	(*Z*)-4-Hexen-1-ol
19	(*Z*)-3-Hexen-1-ol acetate
20	Limonene
21	2-Methylbutanal
22	3-Methylbutanal
23	Methylcycloheptane
24	Nonanal
25	Octanal
26	Octane
27	(*E*)-2-Octenal
28	1-Octene
29	2-Octene
30	(*E*)-2-Octene
31	(*Z*)-2-Octene
32	Pentanal
33	3-Pentanone
34	2-Pentylfuran
35	Styrene
36	Toluene
37	2-Undecenal

Miscellaneous Food Products

Because of the significant impact of flavor compounds on the quality of foods and foodstuffs, they have been analyzed in a considerable amount of other food products using the well-established methods of preconcentration followed by either cGC-MS or HPLC. Thus, 46 volatile compounds have been separated and identified in fresh-squeezed unpasteurized orange juice. Ionones, decalactone, and raspberry ketone have been found in some raspberry cultivars and 17 volatile compounds have been identified in ready-made tomato sauces. The presence of 2-methylpropanal, 3-methylbutanal, and 2-methylbutanal has been verified in chocolate flakes and the enzyme activities involved in the flavor precursor formation in unfermented cocoa beans have been elucidated.

The flavor profile of corn-based snacks and peanuts has also been determined and the data have been used for the better understanding of the sensory characteristics of the products.

See also: **Analysis of Food**; **Chromatography**: Principles; Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; Supercritical Fluid Chromatography; Combined Chromatography and Mass Spectrometry; **Flavor (Flavour) Compounds**: Production Methods

Further Reading

Arnault I, Mondy N, Cadoux F and Auger J (2000) Possible interest of various sample transfer techniques for fast gas chromatography–mass spectrometric analysis of true onion volatiles. *Journal of Chromatography. A* 896: 117–124.

Boulanger R, Chassagne D and Crouzet J (1999) Free and bound flavour components of amazonian fruits, 1: Bacuri. *Flavour Fragrance Journal* 14: 303–311.

Charalambous G (ed.) (1993) *Shelf-Life Studies of Foods and Beverages. Chemical, Biological, Physical and Nutritional Aspects*. The Netherlands: Elsevier Science.

Derdelinckx G, van den Brent K, Masschelein CA, Verachtert H and Delvaux F (2000) Biogeneration of flavour (part III): novel methods for beer maturation. *Cerevisia* 25: 57–66.

Enser M (1999) Nutritional effects on meat flavour and stability. In: Richardson RI and Mead GC (eds) *Poultry Meat Science*, pp. 197–214. CAB International.

Ferreira V, Ortega L, Escudero A and Cacho JF (2000) A comparative study of the ability of different solvents and adsorbents to extract aroma compounds from alcoholic beverages. *Journal of Chromatographic Science* 38: 469–476.

Izco JM and Torre P (2000) Characterization of volatile flavour compounds in Roncal cheese extracted by the 'purge and trap' method and analysed by GC-MS. *Food Chemistry* 70: 409–417.

Le Guen S, Prost C and Demaimay M (2000) Characterization of odorant compounds of mussles (Mytlilus edulis) according to their origin using gas chromatography–olfactometry and gas chromatography–mass spectrometry. *Journal of Chromatography A* 896: 361–371.

Maarse H and van der Heij DG (eds) (1994) *Trends in Flavour Research*. The Netherlands: Elsevier Science.

Piggott JR (2000) Review. Dynamism in flavor science and sensory methodology. *Food Research International* 33: 191–197.

Piggott JR and Paterson A (eds) (1994) *Understanding Natural Flavours*. London: Blackie Academic and Professional.

Poole SK, Kiridena W, Miller KG and Poole CF (1995) Review. Planar chromatographic methods for the determination of the quality of spices and flavors as exemplified by cinnamon and vanilla. *Journal of Planar Chromatography and Modern TLC* 8: 257–268.

Swift KAD (1999) *Current Topics in Flavours and Fragrances*. The Netherlands: Kluwer Academic.

van Willige RWG, Linssen JPH and Voragen AGJ (2000) Influence of food matrix on absorption of flavour compounds by linear low-density polyethylene: oil and real food products. *Journal of Science and Food Agriculture* 80: 1790–1797.

Production Methods

A Blake, P Cadby and F Näf, Firmenich SA, Geneva, Switzerland

Introduction

There is a large and established industry developing and manufacturing flavor systems for use in foods and beverages. Legislation and definition of flavors are complex and vary from country to country. Flavor compounds are produced by single or combined physical, chemical, biochemical, biological, and agricultural techniques (**Figures 1–6**).

A major way of producing natural flavors is by isolation from agricultural food products or waste streams of food-processing units using physical methods such as distillation, solvent extraction, and filtration. The flavors thus isolated range from complex mixtures at one extreme, to pure individual flavor chemicals at the other. Flavors produced by isolation are often designated as essential oil extracts or isolates. The raw materials used in the production of nature-identical flavors are nature-identical chemicals (i.e., identical to those found in natural foods) with the optional use of natural extracts, oils, and isolates. The manufacturing is accomplished by mixing the raw materials under tightly controlled conditions and the resulting product is termed 'compounding' or 'blending.' It is understood that chemical reactions must not take place during this process, or on subsequent storage.

The synthesis of organic compounds as flavors may involve (1) controlled chemical synthesis, (2) transformations using living biological systems, (3) enzyme-catalyzed synthesis, and (4) the reaction flavor approach which mimics established food-cooking techniques.

In addition to creating the blend of molecules necessary to provide a specific flavor effect, it is important to deliver these in a way which optimizes

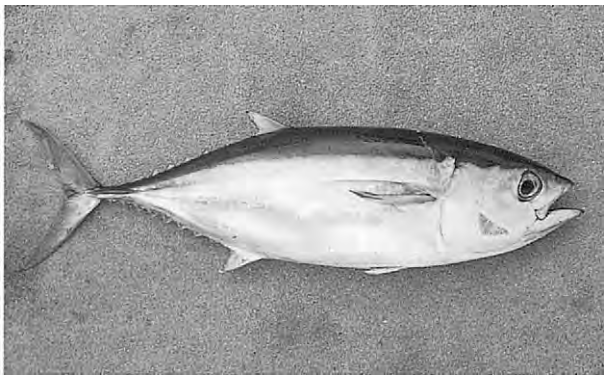

Figure 1 (see color plate 58) Yellowfin tuna. Reproduced from Flavour Compounds, Production Methods, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Figure 2 (see color plate 59) Effluents from food processing. Reproduced from Flavour Compounds, Production Methods, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

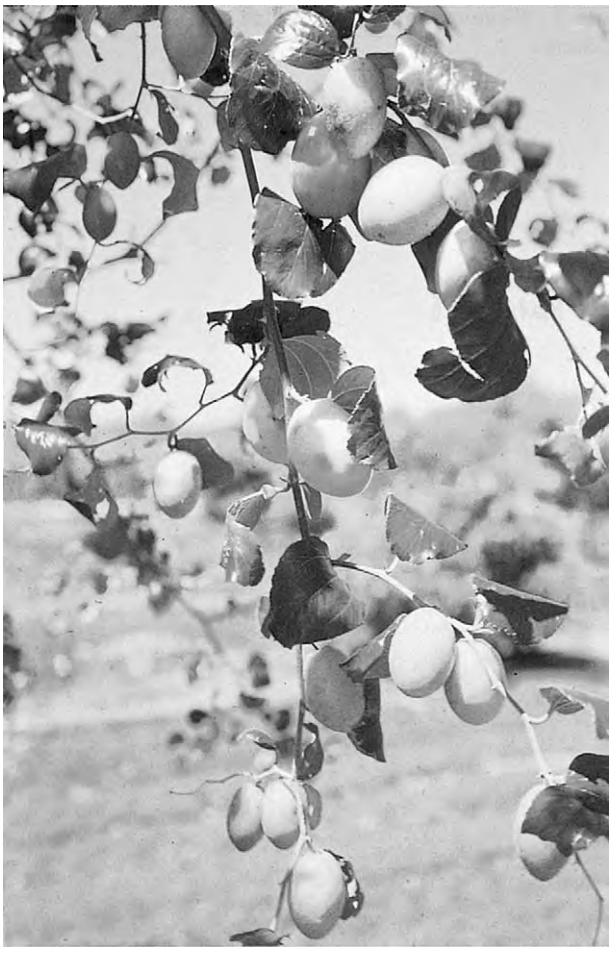

Figure 3 (see color plate 60) Fruits of tropical climates: *Zizyphus mauritiana* Lam. (Indian jujube). Reproduced from Flavour Compounds, Production Methods, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

their functioning. Accordingly, flavors are increasingly sold entrapped in an appropriate delivery system, either in powder or liquid form.

Legislation

Flavor production is highly regulated in order to protect both the consumer and the workers involved in their production. Since its inception in 1973, the International Organization of the Flavor Industry (IOFI) has developed a self-regulatory code of practice covering both of these aspects. Several of its recommendations have been closely modeled by international regulatory bodies. These include standards for natural flavorings and for thermal process flavors. More recently, the Joint Food and Agriculture Organization/World Health Organization Expert Committee on Food Additives (JECFA) has been evaluating the safety of chemically defined flavoring substances under their conditions of use. Since 1997, JECFA has cleared over 100 substances per year in a program that may eventually examine over 2000 flavoring substances claimed to be currently used. Work-place hazard information systems are being put in place in many countries. IOFI regularly updates recommendations on the work-place hazard classification of flavoring ingredients according to the criteria laid down by various governmental bodies.

Isolation of Natural Flavors

Isolation from natural products is the most important and simplest method for producing natural flavors. It is based on the physical properties of the flavors and exploits technologies of a wide range of sophistication. The five principal isolation techniques used

Figure 4 (see color plate 61) Road-side markets in Bangkok displaying citrus, banana, rambutan, mango, and lychee. Reproduced from Flavour Compounds, Production Methods, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Figure 5 (see color plate 62) A collection of the more exotic tropical fruit of South-east Asia. Courtesy of TA Cooke, Plant Pathology, QDPI, Australia. Reproduced from Flavour Compounds, Production Methods, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Figure 6 (see color plate 63) Fruits of tropical climates: *Prunus serotina* var. *salicifolia* (capulin cherry). Reproduced from Flavour Compounds, Production Methods, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

on a production scale are, in decreasing order of importance: (1) distillation; (2) extraction; (3) pressing; (4) filtration; and (5) chromatography. An overview of today's standard technology equipment and their typical applications is presented in Table 1. In certain isolation procedures, acid and/or alkaline adjustments (pH value control) are necessary. Apart from the quality of the raw materials, the sensory properties of the flavors obtained are strongly influenced by the isolation technology (equipment and processes) used. For this reason, the isolation route to many natural flavors remains the unpatented but nevertheless confidential 'know-how' of the various flavor manufacturers (Figures 7–9). (*See* **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **Membrane Techniques**: Principles of Ultrafiltration.)

Production of Nature-Identical Flavors

Nature-identical flavors are mixtures of synthetic nature-identical flavoring substances. These are generally made from available chemical feedstocks of plant or petrochemical origin and are, by definition, chemically identical to substances found in natural foods, herbs, spices, and essential oils. These flavors often contain natural flavoring materials as well but they must not contain flavoring substances which have no counterpart in nature. Over 7000

Table 1 Isolation techniques, equipment, and their application to the production of typical flavor types

Technique	*Equipment*	*Physical property utilized*	*Application to the production of*
Distillation			
Distillation	Still with evaporator, and condenser	Volatility of flavors	
Fractional distillation	Still with evaporator, fractionating columns (1 to several meters high), often with packing	Volatility, differences in boiling point of compounds	Folded citrus oils (concentration by removal of monoterpenes)
Distillation with high-performance condensing device	Recovery units	Low boiling point of flavor	Volatiles from concentration of tomato, apple juice, red fruits, exotic fruits, citrus (essence oils) → 'add-back' systems
Thin-film distillation	Falling-film vaporizer with or without rotating blades	Lowering of evaporation temperature of a liquid thin film as opposed to a bulky mass	Distillation of volatiles from a residue or of high-boiling materials
Molecular distillation	As above, but with extremely low vacuum (10^{-5} bar) and condenser close to vaporizer	Distillation below the boiling point in appropriate equipment and at a certain temperature/pressure (mean free path)	Distillation of volatiles from a residue or high-boiling materials under mild conditions
Steam distillation	Still with condenser and two-phase separator (+ inlet for steam)	Boiling point of a nonmiscible mixture is below the boiling point of the lowest-boiling component	Essential oils, for example, flowers, aromatic plants, and spices
Hydrodiffusion	Vessel with horizontal pallets holding material, steam entering from the top then condensed	Steam volatility of compounds	Seed oils
Extraction			
Extraction by nonpolar, organic solvents and concentration		Solubility of volatiles and certain nonvolatiles (waxes)	Flower and plant concretes (rose, jasmine, tuberose, jonquil)
Extraction of concretes by ethanol, concentration	Extractor[a], filter, evaporator	Solubility difference in ethanol between volatiles and waxes	Flower and plant absolute (rose, jasmine, tuberose, jonquil)
Extraction of plant exudates by methanol, ethanol, or toluene		Solubility	Oleoresins such as pepper, ginger, vanilla, copaiba balsam, oakmoss resin, galbanum resinoid
Extraction by supercritical carbon dioxide	High-pressure equipment	Specific solubility in supercritical carbon dioxide of compounds as a function of temperature and pressure	Various spices; hop oil
Liquid–liquid extraction	Countercurrent solvent extraction column; solvent system: pentane/ethanol–water	Different partition coefficient of terpene hydrocarbons and oxygenated compounds (flavors) in the two solvent phases	Deterpenized citrus oils
Pressing, filtration, and chromatography			
Pressing	Various types of presses	Mechanical destruction of cells and oil glands	Juices, cold-pressed (expressed) citrus peel oils
Mechanical filtration	Sieve	Mechanical separation according to particle size	Removal of solids from liquids (→ clear juices) and/or collection of crystalline compounds
Ultrafiltration	Membrane filters (pore size 1–100 nm)	Separation according to size of molecules, and/or molecular aggregates	Separation of proteins, cell material from low-molecular-weight chemicals, fruit juice clarification

Continued

Table 1 Continued

Technique	*Equipment*	*Physical property utilized*	*Application to the production of*
Gas chromatography	Automated, preparative gas chromatograph	Differences in colligative properties	Isolation and purification of expensive volatile flavor compounds
High-performance liquid chromatography (HPLC)	Preparative HPLC apparatus	Differences in polarity and/or partition coefficient	Isolation and purification of expensive nonvolatile flavor compounds
Column chromatography	Large columns filled with a stationary phase, various fluids	Polarity, partition coefficient, charge, etc.	Mainly for expensive flavors from aqueous reactions/ waste streams

[a]Sometimes equipped with an ultrasonic sound vibrator.

Figure 7 (see color plate 57) Flavor compounds: highly efficient distillation columns. Reproduced from Flavour Compounds, Production Methods, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Figure 8 (see color plate 64) Filtration line for 450 hl output per hour; consisting of FILTROstar candle-type precoat filter and PVPP (polyvinyl pyrrolidone, a filter aid dispersed into a cellulose-based matrix) stabilizing plant Filter-o-mat S. Reproduced from Flavour Compounds, Production Methods, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

volatile substances have so far been detected in common foods. Some 6510 substances are reported in a single compilation alone. However, not all have sufficiently desirable flavoring properties for use in nature-identical flavors. In countries such as the USA, only flavoring materials which are 'generally recognized as safe' (GRAS) by qualified experts may be used. Currently, there are over 2000 flavoring materials on the GRAS lists published by the US Food and Drug Administration and by the Flavor and Extract Manufacturers' Association of the USA. Of these, some 1200 are nature-identical flavoring substances, about 400 are extracts of botanicals, and 400 are artificial (never identified in nature, such as ethyl vanillin, ethyl maltol). (*See* **Essential Oils: Properties and Uses**; **Legislation: International Standards; Codex.**)

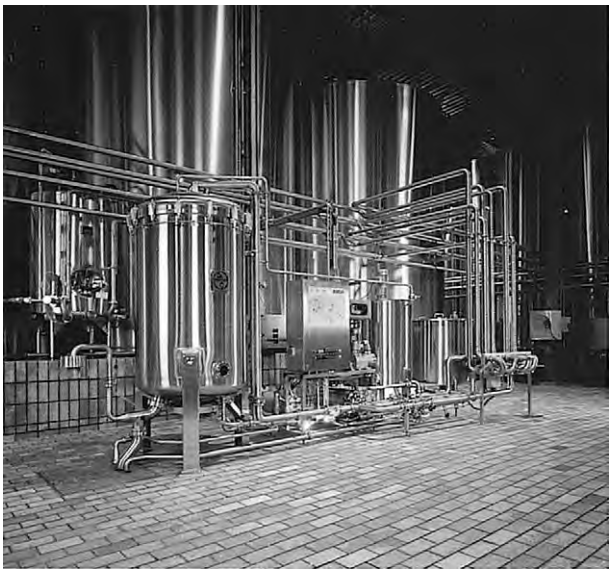

Figure 9 (see color plate 65) Kieselgur filtration plant with 35 m^2 filtering area and slurry feeder DOSIMAT 500. Special execution with automatic filtration monitoring. Reproduced from Flavour Compounds, Production Methods, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Synthesis of Organic Compounds as Flavorings

Controlled Chemical Synthesis

The merits of chemical synthesis for the production of flavor compounds were decisive for the development of today's high-quality flavors. Synthetic flavorings excel because of their constant quality, favorable price/performance ratio, unlimited supply, and toxicological integrity. Chemical synthesis is used to produce compounds in quantities from a few hundred grams up to hundreds of tons. Some examples of typical commodity flavor chemicals are listed in Table 2. They are commonly manufactured by large chemical companies, some of which do not produce and sell compounded flavors themselves. About 1000 small-scale items (with estimated consumptions of 0–500 kg $year^{-1}$) making up ~1.1% by weight of the totally consumed synthetic flavor compounds are mostly produced by the flavor industry itself, some even on laboratory and pilot plant scales. These often strong-smelling items are not profitable *per se*, but are absolutely essential for the creation of good flavors.

Common synthetic methods and reagents are used to transform either petrochemicals or plant-derived intermediates into flavorings. However, toxic reagents or intermediates are excluded. Rigorous control at all stages of manufacture insures that only chemicals of the highest quality are produced. Indeed, process and quality control is comparable to that used in the pharmaceutical industry.

Table 2 Typical commodity flavor chemicals

Chemical	*Estimated annual consumption (tons)*
Vanillin	7000
l-Menthol	9000[a]
Methyl salicylate	1800
Benzaldehyde	500–700
Maltol and ethyl maltol	400–600
Acetaldehyde	400–500
Cinnamaldehyde	400–500

[a]About 20% of this quantity is produced by synthesis.

Transformations Using Living Biological Systems

Microorganisms (bacteria, yeasts, and molds) play a very large role in the production of flavors in prepared foods. Yeasts are responsible for the production of flavor compounds in wine, beer, and bread. Cheese and yogurt flavors are produced by bacteria and molds. Such microorganisms have been exploited in the production of either complete flavors or specific flavor compounds, and the processes are generally referred to as fermentations (Table 3). (*See* **Fermented Foods: Origins and Applications.**)

Plant tissue cultures in bioreactors are being extensively investigated on a research scale as a means of obtaining a variety of flavor chemicals, e.g., citronellal from *Eucalyptus citriodora* roots or mint oil from *Mentha piperita*. However, because of low productivity and differences in the type of compounds produced (as compared to the whole plant), this technology is still far from commercial application. (*See* **Biotechnology in Food Production.**)

Synthesis Involving Isolated Enzyme Preparations

Enzymes are protein catalysts which both perform and control the chemical reactions that occur in all living systems. They are highly selective and can sometimes be used in an isolated or partially purified form to produce flavor compounds from suitable precursors. Many enzyme preparations are commercially available and are used by the food and flavor industries. Table 4 gives some examples with important applications. (*See* **Enzymes: Uses in Food Processing.**)

The Reaction Flavor Approach

Thermal processing of food, such as cooking, broiling, roasting, or baking, has been used since ancient times and is known to give rise to a variety of flavors. These flavor tonalities can also be produced as such without their food systems by the

Table 3 Microorganisms used in biotransformations/biosynthesis

Organism	*Reaction type*	*Precursors and/or carbon source*	*Flavor compound*	*Sensory description*
Bacteria				
Acetobacter spp.	Oxidation	Ethanol	Acetic acid	Vinegar
Corynebacterium spp.		Carbohydrate	Monosodium glutamate	Taste enhancer
Clostridium butyricum		Glucose	Butyric acid	Cheese, dairy
Yeasts				
Candida lipolytica or Sporobolomyces odorus	Oxidative degradation	Ricinoleic acid	γ-Decalactone	Peach
Molds				
Penicillium roquefortii	Hydrolysis, β-oxidation, and decarboxylation	Fatty acid triglycerides	Methyl alkyl ketones	Blue cheese
Ceratocystis moniliformis		Dextrose and urea	Esters	Fruity
Aspergillus niger		Sugar cane	Citric acid	Acidic

Table 4 Commercially available enzyme preparations with their areas of utilization in the flavor industry

Enzyme(s)	*Reaction type*	*Applications*
α-Amylase + glucoamylase	Hydrolysis	Corn syrup (dextrose) from corn starch
D-Glucose isomerase	Isomerization	Fructose-rich sugars (invert sugars) from corn syrup (US production in 2001: ~23.8 million pounds)
Lipases (esterases)	Hydrolysis of esters	Lipolyzed butter fat, free fatty acids from triglycerides, and synthesis of esters from alcohols and carboxylic acids
Proteases	Hydrolysis of proteins and peptides	Hydrolyzed vegetable and animal proteins
		Enzyme-modified cheeses and cheese flavors (via degradation of caseins), debittering of cheese flavors
	Autolysis of yeast	Yeast autolysates
Pectinase	Hydrolysis	Clarification of fruit juices

[a]Data from Corn Refiners Association, Inc., www.com.org; shipment by members, US, updated July 2, 2002.

reaction flavor approach, which involves the carefully controlled heating of a mixture of appropriate precursors in a reactor vessel. Typical flavor tonalities are roasted and broiled beef and chicken, bread, cereal, cocoa, caramel, maple, and vegetables. Chemically speaking, carbohydrates (in particular, their reducing monomers), proteins, free amino acids, and lipids are thermally degraded and transformed into numerous oxygen-, nitrogen- and sulfur-containing five- and six-membered heterocyclic flavor compounds together with α-diketones and aldehydes. More specifically, reaction flavors produced by heating reducing sugars and amino acids (resulting from protein hydrolysis) in addition to various other ingredients such as nucleotides, phospholipids, and fats are called Maillard-type flavors, the annual commercial value of which has been estimated at some US$780 million. The type of flavor thus obtained can be controlled by the nature and ratio of amino acids and sugars which are heated together; for example, cysteine and methionine are essential for obtaining meat flavors whilst proline plays an important role in the formation of bread and cereal flavors. (*See* **Browning**: Nonenzymatic.)

Flavor Delivery Systems

Since flavors are never being consumed as such but have always to be integrated into a food, the best way of delivering flavor is of prime importance to the flavor and food industry. Therefore, various flavor delivery systems have been developed over time. They usually consist of a taste-neutral carrier which combines with a flavor, often stabilizes it, may facilitate its transport and handling, and ultimately provides a better release of the desired amount, at the right place and time. They are sold as powder or liquid flavor systems in addition to conventional flavors.

Some flavor delivery systems are concerned primarily with transport and handling issues and give little protection to the flavor ingredients. More recent encapsulation processes have been designed to extend shelf-life in addition to giving functional flavor delivery in the finished food.

Encapsulation processes to make powder flavor systems trap liquid flavors in a solid carrier matrix. The most important encapsulation technologies are spray drying, extrusion, molecular inclusion, and

coacervation. All these processes lead to products which differ from each other in particle type and size, flavor load, chemical stability, hygroscopicity, and flavor release characteristics. Furthermore, there are notable differences in the process economics. Spray drying is generally considered as the most advantageous encapsulation method, coacervation is expensive, and molecular inclusion technology using β-cyclodextrins has recently been introduced into the marketplace as a result of a cost-lowering of β-cyclodextrin, and its acceptance in the USA as a food additive. Encapsulated flavors are also made by extruding a molten (110–130 °C) low-moisture blend of carbohydrates, which contain a dispersed flavor, into a cooled (−18 °C) solvent. The melt is hardened by the rapid chilling into amorphous glass-like particulates, which provide much better oxidative stability than that obtained by spray drying. Frequently used carriers for encapsulated powder flavors are modified or hydrolyzed starches (e.g., maltodextrin), gum arabic, proteins, and alginates.

Another current way of entrapping and delivering flavor is via liquid, stable emulsions. An oily flavor/water mixture in the presence of an emulsifier is passed through a high-pressure homogenizer. Often the density of the oil phase is brought closer to that of the aqueous phase by the addition of a high-density weighting agent such as sucrose acetate iso-butyrate. (*See* **Extrusion Cooking**: Principles and Practice; **Drying**: Spray Drying.)

Storage, Packaging, and Distribution of Flavors

The storage, packaging, and distribution of flavors requires consideration of numerous issues, including flavor stability during storage, rules on acceptable food packaging materials, regulations on the storage and transport of flammable goods, rules on the avoidance of certain materials (e.g., glass) in specific industries, environmental concerns and worker safety, not to mention the special requirements of food manufactures. (*See* **Sensory Evaluation**: Taste; **Spoilage**: Chemical and Enzymatic Spoilage; **Storage Stability**: Parameters Affecting Storage Stability.)

See also: **Biotechnology in Food Production**; **Browning**: Nonenzymatic; **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **Enzymes**: Uses in Food Processing; **Essential Oils**: Properties and Uses; **Fermented Foods**: Origins and Applications; **Herbs**: Herbs and Their Uses; **Legislation**: International Standards; Codex; **Membrane Techniques**: Principles of Ultrafiltration; **Sensory Evaluation**: Taste; **Spoilage**: Chemical and Enzymatic Spoilage; **Storage Stability**: Parameters Affecting Storage Stability

Further Reading

Ashurst PR (ed.) (1995) *Food Flavourings*, 2nd edn. London: Blackie Academic and Professional Press.

Belitz H-D and Grosch W (1999) *Food Chemistry*, 2nd edn. Berlin: Springer.

Contis ET, Ho C-T, Mussinan CJ, Parliment TH, Shahidi F and Spanier AM (eds) (1998) *Food Flavors: Formation, Analysis and Packaging Influences*. Amsterdam: Elsevier.

Fisher C and Scott TR (1997) *Food Flavours – Biology and Chemistry*. London: Royal Society of Chemistry.

Friberg SE and Larsson K (eds) (1997) *Food Emulsions*, 3rd edn. New York: Marcel Dekker.

Gabelman A (ed.) (1994) *Bioprocess Production of Flavor, Fragrance, and Color Ingredients*. New York: John Wiley.

Maarse H and van der Heij DG (eds) (1994) *Trends in Flavour Research, Developments in Food Science*, p. 35. *Proceedings of the 7th Weurman Flavour Research Symposium*. Amsterdam: Elsevier.

Marsili R (ed.) (1997) *Techniques for Analyzing Food Aroma*. New York: Marcel Dekker.

Nijssen LM, Visscher CA, Maarse H, Willemsens LC and Boelens MH (eds) (1996) *Volatile Compounds in Food*, 7th edn. Zeist: TNO Nutrition and Food Research Institute.

Noble RD and Stern SA (1995) *Membrane Separation Technology: Principles and Applications*. Amsterdam: Elsevier.

O'Brien J, Nursten HE, Crabbe MJC and Ames JM (eds) (1998) *The Maillard Reaction in Foods and Medicine*. Cambridge: Royal Society of Chemistry.

Parliment TH, Morello MJ and McGorrin RJ (eds) (1994) *Thermally Generated Flavors: Maillard, Microwave, and Extrusion Processes*, Washington, DC: American Chemical Society.

Perry RH, Green DW and Maloney JO (eds) (1997) *Perry's Chemical Engineering Handbook*, 7th edn. New York: McGraw-Hill.

Reineccius G (ed.) (1994) *Source Book of Flavors*, 2nd edn. New York: Chapman & Hall.

Risch SJ and Reineccius GA (1995) *Encapsulation and Controlled Release of Food Ingredients*. Washington, DC: American Chemical Society.

Takeoka GR, Teranishi R, Williams PJ and Kobayashi A (eds) (1996) *Biotechnology for Improved Foods and Flavors*. Washington, DC: American Chemical Society.

Ziegler E and Ziegler H (eds) (1998) *Flavourings: Production, Composition, Applications, Regulations*. Weinheim: Wiley-VCH.

FLAXSEED

D H Morris, Mainstream Nutrition, Toronto, Ontario, Canada
M Vaisey-Genser, The University of Manitoba, Winnipeg, Manitoba, Canada

Background

Flaxseed was one of the first plants domesticated by humans about 8000–10 000 years ago. Ancient peoples valued flaxseed for its medicinal properties; its oil, which was used in cooking and in formulating cosmetics; and its fiber, which was used to make linen and sail cloth. The botanical name of flaxseed, *Linum usitatissimum* of the family Linaceae, reflects its versatility and importance in the economic and social development of humans: The word *usitatissimum* is Latin for 'most useful.' Flaxseed, or linseed as it is sometimes called, is valued for many of the same reasons today. This chapter describes the composition of flaxseed and its health benefits, performance properties, safety, and recommended intakes.

Composition

Flaxseed is rich in fat, dietary fiber, and protein. The composition of Canadian-grown flaxseed averages 41% fat, 28% total dietary fiber, 20% protein, 7.7% moisture, 3.5% ash, and 1% simple sugars. The amount of fat in flaxseed can range from 38 to 47%, depending on the variety and growing conditions. The proximate composition for common measures of whole and ground flaxseed and flaxseed oil is shown in Table 1.

Flaxseed has a unique fatty acid profile, being fairly low in saturated fatty acids and rich in α-linolenic acid (ALA), the essential ω-3 fatty acid (see Figure 1). Of the total fatty acids in flaxseed, saturated fatty acids constitute 9%; monounsaturated fatty acids, 18%; and polyunsaturated fatty acids, 73%. Of the polyunsaturated fatty acids, ALA constitutes the majority at 57% of total fatty acids, making flaxseed one of the richest sources of this fatty acid. Linoleic acid, the essential ω-6 fatty acid, is present in a smaller amount (16%). Because of its high ALA content, flaxseed has an ω-6/ω-3 fatty acid ratio of 0.3:1. (*See* **Fatty Acids**: Properties; Metabolism.)

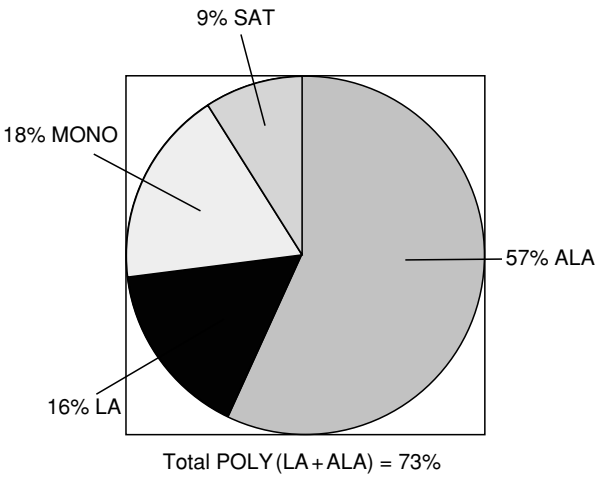

Figure 1 Fatty acid composition of flaxseed oil, based on an analysis conducted by the Canadian Grain Commission, May 2001. Abbreviations: ALA, α-linolenic acid; LA, linoleic acid; MONO, monounsaturated fatty acids; POLY, polyunsaturated fatty acids; SAT, saturated fatty acids.

Table 1 Proximate composition of flaxseed based on common measures[a]

Form of flaxseed	*Weight g*	*Common measure*	*Energy*		*Total fat g*	*ALA g*	*Protein g*	*Total carbohydrate[b] g*	*Total dietary fiber[c] g*
			kcal	*kJ*					
Whole seed	100		450	1890	41	23	20	29	28
	180	1 cup	810	3402	74	41	36	52	50
	11	1 tbsp	50	210	4.5	2.5	2.2	3	3
	4	1 tsp	18	76	1.6	0.9	0.8	1.2	1.1
Ground seed	130	1 cup	585	2457	53	30	26	38	36
	8 g	1 tbsp	36	151	3.3	1.8	1.6	2.3	2.2
	2.7	1 tsp	12	50	1.1	0.6	0.5	0.8	0.8
Flaxseed oil	100		884	3712	100	57			
	14	1 tbsp	124	520	14	8.0			
	5	1 tsp	44	185	5	2.8			

[a]Used with permission of the Flax Council of Canada. Based on a proximate analysis conducted by the Canadian Grain Commission, May 2001. The fat content was determined using the American Oil Chemists' Society (AOCS) Official Method Am 2–93. The moisture content was 7.7%.
[b]Total carbohydrate includes available carbohydrate (1 g) and total dietary fiber (28 g) per 100 g of flaxseed.
[c]Total dietary fiber includes insoluble fiber (19 g) and soluble fiber (9 g) per 100 g of flaxseed.

Flaxseed provides about 28 g of total dietary fiber per 100-g dry weight. (Dietary fiber is defined as the edible parts of plants that are resistant to digestion and absorption in the human small intestine but that are completely or partially fermented in the large intestine.) About one-third of flaxseed fiber is water-soluble, consisting mainly of mucilage gum. The remaining two-thirds is water-insoluble and consists primarily of nonstarch polysaccharides such as cellulose and lignin. The flaxseed lignan, secoisolariciresinol diglycoside (SDG), is a phenolic compound related in structure to lignin.

The amino acid profile of flaxseed protein is similar to that of soybean protein. Flaxseed proteins are albumins and globulins. The predominant globulins, like linin, are less soluble and higher in molecular weight than the albumins. Flaxseed also contains minor amounts of essential vitamins like niacin, folic acid and γ-tocopherol, and minerals like potassium and phosphorus.

Health Benefits

Flaxseed has been called a functional food, because of the health effects associated with its ALA and lignan content. Functional foods have been defined as those that offer health benefits in addition to what would be expected from the food's nutrient content. Because of its wide-ranging health effects, flaxseed may help in the clinical management of chronic conditions such as cardiovascular disease (CVD), stroke, certain types of cancer, and autoimmune diseases like systemic lupus erythematosus.

Laxation

The insoluble dietary fiber in flaxseed is likely responsible for its laxative effect in young and elderly adults. For instance, in a study of 10 healthy young adults, bowel movements increased 30% during the 4-week period when they ate two 25-g flaxseed muffins daily compared with the control period, during which time they ate muffins without flaxseed.

Cardiovascular Diseases

The underlying cause of CVD is atherosclerosis. One risk factor for CVD is elevated blood total cholesterol concentrations. The results of clinical studies indicate that flaxseed has a positive effect on blood lipids. In one study of 10 healthy young men and women, plasma total cholesterol was reduced 6% and low-density-lipoprotein (LDL) cholesterol was reduced 9% following the consumption of flaxseed muffins providing 50 g of ground flaxseed daily for 4 weeks. Plasma high-density-lipoprotein (HDL) cholesterol and triacylglycerol levels were unchanged during the flaxseed supplementation period. In a study of 15 men and women with hyperlipidemia who had just completed a trial of the effect of vitamin E supplementation on serum lipids, adding 15 g of ground flaxseed to their daily diet reduced total blood cholesterol 7% and LDL cholesterol 11% without changing HDL cholesterol levels. (*See* **Cholesterol: Role of Cholesterol in Heart Disease.**)

The reductions in total cholesterol and LDL cholesterol achieved in these studies were likely due to flaxseed's dietary fiber and ALA content. Dietary fiber is widely recognized as an agent that decreases serum cholesterol levels and slows the development of atherosclerotic lesions. Populations with high intakes of dietary fiber have a lower risk of CVD and other chronic diseases than populations with low fiber intakes. ALA also affects the cardiovascular system. In dogs, pure, infused preparations of ALA and the long-chain ω-3 fatty acids, eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), protected against fatal arrhythmias. Although the mechanism of action is not completely understood, ALA and the other ω-3 fatty acids are believed to regulate membrane ion channels and heart muscle fiber excitability, thus stabilizing the electrical signals of the heart and helping prevent ventricular fibrillation and arrhythmias.

Although there is no direct evidence that dietary ω-3 fatty acids are as effective in humans as pure fatty acid preparations are in dogs, indirect evidence for a dietary benefit of ALA is available from epidemiologic studies of diet and CVD risk. For example, researchers in the Lyon Diet Heart Study hypothesized that ALA may have antiarrhythmic actions that significantly reduce fatal CVD events. They recruited subjects who had survived a myocardial infarction. Those subjects who consumed a typical Mediterranean-style diet rich in ALA from many dietary sources had a nearly 70% reduction in fatal cardiac deaths compared with the group who consumed their usual Western-type diets. An extended follow-up of the study revealed that the protective effect of the Mediterranean diet was preserved – ALA, but not EPA and DHA, was associated with protection against recurrence of myocardial infarction. Data from other epidemiologic studies, including the Health Professional Follow-up Study, the Multiple Risk Factor Intervention Trial, and the Nurses' Health Study, suggest that diets high in ALA are associated with a reduced risk of fatal and nonfatal myocardial infarction, fatal CVD, and death from all causes.

ALA also affects the process of atherosclerosis by reducing inflammatory responses. Atherosclerosis is an inflammatory disease that begins in infancy and childhood when the earliest lesions, called fatty

streaks, begin developing in arteries. Flaxseed may influence the progression of atherosclerosis by inhibiting the production of major systemic components of inflammatory activity, including eicosanoids, cytokines, and platelet-activating factor.

Cancer

Flaxseed is a rich source of the lignan and phytoestrogen, SDG, which is a precursor of the mammalian lignans, enterolactone and enterodiol. Phytoestrogens are plant-derived compounds that can interfere with estrogen metabolism in animals and humans. Lignans may protect against certain types of cancer, particularly the hormone-sensitive cancers such as those of the breast, endometrium, and prostate, by interfering with the metabolism of sex hormones. Flaxseed lignans enhance the clearance of circulating estrogen, and they inhibit the binding of sex hormones to cell membranes by blocking cell-membrane receptors. In rat mammary gland tissue, flaxseed lignans decreased cell proliferation, the number of nuclear aberrations, and tumor volume. Preliminary data from one human study suggest a possible benefit of flaxseed consumption by women with breast cancer. Overall, flaxseed lignans, pure SDG, and flaxseed oil rich in ALA affect the carcinogenic process at several points, including initiation, promotion, and progression/growth of tumors. (*See* **Cancer**: Diet in Cancer Prevention.)

Diabetes and Autoimmune Diseases

Flaxseed affects the postprandial blood glucose response in healthy young adults. In a study involving nine young women, blood glucose decreased about 27% after two test meals were consumed. Studies of the hypoglycemic effects of flaxseed among patients with diabetes mellitus are under way. (*See* **Diabetes Mellitus**: Etiology.)

Flaxseed may be useful in the management of autoimmune diseases like rheumatoid arthritis and systemic lupus erythematosus through the actions of ALA on proinflammatory eicosanoids and platelet-activating factor (PAF). The one published study of flaxseed oil's effects on the clinical and subjective assessment of symptoms associated with rheumatoid arthritis found no significant differences between outcomes in patients who consumed flaxseed oil compared with safflower oil for 3 months. However, in studies of normal, healthy volunteers, flaxseed reduced significantly the production of proinflammatory eicosanoids that contribute to inflammation and other immune responses. Dietary flaxseed also inhibits PAF, a powerful mediator of the immune response. In studies involving patients with lupus nephritis, a condition marked by chronic inflammation of the kidney, flaxseed consumption reduced significantly PAF production and kidney inflammation and improved kidney function. Thus, a beneficial role for flaxseed in managing autoimmune diseases has been proposed, but more research is needed to confirm its effects.

Other Health Benefits

Reports that flaxseed aids in treating the symptoms of depression, irritable bowel syndrome, osteoporosis, and menopause are largely anecdotal. Future research will determine the effects of flaxseed's components, particularly the lignans and ALA, on the pathology and treatment of these conditions.

Performance Properties

Market Forms

Flaxseed is marketed primarily as the intact whole seed, which is reddish brown, flat, and oval with a pointed tip. The seed measures about $2.5 \times 5.0 \times 1.5$ mm, slightly larger than a sesame seed. Flaxseed is also marketed in a coarsely ground form. Whole flaxseed may be ground successfully in a home coffee grinder.

Flaxseed oil is available in health food stores either in bottles for use with foods such as salads or in capsules to be used as diet supplements. Oil for the health food market is produced by 'cold-pressing' flaxseed where the maximum allowable temperature is 35 °C. The seed is cleaned, cracked, flaked between rollers, and pressed in expellers filled with water-cooled shafts. It is normally bottled in light-proof containers to avoid photooxidation and refrigerated to limit autoxidation.

Sensory Characteristics

Appearance, texture, and taste are the sensory features that benefit from adding flaxseed to foods. The reddish brown seed provides an attractive color contrast in products like yeast breads and muffins, in cereals such as cream of wheat or corn meal/grits, and in pancakes and potato latkes. Whole flaxseed provides crunchiness and chewiness, textural features that are less pronounced in the ground form. Experienced sensory panelists have described the taste of flaxseed as nutty and cereal-like with a slightly fruity/bitter tang.

Food Uses

Flaxseed has been used in food over the centuries, particularly in the Middle East and North Africa. In Ethiopia, flaxseed is a key ingredient in porridge and, when roasted, crushed and combined with split peas and red pepper, in 'w'et,' a form of stew. There it is also enjoyed as 'ch'lk'a,' a drink made from lightly

roasted ground flaxseed, salt, water, and honey. In China, oil from cold-pressed flaxseed has long been used for stir-frying. In North America, flaxseed continues to be an ingredient of Red River Cereal®, a breakfast staple introduced in 1924.

Acceptability studies show that baked products such as yeast breads, muffins, and cookies containing flaxseed are increasingly well liked by North American consumers. This is in keeping with the growing predominance of whole wheat and multigrain breads in preference to white bread in several regions. Experienced sensory panelists give flaxseed and wholewheat breads comparable ratings in crust color, texture, tenderness, porosity, and loaf shape.

During the 1990's, a 'functional' bread, Burgen® Soy-Lin™, was developed and commercialized in Australia as one outcome from a study of the effects of phytoestrogens on the symptoms of menopause. Four slices of this 10% fiber bread daily contributed 2.5 g of ALA and contained enough soy flour and flaxseed to supply a significant amount of phytoestrogens (220 mg) as isoflavones from soy together with the lignan precursors from flaxseed. By April 1997, Burgen® Soy-Lin™ bread had captured 5.3% of the total Australian bread market and was proving successful in the UK.

Recipe Modifications

Most wheat-flour recipes including yeast breads and quick breads such as muffins, tea biscuits and cookies can comfortably accommodate a 10% replacement of wheat flour by flaxseed. However, when using flaxseed in food products, it is important to recognize two factors. First, flaxseed is an oilseed containing a generous amount of fat, and second, flaxseed is hydrophilic or water-loving.

When substituting flaxseed for wheat flour in baked goods, it is prudent to reduce the amount of fat called for in the recipe by the amount contained in the flaxseed. A useful 'rule of thumb' is the fact that 3 tbsp (1 tbsp = 15 ml) of whole flaxseed, which is 41% oil, replaces 1 tbsp of salad oil. However, 3 tbsp of ground flaxseed will replace only 2/3 tbsp (2 tsp) of salad oil. This is because flaxseed is aerated during grinding so that ground flaxseed weighs less per unit of volume, as shown in Table 1. Accordingly, it replaces a little less fat than an equal amount of whole flaxseed.

The nature of the fat being substituted by the oil in flaxseed may also influence the amount replaced. Hydrogenated shortenings, like salad oils, are 100% fat. However, because they are aerated to make them easier to cream in cake and cookie production, they contain about 10% less fat per unit volume than salad oils. The standard tablespreads butter and margarine are 80% fat with an allowable 16% water, so their fat content per unit volume is less than either hydrogenated shortening or salad oil. Thus, added flaxseed with its 74 g of oil per cup (1 cup = 240 ml) (Table 1) theoretically replaces the volume of a fat type as specified in a standard recipe such that flaxseed would replace a greater volume of butter or margarine than it would of hydrogenated shortening or salad oil. Nevertheless, the balance of ingredients in a sweetened, shortened baked product can tolerate as much as a 25% increase or decrease in an ingredient without making the product fail.

Considering the affinity of flaxseed for water, bakers recommend that whole flaxseed for yeast breads be conditioned with a water presoak for at least 30 min at 18 °C using a 2:1 ratio of water to flaxseed. The soaking water, which contains some of the soluble gums, then serves as the liquid in the recipe. When ground flaxseed is used, water should be added to approximate 75% of the flaxseed weight. Some authors suggest that the yeast used be increased by 25% if 10–15% of whole flaxseed is replaced by ground flaxseed.

Storage Stability

The usefulness of flaxseed as a staple food ingredient is reinforced by the fact that it appears remarkably stable to storage. In one study, flaxseed from a mixture of several varieties grown in Canada in 1996 was ground with a pin-mill and packed in 60-lb (27.3-kg) triple-wrapped plastic-lined paper bags. Samples were stored for 128 days at ambient temperatures under conditions similar to those used by commercial bakeries. The mixed flaxseed contained 43% oil consisting of 60% 18:3*n*-3 (ALA), 16% 18:2*n*-6, and 15% 18:1*n*-9. Table 2 shows that, despite its substantial content of unsaturated fatty acids, ground flaxseed showed little or no evidence of oxidative deterioration during 4 months of storage at room temperature. Peroxide values, free fatty acids, and conjugated double bonds did not change appreciably. These chemical results were verified by the lack of difference in odor intensity measured by a trained sensory panel that examined water/ground flaxseed slurries from all storage intervals concurrently. Furthermore, a consumer panel of 36 volunteers, using a duo–trio test, could not detect any difference between 10% flaxseed yeast breads prepared from ground seed that had been stored at either 0 or 128 days.

While volatile compounds increased during the storage of ground flaxseed, the level of total volatiles after 4 months' storage was 10–25 times less than values reported for stored vegetable oils that contained appreciably less oxidation-susceptible ALA.

Table 2 Chemical and sensory indices of the storage stability of ground mixed flaxseed[a]

Parameter	*Storage day at 23 + 2 °C*				
	0	*33*	*66*	*96*	*128*
Peroxide value (meq kg^{-1})	0.20	0.21	0.12	0.16	0.12
Free fatty acids (%)	0.34	0.48	0.31	0.35	0.31
Conjugated double bonds (absorbance at 268 nm)	0.20	0.19	0.19	0.19	0.19
Odor intensity[b]	3.1	3.6	3.2	3.2	3.5

[a]Adapted from Malcolmson LJ, Przybylski R and Daun JK (2000) Storage stability of milled flaxseed. *Journal of the American Oil Chemists' Society* 77: 235–238.
[b]1, none; 5, intense; nine-member trained panel; two replications.

Increases over time were least in the dienals, which is usually one of the major subclasses resulting from polyunsaturated fatty acid oxidation. Among individual compounds, hexanal, which is often a quality indicator, was found only at the p.p.b. level, marginally close to reported sensory thresholds.

What accounts for the oxidative stability of flaxseed remains a question. In a sequel to the above study, the fatty acid composition and the tocopherols, known antioxidants, in ground mixed flaxseed were examined after 11 and 20 months' storage at ambient temperatures. ALA levels were within the normal variability range after 20 months' storage, and the tocopherols remained essentially unchanged. The authors suggested that they were either not used or in some way regenerated.

Baking Stability

The ALA in flaxseed, whole or ground, appears to be reasonably stable to baking temperatures. For example, the proportion of ALA in a muffin mix where flaxseed made up 28.5% of the formula remained unchanged after baking for 2 h at 178 °C, despite the fact that more oxygen was consumed by flaxseed muffins during baking than those which were flaxseed-free. Biological evidence also supports the baking stability of ALA. No difference was found between the plasma fatty acid profiles of college women eating 50 g of raw, ground flaxseed added to cereal, soup, juice or yogurt daily for 4 weeks and those eating the same amount baked in bread. Both groups showed a lowering of serum total cholesterol and LDL cholesterol.

Several studies suggest that the lignan precursor SDG is stable to baking. For example, SDG levels in the center of flaxseed breads proved similar to those in the crust, even though the latter had a greater heat exposure. Lignan production, measured by *in vitro* fermentation, reflected the amounts of raw flaxseed added in both commercial and home-baked products (baked at 190 °C) and pancakes (205 °C), supporting stability to baking. Furthermore, bioavailability was confirmed in this series of studies when nine women, for periods of 1–8 days, ate 5, 15, or 25 g of flaxseed either added raw to applesauce or baked in muffins or bread. While urinary lignan excretion increased as the amounts of flaxseed eaten increased and with repeated flaxseed consumption, there was no significant difference in urinary lignan excretion due to baking conditions.

Safety

Flaxseed contains two substances that may interfere with the absorption of certain nutrients, although both substances are present in flaxseed at levels that appear to be of little consequence in the human diet. Linatine inhibits the utilization of pyridoxine (vitamin B_6). However, when 30 adults consumed 45 g of flaxseed daily in muffins for 5 weeks, serum pyridoxine levels did not change significantly from baseline. The other substance, phytic acid, is known to form insoluble complexes with calcium, copper, iron, magnesium, and zinc in the intestine, thus reducing their absorption. There is no evidence that this occurs to any appreciable extent in humans.

Flaxseed also contains cyanogenic glycosides, which are a group of natural substances found in more than 2500 plant species that can release cyanide, a poisonous compound, under certain conditions. Several agronomically important crops such as cassava, sorghum, flaxseed, and stone fruits like peaches, plums, cherries, and apricots contain the glycosides. There is no evidence that the cyanogenic glycosides pose a health problem for healthy, nourished populations in developed countries. Baking and other processing methods apparently destroy or volatize the cyanogenic glycosides. Indeed, muffins containing 25 g of ground flaxseed had no detectable levels of the glycosides after baking for 15–18 min at 230 °C.

Recommended Intakes

There is growing concern that Western populations have an unbalanced intake of ω-3 fatty acids relative to the ω-6 fatty acids. The current dietary ω-6/ω-3

Table 3 Contribution of flaxseed to ω-3 and ω-6 fatty acid intakes and the ω-6/ω-3 fatty acid ratio of the North American diet[a]

Food Item	*ω-3 content of food[b] (g)*	*ω-6 content of food (g)*	*ω-6/ω-3 ratio of the diet*
Commonly eaten foods			
Breakfast (cereal, milk, orange juice)	0.106	0.333	
Lunch (roast beef sandwich with pickle and tomato slices plus one large apple)	0.285	2.187	
Snack (4 oz (112-g) of potato chips)	0.216	13.584	
Dinner (pizza, salad)	0.983	8.831	
Total	1.59	24.935	16:1
Adding flaxseed changes the ω-6/ω-3 ratio of the diet			
Flaxseed, whole, 1 tbsp	2.5	0.72	6:1
Flaxseed, ground, 1 tbsp	1.8	0.52	8:1
Flaxseed oil, 1 tbsp	8.0	2.24	3:1

[a]Used with permission of the Flax Council of Canada.
[b]The ω-3 and ω-6 fatty acid contents of foods were calculated using the US Department of Agriculture's Nutrient Database for Standard Reference, Release 13, available on the Internet at www.nal.usda.gov/fnic/foodcomp.

ratio ranges from about 10:1 to 25:1 in typical North American diets. Health Canada recommends a ratio of 4:1 to 10:1, whereas a joint Food and Agriculture Organization/World Health Organization committee recommends a ratio of between 5:1 and 10:1. Flaxseed consumption can increase total ω-3 fatty acid intakes and reduce the ω-6/ω-3 ratio, as shown in Table 3.

The US Institute of Medicine set a daily Adequate Intake of 1.1 g of ALA for women and 1.6 g of ALA for men. These recommended intakes can be achieved by consuming as little as 8 g (1 tbsp) of ground flaxseed or 2.5 g (1/2 tsp) of flaxseed oil per day. Up to 10% of the Adequate Intake can be derived from foods containing EPA and DHA.

In most clinical trials of the health effects of flaxseed, volunteers have consumed 5–50 g of ground flaxseed or 60 ml or flaxseed oil daily. Based on the findings of these published reports, an intake of about 8–16 g (1–2 tbsp) of ground flaxseed daily is recommended. Ground flaxseed is the preferred form because it is easily added to foods and supplies the full complement of nutrients associated with flaxseed's positive health effects. Flaxseed oil is rich in ALA but does not contain dietary fiber or lignans.

Summary

Flaxseed is a functional food valued for its pleasant, nutty taste, unique nutrient profile, and health benefits. Flaxseed is a rich source of ALA, the essential ω-3 fatty acid, and the lignan, SDG. It also provides protein and dietary fiber. Newer research suggests that flaxseed may help in the prevention and management of chronic diseases such as coronary heart disease, diabetes mellitus, certain types of cancer, and autoimmune diseases. Flaxseed's high ALA content can make a significant contribution to total ω-3 fatty acid intake. The lignan precursors and ALA in flaxseed appear to be stable to common storage and baking conditions. Consumers can purchase flaxseed as whole seeds, ground flaxseed, flaxseed oil, in capsule form and in baked goods, cereal mixes, and pasta products.

See also: **Analysis of Food**; **Antioxidants**: Natural Antioxidants; Role of Antioxidant Nutrients in Defense Systems; **Cholesterol**: Factors Determining Blood Cholesterol Levels; **Coronary Heart Disease**: Etiology and Risk Factor; Prevention; **Diabetes Mellitus**: Etiology; Treatment and Management; **Essential Fatty Acids**; **Fatty Acids**: Dietary Importance; **Functional Foods**; **Prostaglandins and Leukotrienes**; **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability

Further Reading

Clark WF, Parbtani A, Huff MW *et al.* (1995) Flaxseed: A potential treatment for lupus nephritis. *Kidney International* 48: 475–480.

Cunnane SC, Ganguli S, Menard C *et al.* (1993) High α-linolenic acid flaxseed (*Linum usitatissimum*). Some nutritional properties in humans. *British Journal of Nutrition* 69: 443–453.

Cunnane SC, Hamadeh MJ, Liede AC *et al.* (1995) Nutritional attributes of traditional flaxseed in healthy young adults. *American Journal of Clinical Nutrition* 61: 62–68.

de Lorgeril M, Renaud S, Mamelle N *et al.* (1994) Mediterranean alpha-linolenic acid-rich diet in secondary prevention of coronary heart disease. *Lancet* 343: 1454–1459.

Jones J (2000) Update on defining dietary fiber. *Cereal Foods World* 45: 219–220.

Malcolmson LJ, Przybylski R and Daun JK (2000) Storage stability of milled flaxseed. *Journal of the American Oil Chemists Society* 77: 235–238.

Payne TJ (2000) Promoting better health with flaxseed in bread. *Cereal Foods World* 45: 102–104.

Przybylski R and Daun JK (2001) Additional data on the storage stability of milled flaxseed. *Journal of the American Oil Chemists Society* 78: 105–106.

Rickard SE, Jenah M, Tou JCL, Nesbitt PD and Thompson LU (1998) Anticancer effects and availability of flaxseed lignans. *Proceedings of the Flax Institute* 57: 8–14.

Stewart S and Mazza G (2000) Effect of flaxseed gum on quality and stability of a model salad dressing. *Journal of Food Quality* 23: 373–390.

Thompson LU, Seidl MM, Richard SE *et al.* (1996) Antitumorigenic effect of a mammalian lignan precursor from flaxseed. *Nutrition and Cancer* 26: 159–165.

Vaisey-Genser M and Morris DH (1997) *Flaxseed – Health, Nutrition and Functionality.* Winnipeg, MB: Flax Council of Canada, also available on the Council's web site at www.flaxcouncil.ca.

FLOCCULATION

B M Baraniak, Agriculture Academy, Lublin, Poland
E Waleriańczyk, Instytut Przemyslu Cukrowniczego, Warszawaul, Poland

Background

Flocculation is a preliminary treatment of colloidal systems, employed to destabilize the sol structure by adding reagents that encourage particle growth, before separation of the solid phase is attempted by filtration or sedimentation. Particles of colloidal dimension (0.5×10^{-5} m) frequently form stable dispersions which do not settle and are not amenable to conventional filtration processes. (*See* **Filtration of Liquids.**)

Mechanism of Flocculation

Colloid Destabilization by Ions

The stability of a colloid is related to its zeta potential, which involves the charged surface of the particles and the oppositely charged layer of counter ions adjacent to it, constituting an electrical double layer. A reduction of zeta potential results in destabilization of the stable colloid structure, and may be caused either by the specific adsorption of counter ions – a nonspecific effect of neutral electrolytes – or by enmeshment of particles within a sweep floc. Since the charge of counter ions is important, it is apparent that polyvalent ions will be more effective than singly charged ions: Na(I) < Ca(II) < Al(III).

Colloid Destabilization by Polymers

A general mechanism for the destabilization of colloidal sols by high-molecular-weight synthetic polymers, termed polyelectrolytes, is described by the theory of bridging. The polyelectrolyte is adsorbed at several sites, not covering the whole area of the colloidal particles. As the polyelectrolyte molecules are relatively long, loops or chains protrude beyond the double layer. Thus, flocculation takes place through hydrophobic or hydrogen bonds, and any compression of the double layer that can be induced by increases in ionic strength should enhance bridging by allowing closer particle approach (**Figure 1**).

Polyelectrolyte adsorption may, in some cases, increase the stability of the colloid. A strongly adsorbed polymer may increase the diameter of the diffuse layer, and the larger double layer will generate repulsive

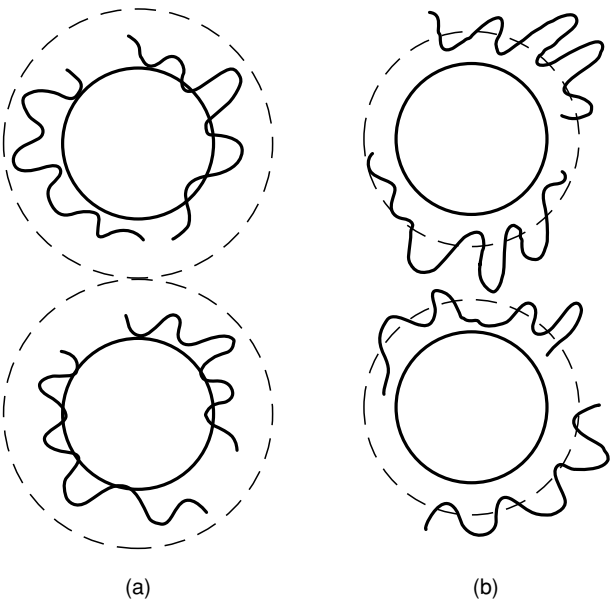

Figure 1 Bridging efficiency of polyelectrolytes: (a) low ionic strength; (b) high ionic strength. Reproduced from Flocculation, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

interreactions, leading to increased stability. The amount of polyelectrolyte applied in relation to the colloidal concentration is critical for destabilization or restabilization of colloid structure. Overdosing is a real danger, and the optimum concentration of polyelectrolyte required for flocculation should be very small (not in excess of $1\,mg\,l^{-1}$).

Aggregation Mechanism

The aggregation of the destabilized colloid structure is the second step in the final process of separation of the solid phase, and it is ultimately dependent on two factors: (1) frequency of collision between particles; (2) effectiveness of the interactions causing particle–particle adhesion. In general, there are two mechanisms by which adhesion contacts occur: (1) aggregation by thermally induced motions, termed 'perikinetic flocculation'; (2) contacts between particles in heterodisperse systems by differential settling under gravity, termed 'orthokinetic flocculation.'

In real systems, the initial stage of flocculation would be perikinetic, but with increasing floc growth the orthokinetic model becomes dominant.

Flocculants and Coagulants

Polyelectrolytes

Polyelectrolytes that are applied practically as flocculants are mainly water-soluble polyacrylamides, polyphosphates and modified natural polymers – gelatines, chitosan, and carrageenan, as well as starch and cellulose derivatives. Polyelectrolytes are classified as cationic, anionic, and nonionic according to the nature of the functional groups along the polymer chain. Predominant among these are polyacrylamide copolymers of acrylamide and acrylate, or monomers containing ammonium groups (Figure 2).

A polyelectrolyte is characterized by its molecular weight, the nature of the functional group, and the charge density. An important consideration in choosing a polyelectrolyte for a desired process is its potential as a coagulant (by destabilization of the colloid via neutralization) and as a flocculant (by interparticle bridging). The pH is also an important parameter to be considered when selecting a polyelectrolyte for a particular application. Sensitivity to pH occurs with cationic polymers in which quaternary ammonium groups are dominant, and with anionic polymers containing sulfonic acid groups. Flocculants with carboxyl or amine groups confer strong pH dependence. The toxicity of polyacrylamides is usually low,

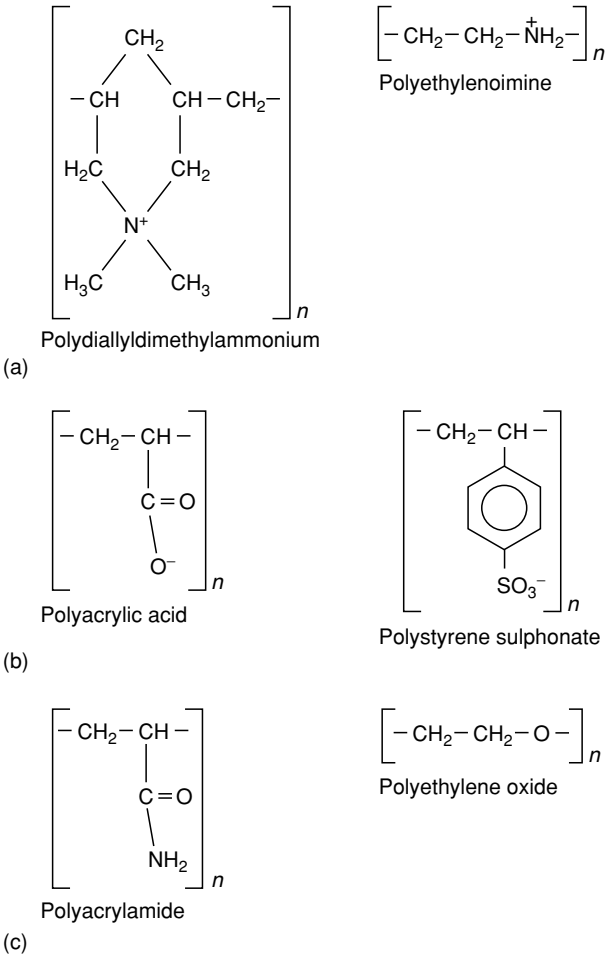

Figure 2 Characteristics of main groups of polymeric flocculants: (a) cationic polyelectrolytes; (b) anionic polyelectrolytes; (c) nonionic polymers. Reproduced from Flocculation, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

usually lower than 0.05%, and results mainly from the presence of free acrylamide.

Coagulants

The reagents most widely used in coagulation processes are polyvalent cations, such as Al(III) or Fe(III) or Fe(II), available as chlorides or sulfates in crystalline forms. Iron salts generally offer advantages over aluminum salts, and Fe(III) is also often superior to Fe(II) because it has a wider pH range of action.

Other Flocculants

Clay materials, such as bentonite or montmorillonite, activated carbon or activated silica are sometimes used to improve the filtration or sedimentation of suspended solids.

Practical Aspects

Selection of Polyelectrolyte

Factors which affect the choice of the particular flocculant or coagulant are the nature of the impurities and the particle size of suspended solids. Selection and dosage required can be determined by laboratory testing of the settling rate, clarity, and volume of sedimented mud. Laboratory testing does not exactly reproduce plant conditions and should be considered as a prelude to plant trials.

Preparation of Polyelectrolyte Solution

Most polyelectrolytes used as flocculants are available as a low dusting powder, and have to be distributed for use as a colloidal sol of concentration 0.05–0.5%. This limitation is due to the high viscosity of more concentrated dispersions (0.1–0.2 Pa s for 0.5%). The casual addition of a polyelectrolyte solid to water results in a jelly-like structure, which is useless as a flocculating agent. Efficient dispersion is achieved by use of an Educator–Venturi system, providing a homogeneous solution. Long-term storage or temperatures greater than 60 °C result in a loss of effectiveness caused by depolymerization.

Point of Addition

The point of addition of a polyelectrolyte is of utmost importance. The flocculant should be added at a point which allows for uniform mixing with the slurry, but it should not be subjected to excessive turbulence, which can disrupt the floc. It is also necessary to arrange for the introduction of the flocculant – as dilute as possible to ensure efficient distribution – at a point which allows time for particle collisions to take place prior to the solid removal stage. In practice, there are two systems which achieve the correct conditions for flocculation: gently agitated tanks, and flocculation chambers.

Removal of Flocculated Solids

The two principal techniques for removal of the solid formed – flotation or sedimentation – are employed, depending on the kind of flocs formed. A light material can be subjected to air treatment for flotation; flocs that tend to settle are generally separated by sedimentation (see Figure 3). Filtration is rarely used – mostly in water treatments, with sand or carbon filters, or rotary vacuum filters for carbonated raw beet juice.

Applications

Waste Water Treatment

Industrial waste water is frequently a complex solids–colloidal system, with soluble compounds of organic or inorganic origin.

Physicochemical treatments of industrial waste water can achieve good results where biological processes are inapplicable, e.g., with nonbiodegradable materials, toxic discharges, or removal of minerals and color. This is cheaper in terms of capital cost, more easily controlled and less space-consuming than biological treatment, but higher in terms of operational costs.

Adsorptive precipitation and flocculation are the main processes for coagulating colloidal particles to form larger aggregates. The principal reagents used are lime, ferric or aluminum sulfate, and polyelectrolytes.

The low-molecular-weight polymers have the advantage, over the inorganic coagulants, that they do not increase the salts' level, but they are often relatively uneconomical. The salts Al(III) or Fe(III) are

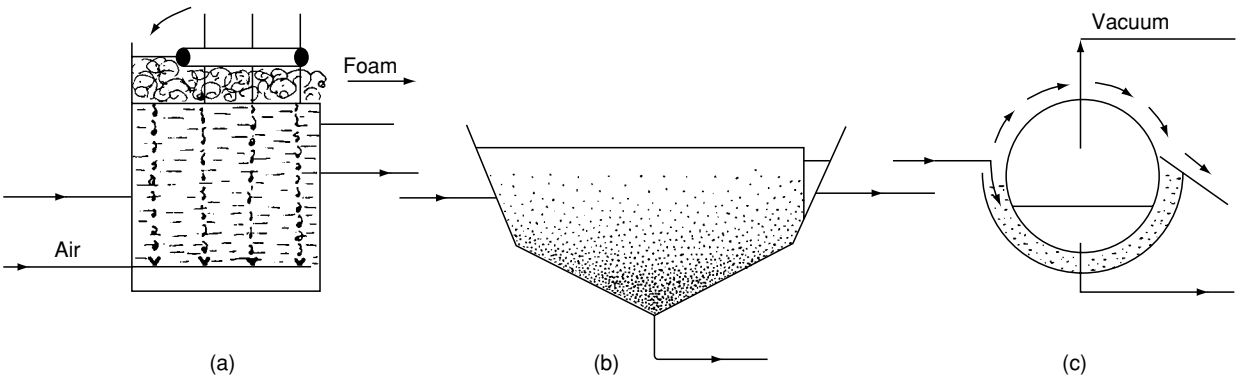

Figure 3 Equipment for removal of solids formed in flocculation: (a) flotation machine; (b) decanter; (c) vacuum rotary filter. Reproduced from Flocculation, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

commonly used first to promote coagulation; then, a high-molecular-weight polyelectrolyte is introduced to the coagulated solids. Highly dispersed solids need either polyvalent ions or low-molecular-weight polyelectrolytes. With the larger particles, high-molecular-weight polymers are more effective; for examples of applications, see Table 1. The flocculation permits the recovery of protein from waste water coming from the food industry, as well as from byproducts, e.g., agglomeration of whey proteins with carboxy-methylcellulose, or protein precipitation by amide polyelectrolytes in waste water of the potato industry. (*See* **Water Supplies**: Water Treatment.)

Raw Sugar Process

Raw sugar juices contain considerable amounts of colloidal materials which have to be removed in clarification. There are two groups of colloid: constituents of the beet cells transferred to the sol form during extraction, and those that originate as a result of condensation, e.g., melanine, melanoidine or invert or saccharose decomposition products. The process of removing colloidal substances is essential, as their presence in thin or dense juices, delays the sedimentation and filtration process, and favors foaming and the increase of coloration, and sugar loss during crystallization.

Lime acts as a primary coagulant, precipitating colloidal and colored matter. A radical improvement in the process of clarification of sugar juice was brought about by the application of flocculants. The application of sodium alginate or modified starch at dosage rates of 10–50 p.p.m. was not very satisfactory. Significant progress was made with the introduction of synthetic polyacrylamides. In general, anionic flocculants are applicable for the treatment of the raw juice from cane or beet. The flocculant dosage rates range from 0.8 to 2 p.p.m., and experience has shown that it is better to dilute the flocculant with raw juice instead of water. The favorable effect of this flocculant introduction was evident in the way the juice with sediment flowed to the filter or decanter.

A proper application is likely to increase the size of the colloidal particles by several times, which facilitates the plate and frame filter press function, and more compact filter cakes than normal are obtained. Polyelectrolytes are also recommended in sugar refining, and the addition to carbonated sugar juice improves both the clarification and filtration processes. (*See* **Sugar**: Refining of Sugarbeet and Sugarcane.)

Technology of Leaf-protein Concentrates

Leaf-protein concentrate (LPC) is one product resulting from green crop fractionation that provides the animal feed industry or food industry with a protein. Green crop fractionation offers the possibility of a greater total biological efficiency in the production of protein than can be achieved by other systems. It is possible to envisage several systems defined by the products of the process, which have as one of their products LPC in a wet or dried form. Freshly expressed juice from macerated green plants contains small amounts of fiber, minerals and proteinaceous material of leaf origin, chloroplasts or fragments thereof, and about 90% water. Removal of the protein fractions is possible through the destabilization of the colloidally suspended particles (Figure 4).

The flocculation of the chloroplastic function is carried out in chambers under light mechanical motion (100 r.p.m.) by introducing the flocculant directly to the green juice which was alkalized with sodium hydroxide to pH 7–8.5. Coagulation occurs almost instantaneously to produce macroflocs that settle down and make subsequent filtration easy.

The effectiveness of flocculation depends on the ionic character of the polyelectrolyte, as well as its

Table 1 Application of flocculation in industrial waste water treatment

Industry	*Pollution parameters*	*Flocculant*	*Benefits*	*Removal of solids*
Processing of poultry, slaughterhouse waste	Fats, grease, blood	Al(III) and polyelectrolyte	Reduction of total solids	Dissolved air flotation
Fish or shrimp processing	Fats, protein	Chitosan or anionic polyelectrolyte	Reduction of total solids	Dissolved air flotation
Pulp and paper	Lignin, color	Cationic polyamines	Color removal	Sedimentation
Fellmongering	Hair and skin products sulfide, ferrous salts	Fe(III) and anionic polyelectrolyte	Reduction of solids	Sedimentation
Vegetable and fruit processing	Carbohydrates, protein	Lignosulfonic acid	Reduction of BOD	Dissolved air flotation
Starch – gluten	Protein, carbohydrates	Hexametaphosphate or synthetic polymers	Reduction of BOD	Dissolved air flotation

BOD, biochemical oxygen demand.

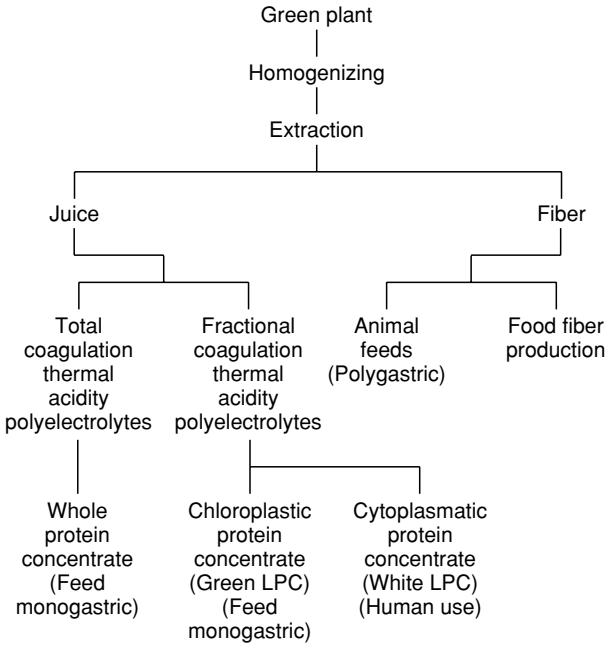

Figure 4 Protocol for obtaining protein fractions from extracted juice. LPC, leaf-protein concentrate. Reproduced from Flocculation, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

charge density and concentration, on the acidity of the plant juice, and on the species of plant. Satisfactory results, in terms of both quantity and concentrates and protein content, were obtained by adding 150–300 mg of Magnafloc LT-27 (Allied Colloids, UK) or Prodefloc C/1 (Prodeco, Italy) per liter of green alfalfa juice. From 100 kg of fresh plants, the process yields 1.6 kg of dry LPC containing 56% of protein.

Cationic and anionic polyacrylamides were applied on a laboratory scale for the flocculation of green juice from the following plants: leaves of sugar beet, clover, buckwheat, vetch, sunflower, and tobacco. The chemical composition of products obtained is similar to that obtained by means of the thermal coagulation method called Pro-Xan.

The application of flocculants in the processing of LPC allows a decrease in the energy consumption and yields proteins which are not denatured thermally.

See also: **Filtration of Liquids**; **Sugar**: Refining of Sugarbeet and Sugarcane; **Water Supplies**: Water Treatment

Further Reading

Eilbeck WJ and Mattock G (1987) *Chemical Processes in Waste Water Treatment*. Chichester, UK: Ellis Horwood.

Fiorentini R, Pachetti G and Galoppini G (1978) *Attualita e Prospettive del Frazionamento umido dei Foraggi*. Pisa: Universita di Pisa.

Ives KJ (1978) *The Scientific Basis of Flocculation*. Alphen aan Rin, The Netherlands: Sijthoff and Noordhoff.

McGinnis RA (1981) *Beet Sugar Technology*. Fort Collins, CO: Beet Sugar Development Foundation.

FLOUR

Contents

Roller Milling Operations

J E Dexter, Canadian Grain Commission, Winnipeg, Manitoba, Canada
A K Sarkar, Canadian International Grains Institute, Winnipeg, Manitoba, Canada

History of Flour Milling

The origins of wheat milling date back thousands of years to the dawn of civilization. Milling technology gradually advanced from the mortar and pestle, used by primitive cultures about 10 000–12 000 years ago, to the invention of the millstone in Roman times. Millstones remained in vogue until the introduction of the purifier, plansifter, and the horizontal roller mill in Europe during the latter part of the nineteenth century. The introduction of improved equipment and the concept of gradual reduction of wheat into flour allowed vast improvements in the degree of refinement of flour. Milling equipment has undergone continual improvement since then, but the basic concept of modern roller milling has remained unchanged.

Structure and Composition of the Wheat Kernel and Flour

The main constituents of the wheat kernel are bran, germ, and endosperm (Figure 1). The endosperm includes the outer aleurone layer, which is removed with the bran during flour milling, and the starchy endosperm. The germ contains the embryo, and a transport organ, the scutellum.

High-quality flour originates from the starchy endosperm. The starchy endosperm comprises about 85% of the wheat kernel. Therefore, it is theoretically possible to convert about 85% of wheat into high-quality flour, free of bran and germ contamination. In practice, the crease, which runs along the length of the grain and reaches almost to the middle of the grain, makes perfect separation of the endosperm impossible.

The constituents of the wheat kernels have very different compositions (Table 1), so that the extent of contamination of the flour by bran and germ alters flour composition and flour functionality. The starchy endosperm contains a low level of minerals compared to other kernel constituents. Accordingly, flour ash (mineral matter left behind after incineration of flour) is widely used as a flour refinement index; the low ash content is indicative of a highly refined flour containing almost pure endosperm. The endosperm contains proteins, known as gluten, which give wheat flour dough the viscoelastic properties essential for high-volume bread and firm pasta.

Wheat kernel hardness has a wide range. Durum wheat is very hard, whereas common wheat ranges from hard to soft. Endosperm structure differs for wheat of different hardness. The endosperm of

Table 1 Approximate composition of common wheat and its constituents[a]

	Wheat	*Bran*[b]	*Endosperm*	*Germ*
Proportion of kernel (%)	100	13	85	2
Protein content (%)	13	14	12	40
Ash content (%)	1.5	7	0.4	5
Thiamin (mg kg^{-1})	3	5	0.4	100
Nicotinic acid (mg kg^{-1})	33	200	7	40

[a]Taken from various sources. Expressed on 13% moisture basis.
[b]Includes the aleurone layer.

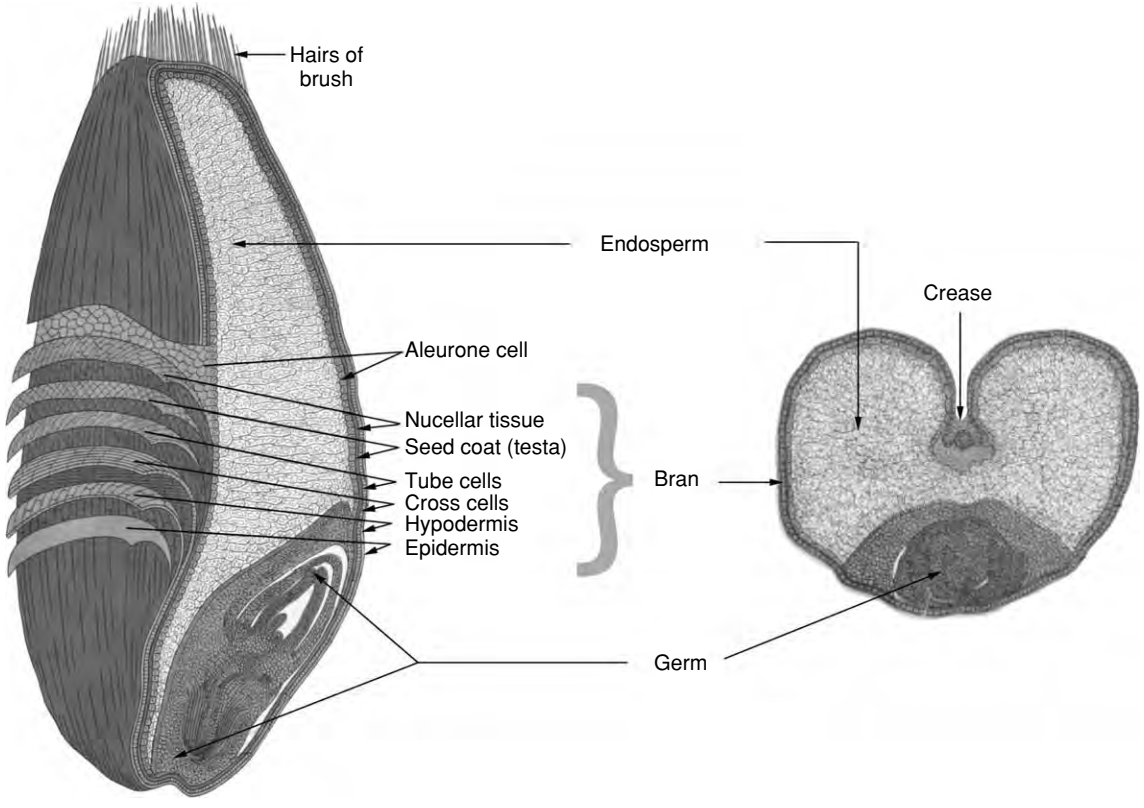

Figure 1 (see color plate 66) The structure of the wheat kernel sectioned longitudinally (left) and transversely (right). The bran layer is comprised of six layers. The aleurone layer, considered part of the endosperm, is removed with bran during roller milling. The crease, which runs the full length of the kernel, prevents fully efficient separation of bran from endosperm.

hard wheat is compact, with starch granules rigidly held within a protein network. The endosperm of soft wheat is more porous.

In general, as the proportion of wheat converted into flour (flour extraction rate or flour yield) increases, contamination by bran and germ increases. As a result the nutritional value of the flour improves as extraction rate increases, but processing quality and shelf-life diminish. The proteins in germ and bran are nongluten proteins that contribute to nutritional value, but do not contribute to processing properties. The germ contains low-molecular-weight sulfhydryl compounds that weaken dough properties by reducing disulfide cross-links, which are key elements of gluten polymer structure. The germ also contains oils and enzymes that reduce storage stability by promoting rancidity. The germ and bran are both rich in nutritionally important vitamins. The bran contains most of the fiber in the kernel, also an important nutritional consideration. However, as the level of bran in flour increases, flour dough functionality is reduced, and flour becomes darker, detracting from end-product appearance.

The Flour Milling Process

Wheat selection, preparation of wheat for milling, and milling technique all play a role in determining flour quality and milling profitability. The remainder of this article describes a modern roller milling operation, and explains how the various stages in the milling process contribute to the yield and quality of final milled products.

Wheat Selection

The milling characteristics and processing quality of wheat are dependent on physical condition and inherent quality traits. Wheat of intrinsically good quality can be rendered of poor processing quality by environmental damage or improper storage. Different classes of wheat, and wheat from different origins, have different quality traits. High-protein hard common wheat is ideal for producing bread flour. Low-protein soft common wheat is preferred for producing cake and biscuit flour. Durum wheat is preferred for pasta products.

In order to meet customer requirements, millers specify wheat type and minimum quality standards that must be met by suppliers. Differences in wheat hardness among wheat classes require alterations in conditioning and milling technique to achieve optimum results. When blends of different wheat types are milled, it is preferable to condition each type separately, but convenience, cost, and storage limitations often require blending prior to conditioning.

Wheat Reception and Storage

Wheat may be received at mills by truck, rail, or ship (Figure 2). Incoming wheat is weighed, and a sample is analyzed prior to storage. The content of foreign material and damaged grains is determined. Analyses such as weight per unit volume (test weight), protein content, and moisture content are performed. This determines whether the incoming wheat is within specification. Wheat is placed in a bin with other wheat of comparable physical condition and similar end-use potential to allow quality characteristics to be preserved, and used to the best advantage. In most cases, the wheat passes through a magnet to remove ferrous metals, and it may be given a cursory cleaning prior to elevation to storage bins.

Wheat Cleaning

Wheat must be cleaned prior to milling to remove foreign material such as stones, metal, unthreshed grain, badly damaged kernels, and foreign seeds that either adversely affect flour quality or pose a hazard to milling equipment.

Figure 2 shows a typical wheat-cleaning scheme. The wheat first passes through a magnet, and then proceeds to a grain separator which uses sieves and aspiration to remove impurities. Large impurities, such as maize, soybeans, and unthreshed wheat are removed by the top screen; small impurities, such as small and broken kernels and weed seeds, pass through the bottom sieve. As wheat falls off the end of the bottom sieve, it is aspirated to remove chaff, dust, and shriveled kernels.

Next, the wheat goes to a destoner where impurities such as mud balls and small stones comparable in size to wheat kernels, but denser, are removed. A destoner is an oscillating inclined metal screen through which an upward air current is passed. Wheat floats on an air cushion and moves down the slope. Denser materials make more contact with the screen, are propelled up the slope by the screen oscillations, and are collected separately.

The wheat now passes through disk separators to remove remaining impurities that differ in length to wheat kernels, but are of similar density and cross-section. Disk separators comprise disks revolving in a vertical plane through the wheat. The surface of the disks has numerous pockets which pick up particles that are short enough to lodge in the pockets. There are generally two units. The first is designed to pick up wheat while rejecting kernels longer than wheat, such as oats and barley. The second picks up small seeds and rejects wheat kernels.

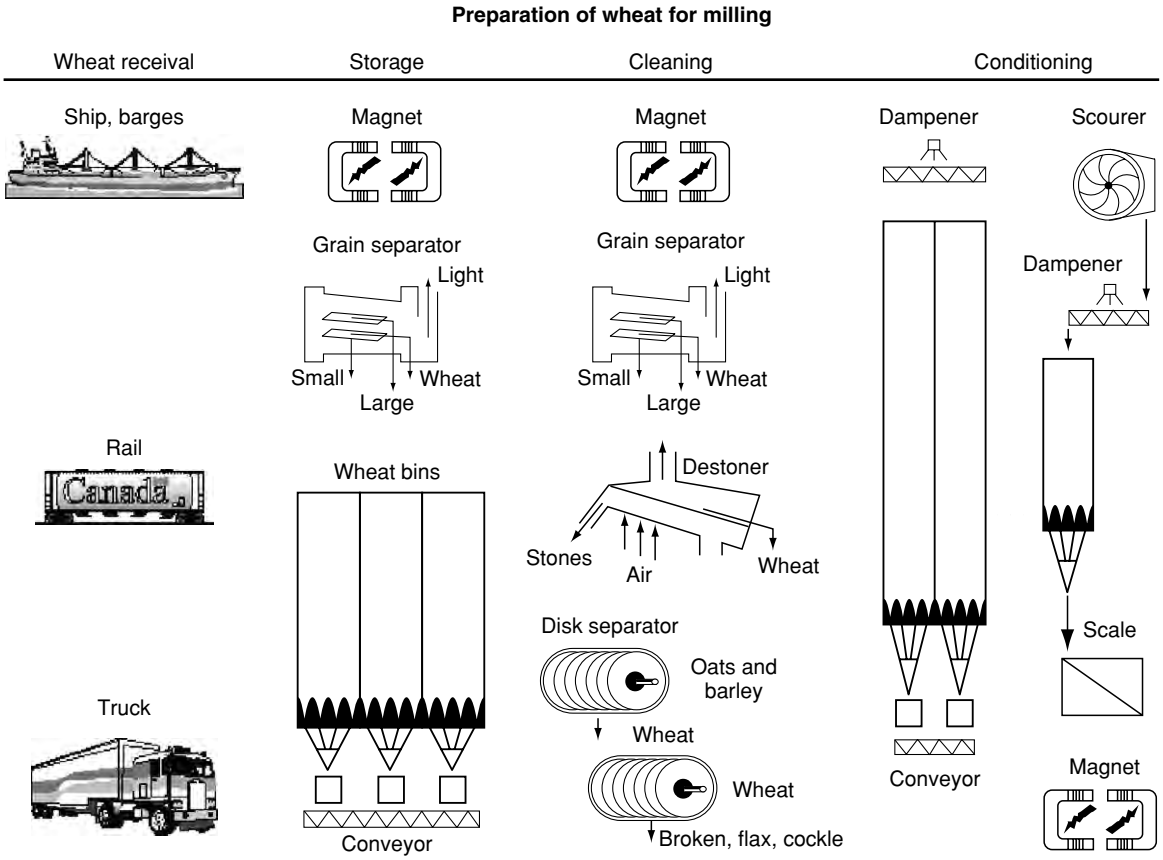

Figure 2 A simplified wheat intake, cleaning, and conditioning diagram.

Wheat Conditioning

After cleaning, the wheat is conditioned (tempered) by the addition of moisture (Figure 2). The precise addition of the correct tempering moisture is of fundamental importance to the milling process.

The primary purpose of tempering is to optimize the separation of bran from endosperm during the milling process. Tempering also reduces bran powdering, thereby reducing bran contamination in flour, and imparting optimum texture to the endosperm for efficient reduction into flour. Too much tempering moisture reduces flour yield because complete separation of bran from endosperm is difficult to achieve, and sieving efficiency is reduced. Too little tempering moisture results in bran powdering which contaminates flour.

Tempering water is added by spray nozzles to wheat in an enclosed screw conveyor. Modern tempering equipment features electronic control and vigorous mixing action to insure that precise and uniform addition of moisture is achieved. Following tempering, wheat is stored in bins for a predetermined time to allow uniform distribution of moisture within the wheat kernel.

Optimum tempering moisture varies with wheat type. Soft wheat is tempered to lower moisture than hard wheat, because soft wheat stocks are inherently more sticky and difficult to sift than hard wheat stocks, and that tendency is exacerbated as moisture content increases. Optimum tempering time may vary from several hours to a day or more depending on incoming wheat moisture content and hardness. Water penetrates into the endosperm of soft wheat more quickly than for hard wheat because soft wheat endosperm is more porous. Therefore, tempered soft wheat does not need to be rested as long before milling as hard wheat. Wheat is often tempered in two stages with a resting time between stages, particularly when a large amount of tempering moisture is added.

After tempering, wheat is scoured. Scourers use steel beaters to throw wheat against a wire screen. Dirt and loosened bran, which would contaminate the flour, pass through the screen. Commonly, as shown in Figure 2, wheat is lightly sprayed just prior to milling. This is to toughen the bran, and reduce bran powdering. Following a brief period in holding bins, the wheat is weighed, and passed

through a magnet to remove ferrous metal fragments prior to milling.

Mill Equipment

Roller milling consists of a succession of grindings, each followed by separation of the ground stock according to its size and composition. The most important pieces of equipment common to all modern flour mills are the roller mill, the plansifter, the purifier, and the bran duster.

The standard modern roller mill is a double-mounted stand consisting of two pairs of steel rollers mounted within a cast iron frame (Figure 3). The roller pairs are mounted back to back. Each roller pair is driven separately and has separate feeding and adjustment mechanisms. Recently, roller mills with four pairs of rollers, comprising two pairs stacked vertically with the upper pairs feeding the lower pairs directly, have become available. They are gaining in popularity, particularly for use early in the break system, and early in the reduction system, which allows an economical way of increasing milling capacity.

The two rollers within each pair rotate in opposite directions and at different speeds. The ratio of roller speeds is known as the differential. Having one roller rotate faster than the other imparts shear to the grinding action. The slower roll holds the stock, while the faster roll compresses and shears it.

The surface of the rollers may be smooth or corrugated, with flutes cut in a spiral pattern along the length of the roller. The spiral prevents the rollers from locking, and imparts a cutting and shearing effect. The number of flutes per centimeter and the shape of the flutes vary widely, depending on the stage of the milling process, the type of wheat being milled, and the desired flour extraction rate. Reduction rollers are usually smooth, although the surface is slightly rough (frosted) to impart some shear to the grinding action. For the most effective performance, break rollers are run at a differential of about 2.5:1, whereas reduction rollers are run at a differential of about 1.25:1.

A modern plansifter consists of as many as 30 layers of sieves, rotating in a horizontal plane. The sieves are arranged in sections, with some of the sifter units comprising up to eight sections. The aperture of the sifter clothing ranges from about 1600 μm down to less than 100 μm. Coarse clothing is composed of wire mesh, while finer clothing is composed of nylon or silk.

Stock enters the sifter at the top. The rotating motion of the sifter works the finer material through the top sieve. The coarse material is drawn away and the finer material proceeds to a series of progressively finer sieves where further size separations are made.

Bran dusters use impact to remove endosperm that remains attached to broad bran at the end of the break system. The same machine, when used to remove endosperm from shorts (fine bran from the end of the break system), is commonly referred to as a shorts duster. Bran or shorts is fed into a horizontal

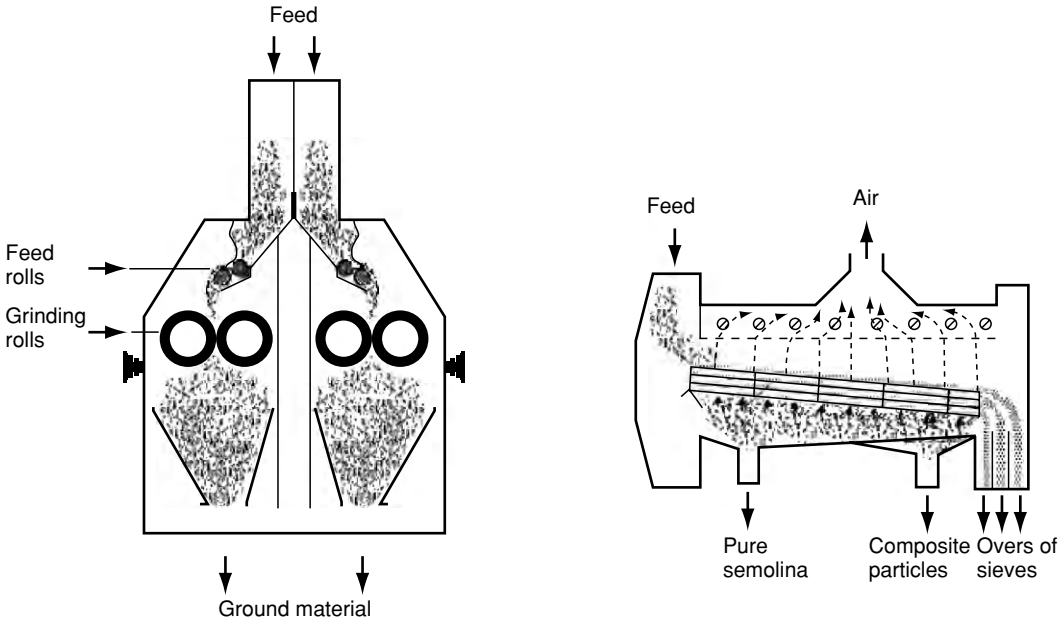

Figure 3 Schematic views of a modern roller mill (left) and a modern three-deck purifier (right).

cylinder, the lower part of which is perforated. Within the cylinder, finger beaters attached to a central shaft rotate, and scrape endosperm from bran or shorts. Endosperm is pushed through the apertures, and bran or shorts passes though the cylinder and is collected separately.

Purifiers separate pure endosperm particles from bran-rich particles of comparable size (Figure 3). This is done by taking advantage of the greater density of pure endosperm particles.

Purifiers comprise a sieve frame mounted with a slight downward slope. The frame is clothed with progressively coarser material from head to tail, and an air current passes upwards through the sieves. The frame oscillates longitudinally. The frame oscillation and the air currents cause the stock to stratify, with the lighter bran-rich particles rising to the surface. The heavier endosperm particles fall though the sieve if the aperture is large enough, whereas the bran-rich material floats over the end of the sieve frame.

The combination of the air flotation and sieving action allows the purifier to separate the incoming mixture into material that is progressively coarser and more contaminated by bran from head to tail. Modern purifiers have two or three decks of sieves to enhance efficiency of separation.

Flour Milling

Figure 4 gives a simplified version of a typical hard common wheat flour mill flow, designed to give a flour extraction rate of about 75%. The milling process comprises the break system, intermediate processing, which includes purification and sizing, and the reduction system.

The wheat kernel is opened on the first break rollers. The ground material passes to the first break sifter, where particles are separated according to size. The largest particles, which consist of wheat bran and adhering endosperm, proceed to the second break rollers. Each successive break and sifting passage separates more bran from endosperm. Immediately following the last break sifter, the bran proceeds to a shorts duster to remove some of the adhering endosperm.

The primary purpose of the break system is to separate endosperm from bran as efficiently as possible, although a small amount of break flour is

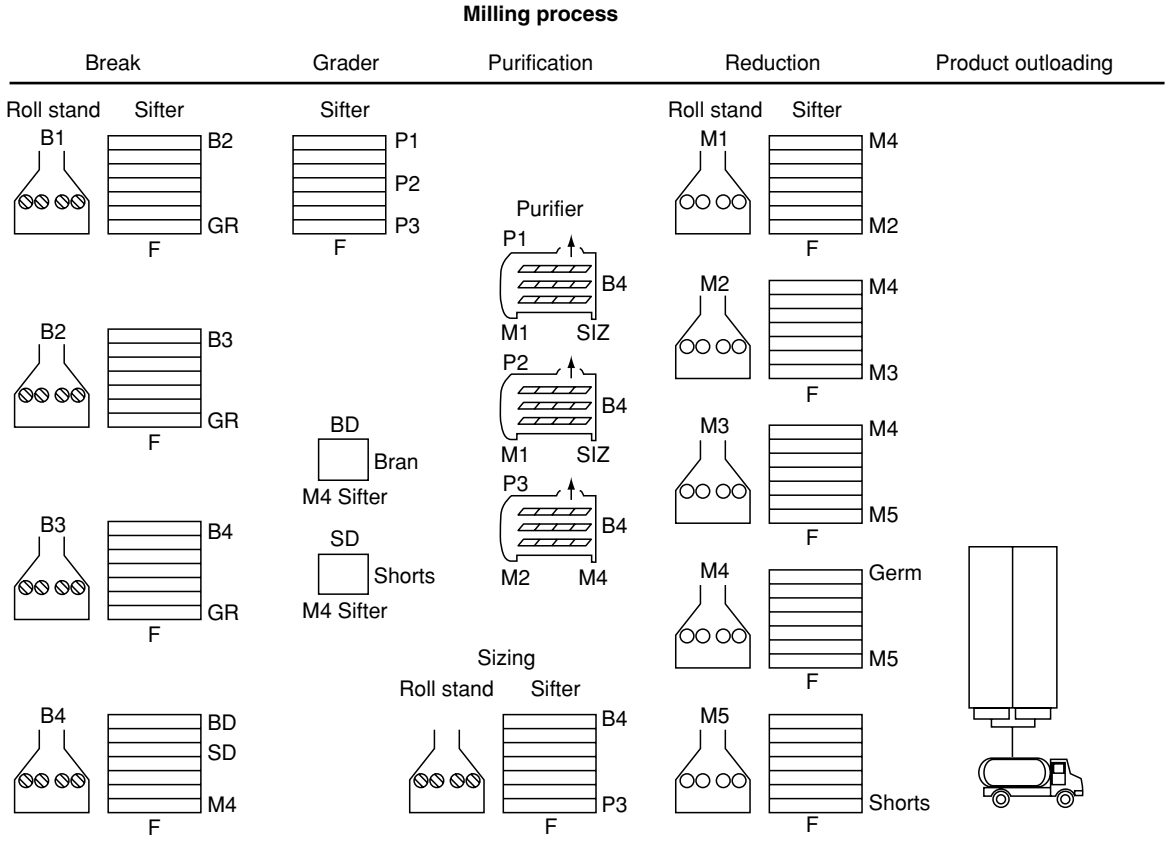

Figure 4 A simplified hard common wheat mill diagram. B, break; BD, bran duster; GR, grader; P, purifier; SD, shorts duster; SIZ, sizing; M, middling. Corrugated rollers are cross-hatched and smooth rollers are open.

produced. Most of the finer stock produced by the break system consists of particles larger than flour known as semolina (coarser particles) and middlings (finer particles). Semolina and middlings are separated by a sifter, referred to as a grader, according to particle size, before proceeding to purifiers.

Purifiers separate pure endosperm particles from brain-rich material. As a result, essentially pure endosperm is sent to the beginning of the reduction system. This allows the production of a large proportion of highly valued, highly refined white flour.

The coarse bran-rich material from purifiers proceeds to the sizing (scratch) system. The objective of the sizing system is not to make flour, but to remove bran adhering to middlings. This is accomplished by light grinding on finely corrugated rollers. The ground stock is then sifted and purified before continuing on to the reduction system.

The heart of the flour milling process is the reduction system, where most of the flour is produced. The reduction system reduces the carefully sized and purified middlings to flour particle size. This is done gradually by a series of successive grinding and siftings. Germ, which is flattened rather than reduced by smooth reduction rolls, can be recovered by sifting.

Grinding conditions must be carefully controlled to produce the maximum amount of highly valued prime-quality flour. If grinding is too severe, bran contamination is increased, reducing flour brightness. Severe grinding can also result in excessive starch damage which adversely affects flour functionality.

Flour Dividing

A flour mill may produce a single or multiple grades of finished flour from the milling of a given wheat or wheat mix. These finished flours are prepared by combining various flour streams that are produced at different stages of milling (Figure 5).

Individual flour streams, when compared with each other, exhibit variable processing quality. Flour millers may combine all flour streams to produce a single flour known as a 'straight-run' or 'straight-grade' flour. A more complex alternative is to produce several flours with different properties by judicious blending of streams. Figure 5 shows that each flour stream leaving a sifter can be directed to any one of three flour conveyors. Each continuous line represents a given flour stream flow to a designated flour conveyor. In this example, three flour streams each go to conveyors for flour #1 and flour #2, respectively, while two streams are directed to the conveyor for flour #3. This is known as 'divide' or 'split-run' milling and allows millers with demanding clients to increase return by closely targeting specific processing requirements.

Flour Additives

Mills commonly add additives to flour to improve nutritional value or to assist in achieving specific processing quality. As mentioned earlier, endosperm, the main component of flour, is poor in vitamins compared to the bran and germ. Many countries have legislation requiring the addition of vitamins and minerals to improve the nutritional properties of flours.

Bleaching agents like benzoyl peroxide may be added to give the flour a whiter appearance. Other additives impart better functionality. These include various enzymes (alpha-amylase of either fungal or grain origin) and oxidizing agents. Additives are available in powder form, and are dispensed directly into flour conveyors using powder feeders.

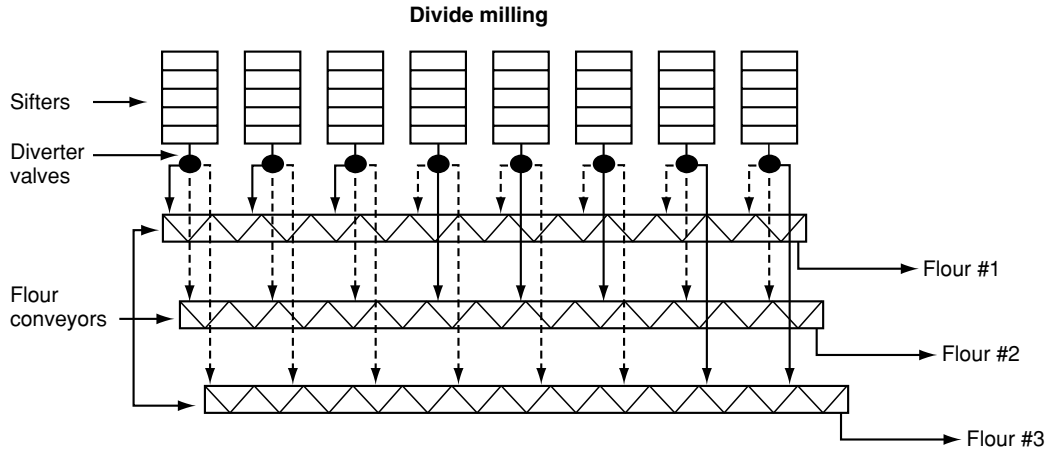

Figure 5 An example of divide milling, where three custom flours are being produced from a single wheat mix.

Soft Wheat Milling

Soft wheat milling is similar to hard wheat milling, but intrinsic differences between soft wheat and hard wheat require some modifications to the milling process. Soft wheat breaks down more quickly than hard wheat, and soft wheat stock is stickier and fluffier, which makes it more difficult to sift than hard wheat stock. In addition, the endosperm of soft wheat adheres more strongly to bran, reducing flour extraction rate expectations by up to 2% unless the break system is extended.

Soft wheat must be fed more slowly to the mill to facilitate sifting and to insure that stock flows freely through the mill. The rapid breakdown of soft wheat increases the yield of flour from the break system and reduces the yield of semolina and middlings. As a result, purifiers are of less importance and are often absent from the flow.

Durum Wheat Milling

The same milling equipment is used for durum wheat milling as for common wheat milling, but the mill flow is very different. The product preferred for premium pasta and couscous is uniformly sized semolina, free of bran. Durum wheat flour produced from semolina mills is a lower-valued byproduct.

Durum wheat is very hard, which facilitates a high yield of semolina with minimal production of flour. The break system for durum wheat is extended to allow gradual breakdown of the wheat kernel to achieve a maximum production of semolina and a minimum production of flour. Purified semolina from the break system is uniformly sized and made free from adhering bran by a succession of sizing passages, gradings, and purifications. Most of the semolina is from the sizing purifiers, making durum mills readily recognized by the large number of purifiers.

Rye Milling

Rye grain is milled into various forms of meal and flour for application in the production of a variety of baked products, particularly in northern and eastern Europe. Dark rye bread, pumpernickel, and a host of other bread products containing rye offer a unique combination of texture and flavor that is truly characteristic of rye.

A rye mill also uses grinding and sifting equipment similar to that used in wheat flour milling. Rye milling is similar to the milling of soft wheat. The endosperm is very soft and sticky, making it very difficult to grind and sift efficiently.

Because rye is so soft, a significant proportion of the total flour is produced at the beginning of the flow following each break passage. Rye endosperm is so soft and plastic that smooth rolls tend to flatten middlings rather than reduce them. As a result, all the grinding passages in a rye mill use corrugated rolls, including the reduction rolls.

Rye endosperm breaks down quickly into fine particles, and these particles lump together. Therefore, purification passages are of no value and purifiers are absent. Rye has gluten, but it imparts much less elasticity to dough than wheat gluten. Rye flour color is inherently dark. As a result, rye flour produces bread loaves that are denser, darker, and heavier in texture than wheat flour bread.

Milling Byproducts

The byproducts of milling receive little attention, but they are an important economical consideration in flour milling. These include impurities from the cleaning house (screenings), mill feed (bran and shorts), and germ.

Screenings, with the exception of metal, stones, and mud balls, are usually ground and sold as animal feed. The mill feed, which comprises large bran flakes produced from the break system and finer bran particles, or shorts, are also primarily used as animal feed. Depending on market conditions, mill feeds also find use for foods such as breakfast cereals and high-fiber specialty wheat products.

Wheat germ is a highly valued byproduct because of its excellent nutritional composition. As a result, many modern mills have sophisticated germ recovery systems.

Debranning (Preprocessing)

Wheat debranning, or preprocessing, a process in which the bran layers are removed sequentially by friction and abrasion operations in modified rice polishers prior to roller milling, has enormous potential in wheat milling. Debranning has gained considerable acceptance in durum wheat semolina milling in particular. The yield and refinement of semolina are significantly improved, allowing more efficient milling of top-quality durum wheat, or, alternatively, the use of lower-quality durum wheat to produce semolina within customer specifications. There is less conclusive evidence that debranning improves the milling performance of common wheat.

Regardless of wheat class being milled, an advantage of using debranning is lower capital investment because simplifications of the mill flow (the break

system is almost eliminated) allows more compact plants for a given capacity. In addition, the byproducts of debranning have the potential to be more valuable than byproducts from conventional roller milling. Rather than removing all bran layers together, as in conventional roller milling, during debranning the individual bran layers are stripped off in sequence. Each bran layer has distinct physicochemical and nutritional properties, giving debranning byproducts great promise as novel food ingredients.

Milling Process Monitoring and Control

All electrical motors driving various pieces of equipment are sequence-controlled and interlocked to provide ease of operation, control, and protection. This allows sequential starting and stopping of the motors and stopping of the process in a failsafe manner in case of interruptions due to a malfunction or for any other reason. Additionally, a number of process control functions are incorporated that may, depending on the extent of use of various types of sensors, switches, high-level and low-level indicators, allow the entire process to be automated. A conventional control system involves a mimic process diagram for visual display of the process, and a hard-wired electromechanical relay control system for control functions.

Since the early 1980s there has been a gradual transition to control flour mill automation with computerization. Mimic panels with small pilot lights are now being replaced by computer graphics for visual display of the process status. Hard-wired control relays are being replaced by programmable logic controllers (PLC). PLC control with PC operator interface enables performance of all the functions described above. Additionally, it offers the flexibility of changing control functions, if required, through simply making programming changes rather than changing the hard wiring. It also allows full integration of the remaining critical functions relating to process control and performance evaluation, including product yield calculations and automated roll gap adjustment.

Another important advance is online quality monitoring. Factors such as moisture content, protein content, and flour color can be monitored continuously using automated online samplers allowing flour millers to meet customer specification efficiently.

Computerized process control has greatly reduced labor requirements in flour mills. Many mills have gone to so-called lights-out operation, where the mill runs for extended periods without staff present in the mill building.

See also: **Biscuits, Cookies, and Crackers**: Nature of the Products; **Bread**: Breadmaking Processes; **Cereals**: Breakfast Cereals; **Dietary Fiber**: Bran; **Food Fortification**; **Milling**: Principles of Milling; **Pasta and Macaroni**: Methods of Manufacture; **Rye**; **Wheat**: Grain Structure of Wheat and Wheat-based Products

Further Reading

Canadian International Grains Institute (1993) *Grains and Oilseeds: Handling, Marketing, Processing*, 4th edn, vols 1 and 2. Winnipeg, Manitoba: Canadian International Grains Institute.

Dexter JE and Wood PJ (1996) Recent applications of debranning of wheat before milling. *Trends in Food Science Technology* 7: 35–41.

Fabriani G and Lintas C (eds) (1988) *Durum Wheat: Chemistry and Technology*. St Paul, Minnesota: American Association of Cereal Chemists.

Hoseney RC (1994) *Principles of Cereal Science and Technology*, 2nd edn. St Paul, Minnesota: American Association of Cereal Chemists.

Matz SA (ed.) (1991) *The Chemistry and Technology of Cereals as Food and Feed*, 2nd edn. McAllen, Texas: Pan-Tech International.

National Association of British and Irish Millers (1976) *The Practice of Flour Milling*, vol. 1. Richmond: Dimbleby Printers.

Palmer GH (ed.) (1989) *Cereal Science and Technology*. Aberdeen: Aberdeen University Press.

Pomeranz Y (ed.) (1988) *Wheat: Chemistry and Technology*, 3rd edn, vols 1 and 2. St Paul, Minnesota: American Association of Cereal Chemists.

Posner ES and Hibbs AN (1997) *Wheat Flour Milling*. St Paul, Minnesota: American Association of Cereal Chemists.

Yamazaki WT and Greenwood CT (eds) (1981) *Soft Wheat: Production, Breeding, Milling and Uses*. St Paul, Minnesota: American Association of Cereal Chemists.

Analysis of Wheat Flours

K R Preston and P C Williams, (formerly of) Grain Research Laboratory, Winnipeg, Manitoba, Canada

Introduction

Flour, produced by the selective removal of the germ and the outer layers (bran) from the endosperm of wheat during the milling process, is used in the production of a wide variety of baked, extruded, and sheeted products throughout the world. Each of these products may require different types of flour

to produce an acceptable product, as shown in Table 1. The ability of a flour to produce a specific product is determined by the inherent characteristics of the wheat, the effects of environment during growth, harvesting, and storage on wheat properties, and the milling techniques used to produce the flour. Flour (and wheat)-testing procedures have been developed to predict or measure the effects of these factors on flour-processing quality and to insure health standards have been met.

The most common applications of flour analysis include:

1. In wheat-breeding programs where new cultivars are evaluated for selection.
2. In government and private laboratories responsible for monitoring quality and health standards, providing information for segregation of wheats and flours and providing information to buyers.
3. In flour mills for selection of wheat lots and the blending thereof and to insure that flours produced meet customer specifications and health standards.
4. In end-user facilities to verify that flour specifications have been met and that health-related standards are maintained.

This article provides an overview of quality and health-related testing procedures commonly used in flour analysis with emphasis on their use and interpretation. This review is restricted to common (hexaploid) wheat flour testing and does not include durum wheat which is normally milled to semolina for the production of pasta and couscous. Detailed information on wheat and flour test procedures are available from approved methods manuals published by the two largest associations involved in cereal chemistry – the American Association of Cereal Chemists (3340 Pilot Knob Road, St Paul, MN 55121, USA http://www.aaccnet.org) and the International Association for Cereal Science and Technology (PO Box 77, Wiener Str. 22a, A-2320 Schwechat, Austria, http://www.icc.or.at).

Table 1 Major uses of common wheat flour classified by protein content and physical dough strength

Flour type[a]	*Uses*[b]
Strong, high-protein	Pan breads, blending
Medium strong, medium-protein	Hearth breads, flat breads, steamed breads, noodles
Weak, low-protein	Cakes, cookies, dumplings Soft wheat noodles (udon)

[a]Approximate protein ranges for high-, medium-, and low-protein (14% moisture basis) flours are >12%, 10–12%, and <10% respectively.
[b]Products listed with each flour type are intended as a general guideline. Many of the products listed are often produced from the other flour types.

Flour Quality Tests

Flour quality tests are designed to predict or determine milling efficiency and end-product suitability. These tests can be subdivided into three categories including analysis of basic constituents (proximate analysis), processing-related parameters, and analysis for specific additives. The most common of these tests are summarized in Table 2.

Flours are also commonly subjected to physical dough tests which provide information involving the viscoelastic properties of flour dough and to small laboratory-scale end product processing tests which provide the most accurate estimate of product quality.

Proximate Analysis

Proximate analysis of flour normally involves the determination of moisture, protein, and ash content. Measurement of fiber components (endogenous and that added from other sources) is also becoming increasingly important in view of their nutritional importance. Determination of oil and carbohydrate composition, the latter of which is determined by difference (i.e., 100 – sum of the five constituents listed above) is not normally carried out and will not be discussed further.

Flour moisture content is fundamental information, since variance in moisture content affects the proportion of all other constituents. Analytical information for flour is normally reported on a specified moisture basis, while the amount of flour required for physical dough or for end-product testing is normally weighed to a specified moisture basis. Moisture content is also important for both economic reasons and flour storage. Selling higher-moisture flour is normally more profitable to mills. Flour will deteriorate rapidly if moisture exceeds 15% in moderate climates and 14% in tropical climates. Flour moisture can be determined by air-oven methods. For more rapid analysis, near-infrared reflectance (NIR) can be used.

The protein content of flour is an essential piece of information concerning the end-use potential of a

Table 2 Common quality flour tests

Constituents	*Parameters*	*Additives*
Moisture	Flour color	Potassium bromate
Protein	Wet gluten	Ascorbic acid
Ash	Gluten index	Riboflavin
Fiber	Starch damage	Thiamin
	Particle size	Niacin
	Amylograph peak viscosity	Iron
	Falling number	Folic acid
	Maltose value	
	Gassing power	
	Fat acidity	

sample within a flour type or class. For example, with bread-wheat flours, protein content is generally positively associated with baking quality parameters such as loaf volume and bread score, as shown in Figure 1. In contrast, with cake and cookie flours the relationship between quality and protein content is negative. However, protein quality, as well as quantity, exerts a major impact upon end-product quality and must also be considered. Protein quality can be assessed by physical dough and end-product tests, described later in this article.

For many years the Kjeldahl test and modifications thereof, involving the measurement of ammonia nitrogen released by acid hydrolysis, has been universally accepted as a method for the determination of the protein content (N × 5.7) of wheat and flour. Since the last edition of this encyclopedia, improvements in the instrumentation of the Dumas test have resulted in the introduction of a number of high-temperature combustion nitrogen analyzers (CNA). The increased accuracy and precision of these instruments (relative to the Kjeldahl method), combined with the elimination of the requirements for fume exhaust, drainage systems, and the use of highly corrosive chemicals, have led to increasing adoption of the Dumas test as the accepted reference method for protein determination in many commodities, including flour. In addition, the Dumas method offers improvements in the time per test (3 min, as compared to 2 h for the Kjeldahl test), while the method uses no offensive chemicals. The instruments are semiautomated, which enables virtually continuous analysis throughout the working day. For more rapid analysis, NIR can be used to predict protein content from whole and ground grain and flour.

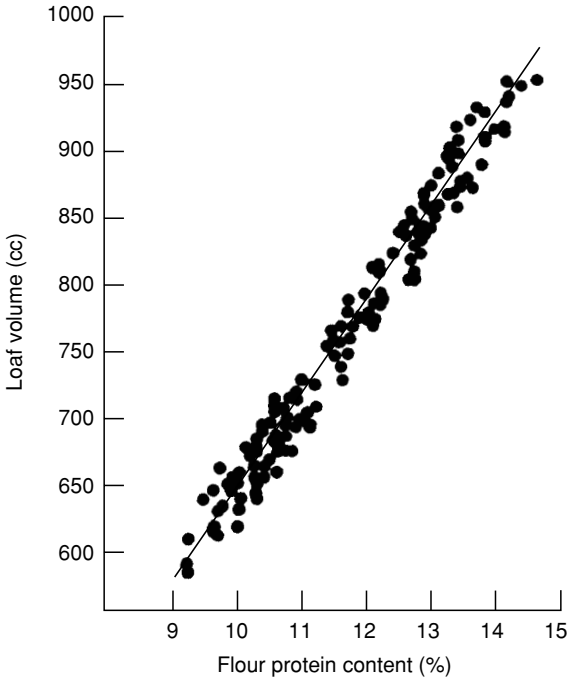

Figure 1 Relationship between loaf volume of remix-processed, straight-dough bread, and flour protein content (14% moisture basis) for Canadian red spring wheat straight-grade flours. Reproduced from Flour, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

The ash test is used in the milling industry as a measure of milling efficiency and in classification of mill streams for the production of various types of flours. Ash content is primarily a measure of the efficiency of separation of the low-ash endosperm from the high-ash bran. Flours with low ash content (less than about 0.40%, 14.0% moisture basis) are termed patent flours and are derived from the inner endosperm of the wheat kernel. This flour is used where a bright white color is needed in the end product.

Straight-run, or straight-grade flours, which are most commonly produced worldwide, can have ash contents varying from approximately 0.48 to over 0.60%. They are derived from the entire endosperm with varying degrees of bran inclusion and normally represent wheat extraction rates of 74–80%. High-ash flours (> 0.60%), involving high wheat extraction rates (80–95%), are produced where a darker-colored product is preferred by the consumer (e.g., brown and wholewheat breads), for traditional high-extraction flour products such as flat breads in the Middle East and chapattis in the Indian subcontinent or where government policy requires high extraction rates for economic reasons. As with protein, ash content is a common flour measurement used by mills to meet customer specifications. In some countries, flour color is used in place of ash content as a measure of milling efficiency. Flour ash content is normally determined by weight difference after incineration of a sample for several hours in a high-temperature (550–590 °C) muffle oven.

Crude fiber and dietary fiber have not been commonly measured in flour as a routine test. However, measurement of dietary fiber is becoming increasingly important in flours to which high-fiber ingredients have been added for nutritional considerations (e.g., high-fiber breads and breakfast cereals). At present, a number of methods are used to measure dietary fiber but no single procedure has gained widespread acceptance.

NIR analysis is now in commercial use as an inline instrument in modern flour mills. The NIR instruments are calibrated to provide continual data for moisture, protein, and ash contents of flour during

the milling process. The inline instruments are used in conjunction with benchtop computers (PCs) to monitor and control flour composition. The information is used by millers to control mill-flow. As well, NIR instruments have been calibrated for the prediction of flour starch damage and wet gluten content. Table 3 summarizes typical data for proximate analysis of flour, using a scanning spectrophotometer.

Flour is an ideal medium for NIR analysis. Due to the sensitivity of NIR analysis to particle size, the technology also affords an excellent means of monitoring the consistency of the production of all products of a flour mill.

Processing Related Parameters

Several tests provide useful information concerning the functionality or end-use potential of flour. The most widely used of these processing-related parameters are listed in Table 2. They can be grouped into parameters that are influenced mainly by milling extraction (flour color), protein content (wet gluten), wheat texture (starch damage and particle size), α-amylase activity (amylograph peak viscosity and falling number), wheat texture and α-amylase activity combined (gassing power and maltose), and storage conditions (fat acidity).

In some countries (e.g., UK) flour color is regarded as being a more reliable indication of the purity of flour than the ash test and is used in place of the latter in flour specifications. Flour color is also used in conjunction with the ash test to give a better indication of potential product quality, in particular flour whiteness and brightness. Measurement of flour color is considered particularly important in many Asian mills producing flour for various types of noodles where consumer preference is strong for a bright white product (e.g., dried noodles) or a bright, light yellow product (e.g., alkali 'Chinese' noodles).

Flour color can be determined with the Satake/Kent-Jones color grader which involves measurement of the reflectance of a flour-water slurry. Other common methods include the Agtron test, and the Minolta, or tri-stimulus test. The determination of tri-stimulus color values of flour or products thereof, using reflectance spectrometers, is now being widely used. Brightness (L*), red/green chromaticity (a*), and yellow/blue chromaticity (b*) can be rapidly and simultaneously measured.

Wet gluten is formed by hand or machine washing of flour in a dilute salt solution. The resulting viscoelastic product consists mainly of insoluble proteins which are of primary importance in determining dough rheological properties. Wet gluten content is closely related to flour protein content (in the ratio of approximately 3:1). The hand-washing method is used widely in many countries in place of protein content for flour specifications due to its simplicity and low cost. The yield or the 'feel' of the resulting gluten is also useful as an overall indication of inherent quality and as an indication of damage caused by proteolytic enzymes injected into the wheat kernel by the wheat bug (Cinch bug, Suni bug, etc., *Eurygaster* sp.).

About 30 years ago Perten Instruments introduced the Glutomatic instrument, as a mechanical washing method for determination of wet gluten content. The test method has been modified to provide a gluten index value. The gluten index value is becoming widely used as an indication of gluten (flour) strength. In this procedure, wet gluten prepared by machine washing of flour or ground grain is centrifuged through a sieve. The percentage of gluten remaining on the sieve after centrifugation, relative to the total gluten content, is defined as the gluten index value. High values (90–100) indicate strong gluten characteristics while low values (up to 30) indicate weak gluten characteristics.

Damaged starch is caused by the rupture and abrasion of starch granules during the milling process. The amount of damaged starch is closely related to the texture of the wheat. Hard-textured wheat gives flour with much higher starch damage than soft-textured wheat. The degree of flour starch damage, in addition to protein content, is a primary determinant of the ability of flour dough to absorb water (damaged starch absorbs approximately three times the amount of water compared to intact starch granules). This is

Table 3 Accuracy and precision of near-infrared analysis of flour

Parameter	r^2	*SEP*	*b*	*a*	*Bias*	*RPD*	*SET*	*CV%*
Moisture	0.990	0.119	1.017	−0.232	−0.006	9.71	0.025	0.25
Protein	0.992	0.132	0.990	0.102	0.009	11.43	0.044	0.38
Ash	0.880	0.0189	1.020	−0.0091	0.0005	2.86	0.012	2.40
Wet gluten	0.908	1.33	0.956	1.17	0.322	3.51	0.79	2.50
Starch damage	0.947	0.318	0.997	0.019	−0.003	4.64	0.057	0.64

SEP, standard error of prediction; *b*, slope; *a*, intercept; RPD, standard deviation of reference data/SEP; SET, standard error per test; CV%, coefficient of variability.

particularly important in bread production where high water absorption is desirable and in the production of 'soft' wheat products such as cakes and cookies where low water absorption is desirable.

Damaged starch is also important in the baking of leavened products since, due to its susceptibility to α-amylase attack, it provides a source of sugar to the yeast. Damaged starch is measured by digestion with an excess of amylases (α and β) followed by measurement of the resulting released sugar. The Tripette & Renaud (Chopin) model RT instrument is also used for the determination of starch damage. The Model RT employs an iodine reaction for the prediction of starch damage.

Flour particle size (granulation) is also strongly influenced by wheat texture. Harder wheat gives larger average particle size than softer wheat. Milling conditions may also influence particle size (in particular, the choice of sieves). Finer flours generally absorb water faster than coarse flours. Flour particle size is measured by a variety of techniques including sedimentation, sieving, light refraction, and light scattering.

Enzymes present in flour can have a major impact upon processing quality. The most important enzyme in flour is α-amylase. High levels of α-amylase are usually detrimental to processing quality, particularly in the case of bread products where dough becomes difficult to handle (sticky) and shows a tendency to collapse during baking. This is mainly attributable to the formation of dextrans, which are products of starch hydrolysis. High levels of α-amylase are most commonly associated with the presence of sprouted wheat in the milling grist. Very low levels of this enzyme may also be undesirable in bread flours since insufficient sugar will be released for gas production by yeast. α-Amylase activity in flour is normally estimated by measuring the starch-pasting characteristics of flours. The most common measurements include amylograph peak viscosity Hagberg (now Perten) falling number and the Rapid Visco-analyzer. In all three methods, lower values indicate higher enzyme activity. Where activity is low, sufficient α-amylase (from fungus or malted wheat or flour) may be added to give a predetermined amylograph peak viscosity equivalent to a falling number value of about 250 Brabender units.

The ability of α-amylase to produce sugars during fermentation is also dependent on the degree of starch damage, since intact starch granules are less susceptible to attack. Gassing power and maltose or diastatic value are measurements of the effects of these two factors. These measurements are commonly used to assess the ability of flour to maintain gas production by yeast and to make adjustments in α-amylase or sugar levels if necessary.

Other enzymes also have an impact on flour quality. However, in most cases differences in activity between different flours do not have a major impact on processing quality. Exceptions include proteases and certain oxidases. Endogenous flour proteolytic enzymes do not appear to have a major impact upon flour quality but, as noted earlier, injection of these enzymes by insects into the wheat kernel can cause extensive damage (weakening) of the gluten proteins. Polyphenoloxidases can cause browning reactions while lipoxygenases can bleach flour dough due to oxidation of pigments. High levels of the former enzymes, due to sprout damage or inherent variety characteristics, can cause product discoloration, particularly in the case of wet noodles. Lipoxygenases are often added to bread flours (as bean or pea flour) as a whitening agent.

During long-term storage of wheat and flour, a gradual deterioration in quality may occur, depending upon conditions. This is indicated by increased levels of free fatty acids, acid phosphates, and amino acids. Fat acidity shows the largest changes during the early stages of deterioration and is used most commonly to diagnose this process. The most common method of measuring fat acidity in flour involves extraction of free fatty acids with toluene followed by titration with dilute alcoholic alkali.

Additives

Ingredients may be added to flour by mills as improvers and to enhance the nutritional value of the flour. The most common improvers include bleaching agents to whiten the flour and oxidants which are added to improve the baking quality of bread flours. Soft wheat flours are sometimes chlorinated (to a specified pH value) for specific products.

Bleaching agents such as benzoyl peroxide are not measured since they react rapidly on contact with flour pigments. Oxidative improvers (added in the p.p.m. range) are measured by millers to insure that correct amounts have been added. The most commonly used oxidants include ascorbic acid, potassium bromate, and azodicabonamide (ADA). The use of the latter two oxidants has been banned in many countries due to health concerns, leaving ascorbic acid as the only widely accepted chemical additive for oxidation. This has resulted in the widespread use by the baking industry of oxidative enzymes obtained primarily from fungal sources to enhance or supplement the action of ascorbic acid. Ascorbic acid is quantified by measuring the decolorization of 2,6-dichlorophenolindophenol while bromate and ADA are measured by reaction with potassium iodide at low pH.

Vitamins and minerals may be added to flours to replace those removed with the bran and germ layers

during milling. In many countries the addition of these nutrients is mandatory. Vitamins of the B complex, including niacin, riboflavin, thiamin, and folic acid are the most important additives in white flours. Niacin can be determined by absorbance at 400 nm after reaction with cyanogen bromide. Riboflavin, thiamin, and folic acid are normally determined by spectrophotometric/fluorometric or microbiological methods.

The ratio of ash in wheat bran relative to the endosperm is approximately 20:1. Consequently, removal of the bran during milling also depletes the mineral content of the flour. Efforts to replenish essential minerals include the addition of calcium in the form of creta praeparata (calcium carbonate) and iron salts. Calcium content in flour can be determined volumetrically by ethylenediaminetetraacetic acid (EDTA) titration. Iron is determined colorimetrically.

Physical Dough and End-Product Tests

Physical dough tests are commonly carried out as part of flour analysis. They provide information on the viscoelastic properties of wheat flour dough which is closely related to flour functionality. The most widely used instrumentation includes the Brabender farinograph and extensograph, the mixograph, and the Tripette & Renaud (Chopin) alveograph.

The farinograph, a torque-recording mixer, is the most commonly used instrument. It measures the mixing properties of a flour-water dough system and provides information on the mixing time required to reach peak torque prior to dough breakdown (dough development time) and on the stability of the dough to overmixing (mixing tolerance index and stability). The instrument also provides an indication of water absorption potential since addition of water is adjusted until the peak torque during mixing reaches a defined value (500 Brabender units). This information is used extensively (usually in combination with protein content and flour ash) to meet bakery specifications for pan bread flours as well as a wide variety of other products (hearth breads, flat breads, steamed breads, noodles, chapattis). Flours with long dough development times (> 4 min) and long stability times (> 12 min) are considered as strong. Flours with development times of 2–4 min and stability times of approximately 4–8 min are considered to be of medium strength and those with shorter development times and stability times are considered as weak. Examples of each are shown in Figure 2.

The mixograph is also a torque-recording mixer that provides information on dough mixing time requirements and dough tolerance to overmixing. It is used extensively in the USA for flour strength classification. The increased use of personal computers (PCs) in laboratories has resulted in development of software which assists in the interpretation of these (farinograph and mixograph) dynamometer-powered instruments.

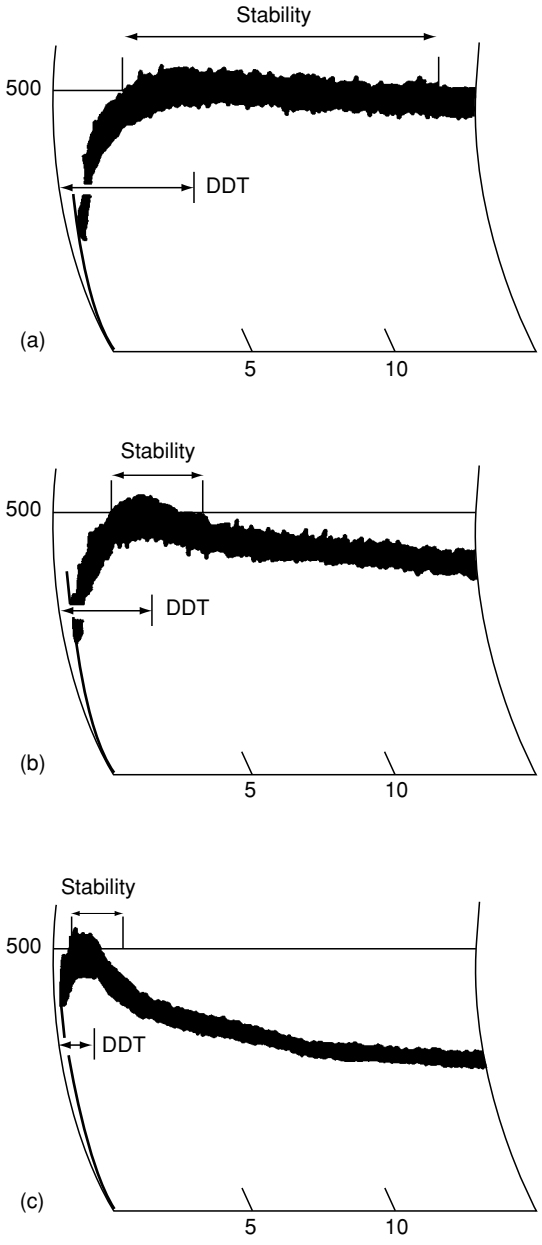

Figure 2 Farinograms of flours showing (a) strong, (b) medium, and (c) weak physical dough properties. Dough development time (DDT) and stabilities measured in minutes are shown as indicated. Reproduced from Flour, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

The extensograph provides a method of measuring the stretching properties of a flour–water–salt dough which has been mixed to defined conditions in the farinograph. The load-extension curve (extensogram) provides a measure of the maximum height

(resistance of dough to stretching), extensibility (length of dough at rupture during stretching), and area of the curve (energy required to stretch dough to breaking point). This method is used to assess flour strength and viscoelastic balance for bread production. The instrument has also been used to estimate the oxidation requirements of bread flours. Due to the large sample requirement (300 g flour) and the long time required to run the test, the extensograph is not so commonly used as it was previously in quality control applications.

The alveograph, like the extensograph, is used to measure the stretching properties of a flour–water–salt dough. Ingredients are mixed in the instrument's mixing chamber, extruded and cut into rounded disks which are clamped into a chamber through which air is pumped, resulting in the stretching of a dough 'bubble' until rupture occurs. The maximum pressure required to stretch the dough and the length (time required to rupture the dough) are obtained directly, while the *W* value (related to energy required to stretch the dough to rupture) and *G* (swelling index) are calculated from the curve. These values, especially the *W* value, are commonly used in specifications for 'French-style' hearth breads in Europe, the Middle East, and South and Central America. Alveograph parameters are also being used increasingly in soft wheat flour specifications (for cakes, cookies, biscuits, etc.).

An alveograph parameter which is receiving increasing attention is the ratio of the peak height to the length of the alveogram (the *P*/*L* ratio). It is related to the *W* value, in that, while a given *W* value can arrive by different combinations of *P* or *L*, the *P*/*L* ratio gives a better picture of the shape of the alveogram. A *P*/*L* ratio of about 1.0 is preferred for the baking of raised breads. Weak, extensible doughs, suitable for cake and cookie production, have *P*/*L* ratios of about 0.2, while *P*/*L* ratios of 2 or higher indicate that the dough is inclined to be 'bucky,' and lacks the desired degree of extensibility.

Detailed information on these instruments is provided in handbooks published by the American Association of Cereal Chemists.

Small-scale end-product tests are regularly carried out in laboratories to assess flour performance. These tests vary widely depending on the product and the local, regional, or national preference of the consumer. A discussion of this topic is beyond the scope of this article.

Health-related Tests

Six categories of toxic substances may affect wheat and wheat flour. These include pesticide residues; molds and mycotoxins, (the toxic metabolites of molds); toxic trace elements; foreign substances such as filth; radionuclides and microorganisms. Of all the types of pesticides that may be used in the production and storage of grain, which include mainly insecticides, herbicides, fungicides, and plant growth regulators, compounds used just prior to harvest and postharvest chemicals such as those used for fumigation and grain protection are the major risks for residues in wheat and flour. The presence of mycotoxins in wheat is associated with high levels of fungi which in turn are usually attributed to growing and harvesting conditions favorable to mold growth and poor storage conditions. The presence of unacceptable levels of toxic elements such as mercury, lead, arsenic, and cadmium in wheat may be due to a combination of factors which include the use of agricultural chemicals containing these substances, environmental contamination of soil, inherent variety characteristics, and natural soil conditions.

The presence of filth such as insect parts and eggs, excreta from insects and rodents, and rodent hairs in wheat and flour is usually the result of poor housekeeping, failure to maintain proper sanitary conditions in grain storage areas and mills, and ineffective grain cleaning. Radionuclide contamination of grain is possible from fallout resulting from nuclear accidents such as the Chernobyl incident in 1986. The presence of undesirable microorganisms at harmful levels in wheat and flour is also a matter of poor sanitation practices and poor grain or flour storage conditions.

Quality assurance of wheat moving into food channels to insure the absence of harmful levels of toxic substances is an ongoing process. It involves screening and control measures to identify contaminated parcels so that they can be removed from food channels and monitoring measures to confirm the absence of objectionable levels of toxic substances from commercial shipments. Insuring the safety of flour with respect to toxic substances is generally the responsibility of government health agencies and is usually accomplished through ongoing monitoring of products. Flour testing by mills and end users is carried out to insure that sanitary conditions are maintained and to meet specific requirements by processors.

Tests for pesticide residues, mycotoxins (especially from *Fusarium* spp.) and toxic elements are sensitive to parts per million or parts per billion levels. The analytical technique followed for determination of pesticide residues and mycotoxins will depend greatly on the requirements for sensitivity, selectivity, rapidity, and economy. For highly accurate and sensitive work, procedures using gas chromatography, high-performance liquid chromatography and mass

spectroscopy are usually followed for determination of pesticide residues and mycotoxins, whereas for toxic elements, atomic absorption spectroscopy is most often employed. Other procedures, involving the use of thin-layer chromatography, column chromatography and enzyme-linked immunosorbent assay (ELISA) test kits are also appropriate for many applications.

Tests for microorganisms in flour normally involve the extraction of the sample with sterile buffer, adding an aliquot of the extract at several dilutions to agar containing appropriate nutrients, and counting colonies after a specified growth period under fixed conditions. Controlled growing conditions and selective media can be used to aid the detection of specific organisms. Colony appearance, biochemical activity, and staining techniques are used in the identification process. Some of the more common tests include those for nonspecific aerobic microorganisms (standard plate counts for bacteria plus molds and yeasts), coliforms, fecal coliforms, *Salmonella*, *Staphylococcus*, *Bacillus cereus* and *B. subtilis* (the cause of ropiness in bread).

Insects and filth (including insect fragments and eggs, insect and rodent excrement) present in wheat are normally removed during the cleaning process prior to milling. However, insects and eggs present in the interior of the wheat kernel may not be removed by this process. Tests for filth are therefore carried out to measure the amount of insect material derived from the wheat and, more importantly, to insure that proper sanitary conditions are maintained in the processing plant. Flour specifications from secondary users may also require verification. The first step in identifying the presence of insects and filth involves visual inspection. Flour can also be examined directly by microscopy. To obtain more accurate results, filth is analyzed by several methods, involving wetting and digestion of the flour with pancreatic enzyme solutions or boiling 5% hydrochloric acid, sieving and filtering, followed by examination of the filter paper by microscopy.

See also: **Ascorbic Acid**: Properties and Determination; **Spectroscopy**: Near-infrared; **Starch**: Structure, Properties, and Determination; **Wheat**: Grain Structure of Wheat and Wheat-based Products

Further Reading

American Association of Cereal Chemists (2000) *Approved Methods of the American Association of Cereal Chemists*, 10th edn. St Paul, MN: The Association.

International Association for Cereal Science and Technology (1998) *ICC Standards; Standard Methods of the International Association for Cereal Science and Technology (ICC)*. Vienna: ICC.

Pomeranz Y (1988) *Wheat: Chemistry and Technology*, 3rd edn., vols I and II. St Paul, MN: American Association of Cereal Chemists.

Dietary Importance

D A T Southgate, Norwich, Norfolk, UK

Background

Flour is defined in most dictionaries as a term used for the ground material from cereal grains and related materials. In the context of Western Europe and most of North America, it most often implies the ground material from wheat, although the flour from other cereals, notably rye, maize, and rice, seeds of noncereals such as buckwheat, sorghum, and millets, and legumes such as soya bean are also sources of flours widely used in human nutrition. This article will focus most attention on the flours from wheat and only comment briefly on the nutritional contribution of the other flours.

The nutritional importance of flours derives first from a source of many nutrients and, second, from the many different types of foods that can be prepared that include flours of one kind or another as ingredients. The cultivation of cereals and the recognition that, when ground, they provide the basis of a wide range of foods from breads, where they are the major ingredients, to foods such as soups and gruel's, where they can be major or minor components, formed a key stage in the development of human civilization in the Middle East.

Mature cereal grains, as harvested from the plant, are hard and rather indigestible, unless they are steamed or soaked in water. Grinding or pounding the grain breaks the hard external coat and makes them more acceptable as foods. There is good archeological evidence for the development of grinding, initially between hand-operated stones, 'querns,' and later between mechanically operated millstones, and since the late nineteenth century between steel rollers to produce flours that greatly extended the range of foods in which cereals could be used.

Structure of the Cereal Grains

While the different cereal grains differ in detail in their structure, the principal components are common across all cereals and indeed most seeds. There is a

toughened outer seed coat, which includes the wall of the ovary in the grass flower.

In the wheat grain, these bran layers amount to about 8% of the grain weight. The seed contains the embryo or germ, which accounts for 2.5% of the wheat grain. The bulk of the grain is the endosperm, which contains the food reserve for the germinating grain. The cells of the endosperm have thin walls and are packed with starch grains as the reserve carbohydrate. Legume seeds also contain lipid as the energy reserve.

In the wheat grain, the endosperm accounts for about 82% of the grain. In the wheat grain, there is distinct layer of cells with thickened walls lying between the endosperm and the bran. This aleurone layer is rich in protein and B-vitamins, and contains most of the lipid in the grain. Grinding the grain breaks the endospermal cell wall and liberates the starch, which is the major component of cereal flours. In wheat, the rotating action of the millstones and rollers has the effect of separating the branny layer as a flake, which can be separated very easily from the edospermal material. The structure of the rye grain is similar, but in barley and rice, the outer layers are fused with the seed coat, and separation requires a different method. The other major cereal, maize, has a very tough coat that separates during pounding.

Production of Flours

Once the cereal has been ground, the ground material is usually fractionated to produce a range of types of flour. The initial grinding between conventional millstones produces wholemeal flour from which the different types of flour based primarily on the proportion of the whole grain present in the flour can be produced by sieving. Wholemeal wheat flour conventionally contains 98% of the starting grain (there is an allowance for cleaning the grain) and is said to have an extraction rate of 100%. Sieving removes progressively more of the branny and germ materials from the flour leading to flours of lower extraction rates where the proportion of endosperm is higher. Low-extraction flours are very fine and white, and are used in the production of some types of cakes. Bread-making flours have typical extraction rates of around 72%.

Roller milling initially separates the branny layers as a flake, and the particle size of the remaining grain material is progressively reduced to produce different streams of flours with different proportions of endosperm. The different types of flour and nominal extraction rates are formed by combining the different streams. Rye flours can be produced in similar ways, but most flours in use tend to be of the higher extraction rates. Corn flours are produced by sieving the ground cereal and, in many parts of the world, are consumed as high-extraction pounded flours. Cornflour is a very fine endospermal fraction.

Barley and rice grains have a different structure in that the outer layers of the seed coats are fused to the grain and cannot be removed by conventional milling. These grains are usually processed by blocking or 'pearling' the grain, in which the outer coats are removed with abrasive discs that grind away the outer layers to leave a grain that is made up mainly of endosperm. This is traditionally made into flour by pounding, although both barley and rice flours form minor proportions of the total use of these cereals.

Oats have a highly silicified seed coat and are typically prepared by steaming and rolling to produce groats rather than flours similar to wheat. The composition of some important flours is given in Table 1.

Table 1 Composition of some different flours (per 100 g)

Cereal	*Water*	*Protein*	*Fat*	*Carbohydrate*	*NSP*[a]	
Wheat	14.0	12.7	2.2	63.9	9.0	Wholemeal, mixed grist
Rye	15.0	8.2	2.0	75.9	11.7	Wholemeal
Barley	11.7	10.6	2.1	64.0	14.9	Wholemeal
Maize	12.0	9.2	3.9	73.7	4.4	Wholemeal
Rice	11.6	6.4	0.8	80.1	na	Low extraction
Oatmeal	8.9	12.4	8.7	72.6	6.8	Wholemeal
Soya	7.0	36.8	23.5	23.5	10.7	Full fat

[a]Nonstarch polysaccharides.
[b]Based on a number of sources. See Southgate DAT (1999) Cereals and cereals products. In: Garrow JS, James WPT and Ralph A (eds.) *Human Nutrition and Dietetics*, 10th edn, pp. 333–347. Edinburgh: Churchill Livingstone.
na, not available

Proximate Composition

Carbohydrates

Carbohydrates form the major constituent of all flours, and starch is the major carbohydrate present. In the unprocessed endosperm, the starch is contained within starch granules, which survive into the flours, although some starch granules are broken during milling. Most cereals contain the two forms of starch, amylopectin and amylose; amylopectin is usually the major component forming 75–82% of the total starch, but waxy maize contains virtually pure amylopectin, and there are also high-amylose maizes containing around 70% amylose, which gives them special physical properties that are important for use in some types of products.

Small amounts of free sugars are present in most flours, in wheat; sucrose is the major sugar, with traces of glucose and fructose. Maltose at low levels may be present, but high maltose levels are usually evidence of grains beginning to germinate at the time of harvest, which makes them less suitable for many processes.

Nonstarch Polysaccharides (Dietary Fiber)

The outer seed coats of the cereal grains contain plant cell wall material (dietary fiber), and cereals are of major nutritional importance in the provision of dietary fiber in the diet. The amounts of the nonstarch polysaccharides (NSP), which are the major components of plant cell wall material, depend on the extraction rate of the flour. Wholegrain flours contain the highest amounts, and the amount present falls as the extraction rate decreases. Table 2 illustrates the effects of extraction on the composition of wheat flours. The nonstarch polysaccharides of wholemeal wheat and rye are rich in the noncellulosic polysaccharides; of these, arabinogalactans have water-binding properties that are important for the physical characteristic of flours. Low-extraction flours contain mainly the cell walls from the endosperm, which are rich in cellulose. The noncarbohydrate lignin is also present in many seed coats; this makes the NSP from the higher extraction flours less digestible to the intestinal microflora of the large bowel and may be nutritionally important for increasing fecal bulk. The NSP from barley and oats are rich in β-glucans, which are water-soluble, giving viscous colloidal solutions that are nutritionally important in the products of these two cereals.

Table 2 Effects of extraction on the composition of wheat flours (per 100 g)

Flour, (extraction rate)	*Water*	*Protein*	*Fat*	*Carbohydrates*[a]	*NSP*[b]	*Energy (kJ)*
Wholemeal (100%)	14.0	12.7	2.2	63.9	9.0	1318
Brown (85%)	14.0	12.6	1.8	68.5	6.4	1377
Bread-making (72%)	14.0	11.5	1.4	75.3	3.1	1451
Patent (45%)	14.1	10.8	1.3	79.2	na	1480

[a]Available carbohydrates sum of sugar and starches.
[b]Nonstarch polysaccharides.
na, not available.

Lipids

Most of the lipid in the cereal grains is located in the germ, and small amounts are transferred to the endosperm component during milling. Oats are distinctive in this respect because they contain higher levels of lipid. The cereal lipids are typically unsaturated.

Proteins

Cereal grains contain a range of proteins, the amounts and types present varying with the variety, which determines the suitability of the flours for use in products. Wheats are classified into a number of types according to the amounts of types of protein present. North American and central European wheats contain higher levels of protein and are said to be 'strong'; these provide the flours that are most suitable for bread-making. British wheats tend to have lower protein contents and are 'weak' and less suitable for bread-making, although the mechanical development of the dough in the Chorleywood Process enables high proportions to be used in the mixed grists used in bread-making. 'Hard' wheats contain higher amounts of protein and differ from 'soft' wheats in their capacity to produce high-extraction and free-flowing flours. Wheats for bread-making are preferably hard and strong. Soft wheats are used in the production of biscuits and other products.

One important group of proteins in many cereals, especially wheat and rye, are the glutens. The glutens are insoluble proteins rich in sulfur amino acids, and they have the property of producing three-dimensional networks within the dough trapping the carbon dioxide formed by fermentation with the yeast; these are responsible for the formation of the characteristic texture of breads, which is a major factor in the nutritional importance in flours, especially in wheat-producing countries. Rye flours also contain a gluten that yields a less well-developed crumb structure. Rye breads tend to be more solid than wheat breads, and wheat/rye mixtures are especially important nutritionally in northern Europe. Barley does not produce breads in this sense and are more equivalent to cake-like structures as do oat flours.

The glutens are also nutritionally significant in being responsible for the intestinal lesions in 'gluten-sensitive enteropathy' and 'celiac disease.' The amino acid composition of cereal proteins is reasonably balanced, but lysine is one of the limiting amino acids that is partially destroyed in baking. Maize proteins are low in tryptophan, and pellagra has been observed in maize-consuming populations.

Vitamins

Cereals are good sources of some B-vitamins and poor sources of vitamin C and lipid-soluble vitamins. The vitamins are concentrated in the germ and the outer layers of the grain. Milling and the separation of the endosperm from the germ and bran lead to fractionation of the vitamins so that the concentrations are much lower in low-extraction white flours (Table 3).

Inorganic Nutrients

Flours, in common with most foods derived from plants, are rich in potassium and low in sodium. The production of self-raising flours increases the amounts of sodium and calcium, depending on the type of raising agent. Cereal flours are also good sources of phosphorus and reasonable sources of magnesium. Calcium levels are modest, and significant levels of iron and zinc are present. The inorganic nutrients are concentrated in the outer branny layers and in the germ, so that milling of the grain leads to the fractionation of the nutrients, and low extraction flours have much lower levels than the wholegrain flours. Table 4 illustrates the effects of extraction on some inorganic nutrients in wheat flours. The phosphorus in cereals is mainly present in phytic acid (*m*-inositol hexaphosphate), which is nutritionally important because it forms insoluble compounds with calcium, iron, and zinc, and thus reduces their bioavailability by preventing absorption in the small intestine.

Table 3 Effects of extraction rate on the B-vitamins in flours unfortified wheat flours (per 100 g)

Grade (extraction rate)	*Thiamin (mg)*	*Riboflavin (mg)*	*Niacin (mg)*	*Vitamin B_6 (mg)*	*Folates (μg)*
Wholemeal, 100%	0.47	0.09	5.7	0.50	57
Brown, 85%	0.30	0.07	1.7	0.30	51
Bread-making, 72%	0.10	0.03	0.7	0.15	31
Patent, 45%	0.10	0.02	0.7	na	na

na, not available.

Various sources modified from Southgate DAT (1999) Cereals and cereals products. In: Garrow JS, James WPT and Ralph A (eds.) *Human Nutrition and Dietetics*, 10th edn, pp. 333–347. Edinburgh: Churchill Livingstone.

Table 4 Effects of extraction rate on the inorganic constituents in wheat flours (mg per 100 g)

Grade	*K*	*Ca*	*Mg*	*P*	*Fe*
Wholemeal	360	35	140	340	4.0
Brown	280	20	110	270	2.5
Bread-making	130	15	36	130	1.5
Patent	100	15	19	89	1

Fortification increases the calcium values by about 120 mg and iron by about 0.7 mg. Modified from Southgate DAT (1999) Cereals and cereals products. In: Garrow JS, James WPT and Ralph A (eds.) *Human Nutrition and Dietetics*, 10th edn, pp. 333–347. Edinburgh: Churchill Livingstone.

Fortification with Inorganic Nutrients

At the start of World War II in 1939, the UK authorities were concerned that the supply of dairy products might be reduced by enemy action and that the proportion of cereal products in the diet would have to increase. The extraction rate of flours would also have to be increased to divert more cereal imports to human foods. These changes in the diet would accentuate the effects of phytic acid on the calcium in the diet. Regulations were introduced to implement fortification of all wheat flours, other than wholemeal, with calcium. After the war, there was a delay in returning to the low-extraction (white) flours for bread-making, but in 1953, when the regulations were relaxed, it was decided that the flour should be fortified with iron, thiamin, and niacin to maintain the concentrations of these nutrients at the levels they had been in the higher-extraction flours used in war time and the immediate postwar period. Although the primary nutritional reasons for fortification of wheat flours have largely disappeared, the fortification continues in the UK so that the calcium, iron, thiamin, and niacin levels in the UK flours are higher than those in most other countries.

Fortification with other nutrients has been considered recently in the UK. These include vitamin D in flours used primarily by immigrant communities in order to counteract the increase in the incidence of rickets in those communities. However this fortification was never introduced because of the difficulty in restricting the fortified flours to those for whom it was intended. Fortification with folic acid has been proposed as a prophylactic measure for reducing the incidence of neural-tube defects. The latter fortification is being implemented in the USA.

Losses of Nutrients on Storage of Flours

Provided the moisture content is below 15 g per 100 g, the nutrients in flours are relatively stable at cool ambient temperatures. Losses due to infestation with mites can occur on prolonged storage, especially at higher temperatures. The lipids in flours slowly oxidize at ambient temperatures, and rancidity can develop in wholemeal flours, which therefore have shelf-lives of about 3 months compared with about 9 months from brown flours and 2–3 years for white flours in paper bags.

Nutritional Contributions of Flours

The major nutritional contributions of flours to the human diet lie in the wide range of products in which flours are components. The range of products derived from flours in the UK diet, for example, is very large, ranging from breads and biscuits where flour is the primary ingredient, to its use in the preparation of pastry for meat, fish, vegetable, and fruit dishes, pizza bases, and pasta. Flours are used as thickening constituents in soups and many other dishes.

This widespread use means that flours are an important source of energy in the diet. Their contribution to protein intake is also significant because the proportion of protein to energy approaches many dietary recommendations. Although the conventional approach is to view cereals as energy sources alone, it is important to recognize that the current dietary recommendation for the proportion of energy from starchy carbohydrates in the diet could not be met without the contributions from flour. High-extraction cereal flours and the products made from them are significant sources of nonstarch polysaccharides.

See also: **Bread**: Dietary Importance; **Celiac (Coeliac) Disease**; **Cereals**: Contribution to the Diet; Dietary Importance; **Cobalamins**: Properties and Determination; Physiology; **Milling**: Principles of Milling; **Riboflavin**: Properties and Determination; Physiology; **Starch**: Structure, Properties, and Determination; **Thiamin**: Properties and Determination; Physiology; **Vitamin B_6**: Properties and Determination; Physiology

Further Reading

Department of Health and Social Security (1981) *Nutritional Aspects of Bread and Flour. Report on the Panel on Bread and Flour of the Committee on Medical Aspects of Food Policy. Reports on Health and Social Subjects No. 23*. London: HMSO.

FAO/WHO (1998) *Carbohydrates in Human Nutrition. Report of a Joint FAO/WHO Expert Consultation*. Rome: FAO.

Food Standards Agency (2001) *McCance and Widdowson's The Composition of Foods*, 6th edn.

Holland B, Unwin ID and Buss DH (1988) *Cereals and Cereal Products*. Nottingham, UK: Royal Society of Chemistry and Ministry of Agriculture Fisheries and Food.

Holland B, Welch AA, Unwin ID *et al*. (1991) *McCance and Widdowson's The Composition of Foods*, 5th edn. Cambridge: Royal Society of Chemistry and Ministry of Agriculture Fisheries and Food.

Kent NL (1983) *Technology of Cereals*, 3rd edn. Oxford: Pergamon Press.

Southgate DAT (1999) Cereals and cereals products. In: Garrow JS, James WPT and Ralph A (eds) *Human Nutrition and Dietetics*, 10th edn, pp. 333–347. Edinburgh: Churchill Livingstone.

Flow *See* **Rheological Properties of Food Materials**; **Sensory Evaluation**: Sensory Characteristics of Human Foods; Food Acceptability and Sensory Evaluation; Practical Considerations; Sensory Difference Testing; Sensory Rating and Scoring Methods; Descriptive Analysis; Appearance; Texture; Aroma; Taste

Fluidized-bed Drying *See* **Drying**: Theory of Air-drying; Drying Using Natural Radiation; Fluidized-bed Drying; Spray Drying; Dielectric and Osmotic Drying; Physical and Structural Changes; Chemical Changes; Hygiene; Equipment Used in Drying Foods

Fluorescence Spectroscopy *See* **Spectroscopy**: Overview; Infrared and Raman; Near-infrared; Fluorescence; Atomic Emission and Absorption; Nuclear Magnetic Resonance; Visible Spectroscopy and Colorimetry

FLUORIDE

A Hardisson, M I Rodríguez and A Burgos, University of La Laguna, Canary Islands, Quebec, Spain

Background

Fluorine is the most electronegative of all the known chemical elements and the most reactive. Its chemical characteristics make it one of most physiologically active ions. The interest in this element undoubtedly comes from the observations made in the 1980s of the toxic and therapeutic effects of fluorine, especially on teeth. In high doses, it can cause disorders in enamel formation, leading to *mottled teeth* or *dental fluorosis*, and in relatively small doses, it can reduce the prevalence of dental caries by over 50%. Dental caries are frequently suffered by the population, even in the most developed countries, and fluorides remain the best remedy. In fact, it has been shown that there is a correlation between communities that have consumed water containing moderate levels of fluoride (around 1.5 mg l^{-1}) and lower rates of dental caries. However, the relationship between consuming water containing high levels of fluoride ions and fluorosis has also been confirmed. Fluorosis is a disease that is characterized by 'mottled teeth,' and high intakes of fluoride are probably related to the appearance of kidney problems, leading to decalcification of the bones due to the flouride-sequestering quantities of Ca^{2+} that are essential for the correct development of the bones. Fluorine is a component of water that, according to the WHO, should be present in concentrations of around 1 mg l^{-1} to prevent dental caries, but if these levels exceed 1.5 mg l^{-1}, toxic effects start to appear, making the margin between essential and toxic levels very slim indeed.

Sources

The body's main source of fluorine is undoubtedly drinking water, as this is the cheapest and most efficient way of providing it, and fluoride plays an important role in human health. There are also traces of fluorine in almost all food products. The fluoride content in food is of great importance, as it can have both beneficial and prejudicial effects on our health when added to the quantities supplied by fluoride-treated water and toothpaste.

Man is naturally exposed to fluorides due to the increasing industrial use of these compounds, the increased use of fluoride toothpastes by the general population, and the treatment of water with fluoride as a fundamental weapon in the arsenal against caries in certain communities. Apart from water, there are other natural sources of fluorine to be found in the animal and vegetable worlds (tea is rich in fluorides); the atmosphere and even some medicines and cosmetics are occasional sources.

Dietary Intake

The level of fluorine in water is often used as a criterion to establish whether or not a population's diet is deficient in this element. However, in order to accurately determine fluoride intake, the contribution from food, both liquid and solid, the air, occasionally contaminated with fluoride either from industry or of volcanic origin, and the widely consumed products that have been treated with fluoride to fight caries should also be taken into consideration. The main tissues that are known to take up fluorine are bone and dental enamel, and uptake is proportional to total intake. Dietary recommendations are between 1.5 and 4 mg per person per day for adults, between 0.1 and 1 mg per person per day for the first year of life, between 0.5 and 1.5 mg per person per day for the next two years, and up to 2.5 mg per person per day for 12-year-old children.

Dental Caries

More than a decade passed after the discovery of the relationship between fluoride and endemic chronic dental fluorosis, before the beneficial influence of fluoride on the prevalence of caries was established. This discovery was the fruit of many years of research on dental fluorosis.

Later studies have shown that the addition of fluoride to the drinking water in areas of low natural concentrations is a practical and efficient method of reducing the incidence of dental caries. The recommendations approved by health organizations, therefore, suggest fluoride concentrations of between 0.7 and 1.2 mg l^{-1}, depending on the average local temperature. Dental health has also improved in recent years, even in areas with low concentrations of fluoride in drinking water. This is probably due to increased intakes of fluoride from other sources, such as food processed with fluoride-enriched water, external applications of fluorine by orthodontists, fluoride supplements, and even unintentional intake of fluoride-treated toothpaste.

Content in Waters and Liquid Food Products

The fact that fluoride is an element that has a beneficial effect on human health in certain dosages, but can become toxic to man in slightly higher concentrations, highlights the need for an exhaustive evaluation of the total quantity of fluoride consumed by a person in a specific community in a given geographic area. This quantity is affected both by community habits and by climate conditions, fundamentally by average temperatures (especially important in areas in which the public water supply has been treated with fluoride).

Methodology

One of the problems to be considered is the choice of the most suitable method for determining fluoride contents. The most widely used techniques include: gravimetry (precipitation as PbFCl, precipitation as calcium fluoride) and volumetry (determining the soluble fluoride and the combined fluoride), indirect colorimetry (SPADNS method, quercetin–zirconium system (III), zirconium–alizarin red system, zirconium–chromium cyanine system, zirconium–alizarin sulfonate system), direct colorimetry, potentiometric methods (based on the use of a selective fluoride ion electrode), and spectrofluorimetric methods. In some cases, especially for complex matrices, prior separation of fluorides may be necessary. The most widely used methods of separation include fluorine distillation, ion exchange, HF microdiffusion, and interference ion precipitation.

Water

Fluoride in drinking water is commonly the main source of total fluoride intake. The intake of F^- from drinking water will depend fundamentally on: (1) the concentration of F^- in the water, (2) the age of the person, (3) weather conditions (temperature), and (4) eating habits. Table 1 shows typical average fluoride concentration in drinking waters obtained in different geographic areas and at different times. Although the contents are variable in general, there are some waters in India that have a high fluoride content, similar to the levels of the north Tenerife (Spain) water supply in the health district of Icod de los Vinos. Based on the average concentration intervals and an average water consumption of 2 l per day, the fluoride intake would be between 5 and 11.1 mg of fluoride per liter per person per day, which exceeds dietary recommendations and explains the fact that this is an area with a high degree of dental fluorosis.

Table 2 shows the average concentration in the most commonly consumed domestic and foreign mineral waters in Spain, at different times in the same geographic area (the island of Tenerife). Some of these concentration values are above the recommended level of $1\,mg\,l^{-1}$, but none is above the maximum ceiling of $1.5\,mg\,l^{-1}$.

Wine

Small concentrations of fluoride can be found naturally in wine, but these quantities do not generally exceed the maximum limit of $1\,mg\,l^{-1}$ recommended by the International Vine and Wine Bureau. In fact, the fluoride contents found in most research work done in this field are between 0.1 and 0.9 mg of fluoride per liter. The natural fluoride content of wine, therefore, cannot be classified as unhealthy in normal circumstances. The fluoride gets into the wine from the fluoride contained in the irrigation water, so

Table 1 Average fluoride content in drinking waters of different geographic areas

Geographic area	*Analytic method*	*Mean concentration ($mg\,l^{-1}$)*	*Year*
Portugal (Ribeiráo Prieto)	Ion-selective electrode	0.1–1	1990
India (Visakhapatnam)	Ion-selective electrode	0.31–8.35	1997
Spain (Basque Country)	Ion-selective electrode	Álava: < 0.05–11.12 Guipúzcoa: < 0.05–0.26 Vizcaya: < 0.05–0.45	1984
Spain (La Rioja)	Ion-selective electrode	0.22–1.1	1986
Spain (Tenerife)	Colorimetry	6% < 1.5	1985
	SPADNS and ion-selective electrode	64% ≤ 0.5–1.5 20% < 0.5	
Spain (Tenerife)	Ion-selective electrode	3.90	1988
Spain (Tenerife)	Ion-selective electrode	4.22	2001
Kenya	Ion-selective electrode	0.28	1995
Poland	Ion-selective electrode	1.2	1996
Spain (Soria)	Ion-selective electrode	0.2	1999

Table 2 Average fluoride concentration in bottled waters consumed in Spain of different origins

Geographic area	*Analytical method*	*Mean concentration ($mg\,l^{-1}$)*	*Year*
Spain (Mainland)	Ion-selective electrode	1.77	1976
Spain (Tenerife)	Ion-selective electrode	0.66	1989
Spain (Tenerife)	Ion-selective electrode	1.47	1990
Spain (Tenerife)	Ion-selective electrode	1.1	1995
Spain (Tenerife)	Ion-selective electrode	1.09	2001

Table 3 Average fluoride concentration in wines from different areas of Spain

Geographic area	*Analytical method*	*Mean concentration ($mg\,l^{-1}$)*	*Year*
Spain (Mainland)	Ion-selective electrode	0.83	1971
Spain (Mainland)	Ion-selective electrode	0.13	1979
Spain (Jerez)	Ion-selective electrode	0.31	1983
Spain (Mainland)	Ion-selective electrode	0.28	1988
Spain (Tenerife)	Ion-selective electrode	0.20	1990
Spain (Mainland)	Ion-selective electrode	0.25	1990
Spain (Tenerife)	Ion-selective electrode	0.20	1996
Spain (Tenerife)	Ion-selective electrode	0.24	2000

when it is found in higher concentrations, one can assume that there has been some form of contamination, either natural (soil or water that is excessively rich in fluoride) or accidental, with two main sources. The first source would be the use of cement vats for making the wine and/or for storing the wine; the second source of contamination would come from the direct addition of fluoride-enriched compounds like antiseptics or antifermentation agents. Although their use is banned in many countries, they are still used, because it is known that, even when diluted to between 1:400 and 1:5000, fluoride still presents a marked antiseptic action.

Another source that must be added to the two main sources is the increasing use of kryolite (sodium fluoroaluminate) in the USA as an insecticide to control certain pests that affect the vineyards. This new source of pollution raises the fluoride content in the wine produced from these grapes to 3 mg of fluoride per liter, which, in principle, makes it necessary to review wines that were not initially considered to have a high fluoride content. On this subject, mention must be made of the *wine fluorosis* discovered in Barcelona in 1965. A joint study by the Departments of Medical Pathology, Bromatology and Toxicology of the University of Barcelona demonstrated that the etiology of some *periostitis deformans* was the consequence of consuming wine with high fluoride concentrations. Situations of this kind require not only a high intake of fluoride, but also the nutritional deficits typically exhibited by alcoholics, making it more probable that consumers of wines containing added fluorides will suffer from this kind of disease. As fluoride has a marked preference for bone, it causes alterations in the skeleton, as well as inhibiting some important enzymes (phosphoenolpiruvase, phosphotases, esterases, and even the cytochrome oxidase system).

According to several results (Table 3), the fluorine contents of wines of different vintages studied exceed neither the established limit of $1\,mg\,l^{-1}$ nor the recommended daily intake.

Beers, Soft Drinks, and Juices

Table 4 gives the fluoride contents in the most popular beers and soft drinks. These results indicate that the fluoride content found comes mainly from the water used in the manufacturing process. According to the literature, the F^- content in beers is low, around 0.3–0.8 mg l^{-1}, unlike wine, which can reach 6–$8\,mg\,l^{-1}$, due to the reasons mentioned in the section on wine; because it is stored in cement vats, and because of the use of compounds with added fluoride like antiseptics and antifermenting agents. With regard to soft drinks, the drinks made from carbonated water have the highest fluoride content ($1.05\,mg\,l^{-1}$), although they do not reach the permitted maximum of $1.5\,mg\,l^{-1}$. Bottled natural fruit juice, however, has a lower content.

Milk and Derivatives

Milk and dairy products are consumed universally. Their importance in our diet has been accepted for thousands of years. Table 4 lists literature data on the fluoride content in milks and derivatives. The values detected in dairy products are below $0.4\,mg\,kg^{-1}$, the recommended ceiling for formula for children, so

their consumption by the child population does not imply any toxicological risk for them. Furthermore, one must remember that some kinds of milk have been enriched with fluoride.

Tea

Tea leaves also have a high fluoride content. These values are above 400 mg kg^{-1} (dry weight), but tea infusions (tea-bags) have concentrations of around 0.5–1.5 mg l^{-1}. The amount of F^- present in a cup of tea will depend not only on its volume but also on the brand of tea, the amount of tea used, how long it is brewed for, whether or not it is a dilution of a previously prepared cup of tea, and whether or not it has been made with fluoride-enriched water. The F^- intake of tea drinkers can vary between 0.04 and 2.7 mg day^{-1}. Table 4 shows the fluoride contents of tea.

Fluorides in Solid Foods

Fluorides can be detected in all food. There does not seem to be a clear relation between the F^- content of most plants and the content found in the soil or the water in the area where the plants are grown. Plants grown in acid soil, however, have a higher concentration of F^-. The F^- content found in the leaves of most plants is between 2 and 120 mg of F^- per kilogram. The fruit and vegetables consumed by man have a fluoride content of around 0.1–0.4 mg kg^{-1} and, therefore, are of little interest, unless there is a prevalence of unusual dietary patterns, although F^- concentrations of up to 2.0 and 2.1 mg kg^{-1} have been found in barley and rice irrigated with nonfluoride-treated water, and concentrations of up to 4.3 and 6.4 mg of F^- per kilogram, respectively, when the water used was fluoride-enriched. Yams and cassava also make up the main diet of many tropical regions, and they contain high levels of F^-. However, fluoride contents in meat tend to be low. Fish products (tinned fish like salmon, sardines) have fairly high contents, up to 40 mg kg^{-1}. However, the consumption of fluorides from fish alone, in a combined diet, does not provide enough fluorides to exceed 0.2 mg day^{-1}.

The amount of the total fluoride intake accounted for by solid foods has been considered as low, although this can vary, depending on the composition of the daily diet and the fluoride content of the water where the food has been prepared.

These quantities of fluorides in food can be increased by the use of pesticides with added fluoride and phosphated chemical fertilizers, and because of the fluoride to be found in the irrigation and washing waters. Furthermore, one must consider the fluoride contribution made by industrially processed foods.

Table 5 Lists the most significant data collected on the fluoride contents of solid foods. It can be seen that the contents are very low, with the exception of fish products.

Table 4 Fluoride ion content in beers, soft drinks, juices, milk, and tea

Drink	*Analytical method*	*Mean concentration (mg l^{-1})*
Beers (1990)	Ion-selective electrode	0.51
Beers (1992)	Ion-selective electrode	0.16
Bitter soft drinks	Ion-selective electrode	0.21
Orange soft drinks	Ion-selective electrode	0.30
Lemon soft drinks	Ion-selective electrode	0.41
Cola soft drinks	Ion-selective electrode	0.34
Carbonated soft drinks	Ion-selective electrode	1.06
Natural juices	Ion-selective electrode	0.18
Mother's milk	Ion-selective electrode	0.18
Pasteurised milk	Ion-selective electrode	0.10
Evaporated milk	Ion-selective electrode	0.04
Cows' milk (raw)	Ion-selective electrode	0.26
Powdered milk	Ion-selective electrode	0.11
Cream cheese	Ion-selective electrode	0.04
Smoked cheese	Ion-selective electrode	0.27
Tea-leaves	Ion-selective electrode	148.03 mg kg^{-1}
Tea-bags	Ion-selective electrode	0.84

Table 5 Fluoride content of solid foods

Solid food	*F^- content (mg kg^{-1})*
Cereals and grain	
Wheat germ	0.12
Oats	0.12
Rice	0.05
Precooked rice	0.05
Chick peas	0.08
Fresh vegetables	
Beetroot	2.19
Cucumber	0.34
Silver beet	0.25
Carrot	0.06
Garlic	0.15
Cabbage	0.07
Aubergine	0.23
Tinned products	
Guayaba purée	0.02
Mango purée	0.02
Guayaba jam	0.23
Pepper purée	0.17
Pastas	
Macaronis	0.14
Noodles	0.12
Spaghetti	0.09
Coditos	0.13
Fish	
Tuna	3.73 mg l^{-1}
Anchovy, fillets	3.02 mg l^{-1}
Sardines in oil	4.27 mg l^{-1}

See also: **Beers**: Chemistry of Brewing; **Dental Disease**: Etiology of Dental Caries; Fluoride in the Prevention of Dental Decay; **Milk**: Analysis; **Tea**: Chemistry; **Water Supplies**: Chemical Analysis; **Wines**: Types of Table Wine

Further Reading

Hardisson A, Rodríguez MI, Burgos A, Díaz Flores L, Gutiérrez R and Várela H (2001) Fluoride levels in public supply and bottled drinking waters in the island of Tenerife (Spain). *Bulletin of Environmental Contamination and Toxicology* 67(2): 163–170.

Kjellevold Malde M, Maage A, Macha E, Julshamn K and Bjorvatn K (1997) Fluoride Content in selected food items from five areas in East Africa. *Journal of Food Composition and Analysis* 10: 233–245.

Martín M, Hardisson A and Álvarez R (1992) Fluoride concentration in beers and soft drinks. *Journal of Food Composition and Analysis* 5: 172–180.

Martín MM, Hardisson A and Álvarez R (1993) Direct potentiometric determination of fluoride in beverage. Comparative study of different buffering solutions. *Food Chemistry* 46: 85–88.

Pérez-Olmos R, Hardisson A, Elías I, Ríos R and Martín M (1990) Determination of fluoride in Tenerife wines. Comparative study of conditioning solutions. *Belgian Journal of Food Chemistry and Biotechnology* 45(6): 210–214.

Sanz Pérez E and Sanz J (1999) Fluoride concentration in drinking water in the province of Soria (central Spain) and caries in children. *Environmental Geochemistry and Health* 21: 133–140.

Foaming *See* **Aerated Foods**

FOLIC ACID

Contents

Properties and Determination

C J Bates, MRC Nutrition Research, Cambridge, UK

Introduction

In the older literature, it was customary to use the terms 'folic acid,' 'folate,' or 'folacin' interchangeably for all of the pteroylglutamic acid derivatives and their γ-polyglutamic acid conjugates. More recently, the custom has changed so that 'folic acid' is now reserved for the parent and commercially available pteroylmonoglutamic acid which is in the oxidized state, whereas 'folate' is used to describe the derivatives which typically occur in foods and tissues. The latter occur in many forms, which may be reduced; they often carry one-carbon substituents and are often poly-γ-glutamates, especially inside cells. The advantage of this new practice is that it permits a shorthand distinction between the commercial folic acid that is present in supplements and is used in food fortification, and the folate that is native in living tissues and in natural foods, and which has a different bioavailability.

Physical Properties

The fully oxidized parent substance of the folate or folacin group, pteroylglutamic acid or pteroylmonoglutamic acid (formula weight 441.4) crystallizes from hot water as yellow-orange platelets. It does not melt, but darkens and chars at about 250 °C. It is only very slightly soluble in cold water (1–10 mg l^{-1}), increasing to 1% in boiling water, and is much more soluble in aqueous alkali, acetic acid, phenol, pyridine, and other basic solvents. A solution for injection is obtained by dissolving it in aqueous sodium bicarbonate, or by using the sodium salt, which is soluble in water. The extinction coefficient ($E^{1\%}_{1\,cm}$) in 0.1 mol l^{-1} alkali is 565 at 255 nm, 350 at 282 nm, and 195 at 365 nm (absorption maxima).

5-Formyl-5,6,7,8-tetrahydropteroylglutamic acid (folinic acid: formula weight 473.4) also decomposes without melting at 250 °C; it is sparingly soluble in water, but is more soluble in aqueous alkali, and the

absorption maximum is at 282 nm in alkali. It is more stable at neutral or slightly alkaline pH than at acid pH, in solution. The 10-formyl derivative is more labile, and more sensitive to oxygen.

Tetrahydrofolic acid (formula weight 443.4) is extremely readily oxidized, and the freeze-dried solid form must be stored *in vacuo*. Solutions in 0.5% ascorbate or in 1.0 mol l^{-1} mercaptoethanol are moderately stable, and it has an absorption maximum at 298 nm in neutral solution.

5-Methyltetrahydrofolic acid (formula weight 459.4) is intermediate in stability to oxidization between pteroylglutamic acid and tetrahydrofolic acid. It can be obtained as a white powder, but in solution it requires ascorbate or another reducing substance to achieve medium-term stability. The absorption maximum at 290 nm has a molar extinction coefficient (ε) of 31.7×10^3 mol^{-1} cm^{-1}, or $E^{1\%}_{1\,cm}$ of 690, at neutral pH.

Chemical Properties

The structural formulae of three of the commonly encountered forms of folic acid are shown in Figure 1.

Pteroylglutamic acid and related pterins are readily hydrogenated in two stages to the dihydro and then the tetrahydro forms, involving firstly positions 7 and 8 of the pteridine ring, then positions 5 and 6. These hydrogenation reactions can be performed either *in vitro*, with hydrogen and a platinum oxide catalyst or by hydrosulfite, or else *in vivo*, by enzyme-catalyzed reactions involving dihydrofolate reductase. Once converted to the fully reduced form, single carbon units at various levels of oxidation can be transferred to the nitrogen atoms at positions 5 or 10, yielding methyl, methylene, methenyl, hydroxymethyl, or formyl units: the second and third of these form a bridge between the two acceptor nitrogen atoms. A combination of a carbon and nitrogen atom (formimino group, –CHNH–) can also be carried. These transfer reactions constitute the fundamental basis of the single-carbon-unit transfer operations, for which the folic acid cofactors are essential *in vivo*. (*See* **Cobalamins**: Physiology; **Coenzymes**.)

Chemical syntheses of pteroylglutamic acid have been achieved, first by condensation of guanidine with ethyl cyanoacetate to give a pyrimidine ring, then continuing either by reacting *p*-aminobenzoylglutamic acid with 2,4,5-triamino-6-hydroxypyrimi-

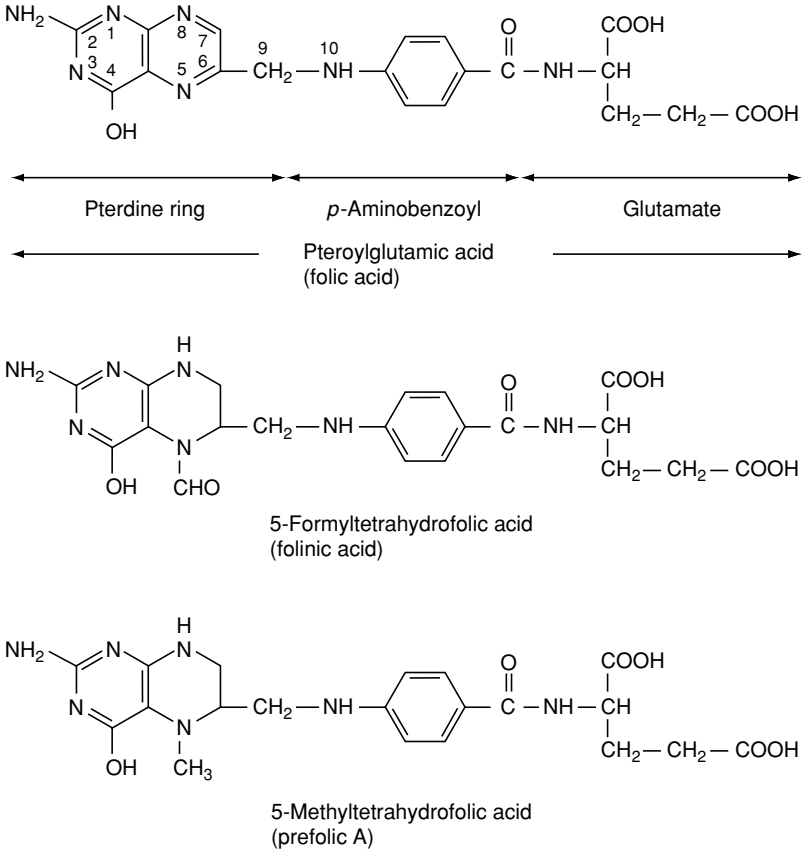

Figure 1 Structural formulae of commonly encountered forms of folic acid.

dine and α,β-dibromopropionaldehyde, or by reacting 2,4,5-triamino-6-hydroxypyrimidine with pyridine, α,β-dibromopropionaldehyde, potassium iodide, and *p*-aminobenzoylglutamic acid. For pteroylglutamic acid there is only one asymmetric center which is within the glutamic acid moiety (L form), but after reduction a second asymmetric center appears at carbon 6, and chemical synthesis produces both D and L forms, whereas the natural product has the L configuration. Degradation of folic acid *in vivo* involves an initial cleavage at the 9–10 bond.

Folate polyglutamates are the forms of folic acid (including the reduced, C_1-substituted cofactor forms *in vivo*) which have γ-linked chains of poly-L-glutamic acid. These chains can extend up to at least 10 L-glutamyl units, and different tissues differ in their spectrum of chain lengths. The glutamate residues are added by the enzyme, pteroylpolyglutamate synthetase (pteroylpolyglutamate ligase), and can be removed by pteroylpolyglutamate hydrolase (alias folate conjugase or γ-glutamyl carboxypeptidase). The latter enzyme is present, *inter alia*, in serum, in duodenal and jejunal mucosal cells and in kidney, especially within the lysosomes. The synthetase is an intracellular enzyme, responsible for building the active forms of the folate cofactors. Tetrahydrofolate and the formyl folates are preferred to methyltetrahydrofolate for the synthetase reaction, which helps to explain the poor retention of folate in vitamin B_{12} deficiency, when such nonmethylated folate species decline in concentration, leading to a change in the balance towards shorter-chain polyglutamates.

Occurrence in Food

A wide variety of forms of folate occur in food, and this complexity has resulted in major problems, both for devising accurate analytical methods to measure the folate content of foods as eaten, and in measuring the bioavailability of the different forms. As noted previously, the source of variability is threefold: (1) the oxidation state of the pteridine ring; (2) the presence and type of the one-carbon substituent carried, and (3) the presence and length of the polyglutamate side-chains. (*See* **Bioavailability of Nutrients**.)

Because monoglutamates and short-chain polyglutamates are more readily available for absorption than long-chain polyglutamates, an operational definition of 'free folate' was established, which included only the monoglutamates and short-chain polyglutamates that are most readily available to promote the growth of the test organism *Lactobacillus rhamnosus* (formerly *L. casei*). However, this category has not, in practice, proved to be very useful, because the proportion of 'free folate' can vary enormously with the history of the food sample, the assay conditions (especially the initial folate extraction, if endogenous conjugases are present), and with the duration of exposure to the assay microorganism. The original studies on this subject made use of polyglutamates derived from the oxidized form, pteroylglutamate, not those of the natural forms of folate, so that the results were not entirely applicable to the forms of folate in foods. (*See* **Lactic Acid Bacteria**.)

Table 1 shows the folate content of 'typical' (average) samples of food items as eaten. These folate values are still being revised, usually in an upward direction, as the assays are optimized. Nearly all of the existing food table folate values have been obtained by microbiological assay procedures.

It is clear that, although folate contents per 100 g vary considerably, folate is fairly widely distributed between different food types. Despite this, there are very considerable variations in folate content between diets which, on the one hand, depend heavily on unfortified cereal staples (traditionally 'poor' diets) and, on the other, include the richer plant and animal sources. Of the folates in a mixed diet, over 90% are in polyglutamates of varying chain length; 60% are methylated, and 30% have a formyl group. Foods containing substantial amounts of vitamin C also tend to be rich folate sources, partly because of the preservative effect of vitamin C on folates during storage, preparation, and cooking. (*See* **Ascorbic Acid**: Physiology; **Cereals**: Dietary Importance.)

In the UK, vegetables and cereals each provide about 30% of food folate intake; milk and milk products provide about 10%; meat, fruit, and beverages each provide about 5%; eggs, fish, and other foods each provide 1–4%.

Use in Food for Fortification

If the vitamin is added to manufactured foods, it is almost always as pteroylglutamic acid (folic acid), because this is by far the cheapest commercial form. Infant formulae usually contain added folic acid, in amounts between 3 and 15 μg 100 ml^{-1}, as fed. The practice of adding folic acid to manufactured foods is becoming much more common than it was, especially for breakfast cereals. In the USA, folic acid fortification of cereal grains for bread making has now become mandatory, at 1.4 μg g^{-1}, from the beginning of 1998. (*See* **Food Fortification**; **Infant Foods**: Milk Formulas.)

Extraction and Clean-up

Two problems dominate the procedures for extraction of folic acid from biological tissues: the prevention of

Table 1 The folate content of typical food items (μg 100 g^{-1} wet weight)

Cereals and nuts		*Dairy products*		*Fruit*	
Boiled rice	4–10	Milk	6	Oranges	30
Boiled spaghetti	4–7	Cream	7–12	Grapefruit	26
Wheat flour	20–50	Butter	trace	Apples	1–4
Peanuts	50–110	Cheese	15–100	Pears	2–4
Walnuts	70	Icecream	7–9	Plums	3
Coconut	9	Eggs	50	Strawberries	20
Cereal products		*Meat and fish products*		*Vegetables*	
Bread	15–90	Roast beef	15–17	Cabbage (boiled)	25–35
Biscuits	7–40	Roast chicken	7–13	Cauliflower (boiled)	50
Cake	7–100	Roast pork	6–7	Brussel sprouts	110
Breakfast cereals	2–400	Fried liver	200–600	Lettuce	55
Porridge	4–7	Fried kidney	80	Spinach (boiled)	90
Pastry	7–35	Sausage	2–4	Potatoes (boiled)	10–25
Custard	5	White fish	10–14	Carrots (raw)	12–30
		Fish fingers	16	Peas (boiled)	27

Data compiled from McCance and Widdowson's *The Composition of Foods*, 4th and 5th edns.

oxidation mainly by molecular oxygen, and the preservation, or alternatively the complete removal, of the polyglutamate chain. Prevention of oxidation can be achieved by a variety of reducing agents, the most common choice being ascorbic acid, usually at 1% w/v final concentration, either as the free acid or else partially neutralized, e.g., to a pH around 4.

Removal of the polyglutamate side chain can be achieved with partly purified preparations of folate conjugase (polyglutamyl hydrolase). The ones which are most commonly used for food folate release are the hog kidney enzyme and the chicken pancreas enzyme. The former hydrolyzes the polyglutamate chain and thus releases monoglutamates from the food folates more completely than the latter, which tends to produce diglutamates. Although different workers have generally taken care to optimize the assay procedures with their particular selected enzyme sources, it is important to consider carefully the implications of using an enzyme which may not liberate monoglutamates, because this can seriously affect the relative molar response of the standard (such as pteroylglutamic acid) against the more complex and variable folate species obtained from the food matrix. One study indicated that the folate content of foods as measured after deconjugation with the hog kidney enzyme can be at least 50% higher than the values achieved after deconjugation with chicken pancreas enzyme. A 'triple enzyme' mixture, comprising chicken pancreas conjugase plus α-amylase and a protease, is widely used.

The release of monoglutamyl folates from the polyglutamates in red cell extracts has generally been achieved by incubating the red cell lysates in the presence of their own plasma, because plasma contains a very active conjugase. This is why red cell folate assays in whole blood, rather than separated red cells, are preferred, since the cellular components contain little or no conjugase. Most of the published procedures for the red cell folate assay have assumed that the monoglutamates are completely released after a brief period of incubation with plasma enzyme. Another assumption, highlighted recently, that is made with the assay of folate in red cell hemolysates, is that it is unaffected by the extent of binding to hemoglobin or the extent of physical trapping, which varies with the degree of oxygenation of the sample. Although red cell folate is a preferred index, because it measures long-term status, there are still serious practical problems in achieving accurate measurements and in obtaining acceptable inter-laboratory agreement.

If the polyglutamyl folates are to be preserved, in order to measure the polyglutamyl profile that is characteristic of the undamaged tissue, then adventitious conjugase activity must be inactivated at the time of extraction. One way of achieving this is by using a boiling ascorbic acid solution during homogenization.

If extremely labile folates, such as tetrahydrofolic acid, are to be extracted and preserved intact, then special precautions such as the exclusion of oxygen during the extraction procedure, and the use of low temperatures, may be necessary.

Analytical Procedures

Once the folates have been extracted from the tissues, there are three main approaches to the problem of their estimation and characterization. The oldest general procedure depends on the extent of growth of a specific folate-dependent bacterial species. Its growth

medium contains optimum amounts of all the necessary growth factors except folate, and a small amount of the food extract is then added, to provide sufficient folate for partial suboptimum growth of the test organism. The second general approach, which has now been available for nearly three decades, depends on the specific interaction of the extracted folate with a fixed amount of a folate-binding protein plus radioactive- or fluorescence-labeled folate, after which the protein-bound folate is separated from the unbound folate and the radioactivity or fluorescence of the bound folate is measured. The amount of label bound depends on the total amount of folate present. This approach can be subdivided into those assays which depend on naturally occurring folate binders obtained from body fluids such as milk, and those which depend on binders which are artifically induced in an animal, for instance by using folate bound to a hapten as an antigen. (*See* **Immunoassays**: Radioimmunoassay and Enzyme Immunoassay.)

The third general approach is the separation of different folate species from each other and from interfering substances, using high-performance liquid chromatography. This approach has the advantage of specificity, but a disadvantage for tissue folate analyses is that the available detectors (ultraviolet absorption, fluorescence, or electrochemical) are barely sensitive enough to detect all of the folates found in tissues, and they do not always give a constant molar response with different folate species, which complicates the problem of calibration. Nevertheless, new data and further refinements continue to emerge. (*See* **Chromatography**: High-performance Liquid Chromatography.)

The microbiological assay procedures have been established for long enough to dominate the literature, for instance, on food analysis and food table compendia. Total folate has, for several decades, been measured using *L. rhamnosus* (previously *L. casei*), because this organism can respond equally well to all the common monoglutumate forms of folate. However, the use of inappropriate assay conditions, e.g., a poor choice and control of pH, has often resulted in the underestimation of total folate contents. Until this source of error has been effectively eliminated, the food table values for folate and the derived estimates of human dietary intakes of folate may be too low. Revised values are gradually being obtained to rectify this problem.

Two other microorganisms have been used, especially in the earlier studies of folate speciation. One is now called *Enterococcus hirae* (formerly *Streptococcus faecalis*). This organism grows on most forms of folate except the 5-methyl derivative. The other is called *Pediococcus pentosaceus* (formerly *P. cerevisiae*, or *Leuconostoc citrovorum*), and grows most readily on 'citrovorum factor,' i.e., N_5-formyltetrahydrofolic acid, or 'folinic acid'. Differential growth of these organisms on tissue extracts has given useful estimates of the different monoglutamate forms of folates. However, it is important to note that some tissue or food extracts contain growth inhibitors, or promoters, which may result in errors in the estimation of folate levels, and species ratios.

The binding protein methods have recently proved very popular, especially for blood status assessment, because of their ease of operation, but they too can suffer from errors. One source of error, mentioned earlier, may be the incomplete deconjugation of polyglutamate side chains; another may be the partial oxidation of folates to their degradation products. Like the microbiological assay, specific binding assays can also suffer from pH-dependent variations in response to different folate species, so that the choice of calibrant and of pH may be critical. There are many variants, using different binding proteins or antibodies; some depend on the binding of radioactive- or fluorescence-labeled folate markers; others depend on enzyme-linked immunosorbent assays. However, there is a constant need for interlaboratory methodological comparisons to achieve good comparability of results, and a careful study of the sources of error in these assays, and of procedures to insure their elimination. External quality assurance schemes are available for folate (and for vitamin B_{12}) assays in serum or plasma, and to a lesser extent for folate in red cells, because these are widely performed, especially by hospital hematology laboratories.

Those studies which have made the most use of high-performance liquid chromatography have been metabolic studies in animals, in which the intertissue distribution, turnover, and speciation of folates have been followed, using radioactively labeled precursors. The use of stable isotope-labeled folates for studies of folate economy in human subjects has recently become feasible, and a number of new studies have begun to examine the relative bioavailability of different folate congeners, with and without a food matrix, by this approach. Urinary excretion of undegraded folates and of the breakdown products derived from *para*-aminobenzoic acid has provided suitable material for isotope ratio analysis. Undoubtedly this approach will expand and be more widely used, especially to study folate bioavailability, in the future.

See also: **Ascorbic Acid**: Physiology; **Bioavailability of Nutrients**; **Cereals**: Dietary Importance; **Chromatography**: High-performance Liquid Chromatography; **Cobalamins**: Physiology; **Coenzymes**;

Food Fortification; **Immunoassays**: Radioimmunoassay and Enzyme Immunoassay; **Infant Foods**: Milk Formulas; **Lactic Acid Bacteria**; **Macrobiotic Diets**

Further Reading

Ball GFM (1998) *Bioavailability and Analysis of Vitamins in Foods*. London: Chapman & Hall.

Friedrich W (1988) *Vitamins*. Berlin: Walter de Gruyter.

Gregory JF III, Williamson J, Liao J-F, Bailey LB and Toth JP (1998) Kinetic model of folate metabolism in nonpregnant women consuming [$^2H_2$] folic acid: isotopic labeling of urinary folate and the catabolite para-acetamidobenzoylglutamate indicates slow, intake-dependent, turnover of folate pools. *Journal of Nutrition* 128: 1896–1906.

Phillips DR and Wright AJA (1982) Studies on the response of *L. casei* to different folate monoglutamates. *British Journal of Nutrition* 47: 183–189.

Wright AJA, Finglas PM and Southon S (1998) Erythrocyte folate analysis: a cause for concern? *Clinical Chemistry* 44: 1886–1891.

Physiology

C J Bates, MRC Human Nutrition Research, Cambridge, UK

Nutritional Significance of Folic Acid and Folates

Like most micronutrients, the nutritional significance of folate was recognized and explored as the result of a deficiency disease, namely megaloblastic anemia, which occurred especially among pregnant women living in India. This was characterized by enlarged (megaloblastic) red cell precursors in the bone marrow, by macrocytic erythrocytes and by abnormal polymorphonuclear leukocytes with increased numbers of nuclear lobes, in the peripheral blood. There may also be lesions of the tongue (glossitis) and of the mucocutaneous margins of the mouth (cheilosis), and malabsorption. All of these lesions are shared with other vitamin B deficiency syndromes, so none of them is unambiguously diagnostic. Normally, the total body content of folate is about 5–20 mg. (*See* **Anemia (Anaemia)**: Megaloblastic Anemias.)

As the biochemical functions of the folate coenzymes became clearer, it was recognized that folate deficiency had implications for many aspects of homeostasis, metabolism, and health beyond the limited confines of blood cell production and maturation. Folate is essential for nucleic acid synthesis at all sites, and for a wide range of biochemical pathways involving one-carbon transfers. Deficiency states are by no means confined to the developing countries, although they are commonest and most severe there. As folate deficiency progresses and becomes more severe, there is a characteristic sequence of events, in which low plasma folate levels are followed by low erythrocyte folate levels, then increased polymorphonuclear leukocyte nuclear lobe numbers, macroovalocytosis of the erythrocytes and bone marrow macrocytosis, then macrocytic anemia.

Conflicting evidence about human requirements has led to considerable fluctuations in official recommendations for folate intakes, and problems with the assay of folates in foods have created parallel difficulties in the estimation of dietary folate intakes by individuals and populations.

Physiology and Functions of Folic Acid

In order to be absorbed efficiently, the polyglutamate folates in food need to be broken down to monoglutamates, either during food preparation, or at the brush border of the jejunum, which is the main site of absorption. It is difficult to give precise figures for the efficiency of deconjugation, and of absorption, because they vary considerably with the type of folate, with other dietary constituents, and with unidentified physiological factors, especially the pH at the site of absorption, the optimum being *c*. pH 6. Many foods contain inhibitors of the intestinal brush border folate conjugase and/or folate transport system, thereby reducing the efficiency of absorption. Others, such as milk, contain folate binders which may stimulate absorption.

Absorption of food folate, which occurs mainly in the jejunum, is largely attributable to active transport against the concentration gradient at moderate intakes, and appears to involve two jejunal brush border folate-binding proteins. This process is pH-, sodium-, and glucose-dependent. Recent estimates suggest that about 50% of the complex mixture of folates that occur, in food can be absorbed and utilized. Although most individual folates, even polyglutamates, are better absorbed than this when given on their own, the effects of partial degradation and food borne inhibitors etc. reduce the overall efficiency, which can be very variable. Small amounts of folate may arise from the intestinal flora *in situ*, but in humans (unlike the rat), this is now known to be only a very minor source of absorbed folate. Although mono and short-chain polyglutamates are more readily absorbed and utilized than the long-chain polyglutamates, the difference between them is not as substantial as was once thought, and the old subdivision into 'free' and 'bound' (i.e., short-versus

long-chain polyglutamates) has largely been superseded. The nutritional status of the subject can affect the efficiency of folate utilization and its economy: deficiencies of vitamin B_{12}, iron, zinc, or vitamin C in particular can have an adverse effect. Folate absorption is also adversely affected by some drugs, including diphenylhydantoin, phenytoin, phenobarbital, primidone (anticonvulsants); cholestyramine, salicylates and nonsteroidal antiinflammatory drugs; and salicylazosulfapyridine (used to treat bowel inflammation).

Once absorbed, the monoglutamate folates in the portal plasma are carried to the liver for processing, the liver being a major repository of folate coenzymes. From there, folate is carried to the other tissues. Hepatic folate is also secreted in bile, and much of this biliary folate is reabsorbed. Absorbed folate in excess of requirements is excreted in the urine, and folate turnover results in several characteristic degradation products, such as *p*-aminobenzoylglutamates, also destined for excretion in the urine. A substantial amount of folate turnover also occurs by fecal excretion of endogenous folate. (*See* **Coenzymes**.)

Folate transport into the cells of other tissues is also mainly by carrier-mediated processes. There are two separate mechanisms: one involves a reduced folate carrier protein which is a transmembrane transporter, and the other is an anchored protein called membrane folate receptor, which has a high affinity for both reduced folates and folic acid. This is especially active in tissues like kidney, placenta, and breast, where efficient folate transport is critically important. It can be inactivated by certain mycotoxins. There are specific folate-binding proteins at various sites; some are involved in folate transport, others may have a protective function.

Within the cells of mammalian tissues, folate is converted by the enzyme, folyl polyglutamate synthetase, to γ-linked polyglutamates of the reduced folates. 5-Methyl tetrahydrofolate polyglutamate is the largest fraction of tissue folate, except in tissues with especially rapid cell division, where the formyl derivative tends to be dominant.

In most tissues, the dominant functions of the folate coenzymes are the synthesis of DNA, thus permitting cell division, growth, and tissue renewal, and the provision of active methyl groups via the methionine cycle and the methyl donor *S*-adenosylmethionine (SAM). Folates have several essential functions in the synthesis of purine and pyrimidine building blocks of DNA, but the one function that is exquisitely sensitive towards folate deficiency is the conversion of deoxyuridylic acid to thymidylic acid, catalyzed by thymidylate synthase (Figure 1). For this reaction, folate must first be converted into 5,10-methylene tetrahydrofolic acid, which undergoes a regeneration cycle each time it participates. This function of folate explains most of the pathological effects of its deficiency in humans and animals, including megaloblastic anemia, leukocytopenia and other white cell abnormalities that arise from disturbances in DNA synthesis and hence of cell division in bone marrow. Folate deficiency also reduces growth (in children), regeneration of intestinal mucosa, and cell division at other sites which have a rapid turnover. (*See* **Nucleic Acids**: Physiology.)

Certain anticancer drugs which interfere in the folate regeneration cycle (e.g., dihydrofolate reductase inhibitors such as methotrexate), and some antibiotics which are antimetabolities of folic acid, act by reducing pathological rates of cell division, thus protecting the host organism against unchecked division of parasitic cells. A major challenge for future research will be to target such antimetabolites specifically to the sites of invasion and damage, in order to reduce their side-effects on healthy tissues. Paradoxically, healthy tissues may also be protected against some carcinogenic agents by adequate folate nutrition, because removal of damaged DNA requires folate-dependent metabolic reactions. Recently, there have been several controlled studies of cancer susceptibility in human subjects, in which folic acid has been tested for possible protective effects. These have focused especially on the large bowel, uterine cervix, breast, pancreas and, to a lesser degree, lung. Although a consensus has yet to emerge, interest in this subject is rapidly expanding. (*See* **Carcinogens**: Carcinogenic Substances in Food: Mechanisms.)

Other important functions of the folate coenzymes include the turnover of histidine, the synthesis of methionine from homocysteine, interconversion of serine and glycine, and other single-carbon transfers between molecules. As noted above, the methionine-SAM pathway is of major importance for provision of methyl groups for a plethora of essential structural and functional molecules. There are complex allosteric (feedback) interactions which closely control these pathways which involve folate coenzyme/cosubstrate participation. The strong metabolic link between the functions of folate and those of vitamin B_{12} occurs at the methionine synthase reaction which transfers the carbon unit from methyltetrahydrofolate to homocysteine to form methionine (Figure 1). The requirement for B_{12} cofactor in this reaction leads to a lack of suitable active single-carbon units for DNA synthesis during B_{12} deficiency, so that either folate deficiency or B_{12} deficiency will result in megaloblastic anemia and the other sequelae of impaired cell division at sensitive tissue sites. (*See* **Amino Acids**: Metabolism; **Cobalamins**: Physiology.)

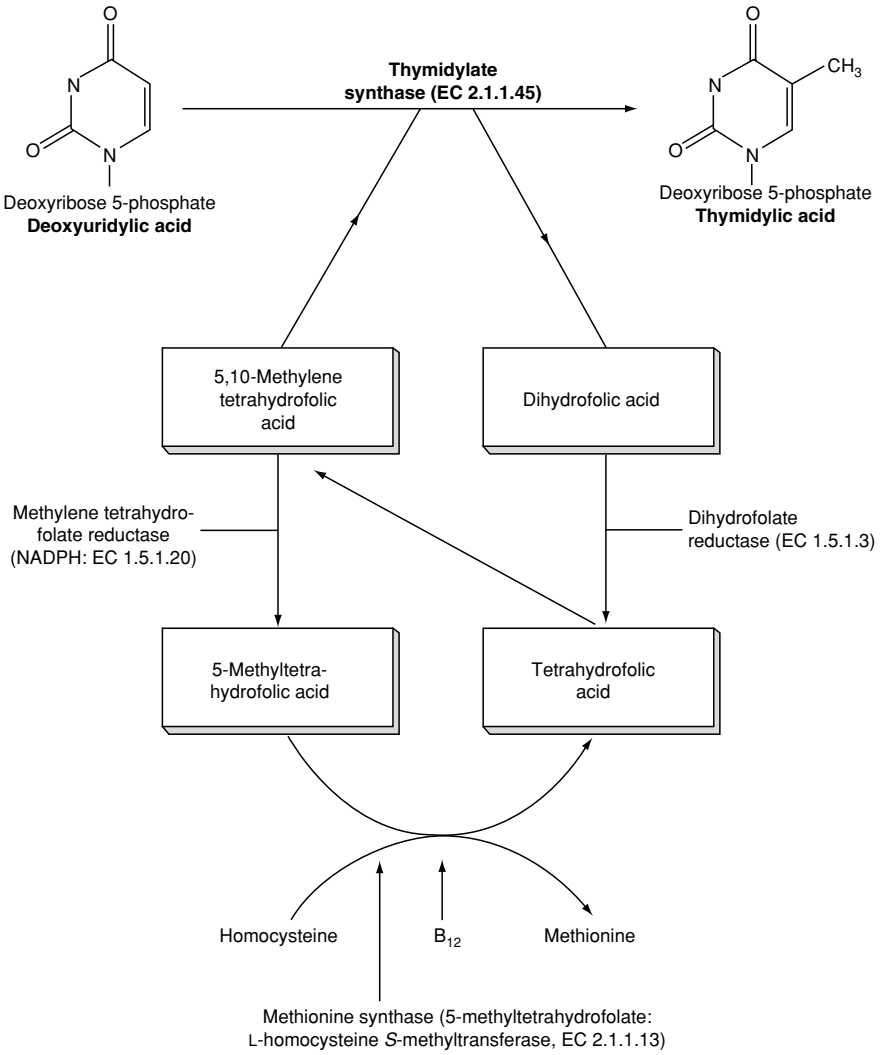

Figure 1 Biosynthesis of thymidylate, regeneration of methylene tetrahydrofolic acid, and biosynthesis of methionine.

Turnover of folate in the body results in part from its cleavage to liberate *p*-aminobenzoyl polyglutamate, which is then deconjugated and excreted as *p*-acetamidobenzoylglutamate and *p*-acetamidobenzoate. These breakdown products, plus some unchanged folate (at higher circulating levels), are excreted in the urine; however the essential body stores of folate are efficiently conserved in the kidney by reabsorption in the proximal renal tubule, and by an efficient enterohepatic circulation. Folate that is found in the feces is mainly what is formed by intestinal bacteria in the large bowel, unless folic acid intakes are very large.

Biochemical Status Measurements

Folate biochemical status is monitored most commonly by the assay of folate concentrations in plasma or in the red cells of the peripheral blood.

Plasma folate essentially reflects recent intakes and the magnitude of intertissue transfer, whereas red cell levels represent longer-term tissue stores, since intracellular (red cell) folate does not readily exchange with extracellular folate. Infants have much higher serum and red cell folate levels than those of adults, and this may be connected with the greater rate of cell division in infancy. Assays for folate which are based on competitive protein binding (radioassay or fluorescence assay), or on antibody binding, are now widely used, and several commercial kits for competitive-binding assays are now available. It is commonly accepted that serum (or plasma) levels below 3 μg l^{-1} or erythrocyte levels below 100 μg l^{-1} are indicative of biochemical deficiency, and should be further investigated or treated.

'Functional' tests, based on the efficiency of catabolism of a histidine load, or on deoxyuridine suppression of preformed thymidine utilization for DNA synthesis in tissue biopsies such as bone marrow cells,

may provide useful functional evidence of tissue folate adequacy, although they are not used for routine measurement of status. Folate deficiency, like vitamin B_{12} or B_6 deficiencies, results in raised plasma homocysteine (see below), and this functional index can provide useful information about folate status and dietary adequacy. (*See* **Immunoassays**: Radioimmunoassay and Enzyme Immunoassay.)

A useful cytological indicator of marginal folate deficiency, which is more sensitive than overt anemia or megaloblastosis, is an increase in the lobe average count of the polymorphonuclear neutrophils in the circulation. Fewer than 3.3 lobes per cell is considered normal; more than 3.65 per cell indicates a significant functional deficiency.

Human Requirements for Folic Acid and Folates

Recommended Dietary Allowances and Dietary Reference Values

Several lines of evidence have indicated that for adult men and adult nonpregnant and nonlactating women, the minimum requirement for folate (given as a supplement of pteroylglutamic acid) which is needed to prevent overt clinical deficiency is not greater than $50\,\mu g\,day^{-1}$. The problem which has faced recommended dietary allowance (RDA) committees is how much more is needed to cover the needs of nearly all individuals, when folate is obtained from food? In 1980, for instance, the US committee recommended $400\,\mu g\,day^{-1}$ and $800\,\mu g\,day^{-1}$ during pregnancy. By the end of the 1980s, the available evidence suggested that the majority of healthy adults in the USA, Canada, the UK, and elsewhere appeared to maintain acceptable folate status on intakes around 150–$200\,\mu g\,day^{-1}$. During the past decade, there has been a great deal of research activity which has once more reopened the debate and has resulted in a shift in the consensus towards a preference for higher requirement estimates once more. Metabolic ward studies of trends in status indices at different folate intakes have provided some of the evidence; studies of neural tube defect risk in very early pregnancy plus the identification of a substantial segment of the population with a genotype that implies possibly increased folate requirements is another, and the observation that plasma homocysteine concentrations are responsive to folic acid supplements even when folate intakes are as high as $300\,\mu g\,day^{-1}$ or higher, is a third.

Dietary Folate Equivalents (DFE)

The bioavailability of food folate differs from that of synthetic folic acid, mainly because the latter, being already oxidized, is not easily damaged, it is already a monoglutamate, and it is not in a tissue matrix. On average, the folic acid in supplements and fortified food is calculated to be 1.7 times as available as food folate; therefore the US National Academy of Sciences has recently introduced the concept of dietary folate equivalents (DFEs), analogous to retinol or niacin equivalents, which is equal to (μg food folate) + (1.7 μg synthetic folic acid). This has been used in calculating the new US (1998) estimated average requirements or EAR (e.g., $320\,\mu g\,day^{-1}$ for nonpregnant, nonlactating adults), and RDA (e.g., $400\,\mu g\,day^{-1}$ for nonpregnant, nonlactating adults), rising to $600\,\mu g\,day^{-1}$ during pregnancy and $500\,\mu g\,day^{-1}$ during lactation. Requirements during late pregnancy clearly must be higher than those of nonpregnant adults, because unsupplemented pregnant women, even in well-nourished societies, often develop folate-responsive bone marrow megaloblastosis during the later stages of pregnancy. Folate supplements effectively insure that this functional abnormality does not occur. (*See* **Dietary Requirements of Adults**.)

The folates that are present in breast milk ($c.50\,\mu g\,l^{-1}$) appear to be more readily available than the pteroylglutamic acid that is added to formula feeds, since breast-fed babies tend to have higher blood folate levels for a given folate intake. Folate-deficiency anemia has been encountered in premature or growth-retarded infants given inappropriate formulas, but these mistakes have now been rectified by giving folic acid-fortified feeds. Erythroblastotic infants have a particularly high folate requirement, and their growth rate can respond dramatically to folate supplements. Breast-milk folate levels are generally well protected, even when the mother's tissues and circulating pool have been partially depleted by moderate dietary folate deficiency. (*See* **Infant Foods**: Milk Formulas; **Infants**: Breast- and Bottle-feeding.)

Certain types of long-term medication, notably with anticonvulsive drugs, can interfere with folate economy, and therapeutic doses of folate may make it more difficult to control epileptic seizures in some patients. Patients with tropical sprue may, in contrast, benefit from folate supplements, at least transiently. Older people are particularly at risk, because their dietary folate intakes often fall below the safe level, and because homocysteine levels and the risk of overt vascular disease rise very steeply in old age. Chronic alcoholics, smokers, and people with chronic intestinal, renal, or hemolytic disease, or with rare inborn errors of folate economy, are at increased risk of deficiency. (*See* **Drug–Nutrient Interactions**; **Elderly**: Nutritionally Related Problems.)

Vitamin B_{12} deficiency results in impairment of folate utilization, and hence low intracellular and

tissue levels of folates. There is evidence of folate abnormalities associated with affective psychiatric disorders, and some patients may benefit markedly from folic acid supplementation. Raised homocysteine levels have recently been reported in the context of dementia, including Alzheimer's disease. A congenital disease associated with changes in the X chromosome and known as fragile X syndrome has been effectively treated with large folate supplements.

Tetrahydrobiopterin, another member of the same pteridine family to which the folates belong, is found in most mammalian tissues. It plays an important cofactor role in certain enzymes, especially those which introduce oxygen into aromatic amino acids. Unlike folic acid, tetrahydrobiopterin can be synthesized *de novo* within the body, and is therefore not classified as an essential vitamin. However, a dietary source of it may be essential in some abnormal metabolic states, particularly for some infants. A congenital lack of dihydropteridine reductase produces severe neurological defects, attributable to a tissue deficiency of tetrahydrobiopterin.

Currently Important Public Health Issues

Neural Tube Defects

During the early stages of embryonic development, failure of the normal closure of the neural tube can result in permanent abnormalities of the brain, cord, and/or spine called spina bifida, anencephaly, and encephalocele, which may be highly debilitating or fatal. This condition was relatively common in parts of the UK and Northern Ireland until the 1970s, averaging around 0.4% of human fetuses; it has now fallen to around 0.1%. Pregnancy termination has usually been advised following its detection by amniocentesis. Evidence of possible involvement of vitamins, and particularly folic acid, was suspected, and then proven by a large Medical Research Council trial, published in 1991. A 70% reduction in recurrence of neural tube defects (NTD) in a high-risk group, namely mothers with a previously affected fetus, was achieved by a 4 mg day^{-1} folic acid supplement during the critical early development period. More recently, studies in Hungary and China obtained substantial reductions in NTD incidence by smaller folate supplements, of 800 μg and 400 μg respectively. It is now recommended in the UK that women planning a pregnancy who are in this high-risk category should receive 4 or 5 mg day^{-1}, and those who are not at high risk should receive 400 μg (0.4 mg) day^{-1}, for a period which lasts from before conception until the third month of pregnancy. In the USA, there has been mandatory fortification of cereal grain products with folic acid at 1.4 μg g^{-1} from the beginning of 1998, one purpose being to reduce the incidence of NTD. In the UK, mandatory fortification has not been introduced, but a recent Commission on Medical Aspects of Food Policy (COMA) report has recommended folic acid fortification of flour at 2.4 μg g^{-1}, calculated to be the best compromise between reduction of NTD births and avoidance of intakes >1 mg day^{-1} in people >50 years.

NTD, like most pathologies, has multifactorial risks, and its etiology is not fully understood. An abnormality in vitamin B_{12} economy has been reported, and NTD occur frequently in women who are homozygous for the thermolabile form of the folate-metabolizing enzyme, methylene tetrahydrofolate reductase (MTHF). Recently, a deficiency of folate has also been linked with the occurrence of hare lip and of facial clefts.

Homocysteine

Folate, together with vitamins B_{12} and B_6, are all B vitamins whose supply and status can affect the circulating concentrations of the sulfur amino acid homocysteine. Moderately raised levels are an independent risk factor for several types of vascular disease, including cardiovascular disease, cerebrovascular disease, and deep vein thrombosis. Relatively high intakes and tissue levels of these three vitamins help to minimize plasma homocysteine levels. Homocysteine, the precursor of methionine, requires methyl tetrahydrofolate as the cosubstrate of vitamin B_{12}-dependent methionine synthase. In most normal adults, the plasma concentration of homocysteine is *c.* 5–15 μmol l^{-1}; moderate hyperhomocysteinemia is defined as 15–30 μmol l^{-1}; intermediate as 30–100 μmol l^{-1}; severe as >100 μmol l^{-1}. Folate, across its physiological range, is a particularly powerful determinant of plasma homocysteine. Although there is a widespread expectation that folate status may influence vascular disease risk, this has not yet been proven. There is the further perceived complication (see below), that high folate intakes can increase hemoglobin synthesis, without preventing neuropathy, in the early stages of vitamin B_{12} deficiency. The 'masking' of B_{12}-deficiency anemia might delay diagnosis and treatment, with irreversible consequences. Therefore, public health interventions designed to optimize the folate status of populations need to proceed with caution.

Toxicity

Folic acid is generally considered to be nontoxic, even in large amounts when taken orally, but some concern has been expressed about the danger of masking, or even possibly hastening the progress of,

the early stages of vitamin B_{12}-deficiency neuropathy, including that caused by pernicious anemia. Most workers now consider that masking is the main hazard. Very high folic acid intakes may have subtle effects on zinc economy, but this is not now considered a serious hazard. Reports of mood changes, irritability, sleeplessness, and gastrointestinal symptoms on intakes of 15 mg of pteroylglutamate per day have been reported, but other observers disagree. Occasional reports of allergic reactions to folic acid have occurred, and high doses may be epileptogenic or reduce the efficacy of anticonvulsant control of epilepsy in susceptible subjects. In general, water-soluble vitamin intakes seem to be tolerated by most people over a very wide range. However, in view of the current publicity about benefits of folic acid, it seems prudent to warn that excessive intakes of any nutrient may produce imbalances which are potentially risky.

See also: **Amino Acids**: Metabolism; **Anemia (Anaemia)**: Megaloblastic Anemias; **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; **Cobalamins**: Physiology; **Coenzymes**; **Dietary Requirements of Adults**; **Drug–Nutrient Interactions**; **Elderly**: Nutritionally Related Problems; **Immunoassays**: Radioimmunoassay and Enzyme Immunoassay; **Infant Foods**: Milk Formulas; **Infants**: Breast- and Bottle-feeding; **Nucleic Acids**: Physiology

Further Reading

Bailey LB (ed.) (1995) *Folate in Health and Disease*. New York: Marcel Dekker.

Department of Health (1999) *Report on Health and Social Subjects no. 50. Folic Acid and the Prevention of Disease*. Report of the Committee on Medical Aspects of Food and Nutrition Policy. London: The Stationery Office.

Food and Nutrition Board. Institute of Medicine. National Academy of Sciences, Subcommittee on Folate, other B Vitamins and Choline (1998) *Dietary Reference Intakes: Thiamin, Riboflavin, Niacin, Vitamin B-6, Folate, Vitamin B-12, Pantothenic Acid, Biotin and Choline*. Washington, DC: National Academy Press.

Gregory JF (1989) Chemical and nutritional aspects of folate research: analytical procedures, methods of folate synthesis, stability and bioavailability of dietary folates. *Advances in Food and Nutrition Research* 33: 1–101.

Gregory JF (1997) Bioavailability of folate. *European Journal of Clinical Nutrition 51* (supplement 1): S54–S59.

Hawkes JG and Villota R (1989) Folates in foods: reactivity, stability during processing, and nutritional implications. *CRC Critical Reviews in Food Science and Nutrition* 8: 439–538.

Kim Y-I (1999) Folate and carcinogenesis: evidence, mechanisms and implications. *Journal of Nutritional Biochemistry* 10: 66–88.

MRC Vitamin Study Research Group (1991) Prevention of neural tube defects: results of the Medical Research Council vitamin study. *Lancet* 2: 131–137.

Picciano MF, Stokstad ELR and Gregory JF (eds) (1990) *Folic Acid Metabolism in Health and Disease. Contemporary Issues in Clinical Nutrition*, vol. 13. New York: Wiley-Liss.

Refsum H, Ueland PM, Nygaard O and Vollset SE (1998) Homocysteine and cardiovascular disease. *Annual Review of Medicine* 49: 31–62.

Zittoun J and Cooper BA (eds) (1989) *Folates and Cobalamins*. Berlin: Springer-Verlag.

FOOD ACCEPTABILITY

Contents
Affective Methods
Market Research Methods

Affective Methods

H L Meiselman and A V Cardello, US Army Soldier Centre, Natick, MA, USA

Introduction

The study of human behavior with respect to food includes food selection, food acceptance, and food consumption. Studies of food acceptability involve both sensory and affective dimensions of food. Sensory studies relate to the perception of the gustatory, olfactory, textural, visual, thermal, and other attributes of food. Affective studies concern people's evaluative reactions on a like–dislike or hedonic dimension. Over the past decade, there has been increased attention to the many variables that influence food acceptability and a growing awareness of the complexity of food acceptability. Along

with these changes have come new sensory and affective methods for assessing food acceptance, and greater appreciation for how to apply existing methods.

Sensory and Affective Judgments

While a model of food selection and consumption based solely on the sensory properties of foods is conceivable, the evaluative dimension is a primary driver of the consumer's behavioral response to food. One of the critical questions in food acceptability research is whether to ask respondents for strictly sensory judgments, or to ask for evaluative or hedonic judgments, or both. This distinction between description and evaluation is key in the selection of test subjects, methods, samples, controls, and other experimental design considerations. Both the sensory and affective aspects of food acceptability are well covered by recent volumes such as Lawless and Heymann, and Meiselman and MacFie. However, the social, cultural, physiological, and contextual aspects of the eating situation must also be considered for their influence in determining choice, acceptance, and consumption. This article will review and compare the primary methods for measuring sensory and affective responses to foods (Figure 1).

Trained, Expert, and Consumer Panels

A trained panel is a group of individuals who have been selected for training based on criteria related to their ability to perceive differences in basic sensory attributes, and then trained in specific sensory evaluation techniques and the use of a standardized lexicon. Preselection criteria commonly include sensory threshold testing. Sensory training is usually conducted in one specific technique, e.g., descriptive texture analysis, rather than in a variety of techniques. An expert panel is usually a group with extensive formal or informal experience with the product or the product category, but without formal or standardized selection and training. Expert panels are often made up of inhouse experts within an organization. Trained and expert panels members may or may not be actual users of the product or product class. Consumer panels are comprised of people without special training or experience; a consumer panel should represent the target population of the product.

Most professionals agree that affective acceptance testing should be done only with untrained, nonexpert consumers. The reason for this is that trained panelists may well have different preferences from the average consumer. A number of studies have compared trained and consumer panels for hedonic judgments. For example, Shepherd and colleagues found that ratings of eight tomato soups were different for trained and consumer panels. Only the trained panel showed significant differences in overall product ratings, and in general, the trained and consumer panels preferred different products.

On the other hand, the extent to which consumers can be used to provide sensory data is somewhat controversial. The reason for this is that the fields of psychology and psychophysics have had a long and successful history of using naive subjects (consumers) to make sensory and perceptual judgments on a wide range of both simple and complex stimuli. This paradigmatic approach has been carried over by psychologists such as Moskowitz to the study of food acceptability, and methods have been developed to apply consumer terminology to the study of the sensory dimensions of food (free-choice profiling). However, traditional food industry approaches focused on either expert or trained panels to provide such sensory data on foods. At present, it can best be said that a consumer panel should be used when affective data are being obtained or when sensory data are to be generalized to a consumer population. Trained panels should be used when there is a requirement to maximize the sensitivity to differences in the product, when detailed descriptive information is required, or when comparisons are to be made to instrumental data. Expert panels, in general, should be restricted to specific, in-plant, quality control problems or to commodity areas in which expert judgments have been shown to be a valid and reliable approach to obtaining necessary data.

Category Scales

Category scales are the most commonly used methods for assessing food likes and dislikes. They are characterized by a series of labeled points or categories, and a food item is rated by assigning it to one of several (usually 3–10) available descriptive categories. One of the earliest and still most commonly used scales of food preference and food acceptance is based on the category method. This is the nine-point hedonic scale developed by Perham and colleagues in the early 1950s at the US Army Quartermaster Food and Container Institute in Chicago which later moved to Natick, Massachusetts.

Empirical evidence indicates that the points on the nine-point hedonic scale do not represent equal subjective intervals. Jones and colleagues scaled a variety of verbal phrases for their meaning on an independent like–dislike scale. These data reveal that the intervals between phrases used as category labels for the

nine-point hedonic scale are unequal. For example, the interval between 'like extremely' and 'like very much' is larger than the interval between 'like moderately' and 'like slightly.' Earlier work on this problem has been reviewed by Schutz and Cardello, and this nonequivalence of scale intervals demonstrated using magnitude estimates of liking; this will be discussed in the next section.

Another problem with category scales is that respondents tend not to use the extreme categories, because they fear that if they use these categories to rate a food item and a more extremely liked (or disliked) item is presented subsequently, they will have no appropriate category left to assign. As such, the nine-point hedonic scale can effectively become a seven-point scale of liking/disliking. Also, many category scales (such as the nine-point hedonic scale) have a neutral point. There is some evidence that eliminating this neutral point may increase the efficiency of the scale and may discourage the complacency in judgment that is provided by this 'safe' category.

In spite of these problems, many researchers have adopted the nine-point hedonic scale for studying food acceptability because of its ease of use, its applicability to both preference and acceptance data, and its good reliability. In addition, category scales are widely used for rating other dimensions of food acceptability, such as the sensory intensity of food attributes, the situational appropriateness of foods, sensory and/or hedonic expectations toward foods, and even preference mapping.

Unstructured or Line Scales

Scales that consist of an unbroken line that is anchored on both ends are called unstructured, graphic, line, or visual analog scales. Line scales can be used to quantify both sensory dimensions and hedonic dimensions. Several different scales were tried in one study by Pangborn and colleagues: a 10-cm line marked 'dislike extremely' and 'like extremely' as the two anchors on a visual analog scale of liking, a similar line scale with the midpoint defined by 'neither like nor dislike,' and a 100-unit scale labeled at the ends and midpoint. Hedonic scores on the line scale tended to be more centrally clustered than scores of sensory intensity. Reliability of the scores from line scaling was usually good, but with some products subjects did not maintain the same order of hedonic judgment. Lawless and Malone found that line scales of sensory intensity were not as restrictive as category scales, and could be completed quickly; however, they may be somewhat more difficult to understand.

Herbert Stone and his colleagues developed the procedure known as quantitative description analysis (QDA), and recommended the use of line scales for quantifying sensory dimensions, maintaining the use of the nine-point hedonic scale for the hedonic dimension of products. In the original description of QDA, both sensory and hedonic dimensions were measured on 6-in (15-cm) line scales with anchors (0–5) located 1–3 cm from each end (labeled 'dislike moderately,' and 'like moderately') and a midpoint ('neither'). It was argued that the scales were interval scales and empirical evidence was presented to support this.

Ratio Scales

In the 1950s, SS Stevens at Harvard University developed a variety of ratio scales that permit ratio comparisons of stimuli (twice as strong, one-third as sweet, etc.). The most widely known (but not the only) ratio methods to be adopted for use in food research are magnitude estimation and its variant, magnitude matching. The latter technique has only been applied to basic sensory problems and will not be further discussed.

Magnitude estimation was originally applied to a large number of sensory modalities to establish psychophysical functions relating physical intensity to perceived (subjective) intensity. Magnitude estimation was probably first applied to foods in 1971 by Howard Moskowitz and Joel Sidel. The practical importance of magnitude estimation was argued as a 'supplement' to category scaling, and it was suggested that the category scale could be used to label the perceptual categories of acceptance, while magnitude estimation could be used to refine the degree of acceptability.

In magnitude estimation the task for the respondent is to rate the sensory intensity or liking of one product in relation to another. If one product is liked twice as much as another product, it is assigned a number twice as large. Many studies have used magnitude estimation to judge the perceived sensory intensity of laboratory food models and of real foods. Far fewer studies have used magnitude estimation to assess the hedonic dimensions of foods, primarily because affective judgments must be made by untrained consumers, and magnitude estimation is a more complex method to understanding and to use than are other methods of scaling, e.g., the nine-point hedonic scale.

Several studies comparing magnitude estimation to category and line scales for rating sensory intensity have shown equal sensitivity, reliability and precision among the scales. When used to scale affect, magnitude estimation was found by some authors to be more sensitive than the nine-point hedonic scale,

1. Category scales

(a) Peryam and Girardot (1952), Jones *et al.* (1955) – the nine-point hedonic scale

1	2	3	4	5	6	7	8	9
Dislike extremely	Dislike very much	Dislike moderately	Dislike slightly	Neither like nor dislike	Like slightly	Like moderately	Like very much	Like extremely

(b) Lawless and Malone (1986) – nine categories

1	2	3	4	5	6	7	8	9
Little								Much

2. Line scales

(a) Pangborn *et al.* (1989) – labeled 0–100

0 Dislike extremely	50 Neither like nor dislike	Like extremely 100

(b) Shepherd *et al.* (1988), Pangborn *et al.* (1989) – 100 mm

Dislike extremely	Neither like nor dislike	Like extremely

(c) Giovanni and Pangborn (1983) – 100 mm

Dislike extremely		Like extremely

(d) Stone *et al.* (1974)

Dislike moderately	Neither	Like moderately

3. Relative-to-ideal scales

(a) Shepherd *et al.* (1988) – 100 mm

Not nearly strong enough	Just right	Much too strong

4. Ratio scales

Magnitude estimation:

(a) Moskowitz and Sidel (1971) – first sample served as standard; no zeros or negative numbers permitted.

(b) Pearce *et al.* (1986) – unipolar scale from zero to infinity (liking).

(c) Pearce *et al.* (1986) – bipolar scale from minus infinity (disliking) through zero (neutral) to plus infinity (liking).

(d) Giovanni and Pangborn (1983) – unipolar scale; no zeros or negative numbers permitted; fixed modulus of 10.

Labeled magnitude scales

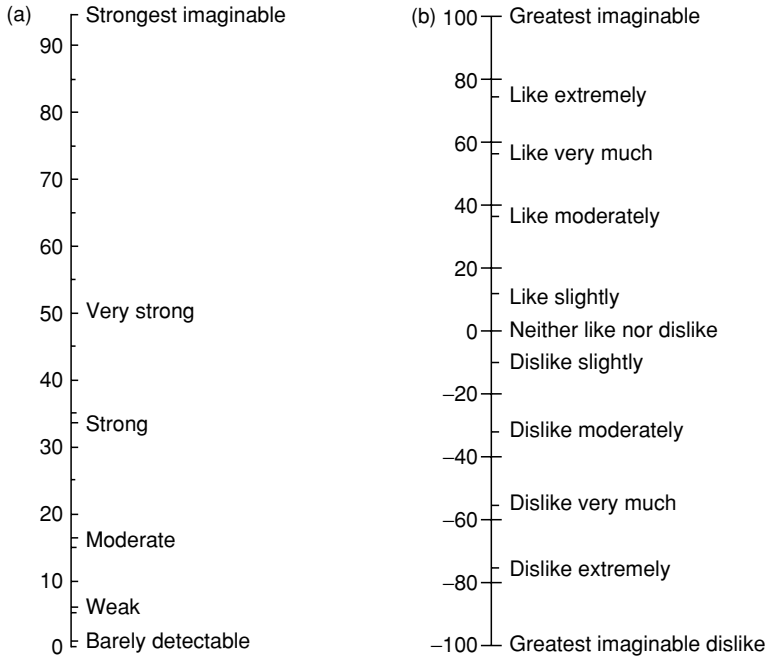

(a) Green *et al.* (1993) – equally spaced numerical labels with ratio-determined verbal labels.

(b) Schutz and Cardello (2000) – equally spaced numerical labels with ratio-determined verbal labels.

5. Frequency scales

(a) Schutz (1965)

I would eat this every opportunity I had.
I would eat this very often.
I would frequently eat this.
I like this and would eat it now and then.
I would eat this if available but would not go out of my way.
I don't like it but would eat it on occasion.
I would hardly ever eat this.
I would eat this only if there were no other food choices.
I would eat this if I were forced to.
Never tried.

(b) Meiselman (1988) – each food can be rated for preferred frequency in terms of days per week (1–7) and weeks per month (7–28 times) or rated 'never.' There are separate ratings for each meal.

	Breakfast		Mid day		Evening meal		
	days per week	weeks/ month	days per week	weeks/ month	days per week	weeks/ month	Never
Food name	1 2 3 4 5 6 7	1 2 3 4	1 2 3 4 5 6 7	1 2 3 4	1 2 3 4 5 6 7	1 2 3 4	0

Figure 1 Examples of sensory, hedonic, and frequency scales. Most of the scales can function as either sensory or hedonic scales by changing from sensory attributes (e.g., salty) to hedonic attributes (e.g., like) or vice versa.

while others found the hedonic scale to be more reliable, although both scales were linearly related. A collaborative study by Pearce was sponsored by the American Society for Testing Materials (ASTM) and compared the nine-point hedonic scale with both unipolar and bipolar (like–dislike) magnitude estimation. Eight fabric samples were rated by 20–30 panelists at 23 different sites. All three scales ranked the eight fabrics in the same order with very similar spread of ratings; they achieved very similar degrees of precision in differentiating samples, and similar discrimination. Reliability of the category scale was higher than for either form of magnitude estimation, although higher reliability would likely be obtained with larger sample sizes.

Thus, it appears that, while magnitude estimation provides valid and reliable data on the sensory and hedonic attributes of food, its practical difficulty in use has limited its broad-based application, in favor of the simplicity of category scales. However, in a recent effort to combine the practical simplicity of category scales with the quantitative sophistication of ratio scaling for rating the intensity of oral sensations, a hybrid, category-ratio scale was developed by Barry Green. This scale places verbal labels of expressed intensity along a linear numerical scale at specific locations that reflect the numerical ratios among their perceived intensities, as obtained by a magnitude estimation procedure. A critical characteristic of this scale is the presence of a phase involving the 'maximal,' 'strongest imaginable' or 'strongest possible' sensation. The use of this phase is assumed to serve as a fixed endpoint of sensation that places judgments of different subjects on a common sensory 'ruler.' Subjects can quickly look at the verbal labels and corresponding numbers and then place a slash mark (/) through the line scale to indicate the perceived strength of their sensation. The resulting data have been shown to have ratio properties. More recent studies of this scale have shown that the scale is most effective for application to chemosensory stimuli that include extremely intense (painful) sensations.

Recently, the use of such labeled magnitude scales has been extended to affective ratings by Howard Schutz and Armand Cardello. Using Barry Green's approach, the semantic meaning of various verbal phrases used to describe the liking/disliking of foods was scaled. By locating these phrases along a visual analog scale in accordance with their determined semantic meaning, it was shown that consumers could use these scales to rate the affective aspects of stimuli with equal or better sensitivity, greater reliability, and equivalent ease of use as with the nine-point hedonic scale.

Alternative Scales: Relative-to-Ideal, Food Frequency

While category, line, and ratio scales are the most commonly used instruments for assessing food acceptability, they are not the only ones. Another approach to scaling affective reactions to foods is the relative-to-ideal scale, which combines the affective or hedonic component with the sensory component. For example, rather than separately scaling sweetness and liking, the relative-to-ideal scale combines the two dimensions to yield a scale from 'much too sweet' to 'not nearly sweet enough.' The midpoint of such scales is usually defined as 'just right.' Such a scale implies that the respondent has an internal ideal of a product against which to compare a stimulus.

Hence it is appropriate for established products, but less appropriate or inappropriate for novel products or for laboratory food models (e.g., sugar water).

A number of studies have shown good agreement between relative-to-ideal scales and hedonic scales. The breakpoint in the hedonic function, at which responses shift from increasing liking to decreasing disliking, is highly correlated with the just-right point, although the latter tends to be slightly higher. Shepherd has argued that relative-to-ideal scales are 'unfolded hedonic scales,' i.e., scales which measure the amount of like or dislike from the just-right point of the scale. One advantage of the relative-to-ideal scale is that it produces a sharp fulcrum, below which are less-liked, weaker points, and above which are more-liked, stronger points. Hedonic judgments across the same range will yield a rounded breakpoint, because different respondents have different maximally preferred points.

The relative-to-ideal scale appears useful for quantifying the hedonic aspect of a product along a clear, definable intensity dimension. The scale has been used widely in the food industry because it tells product developers in which direction to make product changes along a sensory dimension (e.g., make it more sweet). It is unclear how this would be applied to complex food products with multiple physical dimensions (many tastes, smells, textures) and to combinations of food products. It is also unclear what level of training or expertise is needed for optimal use of this scale.

Another approach for quantifying food acceptability is through the use of food frequency scales. The basic concept here is that the frequency with which an individual consumes a food or expects to consume it is a valid index of the acceptability of the food to that person. Similarly, since the practical purpose of any measure of food acceptability is to predict subsequent consumer behavior toward the food, simple measures of affect may not suffice. For example, although cheesecake may receive a rating of eight on a nine-point hedonic scale, while milk may receive a rating of six, this does not mean that the individual wants to eat cheesecake more often than he/she wants to drink milk. Thus, hedonic judgments of the acceptability of foods are not as useful for menu planning or for predicting food choice as are judgments of the desired frequency of serving.

Howard Schutz developed the first form of food frequency scale was developed as a supplement to the nine-point hedonic scale. Known as the food action (FACT) scale, it consists of a series of statements concerning the frequency with which a food item is desired to be eaten, ranging from 'eat every opportunity I get' to 'eat if forced.' By combining hedonic scale data with FACT scale data, information can be obtained about both the affective aspects of a food item and the relative frequency that it is desired. The FACT scale was shown to have high group reliability coefficients of 0.97.

While the FACT scale provides only relative frequency estimates, other frequency techniques can be used to provide absolute frequency measures. A frequency scale was designed to assess food preferences of military men. Respondents indicate the number of times during a 30-day period that they would like to have an item served. Other research from Natick indicated high group reliability for such frequency scales ($r = 0.98$), moderate individual reliability coefficients ($r = 0.58$–0.60), and moderate correlation with consumption ($r = 0.59$–0.66).

Affective Judgments: The Role of Context and Expectations

The study of food acceptability and the measurement of its sensory and affective dimensions are complicated by the fact that foods are typically eaten within a context, i.e., the meal, and that the consumer typically brings to the dining situation a host of expectations that are based on previous experience, product and nutrition information, and packaging, all of which can affect liking for the food. Such influences can limit the predictive validity of laboratory-based measures of food acceptance unless adequate research methods are developed to understand and control these effects. In terms of situational context, few studies have adequately assessed the contribution of the meal context/situation to liking. Herbert Meiselman and his colleagues have studied the influence of a number of contextual variables was studied. All studies showed changes in the probability of food selection, while some studies showed changes in acceptance and some did not. For example, the effort required to obtain food showed large selection rate changes but no reliable changes in acceptance, whereas changes in food cost changed both selection rates and acceptance. Further studies identified a major difference in acceptance ratings of foods in institutional vs noninstitutional food service settings, in agreement with previous observations concerning expectations in these two different classes of environments.

In an attempt to develop a measure of food acceptability that takes into account the situational influences of the dining situation, Howard Schutz developed the concept of 'situational appropriateness' as well as a method to index the degree of appropriateness of any food in a wide range of situations. This scale has been used effectively in a variety of survey formats, and more recently has been shown

to be a valuable adjunct to laboratory affective testing of foods.

Concerning the expectations that consumers bring to the dining situation, Armand Cardello and other researchers have shown that a wide variety of extrinsic variables, such as product information, packaging, brand names, and nutrition statements, can influence the consumer's expected liking for a product. In numerous studies designed to look at the role of these expectations on actual liking and behavior toward the food, expectations were manipulated so as to lead consumers to expect a better-liked or a worse-liked product. By measuring expected liking using a simple category scale and comparing changes in actual liking relative to a baseline rating of the acceptance of the food in a blind (no information) condition, it has been shown that product acceptance assimilates the level of expectation. That is, if expectations are higher than the baseline acceptance of the product, rated acceptability increases. If expectations are lower, rated acceptability decreases. These important influences of consumer beliefs and expectations on product perception have been demonstrated for purely sensory dimensions of food and have been shown to have a direct influence on behavior, e.g., consumption.

Which Method to Use

When conducting basic research or practical product evaluations, the question arises as to which is the best scale method for measuring food acceptability. There are two parts to the answer. First, it is usually far better to select a scale that has been carefully developed and tested in a wide variety of situations, including conditions similar to those planned. Second, if any of the commonly used scales (**Figure 1**) are employed, there is probably not a great deal of difference among them. Category scaling and line scaling appear to be easier for people to use than magnitude estimation, although the use of a semantic ratio scale will provide ratio level data with equal ease of use to that of category and line scales. Relative-to-ideal scales combine a sensory and a hedonic dimension, if it is thought that people have a knowledge of what is 'just right' for the product, while frequency scales provide more meaningful data for menu planning and predicting repeat consumption of foods. Lastly, caution must always be used in generalizing the results of laboratory-based affective studies to real-life eating situations, in order to avoid the influences of both situational context and consumer expectations. In view of the increase in published methodological research on sensory and affective scaling in the past decade, the strengths and weaknesses of different scaling methods under a variety of conditions of use will be better clarified in the future.

See also: **Sensory Evaluation**: Sensory Difference Testing; Sensory Rating and Scoring Methods; Descriptive Analysis

Further Reading

Giovanni ME and Pangborn RM (1983) Measurement of taste intensity and degrees of liking of beverages by graphic scales and magnitude estimation. *Journal of Food Science* 48: 1175–1182.

Green BG, Shaffer GS and Gilmore MM (1993) Derivation and evaluation of a semantic scale of oral sensation magnitude with apparent ratio properties. *Chemical Senses* 18: 683–702.

Jones LV, Peryma R and Thurstone LL (1955) Development of a scale for measuring soldiers' food preferences. *Food Research* 20: 512–520.

Lavenka N and Kamen J (1994) Magnitude estimation of food acceptance. *Journal of Food Science* 59: 1322–1324.

Lawless HT and Heymann H (1998) *Sensory Evaluation of Food*. New York: Chapman and Hall.

Lawless HT and Malone GJ (1986) The discriminative efficiency of common scaling methods. *Journal of Sensory Studies* 1: 85–98.

MacFie HJH and Thomson DMH (1994) *Measurement of Food Preferences*. Glasgow: Blackie Academic and Professional.

McDaniel MR and Sawyer FM (1981) Preference testing of whiskey sour formulations: magnitude estimation versus the 9-point hedonic. *Journal of Food Science* 46: 182–185.

Meiselman HL (1988) Consumer studies of food habits. In: Piggott JR (ed.) *Sensory Analysis of Foods*, 2nd edn, pp. 267–334. London: Elsevier Applied Science.

Meiselman HL (2000) *Dimensions of the Meal: The Science, Culture, Business, and Art of Eating*. Gaithersburg, MD: Aspen.

Meiselman HL and MacFie HJH (1966) *Food Choice Acceptance and Consumption*. London: Blackie Academic and Professional.

Moskowitz HR (1983) *Product Testing and Sensory Evaluation of Food*. Westport, CT: Food and Nutrition Press.

Pangborn RM, Guinard JX and Meiselman HL (1989) Evaluation of bitterness and caffeine in hot chocolate drink by category, graphic and ratio scaling. *Journal of Sensory Studies* 4: 31–53.

Peryam DR and Girardot NF (1952) Advanced taste-test method. *Food Engineering* 24: 58–61.

Pearce JH, Korth B and Warren CB (1986) Evaluation of three scaling methods for hedonics. *Journal of Sensory Studies* 1: 27–46.

Schutz HG (1965) A food action rating scale for measuring food acceptance. *Journal of Food Science* 30: 365.

Schutz HG and Cardello AV (2001) Development of a labeled affective magnitude (LAM) scale for assessing food liking/disliking. *Journal of Sensory Studies* 16: 117–159.

Shepherd R, Griffiths NM and Smith K (1988) The relationship between consumer preferences and trained panel response. *Journal of Sensory Studies* 3: 19–35.

Shepherd R, Smith K and Farleigh CA (1989) The relationship between intensity, hedonic and relative-to-ideal ratings. *Food Quality and Preference* 1: 75–80.

Stone H, Sidel J, Woolsey A and Singleton RC (1974) Sensory evaluation by quantitative descriptive analysis. *Food Technology* 29: 24–34.

Market Research Methods

H R Moskowitz, Moskowitz Jacobs Inc., White Plains, NY, USA

The Market Research Role in the Corporation

Market research as a profession has the job of supporting the marketing function. For the past 40 years corporations worldwide have used market research to identify consumer trends, measure the acceptance of concepts, and measure the acceptance of products.

Acceptance measurement takes on a variety of different aspects. At the one end lies small-scale taste-tests in the corporate facility. These tests are usually executed by the in-house sensory evaluation group. There are slightly larger-scale research studies run with local consumers. These are the so-called church panels. Church panels may be executed either by a market researcher or by the in-house sensory analyst. The researcher contracts with local volunteer organizations to provide a source of panelists. For the larger-scale studies, involving hundreds of consumers, marketing researchers usually play the key role, rather than sensory analysts. Marketing researchers run these types of studies, often across a country to obtain geographic dispersion, and often with a relatively limited number of products. The role of the marketing researcher in these larger-scale studies is to present a report card of the product, in terms of performance, either in an absolute sense or relative to competition.

The Market Research Role versus the Sensory Research Role in Acceptance Testing

In today's corporations market researchers and sensory analysts each have responsibility for primary research on product acceptance. Sensory analysts often run small-scale studies to identify which particular formulation among a set of formulations looks promising for further development. Sensory analysts are therefore called in to execute acceptance tests very early in the project. Sensory analysts also measure acceptance as part of a general profiling exercise wherein panelists rate the different characteristics of products. The role of the sensory analyst in acceptance testing is to support product development, and not to support marketing. Therefore, the acceptance evaluations run by sensory analysts are usually considered directional in terms of ongoing development, rather than constituting a definitive 'go/no' decision for marketing. These guidance tests, none the less, often involve consumers.

Marketing researchers interact most usually with marketers and general management rather than with research and development (R&D). By the time a product has reached the marketing researcher, the product is presumed to be optimally acceptable. The marketing researcher tests the product among many individuals, in multiple locations, and often against a market leader, or against the current in-market product. The goal is not so much to provide guidance for development (such guidance having been assumed to be already in place and implemented), but rather to measure the potential of the product in terms of market share. From time to time the acceptance testing run by market researchers may be used to identify potentially fatal flaws in the product. For the most part if the sensory analysts have done their job properly such fatal product flaws should already have been eliminated.

The Multiple Aspects of Acceptance Measurement

Market research techniques involve a number of different aspects of acceptance measurement. Each of these aspects is investigated regularly by the researcher in an attempt to help the corporation develop the best product possible. These aspects include the following:

1. basic product acceptance without the presence of a concept
2. product–concept fit (does the concept fit the product, and vice versa?)
3. frequency of use
4. purchase intent coupled with price information
5. replacement of different foods by the product
6. appropriateness measurement – when would the product be eaten (e.g., day-part)?

7. relevance of the product as part of a line of products
8. loss of acceptance due to sensory-specific satiety and/or boredom/habituation
9. branded versus blind acceptance

Basic Product Acceptance

One key corporate goal is to offer a product with a high level of liking, packaged attractively, at an affordable (and competitive) price. In order to insure final product acceptance the market researcher uses a variety of procedures. Two of the most common procedures are monadic tests and paired comparison tests.

Monadic tests These tests require the consumer to rate the product on a scale of acceptability. There are many such scales (Table 1). These tests simply ask the consumer to act as a measuring instrument, assigning a number or a verbal label to represent product acceptance. On the analytic side the researcher averages the ratings to obtain a measure of acceptance. The results are reported as a score card of the products. Statistical tests are used in order to determine whether or not the product being tested differs significantly from the competition. Whatever scale is used must be amenable to statistical analysis, such as tests of differences (viz., *t*-tests).

Table 1 Types of hedonics scales

Two-point	*Five-point*	*Nine-point hedonic scale*
Like	Very good	Like extremely
Dislike	Good	Like very much
	Moderate	Like moderately
	Tolerate	Like slightly
Three-point	Never tried	Neither like nor dislike
		Dislike slightly
Acceptable	Very good	Dislike moderately
Dislike	Good	Dislike very much
Never tried	Moderate	Dislike extremely
	Dislike	
Like a lot	Tolerate	
Dislike		*Purchase intent scale*
Do not know		
	FACT scale	Definitely would buy
Well liked	Eat every opportunity	Probably would buy
Indifferent	Eat very often	Might/might not buy
Disliked	Frequently eat	Probably not buy
	Eat now and then	Definitely not buy
	Eat if available	
	Don't like–eat on occasion	
	Hardly every eat	
	Eat if no other choice	
	Eat if forced	

Reproduced from Meiselman HL (1978) Scales for measuring food preference. In: Petersen MS and Johnson AH (eds) *Encyclopedia of Food Science*, pp. 675–678. Westport: AVI, with permission.

Paired comparison tests These tests pit one product against another. One of the products may be the current product, another may be a revised formulation or a competitor, depending upon the business objectives. The consumer selects which of the two products is preferred, either overall or on specific characteristics. Quite often commercial tests require the consumer to indicate preference on many different attributes (up to 20 or more). The researcher computes the proportion of panelists who choose one product over another, for each particular attribute being investigated. The results are reported in the form of percentages. The paired comparison tests are used for several purposes, included measurement of performance against the competition, as well as claims substantiation (viz., product X is preferred to product Y).

Results reported for many subgroups For both monadic and paired comparison tests the market researcher looks for the data analyzed by total panel, and by the different key subgroups in the population. It is not unusual for a market researcher to work with upwards of 10 or 20 different subgroups in the population. Since the researcher deals with 300+ panelists in many of these studies, and since there is a classification questionnaire accompanying the actual evaluations, the researcher can cut the data many different ways.

Product–Concept Fit

Market researchers often test products in the presence of a concept. The concept tells the consumer what to expect. The concept may stress the benefits of the product (positioning), or the features of the product (product concept), or some combination of both. The results of the product concept fit show either the degree to which the product is acceptable based upon expectations set up by the concept, or the degree to which the product delivers what the concept promises. When products are tested in the presence of a concept the researcher gets a much better idea of the actual performance of the product after the consumer better understands the reason for the product and its end uses (e.g., appropriate time of day for consumption; pricing, etc., specific health benefits). Occasionally the product may score poorly without the presence of a concept, especially when the product is a health-oriented product that by itself is not particularly acceptable, but provides other benefits, such as nutrition and energy. The basic

liking rating may remain poor in the presence of the concept, but given the information about the product benefits, the purchase intent may increase dramatically.

Frequency of Use

Another type of acceptance measurement is the estimated frequency of use. Market researchers usually do studies to estimate volume, or amount to be consumed, not just overall liking. Thus, a measure of frequency reveals whether the product is highly liked but might be consumed infrequently, highly liked and consumed frequently, etc. Frequency is measured in at least two ways. One instructs the consumer to assign a number (e.g., times per month or times per year) to show the estimated number of times that the person would eat the product. Another gives the consumer a number of eating occasions (e.g., 10), and instructs the consumer to assign the number of times out of the next set of consumption occasions when the individual would eat the product. This approach is often used with the consumer given an array of products, and instructed to assign to each product the number of eating occasions out of the next 10 (or whatever number). This is called the constant sum method, and is used to determine the impact of introducing a product on the frequency of consumption of other products with which it competes for consumption.

Of all the acceptance measures, purchase intent is probably the most popular, and the one on which many marketing decisions are based. The typical purchase intent scale comprises a limited number of points (the most common being the five-point scale; Table 1). Rather than computing an average purchase score, the market researcher most typically computes the distribution of purchase intent ratings across the population of consumers who rated the product. This generates a set of incidence statistics, treated as percentages. The term 'top-two box' refers to the proportion of consumers who, having evaluated a specific product, rate it as definitely or probably would purchase.

Purchase intent is important for a number of reasons. First, it asks the consumer to describe what action he or she would take – purchase or not purchase. Second, many researchers create a normative database from many such studies, so that when they determine the purchase intent rating for a new product they can also estimate the likelihood of success. Third, a number of proprietary research procedures, aimed at estimating volume sales, use purchase intent as part of the input information to the model.

Purchase Ratings Coupled with Price Information

Purchase intent scales are often used along with pricing information. Although in theory price information should not affect product liking, it should affect purchase intent. Thus, a product may be very well liked, but the purchase intent may increase if the product is offered at a low price, and decrease if the same product is offered at a high price. This implicit demand curve can be captured by a simple equation: Purchase intent $= k_0(\text{liking}^m) \times (\text{price}^n)$. Typically the exponent for price, n, is much higher in absolute value (and negative) than the exponent for liking, (m, which is positive). This means that small changes in price have more effect in lowering purchase intent than the same small proportional change in liking has in raising purchase intent.

Acceptance as Replacement of a Current Product

Another aspect of acceptance deals with the issue of whether the product would replace or simply be added to the consumer's choices of foods to be purchased. Although not strictly a question of degree of liking, replacement questions deal with the issue of how the food will be used. Marketers always search for products that will be added to a person's set of foods, rather than one that has to fight to oust another food. It is much easier to add to a person's list than to replace a food that already has a history of being purchased.

Appropriateness – Day-Part and Situation

Food acceptance possesses a significant cognitive component, viz., that foods have defined parts of the day when they will be eaten, and other day-parts when they are not customarily eaten. Much of this is cultural, and labile, explaining why now cola drinks are being consumed for breakfast. Consumers have no trouble identifying the appropriate day-parts for specific products, and often use implicit rules to assign a product to breakfast, lunch, dinner, and snack. Appropriateness and eating context is emerging as an important aspect of food acceptance. The same food can be highly acceptable for dinner, but not particularly acceptable for breakfast (e.g., soup for North American consumers).

Relevance of the Product as Part of a Line of Related Products

From the manufacturer's point of view a single product (e.g., a single flavor) has several aspects of acceptance. One is the acceptance of the single product by itself. Another is the degree to which the product augments a line of already existing or

planned products. For instance, a fruit-flavored cereal may be highly acceptable itself without reference to other products, but may be redundant. That is, consumers satisfied with one fruit-flavored cereal may not really be interested in purchasing another, equally liked fruit-flavored cereal featured in the same line. To the marketer this means that the product is liked but redundant. A practical outcome would be that the item is rejected as part of the line. A substantial amount of consumer research funds is devoted to line optimization, or creating the proper combination of items in such a way that the line maximizes the number of consumers who would be satisfied by at least one of the products offered in the line. Market researchers call the solution of this problem the TURF issue, or 'total unduplicated reach and frequency.' Product acceptance here is viewed as an issue that transcends the more conventional rating of liking, because it calls into account the presence of, and reaction to, other potential entries in the same product line.

Acceptance in Light of Sensory Satiety and Boredom/Habituation

Very few products enjoy continued consumer acceptance at constant level (given a constant amount of advertising and promotion support). Sooner or later consumers tire of most products, and consume the product less frequently. This reduction in consumption comes about two ways – through sensory-specific satiety, and through boredom and habituation. A food loses its acceptance very quickly after eating, so that during a meal we tire of a food and stop eating it. This is sensory-specific satiety – we may have room for dessert even though we don't want to continue eating the main course. Habituation and boredom occur when the consumer simply tires of a product, and stops selecting it and/or eating it, even if the food has not been consumed recently. To some degree the more salient the sensory characteristics of the products, the more rapidly the product becomes boring. A good example of this is the difference between steak and hamburger. Steak wins against hamburger time after time in preference tests. Yet people tire more rapidly of steak after multiple consumptions, perhaps because of the salient sensory factors. An even more dramatic example of this occurs in the case of bread versus steak. Bread almost never wins in acceptance tests against steak, but we can eat bread day after day, meal after meal. We cannot consume steak to the same degree.

Branded versus Blind Product Acceptance

Most, but not all product testing is done in a blinded environment, wherein the researcher does not identify the particular brand. Branding (the identification of the brand), however, may make a great deal of difference in the acceptance rating when the product is already on the market. For many common products the brand may add substantially to the acceptance rating. The difference between the blind and the branded product shows the impact of the brand itself. In developmental research it is easy to test products branded because the researcher has control over the product, and can instruct R&D to create products in a form that cannot be identified. It is not so easy with in-market products. When the researcher wants to test competitor products that exist in the market, attempts to 'blind' or disguise the product may require unusual precautions. These have included shaving off an imprint on the product, testing the product in a red glass in red light (to prevent visual identification), or somehow cutting the product in order to eliminate a tell-tale shape that would immediately signal the brand.

Implementing a Study of Food Acceptance – Design and Field Considerations

The importance of acceptance as a precursor for product success has led to a large number of procedural recommendations and rules of thumb. Some of these are based in science, others are simply dictated by fiat through company policy.

Choice of Panelists who Participate

The composition of the sample generally is critical for a valid acceptance study. At the early stage of evaluation the particular choice of panelists is not too critical, other than the panelists accept the product. At the later stages, usually conducted by the market researcher, sampling the proper individuals becomes more important. The financial implications of a wrong decision (so-called downside risk) grow rapidly as the product comes closer to being launched in the marketplace. That is, if the study uses the wrong consumers, then the results are immediately suspect and the business decision may be faulty, even if the study is executed perfectly. The consumers may be chosen to be current users of the product, users of the competitor products, and lapsed users (individuals who used to use the product but no longer do so). In the case of children who consume the product but do not buy it, the panelists may comprise a group of parents and children (not necessarily related), in order to determine how the 'gatekeeper' (parent) responds, and how the ultimate consumer (child) responds. The conventional market research sample comprises some or all of these individuals, in sufficiently large base sizes of individuals so that the ratings from each subgroup can be analyzed.

What Type of Questionnaire is Most Appropriate for the Acceptance Test?

Most market researchers embed the acceptance scale (whether liking or purchase) in a larger set of questions dealing with other aspects of the food, such as the sensory and image aspects. For market researchers product acceptance is only one part of the picture (albeit an important part). Other important aspects include image of the product, when the product will be used, etc. The key exception here is a claims test, whose goal is to demonstrate that one product is preferred to another. The rule of thumb for a claims test is that the researcher should *limit* the question to liking or purchase intent alone, and not deal with any other attributes that could cloud the findings, and therefore the resulting claim.

Where Should the Acceptance Test be Conducted?

There are two test venues – central location and home use. A central location venue requires that the consumer come to a central site in order to evaluate the product. The site may be a shopping mall or a specialized test facility. A central location site affords the greatest control over the evaluation, because a trained food professional and/or interviewer can ensure that the product is properly prepared and served. The other test venue is one's home, wherein the product is sent home (e.g., by mail) or given to the panelists at a shopping mall to take home. The product is prepared, eaten, and often rated at home. The panelist may then be telephoned to get the ratings, or may mail back (or e-mail back) the completed questionnaire.

Some Typical Best Practices for Choosing Consumers who Participate

Conventionally, market researchers try to use large representative consumer panels. Often (but not always) this consists of testing panelists in different markets, and of different ages. Historically, most food products have been evaluated by females, except for products specifically designed for males (which are relatively few). In more recent years it has come to be realized that geographical dispersion does not guarantee representativeness in the population, because there are more fundamental ways in which people differ that transcend geography. One newly emerging concept is the existence of sensory preference segments, where individuals in a segment show rather similar sensory preferences to product features. These members of the same segment may live in different locations, have different incomes, education, and purchase history. The jury is still out on how to improve the consumer panelist selection process in light of these overwhelming differences among sensory preference segments.

Anchoring the Scale

Some practitioners feel that every point on the acceptance rating scale should be anchored with a verbal descriptor, because they feel that it is the descriptive term, not the numerical scale, that is important. Other practitioners feel that the scale itself is relevant, and that the only anchors need be at the top and the bottom of the scale.

The Position of the Acceptance Rating in the Order of Attribute Questions

Market researchers have not adopted any hard and fast rules about the nature of the questionnaire, and the location of the acceptance measure within the questionnaire. Some researchers feel strongly that the acceptance rating should come first in order for that rating not to be influenced by other attribute ratings. Other researchers feel just as strongly that the acceptance rating should come last, because otherwise the consumers will attempt to justify their acceptance rating by having all other attribute ratings fall into line. The majority of researchers pay little attention to this issue, but the issue arises from time to time.

Appropriate Statistics for Ratings

Most market researchers come from and/or operate within a sociology-oriented system, where the emphasis is on percentage of people who exhibit certain behaviors (such as accepting a food). Thus the majority of market researchers working with purchase intent scales use percent measures (percent rating the product as probably or definitely buy), with the appropriate statistical analysis being that used for percentages. In some cases (especially with liking ratings), the researchers will use means and standard deviations, so the appropriate statistics there are the ones used in parametric analyses (e.g., analysis of variance, regression).

See also: **Dietary Surveys**: Surveys of Food Intakes in Groups and Individuals; **Food Acceptability**: Market Research Methods; **Nutritional Assessment**: Importance of Measuring Nutritional Status; **Satiety and Appetite**: The Role of Satiety in Nutrition; **Sensory Evaluation**: Sensory Characteristics of Human Foods; Food Acceptability and Sensory Evaluation; Practical Considerations; Sensory Difference Testing; **Wines**: Wine Tasting

Further Reading

Amerine MA, Pangborn RM, and Roessler EB (1965) *Principles of Sensory Evaluation of Food*. New York: Academic Press.

Balintfy J, Duffy WJ and Sinha P (1974) Modeling food preferences over time. *Operations Research* 22: 711–727.

Chakrapani C (ed.) (2000) *Marketing Research: State-of-the-Art Perspectives*. Chicago, IL: American Marketing Association.

Graf E and Saguy IS (1991) *Food Product Development*. New York: Van Nostrand Reinhold/AVI.

Meiselman HL (1978) Scales for measuring food preference. In: Petersen MS and Johnson AH (eds) *Encyclopedia of Food Science*, pp. 675–678. Westport: Avi.

Meiselman HL (ed.) (2000) *Dimensions of the Meal: The Science, Culture, Business and Art of Eating*. Gaithersburg, MD: Aspen.

Moskowitz HR (1983) *Product Testing and Sensory Evaluation of Foods: Marketing and R&D Approaches*. Westport, CT: Food and Nutrition Press.

Moskowitz HR (1985) *New Directions for Product Testing and Sensory Analysis of Foods*. Trumbull, CT: Food and Nutrition Press.

Moskowitz HR (1994) *Food Concepts and Products: Just-In-Time Development*. Trumbull, CT: Food and Nutrition Press.

Smithies RA and Buchanan BS (1990) Substantiating a taste claim. In: *Proceedings of the NAD Workshop*. New York: Council of Better Business Bureaus, and the Advertising Research Foundation.

Thomson DMH (ed.) (1989) *Food Acceptability*. Barking, Essex: Elsevier Applied Science.

Van Trijp JCM (1995) *Variety-Seeking in Product Choice Behavior – Theory with Applications in the Food Domain*, Wageningen, the Netherlands: Mansholt Studies.

Wu L (ed.) (1989) *Product Testing with Consumers for Research Guidance*. Philadelphia: American Society for Testing and Materials.

FOOD ADDITIVES

Safety

S A Jordan, Food Directorate, Ottawa, Ontario, Canada

Background

The relationship between humans and food additives spans thousands of years. The initial use of materials added to foods was likely for the practice of food preservation. Food preservation in ancient times took a variety of forms, from purely physical processes such as drying to the addition of substances such as smoke, salt, vinegar, or alcohol. Archeological and written evidence of food preservation reveals the use of these substances as far back as 5000 years. While the preservation of food was a critical activity for ancient populations, it was not the only reason for adding substances to food. The ancient origins of spices as flavoring agents is well known; however, it is known that naturally derived materials such as saffron and certain berry juices were used by the Egyptians and Romans to color foods. Today, however, synthetic additives have largely supplanted their natural counterparts in producing desirable technical effects.

Generally, food additives are defined as substances added to food to produce a specified and desired technical effect. In some jurisdictions, the definition is somewhat broader and may include 'indirect food additives' such as packaging materials, which may be incorporated into food unintentionally. Taking the USA as an example, at the present time, there are about 3000 food additives approved for use, while Canada, for example, regulates only about 400. While there are obviously international differences as to what is considered a food additive, this summary deals with premarket safety evaluation procedures that may apply to chemicals that can be incorporated into food either by direct or indirect addition. Examples of direct food additives are provided in **Table 1**.

Benefits of Direct Food Additives

The benefits of food additives are many. An obvious benefit of using food additives is increased safety of products. The use of antimicrobial agents to prevent proliferation of bacteria or molds that may result in food poisoning contributes significantly to providing

Table 1 Examples of food additive categories

Anticaking agents	Emulsifiers	Preservatives
Antioxidants	Enzymes	Processing aids
Antimicrobial agents	Firming agents	Propellants
Colors	Formulation aids	Stabilizers
Curing agents	Fumigants	Surface finishing agents
Drying agents	pH control agents	Nonnutritive sweeteners

a safe food supply. Other ultimate benefits of food additives include a greater choice of foods in the marketplace, the availability of lower-priced foods and enhanced organoleptic properties of food products.

The Acceptable Daily Intake

While food additives provide numerous benefits for consumers and manufacturers, there are potential risks associated with the addition of chemicals which may have widespread use. Foods generally are considered to be safe when consumed in an *ad libitum* fashion for long periods of time, and the addition of substances to food must not cause undue risk when consumed in this fashion. In order to minimize any potential adverse effects of a new food additive, a thorough safety assessment is required. This type of assessment is aimed at determining a safe level of intake of material. This level is termed an acceptable daily intake (ADI), a concept first proposed by the Joint FAO/WHO Expert Committee on Food Additives (JECFA) in 1961. The ADI is the level of intake by a human of a food additive, per day, for which no adverse effects are expected when consumed over an entire lifetime. The acceptable daily intake for food additives usually is expressed as milligrams of the substance ingested per kilogram of body weight.

The concept of the ADI is based on the premise that, for most compounds, there exists a discrete threshold of exposure above which adverse effects may be produced; however, the concept of thresholds may not hold for substances that are able to produce cancer. Substances that are known or suspected of being carcinogenic to humans (especially genotoxic carcinogens) are not allowed as food additives.

Toxicity studies on specific substances, conducted as part of an industry submission of data to government agencies, or literature searches of earlier information, are geared to identify levels of these substances that produce no adverse effects (or, indeed, no effects at all). Acronyms used to identify such levels are NOEL (no observed effect level), NOAEL (no observed adverse effect level), LOEL (lowest observed effect level), and LOAEL (lowest observed adverse effect level). Any measured level of effect from experimental studies is associated with uncertainty in terms of how it can be applied to possible responses in 'real-life' situations. An ADI, applicable to humans, is calculated by dividing the NO(A)EL or LO(A)EL, usually derived from animal studies, by numerical factors in order to reduce the uncertainties associated with the available scientific data.

Types of uncertainties that typically result from the evaluation of experimental data are the extrapolation of data derived from animals studies to humans, the susceptibility of human subpopulations that may be overly sensitive to the compound under study, the use of data from which only a LOEL is available, the extrapolation of less than chronic data to chronic exposure, and any other inadequacies or deficiencies of the data. The uncertainty factor for a lack of chronic data would not normally apply to current data requirements for assessing new food additives, since chronic studies in animals would need to be submitted to the regulatory authority. The number of these 'uncertainty factors' (also sometimes termed 'safety factors') that are applied, and their numerical values, depend on the situation and the available data, and are assessed on a case-by-case basis. Table 2 outlines various uncertainty factors currently in use and their usual numerical value.

The application of uncertainty factors is somewhat subjective; however, validation exercises have provided evidence that the value of 10 is adequate for less than chronic to chronic extrapolation and LOEL to NOEL extrapolation, in most cases. Still, the use of uncertainty factors is usually tailored to suit the particular strengths and weaknesses of the database being assessed. Once an ADI has been established, it should not be considered an absolute threshold. Periodic consumption of an additive at levels above the ADI may be without appreciable risk, depending on the circumstances. As the level of exposure increases above the ADI, however, the risk of adverse effects increases as well. Owing to the subjective nature of the uncertainty factor approach in the derivation of ADIs, other mathematical approaches have been proposed.

The criticisms levelled at the uncertainty factor approach include the fact that the NOEL in scientific studies is simply one of the dose levels and may not reflect the true NOEL for the species in question. Also, the size of the study affects the size of the

Table 2 Typical uncertainty factors and their values

Uncertainty	*Uncertainty factor value*
Extrapolation from the average animal to the average human	10
Protection of sensitive populations of humans (intraspecies variation in susceptibility)	10
No NOEL, only a LO(A)EL identified from scientific studies	10
Extrapolation of data from less than chronic exposure to chronic or lifetime intakes (however, data on chronic exposure would be required for current food additive submissions)	10
Inadequacies of the data (e.g., certain types of studies missing, severity of effects associated with a LOAEL)	2–10

NOEL. Studies with large numbers of animals tend to provide a lower NOEL, and vice versa. This, in fact, penalizes the company for conducting a larger study, since a lower NOEL will be associated with a lower ADI. Another criticism is that the NOEL-uncertainty factor approach ignores any dose–response curve generated by multiple-dose studies. Included in the alternative approaches to the NOEL-uncertainty factor approach is the concept of benchmark dose, defined as the lower confidence limit on a dose producing some predetermined small increase (e.g., 1% or 10%) in response rate over the background level for a specific effect. Other variations on this method have been proposed; however, all rely to some degree on the use of uncertainty factors (e.g., for interspecies extrapolation). Advantages of alternative methods are that they take into consideration the dose–response and the variability of data, and in some cases allow the assessment of risk at exposures below the ADI. The disadvantages of these alternative methods are that they do not entirely avoid the use of arbitrary uncertainty factors and that most toxicological studies do not provide sufficient information to adequately characterize the dose–response curve. Based on these concerns, many organizations rely much more on the traditional NOEL-uncertainty factor approach in calculating appropriate ADIs for food additives.

Along with the assessment of an ADI for a food additive, another important step is the determination of how much of the additive is, or is likely to be, consumed in the general population. If the estimated daily intake is sufficiently above the ADI, concerns may be raised regarding the use of that additive in general commerce. Estimated daily intakes are calculated knowing in what foods the additive will be used and the amounts of those food products consumed by different populations of consumers. The use of proposed food additives can be limited to specific food categories to limit the consumption, if this is seen as necessary by the regulatory authority.

Toxicological Assessment of Food Additives

The toxicological assessment of proposed food additives encompasses a wide variety of tests ranging from relatively simple *in vitro* studies to complex multigeneration animal studies. The tests required for each proposed additive will vary depending on the nature of the additive itself. For proposed additives that are the same as either naturally and commonly consumed components of food or normal constituents of the human body (e.g., lecithin), the testing requirements may be much less stringent compared with additives which are novel synthetic compounds. Also, some regulatory authorities have variable minimal data requirements, depending on the expected toxicity of the additive and its anticipated consumer exposure. In this 'Level of Concern' system, increasing anticipated exposure triggers additional data requirements. As the potential toxicity of the product increases (based on structural characteristics), the exposure-based cut-off points for increased data requirements are lower. The potential toxicity is determined by comparing the structure of the proposed additive with those of known biological activity. Substances of low concern would include normal cellular constituents such as carbohydrates and fats, aliphatic hydrocarbons which are saturated and contain few functional groups, and inorganic chemicals, which are normal cellular constituents or inert gases. Those substances of high concern include aromatic hydrocarbons and halogenated and heterocyclic chemicals.

In all cases, the conduct of toxicity studies, and the interpretation of data derived from these studies, may reveal the need for further testing, which may not have been apparent at the outset. In some cases, the regulatory authority may see the need for further testing in order to address specific concerns. Since food additives generally are meant to be consumed in a long-term, *ad libitum* manner, the studies that play the most important role in assessing the safety of proposed food additives are those aimed at determining potential chronic and/or carcinogenic, along with reproductive effects. Other toxicological tests contribute to the overall understanding of the possible adverse effects of a substance on biological systems. Adequate dose levels in studies allow the definition of a dose–response and the assignment of a NOEL to the compound under study. The high dose employed in these studies should elicit some adverse effects. With any type of toxicological testing, there must be adequate methodologies employed to insure the resulting data are meaningful and to reduce any potential criticisms, which may put the results in dispute. Even so, animal methodologies do not always allow extrapolation to human exposure scenarios. For instance, the potential food additive exposure of human newborns and infants through artificial milk substitutes is not mimicked in rodent models where newborns are nourished only via maternal milk. Several regulatory authorities have issued guidelines for the appropriate conduct of toxicity tests. Both the Organisation for Economic Cooperation and Development (OECD) and the US Food and Drug Administration (FDA) have published documents in this area with the FDA's detailed Redbook II currently undergoing revision. The OECD's Guidelines for the Testing of Chemicals are comprehensive in their

coverage of a wide variety of testing methodologies and are updated as needed. The OECD guidelines have been endorsed by member countries, allowing the acceptance of data between all countries belonging to the OECD. In addition to the aforementioned guidelines, the concept of Good Laboratory Practice is important in assuring that testing laboratories are able to generate valid data. Some regulatory bodies publish principles of Good Laboratory Practice.

Types of Toxicity Tests

Short-term tests for mutagenicity/genotoxicity The potential of a test chemical to interact with or damage DNA (genotoxicity) and have carcinogenic activity can be assessed using a variety of tests, following exposure *in vitro* or *in vivo*, with cells obtained from microbial, mammalian, and human sources. Usually, a battery of short-term tests for food additives will include a test for microbial gene mutations, chromosomal aberrations in mammalian cells *in vitro*, and chromosomal aberrations in mammalian cells *in vivo*.

In vitro tests include the bacterial *Salmonella typhimurium* reverse mutation assay (Ames test) to detect base-pair substitutions and frameshift mutations. Mammalian cell lines include those from mouse lymphoma cells, various tissues from Chinese hamsters, and from human cells to measure chromosomal aberrations and gene mutations. Other *in vitro* tests sometimes conducted include DNA repair (unscheduled DNA synthesis) or sister chromatid exchanges and cell transformation. *In vivo* tests include those to detect chromosomal aberrations (including micronucleus formation) in either circulating blood cells or bone marrow cells, and DNA repair assays.

In vitro tests require an exogenous metabolic activation system that may not reflect metabolism in live animals. These tests tend to be overly sensitive and therefore may overpredict activity. However, they are generally based on well-established, easy-to-perform assays that are reproducible and can incorporate high doses of test substance. *In vivo* genotoxicity assays have the advantage of having an intact metabolic system and the ability to assess concurrently target organ distribution of the test substance and its metabolites. Also, in concert with other animal toxicity tests, the results of these types of assays are more readily applicable for extrapolation to humans. However, *in vivo* tests may be insensitive, requiring the expense of large numbers of animals and, frequently, large amounts of test material. Agents that cause cancer by nongenotoxic mechanisms will not be detected by genotoxicity tests, and these assays do not replace bioassays that are aimed at detecting cancer.

Toxicokinetics/pharmacokinetics Results from these studies are useful in the interpretation of data derived from other investigations and provide information on the absorption, distribution, metabolism, and excretion of the test substance. This information also is useful in the development of further studies. For example, if it is known that a substance passes through the gut unabsorbed, the design of other studies can be tailored accordingly.

Acute toxicity tests Generally, acute oral toxicity tests are used where no information on a particular substance is available. The data derived from these single, high-dose, studies are limited, but provide information on maximum tolerated doses and may give information on the mechanism of action of the compound and the organ systems that may be affected.

Short-term, repeated-dose, toxicity tests Usually conducted for 28 (or sometimes 14) days, these short-term, oral, multiple-dose studies are designed to gather preliminary information on the effects of the daily administration of test substances on such parameters as hematology, clinical chemistry, and pathology. Also, the data derived can be used to narrow the focus of future studies to any identified target organs and to determine appropriate doses for further, longer-term, studies. Short-term tests generally are conducted on a rodent (mouse or rat) and a nonrodent species (usually dogs).

Subchronic toxicity tests Subchronic studies are usually conducted for 90 days and generally involve parameters and testing methodologies similar to shorter-term studies. Again, both rodent and nonrodent species are employed in subchronic studies. These investigations will reveal significantly more information on physiological and pathological effects of the test chemical than will shorter-term studies. Focused studies on other parameters such as immunotoxicology and neurotoxicology are included in subchronic testing strategies.

Chronic and carcinogenicity studies These two types of studies may be conducted separately; however, if a sufficient number of animals are used, the chronic and carcinogenicity aspects may be combined in a single study. The length of the studies is to include the majority of the lifespan of the animal; however, animals in the chronic exposure arm of the study may be removed earlier. Generally, two rodent species are employed (usually rats and mice), and the study is conducted for 2 years with daily dosing of multiple dose levels. In some cases, animals are also exposed during gestation. These long-term studies are the

most comprehensive, with a considerable variety of parameters and toxic endpoints being examined and the requirement of high standards of analysis, including pathology and statistics. NOELs derived from chronic studies are important, since the exposure scenario can be extrapolated to the real-life, long-term, human situation with food additives. Important goals of all long-term studies are to develop dose–response information and to identify clear LOELs and NOELs that can be used for ADI calculations. (*See* **Carcinogens**: Carcinogenicity Tests.)

Reproduction, teratology, and developmental toxicity studies The aim of these investigations is to determine the potential effects of the test substance on the ability of animals of both sexes to produce viable and healthy offspring and on the subsequent development of these progeny. Most studies in this regard involve continuous dosing through two generations of test animals (rats or mice). Males are dosed in advance of mating to elucidate any effects of the test compound on one complete spermatogenic cycle. Females should be dosed for at least two estrous periods before attempted mating. In this way, the first-generation offspring are exposed from the moment of conception to the weaning of their own (second-generation) offspring. Reproductive parameters measured include gonadal function, mating behavior, time to conception, fetal viability, parturition, and lactation. Developmental parameters include morbidity and mortality, growth, and behavior. As part of the reproductive studies, or investigated in its own right, the potential of the test material to produce physical or functional birth defects must be determined. These teratology studies are conducted by exposing the fetuses during *in utero* development for the time that corresponds to the development of various organ systems. Just before expected parturition, the fetuses are examined for skeletal and visceral abnormalities.

Other toxicity tests The need for further studies may become apparent during other investigations. These include focused studies on the potential toxicity of the substance on the immune or nervous systems.

Human studies After significant evaluation in animals, human testing is sometimes conducted. Studies in humans usually are proposed in the case of food additives that are expected to have a wide use by a large proportion of consumers (e.g., sugar or fat substitutes). While toxicological studies obviously are not conducted in people, these human studies provide information on possible unwanted effects (such as the gastrointestinal tolerability of the additive) resulting from more realistic doses and use patterns than can be achieved with animal studies.

Interpretation of Toxicological Data

The data derived from the aforementioned studies should provide sufficient information to enable the calculation of an ADI for the particular proposed food additive. It is important, however, to keep in mind that it is not unusual for different professional opinions to exist on the significance of toxicity of a substance based on the same data. Reasons for divided opinion are varied and may include differing views on the potential severity of effects noted and on the extrapolation of animal data to the human situation. In some cases, debate has occurred as to the meaning of results from animal studies where effects noted (including tumor development) are present in biological structures that have no counterpart in humans or where the metabolism of the compound differs between animals and humans. Well-known examples of food additive substances that have produced significant debate are saccharin and butylated hydroxyanisole (BHA).

Saccharin, a sugar substitute that was generally regarded as safe in the USA until 1972, has been linked to bladder cancer in rats when exposed *in utero* or from birth, with males affected more than females. A NOAEL of 3% saccharin in the diet was established in a large two-generation study in rats, and no tumors were noted with 1% saccharin in the diet. The factors contributing to the development of bladder neoplasms by saccharin are species-specific. While saccharin is nongenotoxic, it has been suggested that its ability to cause bladder cancer results from the complexing of saccharin ions with urinary proteins with precipitation and the formation of crystals. The crystals may act to damage the walls of the bladder with increased cell proliferation. The protein suspected of being involved in the crystal formation (α-2-microglobulin) is present in higher concentrations in the urine of male rats compared with female rats, mice, and humans. This and other species differences in urine composition suggest that humans may not be susceptible to the carcinogenic action of saccharin. Epidemiologic studies in human populations using saccharin have not revealed any increase in bladder cancer.

BHA is a phenolic antioxidant used mainly in fats and oils. It has been shown to have a low acute toxicity. In chronic exposure investigations, no consistent adverse effects have been noted with up to several per cent BHA in the diet, with the exception of forestomach hyperplasia and tumor formation. Also, it has been shown to have a low potential of being genotoxic. Malignant tumors of the forestomach

have been noted in rats fed 2% BHA in the diet for 2 years, which equals almost 1 g of BHA per kilogram of body weight. No such tumors were noted when the rats were fed 0.5% BHA in their diet. In other rodent species with forestomachs, BHA has produced hyperplasia, and benign tumors have been noted in the hamster forestomach. No species other than the hamster and rat have been shown to develop carcinomas in this organ. The carcinogenic effect in certain rodent species appears to be due to indirect, rather than direct, effects of BHA on DNA. Since the forestomach is an organ that is not present in humans, the relevance of the findings in animals is questionable. In addition, the effects noted in the positive animal models were observed at very high doses, and a threshold for the formation of carcinomas in rodents was identified.

Because of new developments in the scientific literature or in the interpretation of toxicological data, it is essential that regulators keep up to date with the toxicological literature. The generation of substantial new evidence may allow a refinement or reassessment of the safety of specific food additives.

Emerging Issues

The continuing evolution of the food industry means that new types of products come into use on a continuing basis. Novel types of food additives and packaging methods and materials arise that present unique challenges to the toxicologist and regulatory authorities in terms of appropriate testing strategies.

In recent years, certain plants and microorganisms have been modified through biotechnology, including genetic modification. Food additives, such as enzymes, derived from genetically modified organisms (or the addition of the modified organisms themselves) present challenges in safety assessment. Issues to consider in the assessment of these types of food additives include the exposure to foreign DNA and proteins (and thus considerations of allergenicity), increased exposure to these types of products as the market increases, and the necessity of microbiological purity if products are derived from modified microorganisms. (*See* **Biotechnology in Food Production; Genetically Modified Foods**.)

The advent of 'functional foods' (foods containing ingredients demonstrated to have specific health benefits, in addition to providing basic nutrients and nutritional benefits) has led regulatory authorities not only to develop ways of regulating health claims associated with these products but also to consider the inherent safety of these products. Globalization of markets can result in new types of food and food ingredients being introduced. The addition of medicinal-type herbs, such as *Echinacea* species or St. John's Wort (*Hypericum perforatum*) to conventional foods, for perceived health benefits, necessitates an assessment of the safety of these new food products, since herbal ingredients may be consumed in ways that are very different from their use in traditional systems of medicine.

That young children are different from adults in many aspects of their physiology is well known, and this must be taken into account in the assessment of food additives that may be consumed by this population. In recent years, the applicability of the ADI-uncertainty factor approach to insuring the safety of food additives consumed by children has come under scrutiny, and this subject is one of ongoing discussion in many countries.

See also: **Biotechnology in Food Production**; **Carcinogens**: Carcinogenicity Tests; **Genetically Modified Foods**

Further Reading

Branen AL, Davidson PM and Salminen S (1990) *Food Additives*. New York: Marcel Dekker.

Ditschum TL and Winter CK (2001) Food additives. In: Helferich W and Winter CK (eds) *Food Toxicology*, pp. 187–202. Boca Raton, FL: CRC Press.

Maga JA and Tu AT (eds) (1995) *Food Additive Toxicology*. New York: Marcel Dekker.

Middlekauff RD and Shubik P (eds) (1989) *International Food Regulation Handbook. Policy, Science, Law*. New York: Marcel Dekker.

Organisation for Economic Cooperation and Development (1992) *The OECD Principles of Good Laboratory Practice. Environment Monograph 45*. Paris: OECD.

Organisation for Economic Cooperation and Development (1993) *OECD Guidelines for the Testing of Chemicals*. Paris: OECD.

Rodericks JV (2000) Food constituents, additives, and contaminants. In: Lippmann M (ed.) *Environmental Toxicants. Human Exposures and Their Health Effects*, pp. 377–408. New York: John Wiley.

Rodericks JV, Frankos VH and Plunkett LM (1995) Food additives. In: Chengelis CP, Holson JF and Gad SC (eds) *Regulatory Toxicology*, pp. 51–81. New York: Raven Press.

US Food and Drug Administration (1993) *Toxicological Principles for the Safety Assessment of Direct Food Additives and Color Additives Used in Food. 'Redbook II'*. Washington, DC: U.S. FDA, Center for Food Safety and Applied Nutrition.

US Institute of Medicine (1999) *Workshop Summary: Enhancing the Regulatory Decision-Making Approval Process for Direct Food Ingredient Technologies*. Washington, DC: National Academy Press.

Food Allergy *See* **Food Intolerance**: Types; Food Allergies; Milk Allergy; Lactose Intolerance; Elimination Diets

Food Analysis *See* **Analysis of Food**; **Chromatography**: Principles; Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; Supercritical Fluid Chromatography; Combined Chromatography and Mass Spectrometry; **Immunoassays**: Principles; Radioimmunoassay and Enzyme Immunoassay; **Mass Spectrometry**: Principles and Instrumentation; Applications; **Spectroscopy**: Overview; Infrared and Raman; Near-infrared; Fluorescence; Atomic Emission and Absorption; Nuclear Magnetic Resonance; Visible Spectroscopy and Colorimetry

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS

E Casadei and J Albert, Joint FAO/WHO Food Standards Programme, Rome, Italy

History and Purpose of the FAO

The Food and Agriculture Organization of the United Nations (FAO) is an autonomous agency within the United Nations (UN) system. The organization had a membership of 180 countries in 2001.

In 1943, world leaders met at Hot Springs, Virginia, USA and determined to lay a framework for peace through international cooperation in food and agriculture. The meeting, the UN Conference on Food and Agriculture, decided on the establishment of permanent organization in the field of food and agriculture and set up an interim commission to prepare a specific plan. This commission worked for 2 years and prepared the constitution of the FAO. The signing of the constitution and the opening of the first FAO Conference were arranged to take place in Quebec, Canada, on 16 October 1945. Representatives from 44 nations met at this date and established the FAO, the first of the UN agencies. As the commemorative tablet in Canada states, 'Thus for the first time nations organized to raise levels of nutrition and to improve production and distribution of food and agricultural products.' The FAO's goal is enshrined in its motto, '*Fiat Panis* – Let there be bread.'

The Conference is the FAO governing body and is composed of all member nations. It meets every 2 years and is a major forum for the discussion of food and agriculture issues. It is able to take concerted action on issues of concern to its members. Recent initiatives include the adoption of a World Food Security Compact and an International Code of Conduct on Pesticide Distribution and use.

The Conference elects a Council to which a series of committees, such as the Committees on Commodity Problems, Fisheries, Agriculture, Forestry, and World Food Security, report. Many bodies have been set up by the FAO Conference and Council to foster cooperation in particular subject areas. In adopting the International Undertaking on Plant Genetic Resources in 1983, for example, the Conference provided for the establishment of a commission on Plan Genetic Resources to guide the FAO on policy and to monitor the implementation of the undertaking. The Commission on African Animal Trypanosomiasis was established in 1979 to coordinate a large-scale FAO program to control the disease. Another example is the cooperation of the FAO with the World Health Organization (WHO) through a joint FAO/WHO Codex Alimentarius Commission, which has issued some 200 international food standards. (*See* **Legislation:** Codex.)

Along with these bodies, FAO organizes conferences and meetings on particular issues. For example, the FAO-serviced World Food Conference in 1974

resulted in a series of measures to promote world food security. In December 1992, the first International Conference on Nutrition (ICN) was held in the FAO's Headquarters in Rome. The conference was jointly sponsored by the FAO and the WHO. At this conference, a World Declaration and Plan of Action for Nutrition were unanimously adopted, committing governments to prepare their National Plans of Action for Nutrition containing attainable goals and measurable targets based on the principles and relevant strategies agreed at theICN. World leaders assembled in Rome in November 1996 for the World Food Summit aimed at renewing global commitment to the fight against hunger. The FAO called the Summit in response to widespread undernutrition and growing concern about the capacity of agriculture to meet future food needs.

Besides being a major forum for the discussion of food and agriculture matters, the FAO is an information center collecting data related to its field of work. It also acts as an adviser to governments, being a useful source of independent advice on agricultural policy and planning, on the administrative and legal structures needed for development, and on ways of ensuring that national development strategies are directed towards rural development and the alleviation of poverty and malnutrition. Its role as a development agency should be highlighted: the FAO gives direct, practical help in the developing world through technical assistance projects in all areas of food and agriculture (including fishery and forestry). These field projects strengthen local institutions, assist research and training, and develop and demonstrate new techniques.

Food Security and Nutrition

The idea of international action to achieve world food security was launched by the FAO in early 1973. The international undertaking in world food security contains a set of general principles for stockholding, special assistance to developing countries (including aid to increase production), the exchange of information, and arrangements for consultations. The concept has developed further. Today, the objective of the world food security concept is to ensure that all people at all times have both physical and economic access to the basic food they need. Its more specific aims are related to production, with special emphasis on low-income food-deficit countries, stability in the flow of supplies, and the problems of access to available supplies on the part of needy nations and social groups.

The search for policies that would cope with food shortages and surpluses has been one of the main threads in FAO history. Various activities contributed to the preparation of such a policy, including the following: expanding production, improving nutrition, stabilizing prices, obtaining a fair return for producers, and coping with famine.

Nutrition was a dominant theme at the Hot Springs Conference, with the adoption of eight resolutions dealing exclusively with various aspects of nutrition, far more than on other subjects. There is no single formula for curing malnutrition. The best approaches vary, according to local and national circumstances. The development of policies for food and nutrition is largely a question of orienting sectoral activities to have an impact on nutrition. Thus, from the mid-1970s, the FAO worked closely with a number of governments on the incorporation of nutritional objectives into development plans.

A mass of information on every aspect of food and agriculture (including production, supply, demand, prices, and technology) is needed if government planners, traders, scientists, farmers, and technologists in every country in the world are to make rational decisions on planning, investment, marketing, research, or training. The FAO provides that essential information for food policy planning. For example, the Global Information and Early Warning System for Food and Agriculture monitors production conditions and food supply prospects world-wide and issues warnings when there are indications of trouble. The role of world food surveys cannot be overemphasized in the assessment of world nutrition status. The FAO prepares these surveys for collecting essential data on the world's food problems. These data are analyzed and distributed to all member countries and concerned bodies as reference data for the elaboration of food policies at world or national levels. At the time of writing, the latest survey is the sixth one, published in 1996.

In addition to its tasks of collecting, analyzing and disseminating information, advising governments on food and agriculture (including fishery and forestry) policy and planning, and providing opportunities for governments to meet and discuss food and agriculture problems, the FAO carries out a major program of technical advice and assistance for the agricultural community on behalf of governments and development funding agencies. The FAO is involved at every stage of agricultural development, from demonstrating new cultivation techniques to subsistence farmers, to advising governments on how to achieve more stable and equitable international trade in agricultural commodities.

Food Control

In the past several decades, nations have experienced extensive changes and faced new or greater problems in assuring adequate national food security, trade, and health. This trend is continuing for the following reasons: intensification of farm production and the centralization of food manufacturing; the development of large urban markets remote from the agricultural production areas; markets and consumer demands for consistent supply of safe, uncontaminated and good-quality food and food products; the economic dependency of many countries on revenues from agricultural food exports; and the rapid growth in global food trade.

Within the overall mandate of the FAO with regard to production, distribution and marketing of food, and for nutrition improvement, the necessity of providing a safe, wholesome, and nutritious food supply, and the protection of this supply from losses, waste, deterioration, contamination, and adulteration is implicit. It is interesting that the Hot Springs Conference noted that the FAO would play a role in insuring 'protection of the public, through the medium of pure-food laws, against impurities for adulteration and against unfair competition and undesirable trade practices.' It recommended that the FAO should 'assist governments to extend and improve standards of nutrient content and purity of all important foods.' In view of this, the FAO has developed programs to assist governments to orient their country policies and programs so as to include nutrition improvement and food protection as specific objectives.

Food safety and quality start at the farm level and continue throughout the processing and distibution chain to storage and final preparation by the consumer. Good agricultural, processing, distribution, and marketing practices are essential to ensure consumer protection. Now, more than ever, food control programs are needed in order to assure that foods are safe for consumption and are sold to the consumer on fair terms; furthermore, such programs should help to reduce food losses caused by improper handling or spoilage of raw materials, prevent 'dumping' practices of shipping nonstandard (low-quality or contaminated) food products to countries that do not have the capability of food control, and increase the earnings of foreign currency necessary for the implementation of development plans. Food control, including standards, is therefore an integrated part of the programs developed by the FAO.

Issues relating to food safety and food control draw considerable public attention and represent an important concern for governments, consumers, and the industry alike. They need to be addressed at the national level by adequate legislation, which should be developed in cooperation with all interested sectors and implemented efficiently.

The Economic and Social Department of the FAO, and especially its Food and Nutrition Division (ESN) are the main bodies responsible for the programs oriented to guarantee food quality and safety and consumer protection. Within this division, the Food Quality and Standards Service focuses on activities in the field of food control. These activities are implemented through the Food Standards Group, which provides the Secretariat for the Codex Alimentarius Commission of the Joint FAO/WHO Food Standards Program, and the Food Quality and Consumer Protection Group, which assists developing countries in the formulation and development of effective food control programs and gives scientific and technical support to the Codex Alimentarius Commission. Technical assistance provided by the FAO in food control is based upon the fundamental elements that all effective programs must consider:

- a basic food law complemented by detailed regulations that provide for correct hygienic practices in food and agricultural production and processing, the establishment of food standard and codes of practice, the correct use of food additives, pesticides, etc., and the proper use of labeling to avoid consumer deception;
- an effective administration of such laws and regulations through an organization which includes technical officers, inspectors, laboratory personnel and compliance officers, together with the necessary infrastructure to carry out their tasks adequately;
- a well-informed consumer, aware of the problems related to food control, especially with respect to good handling practices and storage; and
- food producers, processors, and handlers who cooperate with government agencies in assuring food quality and safety.

In order to insure food security, facilitate trade and protect consumers, the development of the following activities receives special consideration from the FAO:

- support of developing countries in strengthening their national strategies for quality and safety, food control systems and contamination monitoring programs;
- preparation of food legislation, regulations, and standards on a national, regional, or international basis accordingly to meet changing consumer demands (e.g., food irradiation, biotechnology),

new food ingredients (e.g., vegetable proteins), and new food safety concerns (e.g., veterinary drug residues, radionuclides); and
- technical assistance in training for food control inspectorate and laboratory personnel and for food control management staff. (*See* **Legislation**: History; International Standards.)

Joint FAO/WHO Food Standards Program Codex Alimentarius Commission

The Hot Springs Conference noted the role of government services in insuring the protection of the public. It recommended that the FAO, when established, should 'consider the formulation and adoption of international standards to facilitate and protect the interchange of such products between countries, and agree upon international methods of determination.'

In 1958, work was begun by an FAO Committee of Government Experts, in collaboration with the International Dairy Federation, on the establishment of a Code of Principles concerning Milk and Milk Products and compositional standards for milk products. A further development that took place in 1958, and which was to prove to be of great importance to both the FAO and WHO in introducing an international program for the elaboration of food standards, was the creation of a body known as the Codex Alimentarius *Europaeus*. In 1961, the Council of the European Codex Alimentarius adopted a resolution proposing that the work should be taken over by the FAO and WHO. From that moment, matters moved quickly, and the Codex Alimentarius Commission was established in 1962 by the decision of the Conference of the FAO and the Executive Board of the WHO.

The basic objectives were precisely those put forward at Hot Springs: removal of nontariff trade barriers caused by differing national food legislation, insuring fair practices in international trade, and protection of the consumer against health risks and fraud. The Commission assumed the role of an international forum for elaborating standards for foods moving in international trade and for providing guidance to countries wishing to create their own national food laws and regulations.

The Uruguay Round of Multilateral Trade Negotiations concluded in Marrakesh in April 1994. It established a new World Trade Organization (WTO). Amongst the agreements were the Agreement on the Application of Sanitary and Phytosanitary Measures (the SPS Agreement) and the Agreement on Technical Barriers to Trade (the TBT Agreement). Both of these agreements are relevant in understanding the international dimensions of food. The SPS Agreement encourages the use of international standards, guidelines, or recommendations where they exist and identifies those of Codex (relating to food additives, veterinary drugs and pesticides residues, contaminants, methods of analysis and sampling, and codes and guidelines of hygienic practices), as consistent with provisions of the SPS. Thus, the Codex standards serve as a benchmark for a comparison of national sanitary and phytosanitary measures. The TBT Agreement also recognizes international standards where they exist. It requires that technical regulations on traditional quality factors, fraudulent practices, packaging labeling, etc. imposed by countries not be more restrictive on import products than they are on products produced domestically.

Membership of the Codex Alimentarius Commission is open to all Member Nations and Associate Members of the FAO and/or WHO. In 30 years, the Commission has grown from the original 30 countries that were present at its First Session in 1963 to 165 member countries in 2001. The distribution of Member Countries is as follows: Africa, 41; Near East 19; Asia, 21; Europe, 40; Latin America and the Caribbean, 31; North America, 2; Southwest Pacific, 11. The Commission has an Executive Committee and 29 subsidiary bodies. All of these bodies, like the Commission, are intergovernmental in character. The composition of the Executive Committee is such as to insure an adequate representation of the various geographical areas of the world to which the members of the Commission belong. The Executive Committee makes recommendations to the Commission regarding the general orientation of the program of the work.

The main activities of the Commission have been concerned with the development of international world-wide standards for foods as well as other texts on related matters, their publication in the Codex Alimentarius, and the registration of acceptances by member governments of standards. In addition, the Commission, through its subsidiary bodies, is concerned with the elaboration of advisory Codes of Practice, which are adopted within the Codex. The Commission determines priorities, and initiates and guides the preparation of draft standards, whether by its own subsidiary bodies or with the assistance of other appropriate organizations.

During its lifetime, the Codex Alimentarius Commission has elaborated more than 230 individual commodity standards, 37 codes of hygienic and technological practices, and maximum residue limits for pesticides covering more than 3000 potential pesticide plant and animal combinations. The Codex Alimentarius itself contains standards on the following:

processed, semiprocessed, or nonprocessed food products; raw materials used in manufacturing; labeling; nutrition information; food additives; contaminants; pesticide residues; veterinary drug and hormone residues in food; methods of analysis; sampling; codes of practice; and guidelines and other recommendations. All the above have been published in 16 volumes of the Codex Alimentarius.

Labeling

Food labeling is the primary means of communication between the producer and the consumer. One of the most vexatious problems that exporters of foodstuffs are likely to encounter is the wide disparity between different countries' requirements with regard to labeling. The work of Codex in devising and issuing a General Standard on the Labeling of Pre-Packaged Foods must be regarded as a landmark, a major achievement, in international recommendations for food legislation.

Codex has issued Guidelines for the production, processing, labeling, and marketing of organically produced foods. These guidelines set out the principles of organic production at farm, preparation, storage, transport, labeling and marketing stages, and provide an indication of accepted permitted inputs for soil fertilizing and conditioning, plant pest and disease control, and food additives and processing aids. For labeling purposes, the use of terms inferring that organic production methods have been used are restricted to products derived from operators under the supervision of a certification body or authority.

Nutrition Information

Consumer demand for nutrition information is growing, stimulated by the media, whose interpretation and presentation of the views of nutrition scientists and on national policies are not always balanced or accurate. Few consumers have any systematic training in nutritional matters or easy recourse to reliable information sources. The natural concern for the family health leads them to use the media as their information source. The problem faced by Codex was how to regulate it so that accurate and balanced information could be presented without creating a rigid system. The Guidelines on Nutrition Labelling have been issued to insure that nutrition labeling is effective:

1. providing the consumer with information about a food so that wise choices can be made;
2. providing a means for conveying information of the nutrient content of a food on the label;
3. encouraging the use of sound nutrition principles in the formulation of foods that would benefit public health; and
4. providing the opportunity to include supplementary nutrition information on the label

Nutrition labeling should not describe a product or present information that is in any way false, misleading, deceptive, or insignificant in any manner, to insure that no nutritional claims are made without nutrition labeling.

Food Additives

The use of food additives is becoming more important in the production and world-wide distribution of food. As the demand for variety increases, and with it the desire to purchase goods from distant lands, the importance of food additives will increase, rather than diminish. None the less, this is a topic of great consumer concern, and Codex has always treated it as such through the Codex Committee on Food Additives and Contaminants (CCFAC). The CCFAC considers technological needs on the one hand from information supplied by Codex Commodity Committees. On the other hand, it applies safety considerations based on the reports of the Joint Expert Committee on Food Additives. The CCFAC follows a risk-based approach to develop the General Standard for Food Additives, which includes a list of additives that can be used in accordance with Good Manufacturing Practice and specific levels of use for a number of additives for which an acceptable daily intake (ADI) has been established.

Contaminants

The term 'contaminants' is used to denote the presence in food of traces of undesirable elements of substances, e.g., traces of lead, cadmium, polychlorinated biphenyls, or even radioactive fallout. The CCFAC is responsible for establishing permitted maximum or guideline levels for individual food additives, for contaminants (including environmental contaminants), and for naturally occurring toxicants in foodstuffs and animal feeds. The establishment of maximum levels for contaminants is a complex subject, and there are specific difficulties to address in order to reach an agreement at the international level in this area. Sometimes, the complete or updated risk assessment is lacking, especially exposure assessment, and there is a need to update the scientific information. In addition, legislation in Member Countries differs widely, especially between exporter and importer countries, and it is difficult to come to a

consensus; there are also several technical problems to be addressed, especially as regards methods of analysis and sampling or inspection procedures.

Pesticide Residues

The use of pesticides is essential in agriculture and horticulture if crop yields are to be economically viable and, indeed, if crops are to be successful. The fact that residues can remain in the food is another topic of considerable consumer and even wider concern. There is well-justified concern that an unrestricted and uninformed use of these substances could damage health. The Codex Committee on Pesticide Residues, on the basis of the work of the FAO/WHO Joint Meeting (Committee of Experts) on Pesticide Residues (JMPR), has established maximum residue limits (MRLs).

Maximum Residue Limit (MRL) is the maximum concentration of a pesticide residue (expressed as milligrams per kilogram) recommended by the Codex Alimentarius Commission to be legally permitted in or on food commodities and animal feeds. MRLs are based on Good Agriculture Practice data, and foods derived from commodities that comply with the respective MRLs are intended to be toxicologically acceptable.

Codex MRLs that are primarily intended to apply in international trade are derived from estimations made by the JMPR as follows:

1. toxicological assessment of the pesticide and its residue; and
2. review of residue data from supervised trials and supervised uses including those reflecting national food and agricultural practices. Data from supervised trials conducted at the highest nationally recommended, authorized, or registered uses are included in the review. In order to accommodate variations in national pest control requirements, Codex MRLs take into account the higher levels shown to arise in such supervised trials, which are considered to represent effective pest-control practices. Consideration of the various dietary residue estimates and determinations both at the national and international level in comparison with the ADI should indicate that foods complying with Codex MRLs are safe for human consumption.

Food Hygiene

Poor food hygiene is a major cause of human illness. Most hygiene problems are caused by faulty handling of food and, in the case of processed or packaged food, often occur after manufacture and packaging. There are powerful reasons for the great attention paid by Codex to food hygiene. In matters of hygiene, it is especially true that good practices and economic benefits go hand in hand. The establishment of effective good manufacturing practice as regards hygiene is relatively simple in principle but is far from simple in practice.

Codex has issued a Code of Practice and General Principles of Food Hygiene. These General Principles lay a firm foundation for insuring food hygiene. The document follows the food chain from primary production through to final consumption, highlighting the key hygiene controls at each stage. It recommends a Hazard Analysis and Critical Control Points (HACCP)-based approach, wherever possible, to enhance food safety. The HACCP system, which is science-based and systematic, identifies specific hazards and measures for their control to insure the safety of food. HACCP is a tool to assess hazards and establish control systems that focus on prevention rather than relying mainly on end-product testing.

Principles for the establishment and application of microbiological criteria for foods have been approved by the Codex Alimentarius Commission. These principles are intended to give guidance on the establishment and application of microbiological criteria for foods at any point in the food chain from primary production to final consumption.

Current Issues

Matters related to biotechnology are at the center of considerable debate in many countries and also at the international level. In the framework of Codex, the Commission decided to establish the Task Force on Food Derived from Biotechnology, in order to insure a comprehensive approach and to consider safety, nutrition, and all relevant aspects of genetically modified organisms (GMOs) and food produced from GMOs.

Several food safety issues are related to animal feeding, especially contamination of feed with mycotoxins and other contaminants. This question drew the attention of the media owing to the recent occurrence of contamination with dioxins in feeds, but there are also several other issues, including the Bovine Spongiform Encephalopathy. The Commission decided to establish a specific Task Force on Animal Feeding to consider all aspects that can affect food safety. Another important emerging issue in relation to animal production is antimicrobial resistance in bacteria, and in view of its complexity, it should be addressed through a multidisciplinary approach. It is currently under consideration by the Committee on Food Hygiene and the Committee on

Residues of Veterinary Drugs in Foods and is also relevant to the work of the Task Force.

Future Directions

The fundamental objective of the Codex Alimentarius Commission is to establish sound internationally agreed guidelines for national food control systems based on the criteria of consumer health protection and fair practices in trade, and taking into account the needs and special concerns of all countries.

The new recognition and status that Codex standards, guidelines, and other recommendations acquired under the WTO Agreement on the Application of Sanitary and Phytosanitary Measures (the SPS Agreement) also brought new challenges and responsibilities, including the need to insure that its standards and related texts are based on scientific principles and meet the needs and mandate of the organization. The WTO Agreement on Technical Barriers to Trade is also of great relevance, given the significance of the provisions pertaining to product description, labeling, packaging, and quality descriptors for consumer information and fair practices in trade. Although quality provisions are fundamentally driven by the market, the CAC has an important role in insuring that provisions relating to quality are sound and based on the criteria of essentiality and do not constitute disguised barriers to trade.

See also: **Legislation**: History; International Standards; Codex

Further Reading

Codex Alimentarius Commission (1989) *Codex Alimentarius Commission Procedural Manual*, 7th edn. Rome: Food and Agriculture Organization/World HealthOrganization.

FAO (1976) *Guidelines for Developing an Effective National Food Control System. FAO Food Control Series No. 1, WHO Food Control No. 1*. Rome: Food and Agriculture Organization.

FAO (1985) *FAO, The First 40 Years*. Rome: Food and Agriculture Organization.

FAO (1987) *Introducing Codex Alimentarius*. Rome: Food and Agriculture Organization.

FAO (1988) *FAO, What it is, what it does*. Rome: Food and Agriculture Organization.

FOOD AND DRUG ADMINISTRATION

A Miller and T Nordenberg, FDA's Center for Food Safety and Applied Nutrition, Maryland, MD, USA

Historical Introduction

The Food and Drug Administration (FDA) is an Executive Branch agency within the US Department of Health and Human Services (HHS). Its regulatory obligations include overseeing the safety of about 80% of the domestically produced and imported food in the USA – specifically, all food except meat, poultry, and processed eggs.

Long before the federal government of the USA began to regulate the country's food and drug products, individual states had made independent efforts to regulate them. But despite the threat of punishment for selling adulterated and mislabeled foods, the state laws, including the first-of-its-kind Massachusetts food and drug law passed in 1785, proved unsuccessful in preventing the marketing of violative food products.

It was not until 1906 that the first federal Pure Food and Drug Act was enacted, which made it a federal crime to sell mislabeled or adulterated food, drink, and drugs intended for interstate commerce. This cornerstone legislation of the future Food and Drug Administration was passed following the 1905 publication of Upton Sinclair's provocative novel *The Jungle*, which exposed the unsanitary conditions in American meatpacking plants, and the public crusade of Dr. Harvey Wiley, the chief chemist of the US Department of Agriculture (USDA) in the late 1800s and early 1900s, to reform the food and drug laws to protect consumers from harmful preservatives and adulterants.

Wiley, who came to be known as the 'Father of the Pure Food and Drugs Act,' was appointed as the first administrator of the landmark 1906 legislation. Under Wiley's leadership, the staff at the USDA's Bureau of Chemistry improved upon existing scientific methods

for analyzing food and refined legal procedures and inspection methods helpful in enforcing the law.

The first separate federal food and drug agency was established in 1927, and took the name 'Food and Drug Administration' in 1931. While the 1906 law laid the agency's foundation, Congress has continued to update the act over the years, ultimately assigning FDA responsibility for assuring the safety of most US food products. The Federal Food, Drug, and Cosmetic Act, signed into law by President Franklin D. Roosevelt in 1938, remains the basic food and drug law of the USA. Three amendments to the 1938 law had important implications for food regulation: the Miller Pesticide Amendment in 1954, the Food Additives Amendment in 1958 (which includes the 'Delaney Clause'), and the Color Additive Amendments in 1960.

Despite the legal amendments, the overall goal of the federal food law has remained constant: protecting US consumers from adulterated and misbranded food products, whether they are produced domestically or imported into the USA. It is the Food and Drug Administration's responsibility to implement the food and drug laws to protect and promote the health and economic interests of the American public.

Recent statistics from the Centers for Disease Control and Prevention illuminate the enormity of this challenge: foodborne illness, according to CDC, causes approximately 76 million illnesses, 325 000 hospitalizations and 5000 deaths each year in the USA. And in their quest to improve the safety of the US food supply, food scientists and regulators continue to be faced with ever-changing hurdles: for example, US consumers are eating more seafood, fresh produce (especially imported produce), and more ready-to-eat foods, as well as more foods prepared at restaurants and institutional settings such as nursing homes and hospitals. Also, up to 25% of the US population is considered 'at risk' for foodborne illness. Vulnerable populations include pregnant women, children, the elderly, and those who are immunocompromised. But perhaps the most perplexing challenge of all comes from evolving bacteria such as new strains of *Escherichia coli (E. coli)* 0157:H7.

Today's FDA: Mission and Organization

Along with setting the standards and ensuring the safety of over 75% of the food consumed in the USA, FDA is responsible for ensuring the safety of animal feed and drugs and all human medications and vaccines, blood products and tissues for transplantation, medical devices, devices that emit radiation, and cosmetics.

Of the five product-oriented centers that regulate these areas, the Center for Food Safety and Applied Nutrition (CFSAN) carries out the mission of ensuring that food products in the USA are safe, nutritious, wholesome, and properly labeled. CFSAN regulates all food products sold in interstate commerce, except meat, poultry, and processed egg products, which fall under the authority of the US Department of Agriculture. Products made and sold entirely within a single state are regulated by that state.

FDA's responsibilities cover an estimated $240 billion worth of domestic food shipments and $15 billion worth of imported foods, from the point of production or entry into the USA to the point of sale. About 46 000 food establishments are listed in the agency's 'Official Establishment Inventory' of regulated companies. These figures do not include the 600 000 or so restaurants and institutional food-service establishments, nor do they include the roughly 235 000 grocery stores and other food outlets in the USA. These are regulated by state and local authorities to whom the FDA's CFSAN provides guidance and technical assistance in the form of a Model Food Code published every other year. FDA also has regional retail specialists who provide technical assistance to state and local governments.

Since 1992, CFSAN has been organized into product-oriented offices, including the Office of Plant and Dairy Foods and Beverages, the Office of Seafood, the Office of Food Additive Safety, and the Office of Nutritional Products, Labeling and Dietary Supplements. Other offices that provide support to the product-based offices (as well as to agency stakeholders) include the Office of Field Programs, the Office of Scientific Analysis and Support, the Office of Applied Research Skills and Safety Assessment, the Office of Constituent Operations, and the Office of Management Systems. In addition to these offices involved in food regulation, CFSAN also has the agency's office of colors and cosmetics.

Two additional CFSAN component offices play crucial cross-cutting roles, including coordination of agency efforts: the Food Safety staff was formed after President Clinton's announcement in 1997 of a national initiative to enhance food safety. The FSI staff focuses on meeting the mandate of reducing foodborne disease to the greatest extent possible, and the Office of Science advises CFSAN's director and other senior managers on applying the best science available to major issues.

The nearly 900 CFSAN employees have specialized training in diverse scientific and other disciplines. Employees include chemists, microbiologists, biologists, toxicologists, food technologists, pathologists, pharmacologists, nutritionists, dietitians,

physiologists, environmental health specialists, geneticists, molecular biologists, mathematicians, sanitarians, economists, risk assessors, lawyers, and educators.

While CFSAN experts play the lead role in food safety, the premarket review of food and color additives, the chemical safety of food and dietary supplements, food labeling, and nutrition, three other FDA centers also play integral roles in food safety:

- The *Center for Veterinary Medicine* is responsible for ensuring that animal drugs and medicated feeds are safe and effective, and that people can safely consume food produced from treated animals.
- The *Office of Regulatory Affairs* (ORA) conducts food safety inspections in production plants, collects and analyzes food samples, and monitors imports for compliance with US requirements. ORA's district offices across the nation conduct 'field assignments' to assess whether imported and domestically produced food products are in compliance with FDA's regulations; remove violative products from the marketplace; provide data on food products for consideration in regulation and policy development; follow up on consumer complaints; assure that recalled or detained products are removed from distribution channels; and participate in investigations when FDA-regulated food is associated with illness.
- The National Center for Toxicological Research undertakes scientific research and testing in support of FDA's mission.

Complementing FDA's regulatory responsibilities are the food-related functions of other federal agencies – notably, the Centers for Disease Control and Prevention (CDC) and the National Institutes of Health (NIH) within HHS, the USDA's Food Safety and Inspection Service, and the US Environmental Protection Agency (EPA). FDA works closely with these and other federal agencies, such as the US Department of Commerce's National Marine Fisheries Service, the US Department of Treasury's Bureau of Alcohol, Tobacco and Firearms, the US Customs Service, the US Federal Trade Commission, the US Department of Defense, and others.

Pursuant to Presidential Order, these agencies have developed a framework to eliminate jurisdictional overlaps and fill potential gaps. A number of initiatives have integrated food-safety regulation among federal agencies and their state and local counterparts, and uniform data collection and reporting systems are being developed to enable the three levels of government in the USA to function as a national food-safety system.

While the agencies have varying approaches to food regulation under their distinct authorizing statutes, they communicate extensively with each other and abide by a series of formal agreements setting forth the agencies' respective food safety and nutrition roles.

FDA also collaborates extensively with state and local authorities, including the Association of Food and Drug Officials and state, county, and city health and agricultural departments.

Increasingly, the agency is also taking advantage of the specialized expertise that outside food-safety partners can contribute. Two academic–industry–government consortia exemplify such collaboration:

- *National Center for Food Safety and Technology.* Located near Chicago, this joint effort of FDA, the Illinois Institute of Technology, and food companies focuses on the safety of food processing and packaging technologies. Established in 1988, the center has contributed to significant improvements in the safety of sprouts and fresh juice products, for example, and to the development of alternative commercial sterilization procedures such as ultra-high-pressure processing.
- *Joint Institute for Food Safety and Applied Nutrition.* Established in 1996 with the University of Maryland, this program provides the foundation for partnerships with other academic institutions, private industry, consumer and trade groups, and international organizations. The integrated academic and regulatory science program focuses on research in the areas of food safety, risk assessment, human nutrition, and animal health and production.

FDA has also established an extramural grant program to support research in topical areas that include produce safety, egg safety, antimicrobial resistance, methods to detect foodborne viruses in foods, and research on food service, transportation and consumer practices.

By helping FDA to stay on the cutting edge in science, such collaborations ultimately support FDA in making well-founded regulatory decisions. To improve the safety of the food supply globally, FDA has undertaken cooperative research with institutions in other countries, as well.

FDA's Regulatory Role

FDA is charged with enforcing the Federal Food, Drug, and Cosmetic Act (FFDCA), which prohibits the distribution within the USA, or importation into the country, of any *adulterated* food (a food containing 'any poisonous or deleterious substance which may render it injurious to health') or *misbranded*

food (a food whose labeling 'is false or misleading in any particular').

To ensure that the food industry knows specifically what is expected of them, all of the agency's regulations are published in the US government's official daily bulletin, the *Federal Register*, pursuant to a process called 'notice and comment' rulemaking. Before finalizing certain policies, standards, and guidelines, FDA publishes a proposed approach in the *Federal Register*. The proposal contains details of the proposal and its intent, and an invitation to the public to provide written comments for agency consideration.

While 90 days is the standard period for receiving comments, the period is often extended to 180 days or longer when needed. On some issues, the agency has received as many as half a million comments. After careful analysis of all comments, appropriate suggestions are incorporated into the final version of a new regulation. The final regulation is accompanied in the *Federal Register* by a point-by-point discussion of all significant suggestions or criticisms and an explanation of the basis on which they were accepted or turned down.

This notice-and-comment rulemaking process is an essential element of FDA's fair and transparent administrative procedure, which enables the public to know how a decision was made, believe in its integrity, and have confidence that it is fair and based in science.

From the perspective of *enforcing* the FFDCA and related regulations, the agency's role can be divided into premarket and postmarket regulatory responsibilities. The premarket system focuses primarily on the safety of food additives and requires that the person or corporation requesting FDA's approval of an additive provide scientific evidence before marketing to support its safety. The requirement covers direct additives, such as a new nonnutritive sweetener, and indirect additives, such as a component of plastic bottles that can migrate into food. (Noncarcinogenic indirect food additives whose migration into food is likely to be trivial, however, are candidates for abbreviated review by FDA.)

The premarket approval requirement is waived altogether in certain cases – for example, when a substance is generally recognized by the scientific community to be safe – that is, deemed 'generally recognized as safe,' (GRAS). The procedure for a determination that a substance is GRAS has become much simpler under a recent proposed rule allowing 'voluntary notification' of GRAS status.

FDA's *postmarket* food-related activities involve primarily compliance and surveillance activities, during processing and in the marketplace. The agency's system of inspection and enforcement places priority on food at high risk of microbiological contamination (such as infant formula, ready-to-eat foods, and seafood).

Unlike in its premarket actions, in its postmarket activities, FDA bears the burden of proving that a food is adulterated (or misbranded) under the FFDCA or related regulations. The agency can collect evidence of adulteration through its inspections, laboratory analyses, and other sources of information.

If a food is determined to be violative, it can be seized or the manufacturer enjoined from producing it. Also, FDA can seek civil and criminal penalties against those who willingly introduce noncompliant foods into interstate commerce. Violative products are sometimes voluntarily recalled by the manufacturer, not by FDA, which backs the legal authority to order a product's recall.

Products imported into the USA are necessarily handled differently than domestically produced products, because it is generally not feasible for FDA inspectors to inspect the facility where an imported food was produced. In the case of imported foods, if a product appears to be adulterated or is being examined as part of a routine surveillance inspection of imported products, it can be detained or placed under bond while samples are evaluated.

If the product is determined to be adulterated under the FFDCA, it is refused entry into the country. Foreign manufacturers or importers with a history of contamination problems may be placed on 'automatic detention,' which requires the importer to provide an independent laboratory analysis showing that the product is free of any hazard. Sometimes, a product from an entire region may be considered to have an excess risk of adulteration, in which case, each lot of food seeking entry in to the USA must be tested before being admitted.

Food Regulation's Scientific Underpinnings

FDA has a longstanding commitment to supporting and participating in scientific research that can inform regulatory decision-making and contribute to a safer food supply. Agency scientists have, for example, developed a state-of-the-art polymerase chain reaction technique to allow them to detect *Cyclospora cayetanensis*, a protozan parasite on produce. The pathogen was associated with foodborne illness from imported raspberries into the USA and Canada in 1996 and 1997. Meanwhile, advances in molecular biology have also improved scientists' ability to detect foodborne viruses such as Hepatitis A and Norwalk-like viruses associated with illnesses

from produce and seafood, water, or sick food workers.

For another example, a team of scientific experts, including FDA scientists, has devised methods to reduce the microbiological hazards associated with seed sprouts, which are an increasingly popular ingredient in the American diet and, beginning in 1995, were linked with a rising incidence of foodborne disease outbreaks. Specifically, researchers evaluated a number of rapid tests that can be used to determine the safety of the irrigation water used to grow sprouts. These tests, conducted on the water after contacting the growing seeds and sprouts, can point to contaminated batches of sprouts early in the process.

Recently, another team of FDA scientists demonstrated that surface heating of fruit is an effective method to improve apple juice safety by reducing the numbers of microorganisms, including *E. coli* 0157:H7. A model was developed that showed that there is minimal penetration of heat below the surface of apples during use of the inactivation technique.

FDA's research is aimed at understanding and controlling a wide variety of potential hazards in the food supply, be they microbiological contaminants or other threats, including chemical contaminants, pesticide residues, or heavy metals. A variety of scientific tools are employed to help ensure that policy development and regulatory decisions are based on sound scientific principles.

While the agency has conducted risk and safety assessments on chemical hazards for several decades, today, FDA also conducts risk assessments to characterize the likelihood and severity of human health risks posed by microbial contaminants. Recently, FDA has undertaken quantitative microbial risk assessments to evaluate the public health impact of *Listeria monocytogenes, Vibrio parahaemolyticus*, and foodborne pathogens associated with the use of antimicrobials in food-producing animals, which can prove hazardous to people if resistant bacteria are transferred to them from the animals.

Two surveillance tools offer additional examples of how FDA takes advantage of modern scientific capabilities in fulfilling its public health mission:

- *National Antimicrobial Resistance Monitoring System*. This system, implemented by FDA, CDC, and USDA, tracks whether certain bacteria are beginning to lose susceptibility to antimicrobial drugs by acquiring antibiotic resistance genes. This objective scientific information can guide people's use of antibiotics in food animals so as to avoid potentially devastating human health consequences in the long term.
- *PulseNet*. Using high-speed internet connections and computer algorithms that can compare pulsed-field gel-electrophoresis patterns, or DNA 'fingerprints,' of foodborne bacteria, can rapidly and reliably link sporadic cases of foodborne illness to larger foodborne outbreaks. The collaborative project among FDA, CDC, and USDA was used in 1998 to support a swift response from FDA that minimized harm from a *Shigella* outbreak by quickly linking cases of infections in Minnesota, Massachusetts, California, and Canada to fresh parsley from a single farm.

Manufacturers themselves play a critical role in assuring that hazards are minimized at the site of food production. A key to achieving the safest product possible is the Hazard Analysis and Critical Control Point (HACCP) system, which identifies the critical steps in the food-production process where contamination is most likely to occur, and puts into place controls and limits to address the threat.

FDA has incorporated HACCP principles for many years in its guidance documents and regulatory programs, starting with the 1979 low-acid canned-food regulations. FDA implemented mandatory HACCP for domestic and imported seafood products in 1997 and, in January 2001, finalized the HACCP rule for fruit and vegetable juices.

FDA also has incorporated HACCP principles into its Model Food Code, a guidance document that serves as a model policy for state, territorial, and municipal and local agencies that license and inspect food service establishments, retail food stores, and food-vending operations in the USA. Recently, FDA has been working with the National Conference for Interstate Milk Shipments to develop a fluid milk HACCP pilot program.

Beyond Contaminants: Nutrition and Dietary Supplements

Besides dealing with pathogenic microorganisms and other issues of contamination or adulteration, FDA is responsible for other food and health issues such as nutrition and dietary supplements.

The system of nutritional labeling regulation dramatically changed in the USA with the passage of the Nutrition Labeling and Education Act (NLEA) of 1990, which required standardized nutrition labeling for most foods (except fresh meat and poultry) and authorized the use of nutrient content claims and FDA-approved health claims.

The FDA published final rules in 1992 and 1993, implementing the NLEA provisions. Under these regulations (and under parallel regulations from the

USDA), the food label offers complete, useful and accurate nutrition information about almost every food in the grocery store, in easy-to-read formats to help consumers make healthful food choices.

Among other features, the regulations define certain nutrient claims (such as *light* and *good source*) and list acceptable health claims about relationships between a nutrient or food and the risk of a disease or health-related condition. Some acceptable relationships are: calcium and osteoporosis, fat and cancer, saturated fat and cholesterol and coronary heart disease, sodium and hypertension, and folic acid and neural tube defects.

For dietary supplements, the framework for FDA regulation is set forth in the 1994 Dietary Supplement Health and Education Act (DSHEA). (Under DSHEA's definition, the term 'dietary supplement' includes, with some exceptions, products intended for ingestion as supplements to the diet. This includes vitamins; minerals; herbs, botanicals, and other plant-derived substances; and amino acids and concentrates, metabolites, constituents and extracts of these substances.)

Under DSHEA, FDA approval of supplement ingredients and products is not required before marketing, a feature that clearly distinguishes supplement regulation from the agency's regulation of other products such as drugs and many additives used in conventional foods. Manufacturers are responsible for ensuring that the ingredients are safe and that the ingredient list is accurate.

Dietary supplements may carry certain claims on their labels – for example, certain nutrient content claims like 'good source' or 'high' are allowed if the product meets the definition, as are FDA-approved health claims.

Under DSHEA, too, once a dietary supplement is marketed, FDA has the responsibility to prove that the supplement is unsafe before it can take action to restrict the product's use. FDA focuses its resources based on risk, so food products associated with public health emergencies are assigned the highest priority, followed by products that may have caused injury or illness. Then, products thought to be fraudulent or otherwise in violation of the law are analyzed.

Information Exchange: Stakeholder Input, Education

FDA strives to get broad input into its decisions, through public meetings, advisory committees, and other sources. By providing the opportunity for exchange of ideas and opinions between FDA and the public, including consumers, patients, health professionals, and the regulated industry, the agency is aided in reaching the best policy decisions.

Chartered advisory committees, such as the National Advisory Committee on Microbiological Criteria for Foods, the FDA Science Board, and CFSAN's Food Advisory Committee, play an important role in providing scientific advice. Internationally, FDA participates in Codex Alimentarius, in expert consultations, and on committees of the World Health Organization and Food and Agricultural Organization, to name a few.

In addition to the agency's notice-and-comment rulemaking process, supplementary outreach efforts are sometimes necessary to achieve optimal industry compliance. The following examples illustrate the types of outreach the agency engages in to inform audiences.

FDA's response to several sprout-associated foodborne illness outbreaks can serve as an example of how the agency reaches out to educate different groups on their respective food safety roles. Following the outbreaks in the late 1990s, FDA issued two guidance documents advising sprout growers and seed suppliers of steps to reduce microbial contamination: *Reducing Microbial Food Safety Hazards for Sprouted Seeds and Sampling* and *Microbial Testing of Spent Irrigation Water During Sprout Production*. In addition, FDA and the California Department of Health Services produced an educational video on good agricultural and manufacturing practices for sprout producers. Also, to educate the general public on ways to protect themselves from falling ill from sprouts, FDA issued advisories directly to consumers on the potential hazards associated with eating raw sprouts.

To disseminate food-safety messages effectively to various audiences, including consumers and food producers and preparers, FDA participates in the Partnership for Food Safety Education, which includes representatives from several government agencies as well as from industry and consumer organizations.

For several years, the Partnership has been undertaking a *Fight Bac* (Fight Bacteria) campaign to educate consumers on four simple principles of preventing food contamination while preparing food at home: clean, chill, cook, and combat cross-contamination. Major corporations have included the 'Bac' character and the food-safety messages in their national consumer-education initiatives.

To improve outbreak detection and response when outbreaks *do* occur, FDA has developed training courses. For example, FDA's satellite courses on food microbiology, foodborne disease epidemiology, and traceback in outbreak investigations have been

attended by thousands of government and industry representatives from around the country. FDA has also conducted presentations at numerous conferences in the USA, Mexico, and Latin America on how to conduct tracebacks to determine the source of an outbreak.

But these examples are a selection of food-safety efforts undertaken by FDA and its food-safety partners. Perhaps the most comprehensive information on food-safety efforts in the USA and FDA's food science and nutrition roles is available on the internet – at FDA's website, www.fda.gov, and the joint federal food-safety website, www.foodsafety.gov.

Efforts Paying Off

Educational efforts and other food-safety programs have reaped dramatic rewards in terms of improved food-safety practices, as evidenced by recent CDC and FDA data. CDC data suggest that there was a 21% decline in foodborne illness in the USA from the year 1996 to 1999. This trend continued in 2000 and 2001.

According to two recent phone-tracking surveys, the incidence of the risky practice of eating raw foods was significantly reduced in 1998 from 1993 numbers – down 33% for pink hamburger, 39% for raw oysters and clams, and 29% for raw egg-containing foods. The surveys showed similarly encouraging improvements in consumer behavior in terms of washing hands and cutting boards to prevent cross-contamination between raw and other food.

Despite the apparent effectiveness of steps to reduce foodborne illness in the USA, food-related hazards may never be eliminated altogether, and so far remain a significant threat to the public health. In an effort to extend the downward trend in foodborne illness, FDA must keep up with changes in the science and technology used in agency-regulated products, in the ways in which consumers obtain information about products, in consumer demographics and consumption habits, and in the composition of global trade and production.

By remaining on the cutting edge of science and by staying true to the tradition of science-based regulatory decision-making initiated by Wiley a century ago, FDA can ensure that its regulatory decisions are made in the best interests of the public health, while at the same time not unduly stifling industry innovation.

See also: **Hazard Analysis Critical Control Point**; **Legislation**: History

Foodborne Diseases *See* **Food Poisoning**: Classification; Tracing Origins and Testing; Statistics; Economic Implications; **Microbiology**: Classification of Microorganisms; Detection of Foodborne Pathogens and their Toxins; **Diarrheal (Diarrhoeal) Diseases**; **Colon**: Diseases and Disorders; **Celiac (Coeliac) Disease**; ***Clostridium***: Occurrence of *Clostridium perfringens*; Detection of *Clostridium perfringens*; Food Poisoning by *Clostridium perfringens*; Occurrence of *Clostridium botulinum*; Botulism; ***Escherichia coli***: Occurrence; Detection; Food Poisoning; Occurrence and Epidemiology of Species other than *Escherichia coli*; Food Poisoning by Species other than *Escherichia coli*

FOOD COMPOSITION TABLES

D A T Southgate, Norwich, Norfolk, UK

Introduction

Knowledge of the composition of foods is essential for the dietary management of patients and for most quantitative research on human nutrition. This information has traditionally been provided by tables of food composition and over the last 20 or so years, increasingly by computerized nutritional databases. The principal features of food composition tables and nutritional databases are very similar, and in this article, compilations will be used to describe both printed tables and computerized databases, and only when the comments relate to one of the types specifically will they be distinguished.

These compilations of the nutritional composition of foods are principally used for estimating the nutritional composition of diets from measurements of food consumption, constructing diets in clinical nutrition, and designing rations and menus. Data on the composition of foods from these compilations are also widely used in nutritional surveillance and in assessments of the adequacy of food supplies and as authoritative sources for nutritional labeling. Compositional data are also widely used in nutrition education and the development of nutritional guidance for public health guidance.

General Description

Food compositional tables and nutritional databases share many common features. Initially, the databases were prepared from the printed tables, but now, the compilation processes are based on computerized systems that form the starting point for preparing the printed tables, often in different formats for different types of use.

The basic format is a matrix of food items against nutrient data. The formats of printed tables are constrained by the dimensions of the printed page and the size of the tables as a whole. Two formats are used, one where the values for a number of foods and nutrients are given on a single page (or a sequence of pages) and a second, where the values for one food are given per page. Databases are most frequently presented as a food/nutrient matrix, the dimensions of which are constrained only by the database management system being used.

Introductory Text

Food composition tables have an introductory text. In databases, this forms part of the documentation of the database and may be available on line or provided as hard copy. These provide the user with the rationale adopted by the compilers in selecting the food items that are included and the range of nutrients and other constituents for which values are given. The definitions of the constituents included describe the modes of expression adopted and the analytical methods that were used to obtain the values. In most cases, the values will be derived by a combination of analytical values measured specifically for the compilation and others obtained from the literature, including other databases, and the text will document the sources.

Coverage of Foods

Most human diets include a very wide range of foods, especially cooked dishes made up from a number of ingredients. Most compilers of nutritional data work within limited budgets, and therefore, the number of food items in all compilations is restricted. The primary objective of all compilers is to present compositional values that are representative for the food items consumed by the population for which the compilation is prepared.

In a national database, priority is given to those foods that form the major part of the food intake of the population. These have been identified using national data on food consumption and a range of other information from market surveys, information on food purchasing patterns, and data from the food industry. The foods consumed by specific groups such as infant formulas, which provide all the dietary intake of young infants, are also included. The compilations often include the foods eaten by ethnic groups within the population and those with different dietary patterns. Many compilations also include food items whose composition can be seen as examples of different types of food. Most national compilations do not provide information on the composition of specific brands of foods, unless the brand name is critical for identifying the food or, more rarely, the brand is a very important brand leader.

Food Groups

The foods are usually arranged in food groups that bring together foods related to one another. Subgroups are often used within these major groupings, but this varies from compilation to compilation. Table 1 gives two widely used groupings. The groups used by the Food and Agricultural Organization (FAO) of the UN are linked to those use in classifying foods in commerce and are widely used in tables prepared for regional use.

Food Identification and Nomenclature

Most compilations assign a code number to every food item. These numbers are widely used when calculations of nutrient intakes are being made when the numbers are added to the food intake records to facilitate entry into the database calculations. Systems of numbering have been developed for assigning numbers that classify and identify the food for international use. These coding systems are used in some national compilations, but not universally.

Food names These are given in a formalized way and include descriptors as facets, which further serve to identify the foods. Unequivocal identification of a food item and relating it with its composition are very important for the users of nutritional databases. These naming facets include details of the way in which a foodstuff has been cooked or prepared for consumption. Even within one country, the naming of

Table 1 Examples of major food groups in nutritional compilations

Groups used in FAO tables	*Groups used in UK tables*
Cereals and grain products	Cereals and cereal products
Starchy roots, tubers, and fruits	(Included within vegetables)
Grain legumes and legume products	(Included within vegetables)
Nuts and seeds	Nuts
Vegetables and vegetable products	Vegetables
Fruits	Fruits
Sugars and syrups	Sugars, preserves, and confectionery
Meat, poultry, and game	Meat and meat products
Eggs	Eggs
Fish and shellfish	Fish and fish products
Milk and milk products	Milk and milk products
Oils and fats	Oils and fats
Beverages	Beverages
	Alcoholic beverages
Miscellaneous	Miscellaneous
The FAO database is most widely used for measurements of food purchases and includes mainly raw foods. The groupings are linked to classes of foods used to categorize economic aspects of food production and consumption	The UK database is widely used for studies of individual food intakes and therefore includes many cooked dishes within the food group that appears most appropriate

foods is not uniform, and the same food may be known by different names and vice versa. Most compilations include a thesaurus giving these alternative names and frequently include a list of the scientific (taxonomic) names for those foods for which this is appropriate.

Descriptions of samples analyzed These give the sources and numbers of samples on which the data are based. Most compilations give the nutrient values on the basis of the edible portion, and these descriptions include details of how the edible material was prepared and the nature of the inedible matter removed. Many compilations include a factor to convert the values on the edible portion back to a value on a raw or 'as purchased' basis. This information is used to give an indication of the reliability and quality of the nutritional compositional values.

Coverage of Constituents

Most compilations include a reasonably comprehensive range of nutrients, but the growth of nutritional research interests in the role of other biologically active constituents is increasing pressures for extending the number of constituents to include these 'nonnutrients.' The coverage is usually determined by the most demanding of the users, which tend to be nutritional epidemiologists and nutritional researchers. The typical coverage in a nutritional compilation is given in Table 2.

The inclusion of a constituent is conditional on there being a satisfactory method for analyzing the nutrient and the existence of a sufficient body of information on the amounts of the nutrient in a range of foods to assess its inherent variability. This latter factor will determine whether or not it is possible to provide representative data for the nutrient in a range of foods. Some data for nutrients (or other constituents) may only be available for a limited number of foods, and in some cases, the distribution of a nutrient is limited, and many compilations include subsidiary tables giving this more limited coverage.

Expression of Nutrient Values

The nutrient values are usually given per 100 g of edible matter, and the international preference is for the nutrient to be given in weight terms. A number of values given in the compilations are derived values.

Derived values Energy values are for metabolizable energy (energy that can be used by the body) and are calculated from the protein, fat, carbohydrates, and alcohol in foods using energy conversion factors. The most commonly used are those of Atwater, in which protein is assigned, $4\,\mathrm{kcal\,g^{-1}}$ ($16\,\mathrm{kJ\,g^{-1}}$), fat, $9\,\mathrm{kcal\,g^{-1}}$ ($37\,\mathrm{kJ\,g^{-1}}$), carbohydrates (total by difference), $4\,\mathrm{kcal\,g^{-1}}$ ($16\,\mathrm{kJ\,g^{-1}}$) and alcohol, $7\,\mathrm{kcal\,g^{-1}}$ ($29\,\mathrm{kJ\,g^{-1}}$).

In the UK and some other compilations, a modified system is used where carbohydrates (as monosaccharides) are assigned $3.75\,\mathrm{kcal\,g^{-1}}$ ($15\,\mathrm{kJ\,g^{-1}}$). Some compilations, the US for example, use 'Specific Factors' that are designed to allow for the differences in the heats of combustion of different food components and differing digestibilities.

Most compilations discount the small amounts of metabolizable energy derived from dietary fiber from the short-chain fatty acids produced by its

Table 2 Coverage of nutrients in nutritional compilations

Constituent	*Mode of expression*	*Coverage*	*Comments*
Water		Most	
Total nitrogen		Many	
Protein	Calculated from total nitrogen	All	Some use common 6.25 factor; others use specific factors for different food groups
Amino acids	mg per amino acid per gram of N or mg per 100 g of food	Usually in subsidiary tables	
Fat, total		All	
Fatty acids	Fatty acid % total fatty acids or g per 100 g of food	Usually in subsidiary tables	Compilations may include amounts of saturated, monounsaturated and polyunsaturated acids
Carbohydrates total	Often 'by difference' or by summation of sugars and starches	All	Summation is the preferred approach
Sugars	May be expressed as the monosaccharides	Total values in many	Individual sugars given in a few databases
Starches	May be expressed as glucose or as the polymer	Many	
Nonstarch polysaccharides (NSP)	Usually expressed as polymers	Limited	
Dietary fiber	NSP in UK; total dietary fiber values in US and many other compilations	Most	Data rather limited; some older compilations give 'crude fiber' values
Alcohol	g per 100 g or per 100 ml		May be restricted to relevant foods
Energy value	kcal or kJ, calculated from protein, fat, carbohydrate, and alcohol		Choice of conversion factors differs between databases
Inorganic nutrients			
Na, K, Ca, Mg, P, Cl	Usually mg per 100 g	Most	
Cu, Fe, Zn		Most	
F, I, Se, etc.	μg per 100 g of food	Limited	
Vitamins			
Fat-soluble			
A	Retinol μg per 100 g	Most	
D	Cholecalciferol μg per 100 g		
E	α-Tocopherol mg per 100 g		
Carotenes	μg per 100 g β-carotene or retinol equivalents	Most	Few give values for different carotenoids
Vitamin K	Phylloquinone	Limited	
Water-soluble			
Vitamin C	Usually as mg per 100 g	Most	
Thiamin			
Riboflavin			
Niacin			
Vitamin B_6		Limited	Few give values for vitamers
Folates	μg per 100 g	Many	Mostly total values
Vitamin B_{12}	μg per 100 g of food	Most	
Pantothenic acid	Usually mg per 100 g of food	Limited	
Biotin			
Other constituents			
Cholesterol	mg per 100 g of food	Many	
Phytic acid	mg per 100 g of food	Limited usually in supplementary tables	
Purines	mg per 100 g of food	Very few	

fermentation in the large intestine. Similarly, the energy contributed by organic acids is discounted.

Protein values are calculated from the total nitrogen values using conversion factors, which allow for the different amounts of nitrogen in proteins from different sources.

These conversion systems are described in the introductory documentation.

Modes of expression of vitamins Most vitamins are expressed on a weight basis, except where there are several vitamers with differing activities when many

compilations use a mode of expression designed to give the vitamin equivalence.

Vitamin A activity is contributed by retional (preformed vitamin A) and the provitamin A carotenoids. The values may be given as retinol equivalents (μg RE); β-carotene is usually divided by 6, and other carotenoids are usually divided by 12 to give retinol equivalents. These conversion factors are currently the subject of scientific debate at the present time, and expression of the carotenoids on a weight basis is preferred.

Some compilations give vitamin E values as milligrams of α-tocopherol equivalents calculated from the biological equivalence of the different tocopherols and tocotrienols.

Many compilations include a value for niacin equivalents, where the tryptophan content (mg) is divided by 60 and added to the niacin value.

Amino acids and fatty acids Amino acid composition may be expressed in one of two ways: either as milligrams of amino acid per gram of total nitrogen or milligrams of amino acid per 100 g of food. Fatty acids may be expressed as fatty acid percent of total fatty acids or milligrams of fatty acid per 100 g of food.

Compilation of Nutritional Data

As mention before, the compilation of food composition tables and nutritional databases follow similar principles. The compilers are faced with three major decisions: the choice of food items to include, the choice of constituents, and how to source the data.

Because the compilations are essential tools to be used, these decisions are taken in consultation with the range of users of the nutritional data. Analogously, decisions about the format of both tables and databases are based on how the data will be used and software developed to meet the users' requirements. This is usually done by forming a steering group to guide the compilers in their work. This group meets regularly during the progress of the work to monitor progress.

Choice of Food Items

Decisions in this area are the most difficult for the compiler because of the very wide range of foods eaten in the human diet. Users want as complete coverage as possible, and many would like the inclusion of branded foods. Complete coverage of the range of foods on sale in a developed country would require a database with tens of thousands of food items. Furthermore, complete coverage of all the foods consumed, which in the main consist of complex, mixed, cooked dishes, expands this number possibly by two orders of magnitude. All compilers therefore have to consider the priorities for inclusion of food items very critically. The priorities for inclusion are to give coverage of the most important foods for the population for which the compilation is intended and the inclusion of foods important in the diets of specific groups such as infants, ethnic groups, and other groups whose dietary patterns differ from that of the population as an whole.

The compilers have to develop the systems of nomenclature used and decide on the descriptions of the food items in consultation with the users.

Selection of Nutrients

This is usually a slightly easier task because the numbers of nutrients are finite. The number of biologically activity constituents in plants, where there is growing research interest, is extremely high. The compilers must be satisfied that reliable analytical methods exist, and the variation in composition of a single food is not too great to prejudice the provision of representative data.

Sources of Compositional Data

Compilers have two major sources of data. First, data that are obtained specifically for the inclusion in the compilation, which has been called the 'direct method.' This involves the development of sampling and analytical programs for obtaining the data. This approach has the advantage in that the compilers have control over the sampling protocols to insure the identity of the food and to ensure that the sample is representative. They can also control the choice of analytical methods and the quality-assurance schemes in operation in the analytical laboratory. The compilers are thus completely familiar with the provenance and quality of the data. This approach is, however, rather expensive. It can also be argued that this approach discounts the large amount of information available on the composition of foods in the scientific and other literature.

Second, the 'Indirect method.' In this, the literature on the composition of foods is searched for suitable data. This approach involves the compilers in making judgements about the applicability of the samples analyzed, their identity, and whether or not the sampling protocol ensured that a representative sample was analyzed. Judgements also need to be made about the choice of analytical methods and the control of their execution. The modes of expression of data may also need to be converted into that which will be used in the database. This approach requires the compilers to make judgements about the data they select from

the literature, and the assessment process is often demanding and time-consuming. In practice, most national compilations are based on a combination of the two approaches with literature searching typically forming the first stage in the data-compilation process. This is followed by the design of sampling and analytical protocols whenever the literature data are judged to be inappropriate for use, or the search shows that few or no data exist for a food item, which for other reasons, is required in the database.

Using Food-composition Data

The uses of databases can be considered to be of three major types: nutritional analyses, where the database is used to calculate nutrient intakes from records of food consumption; nutritional synthesis, where the database is used in calculations to generate diets with specific nutrient contents; and as a primary source of compositional information.

Nutritional Analyses

In these, the calculations may be based on food consumption measured at many different levels. These levels have important implications for the compilers because they set the requirements for compositional information appropriate to the level of measurement. These are summarized in Table 3.

Food-supplies level Governments and international agencies use measurements at this level for assessing the nutritional adequacy of food supplies. The food-balance sheets of the Food and Agricultural Organization of the United Nations are one important example of this level of use, which uses a special compilation that gives the composition of raw food at the wholesale level; e.g., the composition of bulk cereals and animal carcasses. Many national compilations also include values for this type of data.

Food-purchases level These measurements here are of two types; household budget studies, usually undertaken primarily for economic reasons where food consumption is derived from records of household purchases in monetary terms, and Food Surveys, where the amounts of food purchased by a household are measured, of which the UK National Food Survey is an example. Both these types of use require data for the composition of foods 'as purchased.' Household budgetary surveys also require information on retail prices to convert monetary expenditure into quantities of food.

Individual food-intake level These use records of food consumption by individuals either in a weighed intake study or, retrospectively, where the amounts of food consumed are obtained by dietary recall or a food-frequency questionnaire. These types of study provide the most detailed records of foods consumed and have the most demanding requirements for the inclusion of food items in a database because the records include specific branded foods and, more importantly, cooked mixed dishes, that is foods 'as consumed.'

Nutritional Synthesis

The compilations are widely used by dieticians and other health care professionals to calculate diets that will provide specific amounts of nutrients, for example, low-energy diets for the obese, controlled energy, carbohydrate and fats for the diabetic patient, and low-sodium diets in hypertension. This type of calculation requires the database to contain the range of food likely to be eaten by the patients in question, but the quantitative requirements may not be very demanding. Where, for example, in metabolic studies, diets with very precise composition are required, analyses of the foods are necessary.

At the present time, most of these diets are formulated by manual calculation from the database, but nutritional syntheses programs, which include individual food preferences, are becoming available to computerize the calculation.

Table 3 Requirements for food items and nutrients at different levels of data usage

Level of use	*Food items*	*Nutrients*	*General features of compilations*
Food supplies	Foods measured at raw, bulk or commodity level	Usually restricted to macronutrients and energy	Relatively small database is required
Food purchases			
Household budget surveys	Foods as purchased, retail raw, prices	Macronutrients, major inorganic nutrients and vitamins	Information needed on losses of food during preparation
Food-consumption surveys	Foods as purchased retail or obtained raw	Macronutrients, major micronutrients	Information needed on losses of food during preparation
Individual food intakes	Foods as consumed, cooked including mixed dishes	Comprehensive coverage of macro- and micronutrients	Requirements may extend to very detailed coverage of constituents

Sources of Compositional Information

In addition to their use in providing compositional data for guiding food choices for health reasons, nutritional databases are now widely used in nutritional labeling, where it forms part of food regulation. In most countries, where nutritional labeling is required (or permitted), the data taken from an authoritative nutritional database are permitted. The food manufacturer or distributor is thus relieved from carrying out analyses on their products. The regulations have the requirement that the item in the database must correspond to the food being labeled; generic labeling of foods such as meats, fruits, and vegetables is often permitted.

In the USA, the nutritional labeling regulations require manufacturers to carry out their own analyses, using specified methods defined by the Association of Official Analytical Chemists, and also specifies that the sample shall be representative of production within certain limits. This has provided the US National Database with a great deal of manufacturers' data on branded foods. The UK database also uses some manufacturers' data where the requirements about choice of analytical method and that the sampling was representative can be met.

Accuracy in Use

Judgements in this area depend on the use that is being made of the compilation.

Many users of nutritional databases require a semiquantitative guide in order to give general dietetic advice, but some dietetic uses are more critical, for example, in formulating diets for diabetics and low-sodium diets in hypertension; even so, the accuracy required is not as demanding as for those who conduct nutritional analyses with the database in which nutrient intakes are calculated from records of food intakes.

Relatively few large formal studies have been undertaken, but these show that the calculated values for energy, protein, fat, and available carbohydrates (sugars and starch), are usually within a few percent of the analyzed values. The calculated values for potassium, calcium, phosphorus, and magnesium show a similar agreement. Analytical values for sodium are often in excess of calculated estimates because of the addition of salt in cooking, and analytical values for iron are similarly slightly higher, possibly because of iron contamination. Those nutrients for which natural variability is high tend to be the more difficult to predict with accuracy. Thus, the micronutrients are usually predicted with a lower accuracy, say around 10%, except for vitamin C and folates, where the calculated values may be substantial overestimates.

It is important in comparisons of calculated values with analyzed values that the amounts of foods making up the analyzed diets be precisely the same amounts as those used in the calculations; where collections are not supervised closely, the tendency is to undercollect.

Where a compilation is being used as a source of compositional data on a specific food, the accuracy depends on two major factors, first, how closely the food item in the compilation corresponds to the food in question, and second, the natural variability of the food. All users of nutritional data on foods must take into account that foods are biological materials and show natural variations in composition. This is true, even for processed foods that are produced under closely controlled conditions.

Limitations

Many of the limitations have been mentioned briefly in the earlier sections. They can be considered under three headings.

Coverage of Foods

As is clear from the section on Compilation of Nutritional Data, no existing databases are truly comprehensive in the sense of covering all the foods likely to form part of the human diet. The analytical resources required to construct such a comprehensive database are very great indeed, and coverage of all cooked mixed dishes is virtually impossible. Additionally, the production of new food products and the reformulating of existing products are a continuous activity of the food industry. This means that complete coverage of all food items is a practically impossible aim. Users of nutritional compilations therefore have to evolve strategies for dealing with missing foods.

In most cases, choosing a food that is biologically related or similar to the missing food is satisfactory. In the case of cooked dishes, calculation from the recipe using data on the loss of weight on cooking provides an acceptable strategy, but the estimation of vitamin losses in cooking means that the calculated values must be regarded as estimates.

Coverage of Nutrients

Most compilations provide coverage of all the macronutrients and most of the vitamins and major inorganic nutrients. The principal gaps at present are the trace elements for which, despite their nutritional desirability, it is difficult for the compilers to provide representative values for the reasons discussed earlier.

The coverage of the carotenoids in most databases is limited at present, but in the future, one should expect better coverage. The same will be possible for the different folate vitamins, although the analytical methods are possibly less well developed. Many databases give limited coverage of the carbohydrates in foods. Coverage of the nonnutrient biologically active components represents a greater problem because of the great number of potential constituents coupled with the variety of methods, but nutritional research interests in these components should eventually identify the most important constituents to include.

Data Quality

Quality implies fitness for its purpose, so the uses made of databases need to be considered in any discussion of the limitations on the quality of the data in databases.

Most databases give only mean values, and only a few include limited estimates of variation. The accuracy of databases will always be limited by the natural variability in the composition of foods, and this places a limit on the predictive accuracy of any database both for predicting the composition of any one sample of a food and in the calculations made from the database. These limitations decline as the number of food records used in the calculations increases so that, in effect, the foods eaten approach a representative sample size.

The major limitation in producing databases is the documentation of sampling and the analytical methods used, and particularly the use of quality-assurance schemes in analytical laboratories. In many countries, the need for quality assurance schemes in analytical laboratories producing data on food composition, particularly for use in databases, is becoming accepted. Research is needed to develop objective quality-assurance schemes for sampling protocols to ensure that this important aspect of data quality is covered.

However, the quality of the data in most nutritional databases at present is considerably superior to the quality of food-intake data, and this is the major limitation in using nutritional databases to assess nutrient intakes.

Compatibility of Databases

The growth in the study of nutritional epidemiology at the international level has focused special attention on the compatibility of national databases, and in this, data quality is an essential aspect. A number of international initiatives have been undertaken to address these issues of these INFOODS (International Network of Food Data Systems) and FAO have been very prominent, and a number of European activities have been undertaken, including Eurofoods and Cost 99.

See also: **Bioavailability of Nutrients**; **Carbohydrates**: Classification and Properties; **Carotenoids**: Occurrence, Properties, and Determination; **Cereals**: Contribution to the Diet; **Dairy Products – Nutritional Contribution**; **Dietary Fiber**: Determination; **Eggs**: Structure and Composition; **Fermented Milks**: Dietary Importance; **Fruits of Temperate Climates**: Commercial and Dietary Importance; **Meat**: Nutritional Value; **Milk**: Dietary Importance; **Nutritional Surveillance**: In Industrialized Countries; **Protein**: Food Sources; **Plant Antinutritional Factors**: Characteristics

Further Reading

Cashel K, English R and Lewis J (eds) (1989) *Composition of Australian Foods*. Canberra: Australian Government Publication Services.

Cura A, Salvini S, Parpine M *et al.* (1998) *Banca dati composizione degli alimenti per studi epidemiologici in Italia*. Milan: Tipografia F.illib Verderio snc.

Greenfield H and Southgate DAT (1994) *Food Composition Data; Production, Management and Use*, revised edn. London: Chapman & Hall.

Holland B, Welch AA, Unwin ID *et al.* (1991) *McCance and Widdowson's The Composition of Foods*, 5th edn. Cambridge: Royal Society of Chemistry and Ministry of Agriculture Fisheries and Food.

Klensin JC, Feskanich D, Lin V, Truswell AS and Southgate DAT (1989) *Identification of Food Components for INFOODS Data Interchange*. Tokyo: United Nations University.

Loughridge JM, Walker AD and Towler G (1993) *Inventory of Nutritional Software. FLAIR Concerted Action Programme. No. 12*. Wageningen, The Netherlands: Department of Human Nutrition.

McCance and Widdowson's *The Composition of Foods* (2002), 6th edn. Cambridge: Food Safety Agency (UK) Royal Society of Chemistry.

Siminopoulos AP and Butrum RR (eds) (1992) *International Food Data Bases and Information Exchange. Concepts Principles and Designs. World Review of Nutrition and Dietetics*, vol. 68. Basel: Karger.

Sullivan DM and Carpenter DE (eds) (1993) *Methods of Analysis for Nutrition Labeling*. Arlington, VA: Association of Official Analytical Chemists, International.

Trichopoulou A (ed.) (1992) Methodology and public health aspects of dietary surveillance in Europe: the use of household budget surveys. *European Journal of Clinical Nutrition* 46, supplement 5.

Useful Web Sites www.nal.usda.gov/fnic/foodcomp This is the only format (and a corresponding CD-ROM) in which the USDA data are published. www.fao.org/infoods.

FOOD ETHICS

R Early, Harper Adams University College, Newport, UK

Background

Food is necessary to the life of humans, and all human-beings must either work to provide the food they require to live or obtain their food from another person (or organization) who may specialize in the provision of food. When a man[1] works to provide food for himself and his family, independently of others, few ethical issues are found in the way in which he produces food. The way in which animals are kept and slaughtered for food, and the way in which the land and other natural resources are utilized for food production, and maintained for the benefit of future generations, may be the only sources of moral and ethical concern. The situation changes when a man has to obtain food from another person, or an organization established to provide food. In this circumstance, he places himself at risk of exploitation, abuse, and inadvertent and intentional unethical treatment. Examples of food producers intentionally acting against the interests of consumers are found in historical references. Accum and Hassall recorded a variety of unethical food industry practices occuring in nineteenth century Britain, such as the adulteration of tea and coffee with charcoal and the use of red lead to color cheese, and Sinclair chronicled the sorry practices of Chicago's meat-processing industry at the end of the nineteenth century. Both Hassall and Sinclair catalyzed the development of food laws in their respective countries.

Consumers today may believe that the practices of unethical food producers of over a century ago are features of a distant past, but this is not so. The motivation of profit can tempt food businesses to place their interests above those of consumers, with the risk of causing them various forms of harm through immoral and unethical conduct. For instance, false declarations made by food manufacturers or retailers about food products, or the withholding of facts, concern the moral issue of truth-telling. Any intention to mislead consumers amounts to the intention to harm them, and, in this instance, harm may amount to the purposeful erosion of consumers' autonomy. A consumer's consent to enter into a form of contract with a food manufacturer or retailer – as signified by the consumer purchasing a product offered for sale – can be valid only when the consent is competent, informed, and freely given. False declarations and the withholding of facts constitute lying and reduce the consumer's ability to make purchasing decisions that are both competent and informed and, therefore, freely given. Although such conduct on the part of food businesses does not cause physical harm, instances do occur of food businesses knowingly placing consumers at risk of physical harm. During 1999 and 2000 in the UK, cases came to light of food processors 'cleaning up' chicken meat condemned as unfit for human consumption and selling it into food processing and the retail food marketplace. Not surprisingly, this practice contravened the UK's food safety law. Food laws establish boundaries for the conduct of food businesses, and crossing the boundaries entails the risks of social proscription, prosecution and, penalty. In some ways, laws may be considered to define what is ethically acceptable and, certainly, laws may express requirements that relate directly to moral principles. But though laws state requirements for, and set limits on, certain kinds of conduct, what is considered to be morally and ethically defensible ought not to be defined entirely by the law. The process of legislation is subject to bad judgement, political bias, and lobbying by parties intent on protecting their own interests. Laws can be wrong and fail to protect those who ought to be protected. The moral values exercised by food businesses ought to ensure that their conduct reflects what is right, and not just what the law requires.

Generally, consumers can feel secure in the protection afforded by food law, but such are the pressures of the commercial world that the temptation exists constantly for food businesses to engage in behavior that, though within the boundaries set by law, contests the tenets of common morality. The moral character and the conduct of those who challenge moral values and norms, and who may work to stretch the limits of the law without breaking them, will, inevitably, attract moral scrutiny. Those who determinedly break the law and whose conduct is clearly morally indefensible choose to step outside what is acceptable to society, and they are readily recognized as a threat to society. In comparison, those who maintain a sense of moral respectability in order to remain part of society, but whose conduct erodes generally accepted moral values, might be considered a greater threat to the stability of society. Such agents ought to

[1]Use of the term 'man' indicates no gender bias. It is used purely for convenience.

attract greater moral scrutiny than those who are purely dishonest, because their actions may have more far-reaching effects. The modern food industry has created much that is good, and Western consumers today routinely have access to food of a variety, quantity, quality, and safety, unparalleled at any time in the past. By its very nature, the food industry is an agent of social change. It takes and creates opportunities to influence people, and by its conduct, it risks, and may intend, catalyzing social change: consequently, it exists within society as a kind of moral agent. As part of society, the food industry works to maintain respectability within society, but those businesses that constitute the industry may, at times, be tempted by forms of conduct that challenge moral acceptability. Because consumers are becoming increasingly dependent on the industry for their food, and because the industry exercises increasing levels of influence over society, the justification exists to scrutinize the moral values by which the industry operates and the ethical nature of its conduct.

The expression of moral and ethical concern over matters relating to food is not new. Though the term 'food ethics' may have been formally established relatively recently, Zwart suggested that food ethics has a long history. He argues that, while modern food ethics is concerned with food production, historically, it was centered on consumption. Certainly, this resonates with ancient religious moral conventions that prohibit the consumption of certain types of food materials as well as the practices of vegetarians and vegans, whose food choices are often defined by their particular moral values. Today, food ethics is a branch of moral philosophy that encompasses the relationship of people, and society, with food and the food supply system. In wealthy, Western societies, food ethics is concerned mainly with the conduct of the food businesses that make up the food supply system and the methods by which food is produced for, and marketed to, consumers. With the development of the consumer society, *people* have been transformed into *consumers*, whose function is to consume, in accordance with the policies of governments and the strategic plans of businesses. The degree to which people in Western societies have come to rely on food businesses for their food – and survival – has caused, and legitimizes, the development of food ethics as a field of intellectual inquiry. Modern food ethics is a response to the industrialization and commercialization of the food-supply system. Factors such as the increasing utilization of science and technology in food manufacturing and supply, the methods used in food marketing and advertising, and the evolving dynamics a global food marketplace create and energize the food ethics debate. But food ethics does not respond solely to food issues in the developed, Western world. The problems of global food security, and food shortage and malnutrition in developing countries, are primary issues of food ethics, as recognized by the Food and Agriculture Organization (FAO) of the United Nations. The FAO also recognizes the use of biotechnology and genetically modified organisms in the food-supply system as issues of food and agricultural ethics. Food ethics has emerged to guide and inform those who work within the food-supply system, and those with a professional interest in the activities and conduct of the system and the organizations that constitute it.

Morality and Ethics

As human beings, we generally recognize certain moral values and understand that moral principles derived from moral values can act to guide and stabilize the society within which we live, and inform both social and cultural development. The word 'moral' originates from the Latin, *moralis*, while 'ethics' is derived from the Greek, *ethos*. Both words equate to 'custom' and are sometimes used interchangeably, though different meanings can be assigned in relation to the way in which we think about morality and ethical behavior. Often, we consider morality in terms of values, or standards of conduct, that prescribe concepts of what is right and wrong. In contrast, we judge what is ethical in terms of the analysis and assessment of morally acceptable behavior. For instance, an evaluation of the ethical nature of a food business' conduct would lead to a review of the business' moral values as the justification for the conduct. Similarly, though rich nations may recognize their moral duty to help feed poor nations, translation into practice of the moral values that give rise to the notion of duty relies on ethical conduct that ensures the poor are fed. The ethical assessment of conduct enables us to see that moral values and principles exist and have been upheld. Ethics is concerned with the application of moral values and principles and, specifically, with concepts of right and wrong, good and bad, duty, obligation, virtue, freedom, and choice.

People interact with each other at a personal level, and their moral values influence the ethical nature of their conduct. Of particular interest to moral philosophers, since the time of Socrates (*c.* 470–399 BC), has been the question 'How ought we to live?' The question embodies the concern that, for each of us, our lives should be lived as good lives and that the influence we have on other people's lives will then only be good. At its most basic level, ethics is

concerned with our conduct as individuals in relation to other people and, increasingly, in relation to other species and the environment generally. However, the infrastructure of modern society is complex, and though, as private individuals, we may exert some influence on society, the greatest source of influence is often the organizations that operate within society. Yet, organizations are made up by people, but when the conduct of an organization is assessed in terms of its ethical nature, paradoxically, the moral values from which the conduct is derived are judged to belong to the organization, and not the individuals who make up the organization. Any examination of the moral values and ethical conduct of the food industry must concern the businesses that constitute the industry and the executives who lead the businesses. Conduct that is perceived to reflect the moral values of a food business is most likely to be based on the moral values of the executives in charge of the business and, specifically, the chief executive officer.

Ethics is delineated as *metaethics, descriptive ethics*, and *applied ethics*. Metaethics is a form of philosophical inquiry concerned with concepts, beliefs, and meaning. Metaethics seeks to explain what is meant by concepts such as good, bad, right, and wrong in relation to the character or conduct of the moral agent. It is a process of rational inquiry separate from the 'real-world' domain of applied ethics, but has significant bearing on the practice of applied ethics. Descriptive ethics, as the name suggests, is concerned with the study of the belief systems, moral values, and ethical behavior of social groups and organizations. It is observational in method and somewhat disconnected from its subject matter, and so it compares with sociology and anthropology. Applied ethics contrasts with both metaethics and descriptive ethics in that it is concerned with the translation of moral values into ethical practices and the application of ethics to practical problem-solving and decision-making. Food ethics falls within the realm of applied ethics, because it is concerned with the 'real-world' ethics in relation to the operation of the food supply system and the conduct of the food industry.

Many ethical issues exist in relation to the food-supply system and the activities of the food industry. Food businesses must deal with ethical issues in the way in which they operate, e.g., in the sourcing of raw materials from developing countries, in the methods used to gain and hold market share, and in the way in which products are promoted to consumers. Consumers face ethical issues in the way in which the food industry relates to them and, for instance, maintains respect for them. Problem-solving in applied ethics relies on judgements made in relation to facts, and the quality of facts is critical to decision-making that is ethically defensible. Ethical judgements cannot be made in isolation from ethical theory, and an appeal to accepted concepts in moral theory and systems of ethical thinking is usual. Five ethical theories are commonly accepted as approaches to ethical problem-solving and decision-making: the common-good theory; deontological theory; utilitarian theory; the theory of justice; and virtue theory.

The Common-good

Though we exist as individuals with our own desires and goals, we generally recognize that our own good is linked to the good of the community of which we are a part. If we value the community and what it stands for, we must allow that while we may expect others to respect our interests, and in return we may respect theirs, the interests of individuals must not be placed above those of the community as a whole. As individuals, we may strive to achieve our own goals, but collectively, we must also recognize the importance of achieving goals that are shared by, and benefit, the community as a whole. In complex societies, individuals exert only a small influence over the common good compared with governments, institutions, and the organizations that bear upon the social and economic structures of society. Society must place reliance on those in government, for instance, to recognize the common-good as it relates to all people, and to maintain conditions and achieve objectives that are similarly to everyone's advantage. Social and economic policies must be aimed at benefiting all of society equally. Social institutions such as education and public health provision, the legal system, and the police and fire services are all common-goods. Food businesses contribute to the common-good through the provision of food and employment. Sometimes, they become involved with society in ways that appear altruistic and intended to further the common-good, but, we must question the degree to which they are able to place the interests of society above their own interests. A supermarket chain providing computers to schools might appear intent on benefitting the common-good, but such action may, in reality, be self-interested and designed expressly to gain access to children to influence the household food purchasers of the future.

Deontological Theory

The term 'deontological' is derived from the Greek, *deontos*, meaning duty. Deontological theory is concerned with notions of duty, rights, and justice. It proposes that we have a duty to tell the truth, not to harm others, to keep promises and to respect the privacy of others. Consequently, it recognizes the right of others to: know the truth and be informed

of issues that affect them; be protected from any harms that may arise from our actions; receive what has been promised to them in agreements and contracts entered into freely; and be allowed privacy in their personal lives provided their actions do not violate the rights of others. The German philosopher, Immanuel Kant (1724–1804), has had a profound influence on the development of this class of ethical theory. Significantly, his concept of the *Categorical Imperative*, is proposed as a guide to right action. He states that in matters of morality, we should act only in ways that we would want to be universalized, i.e., ways that would be acceptable as the norm, and that we should treat other people as ends-in-themselves and not the means to our ends. In the marketing of 'reduced fat' foods, we often see truth sidelined and weight-conscious consumers made the means to the end of increased sales. Foods marketed with the suggestion that they are healthy, because they have reduced fat levels, often contain substantial amounts of energy from the carbohydrate content: a fact not honestly revealed to consumers.

Utilitarianism

As a *consequentialist* ethical theory, utilitarianism is concerned with the outcomes of actions. Consequentialism takes the position that the value of an action is dependent on the 'good' of the consequences, but the definition of 'good' is dependent on the moral agent and may not be what is commonly agreed to be good. To resolve this difficulty, Jeremy Bentham (1748–1832) defined the 'good' as whatever yields the greatest pleasure and the least pain, and proposed a calculus of utility – hence the name utilitarianism – by which to measure the balance between pleasure and pain. The practical limitations of Bentham's utilitarian calculus led John Stuart Mill (1806–1873) to define the 'good' as happiness. In utilitarian theory, an action is morally right if it results in "*the greatest happiness for the greatest number.*" The conduct of the food industry as a whole appears, on the surface, to accord with notions of utilitarianism. Through the work of the industry, Western consumers are generally able to obtain nutrition of a very high standard, and it would seem that the industry is well disposed to provide for the greatest happiness of the greatest number in matters of food.

Justice

The desires, ambitions, and goals of human beings as individuals can conflict with the good of others and the common-good. Egoism is a form of consequentialism that considers people to be self-interested moral agents who will judge an act to be right if it furthers their own interests, even to the exclusion of others. If it is in their interests, egoists will act with favoritism and discrimination, both of which, by general moral agreement, are regarded to be unfair and unjust. Ideas of justice are central to moral philosophy. Aristotle (384–322 BC) proposed a system of justice and considered that people should be treated equally, unless there is some morally defensible reason to treat them differently. John Rawls has recognized that the stability of a society, or social grouping, depends on the members of the society, or group, feeling they are treated justly. In his book, *A Theory of Justice*, he proposes a *contractarian theory*, whereby social justice can be delivered only by rational agents who have dispensed with particular allegiances and interests, but who retain human needs and dependencies. To enable rational agents to make decisions for the benefit of society without the conflict of self-interest, he creates the metaphor of the 'veil of ignorance,' which shields agents from knowing their role in society. Rawls' ideas on justice, thought of as justice-as-fairness, recognize the difficulty that those in authority have in stepping outside the circle of their own interests to act for the good of others. Accepting that society generates benefits and burdens, he puts forward the 'difference principle' as a moral requirement to ensure that the situation of the least well-off is made as good as it can be, thereby working towards justice-as-fairness. Ensuring that the poor have access to an adequate diet is an issue of justice.

Virtue

A person who is virtuous in the way he lives will be an ethical person. Aristotle's theory of ethics is virtue theory. He considered there are two kinds of virtue: moral virtues or qualities of one's character, and intellectual virtues or qualities of one's mind. Of the moral virtues, he proposed 12, sitting in each case midway between representatives of the vices of excess and deficiency. So, the virtue of courage sits between rashness and cowardice, and the virtue of patience between irascibility and lack of spirit. Of the intellectual virtues, he identified five primary virtues (technical skill, scientific knowledge, prudence, intuition, and wisdom) and four secondary virtues (resourcefulness, understanding, judgement, and cleverness). Today, people who exhibit the virtues of courage, honesty, integrity, fairness, compassion, generosity, fidelity, and self-control are most generally admired, for they are recognized as possessing qualities to which we should all aspire. Though we might believe that virtue is important to society, some food businesses celebrate vices to further their own interests, e.g., wickedness and sin made synonymous with pleasure in chocolate, snack-food, and alcoholic beverage advertisements.

Ethical Decision-making

The resolution of ethical problems will benefit from reference to ethical theories, as well as the ethical principles of respect for autonomy, nonmaleficence, beneficence, and justice. Ethically defensible decision-making must be based on:

- identifying the benefits and harms likely to result from alternative courses of action;
- selecting the course of action likely to yield the best consequences and least harm for those affected;
- maintaining respect for the moral rights of those affected;
- treating everyone equally and justly;
- valuing moral virtues and promoting the common good.

Issues in Food Ethics

A wide variety of food ethics issues exist. Those of growing concern to consumers in the UK have been identified by the Co-operative Wholesale Society, as listed in Table 1. A detailed discussion of issues would be beyond the scope of this article, and the following examples have been selected for illustrative purposes.

Feeding the World's Poor

The United Nations' Universal Declaration of Human Rights states that everyone has the right to a standard of living adequate for the health and well-being of themselves and their family, including food, clothing, housing, etc. The imbalance in the quality and supply of food between the rich and the poor people of the world marks one of the central injustices of our time and a critical issue of food ethics. The lack of food necessary to the health and well-being of the poor is significantly a matter of economics as well as politics, but the degree to which remedies to the problem may be sought depends on the sense of moral duty felt by the rich for the poor. In Western societies, supermarkets exhibit a remarkable array of food products, many of which are unnecessary to either the health or happiness of consumers. With the march of globalization, Western food businesses are increasingly sourcing cheap food products in developing countries to maintain the array of products available in supermarkets and to enhance profitability. The extent to which farmers and growers in developing countries are able to benefit from this global trade is questionable. The growth in 'fair-trade' organizations in response to the development of the global food chain would indicate that justice is not distributed equally throughout the chain. Questions therefore arise about the food industry's recognition of its moral duty to suppliers in developing countries and the ethical nature of the industry's relationships with them.

Table 1 Cooperative wholesale society's 'food crimes'

1. *Blackmail* – ''the insidious targeting of the public by global big business putting huge marketing muscle behind products that fail to fit with healthy eating advice''
2. *Contamination* – ''the unnecessary use of chemicals on land and in livestock – interference with nature's way''
3. *Grievous bodily harm* – ''the disregard of animal rights to keep costs down or, even worse, to pamper our taste buds with so called 'luxuries' ''
4. *Vandalism* – ''the destruction of the planet by the intensification of food production systems''
5. *Cannibalism* – ''the practice of permitting animals to be fed with the remains of their own species, or herbivores with animal by-products, or giving animals feed made from the blood of other animals''
6. *Pillage* – ''the careless exploitation of countries, cultures and creeds by multinational concerns milking the so-called global economy''
7. *Fraud* – ''the deliberate assault on the taste and appearance of our food''

Source: CWS (2000) *The Food Crimes Report*. Manchester, UK: The Co-operative Wholesale Society.

Marketing Food to Children

Given the inequalities of the world's food supply, it is ironic that the public health institutions of many Western societies report increasing levels of disease associated with the overconsumption of food. Fast-foods, convenience foods, snack-foods, and confectionery, as well as carbonated beverages, are placed at the center of concerns over the quality of diets in Western societies. Examination of advertising associated with these product groups shows how much manufacturers and retailers depend on children and young people for their markets. The value of children to some kinds of food businesses is seen in the way in which fast-food organizations operate synergistically with movie companies, such as linking cartoon characters with food products, when conditioning young minds has been set as a marketing priority. Food habits formed during childhood are extremely resistant to modification and change. The 'branding' of children's minds with images of certain kinds of food products to sustain the desire for products throughout life may seem an effective strategy for some food businesses. The practice raises many questions about the moral duty of society to protect children, and about the ethical conduct of the industry in making children the means to its end of profitability.

Consumer Dependency

With the development of the food industry during the last quarter of a century, food and cooking have been

taken out of the kitchen and into factories. Many people today do not learn the skills associated with cooking in the home that were commonly held in the past. Processed and prepared foods are attractive to consumers. Such foods allow consumers to buy time for activities other than cooking. Questions arise about the ability of 'food-ignorant' consumers to manage a balanced diet and any moral responsibility the food industry has for the health and well-being of consumers encouraged into dependence on the industry for their food.

Limits to Added-value Products

Of the problems faced by the food industry, perhaps the greatest is the physical limit on the amount of food a consumer can eat in a day and the consequences this has for the industry's goal of continuous economic growth. The innovation of 'added-value' food products that command higher prices and the creation of a now vast range of such products has generated new areas of profitability. The industry maintains that it only responds to consumer demands. Whether consumers actually demand so many new products that making food choices becomes problematic or whether the industry manipulates consumers using sophisticated marketing techniques to create markets for new products is debatable. In this, there is a possible conflict with the idea that the industry is utilitarian in nature; perhaps, alternatively, it is egoistic, according to the moral values of food-business leaders. This is something of which food ethics may bring a greater understanding.

Conclusion

Numerous questions can be asked about the moral values of the food industry and the ethical nature of its conduct. Food ethics provides a reasoned and structured approach to answering questions and to guiding the conduct of the industry such that it maintains respect for people and a common-good relationship with society. In Western societies, people exist in a state of mutual interdependence with the food industry. People as consumers rely on the industry for their food, and the industry needs people to consume its products to remain in business. The food industry regards people not as individuals but as consumer groups, categorized according to various socio-economic criteria. For society to flourish, people as the individuals who constitute society must also flourish. The food industry has a critical part to play in this, as the products of its endeavors are essential to the life of people. Critically, they are consumed and become part of people, affecting both their health and well-being. The food industry has a moral duty to act at all times for the good of people as individuals, not merely as units of consumption. In this, food ethics is of instrumental value to the industry's decision-making and strategic development, helping to ensure that the basis of its existence is morally sound and its conduct is ethically defensible, for the good of people, society, and the industry itself.

Glossary

Autonomy a person's capacity for self-determination.
Beneficence acting to benefit others.
Consumers people as purchasers and/or users of the products and services provided by industry and commerce.
Food supply system the organization of business entities and their activities operating to bring food and food products to the food marketplace.
Non-maleficence acting so as to cause no harm.
Utility The basic unit of desirability as judged by the agent who desires it.

See also: **Biotechnology in Food Production**; **Food and Agriculture Organization of the United Nations**; **Food Safety**; **Legislation**: International Standards

Further Reading

Accum FC (1820) *A Treatise on the Adulteration of Food and Culinary Poisons. Exhibiting the Fraudulent Sophistications of Bread, Beer, Wine, Spiritous Liquors, Tea, Coffee, Cream, Confectionary, Vinegar, Mustard, Pepper, Cheese, Olive Oil, Pickles and other Articles Employed in Domestic Economy, and Methods of Detecting Them*. 2nd edn. London: Longman, Hurst, Rees, Orme, and Brown.

Beauchamp TL and Childress JF (1994) *Principles of Biomedical Ethics*, 4th edn. New York: Oxford University Press.

Bentham J (1988) *The Principles of Morals and Legislation*. New York: Prometheus Books.

CWS (2000) *The Food Crimes Report*. Manchester, UK: Co-operative Wholesale Society.

FAO (2001) *Report of the Panel of Eminent Experts on Ethics in Food and Agriculture*. Rome: Food and Agriculture Organization of the United Nations.

Hassall AH (1855) *Food and its Adulterations*. London: Longman, Brown, Green and Longmans.

Mepham B (ed.) (1996) *Food Ethics*. London: Routledge.

Mill JS (1960) *Utilitarianism – Liberty – Representative Government*. London: J.M. Dent.

Rawls J (1973) *A Theory of Justice*. Oxford: Oxford University Press.

Sinclair U (1986) *The Jungle*. New York: Penguin Books.

UN (1948) *Universal declaration of human rights. General Assembly Resolution 217 A (III) of 10 December 1948* http://www.amnesty.org/aboutai/ughr.htm

Zwart H (2000) A short history of food ethics. *Journal of Agricultural and Environmental Ethics* 12: 13–126.

FOOD FORTIFICATION

M G Venkatesh Mannar, The Micronutrient Initiative, Ottawa, Canada

Synopsis

Vitamins and minerals (commonly termed 'micronutrients') are essential for efficient energy metabolism and other functions of the human body. Food fortification is increasingly recognized as an effective means of delivering micronutrients through commonly consumed foods. Fortification of foods can provide meaningful amounts of the nutrient at normal consumption of the food vehicle. The level of fortification should take into account variations in food consumption to ensure safety for those at the higher end of the scale and impact for those at the lower end. Fortification needs to be supported by adequate food regulations and labeling, quality assurance, and monitoring to ensure compliance and desired impact.

Prevalence and Impact of Micronutrient Deficiencies

Many vitamins and minerals cannot be synthesized by the human body. They must be provided by the diet. They are necessary for the regulatory systems in the body for efficient energy metabolism and for other functions such as cognition, immune system, and reproduction. The amounts needed are small – micrograms or milligrams a day – so they are called 'micro' nutrients (Table 1). When intake of these nutrients is inadequate, individuals and communities suffer serious consequences, including learning disabilities, impaired work capacity, illness, and death. During rapid growth, micronutrient intake must increase or else growth failure or deficiency diseases develop. It is during these periods that deficiency symptoms are most prevalent. For this reason, preschool-aged children, adolescents, and reproductive age women are high-priority groups.

Nearly one in three people living especially in the developing world are affected by one or more micronutrient deficiency. Iron and iodine deficiencies are the most common and widespread nutritional problems. About half the reproductive age women in most of the developing countries suffer from iron-deficiency anemia (IDA). Other high-risk groups include preschool and school-age children and adolescents. Iodine-deficiency disorders affect 20–60% of people in different areas. Vitamin A deficiency (VAD) has been widely recognized, though at a lower clinical prevalence (around 1% in preschool children). However, VAD at a subclinical level (based on serum retinol assays) is thought to be around 10–25%. This is associated with increased mortality and morbidity risk in children and pregnant women. There are other deficiencies that are certainly widespread but far less well recognized, e.g., rickets in young children, associated with calcium and vitamin D deficiency. Dietary intake surveys and limited biochemical indicators show that zinc deficiency is likely as widespread as that of IDA. Selenium and other micronutrients such as folate and vitamin E may have major role in reducing risks of chronic disease (e.g., cancers, heart diseases).

The major cause of micronutrient deficiencies is a lack of adequate intake of bioavailable minerals and vitamins from the diet. This is exacerbated by the fact that commonly consumed foods and beverages (such as rice, wheat, corn, legumes, tea, and coffee) are high in inhibitors and low in enhancers of micronutrient absorption.

Role of Food Fortification

Micronutrient deficiencies can be prevented and even eliminated if populations consume adequate quantities of the required micronutrients on a continuous and ongoing basis. In addressing micronutrient malnutrition in a country, a combination of interventions involving promotion of a well-balanced diet, improving food availability and micronutrient bioavailability, and increasing food consumption, food fortification, and pharmaceutical supplementation need to be emphasized and implemented in a complementary manner. Among these options, food fortification could play an important role to meet a demonstrable need for a nutrient in the population apparent from dietary, biochemical, or clinical evidence.

Fortification is therefore one part of a range of measures that influence the quality of food that include improved agricultural practices, improved food processing, and storage and consumer education to adopt good food-preparation practices. The role of foods and specific nutrients and nonnutrients in providing beneficial nutritional and physiological effects is also gaining attention. As consumer interest in the link between diet and health increases, there are many

Table 1 Recommended dietary allowances (RDAs) for some essential micronutrients

	Minerals								*Fat-soluble vitamins*				*Water-soluble vitamins*						
Category	*Age or condition*	*Calcium (mg)*	*Phosphorus (mg)*	*Magnesium (mg)*	*Iron (mg)*	*Zinc (mg)*	*Iodine (μg)*	*Selenium (μg)*	*Vitamin A (μg RE)*[a]	*Vitamin D (μg)*[b]	*Vitamin E (mg αTE)*[c]	*Vitamin K (μg)*	*Thiamin (mg)*	*Riboflavin (mg)*	*Niacin (mg NE)*[d]	*Vitamin B_6 (mg)*	*Folate (μg)*	*Vitamin C (mg)*	*Vitamin B_{12} (μg)*
Infants	0–6 months	400	300	40	6	5	40	10	375	7.5	3	5	0.3	0.4	5	0.3	25	30	0.3
	7–12 months	600	500	60	10	5	50	15	375	10	4	10	0.4	0.5	6	0.6	35	35	0.5
Children	1–3 years	800	800	80	10	10	70	20	400	10	6	15	0.7	0.8	9	1.0	50	40	0.7
	4–7 years	800	800	120	10	10	90	20	500	10	7	20	0.9	1.1	12	1.1	75	45	1.0
	8–11 years	800	800	170	10	10	120	30	700	10	7	30	1.0	1.2	13	1.4	100	45	1.4
Males	12–14 years	1200	1200	270	12	15	150	40	1000	10	10	45	1.3	1.5	17	1.7	150	50	2.0
	15–18 years	1200	1200	400	12	15	150	50	1000	10	10	65	1.5	1.8	20	2.0	200	60	2.0
	19–24 years	1200	1200	350	10	15	150	70	1000	10	10	70	1.5	1.7	19	2.0	200	60	2.0
	25–50 years	800	800	350	10	15	150	70	1000	5	10	80	1.5	1.7	19	2.0	200	60	2.0
	51+ years	800	800	350	10	15	150	70	1000	5	10	80	1.2	1.4	15	2.0	200	60	2.0
Females	12–14 years	1200	1200	280	15	12	150	45	800	10	8	45	1.1	1.3	15	1.4	150	50	2.0
	15–18 years	1200	1200	300	15	12	150	50	800	10	8	55	1.1	1.3	15	1.5	180	60	2.0
	19–24 years	1200	1200	280	15	12	150	55	800	10	8	60	1.1	1.3	15	1.6	180	60	2.0
	25–50 years	800	800	280	15	12	150	55	800	5	8	65	1.1	1.3	15	1.6	180	60	2.0
	51+ years	800	800	280	10	12	150	55	800	5	8	65	1.0	1.2	13	1.6	180	60	2.0
Pregnancy		1200	1200	320	30	15	175	65	800	10	10	65	1.5	1.6	17	2.2	400	70	2.2
Lactation	0–6 months	1200	1200	355	15	19	200	75	1300	10	12	65	1.6	1.8	20	2.1	280	95	2.6
	7–12 months	1200	1200	340	15	16	200	75	1200	10	11	65	1.6	1.7	20	2.1	260	90	2.6

[a]Retinol equivalents. 1 retinol equivalent = 1 μg of retinol or 6 μg of β-carotene.
[b]As cholecalciferol. 10 μg of cholecalciferol = 400 IU of vitamin D.
[c]1 mg of D-α tocopherol = 1 α-TE.
[d]1 NE (niacin equivalent) = 1 mg of niacin or 60 mg of dietary tryptophan.
Source: National Research Council (1989) Subcommittee on the Tenth Edition of the Recommended Dietary Allowances. Washington, DC: National Academy Press.

opportunities for developing foods that can have positive effects on health and performance.

Food fortification involves the identification of commonly eaten foods that can act as vehicles for one or more micronutrients and lend themselves to centralized processing on an economical scale.

The Codex Alimentarius provides definitions for terminology relating to the addition of nutrients to foods. Fortification or enrichment means the addition of one or more essential nutrients to a food, whether or not it is normally contained in the food for the purpose of preventing or correcting a demonstrated deficiency of one or more nutrients in the population or specific population groups. Restoration refers to compensation for unavoidable nutrient losses during the course of good manufacturing practice in the preparation and preservation of food or during normal storage and handling procedures. In many countries, there are requirements for the compulsory addition of micronutrients such as iron, niacin, and thiamine to bread and flour.

Fortification, when imposed on existing food patterns, does not necessitate changes in the customary diet of the population and does not call for individual compliance. Often, it can be dovetailed into existing food production and distribution systems. For these reasons, fortification can often be implemented and yield results within 2–3 years and be sustained over a long period of time. It can thus be the most cost-effective means of overcoming micronutrient malnutrition.

Effectiveness of Fortification

The concept of nutrient fortification of staple foods was developed in the early part of this century as a means of dealing with micronutrient and vitamin deficiency diseases that were prevalent in Europe and North America. Salt was iodized in Switzerland in the early part of the century. Margarine fortified with vitamin A was introduced in Denmark in 1918. During the 1930s and 1940s, milk was fortified with vitamin A, and flour was fortified with iron and B vitamins in a number of European countries and in North America. In developed countries, where there is a high dependence on processed foods, and industries are streamlined and automated, food fortification has played a major role in the health of the populations at large over the last 40 years, and several nutritional deficiencies have been eliminated. In addition to the virtual elimination of VAD and iodine deficiency disorders (IDD) and the substantial reduction of IDA, diseases as varied as xerophthalmia, pellagra, beriberi, rickets, goiter, and ariboflavinosis have essentially disappeared. Several studies have demonstrated the effectiveness of food fortification in eliminating micronutrient deficiencies. The current low levels of iron deficiency in the USA are attributable to fortified sources. Almost a quarter of iron intake in the US diet comes from fortified sources, much of that from flour products. In Canada, flour fortification with B vitamins began in Newfoundland in 1944. Within four years, deficiencies that were earlier found in nearly 20% of the population had dropped to negligible levels.

In the developing world, there has been a rapid growth in fortifying a wide range of foods. The most successful global fortification experience is the fortification of salt with iodine. Fortification of other staple foods such as flour, oils, sugar, condiments, dairy products, and a range of processed foods with other micronutrients and vitamins are also increasing. In Latin America, flour fortification is reaching significant numbers of people, and sugar fortification has taken hold in Central America. In Asia, the overall reach of fortified foods and its impact in eliminating micronutrient malnutrition is still small, but growing.

In the last decade, the contribution of salt fortification towards the virtual elimination of IDD in developing countries has been substantial. In most nations where IDD was recognized as a public health problem at the beginning of this decade, an average of 60–70% of all salt for human and animal use is now iodized. The stage is set for the global elimination of iodine deficiency disorders globally by 2005.

Food fortification has also played a major role in substantially reducing vitamin A and iron deficiencies. A national sugar fortification program in Guatemala has virtually eliminated VAD as a public health problem. Flour fortification in Chile and Venezuela are substantially improving iron status across all sectors of the population. Today, in Latin America, Asia, and Africa, consumers of fortified products as varied as margarine, milk, noodles, and cornstarch are substantially protected from a range of micronutrient deficiencies.

Food fortification has to be an integral component of the overall strategy to eliminate micronutrient malnutrition. It involves opening new communication channels among the public health community, research institutions, government regulators, food companies, and a variety of civic and consumer organizations. The inputs from these groups could result in a new alliance focused on accomplishing national development through elimination of micronutrient malnutrition. Food fortification should also be the preferred option to reach preschoolers and primary graders.

Considerations for Planning a Food Fortification Intervention

Food fortification involves the identification of commonly eaten foods in a country or region that can act as vehicles for one or more micronutrients and lend themselves to centralized processing on an economical scale. In developed countries, a range of cereals and cereal products, dairy products, processed foods and beverages are fortified. In most developing countries, staple foods and condiments are the obvious choice for fortification, given their consistent consumption by large sections of the population. In such situations, the choice of vehicles at the present time is limited to a handful of foods and condiments: cereals, oils and fats, sugar, salt, and sauces.

Fortification of foods is aimed to provide meaningful levels of the nutrient at normal consumption of the food vehicle. Variations in food consumption may affect the safety of those at the higher end of the scale and the level of impact for those at the lower end. They should also consider prorated intakes by young children to ensure efficacious and safe dosages. Food fortification also requires careful attention to food regulations and labeling, and a careful assessment of the technical and analytical limitations for compliance with label declarations. The Codex Alimentarius has adopted a number of basic principles set out in Table 2 for the addition of nutrients to foods.

Infants and children under the age of 24 months consume a different dietary pattern than do older individuals. There are no major differences in consumption among children aged 2 years and older. Where there are few centrally processed complementary foods, and consumption of such foods is infrequent and rare, careful research is needed on the best ways to address micronutrient problems of infants.

Key considerations for planning and implementing food fortification programs in countries are discussed below.

Selection of Food Vehicles

The selection of a suitable vehicle and a fortificant that deliver the micronutrient in the right amounts consistent to an entire population is extremely rare. Salt iodization is one example of successful large-scale fortification in the developing world – the primary reason being the simple and low-cost technology and the narrow variation in salt consumption within a given region or population. There exist other opportunities for single and multiple fortification of several commonly eaten foods such as wheat and wheat products, corn, rice, milk and milk products, cooking oils, sugar, and condiments. As processed foods gain popularity in the developing world, with an increasing market outreach, they offer new channels for micronutrient delivery. The decision to use a particular food or beverage must be based on scientific data. Well-designed market research, which includes dietary surveys and habits and practices, should be used as a tool in identifying the commonly consumed foods and beverages.

Potential food vehicles could be visualized as a three-tiered pyramid (Figure 1). Staple foods such as sugar, cereals, grains, fats, and oils form the base; basic foods such as breads and biscuits, packaged cereals and flours, and dairy products are in the middle; and value-added foods such as condiments, snacks, candies, convenience and ready-to-eat foods are at the top. Fortifying less expensive staple foods at the base of the pyramid results in broader dissemination of micronutrients throughout the population. Also, since basic and value-added foods are processed

Table 2 General principles for the addition of nutrients to foods

1. The essential nutrient should be present at a level that will not result in either an excessive or an insignificant intake of the added essential nutrient considering amounts from other sources in the diet
2. The addition of an essential nutrient to a food should not result in an adverse effect on the metabolism of any other nutrient.
3. The essential nutrient should be sufficiently stable in the food under customary conditions of packaging, storage, distribution, and use.
4. The essential nutrient should be biologically available from the food.
5. The essential nutrient should not impart undesirable characteristics to the food (e.g., color, taste, flavor, texture, cooking properties) and should not unduly shorten the shelf-life.
6. Technology and processing facilities should be available to permit the addition of the essential nutrient in a satisfactory manner.
7. Addition of essential nutrients to foods should not be used to mislead or deceive the consumer as to the nutritional merit of the food.
8. The additional cost should be reasonable for the intended consumer.
9. Methods of measuring, controlling and/or enforcing the levels of added essential nutrients in foods should be available.
10. When provision is made in food standards, regulations, or guidelines for the addition of essential nutrients to foods, specific provisions should include identifying the essential nutrients to be considered or to be required and the levels at which they should be present in the food to achieve their intended purpose

Source: Codex Alimentarius Commission (1994) *Codex Alimentarius. General Principles for the Addition of Essential Nutrients to Foods*, vol. 4. CAC/GL 09-1987. Rome: Food and Agriculture Organization.

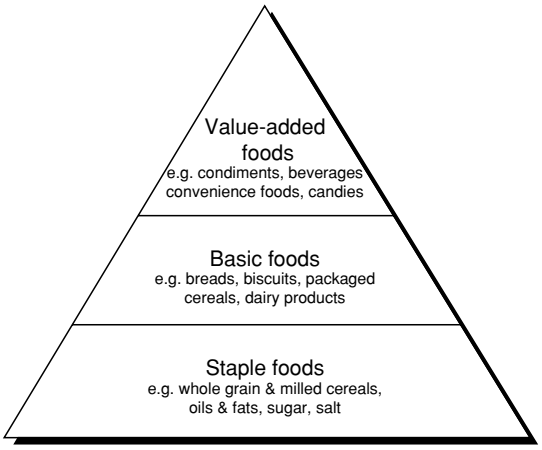

Figure 1 Food-product pyramid.

from staple commodities, fortifying foods at the base of the pyramid results in fortifying products throughout the food chain.

In many countries, a flexible approach using a variety of vehicles each fortified to a specified proportion of recommended dietary intake (RDI) may offer an effective option. If consumption of a particular vehicle is consistent in some groups but only sporadic in others, fortifying several vehicles will likely provide complementary coverage. When considering a multifaceted approach, each food vehicle offers specific opportunities and constraints:

Cereals Cereal grain products such as rice, corn, and wheat flour constitute an important calorie and protein source for many populations. They have proven to be good vehicles in several developed and developing countries, but their application may be limited where cereals foods are often eaten where they are grown and processed at the community level.

Fat and oils Cooking fats and oils may offer an option to deliver a portion of the recommended dietary allowance (RDA), particularly the fat-soluble vitamins such as vitamin A. They have an advantage in that they are often centrally refined and packed.

Dairy products Milk may offer an option where centralized dairy processing exists.

Condiments Sugar, spices, starches and sauces are attractive carriers. Some are processed centrally and consumed in regular quantities.

Value-added products In developed countries such as the USA, fortification of snack foods has been recommended as an appropriate public health measure, given the increasing percentage of dietary energy that the population derives from snack foods. In developing countries, the most vulnerable populations consume these higher-priced products only sporadically. However, consumer awareness, technical breakthroughs, and marketing innovations often emerge from the development of fortified value-added products.

When no universally consumed vehicle is available in a country, the fortification of a number of foods offers several key strategic advantages. When a variety of food sectors are engaged, no single industry can claim unfair treatment. When a variety of foods are fortified, each with a lesser proportion of RDI, the theoretical possibility of consuming dangerous levels of a micronutrient through excess consumption of a single food becomes more remote.

Technologies

Depending on the manner in which foods are processed, different methods for addition of micronutrients have been developed:

- *Dry mixing* for cereal flours and their products, powder milk, powder beverages.
- *Dissolution in water* for liquid milk, drinks, fruit juices and in the water to be used for making bread pastas cookies.
- *Spraying* as in iodization of salt and cornflakes, where the vitamins do not support the cooking or extrusion step.
- *Dissolution in oil* for the lipo-soluble vitamins for enrichment of oily products like margarine.
- *Adhesion* for sugar fortification where the vitamin A in powder form is adhered to the crystals surface by a vegetable oil.
- *Coating* as in rice where the vitamins sprayed over the grain must be coated in order to avoid the losses when washing the grains before cooking.
- *Extrusion* of a blend of rice flour with micronutrients and stabilizers to provide a premix grains that can be blended with rice after milling.

Research and development efforts have enhanced effectiveness of fortification technology. Better refining procedures and packaging have significantly improved the stability of iodine compounds in salt. In the case of iron, stabilizers and absorption enhancers are added along with the fortificant to retain it in an absorbable form or improve absorption. The structure of the iron or iron compounds has been modified to improve absorption. In the case of vitamin A, work is ongoing to retard the loss of potency on storage through addition of antioxidants and other stabilizers. In all these cases, there is scope for further improvement and refinement for better

product stability, absorption, and lower cost, and development efforts need to continue.

Regulations, Guidelines, and Standards for Fortification

The objective of the regulations and guidelines should be (1) to protect the population who buy and consume the product and (2) establish standards and guidelines to which all producers of the fortified products should conform.

The regulation and guidelines should include the following:

A rationale for fortification The decision to fortify a product with particular nutrient has to be based on scientific data obtained through dietary surveys and biochemical and/or clinical studies.

Recommended minimum and maximum level(s) of nutrient added per serving The level of the fortificant added should not be too low to have little or no impact on the nutrition status of the population. At the same time, the level added should not be in excess to cause toxicity.

Establishing RDA This is essential as a reference for fortification and in communicating the amount of the nutrients per serving to the population that consume the fortified product.

Establishing labeling standards The goal is to detail only relevant information.

Establishing guidelines for making nutritional and health claims Claims made on the nutritional and health benefits of the product should be supported by scientific data.

Establishing a monitoring and surveillance system For the regulations and guidelines to be effective, there should be a monitoring and surveillance system established.

Quality Assurance

The success of producing and distributing a micronutrient-fortified product is dependent on the placement of a quality assurance program that includes the following:

Finished product specification and description In order to build quality into the fortified product, the specification and description of the finished product must be established prior to its manufacture. The manufacturing process should be detailed in written operating procedures. Based on the claim to be made, the type and dosage of the micronutrients and the form and source of fortificants are specified.

Raw materials The quality of the raw materials for both the product and packaging play an important role in the quality of the final micronutrient fortified product. The micronutrient fortificants should meet established specifications. In addition, premix purchases are made from approved suppliers that follow good manufacturing practices (GMP). Once received, the fortificants are recorded and stored as specified in an appropriate condition.

Production and Packaging A quality product is ensured by following GMP. The micronutrient sources are added into the product formulation by using already-specified procedure and equipment. The amount of each fortificant added is recorded. Finally, the amount in the finished product is verified by using a validated analytical method.

Release of finished product The product released for distribution should meet all final product specifications. This includes the safety of the product and the level of the micronutrients claimed. There should be a system for (1) monitoring and correcting deviations from specification and (2) identifying the product in case there is a complaint from the consumer.

Documentation Accurate recording is essential for (1) identifying and correcting deviations, (2) recalling a product with serious problems, and (3) making an improvement on the quality of the product.

Training All personnel involved in the development, manufacturing, packing, and storage of the fortified product should be trained in quality assurance.

Advocacy and Social Communications

An effective communications campaign supported by the government should necessarily accompany any major fortification effort in order to gain the understanding and support of key sectors from policymakers and legislators to medical professionals, health workers and consumer groups.

Monitoring and Evaluation

As micronutrient fortification programs expand in countries around the world, the need to monitor the quality of the fortified food (in terms of its micronutrient content) and evaluate its contribution to the alleviation of micronutrient deficiencies will increase. Monitoring activities should ensure that a fortified product contains a sufficient amount of the micronutrient and that it reaches the target population. The

indicators used to monitor fortification efforts reflect the need to determine the quality and safety of the products and their availability to the populations at greatest risk. Monitoring is carried out at various levels – production or importation, wholesale and retail and household. The production and importation levels are the most important point for monitoring the fortification activity.

Monitoring is vitally important to the success of a fortification program for the following reasons:

To improve program effectiveness A fortification program can be effective only if a high-quality fortified is produced, delivered, accepted, and consumed by the target population on a continuous and self-sustaining basis.

To ensure compliance with government standards Monitoring systems will provide information on the extent to which the product conforms to national standards at points of production and consumption. This information is needed to enforce compliance through regulation and encourage voluntary corrective action by the producer.

To identify problem points in the fortification process As food moves from the factory to the household, it generally goes first to distributors, then to retail outlets. In some cases, foods are fortified outside the country and imported for eventual consumption. Monitoring must occur at various critical points starting at the production level and ending at the household level, where the fortified food is stored, prepared, and consumed.

To ensure safety Effective monitoring will ensure that the micronutrient content of the fortified food is within the desired range.

As a final indication of effectiveness, an impact evaluation conducted after 2 or 3 years will determine whether the micronutrients are absorbed well and are preventing recurrences of micronutrient deficiencies in the population.

Progress in Food Fortification

Key global developments to address specific micronutrient deficiencies through food fortification are summarized below.

Iodine Over the past 60 years, a variety of vehicles including bread, flour, sugar, and salt have been proposed as carriers for iodine. Among these, salt iodization has witnessed a remarkable growth in application globally over the past decade. A significant proportion of the populations in over 110 countries have access to iodized salt. Once established in a country, salt iodization is a permanent and long-term solution to the problem of iodine deficiency. Depending upon per-capita salt consumption and anticipated iodine losses between production and consumption, iodine dosage ranges from 20 to 100 mg per kilogram of salt. These losses are dependent upon the quality of salt and packaging. Toxicity issues are negligible, and cost considerations fairly small. Although there is considerable progress in streamlining control programs through salt iodization in several countries, producer compliance, quality assurance, logistic problems, and bottlenecks do remain. The challenge is to systematically identify and tackle these constraints through effective advocacy, social communications, monitoring of salt iodine levels, regulation, and enforcement. In many developing countries, salt iodization is the first large-scale experience in national fortification of a commodity. It has provided valuable lessons in collaboration between government, industry, nongovernmental organizations, the media, and other sectors. It has also given insights into building and sustaining an intervention politically, technically, financially, and culturally.

Iron Although iron has potential for use in more food vehicles than iodine or Vitamin A, fortification with iron is technically more difficult than with other nutrients, since iron reacts with several food ingredients. There are a number of iron compounds that are approved for food fortification. The greatest challenge with iron is to identify a form that is adequately absorbed and yet does not alter the appearance or taste of the food vehicle. The buff-colored, insoluble iron phosphate compounds are stable under a variety of storage conditions but are poorly absorbed. The soluble iron salts like ferrous sulfate are well absorbed but easily discolor by reacting with food ingredients. In addition to the general fortification criteria, iron fortification needs to take into account certain specific considerations:

1. The vehicle should be a component of all meals, because absorption varies inversely with the iron content of the meal.
2. It should not require prolonged storage particularly under hot and humid climatic conditions, since this could cause organoleptic problems.
3. Vehicles that are dark in color or have a strong taste or odor permit the use of more reactive iron compounds.
4. Segregation of iron should not occur during mixing or storage.

Reducing the global prevalence of iron deficiency anemia will require the use of several alternative

fortification compounds and food vehicles, depending upon the local dietary patterns. Some promising possibilities are:

Wheat/corn flour and bakery products In developed countries like the USA, Sweden, and Canada, these foods have had significant success as vehicles for iron. Flour fortification ensures an even meal distribution of the added iron and provides a reasonably constant iron supply to each individual. More recently, all corn and wheat flour in Venezuela was fortified with iron (mixture of ferrous fumarate and reduced iron), vitamin A, and B vitamins.

Condiments In Thailand, fish sauce, which is consumed in many parts of the country, has been tested as a vehicle for iron [in the form of sodium iron ethylenediaminetetraacetic acid (EDTA)] and has shown a good hemoglobin response. Fish sauce and fish pastes have also been tried as vehicles in the Philippines and Vietnam. Curry powder used by an Indian population in South Africa has been fortified with sodium iron EDTA with successful results. In China, the feasibility of using soy sauce as a carrier for iron is being examined.

Processed foods There is a growing trend in several developing countries increasingly to use processed foods such as noodles, cereals, and soup cubes. Weaning foods for children are also becoming popular. Elemental iron of small particle size is being used for the fortification of processed cereals in several developed countries. In Mexico, a chocolate milk powder is fortified with iron. In Tanzania, a fortified orange beverage drink has proved effective in improving the iron and vitamin A status among school children and pregnant mothers.

Salt Recent work has focused on developing a formulation that would permit the addition of both iodine and iron. (The acidic medium in which the iron compound is stable causes the rapid oxidation of iodide/iodate to free iodine, which vaporizes and is lost). A promising approach uses dextrin-encapsulated potassium iodide/iodate mixed with encapsulated ferrous fumarate to provide a barrier between the iron and iodine and improving stability. The successful testing and large-scale application of available technology for the double fortification of salt could represent a major breakthrough for combined iron and iodine delivery.

Zinc Several zinc compounds are available for food fortification ranging in solubility from very soluble (sulfate and chloride), freely soluble (acetate), to almost insoluble (carbonate and oxide). The soluble compounds, especially sulfate and chloride, are most preferred as fortificants. Unlike iron, zinc does not undergo oxidation–reduction. As a result, it does not cause the development of undesirable color and/or flavor when added to foods and beverages. However, like iron, it does cause metallic aftertaste and astringency. Currently, only infant formulas, cereal products (especially infant cereals), and dietetic products (e.g., products used for enteral feeding) are fortified with zinc.

Calcium Foods normally fortified with calcium include cereals and beverages, foods sold as a replacement to milk/yogurt, e.g., soy milk/yogurt, infant formula, weaning foods, enteral foods, slimming foods, and meal replacements. Flour is one of the best examples of fortification of cereal products. A range of breakfast cereals, rice, and corn products are also fortified with calcium in a few countries. More recently, orange juice and concentrates have been fortified with calcium. The compounds commonly used for fortification include carbonate, phosphate/lactate, citrate malate, or gluconate.

Vitamins A and D Vitamin A has been successfully added to foods such as milk (various forms of liquid and dry milk), butter, cheese, flour, bread and cereal product, snacks, and beverages. The most widespread application is in oils or fats that are dietary ingredients of most populations. It is essential that foods fortified with vitamin A are sealed in opaque or dark-colored glass containers with closures to eliminate the effect of light, a stimulant to oxidative processes. Edible antioxidants may also be added to protect both the oil and the added vitamin A from oxidation. Fortification of margarine is widely practiced (at levels of 20 000–50 000 IU kg^{-1}) to equal or exceed the average levels of this nutrient in butter. Often, vitamins D and E are also added. If synthetic β-carotene is also to be used, both color and vitamin A can be added simultaneously. When fat is removed from whole fluid milk, all fat-soluble vitamins, including vitamin A, are removed. All skim milk or low-fat liquid or dry milk products should have vitamin A added and, if given to infants, should probably have vitamins D, E, and possibly C as well. Sugar is another vehicle for vitamin A in Central American countries like Costa Rica, Guatemala, Panama, and Honduras.

B vitamins and folate Thiamine, riboflavin and niacin have been traditionally added to wheat flour in many countries for several decades. As a result, several vitamin B deficiency conditions like beriberi

and pellagra have been eliminated. Most recently, the addition of folate to grain products was made mandatory to reduce the prevalence of neural-tube defects and other congenital abnormalities in children caused by folate deficiency in the mothers during pregnancy.

See also: **Calcium**: Properties and Determination; **Carotenoids**: Occurrence, Properties, and Determination; **Cholecalciferol**: Properties and Determination; **Iodine**: Properties and Determination; Iodine-deficiency Disorders; **Niacin**: Properties and Determination; **Quality Assurance and Quality Control**; **Riboflavin**: Properties and Determination; **Thiamin**: Properties and Determination; **Vitamins**: Overview; **Zinc**: Properties and Determination

Further Reading

Bauernfeind JC and Lachance PA (eds) (1991) *Nutrition Addition to Foods; Nutritional, Technological and Regulatory Aspects*. Trumbull, CT: Food and Nutrition Press.

Clydesdale FM and Wiemer KL (eds) (1985) *Iron Fortification of Foods*. London: Academic Press.

Codex Alimentarius Commission (1994) *Codex Alimentarius. General Principles for the Addition of Essential Nutrients to Foods*, vol. 4. CAC/GL 09-1987. Rome: Food and Agriculture Organization.

Food Fortification to End Micronutrient Malnutrition. State of the Art (1998) *Proceedings of the Satellite Conference of the XVI International Congress of Nutrition*. Ottawa: The Micronutrient Initiative.

Hurrell RH (ed.) (1999) *The Mineral Fortification of Foods*. Leatherhead, UK: Leatherhead International Ltd.

Lotfi ML, Venkatesh Mannar MG, Merx RH and Naber van den Heuvel P (1996) *Micronutrient Fortification of Foods. Current Practices, Research and Opportunities*. Ottawa: The Micronutrient Initiative, Ottawa & Wageningen: International Agricultural Centre.

Micronutrient Fortification of Foods. Current Practices, Research and Opportunities. Ottawa: The Micronutrient Initiative: Ottawa, Canada & Wageningen, The Netherlands: International Agricultural Centre.

Mason JB, Lotfi M, Dalmiya N, Sethuraman K and Deitchler M (2001) *The Micronutrient Report. The Micronutrient Initiative*.

Nestel P (1993) *Food in Fortification Developing Countries*. Washington, DC: USAID/Vital.

Richardson DP (1997) The addition of nutrients to foods. *Proceedings of the Nutrition Society* 56: 807–825.

Food Intake *See* **Dietary Surveys**: Measurement of Food Intake; Surveys of National Food Intake; Surveys of Food Intakes in Groups and Individuals

FOOD INTOLERANCE

Contents

Types

J Gray, Guildford, Surrey, UK
W Chan, London, UK

Introduction

The umbrella term 'adverse reactions to food' has been adopted to describe a range of different types of responses to food, in particular food intolerance, food aversion, and food poisoning (**Table 1**). One way of distinguishing food intolerance and food poisoning is to consider the former as a defect in the individual and the latter as that of the food. The mechanisms underlying food intolerance are numerous, complex, and incompletely understood. It is certain that far fewer reactions are due to true allergy (involving the immune system) than is commonly believed. This article discusses the definitions and

Table 1 Adverse reactions to food

Food aversion	*Food intolerance*	*Food poisoning*
Food avoidance	Allergic reactions	Bacteria and viruses
Psychological intolerance	Enzyme deficiencies	Molds
	Pharmacological reactions	Chemicals
	Toxic and irritant effects	Toxins

mechanisms of the various types of food intolerance, as well as the clinical symptoms, diagnosis, treatment, and prevalence of food intolerances and their predisposing factors.

Much of the confusion surrounding this topic has arisen from the lack of universal agreement about definitions of the terms used to describe food intolerance. There is also a difference in terminology used in the UK and in the USA. For the purposes of this review, the definitions based on those developed by the British Nutrition Foundation Task Force (2001) will be used (Table 1).

Food Intolerance

Food intolerance is the term used to describe a range of responses to food, including allergic reactions, adverse reactions resulting from enzyme deficiencies, pharmacological reactions, and responses to toxic or irritant components of food (Table 1). All the adverse reactions categorized under this heading share one specific characteristic: they are reproducible adverse reactions to specific foods or food ingredients.

Allergic Reactions

These are abnormal reactions of a person's immune system to foreign but normally harmless substances, such as pollen grains or food constituents. These reactions may often be mediated by immunoglobulin E (IgE), but there are some reactions which are delayed and which do not involve IgE.

IgE-mediated reactions

Allergic reactions with IgE involvement are immediate, can be severe, and are mediated by IgE antibodies which trigger mast cells (and other cells). The mast cells are present below the surface of the skin and in the membranes lining the nose, respiratory tract, eyes, intestine, and other mucosal sites. During an IgE-mediated reaction, histamine and other proinflammatory substances which are present in these cells are released, causing symptoms such as rhinitis, asthma, dilation of blood vessels and flushing, swelling (e.g., of the lips), and difficulty in breathing. These allergic reactions are usually localized to one or two organs of the body.

Genetic factors are considered to be of particular importance in the development of IgE-mediated reactions. Atopy is the term used to describe a situation where there is a predisposition to the production of IgE antibodies and development of allergic disease.

Children with IgE-mediated reactions often react to more than one food and high levels of IgE antibodies to common food proteins, such as those present in eggs and cows' milk, are often predictive of later allergy to inhalants such as house dust and pollen. Food allergy does not commonly develop for the first time in adults. However, oral allergy syndrome is sometimes seen in adults who have previously suffered hayfever. Oral allergy syndrome is an IgE-mediated reaction, characterized by an initial irritation to the lips, oral mucosa and palate, followed by edema within minutes of ingestion of the trigger food, which is frequently a raw fruit or vegetable, although the syndrome is also associated with the consumption of tree nuts, peanuts, egg, fish, or milk.

Allergic reactions to food vary considerably in their severity and the discomfort they cause, but the majority are not life-threatening. However, anaphylactic reactions can be very severe and even fatal. The term anaphylaxis is now mainly applied to severe reactions of rapid onset, often accompanied by a fall in blood pressure and severe shock. Peanuts are well known for causing this type of extreme reaction. Other foods that are known occasionally to cause a severe (IgE-mediated) reaction are tree nuts, seeds, eggs, milk, and shellfish. Although most children with sensitization to eggs or milk usually spontaneously recover in 12–24 months, sensitization to peanut protein typically is more persistent.

Non-IgE-mediated reactions

These reactions are delayed and complex. Celiac disease is the best-defined example of this form of response. It is a disease that involves a sensitivity to gluten and appears to be mediated by the cells of the immune system. Most experts agree that celiac disease is a non-IgE-mediated reaction occurring in genetically susceptible individuals and triggered by exposure to the gluten fraction of wheat and other related cereal grains, such as rye and barley.

Delayed reactions to food may take several hours or days to develop. Reactions to cows' milk are often delayed and the majority are not accompanied by detectable IgE antibodies.

Enzyme Deficiencies

The most common intolerance linked to an insufficiency of an enzyme is lactose intolerance (also known as lactose maldigestion). Before it can be

absorbed and utilized by the body as fuel, lactose (the sugar in milk) has to be broken down to its two component sugars, glucose and galactose. This requires the enzyme lactase. If lactase is produced in insufficient quantities, some of the lactose can pass undigested into the large intestine, causing symptoms such as diarrhea and flatulence. The symptoms result from bacterial fermentation of lactose in the colon and the associated osmotic effects.

In most populations, especially where the consumption of milk beyond infancy is culturally uncommon, lactase levels decline in childhood. The deficiency is present in about 70% of the world population, particularly among those of African, Asian, and Jewish descent. In South-east Asia, India, the Middle East and parts of Africa, maldigestion of lactose in adults may affect about one-third of the population. However, in the UK, only 1–2% of adults are affected.

On occasions other enzyme deficiencies can result in an inability to handle specific dietary carbohydrates such as galactose and sucrose. These conditions are very rare. (*See* **Inborn Errors of Metabolism**: Overview.)

Pharmacological Reactions

There are many food components that can produce a pharmacological effect. These effects are usually insignificant in clinical terms, unless the substance is consumed in very large quantities, e.g., the methyl xanthine caffeine. (*See* **Caffeine**.)

Also sometimes responsible for this type of reaction is the group of substances known as vasoactive amines, including histamine and tyramine (Table 2). These are normal constituents of a variety of foods such as cheese, pickled fish, yeast extract, chocolate, bananas, avocados, citrus fruits, and wine. It is often stated that some foods have a histamine-releasing action, e.g., egg whites, shellfish, strawberries, tomatoes, and chocolate.

Histamine may also accumulate in several foods as a result of microbial action during spoilage. During this process, histidine in the food is converted to histamine and it is believed that bacterial decarboxylase enzymes are responsible. (*See* **Histamine**.)

Table 2 Examples of vasoactive amines in foods

Vasoactive amine	*Food*
Tyramine	Cheese
	Pickled fish
Phenylethylamine	Chocolate
Serotonin	Bananas
Octopamine	Citrus fruit
Histamine	Fermented foods, e.g., blue cheeses, pickled fish

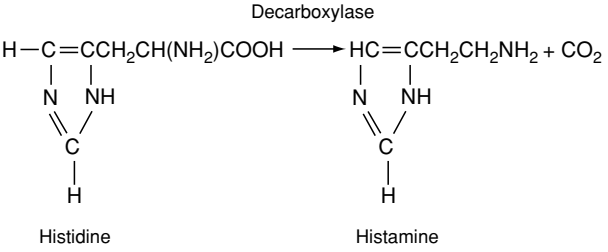

High levels of histamine are responsible for scombrotoxin illness which can occur following the ingestion of scombroid fish, such as mackerel, which has been stored inappropriately and has undergone spoilage. This type of problem can be prevented by insuring that fish are properly refrigerated throughout the processing and distribution chain. (*See* **Fish**: Spoilage of Seafood.)

Histamine-induced symptoms are similar to those found with IgE-mediated food allergy, which also involves the release of histamine and other chemical mediators from mast cells. The two conditions can be distinguished by the relatively large amount of the substance needed to trigger histamine-induced symptoms, in contrast to the minute quantity needed to trigger an allergic reaction.

There is some evidence that reactions to certain food additives, including some coloring agents, such as tartrazine, and some preservatives, such as the benzoates and sulfites, are also pharmacological in origin, and may result from the release of histamine or other chemical mediators. Although these reactions may emulate allergic reactions, convincing evidence to support an immunological basis for them is lacking. Nevertheless, the expression of such reactions is so variable that the possibility of an interaction between biochemically and immunologically mediated mechanisms cannot be excluded until further research is carried out. (*See* **Food Additives**: Safety.)

Toxic and Irritant Effects

Some foods contain natural toxins or substances that can irritate the lining of the intestine. Examples of natural toxins include protease inhibitors in legumes, cyclopeptides, and muscarine in mushrooms, and oxalates in spinach and rhubarb. Spices and chilli may irritate the gut mucosa and cause rapid intestinal transit times. This may be confused with other types of food intolerance. Some people are more susceptible to these effects than others and irritation is more likely if the mucosa is diseased.

Food Aversion

Food aversion is the alteration of eating behavior for psychological reasons. It may manifest as

psychological food intolerance, where there is an adverse physical reaction associated with the ingestion of a particular food, or food avoidance.

Psychological Food Intolerance

This response may be associated with symptoms that may be similar to those associated with true food intolerance. Such reactions are psychosomatic and do not occur when the food or food ingredient is administered in disguised form. Anecdotally, in some cases the bodily reaction can even be reproduced by the mere suggestion that a particular food may have been consumed. Development of an aversion to specific foods is recognized to occur occasionally following vomiting associated with medical treatments, where the food has been eaten at about the same time.

In some individuals food aversion can manifest itself as hyperventilation syndrome in response to food. In such cases, when faced with the food, individuals may hyperventilate to the extent that they suffer severe symptoms and may even lose consciousness. A variant of this form of aversion has been dubbed 'food intolerance by proxy.' This has become increasingly recognized in children. It occurs when an individual's family is unable to accept the symptoms and behavior of the person, except in terms of a physical illness. The belief that the problem is due to an 'allergy' is continuously reinforced within the family, but there is no evidence of a physical intolerance on double-blind challenge.

Food Avoidance

Food avoidance is quite common in the population as a whole, especially in preschool children and adolescents. It may manifest as the avoidance of all food, or of particular foods or food types. It is most prevalent in teenage girls who are frequently obsessed with losing weight and maintaining a particular body image. The extremes of this type of behavior manifest as the eating disorders anorexia nervosa and bulimia nervosa. (*See* **Anorexia Nervosa; Bulimia Nervosa.**)

Clinical Manifestations

The symptoms which can occur as a result of reactions to food fall into three main groups: immediate (within 1 h of ingestion); delayed or late (more than 1 h after ingestion); and remote.

Examples of these symptoms are as follows:

1. immediate gastrointestinal symptoms, e.g., lip and tongue swelling, mouth tingling, vomiting, abdominal pain
2. late gastrointestinal symptoms, e.g., diarrhea, bloating, constipation
3. remote symptoms, e.g., anaphylaxis, angiedema, rhinitis, urticaria, eczema, asthma, joint pains, headaches, depression, irritability

Diagnosis

The diagnosis of food intolerance and specifically of food allergy is not straightforward in most cases because of the nonspecific nature of many of the symptoms. The first stage is to take a careful medical and family history. A positive family history of atopy increases the likelihood of the diagnosis of allergy in an individual.

The timing of development of reactions to foods is also an important part of the diagnosis. The time course of some common symptoms of food intolerance is shown in Figure 1.

A reaction that may involve swelling of the lips and tongue, vomiting, rhinorrhea, urticaria, and asthma within minutes of ingesting a food is evidence of an immunological mechanism. When the symptoms occur more than an hour after the food is consumed, or only after repeated exposure to the food, an immune basis for the reaction is less frequently found.

Laboratory investigations may be used to confirm the diagnosis, especially when food allergy is suspected. The immediate type of allergic response is frequently associated with increased blood levels of IgE and the presence of IgE antibodies to food proteins. Skin prick tests and the radioallergosorbent test (RAST) are used to identify the presence of IgE. The procedure for RAST is as follows: a sample of blood is placed on paper impregnated with purified antigen; the amount of antibody that interacts is estimated by

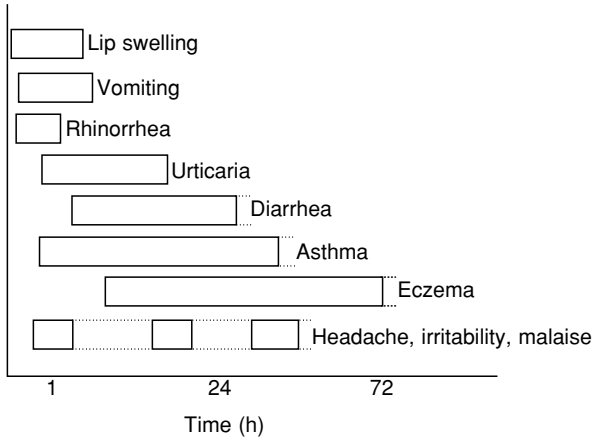

Figure 1 Symptoms and time course of food intolerance. From Lessof MH (1983) Reactions to food in adults. In: Lessof MH (ed.) *Clinical Reactions to Food*, p 112. Chichester: John Wiley.

the use of radioactive markers. These tests are widely used in the diagnosis of allergic disease which is not related to food. The results are likely to be positive in cases of immediate reactions to food, but general use of these tests in the diagnosis of food allergy is controversial because of difficulties in obtaining pure food proteins for testing and frequent false-positive responses.

The measurement of blood levels of histamine, of histamine release, and of other mediators from white blood cells has been investigated as another diagnostic technique. As yet, the results have been too variable for these tests to become part of routine diagnoses.

The only sure diagnostic approach in food intolerance is the use of some form of elimination diet. The nature of the diet will depend on the clinical symptoms and provisional diagnosis. This may simply involve the elimination of individual suspect foods, or the exclusion of all foods commonly associated with adverse reactions. These include milk products, eggs, wheat products, fish, nuts, pork, chocolate, coffee, tea, citrus fruits, and foods containing preservatives and coloring agents. In some cases an extremely restricted or 'oligoantigenic' (few foods) diet of lamb, rice, and pears is used.

If the symptoms disappear over the course of the diet, the patient is 'challenged' with the suspect food or ingredient. If there are many foods involved, these must be reintroduced slowly, one at a time and over a period of weeks or months, to see if the symptoms recur. In some cases it is necessary to confirm the diagnosis by a double-blind challenge, in which the suspect food is given in disguised form so that both the patient and doctor or dietitian are unaware of what has been ingested.

There are also numerous unorthodox diagnostic tests for which extravagant claims are made. In general these have been shown to be of little value in diagnosis and in some cases to be misleading and potentially dangerous because the false diagnoses resulting from their use can lead to inappropriate treatments. These include the pulse test, the sublingual test, the interdermal skin test, cytotoxic tests, and hair sample tests.

Treatment

Once the food or range of foods or ingredients has been identified as the cause of intolerance, a diet which avoids these foods must be followed. If this only involves the removal of a few foods which are 'nonessential', e.g., strawberries or shellfish, then this is straightforward. Where important groups of foods must be excluded, such as cereals or milk products, the advice of an experienced dietitian is essential and supplements may be necessary to insure nutritional adequacy of the diet, particularly in children.

Various drug treatments such as antihistamines and antiinflammatory drugs may be used as an adjunct to the diet, depending on the severity of the condition.

Predisposing Factors to Food Intolerance

There is still much to be learnt about the predisposing factors to food intolerance. It is established that enzyme defects are inherited and that the predisposition to the true food allergy is also inherited. Food allergies are certainly more common in the offspring of atopic parents. In some individuals, more than one type of intolerance may coexist. For example, a person might display an allergic (IgE-mediated) reaction to certain foods and a pharmacological sensitivity to other foods. Much more research is needed to explain the causes of food intolerance and to establish what factors, both genetic and environmental, determine an individual's susceptibility.

Prevalence

Assessment of the prevalence of food intolerance is also difficult, and wide ranges of incidence have been reported in the scientific literature. This is not surprising as many of the estimates are based on selected populations and the diagnostic criteria and testing procedures used in different studies vary considerably.

When adverse reactions to food occur, intolerance rather than true allergy is usually the problem. With the exception of the very rare inborn errors of metabolism, which are usually diagnosed in the neonatal period, there are few reliable estimates of the prevalence of nonallergic reactions.

Food additive intolerance has been studied more carefully and it is estimated that between 0.01% and 0.23% of the general population in the UK may have adverse reactions to some food additives. The incidence of such reactions rises dramatically in sensitive populations, such as those suffering from chronic urticaria. Reports in the scientific literature suggest that as many as 30–50% of such populations may display sensitivity to certain additives.

Estimates of the prevalence of food intolerance in the general population are no more than 5–6% and an estimated 2% of the population may have IgE-mediated food allergies, whereas as many as 20–30% of the UK population perceive themselves to be food-intolerant, or in common parlance, 'allergic' to food. However, as yet there are no systematic data which enable the accurate calculation of the incidence

and prevalence of adverse reactions to food, either in the UK or elsewhere.

Adverse reactions to food are more frequently reported in infants and young children, with varying prevalence estimates from 0.3% to 20% of children. However, the majority of such reactions resolve themselves in childhood and certainly by adolescence. The prevalence of IgE-mediated reactions is higher in young children than adults, with an estimate of about 5%.

It has been suggested that inhalant and possibly food allergies have increased in prevalence, but the swiftness with which apparent change has come about precludes a change in the gene pool being responsible. There has been an increase in the numbers of people suffering from food-induced anaphylaxis; the reason for this is uncertain.

See also: **Anorexia Nervosa**; **Bulimia Nervosa**; **Caffeine**; **Fish**: Spoilage of Seafood; **Food Additives**: Safety; **Histamine**; **Inborn Errors of Metabolism**: Overview

Further Reading

Anon (1987) Medical aspects of food intolerance. *Journal of the Royal College of Physicians of London* 21: 237–266.

British Nutrition Foundation (2002) *Adverse Reactions to Food. Report of the British Nutrition Foundation Task Force*. Oxford: Blackwell Science.

Food Standards Agency (2000) *Adverse Reactions to Food and Food Ingredients. Report from the Committee on Toxicity of Chemicals in Food, Consumer Products and the Environment*. London: TSO.

Lessof MH (1992) *Food Reactions. Royal College of General Practitioners Members' Reference Book* 367–371. London: Royal College of General Practitioners.

Food Allergies

J Gray, Guildford, Surrey, UK
W Chan, London, UK

Introduction

Allergic disorders such as asthma, hayfever, and eczema are common. Adverse reactions to food are also common, particularly in children, but it is recognized that many such reactions are not true allergy, i.e., they are not caused by an adverse response of the body's immune system.

This article examines the specific problem of food allergic disease, including the types of immune reactions and the mechanisms believed to be responsible for them, clinical manifestations of food allergy, genetic and immunological predisposition, prevalence, and the vulnerability of newborn infants to the problem.

The Immune System

The immune system is capable of distinguishing between proteins and other molecules that are part of the body and those that are foreign to it. It is able to protect against harmful foreign particles such as bacteria, viruses, and parasites by generating a response to eliminate them. One of the ways it does so is by producing antibodies against the foreign particle. An antibody is a protein (immunoglobulin) that specifically combines with a foreign substance, referred to as an antigen or allergen. The immune system is also able to ignore harmless foreign matter; otherwise it would be in a constant state of overactivity. (*See* **Allergens**.)

However, in some people there is an abnormality of the immune system so that it produces antibodies in response to harmless and sometimes beneficial molecules, usually proteins, glycoproteins, or polysaccharides. When the antigen is encountered subsequently, an adverse reaction with unpleasant and occasionally dangerous symptoms occurs. Someone whose immune system reacts in this way is said to be allergic. Almost any food, food constituent, food additive, drug, or other substance which is swallowed, such as dust and microorganisms, can act as an antigen within the gut.

The cells which are involved in the immune response are organized into tissues and organs which are collectively referred to as the lymphoid system. This comprises lymphocytes, accessory cells (macrophages and antigen-presenting cells) and, in some tissues, epithelial cells. The lymphoid tissues and organs are described as primary (central) and secondary (peripheral). The primary lymphoid organs, which are the major sites of lymphocyte production, comprise the thymus and bone marrow. The secondary lymphoid organs, which provide an environment for lymphocyte and accessory cell interaction with antigens, comprise the spleen, lymph nodes, and the mucosa-associated lymphoid tissues (MALT). The MALT include the tonsils, and all the tissues associated with the mucosal surfaces of the respiratory tract, of the intestinal tract, referred to as the gut-associated lymphoid tissue (GALT), and including the Peyer's patches of the ileum, and of the genitourinary tract. More than 50% of the body's lymphoid tissues is associated with the mucosal system, especially the GALT, as this is the point of entry for external antigens. The GALT is particularly important in the

context of food allergy because food antigens exert their main immunological effects locally in the gut, although they may also enter the body through the nasal and oral mucosal surfaces, or through the lower respiratory tract and skin.

An immune response can involve various cell types: the lymphocytes, which recognize pathogens or antigens, and the phagocytes, which engulf foreign material and degrade it. The lymphocytes can be divided into two main categories: the B lymphocytes (bursa- or bone marrow-dependent cells), which release antibodies; and the T lymphocytes (thymus-derived cells), which have a wide range of activities. Some T cells interact with the B cells, helping them to make antibody, and another group interacts with phagocytes; these are termed T helper cells (TH). A third group of T cells is responsible for killing infected host cells: they are known as T-cytotoxic cells (TC). The T cells produce soluble proteins (cytokines) which also have regulatory effects in the immune response, acting as signals between lymphocytes, phagocytes, and other cells of the body. The cytokines include the interferons (IFNs) and interleukins (ILs) and tumor necrosis factors (TNFs). The immune response can also involve other soluble mediators, including the complement system, which is a group of about 20 serum proteins that interact and control inflammatory reactions.

Mechanisms of Allergic Reactions

Immune response are frequently classified into four main types, referred to as types I–IV. Types I–III are antibody-mediated, with type I effected by immunoglobulin E (IgE) antibodies; type IV is mainly mediated by T cells and macrophages. In practice, more than one of these types of reactions may occur in an individual at the same time.

Food allergic reactions may be IgE (type I) or non-IgE-dependent, but most common food allergies are mediated by IgE antibodies. IgE-mediated reactions are immediate. This immunoglobulin is normally present in very small quantities, but it is produced in large amounts in some people who are described as 'atopic.' Such people are more prone to allergic reactions such as hayfever, asthma, eczema, urticaria, and food allergy. On the first encounter with an antigen, there is a complex series of events leading to the release of IgE antibodies from the B cells, with assistance from T cells.

The structure of an antibody is shown in Figure 1. It contains a fragment antigen-binding portion (Fab) which is shaped to bind to a specific antigen and a fragment crystallizable portion (Fc) which binds to white blood cells and may also activate complement.

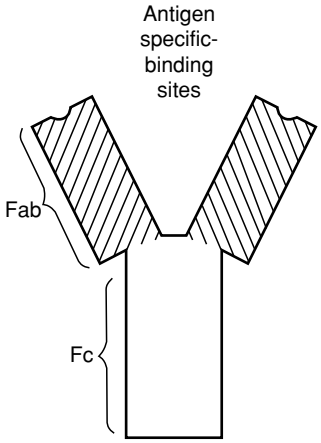

Figure 1 The antibody consists of two portions: the fragment antigen-binding (Fab) portion, shaped to bind a specific antigen, and the fragment crystallizable (Fc) portion, which binds to white blood cells and complement proteins. Reproduced from Food Intolerance: Food Allergies. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

The antigen-specific IgE molecules bind to mast cells, which are granular connective tissue cells containing histamine and other chemical mediators. These cells are present in large numbers in mucosal surfaces of the skin, nose, mouth, eye, respiratory tract, and gut. IgE binds to the mast cells via Fc receptors and the individual is then said to be sensitized. On the next encounter, the antigen interacts with the specific IgE antibodies on the sensitized mast cells, causing an explosive degranulation of the mast cells, with the release of histamine and other mediators, such as serotonin and various cytokines, including IL-3 and IL-4. These mediators act on blood vessels, mucous glands, and smooth muscle, causing the symptoms of allergy. The sequence of events is shown in Figure 2. (*See* **Prostaglandins and Leukotrienes**.)

Delayed reactions to food may take several hours or days to develop and are complex. They are usually not accompanied by an increase in IgE antibodies. Celiac disease is the best defined example of this type of response. (*See* **Celiac (Coeliac) Disease**.)

Clinical Features of Food Allergy

The clinical manifestations of food allergy are variable in location, timing, and severity. Symptoms range from mild and tolerable, e.g., slight abdominal pain, to anaphylaxis – a rapid and extreme generalized allergic response which can result in physical collapse and death. The main organ systems affected are the gastrointestinal tract, skin, and respiratory tract. The clinical features may be influenced by

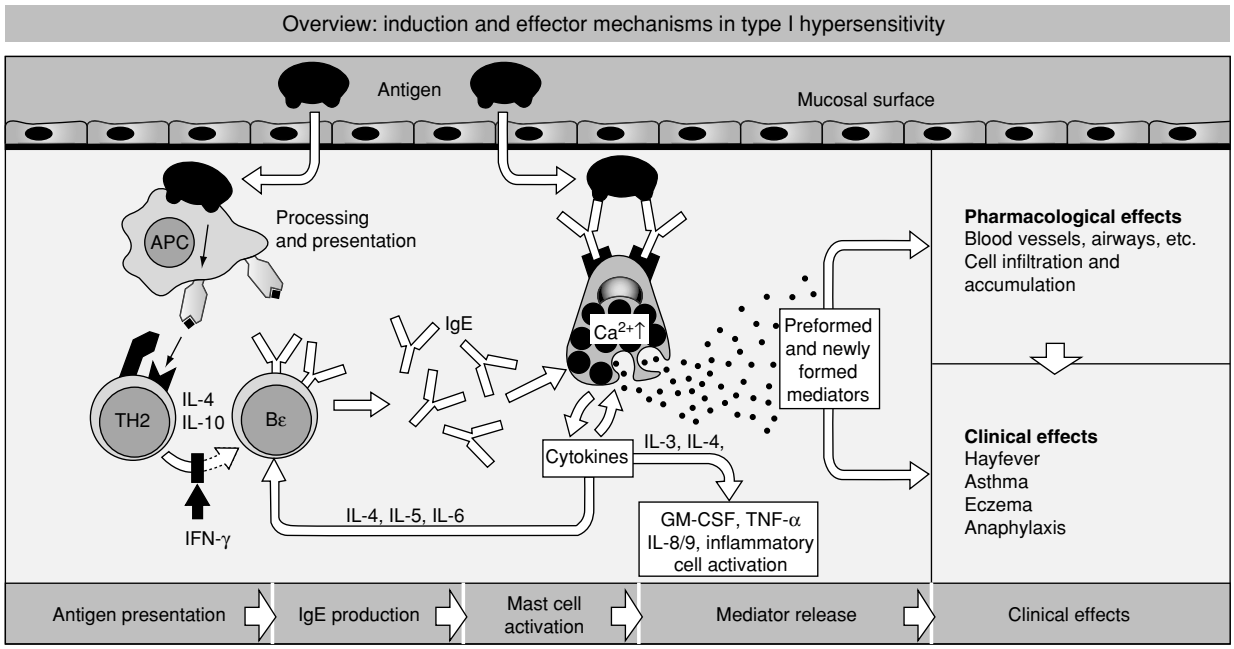

Figure 2 Innocuous environmental antigens (allergens) enter via mucosal surfaces and are taken up by local antigen-presenting cells (APCs), which process and present them to TH cells. TH2 cells secrete cytokines that induce B-cell proliferation and favor the production of an allergen-specific immunoglobulin E (IgE) response. The IgE binds, via Fcε receptors (FcεRI), to mast cells, thus sensitizing them. When allergen subsequently reaches the sensitized mast cell, it cross-links surface-bound IgE, causing an increase in intracellular calcium (Ca^{2+}) that triggers the release of preformed mediators, such as histamine and proteases, and newly synthesized, lipid-derived mediators such as leukotrienes and prostaglandins. These autocolds produce the clinical symptoms of allergy. Cytokines are also released from degranulating mast cells and may augment the inflammatory and IgE responses. IL-4, interleukin-4; IFN-γ, interferor-γ; GM-CSF, granulocyte – macrophage colony-stimulating factor. Adapted, with permission, from Roitt I, Brostoff J and Male D (1998) *Immunology*, 5th edn. Sidcup: Mosby.

other factors such as age, physical exercise, stress, and coexisting medical problems.

Gastrointestinal Tract

Swelling and tingling of the lips, tongue, and throat can present as initial symptoms of food allergy occurring as the antigen comes into contact with these organs. Nausea and vomiting are common acute symptoms, and abdominal pain and diarrhea also occur frequently.

The pathological changes in the small intestine which occur in food allergy have been characterized from biopsy specimens. It is well documented that mucosal damage can occur in cows' milk protein allergy and there is evidence that reactions to soya, egg, fish, and chicken may also result in mucosal damage.

More extensive intestinal mucosal damage is seen in celiac disease in response to the ingestion of gluten. Although abnormal immunological mechanisms are believed to be partially responsible for this condition, the exact mechanism remains uncertain. (*See* **Celiac (Coeliac) Disease**.)

Food allergy can also lead to inflammatory changes in the large bowel and bloody diarrhea (proctocolitis) in infants. However, in adults there is a lack of evidence to suggest an immunological basis for large-bowel disorders. (*See* **Colon: Diseases and Disorders.**)

Skin

Urticaria (an itchy raised skin rash also known as nettle rash) and angiedema (swelling of the skin, mucous membranes, and underlying tissues) are frequently seen as manifestations of food allergy. There is increasing evidence that food intolerance may be a cause of eczema but the basic mechanism of the reaction is still unclear. Infantile eczema is certainly more common in children from atopic families. Such reactions frequently occur in response to cows' milk protein and may be ameliorated by dietary manipulation.

Respiratory Tract

Sneezing and rhinorrhea (running of the nose) can be manifestations of food allergy, either in association with gut or skin symptoms or alone. More rarely, bronchoconstriction occurs when a food is ingested, but food allergy is not a common cause of asthma.

Anaphylactic Shock

Although the majority of food-allergic reactions are not life-threatening, a small proportion of reactions

can lead to anaphylactic shock (anaphylaxis), which is severe and may be fatal. Anaphylaxis is an acute allergic (IgE-mediated) reaction of rapid onset and affecting many organ systems, including the cardiovascular system, the respiratory tract, the mouth, the pharynx, and the skin, either individually or in combination. It is characterized by swelling of the lips and throat, shortness of breath, a rapid fall in blood pressure, progressing to cardiovascular collapse and severe shock. Treatment consists of intramuscular injection of epinephrine (adrenaline), and without this anaphylaxis may cause death. Anaphylactic reactions are relatively rare, but there is evidence that deaths from food-induced anaphylaxis are increasing. In the UK and Europe, peanut is the most common food reported to cause anaphylaxis, but other foods that are known sometimes to cause the condition are tree nuts, seeds, eggs, milk, and shellfish. Inadvertent exposure to potential allergens in foods containing, or contaminated with, minute amounts of the allergen poses a major risk to the small number of people at risk of anaphylaxis.

Predisposition to Food Allergies

It is believed that genetic predisposition and immunological predisposition, notably elevated serum IgE levels, are the most important determinants of allergic disease. Environmental factors are secondary but may still play an important role; they include nondietary factors such as the level of allergen exposure, which may be influenced by season of birth, maternal smoking habit, exposure to pollutants, viral illness, and the nutritional status of the individual.

The importance of genetic factors in the etiology of allergic disease and IgE production is evident from the simple observation that allergic disease is more common in the children of atopic parents. Children from families where both parents are atopic have more than a 50% chance of developing an allergic disorder; where one parent is atopic, the chance is about 30%. However, twin studies have shown that concordance is far less than 100% in monozygotic (identical) twins, indicating that environmental factors are also important in the expression of atopic disease.

Research in atopic families and twins suggests that total IgE levels are influenced by genetic factors. The major control of antigen-specific IgE responses is believed to be linked to the human leukocyte antigen (HLA) immune response genes. However, HLA-linked genes do not appear to be responsible for high IgE levels. There is evidence that general levels of IgE may be determined by other genes, including those responsible for cytokine production.

Inherited immune deficiency disorders, e.g., deficiencies of complement or IgA, can also predispose towards allergy. The presence of a specific type of IgA, secretory IgA, in the gut is believed to reduce antigen absorption. It might be deduced that IgA deficiency could lead to increased antigen absorption and IgE antibody stimulation. However, again the explanation must be more complex, as not all such people develop allergy.

As ingested food travels through the gut, it presents the immune system with a vast array of foreign particles to distinguish. The intestinal mucosa itself is a barrier which can protect against dietary antigens and the secretion of mucus from intestinal goblet cells and the production of secretory IgA are protective mechanisms which help to reduce antigen absorption. Disruption of the intestinal barrier, as occurs after intestinal infection, and increased permeability of the barrier, as in premature infants, the newborn, in malnutrition, and in immune deficiency, may increase the penetration of antigens.

The significance of increased gut permeability in the etiology of food allergic disease is nevertheless uncertain. It has been demonstrated that antigens can cross the gut wall in both allergic and nonallergic individuals, and that immune complexes may circulate in the blood of nonallergic individuals without ill effect. This underlines the complexity, as well as the gaps in our understanding, of the ways in which genetic and environmental factors may interact to cause allergic disorders.

Vulnerability of the Newborn to Food Allergy

It is evident from clinical experience that the newborn is vulnerable to food allergens. For example, there is clear evidence that infants exposed to unmodified cows' milk within the first few weeks of life are particularly susceptible to the development of symptoms of allergy to cows' milk such as eczema. However, our knowledge of the pathogenesis of food allergic disease and the relative importance of the various contributory factors in the neonatal period is still limited.

The immunological response to feeding a 'foreign protein' may result from the immaturity of neonatal digestive and absorptive functions and early penetration of the antigens. It has also been suggested that because newborn infants are temporarily deficient in IgA – secretory IgA does not appear until the age of 2–4 weeks – the gut mucosa is particularly vulnerable to antigen penetration.

The influence of maternal diet during pregnancy and lactation, the protective role of breast-feeding,

and the timing and introduction of solids at weaning may all play their part in the pathogenesis of food allergic disease. The coincident occurrence of infection with the introduction of a specific antigen may also be a crucial factor. (*See* **Pregnancy**: Safe Diet.)

The idea that breast-feeding might protect against allergic disease has existed for some 50 years. Although there are many studies that support this idea, there are many that do not. Such conflict appears to result from the difficulty in controlling studies on infants and differences in methodology. (*See* **Infants**: Breast- and Bottle-feeding.)

On balance, it appears that breast-feeding decreases the likelihood of allergic disease in children who are 'at risk' because of family history. To some extent this is due to the late introduction of foreign protein but it is also attributable to the presence of protective immunological factors, such as IgA, in colostrum and breast milk. Breast milk protects against infection, and breast-feeding should help to maintain the integrity of the gastrointestinal mucosa in the neonatal period.

The level of total IgE in umbilical cord blood is believed to be predictive of allergic disease. The influence of maternal diet on cord blood immunoglobulin levels and the placental transfer of antibodies to specific food proteins in atopic and nonatopic mothers are subjects of active research at the present time. As yet the influence of maternal diet on allergic outcome in the offspring remains uncertain.

The current view is that sensitization *in utero* does occur. However, the influence of dietary manipulation of the mother in pregnancy in preventing food allergy in the infant is contentious and in need of further research before it can be recommended.

Food antigens pass into breast milk in small amounts and these are significant in immunological terms. It is believed that this may lead to sensitization in susceptible infants. Therefore the exclusion of specific foods from the mother's diet during lactation may be beneficial for at-risk infants. Again, this is the subject of research.

The potential severity of the reactions associated with peanut allergy has led to the recommendation, in the UK, that pregnant and breast-feeding women from atopic families (whether they are themselves atopic or if the father or any other sibling has atopic disease) should avoid peanuts and peanut products in order to reduce the risk of peanut allergy developing in their offspring. However, the scientific basis for this recommendation is still open to debate and other countries have not made similar recommendations. (*See* **Food Intolerance**: Types.)

Particularly important for at-risk infants is the timing of weaning and the nature of the foods introduced early on. Exclusive breast-feeding, for at least 4 and preferably 6 months, and avoidance of the early introduction of foods commonly known to cause allergies, e.g., cows' milk and its products, eggs, wheat, fish, nuts, and citrus fruits, would seem to be the best preventive approach. However, more research is needed on the links between weaning and subsequent development of food allergic disorders. (*See* **Infants**: Weaning.)

Prevalence of Food Allergies

It is difficult to arrive at a precise estimate of the prevalence of food allergic disease because of the different testing procedures and diagnostic criteria used in research studies. Figures ranging from 1% to 20% have been cited, but the true prevalence is currently estimated to be towards the lower end of this range, at about 2%. Our present state of knowledge suggests that food allergy is far less common than food intolerances (with a nonimmunological basis), and less common than allergic reactions to pollen and house dust. However, as yet there are no systematic data to permit the accurate calculation of the incidence and prevalence of food allergic reactions, either in the UK or elsewhere.

See also: **Allergens**; **Colloids and Emulsions**; **Colon**: Diseases and Disorders; **Food Intolerance**: Types; **Immunology of Food**; **Infants**: Breast- and Bottle-feeding; Weaning; **Pregnancy**: Safe Diet; **Prostaglandins and Leukotrienes**

Further Reading

British Nutrition Foundation (2002) *Adverse Reactions to Food. Report of the British Nutrition Foundation Task Force*. Oxford: Blackwell Science.

Cant AJ (1991) Food allergy and intolerance. In: Mclaren DS *et al*. (eds) *Textbook of Paediatric Nutrition*, 3rd edn, pp. 204–221. London: Churchill Livingstone.

Food Standards Agency (2000) *Adverse Reactions to Food and Food Ingredients. Report from the Committee on Toxicity of Chemicals in Food, Consumer Products and the Environment*. London: TSO.

Roitt I, Brostoff J and Male D (1998) *Immunology*, 5th edn. Sidcup: Mosby.

Milk Allergy

S L Taylor and J Kabourek, University of Nebraska, Lincoln, NE, USA

Introduction

Cows' milk is one of the most common causes of adverse reactions in foods and can occur through several mechanisms. True cows' milk allergy (CMA) occurs through an immunoglobulin E (IgE) mediated mechanism. The prevalence of allergy to cows' milk is reported to be about 2% of the general pediatric population in developed countries. Treatment of CMA involves the strict avoidance of cows' milk and products containing cows'-milk proteins. Education of both the affected child and parents is also vital to successful management of CMA.

Milk Allergy

Cows' milk is one of the most common causes of adverse reactions in foods and can occur through several mechanisms. True CMA or hypersensitivity is indicative of an adverse reaction to bovine milk proteins that produces symptoms and where an immunologic mechanism may be implicated. CMA is mainly an IgE-mediated, type I immediate hypersensitivity reaction, although type III and IV reactions have also been implicated. CMA is considered the most severe form of intolerance because the threshold dose for elicitation of a reaction is low, and symptoms can be severe in some affected individuals.

Classified by Coombs and Gell, hypersensitivity reaction mechanisms are divided into four types: I–IV. The first three types are antibody-mediated, whereas the fourth type is mediated by T cells and macrophages.

Type I (anaphylactic) reaction is immediate and is due to interaction of antigen (allergen) with IgE-sensitized mast cells and basophils, resulting in histamine and mediator release from these cells.

Type II (cytotoxic) reaction involves both IgG and IgM antibodies. A specific cell or tissue is targeted, activating the complement cascade and causing cytotoxic cells to damage membranes.

Type III hypersensitivity (immune complex) involves an antigen–antibody complex, causing an inflammatory response with fixation of complement.

Type IV (cell-mediated) hypersensitivity involves T cells producing cytokines that recruit phagocytes and other inflammatory cells. Continued stimulation of these cells causes tissue damage.

In true IgE-mediated CMA, exposure of susceptible individuals to one or more of several cows'-milk allergens (naturally occurring proteins) elicits the production of allergen-specific IgE antibodies. These IgE antibodies become fixed to the membranes of mast cells in the tissues and basophils in the blood. This symptomless process is called allergic sensitization. Upon subsequent exposure to cows' milk, the specific cows'-milk allergen cross-links two IgE antibodies on the membrane surface of the sensitized mast cells and basophils. This interaction triggers the release of a host of bioactive mediators from the mast cells and basophils, including histamine, leukotrienes, and prostaglandins. These mediators are responsible for the range of symptoms experienced in an allergic reaction.

Cows'-milk protein intolerance (CMPI), not to be confused with IgE-mediated CMA, is generally a nonimmune-mediated reaction against cows'-milk proteins. CMPI is probably a cell-mediated inflammatory reaction that damages the absorptive function of the intestinal tract. Symptoms are primarily gastrointestinal (nausea, abdominal cramps, and diarrhea) and delayed, appearing several days after milk ingestion. Generally, CMPI is a disorder occurring in infancy that is frequently outgrown.

Lactose intolerance refers to inadequate digestion of lactose, the major carbohydrate found in milk, usually from a deficiency or low levels of β-galactosidase (lactase) enzyme. Lactose intolerance produces gastrointestinal symptoms, including cramps, bloating, gas, diarrhea, and/or nausea. In contrast to CMPI and CMA, lactose intolerance primarily affects adults.

Composition

Cows' milk is a nutrient-dense food consisting of varying amounts of carbohydrate, fat, and protein. The major constituents of cows' milk are water (87.4%) and milk solids (12.6%), which includes vitamins, minerals, carbohydrate, fat, and protein. Cows' milk is known to contain many nutrients, including varying concentrations of fat and water-soluble vitamins, minerals, trace elements, and salts. The principal carbohydrate in milk is lactose. The lactose content of milk varies by species. Cows' milk contains about 4.8% lactose (12–12.5 g lactose per cup), whereas human milk has 7% lactose. In addition to lactose, minor quantities of glucose, galactose, and oligosaccharides are present. Milk fat is a complex lipid existing as an oil-in-water emulsion. Milk lipids are mainly triacylglycerols or fatty acid esters with glycerol, of which more than 400 have been identified. Phospholipids,

sterols, waxes, and free fatty acids have also been identified.

Cows' milk is a heterogeneous mixture of at least 20 proteins. Eighty percent of the total protein is casein proteins and 20% is whey proteins. Caseins are relatively heat-stable, whereas whey proteins are more labile. Casein can be fractionated into four major components: alpha, beta, kappa, and gamma casein. The whey proteins consist of α-lactalbumin, β-lactoglobulin, bovine serum albumin, and several minor proteins, including lactoferrin and lactoperoxidase.

Great variability of the human IgE response to cows' milk exists. Many patients are allergic to more than one milk protein. The major allergens of cows' milk are α-$_{s1}$ casein and β-lactoglobulin. α-Lactalbumin is also involved in some patients. IgE antibodies directed to minor cows'-milk proteins, such as bovine serum albumin and lactoferrin, have been identified in a few patients but their role in CMA is not firmly established.

Prevalence

CMA and CMPI are primarily diseases of infancy. CMA has become one of the most common food allergies in early childhood since cows' milk is usually one of the first foreign antigens encountered in large quantities in infancy. Most infants develop symptoms before 1 month of age, often within 1 week of introduction. Onset of CMA after 1 year of age is extremely rare. Major risk factors for CMA are a positive family history of CMA or other atopic disease and early cows'-milk protein exposure. The onset of disease in most cases is closely related to the time of introduction of cows'-milk products. The prevalence of allergy to cows' milk is reported to be about 2% of the general pediatric population in developed countries. CMA occasionally occurs in adults, affecting less than 1% of the adult population; however, good prevalence estimates for CMA in adults are not available.

Prevention

Because cows' milk is usually one of the first foreign antigens encountered in infancy, preventive measures are ideally employed during this time period, particularly in children of atopic parents. Although conflicting data exist and no single prevention method is completely effective, protective effects have been seen with breast-feeding and delaying introduction of cows' milk.

Prolonged breast-feeding in all infants is the ideal source of nutrition, providing immunological factors, developing intestinal maturation, and enhancing infant–mother bonding. The protective effect of breast-feeding may be due, in part, to avoidance of high doses of cows'-milk proteins or an inherent protective effect of human milk. Support for the benefits of breast-feeding in preventing or delaying the development of CMA has been seen in prolonged breast-feeding practice of at least 6 months. In addition to breast-feeding, delaying the introduction of cows' milk until 6 months of age also has a preventive effect. If breast-feeding is not possible or during weaning, an extensively hydrolyzed formula in combination with delayed milk introduction may be protective. Partially hydrolyzed infant formulas (PHF) are also available. These formulas appear beneficial in the prevention of sensitization to cows' milk and are more palatable than extensively hydrolyzed formula (EHF). But the partial hydrolysates can provoke allergic reactions in previously sensitized infants. There may be a preventive effect of maternal avoidance of common allergenic foods during lactation but there has been no protective effect shown for an altered diet during pregnancy. Occasionally, CMA seems to develop in exclusively breast-fed infants, ostensibly from exposure to small amounts of cows'-milk proteins secreted into the breast milk and arising from the mother's diet. However, cows' milk is so important in the nutrition of lactating women that avoidance of cow's milk during the lactation period is not often recommended. Similarly, the practice of cows' milk avoidance during pregnancy should not be suggested because it may put the health of the mother as well as the fetus in jeopardy of nutritional complications or deficiencies.

Symptoms/Manifestation

Clinical manifestations and severity of symptoms of CMA vary widely from one individual to another. Symptoms arise involving the skin, gastrointestinal or respiratory tracts. The majority of patients have greater than two symptoms affecting greater than two organs/systems. CMA in individuals exhibiting gastrointestinal symptoms and/or dermatitis usually resolves early in life, whereas individuals with urticaria, angioedema, respiratory symptoms, or anaphylaxis may persist with CMA into adult life.

Individuals with immediate reactions often exhibit systemic anaphylaxis and/or dermatological symptoms. Anaphylaxis is a systemic, multiorgan reaction associated with IgE-mediated hypersensitivity. It may progress slowly or rapidly and can involve hypotension, bronchospasm, edema, and cutaneous symptoms. Treatment for anaphylaxis involves the use of

epinephrine (adrenaline) given intramuscularly, antihistamines, and steroids. Although severe reactions to cows' milk are rare, deaths have been documented. Dermatological features may include urticaria, angioedema, and eczema (dermatitis).

Intermediate reactions occurring from 1 to 24 h after ingestion may involve gastrointestinal symptoms, including vomiting, diarrhea, reflux, colic, and bleeding.

Symptoms appearing more than 24 h after cows'-milk ingestion often affect the respiratory tract and skin. Respiratory symptoms may involve nasal stuffiness, coughing, wheezing, tachypnea, and rhinitis. Atopic dermatitis is a rather common manifestation of CMA involving the skin; it can be a chronic condition.

Diagnosis

Differentiation between CMA and CMPI cannot be based on clinical symptoms alone, nor is there a lone clinical test. The first step in evaluation of someone with a history of possible adverse reactions is to take a detailed food history. Along with identifying the suspected food, the quantity of food needed to produce symptoms and the frequency with which reactions occur need to be reported. Physical exams may be helpful in revealing a chronic skin condition such as atopic dermatitis. However, other physical signs, such as respiratory symptoms, may be present but are not necessarily a result of food hypersensitivity and may be attributed to other conditions. Gastrointestinal complaints, in particular, can have numerous causes.

Diagnosis is usually based on cows'-milk protein elimination and challenge tests. Open, controlled cows'-milk challenges are recommended for infants 0–1 year. In children older than 1 year and adults, double-blind placebo-controlled challenges (DBPCFC) are performed. These tests determine if cows' milk is the cause of symptoms but do not identify the mechanism.

Several methods, including skinprick test and radioallergosorbent test (RAST), are available as diagnostic tools in identifying immunological-mediated CMA. Recent advances in RAST have improved the diagnostic value of these tests and lessened the need to perform DBPCFCs. Milk-specific IgE is seen in some milk-hypersensitive individuals but not all. CMPI may be suspected if symptoms are restricted to the gastrointestinal tract and have delayed onset, especially in cases where specific IgE antibodies are absent.

Presently, allergic sensitization to cows'-milk proteins usually occurs during the first few years of life, often disappearing thereafter. The majority of children outgrow CMA and CMPI and will acquire tolerance, with an estimated remission rate of 85–90% before 3 years of age. However, an early increased IgE response to cows' milk proteins is associated with an increased risk of persisting allergy to cows'-milk proteins, development of asthma or rhinoconjuctivitis, or development of atopic disorders later in life.

Treatment/Management

Treatment of CMA and CMPI involves the strict avoidance of cows' milk and products containing cows'-milk proteins. Because even a minute amount of allergen has the potential to trigger sensitization and/or symptoms, awareness is key in treatment. Breast milk is the ideal source of nutrition. If breastfeeding is not possible, numerous formulas are available as cows'-milk substitutes.

Soy protein formulas can be well tolerated by some individuals. However, soy is also a commonly allergenic food and has been reported to cause symptoms in some CMA individuals due to the development of soy allergy.

Meat-based formulas from sources such as chicken or lamb may be used but have also been shown to cause symptoms on rare occasions. Also, a formula based on goats' milk contains β-lactoglobulin and often causes symptoms in CMA individuals.

More commonly, modified-protein formulas are the treatment of choice. Protein hydrolysates are classified as PHF or EHF and may be derived from casein or whey. For the most part, however, these EHFs are generally tolerated and are a good alternative for CMA management. A few CMA infants will respond adversely to trace amounts of milk proteins or peptides present in EHF.

PHF were developed for the purpose of dietary prevention of CMA and are not recommended for individuals who already have CMA. EHF have been in use for over 50 years as a therapeutic diet in CMA management and are the recommended treatment. However, palatability of EHF is an issue, and EHF can be quite costly. In severely sensitive individuals, hypoallergenic, nutritionally complete infant formula composed of individual amino acids (elemental formula) may be the only effective substitute.

Hypoallergenic formula or breast-feeding should be the recommended dietary options for infants with CMA for the first 6 months of life. From 6 months of age, solid foods may be introduced. Generally, highly allergenic foods such as egg, soy, peanut, fish, and wheat are also avoided until the age of 1 year.

Reintroduction of cows'-milk protein may be performed in 3–6-month intervals after the age of 1 year; with infants having severe CMA, pediatric advice should be sought before attempting reintroduction.

Education of both the affected child and parents is vital to successful management of CMA. Both the physician and dietitian can provide guidance and recommend ways for introducing foods in childhood. This helps guarantee a cows'-milk protein-free diet as well as a nutritionally complete diet. Dietitians are also very useful in teaching how to read labels on food packages and identifying milk-derived products.

See also: **Food Intolerance**: Types; Food Allergies; Lactose Intolerance; Elimination Diets; **Infant Foods**: Milk Formulas

Further Reading

American Academy of Pediatrics Committee on Nutrition (2000) Hypoallergenic infant formulas. *Pediatrics* 106: 346–349.

Exl B-M, Müller-Teicher G and Vandenplas Y (2000) Preventative possibilities within the context of cow's milk allergy. *Allergy and Clinical Immunology International* 12: 68–76.

Hosking CS, Heine RG and Hill DJ (2000) The Melbourne milk allergy study – two decades of clinical research. *Allergy and Clinical Immunology International* 12: 198–205.

Novembre E and Vierucci A (2001) Milk allergy/intolerance and atopic dermatitis in infancy and childhood. *Allergy* 56(suppl. 67): 105–108.

Pelto L, Salminen S, Lilius E-M, Nuutila J and Isolauri E (1998) Milk hypersensitivity – key to poorly defined gastrointestinal symptoms in adults. *Allergy* 53: 307–310.

Savilahti E, Kuitunen P and Visakorpi JK (1981) Cow's milk allergy. In: Lebenthal E (ed.) *Textbook of Gastroenterology and Nutrition in Infancy*, pp. 689–708. New York: Raven Press.

Sicherer SH (2000) Determinants of systemic manifestations of food allergy. *Journal of Allergy and Clinical Immunology* 106: S251–S257.

Sprikkelman AB, Heymans HS and Van Aslderen WMC (2000) Development of allergic disorders in children with cow's milk allergy or intolerance in infancy. *Clinical and Experimental Allergy* 30: 1358–1363.

Wal J-M (1998) Cow's milk allergens. *Allergy* 53: 1013–1022.

Zeiger RS (2000) Dietary aspects of food allergy prevention in infants and children. *Journal of Pediatric Gastroenterology and Nutrition* 30: S77–S86.

Lactose Intolerance

F Suarez, Abbott Laboratories, Columbus, OH, USA
C Shannon, The Toledo Hospital, Toledo, OH, USA
S Hertzler, The Ohio State University, Columbus, OH, USA
D Savaiano, Purdue University, West Lafayette, IN, USA

Introduction

Lactose intolerance is best defined as the development of gastrointestinal symptoms after the consumption of the 'milk sugar' lactose. Symptoms of intolerance can occur in most humans and all mammals if the amount of lactose consumed exceeds the ability of the gastrointestinal tract to digest this lactose. Thus, lactose intolerance is intermittent since it depends on both the dose of lactose and the response of the gastrointestinal tract. Lactose, a carbohydrate unique to milks and dairy foods, is a disaccharide composed of glucose and galactose, synthesized in the mammary glands of almost all mammals. Young mammals have a large capacity to digest lactose due to a very high activity of a digestive enzyme (a lactase) found on the lining of the small intestine. Most humans and all mammals exhibit a reduced level of this lactase enzyme (and thus a reduced capacity to digest lactose) as they mature past infancy. Interestingly, approximately 25% of the human population maintains a high level of lactase activity and therefore a large capacity to digest lactose throughout life. These individuals are called lactose-tolerant or lactose digesters. A more scientific name for this group is lactase-persistent since the enzyme activity is maintained or persists. The majority of lactose digesters come from northern European, central African and Middle Eastern backgrounds. The remaining 75% of the world's population, including almost all Asians, Native Americans, and many Africans and Latinos, are thus described as lactose-intolerant or lactase-nonpersistent. The regulation of intestinal lactase activity is under genetic control.

The belief that lactase-nonpersistent individuals will always have abdominal symptoms, such as flatulence, abdominal pain, and diarrhea, following dairy food consumption has resulted in many individuals eliminating dairy products from their diet in an effort to prevent symptoms. However, dairy products provide key nutrients and the majority of calcium found in the food supply. The elimination of milk and milk products from the diet makes it very difficult to obtain the recommended dietary allowance of calcium from natural dietary sources. The average American

woman consumes only 50–60% of the 1000–1500 mg recommended for maximum retention of body calcium. Inadequate calcium intake compromises bone structure and may eventually result in osteoporosis. Thus, it is important to provide clear dietary management that allows lactase-nonpersistent individuals to consume adequate amounts of essential nutrients, including calcium, while avoiding symptoms of intolerance.

Lactose intolerance must not be confused with milk allergy. Milk allergy is caused when the immune system reacts against one or more of the proteins found in milk. Milk allergy usually develops in early infancy, before 1 year of age. In developed countries, the incidence in infancy is estimated at $\leq 5\%$ and the remission rate is about 90% before 3 years of age. Gastrointestinal symptoms are reported in approximately 50% of infants with milk allergy while cutaneous symptoms are more common and respiratory symptoms less so. In contrast, lactose intolerance has no immunological bases, only follows the loss of intestinal lactase activity, which occurs around 3–5 years of age, and is limited to gastrointestinal symptoms.

To understand the dietary management of lactose intolerance, it is important to recognize that multiple factors affect the body's ability to digest and tolerate lactose. These factors include: (1) the amount of lactose ingested; (2) the level of the residual lactase activity that remains after maturation; (3) gastrointestinal transit time; and (4) the capacity of the colonic microflora to ferment (digest) lactose. This article will review these factors in order to describe a dietary approach to eliminate symptoms of intolerance while consuming adequate calcium, as well as to correct the perception that lactose intolerance is a frequent and severe problem.

Lactose Digestion and Lactase Activity

Lactose is present in varying degrees in all mammalian milk, with the exception of the milk of the sea lion. Mammary glands have a unique capability to synthesize lactose via the enzyme lactose synthetase. Lactose sythetase links the two monosaccharides (simple sugars) glucose and galactose by a beta 1–4 bond, forming the disaccharide lactose. Cows' milk contains about $50\,g\,l^{-1}$ of lactose, whereas human milk contains one of the highest concentrations, up to $75\,g\,l^{-1}$. Table 1 lists the lactose content of common dairy foods. Fluid milk products have the highest concentration of lactose. When cheese is manufactured, lactose remains with the whey. Thus, hard cheeses contain very small amounts of lactose since the whey is removed in the elaboration process. Lactose may also be found in medications in biologically insignificant amounts (milligrams) due to its excellent tablet-forming properties.

Dietary lactose cannot be absorbed intact directly across the small intestinal mucosa; it must first be hydrolyzed. The enzyme lactase-phlorizin hydrolase (commonly called lactase), a disaccharidase located on the brush border of the intestinal epithelium, breaks the chemical bond between glucose and galactose. These monosaccharides are absorbed across the intestine and transported to the liver for utilization. The lactase enzyme is found must abundantly in the proximal ileum and specifically only hydrolyzes lactose. Human infant lactase is the last of the disaccharidases to appear during fetal production. The gene responsible for the synthesis of lactase has been located in humans on chromosome 2. Lactase activity can be detected as early as 8 weeks after conception and increases to that of only one-third of a full-term infant by 34 weeks' gestation. During late gestation (35–38 weeks), lactase activity reaches approximately three-quarters of that of a full-term infant. Lactase synthesis and activity are high in nearly all full-term infants and remain high throughout the first 4 years of life.

At some point after weaning, a genetically programmed reduction of lactase synthesis occurs, in approximately three-fourths of the world's population, to a residual level of 5–10% that of infancy. This loss of intestinal lactase activity is not a disease, but rather a normal pattern in human physiology and is transmitted by a recessive gene. The decline in lactase activity is known as lactase-nonpersistence (LNP) or primary acquired hypolactasia and cannot be modified by continued exposure to milk or lactose.

In humans there appears to be a mosaic pattern of lactase activity in the jejunal enterocytes. In LNP individuals, some jejunal enterocytes produce high amounts of lactase while others, even those sharing the same villus, do not produce lactase. Thus, rather than a uniform reduction in lactase production among all enterocytes, a hypolactasic individual may have a patchy distribution of lactose-producing enterocytes that are low in number relative to the nonlactase-producing enterocytes. Current evidence suggests that the regulation of lactase is accomplished primarily at the level of transcription, although post-transcriptional factors may still be elucidated.

Congenital lactose intolerance is an extremely rare inborn error in metabolism in which detectable levels of lactase are absent at birth ($0–10\,IU\,g^{-1}$ protein) and remain abnormal throughout life. An infant with congenital lactase deficiency will have severe diarrheal illness beginning a few days after birth.

Table 1 Lactose and calcium content of selected milks, milk products, and substitutes

Product	*Serving size*	*Lactose (g per serving)*	*Calcium (mg)*
Milk	1 c (237 g)	11	291
Low-fat milk, 2% fat	1 c (237 g)	9–13	297
Lactaid milk, 1% fat, 70% lactose-reduced	1 c (237 g)	3.6	300
Skim milk	1 c (237 g)	12–14	302
Chocolate milk	1 c (237 g)	10–12	280
Sweetened, condensed whole milk	1 c (306 g)	35	868
Dried whole milk	1 c (128 g)	48	1168
Nonfat dry milk, instant	1.5 c (91 g)	46	1120
Buttermilk, fluid	1 c (245 g)	9–11	285
Whipped-cream topping	1 tbs (3 g)	0.4	10
Light cream	1 tbs (15 g)	0.6	14
Half and half	1 tbs (15 g)	0.6	16
Low-fat yogurts	1 c (227–258 g)	11–15	314–415
Cheese			
Blue	1 oz (28 g)	0.7	150
Camembert	1 oz (28 g)	0.1	110
Cheddar	1 oz (28 g)	0.4–0.6	204
Colby	1 oz (28 g)	0.7	194
Cream	1 oz (28 g)	0.8	23
Gouda	1 oz (28 g)	0.6	198
Limburger	1 oz (28 g)	0.1	141
Parmesan, grated	1 oz (28 g)	0.8	390
Cheese, pasteurized, processed			
American	1 oz (28 g)	0.5	174
Pimento	1 oz (28 g)	0.5–1.7	174
Swiss	1 oz (28 g)	0.4–0.6	219
Cottage cheese	1 c (210 g)	5–6	126
Cottage cheese, low-fat, 2% fat	1 c (226 g)	7–8	155
Butter	2 pats (10 g)	0.1	2
Oleomargarine	2 pats (10 g)	0	1
Ice cream			
Vanilla, regular	1 c (133 g)	9	176
French, soft	1 c (173 g)	9	226
Ice milk, vanilla	1 c (131 g)	10	176
Sherbet, orange	1 c (193 g)	4	103
Ice, orange	100 g	0	0

c, cup; tbs, tablespoon.

Treatment with a lactose-free diet eliminates symptoms and promotes normal growth and development.

Because lactase enzyme protrudes from the membrane of the epithelial cell and the greatest activity is found at the tip of the intestinal villi, lactase is extremely vulnerable to intestinal disease or insult. Secondary hypolactasia is a temporary condition caused by damage to the enterocytes via disease, poisons such as alcohol, medications, surgery, or radiation to the gastrointestinal tract (Table 2). This condition is relatively common in developing countries, but most common in Third-World countries where chronic intestinal infections are prevalent. Once the causative disease is resolved and the epithelium heals, lactase activity returns to normal. Total resolution of the insufficiency may require up to 6 months or more of diet therapy.

Assessment of Lactase Levels and Lactose Maldigestion

Lactase level can be ascertained with a modest degree of accuracy simply from the determination of the subject's heritage, because prevalence varies among ethnic and racial groups (Table 3). Individuals of African, Asian, middle Eastern, Mediterranean (Jews, southern Italians, Greeks, Arabs), and Native American origin have a greater than 70% chance of being LNP. Subjects of Scandinavian and middle European origin have about a 5% and 15% chance of being LNP, respectively.

The geographical and racial distribution of lactase persistence has led to the hypothesis that three separate gene mutations occurred several thousands of years ago in places where dairy foods had become

Table 2 Potential causes of secondary hypolactasia

Diseases	
Small-bowel	*Multisystem*
HIV enteropathy	Carcinoid syndrome
Regional enteritis	Cystic fibrosis
(e.g., Crohn's disease)	Diabetic gastropathy
Sprue (celiac and tropical)	Protein-energy malnutrition
Whipple's disease	Zollinger–Ellison syndrome
(intestinal lipodystrophy)	Alcoholism
Ascaris lumbricoides infection	Iron deficiency
Blind-loop syndrome	
Giardiasis	
Infectious diarrhea	
Short gut	
Iatrogenic	
Chemotherapy	
Radiation enteritis	
Surgical resection of intestine	
Medications	
Colchicine (antigout)	
Neomycin (antibiotic)	
Kanamycin (antibiotic)	
Aminosalicylic acid (antibiotic)	

Adapted from: Srinivasan R and Minocha A (1998) When to suspect lactose intolerance: symptomatic, ethnic, and laboratory issues. *Postgraduate Medicine* (1988) 104: 109–123; Scrimshaw NS and Murray EB (1988) The acceptability of milk and milk products in populations with a high prevalence of lactose intolerance. *American Journal of Clinical Nutrition* 48: 1083–1159; Savaiano DA and Levitt MD (1987) Milk intolerance and microbe-containing dairy foods. *Journal of Dairy Science* 70: 397–406.
HIV, human immunodeficiency virus.

an important component of the adult diet. Darwinian theory suggests that mutation yielded a survival advantage to subjects provided they were drinking milk. Herding animals and the use of mammalian milk as a human food originated in northern Europe, the Middle East, and central Africa several thousand years ago. Interestingly, it is in these populations that lactase persistence also appeared. Thus, the current view is that selective advantage in these populations led to the maintenance of a genetic mutation in each of these three populations. Inheritance has spread the persistence trait to other populations. Table 4 shows that approximately 200 million individuals in the USA are lactase-persistent.

Lactase levels can be measured directly or indirectly. Directly assaying the lactase activity requires a mucosal biopsy from the jejunum, since the enzyme is attached to the brush border of the small intestinal mucosa. Such biopsies can be obtained at endoscopies or via the use of a variety of tubes equipped with biopsy capsules. While this test can accurately measure lactase activity, it is an expensive, invasive, and time-consuming test and is therefore rarely used clinically. Thus, lactase activity is usually assessed indirectly from measurements of lactose absorption via blood or breath tests. Digested lactose results in free glucose, which is absorbed rapidly, and ultimately

Table 3 Projections of lactose maldigestion around the world (in millions)

Area	*% Lactose maldigesters (LM)*	*1990*		*2000*		*2020*	
		Population	*LM*	*Population*	*LM*	*Population*	*LM*
Africa	75	629	472	805	604	1172	879
Asia	100	3186	3186	3688	3688	4578	4578
Europe	20	721	144	729	146	722	144
Latin America	70	442	309	520	364	645	452
North America	25	278	70	307	77	363	91
Oceania	25	27	7	31	8	38	10
World		5283	4188 (79%)	6080	4887 (80%)	7518	6154 (82%)

GeoHive, 2002 ONLINE. Available http://www.geohive.com/global/index.php [2002].

Table 4 Projections of lactose maldigestion in the USA (in millions)

	% Lactose maldigesters (LM)	*1990 millions of people*	*LM millions of people*	*2000 millions of people*	*LM millions of people*	*2025 millions of people*	*LM millions of people*
African-Americans	75	30	23	34	25.5	44	33
Asian-Americans	100	7	7	11	11	20.5	20.5
Caucasian	20	188	28	197	39.4	209	41.8
Hispanic (all races)	60	22	13	31	18.6	60	36
Native Americans	100	2	2	2	2	2.5	2.5
Total	29	249	72	275	96.5	336	133.8
(% of population)			29%		35%		40%

Source: US Department of Commerce; 1990 census and estimates.

increases blood glucose levels. The blood glucose test involves feeding a standard 50-g lactose dose and measurement of plasma glucose every 30 min over a period of 2 h. In the presence of lactose maldigestion, blood glucose increases less than 25 mg dl^{-1} above the fasting level. Unfortunately, this test is mildly invasive and with a relatively low reliability.

Because of its ease, low cost, and noninvasiveness, the breath hydrogen test is most widely used to diagnose lactose maldigestion. Undigested lactose remains in the intestine and is fermented by colonic bacteria, producing hydrogen gas, carbon dioxide, and methane in some individuals. Bacterial fermentation is the only source of molecular hydrogen in the body. A portion of the hydrogen produced in the colon diffuses into the blood and is excreted via the lungs. The breath test measures the excretion of this hydrogen. Typically, a subject is given an oral dose of lactose following an overnight ($\geq$ 12 h) fast. Breath samples are collected at regular intervals for a period of 5–8 h and analyzed by gas chromatography. The historical test used 50 g of lactose as a challenge dose, and an increase of 20 parts per million (p.p.m.) or greater above the fasting level as an indicator of lactose maldigestion. More recently, it has been shown that using a sum of hydrogen from hours 5, 6, and 7 and a $\geq$ 15 p.p.m. above-fasting criterion for maldigestion resulted in 100% sensitivity and specificity for carbohydrate maldigestion.

Pathophysiology of Lactose Maldigestion

A positive breath hydrogen test measures undigested carbohydrate in the colon and is indicative of lactose maldigestion. However, the correlation between lactose maldigestion and intolerance symptoms, such as flatulence, abdominal pain, and diarrhea, is unclear. The lactose may or may not result in perceptible symptoms depending on a number of factors, including the amount of lactose entering the colon, the metabolic activity of the colonic flora, the absorptive capacity of the colonic mucosa for the end products of lactose fermentation, and the 'irritability' of the colon.

Colonic bacteria metabolize lactose, thereby reducing osmotic pressure in the colon. If the colonic bacteria did not ferment undigested lactose, the osmotic activity of even relatively small doses of lactose could result in diarrhea. However, diarrhea is virtually never encountered when LNP subjects ingest a cup of milk (12 g of lactose). Colonic bacteria ferment lactose to short-chain fatty acids that are readily absorbed by the colon. However, when massive quantities of carbohydrate are maldigested, colonic absorption of fatty acids may not keep up with production. In this situation the bacteria may actually enhance the osmotic activity and hence aggravate the ensuing diarrhea.

In the fermentation process of lactose, appreciable quantities of gas (carbon dioxide, hydrogen, and sometimes methane) are also produced. The removal of these gases is accomplished by bacterial utilization, excretion in flatus, or by absorption through the colon. With small doses of lactose, gas may be removed as rapidly as it is produced and there will be no symptoms. However, with large lactose loads these removal mechanisms may not keep up with production, and bloating, distention, and flatulence result.

A final important factor that plays a role in the development of gastrointestinal symptoms is the response of the colon to the presence of gas and organic acids. Subjects with an 'irritable' colon might perceive symptoms whereas subject with a 'nonirritable' colon might tolerate the same degree of distention without symptoms. The intensity of symptoms may also vary with the amount of lactose consumed, the degree of colonic adaptation, and the physical form of the lactose-containing food.

The Relationship Between Lactose Maldigestion and Lactose Intolerance

The larger the dose of lactose, the greater the risk that the LNP subject will perceive symptoms of lactose intolerance. Early unblinded studies in which subjects were fed milk or lactose and then asked if they had symptoms suggested that over 45% had symptoms following a glass of milk or its lactose equivalent (12 g). The results of such studies have provided the basis for claims that many subjects require a diet severely restricted in milk and milk products or the use of a variety of commercially available lactose-digestive aids. However, tolerance to milk can be affected by factors unrelated to its lactose content and may be due to psychological factors or cultural attitudes toward milk. There is a psychogenic component to abdominal symptomatology, particularly relating to the minor and nonspecific type of symptoms (bloating, distension), that may result from ingested food. The true symptomatic potential of milk (or other foods) can only be obtained with rigidly controlled double-blind studies. Under these controlled conditions, researchers have found that lactose intolerance is less prevalent than commonly believed. For example, in a well-controlled trial, 50 g of lactose (the quantity in a quart (approx. 1l) of milk), taken as a single dose caused symptoms in most of LNP subjects. On the other hand, several blinded studies indicated that LNP subjects tolerated

up to one cup (237 ml) of milk without experiencing significant symptoms. However, the results of these studies did not gain general acceptance, in part because of failure to utilize subjects with 'severe' lactose intolerance. Some LNP individuals believe that small amounts of lactose, such as the amount used with coffee or cereal, cause gastrointestinal distress.

In 1995, our group conducted a study in 30 self-described 'severely lactose-intolerant individuals.' Initial breath hydrogen test measurements indicated that approximately 30% (9 of 30) of the subjects claiming severe lactose intolerance were digesters, and thus, had no physiological basis for intolerance symptoms. Among the true lactose maldigesters there were no significant differences in symptoms of bloating, flatulence, or number of flatus passages during the period when conventional milk (lactose-containing) versus lactose-hydrolyzed milk was ingested. During both treatment periods, symptoms were extremely mild. The triviality of these symptoms with conventional (or lactose-hydrolyzed) milk was particularly impressive given the subjects' prestudy perception that symptoms would be so bad that they would be compelled to withdraw from the study. These findings further demonstrate how strongly behavioral and psychological factors influence symptom reporting. Additional research is necessary to evaluate the psychological component of symptom reporting in lactose maldigesters.

Lactose Digestion and Calcium

Lactose maldigestion could potentially increase risk of osteoporosis either by decreased milk, and therefore calcium, intake or by decreased calcium absorption. Lactose maldigestion has been associated with lower calcium intakes and consequently a higher prevalence of osteoporosis. A sizable fraction of lactose maldigesters may unnecessarily restrict their intake of lactose-containing, calcium-rich dairy foods. Milk and milk products contribute 73% of the calcium to the US food. A second potential relationship between lactose maldigestion and osteoporosis is that maldigestion of lactose decreases absorption of calcium. Human and animal studies suggest that lactose stimulates the intestinal absorption of calcium. However, there is considerable disagreement regarding the influence of lactose and lactose maldigestion on calcium absorption in adults. Differences in study methodology (milk versus water, dose of lactose, and the choice of method for determining calcium absorption) may explain contrasting results. Physiologic doses of lactose (up to two cups (474 ml) of milk) do not appear to have a significant impact on calcium absorption. Therefore, increased prevalence of osteoporosis in lactose maldigesters is most likely related to avoidance of dairy foods and subsequently inadequate calcium intake rather than impaired intestinal calcium absorption.

Dietary Management for Lactose Maldigestion

There are several dietary strategies that effectively reduce or eliminate intolerance symptoms without compromising nutritional status (Table 5).

Dose–Response to Lactose

There is a strong positive correlation between the dose of lactose consumed and the symptomatic response. In general, increasing the amount of lactose consumed increases the number and severity of symptoms. In lactose maldigesters, 2 g of lactose is almost completely absorbed, whereas 6 g of lactose results in a minimal degree of maldigestion as measured by breath hydrogen. On average, maldigesters absorbed about half of a 12.5-g lactose load, whereas the other half passes to the terminal ileum. Interestingly, unblinded studies have frequently reported intolerance symptoms after consuming only 12 g lactose. In 1995 a double-blind protocol demonstrated that feeding 12 g of lactose (approximately one cup of milk) with a meal resulted in minimal to no symptoms in maldigesters who believed themselves to be extremely intolerant to lactose. Further evidence indicated that lactose-intolerant individuals could consume two cups (474 ml) of milk daily with divided doses at breakfast and dinner, without experiencing appreciable symptoms. Overall, a dose containing 18–25 g of lactose has the potential to produce symptoms (mostly flatulence) in approximately half of a population. However, the frequency of symptoms varies from less than 40% to greater than 90% of a subgroup. Thus, a first important dietary strategy is to limit milk consumption to one eight-ounce glass (12.5 g lactose) per meal. Doing so will dramatically reduce the likelihood of intolerance symptoms.

Factors Affecting Gastrointestinal Transit of Lactose

A key explanation of the large variability in dose–response studies is whether or not the lactose is fed with a meal. Consuming milk with other foods slows the intestinal transit of lactose. Slowing the intestinal transit permits longer contact between residual lactase in the small intestine and lactose, thus improving digestion and reducing the potential for symptoms. It is also possible that additional food in the intestine may slow the rate at which lactose is delivery

Table 5 Dietary strategies for lactose intolerance

Factors affecting lactose digestion	*Dietary strategy*	*References*
Dose of lactose	Consume a cup (237 ml) of milk or less at a time, containing up to 12 g lactose	Suarez *et al*. 1995 Hertzler *et al*. 1996 Suarez and Savaiano 1997
Intestinal transit	Consume milk with other foods, rather than alone, to slow the intestinal transit of lactose	Solomons *et al*. 1985 Martini and Savaiano 1988 Dehkordi *et al*. 1995
Yogurts	Consume yogurts containing active bacteria cultures. A serving, or even more, should be well tolerated. Lactose in yogurts is better digested than the lactose in milks	Kolars *et al.* 1984 Gilliland and Kim 1984 Savaiano *et al*. 1984 Shermak *et al*. 1995
	Pasteurized yogurts do not improve lactose digestion. However, these products, when consumed, produce little to no symptoms	Savaiano *et al*. 1984 Kolars *et al*. 1984 Gilliland and Kim 1984
Digestive aids	Over-the-counter lactase supplement pills, capsules, and drops may be used when large doses of lactose (> 12 g) are consumed at once	Moskovitz *et al*. 1987 Lin *et al*. 1993 Ramirez *et al*. 1994
	Lactose-hydrolyzed milks are also well tolerated	Nielsen *et al*. 1984 Biller *et al*. 1987 Rosado *et al*. 1988 Brand *et al*. 1991
Colon adaptation	Consume lactose-containing foods daily to increase the colon bacteria's ability to metabolize undigested lactose	Perman *et al*. 1981 Florent *et al*. 1985 Hertzler *et al*. 1996

in the colon. A slower entry of lactose into the colon might also improve the fermentation and reduce the potential for symptoms.

Individual foods also affect lactose digestion and tolerance. Whole milk, due to its greater fat and energy content, may slightly decrease breath hydrogen relative to skim milk. Whether this effect is large enough to improve tolerance is unclear. Likewise, the addition of chocolate to milk may improve lactose digestion and symptoms, possibly due to its higher osmolality or energy content and/or the effect of cocoa on intestinal transit. Thus, a second key dietary recommendation is to consume milk and other lactose-containing dairy foods with meals. This strategy will also insure that adequate calcium intake can be achieved without symptoms of intolerance.

Yogurts

Lactose that is found in yogurt with live active cultures (*Lactobacillus delbrueckii* subsp. *bulgaricus* and *Streptococcus salivarius* subsp. *thermophilus*) is digested better than lactose in milk. As a result, yogurt is well tolerated by lactose-intolerant individuals. During the manufacture of yogurts, milk solids are typically added to fluid milk to enhance the product. During yogurt production, the bacteria reduce the lactose level from around 6% to approximately 4%. Lactic acid bacteria in yogurt exhibit a very high activity of β-galactosidase. This activity functions during the manufacturing process, and most importantly also during gastrointestinal digestion of lactose *in vivo* in the stomach and/or small intestine. Clinical studies have showed that consumption of up to 18 g lactose as yogurt (two cups of yogurt) is well tolerated, and results in few symptoms of intolerance.

Pasteurization of yogurt increases the shelf-life but decreases the number of active cultures that are partly responsible for improved lactose digestion. However, pasteurized yogurt is moderately well tolerated, producing minimal symptoms. It is believed that the physical form of the yogurt and the caloric density are additional factors that improve tolerance to lactose found in yogurt. Thus, a third dietary strategy is to include yogurts as a part of a calcium-rich and well-tolerated diet.

Lactase Supplements and Lactose-Reduced Milks

Pills, capsules, and drops that contain lactase derived from yeast (*Kluyveromyces lactis*) or fungal (*Aspergillus niger, A. oryzae*) sources have proved effective for lactose digestion. Since 1984, these over-the-counter preparations have been given the status of generally recognized as safe (GRAS) by the US Food and Drug Administration. Additionally, milk that has been treated with lactase, resulting in a 70–100% reduction in lactose, is commercially available. Lactose-hydrolyzed milk typically has increased

sweetness, due to the presence of free glucose. Lactose hydrolysis can be carried out by the consumer at home, or products can be purchased already in low-lactose form. A number of studies have evaluated the effectiveness of these products. Doses of 3000–6000 FCC (Food Chemical Codex) units of lactase administered just prior to milk consumption decrease both breath hydrogen and symptom responses to lactose loads ranging from 17 to 20 g. Thus, a fourth strategy is to include lactose-reduced products and/or utilize enzyme preparations if symptoms of intolerance are present, despite adherence to simpler and more cost-effective approaches.

Unfermented Acidophilus Milk

Various strains of *Lactobacillus acidophilus* exist; however strain NCFM (North Carolina Food Microbiology) has been most extensively studied and used in commercial products. Unfermented acidophilus milk tastes identical to unaltered milk since the NCFM strain does not multiply in the product, provided that the storage temperature is below 40 °F (5 °C). *L. acidophilus* strain NCFM is derived from human fecal samples and contains β-galactosidase (lactase). However unfermented acidophilus milk does not enhance lactose digestion or reduce intolerance symptoms in doses present in commercially available products. It appears that the ingested bacteria are not disrupted by the bile acids, thus microbial lactase is not released. However, when acidophilus milk is sonicated, which destroys the bacteria membrane, lactose digestion improves.

Colonic Fermentation and Bacterial Adaptation of Lactose

The colonic bacteria ferment undigested lactose and produce short-chain fatty acids and gases. Historically, this fermentation process was viewed as a cause of lactose-intolerance symptoms. However, it is now recognized that the fermentation of lactose, as well as other nonabsorbed carbohydrates plays an important role in the health of the colon, and impacts the nutritional status of the individual. The most compelling evidence for colonic bacterial adaptation to lactose comes from studies where the amount of lactose is carefully controlled and gradually increased over time. Increasing the daily lactose dose from 0.3 g kg^{-1} body weight up to 1.0 g kg^{-1} body weight over a period of several days to 3 weeks results in a several-fold increase in fecal β-galactosidase activity. This elevated enzyme activity returns to baseline levels with in a few days after the lactose-feeding period. Furthermore, lactose feeding dramatically decreases the breath hydrogen response to a lactose challenge dose (Figure 1). In fact, after lactose adaptation, the subjects may no longer appear to be lactose maldigesters according to the breath hydrogen test. The large doses of lactose fed during these adaptation periods (up to 70 g day^{-1}) result in only minor symptoms. Additionally, the severity and frequency of flatus symptoms in response to the lactose challenge dose are significantly reduced.

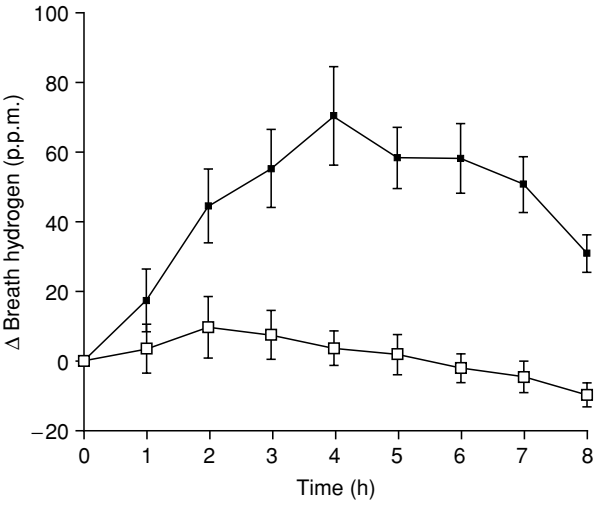

Figure 1 Breath hydrogen response to a lactose challenge after lactose (open squares) or dextrose (filled squares) feeding periods. Data are the means ± SEM, $n = 20$. Reproduced from Hertzler SR and Savaiano DA. Colonic adaptation to daily lactose feeding in lactose maldigesters reduces lactose intolerance. *American Journal of Clinical Nutrition* 64: 232–236, with permission.

Thus it appears that colonic bacterial adaptation to lactose does occur. Although the mechanisms that cause colonic adaptation need further investigation, it is clear that many lactose-intolerant individuals will develop a tolerance to milk if they consume it regularly. This may represent a simpler and less expensive solution than the use of lactose digestive aids. Thus, a final dietary recommendation to lactose-intolerant individuals is not to avoid dairy foods, but rather to include them in the diet on a daily basis. Daily consumption of milk and dairy foods will enhance colon bacterial adaptation and reduce the likelihood for symptoms of intolerance.

Conclusion

Milk and milk products are important sources of many nutrients, including calcium. The avoidance of dairy products to prevent intolerance symptoms jeopardizes bone density, thus increasing the risk for osteoporosis. The inescapable conclusion to be drawn from blinded studies is that virtually all LNP subjects, including even those with severe self-perceived intolerance, can drink milk or milk

products with minimal symptoms if taken in single serving doses with meals. Thus, the role of the health care provider is to convince the patient with self-diagnosed or, less commonly, physician-diagnosed lactose intolerance that various dietary strategies effectively manage lactose intolerance by reducing or eliminating gastrointestinal symptoms. However, beliefs concerning lactose intolerance are not easily reversed, and it may be very difficult to convince patients that their abdominal symptoms are not appreciably aggravated by ingestion of moderate doses of lactose.

See also: **Food Intolerance**: Milk Allergy; **Inborn Errors of Metabolism**: Overview; **Lactic Acid Bacteria**; **Lactose**; **Yogurt**: The Product and its Manufacture; Yogurt-based Products; Dietary Importance

Further Readings

Biller JA, King S, Rosenthal A and Grand RJ (1987) Efficacy of lactase-treated milk for lactose-intolerant pediatric patients. *Journal of Pediatrics* 111: 91–94.

Dehkordi N, Rao DR, Warren AP and Chawan CB (1995) Lactose malabsorption as influenced by chocolate milk, skim milk, sucrose, whole milk, and lactic cultures. *Journal of the American Dietary Association* 95: 484–486.

Florent C, Flourie B, Leblond A, Rautureau M, Bernier JJ and Rambaud JC (1985) Influence of chronic lactulose ingestion on the colonic metabolism of lactulose in man (an in vivo study). *Journal of Clinical Investment* 75: 608–613.

Gilliland SE and Kim HS (1984) Effect of viable starter culture bacteria in yogurt on lactose utilization in humans. *Journal of Dairy Science* 67: 1–6.

Hertzler SR and Savaiano DA (1996) Colonic adaptation to daily lactose feeding in lactose maldigesters reduces lactose intolerance. *American Journal of Clinical Nutrition* 64: 232–236.

Kolars JC, Levitt MD, Aouji M and Savaiano DA (1984) Yogurt – an autodigesting source of lactose. *New England Journal of Medicine* 310: 1–3.

Lin M, Dipalma JA, Martini MC, Gross J, Harlander SK and Savaiano DA (1993) Comparative effects of exogenous lactase (B-galactosidase) preparations on *in vivo* lactose digestion. *Digestive Diseases and Sciences* 38: 2022–2027.

Martini MC and Savaiano DA (1988) Reduced intolerance symptoms from lactose consumed during a meal. *American Journal of Clinical Nutrition* 47: 57–60.

Moskovitz M, Curtis C and Gavaler J (1987) Does oral enzyme replacement therapy reverse intestinal lactose malabsorption? *American Journal of Gastroenterology* 82: 632–635.

Nielsen OH, Schiotz PO, Rasmussen SN and Krasilnikoff PA (1984) Calcium absorption and acceptance of low-lactose milk among children with primary lactase deficiency. *Journal of Pediatric Gastroenterology and Nutrition* 3: 219–223.

Perman JA, Modler S and Olson AC (1981) Role of pH in production of hydrogen from carbohydrates by colonic bacterial flora. Studies in vivo and in vitro. *Journal of Clinical Investment* 67: 643–650.

Ramirez FC, Lee K and Graham DY (1994) All lactase preparations are not the same: results of a prospective, randomized, placebo-controlled trial. *American Journal of Gastroenterology* 89: 566–570.

Rosado JL, Morales M, Pasquetti A, Nobara R and Hernandez L (1988) Nutritional evaluation of a lactose-hydrolyzed milk-based enteral formula diet. I. A comparative study of carbohydrate digestion and clinical tolerance. *Rev Invest Clin* 40: 141–147.

Savaiano DA, AbouElAnouar A, Smith DE and Levitt MD (1984) Lactose malabsorption from yogurt, pasteurized yogurt, sweet acidophilus milk, and cultured milk in lactase-deficient individuals. *American Journal of Clinical Nutrition* 40: 1219–1223.

Shermak MA, Saavedra JM, Jackson TL, Huang SS, Bayless TM and Perman JA (1995) Effect of yogurt on symptoms and kinetics of hydrogen production in lactose-malabsorbing children. *American Journal of Clinical Nutrition* 62: 1003–1006.

Solomons NW, Guerrero AM and Torun B (1985) Dietary manipulation of postprandial colonic lactose fermentation: I. Effect of solid foods in a meal. *American Journal of Clinical Nutrition* 41: 199–208.

Suarez FL and Savaiano DA (1994) Lactose digestion and tolerance in adult and elderly Asian-Americans. *American Journal of Clinical Nutrition* 59: 1021–1024.

Elimination Diets

C M Carter, Great Ormond Street Hospital for Children, London, UK

Definitions and Aims

An elimination diet is one which excludes one or more foods or food additives. The aim of an elimination diet is to diagnose and/or treat food-allergic disease. Food-allergic disease is an intolerance to food(s) resulting from an abnormal immunological response. The title 'food allergy and intolerance' is often used because an immunological basis cannot always be demonstrated. This article discusses the indications for the use of an elimination diet and the dietary manipulations involved.

Indications and Contraindications

The symptoms of food allergy or intolerance may come on quickly or slowly (**Table 1**). Symptoms may

Table 1 Symptoms of food allergy and intolerance

Part of body involved	Usually IgE-mediated	Probably non-IgE-mediated
	Immediate onset	Delayed onset
	Within 2 h	Within a few hours to 3 days
	After small amount of food	After larger amount of food
Gut	Swelling of lips, mouth	Reflux vomiting
	Oral allergy syndrome	Abdominal pain, bloating
	Vomiting	Small-bowel enteropathy
	Colic	Failure to thrive
		Allergic colitis, rectal bleeding
		Diarrhea
		Constipation
Skin	Urticaria	Eczema
	Angiedema	Urticaria (rarely)
	Skin rashes	
Lung	Cough/wheeze/sneeze	Asthma
	Laryngeal edema	
	Bronchospasm	
Systemic	Cardiac arrhythmia	
	Hypotension	
	Anaphylaxis	
Controversial		Migraine
		Migraine with epilepsy
		Attention deficit hyperactivity disorder
		Rheumatoid arthritis

IgE, immunoglobulin E.

appear immediately after a very small amount of food has been consumed and are usually associated with immunoglobulins of the immunoglobulin E (IgE) class. The gut, skin, and/or respiratory tract may be involved and symptoms are listed in Table 1. Skin symptoms include urticaria (also known as nettle rash or hives, an itchy reaction characterized by the appearance of raised red patches or wheals) and angiedema (localized swelling of parts of the face, mainly lips, eyes, and throat) The most serious reaction is known as anaphylaxis, which is a state of shock induced by antigen–antibody (mainly IgE) reaction. It can be life-threatening unless epinephrine (adrenaline) is administered rapidly. People who have these very severe reactions to foods tend to be asthmatic. First signs may be tingling of the lips, tongue, and throat, sneezing, pallor, and feeling unwell and light-headed. Skin rashes, angiedema, and urticaria may or may not occur. First signs are quickly followed by upper or lower airway obstruction, low blood pressure, and cardiac arrhythmia.

On the other hand, symptoms may come on slowly after hours or days and only when a considerable amount of the food has been consumed. Gut, skin, and respiratory reactions are well documented and are likely to be T-cell-mediated without IgE being involved (Table 1). However there is some evidence for a 'late-phase' IgE response. Less well-documented reactions to food include migraine, epilepsy with migraine, attention deficit hyperactivity disorder (ADHD), and rheumatoid arthritis. There is much difference of opinion as to the role of diet, although the small amount of research carried out indicates that the idea that food is involved should be taken seriously, but etiology is not known.

Immediate IgE-mediated Food Allergy

The presence of specific IgE antibodies can be demonstrated quantatively with skinprick tests (SPT) or in the serum with radioallergosorbent tests (RAST). Although the negative predictive accuracy of SPTs is 95%, the positive predictive accuracy is only about 50%. This means that a negative result virtually precludes the possibility of an IgE-mediated allergic reaction, while a positive result indicates an allergic person, but an elimination diet cannot be based on the results of these tests. RAST tests are no more accurate.

However, people (or parents of children) who have immediate reactions to foods usually have no doubt as to the problem food and will need to avoid it. Those who have severe or anaphylactic reactions to food will need to avoid even the smallest trace and this is difficult to do in practice. Any food may be a trigger but the most common provoking foods are peanut, tree nut, milk, egg, wheat, soya, fish, and shellfish. Such elimination diets are hard to achieve for the following reasons:

1. Problems with manufactured food labeling. In Europe and other parts of the world (except the USA and Canada) there is a '25% rule' which means that if a product has a compound ingredient which forms less than 25% of the finished product, its ingredients do not have to be declared. For example, the ingredients of salami on a pizza would not have to be declared. Many food manufacturers are disregarding this rule for peanuts and tree nuts and it is to be hoped that this will happen for milk and wheat also. Another aspect of food labeling involves the use of many terms for one food item. Table 2 shows the many different terms which denote the presence of milk.
2. Factory contamination. Many manufactured foods are not produced on designated lines so there is the possibility of food being contaminated from a previous run or other factory contamination. For this reason most chocolate in the UK and many other manufactured foods carry a 'may contain nuts' label. This precludes a large array of manufactured food for peanut- and tree nut-allergic people.
3. Eating out. Although attempts are being made to educate the catering industry, there are reports of people eating in restaurants or purchasing takeaways being told food was free of the requested items when it was not. There is the possibility of contamination when eating out (or at home) when, for example, a sandwich for a nut-allergic person may be made with an unwashed knife which was previously used for a peanut-butter sandwich. Alternatively a serving spoon could be contaminated with food dished up previously. Cross-contact may also occur through shared grilling surfaces, food containers, and frying oil.

Table 2 Words denoting milk used in ingredient lists on manufactured food labels

Nonfat milk solids
Skimmed milk powder
Casein
Whey
Cheese
Cream
Yogurt
Butter
Ghee
Butterfat
Buttermilk
Caseinates
Dry milk solids
Lactose
Hydrolyzed milk protein
Rennet casein
Sour cream
Whey protein concentrate

Children often grow out of allergies to milk, egg, soya, and wheat but this is much less likely for allergies to peanuts, tree nuts, and fish. Although SPTs may give an indication, the only way to know if an allergy has been outgrown is to perform a food challenge. For people who are at risk of severe reactions challenges are performed in hospital where graded doses of food are given over several hours and resuscitation equipment is available.

Delayed-onset Food Allergy or Intolerance

In the case of delayed reactions to foods, it is much less likely that people know which foods are upsetting them. Indeed, problems with foods which are eaten at nearly every meal such as wheat and milk tend to go completely unnoticed. SPTs and RAST tests are not helpful. Claims of alternative diagnostic tests have not been validated. Although the small-bowel biopsy can be a useful tool in the diagnosis of gastrointestinal disease, it does not indicate which foods are the problem.

The elimination diet is therefore the only available 'diagnostic' tool for suspected delayed-onset food allergy or intolerance. The severity of symptoms should be taken into account before deciding whether or not to embark on an elimination diet. For example, it is inappropriate to use such a diet for eczema unless the child has failed to respond to the optimum skin treatment and the eczema is severe enough to warrant diet intervention. Diets can be socially disruptive and nutritionally inadequate unless supervised properly and may cause more problems than the disease it is intended to treat. However, if symptoms are severe and foods are suspected as being implicated (although this is not a prerequisite), a 'diagnostic' elimination diet is indicated. In the case of severe symptoms, it is important that other more serious disease has been ruled out before embarking on the diet approach.

Elimination Diet Strategy

Diagnosis of food allergy/intolerance by elimination diet involves several phases. The first is an appropriate elimination diet which, if successful, results in relief of some or all the symptoms. The second phase involves sequential reintroduction of foods to identify those foods that cause problems, and a third phase is the final maintenance diet excluding all foods that were demonstrated to have caused problems on open introduction.

Phase 1: elimination diet This may be a diet excluding only one or two foods, a diet omitting a large

number of foods, or even a diet consisting of a hypoallergenic formula only.

The empirical exclusion diet An empirical diet is used where food allergy/intolerance is suspected, causative agents are not known, and one or several of the most commonly provoking foods are avoided. Studies with children who have delayed reactions to food indicate anecdotally that the most common provoking foods are similar to those listed for immediate reactions to foods, except that intolerance to some food additives and chocolate appears more common and intolerance to fish and nuts less common.

In under 1-year-olds the most frequent offender is cows' milk-based infant formula, so a cows' milk-free diet is often used. Diet avoiding egg and milk together with any known problem foods is sometimes used for eczema. It is not unusual for pediatric gastroenterologists to use a diet free of milk, egg, wheat, rye, and possibly soya for children with gastrointestinal symptoms. Several adult centers use a diet free of all grains (some include rice), egg, milk, chocolate, and additives. This is sometimes known as a 'hunter gatherer' diet.

The empirical diet may be followed for up to 6 weeks before a decision is made as to whether it has helped or not. Unhelpful diets must be abandoned.

Problems occur with empirical diets when excluded foods are inadvertently replaced by others which are equally capable of causing adverse reactions. For example, a child on a milk-free diet may drink soya milk or orange juice instead and it is possible to be equally intolerant to these foods. Failure to respond to an empirical diet does not rule out the possibility of food intolerance. More restricted diets therefore have a role to play.

Few-foods diet There are considerable difficulties in teaching people how to avoid a large number of foods. For very restricted diets it is easier to decide on which foods the child can eat and teach the diet in terms of which foods are allowed rather than concentrating on those that are forbidden. This is the basis for the few-foods diet and 3–4 weeks is the longest time one should consider using a very restricted diet, although improvement may occur in a shorter time.

The few-foods diet may consist of no more than 5–10 foods, all of which are least likely to cause problems. Different people working in this field have achieved this in several ways. The simplest few-foods diet consists of lamb, pears, and spring water only. Although an adult may manage this, a child would have great difficulty. Diets using one meat, one carbohydrate source, one fruit, and one vegetable have been used. If no improvement occurs, a second diet containing a different set of foods can be used. This is an ideal diet in theory but, because of compliance problems, especially in children, it is not uncommon to double up and allow two meats, two carbohydrate sources, etc. (Table 3). The diet should be individualized to some extent in that it cannot

Table 3 Few-foods diet and modifications

	Few-foods diet	*Possible variations and/or additions*
Meats	Turkey, lamb	Pork, chicken, rabbit
Starches	Rice, potato	Sweet potato, corn/maize
Vegetables	Two from:1. Brassica	All vegetables except tomato
	2. Carrot, parsnip	Include pulses
	3. Cucumber, marrow, courgette	
	4. Onion, leek, asparagus	
	5. Swede, turnip	
Fruit	Two from:1. Peaches and apricots	Allow all fruits listed on the left
	2. Pears	
	3. Pineapple	
	4. Melon	
	5. Grapes	
Drinks	Tap/bottled water	Hypoallergenic formula for under 2s
	Juice from allowed fruits	
Miscellaneous	Sunflower or olive oil	Spices
	Whey-free margarine	
	Small amount of sugar	
	Jam from allowed fruits	
	Salt, herbs	
Vitamin/mineral supplements	Not essential in short term	
	May be needed in the long term	

Rice can be given as Thai rice noodles, puffed rice cakes, homemade rice cookies, rice 'drink,' boiled and fried rice, and risotto. Potatoes can be prepared in many ways, including plain crisps/chips.

include known problem foods, disliked foods, or foods which are craved and usually eaten in large amounts. Table 3 gives ideas for variations or slightly enlarging the diet to make it more practical. Allergy/intolerance can occur with any food and the few-foods diet is really a 'guesswork' diet using those foods least likely to be implicated.

The few-foods diet can be difficult to follow and should only be adopted if symptoms are frequent and severe. The diet should be carefully taught and the patient given ideas for menus and packed lunches and provided with recipes for cooking with the foods allowed. Close professional support is essential.

Hypoallergenic formula only For such a diet, the only food allowed is a nutritionally complete formula which has no intact protein, the protein source being supplied by either peptides derived from hydrolyzed protein or synthetic amino acids. Such formulas are available for infants and children and adults.

This may be the first choice of diet for the infant but it is very much a last resort for children and adults and it probably means tube feeding (these products do not have a good taste).

Phase 2: reintroduction of foods If the elimination diet has resulted in a worthwhile improvement, foods are reintroduced sequentially in an effort to identify provoking foods. If the effect of the diet is unclear or if there is no improvement it should be abandoned or another type of elimination diet tried. New foods should be given in normal amounts for a week before they are incorporated into the diet. During the period of reintroduction, which may take months, the nutritional adequacy of the diet must be monitored. Vitamin and/or mineral supplements may be required. If staple foods must be avoided, e.g., milk and wheat, alternatives should be found.

If there is a risk of a severe immediate reaction to a food, it should either be avoided or reintroduced in hospital as described. Some children with multiple symptoms and multiple food intolerances have immediate reactions to some foods and delayed reactions to others.

Some people's symptoms are triggered by inhaled substances, e.g., house dust mite, or by contact with substances, e.g., grass as well as food. It can sometimes be difficult during the reintroduction phase to know whether an adverse reaction is due to a food or other allergens in the environment.

Phase 3: maintenance diet Once all foods (and additives) have been introduced and those foods causing problems avoided, the patient has arrived at a maintenance diet which will need to be adhered to.

Children have a tendency to grow out of their problems so the possibility of reintroducing avoided foods should be discussed every 6 months to a year. This may happen in an unsupervised way as a result of an accidental or a deliberate break in the diet.

If the maintenance diet excludes many foods it may be necessary to compromise. During the introduction phase some foods have an adverse effect after normal amounts are consumed regularly over several days. It may be possible to allow these foods from time to time with no adverse effect.

Outcome Assessment

It is extremely helpful for patients (or parents) to keep a daily symptom score diary beginning at least 1 week before the diet starts and continuing during the open reintroduction of foods. It can report on severity of symptoms, dietary indiscretions, and other factors that may be relevant. Diet manipulation is a very crude form of diagnosis and a diary may make it more objective.

Formulas for Infants with Cows' Milk Allergy

The infant who is not breast-fed and who is allergic to the usual infant formulas which are based on cows' milk protein will require an alternative.

Formulas based on whole soya protein isolate have been available for many years. Infants with immediate IgE-mediated allergies often do well on these formulas, although about 15% will be allergic to soya also. Infants with gastoenteropathic symptoms are much more likely to be intolerant of soya and these formulas should not be used.

As previously mentioned, there are several infant formulas available where the protein has been extensively hydrolyzed to form small peptides. These formulas are usually tolerated by cows' milk-allergic infants but a few highly allergic ones cannot manage these either. A formula where the protein consists of pure synthetic amino acids is most useful in these circumstances.

Proof by Blind Food Challenge

Since diet treatment may have a powerful placebo effect, the only way to prove that there is a physical rather than purely psychological basis for the improvement on diet is to administer the foods 'blind', i.e., when the subject is unaware that he or she is consuming it. This is achieved by offering the challenge food hidden in other food known to be tolerated. On a second occasion, in random order with the

first, material is given which looks and tastes like the challenge but does not contain challenge material. When the observer is also blind to the order of administration the procedure is known as a double-blind placebo-controlled food challenge (DBPCFC). This is regarded as the 'gold standard' for proof of food allergy or intolerance.

The DBPCFC has been used in many research studies. Over the last 25 years or so, American workers have used the DBPCFC as an integral part of investigations into immediate allergic reactions to food. Researchers used incremental doses of food over 2–5 h so that placebo and challenge could be completed in one day. They have been recommended as a diagnostic tool outside the research area but are very little used in the UK.

The DBPCFC has also been used to investigate more controversial slow-onset reactions to foods such as ADHD. However, open reintroductions of foods into an apparently successful elimination diet tend to reveal that many people relapse after several days of eating normal amounts of food. This makes a DBPCFC much more difficult to design. One would need to provide challenge and placebo material for a week each with a 2-week washout period in between. This has to be organized on an individual basis and the following crititeria must be fulfilled:

1. Enough of the provoking food must be given for a sufficient length of time. The amount will have been established during open reintroduction.
2. The challenge food must be disguised in food known to be tolerated in the amount to be given.
3. The two must be indistinguishable from one another.
4. For children in particular, the material must be acceptable in taste and quantity.
5. The challenge food should be provided in the same form as that which caused symptoms on open reintroduction. For example, dried powdered milk may not have the same effect on a subject as fresh pasteurized milk.

Although DBPCFCs may be an essential part of research, they are rarely performed in clinical practice. For the more controversial areas of food allergy or intolerance they are time-consuming and difficult to design. They are perhaps best reserved for those cases in which there are serious doubts about the diagnosis.

Elimination Diets for Prevention of Food Allergy

There is much interest in the possibility of preventing the onset of allergies in babies born into allergic families (sibling or parent with allergic disease). Results of research studies are not clear-cut but some tentative guidelines may be drawn.

Diet During Pregnancy

Although research is ongoing in this field there is not enough evidence to suggest that women who are expecting 'at-risk' infants should alter their diets during pregnancy.

Diet of the Mother During Lactation

There is little evidence to suggest that the diet of the mother during lactation has any effect in the long term on her child. However, there is some evidence that in the short term this may be helpful. Indeed, some highly allergic breast-fed children develop allergic symptoms which can be controlled if their mother avoids certain foods. For mothers who wish to do so, a milk- and egg-free diet could be followed, provided the mother has professional support and will need at least a calcium supplement.

Careful Weaning of the Potentially Allergic Child

Breast-feeding should be encouraged for all children. The 'at-risk' infant should have nothing but breast milk for at least the first 4–6 months. If breast milk is not available the best alternative is probably an infant formula where the protein has been highly hydrolyzed. It is not known whether soya formula has a protective effect. It would seem prudent to begin the weaning process with pure baby rice, vegetables, fruits, and meats, and wait until nearer 1 year old to introduce milk, egg, and wheat. There are no specific guidelines on this. Introduction of nuts should be delayed until at least 3 years.

Conclusion

Elimination diets can be expensive, are socially disruptive, and run the risk of being nutritionally inadequate. They should be reserved for individuals with troublesome symptoms who have not responded to symptomatic treatment. The ideal place for this type of treatment is in specialized clinics with interested doctors and dietitians who have the necessary expertise. Immediate reactions to foods can occasionally be very severe or even life-threatening so very careful advice must be given. Despite all these problems, elimination diets can be of great benefit to many people.

See also: **Allergens**; **Food Additives**: Safety; **Food Intolerance**: Types; Food Allergies; Milk Allergy; Lactose Intolerance; **Infant Foods**: Milk Formulas; **Migraine and Diet**

Further Reading

Carter CM (2001) The Immune System. In: Shaw V and Lawson M (eds) *Clinical Paediatric Dietetics*, 2nd edn., pp. 193–214. London: Blackwells.

Devlin J, Stanton RJH and David TJ (1989) Calcium intake and cow's milk free diets. *Archives of Disease in Childhood* 64: 1183–1193.

Host A, Koletzko B, Dreborg S *et al.* (1999) Dietary products used in the treatment and prevention of food allergy. A joint statement of the European Society for Paediatric Allergology and Clinical Immunology (ESPACI) and European Society for Paediatric Gastoenterology, Hepatology and Nutrition (ESPGHAN). *Archives of Disease in Childhood* 81: 80–84.

Hourihane JO'B (1997) Peanut allergy – current status and future challenges (review). *Clinical Experiments in Allergy* 27: 1240–1246.

Joneja JV (1998) *Dietary Management of Food Allergies and Intolerances. A Comprehensive Guide*, 2nd edn. Vancouver: JA Hall.

Sampson HA (1999) Food allergy part 1. Immunopathogenesis and clinical disorders. *Journal of Allergy and Clinical Immunology* 103: 717–728.

Sampson HA (1999) Food allergy part 2. Diagnosis and management. *Journal of Allergy and Clinical Immunology* 103: 981–989.

Sampson HA, Mendelson L and Rosen JP (1992) Fatal and near-fatal anaphylactic reactions to food in children and adolescents. *New England Journal of Medicine* 327: 380–384.

Sicherer SH (1999) Food allergy: when and how to perform oral food challenges. *Pediatric Allergy and Immunology* 10: 226–234.

FOOD LABELING (LABELLING)

Applications

P Berry Ottaway, Berry Ottaway & Associates Ltd, Hereford, UK

Background

Accurate labeling of food has an important role in the marketing and distribution of food, both in providing information to the consumer and in helping to sell the product. Over the past 150 years, labels have developed from being on-pack advertisements for the food to providing information on the composition, nutritional value, usage instructions, and appropriate storage conditions for the food. The increase in consumer awareness of food composition, processing and nutritional content, together with the need for national and international distribution of food products has inevitably led to increased legislation on labeling.

Essential Labeling Requirements

Irrespective of the country, the legislative requirements for labeling contain most, if not all, of a number of essential requirements in common. These requirements are laid down in the Codex Alimentarius General Standard for the Labeling of Prepackaged Foods (CODEX STAN 1-1985, Rev 1-1991) and are:

- an appropriate product name and/or description;
- a list or description of ingredients;
- a net quantity declaration;
- usage or preparation instructions where appropriate;
- address of the manufacturer or marketer;
- batch or lot coding to allow traceability;
- date marking and storage instructions;
- country of origin.

As 165 countries were members of Codex Alimentarius (Codex) in 2000, adherence to the essential requirements for labeling is becoming universal. For example, in the European Union, the directive on food labeling (2000/13/EC) is, in general, consistent with the Codex Standard.

Name of the Food

According to Codex, the name of the food must indicate the true nature of the food and should normally be specific and not generic. Where a name has been established in a Codex Standard or is prescribed in national legislation, the prescribed name should be used. Where a name for the food has not been prescribed, either a common or usual name based on common usage or an appropriate descriptive term for the food that is not misleading or confusing to the consumer can be used. Where a brand name, trade mark, or 'fancy' name is given to the food for marketing purposes, it must be accompanied by a specific name, as defined above.

The Codex Standard also requires that any additional words and phrases necessary to describe the food to avoid misleading or confusing the consumer with regard to the true nature and physical condition of the food should appear on the label in conjunction with, or in close proximity to, the name of the food. This description should include, but is not limited to, the type of treatment the food has undergone (e.g., dried, concentrated, reconstituted, smoked).

List of Ingredients

The Codex Standard requires that a list of ingredients be declared on the label except in the case of single ingredient foods. This list should be headed or preceded by an appropriate title that should consist of, or include, the word 'ingredients.' The list should consist of all ingredients listed in descending order of weight (m/m) at the time of manufacture of the food, often referred to as the 'mixing bowl' stage.

Added water should be declared in the list of ingredients, except when it forms part of an ingredient such as brine, syrup, or broth used in a compound food and which is declared as such in the list of ingredients. If the water, or other volatile ingredients, are evaporated in the course of the manufacture of the food, there is no need for a declaration.

Codex provides that where a compound ingredient for which a name has been established in a Codex Standard or in national legislation constitutes less than 5% of the food, the ingredients that make up the compound ingredient, other than food additives serving a technological function (e.g., antioxidants or preservatives), need not be declared. A compound ingredient at higher levels in the food can be declared as such, provided that its ingredients are listed in descending order by weight in brackets after the name.

The Codex requirement for compound ingredients is not followed in the European Union (EU) where, since 1979, the directive permits that where a compound ingredient constitutes less than 25% of the finished product, the subcomponents of that ingredient do not have to be declared. It is likely that this exemption will eventually be repealed.

In the USA, there are two options for compounded ingredients. The first is to list the ingredient by its established common or usual name with the components given in descending order by weight in brackets. The second is to theoretically break down the compounded ingredient and incorporate each component in the ingredients list in their weight order.

The declaration of permitted food additives is required by Codex to include the class of additive in accordance with the list of additive classes given in the Standard (Table 1). The equivalent list in the European Union labeling directive differs from Codex in that the categories of color retention agent and foaming agent are not included, but modified starch(es) is included in the EU list.

The increasing awareness and concern over allergens in foods are addressed by Codex, which specifies that foods or food ingredients known to cause hypersensitivity should always be declared. These foods and ingredients are given in a list (Table 2). Very few countries have included this requirement into their legislation. The Food and Drugs Administration (FDA) in the USA has not formally defined allergens but has provided examples of foods that are commonly associated with allergic responses. Further to concerns that certain allergenic foods might have been inadvertently excluded from specific declarations, the FDA issued a warning letter to manufacturers on the topic in 1996, which clearly laid out the need to label potentially allergenic ingredients. The letter also acknowledged that potentially allergenic substances (e.g., collective labeling for spices, flavorings, and colorings) may be exempt from specific declaration under the current regulations.

Table 1 Additive classes for labeling (Codex Standard 1–1985 (Rev: 1–1991))

Acidity regulator	Flour treatment agent
Acids	Flavor enhancer
Anticaking agent	Gelling agent
Antifoaming agent	Glazing agent
Antioxidant	Humectant
Bulking agent	Preservative
Color	Propellant
Color retention agent	Raising agent
Emulsifier	Stabilizer
Emulsifying salt	Sweetener
Foaming agent	Thickener
Firming agent	
The following class titles may also be used for food additives falling into the respective classes:	
Flavor(s) and flavoring(s)	Modified starch(es)

Table 2 Foods and ingredients that are known to cause hypersensitivity and must always be declared (Codex Standard 1–1985 (Rev: 1–1991))

Cereals containing gluten, i.e., wheat, rye, barley, oats, spelt, or their hybridized strains and products of these
Crustacea and products of these
Eggs and egg products
Fish and fish products
Peanuts, soybeans, and products of these
Milk and milk products (lactose included)
Tree nuts and nut products
Sulfite in concentrations of 10 mg kg^{-1} or more

EU legislation to require the labeling of a number of allergens has been proposed, but was still not adopted by mid-2001.

One aspect of the ingredient list that tends to be inconsistent is the declaration of processing aids and the carryover of food additives. Codex defines a processing aid as a 'substance or material, not including apparatus or utensils, and not consumed as a food ingredient by itself, intentionally used in the processing of raw materials, foods or its ingredients, to fulfill a certain technological purpose during treatment or processing and which may result in the non-intentional but unavoidable presence of residues or derivatives in the final product.'

In general, processing aids and additives carried over into the food at a level less than that required to achieve a technological function do not have to be declared. However, if a food additive such as an antioxidant or preservative is carried over into the food from an ingredient in an amount sufficient for it to act in that function in the final food, it must be declared in the list of ingredients.

Net Contents

All prepackaged food should carry a declaration of net quantity. Whilst Codex requires that net content declarations are given in the metric system, this still does not apply in some countries. For example, the law in the USA requires the weight to be given in terms of avoirdupois pounds and ounces, and liquid in US gallons, quarts, pints, fluid ounces, and subdivisions thereof. The American gallon is not the same volume as the Imperial gallon originally in use in the UK and many British Commonwealth countries. Declarations for net contents should be:

- by volume for liquid foods;
- by weight for solid foods;
- by either weight or volume for semisolid or viscous foods.

Codex also requires that food packed in a liquid medium should carry a declaration, in the metric system, of the drained weight of the food. A liquid medium refers to the water, syrup, or salt solutions used, for example, in canned fruits and vegetables. Depending on the legislation of the country, the net contents declaration can be based on the 'average weight' system or 'minimum weight' system, or in some cases both.

Instructions for Use

The labels of foods should contain instructions for use, as applicable, to allow the correct or best utilization of the food. Where necessary, details of reconstitution or other ingredients needed to prepare the food should be given.

Name and Address

There is a general requirement that the name and address of a person or company responsible for the food should be included on the label. This can be the name and address of the manufacturer, packer, distributor, importer, or exporter, and should be the most appropriate contact. European food labeling legislation requires that the name and address must be of a person or company established within the EU. The name and address are required to allow the authorities or consumers to contact those responsible for the food in case of complaint or recall.

Batch or Lot Marking

In order to ensure traceability and identification of the food in terms of production, a batch or lot identification must be indelibly printed or embossed on each container. The identification can be in code or in clear and must identify the producing factory, the batch (lot) number, and/or the date and time of production. In some cases, national legislation dictates the method by which the batch or lot number must be declared.

Date Marking

Many countries, including those in the EU, require an easily understood indication of the minimum durability or shelf-life of the food. The Codex Standard on Food Labeling gives the criteria for the date marking of food in cases where there are no special requirements in individual Codex Standards on foods over commodities.

Codex requires that a 'date of minimum durability' must be declared. The form in which this date is given depends on the estimated life of the food and should consist of the day and the month for products with a minimum durability of not more than 3 months, and the month and the year for products with a minimum durability of more than 3 months. If the month is December, under Codex requirements, it is sufficient to indicate the year. For the first case where the day and month are given, the date should be preceded by the words 'Best before . . . ,' and where the month and year is given, by the words 'Best before end . . . ' These statements should be clearly visible on the package or label. Where it is impractical to print the date in conjunction with the statement, there should be a reference to where the date can be found, for example: 'Best before end: see bottom of can.'

Table 3 Products for which an indication of the date of minimum durability shall not be required

Fresh fruits and vegetables, including potatoes that have not been peeled, cut or similarly treated
Wines, liqueur wines, sparkling wines, aromatized wines, fruit wines, and sparkling fruit wines
Beverages containing 10% or more by volume of alcohol
Bakers' or pastry-cooks' wares, which, given the nature of their content, are normally consumed with 24 h of their manufacture
Vinegar
Food-grade salt
Solid sugars
Confectionery products consisting of flavored and/or colored sugars
Chewing gum

According to Codex, the declaration of the day, month, and year should be declared in an uncoded numerical sequence. The month, however, may also be indicated by letters in those countries where such use will not confuse the consumer. The use of letters for the month can be particularly important in the case of trade between the USA and Europe. In the USA, when defining the date in numerical terms, it is common to place the month before the day (i.e., 04/12 for 12 April), whereas in most European countries, the day comes before the month.

The Codex Standards exempt a number of foods and drinks from the date marking requirements (Table 3). An important declaration that should accompany the date of minimum durability on the label is that giving any special storage conditions required if the validity of the date (i.e., the shelf-life) is dependent on the conditions under which the product is kept.

Nutrition Labeling

An increasing awareness of healthy eating and the impact of the diet on heart disease has led to a consumer demand for information relating to the nutritional composition of foods. This demand has led to many food labels containing information on the nutrient content, and to ensure consistency in the manner in which this information is given, most countries have included detailed requirements for nutrition labeling in their legislation.

As with the general labeling of food, Codex also has produced guidelines on nutrition labeling (CAC/GL 2-1985. Rev 1-1993). According to Codex, the information supplied in a nutrient declaration should be for the purpose of providing consumers with a suitable profile of the nutrients contained in the food and which are considered to be of nutritional importance. This information should not lead consumers to believe that there is exact quantitative knowledge of what individuals should eat in order to maintain health, but rather to convey an understanding of the quantities of the nutrients contained in the food. Nutrition labeling should not be used to deliberately imply that a food that carries such labeling necessarily has any nutritional advantage over a similar food that is not so labeled.

The Codex guidelines lay down both the nutrients that should be declared, and the form in which the information should be given. It lists those declarations that should be mandatory as:

- the energy value of the food;
- the amounts of protein, available carbohydrate (i.e., excluding dietary fiber), and fat;
- the amount of any other nutrient for which a nutrient claim is made;
- the amount of any other nutrient considered to be relevant for maintaining a good nutritional status, as required by national legislation.

Where a claim is made for the type and/or quantity of a carbohydrate, the amount of total sugars should be listed, and the amounts of starch or other forms of carbohydrate may also be listed. In the context of this requirement, sugars are defined as mono- and disaccharides present in the food. Similarly, where a claim is made for the quantity and/or type of fatty acids, the amounts of the saturated and polyunsaturated fatty acids should be declared.

If a claim is made for a dietary fiber content, the amount of dietary fiber should also be declared. The Codex guidelines define dietary fiber as 'edible plant and animal material not hydrolyzed by the endogenous enzymes of the human digestive tract as determined by the agreed upon method.' As this definition may differ in detail from that used in some countries, national legislation needs to be considered to ensure that the calculation of dietary fiber is carried out appropriately.

The vitamins and minerals that can be declared, and the method of their declaration, can vary from country to country. The Codex guidelines are general in that they state that, in addition to the information described above, vitamins and minerals may be listed but only for those micronutrients for which recommended intakes have been established and/or which are of national importance in the country concerned.

Where vitamins and minerals are declared, only those present in significant amounts should be listed. As yet, there is no international agreement on what constitutes a significant amount of a micronutrient in a food. Codex states that, as a rule, 5% of the recommended intake of the population concerned, supplied by a serving as quantified on the label, should be taken into consideration in deciding what constitutes

a significant amount. European law, however, states that, as a rule, 15% of the recommended allowance for the vitamin and mineral supplied by 100 g or 100 ml or per serving, if the package contains only a single portion, should be taken into consideration in deciding what constitutes a significant amount.

When the nutrient content of a food is declared, it is essential that the calculations are in accordance with national legislation, as there are differences between countries in the way in which energy and the quantities of nutrients are to be calculated.

Codex and European nutrition labeling law require that protein is calculated as the Total Kjeldahl nitrogen × 6.25. However, in Codex, there is the rider that this factor is to be used unless a different factor is given in a Codex Standard or a specific Codex method of analysis for a food.

The Codex guidelines give the conversion factors to be used for the calculation of energy (Table 4). These factors are not consistently used across the world. For example, European law introduces a further category of polyols contributing 2.4 kcal/10 kJ g^{-1}.

The format for the presentation of nutrition information can vary from country to country, although the 15 EU member states and many non-EU countries closely follow the Codex guidelines. Codex requires that the declarations of nutrient content should be numerical, but additional presentation in other forms (e.g., pie chart) is not excluded. The energy statement should be given in kilojoules and kilocalories per 100 g or 100 ml or per package if the pack only contains a single portion. The energy may also be given per serving or per portion, provided the serving size is given on the label, or the number of portions in the package are given on the label.

The information on the quantities of protein, carbohydrate, and fat should be given in grams and in the same manner as the energy statement. The information on vitamin and mineral content should be given in the metric system (milligrams or micrograms as appropriate) and/or as a percentage of the nutrient reference value per 100 g or per 100 ml or per package, if the product only contains a single portion. The nutrient reference values for nine vitamins and five minerals are given in the Codex guidelines. The EU nutrition labeling directive 90/496/EEC gives reference values (as recommended daily allowances) for 12 vitamins and six minerals, and requires that the micronutrients also be given as a percentage of these values.

In the USA, there is also a requirement for micronutrients to be labeled as a percentage of reference daily intakes. In both European and US law, the order in which vitamins and minerals are to be declared is laid down. However, these requirements differ in detail. There is also some confusion between the EU and USA in the way in which the activity of certain vitamins is calculated. For example, in Europe, the activity of thiamin (vitamin B_1) is measured as the cation, whereas in the USA, it is often measured as the salt. US law also has specific labeling requirements for micronutrients in products that fall into the category of a dietary supplement under the Dietary Supplement and Health Education Act 1994.

Table 4 Conversion factors that should be used for the calculation of energy from food

Protein	4 kcal/17 kJ g^{-1}
Carbohydrate	4 kcal/17 kJ g^{-1}
Fat	9 kcal/37 kJ g^{-1}
Organic acid	3 kcal/13 kJ g^{-1}
Alcohol (ethanol)	7 kcal/29 kJ g^{-1}

Claims

The use of claims on food labels is the most controversial aspect of food labeling with a lack of agreement between countries as to what is permitted.

In most countries, food law prohibits the making of claims that a food can prevent, cure, or treat a human condition or disease. In most cases, this prohibition has been in the law for decades. More recently, there has been increasing awareness that certain diets, and possibly certain foods, may play a significant part in the reduction of risk of a disease such as heart disease or cancer.

Since the mid-1990s, the interest in functional foods has increased the debate as to what claims could be permitted. Claims have been classified into two categories, both with further subdivisions. These are:

- nutrition claims and
- health claims.

Nutrition Claims

A nutrition claim is any representation that states or implies that a food has particular nutritional properties. These include, but are not limited to, the energy value and the contents of protein, carbohydrate, fat, vitamins, and minerals. It is generally accepted that the mention of nutrients in an ingredient list or in a nutritional declaration does not represent a claim.

Nutritional claims can be categorized as:

1. Nutrient claims that describe the amount of a nutrient in a food, such claims can be either for enhanced levels of a nutrient in a food or for reduced levels of certain nutrients, such as fat, or

2. Comparative claims that compare the nutritional components of a food with different versions of the same food or with similar foods, an example would be the comparison of a low fat or low energy food with the normal version, or
3. Nutrient function claims that relate to the function of a nutrient in the body, such claims relate to the biological functions of the nutrient in the body, for example 'calcium aids in the development of strong bones and teeth' or 'protein helps build and repair body tissues.' A nutrient function claim should be based on scientific consensus and should not imply that, by carrying out its function, the nutrient would prevent, treat, or cure a disease or adverse condition.

Health Claims

A health claim is any statement or implication in the labeling or advertising of food that the food is beneficial to health. The area of health claims is controversial, but such claims can be categorized as either:

1. Claims related to dietary guidelines or healthy diets where the composition of the food can contribute to the meeting of the guidelines when consumed as part of the diet; or
2. Disease risk reduction claims where the claim relates to the consumption of a food or food component, which might help reduce the risk of a specific disease or condition. Although it may be considered to come under the concept of 'prevention,' and thus illegal, a claim that a food or food component helps in reducing the risk of a disease is different from the claim that it prevents a disease. In the USA, a limited number of disease-risk reduction claims have been permitted by the Food and Drugs Administration, provided that all the conditions in the permission are met by the food. Examples of these claims are those for:
   - calcium and osteoporosis;
   - dietary lipids and cancer;
   - sodium and hypertension;
   - folate and neural tube defects;
   - fruit and vegetables and cancer.
3. Enhanced function claims that concern the specific beneficial effects of nutrients or other components of a food on the physiological or psychological functions of the body. Such claims can also cover biological activities beyond the established roles of nutrients in growth, development, and other normal functions of the body.

There is little international agreement on which health claims could be permitted, and whilst discussions are taking place within Codex Alimentarius, the European Commission, and a number of countries, there is unlikely to be common consent before about 2005.

Warnings

Through the 1990s, food-related product liability issues and litigation increased significantly. This has led to an increase in warnings on the product labels, mostly related to ingredients. Whereas a number of the warnings are voluntary and relate to possible allergens in the food, some have become statutory. For example, in the USA, food products containing saccharin must, by law, include the statement 'Use of this Product May be Hazardous to Your Health. This Product Contains Saccharin Which Has Been Determined to Cause Cancer in Laboratory Animals.' In addition, products containing the sweetener Aspartame must be labeled with the warning 'Phenylketonurics – Contains Phenylalanine.' A very similar requirement is found in European law when the warning has to be 'Contains a source of phenylalanine.'

In Europe, it is also a legal requirement to include the warning 'Excessive consumption may produce laxative effects,' where a product contains 10% or more by weight of added polyols, regardless of the serving size of the food.

Other Labeling Requirements

There are some aspects of food labeling for which international agreement has not been reached but have already been introduced into the legislation of some countries.

One such requirement is the quantitative declaration of ingredients (QUID), which became part of European labeling law in 2000. A QUID declaration is required when, for example, an ingredient or category of ingredients appears in the name of a product (e.g., ham and mushroom pizza or strawberry yogurt) or where an ingredient is emphasized on the label in words, pictures, or graphics. In such cases, it is required that the percentage of the ingredient in the food product must appear either after the name of the food or in the list of ingredients or category of ingredients in question.

From 1997, there has been consumer-driven concern in a number of countries, particularly some in the EU, about the use of genetically modified organisms (GMOs) or byproducts of GMOs used in foods. As a consequence, from 1998, EU law required the labeling of certain GMO sources if genetically modified protein or deoxyribonucleic acid can be detected. Rapid developments in the use of genetic

modification technology meant that by the end of 2000, the legislation was already becoming outdated, and more comprehensive labeling of GMO sources had to be considered. In many countries, there has not been the same consumer reaction to GMO ingredients, and their labeling has not been a priority.

See also: **Food Intolerance**: Food Allergies; **Nutrition Policies in WHO European Member States**

Further Reading

Blanchfield JR (2000) *Food Labelling*. Cambridge, UK: Woodhead.

Codex Alimentarius Commission (1999) *Food Labelling Complete Texts*, revised. *Joint FAO/WHO Food Standards Programme*. Rome: FAO.

Food and Drug Administration USA (1997) *A Food Labeling Guide*. FDA Center for Food Safety and Applied Nutrition http://vm.cfsan.fda.gov/-dms/flg-1.html

FOOD POISONING

Contents

Classification

D Roberts, Central Public Health Laboratory, London, UK

Introduction

The term food poisoning is used to describe a disturbance of the gastrointestinal tract following the consumption of contaminated food or water. Some regulatory or surveillance agencies prefer to use the terms 'foodborne disease' or 'foodborne infections and intoxications', but for simplicity and ease of understanding by most people the term 'food poisoning' will be used in this article.

A wide range of agents can give rise to food poisoning and the most common of these are listed in **Table 1**. In Europe and North America by far the most common causes of food poisoning are microorganisms, in particular bacteria and viruses (see **Tables 2** and **3**). The involvement of other agents, such as parasites, naturally toxic foods, and chemicals, will vary according to country, diet of the population, level of hygiene in food production, and level of food-processing technology. For example, in countries where there is a shortage of food containers there is opportunity for accidental contamination of food ingredients when barrels and sacks have been previously used for different commodities which may be toxic. (*See* **Parasites**: Occurrence and Detection *and also* individual agents.)

Symptoms

The symptoms of food poisoning may be very diverse and the combinations which are manifest in a patient, together with the incubation period, are often indicative of the agent involved, and thus a useful diagnostic tool. The most common symptoms are nausea, vomiting, and diarrhea with or without fever. In severe cases and with particular agents there may be more serious effects, such as septicemia or involvement of the nervous system.

Table 2 summarizes the incubation period, symptoms, and duration of illness of the main types of bacterial food poisoning. Full descriptions of these illnesses may be found in the relevant articles elsewhere in this encyclopedia. Refer to individual microorganisms. This table also includes *Listeria monocytogenes* which, although it does not usually give rise to typical food-poisoning symptoms, can be transmitted via food.

A short incubation period of a few hours with a vomiting-type syndrome is usually associated with a toxin or poisonous chemical already present in the food, either by accidental contamination (as with some chemicals) or by growth of bacteria such as *Staphylococcus aureus* and *Bacillus cereus*. The

Table 1 Agents of food poisoning

	Examples
Bacteria	*Salmonella, Staphylococcus aureus, Clostridium perfringens, C. botulinum, Bacillus cereus*
Viruses	Parvovirus; Norwalk agent; small, round, featureless viruses
Molds	*Claviceps purpurans* (ergotism)
	Aspergillus flavus (aflatoxicosis)
Marine protozoa	*Gonyaulax tamarensis* (paralytic shellfish poisoning)
	Gambierdiscus toxicus (ciguatera poisoning)
	Prorocentrum lima and several species of *Dinophysis* (diarrheic shellfish poisoning)
Parasites (helminths, protozoa, amebae)	*Trichinella spiralis, Taenia solium, Giardia lamblia, Cryptosporidium, Entamoeba histolytica*
Chemicals	Heavy metals, e.g., lead, zinc, copper, cadmium, mercury, arsenic. Insecticides, herbicides, fungicides, cleaning agents, disinfectants
Toxic plants	Toadstools, hemlock, deadly nightshade
	Raw or undercooked red kidney beans (*Phaseolus vulgaris* lectin)
	Apricot kernels (glucoside-amygdalin), green potatoes (solanine)
Toxic food animals	Puffer fish (tetrodotoxin), red whelk poisoning (tetramine)
	Scombroid fish, poorly processed (histamine)

toxin produced by *Clostridium botulinum* is a neurotoxin and takes longer to act. With agents that produce an infection within the human gut the incubation period is usually longer – 24 h or more; the organisms require this length of time to pass through the stomach and reach the intestine before they multiply and produce their effect.

Chemical Food Poisoning

In the UK chemical food poisoning is uncommon and probably responsible for not more than 1% of all episodes; however, it may be more common in other countries. The source of the chemical may be accidental contamination, leaching out of chemicals from cooking utensils or pipework (copper, cadmium, lead), mistaken identification of powders, e.g., those resembling flour (nicotinic acid, sodium fluoride), or carry-over from harvested produce previously treated with fungicides, herbicides, or insecticides. With the exception of one or two neurotoxins, the main characteristic of chemical food poisoning is vomiting within a few minutes to half an hour of ingestion of the food – a much shorter incubation period than with bacterial food poisoning, where symptoms rarely appear in less than 2 h. Recovery is usually rapid, although with some poisons there may be a high case-fatality rate. Some foods, of both plant and animal origin, are poisonous to humans because they contain certain naturally occurring toxic materials such as the lectin of the red kidney bean and tetrodotoxin of puffer fish. Others become toxic owing to the concentration of chemical poisons produced by the organisms upon which they feed, e.g., shellfish and fish which feed on the marine protozoa which produce the toxins responsible for paralytic and diarrheic shellfish poisonings and ciguatera poisoning. Scombroid fish, which are rich in histidine, become toxic owing to the breakdown of the histidine to histamine by their own natural microbial flora during poor storage following harvesting. (*See* **Contamination of Food; Fish**: Spoilage of Seafood; **Histamine; Shellfish**: Contamination and Spoilage of Molluscs and Crustaceans.)

Bacterial Food Poisoning

The term 'bacterial food poisoning' is usually restricted to the acute gastroenteritis caused by the presence of bacteria, usually in large numbers, or the products of their growth in food. It may be subdivided into three basic types according to the mode of action of the bacteria responsible:

1. Infection-type bacterial food poisoning results from the in vivo multiplication of bacteria consumed with the food, e.g., *Salmonella, Vibrio parahaemolyticus*. The dose of organisms required to initiate infection is usually large (10^6 or more per g); this enables sufficient organisms to pass through the acid region of the stomach and reach the intestine, where they set up their infection. However, infections may result from a low dose of organisms if the food has a protective effect through the stomach, e.g., if it is high in fat or has a high buffering capacity. Thus outbreaks of salmonella infection have resulted from foods such as chocolate and cheese containing levels of 100 salmonellae or less per gram of food.
2. Toxin-type poisoning results from the ingestion of food in which a toxin has already been formed as a result of microbial growth in the food. Examples of this type include *Staphylococcus aureus* (enterotoxin), *B. cereus* (emetic toxin) and *C. botulinum* (neurotoxin). The organisms need to multiply to levels exceeding 10^5 organisms per g in order for sufficient toxin to be produced to cause illness. The emetic toxins which act on the vomiting

Table 2 The main agents of bacterial food poisoning

Organism	*Incubation period (h)*	*Duration*	*Symptoms*
Salmonella (infection)	6–48 (usually 12–36)	1–7 days	D, Abp, V, F
Staphylococcus aureus (toxin in food)	2–6	6–24 h	N, V, D, Abp; collapse and dehydration in severe cases
Clostridium perfringens (toxin in intestine)	8–22	24–48 h	D, Abp, N
C. botulinum (toxin in food)	12–96 (usually 18–36)	Death in 24 h to 8 days or slow convalescence over 6–8 months	Lassitude, fatigue, dizziness, D followed by C; CNS involvement, disturbance in vision and speech
Bacillus cereus (toxin in food)	Vomiting syndrome, 1–6 Diarrheal syndrome, 8–16	6–24 h 18–24 h	N, V, sometimes D Abp, D, sometimes N
B. subtilis (toxin in food)	< 1–14 (median 2.5)	1.5–8 h	V, D, Abp, N
B. licheniformis (toxin in food)	2–14 (median 8)	6–24 h	D, V, Abp
Escherichia coli (infection and toxin)	12–72	1–7 days	D with blood and mucus in stool
Vibrio parahaemolyticus (infection)	2–48 (usually 12–18)	2–5 days	Profuse D, Abp, V, F, often dehydration
Yersinia enterocolitica (infection)	24–36 (3–5 days)	3–5 days	Moderate D to chronic enteritis
Streptococcus (toxin in food? and infection)	3–22	24–48 h	Abp, V, D
Campylobacter jejuni (infection)	3–5 days	Days or weeks	F, headache, Abp, profuse D with blood and mucus
Listeria monocytogenes[a]	1–70 days	Days to years, high mortality rate (30%)	Mild, flu-like illness, septicemia, meningitis, abortion

D, diarrhoea; N, nausea; Abp, abdominal pain; C, constipation; V, vomiting; F, fever; CNS, central nervous system.
[a]A foodborne infection rather than true food poisoning.

center produce symptoms fairly rapidly, within a few hours of consumption of the food, while the neurotoxin of *C. botulinum* takes longer to act.

3. In intermediate-type poisoning the organism (e.g., *C. perfringens*) releases an enterotoxin, but in the bowel; the toxin is not produced readily in the food. Again, a high dose of organisms is required in the food to insure that some survive through the stomach and reach the intestine, where they sporulate and release the enterotoxin, which causes fluid production and diarrhea.

Viruses, which cannot multiply in foods because they require living cells in which to replicate, cause the infection type of food poisoning. The infectious dose level may be very small.

Introduction of Microorganisms into Foods

Table 3 summarizes the principal food sources of the most common food-poisoning organisms and *L. monocytogenes*. Foods of animal origin are the primary source of many of these microorganisms. Those found in the live animal can be carried through to raw meats after slaughter, may persist through further processing, and may ultimately appear in the final retail product if insufficient attention is paid to hygiene and temperature control. Many animals are mass-reared for food, and a large proportion may arrive at the slaughterhouse excreting organisms, such as *Salmonella* and *Campylobacter*, in addition to the pathogens that form part of their normal fecal flora, such as *C. perfringens, Escherichia coli, Yersinia enterocolitica*, and *L. monocytogenes*. Continuous line processing will increase spread from carcass to carcass. Comminution of meat results in a spread of organisms throughout the product. (*See* **Meat: Slaughter.**)

During laying, eggs may be contaminated on the shell surface by *Salmonella* from chicken feces, which may be transferred to the contents during unhygienic breaking out. A small proportion may also be infected while in the hen's ovary. Milk, when freshly drawn from a healthy animal using hygienic dairy practices, should contain very few bacteria. However, various pathogens may be acquired from the milking animal herself (fecal contamination or udder infection) or from the dairy environment and equipment. Some outbreaks of salmonellosis and campylobacteriosis

Table 3 Principal food sources of the common microorganisms causing food poisoning

Agent	*Food source*
Salmonella	Raw meat and poultry, raw milk, eggs
Staphylococcus aureus	Cold meats and desserts (much handled during preparation); dairy products, especially if prepared from raw milk
Clostridium perfringens	Raw meat, poultry, dried foods, herbs, spices, vegetables
C. botulinum	Improperly preserved meat, fish, and vegetables
Bacillus cereus and other *Bacillus* spp.	Cereals, dried foods, dairy products, meat and meat products, herbs, spices, vegetables
Escherichia coli	Many raw foods, particularly of animal origin
Vibrio parahaemolyticus	Raw and cooked fish, shellfish and other seafoods
Yersinia enterocolitica	Raw meat and poultry, meat products, milk and milk products, vegetables
Streptococcus	Raw milk, raw meat and poultry, foods prepared by infected food handlers
Campylobacter jejuni	Raw poultry meat, raw or inadequately heat-treated milk, untreated water
Viruses	Raw shellfish, cold foods prepared by infected food handlers
Listeria monocytogenes (a widespread environmental organism)	Meat, poultry, soft cheese and other dairy products, vegetables, shellfish

are associated with the consumption of raw milk, and a range of organisms may be found in the product (Table 3). This source of infection can be eliminated by heat treatment of the product. (*See* **Milk**: Processing of Liquid Milk.)

Fish and shellfish may become contaminated either from the water from which they are harvested or from the environment during further processing. When fished from water polluted with sewage they may contain many fecal pathogens (Table 3). *Vibrio parahaemolyticus* is a common marine microorganism and is a frequent contaminant of raw fish and other seafoods. Shellfish are filter feeders and thus can concentrate microorganisms (bacteria, viruses, and protozoa) from the water in which they are breeding. Consumption of raw shellfish, e.g., oysters, or those that have been inadequately heat-treated, can lead to bacterial and viral infections.

Some food-poisoning bacteria are common environmental organisms; they may be present on the soil, in river mud, and on vegetation. Thus many raw foods may be contaminated from this source with organisms such as *C. perfringens, B. cereus*, and other *Bacillus* spp., and *L. monocytogenes*. Fruits, vegetables, and cereals will frequently carry a number of these organisms. Where crops require artificial irrigation and/or fertilization, particularly if polluted water and animal or human excreta are used, there is a risk of contamination with enteric bacterial pathogens.

Thus raw foods and ingredients are frequently contaminated with a range of pathogenic organisms. Many of these will be rendered safer by the application of some form of heat treatment, but they are unlikely to be sterile unless a stringent heat process such as the 'botulinum cook', applied to certain canned foods, is used. Pasteurization temperatures will eliminate most vegetative cells, but the spores of *Clostridium* spp. and *Bacillus* spp. will survive. The spores of some of the clostridia are extremely heat-resistant, and the heat processes applied in the canning industry are aimed at eliminating the most heat-resistant spores of the most serious food-poisoning organism, *C. botulinum*. (*See* **Heat Treatment**: Chemical and Microbiological Changes; **Pasteurization**: Principles.)

The food-poisoning toxins produced by bacteria vary in their ability to withstand a heat treatment. The toxins of *C. botulinum* are heat-labile and readily destroyed by boiling, while some of the enterotoxins produced by *Staphylococcus aureus* are very heat-resistant and will withstand boiling for 30 min. The emetic toxin of *B. cereus* is extremely heat-resistant and remains active after 90 min at 121 °C.

Further manipulation of a product after heat treatment will allow entry of a wide range of organisms. Those of particular concern include *Y. enterocolitica* and *L. monocytogenes* both of which can multiply at refrigeration temperatures – refrigeration being the common means of short-term preservation of foods. Thus care must be exercised with respect to hygiene of production, temperature control during storage and display, and shelf-life of the product, if the introduction and growth of these bacteria are to be prevented.

Organisms may also be transferred to food by the food handler, either directly or by cross-contamination through the use of hands, surfaces, utensils, and equipment that have not been adequately cleaned and disinfected between the preparation of different foods, especially between raw and cooked foods. It is in relation to *S. aureus* food poisoning that the food handler has an important role as a direct source of contamination. This organism is commonly carried in the nose and on the skin of humans, as well as causing skin infections, and can be readily transferred to foods by handling. Subsequent

storage at unsuitable temperatures will allow the organisms to multiply and produce their toxins. Food handlers who continue to work with active symptoms of gastroenteritis are also a hazard because there is an increased risk of fecal organisms reaching food.

Prevention of Food Poisoning

An analysis of food-poisoning outbreaks which occurred in England and Wales between 1970 and 1982 showed that the most common contributory factors related to temperature control (Table 4) – either the temperatures at which the food was held after cooking or those achieved during preparation and cooking. A later study published in 1996 drew similar conclusions in relation to outbreaks occurring between 1992 and 1994 in the UK. In order to prevent outbreaks of food poisoning it is important that everyone involved in food preparation is made aware that very few basic ingredients are sterile, but will be contaminated to a greater or lesser extent depending on their origin and amount of heat processing and other manipulation. Poor hygiene during processing and inadequate temperature control can lead to proliferation of organisms to infectious dose levels or the production of toxins in foods. Improvement in agriculture procedures can lead to the reduction in the presence of some organisms, such as *Salmonella* and *Campylobacter*, in our foods, but good hygienic food preparation and education of those involved in food production are the final lines of defense.

Application of the hazard analysis critical control points (HACCP) program is recommended for all food production processes, from small catering procedures to large-scale manufacturing. Once the sources of contamination (the hazards) and the important points at which, and means by which, they can be controlled (critical control points) are identified, and controls introduced, monitored, and verified, then the food manufacturer can have greater assurance that he or she is producing a safe product. (*See* **Hazard Analysis Critical Control Point**.)

Table 4 The most common factors contributing to outbreaks of food poisoning in England and Wales in 1970–1982

Contributing factor	*Percentage of outbreaks*[a] *in which the factor was recorded*
Preparation too far in advance	57
Storage at ambient temperature	38
Inadequate cooling	32
Inadequate reheating	26
Contaminated processed food	17
Undercooking	15
Contaminated canned food	7
Inadequate thawing	6
Cross-contamination	6
Raw food consumed	6
Improper warm-holding	5
Infected food handlers	4
Use of left-overs	4
Extra-large quantities prepared	3

[a]1479 outbreaks studied.
In some outbreaks several factors contributed.
From Roberts D (1986) Factors contributing to outbreaks of Foodborne infection and intoxication in England and Wales 1970–1982. *Proceedings of the Second World Congress of Foodborne Infection and Intoxications*, vol. 1, pp. 157–159. Berlin: Institute of Veterinary Medicine.

See also: **Contamination of Food**; **Fish**: Spoilage of Seafood; **Hazard Analysis Critical Control Point**; **Heat Treatment**: Chemical and Microbiological Changes; **Histamine**; **Meat**: Slaughter; **Milk**: Processing of Liquid Milk; **Parasites**: Occurrence and Detection; **Pasteurization**: Principles; **Shellfish**: Contamination and Spoilage of Molluscs and Crustaceans

Further Reading

Djuretic T, Wall PG, Ryan MJ, Evans HS, Adak GR and Cowden JM (1996) General outbreaks of infectious intestinal disease in England and Wales 1992–1994. *Communicable Disease Report* 6: R57–R63.

Doyle MP (ed.) (1989) *Foodborne Bacterial Pathogens*. New York: Marcel Dekker.

Doyle MP, Beuchat LR and Montville TJ (eds) (1997) *Food Microbiology. Fundamentals and Frontiers*. Washington, DC: ASM Press.

Gilbert RJ and Roberts D (1993) Foodborne bacterial gastroenteritis. In: Parker MT and Collier LH (eds) *Topley and Wilson's Principles of Bacteriology, Virology and Immunity*, 9th edn, vol. 3, pp. 539–620. London: Edward Arnold.

Hobbs BC and Roberts D (1993) *Food Poisoning and Food Hygiene*, 6th edn. London: Edward Arnold.

International Commission on Microbiological Specifications for Foods (1980) *Microbial Ecology of Foods: 1 Factors affecting Life and Death of Microorganisms; 2 Food Commodities*. London: Academic Press.

International Commission on Microbiological Specifications for Foods (1998) *Microorganisms in Foods: 6 Microbial Ecology of Food Commodities*. London: Blackie Academic and Professional.

Lancet Review (1991) *Foodborne Illness*. London: Edward Arnold.

Morsel DAA, Corny JEL, Struijk CB and Baird RM (1995) *Essentials of the Microbiology of Foods*. Chichester: John Wiley and Sons.

Roberts D (1986) Factors contributing to outbreaks of foodborne infection and intoxication in England and Wales 1970–1982. *Proceedings of the Second World Congress of Foodborne Infections and Intoxications*, vol. 1, pp. 157–159. Berlin: Institute of Veterinary Medicine.

Tracing Origins and Testing

A C Baird-Parker, Northampton, Hardingstone, UK

Introduction

Foodborne disease (food poisoning) has been defined by the World Health Organization (WHO) as: 'a disease of infectious or toxic nature, caused by, or thought to be caused by, the consumption of food and water.' It involves a large (>40) and diverse range of agents, the number of which continues to increase yearly as new agents are identified. Table 1 lists those agents currently considered to be involved in disease transmitted by foods.

The main sources of many of these organisms are domestic and wild animals (including fish and birds). Enteroviruses, *Salmonella typhi*, and *Vibrio cholerae* are of human origin and *Clostridium botulinum, Bacillus* spp., and mycotoxigenic molds are general environmental contaminants. (Refer to individual microorganisms.)

Food poisoning is usually thought of as acute gastroenteritis (fever, abdominal pain, and diarrhea, often with vomiting) of the type caused by infectious disease-causing bacteria such as *S. enterica* and toxin-forming bacteria such as *Staphylococcus aureus*. However, a number of the newly identified (often called 'emerging') food-poisoning microorganisms cause a wide range of nonintestinal symptoms. For example, *Listeria monocytogenes* causes illness ranging from a mild flu-like illness to meningitis and life-threatening infections of the central nervous system. In addition, it has been recognized within the last 20 years that a number of classical gastroenteritis-causing bacteria, such as salmonellas and *Yersinia enterocolitica*, may cause extraintestinal infections as sequelae to intestinal disease; these include reactive arthritis, Reiter's syndrome, ankylosing spondylitis, and pericarditis. The wide range of symptoms caused by foodborne disease-causing agents, many of which mimic illness of a nonfood-poisoning nature, may cause some difficulty in identifying a food cause and establishing the extent of a foodborne disease incident.

This article considers the epidemiology of foodborne diseases, the procedures used for testing foods for the causative agents, tracing the source and cause of a food-poisoning incident, and the recall of contaminated food.

Table 1 Foodborne disease-causing agents

Bacteria
Aeromonas hydrophila and *veronii* subsp. *sobria*
Bacillus cereus, B. licheniformis, and *B. subtilis*
Brucella melitensis, B. abortus, and *B. suis*
Campylobacter jejuni, C. coli, C. lari, and *C. upsaliensis*
Clostridium botulinum and *C. perfringens*
Escherichia coli (EIEC, ETEC, VTEC, EAggEC)
Listeria monocytogenes
Mycobacterium bovis
Salmonella enterica
Shigella dysenteriae, S. flexneri, S. boydii, and *S. sonnei*
Streptococcus zooepidemicus
Vibrio cholerae, V. parahaemolyticus, V. fluvialis, and *V. vulnificus*
Staphylococcus aureus
Yersinia enterocolitica
Mycotoxigenic molds
Protozoa
Cryptosporidium
Cyclospora
Giardia
Toxoplasma
Prions
vCJD
Viruses
Hepatitis A and E
Small round structured viruses (SRSV)
Dinoflagellates
Shellfish poisoning (PSP, DSP, NSP, ASP)

Epidemiology

When considering the sources of foodborne disease-causing microorganisms it is important to recognize that virtually all raw agricultural products will from time to time be contaminated with low concentrations of such organisms and, as a result of cross-contamination and recontamination, they will also contaminate cooked and other processed food products. Although salmonella are regarded to be zoonotic, and thus of animal origin, they are frequently found in soils and water, and human carriers may be the source. (*See* **Epidemiology; Zoonoses**.) Organisms of human origin, such as the previously mentioned enteroviruses and *Salmonella typhi*, may also contaminate various parts of the food chain, e.g., via human carriers contaminating food during preparation for consumption, or from human sewage contamination of foods during their growing, e.g., the contamination of vegetables with night soil or molluscan shellfish harvested from waters contaminated with raw human sewage.

Thus the original source of an agent causing foodborne disease may be difficult or impossible to establish with any degree of certainty and it may only be possible, and in many cases only appropriate, to identify the circumstances and actual foodstuff that caused a particular food-poisoning incident. In many cases this will be the result of poor food handling (including

preparation in advance for consumption) allowing survival or growth of the pathogen. (*See* **Food Poisoning**: Statistics.) For example, whereas the main source of the contamination of foods with salmonellas and campylobacters is poultry, provided the poultry is properly cooked, these bacteria will be killed and if the cooked poultry is then protected from recontamination they will not cause illness. Thus, while the source of salmonella contamination of cooked poultry is most likely to be the raw poultry, the cause of any salmonellosis arising from its consumption is most likely to be the result of poor hygienic practice during preparation. In this example, identification of the source of the original contamination is much less important than identifying the circumstances that led to the contamination of the cooked poultry. However, where contamination of a raw ready-to-eat food occurs, it is important to identify the source of the contamination to prevent its recurrence. (*See* **Poultry**: Chicken; Ducks and Geese; Turkey.)

Further, when considering the cause of a foodborne disease incident it is important to recognize that such illness is dose-related (the larger the concentration of the organism or the amount of toxin present in a food at the time of consumption, generally the greater the severity and extent of illness within a group of persons consuming the contaminated food). It is seldom that all persons exposed to an infective or intoxicating dose of a pathogen will show disease symptoms and symptoms will often vary widely in those who are ill. The minimum concentration of an agent necessary to cause illness is frequently referred to as the minimum infective dose or MID. The MID of many food-poisoning bacteria is unknown but it is generally accepted that the infective dose of classical causative agents of enteric disease, such as campylobacters, *Salmonella typhi*, and *Shigella*, is 10–100 cells, whereas the MID of others such as the food-poisoning salmonella (now all classified as *Salmonella enterica*) is generally regarded to be of the order of 100 000–1 million cells for a healthy adult. However, the infective dose and hence the risk of illness arising from a contaminated food is affected by many factors. Factors that may reduce the infective dose include: high virulence of a particular strain; age, status of health of individual exposed to the agent – in particular immune status, nutritional deficiencies in diet, and underlying disease. There is much evidence that there are in some foods substances that protect infective agents from destruction by stomach acids. Thus fats in high-fat foods such as full-fat cheeses, chocolates, and salamis may significantly reduce the infective dose of an agent; for instance, the infective dose of salmonellas may be only a few cells (< 10) in these foods. Therefore, when considering the significance of a particular concentration of an agent in a food, in relation to it having caused illness, it is important to take these facts into consideration and also to remember that in any food collected after a food-poisoning incident, numbers of microorganisms may have increased or decreased depending on how that food has been stored subsequent to the incident. With respect to microbial toxins these are often more persistent in foods than their producing organisms but may increase in amount as the result of the presence of the producing organism, or decrease as a result of biological, chemical, or physically induced breakdown.

Case Histories

Hepatitis A and Frozen Raspberries

Twenty-one persons who consumed a mousse made from frozen raspberries developed jaundice found to be caused by the hepatitis A virus. Investigation of the incident suggested that contamination had occurred during harvesting or packaging of the raspberries. Cases of hepatitis had been reported in the area where the raspberries had been picked and the sanitary conditions in the places where the pickers were housed were found to be poor. Sanitary conditions were improved and basic food hygiene instruction was given to the pickers. There were no more reported incidents from this source.

Salmonella and Chocolate

The uncommon salmonella serotype *Salmonella napoli* caused 245 cases of salmonellosis amongst children. These were traced to the consumption of contaminated chocolate bars from a single factory. The factory was very modern and clean but untreated river water was used for the water circulation system which kept the chocolate molten during transportation around the factory; *S. napoli* had previously been isolated from river water in the region of the factory. Contamination of the chocolate probably occurred as a result of microleaks in the pipes of the water circulation system, resulting in the seepage of water into the chocolate. Chocolate manufactures agreed to a code of hygienic practice following this incident.

Salmonella and Salami Sticks

A number of cases of salmonellosis, mainly amongst infants and children, occurred as a result of the consumption of salami sticks. The salami sticks were manufactured using traditional methods but with a very short maturation period. The short maturation period was not sufficiently long to destroy any

salmonella present in the meat used in the manufacture of the salami. Thus very small numbers survived. This survival was prolonged by the practice in retail outlets in the UK of storing the sticks in chill cabinets; no cases were reported in continental Europe where the sticks were stored at ambient temperatures. The manufacturers reviewed their manufacturing procedures, introducing a heating step in the process, thus guaranteeing the destruction of any salmonella that might survive the usual salami-manufacturing process.

Staphylococcus aureus and Pasta

Epidemiological and microbiological tracing attributed 47 cases of staphylococcal food poisoning, occurring in several European countries, to a single source of lasagne. The source of the problem was traced back to contamination of the eggs used in the manufacture of the pasta and was believed to be associated with the hygiene of equipment used to transport the eggs. As a result of this incident a code of practice was drawn up by the pasta industry requiring the use of pasteurized eggs in pasta manufacture and improved hygienic design of the equipment used for the preparation of the pasta.

Clostridium botulinum and Yogurt

After 27 cases of botulism were traced to the consumption of a locally produced yogurt containing hazelnuts, the source was identified as canned hazelnut purée; this was part of a batch of low-calorie products prepared using artificial sweetener. Unfortunately the manufacturer of the hazelnut purée had not recognized that the heat process used for commercially sterilizing the normal high-sugar product was not suitable for the low-calorie product. Sugars reduce the available water (a_w) to a level at which spores of *C. botulinum* are unable to initiate vegetative growth and are thus unable to grow and form botulinum toxin; the organism grew well in the sugar-free product that had a much higher a_w. The outcome of this incident was to inform producers of these and other products of the need to obtain expert advice on the required heat process before reformulating a canned food. Also, the national code of practice for low-acid canned foods was revised to clarify all requirements for the production of microbiologically safe low-acid canned foods.

Testing Foods for Foodborne disease-causing Microorganisms

Foods may be tested for foodborne pathogens for a variety of reasons; the main reasons are listed in Table 2.

Microbiological testing of foods has been widely applied by health authorities and industry as a means of identifying foods that have levels of microorganisms or their toxins in excess of those regarded to be acceptable and therefore are unsafe or unacceptable for human consumption for other reasons. Such testing was at one time believed to give adequate protection to the consumer and was the linchpin of microbiological control strategies practiced by governments and industry for much of the twentieth century. This was despite the recognition more than 30 years ago that such testing gave little or no assurance of microbiological safety or quality. This can be illustrated by reference to the most stringent testing plan applied internationally for examining baby food for salmonella. In this plan, a sample comprising 60 individual sample units is taken from a batch of baby food and from each sample unit an analytical sample unit of 25 g is taken and examined for salmonella (total analytical sample is thus 1500 g); the pass criterion is no detection of salmonella in any of the analytical units examined. Even when this pass criterion is met, there is 50% probability that up to 1% of the sample units in the food batch will be contaminated. Clearly this degree of confidence is not acceptable, particularly when considering that such a sampling plan may be applied to test the acceptability of a baby food.

It is such considerations that have led to a total reappraisal in recent years of the value of such testing as a means of assuring the microbiological safety of a foodstuff. This has led to the wide-scale introduction of Hazard Analysis Critical Control Point (HACCP)-based control systems whereby the potential hazards associated with manufacture of a food in a particular processing plant are identified using a structured hazard analysis procedure and locations where these hazards can be effectively controlled are identified. These locations are called Critical Control Points and means for their proper control (including means of monitoring that control is maintained) are identified. (*See* **Hazard Analysis Critical Control Point**.)

Table 2 Uses of microbiological testing

Investigational testing	Suspect contaminated food
Compliance testing	Governmental regulations; industry standards and purchasing specifications
Monitoring	Authorities monitoring imported and domestic foods; industry monitoring products and processes
Verification	Good manufacturing practice and Hazard Analysis Critical Control Point

As a result of the introduction of HACCP as a prime means of assuring the microbiological quality and safety of foods, microbiological testing has become less important as a primary control function and increasingly is used mainly by industry for the purpose of verification of HACCP. However, where HACCP is not applied, the industry or a health authority may apply microbiological testing if there is a need to obtain a judgment about the acceptability of a foodstuff. For instance, it may be appropriate for use in a buying specification for a microbiologically sensitive raw material from a new supplier or by a health authority on a consignment of imported food, where microbiological testing is the most appropriate acceptance criterion. For such a purpose, it is important that statistically based sampling plans are used so that the degree of confidence in the result obtained is known; methods of analysis should also be agreed. One of two types of sampling plans may be used for such acceptance sampling. The first and most widely used are the two- and three-class attribute plans established by the International Commission on Microbiological Specifications for Foods. Such plans were developed principally for testing foods in international trade for their acceptability, but are also widely used by industry in buying specifications.

Two-class plans have a single limit which distinguishes between an acceptable and unacceptable concentration of a microorganism or toxin and is used in sampling plans for foodborne pathogens.

Three-class plans are used principally for determining acceptability of levels of general microbial contaminants and indicator organisms and have two limits: one is a lower limit generally established by good manufacturing practice (GMP) requirements, while the other is an upper limit based on some index of unacceptable quality. The other main type of plan used is the standard variables plan which is based on knowledge of microbial distributions in batches of foods and can be used when such distributions approximate to log normality. This limits their use principally to industry, where their advantage over attribute plans is greater accuracy and hence greater confidence in decisions based on their use.

The choice of a sampling plan and testing procedures for investigative sampling which may, for example, be used to identify the level of contamination in a suspect consignment of food, or to investigate food involved in a food-poisoning incident, is more difficult than acceptance sampling. One major problem is that there will usually be no information on either the likely incidence of contaminated product units (or the level of contamination); these factors set requirements for the number of units of product that should be examined and the size of the analytical unit that should be tested. In investigative sampling, unlike acceptance sampling, random sampling may not be the most appropriate procedure. For example, the location of the product units in relation to possible source of the contamination may suggest that cluster or stratified sampling may be more appropriate.

Tracing the Origins of a Foodborne Disease Incident

Investigation of the cause of an outbreak or case of foodborne illness will usually require a number of procedures depending on the type and circumstances associated with the incident. These may include extensive testing of food and clinical samples for likely causative agents; the use of epidemiological analytical techniques to link illness to a common source if a number of cases of illness are involved; field investigations by public health and industry experts to find the cause; and specialist food-processing knowledge to identify what measures need to be taken to prevent its recurrence.

The tracing of the source and cause of an outbreak of foodborne disease associated with a well-defined event, such as a banquet, is usually a relatively easy matter. A list of foods consumed by the ill persons is compared against a list of items consumed by persons who are not ill, and using simple statistical correlation procedures, any significant differences in foods consumed by the two groups are identified and further investigated to find the most likely food source. Tracing the source of sporadic cases in a community is much more difficult and, as well as extensive use of well-founded epidemiological procedures, investigators need a certain amount of serendipity, backed up by specialist laboratory resources, in order to be successful.

Laboratory Investigations

Samples of suspect foods, together with clinical samples, should be collected and transported under conditions that protect them from external contamination and minimize changes in the concentration of any food-poisoning agent that may be present, e.g., in sealed sterile containers in a suitably cooled cool box. Details of the history of the sample, e.g., where and when sampled and by whom and how, together with information on symptoms of illness should be submitted to the laboratory as this will help in selecting the most appropriate method of analysis. On receipt by the laboratory the samples should be documented, properly stored, and examined as soon as possible. Such examinations must be undertaken by laboratory staff who have been trained in the microbiological

examination of foods as well as clinical specimens; they should be particularly knowledgeable of the analytical procedures to be applied. These should be performed in accredited laboratories operating to recognized good laboratory practice (GLP) standards and using procedures and methods that have been approved by an appropriate accreditation organization. The microbiological examination of a foodstuff can be problematic and a number of pitfalls may trap the unwary or careless. For instance, examination of food specimens in the same room where clinical specimens are examined is bad practice; a heavy microbial load in a clinical specimen can easily cross-contaminate a food sample, where only light contamination with the same organism in the clinical specimen is sufficient evidence to implicate a food as the source.

Initial laboratory findings should always be interpreted with care. Isolates and where appropriate original specimens should be submitted to a reference laboratory with expertise in the suspect agent of concern in order to confirm an initial finding and, if appropriate, to apply modern 'fingerprinting' and typing methods such as DNA-based subtyping methods. There are many examples of foods being erroneously implicated in a food-poisoning incident because of poor laboratory practice. There have also been many examples where failure to identify a causative agent, or a food source, has been the result of failure of the laboratory to apply the correct isolation or detection method. The microbiological examination of foods is not an exact science, although continued improvements in detection techniques in recent years have considerably improved our ability to detect pathogenic agents in foodstuffs.

Epidemiological Investigations

Two main analytical tools are used by epidemiologists to trace the source of a foodborne disease incident: case-control studies and cohort studies. For both, the first stage is to review and document all available information. In an investigation of an outbreak, a questionnaire is prepared to detail when and what was consumed and symptoms of illness. This questionnaire is used in preliminary interviews with persons affected by the outbreak to obtain basic information as to the likely cause. A case definition is also agreed, i.e., what symptoms should be searched for, as it is important when case searching (in order to identify the extent of an incident) that only the most likely cases are identified.

In a case-control study a series of questions based on the preliminary questionnaire are asked of the persons identified as corresponding to the case definition (case group) and persons who have not been ill (control group). These two groups are matched as far as possible with respect to age, sex, socioeconomic group, and place of work or residence. Controls are often based on consultation with the general practioners of the persons who have been ill. Based on information in the completed questionnaires, differences between items consumed by the ill and well group are identified and any statistically significant differences subjected to further investigation. One problem with this procedure is that it is common in food-poisoning incidents to have symptomless cases and these may well be mistakenly identified as part of the control group.

In the analysis of the case data the following are the most important.

1. Time to onset of illness. Based on knowledge of the incubation period of the illness it is possible to determine the period of exposure to infection. This information may then be plotted as a histogram relating the frequency of cases chronologically to date of onset of illness. This so-called 'epidemiological' or 'epidemic' curve can then be used to identify whether or not there has been secondary transmission of illness and whether or not it is a point-source outbreak, e.g., a single meal, or a common-source outbreak, e.g., a number of cases from the same source, such as a retail establishment, occurring over a particular time period.
2. The place where infection occurs. A map is drawn of known cases, and researchers look for relationships, such as clusters, which may assist in identifying a source.
3. The persons involved. Analysis of information on individual cases such as age, sex, medical history, occupation, home situation and any travel history may also help one to form hypotheses as to the source of infection.

In a cohort study, a group of people is identified by some other means than illness. For example, this might be all the persons who attended a conference at which it is suspected that a meal was consumed that led to illness. The persons are divided into two groups according to whether or not they ate a particular meal; the attack rate in the group eating the meal and that of the group not eating it is calculated. If there is found to be a statistically significant correlation for the consumption of a particular meal, the next stage is to determine the attack rates for persons eating and not eating a particular menu item. By this means, if successful, a particular menu item will be identified as the source of the agent that caused the illness.

Environmental Health Investigations

As part of an investigation into a foodborne disease incident, health inspectors may visit and investigate the identified source, e.g., the place where a foodstuff implicated in a food-poisoning incident was manufactured or prepared for consumption. This may involve a detailed examination of the hygienic practices, and the collection of product and environmental samples for laboratory examination. If an infectious disease is suspected, food handlers may be asked to submit samples for microbiological examination. Where a source of contamination has been identified, the health authority is most likely to require assurance that the condition that caused the contamination has been corrected prior to allowing the product to be returned to the market place. Such assurance may require evidence that checks have been done by a third party to confirm that satisfactory control measures are in place. If a HACCP plan is in place this plan will need to be revalidated.

Withdrawal of a Suspect Food

If a distributed food is found by a health authority or a producer to be contaminated with a concentration of a pathogen or toxin exceeding a legal limit or exceeding a level generally regarded to be hazardous to health, this food should be withdrawn from sale as soon as possible and if necessary the public informed.

Before announcing the public withdrawal of a food it should be ascertained, as far as is technically possible, that this food is unsafe to eat. Wherever possible, results of microbiological analyses or evidence clearly linking cases of illness to an implicated food item should be available to the health authority and/or producer, as adverse publicity is likely to result from a public recall which may cause much damage to the reputation of the producer. The prudent producer may not wish to run the risk of even worse publicity arising from an incident occurring as a result of failure to inform the public and may voluntarily withdraw the product and inform the public of this.

Authorities will take different actions concerning the extent of a recall depending on a judgment as to the health risk involved by allowing food to remain on sale or in the hands of a consumer. Thus a recall may only be required down to the distribution level, e.g., distribution warehouse, if it is judged that any food remaining on sale is unlikely to cause further illness. A similar decision may well be made for a food containing a level of microorganisms exceeding a legal limit but where no illness has been reported. However, if it is judged that serious illness could result from product on sale or in the hands of consumers, recall down to retail level would be the minimum requirement and a public warning (via the media) would usually also be required. It is important that a recall is properly coordinated by the owner or supplier of the food; for this to be effective a food company should have a documented recall procedure in place. As part of this procedure a recall coordinator should be appointed whose role is to coordinate the recall, to liaise with the public and the authorities on progress with the recall, and to act as the spokesperson in contact with the media.

When carrying out a public recall it is important that this is done as fast and thoroughly as possible. Thus there will be the need for press statements via the media and communication to customers (and retailers) via e-mail or fax, advising them of the recall. In the case of a national or international recall the health authorities will inform their contacts via established networks and if a particular ingredient is involved may work with the supplier to identify and advise producers who may be using this ingredient in further products. Effective coding of food batches and maintenance of records of distribution are essential requirements as production failures will inevitably occur from time to time.

See also: ***Bacillus***: Food Poisoning; ***Clostridium***: Occurrence of *Clostridium botulinum*; Botulism; **Epidemiology**; **Food Poisoning**: Statistics; **Hazard Analysis Critical Control Point**; ***Listeria***: Listeriosis; **Poultry**: Chicken; Ducks and Geese; Turkey; ***Salmonella***: Properties and Occurrence; **Vibrios**: *Vibrio cholerae*; ***Yersinia enterocolitica***: Properties and Occurrence; **Zoonoses**

Further Reading

Baird-Parker AC and Tompkin RB (2000) Risk and microbiological criteria. In: Lund BM, Baird-Parker AC and Gould GW (eds) *The Microbiological Safety and Quality of Food*, vol 11. pp. 1852–1885. Gaithersburg, Maryland: Aspen.

CAST (1994) *Foodborne Pathogens: Risks and Consequences*. Task Force report no. 122. Ames, Iowa: Council for Agricultural Science and Technology.

FAO/WHO (1998) *Codex Alimentarius*, Vol. 1B. *General Requirements (Food Hygiene) Supplement*. Rome: Food & Agriculture Organization.

International Commission on Microbiological Specifications for Food (1986) *Microorganisms in Foods 2. Sampling for Microbiological Analysis: Principles and Specific Applications*, 2nd edn. Toronto: University of Toronto Press.

International Commission on Microbiological Specifications for Food (2002) *Microorganisms in Foods 7. Microbiological Testing in Food Safety Management*. New York: Kluwer Academic/Plenum.

Jarvis B (2000) Sampling for microbiological analysis. In: Lund BM, Baird-Parker AC and Gould GW (eds) *The Microbiological Safety and Quality of Food*, pp. 1691–1733. Gaithersburg, Maryland: Aspen.

Palmer SR (1990) Epidemiological methods in the investigation of food poisoning outbreaks. *Letters in Applied Microbiology* 11: 109–115.

Sharpe JCM and Reilly WJ (2000) Surveillance of foodborne disease. In: Lund BN, Baird-Parker AC and Gould GW (eds) *The Microbiological Safety and Quality of Food*, vol. 11, pp. 975–1011. Gaithersburg, Maryland: Aspen.

WHO (2001) *WHO Surveillance Programme for Control of Foodborne Infections and Intoxications. 7th Report 1990–1992*. Berlin: Federal Institute for Health Protection of Consumers and Veterinary Medicine.

WHO (1997) Food safety and foodborne diseases. *World Health Quarterly* 50: 3–154.

Statistics

S Palmer, University of Wales, College of Medicine, Cardiff, UK

Background

The public health significance of food-poisoning statistics can be properly understood only when set in the context of public-health surveillance.

Modern public-health surveillance of disease has been defined as the continued watchfulness over the distribution and trends of incidence through the systematic collection, consolidation, and evaluation of morbidity and mortality reports and other relevant data, and applying these data to prevention and control.

Surveillance needs to be distinguished from research. The former focuses on problem detection and characteristics, whereas research is mainly to do with hypothesis testing. Surveillance systems generate hypotheses and should not be expected to give detailed answers. Surveillance data provide information for action and as such should stimulate investigation. Emphasis has to be on speed of detection of a potential problem rather than full and accurate documentation. Consequently, great care has to be exercised in interpreting surveillance statistics for food poisoning. Surveillance systems by definition require ongoing collection of data and therefore to be sustained over time have to rely on minimal data which are often incomplete. National surveillance systems have often gone into decline because data collection has been driven by a desire for full documentation rather than the overriding requirements of timeliness and sensitivity to trends.

Specific objectives of foodborne disease surveillance include:

1. early detection of clusters or outbreaks of foodborne disease or new exposures and other risk factors, to trigger rapid investigation and control;
2. measuring trends in microbial agents and risk factors in order to set priorities for interventions, and to evaluate foodborne disease control programs; and
3. to describe the basic epidemiology of foodborne disease, such as its geographical spread and the age distribution of cases in order to develop hypotheses about causation, which can be tested by separate research studies.

Methods of Surveillance

The main steps in surveillance are

- systematic collection of data;
- analyses of these data to produce statistics;
- interpetation of the statistics to produce information;
- distribution of this information to all those who require it so that action can be taken; and
- continuing surveillance to evaluate the action.

Data Collection

Surveillance is, by definition, an ongoing activity and can be sustained only when the burden placed on the data providers is light. Surveillance data should be limited to the minimum required to meet its specified objectives. Reporting methods should be simple and streamlined. Successful systems have been developed using modern information technology. Electronic data collection should be linked to electronic systems for dissemination of high-quality surveillance information.

Statistical Analysis

Usually, the routine analysis of surveillance data is simply the presentation of incidence rates by time, place, and person using graphs, histograms, and maps. However, more sophisticated methods are increasingly being used. Particular statistical issues include the use of time series analysis to model epidemics, the early recognition of unusual events in routine data against a variable baseline rate, small area analysis of clustering, adjustment for delays and incompleteness of reporting, and the use of

surveillance data in risk assessment models and in prediction models to evaluate the consequences of possible interventions.

Reporting

Timely reporting to those responsible for public-health action is an essential part of a bona fide surveillance system. Timeliness is defined by the objectives of the surveillance. For foodborne disease, timely reporting may need to be measured in hours so that contaminated food can be removed from the marketplace and warnings given to consumers, a target which can now be achieved globally through the internet. Typically, foodborne disease surveillance reports appear either as specifically produced publications (e.g., the Communicable Disease Report), or as electronic bulletins.

Sources of Data

Foodborne disease surveillance relies on two main sources of statistics; the statutory notifications of food poisoning and the voluntary reporting of laboratory isolations.

Statutory Notifications

In England and Wales, mandatory notification of infectious disease was introduced nationally in 1899. All registered medical practitioners diagnosing certain diseases have a duty by statute to report to the proper officer of the local authority so that suspected cases can be investigated and control measures taken as appropriate. Currently, food poisoning is notifiable under the Public Health (Control) Act 1984. In England and Wales, weekly summaries of these data are now published in the Public Health Laboratory Service (PHLS) Communicable Disease Report (CDR). The data are later corrected and published quarterly and annually by the Office for National Statistics. In addition to those cases formally notified by medical practitioners, local authorities identify other suspected cases informally. The notification system now has the flexibility to record these 'otherwise ascertained' cases which appear separately in reports.

It is important to note that 'food poisoning' is a notifiable disease, although the bacterial causes of food poisoning are not in themselves notifiable. This has led to some confusion. The Advisory Committee on the Microbiology Safety of Food proposed a working definition of food poisoning which was accepted by government departments and circulated to all doctors in the UK in 1992. Its definition of food poisoning is 'any disease of an infectious or toxic nature thought to be caused by the consumption of food or water.' This definition had previously been adopted by the World Health Organization.

Nevertheless, the way in which this definition is applied is subject to various interpretations. For example, verocytotoxic *E. coli* (VTEC) disease is quite likely to be foodborne, even though person-to-person spread and spread from direct contact with animals and the contaminated environment are significant problems. If the source of infection is unknown, a doctor may or may not suspect foodborne disease. If the source is suspected to be direct contact, technically the disease is not food poisoning, even though reporting to statutory agencies might alert them to a foodborne disease risk. It seems prudent, therefore, that suspected and confirmed VTEC infection, together with all *Salmonella* and *Campylobacter* species, however acquired, should be notified as a suspected foodborne disease.

The purpose of notification is to alert local authorities to take appropriate action to investigate and control incidents; data for surveillance are a secondary benefit. However, it is generally accepted that food-poisoning notifications are a poor indication of the true incidence of food poisoning, even though they may provide useful information on trends. This is because people with the symptoms of food poisoning often do not visit their general practitioners (GPs). When a GP does see a patient with foodborne disease, it is usually not possible to distinguish this from other more common causes of gastroenteritis. In individual cases, the food history is seldom helpful. Only when a feces sample yields a foodborne pathogen is a measure of confidence obtained. However, from a clinical point of view, there is usually little, if any, benefit to the patient from the results of a feces sample, and therefore, GPs do not routinely send samples to the laboratory. When they do, it tends to be because of the severity or persistence of symptoms.

Figure 1 shows that in England and Wales, notifications, ascertained both formally and otherwise, increased greatly from 1982 to 2000. Interpretation of these data is difficult since over this same period, considerable efforts have been made in the National Health Service and local authorities to improve reporting. Inevitably, this means increasing the numbers of cases notified. It is therefore unclear how much of the increase is due to better reporting and how much is due to a real increase in the incidence of food poisoning in the community.

Laboratory Reporting of Microbiological Data

Reports by diagnostic laboratories provide the backbone of infectious disease surveillance in the UK. In England and Wales, the PHLS developed reporting

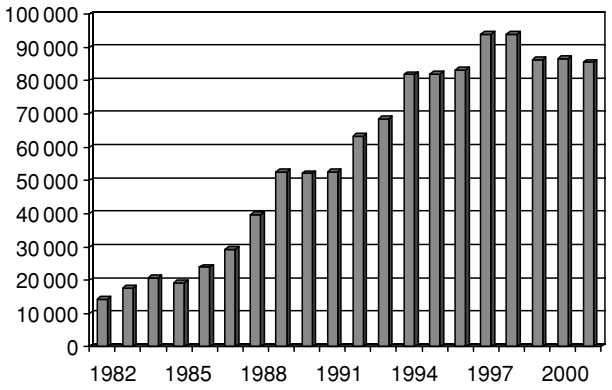

Figure 1 Notifications of food poisoning in England and Wales.

in the 1940s and 1950s. Data are analyzed within a week of receipt by the Communicable Disease Surveillance Centre (CDSC) to produce tables and line lists, which are used in compiling narrative reports for publication to the CDR. The recent introduction of OSURV (business name of electronic reporting system) has substantially replaced manual and electronic reporting.

The PHLS receives reports from over 200 clinical and public health laboratories and also uses the data held by the PHLS Laboratory of Enteric Pathogens, which is the national reference laboratory for *Salmonella*, *Campylobacter*, and VTEC infections. In Scotland, the SCIEH (Scottish Centre for Infection and Environmental Health) receives laboratory reports. Although these data are highly specific because they are based on the most comprehensive diagnostic tests, the number of reports received represents only a small fraction of the true incidence of these pathogens because of both the low level of microbiological investigation in general practice and incomplete laboratory reporting. Information accompanying these reports is limited, and there are no data on the outcome of infected patients. The data do not include patient identifiable details, so the area of residence cannot be mapped.

The main benefits of laboratory reports are that they are highly specific, since they are based on laboratory-diagnosed infections and the fine typing of the infecting organism, they often include clinical and epidemiological details, and they allow free-text comment. The reporting system is flexible, and unusual or new infections can be reported, even though they have not been included in the original reporting instructions. However, the reports are limited to infections for which there is a suitable laboratory test.

Figure 2 shows that reports of *Salmonella* infection in England and Wales tripled from 1981 to 1990. From 1990 to 1997, the high level of reports was maintained. From 1998 to 2000, *Salmonella* reports fell steadily, but in 2001, provisional data suggest an increase. These data correspond to the epidemic of *Salmonella enteritidis* that affected egg production internationally. The effect of the control measures, especially the vaccination of laying flocks, is thought to explain the sudden reversal of the epidemic in 1998.

Figure 3 shows the trends for *Campylobacter* infection in England and Wales. The steady increase in reports in the early years of its recognition could be due to increasing numbers of laboratories involved in testing and reporting. The sensitivity of laboratory methods has continued to improve, but these data are also compatible with a real increase in incidence of the infection in the community.

Figure 4 shows data for *E. coli* O157. In the UK, the incidence of laboratory reported cases rapidly increased through the 1980s following the first report in 1982. This was certainly due to increasing numbers of laboratories introducing screening tests for VTEC.

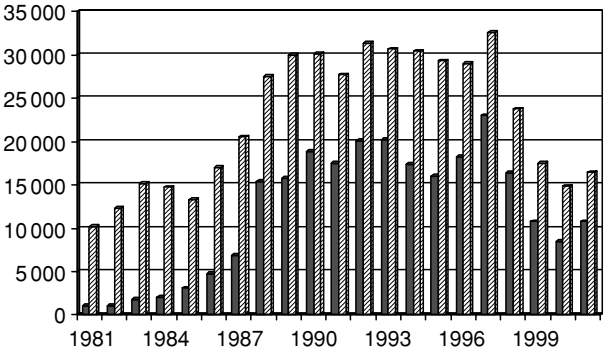

Figure 2 Reports of *Salmonella* infection in England and Wales. ■ : *Salmonella enteritidis*; ▨: total Salmonellas.

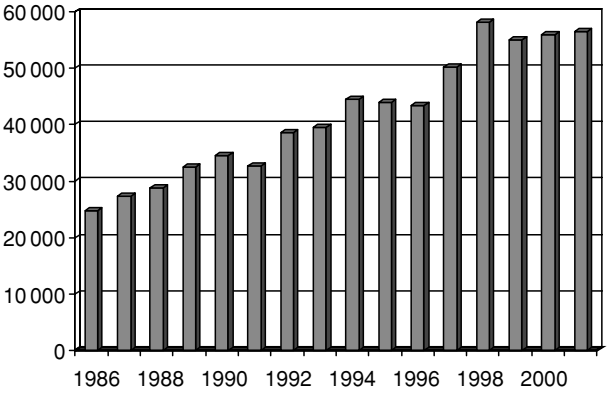

Figure 3 Reports of *Campylobacter* infection in England and Wales.

Throughout the 1990s, the annual number of cases in England and Wales fluctuated, but there is an overall increasing trend. By 1997, the number of reported cases was over 1000.

Reports from Scotland are far higher than those from England and Wales. In part, this may be due to its higher profile there. For example, in 1989, *E. coli* O157 was one of a list of 32 informally selected organisms which were made 'reportable infections' by laboratories in Scotland, in addition to the statutorily notifiable diseases, and this coincided with a doubling (at least) of the recorded incidence in a 12-month period. However, during 1992 and 1993, and against the background of heightened awareness and increased screening, the incidence dropped to about half of the 1991 level (2.3 per 100 000 compared to 4.0 in 1991). The subsequent marked increase, in 1994, also cannot be attributed to changes in reporting rates or screening, suggesting that Scotland does have a higher incidence than the rest of the UK and that there are real significant variations in the incidence of VTEC from year to year.

Figure 5 shows the trends in SRSV infections. Viral gastroenteritis is much more common than these data suggest. Most cases are not investigated through laboratories. When they are, diagnosis requires electron microscopy of feces taken within the first hours of the onset in order to have a high chance of success. Viral gastroenteritis can be foodborne from contamination of any food by dirty hands, or airborne contamination following vomiting or from food contaminated from sewage such as oysters.

Figure 6 shows the laboratory reports of *Listeria* infection, which is usually foodborne in origin. People affected are almost exclusively the elderly, pregnant women, or immunosuppressed patients. The peak in cases in 1987–89 coincided with an outbreak from pâté produced in Belgium and led to health advice for the public and considerable work by the food industry to address the risks from cook-chilled foods.

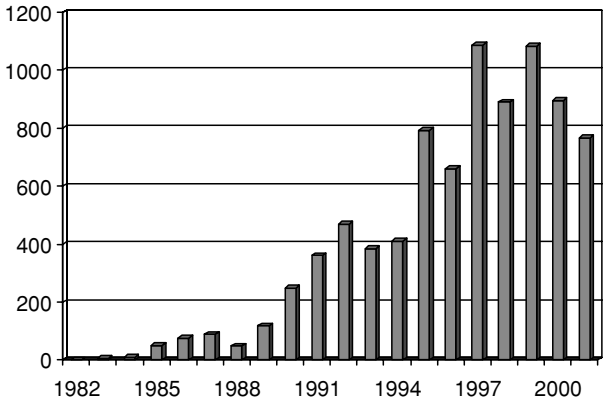

Figure 4 Reports of *E. coli* O157 in England and Wales.

The Reporting Pyramid

In order to gain a better appreciation of the true incidence of gastrointestinal infections, a national study, the Infectious Intestinal Disease (IID) in England study, was commissioned by the Department of Health. This provides the best data to date of the true incidence of enteric pathogens in the country. The objectives of the study were: first to estimate the number, and microbiological causes, of cases of intestinal infectious disease people presenting to their GPs and from whom stool specimens are routinely sent for laboratory examination; second, to compare these data with the data from the national reporting surveillance system; and third, to estimate the prevalence of asymptomatic infection with these agents.

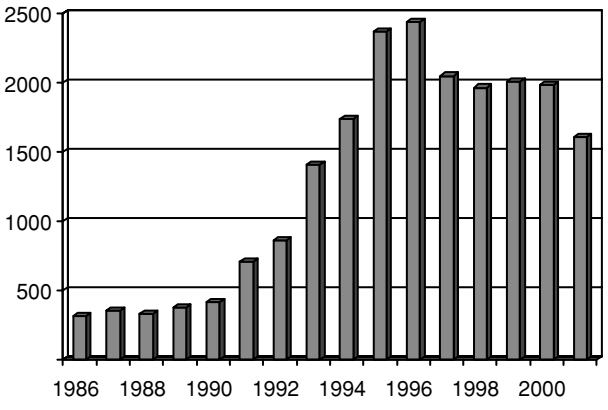

Figure 5 Reports of SRSV infections in England and Wales.

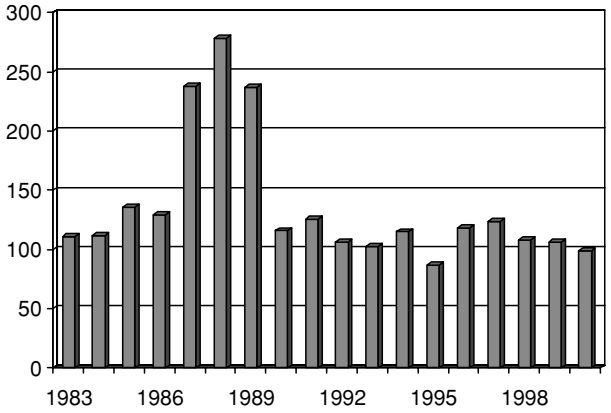

Figure 6 Laboratory reports of *Listeria* infection in England and Wales.

Seventy practices were selected to be representative of the socioeconomic characteristics of the area and to represent an urban and rural mix. Cases of infectious intestinal disease were defined as persons with loose stools or significant vomiting lasting less than 2 weeks in the absence of a known non-infectious cause and preceded by a symptom-free period of 3 weeks. Vomiting was considered significant if it occurred more than once in the 24-h period and if it incapacitated the case or was accompanied by other symptoms such as cramps or fever. One component of the study was to draw at random a cohort of people registered with the 70 GPs, who were followed up for a period of 6 months. These volunteers agreed to fill out diary cards every week and return them to the GPs. These cards stated whether the person had suffered gastrointestinal (GI) illness or not. If someone developed an illness, a stool sample was submitted.

Of particular interest in this study was the estimate of the carriage of potential pathogens in otherwise healthy people. Ten out of 2264 controls (0.4%) were found to be excreting *Salmonella*, 16 out of 2264 (0.7%) were excreting *Campylobacter*, but none out of 2264 were excreting *E. coli* 0157. These data confirm that VTEC is a rare infection in this country at the moment.

The IID study was able to quantify the reporting pyramid referred to above using the cohort component of the study. For all IIDs, it was estimated that for every case reported to the CDSC, six patients were investigated by routine laboratory tests, 23 presented with GI symptoms to their GP, and there were 136 actual cases in the community. For *Salmonella* cases, however, the ratios were much smaller. For every case reported to CDSC, it was estimated that there were 2.3 cases presenting to their GP and 3.2 actual cases in the community (see Figures 7 and 8).

Outbreak Reports

In addition to the reporting of individual cases, national surveillance centers receive reports of foodborne disease outbreaks. A general outbreak is

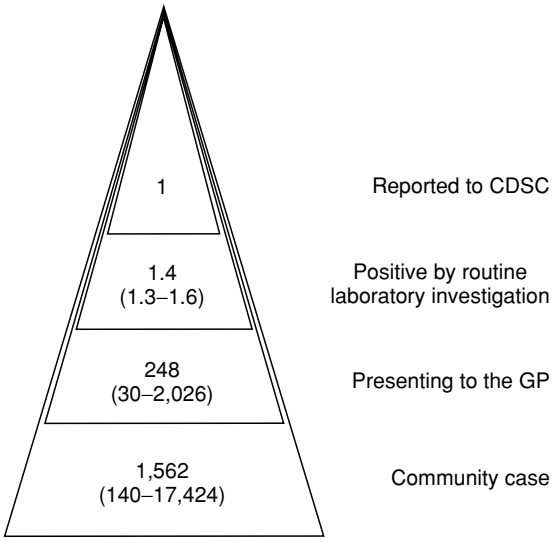

Figure 8 Reporting pyramid for SRSV.

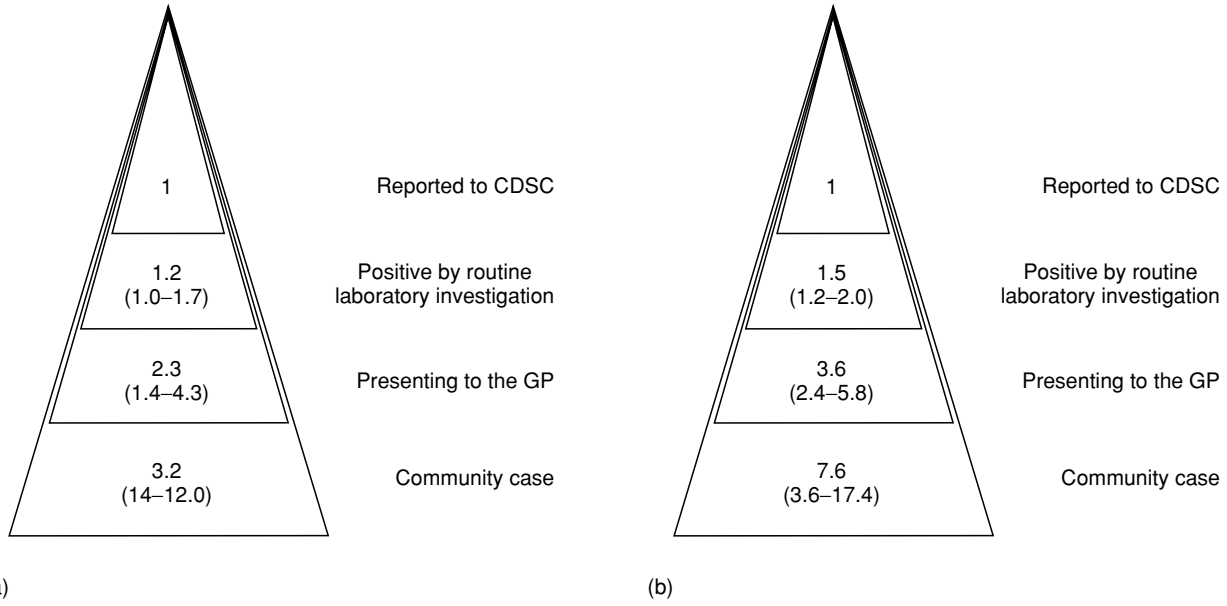

Figure 7 Reporting pyramids for (a) *Salmonella* and (b) *Campylobacter*.

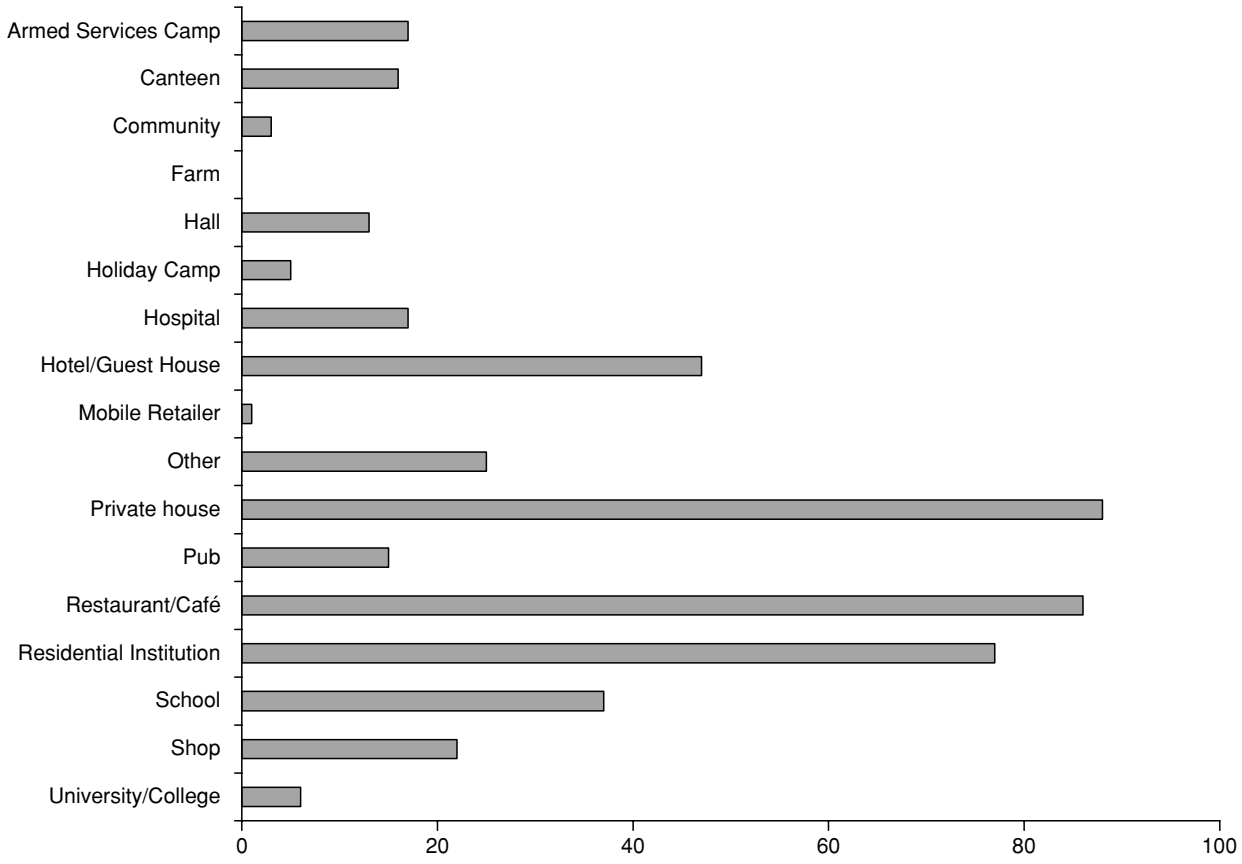

Figure 9 *Salmonella enteritidis* PT4 outbreaks in England and Wales, by place of outbreak, from 1992 to 1997.

defined as 'affecting members of more than one private residence or residence of an institution.' General outbreaks are distinguished from family outbreaks affecting members of the same private residence only. In 1992, an improved systematic surveillance system for outbreaks was developed by the PHLS in England and Wales in response to the Richmond's Committee recommendations. The CDSC is made aware of outbreaks from laboratory reports, consultants in communicable disease control, environmental health officers, and others. Outbreaks are then followed up using a standard request form that documents basic details. Participation in the surveillance scheme is entirely voluntary. Figure 9 shows the number of outbreaks of *Salmonella enteritidis* PT4 in England and Wales for 1992–97.

Enter-net

An international surveillance network (Enter-net) for human gastrointestinal infections involving 15 countries of the European Union and the USA has been set up (see Figure 10). The aim of the network is to conduct international surveillance of human salmonellosis and verocytotoxin producing *E. coli* 0157. This project began first of all with *Salmonella* cases in 1994 and was funded by the EU. The network consisted of the microbiologists in charge of the national reference laboratories in participating countries and epidemiologists responsible for national surveillance. The minimum dataset is collected for each isolate and reported to a central unit.

A dynamic database has been created, which shows that trends in salmonellosis declined throughout Western Europe from 1996 (see Figure 11). As a result of this network, international outbreaks were recognized. Individual countries identifying apparently isolated incidents were linked together to identify wider outbreaks. One of these outbreaks involved 12 cases of *E. coli* 0157 in Denmark, England, Wales, Finland, and Sweden in visitors to Fuerteventura in the Canary Islands.

Other Sources of Statistics

Registrations of Deaths

Most countries have statutory deaths registrations systems that seek to record the causes of death. These are of limited use in foodborne diseases, since

Figure 10 Enter-net participating countries.

such diseases rarely are primary causes of death, and even when they are contributory causes, the registration data may not systematically record this.

Hospital Admissions Data

Many countries have statistics on the numbers of patients admitted or discharged from hospital by disease. Once again, these data are of very limited use in developed countries, since foodborne disease rarely causes admission to hospital. Even when it does, the microbiological causes may not be recorded.

Clinical Reporting Networks

When a new foodborne disease emerges, special reporting systems may be set up, such as the UK collaborative study of childhood hemolytic uremic syndrome (HUS), which began at the end of February 1997 and is conducted by the British Pediatric Surveillance Unit (BPJU), the PHLS, Scottish Centre for Infection and Environmental Health (SCIEH) and the Scottish 0157 Reference Laboratory. Pediatricians report cases of clinically diagnosed HUS directly to the PHLS, CDSC, or SCIEH by telephone at the time of the diagnosis and also follow this up with a standard report card to the BPSU. Clinical and epidemiological data are collected directly from pediatricians. Routine feces and serum samples are sent for standard examination and isolation and subtyping of VTEC. During the first two years of the study, data

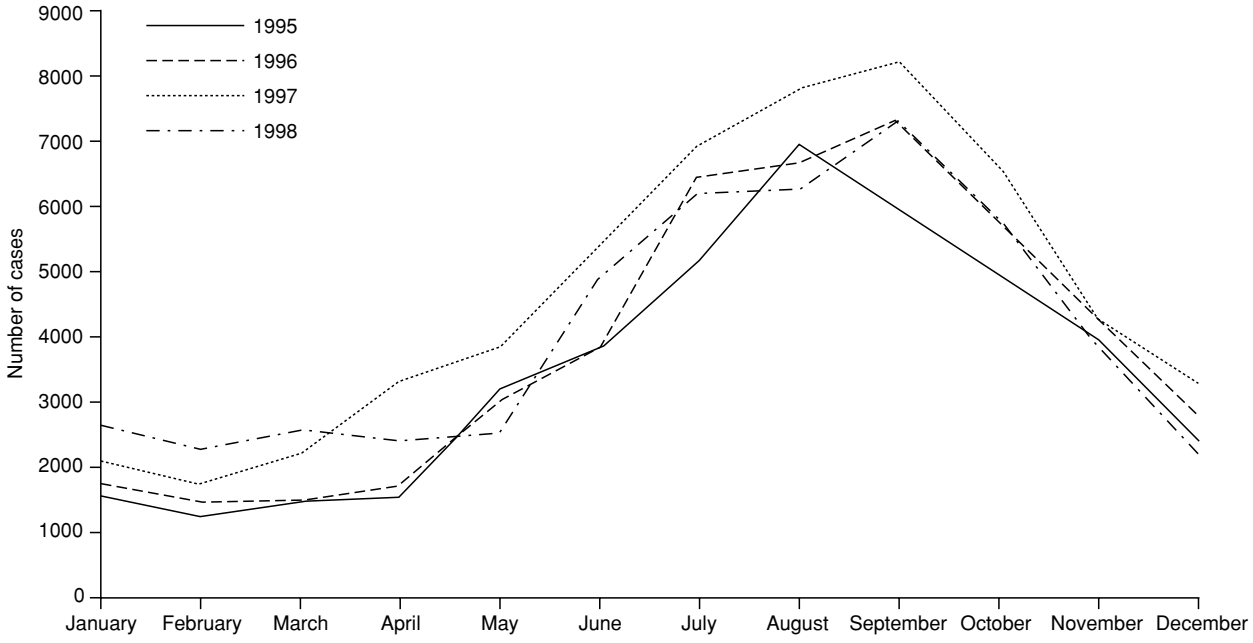

Figure 11 *Salmonella enteritidis* cases in Western Europe from 1995 to 1998.

were collected from 193 clinically confirmed cases, and five children were reported to have died. One hundred and eighty-six cases presented with a diarrheal illness, and 156 of these suffered bloody diarrhea. Stool and/or serum specimens were obtained from 185 cases, and of these, 165 were shown to have suffered VTEC infection.

See also: ***Campylobacter***: Properties and Occurrence; ***Escherichia coli***: Food Poisoning; **Food Poisoning**: Tracing Origins and Testing; ***Listeria***: Properties and Occurrence; ***Salmonella***: Properties and Occurrence

Further Reading

Adak GK, Long SM and O'Brien SJ (2002) Trends in indigenous foodborne disease and deaths, England and Wales: 1992 to 2000. *Gut* 51(6): 832–841.

Advisory Committee on the Microbiological Safety of Food (2001) *Second Report on Salmonella in Eggs*. London: The Stationery Office.

Berkelman RL, Stroup DF and Buehler JW (2002) Public Health Surveillance. In: Detels J, McEwen J, Beaglehole R and Tanaka H (eds) *Oxford Textbook of Public Health. The Scope of Public Health*, Vol. 1, 4th edn., pp. 759–781. Oxford University Press.

Cowden JM (2002) Foodborne infectious risks: do we need a wide system of data collection and survey? The lessons learned from the study of infectious intestinal disease in England. *Rev Epidemiol Sante Publique* 50(1): 89–92.

Fisher IS and Gill ON (2001) International surveillance networks and principles of collaboration. *European Surveillance* 6(2): 17–21.

Food Standards Agency (2000) *A Report of the Study of Infectious Intestinal Disease in England*. London: The Stationery Office.

Frost JA, Gillespie IA and O'Brien SJ (2002) Public health implications of campylobacter outbreaks in England and Wales, 1995–1999: epidemiological and microbiological investigations. *Epidemiology and Infection* 128(2): 111–118.

Herikstad H, Yang S, Van Gilder TJ *et al*. (2002) A population based estimate of the burden of diarrhoeal illness in the United States: FoodNet, 1996–97. *Epidemiology and Infection* 129(1): 9–17.

Palmer SR, Parry S, Perry D *et al*. (2000) The role of outbreaks in developing food safety policy: population based surveillance of salmonella outbreaks in Wales 1986–1998. *Epidemiology and Infection* 125(3): 467–472.

Parry SM and Palmer SR (2002) *E. coli. Environmental Health Issues of VTEC 0157*. Clay's Library of Health and the Environment. Spon Press.

Economic Implications

R B Raybourne and K M Williams, Center for Food Safety and Applied Nutrition, Laurel, MD, USA
T Roberts, Economic Research Service, Washington, DC, USA
Arthritis Working Group[1]

Background

Foodborne diseases are caused by ingesting bacteria, fungi, parasites, viruses or their toxic metabolites through contaminated food or water or through person-to-person contact. Each year, microbial pathogens cause an estimated 76 million cases of foodborne illness, including 5200 deaths in the USA, according to the study by Mead *et al*. at the Centers for Disease Control and Prevention (CDC). The Economic Research Service (ERS) in the United States Department of Agriculture and the Center for Food Safety and Applied Nutrition (CFSAN) in the Food and Drug Administration have estimated the annual human illness costs for a number of causes of foodborne illness. The ERS estimates that the total cost of five major bacterial pathogens is $6.9 billion annually. CFSAN estimates a cost of $27.9 billion for nine major causes of foodborne illness. In their estimates for human foodborne illness costs, both ERS and CFSAN include medical costs, productivity losses from missed work, and an estimate of the value of premature deaths. CFSAN also includes an estimate of the cost of pain and suffering due to illness.

The vast majority of foodborne illnesses are classified as 'acute.' These are usually self-limiting and of short duration, although the cases can range from mild to severe. Gastrointestinal problems and vomiting are common acute symptoms of many foodborne illnesses. Deaths from acute foodborne illness, while rare, are more likely to occur in the very young (including the fetus), the elderly, or patients with compromised immune systems (such as those suffering from AIDS and cancer).

[1]The other participants in the Arthritis Working Group of the Risk Assessment Consortium (RAC) are: ERS – P Frenzen; Centers for Disease Control and Prevention – J Sobel and S O' Connor; and Food and Drug Administration – A Jessup, C McGlaughlin, C Nardinelli, and R Scharff. The RAC's members are United States government agencies with food safety research and management responsibilities. The RAC enhances communication and coordination among the member agencies and promotes the conduct of scientific research that facilitates risk assessments. The RAC's web page is: http://www.foodriskclearing house.umd.edu/Risk_Assessment_Consortium.htm.

Complications of foodborne illness, called chronic sequelae, can occur in any part of the body including the joints, nervous system, kidneys, or heart. These chronic illnesses may afflict the patients for the remainder of their lives or result in premature death. For example, *Campylobacter* infections are estimated to be responsible for 20–40% of Guillain–Barré syndrome (GBS) cases (a major cause of paralysis unrelated to trauma) in the USA. About 1.5% of *E. coli* O157:H7 disease patients develop hemolytic uremic syndrome (HUS), which usually involves red blood cell destruction, kidney failure, and neurological complications, such as seizures and strokes.

We will focus on a group of chronic sequelae of foodborne infections composed of reactive arthritis, Reiter's syndrome, and ankylosing spondylitis, known collectively as the spondyloathropathies. We will use these chronic sequelae as an in depth example of how cost estimates are developed. These arthritis-like conditions occur in the aftermath of acute infection with several genera of Gram-negative Enterobacteriaceae including *Salmonella* spp., and all involve at a minimum some form of joint inflammation. The reported incidence of postacute joint inflammation can vary widely. The degree of severity and duration of arthritis is also highly variable. However, in the most severe or chronic cases, the consequences may be life-long, and there is potential to significantly increase the economic impact beyond the initial cost of acute foodborne illness.

Exposure to Pathogens and Associated Costs

Actions by the food industry, consumers, and the public health sector all influence how food is produced, marketed, prepared, and consumed. These actions influence the probability that a food item contains pathogens. People who consume contaminated food have some probability of becoming ill. Foodborne illness generates costs that are borne by the food industry, households whose members become ill, and/or the public health sector (Figure 1). A full accounting of the costs of all sectors of the economy would include estimates for all the costs listed in Table 1. Depending on the purpose of the study, different cost categories can be selected, though medical costs are those most commonly estimated.

Cost Estimates for Foodborne Disease

The Centers for Disease Control and Prevention estimates that out of a total of 76 million cases of foodborne illness each year in the USA, as many as 62 million cases are of unknown origin. Thus, in 82% of cases of foodborne illness, the disease or agent has not been identified, and costs have not been estimated. ERS and CFSAN have estimated costs for the most common identifiable causes of foodborne illness. As the pioneer of cost estimates for US foodborne illness, ERS has estimated costs for five common foodborne pathogens, likely to be associated with meat and poultry. As the agency with the responsibility for regulating the safety of the most food groups, CFSAN has estimated the cost of almost 20 causes of foodborne illness. Table 2 illustrates the estimated costs of the most common foodborne illnesses that have been studied by the two agencies.

For each agent or disease that causes foodborne illness, ERS and CFSAN estimate an annual cost of illness based on the estimated number of cases, the expected severity of illness, and whether we expect an acute illness to result in chronic sequelae. Estimated costs include medical costs, lost productivity costs, other illness-specific costs (such as special education and residential-care costs), and an estimate of the value of premature deaths. CFSAN also measures quality-adjusted life year (QALY) losses to value the pain and suffering associated with foodborne illness. Costs have not been estimated for the vast majority of complications associated with foodborne illnesses. The cost estimates do, however, cover the following chronic complications in our cost estimates: paralysis following *Campylobacter* spp. infections (called the Guillian-Barré Syndrome), kidney failure following *E. coli* O157:H7 infections (HUS), and chronic disability or impairment following congenital and newborn infections from *Listeria monocytogenes*. CFSAN estimates also include costs for arthritic complications that result from illnesses due to *Salmonella* spp. and *Shigella* spp.

Medical costs ERS and CFSAN use similar methods to estimate medical costs. Disease-outcome trees are developed for each illness to put the medical data in perspective. The total number of cases are divided into different levels of severity of disease:

- cases who recover without seeking medical care,
- cases who visit a physician and recover,
- cases who are hospitalized and recover,
- cases who develop chronic complications (can overlap with an acute survivor category), and
- those who die prematurely because of their illness.

For the different severity groups, the percentage recovering fully, partially recovering, or dying prematurely is estimated and entered into the disease outcome tree. (The disease outcome tree approach will be illustrated later when discussing the medical evidence connecting foodborne illness to arthritis.) Medical costs are estimated for physician and

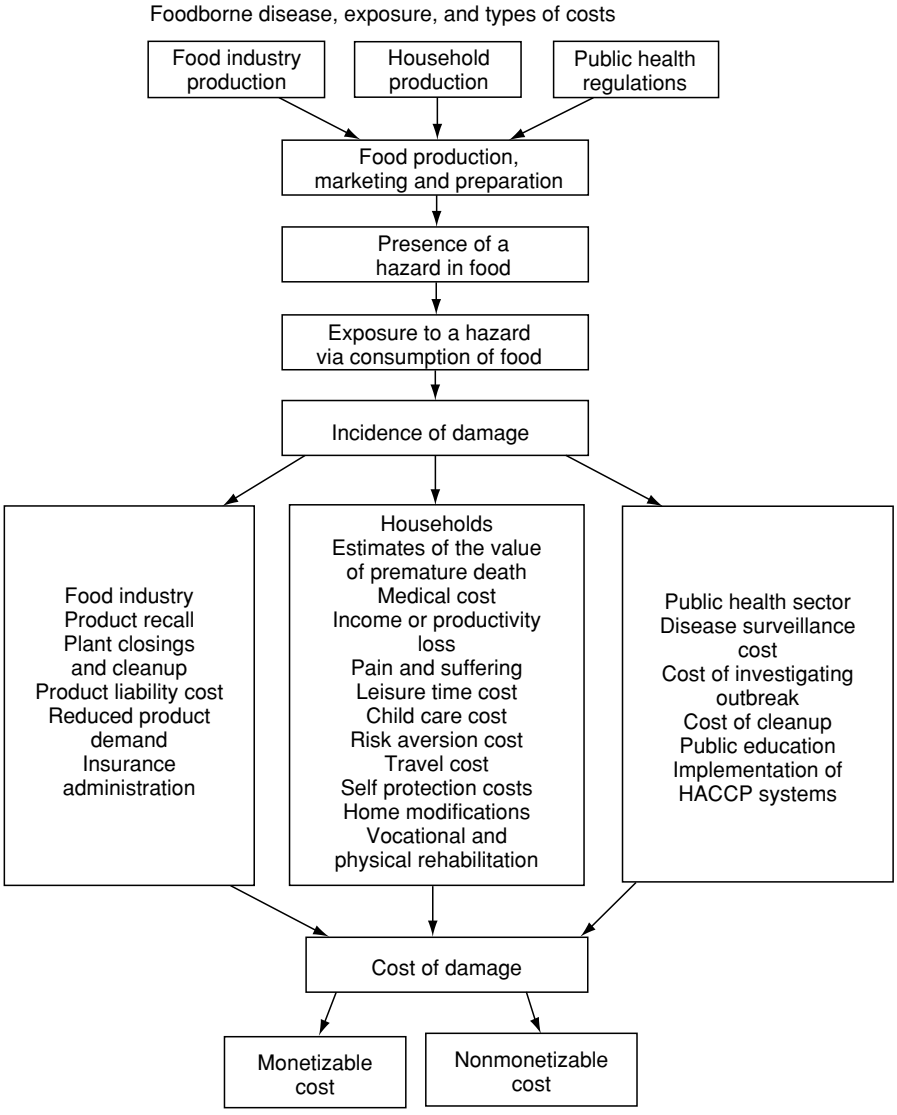

Figure 1 Foodborne disease, exposure, and types of costs.

hospital services, supplies, medications, and special procedures unique to treating the particular foodborne illnesses. Such costs reflect the average number of days or treatments of a medical service, the average cost per service or treatment in the USA, and the number of patients receiving such service or treatment.

ERS estimates of nonmedical costs For the 2000 cost estimates, ERS revised the methodology for valuing premature deaths due to foodborne illness. ERS previously assigned all deaths the same value regardless of age at time of death, based on information about the wage premiums for high-risk occupations (the 'labor market approach'). Now, the updated cost estimates use information about the age distribution of deaths to adjust this value to account for age at death.

In essence, the labor market approach values the economic cost of premature deaths based on the risk premium revealed by the higher wages paid for dangerous jobs. Under this approach, the value of a statistical life equals $6.5 million in August 2000 dollars after updating Viscusi's original 1990 estimate of $5.0 million to account for inflation. The labor market approach assumes that risk preferences observed in job choices are indicative of risk preferences for food safety.

ERS modified the labor market approach by taking the age distribution of deaths from each pathogen

Table 1 Social costs of foodborne illness

Costs to individuals/households
- Human illness costs:
  - Medical costs:
    - Physician visits
    - Laboratory costs
    - Hospitalization or nursing home
    - Drugs and other medications
    - Ambulance or other travel costs
  - Income or productivity loss for:
    - Ill person or fatal case
    - Caregiver for ill person
  - Other illness costs:
    - Travel costs to visit ill person
    - Home modifications
    - Vocational/physical rehabilitation
    - Child care costs
    - Special educational programs
    - Institutional care
    - Lost leisure time
  - Psychological costs:
    - Pain and other psychological costs
    - Risk aversion
  - Averting behavior costs:
    - Extra cleaning/cooking time costs
    - Extra cost of refrigerator, freezer, etc.
      - Flavor changes from traditional recipes (especially meat, milk, egg dishes)
      - Increased food cost when more expensive but safer foods are purchased
  - Altruism (willingness to pay for others to avoid illness)

Industry costs
- Impact of pathogens on animal production costs:
  - Morbidity and mortality of animals on farms
  - Reduced growth rate/feed efficiency and increased time to market
  - Costs of disposal of contaminated animals on farm and at slaughterhouse
  - Increased trimming or reworking at slaughterhouse and processing plant
  - Lost productivity due to illness among workers caused by handling contaminated animals or products
  - Increased meat product spoilage due to pathogen contamination
- Control costs for pathogens at all links in the food chain[a]:
  - New farm practices (age segregated housing, sterilized feed, etc.)
  - Altered animal transport and marketing patterns (animal identification systems, feeding/watering)
  - New slaughterhouse procedures (hide wash, knife sterilization, carcass sterilizing)
  - New processing procedures (pathogen tests, contract purchasing requirements)
  - Altered product transport (increased use of time/temperature indicators)
  - New wholesale/retail practices (pathogen tests, employee training, procedures)
  - Risk assessment modelling by industry for all links in the food chain
  - Price incentives for pathogen-reduced product at each link in the food chain
- Outbreak costs:
  - Herd slaughter/product recall
  - Plant closings and cleanup
  - Regulatory fines
  - Product liability suits from consumers and other firms[b]
  - Reduced product demand because of outbreak:
    - Generic animal product – all firms affected
    - Reduction for specific firm at wholesale or retail level
  - Increased advertising or consumer assurances following outbreak
  - Impact of outbreaks on tourism industry

Regulatory and public health sector costs for foodborne pathogens
- Disease surveillance costs to:
  - Monitor incidence/severity of human disease by foodborne pathogens
  - Monitor pathogen incidence in the food chain
  - Develop integrated database from farm to table for foodborne pathogens
- Research to:
  - Identify new foodborne pathogens for acute and chronic human illnesses
  - Establish high-risk products and production and consumption practices
  - Identify which consumers are at high-risk for which pathogens
  - Develop cheaper and faster pathogen tests
  - Risk assessment modelling for all links in the food chain
- Outbreak costs:
  - Costs of investigating outbreak
  - Testing to contain an outbreak (for example, serum testing and administration of IG in persons exposed to Hepatitis A)
  - Costs of cleanup
  - Legal suits to enforce regulations that may have been violated[b]

Other considerations
- Distributional effects in different regions, industries, etc.
- Equity considerations, such as special concern for children

[a]Some industry costs may fall with better pathogen control, such as reduced product spoilage, possible increases in product shelf-life, and extended shelf-life permitting shipment to more distant markets or lowering shipment costs to nearby markets.
[b]In adding up costs, care must be taken to assure that product liability costs to firms are not already counted in the estimated pain and suffering cost to individuals. However, the legal and court expenses incurred by all parties are social costs.

into account, in effect treating the value of life as an annuity paid over the average US life span at an interest rate of 3.0%. Following age adjustment, the assumed cost of each death ranges from $8.9 million for individuals who died before their first birthday to $1.7 million for individuals who died at age 85 or older.

Since the five microbial pathogens have different health outcomes for different age groups, adjusting for the age of death raises the cost of some foodborne illnesses and lowers the cost of others. For example, the annual cost of foodborne illness caused by *Salmonella* spp. decreases to $2.2 billion, because over two-thirds of the deaths from salmonellosis occur among people over 65 years of age. Adjusting foodborne illness costs for *E. coli* O157:H7 by age at time of death increases the estimates by $200 million, because most deaths are children under the age of five.

Table 2 Economic costs of foodborne illness in the USA: estimated annual costs due to selected US foodborne pathogens, 2000[a]

Pathogen	*Estimated annual foodborne illnesses*[b]			*ERS costs*[c]	*CFSAN costs*[d]
	Cases	*Hospitalizations*	*Deaths*	*Billion 2000 dollars*	
Bacterial infections					
Campylobacter spp.	1 963 141	10 539	99	1.2	–
Salmonella	1 341 873	15 608	553	2.4	17.2
Clostridium perfringens	248 520	41	7	–	0.2
Shigella	89 648	1 246	14	–	0.3
E. coli O157:H7	62 458	1 843	52	0.7	2.2
E. coli, non-O157 STEC[e]	31 229	921	26	0.3	–
Listeria monocytogenes	2 493	2 298	499	2.3	2.5
Parasitic infections					
Giardia lamblia	200 000	500	1	–	0.2
Cryptosporidium parvum	30 000	199	7	–	0.1
Viral infections					
Norwalk-like viruses	9 200 000	20 000	124	–	5.3
Hepatitis A	4 170	90	4	–	0.1
Totals					
ERS total	3 401 194	31 209	1 229	6.9	–
CFSAN total	11 179 162	41 825	1 261	–	28.1

[a]As these new estimates of foodborne illness costs are based on new data and improved methodologies for valuing these costs, the estimates presented here are not directly comparable with earlier ERS and CFSAN estimates of the costs of foodborne disease.
[b]Data from the Centers for Disease Control and Prevention published in Mead PS, Slutsker L, Dietz V *et al*. (1999) Food-related illness and death in the United States. *Emerging Infectious Diseases* 5: 607–625.
[c]The ERS total estimated costs include specific chronic complications in the case of *Campylobacter* (Guillain–Barré syndrome), *E. coli* O157:H7 (hemolytic uremic syndrome), and *Listeria monocytogenes* (congenital and newborn infections resulting in chronic disability or impairment). Estimated costs for *Listeria monocytogenes* exclude less serious cases that do not require hospitalization.
[d]The CFSAN total estimated costs include specific chronic complications in the case of *Salmonella* (arthritic conditions), *Shigella* (arthritic conditions and hemolytic uremic syndrome), and *E. coli* O157:H7 (hemolytic uremic syndrome). CFSAN uses QALYs to estimate the value of pain and suffereing.
[e]STEC stands for Shiga toxin-producing *E. coli*.

One difficult issue is assigning a value to the productivity losses associated with individuals who become ill and are unable to return to work, or for those illnesses that result in a lifetime of disability (such as prenatal exposure to *Listeria monocytogenes*). ERS currently measures the productivity losses due to nonfatal foodborne illness by the value of forgone or lost wages, regardless of whether the lost wages involved a few days missed from work or a permanent disability that prevented an individual from returning to work. Using the value of lost wages for cases resulting in disability probably understates an individual's willingness to pay to avoid disability because it does not account for the value placed on avoiding pain and suffering. The willingness to pay measure derived from labor market studies that ERS uses to value a premature death is not an appropriate measure of willingness to pay to avoid disability because it measures the higher wages paid to workers to accept a higher risk of premature death, not disability. Methods have been suggested to adjust willingness to pay to reduce the risk of premature death downward to estimate willingness to pay to avoid disability, such as the approach based on measuring QALYs. As yet, there is no consensus among economists about how to use these methods to value willingness to pay to avoid the disability, pain, and suffering associated with foodborne illnesses. ERS's conservative estimates of the annual costs due to foodborne illness (particularly the chronic conditions associated with *Campylobacter* spp.) would be substantially increased if willingness to pay to avoid disability, pain, and suffering were also taken into account.

CFSAN estimates of nonmedical costs Medical costs are a small fraction of the total social cost of foodborne illness. In addition to the cost of medical care, economists are concerned with the loss of productivity and the loss of well-being (through pain and suffering) that accompany foodborne illness. CFSAN routinely estimates these nonmedical costs of foodborne illness.

CFSAN was one of the first government agencies to use the labor market approach to estimating the economic cost of premature death. Economists have long favored this method, because it is based on observed tradeoffs that individuals make between dollars and risk. An unadjusted figure of $5.0 million per premature death is generally used in CFSAN analyses. As noted above, this value does not take into account economic changes that have occurred since 1990.

Economists at CFSAN are currently working on an adjustment factor based on both nominal and real changes in society's valuation of health. As a result, the figures that are reported in Table 2 are likely to be an underestimate of the true value of foodborne illness.

The value of productivity losses and pain and suffering are measured by placing values on quality adjusted life years (QALYs). This method uses a two-step procedure for valuing health losses. In the first step, the effect of a condition on health is estimated to be between zero (well-being in the full health state) and one (death). For example, a QALY loss of 0.14 for arthritis means that for every day of suffering with arthritis, the affected individual has a level of well-being 14% lower than they would have had in the absence of arthritis. In the next step, the value of a QALY is estimated and multiplied by expected QALY losses to calculate the value of health losses. QALYs are designed to measure the loss of well-being both from symptoms and from activity limitation. Consequently, both pain and suffering and productivity values are captured by this measure.

ERS and CFSAN use different methods of measuring the nonmedical costs of illness. The net effect of the differing approaches is that CFSAN estimates for the value of foodborne illness include a value for pain and suffering, whereas ERS estimates do not. The inclusion of a value for pain and suffering is demonstrated in Table 2. CFSAN estimates of the cost of foodborne illness are uniformly larger than estimates made by ERS. In the case of *E. coli* O157, CFSAN based its medical cost estimates on estimates calculated by ERS. Nonetheless, the inclusion of QALY losses boosts CFSAN estimates of economic loss due to *E. coli* O157 to $2.2 billion, as opposed to $0.7 billion for ERS.

Chronic Sequelae of Foodborne Illness

A listing of the association of foodborne pathogens with chronic sequelae (Table 3) illustrates that many or most of these agents are associated with some type of complication. Of particular interest is the fact that several bacterial pathogens are associated with arthritic sequelae. In addition to identifying the various disease conditions, determining the economic impact of chronic sequelae associated with foodborne illness requires not only a systematic method of estimating costs but also some understanding of the clinical aspects of these entities. This is particularly true with respect to the degree of variation that may

Table 3 Chronic complications associated with foodborne pathogens

Bacterial and parasitic infection transmitted by foods	*Complications/sequelae*
Bacterial infections	
Aeromonas hydrophila enteritis	Bronchopneumonia, cholecystitis
Brucellosis	Aortitis, epididymo-orchitis, meningitis, pericarditis, spondylitis
Campylobacteriosis	Arthritis, carditis, cholecystitis, colitis, endocarditis, erythema nodosum, Guillain–Barré syndrome, hemolytic uremic syndrome, meningitis, pancreatitis, septicemia
Escherichia coli (EHEC-types) enteritis	Erythema nodosum, hemolytic uremic syndrome, seronegative arthropathy, thrombotic thrombocytopenic purpura
Q-fever	Endocarditis, granulomatous hepatitis
Salmonellosis	Aortitis, cholecystitis, colitis, endocarditis, epididymo-orchitis, meningitis, myocarditis, osteomyelitis, pancreatitis, Reiter's disease, rheumatoid syndromes, septicemia, splenic abscesses, thyroiditis, septic arthritis (sickle-cell anemic persons)
Shigellosis	Erythema nodosum, hemolytic-uremic syndrome, peripheral neuropathy, pneumonia, Reiter's disease, septicemia, splenic abscesses, synovitis
Vibrio parahaemolyticus enteritis	Septicemia
Yersiniosis	Arthritis, cholangitis, erythema nodosum, liver and splenic abscesses, lymphadenitis, pneumonia, pyomyositis, Reiter's disease, septicemia, spondylitis, Still's disease
Parasitic infections	
Cryptosporidiosis[b]	Severe diarrhea, prolonged and sometimes fatal
Giardiasis[b]	Cholangitis, dystrophy, joint symptoms, lymphoidal hyperplasia
Taeniasis	Arthritis, cysticercosis (*T. solium*)
Toxoplasmosis	Encephalitis and other central nervous system diseases, pancarditis, polymyositis
Trichinosis	Cardiac dysfunction, neurologic sequelae

[a]Suspected of being foodborne or waterborne.
[b]Waterborne.
Source: Council for Agricultural Science and Technology (CAST) (1994) *Foodborne Pathogens: Risks and Consequences*. Ames, IA: CAST.

be encountered in severity and duration. An additional factor is whether there are certain populations that are more susceptible to the development of a particular chronic condition. These factors apply in varying degrees to all of the sequelae listed in Table 3, but many of the complications either occur very rarely, or are associated with pathogens that are themselves rare. Arthritis can be a relatively frequent complication of foodborne illness and is associated with some of the more common bacterial pathogens in food. In addition, arthritic sequelae are one of the more complex examples in terms of severity, duration, and susceptible populations. For these reasons, arthritic sequelae will serve as a paradigm for examining various economic and clinical considerations involved in developing cost estimates.

Arthritis Connection to Foodborne Illness

Arthritic sequelae associated with foodborne illness are organized into three entities: Reactive Arthritis (ReA), Reiter's syndrome (RS) and Ankylosing Spondylitis (AS) (Table 4). The incidence of these three conditions is increased in individuals that possess the human Class I histocompatability antigen HLA-B27. This association was one of the first documented examples of genetic predisposition to disease. The interest focused on this connection, the mechanisms involved, and the diseases themselves have resulted in numerous research and review articles. The genetic connection and a commonality of symptoms (joint inflammation) have led experts to group the three conditions together as 'HLA-B27-related spondyloarthropathies.'

- ReA is characterized by the inflammation of one or more joints and generally occurs 1–3 weeks following acute foodborne illness. Viable organisms of the causative pathogen cannot be isolated from the inflamed joints, making it a sterile arthritis, but there are reports that bacterial antigens or DNA may be present in inflamed joints.
- RS is a complicated form of ReA that includes joint involvement as well as at least one extra-articular feature such as inflammation of the urethra (primarily in males) and/or uveitis (eye inflammation). In some cases, no preceding infection can be identified. Those cases that have an identifiable preceding infection are sometimes called Reactive RS. Infections leading to RS can be derived from food (e.g., *Salmonella* and *Yersina*) or from sexually transmitted organisms such as *Chlamydia* spp. Because RS associated with foodborne illness is similar to ReA with respect to articular features and may only have one extra-articular feature, the term ReA will be used to capture both entities.
- AS is the most chronic and debilitating of the spondyloarthropathies, characterized by progressive degeneration of the lumbar spine. AS can take many years to develop, and in most cases, it is difficult to make a specific association with a preceding foodborne infection. The relationship of AS to foodborne illness derives from its strong link to HLA-B27, its inclusion in the spondyloarthropathies, and studies of foodborne disease outbreaks in which patients have been followed for several (many) years. These studies have provided evidence that ReA or RS can progress to AS in some individuals and support the theory that the sponsyloarthropathies are a spectrum of related diseases. Clinically, this relationship has been useful in the early diagnosis of AS, when doing a long-term follow-up of foodborne outbreaks. Progression to diagnosed AS is not the only potential chronic outcome of ReA, which can sometimes lead to recurrent or chronic arthritis of joints other than the spine. The frequency with which these various scenarios occur has a great bearing on the potential economic consequences of foodborne illness.

Before questions surrounding the development of long-term arthritic sequelae can be addressed, the frequency of primary joint inflammation (ReA) after foodborne illness must be estimated. An assessment of the circumstances surrounding the initial episode of illness is also relevant to development of ReA and long-term sequelae. Most of the available information in this area comes from outbreaks of foodborne illness. However, most outbreak investigations focus on the acute, gastrointestinal phase of infection, whereas follow-up surveys to detect arthritic and other sequelae are rare. There is even less information

Table 4 HLA-B27-associated spondyloarthropathies

Condition	*Temporal relationship to infection*	*HLA-B27 association*	*Prognosis*
Reactive arthritis	1–3 weeks after enteric illness	30–80%	Usually self-limiting
Reiter's syndrome	Follows enteric or sexually transmitted infection (1–3 weeks)	60–80%	May be self-limiting non-articular inflammation
Ankylosing spondylitis	Difficult to link	90+%	Progressive

on the occurrence of ReA following sporadic foodborne illness. (Sporadic cases are foodborne illnesses that have not been identified as part of an outbreak.) Clinical information gathered from outbreaks has suggested that several factors influence the incidence and severity of ReA. These include the duration or severity of diarrhea, which is influenced by the virulence of the pathogen and the number of pathogens ingested. The primary host factors appear to be the presence of HLA-B27 or other antigenically related antigens from the HLA-B cross-reactive group of antigens. With respect to factors that influence progression from ReA to chronic joint problems, including AS, severity of initial joint symptoms, and presence of HLA-B27 are the most prominent. Recent studies on AS have also suggested that other host genetic components related to the production of inflammatory mediators such as interferon and tumor necrosis factor play a role as well. With respect to the frequency with which primary and chronic joint problems occur, there is considerable variation in published reports.

Disease-outcome Tree – Node 1

Our goal was to develop a flow chart or disease-outcome tree for arthritic sequelae that expressed the probabilities of various short term and chronic outcomes based on available clinical literature. This tree was developed to distribute the cases into different categories that are likely to have different average costs. Several common variables from published reports of both short- and long-term sequelae were used to develop the tree. These included the duration of patient follow-up, the size of the initial outbreak, response rate to follow-up surveys, the pathogen involved, and bias within the affected population. With regard to potential bias in the populations under study, the incidence of HLA-B27 varies among national and ethnic groups and tends to be higher among Northern European groups where many outbreaks have been studied for arthritic sequelae. Studies from these countries were included, because to exclude them would have omitted several valuable reports. When the data were compiled, there was no obviously higher incidence of *Salmonella*-associated ReA among the European versus North American outbreaks (Table 5).

Numerous studies or case reports involving small numbers of patients were not included, because these studies provided no basis for determining an initial incidence of ReA. Most of the information on the initial incidence of arthritis following infection after *Yersinia enterocolitica, Shigella flexneri*, and *Campylobacter jejuni* was derived from review articles. In general, original reports involving the incidence of ReA following outbreaks of salmonellosis with various serovars were more frequent and were also included. These data were incorporated into the first node of the disease-outcome tree (Figure 2). Only the four pathogens above were considered, because they occur with sufficient frequency in the USA to be captured by FoodNet active surveillance. In Europe,

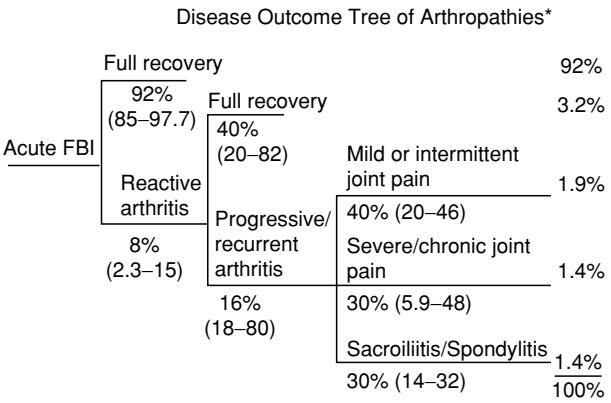

Figure 2 Arthritis–foodborne illness disease outcome tree.

Table 5 Initial incidence of reactive arthritis following foodborne illness

Source	*Type (cases)*	*Location*	*Incidence of ReA (%)*
Keat (1983)	Review	Multiple outbreaks	2–3 (2.5)
Maki-Ikola and Granfors (1992)	Review	Multiple outbreaks	1.2–7.3 (3.5)
Yu and Thomson (1994)	Review Outbreaks	Multiple	6–10 (8.0)
Simonet (1999)	Review	Multiple outbreaks	0.5–15 (4.3)
Ike *et al*. (1986)	Outbreak (565) *Salmonella* – milk	US survey 30% response rate	2.3
Inman *et al*. (1988)	Outbreak (260) *Salmonella*	Canada survey	7.3
Thomson *et al*. (1992)	Outbreak (83) *Salmonella* – luncheon	Canada survey 92% response rate	8.2
Locht *et al*. (1993)	Outbreak (126) *Salmonella* – dinner	Sweden survey 90% response rate	15
Samuel *et al*. (1995)	Outbreak (919) *Salmonella* – restaurant	US survey 33% response rate	14
Thomson *et al*. (1995)	Outbreak (423) *Salmonella*	Canada survey 97% response rate	6.4
Mattila *et al*. (1998)	Outbreak (210) *Salmonella*	Finland survey 91% response rate	11.5

Yersinia pseudotuberculosis has been associated with outbreaks of foodborne illness featuring a high incidence of ReA. This organism was not included, because cases of infection with this pathogen are rare in the USA, with no confirmed cases from 1996–1999, based on FoodNet surveillance. The clinical picture of ReA with all of the four included pathogens appears to be similar. In looking at a number of outbreaks, the overall incidence of initial arthritis also appears to be similar among the above four organisms, but the incidence of arthritis may vary widely between different outbreaks involving the same pathogenic species. This is probably due to variation in pathogen virulence, dose, host population, or even the effectiveness of tracking arthritis. For this reason, in developing the disease-outcome tree, each report or review was treated as a distinct entity and given equal weight, rather than pooling cases from different outbreaks. The reports contributing to the first node of the tree give a range of 2.3–15% for the initial incidence of arthritis following foodborne illness. This results in an average of approximately 8% of foodborne illnesses that result in ReA.

Disease-outcome Tree – Node 2

The second node of the disease-outcome tree distinguishes between ReA that resolves completely and does not recur after the initial episode, and ReA in which the joint inflammation shows any form of progression or periodic recurrence. This node is based on studies that report some progression, or lack of resolution of joint disease, but do not measure or define progression to a particular clinical state. This includes studies with relatively short follow-up times of less than 1 year. Compared with studies on the initial incidence of ReA, studies involving progression or recurrence were fewer and involved outbreaks of salmonellosis more often than other pathogens (Table 6). Therefore, most of the data used for the second node of the disease-outcome tree were also derived from studies involving *Salmonella*. However, our purpose was to quantify, within reasonable bounds, the possible sequelae of foodborne infection from self-limiting to chronic joint disease. To focus exclusively on *Salmonella* would have resulted in omission of two important long-term follow-up studies with *Shigella* and *Yersinia enterocolitica*.

It is conceivable that there is some difference in the methodology and criteria for detection of ReA between older studies and more current reports. For example, a large outbreak of *Shigella* in 1948 in Finland reported a very low 0.2% incidence of arthritis, yet, upon follow-up of 100 of these cases 20 years later, 80% had shown some progression with 32% showing signs of AS. The figure of 80% progression is quite high compared with other studies and may have resulted from initial recognition of arthritis in only the most severe cases. These cases would be the most likely to progress to chronic status. Other review articles reported the proportion of cases that progress to chronic status using a qualitative approach, based on the clinical experience of the authors. In considering the percentage of cases in which initial ReA does not completely resolve, or progresses, the second node of the tree gives a range of 18–80%. This results in an average of 60% of cases that do not evidence full recovery.

Disease-outcome Tree – Node 3

The third node of the tree measures the progression to chronic joint disease, based on the data from relevant studies, making a distinction between AS and other forms of chronic joint inflammation. Fewer data points were available for this node. In the various studies, there was considerable variation in the number of subjects, the duration of follow-up, and the clinical criteria and terminology for progression and for severity. Chronic symptoms were categorized in these reports on the basis of which joints were affected, radiological criteria, patient histories, or a combination of these criteria. If progression was severe enough to cause radiologically detectable

Table 6 Progression of reactive arthritis

Study	*Follow-up (cases)*[a]	*Any progression (%)*	*Severe/chronic (%)*
Sairanen *et al*. (1969)	20 years (100)	80	32
Leirisalo and Suoranta (1988)	10 years (85)	51.8	5.9
Maki-Ikola and Granfors (1992)	> 1 year (164)	78	18.3
Thomson *et al*. (1992)	18 months (6)	33	na
Locht *et al*. (1993)	6 months (17)	47	na
Yu and Thomson (1994)	na (review)	66	33
Thomson *et al*. (1995)	5 years (27)	66	29.6
Leirisalo *et al*. (1997)	5–22 years (50)	60	40
Mattila *et al*. (1998)	4 months (22)	18	na

[a]Original number of ReA cases.

spinal damage, these cases tended to be identified as sacroiliitis or spondylitis (AS). These cases were distinguished in a separate category because AS is readily diagnosed as a distinct entity in the HLA-B27-associated spondyloathropathies. Two additional categories were used to capture other types of chronic or recurrent symptoms. These included cases in which symptoms manifested as periodic flare-ups, and cases in which there was chronic joint pain not associated with the back. Approximately 40% of cases in this node involve minor or recurrent symptoms. Sixty per cent of the cases are more severe or chronic joint inflammation, divided approximately equally between those that show radiologic evidence of spine and lower back inflammation and those having inflammation of other joints.

Conclusions

The disease-outcome tree is a useful tool for economic evaluation that establishes estimates for the likelihood of various arthritic sequelae following an outbreak of foodborne bacterial illness. The pathways are based on reported patterns of clinical disease. The estimates are necessarily inexact, because of limitations of the available data. The estimates are based on incidence data that were all derived from outbreaks of foodborne illness. It is possible that the incidence of arthritis following sporadic cases might be different than the incidence associated with outbreaks. The use of outbreaks might be biased toward more virulent pathogen strains. At present, there is no information on the incidence of ReA following sporadic foodborne illness. An active surveillance system for foodborne illnesses such as FoodNet may prove to be an excellent mechanism from which to determine the incidence of ReA in sporadic illness and to conduct long-term follow-ups for progressive or chronic joint disease. As more definitive data become available, the disease-outcome tree can be updated and modified.

The next issue is the economic valuation method used. The Economic Research Service's annual foodborne illness cost estimates use a cost-of-illness approach for illnesses and a willingness-to-pay (WTP) approach for premature deaths. CFSAN has routinely used the value of QALYs lost in its regulatory impact analyses of food safety regulations. To date, no standardized method has been chosen by the economic profession. Ongoing research is developing WTP estimates for morbidity outcomes. In comparison with cost-of-illness approaches for chronic sequelae, such as arthritis, both the WTP and QALY estimates can substantially increase the economic valuation. The reason is the more comprehensive nature of these estimates that include an implicit valuation for pain and suffering, lost leisure time, and other implicit costs that are not included in the cost of illness estimates based on medical costs, productivity losses, and other out-of-pocket costs.

In the disease-outcome tree developed here, around 8% of the cases of salmonellosis can be expected to develop reactive arthritis. Forty per cent of the reactive athritis cases will fully recover. The remaining 60% can be expected to have recurring episodes of varying severity (40% mild/intermittent joint pain, 30% severe joint pain, 30% spondylitis). In summary, the disease-outcome tree encompasses all the possible lifetime consequences of those in the US population who have been exposed to foodborne pathogens. The tree predicts that on average, 92% of acute cases will have a full recovery; 3.2% will have full recovery with reactive arthritis symptoms for a few weeks; 1.9% will have lifelong mild, intermittent joint pain; 1.4% will have a lifelong progression to severe joint pain; and 1.4% of cases will have a lifelong progression to spondylitis. Valuation of avoiding just the pain and suffering associated with these arthritic outcomes can be expected to be significant. An early study by Thompson suggested that those afflicted would give up 22% of their household income to be rid of rheumatoid arthritis. The disease-outcome tree encompasses all the possible lifetime consequences of those in the U.S. population who have been exposed to foodborne pathogens. The life-long nature and severity of many chronic sequelae, such as arthritis, indicate that these will make substantial contributions to the estimates of the cost of foodborne illness.

See also: ***Campylobacter***: Properties and Occurrence; ***Escherichia coli***: Food Poisoning; Food Poisoning by Species other than *Escherichia coli*; **Food Poisoning**: Classification; Tracing Origins and Testing; Statistics; ***Listeria***: Listeriosis; ***Salmonella***: Properties and Occurrence; ***Shigella***; ***Yersinia enterocolitica***: Properties and Occurrence

Further Reading

Council for Agricultural Science and Technology (CAST) (1994) *Foodborne Pathogens: Risk and Consequences*, *Task Force Report 122* (cochairs, Foegeding PM and Roberts T). Ames, IA: CAST.

Economic Research Service (2001) United States Department of Agriculture http://www.ers.usda.gov/briefing/FoodborneDisease/.

Food and Drug Administration (2001) *Hazard Analysis and Critical Control Point (HAACP); Procedures for the Safe and Sanitary Processing and Importing of Juice; Final Rule, 66 Federal Register 6138–6202*, esp. pp. 6179–6194 (January 19).

Food and Drug Administration (1999) *Preliminary Regulatory Impact Analysis and Initial Regulatory Flexibility Analysis of the Proposed Rule to Require Refrigeration of Shell Eggs at Retail and Safe Handling Labels, 64 Federal Register*, pp. 36516–36539 (July 6).

Food and Drug Administration (1995) *Procedures for the Safe and Sanitary Processing and Importing of Fish and Fishery Products; Final Rule, 60 Federal Register*, pp. 65096–65202, esp. pp. 65179–65192 (December 18).

Locht H, Kihlstrom E and Lindstrom FD (1993) Reactive arthritis after *Salmonella* among medical doctors: Study of an outbreak. *Journal of Rheumatology* 20: 845–848.

Mattila L, Leirisalo-Repo M, Pelkonen P *et al*. (1998) Reactive arthritis following an outbreak of *Salmonella bovismorbificans* infection. *Journal of Infection* 36: 289–295.

Mead PS, Slutsker L, Dietz V *et al*. (1999) Food related illness and death in the United States. *Emerging Infectious Diseases* 5: 607–625.

Sairanen E, Paronen I and Mahonen H (1969) Reiter's syndrome: a follow-up study. *Acta Medica Scandinavica* 185: 57–63.

Samuel MP, Zwillich SH, Thomson TD *et al*. (1995) Fast food arthritis – a clinico-pathologic study of post-*Salmonella* reactive arthritis. *The Journal of Rheumatology* 22: 1947–1952.

Simonet ML (1999) Enterobacteria in reactive arthritis: *Yersinia*, *Shigella*, and *Salmonella*. *Revue du Rheumatisme* (English edition) 66(1, Supplement): 14S–19S.

Thomson GTD, Chiu B, Derubeis D, Falk J and Inman RD (1992) Immunoepidemiology of post-*Salmonella* reactive arthritis in a cohort of women. *Clinical Immunology and Immunopathology* 64: 227–232.

Thomson GTD, DeRubeis DA, Hodge MA, Rajanaygam C and Inman RD (1995) Post-*Salmonella* reactive arthritis: late clinical sequelae in a point source cohort. *American Journal of Medicine* 98: 13–21.

Viscusi WK (1993) The value of risks to life and health. *Journal of Economic Literature* 31: 1912–1946.

Yu DTY and Thomson GTD (1994) Clinical, epidemiological and pathogenic aspects of reactive arthritis. *Food Microbiology* 11: 97–108.

FOOD SAFETY

R Rooney, World Health Organization, Geneva, Switzerland
P G Wall, Food Safety Authority of Ireland, Dublin, Ireland

Background

Contaminated food is one of the most widespread public health problems of the contemporary world and causes considerable morbidity and mortality. Globally, millions of people are affected by foodborne disease. Food poisoning can be very serious in vulnerable groups such as the elderly, infants, young children, pregnant women, and immunocompromised individuals. The proportion of vulnerable groups such as the elderly and immunocompromised is increasing in industrialized countries.

A number of high-profile food-safety emergencies and scares, in recent years, have shaken consumer confidence in the production of food and have focused attention on the way food is produced, processed, and marketed. These incidents have included an epidemic of bovine spongiform encephalitis (BSE) in cattle, possible links between BSE and variant Creutzfeldt–Jakob disease (CJD) in humans, and the discovery of dioxin contaminated animal feed entering the food chain. New foodborne pathogens such as *Escherichia coli* O157:H7, *Cyclospora cayentanensis*, and *Listeria monocytogenes* have also emerged. Infection with *Escherichia coli* O157:H7 can be fatal in young children, the elderly, and the immunocompromised, and within the last 10 years, serious outbreaks have occurred in the USA, Scotland, and Japan. *Listeria monocytogenes* can result in meningoencephalitis and/or septicemia in newborns and adults and abortion in pregnant women. Parasites such as *Cyclospora cayentanensis* and *Cryptosporidium parvuum* have emerged as a serious health threat for immunocompromised individuals, particularly AIDS patients.

Chemicals in food have also caused concern. Veterinary medicines, growth promoters, pesticides, and other man-made chemicals are widely used in the food industry. Improper use of these substances in primary production can result in unsafe residues in food. Chemicals are also added to foods during processing in the form of food additives, flavorings, and colors. Many consumers are worried about the long-term impact of mixtures of chemicals and chronic as well as acute effects on vulnerable groups.

Food safety is a sensitive political and economic issue as well as an important public health issue. This

paper aims to highlight the factors that have contributed to food-safety problems over recent decades such as intensive farming, globalization of the food industry, and changes in consumer demands. The biological, chemical, and physical hazards in food and possible control and prevention strategies for the industry will be discussed.

Food Safety Hazards

The term contaminant covers harmful substances or microorganisms that are not intentionally added to food. Contaminants may enter the food accidentally during growth, cultivation, or preparation, accumulate in food during storage, form in the food through the interaction of chemical components, or may be concentrated from the natural components of the food. There are three main types of hazard: biological, chemical, and physical.

Biological Hazards

Biological agents of concern to public health include pathogenic strains of bacteria, viruses, helminthes, protozoa and algae, and certain toxic products they may produce. Foodborne infections are caused when microorganisms are ingested, and these can multiply in the human body. Intoxications result when microbial or naturally occurring toxins are consumed in contaminated foods. Microorganisms or toxins may be introduced directly from infected food animals or from workers, other foods, or the environment during the preparation or processing of food. Poisonous substances may also be produced by the growth of bacteria and molds in food. Table 1 lists pathogenic organisms of public health importance, which may be transmitted through contaminated food.

Chemical Hazards

Foodborne toxicants can be categorized according to their chemical nature. Some foodborne toxicants are inorganic, e.g., lead and arsenic, whereas others are organic, e.g., saxitoxin and polychlorinated biphenols (PCBs). The main categories of man-made chemicals that occur in foods are shown in Table 2. Chemical hazards may result from pollution arising from industrial and other human activities (e.g., lead, mercury, cadmium, and PCBs), from agricultural practice (e.g., pesticides, fertilizers, and residues resulting from improper use of veterinary drugs in food-producing animals) or from food processing and packaging (e.g., nitrosamines and certain polycyclic aromatic hydrocarbons). These contaminants may present a potential hazard for human health if exposure exceeds tolerable levels.

Physical Hazards

Physical hazards are foreign matter, including dirt and glass, that accidentally get into food They can be the result of environmental contamination during production, processing, storage, packaging, and transport, or from fraudulent practices. The potential for

Table 1 Examples of main biological hazards in foods

Bacteria (spore-forming)	Viruses
Clostridium botulinum	Hepatitis A and E
Clostridium perfringens	Norwalk virus group
Bacillus cereus	Rotavirus
Bacteria (nonspore-forming)	Protozoa and parasites
Brucella abortis	*Cryptosporidium parvum*
Brucella suis	*Diphyllabothrium*
Campylobacter spp.	*Entamoeba histolytica*
Pathogenic *Escherichia coli*	*Giardia lamblia*
(*E. coli* O157:H7, EHEC, EIEC,	*Ascaris lumbricoides*
ETEC, EPEC)	*Taenia solium*
Listeria monocytogenes	*Taenia saginata*
Salmonella spp. (*S. typhimurium, S. enteritidis*)	*Trichinella spiralis*
Shigella (*S. dysenteriae*)	
Staphylococcus aureus	
Streptococcus pyogenes	
Vibrio cholera	
Vibrio parahaemolyticus	
Vibrio vulnificus	
Yersinia enterocolytica	

Source: FAO (1998) *Food Quality and Safety Systems. A Training Manual on Food Hygiene and the Hazard Analysis and Critical Control Point (HACCP) System*. Rome: Food and Agricultural Organization of the United Nations, with permission.

Table 2 Examples of chemical hazard in foods

Industrial/environmental pollution	Additives
Polychlorinated biphenyls (PCBs)	Feed additives
Agricultural chemicals	Veterinary drugs
Pesticides	Food additives
Fertilizers	Vitamins and minerals
Antibiotics	Contaminants
Growth hormones	Lubricants
	Cleaners
Toxic elements and compounds	Pest control chemicals
	Coatings
	Paints
Inorganic chemicals	Chemicals migrating
Lead	from packaging
Tin	Plasticizers
Mercury	Vinyl chloride
Zinc	Printing/coding inks
Cadmium	Adhesives
Arsenic	Lead
Cyanide	Tin

Source: FAO (1998). *Food Quality and Safety Systems. A Training Manual on Food Hygiene and the Hazard Analysis and Critical Control Point (HACCP) System*. Rome: Food and Agricultural Organization of the United Nations, with permission.

Table 3 Examples of physical hazards

Material	*Sources*
Glass	Bottles, jars, light fixtures, utensils, gauge covers, etc.
Wood	Field sources, pallets, boxes, building materials
Stones	Fields, buildings
Metal	Machinery, fields, wire
Insulation	Building materials
Bone	Improper processing
Plastic	Packaging, pallets, equipment
Personal effects	Employees

Source: FAO (1998) *Food Quality and Safety Systems. A Training Manual on Food Hygiene and the Hazard Analysis and Critical Control Point (HACCP) System*. Rome: Food and Agricultural Organization of the United Nations, with permission.

ionizing radiation to have long-term health effects, not only on the people living nearby but also the health of the ecosystem, is considerable. Following the nuclear reactor failure at Chernoble (Ukraine) in 1986, food contaminated with radionucleides with long half-lives, such as cesium 137, is the major source of exposure for people living in this area. Table 3 lists examples of physical hazards in food.

Food Safety Emergencies

The ultimate objective of the food industry and government regulators is to insure that food reaching the consumer is safe and wholesome. Feed manufacturers, farmers, and food operators have the primary responsibility for food safety. Government regulators and consumer associations also have a role to play. However, food-safety emergencies over the last two decades have exposed weaknesses in the food industry from farm to fork. They have shown that considerable progress is needed to come to grips with the preventive aspects of foodborne disease.

Many factors have contributed to recent food emergencies. The food production chain is becoming increasingly complex because of mass production. Modern farming practices are intensive and can result in microorganisms contaminating a large number of crops or infecting a large number of animals. Healthy animals may carry pathogens that cause disease in humans. Examples include *Salmonella* spp., *Listeria* spp., and some strains of *Escherichia coli* and *Campylobacter*. Animals may be infected from feed, from other animals, or from the environment.

Food distribution systems have expanded dramatically over recent decades. This has lengthened the food chain adding opportunities for contamination to occur. Ingredients may come from many different countries and be combined in a single product. The Food and Agricultural Organization (FAO) of the United Nations estimates that over 500 million tonnes of food move in international trade, valued at around $400–500 billion, each year.

Trends show that consumers are relying increasingly on convenience and ready-prepared foods. Such foods have few barriers to microbial growth, such as salt and preservatives, and simple mistakes could easily result in an outbreak of foodborne disease.

World Trade

The Food and Agriculture Organization (FAO) and the World Health Organization (WHO) have complementary food-safety mandates to protect the health of consumers, to prevent the spread of disease and to ensure that the procedures followed in food trade are fair. The Codex Alimentarious Commission (CAC), which manages the joint FAO/WHO food-standards program, sets international standards for food. It was established in 1961 and 1962 by the FAO and WHO and aims to protect the health of consumers and ensure fair practices in food trade.

The World Trade Organization (WTO) rules provide a framework for the application of food-safety measures in international trade. The standards and guidelines of the CAC are recognized by the WTO as the international benchmark standards, guidelines, and recommendations to be used as the basis of harmonizing food safety measures affecting human health and world food trade.

International standards influence trade as well as food safety, and national governments are under pressure to ensure that their regulatory programs are comprehensive and conform to these global standards. Owing to the complexity of the food-production chain, the traditional role of government as the food-safety inspector has changed, with the food industry taking more responsibility for the implementation of quality-assurance programs throughout food production, distribution, and retail. However, a strong national food control agency is required to bring about cooperation between government departments and industry, and to set standards and targets for health protection.

Risk Analysis

The concept of risk analysis is gaining acceptance internationally as a very important component of national food control programs, and its principles are being integrated more fully into the work of food safety regulators and policy-makers. This principle consists of three components: risk assessment, risk management, and risk communication.

Risk assessment is a science-based investigation that can provide an estimate of the probability and impact of adverse health effects attributable to potentially contaminated foods. The aim is to identify, in some quantitative or comparable way, the relationship between hazards and actual exposure to harm. The detail and complexity of a given risk assessment will vary, depending on the availability of time, resources, and scientific information. A full risk assessment may not always be carried out, because data may be incomplete, and resources may be inadequate. However, decisions that are made by well-focused assessments, using whatever data are available, will usually be more reliable and transparent than subjective judgement. Risk assessment links into the other two components of risk analysis, the others being risk management and risk communication. The stages of the risk analysis process are closely linked and cannot be separated easily. Consideration of the ways in which a risk might be managed is necessarily part of the assessment and communication.

Risk assessment consists of four steps: hazard identification, hazard characterization, exposure assessment, and risk characterization (Figure 1). In hazard identification, the association between disease and the presence of a pathogen or contaminant in food is documented. Hazard characterization involves obtaining quantitative information about the hazard, including, where possible, information on dose–response relationships. Exposure assessment estimates the intake or exposure to a chemical or pathogen, in terms of its magnitude, duration, and frequency, for the general population, for subgroups or for individuals. Risk characterization is the integration of hazard identification, hazard characterization, and human intake/exposure assessment. It assesses the likelihood of a particular event. This is the framework adopted by the CAC, the international standard setting body for foods in international trade.

Risk assessment is already fairly well developed for chemical hazards. The estimation of chemical health risks often relies on data obtained in the customary

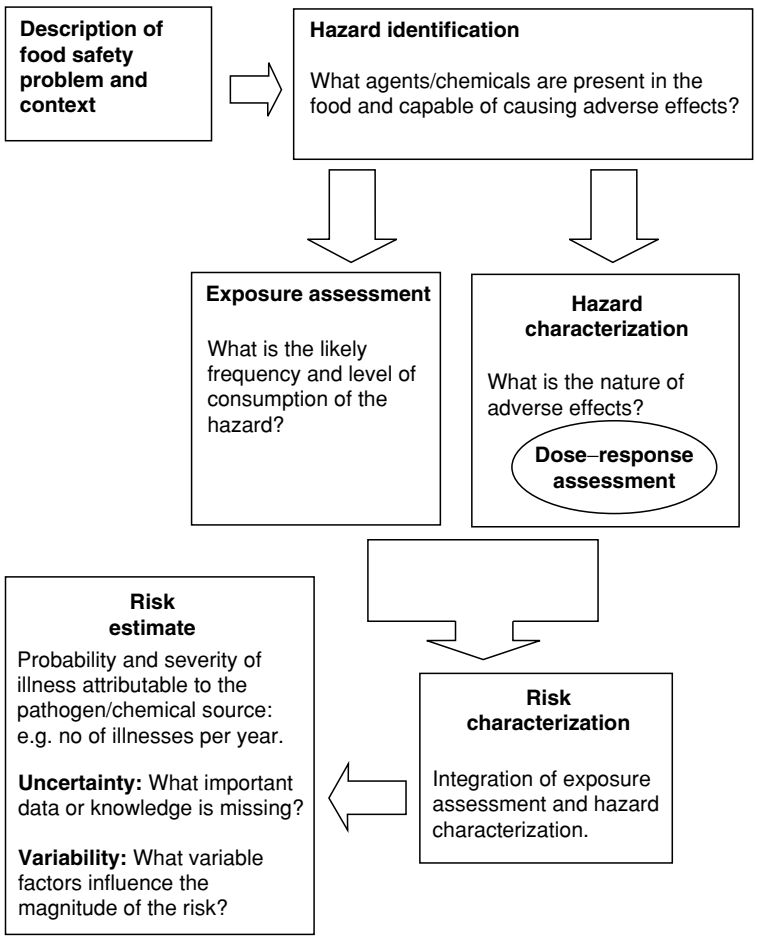

Figure 1 Steps of food-safety risk assessment. Adapted from Lammerding AM and Fazil A (2000) Hazard identification and exposure assessment for microbial food safety risk assessment. *International Journal of Food Microbiology* 58: 147–157, with permission.

rodent feeding assays using standardized safety margins. Such assessments call for an extrapolation of response data derived from animal studies to humans. The setting of advisory standards, such as acceptable daily intakes (ADIs) for food additives, pesticides, and veterinary residues in food, is a key component in the overall risk assessment of many chemicals subject to regulatory control. Quantitative risk assessment is an emerging tool in the field of microbial food safety. Estimation of microbiological hazards presented by foods is more complex, because of biological diversity and variability. However, such estimations are usually more accurate than estimates on adverse effects from chemicals in foods. This is because incidents are frequent, in sharp contrast to those resulting from exposure to chemicals. This allows for a better risk assessment.

Risk Management and Risk Communication

Food scares such as salmonellae in eggs, *Escherichia coli* O157 in beef and BSE have all damaged consumer confidence in regulatory agencies. Such problems were compounded when government departments had the dual role of promoting farming and the food industry as well as protecting the consumer. Such roles resulted in a conflict of interest. A regulatory agency should ideally be independent of trade and economic interests and have consumer interests as its primary function. Although it is the primary responsibility of the food industry to comply with the laws and regulations, the regulatory agency should provide an integrated inspection system from farm to fork. This may best be achieved by establishing a national food standards agency that has responsibility for all sectors of the food chain. Many countries have established such agencies and coordinate the work of all government departments involved in food safety. Such coordination overcomes the problems of duplication of work and conflicts of interest.

A foodborne disease surveillance program is also an essential part of a food control program. Surveillance is necessary to identify foodborne contaminants, their sources, and their modes of transmission in order to determine the best short- and long-term control measures. Interpretation of trends and investigation of outbreaks can help identify the mechanisms by which contamination and disease transmission occurred. When an outbreak occurs or a chemical contamination incident is identified, it is vital to learn from the investigation what went wrong in order to devise strategies to prevent similar events in the future.

Quality Assurance

Food quality and safety can only be insured through the application of quality-control systems throughout the entire food chain. They should be implemented at farm level with the application of good agricultural practices and good veterinary practices at production, good manufacturing practice at processing, and good hygiene practices at retail and catering levels. One of the most effective ways for the food sector to protect public health is to base their food management programs on Hazard Analysis Critical Control Point (HACCP).

HACCP consists of seven principles that outline how to establish, implement, and maintain a quality-assurance plan for a food establishment. It is a systematic approach to the control of potential hazards in a food operation and aims to identify problems before they occur. Control is proactive, since remedial action can be taken. The HACCP principles have international recognition, and the seven principles are listed in Table 4.

HACCP plan adoption has greatly improved the food industry's ability to design programs to insure the safety of food. It was introduced in the early 1970s and has become the premier system for preventing foodborne hazards, particularly those of microbiological origin. The basic process in creating HACCP plans is that the main hazards associated with a food product are determined, and key steps in the production, processing, distribution, marketing, and preparation of food are controlled

Table 4 Seven principles of HACCP

Principle 1
Conduct a hazard analysis. Prepare a list of steps in the process where significant hazards occur and describe the preventive measures. Assess the likelihood of occurrence of the hazard(s) and identify the measures for their control.

Principle 2
Identify the Critical Control Points (CCPs) in the process. Determine the points, procedures or operational steps that can be controlled to eliminate the hazard(s) or minimize its (their) likelihood of occurrence.

Principle 3
Establish critical limits for preventative measures associated with each identified CCP.

Principle 4
Establish CCP monitoring requirements. Establish procedures from the results of monitoring to adjust the process and maintain control.

Principle 5
Establish corrective actions to be taken when monitoring indicates a deviation from an established critical limit.

Principle 6
Establish effective record-keeping procedures that document the HACCP system.

Principle 7
Establish procedures for verification that the HACCP system is working correctly.

Source: Mortimore S and Wallace C (1997) *HACCP: A Practical Approach*. London: Chapman & Hall, with permission.

within preestablished limits to insure the safety of food. A program of monitoring, verification, and record keeping is then implemented to insure that the system functions correctly.

The Codex General Principles of Food Hygiene recommend a HACCP-based approach to enhance food safety. This management tool is internationally recognized as essential to ensure food safety and is recommended by governments, industry, and consumers.

Conclusion

Food safety in the early twenty-first century is an international challenge requiring close cooperation between countries in agreeing standards and in setting up transnational surveillance systems. The lessons of the past two decades are plain to those engaged in the food industry. No longer can farmers grow just what they want or use technical aids to farming without taking into account the effect on the quality of the food produced. Food will always present some risk, and it is the task of the food industry to keep the level of risk to the minimum, which is practicable and technologically feasible. It should be the role of official bodies and the food industry to use risk analysis to determine realistic and achievable risk levels for foodborne hazards and to base food-safety practices on the practical application of the results of these analyses, thus continuously improving the safety of food and thereby lowering the disease risk. By providing effective food-control, countries not only provide public health benefits but enable themselves to participate in international food trade with greater confidence.

The lasting solution to the problems, which have shaken the food industry in recent years, can only be overcome if all those involved – farmers, food processors, wholesalers, transporters, retailers, caterers, scientists, regulators, and government – work together towards common goals.

See also: **Bovine Spongiform Encephalopathy (BSE)**; **Contamination of Food**; ***Escherichia coli***: Occurrence and Epidemiology of Species other than *Escherichia coli*; **Food and Agriculture Organization of the United Nations**; **Hazard Analysis Critical Control Point**; **Legislation**: Codex; ***Listeria***: Listeriosis; **Quality Assurance and Quality Control**; **Risk Assessment**; **World Health Organization**; **World Trade Organization**

Further Reading

Aston G and Tiffney J (1997) *The Essentials to Food Hygiene and Safety*. Surrey: Eaton Publications.

Benenson AS (1995) *Control of Communicable Diseases Manual*, 16th edn. American Public Health Association.

Bhat RV and Moy GG (1997) Monitoring and assessment of dietary exposure to chemical contaminants. Food safety and foodborne diseases. *World Health Statistics Quarterly* 50 (1–2): 132–149.

Bryan F (1992) *Hazard Analysis Critical Control Point. A Guide to Identifying Hazards and Assessing Risks Associated With Food Preparation and Storage*. Geneva: WHO.

Buchanan RL and Whitling RC (1998) Risk assessment: A means for linking HACCP plans and public health. *Journal of Food Protection* 61(11): 1531–1534.

Chesworth N (1997) *Food Hygiene Auditing*, 1st edn. London: Chapman & Hall.

Cliver DO (1990) *Foodborne Diseases*. London: Academic Press.

Commission of the European Communities (1999) *White Paper on Food Safety*. Com, 719 Final. Brussels.

Dillion M and Griffith C (1996) *How to HACCP*, 2nd edn. London: M. D. Associates.

Eley A (1996) *Microbial Food Poisoning*, 2nd edn. London: Chapman & Hall.

Falconer L (1999) Risk Assessment and Risk Management. In: Bassett WH (ed.) *Clays Handbook of Environmental Health*. E&FN Spon.

FAO (1998) *Food Quality and Safety Systems. A Training Manual on Food Hygiene and the Hazard Analysis and Critical Control Point (HACCP) system*. Rome: Food and Agricultural Organization of the United Nations.

FAO (1999) *Report of the Conference on International Food Trade Beyond 2000: Melbourne, Australia. 11–15 October 1999*. Rome: Food and Agriculture Organization of the United Nations.

FAO/WHO (1995) *Application of Risk Analysis to Food Standards Issues. Report of the Joint FAO/WHO Expert Consultation*. Geneva: WHO.

Institute of Environmental and Health (1999) *Developing New Approaches to Assessing Risk to Human Health from Chemicals*. Report Prepared by the Risk Assessment and Toxicology Steering Committee.

Institute of Environmental and Health (1999) *Risk Assessment Approaches used by the Government for Evaluating Human Health Effects of Chemicals*. Report Prepared for the Risk Assessment and Toxicology Steering Committee.

Jay JM (1996) *Modern Food Microbiology*, 5th edn. London: Chapman & Hall.

Jouve JL (2000) Special issue: ILSI Europe Session on Microbiological Risk Assessment. *International Journal of Food Microbiology* 58: 141–142.

Kaferstein FK, Motarjemi Y and Bettcher DW (1997) Foodborne disease control: a transnational challenge. *Emerging Infectious Diseases* 3(4): 503–510.

Lammerding AM and Fazil A (2000) Hazard identification and exposure assessment for microbial food safety risk assessment. *International Journal of Food Microbiology* 58: 147–157.

Lammerding AM and Paoli GM (1997) Quantitative risk assessment: an emerging tool for emerging foodborne pathogens. *Emerging Infectious Diseases* 3.

Leaper S (1997) *HACCP: A Practical Guide*. Technical Manual No. 38. Campden and Chorleywood Food Research Association, UK.
Lupien JR and Kenny MF (1998) Tolerance limits and methodology: effect on international trade. *Journal of Food Protection* 61(11): 1571–1578.
Mortimore S and Wallace C (1994) *HACCP: A Practical Approach*. London: Chapman & Hall.
Mossel DA, Struijk CB, van der Zwet WC, Heijkers FH and Dankert J (1998) Identification, assessment and management of food-related microbiological hazards: historical, fundamental and psycho-social essentials. *International Journal of Food Microbiology* 40(3): 211–243.
Nothermans S, Zwietering MH and Mead GC (1994) The HACCP concept: identification of potentially hazardous microorganisms. *Food Microbiology* 11: 203–214.
Pearson AM and Dutson TR (1995) *HACCP in Meat, Poultry and Fish Processing. Advances in Meat Research Series*, 1st edn., vol. 10. Glasgow: Blackie Academic & Professional.
Philips J (1997) *Food Standards Agency. An Interim Proposal by Professor Philip James*. London: FSA.
Roberts D and Hobbs BC (1993) *Food Poisoning and Food Hygiene*, 6th edn. London: Arnold.
Sprenger RA (1993) *Hygiene for Management. A Text for Food Hygiene Courses*, 6th edn. Highfield.
Tauxe RV (1997) Emerging foodborne diseases: An evolving public health challenge. *Emerging Infectious Diseases* 3(4): 425–434.
Todd EC (1997) Epidemiology of foodborne diseases: a worldwide review. *World Health Statistics Quarterly* 50(1–2): 30–50.
Voysey PA and Brown M (2000) Microbiological risk assessment: a new approach to food safety control. *International Journal of Food Microbiology* 58: 173–179.

FOOD SECURITY

J T Cook, Boston University School of Medicine, Boston, MA, USA

Background

This article is about food security in the USA and the factors that promote it. It is also about poverty-related food insecurity and hunger, their prevalence in the US population, and some of their causes and consequences. The article is not about normal hunger *per se*, although normal hunger is a critical element of the definition and measurement of food insecurity severe enough to include hunger. Instead, this article is about hunger that people in civilized societies agree should not exist. It is about hunger experienced involuntarily in a country that claims to be the wealthiest in the world; hunger experienced by several million people because they lack adequate resources to afford enough food for an active and healthy life. This kind of hunger is most accurately seen as a social problem, and the circumstances and managed process within which it occurs are known as food insecurity.

In pursuit of objectives of the National Nutrition Monitoring and Related Research Act of 1990, the US Department of Agriculture's (USDA) Food and Consumer Services (FCS) and the Department of Health and Human Services' National Center for Health Statistics (NCHS) were assigned responsibility for establishing a research program to develop standardized measures of food security, food insecurity and hunger for the US population. That research program was initiated in 1992 as the Food Security Measurement Project (FSMP). In the first phase of this ongoing research program, a household-level survey instrument was developed for annual administration by the US Census Bureau as a supplement to its nationally representative Current Population Survey (CPS). In the second phase, data from the CPS Food Security Supplement, administered to a sample of approximately 60 000 households, were used to create and validate a food security scale for the US population.

The remainder of this article, briefly describes the food security measures, their application over the period 1995–2000 in the FSMP, and some of the main findings from this research. I also discuss aspects of the social, political, and economic contexts within which the food security measures were applied during this period as they relate to the causes and consequences of food insecurity and hunger in the USA.

What is Food Security?

The term food security came into usage over the past two decades to describe a person's, household's, or community's ability to obtain enough nutritious food for a healthy life. The concept of food security initially emerged during the 1980s in international

development work.[1] It was adopted in the early 1990s as a useful framework for describing, researching, and designing policies to address poverty-related food access problems at the household level in the USA. Research conducted during the 1980s and 1990s in the USA found that household food security is a managed process rather than a state or condition. Within this framework, a household's 'food manager' obtains, prepares, and makes food available to household members in a way determined by the household's overall resources and preferences.

When a household's resources (usually derived from money income and/or benefits from social safety-net programs) are plentiful, it is likely to be food-secure. If, however, household resources become scarce for a period of time, the food manager (usually the mother or another adult in the household) undertakes a managed process aimed at ensuring that sufficient food will be available to enable household members to avoid hunger.

This managed process often involves a variety of coping strategies undertaken to supplement the household's food supply. These include actions such as putting off paying rent or bills, borrowing money or food from friends or family members, reducing the variety of foods prepared and served (often by cutting out fresh fruits and vegetables), and relying on low-cost 'filling' foods. The process of managing household food insecurity often involves considerable emotional stress for members of the household as worries about whether available food will last, or whether there will be money available to buy food, intensify. Households in this condition are said to have become food-insecure.

If a food-insecure household's resource scarcity or food insufficiency persists or worsens, the household food manager may be forced to take actions that lead to reductions in the quality and/or quantity of food available to household members. In food-insecure US households with children, it is typical for adults to reduce their food intake below normal levels (by reducing the size of meals, or skipping meals) to spare the children from experiencing hunger. However, if the household's food insecurity continues or becomes more severe, food intake by children in the household eventually also will be reduced below normal levels. This usual pattern of rationing makes it possible to identify two levels of severity of hunger in US households; moderate hunger and severe hunger. When the former occurs in households with children, usually only adults in the household experience hunger. With the latter, both adults and children experience hunger.

[1]The term food security appears extensively in the work of the United Nations' Food and Agriculture Organization (UN/FAO), the Consultative Group on International Agricultural Research (CGIAR), the International Food Policy Research Institute (IFPRI), the World Bank, and other international development organizations during the 1980s and 1990s. In 1994, the UN/FAO launched the Special Program for Food Security, 'a multidisciplinary program that combines expertise and experience from a wide range of fields to promote an integrated and participative approach to food security'. In November 1996, representatives from 185 countries and the European Community met in Rome for the World Food Summit. Out of that week-long conference emerged the Rome Declaration on World Food Security and an accompanying Plan of Action to reduce food insecurity and hunger by half in all countries by 2015.

What is Hunger?

Hunger is a feeling that everyone has some understanding of because all animals experience hunger, as far as we know. At the individual level, hunger is a physiological state or condition, a set of neurological sensations, and a psychological drive. Hunger is an uneasy or painful sensation caused by a lack of food, and it is a motivation to obtain and consume food. We all know when we are hungry, and we can even say how hungry we are and how it feels. We may be 'so hungry we could eat a horse,' or we may 'just need a little something.' For humans, hunger is usually a normal and healthy response to emptying of food and nutrients from the stomach and upper gastrointestinal tract.

But what about hunger that poor people in developing countries experience, or poor people in developed countries who do not have enough money or other resources to buy, store, and prepare sufficient food for a healthy active life? How is this hunger different from 'normal' hunger? When does hunger become problematic? At what point does it rise to the level of a social concern? Is it when the motivation to obtain and consume food cannot be satisfied due to a lack of resources? How often must this condition exist, for how long, and for how many people and families before it is seen as a social problem that merits a policy response? Historically, concerns about food insufficiency have been articulated in terms of malnutrition, undernutrition, and nutrient deficiencies. How are food insecurity and hunger related to malnutrition? Are these conditions synonymous, or can they exist independently?

These kinds of questions are critical to a clear understanding of food insecurity and hunger as social problems. They are also central to accurate definition and measurement of food security, food insecurity, and hunger, and to the design and implementation of effective policies to prevent or reduce these conditions.

Measuring Food Security

To design an effective research program for developing measures of food security, food insecurity, and

hunger for use in the public policy arena, a federal interagency team first identified an appropriate measurement framework. This measurement framework expressed practical limits for the scope of the measures and clarified the contexts to which they would be applied and the type of instruments, data, and analytical methods from which they would be derived.

Consensus conceptual definitions of food security, food insecurity and hunger for the US context were derived and published by the Life Sciences Research Office (LSRO) of the Federation of Associations for Experimental Biology (FASEB) in collaboration with the American Institute of Nutrition (AIN) in 1990. These conceptual definitions were operationalized, and empirical measures of food security developed under sponsorship of USDA/FCS and CDC/NCHS in the Food Security Measurement Study (FSMS) of 1995–1997.

Data from initial administration of the Food Security Supplement in 1995 were used in a split-sample procedure to create and validate a food security scale. Population-weighted data were also used to produce estimates of the prevalence of food security, food insecurity, and hunger in the US population. The Food Security Supplement has been administered by the Census Bureau annually since 1995 as part of the CPS. A time-series of annual food security data and resulting prevalence estimates for the noninstitutionalized US population by major sociodemographic characteristics is available for the period 1995–2000. In addition, overall state-level prevalence estimates are available for 1995–1998.

Definition and Measurement of Food Security

The conceptual definitions of food security, food insecurity, and hunger synthesized by the LSRO's expert panel have been stated as:

> *Food security*: Access by all people at all times to enough food for an active, healthy life.
>
> *Food insecurity*: Limited or uncertain availability of nutritionally adequate and safe foods or limited or uncertain access to food.
>
> *Hunger*: The uneasy or painful sensation caused by a lack of food. The recurrent and involuntary lack of access to food ... Hunger, in its meaning of the uneasy or painful sensation caused by a lack of food, is ... a potential, although not necessary, consequence of food insecurity. Malnutrition is also a potential, although not necessary, consequence of food insecurity.

Item response theory methods involving nonlinear (logistic) factor analytic scaling models were used to develop the 18-item food security scale from responses to a larger number of questions in the Food Security Supplement to the CPS. When operationalized, the three conceptual categories listed above were shown to represent adjacent levels of severity of a single well-ordered phenomenon. Food security emerged as a continuum, with households at the least-severe level termed food secure, and those at the most severe level termed food insecure with severe hunger, as follows:

> *Food secure*: Household shows no or minimal evidence of food insecurity.
>
> *Food insecure without hunger*: Food insecurity is evident in households' concerns and adjustments to household food management, including reductions in diet quality, but with no or limited reductions in quantity of food intake.
>
> *Food insecure with moderate hunger*: Food intake for adults in the household is reduced to an extent that implies that adults experience hunger due to a lack of resources. If children are present, the quality of food available to them may be reduced, but usually not its quantity.
>
> *Food insecure with severe hunger*: Households with children reduce the children's food intake to an extent that implies that the children experience hunger as a result of inadequate household resources. Adults in households with or without children experience extensive reductions in food intake.

The two categories including hunger are based on survey responses indicating that food intake has been reduced below normal levels (e.g., by reducing the size of meals, skipping meals, or going a whole day without eating) for adults or children in the household, or both, and that these reductions occurred specifically because the household did not have enough food or money to buy food. Cut-point selection procedures required that a pattern of repetition of such intake reductions occur for two or more months out of the previous 12 months for hunger to be confidently inferred. Thus, even though hunger is a normal state experienced by all people, the Food Security Scale was designed to capture recurrent resource-constrained hunger, experienced because a household does not have sufficient food or financial resources to buy food.

Prevalence of Food Insecurity and Hunger in the USA

The Food Security Scale produces continuous interval-level (without a true zero value) household food security scores for all households in which an adult respondent completes the survey. The continuous scores are then used to categorize households into one of the four categories just described on the basis of cutoff values determined in the initial FSMS and

validated with data from the five successive implementations of the scale (in 1996–2000).

In recent years, with high rates of growth in the US economy and very low unemployment, increasingly smaller proportions of most demographic subgroups experienced food insecurity with severe hunger. As a result, USDA analysts combined the two severity levels involving hunger and now report prevalence estimates for three levels of severity only (food secure, food insecure *without* hunger, and food insecure *with* hunger). Table 1 shows the prevalence of these three food security categories for US households and residents by selected characteristics in 2000, the latest year for which estimates are available.[2]

Overall, 11.1 million US households (10.5%) were food-insecure at some level of severity in 2000. In 3.3 million households (3.1%), at least one adult or child experienced hunger.[3] Just over 33.0 million people (12.1%) lived in food-insecure households in 2000, and 9.0 million lived in households where hunger was experienced. Examination of Table 1 shows that people in households with children (ages < 18 years) have almost two and a half times the risk of being food-insecure as those in households without children (relative risk (RR): 0.164/0.067 = 2.4). People in female-headed households with children and no spouse have nearly three times the risk of being food-insecure as those in married-couple households with children (RR = 2.8).

Other notable comparisons of the prevalence of food insecurity and hunger include those by race/ethnicity of household, ratio of household income to poverty, and area of residence. People in non-Hispanic Black and Hispanic households have nearly three times the risk of being food insecure as non-Hispanic White households (RR = 2.8 and 2.9, respectively, compared to non-Hispanic Whites), and people in central-city households have nearly twice the risk of being food insecure as people in metropolitan households not in central cities (RR = 1.9). Households with incomes below 185% of the poverty threshold have over six times the risk of being food insecure as those with incomes greater than or equal to 185% of poverty (RR = 6.3). The prevalence of food insecurity and hunger rises steadily as the ratio of household income to poverty decreases (similar to a dose–response effect), with the highest prevalence among households whose incomes are below 100% of poverty.

Food Insecurity, Hunger, and Malnutrition

It is important to note the relationships among poverty, food insecurity, hunger, and malnutrition. In the consensus conceptual definitions derived by the LSRO/AIN expert panel (listed above) and operationalized in the FSMS, hunger and malnutrition are described explicitly as 'potential, although not necessary' consequences of food insecurity. Resource-constrained hunger is 'nested' within food insecurity, occurring at its more severe levels. It is accurate to say that hunger, so defined, is both a necessary and sufficient condition to imply food insecurity.

Malnutrition in its broadest meaning, however, is neither a necessary nor a sufficient condition to imply food insecurity or hunger, though it may. Malnutrition, both as undernutrition and overnutrition, can occur in the absence of food insecurity or hunger, though either or both of these latter conditions may be associated with malnutrition and may even cause it. Clinically, however, there can be multiple causes of malnutrition, some of which do not involve food insecurity or hunger.

It is unlikely, in the absence of morbidity, congenital anomaly, or pathology, that protein–calorie undernutrition occurs in the US population, unless it is accompanied by relatively severe food insecurity or hunger. However, it is not at all uncommon for micronutrient undernutrition or deficiencies to occur under conditions of food insecurity without hunger, or even under food-secure conditions. Indeed, prominent clinical concerns often arise due to micronutrient deficiencies associated with food insecurity short of measurable hunger, or with only moderate hunger.

In the past two decades, consequences of overnutrition have emerged as important factors in some of the most serious threats to human health (e.g., diabetes, hypertension, cardiovascular disease, orthopedic conditions, sleep apnea, asthma, and cancer). Overweight and obesity are epidemic in the USA across nearly all age levels, with growing concerns regarding emergence of obesity at earlier childhood ages and implications for later body composition and health. Trends in the US economy and society, including effects of technological change on prices and availability of food, patterns of food advertising, marketing, consumption and eating behavior, work, and physical activity, have led to conditions that

[2]As of writing, the data for 2001 have been collected but not analyzed. The USDA and the Census Bureau are preparing for implementation of the 2002 CPS survey.

[3]Since questions in the Food Security Supplement do not ask specifically about the conditions of each member of the household, it is not possible to ascribe hunger status to all members of households with more than one adult and one child. One can, however, ascribe overall food insecurity status to all members of any household that is not food-secure. In addition, if a household is categorized as 'food insecure with hunger,' it is appropriate to say that all members live in a household where hunger is experienced. This minor unintended limitation results from the specific form of the survey and the questions it contains.

Table 1 Prevalence of food security, food insecurity without hunger, and food insecurity with hunger for persons by selected characteristics of households: 2000[a]

Category	*Total*	*Food-secure*		*Food-insecure (all)*		*Food-insecure with no hunger*		*Food-insecure with hunger*	
	(1000s)	*(1000)*	*(%)*	*(1000)*	*(%)*	*(1000)*	*(%)*	*(1000)*	*(%)*
All households	106 043	94 942	89.5	11 101	10.5	7 786	7.3	3 315	3.1
All Persons	273 685	240 454	87.9	33 231	12.1	24 708	9.0	8 523	3.1
Adults	201 922	181 586	89.9	20 336	10.1	14 763	7.3	5 573	2.8
Children < 18 years	71 763	58 868	82.0	12 895	18.0	9 945	13.9	2 950	4.1
Household composition									
With children < 18 years	152 995	127 857	83.6	25 138	16.4	19 489	12.7	5 649	3.7
With children < 6 years	72 810	59 413	81.6	13 397	18.4	10 580	14.5	2 817	3.9
Married-couple families	112 734	99 496	88.3	13 238	11.7	10 782	9.6	2 456	2.2
Female head, no spouse	30 705	20 715	67.5	9 990	32.5	7 204	23.5	2 786	9.1
Male head, no spouse	7 410	5 964	80.5	1 446	19.5	1 123	15.2	323	4.4
With no children < 18 years	120 690	112 597	93.3	8 093	6.7	5 219	4.3	2 874	2.4
More than one adult	93 196	87 756	94.2	5 440	5.8	3 694	4.0	1 746	1.9
Women living alone	16 157	14 526	89.9	1 631	10.1	977	6.0	654	4.0
Men living alone	11 336	10 313	91.0	1 023	9.0	549	4.8	474	4.2
Households with elderly	47 580	44 416	93.4	3 164	6.6	2 396	5.0	768	1.6
Elderly living alone	10 125	9 409	92.9	716	7.1	523	5.2	193	1.9
Race/ethnicity of households									
White non-Hispanic	195 171	178 962	91.7	16 209	8.3	11 759	6.0	4 450	2.3
Black non-Hispanic	33 505	25 755	76.9	7 750	23.1	5 631	16.8	2 119	6.3
Hispanic (of any race)	32 945	24 920	75.6	8 025	24.4	6 365	19.3	1 660	5.0
Other non-Hispanic	12 065	10 818	89.7	1 247	10.3	953	7.9	294	2.4
Ratio of household income to poverty									
Under 1.00	33 447	19 750	59.0	13 697	41.0	9 763	29.2	3 934	11.6
Under 1.30	48 786	30 852	63.2	17 934	36.8	12 991	26.6	4 943	10.1
Under 1.85	71 509	49 402	69.1	22 107	30.9	16 279	22.8	5 828	8.2
1.85 and over	163 288	155 215	95.1	8 073	4.9	6 348	3.9	1 725	1.1
Area of residence									
Inside metropolitan area	221 518	195 268	88.1	26 260	11.9	19 467	8.8	6 783	3.1
In central cities	65 772	54 511	82.9	11 261	17.1	8 451	12.8	2 810	4.3
Not in central cities	117 791	107 349	91.1	10 442	8.9	7 579	6.4	2 863	2.4
Outside metropolitan areas	52 167	45 186	86.6	6 981	13.4	5 241	10.0	1 740	3.3

[a]Totals exclude households whose food security status is unknown.

From Nord M, Kabbani N, Tiehen L *et al*. (2002) *Measuring Food Security in the United States: Household Food Security in the United States, 2000. USDA/ERS Food Assistance and Nutrition Research Report No. 21*. Washington, DC:

support widespread overconsumption of calories relative to daily needs.

Since the 1970s, research literature has accumulated, suggesting that food insecurity may play a role in the onset of overweight and obesity among some low-income subpopulations. Though the evidence is not yet conclusive, there are several lines of research currently underway whose results to date are consistent with hypothesized associations between food insecurity and obesity. To the extent that there are associations between food insecurity and obesity, affected low-income subpopulations may be doubly challenged in their efforts to maintain healthy body composition, diets and lifestyles.

Poverty, Food Insecurity, and Social Welfare Programs

A clear understanding of food insecurity and hunger in the USA requires knowledge of their relationships to poverty, and the range of social policies and programs that have emerged to address both. By definition, resource-constrained food insecurity is primarily an issue for low-income families, or families with limited resources. However, the nature of the US economy, food system, and social welfare system adds to heterogeneity in expression of food insecurity. The way in which poverty and food security are defined and measured in the USA also adds complexity to relationships between these two conditions.

US Poverty Measures

When created in 1963, the US poverty levels were based on the cost of a minimally nutritious diet. Though these thresholds have been updated annually since 1963 for inflation, they have not yet been revised to reflect the change in proportion of total household expenditures spent for food. In the 1960s, the average household spent roughly one-third of its total monthly expenditures for food (averaged over

all households at all income levels). In principle, this implied that multiplying the average cost to households of a minimally nutritious diet by a factor of three would provide an indication of the poverty threshold, or the minimum amount of income needed to meet basic needs.

Using the cost of the USDA's Thrifty Food Plan for different size families as estimates of the costs of minimally nutritious diets, the poverty thresholds were obtained by multiplying these dollar values by three. This remains the basis of poverty measurement in the USA today. Elegant as it is in logic and simplicity, this measure only succeeds when the relative proportion of the average household budget spent on food is accurate. Setting aside the question of whether this is the most accurate approach to estimating the cost of households' basic needs, a more fundamental concern is whether the proportion of household income spent on food has remained constant over time.

Since 1963, the average annual proportion of household expenditures spent for food has declined consistently to about 11% in 2000. Over the same period, the average annual proportion of expenditures spent on other basic needs (most notably housing and transportation) increased consistently. As a result, using the same multiplier logic underlying the initial definition of the poverty thresholds with current proportions of expenditures on food implies a multiplier of approximately 9 instead of 3. Obviously this would lead to much higher income levels being identified as poverty thresholds, and much larger numbers and proportions of people identified as being in poverty.

Another peculiar anachronism in US poverty measurement is that households headed by elderly persons (ages 65 and over) have lower poverty thresholds than similar households headed by people under 65 years of age. This practice dates to an era in which most Americans worked in occupations involving extensive physical exertion, retirement from which ostensibly reduced energy and other nutrient requirements and hence the cost of food for people over age 65. Depending on the size and characteristics of the household (e.g., the number of related children under the age of 18 years), poverty thresholds for households headed by elderly people are 4–10% lower than for comparable households headed by nonelderly. Thus, elder-headed households must have lower incomes (be worse off financially) than nonelder-headed households before they are classified as being in poverty. As a result, the number and proportion of elderly households and persons reported in poverty understate the actual extent and dimensions of poverty among the elderly. In turn, this affects eligibility for assistance from various social welfare programs among some elderly who may be in need.

Support for the view that the current poverty thresholds understate the costs of basic needs and therefore lead to underestimation of the level and proportion of the US population at risk for health issues associated with poverty comes from several sources. The National Research Council (NRC) of the National Academy of Science conducted a review of the US poverty measures in the mid-1990s, finding them deficient in several respects and recommending a number of changes. The overall effect of implementation of the NRC recommendations would be to increase the number and proportion of people identified as living in poverty.

Another relevant set of studies attempts to determine the minimum income levels actually needed by different size and type of families in different geographic locations within the USA for basic economic self-sufficiency. These studies use actual current data on costs of basic goods and services in different states and regions together with existing taxes and tax credits to estimate actual minimum economic self-sufficiency income levels. Generally, the resulting self-sufficiency income estimates fall above 200% of the federal poverty thresholds.

The food-insecurity prevalence estimates summarized in Table 1 for 2000 also provide useful information regarding the relevance of the current poverty thresholds. Data in Table 1 show that, at income levels at or above 185% of poverty, the prevalence of food insecurity and hunger is quite low, with overall food insecurity at 4.9%, food insecurity without hunger at 3.9%, and only 1.1% of households experiencing hunger (numbers do not sum due to rounding). Households with incomes equal to, or greater than, 130% but less than 185% of poverty have higher rates of food insecurity, with 18.4% food insecure overall, 14.5% without hunger, and 3.9% with hunger.

The numbers and proportions of people in poverty by selected household characteristics are shown in Table 2. These data reflect similar patterns to those for food insecurity shown in Table 1. People living in households with children have higher poverty rates than those in households without children, people in households headed by single females have higher poverty rates than those in households headed by married couples, and people in Black and Hispanic households have higher rates of poverty than non-Hispanic Whites.

Table 3 contains state-level average poverty rates and food-insecurity prevalence rates for the years 1996–1998 from the CPS. Calculation of the Pearson correlation coefficient for these two measures indicates a strong positive relationship between poverty

Table 2 Number and percentage of persons in poverty by selected characteristics: 2000

Characteristics	*Total in population/group (1000s)*	*Number in poverty (1000s)*	*Percentage in poverty*
All persons	275 924	31 054	11.3
Children under 18	71 936	11 553	16.1
Ages 18–64 years	171 010	16 142	9.4
Ages 65 years and older	32 978	3 359	10.2
Type of household			
Married-couple families	180 272	10 138	5.6
Related children under 18 years	51 926	4 219	8.1
Related children under 6 years	17 426	1 490	8.6
Female head, no spouse	37 422	10 425	27.9
Related children under 18 years	15 382	6 116	39.8
Related children under 6 years	4 655	2 196	47.2
Women living alone	16 307	3 159	19.4
Men living alone	11 571	1 592	13.8
Race/ethnicity			
White (non-Hispanic)	193 917	14 532	7.5
Black	35 752	7 862	22.0
Hispanic	33 716	7 153	21.2
Ratio of income to poverty			
Under 0.50	275 924	12 099	4.4
Under 1.00	275 924	31 054	11.3
Under 1.30	275 924	45 843	16.6[a]
Under 1.85	275 924	73 329	26.6
Under 2.00	275 924	80 535	29.2
Area of residence			
In metropolitan areas	224 349	24 136	10.8
In central cities	80 144	12 906	16.1
Not in central cities	144 205	11 230	7.8
Outside metropolitan areas	51 575	6 919	13.4

[a]Estimates in this row were obtained by linear interpolation between 1.25 and 1.50.
From Dalaker JD (2001) *Poverty in the United States: 2000. US Census Bureau, Current Population Reports, Series P60-214* Washington, DC: US GPO.

and food insecurity ($R = 0.72$). The resulting coefficient of determination for poverty and food insecurity is also sizeable ($R^2 = 0.51$), indicating that approximately 51% of the variance in food insecurity among states is explained by poverty.

Social Safety-net Programs

Public policies and programs that provide support for low-income individuals and families are increasingly important in reducing and preventing food insecurity, especially in urban areas. One of fifteen food assistance programs funded and overseen by the federal government, the Food Stamp Program (FSP) is the nation's largest nutrition safety-net program for low-income people. Gross income eligibility for receipt of food stamps is 130% of poverty, making most households with incomes below this level, and that fulfill other eligibility requirements, able to receive food assistance if they apply.

However, the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) of 1996, the most recent US welfare reform law, changed several aspects of the FSP, making many legal permanent residents and single adults without children ineligible. In addition, provisions of PRWORA aimed at diverting or deterring applications for cash assistance, and sundry new rules aimed at enforcing compliance with behavioral expectations, appear to have led many households eligible for food stamps not to apply. There is evidence that the magnitude and extensiveness of changes made by PRWORA may have created so much uncertainty about the new welfare system that many current and potential recipients are frequently uninformed or misinformed regarding their eligibility and other critical aspects of program operation.

PRWORA eliminated entitlement status from cash assistance to families with dependent children (AFDC) and replaced it with Temporary Assistance for Needy Families (TANF), while transforming federal funding for AFDC/TANF into a system of block grants to states. New rules and regulations under PRWORA enabled states to place additional restrictions on eligibility and impose a wide range of new behavioral requirements for continued receipt of aid, along with punitive sanctions for failure to comply with these requirements. The new law placed an overall 5-year limit on the receipt of benefits for most recipients and transferred primary responsibilities for design, implementation, and oversight of

Table 3 Average percentage in poverty and food-insecure by state: 1996–98

State	Average percentage in poverty 1996–98	Average percentage food-insecure 1996–98	Difference 1996–98
US	13.2	9.7	3.5
AK	8.8	7.6	1.2
AL	14.7	11.3	3.4
AR	17.2	12.6	4.6
AZ	18.1	12.8	5.3
CA	16.3	11.4	4.9
CO	9.3	8.8	0.5
CT	9.9	8.8	1.1
DC	22.7	11.1	11.6
DE	9.5	6.8	2.7
FL	13.9	11.5	2.4
GA	14.3	9.7	4.6
HI	12.3	10.4	1.9
IA	9.4	7.0	2.4
ID	13.2	10.1	3.1
IL	11.1	8.2	2.9
IN	8.6	7.8	0.8
KS	10.1	9.9	0.2
KY	15.5	8.4	7.1
LA	18.6	12.8	5.8
MA	10.3	6.3	4.0
MD	8.6	7.1	1.5
ME	10.6	8.7	1.9
MI	10.8	8.1	2.7
MN	9.9	6.9	3.0
MO	10.4	8.6	1.8
MS	18.3	14.0	4.3
MT	16.4	10.2	6.2
NC	12.5	8.8	3.7
ND	13.2	4.6	8.6
NE	10.8	7.5	3.3
NH	8.4	7.4	1.0
NJ	9.0	7.3	1.7
NM	22.4	15.1	7.3
NV	9.9	8.6	1.3
NY	16.6	10.0	6.6
OH	11.6	8.5	3.1
OK	14.8	11.9	2.9
OR	12.8	12.6	0.2
PA	11.3	7.1	4.2
RI	11.8	8.7	3.1
SC	13.3	10.2	3.1
SD	13.0	6.4	6.6
TN	14.5	10.9	3.6
TX	16.1	12.9	3.2
UT	8.5	8.8	−0.3
VA	11.3	8.3	3.0
VT	10.6	7.7	2.9
WA	10.0	11.9	−1.9
WI	8.6	7.2	1.4
WV	17.6	9.0	8.6
WY	12.0	9.0	3.0

From Nord M, Jemison K and Bickel G (1999) *Prevalence of Food Insecurity and Hunger, by State 1996–98*. USDA/ERS, Food Assistance and Nutrition Research Report No. 2, September 1999 and Dalaker J (1999) *Poverty in the United States: 1998*. US Census Bureau Current Population Reports, Series P60-207, US GPO, Washington, DC, September 1999.

welfare programs to state and local governments. These changes allowed states wide latitude in imposing even stricter time limits and other eligibility requirements.

A primary goal of welfare reform was to move recipients off the case loads and into jobs. In many states, innovative approaches to delivery of services in support of this transition from welfare to work emerged. As a result of the welfare reform changes and growth in the economy during the 1990s, between 1994 and 1999, the national AFDC/TANF caseload declined by 8.0 million recipients (56.0%). Over the same period, the average national monthly FSP caseload declined by 9.3 million recipients (33.8%).

Several studies have documented extensive declines in participation in the FSP during the 1990s and show that a large part of the decline was due to growth in the economy and falling unemployment raising the incomes of previously eligible households above the income eligibility level. However, these and other studies also show that a majority (56%) of the decline in FSP caseloads resulted from a decline in the participation rates, the proportion of eligible households participating.

Food-security prevalence data from the USDA show that from 1995 to 1999, even though food insecurity declined overall, among households with incomes at or below 130% of the poverty level (the gross income eligibility cutoff for the FSP), food insecurity actually increased from 23 to 28%. This indicates that many low-income households stopped receiving food stamps, or did not apply for them, even though they were food-insecure and felt they needed more food. In addition, studies of the private emergency food-assistance system over the same period indicate that many households leaving the TANF and FSP caseloads are relying more heavily on food from private emergency sources such as food pantries, soup kitchens, and shelters.

Summary and Conclusion

Poverty-related food insecurity and hunger are realities experienced by millions of US households. They are associated with both overnutrition and undernutrition but are not congruent with malnutrition. Food insecurity impacts human development and health throughout the life-cycle, but can be particularly harmful during critical or vulnerable stages early and late in life. Understanding the causes and consequences of food insecurity and knowing how to identify and measure them can improve the quality and effectiveness of social policies, and facilitate prevention and reduction of many kinds of health problems.

Numerous public policies and programs exist to ameliorate and prevent poverty-related food insecurity and hunger. However, the resources to support them ebb and flow with the politics of annual state and federal budgetary cycles. Support and need for these social safety-net programs also vary with business cycles. Unfortunately, need often expands as support shrinks along with employment and government revenues during recessions, and shrinks as support expands along with employment and government revenues during expansions.

In March 2001, the US economy reached the peak of an unprecedented 10-year period of expansion and has been in recession for more than a year. It is worth noting that the record expansionary period from March 1991 to March 2001 began with an unemployment rate at 6.8%, and 8.6 million workers unemployed. During the 10-year expansion, the unemployment rate remained above 6% until September 1994, continued above 5% until April 1997, only dipped below 4% for the two months of September and October 2000, and remained at or above 4% through the peak month of March 2001, when it stood at 4.3% with 6.1 million workers unemployed.

This strongly suggests that the 'full employment' unemployment rate (i.e., the lowest unemployment rate that the US economy can reach and sustain for any appreciable length of time) is probably not lower than 4%, and may be even higher (e.g., 5%). The importance of this is that under known conditions, unemployment is not likely to fall below 4% for very long (if at all), and poverty, food insecurity, and hunger are not likely to disappear altogether, even during economic booms. A clear understanding of the causes and consequences of poverty-related food insecurity and hunger, and of effective policies to reduce and prevent them is critical to the public well-being. Such understanding is more attainable now that valid and reliable food security measures are available and being used.

See also: **Hunger**; **Malnutrition**: The Problem of Malnutrition; Malnutrition in Developed Countries

Further Reading

Anderson SA (ed) (1990) Core indicators of nutritional state for difficult-to-sample populations. Life Sciences Research Office, Federation of American Societies for Experimental Biology. *J Nutr* 120(11S): 1557–1600.

Bickel G, Andrews M and Klein B (1996) Measuring Food Security in the United States: A Supplement to the CPS. In: Hall D and Stavrianos M (eds) *Nutrition and Food Security in the Food Stamp Program*. USDA/FCS/OAE, Alexandria, VA.

Cahalan T, Genser J, Kissmer C, Macaluso TF and Olander C (2001) *The Decline in Food Stamp Participation: A Report to Congress*. USDA/FNS, Report No. FSP-01-WEL.

Dietz WH (1995) Does hunger cause obesity? *Pediatrics* 95(5): 766–767.

Hamilton WL, Cook JT, Thompson WW, Buron LF, Frongillo EA, Olson CM and Wehler CA (1997) *Household Food Security in the United States in 1995: Summary Report of the Food Security Measurement Project*. USDA/FCS/OAE: Alexandria, VA.

Hamilton WL, Cook JT, Thompson WW, Buron LF, Frongillo EA, Olson CM and Wehler CA (1997) *Household Food Security in the United States in 1995: Technical Report of the Food Security Measurement Project*. USDA/FCS/OAE, Alexandria, VA.

Kim M, Ohls J and Cohen R (2001) *Hunger in America 2001: National Report Prepared for America's Second Harvest*. Chicago, IL and Princeton, NJ.

Nord M, Jemison K and Bickel G (1999) *Measuring Food Security in the United States: Prevalence of Food Insecurity and Hunger by State, 1996–1998*. USDA/ERS Food Assistance and Nutrition Research Report No. 2, Washington, DC.

Nord M, Kabbani N, Tiehen L, Andrews M, Bickel G and Carlson S (2002) *Measuring Food Security in the United States: Household Food Security in the United States, 2000*. USDA/ERS Food Assistance and Nutrition Research Report No. 21. Washington, DC.

Olson CM (1999) Nutrition and health outcomes associated with food insecurity and hunger. *J Nutr* 129: 521S–524S.

Radimer K, Olson CM, Greene JC, Campbell CC and Habicht JP (1992) Understanding hunger and developing indicators to assess it in women and children. *J Nut Educ* 24(S): 36S–44S.

Read MA, Frence S and Cunningham K (1994) The role of the gut in regulating food intake in man. *Nutr Rev* 52: 1–10.

Rogers BL, Brown JL and Cook JT (1994) Unifying the poverty line: a critique of maintaining lower poverty standards for the elderly. *Journal of Aging and Social Policy* 6(1/2): 143–166.

Townsend MS, Peerson J, Love B, Achterberg C and Murphy SP (2001) Food insecurity is positively related to overweight in women. *J Nutr* 131: 1738–1745.

Wehler CA, Scott R and Anderson J (1992) The Community Childhood Hunger Identification Project: A Model of Domestic Hunger – Demonstration Project in Seattle, Wahsington. *J Nut Educ* 24(S): 29S–35S.

Foreign Bodies *See* **Adulteration of Foods**: History and Occurrence; Detection; **Contamination of Food**

FREEZE-DRYING

Contents

The Basic Process

J D Mellor and G A Bell, formerly of CSIRO, North Ryde, Australia

Introduction

Freeze drying, which is also known as lyophilization, is the process of removing water from a product by freezing it then subliming the ice to vapor. Sublimation is a physical phenomenon by which solid ice is converted directly into vapor without it passing through the liquid state. Removing water from food, by sublimation, protects the material against loss of important constituents and against chemical reactions that are associated with withdrawing or vaporizing liquid water.

Freeze drying occurs in nature through the combined effects of solar heating, cold dry winds, and rarified atmospheres of mountainous regions. These natural conditions are used to produce freeze-dried 'stock' fish in Norway and a dried potato product called chuno in Peru. Freeze drying is also used in a peculiar way by the North American red squirrel which is known to spread pieces of food in the forks of trees at the beginning of winter, thereby freeze drying its food supply.

Dried foods offer convenience of storage and transport derived from their long shelf-life and low weight. Freeze-dried foods enjoy these properties and are generally of higher quality than products dried by other processes. As a result, freeze-dried foods tend to be preferred in culinary arts over other dried products. (*See* **Drying**: Theory of Air-drying.)

Loss of water vapor from the product depends mainly on two influences: heat available to the product; and the vapor pressure differential across the subliming interface between frozen and dry layers. In practice, these two factors are often compromised in the interest of production economics. Despite widespread use of freeze drying in biological laboratories and in drug and vaccine production, the development of freeze drying as a food technology has been fraught with cost and scale-up problems. Its application to large-scale food processing has been limited to a small number of products including instant coffee, soup mixes, military rations, and herbs.

The final quality of a freeze-dried product can be determined by optimizing conditions in three steps in the process: the primary phase, the drying phase, and the supplementary phase.

Primary Phase

Before freezing, certain preparations are necessary, depending on the product in question. For example, slicing, dicing, and mincing will alter the surface area of the product from where drying can proceed, blanching the product will alter its enzymatic properties, while cooking causes reductions in the free and bound water in the product. These preparations will also tend to enhance heat transfer in the product during the drying phase and, therefore, the speed and quality of the result. (*See* **Freezing**: Operations.)

The water content of a food comprises both free water and water bound to proteins. When a food freezes, a large portion of the free water becomes frozen and is disposable as ice for sublimation. The disposable proportion varies from product to product, and is determined by the final temperature arrived at after freezing. Water remaining in the product after drying may promote microbes and deterioration. Therefore, the temperature to which the

product is frozen can be of critical importance. Table 1 illustrates the influence of temperature on the amount of water frozen out for various foods, determined experimentally. Similar values can be calculated by thermodynamic analysis provided the initial water content and freezing point of the product are known. (*See* **Water Activity**: Principles and Measurement; Effect on Food Stability.)

Drying Phase

Sublimation

In practice, sublimation of ice crystals usually begins at about $-20\,^{\circ}C$ in a vacuum of less than 1.33 mbar, provided sufficient heat is available from the surroundings (benign heat). How long will a product take to dry by sublimation? Ideally, the rate of sublimation is derived from a classical kinetic model of sublimation which assumes: (1) a dynamic equilibrium between the ice surface and its vapor; (2) the rate of escape of water molecules depends on the temperature of the ice; and (3) the tendency of molecules to return to the ice depends on the temperature and pressure of the vapor. This model has limited value in practice. Of greater value in understanding the rate of freeze drying of a food product, is to take account of the rate of vapor diffusion as the water molecules migrate through the vacuum space (containing fewer air molecules than vapor molecules) between the product and the condenser. The migration rate is described by the following equation:

$$G_m = \frac{MD_eP}{RTl} \ln \frac{P - p'_c}{P - p'} \qquad (1)$$

where G_m is the migration rate, M the molecular weight of water vapor, D_e the effective diffusion coefficient, P the vacuum pressure, R the gas constant, T the temperature, l the distance in the vacuum space, p' the vapor pressure outside the product surface, and p'_c the vapor pressure outside the condenser surface. If the freeze drier is considered as a closed system, then the rates for sublimation ($G_s(\alpha p_s - p')$), condensation ($G_c\alpha(p'_c - p_c)$), and migration (G_m) are equal, i.e., $G_s = G_c = G_m$, where p_s is the sublimation vapor pressure and p_c the condensation vapor pressure. Apart from the separate vapor pressure differences ($p_s - p'$ and $p'_c - p_c$), the difference in vapor pressure overall ($p_s - p_c$) is usually regarded as the driving force for sublimation. This model of freeze drying is illustrated in Figure 1.

Vapor Flow

As a piece of food freeze dries, the porous dry product formed around the subliming ice core can cause a severe restriction to vapor flow, and may prevent continuation of the drying process. A useful model of the above conditions assumes the pores in the dry material resemble a bundle of circular cylinders or tubes, with a porosity E expressed as a ratio of pore volume to the intrinsic volume of the layer of dried product, and a tortuosity factor $(L_e/L)^2$ taken as the square of the ratio of the effective flow path L_e to the thickness of the layer L. It can be shown mathematically that:

$$D_e = ED/(L_e/L)^2 \qquad (2)$$

where D_e is the effective diffusion coefficient, and D is the water vapor diffusion coefficient in air. In the

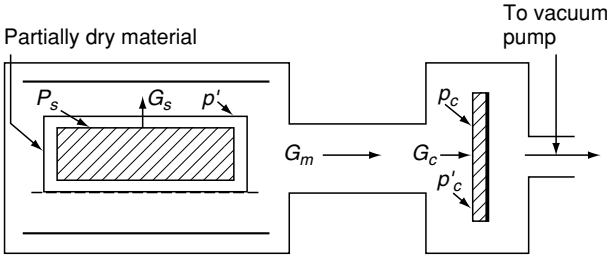

Figure 1 Representation of a common freeze drier. See text for notation of the mathematical variables illustrated. Reproduced from Freeze Drying, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 1 Amount of water frozen in various foods as a function of temperature

Food	*Water content (%)*	*Freezable water (%) frozen at different temperature (°C)*					*Nonfreezable water (kg kg^{-1} solids)*
		−5	*−10*	*−15*	*−20*	*−30*	
Lean beef	74	83	93	97	99	100	0.35
Cod	80.5	85	94	97	98	100	0.39
Egg	74	90	95	98	99	100	0.20
Fruit juice	88	75	87	93	96	100	0.20
White bread	46	50	87	97	99	100	0.30
Peas	78	68	86	92	96	100	0.2–0.3

Data from Kuprianoff J (1964) In: Rey R (ed.) *Advances in Freeze-Drying*, p. 48. Paris.

porous layer, D_e varies inversely with vacuum pressure P and directly with temperature T to the 1.75 power; therefore:

$$D_e = \frac{E}{(L_e/L)^2} D \frac{1013.3}{P} \left(\frac{T}{273}\right)^{1.75} \quad (3)$$

The value D_e increases with decreasing pressure until the mean free path of the gas molecules at intermediate pressures becomes comparable with the pore diameter, so that wall collisions begin to compete with collisions in the gas stream itself in limiting flow by diffusion. This condition is known as transition flow and is the dominating vapor flow in the freeze drying of food.

Heat Flow

In food freeze driers, heating the product is the rate-controlling variable when it is applied to the inner core of ice crystals through an outer layer of dry product. If heat is applied directly through the ice layer by conduction, as is common in laboratory and medical freeze driers, then vapor diffusion is the rate-controlling variable. Food freeze driers usually apply heat by radiation to the product surface which proceeds by conduction through the dry layer. However, the dry product layer insulates the product and limits the radiated heat supply.

Various attempts have been made to compensate for this insulating effect by maximizing heat distribution to the product. These include: (1) the accelerated freeze drying (AFD) technique, illustrated in Figure 2, which employs heat-conducting expanded metal sheets that are hydraulically pressed against the frozen product's surface and serve to conduct heat to the product while also allowing water vapor to escape, and (2) the tempomatic or ribbed-tray method which exploits the conducting properties of a number of ribs that reach into the bulk of the product. Both these methods have seen extensive industrial use. The latter has been used extensively for drying of coffee, berries, and dairy products.

Refrigeration Condensers

It is common to condense the escaping water vapor, produced by sublimation at the product surface, on a refrigerated coil, where condensation is the reverse of sublimation, from vapor to solid without passing through the liquid phase. As ice builds up on a refrigerated coil, it insulates the condenser and renders it less effective. Heat flow through the ice deposit can be modeled to give an indication of the build-up of ice with time, and the resulting temperature deficit (difference between ice surface temperature and coil temperature).

The available surface area of the condenser determines the thickness of ice that will be condensed on to it. The surface area of the condenser is, therefore, an important consideration in designing a food freeze drier.

Condensers have some special characteristics which have been generally neglected in considering the basic physical processes in freeze drying.

It is one of the phenomena of freeze drying that the partial pressure of air near the condenser is always relatively high, this air having been liberated from the product during sublimation, and having come from leaks in the vacuum chamber. The larger the space available to be screened by the condenser, the lower will be the vapor pressure of air in the rest of the

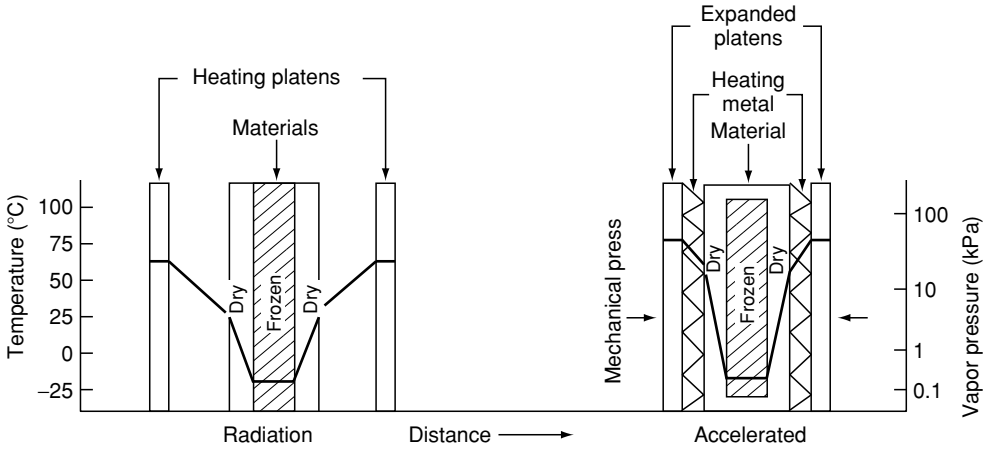

Figure 2 Temperature and vapor pressure gradients (heavy black line) in freeze drying with heating through the dry layers. Reproduced from Freeze Drying, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

vacuum chamber, and hence less air pressure will interfere with the escape of water vapor from the subliming product.

Desorption (Secondary Drying)

After the disappearance of the ice from the freeze-dried product, the residual moisture (nonfreezable water content and, as frequently occurs in practice, incompletely sublimed ice) can be reduced by a secondary drying step called desorption. This consists of heating the product while under a vacuum of between 0.13 and 0.67 mbar. The effect of heating is diminished by the insulating effect of the vacuum in the porous dry product.

An endpoint to secondary drying should be determined to prevent a result in which the product either retains too much water or becomes overdried and damaged by secondary heating. The endpoint can be determined by plotting the ratios of the pressures in the vacuum line before and after isolating the vacuum pump from the chamber by means of a valve. The endpoint is reached when the ratio approaches unity, though values in the range 1.1–1.2 may be practicable. This method assumes the chamber has a low leak rate.

A more exact endpoint can be determined by trapping a sample of the vapor/air mixture from above the desorbing product, then condensing the vapor component outside the vacuum chamber. As in the previous method, the ratio of the partial pressure before and after trapping approaches unity at the endpoint of drying. Precise calibration can be achieved by reference to moisture content predicted by a vapor pressure isotherm for the particular product being dried.

Breaking the Vacuum

A dry inert gas, such as nitrogen, is often used to break the vacuum, and thereby prevent entry of moisture-laden air (and oxygen) into the dry product.

Supplementary (Conditioning) Phase

A freeze-dried product is dry, light, and porous, and ideally retains its original shape. It can be stored for long periods provided its container excludes moisture and oxygen. Residual moisture should be evenly distributed and should not exceed 2% (dry-weight-basis measurement). A conditioning procedure is often necessary to achieve even distribution of residual moisture through the entire mass of the product, since freeze driers with multiple shelves usually have a temperature gradient between the top and bottom shelves and, therefore, do not achieve exactly the same drying conditions on every shelf. Conditioning is achieved by combining the product from all shelves and allowing the mixture to stand for 24 h in an airtight drum before packaging into smaller quantities.

Rehydration

Rehydration restores the lyophile (dry product) to its original physical and organoleptic properties. However, rehydration by immersion of the product in water can only proceed with difficulty in some cases, where pores may have collapsed and protein chains may have changed positions. Protein can be partly digested by adding a proteolytic enzyme to the fluid, and so assist rehydration. The product should be allowed to stand incompletely immersed in the fluid during rehydration, thereby minimizing resistance to rehydration by air becoming entrapped in the porous product.

Technological Improvements

Freeze-drying rates and results for certain food products can be improved by variants of the basic process.

Microwave Heating

This is best applied to remove residual moisture in the desorption stage, since it can speed up this final stage most advantageously. Use of microwave heating in the primary phase is not recommended, because of the danger of ionization of the air under vacuum pressures of less than 1.33 mbar, and possible explosive effects.

Cyclic Pressure

If the vacuum pressure is deliberately increased periodically (two-thirds of the cycle), and then reduced again (one-third of the cycle), a vapor flash occurs at the product surface, as a result of raising the thermal conductivity of the gas, and a small increase in ice interface temperature. The result of this process is a reduction in overall drying time of up to 34% for products such as cooked meat, whole egg, and apple.

Adsorbents

The adsorption process removes water vapor by means of a desiccant rather than refrigeration coils. The desiccant creates a high and well-sustained vapor pressure driving force, particularly at moderately low temperatures, since the equilibrium water vapor pressure of the desiccant decreases as the temperature is lowered.

At $-20\,^{\circ}C$ and 1.04 mbar vacuum, silica gel provides a vapor pressure force to 'drive' freeze drying equivalent to refrigeration coils at $-40\,^{\circ}C$

and similar vacuum pressure. An adsorption system can, therefore, be set at $-20\,^{\circ}C$ and many of the engineering complexities, involving low-temperature refrigeration, high vacuum, and problems of drying the vacuum chamber after defrosting the refrigeration coils, are eliminated.

In an adsorption system, secondary drying can be achieved at $-20\,^{\circ}C$ instead of $+20\,^{\circ}C$, with a greatly reduced risk of overheating the product.

See also: **Drying**: Theory of Air-drying; **Freezing**: Operations; **Water Activity**: Principles and Measurement; Effect on Food Stability

Further Reading

Bell GA and Mellor JD (1986) Development of the adsorption freeze-dryer. *CSIRO Food Research Quarterly* 46: 56–58.

Dalgleish JMcN (1990) *Freeze-Drying for the Food Industries*. London: Elsevier Applied Science.

Goldsmith SA, Rey L and Rothmeyer WW (eds) (1975) *Freeze-Drying and Advanced Food Technology*. London: Academic Press.

King CJ (1970) Freeze-drying of foodstuffs. *CRC Critical Reviews in Food Technology* 1: 379.

Mellor JD (1978) *Fundamentals of Freeze-Drying*. London: Academic Press.

Structural and Flavor (Flavour) Changes

M Marin, Unité Mixte de Recherche de Génie et Microbiologie des Procédés Alimentaires (INRA-INA P-G), Thiverval-Grignon, France

Background

Freeze-drying is a dehydration process especially suited to the conservation of biological products. In comparison with other drying processes, freeze-drying is considered as a reference for manufacturing high-quality dehydrated product. The direct transition of water from solid to vapor (sublimation), without a liquid phase, helps to preserve most of the initial raw material's properties such as appearance, shape, taste, color, and flavor. As an important functional property, the freeze-dried product has a high rehydration capacity. The main limit to the industrial development is its cost due to the low productivity. Consequently, except the application for biological active material (bacteria, vaccine), the use of freeze-drying is restricted in food industry to high added-value products like coffee, ingredients for ready-to-eat foods (fruits and vegetables, meat and fish), and aromatic herbs.

Freezing

Freezing, the first step of the process, determines the structure of the freeze-dried product. The rate of freezing and the temperature level, depending on the nature and the composition of the product, define the number (nucleation) and size (growth) of the ice crystals (**Figure 1**).

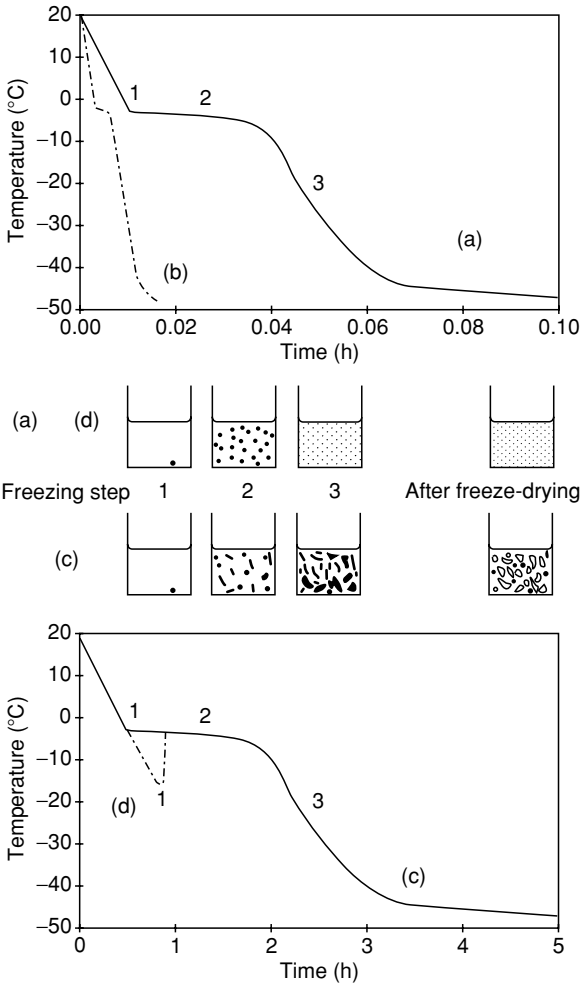

Figure 1 Influence of freezing rate on the final structure of freeze-dried products. (a) Rapid freezing rate; (b) very rapid freezing rate; (c) slow freezing rate; (d) slow freezing with subcooling. The liquid product is shown schematically with black points representing ice. Step 1, appearance of the first ice crystal; step 2, ice nucleation and crystal growth; step 3, end of freezing step. Porosity of the freeze-dried product is shown in relation to freezing rate.

In the case of rapid freezing (more than several centimeters per minute), ice crystals remain small, and the high nucleation rate leads to the creation of numerous crystals. In food products, water crystallizes simultaneously inside and outside the cell, and the microcrystals yield small pores after freeze-drying. This method of freezing preserves the structure of the material as much as possible, but at the same time, the absence of a high porous network tends to decrease the rate of subsequent drying steps and the rehydration capacity of the freeze-dried product.

In the case of slow freezing (less than one centimeter per hour), ice crystal formation is based on two mechanisms. If crystallization begins at a temperature lower than but close to 0 °C (−1 to − 3 °C), the ice crystal growth is faster than the ice nucleation and takes place preferentially outside the cell. Owing to the resulting osmotic effect, water inside the cell tends to move out, and the large crystals obtained in the interstitial areas can cause mechanical injury to the structure. In a few cases, subcooling may occur (several degrees under freezing point without ice), and very rapid crystallization suddenly occurs, yielding a final structure similar to that obtained by quick freezing.

In noncellular systems (gel, juices), the result of slow freezing is a heterogeneous medium with a large dendritic crystal structure and a highly concentrated phase that can form an impermeable layer at the surface of the product. During subsequent drying, if the vapor is thus prevented from leaving the product, the vapor pressure inside the product will increase and make the product foam. To a lesser extent, a nonhomogenous freezing step, based on rapid freezing at the surface (with a refrigerating effect of the superficial water evaporation in a dry cold atmosphere) and a slow freezing inside the product, also creates a less porous layer on the surface. This layer constitutes a barrier to the elimination of the water vapor.

Quenching in liquid nitrogen prevents any crystallization (vitrification) but is not usually carried out in food processing. Meanwhile, if the temperature achieved at the end of the freezing process is sufficiently low, often far below − 20 °C, the frozen product is in an overall amorphous state. This is a good condition to maintain the stability of the frozen structure. Moreover, a lower storage temperature has to be applied in order to preserve the structure obtained after a rapid freezing than after a slow freezing rate.

The freezing step gives the high porosity of the freeze-dried product but explains the main part of the texture loss when damaging the cell wall structure. The volume difference between the fresh and freeze-dried product is generally considered negligible. A rapid (homogeneous) freezing will contribute to preserve the structure and overall texture of the material, while creating the uniform porous network of the freeze-dried product. Slow freezing is nevertheless preferred because the interstices and openings are larger and facilitate mass transfer during drying and rehydration before eating.

Drying

The drying steps (primary – sublimation, and secondary – desorption) may also affect the structure of the product, and the most important factors affecting the criteria of final product quality are the heating temperature and the working pressure.

Temperature

Large differences between the temperature values from the frozen part to the dried layer are observed during freeze-drying.

In the frozen core, it is well known that the temperature must remain under the melting point in order to prevent fusion. Moreover, the structure resulting from the freezing step is stable if the temperature also remains under another lower temperature level, which is characteristic of the glass transition step of the concentrated medium around the ice crystal (Figure 2). As an example, when the vitreous region is obtained (area (3) in Figure 2), the reactions like the recrystallization, also called 'Ostwald maturation,' are highly reduced. Moreover, if

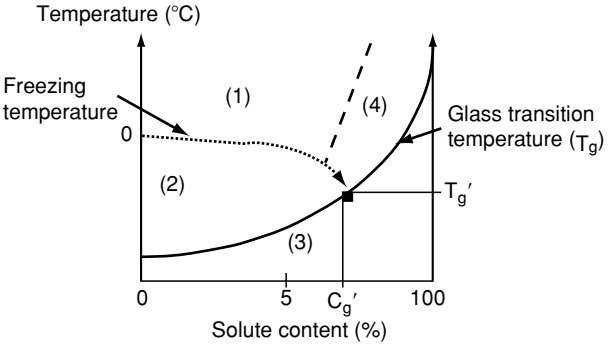

Figure 2 Phase state diagram for a binary mixture. Curves showing transition temperatures – glass transition and melting – are plotted against moisture content. There are four regions: (1) liquid, (2) ice crystal with concentrated liquid, (3) amorphous solid, and (4) liquid with solute crystal. In the amorphous solid region, the viscosity of the matrix is higher than 10^{12}–10^{14} Pa.s. These diagrams may be experimentally determined for a material by differential scanning calorimetry. T_g': Glass transition temperature, of the maximally frozen concentrated solute, C_g'.

the temperature in the frozen core during the primary drying phase is below the melting point but above the glass transition temperature (T_g'), the structural damage may be significant, and the product may collapse when the frozen ice crystal is sublimated. A low temperature (under T_g') maintains a high apparent viscosity; shrinkage and, consequently, collapse are avoided. The product-collapse temperature is closely related to the glass transition temperature but sometimes appears higher than T_g, especially for cellular material (biological tissue). For foods with a high sugar content, the collapse temperature is low (-40 to $-60\,^{\circ}C$) and can be raised by the addition of high-molecular-weight additives (polysaccharides).

The glass transition temperature decreases strongly with the water content, due to the plasticizing effect of water (Figure 2). In the dried layer, a higher temperature, allowing a higher drying rate, can be applied without product damage. But over a critical value of $60\,^{\circ}C$, another kind of damage occurs in the biological product, i.e., protein degradation. This results in a change in the product structure, especially in protein-based foods (meat, fish), and a reduced tenderness and hydration capacity.

During the freeze-drying process, the passage of the product from the frozen state to the dried state is a critical step in which there is a wide variation in water content as well as the temperature. The probability of structural changes (i.e., collapse) is important. As a general rule, to preserve the original structure, the product temperature should remain under the glass transition temperature (T_g), which depends strongly on the water content.

On a practical point of view, the control of temperature is based on the temperature and energy supplied from the heating source. Among the traditional techniques, conductive heating is more preferable for use with thermally sensitive products than a radiation heat source. With microwave heating, it is more difficult to control a homogeneous supply of energy inside the frozen core of the product, and this often leads to a reduction in final product quality.

Pressure

In the atmosphere of the freeze-drier under vacuum, the total pressure is generally assumed to be close to the vapor pressure. Following the equilibrium rule, the vapor pressure is directly related to the temperature at the sublimation front inside the product where the vapor is created. Thus, a control of the total pressure is an indirect way to monitor the product temperature at the sublimation front. At first, the total pressure has to be sufficiently low to avoid the melting of ice. After a rapid abnormal increase in pressure in the freeze-drier, the food product foams and shrinks due to an increase in front temperature above the melting point. Furthermore, during primary drying, if the pressure is sufficiently low to avoid the melting point but higher than the glass transition, shrinkage and collapse mechanisms can be observed. In order to maintain the quality of the product, the total pressure should be sufficiently low in comparison with the sublimation front temperature.

In conclusion, the main modification of the structure of the product is a consequence of the freezing step, assuming that the drying steps are well controlled (no recrystallization and no collapse). During the subsequent drying steps, low values of pressure and temperature (levels depending on the nature of the dry matter of the product) tend to preserve the shape and texture of the frozen product. Knowing that an increase in temperature is a good way to reduce the freeze-drying time, an optimum temperature has to be defined for industrial applications.

Flavor Losses

Freeze-drying is often considered the best drying process as far as flavor compound retention is concerned, due in a first approximation to the low temperature level.

Molecules, responsible for food flavor, usually have a high relative volatility to water and a low molecular weight (less than $300\,g\,mol^{-1}$). Owing to the low concentration of aroma compounds in the food products (p.p.b. to p.p.m.), food flavor is usually evaluated using two kinds of tools: analytical measurements (gas chromatography and mass spectrometry analysis) and/or sensory panel. From these properties, it is difficult to imagine removing water as a vapor phase, without any other volatiles like flavor compounds. Also, flavor losses, even in low proportions, are rapidly considered to be significant. As a matter of fact, the loss of flavor compounds can be limited during freeze-drying with regard to water elimination. The freeze-drying operating conditions that improve the drying rates give an increased retention of volatiles. Rapid primary drying is expected to result in higher levels of volatiles. This can be explained from the principle of selective diffusion. The diffusion coefficient (mobility) of the volatile decreases more rapidly with the water content inside the product than the diffusion coefficient of water itself (Figure 3). The loss of volatile decreases and stops as soon as a critical moisture is attained. One way to decrease the first drying duration is to increase, in a preliminary step, the dry matter (additives, osmotic dehydration). Based on the techniques of encapsulation, additives can also decrease the

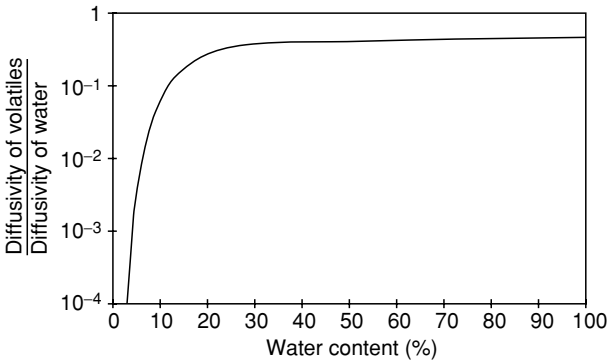

Figure 3 Influence of water content inside food products on the ratio between the diffusion coefficients of volatile compounds and water vapor.

flavor mobility in the matrix during drying. With this approach, a decrease in flavor losses has been observed in coffee, fruits, and vegetables when slow freezing is applied.

Using a constant heating temperature source, retention of flavor decreases when the total pressure increases. Shrinkage and collapse phenomena (high pressure and low apparent viscosity of the medium) are correlated to a poor aroma retention. Considering flavor compounds (or lipids) encapsulated in the amorphous part of the freeze-dried product, recrystallization (either water or solutes) above T_g results in the release of entrapped compounds. Thus, crystallization is related to a loss of flavor. Moreover, exposure of lipids to ambient oxygen is enhanced, which promotes oxidation and creates off-odors. Low-fat meats like chicken, turkey, and ham are easier to freeze-dry. During secondary drying, low heating temperatures are recommended to obtain a freeze-dried product without a cooked flavor, thus preventing new and often undesirable aroma compounds.

Lowering the driving force is another way to reduce the mass transfer of flavor compounds. Reducing the total pressure under vacuum implies that a vacuum pump continually extracts the volatiles from the chamber. Freeze-drying under atmospheric pressure has been proposed to remove less volatiles, but it is necessary to use a highly selective adsorbent that depletes only the water vapor in the freeze-drying atmosphere.

The flavor losses remain low during the freeze-drying process. Under vacuum freeze-drying, the low total pressure prevents oxidation. The storage condition, especially the temperature, can also be detrimental to flavor sensory components.

Packaging and Storage of the Freeze-dried Product

The moisture of the freeze-dried material associated with the temperature during the storage has to be finely controlled in order to preserve the high quality of the freeze-dried product.

The freeze-dried product has a very low water content (less than 5%), and so it is very hygroscopic and necessitates the use of packaging film that is impermeable to water. If there is a gain of a few percent of water, the product can become rubbery at ambient temperature (temperature above the glass transition temperature). Its structure becomes sticky, and the product collapses into lumps after a few weeks. In high-sugar products, there is also a risk of solute crystallization. This loss of structure is associated with an increase in all kinetic reactions inside the product.

Moreover, because of the high porosity of the freeze-dried product, the ambient oxygen is very efficient, and the oxidation reaction can be accelerated (flavor losses, browning).

Depending on the product, a vacuum seems to be better than an inert gas like nitrogen, in order to prevent off-flavors from developing.

A good stability of the freeze-dried product is achieved with a nonpermeable packaging, referring to water vapor and oxygen. Packaging under vacuum and at a low storage temperature (above ambient temperature) is always favorable to increase the time of preservation of the original flavor to several months.

Other Quality Criteria: Nutrients, Color, and Rehydration

Freeze-drying is definitively taken as a reference when quality criteria should be maintained close to the original product. Superior retention of nutrients and a better preservation of the color are also obtained when freeze-drying operating conditions have been chosen to maintain the structure and flavor of the product.

Fruits and vegetables are major sources of vitamins and provitamins, with a reported anticancer activity, which are highly unstable and oxidizable. Freeze-drying at low heating temperatures and under an inert gas gives the best preservation of nutrients compared with other processes (drying, canning). High-activity enzyme extracts from fruits have also been obtained with freeze-drying processes.

In freeze-dried foods, the color is brighter than that of products dried by other techniques. Browning due to Maillard reactions, if compared with other drying

processes, is reduced because of the low temperature used. Nevertheless, a slow freezing rate gives a more pronounced color and higher luminance than a high freezing rate but disappears after rehydration. Referring to the chemical reaction, slow freezing, associated with the cryoconcentration effect, tends to increase the intensity of the red color in freeze-dried products like strawberries and beef meat. This risk of enzymatic browning reactions in fruits is higher with slow freezing. The effect of freeze-drying pressure has been observed after rapid freezing on freeze-dried coffee, with a darker product at high pressure. Secondary drying at more than 60 °C will promote browning (Maillard reactions). Eventually, freeze-dried products are very sensitive to UV light. This is particularly true for vegetables like garlic and carrots, which may become virtually white after a few weeks' storage under light. A layer of aluminum is usually introduced in the packaging film to make it impermeable to UV radiation.

The rehydration capacity is an interesting functional property for industrial use of the product as well as for the consumer. The freezing step, associated with negligible shrinkage, leads to the characteristically high porosity of the freeze-dried product. A low freezing rate promotes a high rehydration capacity. This effect on different freeze-dried food products is not obvious if recrystallization has taken place during freeze-drying. The percentage of sublimed water, which often occupies 80–90% of the volume of biological materials, coincides with the porosity index of the freeze-dried product. The freeze-dried structure explains the instantaneous and high rehydration capacity, which may reach up to 90% of the lost water. Deformation as well as protein denaturation, due to high temperatures during freeze-drying, can decrease this ratio. Nevertheless, it will not achieve the rehydration ratio observed for traditional drying processes, i.e. not higher than 10–20%. During rehydration in boiling water, hot liquid rapidly reenters the pores, and freeze-dried products such as beverages (coffee or tea), vegetables, and all ready-to-eat meal preparations can be instantaneously heated and cooked.

See also: **Drying**: Dielectric and Osmotic Drying; Physical and Structural Changes; Chemical Changes; **Flavor (Flavour) Compounds**: Structures and Characteristics; Production Methods; **Freezing**: Storage of Frozen Foods; Structural and Flavor (Flavour) Changes; Nutritional Value of Frozen Foods; **Oxidation of Food Components**; **Water Activity**: Effect on Food Stability

Further Reading

Dalgleish J (1990) *Freeze-drying for the Food Industries*. London: Elsevier Science.

Goldblith SA, Rey L and Rothmayr WW (eds) (1975) *Freeze-drying and Advanced Food Technology*. London: Academic Press.

Heldman DR and Lund DB (1992) *Handbook of Food Engineering*. New York: Marcel Dekker Inc.

King CJ (1971) *Freeze-drying of Foods*. Cleveland, OH: CRC Press.

Mujumdar AS (1987) *Handbood of Industrial Drying*. New York: Marcel Dekker.

Simatos D, Blond G, Dauvois PH and Sauvageot F (1974) *La Lyophilisation: Principes et Applications*. Paris: Collection de l'ANRT.

FREEZING

Contents

Principles

B R Becker and B A Fricke, University of Missouri-Kansas City, Kansas City, MO, USA

Background

Preservation of food is one of the most significant applications of refrigeration. It is known that cooling and freezing of food effectively reduces the activity of microorganisms and enzymes, thus retarding deterioration. In addition, crystallization of water reduces the amount of liquid water in food items and inhibits microbial growth.

Thermodynamics of the Cooling and Freezing Process

The cooling and freezing of food is a complex process, as illustrated in the cooling/freezing curve shown in Figure 1. Prior to freezing, sensible heat must be removed from the food to decrease its temperature from the initial temperature. As the temperature of the food is decreased, it will pass through its initial freezing point to become undercooled prior to the formation of ice. Then, the formation of ice nuclei releases energy to bring the temperature of the undercooled food back to its initial freezing point. Once ice nuclei have formed, ice crystals will grow as energy is removed.

The number, size, and location of the ice crystals are determined by the rate of energy removal. If the heat transfer rate is much higher than the rate of energy release due to crystal growth, the food will become undercooled, resulting in the formation of additional ice nuclei, and hence, a large number of small ice crystals will form within the food. If the rate of energy release due to crystal growth is more nearly equal to the heat transfer rate, fewer nuclei will form, and the existing crystals will grow to a larger size. In addition, a high freezing rate will not allow sufficient time for water migration from within the cells, thus resulting in intracellular ice crystal nucleation and growth. Conversely, a lower heat transfer rate will provide sufficient time for water migration through the cell walls as extracellular ice crystals remove free water external to the cells, thus resulting in the growth of larger extracellular ice crystals.

Initial Freezing Point

Foods do not freeze completely at a single temperature, but rather they freeze over a range of temperatures, and some foods that are high in sugar content may never be completely frozen. Thus, foods do not have a distinct freezing point but rather an initial freezing point at which the crystallization process begins.

The initial freezing point is important in the determination of proper storage conditions and in the calculation of thermophysical properties. Fresh fruits and vegetables, for example, should be stored above their initial freezing point to avoid frost damage, and since the thermophysical properties of a food change drastically as it freezes, knowledge of its initial freezing point is necessary to accurately model its thermophysical properties.

Freezing Point Depression

The initial freezing point is somewhat lower than the freezing point of pure water because of dissolved substances in the moisture within the food. At the initial freezing point, a portion of the water within the food crystallizes, and the remaining solution becomes more concentrated. Thus, the freezing point of the unfrozen portion of the food is reduced further. As the temperature continues to decrease, the formation of ice crystals increases the concentration

of the solutes in solution and depresses the freezing point further.

Ice Fraction

To predict the thermophysical properties of frozen foods, which depend strongly on the ice and water fractions within the food, it is necessary to determine the mass fraction of water that has crystallized. For temperatures below the initial freezing point, this mass fraction is a function of temperature.

As the food item is cooled to its initial freezing temperature, it will contain ice, unfrozen water, soluble solids, and insoluble solids. As the temperature decreases further, there is an increase in the mass fraction of ice, w_{ice}, and a decrease in the mass fraction of unfrozen water, w_w, which are related as follows:

$$w_{wo} = w_{ice} + w_w, \tag{1}$$

where w_{wo} is the total mass fraction of water within the food.

Assuming that the high moisture content food items can be modeled as ideal dilute solutions, an ice fraction equation can be derived from Raoult's law:

$$w_{ice} = (w_{wo} - w_b)\left[1 - \frac{t_f}{t}\right], \tag{2}$$

where t_f is the initial freezing point of the food (°C), t is the food temperature (°C) and w_b is the mass fraction of water that is bound to solids within the food, and thus is unavailable for freezing. This bound water fraction may be estimated as follows:

$$w_b = 0.4w_p, \tag{3}$$

where w_p is the mass fraction of protein in the food item.

Because eqn (2) underestimates the ice fraction at temperatures near the initial freezing point and overestimates the ice fraction at lower temperatures, the following empirical relationship has been proposed to estimate the mass fraction of ice:

$$w_{ice} = w_{wo}\left[\frac{1.105}{1 + \frac{0.7138}{\ln(t_f - t + 1)}}\right]. \tag{4}$$

Energy Transfer During Freezing

In many food cooling and freezing processes, transient convective heat transfer occurs between a fluid medium and the food item. Knowledge of the corresponding surface heat transfer coefficient is required in order to design the equipment in such processes. Newton's law of cooling defines the surface heat transfer coefficient, h, as follows:

$$q = hA(t_s - t_m), \tag{5}$$

where q is the heat transfer rate, t_s is the surface temperature of the food, t_m is the surrounding fluid temperature, and A is the surface area of the food through which the heat transfer occurs.

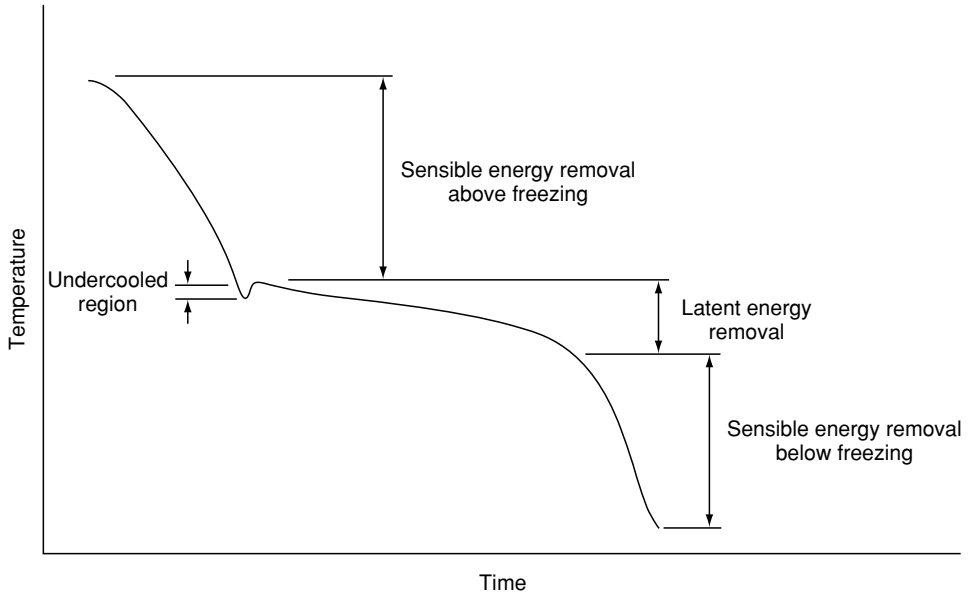

Figure 1 Cooling/freezing curve.

During a convective heat transfer process, energy is theoretically transferred by convection alone. However, in practice, conduction, radiation, and mass transfer may also occur at the same time. In blast cooling/freezing operations, the conductive heat transfer component is usually very small and can be neglected. The significance of the radiation heat transfer and evaporative cooling due to mass transfer must be considered on an individual basis, and their effects upon the total heat transfer must be properly incorporated. Researchers often define an 'effective' heat transfer coefficient, which includes the effects of convection and radiation heat transfer, as well as the energy transfer due to evaporation of moisture from the surface of the food item.

Thermal Properties of Foods

The thermal properties of foods are important in the design of food storage and refrigeration equipment as well as in the estimation of process times for refrigerating, freezing, heating, or drying of foods. Because the thermal properties of foods are strongly dependent upon chemical composition and temperature, the most viable option is to predict these thermal properties using mathematical models that account for the effects of chemical composition and temperature.

Composition data for foods are readily available in the literature. These data consist of the mass fractions of the major food components: water, protein, fat, carbohydrate, fiber, and ash. Food thermal properties can be predicted by using these composition data in conjunction with temperature-dependent mathematical models of the thermal properties of the individual food components.

Equations for predicting the thermal properties of these food components have been developed as functions of temperature in the range of -40 to 150 °C. These equations are presented in Table 1. Because water is the predominant constituent in most food items, the water content of food items significantly influences the thermophysical properties of foods. Therefore, equations for predicting the thermal properties of water and ice have also been developed. These equations are presented in Table 2.

In general, the thermophysical properties of a food item are well behaved when the temperature of the food item is above its initial freezing point. However,

Table 1 Thermal property equations for food components ($-40\,^{\circ}\mathrm{C} \le t \le 150\,^{\circ}\mathrm{C}$)

Thermal property	*Food component*	*Thermal property model*
Thermal conductivity ($\mathrm{W\,m^{-1}\,K^{-1}}$)	Protein	$k = 1.7881 \times 10^{-1} + 1.1958 \times 10^{-3}t - 2.7178 \times 10^{-6}t^2$
	Fat	$k = 1.8071 \times 10^{-1} - 2.7604 \times 10^{-3}t - 1.7749 \times 10^{-7}t^2$
	Carbohydrate	$k = 2.0141 \times 10^{-1} + 1.3874 \times 10^{-3}t - 4.3312 \times 10^{-6}t^2$
	Fiber	$k = 1.8331 \times 10^{-1} + 1.2497 \times 10^{-3}t - 3.1683 \times 10^{-6}t^2$
	Ash	$k = 3.2962 \times 10^{-1} + 1.4011 \times 10^{-3}t - 2.9069 \times 10^{-6}t^2$
Density ($\mathrm{kg\,m^{-3}}$)	Protein	$\rho = 1.3299 \times 10^{3} - 5.1840 \times 10^{-1}t$
	Fat	$\rho = 9.2559 \times 10^{2} - 4.1757 \times 10^{-1}t$
	Carbohydrate	$\rho = 1.5991 \times 10^{3} - 3.1046 \times 10^{-1}t$
	Fiber	$\rho = 1.3115 \times 10^{3} - 3.6589 \times 10^{-1}t$
	Ash	$\rho = 2.4238 \times 10^{3} - 2.8063 \times 10^{-1}t$
Specific heat ($\mathrm{J\,kg^{-1}\,K^{-1}}$)	Protein	$c_p = 2.0082 \times 10^{3} + 1.2089t - 1.3129 \times 10^{-3}t^2$
	Fat	$c_p = 1.9842 \times 10^{3} + 1.4733t - 4.8008 \times 10^{-3}t^2$
	Carbohydrate	$c_p = 1.5488 \times 10^{3} + 1.9625t - 5.9399 \times 10^{-3}t^2$
	Fiber	$c_p = 1.8459 \times 10^{3} + 1.8306t - 4.6509 \times 10^{-3}t^2$
	Ash	$c_p = 1.0926 \times 10^{3} + 1.8896t - 3.6817 \times 10^{-3}t^2$

Table 2 Thermal property equations for water and ice ($-40\,^{\circ}\mathrm{C} \le t \le 150\,^{\circ}\mathrm{C}$)

	Thermal property	*Thermal property model*
Water	Thermal conductivity ($\mathrm{W\,m^{-1}\,K^{-1}}$)	$k_w = 5.7109 \times 10^{-1} + 1.7625 \times 10^{-3}t - 6.7036 \times 10^{-6}t^2$
	Density ($\mathrm{kg\,m^{-3}}$)	$\rho_w = 9.9718 \times 10^{2} + 3.1439 \times 10^{-3}t - 3.7574 \times 10^{-3}\,t^2$
	Specific heat ($\mathrm{J\,kg^{-1}\,K^{-1}}$)[a]	$c_w = 4.0817 \times 10^{3} - 5.3062t + 9.9516 \times 10^{-1}t^2$
	Specific heat ($\mathrm{J\,kg^{-1}\,K^{-1}}$)[b]	$c_w = 4.1762 \times 10^{3} - 9.0864 \times 10^{-2}t + 5.4731 \times 10^{-3}t^2$
Ice	Thermal conductivity ($\mathrm{W\,m^{-1}\,K^{-1}}$)	$k_{ice} = 2.2196 - 6.2489 \times 10^{-3}t + 1.0154 \times 10^{-4}t^2$
	Density ($\mathrm{kg\,m^{-3}}$)	$\rho_{ice} = 9.1689 \times 10^{2} - 1.3071 \times 10^{-1}t$
	Specific heat ($\mathrm{J\,kg^{-1}\,K^{-1}}$)	$c_{ice} = 2.0623 \times 10^{3} + 6.0769t$

[a]For the temperature range of -40 to 0 °C.
[b]For the temperature range of 0 to 150 °C.

below the initial freezing point, the thermophysical properties of a food item vary dramatically with temperature.

Specific Heat Capacity

Specific heat capacity is a measure of the energy transfer required to effect a temperature change and can be used to calculate the heat load imposed on the refrigeration equipment by the freezing of foods. In unfrozen foods, specific heat capacity is relatively constant with respect to temperature. However, for frozen foods, there is a large decrease in specific heat capacity as the temperature decreases.

Unfrozen

The specific heat capacity of an unfrozen food item, c_u, can be obtained from the mass average of the specific heat capacities of its components:

$$c_u = \sum c_i w_i, \tag{6}$$

where the c_i denotes the specific heat capacities of the individual food components, and w_i denotes their mass fractions.

If detailed composition data are not available, a simpler equation for the specific heat capacity of an unfrozen food item can be used:

$$c_u = 4190 - 2300 w_s - 628 w_s^3, \tag{7}$$

where c_u is the specific heat capacity of the unfrozen food item ($J\,kg^{-1}\,K^{-1}$), and w_s is the mass fraction of the solids in the food item.

Frozen

Below the initial freezing point of the food item, the sensible energy due to temperature change and the latent energy due to the fusion of water must be considered. Because latent heat is not released at a constant temperature, but rather over a range of temperatures, an apparent specific heat capacity can be used to account for both the sensible and latent heat effects. To develop expressions for apparent specific heat capacity, first consider the specific enthalpy of a frozen food. The specific enthalpies of the food constituents are additive. Further, assume that the food is composed of solids, liquid water, and ice. Then, the specific enthalpy of the frozen food, h, can be modeled as:

$$h = h_s w_s + h_w w_w + h_{ice} w_{ice}, \tag{8}$$

where h_s is the specific enthalpy of the solid constituents, h_w is the specific enthalpy of water, and h_{ice} is the specific enthalpy of ice. Differentiating eqn (8) with respect to temperature at constant pressure yields an expression for the apparent specific heat capacity, c_a:

$$c_a = \frac{\partial h}{\partial T} = c_s w_s + c_w w_w + c_{ice} w_{ice} + h_w \frac{\partial w_w}{\partial T} + h_{ice} \frac{\partial w_{ice}}{\partial T}, \tag{9}$$

where c_s is the specific heat capacity of the solid constituents, c_w is the specific heat capacity of water, and c_{ice} is the specific heat capacity of ice.

Assuming that high moisture content food items can be modeled as ideal dilute solutions, and noting that $\partial w_w/\partial T = -\partial w_{ice}/\partial T$, the following equation for the apparent specific heat capacity of high moisture content food items can be derived:

$$c_a = c_u + (w_b - w_{wo})\Delta c + E w_s \left[\frac{R T_o^2}{M_w t^2} - 0.8\Delta c\right], \tag{10}$$

where Δc is the difference between the specific heat capacities of water and ice ($\Delta c = c_w - c_{ice}$), E is the ratio of the molar masses of water, M_w, and food solids, M_s, ($E = M_w/M_s$), R is the ideal gas constant, T_o is the freezing point of water ($T_o = 273.2$ K), and t is the food temperature (°C).

A slightly simpler apparent specific heat capacity equation, which is similar in form to that of eqn (10), is given as follows:

$$c_a = 1550 + 1260 w_s + \frac{w_s R T_o^2}{M_s t^2}. \tag{11}$$

Specific Enthalpy

The change in specific enthalpy of a food item can be used to estimate the energy that must be removed from a food item in order to effect a temperature change. Above the freezing point, specific enthalpy consists of sensible energy, whereas below the initial freezing point, specific enthalpy consists of both sensible and latent energy.

Specific enthalpy may be obtained from the definition of the constant pressure specific heat capacity:

$$c_p = \left(\frac{\partial h}{\partial T}\right)_p, \tag{12}$$

where c_p is the constant pressure specific heat capacity, h is the specific enthalpy, and T is temperature. Equations for specific enthalpy may be obtained by integrating expressions of specific heat capacity with respect to temperature.

Unfrozen

The specific enthalpy of an unfrozen food item, h, may be determined by integrating eqn (6) as follows:

$$h = \sum h_i w_i = \sum \int c_i, w_i \mathrm{d}T, \qquad (13)$$

where h_i is the specific enthalpy of the individual food components, and w_i is the mass fraction of the food components.

Integration of eqn (7) from the initial freezing point of a food item, t_f, to a temperature, t, above the freezing point would yield:

$$h = (t - t_f)(4190 - 2300w_s - 628w_s^3). \qquad (14)$$

This equation, however, would predict zero specific enthalpy at the initial freezing point of the food item, t_f. Typically, in the literature for food refrigeration, the reference temperature for zero specific enthalpy is $-40\,^{\circ}\mathrm{C}$. Thus, in order to make eqn (14) consistent with zero specific enthalpy at $-40\,^{\circ}\mathrm{C}$, an additional term must be added to eqn (14), namely, the specific enthalpy at the initial freezing point, h_f:

$$h = h_f + (t - t_f)(4190 - 2300w_s - 628w_s^3). \qquad (15)$$

The specific enthalpy at the initial freezing point, h_f, may be estimated by evaluating eqn (17) at the initial freezing temperature of the food, as discussed below.

Frozen

Integration of eqn (10) between a reference temperature, T_r, and the food temperature, T, leads to the following expression for the specific enthalpy of a frozen food:

$$h = (T - T_r)\left[c_u + (w_b - w_{wo})\Delta c + Ew_s\left(\frac{RT_o^2}{M_w(T_o - T_r)(T_o - T)} - 0.8\Delta c\right)\right]. \qquad (16)$$

Generally, the reference temperature, T_r, is taken to be 233.2 K ($-40\,^{\circ}\mathrm{C}$) at which point the enthalpy is defined to be zero.

By integrating eqn (11) between a reference temperature, T_r, and the food temperature, T, the following expression for specific enthalpy below the initial freezing point may be obtained:

$$h = (t - t_r)\left[1500 + 1260w_s, + \frac{w_s R T_o^2}{M_s t t_r}\right]. \qquad (17)$$

Thermal Conductivity

Thermal conductivity is the material property that relates the conduction heat transfer rate to the temperature gradient. The thermal conductivity of a food item depends upon many factors, including the composition, structure, and temperature of the food item.

Early work in the modeling of food thermal conductivity includes an adaption of Maxwell's equation. This model is based upon the thermal conductivity of dilute dispersions of small spheres in a continuous phase:

$$k = k_c \frac{1 - [1 - a(k_d/k_c)]b}{1 + (a - 1)b}, \qquad (18)$$

where k is the conductivity of the mixture, k_c is the conductivity of the continuous phase, k_d is the conductivity of the dispersed phase, $a = 3k_c/(2k_c + k_d)$, $b = V_d/(V_c + V_d)$, V_d is the volume of the dispersed phase, and V_c is the volume of the continuous phase.

In an effort to account for the different structural features of foods, thermal conductivity models for both homogeneous and fibrous food items have been developed. The differences in thermal conductivity parallel and perpendicular to the food fibers are taken into account in these fibrous food thermal conductivity models.

For an isotropic, homogeneous two-component system composed of continuous and discontinuous phases, in which the thermal conductivity is independent of the direction of heat flow, the following expression may be used to determine thermal conductivity, k:

$$k = k_c\left[\frac{1 - L^2}{1 - L^2(1 - L)}\right], \qquad (19)$$

where k_c is the thermal conductivity of the continuous phase, and L^3 is the volume fraction of the discontinuous phase. In eqn (19), it is assumed that the thermal conductivity of the continuous phase is much larger than the thermal conductivity of the discontinuous phase.

For an anisotropic, fibrous two-component system in which the thermal conductivity is dependent upon the direction of heat flow, two expressions for thermal conductivity have been developed. For heat flow that is parallel to the food fibers, the following expression may be used to determine thermal conductivity, $k_{\|}$:

$$k_{\|} = k_c\left[1 - N^2\left(1 - \frac{k_d}{k_c}\right)\right], \qquad (20)$$

where N^2 is the volume fraction of the discontinuous phase in the fibrous food product. If the heat flow is perpendicular to the food fibers, the following expression for thermal conductivity, $k_{\perp}$, applies:

$$\mathrm{k}_{\perp} = k_c\left[\frac{1 - P}{1 - P(1 - N)}\right], \qquad (21)$$

where $P = N(1 - k_d/k_c)$.

For multicomponent systems, numerous researchers have proposed the use of parallel and perpendicular (or series) thermal conductivity models based upon analogies with electrical resistance. The parallel model is simply the sum of the thermal conductivities of the food constituents multiplied by their volume fractions:

$$k = \sum_{i=1}^{n} \phi_i k_i, \tag{22}$$

where ϕ_i is the volume fraction of constituent i. The volume fraction of constituent i can be found from the following equation:

$$\phi_i = \frac{\frac{w_i}{\rho_i}}{\sum_{j=1}^{n} \frac{w_j}{\rho_j}}. \tag{23}$$

The perpendicular model is the reciprocal of the sum of the volume fractions divided by their thermal conductivities:

$$k = \frac{1}{\sum_{i=1}^{n} \frac{\phi_i}{k_i}}. \tag{24}$$

These two models have been found to predict the upper and lower bounds of the thermal conductivity of most food items.

See also: **Heat Transfer Methods**

Further Readings

Becker BR and Fricke BA (1999) Evaluation of semi-analytical/empirical freezing time estimation methods, Part I: Regularly shaped food items. *International Journal of Heating, Ventilating, Air-Conditioning, and Refrigerating Research* 5: 151–169.

Becker BR and Fricke BA (1999) Evaluation of semi-analytical/empirical freezing time estimation methods, Part II: Irregularly shaped food items. *International Journal of Heating, Ventilating, Air-Conditioning, and Refrigerating Research* 5: 171–187.

Becker BR and Fricke BA (1999) Freezing times of regularly shaped food items. *International Communications in Heat and Mass Transfer* 26: 617–626.

Becker BR and Fricke BA (1999) Food thermophysical property models. *International Communications in Heat and Mass Transfer* 26: 627–636.

Becker BR, Misra A and Fricke BA (1996) Bulk refrigeration of fruits and vegetables, Part I: theoretical considerations of heat and mass transfer. *International Journal of Heating, Ventilating, Air-Conditioning, and Refrigerating Research* 2: 122–134.

Becker BR, Misra A and Fricke BA (1996) Bulk refrigeration of fruits and vegetables, Part II: Computer algorithm for heat loads and moisture loss. *International Journal of Heating, Ventilating, Air-Conditioning, and Refrigerating Research* 2: 215–230.

Chen CS (1985) Thermodynamic analysis of the freezing and thawing of foods: Enthalpy and apparent specific heat. *Journal of Food Science* 50: 1158–1162.

Choi Y and Okos MR (1986) Effects of temperature and composition on the thermal properties of foods. In: LeMaguer M and Jelen P (eds) *Food Engineering and Process Applications*, vol. 1, pp. 93–101. London: Elsevier Applied Science.

Eucken A (1940) Allgemeine Gesetzmassigkeiten fur das Warmeleitvermogen verschiedener Stoffarten und Aggregatzustande. *Forschung auf dem Gebiete des Ingenieurwesens, Ausgabe A* 11: 6.

Franks F (1982) The properties of aqueous solutions at subzero temperatures. In: *Water: A Comprehensive Treatise*, vol. 7, pp. 215–338. New York: Plenum.

Franks F, Asquith MH, Hammond CC, Skaer H leB and Echlin P (1977) Polymeric cryoprotectants in the preservation of biological ultrastruture, I: Low temperature states of aqueous solutions of hydrophilic polymers. *Journal of Microscopy* 110: 223–238.

Fricke BA and Becker BR (2001) Evaluation of thermophysical property models for foods. *International Journal of Heating, Ventilating, Air-Conditioning and Refrigerating Research* 7: 311–330.

Holland B, Welch AA, Unwin ID *et al*. (1991) *McCance and Widdowson's – The Composition of Foods*. Cambridge, UK: Royal Society of Chemistry and Ministry of Agriculture, Fisheries and Food.

Kopelman I (1966) *Transient Heat Transfer and Thermal Properties in Food Systems*. PhD thesis. Michigan State University. East Lansing, MI.

MacKenzie AP (1977) Freezing behaviour of aqueous solutions. *Philosophical Transactions of the Royal Society of London B* 278: 167–189.

Murakami EG and Okos MR (1989) Measurement and prediction of thermal properties of foods. In: Singh RP and Medina AG (eds) *Food Properties and Computer-aided Engineering of Food Processing Systems* pp. 3–48. Dordrecht: Kluwer Academic.

Schwartzberg HG (1976) Effective heat capacities for the freezing and thawing of food. *Journal of Food Science* 41: 152–156.

Slade L and Levine H (1988) Non-equilibrium behavior of small carbohydrate–water systems. *Pure and Applied Chemistry* 60: 1841–1864.

Tchigeov G (1979) *Thermophysical Processes in Food Refrigeration Technology*. Moscow: Food Industry.

USDA (1996) *Nutrient Database for Standard Reference, Release 11*. Washington, DC: US Department of Agriculture.

Operations

B R Becker and B A Fricke, University of Missouri-Kansas City, Kansas City, MO, USA

Introduction

Freezing operations can be grouped by their basic method of extracting heat from foods. Blast freezing utilizes convection heat transfer to remove heat from foods while contact freezing makes use of conduction heat transfer to remove energy from foods. Cryogenic freezing may utilize convection and/or conduction heat transfer to remove heat from foods.

Blast Freezers

Blast freezers use air as the heat transfer medium and depend on contact between the product and the air. Sophistication in air flow control and conveying techniques varies from crude blast freezing chambers to carefully controlled impingement-style freezers. (*See* **Freezing**: Blast and Plate Freezing.)

Blast-Cell Freezer

The blast-cell freezer, shown in Figure 1, is the simplest design and consists of an insulated enclosure equipped with refrigeration coils and fans that circulate the air over the food items in a controlled way. Products are placed on trays, which are then placed into racks in such a way that an air space is left between adjacent layers of trays. The racks are moved in and out of the tunnel manually. Almost all products may be frozen either individually or in cartons in a blast-cell freezer. Because labor requirements are relatively high and product flow is cumbersome, blast-cell freezers are suitable for small quantities of varied products.

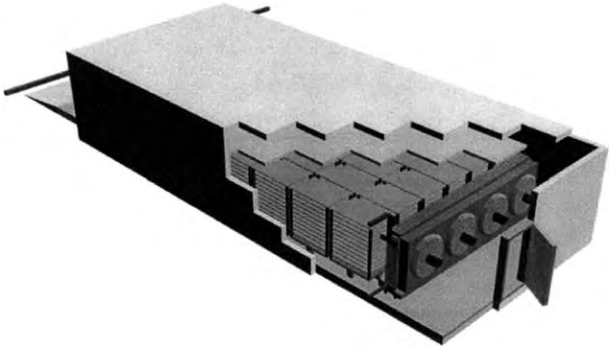

Figure 1 Blast-cell freezer (Ingvar).

Straight-Belt Freezer

The straight-belt freezer, shown in Figure 2, utilizes a wire mesh belt conveyor in a blast room, which provides continuous product flow. Vertical air flow is used, which forces cold air up through the product layer, thereby creating good contact between the air and the product. Straight-belt freezers are generally used with fruits, vegetables, French fried potatoes, cooked meat toppings (e.g., diced chicken), and cooked shrimp.

Straight-belt freezers usually incorporate a two-stage belt composed of two mesh conveyor belts in series. The first stage initially precools or crust-freezes the outer layer of the product before it is transferred to the second stage for final freezing and sensible cooling to −18 °C or below. Freezing times range from 3 to 50 min.

For products with longer freezing times or higher capacity requirements, the freezer footprint can be reduced by stacking belts above each other to form a multipass system, as shown in Figure 3. However, this type of freezing system has the potential for product damage and jams at the belt transfers.

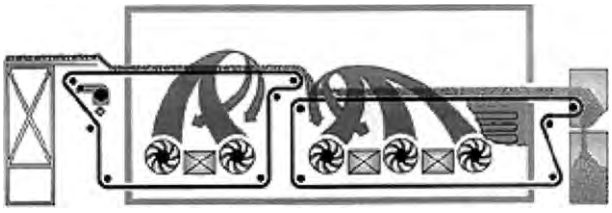

Figure 2 Straight-belt freezer (Advanced Equipment, Inc.).

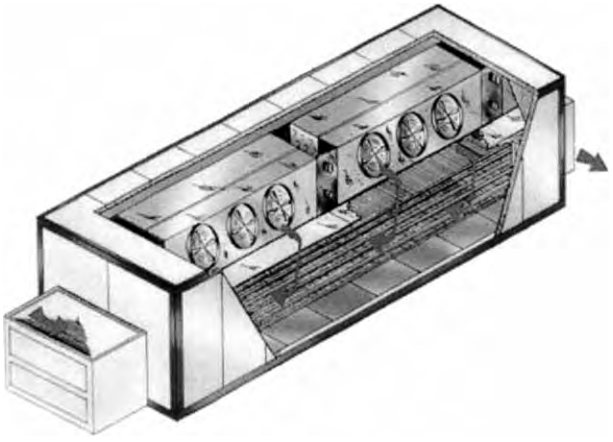

Figure 3 Multipass belt freezer (Advanced Equipment, Inc.).

Spiral-Belt Freezer

The spiral-belt freezer, shown in Figure 4, is often used for products that require gentle handling or a long freezing time (generally 10 min to 3 h). This freezer consists of an endless conveyor belt that is wrapped cylindrically, one tier above or below the last, to form a configuration which requires minimal floor space for a relatively long belt. The number of tiers in the spiral can be varied to accommodate different capacities and two or more spiral towers can be used in series for products with long freezing times. Modular and field-erected spiral freezers are available in a range of belt widths and lengths to accommodate various upstream processes and capacity requirements.

In spiral freezers, horizontal air flow is usually supplied by axial fans mounted along one side of the spiral which blow air horizontally across the spiral conveyor. Some freezers incorporate more sophisticated air flow control utilizing extensive baffling and high-pressure fans.

Typical products frozen in spiral-belt freezers include raw and cooked meat patties, fish fillets, chicken portions, pizza, and a variety of packaged products.

Fluidized-Bed Freezers

The fluidized-bed freezer, shown in Figure 5, utilizes air for product transport and heat transfer. The high degree of fluidization improves the heat transfer rate and is well suited for small, uniform-sized particulate products such as peas, resulting in a freezing time of 3–11 min.

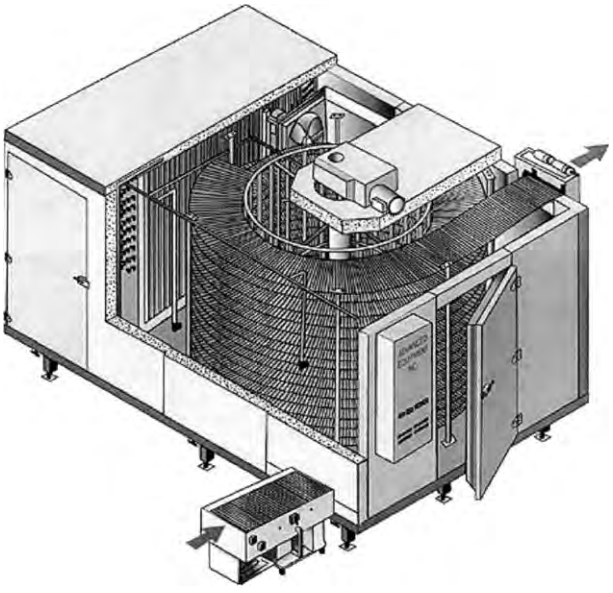

Figure 4 Spiral-belt freezer (Advanced Equipment, Inc.).

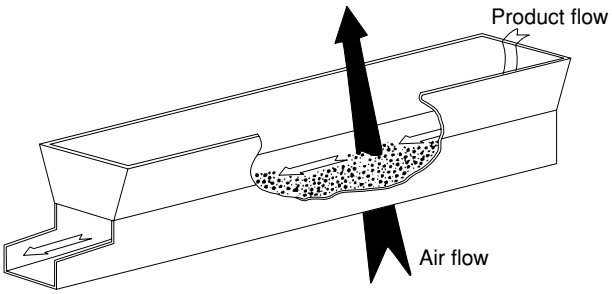

Figure 5 Fluidized-bed freezer (ASHRAE).

Fluidized-Belt Freezers

The fluidized-belt freezer is a hybrid of the belt and fluidized-bed freezers which incorporates a fluidizing section in the first belt stage. It is designed to provide fluidizing conditions for wet incoming product with a belt to assist the transport of heavier products that do not fluidize fully. Once crust-frozen, the product can be loaded deeper for greater efficiency on the second belt.

Contact Freezers

Contact freezers utilize conduction heat transfer whereby the product or package is placed in direct contact with a refrigerated surface. (*See* **Freezing**: Blast and Plate Freezing.)

Plate Freezers

The most common type of contact freezer is the plate freezer, in which the product is pressed between metal plates, as shown in Figure 6. Refrigerant is circulated inside channels in the plates, insuring efficient heat transfer and short freezing times. Plate freezers are especially suited for products that are good conductors of heat, such as fish fillets, chopped spinach, or meat offal. The plate freezer's heat transfer advantage is reduced with increasing product thickness, which is often limited to 5–8 cm.

For packaged products, an additional advantage of plate freezers is that pressure from the plates minimizes any bulging that may occur during freezing, resulting in packages that are even and square. However, packages or cavities should be well filled to insure efficient freezing.

Contact plate freezers are available in horizontal or vertical arrangements with manual loading and unloading. Horizontal plate freezers are also available with automatic loading and unloading, resulting in higher capacities and continuous operation. Automatic plate freezers can accommodate up to 200 packages per minute, with freezing times from 10 to 150 min.

Figure 6 Plate freezer (Ingvar).

Specialized Contact Freezers

A combination of air and contact freezing is often used for wet fish fillets and similar soft, wet products with relatively large, flat surfaces. The product is placed on to a continuous, solid stainless-steel belt and freezing is usually accomplished both by conduction through the belt to a cooling medium below the belt and by convection through controlled air flow above the belt. Another specialized contact freezer conveys food products on a continuous plastic film over a low-temperature refrigerated plate, thus eliminating product deformation or wire mesh belt markings on products which are flat, wet, and sticky, soft, or in need of hand-shaping before freezing. These freezer designs produce attractive product, but a drawback is the physical size of the freezer. Freezing times of less than 30 min can be achieved with these types of freezer.

Cryogenic Freezers

Cryogenic freezers use liquid nitrogen or liquid carbon dioxide as the refrigeration medium, and the freezers may be batch cabinets, straight-belt freezers, spiral conveyors, or liquid immersion freezers. While cryogenic freezers represent a low initial investment, they have a high operating cost. Consequently, cryogenic freezing is often utilized for small-scale production, new products, overload situations, or seasonal products.

The most common type of cryogenic freezer is a straight-through, single-belt tunnel. Liquid nitrogen at $-196\,^{\circ}\mathrm{C}$ or carbon dioxide at $-79\,^{\circ}\mathrm{C}$ is introduced at the outfeed end of the freezer, directly on to the product. As the liquid nitrogen or carbon dioxide vaporizes, the cold vapors are circulated toward the infeed end, where they are used for precooling and initial freezing of the product. The warmed vapors (typically $-45\,^{\circ}\mathrm{C}$) are then discharged to the atmosphere. The low temperature of the liquid and vaporous nitrogen or carbon dioxide provides rapid freezing, which can improve quality and reduce dehydration for some products. However, the freezing cost is high, and the surface of products having a high water content may crack if precautions are not taken. (*See* **Freezing**: Cryogenic Freezing.)

Freezing Time Estimation Methods

In order for food freezing operations to be cost-effective, it is necessary to design the refrigeration equipment optimally to fit the specific requirements of the particular freezing application. This requires estimation of the freezing times of foods, as well as the corresponding refrigeration loads.

Theoretically, the freezing of foods can be described via the Fourier heat conduction equation:

$$\frac{\partial T}{\partial t} = \frac{1}{\rho c}\left[\frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right) + \frac{\partial}{\partial z}\left(k\frac{\partial T}{\partial z}\right)\right] \tag{1}$$

where T is temperature, t is time, ρ is the density of the food, c is the specific heat of the food, k is the thermal conductivity of the food, and x, y, and z are the coordinate directions. For ideal, regularly shaped food items with constant thermophysical properties, uniform initial conditions, constant external conditions, and a prescribed surface temperature or a convection boundary condition, exact analytical solutions for eqn (1) exist which allow for the estimation of the freezing times of foods. However, for realistic freezing processes, the food items are generally irregularly shaped with temperature-dependent thermophysical properties and therefore it is not possible to derive exact analytical solutions for the freezing times of foods.

As a result of this difficulty, most of the research effort has been in the development of semianalytical/

empirical food freezing time prediction methods which make use of various simplifying assumptions. Numerous such methods have been proposed, and the designer is thus faced with the challenge of selecting an appropriate estimation method from the plethora of available methods.

In the following section, the basic freezing time estimation method developed by Plank is discussed first, followed by a discussion of those methods which are based upon modifications of Plank's equation. The discussion then focuses on those freezing time estimation methods in which the freezing time is calculated as the sum of the precooling, phase change, and subcooling times. The last section deals with freezing time estimation methods which account for irregularly shaped food items.

Plank's Equation

The most widely known method for estimating the freezing times of foods was developed by R. Plank in 1913. In this method, it is assumed that only convective heat transfer occurs between the food item and the surrounding cooling medium. In addition, it is assumed that the food item is at its initial freezing temperature and that this temperature is constant throughout the freezing process. Furthermore, a constant thermal conductivity for the frozen region is assumed. Plank's freezing time estimation method is given as follows:

$$t = \frac{L_f}{T_f - T_m}\left[\frac{PD}{h} + \frac{RD^2}{k_s}\right] \tag{2}$$

where L_f is the volumetric latent heat of fusion, T_f is the initial freezing temperature of the food, T_m is the freezing medium temperature, D is the thickness of the slab or the diameter of the sphere or infinite cylinder, h is the convective heat transfer coefficient, k_s is the thermal conductivity of the fully frozen food, and P and R are geometric factors. For the infinite slab, $P = \frac{1}{2}$ and $R = \frac{1}{8}$. For a sphere, P and R are $\frac{1}{6}$ and $\frac{1}{24}$, respectively, and for an infinite cylinder, $P = \frac{1}{4}$ and $R = \frac{1}{16}$.

The geometric factors, P and R, provide insight as to the effect of shape upon freezing time. Plank's shape factors indicate that an infinite slab of thickness D, an infinite cylinder of diameter D and a sphere of diameter D, if exposed to the same conditions, would have freezing times in the ratio of 6:3:2. Hence, a cylinder will freeze in half the time of a slab and a sphere will freeze in one-third the time of a slab.

Various researchers have noted that Plank's method does not accurately predict the freezing times of foods because it assumes that freezing of foods takes place at a constant temperature, and not over a range of temperatures, as is the case in actual food freezing processes. In addition, the thermal conductivity of the frozen food is assumed to be constant, but in reality, the thermal conductivity varies greatly during freezing. Another limitation of Plank's equation is that it neglects the removal of sensible heat above the freezing point. Furthermore, Plank's method is only applicable to infinite slabs, infinite cylinders, and spheres. Researchers have subsequently focused upon development of improved freezing time estimation methods which account for precooling and subcooling times, nonconstant thermal properties, irregular geometries, and phase change over a range of temperatures.

Modifications to Plank's Equation

Modifications to Plank's equation include corrections to account for the removal of sensible heat both above and below the initial freezing point of the food as well as temperature variation during freezing. Regression equations were developed to estimate the geometric parameters, P and R, for infinite slabs, infinite cylinders, and spheres. In these regression equations, the effects of surface heat transfer, precooling, and final subcooling are accounted for by means of the Biot number, Bi, the Plank number, Pk, and the Stefan number, Ste, respectively.

The Biot number, Bi, is defined as follows:

$$Bi = \frac{hD}{k} \tag{3}$$

In the literature for freezing of foods, it is accepted that the characteristic dimension, D, is defined as twice the shortest distance from the thermal center of a food item to its surface. For an infinite slab, D is the thickness. For an infinite cylinder or a sphere, D is the diameter. These definitions will be adopted for this article, unless otherwise noted.

In general, the Plank number, Pk, is defined as follows:

$$Pk = \frac{C_l(T_i - T_f)}{\Delta H} \tag{4}$$

where C_l is the volumetric specific heat of the unfrozen phase and ΔH is the volumetric enthalpy change of the food between T_f and the final food temperature. The Stefan number, Ste, is similarly defined as follows:

$$Ste = \frac{C_s(T_f - T_m)}{\Delta H} \tag{5}$$

where C_s is the volumetric specific heat of the frozen phase.

The freezing times of foods are then calculated with a modified version of Plank's equation. Plank's original geometric factors, P and R, are replaced with the modified values given in Table 1, and the latent heat, L_f, in Plank's equation is replaced with the volumetric enthalpy change of the food, ΔH_{10}, between the freezing temperature, T_f, and the final center temperature, assumed to be $-10\,^{\circ}\mathrm{C}$. As shown in Table 1, the geometric factors P and R are functions of the Plank number, Pk, and the Stefan number, Ste. Both of these parameters should be evaluated using the enthalpy change ΔH_{10}. Thus, the modified Plank equation takes the following form:

$$t = \frac{\Delta H_{10}}{T_f - T_m}\left[\frac{PD}{h} + \frac{RD^2}{k_s}\right] \tag{6}$$

Precooling, Phase Change, and Subcooling Time Calculations

Numerous researchers have taken a different approach to account for the effects of sensible heat removal above and below the initial freezing point. In these methods, the total freezing time, t, is the sum of the precooling, phase change, and subcooling times:

$$t = t_1 + t_2 + t_3 \tag{7}$$

where t_1, t_2, and t_3 are the precooling, phase change, and subcooling times, respectively.

An example of a freezing time estimation method in which sensible heat effects are included by calculating precooling, phase change, and subcooling times separately is given below. The use of a mean freezing point, which is assumed to be 1.5 K below the initial freezing point of the food, accounts for freezing which takes place over a range of temperatures. The freezing time estimation method is stated in terms of the volume and surface area of the food item and is therefore applicable to food items of any shape. This method is given as:

$$t_i = \frac{Q_i}{hA_s\Delta T_{mi}}\left(1 + \frac{Bi_i}{k_i}\right); \quad i = 1,2,3 \tag{8}$$

with the variables defined as shown in Table 2.

Geometric Considerations

Equivalent heat transfer dimensionality A geometric correction factor, called the equivalent heat transfer dimensionality, E, was introduced to calculate the freezing times of irregularly shaped food items. In this method, the freezing time of an irregularly shaped object, t_{shape}, is related to the freezing time of an infinite slab, t_{slab}, via the equivalent heat transfer dimensionality, E, as follows:

$$t_{\text{shape}} = \frac{t_{\text{slab}}}{E} \tag{9}$$

The freezing time of the infinite slab is then calculated from one of the many suitable freezing time estimation methods available for infinite slabs.

Expressions for determining the equivalent heat transfer dimensionality of infinite slabs, infinite and finite cylinders, rectangular bricks, spheres, and two- and three-dimensional irregular shapes are given as follows:

$$E = G_1 + G_2E_1 + G_3E_2 \tag{10}$$

where:

$$E_1 = X(2.32/\beta_1^{1.77})\frac{1}{\beta_1} + [1 - X(2.32/\beta_1^{1.77})]\frac{0.73}{\beta_1^{2.50}} \tag{11}$$

$$E_2 = X(2.32/\beta_2^{1.77})\frac{1}{\beta_2} + [1 - X(2.32/\beta_2^{1.77})]\frac{0.73}{\beta_2^{2.50}} \tag{12}$$

Table 1 Expressions for P and R

Shape	P *and* R *expressions*
Infinite slab	$P = 0.5072 + 0.2018Pk + Ste\left[0.3224Pk + \frac{0.0105}{Bi} + 0.0681\right]$
	$R = 0.1684 + Ste(0.2740Pk - 0.0135)\{$
Infinite cylinder	$P = 0.3751 + 0.0999Pk + Ste\left[0.4008Pk + \frac{0.0710}{Bi} - 0.5865\right]$
	$R = 0.0133 + Ste(0.0415Pk + 0.3957)\{$
Sphere	$P = 0.1084 + 0.0924Pk + Ste\left[0.231Pk - \frac{0.3114}{Bi} + 0.6739\right]$
	$R = 0.0784 + Ste(0.0386Pk - 0.1694)\{$

Table 2 Definition of variables for precooling, phase change, and subcooling time calculations

Process	*Variables*
Precooling	$i = 1$ $k_1 = 6$ $Q_1 = C_l(T_i - T_{fm})V$ $Bi_1 = (Bi_l + Bi_s)/2$ $\Delta T_{m1} = \dfrac{(T_i - T_m) - (T_{fm} - T_m)}{ln\left[\dfrac{T_i - T_m}{T_{fm} - T_m}\right]}$
Phase change	$i = 2$ $k_2 = 4$ $Q_2 = L_f V$ $Bi_2 = Bi_s$ $\Delta T_{m2} = T_{fm} - T_m$
Subcooling	$i = 3$ $k_3 = 6$ $Q_3 = C_s(T_{fm} - T_c)V$ $Bi_3 = Bi_s$ $\Delta T_{m3} = \dfrac{(T_{fm} - T_m) - (T_o - T_m)}{ln\left[\dfrac{T_{fm} - T_m}{T_o - T_m}\right]}$

A_s is the area through which heat is transferred.
Bi_l is the Biot number for the unfrozen phase.
Bi_s is the Biot number for the frozen phase.
C_l is the volumetric specific heat for the unfrozen phase.
C_s is the volumetric specific heat for the frozen phase.
L_f is the volumetric latent heat of fusion.
Q_1, Q_2, and Q_3 are the heats of precooling, phase change, and subcooling, respectively.
ΔT_{m1}, ΔT_{m2}, and ΔT_{m3} are the corresponding log-mean temperature driving forces.
T_c is the final thermal center temperature.
T_i is the initial temperature.
T_{fm} is the mean freezing point, assumed to be 1.5 K below the initial freezing point.
T_m is the freezing medium temperature.
T_o is the mean final temperature.
V is the volume of the food item.

$$X(x) = x/(Bi^{1.34} + x) \tag{13}$$

and the geometric constants, G_1, G_2 and G_3, are given in Table 3. The dimensional ratios, β_1 and β_2, are defined as follows:

$$\beta_1 = \frac{\text{second shortest dimension of the food item}}{\text{shortest dimension of the food item}} \tag{14}$$

$$\beta_2 = \frac{\text{longest dimension of the food item}}{\text{shortest dimension of the food item}} \tag{15}$$

Mean conducting path Knowledge of the Biot number of a food item is required to utilize freezing time estimation methods. To calculate the Biot number of a food item, its characteristic dimension must be known. Because it is difficult to determine the characteristic dimension of an irregularly shaped food item, the concept of the mean conducting path was introduced. The mean conducting path, $D_m/2$, is the mean heat transfer length from the surface of the food item to its thermal center. This mean conducting path is used to calculate the Biot number which, in turn, is used in the calculation of the freezing time. Thus, the Biot number becomes:

$$Bi = \frac{hD_m}{k} \tag{16}$$

where D_m is twice the mean conducting path.

For rectangular blocks of food, the mean conducting path is proportional to the geometric mean of the block's two shorter dimensions. Based upon this result, an equation to calculate the Biot number, Bi, for rectangular blocks of food is given as follows:

$$\frac{Bi}{Bi_o} = 1 + \left\{\left[1.5\sqrt{\beta_1} - 1\right]^{-4} + \left[\left(\frac{1}{\beta_1} + \frac{1}{\beta_2}\right)\left(1 + \frac{4}{Bi_o}\right)\right]^{-4}\right\}^{-0.25} \tag{17}$$

where Bi_o is the Biot number based on the shortest dimension of the block, D_1: $Bi_o = hD_1/k$. The Biot number, Bi, calculated with eqn (17) is then substituted into a freezing time estimation method, such as eqn (8), to calculate the freezing time for rectangular blocks.

It has been noted that for squat-shaped food items, the mean conducting path, $D_m/2$, could be reasonably estimated as the arithmetic mean of the longest and shortest distances from the surface of the food item to its thermal center.

Table 3 Geometric constants for equivalent heat transfer dimensionality

Shape	G_1	G_2	G_3
Infinite slab	1	0	0
Infinite cylinder	2	0	0
Sphere	3	0	0
Squat cylinder	1	2	0
Short cylinder	2	0	1
Infinite rod	1	1	0
Rectangular brick	1	1	1
Two-dimensional irregular shape	1	1	0
Three-dimensional irregular shape	1	1	1

Equivalent sphere diameter The concept of the equivalent sphere diameter can also be used to calculate the freezing time of irregularly shaped food items. In this method, a sphere diameter is calculated which is based upon the volume and the volume-to-surface-area ratio of the irregularly shaped food item. This

equivalent sphere is then used to calculate the freezing time of the food item.

In the equivalent sphere diameter technique, the volume–surface diameter, D_{vs}, is defined as the diameter of a sphere having the same volume-to-surface-area ratio as the irregular shape:

$$D_{vs} = \frac{6V}{A_s} \tag{18}$$

where V is the volume of the irregular shape and A_s is the surface area of the irregular shape. In addition, the volume diameter, D_v, is defined as the diameter of a sphere having the same volume as the irregular shape:

$$D_v = \left(\frac{6V}{\pi}\right)^{1/3} \tag{19}$$

Then, the equivalent sphere diameter, $D_{\mathrm{eq,s}}$, is defined as follows:

$$D_{\mathrm{eq,s}} = \frac{1}{\beta_2 + 1} D_v + \frac{\beta_2}{\beta_2 + 1} D_{vs} \tag{20}$$

Thus, the prediction of the freezing time of the irregularly shaped food item is reduced to predicting the freezing time of a spherical food item with diameter $D_{\mathrm{eq,s}}$.

See also: **Freezing**: Blast and Plate Freezing

Further Reading

Becker BR and Fricke BA (1999a) Evaluation of semi-analytical/empirical freezing time estimation methods, part I: regularly shaped food items. *International Journal of Heating, Ventilating, Air-Conditioning, and Refrigerating Research* 5: 151–169.

Becker BR and Fricke BA (1999b) Evaluation of semi-analytical/empirical freezing time estimation methods, part II: irregularly shaped food items. *International Journal of Heating, Ventilating, Air-Conditioning, and Refrigerating Research* 5: 171–187.

Becker BR and Fricke BA (1999c) Freezing times of regularly shaped food items. *International Communications in Heat and Mass Transfer* 26: 617–626.

Becker BR and Fricke BA (1999d) Food thermophysical property models. *International Communications in Heat and Mass Transfer* 26: 627–636.

Cleland AC and Earle RL (1977) A comparison of analytical and numerical methods of predicting the freezing times of foods. *Journal of Food Science* 42: 1390–1395.

Cleland AC and Earle RL (1979) A comparison of methods for predicting the freezing times of cylindrical and spherical foodstuffs. *Journal of Food Science* 44: 958–963, 970.

Cleland AC and Earle RL (1982) Freezing time prediction for foods – a simplified procedure. *International Journal of Refrigeration* 5: 134–140.

Cleland DJ, Cleland AC and Earle RL (1987a) Prediction of freezing and thawing times for multi-dimensional shapes by simple formulae – part 1: regular shapes. *International Journal of Refrigeration* 10: 156–164.

Cleland DJ, Cleland AC and Earle RL (1987b) Prediction of freezing and thawing times for multi-dimensional shapes by simple formulae – part 2: irregular shapes. *International Journal of Refrigeration* 10: 234–240.

Fricke BA and Becker BR (2001) Evaluation of thermophysical property models for foods. *International Journal of Heating, Ventilating, Air-Conditioning and Refrigerating Research* 7: 311–330.

Ilicali C and Engez ST (1990) A simplified approach for predicting the freezing or thawing times of foods having brick or finite cylinder shape. In: *Engineering and Food*, Speiss WEL and Schubert H (eds) vol. 2, pp. 442–451. London: Elsevier Applied Science.

Ilicali C and Hocalar M (1990) A simplified approach for predicting the freezing times of foodstuffs of anomalous shape. In: Speiss WEL and Schubert H (eds) *Engineering and Food*, vol. 2, pp. 418–425. London: Elsevier Applied Science.

Pham QT (1984) An extension to Plank's equation for predicting freezing times for foodstuffs of simple shapes. *International Journal of Refrigeration* 7: 377–383.

Pham QT (1985) Analytical method for predicting freezing times of rectangular blocks of foodstuffs. *International Journal of Refrigeration* 8: 43–47.

Plank R (1913) Die Gefrierdauer von Eisblocken. *Zeitschrift für die gesamte Kalte-Industrie* 20: 109–114.

Plank R (1941) Beitrage zur Berechnung und Bewertung der Gefriergesch Windikeit von Lebensmitteln. *Beiheft zur Zeitschrift für die gesamte Kalte-Industrie* 10: 1–16.

Blast and Plate Freezing

M F G Boast, MBA Consulting Engineers, Markyate, UK

Background

The provision of low temperature for freezing is obtained from closed-circuit compression–expansion refrigeration systems, or by a cryogenic process of exposing the product to a cold atmosphere of liquid nitrogen or liquid carbon dioxide.

Various refrigerants have been used for the compression–expansion system, which forms the major source of food freezing, probably 90% of all installations.

Ammonia and sulfur dioxide are excellent media for the refrigeration process, but they carry the hazard of being inflammable and the risk of explosion. This led to the development during the 1930s of

the first chlorofluorocarbon (CFC) refrigerants, R11 and R12. These refrigerants replaced SO_2 in domestic refrigeration and NH_3 for small freezing systems. The later development of R22 (a hydrochlorofluorocarbon, or HCFC) and R502 (an azeotrope), displaced NH_3 from all but the largest of industrial freezing systems. Table 1 shows the characteristics of common refrigerants.

During the 1980s, the environmental effect of CFCs and HCFCs, in terms of ozone depletion potential and contribution to the greenhouse effect, became apparent. This has led to the restriction in the use of some refrigerants under the Montreal Protocol, and the development of new refrigerants called hydrofluorocarbons (HFCs).

Refrigerants and the Environment

Ozone Depletion

The ozone-depleting effect of refrigerants is important, because the ozone layer filters out ultraviolet (UV) radiation, which might otherwise reach the surface of the earth. This radiation could have seriously damaging effects on germinating crops and on the photoplankton of the oceans. In addition, increased doses of UV light can cause malignant skin cancer in fair-skinned races.

Greenhouse Effect

Ecological systems of the world are finely dependent on temperature. Small changes in average temperature can affect the distribution of crops and animals. There is evidence that the average global temperature has been increasing since the mid-nineteenth century because of carbon dioxide from the burning of fossil fuels. If it is not halted, mean sea levels will rise.

The consequences of even a small rise in sea level in certain areas of the world, where strong seasonal winds might coincide with high tides, would be disastrous. It is impossible to predict the likely rise of mean sea level, but it could be significant. In the remote past, the earth had no polar ice caps, presumably because of high concentrations of carbon dioxide in the atmosphere, which had not by then been absorbed by forests and oceans.

Chlorofluorocarbons are greenhouse chemicals more potent than the present greenhouse chemicals (carbon dioxide and methane) and should therefore not be released into the atmosphere.

Recommendations

Refrigerants known as CFCs are typically R11, R12, R115, R13B1, and R502 and have now been phased out, although some installations may still exist using the above refrigerants. New systems are not being constructed from the refrigerants.

The destruction of CFCs present problems in terms of safety and the need for environmental protection. Therefore, any system being decommissioned containing CFCs must be addressed in a controlled manner by personnel with the appropriate experience and qualifications.

Table 1 Characteristics of common refrigerants

Substance	*Type*	*Formula*	*Montreal Protocol*	*Ozone depletion potential (R11 = 1)*	*Greenhouse potential (CO_2 = 1)*	*Flammability*
R11	CFC	CCl_3F	Yes	1	3 300	No
R12	CFC	CCl_2F_2	Yes	2	10 000	No
R22	HCFC	$CHClF_2$	No	0.05	1 100	No
R113	CFC	CCl_2FCClF_2	Yes	0.8	4 500	No
R114	CFC	$CClF_2CClF_2$	Yes	1.0	1 300	No
R115	CFC	$CClF_2CF_3$	Yes	0.6	25 000	No
R123	HCFC	$CHCl_2CF_3$	No	0.02	50	No
R124	HCFC	$CHClFCF_3$	No	0.02	300	No
R125	HFC	CHF_2CF_3	No	0	1 900	No
R134a	HFC	CF_3CH_2F	No	0	900	No
R141b	HCFC	CH_3CCl_2F	No	0.08	300	Slight
R142b	HCFC	CH_3CClF_2	No	0.06	1 200	Slight
R152a	HFC	CH_3CHF_2	No	0	100	Moderate
R500	Azeotrope	R12/152a	Yes	0.74	7 400	No
R502	Azeotrope	R22/115	Yes	0.33	13 300	No
R717	Ammonia	NH_3	No	0	0	Moderate
R407C	Azeotrope	R32/125/134a	No	0	1 610	No
R404A	Azeotrope	R125/143a/134a	No	0		No
R410A	Azeotrope	R32/125	No	0	1 890	No

The construction of new systems for small to medium-sized uses refrigerant R404A. Large industrial systems use refrigerant R717 (ammonia).

Refrigerant R717 should only be used by companies and specialists with knowledge of this refrigerant owing to its pungent smell and hazard potential.

Maintenance

Plants should be constructed to the appropriate codes of practice and standards. Routine leak testing must be conscientiously carried out. In addition to being environmentally correct, this will produce cost savings and improve plant reliability.

Leak detectors should be fitted to all systems with a large refrigerant charge and should report to a central monitoring station.

Substitutes for CFCs

The substitute for refrigerant 12 is refrigerant R134A. The substitute for refrigerant R502 is refrigerant R404A. These refrigerants are not suitable for use with mineral oils, and polyester lubricants have been developed. Changing the system from refrigerant R12 or refrigerant R502 requires specialist knowledge and should only be carried out by suitably experienced companies and personnel. Changing the system refrigerant from R12 or R502 requires system flushing and thoroughly cleaning to insure that all traces of the previous refrigerant have been removed.

At the present time, systems using refrigerant R22 can continue to operate and can be serviced with new virgin refrigerant until the year 2005.

Freezing Considerations

Freezing Times

Freezing times vary, and each product has an associated freezing time that will depend on its composition, dimensions, and packaging. Dimensions and packaging have a direct effect on the freezing time, as heat can only leave the product through its surface. Products that are unpacked and have a large surface area-to-weight ratio, such as peas, will freeze in a minimum time, whereas a dense block of meat products, with heavy packaging or boxing, may take days to be fully frozen to the core.

The influence of freezing time is more apparent on some products than others. For strawberries, the drip loss is reduced from 20% to nearly 0% with shorter freezing times. A difference in drip loss from 20% for strawberries frozen in 12 h, to 8% for strawberries frozen in 15 min, is significant. Cryogenic systems perform the same freezing function in 8 min or less, reducing the drip loss to less than 5%.

A modern mechanical freezer or a cryogenic freezer can crust products very rapidly, minimizing the loss of natural juices and locking in the qualities that make the product more marketable.

Packaging

Packaging must be of good quality, odor- and taint-free, and it must totally enclose the product. The container should not be capable of being opened without recognizable damage. It should also protect the product against normal transit and storage hazards and inhibit dehydration by incorporating a moisture vapor barrier. All packaging should carry clear identification and should be coded so that stock rotation can be carried out. The inner carton should be marked with an appropriate identification code that would enable the producer concerned to establish the date of production and the location of the producing factory, and gain easy access to daily production records.

As the product is wrapped, the operator must insure that the inner layer is in good contact with the actual food and that all air has been excluded. Unless this is established, reasonable cooling rates will not be obtained. The packages should then be placed and arranged to insure good contact with the freezing surface in the case of a plate freezer, or be in the air flow of an air blast freezer.

Cooked food must be chilled to around 10 °C or lower before it enters the freezer. The placing of a hot product in the freezer will considerably slow the freezing process as the outer surface will quickly freeze, leaving the hot interior to give up initially larger amounts of sensible heat than would have been necessary had the product been allowed to stabilize and cool prior to freezing.

Food cooked in a domestic kitchen should be first allowed to cool close to ambient temperature and then transferred to the refrigerator to chill down before being placed in the freezer. The use of an immersion thermometer to check core temperatures prior to freezing is recommended. (*See* **Packaging**: Packaging of Liquids; Packaging of Solids.)

Weight Loss and Economics

Freezing equipment will represent the single largest investment in a food production line, yet its operating costs will normally only be 3–5% of the total product cost. Packaging costs vary widely but are normally several times greater than the total cost of freezing.

An important factor to consider when selecting freezing equipment is the weight loss of the product

that occurs during freezing. This loss may be about the same as the operating costs of the freezing facility. This applies to inexpensive products like peas and is even more significant for expensive products, such as seafoods and soft fruits.

Weight loss during freezing may be caused by mechanical losses, downgrading, and dehydration. Mechanical loss refers to products dropping to the floor, sticking to conveyor belts or dripping juice. A modern freezer should have almost no loss in these categories. Downgrading losses refer to damaged products, breakages, and similar occurrences that render the product unsaleable at the top quality price. Dehydration losses will always be present in any freezing system. The evaporation of water vapor from unpacked products during freezing can be seen as frost on the refrigeration plant freezing surfaces. This frost is also caused by excessive infiltration of warm, moist air into the freezing chamber. Still air inside the vapor-proof product carton can create larger dehydration losses than an unpacked product frozen in a cryogenic quick freezer. The heat transfer is poor because no circulation of air occurs within the package. The result is evaporation of moisture, which can be significant, with the frost staying inside the carton.

A poorly designed freezer or freezing tunnel can have dehydration losses as high as 5%, as against a well-designed freezing system in which losses are maintained below 1.5%. Cryogenic liquid nitrogen tunnels normally operate with a dehydration loss of between 0.25 and 1.25%. These losses usually occur when the nitrogen gas circulates over the product at the in-feed end of the freezer. It may be necessary at the in-feed end to temper the product so as to use the heat capacity of the nitrogen more efficiently. Nitrogen immersion freezing tends to have even lower dehydration losses but uses more liquid nitrogen. Carbon dioxide freezing operations use jet impingement, and dehydration losses are similar to that experienced with liquid nitrogen.

Freezing Methods

Freezing equipment can be divided into groups with regard to the basic method of extracting heat from the product to be frozen:

1. *Air blast freezing*: Air at high velocity (between 4 and 7 m s^{-1}) is circulated over the product. The air picks up heat, which is then recooled by means of an air-to-refrigerant heat exchanger before the air is recirculated.
2. *Contact freezing*: Food (packed or unpacked) is placed between metal plates or surfaces, and the heat is extracted by direct conduction through the metal surface, which is refrigerated by the circulation of a cooling medium.
3. *Immersion freezing*: Food is immersed in a low-temperature, nonfreezing solution that is cooled by evaporators in a conventional refrigeration system.
4. *Cryogenic freezing*: Food is exposed to an atmosphere below $-60\,^{\circ}C$, which is achieved by spraying liquid nitrogen (LN) or liquid carbon dioxide into the freezing chamber.

All of the above methods are used in food freezing processes. However, the more favored systems are those that can be operated in line with the proceeding processing and preparation operations and the subsequent packing functions.

Types of Freezing Equipment

Storage Rooms

The storage room must not be considered as freezing equipment, although it may sometimes be used for this purpose. Freezing in a storage room has many disadvantages and should only be used in exceptional cases. The freezing process is slow, and the quality of almost all products suffers. If products are already being stored in the room, their quality is jeopardized, because flavors may be transferred from the warm products yet to be frozen.

Because the storage room is not designed for freezing, the cooling coils may frost up quickly, and total refrigeration capacity may be reduced. The temperature of the products already frozen may rise considerably, affecting their quality. Without doubt, freezing in a storage room leads to a lower product quality.

Blast Freezer

A blast freezing room is usually equipped with more forced-air cooling units or larger coolers than would normally be found. These coolers are often of sufficient size to warrant floor mounting. The air coolers are equipped with fans that create considerable air turbulence. Products may be laid in trays that are then loaded into a freezing trolley or rack; this is moved in and out of the blast freezer by hand or with a forklift truck.

The racks must be designed to provide air space between the trays, which is equal to approximately 50% of the product thickness. Because there is no control over air circulation, the resulting heat transfer of the product to air may be less effective. Blast freezer rooms offer acceptable conditions for a limited range of products with the freezing cycle and

the quality of freezing dependent on the experience of the operator.

Stationary Freezing Tunnels

A stationary freezing tunnel is the simplest type of freezer (Figure 1), designed to produce satisfactory results for the majority of products. It consists of an insulated enclosure, equipped with refrigeration coils and fans that circulate air in a controlled pattern over the product. The air circulation system design influences the freezing rate and the product weight loss.

Products are placed on trays that are then placed in a rack. The racks are arranged to provide an air space between each level of trays and are moved in and out of the tunnel manually. The human element becomes important when positioning the racks inside the tunnel to prevent air bypass; this is essential for efficient freezing processes.

Most products can be frozen in a stationary freezing tunnel. Whole, sliced, or diced vegetables may be frozen either in cartons or unpacked, provided that the layers are limited to 30–40 mm in depth on the trays. Spinach, broccoli, meat patés, fish fillets, and prepared foods are often frozen in packages, using this type of equipment. By using different rack designs, thick packages and whole meat carcasses can also be frozen. The system capacity depends on the product thickness and composition, as well as on the existence of packaging.

This type of freezer is reasonably flexible in use, making it suitable for the initial development stage of the new frozen food product, but does require heavy outlay for manpower and may result in considerable weight loss, if not properly used.

Push-through Tunnels

Some simple mechanization is achieved in a push-through tunnel (Figure 2). Tracks are provided in the floor, and the racks are fitted with casters or wheels. The mechanization system is often hydraulically powered with a special stroking facility that feeds the racks through the freezer. The time of the hydraulic stroking operation can be adjusted to vary the time the product is in the freezer.

This type of equipment may require open ends to allow product entry and exit. If doors are provided, they have to be interlocked with the hydraulic track system, so that they open and close, permitting a rack to be drawn into the tunnel and also ejected at the other end. This type of freezer has a number of advantages over a stationary tunnel: there is improved

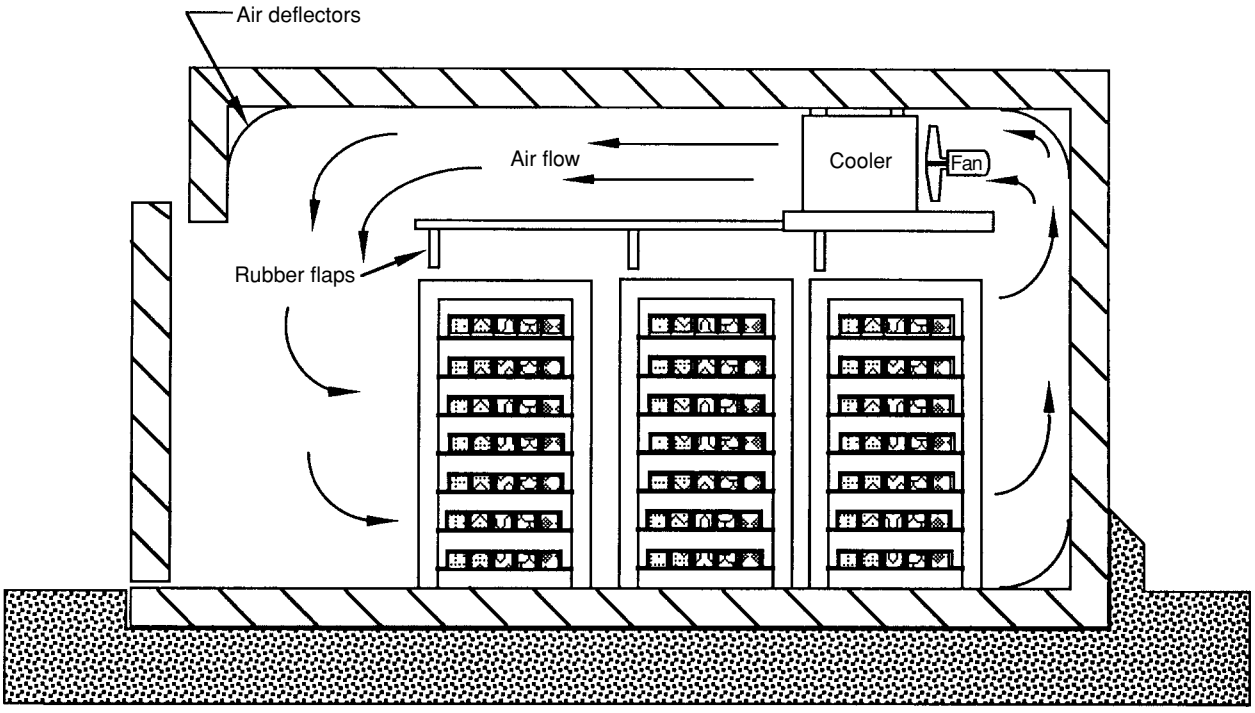

Figure 1 Stationary freezing tunnel (Michael Boast Associates). Reproduced from Blast and Plate Freezing, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

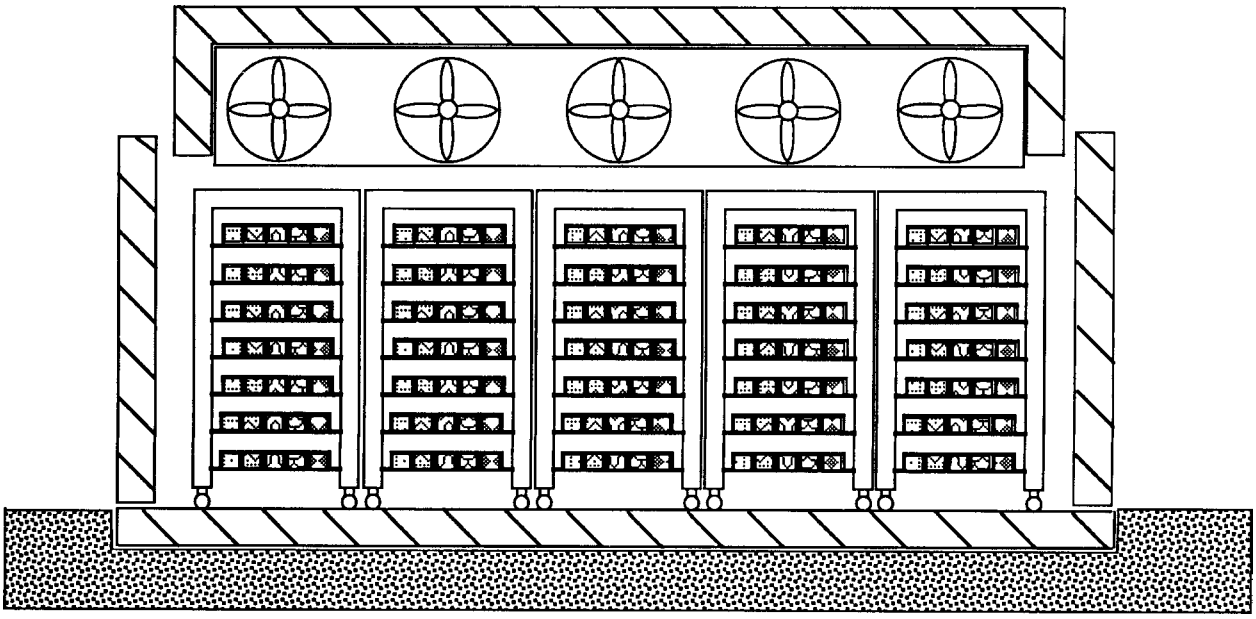

Figure 2 Push-through tunnel (Michael Boast Associates). Reproduced from Blast and Plate Freezing, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

air circulation over the product as it moves at a steady rate through the tunnel; labor costs are considerably decreased; and there is added flexibility with the facility of varying the freezing time.

Some tunnels may have three or four lines, each with independent hydraulic stroking facilities, and this makes different product freezing times available in a single installation.

Belt Freezers

Modern belt freezers use vertical air-flow patterns and require uniform distribution of product over the entire belt surface for effective freezing.

The single-belt freezer, multiple-belt freezer, and spiral-belt freezer are the main types available. The simplest is the single-belt freezer, consisting of the single belt exposed to an updraught of air. This is suitable for deep-fried or relatively dry products that do not tend to freeze to each other or form clumps, i.e., fish sticks, French fried potatoes, and bakery products.

Multiple-belt freezers are suitable for individual freezing of fried fish sticks, fish portions, bakery items and other products.

The spiral-belt freezer (Figures 3 and 4) offers a design that maximizes the belt surface area in a given floor space. It is achieved by using a product belt that can bend around a rotating drum; by stacking up to 40 tiers of belt on a drum, floor space occupied is kept to a minimum. The continuous belt eliminates

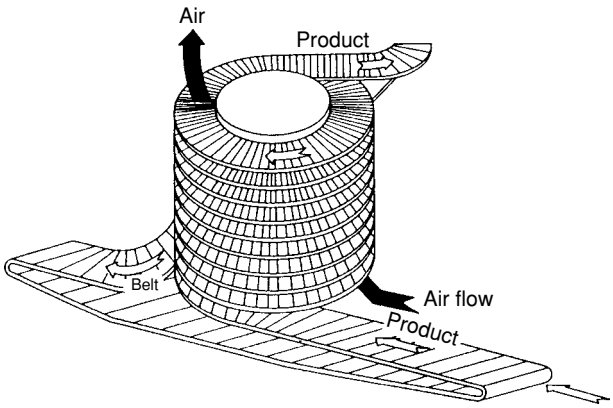

Figure 3 Spiral-belt freezer (Michael Boast Associates). Reproduced from Blast and Plate Freezing, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

product transfer points internal to the system. The product is transferred only at the in-feed and out-feed ends of the freezer. Spiral-belt freezers are suitable for unpacked meat, fish, and poultry products, including meat patés, meat balls, fish fillets and chicken portions.

Fluidized Bed Freezers

Fluidization is defined as a method to keep solid particles floating in an upward direction (Figure 5)

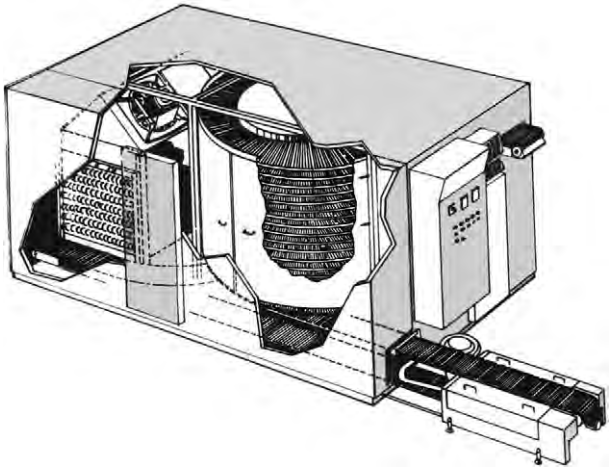

Figure 4 Diagrammatic arrangement of a spiral-belt freezer. Reproduced from Blast and Plate Freezing, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

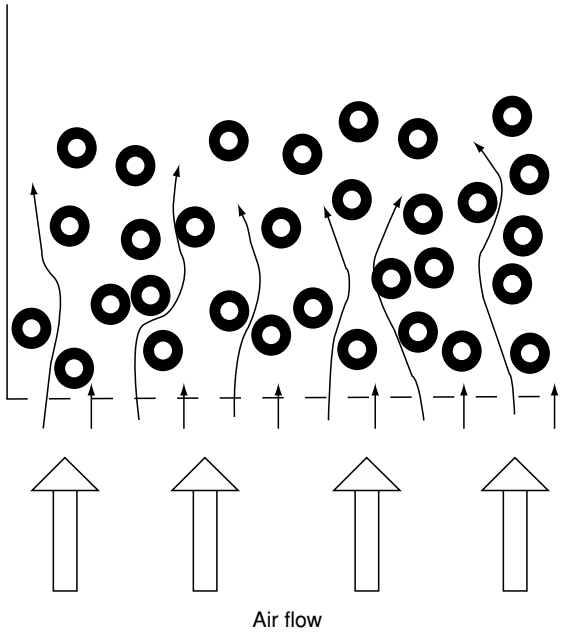

Figure 5 Fluidization principle (Michael Boast Associates). Reproduced from Blast and Plate Freezing, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

in a flow of gas or liquid. In freezing, fluidization occurs when particles of a similar shape and size are subjected to an upward stream of low-temperature air. At certain air velocities, particles float in the airstream, with each particle be separated from the other but surrounded by air and free to move. In this

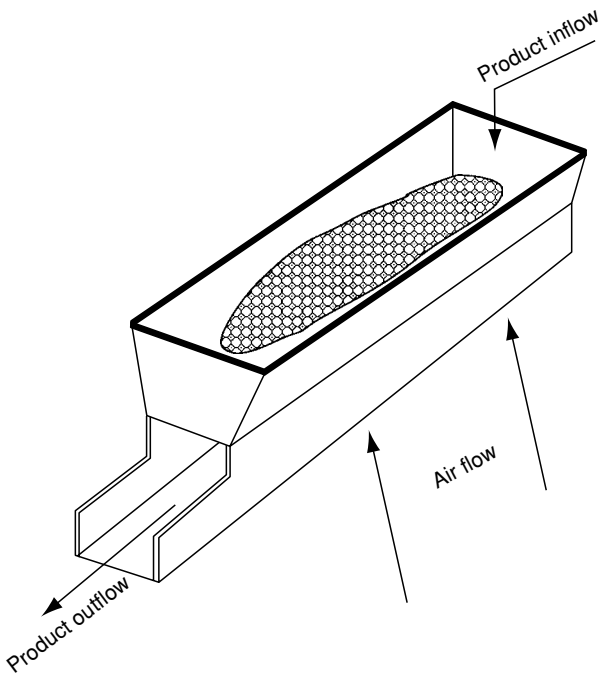

Figure 6 Fluidized bed freezer (Michael Boast Associates). Reproduced from Blast and Plate Freezing, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

state, the particle mass assumes the properties of a liquid or a fluid. By utilizing low-temperature air to achieve fluidization, the product is simultaneously frozen and conveyed by the same air without the aid of a mechanical belt.

Using the fluidization principle for freezing provides several advantages over the conventional belt. Most products that tend to stick together, i.e., sliced green beans, sliced carrots, and sliced cucumbers, are individually quick-frozen. The fluidized bed freezer (Figure 6) is effective and reliable in freezing wet products, as it can accept products with a high surface water content. Another advantage of the fluidized bed freezer is its complete independence from product variations and flow. Even when running at reduced capacity, the same evenly distributed air pattern is maintained.

Plate Freezers

In a plate freezer (Figure 7), the product is pressed firmly between top and bottom metal plates. The refrigerant is circulated in channels, housed inside the plates. This insures good heat transfer and reasonable freezing times.

Because heat transfer at the surface is gradually reduced with increasing product thickness, package

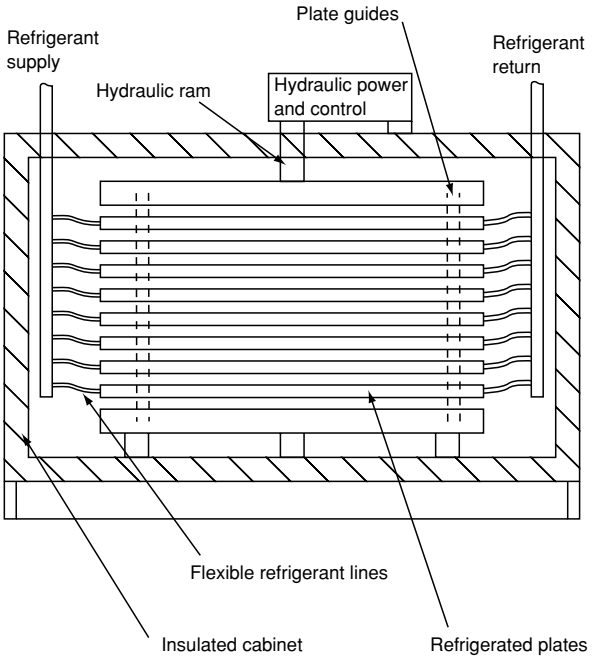

Figure 7 Horizontal plate freezer (Michael Boast Associates). Reproduced from Blast and Plate Freezing, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

thickness has to be limited to a maximum of 50 mm. Pressure from the plates has the additional advantage that it eliminates the bulging of the packages during the freezing cycle, so that the packages are discharged with straight sides and within close tolerances.

The two main types of plate freezers are horizontal or vertical in design; either type can be manual or automatic. In general, manual horizontal plate freezers have 15–20 plates. The product is placed on metal trays at the end of a packaging line, loaded in a rack or trolley, and transported to the freezer. The trays are loaded manually between the plates.

The vertical plate freezer was developed specifically for the freezing of fish at sea; it comprises a series of vertical freezing plates that form partitions in the container. Products are fed from the top, and the finished block of frozen products are discharged to the side, top, or bottom. Usually, this operation is mechanized. Whole unpacked fish are frozen, but fillets may also be handled in a vertical plate freezer. The block thickness varies from 25 to 100 mm.

Drum Freezer

An alternative to the flat plate freezer is the drum freezer, consisting of a rotating drum containing a refrigerant at low temperature, resulting in a cold surface for freezing. The product to be frozen is placed on the stainless steel refrigerated drum, which rotates at a preset speed depending on freezing time required. The product must be frozen, within one rotation. This rotation will last for less than 1 min if the product is only 3 mm thick, such as chopped spinach paste. A 6-min rotation would be required for a 10-mm-thick plaice fish fillet and 16-min rotation for a 20-mm-thick cod fillet.

Immersion Freezing

For irregularly shaped products such as chickens and turkeys, high heat transfer can be achieved in an immersion freezer, which normally consists of a tank that houses a refrigerated brine. The brine is most often of glycol or sodium chloride solution. The product is immersed in the brine or sprayed while it is conveyed through the tank.

Immersion freezers are most commonly used for plastic-bagged chickens or turkeys. The product must be protected from contact with the brine by using a high-quality packaging that gives an absolute tight seal. The brine residues on the packages are normally washed off with water at the freezer exit.

See also: **Freezing**: Principles; Operations; Cryogenic Freezing; Storage of Frozen Foods

Further Reading

ASHRAE (1998) *Refrigeration Systems and Applications*. Atlanta, GA: American Society of Heating Refrigeration and Air Conditioning Engineers.

ASHRAE (2001) *Refrigeration Fundamentals*. Atlanta, GA: American Society of Heating Refrigeration and Air Conditioning Engineers.

Boast MFG (1986) *Refrigeration and Air Conditioning*. Oxford: Butterworth Heinemann.

Boast MFG (1991) *Refrigeration Note Book*. Oxford: Butterworth Heinemann.

Robinson RK (ed.) (1985) *Microbiology of Frozen Foods*. Barking, UK: Elsevier Science.

Cryogenic Freezing

C James and S James, University of Bristol, Langford, UK

Introduction

Freezing methods can be divided into two main classes: (1) those involving direct contact between

the refrigerant and the food; and (2) those involving the use of a secondary medium, e.g., air, brine, or metal plate, which is cooled by the refrigerant.

Cryogenic freezing uses refrigerants, such as liquid nitrogen or solid carbon dioxide, directly. The boiling-off of the refrigerant when it comes in contact with the product brings about cooling. As well as using the latent heat absorbed by the boiling liquid, sensible heat is absorbed by the resulting cold gas.

Most cryogenic systems use total loss refrigerants, i.e., the refrigerant is released to the atmosphere and not recovered. Due to environmental and economic factors, total loss refrigerants must be both readily available and harmless, which limits the choice to atmospheric air and its components.

Table 1 Properties of cryogenic materials

Material	*Boiling point °C at 1 atm*	*Latent heat of vaporization (kcal kg^{-1})*
Helium	−268.8	108.8
Nitrogen	−195.8	47.3
Carbon monoxide	−190.6	53.5
Argon	−184.4	37.0
Methane	−161.1	136.0
Ethane	−88.9	109.0
Nitrous oxide	−88.9	90.0
Propane	−42.2	100.6
Carbon dioxide	−57.6 (5 atm)	75 (approx.)
Carbon dioxide (solid)	−79 (sublimation point)	135.4

History

Cryogenic freezing was first carried out commercially using liquid air, in the 1930s. However, liquid air contains a high proportion of liquid oxygen, which is a powerful oxidizing agent. Theoretically it can be produced on site, eliminating the need to purchase and store 'gas.' Although companies have promoted the use of liquid air, in practice it has been superseded by less harmful liquid nitrogen and liquid, or solid, carbon dioxide.

Food freezing using liquid nitrogen or carbon dioxide became established in Europe from the early 1960s onwards as manufacturers looked for uses of these byproducts. Liquid nitrogen is the byproduct produced when liquid oxygen is made by air distillation, while carbon dioxide is a byproduct of fermentation processes and a constituent of natural gas. By the early 1990s an estimated 10% of all frozen food in the UK was produced using cryogens.

An alternative to the air-based total loss cryogens, liquid dichlorodifluoromethane (food grade R 12, freon, CCl_2F_2), was introduced commercially in 1968 by Du Pont. R 12 had the advantage over other cryogens of being easily recoverable and having a higher operating temperature (−29.8 °C), reducing the problems of thermal shock. It was claimed that such gentle handling allows 'true' individual quick freezing (IQF) of products such as soft berries, meat patties, raw shrimps, breaded scampi, fried onion rings, etc. in less than 6 min. However, its use was not permitted in all countries and it was not widely used.

Physical Properties

The major problem with most cryogenic materials is that they are extremely cold (Table 1). Direct immersion can be inadvisable since an extremely high temperature gradient can be imposed on the food, which is often sufficient to cause it to disintegrate. Research studies have concluded that the ideal cryogen would have a boiling point of −50 °C and a latent heat of evaporation as high as possible.

Liquid nitrogen is produced by liquefaction of air, either as a principal product or as a byproduct of liquid oxygen. Nitrogen is the main constituent of atmospheric air and at atmospheric pressure liquefies at a temperature of −196 °C. It is usually supplied and stored at a pressure of 3–6 bar, with corresponding boiling points of −185 °C to −177 °C. A useful rule of thumb is that 1 t h^{-1} of liquid nitrogen is approximately equivalent to 100 kW of mechanical refrigeration.

Since liquid nitrogen is made from air, its costs are entirely those of manufacture, i.e., the capital cost of the plant, electricity, and transportation. The energy required to produce 1 kg of liquid nitrogen is around 3000 kJ, or eight times the consequent stored refrigerating effect.

Carbon dioxide's physical properties are unusual, in that it does not exist in liquid form at atmospheric pressure. If stored as a pressurized liquid and released into the atmosphere, the liquid changes partly to gas and partly to a frozen solid at −78.5 °C, which sublimes directly into gas without going through a liquid phase.

Liquid carbon dioxide is generally supplied either at ambient temperature (e.g., 25 °C and 65 bar), giving a refrigerating capacity of 199 kJ kg^{-1}, or at −16 °C and 22 bar, giving a refrigerating capacity of 311 kJ kg^{-1}. At the point of use, spray nozzles reduce the pressure of the liquid, generating a mixture of cold vapor and solid carbon dioxide 'snow.'

The majority of the 'refrigeration effect' stored in solid CO_2 is latent, whereas in liquid nitrogen almost half the effect is due to sensible heat transfer to the cold gas. The available refrigerating effect from 1 kg

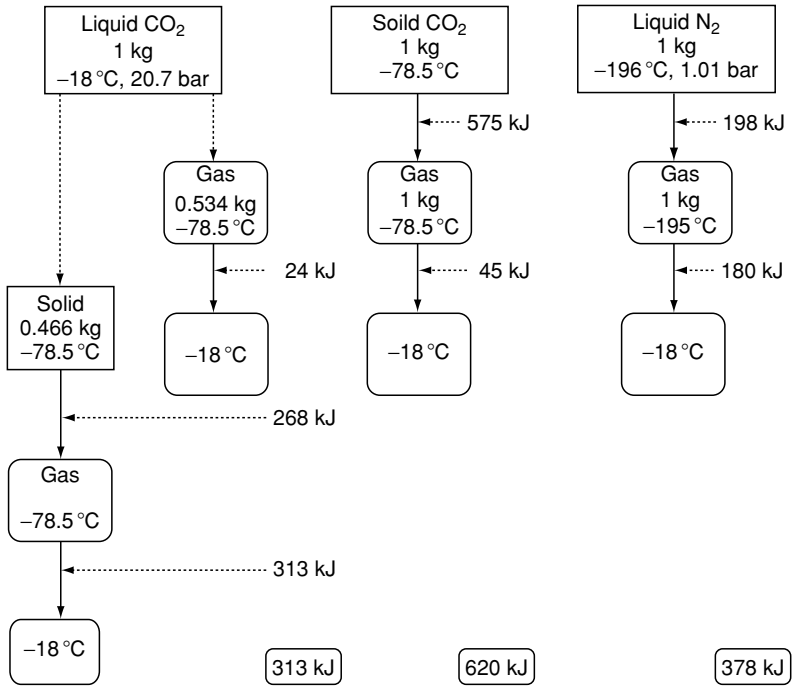

Figure 1 Available refrigeration effect from 1 kg of total loss refrigerant.

of both liquid and solid carbon dioxide and liquid nitrogen is shown in Figure 1.

Comparison of Cryogenic Freezing Systems with Mechanical Freezing Systems

The generally perceived advantages of cryogenic compared with conventional refrigeration are given in Table 2.

Table 2 Advantages and disadvantages of total loss refrigerants in comparison with mechanical refrigeration

Advantages	*Disadvantages*
Low capital investment	High operating cost
High refrigerating capacity	High refrigerating capacity
Compact (small 'footprint')	Poor temperature control
Flexible	Limited duration without filling
Low weight when out of use	High weight at start of use
No residual weight (dry ice)	Reduced humidity
Silent operation	Suffocation hazard
Advantageous storage atmosphere (N_2)	Limited availability
Bacteriostatic affect (CO_2)	
Low maintenance requirements	
Long life	
Foolproof once installed (dry ice)	

Capital Cost

The low capital cost and simplicity of design permit cryogenic systems to be used in a standby or back-up role in emergencies or for 'peak lopping' during periods of exceptional freezing demand.

The overall economics can be favorable for short road deliveries. For static standby installations, or those with very low loading factors, low capital cost can be a major attraction.

Running Costs

In overall energy or fuel terms, there is no way in which cryogenic refrigerants can compete with the best mechanical refrigeration systems. Generally speaking, cryogenic refrigeration offers lower capital and maintenance costs but incurs higher running costs. Many cryogenic freezers are leased from the gas suppliers, reducing capital cost further.

The consumption of the cryogen (Table 3) contributes the major proportion of the freezing costs (about 75%). Refrigerant consumption is measured as a 'consumption ratio' equal to weight of refrigerant used divided by the weight of product frozen. Table 4 shows some approximate quantities of liquid nitrogen and carbon dioxide required to freeze a range of products. High refrigerant consumption can be

Table 3 Major factors affecting cryogen consumption ratio

Factors
Moisture content
Temperature in/out
Residence (freezing) time
Belt loading ratio
Production rate
Operator discipline
Length of production run
Efficiency of freezer

Table 4 The cryogen requirements for freezing various foodstuffs

Foodstuff	*Liquid nitrogen (kg cryogen kg^{-1} foodstuff)*	*Carbon dioxide[a] (kg cryogen kg^{-1} foodstuff)*
Meat and meat products	0.7–1.2	
Beef fat		0.82
Beef lean		1.05
Veal		0.95
Pork		0.73
Lamb		0.88
Beefburgers	0.6–0.8	0.92
Meat pies		0.91
Poultry	1.1–1.3	1.03
Fish		1.03
Shellfish – scampi	1.0–1.3	1.27
Fruit and vegetables	1–1.5	
Beans (green)		1.29
Peas		1.10
Asparagus		1.34
Strawberries		1.30
Fruit pies		1.11
Pastry products	0.4–1.1	
Precooked foods (warm)	1.6	
Icecream	0.7	1.02
Bread		0.66

[a]Theoretical requirements assuming foodstuff temperature reduction from 20 to −18 °C.

significant with cryogenic tunnels, especially following a change in throughput or product type. The reason for this is poor control. A quantity of the cryogen is also required to cool the equipment down prior to start-up and to maintain temperature during idle periods. With efficient plant usage this need not account for more than an additional 2%, over and above that used for actual product freezing.

Maintenance

Cryogenic freezers have few moving parts so maintenance is far easier than for mechanical systems. There are obvious benefits for long rail journeys in using dry ice for which no further attention is required once the cargo and refrigerant have been loaded.

Weight Loss

In several freezing applications, the advantages listed in Table 2 are sufficient to overcome the penalties of cryogenic systems. The potential saving in weight loss in freezing unwrapped products in comparison with air-blast freezing can be substantial. For products of great value such as shellfish this is of particular economic significance, since the lost weight/water has the same value per unit weight as the product.

Flexibility

The flexibility factor has other ramifications. Mechanical systems are limited in their output of cold – both above and below the norm. Overloading affects quality; underloading increases unit costs. With cryogenic systems the 'cold' is on tap. This is particularly important with a new product or a seasonal product when demand is unpredictable and variable.

Cryogenic systems are also often flexible in the type of product that can be fed through them. In a conventional blast freezer, the maximum air velocity inside the freezer limits the amount of heat given up by the evaporator coils. In a cryogenic tunnel, the rate at which refrigerant is added is more or less independent of the gas velocity. This gives cryogenic tunnels considerable flexibility in responding to changes in production rate.

Size

Cryogenic freezers tend to be more compact than those using mechanical refrigeration, in some cases needing only half the floor space for the same throughput. Carbon dioxide systems are extremely compact, for example 1360 kg of product per hour can be produced using an area of 3.6×3.6 m. Also the equipment supplying the refrigeration itself takes up substantially less room than mechanical refrigeration equipment.

Weight

Low weight is a particularly attractive aspect for airline catering applications, and other transport applications.

Noise

Cryogenic systems operate relatively quietly compared with the noise and vibration caused by mechanical compressors. With the growing concern over noise pollution, this may become more of an issue. In

road transport, overnight parking of mechanically refrigerated vehicles is sometimes difficult in noise-sensitive areas, and cryogenic systems overcome this problem.

Safety

In comparison with nitrogen, carbon dioxide is safer. Carbon dioxide is sensed by humans at levels less than 0.5%, although much higher concentrations of nitrogen are tolerable it is not sensed and a potential hazard stems from fainting due to hypoxia (shortage of oxygen to the brain).

Synergism

The rapid freezing rates possible with cryogenic freezing may be employed to enhance mechanical systems. An example is the crust freezing of delicate product, such as fish fillets, using cryogens before freezing in spiral freezers to prevent mechanical damage as the product travels through the spiral. Crust freezing can also be employed to reduce dehydration. However, legislation can complicate this issue. In countries where practices to compensate for weight loss, such as misting of burgers, are allowed, the economics don't favor crust freezing.

Crust freezing can also be used to reduce frosting problems. Moisture loss during freezing can cause frosting of refrigeration coils, reducing the efficiency of the systems and requiring defrosting. Crust freezing traps the water, preventing moisture migration.

Equipment

Cryogens are typically employed as sprays in tunnel or batch cabinet systems, or direct immersion tunnels. Operation of cryogenic systems is relatively straightforward. The two basic functions to regulate are: (1) conveyor speed to give the required retention time, and (2) refrigerant flow rate to give the desired product temperature.

Tunnel Freezers

Essentially a tunnel freezer is a countercurrent heat exchanger. The cryogen is sprayed at one end and the gas passed countercurrent to product traveling on a belt in the opposite direction. This enables the exit temperature to be only 20 °C below the product inlet temperature. Thus product can be precooled before being subjected to the full force of the refrigeration effect. Approximately 50% of the product's heat can be extracted before it reaches the cryogen spray. Straight or spiral tunnels are utilized. As a rule, carbon dioxide tunnels tend to be longer than liquid nitrogen tunnels for the same capacity.

A tunnel freezer operates by conveying the product through an enclosed space within which the cryogen is discharged. Economy of cryogen consumption is improved by insulation and isolation of the freezing tunnel, and by control of the cold gas within the tunnel. The product passes through three distinct zones in the tunnel: the precool zone, the freezing zone, and the postcool zone. In the precool zone, the product is cooled using gas and fans force gas vertically past the product. In the freezing zone, sprays discharge cryogen directly on product and belt, while in the postcool zone, which is a gas-cooled section, with more fan circulation, some temperature stabilization takes place.

Rotary Freezers

Rotary freezers have also been developed to IQF freeze products such as shrimps, cubed chicken meat, or meat balls. A fine mist of cryogen is used to freeze the surface of the product and prevent individual pieces sticking to one another before they enter a rotating drum, which keeps them separated during the freezing process.

Immersion Freezers

Immersion freezers are the simplest and most inexpensive way of using cryogens to freeze foodstuffs. Immersion freezers are best suited to high volumes of product where crust freezing is required or where the product is sufficiently robust to withstand the thermal shock of a high temperature differential. Very fast heat transfers are possible, but the system is wasteful in that only the latent heat of the liquid is used and cold vapor is lost.

Cabinet Freezers

Batch cabinet freezers utilizing cryogens have been developed for freezing prepared and precooked food, particularly for airline meals. Liquid cryogen is injected at time intervals into a cabinet containing the food on racks; circulation fans are usually utilized to provide even and efficient freezing. Freezing times vary from 5 min to 1 h, depending on the product and its temperature.

Pellet Freezers

Liquids and semiliquids are often frozen into pellets. This can be carried out using contact freezing between belts or in molds, or cryogenic freezing using liquid nitrogen and a forming device. Belt freezers employ a similar contact method of freezing to plate freezers. Single-band and double-band freezers are designed to freeze thin layers, usually of liquid or

semiliquid products such as vegetable purées, fruit pulps, egg yolk, sauces, soup, etc. In double-band systems the product is frozen between two endless belts, of which the top is flat and the lower belt corrugated. The product is spread into the corrugations; the top belt encloses the exposed surface, thus freezing the product as IQF pellets.

Combined Freezers

A cost-effective alternative to 100% cryogenic freezing is a combined freezer. The principle is to use a cryogenic freezing unit for initial freezing of the outer surface of the product, followed immediately by mechanical freezing to reduce the temperature of the bulk of the product. One advantage of this type of approach is that the cryogenic stage can often be retrofitted.

Freezing Rate

Most food freezing systems rely on convection as the principal means of heat removal. The rate of heat removal from the product depends on the surface area available for heat flow; the temperature difference between the surface and the medium; and the surface heat transfer coefficient. Much higher coefficients are achieved in cryogenic systems than in most conventional refrigeration (Table 5).

The rate at which heat can be conducted away from the surface is not the sole criterion that governs the time taken to freeze a product. Heat must also be conducted from within the product to its surface before it can be removed. Most foodstuffs are poor conductors of heat and this imposes a severe limitation on attainable freezing times for either large individual items or small items frozen in bulk.

Cryogens are particularly suited to food products that have a high surface-area-to-volume ratio and in which the thermal diffusivity of the food does not restrict the transfer of heat from the product to the freezing medium. Typical examples are fish fillets, shellfish, pastries, burgers, meat slices, sausages, pizzas, and extruded products.

Table 5 Published surface heat transfer coefficients for different cooling processes

Method	*Heat transfer coefficient ($W m^{-2} K^{-1}$)*
Still air	5–10
Air blast, slow	15–20
Air blast, fast	40–50
Fluidized bed	100–4000
Contact, plate	500
Immersion, still	50–200
Immersion, agitated	< 1000
Cryogenic, R 12	1000
Cryogenic, liquid nitrogen	1500

Product

Typical product freezing times in nitrogen systems are shown in Table 6.

Rapid freezing has been claimed to improve taste, texture, aroma, nutritive value and appearance, bacteriological, enzymic, oxidative and chemical degradation, and to reduce drip loss upon thawing.

Ice Crystal Formation

The most apparent influence of the freezing rate on product quality is related to the size of the ice crystals, which are formed during the freezing process.

It has long been shown by photomicrographs that the faster meat and fish are frozen, the smaller the ice crystals produced, and that above a certain rate these are predominantly intracellular, when by implication the damage to cell walls should be minimal. However, extracellular ice crystal formation is not necessarily disruptive because muscle cell wall membranes are elastic, unlike those of vegetable tissue. Nevertheless, largely on the strength of the photomicrographs, plausible advocacy of rapid freezing equipment throughout the 1930s and 1940s resulted, and these views remain.

Quality

The debate about the relationship between freezing rate and quality has gone on for many years. Many trade publications indicate that a considerable improvement in quality can be obtained by ultrarapid freezing with cryogens. However, it is doubtful whether the rate of freezing has any detectable effect on drip or on palatability of meat after thawing and cooking over a wide range of freezing rates. Recent

Table 6 Freezing times of nitrogen freezers

Product	*Freezing time (min)*
Fruit and vegetables	0.5–6
Meat and meat products	3–20[a]
Meat patties, hamburgers	3–5
Pastry products	4–8
Precooked foods (warm)	4–8

[a]Large, bulky meat products.

studies with beefburgers show no differences in eating quality between those frozen to −18 °C in 22 or 2.7 min. For fish, it has been concluded that freezing times of up to 26 h do not significantly influence the quality of the product.

A number of authors have classified foods according to their sensitivity regarding freezing rate into four groups:

1. Products that remain practically unaffected by variations in freezing rate, i.e., products with a high content of dry matter, such as peas, meat products with a high fat content, and certain ready meals.
2. Products which require a minimum freezing rate (0.5–1 °C min^{-1}), but are relatively unaffected by higher freezing rates, i.e., fish, lean meat, and many starch- and flour-thickened ready meals.
3. Products whose quality improves when freezing rates are increased (3–6 °C min^{-1}), e.g., many fruits, egg products, and flour-thickened sauces.
4. Products in which high freezing rates are advantageous for the product quality, but where temperature tensions in the product result in a destruction of the tissue, e.g., fruits and vegetables such as raspberries, tomatoes, and cucumbers.

Some work has indicated that there is an interaction between freezing rate and cooking method. When meat was cooked after thawing, no differences were found between freezing rates (13, 2, or 0.04 cm h^{-1}) in terms of flavor, juiciness, tenderness, thaw loss, or TBA (thiobarbituric acid) values. However, meat that had been cooked from frozen was found to show a favorable effect of faster freezing rates.

With mushrooms, while the quantity of the drip loss is almost independent of the freezing rate, the color of the lost juice changes conspicuously. It contains more coloring and flavoring ingredients the slower the freezing.

Among the products that obtain a decisive and lasting quality improvement by rapid freezing are carrots, corn on the cob, mushrooms, egg products, sauces, cream, shrimps, and prawns.

Appearance

Freezing rate does affect the appearance of many food products. For example, fast freezing of poultry tends to produce a lighter-colored product as the small ice crystals scatter the light more than larger crystals. This lighter color is preferred by many consumers. In fish, rapidly frozen fillets have a dense white, opaque appearance, in comparison with the dark, translucent, vitreous product produced by very slow freezing.

Drip Loss

When ice crystals grow they may puncture the cell walls so the juice bleeds out. This causes 'drip loss' when the product is thawing. The larger the average crystal size is, the greater the number of punctured cells and the greater the drip loss. The loss of juice can result in a loss of firmness and flavor.

The drip loss from some food (i.e., strawberries) can be clearly related to freezing rate. Slow freezing can produce drip losses of 20%. Fast freezing without surface cracking can reduce losses to approximately 5%. With meat and meat products freezing considerably increases drip. However, other factors inherent in the animal and its processing before freezing have more effect on the magnitude of drip than differences in freezing rate.

There are many products that are situated between these two types.

Dehydration–Weight Loss

Lower weight loss, during the freezing of unwrapped food, is one of the main advantages of cryogenic freezing. Weight loss during freezing of beefburgers can be as high at 3% in a poorly designed air blast freezing system compared with 0.4% in a liquid nitrogen tunnel.

However, technological advances in conventional freezing systems can substantially reduce – if not completely remove – this saving. In a modern spiral freezer operating at −30 °C air temperature, the weight loss will be approximately 1.2%. In an air impingement freezer operating at a −46 °C air temperature, the weight loss from a burger can be <0.4%

Cracking

Some products, delicate soft fruits and light pastry products, may crack when they are submitted to very high freezing rates or very low temperatures. Research suggests that crust freezing produces a shell that prevents further volume expansion, when the internal portion of the unfrozen material undergoes phase transition. If the internal stress is higher than the frozen material strength, the product will crack during freezing. Products with high void spaces, which allow internal stresses to dissipate, show a lessened chance of freeze-cracking. Precooling prevents freeze-cracking because it reduces the differences in temperature between the product and the freezing medium, and reduces the difference between freezing time at the center and at the surface. When the phase change of the core region occurs before the surface becomes brittle, food products can support

the internal pressure and freeze-cracking does not occur.

Summary

Cryogenic systems can be used effectively to freeze, transport, and store a wide range of foodstuffs, and a wide range of equipment and techniques is available using either nitrogen or carbon dioxide. While there are many practical, as well as economic, disadvantages to cryogenic refrigeration systems, there are many situations in which low capital or maintenance cost and/or rapid freezing rates make these techniques worthwhile.

See also: **Freezing**: Principles; Operations; Storage of Frozen Foods; Structural and Flavor (Flavour) Changes; Nutritional Value of Frozen Foods; **Transport Logistics of Food**

Further Reading

Anonymous (2001) *Rapid Cooling – Above and Below Zero*. Paris: International Institute of Refrigeration.

Bald WB (ed.) (1991) *Food Freezing: Today and Tomorrow*. London: Springer-Verlag.

George RM (1993) Freezing processes used in the food industry. *Trends in Food Science and Technology* 4: 134–138.

Heap RD and Mansfield JE (1983) The use of total loss refrigerants in transport of foodstuffs. *Australian Refrigeration, Air Conditioning and Heating* 37: 23–26.

Kennedy CJ (ed.) (2000) *Managing Frozen Foods*. Cambridge: Woodhead Publishing.

Mallett CP (ed.) (1993) *Frozen Food Technology*. London: Blackie Academic & Professional.

Storage of Frozen Foods

G Löndahl, Helsingborg, Sweden
T Nilsson, Enebyberg, Sweden

Introduction

The quality of frozen foods depends on the quality and nature of the product to be frozen, on the processing of the product prior to freezing, the freezing process, and the packaging of the product.

During storage, reactions of a physical as well as chemical and biochemical nature will cause changes in the food which gradually lead to a loss of quality. The changes are cumulative and the shelf-life of frozen food is limited by, and dependent on, the nature and speed of the reactions.

Some of the most important changes during storage, which will influence the quality, are as follows:

- loss of water
- loss and transfer of volatile compounds
- oxidative changes in fats
- denaturation of proteins
- color changes
- transfer of aromatic compounds from product to product

These changes, which will be dealt with in more detail in the following sections, are in most cases interacting, and their influence on the overall quality of the food is complex in its nature. It is therefore difficult to find a single objective method to determine the shelf-life of a product. However, much work has been directed with this goal in mind. In some cases, valuable information can be gained from such tests, e.g., determining the rancidity of fatty products, loss of vitamins, weight loss, color changes, etc. However, in most cases organoleptic tests are the most common methods used to determine the shelf-life for frozen products stored under various conditions.

All changes during storage are temperature-dependent. In general, the lower the temperature, the slower the speed of the deteriorative changes. (*See* **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability.)

Loss of Water, Dehydration

During storage, the water vapor pressure over the surface of the product is higher than that of the surrounding air. Since the surface of the cooling coils always has a lower temperature than the air in the storage room, ice will form on the coils. Thus a less than saturated relative humidity is maintained in the air, and also a difference in vapor pressure between the surface of the product and the surrounding air. As a consequence, sublimation of ice from the product will occur and continue as long as this pressure difference is maintained.

Not only does the loss of water mean a loss of weight of the product, which may have economical consequences, but also, in many cases, drying may severely impair the quality of the product. The water evaporates from the surface of small ice crystals. Such crystals have, in relation to volume, a larger area than bigger ones and are therefore prone to evaporate and disappear first. The contact surface between the product tissue and the surrounding air (oxygen) is thereby considerably enhanced, which will increase the

oxidative reactions in the product. In some cases, combinations of oxidative reactions in the proteins and drying will lead to a condition known as freezer burn. This is a severe quality change in the surface of the product. (*See* **Drying**: Theory of Air-drying.)

Prevention of Water Loss and Drying

Packing

If the product is packed in a tight-fitting, water-vapor-proof material, evaporation cannot take place. However, it is important that the packing material is tight-fitting; otherwise, sublimation will take place within the package. The packing material will follow the temperature changes caused by temperature fluctuations in the room faster than the product itself. When the temperature is lowered, evaporation from the product will form ice on the inside of the packing material, and when the temperature conditions are reversed the ice will be deposited on the surface of the product. If there are big variations in the storage temperature, ice formation within the package can be quite important. Small spaces within the package and a storage temperature as even as possible are measures to avoid this condition. If the temperature differences between the surface of the cooling equipment (coils) and the storage temperature are small, sublimation from the product will be low. Good insulation and efficient cooling sytems are effective tools to achieve this. Temperature abuses, which will allow warm air or warm products to enter the room, should be avoided.

Glazing

Glazing is the term used for the application of a thin layer of ice on the surface of a frozen product either by dipping the product in water or by spraying (Table 1). This is an effective method of preventing drying of the product. The sublimation is less pronounced since it takes place from an even surface which occupies a much smaller area than the ice crystals in the product. In addition, the loss of water comes from added water; the amount can be calculated to prevent any sublimation from the product during the intended storage period.

The lower the storage temperature, the lower the sublimation losses as the water vapor pressure decreases with the temperature. At $-10\,^{\circ}C$, it is 2.0 mmHg over ice as compared to 0.3 mmHg at $-30\,^{\circ}C$.

Table 1 Storage life: acceptability for frozen shrimps (*Pandalus borealis*)

Product	*Acceptability (months at −18 °C)*
Block-frozen, vacuum-packed	6–7
IQF, Polybag	3–4
IQF, vacuum-packed	6–7
IQF, vacuum-packed, + 4% glaze	8–9
IQF, vacuum-packed, + 8% glaze	>10

IQF, individually quick frozen.

Loss and Transfer of Volatile Compounds

Water is not the only substance that will escape from a frozen food during storage. Many substances which cause the typical smell and aroma of a food are also volatile at frozen storage temperatures and will evaporate from the surface of the product and enter the surrounding air. As a result, some of the aroma characteristics of the product are diminished or even lost during storage, and this will affect the quality of the product. (*See* **Sensory Evaluation**: Aroma.)

A more severe quality consequence of this phenomenon, however, concerns other products stored in the same environment. If a 'foreign' aroma compound is deposited on a product and dissolved in either the fat or water phase of the product, it can severely affect the quality of the product. Meat, dairy products, including icecream, and other foods with a high fat content are most susceptible to pick up aroma from other products, e.g., fruits, especially citrus fruits and strawberries, some vegetables, fish, etc. This phenomenon is well known during chilled storage, but it is also important during cold storage. It has led in some cases to considerable economic losses.

The only effective method of preventing aroma transfer during cold storage is packing the product in a material which will act as a barrier for the compound in question. Since the compounds responsible for the damage are not easily identified, choice of packing material must in most cases be based on experience and tests. A low storage temperature will reduce the release of volatile compounds and subsequently diminish the problem, but cannot prevent it. A suitable packing material is therefore the only method left. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Oxidation of Fat

One of the most important quality changes in frozen food during storage is the formation of off-flavors caused by rancidity of fats. Rancidity is caused by oxidative changes of polyunsaturated fatty acids. The higher the proportion of polyunsaturated fatty acids, the more prone the fat is to undergo oxidative changes. Some species of fish (e.g., herring, salmon)

are rich in fat with a high content of polyunsaturated fatty acids, and oxidative changes (rancidity) are more pronounced in these species than in lean ones (e.g., cod). In lean species, oxidative reactions take place, but to a much lesser degree. One compound causing off-flavors – 'cold-store flavor' – has been identified as *cis*-4-heptenol, which results from the oxidation of ω-3 polyunsaturated fatty acids. (*See* **Oxidation of Food Components**.)

Off-flavors are the most important quality factors resulting from oxidation of fatty acids but, in addition, color changes in the tissue can occur as a result of reactions between proteins and the oxidation products.

Rancidity changes in fat are also caused by lipolytic enzymes, such as lipases and phospholipases. Since these enzymes are active at low temperatures, lipolytic changes with an accummulation of free fatty acids can take place during cold storage. Interactions between fatty acids and other components in the food can cause changes in fat-rich food which will affect the quality. Normally, however, oxidative changes in fat are the most important factors influencing the quality of fat during frozen storage.

Oxidation of fatty acids takes place not only in fat from fishes, but also in fat from other animals. In beef the fat contains only a small amount of polyunsaturated fatty acids, and rancidity caused by oxidation plays a minor role. In pork, however, the fat contains a much higher amount of polyunsaturated fatty acids, and pork meat is consequently much more prone to oxidative changes during frozen storage. The composition of the fat in pork depends to a large extent on the composition of the fat in the feed, and the keeping quality of different pork lots can differ considerably even if stored under identical conditions.

Oxidation of fat is enhanced by the presence of certain catalytical factors, e.g., special metal ions, heme products, or salt, but the process is slowed down or inhibited by so-called antioxidative substances. If the contact area between oxygen and the tissue is increased, as is the case when ice crystals disappear due to drying, oxidation is considerably enhanced. (*See* **Antioxidants**: Natural Antioxidants.)

Oxidation of fatty acids can be inhibited by preventing oxygen (air) from coming into contact with the fat. As has been shown in numerous investigations, close-fitting packing in an oxygen-proof material, e.g., vaccum-packing, will minimize the oxidative reactions and considerably prolong the shelf-life of the product. The same result can also be achieved by substituting air within the package by an inert gas, e.g., nitrogen. In addition, glazing gives good protection against oxygen reaching the fat.

Changes in Proteins

During frozen storage, changes in the quality of frozen food of both vegetable and animal origin take place as a result of alterations and reactions of the proteins. Many of these protein alterations are still poorly understood, while others have been thoroughly studied in numerous investigations.

Animal Proteins

A well-known characteristic of frozen fish and mammals is a decrease in the water-binding capacity of muscle tissue, resulting in a loss of juice at thawing (drip). In addition, an increase in the firmness of the meat flesh is often observed. These changes, which are more pronounced in fish, also take place in meat (beef and pork) but to a lesser degree. (*See* **Fish**: Processing; **Meat**: Preservation; **Water Activity**: Effect on Food Stability.)

The most important changes take place in the myofibrillar proteins, especially in the myosin globular head of the molecule, and lead to a decrease in its solubility, while the thin filaments show a much higher stability during storage at low temperatures. The alterations have been shown to be temperature-dependent and occur more rapidly at higher storage temperatures. Even if the rate of naturally occurring metabolic changes in the muscle tissue is minimized during frozen storage, some reactions play a role during storage. The activity of the enzymatic calcium adenosine triphosphatase ((Ca^{+})-ATPase) complex, which is of importance for the quality of the meat, is highly temperature-dependent. If the freezing process is slow, a great loss of the enzymatic activity takes place compared to a fast freezing rate. During storage, loss of activity of the enzyme complex continues at a rate dependent on the storage temperature. If the meat has been frozen in prerigor state and kept at a low storage temperature, remaining enzymes are able to cause a very fast reaction when the temperature is raised during thawing. The condition can lead to a phenomenon known as thaw rigor, in which a sudden contraction of muscle fibers leads to a big drip loss.

The changes during freezing and frozen storage in the myofibrillar proteins of muscles from both fish and mammals are of the same biochemical nature. However, the effects on the quality of the product during frozen storage are different. The changes in fish muscles are generally more pronounced than those in meat. During frozen storage of fish, an increased firmness and dryness will result in a tougher, drier product, while beef and pork normally lose much less of their tenderness and juiciness during frozen storage.

Vegetable Proteins

Vegetables and fruits are harvested when ripening of the product is at its peak with regard to flavor, texture, and color. Since these properties are related by enzymatic reactions in the cells of the living plant, which will continue even at low temperatures after harvest, the quality of the product will soon deteriorate if the enzymes in question are not inactivated. In vegetables this is normally achieved by different methods of heat treatment (blanching). Lipoxygenases, catalases, and peroxides are enzymes involved in the deteriorative processes. Of these, peroxide is the most heat-resistant and, therefore, determination of peroxide activity is widely used to check the effect of blanching processes. (*See* **Ripening of Fruit**.)

If some enzymatic activity is left, or if a reaction takes place, changes will occur during storage which may affect the quality of the product. One example is the change of chlorophyls to pheophytines. The desirable green color of some vegetables changes to a yellow or gray color, severely affecting the quality of the product. Enzymatic reactions during frozen storage can also cause off-flavors of the product.

Hygienic Considerations

At normal storage temperatures for frozen food, i.e., $-18\,^{\circ}C$ or lower, all microbiological activity, including growth, has ceased. Some organisms will die or be injured during freezing and storage, the number depending on factors such as the species involved and the nature of the food. From a practical point of view this is of little importance. When the temperature rises at thawing, surviving organisms will resume their activity, including growth. The hygienic quality of frozen food should therefore be regarded as equal to the condition in the product immediately before freezing.

See also: **Antioxidants**: Natural Antioxidants; **Fish**: Processing; **Meat**: Preservation; **Oxidation of Food Components**; **Ripening of Fruit**; **Sensory Evaluation**: Aroma; **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability; **Water Activity**: Effect on Food Stability

Further Reading

Durand MP (2001) History of the application of cold to food. *Bulletin Academie Nationale du Medicine* 185(2): 269–272.

International Institute of Refrigeration (1986) *Recommendations for the Processing and Handling of Frozen Foods*, 3rd edn. Paris: International Institute of Refrigeration.

Jul M (1984) *The Quality of Frozen Foods*. London: Academic Press.

Löndahl G and Danielsson C (1972) *Time Temperature Tolerances for Some Fish and Meat Products*. Proceedings of the International Institute of Refrigeration, Warsaw, Commission C2.

Ludovico-Pelayo L, Hume A and Love RM (1984) Seasonal variations in flavour change of cold-stored trout. In: Zeuthen P, Cheftel JC, Eriksson C *et al.* (eds) *Thermal Processing and Quality of Foods*, pp. 659–663. Essex: Elsevier Science Publishers.

Phillips KM, Simpkins AH, Amama KR *et al.* (2001) Long-term stability of nutrients in a mixed food control material. *Fresenius Journal of Analytical Chemistry* 370(2–3): 297–302.

Wagner JR and Añon MG (1987) *Protein Denaturation in Bovine Muscle during Frozen Storage*. Proceedings of the XVIIth International Congress of Refrigeration, Wien, pp. 717–722. Paris: International Institute of Refrigeration.

Structural and Flavor (Flavour) Changes

C James and S James, University of Bristol, Langford, UK

Background

The main aim of freezing is to extend the storage life of the raw material or food product beyond that achieved at temperatures above the initial freezing point of the material. In a number of cases, freezing is used to produce a product, i.e., icecream, sorbet, ice lollies, etc., that cannot be produced in any other way.

If the temperature of the food is reduced to below $-10\,^{\circ}C$ within a few hours and maintained below that temperature, microorganisms, bacteria, molds, or yeasts will not grow. Microbially, the food will always remain as safe to eat as it was before freezing, irrespective of the length of frozen storage. Some microbiologists consider $-5\,^{\circ}C$ to be a practical maximum temperature to prevent the growth of microorganisms on food.

The textural quality of the food will change with time and conditions of storage. More importantly, the flavor will deteriorate with time. Ultimately, these changes will result in an unacceptable product.

Textural Changes

Most of the textural changes that occur as a result of freezing and frozen storage only become apparent when the food is thawed. A few changes can be observed while the food is still frozen. The most noticeable of these is called 'freezer burn.'

Changes Apparent in Frozen Product

Freezer burn Moisture is lost from the surface of a food when the vapor pressure at its surface is higher than that in the surrounding environment. Freezing considerably reduces the vapor pressure at the surface of a food, but it is still likely to be higher than that in a frozen storage room or display cabinet. Ice from the exposed surface of unwrapped frozen food sublimates directly into the surrounding environment. The rate of the subsequent desiccation depends mainly on the rate of air movement over the surface and its humidity, and extreme desiccation results in a layer of spongy, dry dark material developing on the surface. This is known as 'freezer burn.'

In-package frosting Tight packaging in a heavyweight, moisture-resistant wrapping prevents freezer burn. Loose packaging reduces freezer burn but can allow another problem, 'in-package frosting,' to develop. In-package frosting, occurs when moisture from the surface of the food freezes on the inner surface of the package, and the main cause of this is poor temperature control in the store or cabinet.

Light surfaces Rapid freezing of products such as poultry joints produces very small reflective ice crystals at the surface. These give the product a very light appearance that is very desirable in some markets. Poor temperature control during storage causes the small crystals to grow, merge, and lose their lightness.

Changes Apparent on Thawing

Most of the textural changes caused by freezing are not noticeable until thawing. During the freezing process as the water turns to ice, it expands, and the ice crystals formed cause the cell walls to rupture. Consequently, the texture of the product becomes much softer when the product thaws.

These textural changes are more noticeable in fruits and vegetables that have a higher water content and especially those that are eaten raw. For example, when a frozen tomato is thawed, it turns into mush and liquid. This explains why celery, lettuce and tomatoes are not usually frozen and why frozen fruit are best served before they have completely thawed. In the partially thawed state, the effect of freezing on the fruit tissue is less noticeable.

In the case of vegetables, blanching (exposure of the vegetable to boiling water or steam for a brief period of time) has a marked softening effect, but further changes during freezing or frozen storage are comparatively small. However, if the vegetable is frozen in the raw state or is not correctly blanched, the enzymic action on pectic substances may lead to unacceptable changes in texture.

Textural changes due to freezing are not as apparent in products which are cooked before eating because cooking also softens cell walls. These changes are less noticeable in high-starch vegetables, such as peas, corn, and lima beans.

Consumers expect to purchase frozen meat at a unit price lower than that of a similar chilled product. This price differential results from the belief that the quality of frozen meat is inferior to that of chilled meat. There is little in the published scientific literature to substantiate this view. Some publications show that freezing, frozen storage, and thawing have beneficial effects on the textural properties of pork loins. Taste panels have found the meat to be slightly tenderer with improved juiciness and an easier breakdown of particles, as evidenced by the samples being less cohesive, easier to chew, and releasing more moisture during mastication.

Meat which is frozen just after slaughter before the rigor process is complete and, after being thawed rapidly, contracts vigorously and exudes much drip. The resulting meat is very tough, and this toughness does not disappear with extended cooking. The phenomenon is called 'thaw rigor' or 'thaw shortening' and can be avoided by waiting some time prior to freezing. A long thawing process, in which the meat remains in the -5 to $-2\,^{\circ}C$ range for a long period, alleviates the effect. A similar phenomenon is observed in fish filleted prerigor and frozen individually.

Freezing rate The effect of variations in freezing rate on the tenderness of ground beef patties has been investigated using conditions which achieved freezing times to $-18\,^{\circ}C$ of 24, 48, 72, and 97 h. In one study, immediately after freezing, patties frozen in 24 h were less tough than those frozen in 96 h (Table 1). This difference was still found after 6 months storage. After 18 months, patties frozen in 48 h were tenderer than those frozen at faster or slower rates. Slight reductions in flavor and juiciness were also found in patties frozen in 96 h. Taste panel and sensory scores for the tenderness of ground beef patties frozen in air at $-43\,^{\circ}C$ were higher than for those frozen at $-20\,^{\circ}C$. Freezing rates were not detailed, and patties were stored for 2 weeks at $-30\,^{\circ}C$.

Table 1 Effect of freezing rate on tenderness of ground-beef patties (ranging from 1, extremely tough, to 8, extremely tender)

Storage time	*Freezing rate (hours to −18 °C)*				
	24	*48*	*72*	*96*	*SE*
Just before freezing	6.1[d]	5.6[f]	5.8[e]	6.0[d]	0.03
Just after freezing	5.1[d]	4.7[d,e]	4.8[d,e]	3.7[e]	0.25
6 months	4.9[d]	4.9[d]	4.2[e]	4.3[e]	0.10
18 months	4.5[e,f]	5.2[d]	4.7[d,e]	4.1[f]	0.09

[a, b, c, d, e, f]The same letters indicate no statistical differences.

From Berry BW and Leddy KF (1989) Effects of freezing rate, frozen storage temperature and storage time on tenderness values of beef patties. *Journal of Food Science* 54: 291–296.

Frozen storage Three factors during storage, storage temperature, degree of fluctuation in the storage temperature, and type of wrapping/packaging in which the food is stored, are commonly believed to have the most influence on frozen storage life.

Storage temperature has a marked effect on the behavior of ice crystals that could be detrimental to the ultimate quality of the food. It has been demonstrated that frozen tissue stored for 180 days at −20 °C has small ice crystals of irregular shape in the extracellular spaces. Storage for the same period at −3 °C results in the development of large rounded ice formations with a concomitant compression of the muscle fibers. It is thus desirable that frozen meat should be stored at a sufficiently low temperature to prevent the growth of ice crystals in the extracellular spaces. Such growth occurs if the temperature of the frozen tissue is allowed to rise above its eutectic point. (The eutectic point is the lowest freezing point of all combinations or constituents of the food.) However, quoted values for the eutectic point of meat range from 'probably just below −20 °C' to −52 °C.

Generally, fluctuating temperatures in storage are considered to be detrimental to the product. It has been found that temperature fluctuations produced during storage and transport of frozen food lead to activation of the recrystallization processes, causing enlargement of the ice crystals and diminishing advantages obtained in quick freezing of products. However, it has been reported also that repeated freeze–thaw cycles do not cause any essential change in the muscle ultrastructure and that several freeze–thaw cycles during a product's life cause only minor quality damage (in terms of juice loss) or possibly no damage at all. A slight but significant improvement has been found in samples that have been frozen and unfrozen several times when tested by a taste panel.

If the temperature cycles are so severe that the surface temperature of the food rises above 3 °C during the process, this could create a potential microbiological hazard.

Minor temperature fluctuations in a stored product are generally considered to be unimportant, especially if they are below −18 °C and are only of the magnitude of 1–2 °C. Well-packed products and those that are tightly packed in palletized cartons are also less likely to show quality loss. When ground pork and beef are stored under fluctuating conditions of between −18 and −23 °C for a 6-month period, the well-packed samples show no deterioration. However, poorly packed samples are severely affected by the temperature swings.

The proteins of fish differ from those of meat in their higher susceptibility to damage in frozen storage. Frozen storage results in an increasing tendency to lose moisture on thawing and an increased firmness, leading to toughness and dryness on cooking. These changes are the result of denaturation and cross-linking reactions predominantly of the myofibrillar proteins.

Drip Freezing, always tends to decrease the water-holding capacity and hence increases drip. When meat is frozen quickly, the water, released by the fibrils as the meat has gone into rigor, and that which is still held are frozen simultaneously. Consequently, there is no change in their relative positions or amounts. At slower freezing rates, however, the water balance is altered, the extracellular water freezing first. As freezing continues, the existing ice crystals grow at the expense of water from the intrafibrillar space. This can result in salt crystallization and pH changes which can cause protein denaturation. This has been well documented for frozen fish, which is much more susceptible to freezing damage than meat.

A number of scientific investigations have defined the effect of freezing rate on drip production in meat, demonstrating that the optimal conditions for freezing portioned meat are those that achieve freezing rates between 2 and 5 cm h^{-1} to −7 °C. 'Sow freezing' up to 0.39 cm h^{-1} resulted in a decreased solubility of myofibrillar proteins, increase in weight loss during, freezing thawing, and cooking, lower water-binding capacity, and tougher cooked meat. 'Very quickly frozen' meat (> 4.9 cm h^{-1}) had a somewhat lower

myofibrillar protein solubility, lower water-binding capacity, and somewhat tougher and drier texture. In these studies, samples were thawed after storage periods of 2–3 days at −20 °C, so the relationship between freezing rates and storage life was not investigated.

The results on freezing rate are scientifically very interesting, but in industrial practice, most meat is air-frozen in the form of large individual pieces or cartons containing smaller portions. In commercial situations, freezing rates of 0.5 cm h^{-1} in the deeper sections would be considered 'fast,' and there would be considerable variation in freezing time within the meat. The mutton samples frozen in these studies were much smaller (77.6 g in weight) than most commercial products. Even with such small samples, there was no significant difference in drip after 48 h between cryogenic freezing at −90 °C and a walk-in freezer operating at −21 °C (Table 2).

Experiments have been carried out on freezing samples similar in size to a domestic joint. There was no significant difference in drip loss from 700-g samples of pork l. dorsi frozen in air at −20 or −80 °C. At −20 °C, samples required approximately 6 h to pass from −1 to −6 °C, compared with half this time at −80 °C. Average drip losses were 3.7% at −20 °C and 5.2% at −80 °C. In commercial situations, freezing times typically range from tens of hours to a few days. Freezing rates are therefore outside the values that influence drip potential.

Drip loss in frozen storage has been shown to increase with storage time and increase in storage temperature. After approximately 42 and 63 days, drip from beef stored at −10 or −15 °C had reached 80 and 90%, respectively, of its maximum. At −25 °C, it required over 120 days to reach the 80% value.

Drip loss during thawing from ground beef patties was also found to increase with the length of time the patties had been in frozen storage. For higher-fat-content samples, drip loss increased from 1.8% in fresh samples to 12.5% after 20 weeks in storage. Higher drip losses in thawing were found for samples stored at −12.2 °C than lower temperatures. However, there was no difference between storage temperatures of −23.3 and −34.4 °C.

Flavor Changes

No marked changes in fruit flavor occur on freezing except after an extended process at elevated temperatures. Changes in flavor constituents of frozen fruit are generated during prolonged storage and increase as the storage temperature is increased. The first significant change in stored frozen fruits is a loss of fresh aroma. After this, some off-flavors may develop that are dependent on the type of fruit. Some fruits are known to be more stable than others, but the exact reason is not known.

Enzymes in fruits and vegetables are slowed down, but not destroyed, during freezing. If not inactivated, these enzymes can cause color and flavor changes as well as a loss of nutrients.

Enzymes in vegetables are inactivated by blanching, then rapidly cooled in ice water to prevent cooking. Blanching is essential for top-quality frozen vegetables.

Another type of chemical change that can take place in frozen products is the development of rancid, off-flavors. This can occur when fat, such as in meat, is exposed to air for a certain period of time.

Species has an influence on rancidity development. Investigations have been carried out on rancidity development in cooked and uncooked ground meat patties made from chicken, beef, or pork. In one study, the TBA (thiobarbituric acid) values in uncooked patties were higher in beef than in those made from either pork or chicken (Figure 1). Levels in chicken patties did not change appreciably during 150 days, storage at −20 °C. After 150 days, the TBA values in beef were approximately 7 times higher than those in chicken, and the values in pork were 4–5.5

Table 2 Drip loss (%) from 77.6-g samples of Mm. longissimus lumborum et thoracis frozen under different methods and thawed at 4 °C

Freezing conditions	*Freezing time to − 2.2 °C*	*Freezing rate (cm h^{-1})*	*Storage time at −20 °C*	
			48 h	*2.5 months*
Cryogenic − 90 °C	15 min	6.4	3.34^{a}	9.49^{a}
Cryogenic − 65 °C	22 min	4.4	$4.70^{a,b}$	9.72^{a}
Blast freezer − 21 °C	1.83 h	0.55	5.53^{b}	12.74^{b}
Walk-in freezer − 21 °C	1.88 h	0.53	$4.71^{a,b}$	13.18^{b}
Domestic freezer − 25 °C	1.96 h	0.51	5.26^{b}	11.72^{b}

From Sacks B, Casey NH, Boshof E and van Zyl H (1993) Influence of freezing method on thaw drip and protein loss of low-voltage electrically stimulated and non-stimulated sheeps' muscle. *Meat Science* 34: 235–243.

[a, b, c, d, e, f]The same letters indicate no statistical differences.

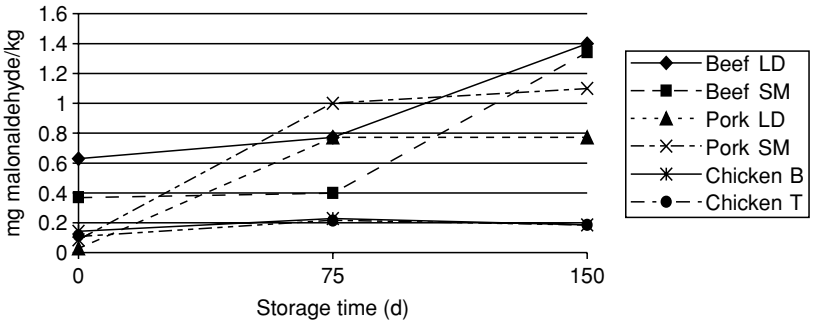

Figure 1 Raw meat TBA values during storage at −20 °C. From Rhee KS, Anderson LM, and Sams AR (1996) Lipid oxidation of beef, chicken and pork. *Journal of Food Science* 61: 8–12 with permission.

times as high as those in chicken. Cooked chicken patties had a higher lipid oxidation potential than cooked beef or pork patties.

Other studies have found a temperature and species interaction for pork and beef. In experiments, the practical storage life (PSL) and high-quality life of pork were superior to beef at −5 °C, but at −10 and −20 °C, the storage life of beef was much longer. In other trials where more than one species was compared, storage times were not fully presented, or direct comparisons were not possible.

It seems fair to conclude that most work indicates differences in frozen storage life between species. However, we currently lack the understanding to use data on the frozen storage life of one species under a given condition to predict the likely storage life of other species under that condition.

Animal-to-animal variation is believed to cause wide variations in storage times. Investigations in New Zealand have found that animal-to-animal variation can lead to storage-life differences as great as 50% in lamb but do not give any definite reason for this variation. Two trials were carried out in which lamb was stored at −5 °C. In the first trial, the lamb was judged rancid after 20 weeks, and in a duplicate trial using identical temperature, processing, and packaging conditions, the lamb was found to keep for 40 weeks. The only variation that could be determined was that different animals were used in the two trials.

Processing of meat prior to freezing generally results in a lengthened storage time. It has been found that the amount of TBA-reactive substances, in prerigor cooked boar, did not change during 3 months, storage at −20 °C. Heating prior to freezing can result in a 50% longer PSL for sausages and is thought to be due mainly to inactivation of some of the enzymes present or, possibly, inactivation of the catalytic activity of hemoglobin. However, the heating process may be critical, since muscles cooked at higher temperatures are most susceptible to oxidative changes during storage. Investigations have also found that pressure-cooking chicken breast at $0.36\,kg\,cm^{-2}$ produce higher TBA values than cooking at atmospheric pressure.

Smoking is generally advantageous due to the antioxidant properties of the smoke. It has been suggested that wood smoke may not only aid frozen storage, due to its antioxidant properties, but also mask off-flavors. Smoked broilers have been found to keep well for over a year without any major quality change. Smoked ham has also been found to store well for a year at −18 °C without any loss of flavor or weight.

Studies on chicken breast and salmon have shown that fast and slow freezing, at undefined rates, have no effect on the rate of development of lipid oxidation during storage, slightly higher levels being found in the fast frozen muscle. In studies on buffalo meat, plate-frozen meat was scored higher for color, texture, and overall quality than blast-frozen samples. Tests were carried out after 3 months' storage at −15 °C, and the freezing times, for 500- to 600-g pieces, from 4 to −30 °C were 280 and 140 min, respectively, for the plate- and blast-freezing operations. After 1 month's storage, the TBA values in plate frozen cuts were significantly higher than in those blast-frozen, but the difference had disappeared after 3 months.

Few would disagree that storage temperature influences frozen storage life and that, in general, the lower the temperature, the longer the PSL. However, there appear to be few papers with data from experiments on the PSL of food at different storage temperatures. Most of the studies published to date concern products that do not meet the lower-temperature longer-storage rule (normal stability).

It has been shown that rancidity in bacon is increased by a higher salt content and that the rates of chemical reactions are accelerated as the temperature is lowered when packed in permeable wrapping.

Studies have shown that free fatty acids accumulate more readily in bacon stored at −30 °C than in bacon stored at −12 °C. The highest rate of free fatty acid accumulation in bacon was found to be at −62 °C. A similar trend was found in trials with liver paste. Cured pork products are known to have an abnormal temperature profile between −5 and −60 °C and do not keep as well between −30 and −40 °C.

Pork chops, Vienna sausage, and pork sausages have been found to have straight-line quality-loss data at lower temperatures but were nonlinear at higher temperatures. Their storage trials were carried out at −40, −30, −24, −18, and −12 °C. Ground veal and pork were also found to have cubic curves at all temperatures.

Improved aroma scores have been found to be only moderately related to lower freezing temperatures but were not related to flavor. Aroma scores for minced beef improved during 6–12-month storage at −12.2, −17.8, or −23.3 °C, although a slight increase in rancidity also occurred. Temperature abuse of −6.7 °C for 48 h caused rancidity on the outside of the samples but did not affect the inner portion. After 24 weeks' storage at −12, −18, or −23 °C, no differences due to storage temperature were found in ground beef patties containing mechanically recovered meat. Ground beef patties stored at −7 °C were found to be tougher than those stored at −18 or −23 °C. Severe quality deterioration was found in patties stored for 6 months at −7 °C. Little difference was found between storage temperatures of −18 or −23 °C over an 18-month period.

Consumer panels are often not very sensitive to quality changes. When a consumer panel was used, it could not tell the difference between samples of lamb stored at −5 and −35 °C. A trained taste panel could differentiate between the two temperatures and scored the samples stored at −5 °C as being rancid.

In recent years, energy-conservation requirements have caused an increased interest in the possibility of using more efficient storage temperatures than have been used to date. Researchers have questioned the wisdom of storage below −20 °C and have asked whether there is any real economic advantage in very-low-temperature preservation. There is a growing realization that storage lives of several foods can be less dependent on temperature than previously thought, and since research has shown that meat and poultry often produce nonlinear time–temperature curves, there is probably an optimum storage temperature for a particular product. Improved packing and preservation of products can also increase storage life and may allow higher storage temperatures to be used. It has been suggested that foods should be split into high-, medium-, and low-stability categories, with −12 °C for products with high stability and −28 °C for products with low stability. Some believe that, based on storage at −18 °C, low-stability meats such as mechanically recovered meat should be stored for 8 months or less. Medium-stability meats such as pork and processed meats should be stored for between 8 and 15 months. High-stability meats, which include all meat and poultry except pork, could be stored for more than 15 months.

See also: **Freeze-drying**: The Basic Process; **Sensory Evaluation**: Texture; Taste; **Smoked Foods**: Principles; Applications of Smoking

Further Reading

Anon. (2001) *Rapid Cooling – Above and Below Zero*. Paris: International Institute of Refrigeration.

Anon. (1986) *Recommendations for the Processing and Handling of Frozen Foods*, 3rd edn. Paris: International Institute of Refrigeration.

Bald WB (ed.) (1991) *Food Freezing: Today and Tomorrow*. London: Springer-Verlag.

George RM (1993) Freezing processes used in the food industry. *Trends in Food Science & Technology* 4: 134–138.

Heap RD and Mansfield JE. The use of total loss refrigerants in transport of foodstuffs. *Australian Refrigeration, Air Conditioning and Heating* 37(2): 23–26.

Kennedy CJ (ed.) (2000) *Managing Frozen Foods*. Cambridge, UK: Woodhead.

Mallett CP (ed.) (1993) *Frozen Food Technology*. London: Blackie Academic & Professional.

Nutritional Value of Frozen Foods

J S Pai, University of Mumbai, Mumbai, India

Background

Freezing of foods has been practiced for a long time, and frozen foods have become very common today. All types of frozen foods are available on the market, from butter, meat, fish, and poultry to semiprocessed foods like frozen vegetables, orange juice concentrate, frozen cake and biscuit dough, French fries, ready-to-eat processed foods like icecream, and some items that simply need to be heated in the oven and are ready to eat. Several decades ago, the frozen food market consisted of only a few vegetables. In recent years, the range and quantity of frozen food have increased tremendously. Some

cooked dishes and meals are also being offered in frozen form.

Freezing has improved the convenience in the preparation of foods. Foods are available out of season and have a very similar quality to that of fresh foods. Frozen foods have been evaluated by experts who observed that when properly prepared and stored, frozen foods have a nutritional value not only as good as the fresh commodity but at times even higher than the commodity procured from the fresh food market and cooked at homes.

Freezing is used as a method of preservation, since most microorganisms do not grow at temperatures below 0 °C. Spoilage-causing microorganisms normally grow on foods and cause spoilage at ambient temperatures. When the foods are chilled to below 20 °C, the growth of most spoilage organisms slows down but does not stop. Enzymes present in foods and microbial enzymes also bring about changes, resulting in spoilage. These enzymes prefer milder temperatures. Their activity is reduced when the temperature is lowered. However, chilling temperatures above 0 °C may allow various microbial and enzymic activities to continue. There are many microorganisms, called 'psychrophiles', that grow quite rapidly at temperatures around 0 °C. Thus, although they have a longer shelf-life than fresh foods, chilled foods are more perishable than frozen foods. Whereas chilled foods may last for a few days to a couple of weeks, frozen foods that are frozen and stored properly may have a shelf-life of a few months to even a couple of years.

The spoilage organisms can be destroyed or stopped from growing and spoiling foods by different methods, including pasteurization, canning, low-temperature storage, including chilling and freezing, chemical preservatives, etc. Heating is the oldest method, but as well as killing the spoilage organisms, heat also degrades the components responsible for color, flavor, texture, etc., affecting sensory properties. It also destroys many nutrients. Chemical preservatives are undesirable, as people prefer natural, additive-free foods. Some preservatives have caused undesirable side-effects in some sensitive individuals. Drying removes the water from foods, and microbes and enzymes need water for their activity. Dried food after rehydration does not completely regain all its original sensory qualities, and there is also a significant loss of nutrients. Chilling and freezing have therefore gained acceptance with consumers as the preferred method of food preservation.

Although chilled and frozen foods may not spoil, their sensory quality and nutritive value may not remain unchanged. Some nutrients are degraded during the processing itself. Losses depend on the process design and rate of freezing. Some nutrients are lost during frozen storage, and here, also, the rate and amount of loss depend on the conditions of storage, including the storage temperature, packaging, type of food being stored, and its composition, etc. Same factors also affect the color, flavor, appearance, texture, etc.

Process of Freezing

Food is at ambient temperature when it is harvested and may undergo deteriorative changes unless it is chilled as soon as possible. If a food is to be stored unprocessed for long periods, it is normally cooled. Pretreatments such as washing, cutting, trimming, sorting, peeling, removal of pit, feathers, skin, bones, etc., removal of certain parts, and blanching are carried out prior to freezing. Washing, trimming, and peeling remove dirt and foreign material and reduce the microbial load. Blanching is a process of heating in water or steam and achieves several objectives in different products. It fixes color in green vegetables, wilts leafy vegetables like spinach, so it can be packed tightly into cartons, and removes undesirable odors associated with some vegetables and fruits. It also deactivates enzymes naturally present in many foods, which, if not removed, cause odor formation during storage, and it removes dissolved gases from food tissues. Finally, it reduces the microbial load, and in the case of certain harmful pathogenic organisms, it might be the only way to control such organisms.

These processes affect the nutrient contents as washing leaches out some water-soluble nutrients, and cutting, trimming, peeling, and pitting, etc. remove parts of the foods that may contain some nutrients. Blanching not only degrades some temperature-sensitive nutrients but also leaches out water-soluble nutrients if hot water is used. Blanching also disrupts the cells and tissues, so the mixing of different components of foods takes place. Some of these components react and degrade each other.

Freezing Damage

The freezing process also causes changes in the tissues of food of animal and plant origin. Some irreversible changes are brought about as a result of freezing, so even after thawing, some changes in the food are evident. When certain commodities, especially those of tropical origin, are chilled, changes take place in membrane lipids and in the structure of proteins. This affects the integrity of the cells, and nutrient loss results from mixing with degrading enzymes or other chemicals.

When food is frozen, the water in both intracellular and extracellular solutions forms ice crystals, increasing the solute concentration and affecting many nutrients. Proteins become denatured as a result of the loss of water from the protein structures as well as the increase in solute concentration. Some vitamins become degraded, as a result of the change in pH and the increase in concentration of degrading chemicals.

Ice crystals can cause mechanical damage by causing excessive stress. Water freezes first from the extracellular area, and ice crystals start forming there, increasing the osmotic pressure outside the cells. This causes the water from inside the cells to move outside, leading to cellular collapse. The loss of cellular integrity not only allows various chemicals to mix and react with each other resulting in degradation of nutrients, but also causes leakage of watery solution as drip loss or exudates when frozen food is thawed. This may lead to further nutrient loss. (*See* **Freezing**: Principles.)

Food Preservation and Nutritive Value

People believe that foods purchased raw and cooked at home are more nutritious than foods prepared in factories. It is generally considered that the housewife or cook at home has a greater expertise and that foods cooked at home are nutritionally superior to factory products, as large-scale cooking may not preserve the nutrients adequately. Although this was probably true about a century ago, today, factory processes can be designed scientifically to preserve nutrients and taste by controlling the processes involved. This is possible because of the advances in scientific knowledge about foods and the better design of the equipment used for processing. The degree of nutrient loss depends on the process used: freezing is by for the least damaging process. (*See* **Preservation of Food**.)

Foods are preserved to feed large populations, especially in urban areas, as well as to make foods available out of season. Traditional preservation processes have included sun drying, smoking, and salting. Later, other processes such as canning, pasteurization, sterilization, and, more recently, chilling, freezing, and freeze-drying were introduced. In the ancient times, knowledge of nutrition was limited and also the motivation of profit overrode concerns for public health. However, gradually, the knowledge of nutrients has evolved, and several food laws have been introduced to ensure the safety and high nutritive value of food. Factory food processed by food technologists today using newer techniques and more efficient processes, is highly nutritious and sometimes even superior to home-cooked food.

Raw food is often associated with freshness. It looks fresh, but it may not have been harvested recently and may have been stored at ambient temperature for several hours or days. This is called 'market fresh', as opposed to 'garden fresh', which means that a food has been harvested recently. After harvest, when a food has been kept at room temperature (which sometimes may be up to 40–50 °C in the summer), there are many changes to nutritive values, including vitamin contents. Many nutrients like vitamin C and folate are degraded quite rapidly in a matter of hours or days after harvest. Green leafy vegetables in warm climates lose 5–18% of their vitamin C 2 h after harvest and as much as 90% within 24 h. Garden peas can lose 7% of their total vitamin C each day, so after a week, they may lose up to 50% vitamin C.

Processing plants, however, have access to freshly harvested foods, and within hours of harvest, they may either process them or precool them and store them until processing, resulting in less nutrient loss. Thus, processed foods may have a higher nutrient content in the raw material than those in fresh food markets. Even after cooking, processed food may often be more nutritious than market fresh food cooked at home, particularly with frozen foods.

There are several different mechanisms of nutrient loss during food processing. Nutrient losses can be due to physical separation, as in peeling, trimming, etc., leaching during blanching, or chemical degradation. When water is involved in the process, water-soluble nutrients like vitamin C, folate, B vitamins, etc. and, to lesser extent, soluble carbohydrates, proteins, and mineral are lost because of leaching. Boiling or steeping in water is an example. Heat damages nutrients that are sensitive to higher temperatures like thiamin, carotene, vitamin A, and vitamin C. However, when food processing includes heat treatment, the food is partly or completely cooked, so it does not have to be cooked again, and brief heating to warm it is enough. Thus, the loss of nutrients incurred in the factory is instead of, and not in addition to, that incurred at home.

Nutritive Value of Frozen Fruits and Vegetables

Fruits and vegetables are important dietary sources of vitamin C, folate, minerals, carotenoids, and anthocyanins, especially in colored fruits, but poor sources of B vitamins and macronutrients. Some years ago, a study was undertaken to evaluate frozen foods from 150 factories with respect to vitamins and minerals. The vitamin contents are listed in Table 1.

Table 1 Vitamin contents of some frozen vegetables (mg/100 g)

Frozen vegetable	*Vitamin*	*Content (approx.)*
Asparagus	Thiamin	0.10–0.25
Broccoli	Carotene	1–3
Brussels sprouts	Ascorbic acid	70–110
Lima beans	Niacin	1–2
Spinach	Ascorbic acid	10–50

The ranges are quite wide, and, besides experimental error, this may be because original nutrient contents vary in fruits and vegetable owing to varietal differences and cultivation and harvest practices. The freezing process includes precooling, washing and other pretreatments, blanching and cooling, freezing, and frozen storage with respect to temperature and time. All these parameters affect the nutrient contents.

Prefeezing storage at warm or ambient conditions causes substantial losses. Raspberries left to stand at ambient temperatures for 24–48 h prior to freezing have been shown to lose 17 and 30% of their vitamin C. Strawberries stored at 0 °C or lower for 2–4 days showed only slight losses in vitamin C.

Freezing itself has very little effect on nutrients, but most studies have indicated losses during blanching and cooling prior to freezing, during frozen storage, and during thawing. Blanching is carried out before freezing to deactivate the enzymes that might be active and cause deteriorative changes in nutrients and sensory quality. Blanching may be carried out in hot water or steam or by using microwave heating.

The losses include water-soluble nutrients such as vitamins C, B complex, and folate, and much less important losses of minerals and a small fraction of proteins and carbohydrates have been reported. Figure 1 shows the losses occurring mostly during blanching and very little during freezing. Not only does water leach out the vitamins, but heat during the blanching causes losses of vitamin C and thiamin, which are heat-sensitive. The type of food, enzyme to be inactivated, size of pieces, etc. determine the length of treatment, which in turn affects the vitamin losses.

Losses depend on the area-to-volume ratio of the material being blanched. Almost half the total vitamin C is lost from vegetables like peas, sliced beans, broccoli, spinach, and diced carrots because of the large surface area-to-volume ratio. Smaller losses of only about 30% occur in whole beans, potatoes, and sprouts, which have smaller ratios. Losses are mostly because of leaching out of water-soluble vitamins. Microwave blanching does not involve water, so much less losses are reported. Microwaving is also more rapid, so heating losses are less. Green beans, when blanched, lose 9% by microwaving, 14% by hot water, and 18% by steam. Some studies have reported almost the same losses by the three methods. Blanching is followed by cooling before freezing, and if water is used, the losses are higher compared with air cooling (see Figure 2).

Although blanching causes losses of vitamins, when vegetables are not blanched, there are greater losses during frozen storage. Without blanching, green beans lose 35% and spinach loses 54% vitamin C during frozen storage at −19 °C for 9 months, whereas blanched vegetables lose only 10 and 24%, respectively. The vitamin B_1 and B_2 losses of unblanched green beans during 1 year's storage at −20 °C are 74 and 39%, whereas after blanching, the losses are 22 and 3% only.

Nutrient losses in fruits and vegetables are negligible during the freezing process. Studies on various vegetables like asparagus, green beans, lime beans, broccoli, sprouts, cauliflower, kale, peas, potatoes, and spinach have shown that the losses of vitamins

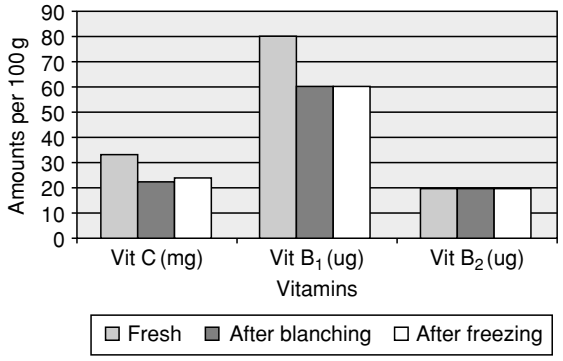

Figure 1 Vitamin losses from green beans.

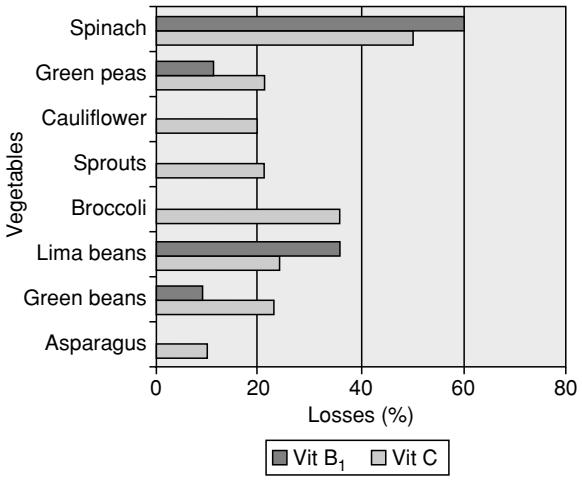

Figure 2 Vitamin losses during blanching.

C, B_1, B_2, niacin, and carotene are either nil or negligible.

Frozen storage, however, causes some losses, and these are affected by the temperature and packaging method. Losses at the usual recommended temperature of −18 °C are quite low. More rapid losses have been reported in vitamin C when frozen peas are stored at higher temperatures. Other vegetables show some losses in vitamins C, B_1, B_2, carotene, etc. when stored at − 18 °C, but besides vitamin C, other vitamins are less sensitive to storage at higher frozen temperatures (see Figure 3).

Losses of vitamin C from many fruits can be large, particularly if fruits are not packed in syrup. Raspberries packed in syrup lose only small amounts of vitamin C. Losses of vitamin C from peaches and strawberries are generally moderate during frozen storage, and this depends also on the oxygen barrier properties of the package. Metal cans are much better packages in that respect compared with composites with metal ends. Citrus juice concentrates are reasonably stable with respect to vitamin C during frozen storage for 9–12 months at − 18 °C. Vitamin C degradation of frozen boysenberries, strawberries, and peaches is greater when they are stored at higher temperatures at − 12, − 7, and − 1 °C, rather than − 18 °C. Citrus juices from grapefruit, orange, and tangerine are more stable at different temperatures. The stability of citrus juices may be due to their low pH and low oxygen content.

Thawing conditions have little effect on vitamin losses, according to the few studies that have been carried out on vegetables and fruits. Losses of vitamin C from citrus juice concentrates during thawing are insignificant. Also, peaches, apricots, nectarines, and fruit salad packed in syrup lose little vitamin C during various thawing procedures of different times ranging from 20 min to 19 h. However, holding the fruits after thawing at ambient conditions will incur losses of vitamin C. Raspberries and peaches after thawing lost up to 15% vitamin in a period from 2 h to 1 day (see Table 2).

Losses of vitamin C in fruits and vegetables during the entire freezing process have been studied. In some vegetables, 10–65% of the original vitamin C is lost. When fruits are protected by syrup and properly packaged to keep oxygen out, the losses are less, as in the case of strawberries, apricots, and peaches. Without sugar syrup, they may lose up to 65–85%, whereas with syrup and good packaging, their losses can be reduced to about 15–30% in up to 9 months of storage at − 18 °C.

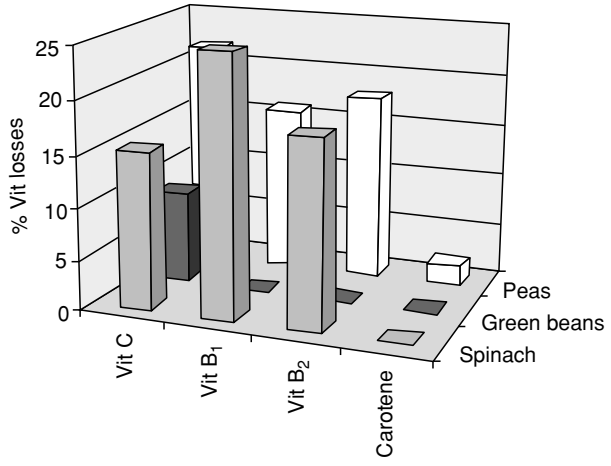

Figure 3 Vitamin losses in vegetables stored for 6 months at − 18 °C.

Nutritive Value of Frozen Animal Foods

Meat is an important source of many vitamins and minerals such as vitamins B_1, B_{12}, niacin, and iron. Other vitamins are available, like B_2 from offals and vitamin A from liver. There is no blanching in meat processing, but an important treatment prior to freezing of meat is aging. In one study, beef was aged for 21 days at 1 °C, resulting in no loss of thiamin and riboflavin but a 35% loss of total niacin. Another study on beef reported no significant losses of vitamin B_6 and pantothenic acid. When pork roast was aged for 7 days at − 1 °C, the amounts of thiamin, riboflavin, pantothenic acid, and nicotinic acid were not affected consistently. Short-term aging with a low nonfreezing temperature did not destroy vitamins significantly. Phosphate treatment of fish did not affect the vitamin and mineral contents, except that the sodium tripolyphosphate treatment increased the sodium content by three to seven times that in untreated samples.

B vitamin losses in various animal meat products such as beef steak, beef liver slices, beef stew, Chicken à la king, Shrimp Newburg, oysters, and pork chops have been investigated during freezing. No significant

Table 2 Losses of Vitamin C in different frozen vegetables stored for 6–12 months at −18°C

Frozen vegetable	*Vitamin in fresh vegetable (mg/100 g approx.)*	*Percent loss after storage (approx.)*
Asparagus	35	10
Green beans	20	45
Lima beans	30	50
Broccoli	110	50
Cauliflower	80	50
Peas	30	45
Spinach	50	65

Table 3 Vitamin B_1 losses from thawing

Frozen food	*Percent thiamin loss (approx.)*
Beef liver, slices	20
Beef stew	1–10
Chicken a la king	2–10
Shrimp Newburg	4–10

losses of thiamin, riboflavin, niacin, pantothenic acid, and pyridoxine have been reported in these products (see Table 3). Some losses have been indicated in pork wherein both freezing and thawing losses were combined.

Losses have been monitored during frozen storage of several meat products at −18 °C for 6 months. Losses have been shown mostly in pyridoxine (18–59%) in beef steak, pork chops and roast, and oysters with minor losses of niacin and pantothenic acid. Oyster incurred significant losses of most B complex vitamins.

Most losses from frozen meats occur when they are improperly thawed. The amount of drip loss or thaw exudates depends on several factors such as time taken to thaw, pH of the meat, state of rigor, ageing prior to freezing, size of cuts, and freezing rate. The amount can range from 1 to 10%. The water-soluble vitamins are lost along with the drip or exudates, and their retention is better, when meat is thawed at room temperature or at 0 °C, than in warm ovens (see Table 3).

The effects of various thawing methods have been investigated. Losses of B vitamins from unpackaged beef steak are generally greater when water is used for thawing, rather than air. Thiamin losses are lowest when a microwave is used for thawing, slightly greater with infrared heating, and greatest with boiling with water. In packaged products, the method of thawing is less significant with respect to losses of B vitamins. Nutrients lost in thaw exudates have been estimated for beef and pork (see Figure 4).

The total losses during the entire freezing process have been evaluated. The meat was aged, frozen, stored for several months at −18 °C to −23 °C before thawing. The losses are shown in Figure 5.

In a more recent study, when rib steaks were aged for 5 days, then frozen and stored at −18 °C and thawed at room temperature for 15 h, the losses were found to be about 10% for each of B_1, B_2, B_6, niacin, and folate, whereas about 30% of pantothenate was lost. Almost all of these losses could be traced to drip losses, and it might be good sense to include drip or thaw juices with meat before cooking, so as to recover most of the lost nutrients. There will of course be a small loss due to heating.

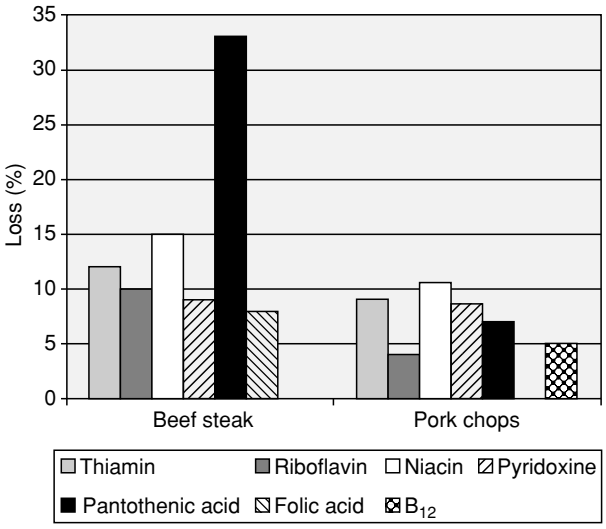

Figure 4 Losses in thaw exudates.

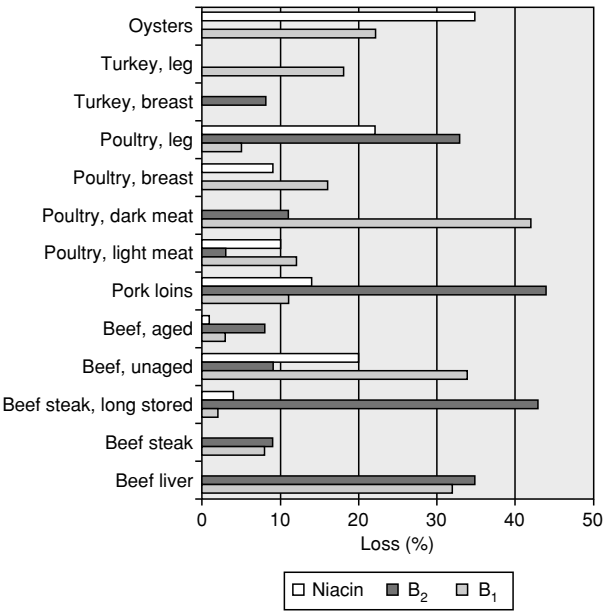

Figure 5 Vitamin losses in the entire process.

Nutritive Value of Frozen Dairy Products

Some products like butter, margarine, and ice creams are stored frozen. Butter loses only a small amount of vitamin D when stored at −10 °C for 2 years, whereas butter and margarine lose about 17% of vitamin A. Ice cream is frozen at lower temperatures. When frozen and stored at −23 °C for 7 months, it loses 5% of riboflavin, 16% of carotene and all of the vitamin C (there is very little vitamin C in ice cream

anyway). When frozen milk is stored for 19 weeks at −14 °C, the losses in biotin and nicotinic acid are nil.

Comparison of Nutritive Value with Other Methods

There have been many types of processed foods available commercially, and there has always been an interest in evaluating these for the losses of nutrients. This is a difficult task, as foods have different varieties as well as degrees of maturity suitable for a particular method of processing. Thus, if the same raw material is used for all the processes and unprocessed for control, it would be an unfair comparison. Also, raw materials of different degrees of maturity would have different amounts of nutrients, and any comparisons would be meaningless. Studies have been undertaken to determine the nutrient losses resulting from different processes. Green peas have been processed by different methods and analyzed for their vitamin C content, the most vulnerable of all the vitamins (see Table 4).

Cooking losses from fresh food have been compared with those from frozen foods, and the losses have been shown to be comparable. Many commercial samples have been purchased and cooked and compared with fresh foods cooked. The amounts of vitamins B_1, B_6, and B_{12} show no difference, whereas some foods such as beef stew, fish cakes, codfish, and Brussels sprouts show higher values for frozen samples. Also, the amount of vitamin C is higher in frozen sprouts. Some experiments have been carried out with cooked meals that have been stored frozen. These have been found to have a higher vitamin C content than conventionally prepared fresh food meals roughly the some amounts of other vitamins like B_1 and B_2, and a slightly higher retention of available lysine.

Frozen Prepared Meals

Frozen prepared meals are either complete meals or portions of meals that are precooked, packed in trays, and frozen for retail. Their popularity is due to the diversity in menu and convenience in preparation. Typical packs may include meat dishes, vegetables, pasta products, sauces, and desserts. Meats are often a major component, which is cooked or fried after treatments like marinating and tenderization. Heat treatment removes pathogens and improves the flavor. While chicken may be completely cooked, delicate foods like fish are heated for a short time, allowing the consumer to complete the cooking while reheating the product. Sometimes, heating may be carried out after packing meat and applying a vacuum. Vegetables are either procured frozen, or raw vegetables are frozen on site after cooking. Other items are normally prepared fresh and included in the meal and frozen.

Different items are assembled on trays composed of aluminum foil, paperboard, or plastic. These trays are embossed to compartmentalize each item. The package is made so that it can be conveniently reheated in either an electric oven or a microwave oven. An alternative is boil-in-bag meals, which have flexible bags in which meals are packed, and the frozen packaged is simply immersed in hot water to heat the contents, which can be poured on to a plate after cutting open the bag.

Each item may be stored frozen before being added to the tray, or the freshly prepared items may be deposited on the tray and then frozen. These products are fed into the depositing machines, which deposit them on to the trays, and these filled trays are then conveyed to the freezer. A belt freezer is commonly used, in which trays are placed on a wire mesh belt, which takes the product through a freezing chamber operated at −40 °C. Freezing is accomplished in a short time, usually about 90 min. Alternatively, a spiral or plate freezer can be used. After freezing, the product is packed into cartons and stored at low temperatures during storage, transport, and marketing.

Freezing *per se* does not damage nutrients to a great extent, but the rate of freezing and the different chemical substances present in the environment of nutrients may affect the retention of nutrients during freezing, frozen storage, and reheating. Nutrients are preserved to a greater extent if individual components of a meal are frozen optimally and separately and then assembled on to a tray. However, some or all of the components are assembled and then frozen because of the ease of production and economics. When different items are to be frozen simultaneously, the ideal conditions are difficult to maintain for each. Thus, losses may occur to a greater extent. Secondly, reheating also affects nutrient retention with the method used (such as oven or microwave heating)

Table 4 Ascorbic acid losses during complete processing by different methods in green peas

Process of preservation	*Percent loss after cooking (approx.)*	*Cooked product content (mg/100 g)*
Dried in air	55	11–12
Canned	60	8–11
Freeze dried	35	13–18
Frozen	40	12–16
Fresh peas	30	16–19

being an important pronounced factor besides the storage conditions. The microwave oven leads to fewer losses, as heating is rapid. Using a higher freezer temperature leads to greater losses of vitamins such as vitamin C.

There are many manufacturers of frozen prepared meals and many different products available. Although seemingly similar, many have different ingredients and processing procedures. The inherent differences make it difficult to generalize the nutrient losses in such products. In one study, it has been shown that a 300-gram frozen chicken dinner supplies half the RDA of protein, more than 100% of the RDA of vitamin A, and substantial amounts of thiamin, riboflavin, vitamin C, niacin, iron, and calcium, whereas a 300-gram beef dinner provides the following RDAs: 50% protein, 25% P, 60% Fe, 14% vitamin A, 25% thiamin, 30% riboflavin, and 95% niacin. A different recipe and processing method might give different results which will be further affected by the storage time, temperature, and reheating procedure. Thus, it is very difficult to state the nutrient contents of these foods and their losses during processing and just prior to consumption.

Conclusions

Nutrient losses in the freezing process are primarily due to blanching in fruits and vegetables, frozen storage in all the frozen foods, and during thawing if water is used with unpackaged frozen food. These losses can be minimized if the frozen storage is maintained at the recommended temperature of $-18\,^{\circ}C$, and if the proper packaging method and material is used to prevent the food coming into contact with oxygen. Blanching can be carried out without water, e.g. by using microwave heating. Thawing should be carried out without contact with water or without removing the packaging material, preferably at room or refrigerated temperature. There will be some drip or thaw exudate losses, which can be reduced if the thaw juices are used with the food during cooking. Modern technology has allowed raw material of the highest quality to be processed under ideal conditions, thus incurring minimum losses of nutrients. Such frozen foods, upon cooking yield highly nutritious products often containing nutrients in quantities at least as much as that of home-cooked unprocessed food.

See also: **Canning**: Principles; **Chilled Storage**: Principles; **Freeze-drying**: The Basic Process; Structural and Flavor (Flavour) Changes; **Freezing**: Principles; Storage of Frozen Foods; Structural and Flavor (Flavour) Changes; **Pasteurization**: Principles; **Preservation of Food**; **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage; Fungi in Food – An Overview; Molds in Spoilage; Yeasts in Spoilage; **Sterilization of Foods**

Further Reading

Bald WB (ed.) (1991) *Food Freezing: Today & Tomorrow*. London: Springer.

Erickson MC and Hung Y-C (1997) *Quality in Frozen Food*. Guithersburg, MD: Aspen.

Jen JJ (ed.) (1989) *Quality Factors of Fruits & Vegetables: Chemistry & Technology*. ACS Symposium Series 405. Washington, DC: American Chemistry Society.

Jeremiah LE (ed.) (1996) *Freezing Effects on Food Quality*. New York: Marcel Dekker.

Karmas E and Harris RS (eds) (1988) *Nutritional Evaluation of Food Processing*, 3rd edn. New York: Van Nostrand Rheinhold.

Kennedy C (ed.) (2000) *Managing Frozen Foods*. Cambridge, UK: Woodhead.

Lawrie RA (1998) *Lawrie's Meat Science*. Cambridge, UK: Woodhead.

Luh BS and Woodroof JG (eds) (1975) *Commercial Vegetable Processing*. Westport, CT: AVI.

Mallett CP (ed.) (1994) *Frozen Food Technology*. Glasgow: Blackie, Academic & Professional.

Salunkhe DK and Kadam SS (eds) (1998) *Handbook of Vegetable Science & Technology*. New York: Marcel Dekker.

Woodroof JG and Luh BS (eds) (1975) *Commercial Fruit Processing*. Westport, CT: AVI.

http://www.easyhomemeals.com

http://www.freweb.org

Fromage Frais *See* **Cheeses**: Types of Cheese; Starter Cultures Employed in Cheese-making; Chemistry and Microbiology of Maturation; Manufacture of Extra-hard Cheeses; Manufacture of Hard and Semi-hard Varieties of Cheese; Cheeses with 'Eyes'; Soft and Special Varieties; White Brined Varieties; Quarg and Fromage Frais; Processed Cheese; Dietary Importance; Mold-ripened Cheeses: Stilton and Related Varieties; Surface Mold-ripened Cheese Varieties

FRUCTOSE

J M Johnson and F D Conforti, Virginia Polytechnic Institute and State University, Blacksburg, USA

Introduction

Fructose as a commercial sugar has been evolving for several years in its application to food processing. Fructose is a monosaccharide and is a natural constituent of many foods. This chapter will discuss fructose in both its syrup and crystalline form and its application in food products. The physical, chemical, and sensory properties will be discussed.

Sources

Fructose is a monosaccharide. Fructose bonded with glucose, another monosaccharide, forms sucrose, or table sugar. Fructose also occurs naturally in abundance in fruits (Table 1) and in lesser amounts in tuberous vegetables such as onions and potatoes. These sources alone contribute some 40–60% of an individual's total fructose intake. However, the major source of fructose as an ingredient in food is from the hydrolyzation of starch to glucose, which is then converted to fructose. (*See* **Carbohydrates**: Classification and Properties.)

Fruits are a rich source of mono- and disaccharides. Dates contain up to 48.5% sucrose, and dried figs contain a mixture of 30.9% fructose and 42.0% glucose. The sucrose content of most fruit and fruit juices is low, though some varieties of melons, peaches, pineapple, and tangerine contain 6–9% sucrose, and mango contains 11.6% sucrose. Reducing sugars (primarily a mixture of fructose and glucose) are the main soluble carbohydrate of most fruits and account for 70% of seedless raisins. Vegetables contain substantially less fructose and glucose than fruits, and the only significant source of sucrose is sugar beets.

Table 1 Free fructose in selected fruits (% fresh weight basis)

Food	*D-Fructose*
Apples, raw, unpeeled	7.6
Apricots, raw	0.7
Apricots, dried	12.2
Bananas, raw	2.7
Blueberries, raw	3.6
Cherries, raw, sweet	6.2
Cherries, raw, sour	3.3
Figs, raw	2.8
Figs, dried	26.0
Grapes, raw, American	6.9
Grapes, raw, European	7.6
Prunes, dried	14.8
Raisins, dried	33.8
Peaches, raw	1.3

Adapted from: The Food Resource, Oregon State University. *http://osu.orst.edu/food-resource/sugar/com2.htm*.

In the late 19th century corn or potato starch was hydrolyzed with dilute acid to yield glucose and dextrins for commercial purposes. In the 1940s, cornstarch was the primary choice for the production of glucose and the introduction of enzyme technology for hydrolysis reactions contributed to the development of glucose syrups to fructose syrups of specified glucose content. The conversion of glucose syrups to fructose syrups by immobilized enzyme technology was introduced in the 1960s.

The major source of fructose as a food ingredient is high-fructose corn syrup. The production of high-fructose corn sweetener requires the following manufacturing steps: (1) wet milling corn to extract the starch; (2) saccharification and liquefaction to hydrolyze polymer starch to monomer dextrose; (3) isomerization to convert dextrose to fructose; and (4) fractionation to enrich the concentration of fructose in the isomerization product stream. An additional step of crystallization is required for the production of dry, crystalline fructose.

Fructose Crystallization

The solubility of fructose at 25 °C is 4 g g^{-1} H_2O. Fructose has the highest solubility of most, if not all, sugars and sugar alcohols. For this reason fructose is difficult to crystallize from aqueous solution. Several processes have been patented for the production of a fructose product in its most stable, pure crystalline form. These procedures include crystallization from methanol, ethanol, and water. The most successful process in use today is based on the crystallization of fructose from aqueous solution. The procedures utilize various combinations of concentration under atmospheric or reduced pressure conditions and seeding to initiate crystallization, followed by a cooling step to allow for crystal growth. Batch and continuous crystallization processes have been devised that incorporate preprogrammed concentrating, cooling, warming, and recooling cycles. Crystalline fructose products are produced in a variety of forms: powdered (90% through a US 200-mesh screen) and products with an average particle size of

300 and 450 μm. Crystalline fructose syrup is similar to physical properties of traditional high-fructose syrups, but has greater clarity, sweetness, and improved color stability. Both crystalline fructose and crystalline fructose syrup are used in many food products.

Isolation and Purification

The manufacture of high-fructose sweetener (HFS) requires the complete depolymerization of starch to its constituent monosaccharide, dextrose. Saccharification and liquefaction are names for the hydrolysis reactions used to accomplish this. Starch from wet milling is prepared for hydrolysis by a process called jetting: the starch slurry is pumped under high pressure through a small orifice with simultaneous injection of steam. Hydrolysis of the starch polymer to oligomers of intermediate molecular weight is initiated with dilute mineral acid or the enzymes α- and β-amylase. Mineral acid and α-amylase make random breaks in the starch backbone, producing soluble oligosaccharides and relatively small amounts of low- and medium-molecular-weight saccharides. β-amylase releases disaccharides units of dextrose (maltose) as it hydrolyzes starch. Glucoamylase completes the enzymatic hydrolysis of di- and oligomeric products of amylase by breaking the α-1, 4 and α-1, 6 bonds that join consecutive dextrose units. Isomerization is the next major refining step. Glucose isomerase is able to catalyze the conversion of dextrose to fructose without the need for cofactor regeneration. This enzyme remains one of the largest industrial uses for immobilization technology worldwide. The amount of fructose enzymatically produced from dextrose at 60 °C is restricted by an equilibrium constant of about 1. Theoretically, the highest fructose yield possible from the 94% dextrose feed stream is 47% at equilibrium. Manufacturing plants typically settle for yields of less than 42% because of the amount of enzyme and reactor time required to achieve equilibrium fructose concentrations. For many applications, the sweetness of glucose–fructose syrup has to be increased to the equivalent of sucrose. This means raising the fraction of fructose in the syrup by chromatographic enrichment to 55% (HFS-55). An attractive alternative method of preparing HFS has now been developed. Fructose is separated from a fructose–glucose mixture using a microporous polypropylene-supported liquid membrane containing a boronic acid carrier (2% wt; flux (10^{-8} mol m^{-2} s^{-1}), −1), glucose 2.9, fructose 55). Furthermore, adding immobilized glucose isomerase to the glucose solution in the source compartment of the liquid membrane transport cell allows glucose isomerization and fructose enrichment to be carried out in a single operation. This gives a receiving phase mixture that is more than 80% fructose.

Chemistry of Fructose

Fructose is a monosaccharide which is found widely in nature both free and combined forms. Unlike glucose, and the related sugars, mannose and galactose, fructose contains a ketone carbonyl. Five hydroxyl groups are also present and the keto group is on the 2-carbon. Fructose is a structional and functional isomer of glucose. Fructose shows a strong levorotation. The equilibrium value of the specific rotation is $[\alpha]_D^{20} = -92°$. A long-used name for this sugar is levulose, corresponding to the use of dextrose for glucose. Fructose shows mutarotation by ring formation through the addition of one of the –OH groups across the ketone carbonyl. The result is the hemiketal and there are two forms, depending on which side of the carbonyl group is used for the addition. However, the system is more complicated than glucose because both five- and six-membered rings are involved, depending whether the –OH group of the five-carbon or the six-carbon reacts in the ring-forming addition to carbonyl.

Fructose is hydroscopic and does not crystallize easily from solution. At 20 °C, approximately 78 g of fructose is soluble in 100 ml of water compared to 65 g sucrose and 20 g lactose. The crystalline hydrate form is less soluble than the anhydrous crystalline form and has very low solubility in nonpolar solvents. Fructose is a reducing sugar and is a reactive component of nonenzymatic browning (Maillard browning). (*See* **Browning**: Nonenzymatic.)

Analysis

In food composition tables, the carbohydrate content is usually given as total carbohydrates by difference, that is, the percentage of water, protein, fat, and ash subtracted from 100. Another widely used term is nitrogen-free extract, calculated as components other than water, nitrogenous compounds, crude fiber, crude fat, and minerals. The characteristics of fructose as a ketose and reducing sugar are utilized for analysis methodology.

A qualitative method for ketose determination is when a purple-pink color develops when 100 μl of a ketose solution is mixed with 0.5 ml of a phenol–acetone–boric acid reagent (5% phenol, 2% acetone, 4% boric acid) and treated with 1.4 ml of 96% sulfuric acid. Absorbency is measured at 568 nm after 60 min at 37 °C. This method allows for the determination of 3–50 nmol of D-fructose with a coefficient

variation of 2.8–7.8%. Carbohydrates other than ketoses do not interfere.

In alkaline solution reducing sugars, which contain an aldehyde or keto group, can reduce copper, silver, bismuth, and mercury salts to compounds of lower valence or to a metallic state. The best known reagent, based on the reduction of copper, is Fehling's solution (two solutions: (1) cupric sulfate; (2) sodium potassium tartrate and sodium hydroxide). Depending on the concentration of sugars in a solution, heating in the presence of Fehling's solution gives a yellowish orange to red solution or precipitate.

The available assay methods for mono- and oligosaccharides include chemical, colorimetric, chromatographic, electrophoretic, optical, and chemical procedures. Today, more and more assay techniques involve preliminary separation by chromatographic and electrophoretic techniques prior to actual assay by classic chemical procedures or colorimetric tests. Microbiological assays have found relatively little application, while the use of enzymes as aids in sugar analysis or in actual assays is gaining in popularity with the commercial availability of pure, selective, and stable preparations.

The separation of sugar by paper chromatography has been developed into the versatile form of qualitative and quantitative carbohydrate microanalysis. Qualitative paper chromatography is the simplest method to distinguish between various sugars in a food. The development of the chromatography is done by the application of compound sprays to identify spots. Separation of 1% sugar solutions containing a maximum of 60 μm of an individual sugar gives best results. Resolution is more rapid with phenol–water or colloidin–water systems. In a phenol-containing system fructose moves faster than xylose.

In enzymatic methods, selective cleavage to monosaccharides and/or enzymatic assay of the monosaccharide is specific and widely used. The use of amylase and amyloglucosidase is an example of the first type; determination of glucose with glucose oxidase is an example of the second type. Microorganisms and enzymes can also be used in the pretreatment of substrate prior to assay by chemical or physical methods. For example, hydrolysis of sucrose by invertase is much more specific than acid hydrolysis, and invertase is used in various procedures requiring sucrose inversion.

Sensory Properties

Fructose is perceived to be sweeter relative to sucrose. β-D-fructose in solution has been rated with values from 100 to 175 compared to a sucrose sweetness value of 100. In the crystalline form, β-D-fructose was rated 180 relative to crystalline sucrose at 100. The most common theories of why some molecules vary in perceived sweetness is based on the hydrogen bonding and geometrical shape of the molecule. Intramolecular hydrogen bonding enhances sweetness, as well as intermolecular bonding with a receptor site that has a complementary configuration to the molecule to allow for efficient bonding. Typically, a proton donor site, designated as AH, is located approximately 0.3 nm from an electronegative site, designated as B. The fructose molecule meets these requirements to elicit the sweet taste. The AH is designated to be the anomeric hydroxyl group and B is designated as the oxygen atom of the primary alcohol group. A third site capable of hydrophobic bonding designated as γ is the methylene carbon atom.

The sweetness of fructose has been reported to decrease with time or age of the solution and is accounted for by mutarotational equilibrium with other forms which are not perceived as sweet. Increased temperature of the solution also depresses the sweetness value of a fructose solution: at the same time the specific rotation is increased and there is a shift in the equilibrium. The age and temperature of the solution may account for the reported differences of the sweetness of fructose when compared with sucrose, ranging from twice to one-third as sweet. The specific rotation of fructose increases with increasing concentration, although not as rapidly as with increased temperature. The relative sweetness of different concentrations is temperature–concentration-dependent. At lower temperatures, the relative sweetness of fructose decreases with increasing concentration but, at higher temperatures (37 °C), the relative sweetness increases. This reflects a complex dependence on the shift in equilibria, shift of forms, and other phenomena.

Sweetness is system-dependent, i.e., the perceived sweetness in a food or beverage system depends on several factors, including temperature, pH, solids content, and the presence of other sweeteners. Fructose exhibits a synergy with other sweeteners: the relative sweetness of 50/50 fructose–sucrose mixture is 128. Similar synergies occur when fructose is used in combination with aspartame, saccharin, and/or sucralose. This synergy allows for a formulator either to obtain higher degrees of sweetness in the finished product without increasing the total level of sweeteners, or to retain a satisfactory degree of sweetness while reducing the amount of sweeteners used and the cost. Either way, fructose both sweetens and improves the sweetness profile.

The sweetness intensity profile of fructose is different from those of sucrose and dextrose. The sweetness

of fructose is perceived earlier than that of sucrose or dextrose and the taste sensation reaches higher than sucrose and diminishes more quickly than sucrose. Because of this, use of fructose most often results in an enhancement of other flavors in the system. Many flavors such as fruit, some spices, and acids come through more clearly and distinctly after the fructose sweetness dissipates, because they are not masked by the lingering sweetness of sucrose. Thus, fructose may improve the flavor profile and possibly reduce cost for these flavors. (*See* **Sensory Evaluation**: Taste.)

Food Applications

Fructose is hygroscopic, which means it readily absorbs water from its environment. Fructose is an excellent humectant, in that it retains moisture for a long period of time at low relative humidity. This humectant can be used to impart improved eating quality, better texture, and longer shelf-life to the food products in which it is used. The use of fructose with other sugars can also help control unwanted crystallization.

Another property fructose imparts because of its relatively small molecular weight is depression of the freezing point of a food. Fructose will lower the freezing point more than will di- or oligosaccharides. In systems in which this may be undesirable, such as soft-serve or hard-frozen dairy desserts, care must be taken to balance the formulation with sugars of higher molecular weights (e.g., corn syrups) or to add the correct combinations of gums and stabilizers. In other products, however, freezing point depression can be used to an advantage. Depressing the freezing point of frozen fruit, for example, helps protect the integrity of the fruit pieces by reducing damaging ice crystal formation.

The presence of a reducing sugar, such as fructose, in a formulation facilitates color development via the classical Maillard browning reaction. Bakers depend on this reaction to give bread crust its distinctive brown color and to provide a baked color to microwaved products. The reducing property is also useful in maintaining the bright red color of tomato ketchup and strawberry preserves. The browning reaction can be controlled to some extent by manipulation of pH and/or time and temperature of processing and storage.

Fructose enhances starch functionality in food systems. Fructose causes viscosity to develop more quickly and the system to achieve a higher final viscosity compared to sucrose. In products that require the starch to cook, fructose will lower the temperature required to gelatinize the starch and ultimately cause a higher final viscosity to be developed.

The major applications of HFS are carbonated beverages and leavened bakery products. Bakers found that the HFS yielded a finished product identical to those sweetened with sucrose. The carbonated beverage industry is the largest user of HFS-42 and -55. The 42% fructose product is used primarily in noncola beverages, often acidified with an organic acid system that is easier to sweeten. Many cola systems, however, use phosphoric acid, which requires the higher sweetness of HFS-55 to give the correct flavor balance; alternatively, an increased amount of HFS-42 may also be used. The National Soft Drink Association has developed stringent guidelines for the improvement in quality of the syrup: color, taste, odor, floc, ash, fructose, other carbohydrate amounts, and microbiological standards. (*See* **Sweeteners**: Intensive.)

Flavor enhancement and a natural compatibility with fruit flavors are two reasons the fruit canning industry has become the third major user of HFS. HFS is frequently blended with liquid sugar and corn syrups to get the right balance of sweetness and fruit flavor. Corn syrups help add eye appeal to the fruit by imparting a shiny, glossy surface appearance when the fruit is served in a dish or tray. HFS-42 is used because of its compatibility with organic acid systems.

HFS is used extensively as a sweetener such as in dairy products like yogurt, chocolate milk, and icecream. Quality and economy are once more the primary reasons, in addition to improved flavor perception and rapid fermentability in yogurt, and mouth feel and viscosity in icecream and chocolate milk. The makers of jams, jellies, and preserves are also major users of HFS. High-solids systems can be formulated by using HFS and corn syrups without the storage problem of crystallization common to sucrose and dextrose. HFS enhances fruit flavors and stabilizes the color of these products throughout their storage life.

Crystalline fructose was positioned early as a nutritionally advantageous sweetener because of the way it is metabolized by the body. Therefore, many of the products first sold that contained crystallized fructose were those that went to diet and/or health-conscious consumers. These included various powdered diet beverages and meal supplements, nutritional candy bars, and other specialty food items. These applications were successful in introducing fructose to many consumers, but cost prevented it from being used as an ingredient in mainstream food products. The development of the technology for crystallization of fructose from HFS-90 led to the commercialization of fructose on a larger scale and at a lower cost than was previously possible.

Application areas for crystalline fructose now include dry-mix beverages, for which the intense sweetness of fructose allows for a reduction in total sweetener content and a parallel reduction in energy. Energy reduction, simple sugar reduction, and flavor profile enhancement are three popular formulation trends that fructose can favorably impact. Other applications that capitalize on the sweetener synergy to reduce energy include 'lite' pancake syrups and 'lite' carbonated beverages. The use of crystalline fructose and sucrose in these products can reduce the energy value by at least one-third, qualifying it for a reduced-energy label. Other application areas include breakfast cereals (flavor enhancement, sweetness synergy); yogurt, chocolate milk, and eggnog (flavor enhancement, energy reduction); baked goods (humectancy, starch synergy, flavor enhancement); fruit packs (flavor enhancement, storage stability); energy supplement in sports drinks (solubility, flavor enhancement, sweetness); and confections (sweetness, starch synergy, humectancy).

Food consumption surveys show an increase in fructose consumption: the ratio of fructose intake as a proportion of total sugars has increased by nearly 30% from the 1977–1978 and 1987–1988 Nationwide Food Consumption Surveys. High intakes of fructose have implications for gastrointestinal health, blood glucose control, and lipid metabolism. Fructose is primarily absorbed from the gut by facilitated diffusion. Persons vary in their abilities to absorb fructose – some experience symptoms of malabsorption with a 20- to 50-g load (a 12-oz (336-g) sweetened soda or fruit drink has between 14 and 22 g fructose; 8 oz apple juice has 14 g fructose). Some of the malabsorption of fructose-containing products may be the result of other nutritive sweeteners that are poorly absorbed (e.g., sugar alcohols in apple juice). Fructose is better absorbed when consumed in sucrose than in products where the amount of free fructose exceeds the amount of glucose (e.g., in honey, prunes, apples and juice, HFS, or crystalline fructose). Because of the method of absorption, fructose intake may lead to a slower rise in blood glucose than sucrose-based sweeteners. This fact, along with rapid clearance of fructose from blood serum, may improve glycemic control. High intakes of fructose could, theoretically, increase production of lipid precursors and increase the risk of hypertriglyceridemia. However, this effect is not consistently seen, even in those who are at high risk of elevated plasma triglycerides.

Only two rare inborn metabolic problems with the ingestion of fructose have been reported: essential fructosuria and fructose intolerance. In essential fructosuria, fructokinase is deficient and, with fructose intolerance, fructose-1-phosphate aldolase is deficient. In both cases, individuals should avoid foods containing fructose and sucrose. Other digestive problems have been reported in individuals who have an allergy to corn products since the primary source of fructose added to foods is HFS from corn.

See also: **Browning**: Nonenzymatic; **Carbohydrates**: Classification and Properties; **Sensory Evaluation**: Taste; **Sweeteners**: Intensive

Further Reading

Anonymous (1998) Position of the American Dietetic Association: use of nutritive and nonnutritive sweeteners. *Journal of the American Dietetic Association* 98: 580–587.

Conforti FD (1993) A suggested corn sweetener blend that serves as a replacement for sucrose in a basic ice cream formula. *Journal of the Society of Dairy Technology* 46: 57–61.

Conforti FD (1994) Effect of fat content and corn sweeteners on selected sensory attributes and shelf stability of vanilla ice cream. *Journal of the Society of Dairy Technology* 47: 69–75.

Hajian HG and Pecsok RL (1990) *Modern Chemical Technology*, vol. II. Englewood Clifts: Prentice Hall.

Hanover ML and White JS (1993) Manufacturing, composition and applications of fructose. *American Journal of Clinical Nutrition* 58: 724S–732S.

Johnson JM, Harris CH and Barbeau WE (1989) Effects of high fructose corn replacement for sucrose on browning, starch gelatinization, and sensory characteristics of cakes. *Cereal Chemistry* 66: 155–157.

Johnson JM and Harris CH (1989) Effects of acidulants in controlling browning in cakes prepared with 100% high fructose corn syrup or sucrose. *Cereal Chemistry* 66: 158–161.

Lee CK (1987) The chemistry and biochemistry of the sweetness of sugars. *Advances in Carbohydrate Chemistry and Biochemistry* 45: 199–351.

Pomeranz Y and Meloan CE (1994) *Food Analysis: Theory and Practice*, 3rd edn. New York: Chapman and Hall.

Shallenberger RS (1982) *Advanced Sugar Chemistry*. Westport: AVI.

FRUITS OF TEMPERATE CLIMATES

Contents

Commercial and Dietary Importance

L G Smith and S M Somerset, Queensland Department of Primary Industries, Hamilton, Queensland, Australia

Commercial and Dietary Importance

Definition

'Temperate-climate fruits' refers to all fruits that normally grow further than 23.5° from the equator, but also includes avocado and citrus, fruits often classified as subtropical. Some temperate fruits can be grown at higher altitudes in tropical latitudes, although poor fruit set may occur in some species. 'Fruits' could be defined as seed-bearing plant organs that undergo a characteristic process, called ripening, changing from unpalatable to palatable (vegetables have no such sequence). Much of the information in this article is applicable to all fruits, and some tropical fruits are listed here for comparison. Horticultural products that could be classified as temperate fruits (such as many nuts and the tomato) are addressed in separate entries.

Production

Temperate fruit production is a major industry. The production and marketing of fresh fruits are much more labor-intensive than that of broadacre crops, so that the temperate fruit industry is a major source of employment. As a consequence of the high labor content of production, large-scale fruit production by major companies is tending to shift towards suitable regions with access to cheap labor.

The estimated world production of the major temperate fruits in listed in Table 1. The production for individual countries is shown in Table 2. For many temperate fruits, large quantities are processed in one form or another (e.g., canned, dried, wine, juice, glacé), and the production data in Tables 1 and 2 include that for processed fruit. The specific details of world fruit production are difficult to determine since, in many countries, fruit-production data are unavailable, and reporting countries also lack standards of uniform data presentation.

Grapes are the major temperate fruit in terms of tonnes produced, with large quantities going into wine and dried fruits. The major grape-producing countries are Italy, France, the USA, Spain, and countries of the former USSR. Second, in terms of quantity produced, are oranges, many of which are processed into juice, and third are apples, probably the temperate fruit most often eaten fresh, although large quantities are also juiced, dried, and canned. Buying patterns in British households for 12 common fruits are presented in Table 3.

Dietary Importance

There are many different aspects to fruit in the diet. Not only is fruit an important source of dietary fiber, carbohydrate, and vitamin C and A, but because of its flavor, it also influences the intake of many other foods and therefore can exert indirect effects on the diet. Although different fruits are morphologically and botanically diverse, in fruit consumption studies, the nutritional composition is generally considered collectively, rather than in terms of distinct fruits. It is therefore more appropriate to discuss the nutritional impact of fruits as a whole.

Historical

Man has been attracted to the sweet taste and full aroma of fruit since prehistoric times, and attempts by man to cultivate various fruits date back over 8000 years. It is not surprising then that the word 'fruit' is often used to imply a pleasure or reward (e.g., in expressions such as 'fruit of our labor'). Their sweet taste is due to the high content of sugars, in particular fructose, sucrose, and glucose. The high sugar content has been recognized for centuries and is exploited, for

Table 1 Characteristics of the major temperate fruits, with four major tropical fruits included for comparison

	World production[d] (× 10^6 t)	Ability to ripen after harvest	Recommended storage temperature[a] (°C)	Likely storage life[a]	Ethylene production[e]	Respiration rates[a,b,c] (mg CO_2 kg^{-1} h^{-1})	
						At storage temperature	At 20 °C
Temperate							
Grapes	57 397	No	0	1–5 m	< 0.1	1–3	20–25
Oranges	66 212	No	5–7	3–12 w	< 0.1	4–7	22–34
Apples	56 060	Yes	2–3	3–12 m	10–100	2–6	20–41
Watermelons	47 646	No	5–10	2–3 w	0.1–1.0	3–4	17–25
Pears	14 379	Yes	−1.5 to −0.5	2–7 m	10–100	3–7	30–70
Cantaloupes	17 849[f]	Yes[g]	0–5[h]	1–4 w	10–100	5–10	45–65
Peaches	11 065[i]	Yes	0	2–3 w	10–100	4–6	59–102
Plums	8 008	Yes	−0.5 to 0	1–4 w	10–100	2–3	18–26
Lemons	9 335[j]	No	10–15[k]	1–2 m	< 0.1	6–23	7–25
Grapefruit	5 072[l]	No	14–16	1–1.5 m	< 0.1	10–18	13–26
Strawberries	2 601	No	0	5–7 d	< 0.1	12–18	102–196
Apricots	2 670	Yes	−0.5 to 0	2–3 w	10–100	5–6	29–52
Avocados	2 325	Yes	8–10	3–10 w	10–100	20–30	74–150
Currants	654	No	−0.5 to 0				
Raspberries	326	No	−0.5 to 0	2–3 d		18–25	> 99
Tropical							
Bananas	58 618	Yes	13.5–15	4–21 d	0.1–1.0	21–75	33–142
Mangoes	23 455	Yes	10–13	2–6 w	< 0.1	45–150	75–200
Pineapples	12 100	No	6–20[k]	2–4 w[k]	0.1–1.0	4–7	28–43
Papaya	4 801	Yes	7–13	1–3 w	10–100	4–10	22–39

[a]Data largely from Hardenburg RE, Watada EE, and Wang CY (1986) *The Commercial Storage of Fruits, Vegetables, and Florist and Nursery Stocks. USDA Handbook No. 96*. Washington, DC: United States Department of Agriculture, with permission.
[b]Values may vary with production area and cultivar.
[c]Lower value would indicate rate for unripe fruit; a higher value would indicate a peak rate.
[d]Data for 1998 from FAO (1999) *Food and Agriculture Yearbook Statistic Series No. 148*. Rome: Food and Agriculture Organization, with permission; total production includes processed, alcohol production and includes FAO estimates where individual country data not available. Missing data unavailable.
[e]C_2H_4 kg h^{-1} at 20 °C. Data largely from Kader AA (1992) Postharvest Technology of Horticultural Crops, 2nd edn. Location: University of California Division of Agriculture and Natural Resources, with permission.
[f]Includes all melons other than watermelons.
[g]Some cultivars (e.g., honeydew) = No.
[h]Not for honeydew melons.
[i]Includes nectarines.
[j]Includes limes.
[k]Very much dependent on production area.
[l]Includes pomelos and similar fruits.
Note that storage temperature, storage life, and respiration rates vary according to fruit ripeness, cultivar, production area and other factors, and are presented here as a guide only.

example, in the drying of fruits, the high sugar content allowing their preservation and a water content sufficient to make the fruit palatable in the dried form. The production of alcohol from grapes and other fruits also depends on a high sugar content. Fruit consumption patterns now relate largely to consumer preferences. In the past, the traditional dietary role of temperate fruits was more of a luxury than a staple, but with the vast improvement of cultural practices, postharvest techniques and distribution systems, the availability of both fruit and fruit products has increased dramatically. Fruit is now more available, at an affordable price, to more of the world's population than ever before. The modern role of fruit in the diet has thus changed from previous generations.

The Modern Dietary Role

In the diets of peoples in temperate zones, the most important nutritional features are fruits of high contents of fiber, carbohydrate, and vitamin C, and low fat and protein contents. Fruit is prominent in the official dietary guidelines of many countries. Most developed countries encourage increased consumption of complex carbohydrates and fiber, which are abundant in fruits, and dietary recommendations in the UK and Germany specifically advise increased fruit consumption. One recent study in the USA found that 41% of the test population had no fruit or fruit juice on the day of survey. Substantial public health benefits might be gained if fruit consumption increased among the less affluent groups of people in the USA, and, probably, in other industrialized nations.

Table 2 World production for 1998 ($\times 10^3$ t) of eight major temperate fruits (includes all uses, processing, drying, alcohol production, etc.)

	Grapes	*Oranges*	*Apples*	*Watermelons*	*Pears*	*Cantaloupes*	*Peaches*	*Plums*
Africa	*2 746*	*4 582*	*1 415*	*3 298*	*470*	*1 125*	*466*	*117*
Algeria	159	280	65	400	47		39	24
Congo		145						
Egypt	870	1 525	410	1 650	57	550	65	4
Ethiopia	4	14						
Libya	37	60	45	210	1	25	9	1
Madagascar	11	84	7		2		8	2
Mauritania				8				
Morocco	262	1 104	285	259	31	415	39	41
Reunion							1	
Senegal		30		261				
Somalia		8		24				
South Africa	1 273	900	515	50	275	23	240	35
Sudan		17		140		26		
Tanzania	17							
Tunisia	111	122	83	296	54	84	65	10
Zimbabwe	2	70	6				1	
North and Central America	*5 585*	*17 839*	*5 892*	*2 545*	*892*	*2 033*	*1 485*	*934*
Belize		170						
Canada	53		510		17	2	56	5
Costa Rica						120		
Cuba		340				30		
El Salvador		36	1	47		20		
Guadeloupe		1		1		3		
Guatemala	4	80	27	100		86		
Honduras		83		12		156		
Jamaica		72		16				
Mexico	452	4 005	390	500	35	490	129	79
Panama		27		20		17		
Trinidad and Tobago		15	1	2				
USA	5 076	12 571	4 964	1 848	840	1 069	1 300	850
South America	*4 704*	*25 833*	*3 226*	*1 492*	*858*	*399*	*833*	*305*
Argentina	2 021	841	1 347	130	540	100	280	132
Bolivia	23	94	9	21	4		36	4
Brazil	738	22 987	787	765	19	92	146	
Chile	1 665	115	880	63	250	77	285	150
Colombia	17	410		54		10	23	
Ecuador		217	25	18	26	5		8
Paraguay	13	208	1	116		30	2	2
Peru	76	234	127	72	3	12	27	6
Suriname		12		3				
Uruguay	140	185	50	14	16	3	25	4
Venezuela	11	528		237		70	9	
Asia	*12 971*	*12 208*	*27 810*	*36 304*	*8 493*	*11 613*	*4 452*	*3 724*
Afghanistan	330	12	18	90	2	22	14	35
Bangladesh						96		
Armenia	106		56	60	16		21	14
Azerbaijan	144	8	125	145	8		15	25
Bhutan		58	6					
Cambodia		62						
China	2 287	2 804	17 508	23 338	6 728	6 372	2 996	2 717
Cyprus	125	45	10	30	2	9	2	1
Gaza Strip	2	105		6		4		
Georgia	275	115	200	225	20		20	38
India	700	2 000	13 000	250	135	640	87	57
Indonesia		614						
Iran	2 200	1 800	2 000	2 200	175	800	132	160
Iraq	305	316	86	455	3	226	26	31
Israel	90	325	111	260	20	74	48	18
Japan	251	131	900	620	428	430	176	136

Continued

Table 2 Continued

	Grapes	*Oranges*	*Apples*	*Watermelons*	*Pears*	*Cantaloupes*	*Peaches*	*Plums*
Jordan	70	40	65	70	3	15	11	4
Kazakhstan	20			306	8		2	2
North Korea			630	100	120	100	100	
South Korea	384		652	1 006	260	300	147	36
Kuwait				1		1		
Kyrgyzstan	22		74	41			3	2
Laos		28				33		
Lebanon	247	155	119	130	64	35	50	40
Malaysia		11		145				
Oman				32				
Pakistan	75	1 410	600	420	36	400	46	80
Philippines	2	20		70		16		
Saudi Arabia	130			460		130		
Syria	450	438		272	19	51	40	29
Tajikstan	120		324	155	19		27	16
Thailand	26	315	82	385				
Turkey	3 650	830	2 500	3 925		1 800	400	190
Turkmenistan	177		30	258	415		5	9
Uzbekistan	585		380	543	32		82	69
Viet Nam		379		200				
Yemen	155	150	2	95		35		
Europe	*30 216*		*16 856*	*3 908*		*2 605*	*3 718*	*2 892*
Albania	68	3	12	210	4		2	12
Austria	234		392		12		10	77
Belarus			215					45
Bosnia and Herzegowina	17		18	3			2	23
Bulgaria	624		152	246	84		65	78
Croatia	421	1	72	60			9	83
Czech Republic	38		309		21		8	33
Denmark			67		6			
Estonia	2		5					1
Finland			11					
France	7 000	1	2 500	9	256	280	470	210
Germany	1 500		2 154		387		15	416
Greece	1 216	814	358	512	71	150	480	8
Hungary	669		500	76	37	6	54	115
Ireland			9					
Italy	9 208	1 921	2 115	590	931	518	1 429	149
Latvia			15					4
Lithuania	25		100		10			6
Macedonia	244		62	120	9		7	20
Moldova	825		485	25	7	6	72	56
Netherlands			470		130	2		6
Norway						3		12
Poland			1 687		83	20		107
Portugal	764	210	146	2	160		62	18
Romania	874		365		64	610	18	404
Russian Federation	300		1 200	520	50		20	155
Slovakia	76		83	30	12	31	9	22
Slovenia	128		122		5		3	4
Spain	4 842	2 403	726	786	561	980	888	150
Sweden			65		16			1
Switzerland					95			10
UK	2		180		25			16
Ukraine	380		1 200	395	160		50	170
Yugoslavia	588	2	235	325	74		46	480
Oceania	*1 176*	*396*	*861*	*99*	*213*	*74*	*111*	*36*
Australia	1 097	369	306	88	172	72	96	33
New Zealand	78	24	501	4	41	1	15	3

Table 3 Fruit-buying patterns of British households[a]

Item	*Percentage buying*	*Times per year*
Bananas	87	22
Apples	86	17
Easy-peel oranges	70	5
Grapes	65	9
Pears	64	9
Regular oranges	61	7
Strawberries	66	4
Melons	53	5
Exotic fruit	44	5
Peaches	39	3
Pineapples	17	2
Avocados	12	3

[a]Source AC Nielson 1998. From Hughes D (1999) Marketing fruit in Europe. In: *Good Fruit and Vegetables*, pp. 34–35. Melbourne: Rural Press, with permission.

Direct Nutritional Attributes

Fruit is generally high in dietary fiber and low in fats and oils, and is therefore an important food in diets designed to reduce the risk of coronary heart disease in developed countries. An exception is the avocado, which has an unusually high oil content (containing up to 30% fat, depending on the production area), but recent studies indicate a beneficial role of the avocado with respect to heart disease, possibly because 50–75% of the fat is monounsaturated.

One dietary feature of fruit in general is the high content of water, ranging from 65% in apples to 88% in papaya. This feature enables fruit to quench thirst in addition to satisfying hunger.

Vitamins

Ascorbic acid (vitamin C) is the principal vitamin supplied by fruit in the diet. Some fruit are particularly rich in this vitamin. Citrus have about 50 mg per 100 g, kiwi fruit about 100 mg per 100 g, and blackcurrants and bell peppers (*Capsicum annum*) have between 200 and 300 mg per 100 g.

The orange-colored fruits such as rockmelon, peach, persimmon, tomato, and apricot are rich in β-carotene, which is converted to vitamin A in the body. Certain fruits can be rich in particular vitamins. For example rockmelon strawberry and orange have significant amounts of folic acid. Pantothenic acid is found in appreciable quantities in watermelons currants and berries, and nicotinic acid in apricots, nectarines, peaches, passion fruit, and guavas.

Fiber

Fruit is a rich source of pectins and gums, components of fiber, and thus fruit constitutes an important source of dietary fiber. Pectins have been shown to delay gastric emptying, which can lead to favorable changes in glycemic responses to particular foods. Pectin may also introduce a satiety effect.

Indirect Nutritional Consequences

Fruit can be consumed in a number of different forms apart from fresh. The drying of fruit has been a popular preservation technique for many generations. The dehydration process concentrates those nutrients that are not heat- or light-labile. Vitamin C is the major casualty of this process. The dehydration process also decreases the water activity, thus increasing the keeping quality by making microbial growth less likely.

Processed fruit is being increasingly used in the dairy industry, as an additive in fermented milk products such as yogurt. This use of fruit has led to a wider acceptability of yogurt and has therefore influenced the intakes of certain nutrients such as calcium and milk protein. This also applies to fruit jams and conserves. Although these fruit products may not have a remarkable nutrient composition themselves, they can be used to improve the palatability of foods such as wholemeal bread, thus increasing the intake of nutrients such as grain fiber and the B-group vitamins. Fruit derivatives are also used in the confectionery-manufacturing industries.

Large quantities of fruit are used in the production of fruit juices. The physical processing to which fruit is subjected for juice production leads to a significant reduction in the content of ascorbic acid, but vitamin C is added to processed fruit juice, primarily as an antioxidant preservative, but also to increase consumer appeal. Fruit juice, either processed or fresh, is a significant source of vitamin C in many developed countries.

Pharmacological and Therapeutic Properties

Many therapeutic drugs in modern use originated as plant products. It is not surprising, then, that certain fruit components exert pharmacological or therapeutic effects. Limonin and nomilin and other liminoids are present in citrus such as orange, lemon, lime, and grapefruit. These compounds are believed to have a role in inhibiting the development of certain forms of cancer, and research has indicated that the antioxidant forms of β-carotene may also play a role in the prevention of some forms of cancer. Prunes (a variety of dried plums) contain hydroxyphenylostatin derivatives that stimulate colonic smooth muscle, thus explaining their traditional use as a laxative.

The effects of some pharmacological agents that occur in some fruits add to the importance of fruit in the diet.

Commercial importance

Production

The suitability of a region for commercial fruit production depends on a combination of climate, adequate water, suitable soils, and access to labor and markets. Climatically, the rainfall pattern is most important, as rain can cause major losses from foliar diseases, fruit splitting, and fruit disease, leading to direct losses or downgrading. Consequently, a dry summer climate with abundant good water for irrigation is ideal for fruit production since the water can be applied precisely as required.

In production areas with high labor costs, much research is directed towards methods of reducing the labor component of both production and postharvest procedures. Such research includes the use of new dwarfing rootstocks to reduce the cost of harvesting and tree management, planting parameters to facilitate machine pruning, trellising parameters to facilitate machine harvesting, tree shakers for fruit harvest, bulk methods for handling, netting for hail and pest control, and improved postharvest practice and facilities.

Varieties and propagation Nearly all major varieties of the major commercially important temperate fruit trees are clonally propagated (rather than grown from seed), either directly by cuttings or by grafting scion varieties on to rootstocks. The grafting process allows a productive variety (the scion) of good eating quality to combine with a different variety (the rootstock) that better resists endemic root diseases. An additional desirable trait of some (dwarfing) rootstocks can be a reduction of tree vigour (wasteful leaf and shoot growth) in the scion, using the incompatibility between rootstock and scion. New varieties introduced from other regions or countries need to be assessed for local suitability, since varieties that produce well in one geographic area often produce poorly in other areas. At rural horticultural research centers ('field stations') around the world, a substantial proportion of the work on all major commercial fruits is continually developing new and improved scions and rootstocks and assessing various rootstock–scion combinations. New varieties are selected from either superior bud mutations in existing varieties or, more commonly, concurrent breeding programs, considering fruit quality, crop yield, and freedom from disease. While some varieties have a long history of productive quality, there is a continual search for new varieties to improve both scion and rootstock performance. Many new varieties of many temperate fruits are now being developed purely to meet changing consumer demands rather than for production factors such as yield or disease resistance.

Some important fruits (e.g., certain apples, plums, cherries) require cross-pollination for adequate fruit set, and orchards of these fruits are interplanted with different varieties.

Fertilization Fruit tree production can require fewer applied fertilizers than broad-acre farming. The deeper roots have access to nutrients that are more abundant deeper in the soil. Calcium and potassium concentrations can also significantly affect fruit quality. Inadequately fertilized trees result in a poorer fruit set, smaller fruit size, and increased susceptibility to pests and disease. Potassium and phosphorus requirements are relatively low in fruit crops, but calcium and trace elements are usually of particular importance to prevent disorders occurring in fruit, although the trees show no other symptoms.

Preharvest pests and disease In most fruit crops, pests (e.g., insects) and disease (e.g., fungi) are major problems. Many commercially bred varieties would be unlikely to survive in nature without human intervention, as they have been selected primarily on the basis of fruit quality and yield, rather than on pest and disease resistance. In most production regions, strict quarantine regulations are enforced to preclude entry of new pests and diseases.

There has been a widespread and general acceptance of the need to greatly minimize chemical pesticides by all sectors of the community, including growers, and recent progress in more environmentally friendly pest control has been quite remarkable. Integrated pest management (IPM), the established approach to controlling preharvest pests, is an integration of chemical, biological physical, and procedural methods. IPM fosters effective populations of the natural predators of the pest species, allowing a well-monitored balance between pest and predator species. If and when the balance is significantly disturbed, a minimum number of carefully applied and precisely timed applications of chemical sprays are applied to keep pests or disease within acceptable limits. The use of insect pheromones that are harmless to humans is now common in controling many pests. These are used either as lures on to treated surfaces containing insecticides or in traps for destruction or by disrupting mating behavior. There still remains much to be done in pest and disease control, and this work constitutes a currently active arena for researchers from many disciplines.

Climacteric and nonclimacteric fruits All fruit are living organs and as such continue to live and respire

after harvest. Fruits such as apples, pears, avocados, tomatoes, and bananas can be harvested slightly immature but 'green' (unripe) without significantly reducing the final eating quality when they subsequently ripen. During this ripening process, complex polysaccharides, such as starch or pectins, hydrolyze to sugar resulting in an increase in sweetness and palatability. All climacteric fruits fall into this category. The fruit are called '*climacteric*' fruit because, concomitantly, their respiration rate develops a climacteric (peak) as they ripen.

Some other fruits such as citrus, cherries, strawberries, grapes, pineapples, and some melons do not get better to eat *after harvest*. Although these fruit may change color, they do not get better to eat the way that a mature but unripe tomato or mango will become notably palatable as ripening proceeds. Concomitantly, these fruit do not have a climacteric peak of respiration, that is, they are '*nonclimacteric*.'

The postharvest ripening nature of major fruits is shown in Table 1.

Fruit Maturity

On the tree or bush, an attached fruit undergoes the natural sequence, whereby it grows (develops), matures, and commences to ripen. During ripening, the fruit becomes attractive to eat: softening, changing colour, losing unpalatable off-flavors such as tannins, increasing in sweetness, becoming less acid, and developing attractive flavors.

When a fruit is 'mature,' it has reached the most appropriate state of development and is now ready for the next stage of its progress, which may be storage, processing, marketing, or ripening for immediate consumption. 'Under-mature' or 'immature' means that the fruit has not yet reached the state of development most appropriate for a particular destiny, being too hard for processing or too unpalatable if consumed. 'Over mature' means that the fruit has developed past the most suitable state of development, for storage, processing, or marketing. Such fruit may be too soft, too colored, or too prone to breakdown during subsequent storage or marketing. A few fruits (e.g., pears and some bananas) do not reach maximum eating quality if left to ripen on the tree, and in that sense, the fresh fruit could become 'overmature' for human consumption.

Normally, the commercial fruit is harvested somewhat before the ripening has commenced to avoid damage during transport, but most climacteric fruit will improve in eating quality and reach an acceptable quality, even if picked substantially immature. Maturity is thus independent of ripeness. Fruits can be immature but unripe, immature and ripe, mature but unripe, mature and ripe.

Nonclimacteric fruits show a dramatic increase in eating quality during the last several days before natural ripeness begins, but this process halts at harvest, and no increase in palatability occurs (or very little). The exact cause for this different response has not been fully determined. If nonclimacteric fruits are harvested even only a few days too early, they most often lack a full-bodied flavor and palatability. However, if left too late, such fruits can suffer greatly increased rots and breakdown during marketing. Hence, the harvest maturity for nonclimacteric fruits is crucial and is often determined using a refractometer to measure the 'soluble solids' (mainly sugar) concentration in the juice of sample fruits. This measurement, either alone (pineapples, strawberries) or in conjunction with titratable acidity measurements (grapes, citrus), is then used to judge the 'maturity' of the fruits.

The maturity of the fruits is generally most important when judging to harvest the first fruit of a district for the season. For avocados, maturity is commonly judged using percentage dry matter.

The maturity indices mentioned (soluble solids; percentage dry matter) are destructive in that the sampled fruit is damaged and usually cannot be marketed. Considerable effort has been applied to develop nondestructive indices of fruit maturity, in particular NIR (near infra-red spectroscopy). This has been quite successfully used in some thin-skinned fruits (stonefruit) where a reflected infrared beam is automatically scanned and analyzed for sugar content. For other fruits such as citrus or pineapple, the thick skins currently pose an obstacle to accurate analysis using NIR. Methods of overcoming this and other problems for different fruits are currently being developed.

Harvesting, Handling, and Packaging

Harvesting constitutes a major production cost. Mechanical harvesting is increasingly used in developed countries, and often where suitable equipment is not currently applicable (e.g., pineapples, melons), research is in progress. Fruits destined for processing are often mechanically harvested because the fruit are processed rapidly before disease can develop from bruised fruits, and partially damaged fruits can be trimmed for processing.

Commercial fruit production in developed countries is often large scale and highly capital-intensive, with sophisticated and specialist equipment used for all processes, including unloading, washing, sorting, treating, size-grading, packing, cooling, handling,

and transporting. Computer tracking and monitoring are essential components.

Disinfestation and Phytosanitation

Insect pests are generally only of minor importance on fruit for domestic consumption but are a very serious issue on produce for export. Most importing countries have strict regulations (phytosanitary) for fresh fruit. Many previous chemical dips or fumigants for pests such as fruit fly have now been proscribed on human health or environmental grounds. In many instances, there are no acceptable treatments available. Nonresidual treatments have been intensively studied, with increasing success. Heat disinfestation (formerly termed vapor heat) is the controlled application of hot humidified air between 40 and 60 °C and between 80 and 100% RH to fruit in a chamber for a sufficient length of time to heat the innermost fruits and fruit core thoroughly to kill the pests without unacceptably damaging the fruit. The time–temperature 'window' for a treatment is very small before damage occurs, and many fruits, or fruits from some growing areas, are unacceptably injured by the disinfestation treatment. Recent research has revealed that 'preconditioning,' a pretreatment temperature regimen, is effective in ameliorating damage in otherwise susceptible fruit. Heat disinfestation is currently commercially used on papaya, lychees and mangoes, but current and ongoing research is very active, endeavoring to find suitable treatments for many other fruits. In some instances, it has been found that heat treatments can greatly reduce postharvest disease and thus improve the market quality.

Postharvest Aspects of Fresh-market Fruit

Fruits are highly perishable but can be marketed vast distances from the site of production with the application of sound postharvest techniques such as appropriate disease treatment; storage-temperature management, humidity control, controlled atmospheres, correct packaging and palletizing, ventilation, and transport carefulness and timely distribution. Small changes in preharvest conditions, harvest maturity, postharvest handling, storage conditions, new varieties, farm practices, or season can have dramatic affects on product quality. With large-scale handling methods, large-scale losses are a constant hazard to novice and experienced handler alike.

Storage temperature and humidity Correct temperature maintenance is the most important factor during marketing, as deterioration from both disease and ripening increases logarithmically with temperature. Fruits generally suffer from disorders if stored for too long or ripened at too low temperatures (the pear is a notable exception). Such disorders include off-flavors, breakdown, mealiness, flesh and skin browning, skin pitting, and increased susceptibility to disease. For example, peaches and nectarines suffer textural dryness (become 'woolly'), whereas avocados develop off-flavors if stored for too long below 13 °C. If tomatoes have been stored for too long at 10 °C or lower, they develop skin diseases from organisms that do not normally infect the fruit ('saprophytic' infections). For most fruits, the optimum ripening temperature is 20–22 °C. Conversely, some fruits can now be stored at lower temperatures than previously believed (and have a longer marketing life) provided that they are removed from cold storage before ripening commences. Recommended storage temperatures and maximum storage times are shown in Table 1.

Different cultivars, different production areas and different seasons can radically affect the tolerance to cold temperatures. Compared with tropical fruits, many temperate fruits can be held at much lower temperatures without any adverse affects, and a few can have relatively long storage lives. Apples can be stored for up to 12 months, citrus for up to 16 weeks, but stone fruit only 2–4 weeks and strawberries only 3–8 days.

Fruit Ripening and Respiration

Harvested fruits, being living tissue, respire during their postharvest life, taking up oxygen, 'burning' sugars, and giving out both carbon dioxide and heat at a rate directly proportional to the ambient temperature. When fruit is stacked, stored, or transported, sufficient refrigeration capacity is required to cope with any climacteric (peak) respiratory heat load that is autocatalytic. If not controlled, total breakdown can occur. Certain fruits (e.g., strawberries, peaches), have high rates of respiration and hence need special attention to cooling, but generally, all fruits destined for distant marketing need some refrigeration. Warm fruit that are not promptly cooled have a reduced market life, and runaway deterioration is a risk. In stacked produce, the cardboard cartons act as insulating containers, preventing cooling. Forcing air through such stacked trays or boxes ('forced-air cooling') is widely used in warm climates to rapidly remove field heat after harvest or prior to transport, reducing the respiration rate and consequently the rate of the heat production.

During ripening, many fruits produce ethylene gas, a potent plant hormone that initiates and catalyzes the commencement of ripening. Ethylene production in fruit follows a similar climacteric production to

respiration, but some fruit produce much higher ethylene levels than others. Table 1 shows the production rates for the different fruits. Ethylene from fruit can trigger senescent changes such as yellowing or withering in vegetables or flowers, or ripening in adjacent fruits. For this reason, during transport, fruits that produce substantial amounts of ethylene are sometimes isolated. Alternatively, ethylene in store room atmospheres is sometimes destroyed with ozone from special generators, or absorbed, using porous alumina beads impregnated with potassium permanganate and used as small disposable sachets in fruit cartons.

Effects of fertilizers on fruit quality and superior taste Whilst soil mineral composition and/or applied fertilizers have major affects on tree vigour and fruit yield, fruit quality is much less directly affected. The perceived 'sweetness' or 'flavor' of superior fruits depends on a three-way balance between the concentrations of sugar, acid, and flavor components. In an orchard, the concentration of sugar per fruit, generally an index of the superior taste, can vary from plant to plant or season to season. Sweetness usually decreases proportionally with increasing fruit mass per unit leaf area per plant. In general, for any particular variety, the flavor components depend on fruit maturity, the sugar concentration depends on ambient light intensity and leaf area, and acid concentration depends on ambient temperatures. The flesh color of some fruits at least (e.g., citrus) depends on day/night temperatures.

Deficiencies of some elements such as calcium, boron, and molybdenum affect shape, internal and external blemishes and storage disorders in some fruits more than others. A high calcium/potassium ratio in apples is used as an index of storage quality as it reduces softening during storage. In citrus fruits, applying extra potassium increases acidity, sometimes seen as a desirable factor. Applied nitrogen encourages leaf growth and photosynthetic efficiency, which can result in more sugars, thus giving sweeter fruit. In general, prevailing weather, applied water, pests, and diseases have a much greater influence on fruit quality than do mineral nutrition or applied fertilizer.

Globalization and the Future of Horticultural Research

Increased economic globalization over recent times has brought about a major decline of both economic viability and research support within horticulture in first-world countries. Third-world countries are becoming the heirs to the resources as their research institutions have often been fostered with using aid from first-world countries. During this same period, large numbers of horticultural research workers in first-world research systems have steadily been not replaced, while third world horticulture and horticultural research has dramatically expanded. The problem of the deinstitutionalization of the horticultural knowledge base in first world countries is now a serious challenge as career researchers are replaced with casual workers, and the institutional structure is eroded away. Concurrently, domestic horticultural production in first-world countries is being forced to compete with produce from third-world regions. The whole issue is contentious and politically challenging, and a situation that has steadily deteriorated for many years without any clear vision of resolution.

See also: **Ascorbic Acid**: Properties and Determination; Physiology; **Dietary Fiber**: Properties and Sources; **Fruits of Temperate Climates**: Factors Affecting Quality; **Fruits of Tropical Climates**: Commercial and Dietary Importance; Fruits of the Sapindaceae; Fruits of the Sapotaceae; Lesser-known Fruits of Africa; Fruits of Central and South America; **Gums**: Properties of Individual Gums; **Pectin**: Properties and Determination; Food Use; **Vitamins**: Overview; Determination

Further Reading

Anonymous (1982) *Fruit and Vegetables Facts and Pointers*. Virginia: United Fresh Fruit and Vegetable Association.

Anonymous (1989) *Fresh Produce Manual. Handling & Storage Practices for Fresh Produce*, 2nd edn. Location: Australian United Fresh Fruit and Vegetable Association Ltd.

Block G (1991) Dietary guidelines and the results of food consumption surveys. *American Journal of Clinical Nutrition* 53: 356S–357S.

FAO (1999) *Food and Agriculture Yearbook Statistic Series No. 148*. Rome: Food and Agriculture Organization.

Hardenburg RE, Watada EE and Wang CY (1986) *The Commercial Storage of Fruits, Vegetables, and Florist and Nursery Stocks. USDA Handbook No. 96*. Washington, DC: United States Department of Agriculture.

Hughes D (1999) Marketing fruit in Europe. In: *Good Fruit and Vegetables*, pp. 34–35. Melbourne: Rural Press.

Kader AA (ed.) (1992) *Postharvest Technology of Horticultural Crops*, 2nd edn. Location: University of California Division of Agriculture and Natural Resources.

Fruits of the Ericacae

J F Hancock and R M Beaudry, Michigan State University, East Lansing, MI, USA
J J Luby, University of Minnesota, St. Paul, MN, USA

Global Distribution

The family Ericaceae is perhaps best known around the world for its showy garden members such as rhododendrons and heathers. However, some 13 genera contain species with fleshy berries, which are consumed locally in many parts of the world (Table 1). The fruits of these species are commonly eaten fresh or sometimes dried. Many are processed into preserves, juice, or wine.

Vaccinium is the most important genus in terms of fruit production. The majority of species inhabit open mountain slopes in the tropics, with the balance being distributed in subtropical, temperate, and boreal regions of the northern hemisphere. Plants of *Vaccinium* vary in form from epiphytes to trailing vines to trees, with the majority being terrestrial shrubs. Some form crowns, whereas others produce new aerial shoots from rhizomes. Flowers may be solitary or in racemes or clusters.

Commercial Importance

Commercial fruit production is mainly from species of section Cyanococcus (cluster-fruited blueberries), including cultivars of *V. corymbosum* L. (highbush blueberry) and *V. ashei* Reade (rabbiteye blueberry) and native stands of *V. angustifolium* Ait. and *V. myrtilloides* Michx. (lowbush blueberries). *Vaccinium macrocarpon* Ait. (large cranberry), a member of section Oxycoccus, is also an important domesticated fruit, although most of the cultivars were selected from the wild. *Vaccinium myrtillus* L. (bilberry, whortleberry), in section Myrtillus, is collected exclusively from the wild. *Vaccinium vitis-idaea* L. (lingonberry, mountain cranberry or cowberry), in section Vitis-idaea, is also collected predominantly from the wild, although it has recently been domesticated.

The highbush blueberry is by far the most important commercial crop, producing over 95 000 t of fruit annually on over 20 000 ha. Highbush production occurs in 36 states in the USA, in six Canadian provinces, and in Europe, Australia, and New Zealand. Highbush plants have recently been established in South American countries, especially Chile. The largest acreages are in Michigan, New Jersey, and North Carolina in the USA, and British Columbia in Canada. Interest is growing in California. Half-high types (*V. corymbosum* × *V. angustifolium*) have made an impact in Minnesota and regions too cold to successfully grow pure highbush.

The commercial production of rabbiteye blueberries is largely confined to southeastern USA, centered in Georgia, and extending from North Carolina to Texas. The estimated area in production in the USA is over 3000 ha with approximately half of this in Georgia. The total annual production is over 5500 t. There is also interest in growing rabbiteye blueberries in warm temperate and subtropical regions of the world, such as southern African and South American countries.

The commercial production of lowbush blueberries is largely confined to approximately 50 000 ha in Maine (USA) and Quebec and the Maritime Provinces of Canada. Maine has only 43% of the hectarage, but generates over half of the production. Annual production now exceeds 55 000 t. Cranberry production is over 200 000 t annually from over 15 000 ha, primarily in Wisconsin, Massachusetts, New Jersey, Washington, Oregon, and Nova Scotia. A major planting of cranberries was also recently made in Chile, along with modest plantings in Ireland, the UK, and northern Europe. Lingonberries are commonly harvested in Scandinavia, the former USSR, Poland and several other European countries, and in eastern Canada, whereas bilberries are gathered and consumed primarily throughout northern Europe, Siberia, and northeastern China.

Varieties

Highbush blueberry breeding was begun by Dr. Fredrick Coville of the US Department of Agriculture in 1908. Over 70 cultivars are now available, although only two, Bluecrop and Jersey, constitute more than half the hectarage. Bluecrop is the most dominant cultivar, owing to its high yields and long storage life, and now encompasses about a third of the total hectareage. It is the leading cultivar in nearly all production regions, and its acreage is still increasing. Most of the Jersey acreage is located in Michigan, and it is no longer being actively planted. Among the newer releases, Elliott and Duke are the most popular. Other important cultivars include Croatan, Blueray, Bluetta, Weymouth, Berkeley, Patriot, Bluejay, and Rubel.

Beginning in the 1970s, the introduction of highbush cultivars with a low chilling requirement for budbreak (so called 'low-chill' or 'southern' highbush) heralded the extension of 'highbush type' blueberry culture into more southern latitudes in the USA and Australia. Sharpblue is now the most widely

Table 1 Species of the family Ericaceae with edible fruits

Genus and species	*Geographic range*	*Uses*[a]
Vaccinium angustifolium Ait.	Eastern N. America	D, F, P, J
V. andringtrense Perr.	Madagascar	D, F, P
V. arbuscula (A. Gray) Mart.	Western N. America	D, F, P
V. arctostaphylos L.	Southern Europe	D, F, P
V. ashei Reade	Southeastern USA	D, F, P
V. berberifolium (A. Gray) Skotts	Hawaii	D, F, P
V. boreale Hall & Ald.	Eastern N. America	D, F, P
V. caespitosum Michx.	N. America	D, F, P
V. confertum Kunth	Mexico	D, F, P
V. consanguineum Klotzch	Mexico, C. America	D, F, P
V. corymbosum L.	Eastern N. America	A, D, J, F, P
V. cylindraceum Sm.	Azores	D, F, P
V. darrowi Camp	Eastern N. America	D, F, P
V. deliciosum Piper	Western N. America	D, F, P
V. dentatum J. Sm.	Hawaii	F, P
V. erythrocarpum Michx.	Eastern N. America	
V. floribundum H.B.K.	Andean S. America	D, F, P
V. hirsutum Buckl.	Eastern N. America	D, F, P
V. leucanthum Schlect.	Mexico	D, F, P
V. littoreum Miq.	Malaysia	D, F, P
V. macrocarpon Ait.	Eastern N. America	D, J, P
V. membranaceum Dougl. ex Hook.	Western N. America	A, D, F, J, P
V. meridionale Sw.	Jamaica	D, F, P
V. mortinia Benth.	Andean S. America	D, F, P
V. myrsinites Lamarck	Southeastern USA	D, F, P
V. myrtillus L.	Europe, Asia, N. America	A, D, F, J, P
V. myrtillodes Michx.	N. America	D, F, P
V. myrtoides (Blume) Miq.	Philippines	F, P
V. oldhamii Miq.	Eastern Asia	D, F, P
V. oxycoccus L.	Europe, Asia, N. America	D, P
V. ovalifolium Smith	Western N. America	D, F, P
V. padifolium Sm.	Madeira	D, F, P
V. pallidum Ait.	Eastern N. America	D, F, P
V. praestans Pamb.	Eastern Asia	D, F, P
V. stamineum L.	Eastern N. America	D, F, P
V. tenellum Ait.	Eastern N. America	D, F, P
V. uliginosum L.	Europe, Asia, N. America	D, F, P
V. vitis-idaea L.	Europe, Asia, N. America	D, P
Arbutus unedo L.	Mediterranean	A, D, F, P
Arctostaphylos arguta Zucc.	Mexico	M
A. manzanita Parry	Southwestern USA	A, P
A. pungens H.B.K.	Mexico, Southwestern USA	D, F, P
A. tomentosa Pursh	Western N. America	A, D, F, P
A. uva-ursi (L.) Spreng.	Europe, Asia, N. America	D, F, P
Gaultheria antipoda Forster	Tasmania, New Zealand	F
G. hispida R.Br.	Australia	
G. hispidula (L.) Torr. & Gray	Eastern N. America	F
G. myrsinites Hook.	Western N. America	P
G. procumbens L.	Eastern N. America	P
G. shallon Pursh.	Western N. America	D
Gaylussacia baccata (Wang.) Koch	Eastern N. America	F, P
G. brachycera (Michx.) Gray	Eastern N. America	F, P
G. dumosa (And.) T.	Eastern N. America	F, P
G. frondosa Torr. & Gray	Eastern N. America	F, P
G. ursina Curtis	Eastern N. America	F, P
Macleania ecuadoriensis Horold	Ecuador	F
M. popenoei Blake	Ecuador	F
Menziesia feriuginea Sm.	Western N. America	D, F
Chiogenes hispidula (L.) Hitchc.	N. America	F, P
Disterigma margaricoccum Blake	Ecuador	F
D. popenoei Blake	Ecuador	F

[a]Uses of fruits: A, fermented for alcoholic beverage; D, dried; F, eaten fresh; J, juice; M, medicinal; P, cooked preserves, jelly, or jam.
From Uphof JCT (1986) *Dictionary of Economic Plants*. Leure, Germany: Cramer J; Usher G (1974) *A Dictionary of Plants used by Man*. London: Constable.

planted southern highbush blueberry. Other important low-chill cultivars are Reveille, O'Neal, Star, Misty, Gulfcoast, and Floridablue. Of the half-high types released for northern cold regions, Northblue and Northland are the most popular.

Over 25 cultivars of rabbiteye blueberries have been released since their breeding began in the 1930s. Tifblue and Climax form the foundation for the present industry, and Brightwell Premier and Powderblue are also being actively planted. In addition, the very-low-chilling cultivars Beckyblue and Bonita are gaining in importance in Florida.

Although lowbush breeding programs existed in Maine, Michigan, West Virginia, Wisconsin, Minnesota, and Nova Scotia at various times, the industry is still based on native selections. Only seven cranberry cultivars account for about 90% of the crop, and all these varieties were selected from the wild in the nineteenth century. Early Black, Stevens, and Searles represent at least half of the area, with most of the remainder being Ben Lear, Pilgrim, Howes, and McFarlin. Most of the new hectareage is Stevens, Ben Lear, and Pilgrim because of their high color.

Selection and breeding of lingonberries was initiated in the late 1960s in Sweden, and the crop is now being domesticated at several locations across Europe, Scandinavia, and, most recently, the USA. At least 12 cultivars have been released: Sussi and Sanna in Sweden, Koralle, Ammerland, and Red Pearl in The Netherlands, Erntekrone, Erntedank, and Erntesegen in Germany, Masovia in Poland, and Splendor and Regal in the USA.

Morphology and Anatomy of the Fruits

The fruits of all flesh-fruited ericaceous species are berries containing many seeds. A waxy cuticle covers the epidermis of the fruit. In blueberries, the fruit are held on corymbs or racemes. Pure inflorescence buds are formed in the late summer and autumn on shoots of the current season. In cranberries, the fruits are borne singly at nodes 3 to 5 on upright shoots that develop from mixed buds on the trailing vines. Fruit of lingonberries are held in drooping racemes on terminal inflorescences. Bilberries are borne singly in the axils of the lowermost leaves of the vegetative shoot.

The *Vaccinium* berry contains one to 50 seeds surrounded by a fleshy and, usually, colorless mesocarp. In most genotypes, seeds are necessary for normal fruit development, although varying levels of parthenocarpy exist. The ultimate size of fruits is strongly correlated with the number of seeds per fruit.

Blueberry fruit enlarge following pollination, according to a double sigmoidal growth curve, and they go through several phases of color development: (1) immature green, (2) translucent greenish white, (3) greenish pink, (4) blue–red, and (5) completely blue. Up to 50% of the increase in berry volume occurs during the shift from greenish pink to blue. Flowering occurs in early spring, and the fruits are ripe in 40–60 days, depending on the variety and environmental conditions. (*See* **Ripening of Fruit**.)

Cranberries also go through several stages of color development, including green, white, and red. However, the size development of the fruit is more linear than blueberry. Once berry growth begins, it continues at a relatively constant rate for 4–6 weeks. Cranberry fruit matures between 60 and 120 days after blossom.

The anthocyanin pigments that give the fruits their characteristic colors are in the cells of a surrounding endocarp layer. A layer of wax often covers the surface of the berry. The light blue color of many blueberry cultivars results from the combination of dark blue pigments overlaid by the translucent wax. Temperature plays an important role in the development of color, as picked fruit will develop a normal color at 16–27 °C whether they are shaded or not, whereas lower temperatures stop normal development. The blueberry and cranberry are mildly climacteric, with only modest elevations in respiration and ethylene production associated with fruit ripening. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments.)

Chemical and Nutritional Composition

An average blueberry fruit is composed of approximately 83% water, 0.7% protein, 0.5% fat, 1.5% fiber, and 15.3% carbohydrate (Table 2). Cranberries contain 88% moisture, 0.2% protein, 0.4% fat, 1.6% fiber, and 7.8% carbohydrate. Blueberries are 3.5% cellulose and 0.7% soluble pectin, and cranberries contain 1.2% pectin. The total sugars of blueberries amount to more than 10% of the fresh weight, bilberries average 14%, and cranberries contain 4%. The predominant reducing sugars in blueberries are glucose and fructose, which represent 2.4%. The edible portion of the cranberry is composed of 2.66% glucose, 0.74% fructose, and 0.14% sucrose. Its pulp contains measurable amounts of lignin, glucose, arabinose and xylose. Refer to individual nutrients.

The overall acid content of *Vaccinium* fruit is relatively high. Ripe cranberries range from 2 to 3%, whereas blueberries fall in the range of 1–2%. The primary organic acid in blueberries is citric acid

Table 2 Composition of 100 g of fresh *Vaccinium* species

Constituent	*Blueberry*	*Cranberry*
Energy value		
Food energy (kJ)	260.4	109.2
Chemical composition (%)		
Moisture	83.20	88.00
Reducing sugars	12.75	4.20
Nonreducing sugars	1.46	0.11
Acids (as citric)	1.15	2.40
Pectin	0.66	1.20
Fat (ether extract)	2.60	0.40
Protein	0.70	0.20
Ash	0.30	0.25
Fibre	1.50	1.60
Mineral content (mg)		
Potassium	81.0	53.0
Sodium	1.0	2.0
Calcium	15.0	13.0
Phosphorus	13.0	8.0
Magnesium	5.3	5.5
Iron	1.0	0.4
Vitamin content		
Vitamin A (IU)	100.0	40
Vitamin C (mg)	22.5	7.5–10.5
Thiamin (B_1)	0.03 mg	13.5 μg
Riboflavin (B_2)	0.06 mg	3.0 μg
Nicotinic acid (μg)	na	33.0
Pantothenic acid (μg)	na	25.0
Pyridoxine (B_6) (μg)	na	10.0
Biotin	na	Trace
Niacin (mg)	0.50	na

na, not applicable.
From Eck P (1988) *Blueberry Science*. New Brunswick: Rutgers University Press; Eck P (1990) *The American Cranberry*. New Brunswick: Rutgers University Press.

(1.2%). They also contain significant amounts of ellagic acid, a compound thought to reduce the risk of cancer. The cranberry contains high levels of several organic acids, including quinic (1.3%), citric (1.1%), malic (0.9%), and benzoic (0.6%); lingonberries also have very high levels of benzoic acid. Ingestion of cranberries or lingonberries leads to increased acidity of the urine through conversion of its high quinic and benzoic acid contents to hippuric acid by the body. The high acidity and possible antibacterial effects of hippuric acid may relieve urinary tract infections and reduce some types of kidney stones. (*See* **Acids:** Natural Acids and Acidulants.)

Compared with other fruits and vegetables, blueberries and cranberries have intermediate to low levels of vitamins, amino acids and minerals (Table 2). Blueberries contain 22.1 mg of vitamin C per 100 g of fresh weight, and cranberries contain 7.5–10.5 mg. Bilberries contain 5 mg of N per 100 g of alcohol-soluble nitrogen, whereas blueberries contain 15–60 mg of N per 100 g. Blueberries are unusual in that arginine is their most prominent amino acid. Glutamic acid and valine predominate in bilberries, and lingonberries contain high levels of serine and aminobutyric acid. Lingonberries also possess appreciable levels of the unusual amino acids 1-aminocyclopane-1-carboxylic acid and 5-hydroxypipecolic acid.

In general, blueberries and bilberries are two of the richest sources of antioxidant phytonutrients among the fresh fruits, with total antioxidant capacity ranging from 13.9 to 45.9 μmol of Trolox equivalents per gram of fresh berry. Berries from the various *Vaccinium* species contain relatively high levels of polyphenolic compounds, with chlorogenic acid predominating. Total anthocyanins in blueberry fruit range from 85 to 270 mg per 100 g, and species in the subgenus *Cyanococcus* carry the same predominant anthocyanins, aglycones and aglycone sugars, although the relative proportions vary. The predominant anthocyanins are delphinidin monogalactoside, cyanidin monogalactoside, petunidin monogalactoside, malvidin monogalactoside, and malvidin monoarabinoside.

Among the other *Vaccinium* spp., cranberries have total anthocyanins varying from 25 to 100 mg per 100 g fruit, with the most important anthocyanins being cyanidin-3-monogalactoside, peonidin-3-monogalactoside, cyanidin-3-monoarabinoxide, and peonidin-3-monoarabinoside. Cowberries contain high quantities of cyanidin-3-galactosides. Bilberries contain high quantities of hydroxycinnamic acid and possess very high levels of quercitin-3-glucoside, rhamnoside, and arabinoside. The various Ericaceae species also contain appreciable amounts of several carotenoids.

The major volatiles contributing to the characteristic aroma of blueberry fruit are *trans*-2-hexanol, *trans*-2-hexanal, and linalool. The predominant volatiles in the bilberry are *trans*-2-hexanal, ethyl-3-methyl butyrate, and ethyl-2-methyl butyrate. In the cranberry, 2-methyl butyrate is rare, but α-terpineol predominates. Benzaldehyde also contributes to the aroma of the cranberry. (*See* **Sensory Evaluation:** Aroma.)

Handling and Storage

Most fresh-marketed fruits of *Vaccinium* are harvested by hand, although lowbush blueberries are commonly removed from the bush with hand-held rakes. Highbush and rabbiteye berries are mechanically harvested for the processed market with over-the-row machines that shake or beat the fruit on to catching pans or conveyors. Fruits are mechanically harvested as a clean-up operation after several hand-pickings, or when 60–70% of the fruit on a bush are

blue. A limited amount of fruit is mechanically harvested for the fresh market, but only with careful sorting to remove defective fruit.

Cranberries are mechanically harvested either dry or wet. Dry harvested fruit are generally stripped from the vines with a picking machine that combs or scoops fruit off the plant. To wet harvest fruit, the bogs are flooded to lift the vines, and the floating fruit are raked or beaten off the surface of the water. Wet harvesting recovers a higher percentage of the fruit, but water-raked berries deteriorate more rapidly.

A typical grading line for fresh-marketed blueberries consists of a 'blower' to remove leaf material and small green berries, a 'tilt belt' to separate soft from firm fruit, and a sorting line, where four to eight people visually scan the fruit for defects. Grading lines for processing berries are usually composed, in order, of a blower, tilt belt, a water tank, where ripe berries sink, a fruit destemmer, and a sorting line. Processed berries are generally frozen either in bulk or individually quick-frozen. The color of blueberries is well preserved by freezing, although they form exudates and redden upon thawing.

Dry-harvested cranberry fruit are generally sorted by bouncing them over a barrier board. Sound berries bounce and travel forward on a conveyor belt, and soft berries are collected below. Small and large berries are then separated over a wire screen and carried along moving belts for further sorting. Water-harvested berries are first dumped on a long, inclined mesh belt that passes through a dryer to prepare them for sorting. Flotation tanks are also used in some instances to sort both wet- and dry-harvested fruit.

Suggested quality standards for sorted blueberry fruit are as follows: (1) pH at 2.25–4.25; (2) citric acid at 0.3–1.3%; (3) soluble solids, greater than 10%; (4) ratio of soluble solids to acid, 10–33%; (5) firmness, greater than 7 g of force for 0.01 cm of deformation on the Instron testing machine; (6) diameter, greater than 10 mm; (7) color, blue with less than 0.5% of the surface having a pink coloration. Cranberry quality is largely determined by its color, particularly in juice products. Other important quality traits are sweeter, darker, uniform-colored fruit for fresh marketing, increased aromas, firmness, uniform size and shape, organic acids, and glossiness. Methods have been devised for sorting *Vaccinium* fruit by firmness and optical density, and these are growing in popularity.

Cranberry fruit can be successfully stored for several months without any significant losses in quality. Fruit were once stored for the fresh market in large ventilated rooms at ambient temperatures, but now refrigerated storage is recommended at 2–4 °C. Processed berries are generally frozen. Most cranberries are stored in bulk, but some berries are prepackaged in perforated 0.45-kg cellophane bags or unsealed cardboard boxes. Storage life can be extended up to 12 weeks if berries are kept cool in bulk and packaged just before harvest. Controlled-atmosphere storage does not appear to have significant benefits, except that treatment with ethylene gas after harvest may increase the anthocyanin content of the fruit. (*See* **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability.)

Blueberry fruit are generally packaged in plastic pint or quart 'clam shell' containers, or open containers that are covered loosely with cellophane or shrink-wrapped. Shelf-life can be greatly extended by lowering fruit temperatures to 0 °C (**Figure 1**). Decreasing the storage temperature slows respiration and other metabolic processes, but even more importantly, it greatly reduces the activity of decay organisms (**Figure 2**). In addition, spore germination of *Alternaria* and *Botrytis* occurs only very slowly at 0 °C, whereas germination of *Glomerella* spores is stopped altogether. During long-term storage (3 weeks or more) at 0 °C, senescent breakdown can occur in the berry flesh, leading to the formation of watersoaked areas and the bleeding of blue skin pigments into the normally colorless mesocarp. However, few precautions other than modest chilling

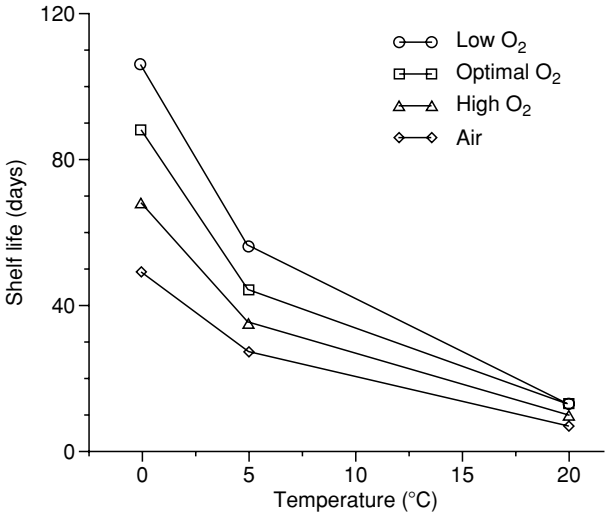

Figure 1 Effect of storage temperature and atmosphere composition on the visual shelf-life (days in marketable condition) of Bluecrop blueberry fruit sealed in low-density polyethylene packages. For the various oxygen (O_2) treatments, the approximate O_2 and carbon dioxide (CO_2) concentrations were, respectively, as follows: low O_2, 0.8% and 12%; optimal O_2, 2% and 7%; high O_2, 7% and 3%; air, 21% and 0%.

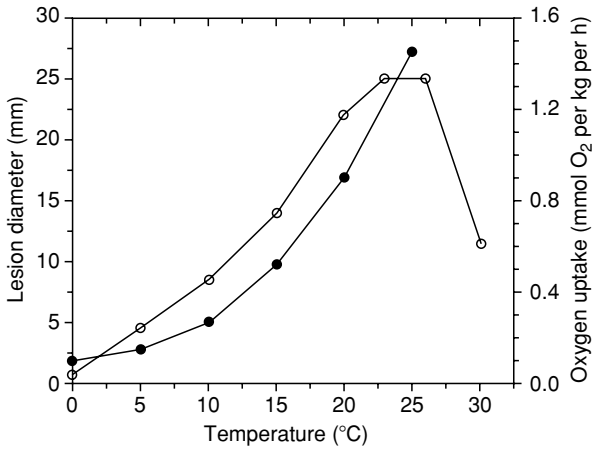

Figure 2 Effect of temperature on (○——○) the development of decay lesions (generalized curve redrawn from Kader AA, Kasmire RF, Mitchell FG, *et al*. (1985) *Postharvest Technology of Horticultural Crops. Extension Bulletin 311*. University of California at Davis.) and (●——●) respiratory metabolism of Bluecrop fruit at ambient oxygen levels. Adapted from Beaudry RM, Cameron AC, Shirazi A and Dostal DL (1992) Modified atmosphere packaging of blueberry fruit: effect of temperature on package oxygen and carbon dioxide. *Journal of the American Society of Horticultural Science* 117: 436–441, with permission.

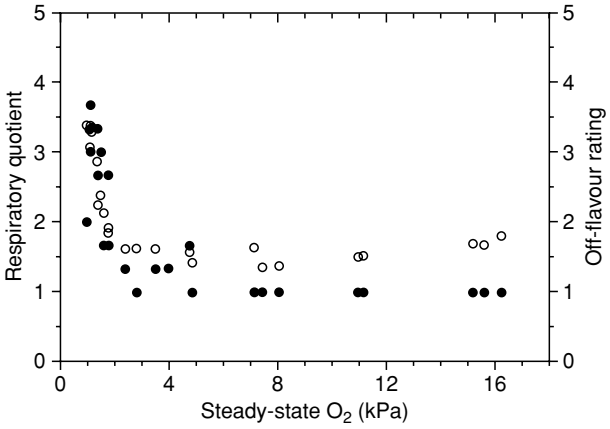

Figure 3 Effect of oxygen (O_2) concentration on (○) the respiratory quotient (RQ) and (●) off-flavor development of Bluecrop blueberry fruit at 20 °C. The increase in the RQ as O_2 levels decline below 2 kPa is indicative of fermentation.

are generally taken to attain a storage life of 7–10 days. Once the fruit have been cooled below approximately 10 °C, they must be kept chilled to avoid condensation ('sweat') upon returning the fruit to warm conditions.

Modified atmospheres (elevated CO_2 and low O_2 conditions) can enhance the visual shelf life of blueberry fruit (Figure 1), although temperature management plays the most important role in maintaining fruit quality. Lowering O_2 to 2% does not have a significant benefit by itself, but combinations of low O_2 and high CO_2 can improve storability. CO_2 levels of 8–10% or greater can significantly reduce decay development of blueberry fruit, but may not slow the decline in internal condition that normally occurs during storage. Packaging systems presently under development rely on the fruit, the packaging film, and temperature control to generate and maintain optimal O_2 and CO_2 levels. Packages must be designed to avoid excessively low O_2 levels, or fermentation and off-flavors will result (Figure 3; note that O_2 partial pressures are in kPa, which can be read as %). Packaging systems that boost package CO_2 levels and/or provide some level of humidity control have been developed for other products and may have considerable potential for packaged blueberry fruit. Chlorine and fungicide dips can also reduce postharvest decay. (*See* **Chilled Storage: Use of Modified-atmosphere Packaging**; **Chill Foods: Effect of Modified-atmosphere Packaging on Food Quality**; **Fungicides.**)

Industrial Uses

Blueberries and bilberries are eaten both as dessert fruits and in processed forms. About 46% of the rabbiteye crop and 50% of the highbush crop are marketed fresh, and the remainder are processed. Nearly all commercially harvested lowbush blueberries, cranberries, and lingonberries are processed.

The first widespread use of cranberries was to make sauce as a speciality item served at Christmas and the American holiday, Thanksgiving. During the 1960s, juice products made their appearance in the USA and now dominate the market. Cranberry 'cocktail' is drunk alone or mixed with other juice products. Cranberries are also made into a syrup, a dried raisin-like product, and a natural red food coloring, which has been used successfully to enhance the color of cherry pie filling.

Blueberries are used primarily in pie fillings, yogurts, icecream, and prepared muffin and pancake mixes. Blueberries are sometimes added to dried products after dehydration using an explosion-puffing process. Syrups, jams, and preserves are also produced, but in limited quantities. The juice of blueberries is rarely consumed directly as it has a very strong flavor and dark color.

Lingonberries are quite tart, but quite edible when cooked and are commonly used for juice, pie fillings, and jam. Bilberries are used fresh or in juice, preserves, or wine. Fruit extracts are also used in pharmaceutical preparations for the treatment of microcirculatory diseases.

See also: **Acids**: Natural Acids and Acidulants; **Chilled Storage**: Use of Modified-atmosphere Packaging; **Chill Foods**: Effect of Modified-atmosphere Packaging on Food Quality; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Fungicides**; **Ripening of Fruit**; **Sensory Evaluation**: Aroma; **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability

Further Reading

Ballington JR, Kirkman WB, Ballinger WE and Maness EP (1988) Anthocyanin, algycone, and aglycone-sugar content in the fruits of temperate North American species of four sections in *Vaccinium*. *Journal of the American Society of Horticultural Science* 113: 746–749.

Beaudry RM, Cameron AC, Shirazi A and Dostal DL (1992) Modified atmosphere packaging of blueberry fruit: effect of temperature on package oxygen and carbon dioxide. *Journal of the American Society of Horticultural Science* 117: 436–441.

Beaudry RM, Moggia CE, Retamales JB and Hancock JF (1998) Quality of 'Ivanhoe' and 'Bluecrop' blueberry fruit transported by air and sea from Chile to North America. *HortScience* 33: 313–317.

Eck P (1988) *Blueberry Science*. New Brunswick: Rutgers University Press.

Eck P (1990) *The American Cranberry*. New Brunswick: Rutgers University Press.

Hancock JF and Draper AD (1989) Blueberry culture in North America. *HortScience* 24: 551–556.

Hulme AC (ed.) (1970) *The Biochemistry of Fruits and their Products*. London: Academic Press.

Kader AA, Kasmire RF, Mitchell FG *et al*. (1985) *Postharvest Technology of Horticultural Crops. Extension Bulletin 3311*. University of California at Davis.

Kalt, W, Forney CF, Martin A and Prior RL (1999) Antioxidant capacity, vitamin C, phenolics and anthocyanins after fresh storage of small fruits. *Journal of Agriculture Food Chemistry* 47: 4638–4644.

Luby JJ, Ballington JR, Draper AD, Pliszka K and Austin ME (1990) Blueberries and cranberries (*Vaccinium*). In: Moore JN and Ballington JR (eds) *Genetic Resources of Temperate Fruit and Nut Crops*. Wageningen, The Netherlands: International Society for Horticultural Science.

Maas JL, Galletta GJ and Stoner GD (1991) Ellagic acid, an anticarcinogen in fruits, especially in strawberries: a review. *Horticultural Science* 26: 10–14.

Uphof JCT (1986) *Dictionary of Economic Plants*. Lehre, Germany: J Cramer.

Usher G (1974) *A Dictionary of Plants used by Man*. London: Constable.

Vander Kloet SP (1988) *The Genus* Vaccinium *in North America. Agriculture Canada Publication 1828*. Ottawa: Agriculture Canada.

Yarborough DE and Smagula JM (eds) (1996) Proceedings of the sixth international symposium on *Vaccinium* culture. *Acta Horticulturae* 446.

Factors Affecting Quality

W S Conway and M Faust, Beltsville Agricultural Research Center, Beltsville, USA
C E Sams, University of Tennessee, Knoxville, USA

Introduction

The primary goal of the fruit grower is to produce a product of such quality that it is attractive to the potential consumer in both appearance and taste. In order for quality fruit to reach the consumer, it must be properly grown in the orchard, harvested at the time of proper maturity, and stored in such a manner that the quality is maintained. Economic losses occurring in postharvest channels are greater than often realized, and the avoidable loss of quality 'between the farm gate and the consumer' is the cause for real concern. Fresh fruit increases several-fold in unit value while moving from the field to the consumer, because of the added cost of careful harvesting and handling. Factors that affect fruit quality can be classified into three groups: (1) nutritional factors during growth; (2) factors associated with the time of harvest; and (3) factors associated with storage. These factors, as they relate to the quality of the major temperate fruits, including apples, apricots, cherries, peaches, pears, and plums, will be discussed.

Tree Nutrition

In order for the fruit tree to produce quality fruit, it must be supplied with the proper mineral nutrients and photosynthates. Orchard trees grow and produce fruit in the same place for 15–50 years. By the time these trees are mature enough to produce a crop, they have developed extensive root and branch systems and are able to acquire nutrients from the soil, intercept light, and produce photosynthates. The perennial nature of the tree and the necessity of producing large crops of fruit regularly imposes demands for nutrition not found in herbaceous plants or forest trees. The tree has changing nutrient needs during the year, particularly when it sets and raises its fruit. These occasional high mineral nutrient needs may not be met by uptake through the roots and must be supplied by foliar application. Fruit nutrition must often be concerned with the nutrition of not only the whole tree but also individual organs. For example, fruit should have a certain nutrient content for maximum storability. This requires techniques other than nutrition for maintaining the tree. First, nutrients must be applied and the translocation and

Table 1 The importance of nutrient effects on fruit quality

Fruit	*Nutrients*					
	Boron	*Calcium*	*Magnesium*	*Phosphorus*	*Potassium*	*Photosynthates*
Apple	+−	+++	−	++	+	+++[a]
Apricot	+	+		+	+++	+++
Cherry					+	
Peach		+			++	+++[a]
Pear	++	++	−		+	++
Plum					+++	++

+Importance of deficiency levels; −the effect of toxicity conditions.
[a]Important during postbloom.

accumulation of these nutrients into the fruit must be understood and horticultural techniques applied to direct such nutrients into the fruit in quantities to attain the goal. Mineral nutrients that affect fruit quality include boron, calcium, magnesium, phosphorus, and potassium. A summary of known effects of nutrients on fruit quality is given in Table 1.

During the development of the tree, branches grow, and leaves intercept more and more light and produce more carbohydrates. At the same time, if the tree is not pruned properly and the inside is shaded, these shaded leaves are unproductive photosynthetically. In apples, one fruit must be supported by 25–30 leaves which are exposed to the sun to produce sufficient carbohydrates to result in sweet, red fruit. Shade-grown fruit are low in sugars and remain green. Small fruit, such as cherries, require less support by the leaves (1.5 leaves per fruit) than larger fruit. To some extent, carbohydrates should be considered as organic nutrients. If less than 75% of the light which is measured above the canopy reaches the leaves, red color will not develop. If the light is less than 50%, sugar accumulation in the fruit is decreased and fruit size will remain small. Similar to changing mineral nutrient demands, the need for photosynthates also changes. Photosynthates are especially needed soon after bloom and again when fruit accumulates sugars close to harvest. (*See* **Carbohydrates**: Classification and Properties.)

Calcium

Calcium is probably the most important mineral element determining fruit quality. It is especially important in apples and pears where it can prevent a number of metabolic disorders. The effect of calcium cannot be substituted by other factors. Calcium is also important in other fruits because it has the general effect of delaying ripening. Fruits high in calcium can be transported better and maintain their quality longer. The tissue calcium concentration at which desirable effects are achieved is usually higher than concentrations that the fruit normally accumulates. Therefore, calcium accumulation into the fruit must be forced by proper horticultural techniques. Calcium nutrition is complicated by the fact that calcium is needed most in the fruit. Consequently, calcium not only needs to be taken up by the tree but also needs to be directed to the fruit. (*See* **Calcium**: Properties and Determination; Physiology.)

Increasing the calcium content of fruit is difficult to achieve through normal fertilization. One method of increasing calcium in the growing fruit is to apply it to fruit trees as a spray. This spray must be applied to the fruit itself because calcium applied to the leaves will not translocate to the fruit or any other part of the tree. Because of dilution during growth, calcium concentration decreases depending upon the growth rate of the fruit. This is an important factor in determining storage quality. Large fruits usually have very poor storage quality. Small fruits, however, possess high calcium concentrations and store well.

The direct application of calcium to the storage organ would appear to be the most effective method for increasing its calcium content. The calcium content of apples has been increased by foliar sprays and postharvest dipping in calcium chloride solutions. Postharvest dipping with additives such as food thickeners in the dip solutions further increases calcium levels in apple fruit. Utilizing temperature differentials, especially dipping warm apples into a cold solution of calcium chloride, also results in additional uptake of the solution. Even more successful was vacuum or pressure infiltration of calcium chloride solutions into the fruit. In studies comparing treatment of Delicious and Golden Delicious apples with solutions of calcium chloride, the calcium concentration of fruit flesh was increased the most using pressure infiltration, followed by vacuum infiltration, with dipping being a distant third. In another study comparing treatment of peaches with calcium chloride solutions, preharvest spray applications were less successful than postharvest pressure infiltration.

Increasing the calcium content of fruit tissue makes the fruit more resistant to postharvest decay caused by soft-rotting fungi. This may be attributable to the fact that calcium ions bound to the pectins present in the cell wall increase the number of cation cross-bridges between pectic acids or between pectic acids and other polysaccharides with acid groups, and may make the cell wall less accessible to enzymes produced by fungal pathogens that cause decay. Fruit with sufficient or high calcium content maintain quality in storage better than fruit which are low in calcium. Calcium content of fruit flesh may be a factor when considering whether to store fruit from a particular orchard following harvest or sell it immediately. Calcium analysis at harvest is often used for making such decisions.

Potassium

Potassium is needed in relatively large quantities. When potassium levels are low, the fruit does not color normally. It is likely that insufficient potassium concentration decreases the photosynthesis of leaves, which in turn lowers sugar concentrations, and this affects the color. It is well established that cell extension is the consequence of the accumulation of potassium in the cells. Maintenance of high turgor pressure associated with potassium is needed for the enlargement of the fruit. This is especially important in peaches and nectarines which need an extra water supply during the last phase of fruit growth. Small peaches are relatively poor in quality. (*See* **Potassium**: Physiology.)

The accumulation of organic acids is often the consequence of unbalanced potassium transport. When protein synthesis is impaired, potassium is in excess, and acids are synthesized for charge balance. Organic acids and potassium also accumulate in the affected spots of the apple disorder Jonathan spot.

Phosphorus

Phosphorus is involved in the energy transfer mechanism, including the generation of adenosine triphosphate (ATP) and the formation of sugar and alcohol esters. In addition to these roles, phosphorus also has a regulatory function in many enzymatic processes where inorganic phosphorus controls the rate of reaction. (*See* **Phosphorus**: Properties and Determination.)

The uptake of phosphorus in fruit of the apple tree closely follows the weight increase of the fruit, and uptake continues until harvest. Phosphorus levels in apples have been positively correlated with fruit firmness and negatively with low-temperature breakdown; it is therefore important to insure high enough levels of phosphorus in the tree to avoid these disorders.

Magnesium

The fruit requires considerable amounts of magnesium. In the leaves of the apple tree the calcium concentration on a dry-weight basis is about five times that of magnesium, whereas in the fruit the magnesium concentration is twice as high as that of calcium. Starch usually accumulates in magnesium-deficient tissues (starch decomposition, sucrose formation, and phloem loading require energy and ATPase activity, which all depend on magnesium), and thus photosynthates are not partitioned into the fruit. Magnesium-deficient trees also usually produce small fruits. The overall photosynthesis of the tree is also severely affected by defoliation. Thus the small fruit size of magnesium-deficient trees may be attributable to more than one factor affecting photosynthesis. (*See* **Magnesium**.)

Some disorders of fruit may be accentuated or their development may be triggered by a high concentration of magnesium. All of these effects are known to be caused by either lower calcium concentration in the fruit or factors which counteract the beneficial effect of calcium. Bitter pit, the collapse of cell clusters usually after storage of the fruit, is generally prevented by sufficient calcium and accentuated by high magnesium.

Boron

Boron-deficiency symptoms are often noticed in the fruit before symptoms appear in the shoot. In apple, the mildest effect of boron deficiency is a flattening of the fruit. If the deficiency is slightly more severe, internal cork, i.e., round or irregular brown regions within the core area, is clearly visible upon cutting the fruit. The dead cell masses become dry, hard, and corky. In pears, similar brown areas are closer to the surface and, if they develop early, the surface above the spots is depressed. Plums show only malformed fruits without browning. If boron deficiency develops early, external cork develops on apple before the fruit is half-grown. In the early stages the areas appear water-soaked, after which they harden and crack. The fruit appears dwarfed and misshapen. (*See* **Boron**.)

There is very little margin between sufficiency and toxicity where boron is concerned. Boron toxicity manifests itself in early maturation of fruit, premature fruit drop, shortened storage life, and senescence breakdown in storage. Boron toxicity is often caused by boron in the irrigation water. Irrigation water should be free of boron.

Fruit Maturity

Fruits developing during the growing season usually reach a point when they are mature. If they are

Table 2 Relative effect of maturity at the time of harvest on fruit quality

Fruit	*Immature*	*Overmature*
Apple	---Poor flavor, metabolic diseases	---Storage breakdown
Apricot	–Poor flavor, low sugars	–Softening
Cherry	---Low sugars	–Softening
Peach	---Poor quality	–Softening
Pear	–Poor flavor	---Storage breakdown
Plum	--Poor flavor, high acids	–Softening

–Relative importance of maturity on fruit quality.

harvested at this time, they can maintain high quality in storage. If they are picked in an immature state, and only a few days separates immature from mature in the case of peaches or apricots, the quality of the fruit remains low. If they are harvested when they are overly mature, the storage quality is poor and breakdown may occur during storage. The transition from the mature to the ripe state is accompanied by a number of enzymatic processes. The fruit softens, the green color (chlorophyl) is destroyed, yellow under-color develops, acids are depleted to a large extent, and starch is converted to sugars. The rates of these processes depend upon the species, and the determination of harvest maturity is more or less critical in influencing the quality of the fruit. The relative importance of maturity on fruit quality is given in Table 2. (*See* **Ripening of Fruit**.)

The time at which fruit mature is generally determined by genetics. In most cases it is well known that fruit matures in a certain number of days after bloom depending upon the variety (e.g., apples mature 80–200 days after bloom; for McIntosh it is 123 days, Delicious 150 days, and Fuji 200 days). Other species follow similar patterns. The temperature during the first 3 weeks after bloom influences the time of maturity. If the temperature after bloom is warm, maturation time is shorter; if the temperature is cool, the time of maturation is longer.

Storage Conditions

Once fruit has been harvested at the right stage of maturity, it is important to maintain the quality of the fruit during storage. Postharvest losses are economically greater than equivalent losses in the field prior to harvest. Fruit with bruises, skin breaks, and other external distortions should not be placed in storage. Not only are these blemishes unsightly, but they also provide avenues of entry for postharvest decay-causing pathogens which will reduce fruit quality. Mechanical damage can also increase the amount of moisture loss during storage, again reducing the quality of fruit. Optimum maturity is a very important factor to consider when deciding which fruit to put in storage. Storage life of fruits can be reduced if they are harvested in an overmature or undermature state. (*See* **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability.)

Refrigerated storage retards senescence caused by ripening, softening, and textural and color changes. Refrigeration reduces unwanted metabolic changes, heat produced through respiration, moisture loss and resultant shriveling, and slows decay caused by postharvest pathogens.

Controlled atmosphere (CA) is a process or system which is used to supplement refrigeration in maintaining fruit quality in storage. It involves a change in storage atmosphere that replaces air with an atmosphere containing 1–2% oxygen, 1–5% carbon dioxide, and 93–98% nitrogen. In conjunction with refrigeration, CA can extend the storage life of apples and pears and result in quality of fruit beyond that of refrigeration alone. Controlled-atmosphere storage (CAS) cannot be applied to cherries, peaches, plums, or apricots. (*See* **Controlled-atmosphere Storage**: Effects on Fruit and Vegetables.)

Effect of Refrigeration and Controlled Atmosphere

In order to obtain maximum results from cold storage, it is necessary for the temperature in storage rooms to be held fairly constant. The best temperature at which to store most apple cultivars is between $-1\,^{\circ}C$ and $0\,^{\circ}C$. If the temperature rises 1 or $2\,^{\circ}C$ above the upper limit of $0\,^{\circ}C$, or if apples are not promptly cooled to the desired temperature, there is a real danger of increased decay and ripening. This danger increases as the period during which the temperature is above the optimum is extended. At 1.5–$2\,^{\circ}C$, 3 or 4 days would have little effect since the temperature of the fruit will not rise as quickly as that of the air, but 10 days at this temperature would reduce the storage life of the fruit by about a week and could result in greater losses owing to decay. Conversely, if the temperature falls several degrees below $-1\,^{\circ}C$ for apples, there is a chance that the fruit will freeze.

Table 3 Storage requirements and storage periods for fruits grown in temperate climates

Fruit	*Cold storage*		*Controlled atmosphere*	
	Temperature[a] *(°C)*	*Duration*[a]	*Temperature*[a] *(°C)*	*Duration*[a] *(months)*
Apple	−1 to 4	3–6 months	−0.5 to 4.4	7–9
Apricot	−0.5 to 0	1–3 weeks		
Cherry	−1 to 0	2–3 weeks		
Peach	−0.5 to 0	2–4 weeks		
Pear	−1.5 to 0	2–7 months	−1.5 to 0	7–9

[a]Depends on cultivar; range is given.

It is important to maintain a uniform temperature throughout the storage room. Temperature variation within the room could result in a mixture of overripe and prime fruit being removed from storage, as well as increased losses owing to decay in some areas of the room. A summary of storage requirements and average storage durations for various temperate climate fruits is given in Table 3.

Apples

Apple cultivars such as Delicious and Golden Delicious can be stored under proper cold-storage conditions for 5–6 months. Early or summer-maturing apples are more perishable then the later-maturing cultivars, and go directly to market rather than being stored for long periods. (*See* **Apples**.)

Refrigeration, however, is desirable for storing apples even for a short period. At 4 °C, apples respire and soften twice as fast as at 0 °C, and three times as fast at 15 °C as at 4 °C.

For the majority of apple cultivars, the optimum storage temperature is between − 1 °C and 0 °C in combination with 90–95% relative humidity (RH). Since the highest freezing point for apples is approximately − 1.5 °C, most apples can be stored at − 1 °C or above. However, a − 1 °C storage temperature requires well-controlled storage temperatures to avoid freezing.

As an example of the importance of proper storage temperatures, for Delicious apples, the cultivar grown in the greatest abundance in the USA, storage at − 1 °C will extend storage life approximately 25% over fruit stored at 0 °C.

While the more popular cultivars such as Delicious and Golden Delicious apples can be stored at − 1 °C to 0 °C, some other cultivars, owing to a greater susceptibility to low-temperature disorders, should be stored at somewhat higher temperatures. Cultivars such as Jonathan and McIntosh should be stored at 2 °C.

CAS, as previously stated, uses refrigeration in a gas-tight room, where the atmosphere within is controlled so that this atmosphere is higher in carbon dioxide and lower in oxygen than normal. Apples stored in CA respire, ripen, and soften more slowly and have fewer disorders than those in air storage. This results in a longer storage life and longer shelf-life after removal than fruit stored in air. These fruit, however, must be harvested at the proper maturity and be of the highest quality to realize the benefits resulting from CAS. CA rooms in which apples are to be stored usually contain an atmosphere of 1.5–3% oxygen, 1–8% carbon dioxide, the remainder being nitrogen. The exact combination of gases is cultivar-dependent. The storage life in CA for McIntosh, for example, may be nearly double that in regular cold storage. The use of CA also allows the fruit to be stored at slightly higher temperatures with good results. Cultivars such as Jonathan and Yellow Newton, which may develop low-temperature disorders at 0 °C in regular cold storage, may be stored at 2 °C and still retain excellent quality. In addition, large quantities of Delicious and Golden Delicious, the most popular apple cultivars, are stored in CA, at − 0.5 °C to 0 °C.

More recently, even better quality retention has been realized with a lower oxygen level of 1% rather than the 2–3% used for McIntosh, Delicious, and Golden Delicious apples. Oxygen at this level, however, must be maintained very precisely at 1%, because when oxygen drops below 1%, injury to the fruit may occur.

Apricots

Although apricots are seldom stored commercially for long periods of time, they may keep well up to 3 weeks at −0.5 °C to 0 °C, with a RH of 90–95% being desirable. Unfortunately, fruit picked when they are mature enough to ship or store are at the same maturity as those used for canning and lack the taste of tree-ripened fruit. The biggest hazard of shipping and handling apricots is decay, and since mature fruit is more susceptible to decay than immature fruit, it is not possible to allow the fruit to reach the stage of maturity for fresh-market use. (*See* **Apricots**.)

Cooling apricots quickly to 4 °C immediately after harvest and then holding them at 0 °C will retard decay and ripening. If the storage temperature is

kept at 4–7 °C, the fruit will have a more 'mealy' texture after ripening then fruit kept at 0 °C.

Cherries

Fresh sweet cherries have a handling limit from harvest to arrival at the market of about 2 weeks if temperatures do not exceed 2 °C. If sealed polythene liners are used in containers, the cold-storage period can be extended by about a week. Sweet cherries covered with polythene and stored at −1 °C to 0 °C for 2 weeks after harvest (90–95% RH is also desirable) will then have a week remaining for distribution in marketing channels. If sweet cherries are held longer than the 2–3 weeks indicated, there is a loss of flavor and brightness. Moisture loss is a critical factor detracting from the fresh appearance of both the fruit and the stems. Cherries harvested in the field should be covered with a moisture barrier at harvest until packaged for shipment. Sweet cherries are extremely perishable and rapid deterioration occurs if the fruit are not refrigerated. Sour cherries are usually unsuitable for storage. They are harvested directly into ice water for immediate cooling and may be held at 0 °C for several days to lengthen the processing period. (*See* **Cherries**.)

Peaches

Peaches are usually stored for only a short period and usually only to escape an overabundance on the market at a given period, or to extend their processing life. Peaches harvested at the proper maturity can usually be stored at −0.5 °C to 0 °C for 2–4 weeks. Early-maturing cultivars have a shorter storage life, while the later-maturing cultivars may be successfully stored for 3–4 weeks. If peaches are stored for longer than 3–4 weeks in cold storage, they often fail to ripen when moved to higher temperatures. The flesh of these fruit may become dry and mealy, or wet and mushy, and may brown internally around the stone. There is a deterioration of flavor, and a dull appearance results. This malady is called internal breakdown. Storage of peaches at 2–5 °C for 7–14 days results in internal breakdown. (*See* **Peaches and Nectarines**.)

Rapid cooling of peaches to 4 °C at harvest is necessary to retard ripening and decay. Brown rot, a fungal disease of peaches, can cause large losses in storage, and refrigeration is an effective method of retarding the development of this disease.

Pears

Pears are very sensitive to the temperature at which they are stored. Most pears in the USA are grown in the Pacific North-west, and are stored at − 1 °C with a RH of 90–95%. The recommended storage temperature for pears is − 1.5 °C to 0.5 °C with a RH of 90–95%. As an example of the sensitivity of pears to storage temperature, Anjou and Bartlett (Williams) pears have approximately a 40% longer storage life at − 1 °C than at 0 °C. It is necessary to maintain precise temperature control at these low temperatures to prevent freezing. It is also necessary to cool the fruit rapidly (remove field heat) to store these fruit successfully for long periods. When pears are being cooled, temperatures as low as − 3.5 °C to − 2 °C can be used, but should then be raised to − 1 °C as the fruit temperatures approach − 1 °C. Since pears also lose moisture rapidly, an RH of about 90% is desirable. (*See* **Pears**.)

Proper maturity at harvest is also conducive to long storage life. If pears are harvested at full maturity, they are less likely to be susceptible to physiological disorders and will ripen properly after storage. However, pear fruit gain much of their weight close to harvest. If growers delay harvest by a week, they may gain 1–2 (US) tons of fruit per acre (2268–4535 kg ha^{-1}), but storage quality will be poor.

The length of time for which pears can be safely stored at − 1 °C varies with cultivar. Holding pears beyond their normal storage period may result in fruit which will not ripen properly. Anjou pears can usually be safely stored for 6–7 months, while pears of the Bartlett (Williams) or Bosc cultivars can only be stored for 3–4 months.

CA, as with apples but to a lesser extent with pears, has been used successfully to extend the storage life of pears and to maintain a greater capacity for ripening. It is generally considered that 2–2.5% oxygen and 0.8–1.0% carbon dioxide is the optimum atmosphere in which to store pears safely.

Plums

In general, plums, including fresh prunes, are not well adapted to long cold-storage periods. Storage period, however, differs according to cultivar. The damson-type cultivars store better than the softer, oriental cultivar fruit. Well-matured fruit of the firmer varieties can be safely stored for 4–5 weeks at − 0.5 °C to 0 °C. Although ripening proceeds very slowly at these temperatures, there is some loss in flavor. If stored beyond 5 weeks, flesh browning and abnormal flavors often result.

See also: **Apples**; **Apricots**; **Boron**; **Calcium**: Properties and Determination; Physiology; **Carbohydrates**: Classification and Properties; **Controlled-atmosphere Storage**: Effects on Fruit and Vegetables; **Magnesium**; **Peaches and Nectarines**; **Pears**; **Phosphorus**:

Properties and Determination; **Potassium**: Physiology; **Ripening of Fruit**; **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability

Further Reading

Conway WS (1989) Altering nutritional factors after harvest to enhance resistance to postharvest disease. *Phytopathology* 79: 1384–1387.

Faust M (1989) *Physiology of Temperate Zone Fruit Trees*. New York: John Wiley.

Hardenburg RE, Watada AE and Wang CY (1986) *The Commercial Storage of Fruits, Vegetables, and Florist and Nursery Stocks*. US Department of Agriculture, Agriculture Handbook 66. Washington, DC: US Government Printing Office.

Improvement and Maintenance of Fruit Germplasm

R M Brennan and S Millam, Scottish Crop Research Institute, Invergowrie, Dundee, UK

Background

Fruit crops have been cultivated for centuries, both commercially and in amateur gardens, and many species covering a wide range of trees, shrubs, and nonwoody perennials have been the long-term subjects of germplasm improvement programs. For example, different varieties of apples (*Malus* spp.) were recorded by the ancient Greeks, and selection and breeding of apple have continued since then. In contrast, the cultivated strawberry *Fragaria* × *ananassa*, an octoploid interspecific hybrid, was developed in the eighteenth century, and the highbush blueberry *Vaccinium corymbosum* was only domesticated in the twentieth century.

Fruit crop species have certain common characteristics: they are long-lived, usually highly heterozygous, and clonally propagated from an elite mother plant. They are also invariably economically minor crops compared with large-scale arable species such as cereals or potatoes, as reflected in the relative research and development resources available. However, fruit is increasingly recognized as an important component of a healthy diet, and consumption of temperate fruit and fruit products, especially juices and dairy derivatives, has risen sharply over the past decade.

Fruit Germplasm Improvement

Breeding Objectives

Breeding objectives in fruit have traditionally concentrated on improving yield or reducing the influence of abiotic factors such as low temperatures on yield. However, resistance to pests and diseases is increasingly important, particularly as chemical controls are reduced on environmental and toxicological grounds. There is also a rising demand for fruit of a defined quality in terms of biochemical characters (content of flavonoids, vitamins, soluble solids, etc.) as well as physical characters (fruit size and shape, skin strength, juice yield). The move towards healthier diets worldwide, but especially in western Europe and the USA, has impacted strongly on breeding and selection priorities, and screening of germplasm, including potential donor species, for nutritional attributes such as antioxidant activity has been reported in several fruit crop species.

Other breeding objectives such as machine harvestability, storage potential, environmental adaptability, and precocity of cropping remain important for most breeding programs.

Financial support for the great majority of fruit breeding programs in western Europe has in the last 15 years moved from the public to the private and commercial sectors. The products from these programmes are more exactly aligned with the sponsors' specific requirements, especially regarding fruit quality, although much of the underpinning genetics research remains in the public domain.

Breeding Methodology

The breeding of new fruit cultivars, as with most clonally propagated crops, begins with the selection of parental genotypes with complementary characteristics suitable for the achievement of the breeding objectives. After hybridization between the parental types, promising phenotypes are selected from the segregating F_1 population, multiplied vegetatively through further screening cycles and eventually released as cultivars. The timescale over which this process operates is substantial: in small fruits, the time from initial hybridization to commercial release is in the order of 15–20 years, and this may be even longer for top fruit. It is therefore vital that (1) the most appropriate hybridizations are made in the first instance, (2) the progenies produced are sufficiently large to allow enough variation for selection of the desired phenotypes, and (3) the most efficient and timely selection criteria are used to select out appropriate phenotypes. The latter process can be

significantly assisted by the use of molecular-marker technologies (see below).

The majority of fruit crop species, particularly tree fruit, produce little or no fruit during the first 2–5 years at least, and these early crops cannot in any case be used as an accurate guide to future cropping worth. It is usual to use this juvenile period for the assessment of vegetative characters such as plant habit and pest/disease resistance. It is clearly important that desirable genotypes be selected as early as possible to maximize the efficiency of the breeding program. The increased emphasis on development of molecular markers will potentially give major advances in terms of early selection of desired segregants from a seedling population. Also, there are sometimes morphological markers that can be used, for example, in *Rubus* where progenies can be screened out at the cotyledon stage for the presence of glandular hairs linked to the later development of undesirable spiny plants.

Once the plants are producing flowers, the evaluation of fruit quality and cropping potential can begin. Promotion of earlier flowering by training to a single stem, glasshouse culture, and high-density planting has been achieved in apple, while grafting of seedling scions on to mature trees has been used extensively to hasten fruiting in both apple and grape.

Interspecific Hybridization

For many fruit crop species, the use of interspecific hybridization is increasing, in order to utilize naturally occurring sources of pest and disease resistance, fruit-quality components, etc. within the available germ plasm (Figure 1 shows a typical hybrid) Some examples are:

- blackcurrant (*Ribes nigrum*): *Ce* gene for resistance to *Cecidophyopsis ribis* introgressed from gooseberry *R. grossularia*;
- highbush blueberry (*Vaccinium corymbosum*): reduced chilling and greater soil pH adaptability from *V. darrowi*, dwarfing habit from *V. angustifolium*;
- apple (*Malus* × *domestica*) – improved winter hardiness from *M.* × *asiatica* and *M. pumifolia*.

The ease with which hybrids can be made across species barriers varies considerably, and the range of techniques for achieving this is described below. Many of the domesticated fruits that are presently in existence are originally hybrids between species, e.g., apple (*Malus* × *domestica*), where several species have contributed to present-day cultivars, strawberry (*Fragaria* × *ananassa*), which is a hybrid between the North American species *F. chiloensis* and *F. virginiana*, and highbush blueberry (*Vaccinium corymbosum*), which is a tetraploid hybrid formed from several diploid races of *Vaccinium* from the southern USA. In many genera, notably *Fragaria* and *Vaccinium*, there are ploidy barriers to interspecific hybridization, but nevertheless there is considerable interest in breeders for overcoming these problems, for example, in the creation of synthetic octoploids in *Fragaria*.

For top fruit species (*Malus* spp., *Pyrus* spp., *Prunus* spp.), there is also considerable effort devoted to rootstock breeding. The earliest progenies of these fruits were grown from seed, and the best selections were then propagated on seedling rootstocks of

Figure 1 (see color plate 67) Interspecific hybridization in *Rubus* breeding (*R. phoenicolasius* is used as the donor of resistance to raspberry beetle (*Butyrus tomentosus*)), showing the wild accession (left), a commercial red raspberry (right), and the resultant F_1 hybrid.

the same or closely related species. Now, the variable seedling rootstocks have been supplanted by the selection and breeding of uniform, often dwarf, rootstocks, and this type of program is of key importance to the future development of these fruits.

Genetic Resources and Genebanks

The elimination in the wild of potentially valuable species material will inevitably have serious repercussions for fruit breeders unless there is adequate conservation of representative genotypes. Most temperate fruits have centers of origin in the northern hemisphere, and it was suggested by Vavilov in the 1930s that wild species show the most genetic variation around or near their center of origin. These centers of diversity, both primary and secondary, are crucially important in the conservation of genetic resources, and the collection of relevant germplasm is seen as a priority by many breeders and researchers. In fruit germplasm collecting expeditions, it is now recognized as important for the wild populations to be both sampled for specific characters relevant to existing needs within the breeding program but also randomly sampled, for less predictable needs in the future.

In addition, there are several more recently domesticated fruits where accessions from the wild have formed the basis of commercial production; for example, some of the first blackcurrant (*Ribes nigrum* L.) cultivars released in Scandinavia were superior selections from the wild, and present day cranberry (*Vaccinium macrocarpon*) production in the USA relies heavily on cultivars selected from the wild – a total of 132 wild selections compared to only seven from controlled hybridization.

Germplasm collections must contain adequate representation of all important taxa and, where appropriate, specific ecotypes and accessions that may possess unique and important characteristics that can be utilized in long-term plant improvement programs. The ability to access new sources of important characters, and to generate sufficient variation in progenies to select out improved seedlings of commercial value, is key to the future of fruit breeding.

Plant genetic resources can be maintained *in situ*, whereby whole environments are conserved and the plants therein preserved through the protection of their habitat. For the maintenance of entire populations, this method of conservation may be appropriate. However, for accessibility, it is invariably preferable to have the germplasm maintained in an *ex situ* form.

Ex situ genebanks, where diverse genotypes are collected and maintained, usually at research centers, in field or laboratory conditions, can provide both a core collection of representative accessions and an active collection that can be easily utilized by the breeder. To be of significant value, material within the collection must be fully phenotyped and characterized, and increasingly, molecular data on each accession are collated.

For clonally propagated material like fruit germplasm, it is best to keep the material as an active field-based collection, or clonal repository, although this has associated problems in terms of cost (land use, culture, labor), pests, and pathogens (causing damage and potential loss of accessions and hence erosion of diversity) and manageability (only relatively low numbers of accessions, particularly of large plants such as fruit trees, can be maintained compared to seed banks).

The use of seed banks is of limited use in clonally propagated crops such as fruits, but it can be important for the storage of wild populations, since often the preferred method of collection in the wild is to obtain fruits of the desired accessions to avoid removing the entire plant. Also, many breeding programs maintain seed of controlled hybridizations, in most cases to allow further populations from interesting crosses to be planted out.

In vitro collections are predominantly used for the long-term storage of base collections of representative genotypes. The plant material can be cryopreserved in liquid nitrogen at $-196\,^{\circ}C$, by which all the normal plant processes are suspended. Systems for successfully cryopreserving a range of plant parts (apical meristems, cell cultures, shoot sections) of several fruit crop species have been developed, but cryopreservation is best considered as a complementary method for situations where other conservation strategies are not appropriate.

The use of *in vitro* storage of plant accessions over short- to medium terms in conditions where growth is reduced, usually at *c*. $4\,^{\circ}C$, is another potentially important method for the maintenance of valuable germplasm. By reducing the temperature at which *in vitro* cultures are grown, often with the presence of osmotic or growth regulatory inhibitors in the culture medium, subculturing intervals can be extended from a few weeks to periods of a year or more. However, the effects on stability have not been quantified for many species.

Overall, *in vitro* techniques are potentially useful for the maintenance of fruit germplasm, but considerably more research is required on long-term stability and the risks of mutations before they can be routinely used. The latter is now fairly easily monitored using molecular techniques, although the risk of mutations varies with species also. The use of

in vitro genebanking has similar requirements in terms of replication and plant health as conventional genebanks maintained in the field or as seed.

Future Breeding Requirements and Prospects

Fruit production in some parts of the world, notably Asia, is presently in a period of significant increases, particularly for apples, pears, and strawberries. However, fruit breeding is not showing a similar increase, either in number of breeding programs or available resources, and in general, there is a trend towards fewer major cultivars worldwide. A highly restricted range of cultivars are grown in a range of locations; for example, cv. Cox apples can be obtained from the UK, continental Europe, and New Zealand, and Elsanta strawberry accounts for the majority of production throughout Europe. This leads inevitably to a loss of local cultivars and a contraction of the genetic base for the crop in general. The breeders' ability to produce new and improved cultivars, responding to developing market requirements, is thus impaired.

The increasing interest worldwide in reduced input production systems, often incorporating aspects of integrated pest management, together with a need to drive down production costs, has led to an increasing emphasis on the development of pest- and disease-resistant cultivars of fruit. Also, processors and marketing groups are demanding specific fruit-quality traits in the available fruit, and the achievement of these objectives through plant breeding represents a major challenge to breeders into the future.

Biotechnological Approaches to Fruit Crop Improvement

Conventional crop-breeding techniques have enabled major advances to be made in the yield and quality of temperature fruit crops this century. However, in recent years, the impact of a range of techniques based on plant tissue culture, recombinant DNA technology, or a combination of both has become realized. This has facilitated the more precise and targeted genetic enhancement of a number of temperate fruit crop species to date and offers the potential for future wider application.

Plant Tissue Culture

The concept of totipotency (first postulated by Schwann and Schleiden in 1838) is unique to plant cells. The capacity for individual cells to regenerate into intact plants underpins all plant tissue culture manipulations. Fruit crops, historically, were among the first species to be investigated in cell culture, with the culture of grapevines *in vitro* (albeit in a relatively unsophisticated form) reported as early as 1944. Most temperate fruit species have now been established in tissue culture, and these systems are of significant economic and scientific importance. Problems associated with the establishment of, particularly woody, fruit germplasm *in vitro* include the long reproductive cycle of the plants, which restricts the availability of the juvenile tissue necessary for the establishment of a tissue culture system to a short time period. However, the advent of increased knowledge in this area in recent years has resulted in the range of species amenable for such technology being widened.

Micropropagation Propagation of plants under sterile conditions, in optimized environmental conditions, at a high density is known as micropropagation. The principal advantage is the provision of clean, clonal material in high volumes at greatly accelerated rates of proliferation compared to conventional propagation. Simple and cost-effective methods for the micropropagation of such species as strawberry, grape, kiwifruit, apple, raspberry, and blackcurrant are in widespread use for the routine commercial production of plants, regardless of climatic or seasonal constraints.

Associated advantages of the use of sterile micropropagation techniques include the potential for cleaning up and rejuvenating diseased or damaged plant material. This has increasing importance, for the provision of clean starting material for application in breeding programs or as high-quality stock for initiating conventional propagation. Meristem culture, in which the apical 12–20 cells are cultured *in vitro*, sometimes with associated heat treatment, is a valuable method for the elimination of viruses from fruit germplasm and has been widely used in the production of virus-free lines of apple, strawberry, and other fruits. A range of other applications can be facilitated by the use of micropropagation. These include the conservation of unique or valuable material *in vitro*, the relatively easy storage of material for programmed cropping systems, and enabling the transport of material in a cost-effective form.

Regeneration from tissue explants Micropropagation is based on the use of preexisting buds, e.g., shoot tips or axillaries. However, methods exist for the *de novo* initiation of shoots from explants such as leaf discs of apples and strawberries, stem sections of *Ribes*, nonclonal material such as immature embryos and cotyledons of *Prunus* spp. (peach, cherry, and apricot), and immature achenes of *Fragaria*. Often, a callus phase involving de-differentiated cells is involved prior to the regeneration of shoots, generally mediated by the application of exogenous growth

regulators. The callus phase provides a possible source of induced variation and novel genetic recombinations in the form of somaclonal variation. This can occur spontaneously or can be induced by the application of mutagens such as γ-irradiation to the initial material. While, in the context of maintenance of a germplasm collection, this type of variation is highly undesirable, in a germplasm enhancement context, there are precedents for its use. These include inducing variation and screening regenerant material of peach germplasm for bacterial resistance, and the production of dwarf cherry rootstocks by applying either γ-irradiation or the chemical mutagenizing agent, colchicine, to *in vitro* cultures to generate novel variants.

Other techniques Several other plant tissue culture-based methodologies are available for the improvement and maintenance of fruit germplasm. These include:

1. Somatic embryogenesis, i.e., the initiation and development of embryos from vegetative cells. This approach has been demonstrated in both grape and apple.
2. Gametic embryogenesis, i.e., the regeneration of plants from cells derived from the gametophytic cycle of the plant, usually pollen or microspores. This technique is a potentially powerful breeding tool, as plants derived from this system would be haploid (or half the native chromosome number), and subsequent chromosome doubling would produce homozygous lines. This represents a considerable time-saving from the many generations of inbreeding required to achieve this by conventional crossing and selection methodology. The method has been demonstrated in *Vitis*, *Prunus*, and *Pyrus* species, and very recent work has successfully applied both anther culture and microspore systems to apple.
3. Cell-suspension cultures can be established from calluses of plants maintained in liquid medium in various forms of rotating vessel. These systems offer great potential for the production of specific compounds such as specific flavorings and colorings for use in the food-processing industry. Of even higher value is the use of such cell-culture systems for the production of secondary metabolites and pharmaceuticals derived from selected plant cells.

Protoplasts Protoplasts (individual plant cells that have been isolated and had their cell wall removed enzymatically) offer an alternative method for genetic manipulation. Earlier work in fruit crops was limited to *Vitis*, *Prunus*, *Pyrus*, and *Malus*, but recent work has been described in a wide range of soft fruit species such as *Ribes* and *Rubus*. Possibly the most significant application of protoplast technology in plant germplasm improvement is in the area of protoplast fusion (somatic hybridization) where hybrid constructions can be made between taxonomically distant species beyond the normal limits of sexual compatibility, and cells with novel nuclear and cytoplasmic construction created. During the last 15 years, numerous reports of successful somatic hybridization at the intraspecific, interspecific, and even intergeneric levels have been published. Although the applications of the techniques in fruit breeding to date have been more limited than in other species, reports of fusions between *Rubus* species and between *Pyrus* and *Prunus* (cherry) protoplasts have been made. The potential for generating novel genetic combinations remains high, however, with useful targets including the introgression of disease resistance from incompatible wild species into cultivated lines. The use of protoplast fusion in creating interspecific *Malus* hybrids is considered by some apple breeders as an important prospect for future development of apple cultivars.

Gene Transfer

Although protoplast fusion enables the potential for whole or partial genome transfer between plants, a more targeted approach for the introduction of known genes or nucleotide sequences into plants is mediated by the application of a range of methods, generically described as transformation technologies.

Agrobacterium-mediated transformation The most widely used method for the transformation of higher plants is by utilization of the natural gene transfer characteristics of *Agrobacterium*. *Agrobacterium tumefaciens* is a naturally occurring soil bacterium that readily infects a large number of dicotyledonous plants. It is unique in that the bacterium integrates its own DNA into the plant genome. Since the first report of the successful transformation of a plant in 1983, over 120 plant species have since been transformed, including a significant number of fruit species. These include *Malus*, *Fragaria*, and *Vitis* species. Transgenic plants of *Malus* have been shown to display stable patterns of expression and Mendelian segregation of transgenes. Modern vectors for transformation have been refined to minimize the amount of foreign DNA incorporated into the recipient plant. Many of the most desirable traits for improvement in fruit crops such as increased yield, better growth habit, and low-temperature hardiness are polygenic (i.e., controlled by more than one gene) and as such, owing to the limitations in gene transfer,

may not be readily improved by transformation technologies. However, single gene traits can be readily introduced into fruit plant species. Examples include:

1. Insect resistance: gene transfer programs have been reported for the insertion of the gene sequence coding for the expression of *Bacillus thuringiensis* (Bt) toxin into *Malus* species. Improved levels of pest resistance have been reported in several studies using this strategy. In addition, the cowpea protease trypsin inhibitor (*CpTi*) gene, encoding for an antimetabolite to a wide range of lepidopteran and coleopteran pests, has been inserted into several fruit species, and the results of field trials of such material have been promising.
2. Virus resistance: the introduction into the plant of part of the nucleotide sequence of the viral genome, e.g., coat-protein sequence has been demonstrated to confer resistance to viral infection. The mechanism is thought to be in much the same way as field-grown plants may be protected against a second viral infection through the cross-protection response afforded by an earlier viral attack, as demonstrated for tobacco mosaic virus. The expression of coat protein genes integrated into the plant via transformation has been shown to result in an elevated degree of protection against viral infection in tomato and other species. The introduction of an *Arabis* mosaic virus coat-protein gene into certain soft fruit species is currently being undertaken, and reports of the implementation of pathogen-derived resistance to plum-pox virus have also been made.
3. Fruit quality: a range of the component factors associated with fruit quality, such as sugar content and various ripening processes have been the subject of a number of recent gene-transfer research programs.

In recent years, not only has the host range of species amenable to *Agrobacterium*-mediated transformation been widened, but also alternative systems for gene delivery and transfer have been developed. These include the use of a helium-powered biolistic gun to propel plasmid DNA on gold bullets into plant cells. This technique is widely used for recalcitrant plant species and assists in overcoming problems both with *Agrobacterium* infectivity and in the necessity of efficient tissue regeneration systems. The main problems precluding the uptake on a wide scale of transgenic crops are probably not technology-based but related to the issues of public acceptance of transgenic products. The scope and potential of transgenic crop technologies are extremely wide, but each case must be considered on its merits.

Molecular Markers

Methods for the identification, characterization, and quantification of genetic variability have enormous potential application in germplasm enhancement programs and can, if applied successfully, offer great timesaving benefits. Previously, the use of isozyme analysis in fruit for genotype identification has proved successful in species such as apple, pear, and cherry, but less successful in species where polymorphism is lower. Isozymes tend to have relatively low genomic coverage and do not always provide sufficient polymorphism to prove useful in genetic analysis. However, the last few years have witnessed a dramatic rise in the sophistication and applicability of a range of DNA-based molecular markers.

The rapid development of genomics technologies, particularly those based on the polymerase chain reaction, has enabled significant progress to be made recently in both genetic mapping and the identification of diagnostic molecular markers that can facilitate breeding and selection of specific characters. The marker systems of choice are now amplified fragment length polymorphisms (AFLPs), simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs). Major advances have been made in the mapping of specific fruit genomes, notably apple but also *Rubus*, *Ribes* and *Vaccinium*, and in the development of markers linked to both single gene traits, such as some pest resistances, and more complex quantitative traits, particularly those relating to fruit quality. The use of these markers in fruit breeding programmes will increase in future and should greatly assist the speed, focus and productivity of fruit germplasm improvement.

See also: **Fruits of Temperate Climates**: Commercial and Dietary Importance; Fruits of the Ericacae; Factors Affecting Quality; Improvement and Maintenance of Fruit Germplasm; **Fruits of Tropical Climates**: Commercial and Dietary Importance; Fruits of the Sapindaceae; Fruits of the Sapotaceae; Lesser-known Fruits of Africa; Fruits of Central and South America; Lesser-known Fruits of Asia; **Nucleic Acids**: Physiology

Further Reading

Abbott AJ and Atkin RK (eds) (1987) *Improving Vegetatively Propagated Crops*. London: Academic Press.

Anon. (1998) *The State of the World's Plant Genetic Resources for Food and Agriculture*. Rome: Food and Agriculture Organization of the United Nations.

Benson EE (ed.) (1999) *Plant Conservation Biotechnology*. London: Taylor & Francis.

Eck P (1990) *The American Cranberry*. New Brunswick, NJ: Rutgers University Press.

Janick J and Moore JN (eds) (1996) *Fruit Breeding* (3. vols.). New York: John Wiley.

Millam S and Spoor W (1999) Biotechnology and crop improvement. In: Simmonds NW and Smartt J (eds) *Principles of Crop Improvement*, 2nd edn, pp. 285–306. London: Blackwell Science.

Moore JN and Janick J (eds) (1983) *Methods in Fruit Breeding*. W. Lafayette, In: Purdue University Press.

Roache FA (1985) *Cultivated Fruits of Britain. Their Origin and History*. Oxford: Blackwell Scientific.

Smartt J and Simmonds NW (1995) *Evolution of Crop Plants*, 2nd edn. Harlow, UK: Longmans.

Stalker HT and Chapman C (eds) (1989) *Scientific Management of Germplasm: Characterization, Evaluation and Enhancement*. Rome: International Board for Plant Genetic Resources.

FRUITS OF TROPICAL CLIMATES

Contents

Commercial and Dietary Importance

S J R Underhill, Horticulture Postharvest Group, Hamilton, Queensland, Australia

Background

Tropical fruit are a botanically diverse group of fruit indigenous to tropical regions. Their representatives come from numerous families, including Anacardiaceae (mango, hog plum, imbu), Sapindaceae (rambutan, taun, lychee, longan), Passifloraceae (passion fruit), Bromeliaceae (pineapple) and Annonaceae (custard apple, soursop, sugar apple). Traditionally, tropical fruit were consumed locally, either grown in a subsistence-style agricultural structure or gathered from the wild. (*See* **Mangoes; Pineapples.**)

Increasing production, a more efficient transportation system, and refrigerated storage have led to increased global consumption in recent years. Considerable quantities of tropical fruit are now exported each year to European, Japanese, and American markets. The vast majority of this production is consumed fresh, although processing industries have also been established. In an increasingly health-conscious population, tropical fruit are seen as an appropriate source of nutrition, added variety and exotic appeal.

Production

Tropical fruit production is a large, rapidly expanding industry (**Table 1**). Bananas and pineapples, particularly from the Philippines, China, Mexico, and Brazil, represent the majority of world trade in tropical fruit. Collectively, these countries produced over 12×10^6 t of bananas and 3×10^6 t of pineapples in 1988. Although world production figures continue to rise, other fruit, such as mango, papaya, and lychee, are becoming increasingly popular. The subtropical regions of the world, particularly Colombia, Florida,

Table 1 Tropical fruit production for fresh consumption in selected growing regions in 1988

	Banana ($\times 10^3$ t)	*Mango* ($\times 10^3$ t)	*Pineapple* ($\times 10^3$ t)	*Papaya* ($\times 10^3$ t)	*Lychee* ($\times 10^3$ t)
Australia	159	12	154		2
Bangladesh	690	160	150		
Brazil	5139	400	1016	1600	
China	2350		430	95	62
Colombia	1300		230	44	
India		9450	790	339	92
Indonesia	1860	410	350	270	
Kenya	148		191		
Malaysia	490		199		
Mexico	1080	780	293	630	
Philippines	3645	378	2250		
Thailand	1606		1771		84

From Anonymous (1989) *Fruit and Tropical Product*. London: Commonwealth Secretariat Publications.

South Africa, Vietnam, Panama, and Mexico, have rapidly expanded their tropical fruit production over the last 20 years. In many of these countries, tropical fruit have become an important source of export revenue. (*See* **Bananas and Plantains**.)

Historically, fruit production was highly regionalized within the tropics. However, much of the current horticultural production is now removed from the place of origin of the particular species. For example, 91 860 t of lychee were produced by India in 1987, substantially more than the fruit's native China production of 61 820 t. Granadilla (*Passiflora quadrangularis*), native to the Americas, is now widely grown throughout the world.

History

The increasing horticultural status of tropical fruit comes from their progressive importation into the more temperate regions of the world. This 'Westernization' of tropical fruit has become prominent over the last 100 years. Its origins extend back to the early 1550s with the transportation of pineapples from South America, and the shipment of bananas across the Caribbean in the 1770s. Many of these journeys have been immortalized in the history books: the expedition sponsored by King George III and Sir Joseph Banks to transport breadfruit (*Artocarpus altilis*) to the West Indies, and the ensuing 'mutiny on the *Bounty*,' is a good example. Although this early distribution was fundamental to the subsequent development of tropical horticultural production, most of these fruit were already widely used by the local population. With transportation by air, it is now possible to move large quantities of tropical fruit anywhere in the world in a matter of days. Increased production and better cultivars will further aid the development of the fresh fruit industries in the future.

Economic Status

To assess their commercial importance, tropical fruits need to be divided into three arbitrary groups. Thus, familiar, well-established fruit, such as banana (*Musa* spp.) and pineapple, represent the majority of current tropical fruit production (Tables 1 and 2). In 1987, the European Community (EC) imported 238 936 t of pineapples, with France (66 304 t), Italy (43 644 t), and Germany (38 877 t) the main markets. Recent importation figures for the USA (1988) are around 100 000 t of pineapples, 2 873 000 t of banana, and 108 000 t of plantains.

Mango and papaya dominate the second group and are characterized by rapid industry growth. In the 5 years from 1983 to 1987, imports of papaya into the EEC rose from 1358 to 5074 t. The UK imported 1556 t of papaya in 1989 alone. These fruit are becoming more widely grown in subtropical regions, such as Jamaica, India, Mexico, Puerto Rico, Israel, the USA and Brazil. (*See* **Papayas**.)

Table 2 Tropical fruit imported by Japan in 1988

Fruit	*Quantity (t)*	*Exporter*
Banana	760 409	Philippines 79%, Taiwan 11%, Ecuador 9%, Colombia 1% (China and Malaysia <1%)
Pineapple	138 156	Philippines 94%, Taiwan 5% (Thailand, Indonesia, Hawaii, China, Malaysia and Australia <1%)
Papaya	5 240	Hawaii 100% (Fiji <1%)
Mango	5 291	Philippines 79%, Mexico 18%, Thailand 1%, Taiwan 1% (Colombia, Fiji, Venezuela and USA <1%)
Lychee	1 149	Taiwan 100%
Durian	104	Thailand 95%, Indonesia 3%, Philippines 2% (Singapore and Sri Lanka <1%)
Passion fruit	15	New Zealand 79%, USA 19%, Colombia 2% (Fiji <1%)
Pitaya (*Hylocereus undatus* B.)	22	Colombia 100% (Mexico <1%)

From Kitagawa H, Matsui T, Kawada K and Agravante JU (1990) Japan as a market of tropical fruit. *Acta Horticultural* 269: 41–45.

The final group is made up of the lesser-known exotic fruits, such as lychee, durian (*Durio zibethinus*), rambutan (*Nephelium lappaceum*), pulasan (*N. mutabile*), mangosteen (*Garcinia mangostana* L.), black sapote (*Diospyros digyna*), jackfruit (*Artocarpus heterophyllus* Lam.), carambola (*Averrhoa carambola*), and abiu (*Pouteria caimito*). These fruit are still relatively unknown in the northern hemisphere, with total imports into the EC around 9000 t in 1988.

Tropical fruit not only represent an important source of income for many countries but also have a valuable social and cultural significance. It is from their diversity of uses that the true socioeconomic value of these fruits becomes evident.

Fresh Consumption

The consumption of tropical fruit is greatest in their place of origin. The average Western European and Northern American consumes about 10 kg of citrus and banana per year. In parts of Africa, plantains may be consumed at the rate of 400 kg per person per year. Similar consumption rates exist for breadfruit in the West Indies. In their place of origin, these fruit represent a cheap and relatively plentiful food source. Historically, much of this production was easily

obtained with relatively little horticultural input. Owing to this large local consumption, it is difficult to assess accurately the true economic value for many tropical fruit, particularly the more exotic types. Production is never fully documented in official records, and in many cases, import–export quantities of tropical fruit are unknown. In Java, for example, 90% of fruit consumed is grown in the home garden. The heavy reliance on subsistence farming makes assessment even more difficult. (*See* **Citrus Fruits**: Types on the Market.)

The processing of tropical fruit has also developed into large industries (Table 3). Fruit have been canned, dried, frozen, or juiced, and marketed either as an individual product or in combination with other fruit types. Canned pineapple is probably the most obvious example. In 1990, over 920 000 t of canned pineapple were produced worldwide. Many of the South-east Asian countries, particularly Thailand, have already developed significant processing industries specializing in the exotic tropicals such as canned lychee and longan, dried mango, and frozen durian. Japan, the USA, and Europe are showing increasing interest in the processing of these more exotic fruit types, particularly mango, passion fruit, and soursop. Although quality is currently highly variable, it is through these products that many people first experience tropical fruit.

Other Uses

The commercial value of tropical fruit is generally equated to their consumption as either a fresh or cooked product. There is, however, a strong dependence on products derived from these fruits. In Africa, large quantities of a weak beer are brewed from banana, representing an important nutritional source. Cashew apple (*Anacardium occidentale*), soursop (*Annona muricata*), and tamarind (*Tamarindus indica*) are all important sources of alcoholic and nonalcoholic beverages in many countries.

Table 3 Production of processed pineapple from selected countries

	Canned ($\times 10^3$ *t*)	*Juiced* ($\times 10^3$ *t*)
Australia	40	29
Malaysia	46	1
Philippines	216	83
South Africa	63	10
Thailand	380	44
USA	133	

From Anonymous (1989) *Fruit and Tropical Product*. London: Commonwealth Secretariat Publications.

Tropical fruits have been extensively adopted in local medicine and religious practices. Bananas are widely favored as a treatment for stomach complaints, particularly ulcers. Tamarind extract is used as the basis of laxatives for bile disorders and in the treatment of fevers. For the more indulgent, carambola has been suggested in the treatment of hangovers.

Much of the medicinal use of tropical fruits is restricted to folk medicine. There are, however, reports in Western medicine of tropical fruit being successfully administered. In 1977, a London hospital reported the control of a postoperative infection in a kidney transplant patient using strips of papaya flesh, where conventional techniques were ineffective. The strong anti-Gram-negative bacterial response of papaya has now been widely documented. Much can be learnt from local folk medicine, and it is likely that with more research, other beneficial chemical compounds will be isolated.

Tropical fruit also have a strong cultural significance. In the Hindu religion, the banana plant is considered a symbol of fertility, and fruit are placed at the doorstep of the dwelling of newly married couples. In South China, bananas are also sometimes planted in the corner of rice fields as a symbol of prosperity for the coming harvest.

Toxicity

Many tropical fruit contain potentially toxic components. In most cases, these are restricted to the nonedible portions – mainly the seed, peel, and latex – and do not pose any real health risk. Many of these products have been widely utilized by the native population in local medicine. Durian seeds are known to induce breathing problems, and an extract made from langsat (*Lansium domesticum*) skin has been used as a hunting poison. The real danger occurs when these so-called nonedible components are inadvertently consumed along with the pulp. For example, jaboticaba (*Myrciaria cauliflora*) skin, which has carcinogenic properties when ingested over a long period, may be accidentally consumed in the same manner as one would eat grape skins. (*See* **Carcinogens**: Carcinogenic Substances in Food: Mechanisms.)

Where ingestion of the pulp itself causes problems, it is usually attributed to the consumption of large quantities. A mild degree of toxicity in mamey fruit (*Mammea americana*) has been known to occur. It has been reported to cause digestive problems in the Caribbean Islands, and extracts fed to dogs have proven fatal. The akee (*Blighia sapida*) and sunsapote (*Licania platypus*) are other such fruit with known toxic properties.

Limitations to Economic Development

Cultivar Selection

The commercialization of tropical fruit has been limited in the past by little plant breeding and the existence of a wide range of cultivar types. Mango, for example, is one of 14 species of the genus *Mangifera*, all of which produce edible fruit. Within the species *M. indica*, there are numerous distinct cultivars. In India alone, there are over 500 named cultivars. Each cultivar may differ in fruit quality, storage life, environmental growth requirements, and disease and insect susceptibility.

Cultivar selection and breeding are essential if the full potential of tropical fruit is to be achieved. Tropical tree fruit breeding is presently both labor-intensive and time-consuming. These limitations may be reduced with the recent advancements in plant biotechnology.

Storage

As more tropical fruit are transported around the world, there is an increasing reliance on storage. Unlike temperate fruit, most tropical fruits are difficult to store. They deteriorate rapidly after harvest and are highly susceptible to pathogen attack. If fruit are stored under ambient conditions, they become inedible within days. Even under optimal conditions, storage life is rarely longer than a few weeks. In the past, this has tended to limit the commercial success of many tropical fruit. In the production areas, rapid local consumption tends to reduce losses, but significant postharvest spoilage still occurs. The development of adequate storage and handling technology is necessary if the tropical fruit industries are to continue to expand. Although considerable postharvest research is under way worldwide, great reliance is still placed on expensive air transport to reduce transportation time. (*See* **Spoilage**: Bacterial Spoilage; **Storage Stability**: Mechanisms of Degradation.)

The accessibility of fruit is also an important criterion in economic success. Many tropical fruits fail to establish a market niche because of irregular, highly seasonal supply, compounded by a short shelf-life. These problems will become less significant as more countries produce tropical fruit and our postharvest technology develops.

Nutritional Value

The consumption of tropical fruit, like temperate fruit, acts to supplement the diet (Table 4). Although tropical fruit are generally a poor source of proteins and lipids, they provide notable amounts of amino acids (tryptophan, methionine, lysine), ascorbic acid (and other organic acids such as malic and tartaric acid), carotenes, vitamin E, carbohydrates, and fiber (hemicelluloses, cellulose and pectic substances, as well as more complex polymers such as lignin). Low concentrations (per 100 g) of various mineral elements, such as calcium (8–125 mg), iron (0.4–1.6 mg), potassium (84–170 mg), sodium (3–28 mg), and phosphorus (24–64 mg), have been reported. (Refer to individual nutrients.)

Tropical fruit with a high starch content, such as plantains and breadfruit, comprise an entirely different dietary niche. These fruit contribute a significant portion of the daily carbohydrate intake. They are widely consumed and have a similar nutrient composition to potato. Compared to the European potato, they actually have a higher carbohydrate content.

Nutritionally, tropical fruit consumption is of greatest importance to populations of the tropical regions. Not only is daily intake of these fruit and their products relatively high, but general dietary habits and total food consumption may be quite poor. This is particularly relevant in the less affluent societies. In Uganda, for example, the laborers' basic rations are based on bananas. In the subtropical and temperate regions, consumption is generally limited to particular social groups. An interesting study of local fruit consumption in Beijing identified the infant component of the population as a significant consumer group (based on daily intake).

Factors Affecting Nutritional Value

The nutritional value of tropical fruit is highly variable. During fruit ripening, storage and subsequent preparation, the nutritional characteristics of the product may be dramatically altered. Other factors, such as the cultivar type, growing region and maturity at harvest, will have a further impact.

Fruit Maturity and Ripeness

The nutritional characteristics of tropical fruit depend on their physiological maturity. During maturation and ripening, fruit undergo numerous biochemical changes, which may affect nutritional status (Table 5). Starch is hydrolyzed into sugars, organic acids are produced, and ash, minerals, and water content are all increased. Fruit maturity becomes an important consideration as many tropical fruit are traditionally consumed at an immature state. Even when fruit are consumed ripe, considerable variability may still exist. The sugar content of pineapples, for instance, increases from 4 to 15% during the last 2 weeks of the ripening process. (*See* **Ripening of Fruit**.)

Table 4 Composition of selected tropical fruit (per 100 g of edible portion)

Botanical name	*Common name*	*Food energy (kcal)*	*Moisture (%)*	*Protein (g)*	*Fat (g)*	*Total carbohydrate (g)*	*Fiber (g)*	*Ash (g)*	*Calcium (mg)*	*Phosphorus (mg)*	*Iron (mg)*	*Vitamin A (μg)*	*Thiamin (mg)*	*Riboflavin (mg)*	*Nicotinic acid (mg)*	*Ascorbic acid (mg)*
Achras zapota	Sapodilla, chico	94	75.0	0.5	1.1	23.0	1.6	0.4	24	10	1.0	10	0.01	0.01	0.2	15
Ananas comosus	Pineapple	52	85.4	0.4	0.2	13.7	0.4	0.3	18	8	0.5	15	0.08	0.04	0.2	61
Annona cherimola	Cherimoya	82	76.6	1.1	0.2	21.3	1.9	0.8	34	35	0.6	0	0.09	0.13	0.9	17
A. muricata	Soursop, guanabana	60	83.1	1.0	0.4	14.9	1.1	0.6	24	28	0.5	5	0.07	0.05	0.9	26
Artocarpus communis	Breadfruit	81	77.3	1.3	0.5	20.1	1.8	0.8	27	33	1.9	Trace	0.10	0.06	0.7	29
A. heterophyllus	Jakfruit	98	72.0	1.3	0.3	25.4	1.0	1.0	22	38			0.03	0.06	0.4	8
Averrhoa carambola	Carambola	36	90.0	0.5	0.3	8.8	0.6	0.4	5	18	0.4	90	0.04	0.02	0.3	35
Calocarpus mammosum	Mamey sapote	121	65.6	1.7	0.4	31.1	2.0	1.2	40	28	1.0	115	0.01	0.02	2.0	22
Carica papaya	Papaya (ripe)	32	90.7	0.5	0.1	8.3	0.6	0.4	20	13	0.4	110	0.03	0.04	0.3	46
Casimiroa edulis	White sapote	65	82.0	1.4	0.4	15.7	1.7	0.5	8	18	0.2	15	0.04	0.07	0.5	23
Chrysophyllum cainito	Star-apple, caimito (ripe)	68	82.8	0.8	1.6	14.5	1.0	0.3	21	17	0.8	5	0.04	0.03	1.0	11
Citrus grandis	Pummelo	34	90.3	0.6	0.2	8.5	0.8	0.4	26	26	0.5	0	0.04	0.02	0.02	35
Diospyros digyna	Black sapote	66	82.0	0.7	1.2	15.0	1.6	1.1	18	26	1.2	40	0.02	0.03	0.2	29
Durio zibethinus	Durian, civet	67	81.1	2.2	0.8	14.8	1.6	1.1	8	38	0.7	10	0.35	0.20	0.7	24
Eugenia dombeyi	Grumichama	53	85.3	0.6	0.3	13.4	0.6	0.4	40	14	0.5	20	0.04	0.03	0.3	19
Garcinia mangostana	Mangosteen	60	84.9	0.5	0.1	14.3			10	20						
Litchi chinensis	Lychee	60	83.1	0.8	1.0	15.7	0.1	0.3	5	31	1.1	0	0.02	0.03	0.9	72
Lucuma caimito	Abiu	140	60.6	1.4	0.4	36.3	0.9	0.9	22	41	1.0	130	0.02	0.02	3.4	49
Mangifera indica	Mango (ripe)	59	83.5	0.5	0.2	15.4	0.8	0.4	12	12	0.8	630	0.05	0.06	0.4	53
Musa paradisiaca	Plantain (mature)	122	65.6	1.0	0.3	32.3	0.5	0.8	8	34	0.8	175	0.06	0.04	0.06	20
M. sapientum	Common banana (mature)	110	68.8	1.2	0.2	29.0	0.4	0.8	7	28	0.5	65	0.04	0.04	0.7	15
Myrciaria cauliflora	Jaboticaba	46	87.1	0.1	0.0	12.6	0.1	0.2	6	9	0.5	0	0.02	0.02	0.2	23
Passiflora ligularis	Sweet granadilla	94	76.3	2.4	2.8	17.3	4.2	1.2	10	64	0.9	5	0.00	0.11	1.6	20
Rheedia madruno	Madrono	46	87.2	0.6	0.1	11.9	1.0	0.2	12	22	0.4	0	0.06	0.04	0.3	6
Sandoricum koetjape	Santol	46	87.0	0.8	0.1	11.8	0.1	0.3	4	17	0.4	0	0.05	0.03	0.7	86
Syzygium malaccense	Malay apple	32	90.9	0.6	0.1	8.0	0.7	0.4	6	16	0.4	Trace	0.03	0.03	0.3	13

From Wilson D and Wilson I (1988) Tropical fruit food composition table. *Rare Fruit Council of Australia Newsletter* 51: 5–9.

Table 5 Changes in the nutritional status of banana during ripening (expressed as a percentage of fresh weight)

Constituents	*Days after harvest*					
	0	*3*	*5*	*7*	*9*	*11*
Water	74.4	75.6	75.4	75.9	76.4	77.4
Protein (N × 6.25)	0.86	0.89	0.88	0.81	0.86	0.86
Crude fat	0.10	0.23	0.31	0.47	0.35	0.17
Reducing sugars (as invert)	0.24	2.81	7.24	10.73	12.98	15.31
Nonreducing sugars (as sucrose)	0.62	4.85	6.52	6.12	3.89	2.60
Total sugars	0.86	7.66	13.76	16.85	16.87	17.91
Starch	20.65	12.85	6.00	2.93	1.73	1.21
Protopectin (as calcium pectate)	0.53	0.56	0.31	0.34	0.21	0.22
Pectin (as calcium pectate)		0.27	0.36	0.34	0.37	0.40
Ash	0.74	0.79	0.79	0.76	0.72	0.74
Total acid (as citric)	0.18	0.25	0.32	0.29	0.26	0.23

From Loesecke HW (1960) Effect of harvesting and handling practices on composition of unprocessed food. In: Harris RS and Loesecke HW (eds) *Nutritional Evaluation of Food Processing*, pp. 58–91. New York: John Wiley.

Preparation

Although the majority of tropical fruit are consumed fresh, there are many cases where consumption is only after some form of preparation. Cooking of the whole fruit, or a paste, is very common for many tropical fruit. This process may be necessary, as some fruit are purgative if eaten raw. The nutritional status of the fruit is dramatically affected as a result. Fresh green plantains, for example, have an energy count of 378–613 J (90–146 cal) per 100 g of edible portion, compared to 1508 J (359 cal) as a dried product and 323 J (77 cal) after cooking.

Cultivar Type and Growing Region

Many cultivar types have distinct nutritional characteristics. Although commonly equated to flavor, color, and taste, differences in ascorbic acid, protein, amino acid, and carotene levels are regularly reported. The growing region is also known to have an effect. Cherimoyas produced in Ecuador have 17 mg of ascorbic acid (per 100 g), whereas a similar cultivar grown in nearby Colombia has just 5 mg of ascorbic acid.

See also: **Bananas and Plantains**; **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; **Citrus Fruits**: Types on the Market; **Mangoes**; **Papayas**; **Pineapples**; **Ripening of Fruit**; **Spoilage**: Bacterial Spoilage; **Storage Stability**: Mechanisms of Degradation

Further Reading

Almeyda N and Martin FW (1977) *Cultivation of Neglected Tropical Fruits with Promise*. Washington, DC: United States Department of Agriculture, Agricultural Research Service.

Anonymous (1989) *Fruit and Tropical Product*. London: Commonwealth Secretariat Publications.

Collins JL (1960) *The Pineapple*. New York: Interscience.

Hulme AC (1971) *The Biochemistry of Fruits and their Products*. London: Academic Press.

Kitagawa H, Matsui T, Kawada K and Agravante JU (1990) Japan as a market of tropical fruit. *Acta Horticulturae* 269: 41–45.

Lindsay P and Cull B (1982) *Fruit Growing in Warm Climates*. New South Wales: Reed Books.

Loesecke HW (1960) Effect of harvesting and handling practices on composition of unprocessed food. In: Harris RS and Loesecke HW (eds) *Nutritional Evaluation of Food Processing*, pp. 58–91. New York: John Wiley.

Menzel CM, Watson BJ and Simpson DR (1988) The lychee in Australia. *Queensland Agricultural Journal* 114: 19–27.

Miller CD, Bazore K and Bartow M (1957) *Fruits of Hawaii*. Honolulu: University of Hawaii Press.

Mukherjee SK (1971) Origin of mango (*Mangifera indica*). *Economic Botany* 26: 260.

Oliviera JFS (1974) The nutritional value of some foods consumed on San Tome Island. *Ecology of Food and Nutrition* 3: 237.

Page P (1984) *Tropical Tree Fruits for Australia, Information Series Q 183018*. Brisbane: Queensland Department of Primary Industries.

Purseglove JW (1974) *Tropical Crops: Dicotyledons*. London: Longman.

Samson JA (1980) *Tropical Fruits*. New York: Longman.

Wills RBH, Lee TH, Graham D, McGlasson WB and Hall EG (1982) *Postharvest: An Introduction to the Physiology and Handling of Fruits and Vegetables*. Sydney: New South Wales University Press.

Wilson D and Wilson I (1988) Tropical fruit food composition table. *Rare Fruit Council of Australia Newsletter* 51: 5–9.

Fruits of the Sapindaceae

C M Menzel, Maroochy Research Station, Nambour, Queensland, Australia

Introduction

The Sapindaceae (or soapberry) family contains more than 1000 species from 125 genera, mostly trees and shrubs, but rarely herbs, and with a wide distribution in the tropics and warm subtropics. The majority of species are native to Asia, although there are a few in South America, Africa and Australia.

Many species are important timber trees or are used as soap substitutes, sources of drugs, beverages, or nuts. However, it is the attractive, eye-catching fruit of the subfamily Nepheleae that are of greater economic interest, especially those that are popular throughout the Orient: lychee or litchi (*Litchi chinensis*), longan (*Dinnocarpus longan*), rambutan (*Nephelium lappaceum*) and pulasan (*Nephelium mutabile*). Lychee is more widely grown than the other species, and is the most important member of the Nepheleae. The four species are similar in growth and fruiting habit, but differ in fruit morphology and climatic adaptation. Lychee is regarded as one of the kings of fruits and has a very long history in China. The fruit are very attractive with a bright red skin covered by angular or conical protuberances. Longan fruit resemble lychee fruit but are smaller, smoother and yellow–tan to brown in color. The fruit are also milder in flavor and less acidic. About a third of people in China and Thailand rate the taste of longan equal to that of lychee or better. Canned longans are also more acceptable than canned lychee. Rambutan and pulasan fruit are similar to lychee, usually with red skin. However, long hairs or spinterns replace the protuberances. Rambutan and pulasan are strictly tropical, cropping only in warm, wet lowland tropical areas. In contrast, lychee and longan bear well in the warm subtropics or at elevation in the tropics.

Origin, Distribution and Commercial Importance

The lychee originated in southern China and possibly in northern Vietnam and the Malay Peninsula, and wild lychee trees can still be found growing in several rainforest areas in these locations. The wild lychee is similar morphologically to cultivated lychee in China, in particular the Jin Feng types. The fruit are edible, but the aril or flesh is relatively thin and sour.

The *Litchi* genus contains two other subspecies that have not been commercialized, *L. chinensis* ssp. *philippinensis* (kumingi), which developed in the Philippines and Papua New Guinea at high elevation (Luzon, Sibuyan, Samar, and Mindanao), and *L. chinensis* ssp. *javensis*, which has been recorded in the Malay Peninsula and Indonesia. Philippines lychee has long oval-shaped fruit with long thorn-like protuberances. Fruit split in the middle when ripe, and the aril only partly covers the seed and is not edible. *L. chinensis* ssp. *javensis* is a rare specimen in Chinese gardens in West Java and Indo-China, and has fruit similar to the cultivated lychee, but the aril is thinner. It is reported to flower and fruit well under wet tropical conditions. Many of the Malayan specimens of *Litchi* belong to ssp. *chinensis* not ssp. *javensis*.

The longan (*D. longana* ssp. *longan* var. *langon*) originated either in subtropical China and Vietnam or in the area between Myanmar and India, and wild trees have been found growing amongst the rainforests on Hainan Island, near northern Vietnam. There are six species in the genus *Dianocarpus*, all from tropical and subtropical Asia, and Australia but longan is the only one grown for its edible fruit. *D. longan* spp. *malesianus* (mata kuching) produces fruit of a similar size to longan. Fruit have a tough skin that is pale dull yellow with dark raised specks. The aril, which envelops a big seed, is whitish, translucent and sweet and in good forms nearly 0.5 cm thick, although it is usually much thinner. Trees grow wild in Malaysia, Borneo, Sumatra, and the Celebes.

Rambutan and pulasan are thought to be native to west Malaysia and Sumatra, although trees that have escaped from cultivation blur the original distribution. They are close relatives of the lychee, and are equally desirable fruit, but are not as well known. The fruit are very similar to each other and may sometimes be confused. Trees are smaller in pulasan, leaflets narrower, and branches more open, and there are fewer fruit in a cluster. The fruit skin is thicker, and the spinterns or tubercles very much shorter. Trees are also reported to be less productive. Other species of *Nephelium* with edible arils that are occasionally consumed include *N. eriopetalum* (lotong), *N. glabrum* (redan), *N. philippense* (bulala), and *N. excrospermoides* (aluao).

The lychee was introduced in the tropical and subtropical world from China from the end of the seventeenth century and is now grown within 15–35° latitude in most countries. The distribution of lychee to the rest of the world has been slow, possibly because of the need for propagation by air-layering and the narrow ecological range of most cultivars. The most important production areas today are China

(1 300 000 t), Vietnam (50 000 t), India (430 000 t), Pakistan, Thailand (80 000 t), Indonesia, Madagascar (35 000 t), South Africa (9500 t), Mexico, and Australia (5000 t). In China, lychee is mainly grown in three southern provinces: Guangdong, Guangxi, and Fujian.

The longan reached Thailand in the late nineteenth century and Hong Kong, Hawaii, Florida, and Australia by the early to mid-twentieth century. However, despite its wide environmental adaptation, longan is commercialized only in China (500 000), Thailand (240 000 t), and Vietnam (365 000 t). Longans are rated 12th in order of total production for tropical fruits in Thailand, with an average of 4–6 times more longans produced than lychees.

Rambutan and pulasan are widely distributed in the humid tropics of South-east Asia. The largest commercial plantings of rambutan occur in Malaysia (80 000 t), Indonesia (148 000 t), Thailand (430 000 t), and the Philippines (20 000 t). In Malaysia, rambutan is the most important fruit crop, whereas in Thailand and Java, rambutan ranks third after mango and tangerine. In the Philippines, rambutan was introduced from Indonesia only in this century and is not listed in the 20 main tree fruits. The pulasan is widely grown in the western part of Java, with smaller plantings in other parts of Indonesia, Malaysia, and Thailand.

Lychees and longans are well known and popular fruit in most Asian countries, and significant quantities are traded each season. The Hong Kong and Singapore markets are well serviced and highly competitive, and have very high standards for imported produce. Consumers know and prefer large highly colored sweet fruit with small seeds and firm flesh.

Hong Kong and Singapore are the main markets for lychees in South-east Asia, with the fruit coming from China, Taiwan, Vietnam, and Thailand. Most of the fruit enter between June and July. The total market is about 20 000 t. Prices range from A$1 to A$11 per kilogram, depending on the fruit quality and supply. In 1994, Thailand exported 6828 t, but mostly as canned fruit (5834 t).

Thailand exports up to 50% of its longan crop in some seasons, and longans are the most important export fruit crop. In 1994, for instance, it exported 46 123 t to Hong Kong, China, Singapore, Malaysia, Indonesia, and the USA, mostly as fresh fruit (32 628 t), but also dried (3335 t), frozen (140 t), and canned (10 020 t). Prices for fresh fruit in Asian markets fluctuate considerably during the season. Early fruit may fetch A$3–5 per kilogram, but as supplies increase, the price drops to A$0.40–1.00 per kilogram. Average prices during the season are about A$1.50 per kilogram. Singapore consumers pay more for Thai fruit because they are normally larger and sweeter than those from China and Taiwan. Retail prices for longan are about half those for lychees.

Lychees share in the growth of exotics in western Europe. Exotics are seen as a bright spot in fruit wholesaling, with increased turnover and higher prices compared with temperate fruits. The UK, France, and Germany import about 15–20 000 t of lychee, mainly from Madagascar and South Africa, with smaller quantities from Reunion, Mauritius, and Israel. Fruit are normally available from November to early March, with peak supplies during mid-December to late January. Prices range from A$4 to A$10 per kilogram.

Exports of rambutan are considerably reduced compared with lychee and longan. Most of the production is sold in the domestic markets. Thailand and Malaysia export about 5000 t, mainly to Singapore and Hong Kong. The export of rambutans does not play a major role in terms of foreign exchange earnings.

Description of Fruit

Lychee fruit are small (3.0–5.0 cm diameter) and variable in shape from round to egg- or heart-shaped. The skin color is bright red, dull red, purple–red or pinkish red, depending on cultivar. Some cultivars also have distinctive yellow blotches on the skin. The skin has protuberances on each segment that can be smooth, sharp-pointed, or hair-like and sharp.

The fleshy edible portion of the fruit is called an aril and is an outgrowth of the seed stalk. It grows as the fruit develops until it completely covers the seed, although in a few cultivars, the aril may not completely envelope the seed, especially during periods of drought. Such cultivars have a low flesh recovery and are poor marketing types, especially in Asian markets. The aril is white and translucent and has the consistency and flavor of a prime muscat grape. Cultivars vary greatly with respect to texture, aroma, and flavor of the aril.

Each fruit normally contains one chestnut brown to dark brown, ovoid to oblong seed about 1.0–3.3 cm long and 0.6–1.2 cm wide. In some cultivars, a high proportion of seeds may be abortive. The abortive seeds are small and shrivelled and are known as ‘chicken tongue’ seeds. Fruit with abortive seeds are preferred, and often attract a higher price, for although they are somewhat smaller than fruit with normal seeds, they usually contain a higher proportion of flesh. The proportion of small or shrivelled seeds is an important characteristic of a cultivar, and varies from season to season and orchard to orchard. Fruit with chicken tongue seed are more susceptible

to environmental stresses and are usually shed before fruit containing normal seeds. The average weight is 16–35 g and flesh recovery 50–70%.

Longan fruit are small (about 1.5–3.0 cm in diameter), globose to round, sometimes with distinctive shoulders. The fruit skin is thin and leathery, and changes from green–yellow to yellow–brown with advancing maturity. The tubercles are typically flattened or indistinct. The aril (flesh) is translucent white to off-white in color, sometimes with a pinkish tinge (after processing), and ranges in texture from juicy to very crisp and in flavor from bland to sweet and aromatic, but seldom acidic.

Seeds are glossy red brown, dark brown to black, small, round to ovoid, and easily separated from the flesh. Most fruit are full-seeded, although some cultivars have up to 20–30% small seeded fruit. Average weight is 12 to 22 g and flesh recovery 60–75%. Fruit size and flesh recovery are related, small fruit normally having a lower flesh recovery.

Rambutan fruit are ovoid or ellipsoid, 3.0–8.0 cm long and 2.5–5.0 cm wide. The fruit has a very attractive appearance with the skin changing from green to various shades of pink, red or yellow during ripening. The skin has hair-like growths or protuberances (spinterns) up to 2 cm in length, densely clustered over the surface. The spinterns vary in color, but are usually red with green tips and are responsible for the common name of the fruit, which is based on the Malay word 'rambut,' meaning hair. The flesh or aril is white or rose-tinted, translucent and juicy, and acid, subacid, or sweet, depending on the cultivar.

The seeds are 2.5–3.5 cm long and 1.0–1.5 cm wide and somewhat flattened. The aril may or may not peel cleanly from the seed. In 'clingstone' cultivars, the aril usually detaches together with somewhat objectionable seed testa. The fruit weight is 20–60 g, and flesh recovery is 30–68%.

Fruit of pulasan are ovoid in shape and 5.0–7.5 cm long and resemble those of rambutan in general appearance. However, the pericarp is thicker, usually dull red, with shorter spines. The aril normally clings to the testa. The flavor and quality of the aril are invariably good, and often sweeter and preferred to that of rambutan.

Chemical Composition and Uses of Fruit

The composition of lychee, longan, and rambutan is shown in Table 1.

Lychees can be eaten directly off the tree or cool-stored, dried, or quick-frozen. Thawed fruit can be used in the same way as freshly picked fruit with only a slight loss of color and flavor. The fruit can be peeled, pitted, and canned in sugar syrup. Fresh or processed fruit can be used alone or with other fruits in tropical fruit salads.

There are substantial canning industries in China, Hong Kong, Taiwan, Vietnam, and Thailand. Canned lychees are not as acceptable as canned longans. This is possibly because the major cultivar that is canned in South-east Asia, Haak Yip, develops an off-flavor with heat sterilization. Fruit of other cultivars are usually acceptable, although the full flavor is lost.

A large proportion of the crop was traditionally dried in China and Vietnam, and often given as presents. Dried lychees are known as 'lychee nuts,' and during drying, the pericarp or outer skin becomes brown and brittle, but retains its shape. The aril or flesh that shrivels around the seed is very pleasant, similar in flavor and texture to a dried date.

Lychees are sometimes fermented for Chinese medicine or used to make wine, jelly, sherbet, yogurt, icecream, and sweets.

Longans can be eaten fresh, dried, or quick frozen. Thawed fruit are quite satisfactory. The fruit can be peeled, pitted, and canned. The juice of most cultivars is sufficiently sweet for processing without adding sugar. Canned fruit are very acceptable and taste much better than canned lychees. Taiwan and Thailand have substantial canning industries. Different cultivars are preferred for canning or drying; others are eaten fresh. Sweet fruit are best used for drying, but those lower in sugar are preferred for canning. Flavor and sweetness are normally related.

Rambutans and pulasans are normally grown for fresh fruit, although a canning industry for rambutans exists in Thailand and Malaysia. Asians generally prefer rambutan cultivars with crisp, nonjuicy fruit, and the testa clinging to the aril is not generally

Table 1 Composition of lychee, longan, and rambutan fruit (data expressed on a fresh-weight basis[a])

Species	*Protein*	*Fat*	*Carbohydrate*	*Fiber*	*Ca*	*P*	*Fe*	*Vitamins (mg kg^{-1})*			
	(g kg^{-1})				*(mg kg^{-1})*			*B_1*	*B_2*	*Niacin*	*C*
Lychee	9	1	131	1	70	410	13	1.1	0.4	3	1670
Longan	10	5	252	4	20	60	3	0.4	0.7	6	80
Rambutan	9	1	145	11	30	60	18	0.4	0.5	6	310

[a]Source: Anonymous (1987) *Fruits in Thailand*. Bangkok: Department of Agricultural Extension, Ministry of Agriculture and Co-operatives, with permission.

regarded as objectionable. Europeans prefer sweet stringless fruit without objectionable testa.

Harvesting and Storage

Immature lychees may sweeten a little once harvested, but do not develop a full flavor. Maturity is judged by a particular shape, skin color, skin texture, and flavor of each cultivar. A maturity index based on a brix/acid ratio of 35 has been developed in Australia. Most fruit can be picked from a tree within 1 week and from a single cultivar in an orchard within 3 weeks. Fruit in local Asian markets are normally sold in clusters within 1–3 days after harvest in baskets weighing about 20–25 kg without any postharvest treatment.

In Australia and South Africa, lychees are detached prior to sale, and a machine for destalking lychees has been developed in Australia. Fruit lose their bright red skin color and turn brown unless moisture loss is controlled within a few days after harvest. Browning is faster under low-humidity conditions. Efforts to maintain humidity tend to cause postharvest rots to develop. Fruit in Australia are generally packed in 2.5-kg plastic bags, with small holes to reduce condensation. Fruit may be air-cooled or hydrocooled to 5–10 °C. Fruit dry out once the bags are opened in the retail outlets. Punnet packs with plastic PVC overwraps allow lychees to be displayed without excess browning and are preferred by many retailers. The film retains sufficient humidity to inhibit browning without condensation clouding the pack. However, fruit are susceptible to rots, especially at higher temperatures. Research is currently investigating the application of pre- and postharvest chemicals, along with packaging and cooling to reduce the rots.

Lychees shipped to Europe from South Africa and Madagascar and, more recently, Israel are normally treated with sulfur to prevent browning. The fruit are placed in a closed container, and 50–150 g of sulfur per cubic meter of air are burnt in the enclosed space for 20–30 min. Alternatively, SO_2 gas can be pumped into the container. The sulfur dioxide bleaches the fruit to a pinkish cream color. Immediately after SO_2, fruit are bright yellow and progressively recover color. Fruit are then often dipped in acid. These treatments have been shown to extend storage life dramatically, inhibit browning and limit the need for specialized packaging and handling. However, there is evidence of sulfur residues and pulp tainting. Treated fruit also never match the bright red appearance at harvest. Recent research in Israel has looked at replacing the sulfur treatment with a cooling/heating cycle, followed by an acid dip. This treatment retains the red skin color. However, eating quality is related to water loss and disease development. Waxes have also been developed in the USA to reduce water loss and fix the red skin pigments. Further research is required in this area.

Longans do not ripen off the tree. They may sweeten a little, but do not develop a full flavor. Maturity is judged by the particular shape, skin color, and flavor of each cultivar. Sending immature fruit to market to gain high early prices damages the reputation of the grower and the longan in the market place.

Most fruit can be picked from a tree in one harvest and from a single cultivar in an orchard within 2 weeks. Consequently, to spread the picking workload, it is essential to plant a range of cultivars in any one orchard. Fruit are harvested by removing the whole cluster plus one or two leaves. Fruit are clipped from the panicles, sorted for size, insect damage, and skin blemishes, and placed in bulk trays.

Disease appears to be the main factor limiting longan storage. With adequate disease control, fruit maintain an acceptable eating quality for 5 weeks at 10 °C and 90% relative humidity, although significant browning of the skin occurs. In the absence of good disease control, storage at 7.5 °C is recommended, but for periods not exceeding about 3 weeks. At lower temperatures, there is a rapid loss of eating quality, principally associated with the presence of off-flavors, whereas above 10 °C, fruit succumb to postharvest diseases.

Unlike lychees, longan fruit do not separate from the fruit stalk easily without the loss of some skin. This could hasten breakdown of the fruit after harvest. Longan fruit, consequently, do not look as attractive as lychees in punnets. Longans marketed attached to the fruit stalk in bunches of about 15–20 fruit have attracted superior prices compared with loose fruit in Australia. A machine for destalking fruit is available in Australia and has a possible use for longan.

In China and Thailand, most fruit are marketed in branches on the fruit stalks in 22–25 kg bamboo baskets and consumed within 3 days without any postharvest treatment. A proportion of the crop is sometimes stored on ice to give a few extra days of shelf-life, but there is generally no precooling or refrigeration. Fruit for sale in Malaysia and Singapore from Thailand are precooled with cold water (hydrocooling) to remove field heat. The fruit usually requires additional top icing to provide further cooling during transportation. This precooling substantially reduces losses due to rots. Fruit exported to Hong Kong are packed with the fruit stalk intact. However, the larger growers consider that average quality is unimpaired when all the raceme stalks are removed.

The Thai industry fumigates longans going to Singapore in SO_2 to control browning and disease.

Sulfur removes blemishes and turns the fruit bright yellow. There is some concern by consumers in Singapore and Hong Kong on SO_2 residues. The optimum SO_2 concentration is a compromise between the rate which is effective in maintaining fruit appearance and the rate which causes tainting of the pulp. The long-term use of sulfur in these crops is under a cloud. Research is continuing in Thailand.

The price of fruit in Thailand depends on the season, cultivar, and fruit quality. Average prices are about A$1.00–1.50 per kilogram, with a range of A$0.50–3.00 per kilogram. The greatest demand is for fruit with a weight of more than 18 g.

As for lychee and longan, fruit maturity in rambutan and pulasan is determined by the color of the skin and aril taste, which vary with cultivar. The whole panicle is normally harvested. In order to harvest all the fruit at correct maturity, it may be necessary to pick trees at 3- to 4-day intervals over a period of 2–6 weeks. The fruit are reasonably hardy, but should be handled carefully to avoid bruising and crushing. In Malaysia and Indonesia, fruit are presented for sale in bunches on panicles, whereas in the Philippines, Indonesia, and Australia, fruit are detached prior to sale. Care should be taken not to rupture the skin when destalking fruit.

Fresh fruit are very attractive and eye-catching. However, ordinarily, the fruit must be consigned to local markets within 3–4 days before browning and decay set in. Removal of the panicle stems hastens deterioration. Moisture loss and browning of the spinterns can be reduced by storage at low temperatures and high relative humidities (95%). Red cultivars can be successfully stored for 14 days at 7.5 °C, and yellow cultivars for 11 days at 12.5 °C. Damage in yellow fruit is much more obvious than in red fruit. Sealed bags are more effective than perforated bags in reducing water loss, but rots and off-flavors may develop. In Australia, fruit are packed in 2.5-kg boxes lined with polyethylene cling wrap or perforated plastic film. Combined with cool storage, fruit may store for a week or more. As with the other members of the Sapindaceae, postharvest handling and storage currently limit domestic and export marketing.

See also: **Canning**: Quality Changes During Canning; **Drying**: Drying Using Natural Radiation; **Fumigants**; **Storage Stability**: Parameters Affecting Storage Stability

Further Reading

Bose TK and Mitra SK (eds) (1989) *Fruits: Tropical and Subtropical*. Calcutta: Naya Prokash.

Groff GW (1921) *The Lychee and Lungan*. New York: Orange Judd Company.

Menzel CM (2002) The lychee crop in Asia and the Pacific. Food and Agricultural Organization of the United Nations, Bangkok, Thailand.

Menzel CM and McConchie CA (1998) Lychee and longan. In: Hyde KW (ed.) *The New Rural Industries*, pp. 288–295. Canberra: Rural Industries Research and Development Corporation.

Menzel CM and Simpson DR (1990) Performance and improvement of lychee cultivars: a review. *Plant Varieties Journal* 44: 107–215.

Menzel CM, Watson BJ and Simpson DR (1989) Longans – a place in Queensland's horticulture. *Queensland Agricultural Journal* 115: 251–265.

Morton JF (1987) *Fruits of Warm Climates*. Winterville, NC: Creative Resources Systems Incorporated.

Page PE (ed.) (1984) *Tropical Tree Fruits For Australia*. Brisbane: Queensland Department of Primary Industries.

Tindall HD, Menini UG and Hodder AJ (1994) *Rambutan Cultivation*. Food and Agricultural Organization of the United Nations Plant Production and Protection Paper. No. 121.

Verheij EWM and Coronel RE (1991) *Plant Resources of South-east Asia No. 2 Edible Fruits and Nuts*. Wageningen, The Netherlands: Pudoc.

Yeap CK (1987) The Sapindaceous fruits and nuts. *Yearbook of the Western Australia Nut and Tree Crops Association* 12: 16–33.

Zhang Z, Yuan P, Wang B, Qui Y and Li J (1997) *Litchi Pictorial Narration of Cultivation*. Guangdong, People's Republic of China: Pomology Research Institute, Guangdong Academy of Agricultural Sciences.

Fruits of the Sapotaceae

S N Vaghani, Gujarat Agricultural University, Navsari, India

Background

Sapotaceae is one of the latex-yielding families, known for the production of Gutta percha, chicle (chewing) gum, timber, edible flowers, fruits, and oil seeds. Some species of *Achras*, *Chrysophyllum*, and *Mimusops* are also cultivated for their edible fruits. The family is included in the order Ebenales, series Heteromerae, subclass Gamopetalae, class Dicotyledon, of division Angiospermae. It is composed of about 50 genera and 1100 species, distributed in the tropical world by habitat. The plants are trees or sometimes shrubs, and the fruit is an indehiscent, one- to eight-seeded berry (sometimes a drupe). The seeds possess crustaceous, shining testas, and are oily and exalbuminous with large, fleshy cotyledons, or

albuminous with flat cotyledons. The family contains very few species that are grown for their edible fruits. However, among these species sapota (*Manilkara zapota* L. van Royen) is the most outstanding edible fruit, and is mainly used for table purposes. The physiology of this fruit is complicated, and very little research work has been performed on its handling, storage, nutrition, and processing.

Manilkara zapota L. van Royen (*Achras zapota* L.)

Vernacular names are sapota, sapota plum, sapodilla, zapotle, chico zapote, chiku, dilly, naseberry, nispero, and chicle tree. Sapota is an evergreen tree, native to tropical America, especially southern Mexico and Central America. In the moist tropics of the world, the trees are often cultivated for their edible fruits. Sapodillas are a favorite desert fruit of tropical America and other warm areas. They are full of translucent yellowish pulp and are delicious. The latex is also tapped from the bark for chicle gum, the base for chewing gum.

Type and Commercial Importance

The fruit of sapota is a fleshy berry. The fully ripe fruit is delicious and sweet, and is chiefly used fresh. In ripe fruit, the skin can also be eaten with the flesh. A small proportion of produce is utilized in various preparations. The fruit imparts a characteristic, pleasant flavor to milk, and is therefore freely utilized in many milk-based products, such as icecream or milkshakes. Chiku halwa, a famous milk-based Indian sweet, is prepared by concentrating sweetened milk to a paste-like consistency, and then adding shreds of sapota fruit previously simmered in ghee. After thorough mixing, the mass is cooked to a thick consistency so as to retain its form. Shrikhand, another milk-based Indian sweet made from curd after draining, is also dressed with cubes of ripe sapota fruit. In coastal areas of India, the fruit slices are soaked in butter for a night and eaten in the morning. It is said to be an excellent prophylactic against biliousness and febrile attacks. In the West Indies, the seeds are known to be aperient and diuretic. The coagulated resinous latex derived from the bark known as chicle is used to make chewing gum. Great forests are cultivated in South America, and tapping the sap of wild trees is quite a cottage industry.

Morphology and Anatomy of the Fruits

The fruit is a fleshy berry, that looks like and is the size of a brown tennis ball, varying in shape from globose to conical to oval, and in size from 5 to 8 cm in length, 3.5 to 7.0 cm in diameter and 75 to 150 g in weight. The skin (rind) is thin, rusty brown, and somewhat 'scurfy,' giving the fruit a striking resemblance to an Irish potato. The flesh is yellowish-brown or red, translucent and soft; it crumbles with a sandy or granular texture. The seeds are endospermic, and vary in number from zero to 12 (or many) depending on the season. Each seed weighs 0.6–1.0 g, and is hard, black and shining, obovate or flattened, about 1.5–2.0 cm long and embedded in pulp around the central axis; the seeds separate easily from the pulp.

Chemical and Nutritional Composition

The principal chemical constituents of the sapota fruit are carbohydrates and tannins. The proximate composition of nutrients in a ripe fruit is presented in Table 1. The major constituent of the fruit is carbohydrate (21%), of which free sugars take a major share: glucose (41%), fructose (38%), galactose (5.4%), maltose (0.5%), lactose (0.1%), and oligosaccharides – as glucose – (6.4%). Galactose, maltose, and lactose are present only in the ripe fruit, but glucose and fructose are present at all stages of fruit development. At the overripe stage, the sucrose content decreases owing to inversion to glucose and fructose. In general, lactose does not occur in fruits; it is mainly of animal origin. However, it occurs in ripe sapota in very small quantities, and differs from that in the mammary gland of animals. The total sugars not only impart sweetness to the fruit, but also mask the astringency caused by the tannins. (Refer to individual nutrients.)

The mature sapota contains polyphenol-like tannins (1.01 g), phenols (4.82 g), flavans (3.11 g), and leucoanthocyanidins (4.81 g per fruit). The fruit is highly astringent when unripe, and this astringency is predominantly caused by the flavans. The polyphenol-like leucodelphinidin and leucocyanidin are the major components, and leucopelargonidin is a minor component. The great resistance of sapota to fungal diseases may be attributed to the presence of large quantities of polyphenols in the ripe fruits. (*See* **Phenolic Compounds**; **Tannins and Polyphenols**.)

Sapota fruits are very low in vitamins (0.1–12 mg%) – as compared to guava (300 mg%) and aonla (700 mg%) – and in minerals. The fruits also contain negligible amounts of fat and protein. Starch is either absent or found in small quantities in the ripe fruit. The seed kernels (50% of the total seed weight) contain liquid fat (20%), saponin (1%), and sapotin (bitter principle; 0.08%). Fruit quality is adversely affected by latex, tannin, aldehyde, and sapotin content, which are present at higher levels in immature and unripe fruits.

Table 1 Nutrient composition of ripe sapota fruits[a] from India and Mexico

Edible portion (%)	*Moisture (g%)*	*Fiber (g%)*	*Carbohydrates (g%)*	*Fat (g%)*	*Proteins (g%)*	*Energy (kJ per 100 g)*	*Minerals (g%)*	*Carotene (μg)*	*Vitamin C*	*Thiamin*	*Riboflavin*	*Nicotlnic acid*	*Calcium*	*Phosphorus*	*Iron*
India 83	73.7	2.6	24.4	1.1	0.7	412	0.5	97	6	0.02	0.03	0.2	28	27	2.0
Mexico 86			18.0	1.1	0.7	160			12	0.02	0.00	0.3	31	9	1.5

[a]Units are mg% unless otherwise stated.

Data from Gopalan C, Sastri BVR and Balasubramanian SC (1977) *Nutritive Value of Foods*. Hydrabad: Nutro.: Indian Council of Medical Research; and Hernandez M, Chavez A, Bourges H and Mendoza E (1974) *Nutritive Value of Foods* (in Spanish). Mexico City: National Institute of Nutrition.

Handling and Storage

Sapota fruits are climacteric in nature, and can be ripened artificially after harvest. The mature fruit, after proper ripening, acquires eating and processing qualities, whereas immature fruits develop poorly and are astringent in taste and very sticky in texture. The fruit should therefore be harvested at full maturity for proper ripening. (*See* **Ripening of Fruit**.)

In general, fruits take 8 months to reach harvest maturity from fruit set, and there are two principal seasons of production: February to April and October to December. However, because of the continuous flowering habit, the fruits do not come to maturity at the same time in a season. Fruits at different stages of development are therefore present on the tree all year round. Maturity can be judged by the appearance of a yellowish streak when the fruit is scratched with the fingernails, or by the oozing of watery latex on slight injury. However, indices of this type are destructive and can only be used for crops for which the total harvest time is of short duration. In the sapota crop, the fruiting is continuous and harvesting is carried out almost throughout the year or for a long duration, and hence such destructive indices are not useful. However, they can help the farmer to acquire expertise in making judgments about physical parameters. Shininess of the fruit surface, development of the fruit, and the extent of 'scruffiness' are good indicators of maturity. A fruit with a smooth surface, shining potato color, and rounded stylar end can be considered mature. The fruit contains latex, which decreases towards maturity (1.80–1.05%), and remains constant during the final stages of fruit growth. Latex almost stops oozing out when fruits are mature, and this change is also used as an indicator of maturity. On maturation, the mellowing in astringency is caused by polymeric changes in the polyphenols, and masking by the sugar. The fruits are harvested with their stalks intact; they are collected in gunny bags to avoid bruising and lowered down to the ground carefully. The fruits are cleaned by rubbing with the hand, vigorous rolling in gunny bags made into a 'cradle,' or washing in a water tub; they are shade-dried. The fruits are then graded on the basis of size (big, medium, and small) manually or with a sapota grader. They are packed in corrugated cardboard, wooden or plastic boxes, or in bamboo baskets, for marketing. Fruits are packed by number, e.g., dozens, or weight, having a final weight of nearly 16 kg per container, and priced accordingly.

Sapota fruits attain eating quality only after proper ripening. The ripe fruit has a pleasant, mellow aroma, and is sweet without any astringency and stickiness. During ripening, starch is hydrolyzed with a concurrent increase in sugar, and there are reductions in levels of starch, tannin, latex, sapotin, aldehydes, and acidity during ripening. The major changes in chemical composition of sapota fruit are presented in Table 2. Carbohydrate metabolism during the various stages of ripening produces oligosaccharides (three types), which finally yield glucose, fructose,

Table 2 Chemical composition of mature and ripe fruits (and latex) of two sapota cultivars: Cricket Ball (A) and Oblong (B)

Composition	*Mature (at harvest)*		*Ripe (eating ripe)*	
	A	*B*	*A*	*B*
Average fruit weight (g)	107.16	20.04	107.07	20.54
Starch (%)	0.90	0.37	0.71	0.35
Sucrose (%)	4.28	5.13	2.12	3.10
Glucose (%)	2.81	4.53	5.65	7.71
Fructose (%)	5.53	3.92	6.30	4.25
Crude protein (%)	0.68	1.02	0.52	0.76
True protein (%)	0.51	0.81	0.50	0.63
Soluble amino acid (mg%)	25	114	43	80
Total acidity (%)	0.15	0.61	0.11	0.41
Nonvolatile acidity (%)	0.07	0.46	0.03	0.31
Malic acid (mg%)	18.95	128.50	17.75	123.91
Vitamin A (IU %)	16 013	1219	410	680
Vitamin C (mg%)	10	5	8	4
Dry matter (%)	21.9	26.5	23.4	25.6
Potassium (%)	0.21	0.17	0.25	0.14
Phosphorus (mg%)	8.0	10.9	11.6	6.8
Calcium (mg%)	10.8	20.9	8.2	16.4
Iron (mg%)	0.42	0.61	0.40	0.52
Tannins (%)	0.27	0.81	0.10	0.20

Reproduced from Selvaraj Y and Pal DK (1984) Changes in the chemical composition and enzyme activity of two sapadilla (*Malinkara zapota*) cultivars during development and ripening. *Journal of Horticultural Science* 59(2): 275–281, with permission.

lactose, and galactose on hydrolysis. The hemicellulose content decreases as the fruit matures, and is further depleted during chemical ripening. Part of the hemicellulose is converted to oligosaccharides during ripening, with a simultaneous increase in sucrose content, followed by an increase in glucose and fructose. However, at the overripe stage, the sucrose content decreases by inversion to glucose and fructose.

In general, the fruits are spread on mats or on the floor in a single layer, or sandwiched in paddy straw, or packed in baskets or plastic crates, or kept in gunny bags for ripening. It takes 5–7 days for ripening, but some varieties, such as the Calcutta Round cultivar, require 9–13 days at room temperature. Fruits can be ripened more uniformly and faster with various ripening media, chemicals, and growth hormones. Amongst the various simple ripening media are sawdust and paddy straw; the latter is better, as it accelerates ripening with maximum organoleptic scores. Ethrel at 1000–2500 p.p.m. concentration is most commonly used as a ripening agent for early and uniform ripening. However, at higher concentrations the organoleptic rating for taste declines.

Since fully ripe fruits become soft and have a very short shelf-life (3–4 days), they need quick disposal. Such fruits are very difficult to handle and transport to distant markets. At the peak of the season, or when the market demand is less, the fruits need an extension of storage life. This can be achieved by retarding the ripening process through the use of various chemicals, growth regulators, wax coatings, irradiation, or low temperature in a modified-atmosphere store. Growth regulators, such as 2,4-dichlorophenoxyacetic acid at 3 p.p.m., prolong storage for 8 days. However, at higher concentrations the organoleptic rating for taste is very poor. Gibberellic acid (300 p.p.m.), kinetin (100 p.p.m.), and silver nitrate ($AgNO_3$; 40 p.p.m.) extend shelf-life by 8, 7, and 6 days respectively. Food-grade wax-waxol (3–6%) increases the shelf-life for 16 days, reduces weight loss, maintains freshness, and improves the appearance of the fruit. A fungicide dip like thiobendazole (1%) or Benomyl (0.2%) prevents storage decay. The fungicide and wax emulsion could be used either in mixed form or separately. Exposure of fruit to irradiation at 10 K rad extends the storage life by 3–5 days at 26.7 °C and by 15 days at 10 °C without any decline in ascorbic acid. Low temperature can extend the shelf-life by reducing the oxidative metabolism and release of ethylene. However, the sapota fruit is highly susceptible to chilling injury. Refrigerated storage at less than 1.6 °C causes irreversible chilling injury to the ripening mechanism of the fruit; it is also adversely affected at 6–10 °C. Even ripe fruits lose their taste and flavor when stored at refrigerated temperatures. (*See* **Irradiation of Foods: Applications.**)

Short-term holding of the fruit at 4 °C, before storage at 20 °C, extends its storage life, yet normal ripening only occurs if the duration of exposure to 4 °C is less than 10 days. This combination can extend shelf-life for up to 24 days, and with satisfactory quality. However, exposure to 4 °C for too long (28 days) results in chilling injuries, and the fruits fail to ripen normally even upon transfer to 20 °C. Sapota fruit stored at 20 °C and 25 °C attain the eating-ripe stage after 10 and 9 days, respectively, and, if retained, total shelf-lives of 15 and 13 days, respectively, have been recorded. At 15 °C, fruit can be stored for 22 days, but normal ripening fails and the taste is not good; hence the recommended optimum for storage and ripening is 20 °C. When an extended storage life is desired, fruit can be stored at 20 °C in an atmosphere containing 5–10% (v/v) carbon dioxide (CO_2) – with exclusion of ethylene (C_2H_4) – and a relative humidity (RH) of 85–90%. Levels of 5% and 10% (v/v) CO_2 in the storage atmosphere extend shelf-life by 18 and 21 days, respectively, and the taste is satisfactory, whereas 20% (v/v) CO_2 adversely affects the taste, and the fruits fail to ripen normally. Ripening is delayed when oxygen is removed from the storage atmosphere, but levels of oxygen higher than those in air have no apparent effect. Removal of C_2H_4 produced by the fruit delays the ripening by 23 days, and RH of 85–90% is optimal for sapota storage.

Fruits are packed in corrugated fiber board (CFB) boxes with honeycomb partitions or in telescopic moisture-proof CFB cartons. Unripe sapota fruits are packed in polyethylene bags (500 gage); placing therein a perforated plastic vial containing calcium chloride (10 g per kg of fruits) has increased the storage life for more than 12 days. When kept in a polyethylene bag with permanganate silica gel at 10–12 °C, the fruit keeps well for 18 days rather than 12 days at room temperature. The storage life of ripe fruit could be extended by 6 weeks at 2–3 °C and 85–90% RH. Fruits ripened under RH that is too high or too low are of poor quality. However, a standard method has been developed for better storage and transportation for the distance and export market. The fruits are harvested (6–9 a.m.) with 80% maturity, weighing 65–70 g (nos 145–150) and with a round to oval shape. Fruits are washed and precooled (hydrocooled) in chilled water (12 °C) containing Benomyl (500 p.p.m.) for 20 min and then dried under a fan. Healthy and firm fruits are packed in CFB boxes ($20.5 \times 20.5 \times 5.5$ cm for 16 fruits and $20.5 \times 30.0 \times 5.5$ cm for 24 fruits) containing honeycomb partitions and air holes. Boxes are closed and

fixed with sellotape. They are stored and transported at 13–14 °C and 90% RH. Generally, the shelf-life of fruit is about 25 days with acceptable quality. Fruit can be ripened at 25 °C to a desired eating quality. (*See* **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs.)

Industrial Uses

Studies conducted in India and elsewhere do not promise a bright future for the industrial processing of sapota fruit. The principal products tested were poor in characteristic flavor and storage stability. However, drying seems to give some promise. Among various drying techniques, sun-drying is the cheapest and easiest. In this method, slices are prepared by steeping in a sugar solution (40%) containing potassium metabisulfite (KMS) (1%) for 20 min (or dipping in 2% KMS for 2 min), and then sun-dried at a maximum tray temperature of 50 °C for 33 h. The dried (16% moisture) slices, packed in 300-gage polyethylene bags, maintain an acceptable organoleptic quality for 9–11 months at ambient temperature (20–35 °C). An analysis of the sun-dried product is given in Table 3. Sun-dried sapota (9.82% moisture) contained carbohydrate (73.83%), protein (2.61%), crude fat (2.30%), crude fiber (9.70%), minerals (1.74%), calcium (94.55 mg%), phosphorus (36.00 mg%), iron (2.16 mg%), vitamin C (3.50 mg%), and vitamin A (466 in per 100 g). The studies conducted to standardize the technique of sun-drying of sapota fruit slices of cv. Kalipatti demonstrated that the prepared ripe sapota fruit slices seeped in sugar syrup (70%) containing KMS (1%) for 3 h and then dried in the sun for 37 h at a maximum temperature range of 45–53 °C gave a superior dry product. The product had 1.5% moisture and better recovery: drying ratio. These dried sapota slices maintained high organoleptic quality in terms of taste, texture, color, appearance, and physicochemical quality. The product retained its acceptability –throughout the period of storage – 11 months. Therefore, this drying technique is promising for commercialization (Table 4).

Dehydrated sapota can also be prepared by drying the fruit halves, initially at 70 °C and finally at 55 °C for 4 days (8 h operation each day); the yield is approximately 30% dried product on the basis of fresh slices. The dried product, cut and packed in 400-gage polyethylene bags, can remain in good condition for 1 year at 5 °C, or for 6 months under ambient conditions (25–35 °C). Dried sapota slices prepared by predrying of sapota slices through osmotic process by steeping the fresh sapota slices in an osmotic medium containing 80% sugar syrup (70 °Brix), 20% glycerol, 0.5% KMS for 6 h, followed by draining and further drying in an electric tray drier

Table 3 Properties of dried sapota fruits

Treatment	*Recovery of dried fruit (%)*	*Moisture (%)*	*Total sugars (%)*				*Taste rating (1–10)*		
			Storage period (months)						
			Ash (%)	*3*	*7*	*11*	*3*	*7*	*11*
Potassium metabisulfite (KMS) (2%)	30.21	16.26	0.123	49.19	51.18	54.05	7.34	6.93	5.00
Sugar syrup (40%) plus KMS (1%)	33.88	16.07	0.103	52.91	54.81	47.85	7.38	7.39	4.75
Untreated	33.54	17.01	0.088	50.47	51.42	49.57	7.42	6.57	3.00

Reproduced from Vaghani SN and Chundawat BS (1986) Sun drying of sapota (*Achras sapota* L.) fruits. *Indian Food Packer* 40(2): 23–28, with permission.

Table 4 Effect of various predrying treatments on physicochemical and organoleptic quality of dried sapota slices

Parameters	*Treatments*		
	Untreated	*SS (70%)*	*SS (70%) + KMS (1%)*
Drying ratio	3.54	3.07	3.02
Dried recovery	28.20	32.53	33.13
Reducing sugars (%)	39.54	31.70	30.60
Total sugars (%)	50.63	56.00	56.39
Color rating (9)	4.60	6.30	6.90
Appearance (9)	5.40	6.30	6.90
Texture rating (9)	6.50	6.80	6.80
Taste rating (9)	6.44	6.98	7.14

SS, sugar syrup; KMS, potassium metabisulfite.
Reproduced from Vaghani SN and Chundawat BS (1997) Processing of sapota: sun drying technique. *Journal of Applied Horticulture* 3(1 & 2): 1–7, with permission.

Table 5 Effect of various methods of drying on physicochemical and organoleptic quality of dried sapota slices

Parameters	*Drying methods*			
	Sun-drying	*Solar-drying*	*Tray-drying*	*Vacuum-drying*
Total suspended solids	62.52	64.32	61.28	60.39
Reconstitution ratio	1.83	1.89	1.93	1.75
Acidity (%)	0.12	0.14	0.14	0.15
Reducing sugars (%)	27.58	28.63	28.51	27.54
Total sugars (%)	40.24	38.34	39.83	37.17
Nonenzymatic browning	71.51	74.56	80.21	67.66
Moisture (%)	12.38	13.95	11.81	13.52
Total solids (%)	87.62	85.96	88.19	86.36
Total SO_2 (p.p.m.)	780	794	728	668
Color (10)	5.40	6.40	6.65	5.98
Texture (10)	6.16	6.26	6.28	6.93
Flavor (10)	5.56	6.03	5.91	6.13
Taste (10)	6.34	6.62	7.12	6.19
Appearance (10)	5.26	6.90	6.86	6.00
Total acceptability (50)	30.54	33.12	35.04	31.44

Reproduced from Vaghani SN and Chundawat BS (1997) Processing of sapota: sun drying technique. *Journal of Applied Horticulture* 3(1 & 2): 1–7, with permission.

at 60 °C for 8 h, are better in eating quality and have longer keeping quality. The recovery of the product by this drying process is 360 g kg^{-1} fresh sapota slices (Table 5). However, the technique of drying is in need of standardization; investment in the costly machinery for dehydration could prove risky and development could result in failure. Therefore, practically no such sapota dehydration units exist on a commercial scale. However, home-scale and cottage-scale sun-drying can be attempted, as it involves no risky investment. In steeping preservation, the product prepared by steeping the ripe sapota fruit slices in a sugar syrup (50 °Brix) containing sodium benzoate or potassium sorbate at 1 g kg^{-1}, KMS (0.5 g kg^{-1}), ascorbic acid (1 g kg^{-1}) and citric acid (1.5 g kg^{-1}), are most stable and organoleptically acceptable with an increased level of drained weight, which kept well for more then 1 year. The quality of product under controlled temperature (4–5 °C) was superior compared to ambient conditions (22–35 °C). This technology can be exploited with advantage both at cottage as well as industrial scale (Table 6).

Freeze-drying has shown commercial promise; one such unit has been installed and is running on a commercial scale in Gujarat State, India. Freeze-drying at 20 °C and 0.30–0.05 mmHg pressure gives a stable and good-quality product. The product retains its natural size and shape, but is lighter in weight owing to its low moisture status (< 2%). Studies have shown that the acceptable quality of wine can be prepared from clarified sapota juice. The juice is extracted from pulp using pectinase enzyme (pectolytic enzyme – Trizyme – 0.3%), incubating for 4 h and cold pressing. The total suspended solid (TSS) of juice is adjusted to 25% and pasteurized. The must is inoculated with old culture of wine yeast (*Saccharomyces cerevisiae* var. *ellipsoideus*) at 2% for 24 h and siphoned, then incubated at room temperature for 10 days. The fermented juice is siphoned out and further kept for stabilization for 10 days using potassium tartrate and again siphoned out. Finally, clarified wine is preserved and bottled using a suitable method. Ripe fruits, especially overripe or 'reject' fruit, can be used in liqueur production, or as raw material for the manufacture of industrial glucose and pectin. Other sapota products are canned pulp or slices, jam, jelly, fruit bars, or ready-to-serve beverages. However, these products are often very poor in terms of characteristic flavors, taste, color, and storage stability. The potential of milk-blended products needs to be examined. Sapota fruits, being rich in sugar, are also suitable for making fermented products. Its sap, known as chicle, after much refining and flavoring, becomes chewing gum. A single tree can yield up to 70 l of chicle in just a few hours. The tree is tapped for this, once in 3 years. The bark of the sapota tree also contains about 12% of tannin, which has excellent industrial value. (*See* **Drying**: Drying Using Natural Radiation; **Freeze-drying**: Structural and Flavor (Flavour) Changes; **Preservation of Food**.)

Chrysophyllum cainito L.

Vernacular names are West Indian star apple, cainito, and tulsiphal. It is a native of the West Indies and Central America and is cultivated in warmer parts of India for its delicious edible fruit. The fruits are shining apple-sized berries. The immature fruits are

Table 6 Effect of different treatments on overall physicochemical and sensory quality of steeped sapota slices during storage (average score values of 1 year)

Treatments	*Storage*	*Drained weight (%)*	*pH*	*Acidity (%)*	*Ascorbic acid (mg 100 g^{-1})*	*TSS (%)*	*Reducing sugars (%)*	*Total sugar (%)*	*Color (10)*	*Appearance (10)*	*Texture (10)*	*Taste (10)*	*Flavor (10)*	*Acceptability (10)*	*Total (60)*	*Overall organoleptic score (%)*
SB (0.85) + CA (2.3) + AA (0.2)	AT	56.21	3.94	0.29	3.15	35.28	22.90	26.63	6.21	6.54	8.14	6.91	6.61	9.06	43.47	72.45
SB (1.0) + KMS (0.5) + CA (1.5) + AA (1.0)	AT	57.62	4.26	0.25	16.82	34.95	22.88	26.35	7.05	7.34	8.54	6.49	6.18	9.50	45.10	75.17
PS (1.0) + CA (2.3) + AA (0.2)	AT	56.43	4.00	0.28	3.36	35.25	22.92	26.66	6.03	6.33	8.24	7.13	6.79	9.21	43.73	72.88
PS (1.0) + KMS (0.5) + CA (1.5) + AA (1.0)	AT	56.77	4.30	0.24	16.99	35.20	22.54	26.57	7.21	7.53	8.86	6.69	6.40	9.50	46.19	76.98
SB (0.85) + CA (2.3) + AA (0.2)	LT	57.07	4.16	0.26	4.87	33.60	22.45	25.63	7.78	8.06	8.03	8.44	8.21	8.21	48.73	81.22
SB (1.0) + KMS (0.5) + CA (1.5) + AA (1.0)	LT	57.78	4.41	0.22	31.96	33.59	21.61	25.58	9.21	9.50	8.76	8.19	7.96	9.89	53.51	89.18
PS (1.0) + CA (2.3) + AA (0.2)	LT	56.43	4.19	0.25	5.17	33.50	21.91	25.42	8.33	8.61	8.10	8.38	8.15	8.42	49.99	83.32
PS (1.0) + KMS (0.5) + CA (1.5) + AA (1.0)	LT	58.17	4.45	0.22	32.65	33.45	21.25	25.31	9.40	9.63	8.81	8.29	8.04	9.94	54.11	90.18

TSS, total suspended solids; SB, sodium benzoate (g kg^{-1}); PS, potassium sorbate (g kg^{-1}); KMS, potassium metabisulphite (g kg^{-1}); CA, citric acid (g kg^{-1}); AA, ascorbic acid (g kg^{-1}); AT, ambient temperature (22–35 °C); LT, low temperature (4–5 °C).

Reproduced from Vaghani SN and Chundawat BS (1999) Standardization of steeping preservation technology of sapota fruit slives of cv Kalipatti. A PG thesis. India: Gujarat Agricultural University.

astringent and contain a sticky white latex. The fruit is bright green and greenish-yellow to red violet when ripe, and should be eaten when fully ripe. They are quite delectable, full of cool, refreshing snow-white pulp with a puzzling flavor. When allowed to ripen on the tree, the fruits develop a sweet, delicious flesh with a sapotaceous flavor and can be eaten fresh or in preserves.

The fruits are hard, smooth, purple or light green, and globular (5–10 cm in diameter), with five or six distinct ribs. The seeds (4–10) are embedded in white, translucent, edible pulp (star-shaped in transverse section) and arranged stellately inside the flesh. An analysis of fruits (flesh) from Hawaii gave the following data: dry matter, 11.47%; protein, 2.34%; fat, 1.39%; acids as malic, 0.17%; total sugars, 4.40%; fiber, 0.86%; ash, 0.39%. The seeds also contain a bitter substance, lucumin (1.2%), as well as oil (6.6%), saponin (0.19%), dextrose (2.4%), and ash (3.75%).

Madhuca indica Hamilton ex. J.F. Gmel. (Syn. *M. latifolia* Macbr; *Bassia latifolia* Roxb; *M. longifolia* L., *B. longifolia* Koenig.)

Vernacular names are illippe, buttertree, mahuva, and mowra.

Type and Commercial Importance

The fruit is a berry. The outer part is eaten raw or cooked and the inner part is made into a flour for cakes. It is valued for its seeds, which yield an oil (vegetable butter) known in commerce as illeppe butter, bassia fat, mahua butter, or mowra fat. The oil is used direct or for cooking purposes in some rural areas, and it is sometimes used as an adulterant for ghee; for this purpose, it is clarified with buttermilk to mask the odor. It is also used in the production of margarine. Madhuca oil has emollient properties and is used to treat skin diseases, rheumatism, and headaches. It also has laxative properties and is considered useful for the treatment of habitual constipation and hemorrhoids (piles); it can be used as an emetic. The seeds are employed as a galactagogue. The oil cake is used as a cheap substitute for shikakai for washing hair. The succulent carolla are a rich source of sugar and also contain appreciable quantities of vitamins, calcium, and essential oils. These are eaten raw or cooked. The flowers are largely used in the preparation of country-made distilled liqueurs. The flowers are also used in the preparation of vinegar and essential oil (0.05%).

Morphology and Anatomy of the Fruits

The berries are ovoid (up to 5 cm long), fleshy and greenish in color, turning reddish-yellow or orange when ripe. The seeds (usually one or two but up to four in var. *latifolia*) are ovoid, pale brown or yellow, compressed and shiny.

Chemical and Nutritional Composition

The fruit contains the following: moisture, 73.64%; carbohydrates, 22.69; protein, 1.73%; fat, 1.61%; mineral matter, 0.69%; calcium, 45 mg per 100 g; phosphorus, 22 mg per 100 g; iron, 1.1 mg per 100 g; ascorbic acid, 40.5 mg per 100 g; carotene (as vitamin A) 512 IU per 100 g; tannins are also present. Seeds available in the trade contain 5–12% moisture and the seed kernel contains the following: oil, 51.1%; protein, 8.0%; N-free extract, 27.9%; fiber, 10.3%; ash, 2.7%. The seed oil has a specific gravity of 0.856–0.870, an acid value of 5–50, a saponification value of 188–200, an iodine value of 53–70; unsaponifiable matter accounts for 1–3%.

Handling and Industrial Uses

In general, the fresh berries are not stored, but the seeds, having a moisture content of less than 7–8%, can be. Traditionally, the madhuca tree was raised, maintained, and saved for its fleshy, sugary flowers which, as they contain 73% sugar and possess a strong flavor, were used to produce liqueur. In India, this tree is saved and maintained by tribal communities only for this purpose (country liqueur). Traditionally, the fresh berries do not have any importance for table use but the seeds are valued for their oil. The fruit ripens in May–June. The seeds are separated from the fruit wall by pressing, and then dried and shelled to obtain the kernel (70% of the weight of the seed). The kernel constitutes the seeds of commerce. The seed kernel yields 20–30% oil when crushed in a primitive crusher, 34–37% in an expeller, and 40–43% when extracted with solvents. Fresh madhuca oil from properly stored seeds is yellow in color without any unpleasant taste. Commercial oils are generally greenish-yellow in color with an offensive odor and disagreeable taste. The oil is fluid, often throwing out a deposit of stearine in cold weather when the oil solidifies to a buttery consistency. The crude oil has a deep color with a high acidity and an unpleasant odor. Refining and hydrogenation yield a product similar to mutton tallow or cacao butter. Oils having an acid value below 13 may be refined by treating with caustic soda; those with higher acid values are extracted with alcohol and further treated with alkali.

The whole berries can be used in the production of absolute alcohol. The fruits, collected near the ripening stage, are rich in starch, which is hydrolyzed into sugars in 2–3 days. The fruit pulp is used as a source of sugar for alcoholic fermentation and 1 t of dry fruits (freed from seeds) yields 130 l of absolute alcohol. Steam distillation of the fruit yields a bright-yellow, volatile oil with a spicy odor.

Mimusops elengi L.

Vernacular names are Spanish cherry, medlar, maulasri, bakul, and borsali. The ripe fruits are edible, and the berry is yellow, pointed, smooth, fleshy, ovoid (2.5–3.8 cm × 1.3–1.7 cm); young fruits are hairy. Seeds are solitary, rarely two, ovoid (1.5–2.3 cm × 1.0–1.3 cm), blackish or grayish-brown, compressed, shining, smooth, with a hard testa and soft, whitish albumen. Ripe fruits are sometimes used to make preserves and pickles. The seed kernel contains a fatty oil (16–25%). The flowers are also used in distilling otto used in perfumes and stimulants. The tree also yields gum and bark tannin (3–7%).

Manilkara hexandra (Roxb) Dubard (Syn. *Mimusops hexandra* (Roxb))

Vernacular names are bulletwood, Ceylon iron wood, Ceylon wood, khirni, raini, and pal. Ripe fruits, which are sweet but astringent, are eaten fresh or dried. The seed oil is useful for cooking, and are sometimes also used for adulterating ghee. The seeds are useful as a tonic, and are prescribed for leprosy and delirium.

Morphology and Anatomy of the Fruits

The fruit is a fleshy berry, 1.5–2.5 cm long, ovoid or ellipsoid. On ripening, the green berry changes to reddish-yellow. The seeds (one or, rarely, two) weigh 13.2 g per 100 seeds, and are 1.0–1.5 cm long, ovoid, smooth, shining, and reddish-brown.

Chemical and Nutritional Composition

The composition of the ripe fruit is as follows: moisture, 68.61%; carbohydrates, 27.74%; fat, 2.42%; protein, 0.48%; mineral matter, 0.75%; calcium, 83 mg per 100 g; phosphorus, 17 mg per 100 g; iron, 0.92 mg per 100 g; nicotinic acid, 0.66 mg per 100 g; ascorbic acid, 15.67 mg per 100 g; carotene (as vitamin A), 675 IU per 100 g; thiamin, 70.33 μg per 100 g; riboflavin, 77.41 μg per 100 g. The seed contains 24.6% edible oil, having a specific gravity of 0.9150, an acid value of 1.34, a saponification value of 191.1, and an iodine value of 65.1; unsaponifiable matter (probably containing ergosterol) accounts for 1.02%. The oil has medicinal value. The plant also yields tannin (from bark) and gum.

Handling and Storage

As the fresh fruit has little commercial value, no special care is required for handling and storage. However, to minimize losses, it is harvested as per market demand. The remaining berries are left on the tree for drying. Thus, as far as possible, the storage of fresh fruit is avoided. However, the dried fruits, which are packed in gunny bags and stored at ambient temperature, remain in good condition for 1 year.

Industrial Uses

The surplus fruits are used for drying. The ripe berries are shaken from the branches of the tree, or fall naturally and are then collected and spread on mats in a thin layer for sun-drying. After proper drying, they are packed in gunny bags and stored. The seeds are used for the extraction of oil by pressing in a primitive crusher. The oil is pale yellow in color with an odor reminiscent of olive oil.

Pouteria campechiana (Kunth) Baechni. (syn. *Lucuma nervosa* A.DC., L. *Salicifolia* (Kunth))

Vernacular names are eggfruit, canistel, sapote borracho, sapote amarillo, and ti-es. The fruit is edible, subglobose or pear-shaped (10 cm long), and orange or yellow; the sweet pulp has a musky flavor.

Pouteria sapota (Jacq.) H.E. Moore & Stearn. (syn. *Calocarpum Mammosum* (L.) Pierre; *C. sapota* (Jacq.) Merrill, *Lucuma mammosa* (L.) C.F. Gaertner)

Vernacular names are sapote, mammees, maneys sapota, maney colorado, marmalade plum, and marmalade fruit. The fruit is very sweet and edible. An important fruit in lowland, Caribbean areas, it is eaten fresh or made into thick preserves, often in combination with guava fruit.

Pouteria gardneriana (A.DC.) Radlk. (syn. *Pouteria suavis* W.D. Hemsley)

The fruits are highly scented and possess a thin, edible pericarp; they have an agreeable taste and are easily digestible. The fruit is about the size of an apricot and the shape of an apple, and is yellow and scarlet when mature. The seed is semiovoid, rather like a large hazel nut.

Synsepalum dulcificum (Schumach. & Thonn.) Daniell

In West Africa, acid citrus fruits are said to taste sweet after chewing the fruits of this plant; whenee is the common name. The fruit is red, oblong to ovoid (2 cm long, 1 cm in diameter).

Xantolis tomentosa (Roxb.) Rafrain (Syn. *Sideroxylon tomontosum*; Roxb; *Panteria tomentosa* (Roxb) Baehni)

Vernacular names are kanta, kumbul, and kumla. Fruit is a berry (2.0–2.5 cm × 1.3–1.7 cm), pubescent when young, afterward glabrous, yellowish-green, ovoid, and smooth. Seeds are usually solitary (sometimes there are up to five), 1.3 cm long, deep brown, smooth, polished and shining, and compressed. The fruit is eaten in curries and made into pickles. The seeds are also reported to yield a fatty oil (41%). It yields a saponin (2%) and fruit extract is credited with anticholeric principles. The black seeds are rubbed in lemon juice and given as an astringent. The plant also yields gutta percha.

See also: **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Drying**: Drying Using Natural Radiation; **Freeze-drying**: Structural and Flavor (Flavour) Changes; **Irradiation of Foods**: Applications; **Phenolic Compounds**; **Preservation of Food**; **Ripening of Fruit**; **Tannins and Polyphenols**

Further Reading

Anonymous (1948–1976) *The Wealth of India – Raw Material*, vols 1–10. New Delhi; CSIR, Publications and Information Directorate.

Broughton WJ and Wong HC (1979) Storage conditions and ripening of chiku fruits *Achras sapota* L. *Scientia Horticulture* 10: 377–385.

Gopalan C, Sastri BVR and Balasubramanian SC (1977) *Nutritive Value of Foods*. Hydrabad: Nutro.: Indian Council of Medical Research.

Hernandez M, Chavez A, Bourges H and Mendoza E (1974) *Nutritive Value of Foods* (in Spanish) Mexico City: National Institute of Nutrition.

Nagy S and Shaw PE (eds) (1980) *Tropical and Subtropical Fruits*. Westport, Connecticut: AVI Publishing.

Selvaraj Y and Pal DK (1984) Changes in the chemical composition and enzyme activity of two sapodilla (*Malinkara zapota*) cultivars during development and ripening. *Journal of Horticultural Science* 59(2): 275–281.

Vaghani SN and Chundawat BS (1986) Sun drying of sapota (*Achras sapota* L.) fruits. *Indian Food Packer* 40(2): 23–28.

Vaghani SN and Chundawat BS (1995) Export of mango and sapota from Gujarat: a case study. *Journal of Applied Horticulture* 11(1 & 2): 45–47.

Vaghani SN and Chundawat BS (1997) Processing of sapota: sun drying technique. *Journal of Applied Horticulture* 3(1 & 2): 1–7.

Vaghani SN and Chundawat BS (1997) *Development of Processed Products Based on Sapota*. Research Report. New Delhi: ICAR.

Vaghani SN and Chundawat BS (1999) Standardization of steeping preservation technology of sapota fruit slices of cv Kalipatt. APG thesis. India: Gujarat Agricultural University.

Lesser-known Fruits of Africa

R J du Preez, ARC-Institute for Tropical and Subtropical Crops, Nelspruit, South Africa

Background

In Africa, there are numerous fruit species that are well known and highly prized by local people who often depend on these trees for food security. Africa has a wide diversity of fruits that provide a source of vitamins, minerals, amino acids, and trace elements to rural populations. These wild resources play an important role in the diet of rural people, especially in times of food scarcity. Although there are numerous different fruits, little is known of them except by the local people. Very few fruits from Africa have been developed commercially.

Some of the trees outlined in this paper have the potential to be developed on a small scale and others on a larger scale, if correct market development is undertaken. The development of value adding products will result in a larger market, job creation, and improved quality of life in rural areas. Conservation of the natural biodiversity will also be achieved.

Some of the most common fruits found in Africa are listed in **Table 1**. This is not a comprehensive list, but includes some of the most common fruits. A more detailed outline is given on 12 of the fruits listed in **Table 1**.

Adansonia digitata (Baobab)

Adansonia digitata (baobab) is a spreading tree, about 10–15 m in height. The trees reach a great age and girth. The leaves are alternate, compound usually three- to nine-foliate. Flowers are waxy white and up to 20 cm in diameter, solitary, and pendulous. The fruit is ovoid, 120 mm or longer in length with a hard woody shell, covered with yellowish gray, velvety hairs (see **Figure 1**).

Table 1 Some of the lesser-known fruits of Africa and their uses

Species	*Common name*	*Family*	*Use*
Adansonia digitata L.	Baobab	Bombacaceae	Edible fruit, seeds used in making a refreshing drink, subsistence from wild
Annona senegalensis Pers.	Wild custard apple	Annonaceae	Edible fruit, subsistence from wild
Annona stenophylla Engl. & Diels		Annonaceae	Edible fruit, staple part of diet in certain areas
Antidesma venosum E. Meyer ex Tul.	Tassel berry	Euphorbiaceae	Edible fruit, jelly, subsistence from wild
Artabotrys brachypetalus Benth.	Purple hook-berry	Annonaceae	Edible fruit, alcoholic beverage, subsistence from wild
Azanza garckeana (F. Hoffm.) Exell & Hillc.	Snot apple/Azanza/ Murojwa	Malvaceae	Edible fruit, jelly, subsistence from wild
Berchemia discolor (Klotzsch) Hemsley	Bird plum	Rhamnaceae	Edible fruit, sold in local markets
Borassus aethiopum Mart.	Borassus palm	Arecaceae	Fruit edible, plant sap used to make wine, subsistence from wild
Bridelia carthatica Bertol. F.	Knobbly bridelia	Euphorbiaceae	Edible fruit, jelly, subsistence from wild
Carissa bispinosa (L) Desf. ex Brenan	Num-num	Apocynaceae	Edible fruit, subsistence from wild
Carissa edulis Vahl	Simple spined num-num/carissa	Apocynaceae	Edible fruit, subsistence from wild
Carissa haematocarpa (Ecklon) A. DC.	Karoo num-num	Apocynaceae	Edible fruit, subsistence from wild
Carissa macrocarpa (Ecklon) A. DC	Large num-num	Apocynaceae	Edible fruit, subsistence from wild, confined to South Africa
Carpobrotus edulis	Sour fig	Mesembryanthemaceae	Edible fruit, subsistence from wild
Cordyla africana Lour.	Wild mango	Fabaceae	Edible fruit, contains ascorbic acid, subsistence from wild
Dialium schlechteri Harms	Zulu podberry	Fabaceae	Fruit made into refreshing drink, subsistence from wild
Diospyros kirkii Hiern.	Pink diospyros	Ebenaceae	Edible fruit, jelly, subsistence from wild
Diospyros mespiliformis Hochst. ex A. DC.	Jackal berry/Ebony diospyros	Ebenaceae	Edible fruit, fruit preserve, beer, subsistence from wild
Dovyalis abyssinica (A. Rich) Warb.	Abyssinica gooseberry	Flacourtiaceae	Edible fruit, jelly, subsistence from wild
Dovyalis caffra Hook & Harvey	Kei apple	Flacourtiaceae	Edible fruit, jelly, jam, cultivated as protective hedge
Dovyalis hebecarpa Gard	Ketembilla/ceylon gooseberry	Flacourtiaceae	Edible fruit, jelly
Dovyalis zeyheri (Sonder) Warb.	Oval kei apple/Wild apricot	Flacourtiaceae	Edible fruit, jelly, subsistence from wild
Ekebergia benguellensis Welw. ex C.DC.	Woodland dog plum	Meliaceae	Edible fruit, subsistence from wild
Englerophytum magalismontanum	Stem fruit	Sapotaceae	Edible fruit with ascorbic acid, wine, 'brandy', syrup, subsistence from wild
Ficus capreifolia Delile	Sandpaper fig/riverine sandpaper fig	Moraceae	Edible fruit, subsistence from wild
Ficus ingens Miq.	Red-leaved rock fig	Moraceae	Edible fruit, subsistence from wild
Ficus kirkii Hutch.	Crown-fruited fig	Moraceae	Edible fruit, subsistence from wild
Ficus natalensis Hochst.	Wild fig	Moraceae	Edible fruit, subsistence from wild
Ficus sycomorus L.	Sycamore fig	Moraceae	Edible fruit, subsistence from wild
Flacourtia indica Merrill	Flacourtia	Flacourtiaceae	Edible fruit, subsistence from wild
Friesodielsia obovata (Benth.) Verdc.	Bastard dwaba-berry/ monkey fingers	Annonaceae	Edible fruit, subsistence from wild
Garcinia huillensis Welw. ex Oliv.	Granite garcinia	Clusiaceae	Edible fruit, contains ascorbic acid, subsistence from wild
Garcinia livingstonei T. Anders.	African mangosteen	Clusiaceae	Fruit eaten raw, jelly, subsistence from wild
Grewia bicolor Juss.	Bastard brandybush	Tiliaceae	Edible fruit, subsistence from wild
Grewia flava DC	Velvet raisin bush/ Brandybush	Tiliaceae	Edible fruit, subsistence from wild

Continued

Table 1 Continued

Species	*Common name*	*Family*	*Use*
Harpephyllum caffrum Bernh. ex Krauss	Wild plum	Anacardiaceae	Fruits eaten raw, jelly, wine, subsistence from wild
Hexalobus monopetalus (A. Rich.) Engl. & Di*els*	Shakama plum/ Baboon's breakfast	Annonaceae	Edible fruit, subsistence from wild
Hyphaene petersiana		Arecaceae	Edible fruit, palm wine from juice of young plants, crafts, subsistence from wild
Hyphaene ventricosa		Arecaceae	Edible fruit, subsistence from wild
Lannea discolor (Sond.) Engl.	Live-long	Anacardiaceae	Edible fruit, subsistence from wild
Lannea edulis (Sond.) Engl.	Wild grape	Anacardiaceae	Edible fruit, subsistence from wild
Manilkara mochisia (Baker) Dubard	Lowveld milkberry	Sapotaceae	Edible fruit, subsistence from wild
Mimusops caffra E. Meyer ex A.DC	Coast red milkwood	Sapotaceae	Edible fruit, subsistence from wild
Mimusops obovata Sonder	Red milkwood	Sapotaceae	Edible fruit, subsistence from wild
Mimusops zeyheri Sonder	Moepel/Common red milkwood Dwaba berry	Sapotaceae	Edible fruit, subsistence from wild
Monanthotaxis caffra (Sonder) Verdc.		Annonaceae	Edible fruit, subsistence from wild
Nylandtia spinosa	Skilpadbessie	Polygalaceae	Edible fruit, thirst quenching drink, subsistence from wild
Olea africana Miller	Wild olive	Oleaceae	Edible fruit, subsistence from wild
Pappea capensis Ecklon & Zeyher	Jacket plum	Sapindaceae	Fruits eaten raw, made into jelly, alcoholic beverage, vinegar, seed produces oil, subsistence from wild
Parinari curatellifolia Planch. ex Benth.	Mobola plum	Chrysobalanaceae	Edible fruit, subsistence from wild
Phoenix reclinata Jacq.	Wild date palm	Arecaceae	Edible dates, subsistence from wild
Rhus lancea L. f.	Karee	Anacardiaceae	Edible fruit, fruit used to make beer, subsistence from wild
Salacia kraussii	Salacia	Celastraceae	Edible fruit, subsistence from wild
Schinziophyton rautanenii	Manketti tree/ Mangongo Marula	Euphorbiaceae	Large, edible oily nut, subsistence from wild
Sclerocarya birrea Sonder		Anacardiaceae	Eaten raw, juice, alcoholic beverage, jelly
Securinega virosa (Roxb. ex Willd) Baillon	White berry bush/ Snowberry tree	Euphorbiaceae	Edible fruit, fruit has medicinal uses, subsistence from wild
Strychnos cocculoides Baker	Corky bark monkey orange	Loganiaceae	Large edible fruits, harvested from wild and sold locally
Strychnos innocua Delile	Dull leaved mukwakwa	Loganiaceae	Edible fruit, subsistence from wild
Strychnos madagascariensis Poir.	Black monkey orange/ shiny leaved mukwakwa	Loganiaceae	Edible fruit, fruit dried and stored, subsistence from wild
Strychnos pungens Solereder	Spiny leaved monkey orange	Loganiaceae	Fruit edible when ripe, subsistence from wild
Strychnos spinosa Lam.	Green monkey apple/ spiny monkey apple Morutla	Loganiaceae	Edible fruit, green fruit used medicinally, subsistence from wild
Syzygium cordatum Hochst.	Waterberry	Myrtaceae	Edible fruit, alcoholic drink, subsistence from wild
Syzygium guineense (Willd.) DC.	Woodland waterberry	Myrtaceae	Edible fruit, subsistence from wild
Tabernaemontana elegans Hochst. ex A. DC.	Small fruited toad tree	Apocynaceae	Edible fruit, subsistence from wild
Tamarindus indica L.	Tamarind	Fabaceae	Fruit used commercially
Terminalia sericea Burch.	Silver terminalia/ Mangwe	Combretaceae	Edible fruit, subsistence from wild

Continued

Table 1 Continued

Species	*Common name*	*Family*	*Use*
Uapaca kirkiana Muell. Arg.	Mahobohobo	Euphorbiaceae	Eaten raw, fruit made into a pleasant wine, subsistence from wild
Uapaca nitida Muell. Arg	Narrow leaved mahobohobo	Euphorbiaceae	Edible fruit, wood makes a good charcoal, subsistence from wild
Uapaca sansibarica Pax	Lesser mahobohobo	Euphorbiaceae	Edible fruit, subsistence from wild
Vangueria infausta Burch.	Wild medlar	Rubiaceae	Edible fruit, subsistence from wild
Vangueriopsis lanciflora (Hiern) Robyns	False wild medlar	Rubiaceae	Edible fruit, subsistence from wild
Vitex doniana Sweet	Black plum	Verbenaceae	Edible fruit, subsistence from wild
Vitex mombassae Vatke	Smelly berry vitex	Verbenaceae	Edible fruit, source of ascorbic acid, subsistence from wild
Vitex payos (Lour.) Merr.	Chocolate berry	Verbenaceae	Edible popular fruit with coffee-like flavor, subsistence from wild
Ximenia americana L.	Small sour plum	Olacaceae	Edible fruit, oil from seed, subsistence from wild
Ximenia caffra Sond.	Large sour plum	Olacaceae	Edible fruit, oil obtained from seeds, subsistence from wild
Ziziphus mucronata Willd.	Buffalo thorn	Rhamnaceae	Edible fruit, subsistence from wild
Zizyphus mauritiana Lamt.	Jujube	Rhamnaceae	Edible fruit, subsistence from wild

Figure 1 Fruit of *Adansonia digitata* (baobab).

The baobab plays an important role in the economy of the local inhabitants, throughout its distribution. Practically all parts of the tree are utilized. In West Africa, the roots are reputed to be cooked and eaten mainly in times of famine. The trunks are often hollowed out and used for a variety of purposes, the most important being the storage of water. A medium-sized tree can hold 1600 l, while a large tree can contain as much as 8000 l.

Bark fibers are commonly stripped from the lower trunk. The tree, however, survives and generates new bark. The fiber from the inner bark is tough and durable and is widely used for rope, harness straps, strings for musical instruments, baskets, nets, fishing lines, and snares. Leaves, especially the young leaves, are popular as a spinach-type vegetable or are dried and powdered and made into soups and sauces. Fresh leaves are rich in ascorbic acid (vitamin C) and contain uronic acids, rhamnose and other sugars, tannin, and potassium tartrate. Leaves of the baobab are also used medicinally – as an expectorant, against fever, as an astringent and to treat kidney and bladder diseases, asthma, diarrhea, inflammations, insect bites, and general fatigue.

The husks of the fruits may be used as dishes or fashioned into vessels. They can also be used as fuel and provide a potash-rich ash suitable for soap-making. Husks, ground into a powder, are also used as a substitute for tobacco. The acid pith, which is rich in ascorbic acid, is used as a substitute for cream of tartar for baking. It is also ground and made into a refreshing drink. The pulp is also used to curdle milk. In East Africa, an extract from the pulp is used as a hair wash.

Seeds are embedded in a powdery white pulp, which contains tartaric acid and potassium bitartrate. The seeds may be eaten fresh or dried, either sucked or ground, and used to flavor foods. They are refreshing to suck and, when soaked in water, make a palatable drink. They are also used as a substitute for coffee. Crushed and roasted seeds can be applied as a paste to diseased teeth and gums.

Carissa bispinosa (Num-num)

The num-num belongs to the Apocynaceae family. *Carissa bispinosa* is the most common species, but

Figure 2 *Carissa macrocarpa* tree (num-num).

Figure 3 (see color plate 68) Flowers and fruit of *Carissa macrocarpa* (num-num).

other *Carissa* species occur and are utilized by local communities. The other species include *Carissa macrocarpa*, the large num-num, which is confined to South African coastal regions, *Carissa edulis*, the simple spined num-num, and *Carissa haematocarpa*, the karoo num-num.

The num-num is indigenous to Southern Africa and is found growing along the coast of KwaZulu-Natal, the Eastern and Western Cape, and also further inland in the Northern Province, Mpumalanga, and Guateng. In Southern Africa, the num-num is pollinated by small beetles, hawk-moths and other night flying insects. Various degrees of unfruitfulness in the USA have been attributed to inadequate pollination. Some seedlings are light croppers, and others never bear at all. It has been found that unproductive plants, apparently self-infertile, will bear fruits after cross-pollination by hand (see Figure 2).

The growth habit of the num-num is usually a somewhat scrambling shrub, or a small tree up to 5 m in height. The tree usually occurs at medium to low altitudes in dry woodland and coastal scrub. The spines are either once- or twice-forked, very occasionally simple and up to 45 mm long. The leaves are broad to narrowly ovate, leathery with a dark glossy green on the upper surface, and paler green below. The flowers are white to pinkish with a slender corolla tube. The flowers are sweetly scented and borne in terminal clusters or singly. Some plants bear flowers that are functionally male. These flowers are larger than normal with larger anthers and stamens that are much longer than the style. Functionally, female flowers have stamens the same length as the style and small anthers without pollen (see Figure 3).

The fruit is ovoid and fleshy, about 15 mm long and red when mature. The skin is thin, enclosing a firm reddish pulp in which numerous small seeds are embedded. When bruised or cut, the flesh exudes white droplets of a gummy juice. The fruit is edible when ripe.

The num-num is subtropical but survives temperatures of $-3\,^{\circ}C$ when well established. Young plants need protection if the temperature drops below zero. The best growth is obtained in full sun. The plant has moderate drought tolerance and a high resistance to soil salinity and salt spray.

The fruit, when fully ripe, is dark red and slightly soft. It can be eaten fresh and can also be used in fruit salads, icecreams, puddings, or for decoration purposes. Fruit can also be cooked to a sauce, used in pies and tarts or preserved whole by pricking, cooking briefly in a sugar syrup, and then bottling. They can also be made into jams or jellies.

Trees are often planted as a protective hedge, as they form an impenetrable barrier, and fruit is then regarded as a by-product. They are, however, slow growing, and a long time is needed before they form an effective barrier. Efforts have also been directed to the development of dwarf, compact, less spiny types for landscape use. Ornamental cultivars such as Bonsai, Dainty Princess, Low Boy, Nana, Nana Compacta, and Prostrata have been developed mainly as garden species. These dwarf cultivars must, however, be kept under control; otherwise, they are apt to revert to the original type, as vigorous shoots may develop and outgrow the compact form.

Dovyalis caffra (Kei-apple)

Dovyalis caffra is subtropical but can survive brief drops of temperatures to as low as $-6\,^{\circ}C$. It does well in almost any type of soil that does not have a high water table. It is drought-resistant and tolerates saline soil.

It grows as a shrub or small tree, usually 3–5 m in height but sometimes reaching 8 m. It is vigorous growing and rather dense and thorny. The spines are

usually 20–70 mm long, although some trees are spineless. It usually occurs in open bush and wooded grassland. Leaves are often clustered on short spurs and are oblong-obovate 20–50 mm long, glossy, and short-petioled. The flowers are small without petals and clustered in the leaf axils. Female and male flowers are usually borne on separate trees. Fruits are almost spherical, 25–40 mm long, and fleshy, with a bright yellow, smooth but tough skin. The pulp is juicy, highly acid, and apricot-textured (see Figure 4).

Figure 4 *Dovyalis caffra* tree (kei-apple).

Figure 5 (see color plate 69) Fruit of *Dovyalis caffra* (kei-apple).

Most people consider the fruit too acidic for eating out of hand, even when fully ripe. It is best cut in half, peeled, seeded, sprinkled with sugar, and allowed to stand for a few hours before serving as a dessert or in fruit salads. The fruits are pleasantly flavored and make excellent jelly and jam when ripe and underripe as pickles. They are widely cultivated both for their fruits and as a hedge. Fresh ripe fruit contains 80 mg of ascorbic acid (vitamin C) per 100 g of fruit. Fifteen amino acids have been identified in the fruit. Investigators in Egypt have demonstrated that the roots, stem, and fruit, but not the leaves and branches, possess antibiotic properties (see Figure 5).

Englerophytum magalismontanum (syn. *Bequaetiodendron magalismontanum*) (Stem Fruit)

Englerophytum magalismontanum, known locally as the stem fruit, is widely distributed from tropical Africa to KwaZulu-Natal in South Africa. It is very common in the Northern Province, Gauteng, Northwest, and Mpumalanga provinces of South Africa and occurs on rocky hills and mountain slopes in close association with granite or quartzite rocks.

It is a small to medium-sized tree 3–10 m in height with a dense bushy habit, with branches often touching the ground. The bark is brown, flaky, or scaly. The stem is straight with horizontal branches. The young twigs, leaf petioles, and floral parts are densely covered with dark brown hairs. The leaves are alternate, oblanceolate to obovate-elliptic, usually 70–140 mm long with a width of 20–30 mm, glossy dark green above with silvery to golden-brown silky appressed hairs below that may wear away on old leaves. The midrib is prominent on the abaxial leaf surface with the lateral veins prominent, immersed, but clearly visible. The leaf apex is broadly tapering, sometimes attenuate; the base is tapering and the margin entire, slightly rolled under and undulated. The petiole is brown to blackish, rough, and up to 14 mm long.

The flowers are inconspicuous, brownish pink in color and about 10 mm long. They are strongly scented and are borne in few to many flowered clusters on the leafless stems and branches. The fruits are ovoid and fleshy, 15–25 mm long, and 10–18 mm in diameter. They have a definite ridge and a small point at the top, and are densely crowded along the old wood. The fruits change from yellow to orange and finally to a dark, glossy red when fully ripe. Some trees, however, bear fruits that are white to light pink. The skin is quite thick and encloses a purplish fruit pulp that contains a sticky, milky juice, and usually one, but sometimes two, flattened seeds.

Figure 6 Fruit of *Englerophytum magalismontanum* (stem fruit).

The fruits, which ripen during the summer, are edible and have a high ascorbic acid (vitamin C) content. They have a sweet, but somewhat astringent, flavor. They may be used to make jelly, jam, or syrup. Local inhabitants eat the fruit raw but also brew a potent alcoholic drink (see Figure 6).

Traditional medicine uses an infusion of the finely stamped fruit, and the roots are regarded as a treatment for epilepsy. The pulverized bark is used as a cure for rheumatism, while powdered roots are also used to relieve headaches. As the wood is tough and durable, it is suitable for fencing poles, implement handles, and hut building, and the Zulu tribe in South Africa uses the wood to make milking buckets.

In gardens, its evergreen foliage and brightly colored fruits are decorative. The seed germinates easily when fresh, and trees may also be propagated from cuttings. The tree is relatively fast-growing and hardy. The fruit is very susceptible to fruit fly, and a spray control program should be followed.

Mimusops zeyheri (Moepel)

The fairly common moepel *Mimusops zeyheri* is also known as the Transvaal Red-milkwood. In South Africa, it is confined to the northern provinces and occurs northwards into Zimbabwe, as well as Botswana and Mozambique. It usually grows at low altitudes in hot areas. It is frequently found on well-wooded rocky hillsides, in riverine fringes, at the margins of evergreen forests, and in dry open woodlands.

It is occasionally a large shrub, but usually a small to medium-sized tree up to 15 m in height. The bark is gray to dark brown, rather smooth in young trees, later becoming rough (see Figure 7).

The leaves are scattered, ovate, lanceolate to oblong, about 110 mm in diameter, thick and leathery, shiny dark green above, and a paler green abaxially. The young leaves and twigs are covered with dense

Figure 7 *Mimusops zeyheri* tree (moepel).

hairs that are lost by maturity, and the leaf apex is broadly tapering to tapering and is often notched. The leaf base is tapering, and margins are entire and slightly thickened. The petiole is up to 30 mm long.

The normal flowering season is during November and December, when fragrant flowers appear among the leaves. The flowers are creamy white and star-shaped, about 10 mm in diameter, and clustered into axillary groups of three or more flowers, often in profusion. They give rise to green fruit, which ripen in the spring. Fruits are ovoid, fleshy, 20–30 mm in length and 10–25 mm in diameter. They are bright yellow when mature. Beneath the rather brittle skin is a yellowish flesh and several shiny brown seeds. Although the edible flesh is rather dry and floury, it contains about 65% water, 28% carbohydrates, and 90 mg of ascorbic acid per 100 g (see Figure 8).

Fruits of the common moepel are an important item in the diets of rural inhabitants, as well as many wild animals. They are usually eaten fresh but often crushed to extract the juice, which is fermented. Fruit can also be used to make delicious jams and jellies. In some instances, the fruits are dried in the sun and then stored. The tree could be an attractive garden subject if situated in a well-drained area, which has a summer rainfall and a mild winter climate. The moepel can be easily propagated from

Figure 8 (see color plate 70) Fruit of *Mimusops zeyheri* (moepel).

seed and, although it is considered to be rather slow-growing, will flourish if well cared for.

Freshly cut wood is pink, hence the name Red-milkwood, but turns pale brown when dry. The wood is moderately hard with a short, wavy grain and is useful as a general purpose timber. It is also suitable for woodwork, since it gives a smooth finish. It is, however, not very strong or durable. Care should be taken when working with green wood, as it contains an irritant, which causes sneezing.

Sclerocarya birrea (Marula)

Sclerocarya birrea (marula) is a large, dioecious, deciduous tree that grows wild in northern and north-eastern South Africa and parts of eastern Botswana. The marula is an erect tree that reaches heights of 15 m. Male and female flowers occur separately, and although they are usually on separate trees, this is not always so. The fruit form varies from almost spherical to oval with a tough, strong skin, which is light yellow when ripe. The fruit flesh is soft, whitish, and contains fibers. The flesh of its fruit is very juicy and aromatic, and is eaten fresh or processed, yielding quality jams, juices, and alcoholic beverages. The nut, which can be eaten fresh or roasted, has a high protein content and contains a nondrying oil that can be used for soap production as well as in food preparations. The fruit usually ripens from January and falls from the tree

Figure 9 *Sclerocarya birrea* tree (marula).

before it is ripe. The fruit is climacteric and will therefore only ripen after harvesting. There are numerous other uses for the marula tree, all of which present opportunities to rural communities (see Figure 9).

The fruit serves as an important source of ascorbic acid (vitamin C) for the rural people. The marula has a high ascorbic acid content of between 54 and 190 mg per 100 g, depending on the selection. This is much higher than the average amount of ascorbic acid found in the orange. In an area that suffers from problems of malnutrition and mineral deficiencies, demand for the fruit will always be high (see Figure 10).

When mature in February/March, the green fruit abscise, followed by a skin color change to yellow, flesh softening, and aroma development. The most popular product obtained through fermentation of the fruit is an alcoholic beer made by rural communities. Rural communities have a high regard for the marula tree, and so it is not used for firewood and other purposes. The marula has become commercialized over the past few years, although it has not yet been planted in orchards. The fruit has been used to make a creamy liqueur, which is now sold internationally. The fruit pulp is used to make jellies and jams as well as locally sold fruit juice and beers. The biggest potential for marula probably lies in the oil derived from the kernel of the nut because of its unique qualities.

Tamarindus indica (Tamarind)

The tamarind is native to tropical Africa and grows wild throughout the Sudan. It was introduced so long ago into India and other parts of southern Asia that it is often reported as being indigenous there. It has now reached the southern limit of its range in Mozambique and has also become naturalized round Durban

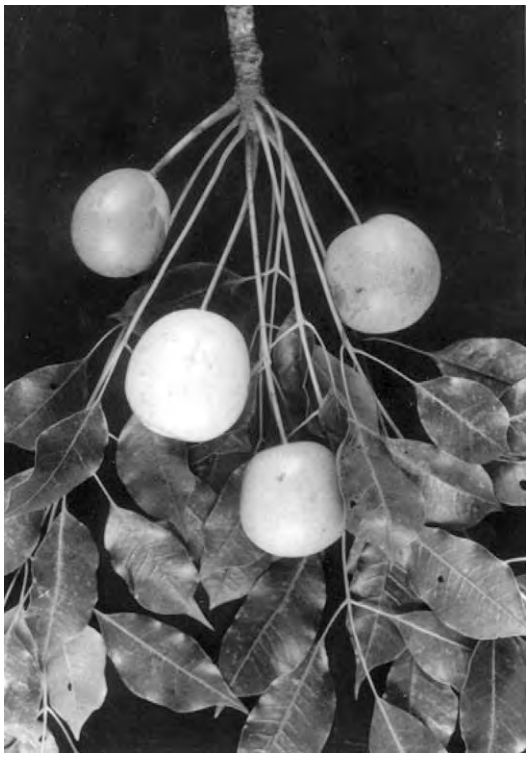

Figure 10 Fruit of *Sclerocarya birrea* (marula).

in KwaZulu-Natal (South Africa). In addition to the usefulness of its fruit, the tamarind has the advantage of being one of the best ornamental trees of the tropics.

The tree is slow-growing and, under favorable conditions, can reach a height of 24–30 m. It is highly wind-resistant with strong supple branches. The mass of bright green, fine, feathery foliage is composed of pinnate leaves, each having 10–20 pairs of oblong leaflets 12.5–25 mm long and 5–6 mm wide, which fold at night. The flowers are inconspicuous and borne in small racemes. They have five petals, of which two are reduced to bristles and are yellow with orange or red streaks. The flower buds are distinctly pink due to the outer color of the four sepals, which are shed when the flower opens.

The fruits are flattish, bean-like, irregularly curved pods, borne in abundance along the new branches, and vary from 50 to 150 mm in length and 20 to 30 mm in diameter. The pods are cinnamon brown or grayish brown externally and initially tender-skinned with green, highly acid flesh and soft whitish underdeveloped seeds. As they mature, the pods fill out, and the juicy pulp turns brown to reddish brown. The skin becomes brittle and is easily cracked, and the pulp dehydrates to a sticky paste enclosed by a few coarse strands of fiber extending lengthwise from the stalk. The seeds are glossy brown and squarish in form.

There are many food uses of the tamarind, ranging from vegetable preparations to flavoring in many types of cuisine and dishes. The most outstanding characteristic of the tamarind is, however, its astringency and its prime use as a souring agent. It is also rich in B complex vitamins. When fully ripe, fruit are also eaten fresh. The pulp is made into a variety of products and is included as an ingredient in many well-known brands of chutneys, curries, and sauces. It also serves as a flavoring in certain fish dishes. Sugared tamarind pulp is often prepared as a confection.

Over-ripe fruits also have their uses; for example, they can be used to clean and brighten silver, copper, and brass. The tamarind also has numerous medicinal uses. The pulp is officially listed in the British, American, and other pharmacopeias, and some 90 000 kg of the shelled fruits are annually imported into the USA for the drug trade. Tamarind preparations are universally recognized as refrigerants in fevers, as laxatives, and as carminatives.

Garcinia livingstonei (African Mangosteen)

Garcinia livingstonei is known as the mangostan or African mangosteen and originates in Southern Africa. It is found growing in the northern coastal region of KwaZulu-Natal, Mpumalanga and Northern Province of South Africa, Mozambique, and Swaziland. It grows northwards as far as Somalia. The mangostan is also found in the northern parts of Botswana and Zimbabwe. It occurs most commonly at low altitudes, in riverine fringes, and in open woodland.

The tree is small to medium with a height of 2–10 m. It has a very distinctive habit, being stiff with rigid branches arising at an acute angle. The tree is evergreen. The leaves are simple and usually in whorls of three, elliptic, oblong to ovate, leathery, glossy dark green above, and paler green below. The young leaves are bright red. The sweetly scented flowers are small, cream to greenish yellow and are borne on slender stalks in groups on short knobbly side spurs clustered along the older stems. The almost spherical fruit are up to 25 mm in diameter and are yellow to orange when mature. Mangostans are usually harvested from November to January (see Figure 11).

The fruit is usually eaten fresh and has a delicious acid-sweet flavor. Fruits are also used to prepare an alcoholic beverage. Potential exists for processing of this fruit into numerous products including beverages, jams, and jellies. Extracts of the leaves and flowers have been tested and show antibiotic properties.

Figure 11 (see color plate 71) Fruit of *Garcinia livingstonei* (African mangosteen).

Cordyla africana (Wild Mango)

Cordyla africana, also known as wild mango, is a large, spreading, deciduous tree up to 25 m in height. It usually occurs at low altitudes in hot areas. The leaves are alternate, compound, imparipinnate with 11–28 leaflets plus a terminal leaflet. Flowers are produced in short, dense axillary sprays, or clustered in the leaf axils. Petals are absent, and the yellow–orange stamens form a spherical head, with the flowers somewhat resembling the characteristic 'powder-puff', but the stamens are stouter and more robust. The fruits usually resemble a semifleshy drupe that is almost spherical and yellow when ripe, with the seeds embedded in the edible fleshy pulp. The fruits vary in size and are usually 40–80 mm in length and 30–60 mm in diameter. They contain ascorbic acid (vitamin C) and are much sought after by the local people, being eaten either fresh or cooked. The wood is brown and hard, and is traditionally used for building purposes or for making African drums. The tree exudes a gum resin.

Harpephyllum caffrum (Wild Plum)

Harpephyllum caffrum is a small to medium-sized evergreen tree, 6–10 m in height. The compound leaves are alternate, crowded at the end of the branches. The leaf consists of four to eight pairs of opposite leaflets plus a terminal leaflet. The leaflets are lanceolate, occasionally ovate, without hairs and dark shiny green. The flowers are small, whitish to yellowish green in small, branched sprays near the end of the branches. Male and female flowers are on separate trees. The fruit is oblong, thinly fleshy, up to 25 mm long, and red when mature.

The fruits are rather sour and relished by children. Good wines and jellies can be made from the fruit. The tree is decorative and is frequently planted as an ornamental.

Parinari curatellifolia (Mobola Plum)

Parinari curatellifolia is an indigenous fruit much sought after by local people. The mobola plum is a large evergreen tree with a spreading habit and can reach heights of 12 m. The leaves are alternate, oblong, and usually about 75 mm in length. The flowers are small and borne in compact heads. The fruits are oval to round approximately 50 mm long and 35 mm in diameter. The skin of the mobola plum is russet to grayish in color and pitted. Its pleasant-tasting yellow flesh has a variety of uses (see Figure 12).

It is eaten fresh and made into a porridge, and a syrup is made from the fruit, providing the basis for a refreshing nonalcoholic drink and also an intoxicating beverage. In traditional medicine, an extract from the bark is used for the treatment of pneumonia, abdominal complaints, and toothache. The root is used for ear and eye complaints. A bark extract is also used in tanning by local peoples. The wood of the tree is hard and resistant to borers and has been used fairly extensively for roof construction, poles, and benches in rural areas.

Vangueria infausta (Wild Medlar)

Vangueria infausta has many common names locally, including wild medlar, mmilo and mispel. It is a small, deciduous tree found in the Eastern and Northern regions of South Africa. It grows well in dry areas and fruits between January and April (see Figure 13).

The fruits are greenish, becoming light brown when ripe. They are almost spherical with a diameter of 25–35 mm (see Figure 14).

Figure 12 (see color plate 72) Leaves and mature fruit of *Parinari curatellifolia* (mobola plum).

Figure 13 Ripe fruit of *Parinari curatellifolia* (mobola plum).

Figure 14 (see color plate 73) Fruit of *Vangueria infausta* (wild maedlar).

The fruits are used by local people, especially during food scarcity, and are relatively high in ascorbic acid (vitamin C). The seed kernels are eaten as a nut and are rich in protein. The potential of the fruit would appear to be in processing, since a number of products can be made from the fruit.

Often, small gall-like growths are noticed on the leaves; these are caused by a fungus, which attacks the leaves.

See also: **Fruits of Tropical Climates**: Commercial and Dietary Importance

Further Reading

Coates Palgrave KC (1977) *Trees of Southern Africa*. Cape Town: Struik.

Morton JF (1987) *Fruits of Warm Climates*. Miami, FL: J.F. Morton.

Palmer E and Pitman N (1972) *Trees of Southern Africa*. Cape Town: Balkema.

Van Wyk BE and Gericke N (2000) *People's Plants: A Guide to Useful Plants of Southern Africa*. Pretoria: Briza.

Wehmeyer AS (1967) Edible wild fruits of the Transvaal. *Journal of Food Industries of South Africa* 19: 49–53.

Fruits of Central and South America

G T Prance, University of Reading, Reading, UK

Background

Tropical America is the source of many familiar fruits, especially avocado, cacao, guava, papaya, passion fruit, and pineapple. However, the region is also the source of a cornucopia of lesser-known fruits of great importance to local peoples. A visit to any market in Central or South America is well worthwhile to see many less familiar fruits that are traded daily. Many of the fruits grown and marketed in the region are from elsewhere in the tropics and are treated in other entries (*See* **Fruits of Tropical Climates**: Commercial and Dietary Importance; Fruits of the Sapindaceae; Fruits of the Sapotaceae; Lesser-known Fruits of Africa; Lesser-known Fruits of Asia), for example, oranges, bananas, jack fruit, carambolas and mangoes. The fruits are arranged here in alphabetical order of the botanical family to which they belong because this brings related fruits together. Some of the lesser known fruits are also described here.

Anacardiaceae (Cashew or Sumac Family)

Cashew Apple (*Anacardium occidentale*)

Cashew is more familiar for the nut, but is also the source of a pseudofruit. The nut is borne on the end of a swollen, pear-shaped, fleshy stalk known as the cashew apple or pear. This technically is not a fruit, since the nut is the actual fruit. The red- or yellow-colored pear is used as a fruit to make a provitamin A and C-rich juice, cashew wine, syrup, preserves, pickles, etc., or is eaten raw. Cashew is widely cultivated throughout Central and Southern America and elsewhere in the tropics, but is a native of the arid lands of northeastern and central Brazil.

Hogplum (*Spondias mombin*)

This is a plum-like yellow fruit of Amazonian origin that is widely cultivated throughout Central and South America, Florida, and elsewhere. It is one of the most

popular fruits in Amazonia as an icecream flavor or for its juice. The ripe fruits rot quickly, and so it is not easy to ship and is used locally where it is grown. A related species is *Spondias dulcis* or the golden-apple from the Pacific is also widely cultivated in tropical America.

Annonaceae (Soursop Family)

This plant family is the source of a wide range of large fleshy fruits, the best-known of which are the cherimoya, the soursop, and the custard apple.

Soursop (*Annona muricata*)

A large compound fruit 15–30 × 10–20 cm in size and weighing up to 3 kg. The fleshy pulp contains many seeds. The fruits are eaten raw or used for juice, icecream or desserts. Soursop originated in northern South America or the Caribbean and is now widely cultivated around the tropics and in Florida. A closely related fruit is the custard apple (*Annona reticulata*).

Sugar Apple or Sweetsop (*Annona squamosa*)

This smaller fruited Annona is also of Caribbean origin, it is cultivated around the tropics and is generally eaten raw rather than processed.

Several other species of *Annona* are used locally, for example, the wild soursop or araticum (*Annona montana*).

Cherimoya (*Annona cherimolia*)

A native of Andean valleys of Ecuador and Peru. The roundish, heart-shaped, green fruits are most popular in Peru and are said to be the tastiest of all the Annonas. The fruit thrives only at higher altitudes.

Biribá (*Rollinia mucosa*)

Although this bright yellow fruit belongs to a different genus, its fruit is similar to the Annonas and can be up to 14 × 16 cm weighing 1.3 kg, although 5-kg fruits have been reported. This species possibly originated in the upper Amazon in the region of the Peru–Brazil border and is widely cultivated in gardens throughout central America, the Caribbean, and Amazonia for its delicious pulp, which is eaten raw.

Apocynaceae (Dogbane Family)

This family contains two little-known but locally important fruits sorva (*Couma utilis*) and mangaba (*Hancornia speciosa*). Sorva is a small round berry with a delicious sweet pulp and many seeds. It is eaten raw. This and other species of *Couma* are an important source of latex used for chewing gum. Mangaba also produces a berry that is usually used in icecream and desserts, and the tree is also a source of latex. Sorva is a native of Central America and mangaba of northeastern Brazil, but both species are cultivated throughout Brazil.

Arecaceae (Palm Family)

Many of the Central and South American palms have edible fruits of considerable local importance as fruit and sources of various types of oils. Six of the most commonly used species are mentioned below.

Assai or Açai (*Euterpe oleracea*)

This multi-stemmed palm is much cultivated in eastern Amazonia for its fruit and as a source of heart-of-palm. The trees produce three to eight branches of fruit a year. The outer pulp of the small round fruits is used to make a drink or icecream flavor. It is colored dark purple and is an important supplement to the nutrition of local people. It has a wide range from Colombia and Ecuador to eastern Amazonian Brazil. Several other species of *Euterpe* are used locally, especially *E. edulis* in the upper Amazon in Peru.

Bacaba (*Oenocarpus bacaba*)

The small round fruits of this Amazonian palm are used in a similar way to assai to make a purple juice that is usually mixed with cassava (*Manihot*) flour.

Buritî (*Mauritia flexuosa*)

This large fan-leafed palm of swampy places in Amazonia is called aguaje in Peru and moriche in Venezuela. The edible fruit is used in all Amazon countries. It is ovoid and covered with small, dark-red, rhomboid scales. Under the scaly exterior is a thin yellow pulp that is often used in icecreams and desserts and is rich in provitamin A.

Patauá (*Jessenia batua*)

The pulp of the fruit of this palm of the upper Amazon in Peru and Colombia is also used to make a beverage similar to that of assai and bacaba. The kernel contains an excellent highly unsaturated oil similar to that of the olive.

Peach Palm (*Bactris gasipaes*)

This is probably the most important of all edible species of American palms. A native of the Andean foothills of Bolivia and Peru, it was spread as far as Central America and the Guianas in pre-Colombian times. The yellow- to red- or orange-colored fruits are much used by indigenous and local peoples. The fruits are boiled and the thick, starchy pulp eaten with salt or honey, and it is often compared with the chestnut

in consistency and flavor. Many other local food preparations and recipes use the peach palm. It is most nutritious, containing both proteins and carbohydrates as well as provitamin A and minerals. Other names for this palm are peijibaye (Central America), pupunha (Brazil), chontaduro (Colombia), pijuayo (Peru), and pijiguao (Venezuela).

Tucumã (*Astrocaryum aculeatum* and *A. vulgare*)

The spiny trunked tucumãs have an ellipsoid fruit with an orangish pulp that is rich in provitamin A. The fruits are usually eaten raw. Both species are native to the Amazon region where they are widely used and often found in markets.

Bombaccaceae (Kapok-tree Family)

Sapote (*Quararibea cordata*)

This is a little-known, but tasty fruit of the upper Amazon around the border of Peru and Brazil. This region is the origin of several Amazonian fruit. The local name sapote or sapota is confusing, because fruits of the Sapotaceae family have the same name. The large irregular-shaped fruit 7–15 × 5–15 cm has a persistent ring around its base formed from the sepals. The thick orange-colored pulp is eaten raw or in drinks. The bitter outer rind must be removed before eating. The sapote trees can reach heights of 45 m.

Bromeliaceae (Pineapple Family)

Pineapple (*Ananas comosus*)

The pineapple is probably the best-known and most widely cultivated fruit of South American origin. Unlike most fruit described here, it is a short herbaceous plant. The fruit is produced at the center of the circle of long, narrow, spiny leaves. The pineapple originated in southwestern Brazil and possibly eastern Paraguay, but was carried from there all over the American tropics before Columbus. Many varieties have been bred from the original wild species and spread around the world. The fruit is formed by the fusion of ovaries and bracts to the axis of the inflorescence, and is technically termed a syncarp. Pineapples are an important source of provitamin A and vitamins B, B_2, and C as well as the minerals iron, potassium, and calcium. The largest producer of this fruit is now Hawaii.

Caricaceae (Papaya Family)

Papaya (*Carica papaya*)

Another popular and well-known fruit of South American origin is the papaya. It probably originated in Central America and was spread southwards into the Andes and entered Amazonia in preColombian times. The fruit is a large berry, filled with black seeds at the center. Papaya trees are usually either male or female, and only the female trees bear fruit (hermaphrodite varieties do occur but are rare). The early travelers to South America soon distributed this delicious and nutritious fruit around the world. The trees are fast-growing and short-lived and produce fruit continuously within a year of sowing the seed. The fruit is a source of provitamin A and vitamins B, B_2, and C, and contains the enzyme papain, which helps the digestion of foods and is used as a meat tenderizer. The papaya is grown commercially around the tropics.

Caryocaraceae (Souari Family)

Piqui (*Caryocar brasiliense*)

The fruit of this low tree of the savanna (cerrado) of Central Brazil is very popular locally and is still mostly harvested from wild trees. The unusual tasting pulp of the fruit is cooked with rice or beans. It is also used to prepare a sweet liqueur. The kernel contains an oil used in cooking and for skin care. The Amazonian species (*C. villosum*) is a much larger tree, the fruit of which is used in the same way.

Chrysobalanaceae (Cocoa Plum Family)

Cocoa Plum (*Chrysobalanus icaco*)

This is a shrub or small tree of coastal sand dunes around the coasts of central and South America and the Caribbean and in similar habitats in Africa. The small plum-like fruit is either eaten raw or, in Venezuela and Colombia, bottled in syrup and is called icacos.

Several other species of this family are used locally and planted in house gardens, especially pajura (*Couepia bracteosa*), a native of the Guianas and Amazonia, marirana (*Couepia subcordata*), a native of Peruvian Amazonia, and oiti-coro (*Couepia rufa*), a native of the Atlantic coastal forests of eastern Brazil. These species of *Couepia* have a large fruit with an oily pulp with an agreeable odor.

Clusiaceae (Mangosteen Family)

Mamey Apple (*Mamea americana*)

A fruit of Central American or Caribbean origin that has now been distributed around the tropics. The large brown fruit can be up to 18 cm in diameter in some varieties. The orange-colored pulp is usually used in compotes and other desserts as well as in a

wide range of liqueurs. This fruit is grown for local use in gardens and small orchards.

Bacuri (*Platonia insignis*)

This most delicious of Amazonian fruits originated in eastern Amazonia. The attractive flowers are about 7 cm long with rose-colored petals and produce a large ovoid berry up to 15 cm in diameter. The pulp is eaten raw or used in icecreams, desserts, and fruit drinks. Some varieties are sweet and preferred for eating, and others quite acid and generally processed.

Another genus of Clusiaceae, *Rheedia*, has several species (e.g., *R. macrophylla*) that produce similar but smaller fruits than bacuri and are generally known as bacupari in Brazilian Amazonia.

Humiriaceae (Humiria Family)

Uchi (*Endlopleura uchi*)

The uchi is a large tree of eastern and central Amazonia. Fruits are generally collected from wild trees after they have fallen from the tree, and are very popular with local people, especially in Pará State, Brazil. The ellipsoid fruit has an unusual but agreeable taste and is eaten raw, or mixed with cassava flour and also used for icecreams, liqueurs, and sweet pastes. The kernel contains an oil similar in composition to olive oil that is used in cooking.

Icacinaceae (Icacina Family)

Umari (*Poraqueiba sericea*)

This low tree is a native of Central and Western Amazonia. The pulp of the plum-like fruit is highly desired by local peoples, and it is always in the markets of Iquitos, Peru and Manaus, Brazil. The pulp has a peculiar smell and flavor that some people find delicious and others repugnant, and is definitely an acquired taste.

Lauraceae (Laurel Family)

Avocado (*Persea americana*)

The avocado is of Mexican and Central American origin, but is now planted around the tropics. This well-known fruit is used as a savory in most places, but is used in sweet drinks and desserts in Brazil. The pulp of the fruit is oily and nutritious and rich in minerals and vitamins. There are three races of avocado recognized: Mexican, Guatemalan, and West Indian. There are also a wide range of cultivars of this important fruit. The name avocado is derived from its Aztec name: ahuacatl. California, Florida, Israel, and South Africa are now important commercial producers of avocados.

Leguminosae (Bean Family)

Inga (*Inga edulis* and Other Species)

Inga is a large South and Central American genus with over 300 species. Several species produce pods with seeds surrounded by a woolly, slightly juicy edible pulp. The most widespread and most cultivated species is *Inga edulis*, which is now so dispersed that its exact area of origin is uncertain. *Inga capitata* has a much thicker and woody pod and is a native of the Guianas and Eastern Amazonia. The fruits of many other species of *Inga* are used locally throughout tropical America. The pulp is eaten raw, and the seeds are spat out.

Malpighiaceae (Barbados Cherry Family)

Acerola or Barbados Cherry (*Malpighia punicifolia*)

This small shrub or tree originated in northern South America or Central America and is widely distributed throughout the Caribbean. The cherry-sized fruits are colored orangish red to scarlet, are slightly three-lobed, and have three stones inside the fleshy pulp. The fruits are one of the richest known sources of vitamin C. They are eaten raw or used in juices, icecreams, jams, and desserts. The acerola is now widely cultivated around the tropics as a source of vitamin C and in recent years has been grown commercially. It is rivaled only by camu-camu (see below) for its content of vitamin C. Consumption of only four fruits meets the daily requirement of vitamin C.

Ciruela or False-coffee (*Bunchosia glandulosa*)

This is a small shrub or tree that produces asymmetric ellipsoid, orange- to scarlet-colored fruits. It is widely cultivated in Central America, the Caribbean, and western South America but is little known elsewhere. The fruits are sometimes eaten raw but more generally made into a sweet paste similar to that of the guava.

Moraceae (Fig Family)

Uvilla (*Pourouma cecropiifolia*)

This fast-growing, short-lived tree is a native of Amazonian Peru, where the grape-like fruit is much used by indigenous peoples. More recently, it has been introduced to Brazil and throughout the Amazon region.

The small, round fruits are borne in clusters and have a tasty sweet pulp. In Brazil, this fruit is called mapati.

Myrtaceae (Myrtle Family)

Perhaps the greatest variety of delicious fruits of tropical America come from the myrtle or guava family.

Guava (*Psidium guajava*)

The guava is a small tree with tortuous branches and a smooth trunk. The fruit is a berry with a fleshy yellow or pink pulp and many seeds near the center. Many different varieties have been developed that vary in color, size and in shape from round to ovoid to pear-shaped. The fruit is often used to make guava paste, but is frequently eaten raw or used for juice or fruit salads. It is rich in provitamin A and vitamins B and C. The guava has been introduced for commercial use to many places in the world, such as Hawaii, where it has become naturalized and an invasive weed.

Several other species of *Psidium* are cultivated for the edible aromatic fruits especially the araçá (*Psidium guineensis*) and the araça-pera (*Psidium acutangulum*). The latter species has unusual quadrangular branches. Both are natives of South America and are widely cultivated.

Feijoa (*Feijoa sellowana*)

The fruits of this small tree are similar in taste to the guava but usually larger and colored green. Feijoa is planted both for its fruit and for its attractive flowers with white petals and a mass of purple stamens. This species is native in the savannas of Paraguay and southern Brazil. It is often grown in subtropical regions such as California and is being strongly marketed by New Zealand.

Araça-boi (*Eugenia stipitata*)

This little-known fruit is a native of the upper Amazon and is probably the most aromatic of all fruits in the myrtle family. The round, yellow fruits can be as large as 12 cm in diameter and weigh 750 g. This very acidic fruit is generally used for its juice mixed with sugar, rather than eaten raw.

Jabuticaba (*Myrciaria cauliflora*)

One of the most popular of all fruits in southern Brazil, the jabuticaba is a native of the Atlantic coastal rainforests. The perfumed flowers followed by the fruits are borne in huge quantities all along the trunk and branches of these low trees. The fruits are round and colored dark purple when mature. Many different varieties with fruits of different sizes have been developed. The fruit is delicious raw and is also used in jams and desserts. Local people make a tasty liqueur from the jabuticaba. Most gardens in southeastern Brazil have a tree where there is always great competition between the owners and birds to pick the fruits.

Camu-camu (*Myrciaria dubia*)

This has only recently been brought to the market and is a native of periodically flooded river banks of the northwestern Amazon. The cherry-sized fruits are borne in great quantities and are reddish when young, maturing dark purple. The fleshy pulp is extremely rich in vitamin C having about 30 times more per gram than citrus. This fruit is used principally for its juice, in icecreams and other desserts. It is very popular in Peru and is still harvested largely from wild trees along the banks of rivers and lakes.

Cayenne Cherry or Pitanga (*Eugenia uniflora*)

This shrub or small tree bears berries of 2–3 cm in diameter that are deeply divided into seven to 10 longitudinal segments. The scarlet fruits are rich in vitamin C and are eaten raw or in juices, jams, and icecreams. The pitanga is native to South America and is widely cultivated in Florida, Hawaii, Central America, and the Caribbean and in many other places around the tropics, especially in the Philippines. It is now grown commercially in a few places.

Passifloraceae (Passion Flower Family)

Passion Fruit (*Passiflora edulis*)

Passion fruits are borne on vines and are one of the best known and most commonly used fruits of tropical American origin. Since they have been widely cultivated throughout the region and elsewhere, many varieties exist. The name passion flower is derived from the cross-like structure of the stigmas emerging from the center of the flowers, and the ring of appendages of the corona that have been likened to the crown of thorns of Christ.

The round to ovoid fruits are usually yellow or purple and contain many seeds surrounded by a gelatinous and tasty pulp. This highly aromatic pulp is extremely popular for juices, icecreams, desserts, and drinks and in numerous other recipes. Passion fruits are a good source of provitamin A. They are often used medicinally as a calmant or sedative.

Several of the other over 400 species of *Passiflora* have edible fruits and are widely cultivated for example *P. nitida*. Passion fruits are cultivated commercially around the tropical and subtropical world.

Granadilla (*Passiflora quandrangularis*)

This is a more robust vine than the true passion fruit and has quandrangular stems that are winged. The

fruits are ovoid and larger, up to 30 × 15 cm in size, and can weigh 2–3 kg. This fruit is cultivated on a large scale in Colombia and the northwest of Amazonia, and is not known from the wild, so its exact place of origin is uncertain. Unlike the passion fruit, it is the pulpy mesocarp rather than the mass around the seeds that is eaten. It is often served as a savory rather than a sweet, but it is also used for juices, icecreams, and other desserts.

Rubiaceae (Coffee Family)

Genipap (*Genipa americana*)

The genipap is a tree up to 15 m tall and is now widely distributed throughout Central and South America and the Caribbean. It probably originated in northern South America but was widely distributed by native peoples in preColombian times. The berry is ovoid, 8–12 × 7–9 cm in size, and weighs 200–400 g. The many seeds are surrounded by a fleshy pulp. The unusual smell and taste make it an acquired taste, but it is nevertheless a very popular fruit with local peoples for drinks and sweets. The fruits must be ripe and soft before eating. The green fruits are the principal source of a black dye used for cloth, fibers, and body paint, and this accounts for the movement of this species over such a wide range by indigenous peoples.

Sapindaceae (Soapberry Family)

Mamoncillo or Limoncillo (*Melicoccus bijugatus*)

This 12–14-cm-high tree produces clusters of small round to ellipsoid fruits with a hard shell. The seeds are surrounded by a juicy and gelatinous pulp, which is eaten and the seeds spat out. This is a native of northern South America and widely cultivated in Florida the Caribbean and Central America. It is often used as a shade tree in cities. In Brazil, much more commonly cultivated is the pitomba (*Talisia esculenta*), which has a similar and equally delicious fruit and is a native of western Amazonia.

Sapotaceae (Sapodilla Family)

This is another plant family that has produced a large number of different edible fruits such as the caimito and the sapota.

Caimito or Abiu (*Pouteria caimito*)

This small tree produces yellow round to elongate-ellipsoid fruit up to 10 cm long. It is native to Amazonia and is cultivated and wild throughout the region. The one to four black seeds are surrounded by a delicious edible pulp. The rind contains a white sticky latex that can adhere unpleasantly to the lips when this fruit is eaten. Many varieties have been developed. Many other species of *Pouteria* are harvested from wild trees throughout Amazonia and are known as *abiurana* in Brazil. The ucuqui (*Pouteria ucuqui*) is a large tree, the fruits of which are very popular in northwestern Amazonian Brazil. The pulp is often mixed with cassava flour to make a porridge-like food.

Sapota or Sapodilla (*Manilkara sapota*)

This is the best-known and most widely cultivated fruit from the Sapotaceae. The trees reach 15–20 m in height and bear round brownish and scaly berries. The one to four flattened seeds are surrounded by a sweet soft pulp that is eaten raw. This species originated in Mexico or Central America and is now cultivated around the tropics. The trunk of this tree is also tapped for its chicle-type latex that is used in chewing gum.

Star Apple (*Chrysophyllum cainito*)

A tree up to 15 m in height with a round fruit 7–10 cm in diameter. When cut in half, the three to 10 seeds appear in a star-like formation, hence the name. These fruits have a bitter white latex in the skin and a pleasant, sweet tasting pulp. The star-apple is a native of the Greater Antilles but is widely cultivated in Central America and elsewhere.

Solanaceae (Nightshade Family)

The Solanaceae of South America have given rise to many plants of extreme economic importance such as potatoes, tomatoes, and bell and capsicum peppers. The latter three are technically fruit, but are used as vegetables and are not discussed further. Several species of *Solanum* are gradually being domesticated for their edible fruit.

Naranjilla (*Solanum quitoense*)

This fast-growing, short-lived subshrub is native to the Andes and grows at 700–2000 m altitude. The orange, round fruits have a greenish pulp around the numerous seeds. The pulp is used mainly as a vitamin-rich drink. It is most popular in Colombia and Ecuador, where it is a native but has also been introduced to Central America. Closely related to this species is the lowland cocona, known in Brazil as cubiu (*Solanum sessiliflorum*), which is a much more widely distributed domesticate than the naranjilla. The cocona berries are larger than those of the naranjilla and are used for juice, as a vegetable, and in salads.

Sterculiaceae (Cacau Family)

Cacau (*Theobroma cacao*)

This species is well known, as the seeds are the source of chocolate. The sweet, aromatic pulp around the seeds is eaten by local peoples. The exact origin of the cacau is uncertain, since it was so widespread in pre-Colombian times, but it possibly came from north-western Amazonia. The large fruits are borne on the trunk and branches of the trees. They are either ellipsoid or round. The rounder fruits are termed the criolo variety. The thick hard outer shell is yellow or red. The edible pulp surrounds 20–40 round seeds arranged in five rows.

Cupuaçú (*Theobroma grandiflorum*)

This species of the cacau genus is cultivated for the aromatic pulp of its most delicious fruits. The small trees produce large ellipsoid football-like fruit 12–25 $\times$ 10–12 cm that can weigh up to 1.5 kg. The exterior has a rust colored scaly covering. Inside, the 20–50 seeds are in five rows and adhere strongly to the thick pulp. The pulp is very popular in Brazil as an ice-cream flavor and for other sweets. The seeds have recently started being used to make cupulate, which tastes much like chocolate but has a different fatty acid profile. The tree is a native of eastern Amazonia, where the fruits are also harvested from wild trees.

The fruits of several other species of *Theobroma* are used locally for the pulp, especially the cacau sacha (*T. speciosum*) in Peru. *Theobroma bicolor* or macambo is commonly cultivated by indigenous peoples of the upper Amazon in Peru for the flat, disc-like seeds, which are roasted and eaten like nuts. This species is widely cultivated from Mexico to Brazil.

See also: **Cocoa**: Chemistry of Processing; Production, Products, and Use; **Fruits of Tropical Climates**: Commercial and Dietary Importance; Fruits of the Sapindaceae; Fruits of the Sapotaceae; Lesser-known Fruits of Africa; Lesser-known Fruits of Asia

Further Reading

Cavalcante PC (1988) *Frutas Comestíveis da Amazonia*, 4th edn. Belém, Brazil: Museu Paraense, Emílio Goeldi.

Popenoe W (1920) *Manual of Tropical and Subtropical Fruits*. New York: Macmillan.

Purseglove JW (1968) *Tropical Crops. Dicotyledons*, vols. 1 and 2. New York: Wiley.

Sampson JA (1986) *Tropical Fruits*, 2nd edn. Harlow, UK: Longman Scientific & Technical.

Silva S and Tassara H (2001) *Frutas no Brasil*. São Paulo: Nobel.

Villachica H (1996) *Frutales hortalizas promisories de la Amazonia*. Lima: Tratado de Cooperación Amazonia.

Lesser-known Fruits of Asia

M N Normah, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

Introduction

Tropical rain forests of South-east Asia in general, and that of the Indo-Malayan region in particular, are known to endow a tremendous variety of plant genetic resources. These include species of valuable and marketable timber species, medicinal plants, and fruit tree species. Many fruits originated from the Malaysian forest. However, out of this wide range of fruits, many are considered rare or lesser-known. They are generally neglected because they have not been exploited commercially and there is a lack of improved varieties. They occur in the wild or grown in scattered home gardens. These rare fruit species are also a source of useful genes for related crop species and promise to add to both nutrition and economic development.

Tropical fruits have been recognized as a rich and natural source of vitamin A and vitamin C, potassium and high dietary fiber. Fruits such as guava (*Psidium guajava*) contain 150 mg 100 g^{-1} vitamin C. Jackfruit (*Artocarpus heterophyllus*) and cempedak (*Artocarpus integer*) are highly nutritious and they are a very valuable carbohydrate source. The seed also has plenty of carbohydrates, as well as a fairly high protein content. The nutritional value of many more tropical fruits is not well documented.

Table 1 shows some examples of the lesser-known species. A more detailed outline is given below on 12 of the fruits listed in **Table 1**.

Artocarpus odoratissimus (Terap, Tarap, Marang)

The terap fruit is a syncarpium, a multiple fruit with numerous carpels fused with the fleshy receptacle. The rind is a mass of closely packed soft spines about 2 cm long with a small nob. It is green when immature and becomes fig-brown and soft when ready to eat. The numerous seeds are embedded in a firm white flesh closely packed together and lightly attached to the central core (receptacle).

Terap is unique to Brunei Darussalam and parts of Sabah and Sarawak in Borneo. Although the fruit was described from and better known in the Philippines, its distribution there is limited and it was probably introduced from Borneo. In the Philippines, the species occurs only cultivated, but in Borneo it is common in the wild state. The large fruit is esteemed

Table 1 Some of the lesser-known fruits of Asia and their uses

Species	*Common name*	*Family*	*Use*
Annona muricata L.	Soursop, durian belanda	Annonaceae	Fresh fruit, juice
Annona reticulata L.	Nona kapri, custard apple	Annonaceae	Fresh fruit
Annona squamosa L.	Nona sri kaya	Annonaceae	Fresh fruit
Artocarpus communis Forst	Sukun	Moraceae	Cooking
Artocarpus odoratissimus Blanco	Marang, terap, tarap	Moraceae	Fresh, ingredient in cakes, seeds roasted or boiled
Averrhoa bilimbi L.	Cucumber tree, belimbing	Ceraniaceae	Pickle, juice
Baccaurea griffithii Hook f.	Larah	Euphorbiaceae	Fresh fruit
Baccaurea motleyana Muell. Arg.	Rambai	Euphorbiaceae	Fresh fruits, juice, jam, organic vinegar, cooking
Baccaurea polyneura Hook f.	Jentik-jentik	Euphorbiaceae	Fresh fruit, novelty
Bouea macrophylla Griff.	Gandaria, kundang	Anacardiaceae	Fresh fruit, juice
Bouea microphylla Griff.	Remia	Anacardiaceae	Fresh fruit, pickle, cooking
Canarium odontophyllum Miq.	Dabai, kembayau	Burseraceae	Fresh fruit (softened in hot water)
Chrysophyllum cainito L.	Kameto, sauh durian	Sapotaceae	Fresh fruit
Citrus aurantifolia Swingle	Limau nipis, sour lime	Rutaceae	Juice, cooking
Citrus microcarpa Bunge	Limau kasturi	Rutaceae	Juice, cooking
Cynometra cauliflora L.	Nam-nam	Leguminosae	Fresh fruit
Dialium indum L.	Tamarind plum, keranji	Leguminosae	Fresh, sugar-coated snacks/titbits
Dimocarpus longan ssp. *malesianus* Leenh. var. *malesianus*	Isau, mata kucing	Sapindaceae	Fresh fruit, canned
Diospyros discolor Wild.	Mentega	Ebenaceae	Fresh fruit
Ficus carica L.	Ara	Moraceae	Cooking
Flacourtia rukam Zoll. and Moritzi	Indian prune, rukam	Flacourtiaceae	Fresh fruit, jam, pickle
Garcinia hombroniana Pierre	Beruas	Guttiferae	Fresh fruit
Garcinia atroviridis Griff.	Asam gelugur	Guttiferae	Cooking
Garcinia cowa Roxb.	Kandis	Guttiferae	Fresh fruit
Garcinia dulcis Kurz	Mundu	Guttiferae	Fresh fruit
Garcinia prainiana King	Cerapu	Guttiferae	Fresh fruit
Gnetum gnemon L. Bulso	Melinjau	Gnetaceae	Chips, curry
Leppisanthes rubiginosa (Roxb.) Leenh.	Kelat layu, mertajam	Sapindaceae	Fresh fruit
Mangifera caesia Jack	Binjai, kemang	Anacardiaceae	Fresh fruit, pickle, culinary
Mangifera cytherea Sonn	Kedondong	Anacardiaceae	Fresh fruit, pickle, cooking
Mangifera foetida Lour.	Bacang	Anacardiaceae	Fresh fruit, jam, pickle, cooking
Mangifera odorata Griff.	Kuini	Anacardiaceae	Fresh fruit, juice
Mangifera quadrifida Jack	Asam kumbang	Anacardiaceae	Fresh fruit
Nephelium rambutan-ake (Labill.) Leenh	Pulasan	Sapindaceae	Fresh fruit
Phyllanthus acidus L. Skeels	Star gooseberry, cermai	Euphorbiaceae	Fresh fruit, pickle
Phyllantus emblica L.	Melaka	Euphorbiaceae	Pickle
Salacca conferta Griff.	Kelubi, asam paya	Palmae	Pickle
Salacca edulis Reinw.	Salak	Palmae	Fresh fruit, pickle
Spondias pinnata Kurz	Amra	Anacardiaceae	Fresh fruit, pickle
Syzygium aqueum (Burm. F.) Alston	Jambu air	Myrtaceae	Fresh fruit
Syzygium malaccense (L.) Merr. and Perry	Malay apple, jambu bol	Myrtaceae	Fresh fruit, juice
Syzygium samarangense (Blume) Merr. and Perry	Jambu semarang	Myrtaceae	Fresh fruit
Tamarindus indica L.	Asam jawa	Leguminosae	Fresh fruit, cooking
Ziziphus mauritiana Lamk.	Bidara	Rhamnaceae	Fresh fruit

for the sweet, juicy, aromatic flesh/pulp surrounding the seeds which can be eaten fresh or used as an ingredient in cakes. The fruit is said to have a finer and more delicate flavor than that of the jackfruit. The seeds are eaten roasted or boiled; boiled seeds have good texture and a delicious nutty flavor. Young fruits are also cooked in coconut milk and eaten as a curried vegetable.

A terap may weigh 1–3 kg and be 13–16 cm long and 10–13 cm wide. The edible portion is the pulp that surrounds the small seed. The rind serves as nutritive roughage for livestock. The terap tree is massive and can attain a height of 30 m or more. Terap trees can thrive on most soils and can serve as useful plants on marginal land to prevent degradation of these areas.

Baccaurea motleyana (Rambai)

Rambai is native to Malaysia and is commonly cultivated in the lowlands. It is also native to Indonesia, for example, Borneo, Sumatra, and Java, where

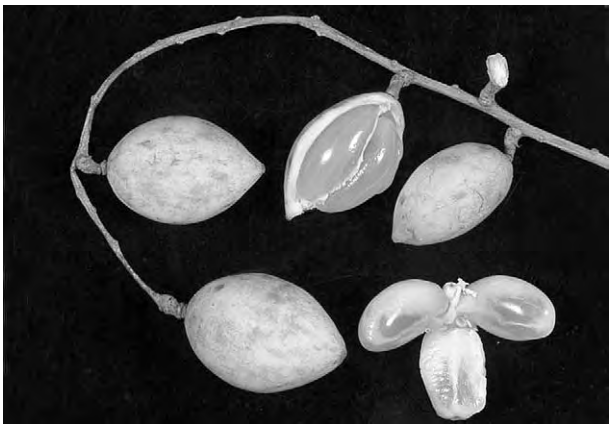

Figure 1 (see color plate 74) Fruits of *Baccaurea motleyana*.

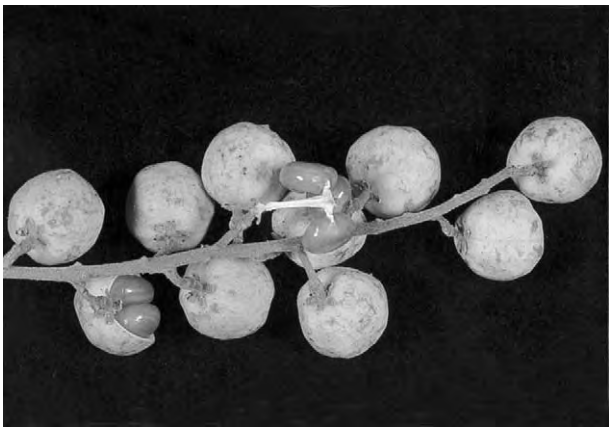

Figure 2 (see color plate 75) Fruits of *Baccaurea polyneura*.

cultivated trees are common. The trees are also found in the wild. The fruits are most pleasant and refreshing; they are eaten fresh and can be processed into juice, jams, and organic vinegar. The fruit is well suited for cooking. The fruits are borne on long stalks hanging down from branches and trunks. The edible portion of the fruits is the segmented juicy translucent pulp. Most rambai is sour but there are trees purportedly producing fruits that are less acid. The fruits are seen in the markets during the season (June to August).

There are usually two compartments in a fruit, with two segments. In a ripe fruit, the segments are enclosed in thin membrane. The juicy pulp adheres firmly to the small flat seed.

The rambai tree grows to a medium size (8–10 m) however, some can reach up to 20 m height, spreading with large green leaves 15–30 cm in length and 7.5–15 cm wide. The thin papery leaves are smooth on the upper surface and darker green than the undersurface. Flowering spikes arise from the trunk and branches but rarely on smaller branchlets. The short spikes grow as the flower buds mature and when the light yellow flowers open, the spikes have already extended and hang down. The strong sweet fragrance of the flowers attracts tiny insects for pollination. Young fruits are green and roundish with straight tapering at the apex. Mature/ripe fruits have a mauve color with less latex present (Figure 1). The rambai tree is hardy and can tolerate a wide range of environments. Fruits are produced in abundance. It would be useful to explore and look into greater use for the fruits, particularly in downstream processing.

Baccaurea polyneura (Jentik-jentik)

Baccaurea polyneura is found only in peninsular Malaysia and Sumatra. The trees are quite rarely cultivated and the fruits are sometimes seen in the market in the northern states of Peninsular Malaysia when in season (July to September). The fruit is an ovoid capsule, 2.0–2.5 cm long, and orange to slightly brown in color (Figure 2). They are juicy and have a sweet-sour taste and are eaten fresh. An interesting characteristic of the fruit is that it can be made to split by flicking and appropriately is called jentik-jentik, i.e., flick-flick.

Canarium odontophyllum (Dabai, Kembayau)

Dabai is a fruit that would seem to have potential as a widespread food source in the tropics. The trees are very heavy-bearing. The fruit is often white when immature, turning black when it is ripe. The oblong fruit, 35–50 × 25–30 mm, is ready for harvest when fruits lose the white sheen covering the fruits. A good-quality fruit has an oily yellow appearance when the black outer skin is removed. The edible portion is the mesocarp which is 2–3 mm thick and surrounds the hard triangular seed. It is a good source of protein and is usually softened in hot water before being eaten with sugar or soy sauce. The sugar or soy sauce enhances the richness of the fruit. By itself, the fruit does not have much flavor but it is popular throughout Sarawak during the season (around October to January). Dabai is also common in Brunei Darussalam, and is found in most parts of the country except in the swamps and coastal sand.

Dimorcarpus longan ssp. *malesianus* var. *malesianus*

Dimorcarpus longan ssp. *malesianus* var. *malesianus* (mata kucing, isau, sau, kakus) trees are found all over Indo-China and Malesia; the greatest variation is found in Borneo. The fruits are globular to slightly

oblong and smooth to warty. The fruit is drupaceous, and is about 2 cm in diameter. Seeds are globular with shining blackish-brown testa enveloped by a thin translucent flesh. In Peninsular Malaysia, the most common form of this taxon is the one with globose smooth fruits which turn brown when ripe. It has a very thin arilloid (flesh) and is hardly worth eating. Superior forms are found in Sarawak, all with densely thick warty fruits and thicker arilloids. The composition of the fruit is not very different from longan, but carbohydrates and energy values are much lower, whereas much higher figures are given for mineral content.

Flacourtia rukam (Rukam, Indian Prune)

Flacourtia rukam is a small tree which is widely distributed but scattered, both cultivated and wild, all over Malesia. The fruit is a round berry, 2–2.5 cm in diameter, light green to pink or purplish-green to dark red with whitish, juicy, acid pulp. There are four to seven flat irregular-shaped seeds. The ripe fruit (Figure 3) can be eaten raw; it is rubbed between the palms of the hand because bruising the flesh eliminates astringency. It is also served in fruit salad (rojak) with a spicy sauce, pickled, or sweetened with sugar to make jam or confectioneries. The young leaves are eaten raw in side dishes. Immature fruit is used to prepare traditional medicine against diarrhea and dysentery. The wood is hard and used to make household utensils such as pestles and furniture.

Garcinia atroviridis

Garcinia atroviridis (asam gelugor, som-maw won) is native to Peninsular Malaysia and Myanmar. It is a distinctive narrow tree growing to about 8–25 m in height with drooping branches, twigs, and leaves.

Figure 3 (see color plate 76) Fruits of *Flacourtia rukam*.

Figure 4 (see color plate 77) Fruits of *Garcinia atroviridis*.

Leaves are commonly about 14–26 cm long and narrow with a pointed tip and upturned edges. They are dark green and glossy, and new leaves are bright pink; all of them droop in a limp way. The female flowers have four cherry-red, slightly convex petals surrounding a thick green ovary which is capped by a dull red stigma. Fruits are solitary on the twig ends and hang down with their weight. They are about 7–10 cm in diameter and fluted from top to bottom with 12–16 deep grooves (Figure 4). The skin between the flutings is smooth and thin. The fruit is green at first, ripening to a bright orange-yellow.

Ripe fruits are acid and astringent in taste. They are sliced and sun-dried, and used in cooking. The young leaves can be eaten fresh or also used in cooking. Although the fruit is not eaten fresh, the narrow tree with drooping branches is conspicuous and would make a beautiful ornamental tree if planted along roadsides.

Garcinia hombroniana

Garcinia hombroniana (manggis hutan, beruas) resembles mangosteen. The tree forms a crown of dark green foliage, some extending 10 m, with tiers of extended branches. The fruit is round, about 5 cm in diameter, like a small mangosteen, and pinkish orange in color when ripe. Fruits are borne at the terminal axil. Between the red fruit rind and the seeds is a sour edible pulp.

The tree thrives on sandy and rocky soil. The hardiness of the tree offers the opportunity for this species to be used as rootstock for the more valuable mangosteen. The tree can also be planted as a shade tree and looks attractive as a roadside tree when planted in a row. Perhaps in the near future, when more information becomes available, the fruits, seeds, and rind may yield exciting products.

Garcinia prainiana

Garcinia prainiana (cerapu, mencupu, cerpu) is another species that is closely related to mangosteen and not less attractive compared to other known species. It is a rather small tree found over a small area of the Peninsular Malay, chiefly about the south end of the main range and also in the state of Perak, often in village orchards. The fruit is round with the top and bottom a bit flattened, and is about 3 cm in diameter. The rind is green, smooth, and shiny when immature and turns into a very attractive color of shiny orange when ripe.

Mangifera caesia syn. Mangifera kemanga

Mangifera caesia (binjai, belunu, kemang) has natural distribution in Sumatra, Borneo, and Peninsular Malaysia. They are rather rare in forests and found more frequently in periodically inundated areas and marshes. In cultivation, binjai has spread to Bali, Peninsular Thailand and, rarely, to western Java. Binjai grows into a huge tree exceeding 30 m with the crown held high above the ground. Branches that hold up the crown are dense. Leaves are 10–30 cm long and 4–11 cm wide, green, and with prominent mid-ribs and veins. At flowering, the crown is covered in a mass of tiny purple, magenta, or pink flowers held together in erect inflorescences.

Mature fruits are 10–15 cm long and 6–8 cm wide, oblong, brownish or yellow in color. The juicy, sweetish-sour binjai fruit can be eaten fresh when ripe. Mature fruits weigh 0.5–1 kg. The flesh is white, juicy, smooth and almost devoid of fiber. The wani form, which is mainly found in Bali but also in East Kalimantan, is much liked this way and fetches a high price in local markets, as the fruit is palatable, juicy, and sweet, almost fiberless, and the strong smell is completely absent. It is excellent for making creamy juices. Binjai is often used to prepare a spice based on chillies (sambal). In some areas, the flesh of ripe fruit is pickled and preserved with salt in jars, to make sambal when there is no fresh fruit. Binjai contains 74.1 mg vitamin C per 100 g edible part.

Binjai has potential as an export fruit, however, there is a need for greater research on the fruit.

Nephelium rambutan-ake syn. *Nephelium mutabile*

The *Nephelium rambutan-ake* (pulasan, kapulasan, bulala, ngoh-khonsan) tree is similar to rambutan, if a little smaller, and the leaves are narrower. Leaves have two to five pairs of leaflets which are about 8–14 cm long and 3–4 cm wide. Fruits are oblong, about 5–7 cm by 4–6 cm. The skin is very dark red when ripe and is covered with short, rather pointed knobs which are thick and fleshy but not elongated, as in the rambutan fruit. The flesh is transparent, whitish or rather yellow and the flavor is very sweet. Pulasan is native to the Malesian region and can be found in Malaysia, Thailand, Indonesia, and the Philippines. There are several different races grown, which produce varying qualities of fruit.

Syzygium malaccense

Syzygium malaccense (Malay apple, jambu bol, jambu merah) originated in South-east Asia but is more restricted to Java, Sumatra, and Peninsular Malaysia. Current distribution ranges from India through South-east Asia to the Pacific Islands. The trees are cultivated in home gardens and along driveways and paths. The tree is about 5–20 m high; the bark is gray and almost corky, and leaves are large, 15–35 cm long, and 7–17 cm wide and are bunched at the twig ends. The base is widened abruptly from leaf stalks which are short and very thick. Flowers, 5 cm across, are very striking, for they are deep pink or cerise, and are in almost stalkless bunches on the upper part of the trunk and on bare mature wood behind the leaves (not at the twig ends).

The fruit is a berry, ellipsoid, 5–8 cm in diameter, usually red with pink or white streaks, or dark red or purplish-yellow or yellow-white. There is one seed in the fruit; this is 2.5–3.5 cm in diameter and brown. There is a central hollow; the flesh is thick, rather dry and scented, and quite pleasant to eat. Eighty percent or more of the fruit is edible. It contains more than 90% water.

See also: **Fruits of Tropical Climates**: Commercial and Dietary Importance; Fruits of the Sapindaceae; Fruits of the Sapotaceae

Further Reading

Aman R (1999) *Buah-Buahan Malaysia*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Tinggal S (1992) *Brunei Darussalam Fruits in Colour*. Brunei Darussalam: Universiti Brunei Darussalam.

Verheij EWM and Coronel RE (eds) (1992) *Plant Genetic Resources of South-East Asia*, no. 2. *Edible Fruits and Nuts*. Bogor, Indonesia: PROSEA.

Yaacob O and Subhadrabandhu S (1995) *The Production of Economic Fruits in South-East Asia*. Kuala Lumpur: Oxford University Press.

FUMIGANTS

R Davis, Pest Management Consultants & Associates, Savannah, GA, USA

Background

In the approach to this discussion of fumigants, an attempt is made to define a fumigant and its uses in the preservation and protection of food, feed, seed, and fiber. This is followed by a listing of common chemicals used as fumigants and their formulations and major characteristics. Finally, there is a discussion of common applications in use today with a listing of suggested readings and resource industries to supplement the presentation.

Most importantly, it must be stressed that anyone planning to use a fumigant should secure the manufacturer's label, application manual, and a Material Safety Data Sheet (MSDS). These documents can usually be secured from the distributor. After reading these completely, all directions should be followed carefully. Local and national government entities may have additional requirements such as training and licensing.

A fumigant can be defined as a chemical that will exist as a lethal gas at ambient temperatures and pressures. As a gas, it will diffuse through air and permeate soil and most products of the farm and forest. This includes food and nonfoods (raw and processed), as well as most packaging materials, with the exception of metals. This permeation interferes with the biological processes of organisms, producing a lethal effect on most living organisms, both pest and nonpests, including man and his domestic animals. This definition excludes aerosols that are particulate suspensions and are dispersed as smokes, fogs, or mists into the space to be treated. Aerosols being particulate in form lack the ability to penetrate the commodity to be treated. Fumigants are rarely commercially available as gases. Availability is usually as cylinders of pressurized liquids that become gas as they are released in normal atmospheric pressure. Hydrogen phosphide (phosphine) is available in a solid form, a metal phosphide, and requires exposure to atmospheric moisture to activate the gas. When safely and effectively used, fumigants are efficacious and important pesticides for those involved in producing, storing, transporting, and processing many products from farms and forests.

A major reason for the wide acceptance of fumigants in the past has been their ease of application and the availability of different application methodologies. Additional modifications of application methods have usually been available to meet varying climatic conditions and storage characteristics. However, in today's approach to pest control, the use of fumigants has been limited in the numbers of fumigants available and in their widespread use. This limitation is due to the adverse effect on our environment, particularly interaction with our food products, resulting in unacceptable residues and in the use of methyl bromide with its adverse effect on our earth's protective atmospheric ozone layer. An ideal fumigant should: (1) have a low cost per effective fumigation; (2) be highly and acutely toxic to all stages of the target pest, but not unduly hazardous to man and his domestic animals; (3) be highly volatile, with a good ability to penetrate, but not be excessively sorbed on and in the commodity, its container or package; (4) have adequate warning properties for easy detection; (5) be noncorrosive, nonflammable and nonexplosive under practical conditions; (6) be of low molecular weight to aid in diffusion and penetration; (7) possess a good storage life; (8) be readily removed by aeration; (9) be noninjurious to seed germination or field and nursery stocks; (10) produce no quality changes in raw and processed foods; and (11) be easy and safe to apply.

Within this area of pest control, fumigants are used in several important areas: (1) in the treatment of wood and wooden structures for infesting insects, primarily termites and boring beetles; (2) in the preplanting treatment of soil for the control of weed seeds, soil infesting nematodes, and insects attacking tobacco, various vegetables and nursery stock; (3) in the control of rodents in and around structures and in orchards; (4) as a preventive treatment in a nation's quarantine activities; and (5) in the treatment of raw and processed stored commodities. It is the use of fumigants in areas 2 and 5 that we are principally concerned with in this presentation.

Today, with our increased ability in chemical analysis, we can now measure minute quantities of pesticides, including fumigant gases and their residues in the air, and on and in the treated commodity. As most pesticides, including all fumigants, are highly toxic to man and the environment, it is now very important that these highly toxic chemicals be only applied by trained and licensed technicians. Anyone planning to use fumigants as a control should investigate what government regulations are in place and what needs to be observed. In the rare event that there is no regulating authority at the location where the

fumigation is to be carried out, the manufacturer's directions accompanying the fumigant should be carefully followed. In the USA, the certification of pesticide applicators and registering of pesticides is vested in the Environment Protection Agency (EPA) and in individual States' Environment Departments.

Physical, Chemical, and Biological Characteristics

The primary factors that will increase or decrease the activity and distribution of fumigants are temperature, atmospheric moisture, application duration, fumigant formulation, dosage, application procedure, storage structure or other fumigation containments, pressure (vacuum), aeration, pest population characteristics and habitat, commodity, and contamination (e.g. dockage in grain).

The safe and effective use of fumigants depends, in large part, on knowledge of the fumigant's primary chemical and physical properties. Table 1 below lists fumigants that are in common use, in limited use, or, in some instances, banned from use in many countries. In the USA, only methyl bromide, hydrogen phosphide (phosphine), and sulfuryl fluoride, Vikane®, are in common use and are approved by the US Environmental Protection Agency. In this presentation, only methyl bromide and hydrogen phosphide will be discussed. Currently, sulfuryl fluoride has very limited approval for use on foodstuffs.

The temperature and atmospheric moisture of the environment and commodity to be treated are the two most important variables in achieving efficacy in fumigation. It is important in a good fumigation for the fumigant to come into equilibrium with the temperature of the commodity being treated. An increase in temperature increases the molecular activity of the fumigant, which facilitates the diffusion and penetration of the gas with decreasing sorption. The opposite effect occurs as the temperature decreases. The biological activity of the pest organism will increase with temperature changes that are within the temperature range of the life processes of the organism. This increase in activity results in a more rapid respiration rate. The acceleration in respiration increases the intake of the fumigant resulting in a quicker kill. Temperature decreases result in a lower biological activity and will lengthen the time of application. Also, it may result in a fumigation failure if the pest does not receive adequate exposure to the fumigant.

The moisture content of a commodity, such as grain, generally increases the sorptive capacity of the kernels, with a resulting loss of fumigant concentration. In general, an increase in fumigant dose or fumigation duration can compensate for sorption and penetration when grain moisture is high. Atmospheric moisture is essential for the activation of phosphine from magnesium or aluminum phosphide. If the atmospheric moisture is too low, insufficient phosphine may be produced in the prescribed time required for insect kill.

In the release of cylinder gases such as methyl bromide and sulfuryl fluoride, moisture can result in increased residues on the commodities treated. Also, with the release of methyl bromide, it may be necessary to heat the pressurized liquid, so when it expands to a gas, it will not use the air's ambient heat, resulting in freezing the moisture that condenses on the nozzle. If this freezing occurs, it may prevent the release of the fumigant. When releasing sulfuryl fluoride, it is necessary to 'shoot' the gas into a warm air stream created by a fan. If this procedure is not used, the cold gas may cause the condensation of the atmospheric moisture and create a fog that will plate out on the

Table 1 Common fumigants and their essential properties

Chemical name	*Molecular weight*[a]	*Boiling point (°C)*[a]	*Specific gravity as a vapor (air = 1)*[b]	*Flammability*[c]
Acrylonitrile	53.06	77.0	1.83	3–17
Carbon disulfide	76.13	46.3	2.63	1.25–44.0
Carbon tetrachloride	153.84	77.0	5.30	Nonflammable
Chloropicrin	164.39	112.0	5.70	Nonflammable
Ethylene dibromide	187.88	164.4	6.48	Nonflammable
Ethylene dichloride	98.97	83.0	3.35	6–16
Ethylene oxide	44.05	10.7	1.52	3–80
Hydrogen cyanide	27.03	26.0	0.93	6–41
Methyl bromide	94.95	3.6	3.27	Nonflammable
Phosphine	34.04	−87.4	1.20	1.79
Sulfuryl fluoride	102.6	−55.2	2.88[c]	Nonflammable

[a]From Weast RC, Atle MJ and Berger WH (eds) (1984) *Handbook of Chemistry and Physics*, 65th edn. Boca Raton, FL: CRC Press.
[b]Harein PK and Davis R (1992) Control of stored grain insects. In: Sauer DB (ed.) *Storage of Cereal Grains and their Products*, pp. 491–534. Minneapolis, MN: American Association of Cereal Chemists.
[c]Vikane Gas Fumigant – Structural Fumigation Manual (1998) Indianapolis, In: Dow AgroSciences.

walls of the enclosure or on the commodity being treated.

Methyl Bromide

Soil Treatments

Methyl bromide has been the fumigant of choice in the preplanting treatment of soils, because of its efficacy for a wide range of plant pests residing in the soil including fungus, nematodes, insects, and weed seeds. Additional advantages include its relatively low cost, ease of application, and short duration needed for effective control and ease of aeration. Preplanting fumigations, for the purpose of discussion, can be divided into two general approaches – area application and raised bed and row applications.

The soil preparation, regardless of the application methods, is similar. The soil must be plowed or otherwise tilled to a depth required for an effective treatment. This depth is usually 15–20 cm. The soil must be worked until it is free of clods or large lumps. Crop residues should be burned or worked into the soil after harvest to allow for decomposition. Soil moisture should be sufficient for seed germination and should be maintained for several days prior to treatment for best results. Soil temperatures should be at 15–27 °C at a depth of 15 cm. Fumigations should not be attempted if air temperatures are below 10 °C.

Area applications Area applications should be accomplished with a chisel-type applicator with a spacing of not more than 30 cm between chisels. The fumigant should be injected to a depth of 15–20 cm. The soil should be covered immediately with polyethylene film by simultaneous film laying equipment. If a soil compactor is used, the soil should be covered with film or other suitable cover within 20 min. Recommended applicator rates vary between 200 and 275 kg Ha^{-1}, depending upon the soil texture. Very heavy soils may require greater dosages; in every instance, the fumigant manufacturer's recommendation must be adhered to. Exposure times are usually 24–48 h with a 72-h aeration; however, this aeration period will vary with the crop to be planted. The aeration period will need to be at least 5–7 days if the treated area is to be used for transplanted seedlings.

Raised bed and row applications These are modifications of the area application treatment. In row applications, the polyethylene film is usually in rolls 90–130 cm wide and 305 m long. Alleyways are left between the beds to allow for working the crop. Application of the fumigant is as described above for the area treatment.

The raised bed applications are used in small fields or in greenhouses where table beds are used. Soil preparation is the same as in the area application. To prepare a raised bed, the area around the bed is trenched with the soil thrown to the outside of the bed so it can be used to bury the edges of the tarpaulin. The tarpaulin or polyethylene film will need to be raised and held off the soil to provide a space for the fumigant to evaporate and distribute. This can be achieved by using burlap bags stuffed with paper or straw, flower pots, landscape timbers, etc. On greenhouse tables, tape or glue can be used to hold and seal the tarpaulin.

The introduction of the fumigant is through polyethylene tubing anchored under the tarpaulin to an evaporation pan. One pan should be of the appropriate size to accept the required amount of fumigant and should be of plastic or metal (*do not use aluminum*). In the absence of oxygen, liquid methyl bromide reacts with aluminum to form methyl aluminum bromide, which ignites spontaneously in the presence of oxygen. The fumigant is introduced through the tubes into the pans. Do not remove the pans or tubes until aeration is complete. The fumigant may be vaporized before introduction by means of a commercial heat exchanger or by using a coil of copper tubing immersed in a vat of hot water.

Other soil fumigants that have been used or are currently available include – chloropicrin, dazomet, dibromochloropropane (DBCP), dichloropropene, and metam-sodium.

Space Treatments

Currently, only two fumigants are approved in the USA and used to treat food, feed, and seed and their processed intermediate and end products. These fumigants are methyl bromide and phosphine (hydrogen phosphide). Sulfuryl fluoride is the only other common fumigant used in the USA where it is the fumigant of choice for the control of termites. Sulfuryl fluoride is a proprietary product of Dow Chemical Company and is registered and sold as Vikane®. It is not very effective for the control of insect eggs and therefore is not believed to have much promise for general use in chambers or other types of structures. Vikane® may have some promise for greater use in insect control. It is an excellent fumigant for the control of the active stages of insects. Termites, when they hatch, need to be fed by the active stages of the species to transfer the digestive fauna to the new generation. If no active stages are available, it results in the death of the colony. Currently, Vikane® is being investigated for possible use in a broader range of pest-control problems.

Methyl bromide has both advantages and disadvantages as a fumigant. Recently, a major disadvantage has come to light. This is that the bromide ion has been determined to be a destroyer of our protective ozone layer. Other disadvantages are that it produces both inorganic and organic residues that often exceed acceptable levels. In its application, its great weight produces pools or puddles in the low parts of chambers and other types of structures. This may create hazardous situations that require special attention during exposure and aeration. When methyl bromide is used, it is essential that the following items are removed before the fumigant is introduced: (1) all exposed food, feed and medicines; (2) seeds, bulbs, and living plants unless being treated; (3) all pets (including fish); (4) furs; (5) horsehair products; (6) natural rubber goods; (7) photographic films and blueprints; (8) automobiles; (9) cinder blocks; and (10) all articles containing sulfur. It is also essential to extinguish all flames such as pilot lights, turn off electric heating elements, open all interior doors and openings to attics and crawl spaces, and open all drawers and cabinet doors when treating structures.

The advantages are numerous but not of sufficient importance to overcome all the disadvantages. Methyl bromide is an excellent fumigant for quarantine purposes. It has a low phytotoxicity and therefore can be used to control pests on live plants. It is a relatively quick fumigant, very quick if used in vacuum chambers, and is effective in the control of most living organisms. It also follows a regime of concentration times time ($C \times T$) for the control of insects under a range of temperatures. This allows the fumigator to vary either the fumigant dose or the application duration to affect the desired control. This fumigant behavior is particularly important in the control of quarantine species. There is also a great volume of $C \times T$ data currently available on a multitude of quarantine organisms.

Chamber (vault) fumigations In these fumigations, the material to be treated is loaded into the vault not exceeding 75% of the available space. All the exhaust ports are closed, circulating fans are turned on, and doors are closed. The fumigant is introduced from outside into the chamber above the material(s) to be treated or into a shallow pan in front of a fan to allow for complete vaporization and distribution.

The exhaust ports should be opened, exhaust fans turned on for aeration at the end of the exposure period, inlet ports opened and/or the door opened slightly to allow for fresh air to wash the fumigant out of the chamber. It may be advisable to close the chamber after a period of aeration and allow the load to further degas. Then, further aeration should be carried out until it is determined to be safe to remove the material(s) and distribute. In the USA, it is safe to enter the chamber and handle the products when the gas level is 19 mg m^{-3} or less.

Vacuum chamber fumigations After the materials to be fumigated have been placed in the steel vacuum chamber, a vacuum is drawn to a level of 100–380 mmHg depending on the product to be treated. The methyl bromide is released into the chamber, usually through a heat exchanger, to insure vaporization. After completion of the required exposure, the vacuum is released and the chamber air exchanged at least twice using 380 mmHg per exchange. Released air should be 19 mg m^{-3} or less to unload the chamber safely.

Railcar, truck, van, and container fumigations In the fumigation of these vehicles, two points must be observed. First, they must not be moved while under fumigation. Movement may reduce air pressure on the outside of the vehicle and suck the fumigant out of the vehicle or break the sealing and reduce the available fumigant for the control of the pests. It may also cause a possible hazardous situation. Second, these vehicles are not manufactured to be used as fumigation chambers. Many are constructed using rivets that cannot be assumed to be airtight, or have wooden floors that have cracks between the boards. Containers (ship-truck type) often have wooden floors and large rear doors. Both are very difficult to seal. These containers are also used for air shipments. This use requires that the containers have pressure vents for the equalization of the atmospheric pressure inside the container as it is moved between different altitudes. If there is any doubt as to whether or not the vehicle is sealable, it should be rejected as not being acceptable for fumigation. Aeration of these vehicle containers should be to 19 mg m^{-3} or less before unloading.

Tarpaulin fumigations The material(s) to be fumigated should be stacked on an airtight surface and covered with a tarpaulin that is of sufficient size to provide a gas distribution dome and also allow for sealing the tarpaulin to the floor. The methyl bromide should be released from outside the stack and into shallow pans in the distribution dome, one pan for each 28 m^3.

Aeration should proceed as follows: if fumigation is being conducted inside a closed structure, all doors and windows should be opened to provide cross-ventilation. The tarpaulin should be opened on one end to allow the stack to aerate for at least 30–60 min before removing the tarpaulin completely. Again, it is

important to monitor the vicinity to insure that the fumigant concentration is 19 mg m^{-3} or less, and the material(s) fumigated is safe to handle and distribute. If the fumigation is conducted outside a structure, the aeration should be conducted in steps as above with monitoring of the immediate surrounding vicinity for safety.

Fumigation of Structures

Outside Release of Fumigant

This type of fumigation is somewhat limited in use. With very large buildings and multistory buildings, this method is not usually practical. When treating the building as well as its contents, exterior sealing of the building is usually required. For treating the contents of the building, interior sealing or tarping the contents will be required. Release of the fumigant should be into shallow evaporating pans (one pan per 35.3 m^3) in front of a fan or through a heat exchanger into the building in front of fans for gas distribution. The building should be locked and guards placed outside if required or needed. Opening the doors and starting the exhaust fans followed by opening all the doors and windows to complete aeration should achieve aeration. Entrance must be prohibited until gas levels are at or below 19 mg m^{-3}.

Inside Release of Fumigant

This is the more commonly used method of fumigating buildings with methyl bromide. In this application, cylinders of gas are placed in the upright position throughout the structure after the building has been sealed. The cylinders are arranged so that, as the cylinders are opened, the fumigators can walk away from the released gas as they leave the building. As methyl bromide is a heavy gas, more gas is usually released on the upper floors of multistoried buildings. It is recommended that fans be placed throughout the building (one fan per 1400 m^3) to assure distribution. The entire operation of releasing the fumigant should be completed in less than 30 min. The fumigators, when releasing the gas, should work from a map and in pairs for safety.

In-transit Shipboard Fumigation

Methyl bromide is rarely, if ever, used for this type of fumigation. The weight of the gas makes it difficult to remove it from a ship's hold, and the duration of the treatment is the duration of the voyage, which can be up to a month or more. During long voyages, the methyl bromide is usually sorbed in the commodity, particularly if it is grain in bulk. This can result in unacceptable residues. The methodology of this fumigation treatment has been largely limited to bulk grain fumigation with hydrogen phosphide (phosphine).

Hydrogen Phosphide (Phosphine)

Phosphine is usually produced from either aluminum phosphide or magnesium phosphide in the presence of atmospheric moisture. This chemical reaction is as follows: metallic phosphide + $3H_2O$ > metallic $(OH)_3$ + phosphine gas. The manufacturer, in order to prevent fire and/or explosion, has appropriately buffered this reaction. After the gas has been evolved, the residue remaining consists of the metallic hydroxide that appears as a gray–white powder. A small amount of undecomposed metallic phosphide may be present in this powder, and the phosphide must be deactivated, as directed by the manufacturer before it can be disposed of. Care needs to be taken that this dust is not inhaled.

Phosphine is available in several formulations that facilitate the ease of using this fumigant. Not all the manufacturers or distributors usually have all the brands or formulations available. The various formulations are as follows:

1. Rounded tablets of 3 g each which produce 1 g of phosphine and usually are available in resealable aluminum flasks of 100 and 500.
2. Pellets of 0.6 g each, which produce 0.2 g of phosphine and are available in resealable aluminum flasks of 1660.
3. Sachets, bags, or blister packs of varying shapes weighting 34 g, which produce 11 g of phosphine. This product comes in steel cans, and the sachets, bags, and blister packs are often in chains or fastened together in some manner for ease of application.
4. Phosphine in a mixture of 2 parts of phosphine in 98 parts of carbon dioxide. This mixture is available by employing a firm that has a phosphine generator or by purchase of the premix in cylinders. The gas is applied through a closed system or released in a sealed structure.

Not all formulations are available or approved for use everywhere

Phosphine is very toxic to all forms of animal life, and even any exposure to man in very small amounts must be avoided. The threshold limit value has been set at 0.42 mg m^{-3} by most countries for the 8-h workday. Phosphine produces a toxic effect on the organism's utilization of oxygen. It is known that in insects, this effect is potentiated by the presence of additional carbon dioxide. In the evolving of the gas from the metallic phosphide, the most important

variables are the ambient temperature and atmospheric moisture. An increase in either or both will increase the rate of gas production. All directions provided by the manufacturer – label, material safety data sheet, and application manual – must be followed carefully.

Phosphine as a fumigant has a number of advantages in addition to its toxicity. The small amount of phosphide necessary to generate the gas required makes the fumigant very portable. Its low molecular weight, which is close to that of air, allows the gas to distribute well and rapidly penetrate packaged commodities such as processed foods and kernels of grain and seeds of legumes. Although very toxic, its mode of action on animal respiration is rather slow and allows for the use of low concentrations, particularly when there is sufficient time for long exposures.

There are also several disadvantages in the use of phosphine. It is corrosive to metals such as copper, copper alloys, and precious metals such as silver and gold. This can result in damage to electronics and electrical office and other equipment. It is important to protect such metals and equipment or remove them from the area being fumigated. Areas with low oxygen levels or where insect respiration is low from low temperatures may result in poor or incomplete control. Phosphine can be spontaneously explosive when it exceeds a concentration of 1.79% of the volume of the atmosphere. This prohibits its use in vacuum chambers or in other situations where the atmosphere may be reduced in volume. Phosphine has not been used successfully in the treatment of soil organisms, as many of them use very little oxygen. It also has been difficult to confine the gas to the soil area long enough to be treated because of the inability to seal the area adequately and because of the great activity of the gas. Although it has been recommended for the fumigation of wood and wood products, it has not been used extensively, because of site safety, difficulty in securing a satisfactory seal, and achieving the required penetration during the exposure duration.

However, the use of phosphine as a fumigant has been used successfully in the control of insects attacking a wide range of dried food commodities both raw and processed. Phosphine has been used very extensively for the treatment of grain and grain products in both static and in-transit conditions. It has also been used successfully for treating many nonfood commodities. With the exceptions noted above, phosphine can be used in most situations where methyl bromide has been used or is currently being used.

In every use of a fumigant, it is incumbent upon the fumigator to read carefully and follow the manufacturer's instructions. In countries with a pesticide regulatory authority, the fumigator should secure all the information made available. In most countries with regulatory authority, their instructions on the use of pesticides have the effect of law.

See also: **Cereals**: Bulk Storage of Grain; Handling of Grain for Storage; **Cleaning Procedures in the Factory**: Types of Disinfectant; **Insect Pests**: Insects and Related Pests; Problems Caused by Insects and Mites; **Pesticides and Herbicides**: Types of Pesticide

Further Reading

Anonymous (1995) *Farm Chemicals Handbook*. Willoughby, OH: Meister.

Anonymous (1984) *Recommendations on the Safe Use of Pesticides in Ships*. London: International Maritime Organization, United Nations.

Anonymous (1990) *Fumigation Handbook*. Washington, DC: Federal Grain Inspection Service, United States Department of Agriculture.

Akesson NB and Yates WE (1979) *Pesticide Application Equipment and Techniques. FAO Agricultural Services Bulletin 38*. Rome: Food and Agriculture Organization of the United Nations.

Bond EJ (1984) *Manual of Fumigation for Insect Control. FAO Plant Production and Protection Paper 54*. Rome: Food and Agriculture Organization of the United Nations.

Davis R (1986) Managing pests in food. In: Bennett GW and Owens JM (eds) *Advances in Urban Pest Management*, pp. 287–311. New York: Van Nostrand Reinholt.

Harein PK (1982) Chemical control for stored-grain insects. In: Christensen CM (ed.) *Storage of Cereal Grains and their Products*, 3rd edn, pp. 319–362. Minneapolis, MN: American Association of Cereal Chemists.

Harein PK and Davis R (1992) Control of stored-grain insects. In: Sauer DB (ed.) *Storage of Cereal Grains and their Products*, pp. 491–534. Minneapolis, MN: American Association of Cereal Chemists.

Jay E, Davis R and Zehner JM (1983) *In-transit Fumigation of Truck-ship Containers with Hydrogen Phosphide: A Feasibility Study. AAT-S-28/April*. New Orleans, LA: Agricultural Research Service, US Department of Agriculture.

Weast RC, Atle MJ and Berger WH (eds) (1984) *Handbook of Chemistry and Physics*, 65th edn. Boca Raton, FL: CRC Press.

FUNCTIONAL FOODS

F Rincón-León, University of Córdoba, Córdoba, Spain

Concept and Definition

It has been known for a long time that a clear relationship exists between the food we eat and our health. The current concept of functional foods has resulted from the gradual recognition that healthy diets result from eating nutritious foods and from the identification of the mechanisms by which foods modulate metabolism and health. When we eat food, our first and most basic aim is to obtain nutrients for our body and to satisfy our metabolic needs. However, some groups of foods, in addition to their nutritional properties, present other additional properties for health. These types of foods are called functional foods and may be defined as any food that has a positive impact on an individual's health, physical performance, or state of mind, in addition to its nutritious value. Other additional considerations have been proposed to define a functional food, including three additional conditions in particular:

1. it is a food (not a capsule, tablet, or powder) derived from natural ingredients;
2. it can and should be consumed as part of the daily diet;
3. it has a particular function when eaten, serving to regulate a particular body process, such as:
   - enhancement of biological defense mechanisms
   - prevention of specific diseases
   - recovery from specific diseases
   - control of physical and mental disorders
   - slowing of the aging process.

Research supporting the beneficial properties of functional foods to combat cancers, coronary heart disease (CHD), diabetes, high blood pressure, inflammation, microbial, viral and parasitic infections, psychotic diseases, spasmodic disorders, ulcers, etc. is based on chemical mechanisms using *in vitro* and cell-culture systems, various disease states in animals, and epidemiology of humans.

Also, it is important to bear in mind that since a food can be classified as functional only after its potentially beneficial effect has been proven by well-designed and properly executed intervention studies in humans, continuous research is essential in order to obtain more information about the functional properties of food through its effect on biomarkers. For example, it has been reported that plasma triacylglycerol concentrations represent a functional indicator of *n*-3 polyunsaturated fatty acid (PUFA) because *n*-3 PUFA exerts a consistent hypotriacylglycerolemic effect, which is dose-dependent and persistent, or that oligofructose and inulin alter several biomarkers, including gastrointestinal transit time, experimentally induced neoplasia, and colonic microflora, suggesting that these nondigestible carbohydrates are naturally occurring dietary constituents that may improve the quality of life and increase resistance to disease in both humans and animals.

Classification

Functional foods may be classified according to the potential medical benefits and properties of their ingredients. On this basis, the following groups may be deemed to be health enhancing: (1) dietary fiber; (2) oligosaccharides; (3) sugar alcohols; (4) amino acids, peptides, and proteins; (5) glycosides; (6) vitamins; (7) cholines; (8) lactic acid bacteria; (9) minerals; (10) polyunsatured fatty acids; and (11) others (e.g., phytochemicals and antioxidants).

Dietary Fiber

Dietary fiber is commonly defined as plant polysaccharides and lignin, which are resistant to hydrolysis by digestive enzymes in man. Plant cell-wall materials containing cellulose, hemicellulose, pectic substances, and lignin are the major components of dietary fiber. In addition, gums and mucilages are also classified as dietary fiber. Depending on its solubility in hot water, dietary fiber may be classified as water-soluble fiber and water-insoluble fiber.

WSF has been shown to reduce postprandial glucose excursions and is deemed to have hypoglycemic properties. Research results also indicate that a variety of different soluble fibers, including guar, psyllium, pectin, and oat bran, have hypocholesterolemic properties.

WIF sources, such as wheat bran and cellulose, are generally considered to provide protection against the development of colon cancer.

Oligosaccharides

The concept of a 'prebiotic' as a functional ingredient was recently defined as a nondigestible food ingredient that beneficially affects the host by selectively stimulating the growth or activity of one or a limited number of bacteria in the colon. Work with prebiotics has been limited, and only studies involving inulin-type fructans have generated sufficient data to enable a thorough

evaluation of their possible use as functional food ingredients. In order for a food ingredient of a food substance to be classified as a prebiotic, it must:

1. be neither hydrolyzed nor absorbed in the upper part of the gastrointestinal tract;
2. be a selective substrate for one or a limited number of beneficial bacteria commensal to the colon, which are stimulated to grow and/or are metabolically stimulated;
3. consequently be able to alter colonic microflora in favor of a healthier composition; and
4. induce luminal or systemic effects that are benefical to the health of the host.

Oligosaccharides are generally defined as carbohydrates from two to 20 monomeric units long. Recently, interest in oligosaccharides has increased not only because of properties that include sweetening ability and fat replacement, but also because of resistance to digestion in the upper gastrointestinal tract and fermentation in the large bowel. These oligosaccharides are the nondigestible oligosaccharides (NDOs). As regards NDOs, there is a general consensus that:

1. there is strong evidence to indicate the prebiotic effect of NDOs in humans; a prebiotic effect may be defined as a food-induced increase in numbers and/or activity predominantly of bifidobacteria and lactic acid bacteria in the large intestine of humans;
2. there is strong evidence to indicate the impact of NDOs on bowel habit;
3. there is promising evidence to indicate that consumption of inulin-type fructans may result in increased Ca absorption in man;
4. there are preliminary indications that inulin-type fructans interact with the functioning of lipid metabolism;
5. there is preliminary evidence in experimental animals of their preventive effect against colon cancer.

For instance, fructooligosaccharides (FOSs) are fructose polymers and may be classified into two main groups depending on the linkage type of the fructose units, such as inulin with β (2–1) linkages and levans with β (2–6) linkages. They occur naturally in foods such as onions, bananas, garlic, and tomatoes. FOSs, like other fermentable dietary fibers, serve as a substrate for colonic bacteria to produce short-chain fatty acids (SCFAs). FOSs help to maintain and restore the balance of healthy gut flora by creating an environment that supports beneficial bifidobacteria but does not support the growth of some pathogenic bacteria. An example is inulin, which is a linear polymer of lined fructose molecules with a glucose moiety at the end, and in crop foods consists of a mixture of various oligo- and polyfructosides with different chain lengths; its functional properties depend partly on its length distribution. Inulin is widely distributed among plants such as many grasses, onions, asparagus, artichoke, chicory, dahlia, and Jerusalem artichoke.

Sugar Alcohols and Other Alcohols

Sugar alcohols are 'polyols.' This term refers to chemical compounds containing three or more hydroxyl groups. Polyols can be divided into acyclic polyols (alditols or glycitols, which are true sugar alcohols), and cyclic polyols. An example of the first group is xylitol, and an example of the second is *myo*-inositol. In general, different functional properties have been attributed to these groups of substances, such as low caloric intake by incomplete absorption, sweeteners in diabetic diet and caries-reducing agents.

The most ubiquitous sugar alcohols are sorbitol and D-manitol, and their concentrations in some plants and plant exudates may be very high. Fruits belonging to the family Rosaceae, such as pears, apples, cherries, plums, peaches, apricots, etc., contain appreciable amounts of sorbitol. Xylitol is a naturally occurring five-carbon sugar alcohol that can be found in many fruits and berries, being most abundant in yellow plums and greengages. Some research papers have documented the noncariogenic and anticariogenic effects in addition to sweetener properties, and recent studies have indicated that xylitol may have prebiotic properties and other effects on health.

Amino Acids, Peptides, and Proteins

The bioactivities of peptides encrypted in major milk proteins are latent until released and activated by enzymatic proteolysis, e.g., during gastrointestinal digestion or food processing. The proteolytic system of lactic acid bacteria may contribute to the release of bioactive peptides, and the further degradation of these peptides by endopeptidases and exopeptidases of lactic acid bacteria may induce the release of bioactive peptides in fermented milk products, these activated peptides being potential modulators of various regulatory processes in the body.

Several studies have shown that cows' milk proteins may stimulate the growth of *Bifidobacterium* species. For instance, it has been shown that α-lactalbumin and lactoferrin are potent growth promoters for several *Bifidobacterium* species, or that trypsin-digested κ-casein presents bifidogenic activity.

Glycosides

Broccoli and other cruciferous vegetables are a good source of one type of glycosides: glucosinolates.

These are hydrolyzed by myrosinase, an enzyme found in plant cells, resulting in a variety of hydrolysis products, including isothiocyanates and indoles. Indole-3 carbinol (I3C) is currently being studied for its cancer-chemopreventive properties, particularly of the mammary gland. In addition to the induction of phase I and II detoxification reactions, I3C may reduce the risk of cancer by modulating estrogen metabolism.

Vitamins

Because there is a substantial loss of folic acid during cooking, supplementation with folic acid, especially during pregnancy and in the elderly, helps to prevent anemia. So folic acid has been added to some foods, such as enriched breads, pastas, rice and cereals.

L-Ascorbic acid (vitamin C), which is synthesized by most animals but not by man, is one of the most important water-soluble antioxidants. Several molecular and biochemical mechanisms of ascorbate-mediated immunostimulation have also been proposed.

Supplementation with vitamin E may help to prevent CHD, due to at least two mechanisms: one mechanism is by protecting blood lipoproteins against oxidation; the other is by inhibiting blood clotting, a process that is involved in the initiation of a heart attack. So both vitamin C and vitamin E are used to prevent CHD. It has been suggested that mortality resulting from to CHD is inversely related to a cumulative antioxidant index, defined in terms of concentrations in plasma, as

$$[\text{vitamin E}] \times [\text{vitamin C}] \times [\beta - \text{carotene}] \times [\text{selenium}]/[\text{cholesterol}],$$

because α-tocopherol (vitamin E), L-ascorbic acid (vitamin C), and β-carotene (which have vitamin A activity) are excellent antioxidants.

Vitamin D (1,25-dihydroxyvitamin D_3) has recently been identified as an immunoregulatory hormone, serving as an immunostimulatory agent of nonspecific immunity and exerting both stimulatory and inhibitory effects on specific immune responses.

Cholines

Lecithin (phosphatidylcholine) is a major component of crude soybean oil. However, raw lecithins are complex mixtures of lipids, namely, phosphatidylcholine, phosphatidylethanolamine, and phosphatidylinositol as main components. Lecithin plays an important role in normal metabolism, for example, forming part of membrane phospholipids; some metabolites are messengers in cells; choline derived from lecithin plays a role in brain cells in the resynthesis of new phosphatidylcholine molecules needed for cells that envelop the axons of neurones; and choline apparently has a protective role, allowing cells to die rather than to turn into malignant cells (cancers) when they undergo mutations.

Lactic Acid Bacteria

In contrast to prebiotics, which represent a new concept, probiotic activities have been the object of numerous hypotheses and have been known since the turn of the century, when it was postulated that lactic acid bacteria provided health and longevity benefits. It is estimated that over 400 species of bacteria, classified in two broad categories, inhabit the human gastrointestinal tract. The categories are: those considered to be benefical (e.g., *Bifidobacterium* and *Lactobacillus*) and those considered to be detrimental (e.g., *Enteobacteriaceae* and *Clostridium* spp.). The bacteria in the first category are used traditionally in food fermentation and are often referred to as probiotics. Probiotics are viable microbial dietary supplements that beneficially affect the host through their effects in the intestinal tract. They are widely used to prepare fermented dairy products such as yogurt or freeze-dried cultures, and may help to improve human health.

Beneficial effects on human health by specific probiotic microorganisms such as prevention of gastrointestinal tract infections, immune stimulation, and balancing of the intestinal microflora have been established in numerous clinical trials.

The hypocholesterolemic effect of fermented milk products was discovered more than 30 years ago during studies conducted in Masai tribesmen in Africa. The Masai have low levels of serum cholesterol and clinical CHD despite a high-meat diet, but they consume daily 4–5 l of fermented whole milk. More evidence supports the role of probiotics in cancer-risk reduction, particularly colon cancer. This observation may be due to the fact that lactic acid cultures can alter the activity of fecal enzymes (e.g., β-glucuronidase, azoreductase, nitroreductase) that are thought to play a role in the development of colon cancer. Some evidence also suggests that probiotics play a role in reducing the risk of rotavirus-induced diarrhea. The increase in the growth of *Bifidobacterium* sp. is also accompanied by the production of nitrogen derivatives, such as ammonia, indole, phenol, and skatole, and by the elimination of carcinogenic substances during fermentation.

Prebiotics and probiotics must be seen as complementary in many aspects, because often a combination of both may prove to be an important clinical method of treating intestinal disorders. A mixture of probiotics and prebiotics may improve the survival

and implantation of live microbial dietary supplements in the gastrointestinal tract. It has been suggested that such an approach may produce the following nutritional benefits:

- improved survival of live bacteria in supplementary or food products and consequently prolonged shelf-life;
- an increased number of ingested bacteria reaching the colon in viable form;
- stimulation in the colon of the growth and implantation of both exogenous and endogenous bacteria; and
- activation of the metabolism of these bacteria (only metabolically active bacteria can promote health).

A new symbiotic concept, or combination of both pre- and probiotics, has been proposed to obtain the healthy benefits of both pre- and probiotic and in addition other new benefits, but this is an attractive concept that needs to be tested in clinical conditions.

Minerals

Minerals play an important role in maintaining muscle and nerve function, regulating water balance and metabolism, mineralization of the skeleton, and transformation of energy. Selenium functions as an antioxidant and is a component of another antioxidant, glutathione peroxidase. It also helps to maintain adequate levels of coenzyme Q in heart muscle. Similar to selenium, several other metalloenzymes are directly involved in immunity and antioxidant defenses, including copper/zinc superoxide dismutase and iron-catalase. Zinc is essential for normal growth, appetite, and immune function. Recommendations for an increased daily dietary intake of calcium have been made in many countries, so the dairy industry has developed different milk products enriched in calcium, up to about 25% on a product basis. It has recently been shown that food products enriched with a milk calcium extract may increase the rate of bone-mass accumulation in prepubertal girls by 1.5% per year.

Polyunsatured Fatty Acids

Dietary polyunsaturated fatty acids (PUFAs) are classified into two families: ω-6, or *n*-6, and ω-3, or *n*-3. Each has different functions in the regulation of physiological processes in the human body. *n*-6 PUFA, but not arachidonic acid, is found primarily in plants; *n*-3 family linolenic acid (18:3*n*-3) is similarly converted into eicosapentaenoic acid (EPA) (20:5*n*-3) and docosahexaenoic acid (DHA) (22:6*n*-3); linolenic acid is found in plant sources, whereas EPA and DHA occur in marine oils.

Research has shown that polyunsaturated fatty acids can play a key role in reducing the risk of cardiovascular diseases by lowering serum triglycerides, and reducing the development of thrombosis and arteriosclerosis.

There is good scientific evidence to indicate that the composition of dietary fatty acid is involved in the etiology of many diseases. Increasing the supply of *n*-3 PUFA may reduce the risk of CHD. In terms of disease status, epidemiological studies have demonstrated that the incidence of CHD is inversely associated with consumption of *n*-3 PUFA. However the cardioprotective effect of fish consumption has been observed in some prospective investigations but not in others. Negative results could be explained by the fact that although *n*-3 fatty acids have been shown to lower triglycerides by 25–30%, they do not reduce low-density lipoprotein (LDL) cholesterol. In fact, a recent review of 72 placebo-controlled human trials showed than *n*-3 fatty acids increased LDL cholesterol. PUFA are derived primarily from fish oils. Therefore, increasing fish intake is the most obvious way of increasing *n*-3 PUFA intake. However, a large percentage (up to 65%) of the population do not eat fish. Thus, there is a need for alternative sources of *n*-3 PUFA, such as functional foods, whose unique fatty acid composition could fortify staple foods, thereby promoting optimal levels of *n*-3 PUFA intake.

Furthermore, an anticarcinogenic fatty acid known as conjugated linoleic acid (CLA) was first isolated from grilled beef in 1987. CLA is a mixture of positional and geometric isomers of linoleic (18:2 *n*-6) in which the double bonds are conjugated instead of existing in the typical methylene interrupted configuration. Nine different isomers of CLA have been reported as occurring naturally in food. Over the past two decades, CLA has been shown to be effective in suppressing forestomach tumors in mice, aberrant colonic crypt foci in rats, and mammary carcinogenesis in rats. More recently, CLA has been investigated for its ability to change body composition, suggesting its role as a weight-reduction agent.

Other (e.g., Phytochemicals and Antioxidants)

Recent epidemiological studies have indicated that a high intake of fruit and vegetables is associated with a reduced risk of a number of chronic diseases. This is attributed to the fact that these foods may provide an optimal mix of phytochemicals, such as natural antioxidants, fibers and other bioactive compounds.

Experts from the US Food Administration have defined phytochemicals as substances found in edible fruits and vegetables that may be ingested by humans daily in gram quantities and that exhibit a potential for modulating human metabolism in a manner

favorable for cancer prevention. Most recently, other beneficial properties, such as cardiovascular-protective properties, have been considered.

Phytochemicals could provide health benefits, such as: (1) substrates for biochemical reactions; (2) cofactors of enzymatic reactions; (3) inhibitors of enzymatic reactions; (4) absorbents/sequestrants that bind to and eliminate undesirable constituents in the intestine; (5) ligands that agonize or antagonize cell surface or intracellular receptors; (6) scavengers of reactive or toxic chemicals; (7) compounds that enhance the absorption and or stability of essential nutrients; (8) selective growth factors for beneficial gastrointestinal bacteria; (9) fermentation substrates for beneficial oral, gastric or intestinal bacteria; and (10) selective inhibitors of deleterious intestinal bacteria. Nutraceuticals have been defined as bioactive phytochemical compounds that have disease-preventing, health-promoting and/or medicinal properties.

Free radicals are unstable molecules that have an unpaired electron, which makes them highly energized and reactive, and which can damage DNA and induce cancer. Food ingredients capable of quenching or stabilizing free radicals are referred to as antioxidant ingredients. There are many antioxidants in foods, but the primary antioxidants that play this very important role in the human body are vitamin C, vitamin E, and carotenoids such as β-carotene, lycopene, lutein, or zeaxanthin. In addition to preventing cancer, some antioxidants such as vitamin E may help to prevent heart disease by preventing oxygen in the blood from combining with LDL cholesterol (oxidized LDL cholesterol is one of the contributors to plaque in the coronary arteries).

Several crops, including grains, oilseeds and horticultural crops, are rich sources of phytochemicals. Cereals such as wheat, barley, and oats, oilseeds such as canola, flaxseed, and mustard, and horticultural crops such as grapes are good sources of phytochemicals. Anthocyanins are flavonoids (pigments) present in fruits such as blueberries and also have antioxidant properties. Isoflavones are a group of phytochemicals that can reduce the risk of heart disease and several cancers, including breast, lung, and prostate cancer. Polyphenols, which are said to be present in large quantities in tea, coffee, red grapes, kidney beans, prunes and red wine, display anticarcinogenic, antioxidant, antibacterial, and antiviral action. Most clinical studies have been carried out on tea; catechins are the predominant and most significant of all tea polyphenols, and the four major green tea catechins are epigallocatechin-3-gallate, epigallocatechin, epicatechin-3-gallate, and epicatechin. Recent studies suggest that benefits from tea consumption are restricted to high intakes in high-risk populations. This hypothesis supports the recent finding that the consumption of five or more cups of green tea per day was associated with a reduced recurrence of stages I and II of breast cancer in Japanese women. Saponins, also known as triterpenoids, are classified as polyphenols, and are abundant in soyabeans more than in any other legume, including various beans and chickpeas. Saponins have a modest hypocholesterolemic effect. Citrus fruits contain particularly high contents of one class of phytochemical known as limonoids, and over the last decade, evidence has been gathered to support the cancer-preventive effect of limonene. Based on these observations, and because it has little or no toxicity in humans, limonene has been suggested as a good candidate for human clinical chemoprevention trial evaluation. Phytosterols (plant sterols), structurally related to cholesterol, reduce blood cholesterol and lower LDL cholesterol levels, possibly by either inhibiting cholesterol absorption by displacing cholesterol from intestinal micelles or altering the activity of enzymes involved in cholesterol metabolism and excretion. The most abundant plant sterol is β-sitosterol (sitosterol), whereas sitostanol, the saturated derivative of sitosterol, occurs at negligible levels in plant lipids.

See also: **Amino Acids**: Properties and Occurrence; **Ascorbic Acid**: Properties and Determination; ***Bifidobacteria* in Foods**; **Cancer**: Diet in Cancer Prevention; **Carbohydrates**: Digestion, Absorption, and Metabolism; **Cholecalciferol**: Properties and Determination; **Choline**: Properties and Determination; **Coronary Heart Disease**: Prevention; **Dietary Fiber**: Properties and Sources; **Fatty Acids**: *Trans*-fatty Acids: Health Effects; **Lactic Acid Bacteria**; **Probiotics**; **Tocopherols**: Properties and Determination

Further Reading

Brassart D and Schiffrin E (2000) Pre- and probiotics. In: Schmidl MK and Labuza TP (eds) *Essentials of Functional Foods*, pp. 205–216. Guithersburg, MD: Aspen.

Duthie GG and Brown KM (1994) Reducing the risk of cardiovascular disease. In: Goldberg I (ed.) *Functional Foods. Designer Foods, Pharmafoods, Nutraceuticals*, pp. 19–38. New York: Chapman & Hall.

Hardy G (2000) Nutraceuticals and functional foods – Introduction and meaning. *Nutrition* 16: 688–689.

Hasler CM (1998) Functional foods: their role in disease prevention and health promotion. *Food Technology* 52: 63–70.

Kwak NS and Jukes DJ (2001a) Functional foods – part 1 – The development of a regulatory concept. *Food Control* 12: 99–107.

Kwak NS and Jukes DJ (2001b) Functional foods – part 2 – The impact on current regulatory terminology. *Food Control* 12: 109–117.

Pszczola DE (1998) ABCs of nutraceutical ingredients. *Food Technology* 52: 30–37.

Salminen S, Ouwehand A, Benno Y and Lee YK (1999) Probiotics: how should they be defined. *Trends in Food Science & Technology* 10: 107–110.

Sawicki J and Slusarki B (1997) Lipids. In: Sikorski ZE (ed.) *Chemical and Functional Properties of Food Components*, pp. 99–118. Lancaster, PA: Technomic.

Sloan AE (1999) The new market: food for the not-so-healthy. *Food Technology* 53: 54–60.

Zubillaga M, Weill R, Postaire E *et al.* (2001) Effect of probiotics and functional foods and their use in different diseases. *Nutrition Research* 21: 569–579.

Fungi *See* **Mushrooms and Truffles**: Classification and Morphology; Use of Wild Mushrooms; **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage; Fungi in Food – An Overview; Molds in Spoilage; Yeasts in Spoilage

FUNGICIDES

N A Smart, Harpenden, Hertfordshire, UK

Introduction

A fungicide is any substance, preparation, or organism intended for destroying or controlling any fungal species during production, storage, or distribution of an agricultural commodity or food, in ornamental plants, or in situations endangering the health of animals or humans. Numerous organic chemicals, as well as some inorganic ones, are formulated for use as fungicides in agriculture and food storage. Fungicides help to increase efficiency of production, by preventing or reducing damage to the growing crop or stored commodity, and to improve the quality of the product, by preventing rotting and a damaged appearance. These chemicals are closely scrutinized for any toxic hazard before they are marketed. Plant breeders can produce cultivars of plants, fruits, and vegetables that are resistant to attack by specific diseases so that a fungicide may not be required. The damage done by fungi (pathogens) depends on the host plant–pathogen–fungicide interaction. Different mechanisms of fungitoxicity can operate in different hosts and environments so that standard measures may not work in every situation. More information on mechanisms that influence host–pathogen interactions, predictive tools for use in the field, and wider options of treatment are desirable to obviate development of resistance to fungicides.

Types of Compound

Fungicides used for preharvest control of fungal and bacterial infections of commercial crops at the beginning of the 20th century were simple inorganic compounds. Organic chemicals began to be developed for use as general agricultural fungicides during the 1950s and many organic compounds, which often have a specific biological action, have been marketed since then. The structural types of chemicals used as fungicides have changed as newer molecules have been found to give more effective control than the earlier ones, and better technologies have been developed for manufacture, making them more cost-effective. However, a significant proportion of inorganic (probably over a third of the total) and older organic compounds (probably about a quarter of the total) are used worldwide.

Fungicides used in agriculture and food storage may be systemic or nonsystemic, eradicant, curative, or protectant, or a combination of these effects. Eradicant fungicides are only active against the later, visible stages of the fungal life cycle. Curative fungicides are active against the early but postpenetrative effects of fungal action. Protectant fungicides prevent infection and are active against spore germination, germination development, and growth.

Although about one-third of organic fungicides can exist in more than one stereochemical form, i.e., can be chiral, only about 1–2% of the products marketed are sold as single isomers. Most of such formulations are mixtures of isomers.

This entry outlines the subject at the end of 1999 and some more recent material has also been included. International common names are used. Fungicides are grouped mainly in terms of their biochemical mechanism of fungicidal action (in as far as this is known):

1. General cell toxicants
2. Inhibition of biosynthesis of microtubules

3. Inhibition of methionine biosynthesis
4. Inhibition of ribosomal RNA synthesis
5. Inhibition of sterol biosynthesis
6. Inhibition of melanin biosynthesis
7. Glutathione system targeting
8. Uncoupling oxidative phosphorylation
9. Inhibition of respiratory chain in the mitochondria
10. Inhibition of mitochondrial electron transport
11. Inhibition of fungal biosynthesis of nucleic acids
12. Inhibition of glycerophospholipid biosynthesis
13. Inhibition of protein synthesis
14. Acquired systemic resistance
15. Biological control
16. Unknown or very uncertain mechanisms of action

This list is not exhaustive and it should be realized that a fungicide may exhibit control by more than one mechanism. Grouping in terms of chemical structure so as to cover most known agricultural fungicides can be arranged within this framework. The chemical structures of some representative organic fungicides within each type are given in Table 1; again, the lists are not exhaustive. Only a few of the many possible compounds of a chemical structural group listed will show sufficient fungicidal activity and be of practical use in the field.

A significant number of fungicides are now generic and off-patent. Such chemicals can be manufactured by any company provided that the manufacturing process satisfies the legal requirements of the country in which it is situated. Health and safety regulations of any country in which the fungicide is then used will, of course, still need to be complied with (see below).

Use

Pathogens Controlled

Fungicides of various types have been successful in controlling most major diseases in growing crops intended for market. The commercially important diseases are (in an order of relative importance): leaf spot diseases, late blight/downy mildew, rice diseases, fruit rots, cereal seed-borne diseases, powdery mildews, cereal stem diseases, rusts, and smuts. Diseases in particular crops which it is commercially essential to control with fungicides are given in Table 2.

Fungicides are also used to control many postharvest diseases that cause rapid and extensive breakdown of high-moisture commodities and pose serious problems. Some infections, such as *Monilinia fructicolis* in stone fruits, will have started in the field. Storage rots in potatoes may be controlled with suitable fungicide treatment in store, e.g., 2-butylamine.

Some crop diseases caused by fungi are still difficult to control with chemicals, e.g., eyespot of cereals can only be partially controlled and a fungicide has only very recently been developed that is effective against take-all of wheat.

Mode of Action

Fungicidal action is not simply interaction between fungicide and fungus; it involves the host (plant) and its environment. Hence plant variety, temperature, humidity, cultivation regime, and several other factors play a part in the control of fungal attack. For effective action, careful timing of fungicide applications in relation to the stages of development of infection is essential. The persistence of a fungicide on the surface or within the plant is an important factor in deciding whether to use that chemical.

Many older-type fungicides, e.g., copper and dithiocarbamates, are surface-acting and protectant in action; they are fungistatic and do not prevent infection. They are also multisite in action.

The highly effective, modern, systemic fungicides are often curative and protectant as well as eradicant, i.e., they are fungitoxic. Systemic fungicides are often selective in the range of pathogens they effectively control: they act on a specific receptor or process, e.g., are sterol C14-demethylation inhibitors. However, there are some uncertainties about the mechanism of action of some fungicides and different mechanisms can operate in different hosts/environments, e.g., carpropamid has two different modes of action.

The build-up of resistance to some groups of fungicides, further discussed in the next section, is considered to be related to single-site attack on the mechanism of fungal cells. Detailed discussion of the biochemistry of the mechanism of action of individual fungicides and of resistance of diseases to control by fungicides is beyond the scope of this article and may be followed up elsewhere (see Further Reading).

The systemic activity of fungicides takes place almost entirely through the xylem. Fungicides are usually applied to the foliage/leaves of plants, although some are effective through root uptake. Attempts have been made to design fungicides that will translocate through the phloem but without success. Systemic fungicides are generally longer-lasting than the older surface-acting chemicals.

The mode of action of the groups of fungicides listed above is now discussed in outline:

1. General cell toxicants. Numerous multiple-site, nonselective older chemicals are widely used. This is particularly the case in some less well-developed countries as these chemicals cost less than more recent selective and systemic fungicides. Their

Table 1 Grouping of fungicides in terms of their biochemical mechanism of fungicidal action and chemical structure

Mechanism of action	*Chemical structural group*	*Example(s) with chemical formula*	
Multisite nonselective		Diverse	Copper (Cu^{2+}) Sulfur (S)
			Dichlofluanid
		Alkylenebisdithiocarbamates	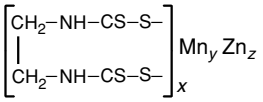 where x is large and $= y + z$ (polymeric) Mancozeb
Inhibition of biosynthesis of microtubules		Benzimidazoles	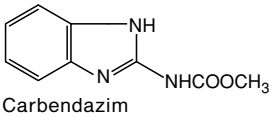 Carbendazim
		N-arylcarbamates	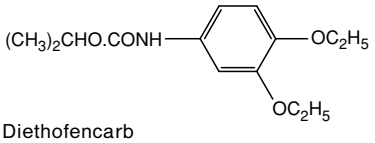 Diethofencarb
Inhibition of methionine biosynthesis		Anilinopyrimidines	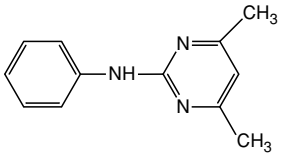 Pyrimethanil
Inhibition of ribosomal RNA synthesis		Acylalanines	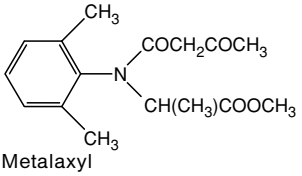 Metalaxyl
Sterol biosynthesis inhibition 14α-demethylation inhibitors		Triazoles	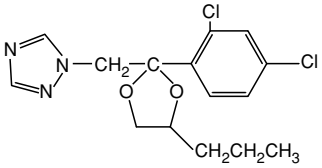 Propiconazole (4 stereoisomers)
		Imidazoles	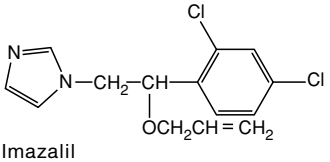 Imazalil

Continued

Table 1 Continued

Mechanism of action	*Chemical structural group*	*Example(s) with chemical formula*
	Some 6-membered heterocyclics	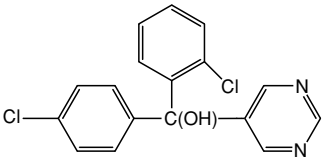 Fenarimol
Δ^{14}-reductase and $\Delta^{8}\rightarrow\Delta^{7}$ isomerase inhibitors	Morpholines	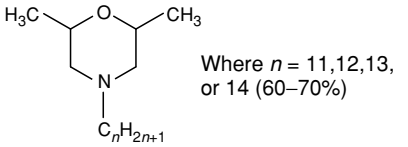 Where n = 11,12,13, or 14 (60–70%) Tridemorph
	Piperidines	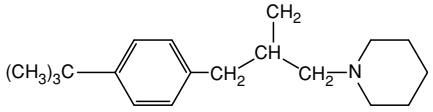 Fenpropidin
Melanin biosynthesis inhibition		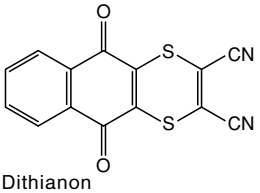 Dithianon
Targeting the glutathione system	Dicarboximides	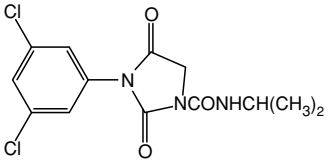 Iprodione
Uncoupling oxidative phosphorylation		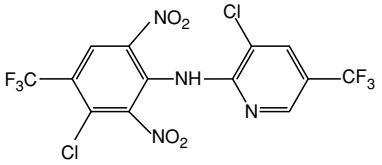 Fluazinam
Inhibition of respiratory chain in mitochondria	*E*-β-methoxyacrylates	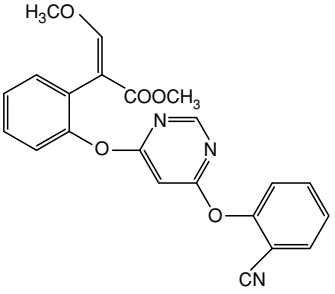 Azoxystrobin

Continued

Table 1 Continued

Mechanism of action	*Chemical structural group*	*Example(s) with chemical formula*
Inhibition of mitochondrial electron transport	Hydroxypyrimidines	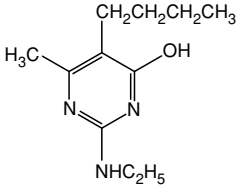 Ethirimol
Inhibition of fungal biosynthesis of nucleic acids	Carboxanilides (oxazolidinones)	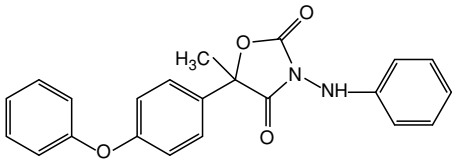 Famoxadone
Inhibition of glycerophospholipid bisynthesis	Diverse	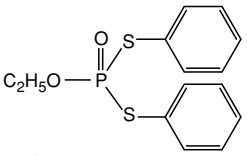 Edifenphos
Inhibition of protein synthesis	Microbiological origin	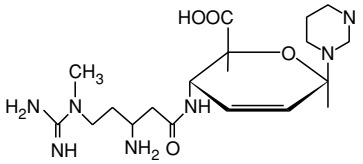 Blasticidin S
Acquired systemic resistance		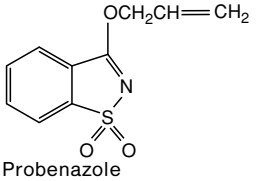 Probenazole
Biological control Unknown or very uncertain mechanisms of action	Organophosphorus compounds	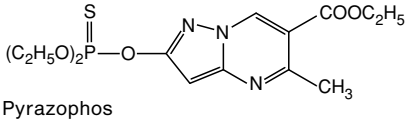 Pyrazophos
	Aromatic hydrocarbons	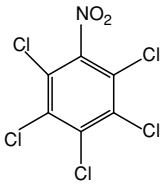 Quintozene

Table 2 Important diseases controlled by fungicides in the field

Crop	*Important diseases against which fungicides are used*
Barley	Loose smut (*Ustilago nuda*), stripe (*Puccinia striiformis*)
Rice	Rice blast (*Magnaporthe grisea*, anam. *Piricularia oryzae*), sheath blast (*Rhizoctonia solani*, anam. *Thanatephorus cucumeris*)
Wheat	Rusts (*Puccinia* spp.), bunts (*Tilletia* spp.), loose smut (*Ustilago* spp.)
Potato	Late blight (*Phytophthora infestans*)
Cocoa	Black pod (*Phytophthora palmivora*)
Coffee	Leaf rust (*Hemileia vastatrix*), berry disease (*Colletotrichum coffeanum*)
Banana	Black and brown (yellow) Sigatoka (*Mycosphaerella* spp.), Panama disease (*Fusarium oxysporum*)
Grapes	Powdery mildew (*Uncinula necator*), downy mildew (*Plasmopora viticola*), *Botrytis cinerea*
Top fruit	Apple scab (*Ventura inaequalis*), pear scab (*V. pirina*), powdery mildew (*Podesphaera leucotricha*)
Cotton	Macrosporiosis (*Alternaria macrospora*), seedling disease complex
Vegetables	Powdery mildew (*Sphaerotheca fuliginea*), early blight (*Verticillium* spp.),damping off (*Pythium* spp.)
Sugar beet	Leaf spot (*Cerospora* spp.)

preventive action is mainly due to inhibition of spore germination related to the blocking of several thiol-containing enzymes involved in spore respiration. Copper (Cu^{2+}) fungicides are among the oldest known fungicides (e.g., as Bordeaux mixture, with lime) and are simply protectant. Sulfur and lime sulfur are also protectant. The alkylenebisdithiocarbamates (EBDCs), e.g., mancozeb, became commercially viable in the 1960s and are essentially surface-acting protectant fungicides. However, in recent years they have been closely scrutinized on toxicological grounds and their use is now restricted in some countries. Dichlofluanid, tolylfluanid, and captan are likewise contact multiple-site fungicides and have protectant and curative action. Orthophenylphenol is used as a dip and on wrapping paper to control storage diseases in citrus fruit.

2. Inhibition of biosynthesis of microtubules. Several benzimidazoles, e.g., carbendazim, benomyl, inhibit biosynthesis through forming complexes with β-tubulin leading to increased acid production and polymerization of tubulin. Some *N*-arylcarbamates, e.g., diethofencarb, are also multiplication inhibitors. These are systemic fungicides having a wide range of action and are both protectant and curative.
3. Inhibition of methionine biosynthesis. Anilinopyrimidines, e.g., pyrimethanil, cyprodinil, can inhibit methionine synthesis. These chemicals are systemic and both protectant and eradicant. They act primarily on the process of penetration of fungi into their host so that the excretion of hydrolytic enzymes is blocked.
4. Inhibition of ribosomal RNA synthesis. Several acylanilines, e.g. metalaxyl, oxadixyl, and related compounds, are systemic, eradicant, and protectant fungicides which are selective and good for controlling *Oomycetes* and *Phytophthora* spp. diseases. (Carboxanilides, although of an acylalanine structure, have a different mode of action from the above and are considered separately below.)
5. Sterol biosynthesis inhibitors (SBIs). There are two groups:
   a. C14α-demethylation inhibitors (DMIs) which interfere with cytochrome P-450_{DM}. The chemical structural types present in these compounds are: triazoles, e.g., bitertanol, propiconazole; imidazoles, e.g., imazalil, prochloraz; and some 6-membered heterocyclics, e.g., fenarimol, triforine. They are systemic, being translocated in the xylem, and have protectant and curative action; some have an eradicant action. They are effective against *Ascomycetes*, *Basidiomycetes*, and *Fungi imperfecti* spp., but are not effective against *Oomycetes* spp.
   b. Δ^{14}-reductase and $\Delta^8 \rightarrow \Delta^7$ isomerase inhibitors. Several morpholines, e.g., tridemorph, and piperidines, e.g., fenpropidin, are systemic fungicides of this type and have eradicant action while affording some protection. The use of tridemorph has been revoked in the UK.
6. Melanin biosynthesis inhibitors (MBIs). These fungicides, e.g., carpropanid or dithianon, are protectant and partly eradicant in action.
7. Glutathione system targeting. Some dicarboximides, such as iprodione, which is a contact fungicide with preventive and eradicant action, and vinclozolin, which is again nonsystemic with protectant action, are effective in controlling *Botrytis* spp. and *Sclerotonia* spp.
8. Uncoupling of oxidative phosphorylation. Fluazinam, which is protectant with good residual effect but little eradicant and systemic activity, acts as a potent uncoupler of oxidative phosphorylation leading to disorder of the mitochondrial membrane (other factors may also be involved). Fenpiclonil, a phenylpyrrole, and fentin fungicides also inhibit oxidative phosphorylation.
9. Inhibition of respiratory chain in the mitochondria. A number of *E*-β-methoxyacrylates are specific inhibitors of the cytochrome bc_1 complex (complex III) in the mitochondrial respiratory chain. Azoxystrobin has protectant, eradicant, translaminar, and systemic properties. It is broad-spectrum in action against the major groups of pathogenic diseases that attack crops and is a

good fungicide for control of downy mildews. This class of fungicide (strobilurins) was derived from the study of edible woodland mushrooms.

10. Inhibition of fungal biosynthesis of nucleic acids. Some hydroxypyrimidines, e.g., ethirimol, were developed during the 1960s and are thought to inhibit fungal biosynthesis by inhibiting adenosine deaminase in the purine reutilization pathway. They have systemic action and are protectant and eradicant.
11. Inhibition of mitochondrial electron transport. Cytochrome bc_1 is involved as the fungicide inhibits electron transport in oxidative phosphorylation in the fungal mitochondria and energy production. Several aryl carboxanilides (oxazolidinones) inhibit succinate ubiquinone reductase and hence fungal respiration. This group of fungicides are specifically effective on *Basidiomycetes*. They are systemic, protectant, and residual. Carboxin is an older example of this class while famoxadone has much more recently reached the market.
12. Inhibition of glycerophospholipid biosynthesis. Edifenphos and validamycin A have this mode of action. Edifenphos is used as a protectant and eradicant fungicide for the control of rice blast.
13. Inhibition of protein synthesis. Some fungicides of microbial origin are used commercially to control disease, e.g., blasticidin S by foliar application to control rice blast, and jinggangmycin to control sheath blight of rice. Since these 'biochemicals' are made biologically, they may well be specific for target organisms. They are also likely to degrade *in vivo* and so be safer than synthetic chemicals.
14. Acquired systemic resistance in the plant and increase in the effectiveness of its defense mechanism. Some chemicals act not by killing the fungal disease but by affecting the defense mechanisms in the metabolism of the plant so that it is more resistant to attack and so continues to develop and function normally; endogenous salicylic acid is believed to play a crucial role in the induction process. Probenazole is a systemic compound used to control rice blast and can be absorbed by the roots. Some antifungal proteins have been found to increase the defense mechanism of the plant, although these have yet to impinge on the commercial market.
15. Biological control using bacteria, virus, and other fungi can work, although its commercial exploitation to date has not been successful. However, this approach can be a useful tool in integrated pesticide management (IPM) – and has been operative in natural systems for millions of years!
16. Unknown or very uncertain mechanism of action.
    a. Certain organophosphorus compounds have been widely used for particular crops for some time without any definitive understanding of their mode of action, although they are now believed to interfere with fungal choline biosynthesis. Pyrazophos is used as a protectant and eradicant fungicide for the control of rice blast. Tolclofos-methyl is a nonsystemic compound of a different type used, for example, on vegetables and fruits.
    b. Aromatic hydrocarbons. Quintozene can be used in dressing seed. Chlorothalonil has a nonsystemic and protective action, giving good control of *Oomycetes*. This group of fungicides has become increasingly suspect toxicologically, e.g., chlorothalonil is now banned in some countries.
    c. Pencyuron is also nonsystemic and protectant and controls rice sheath blight.
    d. Dicloran is protectant for control of *Botrytis*, *Sclerotonia*, and *Monilinia* spp. in fruits.
    e. Cymoxanil, a cyanoacetamide, is protectant and curative in action and is usually used in combination with compounds having contact or local systemic activity.

Resistance

Problems can arise from the build-up of resistance of fungi to control by a fungicide or, more often, by a group of fungicides.

Before 1970, resistance to the effects of the multisite inhibitor fungicides then mostly used was infrequent and only occurred after prolonged use, e.g., in vineyards. Where resistance did occur it was probably nonspecific in nature, e.g., from reduced uptake on the foliage surface.

With the introduction of site-specific systemic fungicides, the incidence of resistance problems became much more frequent and problems emerged more quickly. The risk of resistance from site-specific fungicides may have been increased through selection pressure resulting from their greater effectiveness and eradicant action.

To resolve a resistance problem, a systemic chemical may have to be formulated alongside another active ingredient of a different type, such as a dithiocarbamate, because of a partial loss of effectiveness in controlling fungal infections. Resistance problems with sterol biosynthesis inhibitors can sometimes be overcome by switching to another sterol biosynthesis inhibitor with a low level of cross-resistance or by changing the application method. Growers will also look to plant breeders for cultivars of plant, fruit, or vegetable that are more resistant to attack by disease.

The Groupement International des Associations Nationales de Fabricants de Produits Agrochemiques (GIFAP) has a Fungicides Resistance Action Committee (FRAC) which has working groups covering the six main classes of fungicide for which resistance is a problem: sterol biosynthesis inhibitors, phenylamides, dicarboximides, benzimidazoles, anilinopyrimidines, and strobilurins. Information and recommendations on fungicide resistance are available from this source.

A serious resistance problem has occurred with the control of late blight in potatoes. Aggressive populations of the A2 mating type of *Phytophthora infestans* have developed so that severe losses have occurred. Strategies that have been developed to cope with the problem include using nonsystemic protectant fungicides, translaminar fungicides, and a single *R*-enantiomer formulation of metalaxyl. Resistance to the systemic fungicides normally used has been marked and these failed to control the situation effectively.

Only a limited number of field trials with genetically modified plants have involved resistance to fungi.

Effective fungicide resistance management is crucial. Much management is based on experience but there is a need to develop predictive tools based on mechanistic and other knowledge and data. Many countries have national programs to monitor fungicide resistance.

Extent of Use

The Food and Agriculture Organization gives the world trade in fungicides in 2000 to be worth nearly US$ 3 billion (about 20% of the total pesticide market). This is probably an underestimate of the amounts actually used. The proportion of total fungicide sales, as an estimate of use, by region worldwide in 1995 is indicated in Table 3. In the same study, the total sales of fungicides among various crops worldwide were assessed, as shown in Table 4.

Table 5 shows the sales of fungicides on vegetables and cereals by region worldwide. It should be noted that all global figures for pesticide sales/use are approximate as reliable data may not be obtainable from China or some other South-east Asian countries. Seed treatments are estimated to account for about 10% of the total fungicide market. Although an increasing awareness of the advantages of IPM is evident, use of fungicides globally continues to increase.

Fungicides are available for home garden use on soft fruit, legumes, and other vegetables as well as on ornamentals.

Table 3 Fungicide sales by region in 1995

Region	*Percentage of total worldwide fungicide sales*
Western Europe	41
East Asia	35
North America	10
Latin America	8
Eastern Europe	3
Rest of the world	3

Calculated from data from County NatWest WoodMac for 1995.

Table 4 Relative sales of fungicides for various crops worldwide in 1995

Crop	*Percentage of total sales of fungicides*
Fruits and vegetables	41
Cereals	25
Rice	15
Maize	2
Cotton	2
Soybeans	1
Sugar beet	1
Potato	4
Others	9

Calculated from data from County NatWest WoodMac for 1995.

Table 5 Relative sales of fungicides for use on vegetables and cereals worldwide in 1995

Region	*Percentage of world sales for use on*	
	Fruits and vegetables	*Cereals*
Western Europe	35	83
East Asia	39	6
North America	15	4
Latin America	8	3
Eastern Europe	1	2
Rest of the world	2	1

Calculated from data from Country NatWest WoodMac for 1995.

Methods of Application

Fungicides are applied to growing crops in a wide variety of ways. The active ingredient(s) may be formulated with added wetter, emulsifier, filler, etc. as an emulsifiable concentrate, a suspension concentrate, a fungicidal dust, as granules, or as a wettable powder. Microencapsulated systems are now widely used and have advantages such as reduction in operating hazards, controlled release, and protection of the active chemicals against degradation in the field. Microemulsions are now being developed and offer an advantage of being thermodynamically stable. Reference should be made to the article on pesticide types for detailed information on techniques of application.

Applications to harvested produce are closely linked with procedures for storage and marketing the particular commodity. For example, dipping citrus fruit in a solution of an appropriate fungicide may be

followed by wrapping the individual fruits in fungicide-impregnated paper before packing in trays or boxes for transport.

To control some diseases, the seed is treated before planting by dressing with a dust or liquid or by encapsulating the seed with a nutrient coat which includes fungicide, e.g., sugar beet is normally sown in this way. Fungicidal seed bed drenches are also used.

Surface biocides are applied to empty animal houses and buildings for control of fungal and bacterial infection.

Significant developments in product packaging and pesticide delivery systems have taken place in recent years. Successful development of refillable containers of all sizes, water-dissolvable films or gel tablets, and microencapsulated pesticides substantially reduce or eliminate the need for recycling or disposal of packaging that has been exposed to pesticide and hence potentially may leave hazardous residues.

Examples of Benefits in using Fungicides

1. The quantity, or yield, of the crop or foodstuff is increased, making marketing more commercially viable. Globally, potential losses in crops because of disease are estimated to be about 17% and these are reduced by about a quarter to about 13% by using fungicides. (This averages out the effects of epidemics in which entire crops are lost.) Table 6 shows some of the global advantages of using fungicides in growing rice.

   Work carried out in Germany on the wheat harvests in 1991–93 indicated that the average yield would drop from 6.4 t ha^{-1} to 4.2 t ha^{-1} without the use of fungicides. Furthermore, in a no-pesticide scenario (no herbicides or insecticides either) the loss rate for diseases would increase from 6% to 11%.

   Infestation is related to air temperature and humidity and modern methods can forecast when a fungal infection is likely so that growers can apply fungicides to control proliferation of spores. Forecasting rapid spread of disease and advising on the timing for fungicide treatment was pioneered for the control of potato blight.
2. The quantity of the commodity is increased to obtain a wider market and a better value. Without fungicide dips with citrus fruit for export and of impregnated packing for transport, infections latent from the preharvest environment would severely reduce the proportion of a consignment that would be marketable and lower financial returns markedly. People prefer to buy apples without unsightly apple scab lesions in the skin and will refuse to buy inferior produce.
3. A disease may be eliminated. Bunt or stinking smut (*Tilletia caries*) of wheat now rarely occurs in the UK. It is probable that such diseases have been very efficiently controlled rather than eradicated.

Financial considerations of these benefits have to offset the cost of buying and applying fungicides.

Table 6 Potential losses and actual losses to diseases in rice in 1991–93, by region

Region	*Potential loss from diseases (% of yield)*	*Actual loss to diseases (% of yield)*
Western and southern Europe	21	9
North America and Oceania	24	8
East Asia	22	12
Rest of the world	18	17

These estimates are made alongside estimates of losses to other pests. Based on data from Oerke E-O at the 12th International Symposium on Modern Fungicides and Antifungal Compounds.

Safety Implications and Toxicology

Before a fungicide can be marketed, the manufacturer submits information on the biological, chemical, and physical properties of the active ingredient and of its formulations to pesticide registration authorities of the countries concerned. Possible effects on the operator, the consumer and on the environment are carefully evaluated and registration authorities have to be completely satisfied of the safety of the proposed chemical, its formulations, and the labeling directions for its use, as well as of safety arrangements in transport and storage.

Where sufficient information is available, the Joint FAO/World Health Organization (WHO) Meeting on Pesticide Residues (JMPR) has estimated acceptable daily intakes (ADIs) for pesticides. They are expressed in milligrams of the chemical per kilogram of body weight. Table 7 gives ADIs for some common fungicides.

Nevertheless, some adverse effects on wildlife and the environment have been attributed to fungicides. For example, investigations have suggested that iprodione has altered bird behavior, reduced egg production, reduced hatching rate, and impaired mysid reproduction. Azole fungicides, however, are known to have a low persistence in the environment and a low acute toxicity to birds. Triorganotin compounds are, expecially toxic to aquatic life.

Some fungicides once widely marketed are now withdrawn on toxicological grounds and their movement in trade between countries required prior informed consent (PIC). Organomercury compounds and hexachlorobenzene (HCB) are essentially banned

internationally, having a high toxicological risk to humans. Dinoseb and binapacryl are also suspect on toxicological grounds.

Maximum Residue Limits (MRLs)

The Codex Committee on Pesticide Residues (CCPR) recommends, on the advice of the JMPR, maximum residue limits for pesticides in foodstuffs, whether they are derived from preharvest or postharvest use, where such limits are necessary and whether sufficient information is available. MRLs generally apply to the raw commodity, although (separate) MRLs can be established for foods that have been processed, fabricated, or manufactured, if necessary. Codex MRLs for fungicides are listed in Table 8.

Maximum levels have also been fixed for residues of some fungicides by the European Community and the values corresponding to those of the Codex, where they occur, are given in Table 9. It will be seen that the European limits do not exactly correspond with those of the Codex, although they are generally of the same order of magnitude. This is because the different limits were the result of evaluations at different times, based on different data and by different experts. Some countries also have limits of their own covering particular interests. The overall lack of convergence of pesticide residues limits in foods is less than satisfactory.

Table 7 Acceptable daily intakes (ADIs) of some fungicides used in agriculture and food storage

Fungicide	*ADI (mg kg^{-1} body weight)*
Bitertanol	0.01
Carbendazim	0.03
Chlorothalonil	0.03
Dichlofluanid	0.3
Fenarimol	0.01
Imazalil	0.03
Iprodione	0.06
Metalaxyl	0.03
Propiconazole	0.04
Thiabendazole	0.1
Tolclofos-methyl	0.07
Triadimenol	0.05

Data from *Codex Alimentarius* vol. 28, 2nd edn. *Pesticide Residues in Food – Maximum Residue Limits*. Joint FAO/WHO Food Standards Programme. Rome: Food and Agriculture Organization (1998).

Analysis of Residues

Methods

A necessary precursor to determining any fungicides in a sample of a food is a satisfactory, and in as far as is possible representative, method of sampling. Both the CCPR and the European Community have recently addressed this issue and have revised their recommendations. Many fungicides are one of a small group of compounds having similar chemical structures and/or physical properties. A group, or groups, can often be analyzed using a 'multiresidue' method rather than by employing individual procedures for each possible fungicide.

If present, residues can be identified and quantified using a range of gas chromatographic (GC) columns and detectors, especially mass spectrometric detectors for enforcement purposes. Fungicides not amenable to GC are often determined using high-performance liquid chromatography with ultraviolet or fluorescence detectors. Dithiocarbamates are usually determined as carbon disulfide by a GC head-space procedure. Some fungicides, e.g., chlorothalonil, are

Table 8 Codex maximum residue limits (MRLs) for some fungicides used in agriculture and food storage

Fungicide	*Codex MRL (mg kg^{-1}) in*								
	Apple and pear	*Banana*	*Cabbage*	*Citrus fruits*	*Grapes*	*Potato*	*Peach*	*Tomato*	*Wheat*
Bitertanol	2	0.5					1		0.1
Carbendazim		1				3			
Chlorothalonil		0.2	1	5	0.5	0.2	(1)	5	0.1
Dichlofluanid	5				15	0.1	5	2	0.1
Fenarimol	(0.3)	0.2			(0.3)		(0.5)		
Imazalil	5	2		5		5			0.01[a]
Iprodione	5				10		10	5	
Metalaxyl	1		0.5	5	1	0.05[a]		0.5	
Propiconazole		0.1			0.5		1		0.05[a]
Tolclofos-methyl						0.2			
Thiabendazole	10	3		10		5		2	
Triadimenol	0.5	0.2			2			0.5	0.2

Data from *Codex Alimentarius* vol. 28, 2nd edn. *Pesticide Residues in Food – Maximum Residue Limits*. Joint FAO/WHO Food Standards Programme. Rome: Food and Agriculture Organization (1998).

[a]At or about the limit of determination.

MRLs in parentheses are not yet finalized Codex standards.

Table 9 European Union (EU) maximum limits for some fungicides used in agriculture and food storage

Fungicide	*European Community maximum limits (mg kg^{-1}) in*								
	Apple and pear	*Banana*	*Cabbage*	*Citrus fruits*	*Grapes*	*Peach*	*Potato*	*Tomato*	*Wheat*
Carbendazim	2	1		5	2	1	3[a]	0.5	0.1[a]
Chlorothalonil	1	0.2	3	0.01[a]	1, 3	1	0.01[a]	2	0.1
Fenarimol	0.3	0.3		0.02[a]	0.3	0.02[a]	0.02[a]	0.02[a]	0.02[a]
Imazalil	5	2	0.02[a]	5	0.02[a]	0.02[a]	0.02[a], 5	0.5	0.02[a]
Iprodione	10	3	5	0.02[a], 2, 5	10	5	0.02[a]	5	0.5
Metalaxyl	1		1	0.05[a]	1, 2	0.05[a]	0.05[a]	0.05[a]	0.05[a]
Propiconazole	0.05[a]		0.05[a]	0.05[a]	0.5	0.2	0.05[a]		0.05[a]
Thiabendazole	5	3	0.05[a]	6	0.05[a]		0.05[a], 5	0.05[a]	0.05[a]

The EU has not as yet fixed maximum levels in foodstuffs for many of the fungicides discussed in this chapter.
Where two figures are entered, these are values for different commercial commodities, e.g., table grapes, wine grapes.
[a]Lower limit of analytical determination.

difficult to analyze accurately as they are unstable and degrade during analysis. Sensitive immunoassay techniques have been developed for determining a number of fungicides: benomyl, chlorothalonil, fenpropimorph, metalaxyl, triadimefon, and other compounds. Such approaches have not, as yet, superseded the instrumental chemical approaches for definitive work, although they are increasingly useful for screening purposes.

Monitoring Data

It is not possible to summarize adequately all the selective monitoring data for fungicide residues in foodstuffs in those countries that report their work in the scientific or official publications. However, it appears that of the order of 1% of retail samples may contain residues above a residue limit. In an EU survey of 9700 samples, dithiocarbamates and benzimidazoles were the fungicides occurring most often as residues in foods and exceeded maximum limits the most frequently.

The relatively low toxicity, together with the small amounts, of copper and sulfur formulations likely to be in foods renders residue analysis for these moieties unnecessary. Significant traces of fungicides used for seed dressing, seed bed drenching, or as surface biocides in buildings are unlikely to be found in foodstuffs at market.

See also: **Citrus Fruits**: Composition and Characterization; **Pesticides and Herbicides**: Types of Pesticide; **Sugar**: Sugarbeet

Further Reading

Copping LG and Hewitt HG (1998) *Chemistry and Mode of Action of Crop Protective Agents*. London: Royal Society of Chemistry.

Dehne H-W and Derke E-C (1998) Impact of diseases and disease control on crop production. In: Hutson D and Miyamoto J (eds) *Fungicidal Activity*, pp. 1–21. Chichester: Wiley.

European Union (1998) Directive 98/82/EC of 27 October 1998 amending Council Directive 91/414/EEC concerning the placing of protection products on the market. *Official Journal of the European Communities* L 290/25–54.

European Union websites (2002) http://europa.eu.int/comm/dgs/health_consumer/index_en.htm and http://europa.eu.int/comm/foods/fs/ph_ps/index_en.htm/

FAO (1998) *Joint FAO/WHO Food Standards Programme. Codex Alimentarius*, vol. 28, 2nd edn. *Pesticide Residues in Food – Maximum Residues Limits*. Rome: Food and Agriculture Organization.

FAO (2002) Statistical databases, pesticide trade field. Available at www.fao.org.

Hewitt HG (1998) *Fungicides in Crop Protection*, 221 pp. Wallingford, UK: CAB International.

Kendall SJ and Hollomon DW (1998) Fungicide resistance. In: Hutson D and Miyamoto J (eds) *Fungicidal Activity*, pp. 87–108. Chichester: Wiley.

Kurihara N and Miyamoto J (eds) (1998) *Chirality in Agrochemicals*. Chichester: Wiley.

Lyr H (1987) *Modern Selective Fungicides*. Harlow, UK: Longman.

Lyr H, Russell PE, Dehne H-W and Sisler HD (1999) *Modern Fungicides and Antifungal Compounds II*. 12th International Reinhardsbrunn Symposium. Andover, UK: Intercept.

Oerke E-C (1996) Impact of diseases and disease control on crop production. In: Lyr H, Russell PE and Sisler HD (eds) *Modern Fungicides and Antifungal Compounds*, pp. 17–24. Andover, UK: Intercept.

Ordish G (1967) In: Torgeson DC (ed.) *Fungicides. An Advanced Treatise*, vol. 1, pp. 39–62. New York: Academic Press.

Pesticides 2002. London: HMSO.

Roberts TR, Hutson DH, Jewess PJ, Lee PW, Nicholls PH and Flimmer JR (eds) (1999) *Metabolic Pathways of Agrochemicals. Part 2: Insecticides and Fungicides*. London: Royal Society of Chemistry.

Russell P (1999) Fungicide resistance management: into the next millennium. *Pesticide Outlook* 10(5): 213–215.

G

GALACTOSE

C A Williams, Kingswood, Wotton-under-Edge, UK

Occurrence

Galactose is a monosaccharide and has the same chemical formula as glucose, i.e., $C_6H_{12}O_6$. It is similar to glucose in its structure, differing only in the position of one hydroxyl group. This difference, however, gives galactose different chemical and biochemical properties to glucose.

The major dietary source of galactose is lactose, a disaccharide formed from one molecule of glucose plus one of galactose. Lactose is found only in milk; after weaning, significant quantities of dietary lactose are found only in dairy products (Table 1). Lactose levels are lower than expected in some dairy products, where it has been used by the microbes involved in processing the food.

Lactose, a byproduct of the dairy industry, can be hydrolyzed to produce lactose hydrolysate syrup, which contains lactose, galactose, and glucose. This syrup is used as a sweetener in biscuits, confectionery, and some dairy desserts. Thus, small amounts of lactose and galactose can appear in nondairy foods. (*See* **Lactose.**)

Apart from its presence in lactose hydrolysate syrup, the monosaccharide galactose is seldom found in the diet, although it has been identified as a trace component of some seeds and pulses.

Absorption

It has been established that the monosaccharides are absorbed at different rates from the small intestine. Initial studies, using animal models, showed that glucose and galactose are absorbed at a much faster rate than the other monosaccharides, and galactose is absorbed at a slightly faster rate than glucose. This finding has been confirmed in humans. (*See* **Glucose:** Function and Metabolism.)

In the brush border of the small intestine, there is a carrier that is shared by glucose and galactose; interaction between the carrier and an adjacent sodium pump allows the monosaccharides to be actively accumulated in the intestinal cell prior to their removal into the circulation.

Galactose absorption is an active process in that it requires energy (in the form of adenosine triphosphate, or ATP) and it can be stopped by the presence of metabolic inhibitors. The common carrier system concept is supported by knowledge of a rare clinical condition of glucose–galactose malabsorption, in which neither monosaccharide can be absorbed efficiently because of the absence or impairment of the common carrier.

Prior to absorption, galactose has to be released from lactose. This occurs in the presence of lactase, a disaccharidase, found in the brush border of the mucosal epithelial cells of the small intestine.

Human lactase has a pH optimum of 5.5–6.0, and its highest levels of activity are found in the jejunum, with lower levels occurring in the duodenum and ileum. It is thought that the enzymes that hydrolyze sucrose and maltose are closely aligned with their monosaccharide carriers. This may not be so for lactase. However, in reality, the rate of lactose hydrolysis does not appear to limit the quantity of galactose or glucose absorbed from lactose.

Lactose intolerance occurs when the mucosal cells do not produce sufficient quantities of lactase to hydrolyze ingested lactose. As a result of this insufficiency, undigested lactose remains in the small intestine and, because of its osmotic potential, draws water into the intestinal lumen. On arrival at the large intestine, the lactose can then be fermented by the microflora present to produce various gases and acids. The symptoms produced can range from a mildly bloated feeling to watery, explosive diarrhea. There are three kinds of lactose intolerance:

1. Congenital lactase deficiency is an extremely rare disorder in which lactase is absent from birth. As lactose is a major constituent of the neonatal diet, early diagnosis and instigation of a lactose-free diet are essential if the condition is not to be fatal.
2. Secondary lactase deficiency occurs as a result of damage to the intestinal mucosa, caused, for

Table 1 Lactose content of milk and dairy products

Food	*Lactose content (g per 100 g)*
Cows' milk	4.7
Goats' milk	4.6
Human milk	7.2
Butter	Trace
Cream	2.0–3.2
Cheese (most types)	Trace
Cottage cheese	1.4
Yogurt	3.2–4.8

From Paul AA and Southgate DAT (1978) *McCance and Widdowson's The Composition of Foods*, 4th edn. London: Her Majesty's Stationery Office.

example, by an infection or disease. Lactase levels return to normal on successful treatment of the primary disorder.

3. In primary lactase deficiency, lactase levels are normal at birth but have declined by adulthood. A fall of lactase levels by adulthood has been noted in other mammals, but it was in the early 1960s that the condition was described in humans. Subsequently, it has been found that primary lactase deficiency is the normal state in most races, and with few exceptions, it is only in white Caucasians that lactase levels remain high in adulthood. Most lactose-intolerant individuals do not abstain totally from milk and can take small quantities regularly, but large quantities of milk are to be avoided. The absorption of other nutrients is not affected by drinking small quantities of milk. (*See* **Food Intolerance**: Lactose Intolerance.)

Metabolism of Galactose

The main route of galactose metabolism – the Leloir pathway – results in its conversion to glucose 1-phosphate, which can enter the glycolytic pathway (Figure 1). Intermediates of this pathway be used in glycogenesis, or mucopolysaccharide and glycoprotein synthesis. The pathway exists in the cytoplasm of the cells and is most active in the liver.

The Isselbacher pathway exists as an alternative route for metabolism of galactose 1-phosphate (Figure 1), a potentially toxic metabolite of galactose.

Pathways exist for the reduction of galactose to galactitol and for dehydrogenation to galacturonic acid.

There are two forms of clinical galactosemia, in which an enzyme concerned with galactose metabolism is absent from birth. Classical galactosemia is the result of a deficiency of galactose-1-uridyltransferase, so that galactose is phosphorylated to galactose 1-phosphate (gal 1-p) and no further. If the condition remains unchecked, gal 1-p and galactose accumulate in the tissues and can result in liver malfunction, cataract, mental retardation and failure to thrive.

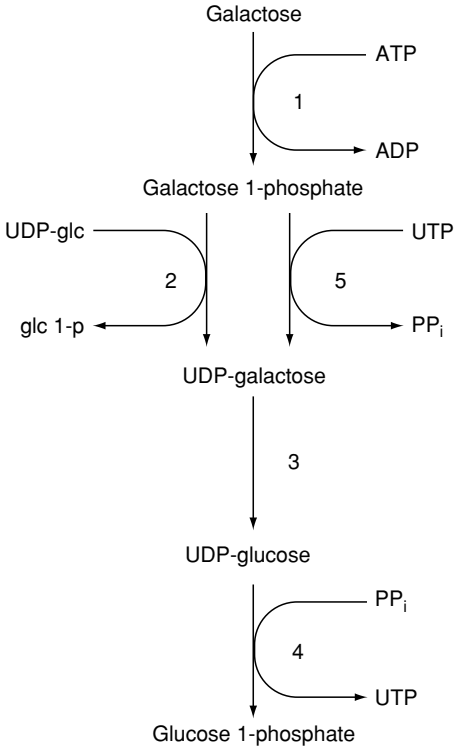

Figure 1 Galactose metabolism – the major pathways. The Leloir pathway: 1, galactokinase; 2, galactokinase 1-phosphate uridyl transferase; 3, UDP-galactose 4-epimerase; 4, UDP-glucose pyrophosphate. The Isselbacher pathway: 5, UDP-galactose pyrophosphorylase. *Note*: although unidirectional arrows have been drawn, most reactions are reversible. ATP, adenosine triphosphate; ADP, adenosine diphosphate; UTP, uridine triphosphate; UDP, uridine diphosphate; glc, glucose; glc 1-p, glucose 1-phosphate; PP_i, inorganic diphosphate. Reproduced from Galactose, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Early diagnosis and initiation of a galactose-free diet are essential. The second type of clinical galactosemia is galactokinase-deficient galactosemia, and this results in the accumulation of unphosphorylated galactose in the tissues. It is characterized by the development of cataracts in the first or second decade of life. Cataracts in both types of condition are caused by an accumulation of galactitol in the lens of the eye.

Metabolic Response to Galactose

The liver is the major site of galactose uptake and metabolism in humans, and practically all circulating galactose is removed by the liver. At one time, the disappearance of galactose from blood following an oral load was used as a test of liver function.

Normally, serum galactose levels are negligible in a healthy fasting individual. It is only after ingestion of galactose or lactose that they increase. The change

in blood galactose levels after ingesting galactose is directly related to the amount of galactose taken. The maximum blood galactose level achieved and the duration of elevated blood galactose levels increase with increasing galactose load. This is in marked contrast to the blood glucose response to differing glucose loads, which is highly controlled, primarily through the action of insulin.

Blood glucose levels increase after galactose ingestion, which is not surprising considering that the major metabolic route for galactose metabolism results in glucose production. Metabolic studies have found the rate of conversion of galactose to glucose to be very efficient in humans. Excretion of $^{14}CO_2$ after ^{14}C-1-galactose is only slightly less than after ^{14}C-1-glucose (galactose would need to be converted to glucose prior to excretion as CO_2). (*See* **Glucose**: Maintenance of Blood Glucose Level.)

A number of studies have reported that galactose metabolism to glucose, as reflected by blood glucose response to oral galactose, is more efficient in young infants than in older children or adults. These findings are, however, disputed.

In humans, oral galactose results in a slight increase in serum insulin levels, although animal studies have found galactose to be incapable of stimulting insulin release from the pancreas. It is believed to be the increase in blood glucose levels after galactose ingestion that brings about the increase in blood insulin levels. In galactosemic subjects, when galactose cannot be converted to glucose, oral galactose does not produce an increase in blood insulin levels.

Various carbohydrates have been found to have differing effects on metabolic rate (MR) and the respiratory quotient (RQ). Equivalent meals of galactose, galactose with glucose, and lactose produced similar changes in MR but had different effects on RQ in man. Galactose alone produces a greater and more rapid increase in RQ than after the other test meals mentioned, or after glucose alone. A larger RQ suggests that a greater level of glucose oxidation is occurring. (*See* **Carbohydrates**: Digestion, Absorption, and Metabolism; Metabolism of Sugars.)

Galactose, lactose, and glucose given with galactose loads have been shown to produce an increase in blood lactate levels in fasting subjects. This is not found after glucose loads. This result suggests a greater increase in glycolysis after galactose-containing meals, as is also suggested by RQ changes.

If glucose is ingested with galactose, either as a monosaccharide mixture or as lactose, the subsequent blood galactose levels are much lower than if galactose had been taken alone. It has been established that this effect is not due to changes in the rate of absorption of galactose, and it cannot be brought about by fructose. Intravenous insulin given at the same time as oral galactose does not influence the blood galactose response to oral galactose, even though the blood glucose levels are low during such a procedure. Intravenous glucagon reduces the blood galactose response to oral galactose and is accompanied by a large increase in blood glucose. It would appear that, in the presence of an elevated blood glucose level, blood galactose levels are lower after a galactose load.

Blood galactose, glucose, and insulin responses to a galactose and glucose mixture are the same in lactose-intolerant and lactose-tolerant individuals.

Alcohol has the opposite effect to glucose on blood galactose levels. Alcohol ingestion with oral galactose increases the blood galactose response to galactose. It is thought that the oxidation of alcohol in the liver increases the level of reduced nicotinamide adenine dinucleotide, which is essential for uridine diphosphate-galactose epimerase activity. Hepatic metabolism and uptake of galactose are therefore inhibited.

The blood galactose, glucose, and insulin levels after lactose are the same as after the equivalent mixture of glucose and galactose. The blood galactose response to lactose is increased if alcohol is taken with the lactose.

Long-term Ingestion of Galactose

Lactose hydrolysate syrup fed at levels of 5–7% of total energy intake for 10 weeks to men and women caused a slight, but insignificant, increase in fasting blood triglyceride levels in males but not females. There are conflicting reports concerning the long-term effects of feeding lactose on serum triglyceride and cholesterol levels.

For many years, dietary lactose has been associated with improved calcium absorption in humans and animals. It would seem that in lactose-tolerant individuals, lactose and glucose given with galactose can increase calcium absorption. It is not known how this effect occurs. (*See* **Calcium**: Physiology.)

See also: **Calcium**: Physiology; **Carbohydrates**: Digestion, Absorption, and Metabolism; Metabolism of Sugars; **Food Intolerance**: Lactose Intolerance; **Glucose**: Function and Metabolism; Maintenance of Blood Glucose Level; **Lactose**

Further Reading

Ali R and Evans JL (1973) Lactose and calcium metabolism: a review. *Journal of Agriculture (University of Puerto Rico)* 57: 149–164.

Cohn RM and Segal S (1973) Galactose metabolism and its regulation. *Metabolism* 22: 627–643.

Dahlqvist A (1984) Lactose intolerance. *Nutrition Abstracts and Reviews* 54: 649–658.

Gitzelmann R and Hansen RG (1980) *Inherited Disorders of Carbohydrate Metabolism*, pp. 61–101. London: MTP Press.

Williams CA (1986) Metabolism of lactose and galactose in man. *Progress in Biochemical Pharmacology* 21: 219–247.

GALLBLADDER

A Madden, London Metropolitan University, London, UK

Introduction

The gallbladder is located under the right lobe of the liver. It is a pear-shaped sac, usually measuring about 9 cm in length, but varying between 4 and 14 cm. Its average volume in health is 50 ml but this may treble in disease conditions. The thin and easily distended walls consist of undifferentiated muscle layers covered by a partial or complete covering of peritoneum. The mucosa is deeply folded and these indentations penetrate the muscle layer.

Structure

The anatomy of the gallbladder and biliary tract, shown in **Figure 1**, may vary between individuals. These variations have little clinical significance except during surgery or radiological investigations.

The fundus, which is the wide end of the gallbladder, may be palpable on physical examination. The body of the gallbladder joins the cystic duct at the neck, where there is an indentation – Hartmann's pouch. The cystic duct, with its spiral mucosal folds, joins with the common hepatic duct to form the common bile duct, which is 5–10 mm in diameter. This passes behind the duodenum, where it is joined by the pancreatic duct before reaching the ampulla of Vater. Around the duct, as it passes through the duodenal wall and ampulla, is a ring of circular muscle fibers called the sphincter of Oddi, which is responsible for controlling bile flow into the duodenum.

The cystic artery, a branch of the hepatic artery, supplies the gallbladder with blood which returns to the portal venous system via the cystic vein. The gallbladder is well innervated by both the sympathetic and parasympathetic nervous systems.

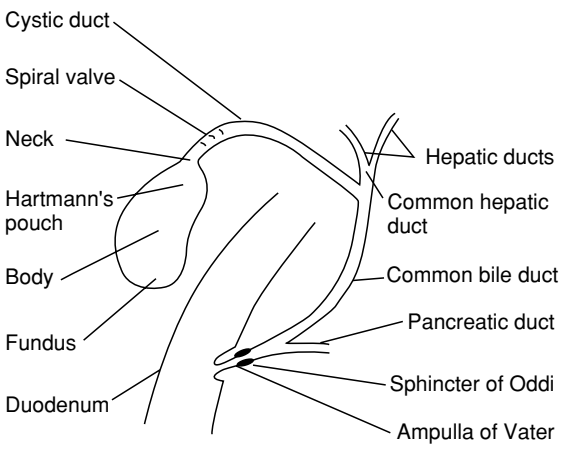

Figure 1 The gallbladder and connecting biliary ducts. Reproduced from Gall Bladder. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Function

The role of the gallbladder is to concentrate and store bile produced by the liver. When the gallbladder contracts, bile passes into the duodenum. Contraction of the muscles in the gallbladder is under both humoral and neurological control. The presence of food, especially fat, in the duodenum and jejunum causes local secretion of cholecystokinin (CCK) by the mucosa; CCK is carried via the blood to the gallbladder, where it stimulates muscle contraction. This contraction may also be brought about by vagal stimulation, which causes a simultaneous relaxation of the muscle in the sphincter of Oddi. In the duodenum and small intestine, bile is essential for the emulsification of dietary fat.

Gallstones (Cholelithiasis)

The most common disorder affecting the gallbladder is the formation of stones. This occurs predominantly in populations consuming a western-style diet, with an incidence of 12–20%.

Gallstones are composed most commonly of a mixture of cholesterol, various calcium salts, phospholipids, and bile pigments. Cholesterol-rich (up to 98% in some cases) stones are the most common, but occasionally stones may be composed almost entirely

of bile pigments. The latter are smaller and colored brown or black.

The differing composition of stones is determined by the factors influencing formation. Stones containing cholesterol occur when the bile stored in the gallbladder is supersaturated with cholesterol. This precipitates out into solid crystals, which settle on a nucleus of mucoprotein or bacteria and gradually increase in size. Stones made predominantly from pigment are more prevalent in individuals suffering from chronic hemolysis, as in sickle-cell disease or thalassemia.

Etiology of Gallstones

Although cholesterol supersaturation of bile is a prerequisite for the formation of cholesterol gallstones, this does not automatically lead to the production of stones. There are several other etiological factors.

Biliary Factors

Bile acids are required for the formation of micelles containing cholesterol and phospholipid in bile. If the concentration of bile acids is low, there may be precipitation of cholesterol. An increase in glycoprotein content may also encourage stone formation. Although bile is often sterile, repeated infections of the gallbladder can promote stone formation by deconjugating bile salts and providing nuclei for cholesterol deposits. Stones are more likely to form in a diseased gallbladder which does not empty fully on contraction, leaving residual bile susceptible to cholesterol precipitation.

Age, Sex, and Family History

Although gallstones may be found in either sex and at any age, the prevalence is twice as common in women as men, and increases with age in both sexes. Women who have never had children are less likely to develop stones than multiparous women. In late pregnancy the gallbladder does not empty completely, leaving residual bile to promote stone formation. Estrogens reduce bile acid secretion, again encouraging stone formation, and there is a higher incidence of their formation in women taking oral contraceptives. The risk of developing gallstones is greater in relatives of gallstone sufferers than in families with no history of the condition.

Physical Activity

Epidemiological studies suggest that the incidence of gallstones is inversely proportional to physical activity. A sedentary lifestyle is a significant risk factor for gallstone formation, independent of obesity, which is frequently associated with a lack of activity. It has been suggested that 30 min of endurance-type activity undertaken five times per week could prevent up to a third of symptomatic gallstone disease in men. Further studies are required to quantify the potential benefits of exercise in women.

Nutritional Factors

Obesity Obese individuals may have slightly raised plasma cholesterol levels, but excretion of cholesterol in their bile is usually greatly increased. As a result, gallstones are more commonly found in individuals weighing above the ideal weight range for their height.

Ironically, bile becomes supersaturated with cholesterol during weight loss and the risk of gallstone formation increases; this is particularly apparent in individuals following very-low-calorie diets. However, once a stable reduced weight has been achieved, cholesterol levels become unsaturated again, so reducing the risk of stone formation.

Western diets The varying geographical prevalence of gallstones supports the suggestion that there is a strong dietary factor involved in their etiology. Africans have a very low incidence of stones, while in western Europe and the USA they are much more common. Worldwide, the incidence is increasing as western-style diets are adopted in developing countries. The dietary characteristics which appear responsible for promoting gallstone formation include a high intake of refined carbohydrate and a low intake of dietary fiber. In experimental animals, gallstone formation may be induced by diets rich in refined carbohydrate which suppress bile acid production.

Fiber If a high-fiber diet is consumed there is a greater excretion of bile acids in the feces and, as a result, less is reabsorbed in the terminal ileum and less reaches the liver for resecretion. This stimulates an increase in hepatic synthesis of bile acids. When less fiber is consumed, an increased transit time allows dehydroxylation in the colon of the bile salt cholic acid to deoxycholic acid. This secondary bile acid is reabsorbed and suppresses hepatic synthesis of further primary bile acids, so reducing the bile acid pool and promoting cholesterol precipitation and thus gallstone formation.

Vegetarians Epidemiological and case-control studies have shown a reduced incidence of cholelithiasis among vegetarians. A high intake of vegetables among women eating mixed diets may be protective. These effects appears to be independent of body weight.

Cholesterol The cholesterol content of bile is only influenced by dietary cholesterol if more than

1000 mg is consumed daily for at least 3 months. This intake is unlikely under normal circumstances and for most individuals the risk of gallstone formation can be considered to be independent of dietary cholesterol intake.

Unsaturated fat There is evidence to suggest that subjects consuming a diet low in saturated fat and cholesterol but high in unsaturated fat may be more susceptible to gallstones. One possible mechanism is that such diets encourage the mobilization of cholesterol from plasma and tissue and, therefore, its excretion in greater quantities in the bile.

Fasting There is diurnal variation in the composition of bile, with cholesterol levels being higher in bile produced at night. This is more likely to influence stone formation if the more lithogenic nocturnal bile is retained in the gallbladder for long periods. As food intake stimulates gallbladder contraction, long intervals of 8–15 h between meals will result in bile stagnation and the increased risk of stone formation.

Alcohol A moderate intake of alcohol may protect against gallstone formation.

Caffeine There is limited epidemiological evidence to suggest that consuming four or more cups of regular coffee per day may help prevent symptomatic gallstone disease in men. This effect has not been observed with decaffeinated coffee.

Clinical Factors

Patients with certain medical conditions are more susceptible to gallstone formation. Approximately 30% of individuals with cirrhosis of the liver have gallstones, predominantly composed of pigment. Resections of the ileum also increase the risk of stone formation as bile acids are not reabsorbed, resulting in a reduced total bile salt pool.

Cholestyramine and clofibrate, prescribed to lower plasma lipid concentrations, may also increase the risk of gallstone formation. Cholestyramine acts by binding bile acids, preventing reabsorption, while clofibrate increases biliary cholesterol excretion.

Long-term parenteral nutrition may result in stone formation in adults and children because in the absence of oral intake there is reduced stimulus for gallbladder, thus stagnant, lithogenic bile is retained.

Treatment

Once gallstones have formed they may remain asymptomatic or cause cholecystitis, discussed below. If an individual has no symptoms, and the stones are diagnosed incidentally, then no treatment may be required. Approximately 15% of asymptomatic stone carriers will develop symptoms during a follow-up of 10–15 years, and a policy of whether or not to treat them while they remain asymptomatic varies from center to center.

Dietary Management

It would be prudent for obese individuals with gallstones to lose weight with a moderate calorie restriction. For those within the ideal weight range, a healthy diet along the lines recommended by the Committee on Medical Aspects of Food Policy would be suitable. This advised that fat in the diet should provide a maximum of 35% of food energy and that fiber intake should be increased. In the UK in 1997 the average diet provided 39% of total calories from fat, and only 12 g of fiber (measured as non-starch polysaccharide) per day. A reduction in fat towards a maximum of 35% energy and an increase in fiber intake to approximately 18 g day^{-1} would be suitable for individuals with gallstones, as well as for the healthy adult population. There is no evidence that a lower fat intake would be beneficial for individuals with gallstones.

A regular intake of three meals a day, including breakfast, which will insure periodic emptying of the gallbladder, may also be helpful.

Cholecystitis

Once gallstones have formed they may not give rise to symptoms, unless they move into the neck of the gallbladder, cystic duct, or common bile duct. Here they can obstruct the release of bile, so that when the gallbladder contracts the pressure rises, causing pain. In addition, the retained bile acts as a chemical irritant on the mucosal wall and makes the area susceptible to bacterial infection. Bile salts may be deconjugated by bacteria to produce further irritation and injury to the mucosa. In severe cases the pressure rise compresses small blood vessels, producing infarction and gangrene.

Cholecystitis is classified as acute or chronic. An acute attack can develop where there is already underlying chronic disease. It may also occur where there has been no previous history of gallbladder disease.

Acute Cholecystitis

This is characterized by severe pain in the right upper abdomen or epigastrium, lasting up to an hour without relief. Nausea may be present and the patient may become pyrexial. If a gallstone has escaped from the gallbladder and lodges in the common bile duct, the

patient may become jaundiced. Eighty-five percent of patients recover spontaneously from acute cholecystitis. However, further attacks are common, although there may be long intervals between them. Complications of acute cholecystitis include gangrenous cholecystitis, liver abscess, peritonitis, and the formation of fistulae to the bowel.

Treatment

The patient is treated with antibiotics to control the infection, usually caused by gut organisms such as *Escherichia coli*, *Streptococcus faecalis* or *Klebsiella* sp. Analgesia and intravenous fluids are also required in the acute phase as the patient may be severely ill and may be unable to take oral fluid.

After the acute attack has subsided, surgical removal of the gallbladder (cholecystectomy) is recommended to prevent further episodes of pain or progression to the complications of empyema. The mortality of cholecystectomy carried out within the first 3 days of acute cholecystitis is 0.5%.

Dietary Management

Traditionally, when the severe pain of cholecystitis has subsided and the patient is able to eat, a light, usually low-fat diet is recommended. However, an evidence-based reevaluation of this practice is required. Dietary fat is a potent stimulator of CCK secretion which causes gallbladder contraction and possibly exacerbates pain. However, CCK is also secreted in response to dietary protein, mixed meals, elemental diets, and both regular and decaffeinated coffee. In addition, the gallbladder will contract in response to sham feeding without any gastric, duodenal, or jejunal stimulation. This cephalic phase of gallbladder contraction is independent of CCK secretion and possibly mediated through a cholinergic mechanism. If a gallbladder is able to contract in the absence of ingested food, there is little evidence to support the avoidance of dietary fat. A study of gallbladder dynamics in patients with gallstones measured the effect of test meals containing varying amounts of fat on the postprandial gallbladder volume. The response was shown to be independent of the fat content in the meals. Most studies of gallbladder contraction have been carried out on subjects with functioning gallbladders but it is unlikely that the response to food stimulation would be greater after acute cholecystitis, if it is changed at all. In the absence of any evidence to the contrary, it would be justified to advise patients to try to eat a normal healthy diet after acute cholecystitis. There is no justification for imposing a virtually fat-free regime in an attempt to avoid pain. If certain foods cause distress to an individual it is reasonable to avoid them. This should not cause nutritional problems if the diet is varied and balanced. Reassurance about eating in general may be needed, as many people are afraid to eat after the severe pain they have experienced. If they remain unwell, supplementary drinks or nutritious 'between-meal snacks' would help to insure that their total nutrient intake is adequate.

Chronic Cholecystitis

This may arise with or without a previously documented attack of acute cholecystitis. The symptoms may be vague, with epigastric discomfort, particularly after eating, and distension of the abdomen. There may be nausea and flatulence which is often relieved by belching or taking antacids. Diagnosis may be difficult because of ill-defined symptoms. Moreover, dyspepsia may be associated with other gastrointestinal diseases. Ultrasound is used to diagnose gallstones and gallbladder disease. Endoscopy may be necessary to rule out peptic ulceration, hiatus hernia, or gastritis.

Treatment

Elective cholecystectomy is recommended if the patient suffers repeated attacks of pain. The mortality is 0.1% in patients less than 50 years old and 0.5% if the patient is over 50 years old.

If patients are unable to undergo surgery because of age, poor general health, or other conditions, they may be treated medically. Dissolution of gallstones using oral bile salts may be attempted if the cystic duct is patent and the stones are radiolucent (i.e., predominantly cholesterol-containing). Patients must be compliant and prepared to follow the treatment for at least 2 years. The aim of dissolution therapy is to expand the total bile salt pool and increase the concentration of bile salts in the bile by giving oral cheno- or ursodeoxycholic acid. Successful dissolution occurs in 14–30% of patients after 2 years of treatment, but side-effects include diarrhea and increased plasma cholesterol. Multiple stones may also be dissolved by topically applying methyl tertbutyl ether. Small numbers of medium-sized stones may be fragmented using extracorporeal shock-wave lithotripsy. However, stone recurrence within 5 years is a problem in approximately 50% of patients undergoing nonsurgical treatment. These results compare unfavorably with cholecystectomy, but should be considered in patients for whom surgery is too great a risk.

Dietary Management

As with asymptomatic gallstones, the first priority for obese patients is to lose weight to within the ideal range for their height. As well as ultimately reducing the cholesterol level in bile, gentle weight reduction

will make potential surgery easier and less hazardous. When normal weight is attained, a healthy diet should be continued, in line with recommendations for the general population: a maximum of 35% of energy from fat and 18 g fiber day^{-1}.

Patients with chronic cholecystitis frequently describe dyspepsia associated with certain foods but there is little evidence to incriminate any particular cause. When subjects describing fat intolerance were challenged with fat without their knowledge, only 8% suffered dyspepsia. Symptoms therefore appear to relate more to preconceived ideas about dietary fat than the fat itself. If particular foods are thought to provoke gastrointestinal symptoms the individual may wish to avoid them, but otherwise a liberal but healthy diet should be recommended.

After Cholecystectomy

Although up to one-third of patients have symptoms after surgical removal of the gallbladder, this is probably because they did not have cholecystitis in the first place: 95% of patients found to have gallstones at surgery suffer no further symptoms afterwards. With the removal of the gallbladder and its function as a bile reservoir there is a redistribution of the bile acid pool. Increased bile acid cycling in the enterohepatic circulation results in greater cholesterol solubility. After surgery, patients should eat a normal healthy diet.

Obstructive Jaundice

Obstructive jaundice may be caused by a gallstone lodged in the common bile duct, carcinoma of the pancreas, bile duct or ampulla and, rarely, benign biliary strictures. A patient presenting with obstructive jaundice will need medical investigation to determine the cause of the obstruction and the treatment required. Ultrasound scanning and cholangiography are needed.

Typically, the patient with a common duct stone has epigastric pain, jaundice, dark urine, and fever. Malignant bile duct obstruction is usually painless and without fever. Once the cause is confirmed, medical or surgical treatment is selected.

Dietary Management

Dietary advice should include both basic nutritional support in patients who may be ill, frail and cachexic, and symptomatic help for those with fat intolerance. Depending on the degree of obstruction, little or no bile may enter the duodenum, so that dietary fat may be inadequately emulsified, resulting in fat malabsorption (steatorrhea). If the obstruction is at the lower end of the biliary tract, e.g., at the ampulla of Vater, the pancreatic duct may also be obstructed, and the lack of pancreatic lipase will exacerbate steatorrhea. If steatorrhea is present, dietary fat should be reduced to levels tolerated by the individual and, if appropriate, pancreatic enzyme supplements prescribed. The calorie deficit from the reduced fat must be made up with an increased carbohydrate and protein intake. Medium-chain triglycerides may be useful as an additional source of calories as these are partially water-miscible and do not require bile for emulsification. When steatorrhea is present, fat-soluble vitamins should be given by intramuscular injection as enteral absorption is unreliable. If a patient does not have steatorrhea there is no benefit from restricting dietary fat, and such a diet should not be recommended prophylactically because of its negative impact on palatability and energy intake.

If obstructive jaundice is accompanied by recurrent infections (cholangitis) and the situation cannot be improved by surgical or medical treatment, for example in primary sclerosing cholangitis, it is essential that adequate nutritional support is given since weight loss and cachexia are commonly seen. Palatable supplements and plenty of encouragement to eat may be needed to help a patient achieve an acceptable nutritional intake and prevent further malnutrition.

Role of Low-fat Diets in Gallbladder Disease

It is a commonly held misconception among both the general public and some professionals that disorders of the gallbladder should automatically be treated with a diet restricted in fat. If this is taken to extremes and continued for long periods (over 3 weeks), several problems may arise. A very-low-fat diet is unpalatable and may require a large bulk of food to be eaten in order to provide sufficient calories. If the low-fat diet results in an inadequate energy intake, undesirable weight loss may occur. Clinical deficiencies of the fat-soluble vitamins, A, D, E, and K, may also eventually arise if the dietary intake is deficient and supplements are not given. An adequate intake of essential fatty acids, linoleic and linolenic, is unlikely to cause the characteristic skin changes associated with deficiency unless a severely fat-restricted diet is followed for several months but, even for shorter periods, such a diet will result in a suboptimal and unacceptable nutritional state.

There is insufficient evidence to support the widespread use of fat restriction in the avoidance of pain in gallbladder disease. However, when pain is associated with fat intake in an individual, a reduced fat intake may help, provided that the total diet remains adequate.

Fat restriction may be beneficial for some patients with obstructive jaundice accompanied by diarrhea. In this case, it should be followed only on medical advice and with the guidance of a state-registered dietitian.

See also: **Caffeine**; **Cholesterol**: Factors Determining Blood Cholesterol Levels; **Dietary Fiber**: Properties and Sources; **Fats**: Requirements; **Obesity**: Etiology and Diagnosis; **Vegetarian Diets**

Further Reading

Andersen T (1992) Liver and gallbladder disease before and after very-low calorie diets. *American Journal of Clinical Nutrition* 56(S1): 235S–239S.

Corazziari E, Shaffer EA, Hogan WJ, Sherman S and Toouli J (1999) Functional disorders of the biliary tract and pancreas. *Gut* 45(S2): 48–54.

Department of Health (1991) *Dietary Reference Values for Food Energy and Nutrients for the United Kingdom*. Report on Health and Social Subjects 41. Committee on Medical Aspects of Food Policy. London: Department of Health.

Hopman WPM, Jansen JBMJ, Rosenbausch G and Lamers CBHW (1987) Cephalic stimulation of gall bladder contractions in humans. *Digestion* 38: 197–203.

Madden A (1992) The role of low fat diets in the management of gall-bladder disease. *Journal of Human Nutrition and Dietetics* 5: 267–273.

Mogadam M, Albarelli J, Ahmed SW, Grogan EJ and Mascatello VJ (1984) Gallbladder dynamics in response to various meals: is dietary fat restriction necessary in the management of gallstones? *American Journal of Gastroenterology* 79: 745–747.

Niebergall-Roth E, Teyssen S and Singer MV (1997) Neurohormonal control of gallbladder motility. *Scandinavian Journal of Gastroenterology* 32: 737–750.

Portincasa P, van de Meeberg P, van Erpecum KJ, Palasciano G and VanBerge-Henegouwen GP (1997) An update on the pathogenesis and treatment of cholesterol gallstones. *Scandinavian Journal of Gastroenterology Supplement* 223: 60–69.

Sherlock S and Dooley J (1996) *Diseases of the Liver and Biliary System*, 10th edn. Oxford: Blackwell Science.

Strasberg SM (1998) The pathogenesis of cholesterol gallstones: a review. *Journal of Gastrointestinal Surgery* 2: 109–125.

Tseng M, Everhart JE and Sandler RS (1999) Dietary intake and gallbladder disease: a review. *Public Health Nutrition* 2: 161–172.

Utter A and Goss F (1997) Exercise and gallbladder function. *Sports Medicine* 23: 18–27.

GAME

Contents

Birds

P J Hudson, University of Stirling, Stirling, UK

Introduction

Game bird production and harvesting are undertaken principally for the financial and social benefits that accrue from shoot-letting rather than as a source of nutrition. However, since the regulations that relate to private property rights vary between countries, so the relative importance of game birds as a food source and as an important form of financial income varies around the world.

Game Bird Production and Revenue

Within the British Isles, the laws relating to hunting season are liberal and restrictions on the level of the harvest are left to the individual who owns the rights to take game from the property, usually the land-owner or, on common ground, the Lord of the Manor. Since game production can result in a profitable harvest, the land-owner has an incentive to manage the land in a manner sympathetic to game production and, when possible, to rear and release game birds to supplement wild birds. People with an interest in shooting then pay for the privilege of shooting the birds, usually on a daily basis, and the dead birds remain the property of the land-owner.

The financial return from the letting of game shooting can provide a significant income to many British

estates. Consider the financial aspects of grouse shooting in Britain. The red grouse is a wild game bird restricted to the heather moorlands of the British Isles; numbers are not supplemented through rearing or releasing but are increased through the care of the heather habitat and the control of harmful predators and parasites. The revenue generated from the letting of grouse shooting depends on the type of shooting undertaken. Driven shooting, where a group of 16 people drive the grouse to a line of 8–10 people standing with guns, can only occur when August grouse densities are sufficiently high and usually only when they exceed 60 birds per square kilometer. During 1991, driven grouse shooting was let at £70 per brace (one brace = two birds) and daily grouse bags of up to 500 and 650 brace per day were recorded. However, this level of harvesting is unusual and a productive day's driven grouse shooting is normally in the region of 100 brace per day. Even so, a day's let grouse shooting which generates £7000 gross per day forms an important source of financial return for upland estates which would otherwise have to depend on poor returns from sheep farming or commercial afforestation.

The sale of the game from these commercial shoots is an insignificant return. During 1991, old grouse were selling at £1 per brace and young grouse at £3.50 per brace, which amounted to approximately 5% of gross margin. Traditionally, each sportsman receives a brace of birds at the end of a day's shooting and the remainder are sold to game dealers who export the meat as a delicacy to continental Europe.

The revenue gathered from game shooting can be significant; within Scotland, the direct and indirect revenue of grouse shooting alone in 1989 generated a total of £10.3 million to the Scottish gross domestic product and created the equivalent of 2323 full-time jobs. These jobs were often in remote rural areas in which alternative opportunities were limited.

The financial returns from birds reared and released provide a much lower net return since the costs of rearing and subsequently feeding are high. In 1991, pheasant rearing costs almost equalled the £15 per bird return at which pheasant shooting is let. Once again the carcass value is low, at only £1 per bird or 7% of the gross margin.

In contrast to the British system, North American hunters have free access to vast areas of land and the game that it may contain and generally must purchase a license to take a limited number of game. Consequently the shooting season is shorter, but more intense, with a limited annual bag for each quarry species. Financially the system is designed so that the money generated from licenses and from taxes on guns and sporting equipment is used to finance the state Fish and Wildlife Services. Unlike the British intensive system of game production, the American system provides an incentive for a larger proportion of the population to be involved in the sport and have access to the game as a source of food.

Harvesting Strategies

The objective of most harvesting strategies is to produce the maximum long-term yield of game from a population (known as the maximum sustainable yield or MSY), be it for economic revenue, quantity of food produced, or even the esthetic benefits arising from the conservation of the game bird resource. The size of the MSY that can be obtained from a population can be estimated through mathematical models using a number of approaches.

The simplest is to consider a population that increases according to a logistic growth curve where the per capita rate of increase falls linearly as the population rises; when this reaches zero the population has reached its carrying capacity. Harvesting this population will result in the size of the population falling to a level where the harvesting rate is equal to the population growth rate. The MSY is reached when the harvest rate equals the maximum growth rate of the population, and for the logistic model this is when the density is at half the carrying capacity. Higher harvesting rates will result in the final extinction of the population, whereas lower harvesting rates allow the population to rise until harvesting rate equals the growth rate of the population.

For any regulated population it is important to realize that any sustainable yield, including the MSY, lies at a level below the carrying capacity of the environment. This point is rarely appreciated by managers of wild animal populations or even by many biologists who are frequently concerned with keepering a breeding population as high as possible in the misplaced belief that this will increase the quantity of the resource available for harvesting. MSYs have been estimated for a number of game bird populations; for example, MSY for bobwhite quail is 55% of the autumn population, which reduces the size of the population to 72% of the unharvested state. For gray partridges the MSY is about 37%, mallard 29%, and pheasants 20% of the female population. For other species MSYs may not be easily calculated since numbers may fluctuate greatly and the optimal solution may depend on whether the population is increasing or decreasing.

The release of reared birds provides a supplement to the wild population but may result in overharvesting of the wild population and in the long term the loss of a wild breeding population. For example,

in Italy, 275 000 gray partridges are reared and released annually and approximately 65% are subsequently shot. Such a level of harvesting is far too high for the wild population to be sustained and yet this is not apparent to the hunters, since the rearing has maintained the size of the harvest. The released birds have a poor survival rate and rarely breed well, so there is insufficient recruitment into the wild population.

Environmental Benefits of Game Bird Production

One of the major benefits of wild game bird shooting are the indirect benefits of game management to the environment. Game management has frequently resulted in the amelioration of the harsher effects of agriculture and commercial afforestation on the environment.

Inevitably farming has produced major ecological changes to the environment, particularly in recent years through the mechanization of farming methods and the application of pesticides, fertilizers, and other chemicals which may have both direct effects in reducing nontarget species and indirect effects through the loss of food species utilized by other organisms. Perhaps the most obvious effect has been the dramatic decline in biodiversity associated with the cereal ecosystem. The effects on game birds (Table 1) have been dramatic but have also included a large number of other species such as the almost total demise of the once-abundant corn crake. Even in the uplands of Britain one of the most important impacts has been the overutilization of the seminatural habitat by domestic stock, which has directly destroyed habitats and in so doing reduced the abundance of the invertebrate food taken by immature game birds (Table 1).

The simple answer to reducing environmental damage and problems of agricultural intensification is to change the direction of the system towards one of extensification. Reduction of the impact of pesticides on cereal fields without significantly reducing the yield can be achieved through the use of conservation headlands, where limited application of pesticides results in the maintenance of beneficial broad-leaved weeds and the invertebrate fauna associated with them.

Large areas of the British countryside are planted with woodland specifically with the objective of producing pheasants. The maintenance of these woodlands, production of rides, and other management activities have proved highly beneficial to song birds, butterflies, and other animals. Similarly, heather moorlands are a habitat of international importance, a large proportion of which is centered in the uplands of Britain and maintained principally through the management activities of estates interested in grouse shooting.

Wise Use of Gamebird Resources

The concept of wise use of renewable resources has a number of interpretations and has yet to be defined in detail. The principle is that the management of the resource, game birds in this case, encourages harvesting at a sustainable level which provides benefits to the conservation of the species harvested and has additional benefits to human society and conservation of other species without exploitation.

One example of this wise use has been in the production of red-legged partridge shooting in the Ciudad Real province of Spain. After the revolution, few partridges existed, but the population was subsequently increased to a level where the population could be harvested through careful management and the cooperation of the local farmers. A harvest is produced which is of value to the whole rural community and provides the farmers with an economic interest in the care and habitat protection of the birds. Two brothers acquired shooting rights over the ground owned by more than 2000 individuals and produced a commercial partridge shoot bringing in nearly $3 million annually. Half of this money is

Table 1 Population size, harvest, and management recommendations for British game bird species

Species	*Number prior to harvest*	*Harvest*	*Status*	*Problem facing population*
Red grouse	1 500 000	30%	Stable after decline	Overgrazing, predation, and disease
Ptarmigan	35 000	5%	Stable after decline	Overgrazing and predation
Black grouse	40 000	5%	Stable after decline	Overgrazing and predation
Capercaillie	2 500	0%	Declining	Overgrazing and predation
Red-legged partridge	1 500 000	30%	Decline after increase	Release of chukar partridges
Grey partridge	1 250 000	15%	Declining steeply	Predation and intensification of cereals
Pheasant	20 000 000	50%	Slight decline	Predation and intensification of cereals
Woodcock	1 000 000	10%	Stable after increase	
Common snipe	720 000	5%	Stable after decline	Drainage
Woodpigeon	13 000 000	18%	Increase after decline	

distributed locally each year by paying a set amount per hectare to the local farmers, an amount for each partridge clutch hatched on the ground, and compensation for any damage. Local taxes are alsopaid which have resulted in local community benefits such as street lighting. The wise use in the development of partridge shooting has had clear economic benefits to the local community and, with a financial interest in each brood produced, encourages farmers to create habitat and reduce intensification of marginal agricultural land.

Scent, Taste, and the Significance of Hanging Game

Game birds have a characteristic taste which can be accentuated through the effects of hanging the dead birds for a period prior to eating. This is a habit developed in colder climates where the daily temperatures during the winter hunting season are usually less than 10 °C. At higher temperatures, and particularly in damp conditions, hanging game can lead to bacterial development and the rotting of game even when care is taken to avoid fly strike.

In Britain, the technique is to examine shot birds, often plucking the lower abdomen to determine that the guts have not been badly punctured. The birds are then hung in a cool game larder with air circulating, away from direct sunlight and so that the birds are not touching each other. The time the birds are left to hang depends on the age of the bird, size of the species, and the weather conditions. In August, young grouse are usually eaten immediately, whereas older birds are left for 2–4 days. Young partridges shot in September are also only hung for a short time so as not to lose the delicate taste of young birds. Young pheasants in October are usually hung for about 3 days, although by January when the temperature is cool the birds can be hung for 14–18 days. Usually birds are examined regularly until the skin below the breast bone turns bluish and produces a slight smell.

One of the characteristic of game birds is the strong scent and taste they emit. The scent is frequently so strong that a pointing dog can locate a bird 50 m downwind and on occasions even further, up to 200 m. Occasionally the scent can be so strong a human observer can smell a grouse at 10 m. One interesting characteristic for most game birds is that scent emission falls and almost ceases once incubation commences, so that trained pointing dogs have problems locating an incubating hen from distances greater than 0.5 m. This change in scent emission coincides with the birds ceasing to produce cecal feces, a brown glutinous feces produced daily from the ceca of the bird which contrasts with the more familiar fibrous feces. It seems likely that much of the scent and taste of the game birds may derive from the birds' ceca.

Most red grouse carry a burden of the cecal nematode *Trichostrongylus tenuis*, which burrows into the cecal wall and causes internal bleeding. Heavily parasitized birds frequently have a stronger taste and may emit more scent. This was demonstrated through a simple experiment. Trained pointing dogs and human observers were used to locate two groups of incubating females, one which had been treated with an anthelmintic to remove the parasitic nematode and a second control group that had received water. Observers and dogs searched for nests, the dogs searching by scent and the human observers by sight. The human observers found the treated birds in the proportions expected but the dogs located relatively few of the treated birds, presumably because they emitted less scent since they carried few parasites. The suggestion is that scent and taste of grouse may be determined by the birds' ceca.

The role of the ceca in digestion is not clear. They were originally thought to function primarily as a suitable environment for the fermentation of plant fiber, but such a hypothesis has been rejected. Ceca are now thought to separate soluble and suspended particulate food components from insoluble material. The adaptive significance is to maximize the rate of

Table 2 Nutrient composition (g per 100 g) of game birds; roasted meat compared with roasted chicken

	Water	*Protein*	*Fat*	*Total N*	*Fatty acids*		
					Saturated	*Mono-unsaturated*	*Poly-unsaturated*
Chicken	68.4	24.8	5.4	3.97	1.6	2.5	1.0
Grouse	61.6	31.3	5.3	5.00	1.2	0.7	3.1
Partridge	54.5	36.7	7.2	5.87	1.9	3.3	1.7
Pheasant	56.9	32.2	9.3	5.15	3.1	4.6	1.1
Pigeon	57.2	27.8	13.2	4.44	N[a]	N[a]	N[a]

[a]N = no data available.

Data abstracted from Holland B, Welch AA, Unwin ID *et al*. (eds) (1991) *McCance and Widdowson's Composition of Foods*, 5th edn. Cambridge: Royal Society of Chemistry and Ministry of Agriculture, Fisheries and Food.

Table 3 Micronutrient composition (per 100 g) of game birds; roasted meat compared with roasted chicken

	Elements (mg)										*Vitamins*													
	Na	*K*	*Ca*	*Mg*	*P*	*Fe*	*Cu*	*Zn*	*Cl*	*Mn*	*Retinol (μg)*	*Carotene (μg)*	*Vitamin D (μg)*	*Vitamin E (mg)*	*Thiamin (mg)*	*Riboflavin (mg)*	*Niacin (mg)*	*Trypt-ophan[b] (mg)*	*Vitamin B_6 (mg)*	*Vitamin B_{12} (μg)*	*Folate (μg)*	*Panto-thenate (mg)*	*Biotin (μg)*	*Vitamin C (mg)*
Chicken	81	310	9	24	210	0.8	0.12	1.5	87	0.03	Tr[a]	Tr	Tr	0.11	0.08	0.19	8.2	4.6	0.26	Tr	10	1.20	3	0
Grouse	96	470	30	41	340	7.6	0.10	1.5	130	0.06	N[a]	N	N	N	0.19	0.80	7.0	5.8	0.64	0.9	37	N	N	0
Partridge	100	410	46	36	310	7.7	N	N	99	N	N	N	N	N	N	N	N	6.9	N	N	N	N	N	0
Pheasant	100	410	49	35	310	8.4	0.10	1.3	110	0.02	N	N	N	N	0.02	0.29	9.2	6.0	0.57	2.5	20	N	N	0
Pigeon	110	410	16	34	400	19.4	0.33	1.7	99	0.05	N	N	N	Tr	0.27	N	7.0	5.2	0.82	N	8	N	N	0

[a]Tr, trace; N, no data available.

[b]Tryptophan values are divided by 60.

Data abstracted from Holland B, Welch AA, Unwin ID *et al*. (eds) (1991) *McCance and Widdowson's The Composition of Foods*, 5th edn. Cambridge: Royal Society of Chemistry and Ministry of Agriculture, Fisheries and Food.

throughput for birds feeding on a high-fiber, low-nutrition diet.

Nutrient Composition

Compared with other forms of poultry, relatively little information has been published for game birds. However, a comparison of the data available with comparable data for chicken shows that the nutrient compositions are not dissimilar (Table 2). In general, meat from game birds is somewhat richer in the sense that it has a higher protein and fat content than chicken. The increased fat content affects the texture of the meat and also its stability on storage. No significant body of data is available to compare the nutritional value of wild birds with those bred in captivity.

The limited information available on micronutrients (minerals and vitamins) shown in Table 3 would also suggest that there are no major differences between the meats of game birds in this respect. (*See* **Meat**: Nutritional Value.) Refer to individual nutrients.

See also: **Meat**: Nutritional Value

Further Reading

Hill DA and Robertson PA (1988) *The Pheasant: Ecology, Management and Conservation*. London: Collins.

Hudson PJ (1986) *Red Grouse: The Biology and Management of a Wild Game Bird*. Fordingbridge: Game Conservancy.

Hudson PJ (1992) *Grouse in Space and Time*. Hants: Game Conservancy.

Hudson PJ and Rands MRW (1988) *Ecology and Management of Gamebirds*. Oxford: Blackwell Scientific.

Johnson P and Wannenburgh A (1987) *The World of Shooting*. Lausanne: Photographex.

Potts GR (1986) *The Partridge. Pesticides, Predation and Conservation*. London: Collins.

Mammals

R J McCormick, University of Wyoming, Laramie, WY, USA

Introduction

Humans obtained food of animal origin exclusively from wild animals before the domestication of goats and sheep which probably began some 8000–9000 years ago. Today, only a small fraction of the meat consumed worldwide comes from nondomesticated or game animals. Utilization of game animals for food continues to varying degrees throughout the world. Hunting for sport, subsistence, or commercial purposes and herding and harvesting are traditional methods of obtaining game meat. Hunting and systematic harvesting of wild populations have been replaced in some locations with the intensive farming of wild ungulates for meat production. The compositional and organoleptic properties of game meat have resulted in increased interest in production of game animals for food. The variable conditions under which game animals are produced, harvested, and marketed often dictate special processing considerations.

Classification

Meat-producing large game species are largely ruminant ungulates which are members of the Cervidae or Bovoidea families (**Table 1**). Wild pig or boar (*Sus scrofa*) and rabbits (*Sylvilagus* spp.) hares (*Lepus* spp.), and other small mammals are also important sources of game meat. The Cervidae family encompasses the various deer species which are characterized by deciduous, solid antlers and generally the absence of a gallbladder. The superfamily Bovoidea includes all the true antelope, pronghorn, sheep, goats, and large wild bovids, such as bison and muskox. They possess hollow permanent horns, with the exception of pronghorn whose horns are deciduous, and the presence of a gallbladder is nearly universal. Indigenous populations of deer are notably absent from Africa and true antelope from North and South America.

Hunting – Sport and Commercial

Large numbers of indigenous and introduced game animals are harvested by hunters worldwide. Estimates

Table 1 Classification of selected ungulate game animals

Family Cervidae	*Family Bovoidea*
White-tailed deer (*Odocoileus virginanus*)	Bison (*Bison bison*)
Mule deer (*O. hemionus*)	Muskox (*Ouibos moschatus*)
Roe deer (*Capreolus capreolus*)	Bighorn sheep (*Ovis canadensis*)
Axis deer or chital (*Axis axis*)	Dall (*O. dalli*)
Fallow deer (*Dama dama*)	Mouflon sheep (*O. musimon*)
Red deer (*Cervus elaphus*)	Pronghorn (*Antilocapra americana*)
Wapiti or elk (*C. elaphus nelsoni*)	Mountain goat (*Oreamnos americanus*)
Sika deer (*C. nippon*)	Chamois (*Rupicapra rupicapra*)
Maral deer (*C. elephus maral*)	Saiga (*Saiga tartarica*)
Rusa deer (*C. timorensis*)	Blesbok (*Damaliscus dorcas*)
Muntjak deer (*Muntiacus muntjak*)	Springbok (*Antidorcas marsupialis*)
Musk deer (*Moschus moschiferus*)	Impala (*Aepyceros melampus*)
Moose (*Alces alces*)	Kudu (*Tragelaphus strepsiceros*)
Caribou or reindeer (*Rangifer tarandus*)	Eland (*Tarrotragus* spp.)

of the numbers of wild game animals harvested annually in North America, Central Europe, the former Soviet Union and Norway, Sweden, and Finland (Fennoscandinavia) are summarized in Table 2. The most prevalent large game species, both in distribution and numbers harvested, is the white-tailed deer. In some areas the actual harvest may be 50–100% greater due to illegal or unreported kills. The game animal harvest in North America is almost exclusively the product of sport hunting. In general, indigenous game animals are considered part of the public domain and private ownership of game animals and the sale of game meat obtained by sport hunting are largely prohibited by law. In Alaska, the western provinces of Canada and the Rocky Mountain region of the USA, game obtained through hunting comprises a significant fraction of the dietary meat.

In Central Europe and Scandinavian regions, with long traditions of sport hunting, significant numbers of ungulate game animals (primarily roe and fallow deer, wild boar, and moose) are harvested annually. A large fraction of the game meat is sold by the hunters to markets and restaurants. In other areas of the world, notably the former Soviet Union and the southern third of Africa, traditional sport and subsistence hunting are distinguished from commercial hunting. In these areas, increasingly larger numbers of game animals are harvested by relatively few hunters, often working in teams with the benefit of all-terrain vehicles, helicopters, or fixed-wing aircraft. In the former Soviet Union, saiga and wild reindeer comprise the bulk of the game animals harvested commerically. In Africa the antelope species, including blesbok, springbok, impala, kudu, and eland, account for the vast majority of wild game cropped for commercial purposes. Commercial hunting of red deer was practiced in New Zealand from the 1930s through the 1970s as part of the effort to control the expanding deer population. In one year (1970–71) an estimated 140 000 red deer were harvested; by 1985 approximately 1.5 million red deer carcasses obtained by hunting had been processed and exported from New Zealand.

A substantial fraction of the wild game harvested by sport hunting or commercially is comprised of small game, primarily rabbits and hares. Argentina exports 10 000–14 000 tonnes of meat from the brown hare (*Lepus capensis*) per year, making it the largest single exporter of game meat in the world. The annual harvest of small game animals in North America and Europe,

Table 2 Annual hunter harvest of game animals in selected locations

Country or region	*Species*	*Estimated annual harvest*
North America[a,b] (excluding Mexico and Central America)	White-tailed deer	3 000 000
	Mule deer	500 000
	Elk (wapiti)	104 000
	Moose	54 000
	Caribou	50 000
	Pronghorn	133 000
	Mountain sheep (dall; bighorn)	1300
	Mountain goat	1800
	Cottontail rabbit	25 000 000
Central Europe[c,d] (including Germany, Switzerland, Austria, and former Czech Republic)	Roe deer	1 221 235
	Fallow deer	17 250
	Red deer	116 300
	Mouflon	8650
	Chamois	46 600
	Wild boar	218 000
	Hare, rabbit, woodchuck	1 542 000
Fennoscandinavia[d,e]	Moose	205 600
	Fallow deer	3000
	Red deer	500
Former Soviet Union[d]	Moose	74 700
	Saiga (Cis-Caspian and Kazakhstan)	356 000
	Roe deer	150 000
	Reindeer (wild)	100 000
	Red deer (maral, Manchurian species)	10 000
	Wild boar	50 000

[a]Wildlife Management Institute, Washington, DC.
[b]*NRA Hunters Planning Guide and Directory* (1984) National Rifle Association of America.
[c]*Deutscher Jagdschutz-Verband Handbuch* (1989) Mainz: Verlag Dieter Hoffmann.
[d]Hudson RJ, Drew KR and Baskin LM (eds) (1989) *Wildlife Production Systems*. Cambridge: Cambridge University Press.
[e]Fennessy PF and Drew KR (eds) (1985) *Biology of Deer Production*. Wellington: Royal Society of New Zealand.

mostly rabbits and hares, is in the millions. Australia exports kangaroo meat, although the quantity has declined over recent years. Of the total game meat marketed internationally, it has been estimated that less than one-half the tonnage is from ungulate species.

Game Farming

Game farming has existed for centuries: reindeer husbandary in Fennoscandinavia and deer farming in China are ancient practices. Herds of bison have been maintained on farms and ranches in North America for most of this century and game ranching in southern Africa involves enclosure of wild species to control numbers and facilitate harvesting. In the thirty years intensive game-farming systems have developed which involve careful reproductive and nutritional management, and rather strict containment of a group of animals. Intensive farming systems exist throughout the world to raise, manage, and process game animals, primarily for meat production or velvet antlers, or both (Table 3).

The greatest number of farmed game animals, mainly red and fallow deer, exists in New Zealand. The farmed deer population is approaching 1 million animals with about 10% slaughtered for meat annually. China, the former Soviet Union, Australia, Korea, and Ireland also have large populations of farmed deer, although in Asia, deer production is mainly for velvet antler used for medicinal purposes, rather than meat. Whenever cervids are intensively farmed, however, velvet antler production is an important byproduct. A large bison population exists in North America with over 70 000 animals in the USA and 5000 in Canada on farms and ranches. Approximately 11 000 bison were slaughtered in 1985 for meat.

The intensive New Zealand deer-farming system evolved in barely two decades from the successful commerical hunting, processing, and export practices of prior years. Capture and containment of free-ranging red deer allowed selection and reproductive management, development of a breeding herd and controlled, but rapid, expansion of the farmed-deer population. The processing and export of deer meat have been facilitated in New Zealand by construction of specially designed deer slaughter and processing plants. Ante- and postmortem inspection of animals and carcasses, and fabrication and packaging of meat are accomplished in a manner similar to that used for domestic livestock. In many other areas where game animals are bred or managed for meat production, notably Africa, animals are often still harvested by shooting with processing of the carcasses carried out by mobile processing units. Absence of suitable handling and processing facilities for bison is a major factor constraining wider development of the bison-meat industry in North America. (*See* **Meat: Slaughter.**)

Processing Considerations

Game is generally subject to handling and treatment different from that of domestic livestock because of the conditions under which carcasses are obtained. Carcass weight and the composition of meat from some common large game and domestic species are shown in Table 4.

Carcass and Meat Characteristics

Most game animals (Cervidae and Bovoidea families) produce carcasses significantly lighter than beef cattle; exceptions are moose, bison, and eland. The most obvious differences between game and domestic species are the lower fat and metabolizable energy levels in cuts of meat from the game species. Differences in leanness between game and domestic species are even larger when whole carcass composition is determined. Carcass fat content of market lambs, for example, may exceed 30% but the carcass fat content of game animals is usually below 10% and often substantially lower. There can be wide seasonal,

Table 3 Major game-farming systems

Location	*Species*	*Approximate numbers managed*	*Main products/byproducts*
New Zealand[a]	Red, fallow deer, and wapiti	> 900 000	Meat and velvet antler
China	Sika and malu deer	300 000	Velvet antler
Former Soviet Republic	Sika and maral deer	90 000	Velvet antler
North America	Bison	88 000	Meat
	Wapiti	6000	Meat and velvet antler
Australia	Fallow, red, and rusa deer	40 000	Meat and velvet antler
Korea	Sika deer	52 000	Velvet antler
	Red deer	2000	
	Wapiti	3000	
UK	Red deer	10 000	Meat

[a]Drew K, personal communication. Other data from Hudson RJ, Drew KR and Baskin LM (eds) (1989) *Wildlife Production Systems*. Cambridge: Cambridge University Press.

Table 4 Composition of meat from selected game and domestic ruminants[a]

	Pronghorn[b,e]	Mule deer[c,e]	Elk (wapiti)[d,e]	Red deer[f]	Lamb[f]	Beef[e]
Carcass weight (kg)	28	43	165	69.4	13.6	275
Protein (g 100 g^{-1})	22.4	22.6	22.4	24.7	17.4	21.7
Fat (g 100 g^{-1})	2.5	2.7	2.0	3.3	18.2	5.0
Water (g 100 g^{-1})	70.0	66.7	73.4	70.8	63.6	
Ash (g 100 g^{-1})	1.2	1.0	0.9	1.4	0.9	
Metabolizable energy (kJ 100 g^{-1})	489	497	468	545	969	568

[a]Data for male game species reported.
[b]Field RA, Smith FC and Hepworth WG (1972) The pronghorn antelope carcass. *University of Wyoming Agriculture Experiment Station Bulletin* 575.
[c]Field RA, Smith FC and Hepworth WG (1973) The mule deer carcass. *University of Wyoming Agriculture Experiment Station Bulletin* 589.
[d]Field RA, Smith FC and Hepworth WG (1973) The elk carcass. *University of Wyoming Agriculture Experiment Station Bulletin* 594.
[e]Miller GJ, Field RA, Riley ML and Williams JC (1986) Lipids in wild ruminant animals and steers. *Journal of Food Quality* 9: 331–343.
[f]Drew KR and Seman DC (1987) The nutrient content of vension. *Proceedings of the Nutrition Society of New Zealand* 12: 49–55.

age-, and gender-related variations in the leanness of game animals. Males postrut are generally much leaner than prerut and older animals often contain more fat than younger.

Game meat is often much darker red in color than the meat from domestic species. The darker color may be due to increased muscle myoglobin concentration decreased intramuscular fat content (marbling), elevated muscle pH, or a combination of all these factors. Myoglobin concentration is a function of animal age and muscle fiber type and muscle from game animals generally possesses a higher proportion of fibers of the red, aerobic type geared to oxidative metabolism than does muscle from domestic species. Red, aerobic fibers contain more myoglobin and are smaller in diameter than less oxidative fiber types. In all species, myoglobin content in muscles increases with age.

Hunter-harvested game animals are particularly subject to excitement and stress which may result in depletion of muscle glycogen levels prior to death. Postmortem lactic acid production in the muscle is decreased, resulting in a higher than normal ultimate muscle pH (5.4–5.7) and a dark appearance. Muscle pH of hunter-harvested pronghorn, mule deer, and white-tail deer when determined ranged from somewhat elevated (5.8–6.0) to normal (5.4–5.6). Both hunter-harvested elk (wapiti) and farmed red deer showed normal ultimate meat pH (< 5.7). (*See* **Exercise**: Muscle.)

Texture and tenderness of meat are primary quality characteristics and are related to muscle connective tissue content and maturity, muscle fiber diameter, and degree of postmortem muscle proteolysis. There can be wide variations in the tenderness of game meat, due primarily to the wide range of ages at which individuals are harvested and to seasonal fluctuations in growth rate. Increasing age results in muscle connective tissue maturation, increased collagen cross-linking, and decreased meat tenderness. Variable growth rate alters collagen synthesis rate and may similarly affect muscle collagen quality and meat tenderness, especially in males during and after the rut. Variable carcass chilling rates and aging temperatures also contribute to differences in tenderness. (*See* **Meat**: Preservation.)

The smaller ungulate game species, particularly fallow deer, produce a fine-textured meat that is highly valued. The fine texture of meat from the smaller game species is due in part to decreased muscle fiber diameter. Both pronghorn and deer can develop an objectionable, mushy texture with increased postmortem hanging time. A comparison of endogenous muscle protease activity and postmortem myofibrillar protein degradation, however, revealed no differences between pronghorn and beef muscle.

Aging

The major differences between processing of conventionally slaughtered domestic livestock and game are the recommendations for reduced length of postmortem refrigerated storage time (aging) for game prior to freezing or consumption. Aging of carcasses with little or no fat results in excessive carcass dehydration, weight loss, and discoloration of the lean meat. The vast majority of game carcasses and meat are not subject to the rigid inspection and slaughter standards applied to conventionally slaughtered animals. With game, high levels of microbial contamination, slow carcass chilling, and elevated muscle pH often result in rapid microbial proliferation. Higher bacterial counts in ground meat from deer and pronghorn than from elk carcasses occur because of the increased carcass surface area relative to total meat in the smaller animals. Intensive game-farming systems and the slaughter and processing of game in specially designed facilities reduce or eliminate many of the difficulties associated with processing and handling hunter-harvested animals. Bacterial counts and refrigerated shelf-life of meat from farmed New Zealand red deer are comparable to values obtained for conventionally processed domestic livestock. Meat from game carcasses often requires addition of fat to enhance palatability.

Table 5 Lipid composition and cholesterol content of meat from selected game species and beef[a,b]

	Pronghorn antelope	*Mule deer*	*Elk*	*Bison*	*Grain-fed beef*	*Grass-fed beef*
Fatty acid (mg per 100 g)						
Saturated[c]	875	972	664	421	2028	933
Stearic	441	401	172	197	651	327
Myristic, palmitic	434	571	492	224	1377	606
Monounsaturated[d]	582	732	508	444	2144	754
Polyunsaturated[e]	530	463	399	182	291	191
ω-6[f]	442	359	343	156	275	139
ω-3[g]	88	104	56	26	16	52
Cholesterol (mg 100 g^{-1})	52	54	48	45	48	49

[a]Uncooked, lean only.
[b]Miller GJ, Field RA, Riley ML and Williams JC (1986) Lipids in wild ruminant animals and steers. *Journal of Food Quality* 9: 331–343.
[c]Myristic, palmitic and stearic acids.
[d]Palmitoleic and oleic acids.
[e]Linoleic, arachidonic and linolenic acids.
[f]Linoleic and arachidonic acids.
[g]Linolenic acid.

Species Identification

Substitution of less valuable meat for game or illegal trade in game species meat poses worldwide concern. Methods for the identification of raw (uncooked) game meat based on protein differences detected by electrophoresis are well established for most species. Heating or cooking denatures and inactivates meat proteins and enzymes, greatly reducing solubility and extractability. Specialized techniques for the identification of game species in processed and cooked meat products based on protein or immunological variations have, therefore, been developed in recent years.

Dietary Significance

Medical and nutritional dietary recommendations currently emphasize decreased overall dietary fat intake and lower saturated fat consumption. There is generally significantly less total lipid associated with meat from game animals than from domestic livestock (Table 4). Substitution of game for meat from domestic livestock in the diet often provides consumers with a means of decreasing dietary energy supplied by fat. At the same time, nutrient intake associated with red meat consumption, including B-vitamins, minerals, and high-quality protein, is maintained. Refer to individual nutrients.

The fatty acid composition and cholesterol content from game animal and beef lean muscle are shown in Table 5. Species differences exist in amounts and proportions of saturated, monounsaturated, and polyunsaturated fatty acids in meat. Decreased dietary intake of saturated fatty acids, particularly myristic and palmitic acids, are associated with lower blood serum cholesterol and, therefore, diminished risk of cardiovascular disease development. Consumption of ω-6 and ω-3 polyunsaturated fatty acids may lower blood serum cholesterol and protect against coronary heart disease development. Game meat is lower in saturated and higher in polyunsaturated fatty acids than grain-fed beef. There are no wide species-related difference in muscle cholesterol content because cholesterol is present in both fat and lean tissues as a component of cellular membranes. Meat from small species is somewhat higher in cholesterol partly because the decreased muscle fiber (cell) diameter in these species favors increased cell membrane mass relative to total muscle mass. (*See* **Fatty Acids**: Properties.)

See also: **Electrophoresis**; **Fatty Acids**: Properties

Further Reading

Chapman JA and Feldhammer GA (eds) (1982) *Wild Mammals of North America: Biology, Management and Economics*. Baltimore: Johns Hopkins University Press.

Fennessy PF and Drew KR (eds) (1985) *Biology of Deer Production*. Wellington: Royal Society of New Zealand.

Hudson RJ, Drew KR and Baskin LM (eds) (1989) *Wildlife Production Systems*. Cambridge: Cambridge University Press.

McCormick RJ, Moore TD, Field RA and Hepworth WG (1988) Identification of deer meat cooked products. *Wildlife Society Bulletin* 16: 433–436.

Patterson RLS (ed.) (1985) *Biochemical Identification of Meat Species*. London: Elsevier.

GARLIC

H Amagase and B L Petesch, Wakunaga of America Co Ltd, Mission Viejo, CA, USA

Introduction

Garlic is a very well-researched herb with centuries of culinary and medicinal uses. This powerful herb appears in the historical writings of numerous cultures. This article touches on the cultivation, history, chemistry, pharmacology, and safety of garlic.

Cultivation

Garlic, or *Allium sativum*, is a member of the lily family. It is a perennial, which is usually grown as an annual. A typical plant consists of one large bulb made up of numerous small cloves. The word 'clove' came from the word 'cleave' which means both to cling together and to divide along natural lines. Garlic is typically planted in autumn and harvested about 8 months later in the summer. Various cultivars can grow under an array of climates, temperatures, rainfall, and soil types.

Historical Effects

Garlic has been utilized as a medicinal for centuries. References to this herb have been recorded in the Old Testament of the Bible, the *Codex Ebers*, an ancient Egyptian medical volume, *Shin ching*, the Chinese *Book of Songs*, and in medical texts of Sumeria, considered to be the world's oldest pharmacopeia. From pharaohs to pharmacies, it is known as the king of the herbs and with more than 2000 scientific publications in the last century alone, it has rightly earned and maintained such a title. A plethora of recent research has confirmed many of the empirical usages of garlic for medicinal purposes. Some of the historical usages of garlic have included remedies for heart problems, headaches, coughs and colds, snake and insect bites, asthma, arthritis, hair growth, worms, respiratory ailments, digestive disturbances, toothaches, and tumors.

Garlic has been used as a poultice to treat respiratory conditions. Since many of its sulfur compounds are modified in the intestinal tract, this was a means of introducing salubrious compounds into the blood stream, as confirmed by garlic breath noted after topical administration. However, careless administration without an adequate vehicle (thick layer of Vaseline) added to the feet prior to administration has led to severe secondary chemical burns.

Tales of garlic as a cure for the great plagues were passed along and written about for several centuries. It was recommended more for protection rather than for actual treatment of the disease. Garlic was coined as 'Russian penicillin' due to its potent antibiotic effects. During both world wars it was used to prevent wounds from growing gangrenous. However, since it can also induce chemical burns, and more practical and effective antibiotics have since been developed, this usage has mostly been abandoned.

Historical interest in garlic has led to extensive research to test its pharmacological effects, resulting in more than 1000 publications over the past decade alone.

Garlic Chemistry

Garlic is quite a complex herb. It is well known with its characteristic odor. Garlic has many intrinsic specific enzymes and chemical constituents that cause a specific odor by crushing the garlic cloves. This odor comes from allicin and decomposes to other compounds instantly. Allicin thus appears to be a transient compound to many of the other sulfur compounds in garlic. It has shown potent antimicrobial effects against even pathogenic and resistant microorganisms in testtubes. However, it is extremely unstable and not bioavailable, absorbed from the intestinal tract, so its testtube effects may not be demonstrated in the body upon consumption. Further, its toxicity is rather high (LD_{50} in $mg\,kg^{-1}$ body weight in mice is 309 in males and 363 in females). It is converted mainly into diallyl disulfide and allylmercaptan in the body. Diallyl disulfide, an oil-soluble compound, has shown remarkable anticancer potential in an array of studies; however, its toxicity is even higher (LD_{50} in $mg\,kg^{-1}$ body weight in mice is 145 in males and 130 in females) than allicin. Chemically garlic is best known for its characteristic sulfur compounds, possessing more than 200 kinds, many of which have scientifically documented pharmacological effects. The water-soluble sulfur compounds derived from γ-glutamyl cysteine, such as *S*-allyl cysteine (SAC), have shown diversified pharmacological activities with little or no toxicity (LD_{50} of SAC in $mg\,kg^{-1}$ body weight in mice is 8890 in males and 9390 in females). Further, they are much more stable and have demonstrated bioavailability in both preclinical and clinical

studies. Garlic also contains nonsulfur compounds with pharmacological effects. A protein fraction, identified in aged garlic extract, has been shown to enhance immune function, nerve growth, to combat cancer, and enhance the growth of friendly bacteria. Steroidal glycosides or saponins in garlic have demonstrated cardioprotective and antifungal activities. And various minerals, such as selenium, germaninum, calcium, copper, iron, and vitamins such as A, B_1, and C, are found in this herb. In fact, garlic is one of the richest sources of selenium. Though garlic is not a rich source of protein, it does contain an array of amino acids.

Pharmacology

Cardiovascular Effects

Extensive research has suggested that garlic intake may reduce various cardiovascular disease risk factors. Randomized, placebo-controlled, double-blind clinical studies have shown modest reductions in elevated blood lipids or fats, such as serum cholesterol, low-density lipoprotein (LDL) cholesterol and triglycerides, and elevated blood pressure. Though the effects are not as pronounced as those of medications designed specifically for such ailments, the side-effects are also nominal or nonexistent, unlike their drug counterparts. Further, garlic is more versatile in its effects on the heart than drugs designed merely to lower blood pressure or lower blood lipids. Garlic has been shown not only to reduce elevated blood lipids, but some studies have even shown an improvement in high-density lipoprotein (HDL)-cholesterol. However, some recent studies utilizing garlic powders and oil have shown no effect on cholesterol, so that standardization of garlic products for effective active ingredients appears to be relevant. Of greater concern than elevated lipids may be their oxidation. Recent work has shown that extracted garlic may prevent the oxidation of blood fats, reducing their ability to adhere to and damage the veins. Garlic has also been shown to reduce the build-up of plaque in the arteries induced by a high-cholesterol diet, as shown in atherogenic models.

Garlic has shown blood-thinning effects, such as inhibiting platelet aggregation and adhesion. These effects have been demonstrated in testtubes and in humans. Thermography has confirmed the ability of garlic to enhance peripheral circulation in humans within 90 min of its intake.

Garlic has also recently been reported to suppress the elevated level of homocysteine, a strong indicator of atherosclerosis, further heart attack, stroke or any kind of stenosis. Thus, garlic has many beneficial effects on the cardiovascular system to reduce multiple risk factors of this disease.

Immune-Enhancing Effects

Garlic has been shown to mitigate infectious diseases through enhancement of the immune system. It has been found to enhance various immune factors such as the phagocyte (cell-killing) activity of macrophages, T-lymphocyte activity, natural killer cell activity, and antibody generation. It has also demonstrated antiviral and antifungal activities. Further, it has been shown to modify, both directly and indirectly, the function of immune cells which play a leading role in allergic cascade reactions, including inflammation. Garlic has also been shown to improve age-related deterioration of the immune response.

Clinical studies have shown that both raw garlic and especially aged garlic extract significantly enhance natural killer cell activity in normal subjects. In a clinical study done on acquired immunodeficiency syndrome (AIDS) patients, who have poor immune function, aged garlic extract was shown to bring natural killer cell activity to normal within only 6 weeks of oral intake. Subjects also experienced improvement in subjective symptoms such as candidiasis and pansinusitus.

Antimicrobial Effects

In testtubes, garlic and some of its constituents have shown potent antimicrobial effects against even pathogenic and resistant microorganisms. Fresh raw garlic and its constituents have been shown to kill *Escherichia coli, Salmonella enteritidis, S. typhimurium, Staphylococcus aureus* and *S. vulgaris, Shigella dysenteriae, S. flexneri, S. sonnei, Mycobacterium tuberculosis, Giardia lamblia, Helicobacter pylori, Entamoeba histolytica*, and other organisms. However, few clinical studies have been done in regards to such antiinfectious effects and some of the effective compounds may not be absorbed from the intestinal tract. Quality research still needs to be done in this area to confirm an effect in humans.

Garlic has demonstrated antifungal effects against *Candida albicans, Aspergillus niger, Cryptococcus neoformans, Coccidiodes immitis, Epidermophyton, Trichophyton*, and other fungi. *Candida albicans* is responsible for secondary yeast infections that often occur after intake of antibiotics. Aged garlic extract was shown to reduce candidiasis, or systemic *Candida albicans* infections, in AIDS patients.

Antiviral effects have also been noted against herpes simplex and influenza. Orally and topically, aged garlic extract has been utilized for genital herpes.

Anticancer Effects

A great deal of epidemiological, chemopreventive, and other research has been done in the area of cancer. Various epidemiological studies conducted in China, Italy, and the USA have shown an inverse relationship between garlic intake and stomach and colon cancer. Various cell culture studies have suggested that compounds in garlic may inhibit the growth of human breast cancer cells, several human melanoma cells, and neuroblastoma cells. Aged garlic extract has been shown to inhibit the growth of breast cancer, bladder cancer, skin cancer, colon cancer, the development of esophageal tumors, and stomach and lung tumors in animal models. The anticarcinogenic effects of garlic are believed to be due to its ability to reduce the rate of activation of chemical carcinogens, that is, preventing the conversion of inactive cancer-causing agents into active ones; its ability to enhance the rate of chemical carcinogen excretion, to protect DNA from carcinogen-DNA adduct formation, to enhance enzymes that detoxify carcinogens, and to suppress tumor cell growth through stimulation of immunoresponder cells. Garlic compounds, injected into cancerous tumors, have resulted in immune cells more aggressively attacking the tumor. Oral intake of garlic has been shown to enhance the activity of various immune cells.

Studies have also suggested that garlic may be a promising adjuvant to cancer therapy. Aged garlic extract was shown to reduce side-effects such as fatigue and anorexia in head-and-neck tumor patients on radio- and chemotherapy and to reduce the cardiotoxicity of the potent anticancer drug doxorubicin/Adriamycin. Aged garlic extract has also been shown to reduce the intestinal damage caused by the anticancer drugs 5-fluorouracil and methotrexate, suggesting that garlic should be included in an overall strategy to reduce/prevent cancer.

Liver Protection

Aged garlic extract and some of its water-soluble compounds, including *S*-allyl cysteine and *S*-allyl mercaptocysteine, have shown significant liver-protective effects against an array of liver toxins, including carbon tetrachloride, aflatoxin B_1, acetaminophen (paracetamol), and others. The protective effects appear to be due to the enhancement of detoxifying phase II enzymes, such as the glutathione-related enzymes. Enhancement of glutathione level by aged garlic extract has been demonstrated in both testtubes and humans.

Other Potential Effects

Recent research has demonstrated antiaging and neurovascular support from garlic and its constituents. Compounds in garlic have been shown to enhance memory in genetically age-accelerated rodents with learning deficits (used for Alzheimer models). Garlic has also been shown to prolong their life span. Enhanced nerve growth and branching have been demonstrated in culture. Garlic, specifically aged garlic extract, has been shown to protect nerve cells from oxidative damage caused by the accumulated neurotoxin beta-amyloid (Aß), which is a hallmark for Alzheimer's disease. One of the water-soluble sulfur compounds, *S*-allyl cysteine, is found in the brain after the consumption of garlic. This may suggest that bioavailable compounds such as *S*-allyl cysteine may work at the target sites of the body and can pass the blood–brain barrier. Moderate but significant ergogenic or energy-enhancing and antistress effects have been noted in athletes following intake of garlic. Garlic has also been shown to reduce stress-induced suppression of immune function and adrenal hypertrophy.

Dosages

Various forms of garlic have shown health-enhancing effects – from fresh garlic, cooked garlic, garlic oils, garlic powders, to aged garlic extract. More than 250 studies have been conducted on aged garlic extract. Some clinical studies have been done on garlic powders that have shown both positive and no effects. Thus, the efficacy of garlic powders, especially in regard to cholesterol reduction, is controversial. The following are dosages used in various studies:

Raw garlic	1–2 cloves (3000–6000 mg)
Garlic oil	2–5 mg
Garlic powder	600–900 mg
Aged garlic extract	4 ml
	1000–7200 mg (as dried form)

Safety

For the most part, garlic is quite a safe herb when taken in reasonable amounts by appropriate means. However, taking garlic as a preventive involves long-term supplementation, which leads to the necessity of consideration of the issue of toxicity. One of the greater concerns associated with long-term garlic supplementation is the general attitude of the public that if a little of something is good for the body, a larger quantity must be even better.

Although garlic has been safely used in cooking as a popular condiment/flavoring and used traditionally for medicinal purposes, it is commonly known that excessive consumption of garlic can cause burning

sensations and diarrhea. Excess intake of raw garlic or garlic powders can cause the following adverse effects/toxicity: stomach disorders and diarrhea; decrease in serum protein and calcium; anemia; bronchial asthma (via inhalation, seen from factory workers); contact dermatitis (skin burns); and inhibition of spermatogenesis. It appears that allicin and some of its degraded compounds may be responsible for some of the harmful effects of garlic, such as anemia and skin burns. The oil-soluble sulfur compounds in garlic have been shown to be much more toxic than water-soluble compounds.

Cooking can reduce the toxicity of garlic but also eliminates its antimicrobial effects. Aging garlic by long-term extraction, by reducing oil-soluble compounds and increasing water-soluble compounds, greatly reduces the toxicity of garlic so that even large amounts can be and have been consumed safely. Further, aging garlic improves its pharmacological activity.

Since garlic has shown potent antiplatelet or blood-thinning activity, it is typically contraindicated with Coumadin (sodium warfarin) or other anticoagulant medications. However, a clinical study has shown that aged garlic extract does not worsen the side-effects of Coumadin in patients. Thus, unlike fresh garlic or other forms of garlic, such as dehydrated garlic powder, aged garlic extract does not have a contraindication with medication.

It may not be suitable for pregnant and nursing women to consume a large amount of garlic products, except aged garlic extract that is approved for those women in Japan as a medication for nourishment pregnant and nursing women.

Conclusions

Garlic is a unique herb loved the world over for centuries. The odorous compounds were formerly believed to be responsible for its beneficial effects. However, recently numerous nonodorous compounds in garlic have been discovered which have potent pharmacological activity. Recent clinical studies support many of the historical usages of garlic, suggesting that its long-term intake may reduce many risk factors for various diseases.

See also: **Aflatoxins**; **Cancer**: Carcinogens in the Food Chain; Diet in Cancer Prevention; **Cholesterol**: Factors Determining Blood Cholesterol Levels; Role of Cholesterol in Heart Disease; **Coronary Heart Disease**: Antioxidant Status; **Liver**: Nutritional Management of Liver and Biliary Disorders

Further Reading

Bradley PR (ed.) (1992) *British Herbal Compendium. A Handbook of Scientific Information on Widely Used Plant Drugs*, vol. 1, pp. 105–108. Bournemouth, UK: British Herbal Medicine Association, Scientific Committee.

Ho C-T and Zheng QY (eds) (2001) *Quality Management of Nutraceuticals*. ACS Symposium Series 803. Washington, DC: American Chemical Society.

Ho CT, Osawa T, Huang MT and Rosen RT (eds) (1994) *Food Phytochemicals for Cancer Prevention II. Teas, Spices, and Herbs*. ACS Symposium Series 547. Washington, DC: American Chemical Society.

Huang MT, Osawa T, Ho CT and Rosen RT (eds) *Food Phytochemicals for Cancer Prevention I. Fruits and Vegetables*. ACS Symposium Series 546. Washington, DC: American Chemical Society.

Lachance PA (1997) *Nutraceuticals: Designer Foods III–Garlic, Soy and Licorice*. Trumball, Connecticut: Food and Nutrition Press.

Lau B (1997) *Garlic and You: The Modern Medicine*. Vancouver, Canada: Apple.

Mathew B (1996) *A Review of Allium Section Allium*. Kew, UK: Royal Botanic Gardens.

Moyers S (1996) *Garlic in Health, History, and World Cuisine*. St. Petersburg, Florida: Suncoast Press.

Saito H (2000) *Garlic Science*. Tokyo, Japan: Asakura Shoten.

Shahidi F and Ho CT (2000) *Phytochemicals and Phytopharmaceuticals*. Champaign, Illinois: AOCS Press.

Woodward P (1996) *Garlic and Friends*. Melbourne, Australia: Hyland House.

Gas Chromatography *See* **Chromatography**: Principles; Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; Supercritical Fluid Chromatography; Combined Chromatography and Mass Spectrometry

Geese *See* **Poultry**: Chicken; Ducks and Geese; Turkey

GELATIN

P Harris, V Normand and I T Norton, Unilever Research Laboratory, Sharnbrook, Bedford, UK

Introduction

Gelatin is a nutritious protein containing essential amino acids and is derived from collagen present in the skin and bones of animals. The main food use for gelatin is as a gelling agent for ready-to-eat products residing in the refrigerator (e.g., mousse, trifles, etc.). Other uses are as a dry ingredient (crystal table jellies) or in a highly concentrated gel form (fruit pastilles, marshmallow). More gelatin is sold to the food industry than any other gelling agent; it is relatively cheap to produce in quantity, and there is a ready supply of suitable raw material. This article will review the source, manufacture, and properties of gelatin.

Collagen

Occurrence

Collagen is an important component of skin and bones, and constitutes as much as 30% of total human protein, with similar proportions being found in many other animals. It is widespread in both vertebrates and invertebrates, differing somewhat in amino acid composition, but providing the same function-strength and support for the tissues and organs of the animal concerned.

Structure

The collagen molecule exists as a triple helix comprising three discrete α-chains stabilized by hydrogen bonding. It is characterized by high percentages of glycine (33%), and the imino acids proline and hydroxyproline (22%).

At least 10 collagen types have been identified and another four tentatively recorded on the basis of differences in the three α-chains; the best known are types I and III, from which all commercial gelatins are produced.

The molecular weight of collagen is around 285 kDa, and that of the α-chains 95 kDa.

In most tissues, collagen is arranged in bundles of four or five molecules to form structures known as fibrils. These are further associated with other fibrils nearby to form further bundles of larger diameter. These fibrils are stabilized by the formation of intermolecular cross-links formed from lysine or hydroxylysine residues; the degree of this cross-linking increases with maturity of the animal.

Gelatin

Manufacture

Three types of gelatin are available commercially – ossein (from bones), bovine hide, and pigskin. Gelatin is extracted from collagen by a controlled acid or alkali hydrolysis. The source, age, and type of collagen each influence the properties of the gelatins derived from them. The following steps make up the manufacturing process:

Washing Initial washing of the starting raw material removes impurities; this step also includes degreasing and demineralization of bones to produce ossein.

Pretreatment The pretreatment process is designed to convert collagen into a form suitable for extraction. To achieve this, a sufficient number of covalent cross-links in the collagen must be broken in order to enable the release of free α-chains. The process is also designed to remove other naturally occurring organic substances such as proteoglycan, blood, mucins, and sugars. Pretreatments are optimized by each manufacturer to give the required physical and chemical properties to the gelatins that are produced. The degree of collagen cross-linking governs the pretreatment process required for gelatin manufacture. In young fetal tissues, the concentration of cross-links is at a minimum, and much of the collagen is soluble in mild acid or salt buffer. As collagen matures, the proportion of soluble fraction (procollagen) reduces markedly as cross-linking progresses. To convert insoluble collagen into soluble gelatine three processes are in current use:

1. Acid pretreatments leading to acid process type A gelatins.
2. Alkaline or basic pretreatments leading to alkaline process or type B gelatins.
3. Acid pretreatment processes, which are less aggressive than those using alkali, are applied to pigskin and fresh ossein from young animals. Alkaline pretreatment processes are normally applied to bovine hide and ossein.

Extraction The extraction process is designed to obtain the maximum yield by optimizing the balance between pH, temperature, and the extraction time. In practice, gelatin is obtained from the raw material in three or four separate extractions, each at an increasing temperature. Typical values are 55 °C for the first extraction, 60 °C for the second, 70 °C for the third, and 80–90 °C for the final extraction, each giving gelatins of decreasing gel strength, viscosity, and increasing color. The simplest result of the denaturation is solubilization of the α-chain, however, the denaturation process also leads to secondary products due to random cleavage of single collagen molecules (small units). The other important point to remember is that hydrolysis is sometimes incomplete, leaving covalent bonds between molecules. When two or three α-chains are linked together they are commonly called β- and γ-chains, and when more then three chains are covalently linked, the macromolecules form what are known as microgels. The molecular distribution produced from any extraction can be divided into three broad molecular weight classes: small units, $\alpha + \beta + \gamma$, and microgels. As the extraction number increases, the fraction of $\alpha + \beta + \gamma$ chains decreases, the fraction of small units increases, and the fraction of microgels decreases.

Purification Following extraction, the gelatins are filtered to remove suspended insolubles such as fat or unextracted collagen fibers. The gelatin is further purified by deionization which both removes inorganic salts left from pretreatment and also adjusts the pH to that suitable for sale. Most commercial gelatins are sold in the pH range 5.0–5.8.

Concentration The final stage is evaporation, sterilization, and drying. These are performed as quickly as possible to minimize loss of properties; after this the gelatins are subjected to laboratory testing for their physical and bacteriological characteristics.

Structure

The primary structure of gelatin closely resembles the parent collagen. Small differences are due to raw material sources together with pretreatment and extraction procedures.

These can be summarized as: (1) Partial removal of amide groups of asparagine and glutamine resulting in an increase in the aspartic and glutamic acid content; and (2) conversion of arginine to ornithine during prolonged alkali pretreatment processes. Typical amino acid composition data of gelatins are shown in Table 1.

Table 1 Average amino acid distribution for the three α-chains found in the collagen precursors of gelatin

Residue	*Residues per 1000 residues*
3-Hydroxyproline	1
4-Hydroxyproline	108
Proline	115
Lysine	25
Hydroxylysine	9
Glycine	340
Cysteine	0
Serine	34
Alanine	104
Histidine	7
Valine	23
Methionine	7
Isoleucine	11
Leucine	24
Arginine	49
Phenylalanine	11
Aspartic acid	42
Threonine	16
Glutamic acid	71
Tyrosine	3

Properties

Commercial gelatins consist of gelatin protein in a high state of purity; the nonproteinaceous matter present is mainly mineral ash and residual moisture. The only additive sometimes incorporated during manufacture is sulfur dioxide, used as a color inhibitor during extraction and evaporation and not as a preservative. There is also a small proportion of carbohydrate (1–1.5%) in the form of glucose and galactose bound to gelatine at hydroxylysine residues.

The main gelatin properties of interest to food users are:

1. Bloom value, which is a function of the gel strength.
2. Viscosity, which gives a measure of the solution properties.
3. Surface activity or polyelectrolyte properties.

These main properties will be discussed in more detail later.

Other properties which are important are as follows.

Particle size and solubility Gelatin, although insoluble in cold water and other liquids such as milk, sugar solution, dilute food acids, etc., will swell and absorb up to 10 times its weight in water, the rate of which is dependent on particle size. Swelling characteristics in cold water are determined by temperature and salt or sugar content of the liquid, all of which can affect the rate of water uptake.

Isoelectric and isoionic points The isoelectric point is defined as the pH at which no net migration takes place in an electric field, while the isoionic point is defined as the pH at which there is no net charge on

the molecule. In a deionized solution, the isoelectric and isoionic points are for most purposes identical.

The isoelectric point of gelatine is dependent on the type of pretreatment applied during manufacture. Type A or acid-processed gelatins have isoelectric points that can vary from 6.5 to 9.0. Acid ossein gelatins are at the lower end of the range with isoelectric point values of 6.5–7.5, while acid pigskin is more likely to have values from 7.5 to 9.0.

Many physical properties of gelatine have either a minimum or a maximum at the isoelectric point. This means that acid- and alkali-derived gelatines can sometimes perform quite differently in the same system.

Solution Properties

Typical commercial gelatins tend not to find use as thickening agents due to their relatively low specific viscosity ($\eta_s < 60$ m.ps., @ 6.66) as there are more efficient polysaccharides available for this purpose. Instead, gelatins in the liquid phase can be used for their surface-active properties, either as a stabilizer and emulsifier or as a polyelectrolyte.

Higher-viscosity gelatins will give higher melting temperatures and somewhat quicker setting times, while those of lower viscosity can be prepared in a much higher concentration without causing problems due to tailing when depositing in molds.

A solution of gelatin in water will possess a viscosity or thickening power proportional to its concentration, pH, ionic strength, and the specific viscosity of the gelatin itself. The relationship of viscosity with concentration is not directly proportional but is approximately logarithmic. There is a straight-line dependence for log viscosity with the reciprocal of absolute temperature. The viscosity is always at a minimum at the isoelectric point, and increases as the overall charge on the molecule is increased.

Gel Properties

Preparation of a gel The gelatin is first dispersed in cold water (never the reverse) and allowed to swell. The dispersion is then heated to 50–60 °C and the resultant solution allowed to cool without agitation. A gel is formed at temperatures below about 25 °C depending on the type of gelatin used and its concentration. The critical concentration for a gelatin gel to form is approximately 1% w/w.

Mechanism of gel formation When a solution of gelatin is hot (> 65 °C), it exists as near to random coil chains. As the solution is cooled, the chains lose some of their mobility and start to order locally as segments: on further cooling, these ordered segments associate into triple helices. The site where this association takes place is referred to as cross-link or junction zone. The structure of these junction zones is similar to that of the native collagen. When sufficient junction zones are formed, a three-dimensional network or gel is produced. The gelation kinetics can be divided into two phases:

1. During the first hour, the gel strength increases rapidly. The reinforcement of the network is believed to be mainly due to formation of new cross-links.
2. From 1 h to several hundred hours, the gel strength increase slows down and is linear with the logarithm of time. The reinforcement of the network is believed to be due to the extension of existing cross-links rather than the formation of new ones.

The exact nature of the junction zones is still under investigation, although it has been shown that (GLY-PRO-HYP) triplets are involved. Hydrogen bonding takes place between the amino acids and nearby C–H groups, e.g., on glycine found on adjacent chains. (It is well known that hydrogen bond breakers, such as thiocyanate or urea, can prevent gelation.)

The equilibrium of junction zone formation and breakage moves to create a dynamic gel structure as the temperature of the solution approaches the setting point. At this stage, the larger branched molecules (microgels) begin to grow via junction zone formation with other smaller α- + β- + γ-chains. The condensation reaction corresponds to a second-order reaction, even if triple helices are formed. This has been visualized microscopically, which suggests that an α-chain loops back on to itself and then reacts with a second α-chain to form a triple helix involving just two chains and the junction zone has four free ends. The current thinking is that the elasticity of gelatin gels comes from the relatively free rotation of the segments before these junction zones.

Measurement of Gel Strength

The gel strength of gelatin is nearly always quoted commercially as the Bloom strength. The Bloom value of gelatin is defined as the weight required to push a cylindrical plunger of 12.7 mm diameter 4 mm into a previously prepared gel of 6.66 w/w concentration matured at 10 °C for 16–18 h. Gel strength determinations not performed under these strict conditions cannot be quoted as Bloom strength.

To determine the gel strength, one of the following three instruments can be used:

1. The lead shot gelometer, also known as the Bloom gelometer.
2. The Boucher gelometer.
3. The Stevens-LFRA texture analyzer.

Large Deformation and Failure Behavior

As the requirements of users of gelatin have become more complex, the gel properties are now commonly determined in other ways. Gelatin gels are viscoelastic and incompressible materials (Poisson's ratio close to 0.5). When submitted to large strains, they always present a strain-hardening behavior before failure. The strain at failure is always greater than 50% of strain and can reach 200% of strain in appropriate conditions of concentration and deformation rate. These properties can be linked to the dynamic nature of the cross-links which are melting and reforming under stress, leading to alignment of the elastically active chains. As might be expected, the strain, stress at failure, and the Young's modulus increase when the concentration increases.

Factors Affecting the Properties of Gelatine Gels

Concentration and temperature The exact relationship between concentration and gel strength depends on the type and origin of the gelatin itself, but using gelatins commonly found in the food industry, i.e., 100–250 Bloom, the relationship becomes $(C_2)^n \times (B_1) = (C_1)^n \times (B_2)$ where C is the concentration of the gel concerned; B, the Bloom of the gel concerned; n is equal to 1.7 for high-Bloom gelatins and 1.8–1.9 for lower-strength gelatins of 150–100 Bloom. More precisely, for well-characterized gelatins, a correspondence between the elastic modulus, the measurement temperature, and the concentration in α- + β- + γ-chains has been introduced at any time of the gelation process (master curve). Two equations are necessary for this model:

$$(C_2)^2 \times (T_1) \times (G_1) = (C_1)^2 \times (T_2) \times (G_2)$$

$$\text{and } (C_1) \times (T_2) \times (\tau_1) = (C_2) \times (T_1) \times (\tau_2)$$

where G is the elastic modulus, T the temperature in degrees Celsius, and τ the time scale.

In principle, this model, together with a single elasticity measurement at one temperature and one concentration, can be used to calculate the whole gelation kinetics of a sample. This includes setting to several months, storage, at concentrations between 3 and 10% w/w and temperatures between 5 and 20 °C. The Bloom value is directly proportional to the elastic modulus at a concentration of 6.66% after 16 h at 10 °C.

Effect of pH The gel strength is only seriously affected by solution pH at extremes of the range; from pH 4 to pH 9 it is not affected to any significant extent. Dilute gels (< 2%) are more affected while stronger gels (> 10%) are relatively insensitive to pH.

Time and temperature of set The gel strength will depend on the time and temperature of set. This will vary from gelatin to gelatin and is also dependent on the proportions of molecular fractions present and thus the viscosity. When a gel is in the process of formation, the more the coils can align themselves, before or during maturation, the greater the number of junctions that will form. This becomes apparent if gelatins are snap-chilled as they are found to have significantly lower gel strengths than expected. The sites involved in hydrogen bonding cannot align themselves efficiently enough to produce the expected number of junctions, whereas if gels are 'tempered' at a temperature just above the setting point, the gel strength is above that which could be expected.

Effect of melting point The unique organoleptic qualities exhibited by gelatin are strongly dependent on its melting point, which in turn is influenced by:

1. The bloom value and viscosity of the gelatin.
2. The concentration of the gel.

These in turn can be altered by other ingredients incorporated in the food concerned e.g., salts, sugars, and other gelling or thickening agents.

Effect of low-molecular-weight compounds The effect of other dissolved ingredients in the gel can be either to weaken the gel or to strengthen it. Most simple sugars, glycerol, and other nonelectrolytes can contribute to an increase in gel strength (fructose and sorbitol being an exception) while the addition of most electrolytes has the opposite effect. It has been shown that these components affect the number and size of the junction zone as they impact on the quantity of the solvent (i.e., in sucrose, the number of cross-links increases and the size of each junction zone decreases).

Compatibility with other food biopolymers Gelatin is incorporated as a functional ingredient in a wide variety of foods. When doing so, it is important to have a knowledge regarding its compatibility with other substances with which it is likely to be blended in order to form the final food product.

Food biopolymers can be classified into two main groups depending on their structural features. Gelatin solutions will either interact constructively or destructively depending on the environment.

Noncharged biopolymers, e.g., locust bean gum, guar gum, starches, agarose This class of hydrocolloid can be added in low concentrations (separately from the gelatin) to most food systems without significantly impairing the gelling properties of the gelatin.

However, as the concentration is increased, phase separation of the two will often occur.

When composites are prepared in the liquid state at high temperature, gelatin is compatible with the other biopolymer molecules. This compatibility is reduced when the temperature decreases. The incompatibility between the two biopolymers is revealed by an increase in turbidity of the solution and by the appearance of two separated phases (droplets). The two phases have different compositions: one phase is enriched in gelatin while the other is enriched in the other biopolymer and a phase diagram and tie-lines can be constructed (Figure 1). Ultimately, when the densities of the two phases are different, the hot solution evolves into two physically separate layers, with creaming of one phase and sedimentation of the other (Figure 2).

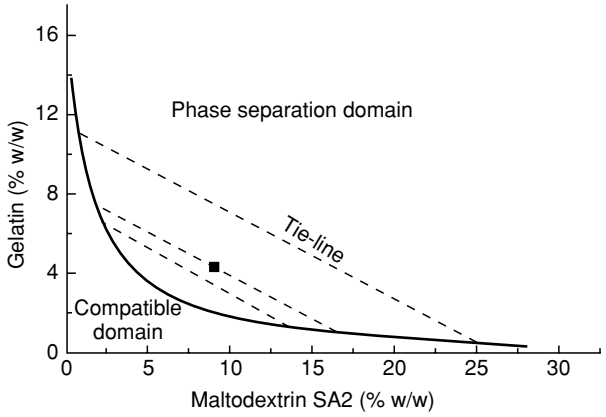

Figure 1 Phase diagram of gelatin (lime hide), maltodextrin SA2 (degraded potato starch), measured at 60 °C. The binodal line separates compatibility and incompatibility domains.

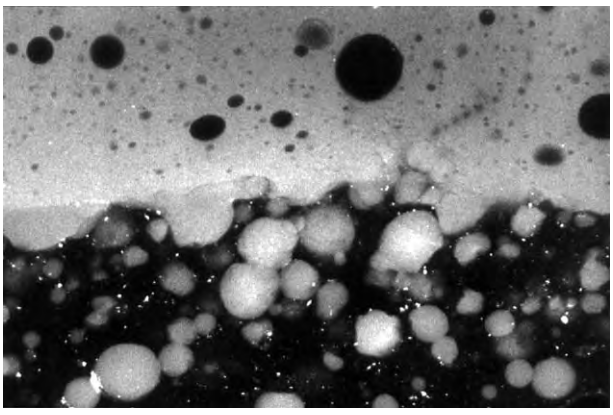

Figure 2 Confocal laser scanning microscopy image of the boundary between two phases in equilibrium. Gelatin is clear and maltodextrin is dark. The composition of the system corresponds to the solid square presented in **Figure 1**. The composition of each phase is determined by the end points of the tie-line (20 mm = 100 μm).

The composite gel formed after cooling of the mixed solution has properties which depend on the microstructure produced. For instance, whether the gelatin forms the included or continuous phase for the same initial system composition, different trapped microstructures can be achieved depending on the processing conditions (i.e., temperature profile, shear, etc.). This can even result in phase inversion of the two biopolymer phases.

Charged biopolymers, e.g., pectins and alginates, agar, carrageenan, and gum arabic This class of hydrocolloid with its negative charge will interact with gelatin, coacervating in a joint phase from solution when the gelatin has a net positive charge and acting synergistically when the gelatin has a negative charge. The degree of synergism is dependent on the negative charge density, and its distribution on the hydrocolloid. However, when the saturation of the charges has been reached, similar behavior as above is reached involving compatibility, incompatibility, and microstructure trapping.

Effect of processing conditions Gelatin will degrade and lose its gelling properties when subjected to conditions of heat, extremes of pH, and exposure to enzyme attack. This has been studied in some detail and modeled.

Also, when gelatin gels under physical disruptions (shear), the final properties of the gel can be very different. For instance, constant shear stress conditions which modify the linear rheological properties at short times of the gelatin gels still result in gels with similar linear properties at long times, but in which the resistance against fracturing is weaker. The model proposed involves particles of gelatin that connect one to the other by cross-link formation to form a network with the same characteristics as the quiescent gel. However, the last cross-links formed are weaker than the earliest and therefore are the first to melt when the gel is submitted to large strains. This highlights the healing properties of the gelatin. Moreover, if constant shear strain is used, instead of a rubber-like material, the final product is more fluid or paste-like as fluid gel particles are produced.

Polyelectrolyte Properties

The gelatin molecule, with its mixture of acidic and basic amino acid side chains, has polyelectrolyte characteristics and is described as amphoteric. At pH levels lower than the isoelectric point the molecule will have a net positive charge, and above this point a net negative charge. Since the isoelectric point in gelatin can vary considerably, the extent of net charge is dependent on the environment pH.

Surface Activity and Related Sol Properties

Gelatin is often used for its surface-active properties, altering the rate of crystal growth in supersaturated solutions, e.g., marshmallow and icecream, where sugar and ice crystal growth are suppressed, giving the desired product. It also finds applications in stabilizing other emulsions such as mayonnaise, pâté, and meat homogenates. Gelatin is added to various types of yogurt formulations, stabilizing the product. It forms a weak gel structure, which imbibes free water, thus inhibiting separation of whey, especially after pasterurization.

The polyelectrolyte behavior of gelatin solutions is widely used for clearing wine and apple juice musts of yeast suspensions, tannins, and other polysaccharides formed during fermentation. It can also be used in the modern 'hot fining process' designed for improved cider and apple juice production. These properties are related on the molecular level to three parameters:

1. Overall charge, and its distribution along the chain.
2. Distribution of nonionic groups.
3. Molecular weight (mean chain length).

Applications

Uses of gelatin in food production Gelatin is widely used as an ingredient in the food industry, due to the following functional properties:

1. It forms high-quality gels in dilute solution with typical clean, melt-in-the-mouth textures.
2. It forms elastic gum-type textures in the concentrated gel, slowly dissolving in the mouth.
3. It forms stable gels on cold storage.
4. It produces emulsification and stabilization of immiscible liquid/liquid, liquid/air or liquid/solid mixtures.
5. It acts as a polyelectrolyte to flocculate suspended particles or unstable colloids from dilute solutions.
6. It acts as a water retainer (it swells easily and there is no syneresis).
7. It acts as an efficient tableting aid and binder.

Gelatin is sold to many sections of the food industry. The major areas of use are in table jellies, confectionery, meat products, icecreams, low-fat and very-low-fat spreads, and chilled dairy products.

Other applications are in the pharmaceutical industry for hard and soft capsule manufacture, and in the photographic industry, which exploits the unique combination of gel formation and surface activity to suspend particles of silver chloride or light-sensitive dyes, without any agglomeration.

Gelatine Derivatives

Cold-Water-Soluble Gelatin

When gelatin solutions are dried without passing through the gel phase, they are of amorphous structure, and do not show any crystalline character, as do gelatins dried from the gel phase. When these dried solutions are rehydrated at a temperature below the gelling point, they will set and form a rigid gel identical in every way to that of a normal gelatin. Unfortunately, when dehydrated, gelatin is extremely hygroscopic and it is difficult to form gels of moderate concentration. In order to overcome this difficulty, mixtures of gelatin and certain carrier substances are available. Glucose syrup and starch are the most popular, as they are ingredients in many of the products for which the gelatin is intended. Drum-dried instant gelatin is also available, but still has to be intimately mixed with other ingredients to insure complete solution.

Hydrolyzed Gelatins

Hydrolyzed gelatins differ from normal gelatins in that they are soluble in cold water, but possess no gelling power.

The hydrolysis of gelatin is complex. There must be strict control of the reaction to avoid the formation of undesirable products such as bitter peptides. Collagen itself is sometimes used, thereby avoiding the gelatin extraction stage. Several manufacturers market a range of hydrolyzed gelatins, each with controlled molecular weight ranges, which can vary from < 1000 to 15 000 molecular weight.

They can be used in tablet binding and granulation agents replacing gelatin, which is not cold-water-soluble under normal conditions of use. Hydrolyzed gelatines give the tablet a better rate of dissolution and disintegration. They are also used as emulsifiers in meat/fat emulsion and as an encapsulation agent for flavor concentrates, taking advantage of its low carbohydrate content.

Hydrolyzed gelatin can be kept as a concentrated solution in a bulk pack after ultrahigh-temperature treatment, or more commonly with preservative added. These are often supplied to the wine industry as ready-to-use fining agents. Hydrolyzed gelatins still possess many of the useful properties of gelatin, e.g., emulsion stability, colloid protection, and flocculation promotion. Much of the production is destined for the cosmetic market where it is added to shampoos as 'protein' and to creams and lotions as 'hydrolyzed collagen,' where use is made of its skin and hair absorption characteristics.

The price of food-grade hydrolyzed gelatine is roughly that of a high-grade gelatin, partially due

to the cost incurred in spray-drying the processed materials.

Analysis of Gelatin Content

The gelatin content of foods can be determined in several ways, each depending on other known constituents present. Kjeldahl nitrogen digestions or biuret reagent determinations can only be applied if it is certain that there are no other proteinaceous compounds present. The procedure for Kjeldahl determinations is well known and suitable conditions have been published. The conversion factor for gelatin determinations is considerably lower than for most other proteins (5.36 for collagen, 5.51 for type B gelatin, and 5.46 for type A gelatin). Hydroxyproline analysis is the definitive method which is applicable in nearly all situations with a high degree of accuracy.

Nutritional Aspects

Because gelatin does not contain any tryptophan, it cannot be used as a 'complete protein;' however it has increased proportions of certain amino acids (e.g., lysine). It can therefore be used to supplement other proteins, to give a mixture with a higher protein value than each component. When mixed with beef protein, the net protein value can rise from 84% to 99%.

The calorific value of gelatin is only 3.5 kcal g^{-1}, which explains its use in low-calorie and diet foods.

See also: **Bone**; **Fish Oils**: Dietary Importance

Further Reading

Amman-Brass H and Pouradier J (eds) (1989) *Proceedings of the Fifth IAG Conference*. Fribourg, Switzerland.

Harris P (ed.) (1990) *Food Gels*. Essex: Elsevier Science.

Mitchell JR and Ledward DA (eds) (1986) *Functional Properties of Food Macromolecules*. Essex: Elsevier Science.

Pearson AM, Dutson TR and Bailey AJ (eds) (1987) *Collagen as a Food. Advances in Meat Research*, vol. 4. London: Van Nostrand Reinhold.

Ross-Murphy SB (1997) Structure and rheology of gelatin gels. *Imaging Science Journal* 45: 205–209.

teNijenhuis K (1997) Thermoreversible networks – viscoelastic properties and structure of gels – introduction. *Advances in Polymer Science* 130: 1–12.

Veis A (ed.) (1964) *The Macromolecular Chemistry of Gelatin*. New York: Academic Press.

Ward AG and Courts A (eds) (1977) *The Science and Technology of Gelatin*. London: Academic Press.

GENE EXPRESSION AND NUTRITION

P Trayhurn, University of Liverpool, Liverpool, UK
J E Hesketh, University of Newcastle, Newcastle upon Tyne, UK

Introduction

Molecular biology has provided a detailed description as to how genes are organized and expressed. A number of technologies have been developed in the period since the late 1970s which enable the factors which influence the expression of genes to be readily investigated. In particular, these techniques permit study of the interaction between nutrients and the genetic information in an organism. This article summarizes some of the key concepts and techniques of molecular biology, and their application to nutritional science.

Structure of DNA and RNA

The two fundamental molecules involved in the transfer and translation of genetic information are DNA and RNA. These polymeric macromolecules are composed of a chain of nucleotides. Each nucleotide consists of a sugar moiety, a phosphate group, and either a purine or a pyrimidine base. In the case of RNA the base is ribose, while in DNA it is deoxyribose. The nucleotides in DNA contain one of four bases – the purines adenine (A) and guanine (G), and the pyrimidines cytosine (C), and thymine (T). Both the purines A and G are also present in RNA, as is the pyrimidine C. However, the second pyrimidine in RNA is uracil (U) rather than T.

The nucleotides in DNA and RNA are linked by phosphate bridges to form a sugar–phosphate backbone. Each base is attached to the sugar moiety. DNA consists of two polynucleotide strands arranged in the double helix structure elucidated by Watson and Crick. The sugar–phosphate is on the outside of the helix, with the bases in the center. The two strands of the helix are held together by hydrogen bonding between specific base pairs. Each base pair consists of a purine and a pyrimidine, the pairings being A–T (two hydrogen bonds) and C–G (three hydrogen bonds).

The information necessary for the synthesis of each individual protein in an organism is encoded by lengths of DNA, termed genes. Large numbers of different genes are linked together linearly, and assembled with specific histone proteins, to form chromosomes, the number of which varies from species to species (e.g., 46 in humans and 8 in fruit flies). In eukaryotes, the chromosomes and the genes within them are localized in the nucleus of the cell. The number of human genes is thought to be between 30 000 and 50 000, but the full realization of the Human Genome project will provide a definitive assessment.

Gene Expression

The precision of base pairing allows the nucleic acids to act as templates for the synthesis of strands containing complementary sequences (Figure 1). This is the critical feature of the transfer of genetic information in all organisms. The sequence of the four bases in DNA is transcribed into a complementary strand of RNA, messenger RNA (mRNA). This transcription process uses the same base pairings as in the two strands of DNA, except that mRNA contains U rather than T. Thus A in DNA codes for U in mRNA.

Genes are now recognized to consist of coding regions (exons), regulatory sequences (promoters), and internal noncoding regions (introns); only a small part ($\sim$2%) of the mammalian genome actually codes for protein. Transcription produces an RNA molecule which is complementary to both coding and noncoding regions. However, sequences corresponding to the introns are removed from the primary transcript to produce the functional mRNA molecule.

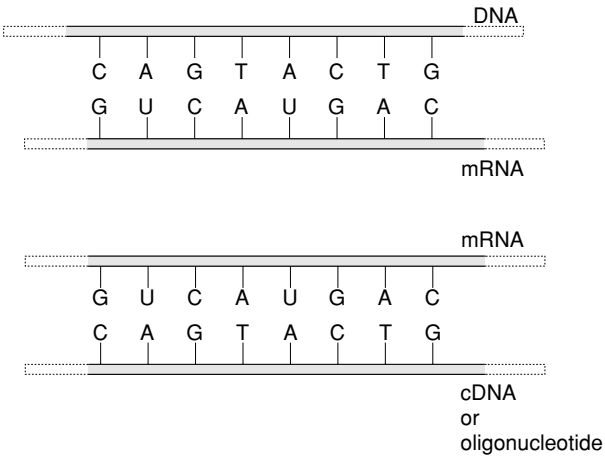

Figure 1 Coding of messenger RNA (mRNA) and complementary DNA (cDNA) molecules.

In eukaryotic cells mRNA is transferred from the nucleus to the cytoplasm, where it becomes associated with ribosomes. The ribosomes, which consist of ribosomal RNA (rRNA) and protein, read the sequence of bases in mRNA and translate this into a sequence of amino acids specific for each individual protein. Not only does the ribosome read the sequence of bases, but it also assembles the amino acids into the protein through the intervention of a further form of RNA, namely transfer RNA (tRNA).

Each amino acid in a protein is coded by a sequence of three bases, and this triplet code is the same in all organisms (universality of the genetic code). The code has some degeneracy in that certain amino acids are coded by more than one triplet.

It is evident from the above that DNA sequences direct the synthesis of protein through the medium of mRNA. Modern molecular biological techniques allow the detection of specific mRNAs, and in practice the expression of a particular gene in a tissue is assessed by identifying the presence of the corresponding mRNA. Strictly, the protein itself should also be identified since an mRNA could, in principle, remain untranslated.

Technology

The ability to detect specific mRNAs depends on two broad types of methodology, one depending on the preparation of DNA sequences which are complementary to part, or all, of the sequence of the mRNA of interest. This is achieved through the application of recombinant DNA technology. Once obtained, a complementary DNA (cDNA) probe will bind, or hybridize, with the corresponding mRNA through the specificity of the base pairings. By tagging the cDNA molecule with a marker, the cDNA–mRNA complex can be readily detected.

The other methodology is dependent on the conversion of the mRNA sequence to DNA using reverse transcriptase (see below) and the subsequent amplification of this DNA by the polymerase chain reaction (PCR).

Isolation of RNA

In order to obtain RNA, tissues and cell membranes must first be disrupted in a denaturing medium, and nucleic acids selectively solubilized. Precautions should be taken throughout to avoid degradation by endogenous and exogenous ribonuclease enzymes. Most procedures yield total RNA, from which mRNA can be obtained by exploiting the characteristic poly(A)$^+$ sequence; this sequence can be bound by synthetic poly(T) oligonucleotides. By linking oli-

go(T) to supports, such as cellulose or magnetic beads, mRNA may be readily isolated.

cDNA cloning

The production of a cDNA probe utilizes the total mRNA complement of a tissue as the starting material. The process requires the synthesis of complementary DNA sequences, their linking to DNA from a plasmid (vector), multiplication of the plasmids in host bacteria, and subsequent screening and selection of clones containing the cDNA of interest (Figure 2).

cDNA sequences are synthesized from mRNAs using the prokaryotic enzyme reverse transcriptase; this involves a fundamental reversal of the normal flow of genetic information. Each cDNA will be an exact replica of parts, or all, of the coding regions of the original DNA from which the mRNA was derived (Figure 1). Since the starting material is invariably a complex mixture of some hundreds or thousands of mRNAs, a wide range of cDNAs will be obtained by the action of the reverse transcriptase.

In the next stage the cDNAs are joined (ligated) to a plasmid DNA (e.g., pBR322), thus generating a recombinant DNA molecule. The plasmids are incorporated into a host bacterium, such as *Escherichia coli*, which is then grown using standard microbiological procedures, resulting in the production of large quantities of the cDNAs. Using appropriate vectors, the cDNAs will direct the synthesis of encoded proteins. Traditionally, clones containing the required cDNAs are identified using antibodies raised against the protein product of the mRNA of interest. Appropriate clones are maintained in culture, and the cDNA excised from the plasmid DNA within the bacteria, using selective restriction enzymes (endonucleases).

Hybridization

A cDNA will bind, or hybridize, specifically to its corresponding mRNA, and this complex can be detected if the cDNA is first tagged with a radioactive (^{32}P) or other label (e.g., enzyme-linked systems which generate chemiluminescence). Hybridization

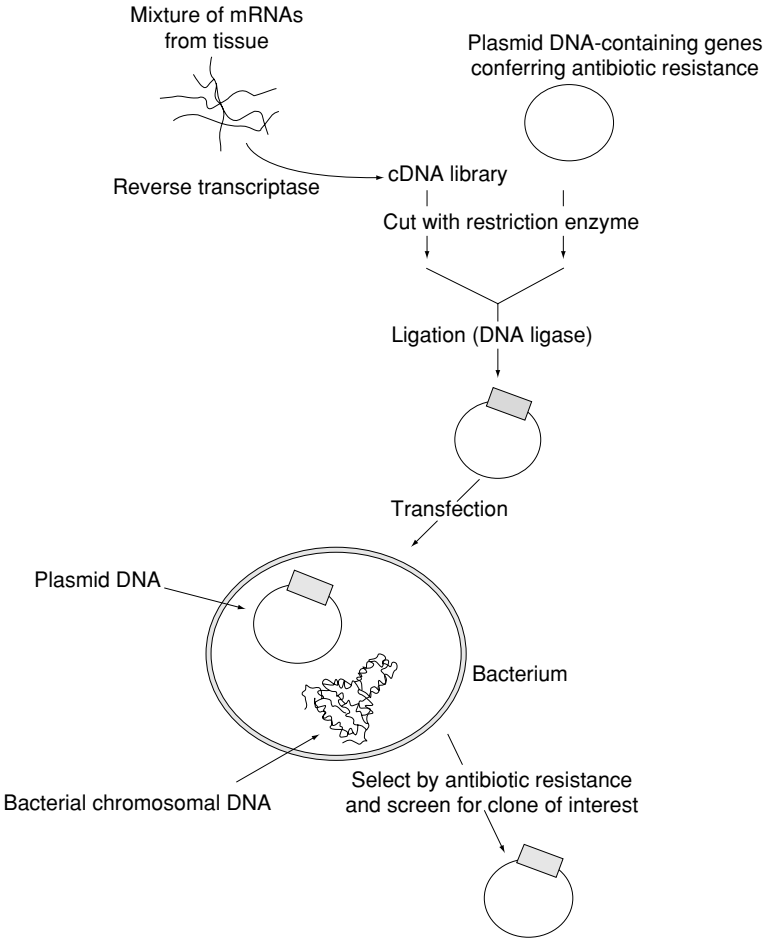

Figure 2 A schematic representation of the application of recombinant DNA technology to produce a cDNA probe.

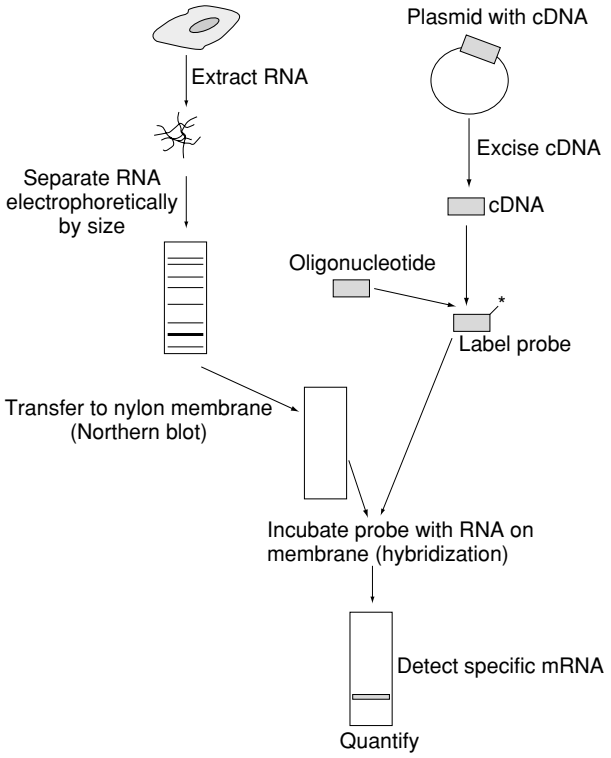

Figure 3 A schematic representation of the detection and measurement of mRNA.

is carried out on solid supports such as nitrocellulose or nylon membranes. This can be achieved by direct application of a mixture of total RNA, or purified mRNA, on to a membrane (dot blot). Alternatively, and more usually, the RNA species are first separated electrophoretically according to size (using agarose gels), followed by transfer to a membrane (Northern blotting); see Figure 3.

Identification of mRNA–cDNA complexes is made by autoradiography, film, or colorimetry, according to the nature of the label attached to the probe. In each of these cases quantification can be achieved by densitometry, or image analysis. A membrane may be sequentially reprobed for the analysis of different RNA species, and this is particularly important in quantitative studies in that it allows the level of a specific mRNA to be related to reference RNAs. In addition, and of special relevance to nutritional studies, the effects of a defined manipulation can be assessed with respect to the expression of different genes associated with linked metabolic processes.

Oligonucleotides

Once the sequence of a gene, or a cDNA probe, is known, computer databases of gene and cDNA sequences can be scanned, so as to design an oligonucleotide probe which will specifically recognize a particular mRNA. Such oligonucleotides provide a convenient alternative to cDNA probes for the detection and measurement of mRNAs (Figure 1). Automated oligonucleotide synthesizers are widely available, and in general an oligonucleotide of 25–30 bases has sufficient specificity to provide a unique probe for each mRNA.

Polymerase Chain Reaction

An alternative to mRNA detection by Northern hybridization is detection by a combination of reverse transcriptase and polymerase chain reaction (RT-PCR). This approach uses short, complementary sequences (primers) to make a cDNA sequence from the mRNA and then, using PCR, to amplify the DNA to produce copies of this DNA sequence. Since PCR uses primers specific to the DNA sequence only the cDNA produced from the original RNA is amplified and produced in sufficient quantity to be detectable after subsequent electrophoretic separation. The method is more sensitive than Northern blotting and more suitable for dealing with large numbers of samples. However, care needs to be taken to make the method quantitative.

In situ hybridization

A further methodology for examining gene expression is *in situ* hybridization, which utilizes probes (often oligonucleotides) for detecting mRNAs in tissue sections. *In situ* hybridization is particularly valuable for identifying the specific region(s) of a complex organ, such as the brain, in which expression of a particular gene occurs. Similarly, the cell types in which expression is taking place within a tissue can be determined.

DNA microarrays

A recent, and rapidly evolving innovation comes from the development of DNA microarrays. One such technique involves the use of arrays of single-stranded cDNA targets immobilized on nylon or glass supports. Each array can contain thousands of cDNA target sequences. Gene expression is monitored by isolation of RNA from a given material, followed by conversion of all the mRNA species in the sample to fluorescently labeled single-stranded DNA probes (using reverse transcriptase) and then hybridization of the labeled probes with the immobilized target. This methodology allows the analysis of the expression of hundreds – or even thousands – of genes in parallel; it offers considerable potential for investigating the integrated effects of a given stimulus or intervention on gene expression.

Nutrition

An appropriate quantity and balance of nutrients is essential for the genetic potential of each individual to be fully realized. Nutrition represents a major interaction with the environment, and an organism has to exhibit both acute and long-term adjustments to changes in nutrient availability. For mammals, these changes are especially marked both at birth, when feeding by the oral route is initiated with the periodic consumption of milks that are characteristically of high fat content, and at weaning when there is normally a switch towards a diet with a lower ratio of fat to carbohydrate.

Gross Nutritional Changes and Gene Expression

Changes in diet induce alterations in the flux of substrates through individual metabolic pathways. Such adaptations are regulated by alterations in hormonal status, and require, on a long-term basis, changes in the amount of associated enzymes. Adjustments in enzyme synthesis generally involve changes in gene expression – either by switching genes 'on' or 'off,' or by 'up' or 'down' regulation of genes that are already being expressed. The genes that code for hormones and their receptor proteins, as well as for transport and carrier proteins, are often also subject to nutritional modulation.

The expression of individual genes is affected by changes in both the quantity and quality of the diet. Thus fasting and refeeding (which represent extreme changes in quantity) lead to alterations in the expression of a variety of genes, including those coding for enzymes of carbohydrate and lipid metabolism, and for receptor and transporter proteins. For example, fasting induces an increase in the level of the mRNAs for both the insulin and the insulin-like growth factor (IGF_1), while subsequent refeeding leads to a rapid fall. Such changes may be tissue-specific, as is the case with the glucose transporters where the level of the mRNA for the insulin-sensitive variety (GLUT 4) alters reciprocally in skeletal muscle and white adipose tissue in response to fasting.

Major changes in the macronutrient composition of the diet can provoke substantial alterations in the expression of particular genes. For example, switching from a high-fat to a high-carbohydrate diet leads to greatly increased levels of the mRNAs for lipogenic enzymes, such as fatty acid synthetase and acetyl coenzyme A carboxylase; in contrast, the levels of the mRNAs for enzymes associated with gluconeogenesis (e.g., phosphoenolpyruvate carboxykinase) are decreased.

It should be noted that, although changes in the level of an mRNA generally imply that the expression of a particular gene has altered, this is not always the case since variations in the stability of mRNAs may also occur.

Transcriptional and Posttranscriptional Control of Gene Expression

In many cases, such as those discussed above, dietary components do not themselves directly affect gene expression, the effects being mediated through changes in hormonal signals. However, in certain instances there is a more direct interaction between specific nutrients and gene transcription. This is especially clear in the case of some micronutrients. Zinc, for example, is required for the loop, or finger-like, structures present in a number of regulatory transcription factors (e.g., nuclear receptors for thyroxine and steroid hormones), and has also been implicated in the regulation of the level of the mRNA for thymidine kinase, an enzyme involved in DNA synthesis.

In addition, the expression of certain genes, notably those coding for proteins involved in the transport, or storage, of trace elements is sensitive to changes in the supply of specific micronutrients. This regulation occurs both at transcriptional and translational levels. Thus transcription of the gene for the metal-binding protein, metallothionein, is regulated by zinc. On the other hand, synthesis of the iron storage protein, ferritin, is controlled by binding of an iron-dependent protein to a specific sequence in an untranslated region of the ferritin mRNA, thereby inhibiting its translation into the protein.

Increasingly, it is being realized that nutrient–gene interactions involve not only the regulation of gene transcription but also posttranscriptional controls such as mRNA stability, translation, and mRNA localization.

Transgenics

Transgenics involve the insertion of genes from one species into the chromosomes of another. The use of transgenic animals has opened substantial opportunities both for the study of metabolic control and for biotechnological applications in agriculture and medicine. One of the earliest and most dramatic examples of the application of transgenics came from the insertion of the gene for growth hormone into mice – such transgenic mice are substantially larger than normal animals. From a nutritional standpoint transgenic animals may have altered nutrient requirements, which will depend in part on the nature of the incorporated gene(s) and the physiological consequences of the change.

Biotechnological applications of transgenics to livestock production include the possibility of inserting genes coding for enzymes associated with the synthesis of essential amino acids, such as lysine, which may be rate-limiting in protein deposition. In medicine, transgenics allows the insertion of genes coding for proteins of therapeutic or pharmaceutical interest. Production of such proteins in milk, through expression in the mammary gland, is likely to provide a particularly potent route for harvesting large quantities of substances such as growth factors and hormones.

Glossary

cDNA

A DNA sequence complementary to part, or all, of the sequence of a particular messenger RNA

Gene Transcription

Conversion of the base sequence of a gene into the corresponding complementary RNA sequence (messenger RNA)

Gene Translation

Conversion of the base sequence in messenger RNA to the corresponding sequence of amino acids in a protein

Hybridization

Formation of duplexes between complementary RNA and/or DNA sequences

Northern Blotting

Transfer of RNA, following separation by molecular size using agarose gel electrophoresis, to a nylon or nitrocellulose membrane

Plasmid

A piece of circular, extrachromsomal DNA which replicates independently; used for the amplification of inserted cDNA sequences

See also: **Cells**; **Enzymes**: Functions and Characteristics; **Nucleic Acids**: Properties and Determination; Physiology; **Zinc**: Physiology

Further Reading

Goodridge AG (1987) Dietary regulation of gene expression: enzymes involved in carbohydrate and lipid metabolism. *Annual Reviews of Nutrition* 7: 157–185.

Hanson RW, Hatzoglou M, McGrane MM, Wynshaw-Boris A, Rottman FM and Wagner T (1989) Molecular biology and nutrition research. *Journal of Nutrition* 119: 957–964.

Hesketh JE, Vasconcelos MH and Bermano G (1998) Regulatory signals in messenger RNA: determinants of nutrient–gene interaction and metabolic compartmentation. *British Journal of Nutrition* 80: 307–321.

Lewin B (1997) *Genes VI*, 6th edn. Oxford: Oxford University Press.

Sambrook J, Fritsch EF and Maniatis T (1989) *Molecular Cloning – A Laboratory Manual*, 2nd edn. New York: Cold Spring Harbor Laboratory Press.

Trayhurn P (1996) Northern blotting. *Proceedings of the Nutrition Society* 55: 583–589.

GENETICALLY MODIFIED FOODS

R L Fuchs and M A Mackey, Monsanto Company, St Louis, MO, USA

Background

Food crops have been genetically improved since the beginning of agriculture by selecting and breeding seeds from superior plants. These traditional breeding methods resulted in significant increases in crop productivity and food and feed quality. More recently, it has become possible to introduce DNA directly into crop plants to produce specific genetic modifications; the techniques used are commonly referred to as plant biotechnology, genetic engineering, or recombinant DNA methodology. The introduced DNA can code for proteins that provide plants with the ability to protect themselves against insect damage and fungal, viral, or bacterial diseases; tolerate preferred herbicides; increase yield and nutritional value; reduce naturally occurring toxicants or allergens; modify the ripening process; and provide fruits or vegetables with superior flavor, and other desirable qualities.

Plant biotechnology is a tool to help address the challenge of producing more food, and more nutritious food, more effectively than was previously possible. Global demand for food is projected at least to double in the next 50 years as the world population increases from the current 6 billion to 10 billion. The World Bank outlined a list of agricultural

improvements that can help meet future food needs, including improving pest control and soil conservation; developing new crop strains with increased yields, resistance to pests, and tolerance to drought conditions; and reducing dependence on pesticides and herbicides. Biotechnology must be integrated with the best crop germplasms and agricultural practices in the various geographies to help meet these future food demands.

Biotechnology Process

There are several ways to insert genes in plants, with the most popular being through use of specific bacterial plasmids or particle bombardment. Plasmids are rings of DNA found in bacteria, such as *Escherichia coli* and *Agrobacterium tumefaciens*, a soil bacteria. A gene coding for a desired trait from a donor cell is spliced into an appropriate plasmid using restriction endonucleases. Appropriate regulatory signals are added to enable the gene to function optimally in the host plant. The engineered plasmids are introduced into *E. coli*, then transferred to *A. tumefaciens* by a conjugation process. *A. tumefaciens* containing the engineered plasmids are incubated with tissue from the host plant under conditions which enable the gene(s) of interest to be transferred to and stably inserted into the genome of the host. The inserted gene(s) can then be transferred to related plant varieties using traditional breeding methods.

Particle bombardment is used to transfer DNA directly into cells by use of a 'particle gun.' Pellets of gold or tungsten are coated with the desired DNA and literally fired through the plant cell walls. As they pass through the cells, some of the DNA coating is left behind and stably incorporated into the plant genomic DNA to create a plant cell that will express the desired trait. This technique has been particularly valuable in plants for which the *Agrobacterium* transformation method was not effective or efficient.

Products on the Market

Since the initial reports of the production of the first genetically improved plants in 1983, almost all economically important plants have been genetically improved, and as of 2001 over 50 different products have completed regulatory review by the appropriate regulatory agencies for marketing in at least one country (Table 1).

Most of these initial products were developed either to be resistant to pests, such as viruses and insects (including potato beetles, cotton bollworm, or corn borers), or to tolerate specific herbicides, such as glyphosate, glufosinate, or bromoxynil. Making plants resistant to pests improves pest

Table 1 Plant biotechnology products that have regulatory approval in at least one country as of 2001

Company	*Trait*
AgrEvo	Phosphinothricin-tolerant canola, corn, soy, rice, sugar beets
	Male sterile/phosphinothricin-tolerant canola and corn
	Phosphinothricin-tolerant/ lepidopteran-resistant corn
	Fertility restorer/phosphinothricin-tolerant canola and corn
Agritope, Inc.	Modified fruit ripening tomato
	Delayed-ripening cantaloupe
Asgrow Seed Co.	CMV/WMV2/ZYMV-resistant squash
BASF	Sethoxydim herbicide-tolerant corn
	Canola with enhanced digestibility (animal feed use)
Bejo-Baden	Male sterility chicory
Calgene Inc.	Delayed-ripening tomato
	Bromoxynil-tolerant cotton
	Bromoxynil-tolerant/lepidopteran-resistant cotton laurate canola
China	Virus-resistant tomato and tobacco
Ciba-Giegy	Lepidopteran-resistant corn
Cornell U./U. of Hawaii	PRSV virus-resistant papaya
DeKalb Genetics Corp.	Phosphinothricin-tolerant corn
	European corn borer-resistant corn
	European corn borer-resistant/ phosphinothricin-tolerant corn
DNA Plant Technology	Improved-ripening tomato
DuPont	Sulfonylurea-tolerant cotton
	High-oleic-acid soybean
Florigene	Carnations with increased vase life and modified flower color
Monsanto	Glyphosate-tolerant soybean, cotton, corn, canola, sugar beet
	Improved-ripening tomato
	Insect-protected tomato
	Colorado potato beetle-resistant potato
	Lepidopteran-resistant cotton
	European corn borer-resistant corn
	European corn borer-resistant/ glyphosate-tolerant corn
	Colorado potato beetle-resistant PVY-resistant potato
	Colorado potato beetle-resistant PLRV-resistant potato
Mycogen	European corn borer-resistant corn
Novartis	European corn borer-resistant corn
	European corn borer-resistant/ phosphinothricin-tolerant corn
Northrup King	European corn borer-resistant corn
Rhône-Poulenc	Bromoxynil-tolerant canola
Pioneer Hybrid	Male sterile/phosphinothricin-tolerant corn
Seita	Bromoxynil-tolerant tobacco
Seminis	Virus-protected squash
University of Saskatchewan	Sulfonylurea-tolerant flax
Upjohn	WMV2/ZYMV virus-resistant squash
Zeneca/Petoseed	Improved-ripening tomato

CMV, cucumber mosaic virus; WMV2, Watermelon mosaic virus type 2; ZYMV, Zucchini yellow mosaic virus; PRSV, papaya ringspot virus; PVY, potato virus Y; PLRV, potato leaf roll virus.

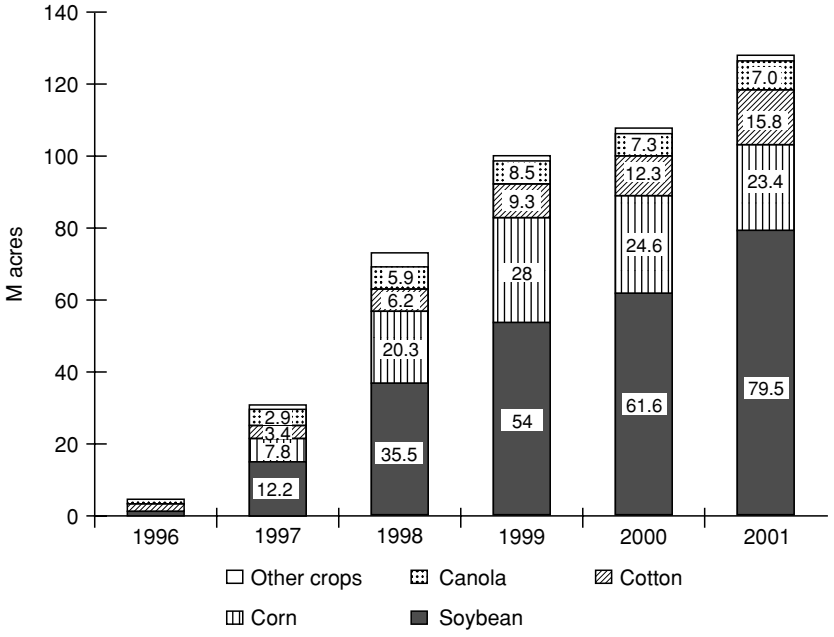

Figure 1 Worldwide acreage (millions) of genetically improved crops planted in 1996 to 2001. Information derived from James C (2001) *Global Review of Commercialized Transgenic Crops*: 2001 Preview ISAAA Brief 24–2001. Ithaca, NY: ISAAA.

management efficacy while reducing the farmer's reliance on and use of various pesticides and increasing crop yield. Herbicide tolerance in crops allows farmers to manage weeds more effectively and with the use of fewer herbicides.

Figure 1 shows the increase in acreage planted worldwide in genetically improved crops from 1996 to 2001. Thirteen countries (USA, Argentina, Canada, China, Australia, South Africa, Mexico, Spain, France, Portugal, Romania, Bulgaria, and Ukraine) grew nine crops (soybean, corn, cotton, canola, potato, squash, papaya, tomato, and tobacco) commercially in 2001. Over 65% of the soybeans, 65% of the cotton, and 25% of the corn grown in the USA in 2000 were genetically improved, while over 95% of the soybeans grown in Argentina were genetically improved (James 2001).

Safety Assessment Process and Regulations for Foods Derived from Genetically Improved Crops

As plant biotechnology products have been developed, so have national and international regulatory oversight measures for them. Appropriate scientific procedures to assess the safety of products from crop biotechnology have been developed over the past decade by scientists and regulatory agencies, including the United Nations (UN) Food and Agriculture Organization (FAO), the UN World Health Organization (WHO), the Organization for Economic Cooperation and Development (OECD), the US Food and Drug Administration (FDA) and others. There is general consensus among scientists on the approach used and information appropriate to assess the safety of foods and feeds from genetically improved plants. These procedures focus on the safety of the final product.

The majority of food and feed products from genetically improved plants introduced to date have been shown to be compositionally and nutritionally comparable (i.e., 'substantially equivalent') to the parent crops, with the exception of the introduced trait(s). Additional safety assessments have focused on the safety of the expressed protein and any products resulting from it. The safety of the protein conferring the desired trait is assessed via internationally accepted procedures, including characterization of the function, specificity, and history of consumption of the same or related proteins, estimation of the level of consumption of the protein, assessment of digestibility, and appropriate animal toxicology studies. Special attention is paid to insure that the introduced protein is not allergenic and does not share physicochemical properties of allergenic proteins. Once food or feed from a crop has been shown to be substantially equivalent to its traditional counterpart except for the introduced trait, and the trait has been shown to raise no safety concerns, the crop (following appropriate regulatory review) may be introduced into the marketplace with the same degree of safety assurance as those products derived from their traditional counterparts.

Table 2 summarizes the information typically provided to regulatory authorities to demonstrate the food, feed, and environmental safety of genetically improved plants and plant products.

In the USA three federal agencies have the authority to regulate biotechnology. The FDA maintains authority to insure the food and feed safety of the products of plant biotechnology; the US Department of Agriculture (USDA) has the authority to insure that these plants do not present any plant pest risks; and the Environmental Protection Agency (EPA) has the authority to assess the safety of plants that produce a pesticidal substance (e.g., plants protected against insects, fungi, or viruses). The EPA also maintains authority over herbicides used in conjunction with herbicide-tolerant plants.

In the FDA's 1992 document *Statement of Policy: Foods Derived from New Plant Varieties*, as well as in the international documents from the FAO, WHO, OECD, and other organizations, there is a general consensus that foods derived from genetically improved plants do not pose any unique safety concerns. The FDA regulates these products in the same way as food and feed products derived from traditionally bred plant varieties.

Other countries, such as Canada, Japan, Korea, South Africa, Switzerland, and Australia also have specific regulations for genetically engineered plant products. In the European Union, the primary control is Directive 2001/18 EEC, which regulates the release of genetically modified organisms (GMO) into the environment. The foods and feeds produced from these GMO are regulated via the Novel Foods Regulation, which requires that before a novel food is marketed, manufacturers must document that the new food does not present a new safety or nutritional risk. Other countries are also developing regulations for these products.

Table 2 Data typically provided to regulatory agencies to support the food, feed, and environmental safety of genetically improved plants and plant products

Food/feed safety assessment
Molecular characterization
Gene source
Transformation system
Insert number
Copy number
Integrity of each inserted genetic element (promoter, coding sequence, polyadenylation sequence)
Genetic stability
Protein safety assessment
Source
Host/processing
Protein expression levels/consumption
History of use of same/similar protein
Safety to nontarget organisms
Function/specificity/mode of action
Homology to known toxins/allergens
Digestibility
Potential toxicity testing (case by case)
Allergenicity assessment
Marker gene/protein safety
Nutritional equivalence
Identification of key nutrients
Levels of key nutrients relative to traditional counterpart
Anticipated uses relative to historical uses
Nutritional assessment of expressed trait (e.g., modified oil, carbohydrate, etc.)
In vitro/*in vivo* nutritional/animal wholesomeness studies (case by case)
Toxicological assessment
Identification of key antinutrients/toxicants in host or organisms related to host
Levels of key antinutrients relative to traditional counterpart
Toxicological assessment of expressed trait
Environmental safety assessment
Molecular characterization (as above)
Outcrossing/gene flow (potential/impact)
Weediness
Competitiveness/survivability/dormancy
Morphological/phenotypic characteristics
Insect/disease susceptibility
Impact on nontarget organisms
Agronomic performance
Resistance management (where appropriate)

Food Quality Enhancements

Although most of the initial genetically improved plant products have improved agronomic traits, the first plant biotechnology product marketed in the USA, the Flavr Savr tomato, was developed to allow the fruit to ripen longer on the vine and have an increased shelf-life. Many additional products with improved quality or nutritional properties are under development (Table 3).

Carbohydrates Modifications in the composition and total amount of starch have been achieved via

Table 3 Products from plants improved to contain improved nutritional or quality traits

High-oleic-acid vegetable oils
High-stearate vegetable oils to replace the hydrogenated oils and reduce *trans* fatty acids
Canola oil high in beta-carotene
Corn and soy with increased levels of lysine, tryptophan, methionine
Corn, rice, wheat, and potato with improved starch characteristics
Rice high in beta-carotene and iron
Sweet potatoes with improved protein quality
Ripening-controlled papayas, cherry tomatoes, bananas, strawberries, and pineapples
Tomatoes with increased lycopene
Soybean with increased vitamin E

biotechnology. For example, amylose-free, amylopectin-enriched potatoes have been produced by inhibiting the production of granule-bound starch synthase to produce starch preferred for specific industrial applications. It was recently reported that by introducing photosynthetic genes from corn into rice, yield increases of up to 25% were observed in new rice varieties. In addition, by redirecting the flow of carbon from starch biosynthesis, increases in sucrose content and production of new components like fructans or other sweeteners also can be realized.

Oils Plant biotechnology efforts have focused largely on altering the fatty acid composition of canola, soybeans, corn, and sunflowers, all major sources of dietary oils, to provide either more oxidatively stable oils or oils with enhanced nutritional characteristics. A genetically improved soybean with increased levels of oleic acid, a monounsaturated fatty acid, and reduced levels of polyunsaturated fatty acids has also been produced. Because this product is more stable, it does not require hydrogenation, thereby eliminating the production of *trans* fatty acids which have been linked to cardiovascular disease. Modification of other fatty acids, such as stearate, is also being evaluated for improved human health.

Proteins Foods that provide sufficient quantities of essential amino acids are critical to meet dietary needs, especially in developing countries where foods low in one or more of these amino acids are often relied upon as staples of the diet. Approaches to increase the amount of lysine in cereals and methionine in legumes are being assessed. For example, total lysine content of canola and soybean was increased two- and fivefold, respectively, by expressing non-feedback-inhibited forms of two key enzymes in lysine biosynthesis. A two- to threefold increase in the total lysine content of corn was also obtained using the same approach. These changes are important both from a human nutritional perspective as well as for improved animal feeds, eliminating or decreasing the need to provide supplemental lysine to animal rations.

Vitamins/minerals Research efforts to increase the levels of certain vitamins and minerals in genetically improved crops are ongoing. In developing countries where malnutrition is a major concern, enhancing these components in staple foods may improve nutritional status and reduce disease. These projects include increasing the levels of beta-carotene, the precursor of vitamin A, vitamin D, iron, lycopene, folate, and other nutrients. The most exciting of these research programs is the production of a variety of rice with significantly increased levels of beta-carotene and iron in a more bioavailable form. Consumption of a typical serving of $300\,g\,day^{-1}$ of this rice could help reduce both xerophthalmia and iron-deficiency anemia, which are devastating nutritional problems in the developing world. Over 2 million children die annually from vitamin A-related illness and over 3.7 billion people globally have a deficiency in iron intake. A canola variety with greatly increased levels of beta-carotene in the oil has also been developed.

Eliminating undesirable food components Plant biotechnology can reduce a number of undesirable food components, such as one of the major allergenic proteins in rice or the major allergens in peanut or soybeans. Other efforts to eliminate allergens in foods by modifying their amino acid sequences have also proven successful. Efforts to reduce or eliminate other undesirable components of foods, e.g., glucosinolates in canola meal, protease inhibitors in beans, glycoalkaloids in potatoes and mycotoxins in corn, are being evaluated. Components such as caffeine from coffee beans can be eliminated or reduced to provide a coffee with no or a very low caffeine level without using chemical extraction.

Impact of Genetically Improved Crops on Agriculture

Insect Protection

Insect pests, which can cause tremendous damage to agronomically important crops, typically have been controlled by a variety of chemical insecticides. Yet 20–30% of total crop product is still lost due to these pests. Thus, genetically improved plants containing insecticidal genes from microbes which have been used safely as microbial insecticides for over four decades were developed to improve the efficiency of pest management.

Approximately four million acres of insect-protected cotton varieties were grown in the USA in 1999 and 2000. Use of these varieties significantly reduced the use of chemical insecticides (by over two million pounds annually in the USA alone), significantly increased cotton lint production, and also increased the number of beneficial insects compared to fields treated with chemical insecticides due to increased specificity of the pesticidal protein. US farmers realized significant economic value, averaging about $40 per acre increase in income. Data from developing countries show even greater benefits; for example, insect-protected cotton in China allowed

reduction of chemical insecticide use by approximately 80% and reduced applications from an average of 15 to less than one per season and increased yields by approximately 25%.

Potato plants protected against Colorado potato beetle, the most damaging potato insect pest, were planted commercially in 1996. Growers used approximately 35% less insecticide to control this beetle compared to growers planting nongenetically protected varieties. In 1999, two new products were introduced which combined protection against the Colorado potato beetle with resistance to either potato leaf roll virus or to potato virus Y, two major viral diseases for potato. These products essentially eliminate the need for chemical insecticide use for potato production.

In 1999 and 2000 approximately 25% of the corn planted in the USA was genetically improved to resist pests. An additional advantage of this genetic modification is that the protection of corn ears from pests also drastically reduces the amount of secondary infection caused by the fungus *Fusarium*. The ability to control *Fusarium* results in up to 95% reduction of the level of the mycotoxin fumonisin under some environmental conditions. Fumonisin is known to cause adverse health effects in swine, horses, rodents, and humans. Significant reductions in fumonisin on insect-protected corn have been demonstrated in a number of countries, including the USA, France, Italy, and Argentina.

Virus Protection

Virus-resistant plants have been produced by using genes derived directly from the virus – so-called pathogen-derived resistance. The first genetically improved plant, tobacco with increased resistance to tobacco mosaic virus (TMV), resulted from the expression of the TMV coat protein in tobacco. This coat protein-mediated resistance has been used extensively to confer viral resistance to numerous food crops, including tomato, squash, melon, papaya, sweet potato, and potato. Squash varieties resistant to zucchini yellow mosaic virus (ZYMV) and watermelon mosaic virus (WMV) were approved in the USA in 1995 and marketed. These two viruses routinely reduce crop yields by 20–80% depending on production season and growing region. More recently, squash plants resistant to these two viruses plus a third destructive virus, cucumber mosaic virus (CMV), were developed. Marketing of these products will allow the more consistent production of squash varieties with significant reductions in the use of chemical insecticides that are typically required to control virus spread by insects. Coat protein genes have also been used to confer resistance to papaya ringspot virus (PRSV) in Hawaii, and to sweet potato feathery mottle virus (SPFMV) in Kenya. Resistance to PRSV has essentially saved papaya production in Hawaii, while resistance to SPFMV has the potential to increase yields in Kenya by up to 80%, based on reported yield losses attributed to this virus.

Selectivity to Herbicides

Weeds are one of the major agricultural pests that can devastate a crop. Weeds compete with crops and reduce yield, decrease harvest efficiency, decrease seed quality, and serve as a reservoir for pests. Agricultural practices used prior to World War II were labor-intensive and contributed to soil erosion. These processes have been replaced by the use of chemical herbicides that are the most effective, reliable, and economic method of controlling weeds on a large scale. Today herbicides are used on essentially 100% of the acreage of the major agronomic crops in developed countries.

Traditionally, herbicides are selected based on the weeds to be controlled and the natural resistance of the crop to the herbicide. Biotechnology provides an opportunity to modify crops so they tolerate selected herbicides with preferred environmental properties, such as glyphosate. These herbicide-tolerant crops allow the farmer to apply herbicide to planted fields, killing weeds but leaving the crop unaffected. This ability provides increased flexibility and cost savings to growers. In addition, farmers can move from using preemergent, soil-incorporated herbicides to postemergent herbicides that are applied on an 'as-needed' basis. This strategy can reduce the number and total amount of herbicides used and enables the application of herbicides that bind tightly to the soil and are less likely to enter the groundwater, thereby providing environmental benefits and encouraging the use of reduced-tillage agriculture.

Soybeans, canola, cotton, and corn tolerant to glyphosate, and canola and corn tolerant to phosphinotricin are examples of herbicide-tolerant crops (Table 1). When introduced in 1996, glyphosate-tolerant soybeans were planted on only 1.5% of the US soybean acres. Because of exceptional weed control and crop safety, these soybeans were rapidly adopted. Approximately 75% of the growers used only one herbicide application, providing a reduction in the amount and number of herbicides used compared to soybeans treated with conventional herbicide programs. By 1999 greater than 65% of the US soybeans, and over 95% of Argentinian soybeans in 2001, were glyphosate-tolerant.

Plants as Production Factories

Plants provide an unmatched capacity to produce products with a maximum efficiency and a minimum of energy – thus, cost-effectively. Plants can be bioengineered to produce insecticidal proteins using just carbon dioxide, water, and nitrogen, thereby reducing the need for expensive and energy-intensive chemical insecticide manufacturing. The potato which has been developed to produce proteins that protect it from both the Colorado potato beetle and the potato leaf roll virus is an excellent example. The combination of these two traits enables potatoes to be grown without the application of chemical insecticides and will result in environmental and economic savings. Approximately 5 million pounds of chemical insecticides are used to control these pests annually in the USA, and 2.5 million pounds of waste are generated in the production process. Less than 5% of the applied insecticide actually reaches the pest. Using the plant to produce these pesticides is more energy-efficient and environmentally acceptable.

One of the most exciting applications of plant biotechnology is the ability to produce edible vaccines in plants. For example, transgenic potatoes containing the nontoxic binding subunit of the *E. coli* enterotoxin as well as transgenic potatoes which express the capsid protein of Norwalk virus have been produced. The expressed proteins function as oral immunogens in humans which, if effective, could immunize people against the diarrheal diseases caused by these organisms. Efforts are under way to introduce such vaccines into banana, which would be eaten raw and thus serve as an ideal source to deliver this technology to developing countries, where inexpensive vaccines are urgently needed.

Improving crop yields is essential to meeting the increased food demands of an ever-increasing world population. Improving crop protection from pests and increasing crop yields through herbicide tolerance are two strategies that will help meet this need. Biotechnological enhancements of crops can insure that these changes occur quickly enough to meet global food needs. At the same time, agricultural biotechnology has advantages for the environment in terms of reductions in soil erosion and tilling, and less agricultural chemical use and run-off. The first generation of crops developed via biotechnology provides these agronomic and environmental benefits. The next generation of products is expected to provide greater direct consumer benefits, such as enhanced food quality and nutritional value.

See also: **Biotechnology in Food Production**; **Food Safety**; **Rape Seed Oil/Canola**; **Retinol**: Properties and Determination; Physiology; **Rice**; **Soy (Soya) Beans**: The Crop; **Rape Seed Oil/Canola**

Further Reading

American Council on Science and Health (2000) *Biotechnology and Food*, 2nd edn. http://www.acsh.org.

Gianessi LP and Carpenter JE (1999) *Agricultural Biotechnology: Insect Control Benefits*. Washington, DC: National Center for Food and Agricultural Policy, http://ncfap.org.

Hammond BG and Fuchs RL (1999) Safety evaluation for new varieties of food crops developed through biotechnology. In: Thomas JA (ed.) *Biotechnology and Safety Assessment*, pp. 61–79. Philadelphia, PA: Taylor & Francis.

IFT (2000) Expert report on biotechnology and foods. *Food Technology* 54: 124–136; 54 (9): 53–74; 54 (10): 61–80.

James C (2001) *Global Review of Commercialized Transgenic Crops*: 2001 Preview, ISAAA Brief 24: 1–20 Ithaca, NY: ISAAA.

Kendall HW, Beachy R, Eisner T *et al.* (1997) *World Bank: Bioengineering of Crops. Report of the World Bank Panel on Transgenic Crops*. Environmentally and Socially Sustainable Development Studies and Monograph Series no. 23. http://www.worldbank.org.

Metcalfe DD, Astwood JD, Townsend R, Sampson HA, Taylor SL and Fuchs RL (1996) Assessment of the allergenic potential of foods derived from genetically engineered crop plants. *Critical Reviews in Food Science and Nutrition* 36: S165–S186.

Organization for Economic Cooperation and Development (1997) *OECD Documents. Report of the OECD Workshop on the Toxicological and Nutritional Testing of Novel Foods*. Paris, 1997; Report of the OECD Task Force for the Safety of Novel Foods and Feeds – Okinawa, Japan. 2000 http://www.oecd.org/subject/biotech/report_taskforce.pdf.

Royal Society of London, USA National Academy of Sciences, Brazilian Academy of Sciences, Chinese Academy of Sciences, Indian National Academy of Sciences, Mexican Academy of Sciences, and the Third World Academy of Sciences Transgenic Plants and World Agriculture (2000) http://www.royalsoc.ac.uk.

World Health Organization (2000) *Safety Aspects of Genetically Modified Foods of Plant Origin. Report of a Joint FAO/WHO Consultation*. Geneva, Switzerland: World Health Organization.

Geriatrics *See* **Elderly**: Nutritional Status; Nutritionally Related Problems; Nutritional Management of Geriatric Patients

GHEE

G S Rajorhia, National Dairy Research Institute, Karnal, India

Introduction

Ghee (synonyms: ghrit, ghritam, havi, rognezard, samn, samna, maslea) is a milk-fat product originally produced in India. The origin of gheemaking lies far beyond recorded history. Indian epics and Vedas relate it to third-century BC. Ghee is also marketed in Australia, Armenia, many African and Asian countries, Belgium, The Netherlands, New Zealand, the UK, and the USA. 'Ghee' is the Indian name for clarified butterfat, and it is usually prepared from cows' milk, buffalo milk, or mixed milks. Milk is separated into cream, which may be directly converted into ghee, or it can be first converted into butter and then into ghee. Milk of other mammals, such as goat, sheep, and camel, can also be used. Cream or butter is heated at a temperature sufficient to evaporate nearly all the moisture content, and to produce the characteristic flavor. Heating destroys most of the microorganisms and enzymes. Low water activity hampers the growth of organisms that survive heat treatment or those recontaminating the product. The suspended curd particles are either filtered off or removed by centrifugation. The clear butterfat is cooled to develop a special physical structure by a controlled crystallization process. (*See* **Buffalo**: Milk; **Milk**: Dietary Importance; **Sheep**: Milk.)

Mode of Ghee Utilization

The butterfat in milk is largely used in the form of ghee, as this method of utilization admirably suits the climatic conditions and food habits prevailing in the tropics and the subtropics. It does away with the need for elaborate arrangements which would otherwise be required for the manufacture and storage of table butter. It is an effective way of avoiding the risk of microbial and chemical deterioration incidental to holding butter at relatively high temperatures.

There are regional and seasonal variations in the mode of ghee utilization which can largely be influenced by local food habits, price structure, number of festivals, and image of the product as a nutritional supplement. A major portion of ghee is utilized for culinary purposes. This may cover its extensive use for the direct dressing of food articles, such as unleavened breads, cooked rice, and lentils (dhal), for flavoring, and as a cooking and frying medium. About 60–70% of the total ghee in India is used for direct dressing, and almost 15–20% for the cooking and frying of foods. A significant quantity of ghee is used by confectioners and bakers. Ghee is recognized as a sacred article, and approximately 5–7% is used in religious rites, such as the burning of ghee dips, lighting of the sacred fire in yaganas, preparation of sacred offerings and prasadam, and even for cremation of dead bodies. It is conjectured that ahutis (burning of ghee in religious functions) purifies the air. This practice is considered environmentally safe. Other usages of ghee of minor significance include ladies' hair dressing, body massage of wrestlers, athletes, and invalids, and in the formulation of indigenous pharmaceutical drugs. Ghee is also used for flavoring snuff by mixing and grinding to a very fine powder.

Method of Manufacture

The methods of ghee manufacture vary according to the base material used (milk, cream, butter), intermediate treatment of raw materials, and handling of the semifinished or fully formed ghee. There are four methods for the production of ghee which are essentially based on batch operation:

1. the indigenous milk–butter process
2. direct cream (DC) method
3. the creamery butter (CB) method
4. prestratification (PS) method

These methods are suitable for the different scales of production.

Indigenous Milk–Butter Process

The indigenous methods of gheemaking usually involve: (1) direct churning of raw milk; (2) lactic acid fermentation of heat-treated milk followed by churning; or (3) removal of thick clotted-cream layers (malai); from continuously heated milk at temperatures around 80 °C followed by grinding of clotted cream, its dispersal in water and, finally, churning. In methods (1) and (2), milk is churned every day. Hand-driven wooden beaters are usually employed for separating the butter. After accumulating sufficient quantity over a period of a few days, the butter is melted in a metal pan or earthenware vessel on an open fire until almost all the moisture has been

removed. The extent of frothing may be used as an index to judge when to terminate heating. After heating, the contents are left undisturbed. When the curd particles have settled at the bottom of the pan, the clear fat is carefully decanted off into ghee storage vessels.

The traditional gheemaking practice contributes about 90% of the total ghee production in India. This method leaves behind a large quantity of buttermilk of varying quality, and also leads to low fat recoveries (75–85%). This is why modern dairies do not use the indigenous method of gheemaking. However, village ghee constitutes the major share of the base material used for the blending operations at ghee grading and packing centers functioning under the Agricultural Marketing Grading (AGMARK) scheme in India. (*See* **Lactic Acid Bacteria**.)

Direct Cream Method

The small dairies use a technologically improved method for gheemaking which involves the separation of cream from milk by centrifugation. This process omits the need for production of butter because cream is directly converted into ghee. The fresh cream, cultured cream, or washed cream is heated to 115 °C in a stainless-steel, jacketed ghee kettle fitted with an agitator, steam control valve, pressure and temperature gages and a movable, hollow, stainless-steel tube centrally bored for emptying out the contents. Alternatively, provision can be made for a tilting device on the ghee kettle to decant off the product. Heating is discontinued as soon as the color of the ghee residue turns to golden yellow or light brown. One of the limitations of the DC method is that it requires a long heating time to remove the moisture. A high content of serum solids in the cream may also produce a highly caramelized flavor in the ghee, and lead to about 4–6% loss of butterfat in the ghee residue or during handling operations, depending upon the fat percentage in the cream. The use of plastic cream or washed cream with about 75–80% fat is recommended for minimizing both fat loss and steam consumption. The final product will have a less intense cooked flavor when low-solids not fat (SNF) cream is used.

Creamery Butter Method

Most dairies use unsalted CB or white butter as a raw material for gheemaking. A typical plant assembly for the CB method comprises the following: (1) a cream separator; (2) butter churn; (3) butter-melting outfits; (4) steam-jacketed, stainless-steel ghee kettle with agitator and process controls; (5) ghee filtration devices, such as disk filters or oil clarifier; (6) storage tanks for cream, butter, and ghee; (7) pumps and pipelines interconnecting these facilities; (8) crystallization tanks; and (9) product filling and packaging lines.

First, the butter mass is melted at 60 °C. The molten butter is pumped into the ghee boiler, and the steam pressure increased to raise the temperature to 90 °C. This temperature remains constant as long as the moisture is being driven off. The scum which collects on the top surface of the product is removed from time to time with the help of a perforated ladle. The temperature gradually rises and the heating at the last stage is carefully controlled. The endpoint shows the disappearance of effervescence, appearance of finer air bubbles on the surface of fat, and browning of the curd particles. At this stage, the typical ghee aroma is also produced. The final temperature of clarification is adjusted to 110 °C. In some cases, heating beyond this temperature is carried out in order to generate a marked 'cooked' flavor, relished by a sizeable section of consumers. The ghee is then pumped, via an oil filter or clarifier, into settling tanks which are cooled by recirculating water at 60 °C. The residue is packed in suitable containers at 60 °C.

Prestratification Method

The PS method consists of keeping molten butter undisturbed in a ghee boiler at a temperature of 80–85 °C for 30 min for stratifying the mass into three distinct layers. The top layer is composed of floating, denatured protein particles and impurities, the middle layer of almost clear fat, and the bottom layer of buttermilk serum. This division helps in mechanical removal of the bottom layer of buttermilk, carrying about 80% of the moisture and 70% of the SNF contained in butter. Removal of the buttermilk eliminates the need for prolonged heating for evaporation of the moisture, and results in the formation of a significantly low quantity of ghee residue, into which a portion of ghee can become absorbed and from where it is irrecoverable.

The middle layer of fat is heated, usually to 110 °C, along with some denatured curd particles floating on the top. This process is essential to promote development of a more chracteristic aroma. This method offers the advantages of economy in fuel consumption up to 35–50%, saving in time and labor up to 45%, and production of ghee with lower free fatty acid (FFA) levels and acidity. The provisions of pressure gage, safety valves, temperature regulators, and condensate outlet pipe make the PS process capable of producing ghee of better quality, and amenable to process controls. Stratification also helps in the production of ghee with a milder flavor. Its application is limited to batch-scale operation.

Continuous Gheemaking (Figure 1)

To overcome the problems of low heat transfer coefficients in batch methods of gheemaking, limitations of scale of operation and excessive exposure of the plant operators to the stresses of heat and humidity, a continuous gheemaking machine was designed using the principles of hydrodynamics and heat transfer in a horizontal, straight-sided, thin-film, scraped-surface heat exchanger. The rate of water evaporation is 75–80 kg h^{-1} m^{-2} of surface area at a steam pressure in the jacket between 450 and 500 kPa. The rotor is provided with four variable-clearance blades revolving at a speed of 2.45 m s^{-1}. The capacity of the prototype developed is 500 kg h^{-1} ghee.

Butter, heated in a tank to 75 °C, is pumped into the heat exchanger, with the flow rate adjusted with a rotameter. The rotor drive is switched on, and steam allowed to flow into the jacket. The centrifugal action of the rotor causes the product to spread uniformly as a thin film on the heating surface. The turbulence induced by the blade action causes rapid evaporation of water from the product film. The vapors removed via vapor outlets are used for preheating the butter in a tank. During the warm-up period, the partially concentrated fat is diverted into a balance tank. After attaining the desired temperature, ghee is collected in a tank and subsequently passed through a residue separator or clarifier.

Quality of Ghee

The quality attributes of ghee are affected by many factors, such as quality of base material, extent of lactic acid fermentation, time and temperature of heating, rate of cooling, package type, filling conditions, presence of oxygen and contaminants such as iron and copper, exposure to sunlight and temperature, and duration of storage. These factors should be controlled to produce ghee of premium quality.

The consumers of ghee always look for most desirable flavor, texture, and color, and freedom from suspended serum residues. They also want an assurance of purity, freshness, and wholesomeness. The above attributes are assigned relative weightings by consumers in the overall assessment of quality. A perfect score of 60 for flavor, 25 for texture, 10 for color, and 5 for freedom from suspended impurities is recommended for the judging of ghee. The preference for quality is also determined by the end use.

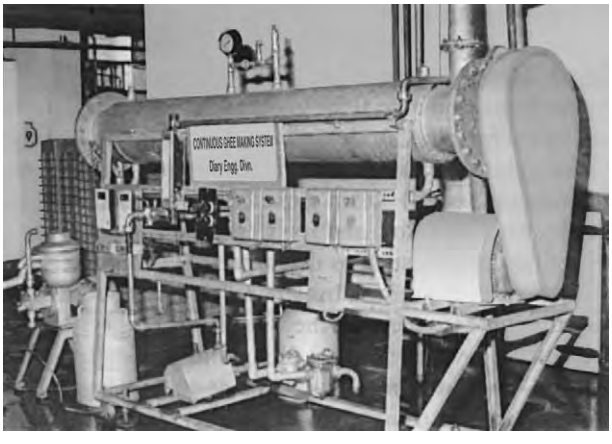

Figure 1 A continuous ghee plant. Courtesy of NDRI, Karnal, India.

Flavor

Aroma and taste constitute the flavor of ghee. A perfect ghee flavor is characterized by a multitude of sensory perceptions which are pleasant, enjoyable, and lingering in the mouth. Consumers always resist any change in the flavor of ghee as this is one characteristic which predominantly determines acceptability. There are regional preferences for flavor in ghee. The preferred ghee flavors range from 'slightly curdy' to 'pronounced curdy', cooked to caramelized and, at times, slightly oxidized in some quarters of the population. The village-produced ghee is characterized by a curdy flavor which lingers in the mouth. (*See* **Sensory Evaluation**: Taste.)

The quality and the amount of SNF present in the base material, as well as the intensity of heating separately and cumulatively, affect the flavor of ghee. Technologies have been developed in the author's laboratory to produce ghee with desired flavors of both 'curdy' and 'cooked' types.

The flavor of ghee is mainly contributed by the heat interaction products formed between the unfermented serum portion, comprising the native carbohydrate and protein system, and by metabolic products of the starter culture when ripened cream is used for gheemaking. The most important flavor components of ghee are as follows: FFAs, 6–12 mg g^{-1} fresh ghee; methyl ketones, 6–12 p.p.m.; a complex mixture of 44 lactones, both δ-lactones and γ-lactones, 29–43 p.p.m.; aldehydes, ketones, and alcohols. About 34 monocarbonyls were found in the flavor volatiles of ghee. The 'curdy' flavor in ghee could be produced by mixing of desi ghee with dairy ghee in varying proportions, by the addition of lactic cultures to butter at heating, by incubating the molten butter with a lactic culture, or by the addition of 'lassi' powder at the time of heating. 'Cooked' flavor could be simulated by clarifying butter at temperatures of 115 °C for 10 min, or 120 °C for 5 min, or 125 °C without any holding time. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Texture

Granulation of ghee is an important criterion for its selection; a good grainy texture is very much appreciated by consumers, and such a ghee develops a lower degree of rancidity than ghee kept in the liquid state. Milk fat has the unique property of forming grains because it is made up of a wide variety of complex triacylglycerol mixtures with varying melting points. The texture of ghee will depend on the source of the fat (animal species), method of preparation, temperature of clarification, rate of cooling, amount of FFAs, rate of seeding, and storage temperature. The presence of FFAs markedly increases the grain size, but the quantity of grains is only increased to a limited extent. Seeding with grains of ghee at the rate of 2% by weight of ghee improves grain formation. The grain shape becomes needle-like, in contrast to the spherical shapes obtained without seeding. The large number of fatty acid residues present in ghee results in a wide variety of crystallization patterns.

The maximum amount of solid fraction (about 74%) is obtained at 28 °C in 20–24 h from buffaloes' ghee, closely followed by cows' ghee (69.5%) and a distinct low in goats' ghee (30%). There can be significant differences in the melting curves of fat from the milk of buffaloes, cows, and goats. The changes in the conditions of cooling can have a pronounced effect on ghee texture. If ghee is cooled rapidly, a larger number of very fine crystals will be formed, all consisting of a mixture of high- and low-melting fats, leading to a smooth, grease-like character. Slow cooling of ghee from a temperature higher than the melting point will lead to the formation of few crystals with a high melting point. As cooling proceeds, more and more fat solidifies, forming a mass of large crystals suspended in liquid fat.

Color

Buffaloes' ghee appears whitish in color owing to the absence of carotene, which imparts a yellow color to cows' ghee. In the village method of gheemaking, the development of a greenish-yellow tinge in buffaloes' ghee is caused by the action of lactic acid bacteria. Ghee produced by the DC method has a brownish color compared to that prepared by the CB process. Stratification results in a light color. A more intense heating in the presence of a high SNF content will result in a darker color, especially if the raw material has been fermented. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments.)

Shelf-Life of Ghee

Although ghee has a better capacity to resist spoilage by elemental and microbial attack than any other milk product, it is common knowledge that, upon prolonged storage at ambient temperature, it undergoes oxidative changes. Reaction of oxygen with the 'unsaturated fat' is a major cause of spoilage. It gives rise to a typical, strong and disagreeable odor. Production of off-flavors accompanies the loss of nutritive value. (*See* **Oxidation of Food Components**.)

Autooxidation of ghee is aggravated by metallic contamination and sunlight. The 'acceleration' effect of light is dependent on its wavelength, and visible light accelerates the decomposition of hydroperoxides. The effect of ultraviolet light on ghee is more pronounced than the impact of other rays. High-energy radiations, such as β- and γ-rays, exert a pronounced acceleration effect because they split hydroperoxides and also generate free radicals from molecules of unoxidized substrate.

The durability of ghee is affected by the degree of unsaturation of the fat, the temperature at which the ghee is stored, the manner in which the milk for gheemaking was handled, uncontrolled fermentation during curdling, uneven heating during manufacture, and insanitary conditions of the vessels used for the production and storage of ghee.

A number of synthetic antioxidants, such as gallates (ethyl, propyl, octyl), butylated hydroxytoluene (BHT), butylated hydroxyanisole (BHA), tertiary butyl hydroquinone (TBHQ), ascorbic acid, α-tocopherol, phospholipids, and some natural antioxidants, namely curry leaves, betel leaves, soya bean powder, safflower, and amla (*Phyllanthus amblica*), can be added in small amounts (permitted legally by different countries) with a view to either complete prevention or partial retardation of the oxidation of fat during storage. (*See* **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants.)

The antioxidative properties of curry and betel leaves are attributed to their phenolic compounds, predominantly hydroxychavicol. These leaves also contain ascorbic acid, which may act synergistically. Curry leaves and betel leaves also contain many amino acids which serve as antioxidants. Studies have proved that the age-old practice of boiling betel and curry leaves with desi butter at the time of clarification helps to improve the flavor, color, and shelf-life of ghee.

Increasing the Durability of Ghee

1. Ghee should be prepared from a good-quality cream or butter.
2. If possible, oxygen should be replaced by nitrogen gas.
3. Some permitted antioxidants should be added before crystallization.

4. Feeding of green fodders and cottonseeds to lactating cows or buffaloes helps to increase the keeping quality.
5. Transparent containers should not be used for packaging ghee.
6. Ghee should be stored in a cool, dark place, preferably at temperatures below 20 °C.

Packaging of Ghee

Ghee is generally packed in lacquered tin cans of various capacities ranging from 250 g to 15 kg. Tin cans protect the product against tampering, and allow transport to far-off places without any significant wastage; they can be printed with attractive and colorful designs. However, tin cans are very expensive. Some plants do pack ghee in metallized, polyester pouches, but these also work out to be expensive owing to the aluminum coating. High-density polyethylene and polypropylene are known to have low water vapor transmission rates and are inexpensive. Such films can be laminated with other suitable, basic packaging materials. Ghee packed in flexible pouches should be placed in cartons that contain some cushioning matter to absorb vibrations during transportation or rough handling. Ghee can also be packaged in polymer-coated cellophane, polyester, nylon-6, food-grade polyvinyl chloride (PVC) or various laminates.

Bulk transport of ghee in tankers, and distribution at the consuming centers in retail containers, could be ideal for interstate or intercontinental marketing. A suitable pumping device is needed to transfer the ghee from the bulk tank without affecting the texture and granular structure of the ghee. Bulk dispensing of ghee, as used for milk, could also become an alternative system for retailing.

Chemical Composition of Ghee

Ghee is largely made up of triacylglycerols (98%) and other minor constituents, such as (1) diacylglycerols (1–2%), (2) monoacylglycerols (0.1–0.2%), (3) FFAs (1–10 mg $100\,g^{-1}$), (4) phospholipids (0–80 mg $100\,g^{-1}$), (5) cholesterol (0.25–0.4%), (6) fat-soluble vitamins, (7) carbonyls (4–6 $\mu g\ g^{-1}$), (8) glyceryl ethers ($0.8\,\mu mol\ g^{-1}$), and (9) alcohol ($1.8–2.3\,\mu mol\ g^{-1}$). (*See* **Cholesterol**: Properties and Determination; **Fatty Acids**: Properties; **Phospholipids**: Properties and Occurrence; **Triglycerides**: Structures and Properties; **Vitamins**: Overview.)

The presence of about 500 fatty acids in milk fat has been reported and the majority possess even-numbered carbon chains, $C_{4–18}$. The total proportion of polyunsaturated fatty acids is only 3–4% in ghee. Buffaloes' ghee is richer in butyric, palmitic, stearic, tetraenoic, and pentaenoic acids than the fat of cows' milk. The levels of other short-chain fatty acids (caproic to myristic) are higher in cows' ghee than in buffaloes' ghee. The fatty acid composition greatly influences the fat constants and rheological properties, such as melting and crystallization behavior, solubility, surface activity, and ability to form emulsions. Table 1 shows the major analytical constants for buffaloes' ghee and cows' ghee.

Table 1 Average analytical constants of ghee

Constant	*Buffalo*	*Cow*
Butyrorefractometer reading	42.0	43.2
Saponification value	230.1	227.3
Reichert–Meissl value	32.3	26.7
Polenske value	1.41	1.76
Kirschner value	28.5	22.1
Iodine value	29.4	33.7
Melting point	32–43	28–41

Nutritive Value of Ghee

Milk fat performs many definite functions in the human diets. It is a rich source of fat-soluble vitamins, essential fatty acids, and other growth-promoting factors. It is believed that ghee improves mental power, physical appearance, and longevity. It increases the capacity to work and enhances the body's resistance to stresses such as cold and undernourishment. Ghee is used as a curative agent for wounds, ulcers, and eye ailments. The conjugated linoleic acid content of milk fat, which is known to be anticarcinogenic, increases substantially during the manufacture of ghee. Addition of ghee to foods improves the satiety value and consumers enjoy the feeling of eating. (*See* **Fatty Acids**: Dietary Importance.)

Anhydrous Milk Fat (AMF)

AMF (anhydrous butterfat, butter oil) is composed of the products exclusively obtained from prime-quality milk, cream, or butter, and resulting from the removal of almost the entire water and SNF contents. The highest-grade AMF suitable for the production of recombined milk and other milk products is produced from fresh milk or cream (rather than from butter) which has not undergone any appreciable microbiological or chemical changes. The moisture content, FFA level, and peroxide value must be very low to guarantee keeping quality. The essential compositional and quality factors for the different grades of AMF products are given in Table 2.

Anhydrous milk fat is obtained from prime-quality raw materials to which no neutralizing substances have been added. The maximum copper content is 0.05 p.p.m., and that of iron 0.2 p.p.m. The peroxide

Table 2 Composition of milk-fat products

Product	*Fat (%)*	*Moisture (%)*	*Maximum peroxide value per kg of fat*	*Fat-free dry mass*	*FFAs (maximum %)*
Anhydrous milk fat					
Grade I	99.8	0.1	0.2	0.2	0.3
Grade II	99.3	0.5	0.3	0.7	
Grade III (butter oil)	99.6	0.3	0.8	0.4	
Dry butter	96.0	0.3		4.0	
Ghee	99.5	0.5			3.0

FFAs, free fatty acids.

value (in mmd O_2 per kg fat) should not be greater than 0.2. Coliforms should be absent in 1 g of product. It must have a clean, bland taste with a smooth and fine-grained structure.

Methods of AMF Production

Only well-equipped dairy plants can produce pure milk fat to international standards, and the production methods, of which there are several, are classified in terms of three basic processes: (1) direct; (2) semidirect; and (3) indirect.

Direct process In the direct process, pure milk fat is made directly from milk. In a typical process, milk is pasteurized, using a regenerative process of heating, and separated into skimmed milk and cream (usually containing 35% fat). The cream is pumped through a plate pasteurizer, using a temperature of at least 75 °C for 15 s, and cooled by regeneration to 55–58 °C. It is then pumped to a clarifixator where it is concentrated to a very high fat content, and the emulsion breaks. The clarifixator has two outlets, one for skimmed milk and the other for oil. The oil phase (80–85% fat) is pumped, via a balance tank, to the fat concentrator, and, after preheating to 90–91 °C in a plate heat exchanger, is pumped to a vacuum chamber for dehydration to below 0.1% moisture. The finished product, extracted from the vacuum chamber, is cooled to 21 °C before packaging in large metal cans or drums. Nitrogen sparging of the AMF is carried out by injecting nitrogen gas at a pressure of 100–140 kPa into the oil-line after the final dehydration stage.

Semidirect process In a semidirect process, cream is fed, with the help of a positive displacement pump, to a continuous buttermaking plant. The butter granules then pass to a jacketed, hopper-dough pump, to a combination in order to condition the fat and make melting by a heat exchanger possible. Using a centrifugation system, desludging is carried out to remove all the solids and most of the aqueous phase. The semidry oil plus added water is returned to a heat exchanger for temperature adjustment. The contents are, thereafter, fed to another centrifuge which removes the last traces of solids and reduces the moisture content of the fat to less than 0.5%. Finally, the oil is passed through a vacuum vessel to reduce the moisture to less than 0.1%.

Indirect process In the indirect process, butter is used as a raw material. Butter is melted in steam-heated melting kettles or revolving plate-type 'melters.' The plate rotates at 150 rpm, throwing the molten butter on to the wall of the melter, where it drains into a collection trough. The molten butter is passed through a stainless-steel mesh filter to remove any extraneous matter, such as wrapping material, and is then pumped to the agitated, water-jacketed holding vats at 55–60 °C. Neutralization, if required, is carried out prior to separation in the holding vat and before pumping to centrifuge. Molten butter is pumped to a Sharples centrifuge for reduction of the total SNF and water content to about 0.5–0.8%. Fat from the centrifuge is washed with hot water at 82 °C. The fat–water mix is then passed to a battery of six Sharples vapor-type super centrifuges, or a set of oil clarifiers, to accomplish the complete removal of SNF and water. The oil is finally vacuum-treated to remove the remaining moisture.

Uses of AMF

AMF can be utilized in a number of ways. It is particularly suitable for use in warm climates and for export. It is used widely in making recombined milk and recombined milk products, and as an ingredient in many processed foods of commercial value. In addition, AMF can be converted into ghee or flavored with simulated ghee flavors.

Conversion of AMF into Ghee

In many countries, surplus milk fat is conserved in the form of butter oil, rather than butter, because of savings in cold storage capacity and the anticipated long shelf-life at ambient temperature. AMF lacks the characteristic flavor, aroma, and texture which are otherwise required in ghee, although the chemical composition is almost identical to that of ghee. A scheme has been suggested for producing ghee from butteroil by the addition of skim milk dahi (20% w/w) or dahi powder (5% w/w) to butteroil and heating the mixture to 120 °C for 3 min. It is also possible to incorporate ghee flavor concentrate in various proportions to butteroil. A synthetic flavor was

constituted to induce a ghee-like flavor in butteroil. The synthetic mixture consisted of δ-(G-10)-lactone (3 p.p.m.), δ-(G-12)-lactone (15 p.p.m.), decanoic acid (5 p.p.m.), and kenanone-2 (10 p.p.m.). The shelf-life of flavored butteroil was improved from 10 weeks to about 6 months by adding BHA (0.02%). A 'cooked' flavor can be induced by heating the ghee residue (10%) with butteroil at 120 °C for 30 min.

See also: **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; **Buffalo**: Milk; **Cholesterol**: Properties and Determination; **Fatty Acids**: Properties; Dietary Importance; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Lactic Acid Bacteria**; **Milk**: Dietary Importance; **Oxidation of Food Components**; **Phospholipids**: Properties and Occurrence; **Sensory Evaluation**: Taste; **Sheep**: Milk; **Triglycerides**: Structures and Properties; **Vitamins**: Overview

Further Reading

Abichandani H, Sarma SC and Bector BS (1991) Continuous ghee manufacture – an engineering solution. *Indian Food Industry* 10(4): 35.

Rajorhia GS and Sherwon JW (1986) Effect of processing conditions on the melting properties of ghee. *Indian Dairyman* 38(8): 381.

Rama Murthy MK (1980) Factors affecting the composition, flavour and textural properties of ghee. *Indian Dairyman* 32: 765.

Ray SC (1959) *Prestratification Method of Ghee Making*. ICAR Bulletin. New Delhi: Indian Council of Agriculture Research.

Wadhwa BK and Jain MK (1985) Simulation of ghee flavour in butter oil with synthetic flavouring compounds. *Journal of Food Science Technology* 22: 24.

Wadhwa BK and Jain MK (1991) Production of ghee from butter oil – a review. *Indian Journal of Dairy Science* 44(6): 372.

GIN

Contents

The Product and its Manufacture

R I Aylott, Tillyrie, Ramoyle, Dunblane, UK

Background

Gin is a distilled spirit drink. It is a colorless liquid containing at least 37.5% alcohol that has its flavor principally derived from juniper berries. It is usually mixed for drinking with tonic water, ice and lemon, or with other alcoholic beverages such as vermouth to make cocktails such as dry Martini.

Distilled gin is traditionally made by distilling alcohol in the presence of juniper berries and other botanical ingredients. Coriander seeds are usually included, together with ingredients such as angelica root, herbs, and orange and lemon peels. As the alcohol boils in the gin still, it extracts flavors from botanical materials, and these condense with the alcohol in the distillate. The alcoholic strength of the gin distillate simply has to be reduced with water down to the required bottling strength, and the gin is ready. There is no maturation process in wooded casks as is used with brown spirits such as whisky, brandy, and rum.

Gin as an English product dates back to the late seventeenth century. It was introduced into England by soldiers returning from wars in the Low Countries where wine and grain spirits were flavored with juniper to make 'jenever' (a corruption of the French word 'genievre'). In those days, the quality of the available alcohol was poor compared with modern alcohol, and the juniper flavor no doubt served to make the pungent alcohol more palatable.

When gin was introduced into England, the most popular alcoholic drinks, other than home-produced beers, were wines and brandies from France. The Dutchman William of Orange became King of England. Strong anti-French anti-Catholic interests at that time helped push the 1690 Distilling Act through Parliament. This act banned the import of French wine and brandy and encouraged the distillation of spirits from grain. Further changes in laws and fashion led to a vast increase in the production and consumption of gin.

The early eighteenth century was the time of a gin craze in England. Gin was often sweetened with sugar or glycerin and was known as 'Old Tom.' Consumption is reported to have increased from 0.5 million gallons in 1689 to 5 million gallons in 1727. By 1733, 11 million gallons were produced in London alone – that was 14 gallons a year for each resident.

Production in London, whose population was 500 000, increased to 20 million gallons by 1743. This enormous consumption of cheap liquor obviously had major social effects, especially on the poor people who already lived under bad conditions. Parliament tried to control consumption with the Gin Acts of 1736 and 1743. Prices rose, and this had the effect of reducing demand.

Quality started to increase, and the end of the eighteenth century/early nineteenth century was the time of the foundation of many of the firms carrying the popular brand names of today. Gin became socially acceptable, being drunk by the servants of the British Empire. Export trade developed, later involving overseas distilling and packaging operations.

Modern neutral alcohols are much purer than their equivalents in earlier centuries and now enable the sensory characteristics of juniper berries and other gin botanicals to be the primary source of gin flavor. These unique qualities help to make gin the internationally popular spirit it is today.

Definitions and Regulations

Gin and other major distilled spirits are defined in law. These definitions serve to protect the interests of consumers, protect gin as a generic product, and protect the interests of the gin distiller.

The law within the European Union (Council Regulation No. 1576 of 1989) is particularly important, as that is the definition where the generic product originated. In summary, a drink may be called *gin* if it is produced by flavoring *ethyl alcohol of agricultural origin* with natural and/or nature identical flavoring substances so that the taste is predominantly that of juniper. *Distilled gin* is produced by redistilling the alcohol in stills traditionally used for gin in the presence of juniper berries and of other natural botanicals provided that the juniper taste is predominant. The term *distilled gin* also applies to mixtures of distilled gin and similar alcohol of agricultural origin. Natural and/or nature identical substances may also be used. *London gin* is a type of distilled gin (and not a geographical location of manufacture), whereas *Plymouth gin* is a geographical designation. Gin simply made by adding essences or flavorings to alcohol may not be called distilled gin – this product is known in the industry as a compounded gin. The minimum bottling strength for all gin is 37.5% vol.

The regulations of other countries reflect much of the European law. The United States Bureau of Alcohol, Tobacco and Firearms distilled gin definition requires all the alcohol to be distilled with the botanical materials and the minimum bottling strength to be 40% vol. (80° US proof). Canada also has a 40% minimum bottling strength, whereas in Australia, it is 37% vol.

Materials for Gin Distillation

The three key ingredients for gin distillation are botanical materials, alcohol, and water. Each gin distiller has a specific botanical recipe for each brand and will have exacting quality requirements.

Figure 1 (see color plate 78) Juniper berries, the essential botanical ingredient in gin, together with commonly used coriander seeds.

Gin botanical ingredients are subject to rigorous quality assessment before purchase, are often stored for 2 years before use, and are carefully blended to ensure minimal seasonal and batch to batch variation. The essential ingredients are juniper berries (*Juniperus communis*). Although juniper is dispersed widely, it is usually harvested for distilling in Italy and the Balkans. Coriander seeds (*Coriandrum sativum*) are sourced from Morocco and Russia, angelica root (*Archangelica officinalis*) from Germany, and orange and lemon peels from Mediterranean countries. Drawings of juniper and coriander are shown in Figure 1. Other botanical materials used in the recipes of specific gins include aniseed, calamus root, caraway seed, cassia bark, cardamom seeds, cinnamon, fennel seed, liquorice root, nutmeg, orris root, and herbs.

The alcohol used for gin distilling is known as neutral alcohol. It is of agricultural origin and is commonly fermented from cereals or molasses. Gin distilled from grain neutral spirits is often sold at a premium. Neutral alcohol is produced on continuous distillation columns and is highly rectified. The resulting distillate is essentially very pure ethanol with only very small traces of congeners (that is, other compounds from the fermentation). The characteristics of ethyl alcohol of agricultural origin are defined within Annex 1 of regulation 1576/89, and the minimum alcoholic strength after final rectification is 96% vol. Production of the neutral alcohol is a separate process from gin distilling and is often conducted at separate distilleries and by different producers.

Water is used in two stages of gin production. First, it is added to the neutral alcohol in the still prior to the start of distillation, and second, it is added to the distillate in order to bring the gin distillate down to the required bottling strength. In both cases, high-purity water is required. It must be clear, odorless, and low in trace dissolved minerals. It is common for such water to be taken from a local supply and purified by demineralization and carbon treatment processes. Historically, many of the London gin distilleries were sited where wells could provide good sources of water.

Gin Distillation

Distilled gin is produced in a batch process in a traditional copper still similar to those used for Scotch malt whisky. Figure 2 shows a schematic diagram of

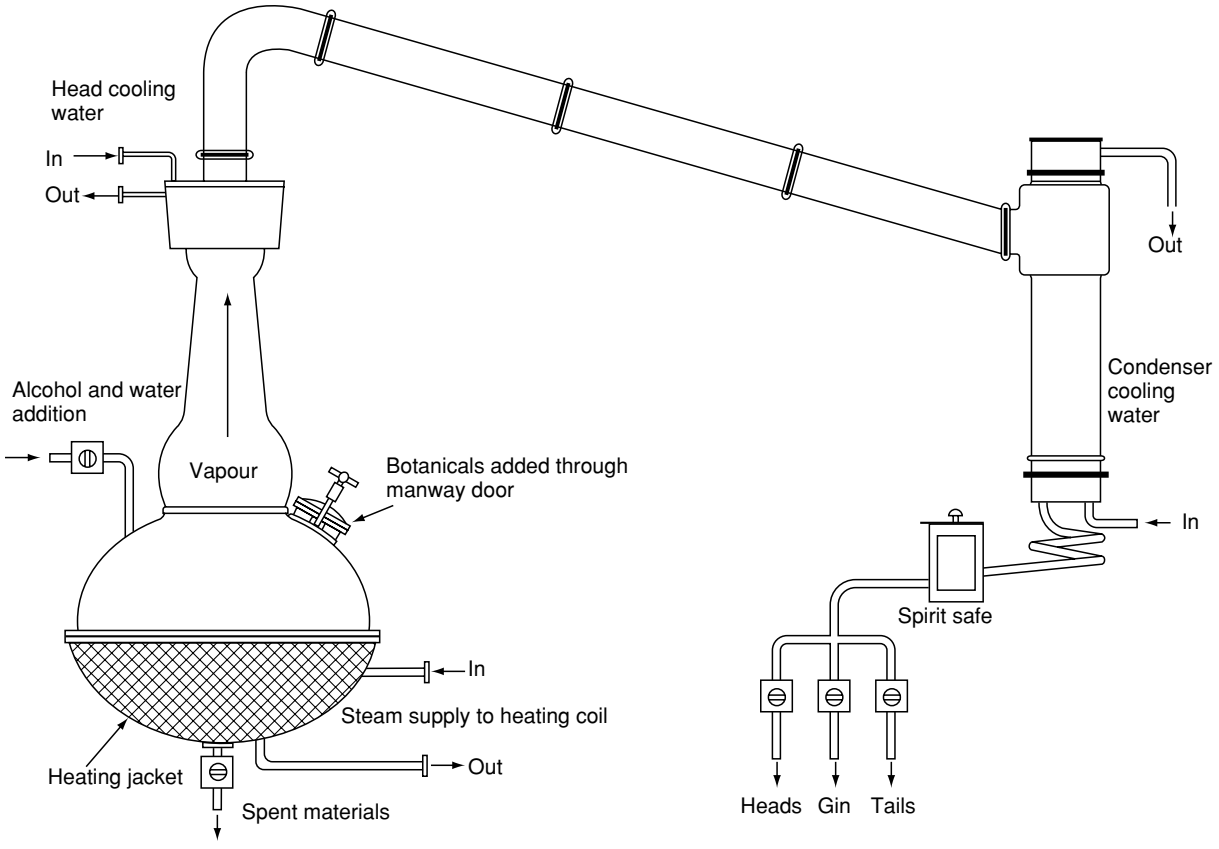

Figure 2 Schematic diagram of a typical gin still.

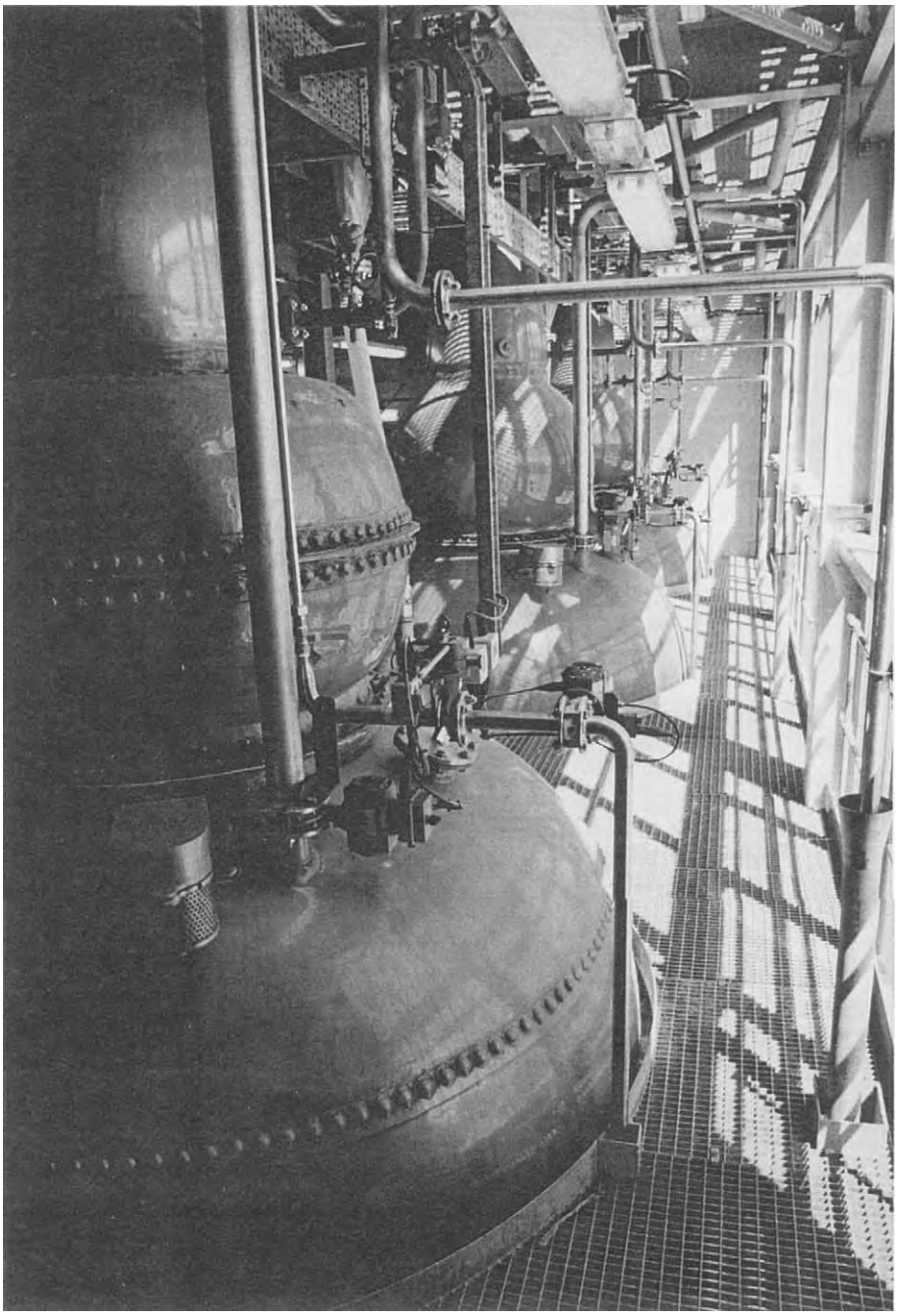

Figure 3 Modern Tanqueray Gordon distillery equipped with old traditional gin stills.

the process, and Figure 3 shows a photograph of the modern Tanqueray Gordon distillery equipped with old traditional gin stills.

The start of the distillation process involves charging the still with water followed by neutral alcohol in order to achieve an alcoholic strength around 60% vol. Botanical ingredients are then added into the still, either loose or in a bag suspended above the liquid. The still is closed, and heat is applied, nowadays using a steam coil. As the liquid starts to boil, the alcoholic strength of the distillate is greatest and contains the most volatile botanical congeners (mainly monoterpenes). This first fraction is known as 'heads' and is kept separate, as it is not required. The second fraction is the main product and is collected as distilled gin at around 80% vol. alcoholic strength. This fraction contains all the congeners found in gin, including monoterpenes, terpineols, and sesquiterpenes). Lastly, as the alcohol content of the distillate starts to fall, the temperature increases,

and the character of the distillate changes as the less desirable low volatile congeners begin to boil. The third 'tails' fraction is then collected separately with extra heat applied to the still to recover as much alcohol as possible. The heads and tails fractions are combined as *feints* and subject to a separate alcohol recovery process.

Throughout the process, the gin distiller will carefully control the rate of heating, monitor the alcoholic strength of the distillate and check the sensory properties of the gin. The gin distillate may then be simply diluted with water down to bottling strength (to produce a 'one shot' gin) or may be blended with other distillates and alcohol to produce the final distilled gin.

Lastly, compounded gin may be made by adding gin essences or flavoring to alcohol. This is the simplest way of making a cheaper product. The essences are normally made by a specialist flavor company according to a defined recipe. No further processing is required except reduction to bottling strength. Such products may not be called distilled gin.

Packaging and Distribution

Immediately before bottling, gin is reduced with water to the required bottling strength and passed through a fine filter. Final alcoholic strengths commonly range from a minimum of 37.5 and 40% in domestic markets up to 47.3% vol. in duty free shops. Gin is normally packaged in glass or PET (polyethylene terephthalate) plastic bottles in common with other popular distilled spirits. The most popular bottle sizes are 700 ml, 750 ml, and 1 l. Fifty-milliliter PET bottles are common in the airline trade. Gin is often distilled at one site and packaged at another. Many brands are shipped overseas at high alcoholic distillation strength in stainless steel tanks for local bottling in order to achieve import tax and transport cost savings by using locally sourced packaging. Gin is a very stable product undergoing no degradation when stored under good conditions in a properly sealed bottle.

Flavored Gins

A range of specialty products are based upon gin and are made by both commercial and cottage industries. Sloe gin is traditionally made by steeping sloe berries (*Prunus spinosa*) in gin in order to extract its natural flavors and juices from the fruit. Similarly, orange and lemon gins are made by steeping the peels of these fruits in gin. After a number of weeks, the steeps are filtered, sweetened to taste with sugar, and reduced in alcoholic strength with water ready for bottling. Ingredient variations on these processes include the use of fruit juices and flavorings.

Table 1 Major eight gin brands, their key markets, and sales in 2000

Brand	*Region*	*Sales (millions of 9 liter cases)*
Ginebra San Miguel	Philippines	27.3
Gordon's gin	International	5.05
Seagram's gin	USA	3.42
Gilbey's	International	2.25
Beefeater	International	2.20
Larios gin	Spain	2.15
Tanqueray gin	International	1.83
Bombay Sapphire gin	International	1.82

From: Drinks International 2001, The Club 2000.

Gin Today

Gin is one of the major international generic distilled spirits with only vodka and Scotch whisky selling greater volumes. Table 1 lists the major eight gin brands in the year 2000.

In addition, there are a large number of other local brands to be found plus many own-label brands produced for sale in supermarkets. The more expensive brands tend to be distilled gins, and the cheapest brands are often made using compounding processes.

As with most distilled spirits, many rival products seek to imitate the leading brands. Analytical methods have been developed by which the authenticity of suspect samples may be checked, thus helping protect the interests of the consumer and brand owner.

See also: **Alcohol**: Properties and Determination; Metabolism, Beneficial Effects, and Toxicology; Alcohol Consumption; **Legislation**: History

Further Reading

Aylott RI (1995a) Flavoured spirits. In: Lea AGH and Piggott JR (eds) *Fermented Beverage Production*, p. 275. Glasgow: Blackie Academic & Professional.

Aylott RI (1995b) Analytical strategies to confirm gin authenticity. *Journal of the Association of Public Analysts* 31: 179.

Clutton DW (1979) The Production of Gin and Vodka. *Brewers Guardian* 108(10): 25.

Coates G (2000) *Classic Gin*. London: Prion Books.

European Communities (1989) Council Regulation 1576/89 laying down general rules on the definition, description and presentation of spirit drinks. *Official Journal of the European Communities* L:160/1.

Simpson AC (1977) Gin and vodka. In: Rose AH (ed.) *Alcoholic Beverages*, p. 537. London: Academic Press.

Composition and Analysis

E Guerra Hernández, Universidad de Granada, Granada, Spain

Introduction

Gin is a colorless spirit obtained by distilling an aqueous mixture of alcohol together with aromatic plant materials, generally juniper berries (*Juniperus communis* L.), to which water and alcohol and at times fruit juices, extracts, and/or essential oils of fruits may be added. According to this definition there are more than 10 different types of gins on the market, produced in various parts of the world. The components of gin are water, alcohol, and compounds derived from the plant materials, essential oils and fruits used in its production.

Composition

Alcohol

The carbohydrate source for obtaining alcohol by fermentation is a grain mash consisting of corn, malt, and rye in varying proportions according to the country of production. Alcohol obtained in this way is purified by distillation to eliminate the flavor of the fermentation products until a clear, clean, and neutral alcohol in taste and odor is obtained. Following this procedure, only minor traces of aldehydes, esters, and higher alcohols are found in the majority of gins, with the exception of Dutch gin, which is made from unrefined alcohol.

Plant Ingredients

The main plant ingredients used are juniper berries together with coriander seeds and angelica root. The precise proportion in which the distilleries use these three components in combination with other plant ingredients, such as anise seed, bitter orange peel, cinnamon bark and cassia bark, is a secret of the distillery, since this is what determines the distinctive taste of each gin.

The essential oil of these ingredients gives gin its characteristic flavor and is obtained by distilling alcohol together with the particular plant ingredients (distilled gin) or, if the essential oil is already in liquid oil form, it is added to the alcohol, as in the case of compound gin.

This essential oil contains terpenes (α-pinene, myrcene, limonene, γ-terpinene, linalool, terpine-4-ol, geraniol, borneol) and their proportion varies according to the combination in which the plant materials are used.

In summary, all types of gin contain, as main components, ethyl alcohol (38–50% alcohol by volume) and water, and in smaller quantities terpenic compounds, which are precisely those which define it and distinguish it from any other alcoholic beverage. Only a few gins contain sucrose, and in others esters, aldehydes, and higher alcohols are found in appreciable concentrations.

Flavor Components

The most common gin is the London type in which the flavor components are derived from the plant ingredients used in its production. A distilled dry gin is a distilled gin with a very small proportion of plant-derived components.

The two main plant ingredients are *J. communis* L. berries and *Coriandrum sativum* L. seeds. The main components of essential oil from these plants are terpenes which constitute the major part of the flavor of the gin. The components found in these plants and in the gins made from them are given in Table 1.

As can be observed, the main terpenic components of the gin, α-pinene and linalool, coincide with those of its plant ingredients. If only the heart of the distilled fraction is used, the proportion of the main components with low molecular weight (unoxygenated monoterpenes) from *J. communis* L. is reduced.

Some gins use not only these components but also other plant ingredients which contain a high proportion of essential oil and, in addition, possess a terpene content near to or exceeding 50%. Their effect, although they are only used in small proportions to give the gin aroma, can significantly affect the final aroma. This is the case with the use of orange oil, which contains 90% *d*-limonene, and cinnamon and cassia bark, whose oils contain 70–95% and 60–70% cinnamic aldehyde respectively.

Table 1 Range of main terpene components (expressed as relative percentage) in berries of *Juniperus communis* L. and seeds of *Coriandrum sativum* L., and main terpenes found in samples of gin

J. communis		C. sativum		*Gin*
α-Pinene	30–35	Linalool	65–75	Linalool
β-Myrcene	18–30	*p*-Cymol	5–7	α-Pinene
β-Pinene	2–14	Borneol	0.2–7	Sabinene
Limonene	4–10	α-Pinene	2.5–6.5	β-Myrcene
Terpinen-4-ol	0.1–10.5	γ-Terpinene	1.4–10	Limonene
Terpinolene	0.5–3	Geraniol	2–4	γ-Murolene
γ-Murolene	1–4.3	β-Pinene	0.3–1.4	Terpinen-4-ol
Sabinene	0.2–3	Limonene	0.1–1.7	Geraniol
γ-Cadinene	0.4–1.6	Camphene	0.4–0.8	γ-Terpinene
p-Cymene	0.1–1.5	Sabinene	0.1–0.2	α-Terpinene

The odor threshold for linalool (a major component of coriander oil) in 20% ethanol is 9 p.p.m., while for γ-terpinene it is 1.5 p.p.m.; as both have an 'orangey' aroma, the latter compound can be more dominant than linalool in the aroma.

The main component of juniper (α-pinene, threshold value 5 p.p.m.) seems to play a decisive role in the aroma, which is not the case for *p*-terpineol (threshold value 13 p.p.m.).

Limonene, although having a high odor threshold (6.5 p.p.m.), contributes decisively to the flavor of 'orangey' gins which are made with oranges and lemons. However, this is not the case with cinnamaldehyde (the major component of cassia and cinnamon bark), which is normally found in gins at concentrations below its odor threshold (3 p.p.m.).

In Dutch gin, together with the terpene hydrocarbons, the components responsible for the flavor include higher alcohols, especially isomyl alcohol and, to a lesser degree, isobutanol, propanol, ethyl acetate, and 2-butanol.

Effects of Storage

Gin does not age and it does not require a period after production to reach maturity. On the contrary, since the aromatic components of the gin are essential oils which tend to oxidize and produce the so-called terpene-like character, it is best to consume the products as soon as possible. Terpenes may also be lost owing to their high volatility.

Analytical Methods (Authentication)

The following factors determine the flavour of bottled gin: quality of initial alcohol, formulation of plant ingredients, and distillation technique. Control of the distillation technique (steam pressure, water flow, etc.) is automated in the majority of distilleries.

Analysis of the alcohol is usually restricted to the determination of extract, acidity, esters, aldehydes, methanol, higher alcohols, nitrogenated bases and furfural, analyses which can be performed using traditional chemical methods. For example, the extract is determined by weighing the samples after evaporation to dryness. The total acidity is determined by neutralization with a standard alkaline solution. Esters (expressed as ethyl acetate) are quantitated by reaction with alkaline hydroxylamine to form a hydroxamic acid which, after acidification, forms a colored complex with ferric ions. Ester concentration is proportional to absorbance at 525 nm at constant alcohol concentrations. Aldehydes (expressed as acetaldehyde) are quantitated with an excess of sodium bisulfite, addition of iodine and analysis of remaining iodine with sodium thiosulfate. Methanol is determined by measuring the absorbance at 575 nm after the addition of potassium permanganate and chromotropic acid to a sample, previously diluted at constant alcohol concentration. Higher alcohols (fusel oil) are determined by means of the formation of colored derivatives with *p*-dimethylaminobenzaldehyde, which are measured at 538–543 nm, or with the sodium salt of 4-hydroxybenzaldehyde sulfonic acid, which is then measured at 445 nm. The content in nitrogenated bases is based on fixing them using phosphoric acid followed by an initial distillation. The bases are liberated with sodium hydroxide and a second distillation is performed. The second distillate is tested with hydrochloric acid. The presence of furfural is detected by the reddish color that appears when this compound reacts with aniline in acetic medium.

These methods are acceptable if there are high levels of the compounds in question, but are not as useful when the levels are very low, in which case it is recommended that the analysis be carried out by gas chromotography using polar columns (Carbowax 1500) under isothermal conditions with flame ionization detection. This allows not only detection of components at very low levels but also individual concentration of each component. With this method aceltaldehyde, ethyl acetate, methanol, and higher alcohols can be determined simultaneously.

When the quantities of the above compounds are high, the alcohol must be purified before it is used in the production of gin.

Analysis of the plant components is normally carried out for juniper berries, coriander seeds, and angelica root. Any other plant components are distilled and compared by taste with accepted standards.

A quality essential oil made from juniper must have a refractive index between 1.4840 and 1.4870; lower levels are indicative of a high content of low-boiling terpenes, which are undesirable in the production of gin. The refractive index for coriander seed oil must be between 1.463 and 1.471 and for angelica root between 1.476 and 1.488.

Measurement of the ultraviolet absorption spectra of these oils also provides quick and useful data with regards to the quality of the oils. Thus, juniper oil absorbs strongly between 220 and 240 nm, coinciding with the absorption of one of its components, terpine-4-ol, while coriander oil begins to absorb at lower wavelengths between 200 and 225 nm. Dilutions of these essential oils in alcohol are compared by taste with a sample standard and, according to the aromatic level required in the final gin, the appropriate plant components are mixed well before proceeding with the distillation stage.

Sensory Analysis

Quality control in large distilleries is carried out by taste panels of 20–30 members. These people examine similar samples in difference tests (triangular test). This panel is used both to insure uniformity of the gin which comes to the market as well as for examining the initial alcohol and the solutions of essential oil extracted from the plant ingredients.

To control the quality and authenticity of a gin that is produced, a sensorial analysis, by means of a taste panel, is used. This provides a profile of the taste and character of the gin. According to these analyses various gins can be identified, e.g., those with a taste derived from juniper berries, and gins with a faint orange taste which corresponds to those gins that contain coriander oil; if the orange taste is strong, this indicates that it contains other ingredients which provide limonene. Finally, those flavors are labeled 'tails' that correspond to samples which contain other components such as those containing cinnamaldehyde. When gins are produced with high levels of cassia or cinnamon, or the samples similar to Dutch gin, their aroma is derived not only from the plant components but also from the initial alcohol, which gives it an aroma similar to that of whisky.

Chemical Analysis

The main ingredient together with water is ethyl alcohol, which can be determined, after distillation, by density measurement with a hydrometer.

The content of extract, total acidity, aldehydes, esters, nitrogenated bases, and furfural can be analyzed by the traditional methods mentioned early. The methanol and higher alcohols (2-butanol, 1-propanol, 2-methyl-1-propanol, 1-butanol, 2-methyl-1-butanol and 3-methyl-1-butanol) are quantified by means of chromatography in Carbowax 1500 column and with 4-methyl-2-pentanol as internal standard. (*See* **Chromatography**: Gas Chomatography.)

Analysis of reducing sugar is done by the reducing properties of the sample, previously clarified, in a cuproalkaline solution. The technique is also applicable to sucrose analysis. The sample content is established by the difference in the reducing capacity before and after inversion with hydrochloric acid. (*See* **Carbohydrates**: Classification and Properties.)

Heavy metals As, Pb, Zn, and Cu, are determined by atomic absorption spectrophotometry. (*See* **Spectroscopy**: Atomic Emission and Absorption.)

Analysis of compounds from the plant ingredients which give the gin its character is mainly by ultraviolet spectroscopy and gas chromotography. Ultraviolet spectroscopy is used to characterize gin in the same way as for the analysis of the plant ingredients; analysis between 220 and 240 nm gives information about the level of juniper oil in the gin when compared with diluted essential oil in alcohol of the same proof. Measurements between 200 and 225 nm give the level found in coriander oil. The results obtained provide a partial indication of 'total flavor level' as 'juniper' or 'coriander' because the oils used for standardization are steam-distilled products, whereas gin is distilled in ethanol and part of the terpene components of plants is rejected as 'feints' (combined heads and tails).

Moreover, this type of analysis has other disadvantages. The first is that it gives only a general idea of the flavor but no information as to the identity of the components. It does not allow a distinction to be made between gins which contain orange and coriander oil from those that only contain the latter, since both absorb at the same wavelength. Finally, this method is only applicable to gins whose initial alcohol has been purified, or else the aldehydes, esters, and higher alcohols that may be present will interfere with the spectrophotometric measurements.

The chromatographic method offers the advantage of identifying individual volatile components of the gin but the disadvantage is that the sample must be concentrated, owing to the low levels of the components, with the risk of obtaining results which may not be accurate since the proportion of the components obtained in the concentrate may not be the same as in the initial sample. There is also a risk of introducing artifacts. Furthermore, the increased time required for the analysis and material waste must be considered.

The two most important techniques for concentration of the volatile components in alcoholic beverages are liquid–liquid extraction and methods based on absorption–desorption. Both procedures have been used on gin. It has also been found that the best solvents are dichloromethane, trichlorotrifluorethane, and a 1:1 mixture of these. The results were reproducible after an extraction time of 5 h and 50 ml of solvent was used to extract 100 ml of the sample; the concentration factor obtained was 20 and the yields were 80% for monoterpenes and 100% for oxygenated monoterpenes. In order to obtain higher concentration factors, continuous extraction methods may be used. For studying concentration through absorption–desorption, charcoal columns have been used.

These two methods have been applied to the analysis of samples of commercial gin using dichloromethane as a solvent for the concentration with an orbital shaker and trichlorofluoromethane for concentration using continuous extraction. The extracts were analyzed by gas chromatography, using a packed polar column (Carbowax 20M). The number

of compounds clearly identified was six with the first procedure and more than 20 with the second, with the latter having the advantage that the extract, once constituted, was identical to the initial sample, from a sensory point of view.

Discontinuous liquid–liquid extraction for the quantitative analysis of volatile components has also been used. The solvent was diethyl ether (200 ml) and was concentrated to a final volume of the extract of 6 ml. The volume of the initial sample was 60 ml and the column utilized was polar (Carbowax 20M), but in this case using a capillary column of 50 m length.

Quantification was carried out by dissolving the terpenes in 40% alcohol by volume at different concentrations and using the same method as with the gin, thereby obtaining standard curves for each component.

In this study 21 different terpene components were quantified (α-pinene, β-pinene, sabinene, δ-3-carene, β-myrcene, limonene, 1,8-cineole, γ-terpiene, *p*-cymene, α-terpinolene, camphor, linalool, linalyl acetate, bornyl acetate, isobornyl acetate, caryophyllene, terpinen-4-ol, myrtenyl acetate, α-terpineol, geranyl acetate, geraniol). **Figure 1** shows a chromatogram obtained by this method.

The main components identified were similar to those obtained in other studies, and were (**Table 1**) linalool, α-pinene, β-myrcene, limonene, terpinen-4-ol and geraniol, with the exception of γ-murolene, a component not found in gins analyzed, but which was detected in the juniper berries in an appreciable quantity and which could have been an indicator of the quality of the gin.

In summary, the only components identified in gin, together with alcohol and water, are the terpenes and these, when found at low levels, require concentration prior to analysis. For the determination of individual compounds a gas chromotographic analysis with high-resolution columns is necessary due to the structural similarity of the terpenes.

Together with these components, the Dutch gins contain higher alcohols (propanol, isobutanol, and isoamy alcohol) in such concentrations (70–90 mg l^{-1}, 70–115 mg l^{-1}, and 315–320 mg l^{-1} respectively) that it is possible to determine them directly in the sample by gas chromotography.

The use of terpenic compounds permits differentiation between gin brands by applying several multivariate statistical methods (principal component; linear, quadratic and nearest neighbor discriminant; cluster analysis and outlier detection). The distilleries reproduce the process of elaboration of gin to obtain uniform terpenic coumpounds and flavor gin.

The chemical components of the gin for which limits have been established, because they reduce the quality of the gin or because they may be toxic are acidity, which must not exceed 10 mg l^{-1}, expressed as acetic acid; methanol, which must not exceed

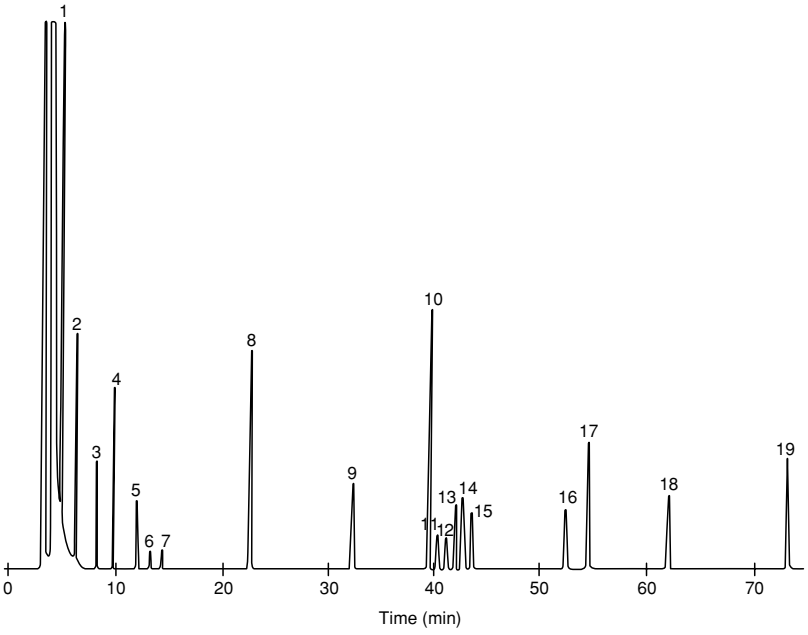

Figure 1 Chromatogram of terpene components found in a sample of commercial gin: (1) α-pinene; (2) β-pinene; (3) β-myrcene; (4) limonene; (5) γ-terpinene; (6) *p*-cymene; (7) α-terpinolene; (8) methyl caprylate (internal standard); (9) camphor; (10) linalool; (11) linalyl acetate; (12) bornyl acetate; (13) isobornyl acetate; (14) caryophyllene; (15) terpinen-4-ol; (16) myrtenyl acetate; (17) α-terpineol; (18) geranyl acetate; (19) geraniol.

$1\,g\,l^{-1}$ (determined chromatographically); and heavy metals As, Pb, Zn, and Cu, which must not exceed concentrations of 0.5 p.p.m. for the first two and 10 p.p.m. for the last two.

The concentration of reducing material must not exceed $35\,g\,l^{-1}$. Finally, commercial gins must not contain apreciable amounts of nitrogenated compounds or furfural.

See also: **Alcohol**: Properties and Determination; **Chromatography**: Gas Chromatography; **Essential Oils**: Properties and Uses; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Herbs**: Herbs and Their Uses; **Sensory Evaluation**: Aroma; **Whisky, Whiskey, and Bourbon**: Composition and Analysis of Whisky

Further Reading

Barros C, Bahlsen C and Marin E (eds) (1998) *Legislacion Alimentaria*, vol. XXX, *Official Methods of Gin Analysis*, pp. 11–24. Madrid: Sid Alimentaria.

Clutton DW (1972) The history of gin. *Flavour Industry* 3: 454–456.

Clutton DW and Evans MB (1978) The flavour constituents of gin. *Journal of Chromatography* 167: 409–419.

Guerra Hernández E, Lopez Martínez MC and García Villanova R (1987) Analytical study in samples of commercial gins. *Anales de Bromatología* 39: 219–227.

Guerra Hernández E, Lopez Martínez MC and García Villanova R (1987) Volatile components identified by gas chromatography of digestion products of *Juniperus* in ethanol. *Anales de Bromatolía* 39: 229–237.

Helrich K (ed.) (1990) AOAC. *Official Methods of Analysis of the Association of Official Analytical Chemists*, 15th edn, pp. 690–707. Virginia: Association of Official Analytical Chemists.

Hui YH (ed.) (1992) *Encyclopaedia of Food Science and Technology* vol. 1, p. 594. New York: John Wiley.

Martínez-Alvarez PJ and Herranz A (1991) Application of multivariate statistical methods to the differentiation of gin brands. *Journal of the Science of Food and Agriculture* 57: 263–272.

Simpson AC (1966) Gin manufacture. *Process Biochemistry* 10: 355–365.

GLUCOSE

Contents

Properties and Analysis

C V K Sreeranjit and J J Lal, Centre for Development of Science & Technology, Thrissur, Kerala, India

Background

Glucose is a word derived from the Greek word 'gleukos' meaning sweet wine. The term glucose was introduced by André Dumas in 1838 to refer to the sweet compound obtained from honey and grapes. Later, Frederich August Kekulé von Stradonitz called it dextrose because the naturally occurring glucose is dextrorotatory in nature. When Fisher studied this sugar, he called it glucose, and chemists have used the same word ever since.

Glucose is a widely distributed, sweet, simple sugar (monosaccharide) synthesized in plants by photosynthesis and usually stored as polysaccharides. In animals dietary carbohydrates, after digestion, diffuse into the blood in the form of glucose and are assimilated by the liver. In the liver, glucose is used for the synthesis of other carbohydrates, used as a major fuel (for energy production) for the tissues of organisms, or converted to other carbohydrates with highly specific functions such as glycogen for storage, ribose in nucleic acid, galactose in lactose of milk, in certain complex lipids (glycolipids), and in combination with proteins in glycoproteins and proteoglycans. The storage form of glucose is glycogen (liver or muscle glycogen) in animals, and starch or cellulose in plants.

Chemical Structure

The structure of glucose can be represented in different ways as a straight chain, simple ring, or chair form. The straight-chain form (aldo sugars are written with the aldehyde group at the top and the primary alcohol at the bottom) can account for some of

the properties of glucose, but a cyclic structure is favored on thermodynamic grounds. X-ray diffraction analysis has revealed that it has chair form (Figures 1–3).

Isomerism

Compounds that have the same structural formula but differ in their spatial configuration are known as stereoisomers. This phenomenon is mainly due to the presence of an asymmetric carbon atom in the compound. In the glucose molecule, there are four such asymmetric carbon atoms, and so there are 16 isomers in nature.

D and L forms The orientation of the —H and —OH groups around the carbon atom adjacent to the terminal primary alcohol carbon determines whether it belongs to the D or L series. In this spatial relationship, if the —OH group on the fifth carbon is on the right, it is D (dextrorotatory)-glucose, and if it is on the left, it is L (levorotatory)-glucose (Figure 4), i.e., the presence of asymmetric carbon atoms confers the optical activity of glucose. When a beam of plane-polarized light passes through glucose solution, it is rotated either to the right (dextrorotatory) (+) or to the left (levorotatory) (−), but if the mixture contains an equal quantity of dextro- and levorotatory forms (a DL mixture), it does not exhibit any optical activity, since the activities of each isomer cancel each other out. This mixture of isomers is also called a racemic mixture.

Pyranose and furanose forms In some cases, glucose is found in a stable ring form similar to pyran (α-D-glucopyranose) or furan (α-D-glucofuranose) as in Figures 5 and 6. In solution, about 99% of glucose is in the pyranose form, and the remainder is in the furanose form.

α and β anomers A ring structure, called a hemiacetal, is formed by the combination of an aldehyde

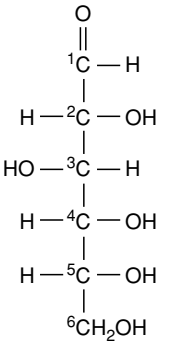

Figure 1 Straight chain.

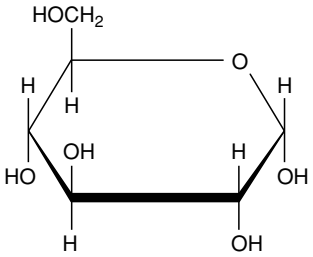

Figure 2 Ring form.

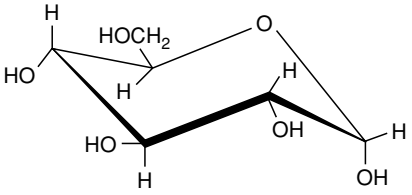

Figure 3 Chair form.

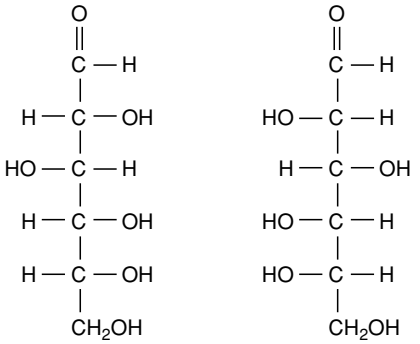

Figure 4 D-Glucose and L-glucose.

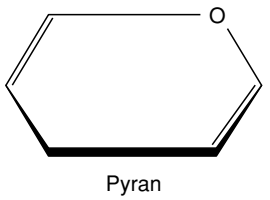

Pyran

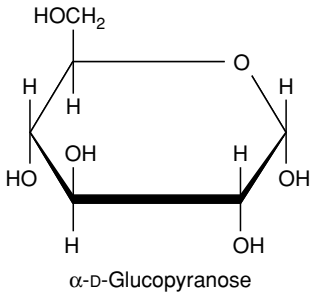

α-D-Glucopyranose

Figure 5 Pyranose form of glucose.

and an alcohol group in glucose. While dissolving in water, the optical rotatory power of the solution of monosaccharides gradually changes until it reaches an equilibrium. The freshly prepared solution of glucose has a specific rotation of $+112°$, and when this solution is allowed to stand, the rotation falls to $+52.7°$ and remains constant. This change in specific rotation is called mutarotation (Figure 7). In solution, the ring structure (α-D-glucopyranose) undergoes isomerism at position 1 or the anomeric carbon atom to form β-D-glucopyranose.

Epimers Epimers are isomers of glucose with different configurations of —OH and —H on the 2, 3 and 4 carbon atoms. The biologically important epimers of glucose are mannose and galactose formed by epimerization at the second and fourth carbon atoms (Figure 8).

Aldoses and ketoses Glucose and fructose have the same molecular formula but differ in their structural formula. In glucose, the first carbon atom has an aldehyde group, while in fructose, there is a keto group in the second carbon.

Configuration of C1 in glucose

The configuration of C1 in glucose may be α (with the —H atom to the left and the —OH group to the right) or β (with the —H atom to the right and the —OH group to the left). The empirical rule states that the α anomer has a higher dextrarotation. Hence, the α-D(+) glucose has a specific rotation of $+111°$, and β-D(+) glucose has a specific rotation of $+18.7°$. In negative rotation, the β anomer has a more negative rotation than the α-anomer.

Properties

Oxidation

Sugars contain numerous oxidizable groups. Mild oxidation converts only the aldehyde into a carboxylic acid (aldonic acid), but more vigorous oxidation results in the oxidation of both aldehyde and primary alcohols to carboxylic acids (alaric acid). Even more vigorous oxidation ruptures the six-carbon framework due to the oxidation of the secondary hydroxyl groups. Mild oxidation of glucose using the enzyme glucose oxidase produces gluconic acid, vigorous oxidation of glucose with strong oxidizing agents produces glucaric acid, and even more vigorous oxidation results in the formation of glucosaccharic acid. Periodic acid treatment results in the complete breakdown of the carbon chain.

Reduction

Similar to the process of oxidation, the product of reduction depends on the nature of the reagent. Sodium amalgam, sodium borohydride, high-pressure catalytic hydrogenation, or electrolyte reduction in an acid solution converts aldoses into the corresponding alcohols, and similarly, glucose is converted to sorbitol. Reduction of glucose with concentrated hydrochloric acid and red phosphorus at 100 °C produces 2-iodohexane, and prolonged heating finally gives n-hexane.

Fermentation

Glucose is fermented by yeast to produce ethanol and forms glucosates with various metallic hydroxides. Treatment with calcium hydroxide produces calcium glucosate ($C_6H_{12}O_6CaO$).

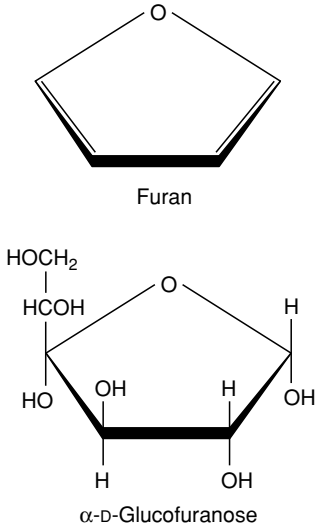

Figure 6 Furanose form of glucose.

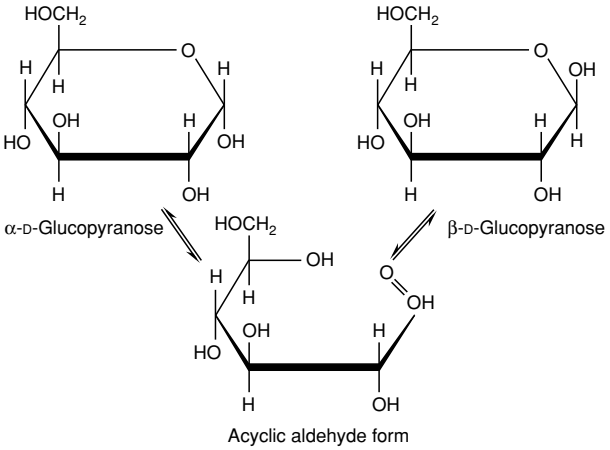

Figure 7 Mutarotation of glucose.

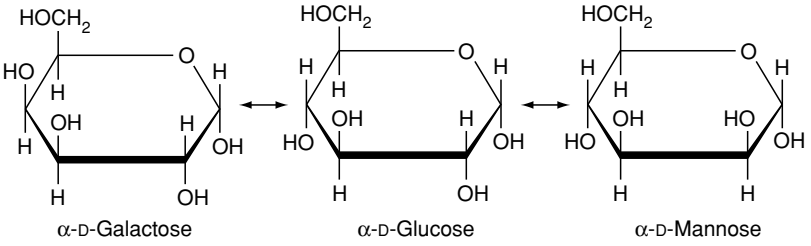

Figure 8 Epimerization of glucose.

Glycosides

Generally, hemiacetals react with another molecule of an alcohol to form acetals. Similarly, sugars (in their ring form) react with a molecule of an alcohol to form an acetal derivative generally called a glycoside. The glycosides of glucose are called glucosides.

Analysis

Monosaccharides contain free aldehyde (—CHO) or ketone (═CO) groups, and so their chemical properties vary, depending on the number of hydroxyl groups and the presence or absence of —CHO/=CO groups. These variations are the basis of color reactions for identifying various monosaccharides. The common methods for conformational (structural), qualitative, and quantitative analyses of glucose are as follows.

Conformational Analysis

Various methods are used to study the conformation of monosaccharides. The common method is to estimate the instability rating of various conformations by using various instability factors such as chair or boat forms, number of axial hydroxyl groups (by considering each axial —OH group result in one instability unit), 1,3 interactions involving the axial —OH group (0.5 instability unit) an axial CH_2OH group (result in two instability units) and if the —OH group on C2 is axial, and the oxygen atom on C1 is equatorial (resulting in 2.5 instability units).

In addition to this instability rating for conformational analysis, various physical methods are used for conformational analysis and to identify and elucidate the structures of various monosaccharides.

X-ray analysis This method is limited to studies on crystals (solid components), and the results may not apply to conformations when the compound in solution. The X-ray analysis of D-glucose has shown that it is in pyranose form.

Infrared spectroscopy This method is used to identify, and determine the structure, configuration, and conformation of, monosaccharides.

Nuclear magnetic resonance spectroscopy This technique is used for the purposes of identification and assignment of configuration and conformation. Deuterium oxide is used as the solvent to examine all C—H protons, and dimethylsulfoxide is used as the solvent to examine —OH protons.

Mass spectrometry This is the most advanced technique used to determine the size of a ring of a monosaccharide (pyranose or furanose), the number and position of methyl, ether, and acetyl groups in methylated sugar, the position of linkage in a disaccharide, etc.

Chromatography This is used to separate, estimate, and identify monosaccharides, using various techniques such as paper, column, gas–liquid (GLC), and thin-layer chromatography (TLC). Column chromatography is used almost exclusively for preparative purposes, and GLC and TLC are particularly useful in analytical work. Definite identification is carried out by isolating the sugar then determining the physical characteristic and precipitating the crystalline derivative. The presence of a reducing sugar on a paper chromatogram is detected using various reagents, e.g., spraying with ammoniacal silver nitrate (black spot formation) or a solution of salt of an aromatic amine, e.g., *N,N*- dimethylaniline hydrochloride, to examine the appearance of colors characteristic of the type of reducing sugar.

Qualitative Analysis

The occurrence of glucose in a carbohydrate solution can be detected by various methods. The commonest tests are as follows.

Molisch's test The principle of this test is that monosaccharides, derived from carbohydrates by hydrolyzing the glycosidic bonds (using concentrated sulfuric acid), become dehydrated to furfural and its derivatives. These products then combine with sulfonated α-naphthol to give a purple complex.

Two drops of α-naphthol solution are added to 2 ml of the test solution in a test tube, and then

about 1 ml of concentrated sulfuric acid is added down the side of the tube so as to form two layers. A color change observed at the interface of the two liquids confirms the presence of carbohydrates in the test solution.

Anthrone reaction The principle of this test is that the aforementioned furfural and derivatives react with anthrone (10-keto, 9,10-dihydroanthracene) to give a blue–green complex.

Five drops of the test solution are added to about 2 ml of anthrone reagent (a mixture of 2 g of anthrone per liter of sulfuric acid) and the solution mixed thoroughly. The appearance of a blue–green color confirms the presence of carbohydrates in the test solution.

Fehling's test This test uses two reagents: Fehling's reagent A (34.65 g of copper sulfate dissolved in distilled water and made up to 500 ml) and Fehling's reagent B (125 g of potassium hydroxide and 173 g of potassium sodium tartrate dissolved in distilled water and made up to 500 ml). The blue alkaline cupric hydroxide present in the solution when heated in the presence of reducing sugar is then reduced to yellow or red cuprous oxide and precipitated.

To 1 ml of the Fehling's solution A and 1 ml solution B few drops of test solution are added and the mixture boiled for a few minutes. The appearance of a yellow or brownish precipitate confirms the presence of glucose in the solution.

Benedict's test Benedict's reagent is made by dissolving 173 g of sodium citrate and 100 g of sodium carbonate in 800 ml of warm water, and the solution filtered and made up to 850 ml with water. Then, 17.3 g of copper sulfate is dissolved in 100 ml of water, the solution made up to 150 ml, and the copper sulfate solution slowly added to the first solution while stirring. The alkaline cupric hydroxide present in the solution, when heated in the presence of reducing sugar, is reduced to yellow or red cuprous oxide and is precipitated.

Five drops of the test solution are added to 2 ml of Benedict's reagent the mixture placed in a boiling water bath for 5 min and the color (green–yellow–orange–red) development against a standard glucose solution. The color of the final solution depends on the concentration of glucose present.

Barfoed's test Barfoed's reagent is a weakly acidic solution and is reduced by monosaccharides. The precipitate of cuprous oxide is less dense than with the previous two tests, and it is best to leave the tube to stand to allow the precipitate to settle. The color of the cuprous oxide is also different, giving a brick-red color in the presence of carbohydrates.

One milliliter of the test solution is added to 2 ml of Barfoed's reagent (made by dissolving 13.3 g of copper acetate in about 200 ml of water and adding 1.8 ml of glacial acetic acid), and the mixture is then boiled for 1 min and allowed to stand and cool. The development of a brick-red color confirms the presence of carbohydrate in the test solution.

Preparation of osazones Compounds containing the —CO—CHOH— group form crystalline osazones with phenyl hydrazine with characteristic shapes and melting points. This can be used as a better method by which to identify a reducing sugar. Phenyl hydrazine reacts with the carbonyl group of the sugar, resulting in the formation of phenylhydrazone and the osazone (Figure 9).

Five milliliters of the sugar solution are acidified with 10 drops of glacial acetic acid, a pinch of phenylhydrazine powder, and 0.1 g of solid sodium acetate. The mixture is then warmed in a boiling water bath for 10 min with occasional shaking and filtered. The filtrate is warmed in the water bath for a further 20 min and cooled very slowly. The osozone crystals can then be examined under a microscope and the shapes compared. Glucose produces needle-shaped yellow osazone crystals.

Quantitative Analysis

Anthrone method This is a very rapid and convenient method for determining hexoses either in free form or in polysaccharides. The resulting blue–green coloration shows a maximum absorption at 620 nm.

Four milliliters of anthrone reagent are added to 1 ml of a protein-free (the presence of protein results in the development of false colors) carbohydrate solution, mixed rapidly, and the tubes with lids (to prevent loss of water by evaporation) placed in a boiling water bath for 10 min. After cooling, the

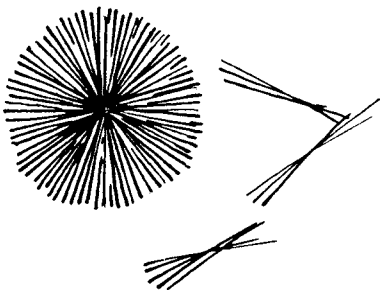

Figure 9 Crystalline appearance of glucosazone (needles or feathery).

optical density is measured at 620 nm against a reagent blank. A standard curve should be prepared using known glucose and glycogen solutions.

Nelson–Somogyi method In this method, the sugar is heated with an alkaline solution of copper tartrate, and cuprous oxide is produced, which reacts with arsenomolybdate to give molybdenum blue; the intense blue color is then measured in a colorimeter. Sodium sulfate is included in the reaction mixture to minimize the entry of atmospheric oxygen into the solution, which would reoxidate the cuprous oxide.

One milliliter of the test solution (containing approximately 50–150 μg of sugar) is mixed with 1.5 ml of water in a test tube and mixed thoroughly. Next, 0.2 ml of barium hydroxide solution ($0.3\,mol\,l^{-1}$) is added to the mixture followed by 0.2 ml of aqueous zinc sulfate solution ($50\,g\,l^{-1}$), then shaken thoroughly and centrifuged. The supernatant is used for analysis.

One milliliter of alkaline copper reagent (made by dissolving 4 g of $CuSO_4 \cdot 5H_2O$, 24 g of Na_2CO_3, 16 g of Na–K tartrate, and 180 g of anhydrous Na_2SO_4 in water and diluting to 1 l) is added and the solution mixed well. The solution is then heated on a boiling water bath for 15 min (in a test tube with a lid), the tubes cooled in cold water, 1 ml of the arsenomolybdate reagent (commercially available) added, and the solution left to stand for a minute until the effervescence ceases. The blue color is diluted with water to a final volume of 10 ml, and the extinction is read at 510 or 660 nm. A standard should be prepared, and the reading should be taken against a reagent blank.

Dinitrosalicylic acid method In this method, the sugar is heated with dinitrosalicylic acid reagent followed by a 40% solution of potassium sodium tartrate, and the dark red color developed is measured in a colorimeter at 510 nm.

The sample is mixed with dinitrosalicylic acid reagent (prepared by dissolving 1 g of dinitrosalicylic acid, 200 mg of crystalline phenol, and 50 mg of sodium sulfate in 100 ml of 1% NaOH and storing the solution at 4 °C) and heated for 5 min in a boiling water bath. Then, 1 ml of 40% potassium sodium tartrate is added and the solution allowed to cool. The intensity of the dark red color developed is read at 510 nm, and the concentration of glucose in the sample is calculated by plotting the optical density against a standard curve prepared using various concentrations of known glucose solutions.

Glucose oxidase method As glucose is a simple sugar, one of the most specific methods for the quantitative estimation of glucose is the use of glucose oxidase (EC 1.1.3.4), which catalyzes the oxidation of α-D-glucose to D-glucono-1,5 lactone (gluconic acid) with the formation of hydrogen peroxide. The oxygen liberated from hydrogen peroxide by peroxidase reacts with the *O*-dianisidine and oxidases it to a red chromophore product.

One milliliter of glucose oxidase peroxidase reagent (made by dissolving 25 mg of *O*-dianisidine in 1 ml of methanol, adding 49 ml of 0.1 M phosphate buffer (pH 6.5), then adding 5 mg of peroxidase and 5 mg of glucose oxidase to the mixture) is added to the sample. The tubes are then incubated at 35 °C for 40 min and the reaction terminated by the addition of 2 ml of 6 N HCl. The intensity of the color developed at 540 nm is read against a reagent blank, and the concentration of the sample is then calculated by plotting a standard curve using different concentrations of known glucose solutions.

Phenol sulfuric acid method In a hot acidic medium, glucose is dehydrated to hydroxymethyl furfural. This forms a green product with phenol and has an absorbed maximum at 490 nm.

To 0.2 ml of the sample solution, 1 ml of 5% phenol solution and 5 ml of 95% sulfuric acid are added and the solution mixed well. After 10 min, the test tubes are shaken and placed in a water bath at 25–30 °C for 20 min. The green color is then read at 490 nm against a reagent blank, and the amount of glucose in the sample is calculated by plotting a standard curve.

See also: **Chromatography**: Principles; Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; Combined Chromatography and Mass Spectrometry; **Glucose**: Properties and Analysis; **Spectroscopy**: Infrared and Raman; Atomic Emission and Absorption; Nuclear Magnetic Resonance

Further Reading

Guthrie RD (1974) *Introduction to Carbohydrate Chemistry*, 4th edn. Oxford: Clarendon Press.

Pigman W and Horten D (1972) *The Carbohydrates*, 2nd edn, vol. 1A. New York: Academic Press.

Sadasivam S and Manikam A (1992) *Biochemical Methods for Agricultural Sciences*, pp. 1–21. New Delhi, India: Wiley Eastern.

Whelam WJ (1975) *Biochemistry of Carbohydrates*. London: Butterworth and University Park, University Park Press.

Whistler RL (1962–1971) *Methods in Carbohydrate Chemistry*, vols. 1–6. New York: Academic Press.

Function and Metabolism

D A Bender, University College London, London, UK

Background

Glucose, arising from dietary carbohydrates, is the major metabolic fuel, providing 50–75% of energy in most diets. In the fed state, glucose in excess of immediate requirements is used for the synthesis of the reserve carbohydrate glycogen in liver and muscle. Between meals these glycogen reserves are used to maintain a supply of glucose for the red blood cells and central nervous system – other tissues use mainly fatty acids in the fasting state, to spare glucose. In the fed state, when glucose is plentifully available, it becomes the major fuel of all tissues, and the respiratory quotient (RQ) rises from about 0.8 to near 1, indicating that glucose is the substrate for most energy-yielding metabolism.

Most glucose is metabolized by glycolysis, leading to the formation of 2 mol of pyruvate per mol of glucose, which then undergoes oxidative decarboxylation to acetyl coenzyme A (CoA), and oxidation in the citric acid cycle. In red blood cells and tissues that are active in fatty acid synthesis (liver, adipose tissue, and lactating mammary gland) the pentose phosphate pathway provides an alternative to part of the pathway of glycolysis.

Glycolysis: The (Anaerobic) Metabolism of Glucose

Overall, the pathway of glycolysis is cleavage of the 6-carbon glucose molecule into two 3-carbon units. The key features of the pathway are:

- Two phosphorylation reactions to form successively glucose-6-phosphate and fructose bisphosphate.
- Cleavage of fructose bisphosphate to yield two molecules of triose (3-carbon sugar) phosphate.
- Two steps in which phosphate is transferred from a 3-carbon substrate on to adenosine diphosphate (ADP), forming adenosine triphosphate (ATP).
- One step in which NAD^+ is reduced to NADH.

The substrate for glycolysis is glucose-6-phosphate; this may arise from two sources:

- Phosphorylation of glucose, catalyzed by hexokinase;
- Phosphorolysis of glycogen in liver and muscle to yield glucose-1-phosphate, catalyzed by glycogen phosphorylase. Glucose-1-phosphate is readily isomerized to glucose-6-phosphate.

The pathway of glycolysis is shown in Figure 1. Although the aim of glucose metabolism is ultimately to phosphorylate ADP → ATP, there is a moderate initial ATP cost; there are two steps in which ATP is used, one to form glucose-6-phosphate when glucose is the substrate, and the other to form fructose bisphosphate.

The formation of fructose bisphosphate, catalyzed by phosphofructokinase, is an important step for the regulation of glucose metabolism. Phosphofructokinase is inhibited by normal intracellular concentrations of ATP. As the ratio of ATP to ADP falls, 5′AMP is formed; it relieves the inhibition of phosphofructokinase by ATP, leading to an increased rate of glycolysis and increased formation of ATP.

Fructose bisphosphate is cleaved to dihydroxyacetone phosphate and glyceraldehyde-3-phosphate, which are interconvertible. The onward metabolism of these 3-carbon sugars is linked to both the reduction of NAD^+ to NADH, and direct (substrate-level) phosphorylation of ADP → ATP.

There is a net yield of 8 × ADP + phosphate → ATP from the oxidation of 1 mol of glucose to 2 mol of pyruvate. Initially, 2 mol of ATP are required to form fructose bisphosphate, but the pathway yields 4 × ATP by direct phosphorylation of ADP, and 2 × NADH (formed from NAD^+), which is equivalent to a further 6 × ATP when oxidized in the electron transport chain.

The glycolytic pathway also provides a route for the metabolism of fructose, galactose (which undergoes phosphorylation to galactose-1-phosphate and isomerization to glucose-1-phosphate), and glycerol. Some fructose is phosphorylated directly to fructose-6-phosphate by hexokinase, but most is phosphorylated to fructose-1-phosphate by a specific fructokinase. Fructose-1-phosphate is then cleaved to yield dihydroxyacetone phosphate and glyceraldehyde; the glyceraldehyde can be phosphorylated to glyceraldehyde-3-phosphate by triose kinase.

Glycerol, arising from the hydrolysis of triacylglycerols, can be phosphorylated and oxidized to dihydroxyacetone phosphate. In triacylglycerol synthesis most glycerol phosphate is formed from dihydroxyacetone phosphate.

The Reduction of Pyruvate to Lactate: Anaerobic Glycolysis

In red blood cells, which lack mitochondria, reoxidation of NADH formed in glycolysis cannot be by way of the electron transport chain, as occurs in other tissues. Similarly, under conditions of maximum exertion, for example, in sprinting, the rate at which oxygen can be taken up into the muscle is inadequate

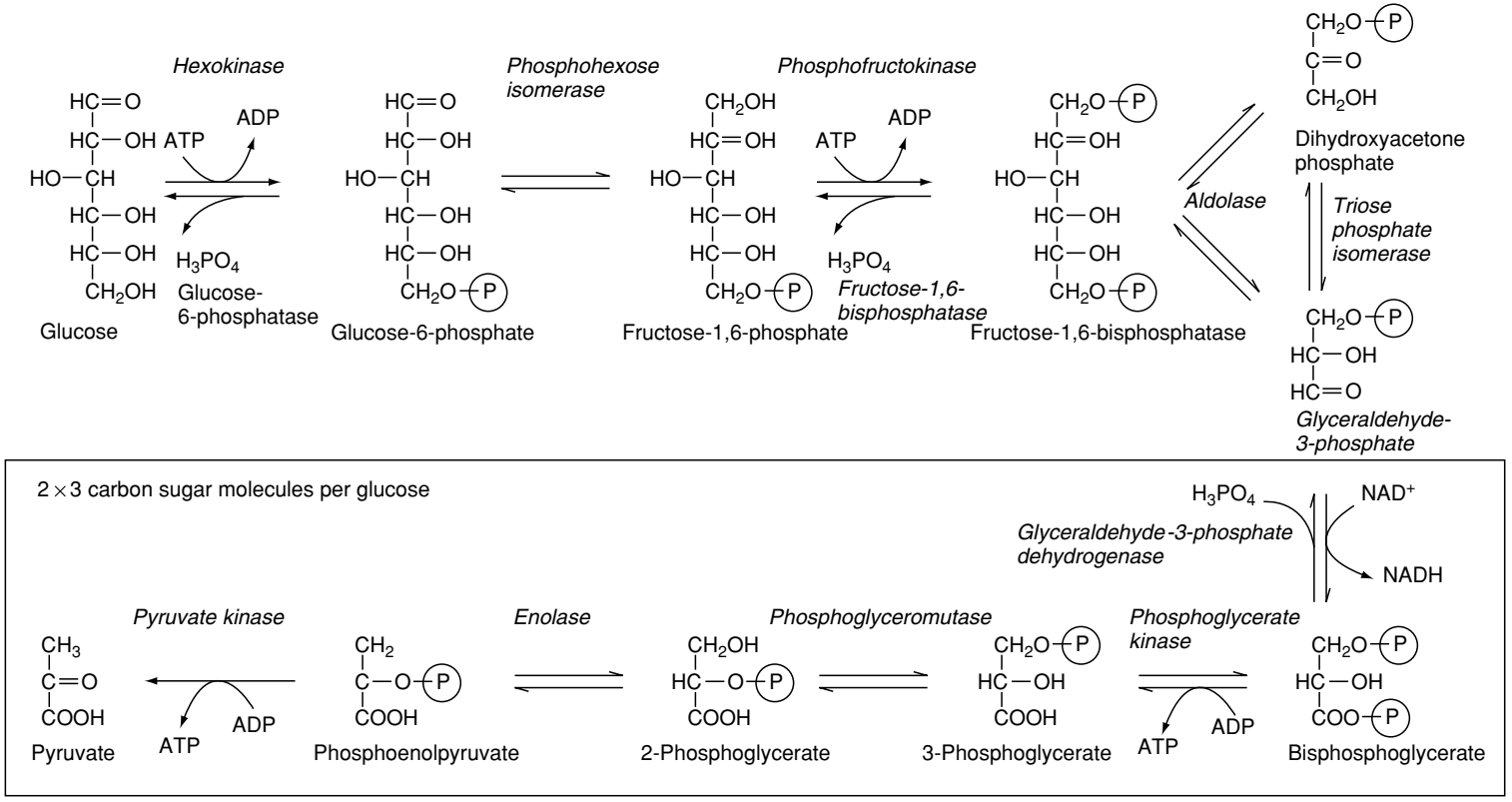

Figure 1 The pathway of glycolysis. Hexokinase and glucokinase EC 2.7.1.1, glucose-6-phosphatase EC 3.1.3.9, phosphoglucomutase EC 5.4.2.2, phosphofructokinase EC 2.7.1.11, fructose-1,6-bisphosphatase EC 3.1.3.11, aldolase EC 4.1.2.13, triose phosphate isomerase EC 5.3.1.1, glyceraldehyde 3-phosphate dehydrogenase EC 1.2.1.12, phosphoglycerate kinase EC 2.7.2.3, phosphoglyceromutase EC 5.4.2.1, phosphopyruvate hydratase EC 4.2.1.11, pyruvate kinase EC 2.7.1.40, ATP, adenosine triphosphate; ADP, adenosine diphosphate.

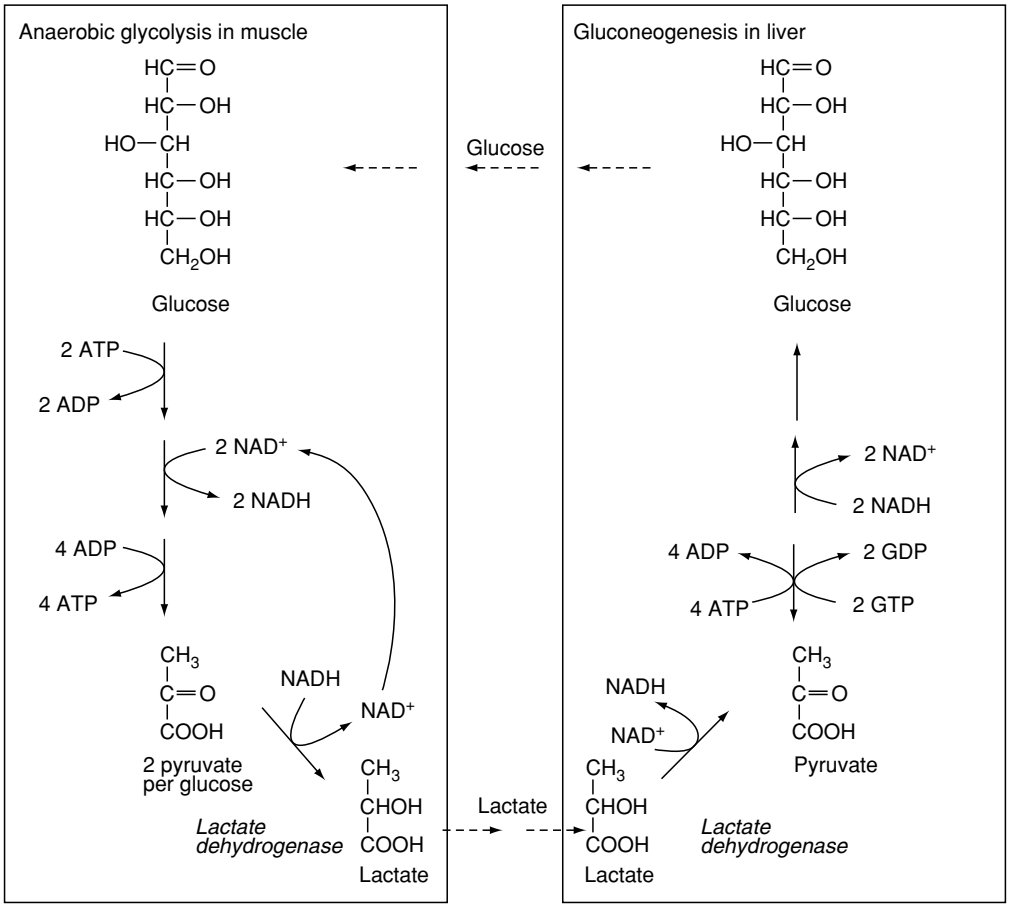

Figure 2 The Cori cycle – anaerobic glycolysis in muscle and gluconeogenesis in the liver. Lactate dehydrogenase EC 1.1.1.28. ATP, adenosine triphosphate; ADP, adenosine diphosphate.

to permit reoxidation of all the NADH which is formed in glycolysis. In order to maintain the oxidation of glucose, and the net yield of 2 × ATP per mol of glucose oxidized (or 3 mol of ATP if the source is muscle glycogen), NADH is oxidized to NAD^+ by the reduction of pyruvate to lactate, catalyzed by lactate dehydrogenase (Figure 2).

Lactate is exported from muscle and red blood cells, and taken up by the liver, where it is used for the resynthesis of glucose – the Cori cycle, shown in Figure 2. Synthesis of glucose from lactate is an ATP (and guanosine triphosphate (GTP))-requiring process. The oxygen debt after strenuous physical activity is due to an increased rate of energy-yielding metabolism to provide the ATP and GTP that are required for gluconeogenesis from lactate. While most of the lactate will be used for gluconeogenesis, a proportion will undergo oxidation to CO_2 in order to provide the ATP and GTP required for gluconeogenesis.

The conversion of glucose to lactate is known as anaerobic glycolysis, since it does not require oxygen. However, it is not true to say that human metabolism (apart from red blood cells) is ever wholly anaerobic. The formation of lactate is the fate of *much* of the pyruvate formed from glucose under conditions of maximum muscle exertion when oxygen is limiting, but as much as possible will continue to undergo complete oxidation.

Many tumors have a low capacity for oxidative metabolism, so that much of the energy-yielding metabolism in the tumor is anaerobic. Lactate produced by anaerobic glycolysis in tumors is exported to the liver for gluconeogenesis; this increased cycling of glucose between anaerobic glycolysis in the tumor and gluconeogenesis in the liver may account for much of the hypermetabolism and consequent weight loss seen in patients with cancer cachexia.

Truly anaerobic glycolysis does occur in microorganisms which are capable of living in the absence of oxygen. Here there are two possible fates for the pyruvate formed from glucose, both of which involve the oxidation of NADH to NAD^+:

- Reduction to lactate, as occurs in human muscle. This is the pathway in lactic acid bacteria, which are responsible for the fermentation of lactose in milk to form yogurt and cheese;
- Decarboxylation and reduction to ethanol. This is the pathway of fermentation in yeast, which is exploited to produce alcoholic beverages.

The Pentose Phosphate Pathway – An Alternative to Glycolysis

There is an alternative pathway for the conversion of glucose-6-phosphate to fructose-6-phosphate, the pentose phosphate pathway (sometimes known as the hexose monophosphate shunt), shown in Figure 3.

Overall, the pentose phosphate pathway produces 2 mol of fructose-6-phosphate, 1 mol of glyceraldehyde-3-phosphate, and 3 mol of carbon dioxide from 3 mol of glucose-6-phosphate, linked to the reduction of 6 mol of $NADP^+$ to NADPH. The sequence of reactions is as follows:

- Three mol of glucose are oxidized to yield 3 mol of the 5-carbon sugar ribulose-5-phosphate + 3 mol of carbon dioxide.
- Two mol of ribulose-5-phosphate are isomerized to yield 2 mol of xylulose-5-phosphate.
- One mol of ribulose-5-phosphate is isomerized to ribose-5-phosphate.
- One mol of xylulose-5-phosphate reacts with the ribose-5-phosphate, yielding (ultimately) fructose-6-phosphate and erythrose-4-phosphate.
- The other mol of xylulose-5-phosphate reacts with the erythrose-4-phosphate, yielding fructose-6-phosphate and glyceraldehyde-3-phosphate.

This is the pathway for the synthesis of ribose for nucleotide synthesis; more importantly, it is the source of half the NADPH required for fatty acid synthesis; tissues that are active in lipogenesis have a high activity of the pentose phosphate pathway.

The pentose phosphate pathway is also important in red blood cells, where the NADPH is required to maintain an adequate pool of reduced glutathione (GSH), the reducing agent for glutathione peroxidase, which reduces $H_2O_2 \rightarrow H_2O$ and O_2. Oxidized glutathione (GSSG) is reduced back to active GSH by glutathione reductase, which uses NADPH as the reducing agent.

Partial or total lack of glucose-6-phosphate dehydrogenase (and hence impaired activity of the pentose phosphate pathway) is the cause of favism, an acute hemolytic anemia with fever and hemoglobinuria, precipitated in genetically susceptible people by the consumption of broad beans (fava beans) and a variety of drugs, all of which, like the toxins in fava beans, undergo redox cycling, producing hydrogen peroxide. Infection can also precipitate an attack, because of the increased production of oxygen radicals as part of the macrophage respiratory burst.

Because of the low activity of the pentose phosphate pathway, there is a lack of NADPH in red blood cells, and hence an impaired ability to remove hydrogen peroxide, which causes oxidative damage to the cell membrane lipids, leading to hemolysis.

The Oxidation of Pyruvate to Acetyl CoA

The first step in the complete oxidation of pyruvate is catalyzed by the pyruvate dehydrogenase multienzyme complex; an oxidative decarboxylation that results in the formation of acetyl CoA. The oxidation involves the reduction of NAD^+ to NADH. Since 2 mol of pyruvate are formed from each mol of glucose, this step represents the formation of 2 mol of NADH, equivalent to $6 \times$ ATP for each mol of glucose metabolized. The acetate is released from the enzyme esterified to coenzyme A as acetyl CoA (Figure 4), which undergoes oxidation in the citric acid cycle.

The decarboxylation and oxidation of pyruvate to form acetyl CoA requires the coenzyme thiamin diphosphate, which is formed from vitamin B_1. In thiamin deficiency this reaction is impaired, and deficient subjects are unable to metabolize glucose normally. Especially after a test dose of glucose or moderate exercise they develop high blood concentration of pyruvate and lactate. In some cases this may be severe enough to result in life-threatening acidosis.

Glycogen Synthesis and Utilization

In the fed state, glycogen is synthesized from glucose in both liver and muscle. The reaction is a stepwise addition of glucose units on to the glycogen that is already present. In muscle insulin stimulates both the uptake of glucose from the blood stream by active transport and also the activity of glycogen synthetase. In the liver, glycogen synthetase is also stimulated in response to insulin, but glucose uptake is insulin-independent and occurs by a carrier-mediated passive process followed by metabolic trapping as glucose-6-phosphate. There are two isoenzymes of hexokinase in the liver:

- An isoenzyme with a low Michaelis constant (K_m) which is saturated, and therefore acting at its maximum rate, at concentrations of glucose very much lower than occur in tissues. This isoenzyme therefore acts at a more or less constant rate regardless of the concentration of glucose available, and

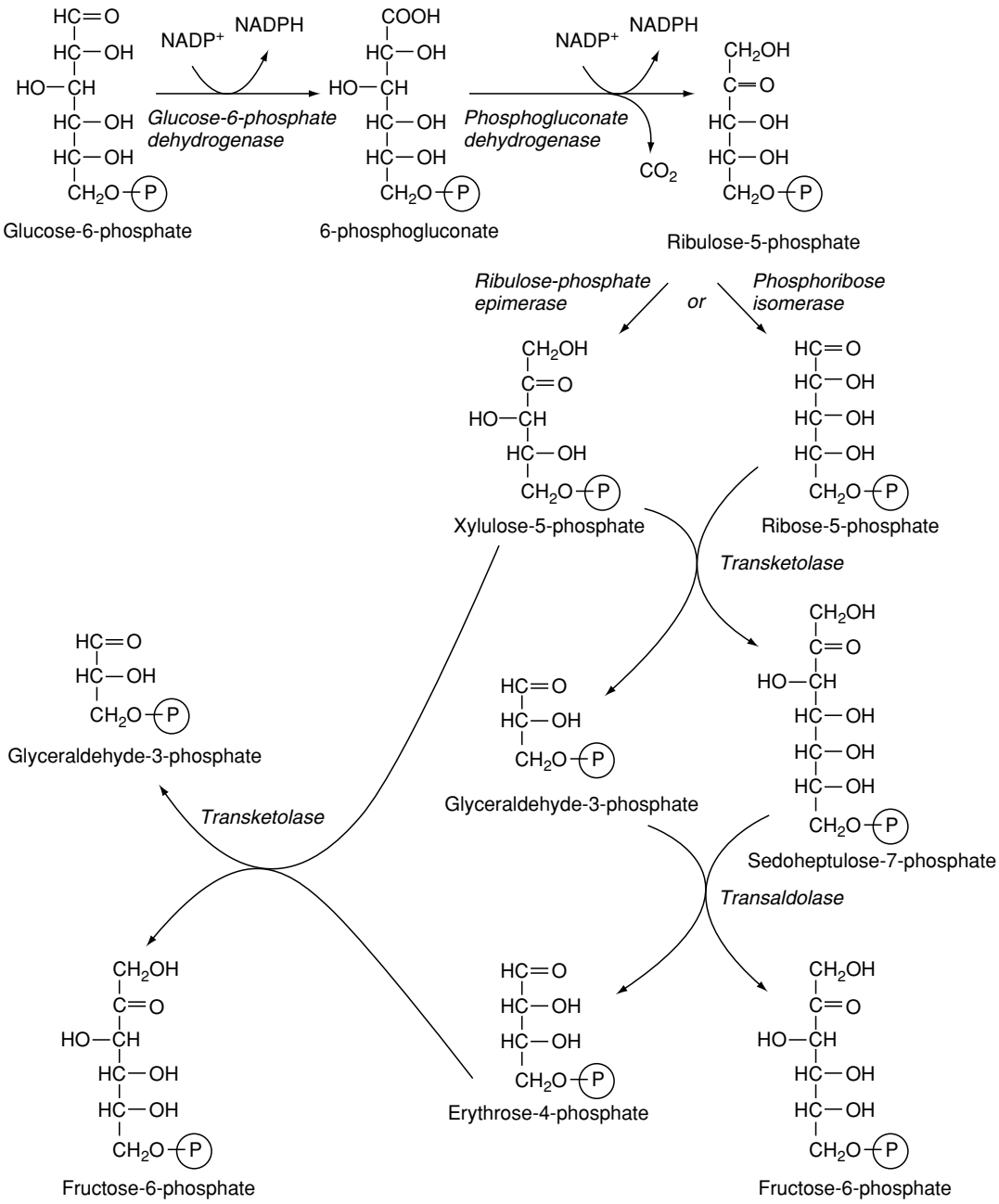

Figure 3 The pentose phosphate pathway. Glucose-6-phosphate dehydrogenase EC 1.1.1.49, phosphogluconate dehydrogenase EC 1.1.1.44, ribulose phosphate epimerase EC 5.1.3.1, phosphoribose isomerase EC 5.3.1.6, transketolase EC 2.2.1.1, transaldolase EC 2.2.1.2.

provides glucose-6-phosphate to meet the liver's requirements for glycolysis.

- An isoenzyme with a high K_m (also known as glucokinase) which only has significant activity as the concentration of glucose coming from the small intestine in the hepatic portal vein rises to above 10–15 mmol l^{-1}. In the fed state this isoenzyme produces glucose-6-phosphate considerably in excess of the liver's requirement for glycolysis; it is used for glycogen synthesis.

As shown in Figure 5, glycogen synthesis involves the intermediate formation of uridine diphosphate (UDP) glucose by reaction between glucose-1-phosphate and uridine triphosphate (UTP). As each glucose unit is added to the growing glycogen chain, so UDP is released, and must be rephosphorylated to UTP by reaction with ATP. There is thus a significant cost of ATP for the synthesis of glycogen: 2 mol of ATP are converted to ADP + phosphate for each glucose unit added, and overall the energy cost

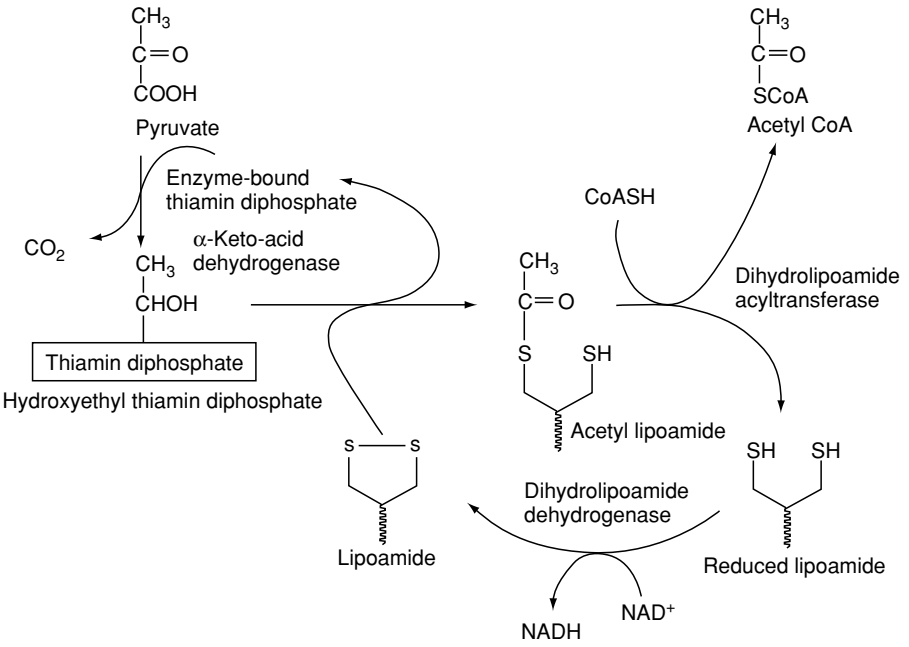

Figure 4 The reaction sequence of the pyruvate dehydrogenase multienzyme complex (EC 1.2.4.1).

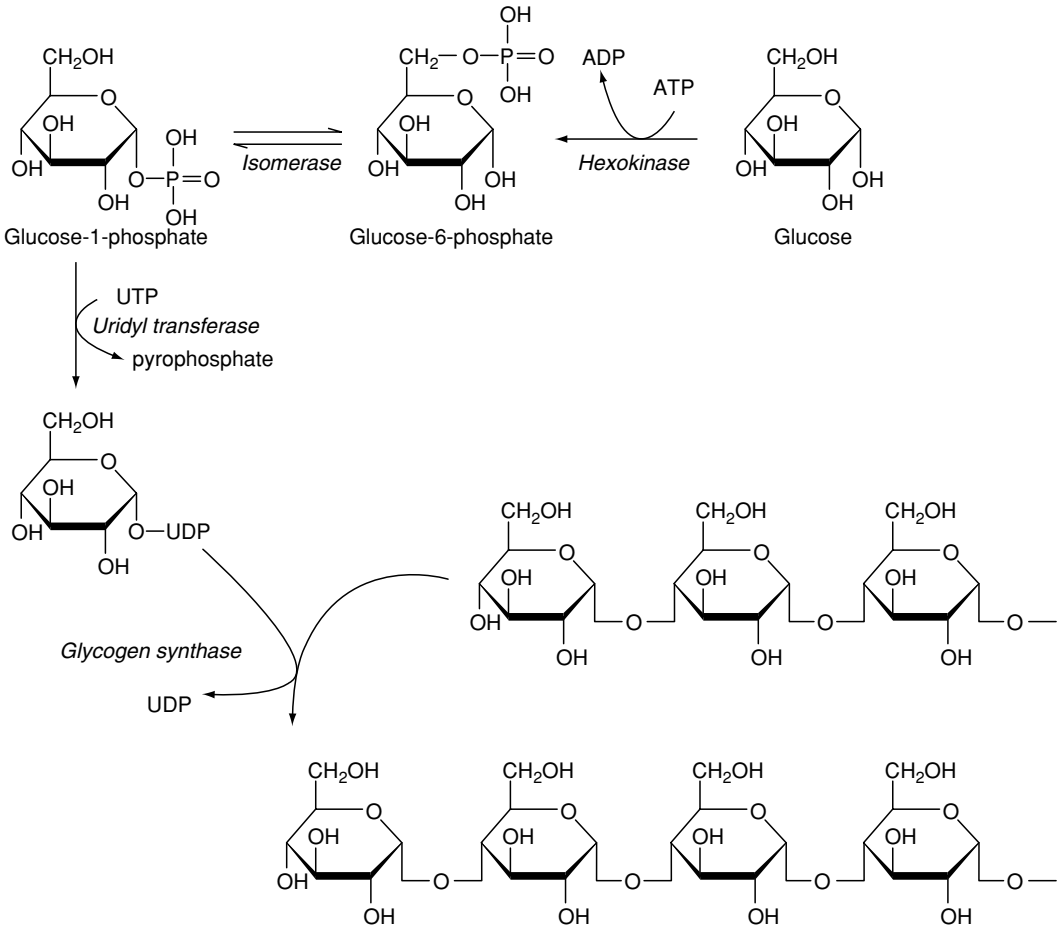

Figure 5 The synthesis of glycogen. Hexokinase EC 2.7.1.1, uridyl transferase EC 2.7.7.9. glycogen synthase EC 2.4.1.11. ADP, adenosine diphosphate; ATP, adenosine triphosphate.

of glycogen synthesis may account for 5% of the energy yield of the carbohydrate stored.

Glycogen synthetase forms only the $\alpha 1 \rightarrow 4$ links that form the straight chains of glycogen. The branch points are introduced by the transfer of 6–10 glucose units in a chain from carbon-4 to carbon-6 of the glucose unit at the branch point.

In the fasting state, glycogen is broken down by the removal of glucose units one at a time from the many ends of the molecule, in response to stimulation by the hormone glucagon. The reaction is a phosphorolysis – cleavage of the glycoside link between two glucose molecules by the introduction of phosphate. The product is glucose-1-phosphate, which is then isomerized to glucose-6-phosphate. In the liver, glucose-6-phosphatase catalyzes the hydrolysis of glucose-6-phosphate to free glucose, which is exported for use by the brain and red blood cells.

Muscle cannot release free glucose from the breakdown of glycogen, since it lacks glucose-6-phosphatase. However, muscle glycogen can be an indirect source of blood glucose in the fasting state. Glucose-6-phosphate from muscle glycogen undergoes glycolysis to pyruvate (Figure 1), which is then transaminated to alanine. Alanine is exported from muscle, and taken up by the liver for use as a substrate for gluconeogenesis.

Glycogen phosphorylase stops cleaving $\alpha 1 \rightarrow 4$ links four glucose residues from a branch point, and a debranching enzyme catalyzes the transfer of a three-glucosyl unit from one chain to the free end of another chain. The $\alpha 1 \rightarrow 6$ link is hydrolyzed by a glucosidase, releasing glucose.

The branched structure of glycogen means that there are many points at which glycogen phosphorylase can act; in response to stimulation by epinephrine (adrenaline) there can be a rapid release of glucose-1-phosphate from glycogen.

Endurance athletes require a slow release of glucose-1-phosphate from glycogen over a period of hours, rather than a rapid release. There is some evidence that this is achieved better from glycogen that is less branched, and therefore has fewer points at which glycogen phosphorylase can act. The formation of branch points in glycogen synthesis is slower than the formation of $\alpha 1 \rightarrow 4$ links, and this has been exploited in the process of 'carbohydrate loading' in preparation for endurance athletic events. The athlete exercises to exhaustion, when muscle glycogen is more or less completely depleted, then consumes a high-carbohydrate meal, which stimulates rapid synthesis of glycogen, with fewer branch points than normal. There is little evidence to show whether this improves endurance performance or not; the improvement that has been reported may be the result of knowing that one has made an effort to improve performance rather than any real metabolic effect.

Gluconeogenesis – The Synthesis of Glucose from Noncarbohydrate Precursors

Because the brain is largely dependent on glucose as its metabolic fuel (and red blood cells are entirely so) there is a need to maintain the blood concentration of glucose between about 3 and 5 mmol l^{-1} in the fasting state. If the plasma concentration of glucose falls below about 2 mmol l^{-1} there is a loss of consciousness – hypoglycemic coma.

In short-term fasting the plasma concentration of glucose is maintained by the use of glycogen, and by releasing free fatty acids from adipose tissues, which are preferentially used by muscle, sparing such glucose as is available for use by the brain and red blood cells.

However, the total body content of glycogen would be exhausted within 12–18 h of fasting if there were no other source of glucose. Gluconeogenesis is the synthesis of glucose from noncarbohydrate precursors: amino acids from the breakdown of protein, and the glycerol of triacylglycerols. It is important to note that, although acetyl CoA, and hence fatty acids, can be synthesized from pyruvate (and therefore from carbohydrates), the decarboxylation of pyruvate to acetyl CoA (Figure 2) cannot be reversed, and pyruvate cannot be formed from acetyl CoA. Since two molecules of carbon dioxide are formed for each 2-carbon acetate unit metabolized in the citric acid cycle, there can be no net formation of oxaloacetate from acetate. It is not possible to synthesize glucose from acetyl CoA, and fatty acids and ketone bodies cannot serve as substrates for gluconeogenesis under any circumstances.

The pathway of gluconeogenesis is essentially the reverse of the pathway of glycolysis, shown in Figure 1. However, at three steps there are separate enzymes involved in the breakdown of glucose (glycolysis) and gluconeogenesis. The reactions of pyruvate kinase, phosphofructokinase, and hexokinase cannot readily be reversed (i.e., they have equilibria which are strongly in the direction of glycolysis and the formation of pyruvate, fructose bisphosphate, and glucose-6-phosphate, respectively).

There are separate enzymes, under distinct metabolic control, for the reverse of each of these reactions in gluconeogenesis:

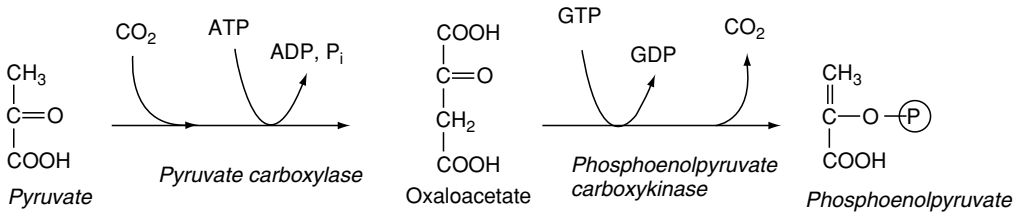

Figure 6 Reactions in the reversal of the pyruvate kinase reaction in gluconeogenesis. Pyruvate carboxylase EC 6.4.1.1, phosphoenolpyruvate carboxykinase EC 4.1.1.32. ATP, adenosine triphosphate; ADP, adenosine diphosphate; P_i, inorganic phosphate; GTP, guanosine triphosphate; GDP, guanosine diphosphate.

- Pyruvate is converted to phosphoenolpyruvate for glucose synthesis by a two-step reaction, with the intermediate formation of oxaloacetate, as shown in Figure 6. Pyruvate is carboxylated to oxaloacetate in an ATP-dependent reaction in which the vitamin biotin is the coenzyme. Oxaloacetate then undergoes a phosphorylation reaction, in which it also loses carbon dioxide, to form phosphoenolpyruvate.
- Fructose bisphosphate is hydrolyzed to fructose-6-phosphate by a hydrolysis reaction catalyzed by the enzyme fructose bisphosphatase;
- Glucose-6-phosphate is hydrolyzed to free glucose and phosphate by the action of glucose-6-phosphatase.

The other reactions of glycolysis are readily reversible, and the overall direction of metabolism, either glycolysis or gluconeogenesis, depends mainly on the relative activities of phosphofructokinase and fructose bisphosphatase.

Many of the products of amino acid metabolism can also be used for gluconeogenesis, since they are sources of either pyruvate or one of the intermediates in the citric acid cycle, and hence give rise to oxaloacetate. The requirement for gluconeogenesis from amino acids in order to maintain a supply of glucose explains why there is often a considerable loss of muscle in prolonged fasting or starvation, even if there are apparently adequate reserves of adipose tissue to meet energy needs.

See also: **Energy Metabolism**; **Thiamin**: Physiology

Further Reading

Bender DA (2002) *Introduction to Nutrition and Metabolism*, 3rd edn. London: Taylor & Francis.

Bender DA and Bender AE (1997) *Nutrition: A Reference Handbook*. Oxford: Oxford University Press.

Murray RK, Granner DK, Mayes PA and Rodwell VW (2000) *Harper's Biochemistry*, 25th edn. New York: McGraw-Hill.

Maintenance of Blood Glucose Level

V Marks, Haslemere, UK

Glucose

Glucose is almost the only simple sugar found in most body fluids in anything more than trace amounts and for all practical purposes is confined to extracellular water. Lactose and fructose are the major sugars in milk and semen, respectively. This article reviews the major factors determining the concentrations of glucose in blood under everyday physiological and pathological conditions.

Body Glucose Pool

The body of an adult subject seldom contains less than 8 g, or more than 28 g, of glucose at any one time (corresponding to blood glucose concentrations of 3.5–10 mmol l^{-1}) despite enormous fluctuations in demand and supply. This quantity of glucose can be looked upon as constituting a hypothetical body pool (**Figure 1**) confined within a glucose space equal in volume to the combined water in blood and the interstitial fluid, i.e., some 35% of total body water.

Glucose enters the cells by facilitated transport utilizing one or more of the genetically determined glucose transporter proteins that have been identified, depending on the tissue. Upon entering a cell, it is immediately phosphorylated and consequently removed from the glucose pool.

Although its subsequent conversion into carbon dioxide and water or other metabolites (most notably glycerol, fatty acids and the glycomoieties of mucopolysaccharides and glycoproteins) is the only way that glucose can ordinarily leave the glucose pool, its loss in the urine may become a major factor in diabetes mellitus.

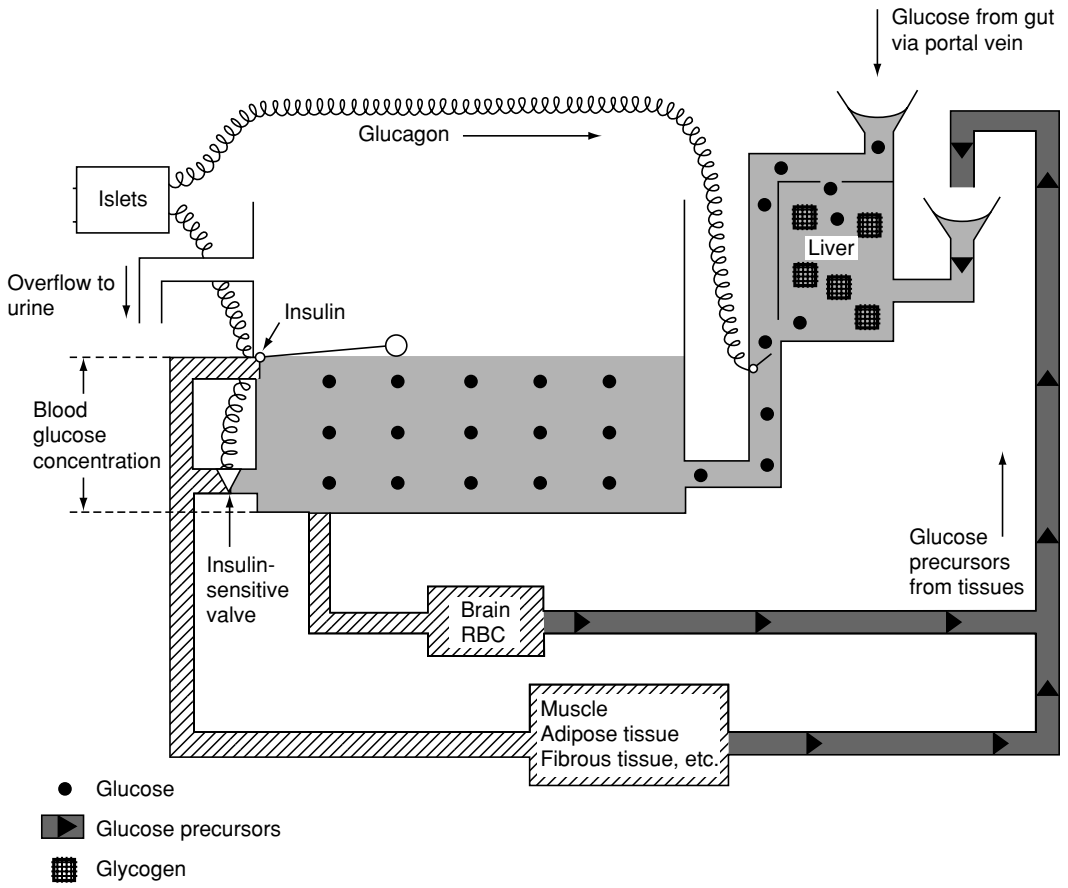

Figure 1 Schematic representation of blood glucose concentration and its relationship to the body glucose pool. The central system represents the hypothetical glucose pool, the actual size of which is represented by the horizontal axis, i.e., volume of distribution multiplied by blood (and extracellular fluid) glucose concentration. The postulated homeostatic switch is the cells of the endocrine pancreas that respond to blood glucose concentration modulated by hormonal and neural factors, themselves controlled by messages received from the gut (enteroinsular axis) and the autonomic nervous system. RBC, red blood cells.

Glucose enters the glucose pool from food via the hepatic vein or, in the postabsorptive subject, by release of glucose from preformed glycogen or molecules newly synthesized by liver cells – also into the hepatic vein.

Glucose Space

The glucose space is constant in any individual, and consequently, the amount of glucose in the pool is directly proportional to its concentration in the blood (see below). It is this concentration that is homeostatically controlled through a series of rather complicated control mechanisms, the most important of which involve individual pancreatic islets of Langerhans that function semiautonomously.

When pool size increases above a threshold, corresponding to a concentration in blood of about $10\,\text{mmol}\,\text{l}^{-1}$, glucose filtered at the glomeruli exceeds the tubular capacity to reabsorb it, and consequently, glucose spills over into the urine. Although temporary increases in glucose pool size (hyperglycemia) are not immediately harmful, decreases in glucose pool size (hypoglycemia) are. Indeed, they are potentially so dangerous that many defense mechanisms have evolved to prevent or overcome it (see below).

Blood Glucose

The brain is the only important drain on the glucose pool in the fasting subject when plasma insulin levels are minimal. It consumes glucose at the rate of about 78 mg per gram of tissue per day. This amounts, in an adult man, to about 110 g per day or 75 g per day in a 1-year-old child. A much smaller loss is into the erythron, which consists mainly of the bone marrow and red blood cells. Estimates of glucose turnover suggest that about 9 g of glucose enter and leave the glucose pool every hour in the average overnight fasting healthy subject.

The concentrations of glucose in venous and arterial blood are similar in the fasting subject because peripheral tissues such as muscle, skin, and connective tissue

do not extract significant amounts of glucose from the blood under these circumstances. In the recently fed subject, however, glucose uptake by peripheral tissues increases markedly under the influence of insulin released in response to the ingestion of a meal. This can produce a difference in arterial and venous blood glucose concentrations of 2 mmol l^{-1} or more.

This fact, known for more than 80 years, is still often forgotten or ignored by both experimentalists and clinicians. It not only has implications as regards our understanding of the physiology of glucose homeostasis, but sometimes has unfortunate consequences for patients who may, if only venous blood is sampled, be misdiagnosed as suffering from hypoglycemia (i.e., blood glucose < 3.0 mmol l^{-1}) when it is not really present. It is, after all, arterial, and not venous blood glucose, that is homeostatically controlled and relevant to brain physiology.

Venous blood is much easier to obtain than arterial blood and explains why, despite its theoretical and practical disadvantages, it is so often used in studies of glucose homeostasis and clinical practice. Fingerprick or earlobe-capillary blood accurately reflects arterial blood glucose levels under most circumstances but is more difficult to obtain in more than small amounts. So-called arterialized venous blood collected from heat-distended veins on the back of the hands can often be used in studies of glucose homeostasis when collection of arterial blood would not be ethical.

Blood glucose concentrations generally lie within the range 3.5–6.0 mmol l^{-1} in healthy fasting adult subjects and seldom rise above 11 mmol l^{-1} in arterial, or 10 mmol l^{-1} in venous blood, even after a large carbohydrate-rich meal. Glucose and other simple sugars given in solution produce greater rises in blood glucose than equal or larger amounts of glucose-yielding carbohydrate taken as part of a solid mixed meal. Conversely, prolonged starvation for as long as several weeks rarely causes the blood glucose concentration to fall below 3 mmol l^{-1}, except in children and adults with impaired gluconeogenesis.

The remarkable ability of the body to regulate the size of the glucose pool under such widely diverse conditions depends mainly upon two organs – the liver and the pancreas – although during prolonged starvation, the kidneys become important generators of new glucose molecules.

Effects of Feeding on Blood Glucose

Glucose

Glucose and the two lesser dietary monosaccharides – fructose and galactose – enter the circulation through the intestinal mucosa. The speed with which they can be absorbed is limited by the rate of transfer from the intestine but rarely exceeds 50 g (0.28 mol) of carbohydrate, as glucose, per hour. This comparatively massive influx of glucose into a pool of ~20 g ordinarily produces a remarkably small perturbation in blood glucose, as the rate of removal from the glucose pool rises to match glucose input.

In healthy people, arterial blood glucose concentrations generally return to fasting levels within 2 h of eating a carbohydrate-rich meal. This remarkable feat of homeostasis is achieved through the prompt, but appropriate, release of insulin into the circulation. This is a consequence of stimulation of pancreatic B cells (the source of insulin) by a rising arterial blood glucose concentration augmented by nervous impulses originating in the brain (cephalic phase), mouth, gut wall, and portal vein, as well as the insulinotropic hormones, GIP and GLP-1, released by endocrine cells in the intestinal mucosa.

Under the influence of the insulinemia so produced, hyperglycemia resulting from glucose inflow from the gut and a reduction in glucagon secretion, the liver reduces its rate of glucose input into the pool and starts extracting it. Peripheral insulin-sensitive tissues, such as connective tissue, skin, fat, and especially striatal muscle start removing glucose. As a result, the arterial blood glucose concentration falls, and the stimulus to insulin secretion declines.

Ordinarily, the rates of change of glucose inflow from the gut into the glucose pool and the outflow of glucose into the tissues are so well aligned that arterial blood glucose levels rarely fall below fasting levels after a meal, and then only temporarily. Under the somewhat unnatural conditions resulting from ingestion of large amounts of sugar in solution on an empty stomach a 'reactive hypoglycemia' may result from persistence of insulin action after plasma insulin has fallen to basal levels and all of the glucose has been absorbed from the gut. It may be, but usually is not, sufficiently severe to produce (neuroglycopenic) symptoms, even in perfectly healthy individuals.

Disposal of an Oral Glucose Load

The exact disposition of glucose absorbed from the gut after a carbohydrate-rich meal by healthy subjects varies widely from individual to individual, and depends on the size, composition, and physical nature of the meal.

All in all, some 70% of a 70-g oral glucose load is taken up by the peripheral tissues, where most of it is used, within 4 h of ingestion, to generate energy by oxidation to carbon dioxide and water. The remaining 30% is removed by the liver during its passage from the gut to the periphery and converted into glycogen, triglycerides, or other metabolites.

Volunteers given a meal consisting of glucose (1 gram per kilogram of bodyweight) as a 45% solution on an empty stomach reduced their normal basal release of glucose from preformed glycogen in the liver from $9\,g\,h^{-1}$ to about $2.2\,g\,h^{-1}$, i.e., by about 75%. This persisted for a period ($\sim 3\,h$), during which glucose was being absorbed from the gut. In other words, although there is a small net uptake of glucose by the liver following a carbohydrate-rich meal, the liberation of glucose from preformed glycogen does not cease completely. Put differently, glycogenolysis and gluconeogenesis take place simultaneously though at different rates, depending on whether glucose is or is not being absorbed, as well as on the amount and nature of the hormones released by the pancreas and intestine in response to the presence of food.

Fructose and Galactose

Galactose and fructose are absorbed from the gut by different mechanisms from one another: galactose shares a transporter mechanism with glucose, whereas fructose has a less efficient one of its own. Once in the portal circulation, both sugars are rapidly taken up by the liver and largely converted into glycogen. This, together with glycogen formed from glucose, provides a small store of carbohydrate that is released as glucose into the body pool during the period when absorption from the gut is no longer occurring, and gluconeogenesis has not yet become fully reestablished.

Starches and their hydrolytic products are converted enzymatically into glucose in the gut lumen and brush border of the mucosa at a rate dependent upon their exact composition. Some starches are absorbed as rapidly as performed glucose, whereas others are absorbed much more slowly. Sucrose is cleaved into glucose and fructose, and lactose into glucose and galactose, before absorption into the body. Intraluminal hydrolysis is rarely a factor in limiting the rate of absorption of simple sugars. Each moiety is dealt with separately.

Postabsorptive Stage

The exact duration of the absorptive phase that follows ingestion of a meal depends on many factors. These include the rate of gastric emptying as well as the size, composition and physical nature of the food. It is unlikely, however, that an average adult eating three meals a day is truly postabsorptive, i.e., absorbing no glucose at all through the intestinal mucosa, for more than a few hours in any 24-h period, and this is mainly between 02.00 and 08.00. During this brief fasting period, glucose lost from the glucose pool by a constant drain into the brain and erythron is replaced by glucose from the liver. This is derived either from reserves of glycogen built up during the absorptive phase of a meal or from new glucose molecules formed from glucose precursors such as lactate, pyruvate, glycerol, and alanine, brought to it in the blood from peripheral tissues. Gluconeogenesis increases under the influence of rising levels of glucagon and fatty acids, both of which are a consequence of falling plasma insulin levels.

The amount of glycogen in the liver varies with the nature of the diet, the size and composition of the last meal, and its timing. The average amount of glycogen in the liver after an overnight fast is about 44 g (range 15–80 g) and does not increase very much after a meal; in other words, gluconeogenesis is already well under way within 12 h of eating the last meal. Nevertheless, after 36 h without food, liver glycogen stores may fall to as low as 4–8 g. Paradoxically, more prolonged fasting has little additional effect: indeed, hepatic glycogen stores may actually be replenished as the brain shifts from using glucose to β-hydroxybutyrate as its main source of energy.

Glycogen probably never disappears from the liver completely except *in extremis*, and there is evidence that it may be an intermediary in the production of glucose by the gluconeogenic pathway. Striatal muscles lack glucose-6 phosphatase, and consequently, although they contain substantial amounts of glycogen, they cannot convert and release it into the blood as glucose. They can, however, release its main breakdown product, lactate, into the blood for reconversion into glucose in the liver.

Gluconeogenesis

The mechanism whereby the liver and, to a smaller but significant extent, the kidneys make new glucose molecules from chemically simpler compounds is referred to as 'gluconeogenesis.' Much attention has been paid to the role of specific hormones, such as glucagon and cortisol, and the enzymes they affect, in determining the rate of gluconeogenesis. The supply of glucose precursors is also important. In humans, lactate is probably the most important precursor, especially during exercise. Others, in descending order of importance, are alanine, pyruvate, glycerol, and, finally, some glucogenic amino acids, including glutamate. The last named is especially important in gluconeogenesis in the kidney. Fatty acids, apart from propionate, do not serve as glucose precursors to any significant degree.

The contribution made by alanine to gluconeogenesis may have been exaggerated in the past, though it does have a role in transporting three-carbon skeletons derived from muscle glycogen to the liver

during fasting for conversion into glucose. It is also formed from amino acids released by proteolysis of nonstructural muscle proteins during periods of prolonged fasting and starvation.

Gluconegenesis is inhibited by eating, mainly through an increase in insulin, and a decrease in glucagon, action. It is enhanced by fasting, probably by the reverse procedure. Alcohol specifically inhibits gluconeogenesis, from lactate but not alanine, by adversely changing the redox potential within the hepatocytes and reducing the availability of nicotinamide adenine dinucleotide (NAD), which is an essential component in the formation of glucose from lactate but not, for example, from other sugars. The inhibition of gluconeogenesis by modest amounts of alcohol can be so profound that severe hypoglycemia may develop in fasting subjects, especially children, who are totally dependent upon gluconeogenesis for glucose needed by their brain and erythron.

Hormones and Glucose Homeostasis

Insulin is the only major hormone capable of lowering blood glucose levels. It does so by inhibiting glycogen breakdown in the liver and inhibiting gluconeogenesis and by encouraging glucose to leave the glucose pool by entering peripheral tissues. It achieves this mainly by activating the glucose transporter protein GLUT 4: an action that is enhanced by exercise and hyperglycemia. Consequently, insulin lowers blood glucose by two independent mechanisms. Which of the two actions predominates depends upon the circumstances: one of the most important is the concentration of insulin in the blood; another is whether it is of exogenous or endogenous (pancreatic) origin. Exogenous insulin reaches peripheral tissues at a concentration in blood greater than in blood perfusing the liver and is unaccompanied by C-peptide. Endogenous insulin, however, reaches the liver at a higher concentration than peripheral tissues and is accompanied by C-peptide, for which there is increasing evidence of synergism with insulin action.

Insulin released into the portal circulation is partially or completely removed by the liver. Insulin injected into skin, muscle, or a vein reaches insulin-sensitive tissues in the periphery at a concentration equal to, or greater than, the liver. Not all tissues on which insulin acts are equally sensitive to its actions: fat cells, for example, are more sensitive to its antilipolytic actions than muscle cells are to its glucose uptake-stimulating properties.

At the concentration at which insulin normally circulates in the peripheral blood of fasting subjects ($\sim 30\,\text{pmol}\,\text{l}^{-1}$), it depresses, but does not completely suppress, the unbridled release of fatty acids from lipocytes. At this concentration, insulin does not encourage glucose uptake by striatal muscle – which falls almost to zero. At insulin concentrations seen in peripheral blood in the absorptive phase of a meal (~ 150–$600\,\text{pmol}\,\text{l}^{-1}$), its effect on peripheral glucose uptake is pronounced and responsible for the marked arteriovenous glucose difference observed at this time. Similar plasma insulin levels are achieved during insulin therapy for diabetes.

The release of insulin from the B cells of the pancreatic islets is highly dependent upon the concentration of glucose in the blood perfusing them. At blood glucose levels below about 3.5–$4.0\,\text{mmol}\,\text{l}^{-1}$, insulin secretion is minimal. This means that as the arterial blood glucose falls towards its basal level in the post-absorptive state, plasma insulin levels also fall. They never fall low enough, however, in the nondiabetic subject to permit uncontrolled liberation of glucose by the liver or fatty acids by adipocytes. This does of course happen when the B cells are destroyed and results in gross hyperglycemia and ketosis, which are the hallmark of insulin-dependent diabetes.

During prolonged starvation (20 days or more without food), small amounts of insulin reach the liver. The amounts reaching the adipocytes are, however, insufficient to prevent seemingly uncontrolled lipolysis leading to hyperketonemia comparable with that seen in diabetic keto-acidosis (~ 10–$20\,\text{mmol}\,\text{l}^{-1}$). The situation differs, however, from that obtained in diabetes, in that the restraining effect of insulin on hepatic gluconeogenesis and glycogenolysis remains. Consequently, blood glucose levels are normal rather than grossly elevated as in insulin deficiency resulting from B cell malfunction.

A consequence of the reduction of insulin release as blood glucose levels fall after absorption of a meal is liberation of the A cells, which are 'downstream' of B cells in the islet, from the suppressive effect of (endogenous) insulin on their own release of glucagon. On reaching its target organ, glucagon promotes glycogenolysis and gluconeogenesis in the liver, in effect reversing the effect of insulin during the absorptive phase. In other words, each islet functions as a miniature homeostat.

Counterregulatory Hormones

Whilst it is possible to explain the control of blood glucose largely by means of the servoregulatory control of insulin and glucagon secretion described above, the body also has many neural and hormonal mechanisms at its disposal to correct or overcome any fall in blood glucose to below the critical level necessary for the maintenance of normal brain function. The sensors for this regulatory function are located in at least two anatomically distinct sites within the brain and in the portal vein itself.

The most important mechanisms involved are:

- stimulation of the sympathetic and parasympathetic nervous systems, which in turn lead, respectively, to release of adrenaline from the adrenal medula and noradrenaline from nerve terminals in the liver, and glucagon from the pancreas;
- secretion of vasopressin by the posterior pituitary gland; adrenocorticotrophic hormone (ACTH), growth hormone, and prolactin by the anterior pituitary gland; and cortisol by the adrenal cortex.

None of these hormones, apart from cortisol, appears to be absolutely essential individually for the maintenance of normal glucose homeostasis, but all are brought into play under adverse conditions. They produce their hyperglycemic effects in a variety of ways that can be summarized as follows:

1. Increasing the liberation of glucose by the breakdown of preformed glycogen in the liver, e.g., glucagon, adrenaline, noradrenaline, vasopressin.
2. Increasing gluconeogenesis in the liver, e.g., glucagon, cortisol.
3. Decreasing peripheral glucose utilization by peripheral tissues, e.g., growth hormone, cortisol, and prolactin.

Glycosuria

Upwards of 100 g of glucose are normally filtered from the blood at the glomeruli of the kidneys each day, more than 99% of which is reabsorbed by the kidney tubules. As a result, healthy people lose less than 150 mg of glucose in their urine each day, an amount too small to be detected by most simple screening procedures for glycosuria. When, for any reason, the amount of glucose filtered at a glomerulus is more than can be reabsorbed by the tubules, glucose appears in the urine at a concentration many times greater than it occurs in blood.

The commonest cause of significant glucosuria, i.e., the presence of glucose in the urine at a concentration greater than $0.15\,g\,l^{-1}$ ($0.8\,mmol\,l^{-1}$) is a blood glucose concentration of $10\,mmol\,l^{-1}$ or more. At or above this concentration, the amount of glucose filtered at the glomerulus is more than its associated kidney tubule can transport back into the circulation. Once this reabsorptive threshold has been exceeded, any increase in filtered glucose load is reflected by a large increase in the amount of glucose eliminated in the urine. The osmotic diuresis so produced is associated with an increased excretion of water, sodium, chloride, and potassium and is often the first clue to the existence of hyperglycemia, the characteristic hallmark of diabetes.

Glycosuria can occur at normal blood glucose levels when, for example, blood flow through the glomerulus is increased, such as during pregnancy or when, owing to either an inherited or acquired defect of renal tubular glucose transport, even the normal amount of glucose filtered at the glomerulus cannot be reabsorbed. Conversely, in people in whom blood flow through the glomeruli is reduced or in whom glomerular filtration is impaired, but in whom normal tubular glucose reabsorption is retained, even gross hyperglycemia may not be accompanied by glycosuria.

Fructose and galactose do not normally appear in urine since their concentration in blood is never sufficiently high – except in disease – for them to do so.

See also: **Carbohydrates**: Metabolism of Sugars; **Diabetes Mellitus**: Etiology; Chemical Pathology; **Fructose**; **Galactose**; **Glucose**: Properties and Analysis; Function and Metabolism; Maintenance of Blood Glucose Level; Glucose Tolerance and the Glycemic (Glycaemic) Index; **Glycogen**

Further Reading

Bolli GB and Fanelli CG (1999) Physiology of glucose counterregulation to hypoglycemia. *Endocrinology and Metabolism Clinics of North America* 28: 467–493.

Marks V, Samols E and Stagner J (1992) Intra-islet interrelationship. In: Flatt PR (ed.) *Nutrient Regulation of Insulin Secretion*, pp. 41–57. London: Portland Press.

Olson AL and Pessin JE (1996) Structure, function, and regulation of the mammalian facilitative glucose transporter gene family. *Annual Review of Nutrition* 16: 235–256.

Wahren J and Johansson B-L (1998) New aspects of C-peptide physiology. *Hormone & Metabolic Research* 30: A2–A5.

Glucose Tolerance and the Glycemic (Glycaemic) Index

G Frost, Nutrition and Dietetic Research Group, London, UK

Introduction

This article examines the influence of dietary carbohydrates on postprandial intermediary metabolism and insulin resistance. The link between high-glycemic-index diets with insulin resistance, type II diabetes, and cardiovascular disease (CVD) is discussed. We highlight the potential benefits of

low-glycemic diets and discuss the inadequacy of today's classification of carbohydrates to identify low-glycemic foods. We conclude by advocating that nutritional guidelines for diabetes should place greater emphasis on the glycemic index.

Glycemic Index – Definition

The glycemic index of several foods is published in international nutritional tables. Methodology on their derivation is available from previous reviews. In summary, the glycemic index is a measure of a carbohydrate's postprandial glucose response (Figure 1). The glycemic index provides a standardized comparison of a carbohydrate's 2-h postprandial glucose response with that of white bread or glucose (Table 1).

$$\frac{\text{Incremental area under blood glucose response curve for the test containing 50 g carbohydrate}}{\text{Corresponding area after equi-carbohydrate portion of white bread}} \times 100$$

Low-glycemic-index carbohydrates have lower 2-h areas under the glucose curve than white bread, while high-glycemic-index foods have higher areas. Although the insulin response is not part of the glycemic index calculation, the lower the glycemic index of a food, the more attenuated is the insulin response.

Type of Dietary Carbohydrate and the Glycemic Index

The present classification of carbohydrates is based on structure rather than physiology. The glycemic index of a carbohydrate, however, is determined more by physiology being highly dependent on a carbohydrate's absorption rate. The glycemic index is influenced by a carbohydrate's composition, tertiary structure, and enzymic digestion (Table 2).

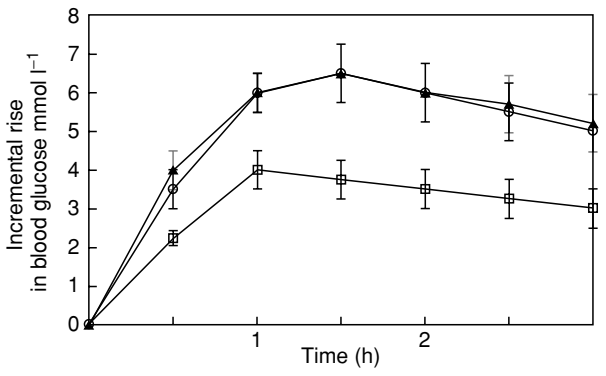

Figure 1 Mean blood glucose increment after equi-carbohydrate meals. Squares, pasta; circles, wholemeal bread; triangles, white bread.

Chain Length and Composition

Complex carbohydrates are polymeric chains of repeating monosaccharide units. Starches are comprised of repeating glucose units. The glycemic indexes of different starches are determined by their susceptibility to enzymic digestion, not chain length. White bread and pasta have similar chain lengths but bread has a higher glycemic index, due to a tertiary structure and solubility that ensure greater exposure to salivary and pancreatic amylases.

Short-chain carbohydrates are rapidly absorbed; however, when they contain nonglucose sugars the glycemic index is lowered proportionally. The

Table 1 An example of glycemic index of carbohydrate-containing foods

Food	*Glycemic index*
Breads	
Rye (wholegrain, i.e., pumpernickel)	68
White and wholemeal wheat bread	100
Pasta	
Spaghetti (white, boiled 15 min)	67
Cereal grains	
Barley (pearled)	36
Rice (polished, boiled 10–25 min)	81
Sweetcorn	80
Breakfast cereals	
All Bran	74
Cornflakes	121
Porridge oats	89
Shredded Wheat	97
Weetabix	109
Biscuits	
Digestive	82
Rich Tea	80
Plain crackers (water biscuits)	100
Root vegetables	
Potato (boiled)	80
Potato (mashed)	98
Potato (sweet)	70
Yam	74
Legumes	
Baked beans (canned)	70
Chick peas (canned)	60
Kidney beans (canned)	74
Lentils (green, dried)	36
Fruit	
Apple	52
Orange	59
Orange Juice	71
Sugars	
Glucose	138
Sucrose	83
Dairy products	
Icecream	69
Skim milk	46

Table 2 Comparison of common carbohydrate-containing foods and their glycemic index

Food type	*Carbohydrate type*						*Glycemic index*
	Refined	*Unrefined*	*Starch*	*Sugar*	*Soluble*	*Nonsoluble*	
White bread	✓		✓				100
Wholemeal bread		✓	✓			✓	100
Weetabix		✓	✓			✓	100
Cornflakes	✓		✓				121
Porridge		✓	✓		✓		89
Baked beans	✓		✓		✓		70
Digestive biscuits		✓	✓			✓	82
Apple		✓	✓		✓		49
Sucrose	✓			✓			83

disaccharides sucrose and lactose consist of 50% glucose and 50% fructose or galactose, respectively, both have a lower glycemic index than maltose, the disaccharide formed from two glucose molecules.

Amylose and Amylopectin

The starches in cereal grains, rice, potatoes, and all green plants are composed of repeating glucose units arranged in straight, amylose, and branched-chain, amylopectin polysaccharides. The absorption rate, and hence the glycemic index, of these starches is influenced by the ratio of amylose to amylopectin. The more compact structure of amylose than amylopectin results in a smaller surface area being available for amylase digestion. Amylose-enriched starches have lower glycemic indexes than those enriched in amylopectin.

Relationship of Insoluble and Soluble Nonstarch Polysaccharides (Fiber) to Glycemic Index

Nondigestible complex carbohydrates are commonly known as dietary fiber. The more correct terminology is nonstarch polysaccharides (NSP). Dietary fiber/NSP is either soluble or insoluble. Clinical studies have shown that diets rich in soluble fiber/NSP, such as guar gum, pectin, and sugar beet fibers lower postprandial blood glucose and insulin levels. Guar gum, a beta-glactomannan from the Indian locust bean, also reduces postprandial lipemia. Nonsoluble NSP has no effect on dietary glycemic index.

Soluble NSP such as pulses, vegetables, whole fruits, oats, and barley form gelatinous gels within the stomach which delay gastric emptying and enzymic digestion, the latter by forming a physical barrier around the carbohydrate. Insoluble NSPs have little effect on gastric emptying and no effect on glucose absorption. High-fiber/high-NSP diets are therefore not necessarily synonymous with low-glycemic foods. Cellulose is the most widely used NSP in household cereals, wholemeal bread, and brown rice, and as cellulose is insoluble, these foods have the same glycemic index whether replete or deplete of their dietary fiber/NSP. For unknown reasons, All Bran is an exception and, despite its high insoluble fiber content, has a low glycemic index.

The solubility of dietary fiber/NSP needs to be specified if benefits on postprandial glycemia and hyperinsulinemia are to be expected. Nonmetabolic benefits are derived from increasing nonsoluble fiber/NSP, and are notable on bowel function and bowel pathology. However, protection against colon cancer has recently been questioned.

Cell Structure, Food Preparation, and Processing

Cooking and food preparation can modify the glycemic index. Highly processed convenience foods tend to have high glycemic indexes. When cooking and processing disrupt the cell wall of starch granules, the glycemic index is increased. Cooked pulse vegetables have low glycemic indexes as their cell walls are resistant to cooking. The intact cereal grains of pumpernickel rye bread, granary bread, and burgule wheat all have low glycemic indexes. However, when granary bread is processed to wholemeal bread these grains are disrupted and the glycemic index rises. Cooling can paradoxically lower the glycemic index of certain starches, such as potatoes, due to the formation of retrograde starches which are resistant to amylase digestion. Food particle size as well as the constituency of the food bolus on swallowing also influence absorption rates, as seen with boiled and mashed potatoes.

Concerns Relating to the Glycemic Index

Despite the 1998 World Health Organization/Food and Agriculture Organization nutritional report advocating greater use of the glycemic index, no American or British dietetic guidelines promote its use. In fact, the only guidelines that do support the use of the glycemic index are the 1998 European guidelines.

Carbohydrates are still broadly grouped on their biochemical basis into refined and complex carbohydrates; starch and NSPs and soluble and insoluble fiber. Part of the failure of the glycemic index not being more widely used is due to concerns relating to its reproducibility, accuracy when part of a mixed meal, and its overall validity. In terms of reproducibility, the glycemic index, like all *in vivo* measurements, such as the oral glucose tolerance test, is dependent on physiological variables and, if these are not controlled for, some variability in its measurement is inevitable. Concerns relating to the glycemic index when part of a mixed meal are theoretical rather than actual, such as the glycemic response of a mixed meal, can be reliably calculated knowing the glycemic index of each constituent carbohydrates and their proportional contribution to that meal. The concerns on validity are centered on the way the glycemic index is calculated, emphasizing the 2-h integrated area rather than the pattern of response. The way the glycemic index is calculated can result in two obviously metabolically different carbohydrates, which would evoke different patterns of glucose and insulin release having similar glycemic indexes (Table 2), for example, a rapidly absorbed and metabolized carbohydrate, such as sucrose, and a complex carbohydrate with poor intestinal absorption and subsequent colonic fermentation. However, in practice the glycemic index of most carbohydrates provides a good indicator of the 2-h plasma glucose and insulin exposure.

Clinical Significance of Postprandial Versus Fasting Hyperglycemia in Diabetic and Nondiabetic Populations

Postprandial hyperglycemia in nondiabetic populations is a stronger predictor of insulin resistance and CVD than fasting glucose. The combined 20-year mortality data on men from the Whitehall, Paris Prospective, and Helsinki Policemen studies showed that the upper quintile compared with the lower for the 2-h postplasma load glucose was associated with a 2.7 times increased risk of CVD mortality. The fasting glucose values were less discriminatory for CVD, with only the upper 2.5% values conferring a 1.8-fold increased mortality risk. Over a 7-year period, elderly women in a study with isolated postprandial hyperglycemia, 2-h value > 11.1 mmol l^{-1} and fasting value < 7.0 mmol l^{-1} on a 75-g oral glucose tolerance test had an approximately threefold increased risk of heart disease when compared with women whose 2-h values were below 11.1 mmol l^{-1}.

In established diabetes postprandial glycemia appears to have a stronger relationship with microvascular and macrovascular disease than fasting blood glucose. Similarly, in gestational diabetes adverse pregnancy outcome is more closely related to postprandial glycemia than fasting and premeal glycemic values.

The Benefits of Low-Glycemic-Index Carbohydrates on Diabetic Control

Three-month clinical studies have shown that low-glycemic-index carbohydrates improve glycemic control in both type 1 and type 2 diabetes and reduce postprandial glucose and insulin values. Good glycemic control and favorable lipid and fibrinolytic profiles have also been reported in individuals with type 1 and 2 diabetes who habitually consume low-glycemic-index dietary carbohydrates. It remains to be seen whether these diets bestow long-term benefits on micro- or macrovascular complications.

The Benefits of Low-Glycemic-Index Carbohydrates on CVD Risk Factors

High-glycemic-index foods induce postprandial hyperinsulinemia, which is a powerful predictor for metabolic risk factors and CVD in epidemiological studies. Both cross-sectional and prospective population studies have shown favorable lipid profiles in association with high-carbohydrate diets. Initially, these benefits were attributed to a high fiber content. However, when the glycemic index is controlled for, it is the low-glycemic-index diets rather than high fiber content that have the greatest influence on high-density lipoprotein (HDL)-cholesterol, insulin sensitivity, and fibrinolytic parameters. In one large cross-sectional study on middle-aged ambulatory subjects, glycemic index was a stronger determinant of HDL cholesterol than any other dietary carbohydrate or fat. Low-glycemic-index diets also lower serum cholesterol and triglyceride levels in hyperlipidemic subjects.

Potentially Detrimental Effects of a High-Carbohydrate Diet

High-carbohydrate diets (> 60% of total dietary energy) that consist predominantly of high glycemic carbohydrates have detrimental metabolic effects. These diets increase serum triglyceride and insulin resistance, having the greatest adverse effect in insulin-resistant states, such as type 2 diabetes or pregnancy. As a consequence of this, the American Diabetic Association recommends limiting dietary carbohydrates to 45% in diabetic pregnancies and to 80% of total energy derived from carbohydrate

and fat outside pregnancy. By contrast, the British Diabetic Association's guidelines advocate that 55% of total dietary energy should come from dietary carbohydrates. The European guidelines are the only ones that specify the use of low-glycemic carbohydrates and the current American Dietetic Association recommendations for the USA give glycemic index an evidence-based rating of B, suggesting more long-term studies are needed.

Several short-term clinical studies have shown that high-carbohydrate/low-glycemic-index diets reduce insulin sensitivity and adversely influence lipid profiles in both normoglycemic and type 2 diabetic subjects. Similar results have been seen in healthy pregnancy and in individuals at risk of cardiovascular disease. In all these studies, substitution of the dietary carbohydrate with a low glycemic index reverses these trends. Interestingly, the very-low-density-lipoprotein (VLDL)-rich triglyceride particles generated from high-carbohydrate diets are thought to be less atherogenic than those associated with the insulin resistance syndrome.

High-carbohydrate diets have been associated with both the development of and the prevention of type 2 diabetes and CVD. These conflicting studies are likely to be explained by differences in dietary glycemic index. Previously many clinical and epidemiological dietary studies have not specified the glycemic index. High-carbohydrate diets in non-industrialized rural populations, such as Africa, are typified by low-glycemic starches and soluble NSP, carbohydrates associated with reduced insulin resistance, and metabolic CVD risk factors. By contrast, high-carbohydrate diets in westernized industrial societies such as North America, with high background incidences of obesity, insulin resistance, and type 2 diabetes, contain high-glycemic carbohydrates due to food processing and soft drinks, and it is these diets which are metabolically disadvantageous.

Obesity and Glycemic Index

Obesity contributes to the pathogenesis and morbidity of type 2 diabetes. Obesity is associated with changes in carbohydrate and fat metabolism that are central to the development of insulin resistance. While low-glycemic-index diets reduce insulin sensitivity and improve metabolic cardiovascular risk factors, they will not reduce weight unless part of an energy-deficient diet. However, in obese subjects, when low-glycemic carbohydrates are incorporated into a hypocaloric diet, there is a greater fall in insulin resistance than can be accounted for by weight loss alone.

Glycemic Index and the Prevention of Type 2 Diabetes

Changes in diet and physical activity levels, both alone and in combination, reduce the progression of impaired glucose tolerance to diabetes. Two large American prospective population studies have demonstrated protection against type 2 diabetes for both men and women when the habitual diet is characterized by having a low glycemic index and low fat content. A similar protective effect against diabetes has been reported in populations consuming high-fiber foods and high quantities of fruit and vegetables. Although it was not calculated, one would predict that these diets would also have a low glycemic index.

Pregnancy and Glycemic Index

Throughout pregnancy in well-nourished urbanized women consuming typical westernized diets, there is a deterioration in glucose tolerance. African women during pregnancy, living in traditional rural populations and consuming high-carbohydrate/low-glycemic-index diets, do not invariably experience deterioration in their glucose tolerance. Clinical studies in the west show that women consuming similar high-carbohydrate/low-glycemic-index diets throughout pregnancy also have no deterioration of glucose tolerance despite the physiological increase in insulin resistance that occurs secondary to maternal and placental hormones. When the proportion of dietary carbohydrate increases above 50% in women with gestational diabetes, if no emphasis no low-glycemic-index carbohydrates is given, glucose tolerance will deteriorate.

Proposed Mechanism by which Dietary Carbohydrates Influence Insulin Resistance

Adipocyte metabolism is central to the pathogenesis of insulin resistance and dietary carbohydrates influence adipocyte function. The previous simplistic view, that insulin resistance resulted from the downregulation of the insulin receptors in response to hyperinsulinemia, is being replaced by the hypothesis that high circulating nonesterified fatty acid (NEFA) levels both impair insulin action and reduce pancreatic β-cell secretion. A plausible explanation of how low-glycemic-index carbohydrates can reduce insulin sensitivity is by their ability to reduce adipocyte NEFA release.

Adipocyte release of NEFA is increased in insulin resistance as a consequence of suppressed hormone-sensitive lipase activity. An inverse relationship between adipocyte NEFA release and insulin sensitivity

has been demonstrated in patients with coronary heart disease. High circulating NEFA levels reduce muscle uptake of glucose and downregulate the insulin receptor. High circulating NEFA levels reduce hepatic insulin clearance; this exaggerates the pre-existing hyperinsulinemia and insulin resistance. Reduced lipoprotein lipase activity increases NEFA levels and hepatic VLDL production; however, VLDL clearance is reduced, and this results in a lower high-density lipoprotein concentration and increased TG-enriched VLDL particles. The latter are catabolized into smaller atherogenic forms of LDL. The metabolic consequences of increased adipocyte NEFA release therefore fuel lipoprotein changes associated with coronary heart disease.

High circulating NEFA levels also adversely affect insulin secretion. High NEFA levels have been postulated to have a lipotoxic effect on the β-cells due to their ability to increase nitric oxide production which is cytotoxic to islets. In addition, when triglyceride accumulates within the β-cell, glucose-stimulated insulin secretion is blunted.

Many of the metabolic benefits associated with low-glycemic-index carbohydrates can be attributed to their ability to reduce adipocyte NEFA release. Low-glycemic-index foods have been consistently shown to improve insulin sensitivity and animal studies have shown that improvements in fat and muscle insulin sensitivity are accompanied by decreases in fatty acid synthatase activity, adipocyte size, and lipid storage. Although human studies have shown that low-glycemic-index diets consumed over 3 weeks increase adipocyte insulin sensitivity, no direct effect on adipocyte metabolism has been identified so far.

Low-glycemic-index diets, by attenuating the insulin response, ensure a more prolonged suppression of postprandial NEFA levels than those following high-glycemic foods. Low-glycemic meals taken in the evening can effectively suppress circulating NEFA concentrations and hepatic glucose output throughout the night. These metabolic effects would be predicted to promote insulin sensitivity.

Our own work has shown that insulin-resistant adults with a history of, or who are at risk of CHD improve their adipocyte insulin sensitivity when consuming a low-glycemic-index diet and in addition their circulating NEFA levels decline. These human studies complement the animal work that show low-glycemic-index diets improve insulin sensitivity by modulating adipocyte metabolism.

Conclusion

The glycemic index of a diet provides a measure of postprandial metabolism. Dietary carbohydrates are absorbed and metabolized differently and therefore influence postprandial glucose, insulin, and NEFA concentrations differently. In western society the proportion of the day that we spend in the postprandial state is increasing as the tendency to snack throughout the day replaces sit-down main meals. The known detrimental consequences of high-glycemic foods on postprandial metabolism should encourage us to advocate low-glycemic diets to counteract the present epidemic of insulin resistance-related diseases, notably CVD and diabetes. The relevance of the glycemic index to these two major preventable diseases of the western world argues strongly for its greater acceptance into our current nutritional guidelines.

See also: **Carbohydrates**: Metabolism of Sugars; **Coronary Heart Disease**: Antioxidant Status; **Diabetes Mellitus**: Etiology; Chemical Pathology; **Obesity**: Etiology and Diagnosis; **Pregnancy**: Nutrition in Diabetic Pregnancy

Further Reading

Alberts DS, Ritenbaugh C, Story JA, Aickin M, Rees A, McGee S, Buller MK, Atwood J, Phelps J, Ramanujam PS *et al.* (1996) Randomized, double-blinded, placebo-controlled study of effect of wheat bran fiber and calcium on fecal bile acids in patients with resected adenomatous colon polyps. *J Natl Cancer Inst* 88: 81–92.

American diabetes association (1999) Nutritional recommendations and principles for people with diabetes mellitus. *Diabetes Care* 22 Suppl 1: S42–S46

Balkau B, Shipley M, Jarrett RJ, Pyorala K, Pyorala M, Forhan A and Eschwege E (1998) High blood glucose concentration is a risk factor for mortality in middle-aged nondiabetic men. 20-year follow-up in the Whitehall Study, the Paris Prospective Study, and the Helsinki Policemen Study. *Diabetes Care* 21: 360–367.

Englyst HN, Trowell H, Southgate DAT and Cummings JH (1987) Dietary fiber and resistant starch. *Am J Clin Nutr* 46: 873–881.

FAO/WHO Report of a joint FAO/WHO report Rome (1998) 14–18 April 1997, editor. Carbohydrates in human nutrition. Paper 66. FAO Food and Nutrition.

Foster-Powell K and Miller JB (1995) International tables of glycemic index. *Am J Clin Nutr* 628: 71S–93S.

Frost G, Leeds A, Trew G, Margara R and Dornhorst A (1998) Insulin sensitivity in women at risk of coronary heart disease and the effect of a low glycaemic index diet. *Metabolism* 47: 1245–1251.

Frost GS, Keogh B, Smith D, Leeds AR and Dornhorst A (1998) Reduced Adipocyte Insulin sensitity in Caucasian and Asian Subjects with Coronary Heart Disease. *Diabetes Medicine* 15: 1003–1009.

Frost G, Leeds AR, Dore CJ, Madeiros S, Brading S and Dornhorst A (1999) Glycaemic index as a determinant of serum HDL-cholesterol concentration. *Lancet* 353: 1045–1048.

Fuchs CS, Giovannucci EL, Colditz GA, Hunter DJ, Stampfer MJ, Rosner B, Speizer FE and Willett WC (1999) Dietary fiber and the risk of colorectal cancer and adenoma in women. *New Eng J Med* 340: 169–176.

Ha KK and Lean MEJ (1998) Recommendations for the nutritional management of patients with diabetes mellitus. *Eur J Clin Nutr* 52: 467–481.

Hanefield M, Temelkova-Kurktschiev T, Schaper F, Henkel E, Siegert G and Koehler C (1999) Impaired fasting glucose is not a risk factor for atherosclerosis. *Diabet Med* 16: 212–218.

Holt SHA, Miller JCB and Petocz P (1997) An insulin index of foods: the insulin demand generated by 1000-kJ portions of common foods. *American Journal Of Clinical Nutrition* 66: 1264–1276.

Jarvi AE, Karstrom BE, Granfeldt Y, Bjorck I, Asp NG and Vessby B (1999) Improved glycemic control and lipid profile and normalized fibrinolytic activity on a low-glycaemic index diet in type 2 diabetes. *Diabetes Care* 22: 10–18.

Jenkins DJA, Kendall CWC and Ransom TPP (1998) Dietary fiber, the evolution of the human diet and coronary heart disease. *Nutrition Research* 18: 633–652.

Kabir M, Rizkalla SW, QuignardBoulange A, GuerreMillo M, Boillot J, Ardouin B, Luo J and Slama G (1998) A high glycemic index starch diet affects lipid storage-related enzymes in normal and to a lesser extent in diabetic rats. *Journal Of Nutrition* 128: 1878–1883.

McGarry JD and Dobbins RL (1999) Fatty acids, lipotoxicity and insulin sectretion. *Diabetologia* 42: 128–138.

GLUCOSINOLATES

B Holst, Nestlé Research Center, Lausanne, Switzerland
G R Fenwick, Institute of Food Research, Norwich, UK

Introduction

Glucosinolates are secondary plant metabolites that are characteristic of the Cruciferae and related families within the order Capparales. Glucosinolate-containing plants have always made major contributions to the diets of humans and farm animals, and include condiments and relishes (mustard and wasabi), traditional leaf vegetables (cabbage, swede), and crops used as animal feedstuffs (rapeseed, kale, turnip).

Glucosinolates are thioglycosides that differ in the structure of the aglycone side chain; they are relatively biologically inactive *per se*, but following tissue disruption they undergo hydrolysis to form a broad range of structurally diverse hydrolysis products possessing varying biological activities. These products are the basis of the sensory characteristics typical of brassicas but also, through their biological effects, affect their nutritional value. Some of the hydrolysis products (goitrin, thiocyanate ion, several isothiocyanates, and nitriles) may have antinutritional or toxic effects; others, especially isothiocyanates (e.g., sulforaphane), are considered to be responsible for the protective, anticarcinogenic effects of a cruciferous-rich diet. These adverse and beneficial effects are highly dose-dependent and the physiological range is relatively narrow. Glucosinolate hydrolysis products also exert antifungal, antimicrobial, and insecticidal properties, and so contribute to the plant's overall defense mechanism.

The large number of glucosinolates, the plethora of hydrolysis products possessing differing biological activities, and the dose dependence of the effects observed make research in this area both challenging and complex. Thus, plant-breeding strategies that concentrated solely on reducing the glucosinolate content of rapeseed for nutritional benefits succeeded in producing very low ('zero') glucosinolate varieties that were more nutritious but these required a significantly increased expenditure for crop protection. On the other hand, pest-resistant crop cultivars that contain high amounts of individual glucosinolates may have unacceptable sensory characteristics and may, possibly, also exhibit antinutritional properties.

Chemical Structure, Properties, and Nomenclature

Naturally occurring glucosinolates are (Z)-*cis*-*N*-hydroximinosulfate esters, possessing a sulfur-linked β-D-glucopyranose moiety and an amino acid-derived side chain ($\mathbf{R_1}$: **Figure 1**). The side chain and sulfate group have an antistereochemical configuration, the structure of the former being highly variable; the glucosinolates may posses aliphatic (alkyl, alkenyl, hydroxyalkenyl, ω-methylthioalkyl), aromatic (benzyl, substituted benzyl), or heterocyclic (indolyl) side chains. Aliphatic- and aromatic glucosinolates share the same degradation pathways but glucosinolates with side chains possessing a terminal double bond or a β-OH group form different products. The major glucosinolates may be subdivided into

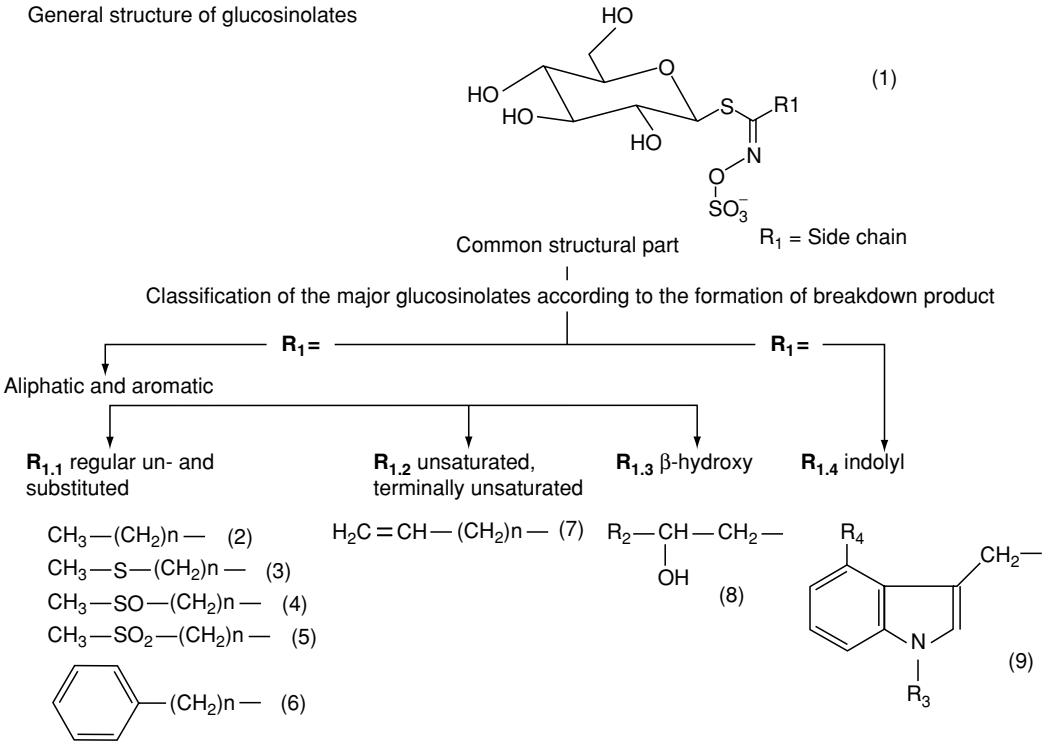

Figure 1 General and individual glucosinolate structures.

aliphatic and aromatic ($\mathbf{R_{1.1}}$), terminally unsaturated ($\mathbf{R_{1.2}}$), β-hydroxy-($\mathbf{R_{1.3}}$), and indolyl ($\mathbf{R_{1.4}}$) groups according to their breakdown. The first group includes a wide range of homologs ($n = 3$ to $n = 11$) and structural classes. The most common side chains include ω-methylthio (**3**), ω-sulfinyl (**4**), ω-sulfonylalkyl (**5**), and aromatic (**6**) groups.

The presence of the sulfate group confers strongly acidic properties on glucosinolates, irrespective of the chemical structure of the side chain, and they occur as salts with the glucosinolate anion usually counterbalanced by potassium, less frequently by sinapinium cation. The sulfate group and thioglucose moiety make glucosinolates nonvolatile and hydrophilic, and the naturally occurring forms exhibit levrotation in solution. Trivial nomenclature was used for many years, but latterly a more precise system has been developed that uses prefixes characteristic of the side chain followed by the word 'glucosinolate.' Although this nomenclature is more informative and prevents confusion, the trivial names are still widely used.

Occurrence

Glucosinolates are limited to a restricted number of dicotyledonous angiosperms and are prevalent throughout 15 botanical families of the order Capparales, e.g., the Brassicaceae, Capparaceae, Resedaceae. As more than 80% of glucosinolates so far identified occur in the Brassicaceae, their presence has been used as a chemotaxonomic criterion for classification within this family. Species of the Brassicaceae are of great economic importance as vegetables (e.g., cabbage, broccoli, cauliflower, Brussels sprouts; root vegetables, e.g., radish, turnip, swede, and leaf vegetables, e.g., rocket salad), relishes (e.g., wasabi, mustard), oilseeds (canola, rapeseed), and forage crops (kale, forage rape).

Glucosinolates have been found in all parts of these plants and, whilst as many as 15 different glucosinolates have been found in the same plant, three to four usually predominate. Occurrence and concentrations vary according to species and cultivar, tissue type, physiological age, plant health, agronomic practice, and nutrition. Levels in reproductive tissues (flowers and seeds) are much higher (up to 10% dry weight) than in vegetative tissues (1–5000 ppm). Glucosinolates are synthesized in the silique walls from relatively few protein amino acids and chain-elongated homologs and translocated to the seeds, where they accumulate inside the vacuole.

The common glucosinolate biosynthetic pathway involves aglycone synthesis via initial *N*-hydroxylation, oxidative decarboxylation to form an aldoxime, incorporation of sulfur to yield a thiohydroximic acid

that undergoes *S*-glucosylation to the desulfoglucosinolate and, finally, sulfation. Modifications of the side chain, including oxidation, hydroxylation, and reduction, are responsible for the observed structural diversity of glucosinolates.

Although several individual steps of the overall biosynthetic pathway remain to be elucidated and several enzymes require purification and characterization, existing knowledge may be exploited for breeding or biotechnological production of new plant varieties with enhanced flavor and functionality, including putative health-promoting activity. Thus breeding has produced rapeseed varieties with very low glucosinolate concentrations, e.g., the 'double zero' rapeseed with a very low content of 2-hydroxy-3-butenylglucosinolate, the precursor of 5-vinyloxazolidine-2-thione (goitrin). Recently there have been attempts to increase levels of certain glucosinolates in crops used as human foods, e.g. enhancing 4-methylsulfinylbutylglucosinolate, the precursor of the anticarcinogenic isothiocyanate, sulforaphene.

The role of the complex myrosinase (β-thioglucosidase, EC 3.2.3.1.)-glucosinolate system in plants is considered to be diverse; glucosinolates may represent a sink for nutrients like nitrogen, sulfur, or growth factors (e.g., indole-3-acetic acid, IAA) and, hence, may be involved in growth regulation, while their hydrolysis products contribute to nonspecific plant defense mechanism against insect, fungi, and microorganism infections (Table 1).

Whilst disruption of cellular tissue leads to hydrolyses of glucosinolates, under some conditions (after plant–fungi, plant–insect, and plant–microorganism interactions, or mechanical wounding) *de novo* synthesis may be induced. Thus, under conditions of low stress and minor cell damage, *de novo* synthesis of glucosinolates may balance – or indeed exceed – losses due to their breakdown. More work is required to establish the underlying biochemical mechanisms controlling these processes.

Table 1 Factors affecting glucosinolate profiles and concentration

Biotic factors	*Abiotic factors*
Fungi	Temperature
Insects	Light (ultraviolet irradiation)
Microorganisms	Water supply
Animals	Fertilizer
Weeds and other competing plants	Harvest conditions
Plant density	Postharvest and storage conditions
Developmental stage	Processing conditions
Part of the plant (organ)	

Hydrolysis

As emphasized above, hydrolysis products – rather than intact glucosinolates – are responsible for the biological activities and flavor characteristic of cruciferous vegetables. Depending on reaction conditions, glucosinolate breakdown may be enzymatic and/or nonenzymatic; since myrosinase co-occurs with glucosinolates in plant tissue, the former predominates, with chemical hydrolysis only taking place where myrosinase is inactivated, e.g., after cooking, or under highly acidic/basic conditions. Myrosinase isoenzymes and glucosinolates are localized in all cells but are compartmentalized; glucosinolates are stored in the vacuole and myrosinases are localized in cytoplasm.

Following tissue damage, enzyme and substrate come into contact, causing hydrolysis of the *S*-glucose bond and, thereby, yielding glucose and an unstable aglycone (**10**, Figure 2) that undergoes spontaneous rearrangement to an isothiocyanate, or nitrile and elemental sulfur. Only allyl-, benzyl-, and 4-methylthiobutyl glucosinolates appear to undergo enzymatic degradation to thiocyanates, although the mechanism remains obscure. Other products, such as epithionitriles and oxazolidine-2-thiones, can also be formed; the nature of the resulting product depending on side-chain structure and reaction conditions (pH, temperature, presence of protein cofactors modifying the action of the enzyme).

Structure of the Side Chain

At a pH of 6–7, most glucosinolates yield stable isothiocyanates (**11, 12**). However, those possessing a β-hydroxylated side chain form unstable hydroxyisothiocyanates that spontaneously cyclize to oxazolidine-2-thiones (**13**), the best-known example being 5-vinyloxazolidine-2-thione (goitrin). Indole glucosinolates also form unstable isothiocyanates; these undergo lysis, initially forming the corresponding alcohol, e.g., indole-3-carbinol (**14**) and subsequently condensing to the dimer (**15**), trimer, or tetramer. In the presence of ascorbic acid, ascorbigen (**16**) is the major product (Figure 3).

Depending on the plant species, autolysis of fresh material (pH $\sim$ 6) can yield nitriles (**17–21**). Several authors have discussed a 'nitrile-forming factor' but this has neither been isolated nor characterized. Under acidic conditions, nitriles are the major degradation products (**17, 18, 20, 21**: Figure 4). In the presence of active epithiospecifier protein (ESP) and Fe^{2+} ions, glucosinolates with terminally unsaturated side chains transfer sulfur from the *S*-glucose moiety to the alkenyl moiety to form an epithionitrile (**19**). ESP is a small protein, which co-occurs and interacts

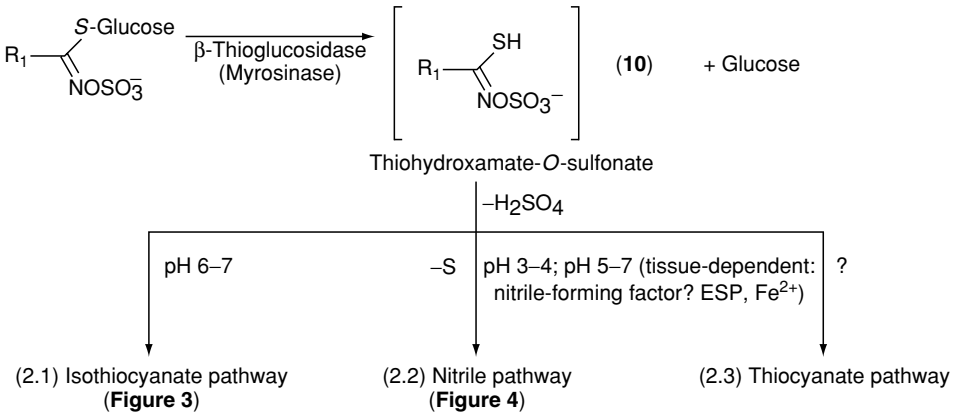

Figure 2 General and initial step of enzymatic glucosinolate hydrolysis.

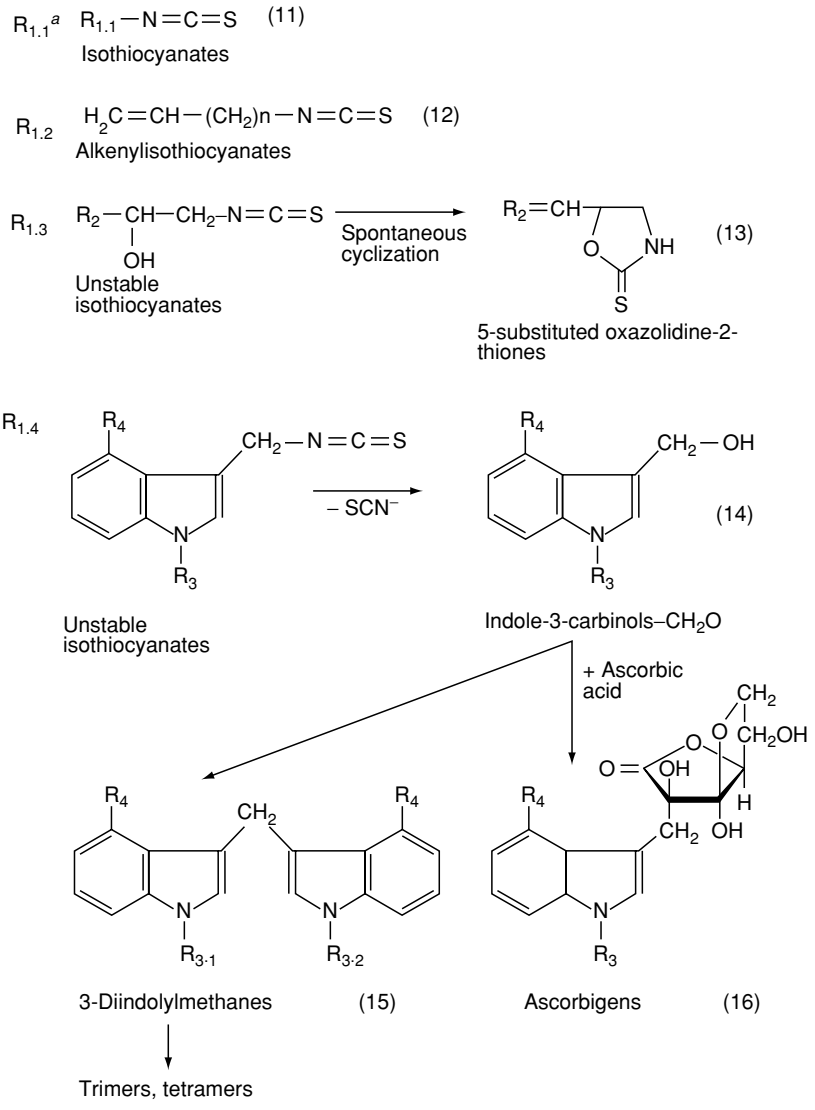

Figure 3 Glucosinolate breakdown products – Isothiocyanates (2.1, Figure 2). [a]Refers to Figure 1.

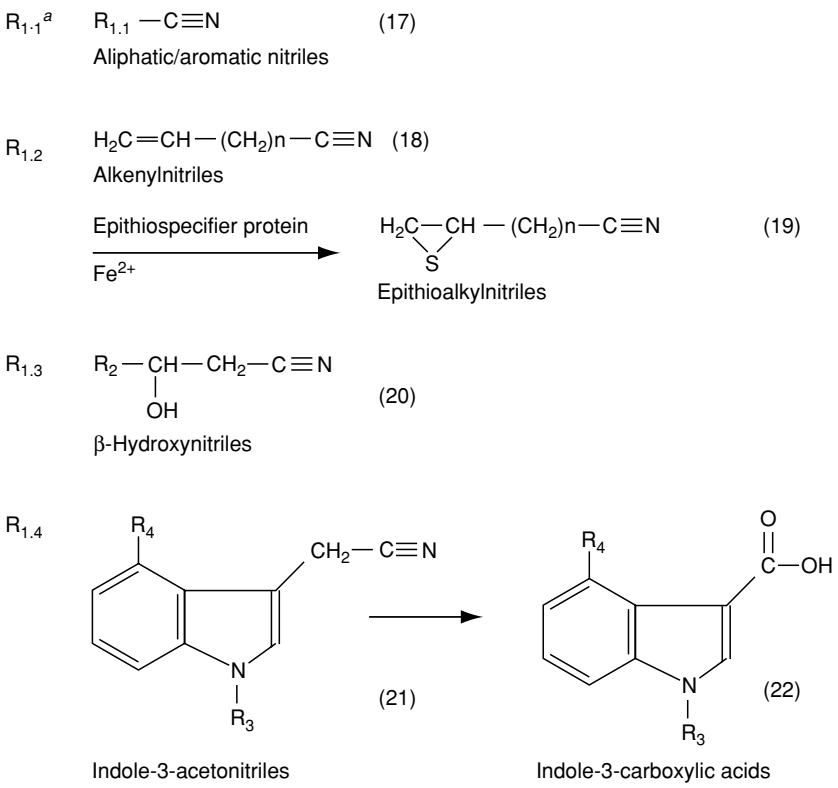

Figure 4 Glucosinolate breakdown products – nitriles (2.2, Figure 2).[a] Refers to Figure 1.

with myrosinase but is not present in all glucosinolate-containing species. In the absence of ESP the addition of ferrous ions to reaction mixtures promotes nitrile formation (18).

Reaction Conditions

Most factors modifying glucosinolate hydrolysis affect either myrosinase activity and specificity, or the activity of ESP, which is a very labile protein, in marked contrast to myrosinase. Low concentrations of ascorbic acid increase myrosinase activity in some species, whilst high concentrations are inhibitory. The broad range of possible degradation pathways and their resulting complexity, as well as the diversity of hydrolysis products, are most likely due to their function in plant defense mechanisms. Hydrolysis products, particularly thiocyanate ion and isothiocyanates, are toxic to animals, fungi, and microbes as well as to other plants. Catabolic detoxification is achieved by transformation to a variety of simple compounds that are either recycled or excreted as volatiles such as methylisothiocyanate, methylthiocyanate, methylthiol, dimethyldisulfide, and disulfide. Plant cell damage, accompanied by competing processes of hydrolysis and *de novo* biosynthesis of specific glucosinolates, especially indole glucosinolates, occurs during storage and processing of cruciferous vegetables. Chopping, cooking, freezing, and fermenting lead to enzymatic and/or nonenzymatic glucosinolate hydrolysis and *de novo* glucosinolate biosynthesis as a wound response. Leaching into cooking liquor is a major cause of glucosinolate loss during processing. There are thus prominent changes in glucosinolate profiles and in the quality and quantity of hydrolysis products formed (and hence flavor characteristics) during processing.

Predicting the degree of glucosinolate loss, and the nature of hydrolysis products formed, is difficult because of the high number of parameters which affect the process; these include the final size of the plant material after chopping, the amount of cooking water, the use of cold or boiling water at the start of the cooking process, the nature of the processing/cooking, cooking time, and precooking/storage conditions (frozen material will be more heat-sensitive than fresh). There are attempts to investigate systematically the effects of single and combined processing parameters on glucosinolate hydrolysis; by modulating these it may eventually be possible to predict the outcome of a given process and identify the major parameters affecting changes in glucosinolates.

Analysis

The type of glucosinolate analysis depends on the particular data required, i.e., overall (total) concentration, amount of individual glucosinolates, quantity and type of degradation products and/or metabolites. For screening purposes in plant breeding and/or the study of environmental influences, the easiest, fastest, and cheapest method to determine total glucosinolate concentrations is by glucose release. Myrosinase-mediated hydrolysis of glucosinolates releases glucose and sulfate in equimolar amounts. Both can be used for total glucosinolate determination but glucose, which can be measured instrumentally with a high sample throughput and low cost, is the analyte of choice if large numbers of samples need to be screened.

The chemical and biological properties of individual glucosinolates and their products can be very different, which is why determination of the levels of individual glucosinolates is often the goal. Intact glucosinolates can be separated by high-performance liquid chromatography (HPLC) on reverse-phase matrices using ion pair reagents to neutralize the strongly ionic sulfate group. Desulfation has been used to circumvent the need for ion pair reagents and also represents a useful – indeed, frequently necessary – clean-up step. Such a method has been declared as a European Union (EU) reference method. Whilst lack of standards makes identification by ultraviolet detection at 227 nm or using diode array detectors difficult, novel coupling techniques such as HPLC – mass spectrometry (MS) or HPLC–MS nuclear magnetic resonance (NMR) have facilitated glucosinolate separation and identification without derivatization.

HPLC–MS is also the method of choice for analyzing glucosinolate hydrolysis products and metabolites; these, together with the parent glucosinolates, can be separated on reverse-phase columns. Since high-performance MS instruments and techniques allow the identification and quantification of compounds incompletely separated, it is now possible to analyze complex mixtures.

Extracts of plant material, biological fluids, and cell culture media or biological tissues are all very complex; depending on the subsequent type of analysis, time- and labor-intensive extraction and purification procedures may be necessary. The most common extraction procedure for glucosinolates uses hot methanol followed by separation on strong ion exchange materials. Novel, liquid chromatography (LC) integrated preparation methods save time and can improve results through on-column analyte enrichment; quantitative and matrix-independent recovery; increased analytical capacity; and low throughput cost per sample.

Bioactivity

While glucosinolates *per se* are biologically inert, their hydrolysis products possess chemical properties rendering them powerful allelochemical agents against microbes, fungi, and plants. They also contribute significantly to the typical flavor of cruciferous vegetables; however, considerable current interest centers on their biological effects on humans and animals, which are highly dose- and structure-dependent, and may elicit positive, protective effects as well as antinutritional and potentially toxic responses.

Cancer-Preventive Potential

The inverse correlation between *Brassica* consumption and cancer risk is well established through many epidemiological studies. However, only limited human intervention studies have been conducted (e.g., with Brussels sprouts) to confirm these results. Mechanistic investigations, largely based on cell culture studies, have demonstrated glucosinolate hydrolysis products to be a source of the anticarcinogenic effects. Thus, indole-3-carbinol derived from 3-indolylmethylglucosinolate modulates the estrogen hydroxylation pathway to produce a less potent form of estradiol and, thereby, might confer protection against estrogen-related cancer.

Individual glucosinolate-derived isothiocyanates, e.g., phenethyl- and 4-methylsulfinylbutyl isothiocyanate (this latter commonly termed sulforaphane), have been shown to prevent formation of chemically induced tumors of the esophagus, stomach, liver, lung, and mammary gland in animals, most likely through mechanisms involving inhibition of cytochrome P-450 carcinogen activation and induction of phase II enzymes such as glutathione-*S*-transferase and NAD(P)H:quinone reductase. Glucosinolate-derived isothiocyanates may alter the balance of phase I/II enzyme metabolism such that phase I oxidative activation of carcinogens is decreased as compared with the increased rate of phase II detoxification.

Several isothiocyanates may also induce cell cycle arrest and apoptosis in colon cancer cells, which represents an alternative mechanism of anticarcinogenic activity. Thus, individual isothiocyanates belong to a small, but important group of compounds exhibiting dual inhibitory action against carcinogenesis. It is not surprising, therefore, that glucosinolate-derived compounds have been identified as targets for functional food or nutraceutical product

development. However, concentration-dependent adverse effects of many glucosinolate hydrolysis products must be considered: dose–effect relationships are not clinically proven and tissue bio-availability remains unknown. It is also unclear if, to what extent, and in which chemical form, they are absorbed, distributed, and metabolized, and also whether active metabolites reach possible target organs. Given these factors, and the absence of information derived from preclinical studies or studies involving animal pharmacology and toxicology, much work is needed before any overall recommendations can be offered regarding their likely beneficial effects on human health.

Antinutritional Effects

In contrast to the above putative beneficial properties, many glucosinolate-derived hydrolysis products are considered as natural toxicants, thereby being the focus of breeding and processing strategies to reduce/remove them from animal feeding stuffs. These compounds elicit a variety of adverse effects on the morphology and function of different cells and organs (*See* **Goitrogens and Antithyroid Compounds**) and antioxidant status. Liver, kidney, thyroid gland, and pancreas are the main target organs; in rats, toxic effects were observed with daily doses of 10–50 mg kg^{-1} body weight. In high concentration, certain isothiocyanates and nitriles may initiate mutagenic, cytotoxic, and carcinogenic processes; thus untreated juice of certain cruciferous vegetables can induce genetic mutations in both bacterial and mammalian systems and subsequent analysis implicated glucosinolate breakdown products as the primary source of mutagenesis. Differences in genotoxicity amongst varieties of cruciferous vegetables have been interpreted on the basis of the nature and amount of individual glucosinolates. Cytotoxicity and genotoxicity studies using, for example, human colon adenocarcinoma HT29 cells, have demonstrated the dose-dependent cytotoxicity of individual glucosinolate hydrolysis products (diindolylmethane (DIM) and sulforaphane) at $< 1.0\,\mu g\,ml^{-1}$; these products were, moreover, able to inhibit quiescent cells from reentering the cell cycle. Although very effective on undifferentiated HT29 cells, DIM and sulforaphane do not appear to affect the viability of differentiated intestinal cells – a significant finding given that cancer cells are undifferentiated.

Bioavailability

The biological effects of glucosinolate hydrolysis products depend greatly on their bioavailability, defined not only by oral dose, but also by their absorption, distribution, metabolism, and excretion (ADME) that, together, offer a measure of tissue bioavailability. Thus, phenylhexylisothiocyanate (a synthetic product) enhances esophageal cancer because of its extreme bioavailability (the highly lipophilic phenylhexenyl group increasing membrane permeability and, hence, absorption) whereas lower homologs decrease esophageal cancer (phenylpropylisothiocyanate > *phenylethylisothiocyanate > phenylbutylisothiocyanate > *benzylisothiocyanate: the asterisk denotes that it is naturally occurring) since reduced bioavailability leads to optimal tissue levels. The broad opinion that even low concentrations of nitriles, arising from glucosinolate breakdown, may exert toxic effects needs to be reconsidered, since their individual chemical structures, and subsequent bioavailabilities, will be key parameters in determining their mode of action. Thus, 1-cyano-2-hydroxy-3-butene may be the most active anticarcinogen in Brussels sprouts, acting via induction of phase II enzymes.

Although chewing will liberate intact glucosinolates, it also causes enzymatic hydrolysis; the extent of these processes has not been investigated. The same is true for the effect of gastric digestion, including the binding of intact glucosinolates/hydrolysis products to peptides, smaller glycoproteins, and other food components. A substantial proportion of the intact glucosinolates in the food may, therefore, not be absorbed in the small intestine and will reach the colon, to be hydrolyzed by endogenous microflora; they might, therefore, have potential for delivering bioactive hydrolysis products directly to the colon and so be effective against colon tumor development. In this context, the precise role of microbial thioglycosidases is controversial and needs further investigation.

Even if absorption is high, bioavailability may be limited by rapid and extensive metabolism. The major form of first-pass metabolism of isothiocyanates is conjugation with glutathione (catalyzed by glutathione-*S*-transferases), but the site of this conjugation is unclear. Little is known about the metabolism of nitriles but structural features appear to influence the formation of cyanide and thiocyanate ions.

Lack of fecal recovery of intact glucosinolates indicates that extensive metabolism is likely and numerous animal studies have shown that, following first-pass metabolism, extensive transformation of glucosinolates and hydrolysis products occurs. The site of metabolism and the nature of intermediates remain unclear. Most studies have been unable to study metabolism quantitatively and 25–65% of metabolites are unidentified. Depending on the animal species, the major urinary metabolites are *N*-acetyl cysteine conjugates (mercapturic acids), thiocyanate

ion, and hippuric acid conjugates. Attempts to identify primary compounds or metabolites in blood failed but the feeding of radiolabeled compounds demonstrated a rapidly occurring peak of radioactivity in the blood that was both species-specific and of short half-life.

There is only very limited knowledge on the metabolism of nitriles; some, e.g., benzyl cyanide, release hydrogen cyanide (HCN) following the action of a microsomal mixed-function oxidase, probably as a result of α-methylene hydroxylation to form an unstable cyanohydrin. The liberated HCN can be converted to less toxic SCN^- by sulfur transferase. The individual structure of the organonitrile affects SCN^- excretion.

The short-term biological activity of phytochemicals (including glucosinolate hydrolysis products) is relatively weak; they have multiple targets and may elicit both beneficial and adverse effects. This poses particular problems in determining the net effect because plant foods contain a complex mixture of phytochemicals with a wide range of physiological effects. One key question to address is: how much does a particular food component contribute to a specific health outcome compared with other dietary constituents? Another is: to what extent does this affect the disease outcome? The availability of a robust and reliable surrogate biomarker should provide answers by facilitating measurement of exposure and by indicating a positive/negative physiological effect prior to the final outcome.

Future Research

The need for future research has been identified in the individual sections above. After a century of glucosinolate study, research into their biological activity and physiological significance remain in their infancy. Determination of their ADME and human studies, following on from animal and cell culture experiments, should be a prime focus of future research. As yet, there are no validated biomarkers of exposure, effect, and susceptibility to dietary glucosinolates that would enable correlation of dose and effect in human intervention studies. Since presystemic metabolism affects tissue bioavailability, knowledge of how individual enzyme status and polymorphism in human glucosinolate-metabolizing enzymes affect ADME is vital. Current data are both insufficient and controversial. Insufficient attention has been paid to the effect of the food matrix, since food contains a range of bioactive compounds, nutrients and non-nutrients, and detailed investigation of their interactions, competition, and synergism with one another and with proteins and carbohydrates will be necessary if an understanding of overall sensory and physiological characteristics is to be achieved.

A second focus for research is elucidation of the complex parameter affecting levels of glucosinolates and hydrolysis products in food. If the biological effects of glucosinolates are to be exploited for social and economic benefit through the breeding or genetic engineering of nutritionally enhanced crops, further information on their biosynthesis, on the genes involved, and on their function and possible linkages is needed. Details of the myrosinase-glucosinolate system is required and factors responsible for directing hydrolysis toward specific products understood; recommendations for processing cruciferous vegetables to optimize glucosinolate hydrolysis products may then follow.

Finally, understanding of the antimicrobial, insecticidal, and fungicidal properties of glucosinolate hydrolysis products will assist targeting of future breeding programs, facilitate integrated crop protection, and provide natural options for plant husbandry in response to consumer concerns.

See also: **Cancer**: Diet in Cancer Prevention; **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; **Chromatography**: High-performance Liquid Chromatography; **Functional Foods**; **Goitrogens and Antithyroid Compounds**

Further Reading

European Commission (1990) *Official Journal of the European Community* no. L170/33. Community regulation (ECC no. 1864/90), pp. 27–34.

Fahey JW and Talalay P (1999) Antioxidant functions of sulphoraphane: a potent inducer of phase II detoxication enzymes. *Food Chemistry and Toxicology* 37: 973–979.

Fahey JW, Zalcmann AT and Talalay P (2000) The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* 56: 5–51.

Fenwick GR, Heaney RK and Mawson R (1989) Glucosinolates. In: Cheeke RR (ed.) *Toxicants of Plant Origin*. Vol. II. Glycosides. pp. 1–40. Boca Raton, FL: CRC Press.

Heaney RK and Fenwick GR (1993) Methods for glucosinolate analysis. *Methods in Plant Biochemistry* 8: 531–550.

Hecht SS (1999) Chemoprevention of cancer by isothiocyanates, modifiers of carcinogen metabolism. *Journal of Nutrition* 129: 768S–774S.

Kelloff GJ, Crowell JA, Steele VE *et al.* (2000) Progress in cancer chemoprevention: development of diet-derived chemopreventive agents. *Journal of Nutrition* 130: 467–471.

Mithen RF, Dekker M, Verkerk R, Rabot S and Johnson IT (2000) The nutritional significance, biosynthesis and

bioavailability of glucosinolates in human foods. *Journal of the Science of Food and Agriculture* 80: 967–984.
Rosa EAS, Heaney RK, Fenwick GR and Portas CAM (1997) Glucosinolates in crop plants. *Horticultural Reviews* 19: 99–215.
Steinmetz KA and Potter JD (1996) Vegetables, fruit, and cancer prevention: a review. *Journal of the American Dietary Association* 96: 1027–1039.
Tawfiq N, Heaney RK, Plumb JA *et al*. (1995) Dietary glucosinolates as blocking agents against carcinogenesis: glucosinolate breakdown products assessed by induction of quinone reductase activity in murine hepa1c1c7 cells. *Carcinogenesis* 16: 1191–1194.
Verhoeven DTH, Verhagen H, Goldbohm RA, Vandenbrandt PA and VanPoppel G (1997) A review of mechanisms underlying anticarcinogenicity by brassica vegetables. *Chemico-Biological Interactions* 103: 79–129.
Wallsgrove RM, Doughty K and Bennett RN (1999) Glucosinolates. In: Singh BK (ed.) *Plant Amino Acids*, pp. 523–561. New York: Marcel Dekker.

Gluten *See* **Wheat**: The Crop; Grain Structure of Wheat and Wheat-based Products; **Celiac (Coeliac) Disease**

GLYCOGEN

M H M Rocha Leão, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

Introduction

Glycogen is a glucose polysaccharide occurring in most mammalian and nonmammalian cells, in microorganisms, and even in some plants. It is an important and quickly mobilized source of stored glucose. In vertebrates it is stored mainly in the liver as a reserve of glucose for other tissues. In hepatocyte cells it is accumulated and mobilized according to blood glucose availability and to extrahepatic cells. Glycogen is also stored in muscles and fat cells. In the muscle it seems to be mainly used for energy purposes as metabolic fuel for glucolysis producing glucose 6-phosphate. Thus, glycogen plays a crucial role as a systemic and cellular energy source and also as an energy store. A great number of enzymes and hormones control the synthesis and degradation of glycogen. Consequently, stores of human body glycogen may vary dramatically due to diet, exercise, and stress.

Structure, Isolation, and Location

The glycogen molecule is formed by (1→4) linked chains of alpha-D-glucose residues and branched at intervals of 8–12 glucose residue units, with (1→6)-alpha-D-linkages which produce a bush-like structure, called beta-particles. Dimers and trimers of beta-particles are linked upon a single protein backbone known as glycogenin. The backbones of several fundamental units are joined to form a macromolecule named an alpha-particle, of 100–150 nm diameter. Due to its polymeric nature, glycogen is osmoticaly inactive and can be stored in large quantities within the cells. The branched glycogen has two advantages: it increases its solubility and facilitates the rapid turnover of sugar residues. The latter is due to exhibition of a greater number of nonreducing ends for simultaneous enzymatic reactions.

Studies on the cellular location of glycogen in several cell types have led to some different interpretations which are due to different procedures employed for glycogen extraction (hot potassium hydroxide (KOH), cold or hot tricarboxylic acid (TCA), hot sulfuric acid (H_2SO_4), and cold or boiling water) used in different laboratories. In yeast, it is accepted that there are two pools of glycogen location. One is water-insoluble or acid-extractable, occurring outside the cellular membrane. This pool corresponds to the major fraction of the yeast glycogen. The second is water-soluble or alkali-extractable, occurring inside the cellular membrane. In the past some authors showed the relationship between the location and their metabolic compartmentalization. However if the two pools have distinct roles in yeast metabolism this remains to be proven.

Recently, an original glycogen structure surrounding the finger-like plasma membrane invaginations in a ring was discovered and visualized by electron microscopy. In tissues, the glycogen is associated with 2–4 times its weight with proteins by adsorption

mechanism. These proteins appear to be mainly the enzymes of its own metabolism. Because glycogen is intimately associated with protein in the cell and the protein–glycogen complex forms an independent functional unit, it is considered as a dynamic organelle called glycosome without a membranous envelope and morphologically similar to ribosome.

Both types of supramolecular structure have protein particles and both are attached to endoplasmic reticulum membranes and cyto-skeleton components. Glycosomes may occur free in cytosol (lyoglicosomes) and are acid-labile, whereas the glycosomes associated with cellular structures (desmoglycosomes) are acid-resistant. Although glycogen structure has been studied after isolation using different mild extraction (cold water, cold TCA, or enzymes to lysis cell walls), or by direct observations by means of electron microscopy, the results from the literature are sometimes controversial and this matter is still debated. The different methodologies used by different groups of researchers have led to different results inducing sometimes different interpretations. Nowadays it is hypothesized that there are two forms of physiological glycogen which differ in the proportion of protein relative to carbohydrate and location in the cells. One form corresponds to protein-rich, and it is acid-insoluble, while another is protein-poor and is acid-soluble. The former is called desmoglycosomes and the latter is called lyoglycosomes. Progress for establishing the relationship between glycosomes (lyo- and desmo-) and glycogen metabolism is expected.

Function

Glycogen metabolism is strongly related to glucose metabolism despite the cell type or the organism where it is found. On the other hand, fats are the major source of energy in the human body but they are only mobilized during prolonged periods of starvation from adipose tissue by lipolysis. The fatty acids produced are degraded to acetyl Coenzyme A (CoA) in the liver via the beta-oxidation pathway. Under these conditions the amounts of acetyl CoA are very high and exceed the degradative capacity of the citric acid cycle. Thus, acetyl CoA is used as a substrate to form ketone bodies which are utilized by all body organs as an energy source under starvation. So, in the human liver, fatty acids are not transformed into blood glucose. If starvation is very prolonged, proteins are also used as an energy source.

Glycogen as an energy reserve stored in the liver is low when compared to fat. The body has more fat tissue and fats yield more energy than carbohydrates. When the energy supply is low the liver responds, mobilizing stored, glycogen and after 6–12 h it is exhausted. When glycogen is depleted there is an increase in the glyconeogenesis rate for glucose synthesis. Humans can store up to 150 g (a third of all glycogen in the body) glycogen in the liver which will serve to maintain the blood glucose level in the postresorptive phase. Therefore, in this organ it falls drastically and can decrease to almost zero during starvation lasting more one day. The buffering effect of liver glycogen to maintain blood glucose level is caused by glucose 6-phosphatase catalysis. Since this enzyme only exists in liver, glycogenolysis in muscles channel glucose 6-phosphate into glycolysis without any adenosine triphosphate (ATP) waste.

Muscle glycogen serves as an energy source during mechanical work. During light exercise glucose, fatty acids, and ketone bodies are completely degraded to CO_2 and H_2O. ATP and creatine phosphate present in resting muscles sustain the necessary energy to muscle contraction for a few seconds.

Glycogen is the most important energetic reserve in muscle tissue. It can form 2% of muscle mass. Glucose 1-phosphate, the first metabolite produced when glycogen is degraded, is processed to glucose 6-phosphate via isomerization. This metabolite yields ATP either via oxidative phosphorylation, when O_2 is present, or through conversion to lactate via anaerobic glucose. There are two types of muscle in animals: the white muscle of rapid contraction and the red muscle of slow contraction. The former works for short periods of time during heavy exercise with high ATP requirements via anaerobic glycolysis. The latter works for long periods of time during prolonged exercise and in continuous contraction, as in cardiac muscle. Red muscle requires low ATP per minute because it is rich in mitochondrial organelles and therefore converts glucose 6-phosphate more slowly to CO_2 and H_2O when compared with white muscle.

Knowledge of microbial nutrition has become more relevant in recent years with the increased use of microorganisms and their products as food additives and supplements. *Saccharomyces cerevisiae* (baker's yeast) is an important microorganism used in food technology which mobilizes glycogen after glucose is exhausted from the medium and in the budding process as carbon and energy sources. Recently, it was found that glycogen plays another important role when *S. cerevisiae* is submitted to stress conditions in the medium of glucose. The effect of stress is to stimulate the recycling of glycogen instead of promoting either its accumulation or its degradation. It is important to note that, under stress, cell growth is arrested, thus there are ATP imbalances in the upper and lower part of the glycolytic pathway. Like a

'turbo' design model, glucose oxidation by glycolysis leads to overproduction of ATP, glucose 6-phsphate, phosphate depletion, and cell death. The solution to this cellular problem is possible because stress induces transcriptional activation and the relative enzymes for glycogen metabolism are under this control. Moreover, as a result, both synthesis and degradation enzymes are expressed to the same extent. Thus, it has been suggested that glycogen turnover may function as a glycolytic safety valve to avoid substrate-accelerated death and also that glycogen in these cells confers survival advantages for cell proliferation and stress resistance. In this manner, the glycogen level or glycogen used as a futile cycle contributes to the fitness of the yeast in different environments and under industrial processes. In brewery fermentation yeast is added at the beginning of each batch for eight or more cycles. Before each cycle it is replaced in proliferation conditions and must break down its glycogen for sterol and fatty acids biosynthesis. Glycogen is the sole carbon source used in membrane component biosynthesis. Naturally, the glycogen content at the beginning of the brewery fermentation has a direct impact on yeast vitality (metabolic activity) and a significant improvement in brewer quality.

Biosynthesis and Degradation

Glycogen phosphorylase is a dimmer catalyzing the first and controlled step in glycogen degradation generating glucose 1-phosphate. The cyclic adenosine monophosphate (cAMP) cascade action yields the active form of glycogen phosphorylase a, which has a phosphoryl group linked to Ser14 in each subunit. The less active form, which is phosphorylase b, is dephosphorylated. Both forms are finely regulated. They contain the vitamin B_6 derivative pyridoxal-5-phosphate, covalently linked, which is required for their activities. The glucose 1-phosphate is converted by phosphoglucomutase to a glucose 1,6-bisphosphate intermediary which is an important enzyme regulator occurring *in vivo* at low concentration to maintain phosphoglucomutase fully active. To debranch the glycogen, two catalytic sites in debranching enzyme for both transferase and alpha (1→6)-glucosidase are required, increasing enzyme efficiency. Although the reaction catalyzed by glycogen phosphorylase is reversible, under physiological conditions glycogen breakdown is a thermodynamically favorable process. Consequently, glycogen biosynthesis and degradation must occur by two separate pathways. This strategy has two advantages: biosynthesis and degradation may occur at the same time and each pathway may be independently regulated. Since under physiological conditions the glycogen biosynthesis from glucose 1-phosphate to glycogen is an unfavorable process, the glycogen biosynthesis needs an exergonic step. So, glucose 1-phosphate reacts with uridine triphosphate (UTP) to form uridine diphosphate glucose (UDPG) via UDPG pyrophosphorylase catalysis. This reaction produces pyrophosphate (PPi) which is hydrolyzed in a thermodynamically highly favorable reaction catalyzed by pyrophosphatase. Both free energy of PPi and free energy of UTP are utilized to drive the glycogen biosynthesis pathway entropically. Glycogen synthase can only act on preexisting alpha (1→4)-linked glucose residues chain. So, the first step in glycogen biosynthesis is the attachment of glucose residue from UDPG to the Tyr group of an autocatalytic enzyme named glycogenin. Sequentially, the autocatalytical glycogenin extends the glucan chain by six or seven alpha-1,4-linked glucose residues, again derived from UDPG. Note that glycogenin is a glucosyl transferase enzyme and serves as an initiator for glycogen biosynthesis. The latter shows that glycogenin has a structural role in glycogen molecule and its biogenesis. Glycogen synthase associates with branching enzyme, elongating the glucan chain and building the glycogen molecule on to the protein backbone. The branching enzyme amylo-(1,4→1,6) transglycosilase breaks segments of at least 11 glucose residues of the linear chain (alpha-amylose), transferring to the C_6-OH groups to the same or another chain.

Allosteric Regulation

Glycogen phosohorylase and glycogen synthase are under finely tuned and reciprocal regulation. The activity of glycogen phosphorylase is allosterically controlled by metabolic effects produced in cellular metabolism. In muscle this enzyme is activated by adenosine monophosphate (AMP) and inhibited by ATP and glucose 6-phosphate. The liver isoform, unlike the muscle isoform, is more tightly controlled by phosphorylation than by allosteric regulation. In addition, in the liver, free glucose readly inhibits glycogen phosphorylase a. So, when cellular energy is high the enzyme is inhibited, and when cellular energy is low it is activated. Naturally, when ATP and glucose 6-phosphate concentrations are high, glycogen biosynthesis is favored. In terms of the symmetry model of allosterism, glycogen phosphorylase may assume the enzymatically inactive T (tense) conformation or the catalitically active R (relaxed) conformation which are differently responsive to allosteric effects. When in R-state, the enzyme has a high-affinity substrate phosphate-binding site, whereas in the T-state this catalytic enzyme site is

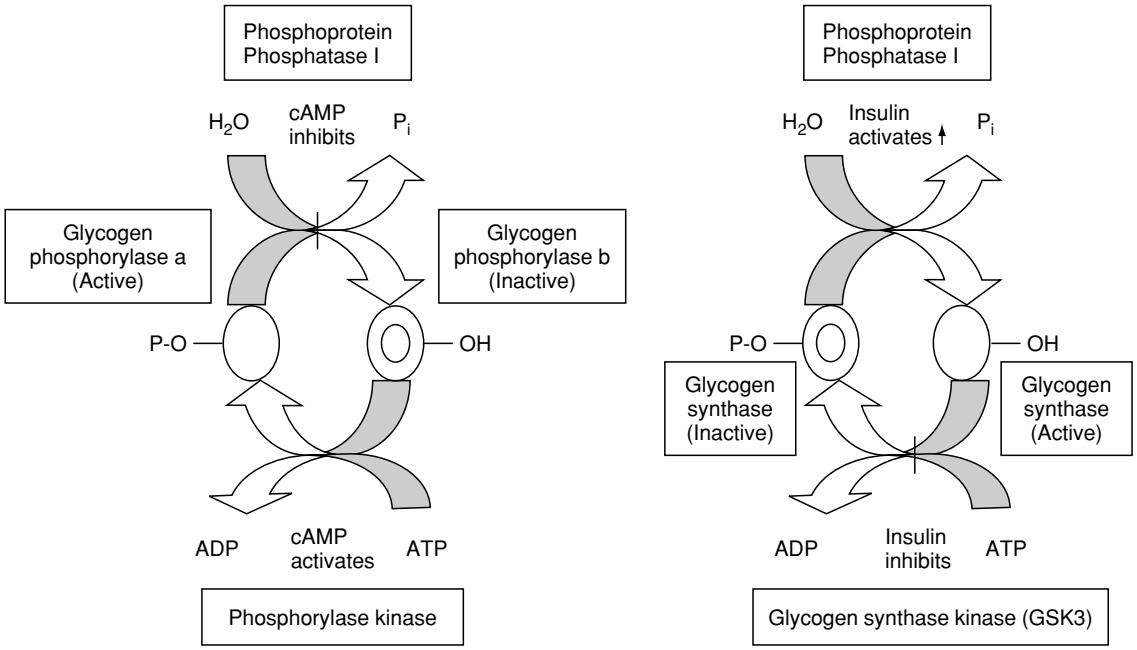

Figure 1 Controls of glycogen phosphorylase and glycogen synthase via enzyme phosphorylase.

inaccessible. Moreover, the R-state is favored by dephosphorylation and allosteric activators, while the T-state is favored by dephosphorylation and by inhibitors. Besides allosteric controls there are enzymatic interconversions through a bicyclic cascade involving three enzymes: phosphorylase kinase, cAMP protein kinase, and phosphoprotein phosphatase, which are themselves under allosteric control (Figure 1). The cascade functions in order to amplify the sensitivity of the system to an allosteric effect. So, a small alteration in cAMP concentration produces large modifications in the cascade action.

Glycogen synthase, a homotetrameric enzyme, contains 1–9 Ser residues to be phosphorylated. Thus, it can exist in several intermediate phosphorylated forms that have depressed ratios but have enhanced sensitivities to glucose 6-phosphate concentration. In muscles a less active form of glycogen synthase b is strongly inhibited by physiological concentrations of ATP, adenosine diphosphate (ADP), and Pi, its allosteric inhibitors, and due to low physiological concentration of glucose 6-phosphate, its allosteric activator, the enzyme b-form is almost totally inactive. In contrast, the more active enzyme a-form is independent of these effects. In this way, the glycogen synthase activity *in vivo* varies with the fraction of the enzyme in dephosphorylated form and the degree of the enzyme phosphorylation. The fraction of a-form is controlled by the same bicyclic cascade involved in glycogen phosphorylase through phosphorylase kinase and phosphoprotein phosphatase 1 enzymes. Moreover, the involvement of multiple protein kinases and protein phosphatases in glycogen synthase activity has been suggested. cAMP protein kinase, protein kinase C, calmodulin-dependent protein kinase, and glycogen synthase kinase-3 are present in this process. The phosphoprotein phosphatase 1 in muscle is only active when it is bound to glycogen via its G subunit. It is regulated by phosphorylation by insulin-stimulated protein kinase and by cAMP-protein kinase. Phosphatase 1 is also inhibited by its binding to the phosphoprotein phosphatase inhibitor 1, which also is modified by phosphorylation/dephosphorylation. Due to this fine control glycogen synthase has been considered to control the glycogen synthesis rate. In muscle it was demonstrated through theoretical quantitative model that glucose transport/hexokinase controls the glycogen synthesis rate and that the role of covalent phosphorylation of glycogen synthase is to adapt the activity of the enzyme to the flux and to control the metabolite level, not the flux. Phosphoprotein phosphatase 1 in the liver can be bound to R and T states of glycogen phosphorylase but only in the T-state enzyme is it accessible for hydrolysis by phosphoprotein phosphatase 1, which converts glycogen phosphorylase a to glycogen phosphorylase b. Thus, phosphoprotein phosphatase 1 results in glycogen phosphorylase inactivation and glycogen synthase activation.

In yeast, two distinct genes *GSY1* and *GSY2* encode two distinct glycogen synthases. The former

is expressed constitutively and the latter is expressed under catabolite depression, heat shock, and nitrogen starvation. The gene product (Gsy2) seems to be the predominant glycogen synthase. Gsy1 and Gsy2 isoforms occur in phosphorylated (less active) and in dephosphorylated (active) forms and have activities also subjected to allosteric control. Glucose 6-phosphate is a direct activator and also controls the phosphorylation state of this enzyme by inhibiting a glycogen synthase kinase. The cAMP-protein kinase and a (Ser/Thr) protein kinase named Snf1 also appear to play a central role in the control of glycogen metabolism in yeast. Using depressed mutant or less repressed sugars it was demonstrated that glycogen accumulation occurs associated to cell growth under both oxidative metabolism and low specific growth rate. Lipase–glycogen interactions have been observed in *Yarrowia lipolytica* lipase *in vitro* as well as during lipase production process. *Y. lipolytica* under stress condition shows a relationship between the lipase liberation rate and glycogen recycling followed decline.

Hormonal Control

In the past, research on glycogen metabolism led to general important discoveries such as the concept of second messenger. The great number of these molecules synthesized after hormone–receptor complex binding in cellular membranes amplifies the hormone signal in target cells. Glycogen metabolism regulation in the body is adjusted with meals at periodic intervals which can vary from short to longer periods. After a meal, blood glucose may increase from 90 to 140 mg dl^{-1} in an hour but reverts to normal in a nondiabetic person in 2 h due to the buffering effect of the liver. Although the general mechanism of glycogen biosynthesis and breakdown is the same in all tissues, the rates of these important pathway are differently regulated by chemical signaling to hormone release into the blood by gland cells. Then, from the blood to target tissues, insulin, epinephrine (adrenaline), and glucagon start the signal transduction in response to available blood glucose and other stimulus such as fight or flight. Under epinephrine and glucagon (Figure 2) a fine cellular control of the signal amplification is mediated by G-protein followed by adenylate cyclase activation, which catalyzes the synthesis of cAMP from ATP. Then, the signal is transduced via phosphorylation/dephosphorylation cascades adjusting glycogen biosynthesis and degradation to several physiological states. The rapid changes in the cAMP concentration are also under phosphodiesterase control, promoting its hydrolysis to AMP. On the other hand, insulin collects a protein family called Ser/Thr kinases in its target cells to advance its anabolic and anticatabolic programs. In the presence of insulin glycogen synthase kinase 3 is inactivated via protein kinase B, thus preventing the latter inhibiting glycogen synthase. Simultaneously, insulin activates protein phosphatase 1, thus activating glycogen synthase. Liver uptake of glucose across hepatocyte plasma membrane by transporter systems is outside the control of insulin. An increase in glycogen synthesis caused by insulin induction of the glycogen synthase in response to blood glucose level also activates the hexose monophosphate shunt, producing NADPH in fatty acid synthesis. The insulin/glucagon rate increase accelerates glycolysis by induction and activation of the glycolytic key enzymes. Besides blood glucose level, the protein and lipids supplied in food producing amino acids and fatty acids in the digestive system place the liver under anabolic states during the absorptive state. In contrast, in the postresorptive state the insulin/glucagon rate decreases causing a direct effect on gluconeogenesis activation as well as on glycogen degradation and on glycolysis deceleration, conducing the liver to enter in to a catabolic state. Glucagon induces the gluconeogenesis enzymes glucose 6-phosphatase, fructose 6-phosphase, fructose 1,6 bisphosphatase, and pyruvate carboxylase and represses the glycolytic enzyme pyruvate kinase, whereas insulin induces the glycolytic enzymes, phosphofructokinase 1, pyruvate kinase, and hexokinase. Glucagon also acts on the glycogen phosphorylation/dephosphorylation enzyme cascades via cAMP control. A similar glucagon effect is found in pyruvate kinase deactivation. Glucose 6-phosphate, ATP, AMP, acetyl CoA and fructose 2,6-bisphosphatase regulate both glycogen biosynthesis and glycogen degradation. The latter is synthesized from fructose 6-phosphate by nonphosphorylated phosphofructokinase 2 under cAMP protein kinase control. In its phosphorylated form it removes phosphate from fructose 2,6 bisphosphate, producing fructose 6-phosphate. Fructose 2,6-bisphosphate is the major glycolysis and gluconeogenesis regulator. It occurs at low cellular concentrations, activating phosphofructokinase 1 and inhibiting fructose 2,6-bisphosphatase. In this way it regulates biosynthesis and the degradation glycogen pathway. In muscle and fat tissues glucose uptake is regulated by insulin and similarly in the liver this hormone activates both glycogen biosynthesis and glycolysis by activation/inhibition, as well as induction/repression of the key enzymes of carbohydrate metabolism. On the other hand, in muscle epinephrine but not glucagon released in fight-or-flight situations activates glycogen degradation and inhibits glycogen biosynthesis at the same time. The

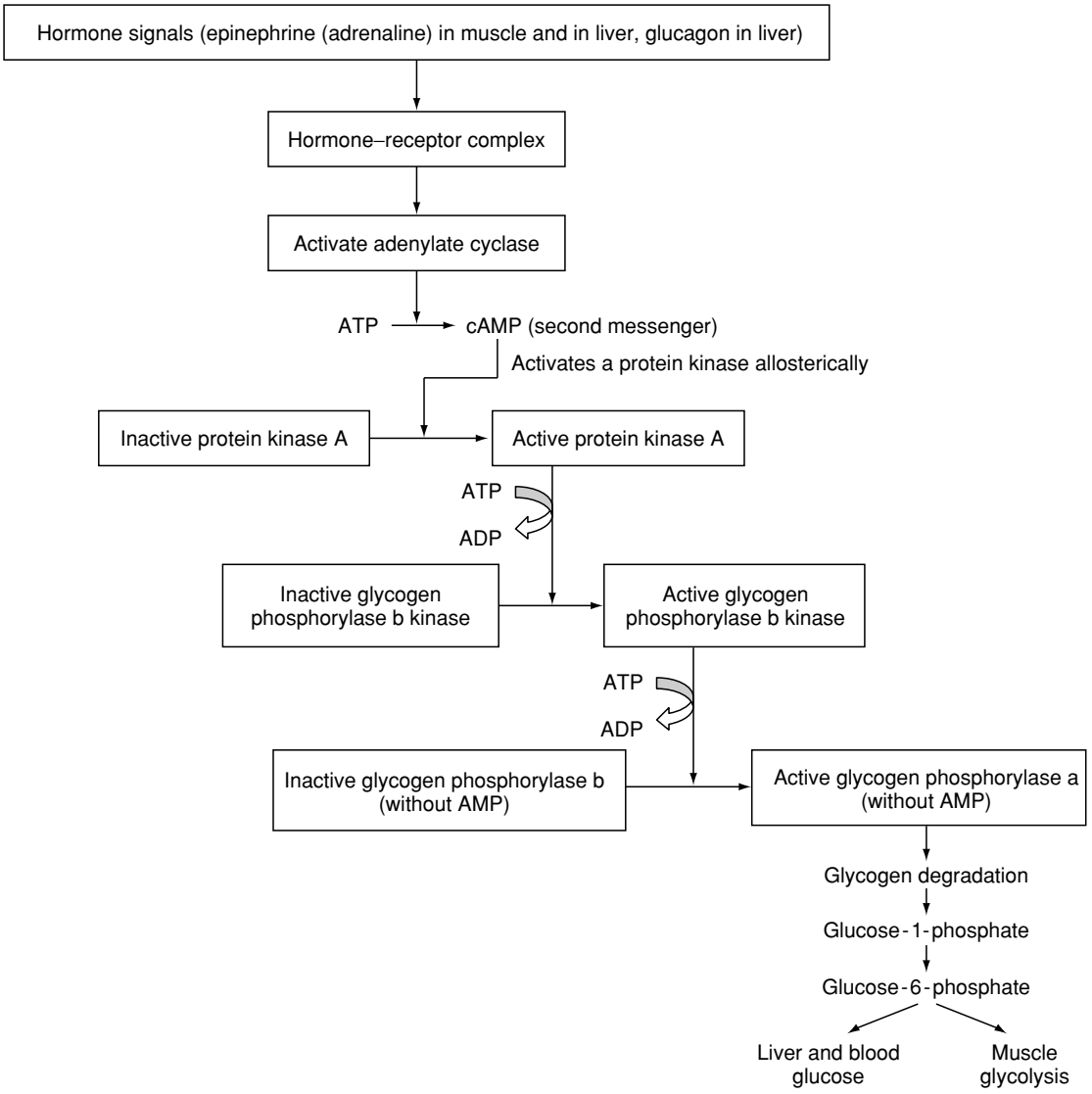

Figure 2 Glycogen degradation via hormone signaling.

activation of glycogen phosphorylase by epinephrine is preceded by cyclic cascade via cAMP protein kinase action, after signaling epinephrine–receptor complex. Ca^{2+} ions are released from the sarcoplasmic reticulum to cytoplasm, after nervous signal, during muscle contraction. Under emergency degradation of glycogen (fight-or-flight), Ca^{2+} ions sustain glycogen phosphorylase in full activity via subunit-calmodulin activation. This process is cAMP-independent. But, under normal muscle contraction, Ca^{2+} ions partly activate phosphorylase kinase which gives partial activation of glycogen phosphorylase.

Glycogenolysis and Physical Activity

The glycogen content stored in the human muscle before exercising is related to the state of preservation. On a high-carbohydrate diet, muscle glycogen may rise to $25\,g\,kg^{-1}$ of muscle. If carbohydrate intake occurs after glycogen is depleted by drastic exercise it may rise to $30\,g\,kg^{-1}$ and in dietary and exercise programs it may reach $40\,g\,kg^{-1}$. This phenomenon is called supercompensation and it is used in baker's yeast production in order to increase the glycogen content of the commercial product. So, baker's yeast is submitted to periods of high-carbohydrate consumption followed by fasting periods.

In athletes either short- or long-duration exercise determines if the glycogen-lactic acid system can provide energy or if an aerobic system is necessary to generate energy. This mechanism may be due both to the lactic acid disappearing rate in the liver for glucose conversion via gluconeogenesis, and the oxygen-replenishing rate in the muscle. Obviously,

in long-duration exercise aerobic glycogen degradation and oxidative phosphorylation are the most important pathways, although fatty acids may also sustain the working muscle. When glycogen is low during exercise an expressive amount of energy comes from fat degradation. Moreover, the muscle macroglycogen (MG) pool, which is a protein-poor glycogen (acid-soluble), as well as the proglycogen (PG) pool, which is a protein-rich glycogen (acid-insoluble) are differently affected by metabolic changes, by diet manipulation, and by glycogen-depleting exercises, respectively. The PG pool accumulates more easily after drastic exercise and the supercompensation phenomenon appears to be related to an increase in MG. The PG pool is less resistant to mobilization and its catabolism is inhibited less rapidly than the MG pool during intensive exercise. Glycogen content and glycogenolysis are positively related, thus glycogen amounts in the muscle before exercise is a regulator of glycogenolysis rate. All these facts suggest that the metabolic regulation of the two pools must be different and that the MG pool is a more stable form than the PG pool. On the other hand, the greater fraction of human muscle MG than PG is utilized in a marathon and accumulation of MG is delayed after this strenuous physical exercise, particularly in type 1 fibers, indicating local control for the glycogen biosynthesis pattern.

Glycogen Storage Disease

Due to the high number of enzymes and isoenzymes involved in glycogen metabolism and because gene expression, enzyme activities, and enzyme-regulatory mechanisms vary between tissues, a large number of diseases are linked to storage and/or structure of the glycogen appearing in a specific tissue but not in another. All of these abnormalities are named glycogenoses or glycogen storage diseases. They are divided into liver and muscle diseases respectively and are classified numerically from I to VIII, including subclassifications. Liver defects manifest hepatomegaly, hypoglycemia and other metabolic disorders, whereas muscle defects manifest cramps, pain, inability to increase the blood lactate level, and exercise intolerance. We must keep in mind that liver glycogen metabolism but not muscle glycogen metabolism is dramatically linked to glucose homeostasis of the human body. Hence, glycogen-liver diseases play a crucial role in human health and survival. Besides glycogen storage diseases, mutations in genes involved in insulin biosynthesis and action affect glycogen-liver metabolism. Diabetes is a chronic metabolic disorder afflicting millions of persons. Type 2 diabetes or noninsulin-dependent diabetes is a less severe form and is a polygenic disease characterized by multiple defects in insulin actions. Patients with type 2 diabetes have excessive hepatic glucose production which contributes to diabetic hyperglycemia in postabsortive state and in postprandial states. Compounds with action antiglycogenolytic that decrease blood glucose *in vivo* such as CP-91149, an allosteric inhibitor of the glycogen phosphorylase which is 200-fold more potent in direct comparison with caffeine, have shown that liver phosphorylase is a pharmacological candidate target for controlling hyperglycemia in diabetes.

See also: **Diabetes Mellitus**: Etiology; Chemical Pathology; **Exercise**: Metabolic Requirements; **Liver**: Nutritional Management of Liver and Biliary Disorders

Further Reading

Aiston S, Hampson L, Gomez-Foix AM, Guinovart JJ and Agius L (2001) Hepatic glycogen synthesis is highly sensitive to phosphorylase activity. *Journal of Biological Chemistry* 276.

Asp S, Daugaard JR, Rohde T, Adamo K and Graham T (1999) Muscle glycogen accumulation after a marathon: roles of fiber type and pro- and macroglycogen. *Journal of Applied Physiology* 86(2): 474–478.

Cahill G, Walsh PK and Donnelly D (2000) Determination of yeast glycogen content by individual cell spectroscopy using image analysis. *Biotechnology Bioengeneering* 69 (3): 312–322.

Calder PC (1991) Glycogen structure and biogenesis. *International Journal of Biochemistry* 23(12): 1335–1352.

Coula RY B, Aigle M and Schaeffer J (2001) Evidence for glycogen associated with plasma mebrane invaginations as visualized by freeze-substitution and the Thiery reaction in *Saccharomyces cerevisiae*. *Journal of Electron Microscopy* 50(2): 133–137.

Elliott WH and Elliott DC (2001) *Biochemistry and Molecular Biology*, 2nd edn. New York: Oxford University Press.

François J and Parrou JL (2001) Reserve carbohydrates metabolism in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiology Reviews* 25: 125–145.

Graham TE, Adamo KB, Shearer J, Marchand I and Saltim B (2001) Pro- and macro glycogenolysis: relationship with exercise intensity and duration. *Journal of Applied Physiology* 90(3): 873–879.

Gunja-Smith Z, Patil NB and Smith EE (1977) Two pools of glycogen in. *Saccharomyces*. *Journal of Bacteriology* 130(2): 818–825.

Kawasse FM, Santos JTP, Pereira-Meirelles FV, Coelho MAS, Lippel Sant Anna G and Rocha-Leão MHM (2000) Lipase–glycogen interactions in *Yarrowia lipolytica* strain. *Electronic Journal of Biotechnology*.

Rybicka KK (1996) Glycosomes–the organelles of glycogen metabolism. *Tissue Cell* 28(3): 253–265.

Rocha-Leão MHM, Meireles DF and Costa-Carvalho VLA (1987) Stimulation of glycogen formation during growth

of non-repressed *Saccharomyces cerevisiae*. *FEMS Microbiology Letters* 44: 423–425.
Rossi-Alva JC, Encarnação APG, Pacheco DFC and Rocha-Leão MHM (2000) Glycogen recycle as prolonged stress response in *Saccharomyces cerevisiae*. CDRON(H73) of the XIII SINAFERM, Rio Janeiro, Brazil.
Teusink B, Walsh MC, van Dam K and Westerhott HV (1998) The danger of metabolic pathways with turbo design. *TIBS* 23: 162–169.
Thomas RE, Brodersen C, Carls MG, Babcock M and Rice SD (1999) Lack of physiological responses to hydrocarbon accumulation by *Mitilus trossulus* after 3–4 years chronic exposure to spilled Exxon Valdez crude oil in Prince William Sound. *Comparative Biochemistry and Physiology Part C: Pharmacology Toxicology and Endocrinology* 122(1): 153–163.
Voet D, Voet JG and Pratt CW (1999) *Biochemistry*, 2nd edn. New York: John Wiley.

GOAT

Contents
Meat
Milk

Meat

N H Casey, W A Niekerk and E C Webb, University of Pretoria, Pretoria, South Africa

Introduction

Goats' meat is an important nutrient source, particularly for people of developing regions, which are situated mainly in the tropics. These regions account for more than 90% of the world goat population of approximately 650 million. Asian, African, and Caribbean peoples in particular consume goat meat. An estimate of the gross distribution of the world goat population is 56% in Asia, 33% in Africa, and 7% in Central America and the Caribbean. Production in these regions is associated primarily with subsistence agriculture. Commercial farming for metropolitan markets is practiced to some degree in Asia, Africa, and in Central and South America.

Domesticated goats, descendants of the wild goat *Capra hircus*, are among the earliest species of livestock to have evolved with pastoralist humans. Although the distinction between wild and domestic forms is vague at best, it is accepted that the bezoar or wild goat originated in the hills of western Asia. Early farmers began herding these wild goats that were drawn to their fields of cultivated grains. The oldest reliable evidence of domestic goats is from a site more than 9000 years old in Iran. Herding of wild goats altered the environmental stresses, allowing animals to acquire the characteristics of domesticated stock. The many goat breeds throughout the world are the result of translocations of humans over the globe, the influences of physiographic and climatic conditions, and of husbandry men breeding and selecting livestock for specific characteristics.

Consumer Preference

Cultural and traditional backgrounds and the socioeconomic status of the community dictate consumer preference for goats' meat. Virtually no cultural or religious taboos restrict the eating of goat meat, with the result that goat meat is readily available to societies in which eating beef, pork, or other meat types is prohibited.

In parts of the world, notably the tropics, subtropics, and arid regions of mainly developing countries, the demand for goats' meat exceeds supplies. Consequently, goat is sold at premium prices, but is increasingly substituted by cheaper mutton.

Goats' meat is not favored in the west. The results of taste panels in which westerners have assessed the acceptability of goats' meat against beef or lamb should be viewed in this perspective. Of prime importance are consumers' own concepts of value, relative to other foods or consumer items. In an attempt to overcome prejudice, Angora goat meat was marketed in the USA under the brand name Chevon. Chevon has now become a recognized commercial term for goats' meat. Goat kid is a favourite delicacy and features as cabrito in the Spanish-speaking world,

simply as goat kid in English, as chevreau in French, and as boklam in South African Afrikaans.

Production

The value of the goat lies in its flesh, milk, hair, and skin. However, meat production has remained the primary reason for goat-keeping. Meat goats therefore constitute the major proportion of the world's goat population and comprise many different breeds. Of these, the Boer goat breed is regarded as exhibiting the most desirable meat production characteristics. A recent study in Botswana concluded that a small ruminant enterprise is profitable and economically viable. Small-scale ruminant enterprises could provide a return of 34% on capital invested. In Zimbabwe, goats are often the only dependable and perennial agricultural output. Many rural farmers now regard goats as an easy income-generating venture since very little investment is required.

Goats' meat is derived from different categories of goat. Of the meat goats, kids, does, and male castrates are the preferred commercial commodities. The main sources of meat from the milch breeds are male kids and older culls. In New Zealand and Australia, dairy-type feral goats are cropped for export. Mohair-producing Angora goats are slaughtered as culls, or marketed when the price of mohair is down. Other goat products, notably milk and cheese (e.g., feta, chèvre, and Gorgonzola), hair, mohair, and cashmere, have important influences on goat populations of specific regions and the related economic value of goats. Goat hides also yield fine leather, which is used for wineskins, water containers, rugs, and clothing of high quality.

Goats generally feed on the leaves and shoots of shrubs, which enables them to thrive in semiarid scrublands and mountains. Traditionally, goats in the developing regions are free-range or are herded communally with little management care, including veterinary care, nutrition, and breeding. Household/personal herds may contain as few as three breeding does, with an average of between five and eight in tropical and subtropical Africa. Commercial herds in Namibia and other regions of southern Africa may number several hundred breeding does in a ratio of one buck to 25 does.

Meat Production Characteristics

General characteristics of goats favoring meat production are:

1. Females are early-maturing, highly prolific, and have good fecundity and mothering ability.
2. An extended breeding season is usual among goats in the tropics, and this is particularly advantageous to meat production.
3. Foraging preferences of goats cause them to graze a wider spectrum of plants than other small stock, which accounts for their ability to thrive in adverse conditions.
4. Foraging may account for relatively low parasite infestation.
5. Goats exploit available feed resources selectively, consuming material with sufficient digestible organic matter at or exceeding their maintenance needs, and their small size enables goats to utilize tropical shrub and shrub pastures more efficiently than cattle.
6. Goats are generally well adapted to hot environments, tolerating the extremes of desert conditions and high-temperature/humidity conditions of the tropics, because of their small size, large surface area to body weight ratio, an ability to conserve water, limited subcutaneous fat cover, and the particular nature of their coats.

Breeding and Performance Testing

Goats can be bred for carcass and meat characteristics by exploiting quantitative and qualitative gene effects. Performance testing as a means of improving productivity of meat goats was introduced in South Africa in 1970. The scheme provides for the performance testing and selection of goats specifically for meat production, according to five phases:

1. characteristics of the dam, her milk production and preweaning growth rate of her progeny
2. postweaning growth rate of progeny at various ages
3. feed conversion efficiency and body weight of male progeny under standardized conditions at a central testing station
4. postweaning growth rate of male progeny under standardized conditions:
   a. on a farm under direction of the Department of Agriculture and
   b. at a central location of a cooperative institution, also under the auspices of the department
5. qualitative and quantitative evaluation of the progeny of a buck

Nutrition and Growth

Well-nourished livestock have a high tolerance to parasites and disease, exhibit good fecundity with low mortalities, grow well, and generally thrive. Nutrient requirements of goats are determined by their physiological state. Specific information can be

obtained from the recommendations of the US National Research Council on the nutrient requirements of goats. Growth rate is an important criterion for efficiency in commercial meat production and is determined by the animal's genetic growth potential, nutrition, and health. Generally, goats do not have a high growth rate, compared with lambs. West African dwarf goats have growth rates of 10–30 g day^{-1} on variable-quality tropical grasses. Poor-quality tropical grasses (e.g., *Gliricidia sepium*) supplemented with a legume (*Leucaena leucocephala*) would improve the growth rate up to 46 g day^{-1}; a concentrate supplement could improve growth further to 70 g day^{-1}. Goat kids, adequately supplemented with small quantities of essential nutrients, will perform equally well on feeds with a low digestibility compared to other animals on forages with higher digestibility. Under favorable conditions, Boer goats may gain weight at more than 200 g day^{-1}. When protein intake is increased above that required for maintenance and growth, puberty and fertility can be achieved at an earlier age in young growing male kids.

Fertility

The good fertility, mothering ability, and extended breeding season of does in the tropics and subtropics permits mating twice per year or three times per 2 years. Thus, the number of kids per doe can be increased dramatically. Twins are more the rule than the exception. Sexual activity of bucks is greatest from sunset to midday. An apparent decline in male libido in spring and summer needs to be taken into account when twice-yearly breeding is practiced.

Angora goats and goats originating from the tropics have limited breeding seasons. In general, small ruminants in Africa are year-round breeders and, as a result, lamb or kid all year round except where a specific management system so dictates. The techniques of caprine embryo transfer, embryo freezing, embryo splitting, sex determination, *in vitro*, and caprine nuclear transfer have all been successful.

Goat Management

Animal productivity is improved by applying even the most basic management practices: herding as opposed to free-ranging, penning at night, removing unproductive animals from the herd, and controlled breeding. More sophisticated management practices, including nutrition supplementation and veterinary care, although yielding high dividends, require additional expenses that may be beyond the reach of subsistence farmers. In these instances, subsidies and state services should be considered.

Carcass and Meat Quality Characteristics

Goats have carcass characteristics slightly different from those of sheep. Goat carcasses are generally lean in appearance, less compact than sheep, and of differing carcass proportions with less total tissue distributed to the hind leg than in sheep. Goats' meat is also different, being coarser-grained, with a detectable different flavor and aroma. Studies on chemical composition and meat quality have suggested that goat meat is not inferior to lamb, but may be darker red in color than lamb. Color of goat meat is important, especially in the case of Capretto carcasses, where flesh should be pale or pink.

Grading

Communicating the expected quality of meat to consumers in a reliable and effective way is important in merchandising meat. This is achieved by grouping carcasses within a series of like quality characteristics into distinct categories, taking into account:

1. physiological age differences due to young animals having more tender meat because of the greater solubility of the collagen in the connective tissue and older animals having a stronger flavor;
2. sex differences due to the coarser grain of males and the strong flavor of the buck's meat;
3. fat differences due to the quality and palatability fat imparts to the meat and the association between fat and meat yield;
4. conformational differences due to its commercial value in some markets and its predictive value, although limited, in determining yield.

Age has an influence on palatability, with the meat from older goats regarded as being juicier. Muscle color may be darker and fat color more yellow with increasing animal age.

Carcass quality generally refers to lean/fat content, conformation, and the extent of damage (bruising, abscesses, and parasite infection). Meat quality includes appearance, composition in terms of lean, fat, and connective tissue and, in commercially displayed cuts, bone content, pH, drip, palatability, nutritive values, ease of processing, and shelf-life.

Dressing Percentage

Carcass yield is an important production criterion and is expressed as the dressing percentage ([dressed carcass weight/live weight] $\times$ 100). However, since it expresses a ratio of live weight to carcass weight and many factors influence these fractions (e.g., alimentary tract size and fill, slaughtering procedures, fleece or skin mass, distribution of body fat, and secondary sex characteristics), the dressing percentage must be

interpreted carefully and comparisons made within species, within breed types, and within slaughtering procedures. The dressing percentage of goats varies between 40% in very young animals and 56% in entire mature males. In regions where body components (head, organs, intestine, and skin) are prized and regarded as part of the carcass, the dressing percentage ranges between 66 and 82%. The dressing percentage increases with age, mass, and fatness, and is a consequence of increasing metabolizable energy per kilogram dry matter (ME kg^{-1} DM) in the ration.

Muscle

Muscle-to-bone ratio Goat carcasses have a higher muscle-to-bone ratio than is apparent from their conformation. A greater carcass and leg length results in a less compact carcass, which may be interpreted erroneously as signifying poor muscling. In addition, goats have less total muscle distributed to the leg and more to the forequarter than sheep. This is illustrated by the muscle distribution of castrated male Boer goats and South African Mutton Merino wether lambs, respectively: forelimb 17.3 and 16.2%; neck 9.3 and 8.3%; ventral trunk 25.8 and 20.9%; dorsal trunk 19.3 and 20.6% and hindlimb 28.4 and 34.1%. (*See* **Meat**: Structure.)

pH Goat muscle contains both aerobic (red) and anaerobic (white) muscle fibers and undergoes the same postmortem biochemical changes as beef and mutton. The decline of goat muscle pH follows a pattern typical of red meat carcasses to stabilize at around pH 5.4. Variations occur due to differences between muscles, sexes, and premortem stress. Exhaustive premortem stress yields dark, firm, and dry meat with a high ultimate pH (pH > 6.0). Postmortem biochemical changes are associated with the loss of water-binding capacity as the pH reaches the isoelectric point of the muscle proteins, the onset of rigor mortis, and the release and activation of proteolytic enzymes, notably cathepsins, responsible for the ripening of meat.

Collagen Goat muscle fibers are thicker and the fiber bundles larger than those of sheep, giving goats' meat a characteristic coarser grain. The attributed toughness of goat has been ascribed to the marketing of mature animals, in which the collagen in the connective tissue has a decreased ability to gelatinize under the influence of heat and moisture. Several studies have indicated that goats' meat is inherently less tender than that of sheep. Muscles of male Boer goat kids have a higher collagen content with lower collagen solubility than those of male lambs of four sheep breeds. Lambs and mutton are more tender with less fibrous tissue residue and a more intense aroma than Angora and Boer goat meat. The species' flavor is also more pronounced. Goat breeds may also differ in their meat quality: meat of Angora goats has a more acceptable flavor and is more tender with less residue than meat of Boer goats, which could be explained by the lower collagen content and slightly better collagen solubility. Evaluation of collagen alone, however, would be insufficient for conclusions on differences in tenderness. Other factors may be involved, especially muscle fiber size, the type of matrix formed by collagen, and the state of muscle contraction.

Electrical stimulation Goat carcasses, having a poor insulating subcutaneous fat cover, are susceptible to muscle toughening through the effects of cold shortening. Cold shortening can be countered by electrically stimulating the carcasses immediately after slaughter. Electrical stimulation increases the rate of postmortem glycolysis, depleting the adenosine triphosphate (ATP) energy source for muscle contraction due to the anaerobic state. The residual contractile properties of muscle are reduced, rigor mortis is advanced, and the enzymes associated with the conditioning of meat are activated. Electrical stimulation of goat carcasses, as with beef, can improve meat tenderness and facilitate accelerated processing of carcasses by hot boning, preferably after a 2-h conditioning period, with no detrimental effect on total bacterial count, tenderness, or cooking loss. Advantages of hot boning are a reduced mass loss in chillers, less carcass chiller space required, and faster throughput and packaging of the meat. (*See* **Meat**: Slaughter.)

Nutrient value Muscle contains the most important quality parameters and nutrient value of meat. Goat muscle is highly nutritious with a high biological value. The percentage moisture in raw, lean tissue ranges between 74 and 76%. Protein, fat, and ash contents are, respectively, 20.6–22.3, 0.6–2.6, and 1.0–1.1%. The essential amino acids determined in the proteins (mg g^{-1}) are arginine (74), histidine (21), isoleucine (51), leucine (84), lysine (75), methionine (27), phenylalanine (35), threonine (48), tryptophan (15), and valine (54). (*See* **Amino Acids**: Metabolism.)

The most important minerals in the muscle and organs of male cross-bred goats are shown in Table 1. Red meat is an excellent source of iron: heme iron is 5–10% as available as nonheme iron. A value of around 2.1 mg g^{-1} reported for lean goat meat from Malaysia compares favorably with values for separable lean beef (2.72 mg g^{-1}), lamb (1.74 mg g^{-1}), and veal (1.11 mg g^{-1}). Refer to individual minerals.

The contents of the vitamins thiamin, riboflavin, and niacin in lean goats' meat compare well with

Table 1 Mean mineral concentrations (mg 100 g^{-1}) in muscle and selected organs of cross-bred goats

Mineral	*Muscle*	*Liver*	*Kidney*	*Heart*	*Spleen*	*Brain*
Ca	11	10.06	13.58	7.7	11.47	46.99
P	155.5	263.9	168.1	111.71	214.03	245.64
Mg	19.7	15.08	10.19	9.63	15.28	12.82
K	350	188.55	122.26	100.15	194.9	277.68
Na	64.48	58.18	148.68	38.52	59.38	136.92
Cu	0.30	8.28	0.52	0.53	0.41	0.40
Zn	3.51	2.99	2.61	1.41	2.19	1.40
Fe	4.37	7.82	9.78	4.40	34.79	3.07
Mn	0.087	0.66	0.19	0.098	0.159	0.122
Dry matter (%)	21.90	25.14	16.98	19.26	19.11	21.36

Table 2 Thiamin, riboflavin, and niacin contents (mg 100 g^{-1}) of goats' meat, lean beef, veal, and lamb

Species	*Thiamin*	*Riboflavin*	*Niacin*
Goat[a]	0.1	0.56	3.6
Beef[b]	0.082	0.218	3.6
Lamb[c]	0.088	0.234	5.33
Veal[d]	0.06	0.3	7.6

From: [a]Abdon I, Del Rosario IF and Olga LG (1980) *Food Composition Tables Recommended for Use in the Philippines. Handbook 1*, 5th edn. Manila: Food and Nutrition Research Institute, [b]USDA (1986) *Composition of Foods: Beef Products – Fresh Processed Prepared. USDA, Human Nutrition Information Service, Agricultural Handbook* No. 8–13. Washington, DC: US Government Printing Office, [c]Ono K, Berry B and Johnson HK *et al*. (1984) Nutrient composition of lamb of two age groups. *Journal of Food Science* 49:1233, [d]Ono K, Berry B and Douglas L (1986) Nutrient composition of some fresh and retail cuts of veal. *Journal of Food Science* 51:1352.

those of lean beef and lamb, as shown in Table 2. (*See* **Niacin**: Physiology; **Riboflavin**: Physiology; **Thiamin**: Physiology.)

The average biological value of goats' meat, based on feeding trials with rats fed a 10% level of protein, is 60.4 and beef 68.6. The digestibility coefficient of meat protein is 97%, giving ingested meat a heat combustion of 17.87 J g^{-1}. The effect of cooking on muscle is method-, time-, and temperature-dependent and the response of muscles to heat treatment varies between muscles and according to pre- and postmortem influences. Generally, temperatures below 100 °C affect palatability, but do not reduce the nutritive value of meat severely. (*See* **Protein**: Quality.)

Fat

Goat carcasses characteristically have a low proportion of subcutaneous fat and a high proportion of intermuscular fat, giving goat carcasses a lean appearance. In terms of total body fat, goats are not necessarily leaner than sheep. At the same degree of total body fatness (21% of empty body weight), dressed Boer goat carcasses recorded 22% carcass fat with 6.7% in the subcutaneous fat depot, while Dorper and South African Mutton Merino carcasses were both 24% fat, with 12.7% and 10.4% subcutaneous fat, respectively. The comparatively poor fat covering of goat carcasses means that the criterion of subcutaneous fatness, which is a reliable predictor of yield in lamb and mutton carcasses, is not suitable for classifying and grading goat carcasses. Lipogenesis occurs in a preferential order between the adipose depots. Visceral fat (omental, mesenteric, kidney, and pericardial) is the earlier-developing depot followed by intermuscular, subcutaneous, and intramuscular fat.

Carcass fat is highly variable and easily influenced by:

1. Sex – males are leaner than females and castrated males.
2. Nutrition – the greater the metabolizable energy intake (ME kg^{-1} DM), the fatter the carcass.
3. Age – increasing fatness to maturity is a phenomenon of normal growth in all sexes.
4. Physiological condition – lactating does will catabolize body fat reserves and become leaner; disease and stressful conditions reduce appetite, causing body reserves to be drawn upon.
5. Physical activity – physical activity, such as foraging or when determining social ranking, increases energy usage, yielding leaner carcasses; penned animals fatten more easily, yielding fatter carcasses.

Fat quality Fat remains an important quality determinant of carcasses and meat. The chemical and physical properties of fat have no direct effect on the commercial value of carcasses, as does fatness, but do influence the organoleptic properties and keeping quality of meat. The degree of saturation of fat, expressed as the iodine number and determined by the number of double bonds in the component fatty acids, is one of the most important characteristics affecting these quality parameters. Saturated fats containing long-chain fatty acids with no or very few double bonds solidify easily upon cooling, thus affecting the palatability of the meat. The less

saturated fats containing a greater number of fatty acids with double bonds are easily oxidized, either by direct chemical action or through intermediary activity of lipolytic enzymes. Direct chemical oxidation is less important in meat than the action of lipases that split fatty acids from triacylglycerols. The rate of autoxidation increases with the number of double bonds, increasing the probability of affecting the flavor and odor of the meat. Chemical oxidation releases peroxides with free radicals (ROO•—, RO•—, OH•—) which may react to damage proteins, enzymes, other lipids, and vitamins. In goat meat, as with other red meat with low lipid- to heme ratios, heme compounds can stabilize peroxides of free radicals and exert an antioxidant effect.

The profile of the long-chain acids of goat meat show oleic acid ($C_{18:1}$) to be the most abundant, with palmitic ($C_{14.0}$) and stearic acids ($C_{18.0}$) being relatively high. Nutritional influences on the fatty acid profile of ruminants are less than on monogastric animals, owing to the biohydrogenation and synthesis of fatty acids in the rumen. It would appear, however, that nutrition could cause subtle changes in ruminants, including goats. The high variance of each fatty acid in kids could be ascribed to the monogastric characteristic of suckling animals, which would be sensitive to nutritional influences.

Visceral fats are more saturated than subcutaneous fats, as is illustrated in the differences between the fatty acids of the subcutaneous and kidney depots. The unsaturated-to-saturated ratio in the subcutaneous fat of goats was 1.11 and of sheep 0.67. Fat from the triceps brachii, biceps femoris, and obliquus internus abdominus muscles of Korean goats contained 55.2–59.6% $C_{18:1}$ and 24.5–25.6% $C_{16:0}$. Unsaturated fatty acids predominated, ranging from 68.5 to 72.3%. Oleic acid levels were slightly higher in the subcutaneous, intermuscular, and kidney fats of Sudanese goats than sheep. Comparing iodine numbers within and between species could be misleading owing to the range of fatty acids occurring in the fat and the exogenous and endogenous influences on the fatty acid profiles. Fatty acids, particularly C_{18} and $C_{18:3}$, influence the flavor of lamb and both 4-methyloctanoic and 4-methylnonanoic acids appear to contribute to the odor of mutton and goats' meat.

Nutrient value Lipids, although recommended to constitute a smaller proportion of the western diet, are an essential nutrient source component in all human foods. In dispute is not whether lipids, including animal fat (triacylglycerols), are a health threat, but the extent to which specific fatty acids, the amount ingested as a proportion of total energy intake or cholesterol, constitute a health threat, in general. Energy-dense triacylglycerols provide readily metabolizable energy. Excessive energy consumption beyond the body's needs for a given physiological state leads to obesity and related health risks. Linoleic ($C_{18:2}$), linolenic ($C_{18:3}$) and arachidonic acid ($C_{20:4}$) acids, constituents of cell walls, mitochondria, and other active metabolic sites, are regarded as essential in the diet since the body cannot synthesize these long-chain polyunsaturates. (*See* **Fatty Acids**: Dietary Importance.)

Fatty acids appear to be closely associated with the body's immune response system. Excessive consumption of dietary fats rich in ω-6 polyunsaturated fatty acids have been shown to possess immunosuppressive properties, influencing immune function in disease, whereas ω-3 polyunsaturated fatty acids have both an antiinflammatory effect and reduce immunosuppressive properties. Goat fats have low ω-6 and ω-9 polyunsaturated fatty acids.

Cooking Normal cooking changes the composition of animal fat and increases the concentration as an energy source. Mass reduction can be 20–35%, of which 70–80% is lipid cookout. Fatty tissue of cooked meat is typically about three times as concentrated a source of energy as is cooked lean tissue, with variances between cuts. Prolonged cooking leads to rendering in which a very high percentage of fat cookout is obtained.

Processing

Traditionally, meat processing is a means of extending shelf-life (preserving) and producing a convenient item for use later and elsewhere. Processing is aimed at reducing the enzyme activity in the meat, retarding oxidation of the fat, and preventing spoilage by microorganisms. These aims have been achieved through drying, curing with salts, or smoking meat. Sausage manufacturing, for which meat is ground to varying degrees, is another form of processing, in principle for the same purpose. Either one or a combination of these procedures in various regions of the world has preserved goats' meat. (*See* **Meat**: Preservation.)

In modern times, meat is processed not only as a means of preserving, but also for producing consumer-acceptable products compatible with modern lifestyles and philosophy of a health-related quality of life. In order to achieve this the processor has to define clearly the image and commercial appeal of the envisaged product. The decision-making process of the consumer needs to be defined in terms of real and perceived value, the convenience the product offers, and its palatability.

Goat meat, regardless of breed, can be used to manufacture processed products of acceptable sensory quality, particularly when a spicy formulation is used. The low appeal of goat meat to some consumers may be due to its lack of tenderness, particularly in meat from older animals. Tenderness should not be a problem for comminuted meat products (e.g., patties) or products that undergo slow cooking or pressure cooking/retorting (e.g., curries and stews). Although goat meat has a less desirable flavor, aroma, tenderness, and juiciness than beef or pork to western taste panelists, a panel found up to 40% substitution of beef by Angora goat meat to be quite acceptable in frankfurters. The goat frankfurters had good physical attributes, being firm, resilient, and springy under forefinger pressure with a firm 'bite' – a desirable textural attribute in quality emulsified sausages. Such frankfurters would maintain their form and shape during peeling, indexing, and packaging operations. (*See* **Meat:** Sausages and Comminuted Products.)

Vienna sausages manufactured from meat of mature does (six teeth) had detectable different palatability characteristics compared with viennas manufactured from beef. The goat viennas also recorded higher shear force values ($P < 0.0002$). Differences in both physical and textural attributes could be related back to the characteristics of the raw meat. Cured and smoked buttocks of mature does (six teeth), compared with cured and smoked beef silverside (two teeth), received higher ratings ($P < 0.01$) in terms of aroma, tenderness, juiciness, and tastiness. No difference in residue could be detected.

Overall, the goat was rated 3.54 with a standard deviation of 0.6 on a scale of 0–5, as opposed to the 2.96 ± 0.97 rating of beef. The results suggest that smoked and cured goat buttock or leg of goat has the potential of being a delicacy and could compete comfortably with other products such as smoked beef and pork gammon.

Processing hot muscle holds decided advantages. Prerigor muscle has a higher water-holding capacity, better fat-emulsifying properties, and produces sausages with less moisture loss and rendering out when cooked. Patties manufactured from hot goat meat (3–4 h postmortem), not electrically stimulated, had lower cooking yields ($P < 0.05$) than patties prepared from chilled meat (24 h postmortem). However, reducing the postmortem time lapse to processing to 1–2 h improved the yield. The organoleptic scores of patties broiled in an oven at 190 °C for 15 min to an internal temperature of 72 ± 2 °C were similar for hot and chilled meat. Reheating precooked, frozen patties reduced sensory scores significantly ($P < 0.05$). The fat content of patties has a significant influence on cooking loss and flavor, texture, and overall acceptability. According to sensory scores a 20% fat content would be the optimum. The quality of warm minced goat meat (3 h postmortem) could be improved by addition of 2.5% sodium chloride and 1% tetrasodium pyrophosphate. These salts resulted in a significant increase in pH, water-holding capacity, and water-soluble proteins, and a decrease in cooking loss and improved redness and overall appearance. The effects on the emulsifying capacity and salt-soluble protein concentration were also significant. Indications are that bacterial growth can be inhibited by treatment with ammonia added at a concentration of 0.402–0.67 mol l^{-1}. The effect was more pronounced on Gram-positive bacteria. (*See* **Sensory Evaluation:** Texture; Taste.)

See also: **Amino Acids**: Metabolism; **Fatty Acids**: Dietary Importance; **Meat**: Structure; Slaughter; Preservation; Sausages and Comminuted Products; **Niacin**: Physiology; **Protein**: Quality; **Riboflavin**: Physiology; **Sensory Evaluation**: Texture; Taste; **Thiamin**: Physiology

Further Reading

Abdon I, Del Rosario IF and Olga LG (1980) *Food Composition Tables Recommended for Use in the Philippines. Handbook 1*, 5th edn. Manila: Food and Nutrition Research Institute.

Breuking HR and Casy NH (1989) Assessing the acceptability of processed goat meat. *South African Journal of Animal Science* 19: 76–80.

Casey NH and Paterson CP (1990) Quality of hot and cold boned, prepacked beef following electrical stimulation. *South African Journal of Food Science and Nutrition* 2: 99–102.

Casey NH and Van Niekerk WA (1985) Fatty acid composition of sub-cutaneous and kidney fat depots of Boer goats and the response to varying levels of maize meal. *South African Journal of Animal Science* 15: 60–62.

Gaili ES and Ali AE (1985) Meat from Sudan desert sheep and goats. II. Composition of the muscular and fatty tissues. *Meat Science* 13: 220–236.

Giménez MS, Baudini PM, Ojeda MS, Molins de Pedernera M and Giménez LA (1985) Some physicochemical properties and composition of adipose tissue of goats fed with different diets. *Nutrition Reports International* 32: 389–397.

Kondaiah N, Anjaneyulu ASR, Sharma N and Joshi HB (1985) Effect of salt and phosphate on quality of buffalo and goat meats. *Meat Science* 15: 183–192.

Lawrie RA (1985) *Meat Science*, 4th edn. Oxford: Pergamon Press.

National Research Council, National Academy of Sciences (1981) *Nutrient Requirements of Goats: Angora, Dairy and Meat Goats in Temperate and Tropical Countries*. Washington, DC: National Academy Press.

Ono K, Berry B, Johnson HK *et al.* (1984) Nutrient composition of lamb of two age groups. *Journal of Food Science* 49: 1233.

Ono K, Berry B and Douglas L (1986) Nutrient composition of some fresh and retail cuts of veal. *Journal of Food Science* 51: 1352.

Padda GS, Keshri RC, Sharma N, Sharma BD and Murthy TRK (1988) Physico-chemical and organoleptic properties of patties from hot, chilled and frozen goat meat. *Meat Science* 22: 245–253.

Panin A and Mahabile M (1997) Profitability and household income contribution of small ruminants to small-scale farmers in Botswana. *Small Ruminants Research* 25: 9–15.

Srinivasan KS and Moorjani MN (1974) Essential amino acid content of goat meat in comparison with other meats. *Journal of Food Science and Technology* 11: 123–124.

USDA (1986) *Composition of Foods: Beef Products – Fresh Processed Prepared. USDA, Human Nutrition Information Service, Agricultural Handbook* No. 8–13. Washington, DC: US Government Printing Office.

Wan Zahari WM and Wahid SA (1985) Mineral concentrations in the blood plasma and various tissue of local crossbred goats. *MARDSI Research Bulletin* 14: 333–340.

Webb EC, Cronjé PB and Donkin EF (1998) Research and training strategies for goat production systems in South Africa. In: Proceedings of a Workshop held on 22–26 November 1998 at Kings Lodge, Hogsback, Eastern Cape, South Africa. Available online at: http://www2.fao.org/agris.

Milk

M Guo, University of Vermont, Burlington, VT, USA

Introduction

More people drink goats' milk than the milk of any other species on this planet, although goats' milk production ranks third after bovine and buffalo milk. Dairy goat farming plays an important role in the diet and economy of many developing countries. Goats' milk provides one of the principal sources of animal protein, calcium, and phosphorus in many regions, such as the Middle East, southern Asia, and some tropical countries. It also serves as an appropriate substitute for cows' milk and is often recommended by medical professionals in situations where bovine milk may cause allergic responses in consumers. Nevertheless, goats' milk products are considered specialty foods but becoming more popular in the USA and other developed countries. The goats' milk industry will continue to grow during the next decade.

The major breeds of dairy goats in Europe and North America are Saanen, Toggenburg, Alpine, La Mancha, Nubian, Oberhasli, and Angora. In the Middle East, Asia, and other regions, various native goats, crossbreeds, and pure imported goats are used for milk production. In this article, chemical, physiochemical, and nutritional properties of goats' milk and goats' milk product technology will be briefly discussed.

Gross Composition

Gross composition includes total solids (TS), fat, crude protein, lactose, and ash. The gross composition of goats' milk in comparison with cows' and human milk is shown in **Table 1**. The levels of TS, fat, crude protein, and ash in goats' milk are, on average, higher than those in cows' milk. The lactose content of goat's milk is 4.5%, i.e., slightly lower than cow's milk ($\sim$ 4.7%). The gross composition of milk is affected by many factors, including breed, individual animal, diet, yield, season, stage of lactation, environment, and management. Fat content is the most variable component in goats' milk, ranging from 2 to 8%. There is an inverse relationship between milk yield and the concentrations of fat, protein, nonfat solids, and ash in the milk. Fat, protein, and ash contents are high in the early stage of lactation, fall sharply to reach the lowest levels during summer months, and then increase gradually towards late lactation. Lactose content does not show considerable changes during lactation. The ash content of both goats' and cows' milk are three to four times higher than that of human milk in general, the ash content of milk is closely related to the growth rate of the species for which it was designed. An inverse relationship between lactose and ash is essential and exists to maintain the osmotic pressure of the milk.

Lipids

The lipids in goats' milk consist mainly of triacylglycerols (98% of total lipids), phospholipids (1%), and cholesterol and its esters (1%). The cream or butter of goats' milk is white in color as there is very little

Table 1 Gross composition of goats', cows', and human milk (g 100 ml^{-1})

Nutrients	*Milk*		
	Goat	*Cow*	*Human*
Total solids	12.90	12.50	12.50
Fat	4.10	3.50	4.40
Crude protein	3.50	3.30	1.00
Lactose	4.50	5.00	6.90
Ash	0.80	0.70	0.20

Table 2 Fatty acid profiles of goats', cows', and human milk fat (%)

Fatty acids	*Goat*	*Cow*	*Human*
C_4	0.7	1.4	0.1
C_6	2.4	2.2	0.2
C_8	3.2	1.8	0.3
C_{10}	8.7	3.6	2.1
C_{12}	4.7	4.0	7.2
C_{14}	10.7	13.0	10.9
C_{16}	28.5	30.2	29.6
C_{18}	13.0	13.7	7.3
$C_{18:1}$	25.2	27.1	35.3
$C_{18:2}$	2.9	3.0	6.7

β-carotene in the milk. Like cows' milk, goats' milk lipids also exist in the form of globules. The average size of the fat globules in goats' milk (2 μm in diameter) is smaller than the globules in cows' milk (3 μm) and buffalo milk (5 μm). Goats' milk has a poor creaming ability at cooler temperatures due to its small fat globules and lack of agglutinin, a clustering agent which causes fat globules in milk to cluster upon cooling.

Goats' milk has a higher level of short-chain fatty acids (C_4–C_{12}) than cows' milk (Table 2): 15% of total fatty acid in goats' milk compared to 9% in cows' milk. Most of the short-chain fatty acids are esterified at the 3-position of glycerol while the longer chains are at 1 or 2 position. Lipases attack ester linkages of short-chain fatty acids more easily than those of longer ones. The level of free fatty acids in goats' milk is also higher than that in cows' milk. The short-chain free fatty acids, especially C_6 and C_8, are responsible for the specific 'goaty' flavor of goats' milk products. There is a strong positive correlation ($r = 0.83$) between the amount of free fatty acids and the goaty flavor score of the cheese made from goats' milk. Fatty acid composition is strongly affected by diet and stage of lactation. The fatty acid profile can be manipulated by feeding different fats. The concentration of essential fatty acids and conjugated linoleic acid (CLA) in goats' milk can be increased by feeding them with canola oil. The proportion of short-chain fatty acids (C_8–C_{14}) in goats' milk is lower, while that of $C_{18:0}$ and $C_{18:1}$ is higher due to does drawing on their body fat at the beginning of the lactation. The long-chain fatty acid profile of goats' milk is similar to that of cows' milk. The lipids of both goats' and cows' milk contain an adequate amount of essential fatty acids for human consumption. (*See* **Fats**: Classification.)

Goats' milk contains 30–40 mg phospholipids per 100 g milk or 8–10 mg g^{-1} fat. About 40% of total phospholipids in goats' milk is in the skim milk, with the remainder in the fat globule membrane. The cholesterol content of goats' milk is in the range of 10–20 mg 100 ml^{-1}. The cholesterol content in the colostrum of goats' milk is much greater (40 mg 100 ml^{-1}). Most of the cholesterol in goats' milk is in the free form, with only a small fraction in the form of esters.

Proteins

Goats' milk contains more crude protein, ranging from 3.0 to 3.8%, than cows' milk. In general, the protein profile of goats' milk is similar to that of cows' milk. However, goats' milk has a lower level of casein nitrogen and a higher level of nonprotein nitrogen than cows' milk. This is responsible for a low cheese yield and a weak yogurt structure/texture. The concentrations of amino acids in goats' milk are compared with those of cows' and human milks in Table 3. (*See* **Protein**: Food Sources.)

The five major proteins of goats' milk are α_{s2}-casein, β-casein, κ-casein, β-lactoglobulin, and α-lactalbumin. Sodium dodecyl sulfate-polyacrylamide gel electrophoretograms of goats' milk show that α_{s2}-casein is the slowest-moving of the three major casein bands, β-casein the major protein in the middle, and κ-casein migrates the fastest of the caseins (Figure 1, lanes 3 and 4). The amino acid compositions of these proteins is shown in Table 4. It was previously believed that goats' milk lacked a protein homologous to bovine α_{s1}-casein. However, recent studies have shown that this protein does exist in the milk of some goat's milk. The level of α_{s1}-casein affects the coagulation properties of the milk. β-Lactoglobulin (β-lg) and α-lactalbumin (α-la) are the

Table 3 Amino acid composition of goats', cows', and human milks (mg 100 g^{-1})

Amino acid	*Goat*	*Cow*	*Human*
Typtophan	44	46	17
Threonine	163	148	46
Isoleucine	207	198	56
Leucine	314	321	95
Lysine	290	260	68
Methionine	80	82	21
Cystine	46	30	19
Phenylalanine	155	158	46
Tyrosine	179	158	53
Valine	240	220	63
Arginine	119	119	43
Histidine	89	89	23
Alanine	118	113	36
Aspartic acid	210	249	82
Glutamic acid	626	687	168
Glycine	50	69	26
Proline	368	318	82
Serine	181	178	43

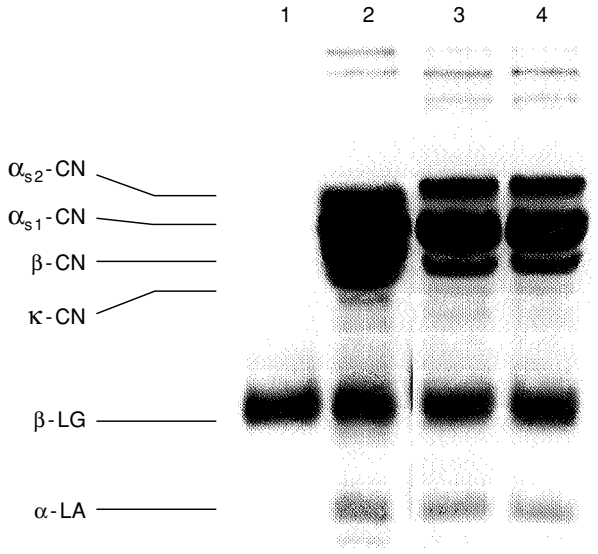

Figure 1 'Sodium dodecyl sulfate polyacrylamide gel electrophoretogram (SDS-PAGE) of samples of goats' milk (lanes 3 and 4), cows' milk (lane 2), and standard β-lactoglobulin (lane 1).

Table 4 Amino acid composition of individual goats' milk proteins (residues per mole)

Amino acid	*α_{s2}-Casein*	*β-Casein*	*κ-Casein*	*β-lg*	*α-la*
Alanine	10	5	16	16	5
Arginine	6	3	5	3	1
Aspartic acid	17	9	16	14	22
Cystine	2	0	3	5	8
Glutamic acid	45	43	26	24	13
Glycine	4	6	1	5	5
Histidine	5	5	4	2	3
Isoleucine	11	9	11	10	8
Leucine	12	20	8	21	13
Lysine	22	12	8	16	13
Methionine	4	6	1	4	0
Phenylalanine	8	9	4	4	4
Proline	18	33	19	8	2
Serine	14	15	13	6	6
Threonine	14	12	15	8	6
Tryptophan	2	1	1	2	4
Tyrosine	11	4	9	4	4
Valine	12	21	11	10	6
Total	217	213	171	162	123

two major whey proteins of goats' milk (Figure 1, lanes 3 and 4).

β-Casein is quantitatively the major component in the casein fraction of goats' milk. It consists of 213 amino acid residues, compared to 209 amino acid residues in bovine β-casein. Goats' κ-casein is a chain of 171 amino acid residues instead of 169 for its bovine counterpart due to valine and histidine being inserted at positions 132 and 133. Like the bovine homolog, goats' κ-casein has phenylalanine at position 105 and methionine in 106. Therefore, chymosin can hydrolyze the molecule between these two residues producing, *para*-κ-casein and caseinomacropeptides. This process leads to rennet-induced coagulation during cheese-making. Goat α_{s2}-casein differs from its bovine counterpart in that it contains a chain of 217 amino acid residues instead of 207.

Goat β-lg, like its bovine homolog, consists of a polypeptide chain of 162 amino acid residues. Like bovine milk, goats' α-la is a chain of 123 amino acid residues. (*See* **Amino Acids:** Properties and Occurrence.)

Like bovine milk, the casein in goats' milk exists as micelles. The majority of goat casein micelles are less than 80 nm in diameter, which is smaller than the casein micelles (120 nm in diameter) in cows' milk. The proportions of small micelles is much greater than in cows' milk. The stability, especially alcohol stability, of casein micelles in goats' milk appears to be lower than those in cows' milk.

Carbohydrates

Lactose is the major carbohydrate in goats' milk. The content of lactose in normal goats' milk is about 130 mmol l^{-1}, or in the range of 4.2–4.8%. This is slightly lower than the level in cows' milk. Other minor carbohydrates in goats' milk include inositol (7 mg 100 ml^{-1}), and trace amounts of galactose and sialic acid. (*See* **Lactose.**)

Minerals

The mineral contents of goats' milk is compared with that in cows' and human milk in Table 5. Goats' milk contains about 133 mg calcium and 100 mg phosphorus per 100 ml. Human milk contains only about 16–25% of these concentrations. Goats' milk has a higher content of calcium, phosphorus, potassium, magnesium, and chlorine and a lower level of sodium than cows' milk. Unlike the major minerals, the concentrations of trace elements in goats' milk are affected by diet, breed, individual animals, and stage of lactation. Goat colostrum has two to four times as much iron (Fe), copper (Cu), manganese (Mn), and zinc (Zn) than mature milk. Most milks, including human milk, are deficient in Fe. Mature fresh goats' milk has about 0.07 mg Fe per 100 ml (Table 4) and the zinc content is the highest among the trace elements, at 0.56 mg per 100 ml. Goats' milk generally contains more Mg, and lower or comparable levels of iodine (I) and Cu than cows' milk. (*See* **Minerals – Dietary Importance.**)

Vitamins

Goats' milk contains a higher level of vitamin A than cows' milk due to the fact that goats convert all

Table 5 Mineral concentrations in goats', cows', and human milks (per 100 ml)

Minerals	*Goat*	*Cow*	*Human*
Ca (mg)	133	119	32
P (mg)	100	93	14
K (mg)	190	151	51
Na (mg)	44	49	17
Cl (mg)	165	104	43
Mg (mg)	16	12	4
Citrate (mg)	130	156	60
Zn (mg)	0.56	0.53	0.38
Fe (mg)	0.07	0.08	0.12
Cu (μg)	50	60	40
Mn (μg)	0.8	3.0	1.2
I (μg)	22	26	7

Table 6 Vitamin content of goats', cows', and human milk (per 100 g)

Vitamins	*Goat*	*Cow*	*Human*
Vitamin A (IU)	185	126	190
Vitamin D (IU)	2.1	2.0	2.2
Thiamin (mg)	0.068	0.045	0.017
Riboflavin (mg)	0.21	0.16	0.04
Niacin (mg)	0.27	0.08	0.17
Pantothenic acid (mg)	0.31	0.32	0.20
Vitamin B_6 (mg)	0.046	0.042	0.011
Folic acid (μg)	1.0	5.0	5.5
Biotin (μg)	1.5	2.0	0.4
Vitamin B_{12}	0.07	0.4	0.03
Vitamin C (mg)	1.5	2.1	4.3
Choline (mg)	15	14	9

carotenes into vitamin A, which gives the milk a whitish color. Goats' milk supplies adequate amounts of vitamin A and niacin, and excesses of thiamin, riboflavin, and pantothenate for human infants (Table 6). Both goats' and cows' milk are equally deficient in vitamin C and D; therefore, these vitamins must be supplemented with other food sources. Goats' milk contains only 20% as much folic acid as cows' milk. The bioavailability of folic acid in goats' milk is also lower than that in cows' or human milk, so goats' milk must be fortified with folic acid for infant feeding in order to prevent infants from getting anemia. (*See* **Vitamins**: Overview.)

Physiochemical Properties

The pH of goats' milk is somewhat lower than that of cows' milk, with values ranging from 6.3 to 6.7. The average pH of goats' milk is about 6.5, compared with 6.7 for cows' milk. The titratable acidity of goats' milk is 0.15%, expressed as a lactic acid, which is slightly lower than that of cows' milk (0.16%).

The osmolality of goats' milk is about 295 mosmol kg^{-1} H_2O. The freezing point of goats' milk is -0.580, compared with -0.540 for cows' milk, indicating that the former contains a higher level of solute constituents, such as salts, than the latter.

The average electrical conductivity of goats' milk is 0.0052 $\Omega\,cm^{-1}$, which is slightly greater than that of cows' milk (0.0048 $\Omega\,cm^{-1}$). The refractive index of goats' milk lies midway between those of cows' and buffalos' milk, in a range of 1.3454–1.3492 at 40 °C.

The surface tension of goats' milk is similar to that of cows' milk, with values ranging from 40 to 55 dyn cm^{-1}, which is much less than water (72 dyn cm^{-1}). The viscosity of goats' milk is about 13 mPa, which is slightly lower than that of cows' milk (15 mPa).

Like milk from other species, goats' milk has a rather wide range of specific gravity – 1.026–1.042. The relationship between percentage of TS and specific gravity (SG) and fat (F) content for goats' milk has been reported by the author's research group as follows:

$$TS = 0.13SG + 1.41F + 4.28 (r^2 = 0.94, P < 0.01)$$

Nutritional Properties

In general, there is little difference in nutritional value between goats' and cows' milk since the chemical composition of both milks is very similar. A randomized double-blind clinical study on the growth of undernourished children using goats' milk as a substitute for cows' milk showed that goats' milk has a nutritional value similar to that of cows' milk.

Goat's, cows', and human milk provide approximately the same amount of energy: each furnishes about 3135 kJ of energy per liter. The outstanding difference is in the proportions of energy derived from lactose and protein. In goats' and cows' milk, fat, protein, and lactose account for about 50, 25, and 25% of the energy, respectively, but in human milk they furnish 55, 7, and 38%, respectively.

A human infant fed solely on goats' milk is oversupplied with protein in relation to calories. In terms of protein quality, profiles for human and cows' milk proteins are similar to those of goats' milk. The three milks differ in the proportions and kind of proteins, but the overall amino acid composition of the mixture of proteins is similar in all three. All three have a satisfactory balance of essential amino acids, meeting or exceeding the Food and Agriculture Organization – World Health Organization requirements for each amino acid.

Both goats' and cows' milk contain adequate concentrations of essential fatty acids (e.g., linoleic acid) for human infants. In either of these milks, linoleic acid provides about 1% of the total calories. The fat in goats' milk has been reported to be more easily

digested than that of cows' milk because the fat globules are smaller in goats' milk. However, there is no scientific evidence to confirm this claim. Nearly 20% of the fatty acids of goats' milk fat are in the category of short- and medium-chain length (4–12 carbons). Cows' milk fat contains only about 12% of fatty acids in this category. This difference may contribute to a greater ease in the digestion of goats' milk fat as lipase attacks ester linkages of such fatty acids more readily than those of longer chains.

One of the most important contributions of milk to human nutrition is the supply of calcium and phosphate. Goats' milk contains about 120 mg 100 ml^{-1} or 300 mg per serving of Ca, and 100 mg 100 ml^{-1} or 250 mg per serving of P. These concentrations are similar to those in cows' milk. Human milk contains much lower levels of these minerals, with only about 25% as much calcium and ~15% as much phosphorus. Therefore, goats' milk provides an excess of calcium and phosphorus in relation to energy for human infants. Most milks, including goats' milk as well as human and cows' milk, are deficient in iron. Iron is absorbed and poorly retained from most food sources. This metal is absorbed better from human milk than from cows' milk due to a higher level of lactoferrin in human milk. It should be noted that both goats' and cows' milks contain higher concentrations of potassium and chloride than human milk. These milks should be diluted before feeding to human infants in order to prevent the onset of acidosis.

Goats' Milk Products

Traditionally, goats' milk was produced on small farms for home consumption more economically than cows' milk. However, nowadays, fresh goats' milk is sold pasteurized in a variety of packaging in supermarkets and health food stores in the USA. Yogurt, icecream, cheese, and powdered milk are also considered to be specialty products in many western countries.

Pasteurized and homogenized goats' milk products are available in US markets, including 2% fat and whole milk, with or without fortification of vitamins A and D.

Yogurt and other fermented milks have been made from goats' milk for centuries. Yogurt products have been shown to be effective in the prevention and treatment of gastrointestinal disorders, such as diarrhea and infantile gastroenteritis, due to the antimicrobial activities of probiotic cultures in the gut. Yogurt and other fermented milks are also reported to have other health benefits, including blood cholesterol reduction and possible cancer prevention. Goats' milk yogurt can be made in a similar manner to cows' milk yogurt. Because of the low casein content in goats' milk, the most realistic and economical approach for the production of goats' milk yogurt with good texture and consistency is through fortification of goats' milk with milk powder and/or stabilizers. A probiotic goats' milk yogurt with improved texture by enzymatic cross-linking has been developed in the author's laboratory at the University of Vermont. Goats' milk is also suitable for the production of other functional fermented products, such as kefir and dietary supplements, e.g., *Lactobacillus acidophilus* capsules or tablets.

Goats' milk icecream is being produced by a number of small companies, in both the USA and other countries. Because goats' milk products are considerably more expensive than comparable cows' milk products, premium icecream is warranted in order to attract consumers. A premium icecream should contain a high fat-to-milk solids nonfat (MSNF) ratio. A typical premium goats' milk icecream mix usually contains 14% fat, 10% MSNF, 18% sweetener, 1.5% egg yolk solids, and 0.25% stabilizer-emulsifier.

Over 400 varieties of goats' milk cheese are listed in the *Agricultural Handbook* of the US Department of Agriculture (No. 54–4). The large number of varieties of cheese made from goats' milk have resulted in a great diversity in the nature of the products. The most important cheeses made from goats' milk include fresh cheeses (chèvre), ripened cheeses (Cheddar, Gouda, and Roquefort), and others such as whey cheese (ricotta). France offers the best goats' milk cheeses, many of which are surface-ripened. The chemical composition of some commercial goats' milk cheeses manufactured in the USA is shown in Table 7.

Other commercial goats' milk products, such as powdered milk, butter (also called ghee in India), and evaporated milk are also being produced commercially in the USA and other countries. The manufacturing

Table 7 Composition of selected commercial goats' milk cheeses in the USA

Nutrients	*Fresh soft*	*Feta*	*Cheddar*	*Shepherd's hard*
Moisture (%)	59.8	52.3	41.7	31.6
Fat (%)	22.5	25.3	26.6	33.1
Protein (%)	19.8	25.1	30.3	28.9
Ash (%)	1.74	4.30	3.60	3.93
Calcium (mg 100 g^{-1})	172	639	599	1035
Phosphorus (mg 100 g^{-1})	275	544	526	737
Potassium (mg 100 g^{-1})	25.8	62.2	11.0	54.4
Sodium (mg 100 g^{-1})	416	916	361	285
Magnesium (mg 100 g^{-1})	14.6	17.6	42.0	61.7
Zinc (mg 100 g^{-1})	0.91	1.55	0.65	1.87

technologies of these products are similar to those for corresponding cows' milk products.

See also: **Carbohydrates**: Classification and Properties; **Cheeses**: Types of Cheese; **Fatty Acids**: Properties; **Infants**: Nutritional Requirements; **Lactose**; **Protein**: Food Sources; **Yogurt**: Dietary Importance

Further Reading

Barillet F and Zervas NP (1999) *Milking and Milk Production of Dairy Sheep and Goats – Proceedings of the Sixth International Symposium on the Milking of Small Ruminants*. Athens, Greece, September 26–October 1, 1998. Wageningen, The Netherlands: Wageningen Pers.

Guo MR (1993) *Goat's Milk and Its Product Technology* (in Chinese). Harbin, PR China: Heilongjiang Science and Technology Publishing House.

Guo MR, Wang SP, Li Z and Kindstedt PS (1998) Ethanol stability of goat's milk. *International Dairy Journal* 8: 57–60.

Guo MR, Dixon PH, Park YW, Gilmore JA and Kindstedt PS (2001) Seasonal changes in the chemical composition of commingled goat's milk. *Journal of Dairy Science* 84: E79–E83.

Haenlein GFW (1996) Status and prospects of the dairy goat industry in the United States. *Journal of Animal Science* 74: 1173–1181.

Haenlein GFW and Ace DL (1984) *Extension Goat Handbook. Goat Milk Versus Cow Milk*. Fact Sheet E 1–3. Washington, DC: US Department of Agriculture.

Jenness R (1980) Composition and characteristics of goat milk: review 1968–1979. *Journal of Dairy Science* 63: 1605–1630.

Juarez M and Ramos M (1986) *Physiochemical Characteristics of Goat Milk as Distinct from those of Cow Milk of International Dairy Federation* Brussels: no. 202, 54–67.

Loewenstein MM (1982) Dairy goat milk and factors affecting it. In: *Proceedings of the Third International Conference on Goat production and Disease*, pp. 226–236. Scottsdale, AZ: Dairy Goat Journal Publishing Company.

Park YW (1990) Nutrient profiles of commercial goat milk cheeses manufacture in the United States. *Journal of Dairy Science* 73: 3059–3067.

GOITROGENS AND ANTITHYROID COMPOUNDS

J Vanderpas, Faculté Médecine, FUNDP, Namur, Belgium

Biochemistry of Thyroid Function and Environmental Factors

Besides iodine deficiency, certain environmental agents, both naturally occurring and synthetic, interfere with thyroid function. They may cause goiter and thyroid dysfunction by acting directly on the thyroid gland or indirectly by altering its regulatory mechanisms and thyroid hormone homeostasis. They are thus called goitrogens or antithyroid compounds, according to their main effect on thyroid volume or on thyroid hormone metabolism.

Figure 1 shows some of the principal steps in thyroid hormone synthesis, secretion, and metabolism which can be influenced by environmental goitrogens. The unique known function of iodine in mammals is its role in the synthesis of thyroid hormones. The uptake of iodide by the thyroid gland is stimulated by the pituitary hormone thyroid-stimulating hormone or thyrotropin (TSH), and its serum concentration can be multiplied > 100 times in some extreme environmental conditions of iodine deficiency and goitrogen toxicity. Some goitrogens, such as thiocyanate and isothiocyanate, interfere with iodide uptake.

After being concentrated by the thyroid gland, iodide is oxidized and bound to some specific tyrosyl residues of thyroglobulin, forming monoiodotyrosine (MIT) and diiodotyrosine (DIT); oxidation of iodide and iodination of selected tyrosyl residues is catalyzed by a heme-containing membrane-bound enzyme, thyroperoxidase (TPO). The same enzyme couples two iodotyrosines to form thyroxine (T_4; four iodine atoms) and triiodothyronine (T_3; three iodine atoms). The ratio of T_4 to T_3 in thyroglobulin is closely dependent on serum TSH concentration and iodine supply: in euthyroidism (normal healthy function of the thyroid gland), 85% of the hormonal content of the thyroid is represented by T_4, while in the hypothyroid state, the ratio of T_4:T_3 within the thyroid is closer to unity. Most of the goitrogens (phenol derivatives, dihydrobenzoic acid, flavonoids, goitrin, disulfides, polycyclic aromatic hydrocarbons, excess iodine) interfere with TPO activity.

Intrathyroid proteolysis of thyroglobulin results in the release of T_4 and T_3 in the systemic circulation; this step is inhibited by lithium and excess iodine. Secreted

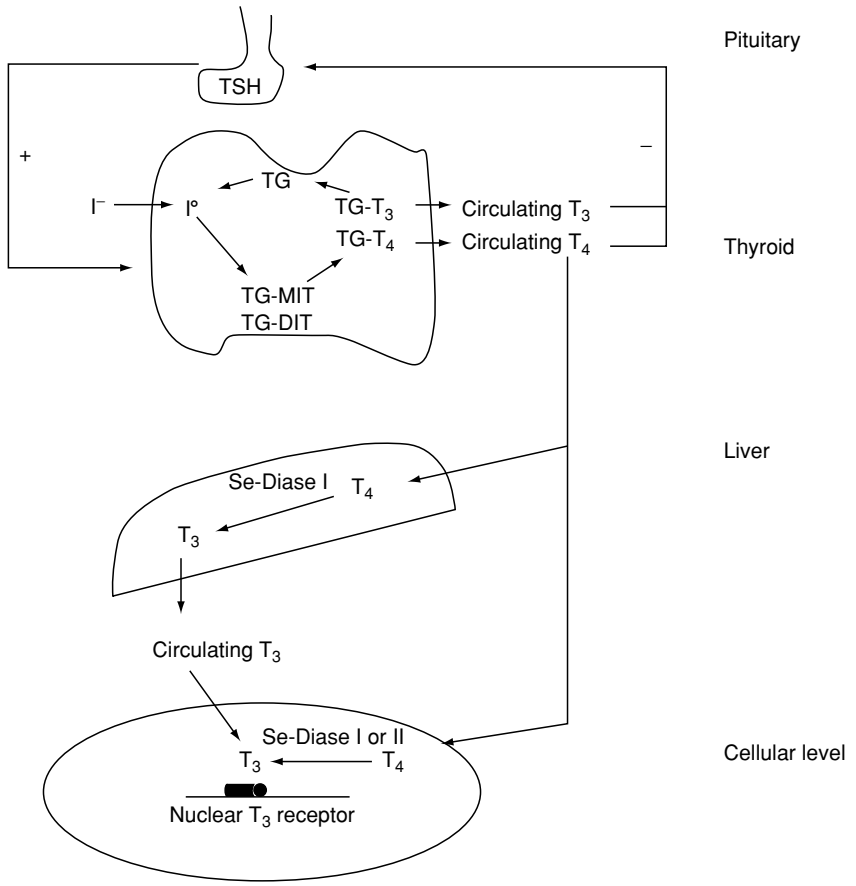

Figure 1 Some metabolic steps in iodine metabolism. Pituitary thyrotropin (TSH) stimulates thyroid iodide (I^-) uptake, and enzymatic organification of iodine (I°) by thyroperoxidase (TPO) in mono- and diiodotyrosine (MIT and DIT) which are further coupled by the same enzyme TPO in tri- and tetraiodothyronine (T_3 and T_4). Part of T_4 is further deiodinated to T_3 in the liver by a selenium-containing enzyme selenium-dependent deiodinase I (Se-Diase I). At the cellular level, the active hormone is T_3, which binds to a specific nuclear T_3 receptor. T_3 comes directly from circulating serum T_3 or from intracellular conversion of T_4 to T_3 TG, thyroglobulin.

in the systemic circulation, serum T_4 and T_3 are bound to albumin, thyroxine-binding globulin and transthyretin (also named thyroxine-binding prealbumin). The binding of thyroid hormones to serum proteins is partly inhibited by flavonoids, resulting in alterations of circulating free hormone concentrations and adaptation of feedback regulation. At the cellular level, the prohormone T_4 is deiodinated to the active hormone T_3 by the action of deiodinases; all three isoforms of deiodinases are selenium-containing enzymes. The type 1 deiodinase is abundant in peripheral tissues like liver, kidney, heart, muscle, thyroid; the type 2 deiodinase is the predominant form in the central nervous system. (*See* **Hormones**: Thyroid Hormones; **Iodine**: Physiology; Iodine-deficiency Disorders.)

Iodine Excess

Since the first descriptions of efficiency of iodine supplementation on goiter in 1820, adverse effects of iodine have been observed.

Currently, a number of drugs have a large iodine content with a very slow metabolism. For example, amiodarone is a very common antiarrhythmic drug, and the usual 200-mg tablet contains 75 mg iodine in an organic form, of which 6 mg is released every day as free iodine; as the molecule is highly liposoluble, excess iodine persists for months after therapy cessation. Other compounds rich in iodine are some X-ray contrast agents, expectorants made from potassium iodide (Quadrinal), and skin disinfectants such as povidone-iodine (Isobetadine). Maternal contamination with povidone-iodine has been documented as a cause of neonatal hypothyroidism.

Some goiter endemias result from iodine in excess in the environment, e.g., ingestion of iodine-rich seaweed in Hokkaido, Japan, or ingestion of water contaminated with iodine in Shanxi province, China (Table 1). An excess intake of iodine, defined as 2 mg or more daily, inhibits the proteolysis and release of thyroid hormones and eventually produces hypothyroidism.

Table 1 Goiter endemias attributed to goitrogens in foodstuffs or to inadequate supply of trace elements

Cause	*Country*	*Source*	*Active compound*
Goitrogens in foodstuffs	Central Africa	Cassava	Cyanogenic glucoside
	Nigeria	Cassava	Cyanogenic glucoside
	Darfur, Sudan	Millet	Flavonoids
			Cyanogenic glucoside
	Maranhao, Brazil	Babassu	?
		Mandioca	Cyanogenic glucoside
	Tasmania	Grass and weeds	Isothiocyanate (cheilorine)
	Finland	Grass and weeds	Goitrin
	Cauca valley, Colombia	Water	Coal-derived water pollutants
	Eastern Kentucky, USA	Water	Coal-derived water pollutants
Inadequate supply of trace elements	Hokkaido, Japan	Seaweeds	Iodine overload
	Zaire	Soil	Selenium deficiency
	Himalayas	Water	Fluoride excess

Iodine excess in nonendemic areas involves in the short term a blockade of iodine organification in the thyroid gland (Wolff–Chaikoff effect); in most instances, there is an escape phenomenon after a few weeks, the thyroid hormonal parameters returning to normal. In a few subjects, a persistent dysthyroid function (hypothyroidism or hyperthryroidism) develops by a yet incompletely explained mechanism. The prevalence of chronic autoimmune thyroiditis in industrialized countries is correlated with iodine intake, with the highest prevalence in countries with the highest intake of iodine, such as the USA and Japan. Iodine supplementation in areas where iodine is deficient increases the prevalence of lymphocytic infiltration of the thyroid threefold, and the prevalence of positive serum tests for thyroid antibodies in such areas rises to over 40% within 0.5–5 years. Amiodarone-induced hypothyroidism usually occurs within 18 months, and is seven times more likely to develop in patients with thyroid antibodies than in those without thyroid antibodies.

Iodine-Induced Hyperthyroidism

Universal salt iodization was the World Health Organization plan for iodine-deficient-countries in Africa, Asia and South America which were previously known as endemic in 1990–91. The marked decrease in the prevalence of goiter during the following years was a striking confirmation of the effectiveness of universal salt iodization in Africa. Nevertheless, iodine-induced hyperthyroidism was clearly documented in two countries. A balance has to be reached between the desirable daily iodine supply necessary to prevent goiter ($> 100\,\mu g\,day^{-1}$) and excessive iodine ($> 400\,\mu g\,day^{-1}$). The risk of developing iodine-induced hyperthyroidism is small (< 1% in the general population); it has been impossible to obtain a precise incidence rate up to now; this incidence rate is increased in the subgroup of women over 40 years of age with long-standing, nodular, and potentially autonomous goiter. From experimental data, it is considered that autonomy of the gland results from a mutation of TSH-receptor in thyroid tissue, which renders the hormonal function independent of its control by TSH. As long as iodine deficiency is present, the multinodular goiter is functionally inefficient. With iodine supplement, the autonomous parts of the gland become active, and the patient develops hyperthyroidism (sometimes named Jod-Basedow hyperthyroidism). Even if the number of cases is low and the clinical signs are asymptomatic or moderate most of the time (profuse sweating, weight loss, increased appetite, ocular changes (upper-eyelid retraction), tachycardia, change of mood (hyperactivity, fatigability, anxiety)), some deaths occurred in relation to the introduction of universal salt iodization in Africa. It is likely that the problem is transient in public health terms, as the *primum movens* cause – presence in the population of autonomous thyroid gland due to iodine deficiency – will decrease with time.

Sulfur-Containing Compounds

Antithyroid sulfur-containing organics identified in vegetable foodstuffs can be divided into two categories according to the way they act on iodine metabolism: (1) thiocyanate and isothiocyanate-derived compounds primarily inhibit the active concentration mechanism of iodide, and their goitrogenic activity can be overcome by iodine administration; (2) goitrin (or 1,5-vinyl-2-thiooxazolidone) is analogous to thiourea in its action, interfering with the TPO organification process, and its action cannot usually be antagonized by iodine. (*See* **Plant Antinutritional Factors**: Detoxification.)

Thiocyanate and Isothiocyanate

Thiocyanate and cyanide do not occur in the intact plant as free anions. They result from the hydrolysis of a thioglucoside by specific enzymes such as the couple linamarine–linamarase in cassava when the roots are crushed. In most populations, crushed roots are detoxified by sinking a few days in river water, or by exposing to sun. In some Central Africa rural ethnic groups (such as Mbanza and Ngbaka from Northern Democratic Republic of Congo and Central Africa Republic), cassava is poorly detoxified, resulting in an elevated content of cyanide. This last anion is rapidly detoxified by sulfur-transferases (also named rodanese) in the liver and kidney (Figure 2).

The principal vegetables containing thioglucosides are kale, cabbage, sprouts, broccoli, kohlrabi, turnips, swedes, rapeseed, and mustard. The main vegetables containing cyanogenic glucosides are cassava, lima beans, linseed, bamboo shoots, and sweet potatoes. One of the most widespread mustard oils is allylisothiocyanate (CH2═CH—CH2—N═C═S), which is the principal mustard oil in cabbage.

The inhibitory action of thiocyanate on iodine uptake is due to a competitive effect of pseudohalide with the mechanism of iodide concentration. However, under experimental conditions, rather high plasma concentrations of thiocyanate are required for inhibiting iodine uptake by the thyroid gland. In the historical use of thiocyanate as a vasodilator in hypertension, administration of 3 g of thiocyanate results in serum levels in the range of 8.0–15.0 mg dl^{-1}, far exceeding the levels attained in populations exposed to environmental thiocyanate overload (around 2.0 mg dl^{-1}). In ethnic groups with the highest prevalence of goiter, cassava is not soaked; the roots are peeled, dried in the sun for 1 or 2 days, then bruised in a mortar with corn that has been steeped for 12–24 h in water. The flour is eaten as a gruel (fuku) prepared in boiling water. In other ethnic groups of Central Africa where goiter prevalence is lower, the roots of cassava are soaked for 2–6 days and then mashed into a purée which is simmered to form a paste of firm and elastic consistency. The paste is enveloped in a palm or banana leaf (chikwangue). The cassava leaves are ground and extensively boiled before consumption (mpondu). Table 2 shows the cyanide content of foods before preparation or ready to eat. Table 3 shows that thiocyanate overload in northern Democratic Republic of Congo is present at birth, resulting from the placental transfer of maternal serum thiocyanate. Thiocyanate is not present in excess in maternal milk as compared to a group of mothers not consuming cassava. Lack of thiocyanate overload associated with a stable iodine deficiency (mammary gland does not concentrate iodine in humans) could explain the relative mitigation of endemic hypothyroidism in infants of Ubangi during the breast-feeding period, and the deterioration of thyroid function after weaning which is chronologically associated with the introduction of cassava meals in food. Figure 3 shows the relationship of iodine

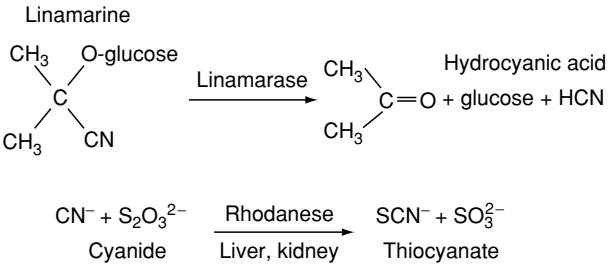

Figure 2 Enzymatic degradation of cassava linamarin by autolysis due to linamarase, and pathway of endogenous conversion of cyanide into thiocyanate with liver and kidney sulfur transferase (rhodanese). Reproduced from Goitrogens and Antithyroid Compounds, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 2 Comparison of cyanide content of foods before preparation or ready to eat, in Ubangi, Zaire

Food	*Cyanide (mg kg^{-1}) (mean ± SEM)*
Before preparation	
Cassava leaves	92.1 ± 21.6
Cassava and corn	19.8 ± 2.5
Cassava flour	4.5 ± 2.7
Wild yams	7.3 ± 0.4
Cultivated yams	8.5 ± 0.3
Bananas (bananes plantain)	4.4 ± 0.7
Ready to eat	
Sun-dried cassava root and corn (fuku)	14.2 ± 0.7
Soaked and boiled cassava root (chikwangue)	3.5 ± 0.4
Boiled assava leaves (mpondu)	8.5 ± 0.7
Cane sugar	2.6 ± 0.1

Reproduced from Ermans AM, Mbulamoko NM, Delange F and Ahluwalia R (1980) *Role of Cassava in the Etiology of Endemic Goitre and Cretinism*. Ottawa: International Development Research Centre, with permission.

Table 3 Thiocyanate concentration in maternal serum and in cord serum at delivery, and in maternal milk during breast-feeding in subjects of Ubangi, Zaire

	Ubangi, Zaire	*Belgium*	*t-Test*
Maternal serum at delivery	0.77 ± 0.53	0.21 ± 0.17	$P < 0.001$
Cord serum	0.78 ± 0.56	0.19 ± 0.22	$P < 0.001$
Maternal milk	0.33 ± 0.15	0.26 ± 0.13	NS

NS, not significant.
Reproduced from Vanderpas J *et al*. (1984) Endemic infantile hypothyroidism in a severe endemic goitre area of Central Africa. *Clinical Endocrinology* 20: 327–340, with permission.

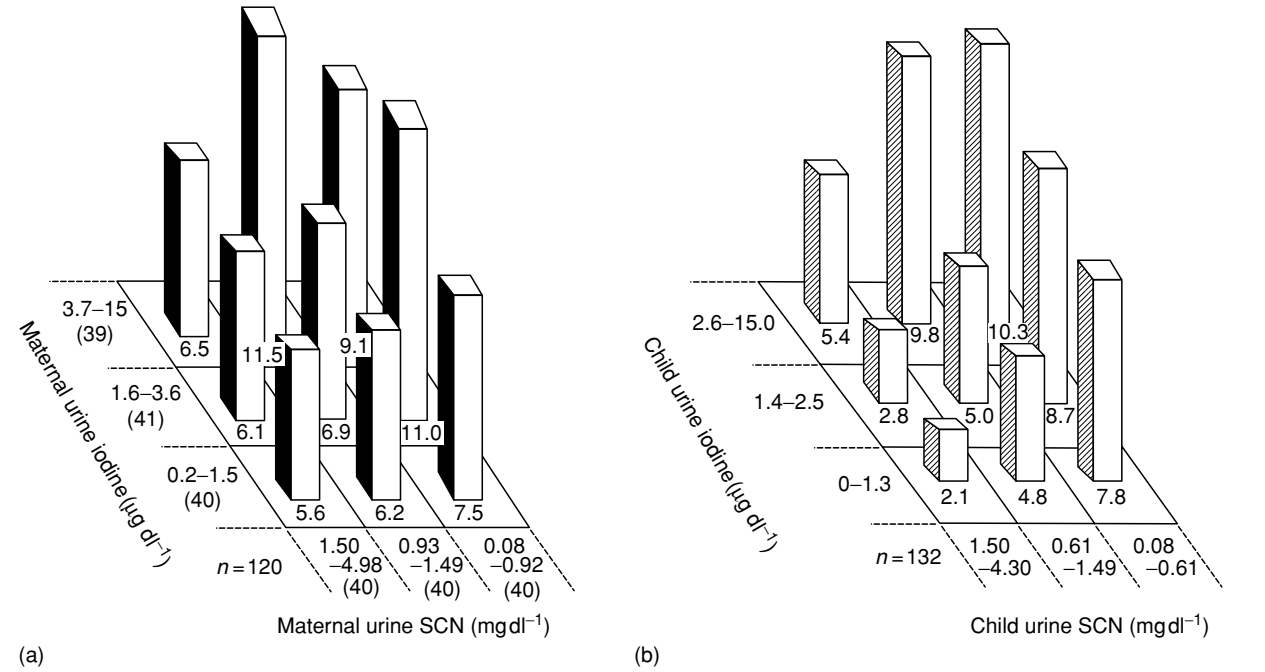

Figure 3 Level of serum thyroxine (T_4) in (a) newborn babies and (b) children aged 0–7 years according to the urine concentrations of iodine and thiocyanate (SCN). Reproduced from Goitrogens and Antithyroid Compounds, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

deficiency and of thiocyanate overload in neonates and in children of Ubangi, Democratic Republic of Congo, with mean serum T_4 concentration. At birth and during childhood, thiocyanate overload and iodine deficiency act in synergy to produce low serum T_4 concentrations. Once iodine deficiency has been corrected, the goitrogenic effect of thiocyanate is no longer present. (*See* **Cassava**: The Nature of the Tuber.)

Smoking

Tobacco smoking is another cause of thiocyanate overload. In a country with borderline iodine deficiency (Belgium), the risk of neonatal hypothyroidism seems to be higher in neonates from smoking mothers. Impaired action of thyroid hormone associated with smoking in women with hypothyroidism has also been shown: smoking was associated with greater serum TSH, total and low-density lipoprotein-cholesterol and creatine kinase concentrations, and prolonged ankle-reflex time. These effects suggest an interaction of smoking with thyroid hormone receptors, different from the effects of thiocyanate. (*See* **Smoking, Diet, and Health.**)

Goitrin and Aliphatic Disulfides

On enzyme hydrolysis, a particular thioglucoside gives rise to progoitrin, which is rapidly converted to form goitrin. Progoitrin is present in swedes and turnips. Administration of goitrin to rats for 20 days induces an enlargement of the thyroid, decreases iodine uptake by the gland, and decreases T_4 synthesis. Other sources of goitrin-like compounds have been detected in various species of herbs and shrubs of the Barbarea and Residea families.

The role of goitrin has been advocated in the development of endemic goiter in schoolchildren of Tasmania; this goiter could not be prevented by iodine supplementation. The endemia was attributed to a goitrogenic factor ingested by cattle from thousand-headed kale and transmitted by milk. However, only 0.05% of goitrin appears in the milk, and is rapidly denatured unless the milk is heated immediately. Goitrin has also been involved as an environmental goitrogen in Finland.

The major components of the volatile compounds from onion and garlic are small aliphatic disulfides which depress uptake of radioactive iodide by the thyroid of rats on a low-iodine diet. Disulfides are also present in high concentrations ($0.3–0.5\,g\,l^{-1}$) in aqueous effluents from coal-conversion processes; it is possible that these water contaminants intervene in the etiology of endemic goiter in the Cauca districts of Colombia.

Flavonoids

Flavonoids are a broadly distributed class of plant pigments, universally present in vascular plants.

They are not synthesized in animal tissues. Their number exceeds 3000 and new structures are reported at a rapid rate. All classes of flavonoids derive their carbon skeleton from cinnamic acid; chalcone is the first common intermediate. More than 1 g of various flavonoids is ingested daily in the western diet. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments.)

Since most flavonoids in foodstuffs are β-glucosides, they are not hydrolyzed by intestinal digestive enzymes and pass largely unaltered into the lower bowel. Microflora in the large intestine hydrolyze the flavonoid glucosides to their corresponding aglycone, and cleave the heterocyclic pyrane ring. Pharmacokinetic investigations in humans showed that 3% of an ingested dose of flavonoids (catechin) is absorbed, with peak plasma levels of 17 mg ml^{-1} after 1 h and biexponential disappearance from plasma with mean half-lives of 8.8 min and 2.4 h.

The effect of flavonoids on thyroid hormone metabolism was discovered in studies of several substances of plant origin, chemically different from ascorbic acid, which have been reported to be beneficial in the control of hemorrhage, and have been referred to as 'vitamin P.' The ability of flavonoids to induce goiter was suggested from many studies involving various plants, e.g., peanuts (arachidoside) and millet (vitexin, glucosylorientine). Millet flavonoids are potent inhibitors of iodine organification, and flavonoids and thiocyanate interact synergistically on organifying enzyme TPO. By their structural homology with thyroid hormones, the flavonoids interact with thyroid hormone transport, by displacing the hormone from their natural transporters in human serum (thyroxine-binding prealbumin and thyroxine-binding globulin).

In human populations, pearl millet (*Pennisetum americanum*) is the staple food for millions of people and livestock in Africa and Asia. Because of its perennial growth, it constitutes the main source of food energy for the very poor in the semiarid tropics. Pennisetum millet is very rich in polyphenols of the flavonoid type which inhibit TPO. Table 4 compares the dietary composition and goiter prevalence among schoolchildren in the Darfur province of western Sudan. The goiter prevalence is closely linked to the level of consumption of millet.

Table 4 Dietary composition and goiter prevalence among schoolchildren in Darfur province, western Sudan

Locality	*Goiter prevalence (%)*	*Energy intake (%)*			
		Millet	*Sorghum*	*Wheat*	*Fat*
Kas	61	74	10	6	10
Tawaila	45	67	12	7	14
Nyala	12	37	27	20	16

Reproduced from Osman AK and Fateh (1981). Factors other than iodine deficiency contributing to the endemicity of goitre in Darfur province (Sudan). *Journal of Human Nutrition* 35: 302–310, with permission.

Selenium Deficiency-Related Thyroid Dyshormonogenesis

Selenium is incorporated as selenocysteine in all three forms of deiodinases by a specific transfer ribonucleic acid (tRNA) at the cotranslational level, with a same anticodon as the tRNA binding to the amber stop codon, TGA. SECYS-tRNA in place of STOP-tRNA is incorporated in case of a stem-and-loop mRNA structure (SECYS-mRNA element) present on the transcribed mRNA. At least for deiodinase 1, the activity is reduced to 10% of the initial level when selenium is not incorporated. The presence, but not its catalytic function, is also demonstrated by molecular biology in deiodinases 2 and 3.

In iodine- (Figure 4) and selenium-deficient subjects of northern Democratic Republic of Congo, selenium supplementation in the absence of iodine supplementation involves a decrease of serum T_4 and of serum reverse T_3 concentrations; in subjects with no loss of thyroid functional capacity (iodine-deficient, otherwise normal schoolchildren), the decrease of serum T_4 was not accompanied by an aggravation of hypothyroidism, as serum T_3 and serum TSH remained stable. In contrast, in subjects with a low thyroid functional capacity (myxedematous cretins), the decrease of serum T_4 was accompanied by an increase of serum TSH and a decrease of serum T_3, clearly showing that an aggravation of hypothyroidism occurred after selenium supplementation in the absence of sufficient iodine supply (Figure 5). According to these results, in areas which are both iodine- and selenium-deficient, it is mandatory to correct iodine deficiency before selenium supplementation; selenium deficiency should be corrected thereafter too.

In Tibetan villages with an elevated prevalence of endemic osteochondrodysplasia, named Kashin–Beck disease, and related to selenium deficiency, iodine deficiency and selenium deficiency are still more severe than in northern Democratic Republic of Congo. Nevertheless, endemic myxedematous cretinism is much less frequent in Tibet than in iodine-deficient areas of Central Africa. The current hypothesis is that at least three factors interact to involve endemic myxedematous cretinism: iodine deficiency + selenium deficiency + thiocyanate overload.

Recent experimental work has shown that selenium deficiency in rats subjected to antithyroid compounds involves thyroid fibrosis: this could be the

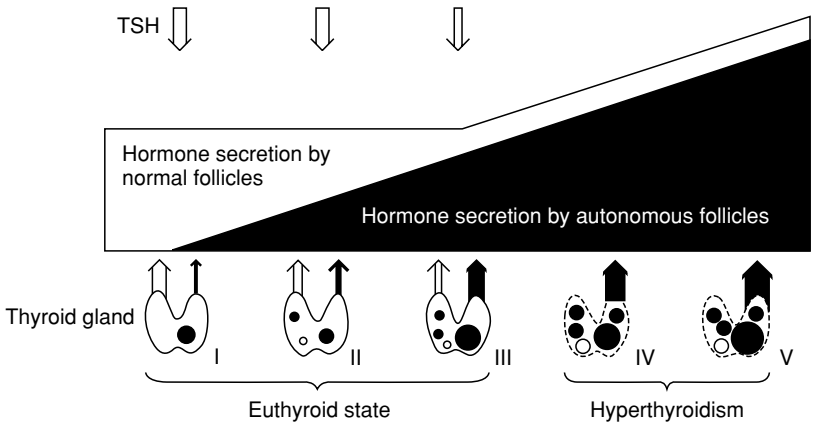

Figure 4 Proposed scheme of iodine-induced hyperthyroidism (also named Jod-Basedow hyperthyroidism). In an iodine-deficient environment, chronic thyroid-stimulating hormone (TSH) stimulation involves a multinodular goiter. This also promotes the development of some autonomous follicles. As long as the iodine substrate is not available, the hyperfunctioning gland remains silent, i.e., without concurrent clinical hyperthyroidism. Once iodine is introduced, for example, with universal iodized salt, some predisposed people (peculiarly women > 40 years) are at risk of developing hyperthyroidism. Reproduced from Braverman LE and Utiger R (eds) (1996) *Werner's and Ingbar's The Thyroid*, 7th edn. Philadelphia: JB Lippincott, with permission.

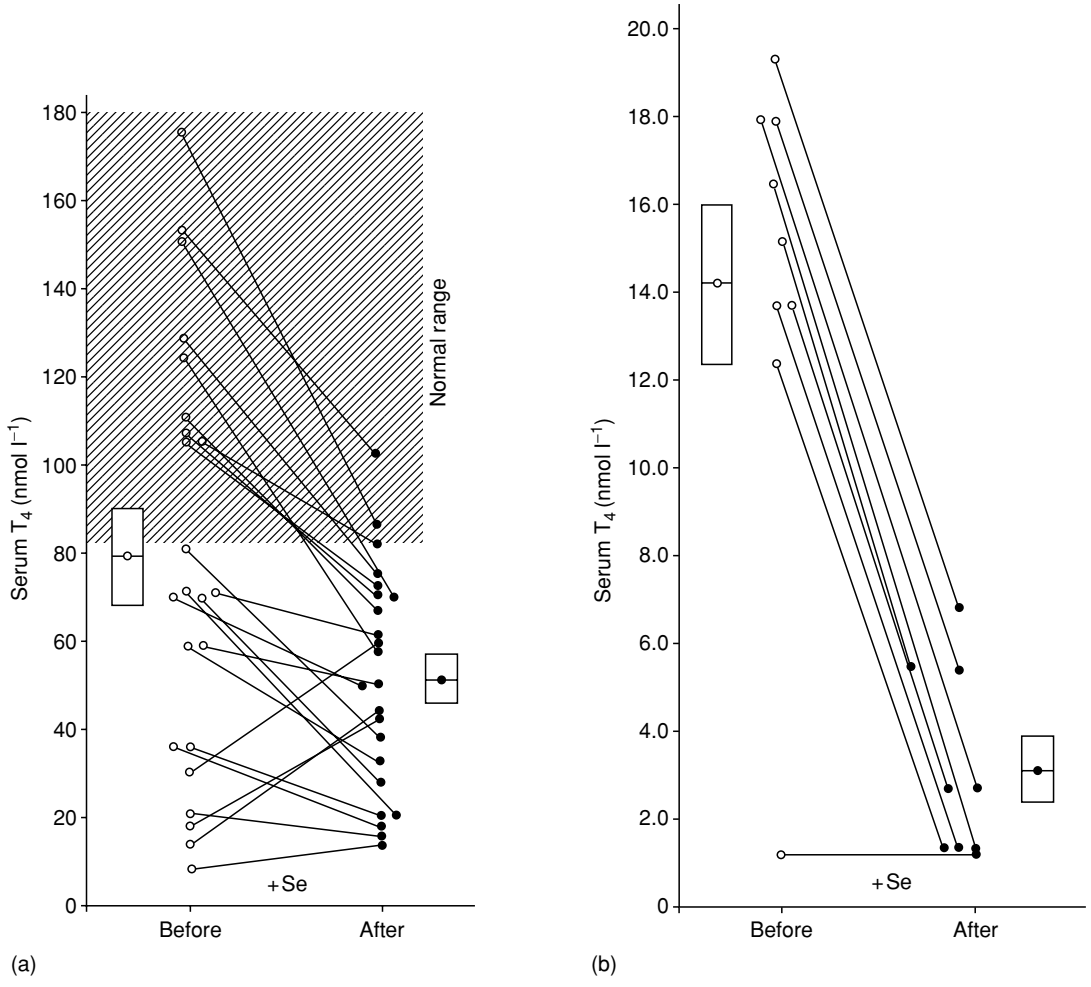

Figure 5 Effect of 2 months of oral selenium (Se) supplementation (50 μg Se daily as selenomethionine) in Se- and iodine-deficient subjects of Zaire on serum thyroxine (T_4) concentration. (a) Normal schoolchildren; (b) myxedematous cretins. Note the different scales of serum T_4. Reproduced from Goitrogens and Antithyroid Compounds, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

clue for the physiological mechanism of thyroid involution in endemic myxedematous cretinism.

Lithium- and Fluoride-Induced Thyroid Dysfunction

Lithium carbonate is used for manic-depressive disorders. Goiter occurs in 5–10% of treated patients, and biological hypothyroidism is reported in 5% of treated patients.

The best recognized effect of lithium on the thyroid is the inhibition of hormone release. The mechanism of action is still controversial and may involve the calcium/1,4,5-phosphatidylinositol pathway.

Lithium excess has been advocated as a cofactor of endemic goiter in Venezuela, where dietary iodine and protein intakes are low.

The consumption of ground water containing higher levels of fluoride has been reported to influence the prevalence of goiter. In 13 villages in the Himalayas, for almost the same iodine intake, the prevalence of goiter varied with different levels of fluoride and hardness of drinking water.

Phenol Derivatives

Humic substances are heterogeneous, high-molecular-weight, complex polymeric, yellow-black compounds which constitute the major organic fraction of soil and water. They are involved in the coalification process and are considered to be both precursors and decomposition products of coals. They include resorcinol, catechol, orcinol, phloroglucinol, pyrogallol, 2,6-dihydroxytoluene, 3,5-dihydroxybenzoic acids, and phtallic acids. These compounds are released into the water or soil, where decomposition occurs.

Resorcinol and hexylresorcinol are present in many topical antiseptic preparations; resorcinol has been used as an antifungal, antimicrobial, and keratolytic agent in concentrations of 2–20% in lotions, creams, and ointments. Hydroquinone and catechol are widely used in the photography industry as developers of black and white pictures or X-ray films, and in the rubber industry as antioxidants or as stabilizers of radical polymerization. Dihydroxyphenols are major components of cigarette smoke. They originate from decomposition of flavonoids.

A variety of phenolic compounds, articularly dihydroxyphenols, and related compounds with meta-substitution in a benzene ring, inhibit TPO-mediated iodide organification, *in vivo* as well as *in vitro*. (*See* **Phenolic Compounds**.)

Investigations into the causative factors in endemic goiter in iodide-sufficient areas of the Cauca valley, Colombia, and in the coal-rich Appalachian area of eastern Kentucky, USA, have demonstrated a significant correlation between watersheds rich in organic matter (coal and shale) and the high incidence of goiter.

Polycyclic Aromatic Hydrocarbons (PAHs), Including Dioxins

Polychlorinated biphenyls (PCBs) and polybrominated biphenyls (PBBs) are commonly incorporated into electrical apparatus, flame-proofing materials, hydraulic fluid, varnishes, and other types of protective coatings. These compounds are almost insoluble in water. (*See* **Polycyclic Aromatic Hydrocarbons**.)

Acute toxicity syndromes of PCB exposire in relatively large quantities have been described after an accidental industrial contamination in Japan in which over 1000 people were poisoned (Yusho disease). Alterations in thyroid function during this epidemic are not documented. In another study, 36 of 86 men who were employed in a PBB-manufacturing plant were reviewed for thyroid abnormalities; the prevalence of antimicrosomal antibodies (4 of 35) was statistically significantly higher than the 2.8% expected in a reference population. These data on the possible effect of goitrogens in thyroid autoimmunity are similar to those reported in rats submitted to other PAHs; goitrous lymphocytic thyroiditis has been observed after administration of the carcinogens, methylcholanthrene and dimethylbenzanthracene, in Buffalo inbred strains of rats, while other strains (Fisher rats) were not susceptible.

See also: **Cassava**: The Nature of the Tuber; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Hormones**: Thyroid Hormones; **Iodine**: Physiology; Iodine-deficiency Disorders; **Phenolic Compounds**; **Plant Antinutritional Factors**: Detoxification; **Polycyclic Aromatic Hydrocarbons**; **Smoking, Diet, and Health**

Further Reading

Arthur JR, Beckett GJ and Mitchell JH (1999) The interactions between selenium and iodine deficiencies in man and animals. *Nutrition Research Reviews* 12: 55–73.

Contempré B, Dumont JE, Ngo B *et al*. (1991) Effect of selenium supplementation in hypothyroid subjects of an iodine and selenium deficient area: the possible danger of indiscriminate supplementation of iodine-deficient subjects with selenium. *Journal of Clinical Endocrinology and Metabolism* 73/1: 213–215.

Day TK and Powell-Jackson PR (1972) Fluoride, water hardness and endemic goitre. *Lancet* 1: 1135–1138.

Ermans AM, Mbulamoko NM, Delange F and Ahluwalia R (1980) *Role of Cassava in the Etiology of Endemic*

Goitre and Cretinism. Ottawa: International Development Research Centre.
Gaitan E (1989) *Environmental Goitrogenesis*. Boca Raton, Florida: CRC Press.
Gembicki M, Stozharov AN, Arinchin AN *et al*. (1997) Iodine deficiency in Belarussian children as a possible factor stimulating the irradiation of the thyroid gland during the Chernobyl catastrophe. *Environmental Health Perspectives* 105 (suppl. 6): 1487–1490.
Joint WHO/UNICEF/ICCIDD Consultation (1997) Review of Findings from Seven-Country Study in Africa on Levels of Salt Iodization in Relation to Iodine Deficiency Disorders, Including Iodine-Induced Hyperthyroidism. Geneva: World Health Organization.
Köhrle J, Spanka M and Hesh RD (1988) Flavonoid effects on transport, metabolism and action of thyroid hormones. In: Cody V, Middleteon E, Harborne JB and Beretz A (eds) *Plant Flavonoids in Biology and Medicine. Biochemical, Cellular and Medicinal Properties*, pp. 323–340. New York: Liss.
Moreno-Reyes R, Suetens C, Mathieu F *et al*. (1998) Iodine and selenium deficiency in Kashin-Beck disease (Tibet). *New England Journal of Medicine* 339: 1112–1120.
Ngo B, Dikassa L, Okitolonda W *et al*. (1997) Selenium status in pregnant women of a rural population (Zaire) in relationship to iodine deficiency. *Tropical Medicine and International Health* 2: 572–585.
Utiger RD (1995) Cigarette smoking and the thyroid. *New England Journal of Medicine* 333: 1001–1002.

Gooseberries *See* **Currants and Gooseberries**

Grains *See* **Cereals**: Contribution to the Diet; Bulk Storage of Grain; Handling of Grain for Storage; Breakfast Cereals; Dietary Importance; **Wheat**: The Crop; Grain Structure of Wheat and Wheat-based Products

Grapefruits *See* **Citrus Fruits**: Types on the Market; Composition and Characterization; Oranges; Processed and Derived Products of Oranges; Lemons; Grapefruits; Limes

GRAPES

R S Jackson, Bic, QC, Canada

Background

Grapes are one of the most important of the fleshy edible crops. They are produced in clusters on members of the genus *Vitis*. Most of the annual production goes into producing wine, with smaller amounts used as fresh fruit or raisins. Because of unique cultural demands, particular varieties are typically grown for a single purpose. Intended use also has influenced the direction of domestication and breeding of grape cultivars.

Commercial Importance of Grapes and Wine

Global grape production averages about 60 million tonnes. By comparison, orange, banana, and apple production averages about 57, 50, and 43 million tonnes. Of this, about 70% of the crop is fermented into wine, 27% consumed as fresh fruit, and 2% dried for raisin production. Regional use varies widely, depending on the climate, but also on religious dictates of the producing or importing country (**Tables 1** and **2**).

Grape production is based primarily on cultivars of *Vitis vinifera*. Cultivation is also largely restricted to a zone between the annual 10 and 20 °C isotherms (**Figure 1**). In addition, grape culture is principally

Table 1 World regional statistics for vineyard coverage and total grape, wine, table grape, and raisin production in 1998[a]

Region	*Vineyard area* (10^3 *ha*)	*%*[b]	*Total grape production* (10^6 *kg*)	*%*	*Wine production* (10^6 *l*)	*%*	*Table grapes* (10^6 *kg*)	*%*	*Raisins* (10^6 *kg*)	*%*
Africa	314	4.0	2 873	4.7	931	3.6	1 398	10.6	27	2.8
Americas	859	11.0	10 391	17.2	4 421	17.1	1 871	14.3	302	31.4
Asia	1 478	19.0	11 891	19.6	612	2.4	6 833	52.1	499	51.9
Europe	5 041	64.6	30 633	50.6	19 112	73.8	2 951	22.5	95	6.8
Oceania	107	1.4	1 176	1.9	802	3.1	66	0.5	38	3.9
Total	7 799		60 585		25 878		13 119		961	

[a]Data from Dutruc-Rosset G (2000) The state of vitiviniculture in the world and the statistical information in 1998. *Bulletin de l'Office International de la Vigne et du Vin* 73: 1–94.
[b]Percentage of world production.

Table 2 Top ten countries in vineyard coverage, total grape, wine, table grape, and raisin production in 1998[a]

Country	*Vineyard area* (10^3 *ha*) *%*[b]		*Country*	*Total grape production* (10^6 *kg*) *%*		*Country*	*Wine production* (10^6 *l*)	*%*	*Country*	*Table grapes* (10^6 *kg*) *%*		*Country*	*Raisins* (10^6 *kg*) *%*	
Spain	1180	15.1	Italy	9208	16.1	Italy	5418	20.9	Turkey	1833	14.0	Turkey	350	36.4
France	914	11.7	France	6877	12.1	France	5267	20.4	China	1792	13.7	USA	255	26.5
Italy	889	11.4	USA	5355	9.4	Spain	3032	11.7	Iran	1724	13.1	Iran	90	9.4
Turkey	602	7.7	Spain	4884	8.6	USA	2045	7.9	Italy	1294	9.9	Greece	87	9.1
USA	364	4.7	Turkey	3650	6.4	Argentina	1267	4.9	Egypt	858	6.5	Australia	38	4.0
Iran	270	3.5	China	2358	4.1	Germany	1083	4.2	India	630	4.8	Chile	35	3.6
Portugal	260	3.3	Iran	2315	4.1	South Africa	816	3.2	Chile	599	4.6	Afghanistan	28	2.9
Romania	253	3.2	Argentina	2002	3.5	Australia	742	2.9	USA	569	4.3	South Africa	25	2.6
Argentina	210	2.7	Germany	1408	2.5	Chile	548	2.1	Syria	413	3.2	Syria	16	1.7
China	194	2.5	South Africa	1300	2.3	Romania	500	1.9	Brazil	398	3.0	Argentina	9	0.9

[a]Data from Dutruc-Rosset G (2000) The state of vitiviniculture in the world and the statistical information in 1998. *Bulletin de l'Office International de la Vigne et du Vin* 73: 1–94.
[b]Percentage of world production.

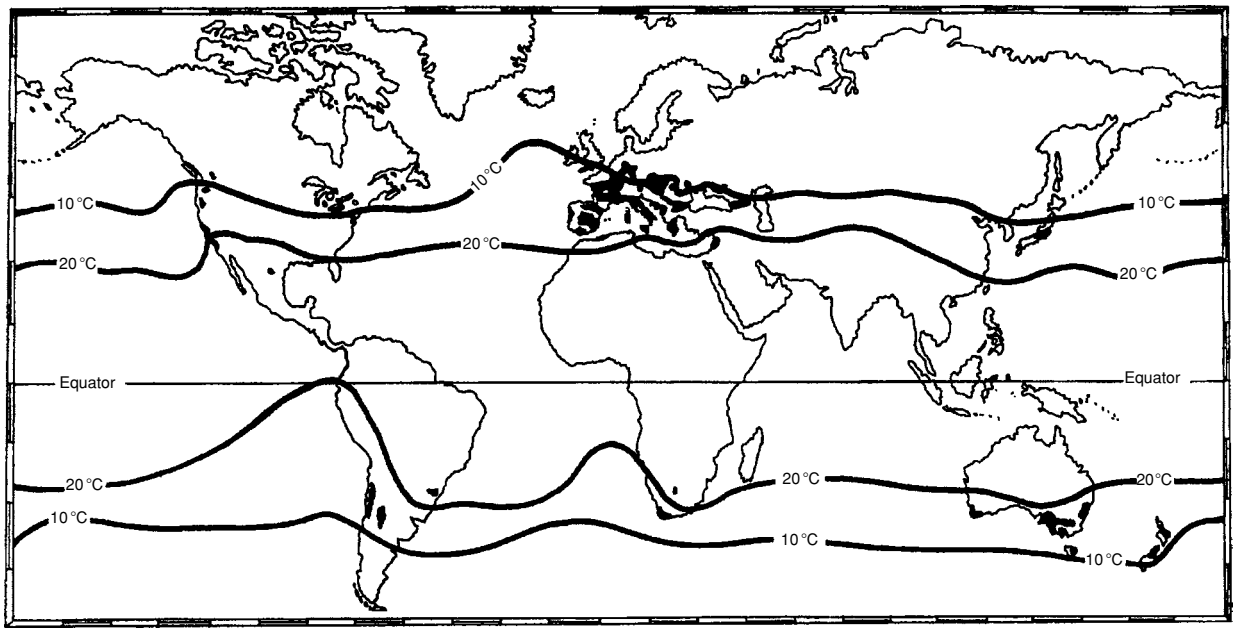

Figure 1 Association between the major grape-producing regions of the globe, with the 10 and 20 °C annual isotherms. Major grape-producing regions are shaded black. Drawing courtesy of H. Casteleyn, reproduced by permission.

located in regions characterized by a Mediterranean climate. Cultivation in cooler, warmer, or more moist regions is commercially feasible only when local climatic conditions or cultural practice compensate for less than ideal conditions.

Approximately 65% of the world's vineyards occur in Mediterranean Europe. Of this proportion, about 60% are located in Spain, Italy, and France. Although the coverage has declined by some 16% since the mid-1980s, grape production has actually increased by about 10%.

Grape Species and Cultivars

Grapes are produced by members of the genus *Vitis* (family Vitaceae). All members are viny and develop swollen or jointed nodes from which leaves, tendrils, or flower clusters form. Each cluster produces many, miniature, nonshowy, bi- or unisexual flowers. These develop into fleshy berries containing up to four seeds. The fruit is normally dark blue–purple owing to the production of anthocyanin pigments. Green to yellow-colored varieties do not produce these pigments.

The genus is divided into two sections or subgenera. The subgenus *Vitis* (bunch grapes) is the larger, containing all species (about 65) except *V. rotundifolia* and *V. popenoei*. These two are placed in the subgenus *Muscadinia* (muscadine grapes). Muscadine grapes are native only to coastal southeastern North America.

The origin of the genus probably involved several interspecies crosses, associated with subsequent chromosome doubling. Modern species are derived from very ancient hexaploids (plants possessing sets of chromosomes from three different species). Bunch grapes (subgenus *Vitis*) possess 38 chromosomes, whereas muscadine grapes (subgenus *Muscadinia*) have 40 chromosomes.

Domestication of *Vitis vinifera* probably started in northeastern Iran and northern Turkey some 4000 years ago. In contrast, the earliest evidence of wine production predates domestication by about two millennia. Archeologic evidence of grapevine domestication appears as changes in seed shape (length relative to width) and an increase in bisexuality (reduction in the presence of infertile pollen from female vines).

Most present-day grape varieties appear to have developed locally, possibly from the crossing of indigenous vines and introduced domesticated varieties. One view of the diffusion routes of the introduction of grape culture to Europe is given in **Figure 2**.

There are more than 5000 named grapevine cultivars. Most are of local interest only, with few having achieved world-wide distribution. Examples are 'Cabernet Sauvignon,' 'Syrah,' 'Pinot noir,' 'Tempranillo' and 'Sangiovese' (red), and 'Chardonnay,' 'Riesling,' and 'Sauvignon blanc' (white). Their productivity is usually lower, and their cultivation typically more demanding. Their reputation comes from their yielding fine, varietally distinctive wines with long-aging potential. In favorable locations, their excellent wine-making properties compensate for reduced yield and increased production costs. Modern cultivars are often crosses between North American *Vitis* species and *V. vinifera* varieties. Examples are 'Concord,' 'Chambourcin,' and 'Cayuga'. Crossing with North American and European grapes has also been crucial in the production of rootstocks that provide resistance or tolerance to many important grapevine pest, diseases, and complicating soil conditions.

Vine Cycle and Vineyard Activity

The end of one growth cycle and the beginning of another typically coincide with the onset of winter. The absence of foliage facilitates the removal of unessential shoot growth and the selection of growth from which to generate the subsequent year's crop.

The first indicator of renewed grapevine activity in the spring is the flow of sap from the ends of pruned shoots. When the average temperature rises above a critical value, buds begin to burst. Once initiated, development and enlargement of the embryonic leaves, tendrils, and flower clusters occur quickly (**Figure 3**). Disease, pest, and weed control measures begin early.

Flower development progresses from the outermost flower parts inward. Maturation of the anthers and anthesis (pollen release) coincides with separation of the fused cap of petals (calyptra) (**Figure 4**). The separation shakes pollen on to the stigma, thus initiating self-fertilization. Flower induction for the subsequent season occurs at this time in newly developing buds. Following fertilization, fruit development begins. Except for seedless varieties, one seed is required for berry development.

Once the shoots have elongated, they are typically tied to a trellis for support. Irrigation, if required, usually ceases at *véraison* (the beginning of mature fruit coloration). This restricts additional vegetative growth. Removing excessive leaf or shoot growth may be conducted to limit shading and a nutrient drain that can retard fruit ripening. Irrigation may be reinitiated after harvest to avoid stressing the vine and permit optimal shoot maturation.

As the grapes near maturity, vineyard activity shifts toward preparing for harvest. Based on chemical analysis of the fruit, environmental conditions, and the desires of the winemaker, a harvest date is set. In cool climates, frost protection measures are set in place.

Once harvest is complete, vineyard activity turns to preparing the vines for winter. In cold climates, this

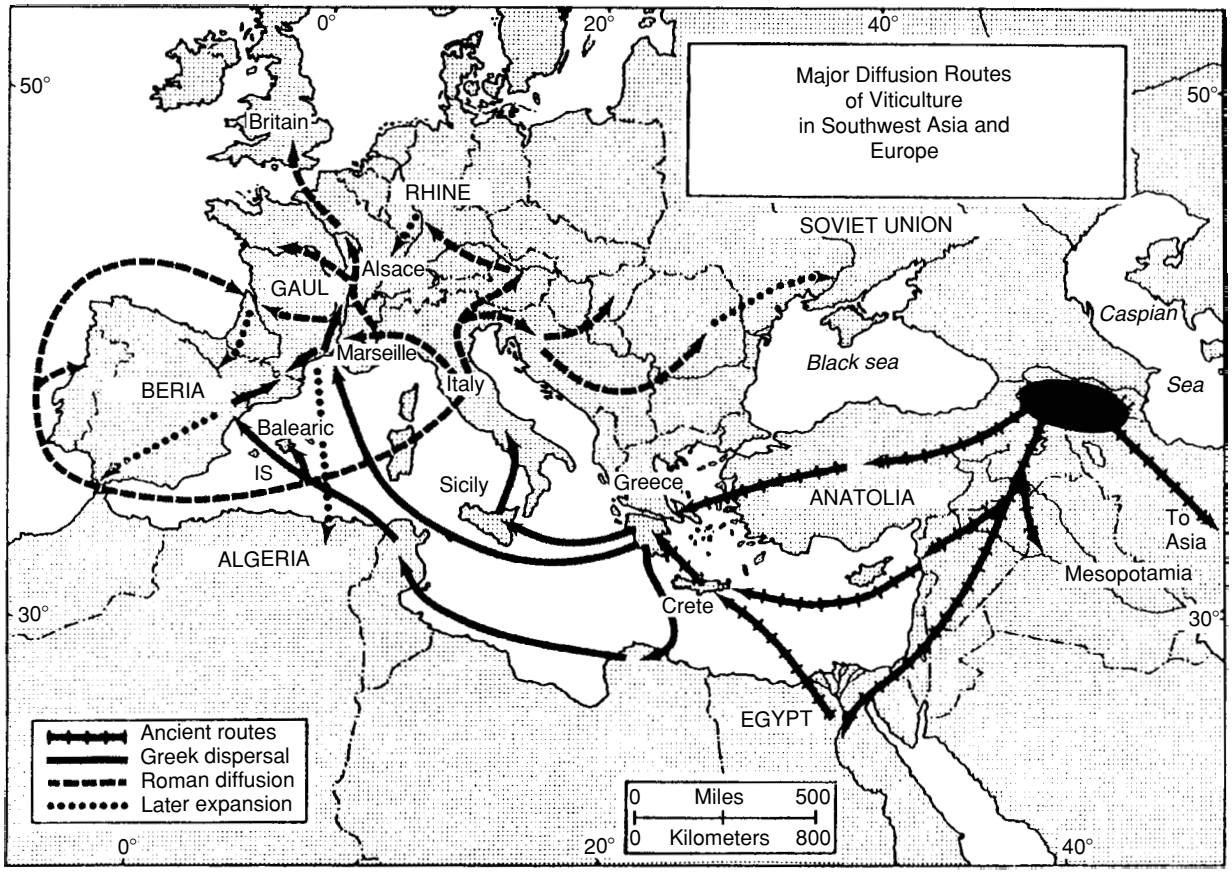

Figure 2 Major diffusion routes of grape growing in southwest Asia and Europe. From de Blij HJ (1983) Wine. A Geographic Appreciation. Rowman and Allanheld, Totowa, NJ, reproduced by permission.

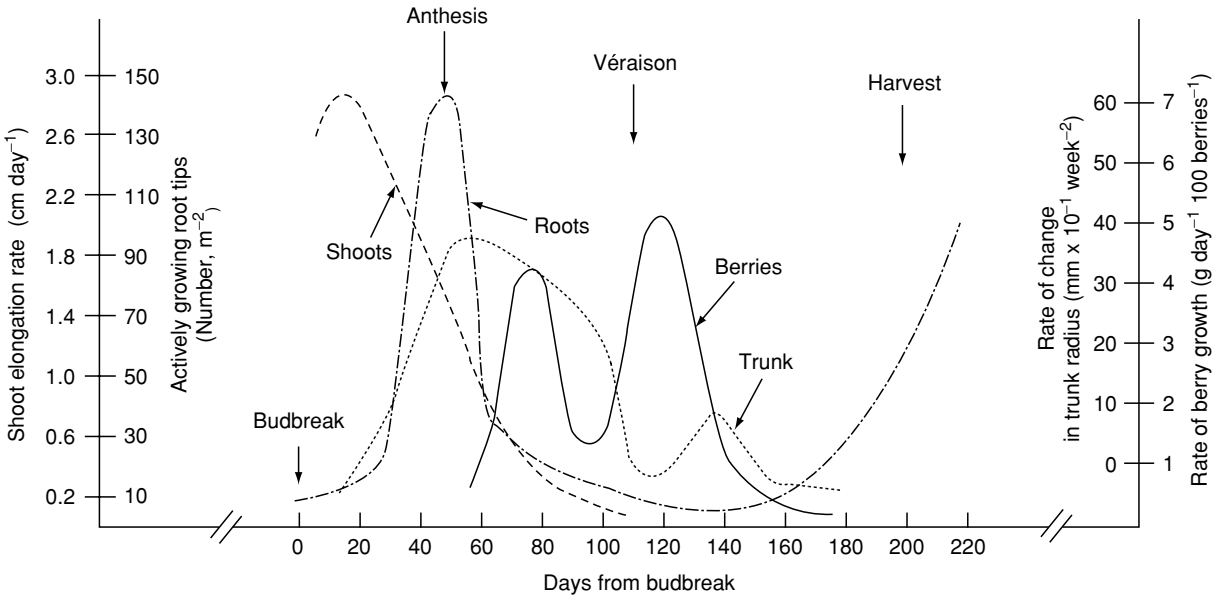

Figure 3 Growth rate of various organs of 'Colombar' grapevines grown in South Africa throughout the season. After van Zyl (1984), from Williams LE and Matthews MA (1990) Grapevine. In: Stewart BA and Nielsen DR (eds), *Irrigation of Agronomic Crops, Agronomy Monograph # 30*, pp. 1019–1055. Madison, WI: American Society of Agronomy, Crop Science Society of America and Soil Science Society of America, with permission.

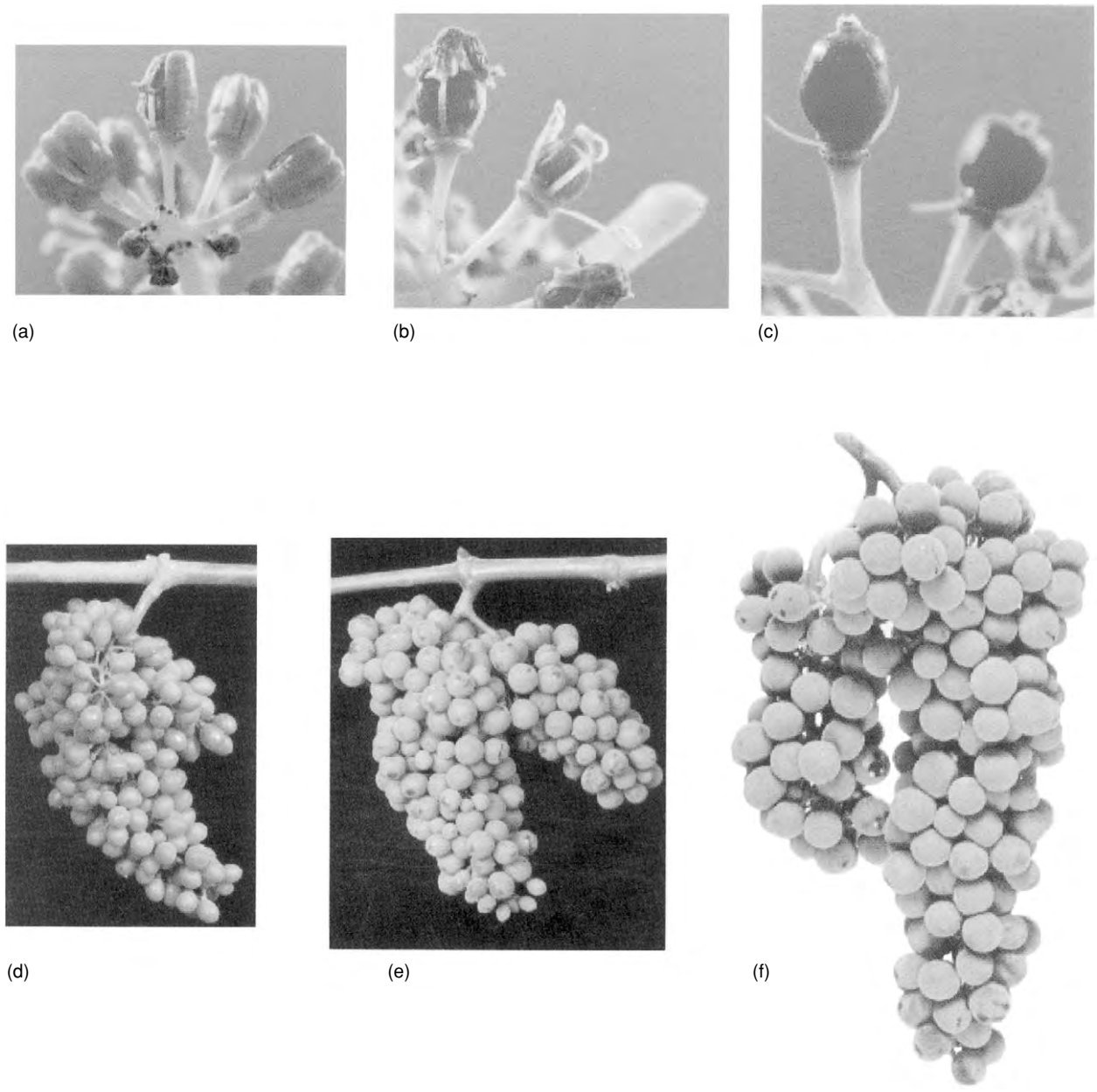

Figure 4 (see color plate 79) Stages in the flowering and fruit development of grapes: A, flowers with the fused petals (calyptra) about to be shed; B, calyptra in last stage of separation and anthers about to shed pollen; C, very young grape still showing filaments of anthers; D, young cluster of green grapes; E, maturing grapes beginning to change color (*vèraison*), the bloom on the berry surface is evident; F, fully mature grapes of 'Cabernet Sauvignon.' From Flaherty DL, Christensen LP, Lanini WT *et al*. (1992) *Grape Pest Management, Publication No. 3343*, 2nd edn. Oakland, CA: University of California, with permission.

may vary from mounding soil up around the shoot–rootstock union to removing the whole shoot system from its support system and burial with soil.

Grape Structure and Development

Fruit growth typically shows an S-shaped curve (Figure 5). In the initial phase, lasting 6–8 weeks, cells divide rapidly, followed by enlargement. In the second phase, lasting 1–6 weeks, berry growth slows, and the seeds develop. At the end of this period, the skin begins to lose its green color (*véraison* – Figure 4). In the final phase, the fruit enlarges to its mature size, tissues soften, acidity falls, sugars accumulate, anthocyanin pigments form (in blue-skinned varieties), and aroma compounds are synthesized.

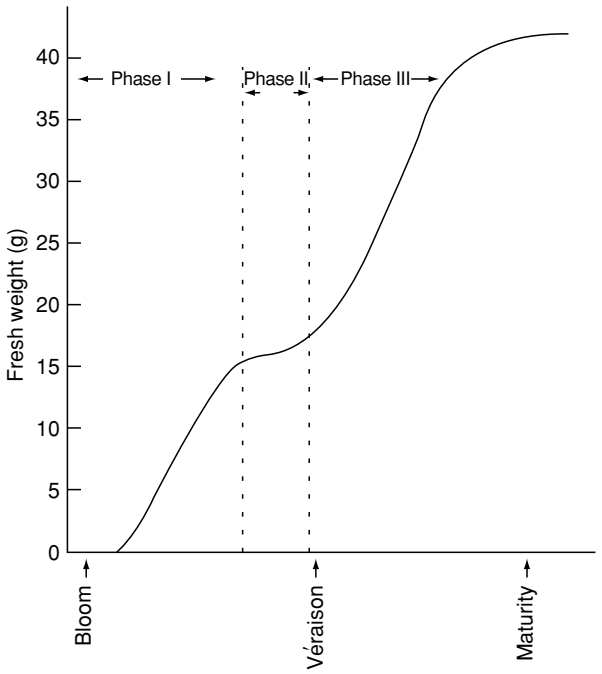

Figure 5 Diagrammatic representation of the growth phases of grapes.

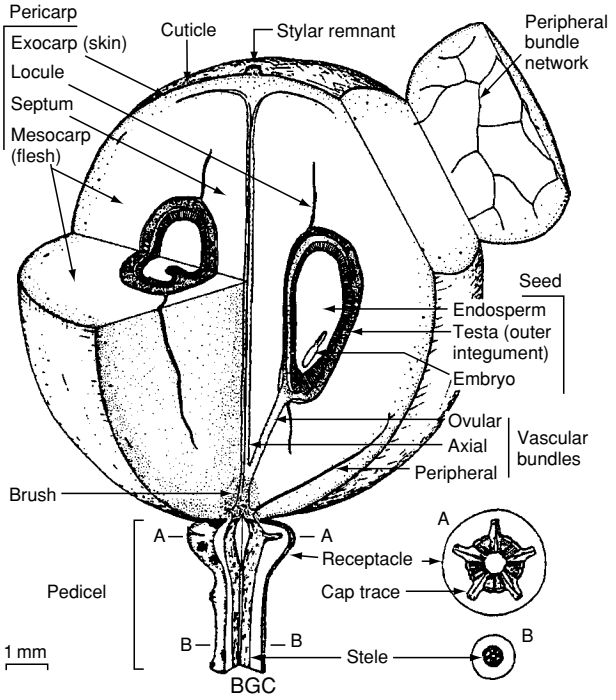

Figure 6 Diagrammatic representation of a grape berry. After Coombe BG (1987) Distribution of solutes within the developing grape berry in relation to its morphology. *American Journal of Enology and Viticulture* 38: 120–128, published by the American Society for Enology and Viticulture, with permission.

The final stage lasts 5–8 weeks. The duration of each phase is cultivar- and climate-controlled.

Mature grapes possess two morphologic regions – the skin and the flesh (Figure 6). The thin skin layer consists of an outer, single-layered epidermis and an inner hypodermis (four to 20 cells thick). The hazy coating (bloom) seen on ripe grapes comes from waxy plates and ridges of the cuticle that cover the epidermis (not yeast cells as commonly believed). Anthocyanins produced in the outer layers of the hypodermis generate the blue–purple color typical of most grapes. White grapes derive their yellowish coloration from the carotenoids and flavonoid pigments in the hypodermis. The hypodermis is also the production site for most aroma compounds.

During ripening, the cell wall of the hypodermis thickens as a result of hydration, loosening bonding between the cells. In 'slip-skin' varieties, however, a layer of cells between the flesh and hypodermis become thinner and lose much of their pectinaceous cell-wall material. This produces a zone of weakness, allowing the skin to separate readily from the flesh.

The internal fleshy tissues of the grape are subdivided into two ill-defined regions – one between the hypodermis and the peripheral vascular strands, and the innermost, being delimited by the peripheral and axial vascular strands (Figure 6).

Grape size is largely dependent on enlargement of the central vacuole of flesh cells. The vacuole is also the primary site for sugar and acid accumulation during ripening.

Chemical composition (see Table 3)

Sugars

Grape chemistry changes most significantly during the later phases of ripening, beginning with *véraison*. This is particularly marked as sugars, translocated from adjacent leaves, accumulate. A simultaneous reduction in vegetative growth, and a metabolic shift to malic acid respiration, also favor sugar deposition.

On reaching the berry, sucrose is hydrolyzed and stored as fructose and glucose. Depending on the cultivar and prevailing conditions, the sugar content may reach 12–28%. For wine-making, the desired sugar content generally ranges from about 21 to 25%. Lower sugar contents are preferred for table grapes. Small, but largely insignificant, amounts of other sugars accrue in ripe grapes – notably raffinose, stachyose, melibiose, maltose, galactose, arabinose, and xylose.

Acids

After sugar, the most marked chemical modification during ripening involves a reduction in acidity.

Table 3 Major chemical constituents in grapes (% fresh weight)[a]

Constituent	*Range*
Water	70–85
Carbohydrates	15–25
Glucose	8–13
Fructose	7–12
Pentoses	0.01–0.05
Pectins	0.01–0.1
Organic acids	0.3–1.5
Tartaric	0.2–1.0
Malic	0.1–0.8
Citric	0.01–0.05
Acetic	0.00–0.02
Phenolics	0.01–0.15
Anthocyanins	0.00–0.05
Tannins	0.01–0.1
Nitrogenous compounds	0.03–0.17
Amino acids	0.01–0.08
Ammonia	0.001–0.012
Minerals	0.3–0.5

[a]Data from Amerine MA, Berg HW, Kunkee RE *et al*. (1980) *The Technology of Wine Making*, 4th edn. Westport, CT: AVI.

Tartaric and malic acids account for about 70–90% of the berry acid content. The remainder consists of variable amounts of organic acids (citric and succinic acids), phenolic acids (e.g., quinic and shikimic acids), amino acids, and fatty acids.

The tartaric acid content of grapes generally remains stable during ripening. Its perceived acidity declines, though, as the proportion of free acid (not associated with potassium and calcium ions) decreases. In contrast, the malic acid content decreases during ripening. This trend is particularly marked in hot climates. Shoot vigor is another factor influencing fruit acidity, partially by increasing fruit shading and cooling.

Phenolics

Phenolics constitute the third most significant group of organic compounds found in grapes. Phenolics not only donate the color to red wines, but also give them their characteristic taste and aging properties. During ripening, the anthocyanin content of red grapes both increases and changes in composition. Because the final distribution can significantly influence the hue, intensity, and color stability of a wine, these changes are of great importance to wine quality. White grapes have a lower total phenolic content and do not synthesize anthocyanins. Aside from seed phenolics, both red and white grapes synthesize most of their phenolic constituents in the skin.

In addition to anthocyanins and tannins, red-skinned cultivars contain flavonols, various benzoic acid and cinnamic acid derivatives, and hydroxycinnamic acid esters of tartaric acid. They are commonly glycosidically linked to glucose, rhamnose, or glucuronic acid.

A small group of phenolic compounds, the stilbene phytoalexins, has attracted considerable attention recently. The major phytoalexin in grapes – resveratrol – is credited with many of the health benefits of moderate wine consumption.

Pectins

Pectins function as the primary glue holding grape cells together. Their retention at the end of ripening provides table grapes with their traditional texture. In contrast, pectin hydrolysis is critical to the softening of wine grapes. Softening facilitates juice and flavor release during grape crushing and minimizes the likelihood of haze production in wines.

Lipids

The lipid component of grapes consists primarily of the surface waxes (mostly oleanolic acid) and cutin fatty acids, membrane phospho- and glycolipids, and seed oils. Of these, the most significant are the long-chain fatty acids on the surface of the skin. These frequently act as important sources of unsaturated fatty acids during wine fermentation. Although they occur in minute amounts, oxidized carotenoid derivatives from grapes generate important aromatic compounds. Examples are damascenone, β-ionone, as well as several norisoprenoids.

Nitrogen-containing Compounds

The concentration of nitrogen-containing compounds in grapes is relatively low and has little direct effect on taste or flavor. Nevertheless, one of the most common amino acids in grapes (proline) may play a significant role in limiting the cytoplasmic damage caused by the rapid drop in osmotic potential associated with the storage of sugar during ripening. Also important, but to wine quality, is the most common peptide in ripe grapes, glutathione, which functions as an important antioxidant early in wine fermentation.

Aromatic Compounds

Compounds that give particular grape varieties their distinctive aroma usually occur in trace amounts, occasionally existing as a few parts per billion. Typically, they accumulate as the grapes ripen. As they collect, an increasing fraction may become bound in nonvolatile complexes, such as glycosides, polymers, or oxidized derivatives. This is particularly noticeable with flavorants such as monoterpenes, norisoprenoids, and nonflavonoid phenolics. In

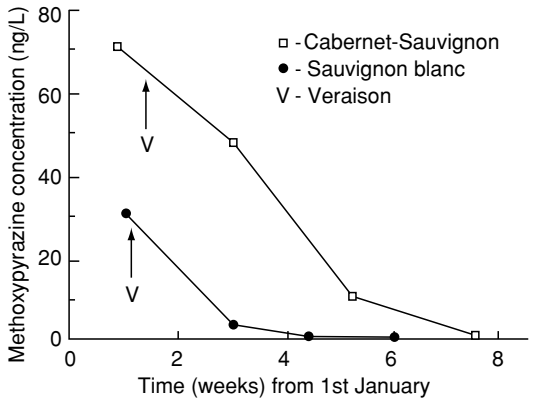

Figure 7 Decline in the concentration of grape methoxypyrazine with ripening. From Allen MS, Lacey MJ and Boyd SJ (1996) Methoxypyrazines of grapes and wines. Differences of origin and behaviour. In: Stockley CS, Sas AN, Johnson RS and Lee TH (eds), *Proceedings of the 9th Australian Wine Industry and Technology Conference, Adelaide, 16–19 July 1996*, pp. 83–86. Adelaide: Winetitles, with permission.

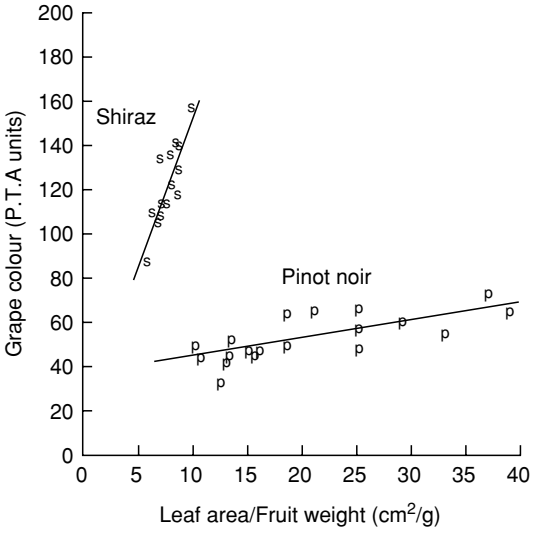

Figure 8 Relationship between grape color and leaf area/fruit weight ratio for 'Shiraz' and 'Pinot noir.' From Hand PG, Gawel R, Coombe BG, and Henschke PM (1993) Viticultural parameters for sustaining wine style. In: Stockley CS, Sas AN, Johnson RS, Leske PA and Lee TH (eds), *Proceedings of the 8th Australian Wine Industry and Technology Conference, Adelaide, 16–19 July 1996*, pp. 167–169. Adelaide: Winetitles, with permission.

contrast, the concentration of a few impact compounds declines near maturity (Figure 7).

Most modern research confirms the belief that most flavorants collect in the skin. Nonetheless, important aroma compounds such as linalool and TDN (1,1,6-trimethyl-1,2-dihydronaphthalene) are stored in the flesh.

For several grape varieties, varietal distinctiveness depends on the synthesis of one or a few related compounds. Examples are the Cabernet group of varieties that produce methoxypyrazines and Muscat cultivars that produce monoterpenes. In other cultivars, though, varietal distinctiveness seems to come from the unique combination of many standard grape aromatics. Examples are 'Chardonnay' and 'Pinot noir.'

Cultivation

Training and Pruning

Vineyard practice aims at obtaining the maximum yield of fully ripened fruit, consistent with long-term vine health. One of the major means of achieving this goal involves training and pruning. Training usually involves a support (trellis). The specific form chosen depends on factors such as the prevailing climate, harvesting practices, and the fruiting characteristics of the cultivar. Associated with training is the positioning of fruit-bearing shoots in a microclimate optimal for grape quality, inflorescence initiation, and shoot maturation. Pruning may involve the selective removal of growing shoots, mature wood, leaves, or excess numbers of flower or fruit clusters. However, pruning typically refers to the removal of excess shoot growth of the previous season during the winter or early spring.

Central to effective grape culture is maintaining just sufficient leaf area (for photosynthesis) to adequately ripen the crop. This leaf area/fruit (LA/F) relationship is related to the more familiar concept of yield per hectare. It is more direct, though, by focusing attention on the fundamental relationship between the supply and demand for photosynthetic energy. However, because of the complex, adaptive nature of the grapevine, excess leaf area is often as detrimental to grape quality as is insufficient foliage. Excessive leaf area tends to promote extended shoot growth, causing undesirable fruit shading and directing nutrients away from the maturing grapes. Thus, it is essential to neither over- nor underprune grapevines. The appropriate level tends to be in the range of $10\,cm^2$ of leaf surface per gram of fruit. However, the value can vary considerably, depending not only on the cultivar (Figure 8), but also on the training system, soil nutrients, water supply, and climatic conditions. The target is to establish the optimal foliage canopy and placement to nourish the crop to its ideal state of ripeness.

On relatively nutrient-poor, dry soils, the vine's inherent vigor must be restrained. This can be achieved by dense vine planting (about 4000 vines ha^{-1}) and severe pruning (removal of > 90% of the

yearly shoot growth). However, on rich moist loamy soils, wide spacing of the vines (about 1500 vines ha^{-1}) with limited pruning is preferable. Under such conditions, the improved growth potential of the vine must be directed into increased yield, not undesirable vegetative growth. When an appropriate LA/F ratio is developed, both fruit yield and grape quality increase. Examples of training systems ideal for rich soils are the Scott Henry trellis, the Smart-Dyson trellis, and the Lyre. These systems not only achieve a desirable LA/F ratio, but also provide better exposure of the grapes to light and air. These features promote ideal berry coloration, flavor development, fruit health, and flower-cluster initiation.

When considering pruning, many features need to be considered, notably the prevailing climate, genetic characteristics of the rootstock and bearing cultivars, soil fertility, and training system. Of the forms available, most pruning systems can be categorized as either spur or cane pruning.

Spur pruning refers to the retention of only a short shoot segment containing about two buds. It is easier to perform, being effectively conducted by machine. In addition, spur pruning restricts fruit production to predetermined locations, making it particularly applicable to mechanical harvesting. The tendency of spur pruning to limit productivity can be either beneficial or detrimental, depending on the vigor and bearing capacity of the vine. Berry size is generally reduced with spur pruning. This has the advantage of increasing the grape surface area/volume ratio and, thus, potentially enhancing a wine's flavor and color.

Cane pruning involves retaining fewer but longer shoots. These may contain upward of 20 or more buds. Cane pruning is generally more popular in cooler, more moist climates and with cultivars bearing small fruit clusters. However, it must be conducted by skilled workers and demands more shoot support than is typical with spur pruning.

Commonly accepted norms for the number of shoots per meter of vine canopy fall in the range of about 15 shoots per meter (positioned more or less equidistantly along the vine). This assumes an average of about 12–18 buds per shoot.

Grafting

Most grapevines in commercial culture are grafted to a rootstock. In much of the world, grafting is essential to obtain resistance from a particularly devastating root-infecting pest, the phylloxera root louse. In the late 1800s, it decimated European grape culture until grafting was introduced. In other areas, rootstocks are required to achieve adequate protection against nematodes, viruses, or saline soil conditions. Certain rootstocks can also limit grapevine vigor in highly fertile, moist soils.

Irrigation and Soil Fertilization

Irrigation is outlawed in most European countries, on the mistaken belief that it is inherently detrimental to grape quality. However, when used wisely, irrigation not only permits grape growing in semiarid to arid regions, but also enhances grape quality. In a process called partial rootzone drying, selectively limited water availability limits vegetative growth and favors optimal fruit ripening. Irrigation can also be used to simultaneously apply fertilizer and pest-control chemicals to vines.

In comparison with other crops, grapevines have relatively limited nutrient demands. This, combined with the accumulation of nutrients in the woody parts of the plant, makes assessing grapevine response to fertilizer application particularly difficult. The primary method of nutrient analysis is based on sampling plant tissue. Optimal fertilizer addition is roughly equivalent to vine requirements minus the amount already available in the soil. Excessive application seldom enhances vine growth or fruit quality, and can be detrimental to both the vine and the environment.

Of inorganic nutrients, only nitrogen is often in marginally short supply. The other two major inorganic nutrients, potassium and phosphorus, are seldom limiting. Vineyard soils are rarely deficient in required micronutrients.

Organic fertilizers, such as compost or manure, have many advantages over inorganic chemical fertilizers. These include improved soil texture and a more even nutrient supply. However, the higher cost and limited availability of organic fertilizers restrict their more extensive use in commercial grape growing.

Improving Table Grape Quality

Grape quality is primarily achieved by the application of the best cultural techniques, under climatic and soil conditions ideally suited to the particular cultivar. However, table grape production requires several practices unique to their cultivation. These include training systems that permit the grapes to hang free, facilitate manual picking, and protect against sunburn, as well as the use of girdling and the application of growth regulators. Of these, the most unique to table grapes are girdling and the use of growth regulators.

Girdling typically involves the removal of a complete strip of bark from trunks or bearing canes below the clusters. This increases the supply of sugars to the fruit by stopping their translocation below the cut. With varieties that set poorly or are seedless, girdling shortly

after pollination increases the number of grapes that develop. Girdling also increases fruit size in seedless cultivars. With table grapes, girdling at *véraison* can improve fruit color and hasten maturation.

Some of the advantages achieved by girdling can be enhanced or replaced by the application of specific growth regulators. Gibberellin is of particular use in seedless cultivars, where it increases grape size. In addition, gibberellin lengthens the stalks of fruit clusters, improving spacing between the grapes. Improved pigmentation can be achieved in colored varieties by the application of ethephon.

Harvesting

Choosing the harvest date is probably the single most important viticultural decision taken yearly by the grower. It sets a limit on the qualities of the grapes. Timing usually is based on one or more chemical analyses of the fruit. In cool climates, sugar content is the primary indicator of maturity, whereas in warmer temperate climates, the sugar/acid ratio is often preferred. In either case, the accumulation of sugar often correlates well with anthocyanin and flavorant build-up in the fruit. Where the correlation is poor, direct measurement of the anthocyanin or flavorant content may be a better indicator of desirable maturity (Figure 8). A new indicator of maturity under investigation measures the glycosyl glucose content. Because most grape flavors are glycosidically bound, assessment of the glycosyl glucose content may provide a direct indicator of flavor potential.

In addition to chemical indicators of maturity, environmental factors that can reduce quality must also be considered in choosing a harvest date. The detrimental effects of early frosts and protracted rainy periods on fruit quality are well known.

Until the development of mechanical harvesters in the late 1960s, all grapes were harvested manually. Except for table grapes and premium wine grapes, labor shortages and market forces have combined to favor machine harvesting. Mechanical harvesters not only reduce harvesting costs (by up to 75%), but also permit rapid harvesting. In comparative tests, no clear sensory difference has been detected between wines produced from hand- and machine-harvested grapes. The main mechanisms used by mechanical harvesters in removing grapes are illustrated in Figure 9.

Packaging and Storage of Table Grapes

Because of the importance of appearance, table grapes are trimmed of unsightly or diseased fruit either in the field or in storage sheds before packing. Packing must be done with due care to avoid damaging the fruit. After packing, the containers are placed in cool storage.

Cooling is particularly critical when the grapes are shipped long distances. Cooling not only increases fruit shelf-life (by suppressing respiration), but also limits microbial activity. Cool temperatures also retard water loss, browning, and separation of the fruit from the cluster. If harvesting has occurred under hot conditions, the fruit is usually precooled to 10 °C, before being placed in storage at near 0 °C. Frequent fumigation of the storage area with sulfur dioxide further helps limit decay organisms, fruit browning, and fruit separation.

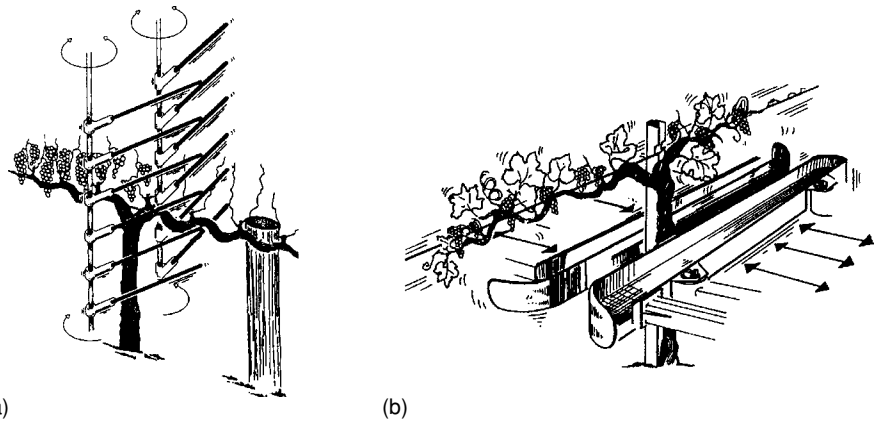

Figure 9 Two main methods of grape removal used by mechanical harvesters: (a) Pivotal strikers that strike the vine. (b) Pulsator rods that shake the vine trunk. From Hamilton RP and Coombe BG (1992) Harvesting of Winegrapes. In: Coombe BG and Dry PR (eds), *Viticulture, Vol. 2, Practices*, pp. 302–327. Adelaide: Winetitles, with permission.

Raisin Production

Table grapes, even with optimal storage, are a perishable crop. Our ancestors solved this problem by processing grapes into either wine or raisins. Raisins have a moisture content of between 10 and 15%, the remainder being almost exclusively sugar. The low moisture content and high osmolarity of the sugar content effectively stop all microbial spoilage activity.

Most grapes used for raisin production are seedless, eliminating the need to remove the seeds. The large-fruited 'Sultana' ('Thompson Seedless') is the most widely grown raisin grape. The small-fruited 'Black Corinth' ('Zante') grape is used for currant production.

Natural drying, either in the sun or under shade, produces the common dark purple–brown raisin of commerce. The browning results from the enzymatic oxidation of phenols released from cell vacuoles during drying. Sun exposure also favors the production of caramelized flavors. Submerging the grapes in a mixture of potassium carbonate and a specially formulated dipping oil disrupts the waxy fruit covering, speeding up drying and producing a golden-brown raisin. Yellow–gold raisins are primarily produced by a hot bleach dip (about 4 s in a boiling solution of dilute sodium hydroxide and sodium sulfate). This is followed by rapid dehydration. Sulfur dioxide in the bleach solution prevents the browning that typically occurs during drying.

See also: **Flavor (Flavour) Compounds**: Structures and Characteristics; Production Methods; **Pectin**: Properties and Determination; Food Use; **Phenolic Compounds**; **Sensory Evaluation**: Aroma; Taste; **Wines**: Types of Table Wine; Production of Table Wines; Production of Sparkling Wines; Dietary Importance

Further Reading

Christensen PL (ed.) (1999) *Raisin Production Manual. University of California, Agriculture and Natural Resources, Publication # 3393*. Davis, CA.

Coombe B and Dry P (eds) (1988) *Viticulture: Resources in Australia*, vol. 1. Adelaide: Winetitles.

Coombe B and Dry P (eds) (1992) *Viticulture: Practices*, vol. 2. Adelaide: Winetitles.

Huglin P and Schneider C (1998) Biologie et Écologie de la Vigne, 2nd edn. Paris: Éditions Tec & Doc Lavoisier.

Jackson D and Schuster D (1997) *The Production of Grapes and Wine in Cool Climates*, 2nd edn. Canterbury, New Zealand: Lincoln University Press.

Jackson RS (2000) *Wine Science: Principles, Practice, Perception*, 2nd edn. San Diego, CA: Academic Press.

Mullins MG, Bouquet A and Williams LE (1992) *Biology of the Grapevine*. Cambridge: Cambridge University Press.

Rantz JM and Lewis KN (eds) (1994) *Proceedings of the International Symposium, Table Grape Production: June 28 & 29, 1994, Anaheim, California*. Davis, CA: American Society for Enology and Viticulture.

Smart RE and Robinson M (1991) *Sunlight into Wine. A Handbook for Winegrape Canopy Management*. Adelaide: Winetitles.

Weaver RJ (1976) *Grape Growing*. New York: Wiley.

Winkler AJ, Cook JA, Kliewer WM and Lider LA (1974) *General Viticulture*. Berkeley, CA: University of California Press.

GROUND NUT OIL

T H Sanders, United States Department of Agriculture, Raleigh, NC, USA

Introduction

Groundnut oil is expressed from the seed of *Arachis hypogaea* L., commonly known as groundnut, peanut, or earth nut, because the seed develops underground. The plant is a legume native to South America and was cultivated as early as 2000–3000 BC. Today, groundnuts are produced on a significant basis in more than 30 different countries with worldwide production figures estimated in excess of 20×10^6 t. The three largest producers of groundnut are India, China, and the USA. In recent years, these countries accounted for approximately 65% of the world production. Although cultivated for many centuries, the economic importance of groundnuts and groundnut oil has increased rapidly only in the past century. Near the middle of the nineteenth century, oil mills in France began importing groundnuts from West Africa for crushing. Because of the excellent

quality of groundnut oil, mills for crushing were soon found in other European countries and then throughout the world. In recent years, the world production of groundnut oil has accounted for about 7–10% of the world vegetable oil total. More than 50% of the groundnuts produced are crushed for oil. In India, 75–80% of the groundnuts produced are crushed for oil, whereas in the USA, groundnuts used for crushing are only approximately 12% of recent production. This low percentage is indicative of the economic importance of groundnuts as a food crop in the USA. Owing to their high content of digestible protein and unsaturated oil, and their exceptional roasted nutty flavor, groundnuts have substantial value and desirability as a food commodity. On a world-wide basis, more than one-third of the groundnuts produced are used as food. In the USA, a high percentage of the limited number of groundnuts used for oil extraction have been separated from edible stocks because of the potential for aflatoxin. The pressed cake, low in oil and high in protein, may be used for animal food if aflatoxin is found to be below acceptable levels. Unacceptable pressed cake is relegated to fertilizer usage. Depending on the extraction methods (hydraulic press, expeller, and/or solvent extractor), residue may contain 1–7% oil, and with inefficient equipment, the percentage may be much higher. Pressed cake from edible-grade groundnuts with low oil content may be ground into flour for human consumption. In many parts of the world, extraction efficiency and lack of hygienic conditions relegate residues unfit for human consumption. In these situations, the residue is utilized either for animal feed or as fertilizer. (*See* **Peanuts**; **Vegetable Oils**: Oil Production and Processing.)

Uses

Peanut oil uses include cooking and frying oil, preparation of shortenings, margarines, and mayonnaise. Some salad oil use is found, and use in pourable dressings is good because of the length of time solids are held in suspension in the oil. However, peanut oil does not meet the strict terminology for salad oil because it solidifies at 0–3 °C. As a cooking oil, especially in deep-fat frying, groundnut oil is excellent since it has a smoke point of 229.4 °C. Refined groundnut oil is odorless, whereas crude oil usually has a nut-like aroma. Groundnut oil develops few off-flavors or odors in use as a frying oil, but degradation of glycerides during frying results in the increase of free fatty acids (FFA) and a decrease in smoke point.

Oil Characteristics

Composition

The seed of the groundnut fruit commonly contain 40–50% total oil, but data indicating a range of 36–56% may be found. Oil content is considered to average about 50% and is generally independent of market type or growth habit. However, some reports suggest that the small-seeded Spanish types tend to have a higher oil content than the large-seeded Virginia types. The composition of the oil changes somewhat among varieties, but the triacylglycerol content is generally in the region of 95%. As peanuts mature, the total oil, triacylglycerol, and oleic acid/linoleic acid (O/L) ratio increase, whereas free fatty acids, polar lipids, monoacylglycerols, and diacylglycerols decrease (Table 1). The relative weight percentage of free fatty acid has been shown to decrease from 4.5 in very immature seed to 0.7 in mature peanut seed. In a study of oil characteristics of peanuts, we reported that free fatty acids decreased (0.8 to 0.05% as oleic acid), the O/L ratio increased, and the oil oven stability increased with maturity. The consistent relationship of these components to shelf-life leaves little doubt as to the relationship of maturity and shelf-life. Diacylglycerol and polar lipid fractions generally account for an additional 2% of the oil weight. Some reports suggest that the total oil percentage decreases

Table 1 Influence of maturity on quantity and composition of selected oil components of Florunner peanuts

Maturity stage[a]	*Oil (percentage dry weight)*	*Triacylglycerol*[b]	*Free fatty acid*[b]	*Diacylglycerol*[b]	*Polar lipid*[b]
5	25.3	85.3	4.5	4.7	2.0
6	30.8	89.3	3.1	3.5	1.4
7	34.4	88.3	2.5	3.6	1.9
8	42.8	90.8	1.8	3.0	1.6
9	45.6	92.6	1.3	2.2	1.3
10	46.7	94.3	0.9	2.0	1.0
11	48.4	94.8	0.7	1.9	0.7
12	48.2	95.8	0.7	1.7	0.6

[a]Relative maturity ranking based on internal shell color. Stage 5 seed are soft and watery. Stage 12 is fully mature.
[b]Relative weight percent.

after full maturation, as lipids are utilized for respiration. Factors such as maturity, environment, cultural practices, variety, and soil temperature affect oil content and composition.

Fatty Acids

As previously indicated, groundnut oil is composed of mixed glycerides and contains a high proportion of unsaturated fatty acids, in particular, oleic (18:1) and linoleic (18:2). Additional fatty acids are palmitic (16:0), stearic (18:0), arachidic (20:0), 11-eicosenoic (20:1), behenic (22:0), and lignoceric (24:0). With maturation, the percentage of oleic acid increases, whereas the percentage of linoleic acid decreases slightly. The stability of groundnut oil is highly correlated with the ratio of oleic acid to linoleic acid. This ratio generally increases with seed maturity, and oil stability increases simultaneously. Fatty acid composition values have been reported to vary widely, as indicated in Table 2. The fatty acid compositions of the lipid classes of groundnut are somewhat variable. The fatty acid composition of three groundnut varieties and composition of lipid classes from those varieties are shown in Table 3. The triacylglycerol composition approximates that of whole oil since the fraction comprises about 95% of the total. Oleic acid, linoleic acid, and palmitic acid comprise about 90% of the total fatty acids. Free fatty acid fractions consistently contained higher percentages of palmitic acid than did the triacylglycerol fractions. Long-chain fatty acids (C20–C24) were generally more predominant in the *sn*-1,3-diacylglycerol than in other fractions, and only traces of long-chain fatty acids were found in the *sn*-1,2 (2,3) diacylglycerol fraction. The phospholipid fractions contained the highest concentrations of palmitic acid. (*See* **Fatty Acids**: Properties.)

Production location has a significant effect on the fatty acid composition of peanut oil, with cooler climates resulting in a greater degree of unsaturation, a lower O/L ratio, and thus, according to published correlations, a shorter shelf-life. These effects have been variously ascribed to temperature, irrigation, and maturity. The relationship between mean temperature at critical growth periods appeared to have a definite effect on oil composition. The relationships examined may provide at least partial explanation for observed problems with oxidative stability in peanuts grown in

Table 2 Reported percentages of major fatty acids in groundnut oil

Fatty acid	*Percentage*
Palmitic	7.4–12.5
Stearic	2.7–4.9
Oleic	41.3–67.4
Linoleic	13.9–35.4
Arachidic	1.2–1.9
Behenic	2.1–3.6
Lignoceric	0.9–1.7

Table 3 Fatty acid composition of various lipid classes in petroleum ether-extracted Starr, Florunner, and Florigiant peanut oil[a]

Variety	*Lipid class*	*16:0*	*18:0*	*18:1*	*18:2*	*20:0*	*20:1*	*22:0*	*24:0*
				Mole %					
Starr	TG	14.1	3.3	44.6	32.1	1.5	0.8	2.7	0.9
	FFA	21.0	4.2	40.4	29.9	1.2	0.4	2.2	0.8
	sn-1,3-DG	17.7	3.8	48.2	22.9	1.5	1.2	3.4	1.3
	sn-1,2(2,3)DG	15.5	2.5	48.9	33.1				
	MG	17.9	2.4	46.5	27.8	0.8	0.8	2.4	1.4
	PL	22.3	3.6	44.8	24.6	0.8	1.0	1.6	1.2
	WO	14.0	2.6	43.9	34.2	1.3	0.8	2.5	0.8
Florunner	TG	11.4	2.3	51.9	28.5	1.2	1.2	2.4	1.2
	FFA	16.9	3.0	45.0	30.1	1.1	1.2	1.9	0.8
	sn-1,3-DG	13.8	2.6	51.5	25.1	1.2	1.6	2.4	1.4
	sn-1,2(2,3)DG	13.6	1.5	48.6	36.3				
	MG	16.1	3.3	47.9	27.4	0.7	0.6	3.2	0.8
	PL	21.3	3.4	45.1	28.9			0.7	0.7
	WO	11.0	1.8	51.7	29.9	1.0	1.1	2.4	1.1
Florigiant	TG	11.2	3.5	52.7	26.6	1.5	0.9	2.3	1.2
	FFA	16.1	3.9	46.6	28.2	1.6	1.0	1.9	0.7
	sn-1,3-DG	13.2	3.9	52.1	22.2	2.4	1.7	3.4	1.2
	sn-1,2(2,3)DG	12.6	2.2	48.9	35.8			0.3	0.3
	MG	16.7	4.4	48.4	26.8	0.8		2.3	0.6
	PL	22.1	3.8	42.8	29.0	0.4		1.0	0.9
	WO	11.0	2.8	54.3	27.2	1.3	0.9	2.0	0.8

[a]All values are the means of three replicate analyses. TG, triacylglycerol; FFA, free fatty acid; DG, diacylglycerol; MG, monoacylglycerol; PL, polar lipid; WO, whole oil.

cooler climates or with cooler temperatures during the later weeks of the growing season.

This research has demonstrated several highly significant variations in oil production from year to year for all three major peanut types, highly significant differences among the same varieties grown in different commercial production area, and highly significant interactions between location and year of production. The study has also demonstrated highly significant correlations between mean temperatures during apparently critical growth periods and the levels of major fatty acids produced.

Triacylglycerol

Studies defining over 20 and up to 36 triacylglycerol species in peanut oil have been conducted. Triacylglycerol species OOL, OOO, OLL, POL, and POO (O = oleic, L = linoleic, P = palmitic) comprised the greatest proportions of the triacylglycerol fraction. This should be expected since oleic acid, linoleic acid, and palmitic acid comprise the greatest percentages of fatty acids in peanut oil. The US Customs Service has utilized triacylglycerol species distribution as one method of identification of some international origins of peanuts. After examination of 18 Argentine and 13 Chinese samples, observations for distinguishing between the two production locations were: (1) the ratio of OOO/OOP was less than 1.0 for Chinese peanuts; it was approximately 1.5–1.7 for the Argentine peanuts; (2) PSL (S = stearic) was not measurable in Chinese peanuts and was present at about 0.2% in Argentine samples; (3) in Chinese samples, OOB (B = behenic) accounted for about 0.4% of the total triacylglycerols, whereas in the Argentine samples, it accounted for about 1.1%; (4) the percentage of OOO was about 10% in Chinese samples versus about 14% in Argentine samples. Triacylglycerol species percentage ranges for Argentine and US samples often overlapped; however, origin identification was possible through trace element analysis. (*See* **Fats:** Classification.)

Data presented in several studies have clearly demonstrated the fact that the environment affects not only the fatty acid composition of peanut oil, but also, although apparently indirectly, the spatial arrangement of those acids on the triacylglycerol molecule. Triacylglycerol composition and structure are important in the areas of nutrition, oil stability, and possible physiological effects. Table 4 provides data demonstrating the differences in triacylglycerol structure that are reasonably expected among different varieties of peanuts. The data shown in Table 4 indicate a nonrandom distribution of fatty acids among the *sn*-1, -2, and -3 positions of the triacylglycerols.

Table 4 Stereospecific analyses of triacylglycerols from six peanut varieties

Variety	*Compound or position*	*Fatty acid distribution (mole %)*[a]							
		16:0	*18:0*	*18:1*	*18:2*	*20:0*	*20:1*	*22:0*	*24:0*
Florigiant	TG	10.8	2.9	53.1	27.3	1.8	1.0	1.9	1.1
	1	20.1	4.9	50.7	22.6	0.5	0.7	0.4	0.3
	2	2.2	0.7	51.5	45.3	0.1	0.3	0.1	
	3	10.3	3.2	57.2	14.0	4.8	2.0	5.3	3.0
Early Bunch	TG	13.3	2.6	42.0	36.3	1.5	0.9	2.4	1.1
	1	24.7	4.4	38.3	31.2	0.3	0.4	0.5	0.2
	2	3.5	1.5	37.2	57.4	0.1	0.2		
	3	11.8	1.9	50.6	20.3	4.0	1.9	6.5	2.9
Florunner	TG	11.4	2.1	50.9	29.1	1.6	1.1	2.4	1.3
	1	20.7	3.5	49.5	24.4	0.3	0.7	0.5	0.5
	2	2.1	0.6	47.8	48.8	0.1	0.4	0.1	0.1
	3	11.4	2.3	55.5	14.1	4.4	2.3	6.5	3.4
Tifrun	TG	12.6	2.4	42.4	36.8	1.5	0.9	2.6	0.8
	1	22.7	4.2	38.8	32.7	0.3	0.4	0.6	0.2
	2	3.2	1.2	36.1	58.8	0.2	0.2	0.1	
	3	11.9	1.7	52.2	18.9	4.0	2.0	7.2	2.2
Starr	TG	14.2	3.3	43.3	33.0	1.8	1.1	2.7	0.7
	1	24.2	4.9	40.4	28.4	0.4	0.6	0.7	0.3
	2	2.4	0.8	39.5	56.9	0.1	0.2	0.1	
	3	16.0	4.2	50.0	13.8	4.8	2.4	7.3	1.8
Spancross	TG	13.5	2.9	44.1	33.4	1.9	1.1	2.2	0.9
	1	24.2	5.2	40.3	28.4	0.3	0.5	0.6	0.4
	2	2.6	0.9	40.2	56.2	0.1			
	3	13.5	2.8	51.8	15.6	5.2	2.5	6.5	2.3

[a]Each value is the mean of three replications.

The percentages of palmitic and stearic acids were generally very low for the *sn*-2 position, higher for *sn*-3, and highest for *sn*-1. The long-chain (C20–C24) fatty acids were located almost exclusively at the *sn*-3 position. The *sn*-2 position of triacylglycerols from all the varieties was high in unsaturated fatty acids. The general patterns of fatty acids found at the *sn*-1 and *sn*-3 positions were similar for all varieties, although the mole percentages of each acid at the two positions frequently differed widely. Mole percentages of palmitic, stearic, and linoleic acids were always higher for the *sn*-1 than for the *sn*-3 position, whereas those of oleic acid were consistently higher for the *sn*-3 position. The patterns of fatty acid distribution at *sn*-2 differed not only from those at *sn*-1 and -3, but also with variety. On the *sn*-2 position, the percentage of oleic acid was higher than that of linoleic acid in Florigiant, but the percentages were about the same in Florunner. In the other four varieties, there was more linoleic acid esterified at the *sn*-2 position than oleic acid. Florigiant and Florunner triacylglycerols contained more oleic acid and less linoleic acid than the other varieties examined, and the concentration effect probably was reflected by the fatty acid placement in the molecule.

Using the data in Table 4, linear regression equations and correlation coefficients were calculated for the plots of the percentage of a fatty acid in the total triacylglycerol *vs*. the percentage of that fatty acid at one of the positions of the triacylglycerol (Table 5). Significant correlations indicate that the total fatty acid present influenced placement of that fatty acid on the triacylglycerol. It has been reported that major saturated, monoene, and diene fatty acids of corn triacylglycerols exhibited a concentration effect in all cases except for saturated acids in the *sn*-2 position. Peanut triacylglycerols exhibited this same pattern, and the low concentrations of the long-chain fatty acids in the triacylglycerol were significantly correlated with percentages found at the *sn*-3 position only. This may be due to the general restriction of the saturated acids (16:0 and 18:0) from the *sn*-2 position and the long-chain acids from the *sn*-1 and *sn*-2 positions. Any substantial deviation from the regression line has been attributed to a change in the mechanism of fatty acid distribution and suggests genetic control of the deviation. No substantial deviation of any fatty acid from the regression lines was detected in the six peanut varieties examined. Although the concentration effects were similar for the varieties, the variation in percentage of a fatty acid at any position is sufficient to indicate possible concentration differences in the various triacylglycerol species found in the total triacylglycerol fraction.

Table 5 Linear regression analyses and correlation coefficients for the relationship of fatty acids in total triacylÙglycerols and fatty acids at each position

Fatty acid	*Position*	*Slope*	*y intercept*	*r*[a]
16:0	1	1.43	4.68	0.95[b]
	2			0.43[c]
	3	1.32	−4.23	0.86[d]
18:0	1	1.28	1.03	0.88[d]
	2			0.09[c]
	3	1.82	−2.23	0.83[d]
18:1	1	1.16	−10.33	0.99[b]
	2	1.28	−16.87	0.99[b]
	3	0.56	26.84	0.95[b]
18:2	1	1.01	−5.07	0.99[b]
	2	1.38	8.83	0.95[b]
	3	0.60	−3.75	0.82[d]
20:0	3	2.78	−0.15	0.99[b]
20:1	3	2.31	−0.16	0.91[d]
22:0	3	2.39	0.88	0.96[b]
24:0	3	2.65	−0.01	0.99[b]

[a] r = correlation coefficient.
[b] Significant at 1% level.
[c] Not significant.
[d] Significant at 5% level.

Phospholipids

Phospholipids of peanut oil have been examined, and it has been determined that oven stability of peanut oil was related to the solvent used for extraction and that removal of phosphatides from the oil by precipitation reduced oven stability. Thus, differences in the relative amounts of polar lipids extracted by different solvents might explain some variation in oven stability. Stability and resistance to oxidation due to sterols and phospholipids are not well documented and warrant further investigation. The concentration of phospholipids in peanut oil is only about 1%. Phospholipids or membrane lipids have been recognized for some time in the form of lecithin materials (primarily phosphatidylcholine (PC)) as a byproduct of oil and fat processing. Phospholipids, especially PC, were found to be important in the treatment of neurological diseases, respiratory distress, liver diseases, and many others. This class of compounds has been shown to be synergistic with tocopherols in delaying onset of lipid oxidation. The major phospholipids of peanut oil are phosphatidic acid (PA), phosphatidylglycerol (PG), phosphatidylethanolamine (PE), phosphatidylinositol (PI), and PC. Phospholipid content and concentrations have been shown to be affected by maturity and postharvest treatment, as shown in Table 6. The higher concentrations of PA and PC in immature peanuts might be explained on the basis that these phospholipids are precursors to other phospholipids. Excessive heat and freezing both affect membrane stability with the result of significant

Table 6 Effect of postharvest treatment on total phospholipid

Treatment	*Phospholipid (% area)*					*Total phospholipid (mg per 100 g DWT)*
	PA	*PG*	*PE*	*PI*	*PC*	
Control	2.2	2.5	13.3	15.7	66.4	500
Immature	4.5	2.3	14.0	7.6	71.7	700
Heat cured	9.5	1.1	16.0	15.4	58.1	900
Freeze-damaged	28.3	14.1	15.2	33.5	8.8	250

PA, phosphatidic acid; PG, phosphatidylglycerol; PE, phosphatidylÙethanolamine; PI, phosphatidylinositol; PC, phosphatidylcholine; DWT, dry weight.

differences in the phospholipid content and distribution. The large increase in PA and great decrease in PC with freezing of nondried peanuts may be related to the fact that freezing induces phospholipase-D activity, with PC being the most susceptible to action of the enzyme. (*See* **Phospholipids**: Properties and Occurrence.)

Sterols

Peanut oil contains several sterols, which are secondary alcohols with 27–29 carbon atoms, and unlike short-chain alcohols, they are crystalline solids at room temperature. Peanuts contain β-sitosterol, campesterol, stigmasterol, Δ5 avenasterol, Δ7 stigmasterol, Δ7 avenasterol, and brassicasterol (Table 7). β-Sitosterol (SIT), the major component in peanut sterols, has been shown to inhibit cancer growth. SIT may offer protection from colon, prostate, and breast cancer. Unrefined peanut oil contains approximately 200 mg of SIT per 100 g of oil, and this value is comparable with soybean oil, which contains approximately 183 mg of SIT per 100 g.

Antioxidants

Few studies on the relative quantity of antioxidants in peanuts have been published, and studies on the relationship of antioxidant content to shelf-life potential are even less abundant. Data on tocopherol content and individual fatty acids from 31 cultivars for 4 years used in a multiple regression equation for the prediction of stability of cold pressed oil revealed that 87% of the stability could be correlated with total tocopherol/percent linoleic acid. A 3-year study of oil composition factors of peanuts from several origins indicated that tocopherol content was consistently different in peanuts from various origins. Peanuts produced in the USA had a consistently higher tocopherol content than peanuts produced in China or Argentina. The highest levels reported in US peanuts were almost 250 p.p.m. in whole seed, whereas the lowest levels were about 100 p.p.m. The tocopherol content of peanut oil has been demonstrated to be as high as 650 p.p.m., depending on the variety of peanut and growing conditions. Vitamin E intake has been shown in numerous studies to be inversely related to the risk of heart disease. The 3-year study on peanuts from the USA, China, and Argentina reported on levels of copper and iron. The copper content was always significantly lower in US peanuts, and the iron content was generally lower. These factors were related to higher O/L ratios, and oil oven stability studies indicated the overall longer shelf-life potential of peanuts produced in the USA. These data, without consideration of the areas of production, demonstrated the consistent close relationship of oil quality factors such as FFA, O/L ratio, peroxide value, total carbonyls, tocopherols, copper, and iron to most measures of shelf-life stability. The data may be extrapolated to the end that any factor that generates inappropriate levels of these compounds will then contribute to a shorter shelf-life of resulting products. (*See* **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants.)

Table 7 Sterol content of unsaponifiables of peanut oil and whole peanuts

	Peanut oil (mg per 100 g)	*Whole peanuts (mg per 100 g)*
Total sterol	337	220
β-Sitosterol	217	142
Campesterol	49	24
Stigmasterol	36	23
Δ5-Avenasterol	26	nr
Δ7-Stigmasterol	6	nr
Δ7-Avenasterol	2	nr
Brassicasterol	Trace	nr
Other		31

nr, not reported.

Properties of Peanut Oil

General

The general properties of peanut oil have been compiled from several references and are provided in Table 8. Crude peanut oil has a bland, but slightly beany, nut-like flavor that is removed during refining. Oxidized oil may have a wet cardboard or painty (oxidized linseed oil) flavor.

Acetyl Value

The acetyl value is the number of milligrams of KOH required to neutralize the acetic acid produced by the hydrolysis of 1 g of acetylated fat and is a measure of the free hydroxyl groups present in the oil. The acetyl number of peanut oil (8.5–9.5) is lower than other vegetable oils but higher than coconut oil, palm oil, and the animal fats and oils.

Table 8 Chemical and physical characteristics of peanut oil

Characteristic	*Value*
Flavor and odor	Bland
Acetyl value	8.5–9.5
Color (visual)	Light yellow
Color (Gardner, maximum)	4
Free fatty acid (as oleic acid, maximum)	0.05%
Heat of fusion (unhydrogenated)	21.7 cal g^{-1}
Melting point	0–3 °C
Peroxide value (maximum)	10 meq of peroxides oxygen per kilogram of oil
Refractive index (n_D 40 °C)	1.46–1.465
Smoke point (minimum)	229.4 °C
Specific gravity (21 °C)	0.915
Unsaponifiable lipids	0.40%

Percent Free Fatty Acid (FFA)

Peanut oil generally has low levels of FFA. The highest natural percentages are found in very immature seed (0.8), and the percentage decreases to *c*. 0.05 in fully mature seed. Improper handling, moisture, fungal invasion and other factors contribute to hydrolysis of triacylglycerols and cause significant increases in FFA. Saponification in the refining process removes FFA, but crude oil used in many parts of the world may contain high levels of FFA with resulting rapid deterioration in quality, flavor, and stability.

Iodine Value (IV)

The iodine value is a measure of the relative degree of unsaturation in oil components, as determined by the uptake of halogen. Because the melting point and oxidative stability are related to the degree of unsaturation, IV provides an estimation of these quality factors. The greater the iodine value, the more unsaturation and the higher the susceptibility to oxidation. Peanut oil (IV 82–107) is more saturated than corn (IV 103–128), cottonseed (IV 99–113), or linseed (IV 155–205) oils; however, it is considerably less saturated than coconut (IV 7.7–10.5), palm (IV 44–54) or butter (IV 25–42) oils.

Color

The light yellow color of peanut oil is due to β-carotene and lutelin, and becomes lighter as peanuts mature. Although oil color may be used to assess maturity, other methods are preferred because many factors such as curing temperature and duration influence oil color. Color measurement is frequently done by visual comparison under a CIE standard light source. The Gardner color is determined using a Gardner-delta color comparator, which has a scale between 1 and 18.

Heat of Fusion

The heat of fusion, or latent heat, is the quantity of heat required to change 1 g of solid to a liquid with no temperature change. This latent heat increases with increasing molecular weight. The heat of fusion of peanut oil is 21.7 cal g^{-1}.

Melting Point

At refrigeration temperatures (0–3 °C), peanut oil sets to a gel. In contrast to safflower, soybean, corn, and olive oils, peanut oil is not filterable during winterization.

Peroxide Value

Elevated peroxide values indicate that lipid oxidation has taken place and is measured as reactive oxygen content in terms of milliequivalents per 1000 g of fat. In raw peanuts, once the cell structure is disrupted by pressing, lipoxygenase reacts with linoleic, linolenic, or arachidonic acid to form hydroperoxides. These oxidation products are correlated with reduced flavor scores and cardboard or painty flavor defects. The peroxide value is often used as an indicator of peanut quality related to oil oxidation.

Unsaponifiable Lipids

The unsaponifiable matter in peanut oil is largely sterols (mainly β-sitosterol and campesterol).

Health Issues

Early work with animals suggested a high atherogenic potential when peanut oil was fed in relatively high doses. Because chemical treatment to randomize peanut oil resulted in a decrease in atherogenicity, the triacylglyceride structure was believed to be involved. However, recent numerous human epidemiological studies have indicated an amazingly high 30–50% reduction in cardiovascular disease in people who ate nuts, including peanuts, four to five times a week. In a recent 4.2-year study, researchers found that compared with those who rarely or never consumed them, study participants who consumed two or more servings of peanuts and nuts per week, decreased their risk of heart disease by 25%. These significant data align with the numerous research studies that have shown that diets high in mono- and polyunsaturated fats and low in saturated fat can be heart-healthy. Recent studies have been conducted in which healthy subjects consumed one of five diets: a low fat diet; one including olive oil; one including peanuts and peanut butter; one including peanut oil, and a typical American diet. Results show that the diet including peanuts and peanut butter, the

diet including peanut oil, and the diet including olive oil (all low in saturated fat and cholesterol, and high in monounsaturated fat) lowered total cholesterol and low-density lipoprotein cholesterol. Further, each of these three diets lowered triacylglycerol levels, but did not lower the beneficial high-density lipoprotein cholesterol.

New Developments – High-oleic Peanut Oil

Peanut lines with a high-oleic acid trait have been identified, and this trait has been incorporated into commercial peanuts. The original two lines identified had approximately 80% oleic and 2% linoleic acid. The lines developed with the high oleic acid trait have O/L ratios of approximately 30, and the lines do not show any meaningful differences in oil content, flavor, color, or texture. Oxidative stability comparisons made on extracted, neutralized, and bleached oil from high oleic (76.3% oleic/4.7% linoleic) and conventional (56.6% oleic/24.2% linoleic) lines differing only in fatty acid composition resulted in up to 15 times more stability in the high oleic oil. The use of high oleic oil in roasting of peanuts resulted in slight increases in shelf-life, as measured by the oxidative stability and peroxide value, and the degree of improved shelf-life was related to the O/L ratio of the peanut roasted. Swine fed a diet incorporating high oleic acid peanuts had increased monounsaturates and lower polyunsaturates in backfat relative to use of a feed incorporating standard peanuts or canola oil. Further, the use of high oleic acid peanuts as the fat source in a low-fat and high-monounsaturate diet produced significant and positive changes in blood lipids in postmenopausal women including a reduction in total cholesterol from 264 to 238 mg dl^{-1}.

See also: **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; Synthetic Antioxidants, Characterization and Analysis; Role of Antioxidant Nutrients in Defense Systems; **Fatty Acids**: Properties; Metabolism; Gamma-linolenic Acid; Dietary Importance; **Peanuts**; **Phospholipids**: Properties and Occurrence; Determination; Physiology; **Tocopherols**: Properties and Determination; Physiology

Further Reading

Ahmed EM and Young CT (1982) Composition, quality and flavor of peanuts. In: Pattee HE and Young CT (eds) *Peanut Science and Technology*. Yoakum, TX: American Peanut Research and Education Society.

Cobb WY and Johnson BR (1973) Physicochemical properties of peanuts. In: *Peanuts – Cultures and Uses*. Stillwater, OK: American Peanut Research and Education Association.

Holaday CE and Pearson JL (1974) Effects of genotype and production area on the fatty acid composition, total oil and total protein in peanuts. *Journal of Food Science* 39: 1206–1209.

Salunke DK and Desai BB (1986) Groundnut. In: *Postharvest Biotechnology of Oilseeds*. Boca Raton, FL: CRC Press.

Sanders TH (1979) Varietal differences in peanut triacylglycerol structure. *Lipids* 14: 630–633.

Sanders TH (1980) Effects of variety and maturity on lipid class composition of peanut oil. *Journal of the American Oil Chemists Society* 57: 8–11.

Woodroof JG (1983) Peanut oil. In: *Peanuts – Production, Processing, Products*. Westport, CT: AVI Publishing.

Young CT (1996) Peanut oil. In: *Bailey's Industrial Oil and Fat Products Edible Oil and Fat Products: Oils and Oil Seeds*, vol. 2, pp. 377–392. New York: John Wiley.

GROWTH AND DEVELOPMENT

R Tojo and R Leis, Santiago de Compostda University, Hospital Clínico Universitario de Santiago de Compostela, Spain

Introduction

Human growth is a product of continous interaction between inheritance and a large number of environmental variables, which act both in fetal period and in the two first decades of the life. Hormones, growth factors, and nutrition play an important role in growth velocity and the attainment of adult size.

Growth is an excellent indicator of health status or illness. Periodic growth assessment must be an obligatory exploration in the supervision of children and adolescents.

Growth and Development

Growth is a complex process. It is the product of continuous interaction between inheritance and environment from conception to adulthood. Each individual has a genetic basis with a well-defined growth potential, modulated by a wide range of extragenic factors. Therefore, body composition and

predetermined final adult size will only be reached if all factors act harmoniously.

Growth and development refer to the transforming process of a fertilized ovum into an adult individual. Growth basically implies an enlargement of body dimension originated by an increase in cell number and size, that is, a process of cell hyperplasia and hypertrophy. Development is a physiological process which makes possible – from a pluripotential and undifferentiated cell – the differentiation, maturation, and function of all the tissues, organs, and apparatus which make up the human organism.

Cell-Growth Stages

Each tissue, organ, or apparatus growth process, in spite of having a development program of its own, is conditioned by cell division and size. This process of tissue growth generally is in three stages:

1. Cell hyperplasia: the initial stages of growth, mainly during organogeny and the fetal period, are characterized by cell division and proliferation, by hyperplasia.
2. Hyperplasia-hypertrophy: as the organ or tissue approaches its predetermined cell total, cell division and increase in cell number are less, whilst cell size rises.
3. Hypertrophy: when total adult cells are reached, cell division stops and cell growth exclusively depends on the increase in existing cell size.

Types of Tissue According to their Growth and Maturation: Tissular Specialization

There are three types of tissue, according to their life span and cellular cycle.

Permanent or Static Tissues

These tissues, once they have reached the adult cell number – which happens precociously – remain theoretically invariable. Their subsequent renovation does not occur, or is only minimal. That is, dead cells cannot be replaced, and cell size and structures can only change or renew their subcellular constituents. Nerve cell, cardiac muscle, and retina cells are the paradigms for this behavior.

Renewing Tissues

These are characterized by a high rate of continuous cell proliferation throughout life. They have a very active turnover of cell populations, which are continously eliminated and renewed, but keep a stable mass. Epidermis, intestinal mucus, seminiferous tubules and hematopoietic tissue are renewing tissues.

Expanding Tissues

Once they have attained adult growth and development level, expanding tissues do not vary unless damaged, in which case a regenerative compensatory phenomenon of cell hyperplasia and hypertrophy takes place. Liver and kidney are typical expanding tissues.

Patterns for Organic and Tissular Growth

General Pattern for Human Growth, or Scammon's General Curve

Scammon's general curve is the curve followed by the whole organism. A good example is size, which, expressed in adult-size percentages, follows a sigmoid curve with two periods of rapid increment (first, prenatal and immediately postnatal ones, and second, pubertal growth) and, between both, a curve of slow and stable growth corresponding to preschool and school age. Many other corporeal components follow roughly this growth pattern (respiratory and digestive systems, kidney, spleen, muscles and osteal system) (Figure 1).

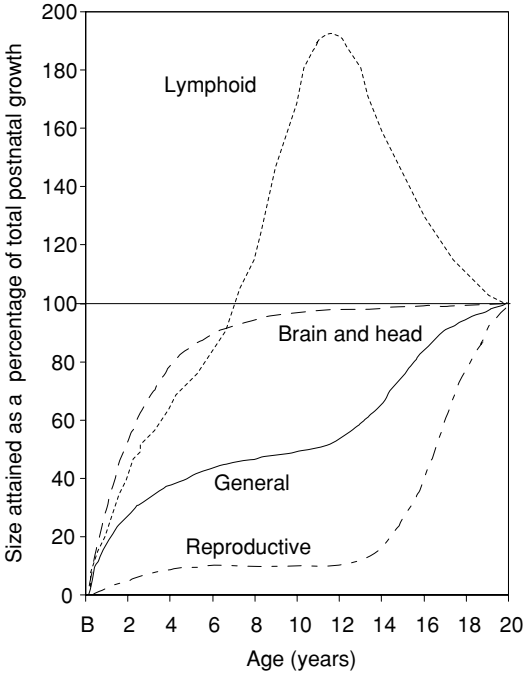

Figure 1 Growth curves of different parts and tisúes of the body. All the curves are of the size attained and plotted as a percentage of total gain from birth to 20 years. Reproduced with permission from Tanner JM (1962) *Growth at adolescence*, 2nd edn. Oxford: Blackwell.

Neural, Cerebral, or Cranial Growth Pattern

The growth of the central nervous system is extremely quick during the prenatal period and in the first years of life, reaching 90% of adult weight at the age of 5.

Lymphoid Growth Pattern

The growth processes of some lymphoid structures, such as thymus, lymphatic nodes, and adenoid glands or tonsils, undergo great development in early life. Thus, at the age of 5, they will have already reached 100% of their adult size, following their enlargement up to 180%, which is achieved at 12–13 years of age. From that time onwards, size nearly halves so that, roughly at 20, it reverts to the 5-year-old size.

Growth Pattern of the Reproductive System

Gonad and external genital growth takes place in two stages: embryonic–fetal stage, permitting sexual difference, followed by a long period when growth is very slow, without significant changes from birth to the end of prepuberty. At the beginning of puberty, and within a 3-year period, the second stage of growth, takes place when gonad and external sexual character growth is completed and reproductive capacity is achieved.

Growth Pattern of Adipose Tissue

This is characterized by rapid growth during the first year of life. Then it decreases, and from 10 years onwards there is a new increase, which is more evident in women's case, which continues in puberty. In contrast, in men it decreases, so that at the end of the growth process women's rate of body fat is almost twice that of men.

Growth and Variation of Body Proportions

After 2 months of fetal life, craniofacial section represents 50% of total body length. This decreases to 25% in the newborn, whilst in adulthood it is only one-eighth of total size. Leg proportion undergoes an opposite evolution to that of the cranium. Trunk proportional changes are much more limited. They affect the superior/inferior body segment ratio, which ranges from 1.7 at birth to 1.0 at 10 years.

Critical or Sensitive Periods of Cell Growth

Teratogens, infections, and nutrient deficiency have different effects on the growth and function of organs, depending on growth speed. That is, during the growth process there are critical or sensitive periods which are characterized by an increase of vulnerability to a specific stimulus. This phenomenon occurs along periods of maximum cell proliferation (cell hyperplasia), mainly in the fetal period and the first year of life, and in puberty, to a lesser extent.

In line with this is the 'programming phenomenon.' Poor nutrition and other adverse biological and environmental influences in the critical periods of fetal development and the first year of life may cause permanent changes in gene expression, cell replication, organic function and structure, hormone action and secretion, and growth factors. These factors, apart from affecting growth and body composition, will also favor the development of degenerative diseases such as cardiovascular disorders, high blood pressure, diabetes or hyperlipidemia (X-syndrome) during adulthood, and these diseases are nowadays the main cause of morbid mortality. Fetal growth is a significant predicting factor for postnatal growth and adult size (Figure 2).

Human Growth Pattern: The Infancy–Childhood–Puberty Model

Currently, the most accurate model for the relationship between growth and the biological factors responsible for growth at different stages of life is Kalberg's ICP (infancy–childhood–puberty) model. Based on experimental data, this model shows a general growth curve representing the additive and partially overlapping effect of its biological basis, from the fetal period to the end of puberty (Figure 3).

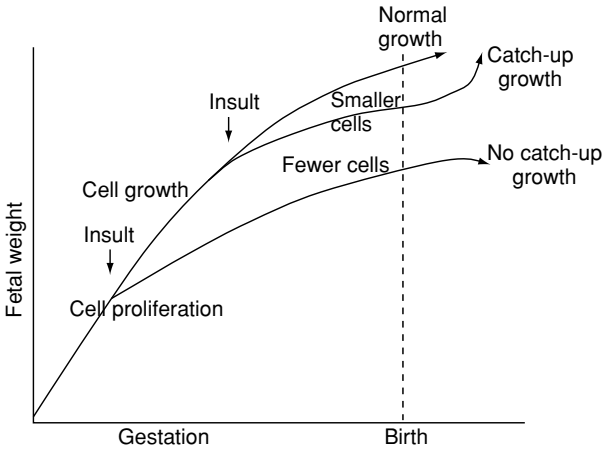

Figure 2 This diagram illustrates how fetal growth is predominantly by cell proliferation in early pregnancy and later by cell growth. For this reason, an early insult may result in a permanent reduction in cell number, whereas one later in pregnancy may reduce cell size and be reversible after birth. Reproduced with permission from Holmes R and Soothill PW (1998) Normal fetal growth. In: Kelmer ChJH, Savage MO, Stirling HF and Saenger P (eds) *Growth disorders*, pp. 143–157. Cambridge: Chapman & Hall.

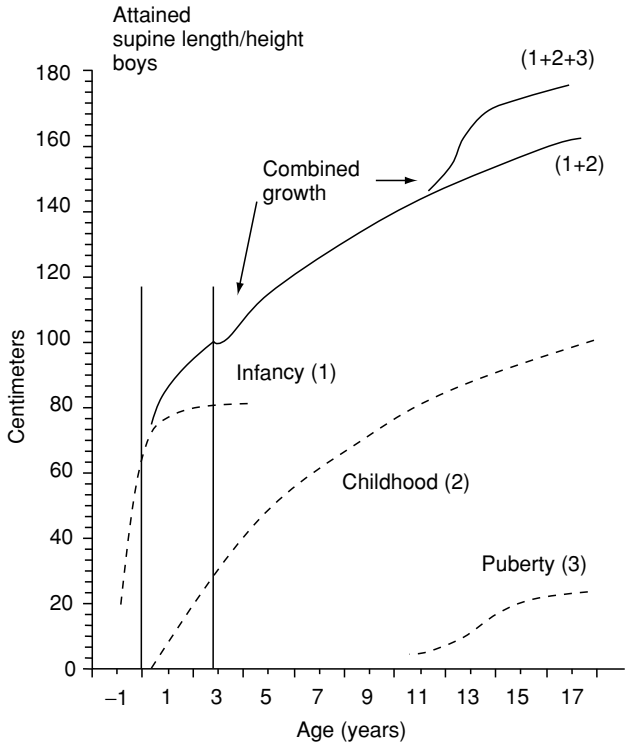

Figure 3 The infancy–childhood–puberty (ICP) model of growth: the infancy component is initially rapid but decelerates as the influence of growth hormone appears by the end of the first year. The effect of growth hormone is sustained through childhood but, in the absence of adequate sex steroid secretion, the pubertal growth spurt is blunted or absent. The combined input of the three components results in the normal pattern of growth until final height is achieved. Reproduced with permission from Karlberg J (1989) A biologically oriented mathematical model (ICP) for human growth. *Acta Paediatrica* 350 (suppl.): 70–78.

Infancy Component (I)

Infancy is from the second half of gestation to the second to third year of life. It is represented by an exponential function ($y = a + b(1-\exp(-ct))$. Therefore, it covers an accelerated period of fetal growth and one of high-speed postnatal deceleration. Growth during the fetal period and the first stage of life depends on the mother to a great extent, since her size, health, nutrition state, placental structure and function, and nutrient contribution to her child exert a great influence. In this sense, growth is mainly 'nutrition-dependent' as growth hormone (GH)-receptor scarcity and tissular insensitivity to their action show their secondary role in fetal growth and during the first period of life.

Childhood Component (C)

Up to the second to third year, the infancy component continues to influence growth. But, from the sixth month onwards, the childhood component starts to exert a progressive influence too, acting in both an additive and complementary way. After the second to third year, the childhood component keeps on exerting this influence until puberty begins. It is characterized by slow and stable growth. The mathematical model for the childhood component is a second-grade polynomial function: $y = a + bt + ct^2$.

In normal conditions, the childhood component starts between sixth and 12th month of life, becoming manifest by means of an increase in growth velocity. Its appearance represents the beginning of a progressive and significant GH effect on linear growth. At this stage, the highest percentage of body growth takes place at the inferior limbs and is related to a greater sensitivity of long bones to GH action in comparison to short ones and vertebrae. The fact that it starts after the 12th month may have a crucial influence on final adult size, as proved by the fact that the biggest difference in adult size between developing and developed countries is established during the very first period of the childhood component. Moreover, socioeconomic, health, and nutritional situation are also involved. Whilst the infancy component is 'nutrition-dependent,' GH action is essential for the childhood component, although nutrition continues to play an important role.

Pubertal Component (P)

The pubertal component is described as a logistic function: $y = a/(1+\exp[b-[t-tv]]$, and is the result of the synergy of two hormonal systems promoting growth, one of them depending on GH and the other one on sexual steroids (estrogens, testosterone). Growth accelerates, ending with a high-speed peak, and then decelerates to adult size.

Growth Velocity According to the following ICP Model

Fetal Growth

The fetal period is the most intensive period of life as regards growth and development. One characteristic is the elevated rate of cell division – there is maximum cell hyperplasia. It is a chronologically short although biologically active time. Let's say that 42 consecutive cell divisions take place in the fetus whilst from postnatal period to the end of growth only five more divisions take place. Thus, from the fertilized cell, the newborn gains 3500 g weight, whereas from newborn to adulthood its weight is only multiplied 20 times (3.5 kg to 70 kg). As regards length, the fertilized cell reaches 50 cm at birth. Subsequently, this length is multiplied 3.5 times (50 cm to 175 cm).

Infant Growth (0–2 years)

Postnatal growth is, essentially, hypertrophic: it is an enlargement of the already existing cells rather than the addition of new ones. The period of maximum postnatal growth corresponds with the first year of life. During the first 12 months of life, size increases 50%, by around 25 cm (50 versus 75–80 cm), weight rises 200%, by around 7 kg (3.3 versus 10 kg), and cranial circumference enlarges by 45%, which is around 11 cm (35 versus 47 cm), relative to birth. During the second year, there is an obvious deceleration of growth: size increases by only 10–12 cm (75 versus 85–90 cm), weight rises 2.5 kg (10 versus 12–13 kg) and cranial circumference enlarges by 2–2.5 cm (47 versus 50 cm).

Growth during Preschool Age (3–5 years)

This is the final stage of neonatal growth deceleration. Growth velocity slows down and becomes more constant; size increases by around 6–8 cm $year^{-1}$. Around the fourth year of life neonatal size has doubled (50 versus 100 cm), weight rises 2 kg $year^{-1}$ and cranial circumference enlarges by 1 cm $year^{-1}$. The increase is in leg length, while subcutaneous fat decreases. This changes the infant's look into a more slender, less rounded one than that of the first 2 years of life.

Growth Velocity from School Age to the Beginning of Puberty (6–10 Years for Girls, 6–12 Years for Boys)

Growth becomes slow and steady. The increment per year is 5–6 cm for size and 3 kg for weight. The child has a proportionally greater increase on leg length. Immediately before puberty, growth velocity reaches its lowest point – 4–5 cm $year^{-1}$. Cranial circumference growth is only 2–3 cm for this whole period.

When they are 6–8 years old, most children experience a slight prepubertal or midchildhood spurt. This mild acceleration (increment 1.5 cm $year^{-1}$) occurs basically at the expense of leg enlargement and coincides with an increase in suprarenal sexual hormone secretion, which may provoke an increase in GH secretion and peak (adrenarchy).

Pubertal Growth

Puberty begins with activation of the hypothalamus–pituitary–gonadal axis, by means of the coordinated action of sex steroids, GH, insulin-like growth factor 1 (IGF-1), and thyroid hormones. It is characterized by growth velocity and acceleration of development of sexual organs. After the slow, steady and prolonged growth stage which takes place in childhood, a progressively accelerating stage (growth spurt) begins, which peaks at maximum speed and then subsequently slows down. Although all body sections take part in the pubertal spurt, it affects trunk size to a larger extent than lower limbs. These stages have been expressed following a using musical system of descriptors (andante, moderato, allegro, lento, lentissimo).

The start of acceleration in pubertal growth happens 18–24 months earlier for girls than for boys. In the first year, growth velocity amounts to 7 cm $year^{-1}$ for boys and 6 cm $year^{-1}$ for girls. During the year of greatest growth (the peak), which happens for girls at 11.5 ± 1 and for boys at 13.5 ± 1 years of life girls gain 8–10 cm and boys 9–11 cm, mainly at the trunk. That is, growth velocity doubles and becomes comparable that which took place in the second year of life. Later on, there is progressive deceleration until adult size is reached. An important increase in weight velocity also occurs: at the peak it is up to 8–9 kg for boys and 7–8 kg for girls.

Before puberty, boys' and girls' size is similar. As girls' puberty begins earlier, they are slightly taller and heavier when aged 11–13. The fact that boys reach puberty 2 years later means that, when they begin it, they are 8–10 cm higher and 6–8 kg heavier than girls. Moreover, 2–5 cm and 2–4 kg must be added to the previously mentioned data because of a more intensive speed peak in boys' case.

Final Adult Height

At the end of growth, which happens 2 years earlier for girls (16 versus 18 years), boys are 10–15 cm taller and 8–12 kg heavier than girls.

Once the closure of long-bone epiphysis occurs, there is still a residual growth of 1–2 cm (no more than 1–2% of total size) which may go on until the 25th year of life, or even longer, when all the spine epiphyses are closed.

Growth Regulation

Growth is neither a constant nor regular linear process, but it is a dynamic one with a chaotic pattern that acts intermittently with static phases when growth is minimal, and jumps about, therefore showing a wide speed range in 1 year.

Determining Factors (Genome)

Growth is genetically determined by means of a polygenic device. Genetic influence on growth is less clear during the prenatal period, when birth weight depends 60% on environmental factors – basically maternal ones – with only 40% due to genetic influence. Nevertheless, the subsequent relationship of adult size with inheritance represents, under ideal health and environment conditions, 80% of

determining factors, so that environment only represents 20%. The fact that to a great extent genetic burden defines size becomes obvious in studies on twins. In the case of monozygotics the difference in their adult size is just 2.5 cm, with a correlation coefficient of 0.99, whilst in the case of dizygotics it is 12 cm, with a coefficient of 0.66. The variability rank between brothers rises to 16 cm, with a coefficient of 0.55, reaching a rank of 25 cm in the general population. Another sample of genetic influence is sexual maturation, since the difference of menarche appearance age in monozygotics is 2 months, in dizygotics is 1 year, while in the general population it is more than 1 year. The correlation between the parents' size and children's adult size is 0.55 when compared to both parents' size, 0.45 for mother and children and 0.42 for father and children. Regarding the correlation between birth size and adult size, this is low in the newborns case ($r = 0.25$–0.30), but increases rapidly and significantly so that in the second year it is already 0.80.

Regulating Factors (Hormones and Growth Patterns)

Each cell is controlled by an enormous variety of hormones, growth patterns, and their carrier proteins. This is accomplished in different ways and by different pathways: endocrine, paracrine, autocrine, juxtacrine, intracrine, and neuroendocrine ones. However, all pathways are necessary to express the whole genetic potential of growth (Table 1).

Prenatal growth regulation Hormonal regulation of fetal growth shows important peculiarities in connection with postnatal growth. GH exerts scarce influence. In contrast, human placental lactogen (hPL) plays a more important role. Although its structure is similar to GH, its growth-promoting action is lower than that of GH.

IGFs play an important role in regulating fetal growth. IGF-2, with insulin-like actions, as well as actions on fetal, placental, and uterine growth, is very important during the fetal period. IGF-1, especially in the last months of pregnancy, keeps a close correlation with birth weight and size, as shown by serum levels. Newborns suffering a delay in their intrauterine growth show significantly lower levels at birth and in their first year of life, especially those whose fetal growth delay was more intense and with a lower adjusting index. This means that IGF secretion is sensitive to nutrient availability, particularly that of energy and proteins. Other growth factors, such as neural growth factor-β (NGF-β), epidermal growth factor (EGF), and transforming growth factor-α (TGF-α), also play a role in cell growth and differentiation. Insulin has an important function, not only in protein synthesis, but also in cell transport and use of amino acids and glucose, in enzymatic induction, in lipogenic activity, and in somatic growth. It also plays a part in the regulation of IGF-1 synthesis. Thyroid hormones, glucocorticoids, and sexual hormones exert a bigger influence on the development of specific organs than on longitudinal growth. The role of thyroxine in the development of the central nervous system is to be noted (Table 2).

Postnatal growth regulation The role of hormones during the postnatal period as executors of the predetermined genetic program for growth is a crucial one. GH is essential for normal linear growth from the end of the first stage of life, when the quantity of receptors and tissular sensitivity to GH becomes evident.

Table 1 Modes of cellular regulation

Mode	*Agent*	*Source*	*Vehicle*	*Target*	*Modulators*
Endocrine	Hormone	Gland	Circulation	Heterologous tissue	Neurons Hormones Growth factors
Paracrine	Growth factor	Local site	Diffusion	Heterologous cell	Hormones Growth factors
Autocrine	Growth factor	Local site	Diffusion	Autologous cell	Hormones Growth factors
Juxtacrine	Growth factor precursor	Local site	Cell contact	Adjacent cell	Proteases
Intracrine	Hormone Growth factor Nutrient	Intracellular	Cytoplasm	Intracellular receptor	?
Neuroendocrine	Neurotransmitter Growth factor	Neuron	Axon	Heterologous tissue	Hormones Growth factors Neurons

Reproduced with permission from Russell WE (1997) Nutrition and the humoral regulation of growth. In: Walker AW and Watkins JB (eds) *Nutrition in Pediatrics*, pp. 248–271. London: B.C. Delker Inc Publ.

During puberty, the increase in sexual hormone secretion – in males this is testosterone while in females it is estrogen – stimulate GH and IGF secretion as well as their local action on growth cartilage. The pubertal growth spurt is, therefore, closely connected with the interaction of GH, IGF-1, and sexual hormones.

GH actions imply a complex interaction of factors, including cerebral neurotransmitters and hypothalamic peptides; genes, receptors, and GH-carrier proteins; IGFs and their receptors; carrier and inhibitor proteins, and nutritional signs. GH exerts direct actions on cell division, protein synthesis, lipolysis, enzymatic induction, and the stimulation of cellular transport of amino acids and glucose. Its effects in promoting longitudinal growth affect growth cartilage directly, inducing maturation of the most undifferentiated and indirect chondrocytes, facilitating the expression of IGF-1 gene, which for its part stimulates chondrocyte maturation and multiplication. The importance of GH, GHBP (growth hormone binding protein), IGF-1 and IGFBP-3 (insulin-like growth factor binding protein-3) carrier proteins is notable in the activity and function on the GH–IGF-1 axis growth (Figure 4, Table 2).

Table 2 Insulin-like growth factors and binding proteins

	Functions
IGF-I	Fetal growth and especially postnatal one Insulin-like actions Important sensitivity to undernutrition
IGF-II	Fetal growth Important insulin-like actions Limited sensitivity to undernutrition
BP-1	Transport of IGFs from vascular space
BP-2	Transport of IGFs from vascular space
BP-3	Sequesters the circulating IGF in an inactive form. Component of 150 kDa serum IGF complex intravascular reservoir
BP-4	Blocks IGF action in cells
BP-5	Potentiales IGF action in cells

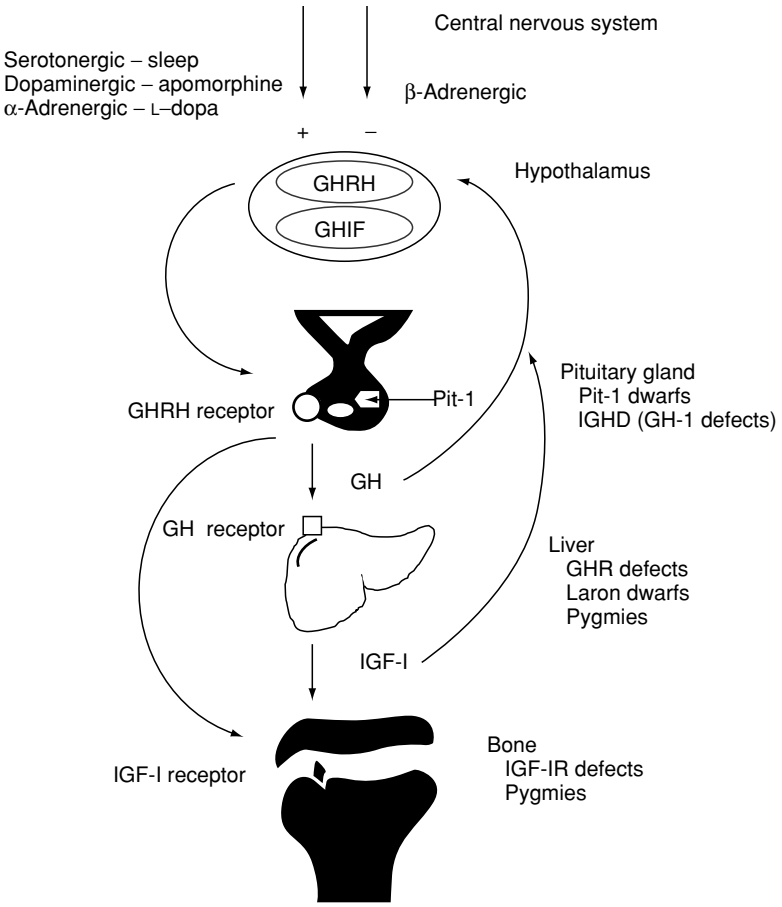

Figure 4 Regulation of growth hormone (GH) secretion: this is regulated by various factors. The sites of derangements responsible for various familial disorders of the GH axis are indicated on the right. GHRH, growth hormone-releasing hormone; GHIF, growth hormone-inhibitory factor, Pit-1, specific pituitary transacting factor; GHR, growth hormone receptor. Reproduced with permission from Mullis PE (1998) Genetic control of growth. In: Kelmer ChJH, Savage MO, Stirling HF and Saenger P (eds) *Growth disorders*, pp. 39–61. Cambridge: Chapman & Hall.

There is a close relationship between IGF-1 and GH. GH, diet, and nutritional state are the main factors regulating IGF-1 synthesis. Although the contribution of proteins and essential amino acids is important for IGF-1 regulation, it is essential in accurate energy input. Liver is the most important source of plasma IGF-1, which acts as an endocrine factor, but many other organs and tissues also synthesize IGF-1 by a local action where it is synthesized or in adjacent tissues, that is, with autocrine and paracrine action. This suggests a diffuse biological action: not only a metabolic but also a mitogenic one, with effects on cell proliferation, protein synthesis, and body enlargement processes.

Thyroid hormones have a role in bone maturation, metabolism, and central neural system (CNS) development. Apart from their direct effect on the cartilage, they are necessary in adequate GH secretion and as a complete expression of their effect on cartilage and bone formation. They stimulate IGF-1 production, while a lack of thyroid hormones restricts the effects of IGF-1 on the cartilage, independent of the effects it exerts.

Insulin favors protein and hepatic glycogen synthesies, as well as lipogenesis. In addition, it exerts an action promoting mitogenic growth mediated by IGF and other growth factors. Sexual hormones play an important role in sexual maturation and pubertal growth, and together with GH are responsible for the pubertal spurt. They exert direct anabolic action and an effect modulating GH and IGF-1 secretion and function.

Peptide growth factors Quite recently a group of messengers has been discovered inducing cell growth, proliferation, and differentiation. Unlike classical hormones, they are delivered in multiple cell types and act in more locally, exerting their actions on a wide variety of cell types or on specific tissues. Epidermal growth factor (EGF) mostly acts on cells of ectodermal origin cells and activating the mitosis of epithelial cells, conjunctive tissue, glial cells, and granulosum ovarian cells. Alfa transformant growth factor (TGF-α) shows similar actions to EGF, and shares the same membrane receptors. Platelet-derived growth factor (PDGF) is a mitogen for many types of mesodermic cells such as fibroblasts, smooth-muscle cells, and glial cells. Fibroblastic growth factors (FGF) stimulate the proliferation of cells of endodermic and mesodermic origin. Together with these growth-stimulating factors, there is an inhibitory system, which is quite unknown so far.

Permissive, External, or Environmental Factors

Nutrition is the main permissive factor, but at the same time socioeconomic, health and psychoemotional levels, environment, habitat, and iatrogenic factors will make it possible or not to achieve the highest genetic potential for growth.

Nutrition has a direct influence on the expression of genes controlling growth and metabolism. It is indispensable for the adequate function of regulating factors and, for their part, these influence nutrient transport and use. Most hormones and growth factors join cell receptors and their ability to influence the genome requires a 'cascade' of reactions, involving proteins, lipids, and ions as secondary messengers. Hormonal control of growth is highly sensitive to nutritional mean, so that intracellular availability of nutrients is the last regulator of growth. Therefore, malnutrition has important negative effects on growth, affecting the GH–IGF axis.

Together with malnutrition, low socioeconomic, and familiar educational levels, the number of children and home characteristics have a negative added influence on growth, not only on prenatal but also postnatal growth. Differences in adult size are due, to a large extent, to the growth delay accumulated during fetal stage and during the first 3 postnatal years, since this is not effectively made up later on. So, in the case of those individuals short for gestational age (SGA) – mainly in symmetric factors (size, weight, and cranial circumference within the same percentile) – who manifest a situation of prolonged malnutrition, their adult size is lower.

Any disease, especially a chronic one, interferes with normal growth and development. The degree affected depends on age at onset, duration, and seriousness. A growing social phenomenon is growth failure due to a serious emotional deprivation producing a lower GH secretion (psychosocial growth delay). An adverse familiar environment and ill treatment are two examples of this condition.

Growth Velocity Secular Trend

Over the last 150 years, since the beginning of the Industrial Revolution, growth velocity has speeded up, producing a significant increase in average adult size in developed countries. The existence of better socioeconomic, hygienic, sanitary, and nutritional conditions has had an important influence. In Spain, the socioeconomic, sanitary, and nutritional changes occurring in the twentieth century marked the evolution of an increase in secular growth. The situation in 1900–50 was that of a developing country nowadays. In 1950–99 the Spanish economic miracle took place, accompanied by a spectacular improvement in lifestyle, medical care, and the availability and accessibility of food of high nutritional value. During the period 1900–55 the height of Spanish soldiers

Table 3 Secular trend in Spain

	Mean Height of Spanish recruits (1956–1996)		
	Height (cm)		
Autonomous Community (regions) CC.AA.	*1956*	*1996*	*Δ1956–1996*
País Vasco	168.7	175.2	6.5
Cataluña	168.6	175.0	6.2
Canarias	167.3	175.1	7.8
Baleares	167.1	174.9	7.8
Madrid	166.8	175.1	8.3
Navarra	166.8	175.6	8.8
Cantabria	166.8	174.8	8.0
Valencia	166.3	174.8	8.5
Aragón	166.1	175.3	9.2
Asturias	165.9	174.5	8.6
La Rioja	165.3	175.0	9.7
Murcia	164.4	174.8	10.4
Castilla-León	164.3	174.8	10.5
Castilla-La Mancha	164.2	174.7	10.5
Andalucía	164.2	174.5	10.3
Galicia	164.2	174.0	9.8
Extremadura	164.0	174.0	10.0
Difference Between (CC.AA.)	4.5	1.6	
Spain	165.9	175.1	9.2

increased only 1.7 cm (0.3 cm per decade), whereas in 1955–95, it increased by 9.2 cm (2.3 cm per decade). A remarkable fact is that differences in adult size among recruits from different regions over the last 40 years have significantly decreased, ranging from 4.5 cm in 1956 to just 1.6 cm in 1996. Those regions which were less developed in 1955 showed a far superior increase with relation to most developed ones, as their income per capita approached the most developed regions (Table 3).

'Catch-Up,' 'Catch-Down,' and Channeling

Birth size reflects maternal environment rather than fetus genotype. Thus, during the first 2 years of life many children suffer some changes in their growth velocity as maternal factors are progressively losing influence and genetic forces determining postnatal growth pattern are gaining in importance. Children born shorter or larger than they should be according to their genetic growth potential gradually change their weight and size percentile upwards or downwards until they reach the most appropriate one, corresponding to their genome. At this point they standardize their growth velocity. These acceleration and deceleration processes, occurring during the postnatal adaptation period, are known as 'catch-up' and 'catch-down,' or 'lag-down,' respectively.

From 2 years onwards, growth becomes quite a steady process, following an extremely regular pattern. It is unusual that a healthy infant changes the channel determined by its genetic antecedants by more than one percentile. So, from this age, size is closely correlated with adult size. This pattern or growth curve has been defined as 'a guided missile' and as a ship led along a channel, so that the infant undergoes a 'canalized' growth process. 'Channeling' has been defined as the capacity of reverting to the original growth curve after moving away from its pattern because of a disease and/or malnutrition. 'Catch-up' is the recuperation growth, which can be three to four times higher than average until the original channel is reached. The earlier and longer an insult is, the more difficult complete recuperation of growth becomes.

Methods for Growth Assessment

Anthropometric Study

The assessment of somatic growth is based on the analysis of those changes produced in time in body size, shape, and composition. In order to value the most important aspects of growth, one of the most reliable methods is the anthropometric study, that is, the application of biomedical techniques expressing quantitatively body shape, which is called 'auxological anthropometry.' The validity of the method depends on a precise and exact measurement with an accurate device or material and by means of a right technique.

Height and size Body height measurement is the best parameter of skeleton growth, since, unlike weight, this is not influenced by fat or water accumulation. Over the first 2 years of life, it is preferable to measure height in the supine position. Ideally, mechanical instruments (Harpender's infantometer, ranging from 300 to 940 mm) should be used to get an exact measure. Size is used instead of height from 2 years onwards, and the chosen instrument is Harpender's stadiometer.

Cranial circumference This one of the most anthropometric measures in children, since it is related to intracranial volume and makes possible an estimate of brain growth velocity. To make this measurement, a nonextensible metal tape measure is used.

Estimate of weight and body mass Weight is not as specific a growth measure as size, since it shows body composition (water, fat, muscle, bones, viscera), that is, a mixture of tissues. But weight, mainly in infants, is a better reflection of health and nutritional state. Precision balances must be used, not only

electronic (which are better for infants), but also manual ones. They should be periodically balanced and calibrated.

Subcutaneous fat is a measure that correlates adequately with body fat content, being higher than weight for adiposity assessment. It is, therefore, an important measure of nutritional state. Triceps subcutaneous fat represents accurately limb fat, and subscapular one shows trunk fatty deposit. The assessment of the deposit of subcutaneous fat tissue must be done with a thickness compass (Holtain calliper).

The weight (kg)/size (m^2) index, or Quetelet's index, or body mass index (BMI) is nowadays considered the most accurate indicator of weight/size.

Bone maturation The most useful way to measure biological maturity is to estimate skeleton age. This is because the changes bones suffer during maturation are very similar in all individuals, and each ossification center undergoes a series of morphological changes which are easily identifiable. Conventionally, an X-ray of the left hand and wrist is used to calculate skeleton maturity simply, because there is a great proportion of long and rounded bones in an easily accessible area.

Atlas methods assess quantitatively the maturity of epiphyseal nuclei. The most widely used one is that of Greulich and Pyle. Numerical methods are ahead of atlas ones, since by transforming a simply qualitative evaluation into a quantitative phenomenon, they allow more precise assessment of the normal process of bone maturation and pathological set of symptoms by means of the mathematical analysis of data. The most widely used method is that of Tanner and Whitehouse (RUSTW2).

Adult Height Prediction

Adult height prediction according to parents' height: genetic potential for height Since genetic constitution is the main determining growth factor, parents' height is of special importance in their children's growth assessment. Each parent's height contributes equally to their children's growth, so their mean height must be a reference. Garn and Rohman, using the data in Fels' longitudinal studies, made tables of mean parental height, which are useful to predict the most probable child's height at a determined age and final height. Tanner constructed some graphics for 2–9-year-old children in which their height is expressed as a percentile of the expected standard, adjusted to mean parental height. A calculation of the expected target size based on mean parental height (genetic growth potential) is achieved like this:

For a boy :

$$\frac{\text{Father's height} + (\text{mother's height} + 13\text{cm}) \quad 10 \times 2}{2}$$

For a girl :

$$\frac{\text{Mother's height} + (\text{father's height} - 13\text{cm}) \quad 8.5 \times 2}{2}$$

Kalberg proposes a new model to estimate target height. It is a simple linear function of mean parents' size (y = estimate of child's height, x = mean parents' size (cm)). It is $y = 45.99 + 0.78x$ for boys, and $y = 37.85 + 0.75x$ for girls. This new model gets better results than the equation of corrected parental height, especially in the case of children whose parents' height is short.

Methods of adult height prediction These methods are based on, or use, skeletal maturation, not only to determine growth velocity but also to make a prediction of adult size. The three predicting models most widely used are:

1. Bayley-Pinneau's method: this uses tables from bone-age data taken from Greulich and Pyle's atlas. These are based on total growth percentage achieved at different ages: there are three subgroups according to whether bone age is advanced, slowed, or accurate in relation to chronological age. These tables can only be applied to bone aged over 6 years.
2. Roche, Wainer and Thissen's method: this uses correlation coefficients existing in different age periods between final size and height, weight, and bone age, according to Greulich and Pyle's atlas, and parental size. This method can be used for 1–16-year-old boys and 1–4-year-old girls.
3. Tanner's method: based on a mathematical equation relating a number of coefficients established for each age and sex under the following parameters: height, bone age, according to the RUSTW2 method, and chronological age. In 1983 the method was revised and more exact predicting equations were proposed.

Graphics and Limits for Normality

As human growth shows a strong tendency to remain 'channeled,' the successive points of a curve are strongly correlated. So, the best method of assessing growth is to measure it regularly and represent it on standard graphics. In this sense, periodic measurement of growth is a sensitive, easy and economical indicator of the state of health and quality of life of children and the population.

Distance graphics Distance graphics are useful to understand a specific growth situation, the height achieved at an individual's age or by the children of any community, by comparison. Distance graphics come from the study of a sample of children of different age who are examined just once ('cross method'). Therefore, it is useful to assess growth, nutrition, health or socioeconomic state, and differences and to monitor height variations in groups or in the whole society. In contrast, the longitudinal method allows the follow-up of a population sample during part or all growth periods. Each child is measured at regular intervals of time. This method is the only one that permits the analysis of individual circumstances and the inflections produced in growth velocity, as well as creating curves for growth velocity. Therefore, it is essential for clinical analysis, diagnosis, and treatment of growth disorders.

To analyze variations, increments, or inflections in the speed of children's growth, it is necessary to have velocity standards in velocity graphics. These can express growth increments experienced in a specific time interval. So, to assess velocity, height increment is divided into the time which has passed between two measurements. The time between two measurements is more easily calculated by means of a decimal calendar.

To assess if a child is within the normal variation limits for growth, percentiles or standard deviations must be used. The third and 97th percentiles define the cutting point between normal and pathological risk, and the first and 99th percentile define pathological risk. Standard deviation, calculated by means of a score standard deviation or *Z*-scoring, allows one to evaluate multiple or fraction standard deviations of a subject separated from the average, by means of the formula: $Z = a - x/\mathrm{TD}$, where a = value to compare, x = mean of standard group and TD = typical deviation. When it is $\pm$ 2 SD, the risk area starts, and $\pm$ 3 SD would be pathology.

The most widely used graphics from an international point of view are Tanner's (longitudinal) and those from the National Centre of Health Statistics (cross ones). However, the most accurate procedure will be to use individual country standards.

See also: **Bone**; **Hormones**: Adrenal Hormones; Thyroid Hormones; Gut Hormones; Pancreatic Hormones; Pituitary Hormones; Steroid Hormones

Further Reading

Barker DJP (2000) *Fetal origins of cardiovascular and lung disease*. Marcel Dekker Inc. New York.

Baxter-Jones ADG, Cardy AH and Helms DS (1999) Influence of socioeconomic conditions on growth in infancy: the 1921 Aberdeen birth cohort. *Archives of Diseases of Childhood* 81: 5–9.

Greulich WW and Pyle SI (1971) *A radiographic standard of reference for the growing hand and wrist*. Chicago: Case Western Reserve University.

Holmes R and Soothill PW (1998) Normal fetal growth. In: Kelmer ChJH, Savage MO, Striling HF and Saenger P (eds) *Growth disorders*, pp. 143–157. Cambridge: Chapmen & Hall.

Karlberg J (1989) A biologically-oriented mathematical model (ICP) for human growth. *Acta Paediatr* 350 (suppl): 70–78.

Karlberg J and Luo ZC (2000) Foetal size to final height. *Acta Paediatr* 89: 632–636.

LeRoith D (1997) Insulin-like growth factors. *N Eng J Med* 336: 633–640.

McCance RA (1976) Critical periods of growth. *Proc Nutr Soc* 35: 309–313.

Mullis PE (1998) Genetic control of growth. In Kelmer CJH and Savage MO (eds) Growth disorders. 39–61. Philadelphia: Chapman Hall Medical.

Prader A, Tanner JM and Von Harnack GA (1963) Catch-up growth following illness or starvation: an example of developmental canalization in man. *J Pediatr* 62: 646–659.

Russell WE (1997) Nutrition and the hormonal regulation of growth. In: Allan Walker W and Watkins Jb (eds) *Nutrition in Pediatris*. 268–271. London: B.C Delker Inc Publ.

Tanner JM (1962) Growth at adolescence. 2nd. ed Oxford: Blackwell.

Tanner JM (1978) *Fetus into man: physical growth from conception to maturity*. Open Books, London.

Tanner JM Whitehouse RH and Cameron N (1983) *Assessment of skeletal naturity and prediction of adult height. (TW2 Methods)*. London: Academic Press.

Tojo R (2001) *Tratado de Nutrición Pediátrica*. Barcelona: Doyma.

Tojo R and Leis R (2001) Crecimiento normal. In: Cruz M (ed.) *Tratado de Nutrición Pediátrica*, 8th edn, pp. 373–394. Madrid: Ergon.

Tojo R, Leis R, Iglesias Diz JL and Pombo M (1996) Secular growth acceleration. Maturation and final adult height in Galicia and Spain. 1900–1993, In: Battaglia F, Falkner F and Garza C (eds) *Maternal and extrauterine nutritional factors their influence on fetal and infant growth*. Ergon, Madrid 21–36.

GUAVAS

S Yusof, Universiti Putra Malaysia, Selangor, Malaysia

Introduction

Guava has long been established in many tropical and subtropical countries of the world, although its origin was from Central America. Much of the interest in guava has been due to its extremely high vitamin content. Nutritionally, it is a healthy fruit, containing relatively high amounts of vitamin C and total fiber. Guava has the potential to become a commercially important tropical fruit crop not only for processing, but also for fresh consumption. Improved propagation techniques have produced fruits with better quality and thus increased the commercial importance of the fruit in many countries, especially in the Far East.

Morphology and Anatomy of Fruits

Guava is the most important fruit plant in the large family, Myrtaceae, which also includes strawberry guava, *Psidium cattleianum sabine*; the rose apple, *Engenia jambos*; Surinam cherry, *Engenia uniflora*; the mountain apple, *Engenia malaccensis*; the Java plum, *Engenia cumini* Merr.; and the spices cinnamon, clove, allspice, and nutmeg. Guava is classified under the genus *Psidium* which contains 150 species, but only *Psidium guajava* has been exploited commercially. Guava is a shrub but under high moisture conditions can grow up to 6–9 m in height and spread, with trunk diameters of 30 cm or more. The trunk is short, freely branching from the base. Branches are pliable and hence are rarely broken by winds. Leaves are usually opposite and simple, arranged in pairs; oblong or oval in shape measuring about 10–18 cm in length; smooth on the upper surface but finely pubescent on the undersurface and prominently veined. Flowers occur singly or in clusters of two to three at the leaf axils of current and preceding growth. The flowers are bisexual or called perfect flowers. They are white in color, 2.5–3.5 cm in diameter, with four or five petals, numerous stamens, and one style. Flowers open between 5 and 7 a.m., depending upon the cultivar and morning temperatures. The tubular calyx encases the bud and splits into four or five segments at anthesis. The ovary is inferior with four or five carpels, each containing numerous ovules in axile placentation. The flowers can be self-pollinated, but cross-pollination takes place most frequently, aided by bees and pollen-carrying insects. After fertilization, the flowers will develop into mature fruits.

Botanically, the fruit is a berry or capsule. It is a many-seeded berry, varying in size from 2.5 to 10 cm in diameter and from 50 to 500 g in weight. The shape of the fruit can be globose, ovoid, elongated, or pear-shaped. The color of the skin of some varieties may turn yellow when ripe; the flesh color may be pink, salmon, white, or yellow. Skin texture may be smooth or rough. The inner walls of the carpels are fleshy and of varying thickness and seeds are embedded in the pulp. The fruits may have thick flesh with only a few seeds in a small center cavity, or thin flesh with numerous seeds embedded in a large mass of pulp. The seeds are small, bony, kidney-shaped, flattened, light yellow or yellowish brown, measuring about 3–5 mm long by 2–3 mm wide. It occupies about 1.6–4.4% of the weight of fruit. Fruit growth follows a simple sigmoid curve and pulp growth parallels total fruit growth. The days from anthesis to harvest can vary from about 120 to over 220 days, depending upon temperature during fruit development. The variation between different cultivars can be up to 60 days.

Cultivars

Guavas can be classified into three types depending on their end use, namely the dessert type, the processing type, and those suitable for both dessert and processing (Table 1). In India a low-acid dessert cultivar, such as Allahabad Safeda an 'apple color,' with white flesh was developed. South Africa's Malherbe and Fan Ratief are mild, sweet, dessert types, with light pink color. They are also suitable for canning as halved fruit. For processing, Beaumont, Ka Hua Kula, and Waiakea are recommended in Hawaii. Beaumont was the first processing cultivar introduced to Hawaiian industry and was the only recommended processing guava until the introduction of Ka Hua Kula. Beaumont produces fruits ranging from 145 to 235 g, averaging about 170 g. Fruits are mildly acid with total titratable acidity ranging from less than 1.0% during the summer to about 1.25% in winter. Soluble solids range from about 7% in winter to 10% during summer. Fruit characteristics of Ka Hua Kula are similar, with a slightly stronger pink color and somewhat higher acidity. Both varieties are good yielders, exceeding 227 kg tree^{-1} year^{-1}. Florida has developed both dessert and processing cultivars of excellent quality. Ruby, Supreme and the hybrid Ruby × Supreme, are excellent dessert types. Pink Acid and Patillo are

two acid cultivars with dark pink color suitable for processing. Other varieties that have been developed in various countries are as shown in Table 1.

Certain guava cultivars exhibit a change of skin color from green to yellow during ripening, whereas in others the color remains green and this is dependent on the cultivar. The flesh color may change from white to either creamy white, yellowish pink, deep pink, or salmon-red upon ripening. The presence of carotenoids, lycopene, and β-carotene may influence the intensity of the flesh color of ripe fruit. Fruit texture may differ between cultivars and it varies with position of the tissue. For example, the Beaumont cultivar is significantly softer when ripe, and its firmness decreases more rapidly during ripening than that of the Vietnamese guava. Irrespective of the cultivar type, outer mesocarp tissue of the fruit is consistently firmer than inner mesocarp tissue.

Fruiting Cycle

The fruiting peaks can deviate with prevailing weather conditions and cultural practices. Since flowers are produced on new growth, factors that stimulate new growth, such as irrigation, fertilization, pruning, and defoliation stimulate flowering. The practice of pruning, fertilizing, and irrigation at the end of harvest is essential. Pruning is repeated every 8–9 months. Initial flowering after pruning occurs in about 5–6 weeks. Defoliation can be accomplished using urea plus ethephon and a detergent. Following defoliation, new leaves appear in about 3–4 weeks, with a peak of flowering occurring 9–12 weeks after defoliation. This is followed by fruit set and a peak of abscission of young fruit. The percentage fruit abscission is correlated with number of fruitlets set per tree. In Hawaii, a small harvest peak occurs between April and November.

Guava plant grows best in warm areas with abundant moisture. The ideal rainfall pattern for guava is alternating dry and wet conditions. The optimum temperature is reported to be between 23 and 28 °C. Temperatures lower than 23 °C and higher than 27 °C during flowering would reduce fruit set significantly.

Chemical Composition

The chemical composition of guava fruits is shown in Table 2. As in most fruits, moisture constitutes a fairly large portion of guava fruit (85%); the energy content is relatively high (230 kJ); the protein and fat contents are low (0.3% and 0.1%, respectively). Carbohydrates are the principal nonaqueous constituents of guava. Of the total carbohydrates (15%), 5.82 g are the sugars fructose, glucose, and sucrose.

Table 2 Proximate composition of common guava, with seeds removed

	Units	*Amount in 100 g edible portion*
Proximate		
Moisture	%	85
Energy	kJ	230
Protein	g	0.3
Fat	g	0.1
Carbohydrate	g	15
Fiber	g	2.4
Ash	g	0.5
Minerals		
Calcium	mg	15
Phosphorus	mg	16
Iron	mg	0.3
Potassium	mg	292
Sodium	mg	6
Vitamins		
Vitamin A	IU	109
Thiamin	mg	0.06
Riboflavin	mg	0.06
Niacin	mg	1.3
Ascorbic acid	mg	190

Table 1 Characteristics of some guava varieties and their use

Variety	*Origin*	*Size*[a]	*Shape*	*Flesh color*	*Seeds*	*Acidity*[b]	*Use*
Vietnamese[c]	Vietnam	Large	Ovoid	White	Few	Low	Dessert
Allahabad Safeda	India	Medim	Round	White	Few	Low	Dessert
Fan Ratief	South Africa	Medium		Light pink	Medium	Medium	Dual purpose
Lucknow-49	India	Large	Round	White	Few, soft	Low	Dessert
Beaumont	Hawaii	Large	Ovoid	Pink	Medium	High	Processing
Ka Hua Kula (097)	Hawaii	Large	Ovoid	Pink	Few	High	Processing
Hong Kong Pink	Hawaii	Medium		Pink	Few	Low	Dessert
Ruby x Supreme	Florida	Large	Ovoid	White	Few	Low	Dessert
11–56	Australia	Medium	Conical	Pink	Medium	High	Processing
11–56 (T3)	Australia	Medium	Round	Strong red-pink	Low-medium	Medium	Dual purpose

[a]Small, < 100 g; medium, 100–200 g; large, > 200 g.
[b]Low, < 0.5% anhydrous citric acid equivalent; medium, 0.5–1% anhydrous citric acid equivalent; high, > 1.0% anhydrous citric acid equivalent. Adapted from Batten (1983).
[c]Yusof S (1989) Characteristics and potential use of guava (*Psidium guajava* L.) for processing and storage. PhD thesis. Universiti Putra Malaysia. Serdang, Selangor, Malaysia.

Fructose is the predominant sugar, constituting about 58.9%, followed by glucose (35.7%), and sucrose (5.3%). During development and maturation the sugars increase gradually and then decrease in overripe fruits (Figure 1). The fiber and ash contents are considered high, with values of 2.4 and 0.5 g respectively. Refer to individual nutrients .

Nutritional Value

Guava is an excellent source of ascorbic acid (Table 3). It is mainly located in the skin and a slightly lower concentration is found in the flesh. The content of ascorbic acid, which is also cultivar-dependent, varies from as low as 60–100 mg 100 g^{-1} in some cultivars to about 200–300 mg 100 g^{-1}, or to even higher levels (800–1000 mg 100 g^{-1}) in other cultivars. Ascorbic acid level generally increases with ripening, and also with decreasing depth of the mesocarp tissue. The fruit has an appreciable amount of minerals such as calcium, phosphorus, iron, potassium, and sodium as well as vitamins like vitamin A, thiamin, riboflavin, and niacin. (*See* **Vitamins**: Overview.)

Table 3 Ascorbic acid content of several guava varieties

Location	*Variety*	*Ascorbic acid (mg 100 g^{-1})*
Florida	Red Cattley	29.1
Florida	Yellow Cattley	39.1
Florida	Common Cattley	23–486
Hawaii	Red or white Cattley	25–50
Hawaii	Common	96–306
California	Common	50–352
Puerto Rico	Common	202–442
India	Country	299
India	Hill	11–19
Australia	Large Yellow	110

Adapted from Jagtiani J, Chan H Jr and Sakai W (1988) Guavas. In: *Tropical Fruits Processing*, pp. 9–43. New York: Academic Press.

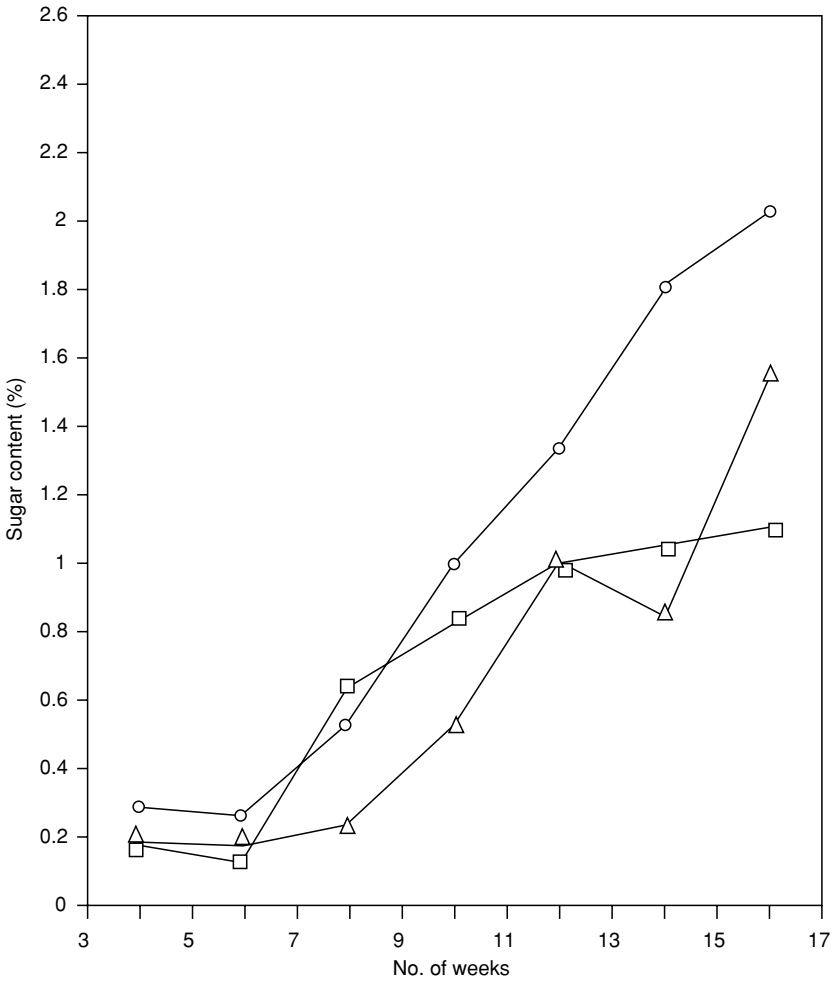

Figure 1 Sugar content of Vietnamese fruits during development and maturation. Circles, fructose; squares, glucose; triangles, sucrose. From S Yusof (1989) Characteristics and potential use of guava (*Psidium guajava* L.) for processing and storage. PhD thesis. Universiti Putra Malaysia. Serdang, Selangor, Malaysia.

Important Biochemical Composition

Acid Content

Glycolic, malic, ascorbic, and citric acids were the major organic acids detected in guavas. Also present is fumaric acid. Total acid level calculated as citric acid ranges from 0.2 to 1.1% fresh weight. The level of ascorbic acid is quite substantial, those of malic and glycolic acids are relatively low, while fumaric acid is only found in trace amounts. The presence of these acids in guava is responsible for the tart flavor and also for its relatively low pH of 3.2–4.1. (*See* **Acesulfame/Acesulphame**.)

Flavor

Flavor and aroma vary widely among the different cultivars. There are low-acid, sweet types, bland types that are low in both sugars and acidity, and high-acid types. The musky aroma is more pronounced in the fully ripened, low-acid, sweet types. The flavor of guava is determined by the types and amounts of sugars, acids, phenolics, and volatile compounds that are present in the fruit. The proportions of these chemical constituents may vary with the fruit age and with the cultivar. Depending on the cultivar, these flavor compounds may accumulate to different extent during ripening and thus may result in guava fruits having a distinctive aroma and taste.

The volatile flavor constituents of guava have been isolated and separated. There are 22 compounds in Hawaiian guavas, of which methyl benzoate, hexanol, *p*-phenylethyl acetate, methyl cinnamate, cinnamyl acetate, and β-ionone are believed to play predominant roles in the flavor and odor of the fruit. More recently, 154 compounds were identified, of which 116 were described in guava for the first time. Later on, 122 volatile compounds in guava were identified. Among the identified compounds were 13 aldehydes, 17 ketones, 31 alcohols, 10 acids, 28 esters, 10 hydrocarbons, and 13 miscellaneous compounds.

Tannins

Tannins or polyphenols are another group of compounds that have been considered in many fruits and lately in guavas. The function of these compounds in plants is unknown, but their ability to combine with proteins has often caused problems for food scientists. Their role in the browning of fruit and fruit products, and a cause of cloudiness in many fruit juices, has long been recognized. Tannins have also been associated with astringency in many fruits, especially in less ripe ones. Changes in polyphenols and polyphenoloxidase activity have been reported during fruit development in guavas. There were decreases in seven categories of phenolics (total phenols, nontannin phenols, tannin phenols, hydrolyzable tannins, simple phenolics, nontannin flavans, and condensed tannins) during ripening; insoluble leucoanthocyanins were the major components. At the mature stage, polyphenolase activity was barely detectable in both pink and white-fleshed guavas. However, at the ripe and fully ripe stages, polyphenolase activity increased dramatically, and this change was stated to be responsible for the decrease in polyphenolics during ripening and the subsequent decline in astringency. (*See* **Dietary Reference Values; Tannins and Polyphenols**.)

Pectin

Several workers demonstrated the potential of guava as a source of pectin. Pectic substances are contained in the intercellular spaces of cell walls where they function as bonding and lining materials. The significance of pectin is in contributing to the rigidity of plant tissues. The pectin content in guavas is highest in the mature stage. The value decreases in the ripe and overripe stage. The methoxyl content is an important factor in evaluating the gelling property of pectin. The values obtained from fresh white and pink cultivars were 3.23% and 3.03% respectively. The equivalent weight of guava pectin varied from 2181 to 3173, which is considerably higher than that reported for mangoes. The presence of pectin in guava impart a viscous property to the guava purée or juice. The pectic substances have been shown to combine with protein or polyphenolic compounds to form a cloud suspension. During the extraction of juice from fruit, pectin is released from the cell wall and since the pH of the juice is lower than the isoelectric point of protein, the positively charged protein is surrounded by the negatively charged pectin. Due to the interruption of the negative charge of the outer pectin molecule, the cloud suspended in the juice is more stable. The pectin contains neutral sugars, rhaminose, arabinose glucose, and fructose in branched chains that can bind to the protein with a hydrogen bond and make the cloud more stable. The content in guava purée was found to be between 5 and 8%.

Stone Cells

Guava contains a considerable amount of stone cells which contribute to an undesirable gritty texture to the processed purée. The number of stone cells that is tolerable varies from region to region. In Brazil, the presence of stone cells is considered a positive quality factor. According to a recent report, there are two types of stone cells in guava: an irregular-shaped type, abundant under the epidermis, and a smooth type found in the core region of the fruit.

Stone cells are developed from parenchyma cells by secondary thickening of the wall. Cell differentiation occurs when fruits are about 14 weeks old. As the fruits increase in maturity, cell wall thickening appears to be more prominent. The stone cells are found in clusters, each measuring approximately 300 μm in length. The presence of stone cells in the flesh of fruits has no specific function. In leaves or other parts of the plant, stone cells may help to give rigidity to the plant part or may act as a protective covering, as in the testa of seeds. The composition of stone cells is reported to be as follows: 0.92% fat; 1.05% ash; 1.50% protein; 37.1% lignin; 53.9% cellulose; 5.49% soluble carbohydrates.

Harvesting

Guavas produce varying amount of fruit throughout the year in the tropics. Fruit yields depend upon cultivar potential, plant density, weather conditions, and all the factors involved in management. On the average the yield ranges from 25 to 40 t ha^{-1} $year^{-1}$. Guava fruit generally takes about 17–20 weeks from fruit-set to reach maturity. The recommended optimum stage for harvesting is about 2–3 weeks before attaining full growth, because guava fruits will experience changes associated with ripening such as skin yellowing, and decrease in tissue firmness and in detachment force while they are still actively growing. It was reported that the Ka Hua Kula guava entered the ripening phase when the fruit was only half the maximum size while in Allahabad Safeda ripening changes occurred towards the end of the growing period. For green ripe guava cultivars, fruit size, moisture content, or chemical attributes such as total soluble solids, titratable acidity, sugar and tannin contents, may also be adopted as additional maturity indices.

Guava fruit must be harvested with great care because of their soft and thin skin. Harvesting is normally done manually to avoid physical injuries, although tests with machine harvesting have been conducted in some countries. Subsequent handling and transportation of the fruit also need extra precautions to reduce bruising. Dessert fruit are hand-harvested mature green and carefully handled to avoid injury, graded to size, and packed carefully in cartons for shipment. Processing fruit should be picked at the firm yellow to half-ripe stage. Overripe fruit and those severely infected with fruit flies and diseases should be destroyed rather than left to fall and rot in the field. When only fully ripe fruits are harvested on a 3-day cycle, losses between 35 and 40% can occur. Harvesting fruits showing some yellow to the half-ripe stage allows the interval between harvest cycles to be lengthened to about 3 days and this would minimize losses. Fruits are harvested into plastic buckets or packing bags worn by the harvester with bottom delivery. The fruits are transferred into larger bins or small wooden boxes and placed in the shade until delivered to processing plant.

Storage and Handling

The common guava (*P. guajava*) is a climacteric fruit. If full-size pale-green fruits are stored at 20 °C, the peak in carbon dioxide and ethylene production occurs about 5–6 days after harvest. Biochemical and color changes accompany the ripening process. Storage of two processing guava cultivars at 20 °C, showed that the fruit developed eating ripe quality after 5 days. After 7 days at 20 °C, the skin was discolored and some rotting occurred. Storage at 0–10 °C extended the postharvest life for about 2 weeks. Storage at 0 °C reduced rotting, but caused chilling injury to the pulp, resulting in darkened flesh. The optimum storage temperature was 5 °C since pulp injury occurred least at that temperature. Several trials have been carried out to prolong the storage life of guavas. Recent experiments indicate that the storage life of mature fruit may be extended up to 1 week at room temperature by using modified atmospheres. Waxing accompanied with low-temperature storage (8–10 °C) increased storage life to 21 days. The use of carnauba-paraffin emulsion imparted a gloss to the fruits that is normally lost during storage. Fully ripe guava should be processed without delay, but if necessary, they can be held for about a week at 2–7 °C with only a small loss in vitamin C content. Storage at 0 °C for up to 2 weeks is used commercially in Hawaii.

Guava fruit is highly perishable; therefore, it is important to select the right type of fruit and the right packaging material for transport. Shallow wooden boxes should be used to protect the fruit from external hazards and from internal crushing of lower fruits if containers are too deep. For long-distance shipments, soft filling or cushioning materials should also be used. Good ventilation is important to prevent build-up of heat and humidity, which promotes microbial spoilage. If ripe fruits need to be shipped, it might be better to hold them in a cool or shady area and transport them during the cooler night temperatures. The best practice would be to ship unripe fruit and ripen it under controlled conditions at the processing plant.

Diseases

Some diseases and causal organisms are specific to certain countries and others are widespread where

guavas are grown (Table 4). Anthracnose is widespread and is considered an important disease in most countries. Algal spots are very common but are not usually serious, except they are of concern in fruits for dessert. Other types of fruit rots are attributed to a number of organisms. *Guignardia* fruit rot becomes serious in Hawaii when fruit is left to overripen on the tree or on the ground. Wilting of guava trees is reported from South Africa and India and attributed to different organisms.

Mucor fruit rot first appears as a water-soaked area and later becomes covered with yellowish, fuzzy mycelia and fruiting bodies. Infection rate can be as high as 80–90% and, as a wound parasite, it is commonly associated with fruit fly oviposition wounds. Culture control is possible by removing fallen fruit from the field at 2–4 days' intervals. Low-acid sweet cultivars are more tolerant to this disease than acid types. Blossom end rot of fruit appears to be widespread. In Hawaii, no organism has been isolated and fungicidal sprays have been ineffective. Calcium application to guavas largely alleviates this disease.

Pests

The types of pests that are common to all guava growing areas are as indicated in Table 5. Report of work done in Hawaii found a list of approximately 45 species of insects and six species of mites attacking guavas. Included among them are species of aphids, thrips, scales, mealy bugs, beetles, moth larvae, false spider, eriophyid, and spider mites. There are several species of parasitic wasps and predators that keep scale insects and mealy bugs under reasonable control. Fruit flies cause serious damage and fruit rot within a day or two upon ripening. Fruit bagging along with thinning 30–40 days after anthesis can significantly reduce the problem and produce high-quality blemish-free dessert fruit.

Thrips can cause silvering of leaves and scarification of the fruit. When young developing fruits are severely damaged, they often fail to develop and become mummified. Natural enemies can keep thrips under fair control, although outbreaks do occur, especially during the fruiting season. Skin of scarred fruit becomes russeted by disruptions of the

Table 4 Some important diseases of guava fruit

Common name	*Organism*	*Parts affected*	*Region*
Anthracnose	*Coiletotrichum Gloeosporioides*	Fruit	Worldwide
Glomereila fruit rot	*Glomereila cingulata* (perfect stage)	Fruit	Puerto Rico
Blossom end rot	*Botrytis cinerea*	Fruit	South Africa
	Phomopsis psidii	Fruit	India, South Africa
	Physiological	Fruit	Hawaii, Australia
Fruit canker	*Pestalotia*	Fruit	Australia
Fruit rot	*Macrophomina* spp.	Fruit	Caribbean
Guava fruit rot	*Rhizopus stolonifer*	Fruit	Hawaii
Guava wilt	*Gliocladium* spp.	Root	South Africa, Australia
Fusarium wilt	*Fusarium solani*	Root	India
Guignardia rot	*Guignardia* spp.	Fruit	Hawaii
Mucor rot	*Mucor hiemalis*	Fruit	Hawaii
Algal spots	*Cephaleuros virescens*	Leaf, fruit	Florida, Hawaii

Adapted from Nakasone H and Paull R (1998) Guava. In: *Tropical Fruits*, pp. 149–172. Wallingford, Oxon: CAB International.

Table 5 Some important insect and nematode pests of guava

Common name	*Organism*	*Parts affected*	*Region*
Mediterranean fruit fly	*Ceratitis capitata*	Fruit	Hawaii, South Africa
Natal fruit fly	*Ceratitis rosa*	Fruit	South Africa
Oriental fruit fly	*Dacus dorsalis*	Fruit	Hawaii
Melon fly	*Dacus cucurbitace*	Fruit	Hawaii
Caribbean fruit fly	*Anastrepha striata*	Fruit	Caribbean, American tropics
No common name	*Monolepta australis*	Leaf	Australia
Red-banded thrips	*Selenothrips*	Leaf, fruit	Universal
Bark-eating caterpillar	*Indarbela quadrinotata*	Bark	India
Root-knot nematode	*Meliodogyne incognita*	Root	Australia
	Meliodogyne aruenaria	Root	Caribbean
	Meliodogyne acrita	Root	Caribbean

Adapted from Nakasone H and Paull R (1998) Guava. In: *Tropical Fruits*, pp. 149–172. Wallingford, Oxon: CAB International.

epidermal layer and corky tissue development in the subepidermal area. This corky tissue must be screened out during the processing; otherwise the purée from russeted fruit appears muddy pink in color. Several sweet cultivars such as Allahabad Safeda, Ruby × Supreme and Lucknow-49 have a higher degree of resistance to thrips than others. This resistance increases somewhat with increasing foliar levels of nitrogen and potassium.

Industrial Use

Guava is one of the easiest fruits to process since it does not show many problems of physical or biochemical nature in relation to texture and shape of fruit or pulp browning during processing. However, compared to other fruits, guavas have found very limited use in the fruit-processing industry. Guava purée, which is also known as guava pulp, is a liquid product prepared by puréeing or pulping whole guavas. A guava-puréeing process, which was developed in Hawaii, and is in commercial use today is similar to that shown in Figure 2. Fruits for processing are placed in a dump tank which serves to soak and wash the fruit. The fruit which then floats is picked up on to a moving conveyer belt; it is inspected and sorted for decay, insect damage, and foreign materials such as leaves and dirt. Clean fruits are then passed through a chopper or slicer to be broken up, and are then fed into a pulper. The pulper removes the seeds and fibrous tissues and forces the remainder of the product through a perforated stainless-steel screen. The puréed material coming from the pulper is next passed through a finisher which removes the stone cells from the fruit. The finisher is equipped with a perforated screen of approximately 0.051-cm-sized holes. Alternatively, the guava purée can be passed through a mustard mill which effectively grinds up the stone cells so that smooth guava purée is obtained. However, the incorporation of an excessive amount of milled stone cells into the purée can discolor it. Other methods of removing stone cells include the use of a finer screen in the finisher, or using a centrifuge. The average yield of purée is approximately 85%. The prepared purée may be used directly for preparation of various guava beverages, or frozen for export. The purée is most commonly manufactured into nectars, various juice drink blends, syrups, icecream toppings, jams, and jellies.

See also: **Acesulfame/Acesulphame**; **Dietary Reference Values**; **Sensory Evaluation**: Aroma; Taste; **Tannins and Polyphenols**

Further Reading

Chyan C, Chen S and Wu C (1992) Different volatile and non-volatile constituents between mature and ripe guava (*Psidium guava* L) fruits. *Journal of Agriculture and Food Chemistry* 40: 846–849.

Itoo S, Yamaguchi T, Oohata J and Ishihata K (1980) Studies on qualities of subtropical fruits. II. Ascorbic acid and stone cells of guava fruits (*Psidium guava* L). *Bulletin of the Faculty of Agriculture, Kogoshima University* 30: 47–54.

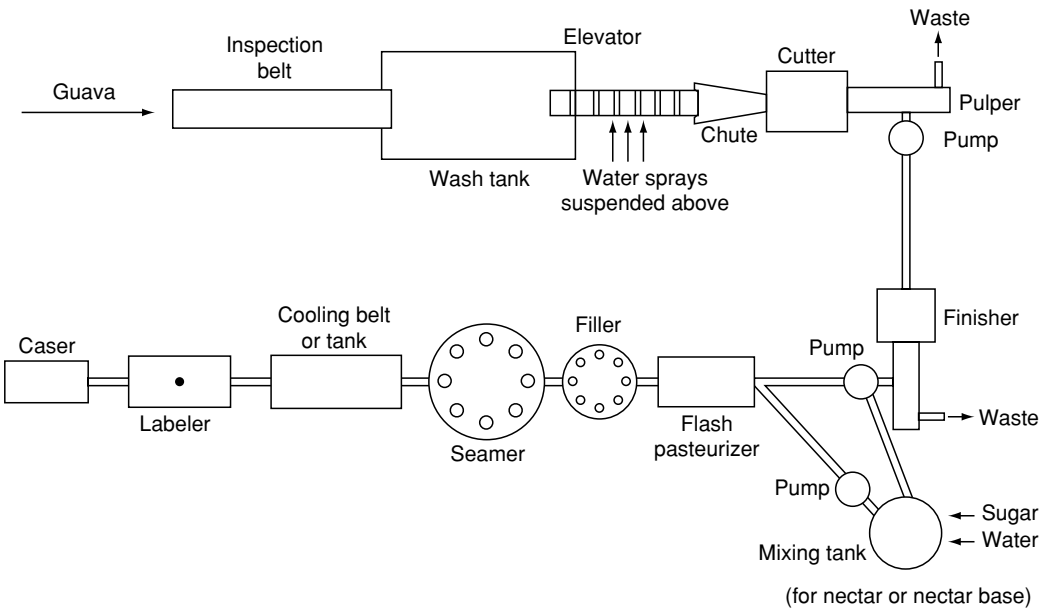

Figure 2 Flow sheet for guava-processing line (for frozen products, bypass pasteurization and cooling belt). Adapted from Boyle F, Seagrave-Smith H, Sakrata S and Sherman G (1957) Hawaii Agric. Exp. Stn., Bull. 111.

Lim K and Khoo C (1990) *Guava in Malaysia. Production, Pests and Diseases*. Kuala Lumpur, Malaysia: Tropical Press.

Mowlah G and Itoo S (1983) Changes in pectin components, ascorbic acid, pectin enzymes and cellulase activity in ripening and stored guavas. *Nippon Shokubin Kogyo Gakkaishi* 30: 454–461.

Nelson PE and Tressler DK (1980) *Fruit and Vegetable Juice Processing Technology*, 3rd edn. Westport, Connecticut: AVI.

Nishimura O, Yamaguch K, Mihara S and Shibamoto T (1989) Volatile constituents of guava fruits (*Psidium guajava* L) and canned puree. *Journal of Agriculture and Food Chemistry* 37: 139–142.

Wilson C (1980) Guava-composition, properties and uses. In: Nagy S and Shaw P (eds) *Tropical and Subtropical Fruits*. Westport, Connecticut: AVI.

Yen G, Lin H and Yang P (1992) Changes in volatile flavour components of guava puree during processing and frozen storage. *Journal of Food Science* 57: 679–681, 685.

Yusof S (1989) Characteristics and potential use of guava (*Psidium guajava* L.) for processing and storage. PhD thesis. Universiti Putra Malaysia. Serdang, Selangor, Malaysia.

Yusof S and Mohamed S (1987) Physico-chemical changes in guava (*Psidium guajava* L.) during development and maturation. *Journal of the Science of Food and Agriculture* 38: 31–39.

GUMS

Contents

Properties of Individual Gums

P A Williams and G O Phillips, The North East Wales Institute, Wrexham, UK

Introduction

The term 'gums' is commonly used to describe a range of polysaccharides that are nowadays widely used in food products to perform a number of functions including, thickening, gelling, stabilization of foams, emulsions, and dispersions, inhibition of ice and sugar crystal formation, and in the controlled release of flavors. The commercially important gums and their source of origin are given in **Table 1**.

The choice of gum for a particular food application is dictated by the functional characteristics required but is inevitably influenced by price and security of supply. It is for these reasons that starches are the most commonly used thickening agents. It is interesting to note, however, that xanthan gum, since its introduction in the early 1970s, has become the thickener of choice in many applications, despite its high price. This is due to the fact that xanthan gum has unique rheological behavior (see later). Gelatin (a protein but commonly considered as a gum) is by far the most widely used gelling agent. However, with the increasing demand for nonanimal products and in particular the recent bovine spongiform encephalopathy (BSE) outbreak in the UK, prices have increased significantly over the last few years. There is currently much activity in the development of a gelatin replacement. The carrageenan market has also been unstable over recent years due to the introduction of cheaper lower-refined grades (processed euchema seaweed or PES) which can compete effectively with the traditional purified grades in applications where gel clarity is not important. The gum arabic market has been particularly erratic due to extreme price fluctuations and security of supply and much effort has been directed at finding alternatives. A number of starch-based substitutes are now available. Gellan gum was approved for food use in Japan in 1988 and later in the USA. More recently it has been approved for food use in Europe and is beginning to establish its own niche market. A summary of the functional properties of the individual gums and the value and volume of the gums market is given in **Table 2**.

Table 1 Source of commercially important gums

Botanical
Trees
Cellulose
Microcrystalline cellulose
Sodium carboxymethyl cellulose
Methyl cellulose
Hydroxypropylmethyl cellulose
Tree gum exudates
Gum arabic (*Acacia senegal, A. seyal*)
Gum karaya (*Stercula urens*)
Gum ghatti (*Anogeissus latifolia*)
Gum tragacanth (*Astragalus*)
Plants
Starch (corn, maize, waxy maize, wheat, pea, potato, rice, sago, etc.)
Pectin (citrus fruits, apple pomace)
Cellulose (cotton)
Seeds
Guar gum (*Cyamopsis tetragonoloba*)
Locust bean gum (*Ceratonia siliqua*)
Tara gum (*Cesalpinia spinosa*)
Tubers
Konjac mannan (*Amorphophallus konjac*)
Algal
Red seaweeds (Rhodophyceae)
Agar (*Gelidium* and *Gracilaria* spp.)
Carrageenan (*Euchema cottonii*, *E. spinosum*, *Chondus crispus*, *Gigartina* sp.)
Brown seaweeds (Phyophyceae)
Alginate (*Laminaria hyperborea*, *Macrocystis pyrifera*, *Ascophyllum nodosum*) Propylene glycol alginate
Microbial
Xanthan gum (*Xanthomonas campestris*)
Curdlan (*Alcaligenes faecalis* var. *myxogenes* 10C3)
Gellan gum (*Auromonas elodea*)
Animal
Gelatin (collagen from animal skins and bones)
Chitosan (shells of shrimps, crabs)

Functional Characteristics

Gums will dissolve or swell in water, although in many cases high temperature and vigorous agitation are required before complete dissolution is achieved. The solutions formed are usually thick and viscous, even at very low concentrations (1%). There are, however, notable exceptions, for example gum arabic solutions only become viscous at concentrations >50%. The viscosity of polymer solutions generally shows a marked increase at a critical polymer concentration, commonly referred to as C^*, which corresponds to the concentration at which the polymer molecules start to touch and interpenetrate due to molecular crowding. C^* decreases with increasing hydrodynamic volume of the gum molecules which in turn is influenced by their molecular mass, shape, and the presence of charged groups. Hence high-molecular-mass linear molecules possessing ionizable groups tend be good thickeners. The main thickeners used in food products are listed in Table 3.

Gum solutions normally exhibit Newtonian behavior at concentrations well below C^*, i.e., their viscosity is independent of the rate of shear. However, above C^* non-Newtonian behavior is usually observed and the viscosity–shear rate profiles show three distinct regions, i.e., a high-viscosity low-shear Newtonian plateau (at shear rates typically $< 1\,s^{-1}$); a shear thinning region ($\sim 1 - 1000\,s^{-1}$), and a low-viscosity high-shear Newtonian plateau (typically $> 1000\,s^{-1}$).

Some gums, notably xanthan gum, have a tendency to undergo weak intermolecular chain association in solution, leading to the formation of a three-dimensional network structure, thus giving rise to a very high viscosity. The junction zones formed can be readily disrupted even at very low shear rates, resulting in a dramatic drop in viscosity. Other polysaccharides self-associate and form stable junction zones and as a consequence strong gel structures are produced. The gels are referred to as 'physical gels' because the junction zones are formed through physical interaction, for example, by hydrogen bonding, hydrophobic association, and cation-mediated cross-linking and differ from synthetic polymer gels which normally consist of covalently cross-linked polymer chains.

Some gums form thermoreversible gels and examples exist where gelation occurs on cooling or heating. Some form nonthermoreversible gels. In such cases gelation may be induced by cross-linking polymer chains with divalent cations. Gels may be optically clear or turbid and a range of textures can be obtained. Gel formation occurs above a critical minimum concentration, which is specific for each gum. Agarose, for example, will form gels at concentrations as low as 0.2%, while for acid-thinned starch, a concentration of ~15% is required. Typically, however, concentrations of less than 1% are required. Gel strength increases with increasing concentration. The main gelling agents are listed in Table 4.

Mixtures of gums are commonly used to impart novel textural characteristics to food products and an added incentive is a reduction in costs. Classic examples include the addition of locust bean gum to kappa-carrageenan to yield softer, more transparent gels and also the addition of locust bean gum to xanthan gum to induce gel formation.

Gums are effective at either inducing or preventing flocculation in particulate dispersions. If insufficient gum is present to coat all the particles fully, aggregation occurs owing to polymer adsorbing on to two or more particles simultaneously, thus inducing bridging. If the particles are fully coated, particle aggregation is prevented by steric repulsion, which arises as a result of interaction between the adsorbed polymer layers. Although nonadsorbed gum is

Table 2 Value and volume of the gums market

Hydrocolloid	*Principal function*	*Thousand tonnes (1998)*	*$ million (1998)*
Agar	Gelling agent		
Alginate	Gelling agent	30	90
Propylene glycol alginate	Emulsifier and foam stabilizer		
Arabic	Emulsifier	30	150
Carrageenan	Gelling agent	35	270
Processed euchema seaweed	Gelling agent		
Carboxymethyl cellulose	Thickener		34
Hydroxypropylmethyl cellulose	Thickener and emulsifier		
Methyl cellulose	Thickener, emulsifier, and gelling agent		
Cellulose (microcrystalline)	Thickener and gelling agent		
Gelatin	Gelling agent	162	350
Guar gum	Thickener	30	80
Karaya	Thickener		
Konjac mannan	Thickener and gelling agent		
Locust bean gum	Thickener	10	100
Pectin	Gelling agent	25	250
Starch (modified)	Thickener and gelling agent	1000	650
Tragacanth	Thickener		
Xanthan gum	Thickener	16	160

Source: Lillford PJ (2000) *Gums and Stabilisers for the Food Industry*. In: Williams PA and Phillips GO (eds) Royal Society of Chemistry Publication 251. Cambridge, UK.

Table 3 Thickeners

Xanthan gum
Very high viscosity at low shear rates (yield stress), highly shear thinning, maintains viscosity in the presence of electrolyte, over a broad pH range and at high temperatures

Carboxymethyl cellulose
High viscosity but reduced by the addition of electrolyte and at low pH

Methyl cellulose and hydroxypropylmethyl cellulose
Viscosity increases with temperature (gelation may occur) not influenced by the addition of electrolytes or pH

Galactomannans (guar and locust bean gum)
Very high viscosity at low shear rates and strongly shear thinning. Not influenced by the presence of electrolyte but can degrade and lose viscosity at high and low pH and when subjected to high temperatures

generally considered to enhance the stability of dispersions (and emulsions) by increasing the viscosity of the aqueous phase, it can have the opposite effect and give rise to depletion flocculation. This phenomenon occurs as a consequence of polymer molecules being excluded from the space between approaching particles. The difference in osmotic pressure between the depleted region and the bulk solution results in weak interparticle attractive forces, which induce aggregation.

Some gums, such as gum arabic, methyl cellulose and propylene glycol alginate, show amphiphilic character which has led to their use in the stabilization of emulsions and foams owing to their affinity to adsorb at the oil–water or air–water interface.

In systems containing sugar or ice crystals, gums can retard crystal growth either by adsorbing on to the crystals or by competing for available water molecules.

Although gums have primarily been used in foods to control texture, organoleptic properties, and stability, consumers are now being made increasingly aware of their nutritional and health benefits. Many gums (e.g., locust bean gum, guar gum, konjac mannan, gum arabic, xanthan gum, pectin) have been shown to reduce blood cholesterol levels. Others (e.g., inulin, gum arabic) have been shown to have prebiotic effects. They are resistant to our digestive enzymes and pass through the stomach and small intestine without being metabolized. They are fermented in the large intestine to yield short-chain fatty acids and stimulate the specific growth of beneficial intestinal bacteria, notably bifidobacteria, and reduce the growth of harmful microorganisms such as clostridia.

Properties of Individual Gums

Tree Gum Exudates

The four main tree gum exudates are gum arabic, gum karaya, gum tragacanth, and gum ghatti, with gum arabic being by far the most important commercially. Gum arabic occurs as a sticky liquid that oozes from the stems and branches of acacia trees (*Acacia senegal* and *A. seyal*) which grow across the Sahelian belt of Africa, principally Sudan. The gummosis process occurs when the tree is subjected to stress conditions such as heat, drought, or wounding. The liquid dries in the sun to form glassy nodules, which

Table 4 Gelling agents

Thermoreversible gelling agents	
Gelatin	Gel formed on cooling. Molecules undergo a coil-to-helix transition followed by aggregation of helices
Agar	Gel formed on cooling. Molecules undergo a coil-to-helix transition followed by aggregation of helices
Kappa-carrageenan	Gel formed on cooling in the presence of salts, notably potassium salts
Iota-carrageenan	Gel formed on cooling in the presence of salts
Low-methoxyl (LM) pectin	Gels formed in the presence of divalent cations, notably calcium at low pH (3–4.5)
Gellan gum	Gels formed on cooling in the presence of salts
Methyl- and hydroxypropylmethyl cellulose	Gels formed on heating
Xanthan gum with locust bean gum or konjac mannan	Gels formed on cooling mixtures
Thermally irreversible gelling agents	
Alginate	Gels formed on the addition of polyvalent cations, notably calcium or at low pH (< 4)
High-methoxyl (HM) pectin	Gels formed at high soluble solid (e.g., 50% sugar) content at low pH < 3.5
Konjac mannan	Gels formed on addition of alkali
Locust bean gum	Gels formed after freezing due to association of galactose-deficient regions

are collected by hand. The gum from *A. senegal* (traditionally the main source of gum arabic) is a complex polysaccharide consisting of galactopyranose (~44%), arabinopyranose, arabinofuranose (~25%), rhamnopyranose (~14%), glucuropyranosyl uronic acid (~15.5%) and 4-*O* methyl glucuropyranosyl uronic acid (~1.5%). It also contains a small amount (~2%) of protein as an integral part of the structure. Analysis of the carbohydrate structure has shown that it consists of a core of (1,3)-β-D-galactose units with extensive branching at the C6 position. The branches consist of galactose and arabinose and terminate with rhamnose and glucuronic acid (Figure 1a). As a consequence of the highly branched structure, viscous solutions are only formed at high gum concentrations (~50% w/w). It has been shown by gel permeation chromatography that the gum consists of three molecular mass fractions which have also been shown to differ principally in their protein contents. Most of the gum (~90%) contains very little protein and has a molecular mass of $\sim 2.5 \times 10^5$. A second fraction (~10% of the total) contains ~10% protein and has a molecular mass of $1–2 \times 10^6$ and has been shown to have a 'wattle-blossom'-type structure where blocks of carbohydrate of molecular mass $\sim 2.5 \times 10^5$ are connected to a common polypeptide chain. The third fraction (~1% of the total) contains up to 50% protein and has a molecular mass of $\sim 2 \times 10^5$. Fraction 2 has been shown to be responsible for the gum's excellent ability to stabilize oil-in-water emulsions. It has been proposed that the hydrophobic polypeptide chains adsorb on to the oil droplets and anchor the molecules to the surface, while the hydrophilic carbohydrate blocks protrude out into solution and prevent droplet aggregation and coalescence due to electrosteric repulsions. Its ability to act as an emulsifier has led to its use in stabilizing flavor oil emulsion concentrates for the soft drinks industry and also in the production of spray-dried encapsulated flavors for use in dry packaged products such as soup and cake mixes. In the latter case the gum forms a film around the flavor particle, preventing oxidation and evaporation, and the gum's high solubility facilitates rapid flavor release. The ability of gum arabic to form concentrated solutions of low viscosity has led to its widespread use in confectionery products, particularly those with a high sugar content such as pastilles where it functions by reducing sugar crystallization.

Gum tragacanth consists of a water-swellable fraction called tragacanthic acid (or bassorin) (60–70%) and a water-soluble fraction called tragacanthin. The former consists of a main chain of (1,4)-α-D-galactopyranosyluronic acid residues with branches linked through (1,3)-β-D-xylose units terminating in L-fucopyranose, as illustrated in Figure 1b. The latter is a highly branched arabinogalactan with a main chain of D-galactopyranosyl units either (1,6)- or (1,3)-linked with side chains consisting mainly of L-arabinofuranose but with a small proportion of D-galacturonic acid and L-rhamnose. Gum tragacanth gives rise to high-viscosity solutions even at 1% concentration. The viscosity decreases irreversibly on heating. The gum solution is stable under acid conditions and shows good emulsification characteristics, which has led to its use in salad dressings and sauces. Its cost and availability, however, have meant that it has been largely replaced in these products by other gums, notably xanthan gum.

Gum karaya is a heavily acetylated polysaccharide composed of chains of α-D-galacturonic acid and α-L-rhamnose. The acid groups are glycosylated with

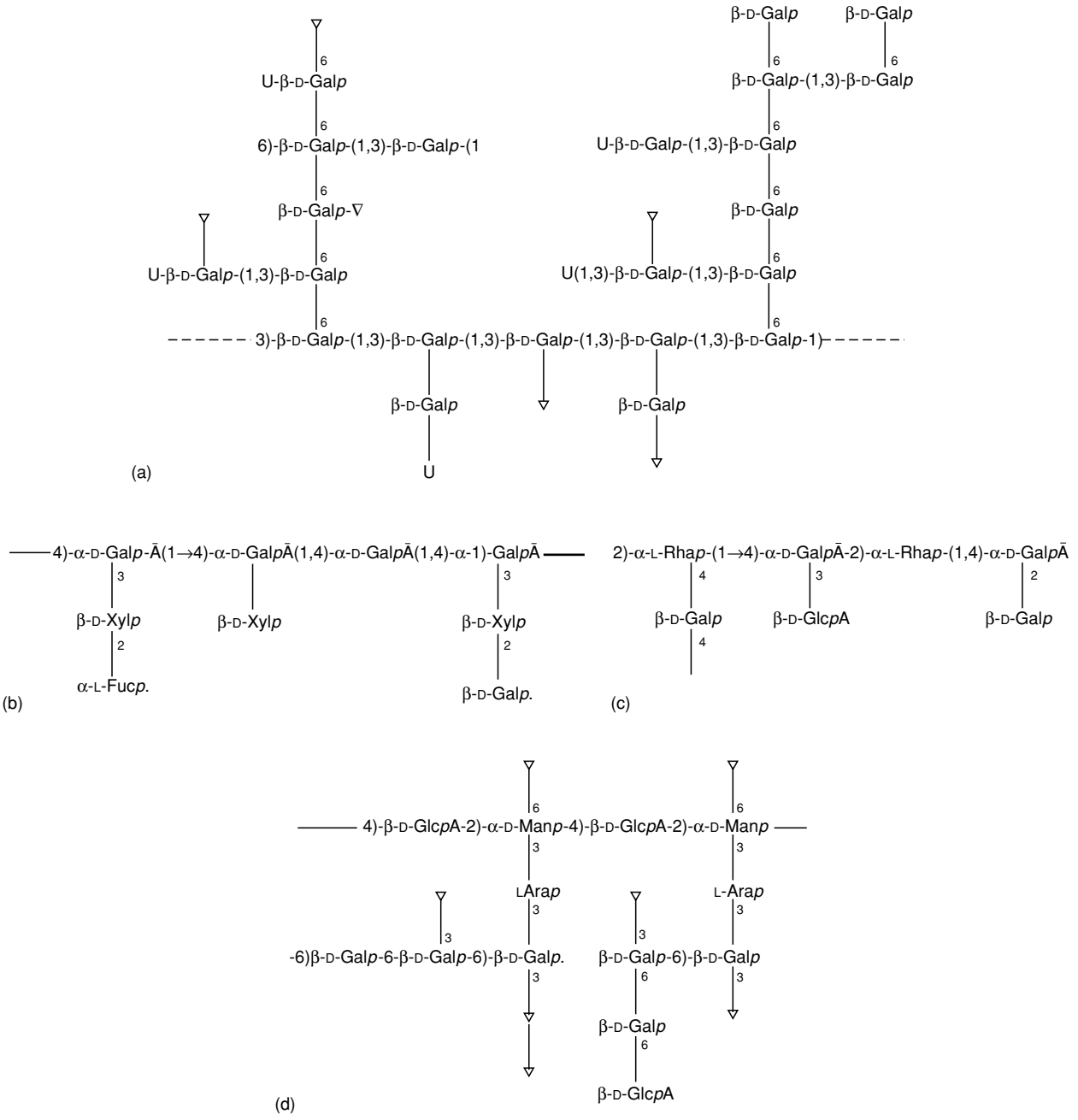

Figure 1 (a) *Acacia senegal* representative segment. U = −2)-α-L-Rha*p*-(1,4)-β-D-GlcA- or −4)-β-D-methyl-Glc*p*A. ∇ = L-Ara*f* or L-Ara*p* terminated (1,3)-linked short chains or α-D-Gal*p*-(1,3)-L-Ara*f*. (b) Tragacanthic acid. (c) Some structural features of gum karaya. (d) Some structural features of gum ghatti: ∇ = L-Ara*f* or L-Ara*p* terminated side chains. Reproduced from Gums: Properties of Individual Gums. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

β-D-galactose or β-D-glucuronic acid residues, whereas about half of the rhamnose groups carry β-D-galactose units as side chains (Figure 1c). Gum karaya swells in water to yield highly viscous solutions, however, solutions show a permanent loss in viscosity on heating. The gum finds some application in salad dressings.

Gum ghatti has a main chain of alternating (1,4)-β-D-glucopyranosyluronic acid and (1,2)-α-D-mannopyranose units and contains numerous side chains of L-arabinose, D-galactose, and D-glucuronic acid (Figure 1d). It contains a water-soluble (> 80%) and water-swellable fraction and was originally developed in the early 20th century as replacement for

gum arabic since it forms solutions of low viscosity at high concentrations and is also a good emulsifier.

Cellulosics

Cellulose is the most abundant of all the polysaccharides, being the structural component of land plants. It is composed of linear chains of (1,4)-β-D-glucopyranose units which are associated through hydrogen bonding, giving rise to a number of crystalline forms. Although it is insoluble in water, it can be chemically modified to form a number of water-soluble derivatives with valuable functional properties. Derivatization usually involves etherification of the reactive hydroxyl groups on the glucose residues. The most common water-soluble cellulose derivatives used in food applications are carboxymethyl (CMC), methyl (MC), and hydroxypropylmethyl (HPMC) cellulose. Derivatization is carried out following conversion of the cellulose to the sodium form by treatment with concentrated alkali. This process destroys the crystalline structure and yields alkali cellulose. The alkali cellulose is then reacted with the appropriate reagent, namely sodium monochloroacetate for CMC, methyl chloride for MC, and methyl chloride/propylene oxide for HPMC. Since the reactions are heterogeneous, substitution can be very irregular.

CMC is the most widely used of the cellulose ethers and can be obtained in various grades, which vary in molecular mass and degree of substitution (DS). The minimum DS is ~ 0.4 (i.e., 4 hydroxyls substituted per 10 glucose residues), since below this the CMC is insoluble. Above this CMC dissolves readily in water to form viscous solutions but, since it is anionic, the viscosity of dilute solutions ($< 1\%$) is reduced on addition of electrolyte and at low pH due to compaction of the polymer chains. A synergistic increase in viscosity occurs in blends with guar gum and locust bean gum. CMC finds widespread use in icecream and other frozen desserts as well as salad dressings, sauces, and gravies.

The nonionic cellulose ethers, MC and HPMC, are also supplied in a range of molecular sizes and DS. For HPMC, since the substituted moieties contain hydroxyl groups they may also participate in the reaction. Hence a molar substitution (MS) is also quoted to characterize the polymers. MC and HPMC dissolve readily in water to produce viscous solutions which on heating form thermoreversible gels. The gelation temperature is dependent on the degree of substitution. For MC containing 30% methoxyl groups, gelation occurs at ~ 50–$55\,^{\circ}C$, whereas for HPMC containing 20% methoxyl and 8% hydroxypropyl, gelation occurs at $\sim 85\,^{\circ}C$. Gelation is believed to be a consequence of molecular association through the hydrophobic methyl and hydroxypropylmethyl substituents. This functional characteristic has led to their use in combination with starch in soups and gravies to prevent undesirable loss of viscosity on heating. A further application is their use in fried foods, where gelation on heating helps retain the structural integrity of the product and additionally serves to prevent moisture loss and oil absorption. Since MC and HPMC are also surface-active they are used in baked goods to aid uniform gas cell formation and in salad dressings to stabilize emulsion droplets.

Plant Extracts

Starch is also a very abundant material and is derived commercially principally from corn and potato and, to a lesser extent, from waxy corn, wheat, tapioca, rice, cassava, sorghum, pea, and sago. It is obtained in the form of granules, which have varying degrees of structural order and consist of two polysaccharides, namely amylose and amylopectin. The proportions of each depend on the source, for example, corn and potato starch contain $\sim 27\%$ and 21% amylose respectively, while waxy corn contains $< 2\%$ amylose.

Amylose consists of linear (1,4)-α-D-glucopyranose chains with very little branching (10 branch points per molecule) and has a molecular mass of typically 2×10^5–2×10^6. Amylopectin also contains sequences of (1,4)-α-glucopyranose units; however, it has extensive branching via (1,6) linkages and has a molecular mass of $\sim 5 \times 10^7$–4×10^8. The starch granules are insoluble in water but on heating they swell and burst, releasing amylose, forming a viscous paste. The temperature at which the granules burst is very characteristic of the source of the starch and is (perhaps inappropriately) referred to as the gelatinization temperature. Typical values for corn and potato starch are $67\,^{\circ}C$ and $60\,^{\circ}C$ respectively. On cooling, the amylose molecules readily self-associate, a process known as retrogradation, and it is then that a gel is formed. Natural starches form turbid gels which are prone to syneresis. Hence most of the commercial starches used are derivatives which have a lesser tendency for retrogradation. Derivatives include hydroxypropyl starches, starch phosphates, oxidized starch, acid, or enzyme-degraded starches.

Pectin is the general term for a group of polyuronans that occur as the structural components of plants. Commercial pectins are obtained from the pomace of apple or the peel of citrus fruits following hydrolysis, which renders the pectin water-soluble. Pectin molecules consist of linear chains of (1,4)-α-D-galacturonic acid residues, up to 80% of which occur as the methyl ester. The hydroxyls in the C2 and C3 positions may be acetylated. The chains also

contain up to 4% (1,2)-α-rhamnopyranose units which give rise to kinks. L-Arabinose, D-galactose, and D-xylose (10–15%) are linked to the rhamnose units forming ramified side chains which are referred to as 'hairy regions' along the otherwise smooth galacturonan backbone. If the degree of esterification (DE) is > 50% it is referred to as high-methoxyl (HM) pectin. Deesterified pectin with DE < 50% is produced by mild acid or alkali treatment and is referred to as low-methoxyl (LM) pectin. If the alkali used is ammonia, amidated pectin is formed. Pectin is soluble in water and is most stable at pH 3–4. At lower pH, the glycosidic bonds are hydrolyzed. In alkaline conditions the glycosidic and ester linkages are cleaved, yielding the free acid. Both HM and LM pectins form gels. For HM pectin (DE 60–75%) gelation occurs at high soluble solids content (typically 50–75% sugar) and at pH < 3.5 over a period of time. The gels are not thermoreversible. Junction zone formation is believed to be as a consequence of hydrophobic association between ester groups coupled with intermolecular hydrogen bonding between hydroxyl groups on the galacturonan backbone. For LM pectin (DE typically 20–40%) gelation is brought about by the addition of divalent cations, and a high solids content and low pH are not a prerequisite. Gelation is rapid and is thermoreversible.

Pectins are mainly used for the production of jams and jellies. High-ester pectins find widespread use in normal jams, whereas low-ester pectins are preferred for low-sugar jams. LM amidated pectin is used to stabilize fruit pulp which is incorporated into products such as yogurt.

Konjac mannan is the soluble extract of konjac flour obtained by pulverizing the dried tuber of *Amorphophallus konjac*. It is a glucomannan and consists of a main chain of (1,4)-β-D-mannopyranosyl and D-glucopyranosyl units with (1,3) linked branches approximately every 10 sugar residues. The ratio of mannose to glucose is 1.6:1. Approximately 1 in 19 sugar residues is acetylated. Konjac mannan forms highly viscous solutions on dissolution in water following heating. In the presence of alkali, thermally irreversible gels are formed as a consequence of deacetylation and aggregation of the glucomannan chains. In Japan, konjac mannan is made into noodles. The process involves pumping a solution of the gum through a spinneret into a hot alkaline bath (e.g., calcium hydroxide solution) which results in the formation of long gel threads. The gum is also sold in blocks of gel, known as konnyaku, for use in traditional Japanese dishes and also as a dessert jelly. Konjac mannan has recently been approved for food use and is now finding application as a thickener and gelling agent. It forms thermoreversible gels with xanthan gum and also increases the gel strength, elasticity, and clarity of kappa-carrageenan gels. For both mixtures it has been argued that the synergistic behavior is due to association of carrageenan and xanthan helices with the glucomannan chains.

Galactomannan Seed Gums

Locust bean (or carob), tara, and guar gums are storage polysaccharides obtained from the endosperms of leguminous seeds of *Ceratonia siliqua*, *Caesalpinia spinosa*, and *Cyamopsis tetragonoloba* respectively. They have a molecular mass of the order of 10^6 and consist of a linear main chain of β-(1,4)-linked D-mannopyranosyl units with galactopyranosyl units linked α-(1,6) to varying degrees. The mannose-to-galactose ratio is approximately 4.5:1, 3:1, and 2:1 for locust bean, tara, and guar gums respectively. The galactose residues are distributed nonuniformly along the mannan chain. The presence of galactose tends to inhibit intermolecular association between the galctomannan chains, hence, whereas guar gum is readily soluble in cold water, tara, and locust bean gums have to be heated to high temperatures to disrupt the aggregation and achieve complete dissolution. Once dissolved, all three yield highly viscous solutions even at 1% concentration and as a consequence their main application is as thickeners.

Locust bean gum will self-associate in solution and will form thermally irreversible gels on freezing. It is commonly used in combination with other polysaccharides, particularly kappa-carrageenan, since it leads to the formation of stronger, more elastic gels, which have improved transparency and are less prone to undergo syneresis. Locust bean gum also forms strong thermoreversible gels with xanthan gum. It is believed that interaction occurs between the carrageenan and xanthan helices and mannose sequences along the backbone which are devoid of galactose residues. This model also explains why guar gum is not as effective.

Algal Polysaccharides

The carrageenans are a family of sulfated galactans obtained from red seaweeds (Rhodophyceae) where they have a key structural function. The traditional method of extraction is by treatment of the seaweed with hot alkali for 10–30 h followed by precipitation with alcohol and then drying. The three major types are kappa-, iota- and lambda-carrageenan. Kappa is obtained from a species of seaweed called *Euchema cottonii* and occurs together with lambda-carrageenan in *Chondrus crispus*. Iota-carrageenan is obtained from *E. spinosum*. They differ essentially

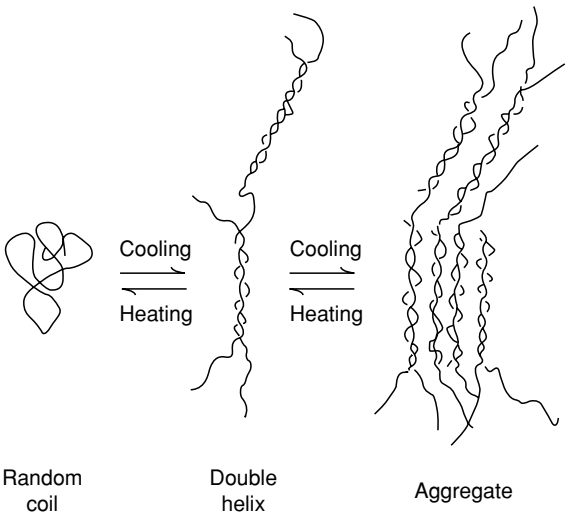

Figure 2 Gelation process for agarose and carrageenan involving double helix formation and aggregation. Reproduced from Gums: Properties of Individual Gums. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

in their degree of sulfation. The idealized structure for kappa-carrageenan is a disaccharide repeat unit consisting of (1,3)-β-galactopyranose-4-sulfate and (1,4)-α-3,6-anhydrogalactopyranose residues. Iota-carrageenan differs only in that the latter residue is sulfated at the C2 position. Lambda-carrageenan is further sulfated and consists of (1,4)-α-galactopyranose 2,6 disulfate and (1,3)-β-linked galactopyranose which is 70% substituted at the C2 position. The carrageenans are all soluble in water but whereas lambda forms viscous solutions, kappa and iota form thermoreversible gels. In solution the molecules undergo a thermoreversible coil-to-helix transition and the helices self-associate, giving rise to a three-dimensional gel structure (Figure 2). The temperature of gelation increases with increasing electrolyte concentration. It has been shown that potassium, rubidium, and cesium ions specifically bind to the helical structure of kappa-carrageenan and hence promote helix formation and gelation at much lower concentrations than other electrolytes. As a consequence, kappa-carrageenan gels are much stronger in the presence of potassium chloride compared to, say, sodium chloride. This ion specificity is not observed for iota-carrageenan, which forms weaker more elastic gels compared to kappa. This is probably due to the fact that the increased charge on the iota-carrageenan chains reduces the extent of helix self-association. Self-association results in the melting temperature being greater than the gelation temperature and this hysteresis is more pronounced in kappa-carrageenan.

Kappa-carrageenan gels are more brittle and turbid than iota-carrageenan gels and are more prone to syneresis. As discussed above, incorporation of locust bean gum or konjac mannan into kappa-carrageenan gels reduces turbidity, brittleness, and syneresis and also increases gel strength.

Kappa-carrageenan is widely used in dairy products because it interacts synergistically with kappa-casein and prevents wheying-off. Carrageenans are also used extensively in meat products such as cooked hams, poultry products, and sausages because of their ability to bind water and control the texture and structural integrity.

Agarose (the major component of agar) is also obtained from red seaweeds, notably *Gelidium* and *Gracilaria* spp. It is a linear neutral polysaccharide and has a similar structure to the carrageenans, consisting of alternating (1,3)-β-D-galactopyranose and (1,4)-linked 3,6-anhydro-α-L-galactopyranose units. It dissolves in near-boiling water and gels on cooling to $\sim < 40\,^{\circ}C$. The gelation mechanism is as described for the carrageenans but, since the agarose chains are nonionic, extensive helix aggregation occurs, resulting in the formation of very strong gels. The gels show pronounced hysteresis and melt only on heating to 80–90 °C. The use of agar is far more restricted commercially than that of carrageenan owing to availability and cost considerations.

Alginate is obtained from brown seaweeds (Phyophyceae) and is a linear (1,4)-linked polyuronan consisting of β-D-mannuronic and α-L-guluronic acids. The acids are present as blocks of separate or mixed sequences along the chain depending on the seaweed source. *Macrocystis pyriferia* and *Ascophyllum nodosum* have a high mannuronic acid content (61% and 65%, respectively), whereas *Laminaria hyperborea* has a high guluronic acid content (69%). Sodium alginate dissolves readily in water to form viscous solutions and forms thermally irreversible gels in the presence of divalent cations (notably calcium). It is the guluronic acid residues that are responsible for gel formation and their proportion and distribution along the polyuronan chain have a major influence on the properties of the gels produced. Adjacent diaxially linked guluronic acid residues form a cavity which acts as a binding site for cations which interact with the carboxyl and hydroxyl groups. Intermolecular cross-linking of sequences results in the formation of junction zones and a three-dimensional gel network. This mechanism is referred to as the 'egg-box model' (Figure 3).

If a sodium alginate solution is added dropwise into a solution containing a calcium salt, gel beads are produced and this process is used for the production of 'artificial fruit' for use in pie fillings. In order

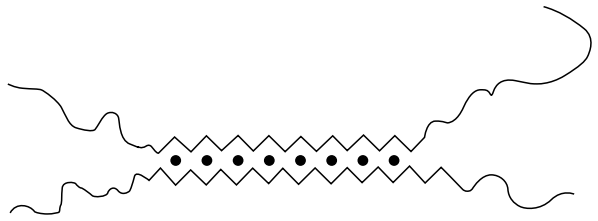

Figure 3 Schematic representation of junction zone formation in calcium alginate gels (egg-box model). The calcium ions are represented by •. Reproduced from Gums: Properties of Individual Gums. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

to produce homogeneous gels, a common practice is to generate the calcium ions slowly *in situ*. Sparingly soluble salts such as calcium phosphate are used and the release of ions is controlled by sequesterants and adjustment of pH. This process is used in the preparation of structured or fabricated foods such as fruit and meat products.

Microbial Gums

Xanthan gum was first discovered in the 1950s and is now finding extensive application in the food industry since its introduction in the early 1970s. The gum is obtained from the genus *Xanthomonas*, notably *X. campestris* by aerobic fermentation. The xanthan molecules have a (1,4)-β-D-glucopyranose backbone as in cellulose, but in addition have a trisaccharide side chain on every other glucose residue linked through the C3 position. The side chain consists of two mannopyranosyl residues linked on either side to a glucuropyranosyl uronic acid group. The inner mannose residue connected to the backbone may be acetylated while the terminal mannose residue may be pyruvated. The molecular mass of the xanthan molecules is very high ($> 3 \times 10^6$) and the gum dissolves in water to yield highly viscous solutions. The xanthan molecules are reported to undergo a thermoreversible coil-to-helix transition in solution, which is shifted to higher temperatures by the addition of electrolyte. In the disordered coil form the side chains are envisaged as protruding away from the backbone into solution, while in the ordered form the molecules form a stiff five-fold helical structure with the side chains folded in and associated with the backbone. Most workers now believe that the helix is double-stranded. The stiff xanthan chains tend to associate in solution, giving rise to a very high viscosity at low shear rates, which is sufficient to prevent particles from sedimenting or oil drops from creaming. The associations are readily broken on applying shear, giving rise to a dramatic drop in viscosity and hence the solutions readily flow. Unlike other polyelectrolytes the viscosity of xanthan solutions is relatively insensitive to addition of electrolyte since the electrolyte will promote helix formation. The helices are also stable over a broad range of pH. The high apparent yield stress and pronounced shear thinning characteristics have led to a large increase in the use of this gum, which now finds application in a broad range of food products such as sauces, salad dressings, and mayonnaises.

Gellan gum has only recently been commercialized. It is produced from *Pseudomonas elodia* by aerobic fermentation and consists of a linear tetrasaccharide repeat unit of (1,3)-β-D-glucopyranose, (1,4)-β-D-glucopyranose uronic acid, (1,4)-β-D-glucopyranose and (1,4)-α-L-rhamnopyranose. In the native form the (1,3)-linked glucose residues contain glycerate and acetate moieties. X-ray fiber diffraction studies indicate that the gellan molecules form a threefold double helical structure and in solution undergo a thermoreversible coil-to-helix transition. The transition shifts to higher temperatures in the presence of electrolyte. Once formed, the helices tend to self-associate, leading to the formation of a gel, as is the situation for agarose and carrageenan. Gels formed in the presence of monovalent ions are usually thermoreversible, although the melting temperature is normally much greater than the gelation temperature – a consequence of the extensive aggregation. If gels are formed by the addition of divalent ions then they can be thermally irreversible. The native form produces soft elastic gels whereas the deacetylated material, which is sold commercially, forms hard brittle gels. The gum is used in fruit fillings.

See also: **Cellulose**; **Gelatin**; **Gums**: Food Uses; Dietary Importance; Nutritional Role of Guar Gum; **Marine Foods**: Production and Uses of Marine Algae; **Pectin**: Properties and Determination; Food Use; **Starch**: Structure, Properties, and Determination

Further Reading

Hill SE, Ledward DA and Mitchell JR (1998) *Functional Properties of Food Macromolecules*. Gaithersburg, MD: Aspen.

Imeson I (1992) *Thickeners and Gelling Agents for Food*. Glasgow: Blackie Academic and Professional.

Lapasin R and Pricl S (1995) *Rheology of Industrial Polysaccharides; Theory and Applications*. Glasgow: Blackie Academic and Professional.

Nussinovitch A (1997) *Hydrocolloid Applications; Gum Technology in the Food and other Industries*. Glasgow: Blackie Academic and Professional.

Phillips GO and Williams PA (2000) *Handbook of Hydrocolloids*. Cambridge: Woodhead.

Ross Murphy S (1994) *Physical Techniques for the Study of Food Biopolymers*. Glasgow: Blackie Academic and Professional.
Stephen AM (1995) *Food Polysaccharides and Their Application*. New York: Marcel Dekker.
Williams PA and Phillips GO (2000) *Gums and Stabilisers for the Food Industry 10*. Cambridge: Royal Society of Chemistry.

Food Uses

P A Williams and G O Phillips, The North East Wales Institute, Wrexham, UK

Introduction

Gums have a major influence on the structural characteristics, texture, organoleptic properties, and overall appearance of foods, even though they are present often at concentrations of less than 1%. On food labels they are commonly referred to as 'stabilizers,' 'thickeners,' or 'gelling agents' and in fact may serve a number of functions, such as enhancing viscosity, preventing particle sedimentation or droplet creaming, inducing gelation, emulsifying oils, stabilizing foams, and inhibiting ice or sugar crystallization. The various food sectors where gums are commonly used are shown in **Table 1** together with an estimate of their usage levels. The high levels of modified starch, gelatin, and guar gum used reflect their relatively low cost.

The use of gums in food products is steadily increasing due to the rapid rise in the consumption of ready-made meals and novelty foods and also because of consumers' growing awareness of the need to increase the amount of fiber and reduce the amount of fat in the diet. Gums are considered as soluble fiber and there is evidence to indicate that they assist plasma cholesterol reduction and promote fermentation in the large bowel. The latter yields short-chain fatty acids, mainly acetate, proprionate, and butyrate, which have a beneficial effect on the colon through stimulation of blood flow and enhancement of electrolyte and fluid absorption and muscular activity. Gums are nowadays widely used for fat replacement in a wide range of low-calorie products. They may be added in their natural form or as specifically fabricated microparticulates and serve to enhance the rheological properties of the lower fat products. Rheological enhancement can also be achieved by incorporating two or more gums. In such systems, phase separation can occur and, if one or more of the gums is able to form a gel, can lead to the development of novel gel structures.

Safety of Gums used in Foods

The Joint Expert Committee on Food Additives (JECFA) was first established in the mid-1950s by the Food and Agricultural Organization (FAO) and World Health Organization (WHO) to assess chemical additives in food on an international basis. The Codex Alimentarius Commission (CAC), an international intergovernmental body, was set up in the early 1960s with the primary aim of protecting the health of the consumer and to facilitate international trade in food commodities. When CAC was formed, it was decided that JECFA would provide expert advice to Codex on matters relating to food additives.

While CAC is the ultimate specification and can provide for approval throughout the world, each country is free to adopt its own standards.

The international numbering system (INS) for food additives is intended as an identification for food additives for use in one or more countries. The

Table 1 Annual usage (tons) of gums in the various food industry sectors

Food sector	*Starches*	*Carrageenan*	*Alginates*	*Gum arabic*	*Guar gum*	*Xanthan*	*Gelatin*	*Pectin*	*CMC*
Confectionery	8000	200	200	4800			13 000	900	
Dairy and desserts	6000	1800	700		3900	250	4500		190
Pet food	3500	150	250		2000	150			
Readymade meals	1500				1500	200			160
Meat products		350					5500		120
Bakery products	2000		400		2100				130
Sauces and dressings					700	200			
Other	2000	1000	300	700	300	180	1500	2900[a]	120
Total usage	23 000	3500	1850	5500	10 500	980	24 500	3800	720

[a]Mainly jams.
Source: Giract based on utilization in the UK, France, and West Germany in 1988.
CMC, carboxymethyl cellulose.

Table 2 International numbering system (INS) and E numbers for food gums

Hydrocolloid	*INS/E number*[a]	*Function*
Carrageenan (including furcelleran)	407	Thickener, gelling agent, stabilizer, emulsifier
Processed euchema seaweed	407a	Thickener, gelling agent, stabilizer, emulsifier
Xanthan gum	415	Thickener, stabilizer, emulsifier, foaming agent
Gellan gum	418	Thickener, gelling agent, and stabilizer
Guar gum	412	Thickener, stabilizer, and emulsifier
Locust bean gum	410	Thickener, gelling agent
Gum arabic	414	Emulsifier, stabilizer, thickener
Pectin	440/E440 (i)	Gelling agent, thickener, stabilizer
Amidated pectin	E440 (ii)	Gelling agent, thickener, stabilizer
Microcrystalline cellulose	INS460 (i)	Anticake, emulsifier, stabilizer, and dispersing agent
Powdered cellulose	INS460 (ii) / E460	Anticake, emulsifier, stabilizer, and dispersing agent
Cellulose	460	Anticake, emulsifier, stabilizer, and dispersing agent
Tragacanth gum	413	Emulsifier, stabilizer, thickening agent
Karaya gum	416	Emulsifier, stabilizer, and thickening agent
Konjac mannan	E425	Gelling agent, thickener, emulsifier, stabilizer
Propylene glycol alginate	405	Thickener, emulsifier
Sodium alginate	401	Thickening agent, stabilizer
Potassium alginate	402	Thickening agent, stabilizer
Alginic acid	400	Thickening agent, stabilizer
Calcium alginate	404	Thickening agent, stabilizer
Ammonium alginate	403	Thickening agent, stabilizer
Methyl cellulose	461	Thickening agent, emulsifier, stabilizer
Hydroxypropyl cellulose	463	Emulsifier, thickener, stabilizer, binder, suspension agent, film coating
Hydroxypropylmethyl cellulose	464	Emulsifier, thickening agent, stabilizer
Sodium carboxymethyl cellulose	466	Thickening agent, stabilizer, suspending agent
Agar	406	Thickening agent and stabilizer
Oat gum	411	Thickener, stabilizer

[a]Where no distinction is made, the INS and E numbers are identical. When only an E number is given, there is no equivalent INS number. From Phillips GO, Williams PA (2000) *Handbook of Hydrocolloids*. Cambridge: Woodhead.

criteria for INS inclusion are: the compound must be approved by a country as a food additive, it must be toxicologically cleared for use by a country, and it must be identified on the final product label.

In the European Union, gums accepted for food use are classified by so-called E numbers. Where both INS and E numbers are available they are interchangeable. INS and E numbers for gums used in foods are given in Table 2, together with their functional characteristics.

Food Applications

Condiments, Mayonnaise, and Dressings

Gum tragacanth was traditionally used in thick sauces, pickles, and relishes because of its ability to maintain the desired rheological properties in the acidic conditions normally associated with these products. However, because of its uncertainty of supply and increasing high cost, modified starches are nowadays the thickeners of choice. Novel products are continually being introduced into the marketplace and some examples of sauces in which gums are incorporated are shown in Figure 1. The choice of gum is dependent on the nature of the product. The tartare sauce, smokey barbecue sauce, and horseradish relish, for example, contain xanthan gum and this acts to thicken the sauces and is stable in the acid conditions found in these products. The creamed garlic sauce contains both xanthan gum and guar gum and these are known to interact synergistically to give enhanced rheological properties. The seafood sauce has a higher oil content than the other sauces and, therefore, propylene glycol alginate is included with xanthan gum. The propylene glycol alginate functions as an emulsifier, preventing droplet coalescence, while the xanthan gum enhances the viscosity of the aqueous phase, thus preventing droplet creaming. The apple sauce, cranberry sauce, and mint jelly all contain pectin, which gives the products gel-like characteristics.

Traditional mayonnaise is a concentrated, oil-in-water emulsion containing about 65% vegetable oil, acidifying ingredients (vinegar or lemon juice) and egg yolk, which acts as an emulsifying agent. The semisolid characteristics associated with this food product are principally controlled by the concentration of oil and the size of the emulsion droplets. Spoonable products can be made at much lower oil content (30–50%) by the incorporation of gums at concentrations as low as 0.5%. The gums used are xanthan gum (sometimes in

Figure 1 Typical sauces incorporating gums.

combination with guar gum or locust bean gum), which functions as a thickener, and propylene glycol alginate, which aids emulsification.

Xanthan gum is widely used nowadays in place of gum tragacanth in pourable dressings because of its stability at low pH, its ability to slow down the separation of the oil-and-water phases, and suspend particulate ingredients, and because of its pronounced shear thinning characteristics.

Coleslaws, potato pasta and Florida-style salads are now commonly available ready-prepared and xanthan gum, perhaps in conjunction with guar gum or locust bean gum, is often incorporated as a thickener.

Icecream, Yogurt, and Other Dairy Products

Icecream Icecream is a complex food system consisting of a liquid phase incorporating fat particles, air bubbles, ice crystals, milk proteins, sugars, and soluble and insoluble salts. Gums are commonly used in combination with emulsifiers in icecream formulations to control the stability of the product. The level of gum used depends on the type and overall composition but is typically in the range of 0.1–0.35%. Even at these very low levels, the gums can have a profound effect, serving to enhance the viscosity of the icecream mix, thus increasing the whipping properties and allowing incorporation of a greater proportion of air prior to freezing. The amount of air incorporated is termed the overrun and it can be 100% or more by volume. The use of gums also has a considerable influence on the organoleptic properties of the product, affecting the body, texture, and mouth feel. Icecreams made without gums also tend to melt much faster. The major influence on texture comes from the ability of gums to inhibit ice crystal growth, which occurs as a result of the phenomenon known as Ostwald's ripening. This process arises from the fact that 'small' ice crystals have a greater solubility than 'large' ice crystals. Hence, the continual melting and freezing of the crystals as a consequence of the temperature fluctuations experienced on prolonged storage in a freezer result in the formation of increasingly large ice crystals. It has been argued that the gums act by binding water molecules, thus preventing them from freezing. However, since they are used in such small quantities, it is more likely that they act by adsorbing on to the ice crystals, thus inhibiting crystal growth. Gums can also reduce 'sandiness' which is found in some icecreams as a result of lactose crystallization.

For batch processing the gum can be added to the cold or hot mix, but for high-temperature, short-time (HTST) pasteurization the gum is dispersed in the mix when cold and should be cold-water-soluble. The gums are usually blended with a proportion of the sugar before incorporation into the mix in order to aid dispersibility.

Today the most widely used gums are alginates, carboxymethyl cellulose, guar gum, and locust bean gum. Serum separation can sometimes occur in the mix before it is frozen and this is avoided by incorporation of kappa-carrageenan as a secondary stabilizer (at, say, 10% of the total stabilizer content). It is believed that the carrageenan interacts with the milk constituents present, notably the kappa-casein, to form a weak gel structure. Microcrystalline cellulose has been shown to be particularly effective at reducing ice crystal formation and enhancing the body, and has application in reduced-fat icecreams.

Yogurt Yogurt is produced from milk following lactose fermentation induced by the action of

microflora. These include *Lactobacillus acidophilus*, *L. bulgaricus*, and *Streptococcus thermophilus*. *Bifidobacterium longum* is also used in the 'bioyogurts' which have now become popular. A culture containing the bacteria is added to the milk after it has been homogenized and undergone heat treatment. The formation of lactic acid causes the milk proteins to coagulate, forming a semisolid gel which entraps the fat globules and the serum containing the dissolved components. For 'set' yogurts, coagulation is carried out directly in retail cartons, resulting in a gel-like product. For 'stirred' yogurts, coagulation is carried out when transferring through a cooler and then packaging. A variety of gums are used in yogurt production in order to reduce syneresis and enhance the rheological and organoleptic properties. It should be noted, however, that legislation in some countries prohibits their use. It is important that the gums do not mask the natural flavor of the yogurt and that they are effective at the product pH of about 4.3. Gums reportedly used are carboxymethyl cellulose, guar gum, locust bean gum, pectin, alginate, agar, carrageenan, and xanthan gum. However, the most common are pectin, gelatin, and modified starch. They are normally used at levels of 0.2–0.5% depending on the gum and the milk solids content. They are often added to the warm milk prior to pasteurization in order to insure complete dissolution. In some cases the gum is added to the coagulum after incubation, in which case its microbiological quality needs to be assured. Fruit-flavored yogurts are produced by adding fruit pulp, consisting of fruit purée (~ 50%), sugar (~30%), and a suitable stabilizer to the yogurt prior to packaging. Low-ester pectins (and low-ester amidated pectins) are particularly effective stabilizers since they are able to provide the optimum rheological properties at low pH and are incorporated at typical concentrations of ~0.6%.

Cheese In cheese manufacture, lactic acid-producing bacteria are added to milk and the mixture is stirred at a constant temperature. When sufficient acidity has developed, rennet, which is a natural enzyme, can be added and this hydrolyzes the kappa-casein in the milk, causing the casein micelles to coagulate and form a gel. The gel is commonly referred to as curd and is separated off from the liquid (known as whey) and may be pressed, salted, and left to ripen according to cheese type. Addition of gums such as guar gum, locust bean gum, and carrageenan increases the rate of coagulation and aids the recovery process, thus increasing the yield of curd. In soft cheeses, where there is a high water content (> 80%), gums can lead to improvement in the body and texture of the product and are also reported to reduce water loss. Blends of carrageenan and galactomannan are particularly effective. Carrageenan is also used in some processed cheeses, and this allows the cheese content to be reduced whilst maintaining good mouth feel and melting properties and improving grating and slice integrity. The use of gums in cheese is, however, subject to food additive regulations and they are generally only allowed in composite cheese products and spreads.

Whipped cream Whipping cream has a high milk fat content (> 30%) and its functional requirements include its ability to incorporate air (i.e., whippability), foam stability, and resistance to serum separation. Sodium alginate, carrageenan, and modified starches can be used to improve these characteristics.

Bakery Products

Bread In the process of bread making, flour, yeast, salt, shortening, water, and so-called 'improvers' are mixed together to form a homogeneous dough. During the process the glutenin and gliadin proteins present in the flour interact to form gluten, giving the dough its elastic characteristics. On standing, the yeast ferments the sugar components present in the flour, producing carbon dioxide which distends the dough, giving it a cellular structure. During baking the starch granules dispersed within the gluten matrix swell irreversibly (gelatinize), releasing amylose and absorbing water, which adds to the structure. The gluten is coagulated by the heat.

Although it is not common practice, gums such as guar, locust bean gum, carboxymethyl cellulose, and xanthan gum can be used in bread manufacture at levels of up to 2%. They have been reported to speed up the development of gluten during mixing, giving a significant reduction in mixing time. In addition, it is claimed that they enhance moisture retention, thus retarding the staling process, which results from the aggregation of amylopectin chains, a process referred to as retrogradation.

Xanthan gum has been used in the preparation of gluten-free bread. It is believed that the xanthan interacts with the starch to form a matrix, allowing structure development comparable to normal bread to occur during baking.

Cakes Gums such as xanthan gum, guar gum, and carboxymethyl cellulose can be used in the production of cakes and they have two major functions. First, they can control the rheological properties of the batter, which is important in large-scale manufacture, where mixing, pumping, and filling stages are encountered. Their shear thinning characteristics are important since a low viscosity is preferable when

mixing and pumping, whereas a high viscosity is required after filling the baking pans in order to prevent splashing and to suspend any particles such as fruit in the batter. The second function is to facilitate more even hydration of the dry ingredients and to control the moisture of the finished product. A more even moisture distribution assists in the stabilization of air cells formed during mixing and this gives rise to a finer, more uniform cell structure, thus improving the volume and texture of the finished cake. Control of the moisture content is of utmost importance since this will influence not only the texture and appearance but also the shelf-life.

Fruit pie fillings Fruit pie fillings usually contain modified starches, which thicken the acidic juices associated with the fruit. Gums such as carrageenan, carboxymethyl cellulose, and low-methoxy pectin and locust bean gum blends are sometimes used in combination with the starch to improve heat stability, to give increased clarity, and to reduce syneresis.

Sodium alginate can be used to produce fruit analogs for pie fillings utilizing the so-called 'internal setting procedure.' In this process a sodium alginate solution containing an insoluble calcium salt (e.g., calcium phosphate) is combined with a fruit purée which contains a sequesterant (e.g., sodium citrate) and acid (citric acid). On mixing, the insoluble calcium salt begins to dissolve in the presence of the acid, releasing calcium ions which initially interact with the alginate, inducing gelation. The uniform gel formed is then cut and a thickened fruit syrup is added to give the finished pie filling.

Sodium alginate can also be used to prepare structured fruits using a coextrusion technique and this is particularly appropriate for fruits such as blackcurrants, which have an outer skin and liquid center. In the process a sodium alginate solution and fruit purée mix are fed separately through nozzles consisting of coaxial tubes positioned above a bath containing a solution of calcium salt (e.g., calcium lactate). The flow of alginate solution is kept constant while that of the fruit purée is intermittent. As each drop of purée breaks away from the nozzle it is coated with a film of the alginate solution which gels instantly, forming a hardened skin on hitting the calcium bath below. The fruit gels formed are then collected and a thickened syrup is added to produce the pie filling.

Gellan gum is now also finding some applications as a gelling agent in fruit fillings for pies and cakes.

Icings, Frostings, and Glazes The icings, frostings, and glazes used for bakery products consist of a saturated solution of sugar in water. Gums such as carboxymethyl cellulose and sodium alginate are used to inhibit sugar crystal growth and to control the rheological and film-forming properties. Such coatings show improved adhesion to the product, have less tendency to crack, and have reduced stickiness.

Gum tragacanth is used in decorative icings for wedding cakes.

Batter Coatings Thickeners such as guar gum, locust bean gum, and xanthan gum are often used in the large-scale production of batter coatings for seafood and poultry. They provide the required rheological properties to the batter and assist in its adhesion to the product.

Desserts

Specialty desserts There is a vast assortment of dessert products available today for the consumer to choose from, and gums are widely used to control the characteristics of the various components, which make up each product. Carrageenan, for example, is often used in milk-based desserts since it interacts with the milk proteins present, giving rise to the formation of a weak gel structure and thus preventing serum separation. Thickeners such as guar and locust bean gum or gelling agents such as pectins are used to control the viscous or gel-like characteristics of the liquor associated with fruits, while sodium alginate is often used to stabilize whipped-cream toppings. Examples of typical dessert products, with details of the additives incorporated into each, are given in Figure 2.

Gums are also commonly used in frozen gateaux in order to confer the desired properties on the individual components and, in particular, to enhance freeze–thaw stability.

Table jellies Jellies are traditionally prepared using gelatin. The material is dissolved in hot water and gelation occurs on cooling in a refrigerator. The gelatin molecules in solution undergo a coil-to-helix transition, which occurs at about 25 °C, and the helices aggregate to form a gel. The gels formed have high clarity and on heating melt at 37 °C, which corresponds to body temperature. The jelly, therefore, melts in the mouth, providing rapid flavor release and a smooth texture, but they are prone to toughening on storage.

Jellies can also be prepared using gums such as carrageenan, alginate, and pectin. Iota-carrageenan gives rise to compliant gels similar to those of gelatin. However, they have a higher melting temperature and hence do not have the same attractive organoleptic properties but since they also have a higher gelling temperature, the gels form without the need for refrigeration. Kappa-carrageenan itself forms very

Figure 2 (see color plate 80) Typical desserts incorporating gums. The additives contained in each dessert include the following: chocolate delight: modified starch, carrageenan, sodium alginate; caramel dessert: modified starch, carrageenan, pectin; raspberry royale: amidated pectin, sodium alginate; pineapple cheesecake: modified starch, gelatin, pectin, guar gum; strawberry fruit fool: gelatin, sodium alginate.

brittle gels but can be used in conjunction with (purified) locust bean gum, which improves the clarity and makes the gels more elastic (less brittle) and less prone to syneresis. Jellies can also be prepared using alginate with a high mannuronic acid content. Soft-textured, nonbrittle gels can be produced with calcium ions under controlled conditions.

Low-acyl gellan gum can be used to prepare dessert jellies with a firm brittle texture. Mixtures of low- and high-acyl gellan gum can be used to produce jellies with a range of textures. Low-acyl gellan can also be used to modify the texture of gelatin gels. In Japan, konjac mannan is commonly used to produce dessert jellies.

Syrups, toppings, and dry dessert mixes Gums are commonly used to enhance the viscosity of syrups for use on pancakes and icecream. They provide the necessary flow characteristics and cling and, in addition, retard sugar crystallization. Xanthan gum and carboxymethyl cellulose give products of higher clarity than those prepared using guar gum or locust bean gum. Propylene glycol alginate is used in some buttered syrups because of its emulsifying characteristics. Xanthan gum, guar gum, and locust bean gum are often used in whipped nondairy dessert toppings in order to control the viscosity of the system prior to whipping. Methyl cellulose and hydroxypropylmethyl cellulose can also be used in this application and they have the added benefit of promoting emulsification of the oil, although other emulsifiers can also be present. Microcrystalline cellulose has been found to be particularly effective at stabilizing foams. The toppings are normally sold frozen and the gums present inhibit ice crystal formation.

Gums are often used in dry dessert mixes. Xanthan gum, carboxymethyl cellulose, and guar gum are particularly effective since they are soluble in cold water. Pectins and alginates are included in some dessert mixes, such as mousses and cheesecakes, in order to provide a gel structure. Xanthan gum is particularly useful in dry sorbet mixes because of its good solubility and stability in acid conditions and also because it can form a weak gel structure, which traps air bubbles formed during mixing.

Carrageenan is widely used in dry milk pudding mixes because it interacts with the casein present, leading to the formation of a gel structure.

Jams, Jellies, and Marmalades

The formulations of jams, jellies, and marmalades in Europe are controlled by a European Union directive which dictates the minimum fruit and sugar solids required for the various categories. Manufacture consists of blending the ingredients, raising the temperature, and evaporating to the correct soluble solids content either at 100 °C at atmospheric pressure or at temperatures down to 60 °C under vacuum. Vacuum boiling avoids heat damage, caramelization, and loss of fruit volatiles, but an additional pasteurization step is required.

The primary characteristic of all preserves is that they are gelled systems, and for conventional preserves pectin is used as the gelling agent. For preserves with a high total soluble solids (TSS) content (> 60%, mainly sugar), one or more high-methoxyl pectins are

used to control gelation. At TSS content of 20–55%, either an amidated or a nonamidated low-methoxyl pectin is used. Gelation of high-methoxyl pectins is pH-dependent and a pH of < 3.5 is required for gelation to occur. This may require the addition of fruit acids to the recipe to reduce the pH, with most jams having a final pH of 2.8–3.4. For low-methoxyl pectins, gelation is controlled by the presence of calcium ions and a high soluble solids content is not required. Gelation is optimized at pH 3–4.5.

In high-sugar preserves the sugar stabilizes the structure and immobilizes the water, thus reducing syneresis. This becomes much more of a problem in low-sugar preserves.

A number of reduced-sugar products are available with TSS contents of < 25%. In these systems gelation is achieved using carrageenan or carrageenan and low-methoxyl pectin blends. Other gums such as guar and locust bean gum may be added to modify the texture and reduce syneresis.

Meat, Stews, Gravies, and Soups

Gums, notably carrageenan and guar gum, are often incorporated into meat products such as sausages, cooked hams, and poultry products where they serve as binders to control the texture and structural integrity and to retain water.

Methyl cellulose is used in fryable products since it gels on heating and hence helps to retain the product's shape during frying.

Gums such as modified starches, guar gum, and carboxymethyl cellulose are used to thicken the gravies for meat pies, canned meats, and stews and to reduce fat migration and water separation during storage.

Modified starches are used to thicken canned soups, while guar gum is commonly used in instant dry soup mixes to add body and aid dispersion of the various ingredients.

Confectionery

Gelatin, modified starch, gum arabic, and pectin are the main gums used in confectionery products. Gum arabic is the major component in traditional wine gums and is present at concentrations of ~50%. Production involves dissolving the gum in water, keeping the temperature as low as possible (< 60 °C) in order to avoid precipitation, and it is then added to a preboiled sugar/glucose solution (70%) followed by flavorings and colors. After standing the liquid is deposited into starch trays and placed in a stoving room for 4–6 days. The gums are then removed from the starch molds, brushed to remove starch, and often glazed with oil or wax. Softer gums or pastilles can be obtained by reducing the stoving time. In view of shortages in gum arabic and severe price fluctuations, considerable efforts have been made to find alternative gums. Nowadays pastilles are commonly made with gum arabic in combination with other gums, notably, starch, maltodextrin, gelatin, pectin, and agar.

Slow-setting, high-methoxyl pectin is used in the preparation of fruit-flavored acidic confectionery jellies and is usually present at concentrations of less than 2%. Some flavors, such as liquorice and vanilla, are unstable in the acid conditions necessary for the gelation of the pectin, and low-ester pectins which gel in the presence of calcium are used instead. These pectins are also used in the manufacture of Turkish delight.

See also: **Bread**: Chemistry of Baking; **Cakes**: Nature of Cakes; **Cream**: Types of Cream; Clotted Cream; **Dairy Products – Nutritional Contribution**; **Dressings and Mayonnaise**: Chemistry of the Products; **Gums**: Properties of Individual Gums; **Ice Cream**: Methods of Manufacture; **Jams and Preserves**: Methods of Manufacture; Chemistry of Manufacture; **Meat**: Hygiene; **Sweets and Candies**: Sugar Confectionery; **Syrups**; **Yogurt**: The Product and its Manufacture

Further Reading

Glicksman M (1982, 1983, 1986) *Food Hydrocolloids*, vols 1–3. CRC Press.

Harris P (1990) *Food Gels*. London: Elsevier Science.

Imeson A (1992) *Thickeners and Gelling Agents for Food*. Glasgow: Blackie Academic and Professional.

Nussinovitch A (1997) *Hydrocolloid Applications; Gum Technology in the Food and Other Industries*. Glasgow: Blackie Academic and Professional.

Phillips GO and Williams PA (2000) *Handbook of Hydrocolloids*. Cambridge: Woodhead.

Stephen AM (1995) *Food Polysaccharides and Their Application*. New York: Marcel Dekker.

Williams PA and Phillips GO (2002) *Gums and Stabilisers for the Food Industry 11*. Cambridge: Royal Society of Chemistry.

Dietary Importance

C A Edwards, University of Glasgow, Glasgow, UK

Background

The dietary importance of gums is related to their effects at various sites in the gastrointestinal tract from the stomach to the large intestine. Most of their actions in the upper gut have been related to

the development of a high viscosity in the gut lumen. However, gums in natural foods may not be released in sufficient quantities to increase luminal viscosity. Isolated gums or foods fortified with gums, however, can have very significant effects of nutrient absorption and postprandial plasma nutrient levels. Other important effects relate to the fermentation of the gums by the microflora in the large intestine. A high intake of particular gums may help in the treatment of constipation and diabetes, and in the prevention of colon cancer. However, only the benefits for treating constipation have been well established, and more research is needed in their use for other conditions.

Physiochemical Properties and Physiological Action

In this context, the term 'gum' covers a wide range of polysaccharide molecules that have varied structures, physical properties and physiological actions. The term, for the most part, could be interchanged with 'soluble nonstarch polysaccharide' or soluble dietary fiber (*See* **Dietary Fiber**: Properties and Sources). There are many different dietary sources of gums. They are found naturally in plant cell walls (pectins and hemicelluloses), as plant exudates (gum arabic) or mucilages (ispaghula[1]) as well as being produced by some bacteria (xylan) and algae (alginates). The effects of gums on human physiology will depend on whether the gum is ingested as part of a plant cell wall, as part of a food complex, or as an isolated polysaccharide.

Most, but not all, gums produce viscous solutions in their isolated form, and some (such as pectin) have cation exchange properties. The extent of the viscosity or cation exchange depends on the length of the molecule and the constituent sugars. The physiological action of such gums in the diet can therefore vary enormously and may differ from one dietary source to another. If the gum is ingested within the framework of a plant cell wall, as with pectin in carrots, the pectin may not be released in sufficient amounts to have a significant action on small intestinal function but may still be available for fermentation in the large intestine. In addition, the cation exchange properties of an isolated pectin may be greater than pectin contained in a food because the potential binding sites may already be taken up by minerals or other charged molecules found in the food matrix.

Another important consideration is the amount of gum ingested. The amount in foods, or added as stabilizers, emulsifiers, and fat substitutes, is small, and most of the effects described below require a substantial intake of isolated gum. For example, the standard dose of ispaghula for treatment of constipation is 3–11 g per day. The dose recommended for the action of guar gum on small intestinal absorption is 15 g per day, whereas the amount of pectin in carrots is 1.5 g per 100 g, and there is only a trace of gums in icecream.

[1]The terms 'ispaghula' and 'psyllium' (used later) are interchangeable.

Table 1 Properties of gums that affect human physiology and health

Property	*Consequence*	*Main examples*
Viscosity	Slower absorption Lower blood glucose Reduced cholesterol Increased satiety	Guar gum, pectin, ispaghula (psyllium), oat gum
Cation exchange	Binding of minerals Binding bile acids Reduced cholesterol	Pectin
Fermentability	SCFA production Low colonic pH Altered microflora	Oat gum, pectin, guar gum
Butyrate production	Colonic fuel Colonic health Possibly anti cancer	Oat gum
Stool bulking	Reduced constipation Dilution of carcinogens	Ispaghula (psyllium), gellan

The physical characteristics of each gum are determined by sugar composition as well as the order of sugars and degree of modification, e.g., the degree of methylation of pectin. Viscosity results from the entangling of free carbohydrate chains. This requires a disorganized network, and so the viscosity of gums is determined by their molecular weight and the degree of irregular versus regular sugar patterns, the latter allowing more ordered cross-linkage between chains.

The gums have a range of physiological and nutritional effects (Table 1) mediated mostly through effects on transit through the stomach, small intestine, and colon, by reducing mixing actions in the gut, by their actions on the interaction between nutrients, enzymes, and mucosal cells, and finally as a result of their fermentation by the colonic microflora.

Effects in the Stomach and Small Intestine

The effects of isolated gums on gastric emptying and digestion and absorption have been well studied. The effects of the same gums in foods are less well established. Viscous gums eaten as part of a mainly liquid meal can slow gastric emptying by making the liquid meal of a more solid consistency. Guar gum, which has one of the highest viscosities, can delay gastric emptying half time by 60%. However, it has also been shown that increased viscosity can facilitate the emptying of discrete solids. In this case, they may

act as lubricants and ease passage through the small 2-mm opening of the pyloric sphincter.

In the small intestine, the main impact of increased viscosity is to resist the movement of the intestinal contractions, reducing the mixing of the digestive enzymes with their substrates and inhibiting the movement of the nutrients to the absorptive epithelium. This can give the impression of an increase in the thickness of the unstirred water layer adjacent to the intestinal mucosa, through which nutrients must pass by diffusion. These two effects, along with slowed gastric emptying, can delay the process of digestion and absorption, resulting in decreased postprandial glycemia and potentially reducing fat absorption.

In addition, the viscous matrix of a gum solution can trap endogenous molecules such as bile acids. Bile acids are essential for the emulsification of fats to allow digestion, and for the formation of micelles to facilitate movement of the products of fat digestion to the mucosa for absorption. A decrease in bile acid concentration will therefore disrupt and delay fat digestion. If cation exchange properties persist, then the bile acids may be bound to the charged gum, and this will have an additional impact in fat digestion. The trapped bile acids, if not released in the large intestine during fermentation, are excreted in feces, which will result in reduced plasma cholesterol as replacement bile acids are synthesized from this precursor.

There are, however, some problems with this simplistic explanation of the effects of viscous gums on postprandial glucose and fat absorption. The viscosity shown in *in vitro* models to delay digestion and reduce mobility of nutrients may not be achieved in the part of the small intestine where most digestion and absorption occurs.

Most digestion occurs in the duodenum and jejunum. When foods enters the duodenum, pancreatic secretion of large amounts of bicarbonate and enzyme rich juice is stimulated. This relatively large volume of fluid will dilute the small intestinal contents and considerably reduce the concentration and therefore the viscosity of the food gum. The relationship between concentration and viscosity is not linear but exponential, and changes in concentration can have a major impact on the viscosity maintained. Experiments in rats have failed to find any large increases in upper small intestinal viscosity after ingestion of large amounts of guar gum. The gums are not digested, however, so the viscosity is regained as the nutrients and fluid are absorbed. Thus, the viscosity in the ileum where bile acids are reabsorbed, as part of the enterohepatic circulation, will be much higher. It is well established, however, that postprandial blood glucose is reduced considerably after ingestion of gums. Bread containing lower-molecular-weight guar still has the action of reducing postprandial glycemia, again suggesting that viscosity may not be the critical factor.

Blood Glucose and Diabetes Mellitus

In the past, the use of gums such as isolated solutions of guar gum to reduce the rise in postprandial blood glucose was thought to be of some therapeutic use in diabetes mellitus. Although the effects of guar gum reducing postprandial hyperglycemia and insulinemia are well established, this approach in treating diabetics was not very successful, as the gum solutions used had poor organoleptic properties with an undesirable mouth feel and aftertaste. Since then, the incorporation of gums such as guar and oat gum in bread products has been explored. It has been found that very palatable breads with proven effects on intestinal absorption rate can be produced, but their use in the diabetic diet has still to be established. There has also been much recent interest in the production and action of cereal products containing oat β-glucans, ispaghula (psyllium), and Konjac Mannan.

Therapeutic Enteral Diets

Sometimes, patients, including children, have a problem in dealing with normal foods and are fed a simplified synthetic fluid diet of easily assimilated nutrients. These diets often have to be liquid to allow administration through tubes directly into the gastrointestinal tract. Until recently, these elemental or polymeric diets have contained no source of dietary fiber. This can result in constipation and, after prolonged use, produce atrophy of the colonic mucosa. Some of these liquid diets have now been formulated with soluble gums as a source of dietary fiber to prevent these symptoms and to maintain a healthy microflora. The gums used include soy polysaccharide, pectin, and carboxymethylcellulose.

Effect of Gums on Plasma Cholesterol

Heart disease is one of the major killers of the developed world, and one of the main risk factors is high plasma low-density lipoprotein (LDL) cholesterol coupled with low high-density lipoprotein (HDL) cholesterol. HDL cholesterol is associated with clearance of sterol from the peripheral tissues to the liver. Many of the viscous food gums (oat, pectin, guar, psyllium, and others) have been shown to have a significant and long-term effect in reducing LDL cholesterol while having little effect on, or increasing, HDL cholesterol. The effect of the gums is modest but may be of significance in reducing the risk of heart disease.

Table 2 Possible mechanisms for the effect of gums on serum cholesterol

Inhibition of fat digestion
Viscosity inhibits mixing of enzymes and lipids
Viscosity/cation exchange traps/binds bile acids
Viscosity inhibits formation of micelles
Viscosity inhibits movement of micelles
Increased bile acid synthesis from cholesterol
Binding of bile acids by cation exchange, viscosity
Increased fecal bile acid losses
Increased fecal cholesterol losses
Inhibited cholesterol absorption
Increased intestinal cell losses
Decreased cholesterol synthesis
Inhibition of hepatic cholesterol synthesis by propionic acid (probably only a pharmacological effect)

It is not clear by which mechanisms the gums act (Table 2). They could reduce fat digestion either directly or by interfering with micelle formation. Although there is likely to be only a small increase in total fat losses, the delaying of fat absorption can result in a change in the type of chylomicra or lipoproteins produced. For those gums like pectin with cation-exchange properties, the bile acids may be bound to the carbohydrate molecule, which are then excreted in the feces. Viscous fibers without cation exchange may also trap bile acids within their matrix. Several gums, including guar, pectin, and ispaghula (psyllium) have been shown to increase fecal bile acid losses, and this has been associated with reduced cholesterol. However, if the fiber is fermented, the bile acid would be released in the large intestine where they could be converted to secondary bile acids and returned to the bile acid pool. Moreover, low-molecular-weight (and therefore low-viscosity) guar gum in bread still reduced cholesterol and not all studies demonstrating an effect on cholesterol show an increase in fecal bile acid. If fecal bile acid losses are increased, cholesterol in the liver is used to replace them, and hence plasma cholesterol falls.

Cholesterol levels could also be decreased due to the higher fecal losses of cholesterol from increased intestinal cell loss seen with viscous fibers, and there has been much discussion about the potential actions of propionic acid produced from the fermentation of the gums. However, the inhibitory effects of propionic acid on hepatic cholesterol synthesis appear to require pharmacological levels of propionate rather than those that are produced from the fermentation of food gums under normal conditions. Acetate is a precursor of fat synthesis and lactulose a highly fermentable saccharide increased serum LDL cholesterol when fed to healthy volunteers. This was associated with higher levels of serum acetate in some subjects. It is evident that the actions of gums on serum cholesterol are mediated by different mechanisms for individual gums and for the same gums under different circumstances.

Satiety and Obesity

If gums had a major effect on satiety, they could be a useful adjunct to the prevention and treatment of the increasing public health problem of obesity. However, although highly viscous gums such as guar have been shown to increase satiety and even promote some weight loss in short-term studies, the unpalatability of such gums and the moderate effects they have on weight make them of unlikely benefit in weight-loss or weight-maintenance programs.

Gums and Mineral Bioavailability

Some food gums with cation-exchange properties have been shown *in vitro* to bind minerals such as calcium. Even those gums without direct binding sites may still entrap these minerals and delay or inhibit their absorption. *In vivo* human studies have shown an increase in fecal mineral output, but the overall mineral balance is not affected in adults. The effect in young children is unknown and is one of the reasons why a high dietary fiber intake has not been encouraged in the under fives. However, there is no evidence of significant mineral imbalance in children on high-fiber diets, and the effects are unlikely to be great. Moreover, there are several studies now suggesting increased calcium absorption in the large intestine, perhaps encouraged by the reduced pH from the bacterial fermentation of soluble fiber.

Effects of Gums in the Large Intestine

Most food gums are readily fermented in the large intestine of man by the diverse and dense microflora. The products and consequences of this fermentation may have a major influence on human health, and these effects may outweigh the impact of the food gums in the small intestine.

Colonic Fermentation

The human colonic microflora is thought to contain over 400 different species (perhaps even more as new molecular techniques, which measure bacterial DNA and rRNA, are discovering that the majority of bacterial groups have never been cultured (*See* **Microflora of the Intestine:** Probiotics). The mainly anaerobic bacteria use the carbohydrates to produce short-chain fatty acids (SCFA), mostly acetic, propionic, and butyric acids, and the gases CO_2, H_2, CH_4,

and H_2S. Colonic pH is also determined by the balance between the rate of fermentation and the buffering of the SCFA by bicarbonate exchanged when the SCFA are absorbed.

The fermentability of each gum depends on its structure and water solubility. The presence of bacteria with the required enzymes for each step of gum degradation is also essential. The degradation and fermentation of a complex carbohydrate structure may be the result of a synergy between several bacterial species, and some of the enzymes may take time to be expressed either by induction of enzymes, in originally predominant bacteria, or by an increase in dominance by new groups. It has been shown in rats that it can take many weeks before the colonic microflora is truly adapted to the introduction of a new gum into the diet.

The extent and rate of fermentation of food gums are important. The rate of fermentation may determine the site of SCFA production. Most colonic disease occurs in the distal colon, and yet most fermentation occurs in the proximal colon. Therefore, food gums that are slowly fermented may encourage prolonged fermentation and SCFA at more distal sites.

The pattern of SCFA produced from each gum also differs. Pectin increases the proportion of acetic acid produced, guar gum and ispaghula tend to promote the production of propionic acid, and oat gum increases production of butyric acid.

Health Impact of Increased Colonic Fermentation

The colonic fermentation of carbohydrates including food gums can result in a substantial fall in colonic luminal pH. There are several benefits of reduced colonic pH, which include the inhibition of undesirable bacterial activities such as 7α dehydroxylase, which forms secondary bile acids that can help to promote cancer. Reduced colonic pH can also lead to precipitation of fatty acids, sterols, and other molecules, which have been implicated as procarcinogens. Low pH also reduces the rate of colonic cell proliferation, thought to be a risk factor for colon cancer. However, the SCFA produced during fermentation can stimulate cell proliferation, and so under normal conditions, the effects of pH here may be canceled out. It has been noted, however, that populations in South Africa with a low risk of colonic cancer also have a lower colonic and fecal pH than those from high-risk populations.

The side-effects of increased colonic fermentation and of ingestion of large amounts of food gums include flatulence and abdominal distention. These are usually associated with eating large amounts of fermentable material but may reduce over time as the colonic flora adapts to the new carbohydrate substrate supply and adopts a more efficient fermentation process. This may take over a week.

The SCFA have many potential actions that are generally beneficial for health. They promote the absorption of water and electrolytes and inhibit the growth of pathogens, reducing the risk of diarrhea. They all stimulate cell proliferation throughout the gut, even though they are produced in the large intestine, and this may be important for wound healing after gut surgery or after gastrointestinal disease. In addition to these general effects, each fatty acid has its own idiosyncratic actions.

Actions of Individual SCFA

Acetic Acid

Acetic acid is always the main SCFA produced, usually making up 50–70% of the SCFA. Acetic acid is the only SCFA to reach the systemic circulation in significant amounts. It can provide energy for muscles and other tissues and is noninsulinogenic. Acetate has less effect on the colonic enterocytes in terms of proliferation and motility than the other SCFA. Pectin has been shown to increase the proportion of acetate when fermented by fecal bacteria *in vitro*. Acetate produced during pectin ingestion may also beneficially affect fibrin network architecture in patients with hypercholesterolemia, reducing the risk of heart disease.

Propionic Acid

Propionic acid is usually the second most abundant SCFA produced during fermentation. Propionic acid is absorbed into the portal vein but is mainly removed by the liver. It is gluconeogenic, unlike the other SCFA, and was previously thought to inhibit cholesterol synthesis by the liver. However, although this effect has been well demonstrated under experimental conditions, it is now thought that not enough propionic acid is formed in the colon and absorbed into the portal vain to have this effect. However, it is worth noting that those gums that are associated with cholesterol-lowering effects are often those that promote propionate production.

Butyrate

Butyric acid is the preferred fuel for the colonic enterocyte. It has also been associated with several potential anticancer properties. An increase in the rate of colonic cell proliferation is believed to be a risk factor for cancer. However, all cells in the colon are

programmed with a time to die by a controlled process called apoptosis. It is probably the balance between the levels of proliferation and apoptosis that determine whether a cancer will or will not develop. Although butyrate has been shown to stimulate cell proliferation of normal colonic cells, it has been shown to stimulate apoptosis in cancer cells *in vitro*. It may also affect other important stages in carcinogenesis such as the promotion of cell differentiation and inhibition of histone deacetylase, which may promote DNA repair.

Gums as Prebiotics

There has been much recent interest in the beneficial health effects of certain strains of lactic acid bacteria or probiotics believed to have various health benefits. These include reducing the duration of rotavirus diarrhea, reducing eczema in atopic infants, and potentially having protective actions against intestinal diseases, including colitis and cancer (*See* **Microflora of the Intestine:** Probiotics). These probiotics are usually bifidobacteria, lactobacilli or streptococci but may include certain strains of *E. coli*. The survival of these bacteria in the gut can be enhanced by the presence of certain fermentable carbohydrates. These are called 'prebiotics,' but to fulfill the definition, carbohydrates must selectively increase the populations of the probiotic organisms. The most accepted prebiotics are fructo-oligosaccharides and resistant starch, but some gums may also increase bacterial populations.

Calories from Fermentation of Gums

SCFA produced from fermentation can be used by the colonic cells and other body tissues for fuel. It is now recognized that dietary fibers, including gums, contribute energy, but the exact amount depends on the fermentability and any increased fecal nitrogen losses. After absorption, the SCFA have to be acylated before entering the major metabolic pathways, and this also has an energy cost. It is now estimated that, on average, fermentation provides 8.4 kJ (2 kcal) per gram of carbohydrate.

Constipation

If a dietary gum is poorly fermented, it may have sufficient residual water-holding capacity to increase stool output. Ispaghula (psyllium), which has a water-holding capacity of approximately 7 g g^{-1}, has been used successfully to treat mild to moderate constipation for many years. Gellan, a bacterially produced gum, has also been shown to be a potent stool bulker but is not used therapeutically. Other gums such as karaya and xanthan, which are slowly fermented, may increase stool wet weight and fecal SCFA without increasing stool solids.

See also: **Colon**: Cancer of the Colon; **Dietary Fiber**: Properties and Sources; Energy Value; **Microflora of the Intestine**: Probiotics; **Pectin**: Properties and Determination; Food Use; **Probiotics**

Further Reading

Blake DE, Hamblett CJ, Frost PG, Judd PA and Ellis PR (1997) Wheat bran supplemented with depolymerised guar gum, reduces the plasma cholesterol concentration in hypocholesterolemic human subjects. *American Journal of Clinical Nutrition* 65: 30–42.

Bonfield C and Kritchevsky D (1997) *Dietary Fiber in Health and Disease*. New York: Plenum Press.

Brown L, Rosner B, Willet WW and Sacks FM (1999) Cholesterol lowering effects of dietary fiber: a meta-analysis. *American Journal of Clinical Nutrition* 69: 30–42.

Canovale E, Tomassi G and Cummings JH (eds) (1995) Topics in dietary fibre research. *European Journal of Clinical Nutrition* 49(Supplement 3).

Eliasson AC (ed.) (1996) *Carbohydrates in Foods*, pp. 265–319. New York: Marcel Dekker.

Nutritional Role of Guar Gum

P Rayment* **and P R Ellis**, King's College London, London, UK

Introduction

It has been recognized for many years that guar gum, a galactomannan-rich legume flour, has considerable potential as a therapeutic agent in the dietary management of diabetes and hyperlipidemia. For example, it is now well established that guar gum can reduce the postprandial rise in blood glucose and insulin concentrations in response to glucose drinks and starch-rich meals in both healthy and diabetic subjects. Furthermore, the addition of guar gum to a normal diet can elicit long-term improvements in metabolic control in people with type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus. The ease of isolation and purification of guar gum explains why it is frequently used as a

*Now at Unilever Research Colworth, UK.

'model' non-starch polysaccharide in physiological and clinical studies. Also, the fact that it is well characterized with respect to other sources of dietary non-starch polysaccharides makes it an ideal candidate for clinical use. This article highlights the nutritional and therapeutic role of guar gum, but mainly focuses on its blood glucose- and cholesterol-lowering actions. Also included is a brief discussion of the physicochemical properties of guar gum, including its rheological behavior in model systems and its influence on the properties of human and animal digesta *in vivo*, which has an important, if not crucial, influence on gut function and metabolism.

Extraction and Composition of Guar Gum

Guar gum, a type of dietary fiber, is extracted from the seed endosperm of a leguminous plant (*Cyamopsis tetragonoloba* (L.) Taub.) indigenous to the Indian subcontinent. Figure 1 shows the pods, seeds, and endosperm extract of the guar plant. To recover guar gum from the seed, the endosperm halves (splits)

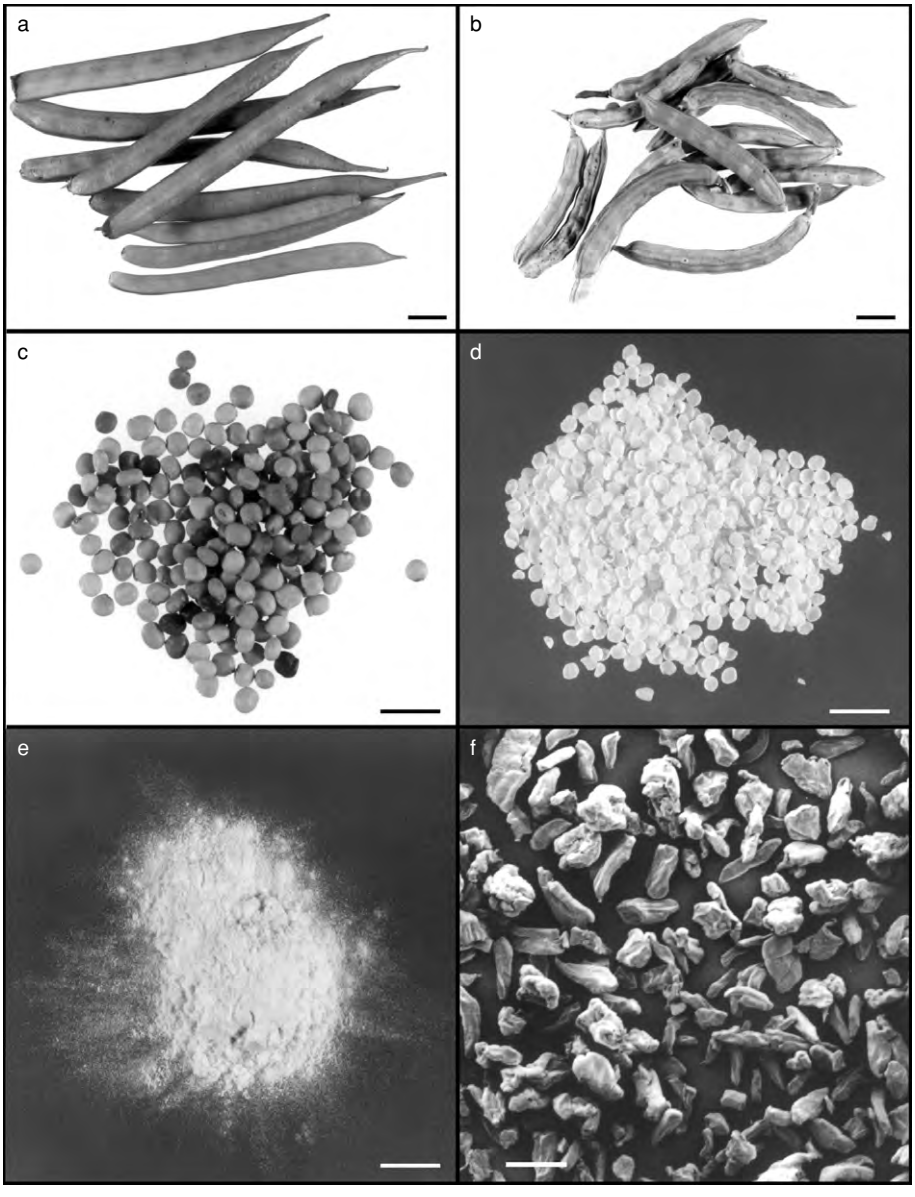

Figure 1 (see color plate 81) Pictures of pods, seeds, and endosperm extract of the guar plant (*Cyamopsis tetragonoloba* (L.) Taub.), a member of the Leguminosae family. (a) Green pods, scale bar = 1.3 cm. (b) Dried pods, scale bar = 1.6 cm. (c) Seeds, scale bar = 1 cm. (d) Endosperm halves (splits), scale bar = 1 cm. (e) Guar gum flour, scale bar = 1 cm. (f) Scanning electron micrograph of guar gum flour, scale bar = 100 μm.

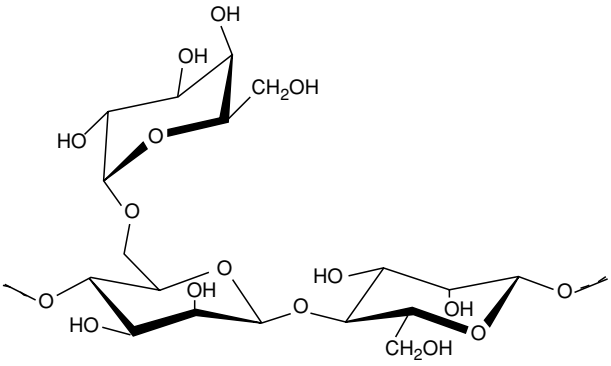

Figure 2 Structure of guar galactomannan showing the mannose backbone and galactose side chains.

must be separated from the hull and the cotyledon. The splits are ground to a fine flour and may then be purified by repeated alcohol washings. In this way guar gum is commercially produced as a flour and is frequently used as an additive (E412) in the food industry for its thickening and stabilizing properties.

Chemically, guar gum mainly consists of a high-molecular-weight polysaccharide, namely galactomannan, which is based on a mannan (M) backbone with galactose (G) side groups, as shown in Figure 2. The ratio of the two components (G:M) seems to vary slightly depending on the origin of the seed, but the gum is generally considered to contain approximately one galactose unit for every two mannose units. The linkages between the sugar groups cannot be hydrolyzed by human digestive enzymes in the small intestine but are fermented by bacterial enzymes in the large intestine. The products of this fermentation include short-chain fatty acids, mainly acetic, propionic, and butyric acids, which have important nutritional implications, notably in relation to colonic function.

The Food Chemicals Codex specifies a standard of not less than 66% of galactomannan in food-grade guar gum in the USA, although generally most samples contain more than 80% (w/w). The other components of guar gum, including fat and protein, which are considered to be impurities, vary between different sources. Table 1 lists the typical composition of a commercially available guar gum flour.

Table 1 Chemical composition of a commercially available guar gum flour

Component	*Concentration (g 100 g^{-1})*
Galactomannan	73.0–86.7
Protein	3.0–6.0
Crude fiber	1.0–4.0
Ash (total minerals)	0.8–2.0
Fat (petroleum ether extractables)	0.5–1.0
Moisture	8.0–14.0
Total impurities[a]	13.3–27.0

[a]Calculated as total nongalactomannan components.

Potential Adverse Effects of Guar Gum

In the food industry, guar gum is normally added to foods in concentrations of $< 1\,g\,100\,g^{-1}$ to provide a thickening, binding, or stabilizing function. Since guar gum is derived from a normal food material, it is classified by the Food and Agricultural Organization/World Health Organization as a substance of low toxicity. The safety of guar gum is substantiated by the results of a number of long-term studies in human subjects. In general, adverse effects, observed in some but not all subjects, appear to be associated with gastrointestinal disturbances. These effects, which sometimes appear to diminish with regular consumption of guar gum, include flatulence, gastrointestinal pain or discomfort, nausea, and diarrhea. Some of these side-effects probably result from bacterial fermentation of guar gum in the large intestine. The extent of the problems is generally related to the dose given, with more adverse gastrointestinal effects occurring when guar gum consumption is excessive (~ 15–$30\,g\,day^{-1}$). It has also been suggested that guar gum may reduce the bioavailability of a number of micronutrients (e.g., vitamins), although there is no evidence to indicate that this is of nutritional significance.

There is more concern about the use of macroaggregates of guar gum that have the potential for causing obstruction in the esophagus and gastrointestinal tract. This concern has arisen from reported incidents of esophageal obstruction and rupture in patients ingesting guar granules. There have also been a number of cases of esophageal and small-bowel obstruction in patients ingesting a particular type of guar-containing diet pill. The USA and Australia have banned such products. In the European Community (EC), food additive legislation prohibits the use of guar gum and other gums for the production of 'dehydrated foodstuffs that will rehydrate on ingestion.' However, the selling of guar products as medicine under license and administered to patients under medical supervision is permitted under EC legislation.

Physicochemical Properties of Guar Gum

Solution Properties

The ability of the galactomannan component of guar gum to hydrate and increase the viscosity, or thickness, of the intestinal contents is an important determinant of its physiological effects. The majority of polysaccharides used in nutritional studies exist in solution as 'random coils' and the number and

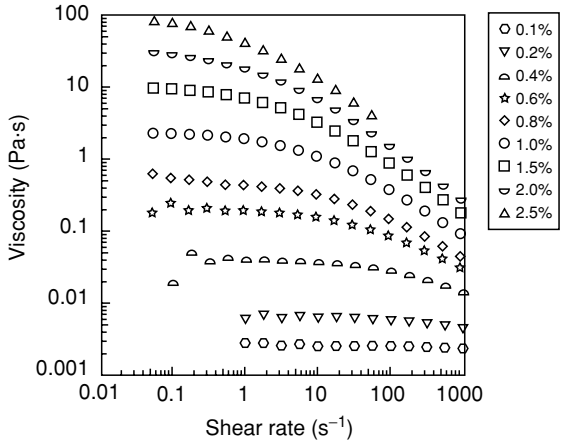

Figure 3 Shear rate and concentration dependence of viscosity for guar galactomannan solutions. Data taken from Rayment P (1996) PhD thesis, University of London.

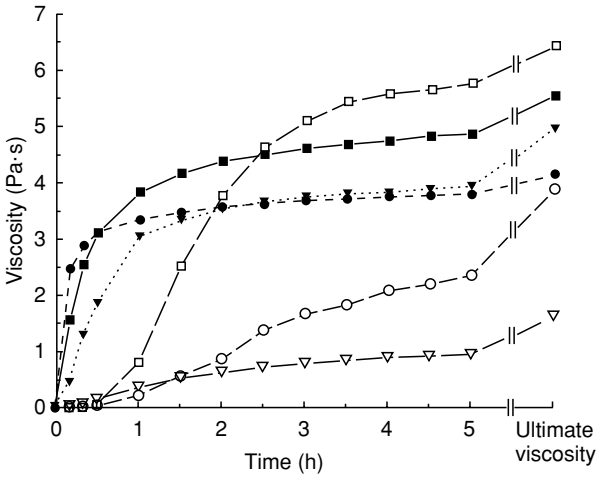

Figure 4 Hydration rate profiles of guar gum samples presented as viscosity (Pa · s) versus time. Experimental points are mean values from measurements on 4 replicate solutions of native guar gum flours M150 (—■—) and RG30 (- -●- -), pharmaceutical granule preparations Guarina (. . .▼. . .), Guarem (—□—) and Lejguar (—○—), and Glucotard minitablets (—▽—). The 'ultimate viscosity' is the measurement of viscosity taken after 24 hours of hydration following homogenisation of the guar solution with an ultra-turrax mixer. The final viscosity is taken only when a consistent maximum reading is recorded on the viscometer (usually after homogenising for 4–6 min). Data taken from Ellis PR & Morris ER (1991) Importance of the rate of hydration of pharmaceutical preparations of guar gum; a new in vitro monitoring method. *Diabetic Medicine* 8, 378–381.

molecular size of the polymer chains are critical determinants of viscosity and thus physiological activity. The high levels of viscosity associated with water-soluble polysaccharides originate predominantly by interpenetration of individual polymer chains to form entangled networks. At low guar gum concentrations, the individual polymer coils occupy a separate domain within the solution with little or no interpenetration. These solutions are described as Newtonian because their viscosity is independent of the rate of stirring or mixing (i.e., related to shear rate). At higher concentrations, where the polymer coils begin to overlap, the system becomes non-Newtonian and viscosity then depends on the shear rate. The more common type of non-Newtonian flow in food systems is shear-thinning (or pseudoplastic) behavior, where viscosity decreases with increased mixing. Figure 3 shows the effect of shear rate and concentration on the viscosity of guar gum solutions.

The viscosity or flow properties of guar gum can be determined by using various rheological techniques. Rheology (from the Greek word *rheos*, meaning stream) is a branch of physical science concerned with the study of the flow and deformation of materials under externally imposed forces. However, to predict the rheological behavior of guar gum in the human intestine, several factors require consideration. These factors include dependence on the mode of administration, the size of the dose and the physicochemical characteristics of the sample used as well as the physical state of the galactomannan in the gut lumen.

Hydration Kinetics of Guar Gum

The realization that viscous polysaccharides can be used for therapeutic benefit has led to the development of various pharmaceutical preparations containing guar gum. These include capsules or macroaggregates of guar gum, such as granules and minitablets, which have been designed to hydrate slowly in the mouth to improve palatability. However, their beneficial effects have been reported to be extremely variable, most probably due to their lack of hydration in the stomach and small intestine. Consequently, the hydration rate of guar gum in the gut lumen is a critical factor. Figure 4 compares the hydration rate profiles of a range of commercially produced samples of guar gum. These samples were hydrated in water and viscosity was measured over a 5-h period. Unsurprisingly, the marked differences observed are mainly attributed to the physical nature of the guar gum preparations. Thus, samples with a smaller particle size (flours < 150 μm) hydrate much more quickly than those with a larger particle size (> 1 mm), e.g., Glucotard minitablets. The results of many clinical trials have shown that the latter types are often physiologically ineffective.

Some research workers have also reported that the clinical efficacy of guar gum is improved when it is incorporated directly into a food rather than when taken as a pharmaceutical premeal supplement. Thus, a key factor in optimizing the glucose-lowering action

of guar gum is to insure that the polymer is in intimate contact with the starch, the main source of dietary glucose.

Nutritional and Physiological Effects of Guar Gum

Many research groups have reported the efficacy of guar gum in attenuating the postprandial rise in blood glucose and insulin concentrations in healthy and diabetic subjects (Figure 5). One reasonable explanation of these postprandial effects is that guar gum reduces the rate (and perhaps the extent) of digestion and absorption of available carbohydrate in the gastrointestinal tract. (Here available carbohydrate is defined as carbohydrate that is digested by α-amylase in the upper gastrointestinal tract and absorbed into the portal blood (mostly as glucose); this includes starch, dextrins and simple sugars (e.g., sucrose, lactose).) There are however problems with obtaining quantitative estimates of glucose absorption in humans; for example, access to the hepatic portal vein is very difficult. Nevertheless, studies in pigs, a useful animal model for studies on dietary polysaccharides, have provided the first direct evidence that guar gum decreases the rate of glucose absorption. In these types of experiment, glucose concentrations were measured simultaneously in the hepatic portal vein and peripheral blood (via the mesenteric artery) and the flow rate of the portal blood was measured also. Significant reductions in glucose absorption over 4 h were observed in the experimental animals given guar gum doses equivalent to those consumed by humans in clinical studies. In the same studies, insulin secretion was significantly decreased over the same period in response to the guar diet. This suggests that the attenuation in the plasma insulin levels seen in humans is attributed to a reduction in insulin secretion from the pancreatic islet β-cells in response to a lowered rate of glucose absorption. Also, a number of human studies have shown that guar gum decreases the postprandial rise in the insulinotropic hormones, plasma gastric-inhibitory polypeptide and glucagon-like peptide-1, which may partly explain the insulin-lowering action of guar gum. Another interpretation of data showing decreases in the peripheral blood insulin concentrations is that there has been an increase in hepatic extraction of insulin, a normal process by which insulin is removed from the blood circulation. Studies carried out to substantiate this have been largely contradictory, however.

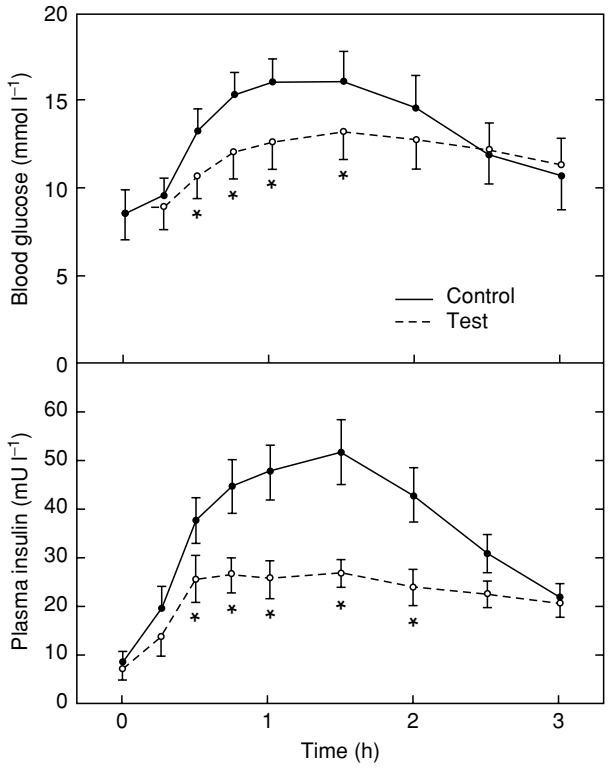

Figure 5 Postprandial blood glucose and insulin concentrations of patients with type 2 (noninsulin-dependent) diabetes mellitus in response to control and test breakfast meals of bread and marmalade (total available carbohydrate, mainly starch and sugars, was 106 g). The test contained 16 g of guar gum (in the bread) and 10 g of pectin (in the marmalade). *Statistically significant differences between control and test meals at individual time points. Figure redrawn from data published by Jenkins DJA, Leeds AR, Gassull MA *et al*. (1976) Unabsorbable carbohydrates and diabetes: decreased post-prandial hyperglycaemia. *Lancet* 2: 172–174.

In relation to long-term studies of insulin action in experimental animals and humans, it has been reported that guar gum increases (improves) insulin sensitivity. A reduction in sensitivity to insulin (i.e., 'insulin resistance') is seen as a decrease in the response of tissues to insulin stimulation and is considered to be an important risk factor for both coronary heart disease (CHD) and type 2 diabetes.

Mechanism of Action of Guar Gum

There is little doubt that the consumption of guar gum, whether mixed in a drink or solid food, can significantly modify digesta properties at all sites of the gastrointestinal tract. One important feature is the capacity of guar gum at relatively low concentrations to form a highly viscous network in solution. As mentioned previously, the behavior of guar gum in a food material is highly complex. In such systems there is unlikely to be one mechanism to explain fully the effects of guar gum in the human gut. A number of physicochemical mechanisms may be

involved, which are largely dependent on the type and form of the guar gum ingested. Much of the literature indicates that guar gum decreases the rate of glucose absorption into the hepatic portal vein by inhibiting the processes associated with digestion and absorption of available carbohydrates. These processes include gastric function, intestinal transit and mixing, α-amylase–starch interactions, and the movement of products of starch hydrolysis to the gut mucosa.

A number of experiments in humans and animals have shown that guar gum reduces the rate of gastric emptying of a meal. However, there are also studies that have shown no effect or even sometimes an acceleration in gastric emptying after the ingestion of guar gum. These contradictory results probably arise from differences in the type of guar gum used, the way in which it was administered, the proportion of solids and liquids in the meals, and the techniques used to measure gastric emptying. In general, it is considered that guar gum may reduce the rate of gastric emptying under some conditions, but this action is unlikely to be the only mechanism. Other groups have shown guar gum has additional effects on gastric function. In studies in dogs, guar gum was reported to impair gastric trituration and sieving by increasing the viscosity of the stomach contents. It was shown that there was a significant increase in the proportion of larger particles of food entering the small intestine when guar gum was incorporated into the diets. Possible explanations for these findings include suggestions that the increased viscosity stabilizes the suspension of larger food particles, which are likely therefore to escape maceration in the antrum of the stomach, and/or that the guar gum alters the contractile pattern of the terminal antrum, thus impairing the ability of the stomach to retain large particles.

There is evidence from human and animal studies that guar gum delays the transit of digesta in the small intestine. The postprandial pattern of gut motility also seems to be influenced by guar gum and other types of dietary fiber. Thus, whereas wheat bran and cellulose-supplemented diets were reported to produce prolonged bursts of intestinal contractions, guar consumption caused continuous contractions with a 50% reduction in amplitude compared with the other types of fiber. The effect of guar gum on the mixing behavior of digesta at different sites of the gastrointestinal tract has yet to be investigated, but it is likely to be extremely complex. It has been suggested that an increase in the viscosity of digesta will produce laminar or 'streamline' flow, rather than turbulent or disorderly flow, which is characteristic of less viscous fluids and facilitates efficient mixing of digesta in the gut. An inhibition of the digesta flow rate will inhibit physical mixing of nutrients and enzymes (e.g., pancreatic amylase). Also, laminar flow behavior would almost certainly have an effect on the rate at which nutrients are exposed to the epithelial surface and then absorbed into the hepatic portal vein.

However, an explanation of the precise mechanism by which guar gum modifies gut function is still elusive, despite the plethora of clinical and physiological studies that have been undertaken. The problem is that researchers understand little about the behavior of guar gum in the human gastrointestinal tract. Digesta is an extremely complex heterogeneous material and the effects of guar gum on this system have not been studied in great detail. Much of the work that has been done has involved investigating digesta in the stomach and small intestine, since it is assumed that guar gum has negligible effects on viscosity in the large intestine due to depolymerization by bacterial fermentation. There are a number of problems associated with measuring digesta rheology *in vivo*. In human subjects, there are obvious practical difficulties in gaining access to the sites of interest in the gut. Notwithstanding the initial problems involved with sampling intestinal contents, particularly in humans, the rheological properties of digesta will depend on the precise location of the sampling site. Thus, digesta is subjected to dilution by gastric secretions and then concentration by absorption of water, by both the intestine and the guar gum, at it passes further along the intestinal tract. Also the samples collected need to be tested quickly to prevent drying or dissolution of particulate materials. In the case of animal experiments, this requires access to rheometers or other measuring devices in the animal house.

Many factors complicate the interpretation of the rheological behavior of digesta containing guar gum. These include a lack of information on the hydration kinetics of guar gum *in vivo* and the contribution of undissolved food particulates to viscosity. Therefore, to enhance our understanding of the physical properties of digesta it is important to study model systems of entangled networks filled with particulates, although *in vivo* other factors such as gastrointestinal motility and fluid secretion/absorption also play a crucial role. In recent years, the rheological properties of guar gum solutions with increasing particulate concentrations have been investigated. The particulates studies included materials that were essentially spherical and rod-like in shape (i.e., starch granules and microcrystalline cellulose, respectively). The effect of particulates or 'fillers' is primarily to increase the viscosity above that of the pure guar gum system, as shown in Figure 6. The initial Newtonian flow

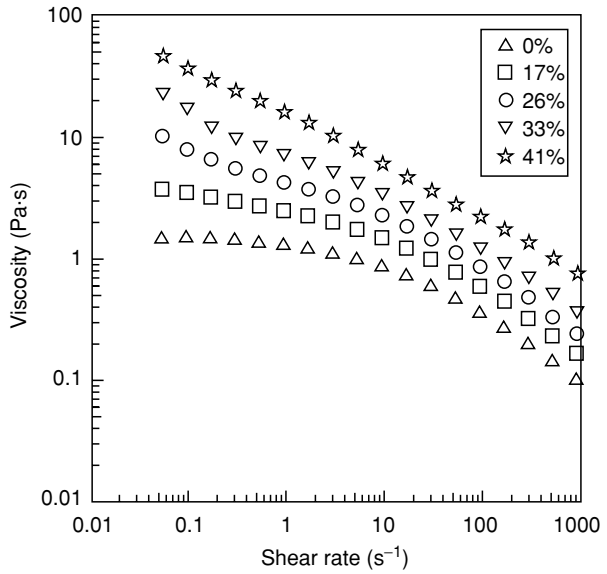

Figure 6 Effect of increasing rice starch concentration on the viscosity–shear rate flow curve of a 1% guar galactomannan solution. Data taken from Rayment P (1996) PhD thesis, University of London.

properties of the pure system at low shear rates become more rate-dependent on increasing particulate concentration – the so-called power-law behavior.

The flow data have been fitted to mathematical equations, which incorporate parameters that describe the shear-thinning nature of guar gum and the increasing solid-like behavior of these filled systems. The model also seems to work well with a more heterogeneous food material containing wheat starch. The extrapolation of such models to real food systems and even digesta is fraught with difficulty, however. The inherent problems with studying digesta have been mentioned already, but to make any real progress in this area, at the very least, information is needed on the fractional volume of particulate material at different sites of the gastrointestinal tract. These data could be obtained by recovering ileal effluent from ileostomy patients and by the recovery of digesta from cannulated animals at appropriate sites of the gut. Further work on the development of more heterogeneous models and techniques to study such systems is crucial if we are to increase our understanding of how guar behaves in the gut environment.

The mechanism by which guar gum reduces plasma cholesterol is still uncertain but, again, the discussion about the effects of guar gum on digesta viscosity and glucose absorption is important here also. The level of viscosity generated in the gut has been suggested as an important determinant of the hypocholesterolemic effect of guar gum in animal studies. The ingestion of guar gum appears to enhance fecal bile acid and neutral sterol excretion, reduce the rate of digestion and absorption of lipids, and inhibit synthesis of cholesterol. *In vitro*, guar gum has been shown to reduce lipid emulsification and the rate of lipolysis of emulsified triacylglycerols, although these effects have not been substantiated *in vivo*. Also, guar gum has been shown to inhibit both the rate of diffusion of cholesterol mixed micelles and hepatic cholesterol synthesis; the latter has been attributed to propionic acid, one of the products of guar fermentation in the large intestine. However, the inconsistency between *in vivo* and *in vitro* studies prevents any conclusive statements being made about the precise mode of action of guar gum.

Dietary Management of Disease

Although guar gum has frequently been used in clinical trials, interpretation of the results can be problematic. There are problems resulting from variable study design, lack of adequate controls, and heterogeneity of subject groups. Furthermore, many nutritional studies reporting the therapeutic effects of guar gum provide very limited information about the type and physicochemical properties of the sample tested, e.g., molecular weight and particle size. Moreover, sometimes it is not even known whether the guar was administered as a pre-meal supplement or incorporated directly into the subjects' meal. Despite these problems, there is sufficient evidence to show that guar gum can be used effectively in the treatment of diabetes and hyperlipidemia, although its role in weight control management has yet to be defined.

Diabetes Mellitus and Hyperlipidemia

The plasma cholesterol-lowering properties of guar gum have been known since the 1960s. Most of the numerous clinical trials published since then have shown reductions in plasma concentrations of cholesterol, mainly the low-density lipoprotein (LDL) fraction, and occasionally triacylglycerols, in both healthy and hyperlipidemic individuals, in some studies for periods of up to 1–2 years. Similar results have been reported in people with diabetes, who are also likely to benefit from any improvements in lipid metabolism given their increased risk for CHD. Currently, although some pharmaceutical preparations are available in the UK, for the management of diabetes, no such products are available for the treatment of hyperlipidemia. In addition to evidence showing the blood glucose-lowering effect of guar gum in response to a starchy meal, long-term improvements have been demonstrated in patients with

type 1 and type 2 diabetes. In one study of a group of type 2 diabetic patients, a 7-g daily dose of guar gum, which had been incorporated into wheat bread, improved glycemic control (as seen by a significant reduction in glycated hemoglobin) and lowered plasma total and LDL-cholesterol levels (Table 2).

As discussed previously, some of the pharmaceutical preparations of guar gum, usually taken as a premeal drink, have been shown to be clinically ineffective due to impaired hydration. Moreover, when guar gum flour is incorporated directly into a meal, the clinically effective dose seems to be significantly lower than that recommended for the pharmaceutical preparations. This has led to the development of a number of everyday foods enriched with guar gum, such as bread, pasta, biscuits, and breakfast cereals. The recommended dose for pharmaceutical preparations of guar gum in diabetes therapy is usually 5 g with each meal (i.e., 15 g day^{-1}). However, it is not known what the precise dose and molecular weight of guar gum should be used in the diet to optimize clinical benefits. A lower and upper daily dose range of 6–15 g of native guar gum has been suggested for use in diabetes therapy. This has been suggested on the basis that doses of guar gum above 15 g day^{-1} may cause side-effects, such as flatulence, and there is a lack of reliable long-term studies at doses of less than 6 g day^{-1}. The average molecular weights of guar galactomannan samples used in nutritional studies are in the range of 2.0–3.0 million. However, more recent studies have shown therapeutic benefits with samples of lower molecular weight (0.5–1 million), produced commercially by partial depolymerization. For example, one such product of approximately 1 million molecular weight has been used to lower plasma cholesterol levels in hypercholesterolemic human subjects (Table 3). Further work is required to define the lowest molecular weight and optimum dose that can be used without a loss in clinical efficacy. Such knowledge would have a number of advantages for any future product development, since low-molecular-weight grades are easier to incorporate into foods and are more palatable.

Obesity

Obesity is associated with an increased risk of diabetes, insulin resistance, CHD, and hypertension. Most short-term studies (often single-meal tests) have shown that guar gum reduces feelings of hunger

Table 2 Mean values (± SEM) of fasting plasma glucose, glycated hemoglobin (HbA_{1c}), and plasma lipid concentrations in 16 patients with type 2 (noninsulin-dependent) diabetes after consuming control wheat bread and bread containing guar gum flour for two 6-week periods, respectively. Control and guar bread were taken by each patient in randomized order

	Control period		*Guar bread period*	
	Start (0 weeks)	*End (6 weeks)*	*Start (0 weeks)*	*End (6 weeks)*
Fasting glucose (mmol l^{-1})	9.7 ± 0.8	9.4 ± 0.7	9.7 ± 1.0	9.1 ± 0.8
HbA_{1c} (%)	11.3 ± 0.8	11.2 ± 0.8	11.5 ± 0.8	10.7[a] ± 0.8
Total cholesterol (mmol l^{-1})	5.7 ± 0.3	5.8 ± 0.3	5.9 ± 0.3	5.4[b] ± 0.2
LDL-cholesterol (mmol l^{-1})	3.8 ± 0.3	3.9 ± 0.2	3.9 ± 0.2	3.5[c] ± 0.2
HDL-cholesterol (mmol l^{-1})	1.2 ± 0.1	1.2 ± 0.1	1.2 ± 0.1	1.2 ± 0.1
Triacylglycerols (mmol l^{-1})	1.7 ± 0.3	1.7 ± 0.3	1.8 ± 0.3	1.6 ± 0.2

[a,b,c]Values after 6 weeks significantly different from control at $P < 0.05$, $P < 0.02$ and $P < 0.01$, respectively.
LDL, low-density lipoprotein; HDL, high-density lipoprotein.
Data taken from Peterson DB, Ellis PR, Baylis JM *et al*. (1987) Low dose guar in a novel food product: improved metabolic control in non-insulin-dependent diabetes. *Diabetic Medicine* 4: 111–115.

Table 3 Mean values (± SEM) of plasma lipid concentrations in 11 moderately hypercholesterolemic humans after consuming control wheat bread and bread containing depolymerized guar gum for two 3-week periods, respectively. Control and guar bread were taken by each patient in randomized order

	Control period		*Guar bread period*	
	Start (0 weeks)	*End (3 weeks)*	*Start (0 weeks)*	*End (3 weeks)*
Total cholesterol (mmol l^{-1})	6.37 ± 0.21	6.53 ± 0.16	6.52 ± 0.17	5.89[a] ± 0.8
LDL-cholesterol (mmol l^{-1})	4.18 ± 0.15	4.31 ± 0.14	4.28 ± 0.19	3.81[a] ± 0.12
HDL-cholesterol (mmol l^{-1})	1.26 ± 0.05	1.31 ± 0.06	1.33 ± 0.05	1.23 ± 0.05
Triacylglycerols (mmol l^{-1})	1.84 ± 0.25	1.86 ± 0.32	1.96 ± 0.27	1.85 ± 0.25

[a]Values after 3 weeks significantly different from control and baseline (start) values at $P < 0.001$.
LDL, low-density lipoprotein; HDL, high-density lipoprotein.
Data taken from Blake DE, Hamblett CJ, Frost PG *et al*. (1997) Wheat bread supplemented with depolymerised guar gum reduces plasma cholesterol concentration in hypercholesterolaemic human subjects. *American Journal of Clinical Nutrition* 65: 107–113.

and appetite and increases feelings of satiety and satiation in all subjects, whether they be normal-weight, overweight, or obese. However, there are problems in designing experiments to investigate feeding behavior, since it is very difficult to make the control food (placebo) look and taste like the test food. One study showed that breads containing high concentrations of guar gum were significantly less palatable than the control breads and this may explain why the subjects were less hungry after consuming the guar-containing bread meal. The energy value of the control and test meals should also be identical in such experiments. Despite the positive effects of guar gum in the short term, many of the long-term intervention studies are contradictory and suggest that a therapeutic role for guar gum in weight control management has yet to be established.

Conclusions and Final Comments

There is now overwhelming evidence to indicate that guar gum has therapeutic benefits for a range of metabolic disorders. At doses of guar gum that are likely to be consumed for therapeutic purposes, there is little evidence that it adversely affects human health. On the contrary, the cholesterol-lowering property of guar gum is of benefit to those who are at an increased risk of CHD, including people with hyperlipidemia and diabetes. Furthermore, the use of guar gum in decreasing postprandial glycemia and insulinemia and in improving long-term glycemic control in diabetic patients is now well established. Any improvement in glycemic control is likely to reduce the risk of microvascular complications (e.g., nephropathy), which are common in people with diabetes, and develop over a number of years.

Recent epidemiological studies have also indicated that diets comprising low-glycemic-index foods may have a protective effect in the development of type 2 diabetes. Guar-containing foods could play an important role in such diets and should be tested in the future for their possible prophylactic benefits. In view of possible beneficial effects of guar gum on insulin resistance, which in itself is a risk factor for both CHD and type 2 diabetes, the role of guar foods needs to be evaluated here also. The effects of guar gum on weight reduction in the long term have yet to be demonstrated, although its use in maintaining weight in weight-reduced subjects merits further study. Obese individuals are likely to benefit however from the consumption of guar gum in terms of improvements in glycemic control, insulin sensitivity, and lipid metabolism.

To assist in the development of new low-glycemic-index foods using guar gum as a major ingredient, it would be advantageous to have a detailed understanding of its behavior in the gut environment. Fundamental information about the way in which guar galactomannan influences the rheological behavior of digesta, through its effects in solution, as an entangled network, or on the swelling of food particles, is of paramount importance. This would need to be closely linked to the effects of galactomannan on nutrient digestion and absorption, including, for example, starch and lipids.

The initial attempts to produce palatable guar foods for the management of diabetes and hyperlipidemia were disappointing. However, the technological difficulties were exacerbated by the perceived need to use high doses of guar gum (15–30 g day^{-1}), since the results of early clinical trials suggested that such doses were needed to elicit metabolic benefits. More recent studies have shown that much lower doses (6–12 g day^{-1}) and partially depolymerized guar gum are not only clinically effective, but have significantly less detrimental effects on sensory qualities of food. From a food technology perspective, it is now possible to produce guar-enriched foods that are both clinically effective and palatable.

See also: **Carbohydrates**: Classification and Properties; **Cholesterol**: Factors Determining Blood Cholesterol Levels; **Coronary Heart Disease**: Prevention; **Diabetes Mellitus**: Treatment and Management; **Dietary Fiber**: Properties and Sources; Determination; Physiological Effects; Effects of Fiber on Absorption; **Glucose**: Maintenance of Blood Glucose Level; Glucose Tolerance and the Glycemic (Glycaemic) Index; **Gums**: Properties of Individual Gums; Food Uses; Dietary Importance

Further Reading

Edwards CA and Read NW (1990) Fibre and small intestinal function. In: Leeds AR (ed.) *Dietary Fibre Perspectives*, vol. 2, pp. 52–75. London: John Libbey.

Ellis PR (1999) The effect of fibre on diabetes. In: Hill M (ed.) *The Right Fibre for the Right Disease*, pp. 33–42. London: Royal Society of Medicine Press.

Ellis PR, Rayment P and Wang Q (1996) A physicochemical perspective of plant polysaccharides in relation to glucose absorption, insulin secretion and the enteroinsular axis. *Proceedings of Nutrition Society* 55: 881–898.

Ellis PR, Wang Q, Rayment P, Ren Y and Ross-Murphy SB (2001) Guar gum: agricultural and botanical aspects, physicochemical and nutritional properties, and its use in the development of functional foods. In: Cho SS and Dreher M (eds) *Handbook of Dietary Fiber*, pp. 613–657. New York: Marcel Dekker.

Flourie B (1992) The influence of dietary fibre on carbohydrate digestion and absorption. In: Schweizer TF and Edwards CA (eds) *Dietary Fibre – A Component*

of Food. Nutritional Function in Health and Disease, pp. 295–332. London: Springer-Verlag.

Gallaher DD and Hassel CA (1995) The role of viscosity in the cholesterol-lowering effect of dietary fiber. In: Kritchevsky D and Bonfield C (eds) *Dietary Fiber, Health and Disease*, pp. 106–114. St Paul, USA: Eagan Press.

Leeds AR (1999) The effect of fibre on coronary heart disease and cholesterol. In: Hill M (ed.) *The Right Fibre for the Right Disease*, pp. 33–42. London: Royal Society of Medicine Press.

Low AG (1990) Nutritional regulation of gastric secretion, digestion and emptying. *Nutrition Research Reviews* 3: 229–252.

Maier H, Anderson M, Karl C, Magnuson K and Whistler RL (1993) Guar, locust bean, tara and fenugreek gums. In: Whistler RN and BeMiller JN (eds) *Industrial Gums: Polysaccharides and their Derivatives*, 3rd edn, pp. 181–226. London: Academic Press.

Morgan LM (1992) Insulin secretion and the entero-insular axis. In: Flatt PR (ed.) *Nutrient Regulation of Insulin Secretion*, pp. 1–22. London: Portland Press.

Read NW and Eastwood M (1992) Gastro-intestinal physiology and function. In: Schweizer TF and Edwards CA (eds) *Dietary Fibre – A Component of Food. Nutritional Function in Health and Disease*, pp. 103–117. London: Springer-Verlag.

Reid JSV and Edwards ME (1995) Galactomannans and other cell wall storage polysaccharides in seeds. In: Stephen AM (ed.) *Food Polysaccharides and their Applications*, pp. 155–186. New York: Marcel Dekker.

Ross-Murphy SB (1994) Rheological methods. In: Ross-Murphy SB (ed.) *Physical Techniques for the Study of Food Biopolymers*, pp. 343–392. London: Blackie Academic & Professional.

Todd PA, Benfield P and Goa KL (1990) Guar gum: a review of its pharmacological properties, and use as a dietary adjunct in hypercholesterolaemia. *Drugs* 39: 917–928.

Truswell AS and Beynen AC (1992) Dietary fibre and plasma lipids: potential for prevention and treatment of hyperlipidaemias. In: Schweizer TF and Edwards CA (eds) *Dietary Fibre – A Component of Food. Nutritional Function in Health and Disease*, pp. 295–332. London: Springer-Verlag.

Gut Hormones *See* **Hormones**: Adrenal Hormones; Thyroid Hormones; Gut Hormones; Pancreatic Hormones; Pituitary Hormones; Steroid Hormones

H

Hard Cider *See* **Cider (Cyder; Hard Cider)**: The Product and its Manufacture; Chemistry and Microbiology of Cidermaking

HAZARD ANALYSIS CRITICAL CONTROL POINT

J Slatter, RSSL, Reading, Berks, UK

Introduction

The Hazard Analysis Critical Control Point (HACCP) system is a program of preventative food safety assurance originally developed in the USA by the Pillsbury Company working with the National Aeronautics and Space Administration (NASA) and the US army laboratories at Natick in order to insure the safety of astronauts' food. In the three decades since then, HACCP has been internationally recognized and accepted as a proven method for food safety assurance. Its original purpose was to insure microbiological safety of foodstuffs; it has since been further broadened to include physical and chemical hazards for a wide variety of food and associated industries.

The role of HACCP in food production and quality management systems is described and legal and training aspects are discussed in this article.

Definition of Hazard Analysis and Critical Control Point

HACCP is a systematic approach to the problems and food safety issues associated with all points of food production, from raw material through to end use by the consumer. HACCP aims to identify the source and potential risk of a hazard occurring at these points and put controls or preventative actions in place, so eliminating the problem before it occurs or reducing it to a safe level and more effectively protecting the consumer from harm.

The HACCP Principles have international acceptance and have been published by the Codex Alimentarius Commission and National Advisory Committee on Microbiological Criteria for Foods (NACMCF). The system of Hazard Analysis, of which HACCP is one method, is part of legislation, through the European Community (EC) directive 93/43/EEC which was implemented in the UK in 1995 through the Food Safety (General Food Hygiene) Regulations. Several industry guides exist which help suppliers and manufacturers to apply the legislation and to comply with the requirements of HACCP.

Principles of the HACCP System

Different food processes are vulnerable to different hazards. There is no single hazard analysis which covers all cases. Each HACCP program must be specific to the operation and take into account every factor that might affect food safety. Although details vary, there is a general sequence of events involved in completing a HACCP program (Table 1).

What are the Principles of HACCP?

The HACCP system consists of seven principles which outline how to establish, implement, and maintain a HACCP plan (Table 2). These principles make up the Codex standard (*See* **Legislation**: Codex), which has become the reference for international food safety and identified as the baseline for consumer protection under the Agreement on Sanitary and Phytosanitary Measures agreed at the General Agreement on Tariffs and Trade (GATT) negotiations in 1995.

Table 1 Glossary of HACCP Terminology

Hazard	A biological, chemical or physical agent in, or condition of, food with the potential to cause an adverse health effect
Risk	The probability that a given hazard will occur and cause an adverse health effect
Critical control point (CCP)	A step at which control can be applied and is essential to prevent or eliminate a food safety hazard or reduce it to an acceptable level
Critical limit	A criterion which separates acceptability and unacceptability
Monitor	The act of conducting a planned sequence of observations or measurements of control parameters to assess whether a CCP is under control
Corrective action	Any action to be taken when the results of monitoring at the CCP indicate a loss of control
Control measures	Any action and activity that can be used to prevent or eliminate a food safety hazard or reduce it to an acceptable level. These were previously known as preventive measures

Table 2 The seven principles comprising the Hazard Analysis Critical Control Point (HACCP) system

1 Conduct a hazard analysis
2 Determine the critical control points (CCPs)
3 Establish critical limit(s)
4 Establish a system to monitor control of the CCP
5 Establish corrective action to be taken when monitoring indicates that a particular CCP is not under control
6 Establish procedures for verification to confirm that the HACCP is working effectively
7 Establish documentation concerning all procedures and records appropriate to these principles and their application

Table 3 The 12 stages of a Hazard Analysis Critical Control Point (HACCP) study

1 Assemble the HACCP team
2 Describe the product
3 Identify the intended use
4 Construct a flow diagram
5 Carry out on-site confirmation of flow diagram
6 Carry out hazard analysis
7 Determine critical control points (CCPs)
8 Establish critical limits for each CCP
9 Establish a monitoring system
10 Establish corrective action
11 Establish verification procedures
12 Establish documentation and record keeping

These seven principles can be broken down into 12 different stages, which can logically be used to develop the HACCP Plan.

Stages in a HACCP study (Table 3)

Assemble the HACCP Team

The HACCP study is best carried out by a multidisciplinary team. This usually comprises individuals covering all aspects of production, e.g., manufacturing, buying, quality assurance, technical, and engineering. Occasionally it may be necessary to enlist the help of an expert to cover specialist topics such as microbiology or toxicology. A person experienced in the HACCP technique should be responsible for leading the team and for managing the study.

Describe the Product

A full description of the product should be drawn up, including any relevant safety information such as composition, processing, durability, storage conditions, and method of distribution.

Identify the Intended Use

The intended use should identify the potential consumer. In some cases, at-risk groups may have to be considered, such as the elderly.

Construct a Flow Diagram

The process must be accurately defined from start to finish and this is normally done in the form of a flow diagram. This a simple stepwise sequence of events, which enables the HACCP team to understand the process.

Carry Out On-site Confirmation of Flow Diagram

It is important to check that the flow diagram is accurate by physically checking it against activities and that it includes exceptional items such as breakdowns, rework, and cleaning. The team should also check that the flow diagram is correct for any shift patterns.

Carry out Hazard Analysis

The terms of reference or scope for the HACCP study need to be set. Ultimately all types of hazards (biological, physical, and chemical) will be examined but in the initial stages it may be useful to examine only one category of hazards or even to concentrate on one type of hazard such as glass or *Salmonella*. The HACCP team will have to consider whether to cover the whole of the process or just part of it and if the process is common for a number of products, then these products must be included in the scope.

The next stage of hazard analysis is not easy, but it can be carried out in a structured way.

Hazard analysis consists of an evaluation of all procedures and raw materials to:

- identify potential hazardous materials or pathogens and their severity
- identify potential sources and points of contamination at each stage of the food chain
- determine the potential for microorganisms to survive or multiply in the food
- assess the likelihood of the hazards occurring

A hazard may be biological (e.g., microorganisms), chemical (e.g., pesticides, cleaning chemicals) or physical (e.g., metal, glass, stones). In order to identify correctly the control for the hazard, the source or cause of the hazard must also be examined. For example, glass from a light fitting would be controlled in a different way to glass from a broken bottle on the filling line. In terms of microorganisms, the hazard may be more adequately explained by identifying either the named organism (e.g., *Salmonella*) or group of organisms (e.g., spore formers) and whether they will be present, survive, or grow at the stage.

There are many ways of identifying hazards based on previous knowledge of the product or process, customer complaints, and information from literature or experts. The team must take care to include not just problems which have previously occurred but also new potential hazards, which can be identified by brainstorming.

Obvious sources of hazards are:

- raw materials
- plant and equipment design
- formulation
- processing
- personnel
- packaging
- storage and distribution

Once the potential hazards have been identified, the severity and possibility of the hazard occurring, i.e., the risk, should be assessed. This will identify the significant hazards. The risk assessment is often a judgment by the HACCP team based on knowledge and information; however, some risk assessment schemes have a scoring system or a weighting of high, medium, or low.

When all significant hazards have been identified, the associated controls for each hazard can be assigned. Controls are the factors that are required to prevent, eliminate, or reduce the occurrence of hazards to an acceptable level. These may be controls that are already in place or it may be that new controls will need to be devised. Also, more than one control may be needed for effective reduction of each hazard.

Determine CCPs

For each step (where a hazard exists) CCPs must be evaluated. A CCP is a stage in a process where food safety hazards must be controlled. They are essential to prevent, eliminate, or reduce hazards to an acceptable level. There may be previous stages where the hazard is controlled or reduced and these are sometimes called control points, but the CCP will be the point where, if control is lost, a health hazard will occur.

Understanding the CCPs can help to focus on the efforts and finances on the key areas of food safety and save money by preventing effort in exercising control at a point of less importance.

Information established by the HACCP team may be sufficient to identify a CCP using judgment and prior knowledge; however, to assist in finding where the CCPs may be, a decision tree is a useful tool (**Figure 1**).

Establish Critical Limits for each CCP

Once the CCPs have been identified, the critical or safety limits need to be established. These define the difference between safe and potentially unsafe foods and of acceptable and unacceptable products.

In the work place, target limits may be required to provide a buffer zone to allow the hazard to be brought back under control before it exceeds the critical limit.

Establish a Monitoring System

The means of testing to insure that the control is adequate must be established for each CCP. This assessment may be chemical, physical, sensory or, rarely, microbiological. Criteria usually include measurements of time, temperature, moisture level, water activity, available chlorine or sensory parameters such as visual appearance, flavor or texture.

Monitoring is a scheduled measurement or observation of a CCP. The monitoring procedure will assess whether there is a loss of control from the critical limits, in which case corrective action will be needed.

Establish Corrective Action

Specific corrective actions are required for each CCP. The actions must insure that the consumer is protected and the CCP is brought back under control. It is effective also to assign responsibilities to individual members of staff for the corrective actions in the

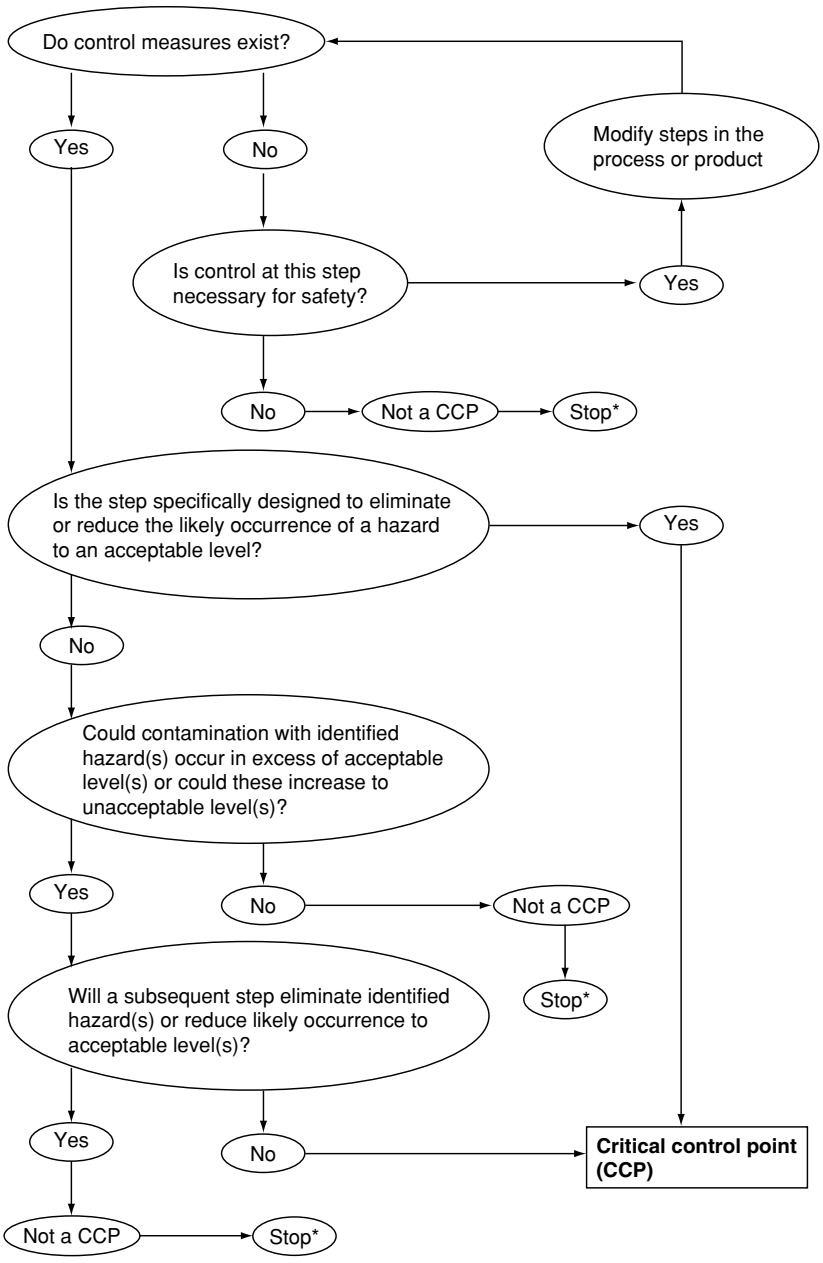

Figure 1 Example of a decision tree for identifying critical control points (CCPs). Reproduced from *Codex Alimentarius Food Hygiene Basic Texts*. (1997).

HACCP plan. All corrective actions should be recorded and reviewed by the HACCP team as part of the verification process.

Establish Verification Procedures

The verification process checks that the HACCP Plan is an accurate record of the activities taking place. The main method of verification is by audit, but other methods may also be used, such as document review and analysis of data.

Establish Documentation and Record Keeping

Documentation and record keeping should be appropriate to the size of the operation. The HACCP will incorporate documents such as the following:

- the HACCP Plan
- hazard analysis
- CCP determinations
- CCP monitoring sheets
- corrective actions
- audit records

- HACCP team meeting minutes
- calibration records

Figure 2 shows a HACCP plan format recommended by Codex Alimentarius. It can be adapted for different operations and size of business and easily facilitates verification.

Implementation

Once developed, the HACCP Plan must be implemented and established in the work place. To be successful the plan must be absorbed into the day-to-day running of the operation with effective training of all personnel, especially those directly involved in the HACCP Plan. The Plan must also stay current and up-to-date if it is to be used as an effective management tool.

Maintenance

Any changes in the product or process which may affect the HACCP system must be identified and the HACCP changed accordingly. This may mean a change in raw material supply, change in process or factory layout, or emergence of further information indicating a health risk associated with the product. Otherwise, the HACCP should be routinely checked annually to ascertain that it is still an accurate reflection of the working environment. This may be done by auditing the HACCP Plan and flow diagrams but can also be carried out by checking process control documents, hygiene audit reports, and customer complaints.

It is also important that the HACCP team remain up-to-date with issues of hygiene and safety affecting other food companies from information in the newspapers, trade press, and membership of trade associations. New technologies may also emerge, providing better controls, and these should be reviewed by the HACCP team.

Prerequisite Programs

The HACCP system will incorporate other existing management systems into its procedures. Typical areas are personal hygiene, good manufacturing practice, supplier quality assurance, and maintenance schedules. These have been given the term 'Prerequisite Programmes' and are normally in place before the HACCP Plan is developed (Figure 3). Prerequisites are systems in their own right and will support the HACCP by taking the control of the general hygiene and good manufacturing practice out of the HACCP Plan.

HACCP and International Quality Systems

Many food companies are using standards such as the ISO 9000 series. There are many similarities between the two systems. Both HACCP and ISO 9000 are concerned with prevention and detection of food safety problems in areas such as calibration, training, document control, control of nonconforming product, and verification through internal audit.

The ISO 9000 standard was updated in March 2001, to ISO 9000:2000 and has guidance notes for the food and drink industries (BS ISO 15161). This includes a requirement for HACCP and provides assistance for incorporating the principles of HACCP into the ISO system.

1. Describe product

2. Diagram process flow

3. HACCP Plan

Step	Hazards(s)	Control measures	CCPs	Critical limits	Monitoring procedures	Corrective actions	Records

4. Verification

Figure 2 Example of a Hazard Analysis Critical Control Point (HACCP) worksheet. CCPs, critical control points.

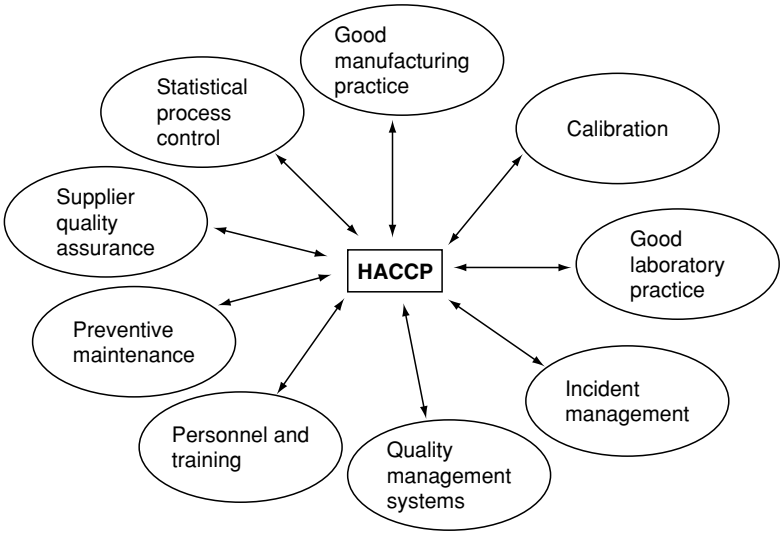

Figure 3 Prerequisites of Hazard Analysis Critical Control Point (HACCP). Reproduced by kind permission of Mortimore and Wallace (1998).

HACCP Training

Training is essential for the effective implementation of HACCP. Staff must understand why the HACCP plan was developed, as well as their role within HACCP, and should be encouraged to suggest any modifications to the HACCP team.

Staff should have an understanding of:

- what hazards are and their importance in food safety
- CCPs and their role in the assurance of product safety
- critical limits which must be met
- corrective actions and responsibilities
- record-keeping requirements
- the objective of verification procedures

It may be necessary to give more specific training to staff involved in CCP monitoring.

Formal training qualifications exist, with examinations introduced by the Royal Society of Health (RSH), the Chartered Institute of Environmental Health Officers (CIEH) and the Royal Institute of Public Health (RIPH). These courses are run by training organizations registered by the above bodies at levels 1, 2 and 3.

A training standard is available from the RIPH steering group to provide a syllabus for training to an advanced level in the principles of HACCP. Training programs have been specified by the World Health Organization to cover the recommended skills and problem encountered in training for HACCP, together with proposed programs for food-processing and manufacturing industries, small food enterprises, and small operations.

See also: **Legislation**: Codex

Further Reading

BS ISO 15161: 2001 *Guidance on Application of ISO 9001: 2000 and for the Food and Drink Industry*. BSI 98/710729.

Codex Alimentarius. Food Hygiene Basic Texts (1997 amended 1999) Rome: Food and Agriculture Organization of the United Nations World Health Organization.

Leaper S (1997) *HACCP: A Practical Guide*. Technical Manual no. 38. Campden & Chorleywood Food Research Association.

Mortimore S and Wallace C (1998) *HACCP. A Practical Approach*, 2nd edn. UK: Aspen.

Mortimore S and Wallace C (2001) *Food Industry Series: HACCP*. Oxford, UK: Blackwell Science.

UK Steering Group on HACCP Training Standards (1999) *HACCP Principles and their Application in Food Safety (Advanced Level) Training Standard*. London: Royal Institute of Public Health.

World Health Organisation (1995) *Training Aspects of the Hazard Analysis Critical Control Point System (HACCP) WHO/FNU/FOS/96.3*. Geneva: Food Safety Unit, Division of Food and Nutrition, World Health Organization.

Health Foods *See* **Functional Foods**

Heart Disease *See* **Coronary Heart Disease**: Etiology and Risk Factor; Antioxidant Status; Intervention Studies; Prevention

HEAT TRANSFER METHODS

J Varley, Imperial College London, London, UK

Introduction

Heat transfer is an essential stage of many food-processing operations, e.g., cooking, sterilizing, freezing, baking, and drying. The fundamental principles of heat transfer and a review of methods for the addition and removal of heat are considered in this article.

Principles of Heat Transfer

There are three fundamental types of heat transfer: conduction, convection, and radiation; these can occur individually or simultaneously.

Heat transfer may be steady-state or unsteady-state. The equations considered here are for steady-state heat transfer and can be solved easily. More complicated solutions arise for unsteady-state heat transfer (i.e., when temperature varies with time), but solutions are available for simple geometries, e.g., slab, cylinder, and sphere.

Conduction

Conduction is the transfer of heat by exchange of molecular energy from hotter to colder regions, under conditions in which there is no turbulent motion of particles. Heat transfer by conduction can take place in a solid, liquid, or gas. Rates of heat transfer by conduction are given by Fourier's law:

$$q = -kA(\mathrm{d}T/\mathrm{d}x) \qquad (1)$$

where q is the heat transfer rate in the positive x direction, k is thermal conductivity, A is the area perpendicular to direction of heat flow, T is temperature and x is distance.

For example, for conduction through a slab the rate of heat transfer is given by the following equation:

$$q = -kA[(T_1 - T_2)/x] \qquad (2)$$

where $(T_1 - T_2)$ is a steady temperature difference maintained across the slab and x is the thickness of the slab (Figure 1).

Thermal conductivity (k) is influenced by temperature, pressure, and composition of the food, including such factors as moisture content and amount of air trapped between cells. Values of thermal conductivity are available for some food materials.

Fourier's law indicates that the rate of heat transfer by conduction is proportional to the driving force for transfer, and inversely proportional to the resistance to heat transfer. For conduction in a slab, the driving force for transfer is the temperature difference, and the resistance to heat transfer is given by x/k.

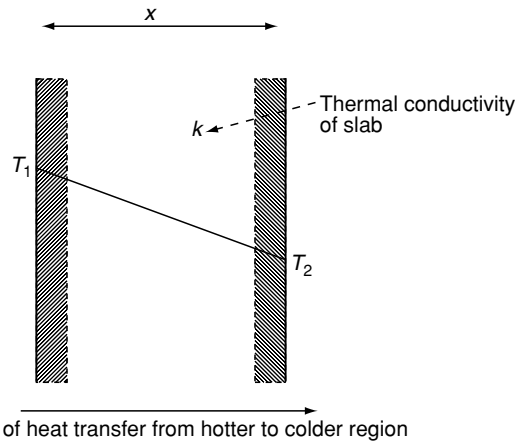

Figure 1 Heat transfer by conduction through a slab. Reproduced from Heat Transfer Methods, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Convection

Convective heat transfer occurs when molecules or groups of molecules in a fluid are mixed. In a solid, molecules are not free to move, and convective heat transfer will therefore only occur in a liquid or gas. Natural convection arises when mixing is a result of density differences caused by temperature gradients. For forced convection, mixing is brought about by mechanical means. Forced convection provides higher rates of heat transfer than natural convection and is therefore more commonly used in the food industry, e.g., in air blast freezers, heat exchangers, agitated retorts, and forced convection ovens. Examples of natural convection include heat transfer in domestic freezers or ovens with natural air circulation (i.e., unaided by fans) and heating below boiling in unstirred, steam-jacketed pans in the food industry.

For heat transfer from a fluid to a surface, rates of convective heat transfer can be calculated from the following equation:

$$q = hA(T_f - T_s) \tag{3}$$

where h is the heat transfer coefficient, T_f is the temperature of the fluid and T_s is the temperature of the surface.

For convective heat transfer between a surface and a fluid, a film (layer of fluid) exists between the bulk fluid and the surface. At the surface, the velocity of the fluid will be zero and there will be minimal mixing of fluid elements. Heat transfer in this film will therefore be by conduction only, and the film will provide the main resistance to heat transfer. The rate of heat transfer can thus be written in terms of the conduction equation:

$$q = (k_f/x_f)A(T_f - T_s) \tag{4}$$

where k_f and x_f are the thermal conductivity and effective thickness of the film, respectively. The film heat transfer coefficient is therefore equal to k_f/x_f for the film, and the resistance to heat transfer is $1/h$.

Typical values of film heat transfer coefficients are available in the literature.

Determination of heat transfer coefficients The heat transfer coefficient can be written in terms of the dimensionless Nusselt number (Nu):

$$(Nu) = (hD)/k \tag{5}$$

where D is a characteristic linear dimension of the heat transfer system, e.g., diameter of a cylinder.

For natural convection, the heat transfer coefficient is found to depend on the temperature difference, acceleration due to gravity, and the physical properties of the food (e.g., density, viscosity, thermal conductivity, and specific heat). The dependence of heat transfer coefficient on these factors can be represented by the following equation:

$$(Nu) = a(Pr)^b(Gr)^c \tag{6}$$

where a, b, and c are constants. Pr and Gr are the Prandtl and Grashof number, respectively:

$$Pr = (c_p\mu)/k \tag{7}$$

$$Gr = (D^3\rho^2 g\beta\Delta T)/\mu^2 \tag{8}$$

where c_p is specific heat at constant pressure, μ is viscosity, D is the linear dimension of the system, ρ is density, g is acceleration due to gravity, and β is the coefficient of thermal expansion. Values for the constants, determined experimentally for some foods at certain conditions, are available in the literature.

For forced convection, gravity and the coefficient of thermal expansion are much less significant than for natural convection. Therefore the heat transfer coefficient does not depend to any great extent on the Grashof number. Of more importance is the influence of forced circulation velocities, which can be represented by the Reynolds number (Re). For forced convection the heat transfer coefficient can, in general, be evaluated from the following dimensionless equation:

$$(Nu) = a(Re)^b(Pr)^c \tag{9}$$

As for natural convection, values for the constants (a, b, and c) in the above equation have been determined for a range of experimental conditions.

Composite Heat Transfer

The equations given above for heat transfer by conduction and convection show that the rate of heat transfer is given by the ratio of the driving force for heat transfer (temperature gradient) and the resistance to heat transfer offered by the material. The resistance to heat transfer for conduction and convection is given by x/k and $1/h$, respectively.

If heat is transferred through a composite material, the heat transfer resistances behave in a similar manner to electrical resistances in series. Consider, for example, heat transfer to a food material from the surroundings, first through a surface (surface 1), then through several composite layers (layers 1 and 2) and, finally, through a further surface film (surface 2). The overall resistance to heat transfer would be given by the following equation:

$$(1/U) = (1/h_{s1}) + (x_1/k_1) + (x_2/k_2) + (1/h_{s2}) \tag{10}$$

where $1/U$ is the overall resistance to heat transfer, U is the overall heat transfer coefficient, h_{s1} and h_{s2} are the surface heat transfer coefficients for surface 1 and 2, x_1 and x_2 are the thicknesses of layers 1 and 2, and k_1 and k_2 are the thermal conductivities of layers 1 and 2.

The overall rate of heat transfer can then be written as follows:

$$q = UA\Delta T \tag{11}$$

where ΔT is the overall temperature difference.

Radiation

Radiation is the transfer of heat by electromagnetic waves. Radiant heat transfer is often found in the food industry, especially in baking and drying operations. Heat transfer by radiation depends on temperature, surface area, and surface properties of the body. Rates of heat emission by radiation can be determined from the Stefan–Boltzmann law:

$$q = A\sigma T^4 \tag{12}$$

where A is the surface area, σ is the Stefan–Boltzmann constant (5.67×10^{-8} W m^{-2} K^{-4}) and T is the absolute temperature of the body.

The Stefan–Boltzmann law determines the maximum amount of radiation that would be emitted at a particular temperature by a perfect radiator (a black body). However, most surfaces do not emit this maximum amount of energy, but many do emit an amount of energy which is directly proportional to the maximum value. The proportionality constant is known as the emissivity of the body (ε) and has a value between 0 and 1. Bodies with emissivities between 0 and 1 are known as gray bodies. For a gray body, the heat transfer equation for radiation is as follows:

$$q = \varepsilon A\sigma T^4 \tag{13}$$

Bodies not only emit radiation but also absorb and reflect radiation. Absorption can be evaluated from the Stefan–Boltzmann law for black bodies, and from the modified equation for gray bodies, with the emissivity replaced by the absorptivity (α). For gray bodies, it can be shown that $\varepsilon = \alpha$. Any radiation which is received by a body but is not absorbed must be reflected. For a gray body, the reflectivity is equal to $(1 - \alpha)$ and the amount of heat reflected can be evaluated by replacing the emissivity in the above equation with the reflectivity.

Radiative heat transfer can take place between bodies and also between a body and its surroundings. As an example, radiative heat exchange between a body and its black surroundings would be given by the following equation:

$$q = A_1\varepsilon_1\sigma(T_1^4 - T_2^4) \tag{14}$$

where A_1 is the surface area of body 1, ε_1 is the emissivity of body 1, T_1 is the temperature of body 1, and T_2 is the temperature of the surroundings.

If the temperature difference $(T_1 - T_2)$ is small, the above equation can be written as follows:

$$q = A_1\varepsilon_1\sigma T_1^3\Delta T \tag{15}$$

The above equation shows that the rate of heat transfer for radiation is directly proportional to the temperature difference and inversely proportional to the resistance to heat transfer (as is the case for convection and conduction). The resistance to heat transfer is given by $1/(A_1\varepsilon_1\sigma {T_1}^3)$. If convection and radiation are occurring simultaneously, this radiative heat transfer resistance can be used to calculate an overall heat transfer resistance (as described above for simultaneous convection and conduction).

Heat Transfer Rates in Heat Exchangers

There are a wide variety of heat exchangers in use in the food industry. These include direct steam injection, plate heat exchangers, and swept-surface heat exchangers. The variety in design of heat exchanger available reflects the wide range of physical properties of food requiring heat treatment. Most heat exchangers can be used for both heating and cooling. Heat transfer rates will, in general, be higher for heating than for cooling, because the viscosity of a liquid film will be higher during cooling than for heating. (*See* **Heat Treatment**: Ultra-high Temperature (UHT) Treatments.)

The surface area of a heat exchanger required to provide a certain heat transfer rate can be determined from the equation for the overall rate of heat transfer (eqn 11), if the overall heat transfer coefficient and the relevant value of the temperature difference are known. In general, two fluids are flowing on opposite sides of the walls of a heat exchanger, either cocurrently or countercurrently (Figure 2). The relevant temperature difference for heat transfer calculations in this case is the log mean temperature difference (ΔT_{lm}), which is defined as follows:

$$\Delta T_{lm} = \frac{(T_1 - T_2) - (T_1' - T_2')}{\ln(T_1 - T_2)/(T_1' - T_2')} \tag{16}$$

where T_1 and T_2 are the temperatures of fluids 1 and 2 at one end of the heat exchanger and T_1' and T_2' are the temperatures of fluids 1 and 2 at the other end of the heat exchanger.

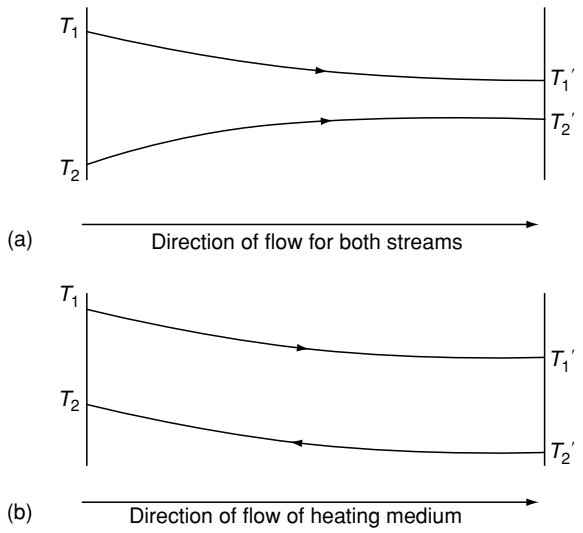

Figure 2 Temperature differences for heat transfer in heat exchangers: (a) co-current flow; (b) countercurrent flow. Reproduced from Heat Transfer Methods, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Addition and Removal of Heat

For addition or removal of heat to a food material, an energy source is required. Sources of energy for bulk generation of heat include gas, oil, solid fuels, electricity (microwave or dielectric heating), or infrared energy. These energy sources may be used directly or indirectly.

Direct methods of heating involve direct contact between the food material and the source of heat energy. The food material therefore comes directly into contact with the products of combustion of the fuel. Direct methods may be used for the addition of heat in food processing, e.g., baking ovens, but they are rarely used for the removal of heat.

In indirect methods, the food material is separated by a barrier from the primary source of heat energy and does not come into contact with the products of combustion. Heat is generally transferred to the food via a heat exchanger. The primary energy source may be used as the heating medium, or it may be used to heat a secondary medium, which will then be passed through the heat exchanger. The heat transfer medium may be a vapor or gas (e.g., steam or air), a liquid (e.g., water), or electricity (as for resistance heating systems). Heating of the secondary heat transfer medium is usually accomplished by means of a boiler, in which the source of fuel is burnt to generate heat. Once the heat transfer medium has been heated, it is pumped to a heat exchanger where heat transfer takes place between this fluid and the food material. In some systems, the secondary heat transfer medium is passed directly over the food, e.g., steam blanching.

Removal of Heat

Removal of heat is commonly referred to as refrigeration or chilling. Removal of heat can be accomplished by mechanical or nonmechanical refrigeration. Mechanical refrigeration is dealt with elsewhere. Typical methods of nonmechanical refrigeration include heat removal using ice, chilled brine, or cryogens. (*See* **Freezing**: Principles; Cryogenic Freezing.)

Heat removal using ice Ice, at 0 °C, is capable of removing large quantities of heat (the heat capacity of ice is 68 kJ kg^{-1}, K^{-1} at 0 °C) and has therefore been used as a means of heat removal for many years. In recent years, the use of ice for heat removal has declined as mechanical refrigeration has grown in popularity. Ice is still used for some chilling applications, particularly when refrigeration is only required for a short period of time (e.g., marine fishing operations) or as a secondary refrigeration system which may be required to deal with peak loads. There are several ways in which ice can be employed as a heat transfer medium. These include the following:

1. Packing crushed ice around the food material.
2. Circulating air over ice blocks before passing it over the food product.
3. Circulating water over ice blocks prior to flow of water through a heat exchanger.

The freezing temperature of ice can be reduced (and hence the potential for heat removal increased) by the addition of salts. However, corrosion can become a problem when ice and salt mixtures are used.

Heat transfer using chilled brine Brine, a solution of salt and water, can be used to remove heat from food materials, e.g., on fishing vessels. As with ice, mechanical refrigeration has largely superseded the use of brine in refrigeration. Sodium chloride and calcium chloride are the most widely used brines for heat removal. The freezing point, and hence capacity to remove heat, depends on the concentration of the brine solution. For typical concentrations used for refrigeration, the minimum freezing point of sodium chloride and calcium chloride solutions are −15 °C and −51 °C, respectively. Problems with corrosion will arise when employing brine solutions, if precautions are not taken. Calcium chloride is much more corrosive than sodium chloride. Appropriate precautions to reduce corrosion include elimination of air, neutralization to pH 7 or 8 with caustic soda or hydrochloric acid, and addition of corrosion retardants, e.g., sodium bichromate. A typical system which

uses brine to remove heat would consist of the following steps:

1. The brine is cooled to the required temperature in direct expansion coils.
2. The brine is then passed through a heat exchange system which is in contact with the food.
3. The brine is returned to the expansion system where heat is removed.

Other media that have been used for chilling include glycol–water mixtures and methylated spirits.

Heat removal using cryogens In cryogenic systems, heat is absorbed by the cryogen as it changes phase. Cryogens commonly used for removing heat include solid and liquid carbon dioxide, and liquid nitrogen. Halogenated hydrocarbon refrigerant gases are also used to a limited extent.

Addition of Heat

Heat can be added by direct or indirect methods (in direct methods the food material comes into direct contact with the energy source, whereas for indirect methods the food material and energy source are separated by a barrier). (*See* **Drying**: Theory of Air-drying.)

Heat transfer media used in indirect heat addition to food materials include vapors, gases, and liquids. Only the most commonly used heat transfer media will be considered in this article.

Heat transfer using air Air is used to a limited extent as a heat transfer medium in indirect heating systems, e.g., in some baking processes, fluidized-bed cooking, and hot-air driers. Air has the advantage of being odorless and nontoxic, although it also has low specific heat and thermal conductivity, which are disadvantageous for efficient heat transfer.

Heat transfer using steam Saturated steam is a widely used heat transfer medium in food processing. Maximum temperatures and pressures of typical steam supplies generated for heat transfer in food processing are 200 °C and 1.7 MPa respectively. Steam is used in a wide variety of heating processes which include steam blanching, pasteurization, heat processing of food in containers, extrusion cooking, and evaporation.

Saturated steam has several physical properties that make it attractive for use as a heat transfer medium, including the following:

1. High heat capacity.
2. Nontoxicity.
3. Stability.
4. No odor.
5. Low cost.

Adverse properties of steam for use as a heat transfer medium include the following:

1. High vapor pressure.
2. Low critical point.
3. Corrosion, formation of scale, or foaming (during steam generation) may occur as a result of impurities being present in the raw material, i.e., the water supply.

Heat transfer using hot water and other liquid media Hot water is used as a heat transfer medium up to temperatures of 200 °C in a wide variety of food processes, including blanching, pasteurization, and in-container sterilization. Advantages of hot water as a heat transfer medium include the following:

1. High specific heat and thermal conductivity.
2. Stability.
3. No odor.
4. Simplicity and ease of control of associated equipment.
5. Abundance and low cost.
6. Limited corrosion, scale formation, and foaming, as compared to steam systems.

Limitations include, for example, the possibility that if hot water comes into direct contact with the food material there may be substantial loss of water-soluble nutrients (vitamins, salts, and minerals).

Other liquids that are used as heat transfer media include mineral oil, chlorinated diphenyls, organosilicate, *o*-dichlorobenzene, and diphenyl or diphenyloxide. All these liquids have the disadvantage of giving off odors, but with the exception of diphenyl and diphenyloxide they have the advantage of possessing low vapor pressures, which allows design of high-temperature, low-pressure systems.

See also: **Drying**: Theory of Air-drying; **Freezing**: Principles; Cryogenic Freezing; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments

Further Reading

Cengel YA (1998) *Heat Transfer: A Practical Approach*. New York: McGraw-Hill.

Cleland AC (1989) *Food Refrigeration Processes – Analysis, Design and Simulation*. Essex: Elsevier Applied Science.

Green DW (ed.) (1984) *Perry's Chemical Engineering Handbook*, 6th edn. New York: McGraw-Hill.

Hayes GD (1987) *Food Engineering Data Handbook*. New York: Longman Scientific and Technical.
Incropera FP and De Witt DP (1981) *Fundamentals of Heat Transfer*. New York: John Wiley.
Loncin M and Merson RL (1979) *Food Engineering – Principles and Selected Applications*. London: Academic Press.
Mohsenin NN (1980) *Thermal Properties of Foods and Agricultural Materials*. New York: Gordon and Breach.
Munton R and Stott JR (1978) *Refrigeration at Sea*, 2nd edn. Essex: Elsevier Applied Science.
Winterton RHS (1997) *Heat Transfer*. New York: Oxford Science Publications.

HEAT TREATMENT

Contents

Ultra-high Temperature (UHT) Treatments

K R Swartzel, North Carolina State University, Raleigh, USA

Introduction

Commercially sterile foods are processed and packaged in a manner that leaves the food free of microorganisms of public health significance, and inhibits the growth of any microorganism under normal nonrefrigerated storage and distribution. Process methods include both thermal and nonthermal. These are accomplished by either: (1) heating the product in a hermetically sealed container so that the package and contents are sterilized at the same time; or (2) separately sterilizing the food and the container and then aseptically filling and sealing the sterile product in the sterile container.

The former sterilization process relates to conventional in-can or in-bottle processes, while the latter refers to ultrahigh-temperature (UHT) processing. By heating product during continuous flow, product temperatures of 130–150 °C are achieved. Due to the relatively high sensitivity of vegetative bacterial cells and spores to these high temperatures, less chemical change occurs than with conventional canning to achieve the same sterilizing effect. This results in increased retention of nutrients, color, flavor, and texture. Time needed for UHT processes is short, in the order of seconds. By understanding product constituent temperature-dependent changes, equipment performance and process design, the desired balance between desirable effects (inactivation of biological materials, i.e., enzymes, microorganisms, and their spores), and undesirable effects (loss of product quality, i.e., flavor, color, and flow properties) can be achieved.

This section discusses thermal aseptic process equipment, processing with and without particulates, temperature measurement, and control considerations. Product quality, microbiological considerations, and aseptic packaging are discussed in separate sections.

Equipment

Equipment required for continuous thermal sterilization of foods includes the following: product supply pump, heat exchanger for preheating product, metering pump, product sterilizer, holding section, heat exchangers for cooling product, and aseptic optional surge tanks (**Figure 1**). Components which may be incorporated into a continuous sterilization system include orifice plate or backpressure valve for maintaining the pressure necessary to prevent the product from boiling; homogenizer; and steam seals on pumps, valves, connection or access ports; and a monitoring and control system to insure the proper thermal treatment.

The product supply pump and preheat exchanger can be conventional food-handling equipment since product has not been sterilized when it moves through these components.

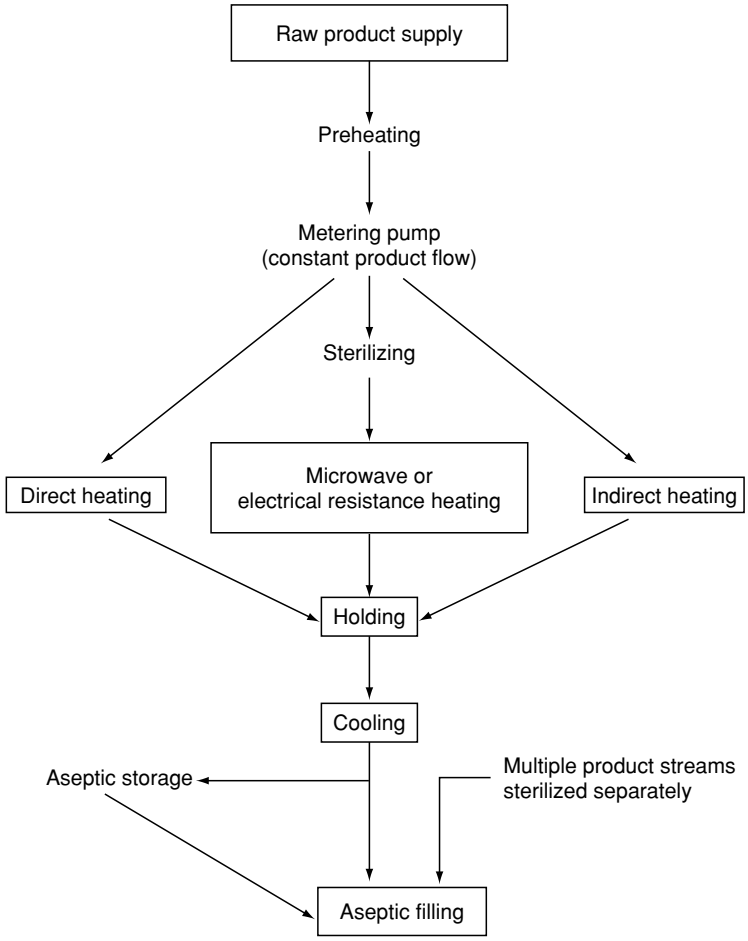

Figure 1 Flow chart of continuous-flow aseptic processing.

Metering Pump

The metering pump determines the flow rate of product through the holding tube. A positive displacement pump is employed to maintain a constant flow rate that is relatively independent of the load on the pump. Rotary pumps are normally specified for pressures up to 68.94 kPa. For higher pressures, plunger or piston pumps are used. The metering pump is located upstream from the holding section. (The holding section will be described in detail later.) If the pump has a speed control device for changing the flow rate, some mechanism is employed to insure that the pump does not operate at a speed higher than that needed to give product the required thermal treatment time in the holding section. A lock on the speed control is the most satisfactory method of preventing unauthorized changes which would reduce residence time.

Direct and Indirect Heaters

The sterilizer is a heat exchanger used to heat product to the sterilization temperature. There are two basic types of commercial sterilizers available: direct and indirect units. Sterilizers may be categorized as direct contact heaters (steam injectors and steam infusers) and indirect heaters (tubular heaters, plate heaters, and scraped surface heaters).

Steam or hot water is normally the source of heat for continuous thermal sterilization. Direct contact heating involves mixing steam and product together. As steam condenses, its latent heat is transferred to the product. When no wall separates the steam and product, the heat transfer rate is high and product is heated rapidly. Product is diluted by the steam condensate and vacuum equipment is generally used for partial cooling and removal of excess moisture. Steam is injected into a continuous stream of product in the steam injector. The steam infuser introduces product with a large surface area into a steam atmosphere under sufficient pressure to achieve the desired sterilization temperature. Steam for direct contact heating must be produced from potable water in a culinary steam generator or in a boiler using accepted boiler treatment compounds.

In the injector steam enters the product in pulses, and excessive vibration, pressure fluctuations, and temperature variation may result. To stabilize operation, orifices may be installed in the product entrance and discharge ports of the injector to isolate the injection chamber. The orifices are approximately equal in size and designed to produce at least a 68.94 kPa pressure drop across the injector during normal operation.

In indirect heating, product is heated by transferring heat from water or steam through a heat exchange wall. The wall may be a tube, a plate, or a scraped surface with each having its own time–temperature profile. Tubular and plate heaters used as sterilizers have relatively small flow channels for product. This gives a large surface area to product volume ratio for rapid heating. The high temperatures used for sterilization require heat exchangers to be built to handle the associated high pressures required. Plate and tube designs incorporate varying machining (plate patterns, tube dembling, tube helical patterns) in an effort to promote mixing, resulting in more uniform product thermal treatment.

Tubular and plate sterilizers lend themselves to regenerative heating and cooling as an energy-saving feature. Regenerative heating uses the heat of the hot product to heat cold incoming product. Longer times are required to heat product to the sterilization temperature than in direct contact heaters, and deposit of product on hot heat exchange surfaces sometimes severely limits the operating time and makes cleaning more difficult. (*See* **Cleaning Procedures in the Factory**: Overall Approach.)

Scrape surface heat exchangers (SSHEs) provide a moving blade inside an elongated circular chamber which scrapes the outer heated surface. SSHEs are used with highly viscous products or products with particulates.

In the direct heating systems (steam injection or infusion), the heating and cooling at higher temperatures take place very rapidly compared with indirect systems (tubes, plates, SSHE). In the direct systems, virtually all of the thermal effects on product occur in the holding section after the product has reached the scheduled process temperature. In the indirect system, substantial thermal effects can occur during heating and some additional effects may occur during cooling. For public safety reasons, however, usually only the heating effects in the holding tube are considered.

Holding Section

The holding section is a pipeline section located immediately downstream from the sterilizer. It is designed to insure that all product is held at, or above, the sterilization temperature for the time equal to, or greater than the scheduled process time. The tube is sloped upward toward the discharge end. The upward slope insures that vapor cannot be trapped. A vapor pocket in the holding tube would permit product to move through the tube faster than if the holding tube contained product only.

For public safety (destruction of pathogenic organisms, mainly *Clostridium botulinum*), regulatory specification calls for a holding tube to be sized to hold the fastest-moving particle in laminar flow for the minimum specified time. (*See* ***Clostridium***: Botulism.)

Heat Exchangers for Cooling Product

Product is cooled either by evaporative cooling in a vacuum chamber or in indirect heat exchangers similar to those used for product heating and sterilization. A vacuum chamber is often used when the sterilizer is a steam injector or infuser. In addition to cooling product, the vacuum chamber removes moisture. When vacuum chamber pressure (or temperature) is controlled properly, the same amount of moisture added during heating by the condensing steam can be removed. The operating pressure or temperature is a function of product flow rate and properties, steam properties and heat losses from the sterilizer, holding tube, and vacuum chamber. The correct pressure (or temperature) must be established for each system.

Since the vacuum chamber is normally controlled at pressures below atmospheric pressure, aseptic seals are used on all ports for pipeline connections, sensors, sight glasses, and manholes. Steam seals are commonly employed.

The heat exchangers used for cooling, whether plate, tubular, SSHE, or vacuum chamber, are designed to withstand temperatures and pressures required for equipment sterilization before product flow is initiated. (*See* **Sterilization of Foods**.)

Aseptic Surge Tank

The function of the aseptic surge tank is to act as a reservoir for product between processing and packaging. This may or may not be a part of all systems. Operation of the tank is as critical as any other part of the aseptic system.

Sterilization of the surge tank is independent of the rest of the system. Four modes are utilized: steam and vent, sterilizing, cooling, and tank sterile. The steam and vent cycle allows the previously cleaned and sanitized tank to be vented of any gases and drained of any liquids or condensates during heat-up. This cycle continues until predetermined sterilizing conditions (temperature and pressure) are met. The sterilization cycle maintains this temperature and pressure for the

sterilization time. The tank is then cooled, but kept pressurized with filtered sterile air or inert gas (i.e., nitrogen). The air pressure also provides a means to move the sterile product to the filler without the use of pumps.

Processing Considerations

Thermal Evaluation

Techniques for evaluating thermal systems have been the subject of elaborate developments in process calculation procedures. Continuous-flow thermal processing offers new challenges to the traditional batch (retorting) process.

Backpressure Requirements

For process temperatures above the boiling point of the liquid, backpressure devices must be used. The backpressure system assures proper pressure to achieve and maintain the desired temperature. For direct steam heating, condensation of the steam may not be complete. Lack of complete condensation inside the injector would cause temperature variations in the holding tube that could lead to some fluid particles being processed below the desired process thermal treatment.

For steam injection units, the product and the steam flows must be isolated from pressure fluctuations inside the injection chamber. Supplementary orifices on the product inlet and the heated product outlet is one method of isolation. The orifices are approximately equal in size and yield a minimum product pressure drop of 68.94 kPa across the injector during normal operations.

For holding tubes in all systems, the product pressure must be of sufficient magnitude to keep heated product in the liquid phase. If this pressure is too low, the resultant vaporization in the holding tube will substantially reduce residence times. A minimum product pressure in the holding tube of 68.94 kPa for operating temperatures from 88 °C through 100 °C is satisfactory. For temperatures above 100 °C, a minimum pressure of 68.94 kPa above the product saturation value pressure is needed to keep the heated product in the liquid phase.

No noncondensable gases can be tolerated in the holding section. Any two-phase flow caused by the noncondensable gases would displace the product in the holding tube, resulting in reduced residence times.

Holding Tube Calculations for Direct Systems

As with the indirect systems, holding tube lengths for direct systems are calculated as twice the length required to hold the average measured flow. However, with the steam injection process, the holding time is adjusted because the product volume increases as the steam condenses to water during heating in the injector. This surplus water is evaporated as the product is cooled in the vacuum chamber. With as much as 60–70 °C increase by steam injection, which is probably the maximum temperature rise that will be used, a volume increase of 12% would occur in the holding section. For holding time calculations, the volumes per unit time are adjusted accordingly.

Consideration for Processing Fluids with Particles

Special equipment is usually incorporated for the processing of fluids containing particles. As with all UHT systems, the required minimum heat treatment for the system must be achieved for all parts of each particle. This means that for systems with particles, the spot receiving the least thermal treatment (the 'cold spot') is generally in the center of the largest particles with the lowest heat of conduction traveling with a minimum of residence time (fastest particle). To achieve this, the carrier fluid is often overprocessed for the desired end result. For many products, this is not a problem and SSHEs are the equipment choice. For heat-sensitive carrier fluids, systems have been developed that keep the particles and final carrier fluid separate during heating. In these systems the solids are processed in a processing carrier fluid. The end carrier fluid is heat-treated in a conventional manner. After the particles have been properly processed the processing fluid is drained off and the end carrier fluid is then added. This also provides for minimum mechanical damage. In other systems, other means of heating are incorporated for rapid, uniform heating, such as electrical resistance heating, microwave heating, and/or radiofrequency heating. With these systems, due to the unique form of heating with conventional cooling, the surface of the particles may define the cold spot.

Numerous special pieces of equipment have been developed for moving the two-phase material and controlling the flow. One such item relates to reciprocating piston pumps capable of pumping 2–3-cm-diameter particulates. Moving chamber devices have been developed to allow for different holding times for different-size particles.

Product Deposition

During the heating of foods, product may deposit on the heat exchange surface. The deposit fouls the heat exchange surface and is undesirable for several reasons. First, the deposit acts as an extra layer of resistance to heat transfer. Product temperature is

raised enough to compensate for the increased resistance. In several cases, the deposit can effectively reduce the cross-sectional area available to flow. Thus, not only is product temperature affected, but so is residence time. Second, flakes of fouled material can break off from the wall of the heat exchanger and be carried and deposited elsewhere by the product flow (i.e., homogenizing values, product packages, etc.). Finally, the burnt material can give off undesirable flavors to the product.

Direct systems are not as susceptible to fouling, though fouled material may be deposited in the holding tube, and sometimes in the vacuum chamber.

Temperature Sensing and Utilities

Temperature sensors are installed at appropriate locations in the product flow stream for controlling and/or recording temperatures to insure sterilization of all contact surfaces. During equipment sterilization, hot water under pressure is circulated through the processing system and interconnecting piping with the aseptic filler.

The sterilization temperature is monitored at the coldest point in the system that must be sterilized. This point is usually in the return line from the filler.

The proper temperature in the holding section is very important in achieving the commercial sterility of a product. It is important that this temperature be closely controlled, monitored, and recorded.

Steam requirements for aseptic systems vary with the type of system being used. Steam supply must be of the proper pressure and flow rate to satisfy process needs. Product heating and steam seals make up the two major uses of steam in UHT processing. Steam that comes in contact with product must be culinary (food-grade).

Steam seals provide a safety barrier at fittings, valves, and other locations on the sterile side of the system. They protect against microbiological contamination from the environment. The seal assembly provides a steam channel between two gaskets, allowing live steam to enter and fill the groove, thus forming a microbial barrier. These seals are located on the potential problem areas of UHT systems. Usually they are found in low-pressure areas (i.e., vacuum chambers) and around moving system components.

Compressed air is used in pneumatic systems such as controllers, recorders, valves, and sensors. Manufacturers of processing and pneumatic systems specify compressed-air requirements for their equipment. Because processing plants have existing air supply systems, it must be determined whether sufficient air volume and pressure are available. If not, corrective procedures should be implemented. Pneumatic systems are very susceptible to entrained particles of moisture, oil, or dirt. Proper filtering devices should, therefore, be installed so specifications for particulates can be adhered to. A pneumatic system jammed by dirt or oil can cause expensive and frustrating shutdown of equipment.

Compressed sterile air or inert gas is also used by manufacturers to maintain positive pressure in aseptic tanks. This positive pressure serves two purposes. First, it acts as a barrier to contamination. Should there be a crack or failure in the tank or valves, the positive pressure would blow sterile air out of the tank, tending to prevent entrance of microorganisms. Second, the pressure is used to push the product from the aseptic tank to the filler.

Since the air is in contact with sterile product, it must also be sterile. Air is sterilized by using filters and/or heaters. Air, to be sterilized, must itself be of good quality so removal of moisture, oil, and particulates is essential. The filters used have filter pores of a size smaller than microorganisms, thus preventing their passage through the filter. In addition, an air-sterilizing assembly consisting of a heater, holding tube, and cooler can be used to sterilize the air.

Safeguards must be developed to insure that product is rejected if all conditions for sterile-aseptic processing are not achieved. These conditions include maintenance of scheduled process time and temperature, correct functioning of steam seal and/or sterile air pressurization systems, or sufficient product pressure downstream from the sterilizer to eliminate any possibility of sterile product contamination. All important and necessary aspects of UHT processes must be documented, with proper records maintained to demonstrate safe operation.

The following sections will examine kinetic aspects of product processing, including quality retention, microbial destruction, and aseptic packaging.

See also: **Canning**: Principles; **Cleaning Procedures in the Factory**: Overall Approach; ***Clostridium***: Botulism; **Packaging**: Packaging of Solids; Aseptic Filling; **Pasteurization**: Principles; **Quality Assurance and Quality Control**; **Sterilization of Foods**

Further Reading

Burton H (1988) *Ultra-High-Temperature Processing of Milk and Milk Products*. New York: Elsevier Applied Science.

Kyereme M, Swartzel KR and Farkas BE (1999) New line intersection procedure for the equivalent point method of thermal evaluation. *Journal of Food Science* 6414: 565–570.

Mermelstein NH (ed.) (1997) Continuous multiphase aseptic processing of foods. *Food Technology* 51: 43–62.

Nelson PE, Chambers JV and Rodriquez JH (eds) (1987) *Principles of Aseptic Processing and Packaging*. Washington, DC: Food Processors Institute.
Reuter H (1989) *Aseptic Packaging of Food*. Lancaster: Techomic.

Chemical and Microbiological Changes

C James and S James, University of Bristol, Langford, North Somerset, UK

Introduction

Temperature controls the rate of all physicochemical reactions and consequently has a profound effect on all biological systems. It influences such fundamental properties as hydrogen bonding, solubility of solute molecules, viscosity, density, and osmotic properties of cell membranes.

The prime purpose of most heat treatments is to make the food safe to eat by reducing the numbers of pathogenic organisms present in the food, if not completely eliminating them. Many food sterilization and pasteurization processes have been developed to achieve this aim without unacceptable changes to the structure or flavor of the food. However, in many cooking processes the aim is to cause structural and chemical changes that will make the food more palatable.

Every operation has, however, side-effects which adversely influence product properties. This is because during every heat treatment many types of process take place simultaneously, i.e., mechanical, physical, chemical, biochemical, and microbial.

Chemical

The adverse chemical effects of heat treatment may be classified as:

- Enzymic and nonenzymic browning reactions
- Vitamin destruction
- Destruction of amino acids
- Oxidation of lipids, flavor components, etc.
- Hydrolysis of carbohydrates, etc.
- Destruction of pigments, such as chlorophyl, carotenoids, and flavonoids
- Development of undesirable flavors by reaction products
- Destruction of carbohydrates and proteins
- Retrogradation of starch
- Cross-linking reactions of polymers, including proteins and carbohydrates, which adversely affect the water uptake of the product.

Changes in Proteins

Heating causes dramatic changes in most foods. The heating of meat is accompanied by changes in appearance, flavor, texture, and nutritive value. The most drastic changes in meat during heating, such as shrinkage and hardening of the tissue, release of juice, and color changes, are caused by changes in the muscle proteins. It is difficult to investigate heat-induced changes in the complicated and highly organized protein structure found in meat. Therefore, many of the experiments have been carried out by heating isolated proteins. These studies show that the most drastic changes of the myofibrillar proteins occur between 30 and 50 °C, reaching almost completion at 60 °C. These changes are characterized by an unfolding of the protein molecules, accompanied by an association of molecules and resulting in coagulation and loss of enzyme activity. At temperatures above 70 °C more severe changes such as oxidation of sulfhydryl groups occur.

Home cooking and commercial canning procedures have little effect on the nutritive value of food proteins. However, other industrial processes that involve the use of excessive heat at low moisture levels, such as drying, and subsequent storage of these products may result in severe nutritional damage. Milk is the only important naturally occurring protein food that has a high content of reducing sugar. Liquid milk appears virtually unaffected by industrial pasteurization and only slightly affected by autoclaving and canning. However, the protein in dried milk can deteriorate during manufacture and storage.

Maillard reactions occur during the baking of bread and biscuits and during the production of breakfast cereals. A 10–15% loss of lysine can occur during the baking of bread with the loss occurring mainly in the crust. Biscuits are an attractive product to enrich the protein content of diets, but are subjected to high temperatures and baked to low moisture contents. Losses of up to 50% reactive lysine can occur during the production of biscuits enriched with 15% skimmed milk powder.

Lysine is the most sensitive amino acid to Maillard damage but all amino acids can be reduced in availability when the heat process is increased.

Changes in Carbohydrates

Heat affects the properties of different carbohydrates in foods to different extents, and not always in the

same direction. The effect can best be considered under four areas – taste, caramelization, pyrolysis, and interaction with other food components. These naturally are not independent but are related to a greater or less extent, depending on chemical structure.

It is important when considering the effect of temperature on taste to distinguish between relative and absolute sweetness. The former is usually determined with reference to sucrose, and the relative sweetness of most sugars decreases with increasing temperature. However, the absolute sweetness of sugars usually increases with increasing temperature.

Caramelization or scorching of carbohydrates is a result of the partial thermal breakdown of carbohydrates. The brown product produced, caramel, consists of dehydration products and colored substances, the proportion of which depend on the extent of heating.

After polymerization, depolymerization, and dehydration reactions, secondary thermal reactions cause carbon–carbon bond cleavage and many pyrolysis products of low molecular weight are produced in a similar way from both mono- and polysaccharides.

The most important reaction of carbohydrates with other food components is the so-called nonenzymic browning or Maillard reaction. This occurs between reducing sugars and nitrogenous compounds, in particular amino acids and proteins. The reaction is highly temperature-dependent and may increase two to three times for each 10 °C rise in temperature in model systems and even more in real foods.

Changes in Vitamins

Three vitamins, vitamin E in its ester form, riboflavin, and niacinamide, are extremely heat-resistant, whereas thiamin is the most labile of the vitamins. Some vitamins show a gradual degradation as the temperature is raised from 100 to 130 °C (vitamins A, D, B_{12}, and C). Other vitamins show a sudden drop above a critical temperature – folic acid at 110 °C, and pyridoxine and pantothenic acid at 120 °C.

Vitamin A and carotene lose activity when heated in the presence of oxygen, but are fairly stable to ordinary cooking. Insignificant losses are found in milk during pasteurization, sterilization, and spray-drying.

Carotene is well retained during the dehydration of vegetables if enzymes have been inactivated by blanching. Synthetic β-carotene has good stability when used as a coloring agent in canned soups. Retention in cakes baked at 175–200 °C was 90–95%, whereas in white bread it was only 50–80%.

Frying of oils and fats may lead to a considerable loss of carotene and vitamin A. Vitamin A appears to be retained as long as water is present but after 40 min at 130–160 °C or 20 min 175–200 °C about 90% had been destroyed. It has been reported that a 40% loss of carotene occurs during the frying of margarine for 5 min with the loss increasing to 60% after 10 min.

Less information is available on the thermal stability of vitamin D in foods but it is though to be similar to vitamin A. Its retention during smoking of fish, spray-drying of eggs, and pasteurization and sterilization of milk is satisfactory.

Vitamin E in food products occurs in the form of free tocopherols. These survive domestic cooking processes far more than commercial processing. Losses of up to 95% have been measured in canned vegetables (Figure 1). However, oils and cereals are a more important source of vitamin E than vegetables. Almost 100% of the free tocopherols were destroyed when vegetable oil was heated for 3 h at 200 °C, whilst 47% of tocopherols in unbleached flour were destroyed during the baking of bread.

High losses occur in water-soluble vitamins during slow heating, long cooking, and slow cooling processes. Much of the loss is due to leaching of soluble matter into the cooking water. When processing of vegetables prior to freezing, steam blanching followed by air-cooling is recommended.

Blanching of fruits and vegetables inactivates the ascorbic acid oxidase and stabilizes vitamin C. Under the most favorable conditions, blanching losses can be less than 10%; however, under severe conditions they can exceed 50%. Similar differences in vitamin C content can result in domestic cooking (Table 1).

The thermal degradation of thiamin vitamin (B_1) results in numerous reaction products from the cleavage of the vitamin molecule to its pyrimidine and thiazole moieties. Vitamin B_1 losses in high-temperature, short-time (HTST) canning processes are far less than in conventional canning. Losses in vegetables are in the 13–16% range compared with

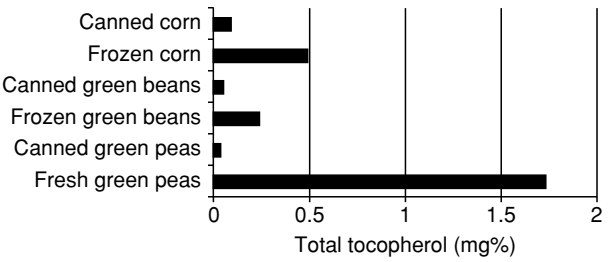

Figure 1 Total tocopherol content after processing of vegetables.

Table 1 Typical vitamin C losses during cooking

Method	Vitamin C (%)		
	Destroyed	Extracted	Retained
Green vegetables	–	–	–
Boiling (long time, much water)	10–15	45–60	25–45
Boiling (short time, little water)	10–15	15–30	55–75
Steaming	30–40	< 10	60–70
Pressure cooking	20–40	< 10	60–80
Root vegetables			
Boiling	10–20	15–25	55–75
Steaming	30–50	< 10	50–70
Pressure cooking	45–55	< 10	45–55

over 40% in conventional systems. With meat, losses are under 10%, compared with approximately 20% conventionally.

Less severe heat treatments such as sous-vide processing have been shown to increase retention of vitamin B_6 and folate acid. After 5 min of boiling, vitamin B_6 and folate losses were 55 and 39% respectively. Steaming for the same time reduced the losses to 17 and 12% and sous vide to 3 and 3% respectively.

Overall, vitamins C, B_1, B_6, and folic acid are the most vulnerable vitamins in thermal processing. However, major losses of vitamins during thermal processing can be taken as a sign of poor treatment methods.

Changes in Enzymes

The purpose of most pasteurization processes is to kil microorganisms and inactivate deteriorative enzymes. Often the inactivation of certain enzymes is used as a processing index to verify that the food has been subjected to the desired heat treatment.

Different heat treatments have different effects on enzymes and within processes, process temperature is critical (Table 2).

Microbiological

The use of heat to inactivate vegetative cells in pasteurization processes and to inactivate both vegetative cells and endospores in sterilization processes is the basis of many food production systems. The aim of thermal processing is to deliver sufficient heat to a food to reduce the chance of survival of an organism that will grow in the food to an acceptably low level. The rational for deciding what is an acceptably low level has developed over many years since initial studies on the thermal inaction of *Clostridium botulinum* in the 1920s.

Table 2 Effect of different heat treatments on enzymes

Process	Influence on enzymes
Sterilization	Extremely thermostable enzymes survive
Pasteurization	Thermostable enzymes survive
Blanching	Quality-changing enzymes inactivated
Cooking/frying/baking	Most enzymes inactivated
Drying	Dependent on process temperature
Extraction/concentration/ distillation	Dependent on process temperature

Heat Activation Kinetics

It is generally observed that the number of viable organisms in a uniformly heated population of a single strain of microorganisms decreases exponentially with time, and over many orders of magnitude, with a slope usually designated as D in minutes. It is also generally observed that the rate of inactivation increases exponentially with rise in temperature, with a slope usually designated z in degrees.

Much careful experimentation has shown that there are clear departures from these simple log: linear relationships in practice. The two most common types of deviation are shown in Figure 2. When large inactivations have been achieved and survivor curves followed down to very small surviving fractions, 'tails' are commonly reported. Rising counts or 'shoulders' during the initial heating periods are also common. A number of reasons for these deviations have been put forward. They include assumptions that they are artifacts caused by counting colony-forming units rather than actual survivors. Alternatively, heterogeneity of resistance within the population or changes in resistance during heating have been suggested to cause deviations from the log:linear relationship.

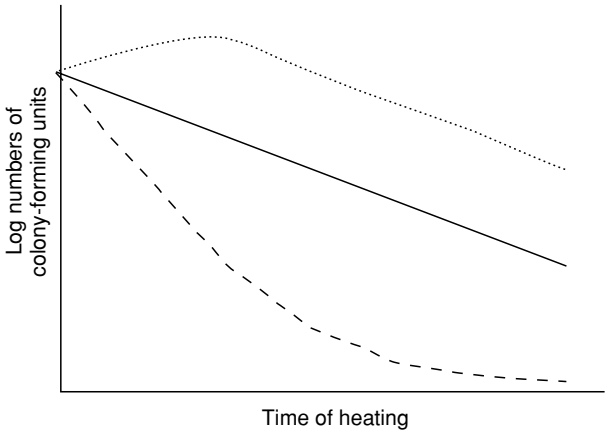

Figure 2 Commonly observed deviations from classical heat-inactivation kinetics. Continuous line, log:linear; dashed line, tail; dotted line, shoulder.

Deviations from the log:linear relationship are more often reported and occur to a greater degree in vegetative cells than in endospores. This is particularly prevalent during lower-temperature, longer-time treatments during which vegetative cells continue to metabolize. This may result in changes in composition and structure and hence thermal tolerance during the heating process.

In vegetative cells the shoulders and tails can be greatly influenced by the heating menstruum and the conditions of growth prior to heating.

Mechanism of Heat Inactivation and Injury

DNA, RNA and ribosomes, cytoplasmic membranes, and specific enzymes are the primary sites for lethal and nonlethal injury of microorganisms.

Studies on spores as well as vegetative cells have been inactivated by heating and have shown single strand breaks in DNA. There is a substantial difference between dry and moist heating. Higher temperatures are needed in dry heating but very high levels of mutants in spore populations are generated. Because of this, heat-injured spores may require additional nutrients for recovery and the production of colony-forming units. It is thought that recovery, in particular with respect to DNA repair, plays an important role in determining the overall resistance of bacterial spores to thermal stress.

Mild heating accelerates the degradation of ribosomal RNA, but this often precedes loss of viability and is not therefore thought to be a prime cause of cell death.

Much early work showed that after heating vegetative cells generally became leaky, losing ions, amino acids, and low-molecular-weight nucleic compounds. The rapid loss of such materials from heated cells has long suggested that heating interferes with the permeability of cytoplasmic membranes. Procedures that minimize leakage have been shown to reduce the level of injury. However, correlations between cell death and leakage are poor. Therefore, membrane damage as a major site of the lethal effects of heat is unlikely.

After some heat treatments, it has been observed that a fall in the rate of germination is preceded by a reduction in colony count. Since DNA is not involved in the initiation of germination this would support the view that damage to DNA is a prime cause of heat inaction. Conflictingly, there are also data that show that in some cases some part of the germination system itself may be especially heat-sensitive and the key site of heat inactivation. The strong inference from such data is that heat inactivates the spores by destroying the activity of an enzyme(s) on the germination pathway. Unfortunately, since the mechanism of germination is not completely understood, the identity of the heat-sensitive step(s) is not known.

Overall, heat is assumed to cause inactivation by destroying a key lethal target in the cell. This is most often directly or indirectly the functionality of the DNA. At the same time, the heat is also destroying many other targets that may be present in larger numbers in the cell. Destroying these targets is therefore less critical unless their numbers are reduced to very low levels or unless the cell becomes further stressed. A heat-treated cell may therefore have a reduced probability of growth after treatment.

Practical Processing

The prime purpose of cooking or other heat treatment is to reduce, if not completely eliminate, any pathogenic bacteria present in the food. To quote from the Chilled Food Association guidelines: 'The heat process must be validated and critical parameters must be controlled and recorded. These must take account of the worse conditions likely to occur with respect to heat transfer (for instance, the use of frozen raw materials or the use of large pieces of meat). All parts of the product must receive a minimum heat process appropriate to the target organism to be controlled and for the targeted shelf life.'

From a pathogenic survival view, heat treatments are divided into three types:

Table 3 Time required to produce a 10^6 reduction in *Listeria monocytogenes* and *Clostridium botulinum* type B at different temperatures

Temperature	*Time (min) to produce 10^6 reduction*	
	Listeria monocytogenes	Clostridium botulinum *type B*
60	43.5	
62	23.3	
64	12.7	
66	6.8	
68	3.7	
70	2.0	
72	1.0	
74	0.6	
76	0.3	
78	0.2	
80	0.09	270.3
82	0.05	138.9
84	0.03	71.9
86		37.0
88		19.2
90		10.0
92		6.3
94		4.0
96		2.5
98		1.6
100		1.0

Table 4 Thermal inactivation rates (*D*-value) of *Listeria monocytogenes* in meats determined using different media for enumeration of surviving *L. monocytogenes*

Type of meat	*Temperature (°c)*	D-*value (min)*	
		Tryptose agar	*LPM agar*
Ground-beef roast	54.4	22.4 ± 1.1	19.3 ± 4.3
	57.2	15.7 ± 4.1	17.6 ± 9.8
	60.0	4.47 ± 1.60	3.5 ± 1.13
	62.8	2.56 ± 1.04	2.82 ± 1.29
Naturally	54.4	12.5 ± 6.5	6.84 ± 1.79
contaminated	57.2	3.41 ± 1.42	1.97 ± 0.80
ground beef	60.0	1.62 ± 1.10	1.11 ± 0.44
	62.8	0.73 ± 0.32	0.47 ± 0.35

1. If the food temperature is less than 70 °C for 2 min then there is a possibility that all pathogens present will survive.
2. In foods heated to 70 °C for 2 min (or equivalent) then all pathogens present will be reduced to an acceptable level (6 log reduction), e.g., *Listeria monocytogenes, Staphylococcus aureus*, salmonella, verocytotoxigenic *Escherichia coli*. However, spores and preformed toxins may persist.
3. In foods heated to 90 °C for 10 min (or equivalent) in addition to vegetative pathogens, spores of psychotrophic *Clostridium botulinum* will be reduced to an acceptable level (6 log reduction). However, more heat-resistant spores, e.g., strains of *Bacillus cereus* and some preformed toxins, may persist.

The equivalent times at different temperatures are shown in Table 3. However, care must be taken when using these equivalent values. Values for temperatures above 70 °C for *Listeria* have been extrapolated from the measured data. The *Clostridium* values have been calculated using a *z*-value of 7 °C below 90 °C and 10 °C above 90 °C.

Rates of thermal inactivation of a five-strain mixture of *L. monocytogenes* were determined in ground-beef roast (Table 4). *D*-values for *L. monocytogenes* Scott A in naturally ground beef were also determined. Large differences in *D*-values were found when using different media for enumeration of the surviving bacteria.

Comparing the results of this study with those obtained for salmonella at similar temperatures in ground-beef roast it was noted that *L. monocytogenes* is about four times more heat-tolerant than salmonella (Table 5).

Although *Salmonella* spp. in general are not as heat-resistant as *Listeria monocytogenes*, one species, *S. senftenberg* 775W (Table 6), has been found to be more heat-resistant than *L. monocytogenes*.

Concern about the manufacture of 'rare' roast beef in the USA (a process requiring relatively low temperatures, usually below 60 °C) initiated studies into the effect of low temperatures on the survival of salmonella in meat systems. Results (Figure 3) showed that adequate processing of meats at temperatures below 60 °C required substantially longer processing times than at temperatures greater than 60 °C.

Campylobacter is the most significant source of food poisoning in the UK, with the number of cases now exceeding those caused by salmonella. In the USA it was found that *Campylobacter jejuni* was present in 38% of pigs, 24% of sheep, and 2% of the beef carcasses examined. Studies carried out in New Zealand found that *C. jejuni* was commonly present in the feces of unweaned calves and from two of four groups of sheep. Poultry is also a major reservoir of *Campylobacter* species. *Campylobacter* species have a low heat resistance, with a *D*-value of 2.12 min at 55 °C and 0.79 min at 57 °C, and should be destroyed by adequate cooking.

Table 5 Comparison of heat tolerance of salmonella with that of *Listeria monocytogenes* in ground-beef roast

Temperature (°c)	D-*value (min)*	
	Salmonella	L. monocytogenes
57.2	3.8–4.2	17.6
62.8	0.6–0.7	2.6–2.8

Table 6 Comparison of heat tolerances (*D*-values) of *Salmonella typhimurium* and *S. senftenberg* 775 W

Temperature (°C)	D-*values (min)*	
	S. typhimurium	S. senftenberg 775W
60		6.3
62.8	0.11	
71.7	0.003	0.09

Reproduced from Shapton DA and Shapton NF (eds) (1991) *Principles and Practices for the Safe Processing of Foods*. Oxford: Butterworth–Heinemann, with permission.

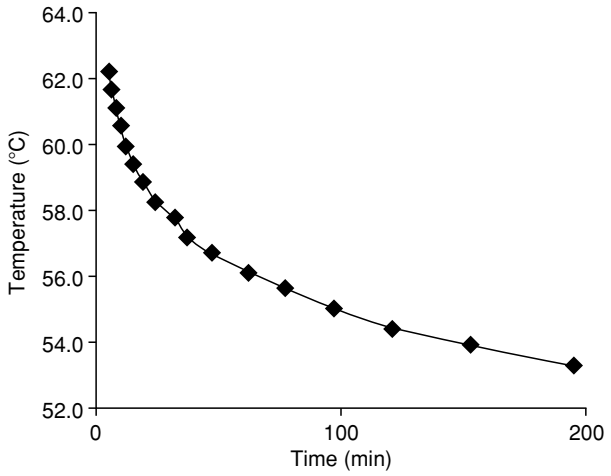

Figure 3 Thermal death curve for a 7-*D* kill of salmonella in cooked beef.

Table 7 US Food and Drug Administration temperature–time recommendations for cooking ground-beef patties to control *Escherichia coli* O157:H7

Temperature (°C)	*Cooking time*
60.0	8 min 20 s
62.2	2 min 7 s
62.6	32 s
68.3	8 s

E. coli is a common organism found in the gut of animals and humans. *E. coli* is mostly harmless in humans, but enterohemorrhagic *E. coli* and, in particular, *E. coli* serotype O157:H7, is recognized as being of significance to human health. Many food-poisoning outbreaks and a number of deaths have been associated with its survival after poor cooking. The US Food and Drug Administration temperature and time recommendations for cooking ground-beef patties to control *E. coli* O157:H7 are shown in Table 7.

Overall

Most heat treatments for food have been established to increase the safe high-quality storage life of the product. The severity of the treatment should be sufficient to reduce, if not completely, eliminate pathogenic organisms likely to be present in the food of interest.

Cooking is a heat treatment with a dual role. Cooking should render the food safe by reducing pathogen levels. At the same time, it facilitates chemical changes that improve the flavor and texture of the food.

Chemical changes in proteins, carbohydrates, vitamins, and enzymes occur in all heat processes. In many cases these changes are undesirable and careful control of the heating regime is required to minimize thermal changes.

See also: **Canning**: Principles; Cans and their Manufacture; Quality Changes During Canning; **Carotenoids**: Occurrence, Properties, and Determination; **Cooking**: Domestic Techniques; **Enzymes**: Functions and Characteristics; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; Electrical Process Heating; ***Listeria***: Properties and Occurrence; **Pasteurization**: Principles; Pasteurization of Liquid Products; Pasteurization of Viscous and Particulate Products; Other Pasteurization Processes; **Protein**: Heat Treatment for Food Proteins; ***Salmonella***: Properties and Occurrence

Further Reading

Anonymous (1997) *Guidelines for Good Hygienic Practice in the Manufacture of Chilled Foods*, 3rd edn. London: Chilled Food Association.

Gaze JE, Brown GD, Gaskell DE and Banks JG (1989) Heat resistance of *Listeria monocytogenes* in homogenates of chicken, beef steak and carrot. *Food Microbiology* 6: 251–259.

Gould GW (ed.) (1989) *Mechanism of Action of Food Preservation Procedures*. Barking: Elsevier Science.

Høyem T and Kvåle O (eds) (1977) *Physical, Chemical and Biological Changes in Food Caused by Thermal Processing*. London: Applied Science.

James C and James SJ (1997) *Meat Decontamination – The State of the Art*. University of Bristol: EU concerted action programme.

McMeekin TA, Olley JN, Ross R and Ratkowsky DA (eds) (1993) *Predictive Microbiology – Theory and Application*. Taunton: Research Studies Press.

Shapton DA and Shapton NF (eds) (1991) *Principles and Practices for the Safe Processing of Foods*. Oxford: Butterworth Heinemann.

Electrical Process Heating

P J Fryer, University of Birmingham, Birmingham, Edgbaston, UK

Introduction

In electrical process ('ohmic') heating of foods, heat is generated by the passage of electrical current through a food. Heat is generated within the food; this overcomes the limitations of slow thermal conduction

within solids, and makes it possible to process solids and liquids at similar rates. This article reviews the principles that govern electrical heating, the factors which control the heating rate, and the design of processes.

Electrical Heating

Thermal processing aims to reduce contamination so that the food will not cause a health hazard during its shelf-life. Each part of the material must be processed to a level prescribed to insure product safety. Heat transfer to solids is slow, however; by the time enough heat is conducted to the centre of a solid to insure sterility, much of the rest is overcooked.

Several sets of reactions occur when a food is heated: those which reduce the level of bacterial contamination, those which result in losses in product quality, in terms of nutrition, taste, and texture, and some that improve taste and texture. The reactions which reduce the number of bacterial spores have a higher activation energy than those which lower product quality: to maximize quality for a given level of sterility it is best to process at as high a temperature as possible. Continuous ultrahigh-temperature (UHT) or high-temperature, short-time (HTST) processes exploit this to produce food of a higher quality than canning. In these processes, food is rapidly heated to *c*. 140 °C, held there for a short period and then rapidly cooled. At 140 °C, food can be sterilized in a few seconds, rather than the several minutes needed at canning temperatures. This requires rapid heating rates to minimize time spent at high temperatures, and thus quality loss, and is best done in a continuous process.

Processing single-phase liquid foods this way is straightforward, because they can be heated and cooled rapidly. High heating and cooling rates ($>1\,^{\circ}\mathrm{C\,s^{-1}}$) are possible. This type of process can efficiently cook liquids such as milks, fruit juices, soups, and sauces, but cannot readily be applied to foods which contain particles, because of the slowness of thermal conduction into the solid. Conduction heating requires a temperature driving force between particles and liquid, and for particles larger than about 2 mm, rapid heating is not feasible. It is possible to use conduction/convection methods to process foods with a high solids fraction (for example, in scraped-surface exchangers), but low heating rates and long hold times are needed, giving poor product quality.

The need to conduct heat is the limiting factor in the thermal processing of particles, and thus for solid–liquid mixtures. Volumetric heat generation techniques can solve this problem. Various techniques are available which use electric fields. In microwave or RF (radio frequency) heating, a high-frequency electric field excites the water molecules within the material, whilst in electrical resistance heating (ohmic heating), however, the passage of electrical current results in heating throughout the material. The process is more energy-efficient than microwave heating, because nearly all of the electrical energy goes to heat the food. It requires the passage of electric current through the material; electrodes that make good contact with the food are required, unlike microwave heating which needs no physical contact. When an electric current flows through a material, heat is generated according to the familiar Ohm's law:

$$W = I^2 R \qquad (1)$$

Here W is power, I is current, and R is the resistance. Heat is generated throughout the material as a result of its inherent electrical resistance.

If the electrical conductivities and thermal capacities of different components of a food are similar, they can heat at similar rates. In practice, as a result of their thermal capacities and electrical conductivities, solids can heat faster than liquids, a result which is impossible by conventional means. Figure 1 shows that rapid heating rates are possible for real foods. Short process times thus make it possible to apply UHT techniques to particulate foods.

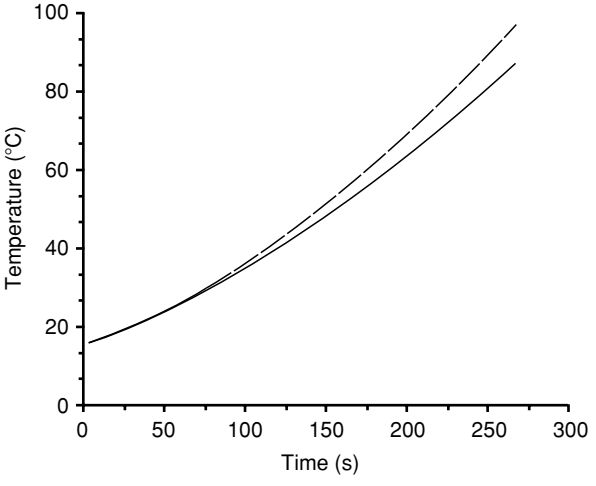

Figure 1 Electrical heating of a piece of lamb meat 20 × 20 × 15 mm, electrical conductivity 0.45 S m^{-1} at 15 °C, in a saline solution, conductivity 0.454 S m^{-1} at 15 °C. Applied voltage gradient 11 V cm^{-1}. Dashed line, solid; continuous line, liquid. Reproduced from Heat Treatment, Electrical Process Heating, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Electrical Heating Processes

The advantages of electrical heating are such that many attempts have been made to commercialize it, in areas such as baking, thawing, and as alternatives to conventional heating processes. An electrical pasteurization process was successful in the USA in the 1930s, and applications have been found in areas such as fish processing. For a process to be commercially acceptable, a number of process and safety criteria must be satisfied:

- electrical design to avoid electrolysis and product contamination;
- effective control of food heating and flow rates;
- efficient aseptic packaging techniques for a two-phase mixture;
- overall cost-effectiveness.

A number of early processes did not meet these criteria, for example, by using direct current supply (which leads to electrolysis), or by using packages containing expensive electrode material. The absence of efficient packing plant also limited success.

The resurgence of interest in the process in the last 15 years is due to the development of a commercial system by APV Baker, which incorporates solutions to the problems described above. The APV Baker ohmic heater was originally developed by EA Technology at Capenhurst, UK. In 1984, APV Baker secured a license for the heating system and then developed a commercial process. Commercial-scale systems are available with power outputs of 75 and 300 kW, corresponding to product capacities of *c.* 750 and 3000 $kg\,h^{-1}$. A schematic process flowsheet is given in Figure 2. The system incorporates heat, hold, and cool sections. Food passes from a product pump into a vertical or near-vertical pipe containing a series of electrodes between which current flows. Sufficient pressure is maintained to insure that the material does not boil; this can be up to 4 bar for sterilization at 140 °C.

The process was originally designed to sterilize foods of high solids fraction, up to about 60%, at heating rates in the region of 1 $°C\,s^{-1}$. Large particles, up to 25 mm in diameter, can be processed, and the technique has found commercial application in the UK, the USA, Europe, and Japan. It has been developed for a number of different applications:

- aseptic processing of high-added-value ready-prepared meals for storage and distribution at ambient temperatures;

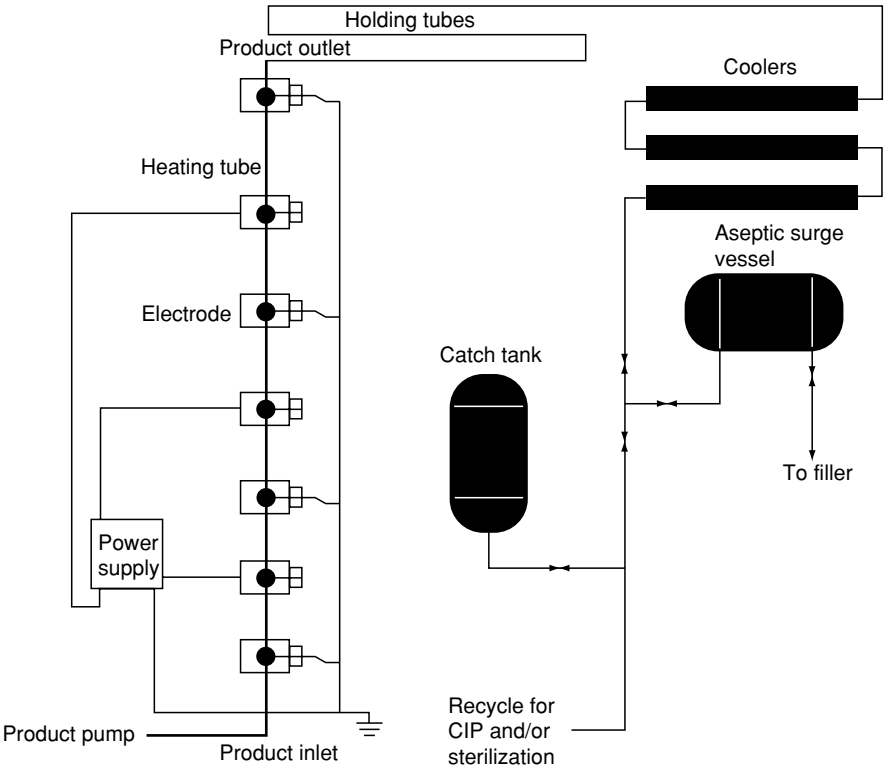

Figure 2 Schematic process flowsheet for the APV Baker ohmic heater. CIP, cleaning in place. Adapted from: Parrott DL (1992) Use of ohmic heating for aseptic processing of particles. *Food Technology* 46: 68–72 with permission.

- pasteurization of particulate food products, for example fruit pieces;
- preheating of food products prior to in-can sterilization;
- the hygienic production of high-added-value ready-prepared meals for storage and distribution at chilled temperatures.

Rapid pasteurization of particles gives very high product quality and has proved very successful. Electrical heating involves short process times: different parts of a meal may require different pretreatments. For example, to produce tender meat and crisp vegetables in the same formulation it will probably be necessary to precook the meat. It is important to prove that an ohmic heater carries out the required treatment. Understanding the principles of electrical heating is necessary both for process design and for proving that it is acceptable to regulatory bodies.

Factors that Control Electrical Heating

Ohmic and conventional processes differ significantly. The process is also significantly different from microwave heating, which has a finite penetration depth of energy into any food; electrical heating has no such limitation. Ohmic heating offers the possibility of rapidly heating solid–liquid mixtures on a commercial scale. The process does however lead to problems for formulation designers and food engineers. Some of these are conceptual; many of the situations possible in ohmic heating are not generally encountered in food processing.

To design a food formulation to exploit the advantages of electrical heating requires understanding of the process and the factors which affect product sterility and quality. The process is controlled by the rate of heat generation, which is governed by a number of factors, the most important of which is the electrical conductivity of the food material. However, temperature is also affected by the way food flows through the heater, and thus the residence time within the plant. To design a food and process requires the factors which affects the heating rate of the material, and the time that it spends in the heater, to be understood.

The heating rate can be found from eqn 1; a more useful form of Ohm's law considers voltage gradients, where the heating rate is given as:

$$Q = \kappa E^2 = \kappa(\text{grad V} \cdot \text{grad V}) \qquad (2)$$

where E is the voltage gradient (V m^{-1}) and κ the electrical conductivity (S m^{-1}). The voltage field within a system in which the electrical conductivity varies with position is found by solving Laplace's equation:

$$\nabla(\kappa \nabla V) = 0 \qquad (3)$$

throughout the material with appropriate boundary conditions, such as constant voltage on the electrodes. The distribution of electrical conductivities thus controls the voltage distribution, and critically affects the heating rates of the different phases. Local voltage gradients may well be different to global ones because of local conductivity changes.

The temperature field that results from heating can be found by solving the equation for thermal transport; for example, in a solid:

$$\frac{\partial T}{\partial t} = \frac{Q}{\rho c_p} + \frac{\lambda \nabla^2 T}{\rho c_p} \qquad (4)$$

where ρ is the density, c_p the specific heat, and λ the thermal conductivity of the material. Eqns 2–4 are coupled through the temperature dependence of the physical properties, particularly electrical conductivity.

Figure 3 demonstrates the interaction between particle shape and alignment of the electric field using experimental results; the same particle will underheat the fluid if placed parallel to the electric field or overheat if placed at right angles to the field. This arises because the solid is less conductive than the liquid: current is diverted past the particle when it is parallel to the field, but has to pass through the solid when it is at right angles to the field. The reverse is seen when the particle is more conductive than the surrounding liquid. Particles with a smaller aspect ratio, such as cubes, will have less heating rate variation with orientation, but some effect will always be seen around sharp edges.

Electrical Conductivity of Foods

In electrical heating it is obviously necessary to know the overall electrical conductivity of the mixture. This is key in determining the power consumption and mean heating rate of the process. It is also necessary to know the local conductivity of the different components of the food. Foods generally have electrical conductivities in the region of $0.01–10\ \text{S m}^{-1}$; these often increase linearly with temperature, with about 2% increase per degree increase. However, where there is a physical or chemical change in the food, complex changes in conductivity can occur, such as when there is a phase change (such as melting of fats) or a change in the structure of the food (such as the breakdown of cell walls).

For a commercial situation, electrical conductivity needs to be measured experimentally for all the components of a mixture. Care should be taken

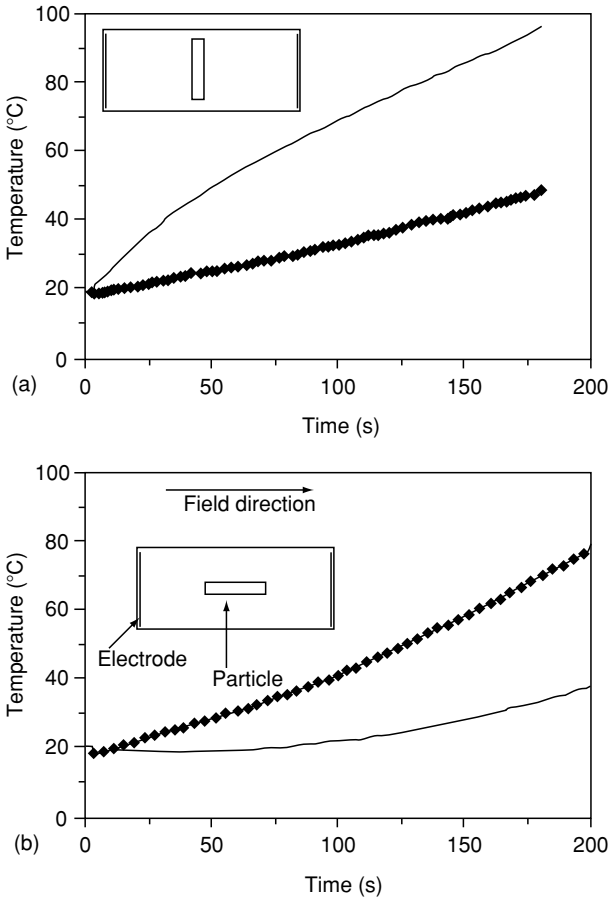

Figure 3 The electrical heating of potato particles, 3 × 4 × 0.75 cm, as a function of orientation to the applied field. Overheating and underheating of a particle (circles), in liquid (continuous line), in different orientations to the electric field (shown in insets). Reproduced from Heat Treatment, Electrical Process Heating, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

here; electrical conductivity can be a strong function of the frequency of the supply, and can be strongly anisotropic, because of the structure of food. For example, the electrical conductivity of carrot at 25 °C was found to be 0.025 S m^{-1} along the axis of the carrot, and 0.02 S m^{-1} perpendicular to the axis. Table 1 shows typical electrical conductivities for foods; these are guides only. It is possible to change the conductivity to some extent, for example by salt infusion, but the result may not be acceptable to consumers. Insulators (such as glass or plastic) and conductors (such as metals) have electrical conductivities several orders of magnitude different to those of foods. In each case, these inclusions will not heat at a significant rate in an electric field; metal has too low a resistance to heat, whilst no current flows through insulators to give rise to heating. Care must be taken to avoid the presence of these inclusions to eliminate cold spots in the system.

Table 1 Electrical conductivity data for typical foods

Product type	*Conductivity at 25 °C (S m^{-1})*
Pickles and chutnies	2.0–3.0
Savory sauces	1.6–1.8
Soups	1.4–1.8
Meats	0.8–1.2
Full-cream milk	0.52
Desserts; vanilla, rice, custard	0.38–0.5
Beaten egg	0.4
Vegetable pieces	0.06–0.1
Fruit pieces	0.05–0.15
Margarine	0.027 (too low for successful heating)

Heating of Solid–Liquid Mixtures

The equations show that the distribution of temperature in a food mixture depends on the distribution of electrical conductivity. This depends on the electrical conductivities of the individual phases, the shape of the particles, and their orientation to the field. Figure 3 has shown this effect for materials where the electrical conductivity of the two phases is significantly different. In practice, if the electrical conductivities of the two phases are within about 10%, then the temperatures will be reasonably uniform.

Eqn 2 can only be solved analytically in very simplistic situations, such as when the physical properties are constant. In practice the commercial ohmic heater involves a flowing mixture where the electrical conductivities vary with temperature: flow through the heater must also be considered in detail. Full solution of the electric, temperature, and flow fields is not feasible given the complexity of the problem and the coupling of the equations via the variation of viscosity and electrical conductivity with temperature.

Various approximations have been tested to predict how flowing particle–liquid mixtures behave. Finite element models have been used to develop solutions for ohmic heating of solid–liquid mixtures: this type of model can simulate both complex shapes and temperature-dependent physical properties. Such models solve the Navier–Stokes equations for flow, but can also solve energy and potential field equations to predict temperatures. In each simulation, the vessel geometry can be discretized by a mesh, and the material properties and boundary conditions given as input data.

In flow situations, the full Navier–Stokes equation for fluid flow must be solved together with the temperature and voltage field. The assumption of no fluid motion can simplify modeling considerably,

shortening the computational run, and gives insight into the 'worst-case scenario': when there is little convective heating, then the temperature differences between different regions of a food mixture will be more pronounced. A number of computational models for the heating of two-phase mixtures have been developed and validated; if the physical properties of the materials are known, it is possible to predict the heating rates and patterns that will result.

The aim of simulations is to demonstrate the nature and extent to which effects might arise in real situations. The finite element method is expensive in computer time, although advances in hardware have made it possible to use high-end personal computers rather than workstations. It is best as a design and research tool rather than as the basis for a control system.

Design of Formulations and Processes

Using electrical heating it is possible to heat a solid–liquid mixture very rapidly. The key physical parameter is electrical conductivity, both of the mixture and of the individual phases. The overall electrical conductivity governs the power that can be supplied to the system. In any given system there will be limits on the voltage and current density that can be sustained. Voltage limits arise through the cost of transformers. Current is limited by the acceptable current densities on the electrode, recommended as $< 4000\ \mathrm{A\ m^{-2}}$ for the APV system. Locally, however, the electrical conductivity of individual regions of the system controls the heating rate.

Once it has reached the required state, it is vital to cool the food as rapidly as possible to avoid loss of product quality. From the heater, material will pass to the holding section and then to cooling. In a conventional process the holding section allows thermal equilibration between particle and liquid and holds the material at temperature long enough for the required level of sterility. In electrical heating equilibration may not be needed because particle and liquid temperatures are closely matched. For process control, it is useful if the particle temperature at the end of the heating section exceeds the liquid, as the liquid temperature is simplest to measure. However, any overheating of particles may impair quality, because it is necessary to remove heat by conduction during cooling. Any design should thus consider the effect of the whole process on the product.

Care must be taken in the design of formulations to eliminate temperature variations, and quality assurance protocols developed to insure that the electrical conductivity of the ingredients lies between acceptable limits. All the physical properties of the system must be known over the whole range of process temperatures, and the heating rates checked under batch conditions. The successful application of electrical heating produces a product of very high quality: developments in electrical equipment design and in process understanding will enhance the ability of food processors to use the technique.

See also: **Canning**: Cans and their Manufacture; Food Handling; Quality Changes During Canning; **Convenience Foods**; **Cooking**: Domestic Use of Microwave Ovens; **Heat Transfer Methods**; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; Chemical and Microbiological Changes; **Quality Assurance and Quality Control**

Further Reading

Davies LJ, Kemp MR and Fryer PJ (1999) The geometry of shadows: effects of inhomogeneities in electrical field processing. *Journal of Food Engineering* 40: 245–258.

de Alwis AAP, Halden K and Fryer PJ (1989) Shape and conductivity effects in the ohmic heating of foods. *Chemical Engineering Research and Design* 67: 159–168.

Halden K, de Alwis AAP and Fryer PJ (1990) Changes in electrical conductivity of foods during ohmic heating. *International Journal of Food Science and Technology* 25: 9–25.

Kim HJ, Choi YM, Yang TCS *et al*. (1996) Validation of ohmic heating for quality enhancement of food products. *Food Technology* 50: 253–261.

Kim HJ, Choi YM, Yang A *et al*. (1996) Microbiological and chemical investigation of ohmic heating of particulate foods using a 5 kW ohmic system. *Journal of Food Processing and Preservation* 20: 41–58.

Parrott DL (1992) Use of ohmic heating for aseptic processing of particles. *Food Technology* 46: 68–72.

Sastry SK and Palaniappan S (1992) Influence of particle orientation on the effective electrical resistance and ohmic heating rate of a liquid–particle mixture. *Journal of Food Processing and Engineering* 15: 213–227.

Stirling R (1987) Ohmic heating – a new process for the food industry. *Power Engineering Journal* 1(6): 365–371.

Yongsawatdigul J, Park JW, Kolbe E, Dagga YA and Morissey MT (1984) Ohmic heating maximises gel functionality of pacific whiting surimi. *Journal of Food Science* 60: 10–14.

Zhang L and Fryer PJ (1994) Food sterilization by electrical heating: sensitivity to process parameters. *American Institute of Chemical Engineering Journal* 40: 888–898.

HEAVY METAL TOXICOLOGY

G L Klein and W R Snodgrass, University of Texas Medical Branch, Galveston, TX, USA

Introduction

This article deals in brief with the toxic effects of the heavy metals, including **Cadmium**: Properties and Determination; **Lead**: Properties and Determination; and **Mercury**: Properties and Determination – each covered in depth in other articles – as well as nickel and bismuth. The latter two, although not obviously dietary toxicants, are nevertheless considered to be heavy metals; they do enter the body and can cause toxicity.

Lead

Toxic Effects

Children appear to be more susceptible to the effects of lead than adults. Manifestations of lead poisoning are neurological, hepatic, hematological, renal and endocrine, and genetic and reproductive.

Neurological Overt lead encephalopathy, consisting of delirium, truncal ataxia, irritability, lethargy, visual disturbances, vomiting, and coma, is less frequently seen than in the past. Peripheral neuropathy and paralysis are still reported in adults as a late sign of lead intoxication. Most striking, however, are the reports of learning difficulties and a small but widely confirmed and apparently irreversible reduction in intelligence quotient in children, leading to poor school performance. Contributing to the learning disability is a lead-associated hearing loss, especially of the high frequencies. These learning disabilities occur with blood lead concentrations previously thought to be safe. Adults also manifest behavioral changes at low blood lead levels. These include fatigue and impaired concentration.

Animal data, primarily but not exclusively from studies using rats, show that exposure to lead from birth can lead to hemorrhages, neuronal edema and necrosis, and mitochondrial dysfunction in both cerebrum and cerebellum. Rats exposed to lead beginning on day 10 of life had these changes only in the cerebellum. Rats exposed to lead beginning on days 20–24 of life had none of these alterations in histology. Other documented effects of lead on the nervous system include inhibition of calcium-stimulated neurotransmitter release at presynaptic terminals while stimulating spontaneous release of neurotransmitters, resulting in mini end-plate potentials. When combined with manganese, lead has produced increased peroxidation in rat brains. Lead has also recently been shown to inhibit nitric oxide synthase activity in mouse brain. In addition, lead has produced necrosis of photoreceptor cells in the rat retina and swelling of endothelial cells lining the retinal vessels, with consequent luminal narrowing. Lead can also cause hearing reduction and possible auditory nerve damage in rats. Moreover, there is a report describing disruption of brain microtubular assembly by organic but not inorganic lead.

Hepatic Animal studies have shown that when lead acetate is added to rat hepatocytes, there is decreased binding of L-tryptophan to hepatocellular nuclei, suggesting that lead may alter certain aspects of hepatocyte function as well.

Hematological Lead produces a microcytic anemia, primarily by interference with two enzymes necessary for heme biosynthesis, δ-aminolevulinic acid dehydratase (ALAD) and ferrochelatase. The latter enzyme catalyzes the incorporation of iron into protoporphyrin. Increased erythrocyte protoporphyrin results from this inhibiting effect of lead. With prolonged elevation of blood lead levels, erythrocyte survival time decreases.

The hematological effects of lead are similar in animals and humans, with the exception that in animals lead has been reported to inhibit the pentose phosphate shunt and glucose-6-phosphate dehydrogenase, thus providing a possible explanation for decreased erythrocyte survival seen with chronic lead poisoning. Also of interest is a lead-binding protein in kidney, which limits lead inhibition of ALA activity in kidney but not in liver.

Renal and endocrine At low blood lead levels, lead may contribute to hypertension. It may also produce proximal tubular dysfunction, leading to glycosuria, aminoaciduria, and hyperphosphaturia. Recent studies by the US National Institute of Occupational Health and Safety (NIOSH) suggest that lead exposure altered glutathione *S*-transferase expression in rabbit kidney, suggesting that the kidney may be more susceptible to peroxidative damage. Gout may develop as a result of hyperuricemia. Furthermore,

lead impairs the function of the renal enzyme 25-hydroxyvitamin D-1α hydroxylase, which converts 25-hydroxyvitamin D to the biologically active compound 1,25-dihydroxyvitamin D.

Renal and endocrine effects in animals include not only the enhancement of lead absorption in calcium deficiency but also the involvement of 1,25-dihydroxyvitamin D_3 in the increased absorption of lead as well as calcium. Furthermore, osteoclasts develop pyknotic nuclei and nuclear and cytoplasmic inclusion bodies that may lead to reduced bone resorption. Lead accumulated in bone can be released by resorption, increasing the local tissue lead concentration. This may be a mechanism by which calcium influx into osteoblasts is blocked by lead. Also, lead appears to reduce the natriuretic response to extracellular volume expansion and reduces binding of follicle-stimulating hormone to receptors in the Sertoli cells of the testes. Lead crossing the placenta in hamsters, chicks, and rats has produced urogenital, rectal, and vertebral malformations in the fetus.

Genetic and reproductive Lead has recently been shown to affect chromatin in cells by inhibiting DNA binding to zinc finger proteins, thus altering transcription. It is also possible that transplacental lead gives rise to congenital anomalies in humans as well as in animals, although the data are inconsistent. There is no convincing documentation of male reproductive organ toxicity of lead in humans, although this has been reported in animals. Similarly, there is no convincing evidence that lead is a carcinogen in humans.

Dietary and Environmental Sources of Lead

Although there is no known biological requirement for lead in humans, it is ubiquitous in an industrial environment. Ingestion or inhalation by children or adults occurs as a result of environmental contamination, including food and water. Children, especially underprivileged inner-city children, may ingest chips from lead-based paint surfaces and can mouth items containing lead from paint, dust, or soil. High levels of lead in soil and house dust have been associated with elevated blood lead levels in children. Lead is still present in unleaded gasoline; thus lead continues to be released into the atmosphere by all gasoline-burning vehicles. Lead from vehicle exhausts may be retained by crops, especially leafy vegetables. Acidic foods may leach lead from lead solder in cans; lead from soldered water pipes can contaminate tap water. The US Environmental Protection Agency (EPA) estimates that 20% of the American population consumes drinking water containing $>20\,\mu g$ of lead per liter ($0.1\,\mu mol\,l^{-1}$); this includes more than 3.8 million children.

Significance of Levels in the Diet

Individuals who are also subject to dietary iron deficiency, calcium and/or zinc deficiency risk increased lead absorption. Adults may absorb approximately 15% of ingested lead, while pregnant women and children may absorb up to 50%. Once in the blood, 99% of lead associates with erythrocytes, while 1% remains in plasma and is not available for tissue entry. Blood lead not retained is excreted in urine or bile. In single-exposure studies, lead has a biological half-life in blood of 25 days, in soft tissue of 40 days, and in bone up to 25 years. The latter pool can be mobilized without further intake under conditions of stress, such as pregnancy, lactation, and chronic disease. The significance of dietary and experimental exposure to lead is the impact these exposures have on blood lead concentrations and their effect on enzymes. Thus inhibition of ALAD occurs at a blood lead concentration of $5–10\,\mu g\,dl^{-1}$ ($0.25–0.5\,\mu mol\,l^{-1}$); decreased erythrocyte ferrochelatase occurs at blood lead concentration of $15\,\mu g\,dl^{-1}$ ($0.70\,\mu mol\,l^{-1}$). Inhibition of the renal enzyme 25-hydroxyvitamin D-1α hydroxylase occurs at $25\,\mu g\,dl^{-1}$ ($1.2\,\mu mol\,l^{-1}$) of blood lead, and subtle neurobehavioral changes will occur with blood lead concentrations of $10–15\,\mu g\,dl^{-1}$ ($0.5–0.7\,\mu mol\,l^{-1}$).

Intake Limits

Because of the neuropsychological changes that can occur with blood lead concentrations between 10 and $15\,\mu g\,dl^{-1}$ ($0.5–0.7\,\mu mol\,l^{-1}$), several states in the USA require physicians to report blood lead concentrations greater than $25\,\mu g\,dl^{-1}$ ($1.2\,\mu mol\,l^{-1}$). The US Occupational Safety and Health Administration has made mandatory the periodic blood lead determinations of workers exposed to ambient air lead concentration of $\geq 50\,\mu g\,m^{-3}$ for over 30 days per year. If workers have blood lead levels exceeding $40\,\mu g\,dl^{-1}$ ($1.9\,\mu mol\,l^{-1}$), the employer is obliged to remove the employees from excessive exposure until blood lead concentrations fall below $40\,\mu g\,dl^{-1}$ ($1.9\,\mu mol\,l^{-1}$). The US EPA requires that the lead content of air be less than $1.5\,\mu g\,m^{-3}$, of unleaded petrol $0.01\,g\,l^{-1}$ (0.05 g per US gal), and of leaded petrol $0.03\,g\,l^{-1}$ (0.10 g per US gal). In the USA the maximum amount of lead permitted in water is $50\,\mu g\,l^{-1}$, but this may be reduced to $5\,\mu g\,l^{-1}$. The US Food and Drug Administration Advisory Panel suggests that no more than 100 μg of lead per day be consumed from food, and the Consumer Product Safety Commission suggests that the lead content

of paint be no more than 0.06% (600 p.p.m.) dry weight.

Management of Lead Poisoning

Management of lead poisoning involves separation of the patient from the lead source, identification and treatment of anyone exposed to the source, and identification and correction of other nutrient deficiencies such as those of iron, zinc, and calcium. Aggressive screening of the population is also required. A recent survey of blood lead levels in Galveston, Texas in 1570 children 6 months to 8 years of age revealed that 19% of them had blood lead concentrations $\geq 10\,\mu g\,dl^{-1}$ ($0.5\,\mu mol\,l^{-1}$). Chelation therapy with oral dimercaptosuccinic acid should be started in anyone with a blood lead concentration of 25–55 $\mu g\,dl^{-1}$ (1.2–2.65 $\mu mol\,l^{-1}$). Two more recent oral chelating agents are dimercaptosuccinic acid and 2,3-dimercaptopropane-1-sulfonate; the latter is not commercially available in the USA. Both are water-soluble derivatives of dimercaprol, a chelating agent. A mobilization test using a chelator and noninvasive measurement of bone lead content by K-type X-ray fluorescence may be performed in selected patients above 18 years of age due to reproducibility of measurements based on bone calcium. (*See* **Lead**: Properties and Determination; Toxicology.)

Cadmium

Toxicity in Humans and Animals

Humans can be exposed to cadmium through either diet or industrial contact. The placenta is generally an efficient barrier against transmission of cadmium from an occupationally exposed mother to fetus. However, exposed mothers may exhibit hypertension during pregnancy and give birth to small-for-gestational-age infants with high blood cadmium concentration. Placental function may also be adversely affected as the presence of cadmium in human placental cultures decreases microvillous vesicle uptake of α-aminoisobutyric acid, a neutral amino acid. In addition, there is increased placental cadmium in women smokers. Thus cadmium may be one of the factors involved in the relationship between low-birth-weight babies and maternal smoking.

In the Jinzu river basin in Toyama Prefecture in Japan, primarily premenopausal women consuming a low-calcium diet were exposed to excessive cadmium as a contaminant of water and rice. These women developed itai-itai disease, manifested by bone pain, osteoporosis, pseudofractures, and a higher incidence of renal disease, including proteinuria (α- and β-microglobulin) and glycosuria. Other nonindustrial exposures have occurred in the past from leaching of cadmium from cadmium-plated containers by acidic drinks such as fruit juices. Cadmium-plated containers are now banned by public health laws in many American states.

Among the symptoms of chronic cadmium food poisoning are increased salivation, persistent vomiting, abdominal pain, and diarrhea. Also of concern is renal tubular damage. Since cadmium absorption is only 2–8% of dietary intake, and acute cadmium ingestion is almost always followed by vomiting, most clinically significant cadmium exposure occurs by inhalation of fumes or dust. Respiratory absorption of cadmium has been reported as anywhere from 12 to 30%. However, it is the total inhaled cadmium that is most likely to result in toxicity. Symptoms of acute toxicity include dyspnea, uncontrollable cough, cyanosis, followed by fever, gastrointestinal symptoms including abdominal pain, diarrhea, and vomiting, emphysema, and heart failure. In nonfatal cases the first symptoms are often throat irritation and chest soreness. Also experienced are headaches, nausea, vomiting, diarrhea, chills and weakness. In chronic respiratory exposure, fatigue, anosmia, dyspnea, weight loss, anemia, renal tubular dysfunction, hepatic dysfunction, and bone marrow involvement have been reported. The most prominent features are emphysema and proteinuria. Urinary cadmium excretion of approximately 200 μg (1.78 μmol) per day and a concentration of 200 μg of cadmium per gram of renal cortical tissue is associated with tubular proteinuria. Chronic rhinitis, osteoporosis, and pseudofractures may also be seen. A recent study in rats suggests that cadmium-associated nephrotoxicity may be mediated by oxidative stress-induced proteolysis of sodium–potassium adenosine triphosphatase. Mouse studies have shown that cadmium administration reduces the erythrocyte count and interleukin-1β and tumor necrosis factor, both of which would play a role in producing osteoporosis.

There is epidemiological evidence that cadmium exposure may be carcinogenic. A study of a cohort of 600 workers exposed to cadmium in a smelting plant for at least 6 months between 1940 and 1978 revealed a higher mortality rate from respiratory tract cancers than the rate in the general US population. Workers exposed for more than 2 years had an even greater mortality rate.

In animals, the range of toxic effects is greater. They include the following: testicular necrosis; placental destruction; abortion; fetal resorption or teratogenic malformation, including exencephaly, renal tubulopathy and glycosuria; liver damage; dental changes; deficiencies of calcium iron, copper, and zinc; skeletal decalcification; anemia; haemorrhagic

lesions of sensory ganglia; hypertension; pulmonary edema; and emphysema. In animals, cadmium is transiently bound to an unidentified plasma component and then appears in erythrocytes as whole blood concentration rises.

After ingestion or injection, cadmium is taken up by most tissues, but liver and kidney contain 50–75% of total body burden. Among the chief histopathological changes are peribronchial and perivascular fibrosis and emphysema in the lungs, and fatty infiltration of kidney and liver. Chronic exposure in rats is manifested by weight loss, anemia, leukocytosis, and infertility. Sarcomas have been observed at the site of subcutaneous injection of cadmium powder to rats.

Cadmium also produces neurological toxicity in some species. Cadmium crosses the blood–brain barrier more readily in neonatal rats than in adult rats. *In vitro* cadmium serves as a blocking agent at both adrenergic and cholinergic synapses. Adult rats given an intraperitoneal injection of cadmium show decreased exploration responses up to 9 days after cessation of cadmium dosing. A single intraventricular injection of cadmium to rats causes alterations in neurotransmitter levels throughout the brain. In addition, some fish show abnormal behavior after cadmium exposure via water contamination. In kidney and liver, cadmium binds to metallothionein. When intracellular accumulation exceeds metallothionein-binding capacity, toxicity is thought to occur. There is also a recent report suggesting that metallothionein may be protective against various toxic effects of cadmium.

Place in the Food Chain

Cadmium is found primarily as cadmium sulfide in ores containing zinc, lead, and copper. Cadmium volatilizes more rapidly than other metals when the ore is being smelted, and it condenses to form fine airborne particles that react immediately with oxygen to form cadmium oxide fumes. Cadmium is used industrially to plate other metals, in pigments, batteries, stabilizers for plastics, metallurgy, nuclear reactor rods, and semiconductors. Cadmium is deposited in soil and water near the industrial source. It is taken up by various marine organisms, especially plankton, molluscs, and shellfish. Brown meat of edible crabs may contain as much as 5–15 μg of cadmium per kg. Certain sea birds, both plankton-eating and fish-eating, have high liver cadmium levels – 20–50 μg per kg wet weight. Otherwise, there is no significant cadmium concentration in marine food chains since cadmium is toxic to fish and fish embryos.

Cadmium is a normal plant constituent and can be absorbed through both leaves and roots. Plants have no excretory mechanism for cadmium. Cadmium from soil is freely exchangeable, especially from acid soils, but its concentration there is normally low – 0.02–0.5 $\mu g\,kg^{-1}$. However, cadmium may be high in plants grown near the contaminating industrial site. Accumulation is greater at the roots than at the top of the plant, restricting the movement of cadmium through food chains. In domestic animals reared for human consumption the cadmium burdens are low. In cattle and other mammals the mammary gland is an efficient barrier to cadmium since only small amounts (1 $\mu g\,dl^{-1}$) are excreted in milk.

Significance of Levels in the Diet

In foodstuffs of animal origin, cadmium is present bound to metallothionein. Its bioavailability after exposure to gastric acid is unknown. Plankton and shellfish have the highest levels of cadmium. After absorption, cadmium is transported to tissue, where it binds to metallothionein, leaving the blood concentration less than 1 $\mu g\,dl^{-1}$ (0.1 $\mu mol\,l^{-1}$). Estimates of biological half-life are 17–33 years in human kidney and approximately 7 years in liver. The absorption of cadmium antagonizes, and is antagonized by, zinc, copper, ferrous and ferric ions, and calcium. Zinc and selenium protect against many acute effects of cadmium in animals, including testicular necrosis, placental damage, and teratogenicity. Copper prevents cadmium-induced anemia. Conversely, with marginal copper intake (3 mg kg^{-1} diet), low cadmium levels (1.5 $\mu g\,kg^{-1}$ diet) interfere with copper metabolism in rats and reduce plasma ceruloplasmin activity. Higher intakes of cadmium (6–18 $\mu g\,kg^{-1}$ diet) exacerbate those effects and reduce bone density. The presence of cadmium at a concentration of 12 $\mu g\,kg^{-1}$ in the maternal diet reduces copper storage in newborn lambs. The mechanism of cadmium antagonism to zinc and copper is unknown but may involve competition for transport sites on metallothionein. Iron transport in chicken duodenum is significantly reduced by dietary cadmium. In addition, calcium deficiency may increase dietary cadmium uptake. Furthermore, cadmium may inhibit conversion of 25-hydroxyvitamin D to 1,25-dihydroxyvitamin D, the biologically active form which acts as a steroid hormone.

Intake Limits

A lethal dose of cadmium following dust or fume exposure is 0.1 $mg\,m^{-3}$ of air for 1 h. The US NIOSH occupational exposure limit is 5 $\mu g\,m^{-3}$ air level per 40-h week. The amount of atmospheric cadmium inhaled by chronically exposed workers varies widely and cannot be correlated at any one

time with atmospheric cadmium. A study of 53 jig solderers in the UK showed that 32 workers employed (exposed) for more than 5 years and 11 of 21 exposed for less than 5 years had elevated blood and urine cadmium levels. Thirty of the 43 workers with elevated cadmium levels had urine cadmium concentrations exceeding 10 nmol $mmol^{-1}$ creatinine. Normal daily urinary cadmium excretion is 2–5 μg (0.02–0.045 μmol), while in unexposed individuals normal blood cadmium is less than 1 μg dl^{-1} (0.1 μmol l^{-1}).

Kidney cadmium, as measured *in vivo* by neutron activation, increased linearly to 15 years of exposure and then reached a plateau. Liver cadmium content increased linearly up to 20 years of exposure. Urinary β_2-microglobulin and retinol-binding protein were greater than the 95th percentile in 20 of 51 subjects. In 9 of the 20, urine protein values were less than twice the upper normal limit, suggesting tubular dysfunction. Both proteins remained within the normal range for up to 15 years of exposure and then increased dramatically. Urinary albumin concentration, a marker of glomerular function, was normal in 48 of 51 subjects. Blood and urinary cadmium concentrations by themselves are only indications of recent exposure, not of cadmium toxicity. Thus *in vivo* neutron activation measurements of tissue cadmium may be a better indicator of toxicity.

Management

The dithiocarbamates appear to be more effective than either dimercaprol or calcium disodium ethylenediaminetetraacetic acid (EDTA) in chelating cadmium in animals. In humans, calcium disodium EDTA has been used, as have dimercaprol, D-penicillamine, and diethyldithiocarbamate (a metabolite of disulfiram). Diethylenetriaminepentaacetic acid also increases renal cadmium excretion. (*See* **Cadmium**: Properties and Determination; Toxicology.)

Mercury

Toxic Effects in Humans and Animals

Mercury may enter the body in two forms: inorganic mercury, absorbed through skin, or from vapor escaping from mercury-containing 'silver' dental amalgam, or through the gastrointestinal tract as an organic salt, primarily of methyl mercury. In children, a symptom complex called acrodynia develops, consisting of redness of lips, throat, and tongue, loss of teeth, swelling, and redness of the skin with pink-red fingertips, palms, and soles. Redness of the conjunctival membranes of the eye and photophobia, enlarged cervical lymph nodes, joint pain, anorexia, and vascular thrombosis are also described. Effects on the nervous system include irritability, withdrawn behavior, proximal muscle weakness, loss of muscle tone, and reduced tendon reflexes.

Occupational exposure to mercury vapor in adults can give rise to acute respiratory distress and renal failure, as well as flu-like symptoms. With more prolonged exposure the chief manifestation is proteinuria resulting from renal tubular damage. This is thought to be caused by a mercury-induced autoimmune response. With long-term mercury vapour exposure in some workers, neurological symptoms develop, including reduced muscle strength and coordination, decreased sensation, and an increase in abnormal reflexes. In individuals who consumed fish contaminated by mercury from an industrial plant near Minamata Bay in Japan and those in Iraq who consumed bread made from wheat contaminated by a mercury-containing fungicide, symptoms of methyl mercury poisoning were primarily neurological, severe, and irreversible.

Exposed infants developed cerebral palsy-like symptoms, including psychomotor retardation, microcephaly, flaccid or spastic paralysis, unsteady gait, abnormal hand motions, visual field narrowing, and tonic convulsions. Exposure of pregnant women led to fetal toxicity, including low birth weight, cerebral and cerebellar histopathology, and severe mental retardation. Mercuric chloride can produce DNA damage in a human fetal hepatic cell line, as can lipid peroxidation.

In animals, mercury toxicity is much the same as in humans. Additional insights into the pathophysiology of mercury have been obtained. Thus mercuric chloride injection into rats can produce renal proximal tubular necrosis. However, in the brown Norway rat, mercuric chloride produces antiglomerular basement membrane antibodies, antiproximal tubular basement membrane antibodies, alterations in helper and suppressor T lymphocytes, and appearance of autoantibodies in peritubular capillaries. With regard to neurotoxicity, injection of pregnant rats with methyl mercuric chloride resulted in appearance of methyl mercury in the fetal nervous system within 2 h. Early changes included mitochondrial degeneration of the endothelium of cerebral capillaries, leading to hemorrhaging, which distorts the pattern of neuron development and migration. Methyl mercury is also believed to disrupt neuronal microtubular assemblage and to interact with sulfhydryl groups of tubular and membrane proteins, causing peroxidative injury. Inhaled mercury can cause respiratory distress and death in mice who lack metallothionein; the latter may be protective, as it is with lead.

Sources of Mercury

Inorganic mercury is primarily inhaled as dust or vapor from dental amalgams or industrial processing. When inorganic mercury is released into either fresh or salt water, it is converted to methyl mercury by methanogenic bacteria and enters the aquatic food chain. Two major epidemics of methyl mercury poisoning in Japan occurred as a result of consumption of contaminated fish from Minamata Bay in Japan following release into the bay of mercury-contaminated waste from an acetaldehyde plant. Agricultural contamination may occur as a result of the use of mercury-contaminated pesticides, such as occurred in Iraq. Finally, poultry products and eggs from hens raised on mercury-contaminated fish meal have varying quantities of mercury. Most of the poultry products fall within the Food and Agriculture Organization/World Health Organization (FAO/WHO) limits for acceptable mercury contents of foods other than fish.

Recently, the adjuvant of the hepatitis B vaccine as well as hepatitis B immune globulin were identified as being contaminated with mercury. Corrective measures are now being taken.

Significance of Levels in the Diet

Dietary intake of mercury is generally not well documented outside of major epidemics. FAO/WHO limits for dietary mercury intake are 0.3 mg mercury per person per week, of which no more than 0.2 mg mercury should be present as the methyl mercury ion. Seafood is probably the primary source of dietary contamination with mercury. Acid rain has increased mercury concentration in the edible tissues of fish, resulting in the closure of many American lakes and rivers to sport fishing. Furthermore, shark and swordfish sales have been restricted because of the high mercury content of these fish. Most poultry products fall below FAO/WHO acceptable limits of mercury of 50 p.p.b. In addition, children, especially toddlers, mouth objects contaminated with soil or household dust, either of which might be mercury-contaminated.

Intake Limits

Whole-blood mercury levels greater than 50 nmol l^{-1} (1 μg dl^{-1}) or urinary mercury concentration of greater than 100 nmol l^{-1} (20 μg l^{-1}) indicate recent mercury ingestion or inhalation. Symptoms of acrodynia have been reported in toddlers with urinary mercury concentration of 249 nmol l^{-1} (50 μg l^{-1}). Whole-blood mercury levels in cases of acute occupational exposure, with nephrotoxicity and neurotoxicity, have been reported to be 24 and 27 μmol l^{-1} (48 and 54 μg dl^{-1}). Hair mercury content is reported to be a reliable indicator of methyl mercury load. Distribution of hair and whole-blood mercury concentrations are in equilibrium, with hair mercury being about 250 times greater than whole-blood mercury. Treatment of choice at present is chelation with dimercaptosuccinic acid to reduce the body burden of mercury. (*See* **Mercury**: Properties and Determination; Toxicology.)

Other Heavy Metals

Nickel

Toxicity Nickel is a heavy transition metal, atomic weight 58.69, to which the general population is exposed primarily as an industrial emission rather than as a dietary component or contaminant. Its concentration in the ambient air in the vicinity of industrial plants is generally less than 1 μg m^{-3}, and this is not considered dangerous to the general public. However, workers exposed to nickel dust in industries such as stainless-steel and nickel alloy production, electroplating, and battery and coinage manufacture are at risk of developing respiratory or nasal neoplasms. Extensive epidemiological studies of cohorts of nickel workers suggest that prolonged exposure to oxides and sulfides of nickel in ambient concentrations exceeding 1 mg m^{-3} is associated with increased risk of lung and nasal tumors. Pulmonary tissue nickel concentration in one study in Germany showed that in high nickel-emission areas in the Ruhr valley pulmonary nickel was threefold greater than in a low-exposure population. Pulmonary nickel concentration in 6 patients who developed bronchial carcinoma ranged from 0.6 to 20.6 μg g^{-1} dry weight, compared to concentrations in low-exposure control lungs of 0.04–0.36 μg g^{-1} dry weight of lung tissue. Half-life studies performed in rats exposed to nickel monosulfide by aerosol inhalation revealed a 20-h half-life in the rat lung. This contrasted with a 221-month half-life for green nickel oxide. No malignancies were seen with the shorter half-life. The mechanism for the causation of respiratory malignancies is uncertain, but may involve nickel induction of chromosome abnormalities and oncogene amplification, as occurs in the production of experimental renal tumors in rats given nickel subsulfide. It has also been shown in rats that nickel administration results in a decreased number of antibody-forming cells in lung-associated lymph nodes, suggesting that alteration of the immune response may play a role in tumorigenesis. In another study, administration of nickel to women showed an increase in serum levels of interleukin-5 4 h after ingestion and a decrease in

CD4 and increase in CD8 lymphocytes 24 h after nickel inhalation. Antioxidants may have a protective effect.

Bismuth

Bismuth is primarily ingested as an antiulcer medication in the form of colloidal bismuth subcitrate or colloidal bismuth subsalicylate.

Toxicity in humans There have been many reports of neurotoxicity in Europe and Australia in individuals ingesting large quantities of bismuth subcitrate or subgallate. There have been isolated cases of patients taking large doses of bismuth subsalicylate and subcitrate, both without impaired renal function. Although bismuth is reported to be poorly soluble in biological fluids, solubility in digestive secretions is increased by the presence of urine or ascorbic acid. Neurotoxic manifestations include numbness and tingling of the extremities, irritability, insomnia, poor concentration, impaired short-term memory, tremors, ataxia, and abnormal electroencephalographic tracings. These symptoms are reversible with discontinuation of the bismuth compounds and with the use of a chelating agent, such as 2,3-dimercapto-1-propane sulfonic acid. Nephropathy following bismuth overdose has also been reported, as has osteoarthropathy. One case of acute reduction in blood platelets has also been reported with bismuth subcitrate. This effect may have been the result of an idiosyncratic drug reaction.

Distribution and toxicity in animals Animal studies indicate that most absorbed bismuth accumulates in the kidneys and only trace amounts in the brain. Tissue distribution of bismuth in humans is unknown. Experimental bismuth neurotoxicity has been produced in mice. Reproduction of tremors, ataxia, myoclonus, and convulsions was associated with hydrocephalus and an average of 8 μg bismuth per gram of brain tissue. Animals with brain bismuth concentration of $4\,\mu g\,g^{-1}$ did not show evidence of neurotoxicity but did manifest hydrocephalus with no other histopathology. When brain bismuth concentration was $1\,\mu g\,g^{-1}$ there was no evidence of hydrocephalus, neurotoxicity, or histopathology.

Sources of bismuth Average dietary intake in humans is not well quantified, but is probably $< 5\,\mu g\,day^{-1}$. It is poorly absorbed from the gastrointestinal tract, but with chronic medication small amounts can accumulate in blood, although generally this amount does not exceed $50\,\mu g\,l^{-1}$. The rate of urinary clearance of bismuth when ingested in the subcitrate form decreases by approximately 2.6% per day and reaches a steady state by 2 weeks after discontinuation of the compound. However, persistence of urinary bismuth concentration of as much as $250\,\mu g\,l^{-1}$ ($1.2\,\mu mol\,l^{-1}$) suggests that there is tissue accumulation of bismuth, and slow mobilization.

Intake limits Data from over 1000 cases of bismuth neurotoxicity in France suggest that a whole-blood bismuth level greater than $180\,\mu g\,l^{-1}$ ($0.86\,\mu mol\,l^{-1}$) is associated with neurotoxicity; if whole-blood bismuth concentration is greater than $100\,\mu g\,l^{-1}$ ($0.48\,\mu mol\,l^{-1}$), bismuth compounds should be discontinued. With whole-blood levels between 50 and $100\,\mu g\,l^{-1}$ ($0.25–0.48\,\mu mol\,l^{-1}$), individuals should be monitored for signs of neurotoxicity, and if less than $50\,\mu g\,l^{-1}$, no monitoring is needed.

See also: **Cadmium**: Properties and Determination; Toxicology; **Lead**: Properties and Determination; Toxicology; **Mercury**: Properties and Determination; Toxicology

Further Reading

Agency for Toxic Substances and Disease Registry (1988) *The Nature and Extent of Lead Poisoning in Children in the United States; A Report to Congress*.

Bierer DW (1990) Bismuth subsalicylate: history, chemistry and safety. *Reviews of Infectious Diseases* 12 (suppl. 1): S3–S8.

Browning E (1969) *Toxicity of Industrial Metals*, pp. 98–103. Oxford: Butterworth Heinemann.

Cooper GP and Manalis RS (1983) Influence of heavy metals on synaptic transmission: a review. *Neurotoxicology* 4: 69–83.

Froomes PR, Wan AT, Keech AC *et al.* (1989) Absorption and elimination of bismuth from oral doses of tripotassium dicitrate bismuthate. *European Journal of Clinical Pharmacology* 37: 533–536.

Kollmeier H, Seeman JW, Rothe G *et al.* (1990) Age, sex and region adjusted concentrations of chromium and nickel in lung tissue. *British Journal of Industrial Medicine* 47: 682–687.

Mertz W (ed.) (1987) *Trace Elements in Human and Animal Nutrition*, 5th edn, vols I and II. London: Academic Press.

National Institute for Occupational Safety and Health (1978) *Occupational Health Guidelines for Chemical Hazards*, pp. 1–5. Cincinnati: US Department of Health and Human Services.

National Institute for Occupational Safety and Health (1984) *Cadmium (Cd)*. Current Intelligence Bulletin 42, pp. 1–9. Cincinnati: US Department of Health and Human Services.

Needleman HL, Schell A, Bellinger D *et al.* (1990) The long-term effects of exposure to low doses of lead in

childhood. An 11-year follow-up report. *New England Journal of Medicine* 322: 83–88.
Report of the International Committee on Nickel Carcinogenesis in Man (1990) *Scandinavian Journal of Work and Environmental Health* 49: 1–648.
Royce SC and Needleman HL (1990) *Agency for Toxic Substances and Disease Registry Case Studies in Environmental Medicine*, pp. 1–20. Atlanta: US Department of Health and Human Services, Public Health Service.
Shukla G and Singhal R (1983) The present status of biological effects of toxic metals in the environment: lead, cadmium, manganese. *Canadian Journal of Physiology and Pharmacology* 62: 1015–1031.
Smith NJ, Topping MD, Stewart JD and Fletcher JI (1986) Occupational cadmium exposure in jig solderers. *British Journal of Industrial Medicine* 43: 663–666.
Wagstaff AJ, Benfield P and Monk JP (1988) Colloidal bismuth subcitrate. A review of its pharmacodynamic and pharmacokinetic properties and its therapeutic use in peptic ulcer disease. *Drugs* 36: 132–157.
Webb M (1975) Cadmium. *British Medical Bulletin* 31: 246–250.

HEMAGGLUTININS (HAEMAGGLUTININS)

G P Savage, Lincoln University, Canterbury, New Zealand

Background

The presence of heat-labile toxic factors in plant products, mainly in leguminous seeds, makes them unsuitable as food for humans unless they are properly cooked. There is evidence, however, that leguminous seeds were an important part of the diet of prehistoric man, and the use of fire for cooking must have had an important bearing on the safe use of leguminous seeds in the diet. In the late eighteenth century, it was noticed that plant lectins, or phytohemagglutinins (hemagglutinins) occurred in castor bean seeds and that these compounds could agglutinate red blood cells. (*See* **Legumes**: Dietary Importance.)

The ingestion of raw or incompletely cooked beans regularly leads to incidents that at first appear to be cases of food poisoning. When no causative organism appears to be involved, high concentrations of hemagglutinins are usually implicated. Vomiting and diarrhea usually occur within 2–3 h after consumption. As few as four or five beans are often the cause. In many instances, the beans have been soaked for a few hours and then eaten without further cooking.

General Properties

Hemagglutinins are proteins that possess a specific affinity for certain sugar molecules. Since carbohydrate units exist in most animal cell membranes, the hemagglutinins may attach to these receptor groups. This attachment will occur only if the lectin molecule has at least two active groups. Hemagglutinins are characterized and detected by their action on red blood cell membranes, causing the blood cells to clump together. Other cells can also be affected. The receptor site on the cell surface must be exposed in order to react with a specific lectin. Pretreatment of red blood cells with a protein-digesting enzyme, such as papain, trypsin, or pronase, activates the cells and presumably exposes the receptor sites.

Structure

All plant lectins are proteins. Some contain a range of covalently bound sugars and can therefore be called glycoproteins. Lectins from different legumes show a remarkable range of specificity towards various animal red blood cells, which suggests that there is a considerable range of structures in this group. (*See* **Protein**: Food Sources.)

Location

Most hemagglutinins of higher plants are found in seeds, but tubers and plant sap are also sources of lectins. Lectins are most commonly found in leguminous seeds. They have been observed in potatoes and cereals, but the amounts in these foods do not appear to have any adverse effects.

Lectin Content of Edible Leguminous Seeds (see Table 1)

The seeds of a wide range of edible legumes are known to contain proteins that agglutinate red blood cells. Some of these hemagglutinins have been suggested to contribute to the poor nutritive value of raw beans. The level of toxicity of kidney beans has

Table 1 Hemagglutinin content of a number of legumes

Legume	*Hemagglutinin (units per gram dry wt)*	*Hemagglutinin removed (%) on soaking in water for 18 h*
Phaseolus vulgaris		
Red kidney beans	37 000–53 000	22–66
White kidney beans	17 000–43 000	18–36
Rose coco beans	39 000	18
Lens culinaris		
Red lentils	77 000	22
Green lentils	18 000	19
Split peas	1 000–16 000	11
Pisum sativum		
Garden peas	5 100	65
Vigna sinensis		
Black-eye beans	1 000	100

From Bender AE (1983) Haemagglutinins (lectins) in beans. *Food Chemistry* 11: 309–320.

been shown to be directly related to the lectin content and hence hemagglutinating activity. Using rat growth experiments and hemagglutinin tests on a range of leguminous seed lines available in the UK, leguminous seeds have been classified into four broad groups. In this study, group A seeds from most varieties of kidney (*Phaseolus vulgaris*), runner (*P. coccineus*) and tepary (*P. acutifolius*) beans showed high reactivity with red blood cells from a range of different species and were also highly toxic (none of the rats survived the feeding experiment). Group B contained seeds from lima or butter beans (*Phaseolus lunatas*) and winged bean (*Phosphocarpus tetagonolobus*), and they agglutinated only human and pronasetreated rat erythrocytes. The seeds did not support proper growth, but the animals did survive the feeding experiment. Seeds included in group C were lentils (*Lens culinaris*), peas (*Pisum sativum*), chickpeas (*Cicer arietinum*), black-eye peas (*Vigna sinensis*), pigeon peas (*Cajanus cajan*), mung beans (*Phaseolus aureus*), field or broad beans (*Vicia faba*), and adzuki beans (*Phaseolus angularis*). These generally had a low reactivity with all cells and were nontoxic. Group D contained soya (*Glycine max*) and pinto (*Phaseolus vulgaris*) beans, which generally had a low reactivity with all cells but caused growth depression at certain levels of inclusion in the diet. It was thought that the growth depression in this group was due to antinutritive factors other than lectins. The experiment showed that no single erythrocyte type could be used as a sole indicator of toxicity. The potential toxicity could, however, be predicted from the response to various erythrocytes. The usefulness of animal feeding trials was clearly shown by these experiments. (*See* **Beans; Peas and Lentils**.)

Assay Methods

The detection of lectins in plant extracts is generally performed by a serial dilution technique followed by visual estimation of the end point. Washed red blood cells are activated by pretreatment with a suitable proteinase (pronase, trypsin, papain), and the test is always carried out in saline solution (0.9%). A more quantitative method is based on the photometric measurement of the density of red blood cells that have not been agglutinated by the extracted lectin. Hemagglutinin activity is usually expressed as hemagglutinin units per milligram of sample. One unit is usually defined as the lowest amount of sample required for agglutination under the test conditions, but has also been defined as the amount of material per milliliter in the last dilution giving 50% agglutination.

The rather subjective nature of this test makes it quite difficult to compare results between different research centers, but it gives good results in the hands of experienced workers. Many workers use a standard reference material to calibrate their own experiments. Wide use of a reference sample between research centers would make comparison of published data more feasible.

Biological Effects of Hemagglutinins

The *in vitro* precipitation of red blood cells does not explain the *in vivo* effects of lectins. Lectins cause agglutination or 'clumping together' of cells by binding to the saccharides branching from lipid and protein molecules of the cells' outer surface. A definite relationship between the lectin content of legume-containing diets and the nutritive value has been made. The feeding of lectin preparations from some legumes has been shown to cause definite growth retardation and, in extreme cases, death. In contrast, it has been shown that white pea hemagglutinin has no growth-depressing or toxic effects on rats when fed at a level of 1% in the diet.

Many ingested lectins cause toxicity by binding to the surface of intestinal epithelial cells, which results in the reduced absorption of nutrients across the intestinal wall. Suppression of intestinal disaccharidase activity and proteolytic activity has also been observed. Since the disaccharidases are involved in the breakdown of carbohydrates to monosaccharides prior to absorption in the small intestine, their reduced activity would be expected to lead to reduced digestibility of dietary carbohydrates. The decreased proteolytic activity would also lead to reduced availability of amino acids and peptides for absorption. The growth inhibition caused by lectins can therefore be attributed to the reduction in absorption of

nutrients due to their disruption of the absorptive area of the intestine. (*See* **Carbohydrates**: Digestion, Absorption, and Metabolism.)

Detoxification

The detoxification of plant lectins is usually achieved by traditional methods of household cooking. There are conditions where this type of processing is not completely effective. When leguminous seeds are used in the manufacture of animal feeds, cooking is usually omitted, although steam pelleting of the final mixed feed may have some effect in the degradation of antinutritive factors. The time and temperature of the feed during the pelleting process are clearly insufficient to degrade a significant proportion of the hemagglutinins. As manufacturers do not want reduced food utilization occurring with their products, they usually limit the amount of leguminous seed protein incorporated in their mix.

Lima bean lectin has been observed to be stable between pH 5 and 7 at elevated temperatures. Addition of *N*-acetyl-L-cystine (NAC) inactivated lectin at a lower temperature than in its absence. This suggests that NAC reacted with the disulfide linkage of the lectin to bring about inactivation. NAC had no effect on lectins extracted from soya bean flour; this supports the proposed role of NAC, as it is known that soya bean lectins do not contain any disulfide bridges. Addition of NAC prior to heatprocessing of some legume flours may reduce the amount of heat processing required and improve the overall quality of the final product.

Hemagglutinins are almost completely destroyed by boiling for 60 min at atmospheric pressure. Autoclaving at 105 °C for 30 min is also effective. If leguminous seeds are previously soaked for 18 h, the hemagglutinins can be completely destroyed by boiling for 2–5 min. Other traditional techniques used to improve the nutritional value of legume seeds involve soaking and germinating. It has been shown that germination of lima beans for 6 days results in a 50% reduction in activity. Soaking for the same time results in an overall 34% reduction in activity. It should be noted that soaking, though not as effective as germination, would only be the first step in preparation; normal cooking would destroy further lectin in the beans.

Dry heat treatment (roasting) to destroy hemagglutinins is remarkably ineffective. For example, winged beans (*Psophocarpus tetragonolobus*) heated at 100 °C for 2 h show less than 5% reduction in lectin activity. The generally low but quite variable levels of hemagglutinins in this seed could be easily destroyed by autoclaving at 120 °C for 5 min.

The treatment of whole cow peas (*Vigna unguiculata*) with microwaves (2400 MHz) for 6 min was almost ineffective in destroying hemagglutinins, while dry roasting (160 °C for 50 min) led to between 28 and 47% reduction in lectin content. Treatment of cow peas with mild alkali (0.5% sodium bicarbonate for 12 h) led to an 80% reduction in hemagglutinin content. (*See* **Cooking**: Domestic Use of Microwave Ovens.)

Infrared heating of winged bean (*Psophocarpus tetragonolobus*) for 60 s was remarkably effective in reducing the hemagglutinating activity of this seed, whereas oven and microwave heating had little effect.

Conclusions

Extensive research has been carried out on the chemical, physical, and biochemical properties of hemagglutinins, but the understanding of their toxicological and nutritive properties is limited. These properties are an important aspect of our knowledge as legumes provide a vital source of protein in developing countries. The preparation of protein isolates from leguminous seeds and the use of these in quick cooking food products need careful consideration of the possible antinutritional effects of residual hemagglutinins. It is possible that these toxic effects may be only partially destroyed, and the resulting, slightly reduced food utilization may be difficult to detect. It is not clear what the long-term effect of consuming low levels of these compounds will be.

Hemagglutinins occur in a wide range of leguminous seeds. Lupins are the only common legume that do not appear to contain any lectins. Legumes contain a wide variety of toxins, including hemagglutinins, that are heat-labile and should therefore present no health hazard when properly cooked. Thus, it is surprising that outbreaks of poisoning following the consumption of incompletely cooked leguminous seeds occur in an otherwise well-informed society.

See also: **Beans**; **Carbohydrates**: Digestion, Absorption, and Metabolism; **Cooking**: Domestic Use of Microwave Ovens; **Legumes**: Dietary Importance; **Peas and Lentils**; **Protein**: Food Sources

Further Reading

Bender AE (1983) Haemagglutinins (lectins) in beans. *Food Chemistry* 11: 309–320.

Boyd WC (1963) The lectins: Their present status. *Vox Sanguinis* 8: 1–32.

Golstein IJ, Monsigny M, Osawa T and Sharon N (1980) What should be called a lectin? *Nature* 285: 66.

Grant G, More LJ, McKenzie NH, Stewart JC and Pusztai A (1983) A survey of the nutritional and haemagglutination properties of legume seeds generally available in the U.K. *British Journal of Nutrition* 50: 207–214.

Hill GH (1977) The composition and nutritive value of lupin seed. *Nutrition Abstracts and Reviews (Series B)* 42: 511–529.

Huprikar SV and Sohonie K (1965) Haemagglutinin in Indian pulses. 2. Purification and properties of haemagglutinin from white pea. *Pisum sativum*. *Enzymologia* 28: 300–345.

Jaffé WG (1960) Effect of feeding lectins and peas on the intestinal and hepatic enzymes of albino rats. *Arzneimittel Forschung* 10: 1012–1016. (Cited by Jindal *et al*.,1982.)

Jaffé WG (1969) Haemagglutinins. In: Liener IF (ed.) *Toxic Constituents of Plant Food Stuffs*, 1st edn, pp. 69–74. New York: Academic Press.

Jaffé WG (1980) Hemagglutinins (lectins). In: Liener IF (ed.) *Toxic Constituents of Plant Food Stuffs*, 2nd edn, pp. 73–102. New York: Academic Press.

Jaffé WG and Vega Lette CL (1968) Heat-labile growth-inhibiting factors in beans (*Phaseolus vulgaris*). *Journal of Nutrition* 94: 203–210.

Jaffé WG, Moreno R and Wallis V (1973) Amylase inhibitors in legume seeds. *Nutrition Reports International* 7: 169–173.

Jindal S, Soni GL and Singh R (1982) Effect of feeding of lectins from lentils and peas on the intestinal and hepatic enzymes of albino rate. *Journal of Plant Foods* 4: 95–103.

Kadam SS, Smithard RR, Eyre MD and Armstrong DG (1987) Effects of heat treatments of antinutritional factors and quality of proteins in winged bean. *Journal of the Science of Food and Agriculture* 39: 267–275.

Liener IE (1955) The photometric determination of the hemagglutinating activity of soyin and crude soybean extracts. *Archives of Biochemistry and Biophysics* 54: 223–231.

Liener IE (1962) Toxic factors in edible legumes and their elimination. *American Journal of Clinical Nutrition* 11: 281–298.

Liener IE (1979) The nutritional significance of plant protease inhibitors. *Proceedings of the Nutrition Society* 38: 109–113.

Muzquiz M, Burbano C, Gorospe MJ and Rodenas I (1989) A chemical study of *Lupinus hispanicus* seed – toxic and anti-nutritional components. *Journal of the Science of Food and Agriculture* 47: 205–214.

Ologhobo AD and Fetuga BL (1984) Haemagglutinin extracts from raw, sprouting and differentially processed lima beans. *Journal of Agricultural Science (Cambridge)* 102: 241–244.

Pusztai A and Palmer R (1977) Nutritional evaluation of kidney beans (*Phaseolus vulgaris*): the toxic principle. *Journal of the Science of Food and Agriculture* 28: 620–623.

Reaidi GB, McPherson L and Bender AE (1981) Toxicity of red kidney beans (*Phaseolus vulgaris*). *Journal of the Science of Food and Agriculture* 32: 846–847.

Romestand B, Corbier F and Roch P (2002) Protease inhibitors and haemagglutinins associated with resistance to the protozoan parasite, *Perkinsus Marinus*, in the Pacific oyster, *Crassostrea gigas*. *Parasitology* 125(4): 323–329.

Sales MG, Maia GA, Teles FFF *et al*. (1984) Cowpea solids prepared by mild alkali treatment, dry roasting, and microwave heating. *Nutrition Reports International* 30: 973–981.

Tan NH, Rahim ZHA, Khor HT and Wong KC (1983) Winged bean (*Psophocarpus tetragonolobus*) tannin level, phytate content, and hemagglutinating activity. *Journal of Agricultural Food Chemistry* 31: 916–917.

Wallace JM and Friedman M (1985) Inactivation of haemagglutinins in lima bean (*Phaseolus lunatus*) flour by *N*-acetyl-L-cysteine, pH and heat. *Nutrition Reports International* 32: 743–748.

HEMICELLULOSES

M T Holtzapple, Texas A&M University, College Station, TX, USA

Introduction

Hemicellulose is the world's second most abundant carbohydrate (after cellulose). An estimated 45 000 million tonnes are produced annually, a daily production of about 20 kg per person. (*See* **Carbohydrates: Classification and Properties.**)

Hemicellulose comprises 20–30% of plant cell walls, which have a composite structure analogous to fiberglass resins. Cellulose provides cell rigidity and strength (like the glass fibers) and is embedded in a hemicellulose matrix encrusted with lignin which bonds the entire structure together. (*See* **Cellulose; Lignin.**)

Definition

Hemicellulose is defined as the cell wall polysaccharides extractable in dilute ($10 \pm 8\%$) sodium hydroxide solutions. It includes all the cell wall polysaccharides except cellulose and pectin. The term 'hemicellulose' was first used by Schulze in 1891,

who believed these extractable polysaccharides to be cellulose precursors. This is now known to be absolutely false, but the name has remained. Staudinger has proposed the name 'polyoses' to prevent this erroneous association, but it has not gained widespread acceptance.

Hemicellulose is quite distinct from cellulose. Hemicellulose polymers are short, with a degree of polymerization (DP) of 50–200. In contrast, cellulose polymers are much longer, with a DP of 500–15 000. Cellulose is a linear polymer without attached side groups, whereas some hemicelluloses are Y-branched and most have attached side groups. Cellulose is a homopolymer of glucose, whereas hemicellulose generally is a heteropolymer composed of many sugars and modified sugars.

Structure

Monomers

Hemicellulose monomers are shown in Figure 1. Three monomers are hexoses (glucose, galactose, and mannose) and two are pentoses (xylose and arabinose). Some of the sugars are acetylated (glucose, mannose, and xylose). Glucose appears as its uronic acid and also as the methylated uronic acid. Most of the hemicellulose sugars appear as D enantiomers except for three L enantiomers (arabinose, fucose, and rhamnose). Galactose is primarily a D enantiomer, although the L enantiomer is occasionally present. Two 6-deoxy sugars (fucose, rhamnose) may be found in small quantities. (*See* **Galactose**.)

Linear free-aldehyde sugars are unstable and cyclize into a ring structure (Figure 1). In reality, the ring resembles a garden chair, but simpler Haworth projections are commonly used. The ring can have six members (pyranose) or five members (furanose). In hemicellulose, arabinose appears as both a pyranose and a furanose, whereas the others appear only as pyranoses. When the linear sugar cyclizes, C1 becomes chiral. Its axial hydroxyl groups are designated α and its equatorial hydroxyl groups are designated β. Both α and β sugars appear in hemicellulose, although the β forms are more common.

Hemicellulose can be classified into three families (xylans, mannans, and galactans), named here according to the backbone polymer. (Note: the naming conventions for hemicellulose vary widely.)

Xylan

Xylan is known by many names – including glucuronoxylan, araboxylan, glucuronoarabinoxylan, and L-arabino(4-*O*-methyl-D-glucurono)xylan – which reflects its heterogeneous nature. Figure 2a shows the backbone polymer is β-1,4-linked xylose. Numerous side groups can attach to the backbone, as shown in Figure 2b. Generally, 10–20% of the xyloses have side groups (i.e., $n = 4$–9), although 30–40% is sometimes observed (i.e., $n = 1.5$–2.5) in cereal flours. Most of the side-group attachments are made through arabinose, which may occur singly or with additional attached groups (xylose, galactose, or 4-*O*-methyl-D-glucuronic acid). Galactose is generally found in the xylans of annual plants. Glucuronic acid and its methylated form may also attach directly to the backbone at xylose C2, although C3 attachment is sometimes observed. Often, the xylose backbone is partially acetylated, particularly in the C2 position, although C3 acetylation is found. Xylans are generally linear, with a DP of 50–200. In some xylans, there are one to two Y branches in a single molecule.

The following generalizations may be made about xylans from different plants:

- monocotyledons – arabinose side groups attached along with glucuronic acid (or its methylated form);
- dicotyledons (including hardwoods) – 4-*O*-methylglucuronic acid attached every 10th xylose;
- softwoods – 4-*O*-methylglucuronic acid attached every sixth xylose with small amounts of arabinose side groups.

One of the few examples of homopolymer xylan is produced by esparto grass (Figure 2c) that contains Y-branched xylose backbone without attached side groups. (Note: esparto grass also makes some heteropolymer xylan with attached arabinose.)

Wheat flour xylan (Figure 2d) has irregularly attached arabinose on about 30–40% of the xylose backbone. 'Open regions,' consisting of about five sequential bare xyloses, occur about every 20–25 xyloses. Wheat bran xylan is similar, except it has more side groups (65% of xylose has attached arabinose) with glucuronic acid on every seven to eight xyloses. Barley husk xylan (Figure 2e) also has attached side groups of glucuronic acid and xyloarabinose. The xyloarabinose side group also occurs in corn cob xylan and perennial rye grass. Maize seed coat xylan (Figure 2f) has many side groups, including arabinose, xylose, galactose, and glucuronic acid. (*See* **Barley**.)

Hardwood xylan (Figure 2g) has 4-*O*-methylglucuronic acid every 10th xylose. Of these 10 xyloses, approximately 3.5–7 are acetylated primarily at the C3 position, although C2 acetylation is also found. Softwood xylan (Figure 2h) has more side groups because 4-*O*-methylglucuronic acid occurs

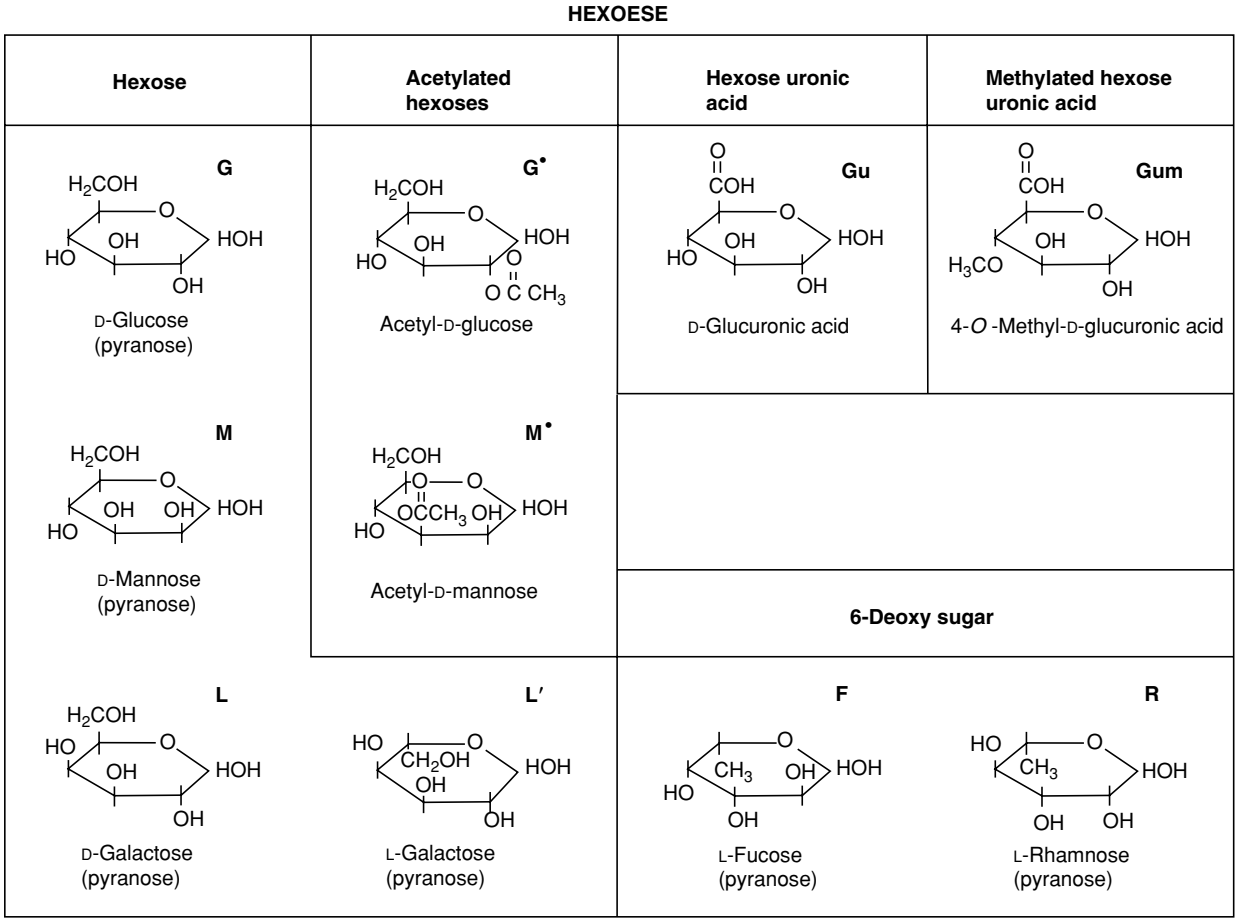

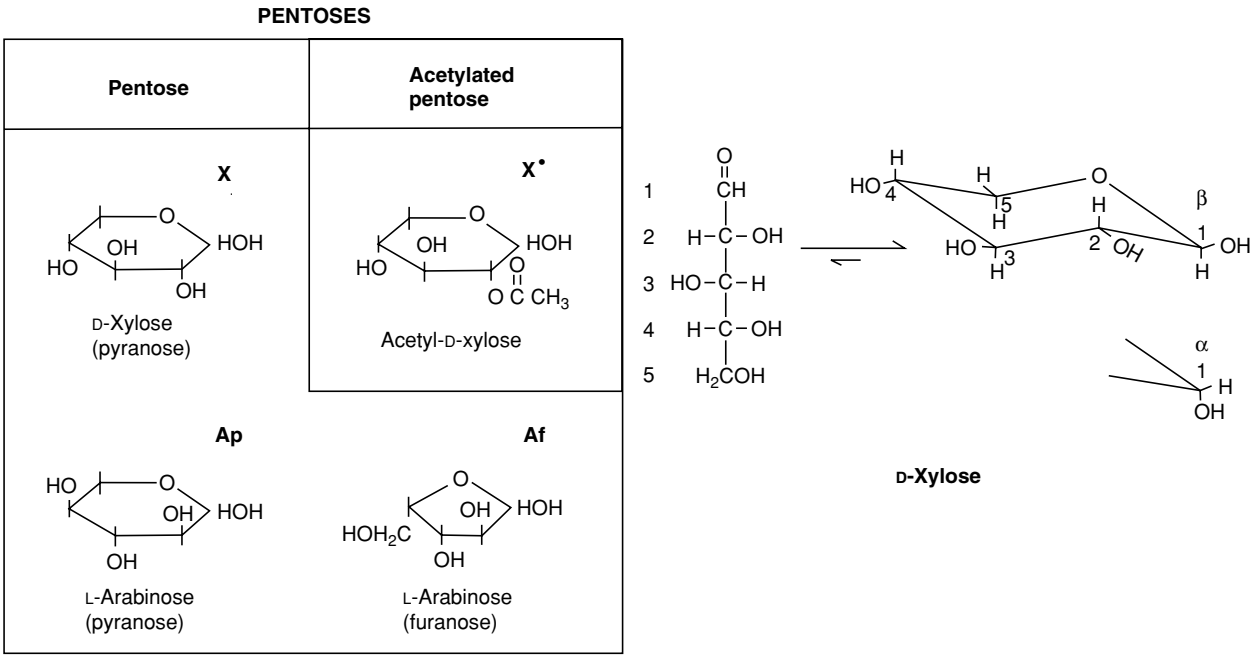

Figure 1 Hemicellulose monomers. Reproduced from Hemicelluloses, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

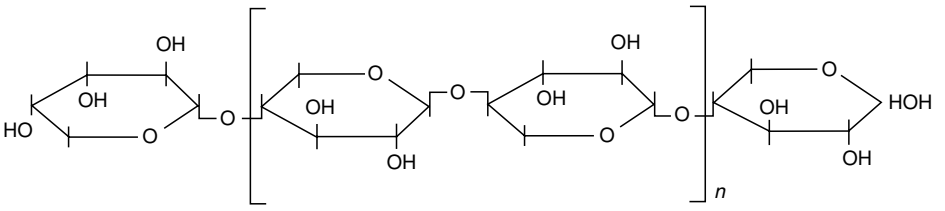

(a)

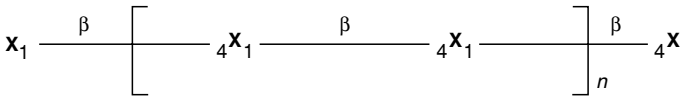

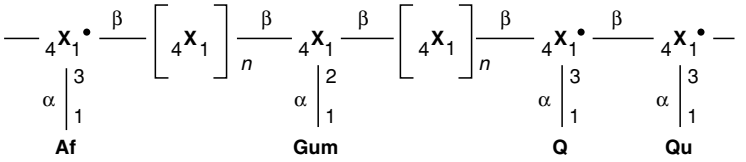
Af
Gum
Q
Qu

(b)

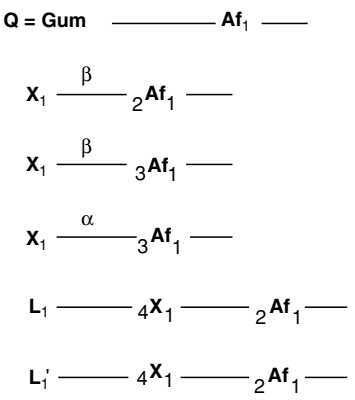
Q = Gum
Af1
X1
2Af1
X1
3Af1
X1
3Af1
L1
4X1
2Af1
L1'
4X1
2Af1

(c)

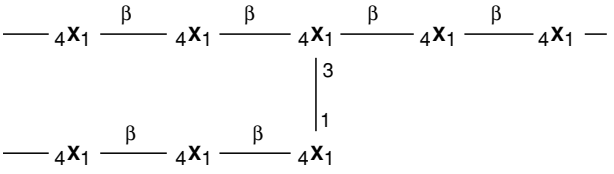

(d)

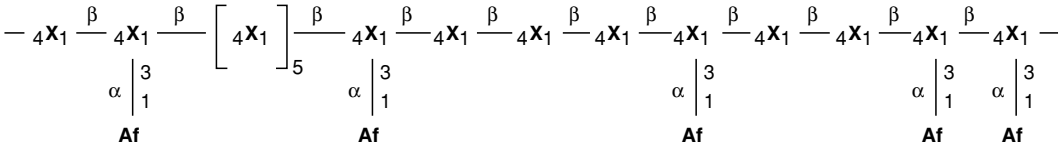
Af
Af
Af
Af
Af

(e)

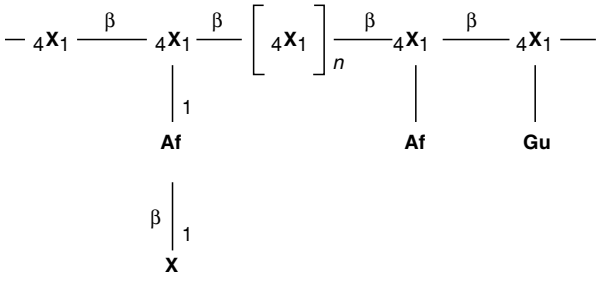
Af
Af
Gu
X

(f)

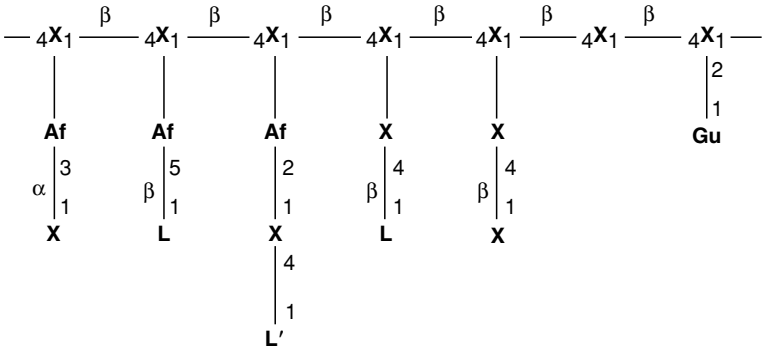

(g)

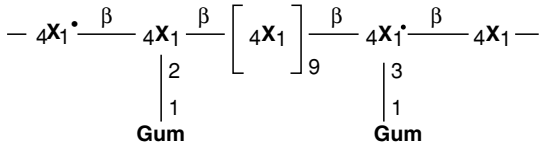

(h)

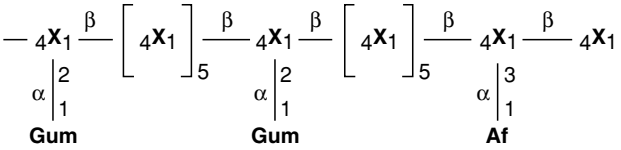

(i)

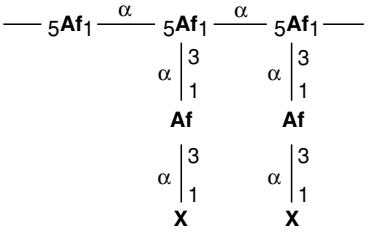

(j)

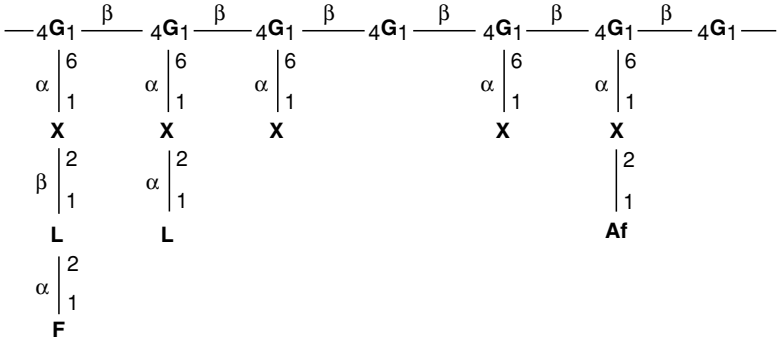

Figure 2 Xylan structures: (a) polymer backbone; (b) general structure; (c) esparto grass; (d) wheat flour; (e) barley husk; (f) maize seed coat; (g) hardwood; (h) softwood; (i) cress seed; (j) runner bean. Reproduced from Hemicelluloses, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

every sixth xylose, and it also has arabinose side groups.

Some unusual xylose-containing hemicelluloses have nonxylose backbones. Cress seed hemicellulose (Figure 2i) has an arabinose backbone with attached xyloarabinose side-groups. Runner bean hemicellulose (Figure 2j) has a glucose backbone similar to cellulose with side groups containing xylose, galactose, arabinose, and fucose. (*See* **Legumes**: Legumes in the Diet.)

Mannans

Mannans generally occur with other sugars, including galactose and glucose. Four distinct groups (**Table 1**) are generally recognized depending on the relative amounts of the additional sugars. In some mannans, particularly from softwoods, the sugars are acetylated.

Mannan (Figure 3a) is a relatively pure (> 95%) mannose polymer found in the food reserves of some seeds. The most famous mannan source is vegetable ivory (i.e., tagua palm seed), a hard crystalline material that may be fashioned into buttons. It is not completely pure, because traces of galactose have been isolated with the mannose. Mannan also occurs in coffee beans and orchid tubers.

Glucomannan is a heteropolymer of random-sequence mannose and glucose. Hardwood glucomannan (Figure 3b) has no side groups. Its mannose-to-glucose ratio is 2:1 (common) to 1:1 (infrequent).

Softwood glucomannan (Figure 3c) has side groups with a mannose-to-glucose ratio of 3:1 (common) to 4:1 (infrequent). A galactose residue is attached to approximately every 15th or 30th mannose unit at C6. Some galactose may also be attached to glucose at C3. Approximately 25% of the backbone sugars are acetylated. Glucomannans also occur in the food reserves of some plants, such as iris seeds (mannose-to-glucose ratio 1:1), lily bulbs, bluebell seeds, and orchids.

Table 1 Sugar and acetyl ratios in various mannans

	Mannose	*Glucose*	*Galactose*	*Acetyl*
Mannan	> 19		1	
Glucomannan				
Hardwood	1–2[a]	1		
Softwood	3[a]– 4	1	0·1–0·2	1
Iris	1	1		
Galactomannan				
Legume	1–5	1		
Galactoglucomannan				
Softwood	3	1	1	1

[a]More common.

Galactomannan is common in the seeds of legumes and carob trees. The most famous form is the food thickener guaran (or guar gum), shown in Figure 3d. Galactose side groups are attached to every other mannose unit. (*See* **Gums**: Properties of Individual Gums.)

Galactoglucomannan is similar to glucomannan, except it has about 10 times more galactose and a lower DP. The backbone from western hemlock galactoglucomannan (Figure 3e) has randomly ordered mannose and glucose. Galactose side groups are attached to either mannose or glucose. Approximately 25% of the backbone is acetylated.

Galactan

Figure 4a shows the β-1,3-linked galactose backbone of galactan. Arabinose is a common side group, so the term 'arabinogalactan' is sometimes used. Larchwood has an unusually high galactan content (10–15% typical, but as high as 25–30%), so it is well studied. Larch galactan (Figure 4b) has side groups of galactose, arabinose, and small amounts of glucuronic acid. The arabinose-to-galactose ratio ranges from 1:4 to 1:8, with 1:6 most common. Maritime pine galactan (Figure 4c) contains a small amount of xylose. Tobacco (Figure 4d) and maple sap galactans contain some rhamnose.

Cellular Structure

Hemicellulose is a major component of plant cell walls, which are well described in the **Cellulose** article. Figure 5 shows the distribution of lignin, cellulose, and various hemicelluloses in the cell wall layers for both softwoods and hardwoods. (The distribution in herbaceous crops would be similar to hardwoods.) Hemicellulose is the dominant carbohydrate in the compound middle lamella, whereas cellulose is more plentiful in the secondary layers (S1, S2, and S3). In hardwoods (and herbaceous crops), glucuronoxylan is the primary hemicellulose, whereas glucomannan is more prevalent in softwood.

Properties

Physical

Because hemicellulose is so heterogeneous, its physical properties are not well studied. The higher heat of combustion for xylan is 17.8 MJ kg^{-1}. Hemicellulose begins to decompose above 200 °C, which makes it one of the least thermally stable components of the cell wall.

In the cell wall, most hemicellulose is noncrystalline. (An exception to this is the mannan from

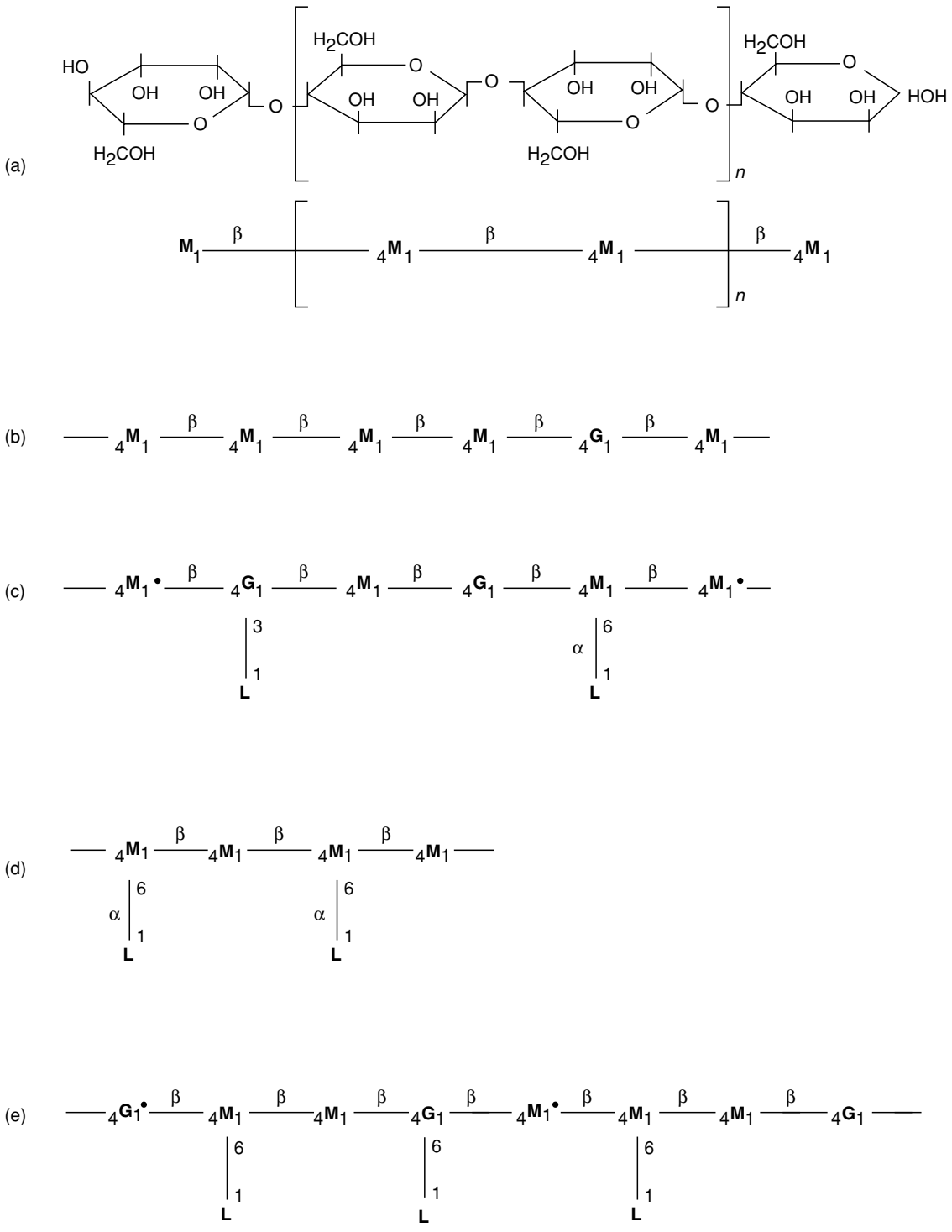

Figure 3 Mannan structures: (a) ivory nut mannan; (b) hardwood glucomannan; (c) softwood glucomannan; (d) guaran; (e) western hemlock. Reproduced from Hemicelluloses, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

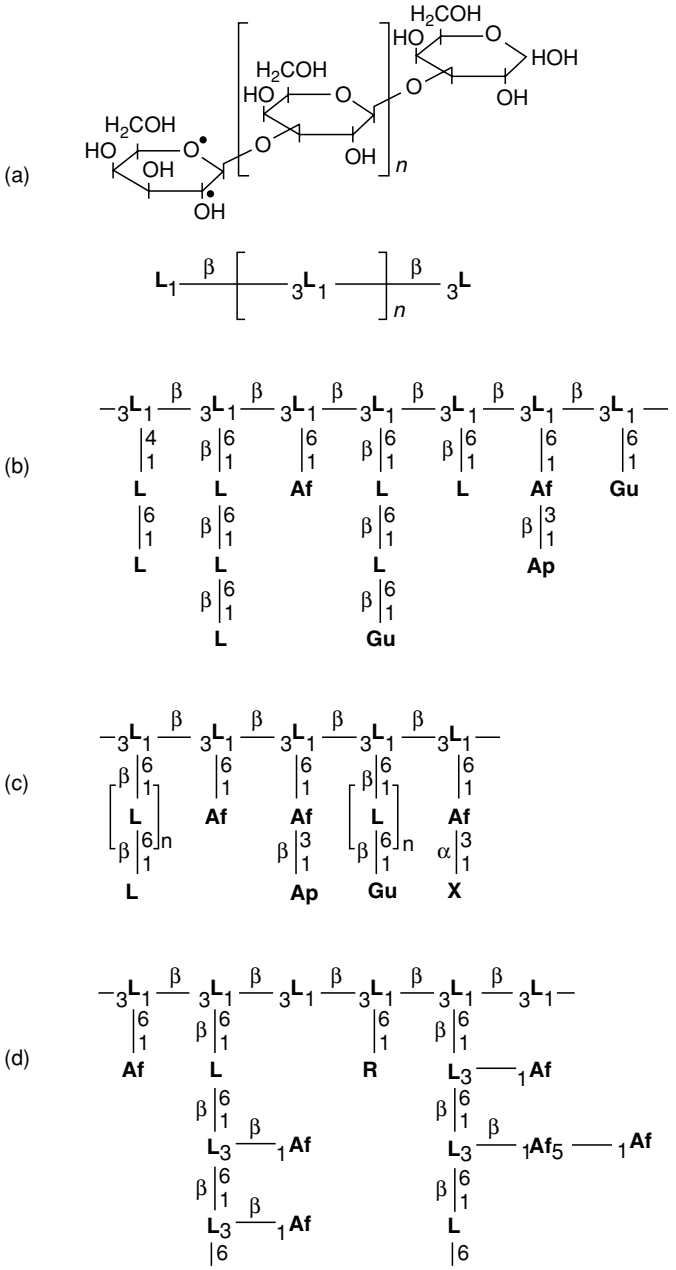

Figure 4 Galactan structures: (a) galactan backbone; (b) larch; (c) maritime pine; (d) tobacco. Reproduced from Hemicelluloses, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

ivory nuts.) Isolated 4-*O*-methylglucuronoxylan is crystalline with a repeating length of 1.48 nm. Adjacent xyloses are rotated 120° with respect to their neighbors in a left-handed helix.

Chemical

Only a few hemicelluloses are water-soluble, such as galactan (from larch) and galactomannan (from guar gum). Most hemicelluloses are soluble in dilute ($10 \pm 8\%$) sodium hydroxide, a property used to define hemicellulose. Alkaline extraction deacetylates hemicellulose, which can be avoided by using dimethyl sulphoxide. The glucomannan *cis*-hydroxyl groups (C2 and C3) impede alkaline solubility, so sodium borate is sometimes added to improve dissolution. A fraction of ivory nut mannan (mannan B) forms microfibrils similar to cellulose. It is insoluble in alkali but, like cellulose, it is soluble in cuprammonium hydroxide solutions.

Dilute acid readily hydrolyzes hemicellulose. The xylose released from xylan degrades to furfural, which further degrades to resins if the hydrolysis conditions are severe.

Alkaline degradation occurs from the polymer 'reducing end', i.e., the end with the free (or unlinked) C1 hydroxyl group. In xylan, successive terminal xyloses 'peel off' and degrade to the saccharinic acid, 2-hydroxymethyl-2,4-dihydroxybutyric acid. The reaction proceeds more readily in calcium hydroxide than sodium hydroxide. This reaction occurs slowly at the usual cold extraction temperatures. At higher temperatures, 4-*O*-methylglucuronic acid loses both its methyl and uronic acid functional groups. Alkaline degradation of hemicellulose can be limited by prereducing with borohydride.

Hemicellulose is much more resistant to oxidation than lignin – a fact exploited in its isolation from natural materials as 'holocellulose.' However, slight changes do occur. Preparing holocellulose using chlorous acid (acidified sodium chlorite) causes depolymerization, oxidation of 2,3-hydroxyl groups, and oxidization of reducing ends to aldonic acids.

Occurrence

A detailed list of the hemicellulose content in various foods and agricultural products is presented in the **Cellulose** article.

Table 2 lists the sugar distribution in various plants determined by acid hydrolyzing the polysaccharides and chromatographically measuring the resulting sugars. Most of the glucose is derived from cellulose, although some comes from hemicellulose mannans. The softwoods have a much higher mannose content because much of their hemicellulose contains mannans rather than xylans. The hemicellulose composition of hardwoods and agricultural residues is similar.

Uses

Furfural can be made by heating xylan in the presence of 12% hydrochloric or sulfuric acid. Oat hulls and corn cobs are traditional sources. Furfural may be used as a solvent in petroleum refining, in manufacturing furfural–phenol plastics (Durite), as a solvent for cellulose nitrate and cellulose acetate, in the manufacture of insecticides, and as a nylon precursor. (*See* **Oats**.)

Xylitol is a sugar alcohol formed by reducing xylose. It is as sweet as sucrose, but is not cariogenic. Its endothermic heat of solution elicits a cool sensation to the mouth, so it has been incorporated into chewing gum. Because it is completely metabolized, it is not a low-calorie sweetener. It has no insulin requirement and has been used for intravenous infusion. (*See* **Sugar Alcohols**.)

Xylose is used as a media ingredient to produce xylose/glucose isomerase which is used in the production of high-fructose corn syrups.

Table 2 Composition of some hemicellulose-containing materials (g per 100 g dry matter)

		Hexosans			*Pentosans*	
		Cellulose	*Hemicellulose*			
	Lignin	*Anhydroglucose*	*Anhydromannose*	*Anhydrogalactose*	*Anhydroxylose*	*Anhydroarabinose*
Softwood[a]						
Balsam fir	29·4	46.8	12.4	1.0	4.8	0.5
Jack pine	28·6	45·6	10·6	1·4	7·1	1·4
White spruce	27·1	46·5	11·6	1·2	6·8	1·6
Hardwood[a]						
Aspen	16·3	57·3	2·3	0·8	16·0	0·4
White birch	18·9	44·7	1·5	0·6	24·6	0·5
Red maple	24·0	46·6	3·5	0·6	17·3	0·5
Agricultural residues[b]						
Barley straw	13·8	37·5	1·26	1·71	15.0	4.0
Corn stover	15·1	35·1	0·25	0·75	13·0	2·8
Cotton gin trash	17·6	18·0	1·9	0·1	4·0	2·0
Rice hulls	19·4	32·5	2·7	0·1	12·3	2·6
Rice straw	9·9	36·9	1·6	0·4	13·0	4·0
Sorghum straw	14·5	32·5	0·8	0·2	15·0	3·0
Wheat straw	14·5	32·9	0·72	2·16	16·9	2·1

[a]Timell TE (1957) Carbohydrate composition of ten North American species of wood. *TAPPI* **40**: 568–572.
[b]Totals are less than 100% since they do not include extractives, ash, protein, or free sugars. From Wiley DF (1985) *Enzymatic hydrolysis of cellulose*. PhD dissertation, University of California.

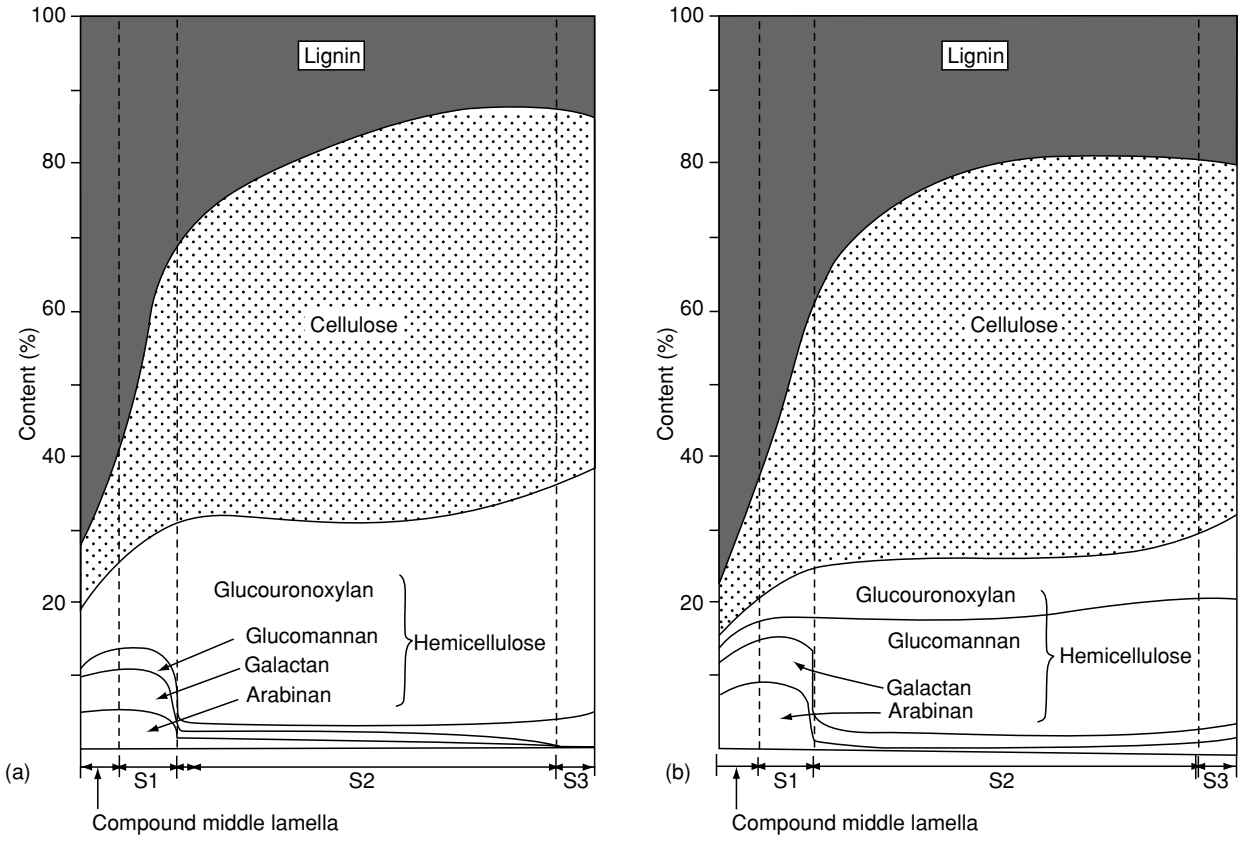

Figure 5 Distribution of lignin, cellulose, and hemicellulose in (a) hardwood and (b) softwood. Data sources: (1) Meier H (1964) In: Zimmerman MH (ed.) *The Formation of Wood in Forest Trees*. New York: Academic Press; (2) Panshin AJ, de Zeeuw C (1970) *Textbook of Wood Technology*, 3rd edn, vol. 1. New York: McGraw-Hill. Reproduced from Hemicelluloses, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Carboxymethylxylan has been prepared with properties similar to carboxymethylcellulose. It has potential uses in detergents, pigment dispersants, or paper coatings. Other hemicellulose derivatives, analogous to cellulose derivatives, include acetates, butyrates, higher-fatty-acid esters, benzoates, and xanthates. The acetates may be used to form films similar to cellophane.

Bread freshness can be improved up to three times by adding hemicellulose to dough. It also improves water-binding capacity, mixing quality, mixing energy efficiency, and loaf volume. (*See* **Bread**: Chemistry of Baking.)

Dietary fiber contains hemicellulose. Approximately 40–60% is digested by microbes in human large intestines. (*See* **Dietary Fiber**: Properties and Sources.)

Hemicellulose Isolation

Holocellulose is prepared by defatting the fiber with hot azetropic ethanol/benzene and delignifying with hot (70–75 °C) chlorous acid ($HClO_2$). Alternatively, it may be treated with chlorine gas and alcohol pyridine, or chlorine gas and 95% alcohol containing 3% monoethanolamine. If the sample contains a starch, it must be heated in water to gelatinize the starch, with subsequent hydrolysis using α-amylase. After washing, the remaining material (holocellulose) contains cellulose and hemicellulose.

Hemicellulose is removed from holocellulose by extracting with alkali (2–18% sodium hydroxide, although 10% is common). Neutralization of the extract precipitates 'hemicellulose A,' which is composed of linear (or nearly linear), high DP hemicellulose without acidic carboxyl groups. 'Hemicellulose B' remains in solution and tends to be more branched with lower DP and more acidic carboxyl groups. Hemicellulose B can be fractionated by gradually adding ethanol and collecting the various components as they precipitate.

An alternate approach to fractionating hemicellulose from holocellulose is to exploit differences in alkali solubilities. For example, 24% potassium

hydroxide extracts holocellulose xylan and galactoglucomannan from both hardwoods and softwoods, but does not extract glucomannan from the residue. The dissolved galactoglucomannan may be precipitated from the extract by adding barium hydroxide. The insoluble glucomannan remaining in the residue can be extracted using 17·5% sodium hydroxide/4% borate with subsequent precipitation by barium hydroxide addition.

Acidic hemicelluloses can be precipitated and isolated from neutral hemicelluloses by using quaternary ammonium salts (e.g., cetyltrimethylammonium bromide) under neutral or mildly alkaline conditions. Then, the neutral hemicelluloses can be precipitated from solution by increasing the alkalinity in the presence of the quaternary ammonium salts.

Hemicellulose Analysis

Gravimetric Method

The **Cellulose** article describes the Van Soest method for preparing acid detergent fiber (ADF), which contains cellulose, lignin, and insoluble ash (mainly silica). This article describes the procedure for making neutral detergent fiber (NDF), which contains hemicellulose, cellulose, lignin, and insoluble ash. The difference between NDF and ADF is a measure of hemicellulose content.

NDF is prepared by placing in a beaker about 1 g of sample (ground to pass a 30-mesh screen), 100 ml of neutral-detergent solution (30 g l^{-1} of sodium lauryl sulfate 18.61 g l^{-1} of disodium ethylenediaminetetraacetatic acid (EDTA) dihydrate, 6.81 g l^{-1} of sodium borate decahydrate, 4.56 g l^{-1} of anhydrous disodium hydrogen phosphate, 10 ml l^{-1} of 2-ethoxyethanol), 2 ml of decahydronaphthalene and 0.5 g of sodium sulfite. The mixture is boiled for 60 min while refluxing to keep the liquid volume constant and then filtered on a previously tared Gooch crucible. After being rinsed twice with boiling water and twice with acetone, the sample is dried overnight at 100 °C and weighed. (If the sample contains starch, it is incubated with α-amylase overnight before washing with water.)

Colorimetric Method

The Technical Association of the Pulp and Paper Institute (TAPPI) Method T 223 os-78 describes a colorimetric method for measuring pentosans. Because pentose sugars are dominant in most hemicelluloses (except for softwood), this procedure indicates hemicellulose content. The sample is boiled in 3.85 mol l^{-1} hydrochloric acid, causing the pentoses to degrade to furfural which is collected in the distillate. The collected furfural is reacted with orcinol which forms a colored compound measured spectrophotometrically. Because of the complexity of the apparatus, only a few samples can be run daily. Holtzapple describes a colorimetric pentosan assay which is performed in a sealed test tube, which speeds up the assay. (*See* **Spectroscopy**: Visible Spectroscopy and Colorimetry.)

Chromatographic Method

TAPPI Method T 249 pm-75 describes a chromatographic technique for measuring mannan, arabinan, xylan, galactan, and glucan. (In starch-free materials, cellulose is the dominant glucan, although some glucan is also present in hemicellulose.) The sample is hydrolyzed using sulfuric acid in a two-step procedure. Some of the hydrolysate is neutralized and the sugars are reduced to alditols (i.e., sugar alcohols) using sodium borohydride. The alditols are acetylated with acetic anhydride and sulfuric acid. The alditol acetates are precipitated in ice water and extracted with methylene chloride for injection into a gas chromatograph. Because sugars degrade under the acid hydrolysis conditions, internal standards of the free sugars are run under identical treatment conditions. This method is extremely time-consuming and is used only when complete characterization of polysaccharides is required. (*See* **Chromatography**: Gas Chromatography.)

See also: **Bread**: Chemistry of Baking; **Carbohydrates**: Classification and Properties; **Cellulose**; **Chromatography**: Gas Chromatography; **Dietary Fiber**: Properties and Sources; **Galactose**; **Gums**: Properties of Individual Gums; **Legumes**: Legumes in the Diet; **Lignin**; **Oats**; **Spectroscopy**: Visible Spectroscopy and Colorimetry; **Sugar Alcohols**

Further Reading

Aspinall GO (1970) *Polysaccharides*, pp. 85–115. New York: Pergamon Press.

Aspinall GO (1982) Analysis of polysaccharides. In: Linebach DR and Inglett GE (eds) *Food Carbohydrates*, pp. 356–374. Westport: AVI.

Dintzis FR (1982) Dietary fiber analysis – concepts and problems. In: Lineback DR and Inglett GE (eds) *Food Carbohydrates*, pp. 312–332. Westport: AVI.

Holtzapple MT and Humphrey AE (1983) Determination of soluble and insoluble pentoses in the presence of glucose. *TAPPI Journal* 66: 90.

Kennedy JF (1988) *Carbohydrate Chemistry*, pp. 229–233. Oxford: Clarendon Press.

Whistler RL and Shah RN (1978) Recent developments in the industrial uses of hemicelluloses. In: Rowell RM and Young RA (eds) *Modified Cellulosics*, pp. 341–356, New York: Academic Press.

Whistler RL and Richards EL (1970) Hemicelluloses. In: Pigman W and Horton D (eds) *The Carbohydrates: Chemistry and Biochemistry*, pp. 447–469. New York: Academic Press.

Whistler RL and Zysk JR (1978) Carbohydrates. In: *Kirk-Othmer Encyclopedia of Chemical Technology*, 3rd edn, vol. 4. New York: Wiley.

Herbicides *See* **Pesticides and Herbicides**: Types of Pesticide; Types, Uses, and Determination of Herbicides; Residue Determination; Toxicology

HERBS

Contents

Herbs and Their Uses

K P Svoboda and T Svoboda, The Scottish Agricultural College, Auchincruive, Ayr, UK

Background

The plant kingdom produces a wide range of valuable products. These include numerous aromatic, dye, spice, medicinal, and other useful substances. Such products have contributed to the quality of human life for thousands of years. From ancient texts and scriptures, it is known that some plants have long served as the principal components of seasonings, flavorings, cosmetics, dyes, and medicines.

Herbs generally include plants whose leaves, stems, flowering tops, and roots are used fresh, dried, or frozen in the seasoning of foods, in the beverage industry, in perfumery and cosmetic products, in the pharmaceutical industry, and for numerous other purposes, including ornamental gardening, pot-pourri, moth and general insect deterrents and in dyeing and fabric cleaning. There is no standard definition of the term 'culinary herbs,' and the distinction between culinary herbs and spices is imprecise. The American Spice Association, the most important body in the spice trade, uses the term 'spices' in such a way as to include both herbs and aromatic seeds. Scientific information on herbs, plant drugs, their active constituents, and their therapeutic and health uses is being published worldwide at an unprecedented rate, in many different journals and in various languages. Scientific and regulatory information is systematically presented in various monographs. Recommendations for new crops which can be used as herbs are given for specialty crop growers. In addition, there is growing interest in organically produced herbs, and such herbs have to conform to organic certification standards. Many herbs are now also being used as dietary supplements and ingredients in foods.

Sources of Herbs

Herb material is obtained from both wild and cultivated plants. In most cases, the demand for culinary herbs is such that collection of material from the wild is insufficient to satisfy market requirements. Cultivation of herbs offers numerous advantages, including:

- the development of high-yielding cultivars, which also leads to a product with more consistent properties than would be obtained from plants growing wild;
- the use of mechanization;
- the control of insects and diseases;
- planned harvesting and transport; and
- efficient drying and storage.

Herb production can be undertaken using either labor-intensive techniques, as for example in north Africa and India, or capital-intensive harvesting and processing methods as in the USA, France, and other Western countries.

The quality of herbs is dependent on several factors, including plant variety, growing conditions, cultural practices, drying and processing methods, storage, and packing. Cleanliness, color, flavor, and aroma are the most important considerations in the selection and utilization of herbs. National quality standards have not yet been established for culinary herbs, but as the market grows, the need for specific quality standards is being recognized. The description and chemical composition of a number of herbs are defined in the *British Pharmacopoeia, British Pharmaceutical Codex, British Herbal Pharmacopoeia, New Cyclopedia of Botanical Drugs and Preparations, British Herbal Compendium, ESCOP Monographs in Herbs or Plant Drugs, WHO Herbal Monographs, and German Commission E Monographs on Herbs*.

Production and Consumption

Accurate data on the area of individual herbs grown and their production in any country are very difficult to obtain. An unofficial estimate of total UK herb production in 1999 was about 3000 ha. The UK growing season is usually from April to November, depending on the herb and the weather in any particular year. Outside the season, fresh herbs are imported largely from Mediterranean countries, especially Israel. A significant proportion of marketed herbs are cultivated on smallholdings or collected from the wild, and do not enter the trade statistics. Most sources agree that the total national consumption of herbs is increasing at around 10% per annum. Her Majesty's Customs and the Ministry for Agriculture are the most important organizations for reporting on the trade and production of herbs. Of the various culinary herbs, they report separately only on thyme and parsley, with other herbs being generally categorised under 'other species' or 'dried vegetables'.

Table 1 shows the estimated annual consumption of the major culinary herbs and volatile oils in the UK. These statistics must be regarded as providing orders of magnitude rather than precise measures. Since commercial herb growing is a relatively small part of horticulture, insufficient research has been carried out to permit the use of more than a narrow range of pesticides, most of them with limited off-label use, which causes some difficulty in growing some crops on a large scale. Traditionally, herbs are hand-harvested, but there are now several purpose-built harvesting machines available, and many people have adapted conventional cereal- and grass-cutting and harvesting equipment for use on herbs, the aim being to cut and immediately lift the crop without bruising. The trade estimates that about 50% of the plants traded are used for the food industry, 25% in cosmetics, 20% for medicinal and 5% for other uses such as insecticides.

Herbs and Volatile Oils

There are 31 plant families made up of 93 species representing the major flavor- and spice-producing taxa. Approximately 10% of the world's major economic plants are sources of flavorings. These may be classified into six main groups, according to their principal flavor components (Table 2).

The volatile (essential) oils consist of a complex mixture of chemical compounds, which combine to give the oil distinctive characteristics. The differences in aroma among plant types result from differences in the volatility and quantity of chemicals in the volatile oil. The volatile oils are products of the plant's metabolic processes and are particularly subject to seasonal and climatic changes, and to the care taken in general husbandry, harvesting, and storage. Volatile oils, oleoresins and oleo-gum resins are produced in various types of secretory tissues. The type of

Table 1 Estimated UK production and imports of selected culinary herbs

Herb	*Area (ha)*	*Yield (dry matter) ($t\,ha^{-1}$)*	*Imports (t)*
Basil	5	1.7	30–50
Marjoram	4	1.9	40–75
Mint	100	3.6	80–175
Oregano	4	1.6	40–55
Parsley	500	3.0	100–250
Rosemary	4	2.5	20–75
Sage	100	1.9	100–500
Savory	4	1.3	10–12
French tarragon	40	0.8	80–250
Thyme	40	0.8	80–250

Table 2 Flavor components of several important commercial herbs

Group	*Flavor component*	*Herb*
1	Cineole or eucalyptol	Bay (*Laurus nobilis*)
		Spanish sage (*Salvia lavandulifolia*)
		Rosemary (*Rosmarinus officinalis*)
2	Eugenol	West Indian bay (*Pimenta racemosa*)
3	Thymol or carvacrol	Origanum (*Origanum vulgare*)
		Wild marjoram (*Origanum marjoranum*)
		Summer savory (*Satureja hortensis*)
		Mexican oregano (*Lippia graveolens*)
		Thyme (*Thymus* spp.)
		Basil (*Ocimum basilicum*)
4	Linalool, methyl chavicol, or terpineol	Sweet marjoram (*Marjorana hortensis*)
		Parsley (*Carum petroselinum*)
		English sage (*Salvia officinalis*)
		French Tarragon (*Artemisia dracunculus*)
5	Thujone	Dalmatian sage (*Salvia officinalis*)
6	Menthol, menthone, or carvone	Mints (*Mentha* spp.)

secretory structure is an important character of a plant family and includes various forms, such as oil and oleoresin cells (ginger, pepper, vanilla), secretory glands (bay, clove, eucalyptus), secretory ducts (tarragon, anise, angelica) and glandular trichomes (mint, sage, lavender, rosemary, chamomile) (Figures 1–4). These structures are associated with many different parts of the plant, such as vascular bundles in the leaves, stems and roots, fruit wall, bark, parenchyma, or epidermal tissue. Leafy herbs are normally used in a well-broken state. Following drying and milling, the finer the powder, the more readily available is the flavor.

Three major types of industrial distillation are used to obtain the volatile oil fraction: water, water and steam, and steam. In water distillation, the plant material is placed in a distilling container with water and heated until the volatile compounds have vaporized. The gaseous mixture of oils and water is piped to a condenser for cooling: the oil, being lighter than water, floats to the top and can be collected. Volatile oils used by the perfumery industry are extracted by the use of organic solvents. Although more expensive than distillation, it avoids elevated temperatures, which may cause the loss of highly volatile components and the denaturation of other chemicals in the oil. The solvent system uses petroleum ether or benzene. In a technique known as enfleurage, the plant material is affixed to a thin layer of odorless fat spread on a glass sheet. The volatile oils are absorbed into the fat and can then be extracted with alcohol.

A more recent development in extraction technology, still at the research and development stage, is the use of supercritical carbon dioxide. Liquid carbon dioxide is used at a range of temperatures and elevated pressure. Under these conditions, carbon dioxide behaves like a solvent and removes different classes of compounds from the plant material depending on the running conditions in force. Moderate pressure (< 50 bar) releases volatile oils from plant material, while at a greater pressure (> 200 bar), waxes and resins predominate in the resultant extracts. By careful manipulation of the operational parameters, it is possible to sequentially remove these different groups of compounds from the raw material. It is also possible to deterpinate certain oils, making them less pungent in taste.

Figure 1 (see color plate 89) Chamomile field in Norfolk.

Volatile oils have long been used in industry in place of the dried herbs. Their advantages over the

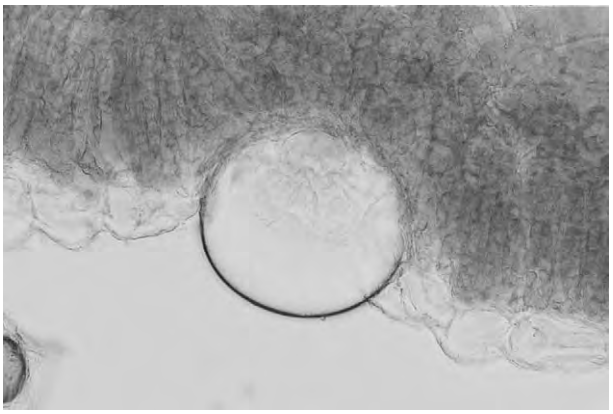

Figure 2 (see color plate 90) Volatile oil gland on the surface of a Greek oregano (*Origanum heracleoticum*) leaf.

plant material include a generally uniform flavor quality and their sterility – they are free from all extraneous contaminants. Oils can be readily incorporated into liquid concentrates and emulsions, the flavor character and strength of which can be standardized. They can be readily mixed with fats and oils which are commonly present in food products and are easily dispersed throughout the mix. Being concentrated, these oils save both storage space and handling costs. They are normally stable and show only a slight deterioration over long periods provided that they are correctly stored.

Cleanliness and Preservation

Spice and herb plants tend to be heavily contaminated with microorganisms and molds, many in spore form. The general negligent and unhygenic way in which they are handled from the time of harvesting until they are eventually presented for sale results in the presence of such contaminants such as rodent hairs and droppings, insect parts, eggs and larvae, hessian sacking fibers, soil, and stones. Cleanliness is regarded as being important by the producer, processor, and consumer, and has to be achieved without detriment to the product itself. In the UK, following the banning of ethylene oxide gas at the end of 1990 as the principle cleaning agent for dried herbs, firms have devised several different methods of reducing the microbial load. In 1991, a licence was granted for including ionizing radiation treatment of dried herbs and spices to a maximum load of 10 kGy. Pesticide contamination and heavy metal content are becoming important issues on the quality market. Some European countries permit irradiation but the laws in the US are different.

Fresh herbs are supplied prewashed, using a sterilized agent such as dilute sodium hypochlorite; frozen herbs are similarly treated, chopped, and then ideally quick-frozen using liquid nitrogen. Freeze-dried herbs go a stage beyond frozen by being subsequently treated at a low temperature in a vacuum, which drives moisture from the herbs without any browning or shrinking, a feature which may occur with conventionally dried herbs. The final product may be stored indefinitely without refrigeration, although packaging requirements are stringent. The process is costly

Figure 3 (see color plate 91) *Agastache* species (Korean mint) at the Auchincruive herb garden.

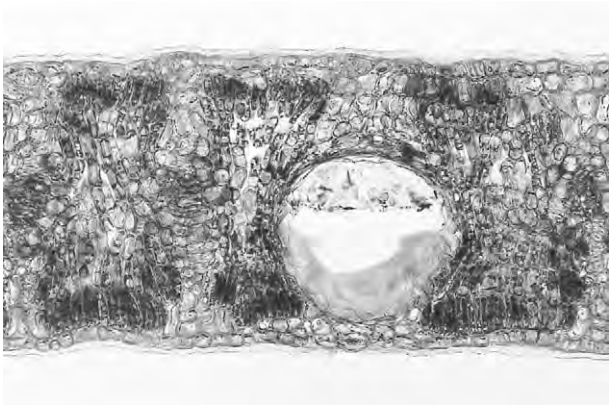

Figure 4 (see color plate 92) Cross-section through a eucalyptus leaf, showing the oil cavity.

and is presently used in commerce to preserve highly degradable substances, such as chopped chives, for which other satisfactory methods of drying are inappropriate. Freeze-drying seems to be better for volatile substances or vitamins but offers no apparent advantage or preservation of certain other substances, such as alkaloids, and may promote the oxidation of fatty acids, especially linoleic and linolenic acid.

Food Industry

Commercially Important Herbs

Commercially important herbs on the UK market include basil, celery, chervil, chives, coriander, dill, marjoram, mint, oregano, parsley, rosemary, sage, savory, tarragon, and thyme. However, many other species are used in cooking, and their range is expanding every year. Mixed herbs usually contain three or four herbs selected from thyme, savory, sage, marjoram, basil, parsley, and rosemary.

Uses of Herbs

Herbs are being incorporated into a variety of different food products: butters, cheeses, biscuits, pastries, popcorn, sausages, salami, meats and stews, fish dishes, soups and sauces, mustards and vinegars, chocolates, and sweets. Herbs can be used fresh, dried, broken, or ground. Alternatively, volatile oils or extracts are readily available on the market. A number of companies have now introduced dehydrated or frozen herbs. Frozen flakes are free-flowing, can be used without any previous defrosting, and have a fresh aroma that is stable for 18 months.

Packaging and Storage

Dried herbs are packaged for the retail trade in a wide variety of small- to medium-sized packs, including glass jars, drums, tubs, cellophane, polythenes, and cartons. Freeze-dried herbs tend to be packed in polyethylene terephthalate. If stored in tubs, the herbs must be free of oxygen and water vapor in airtight, opaque containers and, in the case of freeze-dried material, filled with a dry inert gas. Fresh herbs are usually packed in polystyrene trays with a clear overwrap or in transparent punnets with lids. The wholesale market also accepts polythene bags, open crates, and cartons.

Fresh herbs need to be transported and stored chilled, between 2 and 5 °C, or slightly warmer for more sensitive herbs such as basil. Frozen herbs need to be stored and shipped at temperatures of at least − 18 °C, giving a shelf-life of 12 months, which can be extended to 2 years at − 25 °C. Dried and freeze-dried herbs are stored and transported at ambient temperature, preferably out of direct light.

Pharmaceutical Industry

Although synthetic and semisynthetic organic substances have achieved a substantial share in pharmaceutical applications, plant-derived substances still remain a vital ingredient in modern medicine. As estimated 25% of prescriptions contain one or more active constituents obtained from higher plants. The herbal ingredients are obtained from some 550 different herbs. It has been estimated that about 7000 t of herbs are extracted annually for the ingredients of UK medicines. Some particularly potent plants, such as *Areca catechu, Aconitum napellus, Atropa belladonna*, and others, have been brought under strict pharmaceutical control. Many plant products available to the public are capable of causing serious harm. These include products which can cause allergenic reactions in previously sensitized individuals, preparations which induce gastroenteritis, and products with cardiovascular involvements. Also, many can interfere with anticoagulant treatment due to the presence of substances such as coumarins. A further problem arises from the misidentification of plant products and from contamination of beneficial plant material with toxic species. If the public is to be protected from potentially hazardous plant products available for sale, herbal suppliers must be monitored carefully. Chemical diversity in natural products is a rich source of new pharmaceuticals and other economically important chemicals. A number of new bioassays have been designed to facilitate the isolation of natural products from plants, rapid screening of several thousand samples in one day, and complicated synthesis of complex products based on a natural template. The development of natural product chemistry and the discovery of new natural products from indigenous medicinal plants are becoming increasingly prominent features due to their medicinal, food, and other industrial values.

Product Development and the Future

The herb market has experienced significant growth in recent years, and a considerable amount of product development has taken place, with much input from supermarkets. Despite the effects of economic recession in the early 1990s, this trend is set to continue under the influence of travel abroad, TV cookery programs, and ethnic restaurants. New methods of production and packaging continue to evolve,

offering greater quality and convenience to consumer, caterer, and manufacturer alike. Current developments include new harvesting techniques, improved washing chemicals, and new forms of presentation, such as pelletized frozen herbs. These factors, combined with the demand for more natural flavorings and innovative ideas from all areas of the industry, should ensure a steady growth in the herb market throughout the next decade and beyond.

Research and Development

Research into the biological activity of plant volatile oils has been undertaken for several decades. However, the most interesting results had to await the development of sophisticated analytical instruments such as gas chromatography linked to mass spectroscopy, high-performance liquid chromatography, Fourier-transformed infrared spectroscopy, and Fourier-transformed nuclear magnetic resonance. With such instrumentation available, it is now possible to identify the majority of the chemical constituents in the oils and thus allocate bioactivity to specific compounds. The possibility of synergistic effects should not be overlooked, although the evaluation of such phenomena becomes problematic when many oils have over 100 different constituents. The marked antibacterial and antifungal activities in certain volatile oils provide the food and related industries with a source of natural preservatives. There is some reluctance on the part of the industry to exploit such antimicrobials on a larger scale, owing to the problem of balancing effective antimicrobial activity without rendering food unacceptable because of an overwhelming taste of plant extract. Deterpenation using supercritical carbon dioxide may help to alleviate this aspect.

A number of plant secondary metabolites possess antioxidant compounds in the volatile oil and other fractions. There is now evidence that plant antioxidants taken as diet supplements have a marked beneficial effect upon the lipid metabolism in animal model systems. The long-chained polyunsaturated fatty acids (arachidonic and docosohexaeonic) are well protected by such antioxidants, leading to the reversal of the normal decline in concentration during the ageing process. Many degenerative disease conditions associated with human aging, such as senile dementia and acute memory loss, can be described in terms of lipid inadequacies. Hence, any natural antioxidants capable of slowing down the disease process are of real significance in human health and the quality of life.

Recently, it has become clear that many plant metabolites have potential beneficial effects directly relevant to human diet and health. Some may exert a protective role which reflects similar protective actions like antioxidants, some can prevent cells from mutation (blocking and suppressing anticarcinogens), and certain herbs contain chemicals (isocoumarins, flavonoids, hop acids, and terpenes) which are structural analogs of the natural estrogens. These compounds may interact with the estrogen receptors in human cells. Herbs, by their nature and quality, have the potential to influence both short-term and long-term human health and well-being. An increased level of certain protective factors may be desirable, and both classical plant breeding and biotechnology can provide the improved plant material for human consumption.

See also: **Essential Oils**: Properties and Uses; Isolation and Production; **Herbs**: Herbs of the Compositae; Herbs of the Labiatae; Herbs of the Umbelliferae; Herbs Used in Alcoholic Drinks; **Spices and Flavoring (Flavouring) Crops**: Use of Spices in the Food Industry; Fruits and Seeds; Leaf and Floral Structures; Tubers and Roots; Properties and Analysis

Further Reading

Atta-ur-Rahman and Iqbal Choudhary M (eds) (1998) *New Trends in Natural Product Chemistry*. Karachi, Pakistan: University of Karachi.

Bauer K, Garbe D and Surburg H (1997) *Common Fragrance and Flavour Materials. Preparation, Properties and Uses*, 3rd edn. Weinheim, Germany: Wiley-VCH.

Bown D (1995) *The Royal Horticultural Society: Encyclopedia of Herbs and Their Uses*. London: Dorling Kindersley.

Bradley PR (ed.) (1992) *British Herbal Compendium*, vol. 1. Guildford and King's Lynn, UK: Biddles Ltd.

Deans SG and Ritchie GA (1987) Antibacterial properties of plant essential oils. *International Journal of Food Microbiology* 5: 165–180.

Deans SG, Svoboda KP and Bartlett MC (1991) Effect of microwave oven and warm-air oven drying on the microflora and volatile oil profile of culinary herbs. *Journal of Essential Oil Research* 3: 341–347.

Dewick PM (1998) *Medicinal Natural Products: A Biosynthetic Approach*. New York: John Wiley.

Halva S and Craker LE (1996) *Manual for Northern Herb Growers*. Guildford and King's Lynn, UK: Biddles Ltd.

Hay RKM and Waterman PG (1993) *Volatile Oil Crops*. Harlow, UK: Longman.

Hodge WH (1975) Survey of flavour producing plants. *Food Flavours* 22: 244–245.

Hornok L (ed.) (1992) *Cultivation and Processing of Medicinal Plants*. Chichester, UK: John Wiley.

International Conference on Essential Oils and Aromas (1998) *Global Markets Present and Future*. London: International Federation of Essential Oils and Aroma Trades.

Mason L and Brown C (1999) *Traditional Foods of Britain. An Inventory*. Trowbridge, UK: Prospect Books. Cromwell Press.
Moyler DA and Heath HB (1988) Liquid carbon dioxide extraction of essential oils. In: Lawrence BM, Mookerjee BD and Willis BJ (eds) *Flavours and Fragrances: A World Perspective*, pp. 41–63. Amsterdam: Elsevier Science.
Small E (1997) *Culinary Herbs*. Otawa, Ontario: NRC Monograph Publishing Program.
Svoboda K and Deans SG (1990) A study of the variability of rosemary and sage and their volatile oils on the British market: their antioxidative properties. *Flavour and Fragrance Journal* 7: 81–87.
Svoboda K, Gough J, Hampson J and Galambosi B (1994) Analysis of the essential oils of some *Agastache* species grown in Scotland from various seed sources. *Flavour and Fragrance Journal* 10: 139–145.
Svoboda K and Hampson J (1998) Chemicals from plants. *The Horticulturist* 7: 4, 26–31.
Walton NJ and Brown DE (ed.) (1999) *Chemicals from Plants*. Covent Garden, London: World Scientific Publishing Co. Ltd.
Weiss EA (1997) *Essential Oil Crops*. Wallingford, UK: CAB International.

Herbs of the Compositae

S G Deans and E J M Simpson, Aromatic and Medicinal Plant Group, South Ayrshire, Scotland, UK

Introduction to the Group

The Compositae represents the largest family of flowering plants, with over 900 genera and 14 000 species comprising examples of various ecological types and life forms, such as trees, shrubs, perennial and annual herbs, aquatic, alpine and desert plants, climbers, succulent, and spiny shrubs. Many species are cultivated as vegetables (lettuce, endive, artichoke, salsify, sunflower) and several genera as culinary and medicinal plants (*Achillea, Aster, Calendula, Helenium, Rudbeckia, Tagetes*).

Special Morphological and Anatomical Features

Flowers from small florets are aggregated into heads (capitula) simulating a single large flower and surrounded by a calyx-like involucre of one or more rows of bracts. These vary greatly in number and arrangement. Florets can be all similar (head homogamous) or with central and marginal florets differing (head heterogamous). Pollination is entomophilus, rarely anemophilous; some genera (*Taraxacum, Hieracium*) contain apomictic species.

The family exhibits several specific features, such as secretory resin canals (*Achillea, Ambrosia, Anthemis, Arnica, Artemisia, Centaurea, Cnicus, Echinacea, Ehlinthus, Eupatorium, Inula, Matricaria, Parthenium, Petasites, Pyrethrum, Senecio, Solidago, Tagetes, Tanacetum*), secretory cavities (*Artemisia, Inula, Tagetes*), lactiferous canals, cells and vessels (*Arctium, Carlina, Silybum*), different types of glandular (*Eupatorium, Haplopappus, Helianthus, Olearia*) and nonglandular hairs, and anomalous secondary thickening (*Artemisia, Helianthus, Taraxacum*).

The volatile oil found in the leaves of *Artemisia dracunculus* is produced partly in schizogenic volatile oil canals and partly in characteristic glandular hairs. Differences have been recorded between the oil found in the canals and the oil produced by the glandular hairs. Primary volatile oils extend from the stem into the root system where they are located between the endodermis and the cortex. Excretion canals are also found in the roots, but do not extend to the stem (**Figure 1**).

A blue volatile oil is extracted from the flowers of German chamomile (*Matricaria chamomilla* L.) produced in secretory glandules and canals localized

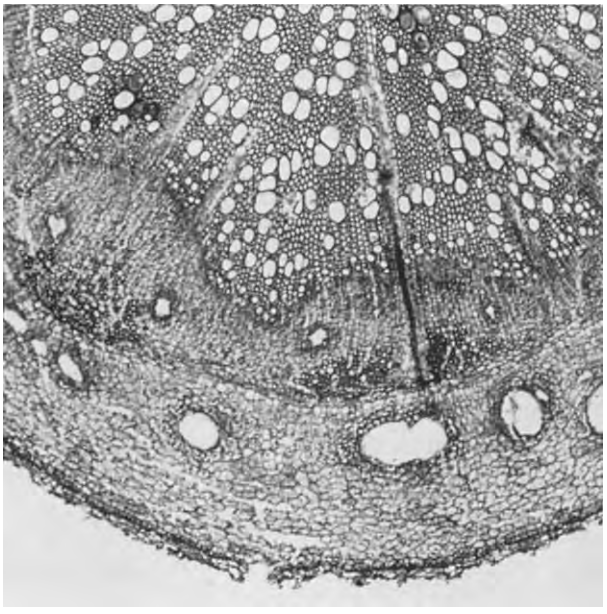

Figure 1 (see color plate 93) Volatile oil canals in root of *Artemisia absinthium* (wormwood). Bright field × 40. Reproduced from Herbs: Herbs of the Compositae, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

mainly in the anthodia. The content of the oil varies considerably during the development of the plant and the flowering of the anthodia. (*See* **Essential Oils: Properties and Uses.**)

Harvesting and Handling Procedures

Harvesting crops at the correct stage of maturity is essential to achieve a product that will be of top quality and meet the buyers' standards. Harvesting of herbs is still carried out by hand in many cases. The development of mechanized harvesting systems has proved to be problematical owing to the wide variety of plant parts to be harvested (roots, rhizomes, stems, leaves, flowers, seeds, and pods) combined with a relatively small crop acreage. Machinery currently in use has either been adapted from equipment constructed to harvest other crops(usually forage, grain, or vegetable products) or specifically designed and constructed for a few herbs or even one specific herb. The construction of new harvesting machinery can only be considered when the market value of the crop or related product is high and the acreage is relatively large, or when large commercial companies finance such equipment development as in chamomile production in Germany.

Drying facilities are essential and may vary from an unused room to an elaborate drying shed. Low humidity is important and using a dehumidifier allows better control over the drying process. The temperature of the drying is critical since many volatile compounds can be lost at elevated temperatures (Table 1).

Use in the Food Industry

Many members of the Compositae are cultivated as edible plants, vegetables, herbs, and spices (Table 2). The edible tubers of the Jerusalem artichoke consist of the parenchyma of the medullary rays and pith, together with the xylem. Fleshy parts of the inflorescence of the globe artichoke are eaten. The roots of scorzonera and salsify are used as vegetable, while dried and roasted roots of dandelion and chicory are used as coffee substitutes and in medicine.

Table 1 Changes in volatile oil yield (%)[a] with rising temperature

Plant	*Temperature (°C)*						
	40	*50*	*60*	*70*	*80*	*90*	*100*
Basil	1.0	0.7	0.7	0.7	0.3	0.3	0.1
Sweet marjoram	1.7	1.3	1.3	1.0	1.0	0.7	0.5
Rosemary	2.3	2.0	2.0	1.7	1.7	1.0	0.3
Sage	2.0	1.7	1.0	0.5	0.5	0.4	0.3
Summer savory	1.3	1.0	0.8	0.7	0.5	0.03	< 0.01
French tarragon	0.5	0.3	0.3	0.2	0.2	0.2	0.1
Thyme	1.0	1.0	1.0	0.3	0.3	0.2	< 0.01

[a]Percentage of volatile oils expressed as value per gram of dry matter of plant material.

Research and Development

An increasing demand from the food, pharmaceutical, and cosmetic industries calls for the development of new varieties with high levels of active constituents, consistent quality, and safe application. The anatomy and the physiology of each plant have to be investigated in order to elucidate the plant's requirements and adaptation to cultivation. Several species are currently being investigated for their bioactivity and a number of representative genera will be discussed. (*See* **Biotechnology in Food Production.**)

Chamomile

Roman chamomile (*Chamaemelum nobile* (L.) All.) and German chamomile (*Matricaria recutita* L.) are two different species of plant commonly called chamomile. Roman chamomile, formerly classified as *Anthemis nobilis* L., is a creeping, herbaceous perennial native to western Europe and North Africa. German chamomile is an erect-growing annual, native to Europe and western Asia. The volatile oil of Roman chamomile consists mainly of chamazulene, angelic acid, tiglic acid and several sesquiterpene lactones. German chamomile contains chamazulene, α-bisabol, α-bisabol oxides A and B, spathulenol and farnesene. Extracts from both plant species are reported to have antiseptic, antibacterial, antifungal, and antitumor activities.

The breeding of chamomile has been targeted towards increased content of two sequiterpenes: chamazulene and α- and β-bisabolol. The diploid and tetraploid varieties have been introduced on the market in Germany and the former Czech Republic, with high and stable contents of these sesquiterpenes. Other important flavanoid secondary metabolites, namely apigenins, are now being investigated in detail. (*See* **Colorants (Colourants): Properties and Determination of Natural Pigments.**)

Achillea millefolium

Yarrow is a hardy perennial plant, native to Europe and Asia. The chemical constituents include a volatile oil comprised of azulene, caryophyllene, thujone, eucalyptol, α- and β-pinene and borneol, as well as lactones, tannins, and alkaloids. Yarrow extracts

Table 2 Uses of herbs of the family Compositae

Plant species	*Medicinal*	*Food industry uses*	*Colors*	*Others*
Achillea millefolium	✓	Salads, substitute in brewing or tobacco		Cosmetics
Antennaria dioica	✓			
Arctium lappa	✓	Salads, candied stalks, liqueurs		
Arnica montana	✓			
Artemisia aborotanum	✓		✓	Insect repellent, cosmetics
Artemisia absinthium	✓	Tea, wine, stuffing		
Artemisia dracunculoides	✓	Flavorings, dressing, vinegar	✓	Cosmetics
Artemisia dracunculus		Flavorings, liqueurs		
Artemisia vulgaris	✓	Tea, stuffing, flavoring, beer		Insect repellent
Bidens tripartita	✓		✓	
Calendula officinalis	✓	Sustitute for saffron, tea, salads, cheese, and butter coloring	✓	Cosmetics
Carlina acaulis	✓			
Centaurea cyanus	✓		✓	
Chamemelum nobile	✓	Beer, tea	✓	Cosmetics
Chrysanthemum balsamita	✓	Salads, beer, soups, cake, poultry		Cosmetics
Chrysanthemum cinerarifolium	✓	Meat and poultry (reduces grease)		Insect repellent
Chrysanthemum parthenium	✓	Meat and poultry (reduces grease)		Insect repellent
Cichorium intybus	✓	Salads, coffee substitute	✓	
Cnicus benedictus	✓	Salads, apertifs		
Cynara scoymus	✓	Cooked vegetable		
Echinacea angustifolia	✓			
Erigeron canadensis	✓			
Eupatorium cannabinum	✓			
Eupatorium perfoliatum	✓			
Eupatorium purpureum	✓			
Gnaphalium uliginosum	✓			
Grindelia camporium	✓			
Helianthus annuus	✓	Oil, edible seeds, coffee substitute		Lamp oil Paper
Hieracium pilosella	✓			
Inula helenium	✓	Candied sweets, liqueurs, vermouths, wines, bitters		
Lactuca virosa	✓			Cosmetics
Lapsana communis	✓	Salads, cooked vegetable		
Liatris odoratissima				Cosmetics
Liatris spicata	✓			Insect repellent, cosmetics
Matricaria chamomilla	✓	Tea, flavor in alcoholic drinks	✓	Cosmetics
Petasites hybridus	✓			
Santolina chaemaecyparissus	✓			Insect repellent
Senecio aureus	✓			
Senecio vulgaris	✓			
Silybum marianum	✓	Cooked vegetables		
Solidago virgaurea	✓			
Tanacetum vulgare	✓			
Taraxacum officinale	✓	Salads, wine, beer, coffee substitute		Insect repellent
Tragopogon porrifolius		Salads, cooked vegetables		
Tussilago farfara	✓	Salads, wine		Herbal tobacco

are generally recognized as safe in beverages, but the finished product must be thujone-free. Extracts of yarrow have also been shown to be effective as mosquito repellents. Contact with yarrow has been reported to cause dermatitis. Field plots and experiments have been established with the objective of extracting chamazulene: 24 t of fresh matter per ha has been harvested at full bloom, yielding 0.3% volatile oil and 4.43 kg of chamazulene per hectare.

Artemisia dracunculus, A. afra, and *A. absinthium*

Artemisia dracunculus L., French tarragon or estragon, is a hardy perennial which grows wild throughout central Europe and Asia, but is also widely cultivated owing to the popularity of tarragon-vinegar-based sauces and dressings. The volatile oil of tarragon, produced in oil canals and glandular hairs, is pale yellow to amber in color, with a delicate spicy odor. The oil consists mainly of methyl chavicol,

typically 65%, as well as β-pinene, camphene, ocimene, limonene, menthol, and sabinene. Extracts of tarragon are reported to have antiseptic, stimulant, antibacterial, and antifungal activities. A mixture of flavones and isocoumarins has also been found in the shoots. French tarragon rarely sets viable seeds and must be propagated from cuttings or by root division. Seed is available from Russian tarragon (*Artemisia dracunculoides*) which is markedly inferior since it lacks the distinct odor and taste required for culinary purposes. Recent micropropagation studies on French tarragon have shown that tissue culture techniques could provide an efficient means of propagation. Optimal growth was achieved using sterile shoot tip explants, placed on mineral salts medium supplemented with sucrose and the phytohormones, benzyl amino purine (BAP, a shoot-promoting phytohormone) and naphthalene acetic acid (NAA, a root-promoting phytohormone). Rooting was maximized by dipping new shoots in NAA or indole-3-butyric acid (IBA, a root-promoting phytohormone).

Artemisia afra Jacq, African wormwood, is indigenous to the mountainous regions of South Africa. It is a bushy, branched perennial shrub growing up to 1 m and can be propagated from seeds, cuttings, and root pieces. It has a thin brownish bark and green-gray leaves with a marked aromatic scent when crushed. It possesses characteristic paniculate aggregations of pale yellow inflorescences. *A. afra* is one of the most widely used medicinal plants. Traditionally, extracts have been applied to a whole range of disease conditions, such as common cold, diabetes mellitus, bronchial complaints, and stomach disorders. The antibacterial properties of the volatile oil are summarized in Table 3. The leaves are used in a tea or as a leaf poultice, and the vapors from boiling leaves can be inhaled. It has also been used as an insect repellent and anthelmintic agent. The volatile oil contains α-thujone, β-thujone, 1,8-cineole, camphor, and α-pinene.

Artemisia absinthium L., wormwood or absinthe, is an erect-growing perennial native to Europe and naturalized in northeastern North America. It is also cultivated commercially in central and northwestern USA. The volatile oil extracted from this very bitter herb has a strong odor and acrid taste, and contains mainly thujone and thujyl alcohols, as well as myrcene, α-pinene, nerol, limonene, and 1,8-cineole. The oil, which can be poisonous, due to its high thujone content, finds limited use in fragrance compounding and external analgesics. Recently, essential oil distilled from *A. absinthium* collected in the Spanish Pyrenees was found to contain no thujone, which may be of significant interest to the food industry. The bitter principle of wormwood is made up of the glycosides, absinthin and absinthinin. The plant also contains lignans, oligosaccharides, flavonol glycosides, and tannins. Owing to its bitterness, wormwood has few culinary uses, but extracts are used to flavor bitter drinks such as vermouth, absinthe, and tonic water, as well as being used medicinally, chiefly as an anthelmintic. (*See* **Vermouth**.) More recently, extracts of *A. absinthium* have shown antibacterial activity against both Gram-positive and Gram-negative bacteria, but not against yeasts.

Table 3 Bactericidal properties (zone of inhibition in mm) of *Artemisia afra* volatile oil

Organism	*Dilution of* A. afra *volatile oil*			
	1.1	*1.2*	*1.5*	*1.10*
Acaligenes faecalis	9.8	6.0	5.2	5.2
Acinetobacter calcoacetica	17.8	9.4	6.6	5.7
Aeromonas hydrophila	4.0	4.0	4.0	4.0
Bacillus subtilis	7.5	7.0	5.6	5.0
Beneckea natriegens	16.2	10.4	8.1	6.3
Brevibacterium linens	19.0	20.0	6.1	5.3
Brocothrix thermosphacta	13.3	9.4	7.0	6.6
Citrobacter freundii	11.5	9.6	8.8	7.6
Clostridium sporogenes	6.1	4.9	4.0	4.0
Enterobacter aerogenes	8.4	5.4	5.4	5.4
Enterococcus faecalis	9.9	7.0	6.0	5.3
Erwinia carotovora	7.6	4.0	4.0	4.0
Escherichia coli	6.4	6.2	5.8	5.3
Flavobacterium suaveolens	9.9	6.7	6.3	4.6
Klebsiella pneumoniae	15.5	10.1	8.1	7.1
Lactobacillus plantarum	7.6	6.2	6.2	4.8
Leuconostoc cremoris	9.4	8.2	4.0	4.0
Micrococcus luteus	8.6	5.6	4.0	4.0
Moraxella spp.	6.0	5.0	4.5	4.1
Proteus vulgaris	4.0	4.0	4.0	4.0
Pseudomonas aeruginosa	6.4	4.8	4.8	4.0
Salmonella pullorum	5.3	4.5	4.3	4.2
Serratia marcescens	11.0	7.7	5.1	4.6
Staphylococcus aureus	8.3	6.9	5.0	4.2
Yersinia enterocolitica	8.3	6.5	6.0	5.8

Recent studies on wormwood have examined the transformed or 'hairy' roots produced *in vitro* by the infection of sterile explants with the soil bacterium, *Agrobacterium rhizogenes*. Differences were noted in the volatile oil profile of the transformed material cultured in flasks and the normal field-grown roots. Microscopic examination revealed a lack of maturity in the transformed root which may account for these differences (Figure 2 and 3).

Poisonous Plants

The family Compositae includes one of the most important poisonous plants in the UK, ragwort (*Senecio jacobaea*). It is abundant on wasteland, beside roads, and in pastures, where it is a troublesome weed. The poisonous principles are pyrrolizidine alkaloids, many

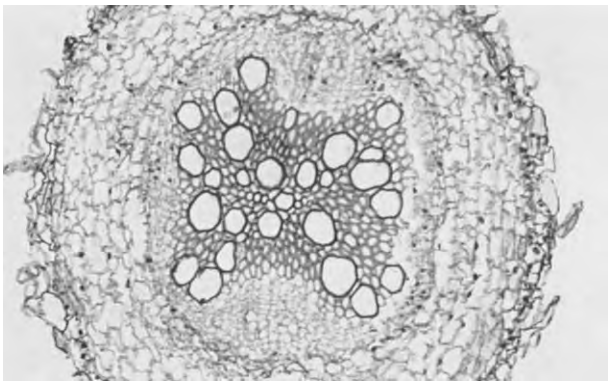

Figure 2 (see color plate 94) Transverse section of *Artemisia absinthium* (wormwood) root showing vascular bundle. × 200. Reproduced from Herbs: Herbs of the Compositae, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

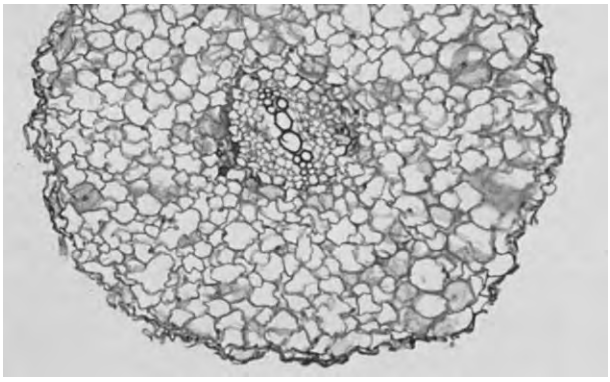

Figure 3 (see color plate 95) Transverse section of *Artemisia absinthium* (wormwood) root showing vascular bundle in transformed root. × 200. Reproduced from Herbs: Herbs of the Compositae, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

of which have been characterized and named, including senecionine, seneciphylline, jaconine, and jacobine. Poisoning in animals causes disorders of the digestive and nervous systems.

Another group of poisonous substances in the Compositae comprises terpenoid compounds, especially sesquiterpene lactones (present in *Lactua virosa, Artemisia douglasii, Achillea millefolium, Matricaria chamomilla*, and *Cichorium endivia*). These compounds can cause allergic skin reactions. *Arnica montana* and great burdock (*Arctium lappa*) can also cause skin irritation owing to attachment of flower heads by hooked bracts.

See also: **Biotechnology in Food Production**; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Essential Oils**: Properties and Uses; **Vermouth**

Further Reading

Arino A, Arberas I, Renobales G, Arriaga S and Dominguez JB (1999) Essential oil of *Artemisia absinthium* L. from the Spanish Pyrenees. *Journal of Essential Oil Research* 11(2): 182–184.

Balza F, Jamieson L and Towers GHN (1985) Chemical constituents of the aerial parts of *Artemisia dracunulus*. *Journal of Natural Products* 48(2): 339–341.

Deans SG and Kennedy AI (2000) *Artemisia absinthium*. In: *Medicinal and Aromatic Plants – Industrial Profiles: Artemisia*. Amsterdam, The Netherlands: Harwood Academic.

Dulger B, Ceylan M, Alitsaous M and Ugurlu E (1999) Antimicrobial activity of *Artemisia absinthium* L. *Turkish Journal of Biology* 23(3): 377–384.

Firouzi R, Azadbakht M and Nabinedjad A (1998) Antilisterial activity of essential oils of some plants. *Journal of Applied Animal Research* 14(1): 75–80.

Hamill JD, Parr AJ, Rhodes MJC, Robins RJ and Walton NJ (1987) New routes to plant secondary products. *Biotechnology* 5: 800–804.

Kennedy AI, Deans SG, Svoboda KP, Gray AI and Waterman PG (1991) Comparison of the volatile oil from fermenter-grown transformed 'hairy' roots and field-grown roots of *Artemisia absinthium* L. (wormwood). *Proceedings of the Phytochemical Society of Europe Symposium, Bioformation of Flavours*. London: Kings College.

Kennedy AI, Deans SG, Svoboda KP, Gray AI and Waterman PG (1991) Volatile oils from normal and transformed roots of *Artemisia absinthium*. *Phytochemistry* 32: 1449–1451.

Lourenço PML, Figueiredo AC, Barroso JG *et al.* (1997) *Achillea millefolium* hairy roots: growth, essential oil composition and antibacterial activity in comparison with parent plant roots. *Proceedings of the IV Iberian Congress on Biotechnology*. Braga, Portugal: University of Minho.

Mackay WA and Kitto SL (1988) Factors affecting *in vitro* shoot proliferation of French tarragon. *Journal of the American Society for Horticultural Science* 113(2): 282–287.

Pino JA, Rosado A and Fuentes V (1998) Chemical composition of the leaf oil of *Achillea millefolium* L. grown in Cuba. *Journal of Essential Oil Research* 10(4): 427–428.

Repcak M, Halasova J, Honcarn R and Podhradsky D (1980) The content and composition of the essential oil in the course of anthodium development in wild chamomile. *Biologia Plantarum* 22(3): 183–191.

Simon JE, Chadwick AF and Craker LE (1984) *Achillea millefolium* L.). In *Herbs, an Indexed Bibliography* 1971–1980. Amsterdam: Elsevier Science.

Thorsell W, Mikiver A, Malander I and Tunon H (1998) Efficacy of plant extracts and oils as mosquito repellants. *Phytomedicine* 5(4): 311–323.

Herbs of the Labiatae

S Kokkini, R Karousou and E Hanlidou, Aristotle University, Thessaloniki, Greece

Introduction

The Labiatae (Lamiaceae) family consists of approximately 3500 species centered chiefly in the Mediterranean area, although some small groups have a local distribution in Australia, South-west Asia, and South America. The family contains many culinary or flavoring herbs, which have been widely used, since antiquity, as spices, teas, or traditional medicines. Several members of the family are also used as sources of essential oils.

Morphological Characteristics

The Labiatae are annual or perennial herbaceous plants or shrubs, often densely glandular and aromatic. Their stems are usually four-angled. The leaves are usually simple and always opposite. The flowers are hermaphrodite or male-sterile (functionally female) and form whorls (verticillasters) that are often arranged in spikes, heads, racemes, or cymes. The calyces are usually five-lobed and two-lipped; rarely, the calyx is one-lipped or actinomorphic. The corollas are sympetalous and two-lipped; rarely, they are one-lipped or more or less actinomorphic. They usually have four stamens, rarely reduced to two; the upper (posterior) pair is usually shorter than the lower (anterior). The ovary is superior, doubly carpellate but appearing equally four-lobed when mature. The style is single, usually gynobasic, shortly bifid above. The fruit consists of four dry (very rarely fleshy), one-seeded nutlets.

Besides the non-glandular or clothing hairs, the leaves and other aerial parts of the Labiatae plants bear a great number of glandular trichomes. Two main types of these secretory structures occur in the Labiatae: the glandular hairs (capitate hairs) and the glandular scales (peltate hairs). The former type is very variable in stalk length and head shape, whereas the latter is more or less stable, consisting of a uni- or bicellular foot, a uni- or bicellular stalk, and an (up to) 18-celled head. The bulk of the essential oil is thought to be produced and accumulated in the glandular scales. Because of their size and number, they can be seen as bright yellowish dots when the lower leaf surface is examined with a hand lens (Figure 1).

Chemical Characteristics

The active ingredients, to which the Labiate herbs owe their flavors and aroma, are the essential oils. They are usually obtained by steam- or water-distilling their fresh or dried aerial parts. However, though the Labiatae are generally considered as an aromatic family, they include several genera whose members produce only traces of essential oils and are almost scentless. The genera of the subfamily Lamioideae consist of essential oil-poor species (e.g., *Phlomis*, *Sideritis*), whereas the subfamily Nepetoideae is quite variable. The different Nepetoideae genera may include oil-poor (e.g., *Lycopus*), oil-poor and oil-rich (e.g., *Salvia*, *Micromeria*) or only oil-rich (e.g., *Mentha*) species.

A variety of highly valued constituents have been found as main components of the Labiatae oils. However, the essential oil profile that determines the economic value of a Labiate herb may vary greatly within a species. As a result, plants belonging to a single species may not be used for the same purpose. However, there are numerous examples of similarity in the oils of widely diverse genera, and thus different species are used similarly.

The total essential oil content and the quantitative and/or qualitative oil composition of a single species depend on the season, the local environmental conditions and the genetic differences. The age and the season at which each herb is harvested are often of considerable importance, since the total oil amount and the relative proportions of its active constituents vary throughout the year (seasonal variation). Sage plants, for example, produce in late summer almost double the amount of the essential oil that is present in spring, whereas the highest

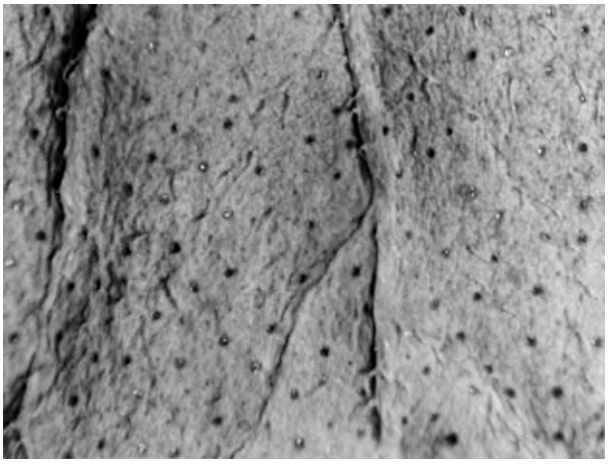

Figure 1 Lower side of *Origanum vulgare* subsp. *hirtum* leaf (Greek oregano) as it appears under the stereoscope (× 32).

thujone content can be obtained in late autumn. Oregano plants grown in areas with mild winter temperatures and a typical Mediterranean summer (low precipitation and high temperature) are much richer in essential oil (up to 30 times more) than those found in areas suffering from low winter temperatures. Similarly, the total phenol content (carvacrol+thymol) of oregano plants is much higher in areas with a Mediterranean type of climate than in areas characterized by a Continental climate (geographic–climatic variation).

Striking differences in qualitative oil composition between individuals of a Labiatae species suggest the presence of chemotypes or chemical races (genetic variation). The different chemical races are distinguished on the basis of the presence or absence of a particular biosynthetic pathway, leading to the formation of a number of compounds and not on the presence or absence of a single compound. Different chemotypes are often encountered in the wild populations of *Mentha*, *Thymus*, and *Satureja* species, and their oils are exploited for different applications. One of the most variable species is *Mentha spicata* with four distinct chemotypes: (1) essential oil rich in linalool and/or linalyl acetate (plants with a fragrant smell), (2) oil rich in carvone/dihydrocarvone (sweet smell), (3) oil rich in piperitenone and/or piperitone oxide (musty smell), and (4) oil rich in methone/isomenthone/pulegone (pungent smell).

It has been established that several of the essential oil constituents are biologically active, with significant antibacterial, antifungal, insecticidal, and antioxidant properties. However, besides essential oils, the members of Labiatae produce other types of biologically active compounds. Diterpenes, which present strong antibacterial and antifungal activities, occur in several genera, such as *Salvia* and *Sideritis*. Iridoids (monoterpene lactones) are mainly found in members of the subfamily Lamioideae (e.g., *Ajuga*, *Phlomis*, *Scutellaria*, and *Leonurus* spp.) and more rarely in Nepetoideae (e.g., *Nepeta*, *Salvia*, and *Satureja* spp.). Several of these show insecticidal, cytotoxic, and cytostatic activities. Among the phenolic constituents, rosmarinic acid is of great interest, especially for its antioxidant and antiinflammatory properties. It is produced by several taxa, including *Rosmarinus officinalis*, *Salvia officinalis*, *Melissa officinalis*, and *Origanum vulgare*.

The commonest Labiatae herbs used in the food industry are listed below under their commercial names. A commercial name may attributed to a specific chemotype or to a particular species or even to a group of species that have a similar oil composition and consequently a similar flavor and use.

Balm

Balm (*Melissa officinalis* L.) is a perennial herb up to 100 cm in height, which is native to South Europe, North Africa, and Anatolia and introduced in almost all temperate regions. It has broadly ovate leaves and white or light pink flowers arranged in axillary whorls. Three subspecies of *Melissa officinalis* are recognized, but only subsp. *officinalis* is cultivated for its aromatic foliage, which has a strong, pleasant, lemon-like odor.

The aromatic balm leaves are often used in beverages and as a seasoning in salads, dressings and sauces, as well as in cooked foods, e.g., in soups and stews. Balm leaves yield a low amount of essential oils, up to 0.5% vol./dry wt. Balm oil is considered as a mild spasmolytic agent and is employed as a flavor ingredient in alcoholic beverages. The main compounds of the commercial oils are citronellal, geranial (citral a), and neral (citral b).

Basil

Basil is obtained from the foliage of *Ocimum basilicum* L. (sweet basil), an annual herbaceous plant up to 100 cm in height. It is native to tropical Asia, Africa and America and widely cultivated in pots and gardens in Europe, South-west Asia and the USA. The leaves are ovate and vary in size, depending on the cultivated variety, from the small leaves of the common basil to the large leaves of lettuce leaf basil. The verticillasters of the white or pink flowers are arranged in terminal racemes. Different cultivated forms, considered distinct by herbalists (e.g., *O. minutum* L.), are not recognized taxonomically.

The dried sweet basil leaves have a sweet, fragrant odor, and their taste is aromatic, warm, and slightly pungent. Basil is considered as the finest of all aromatic herbs and is widely used to flavor cooked vegetables, tomato-paste products, and fish. It is sometimes used with, or as a substitute of, oregano to flavor pizza and spaghetti sauce and is employed together with other spices in the manufacture of vinegar, mustard, and sausages. Though not used in large quantities, sweet basil oil is used quite extensively in the flavoring of several food products, including those for confectionery, alcoholic beverages (liqueurs), baked goods, and condiments. The commercial essential oils are usually methyl-chavicol (an isomer of anethole) and/or linalool rich. However, *O. basilicum* oils are particularly variable and may also have high amounts of methyl-cinammate, geraniol, eugenol, and methyl-eugenol.

Hyssop

Hyssopus officinalis L., an aromatic shrub growing on dry hills and rocky grounds, is native from Portugal and Spain eastwards to India. It is naturalized in the USA, where it grows in gardens and along roadsides. Hyssop plants are up to 60 cm high and have linear or almost linear leaves, actinomorphic calyces and violet–blue, pink, or white flowers in verticillasters, forming a terminal spike-like inflorescence. *H. officinalis* is a morphologically variable species, separated into four subspecies. The cultivated plants belong to subsp. *officinalis*.

The dried aromatic hyssop leaves are considered as a gentle stimulant used in medicines, tonics, bitters, and alcoholic liqueurs. Hyssop oil is employed mostly in the flavoring of beverages, e.g., bitters, tonics, and especially of French-type liqueurs. The essential oils produced by *H. officinalis* have a variable composition, mainly consisting of isopinocamphone (the major component of commercial oils), pinocamphone, pinocarvone, and 1,8-cineole.

Marjoram

The vernacular name marjoram comprises several Labiatae herbs belonging to different species or even genera. The best known are as follows:

Spanish Marjoram

Spanish marjoram is obtained from *Thymus mastichina* L., a dwarf shrub up to 50 cm in height, which grows wild in Spain and Portugal. It has ovate to elliptic leaves, almost actinomorphic calyces, whitish corollas and an inflorescence of somewhat plumose appearance.

Spanish marjoram, also known as wild or forest marjoram, has a different odor from that of sweet marjoram, being harsher and eucalyptus-like, due to the high amount of 1,8-cineole (eucalyptol). Spanish marjoram is used in the manufacture of fragrances, but also in the food industry to flavor canned and processed meats and sauces. The essential oils produced by wild *Th. mastichina* plants may also have high amounts of linalool.

Sweet Marjoram

Sweet marjoram is obtained from *Origanum majorana* L. (syn. *Majorana hortensis* Moench., *M. vulgaris* Miller), a dwarf shrub of dry and rocky places, native on Cyprus and the adjacent part of South Turkey. It is cultivated in several countries in Europe, Africa, America, and Asia. The plants are up to 60 cm high and have ovate, whitish or greyish leaves, one-lipped calyces and white corollas. The flowers are arranged in spikes forming a paniculate inflorescence.

Because of their fragrant odor and pleasant, highly aromatic, slightly sharp and spicy taste, the dried leaves of *O. majorana* are widely used as flavoring agents for dressings, egg and vegetable dishes, soups, stews, cheese, liver, Polish sausages, and fancy meats. The characteristic compounds of the commercially exploited marjoram essential oils are terpinen-4-ol, α-terpinene, γ-terpinene, α-terpineol and *cis*-sabinene hydrate, occurring in variable quantities. *O. majorana* also produces essential oils with different compositions, which are either rich in linalool or *p*-cymene and its biosynthetically related compounds thymol and carvacrol.

Mint

Plants producing essential oils rich in menthane-type compounds are frequently found within the Labiatae, commonly known as the mint family. However, those that today constitute the commercial mint herbs are species of the genus *Mentha*.

Mentha species are perennial herbs up to 150 cm in height, usually growing in damp sites. The white, pink, or lilac flowers are arranged in dense, many-flowered verticillasters, which, in some cases, form a long and spike-like inflorescence or a terminal head. The genus is taxonomically difficult, and the identification of a single plant is often problematic, since, in addition to much morphological plasticity, most of the species are capable of hybridization with each other. Besides the high morphological variability, great differences in chemical composition of the essential oils often occur within a species. As a result, although the genus consists of approximately 25 species and rather fewer hybrids, more than 1000 names of different taxonomic rank have been attributed over the past 200 years to the morphological and/or chemical variants of the mint plants. Among the different *Mentha* species and hybrids, those that have achieved economic importance are as follows.

Cornmint

Extremely menthol-rich essential oils (up to 80% menthol) are produced by *Mentha arvensis* var. *piperascens*, mainly plants cultivated since antiquity in Japan and during recent years in China, India, and Brazil. The cornmint plants are distinguished by their leafy at apex inflorescence, which consists of remote flower verticillasters.

The cornmint oil (also known as Japanese or Chinese or Brazilian peppermint oil) is valued as a source of L-menthol, which is generally obtained by simple

freeze-crystallization. The dementholized oil is used as inexpensive, rather harsh peppermint-like oil.

Pennyroyal

The leaves and flowering parts of *Mentha pulegium* L. (European pennyroyal) plants grown wild in western and central Europe as well as in the Mediterranean region are traditionally used for the preparation of teas and, to a limited extent, for food flavoring (soups, stuffings). *M. pulegium* is distinguished from the other mint species by its inflorescence, which consists of remote verticillasters but is not leafy at apex (Figure 2). Furthermore, *M. pulegium* is distinguished from the other mint species by the calyx, which has unequal teeth (weakly two-lipped), while the corolla tube is gibbous beneath. The North American pennyroyal is *Hedeoma pulegioides* (L.) Pers., an annual plant growing in North America, which has a similar odor to *M. pulegium*.

Since Roman times, herbalists have recommended pennyroyal as an abortifacient herb. Recent data suggested that pulegone, the ketone, often the major constituent of pennyroyal oil, has hepatotoxic effects to animals and man.

Peppermint

Mentha × *piperita* L., a sterile hybrid of the species *M. aquatica* L. (Figure 2) and *M. spicata*, is probably the most important commercial aromatic herb in the world today from the standpoint of the size of the area cultivated for oil distillation. The peppermint leaves have a characteristic, sweetish, strong odor and an aromatic, warm, pungent taste, with a cooling aftertaste.

The essential oils of *M.* × *piperita* cultivated plants are characterized by the preponderance of menthone, isomenthone, and the different isomers of menthol. Peppermint oil finds wide application in the flavoring of chewing gums, sugar confectionery, ice creams, desserts, baked goods, tobacco, and alcoholic beverages. It is also frequently employed in the flavoring of pharmaceutical and oral preparations.

Spearmint

Under the vernacular name 'spearmint,' different *Mentha* species and hybrids producing carvone-rich essential oils are commerically exploited. In particular, two main 'types' of spearmint – Native and Scottish – are widely cultivated, mainly in the USA but also in China, Europe, and South Africa, for the commercial production of their essential oil. The native spearmint oil is obtained either from the species *M. spicata* L. (Figure 2) or from the sterile hybrid *M.* × *villoso-nervata* Opiz. (*M. longifolia* L. × *M. spicata*), whereas the Scottish spearmint oil is produced from the hybrid *M.* × *gentilis* L. nm. *cardiaca* Gray (*M. arvensis* L. × *M. spicata*).

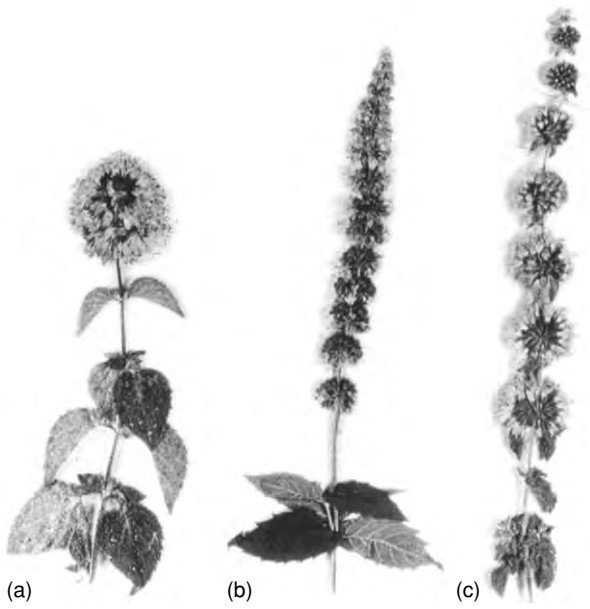

Figure 2 (a) *Mentha aquatica*; (b) *M. spicata*; (c) *M. pulegium*.

Spearmint plants are familiar garden herbs, and they are often used to flavor vegetables, soups, meat and fish sauces, and salads. The spearmint oil, characterized by the high participation of carvone, dihydrocarvone, and their related alcohols and esters, is mainly used in the flavoring of chewing gums, toothpastes, and other oral products.

Oregano

Oregano, one of the most popular herbs throughout the world, presents a serious problem to everyone trying to establish the identity of its botanical source. A vast number of species in over a dozen genera of different families have been referred to 'oregano' in the literature. Oregano is in fact a characteristic flavor produced by a number of plant species that yield essential oils with a more or less high content of carvacrol (noncrystallizable phenol). The flavor of oregano recently became very popular all over the world mainly because the fast food industry supports the ethnically based products. Dried leaves and flowering parts of different oregano plants, commonly known as 'the pizza herb,' are widely used to flavor food products, such as salads, vegetables dishes, fried potatoes, meats, sauces, sausages, and canned goods. Oregano is also employed as an ingredient in chilli powder and is used in chilli con carne and several other

Mexican dishes. Oregano oil sometimes replaces the dried herbs in the food flavoring. Several reports have mentioned its high antibacterial, fungicidal, insecticidal, and antioxidant activity. The species, which are widely used as oregano under different vernacular names, are described below.

Greek Oregano

Greek oregano consists of dried leaves and flowering parts of *Origanum vulgare* subsp. *hirtum* (Link) letswaart (syn. *O. heracleoticum* auct. non L.), and on commercial grounds, it is considered to be of the highest quality. It is a woody perennial up to 100 cm in height, and is found in the Balkan Peninsula and Turkey. It is distinguished from the other *Origanum* species by its dense spikes, clearly distinct from stems and branches, green calyces with five (sub)equal teeth and two-lipped white corollas (Figure 3). Furthermore, it is characterized by the conspicuously glandular-punctate (visible even to naked eye) leaves and calyces. It is mainly collected from the wild from Balkan countries and Turkey. The cultivation in gardens and open fields is nowadays very popular in the Mediterranean countries.

Figure 3 (see color plate 82) *Origanum vulgare*.

O. vulgare subsp. *hirtum* plants often yield an extremely high amount of essential oil (up to 8% dry weight basis), and their carvacrol content may reach the 95% of the oil produced. However, plants having a high thymol content (up to 85%) are not infrequent in the wild populations. In these cases, carvacrol, the compound responsible for characterizing a plant as being of the oregano type, is a minor constituent.

Spanish Oregano

Spanish oregano consists of leaves and flowers of *Coridothymus capitatus* (L.) Reichenb. fil. (*Thymus capitatus* Hoffmans & Link, *Satureja capitata* L.), a shrub 50–150 cm in height, with narrowly elliptical leaves, dorsally flattened calyces, and violet corollas (Figure 4). The flowers are arranged in dense heads. The plants are variable with respect to their essential oil composition, being either carvacrol or thymol-rich. Despite its name, Spanish oregano is common throughout the Mediterranean area.

Turkish Oregano

Origanum onites L. (*O. smyrnaeum* L., *Majorana onites* (L.) Bentham) is a perennial plant, up to 100 cm in height, of the east Mediterranean area. It is abundantly found in many of the Greek islands of the Aegean, extending to western and southern Turkey. Turkish oregano is distinguished from other *Origanum* species by the form of inflorescence (spikes arranged in false corymbs) and the one-lipped calyces. Hybrids between *O. onites* and *O. vulgare* subsp. *hirtum* are often found in the wild. Turkish oregano yields a smaller amount of essential oil (up to 4%) than Greek oregano, and its carvacrol content may reach the 77% of the total oil. It is mainly collected from the wild, but it is also cultivated in west Turkey.

Other species of the Labiatae, locally used as oregano or za'atar (the name used by the Arabs) condiments, are *Satureja thymbra* L., *Thymbra spicata* L., *Origanum compactum* Bentham, *O. elongatum* (Bonnet) Emberger & Maire and the 'Biblical hyssop,' *O. syriacum* L. (syn. *O. maru* L., *Majorana syriaca* (L.) Kost.); all of them are native in the Mediterranean region. Mexican oregano consists of dried leaves of *Lippia graveolens* Humb., Bonpl. & Kunth, a species belonging to the family Verbenaceae. Mexico is the main USA supplier.

Origanum dictamnus L. (Dittany of Crete), an endemic species of the Greek island Kriti (Figure 5) and well known since antiquity for its therapeutical

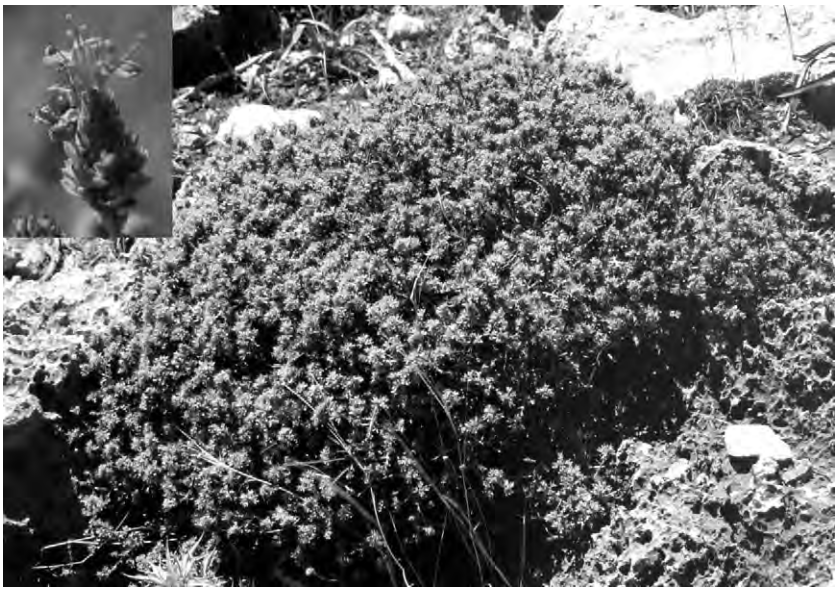

Figure 4 (see color plate 83) *Coridothymus capitatus*.

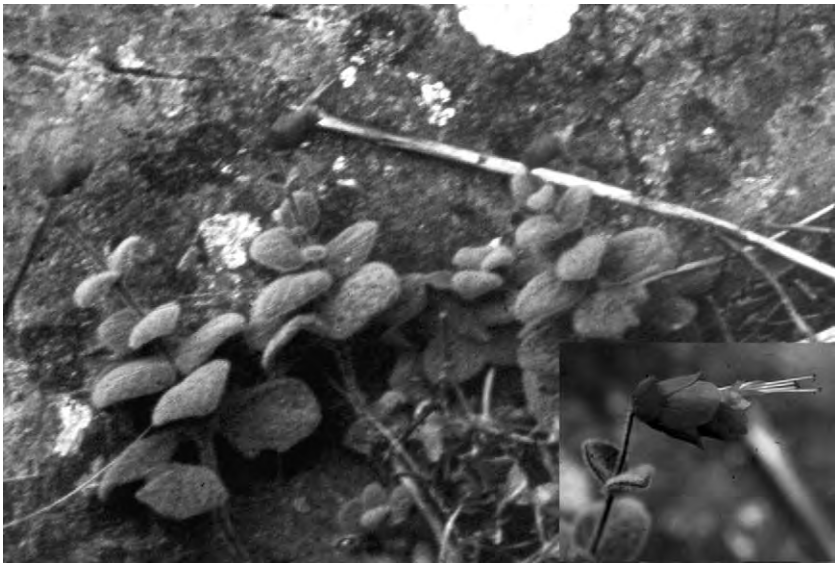

Figure 5 *Origanum dictamnus*.

properties, is employed in the flavoring of alcoholic beverages (Martini).

Rosemary

Rosemary (*Rosmarinus officinalis* L.) is an evergreen shrub, up to 2 m high, native to the Mediterranean region, and cultivated in several European countries and in the USA. It has linear leaves green on the upper surface, with numerous branched hairs making their lower surface whitish. The pale blue, rarely pink, or white flowers are borne in verticillasters produced in the axils of the leaves.

Dried rosemary leaves are aromatic and yield a slightly camphoraceous odor when crushed. They are used to flavour salads, vegetable dishes, soups, meat dishes, sausages, and sauces. Rosemary oil, widely employed in cosmetic products, sometimes replaces the dried leaves in the flavoring of food products. α-Pinene, 1,8-cineole, camphor, and borneol are usually the main rosemary oil constituents. Oils with a decreased amount of α-pinene and

camphor and increased amounts of 1,8-cineole and borneol are judged to be of a better quality.

Sage

The vernacular name 'sage' is ascribed to different aromatic species of the genus *Salvia*, which produce essential oils rich in 1,8-cineole, α- and β-thujone, or camphor. The dried sage leaves are highly aromatic and widely employed in the flavoring of sauces, pork and other meats, poultry stuffings, ground meats (hamburgers), baked fishes, soups, sausages, and canned foods. The essential oils have a fresh-herbaceous odor and are used in perfumery (industrial, soap, or cosmetic perfumes), in flavoring formulations for alcoholic beverages, liqueurs, canned meat, or spicy sauces; they also find application in pharmaceutical preparations.

Dalmatian Sage

Dalmatian sage consists of dried leaves of *Salvia officinalis* L., the most widely used sage plant. Its native occurrence is restricted to Italy and the western part of the Balkan Peninsula, but it is widely cultivated and naturalized in parts of southern and south central Europe. Cultivated as a culinary herb or as an ornament, it may be found all over Europe, where it was possibly introduced by the Romans in ancient times or by monks during the Middle Ages. It is a low (up to 60 cm) shrub, with oblong leaves, white-pubescent beneath, grey–green to silverous above. The violet–blue, pink, or white flowers are arranged in axillary verticillasters (Figure 6).

Salvia officinalis oils are mainly characterized by the predominance of α- and β-thujone (up to 65% of the total oil). However 1,8-cineole and camphor are often detected in considerable amounts (up to 22 and 38% of the total oil, respectively). In commercial grounds, oils with a combined thujyl ketone (α- and β-thujone) content more than 30% and camphor content less than 20% are most valued. Dalmatian sage oil is mainly produced along the Adriatic coast of the former Yugoslavia and Albania, whereas smaller quantities are distilled in Bulgaria, France, Turkey, and Germany.

Greek Sage

Greek sage is obtained from the foliage of *Salvia fruticosa* Miller (syn. *S. triloba* L. fil.). It is an east Mediterranean shrub, extending westwards to south Italy. Introduced by the ancient Greeks and Phoenicians, it was cultivated in the Iberian Peninsula, where nowadays it is naturalized. Greek sage is distinguished from Dalmatian sage by the yellowish green, often three-lobed leaves and the appressed white-tomentose stems.

Although there is a remarkable variation in wild-growing plants, Greek sage oils are generally characterized by high levels of 1,8-cineole (up to 66% of the total oil) and low levels of thujyl ketones and camphor. *S. fruticosa* has been renowned since ancient times for its pharmaceutical properties. Nowadays, the leaf decoction is used in traditional medicine for the treatment of several diseases. The essential oil is a folk remedy for toothache and intestinal complaints.

Spanish Sage

Salvia lavandulifolia Vahl, a west Mediterranean shrub up to 60 cm high, is employed as a flavoring agent under the common name 'Spanish sage.' It can be distinguished from the other sage species by its narrowly oblong to linear leaves. Spanish sage oil is mainly produced in Spain; it is characterized by the preponderance of 1,8-cineole and camphor (up to 59 and 37% of the total oil, respectively), whereas thujyl ketones are detected in negligible amounts (up to 1.3% of the total oil).

Figure 6 (see color plate 84) *Salvia officinalis*.

Cretan Sage

Salvia pomifera L. (Cretan sage) is an east Mediterranean shrub restricted to Greece and Turkey, and mainly distinguished by its membranous, often purple calyces. Cretan sage leaves are exploited on a local scale, often alternatively to Greek sage. However, its oils are always characterized by extremely high amounts of thujyl ketones (58–83% of the total oil), whereas the amounts of 1,8-cineole and camphor are negligible.

Salvia sclarea L. (Clary sage) is a short-lived perennial herb, widely distributed in Europe and South-west and Central Asia. Although it is called sage, its essential oils are quite different from the other sage species, being rich in linalyl acetate and/or linalool (up to 80% of the total oil). Clary sage oil is mainly produced in the former USSR, USA, Bulgaria, and France, whereas minor amounts are distilled in India, Switzerland, Morocco, and several other countries. It is widely used in perfumery but also in flavoring wines and other alcoholic beverages.

Savory

Plants of two species of *Satureja*, which produce essential oils rich in carvacrol, are employed mainly as culinary herbs under the common name savory: *S. hortenis* L. and *S. montana* L. *S. hortensis*, known as Summer savory, is an annual plant, up to 35 cm high, native to south Europe and Anatolia, and widely cultivated as a pot herb. The main area of production is France. It has linear leaves and white or lilac flowers arranged in verticillasters, borne in the axils of the upper leaves. *S. montana*, known as winter savory, is a native plant of south Europe. It belongs to a group of related species, growing wild in parts of the Mediterranean region. Among them, only *S. montana* is used on a larger-than-local scale. It is a small shrub, up to 60 cm high, with lanceolate leaves and white or pale pink flowers arranged in verticillasters, in the axils of the leaf-like bracts.

The value of the two savories is in the high carvacrol content, which gives a fresh, herbaceous, spicy odor with slightly sharp notes, reminiscent of oregano. The two herbs have a similar odor, though summer savory's odour is thought to be more rounded and harsh. They are used to flavor foods, such as sauces, soups, sausages, and canned meats. Savory oil sometimes replaces the dried leaves in food flavoring. The essential oils of *S. montana* wild growing plants are variable and, besides carvacrol, may have high amounts of *p*-cymene, thymol, linalool, borneol, or bornyl acetate.

Thyme

Under the name 'thyme,' plants of *Thymus* species, which produce essential oils rich in thymol, are employed for culinary or medicinal purposes. The species mostly used are *Th. vulgaris* L. (Figure 7) and *Th. zygis* L. and occasionally other species, like *Th. serpyllum* L. or *Th. pulegioides* L.

The herbs have a fragrant odor and a pungent taste, and are used in several food products, such as soups, meat, liver and pork sausages, fish, and poultry

Figure 7 *Thymus vulgaris*.

dressings. Thyme oil is widely employed in many pharmaceuticals and oral preparations. The essential oils produced by wild growing plants of most *Thymus* species are particularly variable. Several compounds have been found as main oil constituents of *Th. vulgaris* and *Th. zygis* plants grown in the wild, such as carvacrol, 1,8-cineole, geraniol, linalool, terpinen-4-ol, α-terpineol, or *trans*-sabinene hydrate.

See also: **Essential Oils**: Properties and Uses; Isolation and Production; **Herbs**: Herbs and Their Uses

Further Reading

Anderson IB, Mullen WH, Meeker JE *et al*. (1996) Pennyroyal toxicity: Measurement of toxic levels in two cases and review of the literature. *Annals of Internal Medicine* 124: 726–733.

Boelens MH and Boelens H (1997) Chemical and sensory evaluation of three sage oils. *Perfumer & Flavorist* 22: 19–40.

Harborne JB, Baxter H and Moss GP (1996) *Dictionary of Plant Toxins*, pp. 310–311. Chichester, UK: Wiley.

Harley RM and Reynolds T (eds) (1992) *Advances in the Labiatae Sciences*. London: Royal Botanic Gardens, Kew.

Kokkini S (1991) Chemical races within the genus *Mentha* L. In: Linskens HF and Jackson JF (eds) *Essential Oils and Waxes. Modern Methods of Plant Analysis*, vol. 12, pp. 61–75. Berlin: Springer.

Lagouri V, Blekas G, Tsimidou M, Kokkini S and Boskou D (1993) Composition and antioxidant activity of essential oils from oregano plants grown wild in Greece. *Zeitschrift für Lebensmitteluntersuchung und -forschung* 197: 20–23.

Lawrence BM (1979–1995) *The Essential Oils 1976–1994*, vols. 1–5. Wheaton: Allured Publishing Corporation.

Padulosi S (ed.) (1997) *Oregano. Promoting the Conservation and Use of Under Utilized and Neglected Crops. 14*. Proceedings of the IPGRI International Workshop on Oregano, 8–12 May 1996, CHIEAM, Valenzano (Bari). Rome: IPGRI.

Werker E (1993) Function of essential oil-secreting glandular hairs in aromatic plants of the Lamiaceae – a review. *Flavour and Fragrance Journal* 8: 249–255.

Herbs of the Umbelliferae

L A Shelef, Wayne State University, Detroit, MI, USA

Background

Herbs of the family Umbelliferae are aromatic plants grown primarily for their use in flavoring of foods, liqueurs, and confectionery. The name Umbelliferae predates Linnaeus and the name Apiaceae was proposed by John Lindley in 1830 for the same family. Both Apiaceae and Umbelliferae are in current use.

The majority of the Umbelliferae plants are indigenous to the Mediterranean countries, where they grow naturally. Brought to northern Europe, they have adapted to climatic and growing conditions of many other countries where they are now cultivated. A point of interest in regard to members of the Umbelliferae plants is that some of them are sources of spices (the seeds) and herbs (leaves), as well as being consumed as vegetables (the bulbs of fennel and celery, for example). The characteristic odor and flavor of members of this family are associated with the major specific essential oil or oils in the edible portions, and particularly in the fruit or seed, in which the content is highest. Within a species, the content of the oil varies, depending on the plant variety, climatic and soil conditions, and storage or processing methods.

Most of the major volatile oils of the Umbelliferae seeds possess medicinal properties that were valued in the past. They have been considered mainly for their stimulant, carminative, pectoral, and palliative properties. Today, the volatile oils are prized less for their medicinal uses than for their flavor and aroma. While their concentration in the seeds is higher than in the leaves, the composition is often different, making the aroma of the herbs generally more delicate.

Production practices of spices and herbs vary widely and these affect the microbial content of the products. Aerobic plate counts of 10^6 microorganisms per gram at retail are common in many spices. Mold growth and formation of mycotoxins, the fungal metabolites, are also of concern in moist products. Cleaning processes and treatment with ethylene oxide eliminate a large portion of the microorganisms present, as well as insects and insect eggs. Propylene oxide is similar to ethylene oxide, although its penetrating ability is lower. While effective, fumigation with gases has been questioned because of toxicity. Irradiation treatment is also effective, and a dose of up to 30 kGy is allowed for dry spices and herbs. The treatment is allowed under Codex Alimentarius, and doses and labeling requirements vary from country to country. Specifications for acceptable contamination levels in products are provided by trade associations (e.g., American Spice Trade Association, European Herb Trade Association, British Herb Trade Association).

Essential oils of a large number of spices have been shown to possess antibacterial and antifungal properties. The ones found in Umbelliferae plants are anethole and, to a lesser extent, also eugenol.

Anethole, a major constituent of the essential oils of anise seed and fennel, has been reported to be effective against various foodborne microorganisms, including pathogens such as *Salmonella* spp. and *Staphylococcus aureus*. Essential oils and extracts of cumin and dill have been reported to be effective in preventing spoilage of wiener-type sausage.

Spices are also known for their antioxidant activity. While umbelliferous seeds such as caraway, celery, cumin, and fennel have some antioxidant properties, they are not as effective as spices of other families in preventing rancidity.

Seeds of the major members of the Umbelliferae family are discussed first in this article, after which the leaf types of the family are reviewed.

Umbelliferae Seeds

Seeds of the Umbelliferae family include anise, caraway, celery, coriander, cumin, dill, and fennel. Drying the seeds is essential for development of full aroma and for storage stability. In traditional methods, the plants are sun-dried after harvest. Commercially dried products are dehydrated under controlled conditions to a specified moisture content, uniform color and, when reconstituted, they retain the flavor and color. The proximate composition of umbelliferous seeds and their mineral content are shown in Table 1.

Grinding the seeds releases the natural flavors and permits control of flavor intensity and uniform distribution. It also improves the appearance of the food products and the economy in use of the seed. Spice-grinding techniques influence the flavor. Heat evolved during grinding causes flavor deterioration. In modern milling techniques such as cryomilling, the process is accomplished at very low temperatures, generally below −70 °C, by direct contact with liquid nitrogen refrigerant. In addition to the low temperature, the process protects the spice oils from oxidation by replacing the air in the mill. It also results in very fine grinding of the spices since the low temperatures render the seeds very brittle. Freeze-ground spices display better retention of aroma, flavor, and color. Moreover, the treatment lowers the microbial load.

Spices are stored in sealed containers under dry, cool conditions. Evaporation of aroma from the intact seed is slow, but once the seed is ground, evaporation is rapid, and oxidation and degradation take place. The shelf-life of the spice is shortened, and storage for no longer than 3 months is generally recommended.

Essential oils are the chief components of spices. The essential oil compounds and contents of the umbelliferous seeds are presented in Table 2. While each of the seeds is characterized by its essential oil composition, similarities in flavor are ascribed to the presence of identical chief compounds. For example, anethole is the chief component of anise seed and fennel, and these two spices are similar in flavor.

The volatile essential oils are extracted from the seeds by steam distillation. Using stainless-steel equipment, steam passes through the ground seeds and carries with it the volatiles. Extraction time varies, depending on the plant, from 6 to 12 h. After cooling, the condensate consists of a mixture of water and oil. The oil fraction is separated by gravity, filtered and dried. Oxidation of the unsaturated compounds (terpenes and sesquiterpenes) is avoided. Terpeneless and sesquinterpeneless extracts with improved storage stability are also produced.

Oleoresins are concentrated, viscous, resinous extracts of the flavor ingredients, obtained by extracting the volatile oils with organic solvents such as hexane and ethylene dichloride, using three to four volumes

Table 1 Proximate composition of umbelliferous seeds (per 100 g edible portion)

Nutrient	*Anise*	*Caraway*	*Celery*	*Coriander*	*Cumin*	*Dill*	*Fennel*
Water (g)	9.5	9.9	6.0	8.8	8.1	7.7	8.8
Protein (g)	17.6	19.8	18.1	12.4	17.8	16.0	15.8
Fat (g)	15.9	14.6	25.3	17.8	22.3	14.5	14.9
Total carbohydrate (g)	50.0	49.9	41.3	55.0	44.2	55.2	52.3
Ash (g)	7.0	5.9	9.3	6.0	7.6	6.6	8.2
Fiber (g)	14.6	12.6	11.8	29.1	10.5	21.1	15.7
Calcium (mg)	646	689	1767	709	931	1516	1196
Iron (mg)	37	16	45	16	66	16	19
Magnesium (mg)	170	258	440	330	366	256	385
Phosphorus (mg)	440	568	547	409	499	277	487
Potassium (mg)	1441	1351	1400	1267	1788	1186	1694
Sodium (mg)	16	17	160	35	168	20	88
Zinc (mg)	5	6	7	5	5	5	4

Data compiled from USDA (1977) *Composition of Foods*, USDA Agriculture Handbook 8–2. Washington, DC: US Department of Agriculture, and other sources.

Table 2 Chief flavor components of umbelliferous seeds

Seed	*Oil content (%)*	*Chief oil, percentage of total*	*Other components*
Anise	2–3	Anethole, 80–90	Pinene, limonene, methyl chavicol, anisaldehyde, aniseketone
Caraway	3–6	D-carvone, 50–65	D-limonene
Celery	1.5–2.5	D-limonene, 60	Selinene, sedanolide, sedanonic anhydride
Coriander	0.1–0.5	D-linalool, 45–70	1,4-*p*-menthadien-7-al
Cumin	2–3	Cuminaldehyde, 35–50	Pinene, *p*-cymene, α-terpinene, cuminyl alcohol
Dill	2.6–3.7	Carvone, 43–63	D-limonene, carvacol, eugenol
Fennel	2–4	Anethole, 50–60	Fenchone, phellandrene, pinene, methyl chavicol, anisaldehyde, anisic acid

Data compiled from Lewis YS (1984) *Spices and Herbs for the Food Industry*, Orpington: Food Trade Press, and other sources.

of the solvent to one volume of the seeds. Steps in the process include grinding, extraction, and distillation. The solvent percolates down through the ground spice, carrying with it the volatile essential oils, the nonvolatile fatty oil (fixed oil), pigments, and resins. The major fatty acids in the fixed oil are oleic, linoleic, petroselinic (*cis*-6-octadecenoic), and palmitic, and their concentration depends on the seed. Heat and vacuum are used to remove the solvent.

Encapsulated oils are available, in addition to spice extracts and oleoresins. These consist of spray-dried emulsions of essential oils with a food-grade, modified, or soluble gum, which forms a protective capsule, preventing evaporation and oxidation. The dry product is the size of fine salt and is water-soluble. It contains about 20% essential oil or oleoresin and has applications in dry mixes such as dehydrated soups and sauces. The oils are released upon dissolution of the capsule.

Discussion of specific umbelliferous seeds follows. It covers historic background, habitat and cultivation, plant description, seed composition, and uses in foods and medicine.

Anise Seed (*Pimpinella anisum* L.)

Background Anise seed is reported to have been found in Egypt as early as 1500 BC. It was once considered an aphrodisiac and a cure for stomach disorders and shortness of breath. It was first used to flavor foods by the Romans in the Middle Ages. The name *Pimpinella* derives from *dipinella*, or twice-pinnate, referring to the shape of the leaves.

Habitat and cultivation Anise is native to Egypt, Greece, and Asia Minor; from there it spread in the Middle Ages to warm parts of Europe, and from there to India, South America, China, and the USA. Flowers appear after 3 months, and green fruit a month later. Yields of about 500–700 kg of seed per hectare are obtained.

Description The plant is an annual, about 50 cm high, with a branching stem, feathery leaves, yellow-white flower clusters, and green fruit (Figure 1). Harvested when the seeds mature, the plants are dried and the seeds are threshed and dried. The anise seed is ovoid in shape, 5 mm long, gray-brown, sweet, spicy, and aromatic.

Composition Upon distillation, anise seed yields 1.5–3.5% essential oil, approximately 90% of which is anethole, with the characteristic anise odor. It crystallizes on cooling to 15–20 °C. The crystallization temperature, which increases with anethole content, is used to evaluate the oil quality. The remaining

Figure 1 (see color plate 85) Herbs of the Umbelliferae: anise.Reproduced from Herbs/Herbs of the Umbelliferae, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

portion consists of 2% methyl chavicol and terpenes such as limonene, pinene, and others. In commercial use, 2.5 kg of the essential oil is equivalent to 100 kg of freshly ground anise seed in seasoning mixes.

Uses Anise seed is used for flavoring baked goods such as biscuits, cakes, sweet rolls and bread, soups, beverages and confectionery (e.g., licorice), and in preparation of cordials and liqueurs. Several Mediterranean alcoholic drinks are flavored with anise, which dominates them. Pastis in France (e.g., Pernod, Ricard, Berger) and anisette contain anise, as does ouza in Greece and arrak in the eastern Mediterranean countries. Meat products, especially Italian style, are flavored with anise seed, as are Indian foods, particularly vegetables and fish dishes from Bengal. The slightly roasted seed is commonly chewed after meals to sweeten the breath.

Medicinal activity The anise seed has aromatic, carminative, and stimulant properties. It is used in cough medicine, as a palliative, to relieve flatulence and colic, and in medicine designed to increase lactation. The oil is considered a good antiseptic.

Caraway (*Carum carvi* L.)

Background The word 'caraway' refers to its origin in the province Caria in Asia Minor. In ancient times, caraway seeds were used to mask the breath and to relieve stomach problems. Bread containing caraway seeds was eaten in ancient Rome.

Habitat and cultivation Caraway is indigenous to Europe, India, North Africa, and other countries. It is cultivated primarily in The Netherlands as well as in Morocco, Germany, Finland, Russia, Norway, Canada, and the USA. Fruit appear within 2 months after the flowers. The seed yield in The Netherlands is about 2200 kg ha^{-1}.

Description The plant is a biennial that grows to a height of 45–100 cm, bearing umbels of white flowers, which are followed by fruit. The plants are harvested when the fruit are still green. The fruit are allowed to ripen, then dried and threshed. The seeds, 6 mm long, are dark brown, slightly curved, ridged, and translucent. The aromatic odor evolves when the fruit is bruised, and the taste is pleasant. The wild seed is smaller than the cultivated one and has a stronger flavor.

Composition The seeds contain 3–6% volatile oil. Distilled oil is prepared from the Dutch, Norwegian, and Russian seeds. Carvone is the chief constituent, comprising 50–65% of the total volatile oil, which also contains the oxygenated oil, carvol. The spent seed after distillation contains protein and fat and is used for feed.

Uses The seeds are used in breads, rolls, and cakes, confections, cheese and soups, and to flavor cabbage, meat dishes, fish, salads, vegetables, and apples. The oil is used in alcoholic liqueurs, particularly in the German cordial kummel. It is also used in nonalcoholic beverages. In addition to the seeds, the tender leaves are used in soups, and the young roots, similar to parsnips, are eaten cooked.

Medicinal activity In the past, caraway was used as a carminative and for relief of indigestion.

Celery Seed (*Apium graveolens* L.)

Background The medicinal properties of celery seed are mentioned in Homer's *Odyssey* (eighth century BC). Its use to flavor foods is first recorded in France at the beginning of the seventeenth century.

Habitat and cultivation Celery is a biennial native to the Near East, southern Europe, and North Africa. It is cultivated for its seeds in India, France, the UK, The Netherlands, the USA, and other countries. Seed yields of around 620 kg ha^{-1} are considered good. World production is about 6000 t, half of which is produced in India.

Description The plant grows to a height of 50–100 cm, with coarsely toothed and segmented leaves and umbels of white flowers. The ovoid seeds are about 1–3 mm in length, brown and ridged. They are crushed and steam-distilled. The root, celeriac, has a pleasant, mild celery flavor and is popular as a vegetable in Europe.

Composition Celery seeds contain 2.5–3.0% volatile oil. Sedanolide (lactone of sedanolic acid) and sedanonic anhydride (1.3%) are responsible for the characteristic odor of the oil, which also contains D-limonene (60%) and selinene (10–20%).

Uses The herb seeds, volatile oil, and oleoresins are widely used in soups, sauces, pickles, eggs, salads, tomato products, and meats. Celery salt is a mixture of ground celery seed and sodium chloride, used as a seasoning in salads, fish and meat dishes, and other foods.

Medicinal activity Celery seeds are said to be carminative, stimulant, diuretic, tonic, and inducers of sleep.

Coriander (*Coriandrum sativum* L.)

Background Coriander was known as far back as 5000 BC and is mentioned in the Bible in Exodus 16:31. The use of the seeds at burials was recorded in 1550 BC. Its use by the Romans and by Hippocrates and other Greek physicians is documented. The Chinese have cultivated coriander since the fourth century, and its use in India and Europe dates back to early times. The name 'coriandrum' is derived from *koros*, a bug, referring to the disagreeable fetid smell of the leaves.

Habitat and cultivation Now indigenous to southern Europe, the plant was brought by the Romans from the east (Egypt). It is also cultivated in North Africa, Malta, India, Central America, Russia, and the USA. The yield of dry seed is 1200–2400 kg ha^{-1}. World production is estimated at 300 000 t annually, of which India produces more than 25%. At least a third of the crop is distilled for the oil. The deodorized fat (about 20% of the seed) is used as lubricant or for soap making. In India the seeds are subjected to mild toasting before use, to improve the flavor.

Description The plant is an annual, 30–100 cm high, branched, bright-green, with erect stems. The flowers are short-stalked umbels, 5–10 rays, pale mauve or white. The seeds are globular, beaked, finely ribbed, yellow-brown, 5 mm in diameter, with five longitudinal ridges. They mature 3 months after planting. The harvested plants initially have a disagreeable scent, which is lost after drying, and then become fragrant. There are two types of coriander, distinguished by the fruit diameter: 3–5 mm (*Coriandrum sativum* L. var. *vulgare*) and 1.5–3 mm (*C. vulgare* L. var. *microcarpum*).

Composition The seeds contain about 1% volatile oil. The alcohol D-linalool is responsible for the aroma of coriander, and a content of 60% or more is desirable. D-α-pinene, β-pinene, and α-terpinene, geraniol, borneol, decylaldehyde, and acetic acid are other constituents of the oil. The essential oil from the smaller fruits is superior in flavor and yield. Overall, coriander oil is more stable than other oils in the group.

Uses In the northern countries of Europe, coriander is added to bread, pastries, cheese, soups, vegetables, and stews. It is further used to flavor processed meat products such as liver sausage and bologna. In Peru, the spice is common in almost all dishes. It is frequently used in soups in Peru and Egypt, and it is an ingredient of curry powder.

Medicinal activity Coriander seeds are primarily used to disguise unpleasant-tasting medicine because of their aromatic taste and characteristic odor. The seed is a stimulant and carminative, and in the past it was considered an aphrodisiac. Extracts from coriander have been reported to improve glucose utilization, as shown by glucose tolerance tests.

Cumin (*Cuminum cyminum* L.)

Background Cumin is mentioned in the Bible in the books of Isaiah and Matthew, and in works of Hippocrates and Dioscorides. It was highly regarded as a condiment by the ancients, and is still used in medicine in India.

Habitat and cultivation Cumin is indigenous to Egypt, and is cultivated in the Mediterranean, India, China, Morocco, southern Russia, and other countries. It grows well on fertile, sandy loam soil free of weeds.

Description The cumin stem is slender and branched, 30–60 cm high. The deep-green leaves are divided into long narrow segments. The flowers are small, rose or white, in stalked umbels with only four to six rays, each about 8 mm long. The seeds are oblong, about 5 mm long, resembling caraway seeds, but lighter in color, bristly rather than smooth, and almost straight, with nine ridges overlapping as many oil channels (vittae). Odor and taste are similar to caraway but stronger, spicy, and curry-like.

Composition Cumin seeds contain 2–4% essential oil, consisting of 40–65% cuminaldehyde, and cymol or cymene and cyminol.

Uses Cumin is an ingredient of curry powder, and is essential in most Indian, Egyptian, and Turkish curries. It is also used in Mexican foods, in meat, poultry, and vegetables. In European countries such as Switzerland and The Netherlands, cumin is used to flavor cheese, while the Germans use it to flavor sauerkraut, bread and other baked products, and kummel liqueur.

Medicinal activity Cumin is considered a stimulant, antispasmodic, carminative, and a remedy for colic, but is not used because of the disagreeable flavor.

Dill (*Anethum graveolens* L.)

Background Dill is referred to in the Bible in Matthew by its original Greek name *Anethum*, and was used in the Middle Ages by magicians against witchcraft. The name is derived from the old Norse

word *dilla*, meaning to lull or soothe, referring to the carminative properties.

Habitat and cultivation Dill is a native of the Mediterranean and Russia. It is cultivated in many European countries, the USA and Canada, India, Pakistan, and Japan, and grows wild in Spain, Portugal, and Italy.

Description The plant, called dill weed, is 60–75 cm high, similar to but smaller than fennel, with feathery leaves on sheathing foot-stalks. The plant has generally one stalk, finely divided leaves, and flat umbels with yellow flowers and small, rolled-inward petals. The numerous flat seeds are approximately 5 mm in length, pungent and bitter. The whole plant is aromatic.

Composition The seeds contain 2.6–4% volatile oil, 40–60% of which is carvone. Additional constituents are dihydrocarvone, D-limonene, α-phellandrene, α-pinene, and dipentene.

Use Dill seed is used in pickles, pickled vegetables, sour cream- or mayonnaise-based sauces, cheese spreads, salad dressings and salads, a variety of vegetables, soups and fish sauces. Its use in seafood is popular in the Scandinavian countries. Some processed meat products, confections, and beverages are also seasoned with dill.

Medicinal activity In addition to the magical powers attributed to dill in ancient times, it was considered a palliative of stomach pains.

Fennel (*Foeniculum vulgare*)

Background The name 'foeniculum,' used by the Romans, is derived from the Latin word *foenum* for hay. This name led to another – *fanculum* – during the Middle Ages, and later to the popular name *fenkel*. The plant was assigned by Linnaeus to the genus *Anethum*, but was later placed by De Candolle in a new genus, *Foeniculum*. *Foeniculum vulgare* is the common or wild variety, *F. dulce* is the sweet (French) fennel, and *F. capillaceum* is the garden variety. Fennel was cultivated by the ancient Romans for its aromatic seeds, used as a vegetable and against witchcraft. It is mentioned in Spanish agriculture records dated 961 AD. Today, all parts of the fennel plant are edible – the seeds, leaves, stalks, and bulb.

Habitat and cultivation Fennel is indigenous to the Mediterranean and grows wild in most parts of temperate Europe, Egypt, India, Japan, China, the USA, and Argentina. It is a hardy perennial that grows particularly well on limestone soil. The yield of seeds is about 1200 kg ha^{-1}.

Description The plant grows to 1.2–1.5 m in height, with finely divided bright-green leaves and umbels of yellow flowers. The cultivated fennel (*F. capillaceum* or *F. officinale*) is taller than the wild variety, with larger stems, bearing fruit 6–12 mm long, double the size of the wild fennel. Several varieties of fennel seed are known, varying in size, volatile oil content, and taste. The most valued seed is elliptical, slightly curved, and pale gray-green in color, with a sweet, fragrant aroma similar to that of anise or licorice. In addition to the perennial varieties, *F. dulce* is an annual, native of Italy, about 30 cm high, with large, finely cut leaves and pale green-white stalks.

Composition Good varieties of fennel contain 4–6% volatile oil, of which the principal constituents are anethole (50–60%) and fenchone (19–22%). Fenchone is colorless, with a pungent camphoraceous odor and bitter taste. Small amounts of α-pinene, phellandrine, camphene, dipentene, methyl chavicole-hydroxyphenylacetone, and limonene are also present in the oil. While fennel grown in eastern European countries and Russia contains 4–5% volatile oil, French fennel yields only 2.1% oil, with lower levels of anethole and a sweet taste owing to absence of fenchone. Indian fennel contains only 0.72% volatile oil.

Uses Fennel seeds, as well as the leaves, are used with fish and seafood dishes. The seeds are used in soups, German and Italian breads, Polish bortsch, Italian sausage such as hot and sweet pepperoni, pickles, salads, poultry and meat dishes, and fruit pie fillings. The Spanish use the seed frequently in cooking and baking, and it is employed in flavoring cordials and liqueurs. Commercially, fennel seed or the extract is used to flavor medicine, soaps, and perfumes.

Medicinal activity Fennel oil has carminative properties, similar to those of dill and anise oil, and is considered beneficial in treating chest, spleen, and kidney diseases. The crushed seeds steeped in boiling water yield fennel water, which is helpful against flatulence. In India, toasted fennel seeds are chewed after meals.

Umbelliferae Leaf Types

Parsley is the most popular leaf-type herb of the Umbelliferae family. Similar, but less recognized

outside France, is chervil. The fresh leaves of parsley and chervil are most frequently used finely chopped on meat, fish, soups, and salads. Leaves of three of the herbs discussed in the first half of this article, namely coriander, dill, and fennel, also find many uses in the food industry and home food preparation. The composition of the fresh leaves is similar; they contain approximately 85% water, 3% protein, 1% fat, 10% carbohydrates, and 2% ash. The flavor substances in fresh leaves are generally present at lower concentrations than those found in seeds of the Umbelliferae plants. For example, dry dill seeds contain 2.6–4% volatile oil, whereas dill leaves or weed contain only 0.3–0.6% volatile oil. The composition of the oil differs as well.

Herb leaves are at their best when used fresh, and ideally should be used as soon as possible after picking. Their limited shelf-life can be extended by appropriate packaging and cold storage in the dark, since the organoleptic properties deteriorate when exposed to air and light. The leaves are preserved by drying to 5% moisture content or less. The composition of dehydrated chervil, coriander, dill, and parsley leaves is presented in Table 3. Drying conditions influence the flavor quality, and improved products are obtained by freeze-drying or lyophilization, as opposed to the conventional drying procedures, since exposure to heat impairs the flavor characteristics. In freeze-drying, the leaves are first frozen and placed in a vacuum, and the ice is sublimated. The freeze-dried leaves contain 1–3% moisture and have a fresh-weight replacement of 1:9. They are spongy and similar in volume and shape to the original material. The bulk density is about 0.3 g cm^{-3}. Stored at a cool temperature in airtight containers, the leaves are easily reconstituted when placed in cold water. Flavor and texture are excellent since maximum temperatures during final dehydration are less than 50 °C. Microwave freeze-drying of spice leaves has also been studied.

Properties of the leaves of chervil, coriander, dill, fennel, and parsley are discussed below.

Table 3 Proximate composition of selected umbelliferous dried leaves (per 100 g edible portion)

Nutrient	*Chervil*	*Coriander*	*Dill*	*Parsley*
Water (g)	7.2	7.3	7.3	9.0
Protein (g)	23.2	21.8	20.0	22.4
Fat (g)	3.9	4.7	4.4	4.4
Carbohydrate (g)	49.1	52.0	55.8	51.7
Fiber (g)	11.3	10.4	11.9	10.3
Ash (g)	16.6	14.0	12.5	12.5
Calcium (mg)	1346	1246	1784	1468
Iron (mg)	32	43	49	98
Magnesium (mg)	130	694	451	249
Phosphorus (mg)	450	481	543	351
Potassium (mg)	4740	4466	3308	3805
Sodium (mg)	83	211	208	452
Zinc (mg)	9		3	5

Data compiled from USDA (1977) *Composition of Foods*, USDA Agriculture Handbook 8–2. Washington, DC: US Department of Agriculture, and other sources.

Chervil (*Anthriscus cerefolium* L.)

Background The Greeks and Romans cultivated the plant during the Middle Ages. It was called *myrrhis* because of the similarity of the aroma to that of myrrh, a biblical spice. Used often in fried cooking, it is considered the gourmet's parsley.

Habitat and cultivation Chervil is an annual or biennial herb, native of western Asia and south-western parts of Russia, from which it reached the Mediterranean about 300 BC. It is cultivated in Europe, France, Italy, Spain, the UK, Russia, and the USA.

Description The plant grows to a height of 30–60 cm and has small, fine leaves and small white flowers (Figure 2). The flavor is similar to that of parsley with a hint of anise.

Figure 2 (see color plate 86) Herbs of the Umbelliferae: chervil. Reproduced from Herbs/Herbs of the Umbelliferae, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Uses Chervil is used frequently in French cooking in combination with parsley, tarragon, and chives, which make up the fines herbes. Alone, it is used like parsley, chopped fine and sprinkled on meat, fish, and soups, in salads and as a garnish. It is also used in baked products.

Coriander (*Coriandrum sativum* L.)

For background, habitat, and cultivation, see the section on coriander seeds, above.

Description Also called Chinese parsley or cilantro in Spanish, the plant leaves are quite similar in shape to the leaves of parsley but more jagged (Figure 3). The aroma is fetid. In contrast to the seeds, the leaves contain only about 0.1% volatile oil, in which aliphatic aldehydes such as decylaldehyde are present. These compounds give the leaves their disagreeable odor, which is so different from that of the seeds.

Uses The fresh young leaves are used like parsley to flavor salads and soups in the Middle East, India, China, Japan, Mexico, and South America. They are also used in chutney.

Figure 3 (see color plate 87) Herbs of the Umbelliferae: coriander. Reproduced from Herbs/Herbs of the Umbelliferae, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Dill (*Anethum graveolens* L.)

For background, habitat and cultivation, and description, see the section on dill seed, above.

Composition Distillation of oil from the plant yields about 0.3–0.6% of volatile oil. The carvone content is less than 20%, and the oil contains a high proportion of hydrocarbons. Phellandrene is responsible for the characteristic flavor.

Use The most important use of dill weed is in pickling cucumbers (dill pickles). The fresh leaves are used for flavoring salads, soups, sauces, and fish dishes, and in pastries. Dill butter is prepared by creaming butter with minced fresh or dried dill.

Fennel (*Foeniculum vulgare*)

For background, habitat and cultivation, and description, see the section on fennel seeds, above.

Uses The tender leaves are used for garnishes and to flavor salads, salad dressings, soups, and sauces. Stems of the cultivated fennel are also eaten in salads (Figure 4). The Italian fennel (*F. dulce*) is a popular vegetable in Italy. Cooked and served in cream sauce, it is reminiscent of celery in taste, although sweeter.

Medicinal activity The leaves of fennel were considered in ancient times to be aphrodisiac, to improve eyesight, stop hiccups, assist in losing weight, relieve stomach pains, and cure wheezing. Syrup prepared from fennel juice was given for chronic coughs. Boiled in water, the liquid is used to increase lactation.

Parsley (*Carum petroselinum, Apium petroselinum L., Petroselinum hortense (crispum)*)

Background The name 'petroselinum,' from which the name parsley is derived, is said to have been assigned by Dioscorides. Although Linnaeus assigned it in 1764 to the genus *Apium*, it is assigned today to the genus *Carum*. The herb was used in funeral rites by the Greeks, who held it sacred.

Habitat and cultivation Parsley is indigenous to the Mediterranean regions. It is cultivated in the Mediterranean countries, Europe, the USA, and Japan. The Germination period is 10–21 days and maturation occurs in 70 days.

Description Parsley is biennial or perennial, with dark-green stems growing to a height of about 60 cm. The plain-leaf parsley is by far the hardier, able to survive a severe winter. Several cultivated varieties exist, characterized by the plain leaf, the

Figure 4 (see color plate 88) Herbs of the Umbelliferae: fennel. Reproduced from Herbs/Herbs of the Umbelliferae, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

broad leaf (Hamburg), and the curled leaf (crispum). Of the latter, the ones with compact leaves are most valuable. The Hamburg variety is grown primarily for its enlarged root. The aroma of the leaves is slightly fragrant, and the taste is tangy and pleasant.

Composition The plant yields no more than 0.06% volatile oil, while the seed contains 3.5%. The main constituent of the volatile oil is apiole, and myristicin, D-pinene, and other aldehydes, ketones, and alcohols are also present. Total solids are 11–12%, of which sucrose is 0.2% and glucose is 0.1%.

Uses The most common uses of parsley are as flavoring to sauces, soups, omelets, stuffings, meat, and other dishes, or sprinkled finely chopped over salads and vegetables and in the preparation of parsley butter. Parsley, in combination with other herbs, is a component of the *bouquet garni* in French cooking. Dried leaves are also used for these purposes. The leaves are frequently used as garnish. The broad-leaf variety (Hamburg) is used for flavoring soups and stews, and the roots are also cooked and eaten like parsnips. Commercially, parsley is used in soup bases, seasoning mixes, salad dressings, and similar foods. Dehydrated parsley is available as flakes, as powder and minced.

Medicinal activity In the past, parsley was considered a remedy for a large number of disorders, including the ability to counteract poisons and for treatment of coronary conditions. Parsley, fresh and dehydrated, has carminative and diuretic properties. Parsley tea is prepared by steeping the dried leaves. The chopped leaves have been used externally as a remedy for bites and stings and as a treatment of cancerous tumors.

See also: **Essential Oils**: Properties and Uses; Isolation and Production; **Herbs**: Herbs and Their Uses; Herbs of the Compositae; Herbs of the Labiatae; Herbs Used in Alcoholic Drinks

Further Reading

Charalambus G (ed.) (1994) *Spices, Herbs and Edible Fungi*. New York: Elsevier.

Craker LE and Simon JE (1986) *Herbs, Spices and Medicinal Plants: Recent Advances in Botany, Horticulture, and Pharmacology 1*. Phoenix, Arizona: Oryx Press.

Duke JA (1985) *CRC Handbook of Medicinal Herbs*. Boca Raton, Florida: CRC Press.

Farrell KT (1990) *Spices, Condiments, and Seasonings*, 2nd edn. New York: Van Nostrand Reinhold.

Lewis YS (1984) *Spices and Herbs for the Food Industry*. Orpington: Food Trade Press.

Pruthi JS (1980) *Spices and Condiments: Chemistry, Microbiology, Technology*. New York: Academic Press.

Rubatskey VE, Quiros CF and Simon PW (1999) *Carrots and Related Vegetable Umbelliferae*. New York: CABI Publishing Books.

Simon JE (1990) Essential oils and culinary herbs. In: Janick J and Simon JE (eds) *Advances in New Crops*, pp. 472–483. Portland, Oregon: Timber Press.

Stobart T (1970) *Herbs, Spices and Flavorings*, New York: McGraw-Hill.

Herbs Used in Alcoholic Drinks

V B Veljković and M Z Stanković, University of Niš, Leskovac, Yugoslavia

Introduction

A wide variety of herbs, rich in specific flavoring compounds, mostly from their aromatic essential oils, are used in industries producing alcoholic drinks.

The herbs impart to an alcoholic beverage the characteristic recognized flavor (taste and aroma) and bouquet. These flavoring attributes have traditionally been one of the most significant qualitative factors in the use of alcoholic drinks. In addition, the essential oils from many aromatic herbs, so-called medicinal herbs, have many beneficial effects on metabolism, which are experienced after the controlled use of alcoholic drinks based on such herbs; for example, medicinal wines (*vina medicata*) have been used for a long time.

The commercially important alcoholic drinks based on herbs are as follows: beers, aromatized wines (vermouths and medicinal wines), liqueurs (bitter, fruit), and hard liquors (bitters, brandies, flavored distilled spirits, etc.).

Use of Herbs in the Alcoholic Drink Industry

In the alcoholic drink industry, both wild and cultivated herbs are used; a summary of the herbs used for this purpose, and known all over the world to a greater or lesser degree is presented in Table 1. Because of the need for larger quantities, the cultivation of some herbs is of great importance. Herbs are most often used dried or processed, and rarely fresh. Different parts of the herb are used: aerial parts, flowers, leaves, berries, roots, and seeds. The soluble essences (essential oils and extracts), obtained by processing either fresh or dried herbs or their parts, are also used in the flavoring of alcoholic drinks.

Herbs are used in alcoholic drink production either as flavorings, or as both fermentation substrate and flavoring agent. The flavor of a herb is dependent on its volatile and nonvolatile constituents; the former gives to the herb its distinctive odor, and the latter can affect some gustatory reaction sometimes coupled with a physiological effect.

Most herbs are used only for flavoring purposes. A base alcohol (beer, spirit, and wine) is combined with one or more herbs, known as the herbal formula, to obtain a special flavor. Herbal formulae usually contain more than one herb. One herb may be a major definitive constituent, while the others are added to develop the overall flavor of an alcoholic drink. For example, juniper berries are always used in gin production, but others, such as coriander seed, angelica root, or caraway seed, play a very important role, also as flavour ingredients, in some gin formulae. In a herbal formula, the selection of ingredients with respect to the major constituent is generally limited, and the proportions of constituents are normally constant and commonly not disclosed by manufacturers. The 'right' proportions are achieved through experimentation.

The herbs of appropriate quality are used for alcoholic drink production. The selection is largely based on organoleptic assessment. Color, overall appearance, odor, moisture content, quality, and content of essential oil or other specific characteristics are assessed for the purpose of selection and approval of quality of herbs or their parts. Although the specific flavor profile is within acceptable limits, quantitative variations in the flavor characteristics are often observed among different batches of the same herb material or even between different parts of the herb. Because of these variations, the herb material should contain the useful part of the herb, with a minimum content of the other parts of that or other herbs, and be of natural odor, color, and proper overall appearance, without any trace of decay or spoilage. The distinctive odor of the herb should be typical and 'clean,' with no trace of mustiness. The acceptable moisture content is usually in the range of 10–20%, mostly 12–14% on a dry-weight basis; it should not be too high, in order to reduce the risk of burning during storage, but too low a moisture content means that the herb material is old. The composition, color, and odor of the essential oil are important factors for the quality of an alcoholic drink, while its content is important for the economic production of the drink.

For the flavoring of alcoholic drinks, herbs are commonly used in the following basic ways:

- By maceration of the herb materials in a base alcohol (spirit, wine, etc.).
- By codistillation of the herb materials or their essential oils with a suitable high-quality spirit (usually grain or other neutral spirit).
- By mixing the essences of the herb materials with a spirit or wine.

Details of these processes differ among the various manufacturers, who usually do not disclose them. The maceration process is applied in the production of fruit brandies with aromatic herbs (usually with medicinal properties), bitters made with bittering herbs, and aromatized wines. Herbs rich in essential oils are macerated with a spirit containing 70% ethanol by volume, while the herbs rich in tannins are treated with a 50% spirit. The herb material is immersed in the chosen alcoholic base in a tank for a certain period of time and occasionally agitated. The liquid is drawn off, and the exhausted herb material is pressed to recover the residual liquid. The codistillation process is used in the production of hard liquors, such as gin. The herbs are placed in a still and covered with the base spirit of approved quality and alcoholic strength. The materials are charged loosely or held in

Table 1 Herbs used in alcoholic drinks

Common name	*Scientific name*	*Parts and essences used* [a]	*Description of the part used and the content of essential oil*	*Use in alcoholic drink*
Angelica	*Archangelica officinalis*	Root, essential oil, macerate	Fleshy tap root, 12–25 cm long; odor sweet and musk-like upon storing; taste sweet, becoming slightly bitter and astringent; oil < 1%	Brandies, gins, flavored spirits, liqueurs, vermouths
Anise	*Pimpinella anisum*	Seeds, essential oil, macerate Root	Gray-green to brown seeds, ovate, laterally compressed, pubescent, 3–5 mm long, 2–3 mm wide; odor anethole-like; taste sweet and aromatic; oil 1.5–4%	Brandies, flavored spirits, liqueurs, medicinal wines, vermouths
Arnica	*Arnica montana*	Root, macerate, percolate	Brown, hard, flexible, knotty root, 3–5 mm thick and 9 cm long; odor faint, aromatic; taste bitter and acrid; oil 0.1%	Bitters, flavored spirits, liqueurs
Balm	*Melissa officinalis*	Leaves	Green, rough cordate, serrated; odor of lemon; taste bitter; oil 0.1–0.3%	Liqueurs, flavored spirits, medicinal wines
Basil	*Ocimum basilicum*	Aerial parts	Low, sweet-scented bush; green, slightly toothed leaves; pink or white flowers; odor strong, pungent, sweet; taste peppery; oil 0.4–1.0%	Flavored spirits
Blessed thistle	*Cnicus benedictus*	Aerial parts	Medium height; dark green, deeply lobed, prickly; large, solitary, yellow heads of flowers; odor faint; taste bitter	Liqueurs, vermouths
Buffalo grass	*Hiërochloe odorata*	Aerial parts	Low, creeping range grass	Vodka
Burdock	*Arctium lappa*	Root	Fleshy, beet-like roots, up to 50 cm long, with brown cork; taste slightly bitter	Vermouths
Calamus	*Acorus calamus*	Root, essential oil, macerate, percolate	Brownish root, spongy texture, 12–25 mm thick, sometimes longer than 1 m; odor slightly sweet and aromatic; taste warm, pungent, rather bitter; oil 1.5–3.5%	Brandies (bitter), flavored spirits, liqueurs, medicinal wines, vermouths
Caraway	*Carum carvi*	Seeds, essential oil, macerate	Dark brown seeds with yellow ridges, crescent-shaped, 4–7 mm long, up to 1 mm broad; odor agreeable; taste pleasant, spicy, but sharp; oil 3–7%	Brandies, gins, flavored spirits, liqueurs, medicinal wines
Cardamom	*Eletaria cardamomum*	Seeds, essential oil	Pale to dark red-brown seeds, sickle-shaped, three-to four-sided, 3–5 mm long, 3 mm broad, hard; odor agreeable; taste strong, spicy, and aromatic; oil 5–9%	Brandies, flavored spirits, liqueurs, medicinal wines
Carnation	*Dianthus caryophyllus*	Flowers, essential oil, macerate	Deep pink, single or double flowers; odor clove-like; taste spicy	Brandy (bitter), flavored spirits, liqueurs, vermouths
Celery	*Apium graveolens*	Seeds, essential oil	Dark brown seeds, roundish-ovoid cremocarp; about 1–1.5 mm long, 1.5 mm wide, 0.5 mm thick; odor characteristic; taste strong and somewhat sharp; oil 2.5–3%	Liqueurs
Clammy sage	*Salvia sclarea*	Flowers, leaves, essential oil	Pale blue or lilac; large flowers; grayish-green, oblong, leaves; odour balsamic; taste aromatic and bitter; oil < 2%	Flavored spirits, liqueurs, vermouths
Common marigold	*Calendula officinalis*	Flowers	Light yellow to orange; odor slight and aromatic; taste salty and slightly bitter; oil in trace	Flavored spirits, liqueurs
Coriander	*Coriandrum sativum*	Seeds, essential oil, macerate	Tawny or reddish seeds, small, rounded (2–4 mm), ridged; odor mellow spicy; taste sweetly aromatic; oil 0.4–1.5%	Brandies, gin, flavored spirits, liqueurs, vermouths
Dandelion	*Taraxacum officinale*	Flowers Roots	Bright golden-yellow, tiny florets Cylindrical, tapering, 10–30 cm long, 1–2 cm wide; odor slight; taste somewhat bitter	Flavored spirits, liqueurs, medicinal wines, vermouths

Continued

Table 1 Continued

Common name	*Scientific name*	*Parts and essences used* [a]	*Description of the part used and the content of essential oil*	*Use in alcoholic drink*
Dill	*Anethum graveolens*	Seeds	Pale brown, oval, 3–5 mm long; odor and taste aromatic, pleasant and characteristic; oil < 4%	Flavored spirits, liqueurs
Elecampane	*Inula helenium*	Root, macerate, percolate	Grayish brown, fleshy root, 6 cm thick, up to 50 cm long; odor aromatic, pleasant; taste slightly bitter and pungent	Flavored spirits, liqueurs, medicinal wines, vermouths
European centaury	*Erythraea centaurium*	Flowering aerial parts	Stems, branched above, 20–30 cm in length; green-brown, oblong or elliptic leaves; rose pink, salver-shaped; odor distinct; taste bitter	Bitters, vermouths
Everlasting	*Helichrysum arenarium*	Flowers	Yellow, rarely orange, grape-like: odor faint, aromatic; taste bitter	Liqueurs
Fennel (wild)	*Foeniculum vulgare*	Seeds, essential oil, macerate	Green to yellow-brown, oblong cremocarps; odor anethole-like; taste sweet; becoming bitter; oil 3.5–6%	Flavored spirits, liqueurs, medicinal wines, vermouths
Fenugreek	*Trigonela foenum*	Seeds	Light to dark yellow-brown, rhomboidal; 2–3 mm thick, 3–6 mm long; odor specific; taste slightly bitter and mucilaginous	Vermouths
Galangal	*Alpinia officunarum*	Root, essential oil, macerate	Red brown on drying; hard, branching, nearly cylindrical root; odor pungent and spicy; taste pungent; oil 0.3–1%	Brandies, liqueurs, medicinal wines, vermouths
Garden thyme	*Thymus vulgaris*	Leaves, macerate	Yellow-green to brown-purple, narrow, ovate; odor thymol-like; taste strong, slightly sharp; oil 0.8–2.5%	Flavored spirits, liqueurs, medicinal wines, vermouths
Gentian	*Gentiana lutea*	Root, macerate	Nearly cylindrical, hard root, up to about 90 cm long, yellow-brown to brown; odor characteristic; taste sweet, becoming bitter	Gentian brandy, medicinal wines, vermouths
Germander	*Teucrium chamaedrys*	Aerial parts	Greenish; ovate, crenate leaves; pinkish-purple flowers; odor faintly aromatic; taste bitter and pungent; oil 0.06%	Liqueurs
German chamomile	*Matricaria chamomilla*	Flowers, essential oil, macerate	Yellow conical flower heads with white margins, odor aromatic; taste bitter; oil 0 ~ 5–1 ~ 5%	Liqueurs
Ginger	*Zingiber officinale*	Root, essential oil, macerate	Grey, striped; hand-like with irregular fingers; 8–12 cm long; odor aromatic and agreeable; taste strongly pungent; oil 1.5–3.3%	Liqueurs, medicinal wines, vermouths
		Aerial parts, flowers	Bine, dioecious; large, toothed leaves; small, yellowish green flowers	Vermouths
Hop	*Humulus lupulus*	Seeds, flowers	From the female plant; yellowish green, ovoid strobiles; odor aromatic; taste bitter; oil in seeds, 1–3%	Beers
Horehound	*Marrubium vulgare*	Aerial parts	Greenish; green-brown, crenate leaves; small, white flowers in dense clusters; odor faint, pleasant; taste bitter and pungent	Vermouths
Hyssop	*Hyssopus officinalis*	Aerial parts, essential oil	Fragrant, shrubby herb; dark green, slender leaves; spikes of blue-pink flowers; odor aromatic; taste bitter, minty; oil 0.3–1%	Liqueurs, medicinal wines, vermouths
Juniper	*Juniperus communis*	Fruits, essential oil	Blue-black, globular (5–8 mm) fruits; odor terpentine-like; taste aromatic, spicy and slightly of pine; oil 0.2–2%	Brandies, gin, geneva, Steinhger, borovička,
Lavender	*Lavandula officinalis*	Flowers, essential oil	Bluish-purple flowers; odor aromatic pleasant; taste aromatic, bitter; oil 0.9–3%	Liqueurs
Liquorice	*Glycyrrhiza glabra*	Root, macerate, percolate	Grayish brown root, brown on drying; odor characteristic; taste sweet	Brandies, liqueurs

Continued

Table 1 Continued

Common name	*Scientific name*	*Parts and essences used* [a]	*Description of the part used and the content of essential oil*	*Use in alcoholic drink*
Lovage	*Levisticum officinale*	Aerial parts, flowers, essential oil	Greenish; dark green-yellow, elliptic flowers; odor strong, pleasant taste, slightly spicy	Liqueurs
Lungwort	*Pulmonaria officinalis*	Aerial parts	Greenish; large spotted leaves; clusters of blue or purple flowers; taste astringent	Vermouths
Marjoram	*Origanum vulgare*	Aerial parts	Greenish; tough, woody stem; gray-green, ovate, crenate leaves; tiny mauve or pink flowers grouped in little round clusters; odor aromatic; taste sweet and spicy; oil > 0.1%	Vermouths
Motherwort	*Leonorus cardiaca*	Aerial parts	Very hairy, upright; deeply palmately lobed leaves with serrated margins; taste bitter	Flavored spirits, liqueurs
Orris	*Iris germanica*	Root, essential oil, tincture	Hard root, up to about 10 cm long, 2 cm thick, pale cream; odor violet-like; taste somewhat bitter; oil 0.1–0.2%	Brandies, liqueurs, vermouths
Pansy	*Viola tricolor*	Aerial parts, flowers	Dark green, spoon-shaped, crenate leaves; blue flowers with a short spur; odor slight; taste faintly bitter	Flavored spirits, liqueurs
Peppermint	*Mentha piperita*	Leaves, essential oil, macerate	Ovate-lanceolate, acute serrate leaves; odor strong, distinctive; taste pungent and cooling; oil 1.5–3.5%	Flavored spirits, liqueurs, medicinal wines, vermouths
Roman chamomile	*Anthemis nobilis*	Flowers	Yellow and white blossoms; with double, white florets; odor strong, aromatic; taste bitter; oil about 1%	Liqueurs, vermouths
Roman wormwood	*Artemisia pontica*	Aerial parts	Greenish; leaves twice- or triple-divided; yellow flower heads; odor and taste aromatic; oil about 1%	Liqueurs, vermouths
Rosemary	*Rosmarinus officinalis*	Leaves, essential oil, macerate	Leaves like coniferous needles; odor strong, distinct; taste pungent; oil 1.5–2%	Liqueurs, medicinal wines, vermouths
Saffron	*Crocus sativus*	Flowers, macerate, percolate, tincture	Bright orange-red stigmas; odor penetrating, distinct; taste aromatic, spicy, pungent; oil about 1%	Brandies, liqueurs, medicinal wines, vermouths
Sage	*Salvia officinalis*	Leaves	Grayish-green, oblong, hairy; odor aromatic; taste warm, pungent, slightly bitter; oil 0.5–2.5%	Flavored spirits, liqueurs, medicinal wines, vermouths
St John's wort	*Hyperycum perforatum*	Aerial parts, flowers	Pale green leaves with oil glands; bright yellow flowers; odor distinct; taste astringent; oil < 0.3%	Flavored spirits, liqueurs
Savory (summer)	*Satureja hortensis*	Aerial parts	Small, narrow, dark green leaves growing sparsely; white, tiny flowers; odor aromatic; taste slightly peppery; oil 2–3%	Flavored spirits, liqueurs, vermouths
Star anis	*Illicium verum*	Seeds	Shiny, light brown, hard, wrinkled; odor anise-like; taste spicy, sweet and pungent; oil 5–8%	Vermouths
Sweet clover	*Melilotus officinalis*	Aerial parts, essential oil	Long, stalked leaf divided into three narrow, ovate, toothed leaflets; small, yellow flowers; odor sweet, hay-like; taste aromatic	Liqueurs, vermouths
Sweet marjoram	*Majorana hortensis*	Aerial parts, essential oil	Branched; hairy leaves; small whitish or rose flowers; odor distinct, pleasant; taste aromatic, bitter; oil vermouths 0.5–1.3%	Flavored spirits, liqueurs,
Tarragon	*Artemisia dracunculus*	Leaves, macerate, percolate	Shrubby; dark, smooth, shiny leaves; odor aromatic; taste sweet and bitter; oil 0.5–2.8%	Flavored spirits, liqueurs, medicinal wines
Valerian	*Valeriana officinalis*	Roots, essential oil	Dark brown root, 2–4 cm long, 1–2 cm wide; odor pungent; taste bitter; oil 0.5–2%	Flavored spirits, liqueurs, vermouths

Continued

Table 1 Continued

Common name	*Scientific name*	*Parts and essences used* [a]	*Description of the part used and the content of essential oil*	*Use in alcoholic drink*
Wild thyme	*Thymus serpyllum*	Aerial parts, flowers	Low, spreading bush; very small, short-stalked, lanceolate obovate oblong leaves; pink-purple flowers; odor distinct; taste aromatic; oil < 1%	Flavored spirits, liqueurs
Woodruff	*Asperula odorata*	Aerial parts, flowers	Low-carpeting plant; slightly shiny leaves; small, white, star-shaped; odor cumarin-like	Brandies, vermouths, vodka
Wormwood	*Artemisia absinthium*	Aerial parts, macerate, percolate	Shrubby; silver-gray leaves; pale greenish-yellow flowers; odor pungent; taste aromatic, extremely bitter; oil < 0.9%	Bitters, flavored spirits, liqueurs, vermouths
Yarrow	*Achillea millefolium*	Aerial parts, macerate, percolate	Pinnate, finely divided leaves; small, white, or pinkish flower heads; odor fragrant; taste pleasant, bitter; oil 0.2–1.3%	Brandies, flavored spirits, liqueurs, vermouths

[a]The essences given are reported in the literature to be used for the flavoring of alcoholic drinks.

porous bags in the spirit, or placed in trays above the free surface of the spirit; the latter two methods have the advantage that the exhausted herbs can be easily removed from the still. The third method is commonly used in production of liqueurs and flavored spirits.

In a few cases, herbs contain not only specific flavoring compounds, but also a significant content of fermentable sugars (juniper berries, about 15–30%; gentian roots, about 25% on a dry-weight basis) which can be converted by strains of yeast into ethanol in an alcoholic fermentation process. For example, juniper berries can be used for flavoring grain spirits (gin, geneva), fruit spirits (fruit brandies; klekovača, from Yugoslavia), but also in the production of special alcoholic drinks which involves the alcoholic fermentation process, for example, Wacholderlüter (the base for Steinhäger), borovička, or brinjevac. The special flavor of the alcoholic drinks obtained in this way is mainly attributable to the flavoring compounds of the herb. Commonly, the herb materials are first comminuted and mixed with water in a specified ratio; sometimes, mineral salts (ammonium salts, phosphates) are also added. Then a certain amount of bakery or brewer's yeast is added. The fermentation conditions are usually regulated; in the fermentation of juniper berries, the temperature is about 25–30 °C, and pH is about 4. Owing to the presence of the aromatic compounds from the essential oil, which act as yeast inhibitors, fermentation is slow and can last up to a couple of weeks. The fermented mash is then distilled to obtain the alcoholic drink. The essential oil also distils and separates as a lighter phase. In the case of juniper berries, it is also possible at first to recover the essential oil by hydrodistillation, then to remove the residual plant material by filtration and finally to use the filtrate as the fermentation medium. By this procedure, the duration of fermentation process is significantly reduced (to about 1 day) and the yield of ethanol is increased.

It is clear that not only the herb materials but also every stage in herb processing and drink preparation affects, at least to some extent, the overall flavor of alcoholic drinks. A delicate balance between what is added and what is lost, transformed, or formed during herb processing and drink preparation is established.

Harvesting and Handling of Herbs

Harvesting

In order to retain their specific valuable properties, herbs have to be harvested at the correct time, which is characteristic for every herb. Table 2 lists the best moments for harvesting some commercially important herbs. The method of harvesting depends on the type and useful part of herb, and can be very primitive, usually by hand, or carried out using special machines, such as combine harvesters. Aerial parts are cut at ground level or somewhat above, and leaves are picked; they are often chopped before drying. Flower heads are picked out whole. Seed plants, after cutting, are left for a couple of hours or days in the field to dry or ripen, and the seeds are then shaken out or threshed for further drying, winnowing, or sifting. Roots are pulled out or turned out with a plough, freed of earth, tops, and rootlets, washed thoroughly and cut into smaller pieces for further drying.

Table 2 Harvesting and handling of herbs used in the alcoholic drink industry

Herb material	*Harvesting time*	*Drying*	*Packaging and storage*
Angelica root	End of autumn	In dark, well-aired places, or in dryers (35–45 °C)	Bags or bales; on shelves in dry, well-aired places
Anise seed	August and September when just beginning to turn brown	In shady or dark, well-aired places with frequent turning	Paper bags; in dry places
Arnica root	In autumn, after the leaves have died down	In shady or dark, well-aired places	Bags or bales; on shelves in dry, well-aired places
Balm leaves	Before flowering, without stems	Immediately after harvesting. In shady, well-aired places, or in dryers (40 °C)	Immediately after drying; hermetic packing; in dark, cool places
Basil aerial parts	During flowering in the early summer	In shady, well-aired places, or in dryers (35–45 °C)	Paper bags; in dry, air-conditioned places
Blessed thistle aerial parts and sowers	At the beginning of flowering in mid-August	Immediately after harvesting. In shady, well-aired places, or in dryers (40 °C)	Paper bags; in dry, air-conditioned places
Burdock root	In autumn, before becoming stiff	In dryers (35–45 °C)	Jute or paper bags; only stored for a few months
Calamus root	At the end of summer or during the autumn	In well-aired places, or in dryers (40 °C)	Bags; in dark, air-conditioned places
Caraway seeds	July to August	In the field until dry	Bags; in air-conditioned places
Cardamom seeds	The fruits are picked up whilst growing	In the sun	Bags; in air-conditioned places
Carnation flowers	May to August	In shady, well-aired places	Bags; in air-conditioned places
Celery seeds	August, in the early morning	Several hours in the field, or in dryers (35–40 °C)	Bags; in dry places
Clammy sage flowers and leaves	About 10 days after flowering	In shady, well-aired places; not recommended before distillation	Paper bags; in dry places
Common marigold flowers	During flowering, June to September	In the shade, or in dryers (40–50 °C) with good air circulation	Cartons; in dry places
Coriander seeds	August; after cutting, the herbs are left to complete ripening	In the shade. In dryers (35–45 °C) with good circulation	Jute bags; in dry, dark places
Dandelion flowers and roots	Immediately after flowering	In shady, well-aired places; often used fresh	Paper bags; when used fresh, the whole herb is transported
Dandelion roots	In autumn	In the open for several days, then in dryers (40–50 °C)	Paper bags
Dill seeds	When the fruits turn from green to brown	In the shade, or in dryers (35–45 °C) with good air circulation	In dry, dark places
Elecampane root	In autumn	In well-aired places, or in dryers (40 °C) with good air circulation	Bags or bales; in dry, well-aired places
European centaury aerial parts	During flowering (July–August)	In trays or other well-aired places, or in dryers (40–50 °C)	Paper or jute bags; in dry, well-aired places
Everlasting flowers	During flowering (June–September)	Immediately after harvest in well-aired places; in dryers (40 °C)	Jute or paper bags; on shelves in dry, cool and well-aired places
Fennel seeds	In summer	First in the shade, then in dryers (40 °C)	Well-closed cartons or paper bags
Fenugreek seeds	End of summer	In well-aired places	Well-packed; in dry places
Galangal root	In early spring	In the sun	Paper bags; in dark, cool places
Garden thyme leaves	During flowering (May–August)	In well-aired places, or in dryers (35 °C)	Paper bags; in dry places
Gentian root	In autumn	In the sun, or in dryers (80 °C)	Jute or paper bags; in dry, cool places
Germander aerial parts	April–July	In shady, well-aired places, or in dryers (35–45 °C)	Jute or paper bags; in dry, well-aired and cool places
German chamomile flowers	During flowering (April–June)	In shady, well-aired places, or in dryers (40 °C)	Cartons, metal boxes or waterproof paper bags; separated from other herbs; in dry, air-conditioned places
Ginger root	About 10 months after planting	In the sun for a day	In dry, dark, and cool places
Hop seeds	In the summer (July–September), when matured	In the shade; in dryers (60 °C)	Textile bags; on shelves in dry, well-aired places

Continued

Table 2 Continued

Herb material	*Harvesting time*	*Drying*	*Packaging and storage*
Horehound aerial parts	During flowering in the summer (June–August)	In shady, well-aired places	Paper, well-closed bags
Juniper berries	August–October	In the sun with frequent turning; in dryers (maximum 40 °C)	Jute bags; in dry, well-aired places
Lavender flowers	During flowering (from mid-June to mid-July)	In dry, well-aired places, or in dryers (40–45 °C)	Bags or bales
Liquorice roots	In the third or fourth year (September–October)	Stacked, in the field; in dryers (50 °C) after thorough cleaning	Stacks covered with hay; the clean roots are packed in bags
Lovage aerial parts and flowers	During flowering	In shady, well-aired places	Bags or bales
Lungwort aerial parts	During flowering (March–April)	In shady, well-aired places	Paper bags
Marjoram aerial parts	During flowering (July to mid-August)	Immediately after harvest; in well-aired places; in dryers (40 °C)	Bags or bales
Motherwort aerial parts	During flowering (June–August)	In well-aired places; in dryers (50–60 °C)	Bales; on shelves in dark, cool and well-aired places
Orris root	Second or third year, in July or August	In the sun for 5 or 6 days; complete drying in well-aired places	Textile bags
Pansy aerial parts and flowers	During flowering (in summer)	In shady, well-aired places	Paper bags or bales
Peppermint leaves	During flowering (in summer)	Left in the field to fade and lose 50% of water (may be used to distil); further drying in shady, well-aired places	Cartons or paper bags
Roman chamomile aerial parts	During flowering (April–June)	In shady, well-aired places; in dryers (40 °C)	Cartons, metal boxes or waterproof paper bags; separated from other herbs; in dry, air-conditioned places
Roman wormwood aerial parts	During flowering (July–September)	In shady, well-aired places; in dryers (40–45 °C)	Bags or bales; in dry, well-aired places
Rosemary leaves	May–October	In shady, well-aired places; in dryers (35–45 °C)	Jute or paper bags; in dry, air-conditioned places
Saffron flowers	During flowering	In shady, well-aired places	Paper bags
Savory aerial parts	During flowering (April–July)	In shady, well-aired places; in dryers (35 °C)	Paper bags
Star anis seeds	When fully ripened	In shady or dark, well-aired places with frequent turning	Paper bags; in dry places
Sweet clover aerial parts	During flowering (June–July)	In shady, well-aired places; in dryers (40 °C)	Paper bags
Sweet marjoram aerial parts	Immediately before flowering (June–September)	In shady, well-aired places; in dryers (40 °C)	Bags
Tarragon leaves	July–September	In well-aired places; must be green	Jute bales; in dry, dark places
Valerian roots	In early spring or in autumn	In well-aired places; in dryers (35–40 °C)	Jute bales or bags; in dry, well-aired places
Wild thyme aerial parts and flowers	During flowering (May–July)	In shady, well-aired places; in dryers (35–40 °C)	Bags or bales; on shelves in dry, well-aired places
Woodruff aerial parts and flowers	During flowering (May–August)	In the shade, with frequent turning	Well-closed containers
Wormwood aerial parts	During flowering (July–August)	In shady, well-aired places; in dryers (40–45 °C)	Bags or bales; in dry, well-aired places
Yarrow aerial parts and flowers	During flowering (June–August)	In dry, shady, well-aired places; in dryers (40–45 °C)	Cartons; in dry, cool, well-aired places

Drying

Fresh herbs are perishable to varying degrees and can be stored for only a limited period of time without spoiling. Spoilage and deterioration in quality of herbs are recognized by changes in color, flavor, and overall appearance. With proper handling procedures, spoilage can be prevented, and shelf-life can be significantly extended.

Drying is a simple and most common method of preserving herbs; dried herbs can be used all year round. Herbs or their parts should be dried immediately after harvesting to decrease the risk of spoilage and the loss of volatile oils. Dried herbs are also relatively immune to microbial spoilage.

Drying a herb or its parts involves removing an amount of water from the herb material by evaporation in the open or in a heated air flow. The principle of drying is the same for all herbs, although specific drying conditions may be required, as can be seen in Table 2.

The oldest method of drying herbs, usually used in the case of wild aromatic herbs, is drying in the open. The herb materials are spread thinly in suitable shady or dark and well-aired places (on the floor, on shelves, in the attic, or in trays). The herb materials should not be turned over too often, to avoid their disintegration. If the herbs are spread too thickly, the danger of burning increases. The duration and efficiency of drying depend on weather conditions. Drying can require a couple of weeks and, if the weather is rainy, the herbs can be spoiled by decay.

Dehydration in a heated air stream offers obvious advantages over drying in the open, notably regulated drying conditions, shorter duration of the process (several hours or even minutes), fewer flavor and color changes as well as a decreased danger of spoilage. When cultivated, herbs are normally dried in heated air driers, such as tray, screen-conveyor, or rotary dryers, to insure efficient drying of large quantities of the herb materials at economic capacity. The drying conditions are regulated by the air temperature, relative humidity, and flow rate. With warm air (40–80 °C), the duration of drying is usually a couple of hours. Drying in hot air (200 °C or even higher) requires a few minutes (2–5 min).

Dried herb materials can be ground, crumbled, rubbed, or treated by any other suitable method for further processing, packaging, or storing.

Packaging and Storing

There is always a risk of spoiling and the loss of volatile compounds during long-term storage, especially under unfavorable conditions. Generally, herbs, leaves, and flowers are sensitive, while roots and seeds endure prolonged storage quite well. Deterioration in quality is accelerated by moisture, light, oxygen, and elevated temperature. Dried herb materials are best kept in a dark, dry and well-aired room at a low temperature. In many cases, the dried herb materials can be stored satisfactorily in paper, linen, or cotton bags placed in well-closed containers or bins. Properly dried and packed herb materials (Table 2) can be kept at room temperature for months, or even a couple of years. Sometimes, storage appears to favor development of the aromatic principles, probably owing to the action of the specific enzymes.

Production of Essences

The processing of herb materials is usually carried out using traditional methods, notably hydrodistillation and solvent extraction, the main goal being to recover their essences: essential oils and extracts. These essences are very suitable for direct application in alcoholic drink preparation as they insure higher consistency in quality of the alcoholic drink. In addition, they are more easily stored, with less danger of spoiling during long-term storage, although their flavor properties seldom compare favorably with those of the herb materials from which they were recovered. In the case of essential oils, this distortion is caused by chemical interchanges resulting in artifacts, the evaporation of the low-boiling components, the dissolution of certain water-soluble components in the still liquors, and the degradation of nonvolatile cellular matter during hydrodistillation.

For complete isolation of valuable essences, the herb material must be properly prepared by comminution to some extent, which depends on the nature of the herb material. Flowers are usually distilled uncomminuted, leaves are ground, seeds must be thoroughly crushed, while roots and arid stalks should be cut into smaller pieces. For this purpose, standard equipment can be used, e.g., roller mills for crushing seeds, cutters for chopping roots, and similar comminuting machines. When comminuted, the herb materials should be immediately processed to hinder the loss of volatile flavorings.

The essential oils are mainly recovered from the herb materials by hydrodistillation. The herb material is put into a still and heated with direct steam (steam distillation); it can also be immersed in warm water (water distillation) or placed on the perforated plate above the bottom of the still partially filled with water (steam and water distillation), and then heated by any of the usual methods. Water distillation is well suited for finely powdered materials and flowers; water and steam distillation is favorable for herb and leaves; and steam distillation is most successful for seeds and roots. The resulting vapor mixture

passes through a condenser, and the condensate is collected in a Florentine flask where the essential oil is separated from the water; this water can be returned to the still or, in some cases, used as fragrant water (*aquae aromaticae*). The duration of a batch distillation depends on the procedure applied, the kind of herb materials, and the extent to which the herb material is comminuted. Steam distillation requires a shorter time for completion than steam and water or water distillation. The essential oils from aerial parts or leaves are usually recovered after 1.5–2 h, from seeds after 4–6 h, and from roots after 6–8 h. In general, the distillation of comminuted herb material requires half the time for complete recovery, as compared with uncomminuted material, and gives higher yields of oil. Different methods of distillation result in essential oils of different composition. The yield of essential oil depends upon several factors, including the age of the herb material, its geographical origin, weather conditions, and the method of distillation. On distillation at atmospheric pressure, thermolabile components can be decomposed and, in such cases, the application of reduced pressure is suggested; in some cases, distillation at an increased pressure is preferred. It is possible to apply solvent extraction first and then to submit the extract thus obtained to steam distillation at reduced pressure, as in the case of orris root.

Traditional extraction processes involve the use of high-proof alcohols, and these are widely employed for dried herbs, leaves, and roots. The extracts obtained have the 'true,' natural odor of the herbal materials. The methods commonly applied are maceration, percolation, and digestion. The extracts obtained are sometimes concentrated by distilling off the alcohol under vacuum. The products, so-called resinoids, are usually dark, viscous, almost solid, and can be used in the flavoring of alcoholic drinks.

Extraction with dense carbon diooxide (CO_2) has also been applied in the removal of essences from some herbs. Dense CO_2 has many advantages over conventional organic solvents. It is without color, odor, or taste, easily removable from the extract, chemically inert, and nonflammable. In the subcritical state, liquid CO_2 behaves as a selective, nonpolar solvent, which under normal operating conditions (0–10 °C, 50–80 bar) dissolves virtually all the useful flavor components. CO_2-extracted juniper berry oil has already been used in the flavoring of gin. The essential oil of ginger roots has also been obtained by liquid CO_2 extraction. Due to the higher pressure and temperature applied, supercritical CO_2 has a higher solvent power than the liquid CO_2. Extraction with supercritical CO_2 is carried out at relatively low temperature (usually 40–60 °C) and high pressure (200–500 bar). It is important that the dissolving power and selectivity can be controlled by selection of adequate pressure/temperature combination. Many aromatic herbs have been extracted with supercritical CO_2, such as coriander seeds, rosemary, sage, and peppermint leaves.

See also: **Beers**: History and Types; Microbreweries; **Brandy and Cognac**: Armagnac, Brandy, and Cognac and their Manufacture; **Gin**: The Product and its Manufacture; **Wines**: Types of Table Wine

Further Reading

Gildemeister E and Hoffmann F (1956–1961) *Die Atherischen Ole*, vols 1–6. Berlin: Akademie-Verlag.

Guenther E (1949–1952) *The Essential Oils*, vols 1–6. New York: van Nostrand Reinhold; reprinted (1972–1992) Carol Stream: Allured Publishing.

Loewoenfeld C and Back P (1979) *The Complete Book of Herbs and Spices*, 2nd edn. London: David and Charles.

Morton ID and Macleod AJ (eds) (1986) *Developments in Food Science 3B: Food Flavours*, part B. Amsterdam: Elsevier Science.

Piggott JR and Paterson A (eds) (1989) *Distilled Beverage Flavour: Recent Developments*. New York: VCH.

Rose AH (1977) *Economic Microbiology*, vol. 1. London: Academic Press.

Shreter AI (ed) (1985) *Pravila sbora i sushki lekarstvennyh rastenii (sbornik instruktsii)*. Moscow: Medicina.

Herring *See* **Fish**: Introduction; Catching and Handling; Fish as Food; Demersal Species of Temperate Climates; Pelagic Species of Temperate Climates; Tuna and Tuna-like Fish of Tropical Climates; Demersal Species of Tropical Climates; Pelagic Species of Tropical Climates; Important Elasmobranch Species; Processing; Miscellaneous Fish Products; Spoilage of Seafood; Dietary Importance of Fish and Shellfish; **Fish Farming**; **Fish Meal**

HISTAMINE

R K Bush, University of Wisconsin-Madison Medical School, Madison, WI, USA
S L Taylor, University of Nebraska, Lincoln, NE, USA

Background

Histamine or β-aminoethylimidazole is formed by enzymatic decarboxylation of histidine, an amino acid that occurs naturally in many tissues in the human body, in the tissue of other animals and plants, and in foods (Figure 1). Once formed, histamine may be metabolized (1) by methylation of the ring by histamine-*N*-methyltransferase to 1-methylhistamine or (2) by oxidative deamination by diaminoxidase (histaminase) to imidazoleacetic acid. Methylhistamine can be deaminated by monoamine oxidase to form methyl imidazoleacetic acid. Imidazoleacetic acid is conjugated with ribose to imidazoleacetic acid riboside. These latter two products are readily excreted in the urine. Many tissues have the capacity to metabolize histamine by one or both pathways.

Distribution

Histamine is present in many human tissues, including skin, intestinal mucosa, heart, lung, and nerve endings in the brain. The usual body storage sites for histamine include mast cells and basophils (Figure 2). Since most of the histamine is stored in the secretory granules of mast cells and basophils, only small amounts (0.2–0.4 mmol) are found in the circulation. A variety of stimuli, both immunological and non-immunological, may trigger the release of histamine from the mast cell or basophil. During tissue stresses of growth or repair, histidine decarboxylase activity is increased; histamine synthesis is also increased, which suggests that histamine may be important in healing processes. The newly formed histamine is rapidly metabolized rather than being stored. Histamine is metabolized by the pathways described above.

Pharmacological Effects

The biological effects of histamine are mediated through three types of receptors, H_1, H_2, and H_3 (Table 1). Histamine binds to these receptors, eliciting a variety of biological and physiological responses. These receptors have different signal transduction mechanisms. The H_1 receptors are coupled to phosphoinositide breakdown and to calcium metabolism, whereas H_2 receptors are linked to adenylate cyclase. Stimulation of the H_1 receptor causes

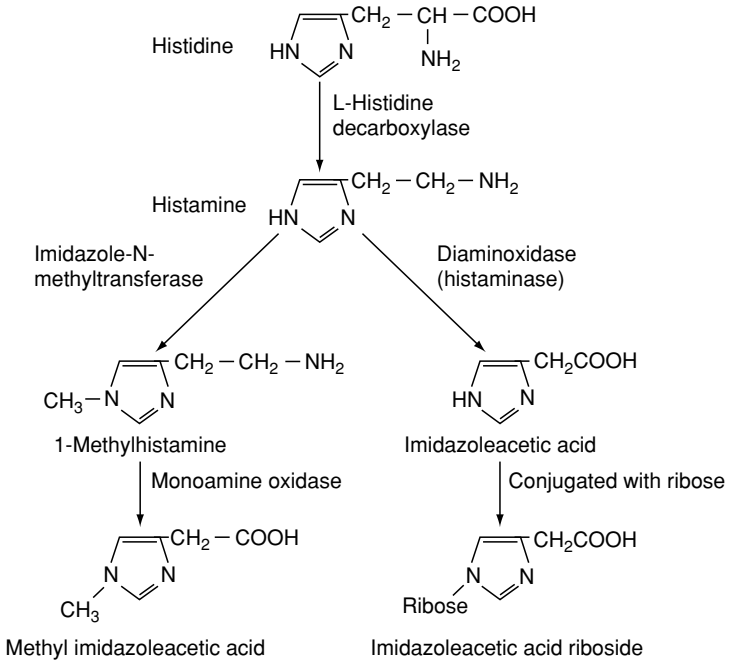

Figure 1 Pathway of synthesis and metabolism of histamine in human cells. The conversion of L-histidine to histamine is catalyzed by L-histidine decarboxylase, which is also found in certain bacteria. The degradation of histamine occurs via two routes; the blockade of one shunts the metabolism to the other. Adapted from Marquardt DT (1983) Histamine. *Clinical Reviews in Allergy* 1: 343–351, with permission.

Table 1 Characteristics of histamine receptor subtypes

	H_1 receptor	*H_2 receptor*	*H_3 receptor*
Effects of stimulation	Bronchial smooth muscle constriction Increased vascular permeability Gut smooth muscle constriction Vasodilation Pruritus Increased cyclic guanosine monophosphate levels	Increased gastric acid secretion Increased pepsin secretion Decreased IgE-mediated basophil histamine release Increased cyclic adenosine monophosphate levels Decreased lymphocyte-mediated cytotoxicity	Regulate synthesis and release of histamine from histaminergic nerves Inhibits endogenous norepinephrine release from sympathetic nerves
Stimulus–response coupling	Linked to phosphoinositide breakdown Increased calcium mobilization	Linked to adenylate cyclase	?
Antagonists	Chlorpheniramine, diphenhydramine, loratidine, cetirizine, fexofenadine	Cimetidine, ranitidine, famotidine, nizatadine	Thioperamide, clobenpropit

bronchial smooth muscle constriction in the airways of the lung (bronchospasm) and increased vascular permeability in the skin, which may be manifested as swelling or hives and produces itching by affecting nerve endings. Stimulation of H_2 receptors in the stomach results in increased acid secretion, which can produce indigestion or peptic ulcer disease. The H_3 receptor stimulus–response coupling is presently unknown. The H_3 receptors were initially described in presynaptic nerve endings and on nerve bodies of histaminergic neurons in the brain, where they function as autoreceptors to regulate the synthesis and release of histamine. H_3 receptors have also been located in the airways and gastrointestinal tract.

These histamine receptors and their responses can be antagonized by specific compounds. For the H_1 receptor, classical antihistamines such as diphenhydramine and chlorpheniramine act as competitive antagonists for the receptor. The H_2 receptor is antagonized by the compounds, cimetidine and ranitidine, whereas the H_3 receptor can be antagonized by thioperamide.

Role in Allergic Reactions

Allergic reactions are among the most commonly recognized effects of histamine. Specific immunoglobulin E (IgE) antibodies (the class associated with allergic reactions), which may be directed at various allergens, including inhalants such as pollens, molds and animal dander, foods, drugs, or insect venoms, are bound to the surface of mast cells and basophils (Figure 2). When the allergen cross-links two IgE molecules attached to the mast cell or basophil surface, histamine release occurs. Histamine then interacts with tissue histamine receptors (predominantly H_1), which can produce local reactions such as

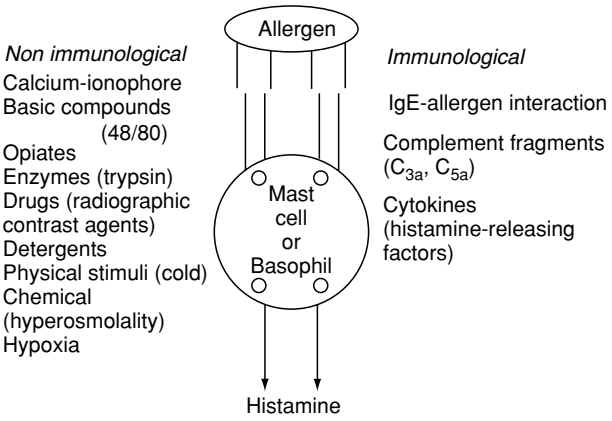

Figure 2 Human mast cell or basophil stimulated to release preformed histamine by a variety of immunological and nonimmunological mechanisms.

sneezing, nasal and ocular pruritus (itching), rhinorrhea (nasal discharge), and nasal congestion. In asthmatics, histamine release may cause cough, wheezing, and dyspnea (shortage of breath) owing to bronchial smooth muscle constriction. More generalized systemic symptoms can occur as a result of the ingestion of foods and drugs, insect stings, or parenteral administration of various medications. These reactions, whether immunological or nonimmunological in nature, result in a release of histamine leading to a prompt and sharp rise in plasma histamine levels. The severity of the reactions is proportional to the magnitude of this rise. Mild reactions are associated with low histamine levels, approximately 1 mmol. Elevation beyond this level can produce generalized reactions with urticaria (itchy skin rash), gastrointestinal disturbance, cardiac arrhythmias, and hypotension (low blood pressure). At levels in excess of

12 mmol, life-threatening responses such as severe hypotension, bronchoconstriction, and ventricular fibrillation can occur. (*See* **Food Intolerance**: Types; Food Allergies.)

Cardiovascular Effects

Histamine causes vasodilation and increased vascular permeability. This can result in flushing and decreased peripheral vascular resistance with subsequent lowering of the blood pressure. Extravasation of fluid results in angioneurotic edema (swelling of the skin on the eyes, lips and extremities). These effects are primarily mediated through H_1 receptor stimulation. Stimulation of the H_1 and H_2 cardiac receptors can lead to a variety of rhythm disturbances.

Bronchial Smooth Muscle Effects

Stimulation of the H_1 receptors in bronchial smooth muscles causes constriction, producing narrowing of the airways. This results in breathing difficulties and wheezing. Stimulation of H_2 receptors may lead to relaxation of airway smooth muscle, although this effect is minimal.

Effect on Gastric Secretion

Gastric mucosal mast cells release histamine, which causes an increase in gastric acid secretion by gastric parietal cells through H_2 receptors. Histamine interacts with acetylcholaine and gastrin to control acid output in the stomach. H_2 receptor antagonists (cimetidine, ranitidine) are widely used to treat peptic ulcer disease because of their ability to decrease acid secretion.

Central Nervous System Effects

Two major sources of histamine in the brain are mast cells and histaminergic neurons. Mast cells are located near blood vessels and may affect local blood flow. Histaminergic neurons (nerve cells whose principal neurotransmitter is histamine) are limited in number, but their axons branch into long ascending fiber networks that ultimately innervate much of the brain. Their role is not clear, but the central effects of histamine include antidiuresis (fluid retention) through the release of vasopressin, induction of adrenocorticotrophic-releasing hormone and prolactin secretion, increase in blood pressure and heart rate, temperature regulation, and control of arousal mechanisms. The sedative affects of the first generation antihistamines (e.g., chlorpheniramine, diphenhydramine), which block H_1 receptors functions are well known. In contrast, second-generation H_1 blockers (loratidine, fexofenadine, cetirizine) that do not readily cross the blood–brain are less likely to induce sedation.

Effect on the Immune System

Histamine may have a more general effect on the immune system than its role in allergic reactions. Stimulation of H_2 receptors suppresses lymphocyte proliferation, T-cell-mediated cytotoxicity, lymphokine production, natural killer cell cytotoxicity, and antibody production by B-lymphocytes. Histamine in certain dose ranges is chemotactic for eosinophils, which are the principal effector cells in killing of parasites such as schistosomes. Parasites that invade body tissues generate profound IgE antibody responses. Interaction of parasite antigens with specific IgE antibodies on the surface of mast cells causes histamine release and the subsequent attraction of eosinophils to the area. Eosinophils can degranulate, releasing their granular proteins, which are potent toxins to parasitic organisms. (*See* **Immunology of Food.**)

Role of Cellular Proliferation

Another physiological role for histamine appears to be in wound healing, since proliferating tissues synthesize histamine at a higher rate than normal tissue. The contribution to wound healing is still not completely understood, however.

Histamine Poisoning

Histamine is present in measurable quantities in certain foods, including cheese, meat, wine, yeast products, and some vegetables such as spinach and tomatoes (Table 2). The ingestion of large quantities of histamine can result in clinical symptoms indistinguishable from those of an allergic reaction. These include flushing, throbbing or severe headaches, palpitations of the heart, abdominal cramps, and diarrhea. Dizziness, burning of the mouth or throat, nausea, hives, and angioneurotic edema can occur. Cough, wheezing, and severe respiratory distress may appear in patients who have an underlying allergic background, particularly in those with asthma. Symptoms generally begin 30–60 min after ingestion of foods containing high concentrations of histamine. (*See* **Migraine and Diet.**)

Measurement of Histamine in Food

A number of analytical methods for measuring the histamine content of foods have been developed including fluorometric methods, thin-layer chromatography, oxygen-sensor-based methods, and biologic methods. In addition, a number of commercially available enzyme-linked immunoassay test kits are marketed in the United States and Europe.

Table 2 Foods with high histamine content

Food	*Histamine content ($\mu g\ g^{-1}$)*
Fish	
Tuna, mackerel, bonito, mahi-mahi, bluefish	ND–9000
Sardine, herring, anchovy	ND–2000
Cheeses	
Swiss	ND–2500
Roquerfort and blue	Trace–4090
Cheddar	ND–1080
Gouda	Trace–200
Mozzarella	ND–50
Wines	
American red	0.2–15.5
American white	0.2–11.4
European red	ND–30
European white	ND–20
Dry sausages	Trace–650
Sauerkraut	< 5.0–200
Soya sauce	ND–2740

ND, not detectable.
Modified from Stratton JH, Hutkins RW and Taylor SL (1991) Biogenic amines in cheese and other fermented foods: a review. *Journal of Food Protection* 54: 460–470, with permission, and Diel E, Bayas N, Stibbe A *et al*. (1997) Histamine containing food: Establishment of a German Food Intolerance Databank (NFID). *Inflamm Res* 46(1): S87–S88.

Scombroid Fish Poisoning

Scombroid fish poisoning results from eating spoiled fish of the Scombridae and Scomberessocidae families, and produces headache, itching, diarrhea, flushing, and palpatations immediately afer ingestion. These fishes include tuna, mackerel, skipjack, and bonito. However, nonscombroid fishes, such as mahi-mahi, blue fish, amberjack, marlin, herrings, sardines, and anchovies, have also been implicated in scombroid-toxin-like illness. Scombroid fish poisoning is the most common cause of ichthyotoxicosis worldwide. In addition to ingestion of fish, histamine poisoning can occur occasionally through ingestion of cheese such as Swiss, Cheddar and Gouda. (*See* **Fish**: Spoilage of Seafood.)

There has been some debate as to whether histamine is the sole cause of scombroid poisoning, but histamine is the most likely causative agent of scombroid fish poisoning. As clinical proof of the role of histamine, a definitive study involved three patients who were exposed to marlin, and later developed the classical symptoms of scombroid fish poisoning. Urinary excretion of histamine and its metabolite, *N*-methylhistamine, was measured. Urine samples collected 1–4 h after ingestion of fish show that the histamine and the *N*-methylhistamine levels were respectively nine to 20 times and 15–20 times more than the normal mean, a likely indication of exposure to high histamine levels. During the subsequent 24 h, the levels fell to four to 15 times and four to 11 times the normal values, respectively, and had returned to normal by 14 days. There was no evidence of mast cell activation as measured by the principal urinary metabolite of prostaglandin D_2. In addition, ingestion of a high concentration of histamine has been demonstrated to produce symptoms of scombroid poisoning. These symptoms are generally self-limiting, lasting a few hours, and can be quickly treated with systemic antihistamine administration. Elevated histamine levels in fish tissue have been associated with poor refrigeration, which results in bacterial spoilage and histamine formation. Bacteria possessing histidine decarboxylase can contaminate the gastrointestinal tract, skin, and gills of fish. In the case of cheese, the entry point to contamination by the histamine-producing bacteria is raw milk. Refrigeration of raw milk and fish, good hygiene during food preparation, and careful control of fermentations will therefore avoid the problem. Cooking, however, will not destroy histamine.

See also: **Fish**: Spoilage of Seafood; **Food Intolerance**: Types; Food Allergies; **Immunology of Food**; **Migraine and Diet**

Further Reading

Chrush C, Sharma S, Unruh H *et al*. (1999) Histamine H_3 receptor blockade improves cardiac function in canine anaphylaxis. *American Journal of Respiratory and Critical Care Medicine* 160: 1142–1149.

Douglas WW (1985) Histamine and *S*-hydroxytryptamine (serotonin) and their antagonists. In: Goodman LS and Gillman AG (eds) *The Pharmacological Basis of Therapeutics*, 7th edn, pp. 605–615. New York: Macmillan.

Helrich K (ed.) (1990) Histamine in seafood: fluorometric method. In: *Official Methods of Analysis of the Association of Analytical Chemists*, 15th edn. Arlington, VA: Association of Official Analytical Chemists, pp. 876–877.

Lehane L and Olley J (2000) Histamine fish poisoning revisited. *International Journal of Food Microbiology*. 58: 1–37.

Marquardt D (1983) Histamine. *Clinical Reviews in Allergy* 1: 343–351.

Marrow JD, Marolies GR, Rowland J and Roberts LJ II (1991) Evidence that histamine is the cause of scombroid-fish poisoning. *New England Journal of Medicine* 324: 716–720.

Pierce FL (1991) Biological effects of histamine: an overview. *Agents and Actions* 33: 4–7.

Price RJ (1999) Compendium of fish and fishery products: processes, hazards and controls. University of California–Davis. National Seafood HACCP Alliance for Training and Education. Available at http://www-seafood.ucdavis.edu

Rafferty P and Holgate ST (1989) Histamine and its antagonists in asthma. *Journal of Allergy and Clinical Immunology* 84: 144–151.

Simon FER (1996) *Histamine and H_1-receptor Antagonists in Allergic Disease*. New York: Marcel Dekker.

HIV DISEASE AND NUTRITION

C Baldwin, Chelsea and Westminster Hospital, London, UK

Introduction

The acquired immunodeficiency syndrome (AIDS) is a disorder characterized by progressive and profound immunodeficiency that renders the infected individual highly susceptible to life-threatening opportunistic infections and tumors. It is caused by infection with the human immunodeficiency virus (HIV). The first cases of AIDS were reported in June 1981. At the end of December 1999 the World Health Organization reported that there were 47 million cases of HIV infection worldwide, of which more than 40 000 were in the UK. AIDS is the now the fourth leading cause of mortality.

In the early days of HIV infection survival was limited by the occurrence of lethal opportunistic infections such as *Pneumocystis carinii* pneumonia and *Mycobacterium avium* complex and the lack of antiretroviral therapies. As understanding of the natural history of the disease improved and effective prophylaxis against the common opportunistic infections became available, the associated morbidity and mortality improved. Initial treatment approaches were limited to the use of nucleoside analog monotherapy, such as with zidovudine (AZT), which provided only transient clinical and immunological benefits. Over 1995 and 1996, availability of a new diagnostic test, viral load measurement, and new drug classes such as protease inhibitors and nonnucleoside reverse transcriptase inhibitors heralded a new approach to HIV management. The era of highly active antiretroviral therapy (HAART), in which suppression of plasma viral replication to below viral load assay limits (currently < 50 HIV RNA copies ml^{-1}) is achieved in a substantial proportion of individuals, has changed the clinical management, prognosis, and survival of people with HIV. The immunologic benefits gained through virus control with 'cocktails' consisting of three or more drugs have led to dramatic reductions in the incidence of the major opportunistic infections and mortality. As the clinical management of HIV infection has altered, so has its nutritional management, although the indications for nutritional support remain less clearly defined. However, concern that many therapy combinations or individual components of therapy may lead to metabolic perturbations in individuals with HIV has led to a new era of nutritional evaluation and investigation in persons with HIV. This article reviews what is known about the nutritional status of people with HIV infection, the nutritional problems specific to this infection and its treatments, and the objectives of nutritional support.

Nutritional Status of People with HIV Infection

Before the introduction of HAART, profound weight loss and wasting were amongst the most common observations in people with HIV infection. In 1987 the Centers for Disease Control (CDC) included HIV wasting ($> 10\%$ loss of baseline body weight + pathogen-negative diarrhea or persistent night sweats), which could not be explained by the presence of concurrent infection, in the list of AIDS-defining illnesses. Two-thirds of men with AIDS were reported to weigh less than 90% of their usual body weight. Since 1995 effective combination therapy has been associated with significant weight gain in the majority of people. Indeed, some patients report gaining excessive amounts of weight on therapy. This may be due to a modest reduction in resting energy expenditure (REE) and improvement in gut function and nutrient absorption, and additionally, the resolution or prevention of minor infections which may contribute to bouts of weight loss. Weight loss and wasting are less common in people receiving HAART but still occur in some people who have an excellent antiviral/immunological response, in some cases as a result of side-effects of therapy. Additionally, not all patients respond or respond durably to HAART.

Weight loss in HIV infection is often multifactorial in its causation, even within individual patients. The management of weight loss is dependent on an understanding of the underlying pathophysiology.

Two major mechanisms dominate the pattern of tissue loss in HIV-related weight loss:

1. Protein-energy malnutrition (PEM) is a condition synonymous with starvation, resulting when the body's needs for protein, energy, or both cannot be met by diet. In HIV infection this could result from inadequate intake or nutrient malabsorption.
2. Cachexia is a clinical syndrome characterized by a mixture of metabolic abnormalities that lead to weight loss through accelerated wasting of host tissue mass, and failure of adequate nutrient intake, absorption, and utilization. This is

different from malnutrition in that there is a marked and rapid depletion of muscle rather than fat.

The major difference between PEM and cachexia is that the nutritional effects of starvation can be reversed by the provision of appropriate feeding whilst the nutritional changes associated with cachexia cannot.

Malnutrition is associated with important clinical consequences independent of the underlying disease. Chronic undernutrition leads to deficiencies in work capacity, immune function, wound healing, organ function, mental state, and growth in children. When secondary malnutrition occurs as a result of disease it is associated with additional problems: reduced quality of life, disease complications, poor clinical outcomes, and increased costs of management. In a study of men with AIDS wasting (mean weight loss 15%), the extent of muscle wasting was highly predictive of the strength and functional capacity of the leg muscles, demonstrating the important relationship between loss of lean body mass and functional capacity. Additionally, in people with HIV infection, weight loss has profound prognostic implications, predicting shortened survival, increased risk of hospitalization, and increased risk of being diagnosed with an opportunistic infection. The effects of weight loss are independent of immune function (as measured by CD4 lymphocyte counts). Recently published observations have quantified the impact of weight loss in HIV infection, demonstrating that a weight loss trend over a 4-month period of even as little as 0–3% was predictive of increased subsequent mortality. Weight loss of more than 5% was associated with a relative risk of 1.88 ($P < 0.001$) of developing an opportunistic infection in the follow-up period.

Characteristics of Malnutrition

Early studies of weight loss and body composition conducted in a group of men with AIDS-related weight loss, compared with an HIV-negative control group, demonstrated that weight loss in this group was associated with a profound depletion of body cell mass that was greater than that expected for the body weight reduction, whereas fat mass was largely preserved. These studies were performed before the first antiretroviral drugs became available and therefore represent untreated HIV infection.

A 1993 study examined patterns of weight change in 30 individuals with HIV over a period of 9–49 months. Two typical patterns of weight change were described – acute and chronic weight loss. Acute weight loss (Figure 1) was usually rapid, occurring at a rate of 4 kg per month or greater, and was associated with opportunistic infections, mainly nongastrointestinal. Recovery to usual or preillness weight was common. Chronic weight loss (Figure 2), in contrast, occurred at a slower rate of less than 4 kg per month, was progressive, and was associated mainly with gastrointestinal disease. These observations have a number of important implications for the nutritional management of people with HIV. First, they demonstrated that HIV infection was not associated with unremitting weight loss. Treatment of the underlying infection leads to the ability to regain weight, often back to preillness weight, and therefore HIV does not prevent an appropriate anabolic response. Second, weight loss may precede specific signs and symptoms of infection and, therefore, is of significant diagnostic value. This potential diagnostic role prompted the recommendation that weight

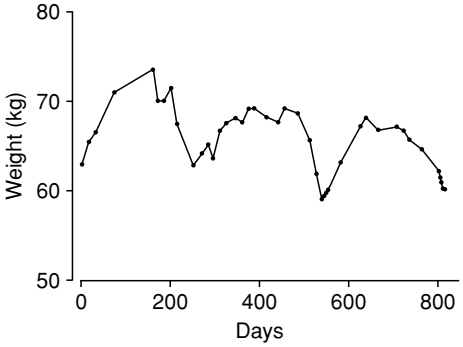

Figure 1 Example of acute weight loss episodes (time 0 = stage IV diagnosis). Reproduced with permission from; Macallan DC, Noble C, Baldwin C *et al*. (1993) Prospective analysis of patterns of weight change in stage IV human immunodeficiency virus infection. *American Journal of Clinical Nutrition* 58: 417–424.

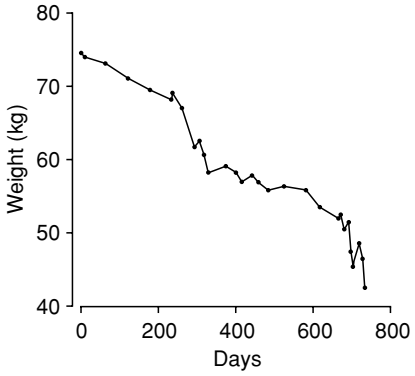

Figure 2 Example of chronic weight loss (time 0 = stage IV diagnosis). Reproduced with permission from; Macallan DC, Noble C, Baldwin C *et al*. (1993) Prospective analysis of patterns of weight change in stage IV human immunodeficiency virus infection. *American Journal of Clinical Nutrition* 58: 417–424.

change be recorded in all patients and be represented graphically in patient notes. Third, weight gain achieved following recovery from weight loss episodes usually occurs without the use of invasive nutritional support and, therefore, could be seen to represent the natural history of weight change in HIV infection.

Body Composition and Effective Antiviral Therapy

Advances in the management of HIV infection have had a significant impact on the incidence of weight loss and wasting but nutritional status has not returned to normal. In 1998 the first reports appeared of altered body composition in people with HIV infection. A syndrome of fat redistribution and associated metabolic abnormalities has been described which has become known as lipodystrophy. The syndrome(s) consists of accumulations of fat (lipohypertrophy) in the abdominal area (intraabdominal in men, breast enlargement in women), posterior neck ('buffalo hump'), and fat loss (lipoatrophy) in the face (particularly the temporalis, nasolabial, and buccal fat pad), arms and legs (leading to vein prominence). Metabolic changes, including hypercholesterolemia, hypertriglyceridemia, hyperglycemia, and insulin resistance, are also reported in association with the syndrome. There is so far no accepted definition for the syndrome of fat redistribution or any accepted set of standardized diagnostic criteria. However, the clinical presentation is characteristic and highly stigmatizing. Reports of the prevalence of the fat redistribution syndrome vary widely, depending in part on definition, from as low as 2% up to 84%. Initially the syndrome was thought to be associated solely with protease inhibitors, but more recently it has been reported in people receiving all classes of antiretroviral therapy. Several authors have proposed mechanisms to explain the development of the syndrome but the causes remain unknown. Identified risk factors include duration of antiretroviral therapy, older age, and white race. The use of protease inhibitors may accelerate time to syndrome onset.

An accurate assessment of body composition in people with fat redistribution is difficult. Bioelectric impedance analysis will not accurately measure body composition in this syndrome as it is unable to detect regional distribution or assess any changes that may occur. Dual X-ray absorptiometry (DEXA) scanning can determine regional fat compartments and may be useful in assessing people with fat redistribution, although the absence of data on normal body composition as well as regional body composition may limit its usefulness in diagnosing the syndrome. Computer-assisted tomography and magnetic resonance imaging can quantify body fat and therefore may be of most use in assessing body composition in this syndrome, although they are not always practical.

Deficiencies of Individual Nutrients

A number of studies have attempted to measure vitamin and mineral levels in HIV-positive subjects compared with a control group. Overall, results suggest that specific nutrient deficiencies are common in HIV infection and especially in the advanced stages of disease. The actual extent of overt and marginal deficiency and their significance is unclear. More information is needed on factors that might influence nutrient status measures, such as nutrient intake, drug–nutrient interactions, the influence of disease, and the presence or absence of infection; improvements in methods of nutrient assessment are also needed.

There is extensive literature on the relationship of nutrient deficiencies to immune function. Nutritional deficiencies can impair immunity and so influence a person's susceptibility to infections. It has been suggested that dietary manipulations might diminish or reverse the immune defects caused by HIV infection and enhance resistance to opportunistic infections. At present there is insufficient evidence to support the use of particular nutrient supplements to modulate immune function. The use of vitamin and mineral supplements containing 100% of the recommended daily allowance for all micronutrients has been proposed in the USA by the Task Force on Nutrition Support in AIDS to insure sufficiency. Although some authors have suggested supplementary levels of some vitamins and minerals in people with HIV in an attempt to achieve normal plasma levels, the benefits of this approach in the short and long term are unclear.

There is no evidence that megadoses of any vitamin or mineral will alter the course of disease or improve the nutritional status of people with HIV.

The Mechanisms of Malnutrition in HIV Infection

There are a number of factors that may lead to the development of malnutrition in HIV infection: altered food intake, malabsorption of nutrients, and changes in metabolism.

Reduced Food Intake

Food intake in an HIV-positive person may be affected by many physical and emotional factors. Anorexia is an accepted response to infection. Loss of appetite is a problem in people with HIV infection and may result from the associated opportunistic

infections, malignancies, gastrointestinal symptoms, side-effects of medication, emotional issues, or a combination of these.

Nutrient Malabsorption

Malabsorption is common in HIV infection and may result from intestinal pathogens and the side-effects of medication.

Metabolic Alterations

Hypermetabolism in association with febrile illness is a familiar concept. A number of investigators have measured REE in asymptomatic HIV infection and demonstrated an increase of around 10%. The presence of opportunistic infections is associated with greater increases in REE of up to 34%. There is still much to be learned about the mechanisms of metabolic regulation that accompany weight loss and wasting in HIV. It is not understood why an increase in metabolic rate is not accompanied by an increase in food intake. Cytokines such as tumor necrosis factor (TNF) and interferons are responsible for many aspects but there is now evidence that they are not alone responsible for the wasting process.

If a person is to be in energy balance, then energy intake must be equal to REE plus energy used in activity (total energy expenditure (TEE)). A study published in 1995 examined the relative contributions of changes in energy expenditure and intake to the pathogenesis of HIV-related weight loss and has provided important information on the underlying causes of weight loss in HIV infection. It examined energy balance in a group of 21 HIV-infected individuals. TEE, REE, activity levels, and food intake were measured during episodes of weight loss, weight gain, and during periods of weight stability. It confirmed that REE is raised in individuals with HIV infection and that during periods of opportunistic infection REE is elevated even further. During periods of weight loss TEE was not elevated but reduced to be almost equal to REE, accounted for by a reduction in activity levels, and was raised during periods of weight gain, therefore TEE could not account for the changes in energy balance necessary to cause weight changes. Measurement of energy intake confirmed that reduced energy intake is the prime determinant of weight change in HIV-associated weight loss. It has been suggested that it is possible to predict the composition of tissue loss during periods of starvation and that there is an inverse relationship between the amount of body fat and the ratio of fat-free mass (FFM) to total tissue lost during food restriction. In the study mentioned above, the amount of FFM lost during periods of weight loss was 58% of the total, which is similar to that predicted using Forbes' equation and is consistent with undernutrition as the underlying mechanism of weight loss in this group of people with HIV.

These findings suggest that an adequate supply of nutrients could prevent weight loss or promote weight gain in people with HIV-related weight loss but simple deficiency of nutrients is not the only possible cause of weight loss. Several other aspects of altered metabolism have been described in people with HIV infection, termed metabolic dysregulation, which may account for relatively greater losses of lean body mass relative to fat in some people (a predominantly cachectic mechanism). First, reduced *de novo* lipogenesis is a classic response to starvation in healthy subjects. In AIDS wasting, *de novo* lipogenesis is increased in the majority of subjects. Second, AIDS wasting is associated with hypertriglyceridemia and low serum cholesterol levels compared with HIV-negative controls, which may be indicative of futile cycling of free fatty acids and triacylglycerols. Futile cycling is thought to be mediated by TNF and is said to occur when free fatty acids are released from fat and returned to the liver, where they are reesterified in triacylglycerols and secreted as very-low-density lipoproteins (VLDL), to return to the fat cell for breakdown and restorage as triacylglycerols. Third, AIDS wasting is associated with serum testosterone levels in the lowest quartile of the normal range. Low testosterone concentrations are observed in about 50% of men with wasting and are usually associated with advanced disease, weight loss, and hypogonadism, suggesting that the low concentrations are stress- or starvation-induced.

The above aspects of metabolic dysregulation do not occur uniformly amongst people with HIV-related weight loss and considerable heterogeneity has been noted in this population. It has been suggested that people with AIDS wasting fall into different pathogenetic subsets and that it may be possible to categorize people into those who have lost weight as a result of 'starvation' and who may therefore respond to provision of nutrients and differentiate them from the subset of patients who are cachectic and may not be able to respond appropriately to nutrition.

Nutritional Problems for People with HIV

People who are HIV-positive are vulnerable to a wide range of pathogens. A person's ability to continue to eat normally may be altered by the disease itself or by the treatment (Table 1). The nutritional problems for people with HIV fall into three main categories:

1. Wasting and weight loss resulting from symptoms that have an impact on food intake, such as dysphagia, taste changes, pain, nausea and vomiting, and shortness of breath
2. Restrictions on food intake imposed by antiretroviral treatments
3. The side-effects of antiretroviral therapy, such as nausea, vomiting and diarrhea and dsylipidemias, impaired glucose tolerance, and body shape changes

Dietetic management can be effective in managing weight loss resulting from a range of symptoms. Drug regimens may demand one of the following adaptations:

1. Eating at the same time as taking the drug
2. Avoiding food for an hour or two after taking the drug
3. Eating or avoiding certain types of food in order to insure maximum absorption of medication

Dietary advice to accommodate antiretroviral therapy is highly individualized and tailored to specific circumstances and to specific combinations of drugs and therefore will not be covered in any detail here.

The metabolic and body shape changes associated with HIV represent a new challenge in the management of HIV infection. The aim of the management of fat redistribution is to replace lost fat and to reduce accumulated fat deposits. So far, management has tended to concentrate on changes to drug therapy which involve switching away from drugs implicated in the causation of the syndrome. A range of interventions have been tried and modest benefits have been reported, which may in part relate to the relaxation of dietary restrictions associated with some medications and to improved well-being. The use of recombinant human growth hormone has been shown to have some positive benefits, predominantly on reducing accumulations of abdominal and dorsocervical fat, in a few small studies. Diet and exercise remain unevaluated as possible management strategies for the fat redistribution syndrome.

The clinical significance of the metabolic complications associated with the fat redistribution syndrome are not known. A small number of reports

Table 1 Common manifestations of acquired immunodeficiency syndrome (AIDS) and potential nutritional problems

AIDS-related infections and cancers	*Potential nutritional problems*
Opportunistic infections	
Fungal	
Candida	
Oral	Sore mouth, altered taste perception, anorexia, reduced saliva production
Esophageal	Dysphagia
Cryptococcus	
Meningitis	Pyrexia, nausea, and vomiting
Protozoan	
Toxoplasmosis	Pyrexia, ethargy, confusion
Pneumocystis carinii pneumonia (PCP)	Pyrexia, dyspnea, anorexia and weight loss, tiredness and lethargy
Bacterial	
Mycobacterium avium intracellulare (MAI)	Pyrexia, anorexia and weight loss, diarrhea and malabsorption
Viral	
Cytomegalovirus (CMV)	Pyrexia, diarrhea, and malabsorption
Herpes simplex (oral)	Dysphagia
Human immunodeficiency virus (HIV)	Pyrexia, diarrhea, and weight loss
Parasitic	
Cryptosporidium	Diarrhea and malabsorption, anorexia, weight loss, nausea and vomiting, pyrexia
Microsporidia	
Cyclospora	
Isospora	
Cancers	
Kaposi's sarcoma (gastrointestinal)	Dysphagia, sore mouth, anorexia, abdominal discomfort and obstruction, diarrhea, and malabsorption
Non-Hodgkin's lymphoma	Anorexia, weight loss, dysphagia, diarrhea
Other	
AIDS enteropathy	Diarrhea and malabsorption, weight loss
AIDS encephalitis	Confusion, dementia, lethargy
Side-effects of HAART	Nausea, diarrhea, dyslipidemias
Coinfections	e.g. hepatitis B and C, alcoholic liver disease

HAART, highly active antiretroviral therapy.

have been published of myocardial infarction in people on HAART. Additionally, fatal pancreatitis with hypertriglyceridemia has been reported. The extent to which lipid disturbances and hyperglycemia are associated with an increased risk of vascular disease and increased survival is unknown and is the subject of a 15 000-patient global study. Results are not expected for another 2 years. Currently, management of isolated hyperlipidemia is through conventional dietary advice and lipid-lowering agents. As some statins (human menopausal gonadotropin coreductase inhibitors) may interact with protease inhibitors, advice should be sought before introducing these agents.

The effective treatment of the fat redistribution syndrome will be most successful only once the mechanisms are understood. However, approaches which in small, mostly cohort studies have yielded benefit in some patients include exercise, growth hormone, antiretroviral therapy modification, and metformin. No uniformly effective management has been established. The use of supplements such as acetyl-L-carnitine, coenzyme Q-10 (ubiquinone), and riboflavin to support mitochondrial function in adipocytes has been widely suggested but not evaluated.

Objectives of Nutritional Support

The objectives of nutritional support in people with HIV infection vary widely depending on the stage of the disease and the presence or absence of symptoms which have an impact on nutritional intake.

Asymptomatic HIV Infection

When a person is diagnosed with HIV infection but is symptom-free, he or she often considers dietary change within a framework of lifestyle changes to maintain health. Dietary information is available in a variety of media, from the advice of friends and family, to magazine articles, television programs, and the internet. The rationale for making dietary changes in people with asymptomatic infection is based on avoiding nutrient deficiencies, which may have an impact on immune function and maintaining an ideal body weight. Advice is aimed at promoting a regular balanced intake of a variety of foods. There is no evidence for increased or decreased requirements of any nutrient compared with the HIV-negative population.

Prior to the introduction of HAART, healthy-eating guidelines designed to promote a reduced intake of fat and sugar and an increase in fiber were considered inappropriate in people with HIV. Since the introduction of HAART and the associated survival advantages and potential cardiovascular side-effects, it may be pertinent to provide dietary advice in line with healthy-eating guidelines. There is no evidence base for such a recommendation, however.

People with HIV infection are at greater risk of contracting food- and waterborne infections since they are immunosuppressed and often suffer from coexisting disease or infections. It is generally recommended that people with HIV are given advice on food hygiene and water safety. Some local policies specify that this happens once a defined level of immunosuppression is reached. Cryptosporidiosis is a waterborne protozoa which can cause chronic diarrhea in immunosuppressed individuals. There have been reports of outbreaks of cryptosporidial infection in people with HIV infection where tap water has been implicated as the cause. The UK Department of Health recommends that people with HIV with a low CD4 lymphocyte count should boil all water (including still, bottled water) used for drinking, brushing teeth, and washing fruit and salad vegetables. This practice has not been adopted throughout the UK and will vary with local policies.

Unproven diet therapies There is currently no cure for HIV/AIDS. Faced with this, many people seek to take control of their own treatment and this often involves the use of diet therapies, which claim to prevent the progression of the disease, or to cure it. There can be tremendous psychological benefit from following a particular treatment, and people should not be discouraged from following a regime of their choice, but helped to evaluate and adjust their eating habits in order to meet nutritional goals. It is important that eating remains an enjoyable experience and that people do not subject their food choice to unnecessarily rigid rules that may detract from enjoyment.

The following considerations are helpful in evaluating dietary options:

- The diet should not contain substances in amounts that may be physically harmful.
- It should not completely replace health care that is generally accepted as effective.
- It should allow for an adequate energy and protein intake along with a variety of foods.
- It should not incur unnecessary expense or hardship.

Symptomatic HIV Infection

Many of the manifestations of HIV infection affect a person's ability to eat normally and maintain body weight, and therefore nutritional support is indicated.

The management of a person with weight loss should involve a number of steps: first, diagnosis and treatment of any underlying opportunistic infections, second, control of any symptoms that may have an impact on food intake; and third, dietary advice.

In the first instance dietary advice should be to maximize food intake by making alterations in the type and timing of meals, including well-chosen snacks between meals, and possibly modifying existing foods to increase their energy and protein content. This approach is a basic tool of dietetics that has received little research attention. A small uncontrolled study in 34 people with HIV demonstrated that dietary advice was associated with significant improvements in nutrient intake, weight, and anthropometric indices. A recent randomized controlled study of dietary counseling with or without oral nutritional supplements showed that nutritional counseling can achieve a substantial increase in nutrient intake in 50% of malnourished people with HIV which after 6 weeks was not associated with changes in weight and nutritional status.

A second line of dietary intervention is the use of specialized proprietary food supplements in combination with a balanced food intake. These usually take the form of prepacked drinks that are nutrient-dense and nutritionally complete. Since the appearance of these products 12–15 years ago, they have acquired the reputation of being able to promote weight gain in a person who is losing weight. A systematic review of energy-dense supplements in illness-related malnutrition has demonstrated significant benefits to weight and nutritional status but the effects on morbidity and mortality are less clear. A small number of studies have looked at the use of oral nutritional supplements in people with HIV and demonstrated short-term benefits to nutritional intake, weight, and nutritional status. More studies are needed to clarify the role of nutritional intervention in HIV-related weight loss and particularly the effects on morbidity, mortality, and clinical outcome.

If adequate nutrition cannot be achieved orally, then enteral or parenteral nutrition should be considered. The major indication for nutritional support in HIV is the failure to maintain body weight. The main goals of nutritional support are to improve survival morbidity by reducing the complication rate associated with secondary infections and to improve quality of life. There are a small number of studies that have looked at the efficacy of nutritional support in HIV. In general they have concentrated on the safety and on the ability to replenish lean body mass. Reports of rates of infection vary considerably with enteral and parenteral feeding but the general conclusions are that nutritional support can be provided with safety in people with HIV infection but should be undertaken with caution because of the serious risk of infection. A small number of studies have looked at the ability to replenish lean body mass using enteral and parenteral nutrition. The results suggest that nutritional support can effectively maintain body composition and reverse the depletion of fat and lean tissue in patients with weight loss in the absence of concurrent infection or when provided in conjunction with effective treatment of an infection. In the stressed catabolic patient with systemic infection, malnutrition is not restored towards normal with nutritional support and weight gains tend to be mainly fat and not lean body mass. It is important to note that in some studies the provision of artificial feeding led to prolonged survival with improved functional capacity and resumption of employment – benefits which should not be overlooked when considering the initiation of nutritional support.

The provision of nutritional support in the terminal stages of disease is a controversial issue. One study demonstrated benefits to quality of life associated with the provision of parenteral feeding in the terminal stages of illness. Recently, studies of the effects on survival of enteral and parenteral feeding have underscored the importance of starting feeding before it is too late.

The area currently receiving the most interest in relation to nutrition and HIV infection is the use of growth hormone and anabolic steroids for their ability to replenish lean body mass. Testosterone and testosterone analogs and recombinant human growth hormone have been demonstrated to replenish lean body mass effectively in people with 5–15% weight loss. Benefits have been demonstrated with and without supervised exercise programs. This is a relatively new area of nutrition and the potential role of anabolic therapies in conjunction with the provision of nutritional support raises many questions:

- Does increasing lean body mass change the prognosis?
- Is long-term anabolic steroid use safe?
- Which patients are likely to benefit – those with mild or advanced disease?
- Should steroids be used intermittently or continuously?
- Which anabolic drug is most effective?

Advances in the understanding of the pathophysiology of weight loss wasting have led to a greater understanding of the role of nutrition in HIV infection. The primary determinant of weight loss is frequently reduced food intake but this cannot alone explain all aspects of HIV-associated wasting and

the observed responses to nutritional support. Maintenance of an adequate food intake is an essential part of treatment but more information is needed on its combination with appropriate anabolic therapies to maintain normal body composition and functioning.

Improvements in the drug therapy of HIV management have had significant impacts on morbidity and mortality but the emergence of the fat redistribution syndrome as a side-effect of therapy represents a new challenge to the clinical and nutritional management of HIV infection.

See also: **Energy**: Energy Expenditure and Energy Balance; **Malabsorption Syndrome**; **Malnutrition**: The Problem of Malnutrition; **Viruses**

Further Reading

Bhasin S, Storer TW, Javanbakht M *et al*. (2000) Testosterone replacement and resistance exercise in HIV-infected men with weight loss and low testosterone levels. *Journal of the American Medical Association* 283: 763–770.

Corcoran C and Grinspoon S (1999) Treatments for wasting in patients with the acquired immunodeficiency syndrome. *New England Journal of Medicine* 340: 1740–1750.

Dobs AS (1999) Is there a role for androgenic anabolic steroids in medical practice? *Journal of the American Medical Association* 281: 1326–1327.

Dwyer JT, Bye RL, Holt PL *et al*. (1988) Unproven nutrition therapies for AIDS: what is the evidence? *Nutrition Today* Mar/Apr: 25–33.

Grunfeld C and Feingold KR (1993) Body weight as essential data in the management of patients with human immunodeficiency virus infection and the acquired immunodeficiency syndrome. *American Journal of Clinical Nutrition* 58: 317–318.

Macallan DC, Noble C, Baldwin C *et al*. (1993) Prospective analysis of patterns of weight change in stage IV human immunodeficiency virus infection. *American Journal of Clinical Nutrition* 58: 417–424.

Macallan DC, Noble C, Baldwin C *et al*. (1995) Energy expenditure and wasting in human immunodeficiency virus infection. *New England Journal of Medicine* 333: 83–88.

Miller TL and Gorbach SL (eds) (1999) *Nutritional Aspects of HIV Infection*. London: Arnold.

Mulligan K, Grunfeld C, Hellerstein MK *et al*. (1993) Anabolic effects of recombinant human growth hormone in patients with wasting associated with human immunodeficiency virus infection. *Journal of Clinical and Endocrinological Metabolism* 77: 956–962.

Safrin S and Grunfeld C (1999) Fat distribution and metabolic changes in patients with HIV infection. *AIDS* 13: 2493–2505.

Strawford A, Barbieri T, Van Loan M *et al*. (1999) Resistance exercise and supraphysiologic androgen therapy in eugonadal men with HIV-related weight loss. *Journal of the American Medical Association* 281: 1282–1290.

Task Force on Nutrition Support in AIDS (1989) Guidelines for nutrition support in AIDS. *Nutrition* 5: 39–46.

Wanke CA (1999) Epidemiological and clinical aspects of the metabolic complications of HIV infection. The fat redistribution syndrome. *AIDS* 13: 1287–1293.

Homocysteine *See* **Coronary Heart Disease**: Etiology and Risk Factor; Antioxidant Status; Intervention Studies; Prevention; **Vitamin B_6**: Properties and Determination; Physiology

HOMOGENIZATION

R A Wilbey, The University of Reading, Reading, UK

Introduction

An emulsion is a stable mixture of two normally immiscible phases as a continuous and a disperse phase. Particles of the disperse phase should be small, ideally 1 μm or less, when Brownian motion becomes a significant factor. The stability of an emulsion also increases with its viscosity. These conditions are the reverse of those needed for efficient separation, the movement of particles above 1 μm following Stokes' law. There are many examples of emulsions in foods, e.g., milk, cream, ice cream, butter, margarine, mayonnaise and salad dressings, sauces, cake batters, and some beverages. (*See* **Colloids and Emulsions**.)

The first treatment of milk to give a finer, more stable, emulsion that would remain homogeneous on standing has been attributed to Auguste Gaulin, whose first patent was granted in 1899. High-pressure homogenization is the most widely used method for stabilizing milk and milk products. It has also been widely used in other sectors of the food industry and in the preparation of nonfood emulsions of low to medium viscosity.

The term 'homogenization' has since been applied to similar emulsification processes, so that the terms 'emulsification' and 'homogenization' may often be treated as synonyms rather than homogenization being regarded as a specialized method of emulsification. A wide range of equipment has been developed for the homogenization of foodstuffs. The simplest of these methods is the high-shear mixer, which has been adapted for both batch and in-line processing. The colloid mill has been widely used for the homogenization of more viscous foodstuffs such as sauces and mayonnaises. Emulsification may also be the result of shear forces acting during other food-preparation processes, such as the cooling of margarine in scraped-surface heat exchangers, the working of butter and in the use of pin mills and disk mills.

The formation of a stable homogeneous foodstuff will depend on both the nature of the foodstuff and the process. There is often an optimum viscosity of the continuous phase for a given homogenizing technique, while lower viscosities in the disperse phase will aid homogenization. The presence of surface-active components may play a critical role in the formation and maintenance of an emulsion. The quantity of disperse phase will also affect the particle size distribution in the emulsion. This article will concentrate on the processes rather than the food components.

High-shear Mixers

Batch Mixers

High-shear mixers are commonly used for the batch mixing of food ingredients and the formation of emulsions. Most high-shear mixers rely on the development of a shear gradient between the rapidly rotating rotor and a static surface or stator. For low-viscosity liquids, the rotor is normally a simple turbine blade, and the stator is perforated. Figure 1 illustrates a mixer with a vertical-slotted disintegrating head that may also be used for breaking up fibrous tissue; a larger number of small square or round holes would normally be used in the stator screen for emulsification alone.

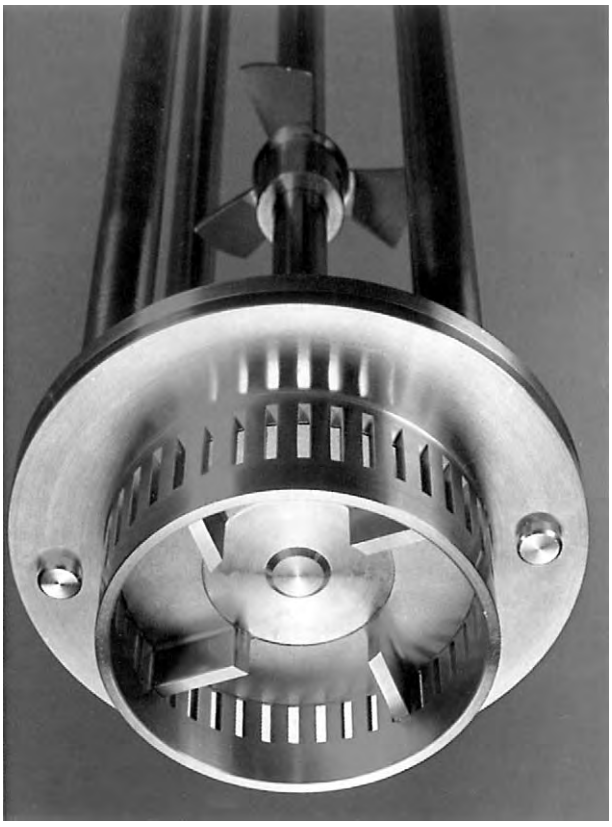

Figure 1 Batch high-shear mixer (courtesy of Silverson Machines Ltd).

The mixer is normally coupled directly to an electric motor, wound to run at 2850 r.p.m. for small mixers or at 1425 r.p.m. for large models, giving rotor tip speeds of approximately 100–150 m s^{-1} for pilot-scale applications. The flow of liquid from the turbine is principally radial, the flow being obstructed by the stator. Some liquid will pass between the rotor tip and the stator: the smaller the clearance, the greater the shear forces that will be developed. The bulk of the liquid, however, will be forced through the perforations in the rotor: the smaller the perforations, the greater the shear exerted on the mixture to be emulsified.

With batch mixing, the mixer may be left running until the desired homogenization/emulsification has been achieved. This method is limited by the viscosity of the foodstuff as adequate flow within the batch mixing vessel is essential, and more viscous foods and large batch sizes are better homogenized by other techniques.

In-line Mixers

In-line mixers for homogenization use the same basic principles as the batch mixers. In the simplest system,

the liquid passes from the turbine through the screen into an outer chamber and is discharged into the pipeline without any recirculation (Figure 2). The turbine has a weak pumping action. For some foods, there may not be sufficient shear in one pass through the homogenizing head so that either two or more mixers need to be placed in series, or the mix recirculated back to the feed tank until a sufficient degree of emulsification has been achieved.

Where recirculation needs to be carried out, this may be incorporated into the design of the mixer by allowing recirculation of part of the discharge from the shear head back to the inlet. The degree of recirculation depends upon the mixer design; the type shown in Figure 3 uses a high degree of recirculation to ensure a homogeneous discharge.

Colloid Mill

The colloid mill derives its emulsifying action from the high shear on the foodstuff as it flows between the rotor and stator, as illustrated in Figure 4. The rotor may be coupled directly to a motor or driven via a variable-speed gearbox. Typical rotor tip speeds are 20–50 ms^{-1}.

Normally, the gap between the rotor and stator are adjustable, typically between 100 and 200 μm, to enable a range of shear rates and throughputs to be achieved. The shear on the product will follow the general relationship:

$$\text{Shear stress } T = \mu \frac{v}{h}$$

where μ = viscosity, v = peripheral velocity, and h = clearance.

Greater disruption may be achieved by using toothed surfaces on the rotor and stator, e.g., for peanut butter, though where a grinding action is needed, a larger clearance may be used. The colloid mill may be used as an in-line emulsifier and is particularly effective with viscous foodstuffs, though the minimum particle size of the disperse phase (1–2 μm) will not reach the submicron sizes that may be achieved with the high-pressure homogenizer.

The High-pressure Homogenizer

The high-pressure homogenizer may be regarded as comprising two parts: a high-pressure pump and one or more homogenizing valves.

The High-pressure Pump

Pressure is used as the driving force for the homogenization process, the liquid mixture to be homogenized being raised to the homogenization pressure by a positive displacement pump. In all but the smallest pumps, there are at least three, sometimes five or

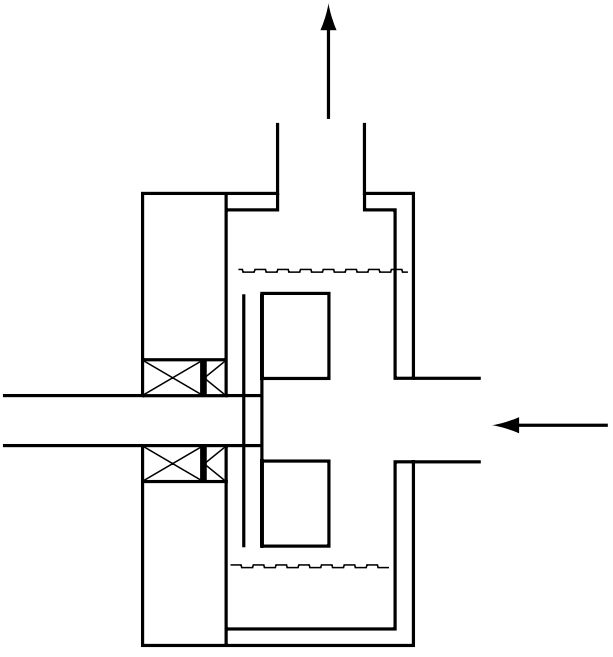

Figure 2 Schematic section through an in-line high-shear mixer.

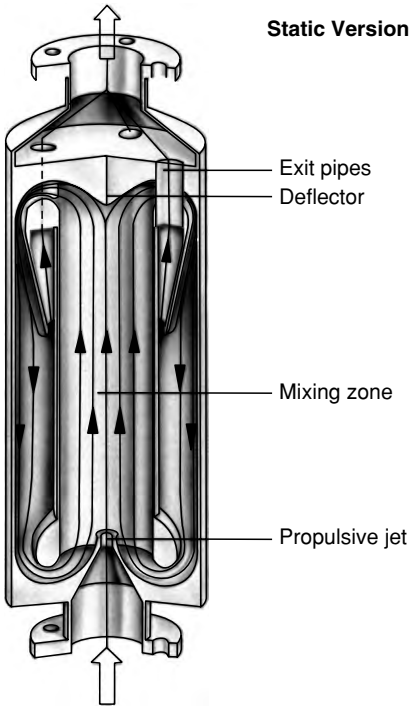

Figure 3 Section through the mixing chamber and shearing zone of a Burdosa Dynamic Loop Mixer (courtesy of dmt Gmbh, Fischbach, Germany).

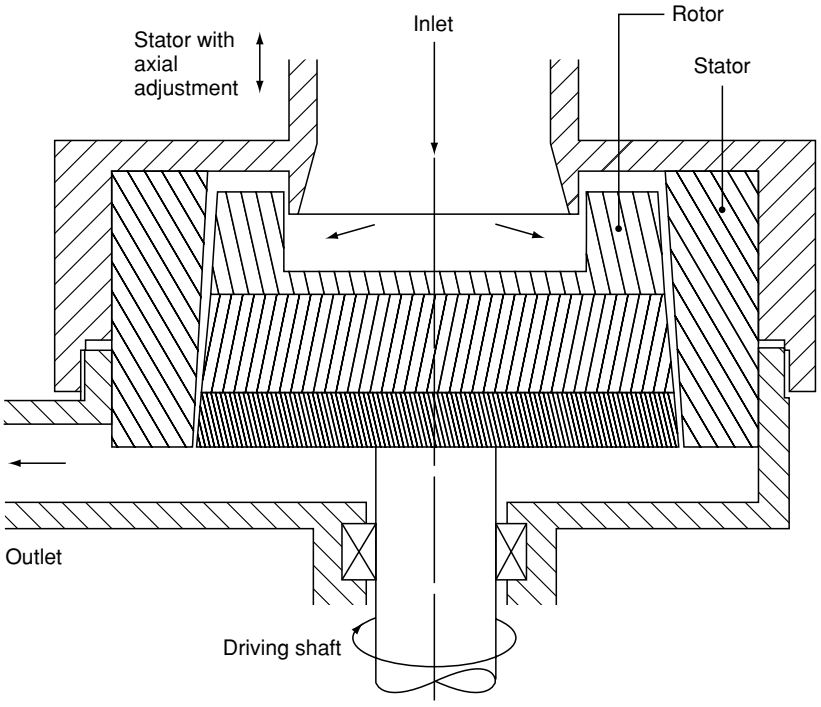

Figure 4 Schematic section through a colloid mill (courtesy of Fryma Ltd, Hemel Hempstead, UK).

seven, pistons. The pistons must be arranged so that they operate consecutively to maintain as even a feed pressure as possible. Both the pistons and the pump block are typically of stainless steel construction, but the piston seal rings are of a soft composite material and can be a source of cleaning problems. The layout of the high-pressure pump is illustrated in Figure 5.

The mixture to be homogenized is fed into the pump block under a slight positive pressure. A gravity feed is sufficient to ensure that the inlet valves are flooded, but continuous operation may be better served by feeding at a small positive pressure (e.g., up to 0.1 MPa) from a centrifugal pump via a preheater.

Two types of valve have been used in the pump block, either ball valves or poppet valves, as illustrated in Figure 6. The upper, discharge, valves are usually spring-loaded to ensure rapid closing.

Ball valves have a smaller contact area with the valve face, and they exert a greater pressure at the point of contact. This is an advantage for viscous liquids and when small particulates may be present in the feed stock, since valve leakage is the limiting factor in the high-pressure homogenization of liquids containing suspended particles. Poppet valves need to be fitted more carefully to the larger valve face, but, when properly maintained, they have a better performance with low-viscosity liquids such as milk.

A third, mushroom, design of valve has been introduced recently, designed as an optimum compromise between the ball and poppet designs. Both inlet and discharge valves are spring-loaded to ensure good seating.

The mixture is discharged from the pumping chambers of each piston into a high-pressure manifold. The pressure in this manifold is normally monitored by a pressure gauge, which must be of hygienic construction, with a diaphragm between the product and the sensor mechanism, plus a means of cleaning the space around the diaphragm. The only exit for product from this high-pressure manifold is through the homogenizing valve.

The Homogenizing Valve

Many designs of homogenizing valve have been invented, but for most valves, the same basic principles apply. The mixture, which must already be a homogeneous mix and itself a coarse emulsion, flows from the high-pressure manifold into the fixed valve seat. The internal diameter of the valve seat is smaller than the manifold, so the velocity of the liquid is increased. To escape from the valve seat, the liquid must flow radially across the face of the valve. The gap between the valve face and the seat is controlled so that a constant pressure drop across the valve is maintained. The opposing force (operating pressure times the valve area) may be provided by a heavy duty spring, a torsion bar, hydraulic pressure or by direct contact with a screw threaded lever. The basic design

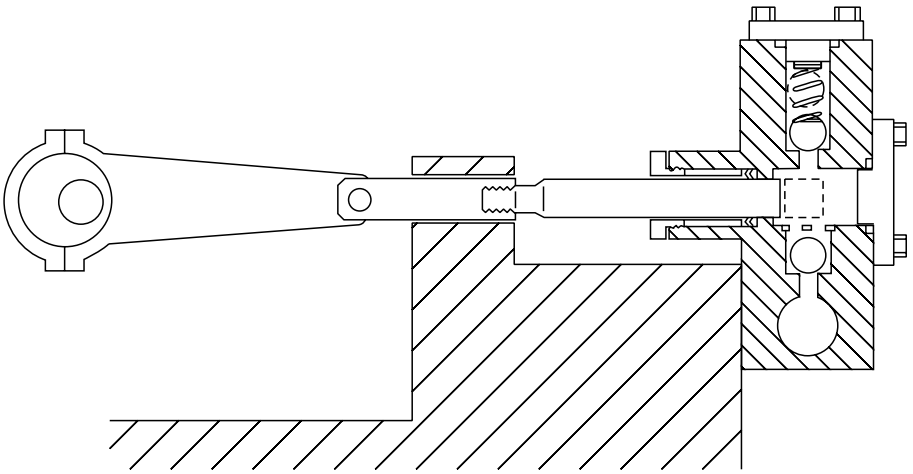

Figure 5 Schematic section of a high-pressure pump.

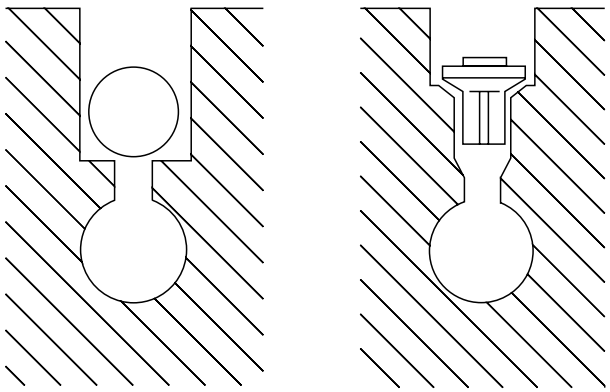

Figure 6 Types of valve: (a) ball; (b) poppet.

is illustrated in Figure 7. Pneumatic actuators have been used in some machines with capacities of less than $10\,000\,l\,h^{-1}$, thus reducing the variance in homogenization pressure. For most dairy products, the homogenization pressure is in the range of 2.5–30 MPa with throughputs of up to $\approx 50\,000\,l\,h^{-1}$ for milk. Machines are available that will produce a feed pressure of more than 100 MPa, though these are restricted to lower throughputs by the power required to drive them.

The gap between the valve and its seat is many times greater than the diameter of the droplets created in the emulsion, e.g., in a normal large-scale dairy application, the gap would be greater than 100 μm, the size of the gap tending to increase with the throughput of the machine. Even so, the cross-sectional area of the flow path within the valve is much smaller than that in the high-pressure manifold, so the velocity of the liquid will rise as it enters the valve. This acceleration can result in distortion of the particles of the disperse phase (often called the wire-drawing effect), which will reduce their stability. Within the valve, there will be a number of effects on the liquid that may contribute to the disruption of the fat globules:

1. Shear effects will result from the high velocity gradients between the liquid and the surfaces of the homogenizing valve. This effect, in combination with the wire-drawing effect, will lead to disruption of globules.
2. Turbulence also results from the high velocity of the liquid in the valve, causing eddy currents within the flow. High-velocity impacts between particles can lead to disruption, and so increases in velocity are associated with decreased fat globule size in the case of oil-in-water emulsions.
3. High-frequency vibrations (> 10 kHz) have been observed in some valves that may be of mechanical origin or a result of turbulence or cavitation. When homogenizing at high temperature, the drop in pressure as the liquid accelerates within the valve may result in the pressure momentarily falling below the saturation vapor pressure of an aqueous continuous phase, leading to the formation of steam bubbles. Collapse of these bubbles as the pressure increases again sets up high-pressure shock waves. Opinions vary about the contribution of cavitation and other sources of vibration to the overall disruptive effect during homogenization.
4. Impact can contribute to the homogenizing effect in some designs of homogenizing valve, where an impact ring is placed perpendicular to the flow of liquid from the valve. With valve exit velocities of 200–300 m s^{-1}, there can be a significant disruptive effect as the jet of emulsion impinges on the ring.

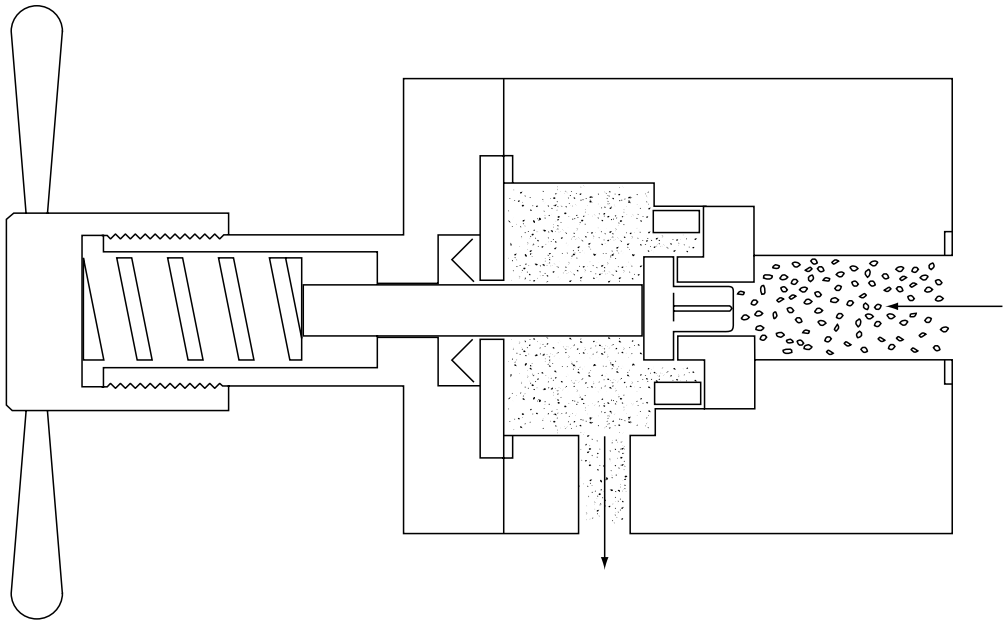

Figure 7 Schematic section through a homogenizing valve assembly.

Homogenizing valves have been produced from very tough corrosion-resisting alloys such as stellite. A better resistance to erosion may now be achieved using tungsten carbide and ceramic valves, though these are more difficult to produce and are more fragile. A very wide range of homogenizing valve designs using both single and multiple homogenizing surfaces have been produced, with the aim of increasing homogenization efficiency and thus reducing the energy costs. Various claims for up to 30% energy savings or a 40% reduction in homogenization pressure for an equivalent size reduction with these new designs have been made, in comparison with the traditional Gaulin valve design.

Practical Considerations in the Use of High-pressure Homogenizers

Since homogenizers operate at high pressures, problems can be encountered with the seals. The pistons are normally lubricated with potable water to minimize wear of the rear seals, though in aseptic machines, the pistons are also surrounded by a steam seal to avoid product contamination. With nonaseptic machines, the homogenization step is normally carried out prior to the heat treatment to minimize hygiene risks. Icecream mix may be homogenized at the heat treatment temperature (66–80 °C). In general, the efficiency of a high-pressure homogenizer increases with temperature, but temperature is normally limited by the product.

Constant homogenizing pressure is necessary for efficient reproducible emulsification. Rapidly fluctuating pressure may be due to a faulty gauge, but the most common cause is poor seating of the valves. Seating problems can result from wear, from incorrect placement of valves in the seatings, or from dirt or foreign bodies on the seats. The presence of air in the feed can also cause problems, and deaeration may be needed in some processes prior to homogenization.

With many dairy emulsions, the fine particles produced by homogenization tend to flocculate. These flocculated globules behave as large particles and separate from the emulsion. The introduction of a second homogenization stage or a multistage valve can overcome this problem. The pressure drop across the second stage valve should be relatively small, typically 3–5 MPa, to break up the aggregates. With two-stage homogenization, it is common practice to set the second-stage pressure first then to set up the first stage, remembering that the gauge will indicate the total pressure. The introduction of a second pressure gauge between the first- and second-stage valves aids control of the process. Where very find dispersions are desired, multiple high-pressure homogenization has been used.

Other Emulsifying Systems

Ultrasonic homogenization uses a transducer or a mechanical wedge to generate high-frequency vibrations at over 10 kHz, typically 18–50 kHz, so that the output is barely audible. These give rise to shock waves plus possibly cavitation in the mixture.

Efficiency declines with increasing frequency, the upper limit being about 50 kHz. It has been suggested that the efficiency of the process is increased by introducing or creating microbubbles or particles that resonate with the frequencies generated in the apparatus. The transducer method has been used on a small scale for producing emulsions and for disrupting cells, whereas the wedge method has been used for emulsification in food products

Emulsification may also be achieved by mixing high-speed streams of the two phases, the resultant mixture then being forced through a small orifice to create a high-velocity impinging jet. This method, known as microfluidization, is essentially a modification of the high-pressure homogenizer principle but can be employed for very small sample sizes.

As with other methods, there can be considerable heat production.

Part of this article was first published in the *Journal of the Society of Dairy Technology*, Volume 45 (1992) and was reproduced by the kind permission of the Society of Dairy Technology.

See also: **Colloids and Emulsions**; **Emulsifiers**: Organic Emulsifiers; Phosphates as Meat Emulsion Stabilizers; Uses in Processed Foods; **Ice Cream**: Microbiology; **Separation and Clarification**

Further Reading

Brennan JG, Butters JR, Cowell ND and Lilley AEV (1990) *Food Engineering Operations*, 3rd edn. London: Elsevier Applied Science.

Kessler HG (1981) *Food Engineering & Dairy Technology*. Freising: Ferlag A. Kessler.

McClements DJ (1999) *Food Emulsions; Principles, Practice and Techniques*. Boca Raton, FL: CRC Press.

Phipps LW (1985) *The High Pressure Dairy Homogenizer*. Reading, UK: National Institute for Research in Dairying.

Walstra P (1985) Formation of emulsions. In: Becher P (ed.) *Encyclopaedia of Emulsion Technology*, vol. 1, pp. 57–127. New York: Marcel Dekker.

Walstra P, Gearts TJ, Noomen A, Jellema A and van Boekel MAJS (1999) *Dairy Technology*. Marcel Dekker, New York.

HONEY

L W Doner, US Department of Agriculture, Wyndmoor, PA, USA

Background

Honey serves to assure the survival of honeybee colonies and, consequently, the pollination and survival of many plants. It was one of man's earliest foods, as honeybees were producing it long before man appeared on the earth. Honey remained the primary sweetener for humans until the late nineteenth century, when its consumption was exceeded by sugar cane-, sugar beet-, and later corn-derived sugars. Honey remains the only sweetener that can be stored and used precisely as it is produced by nature. It is an exceedingly complex substance that is produced by honeybees from floral nectar and, less frequently, honeydew. Honeydew consists of syrups deposited on plants by other insects, such as aphids. Whereas nectar from the many floral sources generally contains just the sugars fructose, glucose, and sucrose, honey contains these and many other more complex sugars. The additional sugars arise largely as a result of honeybees adding several enzymes during the nectar-to-honey ripening process. The presence of active enzymes and the acidity of honey result in its composition being in a constant state of change.

Comprehensive surveys of average honey composition have established that the major components are fructose (38.4%), glucose (30.3%), and water (17.2%). In addition to the two major sugars are an array of more than 20 higher sugars, which are formed by linking the fructose and glucose in various combinations. Honey is therefore primarily a carbohydrate material, and sugars comprise over 95% of its solids content. The principal physical characteristics of honey are due to its sugars, but subtle differences in minor constituents from nectar are responsible for the flavor differences of honey from different floral sources. These constituents include flavoring and aroma compounds, pigments, acids, and minerals. These substances are concentrated during the ripening process, where most of the water from nectar is removed. In order to assure the integrity of honey markets, methods have been developed to detect the addition of inexpensive sugar mixtures as adulterants.

Honey Production and Processing

To summarize, foraging honeybees ingest nectar and honeydew and store it in their honey sac. Pollen is also collected, and this provides the protein portion of the honeybee diet. Pollen, however, is not integral to the production of honey. The honey sac weighs 40–70 mg when full and can account for more than 90% of the weight of the bee. Enzymes are secreted and mixed with the sac contents. Upon returning to the hive, the contents are passed on to the house bees, who carry out the ripening process. This consists of alternately expelling and ingesting the honey sac fluid for 15–20 min. When the nectar has lost about half of its moisture, small droplets of this semiripened honey are deposited in the comb cells. In 1–3 days, numerous enzyme-mediated changes will have occurred and honey density attained. The cells are then capped over with wax and the honey stored as food. Compositional changes continue during ripening, the most prominent result being the production of an increasingly complex array of sugars. This is largely a result of continuing enzyme activity. An enzymatic process is also employed to produce hydrogen peroxide, in order to stabilize the honey against microbial growth.

The primary objectives of commercial honey processing include maintaining the desired physical state, be it liquid or finely granulated, and keeping it free of fermentation. Methods for accomplishing the objectives are well established and have been used for many years. Beekeepers remove the hive units (supers) from the colony and then remove the wax cappings by mechanical slicing. The honey is then centrifugally extracted from the wax comb. Raw honey contains pollen, some wax, sugar-tolerant yeasts, and often some crystals of glucose monohydrate. Flavor and aroma are optimal just after processing, but further treatment is required before large-scale marketing. The process of pressure filtration is used to produce liquid honey, after controlled heating to destroy yeasts and to dissolve sugar crystals. The application of heat to honey requires careful control so that the color, flavor, and aroma of honey are not impaired. Honey is sold in the liquid form, in a finely granulated form known as honey spread or finely granulated honey, and as comb honey. The requirements for the stability of these forms are increasingly stringent, as honey has become a world commodity available in markets throughout the year.

World honey production has increased steadily and, in the year 2000, was estimated to be about 1.2 million tonnes. About 25% of this production is the object of international trade. The three largest honey-exporting countries are China, Argentina, and Mexico, and the largest importers are Germany, Japan, and the USA. The USA is the leading consumer of honey, and honey consumption in Asia has increased substantially in recent years, especially in Japan and China.

General Composition

Honey samples vary in composition, and this variability reflects contributions of the plant source of nectar, beekeeping practices, and climate and environmental conditions. Honeys representing many floral sources in the USA have been analyzed, and the average compositions are listed in Table 1. Many honeys from other countries have been analyzed and generally found to provide similar values. There is wide variability among all the honey components. Among the major components, however, fructose levels exceed those of glucose, the only exceptions being dandelion and blue curls honeys.

Moisture

Moisture content is one of the most important characteristics of honey, as it influences the keeping quality, granulation, and body. The natural moisture of honey in the comb is that remaining from the nectar after it is ripened by the bees. The amount of moisture is a function of factors involved in ripening, such as environmental conditions, original moisture content of the nectar, and strength of the honeybee colony. Normal ripened honey has a moisture content below 18.6%, and when the moisture content exceeds that level, honeys are liable to ferment. This can occur if the count of ubiquitous osmophilic yeasts is sufficiently high. After extraction of the honey from the comb, its moisture content may change, depending on the conditions of storage. Low moisture levels can result in the crystallization of glucose, a process termed granulation. Few

Table 1 Average composition of honey[a]

Property	*Average*	*Range*
Moisture	17.2	12.2–22.9
Fructose	38.4	30.9–44.3
Glucose	30.3	22.9–40.7
Sucrose	1.3	0.2–7.6
'Maltose'[a]	7.3	2.7–16.0
Higher sugars	1.4	0.1–3.8
Ash	0.17	0.02–1.03
Nitrogen	0.04	0.0–0.13
pH	3.91	3.4–6.1

[a]Data for 490 samples; all values in percentages except for pH; 'maltose' includes all reducing disaccharides.
From White JW, Jr. (1962) *Composition of American Honeys. Technical Bulletin No. 1261*. US Department of Agriculture.

beekeepers measure the moisture content of honey, instead relying on guesswork and leaving this measurement to processors.

Several methods are available to measure honey moisture, including direct drying, viscosity measurement, refractometry, and use of the Karl Fischer titration method. The last method is the most accurate, but the hand refractometer is sufficiently accurate and more convenient to use. This instrument must be used with appropriate sugar solutions as standards.

Sugars

The simple sugars fructose and glucose account for over 85% of its solids, so honey is essentially a highly concentrated sugar solution in water. Another 10% of the honey solids includes at least 25 other more complex sugars (Table 2), giving it the most complex sugar profile of any natural product. Some of these sugars are present in very low levels, and they are formed by linking fructose and glucose in many different combinations.

Until the middle of the twentieth century, the only sugars thought to be present in honey were the nectar sugars fructose, glucose, and sucrose, and small amounts of an ill-defined, highly complex carbohydrate. After the ill-defined material was shown to consist of many individual sugars, researchers provided some explanations as to how they could arise from the nectar raw material. It was revealed that while the acids of honey may be responsible for generating some of the sugars, most are generated as a result of the action of enzymes contributed by the honeybees during the ripening process. The enzymatic formation of these sugars is discussed in the following section on the honey enzymes.

A variety of methods have been required to characterize the many honey sugars. Traditional colorimetric methods continue to be useful for determining levels of reducing and nonreducing sugars in honey, but these methods do little to reveal the high degree of complexity. Enzymatic methods can also be used, such as glucose oxidase for measuring glucose levels. Beginning in about 1950, various chromatographic methods have been shown to be extremely useful. Charcoal-celite chromatography continues to be applied for separating the honey sugars into mono-, di-, and higher saccharide fractions. These fractions are then further resolved using paper-, charcoal-celite column-, thin-layer-, high-performance liquid-, capillary gas-liquid-, and high-performance anion exchange chromatography. The latter two chromatographic modes have been especially useful in separating closely related sugars. After their purification, physical methods such as infrared and nuclear magnetic resonance spectroscopy have been applied to identifying the sugars listed in Table 2 as components of honey. Some of these are present at extremely low levels, providing additional challenges to the analyst. (*See* **Fructose**; **Glucose:** Properties and Analysis; **Sucrose:** Properties and Determination.)

Enzymes

Enzymes that have been identified as honey components are listed in Table 3, along with the number of samples analyzed. These are mostly added by the honeybee, although some traces of plant enzymes may be present. Honeybees add these enzymes in order to accomplish the nectar-to-honey ripening process, and they are largely responsible for providing the compositional complexity of honey. The processes involved in the conversion of the three nectar sugars into at least 25 additional sugars of greater complexity are far from being totally understood. It is becoming increasingly clear, however, that enzyme-catalyzed reactions are responsible for the formation of many of these sugars.

Table 2 Sugars that have been identified as honey constituents

Monosaccharides	*Disaccharides*	*Trisaccharides*	*Higher saccharides*
Fructose	Sucrose	Melezitose	Isomaltotetraose
Glucose	Maltose	Maltotriose	Isomaltopentaose
	Maltulose	Isomaltotriose	
	Isomaltose	3-α-Isomaltosyl glucose	
	Nigerose	1-Kestose	
	Turanose	Panose	
	Kojibiose	Isopanose	
	Laminaribiose	Erlose	
	α,β-Trehalose	Theanderose	
	Gentiobiose	Centose	
	Palatinose	Laminaritriose	
	Cellobiose		

From Doner LW (1977) The sugars of honey – A review. *Journal of the Science of Food and Agriculture* 28: 443–456.

Table 3 Enzymes that have been identified as honey constituents

Enzyme	*Number of samples*
α-Glucosidase	1468
β-Glucosidase	11
Glucose oxidase	124
Catalase	38
Acid phosphatase	25
α-Amylase	1746
β-Amylase	1746

From Doner LW (1991) The enzymes of honey. In: Fox PF (ed.) *Food Enzymology*, pp. 143–161. London: Elsevier Applied Science.

The enzyme α-glucosidase is probably the most significant honey enzyme, since it is responsible for many of the important changes that occur during riping. It is often referred to as invertase or sucrase in the honey literature. α-Glucosidase converts the nectar disaccharide sucrose into its constituent monosaccharides, fructose and glucose, a process known as inversion. This reaction is vital because sucrose crystallizes easily and also because its inversion allows the preparation of the super-saturated glucose/fructose mixture found in honey. α-Glucosidase catalyzes additional reactions, leading to the formation of several addition sugars. Researchers have found that incubation of the enzyme with fructose, glucose, and sucrose results in sugars not from hydrolysis reactions but from transglucosylation reactions. By this process, higher sugars are produced from simpler sugars. Sugars from Table 2 that have been identified are kojibiose, nigerose, maltose, isomaltose, turanose, α, β-trehalose, and erlose. The trisaccharide erlose was first discovered as a honey component, and is formed by the reaction of glucose with the glucose moiety of the disaccharide sucrose.

When the most recently discovered honey enzyme, β-glucosidase, has been incubated with fructose, glucose, and sucrose, other honey sugars have resulted. These were the glucose–glucose disaccharides, laminaribiose and gentiobiose. Thus, at least nine of the honey sugars are produced by the action of α- and β-glucosidase on the nectar sugars. The mode of formation of most of the honey trisaccharides remains to be elucidated.

Another honey enzyme contributed by the honeybee, glucose oxidase, is largely responsible for the antibacterial property of honey. This activity was termed inhibine until it was identified to be hydrogen peroxide. This antibacterial agent is produced by the action of glucose oxidase on its glucose substrate. In addition to hydrogen peroxide, gluconic acid is produced by the enzyme. The acidity of this sugar acid is largely responsible for the low pH of honey. In conjunction with the high solids content and resulting low water activity, it helps to stabilize honey against fermentation.

The honey enzymes catalase and acid phosphatase have not been studied in great detail. Catalase was discovered in honey in 1910 and is responsible for converting hydrogen peroxide into water and oxygen. The role of this enzyme appears to be in controlling the levels of hydrogen peroxide in honey, since there is an inverse relationship between the levels of enzyme and substrate.

Acid phosphatase was first reported in 1938 when incubation of β-glycerophosphate with diluted honey was found to result in formation of phosphoric acid and glycerol. The source of this enzyme is unclear. One study suggested that it is only present in slightly fermented honey samples and that yeast is the probable source. A more recent study suggested a plant source for acid phosphatases, since activities were found in pollen and nectar.

Diastase enzyme is added to nectar by bees during ripening, and is a mixture of α- and β-amylase. These enzymes are known to degrade starch, but since nectars are not known to contain starch, their role during ripening is unknown. The activity of honey diastase is known to be reduced upon heating and is easily measured. As a result, its measurement has been used for many years to estimate the extent of heating to which a honey has been exposed. This information is used in countries where such heating is believed to reduce or deplete certain ill-defined health-promoting properties. A problem with this approach is that diastase levels in fresh honeys vary considerably, and extended storage can significantly deplete its level. (*See* **Enzymes:** Functions and Characteristics.)

Minerals

The mineral content of honey is reflected in its total ash content (Table 1). In general, dark honeys have a higher mineral content than light honeys. Like other honey components, ash content is highly variable, ranging from 0.02% to over 1%. The levels of individual minerals that comprise the ash are also variable, but potassium accounts for about one-third of the total amount. Its level exceeds by about 10-fold that of the other major minerals in honey, sodium, calcium, and magnesium. Less abundant elements found in honey are iron, manganese, copper, chlorine, phosphorus, sulfur, and silicon. Sensitive instrumental methods such as atomic absorption or plasma emission spectroscopic analysis can be used to accurately determine levels of even the most trace minerals.

Acids

The great sweetness of honey largely masks flavor contributions of the acids, which generally account

for about 0.5% of the solids. The acids account for the low average pH (3.99) value, which is in part responsible for the excellent stability of honey. Several acids have been found in honey, gluconic acid being present in considerable excess over the others. It arises from glucose through the action of the enzyme glucose oxidase, added by the honeybee. The combined effect of its acidity and the hydrogen peroxide concurrently produced assist in preserving nectar and honey from spoilage. Less prominent acids reported to be in honey are formic, acetic, butyric, lactic, oxalic, succinic, tartaric, maleic, pyruvic, pyroglutamic, α-ketoglutaric, glycollic, citric, malic, 2- or 3-phosphoglyceric acids, and glucose 6-phosphate. (*See* **Acids:** Properties and Determination.)

Proteins and Amino Acids

Protein and amino acid levels in honey are reflected in nitrogen content, which is low, 0.04% on the average, but variable, ranging up to 0.1%. About 40–80% of the total nitrogen in honey is in protein, and most of the remainder resides in the free amino acids. About 20 different nonenzymatic proteins have been identified in honey, many of which are common to all honeys. Some appear to originate in the honeybee, and others in the plant source of nectar. Except for the honey enzymes, little is known about many of the proteins in honey. Their presence results in honeys having a lower surface tension than they would have otherwise, and this encourages the formation of fine air bubbles and a marked tendency to foam. (*See* **Protein:** Food Sources.)

The quantity of free amino acids in honey is small and of no nutritional significance. Honeys have been found to contain between 11 and 21 free amino acids. Proline accounts for about half of the total, most of which originates from the honeybee, rather than from nectar or pollen. Aside from proline, the amino acids glutamic acid, alanine, phenylalanine, tyrosine, leucine, and isoleucine are present in the highest levels. Amino acids react with some of the honey sugars to produce yellow or brown materials, and age-related darkening of honey is probably a result of these reactions.

Flavor, Aroma, and Color

The dominant flavor of honey is sweetness, arising from the blend of the major sugars fructose and glucose, and the wide array of other sugars. Sweetness is common to all honeys, regardless of the floral source of the nectar. Other components common to all honeys, such as gluconic acid and proline, also make contributions to the general flavor of honey. The uniqueness of a given honey is due to contributions of compounds from the wide variety of floral nectar sources. Honey assumes many of the subtle flavor and aroma characteristics of these sources, and there are perhaps as many variations in flavor and aroma as there are nectar sources. Flavors are sufficiently distinctive that experienced honey tasters can identify dozens of different floral sources. Typical flavors range from the most delicate and desirable to those that are harsh and objectionable. (*See* **Flavor (Flavour) Compounds:** Structures and Characteristics.)

As many as 129 components have been separated from extracts of honey, and many of these are volatile organic compounds. Each of these is present in very low levels. About two-thirds have been identified, and many of these probably contribute to honey flavor and aroma, and perhaps color. Low-molecular-weight aldehydes, ketones, alcohols, and esters, make up the bulk of the compounds. One compound present in all honeys is hydroxymethylfurfural. This is produced from decomposition of sugars, especially fructose, in the presence of acid. The high fructose level and low pH of honey combine to favor its formation, which is accelerated when honey is heated. Synthetic honey flavors contain a large proportion of phenylacetic esters, as most of these compounds have a honey taste and aroma. A closely related compound, phenylethyl alcohol, is present in several natural honeys. The flavor and delicate bouquet of honey are vulnerable to heat and improper storage, which can level the subtle distinctions among unique floral sources. Honey is heated during commercial processing in order to delay granulation and avoid fermentation, but careful attention must be given to the time and amount of heat, or valuable flavor and aroma compounds can be lost.

Generally, lighter colors are associated with the milder, more pleasantly flavored honeys. Honey cannot be judged solely on the basis of color, as some of the most distinctively and strongly flavored honeys, such as basswood, are very light, whereas very mild and pleasant honeys such as tulip poplar can be quite dark. Color limits are often specified in trading of honey, however, because flavor is so subjective. Several systems are used for the objective measurement of honey color, and some are designated as official in various countries. Seven classifications are used for international trade; water white, extra white, white, extra light amber, light amber, amber, and dark amber.

Quality Assurance

Much of what has been learned about the composition and properties of honey has resulted from research aimed at verifying its authenticity. Being of

limited supply and relatively expensive, for many years, honey has been a target for adulteration with inexpensive, highly refined sweeteners. High profits are realized when mixtures of these sweeteners and honey are illegally labeled and marketed as pure honeys. Sweeteners that have been mixed with honey over the years have included corn syrup, which contains mostly glucose, and hydrolyzed cane and beet sugars, which contain both fructose and glucose. High fructose corn syrup was the most recent sweetener used to adulterate honey, and it also contains fructose and glucose. Corn syrup is easy to detect, because glucose levels in adulterated mixtures are higher than those normally found in honey. The other three sweeteners are difficult to detect, because adulterated mixtures contain fructose and glucose within the range normally found in honeys.

Methods to detect adulteration of honey have been required to guarantee the integrity of honey markets, therefore assuring an adequate supply of honeybee colonies for crop pollination. In 1977, shortly after the advent of high fructose corn syrup, a method was developed to detect its presence or the presence of sugar cane-derived sweeteners in honey. The test was based on differences in the ratios $^{13}C/^{12}C$ between honey plants and corn and sugar cane plants. The tropical grasses sugar cane and corn employ a different photosynthesis process to fix carbon dioxide, and this results in higher ratios of $^{13}C/^{12}C$ than are found in honeys. These ratios can be precisely measured using isotope ratio mass spectrometry. The application of these tests by government agencies largely discouraged widespread adulteration of honey with sweeteners from corn and sugar cane. Adulteration strategies have shifted to the addition of hydrolyzed beet sugar syrups, which have $^{13}C/^{12}C$ values in the same range as honey. Some progress is being made in its detection, as some oligosaccharides are present in the adulterants that are not present in honey, and these can be detected using sophisticated chromatographic approaches.

Uses of Honey

Most of the honey produced is used directly as a table sweetener or spread. The most significant indirect uses are in baking, cereal, confectionary, and meads. Honey is primarily a high-energy carbohydrate food, possessing distinct flavors that cannot be found elsewhere. The honey sugars are largely the easily digestible simple sugars, similar to those in many fruits. The protein and enzymes of honey, though used as indicators of heating history and hence table quality in some countries, are not present in sufficient quantities to be considered nutritionally significant. Several of the essential vitamins are present in honey, but in insignificant levels. The mineral content of honey is variable, but in the darker honeys, the level can be quite high. Honey is also included in many over-the-counter pharmaceutical preparations, providing flavor and textural properties valued by consumers.

See also: **Acids**: Natural Acids and Acidulants; **Amino Acids**: Properties and Occurrence; **Enzymes**: Functions and Characteristics; **Fructose**; **Glucose**: Properties and Analysis; **Protein**: Chemistry; **Quality Assurance and Quality Control**; **Sensory Evaluation**: Descriptive Analysis; Aroma; Taste

Further Reading

Aureli P, Franciosa G and Fenicia L (2002) Infant botulism and honey in Europe: a commentary. *Pediatric Infectious Diseases Journal* 21(9): 866–868.

Crane E (ed.) (1979) *Honey: A Comprehensive Survey*. London: Heinemann.

Doner LW (1977) The sugars of honey – A review. *Journal of the Science of Food and Agriculture* 28: 443–456.

Doner LW (1991) The enzymes of honey. In: Fox PF (ed.) *Food Enzymology*, pp. 143–161. London: Elsevier Applied Science.

Gheldof N, Wang XH and Engeseth NJ (2002) Identification and quantification of antioxidant components of honeys from various floral sources. *Journal of Agricultural Food Chemistry* 9; 50(21): 5870–5877.

Graham JM (ed.) (1992) *The Hive and the Honey Bee*. Hamilton, IL: Dadant & Sons.

Low NH, Va Vong K and Sporns P (1986) A new enzyme, β-glucosidase in honey. *Journal of Apicultural Research* 25: 178–181.

White JW, Jr. (1962) *Composition of American Honeys*. *Technical Bulletin No. 1261*. US Department of Agriculture.

White JW, Jr. (1978) Honey. In: Chichester CO, Mrak EM and Stewart GF (eds) *Advances in Food Research*, pp. 288–374. New York: Academic Press.

White JW, Jr. (1992) Honey. In: Graham JM (ed.) *The Hive and the Honey Bee*, pp. 869–925. Hamilton, IL: Dadant & Sons.

HORMONES

Contents

Adrenal Hormones

L N Parker, University of California, Long Beach, CA, USA

Background and History

Hormones secreted by both parts of the adrenal gland, the cortex and medulla, affect metabolism of proteins, carbohydrates, fat, minerals, and water. The human adrenal glands are paired structures located at the superior pole of each kidney. They are compound glands, composed of an outer cortex and an inner medulla. The cortex secretes three classes of steroid hormones. In humans, the main examples of these hormones are cortisol, a glucocorticoid; dehydroepiandrosterone (DHEA), an androgen; and aldosterone, a mineralocorticoid.

The adrenal medulla secretes nonsteroidal hormones with a catechol nucleus, among them epinephrine (adrenaline), norepinephrine (noradrenaline), and dopamine. Norepinephrine and epinephrine are primarily involved in the biological alarm mechanism of the sympathoadrenal system. The embryological development of the adrenal cortex and medulla are different, as are their control mechanisms and functions, and in some nonhuman vertebrates, including fish, are located apart from each other.

The adrenal glands were described in the sixteenth century by Eustachius, although their functions remained unknown for approximately three centuries. In the nineteenth century, histological studies revealed three zones in the adrenal cortex, and an inner medullary zone which stained differently from the cortex in the fetal and adult adrenal gland. The functions of the adrenal glands were then gradually elucidated. In 1855, Thomas Addison described a group of patients with anemia, pigmented skin, gastrointestinal symptoms, and marked weakness, who were found at autopsy to have atrophied adrenal glands. Their condition of adrenal failure still bears his name, Addison's disease. At approximately the same time, the physiologists Claude Bernard and Charles Brown-Séquard popularized the concept of internal secretions by glands, and showed by adrenalectomies in animals that the adrenal glands are essential for life.

The next clues about the role of the adrenal glands were discovered beginning at the turn of the twentieth century, when a substance with vasopressor activity was isolated from the adrenal medulla, purified and named adrenaline, or epinephrine. In retrospect, epinephrine thereby became the first hormone with a known chemical structure. Shortly thereafter, it was hypothesized that the adrenal glands controlled salt balance as well as blood pressure when it was found that sodium salts could prolong the life of adrenalectomized animals. A generation later, by the use of lipid, rather than aqueous extracts, steroid hormones were sequentially discovered and purified after it was observed that crude lipid extracts of adrenal glands could keep adrenalectomized animals alive indefinitely. The first adrenal steroid hormone synthesized was deoxycorticosterone (DOC), a weak mineralocorticoid, in 1937. The first glucocorticoid synthesized was cortisone in 1949, and by 1955 the main mineralocorticoid in humans, aldosterone, had been characterized and synthesized. Experiments soon demonstrated adrenocortical control by the pituitary gland, and in turn hypothalamic control of adrenocorticotropin (ACTH) was demonstrated conclusively by the characterization and synthesis of corticotropin-releasing hormone (CRH) in 1981. More recent research has focused on many aspects of adrenal gland function, such as hormone synthesis, hormone receptors, and mechanisms of intracellular action. Clinical research has shown the clinical usefulness of epinephrine and related substances in the treatment of asthma, glucocorticoids in the treatment of inflammatory and other conditions, and mineralocorticoids and related substances in the management of autonomic insufficiency.

Structure of the Adrenal Glands and Hormones Secreted

Adrenal Cortex

The location of the human adrenal glands and the structures surrounding them are shown in Figure 1. The weight range of each adrenal gland is approximately 3.5–5 g, and the cortex comprises about 90% of the gland volume. Chronically increased ACTH secretion causes adrenal weight to increase. The adrenal glands, normally adjacent to the kidneys, are surrounded by fat, and can be separated from the kidney in obesity. Occasionally, accessory adrenal tissue is found in the connective tissue near the main glands, most commonly in locations such as in the retroperitoneal celiac plexus, in the hilum of the spleen, near the ovaries or broad ligament, in the scrotum or in the liver. Embryologically, the adrenal cortex is derived from mesenchymal cells adjacent to the urogenital ridge. The fetal adrenal can be recognized by 2 months of gestation, at which time it is invaded by neuroectodermal cells which will form the medulla. By the second trimester, there is a thin outer definitive zone, which will later form the adult adrenal cortex, but the inner fetal zone comprises most of the adrenal mass. After birth, the fetal zone, which secretes mostly DHEA sulfate (DHEAS), rapidly involutes by 2 months of age, and disappears by 1 year.

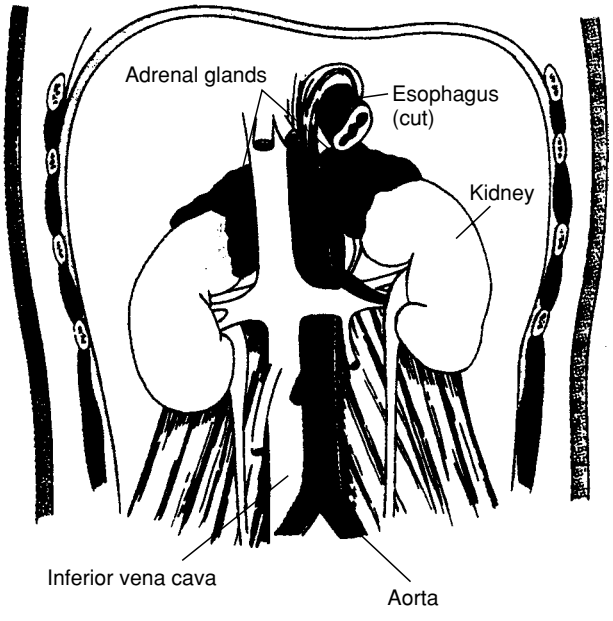

Figure 1 Location and surrounding structures of the human adrenal glands. Reproduced from Gaudin A and Jones K (1989) *Human Anatomy and Physiology*. San Diego: Harcourt Brace Jovanovich, with permission.

By 1 year of age, three zones can be identified by light microscopy in the definitive adrenal cortex, as shown in Figure 2. The outer zone, the zona glomerulosa, is relatively thin, and consists of cells which secrete aldosterone. The middle zone, the zona fasciculata, is usually the thickest layer of the adrenal cortex, and has a columnar structure. Its cells are relatively clear, since they are large and have a high lipid content. The inner zone, the zona reticularis, surrounds the medulla. Its cells are relatively dark-staining and compact in appearance, and often contain lipofuscin pigment granules. Both the zona fasciculata and the zona reticularis produce cortisol and androgens, but in the human, the zona reticularis has sulfotransferase activity, and produces DHEAS. Chronically increased ACTH concentrations result in lipid depletion from the zona fasciculata and an increase in the width of the zona reticularis.

Adrenal Medulla

The adrenal medulla is part of the sympathetic nervous system, which arises from cells of the neural crest during embryonic development. Storage granules which stain brown with chromic acid due to oxidation of catecholamines to melanin give the cells which contain them the name chromaffin or pheochrome cells. These cells are also found on both sides of the aorta, and comprise the paraganglia. The largest collection of these cells is found near the inferior mesenteric artery, where the cells fuse to form a fetal structure called the organ of Zuckerkandl, which undergoes involution within the first year of life. The remainder of the chromaffin cells in the paraganglia and adrenal medulla persist during adult life and secrete norepinephrine, epinephrine, and dopamine. In the adrenal medulla, the cells are arranged in an irregular network with a rich blood supply, and are in contact with sympathetic ganglia. The cells of the adrenal medulla are innervated by preganglionic

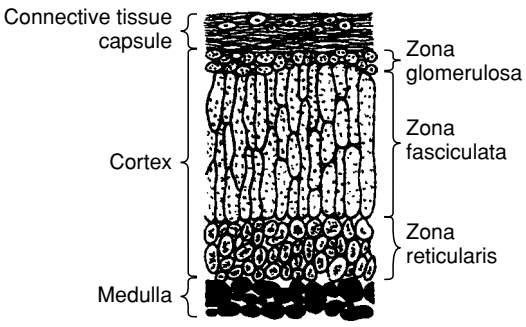

Figure 2 Histology of the human adrenal gland. Reproduced from Gaudin A and Jones K (1989) *Human Anatomy and Physiology*, San Diega: Harcourt Brace Jovanovich, with permission.

fibers of the sympathetic nervous system. Most of the blood supply of the adrenal glands enters through the cortex and drains into the medulla, except for some vessels which supply the medulla directly.

Physiological Effects of Adrenal Gland Hormones

Adrenal Cortex

As in the case of other steroid hormones, glucocorticoids exert their effects on target cells by interacting with soluble cytosolic receptor proteins, after mostly passive diffusion into cells. The glucocorticoid receptor is a single-chain polypeptide with a molecular weight of 94 kDa which binds glucocorticoids with a high affinity. After binding glucocorticoids at the COOH-terminus, the DNA-binding region in midmolecule binds to the nucleus. The glucocorticoid receptor is a member of a family of DNA-binding proteins which act as regulators of gene transcription, including receptors for all types of steroid hormones, the thyroid hormone receptor, and the retinoic acid receptor. The hormone-binding domain of these receptors is well conserved, but the most distinctive feature is the highly conserved central DNA-binding domain. This is the zinc-finger domain, a series of finger-like loop structures of amino acids anchored at the base by a zinc ion chelated between two pairs of cysteine and histidine residues (Figure 3), which interact with coils of the DNA double helix. Glucocorticoid receptor complexes interact with specific glucocorticoid response elements, usually located near the promoter region of target genes.

Glucocorticoids are named for their effect of increasing hepatic glucose production by several mechanisms. They increase activity of hepatic gluconeogenic enzymes such as glucose-6-phosphatase and phosphoenolpyruvate carboxykinase, in the latter case by gene transcription. Glucocorticoids also activate glycogen synthase and inactivate the glycogen-mobilizing enzyme glycogen phosphorylase, with a net increase in hepatic glycogen synthesis. Finally, glucocorticoids decrease peripheral glucose utilization in part by decreasing adipocyte glucose transporter mRNA levels.

With respect to protein and fat metabolism, glucocorticoids stimulate catabolism of nonhepatic proteins with a resultant release of glucogenic amino acids, largely from skeletal muscle, and an increased uptake into liver for gluconeogenesis. In adipose tissue, glucocorticoids stimulate lipolysis and subsequent release of glycerol and free fatty acids, while enhancing the rate of fat oxidation. Pharmacological amounts of glucocorticoids cause osteopenia on a chronic basis, by inhibiting osteoblast function, thereby decreasing new bone formation, and by decreasing intestinal calcium absorption.

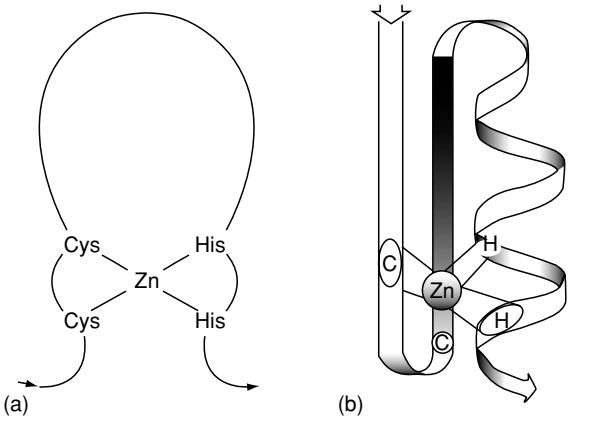

Figure 3 Two postulated models (a, b) of zinc-finger structures in the DNA-binding region of the glucocorticoid receptor. Reproduced from Evans R and Hollenberg S (1988) Zinc Fingers. *Cell* 52: 1. Cambridge, MA: Cell Press, with permission.

Cortisol is a major stress hormone and is secreted in so many different forms of stress that an increase in cortisol secretion is often considered to be part of the definition of a stressful stimulus. It has been speculated that cortisol functions to aid survival in stress by improving the metabolic milieu by means of the energy-producing and biosynthetic pathways described above. The inflammatory process is common in stress caused by illness or injury, and cortisol may also help to minimize damage to the body resulting from excessive inflammation by stimulating mechanisms such as lysosomal membrane stabilization, which prevents release of proteolytic enzymes, and reduction in capillary permeability to avoid leakage of plasma and blood cells into an inflamed area. Among the mechanisms by which cortisol performs these functions is the inhibition of the action of histamine, and the synthesis of prostaglandins. Cortisol inhibits the accumulation of neutrophils at sites of inflammation. Cortisol also has numerous other effects, such as maintenance of normal cardiac output and renal blood flow, and modification of immunological responses. One mechanism for the modification of immunological responses is the promotion of lymphocyte apoptosis.

Adrenal androgens such as DHEAS circulate at high concentrations in young adults, and in the case of DHEAS, the concentration is much greater than that of cortisol. However, their functions are not as clearly elucidated. There is evidence that these steroids have a hepatic receptor, and that they may prevent osteoporosis, facilitate the birth process by

causing cervical softening, modify immunological processes, mediate female libido, and serve as precursors for the more potent sex steroids. In animal studies these steroids have been shown to protect against obesity, diabetes, and certain types of infections and tumors. There is DHEAS in the human brain, and it has been shown to be synthesized in rat brain, but its function in the nervous system of either species is not known.

Aldosterone and other mineralocorticoids maintain normal sodium and potassium concentrations and intravascular volume by regulating electrolyte transport across epithelial surfaces. The specific renal type I mineralocorticoid receptor is unrelated to the glucocorticoid receptor at the amino-terminus, but partially homologous at the steroid-binding COOH-terminus, and highly homologous at the DNA-binding zinc-finger domain. Cortisol displaces aldosterone from this receptor, but *in vivo* is converted locally to cortisone by the microsomal 11-β hydroxysteroid dehydrogenase enzyme, and cortisone has relatively little affinity for the mineralocorticoid receptor. After binding to the mineralocorticoid receptor, mineralocorticoids induce protein synthesis and subsequent activation of a sodium pump, which transports sodium across cell membranes. In the kidney, sodium is absorbed, and potassium and hydrogen ions are secreted. Under the influence of aldosterone, sodium is therefore conserved, and potassium and hydrogen ions are excreted into the urine. Aldosterone has similar effects in sweat glands, salivary glands, and in the intestinal lumen.

Adrenal Medulla

Catecholamines exert their effects via two classes of receptors, α and β. Adrenergic receptors interact with catecholamines on the extracellular side of the plasma membrane, and with G proteins within the cell membrane. Adrenergic receptors are part of a large family of membrane-associated proteins which includes muscarinic acetylcholine receptors. Regions of these proteins show considerable homology with each other, especially the membrane-spanning domains. As with other receptors which are coupled to G proteins, there are seven hydrophobic membrane-spanning domains connected by three extracellular and cytoplasmic loops. Stimulation of the α class of receptors by norepinephrine results in a variety of actions, such as vasoconstriction and blood pressure elevation, pupillary dilatation, and bladder sphincter contraction. Epinephrine also stimulates α receptors. However, in addition, epinephrine also stimulates the β types of receptors, which result in actions such as vasodilatation, acceleration of the heart rate, bronchodilatation, glycogenolysis, and lipolysis. These actions are part of the physiological response to stressful stimuli, and complement those of cortisol. Dopamine is a weak agonist of both the α and β classes of receptors, but in addition, there are several classes of dopaminergic receptors in the central nervous system and peripheral tissues.

In liver, catecholamines promote glucose output by activating glycogenolysis, accelerating gluconeogenesis, and inhibiting glycogen synthesis. Stimulation of adrenergic receptors by catecholamines increases activity of adrenyl cyclase and conversion of glycogen phosphorylase from the inactive to the active form, while there is a simultaneous increase of amino acid uptake by the liver, which increases substrate availability for gluconeogenesis. In adipose tissue, catecholamines also stimulate lipolysis by increasing activity of hormone-sensitive lipase, and the cleavage of triglycerides into fatty acids and glycerol for increased gluconeogenesis and energy availability.

Synthesis and Control of Adrenal Gland Hormones

Adrenal Cortex

The adrenal cortex produces cortisol in response to ACTH secreted by the pituitary gland. The secretion of ACTH occurs in response to decreased circulating concentrations of cortisol, as part of a negative-feedback system, and in response to stressors of many types, including surgery, hemorrhage, thermal injury, and hypoglycemia. In addition, there is a circadian rhythm of pulsatile ACTH and cortisol secretion which in humans results in increased secretion towards the end of the sleep period, and therefore higher levels of circulating cortisol in the morning. ACTH is a 39-amino-acid peptide derived from a larger molecule, 241-amino-acid proopiomelanocortin (POMC). The first 24 amino acids of ACTH are constant and bioactive in many species. POMC secretion is under the control of neurotransmitters and hypothalamic CRH, as shown in Figure 4. POMC undergoes extensive posttranslational processing, producing many peptides, including ACTH, as shown in Figure 5. There is evidence from animal experiments that non-ACTH POMC peptides may synergize with ACTH in control of corticosteroid secretion. Concentrations of CRH, a 41-amino-acid peptide, are high in the hypophysical portal system.

The mechanism of action of ACTH involves many steps, including initial binding to an adrenal cell surface receptor, which ACTH upregulates, thereby increasing the steroidogenic response to ACTH stimulation. ACTH binding activates adenyl cyclase

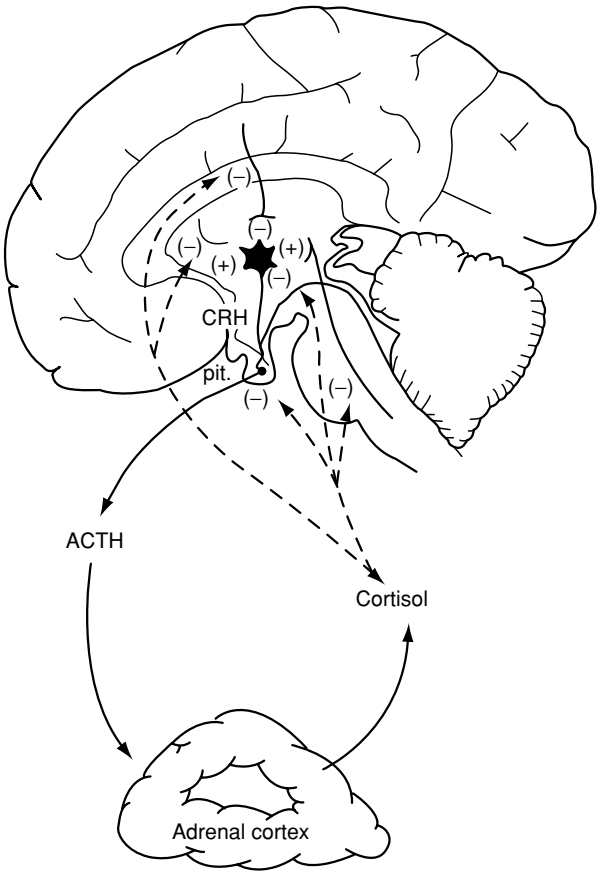

Figure 4 Cortisol feedback loop in the hypothalamic–pituitary–adrenal system. CRH, corticotropin-releasing hormone; pit, pituitary gland; ACTH, adrenocorticotropic hormone. Cortisol negative feedback shown by dashed lines. Reproduced from Darlington D and Dallman M (1990) *Feedback Control in Endocrine Systems. Principles and Practice of Endocrinology and Metabolism*. Philadelphia: JB Lippincott, with permission.

which increases cyclic 3′,5′-monophosphate (cAMP) concentrations, which in turn activate cAMP-dependent protein kinase and phosphorylation of a number of proteins in the presence of extracellular calcium. There is an increase in activity of cholesterol ester hydrolase, which produces free cholesterol for the rate-limiting conversion of cholesterol to pregnenolone. Intracellular delivery of cholesterol to the inner mitochondrial membrane is facilitated by the 30-kDa steroidogenic acute regulatory protein, induced by cAMP. The conversion of cholesterol to pregnenolone then proceeds via the mitochondrial side-chain cleavage enzyme. Plasma lipoproteins also provide cholesterol for steroidogenesis.

As shown in Figure 6, pregnenolone can be converted to mineralocorticoids, glucocorticoids, or androgens. Many of the microsomal enzymatic steps are controlled by ACTH by regulation of the rate of steroidogenic enzyme synthesis. In contrast to the situation in the human, there is little activity of the enzyme 17β-hydroxylase (Figure 6) in rodents, and therefore the main glucocorticoid is corticosterone in this species.

Cortisol and ACTH secretion are closely related, and linked by a negative-feedback loop, as described above. Adrenal androgens, such as DHEA, are also secreted in response to acute ACTH stimulation, but their control is more complex since in some situations they are not secreted in parallel with cortisol. These situations include adrenarche, puberty, aging, polycystic ovarian syndrome, stress, serious illness, anorexia nervosa, and starvation. Adrenarche is the process of adrenal gland maturation which occurs before puberty at approximately age 7, and which involves increased secretion of DHEA and DHEAS along with unchanged secretion of cortisol. In contrast, during aging, while basal levels of cortisol are unchanged, levels of DHEAS peak in young adulthood in both sexes, and then decline markedly, as shown in Figure 7. The decrease in concentrations of adrenal androgens in stress, serious illness, anorexia nervosa, and starvation is somewhat similar to the situation in aging, except that it is reversible with recovery. The reason for the dissociation between cortisol and adrenal androgen secretion is not clear, but may involve regulation by a non-ACTH POMC-related peptide.

Aldosterone is produced only by the zona glomerulosa, because it is the only zone with aldosterone synthase activity. The zona glomerulosa does not have 17β-hydroxylase activity, and does not produce cortisol. Although ACTH causes acute stimulation of aldosterone secretion, there are other more important control mechanisms in addition to ACTH. Angiotensin II is the major regulator of aldosterone secretion. Secretion of angiotensin II is controlled by renin, as shown in Figure 8. Renin release is regulated primarily by the sodium concentration of fluid in contact with the renal juxtaglomerular cells, and renal perfusion, as sensed by renal baroreceptors. Increased renin release is stimulated by decreased sodium concentration or renal arteriolar blood pressure. Renin mediates the conversion of hepatic renin substrate (angiotensinogen) to the 10-amino-acid peptide, angiotensin I, which in turn is converted to the 8-amino-acid peptide angiotensin II, a potent vasopressor, by the converting enzyme in lung and other tissues. The 7-amino-acid peptide angiotensin III is also bioactive in stimulation of aldosterone secretion.

Angiotensins II and III bind to high-affinity zona glomerulosa cell surface receptors and stimulate aldosterone production from cholesterol by a calcium-dependent mechanism involving activation of protein

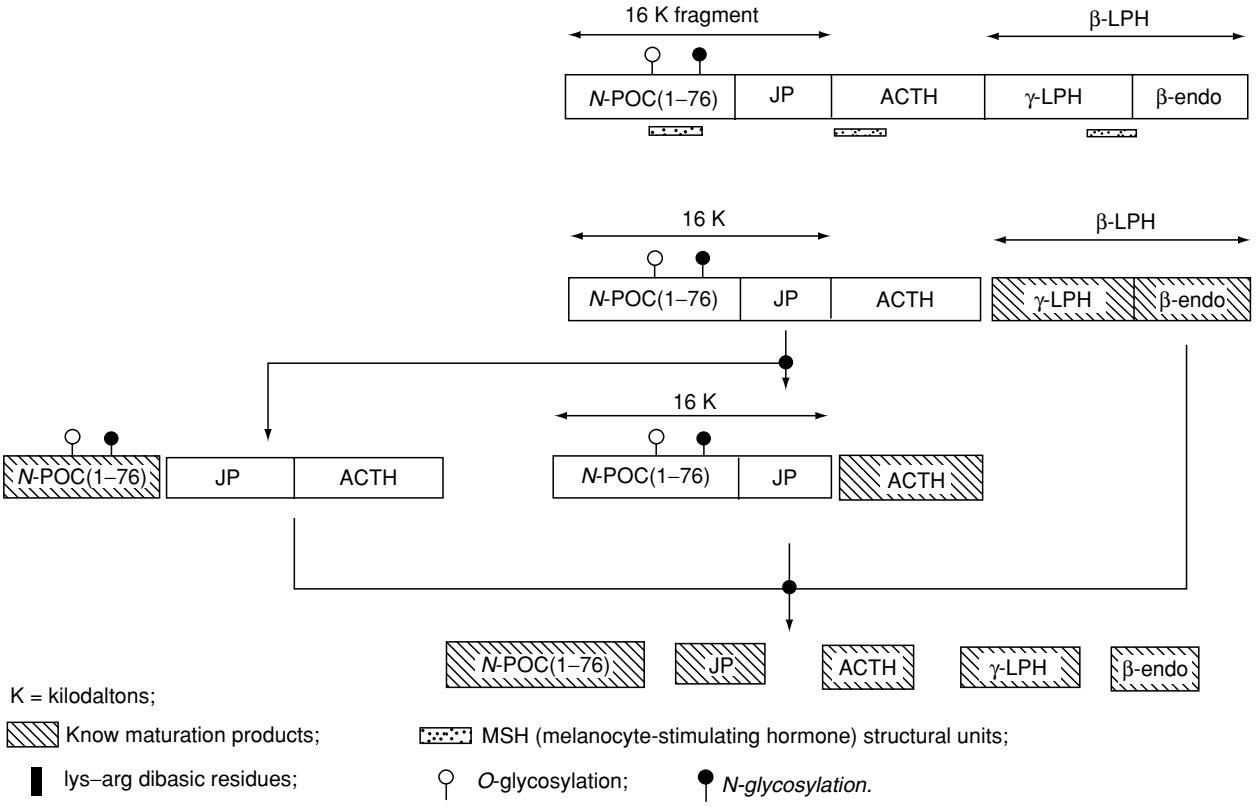

Figure 5 Structure and processing of proopiomelanocortin (POMC; POC). ACTH, adrenocorticotropic hormone; LPH, lipotropin; endo, endorphin; JP, joining peptide. Reproduced from Seidah N et al. (1981) The missing fragment of the pro-sequence of human POMC: sequence and evidence for C-terminal amidation. *Biochemistry Biophys Research Commun* 102: 710, with permission.

kinase C, rather than adenylate cyclase. In addition, aldosterone synthesis may be modified by lipoxygenase pathway metabolites of arachidonic acid. Potassium ion also influences aldosterone secretion by a direct effect on zona glomerulosa cells, and forms the basis for a feedback mechanism to regulate the concentration of extracellular potassium ions. Concentrations of potassium ion have the opposite effect on renin concentrations, but the direct effect on aldosterone secretion is predominant. In addition, as in the case of cortisol and adrenal androgen secretion, there is evidence that an influence on aldosterone secretion may be exerted by non-ACTH POMC-related peptides.

Adrenal Medulla

Control of the adrenal medulla is exerted by the sympathetic nervous system. Whereas preganglionic fibers of the parasympathetic branch of the autonomic nervous system emerge from cranial and sacral spinal nerves, those of the sympathetic nervous system emerge from the thoracic and lumbar spinal nerves and innervate many organs, including the adrenal medulla. Preganglionic nerve impulses are transmitted to postganglionic fibers by liberation of acetylcholine at nerve terminals. This results in secretion of catecholamines by the peripheral sympathetic nervous system and by the adrenal medulla. Norepinephrine is the major secretory product of the peripheral nervous system, and is primarily a neurotransmitter, not a circulating hormone. In the human adrenal medulla, the ratio of epinephrine to norepinephrine secretion is approximately 4:1, and concentrations of epinephrine, a circulating hormone, are sufficient to stimulate adrenergic receptors.

Catecholamine biosynthetic pathways are shown in Figure 9. The rate-limiting step in catecholamine biosynthesis is the initial conversion of tyrosine to dihydroxyphenylalanine (DOPA) by tyrosine hydroxylase (TH), an enzyme which is inhibited by DOPA, dopamine, and norepinephrine. Tyrosine itself is derived from the diet, or converted in the liver from phenylalanine by phenylalanine hydroxylase. The decarboxylation of DOPA to dopamine is catalyzed by aromatic-L-amino acid decarboxylase (AADC). Unlike the other enzymes of catecholamine biosynthesis, AADC is not only found in sympathetic nerve

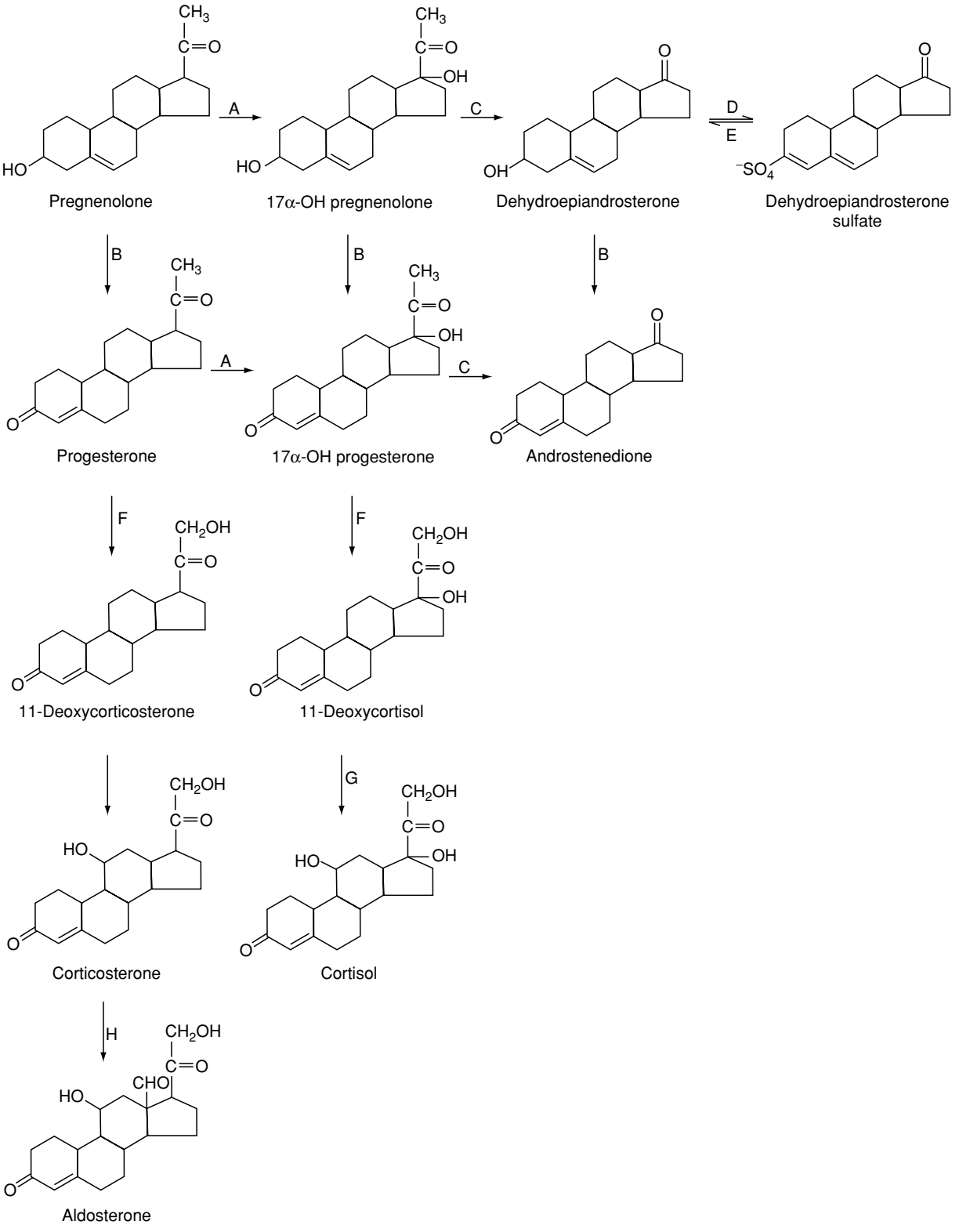

Figure 6 Human adrenocortical steroidogenic pathways. Enzyme activities: A, 17α-hydroxylase (CYP 17); B, 3β-hydroxysteroid dehydrogenase-isomerase (3β-HSD II); C, C17–20 desmolase (CYP 17); D, steroid sulfotransferase; E, steroid sulfatase; F, 21-hydroxylase (CYP21A2); G, 11β-hydroxylase (CYP11B1); H, aldosterone synthase (CYP11B2). Reproduced from Parker (1989) *Adrenal Androgens in Clinical Medicine*. San Diego: Academic Press, with permission.

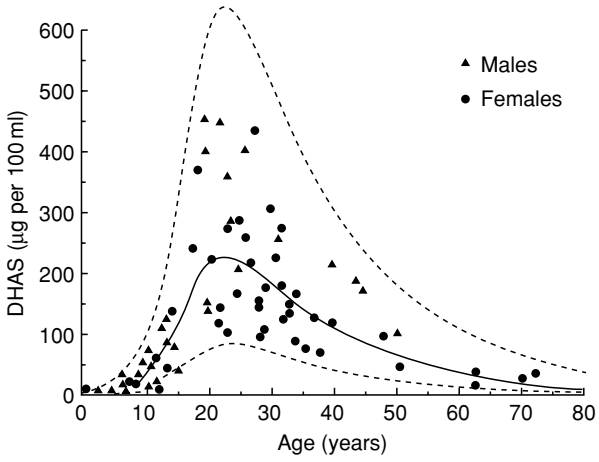

Figure 7 Serum concentrations of dehydroepiandrosterone sulfate (DHAS) in normal subjects 1–73 years of age. Reproduced from Smith M, et al. (1975) A radioimmunoassay for the estimation of serum DHAS in normal and pathological sera. *Clinica Chimica Acta* 65: 5, with permission.

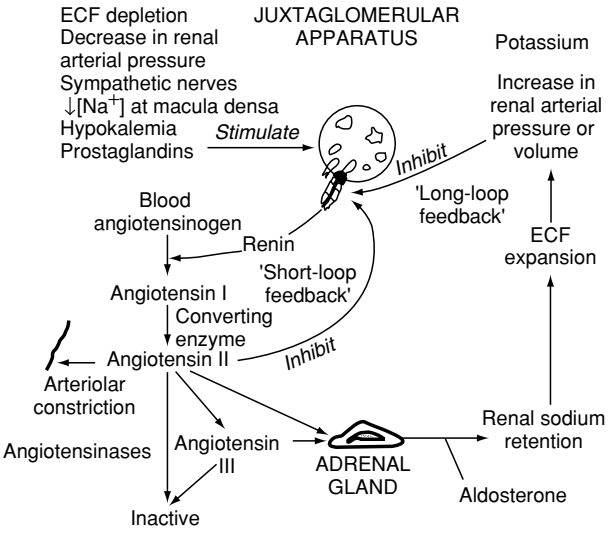

Figure 8 Control of aldosterone secretion by the renin–angiotensin system. ECF, extracellular fluid. Reproduced from Bondy P and Rosenberg L (1980) *Metabolic Control and Disease*. Philadelphia: WB Saunders, with permission.

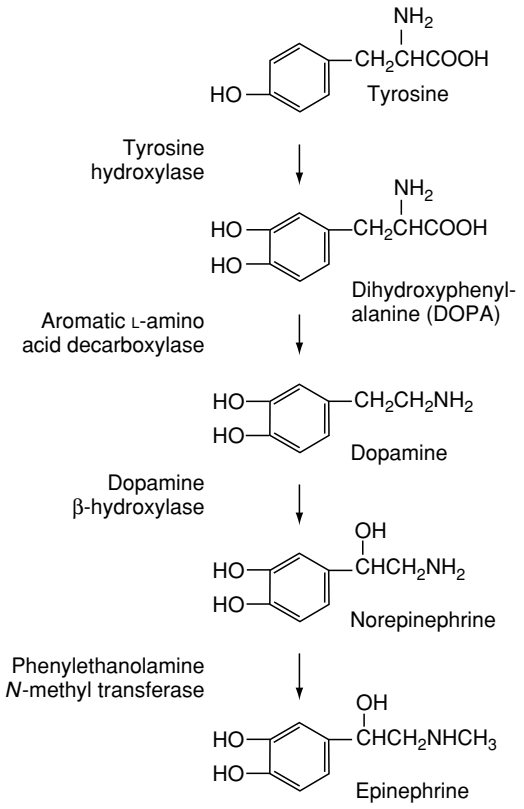

Figure 9 Catecholamine biosynthetic pathway of the sympathetic nervous system. Reproduced from Cryer P (1987) Diseases of the sympathochromaffin system. In: Felig P (ed.) *Endocrinology and Metabolism*. New York: McGraw-Hill, with permission.

endings and the adrenal medulla, but also in nonneuronal tissues, including the kidney, and catalyzes the decarboxylation of other aromatic amino acids in addition to DOPA, including 5-hydroxytryptophan, which is converted to serotonin.

Dopamine β-hydroxylase (DBH) catalyzes the hydroxylation of dopamine to form norepinephrine. The enzyme is found in sympathetic nerve endings, and also in adrenal medullary chromaffin granules, which release DBH along with norepinephrine when stimulated. The major difference between the catecholaminergic pathways of the adrenal medulla and the peripheral nervous system is the presence of the enzyme phenylethanolamine-*N*-methyl transferase (PNMT) in the medulla. This enzyme, also found in small amounts in the brain, catalyzes the conversion of norepinephrine to epinephrine, using *S*-adenosylmethionine as a methyl donor. PNMT is inducible by high concentrations of cortisol which are present in the capillary sinusoidal circulation from the adrenal cortex to the medulla.

A large percentage of synaptically released catecholamines are inactivated by reuptake into storage granules. Metabolism of circulating catecholamines occurs via two main pathways, mediated by the enzymes catechol-*O*-methyltransferase (COMT) and by monoamine oxidase (MAO), as shown in Figure 10. The end product of norepinephrine and epinephrine metabolism, after transformation by both enzymes, is 3-methoxy-4-hydroxymandelic acid (vanillylmandelic acid). Catecholamines in the liver and gut are also inactivated by sulfate or glucuronide conjugation of the phenolic hydroxyl group.

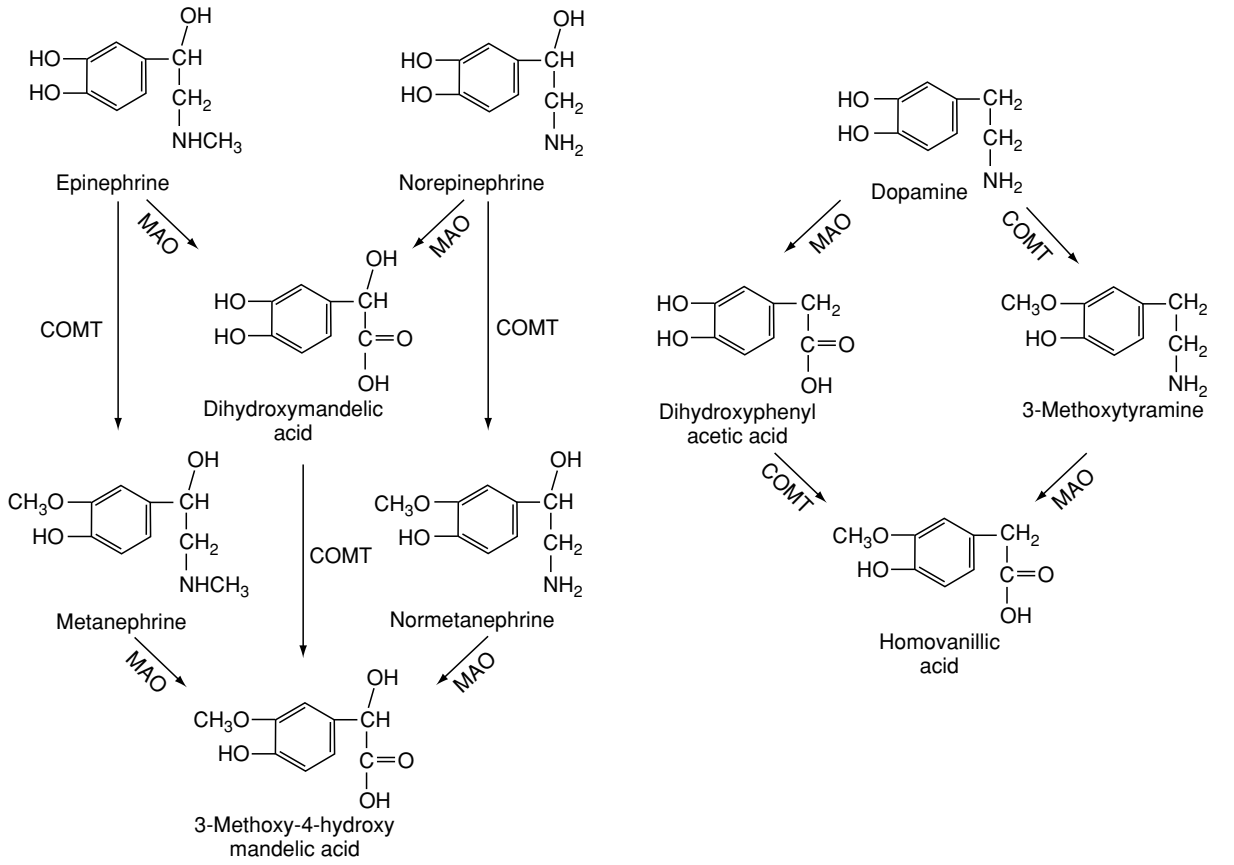

Figure 10 Metabolism of catecholamines by catechol-*O*-methyltransferase (COMT) and monoamine oxidase (MAO). Reproduced from Goldfien A (1986) The adrenal medulla. In: Greenspan F and Forsham P (eds) *Basic and Clinical Endocrinology*. Los Altos, CA: Lange Medical Publications, with permission.

The hypothalamus is the main regulator of sympathetic nervous system function. Impulses from the posterior and lateral hypothalamus result in generalized discharge of the sympathetic nervous system, including the adrenal medulla, although they can also be activated separately. The sympathoadrenal system is characterized by speed and integration, in that catechol-mediated events can take place within seconds, and can coordinate vascular, metabolic, and hormonal components. As discussed above, these responses occur to a variety of noxious, threatening, or stressful stimuli. In addition, the sympathetic nervous system is instrumental in maintaining an appropriate circulating volume and cardiac output during changes of posture from supine to upright. These feedback systems are mediated by sensors in the carotid sinuses, aorta, and medulla, which detect changes in circulatory volume and blood pressure. Although of different embryological origins, and operating via different regulatory mechanisms, the hypothalamic–pituitary–adrenocortical system and the sympathoadrenal system complement each other in the maintenance of homeostasis and a stable metabolic milieu in response to many forms of stress.

See also: **Amino Acids**: Metabolism; **Fatty Acids**: Metabolism; **Renal Function and Disorders**: Kidney: Structure and Function; **Stress and Nutrition**

Further Reading

Baxter J and Tyrrell J (1987) The adrenal cortex. In: Felig P (ed.) *Endocrinology and Metabolism*. New York: McGraw-Hill.

DeQuattro V, Myers M and Campese V (1989) Anatomy and biochemistry of the sympathetic nervous system. In: DeGroot L (ed.) *Endocrinology*. Philadelphia: WB Saunders.

Ganong W (1991) *Review of Medical Physiology. The Adrenals*. Norwalk, CT: Appleton & Lange.

Guyton A (1986) The adrenocortical hormones. In: *Textbook of Medical Physiology*. Philadelphia: WB Saunders.

Hale A and Rees L (1989) ACTH and related peptides. In: DeGroot (ed.) *Endocrinology*. Philadelphia: WB Saunders.

Loriaux DL (1990) The adrenal glands. In: Becker K (ed.) *Principles and Practice of Endocrinology and Metabolism*. Philadelphia: JB Lippincott.

Orth D and Kovacs W (1998) The adrenal cortex. In: *Williams Textbook of Endocrinology*. Philadelphia: WB Saunders.

Parker L (1989) Adrenal androgens: normal physiology. In: *Adrenal Androgens in Clinical Medicine*. San Diego: Academic Press.

Thyroid Hormones

J Vanderpas, Faculté Médecine, FUNDP, Namur, Belgium

Iodinated Thyroid Hormones

Thyroid Follicles and their Development

The thyroid gland displays a peculiar, highly organized architecture characterized by the presence of spheroidal structures – follicles – that are composed of a single layer of epithelial cells (thyroid follicular cells) surrounding a closed cavity (follicular lumen) filled with colloid, a concentrated solution of thyroglobulin (Tg). The follicle has been defined as the morphologic and functional unit of the thyroid. Notably, during intrauterine life, the onset of thyroid function (around 10 weeks in humans) coincides with the appearance of differentiated follicles. It is the follicular organization, together with the polarity of the follicular cells, that allows the several biochemical steps required for thyroid hormone biosynthesis: (1) secretion of a peculiar protein with iodinated amino acids in the follicular lumen as exocrine cells; (2) reabsorption of this peculiar protein, with hydrolysis of its iodinated amino acids; (3) release of iodothyronines into blood by endocrine secretion.

The follicle cell divides the follicular lumen (where hormone synthesis begins) and the blood stream, from where iodine has to be uploaded and where hormones will be released at the end of the process.

The surface of a polarized thyroid follicular cell is divided into two functionally distinct, but physically contiguous regions: an apical and a basolateral domain. Junctional complexes between cells separate these two domains and prevent the mixing of asymmetrically distributed proteins (**Figure 1**). The apical domain displays a differentiated tissue-specific organization characterized by the presence of apical microvilli and pseudopods, and by the localization of thyroperoxidase (TPO). Na^+/I^- symporter (NIS), epidermal growth factor, and thyroid-stimulating hormone (TSH) receptors are located in the basal domain. Thyroid hormone synthesis requires basal-to-apical transport of iodide and Tg. Conversely, hormone secretion is based on apical-to-basal transport of Tg and hormones; in addition, a bidirectional ion transport system controls follicular size.

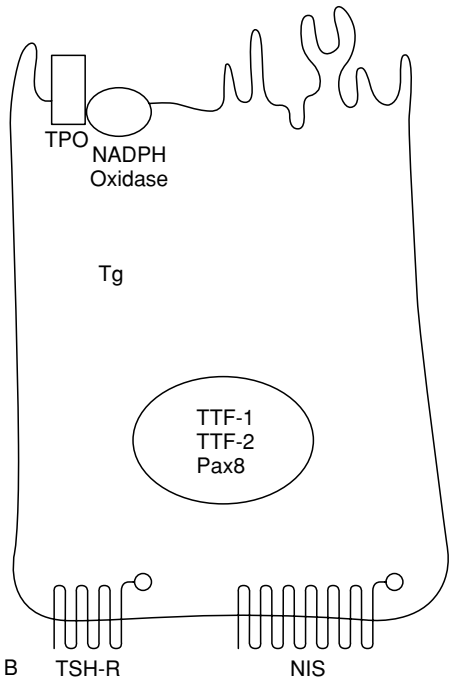

Figure 1 Schematic structural representation of thyroid follicular cell with its polarized architecture. TSH-R, thyroid-stimulating hormone receptor; NIS, Na/I symporter; Tg, thyroglobulin; TPO, thyroperoxidase; NADPH, reduced form of nicotinamide adenine dinucleotide phosphate; TTF-1.

The thyroid primordium of the human embryo is first visible at 20 embryonic days, as a midline enodermal thickening in the floor of the primitive pharynx. It migrates caudally to form a transient thyroglossal duct, and reaches its final position at 35 embryonic days. Tg is detectable at 60 embryonic days, and this step corresponds to the occurrence of fetal thyroid hormone biosynthesis. The early stages of folliculogenesis are independent of thyrotropin. At midgestation (18–20 weeks), the hypothalamopituitary–thyroid axis begins to develop, and thyrotropin is absolutely necessary for thyroid growth and function.

Biosynthesis of Thyroid Hormones

Figure 2 summarizes the process of thyroid hormone synthesis, which includes active concentration of

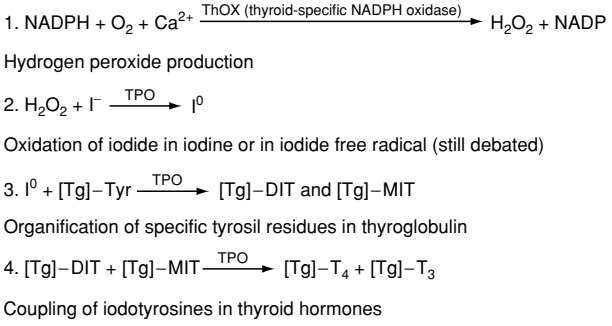

Figure 2 Oxidation of iodide and iodination of thyroglobulin (Tg) to produce thyroxine. DIT, diiodotyrosine; NADP(H), (reduced) and oxidized nicotinamide adenine dinucleotide phosphate; TPO, thyroperoxidase; ThOX, thyroid-specific NADP(H) oxidase; MIT, monoiodotyrosine.

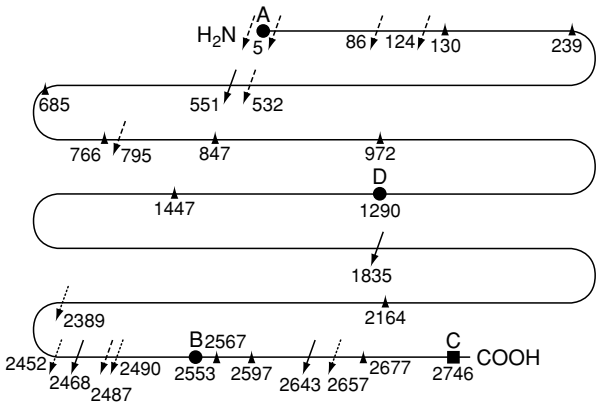

Figure 3 General linear picture of human thyroglobulin polypeptide chain. Numbers refer to cDNA sequence. A, B, D, sites of thyroxine formation; D, site of triiodothyronine formation. Cleavage sites of various cathepsins are symbolized by arrows. Reproduced from Dunn JT (2001) Biosynthesis and secretion of thyroid hormones. In: DeGroot LJ and Jameson JL (eds) *Endocrinology* WB Saunders, with permission.

iodine, incorporation of iodine in thyroglobulin, and breaking down thyroglobulin to release tetraiodothyronine (T_4) and triiodothyronine (T_3).

NIS An active transport mechanism allows thyroid cells to concentrate iodide (I^-) some 20- to 40-fold its level in extracellular space. This transport system is sensitive to other inhibiting anions (thiocyanate, perchlorate). It is found in other tissues (salivary glands, mammary gland). In the thyroid, it is restriced to the basolateral cell membrane. The key protein in the transport mechanism is the human Na/I symporter (hNIS). Its synthesis is stimulated by TSH. It is a 12-loop protein, of 65 kDa. It has homology with the Na/glucose transporter.

hTg Human Tg (hTg) is a 660-kDa protein. It contains 66 tyrosils, and half of them – in maximal conditions – are able to incorporate iodine in a covalent binding, resulting in mono- or diiodotyrosine (MIT, DIT) residues. Some iodothyrosine residues fuse, to form T_4 with its four iodine atoms and T_3 with its three iodine atoms. In normal physiological conditions, Tg contains five residues each of MIT and DIT, 2.5 of T_4 and 0.7 of T_3. These proportions depend on iodine supply, and on Tg accumulation in follicles. Its synthesis and release of thyroid hormones is under thyrotropin control. Iodine by itself, independently of thyrotropin, has an autoregulatory function on Tg synthesis and metabolism.

hTPO Human hTPO is a 103-kDa protein localized in the apical membrane of the follicular cell. It contains heme and, in the presence of hydrogen peroxide, it iodinates tyrosil residues of hTg present in the lumen of the follicles. The iodination process of proteins is not specific to thyroid cells: macrophages and leukocytes iodinate bacteria proteins through a similar process, involving myeloperoxidase in place of hTPO. However, the further step of coupling and thyroid hormone synthesis is quite specific to the thyroid. The same hTPO enzyme is responsible for the coupling of DIT and MIT in T_3 and T_4. It is proposed that the loops of hTg polypeptide make some DIT and MIT residues close, and the TPO enzyme and hydrogen peroxide interact to produce free radicals which form a diphenyl ether across the $OH^\bullet$ free radicals.

hThOX Reduced nicotinamide adenine dinucleotide phosphate (NADPH) oxidase is essential to produce hydrogen peroxide at the apical membrane of the follicular cells. It has recently been identified as a specific thyroid oxidase, with NADPH and calcium as substrates. It has a cytochrome moiety.

The sites of tyrosine iodination, as well as the sites of iodotyrosine coupling on Tg molecule are not stochastic, but are hierarchically defined. The number of iodinated tyrosine residues depends closely on the iodine supply and on thyrotropin stimulation. Figure 3 depicts the general scheme of preferential sites of iodination and coupling. The thyroid stores its hormones as colloid in the follicular lumen, still part of the Tg peptide structure. This process guarantees a stable storage of both iodine and iodinated thyroid hormones in variable-iodine-supply conditions. A normal human adult thyroid accumulates 10–20 mg iodine, which represents a storage for at least 3 months (required daily administration of iodine: $\pm$ 100 µg). This storage depot allows the thyroid to dole out thyroid hormones as the body needs it, and to mobilize it rapidly when called for. Such

mobilization is accomplished by bringing stored Tg back from the follicular lumen into the cell, passing it through endosomal and lysosomal digestive systems, and delivering free hormone to the basal membrane for secretion into the circulation. About 70% of Tg's iodine is in the form of the inactive precursors MIT and DIT, which the thyrocyte deiodinates and then recycles the iodine. The most important enzymes for Tg degradation are the lysosomal proteases. The best characterized are the aspartic endopeptidase cathepsin D and the cysteine endopeptidases cathepsin B and L. Peptides produced by the individual enzymes purified from human thyroids were isolated and the sites of cleavage identified by analogy to the cDNA sequence of hTg. Figure 3 summarizes, also, the major cleavage sites for each enzyme. Cathepsin L (dotted broken arrows in Figure 3) attacked near hydrophobic residues, particularly leucine, and its major cleavage sites were within the C-terminal 400 residues of Tg. Cathepsin D (solid arrow in Figure 3) preferentially cleaved peptide bonds between hydrophobic residues, particularly aromatic ones, similar to its activity in other tissues. Cathepsin B (dashed broken arrows) had its principal cleavage sites in both the N-terminal and C-terminal regions. The presence of major hormonogenic sites at Tg's extreme N- and C-termini may favor selective release of thyroid hormone during proteolysis. The endopeptidases cleave Tg into large peptides that must then be further degraded by exopeptidases. The several exopeptidases identified in the thyroid include dipeptidylpeptidases I and II, lysosomal dipeptidase I, and the carbaxyl exopeptidase *N*-acetyl-L-pbenolalanyl-L-tyrosine hydrolase. Also, cathepsin B has exopeptidase activity, in addition to its role as an endopeptidase. The combination of cathepsin B and lysosomal dipeptidase I is sufficient to release T_4 from its most important site at residue 5.

The thyroid stimulated by TSH shows prompt and vigorous engulfment of material from the lumen to form colloid droplets, which are then internalized into the thyrocyte and passed through its lysosomal-degradative pathway. However, under more usual physiologic conditions, colloid resorption takes place by micropinocytosis, and endocytotic vesicles are formed and pass successively through an endosomal compartment and then into lysosomes. The initial event in micropinocytosis appears to involve the entrance of Tg into coated pits, probably without the intervention of receptors; the amount processed depends on the availability of endocytotic vesicles and the amount of Tg in the lumen.

Tg's structure, specifically, the completion of its carbohydrate side chains, is important for its trafficking at the apical membrane. Immature Tg molecules are deficient in both iodine and glycosylation, so they have exposed *N*-acetylglucosamine residues. These residues are recognized by membrane receptors that apparently direct these immature molecules back to the Golgi for glycosylation and subsequent iodination. These observations suggest that Tg molecules are sorted in the endosomes, perhaps by iodine content, with immature molecules being recycled and mature ones readied for degradation and hormone release.

Some proteolytic processing may take place in endosomes before Tg enters lysosomes. Lysosomal enzymes are also present in endosomes, where the acidic pH would favor their activity. Some proteolytic processing may occur even before endocytosis. For example, cathepsin B activity has been detected at the cell surface of cultured thyrocytes. In the same experimental model, activation of cysteine proteases permitted T_4 release at extralysosomal sites, as well as in the lysosomes, although T_3 release appeared to be restricted to lysosomes.

Specific cleavages in the peptide chain of Tg are also associated with its iodination. Such breaks occurred in the presence of protease inhibitors, thus suggesting that proteolytic enzymes were not involved. However, other specific proteolytic cleavages may also occur as early events in Tg degradation. Limited digestion with trypsin has identified several susceptible regions in the Tg molecule, most notably a region after residue 500 and another one around residue 1800.

T_4 emerges from the lysosomal-degradative pathway, probably as the free hormone or perhaps in small peptides. Most is secreted into the circulation. A specific transport mechanism across the basal membrane has not been identified. While still in the thyroid, some T_4 is converted to T_3 by type 1 selenium-dependent 5′-iodothyronine deiodinase. This enzyme is the same as that in other tissues, such as the liver and kidney, produces T_3, the active form of the hormone, from circulating T_4.

About 70% of Tg's iodine is in MIT and DIT. Although small amounts of MIT and DIT are secreted into the circulation, most is deiodinated within the thyroid and its iodide is returned to the general pool for recycling. This mechanism is important for iodine conservation, as shown by the functional iodine deficiency of patients who have the rare defect that blocks this step. The iodotyrosine deiodinase responsible for this activity is an NADPH-dependent flavoprotein (it is not a selenium-containing enzyme, and it should not be confound with iodothyronine deiodinase). Its partial purification showed a molecular weight of about 42 000, and it consisted of two possibly identical subunits.

Circulating Thyroid Hormone and Free Thyroid Hormone Metabolism

Iodothyronines produced by the thyroid are highly hydrophobic, and need to be bound to specific proteins for transport through the general blood circulation to their target tissues. In blood, thyroid hormones are not covalently bound to specific thyroid hormone-binding proteins (thyroid hormone-binding globulin (TBG) and transthyretin (TTR), previously known as prealbumin). They are also transported aspecifically by albumin. Table 1 shows the properties of the major human thyroid hormone-binding proteins.

Free fraction of thyroid hormones describes the percentage of total hormone that is unbound. About one molecule of T_4 in 4000 and one molecule of T_3 in 400 is free; the remaining molecules are bound to the specific and aspecific proteins. By contrast to corticoid-binding proteins, none of the thyroid hormone-binding proteins has a role in the delivery of thyroid hormones to target tissues. Only free hormones are delivered to target tissues: so, clinical thyroid status is closely dependent on free thyroid hormone concentration, and the general equilibrium is as follows:

$$\text{Free } T_4 + \text{TBG} \Leftrightarrow [\text{TBG}\ldots T_4]$$

The normal free T_4 concentration and TBG concentration are such that about half of TBG sites available to thyroid hormone binding is occupied by T_4. This involves that within a certain range; free hormone concentration is linearly associated with total hormone concentration. So, TBG stabilizes the tissue distribution of T_4.

Modern immunoassays measure the serum free T_4 and free T_3 concentrations, which are directly related to clinical thyroid status.

Table 2 presents the kinetic parameters of circulating total and free thyroid hormones in a 70-kg adult human.

Table 1 Properties of the major human thyroid hormone-binding proteins

Property	*Thyroxine-binding globulin*	*Transthyretin*	*Albumin*
Molecular weight (kDa)	54	55	66
Peptide structure	Glycoprotein	Peptide tetramer	Single chain
Concentration (mol l^{-1})	3×10^{-7}	2×10^{-6}	6×10^{-4}
Half-life (days)	5	2	15
% of tetraiodothyronine carried	75	10–15	10–15

Table 2 Kinetic properties of iodothyronines in humans

Property	T_4	T_3	*Reverse* T_3
Total serum concentration ($\mu g\,dl^{-1}$)	8.1	0.11	0.012
Free serum concentration ($ng\,dl^{-1}$)	1.2	0.29	0.04
Distribution volume (l)	10	35	90
Metabolic clearance rate (per day 70 kg s^{-1})	1.2	25	111
Serum half-life (days)	7	1	0.2
Production ($\mu g\,day^{-1}$ 70 kg s^{-1})	100	31	39
Relative metabolic potency	0.3	1.0	0.0
Nuclear T_3 receptor binding *in vitro*	10^{-9}	10^{-10}	No data

T_4, tetraiodothyronine; T_3, triiodothyronine.

Within some tissues (mainly, liver and kidney), circulating T_4 is deiodinated in circulating T_3 by a selenium-dependent type 1 iodothyronine deiodinase. In normal iodine supply conditions, half of circulating T_3 derives directly from thyroid (breakdown of Tg), and half from liver conversion of T_4 to T_3. The same deiodinase is also able to convert T_4 to reverse T_3 (metabolically inactive). In some physiological conditions such as the neonatal period (cord blood) or during illness, the preferential liver conversion of T_4 to reverse T_3 explains the paradoxical serum profile of low T_3 syndrome without hypothyroidism. The exact mechanism of preferential conversion of T_4 to reverse T_3 in place of T_3 is still unexplained.

Thyroid Hormone Action at the Cellular Level

Free thyroid hormones T_4 and T_3 diffuse through the cell membrane of their target tissues. Up to now, no specific membrane receptors have been identified, and it is considered that this delivery to cells is through a passive diffusion mechanism.

Within the cell, iodothyronine deiodinases act to convert intracellular T_4 to intracellular T_3 or reverse T_3 (Figure 4).

The differential intracellular conversion of T_4 to T_3 in peripheral tissues (thyroid; muscle, including the heart; liver; kidney; adipose tissue, etc.) and in central tissues (central nervous system, including pituitary gland) is explained by the divergent distribution of selenium-containing iodothyronine deiodinases. Type 1 is present mainly in peripheral tissues, while type 2 is present mainly in the central nervous system. Type 2 deiodinase has a much lower affinity constant for T_4 than type 1 deiodinase (1 nmol l^{-1} versus 23 µmol l^{-1}). This explains why, within the central nervous system, in normal physiological conditions, 80% of

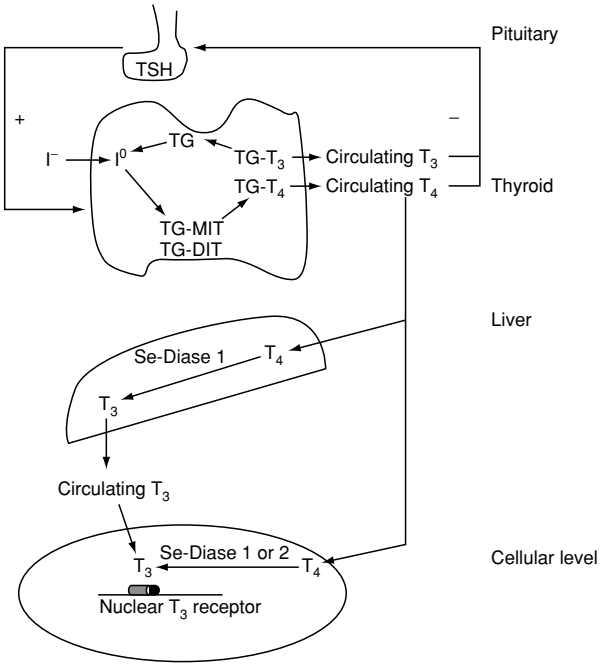

Figure 4 Some metabolic steps in iodine metabolism. Pituitary thyrotropin (TSH) stimulates thyroid I^- uptake, and enzymatic organification of iodine (I^0) by thyroperoxidase (TPO) in a mono- and diiodotyrosine (MIT and DIT), which are further coupled by the same enzyme thyroperoxidase (TPO) in tri- and tetraiodothyronine (T_3 and T_4). Part of T_4 is further deiodinated to T_3 in the liver by selenium-containing enzymes selenium-dependent iodothyronine deiodinases 1 and 2 (Se-Diase 1 and 2). At the cellular level, the active hormone is T_3, which binds to a specific nuclear T_3 receptor. T_3 comes directly from circulating serum T_3 or from the intracellular conversion of T_4 to T_3.

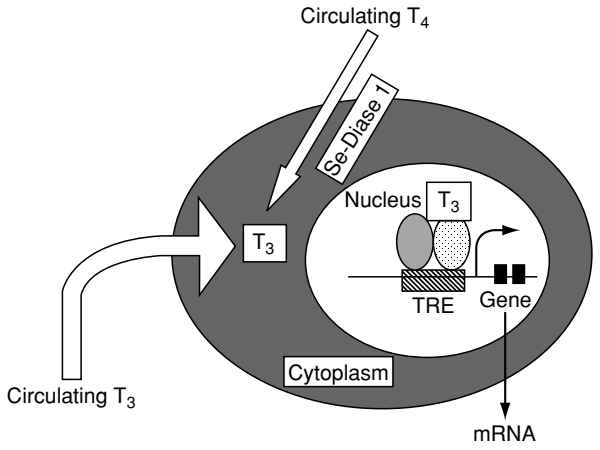

Figure 5 Simplified model of thyroid hormone action on nuclear thyroid hormone receptor. T_4, tetraiiodothyronine; T_3, triiodothyronine; Se-Diase 1; selenium-dependent iodothyronine deiodinase 1; TRE, T_3 response element.

intracellular T_3 derives from local deiodination of T_4, while, in peripheral tissues, this percentage is less than 50%. Moreover, in case of incipient biochemical hypothyroidism, when serum T_4 is already decreased but serum T_3 is still normal, there is increased type 2 deiodinase activity and increased thyroid type 1 deiodinase activity, while there is decreased type 1 deiodinase activity in other peripheral tissues (liver, mucle, heart, kidney). These mechanisms involve central hypothyroidism reflected by increased serum TSH, while peripheral tissues are still euthyroid (there is no evidence of hypothyroidism in peripheral tissues). The advantage of such a complex feedback mechanism is to stimulate maximally the thyroid by TSH in case of incipient biochemical hypothyroidism.

At the nuclear level (Figure 5), thyroid hormone acts through a T_3 response element (TRE). Most thyroid hormone activity described up to now is mediated via these high-affinity nuclear receptors – even if some extranuclear actions of thyroid hormones are also described. The principal differentiation effects of thyroid hormones (brain development, bone maturation) and the metabolic effects of thyroid hormones (uncoupling protein, fatty acid metabolism, malic enzyme activity) are mediated via these receptors. The binding of thyroid hormone T_3 to its receptor involves an induction (positively regulated genes) or a repression (negatively regulated genes) of specific messenger RNAs and the translation of these mRNAs in specific proteins. Examples of positively regulated genes include β-adrenergic receptor, fatty acid synthetase, malic enzyme, growth hormone, uncoupling protein, and type 1 iodothyronine deiodinase (except in thyroid). Examples of negatively regulated genes are TSH, prolactin, type 2 iodothyronine deiodinase, and type 1 iodothyronine deiodinase in thyroid.

The translated protein may also act indirectly, for example, thyroid hormones influence growth via insulin-like growth factor 1 (IGF-1). Nevertheless, this action is not present in hypophysectomized rats. So, the thyroid hormone action on liver IGF-1 is indirect:

$$T_3 \Uparrow \rightarrow \text{Pituitary and circulating growth hormone} \Uparrow \rightarrow \text{Liver and circulating IGF-1} \Uparrow$$

The TRE belongs to the family of *erb*-related receptors (including receptors for estrogen, progesterone, glucocorticoid, mineralocorticoid, androgen, vitamin D, retinoic acid). The receptor binds to the DNA via zinc-fingers. It forms a dimer which is stabilized when bound to T_3 (homodimer) or to 9-*cis*-retinoic acid (heterodimer).

A congenital mutation in TRE is a – rare – cause of resistance to thyroid hormones (200 cases have been

described in the world) characterized by elevated serum free T_4 and elevated serum TSH. Individuals with this syndrome do not in general exhibit signs of hypothyroidism, presumably because the resistant state is compensated for by increased levels of thyroid hormones. In some cases, the syndrome may include attention-deficit hyperactivity disorder, reduced intellectual development, delayed skeletal maturation, tachycardia, and deafness.

Calcitonin

Thyroid also synthesizes calcitonin. The parafollicular or C cells are distributed among follicular cells. These cells derive from the neural crest (by opposition to the endodermal origin of follicular cells). In human adults, these cells represent less than 1% of the follicular cells. Parafollicular cells are also secreting other hormones like gastrin and somatostatin. During ontogeny, around embryonic day 24 in humans, C cells of neural crest origin colonize a vestigial organ – the ultimobranchal organ, derived from the fourth pharyngeal pouch. This ultimobranchial organ, with its embryonic C cells, migrates toward the thyroid, where its cells diffuse among the follicules.

Calcitonin is derived from a 136 amino acid precursor peptide which, after cleavage, leaves an active protein of 32 amino acids.

Calcitonin inhibits bone resorption, peculiarly during growth. For example, it decreases serum calcium after administration of pharmacologic doses. In adults, however, the effect of calcitonin on calcium homeostasis is marginal. Moreover, children born without thyroid (thyroid aplasia) and adequately treated with substitutive T_4 hormonetherapy do not present alterations of calcium homeostasis nor osteoporosis. So, if there is no doubt that calcitonin exerts an effect at pharmacological doses, its role in normal human physiology, even during growth, remains largely unknown.

See also: **Iodine**: Properties and Determination; Physiology

Further Reading

Caillou B, Dupuy C, Lacroix L *et al*. (2001) Expression of reduced nicotinamide adenine dinucleotide phosphate oxidase (ThoX, LNOX, Duox) genes and proteins in human thyroid tissues. *Journal of Clinical and Endocrinological Metabolism* 86: 3351–3358.

Gut Hormones

L M Morgan, University of Surrey, Guildford, UK

Introduction

The gut is the body's largest endocrine organ, both in terms of the number of endocrine cells and number of hormones. The very word 'hormone' was first used by Bayliss and Starling in 1902 to describe the action of the gastrointestinal (GI) hormone secretin. However, the firm establishment of the gut's endocrine status was not achieved until 1970 when the primary structure of the three classical endocrine gut hormones secretin, gastrin, and cholecystokinin (CCK) was finally determined. Since the 1970s there has been a rapid explosion in our knowledge of the variety and complexity of bioactive peptides found in the gut, made possible by new investigative technologies, especially those of molecular biology. The gut has been shown to contain not only endocrine hormones but a huge variety of neuropeptides, and many gut peptides which were formerly believed to be classical hormones have since been shown to be widespread neurotransmitters. This has led to a much broader understanding of the diversity of physiological roles for gut peptides. These roles extend far beyond the absorption and digestion of food in the gut itself to the coordination and regulation of many bodily functions outside the gut. This article reviews the major hormones of the GI tract, their characteristics, physiological effects, and the regulation of their secretion by nutrients and other stimuli. It focuses on those aspects of gut hormone physiology that are concerned with the digestion and metabolism of nutrients and the role of gut hormones in the regulation of appetite and food intake itself.

Distribution and Synthesis

GI hormones were first identified within gut endocrine cells, which are scattered throughout the mucosal epithelium lining the gastric glands, intestinal crypts, and villi. Typically these cells are pyramidal and lie on the basement membrane. Some cells extend to the lumen of the gut by means of tufts of microvilli and some connect to one another via cytoplasmic processes. In addition, the neuronal cells of the gut which form the submucosal and myenteric plexi also produce a wide variety of gut peptides. The peptide-secreting cells share many features, and they were probably derived from a common precursor cell that differentiated to

produce each endocrine or neurocrine cell line, a hypothesis which accounts for the widespread distribution of gut peptides outside the gut itself, in neural tissue and the central nervous system. With the exception of secretin, gastric inhibitory polypeptide (GIP), and peptide YY (PYY), most of the gut-regulatory peptides are indeed also present in brain tissue. Many of the gut peptides thus act as paracrine or neurocrine transmitters rather than as endocrine hormones. This enables many functions of the digestive tract, e.g., secretory or motor functions, to be regulated by interaction between the endocrine and nervous systems. In addition, gut neuropeptides are involved in many centrally mediated functions, such as the perception of appetite and regulation of food intake.

The GI hormones are low-molecular-weight single-chain polypeptides, of chain length generally fewer than 50 amino acids. Most are synthesized by a prohormone mechanism in a manner similar to insulin. However, unlike insulin, GI hormone genes often express several different bioactive peptides. After the initial translation product of the gene is formed, a series of modifications, collectively called posttranslational processing, can result in a number of peptide products of varying chain lengths and biological activities.

Characterization

The major gut hormones of known structure are listed in Table 1, together with the primary attributes by which they have been characterized. Each year

Table 1 Distribution and biological actions of the major gastrointestinal (GI) hormones

Hormone/peptide	*Site of distribution in GI tract*	*Major biological effects*
Secretin family		
Secretin	Endocrine cells, duodenum, and jejunum	Stimulation of pancreatic exocrine secretion
Gastric-inhibitory polypeptide	Endocrine cells, duodenum, and jejunum	Stimulation of insulin secretion
Glucagon-like peptides (GLP-1, GLP-2, glicentin, oxyntomodulin)	Endocrine cells throughout GI tract. Highest concentrations in terminal ileum and colon	Stimulation of insulin secretion and satiety. Inhibition of gastric acid secretion. Trophic effect on gut mucosa
Vasoactive intestinal peptide	Nerve fibers throughout GI tract	Modulation of gut motility. Stimulation of intestinal secretion. Regulation of blood flow
Gastrin family		
Gastrin	Endocrine cells, antrum, and duodenum	Stimulation of gastric acid secretion. Trophic actions on gut mucosa
Cholecystokinin	Endocrine cells, duodenum, and jejunum Nerve fibers, pancreas, and colon	Stimulation of pancreatic secretion, gallbladder contraction, satiety. Inhibition of gastric emptying
Pancreatic polypeptide family		
Pancreatic polypeptide	Endocrine cells, pancreas	Inhibition of pancreatic exocrine, gastric acid secretion
Peptide YY	Endocrine cells throughout GI tract. Highest concentrations in terminal ileum and colon	Inhibition of gastric acid secretion, gastric emptying, intestinal motility
Neuropeptide Y	Nerve fibers throughout GI tract. Pancreas	Inhibition of intestinal and insulin secretion
Tachykinin family		
Substance P	Nerve fibers throughout GI tract	Smooth-muscle contraction regulation of blood flow
Neurokinin A (neuromedin L)	Nerve fibers throughout GI tract	
Neurokinin B (neuromedin K)	Nerve fibers throughout GI tract	
Bombesin-like peptides		
Gastrin-releasing peptide	Nerve fibers throughout GI tract	Smooth muscle contraction. Trophic actions on gut mucosa. Stimulation of gastrin
Neuromedin B	Nerve fibers throughout GI tract	
Somatostatin	Endocrine cells, nerve fibers throughout GI tract, especially stomach, colon, pancreas	Inhibition of many gut peptides, especially gastrin and insulin; gastric, pancreatic, and biliary secretion
Neurotensin	Nerve and endocrine cells throughout GI tract, especially ileum and jejunum	Inhibition of gastric motility and secretion Stimulation of pancreatic exocrine secretion
Motilin	Endocrine cells throughout small intestine, especially duodenum, jejunum	Regulation of interdigestive gut motility. Inhibition of gastric motor activity
Galanin	Nerve fibers throughout GI tract, pancreas	Inhibition of gastric emptying, gut motility, insulin, and other gut peptide secretion

several new bioactive peptides are added to the growing list of GI hormones, and currently the number of regulatory peptides isolated from the GI tract is well in excess of 50. Many of the GI hormones show a large degree of structural homology, and can be divided up into distinct families on the basis of similarities in their amino acid sequence. There is often considerable overlap in the biological activities of hormones in the same group, for example most members of the secretion family have the ability to potentiate insulin secretion. Gut peptides which do not fit into family groups are termed 'orphan' peptides. The ability of GI hormone genes to express multiple phenotypes has led to great molecular heterogeneity and complexity of their hormone products. For example, secretin was originally believed only to exist as a 27-amino-acid peptide. However, by 1990, three additional secretins with full bioactivity had been identified, two of similar size to secretin-27 and formed by variable trimming of secretin-27's carboxyl terminal, and one large bioactive molecule, secretin-71, produced by trimming out the mid-sequence of the prohormone, preprosecretin. Prohormones can also contain more than one active sequence. Within the secretin family, a group of closely related glucagon-like peptides is derived from a single preproglucagon gene. Posttranslational processing is cell-specific; thus in the pancreas glucagon is the hormonal gene product, whilst in the gut, glicentin and oxyntomodulin (extended versions of glucagon) are produced, together with two additional bioactive peptides, glucagon-like peptides 1 and 2.

Some of the neurally distributed gut peptides were first isolated from brain tissue. Somatostatin was originally isolated and characterized from the hypothalamus before its discovery in gastric mucosal cells and the pancreas. The enkephalins, members of the tachykinin family, were first isolated from brain tissue, where they are often detectable in higher concentrations than in tissues within the GI tract.

The gut hormones occurred very early on in evolutionary terms and various members of the gut hormone families are extensively distributed throughout the animal kingdom. Some of the mammalian gut hormones have amphibian counterparts. The bombesin-like peptides, of which gastrin-releasing peptide (GRP) is the chief mammalian example, were originally isolated in amphibian skin. In addition, neurotensin, CCK and the tachykinin, substance P, have their amphibian skin counterparts – xenopsin, caerulein, and physalaemin respectively. Substance P and CCK-like peptides have been found in invertebrate neurons and a CCK-like peptide has even been isolated in hydra, one of the simplest present-day multicellular organisms.

Physiological Effects

Gut peptides can be separated into three functional categories: endocrine, paracrine, and neurocrine, reflecting their cellular location and mode of delivery to target cells (Table 2). Endocrine peptides reach their targets via the general circulation, paracrine by diffusion or via the local circulation, and neural peptides by synaptic transmission.

Essential first steps in identifying the physiological function of a gut peptide are to determine its cellular location, pharmacological properties, and the regulation of its secretion. This is hindered by two features of the neurendocrine system of the gut. The first is the wide dispersion and comingling of peptide-containing cells which make it difficult to excise one endocrine or neural system without interfering with another. The second feature, characteristic of neurons, is the existence of different peptides within the cell along with nonpeptide neurotransmitters, leading to the possibilities of very complex interactions. Nevertheless, the physiological roles of gut hormones in several major functions of the GI tract and the interaction between endocrine and nervous systems have been elucidated and are described below.

Control of Gastric, Pancreatic, and Biliary Secretion

There are three major endogenous stimulators of gastric acid secretion: gastrin, acetylcholine, and histamine; it is probable that all three are interdependent under physiological conditions in humans. The major action of gastrin is to release acid from parietal cells in the fundus and body of the stomach, although it also exerts indirect effects on acid secretion via the stimulation of histamine release from ECL cells. Regulation of gastrin secretion is under the control of both humoral and local factors and also the

Table 2 Functional categories of gut hormones

Hormone	*Endocrine*	*Paracrine*	*Neurocrine*
Secretin	+		
Gastric-inhibitory peptide	+		
Glucagon-like peptides	+		
Vascular-inhibitory peptide			+
Gastrin	+		
Cholecystokinin	+		+
Pancreatic polypeptide family	+		+
Tachykinin family			+
Bombesin-like peptides			+
Somatostatin	+	+	+
Neurotensin	+		
Motilin	+	+	
Galanin			+

autonomic nervous system, illustrating the complex interaction between endocrine, neurocrine, and paracrine systems. Gastrin release is stimulated directly by the ingestion of food, especially protein, and by the neuropeptide GRP, which is itself released by cholinergic stimulation. Gastrin secretion is inhibited by acid, and by somatostatin, acting in a paracrine manner. The ingestion of fat, once the fat has passed though the stomach and into the small intestine, inhibits gastric acid secretion via the stimulation of GI hormones GIP, glucagon-like polypeptide-1 (GLP-1), and CCK. It seems likely that their modes of action vary and include both direct inhibitory effects on parietal cells, as in the case of CCK, and indirect effects on acid secretion via gastrin and somatostatin.

Pancreatic exocrine secretion contains two major components. The first is an aqueous solution containing a high concentration of bicarbonate whose function is to neutralize acid entering the duodenum from the stomach. The second is a solution of enzymes which aid the absorption and digestion of carbohydrate, fat, and protein. Secretin and cholecystokinin are traditionally thought to be responsible for pancreatic exocrine secretion, but other gut peptides, notably vasoactive intestinal peptide (VIP), neurotensin, and pancreatic polypeptide, may also influence secretion. Secretin is responsible for most of the water and bicarbonate secretion. CCK acts synergistically with secretin to augment pancreatic bicarbonate output, and interactions between neurotensin and secretin also appear to be important. During digestion, pancreatic exocrine secretion is partly controlled by efferent vagal fibers, mediated by VIP. Pancreatic polypeptide inhibits pancreatic secretion and may be involved with the damping-down of pancreatic exocrine secretion following a meal.

CCK has been considered a major hormone controlling gallbladder emptying, and release of pancreatic enzymes. However, the lack of reliable assays to measure circulating levels of CCK has caused difficulty in ascribing a physiological, as opposed to pharmacological, role for the hormone. Recent studies using CCK-A (the predominant CCK receptor found in the pancreas and gallbladder) receptor antagonists have confirmed the role of CCK as the primary stimulator of gallbladder contraction, although the effects of CCK receptor antagonists on pancreatic secretion are less clear, probably reflecting multiple pathways of control for pancreatic exocrine secretion.

Gastric Emptying and Gut Motility

Between the striated muscles of the esophagus and the external anal sphincter, which are dependent upon efferent nerves for proper functioning, lie the smooth muscles which form the bulk of the GI tract. They are capable of executing their function without extrinsic innervation, and sympathetic and parasympathetic nerves act only in a modulatory capacity. The intrinsic nerves of the gut form two major networks or plexi: the myenteric plexus which lies between the longditudinal and circular muscles of the gut, and the submucous plexus, which lies between the mucosa and circular muscle layer. Control of gut motility involves many neuronally located gut peptides. Receptors for many peptides have been located on gut smooth muscle cells, including motilin, the opioid peptides, the tachykinins, VIP, and neuropeptide Y.

Regulation of the rate of gastric emptying depends on many factors, including the physical state and size of the meal and its macronutrient content. Three endocrine gut hormones have been implicated in the control of gastric emptying. Secretion of CCK is responsible for the delay in gastric emptying when fat is added to a meal; the presence of fat in the duodenum is a powerful stimulus for CCK release. GLP-1, secreted in response to the presence of all three macronutrients in the upper small intestine, also delays gastric emptying and regulates the influx of nutrients into the upper part of the gut. Peptide YY, which is colocalized and often cosecreted with GLP-1, inhibits gastric emptying and also small intestinal motility. It has been termed the 'ileal brake' and its actions following a meal are thought to enable sufficient time to occur for the meal's absorption. Following the ingestion of food, waves of alternate contraction and relaxation descend from the gut, facilitating the passage of food and aiding its absorption. VIP causes smooth muscle relaxation and its distribution is consistent with the role of a transmitter in the descending inhibitory nerves. Opioid peptides increase the contractile activity of the gut and their net effect is to slow the transit of food. The tachykinins also have potent exitatory and spasmogenic effects on smooth muscle and appear to be involved in regulation of the peristaltic reflex.

When fasting is prolonged, interdigestive myoelectric complexes occur. These are waves of contraction beginning in the stomach which move slowly down the small intestine, whose function is to clear out the intestine between feeds and keep the intraluminal bacterial population low. Motilin is primarily involved in the regulation of this interdigestive activity.

Potentiation of Insulin Secretion and other Metabolic Effects

One of the more important metabolic actions of gut hormones is their ability to modulate insulin secretion. Orally administered glucose leads to far higher circulating insulin levels than the same amount of

glucose administered intravenously. This ability of the gut to stimulate insulin secretion has been termed the enteroinsular axis. Gut peptides potentiate insulin secretin via both neural and endocrine pathways but, in humans, the endocrine role of the gut hormones (collectively termed incretins) is considered to be quantitatively the more important. Although most members of the secretion family of gut hormones have the ability to stimulate insulin in pharmacological doses, only GIP and GLP-1 are both secreted in response to glucose ingestion and stimulate insulin secretion at physiological levels. Together they account for the entire endocrine arm of the enteroinsular axis.

GIP and GLP-1 are both potent stimulators of glucose-induced insulin secretion. On a molar basis, GLP-1 is the stronger stimulator, but as it circulates at lower concentrations than GIP, the contribution of the two hormones to the enteroinsular axis are approximately equivalent. These hormones only stimulate insulin secretion in the presence of hyperglycemia. This can be looked upon as providing a safeguard against inappropriate insulin secretion, as both hormones are also stimulated by oral fat, when insulin secretion would be undesirable unless its hypoglycemic properties were counteracted by simultaneous ingestion of carbohydrate.

The neural arm of the enteroinsular axis is more poorly defined. Both CCK- and VIP-containing neurons have been implicated as neurotransmitters stimulating insulin secretion. The smaller forms of CCK (CCK-4 and CCK-8) also appear to be biologically active, although their precise role in insulin stimulation is not yet fully known.

Consistent with the anabolic role that gut hormones play in carbohydrate metabolism, by stimulating insulin secretion, is a parallel anabolic role in lipid metabolism. GIP has been shown to have an anabolic effect on lipid metabolism, stimulating lipoprotein lipase and enhancing postprandial triglyceride clearance from the plasma. GIP and GLP-1 have also been shown to stimulate *de novo* fatty acid synthesis in adipose tissue. These two GI hormones, which potentiate the secretion of insulin, also exhibit insulin-like actions systemically and provide a direct link between the absorption and subsequent metabolic fate of ingested nutrients.

Appetite Regulation

The regulation of hunger and satiety is a complex process, involving the interaction of both physiological and psychological factors. There is good experimental evidence for the involvement of several of the gut hormones in the short-term regulation of appetite, although the precise mechanisms whereby they exert their effects in human appetite control are less clear. CCK and GLP-1 have been found to mediate satiety in both animal and human studies. The inhibitory effects of CCK on food intake are stimulated by fat ingestion, and appear to be mediated by a local gastrointestinal action, although CCK is partially dependent upon an intact vagal nerve supply in order to exert its satiety effects. Peripheral administration of GLP-1 has been shown to suppress appetite in most, if not all human studies; circulating GLP-1 can theoretically pass the blood–brain barrier to act centrally. Neuropeptide Y is distributed in the central nervous system as well as within the gut. It acts in the brain upon appetite centers in the hypothalamus to stimulate food intake, and it may also be implicated in the regulation of leptin, the peptide secreted from adipose tissue which is involved with the longer-term mediation of satiety.

Regulation of Secretion

Ingestion and absorption of food are the major stimuli to gut hormone secretion, especially the endocrine gut hormones. The major nutrient stimuli for these hormones are listed in Table 3. The type of nutrient that stimulates the secretion of an individual gut hormone usually relates to its physiological role. GIP and GLP-1, for example, the potentiators of insulin secretion, are stimulated by carbohydrate, which in itself stimulates insulin secretion by raising circulating glucose levels. Fat ingestion is a major stimulus to secretin and CCK whose primary function is concerned with stimulating pancreatic and biliary secretion, thus aiding the digestion of fat. GIP secretion is also stimulated by fat absorption, consistent with the role of GIP in aiding the clearance of circulating lipids after a meal, and aiding their uptake into adipose tissue. The secretory response of a gut hormone to nutrients is a very precise one. For example, carbohydrate-induced GIP secretion is stimulated specifically by the absorption of actively transported sugars such as glucose and galactose. The mere presence of nutrients in the gut lumen is insufficient to stimulate GIP secretion, and sugars such as fructose, whose absorption does not require active transport mechanism, do not stimulate GIP secretion. In contrast, GLP-1 secretion is not dependent upon the absorption of nutrients, a property which is consistent with its role as a regulator of gastric emptying and the smooth delivery of nutrients to the upper part of the gut. Chronic dietary changes can also affect gut hormone secretion; a high-fat diet, for example, has been shown to increase GIP secretion in response to both carbohydrate and fat ingestion.

Table 3 Nutritional and other regulation of endocrine gastraintestinal hormone secretion

Hormone/peptide	*Major nutritional stimulants*	*Other modulators of secretion*
Secretin	Fat. Exit of meal from stomach causing duodenal acidification	Stimulation: bile salts Inhibition: somatostatin
Gastric-inhibitory peptide (GIP)	Fat, carbohydrate. Absorption from upper small intestine	Inhibition: insulin
Glucagon-like peptides	Carbohydrate, fat. Presence of food in upper small intestine	Stimulation: gastrin-releasing peptide (GRP) Inhibition: somatostatin
Gastrin	Amino acids and protein in stomach lumen; meal-induced gastric distension	Stimulation: GRP, acetylcholine, norepinephrine (noradrenaline) Inhibition: acid (pH < 3), GIP, secretin, GLP-1, somatostatin
Chdecystokinin	Fat (hydrolysis of triglyceride), phenylalanine and tryptophan	Stimulation: GRP Inhibition: trypsin, bile salts, somatostatin
Pancreatic polypeptide	Ingestion of meal, especially carbohydrate	Stimulation: GRP, motilin, vagal stimulation Inhibition: somatostatin
Peptide YY	Ingestion of meal, especially fat	Inhibition: GLP-1
Somatostatin	Ingestion of meal, especially fat and protein. Meal-induced lowering of gastric pH	Stimulation: GIP, CCK, secretin Inhibition: opioid peptides
Neurotensin	Fat ingestion	Stimulation: GRP Inhibition: somatostatin
Motilin	Stimulated by both oral and intravenous fat (inhibited by glucose, amino acids)	Stimulation: duodenal acidification Inhibition: secretin, GRP, somatostatin

GLP-1, glucagon-like peptide 1.

Interactions between the various gut hormones play an important part in the regulation of gut hormone secretion, particularly in an inhibitory capacity (Table 3). Thus somatostatin is a physiological paracrine inhibitor of gastrin secretion and probably plays an endocrine role in the regulation of other gut hormones, e.g., secretin and neurotensin. GRP is so named because of its role in the stimulation of gastric secretion. GRP infusion will also cause increased plasma concentrations of many other GI hormones, although the physiological relevance of this is unclear.

Less is known about the regulation of gut neuropeptide secretion. These peptides are synthesized by intramural neurons and released from nerve terminals by physiologically appropriate stimuli (e.g., chemical or mechanically induced reflexes). Vagal stimulation has been implicated in the release of several gut neurotransmitter peptides, e.g., VIP and GRP, but its regulatory role has yet to be fully defined.

See also: **Diabetes Mellitus**: Etiology; **Hormones**: Pancreatic Hormones

Further Reading

Holst JJ, Fahenkrug J, Stadil F and Rehfeld JF (1996) Gastrointestinal endocrinology. *Scandinavian Journal of Gastroenterology* 216: 27–38.

Morgan LM (1998) The role of gastrointestinal hormones in carbohydrate and lipid metabolism and homeostasis: effects of gastric inhibitory polypeptide and glucagon-like peptide-1. *Biochemical Society Transactions* 26: 216–222.

Rehfeld JF (1998) The new biology of the gastrointestinal hormones. *Physiological Reviews* 78: 1087–1103.

Schultz H, Makhlouf GM and Rauner BB (eds) (1989) The gastrointestinal system. Neural and endocrine biology. In: *Handbook of Physiology*, vol. II, Section 6. Bethesda: American Physiological Society.

Wilding JPH, Ghatei MA and Bloom SR (1995) Hormones of the gastrointestinal tract. In: DeGroot LJ (ed.) *Endocrinology*, pp. 2870–2894. Philadelpia, London: Saunders.

Pancreatic Hormones

N H McClenaghan and P R Flatt, University of Ulster, Coleraine, UK

Pancreatic Secretions

The pancreas is a secretory gland playing a key role in the metabolism of carbohydrates, proteins, and lipids. Cells of the exocrine pancreas (acini) secrete pancreatic juice into the duodenum, which contains enzymes for the digestion of major food

components. The endocrine pancreas synthesizes and secretes peptide hormones into the blood, thereby controlling nutrient homeostasis by determining the metabolic fate of absorbed glucose, amino acids, and fatty acids, and their mobilization in times of need from body stores. The following sections focus specifically on the endocrine pancreas, and the synthesis, secretion, and action of pancreatic hormones.

Organization of the Endocrine Pancreas

The endocrine function of the pancreas resides within the islet of Langerhans, which constitute 1–3% of total pancreatic mass. The normal adult pancreas contains about 1×10^6 pancreatic islets scattered throughout the body of this elongated organ. Islets are spherical structures less than the size of a pinhead, each comprising a cluster of 2000–6000 individual islet cells surrounded by a connective tissue capsule. The diameter of the islet ranges from 20 to 300 μm, depending on cellular composition.

Pancreatic Islet Cells

Pancreatic islets consist of four major types of secretory cell producing insulin (β cells), glucagon (α_2 cells), somatostatin (α_1 cells; also referred to as δ cells) and pancreatic polypeptide (PP cells) (Table 1). These are sometimes also referred to as B, A, and D cells, respectively. The products of several other minor secretory cell types located in the islets remain unknown. Islets also contain small amounts of additional peptides colocalized with classical islet hormones or present in nerve fibers. These agents have various effects on insulin secretion, as indicated in Table 1.

Each islet is composed of 60–70% β cells located primarily as a central core surrounded by an outer mantle of α_2 cells (5–25%), α_1 cells (5–10%) and PP cells (15–20%). The proportion of α_2 and PP cells varies considerably, depending on islet embryonic origin and location in the pancreas. Islets distributed the tail and body of the pancreas are generally rich in α_2 cells and contain relatively few PP cells, whereas the converse applies for islets located the ventral head region of the pancreas.

Table 1 Hormones, peptides colocalized in hormone-producing cells and neurotransmitters in islets of Langerhans

	Cell type	*Effect on insulin secretion*
Hormones		
Insulin	β cells	↓
Glucagon	α_2 cells	↑
Somatostatin	α_1 cells	↓
Pancreatic polypeptide	PP cells	↓
Peptides colocalized		
DBI	α_2, α_1 cells	↓
IAPP	β cells	↓
Pancreastatin	α_2, β, α_1 cells	↓
TRH	α_2, β cells	↑
Neurotransmitters		
Acetylcholine	Nerve fibers	↑
Noradrenaline	Nerve fibers	↓
CCK	Nerve fibers	↑
Galanin	Nerve fibers	↓
GRP	Nerve fibers	↑
VIP	Nerve fibers	↑

↑, stimulatory effect; ↓, inhibitory effect; CCK, cholecystokinin; DBI, diazepam-binding inhibitor; GRP, gastrin-releasing polypeptide; IAPP, islet amyloid polypeptide; TRH, thyrotropin-releasing hormone; VIP, vasoactive intestinal polypeptide.

Islet Vasculature

Islets are well vascularized and receive a considerable flow of blood from afferent arterioles arising from the splenic and pancreaticoduodenal arteries. Within the islet, the afferent vessels appear to perforate the cellular mantle, supplying a dense network of capillaries that traverse the β cell core, before draining into a small number of venules that ultimately feed into the pancreatic vein. The vasculature of individual islets is believed to constitute a tiny intraislet 'portal' system, with blood flowing progressively past β, α_2, and α_1 cells. This arrangement provides effective delivery to islet cells of bloodborne stimuli, such as glucose, and ensures rapid circulation of the secreted islet hormones to peripheral sites of action. There is also recent evidence that elevation of blood glucose specifically increases islet blood flow, via a cholinergic mechanism.

Islet Innervation

Islets receive an abundant supply of cholinergic, adrenergic, and peptidergic (nonadrenergic, noncholinergic (NANC)) nerves that innervate islet cells and the surrounding capillaries. In addition to the classical neurotransmitters – acetylcholine and noradrenaline – the islets contain a number of neuropeptides that exert stimulatory or inhibitory effects on islet cell secretions (Tables 1 and 2). Parasympathetic innervation is activated by feeding, whereas sympathetic nerves are activated in situations of need such as starvation or fright.

Islet Cell Interactions

Glucagon stimulates the secretion of insulin and somatostatin, whereas both insulin and somatostatin

are general inhibitors of islet hormone secretions (Table 2). This provides the basis for the possible involvement cellular interactions in the integration of islet secretory function (Figure 1). Such interactions may be restricted to adjacent cells within the islet, termed paracrine interactions, or may be more widely mediated through the intraislet portal system, which delivers insulin from β cells downstream to α_2 cells, followed by onward passage of both insulin and secreted glucagon to the somatostatin-containing α_1 cells.

Hormone Biosynthesis and Secretion

Biosynthesis and secretion of islet hormones occur through a similar sequence, involving the production of a precursor polypeptide in the rough endoplasmic reticulum followed by posttranslational enzymatic conversion to the active hormone. The active hormone is stored in secretory granules before ultimate release from the stimulated cell by a regulated process termed 'exocytosis.' This involves movement, aided by microtubules and microfilaments of the granule to

Table 2 Control of insulin, glucagon, somatostatin, and pancreatic polypeptide secretion

Agent	*Insulin*	*Glucagon*	*Somatostatin*	*Pancreatic polypeptide*
Nutrients				
Glucose	↑	↓	↑	↓
Amino acids	↑	↑	↑	↑
Fatty acids	↑	↓	↑	↑
Pancreatic hormones				
Insulin	↓	↓	↓	↑
Glucagon	↑	?	↑	–
Somatostatin	↓	↓	↓	↓
Pancreatic polypeptide	↓	–	↓	?
Intestinal hormones				
CCK[a]	↑	↑	↑	↑
Gastrin	↑	↑	↑	–
GIP	↑	↑	↑	↑
GLP-1[7-36]-amide	↑	↓	↑	?
Secretin	↑	↓	↑	–
VIP[a]	↑	↑	?	↑
Neurotransmitters				
Acetylcholine	↑	↑	↓	↑
α_2-Adrenergic agonists	↓	↓	↓	↓
β-Adrenergic agonists	↑	↑	↑	↑

[a]Also present in nerve fibers innervating islet cells.
↑, stimulatory effect; ↓, inhibitory effect; – no effect; ?, not established; CCK, cholecystokinin; GIP, gastric inhibitory polypeptide; GLP-1[7-36]-amide, glucagon-like peptide 1[7-36]amide (sometimes referred to as 'insulinoptropin'); VIP, vasoactive intestinal polypeptide.

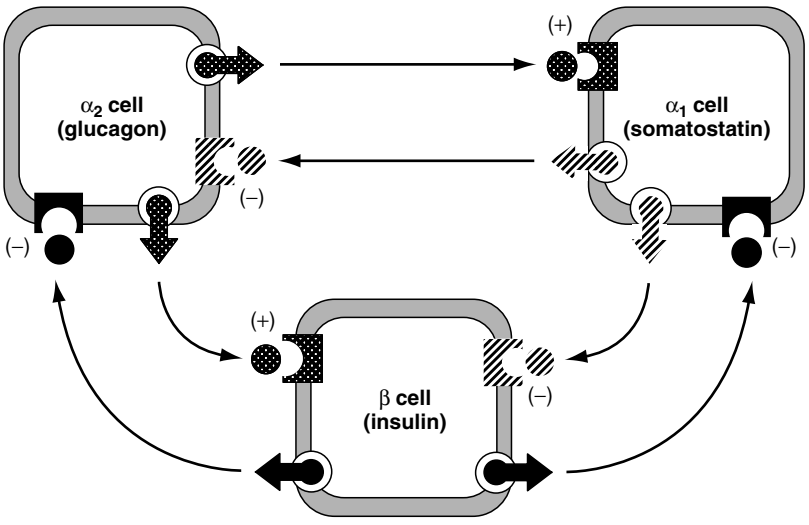

Figure 1 Hormonal interactions of major islet cells. Blood flows from β to α_2 to α_1 cells draining into the portal vein.

the periphery, followed by fusion of the granule membrane with the plasma membrane and liberation of the granule contents into the blood.

More specific details of events concerning biosynthesis and secretion are best known for insulin. In the β cell, transcription of messenger ribonucleic acid (mRNA) gives rise to production of preproinsulin, an 11.5-kDa polypeptide. This is rapidly cleaved within 1 min by proteolytic enzymes to proinsulin (86 amino acid residues, 9 kDa) which is transported by microvesicles to the Golgi apparatus. Proinsulin is then packaged into secretory granules, where it is converted by proteases to equimolar amounts of insulin (51 residues, 6 kDa) and proinsulin C-peptide. A number of second messengers are involved in coupling stimulus recognition to the secretion of insulin by exocytosis. These include cytoplasmic calcium ion concentration ($[Ca^{2+}]_i$), cyclic adenosine monophosphate (cAMP), diacylglycerol (DAG), and inositol 1,4,5-trisphosphate (IP_3).

In pancreatic α_2 cells, the initial product of the proglucagon is a 160-residue polypeptide (18 kDa) that contains the amino acid sequences of several biologically active peptides in addition to glucagon. Posttranslational processing of proglucagon yields glucagon (29 residues, 3.5 kDa) as the main product, plus a proglucagon fragment and, possibly, small amounts of glicentin-related pancreatic peptide (GRPP). In pancreatic α_1 cells, preprosomatostatin (116 residues) is similarly processed to prosomatostatin (92 residues) before conversion to the active, 14-amino-acid form of somatostatin (SRIF-14, 1.5 kDa). In PP cells, pancreatic polypeptide (36 residues, 4.3 kDa) also appears to be produced from a larger, 95-amino-acid, prohormone form. In the intestine, the posttranslational processing of proglucagon and prosomatostatin differs, yielding as major products glucagon-like peptide-1[7-36]-amide (GLP-1[7-36]-amide) and the 28-amino-acid form of somatostatin (SRIF-28), respectively.

Regulation of Secretion

A wide range of nutrients, hormones, and neurotransmitters are capable of influencing the secretion of insulin, glucagon, somatostatin, and pancreatic polypeptide (Tables 1 and 2). In the case of insulin, the major physiological regulator of β cell function is the prevailing plasma glucose concentration. Other metabolizable nutrients can mimic the stimulatory action of glucose, but these other agents basically act as modulators of the glucose response.

Insulin acting locally within the islet is an influential inhibitor of glucagon secretion. Glucose inhibition of glucagon secretion is thus attributable in large part to concomitant stimulation of the β cell with the release of insulin. γ-Aminobutyric acid (GABA), cosecreted with insulin from the β cell, has also been suggested to contribute to inhibition of α_2 cells. The paradoxical ability of amino acids and food-activated elements of the enteroinsular axis to stimulate glucagon as well as insulin secretion is often considered to offer protection from hypoglycemia that might result from the marked stimulation of insulin release. An alternative interpretation is that the stimulation of glucagon release serves to promote insulin secretion by paracrine interaction within the islet. Other situations associated with secretion of glucagon include hypoglycemia, starvation, and stress. In these circumstances, glucagon serves to mobilize nutrient fuels from body stores to meet immediate energy requirements.

Nutrient Regulation of Insulin Secretion

Although many factors influence the secretion of pancreatic hormones, the regulation of insulin secretion is particularly critical for normal metabolism and maintenance of good health. β cell dysregulation leading to deficient insulin secretion results in impaired glucose tolerance, hyperglycemia and type 2 diabetes (also known as noninsulin-dependent diabetes mellitus). Hypersecretion of insulin may result in life-threatening hypoglycemia, as observed in certain individuals with an islet cell tumor (insulinoma).

Nutrients exert direct and indirect effects on insulin secretion. Thus, in addition to the direct stimulatory actions of glucose, amino acids, and fatty acids on the β cell, the ingestion of food and absorption of nutrients trigger neural and hormonal elements of the enteroinsular axis (Figure 2). Activation of the parasympathetic nerve innervating the islets liberates acetylcholine from nerve terminals close to β cells. Stimulation of intestinal endocrine cells leads to the secretion of GLP-1[7–36]-amide, gastric-inhibitory polypeptide (GIP) and cholecystokinin (CCK). These agents are powerful insulin secretagogs at raised glucose concentrations, and they markedly augment the direct stimulatory actions of glucose and other nutrients.

The mechanisms through which nutrients and enteroinsular stimuli trigger insulin secretion from β cells involve elevation of $[Ca^{2+}]_i$ and sensitization of the exocytotic process to normal stimulatory action of $[Ca^{2+}]_i$ (Figure 3). Metabolizable nutrients, including glucose, certain amino acids, and fatty acids, act by virtue of the ability of the β cell to metabolize these substances with the generation of adenosine triphosphate (ATP). ATP generation results in closure of ATP-sensitive potassium channels (K_{ATP}) channels in

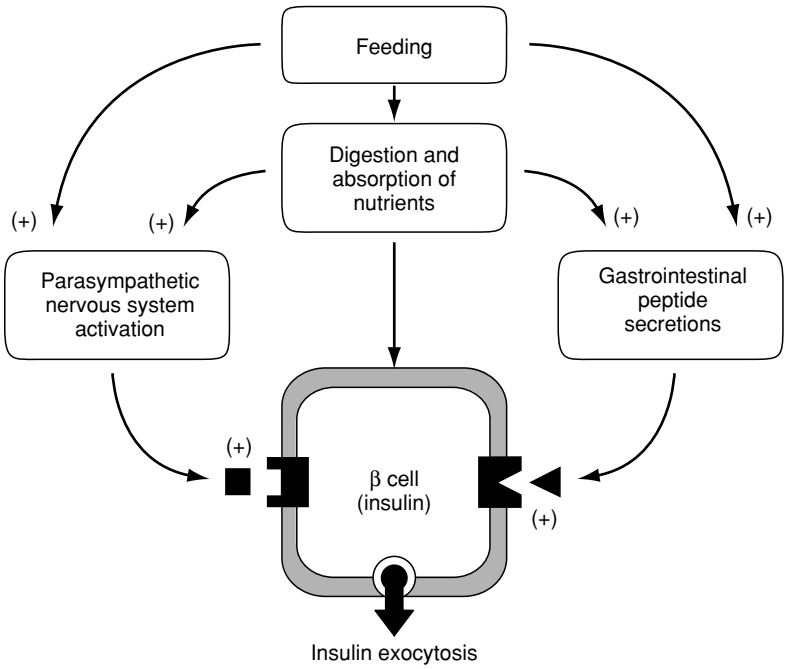

Figure 2 Major neural and hormonal pathways regulating the insulin secretory response to feeding.

the plasma membrane, leading to membrane depolarization. This results in the opening of voltage-dependent calcium channels (VDCCs) allowing Ca^{2+} to enter the β cell down its concentration gradient. The subsequent increase of $[Ca^{2+}]_i$ triggers secretion of insulin by exocytosis. Elevation of $[Ca^{2+}]_i$ helps to restore the membrane potential through opening of Ca^{2+}-sensitive K^+ channels in the plasma membrane. Conditions of elevated $[Ca^{2+}]_i$ increase the sensitivity of the β cell to the stimulatory actions of phospholipase C and adenylate cyclase. Acetylcholine, CCK, GLP-1[7-36]-amide and GIP are particularly potent activators of these enzymes. Generation of IP_3 inhibits Ca^{2+} sequestration by the endoplasmic reticulum, whereas protein kinase C (PKC) and cAMP sensitize the secretory process to the stimulatory action of Ca^{2+}. cAMP also activates protein kinase (PK) responsible for phosphorylation of functionally important proteins.

Physiological Actions

Within the islet, insulin is the principal hormone that appears to directly modulate the secretory activity of other islet cell types. Outside the islet, insulin is the anabolic hormone and driving force behind the regulation of carbohydrate, protein and lipid metabolism (Figure 4). Glucagon is the main catabolic counter-regulatory hormone, but the net effect of glucagon is more dependent on the insulin:glucagon molar ratio rather than the absolute level of glucagon *per se*. The minute-to-minute control of insulin and glucagon secretion from the islets is therefore of paramount importance in the overall regulation of nutrient homeostasis.

Insulin and glucagon exert major effects on the metabolism of carbohydrate, protein and lipid (Table 3). The principal actions of insulin are to lower circulating concentrations of glucose, amino acids and fatty acids, to increase fat storage in adipose tissue, and to promote the formation of muscle protein. Insulin achieves this by stimulating glucose and amino acid uptake and by alteration of the activities of enzymes in liver, muscle, and adipose tissue. These enzymes mediate the effects of insulin on glycogenesis, glycogenolysis, gluconeogenesis, glycolysis, lipogenesis, lipolysis, ketogenesis, and both the synthesis and degradation of protein. The actions of insulin on these metabolic pathways are generally opposed by the action of glucagon. Secretion of this counter-regulatory hormone in conjunction with a decrease in the insulin:glucagon molar ratio serves to generate a catabolic response in times of need, involving mobilization of nutrients from various tissue stores.

Somatostatin and PP secreted from the islets are less significant in physiological terms than the other two pancreatic hormones. Somatostatin and PP are also found in endocrine cells of the intestine, and they are both released into the circulation following feeding.

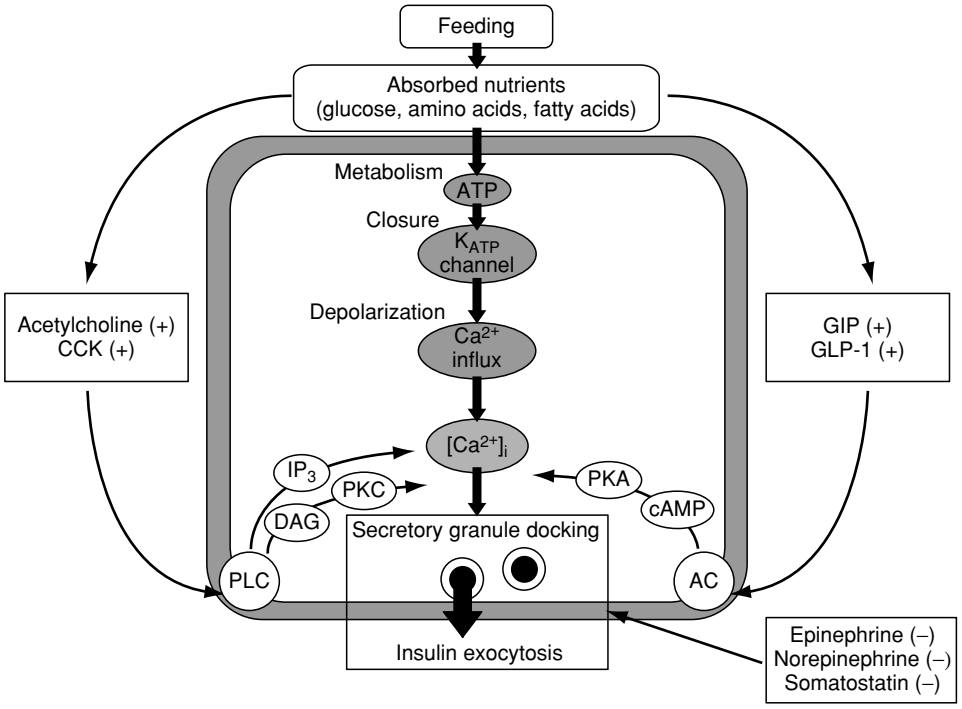

Figure 3 Direct and indirect actions of nutrients in the stimulation of insulin secretion from pancreatic β cells. AC, adenylate cyclase; ATP, adenosine triphosphate; cAMP, cyclic AMP; $[Ca^{2+}]_i$, cytoplasmic calcium ion concentration; CCK, cholecystokinin; DAG, diacylglycerol; GIP, gastric inhibitory polypeptide; GLP-1[7-36]-amide, glucagon-like peptide-1[7-36]-amide; IP_3, inositol 1,4,5-triphosphate; K_{ATP} channel, ATP-sensitive K^+ channel; PLC, phospholipase C; PKA, protein kinase A; PKC, protein kinase C.

Table 3 Major metabolic effects of insulin and glucagon on liver, muscle, and adipose tissue

Carbohydrate metabolism	*Protein metabolism*					*Lipid metabolism*		
	Insulin	*glucagon*		*Insulin*	*glucagon*		*Insulin*	*glucagon*
Blood glucose	↓	↑	Blood amino acids	↓	–	Blood fatty acids	↓	↑
Glucose uptake (muscle, adipose tissue)	↑	–	Protein (muscle)	↑	↓	Fat storage (adipose tissue)	↑	↓
Glycogenesis (liver, muscle)	↑	↓	Amino acid uptake (muscle, liver)	↑	↑	Lipogenesis (liver, adipose tissue)	↑	↓
Glycogenolysis (liver, muscle)	↓	↑	Protein synthesis (liver, muscle)	↑	–	Lipolysis (adipose tissue)	↓	↑
Gluconeogenesis (liver)	↓	↑	Protein degradation (liver, muscle)	↓	↑	Ketogenesis (liver)	↓	↑
Glycolysis (liver, muscle, adipose tissue)	↑	↓						

↑, stimulatory effect; ↓, inhibitory effect; –, no effect.

Somatostatin is a potent inhibitor of secretion from endocrine (pancreas, intestine, pituitary) and exocrine cells, partly acting through paracrine effects, thus contributing to the fine tuning of insulin and glucagon secretion. Somatostatin also inhibits the following aspects of gastrointestinal function, largely as a result of inhibition of gut hormone secretions: gastric acid secretion; secretion of bile and pancreatic juice; gall bladder contraction; intestinal motility; intestinal absorption; splanchnic blood flow.

PP has been shown to exert various actions at pharmacological concentrations. These include the following: inhibiting exocrine and endocrine pancreatic secretions; gall bladder relaxation; increasing gastrointestinal motility; inducing satiety. However, there is presently little convincing evidence for a physiological role of PP.

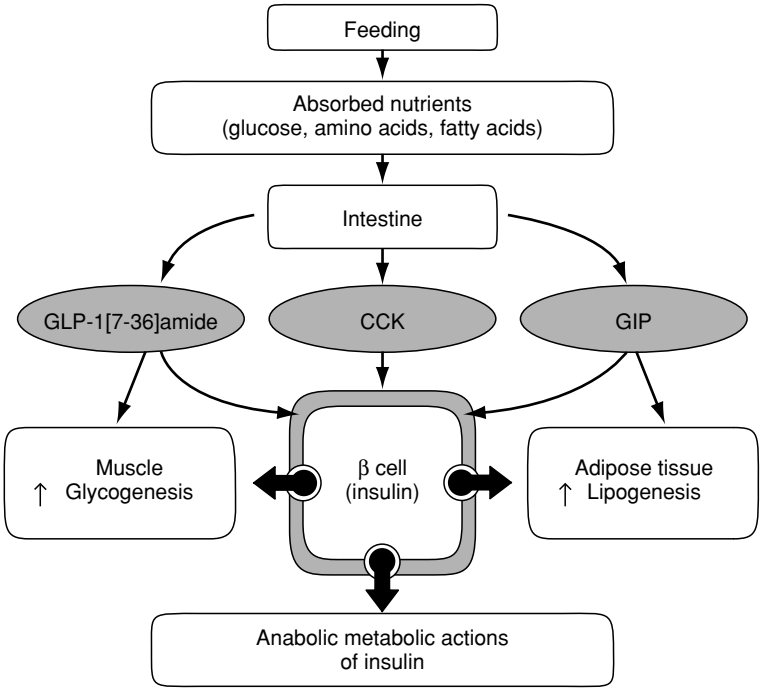

Figure 4 Insulin-mediated anabolic responses to feeding. CCK, cholecystokinin; GIP, gastric inhibitory polypeptide; GLP-1[7-36]-amide, glucagon-like peptide-1[7-36]-amide.

Mechanisms of Action

Insulin and the other pancreatic hormones are secreted into the interstitial fluid; they cross the endothelial barrier and enter the circulation. Effects at the target cells are initiated by the action and binding of the hormone to specific cell surface receptors (Figure 5). The binding of the hormone to these receptors initiates a sequence of postreceptor events ultimately culminating in the various biological actions of the hormone. Rapid events on the secretion of other hormones are mediated by changes of $[Ca^{2+}]_i$ and the activities of adenylate cyclase and phospholipase C (see above). Stimulation of glucose uptake by muscle and adipose tissue results from insulin-induced translocation of glucose transporter proteins (GLUT4) from cytoplasmic vesicles to the plasma membrane. Classical metabolic effects of insulin and glucagon at target cells are achieved by activation and suppression of enzyme activity (so redirecting cell metabolism) or by altering the rate of synthesis of enzymes at the level of transcription and translation.

Insulin binding to the external α-subunit of the glycoprotein insulin receptor leads to a conformational change in the receptor that stimulates tyrosine kinase activity. The activated the β-subunit autophosphorylates at tyrosine sites in addition to phosphorylating intracellular proteins (Figure 5). Indeed it is the tyrosine kinase activity that is essential for insulin signaling. The best characterized postreceptor signal transduction pathway centers around insulin-receptor substrate 1 (IRS-1). Rapid tyrosine phosphorylation following insulin stimulation results in noncovalent binding between phosphorylated sites and specific domains on intracellular protein targets. This ultimately results in the various important biological effects of insulin on target tissues (Figure 5). Following initiation of postreceptor events, the insulin–receptor complex is internalized by the target cell, and whereas receptors are recycled to the cell surface, insulin is degraded intracellularly. Elevated insulin levels, for example in obesity or type 2 diabetes, results in insulin receptor 'downregulation' through decreasing numbers on the cell surface and also by reducing tyrosine kinase activity of the receptor.

Endocrine Pancreas in Diabetes

The two major forms of diabetes, type 1 insulin-dependent diabetes mellitus and type 2 diabetes, are associated with marked changes in the structure of the endocrine pancreas. In long-standing type 1 diabetes, the insulin deficiency is the result of total autoimmune destruction of β cells, while the mass of α_2, α_1 and PP cells remains virtually unchanged. However, in type 2 diabetes, the β cell mass is marginally decreased in association with increases in the α_2 cell

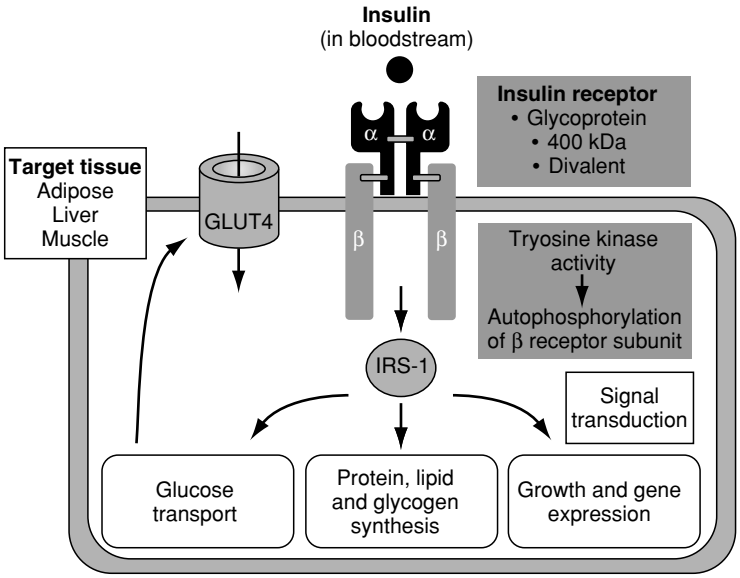

Figure 5 Insulin signalling in target cells. α, α-subunit of insulin receptor; β, β-subunit of insulin receptor; GLUT4, glucose transport protein; IRS-1, insulin receptor substrate-1.

mass and α_2:β cell ratio. Hyperglucagonemia and inappropriate α_2 cell function, both owing to a lack of inhibition of glucagon secretion by insulin, contribute to glucose intolerance and associated metabolic disarray generally observed in diabetes.

Type 2 diabetes is classically associated with defective β cell function, normal or moderately raised insulin concentrations, and insulin resistance. Insulin secretory responses to glucose and other stimuli are impaired in magnitude and kinetics. The β cells in type 2 diabetics are also known to release excessive amounts of proinsulin rather than insulin. Since proinsulin cross-reacts with most antisera used to measure insulin, actual concentrations of insulin in type 2 diabetes are likely to be much lower than previously estimated. Proinsulin and derived fragments exhibit only 10–50% of the biological activity of insulin, and therefore may contribute to the condition of insulin resistance.

See also: **Diabetes Mellitus**: Etiology; Chemical Pathology; Treatment and Management; Problems in Treatment; Secondary Complications

Further Reading

Alberti KGMM, DeFronzo RA and Zimmet P (eds) (1997) *International Textbook of Diabetes Mellitus*, 2nd edn. Chichester, UK: John Wiley & Sons.

Ashcroft F and Ashcroft SJ (eds) (1992) *Insulin*. Oxford: Oxford University Press.

Bailey CJ and Flatt PR (eds) (1990) *New Antidiabetic Drugs*. London: Smith-Gordon.

Flatt PR (ed.) (1992) *Nutrient Regulation of Insulin Secretion*. London: Portland Press.

Flatt PR and Lenzen S (eds) (1994) *Frontiers of Insulin Secretion and Pancreatic B-cell Research*. London: Smith-Gordon.

Frayn KN (1996) *Metabolic Regulation: A Human Perspective*. London: Portland Press.

McClenaghan NH and Flatt PR (1999) Physiological and pharmacological regulation of insulin release: insights offered through exploitation of insulin secreting cell lines. *Diabetes, Obesity and Metabolism* 1: 137–150.

Montague W (1983) *Diabetes and the Endocrine Pancreas: A Biochemical Approach*. London: Croom Helm.

Pickup JC and Williams G (eds) (2002) *Textbook of Diabetes*, 3rd edn. Oxford: Blackwell Scientific Publications.

Porte D, Jr and Sherwin RS (eds) (1997) *Ellenberg and Rifkin's Diabetes Mellitus: Theory and Practice*, 5th edn. Stamford, Connecticut, CT: Appleton & Lange.

Williams G and Pickup JC (eds) (1999) *Handbook of Diabetes*, 2nd edn. Oxford: Blackwell Scientific Publications.

Pituitary Hormones

M Wallis, The University of Sussex, Falmer, Brighton, UK

The Pituitary Gland

In humans the pituitary gland (**Figure 1**) is a small organ about the size of a pea which occupies a bony pocket in the skull (the sella turcica), situated below

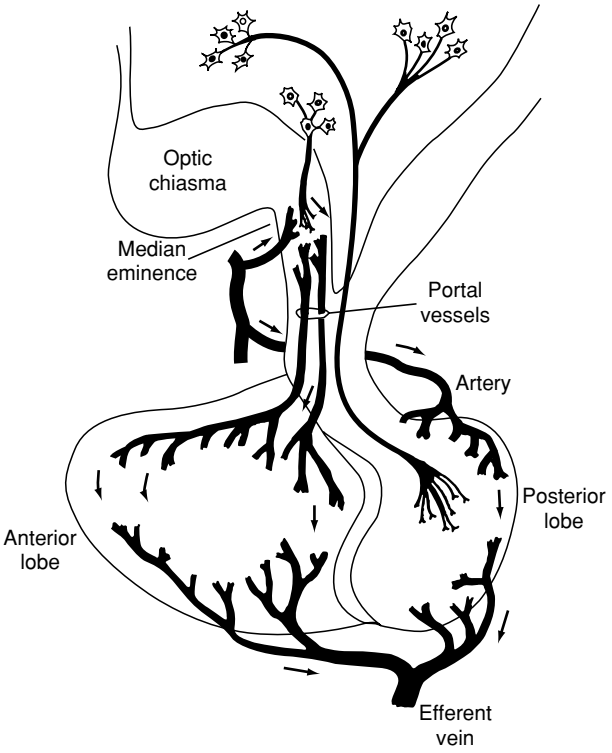

Figure 1 The pituitary gland. Factors secreted at nerve terminals in the median eminence of the hypothalamus are carried via the hypophysial portal vessels to the anterior lobe of the pituitary gland where they regulate the secretion of the various pituitary hormones. Precursors of oxytocin and vasopressin are synthesized in neurons in the hypothalamus and transported via axons (with concomitant processing to the active peptides) to the posterior lobe of the pituitary gland. Modified from Wallis M (1988) The molecular basis of growth hormone deficiency. *Molecular Aspects of Medicine* 10: 429–509.

the midbrain. It is attached to the floor of the third ventricle of the brain (the hypothalamus) by a stalk and this close association between the brain and the pituitary is of general importance to the functioning of the latter. The gland comprises two main parts, with different embryological origins – the neurohypophysis and the adenohypophysis. The neurohypophysis (posterior lobe) originates as a downgrowth from the floor of the brain, the connection with the brain being retained as a major component of the pituitary stalk. The bulk of the neurohypophysis is made up of the swollen nerve terminals of nerve cells which have their cell bodies in the hypothalamus (particularly the supraoptic and paraventricular nuclei), and whose axons pass through the stalk to the neurohypophysis. The adenohypophysis is ectodermal in origin, being formed during development from a pouch (Rathke's pouch) derived from an ingrowing of the cells that will form the roof of the mouth. The adenohypophysis comprises several regions, including the anterior lobe (the source of most adenohypophysial hormones) and the intermediate lobe. The anterior lobe contains about six different hormone-secreting cell types, secreting 6–10 different hormones (Table 1); the main function of several of these is to control the activity of other endocrine glands. Pituitary growth hormone has a particularly important role in regulating the nutritional balance of the organism, and this hormone is emphasized in the second half of this article.

Neurohypophysial Hormones

The neurohypophysis in mammals secretes two peptide hormones, oxytocin and vasopressin (antidiuretic hormone, ADH). Each of these is a nonapeptide, including a disulfide bridge and an amidated C-terminus. Oxytocin and vasopressin are very similar, differing only at two positions. Peptides with slightly different structures but substantially different biological actions are found in lower vertebrates; some invertebrates possess related peptides.

In mammals oxytocin functions mainly to regulate contraction of smooth muscle, especially in mammary gland (leading to milk ejection) and uterus (stimulating uterine contraction in parturition). Several lines of evidence, including recent studies with transgenic (knockout) mice, suggest that the effects on the uterus may be of relatively little physiological importance, although the hormone is widely used clinically to induce birth. The significance of the effects on milk ejection is clearly established – if secretion of oxytocin is prevented (for example, by anesthetizing a suckling animal) the young are unable to obtain milk from the mammary gland. The mechanism of action of oxytocin involves binding to a G-protein-linked receptor on the target cells, followed by activation of intracellular signalling systems leading to elevation of Ca^{2+} levels and myosin light chain phosphorylation.

The main physiological action of vasopressin in mammals is regulation of water resorption by the kidney. A small elevation of blood osmotic pressure leads to increased vasopressin secretion and increased water retention at the kidney. Again, the hormone binds to a specific G-protein-linked receptor on its target cells, followed by activation of intracellular signaling systems, including elevation of cyclic adenosine monophosphate (AMP) levels, leading to insertion of water channels into the luminal membrane of cells of the loop of Henle in the kidney. Vasopressin has various other effects, including actions on blood pressure, hepatic metabolism, and adrenocorticotropic hormone (ACTH) secretion; at least some of these are mediated by a different type of receptor from that found in the kidney.

Table 1 The hormones of the mammalian pituitary gland

Hormone	*Abbreviation*	*Chemical nature*	*Main actions*
Posterior lobe			
Oxytocin		Peptide (9 residues)	Stimulates milk ejection, uterus contraction
Vasopressin (antidiuretic hormone)	VP (ADH)	Peptide (9 residues)	Antidiuretic, raises blood pressure
Intermediate lobe			
α-Melanotropin (α-melanocyte-stimulating hormone)	α-MSH	Peptide (13 residues)	Causes skin darkening, especially in amphibia; functions as neuropeptide in the brain
Anterior lobe			
β-Melanotropin (β-melanocyte-stimulating hormone)	β-MSH	Peptide (18–22 residues)	Causes skin darkening, especially in amphibia
Corticotropin (adrenocorticotropic hormone)	ACTH	Peptide (39 residues)	Stimulates corticosteroidogenesis and growth of adrenal cortex
Lipotropin	LPH	Peptide/protein (~91 residues)	Weakly lipotropic (may be primarily a precursor of β-MSH and β-endorphin)
β-Endorphin		Peptide (31 residues)	Analgesia?
Growth hormone (somatotropin)	GH	Protein (~190 residues)	Promotes growth, many anabolic actions, stimulates insulin-like growth factor 1 production
Prolactin	PRL	Protein (~200 residues)	Lactogenic activity (many other actions, particularly in lower vertebrates)
Thyrotropin (thyroid-stimulating hormone)	TSH	Glycoprotein (2 subunits)	Stimulates thyroid hormone production
Follicle-stimulating hormone	FSH	Glycoprotein (2 subunits)	Stimulates maturation of the Graafian follicle in female, spermatogenesis in male
Luteinizing hormone (interstitial cell-stimulating hormone)	LH	Glycoprotein (2 subunits)	Stimulates maturation and release of ovum and early development of corpus luteum in female, steroid production by gonads in male

Note that some doubt remains as to whether β-MSH, LPH, and β-endorphin are physiologically significant hormones.

Oxytocin and vasopressin are synthesized as precursors in the cell bodies of (different) neurons in the hypothalamus. They are packaged into granules, which pass down the axons of these neurons, through the pituitary stalk, to the swollen axonal endings in the neurohypophysis, where they are stored. Conversion of precursors to active peptides occurs during formation and transport of secretory vesicles, catalyzed by prohormone-converting enzymes included within the vesicles. Also formed by processing of these precursors are neurophysins, which act as binding proteins for oxytocin and vasopressin prior to release. Secretion of the hormones from the nerve endings occurs by exocytosis in response to nerve signals coming down the axons of the same cells. For oxytocin a neural loop links sensory receptors in the mammary gland (detecting suckling) via a number of synapses, to the cell bodies of the neurons that synthesize the hormone. This in turn triggers firing of these neurons, so that an electrical signal passes down their axons and causes depolarization of the plasma membrane of the swollen nerve endings in the neurohypophysis. The entry of Ca^{2+} which follows this provides the signal for exocytosis. The released oxytocin passes via the blood stream to the mammary gland, where it induces milk ejection within a few seconds of the initial suckling stimulus. Similar neural loops regulate secretion of vasopressin, initiated at receptors in the brain that detect changes in osmotic pressure, or in the atrium and carotid sinus that detect changes in blood volume or pressure.

Adenohypophysial Hormones

The hormones of the adenohypophysis are all peptide or protein in nature, and fall into three main families: (1) corticotropin (ACTH), melanocyte-stimulating hormone (MSH) and related peptides (the melanocortin family); (2) growth hormone (GH) and prolactin; (3) thyrotropin or thyroid-stimulating hormone (TSH), follicle-stimulating hormone (FSH) and luteinizing hormone (LH) – the glycoprotein hormone family.

The Melanocortin Family

ACTH is a peptide of 39 residues concerned mainly with regulation of the adrenal cortex, including production and secretion of corticosteroids. It binds to a G-protein-linked receptor in the membrane of target cells in the adrenal cortex, and acts at various points on the pathway leading to steroid synthesis partly via action of cyclic AMP as a second messenger. MSHs

are peptides of 13–22 amino acids, produced characteristically in the intermediate lobe of the pituitary, which regulate skin color in some species by stimulating synthesis and altering distribution of pigment in cells called melanocytes. The actions of MSHs are particularly marked in amphibia. Although they occur in the pituitary gland of many mammals their physiological role there is not clear, although in at least some species they probably do have a role in controlling skin pigmentation. MSHs have been described in the human pituitary, but there is some evidence that in the adult at least they are artifacts of extraction procedures and of little physiological importance. ACTH and MSHs show similarity in amino acid sequences, and it is now known that they are formed from a common biosynthetic precursor, proopiomelanocortin (POMC), together with a larger peptide β-lipotropin and the opioid peptide β-endorphin. Despite this, ACTH and α-MSH are secreted from different cell types (corticotrophs and melanotrophs), reflecting different processing pathways in different cells.

Although MSHs and the other POMC-derived peptides were originally characterized in the pituitary, it is now clear that they are also expressed in the brain, and that as neuropeptides they may play an important role in energy homeostasis and the regulation of obesity. Recent studies indicate that the peptide hormone leptin, produced by adipose cells, acts at least partly by stimulating production of α-MSH in hypothalamic neurons. Binding of α-MSH to specific receptors (MC4 receptors) in hypothalamus and cortex leads to reduced feed intake and increased energy expenditure. Animals with defective MC4 receptors become obese, but the importance of this mechanism for human obesity has not yet been established.

The Growth Hormone/Prolactin Family

GH and prolactin are both medium-sized proteins comprising a single polypeptide chain of 190–200 amino acids. They are structurally similar, and are related to the placental lactogens found in humans, ruminants, and rodents. GH plays a major role in regulating somatic growth and is discussed further below. Prolactin has a number of different actions in mammals, the most important being concerned with regulation of lactation and mammary growth. Prolactin acts with other hormones, including insulin, insulin-like growth factors, corticosteroids, thyroid hormone, and GH (depending on the species) to stimulate growth of the mammary gland and synthesis and secretion of milk components including proteins (caseins, α-lactalbumin, etc.), sugar, and fat. It acts partly at the gene level to stimulate production of specific proteins via a membrane-bound receptor, Jak kinase and STAT proteins, a mechanism similar to that of GH (see below). In ruminants milk production appears to be stimulated more effectively by GH than prolactin, and the use of GH to enhance agricultural milk production is being actively pursued. Whether GH and/or prolactin can influence the nutritional content of milk, as well as total production, is not clear.

The Glycoprotein Hormone Family

The third family of pituitary hormones includes the glycoprotein hormones: the two gonadotropins, FSH and LH, and TSH. These are structurally related to the chorionic gonadotropin (hCG) produced in the human placenta. They are complex proteins, including in each case two subunits (α and β), each of which is about 100 amino acids long and carries one or more carbohydrate moieties. The α-subunit is common to FSH, LH, and TSH (and also hCG), while the β-subunits are different, and provide the hormonal specificities. A three-dimensional (3D) structure has been reported for hCG which shows a number of interesting features, including a 'disulfide knot' in each subunit and a very close association between α- and β-subunits, 'sealed' by a peptide strand and disulfide bridge which act as a kind of 'seat belt.' In the female, FSH and LH control the estrus cycle (menstrual cycle in primates): FSH stimulates growth of the Graafian follicle and production of estrogen by granulosa cells, while LH stimulates maturation of the ovum, ovulation, and development of the corpus luteum. In the male, LH stimulates the steroid-producing interstitial cells to produce androgens, while FSH stimulates spermatogenesis. The actions of both LH and FSH involve binding to G-protein-linked receptors and are mediated in part by cyclic AMP. The actions of TSH are concerned almost exclusively with regulation of the thyroid gland, including stimulation of growth of the gland and production of thyroxine and triiodothyronine.

Regulation of Secretion of Adenohypophysial Hormones

The secretion and synthesis of the hormones of the pituitary gland are largely under the control of the hypothalamus. For the neurohypophysis, as discussed above, such regulation involves nerve signals operating via the axons of the hypothalamic neurons within which vasopressin and oxytocin are synthesized and transported to the neurohypophysis.

Hypothalamic regulation of synthesis and secretion of adenohypophysial hormones in mammals involves primarily blood-borne neuroendocrine factors rather than nerve signals. A capillary plexus in the

hypothalamus allows the blood to pick up neuroendocrine secretions; the capillaries then combine to form the hypothalamic–hypophysial portal vessels which pass down the pituitary stalk and form a second capillary network in the adenohypophysis. This system allows efficient transfer of neuroendocrine factors from the hypothalamus to the adenohypophysis (Figure 1). These factors are mostly peptides, though the catecholamine dopamine is also an important regulator. Most stimulate secretion of one or more adenohypophysial hormones, but some are inhibitors. They are listed in Table 2.

ACTH secretion is stimulated by the hypothalamic peptide corticotropin-releasing hormone (CRH), which synergizes with vasopressin and possibly other factors. α-MSH secretion from the intermediate lobe may be partly under neural control, but is also regulated by hypothalamic peptides and inhibited by dopamine. Secretion of TSH is stimulated by the tripeptide TSH-releasing hormone (TRH). Secretion of FSH and LH is stimulated by the decapeptide LH-releasing hormone (LHRH: gonadotropin-releasing hormone (GnRH)); FSH and LH are synthesized in and secreted from the same cell type (gonadotroph). Prolactin secretion is largely under inhibitory control, by hypothalamic dopamine and possibly other factors. TRH and a number of other peptides can stimulate prolactin secretion, but their physiological significance is not clear. GH secretion is stimulated by the 44-residue peptide GH-releasing hormone (GHRH), and inhibited by the 14-residue peptide somatostatin. Both appear to be physiologically important, though stimulatory control predominates. Recently another peptide (ghrelin) that stimulates GH secretion has been characterized (see below).

Many other factors can influence anterior pituitary secretion, by modulating the actions of the hypothalamic regulators and/or in some cases by direct action at the pituitary. Thus, the secretion of FSH and LH may be modified by hormones produced by the gonads, including gonadal steroids and peptide hormones such as inhibin. Such modulation probably underlies differential release of LH and FSH from a single cell type. Hormones produced from other endocrine organs can also affect secretion of the pituitary hormones regulating those organs. Thus thyroid hormones modulate TSH secretion, and corticosteroids affect ACTH secretion. Secretion of GH is modulated by insulin-like growth factors (IGFs).

At the physiological level, factors controlling the release of hypothalamic peptides are obviously important in the overall control of pituitary hormone secretion. These factors include neural signals (including sensory inputs and signals deriving from endogenous rhythms) and hormonal/metabolic factors, such as glucose or amino acid levels, steroid hormone levels, etc. The hypothalamus provides a mechanism by which the neural and hormonal regulatory systems can be integrated, providing a combined input to the pituitary gland which in turn controls many of the endocrine glands of the body.

Growth Hormone

Each of the pituitary hormones has direct or indirect effects on nutrition. Thus, ACTH and TSH stimulate production of hormones involved in metabolic regulation. Prolactin, by stimulating lactation, affects both the nutrition of the offspring and the nutritional balance of the mother in whom much of the metabolic effort may be diverted towards milk production. However, the pituitary hormone that is most directly involved in regulating the nutritional balance of the organism is GH.

Structure

GH is a protein hormone of about 22 kDa molecular weight. Amino acid sequences of GHs from many

Table 2 The hypothalamic hormones that regulate the adenohypophysis

Hormone	*Abbreviation*	*Chemical nature*	*Main actions*
TSH-releasing hormone	TRH	Peptide (3 residues)	Stimulates release of TSH and possibly prolactin
LH-releasing hormone	LHRH	Peptide (10 residues)	Stimulates release of LH and FSH
Somatostatin (GH release-inhibiting hormone)	SRIF	Peptide (14 residues)	Inhibits release of GH (and many other secreted peptides)
Dopamine	DA	Catecholamine	Inhibits release of prolactin and MSH
GH-releasing hormone	GHRH	Peptide (44 residues)	Stimulates release of GH
Corticotropin-releasing hormone	CRH	Peptide (41 residues)	Stimulates release of corticotropin

TSH, Thyroid-stimulating hormone; LH, luteinizing hormone; FSH, follicle-stimulating hormone; GH, growth hormone; MSH, melanocyte-stimulating hormone.

Note that these are the hypothalamic factors that have now been well characterized. Other factors have been proposed, but either they have not been fully characterized or their physiological significance is in doubt. These include oxytocin-related peptides that may be involved in regulating MSH secretion, a factor that may specifically stimulate FSH secretion, an inhibitor of adenocorticotropic hormone secretion, and factors involved in stimulation of prolactin release.

species are available, and the three-dimensional structures of pig and human GHs have been determined, using X-ray diffraction. The structure comprises a four-helix bundle with an unusual ('up-up-down-down') topology. A similar topology has now been seen in a variety of cytokines and growth factors, indicating that GH is a member of a large protein superfamily (the cytokine family). The receptors for members of this family are also related. There is considerable sequence variation between GHs from different mammals. Most notably, human GH is very different from GHs of nonprimate mammals. Variation in biological properties is also seen – in particular, nonprimate GHs are not active in humans, and human GH, unlike most nonprimate GHs, possesses lactogenic (prolactin-like) activity. The species specificity has meant that it has been necessary for therapeutic purposes to prepare human GH on a large scale in order to treat human hypopituitary dwarfism (see below).

Biological Actions

GH stimulates somatic growth in most vertebrates. Removal of the pituitary gland leads to marked slowing of growth in young animals, and the effect can be reversed by GH treatment. Humans with deficient GH production suffer from retarded growth, leading to dwarfism, which can often be treated by exogenous GH. On the other hand, excess GH, due usually to an adenohypophysial tumor, leads to gigantism (if the tumor develops early in life, before the epiphyses of the long bones fuse) or acromegaly (if the tumor develops later, when only certain bones, characteristically those of the face, hands and feet, can still grow). GH is not the only factor affecting growth, of course. All animals have a genetic program that determines their normal adult size range, presumably optimized by evolutionary pressure to fit a particular niche. The role of GH is presumably to regulate growth of the organism within this optimum range, relative to environmental factors, particularly nutrition. Thus, if food is in short supply, normal growth is slowed, and if the short supply is followed by surplus, growth may be increased to above normal levels (catch-up growth). The final size that an adult mammal achieves will be a compromise between the optimal size that the genetic program determines, and what can be achieved given the environmental circumstances, particularly with regard to nutrition. For an animal to grow in an unregulated fashion towards a predetermined size (defined mainly in terms of bone length) could lead to decreased fitness, due to decreased muscle mass or bone strength. The link between the effects of nutrition and GH on growth is emphasized by the observation that responses to GH in experimental animals may be influenced markedly by the plane of nutrition.

The biological actions of GH are complex, and not well understood. The overall effect of the hormone is to promote growth, and it has direct or indirect effects on many aspects of protein, carbohydrate, and lipid metabolism, as well as effects on division and differentiation of some cell types. Whether all of these effects should be seen as directly related to stimulation of growth, or whether some of the actions of the hormone should be considered quite separately from the growth-promoting effects, is not clear. Some of the actions of GH are mediated by somatomedins/IGFs (**Figure 2**). These are single-chain polypeptides that show structural similarity to insulin. Two main IGFs have been characterized – IGF-1 and IGF-2. IGF-1 is most markedly GH-dependent. IGFs occur in the circulation largely in association with specific binding proteins, of which there are at least six types. These binding proteins extend the half-life and modulate the activity of IGFs; at least one of them is itself GH-dependent.

Circulating levels of IGF-1 are low in hypophysectomized rats or GH-deficient animals or people, and are increased following administration of GH. Unlike GH, IGF-1 stimulates cartilage growth directly, which is a prerequisite for growth of bone. For some time it was considered that circulating IGF-1 (produced mainly in the liver) was the main mediator of the actions of GH on cartilage and bone growth, but it is now clear that GH stimulates IGF-1 production in many tissues, including cartilage, and that the growth factor probably acts locally to stimulate tissue growth, possible synergizing with GH itself.

In addition to its actions on cartilage and bone growth, mediated largely by IGFs, GH also has anabolic effects on various other tissues. For example it stimulates protein synthesis in muscle and liver, including both a general enhancement of the protein synthetic machinery, and induction of gene expression for specific proteins, including IGFs and their binding proteins. GH also has effects on carbohydrate and lipid metabolism. The effects on carbohydrate metabolism include short-term insulin-like effects and longer-term antiinsulin (diabetogenic) effects. Actions on lipid metabolism include stimulation of lipid mobilization and uptake, and decrease in lipid synthesis. Overall the metabolic effect of the hormone is to promote lipid utilization and decrease carbohydrate utilization. The extent to which these metabolic actions are mediated by IGFs is not clear.

Mechanism of Action of Growth Hormone

Understanding of the mechanism by which GH acts has increased substantially over the past decade. As

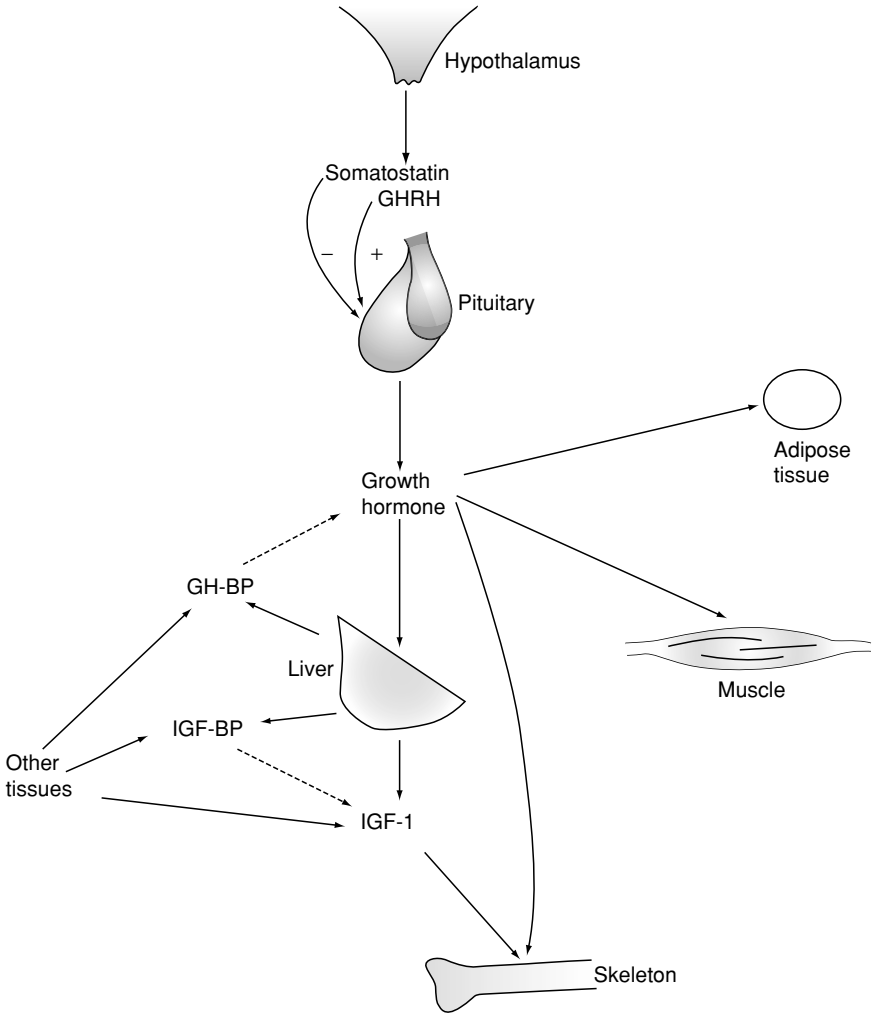

Figure 2 The hypothalamic–pituitary–somatic axis. Circulating insulin-like growth factor 1 (IGF-1) derives both from the liver and other tissues; at least some of the actions of IGF-1 involve local rather than humoral effects. The half-life and actions of growth hormone (GH) and IGH-1 are modulated by corresponding binding proteins (BP). GHRH, growth hormone-releasing hormone. Modified from Wallis M (1988) The molecular basis of growth hormone deficiency. *Molecular Aspects of Medicine* 10: 429–509.

with other polypeptide hormones the interaction with the target cell is via a membrane-bound receptor. When the GH receptor was first cloned and characterized in 1987 it proved to be of a type not previously observed. Since then, receptors of a similar type have been identified for a number of other polypeptide signaling molecules (members of the cytokine receptor family), including prolactin, erythropoietin, and various growth factors and lymphokines. All of these receptors comprise a single membrane-spanning domain, with substantial extracellular (ligand-binding) and intracellular domains. For GH there is evidence that the hormone binds to two receptor molecules, effectively dimerizing or cross-linking the receptor. The most striking demonstration of this has come from X-ray crystallographic studies, which revealed the detailed 3D structure of GH bound to two molecules of the extracellular domain of the receptor. Cross-linking of two extracellular domains brings together the corresponding intracellular domains, and it is thought that interaction between these allows the signal to be transmitted to the interior of the target cell. Thus far there is evidence for only one GH receptor gene in any one mammalian species, but there is generation of receptor heterogeneity by allelic variation, alternative splicing during pre-mRNA processing and/or posttranslational modification. The significance of such receptor heterogeneity is not yet clear, but a notable feature of it is the presence of a soluble form of the extracellular domain in the normal circulation – this acts as a GH-binding protein and may modulate the activity of the hormone and/or prolong its half-life.

The GH receptor is neither a G-protein-linked receptor, nor does it include a protein kinase. However, a major aspect of its mechanism of action is that on activation, following hormone binding and dimerization, it binds and activates a soluble cytosolic protein kinase Jak2. This leads to phosphorylation of the receptor itself and of a number of intracellular effector proteins, including members of the Stat (signal transduction and activation of transcription) protein family, particularly Stat5. Phosphorylated Stat5 moves to the nucleus where it acts as a transcription factor, stimulating the transcription of key GH-dependent genes, including the gene for IGF-1. Activation of Jak2 also leads to stimulation of various other intracellular signal transduction mechanisms, including the MAP kinase pathway.

Regulation of GH Secretion and Synthesis

The secretion of GH is regulated by stimulatory and inhibitory peptides produced in the hypothalamus, including GHRH and somatostatin. Stimulatory control generally predominates. Recently another peptide that stimulates GH release, ghrelin, has been identified in the stomach. Its physiological role is not yet clear, but it also stimulates appetite and may provide a link between GH secretion and food intake. The actions of these peptides are modulated by factors acting directly at the pituitary level, including IGFs, thyroid hormones, and glucocorticoids. Secretion of the hypothalamic peptides is in turn regulated by neural and metabolic factors. GHRH, and hence GH secretion, is stimulated by hypoglycemia, high levels of amino acids such as arginine or leucine (which may result from a high protein meal), and lowered free fatty acid levels. GH levels are lowered in obese subjects, and recent studies suggest that the 'fat hormone' leptin may play some role in modulating GH secretion. GH secretion is normally pulsatile in animals and humans, which complicates the determination of 'normal' levels.

Growth Hormone Deficiency

In humans GH deficiency gives rise to stunted growth/dwarfism and can have a number of causes, some of which are hereditary. Hereditary GH deficiency can be due to: (1) deletion of the gene that codes for GH; (2) mutations that prevent normal development of GH-producing cells (somatotrophs) and GH synthesis; (3) defective GH receptors (so that although normal levels of the hormone are produced they fail to produce normal responses in the target cells); and (4) other reasons, not fully understood. Most cases of GH deficiency in humans are not hereditary in origin, however; they often appear to be due to damage to the link between hypothalamus and pituitary, so that the normal control of GH secretion by GHRH is disrupted. Such disruption can be due to birth trauma, blows to the head in infancy, or other causes. In such cases the pituitary can often respond to GHRH administered exogenously, and this is being explored as a therapeutic approach to treat human hypopituitary dwarfism.

Children suffering from stunted growth due to clearly established GH-deficiency can be treated using human GH, and this has become a standard therapy. Provided GH treatment is started sufficiently early, growth can be restored to near normal, and heights within the normal range can be achieved. For about 20 years the GH used for such therapy was extracted from human pituitary glands obtained postmortem, and in many countries extensive schemes were established for collection of such material. In the mid-1980s, however, concern arose about the contamination of some human GH preparations with infective agents, particularly the prions causing Creutzfeldt–Jakob disease (CJD), and use of GH derived from human pituitaries was largely stopped. Fortunately at that time the possibility arose of using human GH prepared by genetic engineering techniques, and this is now widely employed.

The possibility of using GH to treat forms of short stature other than those due to clear-cut GH deficiency has been considered for some time. The availability of GH produced by recombinant DNA methods in potentially unlimited quantities makes wider usage more feasible, and this is being explored actively. As well as extending treatment to children whose short stature is due to primary causes other than GH deficiency, the use of recombinant DNA-derived GH to treat adults with GH deficiency is being considered, as is its use for promoting anabolism in muscle-wasting conditions such as acquired immunodeficiency syndrome (AIDS), cachexia, and even old age.

Growth Hormone Excess

Oversecretion of GH, usually due to a pituitary tumor, leads to excessive growth. Where this occurs before puberty it can lead to giantism, though this is rare. Where it occurs after puberty (when further growth of the long bones is not possible) it can lead to acromegaly, in which those bones which can still grow in response to GH do so, leading to enlarged hands and feet and distorted facial features. The very high circulating levels of GH and IGF-1 seen in acromegalic patients also lead to thickening of the skin, enlargement of some soft tissues, and often

disorders of the cardiovascular, muscular, and endocrine systems. Diabetes mellitus is relatively common in acromegaly, and shows clinical features rather different from normal, presumably because GH itself has diabetogenic properties.

Acromegaly can be treated by surgical removal of the pituitary tumor, radiation therapy, administration of a drug (e.g., bromocriptine or a somatostatin analog) that inhibits GH secretion and tumor growth and/or use of a GH receptor antagonist (pegvisomant). Very rarely acromegaly is caused by a GHRH-secreting tumor elsewhere in the body, such as the pancreas.

Growth Hormone and Malnutrition; Catch-up Growth

The role of GH in malnutrition and starvation is complex and poorly understood. In humans, GH levels increase on fasting, in both normal and obese subjects, and fall again on refeeding. Paradoxically, IGF-1 levels fall in response to starvation, despite the high GH levels. Such a dissociation of GH and IGF-1 levels is particularly marked in protein-calorie malnutrition (kwashiorkor), where IGF levels may be in the hypopituitary level despite elevated GH levels, but it is also seen in conditions such as anorexia. The mechanism by which the dissociation of GH and IGF-1 levels arises is poorly understood, as is its significance for the control of nutritional balance during malnutrition, but it may effectively allow a dissociation of the actions of the hormone on growth promotion (much reduced during periods of malnutrition) from those on metabolism (still required since the hormone promotes utilization of fat). GH responses to starvation in the rat differ from those seen in humans – both GH and IGF-1 levels fall; responses in ruminants and pig are similar to those in humans.

In children, if nutrition improves after a period of malnutrition, the phenomenon of catch-up growth can occur, in which growth increases to an above normal rate, allowing some or all of the lost ground to be restored. Involvement of the GH/IGF axis in this phenomenon is likely, but the mechanism is unclear.

Nutritional Control of IGF Levels

Although GH clearly plays an essential part in the control of circulating and local IGF-1 levels, it is by no means the only factor. Insulin and thyroxine are important, and nutritional level also regulates both the level and activity of IGF-1. Indeed, IGF-1 levels in the circulation are remarkably sensitive to nutrients, and can provide a marker for nutritional status and effectiveness of treatment. Nutritional influence on IGF-1 levels may be mediated to some extent by insulin or thyroid hormones (but not GH, given the divorce between GH and IGF-1 levels in malnutrition, referred to above), but it is clear that to some extent they are direct. Actions on GH receptor levels, which are lowered on fasting, may be involved in severe dietary restriction, but in less severe protein restriction postreceptor mechanisms appear to predominate. Nutritional level also affects the activity of circulating IGF-1: (1) by altering the balance of the IGF-binding proteins that modulate the half-life, tissue availability, and receptor-binding of IGF-1; (2) by regulating formation of inhibitors of IGF-1; and (3) by altering type 1 IGF receptor levels. The involvement of so many factors in regulating IGF levels and activity probably plays an important part in integrating their various influences on growth.

Growth Hormone and Animal Biotechnology

GH can stimulate growth, meat production, and lactation in farm animals and the availability of large quantities of the hormone produced by recombinant DNA methodology has stimulated work designed to assess and apply the hormone in animal agriculture. Recombinant DNA-derived bovine GH (frequently referred to as bovine somatotropin, BST) has been subjected to extensive trials as an agent for promoting milk production, and is now being used commercially for this purpose in some countries. The ability of GH to promote meat production in various farm animals has been studied extensively. Not only is muscle growth stimulated, but fat content is much reduced, so that a potentially healthier product results. Possible applications of GH in fish farming have also been studied.

The potential for the use of recombinant DNA-derived GH in agriculture is great, but such usage remains controversial because of concerns about food safety, animal welfare, and economic impact. Indirect ways of modulating GH levels and actions in farm animals may prove more acceptable than direct administration of the hormone. Such indirect approaches include immunological modulation, modification of endogeneous hypothalamic activity, and use of the transgenic approach to introduce additional genes for GH into animals. The transgenic approach, coupled with cloning, is particularly powerful, though again welfare and other considerations may limit its usefulness.

See also: **Biotechnology in Food Production**; **Famine, Starvation, and Fasting**; **Growth and Development**; **Malnutrition**: The Problem of Malnutrition; **Renal Function and Disorders**: Kidney: Structure and Function

Further Reading

Boersma B and Wit JM (1997) Catch-up growth. *Endocrine Reviews* 18: 646–661.

Bole-Feysot C, Goffin V, Edery M, Binart N and Kelly PA (1998) Prolactin (PRL) and its receptor: actions, signal transduction pathways and phenotypes observed in PRL receptor knockout mice. *Endocrine Reviews* 19: 225–268.

Cone RD (1999) The central melanocortin system and energy homeostasis. *Trends in Endocrinology and Metabolism* 10: 211–216.

Drake WM, Parkinson C, Besser GM and Trainer PJ (2001) Clinical use of a growth hormone receptor antagonist in the treatment of acromegaly. *Trends in Endocrinology and Metabolism* 12: 408–413.

Grossmann M, Weintraub BD and Szkudlinski MW (1997) Novel insights into the molecular mechanisms of human thyrotropin action: structural, physiological, and therapeutic implications for the glycoprotein hormone family. *Endocrine Reviews* 18: 476–501.

Herrington J and Carter-Su C (2001) Signaling pathways activated by the growth hormone receptor. *Trends in Endocrinology and Metabolism* 12: 252–257.

Hwa V, Oh Y and Rosenfeld RG (1999) The insulin-like growth factor-binding protein (IGFBP) superfamily. *Endocrine Reviews* 20: 761–787.

Jordan SA and Jackson IJ (1998) Melanocortin receptors and antagonists regulate pigmentation and body weight. *Bioessays* 20: 603–606.

Kajima M, Hosoda H, Matsuo H and Kangawa K (2001) Ghrelin: discovery of the natural endogenous ligand for the growth hormone secretagogue receptor. *Trends in Endocrinology and Metabolism* 12: 118–122.

Lapthorn AJ, Harris DC, Littlejohn A *et al*. (1994) Crystal structure of human chorionic gonadotropin. *Nature* 369: 455–461.

Mohr E, Meyerhof W and Richter D (1995) Vasopressin and oxytocin: molecular biology and evolution of the peptide hormones and their receptors. *Vitamins and Hormones* 51: 235–266.

Norman AW and Litwack GD (1997) *Hormones*. New York: Academic Press.

Ohlsson C, Bengtsson BA, Isaksson OGP, Andreassen TT and Slootweg MC (1998) Growth hormone and bone. *Endocrine Reviews* 19: 55–79.

Pell JM and Bates PC (1990) The nutritional regulation of growth hormone action. *Nutrition Research Reviews* 3: 163–192.

Thissen J-P, Ketelslegers J-M and Underwood LE (1994) Nutritional regulation of the insulin-like growth factors. *Endocrine Reviews* 15: 80–101.

Wallis M (1988) The molecular basis of growth hormone deficiency. *Molecular Aspects of Medicine* 10: 429–509.

Wallis M, Howell SL and Taylor KW (1985) *Biochemistry of the Polypeptide Hormones*. Chichester: John Wiley.

Wilson JD, Foster DW, Kronenberg HM and Larsen PR (eds) (1998) *Williams Textbook of Endocrinology*, 9th edn. Philadelphia: Saunders.

Steroid Hormones

A W Norman, University of California-Riverside, Riverside, CA, USA

Introduction

Steroid hormones are well known both physiologically and clinically as regulators of diverse biological responses, including profound effects on cellular metabolism, development, and physiology. The steroid hormones are the estrogens (female sex steroids), androgens (male sex steroids), progestins, mineralocorticoids, glucocorticoids, and vitamin D with its daughter metabolites.

All these different steroid hormones are synthesized from the common precursor, cholesterol, and structurally differ only in the pattern of chemical bonds within the rings and modifications on the side chain. The exquisite specificity of physiological effects that these steroid hormones evoke is mediated by high-affinity intracellular receptor proteins that are exclusively localized in the specific target tissues for each steroid hormone. Specific interaction of the hormone–receptor complex with DNA sequences of the hormone-responsive gene(s) results in the tissue-specific expression of proteins which either directly or indirectly generate the biological responses attributable to the steroid hormones. The application of cellular and molecular biological techniques has allowed greater understanding of the way in which such small molecules exert diverse biological effects with exquisite specificity.

Historical Perspective

The first steroid hormone that was discovered, estrone, was isolated in 1929, before the characterization of the ring structure of the steroid nucleus. Studies by Professor A. Windaus in Göttingen, Germany, in the 1930s led to the chemical characterization and structural determination of cholesterol, using classical organic chemistry manipulations. Methodological advances such as radioisotopes, chromatography, mass spectrophotometry, and nuclear magnetic resonance spectroscopy have all facilitated the elucidation of structures of the other steroid hormones.

Chemistry

In mammalian systems, there are six families of steroid hormones that can be classified on both a structural and a biological (hormonal) basis. These are

the: (1) mineralocorticoids, which instruct the renal tubule to retain sodium; (2) glucocorticoids, which exert manifold effects on carbohydrate metabolism; (3) progestins, which are essential for reproduction; (4) estrogens, which induce female secondary sexual characteristics; (5) androgens, which induce male secondary sexual characteristics; and (6) the vitamin-D hormone, 1,25-dihydroxyvitamin D_3 (1,25-$(OH)_2D_3$), which is important for regulation of calcium and phosphorus homeostasis, bone growth, and development. Also, the bile acids are structurally related to cholesterol and thus constitute a seventh member of the steroid family. All of these steroids are biologically derived from cholesterol.

All steroids are derived from a cyclopentanoperhydrophenanthrene backbone that comprises three C_6 cyclohexane rings designated as A, B, and C, fused to a fourth C_5 cyclopentane ring, notated as the D ring. Within the family of steroid hormones, each member is distinguished by chemical alterations such as isomerization, aromatization, dehydrogenation, and fission in the ring structure, and modifications such as hydroxylations in the side chains of the cholesterol backbone that are catalyzed by specific enzymes present in the endocrine glands. The chemical formulae of these steroid hormones are depicted in Figure 1.

Biosynthesis of Steroids

Figure 2 summarizes in general terms the metabolic pathways leading from cholesterol to the six major steroid hormones (Table 1). The principal tissues of synthesis of the five classical steroid hormones (estrogens, androgens, progestins, glucocorticoids, and mineralocorticoids) are the adrenal cortex, ovaries, and testes. Also, during pregnancy, the fetal–placental unit can serve as a source of estrogen and some other hormones. The sites of synthesis of the sixth class of steroid hormones, the prohormone vitamin D_3 and its metabolites, are the skin, liver, and kidney. The bile acids, which are the seventh important structural class of mammalian steroids, have no known hormonal activity; they are principally synthesized in the liver and are secreted from the gallbladder to participate in digestion and absorption where they function as a detergent to solubilize fats.

Figure 3 describes the relationship between the stimulatory peptide hormone and the six classes of steroid hormones. The biosynthesis of each steroid hormone by its endocrine gland is stimulated by a specific cognate peptide hormone. Thus, for example, the biosynthesis of cortisol in the adrenal cortex is only stimulated by its tropic (stimulatory) peptide hormone adrenocorticotropic hormone (ACTH). ACTH is only secreted by the anterior pituitary gland.

Biosynthesis of Progestins

As evident from Figure 2, the common pathway leading to the production of six classes of steroid hormones primarily involves conversion of cholesterol (which possesses 27 carbon atoms) into pregnenolone and then progesterone (which possesses 21 carbon atoms). Progesterone is synthesized by the corpus luteum of the ovary and placenta and constitutes the chief progestational steroid in humans.

Biosynthesis of Glucocorticoids and Mineralocorticoids

The glucocorticoids and mineralocorticoids are 21-carbon corticosteroids produced by the adrenal cortex (Figure 1). The principal glucocorticoid in males is cortisol, while aldosterone is the main mineralocorticoid.

Biosynthesis of Androgens and Estrogens

The androgens, which possess 19 carbon atoms, are produced in the testes in males, and in the ovaries and placenta in females. Under some circumstances the adrenal cortex can also produce physiologically significant androgens such as androsterone, 4-androstene-3,17-dione, and dehydroepiandrosterone. The major naturally occurring steroids with androgenic activity (in decreasing order of relative potency) are 5α-dihydrotestosterone (5α-DHT; 150–200%), testosterone (100%), androstanediol (65%), androst-4-ene-3,17-dione (25%), androsterone (10%), and dehydroepiandrosterone (10%). The hormonally active form of testosterone in males is 5α-DHT, which is chiefly produced in the prostate gland, although skin, testis, and submaxillary glands also can produce small quantities.

The estrogens, which possess 18 carbon atoms, are produced in the ovarian follicle, corpus luteum, and fetal–placental unit. In both males and females, the adrenal cortex can generate small quantities of estrone from androst-4-ene-3,17-dione. The major naturally occurring steroids with estrogenic activity are estra-3,17β-diol, estra-3,16α, 17β-triol, and estrone.

Biosynthesis of Vitamin D Metabolites

Chemically, the vitamin D steroids are secosteroids because of the breakage of the B-ring, leaving the A-, C-, and D-rings intact (Figure 1). The parent vitamin D is produced in the skin from 7-dehydrocholesterol as a consequence of ultraviolet or sunlight action, which breaks the 9,10 carbon–carbon bond. The sites of synthesis of the vitamin D metabolites are the liver and kidney. The hormonally active forms of vitamin D are $1,25(OH)_2D_3$ and 24,25-dihydroxyvitamin

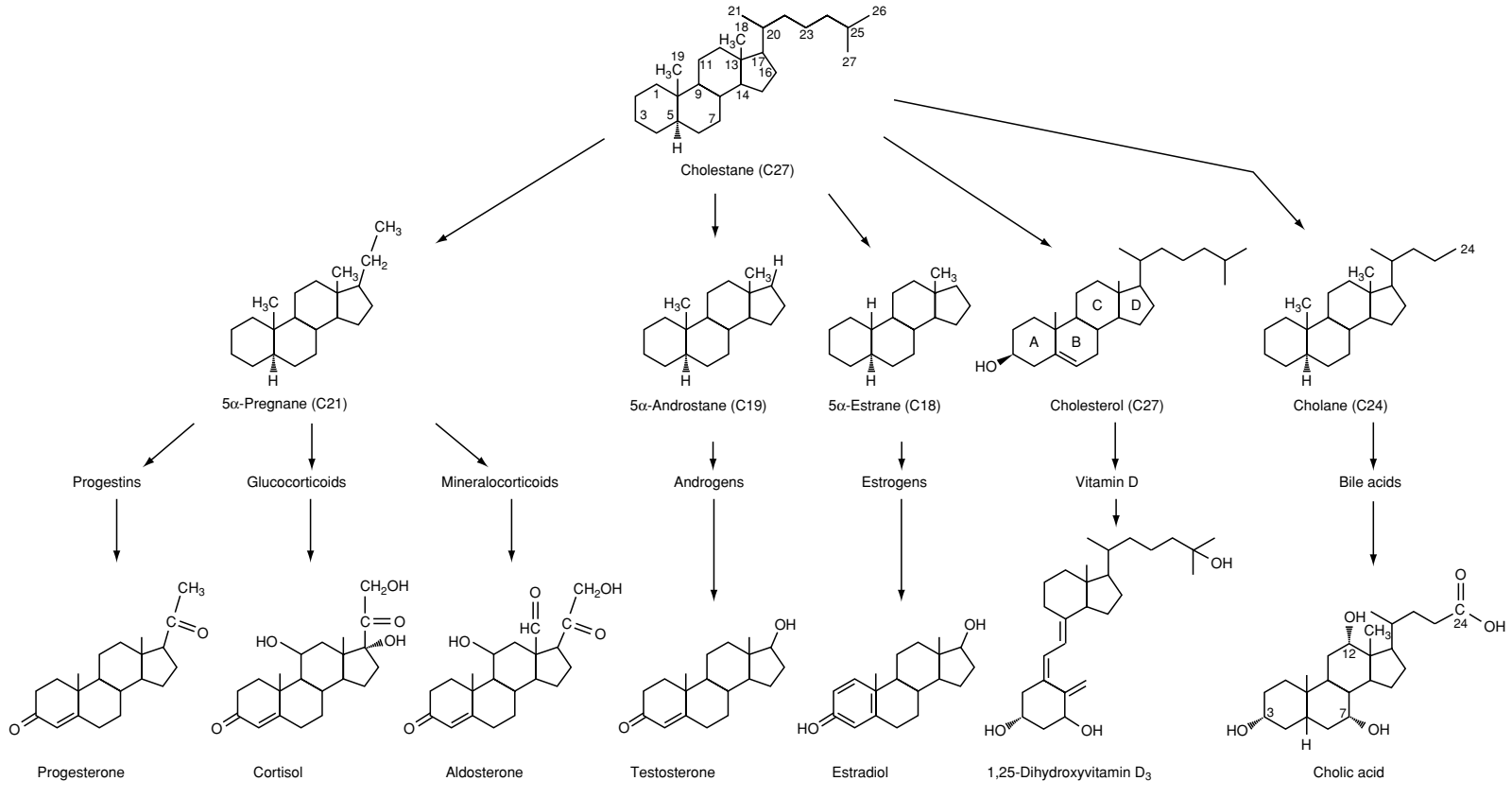

Figure 1 Family tree of the seven principal classes of steroids (bottom row); all of the steroids except cholic acids are steroid hormones. All of the steroids are structurally derived from cholestane (top row), which is closely related to cholesterol.

D_3 (24,25$(OH)_2D_3$), which are instrumental in generating a wide array of cellular responses. These aspects are detailed elsewhere in this volume.

Enzymes involved in Steroid Metabolism

Table 1 summarizes some general properties of the key enzymes involved with conversion of cholesterol into the six key steroid hormones. The various classes of enzymes that participate in the biosynthesis of different steroids include: (1) lyases, which cleave the carbon–carbon bonds bearing vicinal hydroxyl groups in the cholesterol side chain; (2) dehydrogenases and isomerases which convert the 5-ene-3β hydroxysteroid to 4-ene-3-oxosteroid; (3) aromatase enzyme (present in the microsomes of placenta and ovaries), which mediates the aromatization of the A-ring with a phenolic 3-hydroxyl group in the estrogens (Figure 1); and (4) hydroxylases, which play a key role either in the sequential hydroxylation of C22, C20, C17 of the cholesterol side chain or in the stereospecific hydroxylation of the ring structure to yield the different steroids (Figure 1). Thus, for example, the three hydroxylases 17α-, 21-, and 11β-hydroxylase are required for cortisol synthesis, and the 1α and 24-hydroxylases are involved in the production of bioactive vitamin-D metabolites.

All these apparently different hydroxylation reactions are mediated by a family of enzymes called

Table 1 Key enzymes concerned with production of steroid hormones

Enzyme reaction[1]	*Enzyme*		*Reaction(s) mediated*	*Subcellular location*[2]	*Gene location*
	Long name	*Short name*			
①	P450 side-chain cleavage; formerly known as desmolase	(P450scc)	20R-hydroxylation; 22R-hydroxylation; scission of C20–C22 carbon bond	M	15
②	3β-Steroid dehydrogenase-Δ^5, Δ^4-isomerase	(Isom)	Oxidizes 3β-OH to oxo group; isomerizes Δ^5 ene to Δ^4 ene	ER	
③	21-Hydroxylase	(P450c21)	21-hydroxylation	ER	6p
④ ⑤ ⑥	11β-Hydroxylase	(P450c11)[3]	④ 11β-hydroxylation; ⑤ 18-hydroxylation; ⑥ oxidation of C18 hydroxyl to an aldehyde	M	8q
⑦ ⑧	17α-Hydroxylase/ C17–C20 lyase	(P450c17)[4]	⑦ 17α-hydroxylation; ⑧ scission of C17–C20 carbon bond	ER	10
⑨	17-Ketosteroid reductase	(17-oxido reductase)	Reduces 17-oxo to a 17β-hydroxyl as in androstenedione→testosterone; estrone→estradiol; DHEA→androstenediol	ER	
⑩	Aromatase	(P450aro)	Mediates the aromitization of the A-ring	ER	15q 21.1
11	Vitamin D25-hydroxylase	(P450c25)	25-hydroxylation of vitamin D secosteroids	M, ER	
12	25$(OH)D_3$-1α-hydroxylase	(P450c1)	1α-hydroxylation of vitamin D secosteroids	M	
13	25$(OH)D_3$-24-hydroxylase	(P450c24)	24R-hydroxylation of vitamin D secosteroids	M	
14	5α-Reductase		Reduction of Δ^5-ene as in testosterone→dihydrotestosterone	NM	
15	Ferredoxin-oxidoreductase		In the mitochondria transfers electrons from NADPH to ferredoxin	M	17 con→q25
16	P450 reductase		In the endoplasmic reticulum transfers electrons from NADPH to a P450 enzyme	ER	
17	11β-Hydroxy steroid dehydrogenase		Oxidation of 11β-hydroxyl to an oxo as in cortisol→cortisone	ER	
18	12α-Hydroxylase		12α-Hydroxylation of bile acids	ER	
19	6β-Hydroxylase		6β-Hydroxylation of bile acids	ER	
20	7α-Hydroxylase		7α-Hydroxylation of bile acids	ER	

[1]This table summarizes the key enzyme reactions in steroid hormone production (see numbers in circles). Note that, in some instances, the same enzyme catalyzes more than one reaction. (See footnotes 3 and 4.)
[2]Subcellular location coded as M, mitochondria; ER, endoplasmic reticulum; NM, nuclear membrane.
[3]Reactions ④, ⑤, and ⑥ can be mediated by the same enzyme.
[4]Reactions ⑦ and ⑧ are mediated by the same enzyme.
DHEA, dehydroepiandrosterone; NADPH, reduced form of nicotinamide adenine dinucleotide phosphate.

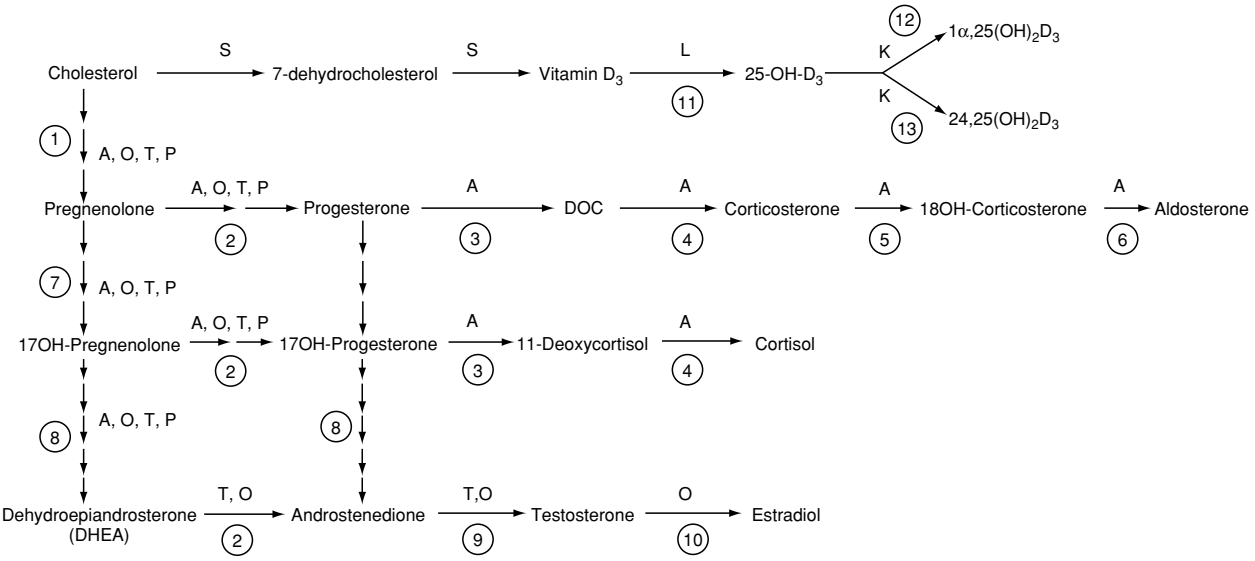

Figure 2 Principal pathways of human steroid hormone synthesis. The details of the enzyme reactions (circled numbers) are summarized in Table 1. Reactions ①, ②, ⑦, and ⑧ are found universally in the adrenals, ovary, testes, and placenta. Reactions ⑨ and ⑩ are found principally in the testes and ovaries. Reactions ⑪ → ⑬ describe the pathway of production from cholesterol of the prohormone vitamin D and its hormonal metabolite products 1,25-dihydroxyvitamin D_3 ($1\alpha,25(OH)_2D_3$) and 24,25-dihydroxyvitamin D_3 ($24,25(OH)_2D_3$). Reaction ⑪ is P450c25 present in the mitochondrial and/or endoplasmic reticulum of the liver. Reactions ⑫ and ⑬: P450c1α and P450c24 are present in the mitochondria of the kidney proximal tubule. Tissue codes: A, adrenals; O, ovaries; T, testes; P, placenta; S, skin; L, liver; K, kidney; DHEA, dehydroepiandrosterone; DOC, deoxycorticosterone.

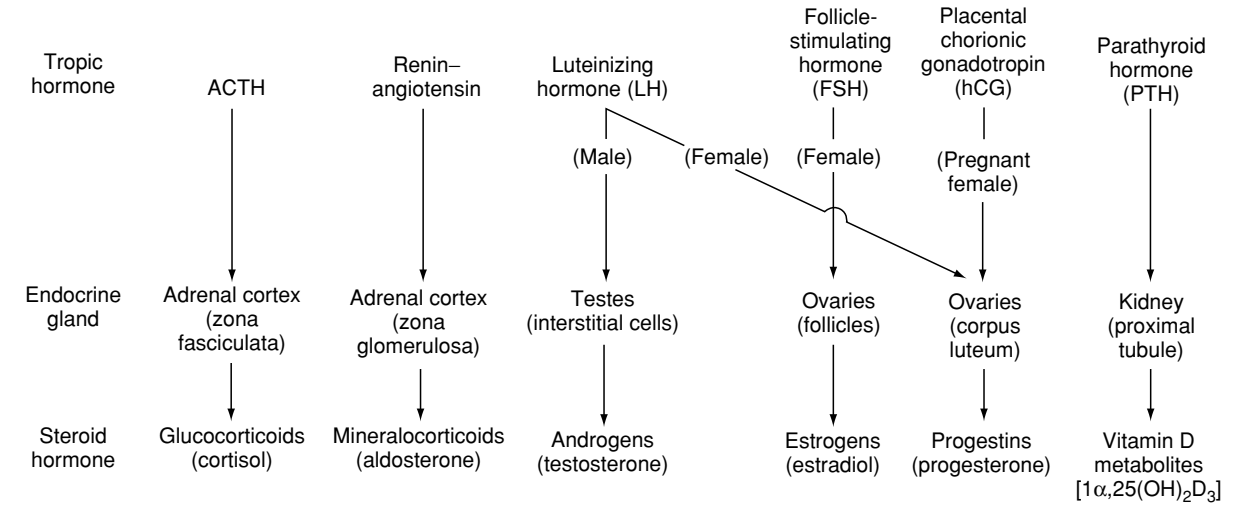

Figure 3 Relationship between stimulatory peptide hormones and the production of steroid hormones. For each of the six classes of steroid hormones, the specific stimulatory or tropic peptide hormone is indicated as well as specification of the organ or cell upon which it acts. Also, in males, follicle-stimulating hormone is known to act upon the Sertoli cells of the seminiferous tubules to increase the production not of a steroid hormone but of several sperm proteins, androgen-binding protein, and inhibin, a peptide hormone. ACTH, adrenocorticotropic hormone.

cytochrome P450 hydroxylases which have a cytochrome moiety at the catalytic center. The cytochrome P450 enzymes are present in either the mitrochondrial or microsomal fraction of cells present in the liver, adrenal cortex, ovary, testis, kidney, placenta, lungs, and intestinal mucosa. The mitochondrial cytochrome P450s are comprised of a flavoprotein dehydrogenase, which accepts electrons from the reduced form of nicotinamide adenine dinucleotide phosphate (NADPH), and the nonheme

iron protein adrenodoxin, which accepts electrons from flavoprotein and transfers them to the cytochrome P450 protein. The microsomal hydroxylase system lacks the adrenodoxin component. Cytochrome P450 hydroxylases (also known as mixed-function oxidases) catalyze oxidation-reduction reactions in which one atom of the substrate (O_2) ends up on the steroid as a part of the hydroxyl and the other oxygen atom as part of water (H_2O); the necessary reducing equivalents are derived from NADPH.

Transport

All steroid hormones, which are quite water-insoluble, circulate in blood, bound to plasma proteins called steroid hormone-binding globulins. The androgens and estrogens are both transported by a plasma β-globulin, which is a hepatic glycoprotein with a molecular weight of 94 000. This carrier protein has a high affinity (K_d 1.5×10^{-10} mol l^{-1}) for testosterone, 5α-DHT, and estradiol but not for progesterone or cortisol. The glucocorticoids are systemically transported via the corticosteroid-binding protein (CSBP) or transcortin, which is also a hepatic protein. Similarly, the vitamin D sterols are bound to the vitamin D-binding protein (DBP), and α-globulin. The primary role of all these different sterol-binding proteins is to provide a reservoir of bound hormone which can effectively dampen oscillations in the 'free' concentrations of hormones.

Catabolism and Excretion of Steroid Hormones

The catabolic pathways that result in the inactivation of the steroid hormones occur mainly in the liver. Conjugation of the hydrophobic steroids either with sulfate or glucuronides renders the hormone more hydrophilic for easy excretion in the urine. The excreted species are usually polyhydroxyl forms with the particular glucuronide or sulfate: these conjugations are mediated by sulfokinase and glucuronyl transferase enzymes.

Functions of Steroid Hormones

Steroid hormones secreted by the endocrine glands exert profound influences on cell growth, development, and physiology, chiefly by controlling the synthesis of particular proteins. All steroid hormones modulate cellular responses by either inducing or repressing the transcription of specific proteins at the genetic level.

A general result of glucocorticoid action is a direct effect of increased glucogen, especially in the liver, and increases in the levels of circulating glucose. Prolonged high levels of glucocorticoids can lead to the death of susceptible cells which accounts for muscle wasting and immunodeficiency. As a powerful antiinflammatory agent, glucocorticoids may act on many cells to induce the synthesis of the protein(s) lipomodulin. Lipomodulin acts as an inhibitor of phospholipase A_2 and thereby prevents the release of prostaglandins, prostacyclins, and leukotrienes, some of which mediate processes of inflammation and pain.

Mineralocorticoids such as aldosterone promote sodium reabsorption by the kidney tubules, usually followed by changes in water balance; in terms of function they mesh with the vasopressin system and atrial natriuretic factor which also control water absorption and excretion by the kidney and also participate in the regulation of blood pressure.

The biological responses of androgens are divisible into four categories: (1) growth-promoting or androgenic effect on the male reproductive tract; (2) stimulating or anabolic effect on body weight (skeletal growth, skeletal muscle growth, subcutaneous fat distribution, growth of prostate and seminal vesicle); (3) development of male secondary sexual characteristics (growth of external genitalia, enlargement of larynx and thickening of vocal cords, hair growth distribution); and (4) actions in the central nervous system and brain (differentiation of brain cortex, hypothalamus, initiation of puberty).

The dominant effect of estrogens occurs in the female reproductive tract. Significant biological activity of estrogens also occurs in the hypothalamus, brain, and mammary tissue. Estrogens primarily initiate the transcription of specific proteins in the uterus, and affect uterine cell proliferation and differentiation. A classical developmental role of estrogen in the liver of the adult hen is in the induction of vitellogenin, which is transported to the ovary and cleaved to yield egg proteins, phosvitin, and lipovitellin.

The biological activity of progesterone is restricted to the female reproductive tract and mammary tissue. Some of the effects mediated by progesterone include thermogenesis in women, regulation of egg movement through the fallopian tubes, preparation of uterus to receive the blastocyst, alteration of electrical activity in the brain, control of uterine contraction (at parturition), and generation of the secretory system of breast during pregnancy.

The steroid hormone $1,25(OH)_2D_3$ is principally involved in the regulation of mineral (calcium and phosphorus) metabolism, bone growth, and differentiation. A detailed account of the manifold effects of vitamin D is given in a subsequent article. (*See* **Cholecalciferol**: Properties and Determination; Physiology.)

Mechanism of Action of Steroid Hormones

The receptors for all the steroid hormones, $1\alpha,25(OH)_2D_3$, thyroid hormone, and retinoic acid, are all proteins which exist exclusively in the cell nucleus or are partitioned between the cytoplasm and nucleus. Each of these receptor proteins functions as DNA-binding proteins, where they regulate the expression of genes related to the biological response of the hormone in question.

Figure 4 presents a generic model of a typical steroid receptor. The protein consists of a single polypeptide chain divided into six domains. Starting at the N-terminus, the domains are respectively (1, 2) the A/B or variable (sequence and length) domain; (3) the C or DNA-binding domain which contains two zinc fingers; (4) the D or variable hinge region; (5) the E or ligand-binding domain; and (6) the F or variable domain.

Figure 5 illustrates the striking structural homology between all the domains of the steroid superfamily. All these proteins are related evolutionarily. The ancestor of these proteins may be the oncogene product v-*erb* A.

The steroid receptors are labile soluble proteins with molecular weight ranging between 50 and 94 kDa. The molecular cloning of steroid receptors has greatly facilitated the delineation of the various functional domains of these receptor proteins (Figures 4 and 5). All members of steroid receptors bear a short DNA-binding domain comprising about 70 amino acids containing many conserved cysteine residues. Eight of the cysteines are organized into two so-called zinc fingers, each containing four cysteine residues tetrahedrally coordinating a zinc ion. Based on the amino acid sequence within the DNA-binding region of steroid receptors, they can be classified into two subfamilies: (1) glucocorticoid receptor is the prototype of the smaller subfamily which includes progesterone, androgens, and mineralocorticoid receptors; (2) the estrogen receptor is the prototype of the larger subfamily which includes the vitamin D receptors, thyroid hormone receptors, and receptors for retinoic acid. Steroid receptors exhibit a large C-terminal domain, responsible for binding hormone, known as hormone-binding domain (HBD). The amino-terminal region contains hypervariable amino acids and certain transactivation sequences. Recent studies have shown that functional properties of steroid receptors are modulated by phosphorylation, dimerization, and association with other proteins such as heat shock proteins (HSPs).

A model of the steroid hormone mode of action is provided in Figure 6. In step 1, the hormone (H) dissociates from the plasma transport protein. In step 2, the hormone enters the target cell by diffusing through the outer cell membrane. Then, depending upon the subcellular localization of the unoccupied receptor, the hormone will either interact with the receptor in the cytoplasmic compartment (glucocorticoid and aldosterone receptors) or continue through

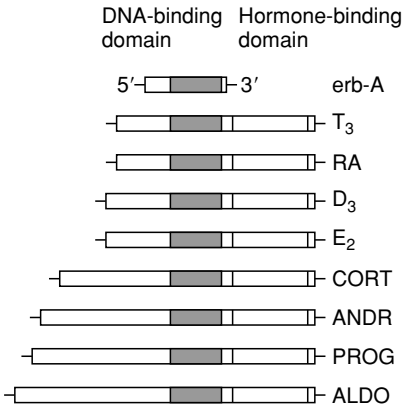

Figure 5 Steroid receptor gene superfamily. T_3, triidothyronine; RA, retinoic acid; D_3, 1,25-dihydroxyvitamin D_3, E_2, estradiol; CORT, cortisol; ANDR, androgen; PROG, progesterone; ALDO, aldosterone. The N-terminal domain (open rectangle to the left of the DNA-binding domain) represents the sequences of least homology among the receptors.

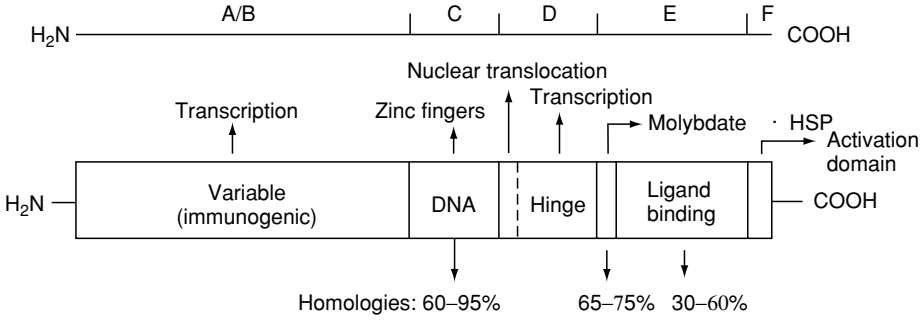

Figure 4 Model of a typical steroid hormone receptor illustrating the six functional domains A → F. HSP, heat shock protein.

the cytoplasm and cross the perinuclear membrane to interact with a receptor in the nucleus (thyroid, estrogen, progesterone, retinoic acid, and $1,25(OH)_2D_3$ receptors).

The unoccupied form of the steroid receptor is believed to exist in the cell as a conjugate with a dimer of the 90-kDa HSP; in Figure 6 the HSP dimer is indicated by a pair of ovals. It has been suggested that the HSP's function is to occlude the DNA-binding domain of the receptor which prevents the unoccupied receptor from binding to DNA.

In steps 3 and 4, the receptor becomes transformed or activated as a consequence of release of the 90-kDa HSP and binding of its cognate ligand, so that the DNA-binding domain is now exposed. In step 5, the cytoplasmic activated receptor translocates to the nucleus, probably through a nucleopore. Then in step 6, the activated receptor, or more likely a homodimer of the receptor, seeks out the correct sequence of DNA that will allow it to form a high-affinity complex between the receptor and the hormone response elements (HRE) of the promoters of a selected set of genes and also any required transcription factors.

As a consequence of the activated receptor binding to the promoters, either induction or repression of that gene will occur, leading to more or less of the mRNA coded for by that gene.

The newly transcribed mRNAs are translocated to the cytoplasm where they become incorporated into polysomes and undergo translation (step 7). In the final step, step 8, the increased or decreased amount of new proteins generates more or less of the biological response(s) dictated by that hormone in that target cell. For any given hormone, a wide array of biological responses can be modulated depending upon the phenotype of the target cell that possesses the cognate receptor. In any given target cell phenotype, only a small subset of genes will have their DNA chromatin in an active or open configuration. Thus, while a hormone may modulate as many as 300 genes in a given organism, in a specific target cell perhaps only a few genes will be available for regulation.

Summary

Steroid hormones impact profoundly on cellular metabolism, development, and physiology. The

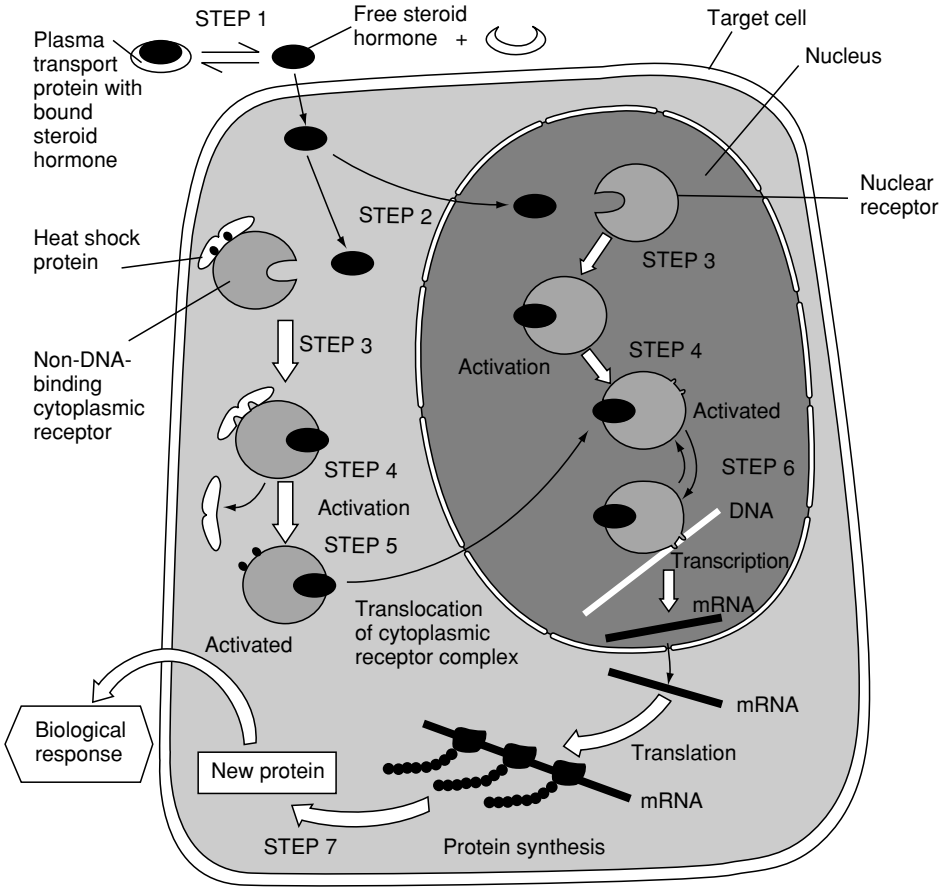

Figure 6 Model of steroid hormone mode of action. A detailed description is given in the text.

repertoire of cellular responses elicited by the various steroid hormones is mediated by specific intracellular receptor proteins localized in target tissues. Cloning of the receptor genes has allowed detailed structure–function analysis of these physiologically important molecules. Greater insight and appreciation of the complex molecular details of steroid hormone action should be possible with the application of modern cellular and molecular biological techniques.

See also: **Cholecalciferol**: Properties and Determination; Physiology

Further Reading

Bouillon R, Okamura WH and Norman AW (1995) Structure–function relationships in the vitamin D endocrine system. *Endocrinology Review* 16: 200–257.

Chambon P (1996) A decade of molecular biology of retinoic acid receptors. *FASEB Journal* 10: 940–954.

Darimont BD, Wagner RL, Apriletti JW *et al.* (1998) Structure and specificity of nuclear receptor–coactivator interactions. *Genes and Development* 12: 3343–3356.

Jenster G, Spencer TE, Burcin MM *et al.* (1997) Steroid receptor induction of gene transcription: a two-step model. *Proceedings of the National Academy of Science of the USA* 94: 7879–7884.

Mangelsdorf DJ, Thummel C, Beato M *et al.* (1995) The nuclear receptor superfamily: the second decade. *Cell* 83: 835–839.

Norman AW and Litwack G (1997) *Hormones*, 2nd edn. Florida: Academic Press.

Weatherman RV, Fletterick RJ and Scanlon TS (1999) Nuclear receptor ligands and ligand-binding domains. *Annual Review of Biochemistry* 68: 559–582.

HORSE MEAT

E Rossier, Formerly of CEREOPA, Paris, France

Introduction

The custom of eating horse meat is probably as old as the world itself and, without venturing into prehistoric times, the majority of the nations of Antiquity (Persians, Greeks, Romans, Chinese, Cossacks, etc.) all ate horse meat.

Paradoxically, in European countries Christianity has strived to abolish the eating of horse meat since the seventh century AD. This ban was based on religious beliefs and was only dispelled after 1870. The eating of horse meat then developed quickly thanks to the vigorous action of several well-known personalities in scientific and, in particular, medical circles.

This article explains the situation of the world market as well as the nutritional and hygienic qualities of horse meat.

The World Horse Meat Market

Horse meat is produced in several countries, although the level of production is often marginal when compared to other types of meat.

The USA is the major producer. The other principal countries are Mexico, Argentina, Canada, China, Mongolia, Ethiopia, Chile, and Brazil, as well as several European countries – Poland, Yugoslavia, France, and Italy. The Commonwealth of Independent States (former Soviet Union) is also a large producer although few data are available.

The majority of these countries are exporters of carcass meat, semicarcass meat, quartered or boned meat, or else of live animals destined for slaughter in the importing country. In practice, a few countries dominate trading, the USA being the lead producer, followed by Argentina and Canada.

There are also a few main importing countries and in these consumption exceeds production which has diminished as a result of mechanization of agriculture and transport. This is particularly so in France, Japan, Italy, Belgium, and the Netherlands. The other markets are much more limited: Switzerland, Austria, Sweden, Norway, and Denmark.

The internal production of horse meat is adequate for domestic consumption in Mongolia, China, Mexico, Ethiopia, and the Commonwealth of Independent States.

In many countries there is a considerable horse population, but little information is available regarding the final use to which the animals are put, except in the production of dog and cat food.

Except for the countries mentioned above, Italy, France, Japan, Belgium, the Netherlands, Spain, Switzerland, and Sweden are the principal consumers (**Table 1**). Only small proportions of these

Table 1 National gross production and total consumption of horse meats in the main consumer countries in 1988

Country	*National gross production ($\times 10^3$ t)*	*Total consumption ($\times 10^3$ t)*
Belgium and Luxembourg	3	28
Denmark	1	1
France	15	60
Germany	5	5
Italy	14	68
The Netherlands	4	21
Spain	7	7
Japan	4	59
Austria	1	2
Finland	1	1
Norway	1	1
Sweden	2	3
Switzerland	1	4
Yugoslavia	9	1

Source: Organisation for Economic Cooperation and Development (1990) *Meat Statistics*, with permission.

populations eat the meat. This attitude might change significantly if the meat were promoted with a specific image in view of the qualities mentioned below.

There are several restraints on consumption of this type of meat, in particular cultural or religious prohibitions, the 'noble animal' symbolism of horses, and the often specific distribution channels of horse butchers, although more widespread distribution is gradually taking over.

Butcher Qualities

Carcass Characteristics

As horses were not intended to produce meat for consumption, unlike other species, the meat that has been eaten has come mainly from grazing animals, working horses and, more recently, sport and racing horses.

The growth in production of young meat, especially from the Trait breed, is new in several European countries and Japan. The horse is in fact a good meat producer. Carcass yields are high (real yield is equal to weight of warm carcass divided by live empty weight), bordering on 70% for young stallions born from the Trait breed and comparable with those for cattle meat.

The carcass contains a very high proportion of muscle (70%) and little fat (10–14%). The distribution of adipose tissue is not the same as in cattle, as horse carcasses contain more subcutaneous and internal adipose deposits. On the other hand, there is a smaller proportion of intermuscular fat. Horse carcasses therefore look more 'glazed,' but studies on the relative growth of different tissues, organs, and body regions have confirmed that the horse is a lean animal. (*See* **Meat**: Structure.)

Changes to the Meat after Slaughter

The physiological principles of postmortem change are identical for the horse as for other species. However, horse meat differs from cattle meat in its speed of maturation. Rigor mortis sets in rapidly (in several hours), and the pH quickly drops to 5 or 6, which limits microbial proliferation and allows autolysis of the muscle. (*See* **Meat**: Slaughter.)

Horse meat contains a large amount of glycogen and adenosine triphosphate (ATP) and therefore retains a degree of plasticity and elasticity for a long time.

Maturation occurs quickly in horse muscle. Most of the maturation is achieved 4–5 days after slaughter, allowing the meat to be eaten very fresh. As for preservation, several factors combine to give horse meat an excellent sanitary quality, including short storage period for maturation, and low pH. But the large proportion of glycogen, lactic acid, and nonprotein nitrogen encourages microbial development. The meat must therefore be sold quickly and stored at low temperature before being eaten.

Organoleptic Qualities

Tenderness

This criterion largely overshadows other quality parameters since the consumer's satisfaction when tasting the meat depends on this. (*See* **Sensory Evaluation**: Texture.)

It is the insoluble protein constituents of the meat which more or less determine its tenderness: the collagen of connective tissue and its degree of polymerization, myofibril proteins, and muscular fibers.

It seems that, with horses as with other species, there are important variations between animals, age, and different muscles of the same carcass.

These variations in toughness result from large differences in the level of collagen present and its fluctuation with the age of the animal. The collagen content is at its maximum at 18 months. The solubility of collagen, which is linked to its aging and hardening, seems to reduce rapidly until 24 months and then stabilizes thereafter. Thus insoluble collagen reduces the tenderness. Horse meat is generally considered to be very tender, however, and this quality of tenderness is one of the consumer's main reasons for purchase.

Color

It is important to consider the color of horse meat, in particular minced horse meat, which constitutes a high proportion of the value of horse carcass. The color is linked to the myoglobin content in the muscles. (*See* **Sensory Evaluation**: Appearance.)

Adult horse meat is very rich in myoglobin and iron and looks very red. This is a quality sought after by consumers in some countries (France), whereas others (Italy) prefer lighter-colored meats. The level of pigment varies according to the particular muscles and the age of the animal. The content of heme iron is still very low at 6 months, giving a very clear, 'brassy' meat. A strong color hardly appears before 18 months.

Nutritional Qualities

The structures – particularly of the muscular fibers – of horse meat are classical and identical to those of other species. On the other hand, the chemical composition of horse muscle is highly characteristic and conveys particularly interesting nutritional qualities.

Average Composition of Horse Meat

At the beginning of the twentieth century the average composition of horse meat was published as follows: water, 74.27%; nitrogenous substances, 21.71%; fats, 2.55%; nonnitrogenous substances, 0.46%; mineral materials, 1.01%. More recent writers give very similar figures, shown in Table 2. Globally, water content is higher in horse meat than in other species; a very low lipid content in horse meat also leads to a low energy value. The high level of water in horse meat can be a handicap for certain processes and can present technological problems for the salted horse meat trades.

Carbohydrates

Horse meat is rich in carbohydrates, especially glycogen. It is one of its main characteristics. It contains from 0.5 to 3 times more than cattle meat. This high level of glycogen gives horse meat a sweeter flavor than beef. (*See* **Carbohydrates**: Classification and Properties.)

Lipids

The main dietary quality of this meat is its low level of lipids. However, this meat is interesting not only in its low lipid content and its distribution of lipids throughout the carcass at different adipose deposits, but also in the structure of these fats. They are mainly comprised of triglycerides with a high proportion of unsaturated and polyunsaturated fatty acids: palmitoleic acid ($C_{16:1}$), oleic acid ($C_{18:1}$), linoleic acid ($C_{18:3}$) and arachidonic acid (C_{20}). (*See* **Fatty Acids**: Properties; **Fats**: Classification.)

Proteins

The proteins are the major constituents after water. They can be divided into extracellular proteins, e.g., collagen, and intracellular proteins. Horse meat is interesting not only in its proportion of proteins but also in the distribution of different amino acids. The few studies on the amino acid content of horse meat show a strong percentage of essential amino acids such as lysine, leucine, and isoleucine. (*See* **Amino Acids**: Properties and Occurrence; **Protein**: Chemistry.)

Minerals and Vitamins

Principal minerals As shown in Table 3, horse meat possesses a low sodium content (21 mg per 100 g on average) and is therefore suitable for those on low-sodium diets; this meat has a strong image of being 'good for the health.' The red color of the adult meat can darken and turn brown rather quickly as a result of oxidation, which makes the meat difficult to present and market. Refer to individual minerals.

Trace elements There are heavy metals such as cadmium, zinc, and lead in horse meat. Practically all heavy metals are present but in very small quantities, with no risk of toxicity. (*See* **Heavy Metal Toxicology**.)

Vitamins The few reported findings in the literature suggest the presence of few fat-soluble vitamins and

Table 2 Average composition of horse meat (per 100 g of consumable meat)

Composition	*Average value*
Water (g)	75
Protein (g)	23
Lipid (g)	2
Carbohydrate (g)	1
Energy (kJ)	462–504

Table 3 Mineral content of horse meat

Minerals	*Content (mg 100 g^{-1} fresh matter)*
Phosphorus	200
Calcium	4
Sodium	21
Potassium	176
Iron	2.7
Magnesium	4
Sulfur	10

more water-soluble vitamins in horse meat. (*See* **Vitamins**: Overview.)

Hygienic Qualities

Horse meat is reputed for its hygienic qualities as it is consumed very fresh. It is even common for minced horse meat to be eaten raw.

Parasitic and Microbial Contaminations

Horse meat is naturally not very contaminated. Although it is claimed to be safe, it must be fresh to be eaten raw or lightly cooked. In fact, its richness in glycogen can encourage the growth of microorganisms. Contamination often occurs at the surface and can easily spread, especially as the meat is often minced. A strict code of hygiene is necessary. (*See* **Contamination of Food**.)

Parasitic infections are rare but must be given special attention. Lesions of the digestive tract must be watched for as well as similar signs of the effects of parasites in both living and dead animals. (*See* **Parasites**: Occurrence and Detection.)

Bacteria such as *Salmonella* are the main agents of contamination of horse meat. They are easily destroyed by cooking heat, but as the meat is often consumed either without being fully cooked through or even raw, they can survive in the food. A very strict code of hygiene is therefore required when working with horse meat in the slaughterhouse and in the butcher's shop.

The transmission of the parasitic disease trichinosis through the consumption of horse meat was unknown until roughly 1975. The disease can be spread amongst carnivores (e.g., dogs, cats, foxes) rodents (e.g., badgers, rats) and omnivores (e.g., pigs, boars, bears, humans), and infestation in horses is quite unusual.

Processing

Horse meat lends itself to butchering and salting processes, either for typical horse meat products (e.g., sausages, saveloy, andouilles, smoked hams) or else in conjunction with other meats. This process has been developed in Belgium, Italy, and France as well as in Japan, where hams are reconstituted, often mixed with other meats. Also in Japan, horse meat is frequently incorporated into corned beef. (*See* **Meat**: Preservation.)

Conclusion

Horse meat shows many advantages in terms of processing and storage, and organoleptic, nutritional, and hygienic qualities.

Some of the special characteristics of this meat are carried over from generation to generation: it is the meat to treat anemia; it is meat suitable for tuberculosis diets; it is the meat for energy; these are special characteristics familiar to nutritional biochemists.

These characteristics, in the current context of competition between meats and other available sources of proteins, and as a result of changes in eating habits in recent years, make horse meat, at the very least, a most interesting meat with the following advantages: high butcher yields; high levels of glycogen and good-quality proteins (essential amino acids); low contents of salt and lipids; tenderness associated with collagen characteristics and their distribution; a characteristic sort of fatty tissue with a predominance of unsaturated fatty acids; safe in terms of residues or contamination provided that the conditions of slaughter, transport, storage, and distribution are satisfactory in order to reduce as much as possible the dangers of microbial infection.

Ultimately, horse meat not only possesses obvious dietary qualities for people in good health and particularly for children, but also is very suitable for those on special diets or everyday diets, e.g., diabetics, obese subjects, those with gout, hypercholesterolemia, hypertriglyceridemia, gastrectomy or kidney insufficiency, and those following a residue-free diet or a low-sodium diet.

See also: **Amino Acids**: Properties and Occurrence; **Carbohydrates**: Classification and Properties; **Contamination of Food**; **Fats**: Classification; **Fatty Acids**: Properties; **Heavy Metal Toxicology**; **Meat**: Structure; Slaughter; Preservation; **Parasites**: Occurrence and Detection; **Protein**: Chemistry; **Sensory Evaluation**: Appearance; Texture; **Vitamins**: Overview

Further Reading

Anderson G and Lee D (1976) Salmonella in horses: a source of contamination of horse meat in a packing plant under federal inspection. *Applied and Environmental Microbiology* 31(5): 661–663.

Catalano AL and Quarantelli A (1979) Carcass characteristics and chemical composition of the meat from milk-fed foals. *La Clinica Veterinaria* 102(6–7): 498–506.

Cattaneo P, Radaelli A and Cantoni C (1979) Solubilité des fractions azotées du muscle de cheval. *Archivio Veterinario Italiano* 30(1–2): 47–48.

Martin-Rosset W, Boccard R, Jussiaux M, Robelin J and Trillaud-Geyl C (1983) Croissance relative des différents tissus, organes et régions corporelles entre 12 et 30 mois chez le cheval de boucherie de différentes races lourdes. *Annales de Zootechnie* 33(2): 153–174.

Organisation for Economic Cooperation and Development (1991) *Meat Balances in OECD Countries*, p. 147. Paris, France: OECD Publications.

Otake Y and Nakazato T (1972) Fatty acid and triglyceride composition of horse meat lipids. *Japanese Journal of Zootechnical Science* 43(11): 631–637.

Renon P, Comi G and Cantoni C (1979) Acides gras du tissu adipeux de cheval. *Archivio Veterinario Italiano* 30 (1–2): 35–40.

Robb J, Harper RB, Mintz HF, *et al*. (1972) Chemical composition and energy value of the body, fatty acid composition of adipose tissue and liver and kidney size in the horse. *Animal Production* 14(1): 25–34.

Robelin J, Boccard R, Martin-Rosset W, Jussiaux M and Trillaud-Geyl C (1984) Caractéristiques des carcasses et qualités de la viande de cheval. In: Jarrige R and Martin-Rosset W (eds) *Le Cheval*, pp. 601–610, Paris: Institut National de la Recherche Agronomique.

Rossier E and Berger C (1988) *Horse Meat: Great but Unknown Qualities*, p. 37. Paris: Centre d'Etude et de Recherche sur l'Economie et l'Organisation des Productions Animales and Institut Technique de l'Elevage Bovin.

United Nations Conference on Trade and Development/General Agreement on Tariffs and Trade (1983) *The World Market for Horsemeat*, p. 120. Geneva: International Trade Centre.

HPLC *See* **Chromatography**: Principles; Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; Supercritical Fluid Chromatography; Combined Chromatography and Mass Spectrometry

HUNGER

J Blundell and A Gillett, The University of Leeds, Leeds, UK

Hunger and the Regulation of Appetite

Hunger refers to a drive or state of motivation that impels animals to search for food. In humans, hunger represents a subjective experience or feeling that is associated with the desire to obtain and eat food. This is how the word 'hunger' is used in everyday social discourse, and scientific methods have been applied to understand the mechanisms underlying this common subjective experience. From a functional point of view, hunger achieves a purpose: that nagging feeling, the presence of which serves to stimulate thoughts about food and eating, is useful and reminds us that the body needs food. In this way, hunger can be seen to possess a clear biological function.

Measurement of Hunger

How do we describe the experience of hunger as it occurs in everyday life? More than 25 years ago, a serious attempt was made to investigate this issue by giving people a questionnaire that asked about physical sensations in a number of bodily areas, moods, urges to eat, and preoccupation with thoughts of food. The questionnaire was completed before and after eating. It was found that the observation 'I feel hungry,' is typically based on the perception of bodily feelings, which at times may be very strong. Gastric sensations, a hollow feeling, or stomach rumbling, are frequent indicators of hunger, although people also report sensations in the mouth, throat, and head. These accompany more diffuse feelings of restlessness and excitability as well as an urge to eat. In one celebrated project, known as the Minnesota experiment, a group of volunteers were placed on a semistarvation diet for 6 months, and their experience of hunger was extreme. Nearly two-thirds reported feeling hungry all the time, and a similar proportion experienced physical discomfort due to hunger. Subjects experienced a marked increase in what was referred to as 'hunger pain.' For some, this pain, vaguely localized in the abdomen, was of mild discomfort. For others, it was intensely painful.

Eating changes both the pattern of physical sensations and the accompanying emotional feelings, with unpleasant and aversive sensations being replaced by pleasant sensations. For example, an aching stomach becomes relaxed, and the feeling of excitement and irritability is replaced by one of contentment.

Despite a great deal of individual variability in the perception of these sensations, hunger can be associated with clear symptoms; it is partly through reference to these that people can make judgements about the intensity of their hunger experience. The measurement of hunger, desire to eat, or urge to eat is most commonly conducted using fixed-point or visual-analog scales. Respectively, these require the subject to choose a number from a scale or a point on a line that corresponds to their current state of hunger. Careful presentation of these scales to people who understand what is being asked of them will yield meaningful information. More importantly, quantifying the subjective experience of hunger makes it a state, which is amenable to scientific investigation. Consequently, hunger can be described qualitatively in terms of the sensations with which it is associated, and it can also be measured quantitatively. This means that the significance of hunger can be understood through its structure and by its intensity. In recent years, the visual analog scale technique has been incorporated into the screen of a small handheld computer. This device, called the Electronic Appetite Rating System (EARS) provides a convenient and user-friendly way of monitoring hunger during the course of a day.

Hunger, Eating Patterns and the Satiating Power of Food

If hunger is the feeling that reminds us to seek food, then the consumption of food is the action that diminishes hunger and keeps it suppressed for a certain period of time, perhaps until the next meal or snack. The capacity of food to reduce the experience of hunger is called the satiating power or satiating efficiency. This power is achieved by certain properties of the food itself engaging with various physiological and biochemical mechanisms within the body that are concerned with the processing of food once it has been ingested. The satiating power of food therefore results from a variety of biological processes and is an important factor in the control of hunger. Some foods have a greater capacity to maintain suppression over hunger than other foods. (*See* **Satiety and Appetite**: Food, Nutrition, and Appetite.)

How is hunger related to the overall control of human appetite and food consumption? The feeling of hunger is an important component in determining what we eat, how much we eat, and when we eat. However, it must be seen in a context of social and physiological variables. On the one hand, eating patterns are maintained by certain enduring habits, attitudes, and opinions about the value and suitability of foods and overall liking for them. These factors, derived from the cultural ethos, largely determine the range of foods that will be consumed, and sometimes the timing of consumption. The intensity of hunger experienced may also be determined, in part, by the culturally approved appropriateness of this feeling.

On the other hand, normal hunger is more importantly associated with the events surrounding meals – so-called periprandial circumstances – and the periods between meals. Thus, hunger can be considered to arise from an interaction between the physiological requirements of the body for food (or particular nutrients) and the capacity of food to satisfy these requirements. Hunger will therefore be successively stimulated and suppressed, giving rise to a diurnal rhythm. This rhythm, and the relationship between hunger and eating, may be modulated by certain social factors (e.g., distressing psychological events) or interrupted by some disease states.

Hunger and the Satiety Cascade

When food consumption reduces hunger and inhibits further eating, two processes are involved. For technical precision and conceptual clarity, it is useful to describe the distinction between satiation and satiety. Both terms may be assigned workable operational definitions (i.e., definitions that depend upon measurable events). Satiation can be regarded as the process that develops during the course of eating and eventually brings a period of eating to an end. Accordingly, satiation can be defined by the measured size of the eating episode (volume or weight of food, or value of the energy content). Hunger declines as satiation develops and usually reaches its lowest point at the end of a meal. Satiety is defined as the state of inhibition over further eating that follows the end of an eating episode and arises from the consequences of food ingestion. The intensity of satiety can be measured by the duration of time elapsing until eating is recommenced, or by the amount consumed at the next meal. The strength of satiety is also measured by the duration of the suppression of hunger. As satiety weakens, so hunger is restored. Figure 1 indicates the changes that occur in the rating of hunger during a meal (as satiation develops) and following a meal (as satiety evolves). It can be seen that the measurement of hunger is an important index of the degree of satiation and satiety.

In the view of some researchers, satiation and satiety can be referred to as intrameal satiety and intermeal satiety, respectively. What mechanisms are responsible for these processes? It is clear that the mechanisms involved in reducing hunger and in maintaining the suppression over hunger range from those that occur when food is initially sensed, to the

effects of metabolites on bodily tissues following digestion and absorption (across the wall of the intestine and into the bloodstream). By definition, satiety is not an instantaneous event but occurs over a considerable time period; it is therefore useful to distinguish different phases of satiety, which can be associated with different mechanisms. This concept is illustrated in Figure 2.

Four mediating processes are identified: sensory, cognitive, postingestive, and postabsorptive. These maintain inhibition over hunger (and eating) during the early and late phases of satiety. Sensory effects are generated through the smell, taste, temperature, and texture of food, and it is likely that these factors inhibit eating in the very short term. Cognitive effects represent the beliefs held about the properties of foods, and these factors may help to inhibit hunger in the short term. The category identified as postingestive processes includes a number of possible actions, including gastric distension and rate of gastric emptying, the release of hormones such as cholecystokinin, and the stimulation of physicochemically specific receptors along the gastrointestinal tract. The postabsorptive phase of satiety includes those mechanisms arising from the action of metabolites after absorption across the intestine and into the blood system. This category embraces the action of chemicals such as glucose and amino acids, which may act directly on the brain after crossing the blood–brain barrier or may influence the brain indirectly via neural inputs following stimulation of peripheral chemoreceptors. The most important suppression and subsequent control of hunger is brought about by postingestive and postabsorptive mediating processes.

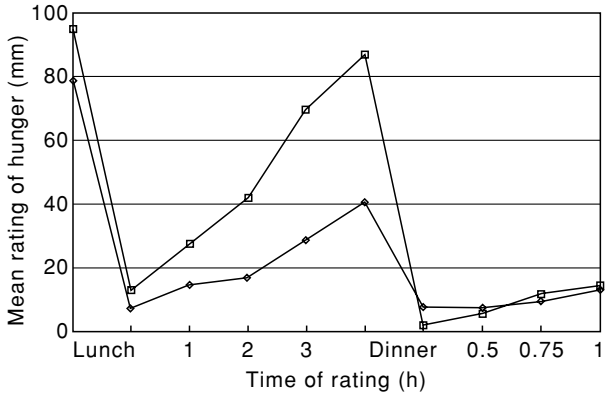

Figure 1 Profiles of the ratings of the subjective experience in two groups of subjects, one of which ate a large lunch (◇) and the other a lunch 50% smaller (□). Each meal suppressed hunger, but the intensity of hunger recovered more rapidly following the small meal. Four hours after the first meal, subjects were given another meal to eat (dinner), and hunger was again suppressed. Hunger ratings reflect the amount of food consumed and the timing of meals.

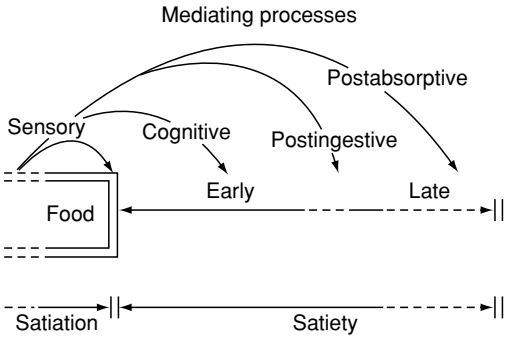

Figure 2 Satiety cascade. This diagram represents the mediating processes, arising from food consumption, that influence the feeling of hunger and determine the satiating efficiency of the food.

Conditioned Hunger

It should be kept in mind that the psychobiological system for appetite control has the capacity to learn, i.e., to form associations between the sensory and postabsorptive characteristics of foods. This means that it will be useful to distinguish between the unconditioned effects of foods, i.e., those in which the natural biological consequences of food processing in the gut are reflected in satiety, and the conditioned effects that come into play owing to the links developed between sensory aspects of food – particularly those that are tasted – and later metabolic effects generated by the same food. The sensory characteristics (or cues) therefore come to predict the impact that the food will later exert. Consequently, these cues can suppress hunger according to their relationship with subsequent physiological events. However, the potency of this mechanism depends upon the stability and reliability of the relationship between tastes (sensory cures) and physiological effects (metabolic consequences) of food. When there is distortion or random variation between sensory characteristics and nutritional properties, the conditioned control of hunger is weakened or lost. Therefore, learned hunger will not be an important factor when the food supply contains many foods with identical tastes but differing metabolic properties.

Is Hunger the Cause of Eating?

In searching for an answer to this question, we can begin by considering whether hunger is a necessary or a sufficient condition for eating to occur. Since hunger may be present, but a person may volitionally prevent

themselves from eating (e.g., someone undertaking a fast for moral or political conviction), hunger is not a sufficient condition. Occasions can also be imagined where a person would eat, if food was particularly tempting, where no hunger was being experienced.

Therefore, it appears that hunger is neither a sufficient nor a necessary condition for eating, i.e., the relationship between hunger and eating is not based on biological inevitability. However, under many circumstances, there is a synchronous relationship between the pattern of food intake and the rhythmic oscillation of hunger. Moreover, the results of many studies confirm the strong relationship between the intensity of experienced hunger sensations and the amount of food consumed. For example, when hunger and eating have been monitored continually, it has been reported that 'the correlation using hunger ratings and intake during the same hour of the day was $r = < 0.5$ ($P < 0.02$). That is, hunger ratings at the start of each hour were correlated with reported intake in the hour following each hunger rating.' In addition, significant correlations have been found between hunger ratings prior to a lunch and subsequent intake of fat, protein, and energy, indicating that these measures of subjective hunger provide a valid index of actual food intake. The correlations found in almost all studies indicate that, in many circumstances, the measured intensity of hunger reliably predicts the amount of food that will be consumed. This fact has led to the proposal that there is a causal connection between hunger and the size of a following meal. Table 1 shows measured correlations between hunger, food intake, and related subjective experiences.

These observations provide the basis for a working conceptualization. The correlations are sufficiently strong and reliable to allow us to act as if hunger is a cause of eating. In actuality, it is probably the case that common physiological mechanisms induce changes both in hunger and in the eating response. Therefore, we can consider an appetite control system in which hunger, eating, and physiological mechanisms (of the satiety cascade) are coupled together. However, the coupling is not perfect, and there will be circumstances where, for example, a physiological treatment will change eating but not hunger. In other words, uncoupling can occur. However, when this happens, e.g., under certain conditions of fasting or in disordered eating, the mechanisms responsible can provide useful information about the role of hunger in the overall control of appetite. It should be noted that, although a number of individuals often deny that sensations of hunger are related to when eating occurs, under scientifically controlled conditions, there is usually a strong relationship between hunger and eating. (*See* **Sensory Evaluation**: Sensory Characteristics of Human Foods.)

Table 1 Intercorrelations between ratings of hunger, other subjective sensations, and measure of food intake observed in experimental studies in lean and obese subjects (all correlation coefficients shown are statistically significant)

	Food intake (kcal)	*Hunger*	*Desire to eat*	*Fullness*
Lean subjects				
Hunger	0.75			
Desire to eat	0.77	0.96		
Fullness	−0.69	−0.93	−0.97	
Prospective consumption	0.80	0.97	0.95	−0.91
Obese subjects				
Hunger	0.65			
Desire to eat	0.68	0.96		
Fullness	−0.54	−0.75	−0.77	
Prospective consumption	0.66	0.88	0.89	−0.78

Disorders of Hunger

The clinical eating disorders, anorexia nervosa and bulimia nervosa, are commonly believed to encompass major disturbances of hunger. Yet the role that hunger may play is not entirely clear. Contrary to the literal meaning of the term, 'anorexia' is not experienced as a loss of appetite. Rather, clinicians recognize that anorexics may endure intense bouts of hunger during their self-imposed restricted eating. For some, their strength in resisting intense episodes of hunger provides a feeling of self-mastery and control that is absent in most other areas of their lives. Anorexics therefore overcome their feelings of hunger. However, there is some evidence to suggest that in conditions of total starvation, hunger may become temporarily diminished. However, once eating is recommenced, hunger may return rapidly and with extreme intensity. (*See* **Anorexia Nervosa**.)

Like anorexia, bulimia finds its literal meaning in an altered state of hunger – 'ox hunger.' But again, the term is imprecise. Close analysis of the physiological states, which characteristically provide an eating binge, show hunger to be lower than it is prior to a normal meal. In addition, while the urge to eat may be strong during a binge, the large amount of food consumed implies some defect in satiation rather than in hunger. Research that has monitored hunger in eating disordered patients during a simple meal describes both great individual variability in the hunger experience and examples of extreme disorganization. In some cases, hunger appears quite dissociated from the circumstance of eating. It is likely that a stable

eating pattern is necessary in order to normalize the experience of hunger, and this may take a long time to establish. (*See* **Bulimia Nervosa**.)

Obesity is another condition in which hunger is often assumed to be impaired. In seeking a cause for excess weight, overeating has been regarded as being caused by high levels of hunger. Although it is now accepted that, on average, obese individuals have a higher energy intake than lean people, this cannot necessarily be assigned to higher levels of hunger. It appears that the experience of hunger is similar in lean and in most obese subjects. Studies suggest that in general, there is no gross fault in the obese food-intake regulation system. Any 'fault' is likely to be subtle. An indication of this comes from the response to a reduced-energy diet. Obese people, in common with those of normal weight, respond to a reduction in their energy intake by increased feelings of hunger. Alternatively, anorectic drugs or nutrients that suppress hunger appear to do so similarly in obese and lean people. Attempts to lose weight by dieting may fail because of an inability to manage hunger. However, it should be recognized that for some very obese people, the sensation of hunger occurs in an eccentric fashion and is not related to periods of eating. Some people can report strong feelings of hunger and fullness simultaneously, (*See* **Obesity**: Etiology and Diagnosis.)

Hunger and Health

The analysis of the phenomenon of hunger, the influencing mechanisms, and its relationship to the regulation of food intake indicate that hunger is a potent factor in the maintenance of general health. In societies where food is plentiful, the relatively mild experiences of hunger play a biological useful role in the orderly regulation of eating patterns. When food is scarce, the power of hunger can drive people to desperate deeds. In certain disease states, a very intense hunger or its absence can indicate the severity of the underlying pathology. In other extreme circumstances, hunger is forced out of synchrony with physiological mechanisms, eating behavior, or both. This labile or disregulated hunger reflects the pathology of the appetite system. The drive to eat, identified as hunger, is therefore a significant marker of health. If the mechanisms underlying this drive are compromised, hunger can become difficult to control. (*See* **Malnutrition**: The Problem of Malnutrition.)

See also: **Anorexia Nervosa**; **Bulimia Nervosa**; **Malnutrition**: The Problem of Malnutrition; **Obesity**: Etiology and Diagnosis; **Satiety and Appetite**: Food, Nutrition, and Appetite; **Sensory Evaluation**: Sensory Characteristics of Human Foods

Further Reading

Blundell JE and Rogers PJ (1991) Hunger, hedonics, and the control of satiation and satiety. In: Friedman M and Kare M (eds) *Chemical Senses*, vol. 4, pp. 127–148. New York: Marcel Dekker.

De Castro JM and Elmore DK (1988) Subjective hunger relationships with meal patterns in the spontaneous feeding behaviour of humans: evidence for a causal connection. *Physiology and Behaviour* 43: 159–165.

Hill AJ, Magson LD and Blundell JE (1984) Hunger and palatability: tracking ratings of subjective experience before, during and after the consumption of preferred and less preferred food. *Appetite* 5: 361–371.

Keys A, Brozek J, Henscher A, Mickelson O and Taylor HL (1950) *The Biology of Human Starvation*, vol. II. Minneapolis, MN: University of Minnesota Press.

King NA, Lawton CL, Delargy HJ, Smith FC and Blundell JE (1997) The electronic appetite rating system (EARS): a portable computerised method for continuous automated monitoring of motivation to eat and mood. In: Wellman PJ and Bartley GH (eds) *Ingestive Behavior Protocols*, pp. 71–76. Society for the Study of Ingestive Behavior.

Monello LF and Mayer J (1967) Hunger and satiety sensations in men, women, boys and girls. *American Journal of Clinical Nutrition* 20: 253–261.

Hydrogenation *See* **Vegetable Oils**: Types and Properties; Oil Production and Processing; Composition and Analysis; Dietary Importance

Hygiene *See* **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant; Overall Approach; Modern Systems; **Food Poisoning**: Classification; Tracing Origins and Testing; Statistics; Economic Implications; **Contamination of Food**; ***Escherichia coli***: Occurrence; Detection; Food Poisoning; Occurrence and Epidemiology of Species other than *Escherichia coli*; Food Poisoning by Species other than *Escherichia coli*; **Effluents from Food Processing**: On-Site Processing of Waste; Microbiology of Treatment Processes; Disposal of Waste Water; Composition and Analysis; **Food Safety**; ***Salmonella***: Properties and Occurrence; Detection; Salmonellosis; **Saponins**; **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage; Fungi in Food – An Overview; Molds in Spoilage; Yeasts in Spoilage

Hyperglycemia (Hyperglycaemia) *See* **Diabetes Mellitus**: Etiology; Chemical Pathology; Treatment and Management; Problems in Treatment; Secondary Complications

HYPERLIPIDEMIA (HYPERLIPIDAEMIA)

J P D Reckless and J M Lawrence, University of Bath, Combe Park, Bath, UK

Lipids

Lipids are a heterogeneous group of substances that play key roles in normal physiology. Two groups are important – that where a main component includes one or more fatty acids (such as triglycerides and phospholipids) and that where the main component is a steroid nucleus (such as cholesterol and steroid hormones). Both groups are relatively insoluble in water. The lipoprotein transport system carries these hydrophobic compounds through the aqueous medium of blood plasma to their site of use, in a carefully regulated process. Lipoproteins consist of a package of triglycerides, cholesterol (free and esterified), phospholipids, and apolipoproteins, with a hydrophilic shell and a hydrophobic core (Figure 1). The sites of synthesis are the gut and liver. The main lipoproteins are high (HDL), low (LDL), and very low (VLDL) density lipoproteins, and chylomicron7s. Each has a different group of surface apolipoproteins, pivotal in determining the site, mechanism, and speed of lipoprotein metabolism. The precise molecular composition in each lipoprotein class varies considerably.

Hyperlipidemia means high levels of fats in the blood, which may be physiological postprandially. Pathological hyperlipidemia results from disordered metabolism of the lipoproteins with either excess production, altered clearance, or both. It is most commonly multifactorial, involving both environmental and intrinsic factors. Many genetic disorders are now well recognized, causing various degrees of hyperlipidemia or dyslipidemia (altered composition). Genetic studies have helped the understanding of lipoprotein metabolism.

This review discusses normal lipoprotein metabolism, focusing on hyperlipidemia. As treating hyperlipidemia has significant benefits in terms of long-term health, having an understanding of normal and abnormal lipoprotein metabolism is becoming increasingly important.

The Intestine

Most dietary fats are triglycerides. These are hydrolyzed to mono- and diacylglycerols, nonesterified fatty acids (NEFA) and glycerol by pancreatic or gut lipases, and absorbed as mixed micelles in the small intestine. In the intestinal cells, triglycerides are reformed and interact with apolipoprotein B-48 (apoB48) to form chylomicrons. ApoB48 is the structural protein for chylomicrons. Each chylomicron has one molecule of apoB48 and carries very large numbers of triglyceride molecules. Precise numbers vary widely, as does the chylomicron size, depending on the diet (diameter 100–1000 nm). Chylomicrons also accommodate any dietary cholesterol absorbed

(usually less than 0.5 g per day). They are released into the intestinal lymphatics and subsequently into the circulation via the thoracic duct.

Chylomicron Metabolism

In the systemic circulation, chylomicrons receive apolipoproteins additional to apoB48, particularly apolipoprotein C (apoC) and apolipoprotein E (apoE), from HDL (Figure 2). One apoC, apoC-II, activates the enzyme lipoprotein lipase (LPL) and facilitates ligand/enzyme binding between LPL and chylomicrons. LPL is found widely on capillary endothelial tissues of many organs including adipose tissue and skeletal and cardiac muscle. The activated enzyme hydrolyzes chylomicron-triglyceride, and NEFA are released for uptake into tissues to be stored or used as energy. As triglyceride is removed, the relative concentration of cholesterol rises, and the lipoprotein becomes smaller and denser. Excess surface components of phospholipids, free cholesterol, and apolipoproteins (but not the apoB48) are transferred to HDL. The efficiency of this process depends on the activity of LPL and also the VLDL triglyceride level in the circulation, as VLDL competes with chylomicrons for LPL. The chylomicron remnant is removed by liver receptor uptake, dependent on multiple copies of apoE on the remnant surface. This receptor is structurally related to the LDL receptor (*see* Low Density Lipoprotein) as a member of the LDL-supergene family and is named the LDL receptor-related protein (LRP). The affinity of the binding depends on the apoE phenotype. Chylomicron remnant removal effectively delivers dietary cholesterol and residual triglyceride to the liver.

Very Low Density Lipoproteins (VLDL)

VLDL are the other major triglyceride-rich lipoproteins (Figure 3). They carry endogenous, hepatically

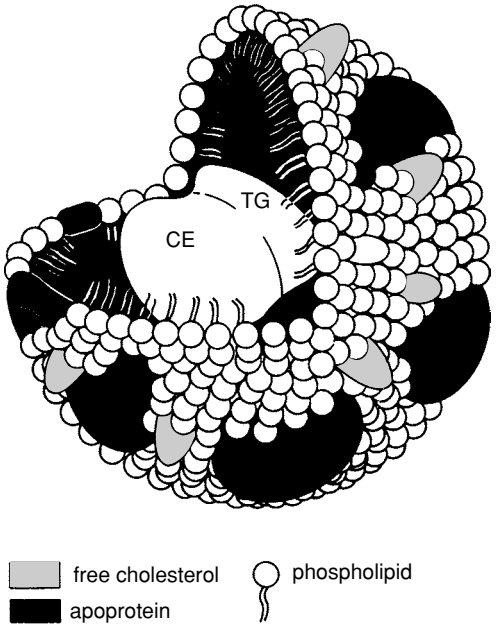

Figure 1 Diagram of a human plasma lipoprotein. From Feher and Richmond, *Lipid and Lipid Disorders*, Gower Medical Publishing, with permission.

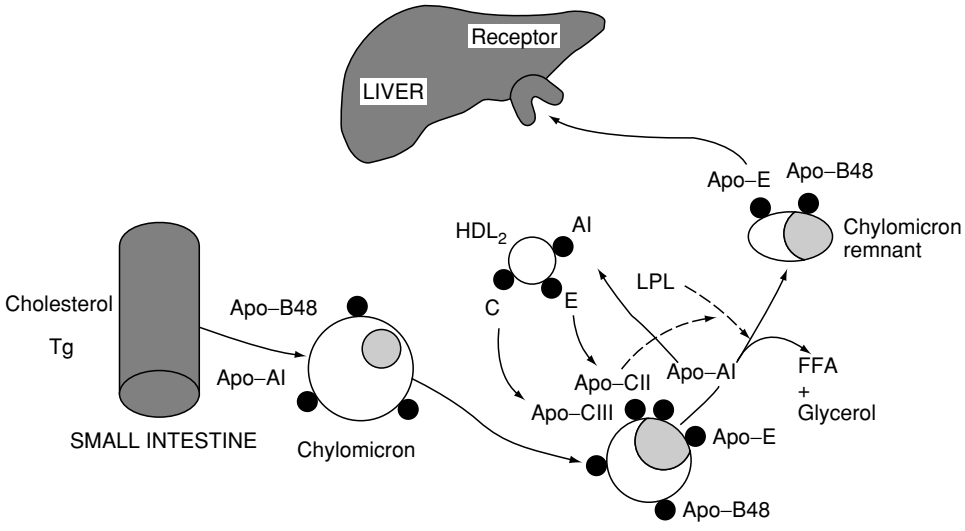

Figure 2 Chylomicron metabolism: synthesis and degradation. The pathway for metabolism of triglyceride of exogenous, dietary origin. □, triglyceride (Tg); ▩, cholesterol; LPL, lipoprotein lipase; FFA, free fatty acid. Reproduced from Hyperlipidaemia, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

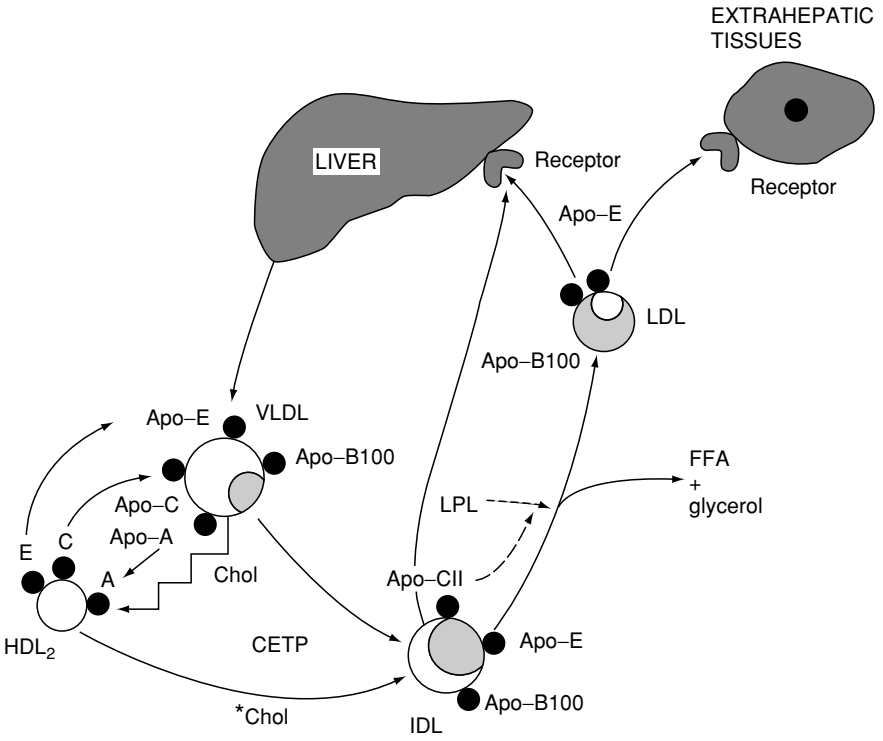

Figure 3 VLDL metabolism. The pathway for metabolism of triglyceride of endogenous, hepatic origin. □, triglyceride (Tg); ▨, cholesterol; LPL, lipoprotein lipase; FFA, free fatty acid; Chol, free cholesterol; *Chol, esterified cholesterol; CETP, cholesterol ester transfer protein. Reproduced from Hyperlipidaemia, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

synthesized triglycerides, largely derived from dietary carbohydrate or from plasma NEFA, and are of a large size (30–60 nm diameter). Each VLDL particle has a single molecule of a larger apoB form, apoB100, with other apolipoproteins (apoE and apoCs) being acquired in plasma from HDL. Like apoB48, apoB100 has a structural role, but also a functional role as a ligand for catabolism. The same gene codes for apoB in both gut and liver, but in the gut, there is post-translational editing of messenger RNA, resulting in a stop codon at position 2153. A truncated (48%) apoB48 is then formed as the structural apolipoprotein, whereas in VLDL, the whole 100% of apoB is synthesized to provide also the functional ligand component.

VLDL size is dependent on the triglyceride content, with large buoyant triglyceride-enriched VLDL being formed at times of triglyceride excess. These large VLDL may be a poorer substrate for the usual path of metabolism involving apoCII-dependent activation of LPL and apoCII-LPL binding similar to that seen with chylomicrons, and thus have a longer residency time in the blood. As triglyceride is hydrolyzed, VLDL particles shrink with loss of surface components to HDL and subsequent catabolism to LDL via IDL. This process is less efficient in the presence of high triglyceride levels and large buoyant VLDL. A longer plasma residency time allows a greater exchange of triglyceride from VLDL into LDL and HDL with reverse exchange of cholesterol ester, mediated by cholesterol ester transfer protein (CETP). This produces cholesterol-enriched VLDL remnants that are less readily metabolized to LDL and can be removed by alternative, but potentially atherogenic, pathways. Additionally, this process produces triglyceride-enriched LDL that shrink as some triglyceride is hydrolyzed in the liver by another endothelial enzyme, hepatic lipase, producing small, dense LDL_3 (see below).

Intermediate Density Lipoproteins (IDL)

Intermediate density lipoproteins are formed as an intermediate step in the normal metabolism of VLDL. They are catabolized rapidly, and the concentration in the plasma is therefore usually very low. IDL are either taken up directly by the liver through binding of surface apoE on IDL to apoE or apoB100/apoE receptors, or are converted to LDL. The exact mechanism of this is unclear but involves the loss of

apoE and the C-apolipoproteins and further triglyceride hydrolysis by hepatic lipase.

Low Density Lipoproteins (LDL)

LDL are effectively responsible for delivery to tissues of cholesterol used for cell membrane synthesis and repair, and for processes such as synthesis of steroid hormones and vitamin D (Figure 4). Entry into cells occurs either in a concentration-dependent nonspecific manner or, more importantly, by a very closely regulated receptor-mediated pathway. LDL bind to specific LDL apoB100/apoE receptors on cell membranes, and LDL–receptor complexes are internalized and undergo lysosomal degradation. The cholesterol ester is hydrolyzed to free cholesterol, while most of the LDL receptors are recycled. The highest concentration of receptors is found in the liver.

In order to maintain cellular function, the cell content of cholesterol is very tightly regulated. High intracellular levels of free cholesterol lead to:

1. Suppression of endogenous cell cholesterol synthesis by inhibition of the rate limiting enzyme 3-hydroxy, 3-methylglutaryl coenzyme-A reductase (HMG CoA-reductase).
2. Suppression of apoB-100/apoE receptor synthesis.
3. Increased activity of acyl-coenzyme A cholesterol acyl-transferase (ACAT), leading to increased cholesterol esterification.

LDL particles are not homogeneous and can be divided on the basis of size and density into LDL_1, LDL_2, and LDL_3. LDL_1, the largest and least dense, is a better substrate for the LDL receptor and is also less easily oxidized. Individuals with higher levels of LDL_1 relative to LDL_3 are said to have a type A phenotype, and those with the converse are said to have a type B phenotype. Excess LDL_3 is associated with an excess risk of coronary heart disease (*see* Hyperlipidemia and the Origin of Clinical Atheroma below).

High Density Lipoproteins (HDL)

HDL are the smallest, most dense of the lipoproteins, and are involved as cholesterol acceptors in reverse cholesterol transport from the peripheral tissues to the liver where the cholesterol is either excreted into the bile or assembled into other lipoproteins (Figure 5). HDL are synthesized in both the liver and gut. Initially secreted as disk-shaped protein and phospholipid bilayers, they mature by acquiring lipid and protein components from intracellular and extracellular fluids and from other lipoproteins to form spherical particles of 5–15 nm diameter. The predominant and structural apolipoproteins of HDL are apolipoproteins A-I and A-II, with significant amounts of apoCs and apoE, but with no apoB. HDL particles are divided by size and density into smaller and denser HDL_3 and larger and less dense HDL_2. HDL_2 particles tend to contain apo A-I only and HDL_3 particles tend to contain both apoAI and apoAII. The presence of larger, less dense HDL_2 correlates with more efficient reverse cholesterol transport.

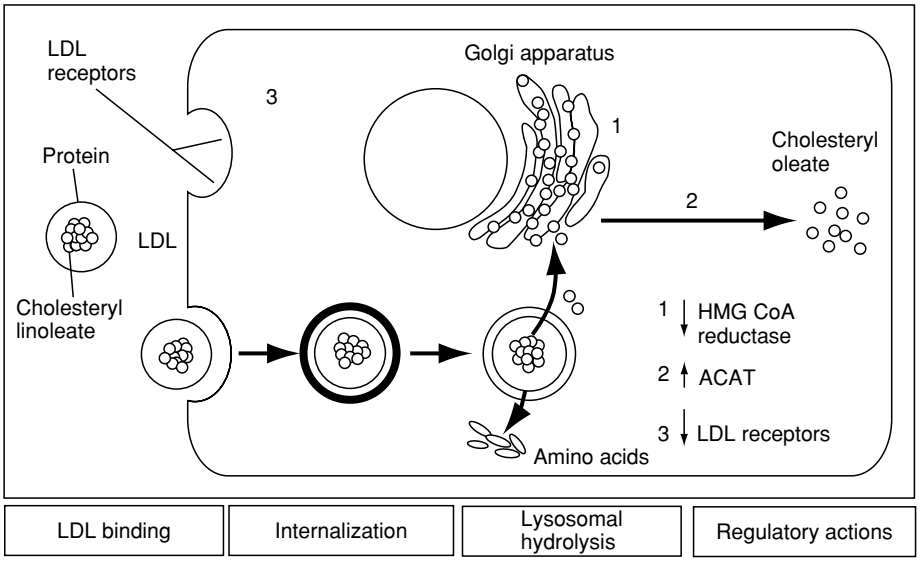

Figure 4 Regulation of intracellular cholesterol concentration by LDL-cholesterol uptake. Reproduced from Hyperlipidaemia, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

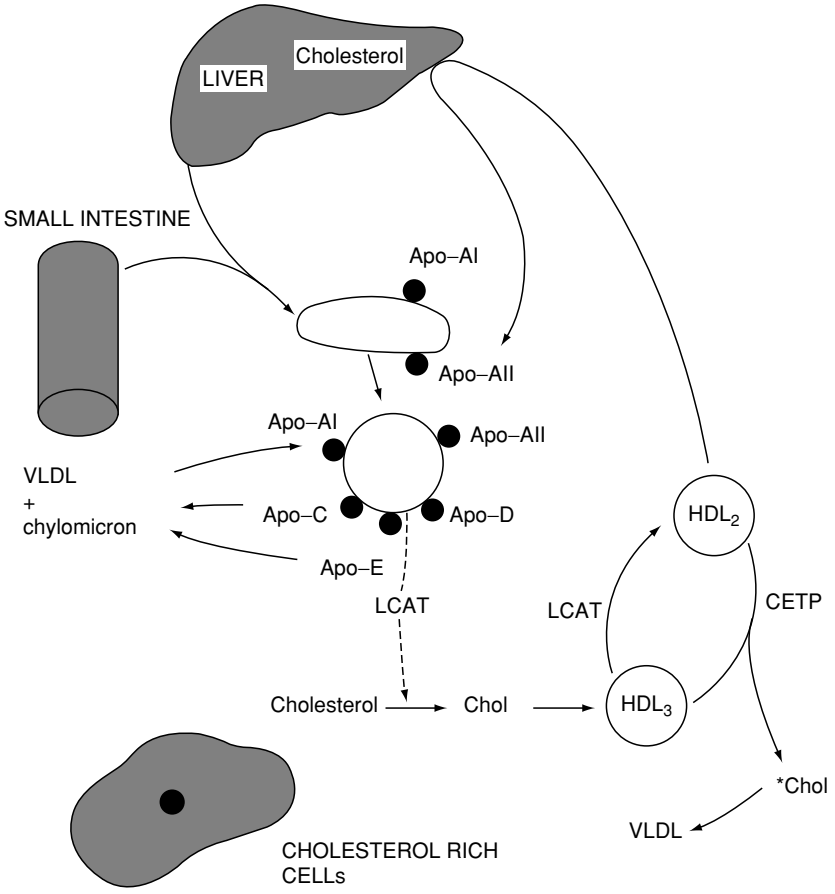

Figure 5 HDL metabolism; Chol, free cholesterol; *Chol, esterified cholesterol; CETP, cholesterol ester transfer protein; LCAT, lecithin cholesterol acyl transferase. Reproduced from Hyperlipidaemia, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

The mechanism of reverse cholesterol transport is gradually being elucidated. Evidence suggests that cells replete with cholesterol express HDL-binding sites on the plasma membrane. Intracellular signaling promotes hydrolysis of stored cholesterol esters, movement of free cholesterol to the cell surface, and transfer down a concentration gradient to HDL. Surface cholesterol is esterified by lecithin cholesterol acyl transferase (LCAT), an HDL-associated enzyme. The resulting hydrophobic cholesterol ester enters the HDL core, leaving the surface cholesterol depleted and thus maintaining the concentration gradient with the cell membrane. Cholesterol esters are transferred from HDL by CETP to VLDL, IDL, and LDL. Cholesterol esters are delivered to the liver substantially by receptor-mediated LDL uptake, and some directly delivered from HDL. Synthesis of HDL depends on transfer of surface components from VLDL catabolism. When there is hypertriglyceridemia, the number and size of VLDL particles rise. As larger VLDL particles tend to be catabolized more slowly, the availability of surface components for HDL is reduced, hence partly explaining the frequent inverse relationship between the concentrations of VLDL and HDL.

Consequences of Hyperlipidemia

While severe hypertriglyceridemia can lead to pancreatitis, the major consequence of hyperlipidemia is accelerated atherogenesis with premature coronary heart disease (CHD), stroke, and peripheral vascular disease. The major risk factors for early CHD are hypertension, cigarette smoking, hyperlipidemia, and diabetes mellitus, risk increasing markedly when multiple risk factors are present.

Hyperlipidemia and the Origins of Clinical Atheroma

Atheroma occurs in large elastic and muscular arteries such as the aorta, coronary, femoral, and carotid

arteries, and especially at predisposed sites such as bifurcations where there is flow disturbance. Plaques form on the basis of fatty streaks that may be present very early in life. Damage to the endothelial lining from high blood pressure, local injury, or poor oxygen supply results in increased permeability, enabling lipoproteins to pass more easily into the subendothelial space. LDL and triglyceride-rich lipoprotein remnants are atherogenic. The influx into the arterial wall depends on both the concentration and the size of the LDL particles in plasma. The higher the concentration of LDL in the plasma, the greater is the degree of influx, and smaller denser LDL_3 particles cross the endothelium more easily. LDL in the arterial wall undergoes a number of changes including oxidation with LDL_3 particles being more susceptible. The oxidized LDL particles are engulfed by scavenger macrophages. These macrophages are formed from monocytes that have traversed the damaged endothelial wall and are immobilized in response to a number of chemical signals. In response to altered LDL, the macrophages express acetylated LDL receptors and take up LDL to become lipid rich foam cells. These undergo apoptosis producing a lipid-rich extracellular medium. Smooth muscle cells migrate from the arterial media in response to this process. These proliferate and produce a connective tissue matrix rich in collagen, elastin, and proteoglycans forming the plaque cap, as part of a damage-healing process.

Plaques may be asymptomatic. In contrast, they may result in a significant reduction in blood flow, producing symptoms such as angina or claudication. A plaque that has a thin cap and a large lipid core may be unstable and susceptible to rupture. In such plaques, the excess macrophage component produces metalloproteinases that break down the collagen matrix (synthesized by smooth muscle cells), making cap thinning and plaque rupture (especially at the shoulder of the cap) more likely. Should this happen, acute occlusion of a vessel may occur, causing syndromes such as acute myocardial infarction.

Classification of Hyperlipidemia

Hyperlipidemia normally refers to high concentrations of the blood lipids, cholesterol, and triglycerides. These high levels reflect underlying changes in lipoproteins, and could be caused by overproduction, reduced catabolism or both. Thus, high levels of triglycerides reflect elevated concentrations of the triglyceride-rich lipoproteins, chylomicrons and VLDL, and their remnants. An increase in cholesterol concentration usually reflects an increase in LDL, with or without an accompanying increase in VLDL. Occasionally, raised total cholesterol may be due to a very high HDL alone.

Hyperlipidemia may be secondary to other medical disorders or therapy, or it may be primary. Primary hyperlipidemia may be caused by a single gene defect or, more commonly, by a polygenic background influenced by environmental factors.

Originally, primary (inherited) disorders were classified on the physicochemical characteristics of lipoproteins defined in the laboratory using ultracentrifugation or electrophoresis (Table 1). Treatment of a patient's hyperlipidemia often results in changes in the lipoprotein pattern and laboratory characteristics, so that such a classification has limited value. It is more useful to define the disease or metabolic defect, and as this is becoming increasingly possible, efforts to do this should always be made. Secondary causes of hyperlipidemia are listed in Table 2. Obesity, diabetes, and excess alcohol commonly act on a genetic background to produce hyperlipidemia.

Hypercholesterolemia

Common or Polygenic Hypercholesterolemia

In the population as a whole, cholesterol levels approach a normal distribution. Epidemiological

Table 1 Classification of hyperlipidemia (Fredrickson's classification, as modified by the World Health Organization)

Type of hyperlipidemia	*Cholesterol*	*Triglyceride*	*Lipoprotein raised*
I		Raised	Chylomicrons
IIa	Raised	Normal	LDL
IIb	Raised	Raised	LDL plus VLDL
III	Raised	Raised	IDL
IV	Normal or raised	Raised	VLDL
V	Raised	Markedly raised	Chylomicrons plus VLDL

Table 2 Commoner secondary causes of hyperlipidemia

Cause	*Example*
Metabolic and nutritional	Obesity Alcohol
Endocrine	Diabetes mellitus Hypothyroidism Pregnancy
Drugs	β-blockers Thiazides Estrogens
Renal disease	Chronic renal failure Nephrotic syndrome
Liver disease	Biliary obstruction

studies have shown that the risk of coronary artery disease rises with the cholesterol level. Several large-scale intervention studies have demonstrated that lowering cholesterol reduces CHD risk in both primary and secondary (after the event) prevention.

In addition to common or polygenic hypercholesterolemia, a number of well-defined inherited syndromes are well recognized resulting in hypercholesterolemia and an increased CHD risk.

Familial Hypercholesterolemia (FH)

Familial hypercholesterolemia is the commonest disease in Western populations to be caused by a single dominant gene defect. It is characterized by a raised cholesterol from birth, subsequent development of cutaneous and tendon xanthomata, and premature vascular disease (Figure 6). The gene coding the LDL receptor is found on chromosome 19. Over 150 different LDL receptor mutations have been characterized, affecting receptor synthesis and function. Heterozygous FH (1 mutant and 1 normal gene) affects 0.2% of the population, the homozygous condition occurring at a rate of 1:1 000 000. A true homozygote will have two copies of the same defect, but compound heterozygosity (with two different defects) is more common.

Homozygous FH is characterized by extreme hypercholesterolemia (cholesterol is approximately four times normal) with cutaneous and tendon xanthomata and corneal arcus beginning in childhood. Significant atheromatous involvement of the aortic root with coronary ostial stenosis is present by puberty, and death from coronary heart disease usually occurs prior to the age of 20 in untreated individuals.

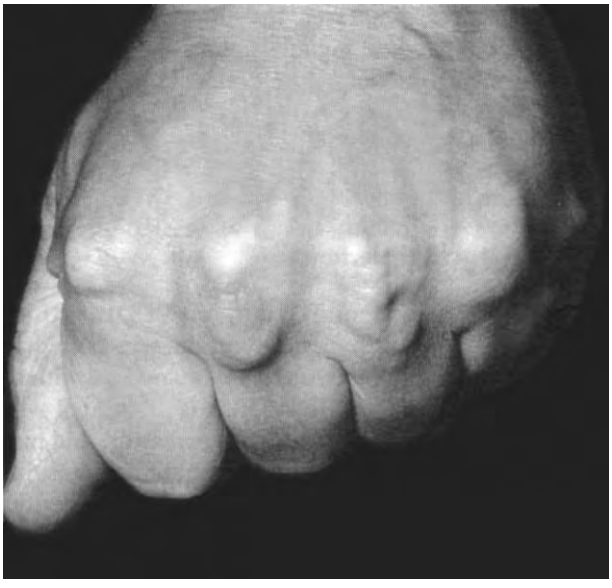

Figure 6 Xanthomata in the extensor tendons of the hands in familial hypercholesterolemia. Reproduced from Hyperlipidaemia, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Tendon xanthomata are also pathognomonic of heterozygous FH occurring in some adults. LDL cholesterol is typically twice normal. Untreated, it leads to CHD in 50% of males by the age of 50 years, and 50% will have died by the age of 60. Fifty per cent of females will have clinical CHD by age 60 when 15% will have died.

The diagnosis of homozygous FH is usually straightforward. Heterozygous FH should be considered in any patient with hypercholesterolemia and personal or family premature CHD. Tendon xanthomata in any family member almost certainly establish the diagnosis, as does high LDL in children.

FH is also caused by abnormalities in apoB such as in familial defective apolipoprotein B-100 where a glutamine-for-arginine substitution in the amino acid 3500 codon in apoB produces an abnormal truncated apoB with impaired binding to the LDL receptor. Like classical FH, it is also an autosomal dominant, affecting about one in 600 of the population.

Familial Combined Hyperlipidemia (FCH)

FCH was first described in 1973 in families of hyperlipidemic patients surviving a myocardial infarction. Less clearly defined than FH, it affects perhaps 1–2% of the population. It is heterogeneous, the etiology remaining uncertain without a single gene defect having been found. Diagnosis requires multiple affected family members and is complicated by variation in time of lipid levels that on occasion may be normal. Cholesterol, triglycerides, or both may be raised with type IIa, IIb, or IV phenotypes.

FCH predisposes to premature atherosclerosis and may be present in 10% of patients with CHD before age 60. Elevated apoB levels are frequently found that may have a central role in atherogenesis in FCH, and apoB concentrations can be a better predictor of CHD than LDL cholesterol. LDL and VLDL each contain one apoB100 molecule, so when apoB is raised, there are increased plasma particle numbers. As LDL numbers increase, the apoB:LDL cholesterol ratio rises, as each particle is relatively cholesterol-poor. LDL particles in FCH are smaller, denser, and more atherogenic (see hyperlipidemia and the origins of clinical atheroma). LDL_3 synthesis is enhanced by the presence of hypertriglyceridemia.

In FCH, no mutations have been found in the apoB structural locus on chromosome 2, but an apoB-raising allele of uncertain location may occur in

some FCH patients. Families heterozygous for LPL or apoC-II deficiency (*see* Severe hypertriglyceridemia below) may exhibit a phenotype similar to FCH. Further work is needed to characterize the FCH genetic background. Hyperapobetalipoproteinemia is a related condition characterized by raised apoB with normal LDL associated with premature CHD.

Hypertriglyceridemia

Hypertriglyceridemia results from overproduction of triglyceride-rich lipoproteins and/or a catabolic defect, due to either a primary lipoprotein disorder or secondary to associated diseases.

Severe hypertriglyceridemia

Severe hypertriglyceridemia (triglycerides > 11 mmol l^{-1}) usually results from secondary factors (e.g., uncontrolled diabetes mellitus, obesity with insulin resistance, and excess alcohol) acting upon a less severe hyperlipidemia. Usually, treatment of the underlying problem will partly correct the hypertriglyceridemia. Inherited abnormalities of lipoprotein metabolism, producing severe hypertriglyceridemia, cause accumulation of chylomicrons with or without VLDL. This gives a pancreatitic, and rarely a CHD, risk.

Lipoprotein lipase deficiency A recessively inherited condition affecting one per million population, this deficiency is caused by mutations in the LPL gene on chromosome 22 resulting in defective function and chylomicron accumulation. Recurrent pancreatitis usually occurs in early adulthood. Eruptive xanthomata are characteristic, and hepatosplenomegaly and lipemia retinalis may occur, with triglycerides ≥ 100 mmol l^{-1}.

Apolipoprotein C-II deficiency ApoC-II activates LPL, and deficiency therefore presents similarly to LPL deficiency, although it is rarer.

Moderate hypertriglyceridemia

Moderate hypertriglyceridemia is often polygenic exacerbated by secondary causes. It is seen frequently with type 2 diabetes and obesity and is an independent CHD risk factor, particularly when associated with a low HDL cholesterol. High triglycerides result in reduced synthesis of HDL, more of which will be HDL_3, increased plasma levels of LDL_3, and a consequent increased risk of atheroma.

The Insulin Resistance Syndrome (see Table 3) This term was coined in 1988 to describe a cluster of abnormalities that frequently occur together in type 2 diabetes. These people are often hypertensive and obese with associated insulin resistance and hyperinsulinemia. A consequence of insulin resistance is increased catabolism of adipose tissue triglyceride with release of nonesterified fatty acids. In the liver, these are reformed into triglycerides and packaged as VLDL, with larger VLDL numbers and size, a lower HDL, and proportionately more LDL_3. This partly explains the threefold CHD increase seen in men with type 2 diabetes and the three- to fivefold increase in women.

Table 3 Features of the Insulin Resistance Syndrome

Syndrome X (The Insulin Resistance Syndrome)
Hyperinsulinemia
Impaired glucose tolerance
Hypertension
Increased triglyceride
Decreased HDL cholesterol

Mixed Hyperlipidemia

Mixed hyperlipidemia with elevated triglycerides and cholesterol is most commonly caused by polygenic elevation of VLDL exacerbated by lifestyle or environmental factors. It can manifest in FCH or FH.

Remnant Hyperlipidemia

Remnant hyperlipidemia is also known as dysbetalipoproteinemia, broad beta disease, or type III hyperlipidemia, and has been recognized as a distinct disease for 40 years. Remnant particles (IDL or β-VLDL) of partial chylomicron or VLDL degradation accumulate in about one in 5000 individuals, although the gene defect affects about 1% of the population. The diagnostic features are palmar crease and tuberous xanthomata (Figure 7). Important complications are premature CHD and peripheral vascular disease, the predisposition to the latter being more than with other dyslipidemias. It is more prevalent in men than women, manifesting at age 25–40. The clinical disorder is brought out by factors such as obesity, diabetes, and untreated hypothyroidism.

Apolipoprotein E(apoE) is a 299-amino-acid glycoprotein on all lipoproteins except LDL, serving as a high-affinity ligand with the LDL, LRP, and apoE receptors. A single gene controls ApoE expression, with three common alleles, ε2, ε3, and ε4, giving rise to the major isoforms, E2, E3, and E4. Six different phenotypes result: E2/E2, E2/E3, E2/E4, E3/E3, E3/E4, and E4/E4. E3/E3 is the most common, being found in 50–70% of populations. Isoforms differ in amino acid composition, apoE2 having a

Figure 7 Palmar crease xanthomata of dysbetalipoproteinemia (remnant hyperlipidemia). Reproduced from Hyperlipidaemia, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

cysteine-for-arginine substitution at amino acid 158. This reduces binding to the apoE receptor to $\sim 1\%$ of normal. Homozygosity for apoE2 is the commonest genotype in remnant hyperlipidemia. However, the abnormality does not usually lead to hyperlipidemia, but rather to marginally lower LDL cholesterol levels, for it is only with secondary factors or an additional primary lipemia that the syndrome is clinically expressed.

Dietary Effects

Body weight

Obesity is an often ignored factor in hyperlipidemia, but it is important, not least because it is linked to other CHD risk factors such as blood, diabetes, and glucose intolerance. Weight control is important in treating hyperlipidemia, coupled with increased exercise.

Fatty Acids and Cholesterol

In Western societies, reducing saturated fat intake as a percentage of total food energy is associated with a reduction in total cholesterol and, to a lesser extent, triglycerides. Levels of LDL fall, possibly as a result of a direct reduction of cholesterol synthesis in the liver, or increased excretion as bile acids. If saturated fatty acids, such as palmitic acid, are replaced by mono- or polyunsaturated fatty acids, such as oleic or linoleic acids, similar falls in serum cholesterol occur.

Polyunsaturated fats of the n-6 series are primarily of plant origin. In addition to the above, they help reduce cholesterol by taking the place in the diet of saturated fats, which tend to be of animal origin and high in cholesterol. The n-3 series of fatty acids come mainly from oily fish. These fatty acids may also have beneficial effects on hemostatic variables of benefit in thrombosis prevention. They can lower triglyceride levels in some moderately hypertriglyceridemic individuals, but may have the disadvantage of increasing LDL cholesterol. Plant sterols and stanols may also reduce LDL by $\geq 10\%$.

The assumption that the amount of cholesterol in the diet is the major factor causing hyperlipidemia is false. It is the total fat intake as a percentage of the total energy intake, and also whether or not the fat is saturated, that matters most. Only a proportion of cholesterol in the diet is absorbed, and for each 100 mg eaten, the serum cholesterol will rise by only approximately $0.1\ \text{mmol}\,\text{l}^{-1}$. This response to dietary cholesterol shows a considerable variation between individuals.

Carbohydrate and Fiber

A high carbohydrate intake (particularly refined carbohydrate and simple sugars) may induce hypertriglyceridemia. Substantial amounts of soluble or mucilaginous fiber in the diet can lower LDL cholesterol, possibly by increasing the fecal loss of sterols. Individuals who eat a high-fiber diet also tend to eat less saturated fat, and less total fat.

Alcohol

Alcohol, a rich source of energy, can contribute to obesity and hypertriglyceridemia, and increase hepatic triglyceride synthesis and VLDL secretion. Substrates that undergo oxidation in the liver have to compete with alcohol, and therefore tend to be available for increased triglyceride synthesis. The tendency to develop severe hypertriglyceridemia secondary to alcohol is variable, but the risk of acute pancreatitis is also increased.

See also: **Bile**; **Cholecalciferol**: Physiology; **Cholesterol**: Role of Cholesterol in Heart Disease; **Choline**: Properties and Determination; **Essential Fatty Acids**; **Fats**: Classification; Occurrence; **Fatty Acids**: Properties; Metabolism; Gamma-linolenic Acid; **Fish Oils**: Dietary Importance; **Hormones**: Steroid Hormones; **Lipoproteins**; **Phospholipids**: Properties and Occurrence; **Triglycerides**: Structures and Properties

Further Reading

Barter PJ and Rye K (ed.) (1999) Plasma lipids and their role in disease. In: *Advances in Vascular Biology*, vol. 5. Chur, Switzerland: Harwood Academic Publishers.

Betteridge DJ (ed.) (1990) Lipid and lipoprotein disorders. In: *Ballière's Clinical Endocrinology and Metabolism*, vol. 4, no. 4. London: Ballière Tindall.
Betteridge DJ, Illingworth DR and Shepherd J (ed.) (1999) *Lipoproteins in Health and Disease*. London: Arnold.
Durrington PN (1989) *Hyperlipidaemia: Diagnosis and Management*. Oxford: Wright Butterworth.
Packard C and Caslake M (1996) Mixed hyperlipidaemia and the lipid shuttle. *Issues in Hyperlipidaemia* 13: 1–6.
Reaven GM (1988) Banting Lecture 1988. Role of insulin resistance in human disease. *Diabetes* 37(12): 1595–1607.
Syvanne M and Taskinen M (1997) Lipids and lipoproteins as coronary risk factors in non-insulin dependent diabetes mellitus. *The Lancet* 350(supplement 1): 20–23.
Thompson GR (1999) The proving of the lipid hypothesis. *Current Opinion in Lipidology* 10: 201–205.

HYPERTENSION

Contents

Physiology

J J Klawe and M Tafil-Klawe, The Ludwik-Rydygier Medical University, Bydgoszcz, Poland

Introduction

An elevated blood arterial blood pressure is one of the more important public health problems, being asymptomatic, easily detectable, and often leading to lethal complications.

Definition

In adults a diastolic pressure below 85 mmHg is normal, between 85 and 89 mmHg is high normal, 90–104 mmHg is mild hypertension, 105–114 mmHg moderate hypertension, and 115 mmHg or greater is severe hypertension. When the diastolic pressure is below 90 mmHg, a systolic pressure below 140 mmHg indicates normal blood pressure, between 140 and 159 mmHg is borderline isolated systolic hypertension, 160 mmHg or higher is isolated systolic hypertension. Patients with arterial hypertension and no definable cause are said to have primary, essential, or idiopathic hypertension. More than 90% of all patients with arterial hypertension are considered to have essential or primary hypertension.

The various forms of arterial hypertension are listed below as a simplified classification of arterial hypertension:

1. Systolic hypertension with wide pulse pressure
   a. Decreased compliance of aorta
   b. Increased stroke volume
2. Systolic and diastolic hypertension (increased peripheral vascular resistance)
   a. Renal
   b. Endocrine
   c. Neurogenic
   d. Miscellaneous
   e. Unknown etiology

Individuals with a specific organ defect which is responsible for hypertension are defined as having a secondary form of hypertension.

The following information relates to the primary form of arterial hypertension. Cardiac output and peripheral vascular resistance are two basic hemodynamic parameters regulating mean arterial pressure. Cardiac output is usually normal in patients with essential hypertension. Elevated blood pressure appears to be associated with increased peripheral vascular resistance. Some investigators have speculated that the increased resistance may be due to lack of nitric oxide, the locally produced vasodilator that is produced by endothelial cells.

Genetic Factors in Hypertension

Animal studies and human investigations have indicated an important role of genetic factors in essential hypertension. Most studies support the concept that inheritance is probably multifactorial or that a number of different genetic defects each have elevated blood pressure as one of their phenotypic expressions.

Three monogene defects – glucocorticoid-remediable aldosteronism (GRA), syndrome of apparent mineralocorticoid excess (AME), and Liddle's syndrome – and susceptibility genes such as angiotensinogen gene (AGT), renin gene (RN), pseudohipoaldosteronism type II, and α-adducin gene (ADDI) have been reported which have as one of their consequences increased arterial pressure.

Environmental Factors

A number of environmental factors are involved in the development of hypertension. These include: obesity, salt intake, alcohol intake, family size, and occupation. The influence of these factors increases with the age of patients.

Salt Sensitivity

In hypertensive population the blood pressure in about 60% is responsive to level of sodium intake. The cause of this special sensitivity to salt varies, and includes primary aldosteronism, bilateral renal artery stenosis, and other factors.

Renin and its Importance in Hypertension

Renin is an enzyme secreted by the juxtaglomerular cells of the kidney. It interacts with aldosterone in a negative-feedback loop. Some hypertensive patients are defined as having low-renin and high-renin essential hypertension. About 20% of hypertensive patients have suppressed plasma renin activity. Approximately 15% of patients with essential hypertension have plasma renin activity levels above the normal range, which suggests that in these patients plasma renin plays a significant role in the pathogenesis of arterial hypertension, activating the renin–angiotensin–aldosterone axis. However other findings have led some investigators to postulate that the elevation of blood pressure and elevated renin levels may be secondary to an increase in adrenergic system activity.

Cell Membrane Defect

The hypothesis about the role of a generalized cell membrane defect in the pathogenesis of salt-sensitive hypertension is derived from data from studies on red blood cells, in which abnormalities in the transport of sodium across the cell membrane have been documented. This abnormality in sodium transport reflects an alteration in cell membrane, and this defect may occur in all cells of the body. An abnormal accumulation of calcium in vascular smooth muscle results in an increase in vascular responsiveness to vasoconstrictor agents.

Reflex Mechanisms in Regulation of Blood Pressure in Arterial Hypertension

The reflex mechanisms regulating blood pressure include: arterial baroreceptors, cardiopulmonary mechanoreceptors, arterial chemoreceptors, and visceral receptors localized in the kidney, liver, and intestine.

The baroreceptor reflex and from cardiopulmonary mechanoreceptors are primary homeostatic controls for blood pressure. Stretch-sensitive mechanoreceptors are tonically active and thereby tonically decrease a sympathetic activity. Additionally, local release of NO from vascular epithelium after each increase in blood pressure counteracts blood pressure elevation. The cardiac response to baroreceptor activation is vagal activation and a decrease in heart rate. Several studies have shown the impairment of baroreceptor function in arterial hypertension.

The carotid chemoreceptors, activated by partial pressure of oxygen in arterial blood and its decrease, showed an augmented resting drive in arterial hypertension. The pressor response to hypoxic stimulation is also augmented in young patients with essential hypertension. Reflex from arterial chemoreceptors is probably involved in the pathogenesis of arterial hypertension in obstructive sleep apnea syndrome. The episodes of nocturnal hypoxia during sleep seem to be responsible for an increase in sympathetic activity and blood pressure.

The kidney–kidney reflex inhibits sympathetic activity, increasing natriuresis and diuresis.

The osmoreceptors in the liver and in the digestive system increase natriuresis and diuresis, decreasing blood pressure.

The disturbance of interaction between the reflex regulatory mechanisms is discussed as a factor which may be involved in the pathogenesis of essential arterial hypertension.

Hypertension is a risk factor for atherosclerosis because high pressure in arteries damages the endothelial lining of the vessels and promotes the formation of atherosclerotic plaques. Additionally, elevated arterial pressure increases cardiac afterload.

See also: **Atherosclerosis**; **Enzymes**: Functions and Characteristics; **Hypertension**: Hypertension and Diet; Nutrition in the Diabetic Hypertensive; **Renal Function and Disorders**: Kidney: Structure and Function

Further Reading

Klawe JJ, Tafil-Klawe M, Peter JH and Smietanowski M (1999) Cardiac response study towards activation and inactivation of the carotid baroreceptor in obstructive sleep apnea patients. *Medical Science Monitoring* 5(3): 449–451.

Narkiewicz K and Somers VK (1997) The sympathetic nervous system and obstructive sleep apnea: implications for hypertension. *Journal of Hypertension* 15: 163–1619.

Person PB (1996) Modulation of cardiovascular control mechanisms and their interaction. *Physiological Review* 76: 193–244.

Przybylski J (1981) Do arterial chemoreceptors play a role in the pathogenesis of hypertension? *Medical Hypotheses* 20: 173–177.

Tafil-Klawe M, Raschke F and Hildebrandt G (1990) Functional asymetry in carotid sinus cardiac reflexes in humans. *European Journal of Applied Physiology and Occupational Physiology* 60: 402–405.

Tafil-Klawe M, Raschke F, Becker H *et al.* (1991) Investigations of arterial baro- and chemoreflexes in patients with arterial hypertension. In: Peter JH, Penzel T, Podszus T and v Wichert P (eds) *Sleep and Health Risk*, pp. 319–334. Springer Verlag.

Trzebski A (1992) Arterial chemoreceptor reflex and hypertension. *Hypertension* 19: 562–566.

Trzebski A, Tafil M, Zoltowski M and Przybylski J (1982) Increased sensitivity of the arterial chemoreceptor drive in young men with mild hypertension. *Cardiovascular Research* 16: 162–172.

Williams GH (2001) Hypertensive vascular disease. In: *Harrison's Principles of Internal Medicine*, pp. 1380–1394. New York: McGraw-Hill Book Company.

Hypertension and Diet

B M Y Cheung and T C Lam, University of Hong Kong, Pokfulam, Hong Kong

Nutrition as a Cause of Hypertension

Hypertension is a common condition that affects more than one-tenth of the adult population. The complications of untreated hypertension include stroke, myocardial infarction, cardiac and renal failure. Its etiology is complex and probably involves the interaction of genetic predisposition and the environment in the widest sense. There is familial clustering of hypertension, which can be a reflection of both shared genes and shared environment or both. As a person's genes is probably not open to modification, especially when the disease-causing gene(s) have not been identified, prevention and treatment of hypertension should be targeted at altering the environmental factors. Results from many population and migrant studies show that environmental factors play an important role in the occurrence of hypertension. One such factor determining the development of hypertension is undoubtedly nutrition. Poor nutritional habits, alcohol consumption, excessive sodium and insufficient potassium intake are some of the nutritional factors leading to hypertension. These, together with socioeconomic status, physical inactivity, psychological stress, and obesity, influence blood pressure in an individual.

Fetal Origin of Hypertension

Barker and colleagues have put forward the 'fetal origins hypothesis' that links cardiovascular disease manifested in adulthood to poor nutrition *in utero* and infancy. Fetal undernutrition in middle to late gestation is thought to result in adaptive changes in fetal growth that predispose to coronary heart disease. The evidence comes from a longitudinal study of 25 000 middle-aged Britons. Those who were small at birth and not premature had relatively high rates of coronary heart disease, hypertension, hypercholesterolemia, and diabetes. Postnatal nutrition also appears to influence blood pressure in later life. Premature babies randomized in a study to breast milk had lower blood pressure in adolescence than those randomized to formula milk. If the fetal origins hypothesis, which is difficult to prove prospectively, is correct, optimizing prenatal care may prevent hypertension and other cardiovascular diseases in adulthood.

Dietary Pattern and Hypertension

There are striking differences in the blood pressure of populations worldwide. Blood pressure is higher and rises more steeply with age in industrialized than in nonindustrialized societies. A predominantly vegetarian dietary pattern is generally present in those populations that have low average blood pressure. In industrialized countries, vegetarians have lower blood pressures than nonvegetarians (**Figure 1**).

Epidemiological and intervention studies including the Multiple Risk Factor Intervention Trial (MRFIT) suggest a multitude of dietary changes that may reduce blood pressure. The Dietary Approaches to Stop Hypertension (DASH) Trial tested three dietary patterns: (1) a control diet that resembled the average US diet; (2) a combination diet high in fruit,

vegetables, whole-grain cereal products, low-fat dairy products, fish, chicken, and lean meats but low in fat and cholesterol; and (3) a fruit and vegetables diet that tested the effects of fruit and vegetables alone and was similar in other nutrients to the control diet. Sodium content was the same in all the diets and the subjects maintained their prestudy body weight throughout the trial. The combination diet was found to reduce the systolic pressure by 11 mmHg and the diastolic pressure by 6 mmHg (Figure 2). The fruit and vegetables diet reduced blood pressure by half this amount. Remarkably, the effect of the combination diet in the hypertensives approaches the effect of treatment with a single antihypertensive drug.

Sodium

Sodium is tightly regulated through a variety of coordinated homeostatic mechanisms involving the sympathetic nervous system, the renin–angiotensin system, and natriuretic peptides. A minute rise in blood pressure is sufficient to augment sodium excretion to achieve sodium balance in normal kidneys. In salt-sensitive hypertension, there is an exaggerated blood pressure response to increased sodium intake. Older patients, overweight type 2 diabetics, or those of African descent tend to have hypertension of this type. They have low renin levels and a lesser pressor response to sodium deprivation. In such hypertensive individuals, sodium restriction is effective and beneficial. The controversy surrounding salt and hypertension stems from the fact that not everyone is salt-sensitive and it has been argued that reducing the salt intake in the general population is neither practical nor effective, and may even be harmful to some people. However, the relationship between salt and blood pressure is now well established (Figure 3), and the higher the salt intake of the population, the steeper the rise in blood pressure with age. Interestingly, in primitive communities, blood pressure does not rise with age.

In many western countries, daily sodium intake is between 150 and 200 mmol. The Trial Of

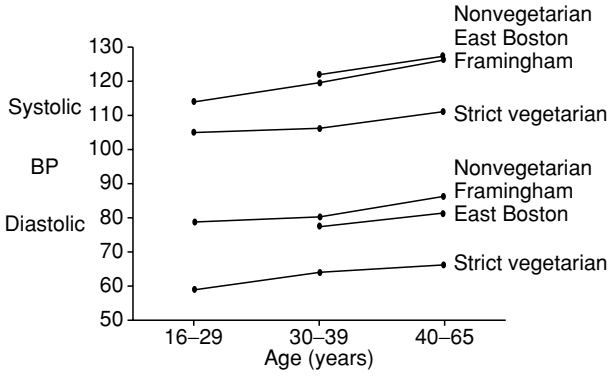

Figure 1 Blood pressure (BP) in a strict vegetarian population in Boston and in nonvegetarian populations in East Boston and Framingham, MA. Adapted from Sacks FM and Kass EH. Low blood pressure in vegetarians: effects of specific food and nutrients. *American Journal of Clinical Nutrition* 48: 795–800, with permission.

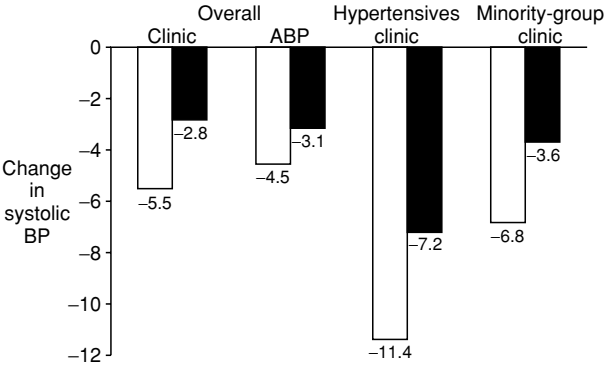

Figure 2 Effect of dietary patterns on blood pressure (BP) in the Dietary Approaches to Stop Hypertension (DASH) study. ABP, ambulatory blood pressure. Open columns, combination diet; filled columns, fruit and vegetable diet. Adapted from Appel LJ, Moore TJ, Obarzanek E *et al*. (1997) A clinical trial of the effects of dietary patterns on blood pressure. DASH Collaborative Research Group. *New England Journal of Medicine* 336(16): 1117–1124, with permission.

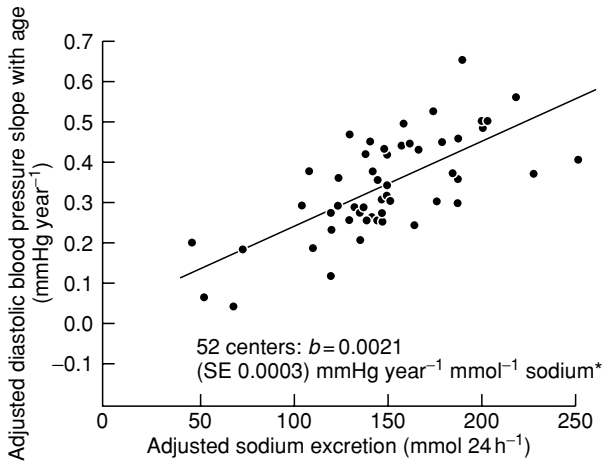

Figure 3 Cross-center plots of diastolic pressure slope with age and median sodium excretion and fitted regression lines for 52 centers, also adjusted for body mass index and alcohol intake. *$P < 0.001$. Reproduced from Intersalt Cooperative Research Group (1988). Intersalt: an international study of electrolyte excretion and blood pressure: results for 24-hour urinary sodium and potassium excretion. *British Medical Journal* 297: 319–328, with permission.

Nonpharmacological Intervention in the Elderly (TONE) showed that a 40 mmol day^{-1} reduction is accompanied by significant reductions in blood pressure (5.3/3.4 mmHg). This modest salt restriction can be achieved by not adding salt at the table and avoiding foods that are high in sodium content (e.g., preserved food). Saltiness is a rather crude taste and is used by the industry to enhance the palatability of food cheaply and to boost the sale of beverages. Historically, salt was needed to preserve food for the winter months but nowadays, refrigeration renders this unnecessary. Fresh food is usually low in salt content and its flavors are better appreciated when not too much salt has been added in cooking.

Carefully controlled clinical studies have demonstrated a dose-dependent relationship between dietary salt intake and blood pressure, although trials of salt restriction do not always show a useful reduction in blood pressure due to their various trial designs. Salt restriction is particularly important when the patient is treated with an angiotensin-converting enzyme inhibitor (ACEI) or an angiotensin II antagonist (sartan). The efficacy of these drugs is diminished in the presence of high salt intake that suppress the renin–angiotensin system.

Potassium

In contrast to sodium, potassium is less tightly regulated physiologically. In diuretic therapy, the renal tubules will conserve sodium at the expense of potassium. Plasma potassium concentration is also linked to acid–base balance. Nevertheless, potassium is also closely associated with blood pressure. Cross-sectional studies in many countries worldwide had identified an inverse relationship between blood pressure and various measures of serum, urine, total body and dietary potassium. Low potassium intake may also underlie the high incidence and prevalence of hypertension in blacks and the elderly. Potassium supplementation alone will decrease blood pressure slightly. The effect is more pronounced in hypertensives than normotensives, in blacks than in whites, and in those with high sodium intake. In reality, higher potassium intake is usually associated with a diet rich in fruit and vegetables. Such a diet will in itself be conducive to a lower blood pressure, so one cannot always disentangle the components of a 'healthy diet' which are actively antihypertensive. High sodium intake and low potassium intake is a common scenario amongst hypertensive patients. Therefore, sodium and potassium should be considered jointly in our dietary recommendations. For example, using a salt substitute that contains potassium chloride increases potassium intake whilst reducing sodium intake.

Calcium

The relationship between calcium and hypertension is controversial. Dietary intake of calcium tends to be lower in hypertensives and calcium supplementation is associated with a modest reduction in blood pressure. A metaanalysis of 23 observational studies showed an inverse association between blood pressure and dietary calcium intake. The effect size was however small and heterogeneous across studies. Metaanalysis of randomized trials of calcium supplementation showed that systolic blood pressure is reduced by around 1 mmHg. Although it is possible that certain subgroups may be more sensitive to the effects of calcium, it is certainly not the major factor in the pathogenesis of hypertension, despite the recent identification of a parathyroid hypertensive factor. Therefore, calcium supplementation, though desirable for other reasons, is not currently recommended as an efficacious means of treating hypertension.

Fat

The relationship between lipids and coronary heart disease is now proven beyond doubt by epidemiological studies and large-scale trials of lipid-lowering drugs (statins) that reduce coronary events. Reduction in cholesterol by pharmacological means may not lower blood pressure, but should be viewed in the context of the overall reduction of cardiovascular risk and the reduction of complications of hypertension such as myocardial infarction. Hypertension and dyslipidemia are independent cardiovascular risk factors and, when they are both present in the same person, the cardiovascular risk is augmented. Therefore, it makes sense to address both risk factors. Nonpharmacological means of lipid lowering through diet and exercise is likely to have a beneficial effect on blood pressure in addition to improving the cardiovascular risk profile. The DASH study suggested that a healthy diet based on less fat and more fruit and vegetables reduces blood pressure, although precisely which component of the DASH diet lowers blood pressure remains to be elucidated.

The effects of saturated, monosaturated, and polyunsaturated fatty acids and carbohydrates have been studied in many clinical trials. Omega-3 unsaturated fatty acids (fish oils) reduce blood pressure but a large intake is needed, so this is not a practical treatment for hypertension.

Obesity

It has become clear in recent years that obesity is one of the major determinants of blood pressure. Our own observations in relatives of hypertensive patients suggest that the occurrence of hypertension is related to whether the relative is obese or not. It therefore appears that those who may be genetically predisposed to hypertension will develop the condition if they become obese.

Recently, two genes involved in the control of body weight, the OB gene and the OB receptor gene, which code for the hormone leptin and its receptor respectively, have been described in the mouse. Deletions in the OB gene or the OB receptor lead to profound obesity of early onset with excessive food intake, decreased energy expenditure, and insulin resistance. In humans, mutations in the corresponding genes have been reported. A high level of circulating leptin is found in the majority of obese people, suggesting that they have leptin resistance. Transgenic mice overexpressing leptin develop hypertension. However, the association of high leptin levels and hypertension has not been firmly established in humans, so the relevance of these genes to the pathogenesis of essential hypertension remains to be elucidated.

Obesity, especially abdominal adiposity, is more closely related to blood pressure than body weight. Moreover, obesity can lead to insulin resistance and ultimately overt type 2 diabetes mellitus. Lean, normoglycemic untreated hypertensive subjects are more insulin-resistant than comparable normotensive subjects. The relationship between insulin resistance and hypertension may involve a variety of mechanisms, including increased sympathetic nervous system activity, proliferation of vascular smooth-muscle cells, altered cation transport, and increased sodium retention. It is worth noting than one-third of diabetics have hypertension and a significant proportion of hypertensives have diabetes. The two conditions overlap to a large extent as components of the 'metabolic syndrome' or 'syndrome X.'

Measurement of blood pressure in those who are obese and have large arms is prone to overestimation if the cuff is not large enough. Nevertheless, the relationship between obesity and blood pressure is well established and indeed, for the overweight hypertensive, losing weight, if it can be achieved, is the most efficacious means of reducing blood pressure amongst the various nonpharmacological treatment options. Our own data and those of other investigators suggest that for each kilogram of body weight lost, there is approximately a 1 mmHg reduction in diastolic blood pressure and a 2 mmHg reduction in systolic blood pressure (Figure 4). Unfortunately, outside clinical trials, this is easier said than done. A coordinated approach involving a physician, nurse, and dietitian and attendance of classes is probably more effective than advice from a single professional. As with other nonpharmacological measures, weight control through diet and exercise will lead to benefits in addition to blood pressure control, such as a better lipid profile, better cardiovascular fitness, less stress on the joints, and so on.

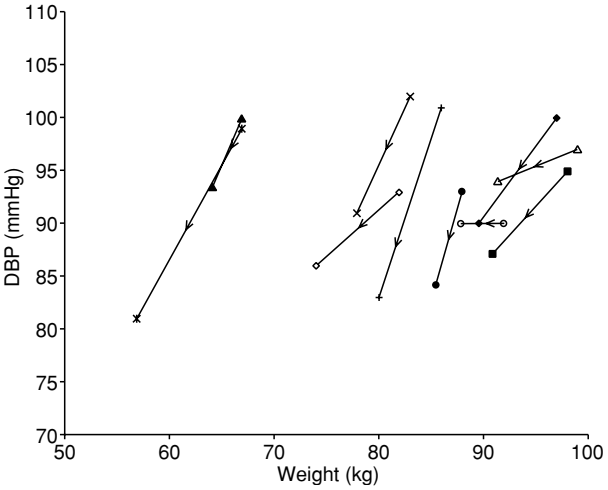

Figure 4 The effect of weight reduction on diastolic blood pressure (DBP). (◆) MacMahon *et al*. (1985) *Lancet* 1: 1233–1236; (■) Gordon *et al*. (1997) *American Journal of Cardiology* 79: 763–767; (▲) Singh *et al*. (1995) *Journal of Human Hypertension* 9: 355–362; (×) Jalkanen *et al. Scandinavian Journal of Social Medicine* 19: 66–71; (★) Singh *et al*. (1990) *Nutrition* 6: 297–302; (●) Wassertheil-Smoller S *et al*. (1992) The Trial of Antihypertensive Interventions and Management (TAIM) study. *Archives of Internal Medicine* 152(1): 131–136; (|) Darne *et al*. (1993) *Blood Pressure* 2: 130–135; (○) Haynes *et al*. (1984) *Journal of Hypertension* 2: 535–539; (△) Fagerberg *et al*. (1989) *British Medical Journal* 299: 480–485; (◇) Andersson *et al*. (1991) *Hypertension* 18: 783–789.

Alcohol

The effect of alcohol on health in general and cardiovascular diseases in particular is dose-dependent. A modest regular intake is probably beneficial because of the increase in high-density lipoprotein cholesterol (HDL-C) and the relief of stress whilst a high intake is definitely harmful. With increasing levels of alcohol intake, more than 50 epidemiological studies from a variety of cultures have reported an increase in blood pressure or a higher prevalence of hypertension. An excessive alcohol intake raises blood pressure and moderation of alcohol intake leads to reduction in blood pressure (Table 1). In practice, hypertensive patients who do not drink alcohol should not be encouraged to start doing so whilst those who do drink large quantities must be encouraged to reduce

Table 1 Randomized controlled trials of the effects of alcohol reduction on blood pressure (BP)

Study, year	*Study population*		*Study results*				
	n	*Age (years) (mean ± SD or range)*	*Duration (weeks)*	*Baseline BP (mm Hg)*	*Alcohol intake difference (drinks[a] per day)*	*BP difference (mm Hg)*	P
Puddey, 1985	46	35±8	6	133/76	3.7	3.8/1.4	< 0.001/ < 0.05
Howes, 1985	10	25–41	0.6	120/66	5.7	8/6	< 0.025/ < 0.001
Puddey, 1987	44	53±16	6	142/84	4.0	5/3	< 0.001/ < 0.001
Ueshima, 1987	50	46±7	2	148/93	2.6	5.2/2.2	< 0.005/NS
Wallace, 1988	641	42±20	52	136/82	1.0	2.1/?	< 0.05/NS
Parker, 1990	59	52±11	4	138/85	3.8	5.4/3.2	< 0.01/0.01
Cox, 1990	72	20–45	4	132/73	3.4	4.1/1.6	< 0.05/ < 0.05
Maheswaran, 1992	41	40s	8	144/90	3.1	Not reported	NS
Puddey, 1992	86	44	18	137/85	3.0	4.8/3.3	< 0.01/ < 0.01
Ueshima, 1993	54	44±8	3	144/96	1.7	3.6/1.9	< 0.05/NS
PATHS, 1998	641	57±11	104	140/86	1.3	0.9/0.6	0.16/0.10

[a]A standard drink is defined as 14 g ethanol and is contained in a 12-oz/350 ml glass of beer, a 5-oz/146 ml glass of table wine, or 1.5/44 ml oz of distilled spirits.

their intake to the recommended levels. Currently, it is believed that men should not drink more than 21 units per week and women should not drink more than 14 units per week. Approximately one unit of alcohol, or 14 g ethanol, is contained in one glass of table wine.

Dietary Approach to Prevent or Treat Hypertension

Alteration of the diet to bring about a lower blood pressure is not only beneficial in those who are hypertensive, but is prudent in those who are not yet hypertensive and even in those who are normotensive. A shift in the distribution of the blood pressures to a lower mean for the population is expected to lead to a massive reduction in the incidence of coronary heart disease and stroke. Achieving a healthier diet will delay or even prevent the onset of hypertension. This may be especially relevant in young people who have a strong family history of hypertension or cardiovascular diseases. In those who have newly diagnosed hypertension, dietary intervention may obviate the need for drug therapy. In patients with mild hypertension who respond to diet and the nonpharmacological approach, cessation of therapy may be possible. In those hypertensive patients who still require drug therapy, an appropriate diet will reduce the intensity of treatment and facilitate blood pressure control.

There is no single diet that will fit everyone's needs. The elderly might be more responsive to salt restriction whilst the young obese hypertensive should have a diet low in fat and calories. Clinicians rarely have the time or the training to go through a diet with the patient so the help of dietitians and other health care professionals is crucial for successful implementation of dietary intervention.

Although most of the dietary recommendations are well accepted by health care professionals and the community, their implementation is a major problem since a change in behavior is needed. Eating and alcohol drinking are normally pleasurable experiences and many do not find regular exercise enjoyable, so diet, abstinence, and exercise require determination and commitment. There is a limit as to how much health care professionals can do in this regard. It is therefore important that the doctor, nurse, and dietitian complement each other and reinforce the message in a team approach.

There is another obstacle to a healthy diet: food sold in fast-food outlets is convenient, inexpensive, and therefore popular. Such food is high in salt and fat content and is generally not fresh. Tinned and other preserved food has a high salt content; McGregor has argued that reducing the salt intake in the general population requires the cooperation of the food industry to reduce the salt content in its products. Health care professionals need to be aware of the economical and social aspects in order to give advice that can feasibly be followed. In developing countries, the highest incidence of hypertension is found in people with high socioeconomic status. As countries become more affluent, hypertension becomes more prevalent in the lower socioeconomic classes. In developed countries, there is an inverse relationship between socioeconomic status and blood pressure. Proposed mechanisms include differences in diet (sodium, potassium, calories), physical activity, body mass, alcohol intake, psychosocial stress, and access

to health care (including acceptance of and use of services). In some respects, unhealthy diet is a form of malnutrition that afflicts poor people in developed countries. The striking differences in diseases between the socioeconomic classes, usually defined in terms of educational level and income, are clearly environmental rather than genetic in origin. Therefore, there is much the society, through governmental and nongovernmental agencies, can do to reverse this.

A healthier diet brings many dividends – it reduces calorie, sodium, cholesterol, and triglyceride intake and increases the intake of potassium and calcium, and the consumption of fruit and vegetables which are rich in fiber and vitamins. Therefore, we should strive to shift the dietary habit of the whole population, including children and adolescents, as much as patients with hypertension and other cardiovascular conditions. A family rather than an individual approach is often called for if the family eats together. The advantage of this is that the whole family can start to adapt to a healthy diet. Primary prevention of hypertension may then become a reality rather than just a possibility.

See also: **Alcohol**: Properties and Determination; **Calcium**: Properties and Determination; Physiology; **Hypertension**: Physiology; **Obesity**: Etiology and Diagnosis; **Potassium**: Properties and Determination; Physiology; **Sodium**: Properties and Determination; Physiology

Further Reading

Alderman MH (2000) Salt, blood pressure, and human health. *Hypertension* 890–893.

Allender PS, Cutler JA, Follmann D *et al.* (1996) Dietary calcium and blood pressure: a meta-analysis of randomized clinical trials. *Annals of Internal Medicine* 124: 825–831.

Appel LJ, Moore TJ, Obarzanek E *et al.* (1997) A clinical trial of effects of dietary patterns on blood pressure: DASH collaborative research group. *New England Journal of Medicine* 336: 1117–1124.

Ascherio A, Hennekens CH, Willet WC *et al.* (1996) Prospective study of nutritional factors, blood pressure, and hypertension among US women. *Hypertension* 27: 1065–1072.

Ascherio A, Rimm EB, Givovannucci E *et al.* (1992) A prospective study of nutritional factors and hypertension among US men. *Circulation* 86: 1475–1484.

Barker DJP, Osmond C, Golding J, Kuh D and Wadsworth MEJ. Growth *in utero*, blood pressure in childhood and adult life, and mortality from cardiovascular disease. *British Medical Journal* 298: 564–567.

Considine RV, Sinha MK, Heiman ML *et al.* (1996) Serum immunoreactive leptin concentrations in normal-weight and obese humans. *New England Journal of Medicine* 334: 292–295.

Fagard RH (1993) Physical fitness and blood pressure. *Journal of Hypertension* 11 (suppl. 5): S47–S52.

Gillum RF, Mussolino ME and Madans JH (1998) Body fat distribution and hypertension incidence in women and men. The NHANES I epidemiologic follow-up study. *International Journal of Obesity Related Metabolic Disorders* 22: 127–34.

Hypertension Detection and Follow-up Program Cooperative Group (1987) Educational level and 5-year all cause mortality in the hypertension detection and follow-up program. *Hypertension* 9: 641–646.

Intersalt Cooperative Research Group (1988) Intersalt: an international study of electrolyte excretion and blood pressure: results for 24-hour urinary sodium and potassium excretion. *British Medical Journal* 297: 310–328.

Kotchen TA, McCarron DA for the Nutrition Committee (1998) Dietary electrolytes and blood pressure. A statement for healthcare professionals from the American Heart Association Nutrition Committee. *Circulation* 98: 613–617.

Law MR, Frost CD and Wald NJ (1991) By how much does dietary salt reduction lower blood pressure? *British Medical Journal* 302: 811–824.

MacGregor GA and De Wardener HE (1998) *Salt Diet and Health*. Cambridge: Cambridge University Press.

Pang PK, Shan JJ, Lewanczuk RZ and Benishin CG (1996) Parathyroid hypertensive factor and intracellular calcium regulation. *Journal of Hypertension* 14: 1053–1060.

Poulter NR, Khaw KT, Hopwood BE *et al.* (1990) The Kenyan Luo migration study: observations on the initiation of a rise in blood pressure. *British Medical Journal* 300: 967–972.

Reaven GM, Lithell H and Landsberg L (1996) Hypertension and associated metabolic abnormalities. The role of insulin resistance and the sympathoadrenal system. *New England Journal of Medicine* 334: 374–381.

Sacks FM, Brown LE, Appel L *et al.* (1995) Combinations of potassium, calcium, and magnesium supplements in hypertension. *Hypertension* 26: 950–956.

Singhal A, Cole TJ and Lucas A (2001) Early nutrition in preterm infants and later blood pressure: two cohorts after randomised trials. *Lancet* 357: 413–419.

Stamler J, Caggiula A, Grandits GA, Kjelsberg M and Cutler JA for the MRFIT Research Group (1996) Relationship of blood pressure of combinations of dietary macronutrients. Findings of the Multiple Risk Factor Intervention Trial (MRFIT). *Circulation* 94: 2417–2423.

Whelton PK, He J, Cutler JA *et al.* (1997) Effects of oral potassium on blood pressure: meta-analysis of randomized controlled clinical trials. *Journal of the American Medical Association* 277: 1624–1632.

Whelton PK, Appel LJ, Espeland MA *et al.* (1998) TONE Collaborative Research Group. Sodium reduction and weight loss in the treatment of hypertension in older persons: a randomized controlled trial of nonpharmacologic interventions in the elderly (TONE). *Journal of the American Medical Association* 279: 839–846.

Nutrition in the Diabetic Hypertensive

S C Bain, University of Birmingham and Birmingham Heartlands Hospital, Bordesley Green East, Birmingham, UK
P M Dodson, Birmingham Heartlands Hospital, Bordesley Green East, Birmingham, UK

Hypertension as a Secondary Complication of Diabetes

Diabetes mellitus and hypertension are frequently associated: hypertension is approximately twice as common in subjects with diabetes mellitus compared with the general population. The prevalence of hypertension increases with age and is relatively greater in those with type 1 (insulin-dependent) diabetes than in those with type 2 (noninsulin-dependent) diabetes after adjustment for age. As in the general population, hypertension occurs more frequently in diabetic males than in diabetic females before the fifth decade and more frequently in diabetic women thereafter.

The etiology of hypertension in association with diabetes is poorly understood and is probably different in the two types of diabetes. Hypertension in type 1 diabetes is generally related to the development of diabetic kidney damage (diabetic nephropathy) which occurs in up to 40% of subjects. Diabetic nephropathy may also occur in type 2 diabetes but is less common (less than 20%). This form of hypertension may rightly be regarded as a secondary complication of diabetes.

In contrast, the majority of type 2 diabetic subjects with raised blood pressure have essential hypertension. This may be a consequence of the diabetic state but is more likely to be an association due to a common etiologic factor such as insulin resistance (see below). Type 2 diabetic patients are also prone to the development of isolated systolic hypertension but other 'secondary' causes of hypertension do not occur with increased frequency in the diabetic population.

Factors Associated with Hypertension in the Diabetic

Essential Hypertension

The reported prevalence of essential hypertension (previously defined as blood pressure more than 160/95 mmHg) among newly diagnosed type 2 diabetic patients is 40% in men and 53% in women. This prevalence is age-related in both sexes, consistently higher in women than in men, and greater than seen in an age- and sex-matched healthy population. Obesity explains part of the increased prevalence since approximately 75% of type 2 diabetics are overweight and obesity is independently associated with raised blood pressure. However, the association of type 2 diabetes and essential hypertension has yet to be fully explained.

In 1988 Reaven introduced the term 'Syndrome X' to describe the cooccurrence of hypertension, impaired glucose tolerance, dyslipidemia (raised triglycerides, low high-density lipoprotein cholesterol), hyperinsulinemia, and insulin resistance (**Table 1**). He proposed this as a cardiovascular risk factor complex and suggested that insulin resistance was causal in the development of the other abnormalities in the syndrome. There are a number of potential mechanisms by which insulin resistance, and the consequent hyperinsulinemia, could induce hypertension. For example, increased sodium retention, disturbed sodium/potassium transport, and sympathetic nervous system overactivity have all been reported. Since obesity is a state of increased peripheral resistance to insulin, this hypothesis provides an explanation for the association of obesity with both type 2 diabetes and raised blood pressure.

Since its original description, Reaven's syndrome has continued to be the focus of much research and has expanded to include postprandial hyperlipidemia, the presence of small dense low-density lipoproteins, hyperuricemia, and abnormal clotting factors. The persistence of the syndrome implies that it is important; however, the causal mechanisms have yet to be dissected and the precise role of insulin resistance remains controversial.

The definition of hypertension is also a rapidly changing area. Controlled clinical trials of drug intervention now suggest that target levels for systolic and diastolic blood pressure in patients with type 2 diabetes should be in the order of 140/85 mmHg. The goals are lower if patients have known ischemic heart disease (130 and 80 mmHg) and are lower still in patients with type 1 diabetes and evidence of diabetic

Table 1 Original components of Syndrome X (Reaven's syndrome)

Insulin resistance
Hyperinsulinemia
Impaired glucose tolerance
Raised triglycerides
Low HDL-cholesterol
Hypertension

HDL, high-density lipoprotein.

nephropathy. These targets are irrespective of age (although data concerning patients over the age of 85 years are lacking) and imply an aggressive treatment for isolated systolic hypertension. Attention should also be given to other modifiable cardiovascular risk factors (currently smoking and hypercholesterolemia; high-density lipoprotein cholesterol may be included in the future).

Renal Hypertension

This form of hypertension is associated with diabetic nephropathy. Diabetic renal damage develops in up to 40% of type 1 diabetics and is associated with high morbidity and mortality. Raised blood pressure may be present at the clinical onset of renal disease (persistent proteinuria), or emerge during the development of nephropathy, when the glomerular filtration rate begins to decline. There is good evidence that the presence of hypertension at this stage accelerates the progression to renal failure. Hence, hypertension may develop secondarily to renal involvement in diabetes and, once manifest, accelerates the decline in renal function. The pathogenesis of hypertension in diabetic patients with overt nephropathy is unclear. However, a predisposition to hypertension and ischemic heart disease (positive family history) increases the susceptibility of type I diabetics to renal disease. This is thought to imply the involvement of genetic factors but these have yet to be identified.

Dietary Management of Hypertension in the Diabetic

Drug treatment of established hypertension is inevitably lifelong. Side-effects of antihypertensive agents are common and metabolic sequelae could theoretically contribute to cardiovascular risk (for example, by increasing insulin resistance). In addition, overzealous lowering of blood pressure with drugs may have deleterious effects, especially in the elderly and in subjects with cardiac disease. The above considerations emphasize that nonpharmacological approaches should be recommended for all subjects with mild hypertension and dietary manipulation is the most important of these.

Weight Reduction

Weight loss is the first-line therapy for obese hypertensive diabetic subjects, since this may have a beneficial effect on both blood pressure and blood glucose control. Weight reduction was shown to lower blood pressure in the Hypertension Control Program study where 39% of patients who had previously required drug therapy for hypertension were controlled by dietary manipulation alone after 4 years. However, the effect of weight loss on blood pressure can be disappointing: a study of 60 mildly hypertensive obese individuals concluded that weight loss was not important in lowering blood pressure. Despite these reservations, weight reduction remains a sensible aim since achievement of a desirable weight for height, as indicated by the body mass index, may normalize glucose tolerance in a proportion of patients with type 2 diabetes as well as reducing insulin resistance.

Dietary recall histories are often used to assess the energy intake of a patient and a deduction of 500 kcal from this figure then determines the recommended energy intake to achieve weight loss of 1–2 kg per month. Unfortunately, this method consistently underestimates daily energy intake (by as much as 800 kcal) and, by setting impossible goals for patients, may lead to poor dietary compliance. It is possible to calculate the energy intake of a weight-stable individual using a nomogram in which age, sex, exercise status, and body mass are taken into account. A realistic weight-reducing diet aims for an energy deficit of not more than 500 kcal from the intake calculated by this method.

When placed on a low-energy diet in a controlled environment all obese subjects lose weight. In the clinical setting and despite realistic dietary targets, a substantial proportion of patients do not achieve weight reduction. Even those who lose weight in the short term often have difficulty in maintaining this desirable state. In both cases, this presumably reflects poor compliance with dietary advice. Second-line therapy, such as behavioral therapy, may then be tried as an adjunct to continued attempts at low energy intake. Behavior therapy is aimed at the identification and avoidance of cues that promote eating. Excellent results can be achieved but at great cost in time to both therapist and patient. In those grossly overweight subjects who are unable to muster the self-discipline to lose more than a few kilograms, surgical procedures may be indicated. Drug therapies have now largely been withdrawn due to their poor side-effect profiles.

The 'Prudent Diet'

Irrespective of the need to lose weight, the nutritional considerations for treating individuals with diabetes mellitus and hypertension are to improve glycemic control, lower blood pressure, and reduce hyperlipidemia. The composition of the most commonly prescribed dietary regimen is high in dietary fiber (40–45 g $24\,h^{-1}$) with more than 50% of total energy as unrefined carbohydrate. Dietary fat intake is reduced to less than 25% of total energy, with less than 10% in the form of saturated fat. This is known as the

'prudent diet.' In a study of 34 diabetic patients with essential hypertension, modification of dietary regimen in this way, without attempts to reduce overall energy intake, led to a significant fall in blood pressure over 3 months. Patients also lost a mean of 2 kg in weight, possibly due to the dietary fiber promoting satiety at a lower level of calorie and fat intake. Longer-term studies (up to 4 years) indicate that the effect on blood pressure is sustained.

Not all authors agree with this form of dietary manipulation. Reaven and coworkers claim that the replacement of fat by complex carbohydrate induces a rise in serum triglyceride levels. Since hypertriglyceridemia is a common abnormality in type 2 diabetes, it is argued that the prescribed diet should not aggravate this situation. It is claimed that replacement of saturated fat with polyunsaturates, without altering the relative proportions of energy derived from fat and carbohydrate, alleviates this problem whilst retaining beneficial effects on glycemic control and serum cholesterol levels. This alternative dietary manipulation is given further support by reports that even dietary enthusiasts find it very difficult to maintain the high fiber and carbohydrate intake of the 'prudent diet.'

Sodium Intake

Sodium intake is also linked with high blood pressure. Double-blind cross-over studies have shown that moderate sodium restriction (not adding salt, avoiding salt-laden foods) produces a beneficial effect on blood pressure. Diabetics appear to be especially sensitive to salt reduction and this may be related to the findings of impaired sodium excretion and a 10% increase in total body exchangeable sodium in diabetics. However, a study of obese patients treated by weight reduction found no difference in blood pressure in those subjects who did not alter their salt intake compared with those who restricted salt.

In the UK, the majority of salt intake comes from the preparation of processed foods and patients may have a large impact on salt intake by adopting a diet rich in fruit and vegetables, thereby reducing consumption of heavily salted products. In this setting, it may be counterproductive to advise further salt restriction and run the risk of diminished dietary compliance. An exception to this is the African-Caribbean diet which is often salt-rich; in this setting the use of other herbs and spices might effect a significant reduction in sodium intake.

Alcohol

Epidemiological studies have convincingly established a link between regular alcohol consumption and high prevalence of hypertension. In individual subjects regular alcohol intake has been shown to raise blood pressure and reduction of alcohol consumption is therefore recommended to reduce blood pressure, as well as to promote weight loss by reducing calorific intake.

Other Dietary Manipulations

There are published studies of other dietary manipulations that have been shown to be of benefit in lowering blood pressure. Calcium supplementation lowers diastolic blood pressure in young people with mild hypertension and potassium supplementation has been shown to reduce blood pressure. Magnesium supplements may also benefit hypertensive patients on long-term diuretics. In untreated subjects, changing to a vegetarian diet may bring about a worthwhile fall in systolic blood pressure. Until further evidence is forthcoming, it is premature to advise supplementation with these minerals; in addition, there is clearly a limited amount of dietary advice with which patients can comply.

Recommendations

At present, most physicians in the UK would advise the following measures as the mainstay of dietary therapy:

- weight reduction, so as to achieve a stable body mass index of less than 27 kg m^{-2}
- moderate alcohol restriction (less than two units of alcohol per day with one alcohol-free day per week)
- moderate salt restriction (no more than 3 g day^{-1})

These maneuvers may delay or limit the use of drugs and in some patients will achieve adequate blood pressure control. In subjects with mild hypertension, nonpharmacological treatment alone should be tried for at least 3 months. If the patient is making good progress, for example with regard to weight loss, these measures may be pursued for longer. If the improvement has been small then drug therapy is appropriate at this stage. In subjects with severe hypertension drug therapy may be commenced immediately, since nonpharmacological treatment alone is unlikely to be effective.

Special Considerations – Diabetic Nephropathy

In type 1 diabetics with established renal disease, there is evidence that protein restriction (to less than 40 g day^{-1}), in addition to the dietary measures outlined above, will retard the progressive decline in renal function. At this stage, however, hypertension should always be treated with drug therapy as well. Indeed, many would argue for the

angiotensin-converting enzyme inhibitor class of antihypertensive drugs to be introduced as a protective measure before hypertension is established.

Prognosis

Hypertension contributes to the leading causes of morbidity and mortality in the diabetic population, namely coronary heart disease, stroke, peripheral vascular disease, lower-extremity amputations, and end-stage renal disease. Raised blood pressure will also worsen diabetic retinopathy, a major cause of blindness in the western world. Clearly, the aim of therapy is to reduce the incidence of these sequelae, especially premature cardiovascular disease, which is quantitatively the most important.

The effectiveness of antihypertensive drug therapy in reducing elevated blood pressure levels is well established. In addition, major multicenter trials of antihypertensive drug therapy have consistently demonstrated a reduction in the rate of strokes by approximately one-third and a decrease in cardiovascular mortality by around 15%. Large long-term studies of antihypertensive therapy have now been performed in diabetic populations and have confirmed that this group of patients benefits from aggressive control of hypertension and other cardiovascular risk factors. Indeed, there are many guidelines suggesting that type 1 diabetics with incipient nephropathy (microalbuminuria at high risk of developing overt diabetic nephropathy) should receive antihypertensive drug therapy whilst their blood pressure is still within the normal range.

Prevention

Until the etiology of essential hypertension and diabetic nephropathy is more fully understood, prevention will remain difficult. As yet, there is no evidence that strict control of blood glucose can prevent the development of either condition, although once present, hyperglycemia may increase proteinuria. This is in keeping with evidence that both essential hypertension and diabetic renal disease have strong genetic components. Nevertheless, it is sensible to advise all diabetics to take a diet low in saturated fat, to moderate their salt intake, and avoid excessive alcohol consumption. Regular exercise is recommended and obesity should be avoided. These measures may delay, or even prevent, the onset of hypertension in type 2 diabetics but are unlikely to affect the development of hypertension in association with diabetic nephropathy.

Finally, since the main aim of treating high blood pressure is to effect a reduction in the number of cardiovascular events, it is appropriate to focus attention on other cardiovascular risk factors such as smoking and hypercholesterolemia. These also have important dietary aspects; stopping smoking is often associated with weight gain; treatment of hypercholesterolemia is through dietary manipulation, at least in the first instance.

See also: **Alcohol**: Properties and Determination; Metabolism, Beneficial Effects, and Toxicology; **Diabetes Mellitus**: Etiology; Chemical Pathology; Treatment and Management; Problems in Treatment; Secondary Complications; **Obesity**: Etiology and Diagnosis; **Renal Function and Disorders**: Nutritional Management of Renal Disorders; **Slimming**: Slimming Diets; **Sodium**: Properties and Determination; Physiology

Further Reading

Bain SC and Dodson PM (1991) Managing the obese patient with hypertension. *Treating Overweight* 21: 2–6.

Barnett AH and Dodson PM (eds) (1990) *Hypertension and Diabetes*. London: Science Press.

Bursztyn P (1987) *Nutrition and Blood Pressure*. Southampton, UK: John Libbey.

Collins R, Peto R, MacMahon S *et al.* (1990) Blood pressure, stroke and coronary heart disease. Part 2, short term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. *Lancet* i: 827–838.

Lean MEJ and James WPT (1986) Prescription of diabetic diets in the 1980s. *Lancet* i: 723–725.

Reaven GM (1988) Role of insulin resistance in human disease. Banting lecture. *Diabetes* 37: 1595–1607.

Reaven GM (1988) Dietary therapy for non-insulin-dependent diabetes mellitus. *New England Journal of Medicine* 319(13): 862–864.

Stamler R, Stamler J, Grimm R *et al.* (1987) Nutritional therapy for high blood pressure. Final report of a four-year randomised controlled trial – the hypertension control program. *Journal of the American Medical Association* 257(11): 1484–1491.

Swales JD, Ramsey LE, Coope JR *et al.* (1989) Report of the British Hypertension Society working party. Treating mild hypertension. *British Medical Journal* 298: 694–698.

HYPOGLYCEMIA (HYPOGLYCAEMIA)

M S Boyne, University of the West Indies, Mona, Jamaica

Background

Hypoglycemia was not clinically appreciated until the discovery of insulin by Banting, Best, Collip, and MacLeod. In 1921, they noted that unpurified insulin led to the sudden death of several test animals. Subsequently, patients with type 1 diabetes treated with insulin described the symptoms now attributed to hypoglycemia. It was soon observed that some nondiabetic patients demonstrated similar symptoms, and the concept of spontaneous hypoglycemia or 'hyperinsulinism' was promoted. The cause remained unclear until 1927, when Wilder *et al*. described a surgeon with a malignant pancreatic tumor that had insulin-like biological activity (i.e., an insulinoma). Hyperinsulinism was also described in children of diabetic mothers in the 1920s and was ascribed to 'compensatory hypertrophy of the pancreas.' By 1933, it was recognized that hypoglycemia was caused by several conditions, including adrenal insufficiency, starvation, cirrhosis, and critical illness.

In 1938, Whipple described the triad of criteria necessary for the diagnosis of hypoglycemia, i.e., symptoms consistent with hypoglycemia, blood glucose concentration less than 2.8 mmol l^{-1} and immediate relief by the ingestion of glucose. The ability to measure plasma insulin by radioimmunoassay in 1959 confirmed that not all hypoglycemia was due to hyperinsulinism. Also, the elucidation of the biochemical pathways of glucose, amino acid, and lipid metabolism has led to the recognition of a plethora of congenital causes of hypoglycemia. Consequently, in this chapter, the regulatory mechanisms of glucose homeostasis are first described to provide the basis for understanding the pathophysiology of hypoglycemia. The common causes of hypoglycemia are then outlined.

Definition

The Third International Symposium on Hypoglycemia in 1986 defined hypoglycemia as a blood glucose concentration less than 2.8 mmol l^{-1}, thus endorsing Whipple's criteria. Many clinicians, while suspicious of hypoglycemia at this threshold, do not diagnose hypoglycemia until levels are below 2.2 mmol l^{-1}. However, others argue that hypoglycemia could be defined at higher levels, as physiological responses to lowering of blood glucose begin at 4.0–4.2 mmol l^{-1} with the suppression of endogenous insulin secretion.

Physiological Defenses against Hypoglycemia

Plasma glucose in humans is maintained within the relatively narrow range of 3.6–7.8 mmol l^{-1}, whether in the fasting (postabsorptive) or fed (postprandial) state. This is accomplished by the tight regulation and coupling of glucose utilization and production, each $\sim$10–14 μmol glucose kg^{-1} min^{-1}. The incidence of hyperglycemia is vastly greater than hypoglycemia, as the body has developed several mechanisms to combat life-threatening hypoglycemia. This maintenance of glucose balance is necessary, since the brain has a critical dependence for glucose as an energy source. Other organs such as muscle, liver, heart, kidney, and adipose tissue can utilize other metabolic substrates to provide the energy necessary for maintenance and growth. The brain, however, has a respiratory quotient of 0.97–0.99 under physiological conditions, indicating its almost exclusive use of glucose. Despite this dependency, the brain has no enzymatic machinery for significant glycogen synthesis or gluconeogenesis. Consequently, it depends upon a continuous, high concentration of glucose in the arterial blood. In fasting conditions, the arterial concentration of glucose becomes limiting, the blood-to-brain glucose transport mediated by glucose transporters (GLUT-1 and 3) decreases, and neuronal function is impaired. The brain can subsist on ketone bodies (i.e., acetoacetate, β-hydroxybutyrate), if their plasma concentration rises to the threshold necessary to pass the blood–brain barrier, but this only occurs during prolonged fasting.

During hypoglycemia, counter-regulatory mechanisms convert the liver into an organ of net glucose export, provide alternative metabolic fuels, and decrease glucose utilization. During fasting (i.e., > 6 h after a meal), the liver maintains physiological concentrations of plasma glucose by glycogenolysis and gluconeogenesis. Initially, glycogenolysis takes precedence, accounting for 70–90% of the total glucose production. However, in the absence of gluconeogenesis, the liver's glycogen stores of 40–90 g are depleted in 8–10 h. Subsequently, hepatic gluconeogenesis becomes the prime source of glucose production. Renal gluconeogenesis only becomes important

during prolonged fasting when it may contribute up to half of the total glucose production. Proteolysis provides the main gluconeogenic precursors (alanine, lactate) catabolized from the carbon skeletons of amino acids via the Cori and lactate cycles.

The rate of lipolysis increases, providing nonesterified free fatty acids for conversion to ketone bodies by the liver. Muscle and fat can efficiently utilize fatty acids and ketones for fuel, so glucose utilization by these tissues almost ceases. After several days of fasting, the plasma concentration of ketone bodies rises to the level that supports transfer across the blood–brain barrier. Ketone bodies then become a source of energy for the brain, decreasing its need for glucose by approximately 50%. Thereafter, gluconeogenesis decreases, resulting in a protein-sparing effect. The rate of total glucose utilization eventually nadirs at $\sim 5\,\mu mol\,kg^{-1}\,min^{-1}$, and correspondingly, plasma glucose levels fall to a plateau of 2.5–3.3 $mmol\,l^{-1}$. These levels can be maintained for several weeks.

Hormonal Mechanisms

The above catabolic pathways are under hormonal regulation. When plasma glucose falls, intracellular signals in β-cells, possibly including glucokinase as a glucose sensor, lead to inhibition of insulin secretion (Figure 1). Consequently, glucose uptake by peripheral tissues, especially muscle, and glycogen synthesis in muscle and liver (via decreased glycogen synthetase activity and increased phosphorylase activity) is halted. Gluconeogenesis is stimulated as phosphoenolpyruvate carboxykinase and fructose-1,6-bisphosphatase activity is increased. Also, insulin's inhibition of hormone-sensitive lipase is released, leading to lipolysis (Figure 2). However, the suppression of insulin secretion is not the predominant factor in the correction of hypoglycemia. This is seen in the spontaneous recovery from acute insulin-induced hypoglycemia prior to the dissipation of insulin. Hence, other counter-regulatory factors must be involved to overcome the effects of the relative hyperinsulinemia.

Using the hyperinsulinemic stepped hypoglycemic clamp technique, it was shown that glucose levels $\leq 3.7\,mmol\,l^{-1}$ stimulate the secretion of the counter-regulatory hormones glucagon, epinephrine, growth hormone, and cortisol. Glucagon and epinephrine serve as the major glucose-raising hormones. Glucagon acts directly and rapidly on the liver, augmenting glycogenolysis (by increasing glycogen phosphorylase activity) and gluconeogenesis (via fructose-1,6-bisphosphatase activity). It has no extrahepatic effect on glucose uptake, and its effects are potent, accounting for approximately 40% of glucose recovery. However, they are transient, as its plasma half-life is only 5–10 min before hepatic degradation occurs.

Epinephrine, acting through plasma membrane adrenergic receptors, stimulates glucagon release, inhibits insulin secretion, promotes lipolysis and proteolysis thereby mobilizing gluconeogenic precursors, and decreases peripheral glucose utilization. Epinephrine's effects become extremely critical if glucagon is deficient. Growth hormone and cortisol contribute

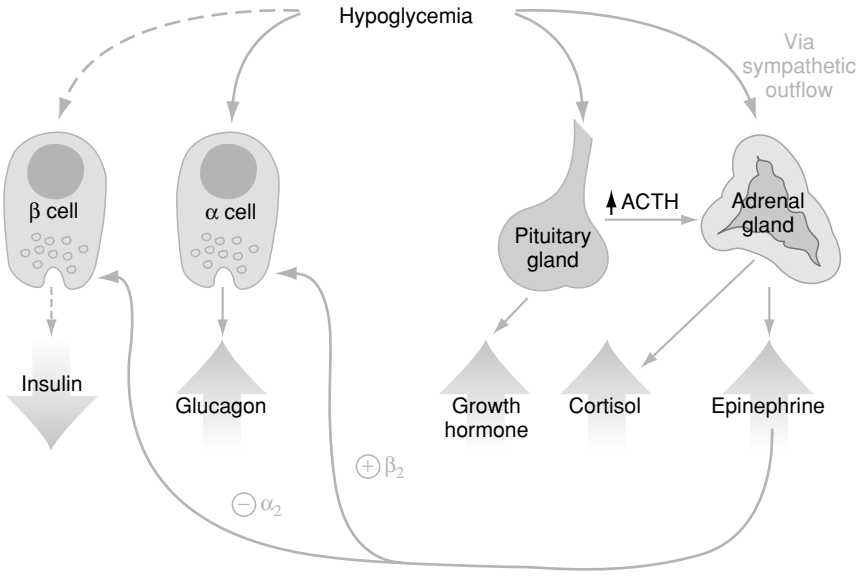

Figure 1 Effect of hypoglycemia on hormonal secretion. α_2 and β_2, receptor-mediated effects of epinephrine.

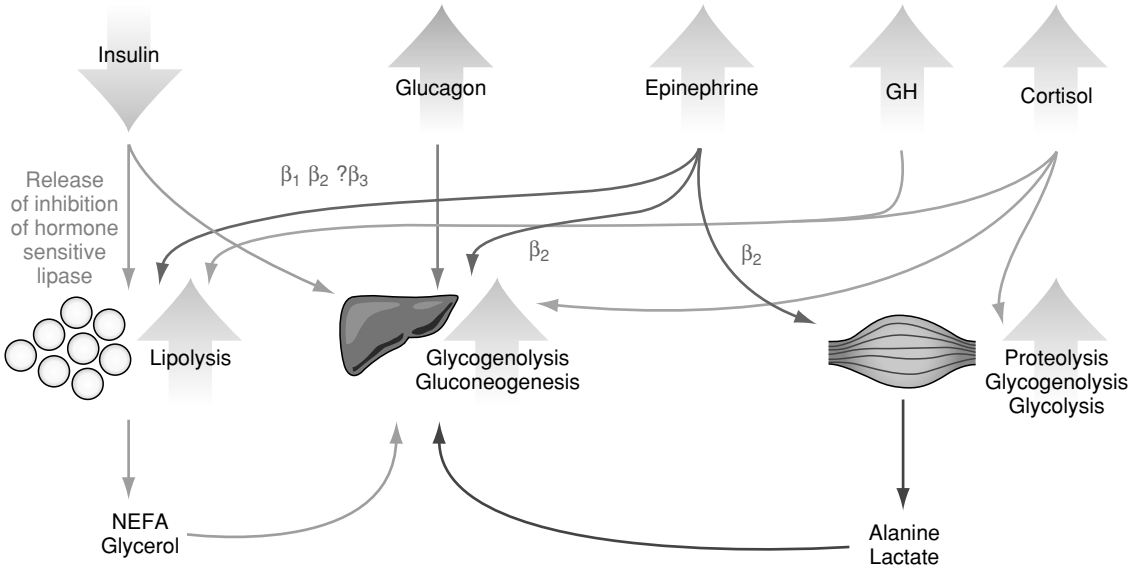

Figure 2 (see color plate 96) Hormonal effects on glucose metabolism during fasting. β_1, β_2 and β_3, receptor-mediated effects of epinephrine; NEFA, nonesterified fatty acids.

if hypoglycemia persists for several hours and is severe. They act by antagonizing insulin effects, inducing a slower recovery from hypoglycemia.

Thresholds for Symptoms of Hypoglycemia

The fall of blood glucose initiates physiological responses prior to the development of any clinical symptoms (Figure 3). A plasma glucose level $< 4.2\,\text{mmol}\,l^{-1}$ suppresses insulin secretion. Counterregulatory hormone secretion begins at $\sim 3.7\,\text{mmol}\,l^{-1}$ when sequential signals within the α-cells promote glucagon secretion. Autonomic nervous system (ANS) discharge is also increased, which augments the glucagon response and initiates the epinephrine response. Infusion of glucose into carotid and vertebral arteries in animal models reduces the epinephrine response by as much as 75%, and the ventromedial hypothalamus is one possible anatomic locus of the ANS signal.

Symptoms are not experienced until the blood glucose levels fall below $3.1\,\text{mmol}\,l^{-1}$. These symptoms are classified as neurogenic or neuroglycopenic. The former includes symptoms induced by the activation of the sympathetic (adrenergic) system and the vagal (cholinergic) system. They include sweating, tremor, anxiety, palpitations, pallor, tingling, and hunger. The neuroglycopenic symptoms occur when the blood glucose levels fall to $\sim 2.6\,\text{mmol}\,l^{-1}$, although there is individual variation. They include poor concentration, inattention, confusion, blurred vision, dizziness, weakness, and fatigue. If the blood glucose falls further, more severe cognitive disturbances develop such as behavioral changes, stupor, seizure, coma, and death.

Classification of Hypoglycemic Disorders

Traditionally, hypoglycemia is classified as either fasting or postprandial (Tables 1 and 2), but it can also be classified on the basis of insulin mediation, state of heath, or age of the patient. Irrespective of the method of classification, most cases of hypoglycemia are attributable to hyperinsulinemic hypoglycemia in individuals with diabetes mellitus from subcutaneous insulin injections, or oral sulfonylureas. (*See* **Diabetes Mellitus**: Problems in Treatment.) It should be noted that in the evaluation of hypoglycemia, artifactual causes i.e., improper sample collection or storage resulting in hemolysis, increased utilization of glucose after phlebotomy by the large number of leucocytes in certain leukemias, and errors in laboratory analysis, need to be excluded.

Postprandial or Food-stimulated Hypoglycemia

Postprandial hypoglycemia is hypoglycemia occurring 2–5 h after a meal. It is a self-limiting condition with no associated mortality, although its morbidity may have been underestimated. It has been described as an epidemic affecting the USA and Europe as more people are presenting self-diagnosed. The majority of

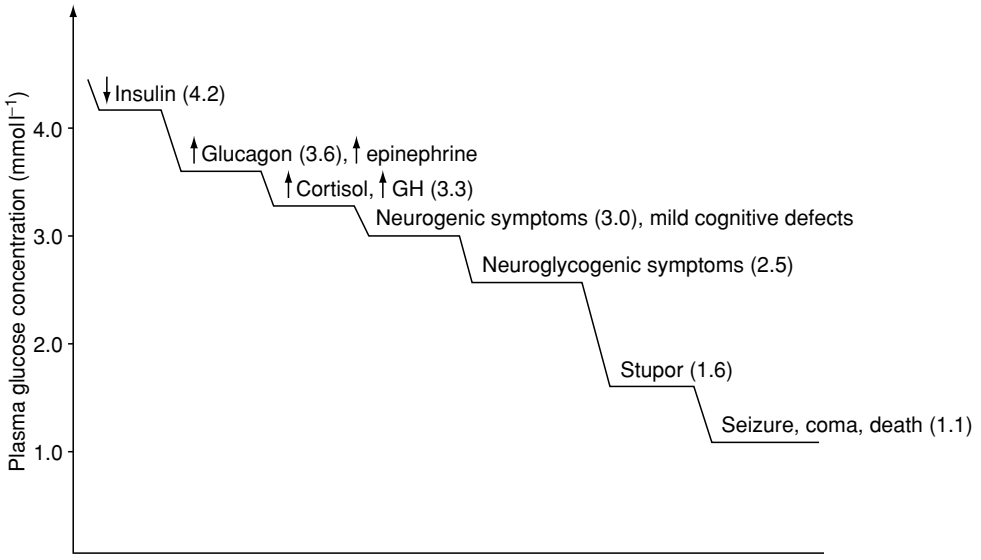

Figure 3 Glycemic thresholds during progressive hypoglycemia.

Table 1 Causes of postprandial hypoglycemia

Alimentary hypoglycemia
Secondary to previous gastrointestinal surgery, especially for peptic ulcer disease and gastrointestinal motility syndromes
Idiopathic postprandial hypoglycemia
Congenital deficiencies of enzymes of carbohydrate metabolism
Hereditary fructose intolerance
Galactosemia
Defective hepatic gluconeogenesis, e.g., fructose-1,6-bisphosphatase deficiency
Gin-and-tonic hypoglycemia
Leucine insensitivity
Causes of fasting hypoglycemia, especially insulinoma

Table 2 Causes of fasting hypoglycemia

Medications
Alcohol
Protein–energy malnutrition
Severe medical illness
Liver failure
Renal failure
Cardiac failure
Sepsis
Endocrinopathies
Adrenal insufficiency
Growth hormone deficiency
Isolated glucagon or epinephrine deficiency
Insulinoma (endogenous hyperinsulinemia)
Isolated
Part of multiple endocrine neoplasia-1 syndrome
Noninsulinoma pancreatogenous hypoglycemia
Noninsulinoma tumor-associated hypoglycemia
Factitious hypoglycemia
Malingering
Psychiatric disorders including Munchausen's syndrome
Pharmacy error
Autoimmune antiinsulin antibody syndrome
Ackee poisoning (Jamaican vomiting sickness)

patients, however, do not fulfill the requirements of Whipple's triad. Misdiagnosis also occurs due to the over-reliance on oral glucose tolerance testing (OGTT) and/or capillary ('finger-stick') glucose for diagnosis. Home reflectance meters for measuring capillary glucose are notoriously unreliable in the hypoglycemic range.

A common cause of postprandial hypoglycemia is *alimentary hypoglycemia*. This generally occurs in individuals who have had gastrointestinal surgery, e.g., gastric bypass, pyloroplasty, and gastrectomy, although a few have normal gastrointestinal tracts. The reported prevalence in postsurgical patients varies between 5 and 37%. The postulated mechanism is rapid transition of ingested food into the upper small intestine, leading to early, severe hyperglycemia (frequently $\geq 14\,\text{mmol}\,\text{l}^{-1}$). This induces the release of local gut factors (incretins), such as glucagon-like peptide-1 and glucose-dependent insulin-releasing peptide, resulting in insulin hypersecretion and glucose levels falling to $< 2.8\,\text{mmol}\,\text{l}^{-1}$ within 2–3 h. Dumping syndrome has to be considered in this setting as it has similar symptoms, but these occur within an hour after meals due to transluminal osmotic/fluid shifts. Treatment of alimentary hypoglycemia involves frequent feedings of high-fiber meals or low-carbohydrate/high-protein meals. Pectin, fructose, uncooked cornstarch, chromium, medium-chain triacylglycerol, octreotide, propanolol, sulfonylureas,

α-glucosidase inhibitors, metformin, and doxepin have also been proposed as alternatives, but none of these therapies have been properly evaluated in controlled trials.

It is controversial whether there really is an entity of *idiopathic postprandial hypoglycemia*. Most have been diagnosed on the grounds of diffuse symptoms and biochemical hypoglycemia by OGTT. In a study of 650 normal subjects, 10% had nadir values of $<2.6\,\text{mmol}\,\text{l}^{-1}$, and 2.5% were $<2.16\,\text{mmol}\,\text{l}^{-1}$ during a 100-g OGTT. Consequently, it would seem unwise to use an arbitrary diagnostic cut-off, since none of these patients had any significant symptoms or long-term adverse effects. In fact, 4.8% of normal individuals have blood glucose values $\leq 3.0\,\text{mmol}\,\text{l}^{-1}$ in free-living situations. When a group of 118 patients with suspected hypoglycemia were studied, only 16 had confirmed hypoglycemia (the mean was $2.2\,\text{mmol}\,\text{l}^{-1}$) during symptoms. The other 102 patients were symptomatic, but there was no relationship to blood glucose levels or meals, and their nadirs were similar to those of the control group. Fourteen of these patients also had placebo-OGTT tests, which provoked characteristic symptoms. Another group of investigators confirmed the invalidity of OGTT and also hypothesized that mixed meal tolerance tests of 550 and 825 kcal should provide a more physiological stimulus to provoke symptoms. The meals did not induce biochemical hypoglycemia, but the patients had characteristic symptoms. In another study, there was a poor correlation between the occurrence of symptoms and plasma glucose concentrations in ambulatory patients. Any abnormalities were not reproducible.

Based on these studies and others, it is generally agreed that an OGTT should not be used in the evaluation of idiopathic postprandial hypoglycemia, as its predictive power is poor (positive predictive value 21–31%, negative predictive value 68–100%). Instead, documenting hypoglycemia after meals that historically provoke hypoglycemia is more useful. It has been suggested that symptomatic individuals without biochemical hypoglycemia should be given the diagnosis of *idiopathic postprandial syndrome* or *pseudohypoglycemia*, since factors other than hypoglycemia are operative, while giving validity to their symptomatology. Other potential mechanisms, such as the rate of glucose decline, excessive insulin secretion, and plasma cortisol responses have not been proven to be significant. Increased tissue sensitivity to insulin may play a role in some individuals. Glucagon responses have not been consistent between studies, but may be subnormal. Exaggerated plasma epinephrine responses have been reported in uncontrolled studies. In any event, other investigators have shown that epinephrine and cortisol concentrations do not increase during the symptoms, indicating that no significant autonomic activation is present. It should be noted that over 50% of patients with idiopathic postprandial syndrome also have evidence of conversion reactions with high scores on the depression, hysteria, and hypochondriasis scales on psychological testing. In one study, 45% had somatoform disorders, 35% had depressive affective disorders, 20% had no evidence of a psychiatric disorder, and only 3% had true idiopathic postprandial hypoglycemia. Hence, idiopathic postprandial hypoglycemia is a rare entity, unlike the more common idiopathic postprandial syndrome. Treatment involves dietary restriction of refined carbohydrates and avoiding caffeine and alcohol.

Other conditions causing postprandial hypoglycemia include *galactosemia* and *hereditary fructose intolerance* that usually present during childhood. Rarely, *leucine* intake can cause hypoglycemia in susceptible individuals. As early as 1924, it was suggested that postprandial hypoglycemia occurs in the early phase of *type 2 diabetes mellitus*. However, this work was based on OGTTs, and only ~5% of patients were symptomatic. It was postulated that abnormal β-cell responses to glucose loads lead to hyperinsulinemia. Individuals with *fasting hypoglycemia*, especially insulinomas, may also have postprandial hypoglycemia. A high intake of alcohol with simple carbohydrates and no complex carbohydrates (such as *gin and tonic*) can lead to symptomatic hypoglycemia within 3–4 h in 10–20% of normal individuals. It is thought that the hypoglycemia is due to an insulinotrophic effect of ethanol in response to sucrose.

Fasting Hypoglycemia

The causes of fasting hypoglycemia are numerous and are outlined in Table 2. In many cases, the cause is evident, such as in an ill patient with multiorgan failure, but in others, especially ambulatory patients, it may not be obvious. In these individuals, it is necessary to measure *simultaneously* plasma glucose, insulin, C-peptide, and sulfonylurea levels during the episode of hypoglycemia. If insulin levels are not available, serum or urine ketone levels can give an indication of insulin mediation, as ketone levels are raised in hypoinsulinemic states.

Drugs

Drugs are a frequent cause of hypoglycemia, and some of the more common drugs are listed in Table 3 along with their putative mechanism. Other drugs have been alleged to cause hypoglycemia, but a causal

Table 3 Drugs associated with hypoglycemia

Drug	*Mechanism of inducing hypoglycemia*
Insulin	Increase glucose uptake by peripheral tissues
Sulfonylureas	Increase insulin secretion via sulfonylurea receptors on β-cells
Pentamidine	Increases insulin secretion
Quinine	Increases insulin secretion
Salicylates	Unknown mechanism – ? increase peripheral glucose uptake; need large doses (4–6 g)
Sulfonamides	Unknown mechanism
Angiotensin-converting enzyme inhibitors	? Increases insulin sensitivity
Acetaminophen	Hepatic necrosis in acute toxic overdoses
Disopyramide	Increases insulin secretion
Monoamine oxidase inhibitors	Unknown mechanism, rare
β-adrenergic antagonists	? Increase peripheral glucose uptake; ? inhibit gluconeogenesis; more common in patients with diabetes

relationship has not been established for many of them. These include benzodiazepines, ciprofloxacin, etomidate, enflurane, fluoxetine, furosemide, halothane, interferon-α, isoniazid, maprotiline, mefloquine, nefazodone, octreotide, phenytoin, and selegiline.

Alcohol

Alcohol inhibits hepatic gluconeogenesis by depleting nicotinamide adenine dinucleotide levels and by decreasing the flux of gluconeogenic substrates (i.e., alanine, glycerol, lactate). However, it does not affect glycogenolysis. Therefore, in the well-fed person, hypoglycemia does not occur until after depletion of glycogen stores by fasting > 12 h, along with consumption of 50–300 g of alcohol during the preceding 6–36 h. In the poorly nourished individual, less alcohol may induce hypoglycemia more quickly. Alcohol may also potentiate insulin and sulfonylurea-induced hypoglycemia. Intravenous glucose causes prompt recovery, unlike glucagon, which is ineffective. Without treatment, the mortality rate is 10% in adults and 25% in children.

Protein–Energy Malnutrition

There are many descriptions of hypoglycemia in severe protein–energy malnutrition (PEM), although there seems to be no consistent difference in the severity of hypoglycemia between marasmus and kwashiorkor. The frequency of hypoglycemia varies geographically, ranging from 10% in Jamaica to 24% in Uganda. Hypoglycemia is a poor prognostic factor, as its presence increases the mortality in marasmic infants by 1.6–3.1-fold, unless intravenous glucose is given immediately. The pathogenesis is not clear, but high rates of glucose utilization in the presence of total body fat depletion, limitation of substrate availability for gluconeogenesis, or diminished activity of gluconeogenic enzymes are possibilities. Hypoglycemia is also seen in anorexia nervosa.

Endocrine Deficiencies

Endocrine deficiencies are rare causes of hypoglycemia. Deficiencies of cortisol and/or growth hormone impair gluconeogenesis, especially in the very young, or if other hypoglycemic stressors are present. These endocrinopathies limit the availability of gluconeogenic substrate availability and decrease the activity of hepatic gluconeogenic enzymes. Combined deficiencies, e.g., in hypopituitarism, cause more severe hypoglycemia. Isolated glucagon or epinephrine deficiencies in infants can also cause hypoglycemia but are extraordinarily rare. They can occur as an acquired condition in individuals with long-standing type 1 diabetes. Hypothyroidism can reduce the rate of glucose absorption from the gut and decrease peripheral glucose utilization. However, although hypothyroidism has been said to be a cause of hypoglycemia, it rarely occurs in isolated thyroxine deficiency. The concurrence of hypoglycemia and hypothyroidism should suggest the presence of adrenal insufficiency.

Severe Medical Illness

Severe liver disease, e.g., hepatectomy or fulminant hepatitis can cause profound hypoglycemia, as the liver is the major source of endogenous glucose production. *Renal failure* is also associated with hypoglycemia. Multiple mechanisms are involved, including cachexia, uremic inhibition of hepatic gluconeogenesis, failure of renal gluconeogenesis, decreased renal clearance of insulin (in patients with concurrent diabetes), and dialysis. Hypoglycemia occasionally occurs in *congestive heart failure*, and it is suggested that it is due to hepatic congestion, cachexia, and inhibition of gluconeogenesis. *Sepsis* increases peripheral glucose utilization and inhibits hepatic gluconeogenesis, mediated by cytokines such as tumor necrosis factor-α, interleukin-1, and interleukin-6.

Ackee Poisoning (Jamaican Vomiting Sickness)

A toxic hypoglycemia syndrome occurs whenever the unripe fruit of ackee (*Blighida sapida*) is consumed, as first reported in 1875. Ackee was imported to the New World from West Africa in 1778 and is found in the Caribbean, Central America, and Florida. Poisoning is more common in children, especially if undernutrition is present. Adults generally have a chronic course with a self-limited cholestatic jaundice. The unripe arillus of the fruit contains hypoglycin, which is sequentially transaminated and decarboxylated to methylenecyclopropylacetyl-CoA. This metabolite irreversibly inhibits acyl-CoA dehydrogenases (especially butyryl-CoA dehydrogenase), leading to inhibition of fatty acid β-oxidation. The metabolic consequences are an acquired organic acidemia (mostly glutaric and ethylmalonic acid), the absence of ketosis or hyperinsulinemia, and impairment of hepatic gluconeogenesis. Patients have severe hypoglycemia, an acute encephalopathy similar to Reye's syndrome, and severe vomiting. Convulsion, coma and death can follow. Treatment involves intensive glucose infusions to maintain normoglycemia. Cooking renders the fruit nontoxic, reducing hypoglycin levels 100-fold.

Insulinomas

Insulinomas are rare tumors with an incidence of four cases per million person-years. Approximately 10% are malignant, 9% are multifocal, 8% are recurrent, and 8% occur as part of multiple endocrine neoplasia-1 syndrome. They are evently distributed throughout the pancreas, and ectopic locations are unusual. Some tumors cosecrete gastrin, adrenocorticotropic hormone, glucagon, and 5-hydroxyindole. The diagnosis relies on documentation of Whipple's triad with inappropriately high insulin ($\geq 36\,\text{pmol}\,\text{l}^{-1}$), C-peptide ($\geq 0.2\,\text{nmol}\,\text{l}^{-1}$), and proinsulin ($\geq 5\ \text{pmol}\,\text{l}^{-1}$) levels during hypoglycemia. Surreptitious use of exogenous insulin and sulfonylureas has to be ruled out, and a 72-h supervised inpatient fast may be required (Table 4). After biochemical confirmation, the tumor may be localized by conventional, intraoperative, or endoscopic ultrasonography. Magnetic resonance imaging or selective intraarterial injection of calcium with venous sampling for insulin may be useful. Some are localized by surgical exploration. Treatment is by surgical enucleation of the tumor, and chemotherapy is necessary for palliation in individuals with metastases.

Noninsulinoma Tumor-associated Hypoglycemia

There are various noninsulinoma tumors that induce hypoglycemia, most commonly mesenchymal cancers, hepatoma, adrenocortical carcinoma, lymphoma, and gastrointestinal cancers. Generally, hypoglycemia is a late presentation, the tumor burden is large (> 500 g), and metastases are commonly present. Many of these tumors produce insulin-like growth factor-II as the putative hypoglycemic agent. They rarely produce ectopic insulin. Occasionally, some induce hypoglycemia by increasing glucose consumption in excess of the liver's gluconeogenetic ability.

Miscellaneous Causes

Factitious hypoglycemia is due to the surreptitious use of insulin or sulfonylureas. This may be due to malingering, Munchausen syndrome, attempted suicide/homicide, or pharmacy error. A proper biochemical evaluation during a 72-h fast (Table 4) should reveal the diagnosis. Insulin levels $> 1200\,\text{pmol}\,\text{l}^{-1}$ should generate suspicion of factitious causes. Very rarely, in patients naïve to insulin, there may be a spontaneous autoimmune *antiinsulin antibody syndrome*. This has been reported in Japanese patients treated with thionamides or with other autoimmune disorders. The mechanism involves an insulinomimetic action when the antiidiotypic antibodies to antiinsulin antibodies interact with the insulin receptor. An insulinomimetic action of antibodies against the insulin receptor can cause hypoglycemia in *type B insulin resistance*. *Severe exercise*, as seen in marathon runners, has occasionally been associated with hypoglycemia.

Childhood Hypoglycemia

As children have an increased ratio of brain to body mass, they have higher rates of glucose flux. There is also limited substrate availability for gluconeogenesis due to their smaller muscle mass. Hence, they are not

Table 4 Interpretation of the results of a 72-h fast

Diagnosis	*Insulin*	*C-peptide*	*Proinsulin*	*Plasma or urine sulfonylurea*	*β-Hydroxybutyrate*
Insulinoma	Increased	Increased	Increased	Negative	None
Exogenous insulin	Increased	Decreased	Decreased	Negative	None
Sulfonylurea ingestion	Increased	Increased	Increased	Present	None
Noninsulin-mediated hypoglycemia	Decreased	Decreased	Decreased	Negative	Increased

able to tolerate hypoglycemia as easily as adults. Their clinical presentation is more variable with hypotonia, cyanosis, hypothermia, tachypnea, feeding difficulties, and poor neuropsychiatric development being prominent. Older children have more adrenergic symptoms and may have focal neurological deficits and movement disorders.

Most of the causes of hypoglycemia in adults are also seen in children. *Transient hyperinsulinism* is the most frequent cause of hypoglycemia in the infant, as seen in the children of diabetic mothers, premature infants, and small-for-gestational-age infants. *Congenital hyperinsulinism*, otherwise called *nesidioblastosis* or *persistent hyperinsulinemic hypoglycemia of infancy*, describes diffuse β-cell hyperplasia with hyperinsulinemic hypoglycemia. Subtotal pancreatectomy or medical treatment with diazoxide or octreotide is advocated for symptomatic infants. This is due to genetic mutations that affect the normal intracellular events controlling insulin secretion. The mutations include:

- mutations inactivating the ATP-sensitive potassium channel in β-cells leading to membrane depolarization (*SUR1* and *Kir6.2 subunit mutations*);
- activating *mutations of the glucokinase gene* increasing glycolysis; and
- activating mutations of the glutamate dehydrogenase gene causing an increased rate of glutamate oxidation in β-cells (*hyperinsulinism/hyperammonemia syndrome*, which includes leucine-sensitive hypoglycemia, as leucine is an allosteric activator of GDH and thus functions as an insulin secretagogue).

Hypoglycemia is also a consequence of inborn errors of metabolism involving gluconeogenesis, glycogen storage, fatty acid metabolism, and amino acid metabolism.

Ketotic hypoglycemia of childhood generally begins between the ages of 1.5 and 5 years, and resolves by age 10. It probably accounts for 90% of children with hypoglycemia beyond infancy, and it may represent a subset of children who are least tolerant of fasting, especially during an intercurrent illness. These children characteristically have relatively decreased muscle mass, and there may be a defect in mobilizing alanine for gluconeogenesis. Although alanine administration increases blood glucose levels, treatment generally involves avoiding any prolonged fasting.

See also: **Alcohol**: Properties and Determination; **Diabetes Mellitus**: Etiology; Problems in Treatment; **Glucose**: Properties and Analysis; Function and Metabolism; Maintenance of Blood Glucose Level; Glucose Tolerance and the Glycemic (Glycaemic) Index; **Hormones**: Pituitary Hormones; **Malnutrition**: The Problem of Malnutrition

Further Reading

Bolli GB (1999) How to ameliorate the problem of hypoglycemia in intensive as well as nonintensive treatment of type 1 diabetes. *Diabetes Care* 22 (supplement 2): B43–B52.

Cryer PE (1997) *Hypoglycemia: Pathophysiology, Diagnosis and Treatment*. New York: Oxford University Press.

Haymond MW and Pagliara AS (1983) Ketotic hypoglycemia. *Clinical Endocrinology and Metabolism* 12: 447–462.

Hofeldt FD (1989) Reactive hypoglycemia. *Endocrinology and Metabolism Clinics of North America* 18: 185–201.

Malouf R and Brust JCM (1985) Hypoglycemia: causes, neurological manifestations and outcome. *Annals of Neurology* 17: 421–430.

Marks V (1987) Functional hypoglycemia: fact or fancy. In: Andreani D, Marks V and Lefebvre PJ (eds) *Hypoglycemia: Serono Symposium*, vol. 38, pp. 1–17. New York: Raven Press.

Seltzer HS (1989) Drug-induced hypoglycemia. *Endocrinology and Metabolism Clinics of North America* 18: 163–183.

Service FJ (1995) Hypoglycemic disorders. *New England Journal of Medicine* 332: 1144–1152.

Service FJ (ed.) (1999) Hypoglycemic disorders. *Endocrinology Clinics of North America* vol. 28(3). Philadelphia, PA: W.B. Saunders.

Sovik O (1989) Inborn errors of amino acid and fatty acid metabolism with hypoglycemia as a major clinical manifestation. *Acta Paediatrica Scandinavica* 78: 161–170.

Sperling MA and Menon RK (1989) Hyperinsulinemic hypoglycemia of infancy. Recent insights into ATP-sensitive potassium channels, sulfonylurea receptors, molecular mechanisms, and treatment. *Endocrinology and Metabolism Clinics of North America* 28: 695–708.

Tanaka K and Ikeda Y (1990) Hypoglycin and Jamaican vomiting sickness. *Progress in Clinical and Biological Research* 321: 167–184.

HYPOVITAMINOSIS A

V Reddy, Panjagutta, Hyderabad, India

Introduction

Vitamin A is an essential nutrient required for normal vision, growth, and development, maintenance of epithelial cellular integrity, immune response, and a variety of metabolic functions. It is a fat-soluble vitamin and is stored in the body when intake exceeds need. Depletion of stores occurs over time when the diet does not provide enough vitamin A to replace the amount used by the tissues. Signs of hypovitaminosis A appear when the stores are depleted to the extent that physiological functions are impaired. A series of problems ranging from ocular changes to growth retardation and increased mortality have been documented in vitamin A-deficient animals, while in human vitamin A deficiency (VAD), attention is focused on xerophthalmia because of the striking changes in the eye. VAD is recognized as the most important cause of childhood blindness in developing countries. Apart from causing eye damage, VAD can also affect host defense mechanisms, increasing susceptibility to infection. There is evidence that VAD, even at subclinical level, increases the risk of morbidity and mortality in vulnerable groups. Recognizing the importance of the problem, heads of state and other high-level policy makers gathered at the World Summit for Children (1990) and the International Conference on Nutrition (1992) adopted the goal of elimination of VAD by the year 2000. Although considerable progress has been made in the past decade, VAD remains a public health problem in many developing countries. There is a need to accelerate the intervention efforts to eliminate VAD and its consequences.

Dietary Sources of Vitamin A

Vitamin A is obtained through the diet as preformed vitamin A or provitamin A carotenoids. Preformed vitamin A is present in animal foods such as milk, butter, eggs, liver, and fish. In most tropical countries, where consumption of these foods is low, the main sources of vitamin A in the diet are the provitamin A carotenoids. These are present in dark-green leafy vegetables, yellow/orange vegetables like carrots and pumpkin, and fruits like mango and papaya. Red palm oil, used mostly in African countries, has a very high concentration of β-carotene. Of the various carotenoids in foods, β-carotene has the maximum vitamin A activity. Other provitamin A carotenoids include α-carotene and β-cryptoxanthine, which have approximately half the vitamin A activity of β-carotene. Values of carotene/vitamin A given in most food composition tables are overestimates, as they are based on old analytical methods which do not distinguish between provitamin A and other carotenoids. Only recently have accurate values for provitamin A carotenoids been obtained, using high-performance liquid chromatography (HPLC) technique.

Recommended Dietary Intakes

The recommended intakes of vitamin A for different age groups are given in Table 1. These intakes are designed to prevent deficiency and provide a safe intake for a majority of the population. The values are expressed in terms of retinol equivalents (RE) per day, taking into account vitamin A activity of provitamin A carotenoids. The Food and Agriculture Organization/World Health Organization (FAO/WHO) expert committee has suggested a ratio of 6:1 for converting β-carotene to vitamin A, assuming that one-third of the carotene in the diet is absorbed and half of it is converted to vitamin A. However, recent studies showed lower bioavailability of carotene, with a ratio of 26:1 for vegetables and 12:1 for fruits. Bioavailability of dietary carotenoids is influenced by a number of factors, including the food source, method of preparation, amount ingested, and dietary fat. Any single/average ratio for the conversion of carotenoids to vitamin A is at best a rough estimate that is not valid for all diets.

Table 1 Recommended dietary intake of vitamin A

Group	*Age (years)*	*Recommended intake (μg RE day^{-1})*
Both sexes	0–1	350
	1–10	400
	10–12	500
	12–15	600
Boys	15–18	600
Men	18+	600
Girls	15–18	600
Women	18+	500
Pregnant women		600
Lactating women		850

RE, retinol equivalent.
Reproduced from FAO/WHO Consultation (1998) *Requirements of Vitamin A, Iron, Folate and Vitamin B_{12}*, pp. 16–31. Rome: FAO, with permission.

Assessment of VAD

Several methods are available for assessing the magnitude and severity of VAD in a population. These include clinical, functional, and biochemical indicators. The WHO monograph on indicators of VAD provides guidelines on the selection and interpretation of practical indicators for population assessment.

Clinical Indicators

Clinical indicators are useful for assessing the extent of VAD in communities where clinical problem or xerophthalmia exists. The core indicators and the minimum prevalence criteria to define a public health problem are given in Table 2. Corneal xerophthalmia indicates a serious problem but the condition is relatively rare and requires a large sample size to estimate the prevalence. Night blindness (XN) and Bitot's spots (X1B) are the two indicators commonly used in community surveys. Conjunctival xerosis is not a reliable sign, while xerosis with Bitot's spot is more specific and a prevalence above 5% in preschool children indicates a public health problem.

Indicators of Subclinical Deficiency

Serum retinol is the most commonly used biochemical indicator. Values $< 0.70\,\text{mmol}\,l^{-1}$ are considered as indicative of VAD in children. The cut-off levels of prevalence used for defining a public health problem of moderate to severe degree are shown in Table 3. It is possible to estimate retinol concentration in finger-prick blood samples using a filter paper technique, which offers logistic advantages. Other indicators include relative dose response (RDR) test for assessing vitamin A stores and conjunctival impression cytology (CIC) for detection of conjunctival changes that precede clinical signs. However, these tests have limited application in field surveys.

Table 2 Classification of xerophthalmia and prevalence criteria constituting a public health problem

Criteria	*Minimum prevalence*
Night blindness (XN)	> 1.0%
Bitot's spots (X1B)	> 0.5%
Corneal xerosis/ulceration (X2, X3A, X3B)	> 0.01%
Corneal scar (XS)	> 0.05%

Data from: WHO (1995) *Global Prevalence of Vitamin A Deficiency*. Geneva: WHO, with permission.

Table 3 Classification of subclinical deficiency (serum retinol $< 0.70\,\text{mmol}\,l^{-1}$)

Public health problem	*Prevalence in children*
Mild	≥ 2– < 10%
Moderate	≥ 10– < 20%
Severe	≥ 20%

Data from: WHO (1995) *Global Prevalence of Vitamin A Deficiency*. Geneva: WHO, with permission.

Global Prevalence

Earlier estimates of VAD prevalence were based on clinical signs of xerophthalmia, while more recent surveys include data on subclinical deficiency based on blood levels of vitamin A $< 0.70\,\text{mmol}\,l^{-1}$. It is now estimated that VAD, including clinical and subclinical forms of moderate to severe degrees of public health significance, exists in 60 countries and it is likely to be a problem in at least an additional 13 countries. On a regional basis, South-east Asia and Africa account for just under 90% of the global problem (Figure 1). Although the prevalence of clinical deficiency has shown a marked decline from 14 million in 1994 to 3 million in 1996, estimates of subclinical deficiency have gone up to 251 million, with the increasing use of serum retinol concentration for assessment. About 254 million children in the world are thus at risk in terms of their health and survival.

Epidemiology

Socioeconomic Factors

VAD occurs within an ambience of ecological, economical, and social deprivations that characterize the poor in developing countries. At the macro level, adverse environment with infertile land and water shortage limit the availability of foods rich in vitamin A. Since animal foods containing preformed vitamin A are expensive, poor families rely on plant sources of provitamin A carotenoids. Illiterate and impoverished mothers tend to follow traditional feeding practices and do not fully utilize even the available resources. High infection rates associated with poor living conditions also contribute to hypovitaminosis A.

Clustering

The occurrence of VAD tends to cluster rather than to be evenly distributed. Clustering within high-risk households and within communities has been described primarily for clinical deficiency. Studies in Asia show that siblings of xerophthalmic children are 10 times more likely to have xerophthalmia than

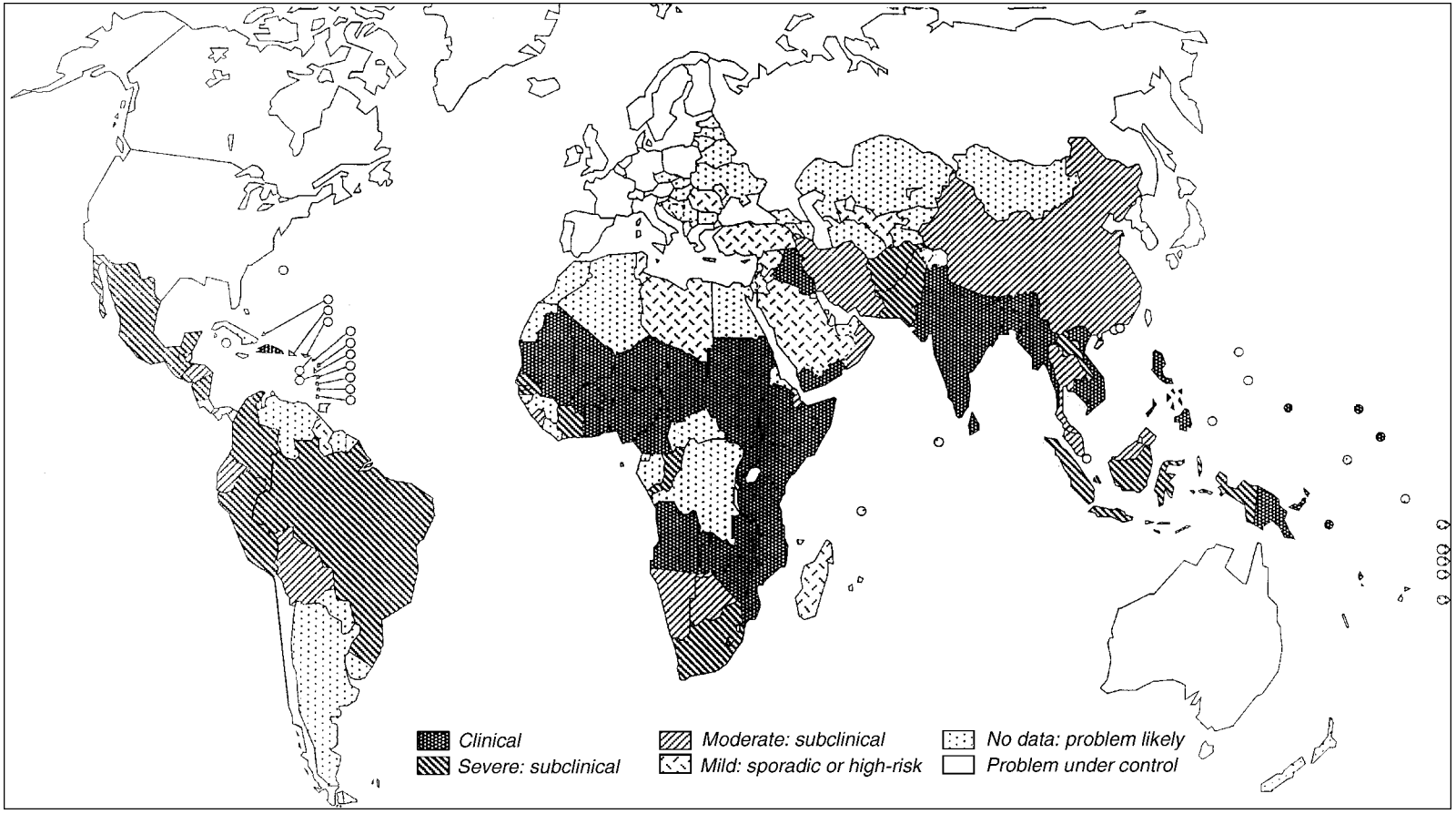

Figure 1 Global prevalence of vitamin A deficiency. Reproduced from WHO (1995) *Global Prevalence of Vitamin A Deficiency. Micronutrient Deficiency Information System*. Working Paper #2. WHO/NUT/95.3. Geneva: WHO, with permission.

those without. Neighboring children of xerophthalmic children are also more likely to develop xerophthalmia. This is due to sharing common sociocultural and economic conditions that result in frequent infections, limited food availability, and poor dietary practices that predispose to VAD.

Seasonal Patterns

The occurrence of xerophthalmia follows a seasonal pattern in different parts of the world. In South Asia, for example, a distinct peak in the incidence of xerophthalmia is observed in the month of May–June. The peak appears to be related to the total food intake and consequent growth spurt of children. It also coincides with a general scarcity of provitamin A-rich vegetables and fruits and a seasonal rise in the incidence of diarrhea and respiratory infection.

Host Factors

Age VAD is more common among children of preschool age. While xerophthalmia rarely occurs during infancy, children between 1 and 4 years are at greater risk. The potentially blinding corneal disease is most prevalent during the weaning period and is associated with severe protein-energy malnutrition (PEM). Symptoms of milder xerophthalmia like Bitot spots and night blindness are more common in older children. Hypovitaminosis A persists throughout adolescence and into adulthood, as judged by low serum retinol levels. Clinical signs of xerophthalmia are rare in adults except for maternal night blindness. Reports from Asia indicate that about 10–20% of women develop night blindness during pregnancy.

Sex No consistent sex difference is demonstrated based on physiological parameters. Differences have been reported in some communities, which are more likely to be related to sex differences in cultural practices of feeding and care rather than to physiological differences.

Dietary Practices

In poor communities, diets are deficient in several nutrients, including vitamin A. Pregnant and lactating women rarely take any extra food to meet their additional needs. As a result, the offspring are born with poor stores of vitamin A. During the first 6 months of life, breast milk is the major source of vitamin A for infants but the vitamin concentration of milk is low in undernourished mothers. Although clinical signs of VAD are rare in infants, their vitamin stores are depleted. Complementary feeding is delayed and foods rich in vitamin A are seldom given, increasing the risk of VAD during the weaning period. Diet surveys in countries of South Asia show that the daily intake of vitamin A among preschool children is less than half the recommended intake and most of it is derived from plant sources of provitamin A carotenoids. Bioavailability of plant carotenoids is low and it is difficult to meet the needs of young children from vegetable foods alone. Dietary fat is important for the absorption of carotenoids but the weaning diets contain little fat. These patterns signal the need to improve child feeding practices in high-risk families to sustain adequate vitamin A status.

Protein-Energy Malnutrition

VAD is often associated with protein-energy malnutrition (PEM) in children. Clinical studies show that the prevalence of xerophthalmia is higher in children with PEM than in those without. Almost all cases of corneal xerophthalmia/keratomalacia are associated with severe PEM. These observations suggest that protein deficiency may aggravate VAD (often associated with protein-energy malnutrition in children). Experimental studies show that there is a close metabolic relationship between protein and vitamin A. While the synthesis of retinol-binding protein (RBP) that carries vitamin A is influenced by protein nutrition, its release from the liver depends on vitamin A status. Studies in malnourished children with corneal xerophthalmia have shown that the concentration of serum retinol was reduced to a greater extent than that of RBP. Administration of vitamin A to such children not only raised serum levels of these components but also showed a rapid clinical improvement. These observations indicate that the role of vitamin A is more critical than that of protein in the development of corneal lesions. Appreciation of this fact has an important bearing on the treatment and prevention of xerophthalmia.

Infections

Keratomalacia is often preceded by an episode of infection: diarrhea and respiratory infection are common precipitating factors. Prospective studies also show that common childhood infections, including diarrhea, respiratory illness, measles, and chickenpox, induce hypovitaminosis A. During acute infection, not only is the food intake reduced but metabolic alterations are also known to occur. These include decreased absorption of vitamin A, impaired retinol transport, and increased renal excretion. Slow normalization of these mechanisms coupled with low dietary intake of vitamin A during the acute phase and recovery lead to VAD. Parasitic infections like ascariasis and giardiasis also aggravate VAD. Intestinal worms may directly compete for uptake of

vitamin A, besides their more general impact on health by suppressing appetite.

Xerophthalmia

Although hypovitaminosis A affects many tissues in the body, the most striking changes are seen in the eye. The conjunctival and corneal epithelium undergo keratinizing metaplasia. Columnar epithelial cells become squamous and mucus-producing goblet cells disappear, resulting in dryness/xerosis of the ocular surfaces. Hypovitaminosis A also affects the retinal function. Vitamin A is required for the synthesis of rhodopsin, a photosensitive pigment in the rod cells of retina that responds to light by releasing vitamin A. If the vitamin is not replaced, rhodopsin production is reduced, resulting in impaired dark adaptation and night blindness.

Night Blindness

Night blindness is one of the early manifestations of VAD and is marked by an inability to move around in the dark. It is common among pregnant women in endemic areas and often disappears after delivery without any treatment. Night blindness is also common among preschool children but its detection is difficult in children below 2 years. In endemic areas, a local term is often used for night blindness that translates into 'twilight blindness' or 'chicken eyes' (chicken have no rods and are therefore night-blind), making the condition readily detectable by history. Impaired dark adaptation is an early functional abnormality which can be detected using a dark adaptometer.

Conjunctival Changes

Conjunctival xerosis appears as a dry, unwettable surface of the bulbar conjunctiva. The conjunctival changes that precede clinical signs can be detected by filter-paper impression cytology. These include keratinization/distortion of epithelial cells and disappearance of goblet cells. In chronic cases, accumulation of keratinized cells appears as white foamy patches of oval or triangular shape called Bitot's spots. They are usually bilateral and seen on the temporal side of the corneal limbus.

Corneal Changes

Severe deficiency of vitamin A involves the cornea, affecting the vision. The earliest corneal epithelial changes appear as punctate defects, which can be detected by a slit lamp. Clinically, the cornea appears dry, hazy, and lusterless (X2). In severe cases, corneal ulcers (X3A) appear as round or oval defects that are usually shallow, but may progress to perforation of cornea. Healed ulcers appear as white scars (XS) and affect vision if they are centrally located. In more advanced cases, there is necrosis and softening of cornea (X3B). This condition, described as keratomalacia, often results in permanent blindness. Corneal xerophthalmia is invariably associated with severe PEM and infection. Hospital studies show high mortality in such cases, ranging from 5 to 25%.

Mortality

Childhood Mortality

Preschool age Vitamin A has been named as 'anti-infective vitamin' as early as 1928, based on the evidence obtained from animal studies. Although there were several reports on the association of xerophthalmia with increased mortality among malnourished children during the subsequent years, the importance of vitamin A in child health and survival was not fully recognized until late 1980s. Community studies in Indonesia showed that preschool children with mild xerophthalmia were two to three times more likely to have developed diarrhea and respiratory illness than those with normal eyes. The risk of mortality was also higher in such cases. The effect of VAD on infection may be mediated through several mechanisms, including impairment of mucosal epithelial barrier and of the immune response. Low T-cell count and impaired blast transformation have been reported in vitamin A-deficient children.

An intervention trial conducted in Indonesia showed that administration of 6-monthly doses of 200 000 IU vitamin A to preschool children resulted in 34% reduction in mortality. This stimulated a lot of interest worldwide and led to a series of field trials in Asia and Africa. Six out of eight controlled trials in children showed mortality reductions ranging from 19% to 54%. However, two studies reported from India and Sudan showed no significant impact. Metaanalysis of data from these studies has estimated a reduction of total mortality by 30%.

Thus, vitamin A supplementation has resulted in significant reductions in mortality in several, though not all, community studies. Vitamin A had a protective effect against death from measles and diarrhea but not from respiratory infection. The impact of vitamin A may vary in different situations depending upon the extent and severity of VAD, existence of concomitant nutritional deficiencies, particularly PEM, and the prevalence of infections in the community. The supplements are expected to have a greater impact in areas where VAD is highly prevalent.

Early infancy A few studies have examined the efficacy of vitamin A supplementation among apparently normal infants under 6 months of age. Administration of a single dose of 50 000 IU vitamin A to neonates in Indonesia resulted in a significant reduction in the risk of death while no effect was observed in a similar study in Nepal. Similarly, in a multicentric WHO study conducted in India, Peru and Ghana, administration of three doses of vitamin A along with diphtheria, pertussis, typhoid (DPT) vaccine in the first 4 months of age had no effect on infant mortality or morbidity.

Measles-related mortality Beneficial effect of vitamin A on measles-related morbidity and mortality in children was reported as early as 1932. In recent placebo-controlled hospital studies conducted in South Africa and Tanzania, administration of large doses of vitamin A to children on admission for severe measles resulted in a 60% reduction in case fatality. The mortality reduction was higher (70%) when measles was associated with pneumonia. The protective effect of the supplements was apparently greater among infants than in older children.

Pregnancy-Related Mortality

Although several studies have examined the impact of vitamin A on child survival, maternal health has not received much attention. Recently, a community trial conducted in Nepal showed that weekly supplements of vitamin A or β-carotene reduced pregnancy-related mortality (mortality during pregnancy up to 12 weeks postpartum). However, differences were only significant when the two groups were pooled. It is surprising to note that the carotene group with lower retinol concentration showed a greater reduction in mortality than the vitamin A group. Furthermore, supplementation was associated with protection from deaths due to injuries and miscellaneous causes but had no impact on mortality ascribable to infections and obstetric causes. Further studies are under way to determine the role of vitamin A in maternal health and survival.

Morbidity

Unlike the impact on mortality, the effects of vitamin A on morbidity have been more difficult to establish. This may be due to variation in disease sensitivity to vitamin A and inherent problems in assessing the prevalence and severity of morbidity. Vitamin A does not have much effect on the prevalence of common childhood diseases but it appears to reduce the severity of potentially fatal infections such as measles, persistent diarrhea, and malaria, especially in the presence of wasting malnutrition. The protective effect becomes stronger with the episode severity. In Bangladesh, administration of vitamin A had no impact on the duration of illness or stool output in children with watery diarrhea, although children with shigellosis who received vitamin A supplements were 32% less likely to be clinically ill as compared to placebo group. In Ghana, vitamin A supplementation was associated with fewer or less severe complications and reduced rates of hospitalization for severe measles. However, vitamin A has no effect on nonmeasles respiratory disease. In some cases, supplementation was associated with an increased risk of lower respiratory infection. This appears a paradox, considering the beneficial effect of vitamin A on the epithelial structures.

Recent studies suggest that vitamin A may have a role in other infections, particularly human immunodeficiency virus (HIV) infection. In adults, low plasma vitamin A concentration was associated with increased clinical progress of HIV infection and greater mortality. It was also found that low vitamin A concentration during pregnancy was associated with increased transmission of infection from mother to child and also increased infant mortality. However, clinical trials in South Africa and Tanzania showed that vitamin A supplementation during pregnancy had no effect on vertical transmission of the disease. In another study, vitamin A supplements reduced diarrheal morbidity in HIV-infected infants. Further research is needed to explore the role of vitamin A in HIV infection.

Treatment of Xerophthalmia

All individuals with xerophthalmia (except pregnant women) should be treated with large oral doses of 200 000 IU vitamin A according to the WHO guidelines (Table 4). The age-specific dose should be given on the first and second days and again 2 weeks later. This applies to all stages of active xerophthalmia, including night blindness, Bitot's spots, and corneal lesions. Corneal xerophthalmia is a medical emergency and should be treated immediately on diagnosis and then patients should be referred to a hospital

Table 4 Treatment of xerophthalmia

Age (months)	*Dose of vitamin A (IU)*
< 6	50 000
6–12	100 000
> 12	200 000

Three doses: on diagnosis + next day + 2 weeks later.

for further treatment as they often present complex problems. Topical application of antibiotic ointment (tetracycline or chloramphenicol) is recommended for corneal lesions to prevent secondary infection. In addition, patients should be provided nutritional support and medical therapy.

Corneal xerosis shows improvement within a week but in more advanced cases (corneal ulcer/keratomalacia), healed scars remain as white opacities, resulting in partial or total blindness. Night blindness responds in a couple of days while Bitot's spots take 3–10 days and some may persist, particularly in older age groups.

Pregnant women should be treated with small doses of vitamin A, not exceeding 10 000 IU daily or 25 000 IU weekly, for about a month. Large doses of vitamin A should be avoided during pregnancy to avoid the risk of teratogenicity.

Prevention of VAD

Inadequate dietary intake of vitamin A is the primary cause of VAD and therefore the most rational approach to prevent the condition would be to improve the diet and increase the intake of vitamin A. As a short-term measure, however, vitamin A supplementation of the high-risk groups is suggested for providing immediate relief. Food fortification is also an effective way of increasing vitamin A intake in the population.

Vitamin A Supplementation

Since vitamin A can be stored in the liver, periodic administration of large doses of vitamin A has been recommended for prevention of VAD. Children between 1 and 6 years of age are given an oral dose of 200 000 IU vitamin A every 4–6 months, while those between 6 and 12 months are given half the dose. Large-scale supplementation programs are now in operation in many developing countries where VAD is a major public health problem. Universal distribution, involving administration of the dose to all preschool children, is adopted by most countries. In addition, targeted distribution of vitamin A is recommended for the high-risk groups (severe PEM, measles, siblings of children with xerophthalmia). Even in areas where a vitamin A program is in operation, these children can be given an additional dose of vitamin A if they had not received the supplement in the previous month.

Although vitamin A supplementation is widely implemented, the coverage is inadequate in most areas. It has been suggested that supplementation may be linked with the national immunization program to take advantage of immunization contacts for the delivery of vitamin A. In most national programs, the first dose of vitamin A is given along with measles vaccine at the age of 9 months. In Bangladesh, vitamin A supplements are given earlier along with DPT immunization. A higher incidence of side-effects like nausea, vomiting, and bulging fontanel were observed with this regimen, raising concern about supplementation in early infancy. However, the symptoms were mild and transient, disappearing within 48 h and follow-up studies showed no long-term sequelae. A WHO multicentric study conducted more recently confirmed the safety of vitamin A supplements but this regimen had no significant benefit for the infants.

Administration of a large dose of 200 000 IU vitamin A has been recommended for lactating mothers within 8 weeks of delivery to raise vitamin A concentration of milk and thus improve vitamin A status of the infant. However, recent studies show that the dose is not enough to produce any beneficial effect. Increasing the amount of vitamin A (two doses of 200 000 IU each) has been suggested. Further studies are needed to determine the feasibility and safety of administering two large doses within 8 weeks.

Food Fortification

Although vitamin A supplementation is a simple and effective intervention, it does not correct the underlying dietary causes. In recent years, food-based approaches have been receiving increasing attention as a long-term sustainable strategy. Food fortification offers a direct, effective, and potentially sustainable way to correct VAD. Vitamin A fortification of sugar has been successfully implemented in Guatemala and other Central American countries. However, a recent survey in Guatemala showed a high prevalence of VAD in children < 3 years, suggesting that fortification of a single food may not ensure adequate vitamin A status of the entire population. Fortification of multiple foods, including complementary foods targeted to young children, has been suggested as an alternate strategy. Other foods fortified with vitamin A include margarine in the Philippines, monosodium glutamate in Indonesia, and edible oils in India. Selection of appropriate vehicle and adequate level of fortification are important to meet the population needs.

Since VAD often coexists with deficiencies of other micronutrients like iron, folate, and zinc, fortification of foods with multiple nutrients has been suggested. Further research is needed to understand the interactions of various nutrients and to establish whether fortification with multiple nutrients would facilitate each other's beneficial effects.

Dietary Diversification

Dietary diversification, including vitamin A-rich foods, is the most effective long-term sustainable solution. Since animal foods containing preformed vitamin A are expensive, attention is focused on plant sources of provitamin A carotenoids. Green leafy vegetables, yellow/orange vegetables and fruits like papaya and mango are common sources of vitamin A. Green leafy vegetables are abundant in most countries where VAD exists, but their consumption is low due to lack of knowledge. Dietary intake can be improved through nutrition education and social marketing strategies. In areas with limited availability, horticulture programs should be undertaken to increase the production of vegetables and fruits. Large-scale home gardening programs have been successfully implemented in some countries like Bangladesh and Thailand. Evaluation studies have shown that the consumption of vegetables was higher in families with home gardens.

Bioavailability of plant carotenoids is relatively low and it is difficult to meet vitamin A requirements of young children through vegetable foods alone. Programs to improve vitamin A status must encourage consumption of a variety of foods, including animal foods like milk and eggs, at least in small amounts to improve the quality of diet. Bioavailabilty varies not only with the food source but also the way it is prepared. Intervention programs should focus on high-carotene varieties of vegetables and promote appropriate methods of processing/cooking to minimize carotene losses. Fat content of diet should be increased to promote carotene absorption. However, effective dietary changes require a thorough understanding of the local sociocultural and behavioral factors that predispose to hypovitaminosis A.

Apart from dietary interventions, improvement in environmental conditions and control of parasitic infection are important to improve the vitamin A status of the population. VAD must be seen in the broader context of malnutrition that afflicts the poor. Greater emphasis should be laid on integrated multisectoral comprehensive strategies, taking into account the community's overall needs.

See also: **Carotenoids**: Occurrence, Properties, and Determination; Physiology; **Food Fortification**; **HIV Disease and Nutrition**; **Malnutrition**: Malnutrition in Developed Countries

Further Reading

ACC/SCN Consultative Group (1994) *Controlling Vitamin A Deficiency*. ACC/SCN State-of-the-art series, nutrition policy discussion paper no. 14. Geneva: United Nations.

Beaton GH, Martorell R, Aronson KJ *et al*. (1993) Effectiveness of vitamin A supplementation in the control of young child mortality in developing countries. ACC/SCN State of the art series, nutrition policy discussion paper no. 13. Geneva: United Nations.

FAO/WHO Expert Consultation (1998) *Requirements of Vitamin A, Iron, Folate and Vitamin B_{12}*, pp. 16–31. Rome: FAO.

McLaren DS and Frigg M (1997) *Sight and Life Manual on Vitamin A Deficiency Disorders*. Basel, Switzerland: Sight and Life.

Reddy V (1986) Vitamin A deficiency control in India. In: Bauernfeind JC (ed.) *Vitamin A Deficiency and its Control*, pp. 389–404. Orlando, FL: Academic Press.

Reddy V (1994) Strategies to control vitamin A deficiency – Indian experience. In: Puri RK (ed.) *Current Concepts in Pediatrics*, pp. 211–216. New Delhi, India: Jaybee Brothers.

Sommer A and West KP (1996) *Vitamin A Deficiency, Health, Survival and Vision*. New York: Oxford University Press.

Villamor E and Fawzi WW (2000) Vitamin A supplementation: implications for morbidity and mortality in children. *Journal of Infectious Diseases* 182(suppl. 1): S122–S123.

West CE (2000) Meeting requirements for vitamin A. *Nutrition Reviews* 58: 341–345.

WHO (1995) *Global Prevalence of Vitamin A Deficiency. Micronutrient Deficiency Information System*. Working Paper #2. WHO/NUT/95.3. Geneva: WHO.

WHO (1996) *Indicators for Assessing Vitamin A Deficiency and Their Application in Monitoring and Evaluating Intervention Programmes*. Micronutrient series, WHO/NUT/96. 10. Geneva: WHO.

WHO (1997) *Vitamin A Supplements: A Guide to their Use in the Treatment and Prevention of Vitamin A Deficiency and Xeropthalmia*, 2nd edn. WHO/UNICEF/IVACG Task Force. Geneva: WHO.

WHO/CHD Immunization-linked Vitamin A Supplementation Group (1998) Randomized trial to assess benefits and safety of vitamin A supplementation linked to immunization in early infancy. *Lancet* 352: 1257–1263.

ICE CREAM

Contents

Methods of Manufacture

R T Marshall, University of Missouri, Columbia, MO, USA

Introduction

Ice cream is a frozen emulsion. The continuous phase consists of unfrozen syrup containing the dissolved substances, mostly sugars and minerals. The disperse phase consists of air cells, globules of milk fat (and in some countries other sources of fat), ice crystals, and insoluble substances including proteins and hydrocolloids.

Steps in the manufacture of ice cream are as follows: decide on composition; determine the availability and characteristics of ingredients needed to make the mix; calculate amounts of each ingredient needed; combine the ingredients of the mix; process the mix; store the mix cold until certain components have hydrated; add liquid flavorings; freeze the mix to a soft consistency while stirring and introducing air; to the soft-frozen product add desired syrups or solid types of flavorings; package; and harden and store the packaged product.

A product called Dippin Dots™ is made from ice cream mix but is not stirred while being frozen. Instead, droplets of ice cream mix are caused to fall into a bath of liquid nitrogen where they are frozen rapidly into small spheres. The frozen bead-like particles are removed, packaged, and stored at a very low temperature to prevent them sticking together. The beads can be dipped into containers for serving. These beads contain little or no air, and the ice crystals within them are very small compared with those in ice cream.

Deciding on Mix Composition

Ice cream is composed of milk fat, nonfat milk solids, sweetener, flavoring, stabilizer, and emulsifier. In a simple formulation cream provides the milk fat and some nonfat milk solids, the remainder of which is supplied by concentrated or dried skim milk. Sugar and corn syrup are the major sweeteners, vegetable gums are the stabilizer and mono- and diacylglycerols are the emulsifiers. There are many other sources of these basic ingredients (*See* **Ice Cream: Properties and Analysis**). Costs, availability, quality, stability and ease of handling are considered in deciding which ingredients to use. Once the ingredients have been chosen, amounts of each component must be calculated.

Calculating Ice Cream Mixes

As long as the sources of the major components of ice cream are few, mix calculation is relatively easy. However, multiple sources of components make the process complex. The illustration provided below is for a simple mix.

Let us assume a mix composition of 12% milk fat, 10% nonfat milk solids, 15% sugar, 0.2% stabilizer, and 0.1% emulsifier. Further assume that ingredients used are cream containing 30% fat and 6.3% nonfat solids, nonfat dry milk containing 96% solids (4% moisture), sugar, gum stabilizer and mono- and diacylglycerol emulsifier. To make 100 units of the mix we need 40 units of cream (12/0.3) and it provides 2.52 units of nonfat milk solids (40×0.063). We subtract 2.52 from the 10 units of nonfat milk solids needed and find we need to add 7.79 units of nonfat dry milk [$(10 - 2.52)/0.96$] to supply the remainder

of the nonfat milk solids. We add 15 units of sugar, 0.2 unit of stabilizer, and 0.1 unit of emulsifier. The latter two contain insignificant amounts of moisture. On addition ($40 + 7.79 + 15 + 0.2 + 0.1$) we find we have only 63.09 units of mix. Therefore, we make up to 100 units with 36.91 units of water. Had the combination of cream of lower fat content and concentrated skim milk been used, little or no water would have been required.

Compounding the Ice Cream Mix

In making the mix, liquid ingredients are combined in a vessel and solid ingredients are added while stirring. Because gums are in the form of particles that adsorb water but do not dissolve, they tend to clump when added to mix in the dry form. To overcome this they can be suspended in sugar syrup by blending for a short time. The suspension is then stirred into the mix and particles of stabilizer disperse well. Similarly, nonfat dry milk tends to form lumps when added in large amounts to a liquid. Instantizing of nonfat dry milk overcomes this problem, making it a desirable form of concentrated nonfat milk solids for use in the home. However, the cost of instantization and the high bulk of the product, compared to regular nonfat dry milk, restrict its use in industry. Instead, a device called a powder funnel delivers the dried milk into a rapidly flowing stream of water or of liquid ingredients of the mix where it disperses well and dissolves quickly. In the home or small industry, nonfat dry milk or stabilizers can be mixed with sugar before adding them to the liquid portion of the mix. When dry ingredients that tend to lump are added slowly to liquid that is being rapidly agitated, they usually disperse adequately. Therefore, some processors use this approach to combining the dry ingredients. Usually the mix is being heated as the ingredients are being combined.

Processing the Mix

Two processes are normally used in preparing the mix for freezing, namely, pasteurization and homogenization. Pasteurization destroys undesirable bacteria and enzymes while homogenization reduces the sizes of fat globules and particles of stabilizer, thus improving smoothness of the finished product. These processes necessitate heating and cooling. The minimal temperature for pasteurization depends on the amount of time the mix is to be held at pasteurization temperature.

When the process is done by the 'batch' method, the entire mix is heated in a vessel to at least 69 °C (155 °F) and held at that temperature for 30 min. Following this treatment, the mix is pumped through a homogenizer, then to a cooling device to lower the temperature to 4 °C or lower. Small batches of mix often do not receive the homogenization treatment.

The most common method of pasteurization involves pumping the mix through a high-temperature, short-time (HTST) pasteurizer (Figure 1). This device, called a platetype heat exchanger, employs hot and cold water to heat and cool the mix, respectively. In addition, a section of plates called the 'regenerator' is used to transfer heat from the outgoing pasteurized mix flowing on one side of thin stainless-steel plates to the incoming cooler mix flowing in an opposite direction on the opposite side of the same plates. This device can save up to 90% of the heat and refrigerant that would otherwise be needed. Minimal time and temperature requirements for this process are 25 s at 80 °C (175 °F). Higher temperatures with shorter times are also permitted in many countries.

Mixes must be heated before homogenization because fat must be in the melted rather than the crystalline state to be broken into smaller globules. Diameters of the fat globules of cows' milk range from about 1 to 10 μm. Homogenization reduces diameters to 2 μm and less. One globule 10 μm in diameter yields eight new globules 5 μm in diameter, and they have a combined surface area twice that of the original globule. Thus, homogenization increases the creaminess attribute of the product. The process is accomplished by pumping the mix through exceedingly small space (the homogenizer valve) at pressures of about 14 MPa.

Storing the Mix

The purpose of cooling the mix to lower than 5 °C and holding it for a few hours is to permit the added hydrocolloids, the gum-type stabilizers, to become hydrated and to permit some crystallization of the fat so it will partially churn in the freezer. The minimal time for this treatment is usually 2 h. Cooling at this stage of production decreases the amount of heat that will need to be removed in the freezer, thus increasing the capacity of the freezer by permitting a shorter time to reach the desired exit temperature.

Freezing the Mix

The Batch Process

Batch freezing involves introducing ice cream mix, often by gravity, into a cylinder containing a 'dasher' with scraper blades (Figure 2). The cylinder is open to the atmosphere and is surrounded by refrigerant. The dasher is caused to rotate within the cylinder so that

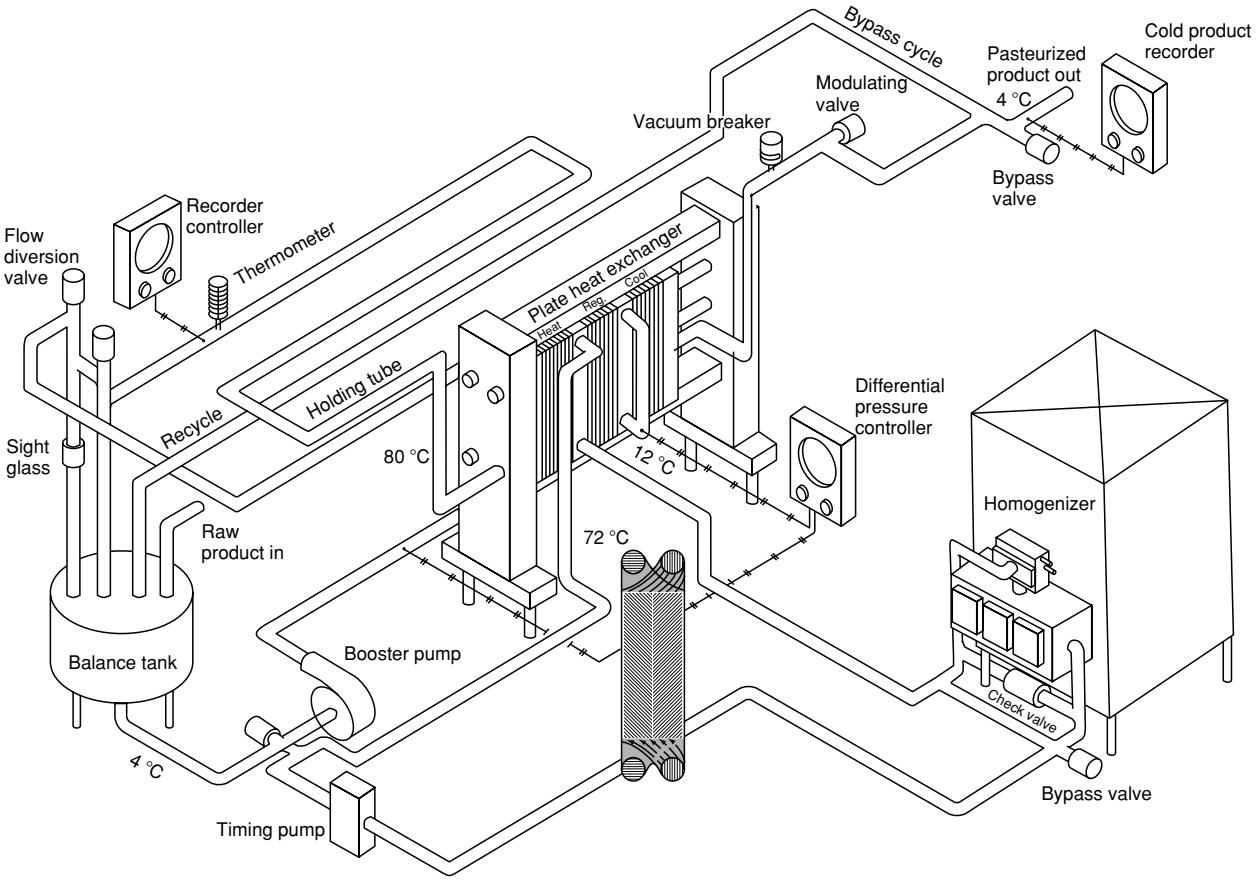

Figure 1 System for high-temperature, short-time pasteurization and homogenization of ice cream mix. The rippled surface of the heat exchanger plates is shown in the lower center. Temperatures (°C) appear near critical locations.

air will be incorporated into the mix, fat globules will be partially churned, and ice crystals that form on the inside wall of the cylinder will be scraped off by the sharp edges of the dasher blades. As the water is frozen into ice, the viscosity of the mix increases. Air bubbles are trapped in the viscous mix, increasing its volume and producing overrun. The maximal overrun permitted in ice cream is about 100%, but that amount of overrun is difficult to obtain in the batch-freezing process.

Capacities of batch freezers vary from less than 1 l (home freezers) to about 401 (commercial operations). Those freezer buckets that are refrigerated with ice and salt, mostly home freezers, must be completely closed to avoid entry of the salt brine into the ice cream. Therefore, any additives such as fruit must be placed in the mix prior to freezing or must be mixed into the soft-frozen product postfreezing. Batch freezers that are cooled with liquid refrigerant, such as the freons, usually have openings above the cylinder through which flavorings, such as fruits, nuts, and bakery products, can be added near the

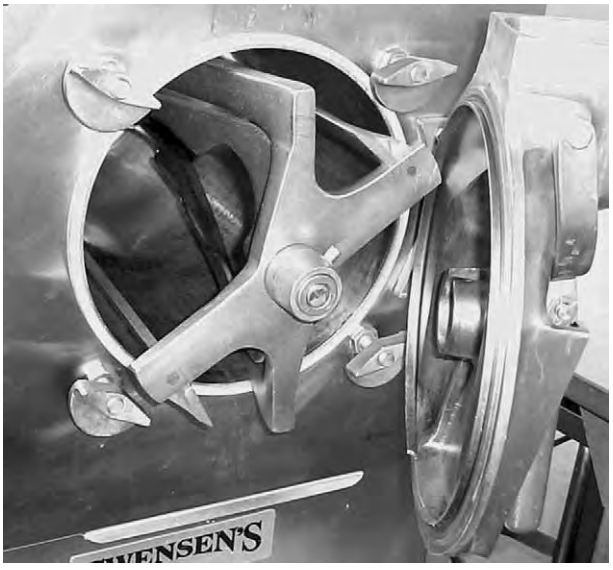

Figure 2 Batch-type ice cream freezer barrel containing beater and blades. The door has an opening through which particulate-type flavorings can be added near the end of the freezing cycle. Frozen product exits through a hatch in the door.

end of the freezing process. This allows the particles of additives to maintain their shape and reduces the adverse effect on incorporation of air bubbles. The refrigeration units are mostly contained within the freezer cabinet. These units pump liquid refrigerant into a chamber surrounding the freezing cylinder where a drop in pressure and migration of heat into the refrigerant change it to a gas. The gas is then mechanically compressed and cooled. On cooling, the refrigerant changes back to a liquid and begins another cycle through the freezer.

The Continuous Process

Most of the ice cream produced commercially is frozen in continuous freezers (Figure 3). They range in capacities from about 100 to 3000 l h^{-1} per freezer barrel. These freezers, except for small ones, are commonly refrigerated with ammonia supplied from a remote compressor unit. The refrigeration cycle starts with liquid ammonia being released into the barrel surrounding the freezing cylinder that contains the mix. The mix, being warmer than the ammonia, gives up heat to the liquid ammonia, transforming it into a gas. This gaseous ammonia is pulled into a compressor that compresses it and forces it through condensing coils where sufficient heat is removed to change the gas into a liquid. This liquid is then ready to be circulated into the freezer to take on more heat from the ice cream mix. At atmospheric pressure ammonia boils at a temperature of $-15.6\,^{\circ}C$. Reducing the pressure enables it to boil at much lower temperatures. Ice cream is frozen to -5 to $-7\,^{\circ}C$ in the continuous freezer. The internal structure of the ice cream never improves after the ice cream leaves the freezer. All of the air cells and ice crystals that can be formed are formed within the freezing cylinder. Subsequent treatments only reduce the numbers of these inclusions by forcing air cells to collapse or merge and ice crystals to grow in size while being reduced in number.

The design of the dasher of the continuous freezer affects the throughput and physical properties of the ice cream. Dashers with a solid core on to which scraper blades are attached may displace as much as 80% of the freezer cylinder volume. Rates of rotation of these dashers tend to be high. They churn the fat quickly, producing stiff ice cream that melts slowly and is favored by producers of extruded novelty products. However, the small annular space between the dasher and freezer cylinder limits the amount of mix that can be contained and increases the probability of freeze-up within the cylinder. Open dashers (Figure 4) that displace 15–50% of the freezer volume are commonly placed into freezing cylinders of comparatively large diameter. The large volume of mix within the

Figure 3 (see color plate 97) Continuous-type freezer with front removed revealing the freezing cylinder and dasher. To regulate the flow of mix and the extent of freezing, pressure is exerted on the freezing chamber by turning the handle located at the outlet for the frozen product in the cylinder cover.

Figure 4 (see color plate 98) Open-type dasher with sharp scraper blades. The dasher is partially inserted into the freezing cylinder of the continuous freezer.

chamber limits the effects of variations in refrigerant supply. Output is more uniform in temperature and overrun than the output from cylinders that have solid dashers, but the ice cream tends to appear wet and to melt rapidly. To prevent freeze-up of cylinders, most modern freezers provide for hot refrigerant gas to be turned into the space surrounding the freezing cylinder when current flow through an ammeter to the dasher motor gets excessively high, indicating very high mix viscosity. This procedure can also be done manually.

Quality of the finished ice cream can be improved by decreasing the sizes while increasing the numbers of ice crystals and air cells during freezing. Two approaches are taken to achieve these conditions. The first method is to use two freezing cylinders. The first is essentially the same as the usual chamber while the second cylinder has a dasher designed to agitate the mix vigorously and to cool it to about 3 °C lower than does a conventional freezer. Pressure within the freezing chamber is kept 4–6 times higher than in a conventional cylinder. The second method is to preaerate the mix by passing it through an emulsifier prior to entry into the freezing cylinder. The device adds air and breaks it into fine bubbles.

All continuous freezers have a variable-speed pump that supplies mix to the freezer chamber. Some also have a pump that operates at a higher speed to remove frozen mix from the freezer. A single pump must force the mix all the way to the filler, whereas placement of a pump at the outlet of the cylinder permits the mix pump to work against only the pressure in the cylinder. The outlet pump then works against the pressure within the cylinder and the downstream line pressures.

Programmable freezers are in wide use. Sensors feed data to the microprocessor that controls the system. Air input is set in proportion to the amount of mix entering the chamber. Some programmable freezers can be linked to filler machines so they sense the appropriate flow rate of the mix into the freezer.

Compared with batch freezers, the advantages of continuous freezers are: (1) volume frozen per freezer is much higher; (2) the product is usually much smoother in texture; (3) overrun can be more closely controlled and higher; (4) ingredient feeders and variegating equipment can be placed downstream from the freezer to introduce additives; and (5) ice cream can be more easily formed into special shapes and packaged with multiple flavors per package. Some continuous freezers have multiple cylinders, making it possible to extrude two or more flavors simultaneously (Figure 5).

Packaging the Ice Cream

Once the ice cream exits the freezer it can be eaten in the soft-frozen form or packaged and hardened. In the USA about 75% of frozen desserts are consumed

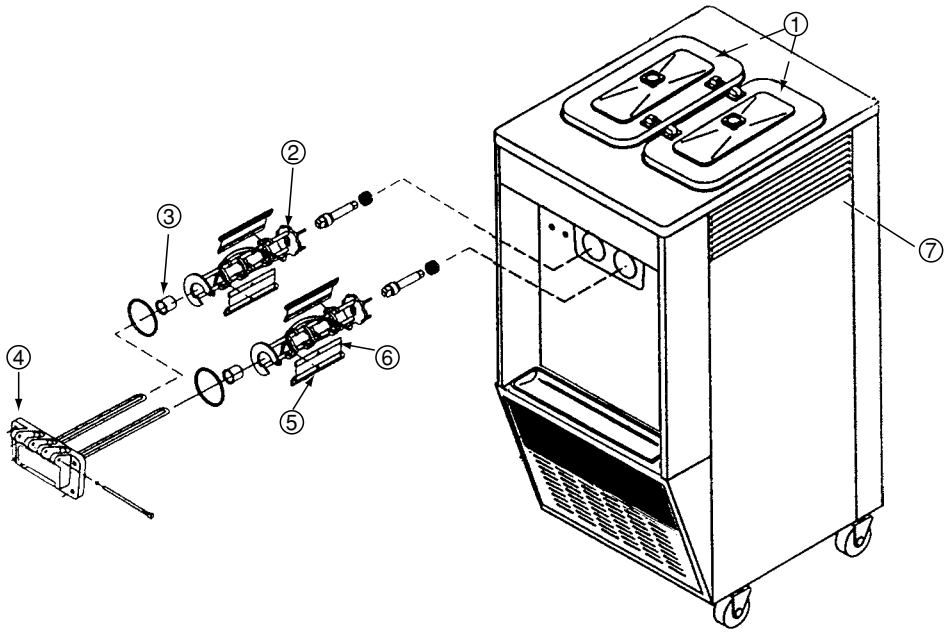

Figure 5 Twin-barrel ice cream freezer: 1, hoppers for mixes; 2, dasher (beater) assembly; 3, front bearing; 4, front cover with two outlet valves; 5, one of four scraper blades; 6, clip to attach scraper blade; 7, cabinet containing refrigeration unit. Adapted from Taylor drawing of Model H84.

in the hard-frozen form. It is imperative that packaging be done immediately after freezing to prevent melting of the frozen product and that the packages be transferred to the hardening facility quickly.

Container sizes vary from 90 ml to 11.36 l (3 to 384 fl. oz.), with most packages of the 1–2-l capacity (more than 80% of the ice cream sold in retail stores in the USA is of the 1.89-l (1/2 US gallon) capacity). Rectangular containers dominate the market. However, higher-priced premium and superpremium frozen desserts are commonly packaged in straight-wall cylindrical or nested cylindrical containers, because these containers connote high quality. A relatively new package is a tapered rectangle with rounded corners. The most popular packaging material is paperboard coated with plastic, but all-plastic containers hold a share of the market.

The package should be convenient to use; attractive; appropriately labeled, including the lot number; and capable of protecting the product from tampering, contamination, and chemical changes, especially oxidation.

Filling is done by volume, but weight per unit volume is controlled to prevent fraud. In the USA the minimal weight is $540\,g\,l^{-1}$ ($4.5\,lb\,gal^{-1}$) for regular ice cream, whereas ice creams with lowered fat content (reduced-fat, light, low-fat, and nonfat) may weigh as little as $480\,g\,l^{-1}$ ($4.0\,lb\,gal^{-1}$). Volume usually varies as little as 1% within a filling operation. In most commercial operations the container is formed on the premises from preprinted stock. Small operations may receive and fill preformed nested containers.

Hardening Frozen Desserts

It is essential to reduce the temperature of packaged ice cream to not more than $-18\,°C$ ($0\,°F$) within a few hours of packaging. Otherwise, ice crystals grow excessively in size and the product becomes coarse in texture. Lower temperatures of approximately $-30\,°C$ ($-22\,°F$) are needed if the product is to be held for more than a few days. To achieve the optimal rate of hardening, the environmental temperature is $-40\,°C$ ($-40\,°F$) or lower. Speed of hardening depends on the environmental temperature, rate of heat conductivity, distance the heat must travel, and speed of removal of heat from the surface of the container. Therefore, large volumes cool slowly because heat must travel far to get to surfaces and there is a low surface area per unit volume with large packages. Rate of removal of heat can be increased by placing packages in contact with cold surfaces. This is done effectively with rectangular packages that are placed between cold steel plates. Another way to remove heat quickly from package surfaces is to force cold air over the surfaces rapidly. This is best done with conical-shaped packages that allow passage of air between stacked packages. Stacking of packages reduces air movement, especially when package surfaces fit well together and when multiple packages are overwrapped with films to facilitate handling. Therefore, stacking and overwrapping should await completion of the hardening process. To maximize heat transfer in minimal space, manufacturers use hardening chambers through which packages are conveyed on trays, shelves, or belts as cold air is blown on to them (Figure 6). Since the environment of a

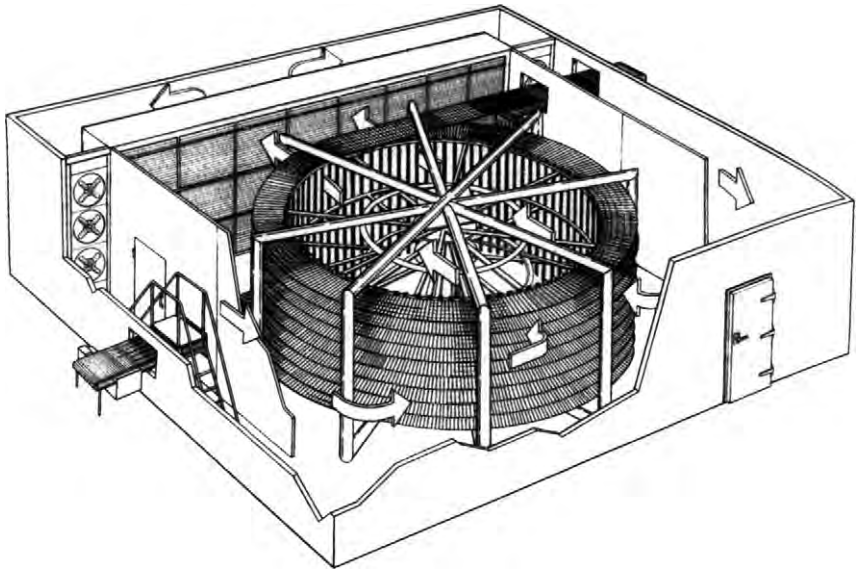

Figure 6 Hardening freezer equipped with a belt-type conveyor that spirals upward toward the storage freezer. Courtesy of Frigoscandia, Northfield, MN.

hardening facility is unfavorable to humans, it is desirable that the frozen product be unloaded mechanically so it can be stacked either mechanically or by hand on to pallets for transfer to the storage facility.

Storage of Frozen Desserts

The temperature of the storage facility need not be as low as that of the hardening facility, but the longer the product is to be held, the lower the temperature needs to be. Usually, $-30\,^{\circ}C$ ($-22\,^{\circ}F$) is satisfactory. Since the environment within a storage facility is unfavorable to humans, storage retrieval systems have been developed that allow workers to move product in and out of storage using computer-controlled mechanical devices. The system keeps a current inventory of numbers and locations of each flavor, package size, lot number, and time of production. Proper quality control and risk management procedures make it necessary that records be kept of ingredients and their sources, results of tests of composition and bacterial counts, and the net weight per unit volume. Temperatures of the storage facility should be monitored continuously, and alarm should be sounded when the temperature exceeds a specified maximum.

See also: **Freezing**: Operations; Storage of Frozen Foods; **Ice Cream**: Properties and Analysis; Dietary Importance; Microbiology

Further Reading

Andreasen TG and Nielsen H, revised by Christensen ES (1998) Ice cream and aerated desserts. In: Early R (ed.) *The Technology of Dairy Products*, 2nd edn, pp. 301–324. Gaithersburg, MD: Aspen Publishers.

International Dairy Federation (1998) Ice cream. In: Buchheim W (ed.) *Proceedings of the International Symposium*, Athens, Greece, Sep. 18–19, 1997. Brussels: International Dairy Federation.

Koxholt MMR, Eisenmann B and Hinrichs J (2001) Effect of fat globule size on the meltdown of ice cream. *Journal of Dairy Science* 84: 31–37.

Li Z, Marshall RT, Heymann H and Fernando L (1997) Effect of milk fat on flavor perception of vanilla ice cream. *Journal of Dairy Science* 80: 3133–3141.

Marshall RT (1998) Ice cream and frozen yogurt. In: Marth EH, Steele JL (eds) *Applied Dairy Microbiology*, pp. 81–108. New York: Marcel Dekker.

Marshall RT and Arbuckle WS (1996) *Ice Cream*, 5th edn. Gaithersburg, MD: Aspen Publishers.

Pappas CP (1988) Comparative laws and regulations on compositional requirements for ice-cream in the EEC. *British Food Journal* 90: 250–254.

Prindiville EA, Marshall RT and Heymann H (1999) Effect of milk fat on the sensory properties of chocolate ice cream. *Journal of Dairy Science* 82: 1425–1432.

Properties and Analysis

R T Marshall, University of Missouri, Columbia, MO, USA

Introduction

Ice cream (glaces à la crème in French; Eiskrem in German; helado in Spanish; morozhenoe in Russian; roomijs in Dutch; fløde is in Danish) is a frozen emulsion. The continuous phase consists of unfrozen syrup containing the dissolved substances, mostly sugars and minerals. The disperse phase consists of air cells, globules of milk fat (and in some countries, other sources of fat), ice crystals, and insoluble substances, including proteins and hydrocolloids.

This article discusses the unique chemical, nutritional, physical, microbiological, and sensory properties of frozen desserts. Both objective and subjective methods are used to measure these properties.

Definition of Ice Cream

According to the US standard for frozen desserts, plain ice cream must contain at least 10% milk fat and 20% total milk solids. Manufacturers are permitted to substitute up to 25% of the nonfat milk solids content by the solids of cheese whey. Furthermore, milk fat may be substituted for nonfat milk solids in 1% increments up to 14% milk fat. For example, a product containing 12% fat need contain only 8% nonfat milk solids and 2% of this amount may be whey solids. Ice cream must weigh at least $540\,g\,l^{-1}$ ($4.5\,lb\,gal^{-1}$) and contain at least $192\,g\,l^{-1}$ of food solids ($1.6\,lb\,gal^{-1}$). This limits the maximum amount of air that can be whipped into the mix during freezing to approximately 100% or the same volume as the volume of the mix that goes into the package. Standards for ice cream having lowered fat content, namely reduced fat, light, lowfat and nonfat ice creams, permit the weight to be as low as $480\,g\,l^{-1}$ ($4.0\,lb\,gal^{-1}$).

The following are the US Food and Drug Administration requirements for labeling of foods with a lower fat content:

- Reduced fat: at least 25% less fat than the reference product.
- Light: 50% reduction in total fat from the reference product, or one-third reduction in calories if less than 50% of the calories are from fat.
- Lowfat: not more than 3 g of total fat per serving.
- Nonfat or fat free: less than 0.5 g of fat per serving.

When large amounts of flavoring ingredients, such as chocolate, nuts, and fruits, are added to plain ice cream mix, the requirements for percentages of milk fat and total milk solids are reduced in proportion to the weight of the 'bulky flavoring.' In no case, however, may there be less than 8% milk fat or 16% total milk solids.

Canadian standards differ somewhat from those of the USA. Whereas plain ice cream must contain at least 10% milk fat, Canada has no minimum for total milk solids. The minimum amount of solids permitted is essentially the same in the two countries. Canada requires a minimum of 36% of total solids in the ice cream and a solids weight of $180\,g\,l^{-1}$ in the finished product. The US standard calls for minimal foods solids of $1.6\,lb\,gal^{-1}$ ($192\,g\,l^{-1}$). As in the USA, Canada permits cocoa or chocolate, fruit, nuts, and confections to displace up to 2% of the milk fat.

Canadian standards permit sales of light ice cream with 5–7.5% milk fat and ice milk with 3–5% milk fat. Total solids in each must be at least 33% and comprise at least $160\,g\,l^{-1}$. The maximal fat permitted when bulky flavors are added to light ice cream is 6%. No reduction is permitted in fat content when bulky flavors are added to ice milk.

Although countries in the European Union (EU) are attempting to agree on common compositional standards, the current minimum fat content of ice cream ranges from 5% in the UK to 10% in Germany. Italy, where ice cream is called gelato, has no requirement. Most EU countries permit the sale of milk ice (ice milk) containing, variously, at least 2.5–5% fat. Some countries permit vegetable fat to replace milk fat, provided disclosure is by obvious labeling. However, products labeled with dairy or milk as part of the name must contain milk fat only. Most of the EU countries have a minimum total solids (dry-matter) requirement of 29 or 30% for ice cream. European standards generally do not specify minimum weights per unit of volume. Most do specify which nondairy ingredients are permitted, including fruits, nuts, flavorings, sweeteners, color additives, stabilizers, and emulsifiers.

In India, ice cream can be made from cows' or buffalo milk. The minima for content of fat, protein, and total solids are 10%, 3.5%, and 36%, respectively. Fat content may be reduced in proportion to the amount of bulky flavorings added but not below 8%. The limit for added stabilizers and emulsifiers is 0.5% and for added starch is 5%.

Chemical Properties

Since ice cream is composed primarily of cream, nonfat milk solids, sweeteners, and flavorings, the major components are milk fat, lactose, milk protein, sugar, and corn syrup solids. Minor ingredients are vegetable gums and mono- and diacylglycerols. Tests commonly used to determine the chemical composition of ice cream and to determine whether it meets regulatory standards include ether extraction for fat and the hot air oven for total solids.

The fat content of ice cream is measured by extracting the lipids from a weighed sample with organic solvents, evaporating the solvents, weighing the extracted fat, and calculating the percentage. Protein is measured with the Kjeldahl test, involving digestion of the sample with strong sulfuric acid, a catalyst, and heating then distilling and measuring the nitrogen in the sample. Nitrogen content is then multiplied by a factor of 6.38 to calculate the protein value. Calories are assayed by burning a sample in a bomb calorimeter to determine how much heat is produced per unit sample. Minerals are quantified using flame photometry or atomic adsorption spectroscopy. Biometric or spectrophotometric assays are used for several vitamins.

To analyze for total solids, about 3 g of melted ice cream is weighed into a dried and weighed pan. Moisture is evaporated on a steam bath and then in an oven at 100 °C for 3.5 h. On cooling, the pan with its dry contents is weighed and the percentage of total solids is calculated.

Tests for total milk solids are not performed by regulatory agencies because of the complexity of ice cream formulae. Rather, control agencies have the option to check records of production and of dairy ingredient utilization to ascertain whether, on a general basis, a firm has met the ice cream standard.

Nutritional Properties

Ice cream products vary widely in nutrient content, primarily because of the differences in content of water, fat, nonfat milk solids, and air per unit of serving. Water content varies from about 70% in soft-serve low-fat or light ice creams to 57% in superpremium ice creams. Fat content varies from 0.5 to 16%, and air displaces ice cream mix on a volume basis from 10% in some superpremium ice creams to over 50% in fat-free ice creams. Based on nutrition labels, a serving of 120 ml (4 fl. oz. and a minimum of about 70 g) of vanilla ice cream varies in caloric content from 90 in the nonfat to 240 in the premium full-fat type.

Ice cream mix contains 3–4.5% protein so that a 70-g serving furnishes 2–3 g. Ranges of other components per 120-ml serving of nonfat through premium plain ice creams, excluding the sugar-free types,

are: carbohydrate 13–19 g; sugar 13–15 g; fat 0–10 g; and calcium 58–135 mg. Because milk fat is a good source of vitamin A, manufacturers are required to add this nutrient to lowfat and nonfat ice creams. Additionally, a serving of ice cream provides about 10% of the daily value of riboflavin plus significant amounts of thiamin, pyridoxine, and pantothenic acid. (*See* **Vitamins**: Overview.)

A serving of regular ice cream contains about 3.5 g of lactose, enough to cause intestinal discomfort in some persons who digest lactose poorly. Such persons produce insufficient lactase (β-galactosidase) in the small intestine. Lactose can be removed from nonfat milk by ultrafiltration and diafiltration or can be hydrolyzed to glucose and galactose by the addition of lactase enzyme. By using a combination of the two treatments, the freezing point of a mix can be maintained while reducing the amount of lactose below biologically significant levels.

Analyses for lactose in milk can be done by both chemical and instrumental methods. However, the presence of significant amounts of other sugars in ice cream precludes using the polarimetric method for it. In milk and milk products to be used in ice cream the concentration of lactose can be determined by removing the fat and protein by precipitation and filtration followed by measuring the degree of rotation of polarized light, which is proportional to the concentration of lactose. The 9.610-nm wavelength of infrared light is highly absorbed by the hydrogen bonds that are abundant in lactose, and this is the principle of a widely used instrumental method of analyzing for lactose in milk. For ice cream itself, high-pressure liquid chromatography is the method of choice for determining the concentration of lactose. (*See* **Lactose**.)

Physical Properties

The major physical characteristic of frozen desserts that concerns regulatory agencies is weight per unit volume of product, and this is affected by the overrun developed in the product. When enough air is whipped into a mix during freezing to cause doubling of the volume, 100% overrun has been obtained. Thus, a mix weighing 1.1 kg l^{-1} would produce 2 l of ice cream weighing 0.55 kg each. Ice creams are classified by industry as superpremium, premium, regular or trade-brand, and economy. The overrun of superpremium ice creams may be as low as 20%, whereas that of economy ice creams is usually at the maximum limit.

Overrun can be calculated for individual containers by determining the portion of mix displaced by air in the specific package as follows:

$$\%\text{Package overrun} = [(\text{mix weight} - \text{product weight})/\text{product weight}] \times 100$$

'Plant overrun' is an expression of the increase in volume of an entire lot of ice cream mix. For example, if 5000 l of mix is frozen into 9000 l of ice cream, the overrun is 80%.

$$\%\text{Plant overrun} = [(\text{volume of product} - \text{volume of mix})/\text{volume mix}] \times 100$$

Hardness of the product at the temperature at which it has the optimum consistency for dipping is an important consideration. Hardness is affected by several factors: principally, melting point, total solids, overrun, and amount and type of stabilizer. When an ice cream store keeps several containers of product in a single cabinet from which each is to be dipped or scooped, only one temperature setting is available. Therefore, it is desirable to have the melting points and overruns of all flavors of ice cream nearly the same. This is not easy to accomplish, because formulae involve several variables that affect the concentration of dissolved substances and, therefore, the melting point. The higher the concentration of water-soluble substances, the lower the freezing and melting points. The ice cream formulator must carefully choose amounts of monosaccharide sweeteners, such as glucose, to use. On an equal-weight basis, glucose lowers the freezing point twice as much as does the disaccharide sucrose while providing only three-fourths as much sweetness.

Choice of the amount and type of stabilizer depends on factors other than hardness, and especially on the need to bind water to increase shelf-life. Hardness can be determined with a puncture probe placed on a texture analyzer. The resistance to penetration of the probe is measured repeatedly in ice cream that is adjusted to a precise temperature. Comparisons to a reference curve enable decisions to be made regarding needs for changes in formulation.

When fat globules are very large, the melting rate of ice cream becomes quite slow because of excess churning of the fat during freezing. The critical mean fat globule diameter is slightly more than 1 μm. Single-stage homogenization at a pressure of 10 MPa is sufficient to produce a favorable size distribution.

Characteristics of the emulsified components of ice cream are not easily determined in tests. Sizes of ice crystals, air cells, and fat globules can be measured by examining a thin film of product using a light microscope equipped with a 'cold stage.' Application of image analysis enables large data sets to be analyzed, thus increasing accuracy in determinations of sizes of emulsified components. Electron microscopy

is a sensitive tool. However, the method is costly and time-consuming. To prevent or minimize changes in structure of ice cream during preparation for scanning electron microscopy, samples are frozen and fractured in liquid nitrogen. They are then freeze-dried and coated with a 3.0-nm-thick layer of platinum while under a high vacuum. Confocal scanning microscopy has the advantage of permitting examination of the ice cream in the native state. The instrument scans thin parallel layers of a sample and develops a three-dimensional image from the data. Laser diffraction spectroscopy is reliable for measuring fat globule size distributions in ice cream. Images obtained by microscopy can be transferred to a microcomputer for analysis. The extent of destabilization of the fat in ice cream can be estimated by comparing the absorbance of diluted samples of both melted ice cream and the mix from which the ice cream was made.

Other approaches to determining the physical characteristics of ice cream mixes include: (1) measuring the electrical charge on the fat globules by observing their electrophoretic mobilities in charged fields; (2) measuring mix viscosities over a wide range of shear and calculating the mean diameter of the fat globules from the data; and (3) determining the stability of the emulsion of fat by observing the rate of coalescence of fat globules.

Microbiological Properties

Although ice cream is not a sterile product, it contains no harmful microorganisms when it is produced by an approved process and under hygienic conditions. Furthermore, freezing destroys many bacteria, and the frozen condition of the product prevents growth of microorganisms. Except for a large outbreak of salmonellosis (see below), few cases of food-borne illness have been traced to commercially manufactured ice cream. Sporadic cases of salmonellosis have been traced to home-produced ice cream, especially that containing raw eggs. (*See* ***Salmonella***: Salmonellosis.)

Most regulatory agencies require that ice cream mixes be pasteurized. Additionally, they exercise controls over manufacturers to see that mixes receive an adequate heat treatment and that hygienic practices are followed throughout the processing plant.

The primary area of concern about microorganisms in frozen desserts is contamination of the frozen product during the addition of fruits, nuts, and flavorings, as well as during filling of containers. The microorganisms that present significant risk are *Listeria monocytogenes* and *Salmonella*.

Although illnesses are rarely, if ever, traced to *L. monocytogenes* in frozen desserts, there have been numerous recalls of frozen desserts found to contain these bacteria. Presently, in many countries detection of any *L. monocytogenes* in ready-to-eat foods is reason for recall by the processor. Infants and persons with low immunity are susceptible to infections by *L. monocytogenes*, whereas the normal healthy adult tolerates them well. They are widely distributed in the environment, including plants and animals, and are capable of living in moist, cool places. Unlike most pathogenic bacteria, *L. monocytogenes* grows at temperatures as low as 3 °C. These factors raise the probability of contamination of frozen desserts from the environment by *L. monocytogenes* compared with other foodborne pathogens. A 1989 survey of 530 samples of ice cream mix (85), ice cream (394), and ice cream novelties (51) by Health and Welfare Canada revealed only two samples that contained *L. monocytogenes*. The Working Group of the World Health Organization found that the incidence of *L. monocytogenes* in frozen desserts varied from 0 to 5.5% and that the number of colony-forming units was usually less than 15 per gram of sample.

Salmonella cells are easily killed by pasteurization, but their frequent presence in raw milk and eggs means that all frozen dessert mixes containing these ingredients should be treated to destroy the organism. The multistate outbreak of salmonellosis from ice cream that occurred in the USA during 1994 was believed to have been caused by the transport of pasteurized ice cream mix in tanks of trucks in which liquid raw eggs had been hauled. Apparently, the tanks that were used had been inadequately cleaned and sanitized prior to receiving the ice cream mix. The mix was not repasteurized prior to freezing at the plant to which it was delivered.

Microbiological analyses routinely done with frozen dairy desserts are the aerobic plate count (APC), also called the standard plate count, and the coliform count (CC) or the Enterobacteriaceae count. Standards are commonly set at 20 000 g^{-1} for the APC and 10 g^{-1} for the CC. Whereas the APC method reveals numbers of most of the bacteria capable of growing in an atmosphere of air and at moderate temperatures of about 30 °C, the CC method selects for Gram-negative, facultatively anaerobic bacteria capable of fermenting lactose and of growing in the presence of certain dyes and bile. The enterobacteria include the coliform bacteria plus other similar Gram-negative bacteria that are able to ferment glucose.

Counts can be made of culture bacteria in frozen yogurt by using a medium that is both selective and differential in function and incubating the plates at 37 °C in an atmosphere of CO_2. On modified Elliker's lactic agar colonies of *Streptococcus thermophilus* appear small and red, whereas those of *Lactobacillus*

bulgaricus appear large and white. (*See* **Lactic Acid Bacteria**.)

Sensory Properties

Ice cream has many attributes that make it a favored food of most people. Among these are a rich sweet flavor, a smooth and resistant texture, and a cold sensation that contrasts with the warmth of most other foods. The chemical and physical properties imparted by the ingredients and the processes used in manufacture and handling largely determine the sensory properties.

The human senses most used in evaluating ice cream are taste, smell, touch (feel), and sight. Only infrequently does a frozen dessert affect hearing. Although it may seem strange that smell is involved with a frozen food, it is only by smell that humans can detect flavors other than sour, bitter, salty, and sweet. These four flavors are detected by taste buds on the tongue, whereas odors are sensed in the back of the mouth and in the nasal cavity. Substances detected by smell are volatile, and the warmer the food, the higher the degree of volatility of these substances. Therefore, for the most flavors to be sensed, a frozen dessert should be consumed at as high a temperature as practicable, and it should be eaten slowly so it does not cool the mouth excessively. (*See* **Sensory Evaluation**: Practical Considerations.)

Since physical abilities vary among individuals, as do preferences for various attributes, it is necessary to select the proper sensory analysts and to insure that they meet the requirements imposed by the purpose of the analysis. Whereas trained analysts are needed to describe sensory attributes and to determine whether products meet specifications, untrained analysts are desired to determine acceptability to the public.

High-quality unflavored ice cream tastes sweet, heated (cooked), nutty, and creamy or buttery. Depending on the type and amount of flavoring added, all of these flavors except sweet may become imperceptible except to the expert evaluator. Mild flavors like vanilla tend to mask little of the background flavor, whereas chocolate masks other flavors well. The task of the ice cream formulator is to describe a mix that will provide the balance of ingredients that will give consumers a product with the most desirable flavor, texture, color, appearance, and keeping quality at a price that is acceptable. The engineer must provide a process that enhances, or at least does not damage, these properties. The manufacturer must insure that the ingredients are of high quality and that the process is performed correctly. Unfortunately, there are many factors that can affect the sensory properties adversely.

The International Dairy Federation (IDF) has established standards for evaluation of dairy products, including ice cream (IDF standard 99C). This standard defines an expert assessor as a person 'with a high degree of sensory sensitivity and experience of sensory methodology, who is able to make consistent and repeated sensory assessments of various products.' A panel consists of 'a group of expert assessors.'

The standard gives instruction on the sampling and preparation of samples; the test room; recruitment, selection, training, and monitoring of assessors; requirements for a panel; supervision; documents and equipment; and assessment. Furthermore, a scale is recommended that rates each chosen attribute in relation to the preestablished sensory specification as follows: 5, conforms; 4, deviates minimally; 3, deviates noticeably; 2, deviates considerably; 1, deviates very considerably, and 0, unfit for human consumption. Descriptions of the deviations should be given for samples scoring 1, 2, or 3. Furthermore, when scores of individual assessors fail to agree within one point, rescoring of the attribute is required. The mean value of the final score is then recorded. Terms used to describe sensory attributes should be listed in the order of most to least significance for the quality of the product.

Samples within a type should be examined in random order and presented so analysts do not know the identity. Products with a mild flavor and low fat content should be examined ahead of those with strong flavors and/or high fat content. Prior to assessment, samples should be tempered from their storage temperature of $-18\,^{\circ}C$ or below to $-13\,^{\circ}C \pm 2\,^{\circ}C$.

For calibration, at least one negative and one positive sample should be presented to the analysts who should discuss them prior to assessment. These samples should be available during the assessment session, but there should be no intercommunication among analysts during this time.

Appearance is evaluated by observing the cut surface of the frozen sample. Texture and body are determined by cutting the sample with a spoon then chewing it and letting it melt in the mouth. Flavor and odor are observed as the sample melts in the mouth. Melting quality is assessed by observing samples of equal size and temperature as they melt in flat dishes at $22 \pm 2\,^{\circ}C$.

Defects in the flavor of ice cream

These defects result from an improper balance of ingredients, the use of ingredients of poor quality, or improper processing or storage. Defective dairy ingredients can impart the following off-flavors: sour (acid), stale (oxidized), rancid (lipolyzed),

lacking freshness, or unclean. The amounts of sweeteners and flavorings used can be excessive or insufficient. Furthermore, flavorings can cause an unnatural flavor when they fail to match the true flavoring. For example, vanillin is the major component of vanilla extract, but it fails to provide many other flavor notes of vanilla extract. Additionally, some sweeteners, such as corn syrup and aspartame, impart flavors other than sweetness. Ice creams, especially the mildly flavored types, tend to absorb off-flavors during storage if not protected. Surfaces exposed to air tend to dry and to become oxidized.

Defects in texture of ice cream

Most consumers prefer ice cream that feels smooth and moderately firm when eaten. Degree of smoothness (opposite of coarseness) depends on the sizes of the substances suspended in the disperse phase of the product. These suspended substances are ice crystals, fat globules, casein micelles, particles of other proteins, and air cells. Particles that are 0.1–2 μm in diameter impart creaminess to ice cream whereas particles larger than about 3 μm cause a gritty texture. Ice crystals in freshly frozen ice cream are generally too small to be detected in the mouth, but they grow during storage when minute ice crystals melt and the water migrates to other crystals on to which it is frozen. Fat globules provide a creamy mouth feel but are usually too small to be detected individually. However, when the churning action of the ice cream freezer causes aggregation of too many of the emulsified fat globules, a slick and greasy mouth feel is imparted. Casein micelles are too small to be sensed individually, but collectively they add to firmness. Air cells affect primarily the firmness of frozen desserts. Too little air results in a heavy product, whereas too much air results in fluffiness. When air cells are quite large, smoothness is diminished.

Texture is affected by the melting point of the unfrozen dissolved phase of the frozen product. A high concentration of dissolved substances results in a low freezing point, a relatively small amount of ice crystals formed, and soft ice cream. Conversely, a low concentration of dissolved solids results in hard ice cream. Stabilizers and emulsifiers affect texture. Stabilizers bind water, reducing the amount available to form ice crystals. Emulsifiers contribute to formation of small air cells as well as moderate clumping of the fat. When used in desired amounts, emulsifiers provide stability to the emulsion, dryness, and a slowed rate of melting.

Defects in color and appearance

Consumers expect the color of foods to be properly distributed and suggestive of the flavor and freshness. Colors that are dull, faded, or uncharacteristic of the labeled flavor are defective. Inclusions, which are usually fruit pieces, nuts, candies, or bakery products, need to be abundant and evenly distributed. Syrups and preserves that are pumped into the soft-frozen dessert should be distributed evenly and with minimal spreading of the column of variegate (swirl).

Defects in melting quality

Ice cream should melt to a smooth and uniform consistency within about 20 min at room temperature. The ice cream on cones or stick novelties should melt relatively slowly to permit consumption in the frozen state. However, adjusting an ice cream formula to produce slow melt can cause slow release of delicate flavors. Slow melting can result from overuse of stabilizer gums, emulsifiers, and corn syrup solids. Products containing a high amount of air (high overrun) or fat tend to melt slowly. Air cells act as an insulator. Fat stabilizes foam structure. Treatments that destabilize proteins cause the curdy and free whey appearance of melted ice cream.

Methods of determining sensory quality

Three basic approaches are used to evaluate frozen desserts for sensory quality. These are: (1) by trained experts working alone or in small numbers; (2) by trained panelists; or (3) by untrained consumers in large numbers. Research into the effects of various treatments on specific sensory properties of ice creams usually requires use of trained panelists who employ techniques such as descriptive analysis and free-choice profiling. Development of new products commonly involves screening of several formulae by a few experts followed by tests of a small number of prospective products by large numbers of consumers. A small panel of experts usually monitors the quality of finished products.

Descriptive analysis involves selection of qualified panelists, training these panelists and using them to describe the product attributes. Training consists of several sessions during which panelists, through examination of the type of product to be evaluated and related samples, produce a list of applicable descriptive terms on which they have come to agreement by discussion and consensus. Reference samples are developed for each descriptor. Similarly, free-choice profiling involves development of descriptors, but each judge develops a personal list of terms. Panelists then rate on a point scale the intensity of each of the terms selected, with 1 = low and 9 = high intensity. Results of these types of analysis are tested statistically by analysis of variance and multivariate analysis of variance. Maps of the attributes, as they relate to treatments and to each other, can be drawn

by canonical variate analysis for descriptive analysis or by general Procrustes analysis for free-choice profiling.

A useful technique for testing the time and intensity of flavor release in a food is called time–intensity evaluation. The technique can be applied in evaluating the effect of ingredients on the release of volatile flavors such as vanilla and chocolate. In one application to ice cream, it was determined that the higher the amount of milk fat, the slower the release of vanilla flavor.

Experts who evaluate samples on an informal basis commonly refer to score cards and scoring guides such as those developed by the Committee on Dairy Products Evaluation of the American Dairy Science Association. The score card lists the defects most often found in ice cream, and the scoring guide suggests scores that reflect the seriousness of the respective defects. For example, a sample that has no observed defects is given a flavor score of 10 and a body and texture score of 5. A deduction of one point is made for a minor defect of flavor, such as lacking flavoring, or of texture, such as slightly coarse (icy). Serious defects in flavor, such as oxidized, rancid, or acid, merit lower scores of no more than 6. Color, melting quality, and package are also scored on a scale of 1–5.

Consumer panelists are chosen to represent the target market. Because they are untrained, numbers of panelists should be more than 50 and preferably 100. Panelists may be asked to indicate preferences by ranking a few samples, or they may indicate their degree of liking on a numerical scale (called hedonic evaluation). A typical scale consists of seven points where 1 = dislike extremely, 2 = dislike moderately, 3 = dislike slightly, 4 = neither like nor dislike, 5 = like slightly, 6 = like moderately, and 7 = like extremely. Hedonic tests can be analyzed statistically when sufficient panelists are used.

See also: **Freezing**: Principles; **Ice Cream**: Methods of Manufacture; Dietary Importance; Microbiology; **Lactose**; ***Listeria***: Properties and Occurrence; **Pasteurization**: Principles; ***Salmonella***: Salmonellosis; **Sensory Evaluation**: Appearance; Texture; Taste

Further Reading

Andreasen TG and Nielsen H (1998) Ice cream and aerated desserts. (Revised by Christensen ES) In: Early R (ed.) *The Technology of Dairy Products*, 2nd edn, pp. 301–324. Gaithersburg, MD: Aspen Publishers.

Bodyfelt FW, Tobias J and Trout GM (1988) *The Sensory Evaluation of Dairy Products*, pp. 166–226. New York: Van Nostrand Reinhold.

Koxholt MMR, Eisenmann B and Hinrichs J (2001) Effect of fat globule size on the meltdown of ice cream. *Journal of Dairy Science* 84: 31–37.

Lawless HT and Heymann H (1996) *Sensory Evaluation of Food: Principles and Practices*. New York: Chapman and Hall.

Li Z, Marshall RT, Heymann H and Fernando L (1997) Effect of milk fat on flavor perception of vanilla ice cream. *Journal of Dairy Science* 80: 3133–3141.

Marshall RT (ed.) (1993) *Standard Methods for the Examination of Dairy Products*, 17th edn, pp. 304–306, 469–479, 487–492. Washington, DC: American Public Health Association.

Marshall RT (1998) Ice cream and frozen yogurt. In: Marth EH and Steele JL (eds) *Applied Dairy Microbiology*, pp. 81–108. New York: Marcel Dekker.

Marshall RT and Arbuckle WS (1996) *Ice Cream*, 5th edn. Gaithersburg, MD: Aspen Publishers.

Ogden LV (1993) Sensory evaluation of dairy products. In: Yui YH (ed.) *Dairy Science and Technology Handbook 1. Principles and Properties*. New York: VCH Publishers.

Pappas CP (1988) Comparative laws and regulations on compositional requirements for ice-cream in the EEC. *British Food Journal* 90: 250–254.

Prindiville EA, Marshall RT and Heymann H (1999) Effect of milk fat on the sensory properties of chocolate ice cream. *Journal of Dairy Science* 82: 1425–1432.

Stone H and Sidel JL (1993) *Sensory Evaluation Practices*. New York: Academic Press.

Welty WM, Marshall RT, Grün IU and Ellersieck MR (2001) Effects of milk fat, cocoa butter or selected fat replacers on flavor volatiles of chocolate ice cream. *Journal of Dairy Science* 84: 21–30.

Dietary Importance

M J Tait, Birds Eye Wall's, Gloucester, UK
D J Finney and S K Narhan, Unilever Research Laboratory, Sharnbrook, UK

Energy Content

Ice cream is a much maligned product, frequently being considered, at best, a 'fun food' – unworthy of serious consideration – and, at worst, a 'junk' food, or even downright harmful! These perjorative statements have dogged the industry virtually from its inception.

In reality, ice cream is a relatively well-balanced, wholesome, easily digestible, and delicious food. It is

because ice cream is a nutritious element of the diet that it is frequently used as a meal component for hospital patients.

The composition of ice cream varies according to legislation and to the complexity of ice cream products, e.g., 'Cornetto', fancy gateaux, etc. It is therefore difficult to define precisely an average composition and energy content.

Most reputable sources seem to agree on energy values of between 180 and 210 kcal per 100 g. An average portion of ice cream – which is measured by volume, rather than weight – is taken to be between 100 and 125 ml (50–79 g), giving a calorie count of roughly 100 kcal per portion. This calorie density is modest in comparison with other desserts and confectionery, and there is a good nutritional balance of macronutrients.

Macronutrient Composition

Classically, the major ingredients of ice cream are milk solids, milk or vegetable fat, and a variety of added carbohydrates. The macronutrient composition of a typical ice cream, according to Souci, Fachmann, and Kraut, and that according to the 'Euroglace' proposal, are given in Table 1. In both instances, the nutritional quality of the milk-derived protein enhances its value. In comparison with many other dairy-based products, ice cream offers a more nutritionally balanced dietary contribution. (*See* **Carbohydrates**: Requirements and Dietary Importance; **Fats**: Requirements; **Protein**: Requirements.)

Finally, ice cream, derived as it is from milk, can be an important source of minerals, notably of calcium, and especially for children.

Micronutrient Composition

The vitamins, macrominerals, and microminerals present in dairy ice cream are listed in Table 2. Refer to individual nutrients.

Sources of Micronutrients

The total micronutrient composition of ice cream varies with the amounts and types of ingredients used in its manufacture. Because of this variation, the data given are only approximate.

Water-soluble Vitamins

In ice cream, the concentration of water-soluble vitamins, such as vitamin C and vitamins of the B complex, are similar to, or slightly higher than, those in pasteurized milk. The concentrations of these vitamins in both dairy and nondairy ice creams are generally similar. Ice cream is rich in vitamins B_{12} and B_2: a 100-g portion supplies 20 and 14% of the US recommended dietary allowance (RDA) respectively.

Fat-soluble Vitamins

The type and concentration of fat-soluble vitamins in ice cream depend on the type and level of fat used. A 100-g portion of dairy ice cream provides 11% of the US RDA of vitamin A.

Minerals

The minerals found in ice cream are similar to those in milk (Table 2). The one mineral that is clearly abundant in ice cream, and which would contribute over 10% of the US RDA if 100 g of ice cream were consumed, is calcium.

Effect of Storage on Micronutrients

Very few studies have been conducted on micronutrient losses in ice cream during storage at constant temperatures and at fluctuating temperatures. Freezing is generally regarded as the best method of preservation when judged on the basis of retention of nutrients. Substantial losses in micronutrient content during the storage of ice cream are unlikely.

Health and Nutritional Issues

As well as being rich in calcium, ice cream contains significant levels of lactose and amino acids such as

Table 1 Ice cream composition and energy levels compared to 'Euroglace' proposal

Nutrient	*Ice cream*[a]		*Ice cream*[b]	
	(g per 100 g)	*(Percentage energy)*	*(g per 100 g)*	*(Percentage energy)*
Energy (kcal)	210	100	173	100
Protein	3.9	8	3.9	10
Fat	11.7	50	8.0	41
Available carbohydrate	21.0	40	21.0	49

[a]From Souci SW, Fachmann W and Kraut H (1989) *Food Composition and Nutrition Tables*, 4th edn., p. 947. Stuttgart: Wissenschaftliche Verlagsgesellschaft GmbHXL.
[b]'Euroglace' proposal.

Table 2 Micronutrient composition of ice cream and milk

Micronutrients (per 100 g)	*Dairy ice cream*[a]	*Pasteurized whole milk*[a]	*Adult US RDA*[b]	*Percentage of US RDA supplied by 100 g of ice cream*
Water-soluble vitamins				
Vitamin C (mg)	1	1	60	1.7
Vitamin B_1 or thiamin (mg)	0.04	0.04	1.5	2.7
Vitamin B_2 or riboflavin (mg)	0.25	0.17	1.8	13.9
Vitamin B_6 or pyridoxine (mg)	0.08	0.06	2.0	4.0
Vitamin B_{12} or cobalamin (μg)	0.4	0.4	2.0	20
Nicotinic acid (mg)	0.13	0.08	20	0.7
Pantothenate (mg)	0.44	0.35		
Biotin (μg)	2.5	1.9		
Folate (μg)	7	6	200	3.5
Fat-soluble vitamins				
Vitamin A or retinol (μg)	115	52	1000	11.5
Carotene (μg)	195	21		
Vitamin D (μg)	0.12	0.03	10	1.2
Vitamin E (mg)	0.21	0.09	10	2.1
Macrominerals				
Calcium (mg)	130	115	1200	10.8
Phosphorus (mg)	110	92	1200	9.2
Magnesium (mg)	13	11	350	3.7
Sodium (mg)	69	55		
Chloride (mg)	110	100		
Potassium (mg)	160	140		
Sulfur (mg)	N	30		
Microminerals				
Iron (mg)	0.1	0.05	15	0.7
Zinc (mg)	0.3	0.4	15	2
Copper (mg)	0.02	Trace		
Selenium (μg)	N	1	70	
Manganese (mg)	Trace	Trace		
Iodine (μg)	N	15	150	

N, the nutrient is present in significant quantities, but there is no reliable information on the amount.
[a]From Holland B, Unwin ID, and Buss DH (1989) *Milk Products and Eggs. Fourth Supplement to McCance and Widdowson's The Composition of Foods*. London: Royal Society of Chemistry and Ministry of Agriculture, Fisheries and Food.
[b]From National Research Council (1989) *Recommended Dietary Allowances*, 10th edn. Washington, DC: National Academy Press.

lysine and arginine. These are known to enhance the absorption of calcium and to promote bone mineralization. Although rich in some micronutrients, ice cream is, in common with other milk-based products, comparatively deficient in iron and vitamins C, D, and E. (*See* **Amino Acids**: Properties and Occurrence; **Lactose**.)

Comparison of Dairy and Nondairy Ice Cream

Ice Creams made from cream or butter (dairy ice creams) and from vegetable fat (nondairy ice creams) are essentially comparable in nutritional value. There are differences, however, in the vitamin A content, which is higher in dairy ice cream, and the vitamin E content, which is higher in nondairy ice cream.

There are also differences in the proportion of saturated and unsaturated fats present. Ice cream made from palm oil is lower in saturated fat than dairy ice cream, whereas ice cream made with coconut oil has a higher saturated fat content than a butterfat product. (*See* **Fatty Acids**: Properties; **Vegetable Oils**: Dietary Importance.)

Sugar Content

Sugars are major ingredients in ice cream manufacture and have an important influence on the nutritional status of the product. Carbohydrates contribute 40–50% of the calories in ice creams. Sucrose is most widely used, but most ice creams contain at least three types of sweetener: sucrose, lactose (from milk) and maltodextrins (from corn syrup). The preferred level of sweetness varies form country to country, with products containing the equivalent of between 13 and 19% of sucrose, and a total of up to 25% carbohydrate by weight. (*See* **Sucrose**: Dietary Importance.)

Nutritional and Health Issues

The nutritional issues relevant to sugars in ice cream are as follows:

- the need for functional lower-calorie sweeteners;
- the interest in products more suitable for diabetics;
- the search for noncariogenic products.

Cariogenicity has not been discussed here because of a lack of information specific to ice cream. Previous extensive reports on nutritional aspects of the diet in the UK and USA have commented on the role of sugars in general in promoting the incidence of dental caries. However, to put these problems in perspective, it should be borne in mind that ice cream contributes only 1–4% of the total intake of sugars in the diet. (*See* **Dental Disease**: Role of Diet; **Sweeteners**: Intensive.)

Nutritional Claims

A number of products with specific nutritional claims have been developed, e.g., lower-calorie ice creams and diabetic products. The basis of these nutritional claims, and the compositional changes needed to support them, are considered below.

Lower-calorie Ice creams

Standard ice creams already have a fairly low-calorie density because of the high percentage of air and water in the product. Nevertheless, further calorie reduction is possible by reducing the fat or sugar content, or by replacing some of the sugars by sugar alcohols, or other synthetic sweeteners such as saccharin or Nutrasweet (aspartame).

Three sugar alcohols – lactitol, maltitol, and Palatinit (isomalt) – are now in manufacture as lower-calorie substitutes for sugars. Compared to sugars at $4\,kcal\,g^{-1}$, maltitol provides only $3\,kcal\,g^{-1}$, and lactitol and Palatinit (a mixture of mannitol and glucitol) only $2\,kcal\,g^{-1}$. (*See* **Sugar Alcohols**.)

Total replacement of sugars by zero-calorie synthetic sweeteners such as saccharin is, however, only practical if some other ingredient can substitute for the other functional characteristics normally provided by the sugars. Polydextrose, a new food ingredient, exhibits some of the functional properties of sugars without contributing any sweetness to the product. It is a randomly bonded polymer of dextrose, resistant to attack by acid and digestive enzymes. Consequently, it is only partially metabolized, with a calorie contribution of only 1 compared to $4\,kcal\,g^{-1}$ for saccharides. Polydextrose has been cleared for use in ice cream by, for example, the Food and Drug Administration (FDA) in the USA and by the UK authorities, although regulations in the UK prohibit its use in foods specially prepared for babies or young children. (*See* **Saccharin**.)

Diabetic Products

Sorbitol and fructose are the most common substitutes for sugars in products designed for diabetics. However, synthetic sweeteners and the more recently developed sugar alcohols are available as alternatives (see above).

Importance of Ice Cream

The recommendations for a healthy and balanced diet made by many governmental bodies emphasize the importance of eating a variety of different foods and the enjoyment of food. Ice cream can be delicious and a pleasure to eat, but it is also a nutritious food. It contains all the macronutrients and many micronutrients, and is a valuable source of energy.

In terms of energy, a 50-g serving of ice cream yields approximately 100 kcal. This is about that of the same weight of whole milk yogurt but two and a half times less than that of chocolate. Volumetrically, owing to the inclusion of air, the ice cream portion is twice as large as that of the yogurt and chocolate. For a child with a RDA of 2000 kcal, a 50-g portion of ice cream would provide 5% of the RDA of energy.

In conclusion, ice cream is particularly valuable for its content of high-quality protein and easily assimilated calcium. It provides many nutrients and can make a useful, as well as enjoyable, contribution to the daily intake of energy in a healthy and balanced diet.

See also: **Amino Acids**: Properties and Occurrence; **Carbohydrates**: Requirements and Dietary Importance; **Dental Disease**: Role of Diet; **Fats**: Requirements; **Fatty Acids**: Properties; **Lactose**; **Protein**: Requirements; **Saccharin**; **Sucrose**: Dietary Importance; **Sugar Alcohols**; **Sweeteners**: Intensive; **Vegetable Oils**: Dietary Importance

Further Reading

Arbuckle WS (1972) *Ice Cream*, 2nd edn. New York: AVI Publishing.

Olsen RE, Broquist HP, Chichester CO *et al.* (1984) *Present Knowledge in Nutrition*, 5th edn. Washington, DC: The Nutrition Foundation.

UK Committee on Medical Aspects of Food Policy (1989) *Dietary Sugars and Human Disease*. London: Her Majesty's Stationery Office.

US FDA (1986) *The Evaluation of Health Aspects of Sugars Contained in Carbohydrate Sweeteners*. Washington, DC: US Food and Drug Administration.

Microbiology

A Kambamanoli-Dimou, Technological Education Institute, Larissa, Greece

Introduction

Ice cream is composed of a mixture of milk products, sweetening materials, stabilizers, flavors, or egg products, which are referred to as ingredients. The wide variety of ingredients that may be used to produce different kinds of ice cream and their microbiological standard and quality greatly affect the quality of the ice cream. The specific production, processing and postprocessing storage, and storage or distribution conditions also considerally affect the quality of the final product. So, the microbiological quality of ice cream depends on many factors, the most important of which are described here.

Raw Materials (Major Components and Additives)

Although the best ice cream can be made from cream, concentrated or condensed milk, sugar, stabilizer, and high-quality flavoring, ice cream manufacturers have found it highly desirable to use a variety of ingredients available from numerous sources. Any of these can contribute bacteria, molds, and yeasts to the final product. Of course the ice cream mix is pasteurized before freezing. This heat treatment gives an adequate reduction in bacterial numbers as well as the destruction of pathogenic organisms. It cannot, however, entirely correct the hygienic quality of poor ingredients. So, carefully selected ingredients are essential for the manufacture of ice cream of the highest quality.

Liquid milk, cream, skim milk, and concentrated skim milk should have been subjected to efficient heat treatment, and must be kept under refrigeration and be used properly. Otherwise, they may contain appreciable numbers of bacteria, including some which are pathogenic (*Mycobacterium* spp., *Streptococcus* spp.). Naturally, the main organisms present in these dairy materials, which have been handled and stored under sanitary conditions, are a few spore-forming bacilli, micrococci, psychrotrophic, and thermoduric microorganisms. They are not a major health hazard but they may sometimes spoil the mix. Milk powders may also contain large numbers of spore-forming bacilli or may be contaminated by salmonellae. Numerous outbreaks of food poisoning attributed to salmonellae or staphylococci from milk powder provide evidence that these pathogens do on occasions survive in the final product. A special hazard of staphylococcal enterotoxin in ice cream may be present if whey powder is used as a source of milk solids. Careful control and storage of these powders under dry and cool conditions are necessary. (*See* **Cream**: Types of Cream; **Milk**: Liquid Milk for the Consumer.)

Several sugars are used as sweetening ingredients. These products rarely undergo microbial spoilage if properly prepared, processed, and stored. Certain yeasts and molds could be the principal spoilage flora, though some species of bacteria have also been suggested as possible spoilage problems, including *Bacillus* and *Leuconostoc* species. Also, sugar syrups are used. Osmophilic yeasts may be able to grow in these syrups, and molds may grow on the surface, so tests for yeasts should be carried out in order to minimize contamination of the ice cream from these ingredients.

Butter and butter oil (anhydrous milk fat), used as a source of milk fat, are made from pasteurized cream so they should not contain pathogenic bacteria, and a high microbiological quality is to be expected when satisfactory hygienic conditions are applied during their production and if they are properly handled. Relatively small numbers of mesophilic bacteria, coliforms, and lipolytic organisms, particularly the *Pseudomonas* species responsible for butter spoilage, as well as molds and yeasts, can be found. Butter commonly is kept refrigerated and during commercial storage is kept at about $-20\,^{\circ}C$, at which temperature no microbial growth can occur. For these reasons bacteria usually do not grow in butter, and when they do, their growth is not extensive. The flavor of good butter is so delicate, however, that small amounts of growth may cause appreciable damage to the flavor. When satisfactory hygienic conditions are applied during the production of the above ingredients and if they are properly handled (storage temperatures of no more than $-20\,^{\circ}C$ for butter, and dry, refrigerated conditions for butter oil), combined with tests for the above-mentioned microorganisms, a high microbiological quality will be insured. Fats other than milk – commonly vegetable fats – may also be used. The high temperature used during their processing and their low moisture content give raw materials containing very few microorganisms. Dry, refrigerated conditions should be used for their storage. (*See* **Butter**: The Product and its Manufacture.)

Stabilizers do not constitute an important source of bacteria, as they are usually produced by methods involving high temperatures. So, if they have been packaged under hygienic conditions they do not contribute to microbiological contamination of the ice cream mix. Attention must be paid to gelatin, which, as an animal product, may be a hazard if

produced under insanitary conditions, so it should be obtained from a reputable supplier and kept under cool, dry conditions. Emulsifiers should not present a problem, except for eggs, which may be contaminated by *Salmonella* spp., and pasteurization is required to avoid any hazard arising from their use. (*See* ***Salmonella***: Properties and Occurrence.)

Fruits (fresh, canned, or frozen) and nuts, chocolate, and cocoa, flavoring materials such as vanilla, colors and so forth are added to ice cream or are used as coatings. All of them are potential sources of hazard, particularly if they are added after the heat treatment of the mix. Fresh fruits are used and yeasts and molds, which predominate their microbial flora, may be a problem. Fruit pulps and concentrates are commercially available and manufactured at a high standard of hygiene. Canned fruits, because of their heat treatment, should be of satisfactory microbiological standard. Several types of nuts may be used in the ice cream or as a coating. They could be a hazard because they may contain molds and, possibly, mycotoxins. Coconuts can also contain salmonellae. For these reasons, it is much better if all these materials are used after heat treatment (roasted nuts, for example, or pasteurized chocolate), especially if they are added to the mix after its pasteurization. Also, they should be stored under cool, dry conditions. The examination of these materials should include a visual inspection and the enumeration of mesophilic bacteria, coliforms, yeasts, and molds.

Colors, if manufactured carelessly, may cause problems of microbial origin, especially when they are added after pasteurization. This can be avoided if they are obtained from a good supplier and stored properly.

In conclusion, a very wide variety of ingredients are used in ice cream manufacture and variations in their quality affect the microbial load of the final product. The need to use high-quality raw materials, purchased from a reputable supplier, carefully stored under good conditions which will not allow the proliferation of microorganisms, is evident. In addition, it is suggested that appropriate microbiological tests should be performed on raw materials, and the use of strict stock rotation is essential.

Hygiene During Production

The manufacture of ice cream is a relatively complex operation which includes a series of steps, all of which have some effect on the microbiological quality of the final product. So, a high-quality ice cream demands good ingredients along with sanitary conditions of processing, freezing, hardening, and distribution. As all these factors are important, they must be carefully controlled in order to produce a safe product, safeguarding the consumer's health.

All the ingredients, after they have been weighed or measured, are blended together to make a liquid mix. This mixture is then subjected to a heat treatment process, which in most countries is specified by legal requirements. This heat treatment renders the mix substantially free of vegetative microorganisms, killing all pathogens likely to be present. The ice cream mix is always homogenized, often as a step in the pasteurization process. The homogenizer is a complex piece of equipment and must be carefully cleaned and disinfected each time it is used, or the mix may be seriously contaminated. It is therefore suggested that homogenization of the ice cream mix is carried out before it is finally heat-treated, where possible. Then follow cooling and aging of the mix. After the aging period, the mix is passed to the freezer where it is subjected to considerable agitation and reduction in temperature, as well as incorporation of air. On leaving the freezer, the ice cream will normally be packaged (in family packs, individual retail packs, or other forms), frozen hard in wind tunnels at $-40\,^{\circ}C$ or in hardening rooms, and then kept at a temperature of about $-30\,^{\circ}C$ both until and during distribution. Some ice cream is sold directly from a dispensing freezer as 'soft-serve' ice cream in cones, or in various types of made-up desserts in restaurants and cafés, or from vehicles complete with their own electricity generation equipment.

The overall operation is relatively complex and each of the steps taken may have some influence on the microbiological quality of the final product. The processing methods or equipment may be responsible for wide variations in bacterial counts.

Ice cream mixes should not be kept for more than 1 h at a high temperature (more than $7\,^{\circ}C$) before being pasteurized, in order to avoid the proliferation of organisms carried in the ingredients. During pasteurization, time as well as temperature should be strictly observed in order to avoid on the one hand excessive heat treatment, which may lead to undesirable flavor changes, and on the other hand, to insure the destruction of pathogenic organisms and the adequate reduction of total bacterial numbers. Cooling of mixes to 0–$2\,^{\circ}C$ almost completely eliminates the possibility that microorganisms can grow in the mix. Cooling of mixes to such low temperatures is done efficiently in direct-expansion-type swept-surface heat exchangers. Such cold temperatures add several days to the shelf-life of mixes sold to other firms for freezing. The cooling must also be rapid and the mix must be kept at that low temperature until it is frozen, otherwise, any viable organisms may proliferate. This can lead to a product with a high microbial count and

possibly to a disease outbreak. The same danger exists if the cooling system fails during the aging of the mix. In this case, the mix must be discarded because, although a repasteurization will kill the organisms, it will not destroy possible toxins already present. Normally, the mix should be frozen within 24 h of heat treatment, as undue prolongation of storage may lead to proliferation of psychrotrophic organisms, with a serious risk of spoilage of the mix.

The incorporation of air into the ice cream mix during the freezing process may cause airborne contamination, so air must be admitted through filters to prevent the ingress of organisms. For some products, the ice cream is blended with water ice on leaving the freezer, and may be frozen on a stick or in a cone (single portions) and covered with a chocolate or flavored couverture, together with broken biscuits or nuts. Addition of ingredients to the pasteurized mix constitutes a critical control point in ice cream manufacture, because no further lethal process is given to the product. Therefore, it is important that all operations at this point be done in a sanitary manner and that ingredients be free of pathogenic microorganisms. The most likely pathogens to be encountered at this point are *Listeria monocytogenes* and *Salmonella*.

The packaging materials may occasionally cause trouble, but there should be no problem if they have been handled and stored under hygienic conditions.

In addition to careful operation of the equipment, proper cleaning and sanitizing of the plant and equipment are of great importance; therefore, special care is demanded. All plant equipment that comes into contact with ice cream or ice cream ingredients must be carefully cleaned and sanitized immediately after use; this is usually done at the end of each day's operation and before it is used again or it may even be necessary to clean the plant thoroughly before further processing during the same day. Not only is the processing equipment important, but ancillary equipment, in particular at the final sales point, must be kept in a satisfactory hygienic condition. Poor cleaning and sanitizing of the plant and equipment may lead to pockets of ice cream residues where intense proliferation occurs, which results in recontamination of the pasteurized mix with a large number of bacteria. For the ordinary soft-serve or dispensing freezer, manufacturers usually lay down strict cleaning routines which must be followed carefully. Isolation of *L. monocytogenes* from an ice cream plant shows that dangerous microorganisms are able to survive in the food-processing environment and improved and precisely targeted cleaning and disinfection practices are of great importance in the eradication of those microorganisms.

Clean surroundings are essential if equipment is to be kept in a hygienic condition. All rooms, especially toilets and locker rooms, must be kept as clean and sanitary as the area immediately surrounding the packaging equipment. Surrounding activities (e.g., sewerage plants, rubbish tips) often represent potential sources of contamination, and birds, rodents, and insects are very important vectors of such contamination. In addition to preventing access of pests to the process area, it is important that the factory yard is kept free of food waste, rubbish, and spilled material that might attract birds, rodents, and insects. All such material should be kept in lidded containers and removed on a daily basis. Pet animals, such as dogs and cats, similarly have no place in a food production factory.

Operations must be segregated to minimize the chances of pathogenic microorganisms being carried from raw materials to finished products. Persons handling raw milk or cream must not be allowed access to rooms where pasteurized products are exposed unless those persons have first changed their clothes completely and have disinfected themselves. Room air pressures should be maintained at successively higher levels, from the mix room to the processing room, to the freezing operation, and to the packaging operation. Thus, the flow of air will be away from the most critical area, i.e., the packaging room. The supply of hot and cold water must be unrestricted and facilities for disposal of both liquid and solid wastes must be adequate. All water that is used in food formulation, or will be used on (or could gain access to) food contact surfaces should be of potable quality and be stored in enclosed tanks and distributed in piping that is completely segregated from other pipe systems. Potable water may be derived from public mains supplies or other sources such as boreholes which must be protected against contamination by surface water or underground contamination from drains, seeping from farm or industrial tips, and similar potentially hazardous areas. Whatever its origin, the water should be routinely examined microbiologically at point of entry to the site and at the point of use, particularly if there is on-site storage.

The hygienic standards of the workforce are crucial to the ice cream manufacturer. No worker who has not been adequately taught the necessities of personal hygiene and approved practices should be allowed to perform tasks in the plant. Every employee must dress in a clean uniform, wear a hair restraint, wash and sanitize hands, disinfect footwear on entry to the process area, and refrain from touching any product contact surface without properly sanitizing the hands, or gloves if worn. Proper sanitary practices are essential to

the ice cream plant. No one should be allowed to enter the processing environment who is not familiar with the required sanitary procedures or who does not conform to the required dress and personal hygiene measures. Freedom from chronic contagious diseases should be confirmed yearly by medical examination. Provided that the preparation of ice cream is conducted in a closed processing cycle with modern industrial equipment of high hygiene standards, the opportunities for contamination by human contact are few.

Tests carried out during the production process should be indicative of the standard of hygiene.

Storage (Quality Deterioration)

Microorganisms are unable to grow in ice cream stored at correct temperatures, although many survive for extended periods. It is a perfect substrate for microbe development as far as its composition is concerned. It has all the essential components (sugar, proteins, water), oxygen, as well as a relatively high and suitable pH. High temperature is the only agent not available. If a rise in temperature does occur, all the conditions are in place for the development and proliferation of microbes that may exist. Microbiological considerations, therefore, primarily involve the elimination of vegetative pathogens by pasteurization and the prevention of recontamination due to insanitary processing and packaging, at all stages up to point of sale, the microbiological status of ingredients with particular reference to thermoduric organisms and preformed toxins, and the prevention of microbial growth before freezing.

Ice cream may be sold direct from the freezer as a soft-serve product, or it may be further reduced in temperature and frozen in wind tunnels at $-40\,^{\circ}C$ or in hardening rooms, to produce 'hard' ice cream, which will be stored at a temperature of about $-30\,^{\circ}C$ until it is sold. Deep-freezing stabilizes the microbial content of ice cream: microorganisms found in it no longer proliferate. Some sensitive species (Gram-negative) die and their populations decrease. Even if the period between freezing and final sale is several months, there will be little change, if any, in the microbial content of the ice cream. Extensive research has shown that both *Mycobacterium* and *Salmonella*, as well as many other less harmful but often more resistant types, can survive at the low temperature of storage for very long periods. They do not multiply provided that the temperature is low enough for the ice cream to remain hard; in effect, the microbial quality of ice cream is 'locked in' by the hardening process.

Problems of quality deterioration can appear when there is a delay between pasteurization and freezing, mainly a delay from heat treatment to cooling of the mix, or the aging of the mix at a temperature higher than $4\,^{\circ}C$ due to ignorance or failure of the freezing system. Spoilage can also occur, in cases of melting and refreezing of the product resulting from temperature fluctuations or failure of the freezing systems. Special care is needed with the mix for soft-serve ice cream that has to be transported, often for long distances, in trucks to retail soft-serve stores or stands where it is kept soft-frozen and dispensed to consumers. Both contamination and temperature abuse of the mix may easily occur. Furthermore, refrigeration space is usually limited, and adequate facilities for cleaning and sanitizing the freezer and the associated equipment are often lacking or are, at best, marginal.

Under these conditions, especially when inadequate practices have preceded storage, the ice cream is overloaded with microbes which lead to quality deterioration or even to cases of food poisoning. Reference is made in the Further Reading section to cases of food poisoning following the consumption of ice cream contaminated by microbes, such as *Staphylococcus, Salmonella, Shigella, Listeria*, and *Streptococcus* group A organisms. More serious cases may occur with homemade ice creams, where a combination of faulty practices, such as the use of raw milk, eggs contaminated with *Salmonella*, inadequate heat treatment, no rapid cooling following heat treatment, and contamination by infected persons give rise to products with high microbial loads, especially of pathogenic bacteria which can survive many months in contaminated ice creams. (*See* ***Listeria***: Properties and Occurrence; ***Shigella***; ***Staphylococcus***: Properties and Occurrence.)

Problems at Point of Sale

The largest proportion of microbiological problems with ice cream products, in general, is due to poor techniques of selling and serving at the final point of distribution. This plays a vital role in the microbiological load of ice cream. Even when the greatest care has been taken to produce an ice cream of the highest quality, it is still liable to contamination at the point of sale.

The method of sale has a major bearing on the level of contamination to which the product is subjected. Although much ice cream is retailed in its final packaging, a significant quantity is portioned from bulk packs at the point of sale.

Prepackaged ice cream that is sold in individual package is considered safer because it has to be handled only by the consumer in its wrapping and can be contaminated only by the consumer. Attention

must be paid to the right storage temperature without temperature variations or failure of the freezing system. Greater degrees of contamination may occur in ice cream which is portioned from bulk packs in restaurants or coffee shops, mobile vans (complete with their own electricity generation equipment), and kiosks. This process is the weak link from the hygiene viewpoint since many outlets have only limited facilities for hand washing and utensil sterilization. In order to minimize the possibility of contamination, the equipment (servers, wafer holders, and so on) has to be kept free of all residues of ice cream, which might otherwise melt and allow the growth of bacteria to recommence. Whenever possible, these items of equipment should be kept in running cold water. If they have to be kept in a jug of water, this water must be changed regularly to avoid it becoming a source of contaminating bacteria. The personal hygiene of the server is also important. Persons handling and serving food should be trained to live up to definite and rigid standards of personal hygiene. Hair, hands and fingernails, uniforms and shoes have to be clean. Soap, water, and single-service towels should be easily accessible.

Soft-serve ice cream sold directly from a dispensing freezer can become contaminated very easily unless stringent precautions are taken. The product is usually manufactured at the point of sale, which may be a specialist outlet or café, a nonfood outlet such as a gas (petrol) filling station, or a mobile outlet or kiosk. Soft-serve ice cream may be manufactured from a conventional mix produced on the premises, from an ultrahigh-temperature-processed, aseptically packaged mix, or from a spray-dried powder mix. Powdered mixes may be formulated for reconstitution in either hot or cold water. Hot-water mixes are preferable with respect to hygiene, but cold-water mixes are often considered more convenient. Attention must be paid to the reconstitution of the mixes; this must be done under satisfactory hygienic conditions to prevent the proliferation of *Salmonella*, which may survive if mixes are not prepared carefully.

Soft-serve ice cream freezers and ancillary equipment should be dismantled, cleaned, and sanitized daily. It must be recognized that maintenance of the necessary hygiene standards can be more difficult in an environment primarily concerned with retailing than in one wholly concerned with manufacturing. Particular difficulties may be encountered in outlets which are predominantly nonfood, such as filling stations, and those with inherently limited facilities, such as kiosks. The self-pasteurizing soft-serve freezer offers at least a partial solution. Such equipment is designed to heat the mix and machine surfaces in contact with the mix to an appropriate temperature and time before cooling rapidly to about 4 °C. It is usual to 'pasteurize' daily, usually at the end of the working day, and restrict full cleaning and sanitization to a weekly basis. It must be emphasized that self-pasteurizing freezers are not intended to process unpasteurized mix. Generally, special dispensing freezers should be kept in constant operation, and placed inside the shop with the taps facing the interior. In addition, they must not be directly exposed to the sun, dust, or flies.

Probably the most serious and dangerous sources of contamination are operation and serving. Many major food-poisoning outbreaks have been caused by human contamination. Cases of typhoid fever, including deaths, have been reported to be caused by ice cream contaminated by the manufacturer who was a urinary excretor of *Salmonella typhi*. There has been a case of *Shigella* dysentery caused by an ice cream that was accidentally touched by a monkey. Also, outbreaks involving salmonellae and staphylococci have been reported. The personal hygiene and habits of vendors at sale points are important. Education, in addition to medical inspection, is absolutely necessary and no employee must be allowed to work without full medical clearance.

Finally, birds, rodents, insects, and pet animals have no place at the retail selling point.

Many countries, recognizing the significance of ice cream in relation to public health, have legislated as to the conditions and methods to be used for heat treatment and subsequent storage and sale.

See also: ***Bacillus***: Occurrence; **Freezing**: Storage of Frozen Foods; ***Listeria***: Properties and Occurrence; ***Mycobacteria***; **Pasteurization**: Principles; **Spoilage**: Bacterial Spoilage; ***Staphylococcus***: Properties and Occurrence; **Storage Stability**: Mechanisms of Degradation

Further Reading

Frazier WC and Westhoff DC (1988) *Food Microbiology*, 4th edn. New York: McGraw-Hill.

Hennessy TW, Hedberg CW, Slutsker L *et al*. (1996) A national outbreak of *Salmonella enteritidis* infection from ice cream. *New England Journal of Medicine* 334: 1281

International Commission on Microbiological Specifications for Food (1974) *Microorganisms in Foods 1*. Toronto: ICMSF/University of Toronto Press.

International Commission on Microbiological Specifications for Food (1980) *Microbial Ecology of Foods*, vol. 2. *Food Commodities*. London: ICMSF/Academic Press.

Jervis DI (1992) Hygiene in milk product manufacture. In: Early R (ed.) *The Technology of Dairy Products*, pp. 272–299. Chichester: John Wiley.

Mantis AI (1993) *Hygiene and the Technology of Milk and its Products*. Thessaloniki: Kiriakidis.

Marshall R and Arbuckle WS (1996) *Ice Cream*, 5th edn. New York: Chapman & Hall.

Miettinen M, Björkroth J and Korkeala H (1999) Characterization of *Listeria monocytogenes* from ice cream plant by serotyping and pulsed field gel electophoresis. *International Journal of Food Microbiology* 46: 187

Rothwell J (1985) Microbiology of frozen dairy products. In: Robinson RK (ed.) *Microbiology of Frozen Foods*, pp. 209–232. London: Elsevier.

Rothwell J (1990) Microbiology of the ice cream and related products. In: Robinson RK (ed.) *Dairy Microbiology*, vol. 2, pp. 1–39. London: Elsevier.

Varnam A and Sutherland J (1993) *Milk and Milk Products*. Food and Commodities series no. 1. London: Chapman & Hall.

Immobilized Enzymes *See* **Enzymes**: Functions and Characteristics; Uses in Food Processing; Uses in Analysis

IMMUNOASSAYS

Contents

Principles

F S Chu, University of Wisconsin, Madison, WI, USA

Introduction

Antibodies are a group of globular proteins called immunoglobulins (Ig), which are formed in response to the invasion of foreign substances in vertebrates. Igs are classified into five groups: IgA, IgG, IgE, IgD, and IgM. IgG, which has a molecular weight of 150 kDa and is composed of two heavy chains (50 kDa) and two light chains (25 kDa), is one of the most common serum Igs (80% of total Igs). Antibodies are capable of binding with foreign substances noncovalently with high affinity and, thus, to inactivate them.

Substances that induce the defensive action by producing antibodies in the host are called antigens, and are characterized by their immunogenicity and their antigenicity. Immunogenicity is the capability of a molecule to induce antibody formation. Thus, the term 'immunogens' specifically refers to those substances capable of inducing production of antibodies. The antigenicity, or antigenic specificity, however, is the characteristics/capabilities of the interaction of the antigen with antibodies. The interaction involves small sites, which are called 'antigenic determinants' or 'epitopes' on the surface of the antigen molecules.

In contrast to antigens, low-molecular-weight compounds (less than 2500 Da) are not immunogenic. These compounds are called haptens and can only become immunogenic after conjugation to a high-molecular-weight protein carrier.

Although it has been known for more than a century that vertebrates are capable of producing specific antibodies after invasion by foreign substances, the complexity of immunoreactions has only been revealed in the last two decades. With the advances in understanding of the mechanism of immunoreactions, the specific interaction between antibody and antigen/hapten has been used as the basis for several analytical techniques called immunoassays. These assays, which are characterized by their high sensitivity, specificity, and simplicity, were originally developed for the analysis of clinical samples. Recent development has led to the application of immunoassays for food and agricultural chemicals, including both normal and abnormal food components, additives and contaminants, as well as agricultural chemical residues.

In contrast to clinical samples, immunoassays for specific substances in foods and feeds are more

complicated because: (1) specific antibodies must first be raised against a wide range of food components and chemicals. Some are antigens, whereas others are nonimmunogenic haptens, most of which are lipid-soluble; (2) foods are very complex systems and sample matrices often interfere with the assay; (3) changes in physicochemical properties of food components/contaminants during processing of foods often complicate immunoassays because the antibody may not recognize the denatured species. Nevertheless, much progress has been made and numerous review articles on the immunoassay of food components/contaminants have appeared. The reader should consult these reviews for details.

Principles

All immunochemical techniques are based on the specific noncovalent interaction between antigens/haptens and specific antibodies. Earlier applications of immunochemical methods for analytical purposes were confined to the interaction between antibodies and high-molecular-weight antigens that led to the precipitation of the antigen–antibody complex for quantitation. However, precipitates are also not formed between antibody and haptens. Investigations in the last three decades have led to the development of more sensitive immunochemical methods. Thus, the term 'immunoassay' now also includes the modern immunochemical techniques which involve the use of labeled marker for the antigen or antibody and innovative approaches for the separation of free antigen and antigen–antibody complex. As a result of the development of new instruments and incorporation of label markers in the assay system, new innovative precipitation methods with high sensitivity are now also available.

For simplicity, the binding of antibody with a simple hapten, which has only one antigenic determinant site (univalent), is used to illustrate the antigen–antibody interaction. As shown in eqn (1), the interaction depends on the equilibrium or affinity constants ($K = k_a/k_d$), which is the ratio of the association (k_a) and dissociation rate constants (k_d) and the concentrations of the reactants, i.e. [antibody] and [antigen].

$$[\text{Antibody}] + [\text{Antigen}] \underset{k_d}{\overset{k_a}{\leftrightarrow}} [\text{Antibody–Antigen}]$$

In general, the K values for antibody–antigen interaction are in the range of 10^6–$10^9\,\text{l mol}^{-1}$, but they may vary from 10^3 to $10^{14}\,\text{l mol}^{-1}$. Antibody with a high K value is more preferable for the assay, but any reaction with a K value above $10^9\,\text{l mol}^{-1}$ would be feasible. For analytical purposes, either a labeled antigen or a labeled antibody marker is present in the system. With an adequate method to separate the antibody–antigen complex from the free antigen and antibody, as well as a sensitive method for detection of the markers in the free or bound form, the concentration of the analyte can readily be determined. In general, a standard curve is established and then used to calculate the concentration of unknown sample. Thus, it is critical to have a pure standard from a reliable source which should also be tested frequently for its stability.

Preparation of Immunochemical Reagents

As indicated above, an effective immunoassay depends on the availability of: (1) a specific antibody; (2) a sensitive marker antigen or antibody; and (3) an effective separation method. Methods for the preparation of marker antigen or antibody and methods for the separation of the free and bound species of antigen and antibody vary considerably with different immunochemical methods used. Approaches for the preparation of immunogen and antibody as well as methods for characterization of antibody can be generalized as follows.

Preparation of Immunogens

Antibodies against major food proteins and naturally occurring foodborne microbial toxins/contaminants have been produced and used in different types of immunoassays. High-molecular-weight substances such as proteins and polysaccharides can be used directly as immunogens. In general, well-characterized and pure immunogens are used for generating antibodies. With the development of monoclonal antibody technology, partially purified proteins or even crude cellular extracts from pathogenic bacteria have also been used successfully for immunization as long as an effective method in selecting specific hybridoma cell lines is available. Because of structural changes for some proteins and other macromolecules during food processing that alter the immunogenic site, purified proteins are not necessarily used to generate antibodies from components in food-related products. Rather, for determination of such materials in processed foods, specific antibodies must be raised for corresponding denatured proteins. Thus, depending on the specific components to be analyzed, one must consider selecting an immunogen which has comparable properties to that actually found in the food.

Haptens must first be conjugated to a protein/polypeptide carrier before immunization. Methods for

the conjugation of various low-molecular-weight naturally occurring toxins, including antibiotics, mycotoxins, phycotoxins (marine toxins), pesticides, and other agricultural chemicals, to protein carriers have been established. Proteins, including bovine serum albumin (BSA), modified BSA, and keyhole limpet hemocyanin, are most commonly used. In general, if the hapten has a reactive group such as a carboxylic or an amino group, it can be conjugated to the protein directly via one of the following approaches: water-soluble carbodiimide method, mixed anhydride method, condensation method in the presence of formaldehyde (Mannich reaction), cross-linking with glutaraldehyde, activated ester method through the formation of *N*-hydroxysuccinimide (NHS) esters, 1,1′-carbonylimidazole and m-maleimidobenzoyl-*N*-hydroxysuccinimide ester, and many other innovative methods. Derivatization is necessary if no reactive group is present. Before derivatization or conjugation, one must select appropriate strategies to insure that certain portions of the hapten molecule are exposed. Not only the amount of hapten conjugated to the protein molecule is important in determining the effectiveness of the conjugates as an immunogen, but also the position of the side chain in the hapten where conjugation is made, as well as the space between the hapten and protein molecule are critical. However, a conjugate containing a large amount of hapten may not necessarily be a better immunogen. Antibodies have been obtained from rabbits immunized with immunogen containing only 2 mol of hapten per mole of carrier protein. The optimum amount of hapten coupled to a protein that yields a good immunogen should be experimentally determined, but generally is in the range of 10–20 mol per mole of protein.

Rapid progress in cloning genes for many proteins/enzymes/toxins in food and agricultural commodities has led to another approach for the preparation of immunogens. By knowing the DNA sequence of a gene, the amino acid sequence for a specific gene product in which we are interested could be deduced. Through structural analysis, the highly immunogenic fragment in this protein could be identified and chemically synthesized. These peptides could then be conjugated to a branched-chain lysine to form a multiple-chain immunogenic polypeptide for immunization. Using this approach, it is possible to generate an antibody for a protein even before it is purified and characterized.

Preparation of Antibodies

Two approaches have been used for the production of antibodies. The classic method involves immunization of animals and then obtaining the antibodies from the sera of the immunized animals. Antibodies obtained by this approach are polyclonal antibodies (PAbs) because they are derived from different B lymphocyte clones. In this method, immunogens are mixed with adjuvant and then injected into animals, most commonly rabbits, intradermally or subcutaneously through the thigh or on the back at multiple sites. For production of a large quantity of antiserum, large animals such goats and horses are used. Other animals, such as pigs and mice, have also been used. Production of antibody has been optimized by immunization of chickens and then isolation of antibody from eggs.

The amount of immunogen used in the immunization varies considerably. In general, 1–2 mg protein is used for each rabbit. To avoid toxicity of some food-borne bacterial toxins, a smaller amount of toxins or toxoids are used in the initial immunization (e.g., 5 μg) and this is followed by a gradual increase in dose for subsequent booster injections. For haptens, 200–500 μg of conjugate is generally used in each injection. Although antiserum obtained from animals could be used directly in the immunoassay, simple purification such as ammonium sulfate precipitation is generally necessary. Further purification of antiserum through various chromatographic methods can improve the efficacy of immunoassays.

If the immunogen is highly immunogenic, antibody titers, as determined by different approaches such as immunodiffusion, radioimmunoassay (RIA) or enzyme immunoassay (EIA), usually start to increase at 4–7 weeks. Subsequent booster injections are generally performed once a month. Since PAbs are heterogeneous, it is advisable that the purified antigen be used in the titer determination. For haptens, a specific hapten-marker must be used. Use of inadequate antigen or hapten-markers usually fails to provide information regarding whether the immunogen and immunization protocols are effective.

Although production of PAbs is relatively simple and the affinity of Pabs is generally high, they are heterogeneous, and their supply is limited. With increasing demand for large amounts of antibodies with homogeneous properties for immunoassays, as well as advances in hybridoma technology, specific monoclonal antibodies (MAbs) for a number of food components and contaminants in foods have also been produced. In this method, spleen cells obtained from mice immunized with immunogen or hapten–protein conjugates are fused with myeloma cells, followed by propagation of the hybrid cells, and selection and characterization of the clones, as well as isolation and characterization of the antibody. Production of mAbs can either be carried out in tissue culture or in mice, in which ascites fluids are collected. Thus,

mAbs are derived from a single B lymphocyte clone and are immortal. Once a hybridoma cell line is obtained, there is an unlimited supply of homogeneous antibodies with unique properties. Similar to the production of PAbs, the key to success in obtaining a useful hybridoma cell line lies in the effectiveness of immunogens as well as the immunogen-marker or the hapten-marker. The importance of an effective immunoassay protocol in selecting the positive clones cannot be overemphasized. Many new stable hybridoma cell lines for production of antibodies against various food components, contaminants, naturally occurring toxicants, and agricultural residuals are now available. Rapid progress in molecular cloning of MAbs has also led to the application of this technology for the production of antibody important to agricultural and food scientists and several laboratories have succeeded in this regard.

Antiidiotype Antibodies

The development of immunochemical methods for agriculture and food analysis has led to a great demand for specific antibodies and related immunochemical reagents for the assay. An alternative approach for preparing immunochemical reagents is through generating antiidiotype antibodies (Ab2), which have gained wide application in diagnostic and therapeutic areas for large molecules. In the last few years, Ab2 against small-molecular-weight haptens, including insecticides, herbicides, hormones, mycotoxins, and phycotoxins, have been successfully developed. Some Ab2 have not only been shown effectively to mimic the biological functions of the haptens, but have also been used as the immunogen to generate antiantiidiotype antibodies (Ab3) with specificity similar to the original antibody (Ab1). Both Ab2 and Ab3 have been shown to be effective in immunoassays.

Characterization of Antibodies

Antibodies obtained from the above methods should be well-characterized before they are used in the immunoassay. In general, both the type of immunoglobulin and the specificity of the antibody to various analogs of the immunogen/hapten should be determined. The typing of immunoglobulin is generally done by using commercially available kits. Information on the immunoglobulin type is important in selecting adequate markers as well as an appropriate method for separation of free antigen or antibody from the antigen–antibody complex in the immunoassay. Information on the specificity will provide an accurate assessment as to whether structurally related compounds will interfere in the immunoassay. They could also be used to estimate the affinity of the antibody to the immunogen/hapten and its structurally related analogs.

The specificity of the antibody, usually expressed as the cross-reactivity of an antibody, is primarily determined by the type of immunogen that has been used. A minor change in the immunogen structure could generate a different antibody. Some antibodies are very specific, whilst others show a broad specificity. The apparent cross-reactivity of the antibody also varies with the type and concentrations of the marker antibody or antigen and the format used in the immunoassay. Thus, in addition to knowing the specificity of the antibodies generated by a specific immunogen, the conditions under which that specificity was determined should be stated. Preferably, the 'apparent cross-reactivity' should be reevaluated under the conditions that are to be used for the assay. The specificity of the polyclonal antibodies can be improved by removal of nonspecific antibodies through immunoabsorption.

For haptens, the site in the hapten molecule linking to the carrier protein is important. The degree of cross-reactivity of these antibodies with structural analogs of the hapten used in the conjugation should be determined. The accuracy of immunoassay of hapten in naturally contaminated samples is generally affected by both the specificity of the antibody and the possible presence of structurally related analogs of the hapten in the sample as well as by some substances with similar structures of the hapten.

Experimentally, the cross-reactivity of antibody is determined by an immunoassay where various structurally related analogs of immunogen/haptens at a wide range of concentrations are used to compete with the binding of the marker ligand with the antibody in the assay. The concentration at 50% inhibition (I_{50}) of the binding is generally used as the basis to calculate the relative cross-reactivity for each analog. A typical example of such a competitive RIA and enzyme-linked immunosorbent assay (ELISA) for aflatoxins is shown in Figure 1. The specificity of the antibody is generally documented with data describing antibody production and characterization for a particular immunogen.

Immunoassays other than RIA and EIA used in Food Analysis

With the availability of different antibodies against various food components and contaminants, many types of immunoassays have been developed. Before modern immunoassays such as RIA and EIA were developed, several classic methods, most of which are based on the precipitation of the immunocomplex, were used widely for the analysis of high-molecular-weight

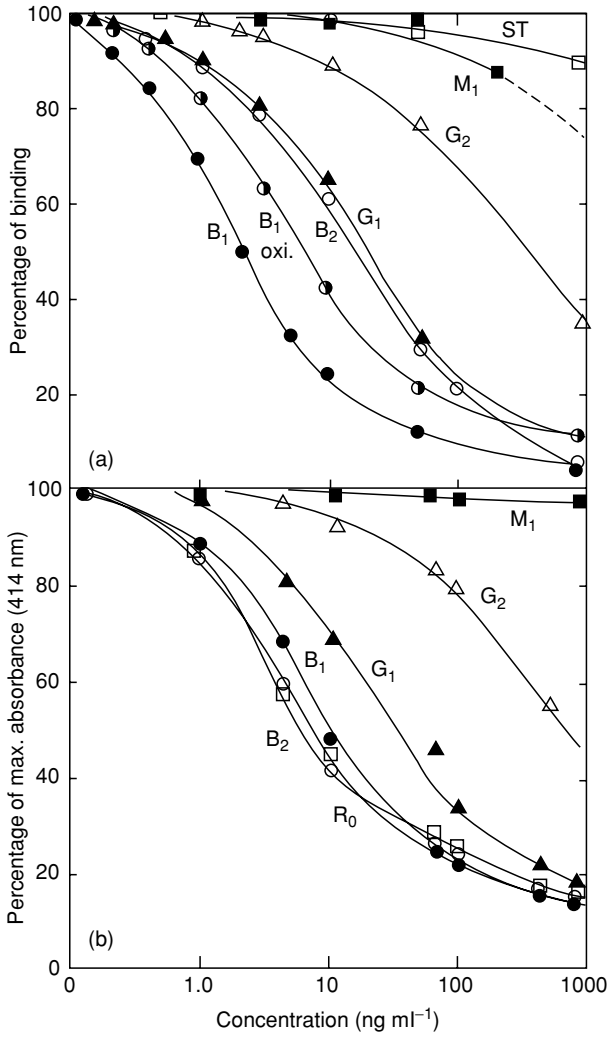

Figure 1 Cross-reactivity of polyclonal antibody against aflatoxin B_1 with various aflatoxins as determined by (a) radioimmunoassay and (b) enzyme-linked immunosorbent assay. (●), Aflatoxin B_1; (◑), aflatoxin B_1-carboxymethyl-oxime; (○), aflatoxin B_2; (▲), aflatoxin G_1; (△), aflatoxin G_2; (■), aflatoxin M_1; (□), sterigmamatocystin.

immunogens. Considerable improvement has been made over the years to some of these methods. With the availability of new instruments and enhancement of the particle size, both the precision and sensitivity have been improved in recent years. Thus, they are still frequently used for the analysis of food constituents. Immunochemical methods other than RIA and ELISA in food analysis are briefly described here.

Precipitation

This method is based on the formation of precipitate as a result of antibody–antigen interaction. Several formats, which require several minutes to several days to perform and have a sensitivity range from 0.1 to 10 $\mu g\,ml^{-1}$, have been used.

Precipitation in solution or quantitative precipitin method In general, an aliquot of different concentrations of antigen or antibody is added to a series of tubes containing a constant amount of antibody or antigen. A precipitate is formed after incubation. The concentration of the unreacted antigen or antibody in the supernatant solution or the weight of the precipitate of the antigen–antibody aggregate is determined after centrifugation. Maximum precipitate is only formed at the equivalent zone where none of the reactants is at excess. This approach is very tedious and subject to error, and is not commonly used at present. Instead, turbidimeter and nephelometer are used in most laboratories. Recent developments have led to measuring the light-scattering effect of the particles instrumentally; the term 'light-scattering immunoassay' is used for this type of assay.

Methods based on simple diffusion of reactants and subsequent precipitation of immunocomplex in agar gel These methods, which are based on the formation of precipitin bands or lines in agar gel, were originally developed by Oudin and Ouchterlony. Three formats are generally used: simple gel diffusion, double diffusion, and radial immunodiffusion (RID).

Simple gel diffusion This method involves the use of a small tube (about 0.5×8 cm; Oudin tube) partially filled with agar containing antibody (occasionally antigen). The test solution is placed on top of the agar column. After antigen diffuses into the agar, a precipitation band is formed at the interface as a result of the antigen–antibody interaction.

Double diffusion (double immunodiffusion) This method involves the use of an agar gel plate (sometimes called an Ouchterlony plate) where cups or wells containing test antibody and antigen solutions are placed in such an order that the antigen and antibody solutions diffuse toward each other through the gel. Most commonly, the antibody well is placed in the center, with the antigen wells around it. After incubation, a band is formed between the antibody and test antigen solution. The intensity of the precipitin band is related to the concentration of the reactants. With appropriate arrangement of the wells for the antigen and antibody, the cross-reactivity of the antibody with structurally related antigen could also be determined. If the precipitin band of the antigen is completely connected with the band of the next well, these two antigens are identical. If they are partially connected to form a spur (arc) type of band, then they are partially identical. However, if both bands cross each other, these two antigens are not identical.

Radial immunodiffusion In this assay, antibody is distributed throughout the gel matrix. The test antigen is placed in wells punched into the gel. After incubation, the antigen is diffused into the gel to form an opaque zone where the immunocomplex occurs. The diameter or area of the circular zone is proportional to concentrations of the test solutions.

Combination of other separation methods with immunodiffusion and precipitation reaction One of the most common approaches is immunoelectrodiffusion (IED) or immunoelectrophoresis, which involves the separation of antigen electrophoretically, followed by immunodiffusion. In practice, a trough parallel to the direction of electrophoresis in a gel is cut immediately after the electrophoresis. The trough is then filled with specific antibody. After incubation, precipitin bands/arcs are formed in the gel.

Several variations of this method have been developed. For example, the IED or so-called 'rocket' immunoelectrophoresis method involves the combination of electrophoresis with RID. In this assay, the antibody is distributed in the gel and antigen or analyte is placed in the well at the end of the gel. Electrophoresis is carried out under conditions which do not permit the migration of antibody. After electrophoresis and diffusion, rocket-shaped precipitates are formed. The height of the rocket is generally related to the concentrations of the test solution. Other variations such as counterimmunoelectrophoresis or immunoosmophoresis, or electrosyneresis and zone immunoassay (ZIA), have been reported, but have not been widely used.

Agglutination Assay/Particle Immunoassay

The unique ability of immunoglobulin to agglutinate cells with multiple antigens has been used in an assay system to determine antigen through competition. Different from the precipitation methods, either an antigen or the corresponding antibody is attached to an inert particle via direct binding or covalent interaction, which serves as the label. Whereas both red blood cells (RBCs) and latex particles have been used for this purpose for some years, recent developments have led to the use of other particles. Thus, a general term of 'particle immunoassay' is also used for this type of assay.

Hemagglutination In hemagglutination (or hemagglutination inhibition assay, HIA, or passive hemagglutination assay, PHA), RBCs (generally sheep), which have been coated or coupled with antigen (thus called sensitized RBCs), are incubated with antibody and sample. After incubation, the degree of agglutination is determined. In this assay, antigen in the sample competes with the antigen coated on the RBCs for their interaction with antibody. Another variation of the HIA is reverse passive hemagglutination assay, in which the purified antibody is coupled to the RBCs. Agglutination is observed when antigen or sample extract is added to the system. Although HIA is relatively simple, food extracts often interfere with the assay and a large amount of antiserum is needed. Only semiquantitative information can be obtained.

Latex agglutination Instead of using sheep RBCs, latex particles coated with antigen have been used as the agglutinating agent in the antigen/antibody reaction. Both regular inhibition and reversed type of agglutination methods have been used.

Other particles and new intrumentation methods for particle counting Several other particles, including microcapsule gelatin, silicate, gold, and emulsified fluorocarbon, have also been introduced as labels in recent years. Among these, the microcapsule gelatin is used most often. Instruments to measure the agglutinated particles are now available and are termed 'particle-counting immunoassay'. Both photometric and light-scattering methods have also been used. The availability of sophisticated instruments has not only improved the sensitivity to a level in the subnanograms per milliliter, but has also led to an automated system that can handle large volumes of samples with good precision.

Other Immunochemical Approaches

Several other immunochemical methods that have gained wide applications in recent years are cited below. Immunoadsorption methods are based on temporary immobilization of immunocomplexes on a solid support; immunoaffinity chromatography is a means of removing undesirable antibodies from an immune serum using the corresponding antigens. Conversely, immunoaffinity chromatography is also used for retaining a given antigen from a complex matrix by using the antiantigen antibodies. Usually, the antigens or the antibodies are immobilized on a support and packed in a small column for immunoaffinity chromatography, which is then used in different types of immunoassays and automation. Some rapid-screening test kits are based on this principle and have gained wide applications. In addition to this technique for haptens, immunomagnetic beads have been used in affinity enrichment and isolation of pathogenic bacteria, including *Escherichia coli* O157:H7. The separation is quick and convenient, requires no elaborate equipment, and can be used instead of precultivation as a precursor for other tests to extract the

target organism. Another powerful immunochemical method is the Western blot, which involves the separation of food components electrophoretically, followed by transfer of the separated components on to nitrocellulose membrane and then staining immunochemically with either enzyme-labeled or radiolabeled antibody. Using a similar principle and in combination with microscopy, a rapid antibody-direct epifluorescent filter technique (antibody-DEFT) for screening of *E. coli* O157:H7 in beef at a level of 0.1 CFU g^{-1} was developed recently. The antibody-DEFT involved membrane filtration, fluorescent antibody staining, and epifluorescence microscopy and was accomplished in less than 1 h.

See also: **Immunoassays**: Radioimmunoassay and Enzyme Immunoassay

Further Reading

Chu FS (1997) Recent development and application of immunochemistry in agriculture and food analysis. In: Dewarti-Hariyadi R and Zakaria F (eds) *Proceedings of the International Symposium of Biomolecular Reactions and Industrial Applications of Immunology in Foods and Agriculture,* pp. 21–40. Jakarta: IPB, Bogor and the French Embassy.

Chu FS (2000) Immunoassays for mycotoxins. In: Hui YH, Smith RA and Spaerke DG *et al.* (eds) *Foodborne Disease Handbook*, vol. 3, pp. 683–713. New York: Dekker.

Fung DYC and Pabon R (1998) Rapid methods and automation in microbiology. *Food Testing and Analysis* 6: 20–26.

Kurtz DA, Skerritt JH and Stanker JH (1995) *New Frontiers in Agrochemical Immunoassay.* Arlington, VA: AOAC International.

Langone JJ and Van Vunakis H (1980–1983) Immunochemical techniques, parts B–F. Monoclonal antibodies and general immunoassay methods. In: *Methods in Enzymology*, vols. 73, 74, 84, 92, 93 . New York: Academic Press.

Nakamura RM, Kasahara Y and Rechnitz GA (1992) *Immunochemical Assays and Biosensor Technology for 1990s.* Washington, DC: American Society for Microbiology.

Perkin-Elmer www.perkinelmer.com.

Rittenburg JH (1990) *Development and Application of Immunoassay for Food Analysis.* London: Elsevier Applied Science.

Tijssen P (1985) *Practice and Theory of Enzyme Immunoassays*. Amsterdam: Elsevier Scientific.

Vanderlaan M, Stanker LH, Watkins BE and Roberts DW (1991) *Immunoassays for Trace Chemical Analysis: Monitoring Toxic Chemicals in Humans, Foods and the Environment*. ACS symposium series 451. Washington, DC: American Chemical Society.

Van Vunakis H and Langone JJ (1980) Immunochemical techniques, part A. In: *Methods in Enzymology*, vol 70. New York: Academic Press.

Radioimmunoassay and Enzyme Immunoassay

F S Chu, University of Wisconsin, Madison, WI, USA

Introduction

Immunochemical methods discussed in the previous article generally involve no labeled antigen or labeled antibody. Concentration of the antigen–antibody complex is estimated from the secondary reaction that leads to precipitation or agglutination. These methods are not sensitive, are subject to nonspecific interference, and are primarily used for analysis of high-molecular-weight proteins. However, the development of radioimmunoassay (RIA) of insulin by Yalow and her colleagues in the late 1950s has widened the scope of immunoassays. This method revolutionized the classic approaches by combining the unique properties of specific antibody–antigen interaction and the use of a radioactive labeled marker to monitor complex formation. Thus, RIA provides specificity, sensitivity, and simplicity, and can be used for analysis of both antigen and haptens. With the introduction of enzymes as markers in the assay, the sensitivity is further amplified. Using different labeled markers, a variety of immunoassays, including fluorescence immunoassay (FIA), time-resolved FIA, FIA polarization immunoassay, enzyme immunoassay (EIA), luminescent immunoassay (LIA), metalloimmunoassay (MIA), and viroimmunoassay (VIA) have been developed. Since RIA and EIA are most commonly used in food analysis, only these two and immunochemical biosenors or immunosensors, which have recently been developed, will be discussed in detail.

Radioimmunoassays

Principles

RIA involves the use of a radioactive marker in the assay, which competes with an analyte in the sample for binding to an antibody. For RIA of high-molecular-weight antigen, either the antigen or antibody molecules can be labeled. It is also common to use a radiolabeled second antibody, i.e., antibody against the primary antibody. In contrast, labeled hapten is typically used in RIA for low-molecular-weight substances. For simplicity, RIA is explained in eqns (1) and (2).

$$\text{Antibody} + \text{Antigen}^* \rightleftharpoons \text{Antibody} - \text{Antigen}^* + \text{Antigen}^* \quad (1)$$

$$\text{Antibody} + \text{Antigen}^* + \text{Analyte} \rightleftharpoons \text{Antibody}- \text{Antigen}^* + \text{Antibody} - \text{Analyte} + \text{Antigen}^* \quad (2)$$

At a constant amount of antibody and labeled antigen (denoted by an asterisk), the presence of analyte in the sample (eqn 2) will result in a decrease in radioactivity in the bound form and an increase in radioactivity in the free form because the analyte is also bound with antibody. In practice, antibody is incubated simultaneously with a solution of unknown sample or known standard, and a constant amount of labeled antigen or hapten. After separation of the free and bound forms, the radioactivity in those fractions is determined. The concentration of the substances to be analyzed in the unknown sample is determined by comparing the results to a standard curve, which is established by plotting the ratio of radioactivities in the bound fraction and free fraction versus the logarithm of the concentration of unlabeled standard analyte.

Although RIA is simple and sensitive, it is limited by the need for a marker with high specific radioactivity, instruments for measuring radioisotopes, and licenses for using radioactive materials and disposal of radioactive materials. Consequently, enzyme-linked immunosorbent assay (ELISA) has become more popular. Nevertheless, because the radioactive marker has the same structural features as the compound to be analyzed, RIA provides good accuracy and is an effective method in the initial phase for screening of antibodies.

Preparation of Radioactive Ligands

The specific activity of the radioactive ligand plays an essential role in the sensitivity of RIA in addition to the affinity constant of antibody and antigen interaction. Because of the availability of high-specificity radioactive iodine, immunoglobulin and protein antigen are generally labeled with ^{125}I. Reagents for iodination such as Bolton-Hunter reagent, chloramine T, iodo-gen, lactoperoxidase, and iodo-beads are commercially available. Thus, most laboratories can prepare their own labeled marker. Since the half-life of ^{125}I is short, iodinated markers generally should not be stored very long. For haptens such as mycotoxins or antibiotics, both tritiated and iodinated derivatives are used. Occasionally, ^{14}C-labeled compounds have also been used. Some high-specific-activity compounds are made commercially by a tritium exchange method. Reduction of a hapten with high-specific-activity-tritiated sodium borohydride is commonly used. For hapten containing hydroxyl groups, the secondary hydroxyl group is first oxidized and then reduced with tritiated sodium borohydride. Iodinated hapten markers are prepared by direct iodination or by iodination of hapten derivatives containing a linker such as tyrosine, histamine, tyramine, or others. For example, 3′-[^{125}I]tyramine-aflatoxin B_1-*O*-carboxymethyl oxime (2300 Ci nmol^{-1}) was used in the RIA of both aflatoxins B_1 and M_1 with high sensitivity (10 pg per assay).

Separation of Antigen–Antibody Complex from Free Antigen

To assess the amount of complex formation, effective methods for the separation of free and bound antigen/antibody must be established. Approaches for the separation of high-molecular-weight antigen and immunocomplex generally can also be used for RIA of hapten as well as for EIA; however, the converse is not always true.

Separation of hapten and hapten–antibody complex Methods such as equilibrium dialysis, ammonium sulfate precipitation, precipitation with organic solvent, polyethylene glycol 6000, membrane filtration, dextran-coated charcoal, and albumin-coated charcoal have been used. Except for the coated charcoal methods, all these methods are based on the separation of molecular weight.

Separation using solid-phase matrix In this approach, antibody is coated to a solid matrix such as polystyrene tubes/beads, microtiter plate, or modified nylon tubes or beads noncovalently, or conjugated to various matrices such as acetylbromocellulose, various types of sepharose gel or agarose gel, and controlled-size fine-particle magnetic gels. Separation is achieved by filtration or centrifugation after the reaction.

Separation using a secondary antibody or other reagents In this approach, a second antibody is added to the reaction mixture to separate the free antigen and immunocomplex. Although the formation of precipitate of the primary antibody with second antibody was used in earlier studies, a solid matrix coupled with the second antibody is now more commonly used. Because of their high affinity to immunoglobulin G (IgG), staphylococcal protein A and protein G coupled to a solid-phase matrix are also commonly used for separation in both hapten and macromolecule assays.

Enzyme Immunoassay

General Considerations and Assay Configurations

EIA is a general term for immunoassays involving use of an enzyme as a marker for the detection of

immunocomplex formation. Whereas the general principle of EIA is similar to RIA, there is an amplification system present in this assay, and thus it is more sensitive. Since no radioactive substances are used, the assays avoid the problems encountered in handling radioactivity. Enzyme labeling can be achieved by conjugation of the enzyme to an antigen or antibody via periodate oxidation with subsequent reductive alkylation method, cross-linking using glutaraldehyde or others. Some of the methods used in the conjugation of hapten to proteins can also be used. However, to avoid nonspecific interaction, methods for coupling of protein/hapten to enzyme should be different from the one that had been used for conjugating the hapten to protein for immunization purpose. Although horseradish peroxidase and alkaline phosphatase are the two enzymes most commonly used, others, such as glucose-6-phosphate dehydrogenase, coupled with oxidoreductase and luciferase, glucose oxidase, beta-galactosidase, urease, and others have also been employed.

Depending on whether or not the immunocomplex is separated from the free antigen, EIAs are further divided into two types. One type is a homogeneous system, which is based on the modification of enzyme activity occurring when antibody binds with the enzyme-labeled antigen/hapten in solution. No separation is necessary in this assay. One such type of this system, which is called enzyme mulitiplied immunoassay (EMIT), has been used for the analysis of some antibiotics and hormones in the clinical diagnosis area. Because modification of enzymatic activity is generally not significant, this system is not very sensitive ($\mu g\,ml^{-1}$ to $mg\,ml^{-1}$ range) and has not been widely used in food analysis.

The other is a heterogeneous system involving separation of free and bound antigen–antibody. In this system, either antigen or antibody is bound to the solid matrix noncovalently, or conjugated to it covalently. Unreacted antibody or antigen/hapten is simply removed by washing or centrifugation. The term 'enzyme-linked immunosorbent assay' (ELISA) is used for this type of assay. Solid phases such as microtiter plates, cellulose, nylon beads/tubes, nitrocellulose membrane, polystyrene tubes/balls, and modified magnetic beads have been used. In some cases, staphylococcal protein A or protein G is coated on the solid surface, entrapping the antibody for subsequent analysis. This method is further divided into two major types. One type is competitive ELISA, which can be used for the analysis of both hapten and macromolecule; the other is noncompetitive sandwich-type ELISA, which is only used for divalent and multivalent antigens.

Competitive ELISA

Depending on whether enzyme-labeled antigen or antibody is used or whether antigen or antigen is coated to the solid phase, several types of competitive ELISAs have been developed. Two major types, i.e., direct competitive ELISA (dC-ELISA) and indirect competitive ELISA (inC-ELISA), are used most commonly in food analysis.

Direct competitive ELISA In this assay, specific antibodies are first coated to a solid phase. The sample or standard solution of analyte is generally incubated simultaneously with enzyme–conjugate or incubated separately in two steps. The amount of enzyme bound to the plate is then determined by incubation with a chromogenic substrate solution. The resulting color/fluorescence, which is inversely proportional to the analyte concentration present in the sample, is then measured instrumentally or by visual comparison with the standards. A typical standard curve for dC-ELISA of aflatoxin is shown in Figure 1 of Immunoassay: Principles and general considerations. Excluding the time for sample preparation, the assay itself can generally be completed in 1–2 h. In general, dC-ELISA is approximately 10–100 times more sensitive than RIA when purified standard is used.

Indirect competitive ELISA (or double-antibody ELISA) In the inC-ELISA, an antigen or a hapten-protein (or polypeptide) conjugate (preferably with a carrier protein/polypeptide that is different from the original carrier protein) is coated to the solid phase, e.g., microtiter plate, to facilitate antibody capture. Antibody and analyte are then incubated together in the wells of the plate. The amount of antibody bound to the wells is then determined by reaction with a second antibody–enzyme complex such as goat anti-rabbit IgG–enzyme complex or goat antimouse-IgG (IgM)–enzyme complex and by subsequent reaction with substrate. Thus, analyte in the sample and antigen in the solid phase compete for the same antibody-binding site in the solution. The sensitivity of the in C-ELISA is comparable to or slightly better than the dC-ELISA, with the advantage that less antibody is required. In addition, it is not necessary to prepare an antigen/hapten–enzyme conjugate. The disadvantage of the method is that an additional incubation step is necessary for the assay and, thus, it requires longer analytical time (2–3 h).

Two modifications of the inC-ELISA have been made. One involves the conjugation of antibody, especially monoclonal antibody, to an enzyme; the conjugate is then used directly in the assay. Thus, this modification converts the inC-ELISA format to a direct assay. The second involves premixing

the antibody with a second antibody–enzyme conjugate. Both modifications have shortened assay time without sacrificing sensitivity. Because only small amounts of antibody are needed, inC-ELISA has been used extensively for screening hybridoma cultures for monoclonal antibody production.

Noncompetitive Sandwich Immunoassay (or Two-Site ELISA)

In this assay, the antibody is coated on to the solid phase, most commonly to the wells of a microtiter plate. The analyte is added and incubated for an appropriate time. After washing, enzyme-labeled antibody is added and incubated, followed by reaction with substrate to develop the color. Under optimal antibody and enzyme antibody concentrations, the color intensity is directly proportional to the concentration of analyte. Thus, it is not necessary to prepare an enzyme-labeled antigen in this assay. A number of modifications of this method have been made.

Monoclonal antibody-based sandwich assay In this assay, a monoclonal antibody (type 'a') is coated on the plate. The plate is sequentially incubated with analyte and another monoclonal antibody (type 'b') labeled with enzyme. The antibody–enzyme conjugate must recognize a different epitope on the analyte than does the antibody (type 'a') used to coat the plate. Determination of analyte–antibody–enzyme–complex is then made by incubation with substrate. With this format, one can select appropriate pairs of monoclonal antibodies for special analytical objectives.

Double-antibody sandwich assay In this assay, antibody originating from animal species 'a' (e.g., rabbit) is coated on to the plate. The plate is sequentially incubated with analyte and antibody against the same analyte from animal species 'b' (e.g., mice), followed by antibody against animal species 'b' (goat antimouse) labeled with enzyme, and then with substrate.

Quick Immunoscreening Tests

Based on the ELISA principle described above, a number of quick screening tests have been developed. For example, by adjustment of reagent concentrations, microtiter plate ELISA assay can be completed in less than 20 mins. Other approaches involve immobilizing the antibody on a paper disk or other membrane which is mounted either in a plastic card or in a plastic cup, or on the top of a plastic tube. In the 'dipstick' assay, antibody or antigen is coated on to a stick, which is then dipped in various reagents for subsequent reactions. Substrates leading to the formation of a water-insoluble chromogenic product are used in these assays. Most of these screening tests are very simple and easy to perform and take 10–15 min to complete. They are designed to provide semiquantitative information at certain cut-off concentrations for the substance in which one is interested. The immunoscreening tests have gained wide application for monitoring microbial toxins in foods and food contaminants such as mycotoxins in the field and versatile assay kits are commercially available.

New Developments and Biosensors:

Nonisotopic Labeling Other than with Enzymes

Among many new nonisotopic labeling systems developed in recently years, FIAs in which the immunoreactants are tagged with fluorescent probes are used most often. For example, in the so-called 'hit-and-run' assay for T-2 toxin, a T-2 toxin column was equilibrated with a fluorescein isothiocyanate (FITC)-labeled Fab fragment of IgG (anti-T-2 toxin). Samples containing T-2 toxin were injected into the column. The FITC-Fab that eluted together with the samples containing T-2 toxin was then determined in a standard flowthrough fluorometer. While fluorescein isothiocyanate (FITC) has been used most often for this purpose, a number of other high-quantum-efficiency reagents, such as the phycobiliproteins, which include phycoerythrins, phycocyanins, and allophycocyaninins, have gained wide applications in recent years. These reagents replace the enzyme as the markers in both the heterogeneous and homogeneous immunoassays.

Fluorescence-based substrate has also been used to improve the sensitivity. New instrumentation further led to the development of several unique FIAs. For example, the fluorescence polarization method can detect the antigen/antibody reaction in solution with good sensitivity in the homogeneous system (ng to μg ml^{-1} range). A time-resolved fluoroimmunoassay, which involves the use of europium ion (Eu)-labeled antibodies, has a sensitivity similar to most ELISAs for aflatoxin analysis (with an IC_{50} of 0.2 ng ml^{-1}). In the particle concentration FIA (PCFIA), the antigen/antibody is bound to the solid phase such as submicrometer polystyrene particles and the fluorescence-labeled reagent and analyte bind to the particle surface competitively. The sensitivity of this for both high-molecular-weight proteins and haptens reaches the ng and sub-ng range in each assay. Similar to FIA, both bioluminescent and chemiluminescent reagents have been used as the markers as well as enzyme substrates in different types of immunoassays.

Enhanced Amplifications Leading to more Sensitive and Ultrasensitive Immunoassays

The sensitivity of different immunoassays has also improved considerably by incorporation of various amplification systems in conjunction with EIA and RIA. Whereas biotin (BT) has been used commonly as a secondary label in systems in conjunction with labeled avidin (AV) or streptavidin, amplification with multiple labeling further improved the sensitivity. The power of amplification is well demonstrated in the ultrasensitive immunoassay where several steps of amplifications and solid-phase concentrations are involved. For example, in a noncompetitive IA of hapten, the hapten is first biotinylated and then reacted with antihapten IgG-coated solid phase. After washing, the bound hapten–BT is removed from the solid phase at pH 1.0. The eluted solution is adjusted to neutral pH, then reacted with antihapten-Fab'-enzyme conjugate, followed by incubation with AV (or streptavidin)-coated solid phase. After washing, the hapten-BT-antihapten-Fab'-enzyme-solid phase is reacted with the substrate. Although this involves many steps, the sensitivity of this system for the hapten reached to 10 attomoles (10^{-18} mol) per assay.

Automated Immunoassay and Development of Antigen/Antibody-Based Biosensors

Development of new sensitive immunoassays that require repeated precise handling of reagents as well as the demand for large numbers of samples for analysis have led to the establishment of automated immunoassay systems. Rapid progress on the antigen/antibody-based biosensors for the detection of various food components, toxins, and contaminants has also emerged. Only a few examples are cited here to show the dimension of this type of technology. A biosensor for a homogeneous immunoassay for T-2 toxin, which involves the use of liposomes and complement, was developed. A fiberoptic immunosensor has been demonstrated to be effective for several microbial toxins, including staphyloccocal enterotoxin B (SEB), T-2 toxins, and fumonisins, in foods. In the assay of hapten-related toxins, monoclonal antibodies are covalently bound to an optical fiber and an evanescent wave effect was utilized to excite the fluorescent-tagged toxin near the surface of the monoclonal antibody fiber as the tagged toxin bound to the fiber. In the presence of unlabeled toxin, it competes with the labeled toxin for binding with monoclonal antibody and results in a decrease in signal. Thus, it is also a competitive assay. For SEB, the sandwich system rather than the competitive format was used and an automated immunosensor detected 5 ng of SEB g^{-1} of cream in 10 min. In the automated particle-based immunosensor (API), the antigen or hapten–protein conjugate is coated on to the polymethylmethacrylate beads (98 μm), and this is then pumped into the 1.5-mm diameter glass capillary to serve as the flow cell within the final lens and trapped on a filter. The sample or calibrated standard solutions that had been incubated with antibodies were then allowed to pass through for a short period (120 s), followed immediately by FITC-labeled goat antirabbit antibody (120 s), and then finally washed with buffer to remove excess label. The fluorescence during each step of the reaction was recorded. Thus, this is a kinetic exclusion assay and can easily detect down to a level of 4 ng g^{-1} of sample (4 p.p.b.) for aflatoxin in 8 min. In a miniaturized ELISA, urease acts as an enzyme marker to produce a pH-shift that can be detected by the ion-sensitive field effect transistor. This system consisted of a flowthrough set-up with a pretreated activated fused silica capillary as a reaction cartridge to serve as a solid phase for the reversible inC-ELISA with a detection limit of 10 pg ml^{-1} for T-2 toxin. In the surface plasmon resonance (SPR) immunosensor, antibody is adsorbed on to a thin gold film substrate coupled to a glass prism. The output beam of a planar light-emitting diode is focused through the prism to excite SPR at the surface of the gold film. When a sample containing antigen/hapten is added to a cell on the outside of the gold film, the angular profile of reflected light intensity shifts. This changes the resonance angle and the reflected beam intensity at a selected angle, both of which are proportional to the antigen/hapten concentration. A detection limit of 50 ng ml^{-1} is obtained for the mycotoxin fumonisin with an analysis time under 10 min. A commercial electrochemiluminescence (ECL) sensor can detect 100 and 1000 of *Escherichia coli* O157 and *Salmonella typhimurium*, respectively, in 1 ml of pristine buffer (or 1000–2000 bacteria ml^{-1} of food samples) in less than 1 h.

Application of Immunoassays in Foods

Immunoassays for Normal Food Components and Adulteration in Foods

Whereas immunochemical methods have been established for most normal food components, one of the major applications of immunoassay in foods is to detect adulteration. This includes the detection of nonmeat proteins in meats and meats from different species. Combinations of electrophoresis with various classic methods, which provided only some qualitative information, were used in early investigations. Both RIA and ELISA are more commonly used at present. For example, RIA has been used to differentiate cattle

meat from sheep, donkey, pig, horse, and kangaroo at 5% contamination level. Various ELISA methods were used to detect albumin and immunoglobulins as a means of detecting blood residues in the muscles of various species. Sensitive RIAs and ELISAs for nonmeat proteins, including soybean proteins, cows' milk, avian egg white protein, wheat gluten, and other nonmeat proteins (legume, cotton seed, peanut, sunflower), have also been developed. These assays are specific and could be used to detect the presence of these proteins in meat products. Since heat treatment or acid hydrolysis can result in a change of protein structure, immunoassays for some of these proteins have also been developed by using antibodies against the denatured proteins and soluble proteins. Immunoassays have also been established for some antinutrients such as trypsin inhibitors, lectins, and allergens. The so-called radioallergosorbent inhibition test (RAST) involving the use of ^{125}I-labeled antihuman IgE, has been used for the detection of allergens in foods. Allergens in foods are also detected with double-antibody sandwich ELISA as well as dC-ELISA and inC-ELISA.

The application of immunoassay for quality control in food processing has also emerged. For example, immunoassays of certain forms of casein, lactoserum, α- and β-lactoalbumins, IgG, lactoferrin, and certain enzymes produced by psychrotropic bacteria in milk were used to determine quality of milk and dairy products. A sensitive ELISA method is also available to differentiate animal rennet and other rennet substitutes. Immunoassay of barley proteins and other soluble proteins and enzymes has been used by the beverage industry. Immunoassays of cortisol and corticosterone were used to monitor meat quality. An ELISA was established for insect myosin, which serves as an indicator for the presence of insects in foods.

Immunoassays for Naturally Occurring Contaminants and Toxins in Foods

Foodborne pathogens and toxins Immunochemical methods have been used for more than 30 years for the detection of foodborne pathogens and high-molecular-weight microbial toxins in foods. Different immunochemical methods for foodborne toxins, including staphylococcal enterotoxins, *Clostridium perfringens* type A enterotoxins, *C. botulinum* toxins, *Bacillus* enterotoxins, and *E. coli* toxins, have been developed. Immunoassays of these toxins in foods have moved from classical immunodiffusion and hemagglutination methods to modern RIA, and now, more commonly, to ELISA. Increasingly, monoclonal antibodies are used. The sensitivity of various types of RIA for the detection of staphylococcal enterotoxins is in the range of 1–5 ng ml^{-1} as compared to that of ELISA of 0.1–5 ng ml^{-1}. An EMIT assay using β-amylase with a sensitivity of 5 ng ml^{-1} has been developed for SEB. The ELISA sensitivities for the detection of botulinum toxin and *C. perfringens* type A toxin are in the range of 0.125–5 ng ml^{-1} and 0.5–500 ng ml^{-1}, respectively. Recent developments have led to more sensitive ELISA methods for monitoring important foodborne bacterial pathogens such as *Campylobacter jejuni, E. coli, Listeria, Salmonella*, and *Yersinina enterocolitica*.

During the last few years, many studies focused on the development of immunoassays for *E. coli* O157:H7. For example, a bead-ELISA with a sensitivity of about 200 pg ml^{-1} was developed for specific detection of the VT2 variant VT2e and a monoclonal antibody-based sandwich ELISA for VT1 and VT2. Using an antigen competition format, a solid-phase fluorescence-based immunoassay was developed. In this assay, a soft glass capillary tube served as the solid support, to which heat-killed *E. coli* O157:H7 were adsorbed. Polyclonal anti-*E. coli* O157:H7 antibody, conjugated with biotin, was used; the bound antigen–antibody complex was detected using avidin molecules labeled with a fluorescent cyanine dye Cy5. Anti-*E. coli* O157 antibody coated to the inexpensive macroporous polyester fabric Polymacron was used to capture *E. coli* O157 antigens for subsequent immunoenzymatic detection of the bacteria. In a rapid and inexpensive sandwich enzyme-linked receptor-based immunodot assay method, the purified glycosphingolipid digalactosylceramide (diGalCer) was used as in the solid-phase matrix to entrap for the assay.

Naturally occurring food toxicants Rapid progress on the development and application of immunoassay for low-molecular-weight naturally occurring toxins, contaminants, and agricultural residues in foods has also occurred during the last two decades. Both monoclonal and polyclonal antibodies against low-molecular-weight food toxins, including marine toxins (brevetoxin, palytoxin, okadaic acid, paralytic shellfish toxins, and ciguatoxins) and mycotoxins (aflatoxins, ochratoxin A, fumonisin, sterigmatocystin, trichothecenes, and zearalenone), are available. Various immunoassay formats have been developed. These methods are generally very simple and sensitive. For example, the lower limits of RIA and ELISA for purified mycotoxin standards are in the range of 0.05–0.5 ng and 2.5–5 pg in each assay, respectively. Because of the interference of sample matrix, the detection limits of most of the mycotoxins in foods/feed are generally about 1–5 p.p.b. for RIA and 0.05–0.5 p.p.b. for ELISA with no sample

clean-up treatment. The detection limits for other food contaminants by RIA and ELISA varied considerably, but generally are less sensitive than for mycotoxins.

Since immunoassays for these haptens in foods share some common problems, strategies to overcome these problems are discussed here. In general, the use of affinity purified antibody in combination with the use of good radioactive-labeled or enzyme-conjugate marker, or with other systems such as chemiluminescence, or amplification systems such as biotin-avidin, can improve the assay sensitivity. However, one of the major concerns is still interference by sample matrix. This problem is generally overcome by dilution of the sample to a range which does not affect the assay or by using sample extract in the preparation of the standard curve. Sample clean-up may sometimes be necessary. For example, the sensitivity of the RIA, dC-ELISA and inC-ELISA for mycotoxins in foods has improved after such treatment. The combination of using high-specific-activity radiolabeled compounds and clean-up samples has allowed the sensitivity for the detection of aflatoxin M_1 in tissues and milk to reach the $0.01\,ng\,g^{-1}$ level in both RIA and ELISA.

Other naturally occurring contaminants In addition to mycotoxins and phycotoxins, sensitive immunoassays have also been developed for many naturally occurring toxicants and contaminants, including DNA-adducts of psoralens, caffeine, naringin, limonin, and selected isoflavones. Sensitive ELISA ($ng\,g^{-1}$) for ergot alkaloids can detect about one sclerotium $20\,kg^{-1}$ of grain. Immunoassay for potato glycoalkaloids such as α-solanine, α-chaconine, and solasonine, and tomato glycoalkaloids such as α-tomatine, tomatidine (aglycon), and digitonin are also available. The sensitivity of these assays is in the $1–200\,ng\,g^{-1}$ range. Contaminants such as mercury ($100\,ng\,g^{-1}$ in scallops) and histamine ($10\,\mu g\,g^{-1}$ salmon) are now detectable by immunoassay.

Agricultural chemicals and residues and pesticides This is another area for which immunochemical methods have been established in the last two decades and have gained wide application. Immunoassays are available for almost all the important antibiotic residues that might be present in our foods. For example, β-lactone antibiotics such as ampicillin, cloxacillin, and penicillin G, in the range of $1–10\,ng\,g^{-1}$, could easily be measured in milk. The sensitivity for the detection of other antibiotics by immunoassay falls in a similar range. Likewise, immunoassays for many pesticides are available. The sensitivity for measuring ethyl parathion, diflubenzuron, and dieldrin in milk are around $5\,ng\,g^{-1}$. Less than $1\,ng\,g^{-1}$ of benomyl/carbendazim in fruits and vegetables could be determined by immunoassay. ELISA kits with a sensitivity in the range of $0.1–5\,ng\,ml^{-1}$ are available for alachlor, aldicarb, atrazine, carbaryl, carbendazim, carbofuran, cyanazine, 2,4-D, and metolachlor.

Concluding Remarks

The development of immunoassays for the detection of food components and contaminants has progressed rapidly in the last few years. Antibodies against almost all the important agricultural and food-related compounds are currently available. Sensitive, simple, and specific immunoassays have been established. Modern immunoassays such as RIA and ELISA are gradually replacing the classic immunodiffusion and agglutination methods for food analyses. These methods are very simple, sensitive, and specific. To avoid the use of radioactive compounds, ELISAs are more commonly used than RIA. The development of new labeling reagents and enhanced amplification systems in conjunction with solid-phase technology further improves the sensitivity and precision of different types of immunoassays, including those of precipitation and agglutination. With the availability of immunochemical reagents and sensitive instrumentation, automation of immunoassays has become a reality. Versatile immunoassay kits that can be used in the field are commercially available and are widely used for the detection of toxins and contaminants in foods. A number of immunoassay protocols have been adopted as first action by the Association of Official Analytical Chemists.

Whereas immunoassays have been used extensively for the analysis of contaminants/toxicants/residues in foods, applications of these methods for the analysis of food components and food adulteration are also merging. However, immunoassays for the control of food processing are not common. Future developments in the area of rapid test, automation, and immunosensors will certainly widen the scope of applications to some of these areas. As such methods are developed, more collaborative studies are needed to validate the efficacy of the protocol and assay kits. Since all immunoassays are based on the specific interaction of antibody with antigen, it is imperative for those who run immunoassays to have a clear understanding of the specificity of the antibody used in the assay as well as the purity and authenticity of the standards. Immunochemical methods, including affinity chromatography as a clean-up tool and immunoassay as a monitoring system for high-performance liquid chromatography, thin-layer chromatography, and electrophoresis have been used in

conjunction with other analytical methodologies. The innovative approaches have led to a new dimension in the immunochemical methods for food analysis. Unquestionably, immunoassays will be advanced to another new era in the new millennium and they will be used extensively in food and agriculture in the coming years. It will not be surprising to see immunoassay screening kits will some day reach the consumer level. Such progress will certainly not only provide safer foods to eat and a better environment in which to live, but will also enhance agricultural productivity by using these new technologies to improve crop quality and yield and enhance animal and plant health.

See also: **Biosensors**; **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography

Further Reading

Beier RC and Stanker LH (1996) *Immunoassays for Residue Analysis: Food Safety*. ACS Symposium Series book no. 621. Washington, DC: American Chemical Society.

Chu FS (1991) Development and use of immunoassays in detection of the ecologically important mycotoxins. In: Bhatnagar D, Lillehoj EB and Arora DK (eds) *Handbook of Applied Mycology, Mycotoxins*, vol. V, pp. 87–136. New York: Dekker.

Fukal L and Kas J (1989) The advantages of immunoassay in food analysis. *Trends in Analytical Chemistry* 8: 112–116.

Maggio ET (1980) *Enzyme-immunoassay*. Boca Raton, Florida: CRC Press.

Morgan MRA, Smith CJ and Williams PA (1992) *Food Safety and Quality Assurance Application of Immunoassay Systems*. London: Elsevier Applied Science.

Pestka JJ (1994) Application of immunology to the analysis and toxicity assessment of mycotoxins. *Food and Agriculture Immunology* 6: 219–234.

Rittenburg JH (1990) *Development and Application of Immunoassay for Food Analysis*. London: Elsevier Applied Science.

Schuurs AHWM and van Weemen BK (1977) Enzyme-immunoassay. *Clinica Chimica Acta* 81: 1–40.

Tijssen P (1985) *Practice and Theory of Enzyme Immunoassays*. Amsterdam: Elsevier Scientific.

Vanderlaan M, Stanker LH, Watkins BE and Roberts DW (1991) *Immunoassays for Trace Chemical Analysis: Monitoring Toxic Chemicals in Humans, Foods and the Environment*. ACS symposium series 451. Washington, DC: American Chemical Society.

Yallow RS (1978) Radioimmunoassay: a probe for the fine structure of biologic systems. *Science* 200: 1236–1245.

IMMUNOLOGY OF FOOD

K M Burghardt and F W Scott, Ottawa Health Research Institute, Ottawa, Canada

Introduction

The process of eating brings us into intimate contact with a wide variety of foods derived from plants and animals. These foods contain carbohydrates, proteins, lipids, micronutrient vitamins and minerals, dietary fiber, other nonnutrient chemicals, bacteria, fungi, viruses, environmental contaminants, food additives, and byproducts of processing and cooking. Food is the most important modifiable environmental exposure to which we are subjected. The aim of this chapter is to explore, in a selective manner, food components that have the potential to alter the activity of the immune system. The gut immune system is continously challenged by a wide spectrum of dietary macromolecules, their metabolic products, resident bacteria, and invading pathogens. Luminal antigens and other molecules are thus a source of information about our immediate environment. The main interaction between food and the immune system occurs in the gastrointestinal tract, the largest lymphoid organ in the body. We shall examine the arsenal of immune cells in the gut-associated lymphoid tissue (GALT), the defense mechanisms, and selected examples of immunomodulatory foods and food components that contribute to health or disease.

Dietary Modulation of the Immune System

Human milk provides optimal nutritional, immunological, physiological, and psychological benefit to a newborn child. At this vulnerable time when its immune system is ill equipped to protect itself, the baby relies on antiidiotypic antibodies and lymphocytes it received *in utero* and which it continues to acquire through mother's milk. Breast-fed infants have been reported to experience significantly fewer gastrointestinal, respiratory, urinary, and ear infections. Some studies have shown long-term enhanced

protection against *Haemophilus influenzae* type b infections, otitis media, diarrhea, respiratory tract infections, and wheezing bronchitis. Mother's milk provides an immediate boost for the young immune system because it delivers antibodies, cytokines, hormones, growth factors, and cells that promote the maturation of the intestinal mucosa as well as modulate immune responses. It contains factors that establish a normal, symbiotic microbial gut flora that protects the newborn from pathogenic bacteria. There is a debate whether infants at risk for development of atopy should be breast-fed because small amounts of food allergens have been detected in human breast milk. A prospective study of neonates identified prolonged (> 9 months) exclusive breast-feeding as the most likely risk factor for atopy by 2 years of age. In contrast, another study showed that supplementation with cows' milk prior to breast-feeding resulted in fewer cases of atopy by 18 months. Numerous other studies have examined this question, and it is yet to be established whether exposure to antigens acquired via the mother's milk induces oral tolerance or priming. Several confounding factors such as antigen dose, timing, and frequency of exposure contribute to the variable results obtained.

The type and quantity of essential fatty acids, such as linoleic acid (*n*-6), dictate the structure and function of the cell membrane. Long-chain and saturated fatty acids reduce fluidity, while short-chain and *cis*-unsaturated fatty acids increase membrane fluidity. Changes to membrane fluidity influence cellular permeability as well as reorganize the distribution and activity of membrane-associated enzymes and receptors. Consequently, this alters the presentation of membrane surface antigens, which affects cellular communication, migration, and immune stimulation.

Linoleic acid is a precursor of arachidonic acid (AA). AA serves as the starting material for eicosanoid synthesis. The membranes of lymphocytes and macrophages contain high concentrations of AA, > 25% of all esterified membrane fatty acids. Mitogens stimulate the release of AA from lymphocytes, which will be converted to eicosanoids by macrophages. Other stimuli that release AA from the cell membrane are reactive oxygen intermediates (ROI) and cytokines such as interleukin-1 (IL-1) and tumor necrosis factor (TNF). Eicosanoids are composed of two families of molecules, the prostanoids (prostaglandins and thromboxanes) and the leukotrienes (lipoxins and hydroxy acids). Release of these oxidation products has physiological and immunological consequences. Prostaglandins regulate immune responses such as IL-2 production, T-cell mitogenesis, and antibody production. Leukotrienes cause leukocyte infiltration, stimulate vascular permeability, and induce smooth-muscle contraction.

Linolenic acid, of the *n*-3 series, was reported to inhibit leukotriene B_4 synthesis and decrease inflammation. Linolenic acid is a precursor for eicosapentaenoic acid (EPA), which is structurally similar to AA, but has antiinflammatory effects. In fact, dietary supplementation with *n*-3 fatty acids increases the incorporation of EPA into the membrane with a proportional decrease in the amount of AA. EPA is metabolized to prostacyclins and thromboxanes which have a net antiinflammatory and antiaggregatory effect. Thus, the amount and type of essential fatty acids in the diet can regulate the type of eicosanoid synthesis. Furthermore, polyunsaturated fatty acids (PUFA) modulate immune responses by enhancing antibody production and decreasing cell-mediated responses, as demonstrated by a shift from T helper 1 (Th1) to Th2 responses that is consistent with IL-2 receptor inhibition. Incorporation of EPA, derived from dietary *n*-3, at the expense of AA reduces the level of PKC activation and thus decreases IL-2 receptor production. There is now considerable evidence that atherosclerosis is an inflammatory disease of the vascular system, the prevention of which involves not only the modification of serum lipids but also a decrease in inflammatory reactions with the endothelium. Dietary *n*-3 PUFA and vitamin E have received particular attention as protective agents.

Dietary fat is also important as a delivery vehicle for the lipid-soluble vitamins, D, E, A, and K. These lipophilic vitamins have immunomodulatory functions. Vitamin A, D, and β-carotene regulate leukocyte activity. Vitamin A controls the recruitment of cells to lymphocyte and macrophage lineages. Vitamin E has antioxidant properties that prevent excessive tissue damage by ROI during stress or infection. Since ROI are also involved in activation of proinflammatory gene expression (e.g., nuclear factor kappaB (NFκB)), antioxidant vitamins can inhibit the inflammatory cascade. These vitamins have been shown to scavenge free radicals that initiate lipid peroxidation. Lipid peroxidation in polymorphonuclear leukocytes, macrophages, and mast cells also releases AA.

Garlic has been used since the times of the early Egyptians as a medicinal herb. There are several biologically active ingredients, including protein fraction 4, diallyl disulfide, and *S*-allyl cysteine in garlic. *S*-allyl cysteine can inhibit activation of NFκB in neuroblastoma and melanoma cells. Thus, garlic is reported to inhibit directly tumor cell growth and may potentiate the immune system by stimulating macrophages, NK (nuclear killer cells), and killer cells, lymphokine-activated killer cells, and increasing production of

IL-2, TNF (tumor necrosis factor), and interferon-γ (IFN-γ). This could be considered an example of inducing a Th1 response that benefits the immune system fight against certain tumors. It is important to be aware that individual foods are unlikely by themselves to be 'magic bullets' that boost the immune response against cancer or regulate other complex chronic diseases. It is the aggregate effect of all dietary exposures that is important. There are several other examples of foods and food components that influence the immune response, including glutamine, arginine, nucleotides, probiotics, vitamins (e.g., vitamin D_3), minerals (e.g., zinc, iron, copper), and oligosaccharides (e.g., 1,3 β-glucans), to name a few.

Micronutrient Modulators of the Immune System

Nutrient status may determine the defense capacity of the immune system. Many vital relationships between trace elements such as zinc, iron, selenium, copper, and cellular function have been described. Most often these elements are an integral part of enzymatic reactions as cofactors and molecular scaffolds. Recently, micronutrients have been shown to regulate directly the immune system.

Zinc has catalytic, structural, and regulatory functions that are well described in cellular metabolism. Zinc deficiency also has immunological consequences, such as lymphopenia, poor cell- and antibody-mediated responses, adrenal hypertrophy, and thymic atrophy. When zinc becomes limiting, it is redistributed to vital organs such as the liver and diverted to the innate arm of the immune system. Zinc deficiency induces a chronic increase in glucocorticoid synthesis, a steroid which promotes apoptosis in precursor lymphoid cells. Thus myeloid cells such as macrophages and neutrophils are spared while lymphoid compartments are depleted. Although macrophages are spared from apoptosis, they are still sensitive to zinc availability. Macrophages isolated from zinc-deficient mice have a reduced capacity to phagocytose and generate an oxidative burst, likely due to the dependence of superoxide dismutase and certain phospholipases on zinc as a cofactor. Similar findings were shown in human monocytes. Clearly, zinc status can alter the choice of defense mechanisms of the immune system.

Iron is an essential trace element because it is part of heme compounds, hemoglobin and myoglobin, which are responsible for oxygen transport and storage, as well as being part of nonheme compounds, and acting as a cofactor for certain critical enzymes like mitochondrial aconitase and ribonucleotide reductase. Moreover, iron availability influences immune effector functions. In the event of chronic infection, iron uptake and storage are increased by proinflammatory cytokines to limit the availability of this essential cofactor to invading pathogens. Increased iron uptake and storage by activated macrophages and monocytes also serve to increase IFN-γ and TNF-α activity and thus enhance their cytotoxic effect. These cytokines also induce the production of nitric oxide, and can form iron–nitrosyl complexes and further decrease iron availability. The counter-effect of the diversion of iron traffic is a reduction in heme synthesis, which contributes to a reduced oxygen transport capacity and possibly chronic anemia. A case in point is anemia of chronic disease, whose pathogenesis has been linked to activated macrophages and monocytes.

Dietary selenium is covalently bound to methionine and cysteine and converted to selenide. Selenide provides an essential cofactor for the synthesis of selenoproteins such as glutathione peroxidases and type 1 iodothyronine deiodinase. Selenium has been exploited in rheumatoid arthritis therapy to reduce oxygen-derived free radicals that damage inflamed joints. Furthermore, selenium also has immunomodulatory properties. Selenium-deficient mice show decreased lymphocyte proliferation in response to mitogen. Selenium supplementation has been shown to enhance cell-mediated immune responses in clinical trials with elderly individuals or subjects receiving parenteral nutrition.

A copper deficit is associated with a small thymus, anemia, decreased antibody production, and neutropenia. In more extreme cases, severe copper deficiency compromises the number and function of neutrophils and macrophages. Individuals fed a copper-deficient diet exhibit reduced T-cell proliferation following mitogen stimulation, an effect that is attributed to decreased IL-2 production. Acquired immunity is also impaired in an animal model of marginal copper deficiency. Pregnant rats were placed on a marginally copper-deficient diet halfway through the gestation period. The rat pups showed reduced cell-mediated responses, yet they had normal copper status as determined by common indices: tissue copper, serum copper, and ceruloplasmin activity. This highlights the exquisite sensitivity of the immune system to nutrient deficiencies and the inherent difficulties in defining 'normal' nutrient requirement for optimum immune function.

How Luminal Antigens are Handled by the Gut Immune System

Nonspecific barriers, such as gastric acid, mucus, digestive enzymes, and peristalsis, as well as specific

immunological barriers, including secretory immunoglobulin A (sIgA) and sIgM, limit antigen access to the GALT. It is estimated that 1 out of 10^5 ingested antigens is absorbed intact by the gastrointestinal tract, which is in the order of microgram per liter quantities. Antigens that cross the gut epithelium are surveyed by immune cells and classified as harmless or harmful. The controlled manner by which luminal antigens gain entry to the GALT is depicted in Figure 1. Lymphoid follicles, called Peyer's patches, are overlaid by membranous epithelial cells (M cells) whose morphology is typified by a smooth apical surface, lack of glycocalyx, and deep invaginations of the basolateral membranes, particularly suited for antigen uptake and delivery to the closely associated lymphocytes and macrophages. Other lymphocytes and dendritic cells are scattered throughout the lamina propria in anticipation of antigens delivered through enterocytes. In healthy epithelia, the uptake of these macromolecules is through active vesicular transport. When the junctions between gut epithelial cells become less restrictive, permeability increases and additional lumen antigens can penetrate the lamina propria and activate resident immune cells. The immune response is based on multiple factors.

A dietary antigen that gains access to the GALT has the potential to induce systemic responses, such as active immune suppression (oral tolerance), immune priming, or it can prompt local immune responses such as the production of sIgA. Oral immune tolerance refers to the unresponsiveness of the immune system, specifically the hyporesponsiveness of mature lymphocytes in peripheral lymphoid tissue that stems from prior oral exposure to the antigen. Recent studies suggest $CD4^+$ Th cells are essential for the induction of oral tolerance because $CD8^+$ cytotoxic T lymphocyte knockout mice can still develop oral tolerance. Furthermore, $CD4^+$ T cells can transfer oral tolerance to a naive individual. Immune deviation from Th1 toward Th2, Th3, or regulatory Tr-1 cells has been proposed to favor normal production of IgA with diminished risk of immunopathology. The major mechanisms, likely acting in concert to establish and maintain tolerance, are clonal anergy (the lack of costimulatory molecules), clonal deletion, and suppression by cytokines such as TGF-β. This depends on the nature of the antigen, dose, maturity of the host immune system, nature of the flora in the digestive tract, and poorly understood host factors. Under certain conditions, the immune system is primed for further immune reaction instead of immune tolerance. Moreover, controversy regarding local immune responses to oral soluble food antigens is not fully resolved. Some suggest that IgA production and secretion accompany oral tolerance, while others contend that healthy individuals do not produce measurable quantities of food-IgA antibodies. Indeed, the decision to commit to any of these pathways is the result of a complex set of cellular interactions and the summation of many bidirectional messages (Figure 1).

Particulate or replicating antigens (e.g., bacteria) are primarily taken up by M cells and diverted to

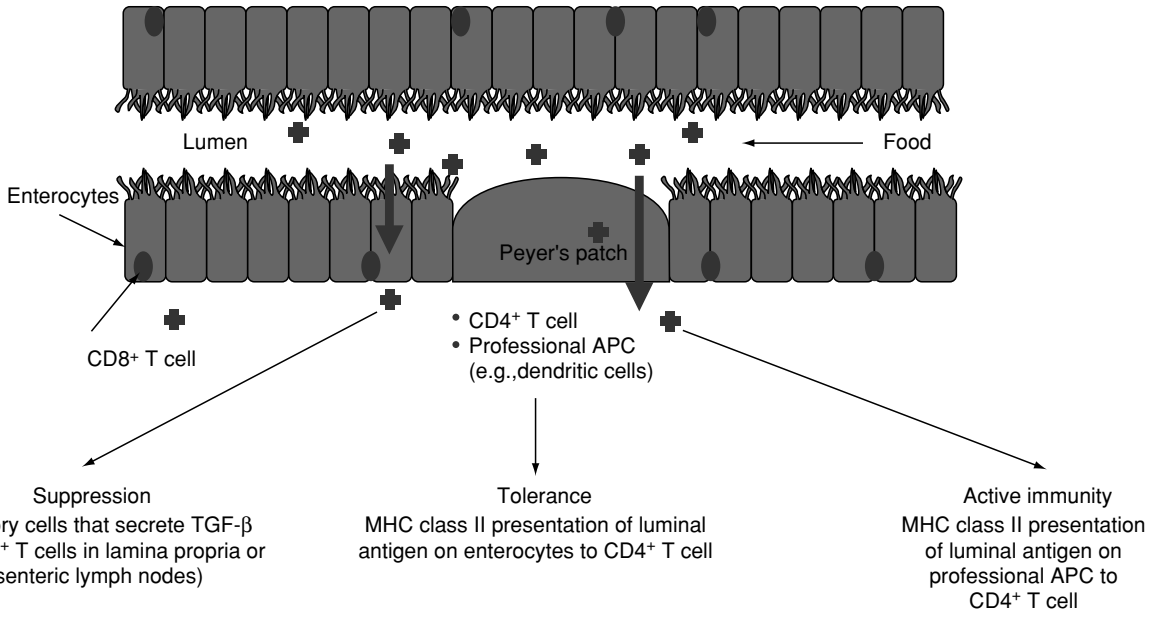

Figure 1 Routes of antigen transport and mechanisms of immune regulation in the gut. APC, aerobic plate count; TGF-β, transforming growth factor-β; MHC, major histocompatibility complex.

the immune cells of Peyer's patches where they will most likely prime the immune system for an assault on these foreign intruders. Most food antigens are digested in the intestinal lumen, which results in various tolerogenic forms of the antigen. The requirement for tolerance induction will depend on the antigenic mixture and its breakdown products. Most soluble antigens induce systemic tolerance.

Oral tolerance has been successfully induced through different regimens and doses of antigens. It has been suggested that a single high dose of antigen ($> 0.5\,mg\,g^{-1}$ body weight) will abrogate antibody, cytokine, and cell-mediated responses in naive individuals as a result of T-cell inactivation. Regulatory T cells are more likely to be activated following multiple low doses ($< 0.1\,mg\,g^{-1}$ body weight). In contrast, very low doses of antigen ($< 0.005\,mg\,g^{-1}$ body weight) were shown to prime systemic and local immune responses.

The age at first encounter with the antigen is an important determinant of the type of immune response that ensues – oral tolerance or immune priming. In general it has been observed that it is easier to tolerize young mice than adult mice. The mechanisms leading to oral tolerance also vary with age. Anergy and/or deletion may be predominant in the immature gut immune system, whereas in adults, regulatory mechanisms have been established and tolerance is more likely the result of active immune suppression. These characteristics of the neonatal immune system may be ascribed to the less restrictive absorption in the neonatal gut, less competent or 'immature' GALT, or differences in the regulation of neonatal lymphocyte responses. A breakdown or failure to establish oral tolerance results in various forms of food hypersensitivity.

Abnormal Immune Reactions to Food

In some individuals, ingestion of certain foods results in a disproportionate immune reaction or allergic reaction. Allergenic foods are commonly resistant to proteolysis, undergo posttranslational glycosylation, and may exhibit enzymatic activity. A distinction should be made between allergy, which is immune-mediated, and intolerance. Intolerance refers to a nonimmunologically mediated reaction stemming from abnormal metabolism (e.g., enzyme deficiency) or impaired excretion of a food constituent or product. Allergy or immediate type 1 hypersensitivity is usually the result of sufficient numbers of IgE molecules binding to some dietary macromolecules, thus initiating the complement cascade and inducing mast cell degranulation. The release of preformed immune mediators stored in the mast cells instigates the immediate onset of symptoms ranging from cutaneous, and gastrointestinal to respiratory, as well as systemic anaphylactic reactions. The severity of the allergic response is associated with the number of immune cells being activated. An allergen that is distributed by the circulatory system has the potential to stimulate a greater number of mast cells, releasing large quantities of immune mediators. Studies of intestinal biopsies have suggested that food allergy in animals and humans is characterized by an increased number of intraepithelial lymphocytes, mucosal mast cells, and eosinophils.

Food allergies and other inflammatory disorders of the intestine are associated with a dysfunction of the epithelial barrier. However, whether the epithelial barrier dysfunction predisposes an individual to allergic disease or is a consequence of a proinflammatory state is controversial. One hypothesis that has sought to explain food allergy suggests that excessive hygiene may be preventing the gut immune system from learning to differentiate between harmful and harmless macromolecules, leading to an inappropriate inflammatory response. Another theory suggests that exposure to high levels of antigen in childhood in individuals with an inherent dysfunction of the gut mucosa may promote allergic diseases. IgE cross-reactivity between food allergens and pathogens is another possible mechanism that may explain the observation that risk of allergy may be proportional to the rate of infection and vaccination. However, the pathogenesis of food hypersensitivity reactions remains unknown.

Diet–Gene Regulation

Food components can influence directly or indirectly the transcription of genes. A case in point is curcumin, which directly influences acute-phase immune responses. Curcumin, a component of turmeric powder from the rhizomes of the *Curcuma longa* L. plant, has antiinflammatory properties that stem from its capacity to inhibit the expression of proinflammatory genes. The transcription of these genes is regulated by NFκB, which in turn is regulated by an endogenous inhibitor called IκBa. Curcumin prevents the degradation of IκBa, which blocks NFκB from translocating to the nucleus to initiate transcription of κB-dependent genes such as TNF-α, IL-6, IL-8, Cox-2 (cyclooxygenase), and nitric oxide synthase. Similarly, some nutrients, such as vitamins A and D, fatty acids, sterols, and zinc, bind to regulatory elements of genes. For example, the metallothionein gene promoter requires the binding of zinc–protein complexes in order to initiate transcription. In addition to the metallothionein gene, 47 other potential zinc-regulated genes have

been identified, which underscores the complexity of nutrient–gene regulation.

In contrast, an indirect effect may be exerted by some nutrient metabolites through secondary metabolism. The effect of dietary fiber is a well-studied example of indirect gene modulation. Intestinal bacteria metabolize undigested fiber and release butyric acid. Although the protective action of butyric acid on colonocytes has not been fully elucidated, *in vitro* studies suggest that butyric acid may modulate gene expression by direct and indirect means. It may acetylate or phosphorylate histones in the nucleus, which would affect the binding of transcription factors to regulatory elements of genes. In particular, elevated butyrate has been shown to inhibit histone deacetylase, which results in the upregulation and secretion of IL-8, a potent chemotactic factor for neutrophils. In addition, butyrate could also act as an oxidative substrate, which would modulate cellular signaling networks via G-proteins. Investigation of these mechanisms may uncover new dietary treatments and better understanding of the progression of colorectal cancer. Moreover, butyrate-producing bacteria in the intestine can vary depending on the diet. Infants fed a casein-based formula predominantly excrete butyric acid and propionic acid, in contrast to breast-fed infants whose stool contains mostly acetic acid. Thus, the presence of butyrate can relay signals to the mucosal immune system, providing vital information about luminal contents and events.

Expression of genes for proteins involved in antigen presentation, namely major histocompatibility complex (MHC) class II molecules and the invariant chain Ii on enterocytes, are also regulated by diet. In mice, prolonged exposure to dams' milk delayed the expression of MHC class II and Ii molecules, suggesting that early oral dosing of antigens could modify the level of antigen presentation. Furthermore, the appearance of MHC class II and Ii molecules was delayed in mice fed an elemental diet composed of synthetic amino acids and fats as compared to littermates fed standard cereal-based rodent diets. Thus, gene expression in intestinal epithelial cells can be manipulated via the presence of specific dietary proteins in the gut lumen.

The complexity of nutrient–gene interactions involves not only interaction with the genome and gene transcription but also affects posttranscriptional events. The regulatory factors involved in posttranscriptional control may be the dietary components themselves, metabolites, or changes in hormones, making it difficult to distinguish the regulatory pathway. There are some examples of nutritional modifications where the regulatory sequences that are targeted are from 3′ and 5′ untranslated regions (UTR) of mRNA that bind specific proteins, a process that in turn regulates polyadenylation, translation, localization, and stability of mRNA. It may be possible in the future to define nutrient–gene interactions in order to tailor dietary requirements to individuals with different disease risk profiles.

Immunonutrition

Lessons learned from studying the immunomodulatory potential of individual nutrients in patients undergoing surgery or recovering from various tissue damage have led to the development of a therapeutic approach called immunonutrition. These special enteral diets are formulated with components that have been shown to enhance or preserve host immune reactions, or suppress exaggerated inflammatory responses, with the aim of decreasing the morbidity and mortality of critically ill patients. Immune-enhancing enteral diets have been assessed in 12 prospective, randomized clinical trials. Although these studies show a strong and statistically significant benefit to most patients, the ideal composition of such a diet still needs to be established. Furthermore, a better definition of the type of patient who would benefit most from immunonutrition is needed before this therapy is widely adopted.

Autoimmunity

Nutrition-based therapies of autoimmune disease are a novel approach that exploits the immunomodulatory nature of certain foods. For example, vitamins have been reported to act as immune regulators in multiple sclerosis. 1,25-Dihydroxy vitamin D_3 supplementation of mice with experimental autoimmune encephalomyelitis (EAE), a murine model of multiple sclerosis, prevented the progression of the disease, while withdrawal caused regression to multiple sclerosis-like symptoms. Nutrition-based therapy using gamma-linolenic acid (18:3*n*-6) supplementation has also reduced the signs and symptoms of disease activity in rheumatoid arthritis. Similarly, dietary modification has been shown significantly to inhibit the development of diabetes in both rodent models of spontaneous autoimmune diabetes, BioBreeding rats and nonobese diabetic (NOD) mice. It was shown that hydrolyzed casein diets had less diabetes-inducing potential than cereal-based diets. Of particular interest, it was also observed that early oral exposure to antigens from known diabetes-promoting diets delayed and inhibited development of diabetes in about one-third of neonates. The identity of the immunomodulatory agent(s) in the diet is currently being investigated.

In other inflammatory disorders of the human intestine, such as celiac disease, ulcerative colitis, and Crohn's disease, a cytokine imbalance may trigger and potentiate the destructive process. In the case of celiac disease, the gut immune system is sensitized to certain wheat peptides. Upon ingestion of wheat, individuals with certain disease-risk haplotypes, particularly the antigen-presenting molecules human leukocyte antigen (HLA)-DQ2 and HLA-DQ8, may display gliadin proteins in a conformation that is recognized by lamina propria T cells. Once activated, these T cells could initiate a destructive inflammatory process that releases the endomysial autoantigen, tissue transglutaminase, and thus potentiates the immune attack and atrophy of intestinal villi. When wheat proteins are eliminated from the diet, the immune attack stops and the symptoms disappear. Of note, is the observation that celiac disease predisposes individuals to other autoimmune disorders such as type 1 diabetes, dermatitis herpetiformis, autoimmune thyroiditis, collagen diseases, autoimmune alopecia, and autoimmune hepatitis. This correlation is stronger in cases of long-term undiagnosed and untreated celiac disease, further suggesting that the GALT is an important site for immune regulation and potential stimulation of autoreactive cells.

New Technologies

New approaches are being explored to assess the immunomodulatory nature of food components. For instance, genomic analysis using microarray chip technology can detect differences in the level of transcription of genes. Alternatively, protein mixtures can be characterized using two-dimensional electrophoresis and mass spectrometry to obtain amino acid sequences of proteins that can be matched with those in existing databases. This approach can be combined with immunoblotting using sera from susceptible individuals or specific monoclonal antibodies. Such proteomic analysis has been used to characterize wheat allergens in baker's asthma. Since there is debate over the risks of toxins, allergens, and other chemical hazards stemming from new food production and processing technologies, these modern tools may be invaluable in food safety assessment. These may also serve as predictive tools for predicting risk of food immunogenicity and allergenicity.

Transgenic plants are being examined as vehicles for vaccine delivery. A case in point is transgenic lettuce plants expressing hepatitis B virus surface antigens that were ingested by human volunteers and resulted in a specific serum-IgG response to the plant-produced viral proteins. Other examples include genetically modified potatoes expressing Norwalk virus capsid protein or autoantigen for autoimmune therapy. Transgenic plants may serve as an inexpensive alternative to fermentation systems for production of antigens. They may also avoid costly purification processes because the subject ingests directly the modified plant. Oral vaccination with bioengineered plants is currently being evaluated for long-term safety, effectiveness, and cost-effectiveness. Transgenic tobacco plants containing the autoantigen GAD-65 (glutamic acid decarboxylase-65) were fed to diabetes-prone NOD mice and this treatment prevented animals from developing diabetes. These data open the possibility of using bioengineered plants as vehicles for the delivery of molecules that may be tolerogenic.

Summary

The nutrition recommendations that are currently suggested by government agencies and health professionals are based on epidemiological studies of discrete populations. The weakness of these studies is that they do not take into account our inherent biochemical and genetic individuality, as was already recognized in 1956 by Williams. The food we eat brings us into direct contact with a complex mixture of nutrients, nonnutrients, toxins, microbes, and molecules that are similar to those that make up our own bodies. Many of these have biological activity that can modify host immune system reactivity. This multitude of food components can alter the nutritional milieu, provide antigens that must be dealt with, act as modifiers of lymphoid cell membranes, alter the balance of immune regulatory (or other) cell compounds such as growth factors, hormones, metabolites, cytokines, chemokines, inflammatory mediators, or interact directly with the genes that are important controllers of the immune response. We are at a stage when new tools for high-throughput screening and analysis of gene and protein expression are providing opportunities for understanding complex interactions among food, genes, and proteins. These technologies and our increasing ability to categorize and comprehend the large amounts of information they generate will permit new insights into the mechanisms by which our exposure to hundreds of chemicals from foods affects our immune systems. This will eventually permit us to understand the range of immune system reactivity and how it is influenced by foods we eat as well as other environmental factors. The advent of genome and proteome exploration may yield information that will permit nutrition-based therapies designed for individuals.

Acknowledgements

The authors thank the following agencies for supporting the research program in Dr. Scott's laboratory: Juvenile Diabetes Research Foundation (JDF), Canadian Inistitutes of Health Research (CIHR), Ontario Research and Development Challenge Fund, Canada Founation for Innovation, Health Canada. Karolina Burghardt was the recipient of a scholarship from the Diabetic Childrens Foundation and the CIHR.

See also: **Anemia (Anaemia)**: Iron-deficiency Anemia; **Copper**: Properties and Determination; Physiology; **Essential Fatty Acids**; **Fats**: Classification; **Food Intolerance**: Types; Food Allergies; Milk Allergy; Lactose Intolerance; Elimination Diets; **Garlic**; **Infants**: Breast- and Bottle-feeding; **Nutrition Education**; **Selenium**: Properties and Determination; **Zinc**: Deficiency

Further Reading

Brandtzaeg P (1998) Development and basic mechanisms of human gut immunity. *Nutrition Reviews* 56: S5–S18.

Cousins RJ (1999) Nutritional regulation of gene expression. *American Journal of Medicine* 106(1A): 20S–23S.

Cunningham-Rundles S (2001) Nutrition and the mucosal immune system. *Current Opinion in Gastroenterology* 17: 171–176.

Hanson LA (1999) Human milk and host defense: immediate and long-term effects. *Acta Paediatrica* 88(430) (Suppl.): 42–46.

Hesketh JE, Vasconcelos MH and Bermano G (1998) Regulatory signals in messenger RNA: determinants of nutrient–gene interaction and metabolic compartmentation. *British Journal of Nutrition* 80: 307–321.

Heyman M (1999) Evaluation of the impact of food technology on the allergenicity of cow's milk proteins. *Proceedings of Nutritional Society* 58: 587–592.

Koletzko B, Aggett PJ, Bindels JG *et al.* (1998) Growth, development and differentiation: a functional food science approach. *British Journal of Nutrition* 80(suppl. 1): S5–S45.

Lamm DL and Riggs DR (2001) Enhanced immunocompetence by garlic: role in bladder cancer and other malignancies. *Journal of Nutrition* 131: 1067S–1070S.

Scott FW (1996) Food induced Type 1 diabetes in the BB rat. *Diabetes Metabolism Reviews* 12: 341–359.

Smith KM, Eaton AD, Finlayson LM and Garside P (2000) Oral tolerance. *American Journal of Respiratory Critical Care Medicine* 162: S175–S178.

Strobel S and Mowat AM (1998) Immune responses to dietary antigens: oral tolerance. *Immunology Today* 19: 173–181.

Voelker R (2000) The hygiene hypothesis. *Journal of the American Medical Association* 283: 1282.

Weindruch R, Kayo T, Lee CK and Prolla TA (2001) Microarray profiling of gene expression in aging and its alteration by caloric restriction in mice. *Journal of Nutrition* 131: 918S–923S.

Williams RJ (1956) Biochemical variations; its significance in biology and medicine. In: *Biochemical Individuality*, pp. 1–7. New York: Wiley.

Ziboh VA (2000) Nutritional modulation of inflammation by polyunsaturated fatty acids/eicosanoids. In: Gershwin ME, German JB and Keen CL (eds) *Nutrition and Immunology: Principles and Practice*, pp. 157–167. Totowa, NJ: Humana Press.

INBORN ERRORS OF METABOLISM

Overview

K de Meer, Vrije Universiteit Medical Center, Amsterdam, The Netherlands

Scope

In clinical medicine and food science, metabolic disorders form a small but important field. Removal of nutrients which can have a critical effect on development of toxic crises or irreversible detrimental effect on organ function is a potentially simple and sometimes life-saving treatment. Simple measures, such as avoidance of prolonged fasting, can be very effective in selected cases. Special diets and costly supplements are necessary in others. Although inborn errors of metabolism are collectively numerous, most disorders are individually rare. Clinical presentation can vary substantially over the spectrum of inborn errors of metabolism and patients can be seen at any age by pediatricians, neurologists, cardiologists, immunologists, hepatologists, dermatologists, and gynecologists. There is no common single test to screen for the whole group of diseases. Collaboration between specialized clinicians and biochemists, metabolic laboratories, and the food industry is required for adequate procedures in the diagnostic work-up and for proper treatment.

Much expert and scientific information on individual genetic variation in relation to human disease is available online (**Table 1**). This article offers a

Table 1 Inborn errors of metabolism: online information sources

OMIM (Johns Hopkins University, Baltimore)

Extensive information on clinical and molecular topics of inborn errors of metabolism and other inherited diseases is available from www.ncbi.nlm.nih.gov in the Online Mendelian Inheritance in Man (OMIM) library. It combines the human gene map with a morbidity map. The OMIM database assigns a code (MIM) number, gives the locus symbol, and provides a complete list of clinical phenotypes for each disease entity. Experts also describe the clinical picture. OMIM provides a morbidity search machine for differential diagnosis on the basis of clinical symptoms and signs. At present, the number of gene loci with an associated clinical disease entity exceed 1700, and more than 2250 human phenotypes have been mapped on the genome. In more than 70% of these disease phenotypes, the molecular basis has been defined. The National Center of Biomedical Information (NCBI) website of the National Library of Medicine at the National Institute of Health also hosts the medical scientific literature (PubMed) and human genome (GenBank) library databases. Together these resources provide an intertwined and up-to-date online source for practitioners and scientists

HUGO Mutation Database Initiative (University of Melbourne)

The website ariel.ucs.unimelb.edu.au/~cotton/mdi.htm provides a list of locus-specific databases according to gene designation, available from university and research institute sources around the world. Connected by the MIM number of each locus, the information from OMIM is only one click away. The drawback is that the visitor should already know the locus lettercode prior to starting up the link. HUGO also lists patient information sites. Information on nonhuman genetic variation is also available

conceptual approach to the field, and describes a selection of diseases in which nutritional intervention is possible.

Genetics and Inborn Errors of Metabolism

Inherited metabolic diseases cause structural changes or metabolic derangements in the body. The term 'inborn error of metabolism' was presented by John Garrow in 1906 in a lecture about alcaptonuria, in which he also hypothesized the theory of one gene, one protein as its cause. Although the concept of inborn error of metabolism implies that a deficiency of a single enzyme is present, the term has also come into use to describe inherited disorders in proteins not involved in metabolism. Mutated genes and their translation products, proteins, can cause derangements during body and tissue formation, and can also be present as abnormal structural proteins and diminished enzyme activity in a wide range of processes. Autosomal recessive inheritance is the most common form of inheritance encountered in inborn errors of metabolism, but not all affected individuals necessarily have symptoms and other inheritance patterns are all but rare.

Mutations in the estimated 30 000 human genes can in principle cause a very large number of diseases. More than 2200 disease loci are now known. In the human genome, mutations in a number of genes are not compatible with survival at the time the embryo develops. In some cases the same mutation in an allele is known to be causally related to more than one disease. Within one disease the same genotype can be associated with more than one phenotypical presentation. Given the magnitude of the human genome, and the multitude of possible mutations and phenotypic disease states, it is not possible to recognize all inborn errors of metabolism by neonatal screening. Apart from dedicated neonatal screening for the more frequent and treatable disorders, simple methods using clinical diagnosis and targeted laboratory investigations are needed to identify the inborn error of metabolism in an individual patient.

From a pathophysiological point of view, the metabolic disorders can be divided into a number of groups that can be used to facilitate diagnostic decision making, including the initial differential diagnosis by the general practitioner and specialist. Specialists in pediatrics, internal medicine, neurology, clinical chemistry, and genetics collaborate in the clinical field of metabolic diseases for establishing the diagnosis on the level of the protein or gene, and to provide treatment and genetic counseling. The focus of this chapter is on inborn errors of metabolism in which nutritional modulation is an important, and in some cases the only, is of treatment. They are summarized in Table 2.

Not all inborn errors of metabolism with nutritional consequences are covered in this article. Familial hyperlipidemias are, also in terms of numbers of patients, a very important group of inborn errors of metabolism, and are dealt with elsewhere in this encyclopedia. Some inborn errors of metabolism in which food components play a role, notably glucose-6-phosphate dehydrogenase deficiency, in which fava beans can trigger symptoms, and porphyrias in which low carbohydrate intake can aggravate symptoms, are not further explored here.

Pathophysiology

Metabolic disorders can be divided into groups, according to the pathophysiological mechanism that is involved.

Table 2 Selected metabolic disorders for which nutritional therapy is used

Amino acid disorders	
Phenylketonuria	Protein restriction, tyrosine and docosa-hexanoic acid supplementation
Tyrosinemia type I and II	Protein restriction, NTBC in type I
Maple syrup urine disease	Valine, leucine, isoleucine restriction, and thiamin supplementation
Urea cycle defects	Protein restriction, arginine and sodium benzoate and phenylacetate supplementation
Homocysteinuria (CBS-deficiency)	Methionine restriction, vitamin B_6, and betaine supplementation
Isovaleric acidemia	Leucine restriction, glycine and carnitine supplementation
Serine synthesis defect (3-PGDHD)	Serine and glycine supplementation
Hartnup disease	High-protein, nicotinamide supplementation
Carbohydrate disorders	
GSD type I and III	High carbohydrate/low fat and cholesterol
Galactosemia	Galactose (and lactose) restriction
Hereditary fructose intolerance	Fructose restriction
Glucose-galactose malabsorption	High fructose/low glucose-galactose
Congenital disorder of glycosylation Ib	Mannose supplementation
Organic acidemias and acidurias	
Propionic acidemia	Protein restriction, avoidance of dehydration, Inhibition of colonic flora with metronidazole
Methyl malonic acidemia	Protein restriction, vitamin B_{12} supplementation
Multiple acyl coenzyme A dehydrogenase deficiency	Fat restriction, riboflavin and carnitine supplementation
Glutaric aciduria type I	Lysine restriction, riboflavin supplementation
Fatty acid oxidation and ketolysis defects	
β-Ketothiolase deficiency	Low protein, avoid fasting
MCAD, LCAD, LCHAD deficiency	Low fat, avoid fasting, ± carnitine
Mitochondrial disorders	
Respiratory-chain disorders	Increased energy need in early life
Pyruvate dehydrogenase deficiency	High fat/low carbohydrate and thiamin
Micronutrient disorders	
Biotinidase deficiency	Biotin supplementation
Vitamin B_6-dependent epilepsy	Vitamin B_6 supplementation
Transcobalamin II deficiency	Vitamin B_{12} supplementation
Immerslund-Grsbeck disease	Vitamin B_{12} supplementation
Hereditary folate malabsorption	Folate supplementation
Methylene tetrahydrofolate reductase deficiency	Folate supplementation
Acrodermatitis enterohepatica	Zinc supplementation
Wilson disease	Zinc supplementation

NTBC, nitro-trifluoro-methylbenzoyl-cyclohexanedione; CBS, cystathionine β-synthase: homozygous patients cannot metabolize homocysteine which is formed after transmethylation of methionine; 3-PGDHD, 3-phosphoglycerate dehydrogenase deficiency: the enzyme block prevents the biosynthesis of serine; GSD, glycogen storage disease; MCAD, medium-chain acyl coenzyme A dehydrogenase; LCAD, long-chain acyl coenzyme A dehydrogenase; LCHAD, long-chain hydroxy acyl coenzyme A dehydrogenase: these mitochondrial enzymes form part of the oxidation of fatty acids.

Group 1 is comprised of disorders characterized by the disturbed synthesis or catabolism of complex molecules. Symptoms are not related to food intake, and are permanent, progressive, and independent of intercurrent infections. Lysosomal storage disease, peroxisomal disorders, α_1-antitrypsin deficiency, and congenital disorders of glycosylation are in this group. With some notable exceptions, nutritional therapy is not effective in these disorders.

Group 2 includes all disorders of intermediate metabolism. Because multiple enzymes are involved in the metabolism between macronutrients and end products, these disorders lead to accumulation of toxic compounds proximal to the metabolic block. There can be progressive or acute intoxication due to these compounds, and intercurrent infections or excessive intake of foods can lead to a metabolic crisis. A symptom-free interval is characteristic. Late onset or an intermittent clinical presentation is typical. Aminoacidemias, most organic acidurias, urea cycle defects, and sugar intolerances belong to this group.

Group 3 consists of inborn errors involved in intermediate metabolism which is directly related with cellular energy transfer. Enzyme defects can affect the supply with macromolecules involved in energy production as well as the transfer of energy by molecules in the matrix and in the inner membrane respiratory chain system within the mitochondria. Glycogen storage diseases, gluconeogenesis defects, defects in pyruvate carboxylase and dehydrogenase,

fatty acid oxidation defects, and mitochondrial respiratory chain disorders are present in this group.

Group 4 is comprised of diseases in which disturbances are present in: (1) membrane transport; (2) intracellular signaling; or (3) critical developmental periods. Defects of transmembrane transporters for carbohydrates, amino acids, and lipids in the intestine are suspected by clinical observation (diarrhea, failure to thrive) and absorption studies. Similarly, defects can be present in tubular reabsorption in the kidney. Other clinical entities are related to disturbed transport of metals (e.g., of copper, such as Menkes and Wilson disease). New insights have been developed that complex compounds, produced in specialized metabolic tissues and necessary for specialized functions in the brain, must undergo a critical step in transport over the cell membrane for proper function. Similarly, with respect to intracellular signaling functions, the assembly or metabolism of complex compounds in the central nervous system and other organs can be deranged. Defects of intracellular membrane transport comprise another group of newly recognized disorders. The transport defects generally result in the dysfunction of one or more affected organs (e.g., diarrhea in carbohydrate malabsorption, liver failure and brain dysfunction by copper storage in Wilson disease, brain dysfunction in creatine transporter deficiency, X-linked nonspecific mental retardation, and abnormal neurotransmitter synthesis) or deficiency syndromes (due to malabsorption).

Recently, insights in developmental biology led to the discovery that certain compounds (e.g., cholesterol) with known functions during extrauterine life are critical for developmental gene description in the embryo (e.g., absence of 7-dehydrocholsterol biosynthesis causes Smith–Lemli–Opitz syndrome, characterized by dysmorphic features and severe mental retardation). Developments in molecular biology and the genome project, human genetics, and clinical chemistry enable discovery of the etiology and pathophysiology of newly recognized disorders with a range of clinical presentations.

Clinical and Laboratory Expertise in Diagnosis

Children with inborn errors of metabolism may present with one or more of a large variety of symptoms and signs. Although most Mendelian phenotypes are expressed early in life, adult presentations are recognized in increasing numbers. The patient history, clinical assessment (for color, odor, hepatomegaly, neurological abnormalities, including hypotonia, and dysmorphic features), and immediate laboratory investigations are necessary for initial judgment in terms of the pathophysiology group and index of suspicion for an inborn error of metabolism. A full medical and genetic family history is important and should include the circumstances of any stillbirth, sudden infant death, unusual death in childhood or early adulthood, and information on consanguinity between the parents.

Blood investigations should include routine hematology and electrolytes (search for anion gap), glucose, creatinine, liver enzymes, bilirubin, ammonia, calcium, phosphate, lactate, pyruvate, ketone bodies (3-hydroxy- and ketobutyrate), fatty acids, uric acid, blood-gas analysis, and prothrombin time. Urinalysis should include acetone, reducing substances, pH, sulfite, electrolytes, and uric acid. Further investigation may comprise lumbar puncture, chest X-ray, and cardiac and central nervous system function studies. Further laboratory and other investigations are guided by the pathophysiological group, abnormalities in the routine investigations, and suspicion of an individual disorder.

The metabolic laboratory is specialized in the investigation of body fluids in the search of abnormal metabolites which are present in many inborn errors of metabolism. Apart from the routine investigations mentioned above, liquid, gas, and thin-layer chromatography are performed on plasma, urine and cerebrospinal fluid, and mass spectrometry, electrophoresis and spectrometric analysis techniques are other methods available for investigation. The use and distribution of these techniques in the investigation of patients suspected of an inborn error of metabolism are depicted in **Figure 1**.

A systematic description of the metabolites and abnormalities frequently observed in inborn errors of metabolism is beyond the scope of this overview. In practice, the inborn errors in which laboratory abnormalities are found are grouped using three different conceptual approaches. Grouping takes place by: (1) the main laboratory abnormality in a body fluid (e.g., aminoacidemia or organic aciduria) as a result of the enzyme defect; (2) the cell organelle where the abnormality is located (e.g., lysosomal storage disease, perixisomal disorder); and (3) the abnormal structural molecule (e.g., defect of purine metabolism, mucopolysacharidosis, or congenital disorder of glycosylation). **Figure 2** shows the distribution of diagnosis of inborn error of metabolism based on records in one laboratory of metabolic diseases where both regional and international patient investigations take place.

The diagnosis system of an inborn error of metabolism is eventually based on the enzyme nomenclature, the MIM number (Mendelian inheritance in

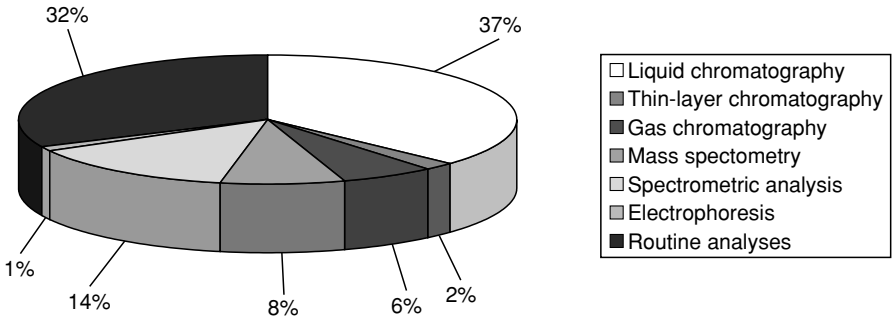

Figure 1 (see color plate 99) Laboratory analyses performed in body fluids of 2150 patients investigated over a period of 5 years in the laboratory for metabolic diseases of the Vrije Universiteit Medical Center in Amsterdam. Investigations were performed in whole blood, plasma, urine, amniotic fluid, and cerebrospinal fluid.

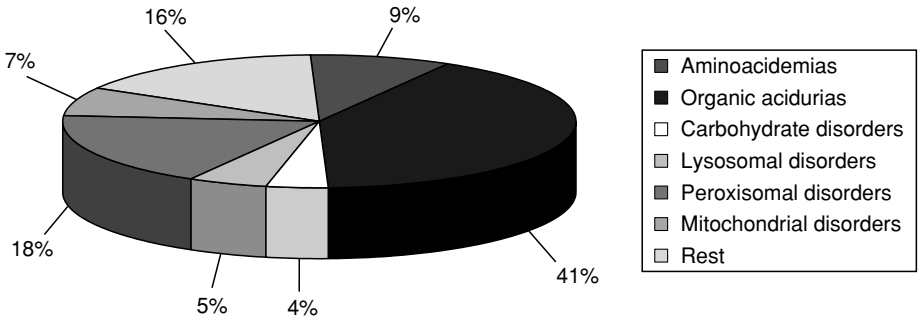

Figure 2 (see color plate 100) Frequency distribution according to diagnostic group in 263 patients newly identified with an inborn error of metabolism over a period of 5 years in the Laboratory for Metabolic Diseases of the Vrije Universiteit Medical Center in Amsterdam. Fatty acid oxidation defects included in organic acidurias.

man, formerly the McKusick number) for each disease entity, and the mutations linked with the clinical entities in the OMIM database (Online Mendelian inheritance in man: Table 1). Enzyme diagnosis, using leukocytes, cultured fibroblasts, or biopsies from chorionic villi or other affected tissue, is increasingly used to confirm the diagnosis. DNA diagnosis is also developed for an increasing number of inborn errors.

In the following paragraphs, some important disorders and clinical problems are presented, to illustrate the interactions between nutrition and inborn errors of metabolism.

Amino Acid Disorders (Group 2 Pathophysiology)

Amino acids are the building blocks of body protein and are mainly derived from the diet. Ingested protein is absorbed almost completely. Amino acids which become available in the body, in excess of the needs for protein synthesis and growth, are eventually oxidized through several pathways. For the nonessential amino acids there are also pathways of biosynthesis. Some of these pathways are shared between two or more amino acids, and others are unique to one amino acid. Inborn errors of metabolism are known for each amino acid. Here we discuss two of the most frequently encountered aminoacidemias.

Hyperphenylalaninemia

Phenylalanine is an essential amino acid. It is normally degraded via the tyrosine pathway. To enter the tyrosine pathway, phenylalanine is converted into tyrosine by the enzyme phenylalanine hydroxylase, which has tetrahydrobiopterin as a cofactor. Deficiency of the enzyme or of its cofactor causes accumulation of phenylalanine in the body fluids and tissues. Hyperphenylalaninemia is present, and detection in plasma is a reliable way of establishing the suspected diagnosis.

Classic phenylketonuria (PKU) is caused by the deficiency of the enzyme phenylalanine hydroxylase. In the first weeks after birth patients have no symptoms, although in the neonatal period vomiting can be an early symptom. Mental retardation is the major abnormality in untreated patients, with an estimated loss of about 50 IQ points by the end of the first year

of life. Untreated patients develop behavioral and neurological abnormalities, including hypertonicity and athetosis, and 25% of patients develop epilepsy. On clinical examination they have an unpleasant smell, probably due to the accumulation of abnormal metabolites of phenylalanine such as phenylacetic acid (which has a musky odor). Microcephaly and growth retardation are common. If protein and phenylalanine restriction is instituted early after birth, the patients remain free of these symptom. Special PKU infant formulas are available, but careful follow-up of the patient's physical growth and development is needed and regular measurement of phenylalanine plasma levels are mandatory in order to prevent under- and overtreatment. Metabolic derangement can take place under catabolic conditions, e.g., during infections.

Proper treatment requires early detection. Mass newborn screening for PKU became effective when Guthrie developed a bacterial inhibition assay which could be done on dried blood collected on paper obtained after a heel puncture during the second half of the first week of life.

About one in 50 patients with hyperphenylalaninemia have a defect in one of the enzymes necessary for the synthesis or recycling of the cofactor tetrahydrobiopterin. If undetected and treated as PKU, these patients show normalized phenylalanine plasma levels but develop severe neurological symptoms. The reason is that tetrahydrobiopterin is a cofactor for other hydroxylases, involved not only in hydroxylation of phenylalanine but also of tyrosine and tryptophan. The latter two are involved in the biosynthesis of the neurotransmitters dopamine and serotinin. For this reason all neonates with hyperphenylalaninemia are required to undergo tests for deficiency of the cofactor. Treatment of these patients requires protein restriction and supplementation with tetrahydrobiopterin (because they also have nonclassical PKU), and supplementation of L-dopa and 5-hydroxytryptophan (because the supplemented cofactor does not cross the blood–brain barrier).

Research has demonstrated that *in vivo* hydroxylation in children with PKU can be so low that patients may suffer from the effects of a diminished availability of the metabolic product of phenylalanine hydroxylase, tyrosine. Tyrosine is an essential amino acid required for physical growth but also a precursor for dopamine biosynthesis. Tyrosine supplementation thus is required in some patients with PKU.

PKU is one of the more common inborn errors of metabolism and this autosomal recessive disease is a major cause of preventable hereditary mental retardation and neurological debilitation in the population. Neonatal screening programs have saved thousands of patients who would otherwise have spent their life in asylums. Nowadays, many adult patients cease the diet and report no psychological or neurological problems. None the less, new problems have emerged in the management of these patients. Pregnancies in mothers with PKU who no longer use the low-phenylalanine diet are complicated by spontaneous abortions, and their living offspring are often mentally retarded and show microcephaly or congenital heart disease. It has become clear that women with PKU who are of child-bearing age should start a low-phenylalanine diet before conception and should be closely monitored for their phenylalanine blood levels before and during pregnancy. Also, deficiency of *n*-6 polyunsaturated fatty acids has been described in patients with PKU (and supplementation has been advocated). The complications during pregnancy and of polyunsaturated fatty acid synthesis demonstrate one of the concepts of group 2 pathophysiology, i.e., that metabolites accumulating proximal to the metabolic block can cause derangements in distant organs and metabolic pathways.

Tyrosinemia Type 1

Tyrosine is likewise an essential amino acid derived from ingested protein. Excess tyrosine is oxidized. Several inborn errors of metabolism in the degradative pathway are known. Deficiency of fumarylacetoacetate hydroxylase causes type 1 tyrosinemia, and has an acute or chronic clinical manifestation. The acute form presents in infancy, and comprises most reported cases. Some symptoms resemble that of PKU with failure to thrive, developmental delay, and vomiting, but hepatic manifestations with organ enlargement, jaundice, and bleeding tendency are common findings as well. Apart from elevated tyrosine, elevated methionine is also found in many patients. Hypoproteinema and low prothrombin are often present. Many patients develop end-stage hepatic failure in childhood and die unless appropriate therapy is started. Patients with the chronic form also present with failure to thrive and development delay, but generally not during the first year of life. Cirrhosis (which can be associated with acute liver failure during catabolic episodes), renal tubular dysfunction, and vitamin D-resistant rickets are often found. Acute episodes of polyneuropathy may complicate the disease. In the long term, a substantial proportion of patients with type 1 tyrosinemia develop hepatic adenoma, and over time hepatocellular carcinoma can develop in these lesions. The wide spectrum of organ pathology, including malignant disease in tyrosinemia, demonstrates the effect of toxic substances that accumulate due to the inability to catabolize tyrosine. Although dietary restriction of tyrosine and

methionine has been advocated over the years, many patients have shown progression of hepatorenal and hematological symptoms and developed malignant liver disease under this treatment regimen. Liver transplantation is a cure for patients with liver cirrhosis or (solitary) tumors, as the donor liver provides the patients with a normally functioning liver, including normal activity of the deficient enzyme. Liver transplantation has a high risk of acute and chronic complications and thus is only an option for patients with severe liver disease. Since 1994 a promising new pharmacological therapy has been successful in reducing morbidity and liver complications: nitrotrifluoro-methylbenzoyl-cyclohexanedione (NTBC), an inhibitor of *p*-hydroxyphenylpyruvate dioxygenase, prevents the formation of (cytotoxic and carcinogenic) metabolites distal of this enzyme and proximal of the inborn error. Although the elevated tyrosine levels do not disappear during NTBC therapy, and tyrosine and methionine restriction remain necessary, this treatment appears remarkably effective. Patient follow-up of this treatment has been less than 10 years at present. There is hope that the long-term life-threatening complications such as liver failure, bleeding disorders and tumor induction are preventable with NTBC.

Table 3 Metabolic causes of hypoglycemia in childhood

Decreased production of glucose
Decreased release of glucose from the liver
Glycogen synthase deficiency
Glucose-6-phosphate deficiency (GSD type Ia or Ib)
Amylo-1,6-glucosidase deficiency (GSD type III)
Galactose-1-phosphate deficiency (galactosemia)
Fructose-1-phosphate aldolase deficiency (hereditary fructose intolerance)
Decreased rate of gluconeogenesis
Pyruvate carboxylase deficiency
Phosphoenolpyruvate carboxylase deficiency
Fructose-1,6-diphosphatase deficiency
Glycerokinase deficiency
Decreased availability of alternative fuels, resulting in increased use or decreased conservation of glucose
Impaired oxidation of fatty acids
Medium-chain acyl-coenzyme A dehydrogenase deficiency
Long-chain acyl-coenzyme A dehydrogenase deficiency
Carnitine acyltransferase I deficiency
Multiple acyl coenzyme A dehydrogenase deficiency
Impaired synthesis or use of ketones
β-Ketothiolase deficiency
Hydroxymethylglutaryl coenzyme A lyase deficiency
Decreased fat stores
Prematurity
Malnutrition

GSD, glycogen storage disease.
Other metabolic disturbances resulting in hypoglycemia are toxic (exposure to ethanol or salicylate), endocrine (hyperinsulinemia, growth hormone deficiency, cortisol deficiency), hyperleucinemia, and a self-limiting ketotic hypoglycemia.

Energy Disturbance (Group 3 Pathophysiology)

Hypoglycemia

A low blood sugar level is potentially dangerous (for central nervous system function) and can be life-threatening. Some organs (the heart, the brain) have a high energy expenditure while using only a limited selection of metabolic fuels. For the brain the balance between energy supply and expenditure is the most critical because glucose and ketone bodies (and, under certain conditions, lactate) are its only energy substrates and the specific energy demand is high. Energy stores in the neural cells are very low and rapidly consumed. Although hypoglycemia can be caused by nonmetabolic disorders it is a hallmark symptom for a large number of inborn errors of metabolism, particularly those of intermediate carbohydrate and triacylglycerol metabolism (Table 3).

Symptoms can be present within hours after birth (and, in the case of low liver stores, even within minutes). The time of presentation after the last meal is indicative of the relevant group of disorders, with glycogen storage disease, gluconeogenesis defects, and fatty acid oxidation defects and impaired availability of ketones successively being the most probable cause when a wider gap is reported between the time of the last meal and the presentation of the hypoglycemia. For the diagnosis of the inborn error, and for the differential diagnosis, including endocrine disorders, blood and urine sampling at the time of the hypoglycemic crisis is essential. In life-threatening situations such as coma, treatment prevails over the complete collection of samples for the diagnostic work-up. Some abnormalities may still be present in blood and urine even hours after intravenous administration of glucose and restoration of the blood glucose level to the normal range. In some disorders symptoms can develop only months or even years after birth. The clinical spectrum is wide, with liver enlargement present in some (e.g., glycogen storage type I and III, hereditary fructose intolerance) but not all disorders and toxic crises present in some (e.g., coma and severe liver function abnormalities in medium-chain acyl-coenzyme A dehydrogenase deficiency) but not all disorders. Hyperlipidemia, granulocyte dysfunction, hyperuricacidemia and gout, hyperlactatemia and osteoporosis, liver adenomas, and pancreatitis are complications in just one of these disorders (glycogen storage disease type I). Cataracts with or without liver dysfunction are seen in galactosemia. This illustrates that virtually any organ

system can be affected in the wide spectrum of disorders encountered as the cause of a decreased blood level of just one metabolite, glucose.

A systematic overview of the spectrum of disease and complications of inborn errors of metabolism causing hypoglycemia cannot be further pursued here. Treatment depends on the cause, and ranges from strict avoidance of certain foods (e.g., galactose and lactose in galactosemia) to continuous drip-feeding (e.g., in glycogen storage disease type I). Avoidance of prolonged fasting is important in a large proportion of disorders. Early diagnosis is essential to prevent brain damage and other irreversible organ failure. Thus recurrent or severe hypoglycemia should bear a high degree of suspicion of a metabolic cause.

Elevated Resting Energy Expenditure

Mitochondrial respiratory chain enzyme defects are associated with myopathy and failure to thrive in early life in mildly affected patients and severe multiorgan involvement, including brain damage and early death in other patients. The failure to thrive in infants with the disease can result from neurological impairment (vomiting, difficulty with swallowing) but also elevated energy expenditure is often found. The pathophysiology of the increased resting energy expenditure is of conceptual interest for this overview on human inborn errors and nutrition. The mitochondrial respiratory chain is comprised of five enzyme complexes which together drive the process of oxidative phosphorylation and maintain the proton gradient through which adenosine triphosphate (ATP) production is maintained. Complex I, III, and IV have proton-pumping capacity, and the efficiency of the protons (with eventually energy transfer to ATP, P) pumped and coupled electron flow through the respiratory chain to oxygen (O) can be expressed as the P/O ratio. In normal mitochondria the ratio is about 1.75 for reduced nicotinamide adenine dinucleotide (NADH)-linked substrates and 2.75 for reduced flavin adenine dinucleolide (FADH)-linked substrates. However, in patients with deficiency of complex I, II, or III, the respiratory chain is substantially less efficient as one proton (ATP) less is produced by the electron's energy transfer between complex I and V. Adaptive changes, including a higher adenosine diphosphate (ADP) concentration (a known stimulus of oxidative phosphorylation activity) and higher resting oxygen consumption are present in many patients. To find out whether, and to what extent, this pathophysiological mechanism is present in an individual patient diagnosed with a respiratory chain defect indirect calorimetry can be useful. Therapy exists in high-energy feedings, and growth monitoring is important. This is particularly important during infancy when energy demands and growth rates are already high and patients depend on food offered by the mother. In later life many patients expend less energy due to muscle fatigue and lower physical activity levels, and the balance between total daily energy expenditure and energy intake is more easily maintained.

Lipids and Essential Fatty Acid Deficiency

In the affluent world and in poor parts of the world, nonhereditary causes of insufficient essential fatty acid status prevail and are worldwide much more frequent than hereditary causes. Disturbance of *n*-6 and *n*-3 polyunsaturated fatty acid status is caused by a number of inborn errors. The most frequent ones are shown in Table 4. Single-enzyme deficiencies in the metabolic pathways of *n*-6 and *n*-3 polyunsaturated fatty acid synthesis and oxidation are extremely rare, probably because essential fatty acids from food and the interconnections between the *n*-6 and *n*-3 desaturase pathways provide alternative provisions and overflow routes for affected metabolites. These rare patients are not reviewed here or included in Table 4.

Table 4 Causes of essential fatty acid deficiency

Mechanism and disease state	*Affected gene(s)*
Hereditary causes	
Intestinal malabsorption	
Abetalipoproteinemia	MTTP
Hypobetalipoproteinemia	Apolipoprotein B
Anderson disease	?
Exocrine pancreas insufficiency	
Cystic fibrosis	CFTR
Shwachman syndrome	Mapped to 7p11-q11
Pearson syndrome	Mitochondrial DNA deletions
Deranged metabolism of polyunsaturated fatty acids	
Decreased synthesis of docosahexanoic acid	
Phenylketonuria	PAH
Zellweger syndrome	PEX
Decreased inhibition of leukotriene B_4	
Sjögren–Larsson syndrome	FALDH
Nonhereditary causes	None
Malnutrition	
Fat-free diet	
Parenteral nutrition without essential fatty acid supplementation	

MTTP, microsomal triglyceride transfer protein; CFTR, cystic fibrosis transmembrane conductance regulator; PAH, phenylalanine hydroxylase; PEX, peroxin, involved in peroxisome biogenesis; FALDH, microsomal fatty aldehyde dehydrogenase.

In patients with disorders in which the assembly of peroxisomes is deficient (mutations in PEX protein complementation groups, e.g., Zellweger syndrome), deficiency of some essential fatty acids has been documented. Enzymes required for the oxidation of very-long-chain fatty acids (including arachidonic acid) and in the synthesis of some polyunsaturated compounds are located in the peroxisome and these are absent or have a very low activity in these patients.

PKU is a single-enzyme disease in phenylalanine hydroxylation in which low essential fatty acid status has been found. It is associated with neuropathy in some patients. Circulating abnormal metabolites affect the synthesis of *n*-6 (arachidonic acid) and *n*-3 (docosahexanoic acid) fatty acids.

Secondary deficiency of essential fatty acids has been found in relatively rare disorders causing lipid malabsorption (transport defects, group 4 pathophysiology) and hereditary exocrine pancreas insufficiency. In the latter group, cystic fibrosis is the most frequently encountered disorder. Sjögren–Larsson syndrome (in which microsomal conversion of medium- and long-chain fatty aldehydes to their carboxylic acids is defective) causes a unique inactivation in leukotriene B_4 inhibition. Patients have elevated leukotriene levels and inflammatory skin changes. Symptoms of essential fatty acid deficiency in other patients are nonspecific and range from dermatitis to severe failure to thrive. Diagnosis is often delayed. Treatment of exocrine pancreas insufficiency by pancreas enzyme supplementation often corrects the essential fatty acid profile. In other disorders skin ointments (on the basis of vegetable oil rich in polyunsaturated fatty acids) are a treatment option. The course of several severe and fatal disorders, such as abetaliproteinema or Zellweger syndrome (in which oral supplementation is given), is not clearly altered by such treatment however.

See also: **Amino Acids**: Properties and Occurrence; Determination; Metabolism; **Diabetes Mellitus**: Problems in Treatment; **Fatty Acids**: *Trans*-fatty Acids: Health Effects; **Hyperlipidemia (Hyperlipidaemia)**; **Nucleic Acids**: Properties and Determination; Physiology

Further Reading

Blau N, Duran M, Blaskovics ME and Gibson KM (eds) (2003) *Physician's Guide to the Laboratory Diagnosis of Metabolic Diseases*. London: Springer.

Fernandes J, Saudubray JM and van den Berghe G (eds) (2000) *Inborn Metabolic Diseases*. Berlin: Springer.

Leonard JV and Morris AAM (2000) Inborn errors of metabolism around time of birth. *Lancet* 356: 583–587.

Scriver CR, Beaudet AL, Sly WS *et al*. (eds) (2001) *The Metabolic and Molecular Bases of Inherited Disease*, 8th edn. New York: McGraw-Hill.

INFANT FOODS

Contents

Milk Formulas

D Hileti-Telfer, The Hospitals for Sick Children, London, UK

Background

Until the early 1900s, there was no alternative to human breast milk for feeding newborn infants. Consequently, over 80% of infants were breast-fed. If a mother was unable to breast-feed her baby, a wet nurse had to be found, or diluted cow's milk had to be given. The 1940s marked a major turning point in the history of infant feeding in the UK, with an increasing proportion of babies being bottle-fed. A number of social, economic, and practical factors contributed to this radical change in infant feeding practices. The major influence was the introduction of the Welfare Food Scheme during World War II. This allowed mothers and children to have free or subsidized National Dried Milk (NDM) at a time when cows' milk was rationed. Inevitably, the use of NDM spread to infants and escalated as social attitudes gradually changed towards women working outside the home.

Unfortunately, NDM and most commercial infant formulas available over the succeeding years were not

adapted to meet the specific nutritional and physiological needs of newborn infants. Excess concentrations of sodium, chloride, phosphorus, and calcium caused many problems, including hypernatremic dehydration, neonatal tetany, and even death in some infants. It is only recently, with advances in food technology, that infant formulas have been developed, providing an acceptable alternative to human milk.

Need for Infant Formulas

Humans are the only mammals who have attempted to feed their young on something other than their own milk. It is now recognized that suitable infant formulas must be made available for infants whose mothers choose not to breast-feed or are unable to establish lactation, or where breast-feeding is contraindicated owing to a medical condition, such as galactosemia and congenital alactasia.

Despite the well-documented advantages of breast-feeding, many women choose not to breast-feed. The UK Office of National Statistics (ONS) 1995 Infant Feeding Survey showed that 39% of mothers gave infant formula from birth, and by 6 weeks, 62% of infants were exclusively bottle-fed. One of the main differences between the ONS surveys of 1990 and 1995 was the decrease in the use of cow's milk at 8 months from 42% in 1990 to 16% in 1995 owing to the continued use of milk formulas.

In the UK, mothers who are carriers of the human immunodeficiency virus (HIV) are discouraged from breast-feeding, as the virus can be passed to an infant via the breast milk. Breast-feeding by HIV mothers is not, however, discouraged in less developed countries such as sub-Saharan Africa, where safe preparation and provision of infant formulas cannot be ensured. Maternal medication can also be found in breast milk. Even low doses of some drugs need to be avoided during lactation, including those that are radioactive, cause allergy or bleeding disorders, are poorly metabolized by the newborn, or inhibit lactation. The relative risks of drugs that pass into the breast milk are reviewed annually in the British National Formulary. If a suitable alternative drug is not available, a mother would need to use an infant formula.

Compositional Guidelines

The composition of all infant formulas in the UK comply with the guidelines of the COMA report 'Artificial Feeds for the Young Infant' and the legal requirements of 'The Infant Formula and Follow-on Formula Regulations 1995'. The 1995 regulations implement the European Commission Directives 91/321/EEC and Council Directive 92/52/EEC, and represent the first mandatory controls on composition and marketing of infant and follow-on formulas. They set upper and lower limits for the concentration of most nutrients in a reconstituted feed (Table 1).

Table 1 Compositional guidelines for infant and follow on formulas

	Infant formulas	*Follow-on formulas*
Major nutrients		
Energy (kcal per 100 ml)	60–75	60–80
Protein (g per 100 kcal)	2.25–3[a]; 1.8–3[b]; 2.25–3[c]	2.25–4.5
Taurine (μmol per 100 kcal)	≥42	
L-Carnitine (μmol per 100 kcal)	≥7.5	
Lipids (g per 100 kcal)	3.3–6.5	3.3–6.5
Linoleic acid (mg per 100 kcal)	300–1200	≥300
α-Linolenic acid (mg per 100 kcal)	≥50	
Linoleic/α-linolenic acid ratio	5–15	
Carbohydrate (g per 100 kcal)	7–14	7–14
Minerals		
Sodium (mg per 100 kcal)	20–60	[d]
Potassium (mg per 100 kcal)	60–145	[d]
Chloride (mg per 100 kcal)	50–125	[d]
Calcium (Ca) (mg per 100 kcal)	≥50	[d]
Phosphorus (P) (mg per 100 kcal)	25–90	[d]
Ca:P ratio	1.2–2	≤2
Magnesium (mg per 100 kcal)	5–15	[d]
Iron (mg per 100 kcal)	0.5–1.5[ab]; 1–2[c]	1–2
Zinc (mg per 100 kcal)	0.5–1.5[ab]; 0.75–2.4[c]	≥0.5[ab] ≥0.75[c]
Copper (μg per 100 kcal)	20–80	[d]
Iodine (μg per 100 kcal)	≥5	≥5
Selenium (μg per 100 kcal)	≤3	
Vitamins		
Retinol (μg per 100 kcal)	60–180	60–180
Cholecalciferol (μg per 100 kcal)	1–2.5	1–3
Thiamin (μg per 100 kcal)	≥40	nr
Riboflavin (μg per 100 kcal)	≥60	nr
Nicotinamide (μg per 100 kcal)	≥800	nr
Pantothenic acid (μg per 100 kcal)	≥300	nr
Vitamin B_6 (μg per 100 kcal)	≥35	nr
Biotin (μg per 100 kcal)	≥1.5	nr
Folic acid (μg per 100 kcal)	≥4	nr
Vitamin B_{12} (μg per 100 kcal)	≥0.1	nr
Vitamin C (μg per 100 kcal)	≥8	≥8
Vitamin K (μg per 100 kcal)	≥4	nr
α-Tocopherol (μg per gram of polyunsaturated fatty acids expressed as linoleic acid)	≥0.5	≥0.5

[a]Unmodified cow's milk.
[b]Modified cow's milk.
[c]Soya milk.
[d]Concentrations at least equal to those found in cow's milk, reduced in the same ratio as the protein concentration of the follow-on formulas to that of cow's milk.
nr, no recommendation.
Data from MAFF (1995) *The Infant Formula and Follow-on Formula Regulations 1995*, No. 77. London: HMSO.

In setting these, the normal physiology of young infants and biological availability of nutrients were considered, but the composition of mature human milk provided the main reference standard. Currently, 10 infant formulas based on cows' milk are available in the UK.

Types of Infant Formulas

Infant formulas should provide the sole source of nutrition during the first few months of life. Although manufacturers can now simulate the nutrient composition of mature breast milk, they are unable to reproduce other aspects such as its enzyme content or immunological properties. Normal infant formulas are manufactured from cows' milk, but this requires substantial alteration to parallel the composition of breast milk (Table 2). These modifications are a reduction in protein and minerals, an increase in carbohydrate, and the addition of vitamins and trace elements. Extra iron and zinc need to be added, owing to their poor absorption from cow's milk. Vitamin activity may be lost during the manufacturing process or on storage, and this must be compensated for.

The other adaptions made to cows' milk are to improve its absorptive characteristics, which is achieved by altering the casein:whey ratio and fat blend.

The protein in cow's milk is predominantly casein, with a whey:casein ratio of 20:80. Two categories of infant formulas are produced: whey-dominant and casein-dominant. The newborn infant more readily absorbs whey protein. Whey-based milks are a mixture of demineralized whey and a small amount of skimmed milk, which achieves a whey:casein ratio similar to breast milk (60:40). The use of whey proteins alters the amino acid pattern, making it closer to human milk. In addition, demineralization of the whey reduces the sodium, potassium, and phosphate contents.

Casein-dominant milks are manufactured using skimmed milk and in a few, full-cream milk as the protein source. Their whey:casein ratio is the same as cows' milk. Although substantially modified during manufacture, the levels of sodium, phosphorus and potassium are usually higher than in whey-dominant milks.

In recent years, studies have indicated that young infants may have an impaired ability to synthesize taurine and carnitine, and a dietary source is therefore required. Carnitine is necessary for the transport of long-chain fatty acids into the mitochondria, where they undergo β-oxidation. Artificial formulas based on cows' milk contain adequate amounts of carnitine, without the need for fortification. Taurine has a major role in the conjugation of bile acids, which is essential for fat absorption, and it may be involved in retinal, cardiac, and central nervous system function. Cows' milk contains a low concentration of taurine. Infant formulas are now fortified with taurine to a level similar to that found in breast milk (3.3–5.2 g per 100 ml).

The carbohydrate present in cows' milk and human milk is lactose. Extra carbohydrate must be added to cows' milk to equate with human milk. This is usually added in the form of lactose, but maltodextrins and occasionally amylose are used in casein-dominant milks. Lactose enhances calcium absorption and helps to inhibit the growth of gut pathogens by fermentation in the colon. Lactose and maltodextrins are efficiently digested and absorbed by the young infant, but the situation with amylose is less clear. Sucrose is rarely used in infant formulas, as it would make the feed very sweet compared with breast milk, and it is also more cariogenic. Glucose and other monosaccharides are avoided, as they would significantly increase the feed osmolality.

The concentrations of fat in mature breast milk and cows' milk are similar. However, the chemical composition, in terms of degree of saturation, chain length, and configuration of fatty acids on the triglyceride molecule is different. These differences influence fat absorption. Short and medium-chain fatty acids are more efficiently absorbed than saturated fatty acids of the same chain length. In both milks, palmitic acid is the most common saturated fatty

Table 2 Comparison between cow's milk and mature human milk

	Cow's milk (per 100 ml)	*Mature human milk (per 100 ml)*
Energy (kcal)	66	69
(kJ)	275	289
Carbohydrate (g)	4.8	7.2
Protein (g)	3.2	1.3
Fat (g)	3.9	4.1
Sodium (mmol)	2.4	0.6
Potassium (mmol)	3.6	1.5
Calcium (mmol)	2.9	0.9
Phosphorus (mmol)	3.0	0.5
Zinc (μmol)	6.1	4.6
Iron (μmol)	0.9	1.3
Vitamin D (μg)	0.03	0.04
Vitamin C (mg)	1.0	4
Vitamin B_1 (mg)	0.04	0.02
Vitamin B_2 (mg)	0.17	0.03
Nicotinic acid (mg)	0.1	0.2
Vitamin E (mg)	0.09	0.34
Vitamin A (μg)	52	58

Data from Holland B, Welch AA, Unwin ID *et al*. (1991) *McCance and Widdowson's The Composition of Foods*, 5th edn. Cambridge: Royal Society of Chemistry.

acid. In human milk, palmitic acid is mainly in the 2 position on the triglyceride molecule, and this is absorbed more easily than palmitic acid in the 1 and 3 positions, as found in cows' milk.

No single fat source can adequately reflect the characteristics of the fat profile in mature breast milk, so a blend is used. Most manufacturers use a mixture of vegetable oils.

The fat source must also provide the essential fatty acids linoleic (C18:2, ω-6) and α-linolenic acid (C18:3, ω-3). Current research suggests that a balance between these two fatty acids is necessary, as excess linoleic acid suppresses the formation of α-linolenic acid derivatives. A ratio of 5:1 of ω-6:ω-3, as occurs in breast milk, is being suggested. However, there is a need for more short- and long-term studies before the optimum ratio and its effects on growth are evaluated. Not all infant formulas provide a ratio of 5:1.

Linoleic and α-linolenic acid are the precursors of the very-long-chain (C20–C22) polyunsaturated fatty acids (LCPUFA), arachidonic and docosahexaenoic acid. LCPUFA are involved in the neural and vascular development of the fetus and neonate and are present in human milk. Recently, there has been much interest in the possible benefits of the addition of LCPUFA to infant formulas. A recent systematic review concluded that ω-3 LCPUFA increases the early visual maturation in preterm infants, but no long-term benefit has been shown for preterm infants receiving formula supplemented with LCPUFA. However, another systematic review prepared by the same author, examining the effect of LCPUFA supplementation in term infants, concluded that there is little evidence that it confers any benefit on visual or cognitive development. Currently, some, but not all, infant formulas are supplemented with LCPUFA.

Nucleotides, a component of the nonprotein nitrogen in human milk, may be important for normal immune function. Nucleotides can also be synthesized endogenously. Nucleotide nitrogen accounts for 0.1–0.15% of the total nitrogen content of human milk, but is present in much smaller concentrations in cows' milk-based formulas. Supplementation of infant formulas with nucleotides seems to be beneficial in clinical trials, although further research is needed before routine nucleotide supplementation of infant formulas can be considered.

Since 1984, a further range of formulas have been manufactured known as 'follow-on' milks. These are based on cows' milk and are intended for use from 6 months of age in the UK, as part of mixed feeding. They are not suitable as a replacement for either breast milk or infant formulas before this age.

'Follow-on' milks have a higher protein and iron content than infant formulas, but the energy density is similar. The main 1995 compositional guidelines for infant and 'follow-on' formulas set by the European Commission Directives 91/321/EEC and 92/52/EEC are shown in Table 1. The iron, calcium, and sodium contents are higher than in infant formulas. Generally, all brands are supplemented with the full range of vitamins and minerals.

Soya infant formulas are manufactured from soya protein isolate and used for infants on cows' milk-free diet. Their carbohydrate source is a glucose polymer, and the protein is derived from soy. They are supplemented with methionine, carnitine, minerals, vitamins, and trace elements, so that the available nutrients are similar to those in human milk (Table 1).

Production

Advances in food technology since the early 1960s have enabled manufacturers to produce infant formulas that are in line with present scientific knowledge about infants' nutritional requirements.

Infant formulas are typically composed of the following: skimmed milk, demineralized whey (in whey-dominant formulas only), a carbohydrate source, fat blend, and vitamins and minerals. Each raw material undergoes stringent quality-control checks on arrival at the factory and throughout all stages of manufacture. These tests are for nutritional composition, purity, microbiological safety, and physical properties. The manufacturing process and the cleaning of the manufacturing plant are partially or fully automated.

Production methods vary, but all aim to blend the raw ingredients together to produce a safe, homogeneous, and stable powder or liquid. A typical example of an infant formula manufacture is as follows:

Pasteurized skimmed milk is bought in as liquid or powder. On arrival, the liquid milk is usually repasteurized by heating to 72 °C for 15 s, which destroys about 90% of all microorganisms present. Throughout the manufacturing process, the product is repasteurized several times. Whey is purchased in a demineralized form or is demineralized on site. The technique of demineralization, which was perfected in the 1960s, was a major advance in infant formula production and is performed by electrodialysis, ultrafiltration, or ion exchange. The repasteurized skimmed milk and demineralized whey are mixed with the carbohydrate source, vitamin premix, minerals, and the fat blend. The minerals required will vary, depending on their concentration in other raw materials. The fat-soluble vitamins are usually added to the fat prior to this process.

The bulk ingredients usually undergo centrifugal clarification, to remove small particle matter, and

then homogenization to ensure that all the particles are uniform in the final product. The liquid infant formula is further subjected to two heat processes: pasteurization and heating to a suitable temperature for spray-drying. Spray-drying produces a powder that will reconstitute easily into solution. Some companies use additional wetting and drying cycles to form granules, which, they claim, are easier to measure and dissolve.

In the UK, powdered infant formulas are packaged in a protected atmosphere using an inert gas to help remove the oxygen, thus reducing any oxidation of the powder. On the packing production line, the cans (already sealed at one end) are inverted and subjected to intense blower and vacuum systems to remove any contamination. A machine drops the measuring scoop into the can, prior to filling with powder. Air is then removed from the pack and replaced with an inert gas mixture, and the container is sealed. Foil sachets are filled in a similar manner. The shelf-life is 2–3 years for sealed tins and 15 months for sachets; when opened, tins need to be used within 4 weeks, and foil sachets within 2 weeks.

Several companies produce ready-to-feed (RTF) infant formulas for community and hospital use. Glass bottles (100 ml) are produced for hospital use and laminated cartons (250 and 1000 ml) for home use. The major ingredients are the same as those used in powdered infant formula production. For the production of RTF bottles, the liquid is produced using water that has been ultrafiltered or deionized to remove all mineral elements. The feed is then repasteurized and decanted into bottles, capped, and sealed. Finally, the bottles undergo terminal sterilization in a rotary autoclave.

For the production of RTF cartons, milk is subjected to UHT sterilization. This quick, high-temperature process (typically 142 °C for 2 s) aims to retain flavor (by reducing caramelization) and allows an extended shelf-life. The cartons are multilaminate, composed of polyethylene, aluminum, and cardboard layers. The base is heat-treated and the packs sterilized internally by hot hydrogen peroxide, before being aseptically filled with milk. The manufacturers recommend that opened 250-ml cartons are used within 24 h, and opened 1000-ml cartons are used within 48 h. The quality checks on these products are extensive, possibly more so than those performed on powders, with up to 5% of the final product destructively tested.

Preparation

In response to the 1974 government recommendation, reconstitution of infant formula in the UK was standardized, requiring only the addition of water. Previously, feed reconstitution was not uniform, which sometimes resulted in over- or underconcentration of feeds. Nowadays, the dilution is always one level scoop of powder added to 30 ml (1 fl. oz.) of water. Scoops from different brands are not interchangeable. However, the relative risk from using the wrong scoop is minimal, unless a scoop for granulated powder is used for measuring nongranulated powder (or vice versa).

Freshly boiled water that has been allowed to cool should be used to reconstitute infant formulas. Domestic softened water, or water repeatedly boiled, is unsuitable, owing to unacceptably high levels of certain minerals. Water filters remove organic compounds, chlorine, and pesticides from water. The resulting water may not be microbiologically safe once the chlorine has been removed, and must therefore be boiled prior to use in feed preparation. Some bottled waters have an unacceptably high sodium content; only those that contain 20 mg of sodium per liter or less, should be used for reconstituting infant formulas. They will also require boiling before use, as they are not sterile.

Immediately before preparing feeds, hands and work surfaces need to be thoroughly washed. Written and pictorial instructions for reconstituting milk powders are given on all infant feed packaging. The first stage is to measure the cooled, boiled water into a graduated milk bottle. The optimum temperature for the water is usually about 50 °C, as this achieves the best dissolution, but specific instructions are given for each formula. The powder is heaped naturally, not compressed, into a scoop and leveled off with the straight edge of a clean knife or spatula. The appropriate number of scoops are added to the water, the cap replaced, and the bottle shaken to achieve thorough mixing of the feed. This can be fed to the infant or kept refrigerated for a maximum of 24 h. Feeds can be warmed by standing the bottle in a jug of hot water. Microwave ovens should not be used for this purpose, as they may result in uneven heating and risk scalding the baby. The temperature of warmed feeds can be checked by testing a few drops on the inside of the wrist.

Sterilization

Bottle-fed babies do not have the same degree of immunological protection as those that are breast-fed, and so strict attention to hygiene practices is vital during feed preparation.

Used bottles and equipment should be thoroughly washed in warm, soapy water, using a bottle brush, if necessary, to remove all traces of milk, then rinsed in

cold water and sterilized. Sterilization is most commonly performed by the addition of sterilizing tablets or liquid to water to produce a hypochlorite sterilizing solution. All bottles and equipment, except metallic objects, need to be completely immersed for the time specified by the manufacturer. To ensure complete sterilization, the bottles should not contain any air bubbles or milk residues. A fresh solution must be made every 24 h.

Alternatively, items can be disinfected by boiling for 20 min in a lidded saucepan kept exclusively for this purpose. Again, the bottles must be fully submerged and free of air bubbles. Teats need only be boiled for 5 min.

Steam sterilization at home is a fairly recent innovation, relying on moist heat as the means of disinfection. Several companies now produce simple electric steam sterilization units. Steam is generated by the addition of a small amount of water on to a heated base-plate; the water gradually turns to steam, sterilizing the bottle surfaces. The machine automatically switches off on completion of its 6–12-min cycle. A temperature in excess of 98 °C is achieved for 3 min. This conforms with the current UK Department of Health standards for equipment used for disinfection: 80 °C for a minimum of 1 min. Microwave ovens are not effective for sterilizing equipment.

Marketing

The Infant Formula and Follow-on Formula Regulations regulates the marketing of milk formulas. Before this was adopted, manufacturers adhered to the Food Manufacturers' Federation code of practice. This was a voluntary code compiled by the infant formula manufacturers in consultation with the Ministry of Agriculture, Fisheries and Food and the Department of Health and was based on the International Code of Marketing of Breast Milk Substitutes. The main recommendations of the code are as follows: advertising of infant formulas should be restricted to publications, which are distributed by health care professionals; all packaging and advertisements should clearly state the superiority of breast-feeding; the provision of free samples of formulas to mothers should be prohibited; and maternity hospitals should not promote the use of infant formula. Furthermore, in 1989, the UK government advised that RTF infant formulas should no longer be subsidized for National Health Service hospitals as the availability of these cheap products undermined breast-feeding practices. These are now purchased at a price decided by national contract. The WHO code is endorsed by most industrialized countries, but it is not universally implemented.

See also: **Fatty Acids**: *Trans*-fatty Acids: Health Effects; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; **Infants**: Breast- and Bottle-feeding; **Milk**: Dietary Importance; **Powdered Milk**: Characteristics of Milk Powders; **Sterilization of Foods**

Further Reading

Carver JD (1994) Dietary nucleotides: cellular immune, intestinal and hepatic system effects. *Journal of Nutrition* 124 (1 supplement): 144S–148S.

Cosgrove M (1998) Perinatal and infant nutrition: Nucleotides. *Nutrition* 14: 748–751.

Department of Health and Social Security (1977) *The Composition of Mature Human Milk. Report on Health and Social Subjects 12*. London: HMSO.

Department of Health and Social Security (1980) *Artificial Foods for the Young Infant. Report on Health and Social Subjects 12*. London: HMSO.

FMF (1983) *Code of Practice for Marketing of Infant Formulae in the United Kingdom and Schedule for a Monitoring Committee*. London: Food Manufacturers' Federation (now Food and Drink Federation).

Foster K, Lader D and Cheesbrough S (1995) *Infant Feeding*. London: Office for National Statistics, The Stationary Office.

Gibson RA, Makrides M, Newmann MA *et al*. (1994) Ratios of linoleic acid to alpha-linolenic acid in formulas for term infants. *Journal of Pediatrics* 125: S48–S55.

Jensen CL, Prager TC, Fraley JK *et al*. (1997) Effect of dietary linoleic/alpha-linolenic acid ratio on growth and visual function of term infants. *Journal of Pediatrics* 131: 200–209.

Joint FAO/WHO Food Standards Programme Codex Alimentarius Commission (1976) *FAO/WHO Recommended International Standards for Foods for Infants and Children*. Rome: Food and Agriculture Organization.

MAFF (1995) *The Infant Formula and Follow-on Formula Regulations 1995, No. 77*. London: HMSO.

MAFF (1997) *The Infant Formula and Follow-on Formula (Amendment) Regulations 1997, No. 451*. London: HMSO.

Makrides M, Newmann MA, Jeffrey B, Lien EL and Gibson RA (2000) A randomized trial of different ratios of linoleic to alpha-linolenic acid in the diet of term infants: effects on visual function and growth. *American Journal of Clinical Nutrition* 71: 129.

National Dairy Council (1995) *Factfile 2: Nutrition of Infants and Pre-school Children*. London: NDC.

Pickering LK, Granoff DM, Erickson JR *et al*. (1998) Modulation of the immune system by human milk and infant formula containing nucleotides. *Pediatrics* 101: 242–249.

Simmer K (2000a) Long-chain polyunsaturated fatty acid supplementation in preterm infants. *Cochrane Database Syst Rev*, -HD-(2):CD000375 2000.

Simmer K (2000b) Long-chain polyunsaturated fatty acid supplementation in infants born at term. *Cochrane Database Syst Rev*, -HD-(2):CD000376 2000.

Taitz LS and Wardley B (1989) *Handbook of Child Nutrition*. Oxford: Oxford University Press.
WHO (1981) *International Code of Marketing of Breastmilk Substitutes*. Geneva: World Health Organization.

Weaning Foods

A M Fehily, Tinuviel Software, Warrington, UK

Introduction

Weaning is the name given to the process by which infants progress from a diet composed only of breast milk or infant formula milk to a family diet containing a wide variety of foods. This is necessary to insure that nutrient intakes continue to be adequate for healthy growth and development throughout childhood.

Infants vary in the age at which they are ready for weaning. Advice from medical experts is that the majority of infants should not be given solid foods before 4 months of age and that a mixed diet should be offered by 6 months of age. Ideal first foods are rice- or maize-based cereals, fruit, vegetables, and potatoes. The range of foods offered can then be gradually increased, such that by the age of 1 year the diet should be mixed and varied. (*See* **Infants**: Weaning.)

This article discusses the types of foods used during weaning, special diet considerations, the nutrient composition of weaning foods, legislation requirements for the composition of commercial weaning foods, and methods of preparation of weaning foods.

Consistency and Taste

At the start of weaning, foods should be of a very smooth, thin consistency such that they can be sucked from a spoon. The food needs to be sufficiently liquid for the infant to transfer it to the back of the mouth for swallowing, but not so runny that it becomes unmanageable. Foods should be of small particle size and have a bland flavor. At this stage, the aim is to introduce the infant to a wide variety of tastes.

As weaning progresses, the infant is introduced to an increasing variety of foods, providing a range of tastes and textures. Once the infant has become used to taking food from a spoon, puréed foods of a thicker consistency can be introduced, and later foods with small soft lumps and soft finger foods. From about 1 year of age, most infants will be enjoying a regular pattern of meals and snacks made up of family foods.

Types of Foods

During weaning the infant is introduced to a wide variety of foods. In this way, the eventual diet is more likely to provide all the nutrients needed for continued healthy growth and development. A few foods are worthy of special mention and these are discussed below.

Cows' Milk

The UK Department of Health recommends that whole cows' milk should not be used as the main milk drink for infants under the age of 1 year. This is because the main drink makes a major contribution to overall nutrient intakes of infants and cows' milk is not a good enough source of iron or zinc.

The Department of Health advises that semi-skimmed milk should not be used as the main milk drink before 2 years of age and that skimmed milk should not be used as the main milk drink before 5 years of age. This is because milk provides a significant proportion of a child's energy intake, and if skimmed or semiskimmed milk is used as the main drink it may be difficult to insure that total daily energy intake is sufficient.

This advice only applies to the use of these milks as the main drink. It does not apply to the small amounts of milk used in composite dishes. Milk and dairy products are a very important part of a young child's diet, providing an excellent source of protein and calcium. The use of milk and milk products as part of the weaning diet helps to accustom the infant to this important group of foods.

Eggs

For infants, eggs should be thoroughly cooked until the white and yolk are solid.

Meat and Fish

Meat, especially poultry, should be thoroughly cooked. For meat and fish, care needs to be taken to insure that no traces of bone remain in the food.

Nuts and Seeds

For infants and young children, nuts and seeds should be finely ground and not whole, to prevent choking.

Where there is a family history of allergy, the Department of Health advises that peanuts and peanut products should be avoided for the first 3 years.

Honey

It has been recommended that honey should not be given to children under 1 year of age. This is because honey is a known source of certain bacterial spores

(*Clostridium botulinum*). In older children and adults, the natural gut flora suppresses the growth of these spores. However, an infant's gut flora may not adequately suppress the growth of these spores.

Salt

Salt (sodium chloride) is the major source of sodium in the diet of older children and adults. However, infants are less efficient than adults in excreting an excess of sodium in the diet. This is because their kidneys are not fully mature. Salt should therefore not be added to baby foods.

In the UK, commercial baby foods have no added salt. However, many foods intended for older children and adults contain added salt and are therefore not suitable for infants.

Sugar

Frequent consumption of sugary foods can increase the risk of tooth decay.

Sugar should only be added to baby foods where this is necessary to provide an acceptable flavor, and the amount added kept to the minimum required. For example, a dessert such as chocolate pudding would taste bitter if no sugar were added.

Many foods contain sugars which are naturally present, e.g., fruits, fruit juices, vegetables, and milk. Both added sugar (table sugar or sucrose) and sugars naturally present in foods such as fruits and fruit juices can increase the risk of tooth decay if care is not taken in use of these foods. Frequent consumption of these foods should be avoided. Good dental hygiene is also very important: tooth cleaning should start as soon as teeth appear.

Additives

Foods intended for older children and adults may contain additives such as artificial colors, preservatives, artificial flavors, and sweeteners. Such foods are unsuitable for infants. These additives are not used in commercial baby foods.

There are some additives legally permitted for use in weaning foods within the European Union (EU). For example, natural flavorings such as vanilla are sometimes used. Certain emulsifiers may be used where needed to insure complete dispersion of oils and fats.

Modified starches are sometimes used to insure that products have the right consistency. These cross-linked amylopectin starches are used in preference to native starches as they perform better in the manufacturing process and do not lose their thickening properties. The amount used is small and the starch is easily digestible by the infant.

Special Diets

The Department of Health advises that where there is a family history of atopy or gluten enteropathy, mothers should be encouraged to breast-feed for 6 months or longer. Weaning before 4 months should particularly be discouraged and the introduction of foods traditionally regarded as allergenic, for example cows' milk and egg, should be delayed until 6 months at the earliest. For peanuts and peanut products, advice is that, where there is a family history of allergy, these should be avoided for the first 3 years.

For the vast majority of infants, where there is no family history of allergy, there is no need to restrict the weaning diet.

It is important to note that the avoidance of any food increases the risk that the infant may not get enough of all the nutrients he or she needs for healthy growth and development. Thus, if it is necessary to exclude any food from an infant's diet, it is essential that expert nutritional advice is obtained from a state-registered dietitian.

For infants being weaned on to a vegetarian diet, care needs to be taken to insure that the quantity and quality of protein are adequate and that energy intake is adequate. Special care also needs to be taken to avoid iron deficiency.

If using commercial baby foods, it is usually easy to identify products suitable for a milk-free, egg-free, gluten-free, or vegetarian diet, as labels of suitable products often carry a statement to this effect.

Nutrient Composition

The nutrient requirements of infants differ from those of older children and adults and are high relative to body size. (*See* **Infants**: Nutritional Requirements.)

A healthy diet recommended for older children and adults, i.e., low in fat and saturates and high in fiber, is not appropriate for infants.

Weaning accustoms the infant to a wide range of tastes and textures. As solid foods form an increasing part of the diet, the diet becomes varied and the nutrient density increases. This helps to insure continuing adequate nutrient intakes for healthy growth and development.

At the start of weaning, the nutrient composition of the foods is not very important as the amounts of foods consumed are small. As weaning progresses and solid foods make a larger contribution to overall nutrient intake, the nutrient composition of weaning foods becomes more important. Nevertheless, milk continues to make a substantial contribution to overall nutrient intake throughout early childhood.

Energy and Protein

An adequate energy intake is essential for healthy growth and development. If energy intake is inadequate, dietary protein may be broken down by the body and used to supply energy instead of being used for growth.

An adequate intake of protein, with a proper balance of essential amino acids, should also be insured during weaning. The protein content of the diet needs to be sufficient to support the fast growth rate of the infant, but not so much as to put a strain on the excretory capacity of the infant's immature kidneys. If the diet is restricted for any reason, care needs to be taken to insure that a variety of foods is offered at each meal, providing a mixture of protein sources.

All foods provide energy, the best sources being milk, rice, pasta, potatoes, bread, meat, poultry, fish, eggs, cheese, beans, and lentils. The main sources of protein in the diet are meat, poultry, fish, milk, cheese, eggs, cereals, beans, peas, and lentils.

Carbohydrate

Carbohydrate includes starch and sugars and these are important sources of energy. Sugars also enhance the palatability of the diet.

Starch is found in foods such as rice, pasta, potatoes, and bread. Apart from table sugar (sucrose), sugars are naturally present in many foods such as fruits, fruit juices, vegetables, and milk. Fruits, fruit juices, and vegetables contain fructose, glucose, and sucrose. Milk contains lactose.

Care needs to be taken in the use of sugary foods, whether the sugar is added or is naturally present in foods such as fruits and fruit juices, to minimize the risk of tooth decay. Frequent consumption of these foods should be avoided and good dental hygiene is essential.

Fruit juices and other sugary drinks should not be allowed to remain in the mouth for long periods. In the UK, the labels of fruit juices and similar drinks intended for infants carry advice supported by the British Dental Association to insure appropriate use of these drinks:

- Always follow the manufacturer's instructions on usage and dilution
- Serve only at meal times and keep drinking times short
- Never leave infants alone with any drink
- Never use drinks on a dummy or as a comforter
- Ideally serve from a spoon, trainer cup, or beaker, but not from a bottle
- Do not give at bedtime or during the night.

Fat

For adults, a healthy diet is one that is low in fat. This is not true for infants. Fat is an important source of energy in an infant's diet. If total fat intake is low, it can be difficult to insure that energy intake is adequate for healthy growth and development.

Fat is also an important source of essential fatty acids and of fat-soluble vitamins (A and D). Essential fatty acids are necessary for growth and for maintaining the skin in a healthy condition. Vitamin A is essential for vision and for healthy skin. Vitamin D helps the body to absorb calcium, which is essential for healthy bones.

Vitamins and Minerals

One of the best ways of insuring that an infant's intake of vitamins and minerals is adequate is to offer a wide variety of foods.

Two nutrients are worthy of special note: iron and vitamin C. Infants have high iron requirements and it can be difficult to insure that intake is sufficient. Iron deficiency results in anemia and may also result in impaired mental and physical development. The effects of deficiency may be long-lasting.

Iron is better absorbed from meat and meat products, particularly red meat (heme iron), than from vegetable sources (nonheme iron). Some components of the diet bind to iron and thereby hinder iron absorption: phytic acid from cereals, legumes, and other vegetables; phosphoproteins in cows' milk and egg yolk; tannin in tea; and polyphenols in spinach and coffee. Vitamin C, on the other hand, improves the absorption of iron from the diet. Vitamin C also enhances the absorption of other minerals from the diet – zinc and copper. Protein also helps to improve iron absorption, by reducing the inhibitory effect of phytic acid.

Foods containing iron, particularly heme iron and foods containing vitamin C, are therefore important components of a weaning diet. Meat, eggs, beans, lentils, peas, bread, and dark green vegetables provide the main sources of iron in the diet. Fruits, vegetables, and fruit juices provide vitamin C. Some commercial weaning foods have added iron and many have added vitamin C.

Fiber

Fiber is an important part of a healthy diet for adults. This is not the case for infants. A high-fiber bran cereal, for example, would not be appropriate for infants. This is because fiber makes the diet more bulky. If the fiber content of the diet is high, the infant may feel full after eating fewer calories. It may therefore be difficult for

him or her to obtain sufficient energy to insure healthy growth and development.

Commercial Weaning Foods

A wide variety of commercial weaning foods is available. Weaning foods in cans and jars are available either as homogeneous purées for young infants (4–6 months of age), or preparations with larger particle size for older infants (from about 7 months of age). A wide range of dried weaning foods is also available, some intended for infants from about 4 months of age and others intended for infants from about 7 months of age.

Types of product available include baby rice cereal, breakfast recipes, savoury recipes, desserts, rusks, and drinks. Breakfast recipes are composed mainly of fruits and cereals. Savoury varieties are composite dishes containing ingredients such as meats, fish, vegetables, and cereals. Dessert varieties consist of fruits, dairy products, and cereals. Available drinks include pure fruit juices with added vitamin C and purified water with a hint of fruit flavor.

Within the EU, there is legislation governing the composition and labeling of processed cereal-based foods and baby foods for infants and young children, i.e., foods intended for children up to 3 years of age. Details of the compositional requirements are presented in Tables 1 and 2. Some companies also have their own additional nutritional guidelines for weaning foods to insure that these foods help mothers to provide a varied and nutritionally balanced diet for their babies.

Certain vitamins and minerals may be added to weaning foods, either to fortify the foods and thus help insure that an infant's nutrients intakes are adequate, or to restore losses which occur during processing. Manufacturers take professional advice when deciding on the levels of vitamins or minerals to add. There is also legislation within the EU setting down the maximum levels which may be present in foods intended for infants and young children where vitamins and minerals are added.

With regard to pesticide residues, EU legislation is more strict for weaning foods than for foods intended for older children and adults. Foods intended for infants and young children may not contain pesticide residues greater than 0.01 mg kg^{-1}, which in practice is the minimum detectable level.

The labels of commercial baby foods provide information on the ingredients, nutrient composition, and suitability for certain special diets. Nutrition information is given per 100 g for energy, protein, fat, and carbohydrate, although many companies also given information on the sugars, saturates, fiber, and sodium contents and provide data per serving as well as per 100 g.

Methods of Preparation

Commercially Made Weaning Foods

Commercially made weaning foods are produced by a variety of methods, some of which are described below.

Cereal-based foods For cereal-based foods, all the ingredients are mixed together with water to produce a liquid paste which is pregelatinized and dried using a roller drier. The dried flakes are then milled to the required particle size and the product then packed.

When using baby rice or other baby cereals at home, one teaspoon of the cereal should be measured into a clean bowl and mixed with one tablespoon of the infant's usual milk. The amount can be increased as necessary.

For baked cereal-based foods, such as rusks, fat and water are added to the dry ingredients and mixed to form a dough. This is converted into a continuous sheet or ribbon which is then cut, shaped and baked in a traveling-band oven, and the product then packed.

When using rusks at home, the back of a spoon should be used to crush a small piece of rusk in a bowl and then mixed with the infant's usual milk, or cooled previously boiled water, to a creamy consistency.

Dried foods For other dried weaning foods, dried ingredients are blended, mixed with water, and cooked on a roller drier to produce flakes. These are reduced to the required particle size and then packed.

When using dried weaning foods at home, the required amount of food should be measured into a clean bowl with a dry spoon. The feeding table presented on the pack should be used as a guide on the quantities to offer. The required amount of warm, previously boiled water should then be added and mixed thoroughly. For some foods intended for older infants, it may be necessary to add hot water to rehydrate the food. If so, then the food must be left to cool before feeding to the infant.

Foods in cans and jars Ready-to-feed weaning foods in cans and jars are produced as follows. Ingredients are prepared and weighed, then mixed with water. Starches, when used, are premixed with water to insure complete dispersion before addition to the main batch. Cooking of the mixture is effected via a steam jacket surrounding the cooking vessel, or via direct steam injection of the mixture. The cooked

Table 1 Composition of processed cereal-based foods for infants and young children: requirements of European Union legislation[a]

Type of product	*Compositional requirements*
Simple cereals which are or have to be reconstituted with milk or other appropriate nutritious liquids	Carbohydrate: the amount of added carbohydrate from sucrose, fructose, or glucose shall not exceed 1.8 g 100 kJ^{-1} (7.5 g 100 kcal^{-1}). The amount of added fructose shall not exceed 0.9 g 100 kJ^{-1} (3.75 g 100 kcal^{-1}) Fat: content shall not exceed 0.8 g 100 kJ^{-1} (3.3 g 100 kcal^{-1}) Sodium: content may not exceed 25 mg 100 kJ^{-1} (100 mg 100 kcal^{-1}) Thiamin: content shall not be less than 25 μg 100 kJ^{-1} (100 μg 100 kcal^{-1}) Vitamin A: if added, vitamin A content shall be a minimum of 14 μg RE 100 kJ^{-1} (60 μg RE 100 kcal^{-1}) and a maximum of 43 μg RE 100 kJ^{-1} (180 μg RE 100 kcal^{-1}) Vitamin D: if added, vitamin D content shall be a minimum of 0.25 μg 100 kJ^{-1} (1 μg 100 kcal^{-1}) and a maximum of 0.75 μg 100 kJ^{-1} (3 μg 100 kcal^{-1})
Cereals with an added high-protein food which are or have to be reconstituted with water or other protein-free liquid	Protein: content shall not exceed 1.3 g 100 kJ^{-1} (5.5 g 100 kcal^{-1}). The added protein shall not be less than 0.48 g 100 kJ^{-1} (2 g 100 kcal^{-1}). There is an additional requirement to insure high quality of the added protein Carbohydrate: the amount of added carbohydrate from sucrose, fructose, or glucose shall not exceed 1.2 g 100 kJ^{-1} (5 g 100 kcal^{-1}). The amount of added fructose shall not exceed 0.6 g 100 kJ^{-1} (2.5 g 100 kcal^{-1}) Fat: content shall not exceed 1.1 g 100 kJ^{-1} (4.5 g 100 kcal^{-1}) Sodium: content may not exceed 25 mg 100 kJ^{-1} (100 mg 100 kcal^{-1}) Calcium: shall not be less than 20 mg 100 kJ^{-1} (80 mg 100 kcal^{-1}) Thiamin: content shall not be less than 25 μg 100 kJ^{-1} (100 μg 100 kcal^{-1}) Vitamin A: content shall be a minimum of 14 μg RE 100 kJ^{-1} (60 μg RE 100 kcal^{-1}) and a maximum of 43 μg RE 100 kJ^{-1} (180 μg RE 100 kcal^{-1} Vitamin D: content shall be a minimum of 0.25 μg 100 kJ^{-1} (1 μg 100 kcal^{-1}) and a maximum of 0.75 μg 100 kJ^{-1} (3 μg 100 kcal^{-1})
Pastas which are to be used after cooking in boiling water or other appropriate liquids	Sodium: content may not exceed 25 mg 100 kJ^{-1} (100 mg 100 kcal^{-1}) Thiamin: content shall not be less than 25 μg 100 kJ^{-1} (100 μg 100 kcal^{-1}) Vitamin A: if added, vitamin A content shall be a minimum of 14 μg RE 100 kJ^{-1} (60 μg RE 100 kcal^{-1}) and a maximum of 43 μg RE 100 kJ^{-1} (180 μg RE 100 kcal^{-1}) Vitamin D: if added, vitamin D content shall be a minimum of 0.25 μg 100 kJ^{-1} (1 μg 100 kcal^{-1}) and a maximum of 0.75 μg 100 kJ^{-1} (3 μg 100 kcal^{-1})
Rusks and biscuits which are to be used either directly or, after pulverization, with the addition of water, milk, or other suitable liquids	Protein: protein content shall not exceed 1.3 g 100 kJ^{-1} (5.5 g 100 kcal^{-1}). For biscuits made with the addition of a high-protein food, and presented as such, the added protein shall not be less than 0.36 g 100 kJ^{-1} (1.5 g 100 kcal^{-1}). There is an additional requirement to insure high quality of the added protein Carbohydrate: the amount of added carbohydrate from sucrose, fructose, or glucose shall not exceed 1.8 g 100 kJ^{-1} (7.5 g 100 kcal^{-1}). The amount of added fructose shall not exceed 0.9 g 100 kJ^{-1} (3.75 g 100 kcal^{-1}) Fat: content shall not exceed 0.8 g 100 kJ^{-1} (3.3 g 100 kcal^{-1}) Sodium: content may not exceed 25 mg 100 kJ^{-1} (100 mg 100 kcal^{-1}) Calcium: for products manufactured with the addition of milk, and presented as such, the amount of calcium shall not be less than 12 mg 100 kJ^{-1} (50 mg 100 kcal^{-1}) Thiamin: content shall not be less than 25 μg 100 kJ^{-1} (100 μg 100 kcal^{-1}) Vitamin A: if added, vitamin A content shall be a minimum of 14 μg RE 100 kJ^{-1} (60 μg RE 100 kcal^{-1}) and a maximum of 43 μg RE 100 kJ^{-1} (180 μg RE 100 kcal^{-1}) Vitamin D: if added, vitamin D content shall be a minimum of 0.25 μg 100 kJ^{-1} (1 μg 100 kcal^{-1}) and a maximum of 0.75 μg 100 kJ^{-1} (3 μg 100 kcal^{-1})

[a]Commission Directive 96/5/EC, as amended.

The amount of cereal and/or starchy root shall not be less than 25% of the final mixture on a dry weight-for-weight basis. RE, retinol equivalents.

food is then reduced to the appropriate particle size, and hot-filled into cans or jars which are then sealed. The cans or jars are then heat-treated in pressure cookers, using a temperature–time–pressure regime, depending upon the type of product being processed. Finally the cans or jars are cooled and then labeled.

Ready-to-feed weaning foods in cans or jars do not need further cooking at home, and can be safely fed to infants without heating them up. If it is desired to heat the food, the required amount should be removed from the can or jar and placed in a clean cup or small bowl. This should be stood in a saucepan or bowl of hot water and covered. The food becomes lukewarm and ready to serve in a few minutes. The food should be stirred and the temperature tested with another spoon before feeding.

Any remaining unwarmed food should be removed from the can or jar, placed in a clean container, covered and stored in the refrigerator for up to 48 h.

Table 2 Composition of baby foods for infants and young children: requirements of European Union legislation[a]

Type of product	*Compositional requirements*
Meat, fish, or other traditional source of protein are the only ingredients mentioned in the name of the product	Protein: The named meat, fish, or other traditional source of protein shall constitute not less than 40% by weight of the product Each named meat, fish, or other traditional source of protein shall constitute not less than 25% by weight of the total named protein sources Protein from the named sources shall not be less than 1.7 g 100 kJ^{-1} (7 g 100 kcal^{-1}) Fat: If meat or cheese are the only ingredients mentioned in the name of the product, or are mentioned first, total fat content should not exceed 1.4 g 100 kJ^{-1} (6 g 100 kcal^{-1}) Sodium: If cheese is the only ingredient mentioned in the name of the product, sodium content shall not be more than 70 mg 100 kJ^{-1} (300 mg 100 kcal^{-1}). For other products, sodium content should be either not more than 48 mg 100 kJ^{-1} (200 mg 100 kcal^{-1}) or not more than 200 mg 100 g^{-1} Vitamin A and vitamin D: may not be added
Meat, fish, or other traditional source of protein is mentioned first in the name of the product	Protein: The named meat, fish, or other traditional source of protein shall constitute not less than 10% by weight of the product Each named meat, fish, or other traditional source of protein shall constitute not less than 25% by weight of the total named protein sources Protein from the named sources shall not be less than 1 g 100 kJ^{-1} (4 g 100 kcal^{-1}) Fat: Content should not exceed 1.1 g 100 kJ^{-1} (4.5 g 100 kcal^{-1}) Sodium: Content should be either not more than 48 mg 100 kJ^{-1} (200 mg 100 kcal^{-1}) or not more than 200 mg 100 g^{-1} Vitamin A and vitamin D: may not be added
Meat, fish, or other traditional source of protein is mentioned, but not first, in the name of the product	Protein: The named meat, fish, or other traditional source of protein shall constitute not less than 8% by weight of the product Each named meat, fish, or other traditional source of protein shall constitute not less than 25% by weight of the total named protein sources Protein from the named sources shall not be less than 0.5 g 100 kJ^{-1} (2.2 g 100 kcal^{-1}) Total protein shall not be less than 0.7 g 100 kJ^{-1} (3 g 100 kcal^{-1}) If cheese is mentioned in the name, protein from dairy sources shall not be less than 0.5 g 100 kJ^{-1} (2.2 g 100 kcal^{-1}); total protein shall not be less than 0.7 g 100 kJ^{-1} (3 g 100 kcal^{-1}) Fat: Content should not exceed 1.1 g 100 kJ^{-1} (4.5 g 100 kcal^{-1}) Sodium: Content should be either not more than 48 mg 100 kJ^{-1} (200 mg 100 kcal^{-1}) or not more than 200 mg 100 g^{-1} Vitamin A and vitamin D: May not be added
Product is designated as a meal, but meat, fish, or other traditional source of protein not mentioned in the product name	Protein: Total protein shall not be less than 0.7 g 100 kJ^{-1} (3 g 100 kcal^{-1}) Fat: Content should not exceed 1.1 g 100 kJ^{-1} (4.5 g 100 kcal^{-1}) Sodium: Content should be either not more than 48 mg 100 kJ^{-1} (200 mg 100 kcal^{-1}) or not more than 200 mg 100 g^{-1} Vitamin A and vitamin D: May not be added
Fruit-only dishes	Carbohydrate: Shall not exceed 20 g 100 g^{-1} Fat: Content should not exceed 1.1 g 100 kJ^{-1} (4.5 g 100 kcal^{-1}) Sodium: Content should be either not more than 48 mg 100 kJ^{-1} (200 mg 100 kcal^{-1}) or not more than 200 mg 100 g^{-1} Vitamin A and vitamin D: May not be added
Desserts and puddings	Protein: If a dairy product is the first or only ingredient mentioned in the product name, protein from dairy sources must be at least 0.5 g 100 kJ^{-1} (2.2 g 100 kcal^{-1}) Carbohydrate: Shall not exceed 25 g 100 g^{-1} Fat: Content should not exceed 1.1 g 100 kJ^{-1} (4.5 g 100 kcal^{-1}) Sodium: Content should be either not more than 48 mg 100 kJ^{-1} (200 mg 100 kcal^{-1}) or not more than 200 mg 100 g^{-1}
Drinks	Carbohydrate: Vegetable juices and drinks based on them: shall not exceed 10 g 100 ml^{-1} Fruit juices, nectars and drinks based on them: shall not exceed 15 g 100 ml^{-1} Other nonmilk-based drinks: shall not exceed 5 g 100 g^{-1} Fat: Content should not exceed 1.1 g 100 kJ^{-1} (4.5 g 100 kcal^{-1}) Vitamin C: Fruit/vegetable juices – content should be at least 6 mg 100 kJ^{-1} (25 mg 100 kcal^{-1}) or at least 25 mg 100 g^{-1} Vitamin A: Vegetable juices – content should be at least 25 μg RE 100 kJ^{-1} (100 μg RE 100 kcal^{-1}). May not be added to other foods Vitamin D: May not be added

[a]Commission Directive 96/5/EC, as amended. RE, retinol equivalents.

Homemade Weaning Foods

Preparing weaning foods at home is largely a matter of adapting normal cooking processes. If a meal is being cooked for other members of the family, it is usually possible for a suitable weaning food to be prepared from the same basic ingredients.

Preparation of homemade weaning foods usually involves omission of certain ingredients, such as chilli peppers, garlic, ginger, and other spicy ingredients, which might irritate the infant's digestive tract. Salt should not be added during cooking.

Grilling, baking, poaching, or boiling are all suitable cooking methods, but frying with added fat is not recommended. When the food has been cooked, it is necessary to reduce it to a suitable consistency, according to the infant's stage of development. Many hand and electrical gadgets are available for this purpose.

Using a microwave oven is not recommended to heat foods for infants as microwaves can cause uneven heating and food could burn an infant's mouth. However, if a microwave is used to heat foods for infants, the manufacturer's instructions must be followed exactly. The food should always be stirred to insure an even distribution of heat and the temperature checked before the food is fed to an infant.

Weaning Foods in Developing Countries

The first foods added to the diet are generally staple foods. These might include cereals and tubers such as maize, rice, millet, or sweet potato. In China, the staple of the diet is rice, whereas semolina is used in India. In Kenya, the infant diet is based on maize and in Nepal on soya, wheat, and maize.

The simplest recipe for weaning foods is one which has only two ingredients, for example a cereal or root mixed with a legume. Other foods can be added to this basic mix to make a complete meal. Such recipes are known as multimixes and consist of four basic ingredients:

1. A staple as the main ingredient – preferably a cereal.
2. A protein supplement from a plant or animal food – beans, groundnuts, meat, milk, chicken, fish, eggs, etc.
3. A vitamin and mineral supplement – a vegetable and/or fruit.
4. An energy supplement – fat, oil, or sugar to increase the energy concentration of the mix.

Hygiene

Since infants are particularly susceptible to infection during the first year of life, special care must be taken to avoid introducing a source of infection via weaning foods. The scrupulous washing of hands and all equipment used in the home preparation of food is essential. Plastic or glass items, such as plates, spoons, and mugs, can be sterilized chemically, but metal utensils must be boiled.

All weaning food manufacturers have very strict specifications to insure the safety of their products. Cleanliness is maintained in the manufacturing environment with rigorous standards of hygiene.

See also: **Infants**: Nutritional Requirements; Weaning

Further Reading

Commission directive 96/5/EC (1996) Processed cereal-based foods and baby foods for infants and young children. *Official Journal of the European Communities* L49: 17–28.

Commission directive 98/36/EC (1998) Amending directive 96/5/EC on processed cereal-based foods and baby foods for infants and young children. *Official Journal of the European Communities* L167: 23–24.

Department of Health (1999) *Committee on Toxicity of Chemicals in Food, Consumer Products and the Environment. Peanut Allergy.* London: Department of Health.

Department of Health (1994) Report on Health and Social Subjects No. 45. Weaning and the Weaning Diet. London: HMSO.

Mills A and Tyler H (1992) *Food and Nutrient Intakes of British Infants Aged 6–12 Months*. London: HMSO.

INFANTS

Contents

Nutritional Requirements

J Morgan, University of Surrey, Guildford, Surrey, UK

Introduction

For the purposes of this article, infants are defined as aged from birth to 12 months.

For many years estimating energy and nutrient requirements in infancy has relied on data derived from feeding studies and growth patterns of healthy full-term wholly breast-fed infants. Thus requirements in infancy were based on food intake data. However, more direct evidence for requirements (for example, energy expenditure data) has provided more direct data on needs. Thus, recent national and international committees have reflected this knowledge when proposing the best estimate of infant nutrient requirements and reference intakes.

Individual Variation in Food Intake

The wide individual variation in energy and nutrient intakes at any age has been recognized for 40 years. At 1 year of age there are infants at the upper end of the range consuming twice as much food as those at the lower end – some infants at 6 months are consuming energy intakes equivalent to those of young women (Table 1). The phenomenon of the variation in food intake (and hence nutrient requirement) highlights the need to distinguish between the terms 'requirement' and 'recommended intake.'

The effect of gender is well recognized: On average, boys have a greater intake of energy than girls. The US *Recommended Dietary Allowances* report (1989) provides an average energy allowance (AEA) for energy requirements from birth to 10 years which does not distinguish between sexes for energy requirements. However, other authorities, such as the UK Department of Health, do distinguish between the sexes with regard to energy requirements.

In young infants the nutrient requirement for growth itself is substantial, representing up to 30% of the energy requirement, but there exists a large variation within the normal range on the rate of growth and probably also in the composition of the tissue laid down.

The 'average' composition of human milk has been used as a basis for estimating requirements and also for calculating the composition of infant milk formulae. It is as well to remember, however, that the composition of human milk varies, from mother to mother, and from day to day. It depends on the stage of lactation and on the mother's diet (particularly the content of polyunsaturated fatty acids). As well as providing nutritional substances, milk contains several growth factors, including epidermal growth factors, insulin, and insulin-like growth factor (IGF-1). The physiological importance of these substances has not yet been established.

Table 1 Energy intake of boys and girls, illustrating average intake and maximum and minimum values

	Age group (years)	*Average (kcal/kJ day^{-1})*	*Maximum (kcal/kJ day^{-1})*	*Minimum (kcal/kJ day^{-1})*
Boys	1	1154/4824	1547/6466	811/3390
	2	1406/5877	1929/8063	959/4009
	3	1691/7068	2340/9781	1250/5225
Girls	1	1152/4815	1631/6818	842/3520
	2	1431/5982	2048/8561	1055/4410
	3	1533/6408	2328/9731	991/4142

Calculated from Widdowson EM (1947) *A Study of Individual Children's Diets*. London: Medical Research Council/HMSO.

Requirement for Energy

Recently, recommendations for energy have been given as the estimated average energy requirement (EAR) – about half the population will usually need more than the EAR and half less. A recommendation for an energy intake that is adequate for all infants could result in obesity in those infants who have lower than average requirements, therefore the EAR reflects the estimated average energy requirement.

Energy requirements for infancy should be estimated from data on energy expenditure. The 1985 World Health Organization (WHO) / Food and Agriculture Organization report on energy requirements in humans based its recommendations for infants on the observed intake of healthy infants growing normally. It is recognized, however, that observed and desired energy intake may differ, and that a better estimate of energy needs is provided by energy expenditure. The available database on energy expenditure in infants has greatly expanded since 1985, thanks in large part to the application of the doubly-labeled water technique. This noninvasive method permits a reliable quantification of total daily expenditure in infants under normal conditions of everyday life. There are now thousands of data points on energy expenditure in infants obtained with this method. In addition, new data on growth and energy intake of exclusively breast-fed infants became available. Taken together, these data on intake, growth, and energy expenditure suggest that the actual energy requirements for breast-fed infants may be 10–40% lower than the 1985 estimates. One explanation for the overestimation of the previous recommendation may be first, that the data set of observed intakes included a large proportion of formula-fed infants, who consume higher amounts of energy than breast-fed infants, and second, that the reference standard for desirable growth is also based largely on a population of predominantly formula-fed infants.

The fact that healthy, exclusively breast-fed infants grow at a slower rate than the current reference standards is the subject of ongoing research. There is evidence, however, that these infants are no different than infants consuming more energy and growing at a faster rate in terms of morbidity, sleep patterns, or physical activity.

As a result of these studies, efforts are under way to develop a growth reference chart for exclusively breast-fed infants, which presumably would reflect a more physiological growth pattern than that of formula-fed infants.

Dietary reference values (DRV) for infants (males and females) from reports published under the auspices of UK, USA, and European committees are given in Table 2.

Requirement for Protein

Estimates of protein requirements for young infants are based on milk intake data. The WHO, the National Research Council, and the Department of Health have utilized this method.

Table 3 shows the reported intake of protein in human milk-fed male and female babies between birth and 4 months. Assumptions have been made regarding the protein content of milk (and complementary foods after 4 months) and the figure of 1.15 g of protein per 100 ml (for milk) is generally accepted. Because of the high quality and easy digestibility of human milk protein, few infants fed human milk develop protein deficiency.

In estimating protein requirements from 6 to 12 months, the WHO Committee used the factorial approach. A gray area of knowledge exists for protein requirements between 4 and 6 months of age. Estimates of protein requirements have been extrapolated from energy requirements, i.e., 1040 ml of breast milk per day is needed to fulfill the theoretical energy needs, and hence protein requirements, of male

Table 2 Dietary reference values for energy for males and females

Age (years)	*UK EAR (kcal/MJ day^{-1})*	*USA AEA (kcal/MJ day^{-1})*	*EU EAR (kcal/MJ day^{-1})*
Males			
0–3	545/2.28	650/2.7	550/2.3
4–6	690/2.89	650/2.7	715/3.0
7–9	825/3.44	850/3.5	850/3.5
10–12	920/3.85	850/3.5	930/3.9
Females			
0–3	515/2.16	650/2.7	500/2.1
4–6	645/2.69	650/2.7	670/2.8
7–9	765/3.20	850/3.5	790/3.3
10–12	865/3.61	850/3.5	885/3.7

EAR, estimated average requirement; AEA, average energy allowance.

Table 3 Average daily intake of protein by breast-fed infants, 0–4 months

Age (months)	*Boys (g/kg day^{-1})*	*Girls (g/kg day^{-1})*
0–1	2.46	2.39
1–2	1.93	1.93
2–3	1.74	1.78
3–4	1.49	1.53

Modified from World Health Organization (1985) *Energy and Protein Requirements*. WHO Technical Report Series no. 522 and FAO Nutrition Meeting Report Series no. 724.

Table 4 Dietary reference values for protein

Age (years)	*UK RNI (g day^{-1})*	*USA RDA (g day^{-1})*	*EU PRI (g day^{-1})*
0–3	12.5[a]	13.0	
4–6	12.7	13.0	14.0
7–9	13.7	14.0	14.5
10–12	14.9	14.0	14.5

RNI, reference nutrient intake; RDA, recommended dietary allowance; PRI, population reference intake.
[a]Based on egg and milk protein: assume complete digestibility.

infants aged 5–6 months. This approach is obviously not ideal.

Table 4 outlines the dietary reference values for protein (g day^{-1}) for the period 0–12 months from various reports. A major point of issue when estimating requirements is the variation in protein needs for growth. The uncertainty related to this factor has resulted in the addition of 50% to the theoretical nitrogen accretion in the calculation of protein requirements. In fact, adjustment by 50% may even be too small. This is obviously an area where research is urgently needed, i.e., to investigate the daily variation in growth rate and the constraints in storage of essential amino acids to be available when maximum growth occurs. To adjust for interindividual variation in growth and maintenance, a factor of 12.5% has been incorporated. A theoretical basis for calculating the variability of protein requirement for growth does not exist. Work by Fomon indicates that over a period of 1 month children vary in their average daily rate of growth, with a coefficient of variation of 37%. This variation in weight gain is much greater than the variability in protein intake.

Human milk contains characteristic nitrogen-containing components. In fact, 25% of the nitrogen is derived from nonprotein sources: free amino acids, small peptides, amino sugars, creatine, creatinine, and glycolipids. The content of free taurine is 30–40 times higher in human milk compared with cows' milk. Although the nutritional or biochemical role of some of these substances is not fully understood, their presence, like that of the growth-promoting factors previously mentioned, should not be dismissed.

The specific amino acid requirements of infants are of interest. It appears that, beside the eight essential amino acids, three other amino acids (cysteine, taurine, and histidine) are essential in the postnatal period. This has implications for one aspect of infant feeding – the provision of total nutrition by intravenous (IV) feeding. Only recently have amino acid preparations for IV nutrition been tailored to the requirements of infants, including an appropriate amino acid profile.

Requirements for Fats

In the broadest sense, infants have a requirement for fats to provide energy, to provide essential fatty acids, to facilitate the absorption of the fat-soluble vitamins A, E, and D, and as a precursor of structural lipids and eicosanoids. Triacylglycerols in human milk contain more than 150 different fatty acids, but the origin and role of many of these are largely unknown. The average intake of fat can be calculated for breast-fed infants if the fat content of the breast milk is known (Table 5).

The specific requirements of long-chain polyunsaturated fatty acids (PUFA) and cholesterol in infants are not often considered. Infants have specific requirements for PUFA. In the 1950s and 1970s there were reports of deficiency in long-chain fatty acids in premature infants fed parenterally with no lipid source, and in full-term infants fed on formula devoid of linoleic acid. Human milk contains on average 7.2 g of linoleic acid per 100 g total fatty acids or 285 mg per 100 ml, i.e., 3.7% of total energy. The DRV report panel recommended that for infants and young children: 'linoleic acid should provide at least 1 per cent of total energy and alpha linolenic acid at least 0.2 per cent of total energy.'

Table 5 Average daily intake of fat by breast-fed infants aged 0–4 months

Age (months)	*Breast milk consumed (ml)*	*Weight (kg)*	*Average fat intake*[a]	
			(g)	*(g/kg day^{-1})*
Boys				
0–1	719	3.8	30	8
1–2	795	4.75	33	7
2–3	848	5.6	36	6
3–4	822	6.35	35	5
Girls				
0–1	661	3.6	28	8
1–2	731	4.35	30	7
2–3	780	5.05	32	6
3–4	756	5.7	32	5

[a]Fat content of breast milk 4.2 g per 100 ml.

Human milk is relatively rich in cholesterol, providing an average 16 mg l^{-1}. The existence of this sterol in breast milk is indicative of a biological role, although no authority to date has made recommendations on desirable intakes in infants not receiving human milk.

Several studies have suggested a specific role for certain PUFA in visual development. Eicosapentaenoic acid and docosapentaenoic acid have been evaluated in term and preterm infants; these fatty acids are found in high concentration in human milk, but they have not been added to infant formulae until recently. The requirement for them may be higher in preterm infants, but their contribution to visual development of the full-term healthy infant is not fully documented at this point. Several commercial infant formulae, however, are now fortified with eicosapentaenoic acid.

Requirement for Carbohydrate

Eighty percent of the total carbohydrate in human milk is a sugar, lactose (approximately 7 g per 100 ml of human milk). In early infancy, when the diet consists wholly of milk, about 40% of the energy intake is derived from lactose.

As the diet of the infant becomes more varied, the carbohydrate content also varies. Table 6 provides detailed information on the carbohydrate contribution of foods given to infants. After about 4–6 months human milk or infant milk formulae alone cannot meet an infant's energy and other nutrient requirements (notably iron). It is important to remember that infants are not miniature adults and that they have their own specific nutrient needs. There is evidence, however, that dietary guidelines for adults are being inappropriately applied to infants. There is a particular concern regarding the provision of nonstarch polysaccharides (NSP) and also the use of low-fat products in the diet of young children. The UK panel recommended in 1991 that children should have proportionally lower NSP intakes than adults, and that consumption should not be at the expense of energy-rich foods. However, no specific guidelines were given regarding safe intakes for NSP for infants.

Table 6 Average carbohydrate content of several foods for infants

Food	*Carbohydrate (% energy)*
Human milk	37
Cows' milk	29
Strained meat	1
Strained fruits	96
Strained desserts	89
Strained vegetables	80
Strained soups and dinners	56
Strained high-meat dinners	29

Data from Fomon SJ (1974) *Infant Nutrition*, 2nd edn. Philadelphia: WB Saunders.

Requirement for Vitamins and Minerals

In the UK, supplementation of vitamins is recommended for breast-fed infants and young children from 6 months to 5 years unless the child's diet is diverse and plentiful. The daily dose of five vitamin drops provides 7 μg of vitamin D, 200 μg of vitamin A, and 20 mg of vitamin C.

Requirement for Vitamin A

Dietary requirements for vitamin A in infants have been usually based on and extrapolated from intakes calculated from values obtained from a wholly breast-fed infant receiving milk from a well-nourished mother. The breast milk from women living in the USA and Europe contains 40–70 μg of retinol per 100 ml milk and between 20 and 40 μg of carotenoids per 100 ml of milk. Assuming milk is consumed at 750 ml per day (and the concentration of retinol is 40 μg per 100 ml), the infant would receive about 300 mg of retinol per day. After taking into account factors such as individual variation in retinol content of milk, the US committee (NRC 1989) recommend an intake of 350 μg day^{-1} up to 1 year. The UK recommendations (DoH 1991) are set at 350 μg of retinol equivalents per day. In developing countries, vitamin A deficiency is uncommon in the first year of life. A drastic fall in the vitamin A content of milk has to be preceded by the exhaustion of the mother's stores of this vitamin.

Requirement for Vitamin D

Only small amounts of vitamin D are secreted in human breast milk, which indicates that a dietary source of this substance is teleologically unimportant; the major source of vitamin D is the action of ultraviolet (UV) light on the skin. The amount of exposure to UV light will differ from infant to infant. Because of this uncertainty, and to safeguard against lack of exposure to sunlight in the UK, the reference nutrient intake (RNI) is 8.5 μg of vitamin D per day (from food or supplements) from 0 to 6 months, and 7 μg day^{-1} from 7 to 12 months. The US recommended dietary allowance (RDA) is set at 7.5 μg day^{-1} for infants from birth to 6 months of age, and 10 μg day^{-1} from 6 months onwards because of increased body mass. The European 'safe range' from 0 to 3 years is 10–25 mg day^{-1}.

Thus, three authorities recommend a target intake for vitamin D, although there is no doubt that healthy,

wholly breast-fed infants who are exposed to sunlight do not require oral supplementation. Vitamin D deficiency in the material diet, lack of sunlight, adverse social or environmental factors, and infant prematurity indicate the need for supplementation.

Excessive intake of vitamin D is potentially harmful: only five times the RNI (45 μg day^{-1}) has been reported to be associated with hypervitaminosis in young children.

Requirement for Ascorbic Acid (Vitamin C)

Scurvy occurs primarily in infants fed diets consisting exclusively of cows' milk; as this mode of feeding is very uncommon, scurvy is rare in infants under 1 year of age. There are reports that infants receiving as little as 7 mg of vitamin C per day are protected from scurvy. Human milk contains approximately 4 mg of vitamin C (ascorbic acid and dehydroascorbic acid) per 100 ml. An infant receiving 750 ml of human milk per day will receive in the order of 30 mg of vitamin C. The UK recommends a daily intake of 25 mg of vitamin C up to 12 months; the USA recommends a daily intake of 30 mg and 35 mg of vitamin C from 0 to 6 months and 7 months to 1 year of age respectively.

Requirement for Iron

A normal healthy fetus stores iron in the last trimester of pregnancy. Because of this, a full-term infant can maintain satisfactory hemoglobin levels, without any other sources of iron, for about the first 4–6 months of life. Breast milk is noticeably low in iron – 76 μg 100 ml^{-1}. However, because the iron is well absorbed (50–70%, compared with 10–30% from cows' milk), and because an infant has hepatic hemoglobin stores, a dietary source is unnecessary. After 4–6 months of age, and in nonbreast-fed infants from birth, a dietary iron source is needed. Infant milk formulae and many infant food products contain added iron, thus insuring an appropriate dietary intake; in addition food commodities such as red meat can be included to provide a source of bioavailable iron. Dietary reference values for iron differ from country to country. At ages 4–6 months the UK value is 4.3 mg day^{-1}, the USA RDA is 6.0 mg day^{-1}, and the WHO recommendation is 8.5 mg day^{-1}.

Requirement for Calcium

Adequate mineralization of the skeleton occurs in breast-fed infants receiving an intake of calcium as low as 200–300 mg day^{-1}. The most important factor governing calcium absorption is the degree of skeletal mineralization, mediated by a feedback mechanism, controlled by synthesis of 1,25-dihydroxycholecalciferol (a derivative of vitamin D). Others factors include a high dietary lactose content, the composition and structure of dietary fat, and a low buffering effect – all provided by human milk. Approximately 66% of calcium intake is absorbed. Although the requirement of calcium is probably 250 mg day^{-1}, the recommended intake is usually set in the region of 500 mg day^{-1}. The safety margin is necessary to allow for a variance in the absorption rate and to allow for individual variation in requirement.

Conclusion

An infant is not a small adult. During the first year of life, in particular, the infant has specific requirements related to physiology. As these requirements are becoming more fully understood, so more completely are we able to meet the requirement by adjusting infant milk formula. There are still gaps in our knowledge and these have been highlighted.

See also: **Amino Acids**: Metabolism; **Ascorbic Acid**: Physiology; **Calcium**: Physiology; **Carbohydrates**: Requirements and Dietary Importance; **Energy**: Intake and Energy Requirements; Measurement of Energy Expenditure; Energy Expenditure and Energy Balance; **Growth and Development**; **Infant Foods**: Milk Formulas; Weaning Foods; **Iron**: Physiology; **Lactation**: Physiology; **Protein**: Requirements; **Retinol**: Physiology

Further Reading

Butte N (1996) Energy requirements of infants. *European Journal of Clinical Nutrition* 50: S24–S36.

Department of Health (1991) *Dietary Reference Values for Food Energy and Nutrients for the United Kingdom*. Report on Health and Social Subjects, no. 41. London: HMSO.

Department of Health (1994) *Weaning and the Weaning Diet*. Report on Health and Social Subjects, no. 45. London: HMSO.

European Community (1995) *Nutrient and Energy Intakes*. Luxembourg: European Communities: The Stationary Office.

Fomon SJ (1993) *Nutrition of Normal Infants*. St Louis: Mosby.

Morgan JB, Redfern AM and Stordy BJ (1995) Healthy eating for infants – mothers' attitudes. *Acta Paediatrica* 84: 512–515.

National Research Council (1989) *Recommended Dietary Allowances*, 10th edn. Washington, DC: National Academy Press.

Paul AA, Whitehead RG and Black AE (1990) Energy intakes and growth from two months to 3 years in initially breast-fed children. *Journal of Human Nutrition and Dietetics* 3: 79–92.

World Health Organization (1985) *Energy and Protein Requirements*. WHO Technical Report Series no. 522 and FAO Nutrition Meetings Report Series no. 724. Geneva: WHO.

Breast- and Bottle-feeding

S Flack and V Shaw, Great Ormond Street Hospital for Sick Children, London, UK

Introduction

Breast milk is recognized as the best source of nutrition for healthy infants. It is species specific. Optimum infant nutrition from breast milk is dependent on adequate volume and mother's good health and nutritional status. The practice of demand feeding helps ensure adequate milk supply is maintained. Insufficient supply may be corrected by encouraging more frequent feeds or expressing milk. The infant's suckling and milk removal stimulate increased breast milk production.

Breast milk composition is not standard. It is influenced by maternal diet and by the mother's nutritional status. It changes during the course of lactation to meet the requirements of the developing infant. Milk shows diurnal variation; it varies from mother to mother, day to day, and even during a feed. At the initiation of a feed the milk has a low fat content (fore milk) and this increases up to fourfold during the feed (hind milk), to facilitate satiety.

The major change in breast milk composition occurs as the colostrum, the initial milk produced, changes to mature milk. Milk can be considered mature at about day 10 postpartum (Table 1).

The use of term pooled human milk as a sole source of nutrition for preterm infants remains controversial. It is recognized that early preterm mother's milk is more nutrient dense than term milk and is therefore more suitable for the premature infant's requirements. However, additional energy and nutrient fortification may be necessary.

Colostrum

This is the milk produced in the first few days after birth. It is compositionally distinctive from mature human milk (Table 1). It contains a higher concentration of protein, vitamin A, vitamin B_{12}, sodium, and some trace minerals. Initial colostrum has a high sodium content (60 mmol l^{-1}), which decreases rapidly as the paracellular pathways close. Much of the protein is present as secretory immunoglobulin A (sIgA) which functions to protect the gut from infection. It has a lower fat content than mature breast milk and, therefore, a lower energy content – 236 kJ (56 kcal) per 100 g.

Mature Human Milk

Energy

The energy value of mature human milk is quoted as 289 kJ (69 kcal) per 100 g in standard food tables in the UK. It is accepted that this is an average value, as milk composition is variable. Advances in technology have led to new estimates of its energy content. Doubly labeled water analysis has shown the metabolizable energy content of breast milk to be 2219 kJ (530 kcal) per liter at 6 weeks and 2428 kJ (580 kcal) per liter at 3 months. The clinical significance of this finding is unknown. Infant formulas are manufactured to provide a similar energy content to that of breast milk and contain approximately 2805 kJ (670 kcal) per liter; this has lead to speculation that infants receiving formula may be being overfed.

Protein

Protein in human milk is not homogeneous. It is traditionally subdivided into two main groups: casein and whey. Varying figures exist in the literature for total whey and casein content at different stages of lactation. The concentration of whey protein decreases during lactation. Some work suggests that the casein to whey ratio is 10:90 in colostrum, changing to a ratio of 45:55 in mature milk. Other literature suggests that whey proteins continue to represent about two-thirds of the protein in mature breast

Table 1 Comparison of the composition of colostrum and mature human milk

Nutrient per 100 g	*Colostrum*	*Mature human milk*
Energy (kJ)	236	289
Protein (g)	2.0	1.3
Fat (g)	2.6	4.1
Carbohydrate (g)	6.6	7.2
Sodium (mg)	47	15
Potassium (mg)	70	58
Calcium (mg)	28	34
Iron (mg)	0.07	0.07
Retinol (μg)	155	58
Carotene (μg)	(135)[a]	(24)[a]
Vitamin D (μg)	N[b]	0.04
Thiamin (mg)	Trace	0.02
Vitamin B_{12} (μg)	0.1	Trace
Folate (μg)	2	5
Vitamin C (mg)	7	4

[a]Estimated amount.
[b]Significant quantities, but no reliable information.
From Macdonald S and Shaw V (1993) Breast- and bottle-feeding. In: Macrae R, Robinson RK and Sadler MJ (eds) *Encyclopaedia of Food Science, Food Technology* and Nutrition, pp. 2507–2511. London: Academic Press, with permission.
Data from Holland B, Welch AA, Unwin ID *et al*. (1991) *McCance and Widdowson's The Composition of Foods*, 5th edn. Cambridge: Royal Society of Chemistry/Ministry of Agriculture, Fisheries and Food.

milk. The whey fraction's major components are α-lactalbumin, lactoferrin, and sIgA. The concentration of sIgA is approximately $9\,g\,l^{-1}$ in colostrum, falling to $1–2\,g\,l^{-1}$ in mature milk, representing about 10% of the protein. sIgA acts against enteric and respiratory immunogens. Lactoferrin also has an antibacterial affect. Its binding of free iron inhibits bacterial growth, enhances iron absorption, and stimulates proliferation of the mucosa. Casein and α-lactalbumin supply the essential amino acids.

Breast milk also contains nonprotein nitrogen, e.g., urea, nucleotides, and free amino acids such as taurine. Some of these constituents are believed to be important in the development of the infant. Taurine is the end product of methionine and cystine metabolism, but is considered a conditionally essential amino acid in infants. It is required for retinal development and digestion. It may also be protective against oxidative cellular injuries.

Fat

Fat is the main energy source in human milk, contributing 54% of the total energy in mature milk. The amount of fat and fatty acid composition in breast milk is variable and affected by maternal diet, nutritional status, and starvation (by the catabolism of her adipose tissue). It provides the essential fatty acids, i.e., linoleic (C_{18}:2*n*-6) and α-linolenic (C_{18}:3*n*-3) acids, contains long-chain polyunsaturated fatty acids (LCPs), and is a carrier for the fat-soluble vitamins (A, D, E, and K) and prostaglandins. The two principal functions of LCPs are as structural components of membranes and as the precursors of the biologically active eicosanoids.

Fatty acid composition also changes during the course of lactation. These changes can be linked to the requirements for neonatal brain development. Initial rapid cell division requirements are different from that of myelination.

Breast milk contains a form of inactive lipase, which is activated by bile salts in the duodenum and aids fat absorption in the infant. This is important, as pancreatic lipase secretion and bile salt conjugation are inefficient in the newborn.

Carbohydrate

Lactose, a disaccharide composed of glucose and galactose, is the main carbohydrate in human milk. It constitutes approximately 80% of the total carbohydrate. Oligosaccharides are also present in both colostrum and mature milk. The concentration of oligosaccharides in breast milk decreases as colostrum changes to mature milk. Oligosaccharides and lactose promote the growth of bifidus flora in the gut. Unabsorbed lactose is converted by intestinal microflora to lactic acid, creating an acidic pH which inhibits the growth of many pathogens and helps to protect the breast-fed infant from gastrointestinal infections. The additional benefit of this low pH is its effect on calcium. It increases the solubility of calcium and hence its absorption.

Vitamins and Minerals

Vitamin requirements can usually be met by adequate volume of breast milk from a well-nourished mother. Maternal diet influences the concentrations of most vitamins in human milk. Cases of clinical deficiencies of vitamins B_{12}, D, and K have been observed. Clinical rickets is rare but may occur in breast-fed infants, usually related to the infant or mother having limited exposure to sunlight. In the UK vitamin D supplements are recommended if there is concern about maternal diet or the infant's exposure to sunlight. Vitamin B_{12} is only found in animal products, therefore there have been case reports of deficiency (megaloblastic anemia) in breast-fed infants of vegan mothers. Breast milk has a low vitamin K content. Hemorrhagic disease occurs in approximately 2% of breast-fed infants. Additional vitamin K supplements are required during the period of exclusive breast-feeding. In the UK it is usually given intramuscularly at birth.

Breast milk provides all the major minerals and trace elements known to be essential for healthy infants in sufficient quantities for at least the first 4 months. The levels of calcium, phosphorus, and iron are low, but their bioavailability from human milk is excellent. However by 6 months additional iron and zinc supplements may be necessary; these requirements will be met by a balanced weaning diet.

Water

Mature human milk is 87% water. The fluid intake of an exclusively breast-fed infant will always be adequate provided demand feeding is practised. Additional water supplements are unnecessary even in very hot climates.

Nutritional Value of Cows' Milk and Infant Formula

Unmodified cows' milk is not a suitable feed for infants. It has a high renal solute load, inadequate vitamin and mineral profile, and is poorly absorbed. Infant formula is manufactured from modified cows' milk as a substitute for human breast milk. Available formulas in the UK conform to European Community compositional guidelines. These formulas may be nutritionally adequate but they lack the human anti-infective, humoral, and cellular factors found in breast milk.

Advantages of Breast-Feeding

There are numerous advantages of breast-feeding, including nutritional, emotional, physiological, and economic factors. Breast-feeding guarantees regular close contact between mother and child, promoting the emotional bonding process.

Many of the nutritional advantages of breast milk have already been discussed. In summary, milk from a healthy mother provides all the nutrients in an appropriate and easily absorbable form to meet the requirements of the infant.

Additionally, this feeding method is cheap, safe, and offers anti-infective advantages. Cross-infection from adults to the infant may be reduced; as only the mother is able to breast-feed, the number of adults handling the child may be reduced. Breast-feeding is convenient, as the milk will always be available at the correct temperature and ready to feed. The mother does not need to buy equipment, sterilize bottles and teats, or prepare infant formula in hygienic conditions. Many constituents in human breast milk stimulate the development of the infant's own immunity. Through the milk the mother is able to transfer to her infant antibodies (IgA) which are specific antigens against their environment. This fills the 'immunity gap' whilst the infant's immune system develops and its gut epithelial cells mature to be able to produce IgA. Other important defenses include macrophages (ingest foreign particles), lactoferrin (bacteriostatic), milk leukocytes (phagocytosis of bacteria, viruses, and fungi) and lysozyme (bactericidal and anti-inflammatory activity). Whilst the baby's own immune system matures, these bioactive milk constituents act synergistically to provide a passive immunological support system for the infant.

Epidemiological evidence suggests that breast-feeding has a protective effect against diarrhea, respiratory infections, necrotizing enterocolitis, and otitis media. Research has recently focused on possible links between infant feeding practice and chronic illness such as Crohn's disease and childhood lymphoma in later life. The American Academy of Pediatrics reviewed feeding method and incidence of type 1 insulin-dependent diabetes mellitus (IDDM). It recommended breast-feeding as the main source of infant nutrition and avoidance of intact cows' milk protein in the first year of life if there is a strong family history of IDDM.

The role of breast-feeding in protecting the infant against the development of atopic and allergic disorders in later life is controversial. Further research is needed in this area. It is hypothesized that the immunological factors in breast milk may protect the infant from absorbing intact allergens. Breast-feeding results in delayed exposure to many allergenic compounds in foods. However, components of the mother's diet may be passed to the infant via breast milk. The role of maternal exclusion diets in exclusive breast-feeding for high-risk infants is also controversial. Allergy to peanut is potentially life-threatening and for this reason, in the UK, the Committee on Toxicity of Chemicals in Food, Consumer Products and the Environment has suggested that pregnant and lactating women should avoid peanuts if they themselves, the father, or any sibling of the baby is atopic.

Enhanced cognitive development is associated with breast-feeding. The role of LCPs in breast milk and infant development has been the subject of recent research. Docosahexaenoic acid (DHA) (C_{22}:6*n*-3) and arachidonic acid (C_{20}:4*n*-6) are considered essential nutrients in early infancy. Dietary DHA appears to be essential for optimal membrane function in the nervous system and thus cognitive development in premature infants. Many infant formulas now contain added LCPs but other constituents of breast milk, e.g., growth factor, hormones, and choline, may additionally aid brain maturation.

Malnutrition can result from either the overconcentration or overdilution of infant formulas. There is no risk of this with breast milk. Breast-feeding is linked with child survival in developing countries due to its protective effect against diarrheal illness, environmental factors (e.g., lack of clean water supply), and economic factors (e.g., expense of buying adequate volumes of infant formulas).

It is known that breast-feeding has advantages for maternal health. Breast-feeding stimulates contraction of the uterus. It is associated with weight loss and thus a more rapid return to prepregnancy weight. Some epidemiological evidence suggests a link between extended lactation and reduced risk of premenopausal breast, ovarian, and endometrial cancers.

Lactation produces hormonal changes which suppress fertility. At the population level, this lactational amenorrhea enables prolonged birth spacing. Additionally, in some societies there are taboos against sexual intercourse during the period of lactation. Breast-feeding therefore has a major contraceptive effect.

Reasons for Bottle-Feeding

There are few contraindications to breast-feeding; the main reasons are related to the health and medication requirements of the mother. Maternal medication usage may prevent breast-feeding. Most drugs can be transmitted to the infant via breast milk. Drugs may cause toxicity in the infant, e.g., ergotamine, and others inhibit the suckling reflex, e.g., phenobarbitone.

The *British National Formulary* is an important resource for health professionals to check the suitability of a medicine for breast-feeding women. UK recommendations are that only essential medications should be prescribed to breast-feeding mothers.

There are certain medical conditions in the baby where breast-feeding is contraindicated: galactosemia, long-chain fatty acid oxidation disorders, and glucose-galactose malabsorption.

Breast-feeding is known to protect against bacterial infections. Certain viruses however can be transmitted from mother to infant in breast milk. The largest global health problem is related to human immunodeficiency virus (HIV) transmission via breast milk. The rate of vertical transmission from mother to infant in human milk is between 5 and 15%. In England and Wales the incidence of vertically transmitted HIV infection is increasing. In the UK the current recommendation is that mothers who are HIV-positive, or in a high-risk category but unsure of their serological status, should not breast-feed. Experts argue that if antenatal HIV testing was more widespread, there would be an opportunity to counsel against breast-feeding and insure early diagnosis and treatment of infected children.

In the UK bottle-feeding is a safe alternative to breast-feeding. In developing countries with high infant mortality rates the situation is very different. The known protective effect of breast-feeding against gastrointestinal infection is very important. It has been estimated that bottle-fed infants are 14 times more likely to die from diarrheal diseases and four times more likely to die from pneumonia than exclusively breast-fed infants. In these countries the risk of vertical HIV transmission needs to be weighed against the risk of gastrointestinal infection caused by inadequate water supply and sanitation. The United Nations has recently issued a new directive stating that all women should be informed of the risk of vertical transmission via breast milk due to the escalating problem of HIV. It recommends the decision to breast-feed should be made by the mother due to the diversity of situations in developing countries.

There remain concerns that women who do not breast-feed in developing countries may be stigmatized, and there are concerns over the safety of alternative feeds. Previously, infant deaths in developing countries resulted from the problems of mixing infant formulas in unsuitable environments. Recently it has been hypothesized that rates of HIV transmission are affected by mastitis infection and vitamin A deficiency in pregnancy. A 1994 study of HIV-infected women in Malawi found that 32% of those who were vitamin A-deficient during pregnancy had passed HIV on to their infants. In contrast, only 3% with adequate vitamin A levels did so. If health strategies are implemented to alter these states, breast-feeding may be a recommended feeding method for infants of HIV-infected women in the future.

In the UK women's choice of feeding method is influenced by social and cultural factors. The Office for National Statistics identifies that 94% of mothers have decided on feeding method antenatally. They identified that 36% of the mothers surveyed in the UK intended to bottle-feed. Reasons for planning to bottle-feed included disliking the idea of breast-feeding and embarrassment associated with breast-feeding. In western societies taboos continue to be associated with breast-feeding in public. The other major reason for choosing to bottle-feed is that the mother is not the sole feeder of the child, so other people can help with the feeding.

Current Practices

Breast-feeding in developing countries remains almost universal. It remains common in Africa for most infants to be breast-fed into the second year of life. As these countries become more urbanized, breast-feeding trends alter. Generally, women in rural areas are more likely to breast-feed than their urban counterparts. This change may reflect the availability of infant formula, limited support networks, and variations of maternal employment patterns. The wealthiest and most highly educated in developing countries are less likely to breast-feed. The converse is true in developed countries. Immigrants' feeding choice is affected by the norm in both their country of origin and resident country.

In the UK, government policy consistently supports the promotion of breast-feeding. Since the 1970s infant feeding practices have been surveyed and results analyzed, leading to Department of Health initiatives to promote breast-feeding, e.g., National Breastfeeding Awareness Week. This has been an annual event since 1993. In 1995, the Office for National Statistics survey reported an increased incidence of breast-feeding (**Table 2**). The composition of the sample population contained more mothers aged over 30 from nonmanual social class groups and with higher education levels. Each of these factors is associated with increased incidence of breast-feeding. When the results are standardized for the age and level of education, no increase in breast-feeding incidence in England and Wales is demonstrated. In Northern Ireland, however, standardized rates increased from 36% in 1990 to 41% in 1995.

The World Health Organization recommends exclusive breast-feeding for at least 6 months. A reduced

Table 2 Prevalence of breast-feeding (%) up to 6 months of age by country

	England and Wales		*Scotland*		*Northern Ireland*	
	1990	*1995*	*1990*	*1995*	*1990*	*1995*
Birth	64	68	50	55	36	45
2 weeks	51	54	39	44	27	32
6 weeks	39	44	30	36	17	25
4 months	25	28	20	24	8	12
6 months	21	22	16	19	5	8
Base (number of infants)	4942	4598	1981	1863	1497	1476

From Foster K, Lader D and Cheesbrough S (1997) *Office for National Statistics – Infant Feeding 1995*. London: Stationery Office. Crown copyright is reproduced with the permission of the Controller of Her Majesty's Stationery Office.

Table 3 Duration of breast-feeding (%) for mothers who breast-fed initially in the UK

	1990	*1995*
Birth	100	100
1 week	85	85
2 weeks	80	80
6 weeks	61	65
4 months	39	42
6 months	33	32
Base (number of infants)	3438	3410

From Foster K, Lader D and Cheesbrough S (1997) *Office for National Statistics – Infant Feeding 1995*. London: Stationery Office. Crown copyright is reproduced with the permission of the Controller of Her Majesty's Stationery Office.

incidence of gastrointestinal infection is associated with breast-feeding for a minimum of 13 weeks. Results for the UK in 1995 showed that 20% of women had stopped breast-feeding at 2 weeks postpartum and 60% by the age of 4 months (Table 3). Initiatives are clearly needed to improve the duration of breast-feeding.

Reasons for stopping breast-feeding included 'insufficient milk.' This may be related to anxiety and lack of professional support and advice. Establishing breast-feeding is no longer instinctive in developed countries. Painful breasts, baby not suckling, not liking breast-feeding, and length of time taken to feed were other common reasons for cessation of breast-feeding.

The mother's experience of breast-feeding in hospital influences its duration. The giving of infant formula complementary feeds continues to exert a strong effect on the length of breast-feeding. In 1995, 34% of mothers whose infants received a bottle in hospital stopped breast-feeding within the first 2 weeks postpartum. Delays in first holding the baby significantly affected the likelihood of establishing breast-feeding. Only 51% of mothers who were separated from their baby for more than 12 h successfully initiated breast-feeding. These issues are being tackled by educating all health care professionals and the adoption of UNICEF's Baby Friendly Hospital Initiative. Between the 1990 and 1995 Office of National Statistics surveys, an increase in the number of babies kept at the mother's side at all times and the reduction in the number of breast-fed babies offered bottled formula during their hospital stay were seen. This is an encouraging trend.

Education and Advice

The 1995 Office of National Statistics survey showed that nearly all mothers (99%) had antenatal checkups and approximately 70% of new mothers attended antenatal classes. These are ideal opportunities in which to offer advice about feeding methods. The survey also showed that either a midwife or nurse gave 86% of new mothers help and advice at their first breast-feed in hospital. During their hospital admission, 35% of breast-feeding mothers experienced difficulties. The majority of these women (82%) received help and advice from health professionals and this has been shown to influence feeding duration. As about one-third of first-time mothers leave hospital within 48 h, access to professional support is restricted and potentially the declining lengths of hospital postpartum stay may impact on initiation and duration of breast-feeding.

After discharge from hospital, the mother is supported at home for the first 10 days by the midwife. Care is then transferred to the health visitor. Additionally several voluntary organizations can offer support on breast-feeding; these include the National Childbirth Trust, La Leche League and the Association of Breastfeeding Mothers.

See also: **Anemia (Anaemia)**: Iron-deficiency Anemia; **Atherosclerosis**; **Carotenoids**: Occurrence, Properties, and Determination; **Celiac (Coeliac) Disease**; **Cholecalciferol**: Properties and Determination; **Food Intolerance**: Food Allergies; **Infant Foods**: Weaning Foods; **Iron**: Properties and Determination; **Metabolic Rate**; **Protein**: Requirements; **Retinol**: Properties and Determination; **Rickets and Osteomalacia**; **Zinc**: Properties and Determination; Deficiency

Further Reading

Forsyth JS (1992) Is it worthwhile breast-feeding? *European Journal of Clinical Nutrition* 46 (suppl. 1): S19–S25.

Foster K, Lader D and Cheesbrough S (1997) *Office for National Statistics – Infant Feeding 1995*. London: Stationery Office.

Goedhart AC and Bindels JG (1994) The composition of human milk as a model for the design of infant formulas: recent findings and possible applications. *Nutrition Research Reviews* 7: 1–23.

Golding J, Rogers I and Emmett P (1997) Breast feeding: benefits and hazards. *Early Human Development* 49 (suppl.): S1–S204.

Henig MJ and Dewey KG (1996) Health advantages of breast-feeding for infants: a critical review. *Nutrition Research Reviews* 9: 89–110.

Henig MJ and Dewey KG (1997) Health effects of breast-feeding for mothers: a critical review. *Nutrition Research Reviews* 10: 35–56.

Holland B, Welch AA, Unwin ID *et al*. (1991) *McCance and Widdowson's The Composition of Foods*, 5th edn. Cambridge: Royal Society of Chemistry, Ministry of Agriculture, Fisheries and Food.

Macdonald S and Shaw V (1993) Breast- and bottle-feeding. In: Macrae R, Robinson RK and Sadler MJ (eds) *Encyclopaedia of Food Science, Food Technology and Nutrition*, pp. 2507–2511. London: Academic Press.

Statement of the Standing Committee on Nutrition of the British Paediatric Association (1994) Is breast feeding beneficial in the UK? *Archives of Diseases in Childhood* 71: 376–380.

Wharton B (1997) *British Nutrition Foundation Briefing Paper, Nutrition in Infancy*. London: British Nutrition Foundation.

White A, Freeth S and O'Brien M (1992) *OPCS Report – Infant Feeding 1990*. London: HMSO.

Weaning

A MacDonald, The Children's Hospital, Birmingham, UK

Introduction

The term 'weaning' comes from the word 'wenian,' which means 'to accustom.' It is a gradual process through which an infant moves from total dependence on milk to eating normal family meals. The most widely used definition is the process of introducing semisolid food into the infant's diet in addition to milk. Definitions of weaning vary from country to country and include 'the process of expanding the diet to include food and drinks other than breast or infant formula,' 'cease to be suckled,' or 'complete discontinuation of breast-feeding.' As many of the foods initially introduced are not truly solid, the German word 'beiokost' is sometimes used to refer to these foods as a group.

History of Weaning

In ancient Egyptian and Babylonian times, the common age of weaning was 3 years. In the sixteenth and seventeenth century, pap (mixtures of flour, water, and sometimes milk) and panada (a broth of bread, flour, meat or lentils, and sometimes milk) were described as infant foods in the literature. Weaning was quite a dangerous time when started in early infancy, and Queen Anne is said to have lost 18 children from spoon-feeding and wet nursing. The infant food industry started in 1867. By the latter part of the nineteenth century, there was a more cautious approach to solid-food introduction. It was widely recommended that vegetables, particularly green, were not given to infants before 2–3 years of age, although cereals and custards were commonly introduced by 4 months. The situation changed again in the early part of the twentieth century with the trend towards earlier weaning in developed countries. This continued in the UK until the 1970s and 1980s, although the latest Office of National Statistics (ONS) report would suggest that more parents are delaying the introduction of solids until at least 3 months of age.

Why Wean?

There are four important reasons for weaning. These are:

- If an infant is still unweaned by 6 months, the intake of energy is likely to be inadequate. Infants double their birth weight in the first 4–5 months of life (DOH, 1994), and breast or formula milk can no longer meet energy needs. Late weaning is associated with a failure to thrive in infants.
- Breast milk provides little iron and vitamin D. In the later stages of lactation, zinc and copper concentrations also begin to fall. However, the bioavailability of nutrients in breast milk is good, and no other foods are usually needed until the age of 4 months, providing the mother is taking a good mixed diet. In addition, for the first 4 months, the full-term infant relies on iron stores to help meet requirements. Sooner or later, these stores will be depleted, and iron-deficiency anemia will occur unless there is a supply from the weaning diet. Formula milk contains higher quantities of iron, vitamin D, zinc, and copper, but with a different bioavailability than breast milk, but for other reasons, it is still recommended that solids be introduced by 6 months.
- There are important developmental reasons for introducing solid food. The neuromuscular coordination needed to eat solid foods develops during

Table 1 Developmental changes relevant to weaning

4 months
Can maintain posture if supported in a chair
5 months
Can take soft, puréed food from a spoon, form a bolus, and swallow it
Can hold objects and put them to their mouths to suck
6 months – 8 months
Teeth begin to erupt
Sits without support
Can chew
Can hand feed items like biscuits
Learns to shut mouth, turn head, and indicate feed refusal
9–12 months
Can sit easily
Attempts to use spoon
Indicates and understands 'no'
Pincer grasp of small objects

the early months of life. At 4 months, most infants can put objects to their mouths, take soft foods from a spoon, from a bolus and swallow (Table 1). From about 5 months, teeth begin to erupt, and infants are able to chew. Feeding behavior therefore proceeds from sucking to biting and chewing. Chewing improves the mouth and tongue coordination, which is important for speech development. Introduction of different tastes and textures promotes biting and chewing skills. However, if introduction of solids is delayed beyond 6 months, the weaning process becomes more difficult. Infants are less receptive to new tastes, and it is more difficult to teach older infants to chew. From 7 months of age, infants learn to shut their mouths, turn their heads, and refuse to feed.

- The fourth reason is to provide a variety of tastes and flavors. This will enable the infant to eventually accept and enjoy normal family meals.

When to Wean?

The ideal time to introduce solids will vary between infants and differ between cultures. In traditional societies, it is still practice for infants to be almost entirely breast-fed for the first year of life. In many Western countries, particularly Europe and North America, it is recommended that 'the majority of infants should not be given solids before the age of four months but a mixed diet should be offered by the age of six months'. Early weaning is discouraged for a number of reasons:

- Poor muscular coordination. Before 3 months of age, infants have poor head control, and back support is weak, so it is difficult to hold infants in a position where they can be fed and can swallow semisolid food.
- The kidney is immature and may not be able to cope with a higher renal solute load, and this could lead to higher plasma concentrations of sodium and urea if there is insufficient water available for their excretion.
- The gastrointestinal tract is immature, intestinal and pancreatic enzyme production is not fully developed, and the gut is more permeable to some of the foreign proteins introduced with food, thus increasing the risk of allergic reactions.
- Early weaning in the developing countries has been associated with a poorer growth velocity than in infants who are exclusively breast-fed. This is probably associated with the lower energy density of the cereal-based gruel offered, resulting from the addition of extra water to produce a food of suitable consistency. Also, iron bioavailability from cereal-based gruel is low. Another major problem is that contaminated weaning foods causes infection and diarrhea. This, in turn, increases nutritional requirements but often reduces an infant's appetite and overall nutritional intake.
- Solid feeding before 15 weeks is associated with increased wheeze during the first seven years of life. There is also evidence that early exposure to solids is associated with an increased risk of eczema. In a group of children studied until the age of 10 years, it has been reported that infants given solids during the first 4 months were 1.6 times more likely to develop eczema than infants not given solids before 4 months.
- There is evidence that early introduction of solids is associated with increased body fat and weight in childhood.
- There is some suggestion that early introduction of gluten in solids may contribute to a higher incidence of coeliac disease. In Glasgow, for example, records indicate that increased numbers of children were admitted to hospital with celiac disease and were diagnosed earlier during 1958–1965 and 1966–1973 than during 1950–1957. Cereals containing gluten were also introduced earlier in Glasgow during 1958–1963, and 91% of infants were given gluten by 3 months of age in 1965.

In Western countries, there is still a trend towards early weaning. The ONS survey reported that 55% of infants had been given solids by 3 months and 91% by 4 months. Babies of Asian origin in the UK were given solids slightly later than white babies, but 90% still had solids by 4 months.

Nutritional Requirements of the Weaning Infant

In the early months of infancy, the nutrient intake is derived almost totally from human or infant formula,

but in later months, the diet becomes more diverse with the introduction of solid foods. In ensuring that the provision of all nutrients is met for infants, a careful selection of solid foods given in the appropriate amount is required.

Energy

Energy should be provided in adequate quantities throughout infancy to provide fuel for the basal metabolic rate (BMR), growth, and physical activity. The amount of energy required per kilogram of body weight declines gradually, owing to a decreasing growth rate and BMR, even though the energy requirements for activity will increase as the infant becomes older.

The UK estimated average requirements (EAR) for energy are given in Table 2. These are based on results of studies using the doubly labelled water technique of measurement that reflects the total energy expenditure in free living infants and children with additions made for the energy value of new tissue deposited during normal growth.

Despite the introduction of solids, milk still needs to provide a significant proportion of the total energy intake throughout the first year of life. Fat contributes 50% of the energy in breast and infant formula milk and is the main source of energy in infants less than 6 months of age. It still continues to be important in childhood, and infants over 6 months of age will still need 500–600 ml of milk daily. From studies estimating the nutrient intake of older infants, solids contributed 52, 57, and 64% of the total energy intake at between 7 and 8 months, 9 and 10 months, and 11 and 12 months, respectively, whilst milk made up the rest of the energy deficit.

However, adequate energy from weaning foods should be assured, and the energy density of weaning foods is important. If it is high in carbohydrate and low in fat, the food will be very bulky, and it may be difficult for the infant to eat sufficient quantities to meet their energy needs. Failure to thrive, which is a common problem in developing countries, may result.

Protein

Protein is a crucial determinant of linear growth; it provides nitrogen and essential amino acids for synthesis of body tissue. The DOH (1994) recommends an adequate intake of protein with a proper balance of essential amino acids during weaning (Table 3). The protein/energy ratio is also important, and it is recommended that infants between 6 and 9 months have an intake of not less than 1.74 g of protein per 100 kcal. In developed countries, there should be few problems in achieving an adequate energy/protein ratio, but in developing countries, if weaning foods such as rice and maize are supplemented with oil to improve energy density, this would significantly reduce the energy/protein ratio below the minimum safe ratio.

Vitamins and Minerals (Table 4)

Vitamin A This is required for growth, normal development, and differentiation of tissues. The 1988 FAO/WHO Expert Group considered that a daily intake of 350 μg retinol equivalents would meet the needs of all healthy infants and allow the building and maintaining of sufficient liver stores.

Vitamin D This has an important metabolic role in the control of calcium homeostasis and deposition of calcium in bones and teeth. There are two main forms of vitamin D: cholecalciferol, formed by the action of sunlight on 7-dehydrocholesterol, which occurs naturally in the skin of man, and ergocalciferol, a synthetic source of vitamin D. Weaning infants are vulnerable to poor vitamin D status. Weaning diets are generally low in vitamin D, and breast milk contains little vitamin D. Infant formula is fortified with vitamin D, and provided adequate amounts are consumed, infants fed these products are unlikely to become depleted. In the UK, it is recommended that

Table 2 Estimated average requirements for energy during weaning (DOH, 1991)

Age (months)	*Energy requirement (kcal kg^{-1} day^{-1})*
3	100
6	95
9	95
12	95

Table 3 Reference nutrient intake for protein during weaning (DOH, 1991)

Age (months)	*Protein (g^{-1} kg^{-1} day^{-1})*
4–6	1.6
7–9	1.6
10–12	1.5

Table 4 Reference nutrient intake for vitamins and minerals during weaning (DOH, 1991)

Nutrient	*Age*		
	0–3 months	*4–6 months*	*7–12 months*
Vitamin A (μg)	350	350	350
Vitamin D (μg)	9	9	7
Vitamin C (mg)	25	25	25
Calcium (mg)	525	525	525
Iron (mg)	1.7	4.3	7.8
Zinc (mg)	4.0	4.0	5.0

vitamin supplements including vitamin D be given to breast-fed infants from the age of 6 months and to formula-fed infants when the volume of milk decreases below 500 ml daily.

Iron This is an important nutrient during weaning, which is necessary for the structure of the heme part of hemoglobin, as well a component of myoglobin and many enzymes. For breast-fed infants, after the age of 6 months, the amount of iron from breast milk is insufficient to meet increasing needs. Therefore, adequate intakes of iron, from heme (meat) and non-heme sources (cereals, pulses, and vegetables), should be encouraged with foods rich in vitamin C to aid absorption. Furthermore, the absorption of iron from breast milk is greater if breast feeds are given separately from solid food as inhibitors in the food bind the iron from breast milk in unabsorbable complexes. Formula milks in the UK are fortified with iron, and providing infants are taking over 500 ml daily, they are at less risk of iron-deficiency anemia. However, it is still important that iron-rich weaning foods be encouraged.

Zinc This is a component of several mammalian enzymes and is active in many metabolic pathways. Zinc deficiency may limit growth and may be a vulnerable nutrient during the weaning period. In the UK, there is evidence that the average daily intake of zinc is below the reference nutrient intake. Although meat and meat products are the richest sources of zinc, these foods provided only 10% of the dietary zinc intake.

Which Foods Should be Given?

Initial foods should be semisolid and have a smooth consistency and bland taste, and suitable first foods are cereal, such as baby rice or puréed home cooked rice, mashed potato, custard, plain (unsweetened) yogurt and puréed fruit, and vegetables. This should then progress to puréed meat, pulses, and a wider variety of cereals. In many parts of the world, the first food to be introduced is traditionally a cereal. In western European cultures, a commercially prepared infant cereal containing added vitamins and minerals is commonly the first food. In developing countries, the local staple cereal is more likely to be given, and this may include cereals such as corn, barley, oatmeal, millet, sweet potato, cassava, rice, plantain, yam, arrowroot, potato, or sago. These are mixed to a porridge or paste with milk or water.

Between the age of 6 and 12 months, the texture of foods given should gradually progress from semisolid to soft family foods. From the age of 6 months, minced or mashed food with soft finger foods can be encouraged. Meat should be coarsely puréed; wholemeal cereals or instant oat cereal can be given in addition to well-cooked egg. Suitable finger foods include soft fruits, e.g., banana and melon, cheese fingers, soft cooked carrot and fingers of toast, chappatti, and pitta bread.

From the age of 10 months, foods such as meat should be minced or finely chopped. Cooked vegetables need only be chopped, and pasta dishes with small shaped pasta can be given. Finger foods are particularly important at every meal to encourage the infant to participate in self-feeding.

Commercial Versus Home-made Family Weaning Foods

Homemade family weaning foods The use of appropriately home-made foods is encouraged, and in the UK, about 40% of infants are given some home-made weaning foods. They have several advantages: (1) they accustom an infant to the taste of adult-based food; (2) they are cheap and easy to prepare; (3) wastage is low; and (4) meat-based Halal weaning foods can be prepared for Muslim infants. Unfortunately, analysis of home-made weaning foods from the UK has revealed that some meals contain low amounts of energy, iron, and zinc but high amounts of nonstarch polysaccharides, protein, and sodium. Intakes of iron are also lower in infants receiving mainly family foods. Typical home-made foods given to a group of 6- to 8-month-old infants from a deprived inner city area in the UK were mashed potato, tinned spaghetti, and baked beans. However, breast or infant formula milk or cereal can be added to home-made weaning foods to the improve energy and nutrient density.

Commercial baby foods In the developed world, commercial weaning foods (tins, jars, and packet foods) are popular, especially cereals, ready meals, and desserts. They may be fortified with extra vitamins and minerals, and contain no added salt. Even with commercial baby foods, there are some cultural differences concerning manufactured foods given to infants. The Italians prefer individual constituents, i.e., a meat dish to which pasta can be added, whereas the British prefer composite sweet and savory dishes.

The Codex Alimentarius Commission (WHO/FAO) has set compositional standards for baby foods. In 1996, the European Commission issued a directive on processed cereal-based foods and baby foods for infants and young children (96/5/EC); it was amended in 1999. The regulations prescribe manufacturing, compositional and labelling requirements

and prohibit the sale of any processed cereal-based food or baby food, unless it complies with the regulations. The UK government has also published an expert report on additives in food specially prepared for infants and young children, which recommended that additives should only be used when there is a strong justification on technological grounds.

Although not essential, most mothers in developed countries use commercially produced weaning foods at some stage because they are so convenient. In the early stages of weaning, dried products can be made up in small quantities, so avoiding waste. However, commercial weaning foods are relatively expensive, there is a limited range of vegetarian foods, and there is a tendency for some parents to keep infants on strained varieties for prolonged periods of time, which may contribute to the delayed acceptance of more lumpy food.

Foods Avoided During Weaning

Salt

In the West, it is widely recommended that salt should not be added either to weaning foods during the cooking process or to the food at the table, as the developing kidney is unable to excrete large sodium intakes. Excessive sodium intake in young infants is associated with hypernatremia. In addition, there is some concern that a high salt intake in infancy and childhood may cause hypertension in later life; the evidence for this is unconvincing.

Sugar

It is recommended in North America and Europe that sugar should be kept to a minimum and should not be added to foods during preparation. The preference for sweet foods and drinks has been demonstrated even in preterm infants. To avoid encouraging the taste for sweet foods and to reduce the risk of dental caries, sugar should be avoided as much as possible. Even so, in the UK, it is common to introduce sugar-based foods early in the diet. In the MAFF survey looking at infant feeding practices between 6 and 12 months, 62% had been given sweets and chocolate, 70% received biscuits, and 45% of mothers gave milk-based dishes sweetened with sugar.

Gluten

There are many disputes about the best time to introduce gluten into the diet. Delayed introduction is associated with an increased age of onset, although there are those who maintain that it decreases the incidence. Challacombe *et al.* (1997) suggest that later introduction of gluten, increased use of gluten-free foods for weaning, and increased incidence of breast-feeding are associated with a decreasing incidence of celiac disease and transient gluten intolerance. However, in Sweden, a doubling of the gluten content of commercial baby foods was associated with an increased incidence of celiac disease from 1.7 children per 100 live births during 1978–1982 to 3.5 per 1000 thereafter. This occurred in spite of an increased prevalence and duration of breast-feeding and the delay in the introduction of dietary gluten from 4 to 6 months. Gluten-sensitive individuals who are breast-fed and who are introduced to gluten relatively late may not present clinically during weaning, although the disorder is still thought to manifest at an older age, and the cumulative prevalence is unaffected. Therefore, to prevent celiac disease developing at least during the early weaning period, when the infant is particularly vulnerable nutritionally, it is recommended by most health professionals that cereals given to infants under 6 months should be gluten-free.

Food Allergens

Infants are most vulnerable to the initiation of food allergy in the first months of life, and the risk of food allergy is greatly increased by family history of atopic disease such as eczema and asthma. When there is a strong family history of allergy, mothers should be encouraged to breast-feed for 6 months or longer. Weaning should be particularly discouraged before 4 months, and foods traditionally regarded as allergenic (e.g., cows' milk, soya protein, gluten, eggs, fish, and nuts) should not be introduced until 6 months of age at the earliest. In the UK, a working group for the Committee on Toxicity of Chemicals in Food, Consumer Products, and the Environment recommended that peanuts or peanut products should be avoided until 3 years of age in all children with a parent or sibling with an atopic disease.

Suitable Milks During Weaning

Breast milk This should be encouraged, but from 6 months of age, breast milk cannot supply the requirements for several micronutrients, including iron, copper, zinc, and vitamin D. Dietary sources should be given either as part of the weaning diet or as specific supplements, e.g., vitamin D supplements.

Infant formula Providing it is iron-fortified, it should be continued until the age of 1 year. It is a particularly good source of vitamin D, and infant formula sold in the UK is also rich in iron. Providing that a volume of 500 ml daily is taken, there should be no need to give extra vitamin supplements.

Follow-on formula This is intended for older infants over the age of 6 months. It is based on modified cows' milk. It contains less protein, calcium, and phosphorus than cows' milk but more than standard infant formula. Most follow-on formula contains almost double the iron and 45% more vitamin C than standard infant formula. It can provide a useful drink for older infants if breast-feeding has been stopped or the volume of standard infant formula is less than 400 ml daily. However, follow-on formula should not replace breast milk or standard infant formula if the volume of either is adequate.

Whole cow's milk This is not suitable for infants under 1 year as the main drink. It is low in iron, vitamin C, and vitamin D. It may be used in small quantities in the preparation of solid foods.

Goat's and sheep's milks These are low in iron, vitamin D, and folic acid and should not be given to infants under 1 year of age. Although these milks may be perceived as less allergenic or else providing special nourishment, none of these claims have been substantiated.

Problems Associated with Weaning

Iron-deficiency Anemia

Iron-deficiency anemia in later infancy and the toddler years is common throughout the world, and simple dietary iron deficiency is the most common cause. In the UK, it is particularly frequent in ethnic and socio-economically deprived toddlers, and 4–28% of older infants have been reported to have iron-deficiency anemia. It can cause psychomotor delay, apathy, irritability, reduced attention span, pallor, decreased resistance to infection, reduced exercise capacity, and poor appetite. It is associated with prolonged use of breast-feeding, early introduction of whole cow's milk and inadequate dietary iron during weaning. There have been many initiatives to try and reduce the prevalence of iron-deficiency anemia in the Western world. These include screening, health education, iron fortification of food and extension of the use of iron-rich infant formula. Although large-scale dietary education has failed to reduce iron-deficiency anemia in the UK (Child *et al*., 1997), programs to provide supplemented food containing iron to target groups have been shown to be successful in the USA. In addition, in the UK, a project extending the use of iron-fortified follow-on formula beyond 1 year significantly reduced the prevalence of iron-deficiency anemia.

Rickets

Prolonged deficiency of vitamin D can result in rickets. The main symptoms include skeletal deformity, delayed dentition, bone pain, muscle weakness, hypocalcemia, misery, and failure to thrive or short stature. Children under 3 years are particularly vulnerable to a poor vitamin D status because of demands resulting from the high rate at which calcium is being laid down in the bone. Although it is uncommon in Europe and North America, new cases are still being reported particularly in northern climates (Gessner *et al*., 1997) and immigrant children (Brundvand and Nordshus, 1996). In the UK, low-serum 25-hydroxycholecalciferol concentrations have been noted in 20% of Bangladeshi, 34% of Pakistani, and 25% of Indian 2-year-old children living in England (Lawson and Thomas, 1999). Low 25-hydroxy vitamin D levels have been noted in the winter in British children not taking vitamin D supplements.

Rickets is commonly reported in some hot climates. In a recent Nigerian study, overt rickets was found in 2.4% of households. Factors in hot climates such as increasing urbanization and the tendency of mothers and their young children to avoid exposure to sunlight, either by remaining indoors or in the shade or by covering the skin, must all contribute to a decreased endogenous synthesis of vitamin D, in combination with a poor vitamin D or calcium intake. In western countries, prolonged breast-feeding by mothers of poor vitamin D status, early introduction of whole cows' milk, poor weaning sources of vitamin D, little exposure to sunlight, and even the use of sunscreen are contributing factors.

Vitamin A Deficiency

In the developing world, vitamin A deficiency is one of the most common causes of blindness in children, particularly in the Middle east, the Indian subcontinent, and South-east Asia. It causes xerophalmia, keratomalacia, and eventually blindness. This is due to reduced infant stores of vitamin A due to maternal deficiency, low vitamin A-containing breast milk, and weaning diets low in preformed retinol from animal food. Vitamin A deficiency is often accompanied by protein-energy malnutrition. Vitamin A supplements, even in the absence of clinical overt deficiency, reduce child mortality.

Weaning Practices in Developing Countries

The weaning infant is potentially at risk in developing countries, and many nutritional problems arise with

the introduction of solids. The need to introduce nonmilk food sources with only the availability of rather crude preservation processes, poor hygiene, sanitation, and inadequate water supplies introduces the risks of gastrointestinal and parasitic infection because of the heavy contamination of foodstuffs with infecting organisms. In rural Bangladesh, a significant association has been found between the frequency of isolation of *Escherichia coli* in the weaning foods and *E. coli* diarrhea in the infants consuming them, and there can be little doubt that food plays an important role in the epidemiology of diarrhea in young children. The effects of infection lead to anorexia and pyrexia with increased metabolic requirements, and this can only serve to exacerbate any effects of a poor nutritional intake. In addition, many infants have been breast-fed for prolonged periods of time without weaning solids, and when they are introduced, it is probably a low-energy-density, cereal-based gruel that may be given only once or twice a day. The continued effects of deteriorating nutritional status and infection may lead to protein-energy malnutrition.

See also: **Anemia (Anaemia)**: Iron-deficiency Anemia; **Atherosclerosis**; **Carotenoids**: Occurrence, Properties, and Determination; **Celiac (Coeliac) Disease**; **Cholecalciferol**: Properties and Determination; **Food Intolerance**: Food Allergies; **Infant Foods**: Weaning Foods; **Iron**: Properties and Determination; **Metabolic Rate**; **Protein**: Requirements; **Retinol**: Properties and Determination; **Rickets and Osteomalacia**; **Zinc**: Properties and Determination; Deficiency

Further Reading

Akpede GO, Omotara BA and Ambe JP (1999) Rickets and deprivation: a Nigerian study. *Journal of the Royal Society of Health* 119: 216–222.

Arneil GC (1967) *Dietary Study of 4365 Scottish Infants: 1965. Scottish Health Service Studies No. 6*. Edinburgh: HMSO.

Brunvand L and Nordshus T (1996) Nutritional rickets – an old disease with new relevance. *Nordisk Medicin* 111: 219–221.

Cavell B, Stenhammar L, Ascher H *et al.* (1992) Increasing incidence of childhood coeliac disease in Sweden. Results of a national study. *Acta Paediatrica* 81: 589–592.

Challacombe DN, Mecrow IK, Elliott K, Clarke FJ and Wheeler EE (1997) Changing infant feeding practices and declining incidence of coeliac disease in West Somerset. *Archives of Disease in Childhood* 77: 206–209.

Childs F, Aukett MA, Darbyshire P, Illett S and Livera LN (1997) Does nutritional education work in preventing iron deficiency in the inner city? *Archives of Disease in Childhood* 76: 114–147.

Commission Directive on the Processed Cereal Based Foods and Baby Foods for Infants and Young Children (1996) *96/5/EC. Amendment Regulations 1999*. London: HMSO.

Committee on Toxicity of Chemicals in Food, Consumer Products and the Environment (1998) *Peanut Allergy*. Department of Health.

Daly A, MacDonald A, Auckett A *et al.* (1996) Prevention of anaemia in inner city toddlers by an iron supplemented cows' milk formula. *Archives of Disease in Childhood* 75: 9–16.

Daly A, MacDonald A and Booth IW (1998) Diet and disadvantage: observations on infant feeding from an inner city. *Human Nutrition & Dietetics* 11: 381–389.

Davies PS, Bates CJ, Cole TJ, Prentice A and Clarke PC (1999) Vitamin D: seasonal and regional differences in preschool children in Great Britain. *European Journal of Clinical Nutrition* 53: 195–198.

Department of Health (1984) *Weaning and the Weaning Diet*. Report on Health and Social Subjects No. 45. London: HMSO.

Department of Health (1991) *Dietary Reference Values for Food Energy and Nutrients*. Report on Health and Social Subjects No. 41. London: HMSO.

Ehrhardt P (1986) Iron deficiency in young Bradford children from different ethnic groups. *British Medical Journal* 292: 90–93.

Fergusson DM and Horwood LJ (1994) Early solid food diet and eczema in childhood: a 10-year longitudinal study. *Pediatric Allergy and Immunology 5* (Supplement 6): 44–47.

Fomon SJ (1993) *Nutrition of Normal Infants*. St Louis, MO: Mosby.

Food and Agriculture Organisation (1988) *Requirements of Vitamin A, Iron, Folate and Vitamin B_{12}*. Report of Joint FAO/WHO Expert Consultation (FAO Food and Nutrition Series: 23) Rome: Food and Agriculture Organization.

Foster K, Lader D and Cheesbrough S (1997) *Infant Feeding 1995*. Office of National Statistics. London: The Stationery Office.

Gessner BD, deSchweinitz E, Petersen KM and Lewandowski C (1997) Nutritional rickets among breast-fed black and Alaska Native children. *Alaska Medicine* 39: 72–74.

Golden BE (1999) Infancy, childhood and adolescence. In: Garrow JS, James WPT and Ralph (eds) *Human Nutrition and Dietetics*, pp. 449–464. London: Churchill Livingstone.

Hevada AR and Newman DR (1992) Weaning: historical perspectives, practical recommendations and current controversies. *Current Problems in Pediatrics* 22: 223–240.

Kaper BP, Romness MJ and Urbanek PJ (2000) Nutritional rickets: report of four cases diagnosed at orthopaedic evaluation. *American Journal of Orthopsychiatry* 29: 214–218.

Lawson M and Thomas M (1999) Vitamin D concentrations in Asian children aged 2 years living in England: population survey. *British Medical Journal* 318: 28.

Mills A and Tyler H (1992) *Food and Nutrient Intakes of British Infants 6–12 Months*. London: HMSO.

Ministry of Agriculture, Fisheries and Food (1992) Report on the Review of the use of Food Additives in Foods specially prepared for Infants and Young Children FdAC/REP/12. London: HMSO.

Morton RE, Nysenbaum A and Price K (1988) Iron status in the first year of life. *Journal of Pediatrics, Gastroenterology and Nutrition* 7: 707–712.

Mughal MZ, Salama H, Laing I and Mawer EB (1999) Florid rickets associated with prolonged vitamin D supplementation. *British Medical Journal* 318: 39–40.

Pugliese MT, Blumberg DL, Hludzinski J and Kay S (1998) Nutritional rickets in suburbia. *Journal of the American College of Nutrition* 17: 637–641.

Stordy BJ, Redfern A and Morgan JB (1995) Healthy eating for infants – mothers' actions. *Acta Paediatrica* 84: 733–741.

Thomas M and Avery V (1997) *Infant Feeding in Asian Families*. London: Office of National Statistics.

Wharton B (1996) Weaning: pathophysiology, practice and policy. In: Walker WA and Watkins JB (eds) *Nutrition in Paediatrics*, pp. 423–435. Hamilton: BC Decker.

Wharton B (1997) *Nutrition in Infancy*. British Nutrition Foundation Briefing Paper, BNF.

Wilson AC, Forsyth JS, Greene SA *et al*. (1998) Relation of infant diet to childhood health: seven year follow up of cohort of children in Dundee infant feeding study. *British Medical Journal* 316: 21–25.

Zlotkin S (1999) Vitamin D concentrations in Asian children living in England. Limited vitamin D intake and use of sunscreens may lead to rickets. *British Medical Journal* 318: 1417.

Feeding Problems

E M E Poskitt, Department of Public Health Nutrition, London School of Hygiene and Tropical Medicine, UK

Introduction

Nutrition is of great importance in infancy. Infants are entirely dependent on others for their nutrition and carers may have difficulty recognizing their needs. This dependence and infants' high nutritional requirements for growth and development make malnutrition a relatively common problem. Prolonged malnutrition can impair both somatic growth and brain development.

Failure to Thrive

Failure to thrive (**Figure 1**) is failure to gain in height and weight at the expected rates (determined from recognized growth data). The causes are numerous (**Table 1**) but if the precipitating causes can be successfully treated, failure to thrive is followed by accelerated or 'catch-up' growth (**Figure 1a**), provided extra food meets the nutritional needs of accelerated growth. Without extra nutrients, catch-up growth is likely to be incomplete. Affected infants may then resume normal growth rates, but along lower growth centiles than previously (**Figure 1b**).

Some children with growth below the third centile or ≈ -2SD weight-for-age or below (outside the recognized 'normal' range) may be labeled as failure-to-thrive although growing at more or less normal rates for age. They may have been low-weight for gestational age at birth and always small (**Figure 1c**). These constitutionally small children may show delayed bone age, delayed adolescence, and sometimes late catch-up to more normal adult size.

Careful dietary histories should indicate which infants are failing to thrive because of inadequate intakes, inappropriately constituted feeds, or inappropriate weaning practices. Average milk or formula intakes are 150 ml kg^{-1} day^{-1} (441 kJ kg^{-1} day^{-1}) for full-term normal infants. Around 6–8 weeks of age intakes often stabilize despite continuing

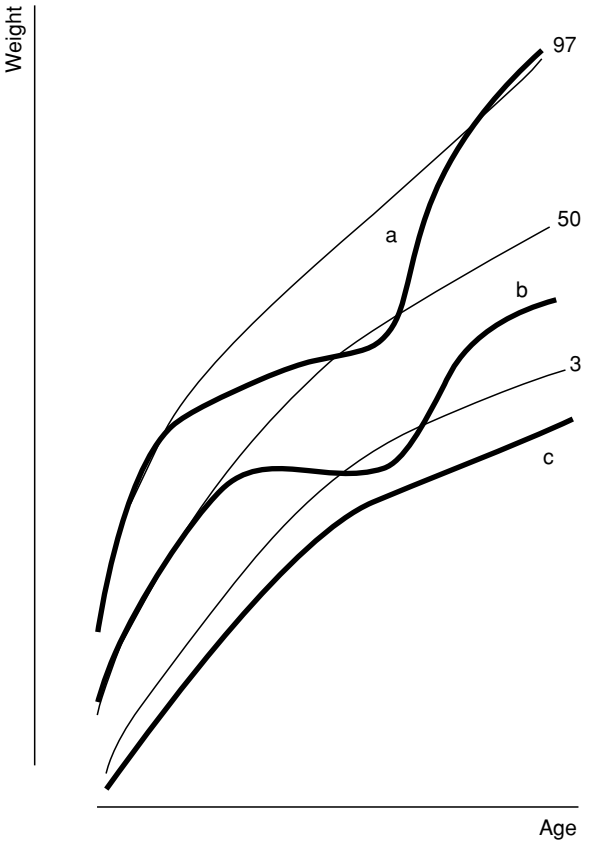

Figure 1 Weight of children with (a) failure to thrive and complete catch-up growth; (b) failure to thrive and failure of complete catch-up growth; and (c) constitutional small size plotted against the 3rd, 50th, and 97th weight-for-age centiles.

Table 1 Classification of nutritional problems associated with failure to thrive

Basic problem	*Examples*
Inadequate intake	Poverty; famine; ignorance of nutrient needs
	Vomiting: all causes
	Feeding problems, including neurological damage
	Anorexia
Increased losses	Protein-losing enteropathy
	Glycosuria; proteinuria
	Exfoliative skin conditions
	Suppuration
Failure to absorb food	Cystic fibrosis
	Gluten-sensitive enteropathy
	Giardiasis
	Lactose intolerance
Failure to utilize food	Inborn errors of metabolism
	Endocrine problems
	Micronutrient deficiencies, e.g., zinc
	Chronic renal failure
	Emotional deprivation
Increased requirements	Small-for-dates and premature infants
	Catch-up growth
	Congenital heart disease: left-to-right shunt

increases in weight and intakes $kg^{-1}\,day^{-1}$ thus decrease.

It may be difficult to assess the adequacy of milk intake in breast-fed infants. Normal growth velocities and general contentment after feeding suggest intakes are adequate. If there is concern, weighing infants in clean diapers prior to breast-feeding and then weighing them again after feeding without diaper change provides some indication of the weight, and thus the volume, of milk consumed. This test weighing must be for all feeds over 24 h since infants may consume large feeds early in the morning and in the evening, but relatively small amounts at other times of day.

When breast milk output seems insufficient for an infant's needs, improving maternal nutrition and increasing the frequency of suckling may improve milk secretion. Current recommendations are that infants should be exclusively breast-fed until at least 4 months of age and preferably until 6 months of age. Usually breast-feeding failure arises because of inappropriate feeding practices (e.g., infrequent suckling, especially in early lactation, or incorrect positioning and failure by the infant to latch on). Advice on feeding technique may resolve the problems. However, mothers who are sick or very malnourished may sometimes fail to produce sufficient milk to support normal infant growth. The decision whether to continue struggling to increase breast milk production, to introduce infant formula feeds, or to introduce weaning foods early is determined by individual circumstances such as the age of the child and cause of breast-feeding failure. Failure to thrive without evidence of nutritional inadequacy, or secondary to persistent anorexia and vomiting, requires urgent medical investigation.

Occasionally failure to thrive occurs without evidence of clinical disease or nutritional inadequacy. Children with such nonorganic failure to thrive may eat well, sometimes excessively, but grow poorly. Typically they are apathetic, showing little social interaction, although some are restless and inattentive. Developmental delay, particularly in social and cognitive skills, is common. These children come from home environments of social and/or emotional deprivation. Environmental stress and lack of love inhibit normal growth. Changing the home environment without necessarily changing the diet can induce accelerated growth giving a dramatically improved effect – confirming the diagnosis.

Vomiting

Most normal infants vomit occasionally and some vomit frequently. Probably the most common cause of persistent vomiting in infancy is laxity of the gastroesophageal sphincter. The sphincter is the spiral muscular area, 2–3 cm in length, at the lower end of the esophagus. Sphincter competence is related to the length of intraabdominal sphincter (which increases with age) and to intragastric pressure. Competence tends to be poor at birth but improves after the first 2 weeks of life, regardless of gestational age. The fluid diets of young infants and the high volumes of liquid needed to meet nutritional requirements produce gastric distension, high intraabdominal tension, and easy regurgitation. Thickening feeds to increase resistance to regurgitation, feeding smaller feeds more frequently, and raising the head end of the cot help reduce gastric distension and regurgitation. As infants grow, consume more solid food and adopt an upright posture, reflux ceases to be a problem.

Feeding problems, including gastroesophageal reflux, are common in children with neurological handicap. Failure to thrive can be severe and may improve with intragastric feeding through gastrostomy, but this exacerbates gastroesophageal reflux. Surgical fundoplication is usually performed, prior to gastrostomy, to control reflux.

Vomiting can be a symptom of serious gastrointestinal or parenteral disorder. When infants vomit and fail to thrive, vomit bile or blood, or vomit to the extent of causing dehydration, medical investigation is required urgently.

Diarrhea

Diarrhea is an increase in total stool output, usually from increase in water content, volume, and frequency of stools. The color, consistency, and frequency of stools vary widely both between different infants and in the same infant on different occasions. Exclusively breast-fed infants tend to have frequent watery yellow stools. Formula-fed infants have less frequent, greenish, drier-looking stools.

In infancy diarrhea is commonly part of either a specific gastrointestinal or a generalized (usually viral) infection. Infectious diarrhea has an acute onset but is usually short-lived and associated with other signs of infection: pyrexia, vomiting, anorexia, and general malaise. Weight loss from anorexia and dehydration are common. Weight recovery is generally rapid and complete. Acute diarrhea has little effect on long-term weight-for-age and even less impact on long-term height-for-age. However, chronic diarrhea and significant growth faltering are common with some gastrointestinal infections (especially rotavirus, *Giardia lamblia*, and *Entamoeba histolytica*).

Malnutrition causes profound changes to gastrointestinal mucosal structure and function. This predisposes malnourished children to severe diarrheal infection and high mortality from diarrheal illness.

All diarrhea should be managed with extra fluids, usually in the form of oral rehydration solution (ORS). ORS suitable for most infant diarrhea has lower sodium and chloride content than that intended for rehydration in adults.

Diarrhea which is persistent and associated with failure to thrive, dehydration, bleeding or other obvious illness, requires medical investigation.

Toddler diarrhea

Some young children persistently pass watery stools or have bouts of loose watery stools alternating with episodes of constipation without evidence of infection, malnutrition, or other ill health. The diarrhea stools are typically voluminous and watery with bits of undigested food in them, indicating the visible bean husks, tomato skins, and other vegetable matter present, although less obviously so, in normally formed stools. Children with toddler diarrhea are active and thriving but their parents are usually very concerned.

Toddler diarrhea has no recognized specific pathology and probably has many causes. Whole-gut transit time is reduced but there is no evidence of abnormal duodenal or jejunal secretion, nor of malabsorption. Minor infection may set off bouts of diarrhea and in some children there is a family history of response to stress with diarrhea. Poor hygiene and low-grade infection associated with constant pacifier (dummy) sucking and prolonged bottle-feeding may contribute to the problem.

Many affected children have high intakes of sweetened fruit juices and/or low intakes of dietary fat. High fluid intakes with high simple carbohydrate content encourage rapid intestinal transit which overwhelms small intestinal function, so sugars reach the large intestine unabsorbed. These and their fermentation products in the colon stimulate osmotic diarrhea. Reducing fluid intakes, particularly of sweetened juices and carbonated drinks, reduces the symptoms in some toddlers. Changing to more adult diets with higher concentrations of complex carbohydrates and fats can also reduce diarrhea. Lipids in the ileum slow duodenal and jejunal transit time (the ileal brake) and may modulate mucosal contact time and the absorption of intestinal luminal contents. Graduating to cup-feeding usually results in less excessive fluid intakes and, particularly if pacifiers are avoided, encourages more hygienic feeding practices.

Toddler diarrhea usually resolves as children move on to more mature diets and develop continence. It is unusual after the age of 6 years.

Micronutrient Deficiencies

Significant micronutrient deficiencies, other than for iron, are rare in normal-weight, full-term healthy infants in developed countries. Vitamin D deficiency rickets can be a problem in low-birth-weight (LBW) infants and noncaucasian infants. In developing countries low levels of vitamin A are associated with high morbidity and mortality from diarrheal diseases, respiratory infections, and measles. Iodine deficiency is also common. Many countries now have supplementation programs to prevent vitamin A and iodine deficiency.

Iron Deficiency

The human fetus has high hemoglobin levels to accommodate low oxygen tension *in utero*. After birth, in response to the high oxygen tension in the extrauterine environment, the bone marrow becomes relatively quiescent and hemoglobin levels decline. Eventually (usually at 4–6 weeks of age) the bone marrow becomes more active again but by then infants have grown and their blood volumes have expanded. Hemoglobin levels remain below adult levels, at 11–12 g dl^{-1}, until late childhood.

Where infants are small and growing very rapidly (e.g., premature and small-for-dates infants), with initially low total blood volume and low total body iron, the need for iron for new blood formation may soon exceed the iron stored, so anemia results. Once body weight has doubled, all infants are at risk of iron deficiency since, without good external sources of

Table 2 Classification of problems predisposing to iron deficiency in childhood

Basic nutritional problem	*Example*
Inadequate intake	Early introduction to cows' milk
	Prolonged exclusive breast-feeding
	Poor quality of weaning diet
	Strict vegetarian diet
Increased losses	Bleeding disorders
	Reflux esophagitis
Failure to absorb	Celiac disease
	Pyrexia
	Iron in ferric form
	High-phytate diet
Failure to utilize	Chronic infection
	Hypothyroidism
	Riboflavin deficiency
Increased requirements	Rapid growth: low-birth-weight infants
	Catch-up growth

iron, iron stores of the body will be distributed over twice the blood volume, and hemoglobin levels will have halved. However, iron supplementation should probably be avoided in exclusively breast-fed infants unless they show clinical evidence of deficiency, since lactoferrin is a major immunological component of breast milk. It binds iron in the small intestine, speeding absorption and inhibiting multiplication of iron-dependent bacterial pathogens. Iron fortification of infant formulas reduces the prevalence of severe iron deficiency in nonbreast-fed infants.

Iron sufficiency is important. **Table 2** lists some factors predisposing to iron-deficiency anemia. The frequency of iron-deficiency problems is an indication of the small margin of safety between requirements for iron and the quantity of iron young children are likely to ingest and absorb from their diets. Low-grade iron deficiency can impair growth rates, intellectual function, and the immunological response to infection. In countries where malaria is common, anemia resulting from iron deficiency puts infants at great risk of life-threatening severe malarial anemia.

Vitamin D Deficiency: Rickets

Rickets used to be common in weanlings. The Clean Air Act (1956) probably did as much to reduce rickets in the UK as any medical intervention, since vitamin D metabolites are predominantly derived from irradiation of 7-dehydrocholesterol in the skin by sunlight of wavelength around 300 nm.

Rickets is still seen in the UK in newborn Asian infants and mothers. The mothers have both inadequate absorption of dietary vitamin D and inadequate skin synthesis of endogenous vitamin D because of indoor habits and clothing cover when out of doors. Low calcium absorption due to low dietary intake and the high phytate content of traditional flours exacerbates the need for vitamin D to optimize calcium absorption from the small intestine. Skin pigmentation increases the need for sunlight exposure when levels of ultraviolet light of appropriate wavelength are low (e.g., winter in northern latitudes).

Infants born to vitamin D-deficient mothers may present with hypocalcemic tetany and convulsions soon after birth or with rickets in the weaning period. The incidence of rickets in Asian children is diminishing in the UK as Asian women in this country adopt lifestyles more suited to northern latitudes and their need for vitamin D supplementation is better recognized. However, signs of vitamin D deficiency in mothers and children are widely reported from the Middle East where there is plenty of sunshine of vitamin D synthesizing wavelength, but little social opportunity for women and children to expose their bodies to the sunlight.

Vegetarian Diets in Infancy

Infants weaned on to vegetarian diets may become deficient in a variety of nutrients. If diets include eggs and milk, it is not difficult to achieve dietary requirements, particularly if breast milk or infant formula is continued as a milk source beyond the first year of life. Milk usually provides a substantial proportion of riboflavin, calcium, and total energy intake in weanling children. Infant formula can provide a fairly absorbable dietary source of iron which may otherwise be deficient.

Vegetarian diets tend to be of low energy density because of high-fiber and low-fat content. Children should be fed frequently (at least four times a day) to enable them to consume the volumes of food necessary to meet nutrient needs. No vegetable protein contains the full range of essential amino acids so diets should include a variety of protein sources (e.g., beans, lentils, cereals, dark green leaves, soya and homogenized nut products).

It is difficult to achieve adequate nutrition in very young children with vegan, macrobiotic, and other strict vegetarian diets. These diets are not recommended for infants unless there is close pediatric dietetic supervision.

Low-birth-weight Infants

The problems of both premature and LBW for gestational age (small-for-dates or SFD) infants are similar, although the greater immaturity of the premature infants' gastrointestinal tracts and of other organs makes the nutrition of the former more difficult than that of the latter. In both groups there is a need for more nutrients per unit of body weight to cope

with the rapid growth to parallel intrauterine growth in premature infants and catch-up growth in SFD infants. The small stomach volumes of LBW infants restrict the volumes which can be fed at any one time without causing apnea, vomiting, or necrotizing enterocolitis. Respiratory problems in premature infants may also restrict the amount of enteral food tolerated. Thus there is a conflict between increased (proportionate to body weight) nutrient needs, especially for protein, calcium, phosphate, fluid, and energy, and the amounts of food that can be administered safely. Even when LBW infants are fed parenterally there are restrictions on the fluid volumes, and thus the quantity of nutrients, that can be given safely. The high demands for growth and development, particularly for the brain, lead to deficiencies of some nutrients which are usually synthesized by the body. This critical nutrient status applies to the long-chain polyunsaturated fatty acids, docosahexaenoic and arachidonic acids, and possibly also carnitine, nucleotides, and other nutrients. For this reason breast milk, which has high levels of these nutrients, may have advantages over infant formulas for the nutrition of LBW infants, if supplemented with energy to meet the high requirements of these infants. However, manufacturers, aware of the physiological needs of LBW infants, are gradually developing formulas which aim to meet these special needs.

Table 3 lists the problems of LBW infants which are discussed elsewhere. LBW infants should be supplemented at least with vitamins A, C, D, and folate over the first year of life. Iron supplementation in the neonatal period risks encouraging bacterial infection and damage by free iron radicals, since iron stores are replete following temporary bone marrow inactivity associated with high hemoglobin levels at birth. However, iron supplementation is advisable in formula-fed infants after the first 4 weeks. Iron supplements should probably not be given at all to wholly breast-fed infants unless they show hematological evidence of iron deficiency, since free iron in the gastrointestinal tract can counteract the protective effects of lactoferrin.

Table 3 Nutritional problems of low-birth-weight infants

Problem	*Precipitating factors*
Immediate hypoglycemia	Limited liver glycogen Delay in mobilizing fat stores Problems with supplying nutrition
Immediate hypocalcemia	Loss of high calcium flux across placenta Immaturity of calcium homeostasis Occasionally maternal vitamin D deficiency
Later poor growth	Low energy density of breast milk Poor tolerance of other energy/nutrient-dense feeds Poor fat absorption by immature infants Increased needs due to bronchopulmonary dysplasia
Bone disease of prematurity	Substrate deficiency, especially phosphate Renal calcium loss from acidosis and/or diuretics Vitamin D, protein, or copper deficiency
Anemia	Vitamin E deficiency increasing early hemolysis Folate deficiency Iron deficiency

Allergy

Cows' Milk Protein Intolerance (CMPI)

Food allergies, particularly CMPI, are frequently diagnosed in infants. True CMPI affects between 1.5% and 5% of the population of western Europe. In families with a strong atopic history the incidence may be much higher. Cows' milk protein intolerance most commonly develops in children under 2 months of age, although it may develop at a later age after gastrointestinal surgery, malnutrition, or gastrointestinal infection.

Children with diarrhea and/or blood loss in association with CMPI may show villous atrophy and increased inflammatory cell infiltrations on jejunal biopsy, or acute colitis on sigmoidoscopy. Those with respiratory symptoms may present with pulmonary hemosiderosis which resolves on removal of all cows' milk-containing products from the diet. However, for most children, manifestations of CMPI are varied and nonspecific.

Infants who vomit, have diarrhea, have a rash, or cry a lot, are liable to be diagnosed as CMPI. They are put on to commercial cows' milk-free formulas and may seem to improve. Improvements may have been coincidental or as a result of mothers feeling reassured that 'something was being done.'

Commercial cows' milk protein-free formulas are also lactose-free. Excluding formula and cows' milk products from the diet does not distinguish between CMPI and lactose intolerance. Children with lactose intolerance commonly have a preceding history of gastrointestinal infection, surgery, or malnutrition. They have loose stools, weight loss, and abdominal distension. Stools are acid and contain reducing substances. Thus, if infants' stools show $pH > 5.1$ and have no reducing substances in them and symptoms resolve on cows' milk protein-free diet but recur when milk protein is reintroduced, CMPI is likely.

Where the child's symptoms have been dramatic (e.g., angioneurotic edema; profuse life-threatening

diarrhea), challenging the diagnosis of CMPI by reintroduction of milk and milk products can be dangerous. Children should only be challenged with milk under hospital supervision and even then with great care.

The particular elements of the cows' milk protein causing allergenic responses vary. At least 18 components that can induce antibody formation have been identified. When purified protein fractions are used as challenges, β-lactoglobulin levels and casein produce symptoms in over 60% of affected individuals. Antibodies to cows' milk protein are widespread and neither their presence, nor their levels in the blood, correlate well with reactions to cows' milk.

Breast-fed infants often show circulating cows' milk protein antibodies. Changing from breast-feeding to cows' milk formula feeding can, very rarely, precipitate dramatic anaphylactic symptoms of angioneurotic edema with acute and very dangerous swelling of tissues around the neck and airway. Relevant proteins from the cows' milk consumed by the mother are thought to be transferred to affected infants either across the placenta or in breast milk, causing sensitization. The infants are then desensitized by the small amounts of antigen in maternal milk but are overwhelmed when confronted with a large load of antigen from formula feeding.

The management of CMPI involves removal of all cows' milk protein-containing products from the diet. Pediatric dietetic advice is advisable since milk and milk products continue to contribute significant amounts of energy and other nutrients to normal children's diets, even after the start of mixed feeding. Difficulties in meeting nutritional needs on milk-free weaning diets can lead to undernutrition. Soya-based infant formulas can be helpful but infants with CMPI may also develop soya protein intolerance. Hydrolyzed cows' milk protein formulas are more suitable for treatment of CMPI but are rather unpalatable (to mothers anyway), not available in supermarkets in the UK, and costly.

Tolerance to cows' milk protein usually develops as children move into the second or third year of life.

Eczema

The extent to which food intolerance is relevant to the development of infantile eczema remains controversial. Some infants with atopic eczema show significant improvement in their skin problem when all milk products are withdrawn from the diet. Others only improve when put on very restricted lists of foods. Studies of infants with familial high risk of atopic eczema suggest that eczema may be prevented in some infants by exclusive breast-feeding for the first 6 months of life. Moreover, exclusively breast-fed infants who do develop eczema may have become sensitized to foreign proteins transferred through the breast milk.

Other Food Allergens

Other food allergies are quite common in infancy, especially in those with an atopic family history. Common precipitants are egg protein, citrus fruits, nuts (not advised for young children anyway because of the risks of inhalation), and the food colorants tartrazine and quinoline yellow. Diarrhea, vomiting, rashes, wheezing and – rarely – behavioral disturbances may occur. (Most behavioral problems are disciplinary rather than allergic.) The pattern of response in each child is usually repeated on subsequent contact with the allergen but there is great variation in the pattern of individual responses.

Colic

Around 3 months of age healthy infants often develop bouts of fractiousness and inconsolable crying, usually in the early evening. Their abdomens become tight and distended, probably because of the air swallowed whilst crying. Infants tend to flex their legs when crying and this may be interpreted as abdominal pain. 'Three-month colic' is an ill-defined problem with a wide range of symptomatology. In the long term, colic resolves without specific treatment. It may be a behavioral rather than a medical problem. Nevertheless, it causes great distress to the child and parents. Removal of cows' milk formula and substitution with nonallergenic formula or soya protein formula has sometimes been associated with improvement. However it is difficult to know if this has anything to do with resolution of symptoms in a benign self-limiting condition.

Cot Death: Sudden Infant Death Syndrome

Sudden unexpected infant death (SUID) affects about 1 in 500 infants in the UK. The condition is of mixed etiology, since detailed history and autopsy reveal cause of death or evidence of illness in many of the infants. There remains a group of SUID infants for whom detailed investigation shows no adequate explanation for death. These are the cases of true 'sudden infant death syndrome' (SIDS). In the past, anaphylactic shock from acute CMPI was sometimes cited as possible etiology for SIDS since SUID is more common in formula- than in breast-fed infants. Some infants have postmortem evidence of milk/formula in the respiratory tract, perhaps as a terminal event. However, true SIDS is probably not more common in formula-fed

than breast-fed infants. It now seems likely that a major component of death in SIDS relates to impaired cardiorespiratory arousal. Putting infants to sleep supine, rather than prone or on their sides, has reduced SIDS by about 35% in many countries.

Another nutritional explanation proffered for some forms of SUID is undiagnosed inborn errors of metabolism involving inhibition of breakdown of long-chain fatty acids. Children with inborn errors of this kind are unable to metabolize fats to cope with the stress of fasting and die of hypoglycemia. Postmortem examination reveals acute fatty degeneration of the liver. In such very rare cases, SUID may be familial since inborn errors of metabolism involving fatty acid breakdwon are inherited, usually by autosomal recessive genes.

Teething Problems

Teeth are erupting more or less continuously between 6 and 30 months. Teething children may show excessive salivation, gingival swelling, and general irritability. These symptoms may exacerbate existing illness but there is no evidence that teething causes fever or diarrhea. Rubbing the gums with cool cloths or giving infants firm objects to bite may reduce teething discomfort.

Most children develop their first teeth around 6 months of age but some do not develop teeth until around 2 years. Late tooth erupters usually cope well with chewing and should be encouraged to progress in the weaning process in the normal way.

Nursing Bottle Syndrome: Rampant Caries in Infancy

Nursing bottle syndrome is the rampant carious destruction of the upper incisors in young children. Although associated with offering bottles of juice/milk or pacifiers sweetened with sugary solutions to infants when they are ready to sleep, rampant caries in this age group probably also reflects underlying enamel hypoplasia in the primary teeth which is secondary to prematurity or to maternal nutritional deficiencies in pregnancy.

Constant bathing of upper incisors with sugary fluids, together with reduced salivary secretion and reduced mouth movements at night, encourages the development of plaque. *Streptococcus mutans*, which survives well at low pH, collects under plaque and adheres to tooth enamel. Enamel is eroded, allowing opportunity for further tooth infection by other bacteria. Excessive vomiting, by creating an acid environment in the mouth, may contribute to the risk. Teeth in the front of the upper jaw are reduced to carious discolored stumps. Children may become faddy eaters due to the discomfort of chewing meat and biting vegetables and fruit.

Fluoridation of water supplies with one part per million of fluoride protects against dental caries. If water is not fluoridated, children can be supplemented with fluoride 1.25 mg day^{-1} in infancy and 1.5 mg day^{-1} between 1 and 3 years. Regular, at least daily, tooth brushing should be taught from the first appearance of teeth. Adults should supervise loading children's toothbrushes with small pea-sized lumps of toothpaste since small children may eat toothpaste rather than spit it out after cleaning their teeth. As much as 1.66 mg of fluoride could be consumed by small children brushing their teeth inexpertly twice a day with fluoridated toothpaste.

All food, and particularly sweetened drinks, should be discouraged after tooth brushing at night.

See also: **Anemia (Anaemia)**: Iron-deficiency Anemia; **Cholecalciferol**: Properties and Determination, **Dental Disease**: Etiology of Dental Caries; **Food Intolerance**: Types; **Growth and Development**; **Inborn Errors of Metabolism**: Overview; **Malnutrition**: The Problem of Malnutrition; **Preterm Infants – Nutritional Requirements and Management**; **Tocopherols**: Properties and Determination; **Vegetarian Diets**

Further Reading

Department of Health (1994) *Weaning and the Weaning Diet. Report on Health and Social Subjects*, no. 45. London: Her Majesty's Stationery Office.

Hunt CE (2000) Sudden infant death syndrome. In: Behrman RE, Kliegman RM and Jenson HB (eds) *Nelson's Textbook of Pediatrics*, 16th edn, pp. 2139–2143. Philadelphia: WB Saunders.

Filer LJ, Jr. (ed) (1989) *Dietary Iron. Birth to Two Years*. New York: Raven Press.

Johnson D and Tinanoff N (2000) Dental caries. In: Behrman RE, Kliegman RM and Jenson HB (eds) *Nelson's Textbook of Pediatrics*, 16th edn, pp. 1114–1116. Philadelphia: WB Saunders.

Simeon DT and Grantham-McGregor SM (1991) Nutritional deficiencies and children's behaviour and mental development. *Nutrition Reviews* 3: 1–24.

Skuse D, Reilly S and Wolfe D (1994) Psychosocial adversity and growth during infancy. *European Journal of Clinical Nutrition* 48 (suppl. 1): S113–S130.

Standing Committee on Nutrition, British Paediatric Association (1988) Vegetarian weaning. *Archives of Disease in Childhood* 63: 470–478.

Stevenson RD and Allaire JH (1991) The development of normal feeding and swallowing. *Pediatric Clinics of North America* 38: 1439–1453.

Walker-Smith J and Murch S (1999) Toddler diarrhoea. In: *Diseases of the Small Intestine in Childhood*, 4th edn, pp. 364–366. Oxford: Isis Medical Media.

Widdowson EM (1947) Mental contentment and physical growth. *Lancet* i: 1316–1318.

INFECTION, FEVER, AND NUTRITION

C J Field, University of Alberta, Edmonton, Alberta, Canada

Background

Infection is a broad term that encompasses a number of pathological situations initiated by invasion of infectious organisms. An effort to reduce the number of infectious incidences is a primary focus for those who work in public health but also for those who work in the clinical setting. Infections are responsible for worsening the nutritional status of patients, doubling the length of hospital stay, increasing the number of medical and surgical procedures required, delaying the provision of adjuvant therapies, and significantly reducing patients' quality of life. This article will briefly review the immune response/role in infection and fever, the importance of nutrition to the immune system, and the role of nutrition and specific nutrients on the response to fever and infection.

The Immune Response to Infection

The immune system (cells and molecules responsible for immunity) is defined as part of the host's defense against destructive forces from outside the body (i.e., bacteria, viruses, and parasites) and those from within, such as malignant and autoreactive cells. Appropriate interactions between the different cells of the immune system are necessary to eliminate infectious organisms (Table 1). The purpose is to disadvantage and destroy invading organisms, while protecting healthy tissue from the damaging influences of the compounds produced during the response. Cytokines and other compounds play a central role orchestrating the immune response to infectious organisms. However, the proinflammatory cytokines and other molecules produced during the inflammatory response also pose a danger to the host. The host response (cytokines and acute phase response) designed to fight infection is partially responsible for a serious and life-threatening metabolic condition called sepsis. Sepsis is characterized by an unregulated host inflammatory response and the presence of bacteria (though primary invasion or secondary to other infections) or their products in the bloodstream.

The Innate Immune System and Infection

The innate immune system comprises defenses that are present early in life that function without depending on previous exposure to a particular pathogen, thereby providing the first line of defense that protects the host during the 4–5 days it takes for lymphocytes in the acquired immune system to become activated. The major components of the innate immunity are illustrated in Figure 1. The resistance to bacterial invasions is thought to depend primarily on nonT-cell branches of immunity such as physical barriers (skin and mucous membranes),

Table 1 Immune responses to some common infectious agents

Infection type	*Examples*	*Immune response*
Viruses	HIV, influenza, rhinovirus	• Humoral immune system blocks initial infection and recruits other immune cells • Cell-mediated response is most effective at destroying cells once infected with virus
Parasites	*Trichinella*, Helminths	• Low immune response, as there is no of the infectious agent inside the host's cells • May be some humoral and inflammatory response from B-cells and macrophages/neutrophils near the invasion site
Protozoa	Malaria	• Bloodborne protozoa are destroyed by B-cell antibodies • Once cells are infected, cells are destroyed by cell-mediated functions
Intracellular bacteria	*Listeria monocytogenes, Mycobacterium tuberculosis*	• Cell-mediated, delayed-type hypersensitivity response important in destroying intracellular bacteria • Stimulates a type-2 helper T-cell response (pattern of cytokines)
Extracellular bacteria	*Streptococcus pneumoniae*	• Humoral immune system (located in lymphoid tissues throughout the body) secretes antibodies
Sepsis (systemic inflammatory response system)	Endotoxin, lipopolysaccharides	• Caused by a dysregulated inflammatory response to an infectious agent

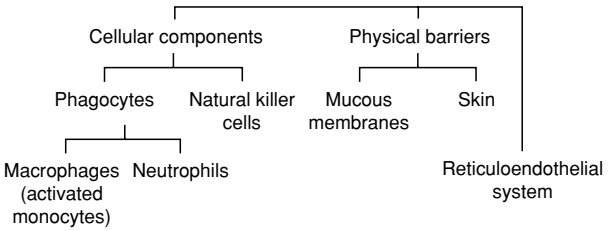

Figure 1 Defenses of the innate immune system.

neutrophils, macrophages, complement, and acute phase proteins. If infectious organisms penetrate this barrier, the innate immune system responds immediately by activating the acquired immune system. If the pathogens are bloodborne, the reticuloendothelial system becomes involved.

The Acquired Immune System and Infection

The acquired (or adaptive) immune response develops over one's lifetime in response to environmental challenges (pathogens/antigens). The cells that comprise the acquired immune defense are illustrated in Figure 2. T- and B-lymphocytes (cells) are the primary cells of this arm of immune system. The T-cell population is divided into two basic subsets CD4 (helper/inducer) and CD8 (cytotoxic/suppressor) cells. T-cells play multiple roles functions including providing help to B-cells, stimulating the inflammatory response, maintaining antigen-specific memory and producing cytokines. Once activated, some T-cells can directly destroy cells (via cytotoxic killing) infected with intracellular pathogens. B-cells produce and secrete antibodies for the elimination of extracellular pathogens and parasites. B-cells can also serve as antigen-presenting cells and activate T-cells.

The Mucosal Associated Lymphoid Tissue

Mucosal tissues (contain components of both the innate and acquired immune system) are strategically located in areas where external pathogens enter the body. Immune cells that reside in mucosal tissues protect against the entry of infectious agents. The immune responses arising in these regions must be very selective so as to avoid mucosal damage that would impair functions such as gas exchange (lungs). The gastrointestinal system serves as a primary reservoir for bacteria that have the potential, if not properly controlled, to induce life-threatening infections. A vast immune system exists in the gastrointestinal tract with immune cells located in intraepithelial and lamina propria regions of the gut and in specialized lymph tissues, Peyer's patches, lining the large and small intestine. The main defense in the gut comes from the large number of activated B-cells present in the lamina propria region that secrete IgA.

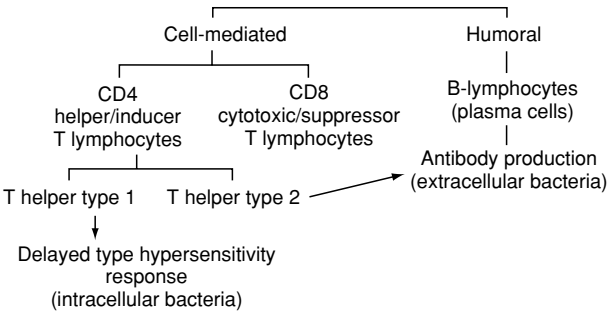

Figure 2 Defenses of the acquired immune system.

Mediators of the Immune System

The response of immune cells to infectious agents is mediated by a number of compounds. Cytokines are glycoproteins that are secreted by immune and other cell types and are a means of communication between cells of the immune system and with other cells in the body. The localized concentration of various cytokines will determine what immune cells are activated. T-helper cells type 1 cells (Th1 CD4+ cells) produce interleukin (IL)-2 and interferon-γ (IFN-γ), which generally promote a cell-mediated inflammatory response. Th2-type CD4+ cells produce IL-4, IL-5, IL-10, and IL-13, which support a humoral antibody response. It has been hypothesized that the balance of these cytokines offers a mechanism by which the immune system can respond appropriately to a specific type of infection.

Other important mediators of immune function include epinephrine, glucagon, and cortisol, which are released in response to stress and infection. These hormones have been shown to have immunosuppressive effects on immune cells. Lipid-derived factors such as eicosanoids, leukotrienes, and platelet-activating factor, made by a variety of cell types in response to cytokines and other signals, also modulate immune function. When tissue is injured during infection, the complement cascade is activated, and fibronectin, vasoactive amines, and chemoattractants are released. The complement cascade is important in mediating inflammation and promoting phagocytosis of microorganisms. Nitric oxide is a potent immunobiological compound produced by macrophages and neutrophils to combat intracellular pathogens (i.e., parasitic fungi, protozoa, helminths, and mycobacteria) and viruses.

Fever

Fever is a host response that is characterized by an elevation of core body temperature. The most

frequent cause of fever in acutely ill patients is infection (by bacteria, viruses, and/or their toxins). In most cases, fever is initiated by the production of cytokines (IL-1, IL-6, tumor necrosis factor (TNF)-α) from macrophages and neutrophils. In general, it is believed that proinflammatory cytokines stimulate the production of arachidonic acid metabolites (PGE_2 and thromboxane A_2), which affect the temperature-regulating areas in the brain (hypothalamus, medulla). Subsequently, the brain signals mechanisms that generate heat and reduce heat loss, thus increasing core body temperature.

It is widely accepted that fever has three distinct phases. The first is known as the chill phase where the individual experiences shivering and feeling cold. The second phase is a plateau phase when the thermostatic set point and body temperature are at the same level. This is followed by a third phase known as 'defervescence' when the fever 'breaks,' resulting in flushing and sweating. This increase in body temperature is important as it increases the speed at which neutrophils move and secrete antibacterial substances, increases B-cell and T-cell replication, and results in a higher production of IFN-γ (antiviral and antibacterial properties). High fever, however, is clearly maladaptive to the host resulting in loss of appetite, increased sleepiness, muscle aches, and hypermetabolism. Resting energy expenditure rises by as much as 10–15% for each degree Celsius increase in body temperature.

Interaction Between Nutrition, Infection and the Immune System

In children and adults with marginal or poor nutritional status, episodes of infectious diseases are more severe and result in more complications. Almost any nutrient deficiency, if sufficiently severe, will impair resistance to infection. The impact of infection on nutritional status depends on the previous nutritional health of the individual, the nature and duration of the infection, and the diet consumed during the recovery period. Thus, infections themselves promote malnutrition (appetite loss and anorexia, malabsorption, and elevated metabolism of energy, and other nutrients), creating a complex cycle of immunosuppression and worsened nutritional status. Worsened nutritional status will then further suppress immunity and predispose the individual to opportunistic infections (Figure 3).

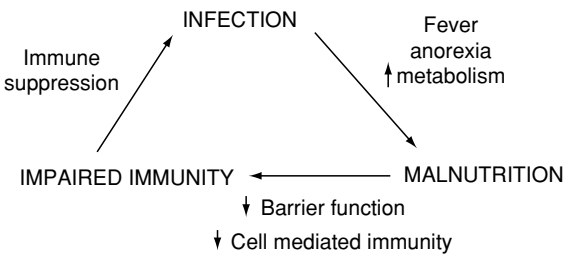

Figure 3 Relationship between malnutrition, immunity, and infection.

The interaction between nutrition and infection is complex. The physiological mechanisms responsible for metabolic changes during infection are not completely understood, although cytokines are clearly involved. Some degree of appetite loss (anorexia) is present during most infections. In some cases, this anorexia is due to nausea and vomiting; in others, gastrointestinal lesions. Additionally, the presence of a fever can result in appetite loss (anorexia) resulting in a 10–40% decrease in dietary intake, not only of protein and energy but also of most nutrients. Anorexia will precipitate clinical nutrient deficiencies of any nutrient in which body stores are limited. The extent of the depletion of nutrient status will subsequently increase the risk of damage to the host's tissues from the inflammatory response. To sustain a hypermetabolic rate (i.e., in fever), there is an acute mobilization of endogenous energy stores (glucose and fat). However, during infections, there is also an impaired ability to use these substrates. If body stores are used to provide for the metabolic needs of infection and fever, weight loss occurs. In fact, this sort of observation led to the introduction of the lay-term 'consumption' to describe tuberculosis, the classic chronic wasting infectious disease. The high prevalence of infections among children living in poor areas of developing countries results in impaired linear growth. In addition, the acute phase response (e.g., proinflamatory cytokine release) to fever directly affects bone remodeling that is required for long bone growth.

Nutrition and the Immune System

The past 25 years have resulted in an explosion of scientific literature defining the role of specific nutrients on immunity. Most host defenses are breached in protein-energy malnutrition, particularly cell-mediated immunity. It is now generally accepted that nutrition is an important determinant of the immune response. It appears that almost every known nutrient deficiency can affect disease resistance if it is sufficiently severe, and this may also be true for many nutrients consumed in excess. Nutritional state and specific nutrients may impact on the immune system directly (e.g., triggering immune cell activation, or altering immune cell interactions) or

indirectly (e.g., changing substrate for DNA synthesis, altering energy metabolism, changing physiological integrity of cells, altering the synthesis and release of cytokines/hormones, or changing the sensitivity of target tissues to various mediators). These changes have widespread effects on lipid, carbohydrate, and protein metabolism.

Knowledge of the importance of nutrition to the functioning of the immune system has led to several practical applications. These include the use of immunological tests: as prognostic indexes in surgical patients, for assessing nutritional status, and for measuring the efficacy and adequacy of nutritional therapy. However, our knowledge of the effects of nutrition on immune function now extends beyond clinical nutrient deficiency, and there is a growing body of literature demonstrating the immune benefits of increasing the intake of specific foods and nutrients.

Nutritional Management of Infectious Disease

Many interventions are possible to reduce the effect of infection on morbidity and mortality. Prevention of disease through sanitation, vector control, promotion of breast-feeding, and vaccination is crucial. Appropriate treatment of infections (e.g., antibiotics) as well as supportive nutritional therapy during and after recovery is also important. However, once the stress response has been activated, and inflammatory mediators released, it is unlikely that nutritional intervention alone can stop this process. It was not until 1980 that the first prospective randomized clinical trial was reported that clearly demonstrated that nutritional intervention could improve patient survival and reduce infectious incidences. Maintaining nutritional status is an essential component of the management of infectious illness. Recent advances in our understanding of the role of specific nutrients in modulating immunity and inflammation have stimulated the development of diets and supplements aimed at improving recovery from trauma and critical illness.

Even when efforts are made to maintain intake and nutritional status during illness, the anorexia associated with infections, even those as mild as those produced with immunization, is a major obstacle. As measured by the occurrence of infectious complications, most clinical studies find that the enteral (oral) route of providing nutrition is superior to the parenteral (intervenous) route. Reasons for this phenomenon are not clear, but it appears that enteral nutrition supports the gut barrier and gut-associated lymphoid tissue. However, when nutrient intake via the enteral route is not possible, total parenteral nutrition (TPN) has become a life-saving therapy to meet the nutritional needs of patients.

In the management of sepsis, the most severe complication of infectious disease, the effects of the resulting hypermetabolism can be at least partly thwarted through delivery of enteral nutrients. The introduction of hyperalimentation, together with a diet high in protein content, has been shown to improve survival by decreasing the number of septic episodes. General considerations for the nutritional management of infection are listed in Table 2 and the need for specific nutrients are discussed below.

Energy

During starvation, there is a specific adaptive response that is designed to preserve lean body mass. However, during the stress response, hormones, cytokines, and inflammatory mediators heavily influence the way in which energy substrates are metabolized. Energy substrates are mobilized to support inflammation, repair of tissues, and maintenance of immune function, usually at the expense of lean body tissues. Lean tissue is broken down to its constituent amino acids, which provide precursors for the synthesis of glucose in the liver (gluconeogenesis).

Resting energy consumption during sepsis is reported to increase as much as 30–40%. Thus, providing additional energy to meet the increased needs during infection and fever is an essential part of nutritional management. The energy needs of adult patients are usually based on estimations of energy

Table 2 General considerations for the nutritional management of infection

- Attempts should be made to prevent chronic or multiple bouts of even mild infections. The cumulative effects of repeated infections result in a greater net energy deficit than one infectious episode
- Encourage breast-feeding, as it is a useful and cost-efficient dietary intervention to prevent infections diseases in small children
- Begin feeding early in acute infection, as it has been shown to have a significant, positive effect on the course of the disease
- Efforts should be made to continue provision of nutrition during mild or moderate episodes of common infections, including diarrhea
- Repletion of nutrients lost during the acute phase of infection will require high intakes of most nutrients for extended periods of time. Above-normal nutrient requirements continue after normalization of the clinical manifestations of infection
- After an infectious episode, the energy density of habitual diets should be assessed, to determine if they can sustain recovery growth after common illnesses. Enrichment or supplementation may be required. Patients with chronic infections and a poor appetite may require supplemental feeding of high-energy formulas, either orally or by tube

expenditure by the Harris–Benedict equation, with the addition of energy losses or increased requirements due to the infection. The metabolic needs of any other underlying disease and the effects of drugs used to treat infection should also be considered in estimating energy requirements. In children, dietary intake during infection and recovery must be sufficient to sustain rates of growth that may be several times higher than in healthy children of the same age (catch-up growth). Most studies have found little correlation between dietary protein intake at adequate ranges and rate of weight gain, while there is a clear association between energy intake and nitrogen accretion.

Protein

Acute infection has a negative impact on nitrogen balance. It has been estimated that approximately 45 g of protein is required daily during a severe infection just to support the increased synthetic demand by immune cells and related tissues. Amino acids from skeletal muscle, skin and possibly bone are released to provide substrate for the synthesis of cells and proteins associated with the response to infection. This biosynthetic response by the host is choreographed by the cytokines, which both stimulate catabolism of protein stores and reduce the availability of amino acids for other processes in the body. The specific effect of infection on muscle protein loss varies depending on the causative agent and the initial nutritional status of the host. In the well-nourished host, mild infection is associated with an increased protein turnover. However, in a more chronic and prolonged infection, the modest rise in protein synthesis, is overcome by the marked increase in the rate of protein breakdown. Additionally, chronic diarrhea, intestinal parasitism, and protein-losing enteropathy can result in additional protein losses.

Most often, the supply of amino acids from the diet does match the increased demand during an infection, and this results in depletion of body stores. The inhibitory effects of infection upon growth, pregnancy, and lactation are well recognized. During infection-induced weight loss, there are reductions in the plasma concentrations of sulfur amino acids, glycine, serine, and taurine. These amino acids are found in high concentrations in many compounds associated with immune and inflammatory responses, most notably cytokines, glutathione, metallothionein, and acute-phase proteins. Early provision of sufficient calories and protein has proven beneficial and is now a common feeding protocol in critically ill patients. However, it has been reported that, apparently, adequate nutritional support in the presence of a severe inflammatory stimulus only attenuates the gluconeogenic process and the breakdown of lean tissue continues.

Fat

Hypertriglyceridemia, resulting from accelerated fat mobilization from stores, decreased clearance, increased synthesis, and/or release from the liver, is characteristic of septic patients. Additionally, studies have shown fatty acids to undergo inefficient cycling between organs, which may also contribute to the elevated metabolic rate. Thus, the energy that, under normal circumstances, can be obtained from a concentrated energy source such as fat is reduced. At present, there are limited clinical data to support that providing additional calories in the form of fat will be beneficial during infections.

Metabolic pathways for *de novo* synthesis of fatty acids may be impaired during stress states, and supplementation of certain fatty acids in the diet may be required. The dietary requirements for specific fatty acids during infection are not known, and this issue is heavily debated. Studies have demonstrated that feeding specific fatty acids can impact on oxidation, metabolism, and the inflammatory response. At present, however, the extent to which fatty acid metabolism, particularly that of the essential fatty acids, is altered during infection and cancer and the ability to improve patient outcome with supplementation of fatty acids in the diet are not known.

Extensive studies carried out, mostly with animal models, feeding different amounts and types of fats have shown that fatty acid content of the diet affects most immune functions. In general, fats rich in *n*-3 polyunsaturated or monounsaturated fatty acids suppress inflammation, and fats rich in *n*-6 polyunsaturated fatty acids exert the opposite effect. The ability of inflammatory cells, such as macrophages, to produce cytokines may also be influenced in a similar manner by feeding different fatty acids. Adding *n*-3 fatty acids to the diet have been shown to suppress the thermic (fever) response to experimental infection. There are a number of levels at which fats may modify host defense and cytokine biology. Most relate to the ability of fats to change the fatty acid composition of membrane phospholipids. Subsequently, changes in membrane lipid composition will change cell signaling, gene expression, and the synthesis and release of inflammatory mediators, which will modify the intensity of the inflammatory response.

Vitamins

Vitamins and minerals play widespread and complex roles in the response to fever and infection. Some of these micronutrients are incorporated into substances

that are released during the response to infection, others are components of antioxidant defense, while others support the response by the host.

Vitamin A The role of vitamin A deficiency in vulnerability to infection was recognized early in the twentieth century. The impact of vitamin A deficiency on increasing the susceptibility to, and severity of, infections is related to its essential functions in metabolism, glycoprotein synthesis, cellular differentiation, and immune function. Vitamin A deficiency is associated with impaired T-cell, B-cell, and neutrophil function and decreased epithelial defense against bacterial invasion. Lower serum levels of both carotene (provitamin A) and vitamin A have been reported during various infections, including hookworm, acute respiratory infections, gastroenteritis, and measles. Intestinal infections reduce the absorption of vitamin A from the diet and can precipitate the clinical appearance of xerophthalmia and keratomalacia in those who have a marginal vitamin A status.

The efficacy of vitamin A supplementation during infections has been examined in randomized, double-blind, placebo-controlled trials of malnourished children in various regions of the developing world. Supplementation with this vitamin is reported to improve the response to measles vaccines, maintain the health and barrier function of the intestine, lower the incidence of respiratory-tract infections, and reduce mortality associated with diarrhea and measles. The World Health Organization currently recommends that vitamin A be given to all individuals who develop measles in developing countries, whether or not they have symptoms of vitamin A deficiency. There are also clinical data suggesting that vitamin A deficiency in HIV-1-infected individuals contributes to increased mortality, disease progression, and a higher rate of maternal–infant disease transfer.

Excessive intakes of vitamin A are potentially toxic in humans. Although there has been little work on the relationship to infection in humans, chronic megadosing with vitamin A increases the susceptibility to infectious organisms, suppressed hematopoiesis, T-cell proliferation, and antigen-specific antibody production in animals.

B vitamins A clinical deficiency of thiamin (beriberi) or niacin (pellagra) is associated with an increased risk of infections. The B vitamins are involved in energy and protein metabolism and are needed for cell division and the formation of both red and white blood cells. Although there are no specific requirements for these nutrients determined during infection, the increase in metabolic rate and immune activity that occurs during an infection would likely increase their requirement above that of a healthy individual.

Vitamin C The essentiality of vitamin C (ascorbic acid) to protection from infections is most clearly illustrated during its clinical deficiency disease (scurvy), where immunosuppression and infections almost always occur. Poor vitamin C status results in a loss of integrity of epithelia and impaired bactericidal capacity of neutrophils. Vitamin C is highly concentrated in leukocytes and is rapidly utilized during infection (e.g., to prevent oxidative damage). Reductions in vitamin C levels in plasma have been reported during even mild infections such as the common cold and after immunization against smallpox and measles. A reduced concentration of this vitamin in leukocytes is associated with reduced immune function.

The immune system is a major producer of reactive oxygen radicals and is particularly sensitive to oxidative stress. The actions of vitamin C as a reducing agent and oxygen-radical quencher are well established. If these oxidants are not rapidly neutralized, they can inhibit chemotaxis, phagocytosis, and other antimicrobial activities of immune cells. There has been a long-standing debate concerning the possible function of high doses of vitamin C in 'boosting' immunity to combat infections. Supplementation of this vitamin was recently reported to reduce the incidence of respiratory infections in ultramarathon runners. The results of animal studies and a few human studies have suggested that the function of T-cells and neutrophils and antiviral resistance are increased with large intakes of vitamin C. Since the early interest and advocacy by Linus Pauling, subsequent studies on megadosing with vitamin C have not unequivocally shown that vitamin C can prevent and/or treat upper respiratory-tract infections. Some experts suggest that positive effects of vitamin C supplementation, in early efficacy studies, were complicated by the use of a mixture of antioxidant nutrients, not controlling for physical stress or studying subjects with low overall nutritional intakes.

Unlike many other dietary antioxidants, even when consumed at very high levels ($5000\,mg\,day^{-1}$), vitamin C appears to be safe. However, one might still remain cautious, as some *in vitro* experiments demonstrate that very high levels of vitamin C suppress the T-cell and neutrophil response to infectious agents.

Vitamin E Vitamin E-deficient rats demonstrate an enhanced inflammatory response to endotoxins and addition of the vitamin to their diet suppressed this effect. Vitamin E supplementation has been shown to increase T-cell response, reduce the production of

some inflammatory/stress cytokines, and improve the ability of macrophages to destroy bacteria. Although the data are still limited, there is evidence that vitamin E may play roles beyond basic nutrition function to aid in host defense to infection.

Minerals

Dramatic changes in mineral distribution occur at the outset of infections, with particularly striking shifts of iron and zinc to intracellular compartments. Because iron and zinc are essential for growth and replication of microorganisms, it is possible that the rapid decrease in circulating iron and zinc during infections would limit the supply of these essential minerals for the invading pathogens. The importance of this nutrient as a host strategy to deprive the invading pathogens of the essential divalent cationic factors they need to thrive is controversial, as these nutrients are also essential to the host.

Iron Experimental and clinical data suggest that there is an increased risk of infection in the presence of iron-deficient anemia, although a small number of reports indicate otherwise. Experimental studies in laboratory animals uniformly report reversible deleterious effects of iron deficiency on most measures of functional immunity. It is well documented that iron regulates the function of T-cells, and in most studies (*in vivo* and *in vitro*), a deficiency results in impaired cell-mediated immunity. Humoral immunity may be less affected by iron deficiency than cellular immunity. Iron is a component of the cytotoxic enzymes that are used by neutrophils, macrophages, and natural killer cells. It has been proposed that the immunosurveillance role of macrophages may be mediated in part by modulation of the iron status in cells. This sequestration of iron into macrophages has been proposed to be beneficial during the early acute stages of infectious disease, as it would limit the availability to microorganisms (particularly intracellular microorganisms).

The relationship between iron repletion and infection is somewhat controversial. The major dilemma focuses around the studies demonstrating a positive relationship between iron repletion/supplementation and increased morbidity from infections. Administration of parenteral iron during an infection has been demonstrated to heightened susceptibility to infection. These clinical observations have been supported by studies *in vitro*, in which the rate of microbial growth diminished when iron was removed from culture media. Although animal studies have shown that excess iron promotes the growth of some microorganisms, there is no compelling clinical evidence (except perhaps for tuberculosis, malaria, and brucellosis) that this occurs in adult humans receiving oral iron supplementation. In a recent review, it was concluded that there is little evidence that oral iron supplementation or provision of iron fortified foods to iron deficient individuals inhibits immune function or increases the susceptibility to most infectious agents (with the possible exceptions of malaria-related disease, HIV, and tuberculosis). Thus, the relationship between iron and infection may depend on the ability of the infectious organism to acquire iron and the state of iron nutriture of the host. In conclusion, both iron deficiency and iron excess exert adverse effects on immune responses and alter the metabolism and growth of the pathogens. The balance of these various effects ultimately will determine clinical outcome.

Zinc Zinc is a constituent of key enzymes involved in cell replication (DNA and RNA polymerase). Thus, deficiency will impact on any antimicrobial activity that depends on cellular replication and differentiation. Zinc plays a role in regulating the activation of acute-phase genes (via its ability to bind to RNA finger loop domains known as 'zinc fingers'), such that zinc deficiency impairs translation of genes normally activated during the acute-phase reaction. Another essential role of zinc in the immune response is its binding to certain thymus-derived peptides (thymulin) that appear to function as hormones in the differentiation and maturation of T-cells. In zinc deficiency, the depletion of mature T-cells heightens susceptibility to infection. Thus, clinical zinc deficiency results in thymic involution, lymphoid tissue atrophy, lymphopenida, decreased cell-mediated immunity, and impaired natural killer cell activity. Defects in cell-mediated immunity and increased susceptibility to infections are well described in the human disease acrodermatitis enteropathica, an inherited defect of zinc absorption. Animal studies have confirmed that zinc deficiency impairs cellular immunity, compromises B-cell development and antibody production, inhibits macrophage function, decreases natural killer activity, and increases the mortality to various infectious organisms.

Plasma zinc decreases during infection, even in the zinc-sufficient individual. The hypozincemia of infection is, like iron, part of the acute-phase response mediated by cytokines, which induce the synthesis of the intracellular zinc-binding proteins liver, thymus, and bone marrow. The physiological role of this acute transfer of zinc from the extracellular to the intracellular compartment is not clear but is hypothesized to be an attempt to limit viral and bacterial DNA transcription and RNA translation.

The impairment in immune response associated with zinc deficiency is readily reversed by supplementation.

A number of experimental trials have illustrated the efficacy of zinc supplementation to improve immune function during recovery from infections. For example, zinc supplementation was reported to reduce the duration and severity of cold symptoms. Community-based prospective zinc supplementation of poorly nourished infants and children, and therapeutic trials in acutely ill children, suggest a decrease in diarrhea duration, pneumonia, growth-stunting, acute lower respiratory infections, dysentery altered intestinal permeability and improved food intake, and a more rapid improvement in patients.

Other minerals The diarrhea associated with gastrointestinal infections reduces the absorption and increases the endogenous loss of many of the other trace minerals, notably copper. It is recommended that treatment regimens for diarrhea should include supplementation with copper and other minerals in addition to zinc. Recently, it was found that elderly residents in long-term care facilities taking a micronutrient supplement containing zinc and selenium had up to four times fewer infections than those taking a placebo. In addition to zinc and iron, other trace elements are present in several acute-phase proteins and enzymes associated with antioxidant defense. These proteins include caeruloplasmin (copper), superoxide dismutases (manganese, copper, and zinc), and glutathione peroxidase (selenium). Deficiencies in these micronutrients alter the ability of immune cells to respond to inflammatory agents.

Immunonutrients

An immunonutrient is loosely defined as a nutrient that provides specific benefits to the immune system. It is hypothesized that, under certain physiological conditions (e.g., infection), the demand for the nutrient may exceed the ability of the body to synthesize it. There is growing evidence, from both animal and human studies of infectious diseases, that immunity might be improved by providing immunomodulatory nutrients (e.g., arginine, glutamine, docosahexaneoic acid, eicosapentanoic acid, gamma linolenic acid, zinc, and/or nucleotides). Based on these studies, there are now commercially available enteral and parenteral solutions that contain mixtures of immunomodulatory nutrients. Complete enteral diets containing combinations of immunonutrients (arginine, glutamine, nucleotides, ω-3 fatty acids), when given postoperatively to surgical patients, have been reported to reduce hospital stay, medical treatment costs and the incidence of wound complications and infections. The promising experimental evidence of the metabolic and immune effects of these nutrients offers therapeutic potential for the future, and ultimately might help prevent and reduce the mortality and morbidity associated with infections. Owing to the inclusion of multiply immune-enhancing nutrients in experimental studies, it is difficult, at present, to isolate the specific nutrient or combinations responsible for improvement. Furthermore, one nutrient may work synergistically with others to produce an effect over and above what would be observed with a single nutrient alone. At present, more studies are necessary to find the dosage and duration of the feeding to obtain the desired clinical outcome.

Conclusions

Infectious diseases are on the increase world-wide and are a major cause of morbidity and mortality and contribute to malnutrition. Today, protein energy malnutrition is accepted as a major cause of immunodeficiency world-wide. Infections occur frequently in both the healthy population and with more severe consequences, in individuals who are immunosuppressed through malnutrition. Thus, nutrition plays an important role in the prevention and treatment of infections. There are many levels at which nutrient intake can modify the intensity and characteristics of the response of animals and humans to inflammatory stimuli. Recently, it was demonstrated, that for an RNA virus, the host diet influenced the genetic make-up of the pathogen and thereby alter its virulence. New knowledge on the interaction between nutrition and the host immune function and perhaps also the infectious agent will permit the development of designer feeding formulas/regimens with selective ingredients to optimize immunity to reduce the risk of infection and promote the recovery from infections and fever.

See also: **Elderly**: Nutritionally Related Problems; **Enteral Nutrition**; **Famine, Starvation, and Fasting**; **Fatty Acids**: Dietary Importance; **HIV Disease and Nutrition**; **Hypertension**: Hypertension and Diet; **Malnutrition**: The Problem of Malnutrition; Malnutrition in Developed Countries; **Minerals – Dietary Importance**; **Nutritional Surveillance**: In Industrialized Countries; **Stress and Nutrition**; **Trace Elements**

Further Reading

Alexander JW (1998) Immunonutrition: the role of omega-3 fatty acids. *Nutrition* 14: 627–633.

Bhaskaram P (2001) Immunobiology of mild micronutrient deficiencies. *British Journal of Nutrition* 85 (supplement 2): S75–S80.

Field CJ, Johnson IR and Pratt VC (2000) Glutamine and arginine: immunonutrients for improved health? *Medicine and Science in Sports and Exercise* 32: S377–S388.

Field CJ, Johnson IR and Schley PD (2002) Nutrients and their role in host resistance to infection. *Journal of Leukocyte Biology* 71: 16–32.

Grimble RF (1998) Nutritional modulation of cytokine biology. *Nutrition* 14: 634–640.

James MJ, Gibson RA and Cleland LG (2000) Dietary polyunsaturated fatty acids and inflammatory mediator production. *American Journal of Clinical Nutrition* 71: 343S–348S.

Mainous MR and Dietch EA (1994) Nutrition and infection. *Surgical Clinics of North America* 74: 659–676.

Mizock BA (2000) Metabolic derangements in sepsis and septic shock. *Critical Care Clinics* 16: 319–336.

Schrimshaw NS and SanGiovanni JP (1997) Synergism of nutrition, infection and immunity: an overview. *American Journal of Clinical Nutrition* 66: S464–S477.

Shankar AH and Prasad AS (1998) Zinc and immune function: the biological basis of altered resistance to infection. *American Journal of Clinical Nutrition* 68: 447S–463S.

Stephensen CB (1999) Burden of infection on growth failure. *Journal of Nutrition* 129: 534S–538S.

Woodward B (1998) Protein, calories, and immune defenses. *Nutrition Reviews* 56: S84–S92.

INFLAMMATORY BOWEL DISEASE

A Bousvaros, Children's Hospital Boston, Boston, MA, USA

Background

Inflammatory bowel disease (IBD) is characterized by chronic relapsing intestinal inflammation without an infectious or dietary cause. The term idiopathic IBD is traditionally limited to the illnesses Crohn's disease (CD) and ulcerative colitis (UC), which will be the focus of this chapter. However, there are many other conditions (chronic granulomatous disease, BehÓet's disease, graft- vs. host-disease, HIV) which may cause immune-mediated gastrointestinal inflammation.

CD and UC are two separate illnesses that may share some common genetic background, because a family history of either disease increases an individual's risk of developing both diseases. CD is characterized by a transmural granulomatous intestinal inflammation, which may occur anywhere in the intestinal tract from the mouth to the anus; extraintestinal granulomas in the lymph nodes, skin, and genitalia may also be seen. The most common disease locations include the terminal ileum, cecum, entire colon, and perianal regions. Intestinal strictures, fistulae, and abscess are common in CD but not in UC. Up to 30% of individuals with CD may have microscopic gastroduodenal inflammation. In contrast, ulcerative colitis is characterized by mucosal inflammation limited to the large intestine. UC extends from the rectum proximally. Three major phenotypes of UC have been described: proctitis (limited to the rectum), left-sided disease (distal to the splenic flexure), and pancolitis (extending past the splenic flexure). Pharmacotherapy of both CD and UC includes the use of aminoslicylates, corticosteroids, 6-mercaptopurine, and other immunosuppressive agents. Nutrition may be used as primary therapy in CD but is only palliative in UC.

The precise etiology of these two diseases is unknown. However, it is hypothesized that an underlying genetic mutation involving regulation of the immune system predisposes an individual to the development of CD or UC. When that individual is exposed to a dietary or infectious antigenic trigger for the disease, excessive and aberrant intestinal inflammation results. In support of this hypothesis, mutations in an immunomodulator gene (NOD2) have been identified in a subset of individuals with CD.

Nutritional Complications in IBD

Before determining the extent of evaluation and therapy in a patient with inflammatory bowel disease, the clinician must ask several questions. First, does the patient have CD or UC? Nutritional complications are far more common in patients with CD, though anemia is common in both. Second, where is the location of the disease? A patient with CD and ileal disease is more likely to develop steatorrhea and vitamin B_{12} deficiency than a patient with disease limited to the colon. Third, what is the patient's age? Children and teenagers with chronic malnutrition from Crohn's disease are more likely to present with slowed growth and short stature, whereas adults present with weight loss. Finally, how active is the patient's disease? A patient with IBD and multiple relapses is at greater risk for nutritional complications than a patient who has been in remission for

Table 1 Nutritional complications of inflammatory bowel disease

Global
Inappropriate caloric restriction – anorexia
Weight loss
Growth failure (in children)
Steatorrhea (in Crohn's disease)
Protein-losing enteropathy
Cholestasis (in primary sclerosing cholangitis)
Osteopenia/low bone density
Common micronutrient deficiencies
Iron-deficiency anemia
Hypoalbuminemia
Vitamin B_{12} deficiency (in ileal Crohn's disease)
Folate deficiency (especially in patients taking sulfasalazine)
Hypovitaminosis (vitamins A, D, and E)
Zinc and selenium deficiencies

years. Nutritional complications of IBD are summarized in (Table 1).

Caloric Deficiency and Growth Failure

Acute severe exacerbations of CD and UC may cause a sudden cessation of food intake, with weight loss and dehydration. More commonly, however, low-grade disease activity causes subtle changes in appetite and growth, especially in CD. Adults with CD have a significantly lower percent body fat (as determined by bioelectrical impedance) compared with patients with UC and controls. In children, growth failure (decreased height velocity leading to short stature) results primarily from inadequate caloric intake. Approximately 50% of children with CD and 10% of children with UC demonstrate decreased height velocity in childhood and adolescence, with accompanying pubertal delay. Growth failure may precede intestinal symptoms of CD in childhood. The etiology of the decreased intake is multifactorial, and may include gastritis and esophagitis, cytokine-mediated anorexia, fear of eating secondary to intestinal inflammation, and decreased taste sensation secondary to micronutient deficiencies (e.g., zinc deficiency). Chronic corticosteroid usage, whether with prednisone or newer agents such as budesonide, may also impair linear growth. While levels of growth hormone and IGF-1 may be low in children, the levels rise with nutritional restitution.

Patients with CD of the small bowel may experience fat malabsorption and steatorrhea, because terminal ileal inflammation leads to decreased enterohepatic circulation of bile acids. Steatorrhea is less common and less severe than in patients with pancreatic insufficiency.

In an anthropometric study of 24 children with moderately active CD, requiring hospitalization, researchers identified a depletion of lean body mass and expansion of the extracellular water compartment. The mean resting energy expenditure (REE), as estimated by indirect calorimetry, was 1250 kcal per day, which was significantly higher than patients with anorexia nervosa and similar degrees of malnutrition. The authors hypothesized that malnourished patients with CD may have a 'lack of REE adaptation' mediated by tumor necrosis factor and interleukin-6.

Osteoporosis

Patients with IBD, especially those treated with long-term corticosteroids, are at increased risk for pathologic fractures, including vertebral compression fractures. The bone mineral density (BMD) of the lumbar spine, as assessed by quantitative computed tomography and dual energy X-ray absorptiometry, demonstrates that approximately 30% of patients with IBD have a BMD *Z*-score of less than -1, compared with approximately 15% of age-matched controls. Patients with CD are at a higher risk of osteopenia than UC patients. Reports suggest that up to 70% of children may have a low BMD, but these reports must be interpreted cautiously because of the lack of well-established pediatric standards. The etiologies for osteopenia in IBD are multifactorial and include hypovitaminosis D, decreased calcium intake, calcium malabsorption from steatorrhea, corticosteroid effects, and cytokine-medicated bone resorption. In addition, primary sclerosing cholangitis, an immune-mediated disease causing bile duct strictures and cholestasis, occurs in 1–4% of IBD patients and may result in additional impairment of vitamin D and calcium absorption.

Anemia and Iron Deficiency

Anemia in IBD is multifactorial and may reflect iron deficiency, folate or vitamin B_{12} deficiency, gastrointestinal blood loss, impaired utilization of iron, or myelosuppression caused by medications (e.g., sulfasalazine, 6-mercaptopurine). The most common cause of anemia in IBD is iron deficiency secondary to impaired iron absorption and intermittent GI bleeding. Laboratory features include microcytosis, serum iron levels $< 30\,\mu g\,dl^{-1}$ and ferritin levels $< 12\,ng\,ml^{-1}$. Iron indices may be misleading, because the chronic inflammation may cause a decrease in the total iron binding capacity and an elevation of serum ferritin.

Micronutrient Deficiencies

While a laboratory studies of micronutrient deficiencies are commonly reported in patients with CD and UC, clinical manifestations of such deficiencies are fortunately quite rare and largely limited to case

reports. Vitamin B_{12} deficiency complicating Crohn's ileitis or ileal resection is the most clinically important vitamin deficiency. In patients with extensive small bowel CD, vitamin A deficiency causing impaired dark adaptation and night blindness may occur. Zinc deficiency causing dermatitis has also been reported. Studies of IBD patients both with active disease and in remission have demonstrated reduced plasma levels of vitamins A, C, D, and E, as well as selenium, magnesium, and zinc. The risk of micronutrient deficiency may be increased with greater disease activity. Folic acid deficiency may occur in individuals taking sulfasalazine. In addition, some of the newer immunomodulatory agents used to treat severe UC or CD (e.g., cyclosporine or tacrolimus) may cause hypocalcemia or hypomagnesemia.

Other Nutritional Complications

CD patients with ileitis or ileal resection may develop hyperoxaluria and renal stones. Increased gut oxalate absorption occurs because the luminal bile acid deficiency results in steatorrhea, and the increased intraluminal fat binds calcium, so there is no free calcium to bind oxalate. High levels of homocysteine have been reported in CD patients from countries where there is no supplementation of grains with folate.

Nutritional Therapy

Nutritional Therapy of the Acute IBD Patient

Hospitalized patients with IBD are typically severely ill. A patient with a severe exacerbation of ulcerative colitis may be passing 15–20 bowel movements per day and have severe abdominal cramping. Eating and drinking may exacerbate the pain, so patients tend not to eat. Thus, while there is no evidence that use of parenteral nutrition helps heal the colitis, bowel rest and parenteral nutrition will help forestall weight loss and may decrease pain. Thus, nutritional support of these patients is a useful adjunct to medical therapy (intravenous corticosteroids and/or cyclosporine). If no response to medical therapy occurs within approximately 14 days, the patient should be considered for colectomy.

Hospitalized patients with CD present with small bowel or colonic mucosal inflammation, abdominal abscess, or small bowel stricture. While ill patients commonly require treatment with corticosteroids, antibiotics, and (in the case of abscess and stricture) surgery, enteral feeding and/or parenteral nutrition is once again a useful adjunct. Hospitalized CD patients who are less ill, whose disease is primarily limited to the small bowel, whose principal symptoms are weight loss and diarrhea, and who do not have stricture or abscess may be candidates for parenternal nutrition as primary therapy (see below).

Enteral Nutrition as Primary Therapy for CD

Multiple randomized controlled trials demonstrate that in patients with active CD, nutritional therapy may induce a remission. Such therapy involves cessation of eating food and exclusive intake of elemental and/or polymeric formula by mouth, nasogastric tube, or gastrostomy. The typical elemental formula has a density of $1\,kcal\,ml^{-1}$, and provides 15% of calories as free amino acids, 25% as medium-chain triglyceride, canola or soybean oil, and 63% maltodextrin or starch. Polymeric formula has a density of $1–1.5\,kcal\,ml^{-1}$, contains similar fat and carbohydrates, but has sodium and calcium caseinate instead of free amino acids.

While the exact mechanism by which formulas induce remission in CD is unknown, postulated mechanisms include decreased dietary antigenic load and modification of the intestinal bacterial flora. In a metaanalysis of eight studies conducted prior to 1995, the remission rate ranged from 50–80% after 4 weeks of therapy. However, it was concluded that corticosteroids were more effective than enteral nutrition at bringing about remission. Trials comparing elemental with polymeric formula as a primary therapy have not demonstrated any clear benefit of one over the other. It is therefore reasonable to use a polymeric formula first, given its lower cost and greater palatability. While liquid diets have the advantage of bringing about remission without exposing patients to corticosteroid side-effects, the disease may relapse within 1–2 months once patients return to a regular diet. In addition, the passage of the nasogastric tube may be uncomfortable for children, and children who go to school with an indwelling nasogastric tube may be stigmatized socially. Enteral nutrition as a primary therapy is commonly utilized in Europe, but less so in the USA.

Enteral Therapy for Growth Failure in Children

Since growth failure and short stature in children and adolescents with CD are principally caused by decreased caloric intake, nocturnal supplementation with formulas increases the height velocity and improves growth. The advantage of nocturnal supplementation is that a child can eat normally during the day, and the nasogastric tube can be removed before a child goes to school. In a typical adolescent or adult, 1000–1500 calories of an elemental or polymeric formula are delivered overnight at a rate of $100–150\,ml\,h^{-1}$. In a controlled study of children aged 8–15 years, nocturnal nastogastric infusion of elemental

formula for 1 out of every 4 months for 1 year resulted in a mean weight gain of 6.9 kg and a height gain of 7.0 cm. In addition, it has been suggested that children who respond initially to 4 weeks of exclusive enteral feeding have a higher probability of staying in remission if they continue with nocturnal supplementation after they resume eating.

Home Parenteral Nutrition

While total parenteral nutrition was a commonly utilized therapy for CD patients in the 1970s and 1980s, home parenteral nutrition is infrequently utilized now. Improvements in medical therapy with the introduction of immunosuppressive agents (azathioprine, 6-mercaptopurine, methotrexate, infliximab) have decreased the number of patients with chronically active disease. In addition, surgeons have become more cautious about resecting bowel in patients with CD, so there are fewer patients with short bowel syndrome. Lastly, it is now realized that enteral nutrition is as effective as parenteral nutrition in bringing about remission in most patients with CD. Home parenteral nutrition may have a limited role in the rare patient with CD refractory to other therapies, intestinal stricture, fistulae, or abscess awaiting surgery. Given that many of these patients are commonly maintained on immunosuppressive drugs, the risk of bacterial or fungal sepsis is significant with prolonged PN.

Nutritional Therapy of Nonmalnourished Outpatients

Most adults and children with CD and UC feel well, are not interested in nasogastric tube feeding, and are either of normal weight or minimally underweight. While clinically asymptomatic, they remain at risk for nutritional complications of IBD and desire dietary counseling. There is no specific diet universally recommended for patients with CD and UC. Low-residue, low-fiber diets may be palliative for patients with narrowed regions of bowel (i.e., CD patients with inflammatory or fibrotic strictures). While there are anecdotal reports of elimination diets (e.g., milk-free, low-carbohydrate) maintaining remission in some patients, there is not enough evidence in the literature to encourage patients to attempt these diets. Studies of fish oil capsules suggest some efficacy in the treatment of ileocolonic CD and mild ulcerative colitis. The efficacy of fish oil may be due to the polyunsaturated fatty acids, which may decrease the production of leukotriene B4 and inflammatory mediator seen in UC.

Given the many studies documenting micronutrient and antioxidant deficiencies, patients should be counseled to take a daily multivitamin. Most such vitamins contain 400 μg of folic acid, but in patients receiving sulfasalazine, 1 mg of folate per day is recommended. Given the risk of osteopenia, patients should be counseled to take 1200–1500 mg of total daily calcium, either through their diet or as supplements. Studies of complementary medicine use in IBD suggest that 20–40% of CD and UC patients routinely utilize herbs and/or megavitamins and fail to tell their physician. Thus, the physician should specifically inquire about such use and inform the patient of the potential toxicities of megavitamin therapy. Patients with chronic ileitis and/or ileal resections should be monitored for vitamin B_{12} deficiency and supplemented with oral or intramuscular vitamin B_{12}.

Prevention of IBD

Both CD and UC have strong genetic components. It is estimated that a child born to a parent with CD or UC has a 5% chance of developing IBD over their lifetime. However, the fact that IBD varies so dramatically by region suggests that environmental factors may also confer risk. Environmental factors already identified include smoking and oral contraceptive use, both of which increase the risk of CD. Since the potential exists that food antigens may be a trigger for the onset of IBD, investigators have tried to identify modifiable dietary factors that may increase risk. Unfortunately, because of the low incidence of CD and UC, prospective cohort studies of risk factors are nearly impossible. Case-control studies suggest that patients with UC have higher intakes of sucrose and fat, and a reduced intake of fructose, fruit, and vitamin C. Given the limitations of these studies, no formal recommendation for a preventative diet can be made at this time.

See also: **Anemia (Anaemia)**: Iron-deficiency Anemia; **Carotenoids**: Occurrence, Properties, and Determination; Physiology; **Cobalamins**: Properties and Determination; Physiology; **Enteral Nutrition**; **Folic Acid**: Properties and Determination; Physiology; **Osteoporosis**; **Parenteral Nutrition**; **Zinc**: Deficiency

Further Reading

Belli DC, Seidman E, Bouthillier L *et al.* (1988) Chronic intermittent elemental diet improved growth failure in children with Crohn's disease. *Gastroenterology* 94: 603–610.

Geerling BJ, Badart-Smook A, Stockbrugger RW and Brunner RM (1998) Comprehensive nutritional status in patients with long-standing Crohn's disease currently in remission. *American Journal of Clinical Nutrition* 67: 919–926.

Geerling BJ, Dagnelie PC, Badart-Smook RD, Russel MG, Stockbrugger RW and Brummer RJ (2000) Diet as a risk

factor for the development of ulcerative colitis. *American Journal of Gastroenterology* 95: 1008–1013.
Gokhale R, Favus MJ, Karrison T, Sutton MM, Rich B and Kirschner BS (1998) Bone mineral density assessment in children with inflammatory bowel disease. *Gastroenterology* 114: 902–911.
Griffiths AM, Ohlsson A, Sherman PM and Sutherland LR (1995) Meta-analysis of enteral nutrition as a primary treatment of active Crohn's disease. *Gastroenterology* 108: 1056–1067.
Han PD, Burke A, Baldassano RN, Rombeau JL and Lichtenstein GR (1999) Nutrition and inflammatory bowel disease. *Gastroenterology Clinics of North America* 28: 423–443.
Oliva MM and Lake AM (1996) Nutritional considerations and management of the child with inflammatory bowel disease. *Nutrition* 12: 151–158.
Schreiber S and Webel S (1997) Diagnosis and treatment of anemia in inflammatory bowel disease. *Inflammatory Bowel Diseases* 3: 204–216.
Wilchanski M, Sherman PM, Pencharz P, Davis L, Corey M and Griffiths A (1996) Supplementary enteral nutrition maintains remission in paediatric Crohn's disease. *Gut* 38: 543–548.

Infrared Spectroscopy *See* **Spectroscopy**: Overview; Infrared and Raman; Near-infrared; Fluorescence; Atomic Emission and Absorption; Nuclear Magnetic Resonance; Visible Spectroscopy and Colorimetry

Insecticides *See* **Pesticides and Herbicides**: Types of Pesticide; Types, Uses, and Determination of Herbicides; Residue Determination; Toxicology

INSECT PESTS

Contents

Insects and Related Pests

R Davis, Pest Management Consultants and Associates, Savannah, GA, USA

Background

All animals, including our insect pests and their relatives, are here on earth as a part of the balance needed to maintain the health of the biosphere. This need is often questioned, particularly when we so often find insects and mites infesting our food, feed, seed, and fiber. It sometimes seems that these pests can invade and infest almost any commodity under any storage situation despite our most conscientious efforts to exclude them.

The fossil evidence supports the success of these animals which are recorded as having remained relatively unchanged for at least the last 400 million years. We can point to several characteristics, common to all insects, which contribute to this success. Such attributes as the various types of metamorphosis, ability to fly, a high reproductive potential, the ability to survive for long periods of time without food and water, and tolerance to adverse low temperatures all have survival value. Many insects can and will slow their development as a result of population pressure, lack of food, or adverse temperature. There is now evidence of different life-cycle forms within strains of at least two species of flour beetles (*Tribolium* spp.) that may be of value for the establishing of new infestations. These two forms are ones with six larval instars – the invaders – and one with seven

larval instars – the colonizers. Such a division is of great survival value to a species.

Insects are the most successful animal class. Their great variation in food and other environment requirements is of great survival value. Additionally they have enormous dispersal abilities by wing, wind, and piggybacking on higher animals such as birds, bats, and flightless animals, including humans, and this has led to wide distribution for many species. Today, with the assistance of worldwide commerces, insects have now colonized every continent except Antarctica. Many species which now find human food, feed, seed, and fiber as their food and environment are cosmopolitan.

Classification

In order to explain the success of these groups of pests of humans, it is essential to learn something of their classification and their developmental biology. An appreciation is needed of these insect and mite pests in their proper place among all the other animals of the world. In an effort to arrange the world's animals in an orderly manner, humans have developed a hierarchical classification system of seven primary groupings. Each of these groups may have supra- or sublevels, and in some instances additional primary groups are added to accommodate animal groupings that have a large number of species. For our purposes, simply recognizing the seven classification groups presented in Table 1 will suffice for this discussion.

The individuals in the animal kingdom are distinguished from those in the plant kingdom by their ability to move from place to place and their ability to make food from inorganic materials. While there are some exceptions to these characteristics for the animal kingdom, they will not affect our discussion of the phylum Arthropoda. It is in this phylum that we find our insect and mite pests along with their close relatives. Arthropoda is the animal kingdom's largest phylum, in terms of both number of species and the number of individuals. Six distinct classes of animals are readily recognized in the phylum Arthropoda.

Table 1 Classification of the animal kingdom

Classification	*Examples*	
	Insects	*Mankind*
Kingdom	Animalia	Animalia
Phylum	Arthropoda	Chordata
Class	Insecta (Hexapoda)	Mammalia
Order	Coleoptera	Primata
Family	Tenebrionidae	Hominidae
Genus	*Tribolium*	*Homo*
Species	*confusum*	*sapiens*

Several other classes have been placed in this class from time to time by various specialists, but again the placement of these questionable classes will not affect our discussion.

The major characteristics of the phylum Arthropoda are as follows: (1) the body is segmented; (2) the appendages are paired and segmented; (3) the body possesses an exoskeleton; and (4) the body is bilaterally symmetrical.

By using these four characteristics for the phylum we can develop a simple key that will distinguish for us the main classes of Arthropoda and show us something of their relationships.

1a. Aquatic species (2 go to couplet)
1b. Aerial species, but some may be limited to very moist environments (3 go to couplet)
2a. Marine species without antennae. This class is represented by a single species: *Xiphosura polyphemus*, the horseshoe crab: class Merostomata
2b. Both marine and fresh-water species, possessing two pairs of antennae and at least five pairs of legs, e.g. crayfish, lobsters and shrimps: class Crustacea (in part)
3a. Possess two pair of antennae and live in moist environments, e.g., sowbugs, pillbugs (order Isopoda) and shrimp fleas (order Amphipoda): class Crustacea (in part)
3b. Does not possess two pairs of antennae (4 go to couplet)
4a. Possess no antennae, four pair of legs as adults, and have only two apparent body regions (cephalothorax and abdomen), e.g., spiders, scorpions, ticks, and mites (order Acarina): class Arachnida
4b. Possess one pair of antennae (5 go to couplet)
5a. Possess more than three pairs of legs and no wings (6 go to couplet)
5b. Possess three pairs of legs and usually wings in the adult stage, e.g., beetles, moths, cockroaches, and flies: class Insecta
6a. Possess two pairs of legs on the same body segment, e.g., millipedes: class Diplopoda
6b. Possess one pair of legs on each body segment, e.g., centipedes: class Chilopoda

In the protection of agricultural commodities (food, feed, seed, and fiber) we are primarily interested in the mites, class Arachnida, order Acarina, and the insects, class Insecta, primarily the orders Lepidoptera (moths) and Coleoptera (beetles).

The classes Insecta and Arachnida are very large animal groupings from the standpoint of both numbers and species. While the total number of animals species worldwide is often debated, it can really be stated that the arthropods, principally the insects and mites, represent over 90% of the known

and described species. There are various estimates of this number, but it is safe to say that the known and described species would easily exceed one million. Fortunately, the number of species infesting our stored subsistence and fiber is quite limited. A US Department of Agriculture publication provides an illustrated key listing 650 species of insects and mites known to infest food in the USA. As most insects attacking stored subsistence have traveled with their food around the world, they are considered as cosmopolitan in distribution. A publication of the American Peanut Research and Education Society lists 120 species collected in the south-eastern USA on stored-in-shell peanuts, an excellent host for many insects attacking stored food. However, they only elected to discuss 21 of these species as major pests. Many of the others collected were parasites and predators on the pests, or insects that attack peanuts which are out of condition, e.g., possessed a high moisture content, were moldy, or had an excessive amount of debris mixed with the peanuts.

Mite species attacking stored commodities would likewise be quite limited. Only 16 mite genera have been listed as possessing species attacking stored commodities. Here again most other species found in stored commodities are predators of insects and mites infesting the commodity. These naturally occurring parasites and predators, e.g., members of the mite genus *Pyemotes* (family Pyemotidae), which prey on insects infesting the commodity, are often as serious a problem as the phytophagous species. Members of this family will often attack humans and cause a dermatitis known as 'grocer's itch.' These mites and insects are food contaminates and will be considered as filth by sanitarians and food inspectors. However, it must be realized that the management of phytophagous insects and mites will in turn control these insects and mite contaminates. (*See* **Parasites**: Occurrence and Detection.)

When infestations of birds, rodents, or bats exist within or close to storage facilities, there may occur in and on the commodities, and its packaging, insects and mite parasites, as well as predators of these higher animals. Some of these parasites and predators may also attack humans, and their occurrence will always be considered to be evidence of unsanitary storage conditions.

It is probably safe to say that for most stored subsistence we will find a limited number of major pests among those found infesting the commodity. There will be a slight variation in the species complex for each commodity and for that part of the world where the storage takes place. However, this latter variation will usually be quite small and perhaps go unnoticed because most of these pests are cosmopolitan, resulting from world trade in commodities.

For the convenience of humans, the insects and mites attacking stored commodities have been grouped in nonscientific ways by those responsible for handling the commodities and those responsible for controlling these pests. These groupings are artificial and primarily designed to assist warehouse-workers, pest control operators, sanitarians, and farmers in discussions of their observations. In the protection of stored whole grain, one often speaks of internal feeders, external feeders, scavengers, and secondary and tertiary feeders. If the commodity is a packaged product the terms become penetrators and invaders. The terms long-lived and short-lived will, on occasion, be used in relation to adults. Other terms heard will be grain beetles, flour beetles, grain moths, flour moths, mealworms, warehouse beetles, fabric pests, hide beetles, and many others, depending upon their environment and/or their food habits. Most of these terms are self-explanatory. Several are not and will be defined below and followed by examples.

Pests of Stored Plant Products (Seeds)

Internal Feeders

These insect larvae feed entirely within the seed. Examples of internal feeders include the rice, granary, and maize weevils, *Sitophilus* spp., the Angoumois grain moth, *Sitotroga cerealella* (Olivier), and all the bean and pea weevils.

External Feeders

All stages of these insects feed outside the seed. Examples of external feeders include the lesser grain borer, *Rhyzopertha dominica* (F.) the khapra beetle *Trogoderma granarium* Evert, and the cigarette beetle, *Lasioderma serricorne* (F.)

Scavengers

These insects feed on seeds only after the seed coat is broken. Examples of scavengers include the flour beetles, *Tribolium* spp., flat grain beetles, *Cryptolestes* spp., and grain beetles, *Oryzaephilus* spp.

Secondary Feeders

These insects feed only on products that are deteriorated, damp, and moldy. Examples of secondary feeders include *Tenebrio* and *Alphitobius* spp., mites *Acarus siro* L. and *Tyrophagus* spp., and psocids, *Liposcellis* spp.

Penetrators

These insects have the ability to penetrate many of the types and combinations of paper, films, and foils used

in packaging. Examples of penetrators include the lesser grain borer, cigarette beetle, and many of the moth larvae.

Invaders

These insects usually depend upon openings in the packages resulting from the manufacturing process, or openings made by penetrating insects. Examples of invaders include the flour beetles and the grain beetles.

Pests of Animal Products

Some insects and mites need or prefer a diet higher in protein than they can secure by feeding exclusively on plant-derived foods. These pests feed on dried meats, hair, feathers, dried fish, and on processed animal products such as woolen cloth, dried milk and eggs, fishmeal, cheese, and leather. These pests can be grouped according to the foods eaten. The following are some of the more common and important groupings.

Fabric Pests

Fabric pests feed on keratinous materials such as wool cloth, mohair, and hair, and include such species as the clothes moths, *Tinea* and *Tineola* spp., and the carpet beetles, *Attagenus* and *Anthrenus* spp.

Pests of Hides and Skins

These pests will feed on green hides and skins, and can cause considerable damage and defects. Pests of hides and skins include several species of the genera *Dermestus* and *Trogoderma* spp., the ham beetle, *Necrobia rufipes* (DeGeer), the spider beetles, *Ptinus* spp., flesh flies *Sarcophaga* spp., carpet beetles of the genera *Attagenus* and Anthrenus spp., and the cheese skipper, *Piophila casei* (L.).

Pests of Processed Animal Products

These insects can be represented by almost any of the above species as well as many of the species that commonly feed on processed plant products. Many animal feeds are mixtures of both animal- and plant-derived products.

Pests of Root Crops

In the part of the world where root crops are commonly stored dry, they are attacked by many of the boring beetles that are known to attack dried grain and harvested wood, e.g., the lesser grain borer.

Pests of Perishables

The most common pests of perishables such as fresh fruits, particularly those fruits of tropical and subtropical origins, are the fruit flies of the insect family Trypetidae, e.g., the Mediterranean fruit fly.

Pests of Dried Fruits and Nuts

These commodities have many insect pests, many of which come in from the field at harvest time. They include pests that feed on the commodity and also on the fungus and molds that grow on and in the commodity. Major pests include several moth species, e.g., the almond moth, *Cadra cautella* (Walker), beetle species of the genera *Tribolium* spp., *Oryzaephilus* spp., and *Carpophilus* spp.

Mite Pests

The phytophagous mites attacking and infesting stored commodities must be classified as scavengers. In some instances they have been recorded as attacking sound kernels of grain and consuming the germ. This is questionable, but most probably the grain possessed a high moisture content, and fungi has infested the germ and the kernel. Like insects, mites contaminate much more commodity than they consume – with their feces, odors, cast skins, and dead bodies. As with the insects, scavengers and secondary feeders, contamination by mites suggests that the commodity is out of condition.

Unfavorable temperatures and relative humidities will result in significant deterioration of the commodity for further spread of insect, mite, and fungi infestations.

Developmental Biology

It is important that we understand the biology and life cycles of these pests of our food and fiber. All insect and mite pests develop from eggs. Most young hatch from the egg after the female lays it, but a few are born alive from eggs that hatch inside the mother. Following the hatching, these pests develop in a stepwise fashion by stages. Each stage is limited in size by the space provided inside the exoskeleton. As the pest develops it lays down a new exoskeleton beneath the current one. At specific times, determined by hormonal action, the old exoskeleton splits down the back of the pest and the next stage emerges and expands to its new size limitation. The process of shedding an old exoskeleton and emerging with a new exoskeleton is called moulting. Usually an insect will have four to eight moults but some insects, e.g., members of the family Dermestidiae, can have more than 20 moults. The period of time between moults is called an instar.

During growth, insects change not only in size but also in form. In some instances the only change will be maturity and development of the sexual organs

while in other species the changes will be quite startling. This process of change accompanying growth is called metamorphosis. There are four distinct types of metamorphosis in insects: ametamorphosis, gradual metamorphosis, incomplete metamorphosis, and complete metamorphosis.

In ametamorphosis there is very little change in form as the insect grows, and the only visual changes will be size and sexual maturity. An example of an insect possessing this type of metamorphosis is the silverfish.

In gradual metamorphosis the insect develops from egg to nymph to adult. The nymphs resemble the adult except that they are smaller, not winged or sexually mature. An example of an insect with this type of metamorphosis is the grasshopper.

Incomplete metamorphosis involves much greater body changes than those in gradual metamorphosis. The young have a completely different body structure and habitat from the adults. The young are aquatic and called naiads. The adults are terrestrial and often winged. An example of an insect with this type of metamorphosis is the dragonfly.

In insects with complete metamorphosis we find four distinct life stages: egg, larva, pupa, and adult. The larva is worm-like in appearance, but its activity varies from a rapidly moving individual or free-living larva to legless grubs found inside seeds. The pupal stage, which is a resting stage between the larva and adult, is really a very active time. It is at this time that the larval body is broken down and the adult body constructed. The last stage, the adult, which is often very different from the larva and pupa, has as its main function dispersal, mating, and extending or finding new habitat.

Members of the class Arachnida also possess a form of metamorphosis, but changes are not so extensive as those in insects. In the Acarina (mite) there are usually four life stages: egg, larva, nymph, and adult. The larva has only six legs and there are three nymphal stages. The nymphs appear to resemble the adults, but are not sexually mature.

Some species in both the insects and mites have or can develop forms, during the larval and nymphal stages respectively, that are resistant or tolerant to adverse environmental conditions. These resistant or tolerant stages are referred to as a condition of diapause in insects and as a hypopus in a mite.

Most of the insect pests of food, feed, seed, and fiber have complete metamorphosis. This type of development allows for the larva and adult to have separate ecological niches on the same food or separate habitats for all their life activities. This developmental method has great survival value. The various life stages within the species development usually do not cause any population pressure on each other.

The study of insects and mites and their relationship to the welfare of humans has been both a valued and an interesting pursuit of science. Elucidating their behavior, physiology, and genetics has led and will continue to lead to newer and safer ways to control these animals with a penchant for our food and fiber.

See also: **Insect Pests**: Problems Caused by Insects and Mites; **Parasites**: Occurrence and Detection

Further Reading

Anonymous (1986) *Stored-Grain Insects. Agriculture Handbook no. 500*. Washington, DC: US Department of Agriculture, Agricultural Research Service.

Baker EW and Wharton GW (1952) *An Introduction to Acarology*. New York: MacMillan.

Comstock JH (1950) *An Introduction to Entomology*, 9th edn. New York: Comstock.

Gorham JR (ed.) (1978) *Training Manual for Analytical Entomology in the Food Industry*. FDA Technical Bulletin no. 2, Washington, DC: Food and Drug Administration, US Department of Health, Education and Welfare.

Gorham JR (ed.) (1987) *Insect and Mite Pests in Food: An Illustrated Key*. Agricultural Handbook no. 655. Washington, DC: US Department of Agriculture.

Hughes AM (1961) *The Mites of Stored Foods*. Ministry of Agriculture, Fisheries and Food, Technical Bulletin no. 9. London: Her Majesty's Stationary Office.

Monro JW (1966) *Pests of Stored Products*. The Rentokil Library. London: Hutchinson.

Redlinger LM and Davis R (1982) Insect control in postharvest peanuts. In: Pate E and Young CT (eds) *Peanut Science and Technology*, pp. 520–570. Texas: American Peanut Research and Education Society.

Problems Caused by Insects and Mites

R Davis, Pest Management Consultants and Associates, Savannah, GA, USA

Background

In 1978, contributors to the US National Academy of Sciences report, *Postharvest Food Losses in Developing Countries*, pointed out that to quantify food losses to insects and other postharvest pests was virtually impossible. They set 10% as an average minimum overall loss for cereals and legumes and 20% as the minimum for perishables and fish. Because diets in developing countries contain little, if any, dried or

smoked red meat or poultry, no loss figures were considered.

In 1975, worldwide losses due to insects and rodents were estimated by the Food and Agriculture Organization (FAO) of the United Nations as about 20%. These figures ranged from about 10% in developed countries to 30% in African and Asian countries. However, in the literature, authoritative sources often refer to losses of 50% in grain, with much of the remainder so contaminated or damaged as to be nearly worthless. Whether we will ever be able to quantify the losses correctly may not be as important as we believe. In any of the underdeveloped or developing countries of the world there are instances where pests have ruined entire storages; there is clearly a need for correction.

The effort to secure adequate loss data to estimate worldwide losses requires a study of the continuum from the farm to the consumer. This has proved to be a major logistics problem in even the most developed countries. In addition, social and cultural factors will also limit the methods of preparation and utilization. Politics will affect the storage, distribution, and availability. It requires a concerted effort on the part of all levels of government and the consuming public to insure an adequate supply of wholesome and nutritious food and quality fiber.

The seriousness of the postharvest losses has been recognized by the United Nations. In 1975, the FAO organized a conference on postharvest food losses. This conference stressed the losses in the tropical regions and established a goal of reducing postharvest food losses by 50% by 1985. To achieve this goal it was necessary to develop methods of loss assessment. Neither goal has been reached as yet, but many improvements have been made in protecting harvested foods during their movement from the farm to the ultimate consumer.

Sites of Possible Food Losses

Losses may occur at any point during the interim between the harvest and gathering up to the point of consumption. For the convenience of this discussion this continuum will be broken into the following intervals:

1. Harvest – the separation of, or collection of, the part of the plant to be utilized by humans or the slaughter of animals.
2. Initial processing – the separation of, or extraction of, edible from nonedible, e.g., peeling of fruits and vegetables, removing corn (maize) from the cob, and other activities associated with cleaning, drying, and grading, or cleaning and cutting-up of animal carcasses.
3. Preservation – the prevention of loss and spoilage, e.g., providing proper storage conditions and facilities, drying to inhibit fungus growth, and the use of pesticides to protect against insect attack.
4. Storage – the management of the storage facility, e.g., protecting food from damage by moisture, temperature, and pests through the use of aeration, pesticides, refrigeration, controlled atmospheres, etc.
5. Transportation – the use of all kinds of vehicles and containers to move foods from point of production to the point of final consumption.
6. Processing – the conversion of edible forms of plant and animal products into another form which is more acceptable or more convenient to the consumer, e.g., reducing the fat content in meats, packaging and canning fruits and vegetables, and quick-freezing raw and partially prepared individual foods or even complete meals.

Industries at Special Risk

The loss of a food for any reason is significant to those who suffered the loss. However, when losses are very large and involve significantly large areas, such as regions and countries, the economics and logistics of supplying these areas with replacement food stocks become a major problem which usually results in suffering and starvation. In the underdeveloped countries and, to a somewhat lesser extent, in the developing countries, the most important industry is farming.

The farmers must store sufficient food to feed themselves and their families for about a year, and produce enough surplus first, to sell in order to secure needed staples and second, to meet their seeding needs for the next cropping season. To some extent farmers have to backstop themselves against catastrophic losses by joining in cooperative storages at the village or other local levels. In many countries the national governments have established programs of food storage administered by departments or ministries of food, agriculture, or supply. However, such programs are dependent upon food production within the country, or the availability of currency or credit that will allow for importation in the event of harvest shortfalls.

It is well documented that many pests attacking stored products, particularly the grains and oilseeds, first attack these seeds in the field. Similarly, it is also known that many transportation vehicles and containers often harbor pests left from previous uses of these vehicles and containers. While these insects and mites are, without question, the source of many subsequent infestations that arise in storage and processing facilities, they are rarely, if ever, adequately

controlled. Proper insect-resistant packaging and intransit and static fumigation of ships, barges, railcars, and containers are examples of procedures that have helped to minimize the effect of residual insects and mites in such vehicles. (*See* **Cereals**: Handling of Grain for Storage; **Fumigants**.)

In the continuum of food moving from the farm to the ultimate consumer, there are two primary areas where the industries involved in the interim can be at special or high risk. These two areas are storage and processing. While it is recognized that some firms may well be involved in both storage and processing, the two areas will be discussed separately.

Storage Industries

It is in the industries that provide and maintain storage facilities that those rather large losses due to pests occur. In these facilities raw or partially processed products are stored in bulk, providing a large, readily available food supply for any pest that can gain entry. In the case of insect and mite pests the process of exclusion is often very difficult, or expensive, or both. Fortunately, these pests are poikilothermic (cold-blooded) and cannot adjust their body temperatures from that of the environment. Therefore proper regulation of the temperature and the moisture content of the commodity can be used to prevent any incipient infestations from getting out of hand and causing large amounts of damage. (*See* **Storage Stability**: Parameters Affecting Storage Stability.)

The problems of providing proper storage and educating those responsible for managing such storage have occupied an important place on the agenda of international organizations such as the FAO, the International Atomic Energy Agency (IAEA) and the World Bank, as well as international consortiums such as the Group for Assistance on Systems Relating to Grain After Harvest (GASGA). Placing emphasis on well-managed storage is, of course, a proper approach; logistics, transportation, and processing are rather unimportant if there is nothing to move, distribute, or process. Storage is also the point at which significant losses in both quantity and quality can occur in developing countries, unless trained and experienced personnel are responsible and alert to such potential problems.

Processing Industries

In the industries preparing and modifying foods for increased acceptance and convenience, we rarely see major losses in quantity of food products. However, losses in quality will occur unless those responsible for the actual processing are not constantly alert. The loss in quality will usually come about through an inspection or regulatory procedure, and will usually be in the form of adulteration and contamination. Most developing countries have regulations, supported by an inspection procedure that will establish the quality of food produced as a result of processing. Industry, to protect itself, develops good manufacturing practices that, if followed, meet their country's regulations for clean, safe food.

Therefore it is essential for the maintenance of a safe food industry that correct identifications of insects and their related pests are readily available. Such identifications must include all life stages – eggs, larvae, pupae, and adults – found in both raw and processed food. In fact, this need for rapid and accurate identifications extends to fragments of insects found in processed foods.

In the USA, this regulation of foods is found in the Food and Drug Act (FDA). Regarding insects, mites, and other pests, this law says that a food is adulterated if it has been prepared, packed, or held under conditions whereby it may have become contaminated. The food inspected does not have to contain pests. The FDA inspector noting excessive evidence of insects will try to find the infested product, but such a finding is not required for regulatory action. (*See* **Legislation**: Contaminants and Adulterants.)

Processing industries may also confront another regulatory process in many countries, that of plant and animal quarantine. These laws are not promulgated to protect humans and their animals from adulterated foods but rather to prevent the importation of pests that will affect the nation's agriculture. However, these laws may impact upon the availability of a product scheduled to be processed.

Problems and Control of Beetles and Weevils

Beetles and weevils are most often thought of as pests of grains and legume seeds. This is apparently due to the large quantities of these seeds stored and our heavy dependence upon them. However, they are also major pests of roots and tubers, dried fruits and nuts, and dried meats and other animal products. In Table 1 is a listing of the major beetle and weevil pests of these stored products.

The control of beetles and weevils has traditionally been to maintain a low and uniform temperature and moisture content throughout. Most insect and mite pests have quite specific optimum temperatures for activity and reproduction, and a lower level with regard to moisture content of the product for survival. This optimum temperature is between 21 and 35 °C, with minimum moisture content of 10–10.5%.

Table 1 Beetle and weevil pests of stored products

Pests of grain and legume seeds and their processed products	
Khapra beetle (*Trogoderma granarium*)	Siamese grain weevil (*Lophocateres pusillus*)
Larger grain borer (*Prostephenus truncatus*)	Rusty grain beetle (*Cryptolestes ferrugineus*)
Lesser grain borer (*Rhizopertha dominica*)	Foreign grain beetle (*Ahasverus advena*)
Cigarette beetle (*Lasioderma serricorne*)	Longheaded flour beetle (*Latheticus oryzae*)
Sawtoothed grain beetle (*Oryzaephilus surinamensis*)	Broadhorned flour beetle (*Gnathocerus cornutus*)
Merchant grain beetle (*Oryzaephilus mecator*)	Drugstore beetle (*Stegobium paniceum*)
Confused flour beetle (*Tribolium confusum*)	Bean weevil (*Acanthoscelides obtectus*)
Red flour beetle (*Tribolium castaneum*)	Cowpea weevils (*Callosobruchus* spp.)
Granary weevil (*Sitophilus granarius*)	Pea and bean weevils (*Bruchus* spp.)
Rice weevil (*S. oryzae*)	Groundnut beetle (*Caryedon serratus*)
Maize weevil (*S. zeamais*)	Coffee bean weevil (*Araeceres fasciculatus*)
Pests of roots and tubers	
Powder-post beetles (*Lyctus* spp.)	Sweet potato weevils (*Cyclas* spp.)
False powder-post beetles (*Bostrichid* beetles)	Yam beetles (*Prionoryctes* spp.)
Siamese grain beetle (*Lophocateres pusillus*)	Broadnosed grain weevil (*Caulophilus oryzae*)
Flat grain beetles (*Cryptolestes* spp.)	
Pests of dried fruit and nuts	
Cigarette beetle (*Lasioderma serricorne*)	Red flour beetle (*Tribolium castaneum*)
Sap beetles (*Carpophilus* spp.)	Nut weevils (*Curculio* spp.)
Sawtoothed grain beetles (*Oryzaephilus surinamensis*)	
Pests of dried meats and other animal products	
Hide beetles (*Dermestes* spp.)	Carpet beetles (*Anthrenus* spp.)
Carpet beetles (*Attagenus* spp.)	Ham and copra beetle (*Necrobia* spp.)

At temperatures below 13–15 °C reproduction is severely limited in most species, and at lower temperatures it is prohibited altogether.

In conjunction with managing the storage environment there have been two approaches – preventive and corrective – to the effective use of chemical pesticides. The preventive approach has involved the use of grain and oilseed protectants, residual pesticides sprayed or dusted directly on to the seeds, or into raw foods that have a legal tolerance for the pesticides. Residuals will be applied most often into the storage in the form of crack and crevice, spot or general sprays or dusts, to prevent crawling pests from reaching the product. Often a program of periodic applications of insecticide fogs or aerosols is used to supplement these residual pesticide applications. These fogs and aerosols are contact pesticides and provide little, if any, residual, but provide excellent control of exposed crawling and flying pests. (*See* **Pesticides and Herbicides**: Residue Determination.)

The corrective approach involves the use of fumigants, particularly when an infestation gets out of hand, or when there is a need for a quick kill to facilitate a sale or meet a quarantine requirement. When a food requires this treatment, it is generally the result of the preventive program being ineffective as a result of some breakdown in the pest management program.

There are several fumigants available throughout the world, but many are now being phased out or banned because of their adverse effect on humans or the environment. Two fumigants, methyl bromide and phosphine, are still readily available and used throughout most of the world for the protection of foods. Phosphine or hydrogen phosphide gas, a product of atmospheric moisture reacting with a metal phosphide to release the gas PH_3, is the easiest to use, readily available and leaves little, if any, residue. It does require 3–5 days' confinement, with the product under ideal temperature conditions of 15–35 °C, and longer if the gas is to penetrate to depths of 10–20 m without the assistance of a recirculation system. Methyl bromide is still the fumigant of choice when fumigations must be effected within 24 h or less. It is also the fumigant of choice by most quarantine agencies. However, methyl bromide has been determined to be an ozone depleter and therefore is considered to be detrimental to the earth's protective ozone layer (in an international treaty referred to as the Montreal Protocol). Currently methyl bromide is being phased out of production and use and most probably will not be available as a fumigant in developed nations after the year 2005 and developing nations after the year 2015.

The major problem in effecting satisfactory control is the difficulty of detecting infestations of the weevils and borers. Shortly after an infestation is established, more than 95% of the population – larvae and pupae – is hidden within the seeds. This results in a large amount of damage for the small number of adults that may be visible. It is therefore important to maintain an effective surveillance and inspection program if populations of these insect pests are to be managed and damage is to be minimized.

Problems and Control of Moths

As with the beetles, moth pests of stored products are also primarily pests of grains and legumes. However,

Table 2 Moth pests of stored products

Moth pests of grain and legume seeds and their processed products	
Almond moth (*Cadra cautella*)	Meal moth (*Pyralis farinalis*)
Indianmeal moth (*Plodia interpunctella*)	Rice moth (*Corcyra cephalonica*)
Angoumois grain moth (*Sitotroga cerealella*)	Mediterranean flour moth (*Ephestia kuehniella*)
Tobacco moth (*Ephestia elutella*)	Pink bollworm (*Pectinophora gossypiella*)
Moth pests of roots and tubers	
Potato tuber moth (*Phathorimaea operculella*)	Gray pyralid (*Pyralis manihotalis*)
Moth pests of dried fruits and nuts	
Almond moth (*Cadra cautella*)	Hickory shuckworm (*Lespeyresia caryana*)
Dried fruit moth (*Ephestia calidella*)	Naval orangeworm (*Paramyelois transitella*)
Raisin moth (*Ephestia figulilella*)	Indianmeal moth (*Plodia interpunctella*)
Moth pests of dried meats and other animal products	
Gray pyralid (*Pyralis manihotalis*)	White-shouldered house moth (*Endrosis sarcitrella*)
Case-bearing clothes moth (*Tinea pellionella*)	Brown house moth (*Hofmannophila pseudospretella*)
Common clothes moth (*Tineola bisselliella*)	

they are major pests of dried fruits and animal products. In Table 2 is a listing of the major moth pests by commodity groupings. (*See* **Legumes**: Legumes in the Diet.)

The control of moths in storage situations is effected in much the same way as the control of beetles and weevils. Their presence is readily recognized by the presence of flying adults. The larvae of the moths, the caterpillars, do not penetrate deeply into the stored product and usually are found feeding close to the surface. The larvae trail out a dragline of silk as they move about. In very heavy infestations these silk lines can coalesce and form a cover over the surface like a tarpaulin and even affect the penetration of fumigant gases. However, when any amount of silk occurs it is the result of a neglected or mismanaged situation. One pest, the Angoumois grain moth, has a behavior somewhat similar to that of weevils. It often infests the grain in the field, and the larvae mine inside the seed, making it difficult to assess the population. The adults of the moths do not feed; the larvae cause all the damage.

Today the monitoring for infestation of moths is made much easier by the use of pheromones and traps. However, their acceptance by warehouse staff has been slow because of the low dollar value per unit of raw commodity stored. There also remain some questions as to how to space these traps in a warehouse and exactly what the catch means. It is only a matter of time before these questions are answered, at least for our major commodity groups.

Problems and Control of Mites

Mites are not primary pests of most foods, except possibly for crops such as mushrooms and cheeses, and similar products associated with fungi and fermentation. There are relatively few mites of importance to foods, but those that do attack often cause very serious problems and great loss of product. In Table 3 is a list of some of the major mite pests occurring in foods. (*See* **Spoilage**: Yeasts in Spoilage.)

When mite infestation occurs, it is usually the result of unfavorable storage conditions. Mite populations flourish in cool, moist environments which are also conducive to the growth of fungi and yeasts. Under these conditions, the mites feed on both the fungi and yeast and on the substrata that have been prepared by these fungi and yeasts. The fungi and yeast break down the seed coat on the grain and the legumes, and moisten or otherwise prepare the food for the mite to feed on.

The control of mites is largely one of correcting the environmental conditions so that it is unfavorable for the mites. The use of residual pesticides is usually ineffective because the fungi, yeasts, and moisture degrade the pesticides. Residues on adjacent surfaces, e.g., walls and floors, are quite ineffective, as the mites are small; they move slowly and rarely, if ever, encounter these surfaces. The use of fumigants is effective, but they may not always be legally used on the food product involved. Some success has been achieved by use of high temperatures, above 38 °C, but some foods may not accept such high temperatures.

Past, Present, and Future of Insect Pest Control

Pest control methodology can be found in the ancient writings of biblical times. Grain was stored in sealed

Table 3 Mite pests of stored products

Mite pests of grain and legume and their processed products	
Flour mite (*Acarus siro*)	*Suidasia nesbitti*
Grain mite (*Glycyphagus destructor*)	*Calogylyphus rhizoglyphoides*
Tyrophagus longior	
Mite pests of dried fruits and nuts	
Flour mite (*Acarus siro*)	*Carpoglyphus lactis*
Mite pests of dried meats and other animal products	
Flour mite (*Acarus siro*)	Cheesemite (*Tyrophagus casei*)
Lardoglyphus konoi	

jars, providing a lethal atmosphere devoid of oxygen and high in carbon dioxide, or grain was treated with abrasive minerals or diatomaceous earth or fumigated by sulfur dioxide produced by burning sulfur. During the last century dependence has also been on the arsenicals and botanicals and several fumigant gases. Since World War II, we have been heavily dependent upon chemicals, e.g., chloronated hydocarbons, organophosphates, carbamates, and pyrethroids. We have now learnt that the effects of these chemicals on the environment and their efficacy for insect control have been counterproductive to our goals for a quality of life.

During the last 30–40 years our research efforts have been ever-increasing in the search to protect our food, feed, seed, and fiber by better detection and damage assessment and the implementation of programmes of integrated pest control. Integrated pest control is described as including an integration of pest control tactics that emphasize the prevention and avoidance of pest problems and meet economic, public health, and environmental goals.

The following list describes some pest control tactics of the present and future for possible use in integrated pest control:

1. Modified atmospheres – the use of reduced oxygen concentration by the addition of carbon dioxide, nitrogen, or combustion gases.
2. Gamma-radiation – the use of ionizing electromagnetic energy, which causes cell and tissue damage.
3. Accelerated electrons – the use of high-speed electrons, which upon impact destroy cells and tissues.
4. Hermetic storage – the sealing of a food in a container so that natural respiration of the food and any pests present will reduce the oxygen content to a level that will not sustain life.
5. Parasites and predators – the addition of insect parasites and predators to raw commodities to control the pest feeding on the food.
6. Insect pathogens – the use of bacteria, viruses, and fungi that naturally attack the pests of foods but are not harmful to humans or animals.
7. Temperature – the use of high or low temperatures, including infrared and microwave radiations.
8. Genetic engineering – the insertion of a gene that produces a material that is toxic to the insect pests but not to other animals.
9. The safe and responsible use of chemical insecticides and acaricides in combination with other pest control tactics. (*See* **Chill Foods:** Effect of Modified-atmosphere Packaging on Food Quality; **Irradiation of Foods:** Basic Principles.)

See also: **Cereals**: Handling of Grain for Storage; **Chilled Storage**: Use of Modified-atmosphere Packaging; **Fumigants**; **Irradiation of Foods**: Basic Principles; **Legislation**: Contaminants and Adulterants; **Legumes**: Legumes in the Diet; **Pesticides and Herbicides**: Residue Determination; **Spoilage**: Yeasts in Spoilage; **Storage Stability**: Parameters Affecting Storage Stability

Further Reading

Anonymous (1978) *Postharvest Food Losses in Developing Countries*. Washington, DC: National Academy of Sciences, National Research Council, Board of Science and Technology for International Development.

Anonymous (1986) *Stored-Grain Insects*. Agriculture Handbook no. 500. Washington, DC: USDA, Agricultural Research Service.

Anonymous (1998) *Determining the Practice of Integrated Pest Management: A Working Definition for the Year 2000 Goal*. IPM Committee. Washington, DC: US Department of Agriculture.

Arbogast RT, Lecato GL and Byrd RV (1980) External morphology of eggs of stored-product moths (Lepidoptera: Pyralide, Gelechiidae, Tineidae). *International Journal of Insect Morphology and Embryology* 9: 165–177.

Baur FJ (ed.) (1984) *Insect Management for Food Storage and Processing*. Minneapolis: American Association of Cereal Chemists.

Davis R (1986) Managing pests of food. In: Bennett GW and Owens JM (eds) *Advances in Urban Pest Management*, pp. 287–311. New York: Van Nostrand Reinhold.

Gorham JR (ed.) (1978) *Training Manual for Analytical Entomology in the Food Industry*. FDA Technical Bulletin no. 2. Washington, DC: Food and Drug Administration, US Department of Health, Education and Welfare.

Gorham JR (ed.) (1987) *Insect and Mite Pests in Food: An Illustrated Key*. Agriculture Handbook no. 655. Washington, DC: US Department of Agriculture.

Hughes AM (1961) *The Mites of Stored Foods*. Ministry of Agriculture, Fisheries and Food. Technical Bulletin no. 9. London: Her Majesty's Stationery Office.

LeCato GL and Flaherty BR (1974) Description of eggs of selected species of stored-product insects Coleoptera and Lepidoptera. *Journal of Kansas Entomological Society* 47: 308–318.

INSTRUMENTATION AND PROCESS CONTROL

R G Moreira, Texas A&M University, TX, USA

Introduction

Process control involves the regulation of variables in a process. It is the combination of raw materials and equipment that produces a desirable final product through changes in their energy and physicochemical properties. These processes can be continuous or batch.

In the food-processing industry, automatic control is one of the key elements that affect productivity, quality, safety, and competitiveness. Automation has become an integral part of modern food manufacturing today. It is essential, for example, in controlling pressure, temperature, viscosity, humidity, and flow rate as well as product quality attributes such as color, texture, bulk density, moisture, and oil contents.

In the past, food-processing systems used only crude control of flow rates, ampere meters to estimate the power drawn by drive motors, and thermocouples and visual observation at the process discharge to maintain product specifications. Nowadays, new applications require precise control over the input variables to achieve a consistent product within narrow specification limits and to maximize throughput. Food process operations today are monitored and controlled by computer systems, using accurate information of input and output variables and advanced control algorithms.

Description of Control Systems

The Control System

A control system is considered to be any system that exists for the purpose of regulating and controlling a process output. Consider, for example, the air-conditioning system used to keep fruits and vegetables fresh in a warehouse. The air-conditioning unit (consisting of a condenser and a compressor) is a system that produces cold air as a result of the flow of refrigerant. This process is assembled with subsystems such as valves and valve actuators that regulate the temperature of the warehouse by controlling the cold air output from the air-conditioning unit. Other subsystems, such as thermostats, which act as sensors, measure the room temperature.

Simply stated, a control system provides an output or response for a given input or stimulus, as shown in Figure 1. The input–output relationship represents the cause-and-effect relationship of the process.

The Input and Output

The input represents a desired response: a factor (variable) that is used to modify the system behavior. It is also called manipulated variable, i.e., a dependent variable of the process used to cause a change in the process.

The output is the actual response – a factor (variable) that is caused by a system and is used as a measure of performance for the given system. It is also called controlled variable, i.e., an independent variable of a process, and can be defined as the parameters of the process that indicate product quality or the operating condition of the process.

The input–output relationship represents a processing of the input signal to provide an output signal variable, often with power amplification.

Open-Loop and Closed-Loop Systems

An open-loop control system (Figure 2a) utilizes a controller (or control actuator) to obtain the desired response (also called reference input or setpoint). The input is sometimes called the reference (or target). An example of such a system is a bread machine, where mixing, raising, and kneading operate on a time basis. The bread machine does not measure the output that is the final product – the bread quality (color, volume, doneness).

Open-loop control systems are used when the relationship between the input and output is known and when internal or external disturbances are not present in the process. Disturbances such as machine wear and ambient humidity can affect the final product quality and the system cannot correct for these disturbances.

Therefore, open-loop systems do not correct for disturbances and are simply commanded by the input. A toaster is another typical example of an open-loop system. In this case, the controlled variable (output) is the color of the toast. The machine

Figure 1 Simplified description of a process to be controlled.

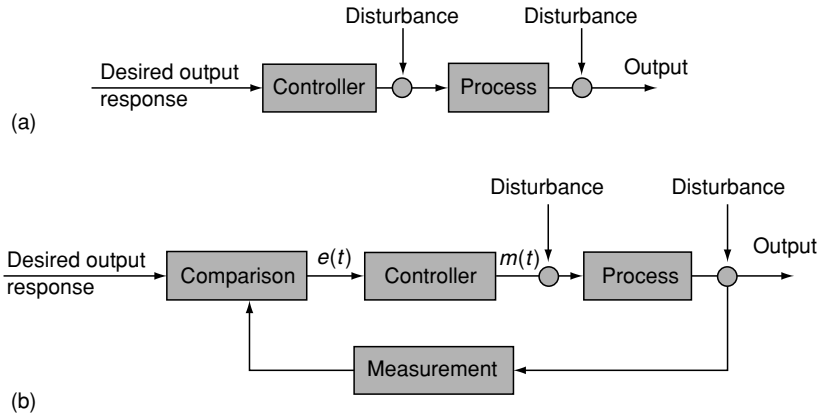

Figure 2 (a) Open-loop control system; (b) closed-loop control system.

operates by applying heat to turn the toast darker. The toaster does not measure the color of the toast. It does not correct for the fact that the toast is rye, white, or sourdough, nor if the toast comes in different thicknesses.

Sensitivity to disturbance and the system's inability to correct for disturbances are the main disadvantage of the open-loop system. Closed-loop control systems may be used to overcome the drawbacks of open-loop systems. A closed-loop control system uses an additional measure of the actual output to compare the actual output with the desired output response. The measure of the output is called the feedback signal. A simple closed-loop feedback control system is shown in Figure 2b.

A feedback control system uses a function of the prescribed relationship between the output and the reference input to control the process. At the comparison box (Figure 2b), the difference between the output of the process under control (which arrives via the feedback path) and the reference input is amplified and used to control the process so the difference is continuously reduced. This difference is referred as the input error ($e(t)$ in Figure 2b).

The comparison box generally consists of an input transducer that converts the input signal to a form used by the controller. At the measurement box, an output transducer (sensor) is used to measure the output response and also to convert this response signal to the form used by the controller. For example, if the controller uses electrical signals to operate the valve of a temperature control system, the input position and the output temperature are converted to electrical signals. Potentiometers are used to convert input position to voltage. A thermistor, a device that changes its electrical resistance with temperature, is used to convert output temperature to voltage.

The closed-loop system compensates for disturbances by measuring the output response, feeding the measurement back through a feedback path, and comparing that response to the input at the comparison box (Figure 2b). If there is any difference between the two responses ($e(t) \neq 0$), the system drives the process to make a correction. If there is no difference ($e(t) = 0$), the system does not drive the process, since the process response is already the desired response. Closed-loop systems are also called feedback control systems.

To illustrate the concept of feedback control, Figure 3a shows a simplified manual control of a baking process. On the outlet side of the oven, there is an indicator (colorimeter) to provide the operator with information on the actual value of the biscuit outlet color (controlled variable). The operator wants to keep the biscuit final color close to the desired value. If the operator watches the colorimeter and finds that the color value is higher (the biscuit is darker) than the desired response, he/she will increase the conveyor speed, i.e., the auger r.p.m. It is considered in this example that the oven temperature and air flow rate are maintained constant for simplicity. The desired response, of course, is in the operator's mind and the operator makes all of the control decisions. It is quite possible that the color becomes lighter so the operator will have to repeat the operation in the opposite direction until the desired value is reached and kept constant during the operation.

If the operator is replaced by an automatic controller, as shown in Figure 3b, the system becomes an automatic feedback control system. The actual outlet biscuit color (the process output), measured by the colorimeter, is compared to the desired response and an acting error signal generated. The output color is converted to the same units of the desired response by

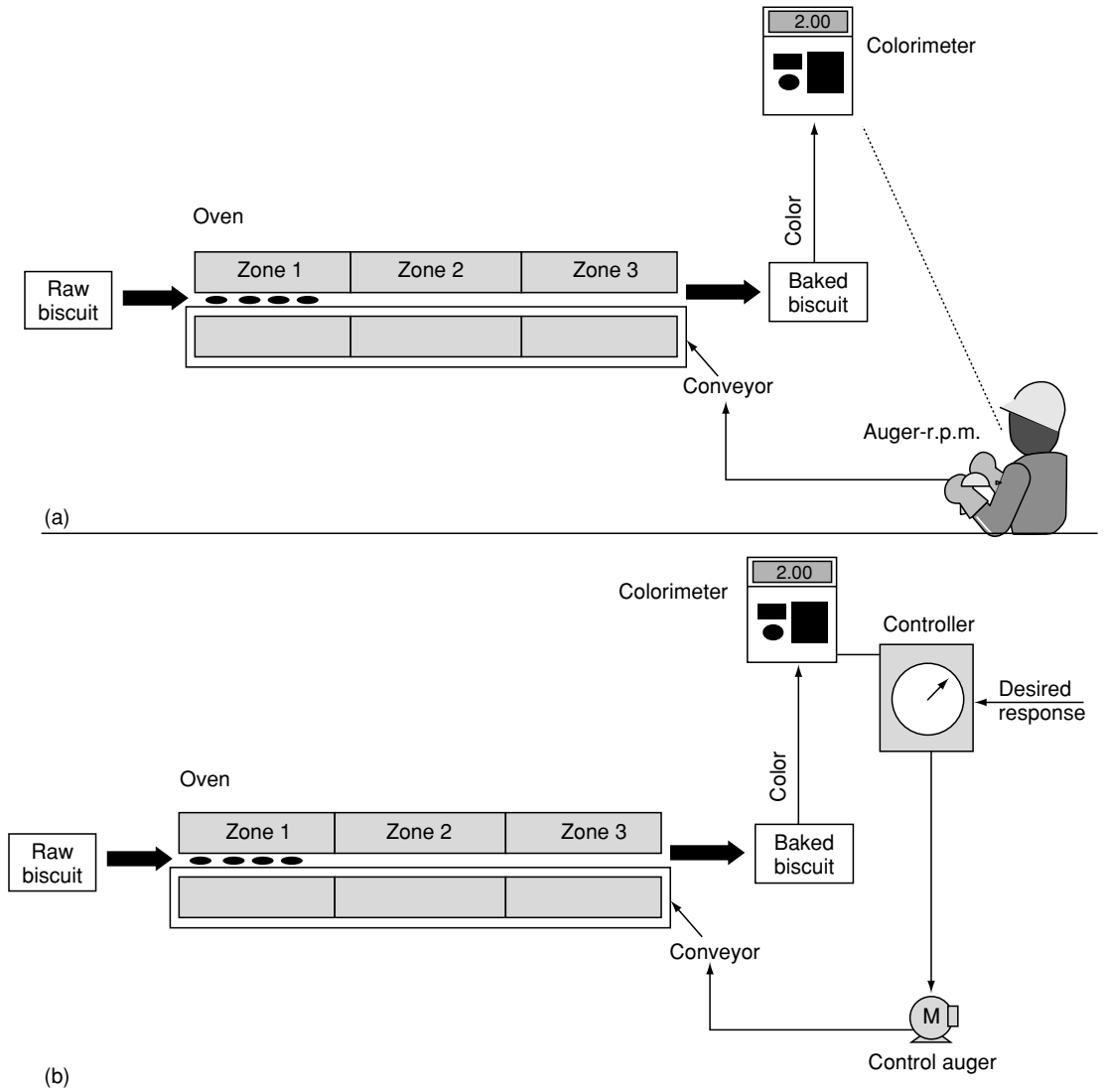

Figure 3 Manual and automatic control of a baking process.

a transducer. Based on the acting error value, the controller calculates the changes that need to be made in the auger r.p.m. to reduce and then eliminate that error.

Both manual and automatic controls operate likewise. In this case, the operator's eyes correspond to the acting error device, the brain corresponds to the automatic controller, and the operator's muscle represents the actuator.

Advantages of closed-loop systems as compared to open-loop systems include greater accuracy and less sensitivity to noise and disturbances. The transient response and the steady-state error can be controlled easily and with great flexibility in closed-loop systems. On the other hand, closed-loop systems are more complex and expensive than open-loop systems.

Computer-Controlled Systems

Since the 1970s, process controls have evolved from pneumatic analog technology through electronic analog technology to microprocessor-based controls.

Nowadays, many systems use a digital computer as the controller. The use of a computer in the control-system loop allows for the control and compensation of many loops by the same computer through time sharing. In addition, any adjustment of the controller parameters can be made by changes in software rather than hardware.

One important aspect of computer-controlled systems is its real-time computing application. It allows initialization of data acquisition operations and other tasks as control output calculations, supervision, scheduling, and so on. Real time means

that an operation is to be completed within a specified time constraint.

The Control Problem

The control problem can be stated as follows: determine the one value of the manipulated variable (input) that establishes a balance among all the influences (loads/disturbances) on the controlled variable (output) and keeps this variable constant at the desired value.

The steps required to design a control system are: process description, process and signal modeling, design method, and control algorithm. Once the controlled, manipulated, and load variables have been identified, the next step is to describe mathematically the process and signals (disturbances, reference variables, initial values). The design methods for linear control systems can be classified into parameter optimized control systems and structure optimized control systems. Once the design method has been selected, designing principles are applied to determine the control parameters.

Closed-Control System Analysis and Design

From the control point of view, a system is composed of controlled variables, manipulated variables (process input), and disturbances. In a control scenario, the number of manipulated (independent) variables must be at least as large as the number of controlled (dependent) variables. The controlled variables should measure product quality directly or strongly affect it. The manipulated variables should have a large effect on the controlled variables (large gain), should rapidly affect the controlled variables (minimum delay, small time constant), and should affect the controlled variables directly rather than indirectly.

Control systems are dynamic systems. That is, the response of a control system to an input is generally transient before reaching a steady-state response. Consider, for example, the baking process (Figure 3b) where a new biscuit is baked to a darker color by changing the conveyor speed. When the desired response (color of the biscuit) is changed to a higher level (darker), the biscuit entering the oven will leave the oven with a speed and color accuracy designed for product quality. Figure 4 shows the input and output for the baking system. The change in the desired color value is the input and is represented by a step command. Due to the limited power available and to keep the products already in the oven from breaking apart, it is not advised to have the oven

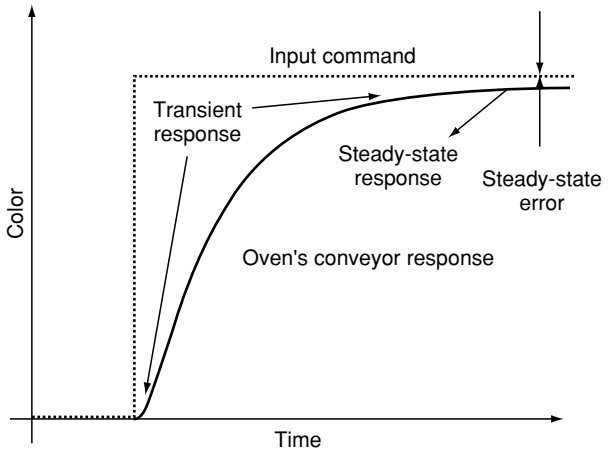

Figure 4 Oven input and output.

mimicking the suddenness of the input. The input represents what the output must be after the conveyor speed has changed; the conveyor follows the change described by the curved region in Figure 4; this is termed the oven's conveyor response. This part of the response is called the transient response.

After the transient response, the process approaches its steady-state response, i.e., approximates the desired response. For the baking process example, this response occurs when the biscuit reaches the outlet side of the oven. The accuracy of the biscuit's final color is a second factor that could make the output different from the input. This difference is called steady-state error. Steady-state errors exist when the control systems are defective.

Transient Response

In the baking process presented above, if the transient response were too slow, the process would be inefficient (low throughput), whereas an excessively rapid response would not be sufficient to bake the product to the desired color. If the conveyor speed oscillated around the desired response, the product color would be different from one biscuit to the another, resulting in poor quality. Therefore, transient response is very important in the design and analysis of control systems.

Steady-State Response

Another important objective in the analysis and design of control systems is that the system has a 'good' steady-state error performance. The steady-state response resembles the input and is usually what remains after the transients have decayed to zero. In the baking process, this response is the conveyor speed modified to a speed that produced the desired biscuit final color.

Stability

System stability under all system operating conditions is another control system specification. A system is unstable when its natural response (the way the system dissipates/acquires energy independently of the input) is much greater than the forced response (dependent upon the input) making the system uncontrollable. This condition could lead to self-destruction of the physical device if limit stops are not provided as part of the design. Therefore, control systems must be designed to be stable, i.e., their natural response must decay to zero or oscillate as time approaches to infinity.

Process Control

The objective of a process control system is to maintain a variable in the process at its setpoint (desired response). A process control system may be either open-loop or closed-loop, but closed-loop systems are more common. The process control industry has developed controllers that are standard and flexible for closed-loop systems.

Examples of common features shared by many process controllers include: (1) the values of the setpoint, process, and output variables are shown in either analog or digital format; (2) the operator is allowed to adjust the setpoint and switch between automatic and manual control; (3) when manual control is selected, the operator is allowed to adjust the controller output to vary the manipulated variable in an open-loop control mode; (4) the operator is allowed to adjust the control model settings to 'tune the controller' for optimum response; and (5) to provide for remote setting of the setpoint by an external signal. Many additional features are provided by microcontrollers (microprocessor-based digital controllers).

A control system (feedback or feedforward) performs essentially three operations: (1) measurement; (2) manipulation; and (3) signal transmission. Measurement is an integral part of a control system. To control a variable, its values need to be measured and converted to suitable signals. Manipulation consists of use of the control action to manipulate some variable in the process in a way that will tend to reduce the error. Signal transmission refers to signal communications between controllers and instruments.

Measuring Transmitter

A measuring transmitter (sensor) senses the value of a controlled variable and converts it into a usable signal. Example of sensors include temperature transmitter, flow transmitter, pressure transmitter, etc. Consider, for example, a temperature transmitter. The primary element of this sensor could be a thermocouple, a resistance element, a thermistor, or a filled thermal element. The signal transducer receives the output of the primary element and produces an electric current signal. In the case of a thermocouple, for example, it converts temperature into a millivolt signal, and the thermocouple transducer converts the millivolt signal into an electric current.

Manipulating Element

A controller has two interfaces with the process it controls: the input interface to the controller, and the output interface from the controller. The input side is handled by sensors and signal conditioners. Various types of manipulating elements handle the output side.

A manipulating element uses the controller output to regulate the manipulated variable and usually consists of two parts: actuator and final controlling element. An actuator translates the controller output into an action on the final control element, and the final controlling element changes the value of the manipulated variable. Manipulating elements include valves, dampers, fans, pumps, heating elements, and electric motors.

The pneumatic control valve is the most common final control element used to manipulate the flow rate of a fluid. Figure 5 illustrates a typical pneumatic control valve. The input to the valve is air pressure that acts on the diaphragm, compressing the spring, thus pulling the stem out and finally opening the valve. The actuator of a pneumatic valve can be

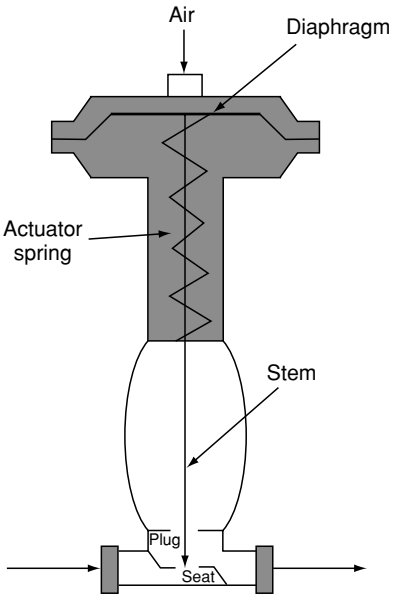

Figure 5 Pneumatic control valve.

described as direct actuator (if the actuator spring is located below the diaphragm) or reverse actuator (when the spring is above the diaphragm). The valve position is the displacement plug from the fully seated position. The valve action is in the direction in which the valve plug moves as the air pressure increases. If the valve closes as the pressure increases, an air-to-close valve results. If the increased air pressure opens the valve, it becomes an air-to-open valve.

Signal Transmission

Most signals can be described as analog, digital, or pulse. Analog signals typically vary smoothly and continuously over time; digital signals are present at discrete points in time. In most control applications, analog signals range continuously over a specified current or voltage range, such as 4–20 mA DC or 0–5 V DC. Digital signals are essentially on/off (the pump is on or off), while analog signals represent continuously variable entities such as temperatures, pressures, or flow rates. Computer-based control systems understand only discrete on/off information, therefore conversion of analog signals to digital representations is necessary.

Transduction is the process of changing energy from one form into another. Hence, a transducer is a device that converts physical energy into an electrical voltage or current signal for transmission. There are many different forms of analog electrical transducers. Examples of transducers include thermocouples and resistance temperature detectors (RTDs) for measuring temperature via voltage and resistance measurement, respectively. Transmission channels may be wires or coaxial cables. For noise-resistant transmission over significant distances, the raw transducer signal is often converted to a 4–20 mA signal by a two-wire, loop-powered transmitter.

Industrial networks that transmit data using digital signals are often an integral part of a process control solution. Signals from digital instruments and controllers are transmitted in digital format (on–off pulses) using, for example, a number of parallel wires. New instrumentation systems use a single data highway to transmit digital signals to a series of instruments and controllers. These data highways are usually coaxial cables linked in serial or daisy-chain (serial with complete loop). A microcomputer built in the instrument and controller communicate periodically over the highway by sending to or requesting from information between different devices installed in the system.

Examples of data highway are wide area network (WAN) and local area network (LAN). Networks are communication channels that connect large numbers of stations to one or more central stations. WANs are used to connect process systems separated by considerable distance (branch offices in different cities). LAN is a data highway used to connect local process units with central operator displays and high-level computers. LANs are used to connect process systems located within a radius of about 1.5 km.

Nearly all digital network descriptions start with the open systems interface (OSI) model. The OSI model is based on several functional levels. The lower levels are responsible for transmitting messages between nodes on the same cable and from one cable to another. The upper levels handle data formatting and access security. A number of LANs are based on the OSI model, such as the manufacturing automation protocol (MAP) and the Ethernet.

The MAP is a multilevel communication network whose objective is to link together all the computers, controllers, equipment, devices, and offices in an entire factory. The Ethernet is able to handle error detection and address the source of destination.

Control Modes

The operations performed by a feedback control system include measurement, decision, and manipulation. A closed-loop control system is shown in Figure 2b. Measurement and manipulation have been described above. Decision consists of calculating the error (desired value minus measured value) and using the error to form a control action.

The control modes convert the error into a control action (controller output) that will reduce the error. The most common control modes are the proportional mode (P), the integral mode (I), and the derivative mode (D).

The simplest of the three modes is the proportional mode. As the name indicates, the proportional mode produces a control action that is proportional to error. If the error is small, the control mode will produce a small action. If the error is large, the control action will be large. To accomplish the proportional mode, the error is multiplied by a gain constant, K_p. In a proportional control, the controller output is algebraically proportional to the error input signal to the controller as:

$$m(t) = K_p e(t) \tag{1}$$

where m(t) is the controller output, $e(t)$ is the error input signal, and K_p is the controller gain. Eqn (1) is called the control algorithm.

The integral mode produces a control action that continues to increase its corrective effect as long as the error persists. The integral mode increases slowly when the error is small, and corrects more rapidly if the error is large. Integral control action is based on

the principle that the controller output *m(t)* is proportional to both the size and the duration of the actuating error signal *e(t)*:

$$\frac{dm(t)}{dt} = K_i e(t) \tag{2}$$

or:

$$m(t) = K_i \int_0^t e(t)dt \tag{3}$$

where K_i is an adjustable constant.

The derivative mode produces a control action that is proportional to the rate at which the error is changing. If the error is increasing rapidly, the derivative mode will prevent the error becoming larger by producing a correction action proportional to how fast the error is changing. Derivative action (also called rate control action) is based solely on the rate of change of the error signal. An ideal derivative action can be described as:

$$m(t) = K_d \frac{de(t)}{dt} \tag{4}$$

where K_d is an adjustable constant.

These three control modes can be used alone (with the exception of the derivative mode) or combined. The common control mode combinations are: P, PI, PD, and PID. The proportional plus integral (PI) control mode is used on processes with large load changes when the proportional mode alone is not capable of reducing the offset. The proportional plus derivative (PD) control mode is used on processes with sudden load changes. The proportional plus integral plus derivative (PID) control mode is used on processes with sudden, large load changes.

Figure 6 shows the typical response behavior of a controlled process after a step change in the setpoint occurs. The output variable is shown as a deviation from the new setpoint change. If no feedback control is implemented, the process slowly reaches the new steady state. Proportional control speeds up the process and reduces the offset. Adding integral action (PI) eliminates the offset but causes the loop to oscillate. The addition of derivative control (PID) made the response faster and less oscillatory.

Advanced Control

The PID controller has been around for more than 50 years and is still today the most widely used industrial controller. It is simple, easy to understand, and does not require a precise process model to start up or maintain. The PID works well if the process system is simple. However, if the process dynamics change, it needs to be retuned, and it has trouble controlling complex systems, i.e., processes that are nonlinear, time-variant, coupled, and have parameter or structure uncertainties.

Several advanced control strategies exist that can provide improved process control over PIDs. Some of these methods include cascade control, robust control, adaptive control, model predictive control, and intelligent control. With the advance of computer control, many of these techniques have been successfully applied in the food industry.

Cascade control involves two controllers, with the output controller providing the setpoint of the

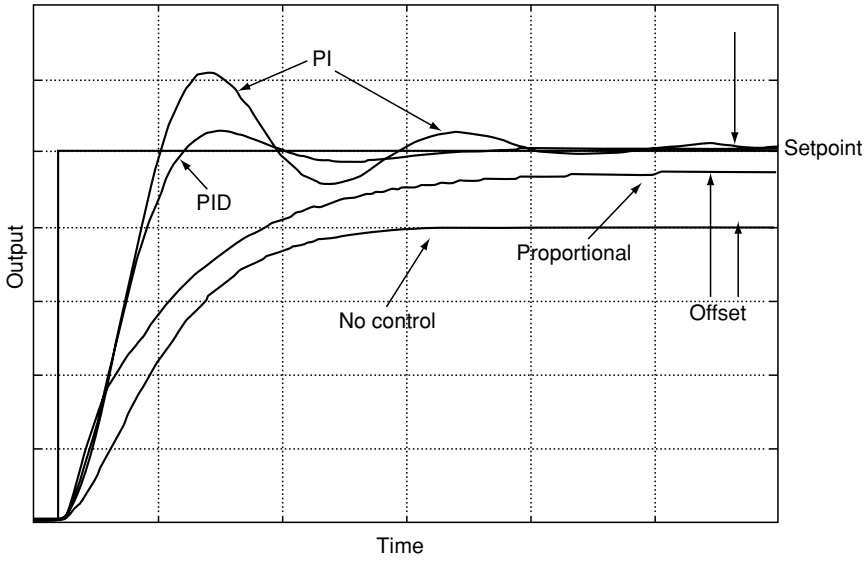

Figure 6 Step response of a feedback control system. PID, proportional plus derivative; PI, proportional plus integral.

secondary controller. The design of controllers or control algorithms described up to now assumed that only the control variable determines the process input. This results in single control loops. By connecting additional measurable variables to the single loop (i.e., disturbances or auxiliary variables) it is possible to obtain improved control behavior. These additions to the single loop lead to interconnected control systems.

Robust control is a controller design method that focuses on the reliability (robustness) of the control algorithm. Robust control methods are suited to processes that have large uncertainty ranges and small stability margins.

Adaptive control is defined as an intelligent feedback control system that adjusts its characteristics to changes in the environment so as to operate in an optimal manner. In process control applications, the PID self-tuning scheme is widely implemented in commercial products.

Model predictive control (MPC) is one of the few advanced control techniques used successfully in industrial control applications. The essence of MPC is based on three key elements: (1) predictive model; (2) optimization in range of a temporal window; and (3) feedback correction. These steps are carried on continuously by computer programs online.

Intelligent methods in control are becoming part of mainstream control approaches. They integrate concepts and methods from areas such as control, identification, estimation, communication theory, and artificial intelligence. Intelligent control paradigms can endow control systems with functionality that traditional methods could hardly make available. They can also improve the analysis, design, operation, and maintenance characteristics of control systems. Examples of this type of control strategy include fuzzy logic and neural network.

Fuzzy logic and neural network aim at mimicking the operation of the human brain to some extent. A control law described by fuzzy rules is more understandable by a human operator who knows the process but who is not an expert control engineer. It facilitates the integration of heuristic knowledge about the system and theoretical control laws from mathematical models, that are always an idealization of the process.

Neural networks go further by aiming at building models of the way a human being thinks and reaches conclusions. In general, neural networks are useful for representing and approximating nonlinear relationships. They are applicable for multivariable systems and require less restrictive assumptions on the process than conventional modeling and control techniques.

See also: **Plant Design**: Basic Principles; Designing for Hygienic Operation

Further Reading

Bateson RN (1999) *Introduction to Control System Technology*. Upper Saddle River, NJ: Prentice Hall.

Brescia L and Moreira RG (1997) Modeling and control of a continuous frying process: II Control development. *Transactions of Chemical Engineering* 75(c): 12–20.

Clarke DW, Mohtadi C and Tuffs PS (1987) Generalized predictive control – part I. The basic algorithm. *Automatica* 23(2): 137–148.

Clarke DW, Mohtadi C and Tuffs PS (1987) Generalized predictive control – part II. Extensions and interpretations. *Automatica* 23(2): 149–160.

Isermann R (1989) *Digital Control Systems*, vol. II. *Stochastic Control, Multivariable Control Adaptive Control, Applications*. New York, NY: Spring-Verlag.

Moreira RG (2001) *Automatic Control for Food Processing Systems*. Aspen.

Ogata K (1970) *Modern Control Engineering*. Englewood Cliffs, NJ: Prentice-Hall.

Prett DM and Garcia CE (1988) *Fundamental Process Control*. Boston, MA: Butterworth-Heinemann.

Schonauer S and Moreira RG (1995) Development of a fixed-GPC controller for a food extruder based on product quality attributes. Part I: control development, implementation and analysis. *Transactions of the Institute of Chemical Engineering Part C* 73(C4): 200–210.

Seborg DE, Edgar TF and Mellichamp DA (1989) *Process Dynamics and Control*. New York, NY: John Wiley.

Takahashi Y, Rabins MJ and Auslaner DM (1972) *Control and Dynamic System*. Menlo Park, CA: Addison-Wesley.

Insulin *See* **Diabetes Mellitus**: Etiology; Chemical Pathology; Treatment and Management; Problems in Treatment; Secondary Complications

Interesterification *See* **Vegetable Oils**: Types and Properties; Oil Production and Processing; Composition and Analysis; Dietary Importance

INTERMEDIATE-MOISTURE FOODS

J Sych, Université Laval, Québec, Canada

Introduction

Lowering of water activity (a_w) to a level where most microorganisms cannot grow is one of the oldest methods of food preservation and forms the basis for intermediate-moisture foods (IMF). This article will review the nature and uses of IMF by presenting the principles behind formulation, some typical examples on the market, factors limiting stability, and possible future developments of IMF.

Principles Behind Formulation of IMF

IMF have a_w reduced to approximately 0.65 to 0.90 (i.e., moisture content of 15–40%), are plastic enough to be consumed without rehydration, and are stable without refrigeration. Other preservative factors act in association with a_w to insure the shelf-life of IMF.

a_w is the amount of water available for deteriorative processes in a food product and is defined as the ratio of water vapor pressure of the product (p) to the saturation pressure of water (p_o) at the same temperature: $a_w = p/p_o$. At low a_w in food, at or below the monolayer value, water is strongly bound to specific sites and unavailable as a solvent for reaction. In the a_w range of IMF, the growth of microorganisms is largely restricted, compared with high-moisture foods, but certain reactions, such as Maillard and oxidation, can still proceed, due to the increased solvent and mobilization properties of water above the monolayer. A reduction of a_w is achieved by chemical means (addition of humectants) and/or physical dehydration (evaporation, drying, ultrafiltration, etc.). (*See* **Water Activity**: Principles and Measurement; Effect on Food Stability.)

Processing Technology for IMF

Traditional IMF are processed by the withdrawal of water, by adsorption or desorption, and/or by the addition of conventional humectants, such as salt (sodium chloride) and sugar (sucrose). Although some traditional IMF, such as dried fruits, are simply dried to the desired a_w, most IMF are prepared by a combination of drying with added humectants. Bakery products are prepared by adding water and other ingredients to low-a_w cereals, followed by various processing and heating procedures to yield a_w in the range of IMF. (*See* **Sucrose**: Properties and Determination.)

Novel IMF are often processed by using humectants, such as glycerol, propylene glycol, and sorbitol, for a_w adjustment. Combinations of several humectants, each at lower concentrations, are often used to reduce off-flavor problems. Novel IMF may be prepared by:

1. Moist infusion (desorption): solid food pieces are equilibrated in a solution of lower a_w, often with heating. Drying may be used to reduce further a_w (e.g., intermediate-moisture fruits)
2. Dry infusion (adsorption): solid food pieces are dehydrated and then infused by soaking in a solution containing humectants (e.g., bite-sized cubes of meat for space or military rations)
3. Blending: food components are weighed, blended, cooked, and extruded using various processing techniques (e.g., fermented meats)

IMF processed by adsorption normally have slower microbial growth and lipid oxidation due to a lower moisture content than IMF prepared by desorption with identical a_w. This is because of sorption hysteresis – the phenomenon in foods whereby the moisture content of a food is lower during equilibration by adsorption than by desorption. In spite of these hysteresis effects and the waste of residual infusion solutions, the desorption method is generally used more, due to its advantages of lower cost, enzyme inactivation, and better retention of flavor and structural integrity. With this method, it is possible to reduce food weight by 50% by the simultaneous countercurrent diffusion of water out of the food and solute into the food.

In the production of novel IMF, pasteurization and blanching are commonly used to reduce initial levels of microorganisms and to inactivate enzymes. Antimycotics, such as sorbic and benzoic acid and their salts; methyl and propyl paraben; or calcium propionate, assist in the control of mold and yeast growth. Other chemical additives may also be incorporated to slow down physical and chemical deterioration during storage. (*See* **Pasteurization**: Principles; **Spoilage**: Chemical and Enzymatic Spoilage; Molds in Spoilage; Yeasts in Spoilage.)

Effects of a_w on Microorganisms

Minimum a_w for growth and toxin production of some microorganisms capable of growing in IMF

are shown in **Table 1**. Generally, bacteria are less tolerant of decreased a_w than are yeasts, and yeasts are less tolerant than molds (Figure 1). At $a_w < 0.95$, the growth of most Gram-negative bacteria, as well as spore-forming bacteria, such as *Bacillus* and *Clostridium* spp., is inhibited. Gram-positive bacteria, such as *Lactobacillus* and *Micrococcus* spp., tolerate lower a_w and act as competitive microflora in fermented intermediate-moisture (IM) meat. Among the foodborne pathogens, *Staphylococcus aureus* has the lowest minimum a_w for growth (**Table 1**). Yeasts and molds, referred to as xerophiles, are capable of growth in high concentrations of solute, or at $a_w \leq 0.85$. The control of bacterial growth in IMF eliminates competition, which enhances mold and yeast growth. (*See* ***Bacillus***: Occurrence; ***Clostridium***: Occurrence of *Clostridium perfringens*; **Spoilage**: Bacterial Spoilage; ***Staphylococcus***: Properties and Occurrence.)

Table 1 Some microorganisms of concern in intermediate-moisture foods[a]

Microorganism	*Toxin produced*	*Limiting* a_w	
		Growth	*Toxin*
Bacteria			
Staphylococcus aureus	Enterotoxin A, B, C	0.86	0.87–0.97
Halophilic bacteria		0.75	
Yeasts			
Saccharomyces rouxii		0.62–0.81	
Debaryomyces hansenii		0.83–0.88	
Molds			
Aspergillus flavus	Aflatoxin	0.78–0.80	0.83–0.87
A. chevalieri	Xanthocillin X	0.65	
A. ochraceus	Ochratoxin	0.77–0.83	0.83–0.87
	Penicillic acid	0.81	0.88
Penicillium viridicatum	Ochratoxin	0.83	0.83–0.86

[a] a_w range of 0.65–0.90.

Reduced a_w in foods affects the growth of bacteria, yeasts, and molds by increasing the lag phase, decreasing the rate of growth in the logarithmic phase, as well as decreasing the maximum level of development. Sodium chloride is the most effective humectant on a mole-to-mole basis for reducing a_w, followed by glycerol, sorbitol, and sucrose. However, at the same a_w, microbial growth varies with the nature of the a_w-controlling humectant. For relatively salt-sensitive bacteria, sodium chloride exerts a greater inhibitory effect than polyhydric alcohols, such as glycerol, with sugars in between. In contrast, for relatively salt-tolerant bacteria, such as *S. aureus*, glycerol has a greater inhibitory effect than sodium chloride. Many yeasts and molds are more inhibited by sodium chloride than by sucrose, glycerol, or glucose, at the same a_w. The inhibitory effects of sucrose and sodium chloride are attributed to their ability to lower a_w, whereas polyhydric alcohols have less of an effect on a_w, but appear to cause damage to enzymes, and may have antimycotic activity.

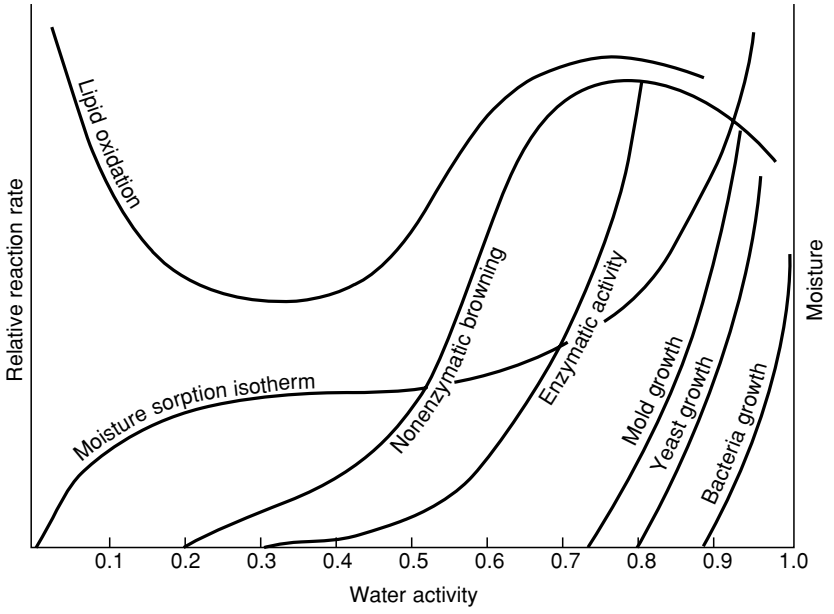

Figure 1 Relative reaction rate as a function of water activity for foods. From Taoukis PS, Labuza TP, Saguy S (1997) Kinetics of food deterioration and shelf-life prediction. In: Valentas KJ, Rotstein E, Singh RP (eds) *Handbook of Food Engineering Practice*, p. 381. New York: CRC Press, with permission.

For several genera of molds, higher a_w is required for spore formation than for germination of spores. The time required for spore germination of microorganisms increases with reduced a_w. Germination of spores of some microorganisms occurs at a_w values below that required for growth, which suggests that decreases in spore numbers can occur during storage of some IMF. The type of solute also affects the minimum a_w for spore germination.

The heat resistance of some bacteria and spores increases under stress, such as reduced a_w; this effect is solute-specific and may be associated with the synthesis of protective proteins. For heat-sensitive strains of *Salmonella*, the increase in heat resistance can be significant, particularly at high levels of small-molecular-weight solutes. Such increases in heat resistance must be taken into account in IMF production, since heat treatments are often applied after incorporating humectants.

The minimum a_w for toxin production of microorganisms is always higher or equal to that for growth (**Table 1**). Mycotoxin-producing molds present the greatest risk since many of their metabolites (e.g., aflatoxins) cause toxic effects in humans. Mold growth and toxin production on IMF surfaces can lead to penetration of toxins inside the product. In addition, data on a_w limits for mycotoxin production in IMF are limited.

Other Preservative Factors of IMF

Generally, water requirements of microorganisms for growth and toxin production increase as environmental factors are changed from their optimum. IMF are most often stabilized by reductions in a_w in combination with other preservative factors, also referred to as hurdles. Current views hold that the presence of hurdles may interfere with the homeostasis or internal equilibrium of microorganisms and thus inhibit their growth. A greater understanding of these principles has led to benefits of improved product quality and microbial safety; reduced additive levels; and reduced refrigeration. Certain combinations of hurdles can also result in additive or synergistic effects allowing for a reduction of individual factors. The most important hurdles used in varying intensities to stabilize IMF are: available water (a_w), heat processing, acidification, redox potential, modified atmospheres, preservatives, and competitive microflora. **Table 2** shows the a_w of some IMF as well as the other factors contributing to microbial stability.

The combined effect of pH and reduced a_w is used in many foods to inhibit the growth of foodborne pathogens (**Table 2**), as in IM meats where a pH of less than 5 or a_w of less than 0.85 is recommended to inhibit *S. aureus*. The removal of oxygen during packaging, i.e., low redox potential, inhibits obligate aerobes (molds and some bacteria), and increases the sensitivity of other microorganisms, including yeasts, to a_w decreases. This is used to delay rancidity and mold growth in novel IMF, such as German mini-salami with a_w between 0.82 and 0.85. Protective packaging is also an important hurdle used to maintain the stabilizing effects of redox potential and modified atmospheres.

Table 2 Water activity (a_w) and other preservative factors of some intermediate-moisture foods

Product	a_w	*Preservative factors*
Traditional IMF		
Jams	0.82–0.90	a_w, pH, F, Eh, pr[a]
Dried fruits	0.60–0.75	a_w, pH, F[a], pr
Baked products	0.60–0.90	a_w, F, pr
Meats	0.65–0.85	a_w, pH, Eh, pr, cf
Novel IMF		
Pet food	0.80–0.85	a_w, F, pr
Ready-to-eat meat cubes	0.66–0.80	a_w, F, pH[a], pr
Fruit pastes	≤ 0.70	a_w, pH, F, pr
Liquid tea concentrate	0.75–0.85	a_w, pH
Fruit-filled cereal bars	≤ 0.85	a_w, F, pH, pr
Liquid whole eggs	≤ 0.88	a_w

[a]Possible preservative factor.
F, heat processing; pH, acidification; Eh, redox potential; pr, preservatives; cf, competitive microflora.

Many IMF contain added or natural chemical preservatives to control the growth of molds and yeasts. Legislation on the types of additives and their permitted levels varies among countries. Although frequently used, benzoic and sorbic acid require the undissociated form ($pH \leq 4.5$) for optimum antimicrobial activity, which is often organoleptically unacceptable in IMF. Another problem is the high tolerance of some molds and yeasts to maximum permissible levels of preservatives. Furthermore, interactions between antimicrobials and a_w may exist; for example, *Saccharomyces rouxii* was found to have greater resistance to sorbate at reduced a_w. The increasing interest in natural preservatives has oriented much research towards the identification of natural substances, such as those in garlic which improve the stability of many traditional IM meats. Edible coatings containing antimicrobials or antioxidants can be applied to IMF to obtain specific functional effects on food surfaces. For example, an edible paste of ground garlic and other spices is used to control surface mold growth during storage of Turkish pastirma, a fermented IM meat.

Microstructure is also an important factor which affects microbial growth, survival, and death in comminuted and blended IMF, such as fermented sausages. The accumulation of natural flora as well as

starter cultures in small cavities has significant implications on the entire fermentation and on the inactivation of pathogens, which are often located in distant areas of the food matrix. An even distribution of bacteria in the sausage is important for optimizing the shelf-life of fermented sausages, obtained by adequately mixing the sausage ingredients and using liquid starter cultures.

Some Typical Examples on the Market

Pet foods have been on the market since the 1960s; they are processed by blending and pasteurizing meat byproducts, cereals, and other ingredients including humectants (sucrose and propylene glycol), followed by extrusion cooking techniques to form a variety of products. Their shelf-life stability of more than 12 months is based on reduced water activity ($a_w \leq 0.85$), heat processing, and addition of antimycotics, most commonly potassium sorbate. Other ingredients, such as sodium caseinate and starch, are sometimes added for specific textural characteristics.

There are many different traditional IM meats. One example is salami, which is a fermented sausage originating from Europe. Its shelf-life stability depends on a specific sequence of the stabilizing factors preservatives, redox potential, competitive microflora, acidification, and a_w. Early in sausage ripening, nitrite added with curing salts inhibits the growth of many bacteria, such as *Salmonella* spp. Nitrite levels eventually decrease, but other bacteria continue to multiply and cause a reduction in redox potential. This inhibits the growth of aerobic organisms, while favoring lactic acid bacteria, which produce acid and reduce the pH of sausages. In long-ripened fermented sausages, a_w (≤ 0.90) becomes the main stabilizing factor, as numbers of lactic acid bacteria decrease with storage time, resulting in slight increases in acidification and redox potential. Other fermented IM meats are French saucisson, and Turkish pastirma.

Beef jerky or jerked beef (North America), lup cheong (China), biltong (South Africa), and carne de sol (South America) are examples of nonfermented IM meats which are produced in different regions of the world. Various combinations of factors contribute to product stability, such as a_w reduction by addition of salt and sugar, mild heat treatments, vacuum packaging, and preservatives. A variety of other ingredients, such as vinegar, pepper, coriander, and garlic, are used for flavor and shelf-life improvements in IM meats.

IM fruit pastes and purées, intended for use in candy, cookies, icecream, etc., have been developed by treating various fruits (apricots, peaches, papayas, etc.) with sulfur dioxide, dehydrating to $a_w \leq 0.70$ and employing extrusion techniques. Meal replacement bars (a_w of approximately 0.60) are IMF which contain a variety of ingredients such as milk protein, soya protein, fat, flour, starch, sucrose, salt, propylene glycol, and preservatives. Multitextured shelf-stable products have been introduced as well, including moist puddings in éclair-type shells; ready-to-eat breakfast cereals containing dried fruits; snack bars with fruit fillings; and cookies with crunchy exteriors and soft-moist interiors. Components of multitextured foods initially have different a_w and equilibrate to a common a_w during storage.

Factors Limiting IMF Stability

The significant effects of a_w on chemical reactions have important implications for the preservation of IMF. The shelf-life of IMF can be limited by nonenzymatic browning reactions, particularly the Maillard reaction which is a condensation reaction between the carbonyl group of a reducing compound, mainly sugar, and a free-amino group. The reaction can greatly affect food color due to the formation of brown pigments during processing and storage. It is also associated with toughening of texture, decreases in protein quality, and slight increases in a_w. Inversely, certain Maillard reaction products have bactericidal effects which contribute to product stability. (*See* **Browning**: Nonenzymatic; **Storage Stability**: Parameters Affecting Storage Stability.)

Generally, as for most deteriorative reactions, browning reaction rates increase exponentially with increasing a_w above the food monolayer due to an increased mobility of reactants (Figure 1). Maximum browning rates occur at approximately 0.65–0.75, which is in the IMF a_w range, and then the reaction decreases as a result of reactant dilution and product inhibition.

Variations in browning rate maxima are due to differences in the physical and chemical structure of IMF. Liquid humectants, such as glycerol or propylene glycol, reduce browning in IMF by shifting a_w maximas from approximately 0.7 to 0.4–0.5. These humectants appear to act as aqueous solvents, allowing greater reactant mobility at lower moisture contents. Maillard reaction rates also decrease with decreasing pH and the addition of sulfur dioxide or sulfites, which are used in IMF as browning inhibitors.

Free radical autoxidation of unsaturated lipids, and other compounds, is another important shelf-life problem of IMF. In the IMF a_w range, lipid oxidation rates generally increase with increasing water content (Figure 1) due to greater mobility of reactants, dissolution of precipitated catalysts, and swelling of solid

matrices, resulting in the exposure of new surfaces for reaction. Hydroperoxides, formed during lipid oxidation, decompose into secondary oxidation products, many of which result in rancidity in IMF. In IM meats, hydroperoxides and their products can react with proteins and reduce their solubility. Carbonyl-reducing compounds formed during oxidation reactions can act as reactants in the Maillard reaction.

Lipid oxidation in IMF is controlled by elaborate packaging technology, using vacuum, modified or controlled atmospheres to reduce oxygen contents; and using oxygen-impermeable and opaque films to reduce exposure to oxygen and light. In addition, IMF often contain lipid-soluble antioxidants (e.g., butylated hydroxyanisole, BHA), which combine with free radicals, or metal chelating agents, such as citric acid. Sulfur dioxide treatments are used to reduce losses of carotene and ascorbic acid in dried fruits. (*See* **Antioxidants**: Synthetic Antioxidants; **Chill Foods**: Effect of Modified-atmosphere Packaging on Food Quality; **Chilled Storage**: Packaging Under Vacuum; **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs.)

The major microbiological spoilage problem, limiting the shelf-life of many IMF, is xerophilic mold growth with the associated risk of mycotoxin production. Mold strains of *Aspergillus* and *Penicillium* occur widely in the environment and cause storage problems in many IMF (**Table 1**). Xerophilic yeasts do not produce toxins, but may cause fermentation and spoilage of some IMF.

Temperature fluctuations in the commercial production and distribution of IMF can cause condensation and increases in a_w on product surfaces, which are detrimental to product stability. This occurs when products are transferred between environments of different temperatures or are packaged while warm. In addition, temperature differentials, which lead to moisture migration within products, may occur in lighted, refrigerated display shelves. These effects have been associated with increased sporulation and mycotoxin production of molds, such as *Aspergillus flavus* and *A. parasiticus* at storage temperatures between 25 and 40 °C.

Progress and Future Developments

Some of the obstacles preventing large-scale developments of novel IMF are poor palatability due to the relatively high levels of humectants required to lower a_w, as well as consumer health concerns and legal difficulties associated with the use of antimicrobials. There is also a growing trend in industrialized countries towards minimally processed foods, which require refrigeration.

However, certain developing countries are interested in simple and inexpensive conservation processes based on combined factors, which give stable products without refrigeration. A better understanding of the underlying principles of hurdles in traditional IMF recently led to the development of high-moisture fruit products (HMFP) in Latin America. Similar to most IMF, high-moisture fruits are stabilized by several hurdles, including heat treatments, pH reduction, and the addition of antimicrobials; however, the predominance of the a_w hurdle for preservation is reduced. These innovative products are tropical fruits, whether whole, pieces, or purée; with a_w above 0.92; and stable without refrigeration. The developed processes are energy-efficient and allow a better utilization of tropical fruits. Other similar, higher-moisture products may attract further research and consumer interest, due to their reduced levels of humectants and improved organoleptic quality compared with IMF or fully dried products.

Certainly the enhanced knowledge of hurdle technology, allowing more effective combinations of preservative factors, will be used to improve the quality and safety of existing IMF and preserved foods in general. Of great benefit to this research area and to product design are the increased possibilities for predicting effects of combined factors by computer technology. As for other processed foods, well-defined guidelines of good manufacturing practice and Hazard Analysis Critical Control Point (HACCP) principles must be a priority in the production of IMF and other hurdle-stabilized products to insure food quality.

Another possible area of development is the use of edible coatings containing antimycotics or antioxidants to enhance microbial or oxidative stability, and reduce the overall concentration of additives in IMF. Food films, often based on a high-molecular-mass compound, such as starch, cellulose, proteins, waxes, or fats, reduce the diffusion of preservatives from food surfaces into food bulk. This can lead to improved surface microbial stability allowing for product formulations with slightly higher a_w, lower levels of humectants and improved palatability. Developments in barrier films, which reduce moisture transfer between components of different a_w, will also find applications in heterogeneous IMF.

In addition, research and development will continue to focus on naturally occurring food preservatives. For example, bacteriocins and other antimicrobial substances produced by a variety of lactic acid bacteria may find applications in IMF. Natural antioxidants from plant sources, such as rosemary (*Rosmarinus*), are also being actively

investigated for potential applications in food preservation. There is much research interest in encapsulation and controlled-release systems for increasing the effectiveness and range of application of many natural preservatives.

Finally, there will probably be more developments of IMF for industrial use, as for example concentrated food preparations (e.g., soup, sauce, or meal concentrates) that can be rehydrated, and directly used in fast foods, restaurants, and other food service industries. Certainly, novel IMF bakery products with a variety of characteristics, such as moist texture and long shelf-life, will continue to appear on the market.

See also: **Antioxidants**: Synthetic Antioxidants; ***Bacillus***: Occurrence; **Chilled Storage**: Use of Modified-atmosphere Packaging; Packaging Under Vacuum; **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Pasteurization**: Principles; **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage; Molds in Spoilage; Yeasts in Spoilage; **Storage Stability**: Parameters Affecting Storage Stability; **Sucrose**: Properties and Determination; **Water Activity**: Principles and Measurement; Effect on Food Stability

Further Reading

Baldwin EA (1999) Surface treatments and edible coatings in food preservation. In: Rahman S (ed.) *Handbook of Food Preservation*, pp. 577–610. New York: Marcel Dekker.

Davies R, Birch GG and Parker KJ (eds) (1976) *Intermediate Moisture Foods*. London: Elsevier Science Publishers.

Erickson LE (1982) Recent developments in intermediate moisture foods. *Journal of Food Protection* 45: 484–491.

Labuza TP and Saltmarch M (1981) The non-enzymatic browning reaction as affected by water in foods. In: Rockland LB and Stewart GF (eds) *Water Activity: Influences on Food Quality*, pp. 605–650. New York: Academic Press.

Leistner L (1999) Combined methods for food preservation. In: Rahman S (ed.) *Handbook of Food Preservation*, pp. 457–485. New York: Marcel Dekker.

Rockland LB and Beuchat LR (eds) (1987) *Water Activity: Theory and Applications to Food*. New York: Marcel Dekker.

Seow CC (ed.) (1988) *Food Preservation by Moisture Control*. London: Elsevier Science Publishers.

Simatos D and Multon JL (eds) (1985) *Properties of Water in Foods*, pp. 405–481. Dordrecht: Martinus Nijhoff.

Taoukis PS, Labuza TP and Saguy S (1997) Kinetics of food deterioration and shelf-life prediction. In: Valentas KJ, Rotstein E and Singh RP (eds) *Handbook of Food Engineering Practice*, p. 381. New York: CRC Press.

INTERNATIONAL UNION OF FOOD SCIENCE AND TECHNOLOGY

K A Buckle, The University of New South Wales, Sydney, Australia
J S Meech, Oakville, Ontario, Canada

WHAT IS IUFoST?

The International Union of Food Science and Technology (IUFoST) is the sole global, country-member organization that represents more than 200 000 food scientists and food technologists. It is a voluntary, nonprofit association that links the world's food scientists and food technologists through international cooperation and exchange of scientific and technical information. IUFoST is a full member of the International Council for Scientific Unions (ICSU), now the International Council for Science. It works closely with several United Nations agencies, such as the Food and Agriculture Organization, World Health Organization, Codex Alimentarius Commission and International Atomic Energy Agency, as well as other international organizations, to transfer technical knowledge and strategies for basic and applied research to the developing world.

IUFoST sponsors international and regional congresses, conferences, symposia, and workshops held in member countries, and promotes a safe, secure, and wholesome food supply for the world's increasing population.

Origin

The profession of food science and technology is generally agreed to have emerged in the USA in the early decades of the twentieth century. The subsequent formation of the (US) Institute of Food Technologists (IFT) in 1939 and its successful development and expansion led to the emergence of similar associations in many countries during the following 20 years. In 1960, the UK government and several British scientific societies hosted a conference in London in commemoration of the centenary of the UK Food and Drugs Act (1860) and the 150th

anniversary of the publication in 1810 of Nicolas Appert's book (*L'art de conserver, pendant plusieurs années, toutes les substances animales et végétales*) on the heat preservation of foods in sealed containers. In the preceding week, Professor John Hawthorn, then at the Glasgow Royal College of Science and Technology (later to become University of Strathclyde), convened a symposium on recent advances in food science, and discussions between several eminent UK and US food scientists raised the concept of an international food society. The first outcome of discussions on the formation of an international organization was the First International Congress of Food Science and Technology held in London in 1962 and organized by the Food Group of the Society of Chemical Industry.

The President of the Congress, Lord Rank, stated in his Presidential address: 'If the potentialities of... food science and technology are to... culminate in the peoples of the world receiving a sufficiency of food that is... appealing and nutritionally adequate, then there must be international collaboration.' Following the Congress, an *ad hoc* International Committee of Food Science and Technology (ICFoST) was formed with eminent scientists from several countries, to promote and foster international cooperation in food science and technology. A Second International Congress of Food Science and Technology was held in Warsaw, Poland, in 1966. By the time the Third International Congress of Food Science and Technology (*Science of Survival: SOS/70*) was held in Washington, DC, USA in 1970, ICFoST had developed a constitution and working arrangements for an International Union. Professor George Stewart of the University of California at Davis, USA, one of the pioneering scientists associated with the preparations of the previous decade, was the Founding IUFoST President.

Membership

IUFoST was inaugurated on 14 August 1970 with 20 nations as the founding membership (Table 1). The present membership is 64 (Table 2). Membership is open to all countries, and consists of one representative body, termed an 'adhering body,' from each country. The term 'country' is understood to be a defined geographical area, or areas, that has an independent science budget. Only one adhering body can represent each such area. Adhering bodies can be a national food science and/or food technology group, society, or institute composed of food scientists and/or food technologists and/or other professionals engaged in food science and/or technology; an academy of sciences, national research council, or similar organization of scientists; an intersociety committee or similar group representing two or more societies composed of food scientists, and/or food engineers, and/or food technologists; or a national committee representing food scientists, food engineers, and food technologists working in the field of food science and technology.

Table 1 IUFoST founding member countries 1970

Australia	Mexico
Canada	Netherlands
Finland	New Zealand
France	Philippines
Hungary	South Africa
India	Spain
Ireland	Sweden
Italy	Thailand
Japan	UK
Korea	USA

Table 2 IUFoST member countries 2001

Albania	France	Malawi	Singapore
Angola	Germany	Malaysia	Slovenia
Argentina	Greece	Mexico	South Africa
Australia	Hungary	Mozambique	Spain
Austria	India	Namibia	Sweden
Belgium	Indonesia	Netherlands	Switzerland
Botswana	Ireland	New Zealand	Taiwan
Bulgaria	Italy	Nigeria	Tanzania
Brazil	Japan	Norway	Thailand
Canada	Kazakhstan	People's Republic of China	Trinidad and Tobago
Chile	Kenya	Philippines	Uganda
Costa Rica	Korea	Poland	UK
Czech Republic	Kuwait	Portugal	USA
Denmark	Lesotho	Qatar	Vietnam
Egypt	Lithuania	Russia	Zambia
Finland	Macedonia	Saudi Arabia	Zimbabwe

Institutions may be admitted as Institutional Associate Members, and commodity associations and commercial enterprises as Corporate Associate Members.

In addition to member country adhering bodies, four Regional Groupings of adhering bodies have been formed within IUFoST, each with an IUFoST-approved constitution:

- European Federation of Food Science and Technology (EFFoST, formed in 1982);
- Federation of Institutes of Food Science and Technology in ASEAN (FIFSTA, formed in 1988);
- Eastern, Central and South African Associations of Food Science and Technology (ECSAAFoST, formed in 1991);
- La Asociación Latinoamericana y del Caribe de Ciencia y Tecnologia de Alimentos (ALACCTA, the Latin American and Caribbean Association of Food Science and Technology, formed in 1986).

Table 3 Member countries of IUFoST Regional Groupings

Regional grouping	*Member countries*
EFFoST	Austria, Belgium, Bulgaria, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Lithuania, Netherlands, Norway, Poland, Portugal, Spain, Switzerland, UK
FIFSTA	Indonesia, Malaysia, Philippines, Singapore, Thailand, Vietnam
ECSAAFoST	Angola, Botswana, Kenya, Lesotho, Malawi, Mozambique, Namibia, South Africa, Swaziland, Tanzania, Uganda, Zambia, Zimbabwe
ALACCTA	Argentina, Brazil, Chile, Costa Rica, Mexico, Trinidad and Tobago

ASEAN, Association of South-East Asian Nations; EFFoST, European Association of Food Science and Technology; FIFSTA, Federation of Institutes of Food Science and Technology in ASEAN; ECSAAFoST, East, Central and South African Associations of Food Science and Technology; ALACCTA, La Asociación Latinoamericana y del Caribe de Ciencia y Tecnologia de Alimentos (Latin American and Caribbean Association of Food Science and Technology).

Member countries of these groupings are shown in Table 3.

Purposes and Objectives

Vision

The vision of IUFoST is: 'To be recognized as the international agency that most effectively supports food professionals and other food sector constituencies, meeting their needs and aspirations by continuous refinement of a range of services, delivered according to the specific needs of regions, countries and of those we serve.'

Mission

The mission of IUFoST is: 'To be the leading international agency in developing and facilitating co-operative arrangements that bring sought-after services to food professionals and other food sector constituencies in different countries, thereby equipping them to maximize the benefits of food science and technology.'

Objectives

The primary purposes or objectives of IUFoST are to encourage and foster:

- international cooperation and exchange of knowledge and ideas among food scientists and technologists;
- further development of and support for food research;
- progress in the fields of theoretical and applied food science for improvements in the processing, manufacturing, preservation and distribution of food;
- the education and training of food scientists and technologists; and
- development of both individual professionalism and professional organization among food scientists and technologists.

Specific objectives aim:

- to encourage world-wide exchange of ideas and experience in the scientific disciplines and technological fields mentioned above, and specifically by sponsoring international congresses and other such regional and/or world-wide conferences and symposia as may usefully serve this purpose;
- to establish commissions and/or other functioning bodies, as may be required, to deal with scientific and technical problems in these fields on an international scale, and to provide other means of effective international cooperation;
- to maintain, on a world-wide basis, cooperation and coordination of effort with other international scientific and technological bodies with common or similar interests; and
- to engage, as appropriate, in the publication of journals, newsletters, proceedings of congresses, symposia and conferences, and reviews of achievements in food science and technology.

In dealing with external constituencies, IUFoST is:

- committed to upholding the highest professional standards of scientists working in food science and technology;
- committed to nonaligned/global leadership in food science and technology;
- committed to a safe, wholesome, and accessible food supply;
- ethically unquestionable, unbiased, balanced, independent, and honest;
- a positive force for desirable change, forward-looking, and professional;
- sensitive to, and inclusive of, cultural differences;
- financially independent of food industry lobby groups; and
- authoritative/learned in both scientific and industrial food issues.

In dealing with external agencies, groups, and individuals, IUFoST recognizes that it has important relationships with, and responsibilities to, a number of diverse organizations and groups: adhering bodies of IUFoST; regional groupings within IUFoST; international agencies (such as FAO, WHO, Codex, IAEA); community and consumer organizations; food professionals; students; international scientific

organizations such as ICSU; government agencies; industry organizations and groups; opinion formers; associate members; and countries and areas with an insecure food supply.

Structure and Operation

Structure

The principal structures within IUFoST are the General Assembly, Governing Council, Management Committee, Scientific Council, and International Academy of Food Science and Technology (IAFoST). Advisory and other committees are also constituted (Table 4). The Officers of IUFoST are the President, the President-elect, the Scientific Council Chair, and the Secretary-General, who also serves as Treasurer and is an employee selected by the Management Committee (Table 5). The Past President serves as an *ex officio* member for the two years following tenure as President. No more than two officers may be citizens of the same nation.

General Assembly Members of the General Assembly consist of delegates with voting rights representing each adhering body (one for each country) on the basis of its annual financial contribution. The number of delegates and ordinary votes (range 1–7) is determined on the basis of the annual fee payable.

The General Assembly meets biennially (previously every four years until 1999) in conjunction with each World Congress of Food Science and Technology. It has final authority with respect to all matters of official policy for and action by IUFoST. It is responsible for setting policy and broad strategy; electing officers (except the Secretary-General) and the Governing Council for a two-year term; approving the fee schedule for adhering bodies; appointing the Audit, Constitution and Nominations Committees; electing the Scientific Council from candidates nominated by

Table 4 Composition of IUFoST governing structures

General Assembly	Governing Council and 1–7 Delegates of adhering bodies (one for each country) allocated on the basis of a country's financial contribution
Governing Council	President, President-elect, Chair of Scientific Council, Secretary-General, Immediate Past President (*ex officio* and nonvoting)
Management Committee	President, President-elect, Chair of Scientific Council, Immediate Past President (*ex officio* and nonvoting), Secretary-General (*ex officio* and nonvoting), Chair of Audit Committee (*ex officio* and nonvoting)
Scientific Council	Chair, Chair-elect, Past Chair, two Scientific Councillors, President (*ex officio* and nonvoting), Secretary-General (*ex officio* and nonvoting)
International Academy of Food Science and Technology	Elected group of independent distinguished food scientists and technologists (Fellows) from whom IUFoST may draw nonaligned advice on scientific matters
Audit Committee	Five members, each from a different country and not holding an elected position
Nominations Advisory Committee	President-elect (Chair) and four members, each from a different country and not holding an elected position
Constitution Advisory Committee	Five members, each from a different country and not holding an elected position

Table 5 Officers of IUFoST 1970–2003

Years	*President*	*Secretary-General*	*Treasurer*
1970–74	G. Stewart (USA)	E. von Sydow (Sweden)	E. von Sydow (Sweden)
1974–78	J. Hawthorn (UK)	E. von Sydow (Sweden)	E. von Sydow (Sweden)
1978–83	J.H. Hulse (Canada)	J.F. Kefford (Australia)	J.F. Diehl (Germany)
1983–87	R.L. Hall (USA)	J.F. Kefford (Australia)	J.F. Diehl (Germany)
1987–91	E. von Sydow (Sweden)	D.E. Hood (Ireland)	F.E. Escher (Switzerland)
1991–95	D.E. Hood (Ireland)	1991–92 G.E. Timbers (Canada) 1992–95 J.P. Meyers (Canada)	F.E. Escher (Switzerland)
1995–99	P.A. Biacs (Hungary)	J.P. Meyers (Canada)	F.E. Escher (Switzerland)
		Secretary-General/Treasurer	*Chair, Scientific Council*
1999	L. Marovatsanga[a] (Zimbabwe)	J.S. Meech (Canada)	D.R. Lineback (USA)
1999–2003	W.E.L. Spiess (Germany)	J.S. Meech (Canada)	D.R. Lineback (USA)

[a]Deceased 1999.

IAFoST; maintaining and approving changes to the Constitution and By-laws; and ratifying decisions of the Governing Council. It considers proposals involving international co-operative scientific programs and activities, and approves methods to finance IUFoST activities and reports of expenditures.

Governing Council The Governing Council (previously the Executive Committee) meets annually to set specific objectives and operational strategies for IUFoST. It has general oversight of the activities of the Officers and Management Committee, adopts budgets upon the recommendation of the Management Committee, sets fees for adhering bodies, appoints external auditors, selects the time and place of international congresses and other such meetings sponsored by IUFoST, appoints advisory committees, and ratifies ongoing actions of the Scientific Council and Management Committee. It often meets in conjunction with symposia organized by adhering bodies or regional groupings.

Management Committee The Management Committee implements strategies to achieve the objectives set by the Governing Council, including exercising general supervision of the activities of the Secretary-General, approving or modifying membership of working groups and committees, preparing annual reports and expenditures, and approving country or individual member applications. It meets at least every three months.

Scientific Council The Scientific Council is independent of the Governing Council and Management Committee. It is responsible for maintaining the scientific standard and integrity of all IUFoST activities, including publicly released scientific statements and the award of prizes. It recommends membership of working groups and scientific commissions to carry out specific tasks, and supervises the scientific output of such groups. It meets at least annually.

International Academy of Food Science and Technology The General Assembly, at its 7th meeting in Budapest in 1995, accepted the recommendation of the (then) Executive Committee to form IAFoST. The Academy, which was formed in 1997, is a learned society composed of elected members (Fellows) from all parts of the world. Fellows are usually, but not necessarily, affiliated with adhering bodies of IUFoST. The Academy statutes provide that the President, Secretary-General, and all past Presidents and Secretaries-General are automatically Fellows. The first meeting of the Academy was held at the 10th World Congress of Food Science and Technology held in Sydney in October 1999. The Academy members (Fellows) represent no organization and serve as independent persons to provide a pool of scientific expertise in food science and technology from which IUFoST may draw advice. Nominated candidates are elected in advance of each General Assembly, requiring an approving vote by two-thirds of the voting Fellows. The Academy nominates to the General Assembly candidates for the Scientific Council, provides scientific advice to the President and Scientific Council, and provides scientific advice to the President and Scientific Council, and proposes membership of working groups to the Scientific Council.

Audit Committee The Audit Committee establishes appropriate accounting control procedures, recommends external auditors, and ensures their independence. The Chair presents the annual audit of the IUFoST's financial affairs to the Governing Council.

Nominations Advisory Committee The Nominations Advisory Committee ensures that there are nominated candidates of sufficient merit for all positions to be filled, and that the candidates collectively meet the criteria for regional representation.

Constitution Advisory Committee The Constitution Advisory Committee formulates and presents to the Governing Council and thence to the General Assembly changes to the Constitution and By-laws for the more effective fulfillment of IUFoST's policies and the conduct of its business. It also provides advice to the General Assembly on the suitability of proposed changes to the Constitution, as proposed by adhering bodies.

Operation

The *Constitution* of the IUFoST was completely revised and accepted by the General Assembly in October 1999 at the 10th World Congress in Sydney. The Governing Council replaced the Executive Committee, and a Management Committee was established to execute IUFoST business between Governing Council meetings.

The General Assembly also adopted a *Strategic Plan* to reposition the organization, so that it can operate in a financially viable manner, while at the same time adapting and growing to meet the challenges of the years ahead.

An *Operational Plan* describes the accomplishments and challenges of IUFoST, and outlines procedures taken by IUFoST to develop a management system of *service delivery centers* (SDCs), each with measurable objectives and a budget to achieve these.

With the review of the Strategic Plan in 1999, the Governing Council established six SDCs identified by IUFoST delegates at the 8th General Assembly in October 1999. The SDCs are:

- SDC1: Continuing Professional Development (conferences, workshops);
- SDC2: Information Services (journals, technical publications, internet services);
- SDC3: Advice and Expert Opinion (consulting, expert advice, international databases);
- SDC4: Food Science and Technology Development (new technologies applications, regulatory technical standards, scientific needs);
- SDC5: Development of Professionalism (international standards on professional skills, strategy development, relation to other unions, position papers, status reports); and
- SDC6: Marketing and Support Services (promotion of collaboration/interaction between researchers in different IUFoST countries).

Activities

IUFoST conducts a wide range of activities in order to achieve its objectives of fostering international cooperation and exchange of knowledge and ideas between food scientists and technologists throughout the world. These are conducted in conjunction with the adhering bodies and their regional groupings, or with a range of other local, national, or international organizations.

Congresses

Until 1999, World Congresses of Food Science and Technology were conducted approximately every 4 years in conjunction with the adhering body of the host country (Table 6). From 2001, they have been held biennially, commencing with Congress 11 in Seoul, South Korea.

Table 6 World Congresses of Food Science and Technology

Congress No.	*Year*	*Location*
1	1962	London, UK
2	1966	Warsaw, Poland
3	1970	Washington, DC, USA
4	1974	Madrid, Spain
5	1978	Kyoto, Japan
6	1983[a]	Dublin, Ireland
7	1987	Singapore
8	1991	Toronto, Canada
9	1995	Budapest, Hungary
10	1999	Sydney, Australia
11	2001	Seoul, South Korea
12	2003	Chicago, USA
13	2006	Nantes, France

[a]Congress dates were adjusted to avoid clashes with International Union of Nutritional Sciences Congresses.

Regional Meetings

IUFoST has been associated with over 200 conferences, symposia, seminars, workshops, training programs, and short courses since its inception in 1970, covering diverse topics in areas such as food science, food technology, food engineering, food microbiology, human nutrition, national and international food law, food toxicology, product development, and food security. Proceedings of many of these activities have been published.

Declarations on Food Security

IUFoST has long recognized the role of food science and technology in improving food security in the developing world, especially in Africa and Eastern Europe. The *Victoria Falls Declaration*, supporting the plight of consumers in the developing world, especially subSaharan Africa, was approved by delegates at ICSAFoST '94, the first international food conference held in Zimbabwe in September 1994. The declaration was based on the original declaration from the FAO/WHO International Conference on Nutrition held in Rome in December 1992, which was endorsed by 159 states and the European Economic Community (EEC). Following the 9th World Congress of Food Science and Technology in Budapest, Hungary in July–August 1995, the *Budapest Declaration*, based on the FAO/WHO and Victoria Falls declarations, was approved by delegates to the IUFoST General Assembly. It states:

1. We, the delegates to the 7th General Assembly of the International Union of Food Science and Technology (IUFoST), in reference to the Joint FAO/WHO International Conference on Nutrition (Rome, 1992) and its World Declaration on Nutrition, declare our determination to work for the *elimination of hunger and reduction of all forms of malnutrition* throughout the world. *We recognize that access to nutritionally adequate and safe food is the right of each individual*. We also declare our commitment to work with all other organizations *to ensure sustained nutritional well-being for all people in a peaceful, just and environmentally safe world*. In this endeavor, we recognize the central role of food science and technology in ensuring the year-round availability of the quantity and variety of safe and wholesome foods necessary to meet the nutritional needs of the world's growing population.
2. *Despite appreciable worldwide improvements in life expectancy, adult literacy and nutritional*

status, we all view with the deepest concern the unacceptable fact that about 800 million people in developing countries – twenty percent of their combined population – still do not have access to enough food to meet the basic daily needs for nutritional well-being. We are especially concerned *by the high prevalence and increasing numbers of malnourished children under five years of age in parts of Africa, Asia and Latin America and the Caribbean as well as other parts of the world. Moreover, 2000 million people, mostly women and children, are deficient in one or more micronutrients; babies continue to be born mentally retarded as a result of iodine deficiency; children go blind and die of vitamin A deficiency; and enormous numbers of women and children are adversely affected by iron deficiency. Hundreds of millions of people*, particularly young children, pregnant women, the immunocompromised and the elderly, *also suffer from communicable and non-communicable diseases caused by contaminated food and drinking water. At the same time, chronic non-communicable diseases related to unbalanced dietary intakes often lead to premature deaths*. Taken together, these diseases result in enormous health care costs as well as important losses due to reduced economic productivity. We recognize that the contributions of food science and technology are essential in solving many of these problems, especially those related to inappropriate dietary intakes, food hazards and micronutrient deficiencies.

3. *We recognize* that *poverty* and limited access to *education, which are* commonly *the effects of underdevelopment*, are often major *causes of hunger and undernutrition. There are poor people in most societies who do not have adequate access to food, safe water and sanitation, health services and education, which are the basic requirements for nutritional well-being*. To overcome these problems, there is an urgent need to strengthen the food science and technology base to support food and agriculture sectors, especially in low-income, food-deficit countries, in order to expand and diversify food supplies, create income-earning opportunities and generate local resources for development.
4. *Slow progress in solving nutrition problems in many countries reflects the lack of human and financial resources, institutional capacity and policy commitment needed to assess the nature, magnitude and causes of nutrition problems and to implement concerted programs to overcome them. Basic and applied scientific research, as well as food and nutrition surveillance systems, are needed to more clearly identify the factors that contribute to the problems of malnutrition and the ways and means of eliminating these problems, particularly for women, children and aged persons*. We recognize that food science and technology inputs together with modern biotechnology are essential in many basic and applied scientific research programs. We further recognize our responsibility to promote public understanding of issues involving food science and technology, so that public support and acceptance allow the timely introduction of beneficial applications of present and emerging new technologies.
5. *The right of women and adolescent girls to adequate nutrition is crucial. Their health and education must be improved. Women should be given the opportunity to participate in the decision-making process and to have increased access to and control of resources*. We recognize that, in many parts of the world, women have a particularly important role to play in food production, handling and preparation.
6. *We acknowledge the importance of further liberalization and expansion of world trade, which would increase foreign exchange earnings and employment in developing countries*. We recognize the important role of food science and technology in the establishment of standards, guidelines and other recommendations of the Codex Alimentarius Commission, including their acceptance by national food authorities.
7. We recognize aspects of food science and technology as integral components, both in the preparation of National Plans of Action for Nutrition and in the formulation of economic development plans and identify the following areas as being especially significant:
   - promotion of the safety and quality of all foods;
   - reduction of pre- and postharvest food losses;
   - adaptation and improvement of traditional foods and processes;
   - beneficial application of biotechnology and other new technologies;
   - development and dissemination of improved knowledge of food composition;
   - promotion of domestic and international food trade;
   - development of food materials with improved functionality;
   - more efficient and environmentally sustainable food production and processes; and
   - education in nutrition, food science, and technology at all levels.

8. We support the *active cooperation among governments, multilateral, bilateral and non-governmental organizations, academic bodies, the private sector, communities and individuals to eliminate progressively the causes that lead to the scandal of hunger and all forms of malnutrition in the midst of abundance*. We, the adhering bodies of IUFoST, representing professional food scientists and technologists from 58 nations throughout the world, are fully committed to cooperate and collaborate with all parties in achieving these objectives.
9. We recognize that efforts to ensure safe and adequate diets for the people of the world can only succeed when social, cultural, political, economic and educational components of the problem are also addressed. We further recognize that, in addition to adequate resources to assure household food security, consumers must have an adequate knowledge of the basics of food safety and nutrition to protect their health and to promote their nutritional well-being.
10. *With a clear appreciation for the intrinsic value of human life* and in recognition of the right of all individuals to safe and nutritionally adequate food, we adopt this Declaration and affirm our commitment as food scientists and technologists *to protect and promote the nutritional well-being for all*.

*Note: Words in italics are included in the FAO/WHO Declaration.

Publications

The IUFoST newsletter was first published in 1970. It became *Newsline* with the 25th edition in March 1993 when it was first published as part of *Lebensmittel-Wissenschaft & Technologie (LWT, Food Science and Technology)*, a scientific journal owned by the Swiss Society of Food Science and Technology and published by Academic Press. *Newsline* is now published three times per year and is available on the IUFoST website (http://www.iufost.org). The Elsevier journal *Trends in Food Science and Technology* is also an official journal of IUFoST and EFFoST.

In September 2000, IUFoST and its US adhering body, the IFT, jointly launched the online electronic magazine *The World of Food Science*, directed at food scientists and food technologists at all levels in the international arena. Its mission is:

> To inform the world community of food scientists, technologists, food regulators, governmental advisors, reporters and others interested in the promotion of a safe and sufficient food supply through timely news of events, business activities such as acquisitions, new products and new markets, regulatory measures, government policies, scientific and technological advances, and regional business reports and developments. The magazine has a commitment to be a premier vehicle for information about food science and technology, to encourage interaction of food scientists in the international community and to promote the fields of food science and technology.

An international Editorial Advisory Board contains members representing both IUFoST and IFT, together with a senior staff officer of both organizations.

Relationship to other Organizations

ICSU

IUFoST was admitted as a Scientific Associate member of the International Council of Scientific Unions (ICSU) in 1981. During a hard-fought campaign for recognition of food science, which was led by Dr Ted Hood as President in the early 1990s, the direct support of 12 International Unions and 12 National Academies of Science was eventually achieved. IUFoST was subsequently elected to full membership during the 25th General Assembly of ICSU held in Washington, DC in 1996.

FAO/WHO/United Nations

IUFoST has close relationships with both the FAO and WHO and has a standing invitation to contribute to food-related programs and to make recommendations on FAO and WHO policies and activities. It contributes to activities of the Codex Alimentarius Commission, United Nations University and other UN agencies such as the Organization for Economic Cooperation and Development (OECD).

IUNS

IUFoST is collaborating with IUNS to compile a dictionary or directory of professional terminology in food science/technology for nonEnglish speaking experts.

See also: **Food and Agriculture Organization of the United Nations**; **World Health Organization**

Further Reading

Anon (2000) International Union of Food Science and Technology. In Merle E (ed.) *ICSU Year Book 2000/2001*, pp. 186–187. Paris: International Council for Science.

Anon (2000) IUFoST activities. *Food Technology* 54: 24.

Blanchfield JR (2000) The best-kept secret: IUFoST revealed. *Food Technology* 54: 48–52.

Fennema OR (1995) IUFoST – An important organization with lofty goals and limited means. *Food Technology* 49: 12–13.

IODINE

Contents

Properties and Determination

M R L'Abbé, Nutrition Research Division, Health Canada, Ottawa, Ontario, Canada

Background

Iodine is a blueish-black solid of the halogen family of elements which readily sublimes to the deep violet I_2 vapor. The iodine concentration of human foods is usually low and extremely variable. Foods naturally high in iodine include seaweeds and marine fish. In addition, iodine can be added to foods from sources such as fertilizers, animal feed supplements, iodine-containing sanitizers and disinfectants or through the use of iodine-containing food additives and food colorants. Iodine is also added in many countries to foods such as table salt for iodine fortification for the prevention of iodine deficiency disorders (IDD) or goitre. As iodine is volatile and can be easily oxidized or reduced, care must be taken to prevent volatility losses during analysis. The analyst must also be aware of the chemical form and oxidation state of iodine under various conditions to avoid losses.

Physical and Chemical Properties

Iodine (I), atomic number 53, atomic weight 126.91, is a nonmetallic element of the halogen family (group VIIA) after fluorine (F), chlorine (Cl), and bromine (Br) in the periodic table. Of these, iodine is the heaviest and least electronegative. Solid iodine is a bluish-black crystalline solid with a metallic luster in the form of shiny flakes which can be readily crushed into a fine powder. Iodine crystallizes in an orthorhombic structure and is arranged in a unit cell of eight atoms arranged as a symmetrical bipyramid.

Solid iodine readily sublimes, as evidenced by its pressure below melting point (Table 1). The vapor is made up of diatomic iodine molecules (I_2) which dissociate readily at elevated temperatures. Iodine vapor is characterized by its familiar violet color. The specific gravity of I_2 vapor is approximately nine times that of air. Solid iodine is only slightly soluble in water and no hydrates form on dissolution, although it readily dissolves in many organic solvents. The color of iodine in solution varies with the solvent; aliphatic hydrocarbons and carbon tetrachloride give violet solutions, aromatic hydrocarbons give pink or brownish pink solutions, and alcohols, water, ethers, and some amines give brown colors.

The only stable isotope of iodine has a mass of 127. There are 22 other isotopes having masses between 117 and 139, with 14 of these yielding significant radiation. The electron configuration of the iodine atom is $4d^{10}5s^25p^5$. The chemistry of iodine is rather complex as it can exist in a number of valence states, with its principal oxidation states being -1, $+1$, $+3$, $+5$, and $+7$ and an oxide IO_2, with an oxidation state of $+4$. All these forms are thermodynamically stable except the $+4$ state. Iodine dissolves without reaction in concentrated sulfuric acid and with concentrated nitric acid it reacts to form iodine pentoxide; in acidic solution it has mild oxidizing properties. Iodine readily forms covalent iodides of metals such as aluminum, tin, titanium, and many organic substances, but not with S, Se, and the noble gases, and reacts only indirectly with C, N, O, and some noble metals such as Pt. Iodine as a vapor or solid reacts

Table 1 Selected physical properties of iodine[a]

Property	
Color	Bluish black (solid) Varies with solvent[b] (liquid) Violet (gas)
Melting point, °C	113.6
Boiling point, °C	185 (liquid)
Crystal structure, 4 mol I_2	Orthorhombic
Density, g ml^{-1} (20 °C)	4.93
Vapor pressure, kPa (25 °C)	0.04133
Vapor pressure, kPa (113.6 °C)	12.0655
Solubility, g kg^{-1} (25 °C), water	0.34
Solubility, g kg^{-1} (25 °C), benzene	164.0
Solubility, g kg^{-1} (25 °C), chloroform	49.7
Solubility, g kg^{-1} (25 °C), ethyl alcohol	271.1
Solubility, g kg^{-1} (25 °C), *n*-hexane	13.2

[a]Solid form except where indicated otherwise.
[b]Color varies with solvent; see text for details.

readily with Mg, Ca, Al, Zn, Sn, Ni, and Fe, forming the corresponding iodides. The metal iodides and polyiodides are bases, whereas the iodine halides are acids.

Sources of Iodine

Iodine is one of the rarest nonmetallic elements on earth. Although not abundant, iodine is distributed widely in rocks, soils, waters, plants, animal tissues, and foodstuffs. Of these, only a few substances characteristically contain large quantities of iodine – these being seaweeds, sponges, and corals, underground waters from deep wells, and the vast natural deposits of sodium nitrate found in northern Chile which contain high amounts of iodine. Sea water contains the largest total amount of iodine, mostly in the form of iodide and iodates. Iodine is readily volatilized by the action of sunlight and heat. Sunlight oxidizes iodide to iodine and about 400 000 tons iodine of escape from the oceans into the air each year. Atmospheric iodine is then deposited on land and vegetation. Many soils are low in iodine, particularly crystalline soils exposed to glaciation and leaching, and soils lacking humus cannot retain iodine. Heavy rains also promote iodine-deficient soils, as does frequent flooding of low-lying coastal regions. Areas of low iodine and consequently a higher prevalence of iodine-deficiency goiter have been located on every continent throughout the world.

Iodine has a wide range of chemical and related industrial uses, some of which can make it into the food supply. In 1996, approximately 39% of iodine production was used for sanitation, pharmaceuticals 24%, heat stabilizers 13%, catalysts 9%, animal feed 7%, and other uses 8%.

Water and Soil

Sea water is high in iodine and contains approximately 50 $\mu g\,l^{-1}$, mainly as iodide salts, although the level in oceans can vary widely (Table 2). The level of iodine in drinking water is much lower than that of sea water and usually reflects the iodine content of the rocks and soils in the region, although water from deep wells can provide significant amounts of iodine. Iodine in soil is in the form of iodide ions which are readily oxidized by sunlight to volatile elemental iodine, which is then released into the atmosphere, thereby reducing soil levels in many areas. Locally grown plants and crops and meat and milk products usually reflect the iodine content of soils and hence, foods from large parts of the world are iodine-deficient unless fortified with iodine or iodine is added to animal feeds or added during food processing.

Dietary Sources

Foods The iodine concentration of human foods is extremely variable. These variations are due to a number of factors, including differences in the iodine content and availability of iodine in the soil and water, as described above. In addition, iodine is added inadvertently to foods from sources such as fertilizers, animal feed supplements, iodine-containing sanitizers and disinfectants, or through the use of iodine-containing food additives and food colorants. Iodine is also added in many countries to foods such as table salt for iodine fortification for the prevention of iodine-deficiency disorders (IDD) or goiter.

Although the concentration of iodine in foodstuffs is extremely variable, oilseed protein supplements (soybean, cottonseed, linseed, and peanut) typically contain 0.11–0.20 p.p.m. iodine, whereas cereal grains range from 0.04 to 0.10 p.p.m. iodine. Milk, milk products, eggs, and meat usually contain more iodine than plant foods, although actual levels may reflect iodine provided to the animals either naturally, or through feed supplements, veterinary uses, or during food processing.

The food sources containing the most iodine tend to be marine products, although in recent dietary surveys in some countries, dairy products and bakery products were found to provide significant amounts of iodine (Table 3). Ocean fish are especially rich in iodine and may contain 300–3000 p.p.b. iodine compared to 20–40 p.p.b. for freshwater fish. No plants are known to preferentially, concentrate iodine although seaweeds are particularly rich sources

Table 2 Typical iodine levels in goitrogenous and nongoitrogenous regions

	Area	
	Goiter-free areas	*Endemic goiter area*[a]
Drinking water ($\mu g\,l^{-1}$)	5–15[b] (4000–16 000)[c]	0.1–2
Sea water ($\mu g\,l^{-1}$)	50	
Air ($\mu g\,m^{-3}$)	0.7	0.02
Soil ($\mu g\,kg^{-1}$)	300[d]	< 100
Dietary intakes ($\mu g\,day^{-1}$)	100–150[e]	< 25
Milk ($\mu g\,kg^{-1}$)	50–1000	< 25
Bread ($\mu g\,kg^{-1}$)	15	< 5
Animal tissues ($\mu g\,kg^{-1}$)	100–400	< 50

[a]Areas of moderate to severe iodine-deficiency disorder (IDD) according to the World Health Organization (1996) classification for prevalence (severity) of IDD; median urinary iodine levels ($\mu g\,l^{-1}$) as mild (50–99), moderate (20–49), or severe (<20).
[b]In some areas, levels as high as 40–100 $\mu g\,l^{-1}$ have been reported.
[c]Final concentration in water after using iodine preparations for water disinfection when used at recommended levels; recommended for short-term use only.
[d]Highest concentrations found in Chilean nitrate deposits (0.5–2 g kg^{-1}).
[e]Intakes should be higher (200–300 $\mu g\,day^{-1}$) in the presence of large amounts of goitrogens.

(50–4500 $\mu g\,g^{-1}$) due to the ocean environment in which they grow. In Canada, milk and dairy products contributed the greatest amounts to intake (239 μg day^{-1}), followed by cereals (84 μg day^{-1}) and meat, eggs, and fish (28 μg day^{-1}).

As noted above, the iodine content of foodstuffs is often modified by food-processing technology and manufacturing practices. Iodine and its compounds are used in a variety of food-related applications, including nutrient fortification (e.g., iodized salt), food additives (e.g., dough conditioners, maturing agents, and food colorants), agricultural chemicals (e.g., herbicides and fungicides), animal feed supplements, and veterinary drugs (e.g., ethylenediamine dihydroiodide, EDDI, which is 80% iodine and is used to treat and/or prevent foot rot in cattle), and sanitizers (e.g., iodophores) (Table 4).

Iodine-containing supplements in the feed of dairy cattle, iodine-containing veterinary pharmaceuticals, teat-dips and udder washes, as well as iodophores used as sanitizing agents in dairy-processing establishments can contribute to high iodine levels in dairy products in countries where such practices are prevalent. Milk from dairy cows where no iodophor sanitizers or iodine-containing veterinary pharmaceuticals are used usually contains less than 50 $\mu g\,l^{-1}$ compared to 200–800 $\mu g\,l^{-1}$ or higher measured in some recent surveys where these practices are common. Iodine concentrations in milk are usually higher in winter than summer, likely reflecting the greater use of supplements and formulated feedstuffs compared to pasture feeding. Very high levels of iodine in milk have been observed with the feeding of relatively high amounts of EDDI (which is 80% I). For example, the

Table 3 Comparison of iodine intakes in various countries by food group (μg day^{-1})[a,b,c]

Food group	*Canada*[d]	*UK*[e]	*China*[f]
Milk and dairy products	239	91	1
Meat, fish, poultry, eggs	28	38	21
Breads, grains, cereal products	84	4	66
Fruits and fruit products	1	3	1
Vegetables, including potatoes	2	13	38
Fats and oils	2	1	NR[g]
Nuts and legumes	<1	0	5
Sugars and condiments	14	12	<1
Beverages	Included above	17	5
Estimated total intake[c] (μg per person per day)	371	180	166

[a]Values are for foods as consumed.
[b]Comparisons by food groups are only approximate as the compositions of the various food groups are not exactly the same in different countries.
[c]Does not include iodine from iodized salt, iodized oil, iodine supplements, or from drinking water.
[d]Calculated from Fischer PWF and Giroux A (1987) Iodine content of a representative Canadian diet. *Journal of the Canadian Dietetic Association* 48: 24–27.
[e]Calculated from Ministry of Agriculture, Fisheries and Food (2000) Food Standards Agency UK–1997 *Total Diet Study–Fluorine, Bromine and Iodine. Food Surveillance in Formation Sheet, 127*. London: MAFF.
[f]Average from five regions, from Hou X, Chai C, Qian Q, Lin G, Zhang Y and Wang K (1997). The study of iodine in Chinese total diets. *Science of the Total Environment* 193: 161–167.
[g]NR, not reported.

Table 4 Some examples of other sources of iodine contributing to food iodine content or to iodine intakes

Use	*Example (s)*
Therapeutic drugs	Preparations to treat thyroid disorders, radiopharmaceuticals and contrast media, topical antiseptics (povidone-iodine solutions and ointments, betadine – an organically bound iodine), potassium iodide used as an expectorant in cough and cold medications
Fortification	Iodized salt, iodized oil
Food additives	
Dough conditioners	Potassium iodate (KIO_4) in bread and bakery products
Food colorants/dyes	Erythrosine (red dye #3; 2,4,5,7-tetraiodofluorescein)
Agricultural chemical residues	4′-hydroxy-3′5′-diiodobenzoic acid (herbicide)
Veterinary drug residues	Ethylenediamine dihydroiodide (EDDI)
Animal feed supplements	Provided as salt licks, or added as part of animal feed supplement. Dietary requirements for livestock range from 0.14 (swine) to 0.5 (beef and dairy cattle) mg kg^{-1}
Sanitizers	Iodophores used in milking equipment, dairies, food processing, and restaurant grills
Water purification	Effective at 0.5–1 p.p.m. (although water purification with iodine can often contribute 1000–2000 $\mu g\,l\,day^{-1}$)

iodine concentration in milk of cows fed 40 mg EDDI was 361 $\mu g\,l^{-1}$, compared to 1559 $\mu g\,l^{-1}$ in milk from cows given 162 mg EDDI. In Scandinavian countries (except Iceland), cows' milk is a good source of iodine because cattle fodder is iodinated by law.

Iodine fortification and supplements Another potential large contributor to iodine intakes is iodized table salt, although iodized salt is not universally available, nor is fortification mandatory in many countries. According to the World Health Organization in 1994, about 1600 million people were at risk of IDDs because they live in iodine-deficient environments. Of these, 656 million had goiter and 43 million had some degree of mental defect, including 11.3 million overt cretins. Endemic goiter is the most visible sign of iodine deficiency, but impairment of nervous system development and function is the most serious consequence of iodine deficiency. Iodine deficiency is still the world's greatest single cause of preventable brain damage and mental retardation.

In order to combat IDD, iodized table salt has been used in many countries since the 1920s and can contribute significant amounts of iodine to daily intakes. A variety of other vehicles have also been used for iodine supplementation of populations whose dietary intakes of iodine are low. For example, injection with iodized oil has been used in Papua New Guinea and was found to be particularly effective for treatment of goiter in remote mountain villages. In some countries teas, rice, or oils have been fortified with iodine, although problems have been encountered with non-uniform intakes, stability, or bioavailability of some of these forms. Before the advent of iodine fortification, prevalence rates of endemic goiter in many parts of the world were high and still remain so in areas without universal salt fortification or iodine supplement programs.

Iodized salt In Canada, iodide is added at the rate of 76 $\mu g\ I\,g^{-1}$ table salt while, for example, in Norway iodide is added at 5, France 15, Finland 25, Sweden 50, USA 100, and in China at 12–30 $\mu g\ I\,g^{-1}$ salt. In Iceland salt is not iodized because iodine intakes are already sufficiently high (300–350 $\mu g\,day^{-1}$) due to the high consumption of salt-water fish. Intakes of table salt also vary greatly in population surveys. In Canada, some individuals use virtually no table salt, compared to many others who consume more than 2 $g\,day^{-1}$. This contrasts greatly with countries such as China where the consumption of table salt averages 10–30 g per person per day.

Goitrogenic substances Goitrogenic substances are substances that have antithyroid activity and interfere with thyroid hormone production or utilization. The most important goitrogens are thiocyanates, which can compete with iodine for uptake into the thyroid gland, especially when iodine intakes are low. Many goitrogenic substances or their precursors are present in plants of the *Brassica* spp. (the crucifer or mustard) family and at least 10 other dicotyledonous plant families. *Brassica* species such as cabbage, broccoli, cauliflower, rutabaga, mustard, Brussels sprouts, and turnip contain glucosinolates (previously called thioglucosides) which are hydrolyzed to form isothiocyanates, nitriles, and thiocyanates. About 100 different kinds of glucosinolates are known to exist in the plant kingdom, but only about 10 are present in *Brassica*; high levels are also found in a number of seed meals, e.g., some rapeseed and soybean meals. The first toxic effects of isothiocyanates and other hydrolytic products from glucosinolates are goiter and a general inhibition of iodine uptake by the thyroid.

Goiter and IDDs have also been attributed to cassava consumption in many parts of the underdeveloped world. Cassava is a widely grown root crop which accumulates cyanogenic glucosides, linamarin, amygdalin, and lotaustralin. Linamarin accounts for more than 80% of the cassava cyanogenic glucosides and cyanide intoxication can occur when poorly processed cassava meal is eaten.

In some parts of the world, water has been shown as the cause of endemic goiter irrespective of its iodine content, particularly water from shallow wells. Such water is often grossly polluted and/or contains high levels of nitrate or humic substances (and some of their degradation products) which have also been shown to be goitrogenic by blocking thyroidal iodination.

Iodine Determination

Iodine analysis presents some of the most challenging of analytical situations. Iodine is volatile and can be easily oxidized or reduced. Thus, with iodine, care must be taken to prevent volatile losses and the analyst must be aware of the chemical form and oxidation state of iodine under various conditions. Conditions which are usually used to prepare samples and dissolve the ash during the analysis of other trace elements of nutritional concern may not be appropriate for iodine, so that iodine values are usually not available in many multielement food surveys. When iodine data are needed, a separate sample preparation phase specifically designed for iodine is usually employed. In addition, the content of iodine in diets and biological samples is often very low (normally well below $< 1.0\,\mu g\,g^{-1}$; usually in the $ng\,g^{-1}$ range), making accurate determination of iodine a difficult task.

Sampling and Sample Preparation

The complex nature of food samples compared to biological matrices necessitates extra care in preparing food composites representative of different geographic regions, seasons, food preparation (fresh, canned, frozen), manufacturers, and lots. Care must also be taken in thoroughly mixing, blending, and homogenizing a representative sample. It is important to transport and analyze samples as quickly as possible. Homogenized food and biological samples may be stored in a refrigerator for a short time (1 or 2 days prior to analysis). If samples are to be stored for longer periods, samples should be appropriately packaged and frozen at low temperatures (usually $-20\,^{\circ}C$ to $-80\,^{\circ}C$) to minimize iodine loss or sample degradation. As for all trace element analyses, appropriate storage containers should be used, although iodine contamination during preparation is less of a concern than for many other trace elements.

Analytical Quality Control and Reference Standards

Prior to implementation of any analytical method, method verification should form the first essential step in the analysis of food or biological samples. In its simplest form, method verification may consist of low-level standard additions to representative food samples along with the analysis of a certified standard reference material of similar composition and analyte levels as the sample of interest. Unfortunately, until recently, many of the reference standards did not have certified values for iodine, which severely limited the choice of reference standards with similar matrix properties and encompassing the range of iodine seen in diet and biological samples. In recent years, however, certified iodine values have been added for many more reference materials (Table 5).

For implementation of new or previously untried methods, more extensive preliminary testing is required, including a range of recovery testing, analysis of a variety of reference standards, and comparison of results with other established methods. Adoption of new analytical techniques usually involves more rigorous and involved analytical accuracy procedures such as interlaboratory verification.

Choice of standards Volatilization losses are minimized by maintaining iodine in an alkaline mixture or in $2\,mol\,l^{-1}$ H_2SO_4. Iodic acid standards are more stable than other forms of iodine such as sodium or potassium iodide or iodate. Iodine solutions are unstable in light and should be stored away from extended exposure to white light. If sodium or potassium iodide standards are used, stock solutions should be prepared, stored at room temperature, protected from white light, and discarded after 30 days. We have found that iodic acid stock solutions are be stable up to 5 years when protected from light. Diluted working standard solutions are normally prepared fresh each week.

Analytical Techniques for Determination of Iodine

There is a limited choice of methods for the determination of iodine in biological samples or food samples, with the exception of certain seaweeds or thyroid tissue which contain high concentrations of iodine. The most common methods are based on catalytic reactions, ion-selective electrodes, or neutron

Table 5 National Institute of Standards and Technology (NIST) reference standards suitable for analysis of iodine in food samples and agricultural commodities

NIST identification	*Sample*	*Certified value for I*[a] *(mg kg*$^{-1}$*)*
SRM 1566a	Oyster tissue	4.46 ± 0.42
SRM 1549	Nonfat milk powder	3.38 ± 0.02
RM 8435	Whole milk powder	2.3 ± 0.4
RM 8415	Whole egg powder	1.97 ± 0.46
SRM 1846	Infant formula (milk-based)	1.11 ± 0.17
SRM 1548a	Typical diet	0.759 ± 0.103
SRM 1571	Orchard leaves	Discontinued
RM 8418	Wheat gluten	0.060 ± 0.017
RM 8414	Bovine muscle powder	0.035 ± 0.012
RM 8433	Corn bran	0.026 ± 0.006
RM 8436	Durum wheat flour	0.006 ± 0.004
SRM 1577a	Bovine liver	Data for iodine not available
SRM 1567a[b]	Wheat flour	(0.0009)
SRM 1568a[b]	Rice flour	(0.0009)

[a]Certified value from the NIST certificate. Values in parentheses are not certified but are provided by NIST for information only.
[b]Reference material discontinued; replacement indicated by same SRM number followed by next letter designation, but replacement SRM is not yet certified for iodine.

activation analysis. A few others such as spot tests are useful in a limited number of situations.

The method used in the analysis of samples should be as sensitive as possible relative to the levels of analyte in the samples under investigation. Prior to the implementation of any iodine analysis, it is essential that the analytical method under consideration be shown to be adequate for the range of food or biological samples to be analyzed. The limit of detection is usually defined as the minimum concentration of analyte that can be qualitatively detected, but not quantitatively determined, under a predefined set of conditions. Limit of quantification, on the other hand, is the minimum concentration which can be determined quantitatively with acceptable accuracy and consistency. Quantities measured between the two are usually referred to as 'trace.' Unfortunately, in large surveys of food samples when multielement analyses are performed, limits of quantification for iodine are often not sufficiently sensitive to determine accurately iodine levels in a wide range of food samples. For example, in the 1995 US total diet study, the limit of quantification for iodine was 0.002 mg 100 g^{-1} (20 ng g^{-1}), so that a large number of samples had reported iodine values of zero. Food iodine values are not currently available on the US Department of Agriculture nutrient database.

Sample Preparation

The two methods commonly used to destroy organic matrix prior to iodine analysis are based on dry alkaline combustion or wet acid digestion. The former are usually done in open vessels, but can be done in a closed-vessel microwave digestion apparatus. In general, both methods are time-consuming, but necessary prior to most analytical techniques. There are a number of possible sources of iodine loss during the sample decomposition phase, including evolution of gaseous I_2, absorption or adsorption on to surfaces, or precipitation of undissolved material.

Dry ashing typically involves treating samples with 2.5 × volume 2 mol l^{-1} KOH in ethanol, drying, gradual heating in a combustion oven to 600 °C for 2 h, cooling, dissolution (often ultrasonic dissolution is necessary), followed by slurry filtration or centrifugation prior to analysis of the filtrate. Details can be found in the Association of Official Analytical Chemists method 935.14.

For wet acid digestion, samples are digested in borosilicate tubes in aluminum heating blocks or in Kjeldahl flasks on Kjeldahl digestion racks with a mixture of acids. Approximately 0.5–2 g of sample is added to the flasks and an equal volume of water if the sample is dry. Ten milliliters concentrated HNO_3 is added, followed by 20 ml $HClO_4$ and 5.0 ml H_2SO_4. Samples are covered and left overnight at room temperature. Digestion is performed slowly for the first 30 min with a cold finger inserted into the top of the flask to prevent volatile losses of iodine with the volatile nitrogen oxides. Thus the iodine is refluxed back into the flask and oxidized by the large excess of perchloric acid to nonvolatile iodates. If insufficient perchloric acid is used, elemental iodine will be lost. Once the evolution of nitrogen oxides subsides, the remaining nitric acid is distilled off and the temperature rises. Any lipid material will then be converted to CO_2 and the rest of the perchloric acid distilled off, resulting in a solution of pale straw color. At this point dense sulfuric acid vapor will begin to form and it is imperative that digestion ceases. Samples are cooled and diluted to 100 ml with water (final concentration 2 mol l^{-1} H_2SO_4). Iodic acid (HIO_3) standards (1 ml each of 10, 50, 100, 250, 500, and 1000 p.p.b) and blanks are digested and then analyzed with each batch of samples. Aliquots of the digestate solution are then analyzed for iodine using the iodine-catalyzed reduction of Ce^{4+} by As^{3+}. A microplate modification of this colorimetric method is described in the following section.

Determination of Iodine

Chemical or catalytic methods The spectrophotometric determination of iodine is based on the catalytic effect of iodide on the redox reaction between the yellow ceric (IV) and arsenic (III) which yields the colorless cerium (III) and arsenic (V), first reported by Sandell and Koltoff in 1934. This is a two-step reaction, where the rate of reaction is measured by the determination of the loss of the yellow ceric (Ce^{+4}) ions at 410 nm. Of the two steps, reaction (1) is generally regarded as slow compared to the rapid reaction (2). The net reaction (3) proceeds with extreme slowness in dilute sulfuric acid at room temperature, but is greatly increased in the presence of I^-.

$$2\,Ce^{+4} + 2\,I^- \rightarrow 2\,Ce^{+3} + I_2 \qquad (1)$$

$$As^{+3} + I_2 \rightarrow As^{+5} + 2\,I^- \qquad (2)$$

$$\text{Net} \quad 2\,Ce^{+4} + As^{+3} \rightarrow 2\,Ce^{+3} + As^{+5} \qquad (3)$$

A couple of points are important to remember when using this reaction. The catalytic coefficient of iodide is a function not only of iodide concentration, but also of the Ce^{+4}/As^{+3} concentration ratio, the chloride concentration, and the sulfuric acid concentration. The reaction is linear with respect to iodide when there is a large ratio of As III to Ce IV. Chloride can increase the velocity at low concentrations and also inhibit the catalysis by iodide at high sodium chloride concentrations, leading to large fluctuations in

background rates which interfere with iodine determination of food samples with widely differing salt contents. This interference is overcome by using a large excess of NaCl in the reaction mixture, typically 10%, although this results in a slight loss of sensitivity.

The iodine-catalyzed reduction of Ce^{4+} by As^{3+} colorimetric method has been semiautomated using a Technicon autoanalyzer, which is no longer available. More recent modifications have adapted the colorimetric assay to 96-well microplate assay systems. In the assay, 300 μl Ce^{4+} (0.015 mol l^{-1} $(NH_4)_4Ce(SO_4)_4 \times 2H_2O$ in 2 mol l^{-1} H_2SO_4), 300 μl As^{3+} (0.015 mol l^{-1} As_2O_3 in 2 mol l^{-1} H_2SO_4), 300 μl NaOH (10% NaOH) are added to each well. One hundred microliters of sample or standard are added to each microplate well at timed intervals. The plates are mixed, centrifuged at 110 **g** for 40 s (IEC Centra MP4R, Fisher Scientific, Nepean, ON) and incubated for 35 min at 37 °C. The color change of samples, which is directly proportional to the concentration of iodine in the sample, is measured at 410 nm using a microplate spectrophotometer and compared to the color change from the standard curve prepared from the ashed standard solutions. Standard curves used for food samples typically range from 10 to 1000 ng ml^{-1}.

Alternative colorimetric procedures based on the destruction of the intensely colored orange iron thiocyanate complex by nitrite in the presence of iodide have also been employed and have the same range of iodide sensitivity as the iodide catalyzed ceric/arsenic reaction.

The acid digestion followed by semiautomated colorimetric determination of iodine has several advantages for the analysis of large numbers of food samples. There are few losses of iodine; typical recoveries of iodine added to a variety of food composites are > 95%, and similar recoveries are found for National Institute of Standards and Technology (NIST) reference materials within the range of the standard curve. An analyst can typically digest 36 samples per day for 3 or 4 days per week followed by colorimetric determination of all sample digestates on a single day. The disadvantages are that one must take appropriate safety precautions when working with concentrated acids, a perchloric acid fume hood is required, and the digestion must be monitored continuously by the analyst.

Iodide electrodes Iodine in water, milk samples, and other liquids in which the only source of iodine is iodide (I^-) can be determined using an ion-specific electrode. In short, the protein in milk is removed by acid precipitation, samples are filtered, and iodine in the filtrate is determined by an iodide-specific electrode using the standard addition procedure. Nickel nitrate is added to reduce interferences to the iodide response. Ion-specific electrodes, however, are only suitable for liquid samples, although samples do not have to be digested prior to analysis. Because the gels get coated quickly when analyzing samples such as milk and infant formula, the electrodes must be carefully cleaned and conditioned often to ensure accurate results. Full details for the determination of iodide in ready-to-feed milk-based infant formula using an ion-selective electrode have been described in the Association of Official Analytical Chemists method 992.24.

Spot tests Several rapid spot test kits have been developed for the semiquantitative or qualitative estimation of iodide in iodized table salt. In this procedure the presence of iodine in salt is determined with a standard starch solution, provided in a ready-to-use formulation. In the standard test procedure, one drop of test solution is added to a spoon of salt. The change in color of the salt sample ranges from white to light blue to dark violet depending on the iodine content of the salt (usually classified as nil, 7, 15, or 30 p.p.m.) when compared to the standard color chart provided with the kit. These kits are useful in field situations because results are available immediately, particularly in underdeveloped countries where analytical lab facilities are not readily available. Errors are usually associated with the subjective nature of the color grading of the reaction.

Neutron activation analysis (NAA) When high sensitivity and freedom from reagent blanks are desired, neutron activation analysis (NAA) is often the method of choice. In general, NAA is the best confirmation technique for trace element analysis as it is independent of potential chemical interferences from the sample matrix, although its widespread use in food analysis is often limited due to high cost and limited sample throughout capacity. NAA measures radioactive iodine via gamma-ray spectroscopy after a period of irradiation. Various forms of NAA have been used, including: (1) radiochemical NAA; (2) epithermal NAA, which permits nondestructive analysis of samples; and (3) more recent modifications such as preirradiation combustion (PC-NAA), which permits extraction of the iodine from the sample matrix without using strong chemical reagents, followed by NAA.

For diet samples, conventional NAA of iodine is critically interfered with by the strong radioactivities of ^{24}Na, ^{38}Cl, and ^{42}K. In order to improve the detection limits of iodine, epithermal NAA using boron nitride as a shield material has been used.

In a further refinement using PC-NAA, iodine is first volatilized from the sample matrix by

combustion and adsorbed on charcoal after a series of combustion steps. The segment of quartz containing the charcoal is sealed under vacuum and NAA performed on the material. Detection limit is typically 50 ng and recovery using iodine-125 as a tracer is usually 80–100%. Stable iodine is determined following a 60-s irradiation to produce iodine-128. After irradiation, samples and standards are counted using a gamma-ray detector and sample concentration determined by comparing the peak heights of iodine-128 in the standards and samples. Due to the high equipment expense and low throughput (typically an analyst can only analyze two samples per day), this method is not useful for routine applications, although it may be useful for verification of results.

Inductively coupled plasma–mass spectrometry (ICP-MS) ICP-MS has also been used for iodine analysis. Samples are extracted with tetramethyl ammonium hydroxide (TMAH) at 90 °C prior to analysis by ICP-MS using metal compounds such as tellurium as an internal standard. Although recoveries are usually very good, ICP-MS, like NAA, is expensive and too time-consuming for large-scale food surveys.

See also: **Analysis of Food**; **Food Fortification**; **Glucosinolates**; **Goitrogens and Antithyroid Compounds**; **Hormones**: Thyroid Hormones; **Iodine**: Physiology; Iodine-deficiency Disorders; **Spectroscopy**: Visible Spectroscopy and Colorimetry; **Trace Elements**; **Water Supplies**: Water Treatment

Further Reading

Fischer PWF, L'Abbé MR and Giroux A (1986) Colorimetric determination of total iodine in foods by iodide-catalyzed reduction of Ce^{+4}. *Journal of the Association of Official Analytical Chemists* 69: 678–689.

Lauterbach A and Ober G (1995) Iodine and iodine compounds. In: Howe-Grant M (ed.) *Kirk-Othmer Encyclopedia of Chemical Technology*, 4th edn, vol. 14, pp. 709–736. New York: John Wiley.

MacKay GM, Points J and Pritchard E (eds) (1996) *Trace Analysis: A Structured Approach to Obtaining Reliable Results*. Cambridge: Laboratory of the Government Chemist, Royal Society of Chemistry.

Norman B and Iyengar V (1998) The application of pre-irradiation combustion and neutron activation analysis technique for the determination of iodine in food and environmental reference materials. *Biological Trace Element Research* 63: 221–229.

Pennington JAT, Schoen SA, Salmon GD *et al.* (1995) Composition of core foods of the US Food Supply, 1982–1991. III. Copper, manganese, selenium and iodine. *Journal of Food Composition and Analysis* 8: 171–217.

WHO (1996) *Trace elements in human nutrition and health, Chapter 4: Iodine*, pp. 49–71. Geneva: WHO.

Physiology

J A T Pennington, Division of Nutrition Research Coordination, National Institutes of Health, Bethesda, MD, USA

Absorption, Bioavailability, Transport, and Storage of Iodine

Iodine in foods is primarily in the inorganic iodide (I^-) form, which is readily absorbed form the stomach and upper small intestine. Other forms of iodine in foods are reduced to iodide before or after absorption. Absorbed iodine is rapidly cleared from the blood plasma by the thyroid gland, kidney, salivary glands, gastric mucosa, and lactating mammary glands. Iodine intake in excess of body need is excreted primarily through the urine.

The total amount of iodine in the body is about 10–50 mg for an adult, and most of this is concentrated in the thyroid gland. Thyroid-stimulating hormone (TSH), from the anterior lobe of the pituitary gland, stimulates the active transport of inorganic iodide from the blood into the thyroid gland. Within the thyroid cells, iodide is oxidized to iodine and combined with tyrosine to form the thyroid hormones, thyroxine (T_4) and triiodothyronine (T_3). The formation of thyroid hormones is also stimulated by TSH. Thyroid hormones are stored in the thyroid gland bound to the glycoprotein, thyroglobulin.

Proteolytic enzymes release the thyroid hormones from thyroglobulin so that the hormones may diffuse into the blood. In the blood, thyroxine is found primarily bound to globulin and albumin. The secretion of thyroid hormones into the blood is regulated by TSH through a biofeedback system. When dietary iodine is adequate, thyroid hormone is produced in normal amounts at a low level of TSH; if dietary iodine is limited, TSH secretion is increased, which promotes iodine uptake by the thyroid. The thyroid gland stores enough thyroid hormone to last several months in the event that dietary iodine is not available. (*See* **Hormones**: Thyroid Hormones.)

Functions of Iodine in the Body

Iodine is a structural constituent of the two thyroid hormones thyroxine (T_4) and triiodothyronine (T_3). The chemical names of T_4 and T_3 are, respectively, 3,5,3′,5′ tetraiodothyronine and 3,5,3′ triiodothyronine. T_4 and T_3 are iodinated molecules of the essential amino acid tyrosine. Iodine accounts for about 65% of the weight of T_4 and 49% of the weight of T_3. The thyroid secretes T_4 and small amounts of T_3. The

principal role of these hormones is to regulate cellular metabolism. The hormones are also required for normal growth and bone maturation in children. In infants, adequate supplies of thyroid hormone are necessary for the development of the central nervous system in the first 1–2 years of life. (*See* **Children**: Nutritional Requirements; **Infants**: Nutritional Requirements.)

Thyroid hormones accelerate cellular reactions in nearly all cells of the body, resulting in increased oxygen consumption and an increase in basal metabolic rate. The effects of the hormones are partially the result of increases in enzyme activity, as many enzymes are reported to be affected by T_4. Within body cells, thyroid hormones increase glycogen synthesis, gastrointestinal uptake of glucose and galactose, uptake of glucose by adipocytes, lipolysis, oxidation of free fatty acids, and protein synthesis. Thyroid hormones decrease serum cholesterol by increasing the removal of low-density lipoprotein from the blood.

Iodine Requirements

Approximately 60–100 μg of iodine per day is needed to produce an adequate supply of thyroid hormones. The Recommended Dietary Allowances (RDAs) for iodine in the USA are 40–50 μg per day for infants, 70–120 μg per day for children 1–10 years of age, and 150 μg per day for adolescents and adults. The USA RDAs are 175 μg per day during pregnancy and 200 μg per day during lactation. Similar recommendations for iodine intake are found in other countries. Reported average daily intakes of iodine in the USA, Australia, New Zealand, Canada, and European countries generally meet or exceed these recommendations.

Iodized salt is a convenient and effective means of meeting recommended intakes of iodine and preventing iodine deficiency in countries where the usual intake of iodine from food sources is low. In many countries, the level of iodine in salt ranges from 0.002 to 0.005%; in the USA, it is 0.01%, or approximately 416 μg of iodine per teaspoon of salt. Thus, one teaspoon of iodized salt provides 277% of the adult RDA.

Dietary Sources of Iodine

The inherent or natural iodine content of foods is variable, depending upon the iodine content of soil and water (for plant-based foods) and the iodine content of feed (for animal-based foods). The highest natural source of iodine is likely to be some species of seaweed.

Milk and products containing milk may be high in iodine due to the presence of indirect (i.e., unintentional) food additives. The iodine in milk is derived from ethylenediamine dihyroiodide (EDDI), a dietary iodine supplement added to dairy cattle feed and from iodophor solutions used as sanitizing agents on dairy cattle and dairy equipment. The use and extent of use of EDDI and iodophor solutions by the dairy industry cause great variation in the iodine content of milk. Seasonal and regional variations in the iodine content of milk may reflect different feeding and supplementation practices. During the summer or in warmer climates, dairy cattle may be pasture-fed. During the winter or in colder climates, dairy cattle may be given feed with EDDI supplementation. Feeding during spring and autumn or in moderate climates may reflect mixtures of pasture feeding and indoor feed supplementation.

Other foods high in iodine are those with iodine-containing food additives such as iodate dough conditioners, the red food dye erythrosine, and iodized salt. Iodate dough conditioners may be used in breads and other grain products. Erythrosine (tetraiodofluorescein) is 58% iodine (w/w) and may be present in some red-colored candies, pastries, and breakfast cereals. Although the concentrations of iodine in foods with erythrosine are quite high, the iodine from this source is only about 2–5% bioavailable. Regulations regarding the mandatory or voluntary iodization of salt and the level of iodization vary among countries. In the USA, for example, both iodized and noniodized salt are available in the marketplace, and the salt used by food manufacturers and by restaurants is usually noniodized. In other countries, the iodization of salt is mandated by government regulations.

Iodine is also available in dietary supplements as single supplements, in multiple mineral pills, and in multiple vitamin–mineral pills. The level of iodine in dietary supplements varies but is often 100% or more of the RDA. Iodine may also be present in other nutritional supplements and botanicals, especially if they include dried extracts of seaweed, plants, or fish.

Iodine Intakes

Iodine intake varies widely among population groups. Intakes may be less than 50 μg per day in iodine-deficient areas, over 400 μg per day in areas with iodized salt, and 1–5 mg per day or more in coastal areas of Japan where seaweed is commonly consumed. The main sources of dietary iodine for adults in the 1982–1991 US Total Diet Studies (TDS) were grain products (32% of iodine intake); milk and cheese (18%); animal flesh, especially fish

(13%); and desserts made with milk (10%). These results for the US TDS were based on a yearly analysis of 234 core foods for iodine content. Iodine intakes from the 1982–1991 US TDS ranged from 170 μg per day for infants to 386 μg per day for 25–30-year-old men. Average intakes for the more recent (1990–1991) TDS, were somewhat lower, ranging from 117 μg per day for infants to 317 μg per day for 14–16-year-old males, and there seems to have been a declining trend in iodine intake from 1982 to 1991. Nonetheless, these more recent iodine intakes are well above the RDAs without including iodine from discretionary salt (that added during cooking or at the table). Thus, the TDS estimates of iodine intake would be underestimated for individuals using iodized salt. The use of a half a teaspoon of iodized salt per day (iodized at the US level of 76 μg of iodine per gram of salt) would add an additional 208 μg of iodine.

Iodine Deficiency

Dietary iodine deficiency leads to a decrease in plasma T_4 and an increase in TSH. Under the influence of TSH, the thyroid becomes overactive, leading to hypertrophy and hyperplasia of the gland. This is known as 'endemic goiter.' Most individuals with goiter are euthyroid (i.e., have normal thyroid levels). Goiters range in size for modest enlargements to those that are several times the normal size and are frequently nodular in appearance. Large goiters may cause obstruction of the trachea and esophagus, and interfere with breathing and speaking. Prolonged deficiency of iodine, resulting in decreased levels of thyroid hormones, causes hypothyroidism or myxedema. The characteristics of this disease include decreased metabolic rate, fatigue, cold intolerance, weight gain, puffy face, dry skin, hoarse voice, edema, easy bruising, drooping upper eyelids, mental apathy, and problems with coordination. These symptoms are reversible with iodine therapy. Iodine deficiency during fetal development, infancy, or early childhood may result in cretinism, a condition that is not reversible and is characterized by retarded growth, mental retardation, deafness, and neurological problems.

In a few geographical areas, the development of endemic goiter as a result of iodine deficiency may be exacerbated by the ingestion of naturally occurring compounds in foods, known as 'goitrogens.' Goitrogens act by impairing iodine uptake by the thyroid or impairing the incorporation of iodine into tyrosine. Foods that contain goitrogenic substances include cassava, sweet potatoes, millet, and vegetables of the genus *Brassica* such as cabbage, broccoli, and turnips. The goitrogenic effect of these foods is not usually apparent if the diet is adequate in iodine. (*See* **Goitrogens and Antithyroid Compounds**.)

The FAO/WHO estimates that more than 20% of the world's population lives in geographic areas that are iodine-deficient. Iodine deficiency affects about 685 million people in Asia and the Pacific, 150 million in Africa, almost 55 million in the Americans, 33 million in North Africa and the Near East, and 82 million in Europe. Goiter from iodine deficiency affects about 200 million people world-wide. Prevalence rates for goiter of 5–20% are considered mild; in some communities, in Africa, Asia, and Latin America, more than 60% of the people have goiter. Children and women are more vulnerable to iodine deficiency than other population subgroups. About 26 million people have some degree of brain damage and mental retardation, and six million are affected by cretinism because of iodine deficiency.

Iodine Toxicity

Some individuals can tolerate very high levels of iodine with no apparent side-effects, and iodine intakes of up to 1 mg per day are probably safe for the majority of the population but may cause adverse effects in some individuals. Those who are most likely to respond adversely to iodine are those who have lived in areas of endemic goiter or who for other reasons have a habitually low intake of iodine, those with other thyroid disorders, and those who are sensitive to iodine.

Exposure to excessive iodine has occurred through foods, water supplies, drugs, dietary supplements, topical medication, and/or iodinated contrast media (used for X-rays). In the past, iodine-containing drugs (most commonly potassium iodide solutions) have been prescribed as expectorants for respiratory disease such as asthma, bronchitis, cystic fibrosis, and chronic pulmonary obstruction. Such drugs have also been prescribed for goiter, hyperthyroidism, rheumatism, and syphilis. Amiodarone, an iodine-containing drug, was prescribed for cardiac problems. The medical literature reveals that excessive exposure to iodine has resulted in thyroiditis, goiter, hypothyroidism, hyperthyroidism, sensitivity reactions, or acute responses for some individuals. Maternal iodine exposure during pregnancy has resulted in goiter, respiratory problems, enlarged hearts, and death of infants.

Iodine-induced hyperthyroidism may occur in some people receiving iodine supplementation. This condition is referred to as Jod–Basedow disease. These individuals may have underlying Graves' disease (hyperthyroidism) and react adversely when

given adequate supplies of iodine. Another explanation for the hyperthyroidism is that thyroid nodules of iodine deficiency become autonomous of normal control and produce excess thyroid hormone when iodine becomes available. High levels of iodine appear to interfere with the synthesis of thyroid hormones. This is referred to as the Wolff–Chaikoff effect. The block is usually temporary, but it persists in some subjects and leads to development of goiters.

Individuals who are sensitive to iodine may react to even small quantities of this element with fever, salivary gland enlargement, ioderma, visual problems, and/or other symptoms. Ioderma (dermatoses) resulting from iodine sensitivity vary from acne-type eruptions, rashes and urticaria, to eruptions described as bullous vesicular, purpuric hemorrhagic, pustular, or tuberous fungating. Deaths from these more severe forms of ioderma have been reported. Acute responses to the ingestion or injection of large doses of an iodine-containing solution have included cardiovascular collapse, convulsions, asthma attacks, and death.

Assessment of Iodine Status

Assessment of iodine status for an individual may be done by a physician checking for the presence of goiter, by measuring serum T_4 and TSH, by measuring urinary iodine, and by checking for other signs and symptoms of iodine deficiency or excess (e.g., level of fatigue, blood pressure, heart rate). Urinary iodine alone is not sufficient for determining the iodine status of an individual because it is a 'short-term' measure of iodine status reflecting the previous day's iodine intake, which may not be consistent with long-term measures of iodine intake. Individuals should also be asked about their usual diet, use of iodized salt, and use of dietary supplements that might contain iodine. Dietary assessment may be inaccurate due to the difficulty of obtaining precise information about iodine used as indirect or direct food additives (e.g., in milk, salt, bread) and in supplements.

Assessment of the iodine status of population groups requires methods that are both rapid and inexpensive. The methods used include the rate of palpable and/or visible goiter, urinary iodine excretion, and blood T_4 and TSH. In countries where assessment of iodine status is especially important for public iodine supplementation programs, iodine status may be measured by assessing the level of TSH in neonatal blood. Even though urinary iodine is a short-term measure of iodine status, it gives a useful indication of the distribution of iodine intakes for a population. Urinary iodine levels of $<20\,\mu g\,l^{-1}$ are considered to reflect severe deficiency, $20–49\,\mu g\,l^{-1}$ are moderate deficiency, and $50–99\,\mu g\,l^{-1}$ are mild deficiency. Urinary iodine levels between 100 and $200\,\mu g\,l^{-1}$ are considered satisfactory.

See also: **Children**: Nutritional Requirements; **Goitrogens and Antithyroid Compounds**; **Hormones**: Thyroid Hormones; **Infants**: Nutritional Requirements; **Iodine**: Iodine-deficiency Disorders

Further Reading

Bogden JD and Klevay LM (2000) *Clinical Nutrition of the Essential Trace Elements and Minerals*. Totowa, NJ: Humana Press.

Hetzel BS (1989) *The Story of Iodine Deficiency: An International Challenge in Nutrition*. Oxford: Oxford University Press.

Hetzel BS and Clugston GA (1999) Iodine. In: Shils ME, Olson JA, Shike M and Ross AC (eds) *Modern Nutrition in Health and Disease*, 9th edn. Baltimore, MD: Williams & Wilkins.

Pennington JAT (1988) Iodine. In: Smith K (ed.) *Trace Minerals in Foods*. New York: Marcel Dekker.

Pennington JAT (1990) A review of iodine toxicity reports. *Journal of the American Dietetic Association* 90: 1571–1581.

Iodine-deficiency Disorders

E N Pearce and S L Lee, Boston Medical Center, Boston, MA, USA

Introduction

Iodine is a chemical element found in trace amounts in the human body (15–20 mg). Iodine is stored almost entirely within the thyroid gland. The amount of iodine found in extrathyroidal tissues is minute and, therefore, it is considered a 'trace element' or a 'micronutrient.' The only function for iodine in human physiology is in the synthesis of thyroid hormones. Severe iodine deficiency will result in impairment of thyroid hormone synthesis that is necessary for normal childhood growth and adult metabolic function. Population effects of severe iodine deficiency include endemic goiter (enlarged thyroid gland), cretinism, decreased fertility rate, increased infant mortality, and mental retardation. The term iodine deficiency disorders (IDD) refers to the full clinical spectrum of disease caused by iodine deficiency.

Iodine Deficiency

Worldwide, there are large geographic areas with soil deficient in iodine. Human iodine intake is derived approximately 90% from food and 10% from water. The iodine content of foods and water generally reflects the iodine content of the local soil in which the food is grown. Current calculations by the International Council for Control of Iodine Deficiency Disorders (ICCIDD) have suggested that 29% of the world's population, or approximately 1.6 billion people, live in areas of iodine deficiency and are at risk for IDD. Those people at highest risk for IDD are located primarily in mountainous regions such as the Himalayas, the European Alps, the Andes, and the mountains of China. Iodine in the soil at these locations has been washed away by millions of years of glaciation and flooding. Iodine deficiency also occurs in lowland regions which are far from the oceans, such as central Africa and eastern Europe. Human iodine intake is obtained only through the diet. Crops grown in iodine-deficient areas will have decreased iodine concentrations. Animals in these areas will also be deficient, which may lead to decreased yields of meat, milk, and wool. In all these areas, those who consume only locally produced foods are at risk for IDD.

The dietary iodine requirement for normal thyroid physiology in adults is 100–150 $\mu g\,day^{-1}$. Two similar sets of recent recommendations from the US National Academy of Sciences, UNICEF (United Nations Childrens Fund), ICCIDD, and the World Health Organization (WHO) for different populations are shown in Tables 1 and 2.

In areas where iodine is not added to food products meant for humans or domesticated animals the primary sources of dietary iodine are saltwater fish and seaweed, because of the high iodine concentration in the oceans. Trace amounts are also found in grains and vegetables.

In countries such as the USA and the UK, iodine deficiency has been eliminated by means of 'silent prophylaxis.' Iodine has been voluntarily supplemented in table salt ($\sim 70\,g\,kg^{-1}$). After initial success in the 1920s to 1960s, areas that became iodine-sufficient after voluntary dietary supplementation have had a measurable decrease in iodine intake. Data collected for the first US National Health and Nutrition Examination Survey (NHANES I), conducted from 1971 to 1974, showed that the average urinary iodine concentration was 293.3 $\mu g\,g^{-1}$ creatinine, reflecting adequate dietary iodine intake. By the time of NHANES III in 1988–94, however, the average urinary iodine had fallen to 124.6 $\mu g\,g^{-1}$ creatinine. In particular, between the two surveys, the risk for insufficient dietary iodine intake of women of child-bearing age (15–44 years) increased 3.8 times.

It is not clear why iodine intake has fallen again in these areas but the etiology is likely multifactorial. This decrease in dietary iodine may be a result of reduced intake of eggs and salt, decreased iodate conditioners in bread, decreased use of iodized salt in manufactured foods, and poor education about the use of iodized salt. Currently in the USA only 50% of the salt sold in supermarkets is iodized. Noniodized salt contains little or no iodine. Another major source of dietary iodine in industrialized countries is egg yolks, because of iodine supplementation of chicken feed. Dairy products and meats have variable amounts of iodine, as some dairy cows and cattle are fed extremely high levels of iodine in supplements to maintain fertility and prevent hoof root. Levels of iodine in cows' milk reflect the animal's iodine oral intake. It has been often erroneously concluded that iodine in milk is from the iodophors used in the dairy industry to cleanse cows' teats. Small published studies suggest that, when properly used and rinsed off the cows' teats, very little iodine is transferred from iodophors into milk. Starting in the 1940s, iodate had been used as a bread stabilizer by

Table 1 WHO/UNICEF/ICCIDD recommended daily iodine intake[a]

Population	*Recommended daily intake (μ/day)*
Children	
0–59 mo	90
6–12 yr	120
Adults > 13 yr	150
Pregnant and lactating women	200

[a]WHO/UNICEF/ICCIDD (2001).

Table 2 US National Academy of Sciences recommended daily iodine intake[a]

Population	*Recommended daily allowance (RDA) (μ/day)*
Infants	
0–6 mo	110 AI[b]
7–12 mo	130 AI[b]
1–8 yr	90
9–13 yr	120
Children 6–12 yr	120
Adults > 13 yr	150
Pregnant women	220
Lactating women	290

[a]US National Academy Sciences Food and Nutrition Board and Institute of Medicine (2002).
[b]RDA not established for infants. These are values are adequate intake (AI) for infants based on the minimal iodine intake of infants exclusively fed human milk.

the commercial baking industry; this practice was abandoned in the USA in the 1960s but has reappeared in some breads in the US. One particular type of red dye, erythrocine (US Food, Drug and Cosmetic (FD&C) no.3), contains large amounts of iodine. This dye was commonly used until the 1960s in candy and breakfast cereals in the USA and is still found in some medications. In most other countries, iodination of salt, bread, or drinking water is mandated by law, and is performed under government supervision.

Prevalence

More than 13% of the world's population has some form of IDD. In 1990, at least 1.572 billion people worldwide were at risk for IDD. These included 710 million people in Asia, 60 million in Latin America, 227 million in Africa, and 20–30 million in Europe. Of these, 655 million were known to have goiter, and 11.2 million had cretinism. IDD is known to affect 50 million children: 100 000 cretins are born each year. The ICCIDD maintains a database of iodine nutrition in different countries (CIDDS database) available at the website: http://www.med.virginia.edu/~jtd/iccidd.

Pathophysiology of IDD

Dietary iodine is readily taken up through the gut in the form of iodide. From the circulation, it is concentrated in the thyroid gland by means of an energy-dependent sodium-iodate symporter. In the follicle cells of the thyroid gland, four atoms of iodine are incorporated into each molecule of L-thyroxine (L-T_4) and three atoms into each molecule of triiodothyronine (T_3). These hormones are essential for neuronal development, sexual development, and growth, as well as for regulating metabolic rate, body heat, and energy.

When dietary iodine intake is inadequate, the serum T_4 level initially falls, and a number of physiological processes occur to restore adequate thyroid hormone production. The pituitary gland detects low T_4 levels and releases more thyroid-stimulating hormone (TSH). TSH stimulates the growth, iodine uptake, and metabolic activity of thyroid follicular cells. TSH stimulates the thyroid to increase iodine uptake, thyroid hormone synthesis, and thyroid hormone secretion into the circulation. Increased TSH and the reduction of iodine stores within the thyroid result in an increased ratio of T_3 to T_4 production. T_3 is 20–100 times biologically more potent than T_4, and requires fewer atoms of iodine, so this tends to conserve iodine stores and maintain normal thyroid function. In addition, thyroid hormones are deiodinated in the liver and the iodine is released back into the circulation for reuptake and reuse by the thyroid gland. Under these circumstances of an efficient reutilization process, approximately 100–150 μg of iodine daily is passively lost in the urine, with additional small losses from biliary secretion into the gut, and must be replaced by dietary intake.

Enlargement of the thyroid gland is a normal adaptive reaction to low iodine intake. Iodine deficiency is the most frequent cause of endemic goiter, which is defined as a goiter prevalence of > 10% in a population. Thyroid gland volume measured by the sensitive and accurate method of ultrasonography shows that at any age the glands of an iodine-sufficient population are smaller than an iodine-deficient population. Iodine-deficiency goiter is initially diffuse with homogeneous hyperplasia. With time, the hyperplasia becomes nonuniform and interspersed by areas of regression or degeneration and fibrosis, resulting in nodular changes. Some nodules may become autonomous and secrete excess thyroid hormone regardless of the TSH level. Autonomous thyroid nodules in areas of iodine deficiency frequently contain TSH receptor-activating mutations that keep the TSH receptor in the activated or 'on' state, regardless of TSH ligand binding. With time (years to decades), growth and accumulation of the autonomous nodules may result in mild thyrotoxicosis. When a goiter is associated with hyperthyroidism, it is called a toxic multinodular goiter. In addition, autonomous nodules in goiters of euthyroid patients subsequently exposed to normal or high iodine levels may also lead to thyrotoxicosis. Depending on the exposure time and severity of the iodine deficiency, goiters may enlarge sufficiently to cause tracheal and esophageal narrowing and obstruction. Generally, treatment with iodine supplementation or T_4 therapy will not decrease the size of large nodular goiters in adults but will shrink small diffuse goiters of young children with iodine deficiency.

If iodine deficiency is extremely severe, thyroid hormone production will fall, and patients will become hypothyroid. Adults will have the usual signs and symptoms of hypothyroidism with goiter. Hypothyroidism in the fetus, neonates, and in young children prevents central nervous system maturation, especially neuronal myelination, leading to permanent mental retardation. In the most severe form, the constellation of mental retardation and growth abnormalities is called cretinism.

The severity of iodine deficiency in a population can be estimated by several measurements that can be obtained in the field, as shown in Table 3.

Table 3 Measures of iodine deficiency disorder severity in populations

	None	*Mild*	*Moderate*	*Severe*
Urinary iodine, median ($\mu g\,dl^{-1}$)	> 100	50–99	20–49	< 20
Goiter prevalence	< 5%	5–20%	20–30%	> 30%
Neonatal thyroid-stimulating hormone > 5 mIU ml^{-1} whole blood	< 3%	3–20%	20–40%	> 40%
Cretinism	0	0	+	+++

Adapted from WHO-UNICEF-ICCIDD (1993).

Dietary iodine intake is not the sole determinant of goiter development. Even in areas of severe iodine deficiency, there can be substantial variations in the incidence of goiter. One reason for this is the presence of goitrogens in the diet. The clinical disorders of iodine deficiency are magnified in the presence of dietary goitrogens. Goitrogens are substances that decrease the bioavailability of iodine or interfere within the thyroid hormone biosynthethic pathway. The presence of these substances was first established through animal experiments in the 1920s. Thiocyanates inhibit the thyroid gland's iodine-concentrating mechanism. Vegetables of the genus *Brassica*, which include cabbage, Brussels sprouts, broccoli, cauliflower, turnips, rape and rapeseed, and horseradish, all contain thioglucosides which are converted to thiocyanate after ingestion. Foods that contain cyanogenic glucosides release cyanide when digested; this is detoxified in the body to become thiocyanate. These goitrogenic foods include many major dietary staples of the developing world, such as cassava, maize, sweet potatoes, pearl millet, and lima beans. Finally, flavonoids (polyphenols), organic compounds found in many food plants, including millet species, also have antithyroid activity that inhibit thyroid hormone synthesis. Additionally, the clinical disorders of iodine deficiency tend to be more profound in geographic areas with coexisting selenium and vitamin A deficiencies, or protein-calorie malnutrition. Generally, the foods listed here by themselves do not contain sufficient goitrogens to be clinically important for thyroid physiology until the intake constitutes a major portion of the diet or is combined with iodine or other dietary deficiencies.

Clinical Manifestations

Affected patients come from geographic regions where IDD is endemic. The first sign of iodine deficiency is diffuse thyroid enlargement, which becomes nodular or multinodular with time. If a goiter is large enough, individuals may complain of compressive symptoms such as hoarseness, shortness of breath, or dysphagia. Goiters can be assessed by palpation and can be graded from 0 to 2, as shown in Table 4.

Table 4 Clinical grades of thyroid enlargement

Grade 0	No visible or palpable goiter
Grade 1	Enlarged thyroid that is palpable but not visible when the neck is in the neutral position
Grade 2	Enlarged thyroid swelling that is palpable and visible when the neck is in a neutral position

Reproduced from Delange FM and Utiger RD (eds) (2000) In: *Werner & Inbgar's The Thyroid*, 8th edn. Philadelphia: Williams & Wilkins, with permission.

Thyroid size, determined by ultrasound, has recently been shown to reflect the iodine sufficiency of a population. Unlike size estimates determined by palpation alone, ultrasound measurements are reproducibly accurate to within 10%. When goiter appears in >5% of a regional population, iodine deficiency should be considered (Table 3).

Iodine requirements increase during pregnancy, in part because of fetal requirements but also because of the mother's increase in total body volume. This increased iodine requirement will increase maternal goiter size in areas of mild, moderate, or severe iodine deficiency. Even in areas of borderline iodine intake, up to 10% of women may develop goiter during pregnancy. This pregnancy-related increase in thyroid volume has been carefully documented by ultrasound scans and may not be fully reversible after pregnancy, so goiter may increase in size in the mother after each pregnancy.

Patients with hypothyroidism due to severe iodine deficiency may experience fatigue, weight gain, cold intolerance, dry skin, constipation, or depression. Signs on physical exam may include dry skin, periorbital edema, and delayed relaxation phase of the deep tendon reflexes. Neonates and young infants are much more severely affected by iodine deficiency than are adults, and are more likely to become overtly hypothyroid.

Children are profoundly affected by iodine deficiency. In addition to the risk of goiter formation, neurological development and linear growth are critically dependent on prenatal and postnatal iodine intake and thyroid hormone sufficiency. Significant neuronal maturation and myelination

Table 5 Neurologic and myxedematous cretinism

Neurologic cretinism	Mental retardation, abnormal gait, and deaf-mutism, but not goiter or hypothroidism
Myxedematous cretinism	Mental retardation, short stature, goiter, and hypothyroidism

occur postnatally. Cretinism is the most extreme manifestation of IDD. It is rare, but is still seen in regions of severe iodine deficiency in southern and eastern Europe, Asia, Africa, and Latin America. Despite efforts by many international health organizations, including the WHO, 100 000 cretins are born each year. Cretinism can be divided into neurologic and myxedematous subtypes, although these subtypes have considerable clinical overlap, as shown in Table 5. Neurologic cretinism is more common and is thought to be caused by maternal hypothyroidism secondary to iodine deficiency during pregnancy (especially between gestational weeks 12 and 30). Myxedematous cretinism is found primarily in Nepal, western China, and Zaire. It is considered a result of iodine deficiency and hypothyroidism in the fetus during late pregnancy or in the neonatal period. Both conditions can be prevented by adequate maternal and childhood iodine intake.

Although overt cretinism is rare, populations in which severe iodine deficiency is prevalent are at risk for lower IQ and mental retardation, sometimes known as IDD developmental retardation. In fact, iodine deficiency is the leading cause of preventable mental retardation worldwide. This decrease in mental acuity in afflicted communities, combined with decreased agricultural yields of meat, milk, and wool from iodine-deficient animals, can have a devastating impact on local economies.

Severely iodine-deficient women are more likely to experience infertility, and pregnancy in this group is more likely to result in miscarriage or congenital anomalies. Children born to these women are more likely to have low birth weights, and to suffer infant mortality. Recent studies in the USA suggest that infants born to mothers with untreated hypothyroidism have a small reduction in intelligence as measured by standardized tests.

Finally, it is unclear whether iodine deficiency increases the incidence of thyroid cancer. There are clearer data that iodine intake affects the histiological subtypes of thyroid cancers. In areas where iodine deficiency is endemic there is a higher proportion of more aggressive thyroid cancers (follicular and anaplastic thyroid carcinoma) and an increased thyroid cancer mortality rate. Iodine prophylaxis has been demonstrated to reduce the proportion of follicular thyroid carcinomas and increase the proportion of the less aggressive form of thyroid cancer, papillary thyroid carcinoma.

Laboratory Data

The kidneys excrete approximately 90% of ingested iodine. The best diagnostic test for IDD, therefore, is a 24-h urine iodine collection. Based on this, IDD can be classified as mild (3.5–5 $\mu g\,l\,dl^{-1}$), moderate (2.0–3.5 $\mu g\,l\,dl^{-1}$), or severe (< 2.0 $\mu g\,l\,dl^{-1}$). If a 24-h urine collection is not practical, a random urine iodine-to-creatinine ratio can be substituted. In this case, 50–100 $\mu g\,l\,g^{-1}$ creatinine is consistent with mild iodine deficiency, 20–49 $\mu g\,l\,g^{-1}$ creatinine with moderate deficiency, and < 20 $\mu g\,l\,g^{-1}$ creatinine with severe deficiency.

Population studies have shown that neonates with IDD will have elevated TSH levels at birth but normal levels when evaluated weeks later. The extent of their transient hypothyroidism correlates with the severity of the iodine deficiency (Table 2). Blood from the umbilical cord or from a heelstick can be saved on filter paper and sent to a laboratory for screening within the first few days of life. Neonatal TSH screening is used widely in Europe, Japan, and North America. In areas of adequate iodine intake, about one in 4000 infants will be hypothyroid. In iodine-deficient areas, by contrast, hypothyroidism may be present in up to 10% of newborns. Early treatment of these hypothyroid infants with exogenous thyroid hormone can reduce the risk for developmental delays and mental retardation. Unfortunately, neonatal TSH screening is difficult and expensive and is not done in much of the developing nonindustrialized world.

Thyroid function studies are usually within the normal range in the presence of mild iodine insufficiency. In euthyroid patients with iodine deficiency, however, serum TSH levels may be normal to increased, T_3 levels normal or slightly elevated, and T_4 levels normal or decreased. Only in very extreme iodine deficiency does hypothyroidism develop; in this setting serum TSH will be elevated and both T_3 and T_4 decreased.

Twenty-four-hour radioactive iodine uptake is used in developed countries to help distinguish between different types of thyroid disease. Uptakes will be substantially increased in the presence of IDD due to increased TSH stimulation and reduction in the nonisotopic iodine pool. Therefore, thyroid uptake values in iodine-sufficient areas such as the USA are significantly lower than in areas with iodine deficiency such as western Europe.

Thyroid size as determined by ultrasound reflects the iodine sufficiency of a population. When goiter appears in > 5% of a regional population, iodine deficiency

should be considered (Table 3). Ultrasound scans of a goiter can be used to document size, volume, and normal size of individual nodules within the gland.

Medical Treatment

Prophylactic iodine replacement therapy will prevent IDD, including goiter, hypothyroidism, and cretinism, and is indicated for the entire population within an area of iodine deficiency. Once IDD has been established, long-term dietary iodine replacement may decrease the size of small diffuse goiters of short duration in infants, young children, and pregnant women. The supplementation of iodine will not reverse cretinism or reduce the size of large nodular endemic goiters.

Otherwise, patients do not routinely require specific therapy unless goiter is large enough to cause compressive or obstructive symptoms (for example, tracheal obstruction, thoracic inlet occlusion, or hoarseness). Radioactive iodine (I-131) has been used, primarily in Europe, to decrease thyroid volume of euthyroid goiters by up to 40–60%. Exogenous (L-T_4) can also be used to decrease goiter size, but is generally not effective in adults and older children. Supplemental L-T_4, when added to T_3 and T_4 secretion by the autonomous nodules in the endemic goiter, may cause thyrotoxicosis. Long-term L-T_4 therapy that results in the suppression of TSH to below-normal levels may have deleterious effects on cardiac and bone health and, therefore, is no longer routinely given to patients with goiter.

Surgical treatment by thyroidectomy may be indicated for compressive symptoms of a large goiter.

Prevention

At a population level, IDD can be prevented by the iodination of food products or the water supply. Optimal supplements should provide physiologic iodine levels, must be able to reach all of the affected population, and must be affordable. Different approaches have been required in different areas of the world.

Iodized salt was first instituted in the USA and Switzerland in the 1920s. In practice, the iodination of salt has proved to be the best method of iodine supplementation. Everyone needs salt, it is consumed at the same rate throughout the year, and there is little variation in patterns of salt use with differences in socioeconomic status. Salt can be used in animal feeds as well as in table salt for human consumption. Additionally, in most areas salt comes from a limited number of production centers. The technology involved in salt iodination is relatively straightforward. The iodine does not affect the appearance or the taste of the salt. Either iodide or iodate can be used; the iodate is more stable in tropical climates. Salt iodination programs cost US$0.05 per person annually. Iodine dose recommendations for each region are based on the average salt intake of the local population, local iodine requirements, and average losses of iodine during salt transportation and storage. Despite the relative simplicity of this approach, salt iodination can take years to reach the most severely affected areas of an iodine-deficient region.

An alternative to salt iodination in some developing countries has been the periodic use of iodized oil supplements. This is frequently used as an interim measure while iodized salt programs are being implemented, and is especially used for women of child-bearing age and for children. The iodized oil can be given as an intramuscular injection of 0.5 ml (240 000 μg iodine) during the first year of life or 1.0 ml (480 000 μg iodine) thereafter. A single dose is effective for 3–5 years. However, the injections require individual contact with a skilled caregiver. There is also a risk of disease transmission if sterile syringes and needles are not readily available. Therefore, oral iodized oil administration has been gaining favor instead. One oral administration can be effective for up to a year, although it does not provide the constant levels attained by injection. The usual oral dose is 0.1–0.25 ml (about 380 000 μg iodine).

Iodized bread has been used with some success in several countries, including the Netherlands, Tasmania, Russia, and Australia. Iodination of other substances such as tea and sugar has also been tried, though these substances tend to be consumed in variable quantities. An alternative approach is the iodination of drinking water. Relatively constant daily amounts of iodine can be added to either household or school water supplies. However, this is generally more expensive than large-scale iodized salt programs and requires a significant amount of monitoring and oversight. In some remote areas of China, where other approaches were not feasible, the iodination of irrigation and well water was highly effective over several years. This intervention was shown to decrease infant mortality by about 50%. This approach was extremely cost-effective, improving not only human health but livestock production.

Education of government officials and of affected populations is a critical component of any iodine supplementation program. It is important for the public to understand the importance of using iodized salt, especially in countries such as the USA and the UK, where iodination of salt is not mandated by law (only 50% of salt sold in US supermarkets and 2.5% of salt sold in the UK is iodized).

Complications of Iodine Treatment

The primary complication of iodine prophylaxis for IDD is the development of hyperthyroidism. This is known as the Jod–Basedow phenomenon, or iodine-induced hyperthyroidism (IIH). The presentation of hyperthyroidism may be subtle. Typical symptoms include anxiety, heat intolerance, palpitations, weight loss, increased appetite, tremor, and menstrual disturbances. IIH is caused by the hyperfunctioning areas of autonomy that tend to develop in long-standing iodine-deficient goiters. It tends to be observed in patients over age 45 during the early phase of IDD prevention programs. It has been seen in almost all iodine supplementation programs worldwide. Well-documented IIH epidemics occurred in the USA from 1924 to 1928, in Tasmania in the 1960s, and, more recently, in Zaire and Zimbabwe. The outbreaks in Zaire and Zimbabwe were thought to be due to iodine overload secondary to insufficient monitoring of salt iodination programs. IIH occurs, however, even with physiologic iodine supplementation levels (100–200 $\mu g\,day^{-1}$). IIH is usually benign and easily treatable, but in rare cases severe cardiovascular complications can lead to fatalities. For this reason, iodized oil administration is not recommended for adults over age 45. In general, doses of up to 1100 μg of iodine per day are considered safe for adults.

Eliminating IDD: Worldwide Progress

At the United Nations World Summit for Children in 1990, a plan was adopted for the elimination of IDD by the year 2000. Although substantial progress has been made toward this goal, it has not been achieved. According to WHO data, the proportion of the world population at risk for IDD dropped from 28.9% in 1994 to 13.7% in 1997. The proportion of the population in developing countries regularly consuming iodized salt increased from only 5% in 1990 to 60% in 1994. Currently 68% of the 5 billion people residing in countries with iodine deficiency have access to iodized salt. Rates of goiter, cretinism, and mental retardation are decreasing worldwide. However, programs must be continually monitored, and must be sustainable.

See also: **Hormones**: Thyroid Hormones; **Iodine**: Properties and Determination; Physiology

Further Reading

Delange FM (2000) Iodine deficiency. In: Braverman LE and Utiger RD (eds) *Werner & Ingbar's The Thyroid*, 8th edn, pp. 295–316. Philadelphia: Lippincott Williams & Wilkins.

Delange F and Hetzel BS (2002) The iodine deficiency disorders. Available at: http://www.thyroidmanager.org/ Chapter 20. Endocrine Education Inc., Boston, MA.

Dunn JT (ed.) *IDD Newsletter.* International Council for Control of Iodine Deficiency Diseases.

Dunn JT and Van der Har F (1990) *A Practical Guide to the Correction of Iodine Deficiency.* The Netherlands: ICCIDD.

Hetzel BS (1989) *The Story of Iodine Deficiency: An Intellectual Challenge in Nutrition*. New York: Oxford University Press.

Hollowell JG, Staehling NW, Hannon WH *et al.* (1998) Iodine nutrition in the United States: trends and public health implications: iodine excretion data from the National Health and Nutrition Surveys I and III (1971–1974 and 1988–1994). *Journal of Clinical Endocrinology and Metabolism* 83: 3401–3408.

International Council for the Control of Iodine Deficiency Disorders (ICCIDD) Database of iodine nutrition in different countries (CIDDS database). Available online at http://www.med.virginia.edu/~jtd/iccidd.

Lee K, Bradley R, Dwyer J and Lee SL (1999) Too much versus too little: the implications of current iodine intake in the United States. *Nutrition Reviews* 57(6): 177–181.

United States National Academy of Sciences (2002) *Dietary reference intakes for vitamin A, vitamin K, arsenic, boron, chromium, copper, iodine, iron, manganese, molybdenum, nickel, silicon, vanadium, and zinc.* Chapter 8, pp. 258–289.Washington, DC: A joint Food and Nutrition Board and Institute of Medicine publication.

WHO Nutrtion Unit (1994) *Indicators for assessing iodine deficiency disorders and their control through salt iodization*. Geneva: WHO.

WHO/UNICEF/ICCIDD (1993) *Indicators for assessing iodine deficiency disorders and their control programmes*. Report of a joint WHO/UNICEF/ICCIDD consultation. Geneva: WHO.

WHO/UNICEF/ICCIDD (2001) *Assessment of the iodine deficiency disorders and monitoring their elimination*. Geneva: WHO.

IRON

Contents

Properties and Determination

J R Chipperfield, University of Hull, Hull, UK

Physical and Chemical Properties

The element iron (Latin *ferrum*), atomic number 26, relative atomic mass 55.85, has an electronic configuration [Ar] $3d^6\,4s^2$. Iron normally occurs in one of three oxidation states iron(0) (metallic iron), iron(II) and iron(III). Iron(II) and iron(III) compounds were formerly called 'ferrous' and 'ferric' respectively. Iron oxides are high-melting-point solids. Iron(II) oxide (FeO), is black, whereas iron(III) oxide (Fe_2O_3) is orange-red. Both iron(II) and iron(III) form many ionic salts, such as halides ($FeCl_2$ and $FeCl_3$), sulfates ($FeSO_4$ and $Fe_2(SO_4)_3$), and nitrates ($Fe(NO_3)_2$ and $Fe(NO_3)_3$). Iron(II) sulfate, iron(II) gluconate, and iron(II) fumarate are used medicinally for oral iron supplementation. In acidic solution these salts dissolve to give the aquated cations $[Fe(H_2O)_6]^{2+}$, and $[Fe(H_2O)_6]^{3+}$ (Figure 1). These octahedral aquated ions are also found in solid hydrated salts such as hydrated iron(II) sulfate ($FeSO_4.7H_2O$), and hydrated iron(III) sulfate, $(Fe_2(SO_4)_3.9H_2O)$. As pH increases, insoluble hydroxides $Fe(OH)_2$ and $Fe(OH)_3$ precipitate, the latter being so insoluble that, at pH 7, the equilibrium concentration of iron(III) in solution is only 10^{-9} mol dm^{-3}. Solutions of iron(III) salts are strongly acidic due to hydrolysis of the $[Fe(H_2O)_6]^{3+}$ cation (eqn 1).

$$[Fe(H_2O)_6]^{3+} + H_2O \rightleftharpoons [Fe(H_2O)_5(OH)]^{2+} + H_3O^+ \qquad (1)$$

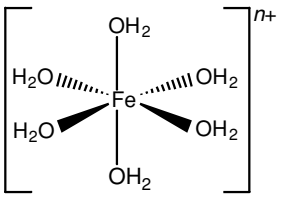

n = 2[iron(II)] or 3[iron(III)]

Figure 1

The pK_a of $[Fe(H_2O)_6]^{3+}$ is around 2, similar to that of phosphoric acid, and therefore solutions of iron(III) salts are as acidic as a solution of phosphoric acid of the same molarity.

Both iron(II) and iron(III) form complexes with many ligands. Simple monodentate ligands such as Cl^- and CN^- give complex ions such as tetrahedral $[FeCl_4]^{2-}$ and $[FeCl_4]^-$ (Figure 2) and octahedral $[Fe(CN)_6]^{4-}$ and $[Fe(CN)_6]^{3-}$ (Figure 3). Chelating ligands such as phen (1,10-phenanthroline), (Figure 4), and dipy (2,2′-dipyridyl), (Figure 5), give the chelated complex ions $[Fe(phen)_3]^{2+}$, $[Fe(phen)_3]^{3+}$, $[Fe(dipy)_3]^{2+}$, and $[Fe(dipy)_3]^{3+}$ (Figure 6).

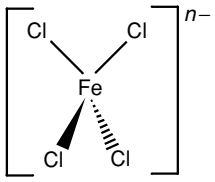

n = 2[iron(II)] or 1[iron(III)]

Figure 2

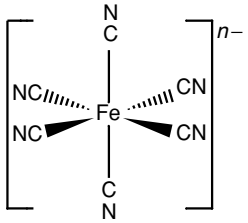

n = 4 [iron(II)] or 3[(iron(III)]

Figure 3

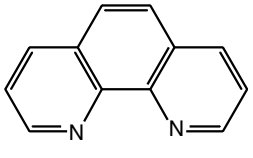

1,10-phenanthroline (phen)

Figure 4

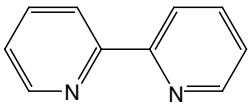

2,2'-dipyridyl (dipy)

Figure 5

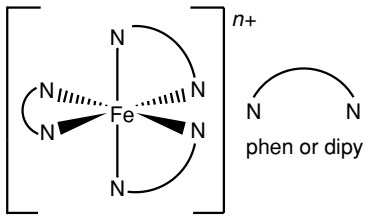

n = 2[iron(II) or 3[iron(III)]

Figure 6

In the absence of ligands, aerial oxygen will readily oxidize solutions of iron(II) compounds to iron(III). When ligands are present they most commonly (e.g., CN^-, F^-) stabilize iron(III) more than iron(II), or more rarely (e.g., phen) stabilize iron(II) relative to iron(III) such that aerial oxidation of iron(II) to iron(III) is prevented. Iron(II) has the electron configuration [Ar] $3d^6$, and in octahedral complexes the six 3d electrons may be found in two configurations – high-spin, with four unpaired electrons, or low-spin, with no unpaired electrons. The presence of unpaired electrons imparts paramagnetism to high-spin iron(II) compounds. Iron(III) is similarly found in a high-spin (five unpaired electrons) or low-spin (one unpaired electron) configuration. The nature of the ligands determines whether complexes are high- or low-spin.

In Pearson's terminology of 'hard' and 'soft' acids and bases, iron(III) is a 'hard' acid and forms its most stable complexes and compounds with 'hard' ligands (i.e., those containing oxygen or nitrogen donor atoms). Iron(II) is 'softer' than iron(III) and, as well as bonding to both oxygen and nitrogen donor atoms, is also found bonded to 'soft' ligands such as sulfur donor ligands.

In life processes, iron(II) complexes are needed which are not irreversibly oxidized to iron(III) by aerial oxidation.

Occurrence and Speciation in Foods and in the Environment

Foods

Iron is found in foods derived from both animals and plants. Table 1 shows typical concentrations of iron found in a variety of common foods. The iron may be present either as heme-based compounds or as non-heme compounds.

Heme compounds In mammals the two main iron-containing compounds are the heme proteins hemoglobin and myoglobin, which are used to transport oxygen. Heme proteins consist of an iron–porphyrin unit (Figure 7) and a protein unit. The porphyrin unit coordinates to the iron in a plane. In myoglobin and hemoglobin the fifth coordination position of the iron is occupied by a nitrogen atom (from a histidine residue) which is linked to the protein chain, and the sixth position is occupied by a water or oxygen molecule. This is shown diagrammatically in Figure 8. Other heme proteins are the cytochromes in which, as part of redox reactions, the iron atoms shuttle between iron(II) and iron(III). They have a similar porphyrin–iron core to that shown above, but both the fifth and sixth coordination positions of the iron are linked to the protein chain (Figure 9).

Although most familiar in animals, photosynthesis in plants involves similar heme proteins.

Nonheme proteins Nonheme proteins in animals and plants are used to store and transport iron. The most abundant in mammals is ferritin. This consists of an iron-containing core surrounded by a protein sheath (Figure 10). The core may contain up to 4500 iron atoms (as iron(III)), and has a structure somewhat similar to the hydrated iron(III) oxide ferrihydrite $5Fe_2O_3.9H_2O$. The surrounding protein sheath consists of 24 similar subunits. Another iron transport protein, transferrin, is a glycoprotein containing a single polypeptide chain folded into two lobes. Each protein can strongly bond two iron(III) ions and each iron(III) ion has a carbonate or bicarbonate anion associated with it. The iron(III) atoms are octahedrally coordinated.

A variety of microorganisms produce and release siderophores. These are powerful chelating agents which bond to iron atoms through oxygen atoms, and yield neutral molecules which can pass through cell walls. One particular example is ferrichrome (Figure 11) which is a cyclic hexapeptide containing three hydroxamate groups.

Iron–sulfur proteins These are widely distributed and encompass rubredoxins, ferredoxins, and high-potential iron proteins (HIPIPs). Rubredoxins (Figure 12) contain a single iron atom bonded through four tetrahedrally arranged sulfur atoms to the rest of the protein. Ferredoxins can have two iron atoms (Figure 13), or four iron atoms (Figure 14). In each case there are sulfur atoms bridging between the

Table 1 Iron content of common foods

Cereals	*Iron content (mg Fe per 100 g)*	*Milk products*	*Iron content (mg Fe per 100 g)*	*Meat, fish*	*Iron content (mg Fe per 100 g)*	*Vegetables*	*Iron content (mg Fe per 100 g)*	*Miscellaneous*	*Iron content (mg Fe per 100 g)*
White plain flour	1.5	Whole milk	0.06	Bacon (raw, lean)	1.2	Potatoes (new)	0.3	Raisins	3.8
Whole meal flour	3.9	Cheddar cheese	0.3	Ham (canned)	1.2	Lentils (dried)	11.1	Currants	1.3
Savory rice (cooked)	0.5	Yogurt	0.1	Beef (raw, lean)	2.1	Peas (raw)	2.8	Mixed nuts	2.1
White spaghetti (boiled)	0.5	Dairy icecream	0.1	Lamb (raw, lean)	1.6	Cabbage (raw)	0.7	White sugar	trace
Brown bread	2.2	*Eggs and fats*		Veal (raw)	1.2	Swede (raw)	0.1	Honey	0.4
White bread	1.6	Chicken egg	1.9	Chicken (raw)	0.7	Tomato	0.5	Instant coffee	4.4
Cornflakes	6.7	Butter	0.2	Liver (pig, raw)	21.0	Mushroom	0.6	Coffee (infusion)	trace
Christmas pudding	1.5	Margarine	0.3	Cod (raw)	0.3	Apples	0.1	Tea (infusion)	trace
		Olive oil	0.4	Herring (raw)	0.8	Bananas	0.3	Beer (canned)	0.01
				Prawns (boiled)	1.1	Plum	0.4	Red wine	0.9

Data abstracted from Holland B, Welch AA, Unwin ID, Bun DH and Paul AA (1991) *McCance and Widdowson's The Composition of Foods*, 5th edn. Cambridge: Royal Society of Chemistry.

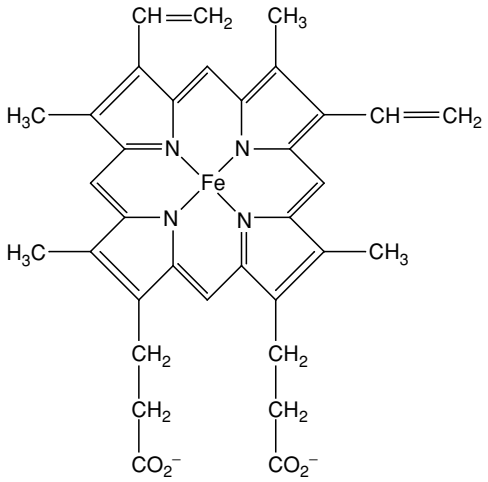

Figure 7

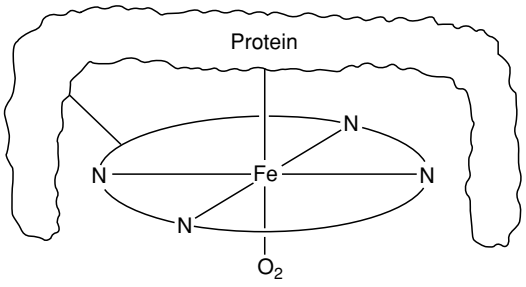

Figure 8

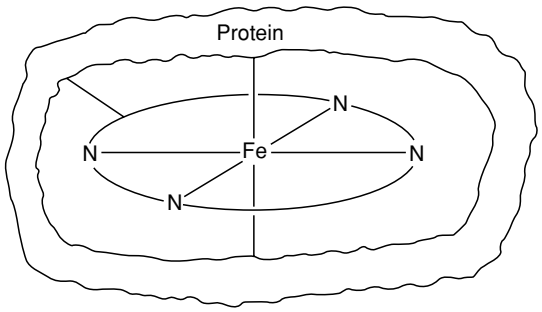

Figure 9

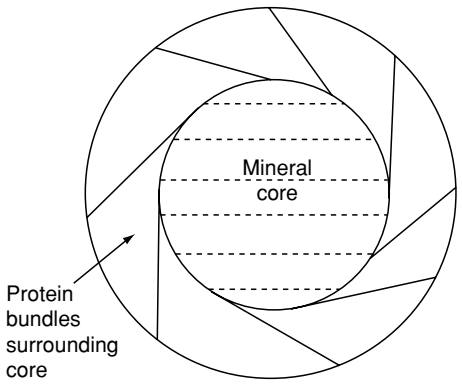

Figure 10

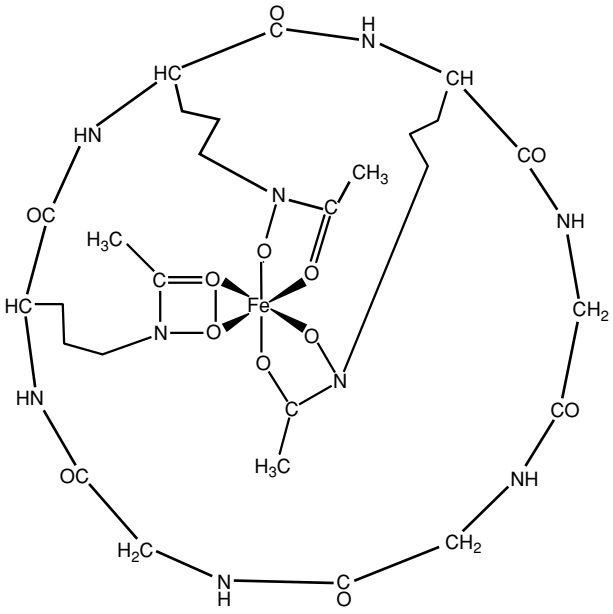

Figure 11

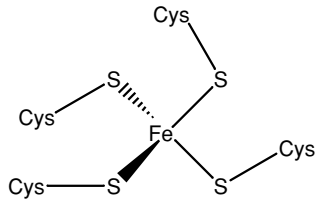

Figure 12

iron atoms, while the iron atoms are linked to the rest of the protein via the sulfur atoms of cysteines. In each of these three types of iron–sulfur protein each iron atom is tetrahedrally surrounded by four sulfur atoms. Compounds containing more than one iron atom show complex redox behavior as each iron atom can be either iron(II) or iron(III).

Environment

Terrestrial The earth's core is composed of metallic iron, and iron compounds form 1.5% of the lithosphere, making iron the eighth most abundant element. The most common iron mineral is hematite (Fe_2O_3), with magnetite (Fe_3O_4), siderite ($FeCO_3$), limonite (FeO(OH)) and iron pyrites (Fool's gold, FeS_2) also widely distributed. At neutral pH, aerial oxygen readily oxidizes iron(II) in solution to iron(III). To be stable, iron(II) compounds must, like iron pyrites, be insoluble.

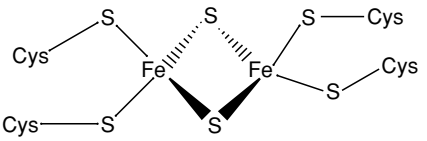

Figure 13

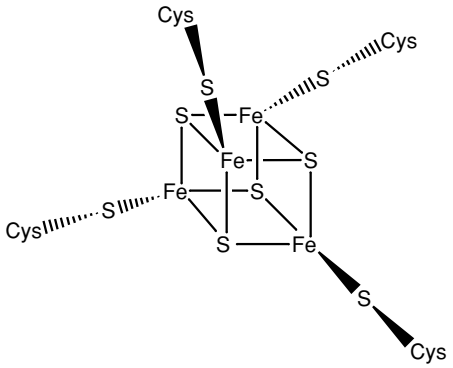

Figure 14

Aquatic Sea water contains low concentrations of iron (0.002–0.02 ppm) complexed by chloride ions. Fresh water, with pH around 7, can contain suspensions of hydrated iron(III) oxide, or iron complexed to organic material.

Losses/Gains during Food Processing (Including Changes in Speciation)

Iron compounds in foods are not appreciably volatile so iron cannot be lost by vaporization during normal processing. As iron compounds occur naturally associated with proteins, the iron compounds found in foods are generally changed on heating, so that the tertiary structure of the protein is destroyed. This will cause the iron compound to lose its functionality. Thus, after heating, myoglobin will no longer reversibly bind oxygen – the iron(II) is irreversibly oxidized to iron(III). Normal cooking will not break the peptide links in the protein, or free iron from its coordination shell.

In principle, foods cooked or processed in iron or steel vessels could potentially pick up significant amounts of iron. However, as commercial processing is now largely carried out in equipment constructed of stainless steel, this is not a major problem. The most likely situation where significant iron pick-up may occur is with acidic foods or those containing high levels of chloride ions. In domestic situations where cast iron pots are used, iron uptake may be appreciable, in relation to the amounts naturally present.

The levels of iron resulting from contamination are unlikely to be significant from a toxicological (or even nutritional) point of view. However, they may lead to metallic taints in some foods and they may accelerate formation of rancidity by acting as a catalyst of lipid oxidation.

Iron Fortification of Foods

Iron compounds used for fortification of foods are varied. These additives are still widely referred to by their older (ferrous/ferric) names. Iron(II) (ferrous) sulfate is the cheapest iron source, and iron introduced this way is well absorbed. Iron absorption depends on the nature of the food. Sources of ascorbic acid enhance iron absorption from additives, while tea inhibits absorption. Choice of fortificant is a balance between using iron in a form which is well absorbed, while not causing unacceptable color or taste changes in the food. Many countries add iron compounds to flour and cereals to counteract iron-deficiency anemia. Flour is fortified to iron levels of between 16.5 ppm (UK) and 44 ppm (US).

Dependent on the liquidity of the product, one can use soluble compounds with good iron bioavailability (iron(II) (ferrous) sulfate, iron(II) (ferrous) gluconate, iron(II) (ferrous) lactate, iron(III) (ferric) saccharate, iron(III) (ferric) ammonium citrate, soluble iron(III) (ferric) pyrophosphate). Care has to be taken to avoid unpleasant taste, undesirable coloring and catalysis of fat oxidation. Less soluble iron sources (iron(II) (ferrous) fumarate, iron(II) (ferrous) succinate, iron(III) (ferric) pyrophosphate, elemental iron, sodium iron(III) pyrophosphate, iron(III) (ferric) orthophosphate) may be better with more solid foods.

Analysis: Isolation and Extraction

It is relatively easy to analyze food samples for their iron content. The method of analysis depends on the nature both of the samples, and of the detection technique selected. It is usual to remove the organic part of foods from the inorganic constituents before analysis. This can be done by wet digestion or dry ashing methods.

Wet digestion

Many wet analytical techniques require the iron to be present as uncomplexed iron(II) or iron(III) ions in the analyte with absence of organic residues. A universal method to achieve this is to heat the sample with a digestion mixture containing concentrated nitric, perchloric and sulfuric acids in a Kjeldahl flask for some hours. All organic matter is oxidized

to water and carbon dioxide, while metal ions are left behind as uncomplexed ions (the anions from these acids are poor ligands). While not necessary for oxidation, the presence of sulfuric acid insures that samples do not dry out to an explosive perchlorate. After digestion, the sample is diluted appropriately.

Dry Ashing

The simplest way to remove organic matter from a sample prior to analysis is by aerial oxidation at red heat in a silica, porcelain, or platinum crucible. Samples (1–10 g) are normally dried carefully at 100–110 °C before heating in a muffle furnace. All iron is oxidized to Fe_2O_3, and organic matter present is oxidized to carbon dioxide and water vapor. At high temperatures, iron porphyrin compounds and others may volatilize before oxidation and iron may be lost, so ashing temperatures must be kept as low as possible – temperatures of 500–550 °C are generally used. Dry ashing methods must be tested for iron recovery before routine use. After ashing, the cooled ash is dissolved in dilute hydrochloric acid.

Analysis: Spectroscopic and Electrochemical Methods

The simplest way to assay the iron content of the solutions resulting from ashing is to reduce all iron to iron(II) using metallic zinc, and then to titrate the iron(II) with potassium permanganate.

Spectroscopic Methods

Instrumental methods, mainly spectroscopic, are now widely used.

UV/Visible Absorbance Both iron(II) and iron(III) form colored complexes, which can be used to assay iron. Samples from ashing normally contain iron as iron(III). Addition of thiocyanate produces a blood-red color of $[Fe(H_2O)_5SCN]^{2+}$, which can be spectrophotometrically measured at 480 nm. Normally samples are reduced to iron(II) with a reductant such as ascorbic acid, or hydroxylamine hydrochloride, the complexing ligands 1,10-phenanthroline (phen) or 2,2′-dipyridyl (dipy) added, and absorbances of the colored complex ions $[Fe(phen)_3]^{2+}$ or $[Fe(dipy)_3]^{2+}$ measured at 505–510 nm.

Atomic Absorbance Spectrophotometry Atomic absorption spectrophotometry (AAS) has been widely used to assay for iron. Solutions containing iron(II) or iron(III) are introduced as an aerosol into the flame where atomization takes place. The method has a detection limit of 0.003 ppm, and is easily automated for the analysis of many samples. Interference from other metals in the sample is minimal. Normally only a single element, e.g. iron, can be analyzed for at a time on each sample.

Inductively Coupled Plasma Optical Emission Spectrophotometry (ICPOES) Solutions containing iron(II) or iron(III) are introduced into a plasma as an aerosol, and the emissions from atomic transitions specific to iron are measured in a suitable spectrophotometer. This technique has a similar sensitivity to AAS, but is more convenient for multielement assays. As atomic transitions are detected, the technique is sometimes referred to as Inductively Coupled Plasma Atomic Emission Spectrophotometry (ICPAES).

Spark Emission Spectroscopy This works in a similar way to ICPOES. The liquid sample is placed so that sparks are formed between the solution and an inert electrode. In the sparks, iron atoms are excited to high energy levels, and emissions at wavelengths specific to iron are measured.

Electrochemical Methods

Spectroscopic methods are now widely used for iron analysis. Electrochemical methods, though effective, are of secondary importance. Iron-containing solutions may be reduced to convert all iron to iron(II) which can be titrated with ceric sulfate, and the endpoint detected potentiometrically. Low concentrations of iron may be determined by polarography, but advanced polarographic techniques such as differential pulse polarography generally give better results.

See also: **Iron**: Properties and Determination; **Spectroscopy**: Atomic Emission and Absorption; Visible Spectroscopy and Colorimetry

Further Reading

Clydesdale FM and Weimer KL (1985) *Iron Fortification of Foods*. Orlando, FL: Academic Press.

Cotton FA, Wilkinson G, Murillo CA and Bochmann M (1999) *Advanced Inorganic Chemistry*, 6th edn. New York: Wiley-Interscience.

Cunniff P (1995) *Official Methods of Analysis*, 16th edn. Arlington: Association of Official Analytical Chemists.

Greenwood NN and Earnshaw A (1997) *Chemistry of the Elements*, 2nd edn. Boston: Butterworth-Heinemann.

Holland B, Welch AA, Unwin ID, Bun DH and Paul AA (1991) *McCance and Widdowson's The Composition of Foods*, 5th edn. Cambridge: Royal Society of Chemistry.

Hughes MN (1987) Coordination compounds in biology. In Wilkinson G, Gillard RD and McLeverty JA (eds) *Comprehensive Coordination Chemistry*, vol. 6. Oxford: Pergamon.

Kaim W and Schwederski B (1994) *Bioinorganic Chemistry: Inorganic Elements in the Chemistry of Life*. Wiley: Chichester.
Neilson SS (1998) *Food Analysis*, 2nd edn. Aspen: Gaithersberg.
Paré JRJ and Bélanger JMR (1997) *Instrumental Methods in Food Analysis*. Amsterdam: Elsevier.
Reilly C (1991) *Metal Contamination of Food*, 2nd edn. London: Elsevier.
Sigel A and Sigel H (1998) *Metal Ions in Biological Systems*, vol. 35. *Iron Transport and Storage in Microorganisms, Plants and Animals*. New York: M Dekker.

Physiology

S R Lynch, Eastern Virginia Medical School, Norfolk, Virginia, VA, USA

Introduction

Iron plays a central role in metabolic processes involving oxygen transport and storage as well as oxidative metabolism and cellular growth. The fact that it readily serves as an electron donor or acceptor accounts both for its critical metabolic role and its potential toxicity. Iron-containing compounds function as carriers for oxygen and electrons and as catalysts for oxidation and hydroxylation reactions. Ionic iron can also participate in reactions that produce toxic free radicals. Free radicals may in turn damage cellular constituents. It is therefore not surprising that body iron content is controlled within narrow limits, and that the metal itself is transported and stored as a component of specific iron-binding proteins rather than as the free cation.

This article reviews the functional role of iron in the body, the physiological processes responsible for the control of internal exchange and balance, and the consequences of iron deficiency and excess. Some aspects of the physiology of iron that are specifically related to iron-deficiency anemia are discussed later. (*See* **Anemia (Anaemia)**: Iron deficiency Anemia.)

Body Iron Distribution

The normal human body contains 3–4 g of iron (40–50 mg per kg of body weight; **Table 1**), 75% (approximately 36 mg kg^{-1}) of which is present in metabolically active compounds. The remainder is contained in a storage pool (approximately 10 mg kg^{-1} in men and 5 mg kg^{-1} in menstruating women) which is readily available if metabolically active iron is depleted for any reason.

Table 1 Distribution of iron in adult human beings

	Men: total body content		Women[a]: total body content	
	mg	mg kg^{-1}	mg	mg kg^{-1}
Functional				
Hemoglobin	2300	31	1700	28
Myoglobin	320	4	180	3
Heme enzymes	80	1	60	1
Nonheme enzymes	100	1	76	1
Storage				
Ferritin	540	7	200	3
Hemosiderin	235	3	100	2

[a]Age range 18–44 years.

Functional Iron Compartment

Heme Proteins

Most of the functional iron in the body is present in the form of heme proteins, i.e., proteins with an iron protoporphyrin IX prosthetic group.

Hemoglobin, which is made up of four globin chains, each with an attached heme group, transports oxygen to the tissues. It is quantitatively the most important heme protein and contains 80% of all functional iron.

Myoglobin is found in the sarcoplasm of muscles. It has a structure similar to hemoglobin but contains only one globin chain attached to a single heme group and accounts for a further 10% of functional iron. Myoglobin acts as an oxygen store, insuring an adequate oxygen supply during muscle contraction. (*See* **Exercise**: Muscle.)

Despite its vital metabolic role, all other tissue iron represents only a small fraction of total body iron. The cytochromes are a group of heme-containing electron transport enzymes that are essential for the oxidative metabolism necessary to generate adenosine triphosphate (ATP) as well as for the oxidative degradation of drugs and endogenous substrates. Catalase and peroxidase are involved in the reduction of endogenously generated hydrogen peroxide.

Nonheme Tissue Iron

In mitochondria, nonheme compounds account for more iron than do those containing heme. This group of enzymes includes the metalloflavoproteins, the iron sulfur proteins, and ribonucleotide reductase. In addition, iron is necessary in a loosely bound form for the activity of other enzymes, such as those responsible for the hydroxylation of proline and lysine in protocollagen, during the synthesis of collagen. (*See* **Coenzymes**.)

All functional iron compounds are constantly being degraded and replaced by newly synthesized material.

Internal iron exchange therefore plays a crucial role in preserving normal iron-dependent metabolic processes.

Iron Transport and Storage

Iron entering the plasma is rapidly bound to the specific iron transport protein, transferrin. The iron-free protein apotransferrin is a single-chain glycoprotein (mol wt 79 570) with two nonidentical iron-binding sites that have a high affinity for ferric iron under physiological conditions (effective stability constant, $10^{24}\,\mathrm{mol\,l^{-1}}$). Plasma apotransferrin is synthesized predominantly in the liver. It exists in the plasma in the iron-free form or as monoferric or diferric transferrin since iron loading at each binding site is a random process.

Iron delivery from plasma transferrin to the tissues is mediated by a specific transferrin receptor which is a transmembrane glycoprotein dimer composed of two identical subunits (each with mol wt 94 000) linked by a disulfide bond. Transferrin receptors are expressed on the surfaces of all cells in proportion to their iron requirements. Large numbers are present in tissues with high requirements, e.g., developing red blood cells and placenta.

At the pH of plasma and extravascular fluid bathing cell surfaces, receptors have very little affinity for apotransferrin and the highest affinity for diferric transferrin ($2–7\times10^{-9}\,\mathrm{mol\,l^{-1}}$). Once bound, the transferrin–transferrin receptor complex, together with its attached iron, is internalized by the cell in a clathrin-coated pit that closes to form an endosome. The endosome then fuses with an acidic vesicle ($\mathrm{pH} < 5.5$). The fall in pH results in release of iron from the transferrin. The iron is transferred across the endosomal membrane into the cell. Natural resistance-associated macrophage protein 2 (Nramp 2, now also called divalent metal transporter 1, DMT-1) has recently been identified as the putative transmembrane iron transport protein that transfers iron out of the endosome. The rat isoform of Nramp 2 is a divalent cation transporter (DCT-1) with a broad substrate range that includes ferrous iron. The transferrin receptor remains intact. Its affinity for apotransferrin increases, becoming equal to that for diferric transferrin, because of the lower pH. The complex is transported back to the cell surface, where the apotransferrin is released back into the circulating plasma.

Any iron entering a cell that is not immediately used for the synthesis of metabolically active compounds is stored in the form of ferritin. Apoferritin is a hollow, spherical protein shell composed of 24 subunits that may be of two types, differing slightly in molecular weight – L (mol wt 19 700) and H (mol wt 21 100). Each complete apoferritin molecule can store as many as 4500 iron atoms within its central core as ferric hydroxyphosphate. Iron enters and leaves the intact protein through channels in the shell.

Catabolism of cellular ferritin may result in the formation of a second type of iron storage protein, hemosiderin, which is water-insoluble. It has both a higher iron content and a slower turnover than ferritin.

The acquisition and storage of iron by cells is regulated by the translational control of the synthesis of transferrin receptors and of apoferritin. Two iron-regulatory proteins (IRP-1 and IRP-2) that both sense and adjust cellular iron supply have been identified. They are cytoplasmic RNA-binding proteins that modulate the expression of messenger RNA (mRNA) for transferrin receptor and apoferritin by binding to iron-responsive elements (IREs) on the 3′ and 5′ untranslated regions of the mRNAs for transferrin receptors and the H- and L-chains of ferritin respectively. Low cellular iron levels favor increased binding of the IRPs to the IREs, repressing the synthesis of ferritin, but stabilizing transferrin receptor mRNA against cellular ribonucleases, thereby increasing transferrin receptor expression and cellular iron uptake. High cellular iron leads to decreased IRP binding with a decrease in iron uptake and increased ferritin synthesis and iron storage.

The two IRPs are functionally similar, but are regulated in different ways. In iron-replete cells IRP-1 has been identified as cytosolic aconitase, an iron-sulfur protein that mediates the enzymatic interconversion of citrate to isocitrate in the tricarbocylic acid cycle. Low cellular iron results in reversible conversion of enzymatically active aconitase to a form with high RNA-binding affinity and no enzymatic activity. Transferrin receptor synthesis is enhanced, apoferritin formation is suppressed, and functional iron homeostasis is restored. IRP-2 is functionally similar to IRP-1, but lacks aconitase activity. Unlike IRP-1, the binding of IRP-2 to the IREs is regulated by degradation of the protein when cells are iron-replete.

The above description characterizes iron transport and storage in most cells of the human body. However, erythroid cells have very high iron requirements and appear to possess mechanisms for regulating iron uptake at a transcriptional level that can override posttranscriptional control. Iron transport and storage by the macrophages of the spleen, bone marrow, and liver are also different. These are cells primarily involved in processing hemoglobin derived from senescent red blood cells. At the end of their life span, erythrocytes are phagocytosed by macrophages,

predominantly in the spleen. The heme is separated from the globin and rapidly catabolized by the enzyme, heme oxygenase. Iron is released and either returned to the plasma within a few hours or incorporated into the storage compartment of the cell with a more gradual return to the plasma (half-life about 7 days). Normally, two-thirds of the iron is released immediately, but this fraction may be either increased when demands are high or reduced when less iron is required.

Internal Iron Exchange

The development of radioiron tracers made it possible to quantify the processes involved in the internal iron exchange described above. Since 80% of the body's functional iron is in the hemoglobin of the circulating red blood cells, measurements of internal iron exchange are dominated by the requirements of this compartment (Figure 1). Complete exchange of the iron in the circulating red blood cell compartment occurs every 4 months. This involves rapid transfer of iron by plasma transferrin. Only 3–4 mg iron is found in the plasma at any one time, but 35 mg is transported through this compartment each day. Most of it comes from hemoglobin catabolism in macrophages. Two-thirds (24 mg day^{-1}) is delivered to erythroid precursors in the bone marrow for the synthesis of new hemoglobin. While there is some iron loss owing to ineffective red cell production or the removal of iron not used for hemoglobin synthesis from red cell precursors, most of the iron (70%) is returned to the circulation as hemoglobin in erythrocytes. Erythrocytes have a life span of about 120 days. The senescent cells are catabolized in the macrophages of the liver, spleen, and bone marrow, completing the cycle.

Most of the iron lost during red cell production and the iron derived from senescent red cells is processed by macrophages. Thus a quantity of iron equivalent to about 1% of the functional hemoglobin iron in the circulating red cell compartment (22 mg iron) enters the macrophages each day with a corresponding quantity being released to the plasma.

A smaller but quantitatively significant daily exchange occurs between the plasma iron and storage iron in hepatocytes. The rate of exchange and direction of net flow is dependent on serum iron concentration and transferrin saturation. A high serum iron concentration favors the accumulation of iron in the liver.

Finally, a minor fraction of the daily iron turnover (about 3 mg day^{-1}) is transferred between the plasma and the extravascular transferrin compartment and approximately two-thirds of this (2 mg day^{-1}) exchanges with the tissues, supplying iron to heme and nonheme iron-dependent enzymes.

Iron Absorption and Excretion

The total body iron content of healthy human beings is held within narrow limits. This is primarily a consequence of the fact that most of the iron is located in tightly regulated functional compartments. Iron stores usually account for only 15–30% of total body iron. The size of the storage compartment may vary 25-fold without any apparent physiological impairment.

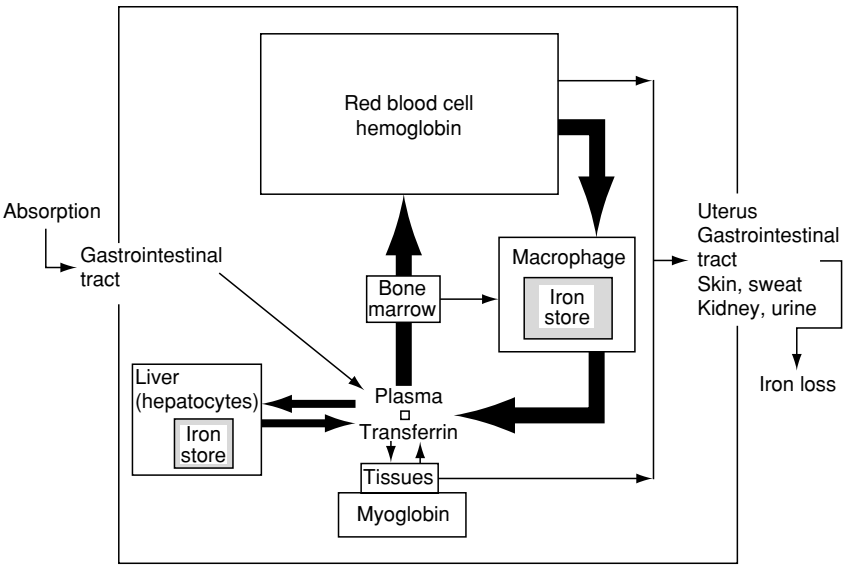

Figure 1 Body iron exchange.

Iron Loss

In contrast to the dynamic transfer of iron between body compartments, exchange with the external environment is minimal. No adjustable excretory mechanism exists, but a small obligatory loss occurs with the physiological turnover of skin epithelium and the cells of the gastrointestinal and urinary tracts (Figure 1). Low concentrations of iron are also present in sweat, bile, and urine. In addition, the feces contain small quantities of blood. Finally, menstruation accounts for a significant proportion of the iron lost by women of child-bearing age.

In normal men the loss of iron from the body is about 1 mg day^{-1}. This is balanced by an equivalent absorption. Complete iron exchange with the environment in a normal man would therefore be expected to take about 10 years. In women menstrual losses account for an extra 0.5 mg day^{-1}. The higher iron loss is matched by a higher rate of absorption.

Iron Absorption

Typical western diets contain approximately 1.5 mg iron MJ^{-1}. From the point of view of absorption, food iron may be considered to exist in one of three forms. Ten percent or less is present as heme derived chiefly from hemoglobin and myoglobin in meat. Heme iron is readily absorbed and little affected by dietary factors. This small fraction may supply a third of the iron requirements of individuals eating a mixed meat-containing diet. The remainder of the iron (90% or more in western diets and often virtually 100% in the diets of developing countries) must be solubilized for absorption. If solubilized, nonheme iron derived from all dietary sources enters a common pool in the lumen of the upper small intestine before absorption. A variable proportion of nonheme iron is insoluble and unavailable for absorption. This iron is generally regarded as contaminant iron. Much of it enters food products during storage or processing, particularly in developing countries. (*See* **Meat: Nutritional Value.**)

Absorption of soluble nonheme iron from the common pool is quite variable. It is governed both by the body's requirement and the balance between enhancing and inhibiting ligands in the diet that either promote uptake by mucosal cells (increase bioavailability) or render the iron in the pool unavailable for absorption. Enhancing factors include ascorbic acid, other organic acids, and meat and fish tissue. Low gastric luminal pH resulting from hydrochloric acid secretion by the stomach or from the ingestion of acidic foods also promotes nonheme iron absorption. The most powerful inhibitors are found in vegetable foods and include phytates, polyphenols, and vegetable proteins. (*See* **Bioavailability of Nutrients.**)

Absorption of both heme and nonheme iron is maximal in the duodenum, partly because the luminal conditions (particularly lower pH) favor solubilization and absorption of nonheme iron, but also because cells in his region show the greatest ability to respond to changes in body iron needs.

Absorption may be considered to occur in three phases: first, uptake across the mucosal cell brush border; second, a phase that involves either rapid transfer of iron through the mucosal cells or its retention in the cellular ferritin store; and third, transfer from the mucosal cell to plasma transferrin. Most of the iron retained in the cell is lost when the cell exfoliates. The absorption of both heme and nonheme iron is regulated by the size of the body iron stores.

Heme iron enters the mucosal cell through a pathway different from that for nonheme iron. It is taken up as the intact heme moiety. Once within the cell, iron is released from heme by the enzyme, heme oxygenase. It joins an absorption pathway common to both heme and nonheme iron.

Percentage absorption from the small heme pool varies about twofold. The primary adaptation to changes in iron requirements occurs in the adjustment of absorption from the larger nonheme pool. Percentage absorption is inversely correlated with the size of the body iron stores and may vary as much as 20-fold. Regulation occurs both during mucosal iron uptake by the enterocyte and during release of iron from the cell to circulating transferrin. Mucosal uptake is the rate-limiting step.

Despite years of research, the precise molecular mechanisms involved in the uptake of elemental iron into mucosal cells and its transfer to the plasma remain a subject of controversy. However, it was recently discovered that two isoforms of Nramp 2 (DMT-1) are expressed at the brush borders of the apical poles of enterocytes in the duodenum, the site of maximal iron absorption. Dietary iron starvation leads to a marked upregulation of Nramp 2 isoform 1 in proximal duodenal enterocytes. Cells in the rest of the small intestine are not affected. Nramp 2 mRNA has an IRE in its 3′ untranslated region.

It has been postulated that the iron status of duodenal enterocytes is determined by their iron uptake from the circulation during development in the duodenal crypts. They only become functional iron-absorbing cells 48 h later when they reach the duodenal villous tips. According to the theory the cells are programmed to regulate iron absorption during this period of time by an IRP–IRE mechanism. In the face of iron deficiency, enterocytes acquire less iron from the plasma during their early development.

DMT-1 is upregulated and apoferritin synthesis limited. The cells absorb and transfer dietary iron through the basolateral membrane to circulating plasma transferrin efficiently. If the individual is iron-sufficient, the enterocytes acquire more iron during early development. They will express less DMT-1, take up less iron from the duodenal lumen, and store some of the absorbed iron intracellularly as ferritin. Very little of the iron stored as ferritin reenters the absorptive pathway at a later stage. Most of it is lost when the cell exfoliates.

Iron Deficiency

In western countries iron deficiency occurs most often when there is a relatively sudden increase in iron requirements or iron losses, e.g., during pregnancy or in association with pathological blood loss.

The physiological importance of both the storage iron compartment and the capacity for the rapid transfer of iron through the plasma in preventing overt functional iron deficiency is illustrated by the effect of blood loss. In response to significant anemia caused by bleeding, individuals with an iron store of 1000 mg can mobilize 40 mg day^{-1}, allowing rapid restoration of the functional deficit. On the other hand, an individual who lacks storage iron and depends on absorption from an average diet for the additional iron will increase delivery to the bone marrow by only 2–4 mg day^{-1}. Red cell production increases by a small margin. Thus, loss of iron from the major functional compartments is rapidly corrected when stores are adequate, but very slowly replaced by absorption, even when dietary iron bioavailability is relatively high. If meal iron bioavailability is low, positive balance may not be achieved, leading to chronic iron-deficiency anemia. Low bioavailability diets also lead to iron deficiency when requirements are increased by rapid growth or menstruation.

Negative iron balance leads to a reduction in the iron content of all functional compartments. Anemia caused by reduced hemoglobin synthesis is the most easily documented. Nevertheless, the availability of iron to support metabolic systems in the tissues is reduced concurrently. The physiological consequences of iron deficiency therefore result both from impaired oxygen delivery and functional abnormalities in the tissues. However it is noteworthy that clinically significant tissue iron deficiency does not appear to occur in the absence of anemia.

Effects of Iron Deficiency

The clinical manifestations of moderate to severe anemia are well known and include pallor, fatigue, weakness, dizziness, and reduced maximal work capacity. The consequences of anemia *per se* and the adverse effects of iron-deficiency anemia on the outcome of pregnancy are dealt with in the section on iron-deficiency anemia. (*See* **Anemia (Anaemia)**: Iron-deficiency Anemia.)

Several additional consequences are thought to result primarily from the associated tissue iron deficiency. From the epidemiological point of view, the most important are developmental delay and cognitive impairment in childhood and impaired work performance at all ages.

Mental and Motor Development

Iron-deficiency anemia is associated with impaired mental development and physical coordination in children under the age of 2 years, difficulty with visual discrimination tests, and the ability to maintain selective attention in preschool children and poor school achievement in later childhood. Short-term memory and emotional health may also be affected. These functional abnormalities are thought to be the result of the effects of iron deficiency on brain neurotransmitters, but their biochemical basis is poorly understood. In older children they appear to be reversible by successful iron supplementation.

Skeletal Muscle Dysfunction

Significant limitation of the ability to perform endurance physical activity has emerged as an important consequence of chronic iron deficiency. Animal studies conducted by Finch and coworkers demonstrated that iron-deficient rats show marked impairment of running ability which is unrelated to hemoglobin level. It results from impaired oxidative metabolism in iron-depleted muscles. Field studies from many developing countries suggest that a similar disability reduces an iron-deficient individual's ability to carry out prolonged physical work.

Immunity and Infection

Several laboratory tests of immune function are abnormal in patients with iron-deficiency anemia. Lymphocyte proliferation in response to the mitogens, phytohemagglutinin, and concanavalin A is reduced, demonstrating defective T-cell immunity. Intracellular bacterial killing by polymorphonuclear leukocytes is impaired. The clinical importance of these laboratory observations is uncertain, although some studies have suggested that the administration of iron to iron-deficient children may reduce the prevalence of enteritis and influenza-like illnesses. Iron deficiency also appears to be a predisposing factor in chronic mucocutaneous candidiasis.

Miscellaneous Liabilities

Angular stomatitis, glossitis, postcricoid webbing of the esophagus associated with painful dysphagia (Patterson–Kelly or Plummer–Vinson syndrome), atrophic gastritis, and koilonychia (spoon-shaped finger nails) have all been attributed to tissue iron deficiency. These clinical findings appear to be less common in recent years and a marked geographic variation in prevalence has frequently been noted, suggesting that factors other than iron deficiency may play an important role.

An intriguing sensory disturbance encountered in both children and adults who are iron-deficient is the perversion of taste leading to the consumption of nonfood items (pica) or compulsive ice eating (pagophagia). The specificity of pagophagia as a symptom of iron deficiency has been confirmed by the study of patients in whom iron deficiency was induced by phlebotomy alone, making the contribution of confounding nutritional and social factors unlikely. It is corrected by iron repletion.

Iron deficiency may also lead to impaired thermogenesis as well as abnormalities in thyroid metabolism and catecholamine turnover. Finally, lead absorption may be enhanced, increasing the risk of lead toxicity, particularly in children.

Iron Toxicity

Acute Iron Toxicity

The ingestion of large quantities of elemental iron can cause acute iron poisoning. This occurs most often in young children who may eat iron tablets as 'candy.' The pathological consequences include a severe necrotizing gastroenteritis as well as disseminated intravascular coagulation, liver and cardiac injury. Death may ensue if the dose is large or treatment is not instituted immediately.

Chronic Iron Toxicity

More important from the nutritional point of view is the gradual accumulation that occurs when the quantity of iron entering the body exceeds requirements by even a small margin. The body has no means of increasing iron excretion significantly, making a positive iron balance inevitable if regulation of absorption is impaired, or if the diet contains a quantity of available iron that overwhelms the absorptive control mechanisms. Excess iron can also be introduced through the parenteral route (blood transfusion, parenteral iron administration).

Impaired regulation of absorption Hereditary hemochromatosis is the commonest cause of iron overload resulting from the impaired regulation of iron absorption, and the commonest form of iron overload in the USA, western Europe, and Australia. It is an autosomal recessive disorder with a gene frequency reported to be as high as 1 in 10 in people of northern European descent. One in 300 are homozygous and at risk for developing severe iron overload with its associated pathological consequences. A single G-to-A mutation resulting in a cysteine-to-tyrosine substitution (C282Y) in the HFE gene has recently been shown to be responsible for the majority of cases of human leukocyte antigen (HLA)-linked hereditary hemochromatosis. The HFE gene is located on the short arm of chromosome 6 and is tightly linked to the HLA locus. It has been identified in between 60% and 100% of patients depending on whether their ethnic origin was southern or northern Europe. A second mutation, H63D, has been linked to an additional 1–10% of cases, especially when present with C282Y in the compound heterozygous form. The mechanisms by which mutations in the HFE gene lead to tissue iron overload have not yet been established.

HFE-associated hemochromatosis is characterized by a rate of iron absorption that is inappropriately high for the size of body iron stores. Downregulation of nonheme iron absorption in the face of rising storage iron is impaired and heme iron absorption virtually unregulated. Equally important to the pathogenesis of the condition is the presence of an abnormally high serum iron and transferrin saturation associated with disordered iron distribution within the body, favoring iron accumulation in the parenchymal cells of organs such as the liver, heart, and pancreas. Comparatively less iron is located in the normal macrophage iron store.

Although the defective control of absorption is an inborn error of metabolism manifest from birth, the prevalence of the clinical syndrome is highest in men over the age of 40 years. Excess iron is accumulated slowly because maximal absorption from the average western diet is only 3–5 mg day^{-1}. Once the total body iron load has increased to 15 g or more (a process which takes 15–20 years), organ damage becomes evident.

The exact mechanism by which the iron injures tissues has not been established, but lipid peroxidation in membranes and subcellular organelles, as well as iron-induced lysosomal disruption, are probably involved. The clinical consequences include increased skin pigmentation, cirrhosis of the liver associated with an increased risk of developing liver cancer, congestive cardiac failure and cardiac arrhythmias, diabetes mellitus, and hypogonadism owing both to end organ and pituitary dysfunction. In addition,

patients with hemochromatosis often suffer from an arthropathy characterized by chondrocalcinosis and a predilection for involvement of the second and third metacarpophalangeal joints of the hand.

If HFE-associated hemochromatosis is identified before significant iron overload has occurred, all of the clinical findings may be averted by iron removal. Therapeutic phlebotomy is usually employed. Unchecked progressive iron overload is associated with severe irreversible morbidity and a significantly shortened life span.

Excessive iron absorption from a normal diet may also occur in association with certain iron-loading anemias, chronic liver disease, and porphyria cutanea tarda.

Dietary iron overload is a form of chronic iron toxicity unique to the indigenous population of several countries in southern Africa (subSaharan iron overload). Traditional acidic fermented beverages brewed in containers made from iron become contaminated with iron at concentrations of 15–40 mg l^{-1}. The beer has a low alcohol content and the large volumes consumed may supply 50–100 mg of available dietary iron per day. Total body iron burdens comparable to those encountered in hereditary hemochromatosis are encountered. Organ damage with progression to cirrhosis and the development of diabetes mellitus occur. However in the earlier stages cellular iron distribution tends to be different from that seen in HFE-associated hereditary hemochromatosis with greater accumulation in the macrophages of the spleen, liver, bone marrow, and muscle. Two other clinical conditions are commonly associated with the presence of excess iron – ascorbic acid deficiency and osteoporosis. Ascorbic acid deficiency results from accelerated catabolism in patients with severe iron overload, often in association with a low dietary intake. The factors responsible for the osteoporosis remain poorly understood.

SubSaharan dietary iron overload was originally believed to be solely a consequence of excessive iron consumption. Recent investigations have challenged this belief. Family studies suggest that iron loading occurs only when iron intake is excessive and a putative non-HLA-linked iron-loading gene is present.

See also: **Anemia (Anaemia)**: Iron-deficiency Anemia; **Bioavailability of Nutrients**; **Coenzymes**; **Exercise**: Muscle; **Immunology of Food**; **Meat**: Nutritional Value

Further Reading

Brock JH, Halliday JW, Pippard MJ and Powell LW (1994) *Iron Metabolism in Health and Disease*. London: WB Saunders.

Canonne-Hergaux F, Gruenheid S, Ponka P and Gros P (1999) Cellular and subcellular localization of the Nramp 2 iron transporter in the intestinal brush border and regulation by dietary iron. *Blood* 93: 4406–4417.

Dallman PR (1982) Manifestations of iron deficiency. *Seminars in Hematology* 19: 19–30.

Finch CA and Huebers H (1982) Perspectives in iron metabolism. *New England Journal of Medicine* 306: 1520–1528.

Klausner RD, Rouault TA and Harford JB (1993) Regulating the fate of mRNA: the control of cellular iron metabolism. *Cell* 72: 19–28.

Ponka P (1997) Tissue-specific regulation of iron metabolism and heme synthesis: distinct control mechanisms in erythroid cells. *Blood* 89: 1–25.

Biosynthesis and Significance of Heme (Haem)

J J Klawe, The Ludwik-Rydygier Medical University, Bydgoszcz, Poland

Iron – Biological Significance

Iron is important for the synthesis of hemoglobin, myoglobin, and as a cofactor of numerous enzymes, including cytochromes, cytochrome oxidase, peroxidase, and catalase.

The total quantity of iron in the human body averages about 4.5 g, 65–70% of which is present in the form of hemoglobin. Several forms of iron 'biological container' are presented in **Table 1**.

The maintenance of iron balance is essential and necessary for health. Iron deficiency causes anemia. Prolonged iron overload causes an accumulation of hemosiderin in the tissues, clinically manifested as hemosiderosis, and its severe form with damage of tissues – hemochromatosis.

Table 1 Forms of iron 'biological containers'

Biological iron storage	*% of total iron quantity*
Hemoglobin	70
Myoglobin	3–4
Various heme enzymes controlling intracellular oxidation	1
Transferrin in the blood plasma	0.1
Ferritin or hemosiderin	15
Other forms	10–15

Iron – Absorption, Transport, and Loss

Iron is readily absorbed in the state Fe^{2+}, but most dietary iron is in the ferric form Fe^{3+}. The gastric secretions dissolve the iron and permit it to form soluble complexes with ascorbic acids and other substances that aid its reduction to the Fe^{2+} form. In humans iron-deficiency anemia is a relatively frequent complication of partial gastrectomy. Heme is also absorbed because of its solubility. Fe^{2+} contained in heme is released into mucosal cells.

Iron is absorbed from the upper part of the small intestine, because duodenum and adjacent jejunum contain most mucosal cells suitable for iron absorption. Various dietary factors influence iron absorption, for example, phytic acid from cereals, which forms iron-insoluble compounds in the intestine, as do phosphates and oxalates.

Nonheme iron binds to mucin. Cell adhesion molecule integrin transfers the iron to mobilferrin. Mobilferrin carries iron into the cell and binds more iron in iron-deficiency states. The mucosal cells contain iron-binding protein, or apoferritin. Some iron is utilized by mitochondria, and the remainder is partitioned between apoferritin in the mucosal cells and transferrin. Transferrin is a protein that carries iron in the blood. Specific cell-surface receptors deliver the transferrin with its bound iron to endosomes. Here the low pH induces the separation of transferrin and the iron. Excess iron in the blood is deposited in all cells of the body, especially in liver cells. Apoferritin is also found in many other tissues, where, together with iron, it forms ferritin. Apoferritin is a globular protein. Iron forms a micelle of ferric hydroxyphosphate. In ferritin 24 subunits of protein surround each micelle. The ferritin micelle can contain 4500 atoms of iron. It is the iron storage form in tissues. Molecules of ferritin may aggregate in lysosomal membranes as deposits, containing up to 50% iron. These deposits are called hemosiderin.

Iron is also transported across the placenta from mother to fetus. Maternal iron plasma concentration is usually lower than fetal concentration. However, the binding capacity of maternal plasma for iron is much higher, even though iron is actively transported into the fetal compartment. A daily transport of about 1.8 μmol is necessary to meet total fetal requirements.

In adult humans, the amount of iron lost from the body is relatively small. It is unregulated, and total body stores of iron are regulated by changes in the rate at which iron is absorbed. Men lose about 0.6 mg day^{-1}, whereas women have a variable, larger loss because of the additional iron lost in blood shed during menstruation.

Biosynthesis and Significance of Heme

Evolution of higher organisms has been accompanied by the development of oxygen transport protein. In humans this protein – hemoglobin – is concentrated in specialized cells – erythrocytes. Each erythrocyte contains about 300 million hemoglobin molecules. The organic structure of a protein is unable to direct binding of oxygen. However, certain metals, in their lower oxidation states (for example, Fe^{2+}) have a strong tendency to bind oxygen. Thus, in the evolution of the hemoglobin–myoglobin family of proteins, Fe^{2+} has been utilized in the O_2-binding site.

There are a number of possibilities in which various iron-containing proteins hold iron in the form Fe^{2+}. Hemoglobin and myoglobin are a family of proteins, in which iron is chelated by the tetrapyrrole ring system – protoporphyrin IX. It is one protein in a large class of porphyrins, which are also encountered in chlorophyl, cytochrome proteins, and other natural pigments. The porphyrins are colored. The iron porphyrin in hemoglobin accounts for the red color of blood. The complex of protoporphyrin IX with Fe^{2+} is called heme. Ferrous iron is octahedrally coordinated, which means it should have six ligands or binding groups attached to it. The nitrogen atoms of the porphyrin ring account for only four ligands. Two remaining available coordination sites lie along an axis perpendicular to the plane of the ring. In myoglobin one of these sites is occupied by nitrogen of histidine (part of the protein helix).

When oxygen is bound, the O_2 molecule occupies the vacant side. Hemoglobin has evolved from myoglobin and forms a tetrameric structure. Each of the four chains in hemoglobin has a folded structure similar to that of myoglobin, and each carries a heme. Hemoglobin contains four subunits: two α-chains and two β-chains. Both chains are very similar but are distinguishable in primary structures and folding. Each subunit has primary, secondary, and tertiary structures. Amino acid side chains in hemoglobin provide hydrophobic, salt bridges, and hydrogen bond interactions. These are necessary to stabilize a particular quaternary structure.

Each hemoglobin molecule can bind four oxygen molecules, in four myoglobin-like sites. To simplify, we can consider hemoglobin as having two states of quaternary structure: one characteristic for the deoxy form and the other for the oxygenated form. The oxy form has a higher affinity for O_2. The transition from the deoxy to the oxy conformation involves changes in the interactions between the subunits. A ligand oxygen, binding to its side, tends to pull the Fe^{2+} a very short distance down into the heme and flattens the heme. Consequently, a molecular rearrangement

occurs, and pulling the Fe into the heme produces a much larger shift in the surrounding structure, particularly at the αβ interfaces. Cooperative binding, oxygen transport, and allosteric transition between structurally different high-affinity and low-affinity states mean that hemoglobin behaves in an allosteric way. Several factors, such as carbon dioxide, protons, and other substances, promote these changes. They are called allosteric effectors.

For each polypeptide chain produced by an organism there exists a corresponding gene. The nucleotide sequence in the gene dictates, via the genetic code, the amino acid sequence of the protein. This process is exemplified by the β-globin gene. In the gene for human β hemoglobin, introns (noncoding regions) alternate with exons (regions that are expressed in the polypeptide sequence). From this gene is produced a primary transcript (pre-mRNA), which is spliced to yield the final mRNA. This mRNA is translated into the β-chain, which then adopts its favored three-dimensional structure. Very primitive animals had only a myoglobinlike, single-chain globin for oxygen storage (transport was not necessary; most of these animals were very small). More than 800 million years ago the primitive globin gene was duplicated. The one copy evolved into the gene for an oxygen transport protein and to hemoglobins. The evolutionary line led to mammals carrying both α- and β-globin genes and capable of forming tetrametric hemoglobins. Gene duplication has also occurred in the hemoglobin line, leading to embryonic and fetal forms, and accounting for the special hemoglobin, which is adapted to promote oxygen transfer through the placenta from the mother to the fetus.

Evolution of hemoglobin genes continues. This fact is seen in the existence of several hundred recognized mutant hemoglobins in the human population. Each of these mutant forms exists in only a small part of the total human population. Many give rise to recognized pathologies, while others are referred to as neutral mutations.

See also: **Anemia (Anaemia)**: Iron-deficiency Anemia; **Ascorbic Acid**: Properties and Determination; Physiology; **Iron**: Properties and Determination; Physiology

Further Reading

Ganong WE (1999) *Review of Medical Physiology*. New York: Appleton & Lange.

Greger R and Windhorst U (1996) *Comprehensive Human Physiology*. London: Springer.

Mathews CK and Van Holde KE (1988) *Biochemistry*. London: Benjamin/Cummings Publishing.

IRRADIATION OF FOODS

Contents

Basic Principles

P B Roberts, Institute of Geological and Nuclear Sciences Ltd, Lower Hutt, New Zealand

Introduction

The absorption of energy during irradiation can provide foods with desirable new benefits. Irradiation is therefore analogous to more common processes such as pasteurization or drying that use heat energy. The energy absorbed during irradiation is high-energy, ionizing radiation. (*See* **Drying**: Theory of Air-drying; **Pasteurization**: Principles.)

Irradiation of food must not be confused with the contamination of food by radioactive materials, which themselves emit radiations that may harm a consumer. Such contamination followed nuclear weapon tests and the Chernobyl accident. Food irradiation cannot make food radioactive since the radiation used, though of high energy, is not powerful enough to induce the necessary changes in atomic nuclei. (*See* **Radioactivity in Food**.)

This article discusses the history and current status of food irradiation. The radiation sources and

potential applications are outlined and the general chemical, microbiological, and nutritional effects summarized.

History

Before 1950

Röntgen discovered X-radiation in 1895 and Becquerel observed natural radioactivity in 1896. The radiations produced were soon found to have many biological effects, including the ability to inactivate bacteria. In 1904 Prescott suggested that spoilage bacteria in food could be destroyed by radiation. Patents were obtained in the US and France in 1921 and 1930. However, no practical source of radiation to treat significant food volumes existed until the 1940s when high-energy electron accelerators and nuclear reactors able to produce large quantities of artificial radioactive nuclides became available.

1950–1983

The US Army Quartermaster Corps sponsored the first cooperative effort to develop food irradiation as part of research into light-weight, shelf-stable army rations. By 1966 safety studies on 21 foods were complete and sterilization of bacon and pork, disinfestation of wheat and its products, and inhibition of potato sprouting were uses of radiation approved by the US Food and Drug Administration (FDA). After this, safety studies in the US stalled for 10 years and commercial interest waned.

European interest increased, however, and in 1970 the International Project in the Field of Food Irradiation (IFIP) was initiated. IFIP compiled safety data on many foods to be considered by expert committees that were sponsored by the United Nations organizations responsible for secure, safe food supplies and for the peaceful uses of atomic energy.

In 1980 a joint expert committee on food irradiation stated that 'the irradiation of food up to an overall average absorbed dose of 10 kGy presents no toxicological hazard' and 'introduced no special nutritional or microbiological problems.' This advice prompted the Codex Alimentarius Commission, the body responsible for world food standards, to adopt a worldwide general standard for irradiated food in 1983. The standard sets general conditions for the irradiation of food. It does not imply that irradiation is appropriate for all foods but that irradiation should be regarded and controlled like other physical processes such as canning, dehydration, etc.

Current Status

The Codex standard was issued at a time of increased awareness of food losses in developing countries and a growing aversion to chemical contamination of food via chemical fumigation and preservation in developed countries. By 1999, 41 countries had cleared one or more foods for irradiation. There are over 50 irradiation plants in nearly 30 countries irradiating food, although the volume treated in most is small.

However, the 1980s saw a rise in public suspicion about any technology associated with radiation. Demand for 'natural,' minimally processed foods has increased and official assurances on the wholesomeness of processed foods and additives are increasingly questioned. There has been sufficient public opposition to food irradiation to insure a very cautious attitude to it within the food industry. The annual volume of food irradiated is probably only 500 000 tonnes worldwide. Foods that are irradiated include dried herbs and spices, potatoes, onions, garlic, dried vegetables, tropical fruits, and chicken. Countries that regularly irradiate foods on a commercial basis include the US, The Netherlands, France, Belgium, Japan, South Africa, and China.

The Radiations Used

The changes in food or its contaminating organisms achieved by irradiation processing are caused by chemical reactions induced between the constituent atoms or molecules. These reactions are initiated by radiations that can strip electrons away from atoms or molecules that are left positively charged (ionized). The charged species rapidly split into fragments called free radicals that have a free or unpaired electron. Free radicals react extremely rapidly with each other and nearby molecules as they seek to become more stable by gaining or losing an electron. It is these free radical reactions that trigger the chemical effects leading to the ultimate changes in the food.

There are two principal forms of ionizing radiation: accelerated subatomic particles and high-energy electromagnetic (EM) radiation.

Accelerated Particles

Particles travelling at speeds approaching the speed of light can cause ionization. The only particle used for food irradiation is the electron. An electron accelerator works on similar principles to a television tube (Figure 1). Electrons produced by heating a filament are injected into a vacuum chamber and attracted to a positive terminal. An electron beam is focused and then accelerated within the chamber by the voltage difference experienced by the electrons between the

start and finish of their travel in the chamber, a difference generated electrostatically or by a radio-frequency field. The beam emerges through a thin window at a speed close to that of light. It is swept from side to side by a changing magnetic field while irradiating large targets such as food.

An electron accelerated by a voltage difference of 1×10^6 V acquires an energy of 1 MeV (megaelectron volt). Energies above 1 MeV provide sufficient penetration in food to be useful. An upper limit of 10 MeV is recommended by the Codex general standard to avoid inducing radioactivity in the food.

Electromagnetic Radiation

Figure 2 shows that there is a range of EM radiations, each characterized by different wavelengths or energy and with different uses. Only X- and γ-rays have sufficient energy to cause ionization.

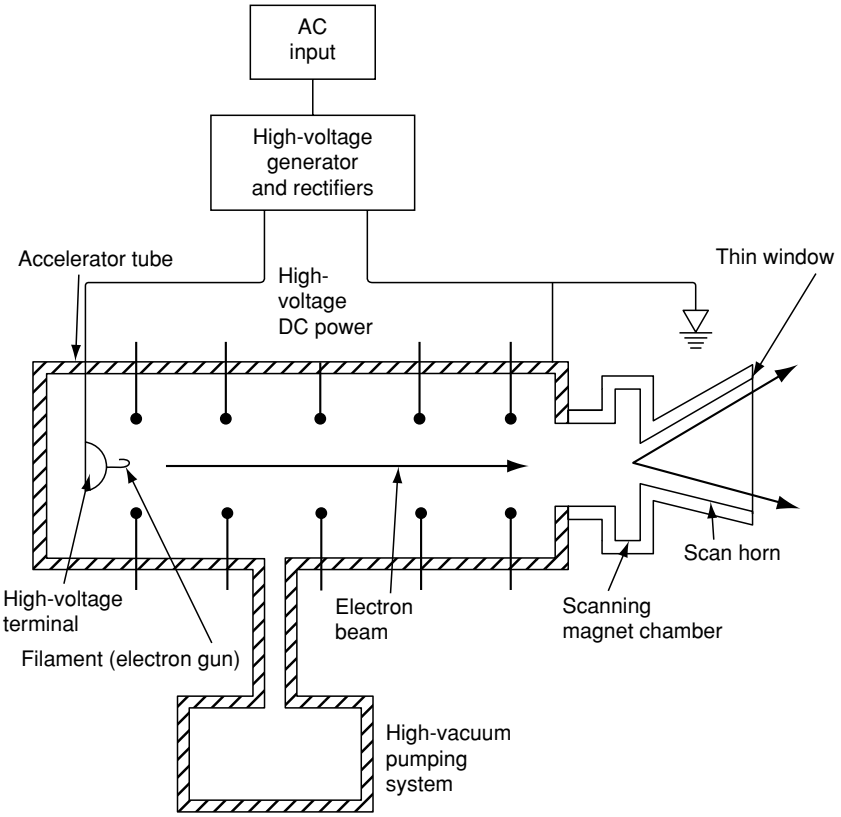

Figure 1 Schematic diagram of an electron accelerator with the voltage difference obtained electrostatically. Reproduced from Irradiation of Foods: Basic Principles. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

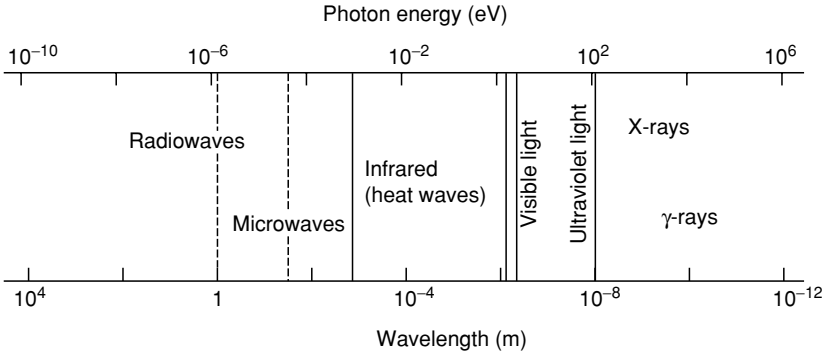

Figure 2 The electromagnetic spectrum. Reproduced from Irradiation of Foods: Basic Principles. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

X-rays are produced when high-speed particles strike a metal target and are rapidly decelerated. To avoid inducing radioactivity in food, X-ray sources must operate below 5 MeV. The efficiency of the conversion of 5 MeV electrons to X-rays is under 10%, making X-ray generators more costly to operate than simple electron machines.

γ-Rays are emitted by many atomic nuclei undergoing radioactive decay. Sources for food irradiation are limited to cobalt-60 and caesium-137, which emit γ-rays with energies around 1 MeV that cannot induce radioactivity.

Cobalt-60 is produced in a nuclear reactor by neutron bombardment of the stable element. It is a highly insoluble, high-melting-point metal. Caesium-137 is one of the fission products produced in the fuel rods of nuclear reactors from which it is chemically separated as a potentially volatile, soluble salt. Cobalt-60 dominates the γ-processing industry because of its more favorable production and environmental features.

Commercial decisions about the type and source of radiation to use are dictated by cost, beam penetration, and the form of the food to be treated. Most irradiation plants use cobalt-60. A few use electron beams and the first plants using X-rays are being constructed.

Dose and General Radiation Effects

Radiation effects in food are caused by the energy that is transferred to the food. The transferred energy is called the absorbed dose (commonly, just dose). Its unit is the gray (Gy) which is the absorption of one joule per kilogram. An older unit is the rad; 1 Gy equals 100 rad.

As the absorbed dose increases, the number of chemical reactions and the severity of the effects caused will increase. Food irradiation usually involves doses of 0.05–30 kGy, depending upon the desired result, although the upper limit recommended by the Codex Alimentarius is 10 kGy.

At low doses chemical reactions can disrupt biochemical and hormonal processes in fresh produce. As the dose increases, cell division is prevented as chemical disruption of DNA becomes severe. In this way microorganisms are inactivated and insects sterilized. At the highest doses used, all the microorganisms are killed and the food is sterilized.

Such effects indicate that extensive chemical changes are induced by irradiation. It may seem surprising that any food can withstand irradiation without being altered to an unacceptable extent. However, radiation deposits energy randomly throughout the food while inactivation of microorganisms, for example, requires that energy be deposited at a few critical points in a vital molecule such as DNA or a chromosome. Such molecules are relatively large targets and present in very few copies. In contrast, significant depletion or alteration of individual food constituents (e.g., amino acids) requires energy to be deposited in a large fraction of molecules that are far smaller and far more numerous than DNA. Frequently, therefore, food irradiation results in a product that is virtually indistinguishable from unprocessed food. However, as with conventional processes, some foods will suffer from undesirable changes in taste, smell, or texture.

Beneficial Uses of Irradiation

The major potential benefits of food irradiation are summarized in **Table 1**. Each benefit requires a minimum dose to be effective. For a particular food the maximum dose is limited by the onset of unacceptable changes in quality. Before a commercial decision is taken to irradiate a food, it is necessary to perform a cost-versus-benefit analysis and to assess the benefit to consumers against any potential loss in sensory qualities.

Table 1 Major technical benefits of food irradiation

Benefit	*Dose range (kGy)*	*Candidate foods*[a]
Inhibit sprouting	0.05–0.15	Potato, onion, garlic
Delay ripening	0.05–0.15	Some tropical fruit
Reduce parasites	0.1–0.3	Pork
Disinfest insects	0.1–1	Grain, rice, some fruit and vegetables
Delay spoilage (ambient)	0.5–5	Strawberry
Delay spoilage (refrigerated)	0.5–10	Meat, poultry, fish
Reduce pathogens	2–10	Meat, poultry, seafood, dried foods
Sterilization	10–30[b]	Spices, herbs, special diets

[a]Not a comprehensive list.
[b]Doses above 10 kGy exceed a Codex recommendation and are permitted only in a few countries for special purposes.

Safety Considerations

When considering whether any processed food is safe and wholesome the chemical, microbiological, and nutritional changes must be evaluated. Several expert international and national committees have considered the safety of irradiated foods. The World Health Organization (WHO) completed its latest reviews in 1994 and 1999. All the reviews reinforce the 1980 joint expert committee finding that, provided good manufacturing practice is followed, the irradiation of any food up to an overall average absorbed dose up to 10 kGy presents no toxicological hazard and no special microbiological or nutritional hazards.

The 1999 WHO report examined the safety of foods irradiated above 10 kGy. It concluded that foods treated above 10 kGy should also be considered safe and nutritionally adequate when produced under established good manufacturing practice.

Chemical Changes

Chemical reactions induced by a 10 kGy dose lead to about 300 mg of products per kilogram of food. Radiation, therefore, produces a large number of compounds but at very low individual concentrations. A few products are species that were not present significantly before irradiation. Some could conceivably be unique to irradiated foods and of unknown toxicity.

Any unique products must be present in extremely low concentrations since modern analytical methods can identify all the radiation products (or very similar compounds) in other processed or unprocessed foods. Several expert committees have considered a huge database of toxicological experiments. Although a small number of experiments initially raised some questions, the committees are satisfied that any doubts were due to inadequate experimental or analytical methods and that any toxicological risk is negligible.

Microbiological Changes

All physical processes have the potential to mutate microorganisms and lead to increased resistance, enhanced pathogenicity, or changed physiological traits important to their identification. If the process does not sterilize the food the surviving organisms will be those most resistant to the process. Subsequent regrowth can lead to microbial populations and risk different to those originally present.

However, irradiation appears little different to other physical processes in its potential for microbiological change. An exception may be its inability to destroy toxins produced by microorganisms that were present before processing, a useful property of heat for some toxins. In practice the risks are small. Following codes of good manufacturing practice developed for other processes and for storage will insure microbiologically safe irradiated foods.

Nutritional Changes

Irradiation up to 10 kGy does not alter significantly the nutritional value of proteins, carbohydrates, minerals, or saturated fats. Oxidation reactions can lead to the loss of essential unsaturated fatty acids. These reactions also induce rancid flavors and some foods with a high unsaturated fat content will not be suitable for irradiation. (*See* **Oxidation of Food Components**.)

In common with heating, freezing, dehydration, and storage, irradiation causes vitamin loss. Vitamin E and thiamin are the most radiation-sensitive, with vitamins A, C, and K also quite sensitive. Losses below 1 kGy are insignificant. Above 1 kGy the losses are comparable to those caused by other physical processes and may reach 10–20% for the more sensitive vitamins. If irradiated foods are to be further processed or cooked, the combined losses must be taken into account. (*See* **Ascorbic Acid**: Properties and Determination; **Retinol**: Properties and Determination; **Tocopherols**: Properties and Determination.)

The overall nutritional consequences of irradiating a particular food will depend upon:

- whether the food is a significant source of particular nutrients
- whether those nutrients are radiation-sensitive
- the dose
- the overall proportion of the food in the diet that is irradiated

Although there may be consumers with particular dietary habits or needs who could be adversely affected, the nutritional consequences of food irradiation will be insignificant for people eating a reasonably balanced diet.

See also: **Ascorbic Acid**: Properties and Determination; **Cheeses**: Manufacture of Hard and Semi-hard Varieties of Cheese; **Drying**: Theory of Air-drying; **Oxidation of Food Components**; **Pasteurization**: Principles; **Radioactivity in Food**; **Tocopherols**: Properties and Determination

Further Reading

Goresline HE (1982) Historical aspects of the radiation preservation of food. In: Josephson ES and Peterson MS (eds) *Preservation of Food by Ionizing Radiation*, pp. 1–46. Boca Raton, FL: CRC Press.

McLaughlin WL, Boyd AW, Chadwick KH, Mcdonald JC and Miller A (1989) *Dosimetry for Radiation Processing*. London, UK: Taylor and Francis.
Murray DR (1990) *Biology of Food Irradiation*. New York: Wiley.
Ramler WJ (1982) Machine sources. In: Josephson ES and Peterson MS (eds) *Preservation of Food by Ionizing Radiation*, pp. 165–188. Boca Raton, FL: CRC Press.
Satin M (1993) *Food Irradiation: A Guidebook*. Lancaster, USA: Technomic.
Thayer DW (1990) Food irradiation: benefits and concerns.*Journal of Food Quality* 13: 147–169.
World Health Organization (1988) *Food Irradiation: A Technique for Preserving and Improving the Safety of Food*. Geneva: World Health Organization.
World Health Organization (1994) *Safety and Nutritional Adequacy of Irradiated Food*. Geneva: World Health Organization.
World Health Organization (1999) *High Dose Irradiation: Wholesomeness of Food Irradiated above 10 kGy*.Report of a Joint FAO/IAEA/WHO Study Group. WHO Technical Report Series 890. Geneva: World Health Organization.

Applications

P B Roberts, Institute of Geological and Nuclear Sciences Ltd, Lower Hutt, New Zealand

Introduction

This article discusses for various foodstuffs the practical benefits of irradiation, including reduction of pathogens, insect disinfestation, extension of shelf-life, and delay of ripening or senescence.

General Principles

Food presented for irradiation should be of good quality and the process must not substitute for good manufacturing practice (GMP) at any stage in the production of food. When food is treated to reduce pathogen or insect numbers, appropriate packaging should be used to prevent recontamination.

The maximum dose recommended by the Codex Alimentarius, 10 kGy, causes a temperature rise of under 2.4 °C, resulting in an unchanged appearance and texture in many temperature-sensitive foods. Oxidative stress does occur and not all foods withstand the doses required to achieve the intended benefits without a loss of quality. The doses suggested here are guides only; the optimum dose varies with conditions and varieties of food. (*See* **Oxidation of Food Components**.)

Combining irradiation with another process may result in an overall improvement in quality. For example, fruit is less susceptible to softening when lower radiation doses are combined with a mild heat treatment, and insect disinfestation is still achieved. Combination treatments are not discussed further.

Major Applications

Applications of food irradiation fall into several major categories.

Improved Food Safety

Foodborne pathogens are a significant and increasing cause of ill health and economic loss. Irradiation has great potential to control pathogen numbers in at-risk foods. For example, it is being considered for use as part of Hazard Analysis Critical Control Point (HACCP) systems for control of pathogens in meat. Irradiation is a 'cold pasteurization' process suitable for solid foods, especially for foods of animal origin and consumed raw. Concern in some countries about the increase in the presence of *Escherichia coli* 0157 H7 in foods has added considerable impetus to the use of irradiation as a decontamination measure.

Improved Food Security

Irradiation can control insect numbers, inhibit sprouting of bulbs and tubers, and delay maturation, decay, and senescence of some fruits and vegetables. Therefore irradiation can help to combat postharvest food losses that may range from 10 to 50% for some crops, depending on climatic conditions and available storage technologies.

Quarantine Security

Most countries require that insect species not found within their borders do not 'hitchhike' on imported foods. The available chemical fumigants are being phased out because of health and environmental concerns. Irradiation is being accepted by plant protection authorities as a technically viable, broad-spectrum alternative to fumigation.

Reduction in Chemical Residues

Irradiation leaves no treatment residues. Consumer demand for foods free of chemical residues is therefore an opportunity for irradiation to supplant chemical fumigation and preservation methods.

Increased Trade

The four applications above provide increasing opportunity for increased international trade in foods.

Irradiation can assist in meeting the hygiene, quarantine, and quality demands of food-importing countries. This is likely to become economically important for developing countries seeking to expand food exports to markets in developed countries.

Raw Plant Products

Fresh Fruits

Insect disinfestation The doses required to guarantee quarantine security are far lower than the Codex-recommended maximum (Table 1) and many fruit suffer no loss of quality (Table 2) or vitamins. Fruit still require proper storage conditions and ozone produced during treatment should be vented to reduce damage to the fruit. (*See* **Insect Pests**: Insects and Related Pests.)

Irradiation up to 1 kGy guarantees that insects cannot reproduce. However, not all insect stages will be rapidly killed. US authorities have a regulation in place on the use of irradiation as a phytosanitary treatment of imported fruits and vegetables.

Extension of shelf-life

Reduction of spoilage Spoilage of fruit is usually caused by fungal or yeast infections. These organisms are inactivated by doses above 1.75 kGy, which approaches the maximum dose tolerated by most fruit. Above the tolerated dose, softening and texture loss occur because of degradation of pectin and cellulose. Spoilage may even be accelerated since infection of the weakened tissue becomes easier or natural antibiotic production is inhibited. (*See* **Spoilage**: Yeasts in Spoilage.)

Sweet cherries and apricots may be protected from fungal rots by a 4 kGy dose without loss of quality but the only fruit generally considered suitable for such treatment is the strawberry. Grey mold rot (*Botrytis cinerara*) in temperate climates and rhizopus rot (*Rhizopus stolonifer*) in subtropical areas can be controlled. Treatment with 2–2.5 kGy and refrigerated storage can increase strawberry shelf-life by a factor of 2–4. (*See* **Apricots; Cherries; Strawberries**.)

Delay in ripening Nonclimacteric fruit are picked ripe and then steadily deteriorate. Climacteric fruit, which are mature but unripe at harvest, undergo an initial decrease in respiration followed by a burst of respiration associated with ethylene production and the start of ripening. (*See* **Ripening of Fruit**.)

Irradiation can suppress ethylene production and the ripening process in preclimacteric fruit. The time between harvest and full ripening is an important marketing consideration and irradiation could induce a useful extension.

Doses above 1 kGy are required to delay ripening of apples and pears and result in fruit of poor quality. In peaches, nectarines, and apricots ripening is accelerated, not delayed. (*See* **Apples; Peaches and Nectarines; Pears**.)

Ripening can be delayed successfully in some subtropical fruits. A delay of a week (30 °C storage) or 10–12 days (20 °C) is observed in bananas after a dose of 0.4 kGy. Higher doses cause softening and discoloration. The fruit should be treated before full maturity – an exception to the general rule. Mangoes in the hard green state are delayed from ripening by about 7 days (25 °C storage) after 0.25–0.75 kGy. Higher doses cause skin darkening. Papayas have a

Table 1 Irradiation treatments to provide quarantine security against insect pests[a]

Insect species	*Stage*	*Minimum dose (kGy)*	*Criterion*
Bactrocera tryoni (Queensland fruit fly)	Eggs/larvae	0.075	Nonemergence of normal adults
Tephritidae	Eggs/larvae	0.15	Nonemergence of normal adults
Cydia pomonella (coding moth)	All	0.25	Nonemergence of normal adults
All other	All	0.3	Sterilization of any adults present, and nonemergence of normal adults from preadult stages

[a]Treatments and criteria extracted from the findings of an international task force meeting on irradiation as a quarantine treatment organized by the International Consultative Group on Food Irradiation (IAEA) in Changmai, Thailand, February 1986.

Table 2 The ability of fruits to retain quality after doses up to 1 kGy[a]

Quality retention	*Fruit*
Good	Apple, cherry, date, guava, mango, nectarine, papaya, peach, raspberry, strawberry, tamarillo, tomato
Reasonable	Apricot, banana, cherimoya, fig, grapefruit, lychee, orange, passionfruit, pear, pineapple, plum, tangelo, tangerine
Poor	Avocado, grape, lemon, lime, olive
Unknown	Pomegranate, kiwifruit

[a]Adapted from Kader AA (1986) Potential applications of ionising radiation in post-harvest handling fresh fruits and vegetables. *Food Technology* June: 117–121.

ripening delay of 3 days at ambient temperature induced by 0.5–0.75 kGy. Skin scald results from higher doses. (*See* **Bananas and Plantains; Mangoes; Papayas.**)

Delay of senescence All fruit, even if not susceptible to rots, eventually deteriorate as the carbohydrate polymers responsible for firmness and texture break down. Generally, irradiation accelerates breakdown but senescence is delayed in sweet cherries, apricots (3 kGy, 4 °C storage) and papayas (0.75 kGy, 25 °C storage).

Fresh Vegetables

Inhibition of sprouting in tubers and bulbs The saleable life of potatoes, onions, garlic, etc., is reduced by sprouting. Low doses (0.02–0.2 kGy), allied with cool, dry storage can inhibit sprouting for months. The mechanism is uncertain but the likely causes are interference with nucleic acid and protein synthesis, chromosomal aberrations in meristematic cells, and inhibition of the production of hormones that control dormancy and sprouting. At even lower doses (< 0.01 kGy) sprouting may be stimulated. Irradiation inhibits wound healing in potatoes which are therefore stored 3–4 weeks before processing. (*See* **Onions and Related Crops.**)

Texture and overall appearance are unaffected but other features can be affected adversely depending upon dose, storage conditions, and variety. A small brown discoloration near the stem end may occur in onions. A blue-gray discoloration that occurs on cooking potatoes that have been stored for several months may be increased by irradiation. Many potato varieties display a temporary rise in sugar content after irradiation. Some varieties are unsuitable for processing into French fries or potato chips since the sugar content increases permanently on storage.

Inhibition of greening Greening of potatoes on exposure to air is inhibited by doses that prevent sprouting. There is conflicting evidence on whether the production of the toxic alkaloid solanine which is associated with greening is also prevented. Greening of endive can be reduced by a dose of 2 kGy.

Other effects Undesirable changes during storage of mushrooms include opening and browning of the cap, darkening of the gills, and browning and elongation of the stem. Doses in the range of 0.05–0.5 kGy inhibit these changes with little effect on flavor. (*See* **Mushrooms and Truffles:** Use of Wild Mushrooms.)

Postharvest curvature of asparagus is inhibited by 0.15 kGy. Higher doses produce splitting of the ends and a slimy, darkened appearance. (*See* **Vegetables of Temperate Climates:** Stem and Other Vegetables.)

Cereal Grains

Dry cereals such as barley, wheat, rice, and maize are subject to depredation by insects (e.g., coleoptera, lepidoptera, and mites). Doses of 3–5 kGy are needed to kill the insects within 24 h. However 0.5 kGy reduces their ability to feed and renders them sterile. This is useful as the grains are usually stored for long periods. Grains should be stored cool and dry to avoid mold growth. (*See* **Cereals:** Handling of Grain for Storage.)

Processed Plant Products

Dried Fruits and Vegetables

Low moisture content acts as a sufficient preservative measure in dried produce. The major benefit of irradiation is to disinfest them of insects such as the saw-toothed grain beetle, the fig moth, and the Indian meal moth. Doses up to 1 kGy are required and generally cause no detectable sensory changes. Darkening may occur upon storage in some dried fruits.

An additional benefit of irradiation is that rehydration and swelling properties tend to be improved and cooking times reduced.

Dried Herbs, Spices, and Vegetable Seasonings

These useful food ingredients are usually dried and ground, chopped, or finely divided. They may be sold as such or used as an ingredient of other processed foods. (*See* **Herbs:** Herbs and Their Uses.)

The initial microbial population is often high (over $10^6\,g^{-1}$) and irradiation at the recommended Codex limit of 10 kGy is used as a decontamination measure. Some countries permit treatment at 30 kGy because these ingredients are of minor dietary importance.

The changes in aroma, color, and flavor caused by irradiation are insignificant at 10 kGy and slight at 30 kGy.

Fresh Meats

Red Meats

Inactivation of Trichinella spiralis The infectious cycle of the helminthic parasite *Trichinella spiralis* in pork can be broken by a dose of 0.3 kGy. No special handling procedures are required and meat quality is retained.

Extension of Shelf-life and Pathogen Reduction

Meat is sterile at death but bacterial contamination may occur during slaughter and processing. Adequate

Table 3 Sensitivity to ionizing radiation of some pathogenic bacteria when present in red meat[a]

Genus	D_{10} (kGy)[b]
Campylobacter	0.08–0.16
Escherichia	0.30–0.55
Listeria	0.20–1.10
Salmonella	0.31–1.30
Staphylococcus	0.34
Streptococcus	0.69–1.20
Yersinia	0.04–0.21

[a]Data adapted from Farkas J (1987) Decontamination, including parasite control, of dried, chilled and frozen foods by irradiation. *Acta Alimentaria* 16: 351–384, and as published by the US Council for Agricultural Science and Technology in Task Force Report no. 115, June 1989.
[b]The dose required to reduce viable cell numbers to 10% of the original numbers.

GMP should insure sufficient retail shelf-life, and vacuum packaging technology can extend the shelf-life of lamb and beef by several weeks. (*See* **Meat**: Preservation.)

If a further guarantee of product hygiene is required, irradiation may be used to reduce pathogen numbers. This will coincidentally extend the shelf-life by reducing spoilage organisms. The induction of off-flavors limits the maximum dose applied. Highly trained taste panels can detect flavor changes in some meats at quite low doses but the threshold dose for unacceptable changes in pork is 1.75 kGy. Doses in the range 1–2.5 kGy are acceptable for beef and lamb.

Such doses significantly reduce the population of many pathogens (Table 3). The meat will not be pathogen-free if the initial contamination is substantial or if more resistant pathogens are present. Spoilage is usually caused by Gram-negative bacteria of the family Enterobacteriaceae and genus *Pseudomonas*, which are radiation-sensitive. Irradiation shifts the population to Gram-positive *Lactobacillus* and *Achromobacter* (*Moraxella-Acinetobacter*) sp. with different spoilage characteristics.

Other processes that reduce shelf-life are 'drip,' an unsightly exudate, and oxidation which causes discoloration and flavor changes. Irradiation does not inhibit such processes.

Fresh Poultry

The principles of decontamination and shelf-life extension are similar for poultry and red meats. Poultry have generally been considered the more likely candidate for irradiation since poultry are a major vehicle for the spread of foodborne pathogens. (*See* **Poultry**: Chicken; Ducks and Geese.) However, concerns about *E. coli* 0157 H7 have put more emphasis on red meats.

Doses of 1–2.5 kGy significantly reduce pathogen numbers and double shelf-life. Below 2.5 kGy, sufficient spoilage organisms will survive to insure spoilage occurs prior to the possible, though unlikely, production of botulism toxin. Other safety measures include storage below 4 °C and oxygen-permeable packaging.

Frozen Red Meats and Poultry

Irradiation can provide extra protection against the presence of pathogens. Doses in the range 3–7 kGy are effective, require no special handling, and retain quality.

Processed Meats

Processed meats are often vacuum-packaged to insure a long shelf-life. However, they may be highly spiced and have a high surface/volume ratio if sliced. During long storage these factors favor pathogen growth which can be inhibited by 2.5–3 kGy.

Seafoods

Dried or Salted and Dried Fish

Irradiation at 0.5 kGy disinfects the stored fish of insects (*Dermestes maculatus, Necrobia rufipes* and members of the family Sarcophagidae) but will not inhibit spoilage due to mold growth. Sensory quality is unaffected.

Fresh Fin Fish

Fresh fish has a relatively short shelf-life due to microbial spoilage, enzyme action, and oxidation. Irradiation (1–2.5 kGy) reduces microbial spoilage, and the shelf-life can be extended safely by several days. Other spoilage mechanisms are unaffected and fish still lose quality during storage. Ideally the fish should be irradiated on board the fishing boat; alternatively, the fish should be put on ice before irradiation on shore. (*See* **Fish**: Spoilage of Seafood.)

Both marine and fresh-water fish can be treated successfully. Accelerated rancidity is a greater problem in fish with a high oil content (herring, salmon, tuna), though in some oily fish (mackerel) their strong flavor can mask taste changes. Storage below 4 °C is essential to prevent growth of *Clostridium botulinum* and toxin production. This can also be inhibited by oxygen-permeable wrapping which will, however, not inhibit oxidation. Below 2.5 kGy sufficient other microorganisms such as radiation-resistant *Achromobacter* spp. will survive and cause spoilage before toxin is produced.

Table 4 Other suggested benefits of the irradiation of foodstuffs

Foodstuff	*Proposed benefit*	*Likely dose range (kGy)*
Wheat flour	Increased loaf size from low-sugar formulations	0.5–1
Nuts	Insect control	1
Beans	Oligosaccharide reduction	2–3
Egg powder/dried whole egg	Decontamination	2–6
Frog legs (frozen)	Decontamination	2–7
Grapes	Increased juice yield	4–5
Fruit juices/purees	Stabilization	5–10
Cured meats	Nitrite reduction	7.5–30
Animal feed	Decontamination/sterilization	7–30
Complete meals	Sterilization (for immunosuppressed patients)	30–50

Shellfish

Shellfish are prone to heavy contamination by pathogenic and spoilage bacteria. Frozen shellfish can be decontaminated using 2–5 kGy with little quality loss. Fresh shellfish (peeled/unpeeled) can be treated with 1–2 kGy, doubling shelf-life, but quality decreases at higher doses. Refrigeration below 4 °C is essential. Viral contamination of shellfish cannot be reduced by these dose levels. (*See* **Shellfish:** Contamination and Spoilage of Molluscs and Crustaceans.)

Black spot occurs on shrimps when they are stored. Radiation inhibits its formation provided the shrimp is very fresh when irradiated.

Other Possible Uses of Irradiation

A number of other uses of irradiation have been implemented or proposed (Table 4).

See also: **Cereals**: Handling of Grain for Storage; ***Escherichia coli***: Occurrence and Epidemiology of Species other than *Escherichia coli* **Fish**: Spoilage of Seafood; **Legumes**: Legumes in the Diet; **Meat**: Preservation; **Oxidation of Food Components**; **Poultry**: Chicken; Ducks and Geese; **Ripening of Fruit**; **Shellfish**: Contamination and Spoilage of Molluscs and Crustaceans; **Spoilage**: Molds in Spoilage; Yeasts in Spoilage; **Vegetables of Temperate Climates**: Stem and Other Vegetables

Further Reading

Anonymous (1989) *Ionizing Energy in Food Processing and Pest Control. II. Applications*. A Task Force Report no. 115. Ames, Iowa: Council for Agricultural Science and Technology.

Anonymous (1992) *Irradiation of Spices, Herbs and Other Vegetable Seasonings; a Compilation of Data for its Regulation and Control*. IAEA-TECDOC-639. Vienna, Austria: International Atomic Energy Agency.

Anonymous (1993) *Irradiation of Poultry Meat and its Products; a Compilation of Data for its Regulation and Control*. IAEA-TECDOC-688. Vienna, Austria: International Atomic Energy Agency.

Anonymous (1997) *Irradiation of Bulbs and Tuber Crops; a Compilation of Data for its Regulation and Control*. IAEA-TECDOC-937. Vienna, Austria: International Atomic Energy Agency.

Elias PS and Cohen AJ (eds) (1983) *Recent Advances in Food Irradiation*. Amsterdam, The Netherlands: Elsevier.

Josephson EJ and Peterson MS (eds) (1982) *Preservation of Foods by Ionizing Radiation*, vol. III. Boca Raton, FL: CRC Press.

Shay BJ, Egan AF and Wills PA (1988) The use of irradiation for extending the storage life of fish and processed meats. *Food Technology in Australia* 40: 310–313.

Thomas P (1986) Radiation preservation of foods of plant origin. Part V. Temperate fruits: pome fruits, stone fruits and berries. *CRC Critical Reviews in Food Science and Nutrition* 24: 357–400.

Urbain WM (1986) Food irradiation. *Food Science and Technology: A Series of Monographs*. Orlando, FL: Academic Press.

World Health Organization (1994) *Safety and Nutritional Adequacy of Irradiated Food*. Geneva: World Health Organization.

Processing Technology

P B Roberts, Institute of Geological and Nuclear Sciences Ltd, Lower Hutt, New Zealand

Introduction

This chapter describes the facilities required to irradiate food, how to measure the absorbed dose, and the likely processing costs. Suitable packaging and methods to detect irradiated foods are also outlined.

General Description of Irradiation Plants

All irradiation facilities have the following components (**Figure 1**):

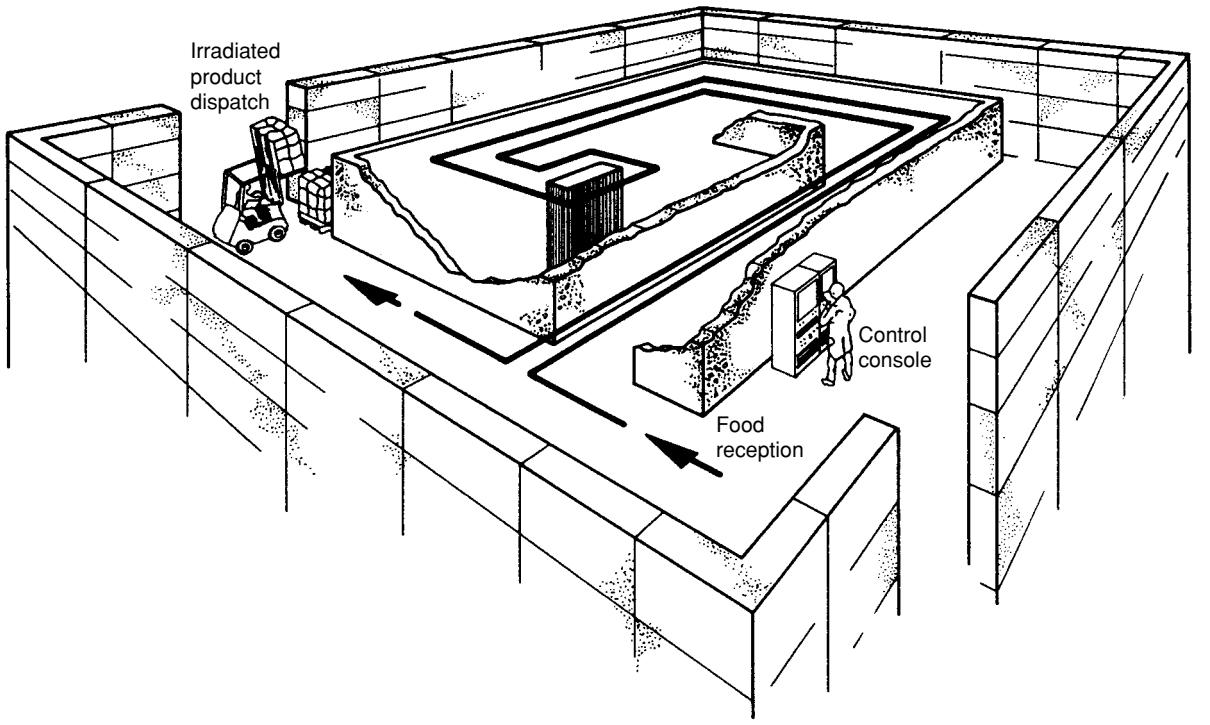

Figure 1 A diagrammatic representation of a cobalt-60 food irradiation plant. The arrows depict the movement of the food through the irradiation chamber and around the central source. Reproduced from Irradiation of Foods: Processing Technology. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

- a source of radiation in an irradiation chamber
- a protective shield
- a conveyor system
- a general area containing a control room or control console and warehouse space

Source

The commonly used sources of radiation are cobalt-60, a radioactive element emitting γ-rays, or an accelerator producing a beam of electrons.

Cobalt-60 is produced by neutron bombardment of stable cobalt in a nuclear reactor. Small nickel-plated slugs of the radioactive metal are loaded into a sealed alloy cylinder typically 10×450 mm and doubly encapsulated in a corrosion-resistant steel pencil. An array of such pencils is built into a rack typically 1–2 m^2. Cobalt-60 decays continuously. The time taken to lose 50% of its initial activity, i.e., its half-life, is 5.26 years. Usually 10% of the cobalt-60 is replenished annually.

There are several types of electron accelerator but all produce a fast-moving stream of electrons in a narrow beam that is rapidly scanned across a 0.2–1.2 m wide 'horn' positioned above the food to be irradiated. Radiation is produced only when the machine is switched on. The source strength can be varied within limits by altering the operating current or voltage.

Shielding

The irradiation chamber is enclosed by a 1.5–2 m thick concrete shield. This incorporates a maze system through which the food is conveyed to the source. The shield and maze reduce direct and reflected radiation levels in the general areas of the facility to safe levels and prevent exposure of the plant staff to radiation. In cobalt-60 plants penetrating γ-rays are emitted continuously in all directions. The radioactive source requires a nonoperating position in which all the radiation can be absorbed, making the chamber safe for staff to enter. This position is at the bottom of a 7 m deep pool of water set into the chamber floor. A hoist moves the source between operating and nonoperating positions.

Electrons (10 MeV) are absorbed within a few centimeters in condensed matter. However, penetrating X-rays are produced when electrons strike objects within the chamber. Therefore, shielding is just as necessary in accelerator facilities as in γ-plants, although its design may be simpler. The accelerator is simply switched off when staff enter the chamber.

Conveyor

Some γ-plants operate in a batch mode. Food is carried into the chamber with the source nonoperational, then irradiated and removed. However, a

conveyor system carrying food continuously past the source is more usual. Conveyor design must insure that the food is irradiated with reasonable uniformity. The speed of the conveyor and, hence, the time spent near the source is a major determinant of the dose received.

A railed system is used to transport large containers (up to $1 \times 1.2 \times 1.8$ m) past cobalt-60 sources in a single- or multiple-pass arrangement. The latter provides more efficient absorption of radiation and more uniform dose distribution.

An electron beam is emitted in one basic direction. This and its limited penetration permit only thin layers of food to be treated under the beam. A simple moving belt acts as the conveyor. Electron beams have been used mainly to treat grains, seeds, and homogenous meat products up to 70 mm thick.

General Area

Controls and servicing equipment are housed outside the shielded area. Areas for receiving incoming food and storing irradiated product are kept separate and contain temperature-controlled rooms if necessary.

Beam Penetration

Figure 2 illustrates the ability of γ-rays and electrons to penetrate food packages. Two-sided irradiation doubles the thickness treatable and reduces the dose variations within the package.

A completely uniform dose throughout a package is not feasible. The ratio of maximum to minimum dose is called the dose uniformity (DU) and should be as close to unity as possible. This is achieved more easily by treating thinner packages, a strategy that wastes radiation energy and reduces the volume treated in a given time. DU values close to unity in bulky packages are achieved more easily with γ-rays than with electrons (Figure 2).

Acceptable DU values differ for different foods and processes. The minimum dose must guarantee the intended effect and the maximum must not be detrimental to the food. Initial plant design must balance the acceptable DU, the throughput required for economic operation, and the range of doses required.

Dosimetry

The energy absorbed by the food (i.e., the dose) is the most important parameter in the treatment process.

Standards and Reference Dosimeters

The dose can be measured absolutely by calorimetry or ionization chamber methods. Calorimetry measures the heat released into water; ionization chambers record the number of ion pairs produced in air. Specialized laboratories have used both methods to provide national standards and reference dosimeters.

However, any effect that varies detectably and reproducibly with dose can be used as a reference dosimeter provided its response can be calibrated against a national standard. The Fricke dosimeter is a particularly well-characterized and common reference dosimeter. It consists of an aqueous solution of ferrous sulfate and sulfuric acid. Free radicals produced during irradiation oxidize ferrous ions to ferric ions, a change detected as an increase in optical absorption. Doses in the range 10–400 Gy can be measured by Fricke dosimetry. A small modification to the solution extends the range to 2 kGy.

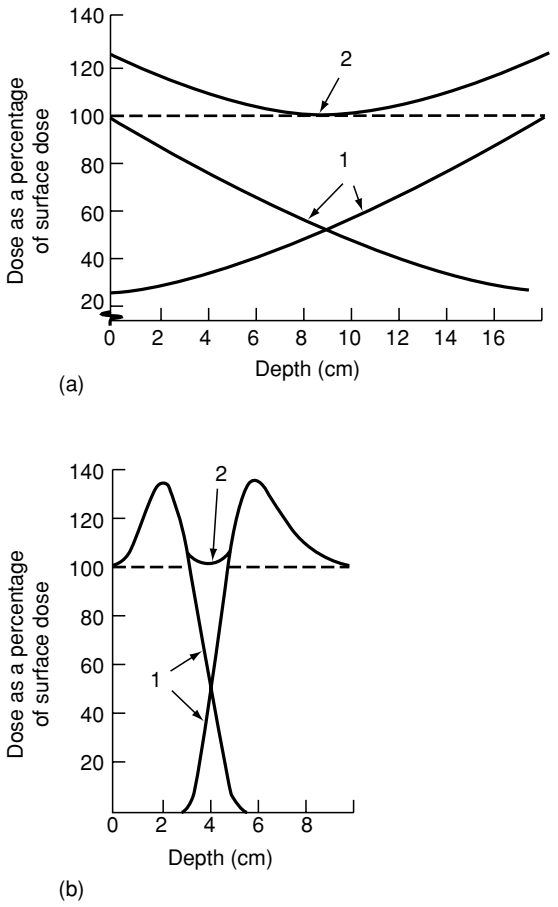

Figure 2 The relationship between dose and depth for unit density packages exposed to (a) cobalt-60 γ-rays or (b) 10 MeV electrons either from one side (1) or two sides (2). Adapted from Cleland MR, Pageau GM (1985) Electron versus gamma rays – alternative sources for irradiation processes. *Food Irradiation Processing. Proceedings of a Symposium*, Washington DC, March, pp. 397–406. Vienna, Austria: International Atomic Energy Agency, with permission.

In-plant Dosimetry

The reference dosimeters are difficult to use routinely in commercial operations. More practical dosimeters comprise plastic materials such as polymethylmethacrylate. Small pellets of the plastic, either clear or colored by dye, can be easily handled and change color over a wide dose range. The color induced gradually fades. Such dosimeters must be calibrated against a reference dosimeter that has itself been calibrated against a national standard.

When any food is presented initially for irradiation, potential dose variations throughout the food must be assessed in trial runs. Plastic dosimeters are placed strategically throughout the food container to map the dose distribution. The positions of maximum and minimum dose (D_{max}, D_{min}) are determined and checked against the prescribed treatment protocol.

During routine operations some packages that contain dosimeters at the points receiving D_{max} and D_{min} are conveyed past the source. Most operators provide a label that is impregnated with a radiation-sensitive dye on the outside of every package. The label changes color upon radiation. It is a simple yes/no indication that the package has been treated rather than a true dosimeter.

Overall Average Absorbed Dose

Technically, the important dose parameters are D_{max} and D_{min} since these will determine the quality and effectiveness of the treatment. However, the Codex General Standard for Irradiated Food refers to an overall average value, D_{av}. This is rigorously defined as:

$$D'_{av} = \frac{1}{M}\int p(x,y,x)d(x,y,z)dV$$

where M is the total mass of product, p is the density at a point defined by the coordinates (x, y, z), d is the absorbed dose at (x, y, z) and dV is the infinitesimal volume element. D_{av} can only be determined in an extensive dose-mapping exercise. However, the simple equation:

$$D_{av} = \frac{D_{max} + D_{min}}{2}$$

is often sufficient.

The Code General Standard for Irradiated Foods recommends a maximum dose in terms of a D_{av} value of 10 kGy. At least 97.5% of the food should receive less than 15 kGy. The dose uniformity should also be optimized. This is usually interpreted as DU values less than 2, but higher values are acceptable for some applications using low doses.

Plant Capacity and Costs

If all the radiation emitted by the source is absorbed in the food then a source of 1 kW can treat 360 kg of food per hour to a dose of 10 kGy. However, radiation is wasted when some of the beam misses or passes completely through the food. Some radiation is absorbed within the source or in the conveyor system. The percentage of radiation actually used, the efficiency factor (E), is usually 20–50%. Simple cobalt plants have relatively low-efficiency factors. Large multiple-pass cobalt plants and electron beam facilities can be designed with an E value approaching 50%.

A useful equation for the production rate of a plant is:

$$T = \frac{3600S}{D}E$$

where T is the throughput (kg h^{-1}), S is the source power (kW), D is the minimum dose (kGy), and E is the efficiency factor.

The source power is given by the cobalt-60 activity (1 kW = 67 kCi or 2.5 PBq) or the accelerator voltage and current (1 kW = 1 MeV × 1 mA).

Plant costs depend on the processing capability (volume and dose) required and, hence, on the source power and warehouse space. Costs can be in the range US$1–6 million. For a typical 1 MCi (37 PBq) cobalt-60 plant the source, the conveyor system, and the building each account for about one-third of the total cost.

Processing costs depend upon whether a high or low dose is required and on the source power. If sufficient volumes of food are available to dedicate the plant to a paticular purpose (e.g., poultry decontamination), then the plant can be designed to maximize efficiency. Most plants treat a variety of foods for different purposes and require greater flexibility at added cost.

In general, processing costs increase in proportion to the dose required and are approximately US$10–100 per tonne over a range of low- to high-dose applications.

Food Handling

Irradiation resembles pasteurization and fumigation in its ability to reduce the bioburden of pathogenic organisms in food. However, such methods must not be used to substitute for good hygiene practices at any stage in the pre- or posttreatment handling of food. Ideally, irradiation should be integrated within a Hazard Analysis Critical Control Point (HACCP) system. Codex has prepared guidelines for the use of HACCP systems.

Overall standards of food handling are discussed in the Codex General Principles of Food Hygiene. Operating standards within an irradiation plant are covered in a Codex Recommended International Code of Practice for the Operation of Irradiation Facilities Used for the Treatment of Food. (*See* **Fumigants; Pasteurization**: Principles.)

Irradiation can penetrate foods prepackaged in plastic or simple insect-proof wrapping. Food such as poultry, fruit, or vegetables can be decontaminated or disinfested in their final package, ready for transport, storage, or display. This eliminates the risk of recontamination until the package is opened by the purchaser.

For some irradiation treatments (e.g., inhibition of sprouting), foods such as potatoes can be treated loose in suitable totes or containers. If foods for which temperature control is an essential part of good storage technique are irradiated, then the plant must have suitable areas set aside to control the temperature. Special care may be needed with seafoods and some meats in which resistant *Clostridium botulinum* spores can survive irradiation. Subsequent growth of the bacterium would produce a fatal toxin, but this is prevented by storage below 4 °C. (*See* ***Clostridium***: Occurrence of *Clostridium botulinum*.)

Irradiation while food is frozen or vacuum-packed inhibits the free radical reactions leading to oxidation rancidity and off-flavors. Treatment while frozen and under vacuum was originally proposed for foods intended for very-high-dose treatments (50–60 kGy) to make them shelf-stable indefinitely at room temperature. The product is an attractive alternative to canned food. Since Codex presently recommends that dose is limited to 10 kGy, the application of food

Table 1 Approvals of packaging materials for use in conjunction with irradiation of food[a]

Packaging material	*Maximum dose (kGy)*	*Country (approval date)*[b]
Cardboard	10; 35	UK; Poland (1991)
Polyethylene coextruded polyvinylacetate	30	US; Canada (1988)
Polyethylene-co-vinylacetate	30	US (1989)
Fiberboard	10	India (1997)
Fiberboard, wax-coated (boxes)	10	US; Canada (1989)
Glassine paper	10	US (1975)
Glass	10	India (1997)
Hessian sacks	10	UK (1991)
Kraft paper	0.5	US (1975)
Nitrocellulose-coated cellophane	10	US: India (1975)
Nylon 11	10	US; India (1975)
Nylon 6	60; 10	US; India (1975)
Paper	10; 35	UK; Poland (1991)
Paper coated or laminated with wax or polyethylene	10; 35	India; Poland (1990)
Paper laminated with aluminum foil	35	Poland (1990)
Polyamide film or polyamide coextruded with polyethylene	35	Poland (1990)
Polyester-metalized-polyethylene laminate	35	Poland (1990)
Polyester-polyethylene laminate	35	Poland (1990)
Polyethylene film (various densities)	60; 35; 10	US; Poland; India (1975)
Polyethylene-paper-aluminum laminate	35	Poland (1990)
Polyethylene-terephthalate	60	US (1975)
Polyolefin (low-destiny as middle or sealant layer)		Canada (1989)
Polyolefin (high-density as external layer)		Canada (1989)
Polyolefin film	10	US (1975)
Polypropylene sacks	10; 35	UK; Poland (1990)
Polypropylene – metalized	35	Poland (1990)
Polystyrene film	10	US; India (1975)
Polystyrene foam trays (Styron 685D)	10	Canada; India (1989)
Rubber hydrochloride film	10	US; India (1975)
Steel, tin plates or enamel-lined	10	India (1997)
Vegetable parchment	60; 10	US; India (1975)
Vinylchloride-co-vinylacetate film	60; 10	US; India (1975)
Vinylidenechloride-coated cellophane	10	US (1975)
Vinylchloride-co-vinylidenechloride film	10	US; India (1975)
Wood	35; 10	Poland; India (1990)
Viscosa	35	Poland (1990)

[a]Data available through the International Consultative Group on Food Irradiation, International Atomic Energy Agency, Vienna.
[b]The earliest approval date is given.

Table 2 Major methods of detecting irradiated foods

Method	*Basis*	*Suitable foods*	*Status*
Electron spin resonance	Free radicals, paramagnetic centers in rigid matrices (e.g., bone, shell, seeds)	Meat (bone in), shellfish, some fruit	Established
Thermoluminescence[a]	Centers of energy excess or deficit in dry material	Herbs, spices, shellfish	Established
Gas chromatography–mass spectrometry	Radiation products from volatile fatty acids	Meat, lipid-containing foods	Advanced, standardization required
Impedance	Change in electrical properties	Potato	Advanced, wider testing required
Microbiology	Change in microbial population	Most uncooked foods	Well developed, indicative test only, not definitive
DNA alterations	Damage to DNA bases and strand breaks	Most uncooked foods	Experimental

[a]The signal arises mainly from adherent minerals (soil, dust). Mineral separation leads to better discrimination and wider applications. A related method using light rather than heat to release stored energy is very promising but suitable equipment is lacking in most laboratories.

irradiation to produce sterile shelf-stable foods in light plastic packaging has received little attention. However the conclusion of a World Health Organization Study Group in 1999 that foods irradiated to any practical dose above 10 kGy are safe and nutritious may lead to more interest in this application.

Packaging

Packaging must meet standards for the retention of postirradiation stability, strength, and permeability. The risk of low-molecular-weight chemicals in the irradiated packaging (e.g., stabilizers) diffusing into the food must also be negligible.

Table 1 lists the packaging materials approved by the US Food and Drug Administration and by regulatory authorities in Canada, Poland, India, and the UK for concurrent use when food is irradiated.

Labeling and Detection of Irradiated Foods

Irradiation is one of a group of food processes that leave no obvious signs to a purchaser, food inspector, or retailer that a food has been processed. Labeling of irradiated foods is generally considered desirable. In its General Standard for Labelling of Pre-packaged Food, Codex recommends use of the phrase 'irradiated [name of food]' or 'treated by ionizing radiation.' Most countries legislate for this wording or a near equivalent to be used with complete foods that have been irradiated. The words are placed on the food package or are prominently displayed near foods sold in bulk containers.

There is considerable debate about the labeling of irradiated minor ingredients or components of a food. The practicality, cost, and need for such labeling are questioned. National regulations vary in their requirements.

Methods to detect irradiated foods would boost consumer confidence in legislative controls, particularly that labeling information was accurate. Unambiguous detection of an irradiated food proved to be a difficult problem. No single method for all foods seems feasible. However, international collaborative trials have established a number of detection methods. Table 2 summarizes the methods that are established or under further investigation. Germany and the UK have official protocols in place for the use of some methods to enforce labeling requirements. Postprocessing detection is not a suitable way to control actual processing conditions.

See also: ***Clostridium***: Occurrence of *Clostridium botulinum*; **Fumigants**; **Hazard Analysis Critical Control Point**; **Legislation**: Codex; **Pasteurization**: Principles; **Plant Design**: Process Control and Automation

Further Reading

Anonymous (1991) *Code of Good Irradiation Practice for Insect Disinfestation of Cereal Grains. International Consultative Group on Food Irradiation Document no. 3* and related codes in this series. Vienna, Austria: International Atomic Energy Agency.

Beers EW (1990) Innovations in irradiator design. *Radiation Physics and Chemistry* 35: 539–546.

Cleland MR and Pageau GM (1985) Electrons versus gamma rays – alternative sources for irradiation processes. *Food Irradiation Processing. Proceedings of a Symposium, Washington DC, March*, pp. 397–406. Vienna, Austria: International Atomic Energy Agency.

Food and Agriculture Organization (1984) *Codex Recommended International Code of Practice for the Operation of Radiation Facilities Used for the Irradiation Treatment of Foods* (CAC/RCP 19–1979 Rev 1) *Codex Alimentarius*, vol XV. Rome, Italy: Food and Agriculture Organization of the United Nations.

Food and Agriculture Organization (1991) *Codex General Standard for the Labelling of Prepackaged Foods*

(Codex STAN 1–1985, Rev. 1–1991) *Codex Alimentarius*. Rome, Italy: Food and Agriculture Organization of the United Nations.
McKeown J and Sherman NK (1985) Linac based irradiators. *Radiation Physics and Chemistry* 25: 103–109.
McLaughlin WL, Boyd AW, Chadwick KH, McDonald JC and Miller A (1989) *Dosimetry for Radiation Processing*. London, UK: Taylor and Francis.
McMurray CH, Stewart EM, Gray R and Pearce J (1996) *Detection Methods for Irradiated Foods: Current Status*. London, UK: Royal Society of Chemistry.
Morrison RM (1990) Economics of food irradiation: comparison between electron accelerators and cobalt-60. *Radiation Physics and Chemistry* 35: 673–680.
Swallow AJ (ed.) (1989) Food irradiation. *Radiation Physics and Chemistry* 34: 875–1030.
Thayer DW (1988) Chemical changes in food packaging resulting from ionising irradiation. In: Hotchkiss JH (ed.) *Food and Packaging Interactions*. ACS Symposium Series, no. 365. Washington, DC: American Chemical Society.
Thompson CC and Cleland MR (1989) High power dynamitron accelerators for X-ray processing. *Nuclear Instruments and Methods* B40/41: 1137–1141.

Legal and Consumer Aspects

P B Roberts, Institute of Geological and Nuclear Sciences Ltd, Lower Hutt, New Zealand

Background

Governments have a responsibility to ensure the safe use of facilities that involve radiation sources, including those used to process food. An increase in the present limited commercial use of food irradiation will depend on the regulatory environment at national and global level, and on the attitude of the food industry and consumers to the product. This article discusses these issues.

Legislative Principles

Governments control the processing of food by various legislative methods. Some governments approve a particular treatment, whereas others may license a facility. In some countries, a single agent or minister is responsible. In other countries, multiple agencies must approve different aspects of the process and the sale of products.

Although regulatory mechanisms differ, there are a few basic principles that authorities should adopt to ensure the safe irradiation processing of food. These are summarized within a number of international documents, in particular:

- Codex General Standard for Irradiated Foods;
- Codex Recommended International Code of Practice for the Operation of Irradiation Facilities for the Treatment of Food;
- Codex General Standard for the Labelling of Prepackaged Food;
- Codex General Principles of Food Hygiene, and related Codes of Practice for specific commodities;
- Codex Hazard Analysis and Critical Control Point System and Guidelines for its Application.

Radiation Safety

Staff within the plant, as well as the public, must be protected against the radiation emitted by the source. Most countries already have effective controls in place to deal with nonfood uses of irradiation facilities. Authorities refer to international guidelines of the International Atomic Energy Agency for the following:

- permissible exposures of the workforce and public;
- safe handling, transportation and disposal of radioactive material;
- safe operation of high-energy accelerators and sources containing large amounts of radioactivity activity;
- emergency procedures in the event of accidents.

Radioactivity cannot be induced in food unless it is irradiated with radiation above a threshold energy. Therefore, the types and maximum energy of the radiation used to process food are restricted to the following:

- γ-rays emitted by the radionuclides cobalt-60 and cesium-137;
- X-rays generated by machine sources operated at or below 5 MeV;
- electrons generated by machine sources operated at or below 10 MeV. (*See* **Radioactivity in Food**.)

Food Hygiene

The irradiation facility must meet the design, construction, and general operating criteria imposed on all well-regulated food handling plants. Authorities must also take into account the packaging and temperature requirements of the food while in the treatment plant.

Irradiation must not be used to relax current standards of hygiene during any stage in the production, storage, transportation, processing, or sale of foodstuffs. The food presented for irradiation must meet all normal standards of microbiological and general

quality, and must have been produced under conditions of good manufacturing practice for which Codex standards and other guidelines are available. In this way, irradiation will confer further benefit or enhanced safety on high-quality food.

Prior Approval and Quality Control

There is no routine method of checking whether or not food has been irradiated, or of estimating the dose received once the food has left the facility. Therefore, the management and operation of the plant must be adequate at all times to guarantee proper control of the process. This is achieved by requiring prior approval of the irradiation of foodstuffs and of the facility. The approval process considers the following factors:

- staffing (numbers, training, skills, supervision);
- dosimetry methods and standards;
- ability to assess initial, especially microbiological, quality;
- separation of incoming and treated food;
- measures to prevent accidental reirradiation;
- hygienic storage and handling techniques and conditions;
- record-keeping and documentation;
- regular inspection by authorities.

The overall goal of the approval process should be to meet the principles of the *Codex Recommended International Code of Practice for the Operation of Radiation Facilities used for the Treatment of Foods*. (*See* **Quality Assurance and Quality Control**.)

International Trade

Foods are subject to documentary requirements when traded internationally. Importing countries often require assurance that the regulatory environment in which the processors operate is at least as stringent as their own. Minimum information on documentation should be as follows:

- an identification number for the batch of food;
- the treatment given (dose, purpose, place, and date of processing);
- the national regulations under which treatment was approved;
- any special handling requirements such as temperature control.

Laboratory methods are becoming available to prove that foods have been irradiated, although it remains a major problem to estimate absorbed doses once the product has left the irradiation plant. Therefore, it remains necessary for countries to ensure that irradiated foods are properly tracked as they move in trade, and that documentation accompanies each shipment. This will require effective cooperation between countries.

Labeling

A label is the only way by which consumers can be informed about whether a food has been irradiated. There is no obligation to inform consumers about some competing technologies (e.g., chemical sprout inhibition, fumigation). However, commercial use of irradiation has coincided with a growing recognition of the consumer's right to be informed and make a considered choice about which foods to purchase.

A secondary reason for use of labeling is the prevention of fraud (i.e., false claims in regard to whether foods have or have not been irradiated).

The Codex General Standard for Pre-packaged Food recommends the use of terms such as 'irradiated (name of food)' or 'treated with ionizing radiation' for English-speaking countries. In France, the term 'ionization' is favored. If a food is sold in a prepackaged form, the label will be on the package. Foods sold loose will have the information provided close to the food container.

Figure 1 illustrates the international symbol or logo for irradiated foods. It would be a simple visual guide to consumers and the international food trade that a food or ingredient had been irradiated. However, its use as the sole identifier of irradiated foods has not been supported by consumer organisations. A logo may be considered as a voluntary addition to the mandatory text.

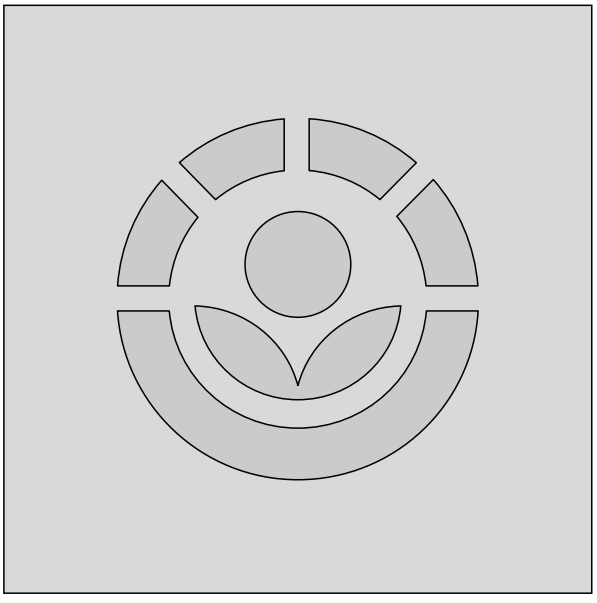

Figure 1 The international logo for irradiated foods.

Role of International Organizations

International organizations such as the Food and Agriculture Organization (FAO), the World Health Organization (WHO), and the International Atomic Energy Agency (IAEA) support the irradiation processing of food as a contribution to safe and secure food supplies. In 1984, these organizations agreed to sponsor the International Consultative Group on Food Irradiation (ICGFI) to act as an international group of experts designated by governments to evaluate and advise on global activities in food irradiation.

A priority for ICGFI is to encourage trade in irradiated foods by working towards regulatory requirements that are as consistent as possible in different countries. The ICGFI has issued a number of documents that advise on the control of food irradiation processing. These include:

- Guidelines for Preparing Regulations for the Control of Food Irradiation Facilities;
- International Inventory of Authorized Food Irradiation Facilities;
- Guidelines for the Authorization of Food Irradiation Generally or by Food Classes;
- Codes of Good Irradiation Practice for individual commodities (amalgamated into a single Code in 2000) and associated monographs of technical data.

The Codex Alimentarius Commission has also issued important documents related to regulating food irradiation, as discussed elsewhere in this series of articles.

International Status of Food Irradiation (1999)

Forty-one countries have informed the Joint Division of the FAO/IAEA that they have listed at least one type of food as cleared for irradiation for one or more purposes. Although the Codex General Standard declares that *any* food can be irradiated up to an overall average dose of 10 kGy, most countries regulate irradiation using the principles of food additives and still require each food or application to be approved on a case-by-case basis.

Some countries have specific regulations on food irradiation, and the list of approved items is part of the regulation. In other countries, the list of approved items has been drawn up by a statutory agency or special committee. There is no specific food irradiation regulation, but an authorization can be given within general food regulations.

A few governments (e.g., Luxembourg, Sweden) have a policy that precludes the irradiation of food or the sale of irradiated food products, though this is not always reflected directly in legislation. Others have a regulation that provides for an overall prohibition on irradiation of foods but with the possibility to seek an exemption for specified foods and purposes (e.g., Australia, New Zealand).

The listing of countries and approved foods appears extensive. However, the existence of an approval does not necessarily mean that there is automatic authorization to proceed with commercial processing. Usually, it is still necessary to make a specific application to process food, to provide justification, and to show that good manufacturing practice will be carried out.

Over 40 different foods or food classes have been approved for irradiation for a variety of purposes, and about 50 irradiation facilities worldwide treat food. However, most of these facilities treat only very small volumes of food.

Two aspects of national regulations, in particular, can vary greatly between countries. One is labeling, especially the labeling of food ingredients. The Codex General Standard on Labeling of Pre-packaged Foods is being interpreted in different ways. Some countries require no labeling of irradiated foodstuffs once the food becomes an ingredient in another food; others require labeling regardless of how small a proportion that an irradiated ingredient constitutes of the complete food; other countries make no reference to labeling.

Another cause of nonuniformity concerns the allowable doses of radiation applied to the food. The Codex Standard allows irradiation of food up to a maximum overall average dose of 10 kGy. This maximum is based on very conservative safety and nutritional considerations. In practice, it is also necessary for processors to limit the dose range applied to foods for technological reasons. These reasons include ensuring that the intended benefit is achieved, for which a minimum dose is necessary, and retaining the quality of the food, for which a maximum below 10 kGy may be necessary. These technological limits, which are basically advisory and the business of the radiation processor, are often included in national regulations in different ways in different countries.

In 1999, the WHO published the findings of a Study Group on the wholesomeness of food irradiated above 10 kGy. It concluded that foods treated above 10 kGy were safe and nutritionally adequate when produced under conditions of good manufacturing practice. Thus, on the grounds of safety, there is no need to legislate a maximum safe dose for irradiation of food. The maximum can be determined by the need to retain the quality and sensory properties of food acceptable to consumers.

During 2000, the Codex Alimentarius Commission will review the implications of the Study Group finding for its General Standard on Irradiated Foods, particularly the recommended maximum dose limit. This may then influence the way in which national regulations deal with dose limits.

Two exceptions to the 10-kGy maximum are already recognized in some countries. Higher doses may be permitted to sterilize minor food ingredients such as herbs and spices or foods for special groups such as hospital patients, astronauts, or the armed forces.

Consumer Issues

Food irradiation has remained a specialized process used for small volumes of food. In developed Western countries, this is mainly because of reluctance on the part of retailers and the food industry to adopt the process. This industry position has arisen from initial doubts or a lack of understanding about the process by consumers. Without widespread acceptance of the process in US and European markets, developing countries have little incentive to invest in irradiation facilities.

Formation of Consumer Opinion

Resistance to irradiated foods was initially based on a misapprehension that the food became radioactive. Among the public and the media, there persists an inability to distinguish between radiation and radioactivity and between eating an irradiated product and being irradiated. Accidents such as Three Mile Island and Chernobyl reinforced these misapprehensions. However, if such confusion were the sole reason for consumer resistance to irradiated foods, it would be relatively easy to counter.

There are a number of general issues related directly or indirectly to food irradiation:

- Interest in natural, unprocessed foods has grown. Associated with this has been a perception of the harmful health consequences of additives and processing. Doubts about the safety of irradiated foods continue to be expressed in spite of expert opinion.
- Consumers' rights to choose and to be informed about their food are widely accepted. The possible additional costs and requirements for accurate and informative labeling reinforce the reluctance of food retailers to stock irradiated foods.
- There remains a general opposition to any new technology that uses radioactive material or radiation.

This mix of issues has ensured that food irradiation has maintained a profile as a controversial process in the media and among consumer activist groups that is hardly warranted by its present importance as a food technology.

Consumer Attitudes

Consumer attitudes vary greatly by country and are dependent on development status, the extent to which irradiated foods are available and media exposure. Surveys of consumer attitudes (e.g., Table 1) have shown that most members of the public are unaware or have little knowledge of food irradiation. When surveyed for an initial reaction, most consumers are either unwilling to purchase irradiated foods or express strong reservations. Individual opposition may decrease when the process is explained more fully. General opposition tends to increase when conflicting pro- and antiirradiation views are publicized.

Consumer opinion has been studied most closely in the USA, where there has also been a substantial public debate on the issue to help inform the consumer. As information has been made available, consumer concern has decreased. In the late 1980s, over 40% of consumers surveyed expressed serious concern about irradiation. By 1996, this had fallen to less than 30%.

Several studies in the USA indicate that up to 20% of consumers oppose the irradiation of foods and are unlikely to change that view. A similar percentage is willing to try irradiated foods on the basis of minimal information. The remainder are open to persuasion in either direction on the basis of the information placed before them.

Some surveys have found that concern about irradiation ranks far lower than other concerns related to food safety (Table 2). Irradiation was classed as a serious concern by 29% of US consumers surveyed, compared with 77 and 66% seriously concerned about bacteria and pesticides, respectively. Nitrites, additives, and other preservatives elicited slightly less concern than irradiation. In the longer term, the fact that

Table 1 Concerns and awareness about food irradiation in the UK

	1986 (%)	*1989 (%)*
Very concerned	25	36
Fairly concerned	44	42
Neither concerned nor unconcerned	8	7
Not very concerned	13	8
Not at all concerned	8	6
Don't know/not stated	1	1
Previously aware of the term 'food irradiation'	48	55

The survey base was 2000 adults.
Data from Consumers' Association (1990) *Food Irradiation: The Consumer's View*. London: Consumer's Association.

Table 2 Consumer perceptions in the USA of potential health hazards from food

	Serious concern	*Some concern*	*Slight concern*	*No hazard*	*Unsure*
Contaminated by bacteria	77	17	5	1	1
Residues such as pesticides	66	23	7	2	2
Antibiotics/ hormones	42	35	13	4	6
Food handling in supermarkets	41	39	16	3	1
Irradiation	29	26	11	8	27
Nitrites	24	41	13	5	17
Additives/ preservatives	20	50	21	7	2

Adapted from *Opinion Research, Trends, Consumer Attitudes and the Supermarket* (1996). Washington, DC: Food Marketing Institute.

irradiation reduces the risk from such high level concerns as bacteria and chemical residues may become more important than fears about irradiation itself.

Another factor shaping public opinion has been increasing opposition among communities to the siting of irradiation plants within their area. This opposition is usually based on exaggerated fears of the transport of radioactive material to the plant or of possible accidents involving environmental contamination.

Market Trials

Between 1984 and 1990, about 40 market trials of 17 irradiated foods were conducted in 13 countries (Table 3). In some trials, food technologies or government officials were available to explain the process and its potential benefits. Sometimes, both irradiated and nonirradiated products were available for comparison.

All of these trials have shown that most people expressed no objection to, or actually preferred, the quality of the irradiated product, and would buy such products again. These results have been borne out when irradiated food goes on retail sale. Irradiated chicken and fruit has been available in small retail outlets in the USA since the mid-1990s. The irradiated food has sold well, and there has been no significant consumer resistance, although some individual consumers may have chosen not to purchase.

The marked difference between attitudes revealed in surveys of general opinion and the experience from market trials or retail sale of irradiated food indicates that food irradiation gains wide acceptance only when irradiated products are available for consumers to sample.

A few trials or surveys have considered the extra cost of irradiated food. Cost is not a major barrier, with many consumers willing to pay extra for obvious improvements in quality. This willingness is dependent on the type of benefit, the cost of food (a basic or luxury item) and the socioeconomic status of the potential purchaser.

Table 3 Examples of market trials of irradiated foods[a]

Food	*Quantity used in trial (tonnes)*	*Date*	*Country (district)*
Potatoes	82.3	1988	Cuba (Havana)
	800	1984	Pakistan (Peshawar)
	8	1987–88	Poland (Proznan)
	8.2	1984–89	China (Shanghai, Hengshan)
Onion	55	1985–88	Argentina (Buenos Aires, Bahia Blanca)
	85	1984–88	Bangladesh (Dacca, Chittagong)
	1250	1984–89	China (Shanghai, Tientsin)
	16.2	1988	Cuba (Havana)
	16	1986–89	Parkistan (Peshawar)
	7	1984–86	Philippines (Davao, Manila)
	6.5	1986–88	Poland (Poznan, Warsaw)
	800	1986–87	Thailand (Bangkok)
Dried fish	11.5	1985–90	Bangladesh (Dacca, Chittagong)
	1.4	1986–89	Indonesia (Jakarta)
Strawberry	13	1987–88	France (Lyon)
	1	1989	Germany (Leipzig)
Mango	2	1986	USA (Miami)
Apple	500	1984–88	China (Shanghai, Tientsin)
	0.3	1988	USA (Missouri)

[a]Data obtained from information held by the International Consultative Group on Food Irradiation, IAEA, Vienna, from which a complete listing may be obtained.

Summary – The Irradiated Food Dichotomy

Many national authorities are convinced of the safety and benefits of irradiated foods and have approved the process for some specific foods. The public have remained inherently suspicious, however, and opposition is generally reinforced by debates involving opponents and proponents of the process over issues such as safety and labeling.

Favorable responses to irradiated products are elicited if consumers are provided with the products to sample and compare, and if information is available simultaneously. However, little irradiated food is actually available. The food trade remains cautious, fearing adverse consumer reaction and the potential questioning of why existing products need to be irradiated.

Few companies wish to be first to market a product that, despite all its technical advantages, continues to

generate controversy. However, as concerns increase about the safety and security of the food supply and about the effects of chemical treatments, the situation may change. Growth in the volumes of irradiated foods available may be slow for some time yet, but eventually, the process should take its place as one of the key treatments available to the food industry.

See also: **Consumer Protection Legislation in the UK**; **Food Safety**; **Hazard Analysis Critical Control Point**; **Irradiation of Foods**: Basic Principles; Applications; Processing Technology; Legal and Consumer Aspects; **Legislation**: History; International Standards; Additives; Contaminants and Adulterants; Codex; **Quality Assurance and Quality Control**; **Radioactivity in Food**

Further Reading

Anonymous (1991) Code of Good Irradiation Practice for insect disinfestation of cereal grains. In: *International Consultative Group on Food Irradiation Document No. 3*. Vienna: International Atomic Energy Agency.

Bruhn CM (1995) Consumer attitudes and market response to irradiated food. *Journal of Food Science* 58(2): 175–181.

Bruhn CM, Sommer R and Schutz HG (1986) Effect of an educational pamphlet and posters on attitudes toward food irradiation. *Journal of Industrial Irradiation Technology* 41: 1–20.

Consumers' Association (1990) *Food Irradiation: The Consumer's View*. London: Consumers' Association.

Engel RE, Derr DE, Englejohn DE and Keppler HM (1990) Regulatory view of the radiation processing of food. *Radiation Physics and Chemistry* 35(1–3): 232–235.

Food and Agriculture Organization (1984a) *Codex General Standard for Irradiated Foods (CODEX STAN 106-1983) Codex Alimentarius*, vol. XV. Rome: Food and Agriculture Organization of the United Nations.

Food and Agriculture Organization (1984b) *Codex Recommended International Code of Practice for the Operation of Radiation Facilities Used for the Treatment of Foods (CAC/RCP 19-1979, Rev 1). Codex Alimentarius*, vol. XV. Rome: Food and Agriculture Organization of the United Nations.

Food and Agriculture Organization (1997) *Codex Recommended International Code of Practice – General Principles of Food Hygiene (CAC/RCP 1-1969, Rev 3-1997). Codex Alimentarius*. Rome: Food and Agriculture Organization of the United Nations.

Satin M (1993) *Food Irradiation: A Guidebook*. Lancaster, PA: Technomic.

World Health Organization (1989) *Food Irradiation: A Technique for Preserving and Improving the Safety of Food*. Geneva: World Health Organization.

Irritable Bowel Syndrome *See* **Inflammatory Bowel Disease**

ISOMALT

K McNutt, Consumer Magazines Digest, Santa Cruz, CA, USA
A Sentko, Sentko Consult GmbH, Bühlertal, Germany

Background

Isomalt is a sugar-free sweetening ingredient. It is a suitable ingredient for candies, cookies, cough drops, and other products that require the volume and texture properties of sugar, as well as sweetness. Because the same amount of isomalt is used in sugar-free products made with isomalt, as would be used if products were sweetened with sugar, isomalt-containing products have the same appearance and texture as those made with sugar.

Isomalt belongs to the category of carbohydrates known chemically as polyols or sugar alcohols. The term 'sugar replacers,' which more descriptively explains their usage, is now used by many food technologists and nutritionists for this group of sweeteners when communicating with people who are not familiar with chemical nomenclature. The consumer benefits of isomalt include:

1. Its energy value is at most 2 kcal g^{-1}, rather than 4 kcal g^{-1}, as for sugars.
2. It does not promote dental caries.
3. It does not significantly increase blood glucose or insulin levels.

Synthesis and Structure

Development of isomalt began when a patent was issued in 1957 for a process in which *Protaminobacter rubrum* transformed sucrose into the reducing disaccharide 6-*O*-α-D-glucopyranosyl-D-fructose, which is chemically named isomaltulose.

Isomalt is produced as shown in Figures 1 and 2. First, sucrose (2-*O*-α-D-glucopyranosyl-β-D-fructofuranosyl) is enzymatically rearranged into an intermediary product, the more stable isomaltulose (6-*O*-α-D-glucopyranosyl-D-fructose), which is a reducing disaccharide. Then, isomaltulose is hydrogenated with hydrogen and a metal catalyst. The result is isomalt, a mixture of the isomers 1-*O*-α-D-glucopyranosyl-D-mannitol dihydrate (1,1-GPM) and 6-*O*-α-D-glucopyranosyl-D-sorbitol (1,6-GPS).

Technical Properties

Isomalt is an odorless, white, crystalline, and low-hygroscopic substance. Isomalt tastes like sugar, but it is less sweet. In a 10% solution, its sweetening power is 50–60% that of sucrose. Although it has less sweetening power, it has a similar sweetness profile.

One of the advantages of using isomalt is that it can be combined with intense sweeteners to achieve various sweetness profiles, thus optimizing sweetness without masking flavor, which is a limitation of some other sweeteners. Furthermore, isomalt blends well with many flavors, including fruity, menthol, and minty. Isomalt's negative heat of solution ($-39.4\,\mathrm{kJ\,kg^{-1}}$) results in a cooling effect lower than all other sugar replacers. Therefore, products such as isomalt-containing chocolates have no unpleasant cooling effect in the mouth.

Standard isomalt has a solubility of 24 g per 100 g of solution at 20 °C, increasing at higher temperatures. Based on its sorption isotherm, an index of raw material hygroscopicity, isomalt is a very stable ingredient. It absorbs virtually no water up to a relative humidity of 85% at 25°C.

Isomalt and its two isomers are very stable to acidic and enzymatic hydrolysis. Since isomalt's disaccharide bond cannot be easily cleaved, it does not provide a substrate for most microorganisms associated with food. As is true for all polyols, isomalt is nonreducing, and therefore does not react with amino or peptide groups. Therefore, no browning reaction occurs during boiling, baking, or extrusion.

Synergistic effects in sweetening power occur when isomalt is combined with other volume-providing sweeteners such as maltitol syrup, xylitol, mannitol, or sorbitol. Because isomalt is less sweet than sugar, it is often used in combination with intense sweeteners such as aspartame or acesulfame-K. Isomalt gives products texture, volume, and mild sweetness. The intense sweetener brings the level of sweetness up to what would occur if sugar were used. Synergistic effects, similar to those occurring with other volume-providing sweeteners, are observed when isomalt is used in combination with intense sweeteners. An additional advantage of such combination usage is that isomalt tends to mask the bitter aftertaste of some intense sweeteners.

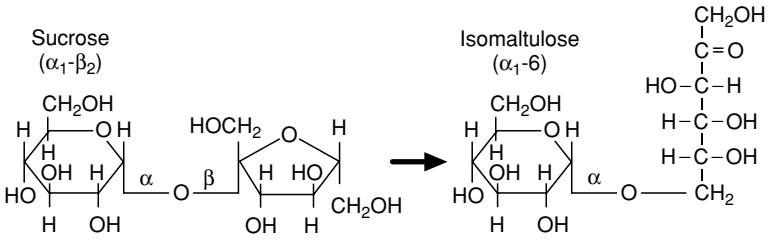

Figure 1 Production of isomaltose by enzymatic transglucosidation.

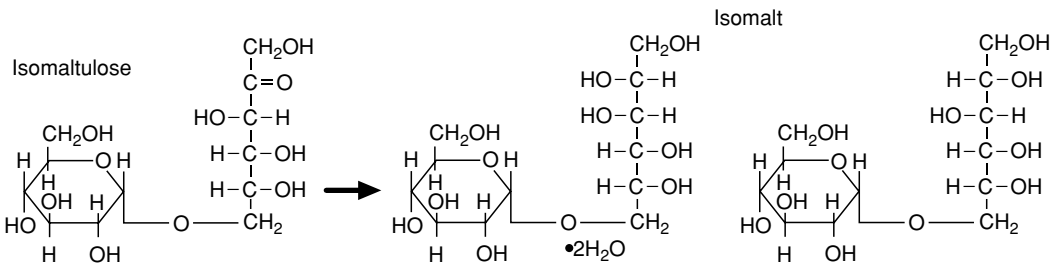

Figure 2 Production of isomalt by hydrogenation.

Consumer Benefits

Isomalt offers benefits that fit health-oriented lifestyles, contemporary weight management guidelines, and the functional foods or nutraceuticals product category.

The current focus of health professionals regarding obesity prevention and weight maintenance is to help people set reasonable weight goals, rather than striving for a defined ideal weight. People are encouraged to develop moderate but regular exercise patterns and to consume a variety of healthful foods. Although no foods should be considered forbidden, the total diet should have a caloric content appropriate for one's energy expenditure. In this context, isomalt expands food-choice options for the growing number of people who are willing to make moderate, but not extreme, lifestyle changes and improvements in their diet. Because of its lower calorie value and other health benefits, isomalt is an ideal ingredient of products for consumers who will adopt healthier eating patterns, as long as foods still taste good and as long as they can still occasionally enjoy desserts, candy, and other sweetened foods.

Another consumer benefit arises from the fact that isomalt is frequently used as an ingredient for many products consumed as snacks or between meals, rather than eaten with other foods. Therefore, it is useful for people who wish to consume sweetened foods without increasing their cariogenic risk or elevating blood glucose and insulin levels.

Isomalt's benefits extend beyond the health arena. Many consumers already factor into their product-choice decisions their concerns about environmental waste. This consumer expectation will, no doubt, become increasingly important as it relates to packaging. Since isomalt absorbs very little water, products made with it tend not to be sticky. Therefore, small items such as candies can be put into a package without each being wrapped separately, an appealing attribute for the growing market of environmentally conscious consumers.

Physiological Properties

The stability of the disaccharide bond between GPM and GPS explains many of the physiological properties of isomalt. From the beginning of the gastrointestinal tract to the end, this enzymatic stability is the key to understanding isomalt's health benefits (see Table 1)

- The reason isomalt does *not promote dental caries* is that oral bacteria cannot readily break its disaccharide bonds and therefore cannot convert it into decay-causing acids.
- Isomalt's *lower caloric value* is mainly due to the fact that intestinal enzymes are not able to hydrolyze its disaccharide bond easily. It is only partially digested and absorbed from the small intestine into the blood, and therefore only partially metabolized.
- The *low-glycemic* benefit of isomalt occurs because of its partial digestion and slow absorption.
- The possible *gastrointestinal* effects of isomalt also result from its incomplete absorption.
- Isomalt's more stable structure also contributes to its usefulness in a *large variety of products*. (*See* Product Applications Section.)

Lower Caloric Value

Based on scientific studies evaluated by the Federation of American Societies for Experimental Biology, an energy value of $2\,\mathrm{kcal\,g^{-1}}$ is used for isomalt for food labeling purposes in the USA. This value is also accepted for food labeling purposes in Canada and other countries where an individual approach to energy values based on scientific evidence is used. The European Union labeling regulation administratively prescribes an average unitary energy value for isomalt and all sugar replacers of $2.4\,\mathrm{kcal\,g^{-1}}$.

Determining the exact amount of food energy that actually becomes available to body cells is more difficult for low-digestibility carbohydrates such as sugar replacers and high-fiber foods than for totally digestible carbohydrates. The energy actually available for body functions depends on:

- the amount of a food digested and absorbed across the intestine wall into the blood;
- how the absorbed products of a food component are metabolized by the body; and
- whether intestinal bacteria use, for their own energy needs, food components that are not digested or absorbed.

Table 1 Isomalt's stable bond explains its health benefits

	because of isomalt's more stable bond . . .		*Therefore, isomalt . . .*
Mouth	Isomalt is not broken down by oral bacteria.	→	*. . . does not cause tooth decay.*
Small intestine	Isomalt is only partially and slowly digested.	→	*. . . provides 50% less kilocalories.*
		→	*. . . does not raise blood glucose and insulin levels.*
Large intestine	Non-absorbed isomalt can pull water into the intestines and can be metabolized by some colonic bacteria.	→	*. . . Eating excessive amounts may cause more gas and looser stools in sensitive people.*

Although isomalt is only partially digested and, therefore, only partially metabolized, the nonabsorbed portion may be used as an energy source for normal colonic bacteria, and some of the end products of this metabolism may be absorbed and converted into biological energy.

Several methods, each with certain advantages and disadvantages, have been utilized to determine the energy value of isomalt. The types of studies include body composition of laboratory animals, indirect calorimetry in human subjects, gastrointestinal absorption research, and studies of the effects on blood glucose, insulin levels, and other metabolic parameters.

Lower Dental Caries Risk

Isomalt is used in many sugar-free products that are frequently consumed between meals such as candies, chewing gum, throat lozenges, and cough drops. This usage increases the importance of its noncariogenic property. A recent review of studies in laboratory animals and humans, using several different techniques to test the cariogenic potential of isomalt, concluded, 'The evidence shows clearly that isomalt has a low acidogenicity and a low cariogenic potential according to criteria established within the dental profession.'

Isomalt does not promote dental caries, because:

- It cannot be fermented by most oral bacteria. Consequently, little or no acid is produced from isomalt metabolism within plaque. The critical pH, 5.7, at which demineralization occurs, is not reached after isomalt ingestion.
- It is not used for polyglucan synthesis, and therefore, it reduces plaque formation.
- Consumption of isomalt-containing products helps remineralize early caries lesions.

Considerable research has been conducted regarding the noncariogenicity of isomalt. The types of studies include *in-vivo* pH studies, *in-vitro* animal studies, and *in-vitro* acid-production studies.

Isomalt solutions alone and a large number of isomalt-containing sugar-free confectionery products have been tested extensively by an *in-vivo* pH telemetry test developed at the University of Zurich. This methodology measures pH changes in interdental plaque and in intraoral fluid after consumption of sugar replacer-containing solutions and sugar replacer-containing food products. The rationale on which it is founded is that the more acidic the environment on the tooth surface, the greater the chance for enamel demineralization and caries formation. Figure 3 shows the details of this procedure.

Imfeld studied the cariogenicity of several confectionery products prepared with isomalt versus similar sucrose-containing products. Isomalt-based chocolate did not cause interdental or intraoral pH levels to drop below 6.0. The sucrose-based chocolate, however, resulted in acid production below the critical level at which tooth decay occurs.

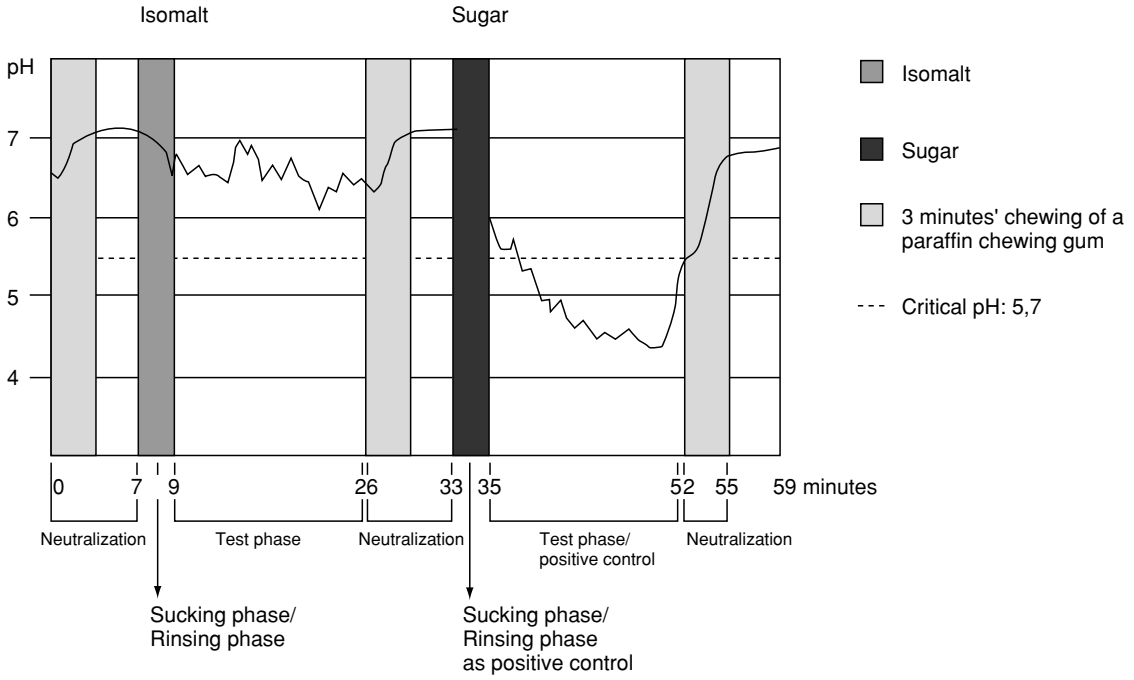

Figure 3 Comparison of oral pH changes after isomalt versus after sucrose using pH telemetry.

Figure 4 The Happy Tooth logo issued by Tooth-Friendly Sweets International.

In 1996, the US Food and Drug Administration (FDA) published a regulation that allows, under specific circumstances, the use of a 'health claim' based on the association between polyols (including isomalt) and the nonpromotion of dental caries. A nonpromotion of dental caries statement is used for food labeling in many other countries. In Switzerland and several other countries, confectionery products that meet criteria similar to those of the US FDA may use the claim 'tooth-friendly.' Manufacturers of products that pass the pH telemetry test may advertise foods as 'safe for teeth' and may print on those products' packaging a logo showing a Happy Tooth, shielded by an umbrella, as shown in Figure 4.

The question of adaptation has been addressed in several studies. The combined results of this research suggest that feeding isomalt to rats does not stimulate the proliferation of those strains that have the ability to ferment isomalt.

Isomalt and other sugar replacers can reverse the initial stages of dental caries, because, owing to their sweet taste, sugar replacers stimulate saliva flow rate and change its composition. Stimulated saliva is an important remineralization agent, because it causes an increase in pH and in salivary calcium concentration, while decreasing inorganic phosphate concentration.

Low Glycemic Response

Isomalt availability has significantly increased the quality and variety of good-tasting, sugar-free products that do not elevate blood glucose and insulin levels. This expansion of low-glycemic response options is especially important for cough drops, throat lozenges, and candies that are usually not consumed at meals, along with foods that contain protein, fat, and complex carbohydrates.

Numerous studies in diabetic and nondiabetic subjects have been conducted on the effect of isomalt alone and of isomalt in comparison with other sweeteners. This research shows that, after ingestion of isomalt, blood glucose and insulin values do not differ significantly from baseline levels.

Two physiological reasons explain isomalt's lower glycemic response. First, isomalt is only partially digested and therefore only absorbed to a limited degree in the small intestine. Second, the portion that *is* absorbed is absorbed slowly, thus resulting in a small increase in blood glucose and insulin levels.

In one study, using nondiabetic subjects who had fasted overnight 43 people were given 50 g of isomalt and seven people were given 100 g. No significant increase in blood glucose was observed during the 2 h after ingestion. In a double-blind cross-over study, 10 healthy young men of normal weight were given comparable amounts of sucrose and isomalt. Blood glucose and insulin showed a marked increase after 30 g of sucrose, whereas after isomalt, these parameters did not change appreciably from the baseline. The differences between isomalt and sucrose responses were highly significant (see Figure 5).

In a randomized cross-over study with 24 Type II diabetics who had fasted for 12 h, subjects were given isomalt or glucose and then blood samples were collected over the next 3 h. In contrast to glucose, after ingestion of isomalt, the blood glucose and insulin curves were not significantly different from those for fasting levels. Petzhold and coworkers confirmed these findings in a comparably designed study. In addition, they found no increase in C-peptides after isomalt, whereas a significant increase occurred after glucose.

In order to evaluate the long-term effects of isomalt, Pometta *et al*. conducted a randomized controlled study with two groups, each composed of 12 Type II diabetics who consumed their usual diet for a period of 3 months. The experimental subjects received 24 g of isomalt daily, offered throughout the day. No significant differences were found between the two groups, either in metabolic parameters (fasting and postprandial blood glucose, hemoglobin A1C, total cholesterol, high-density-lipoprotein cholesterol, triglycerides, phospholipids) or in liver and kidney function tests. In another cross-over study, 60 insulin-dependent children with Type I diabetes were treated for 6 weeks either with a standard diet or with a diet with the daily addition of 20 g of isomalt. This daily dosage level had no effect on insulin requirement, frequency of hypoglycemia, or frequency of glucosuria.

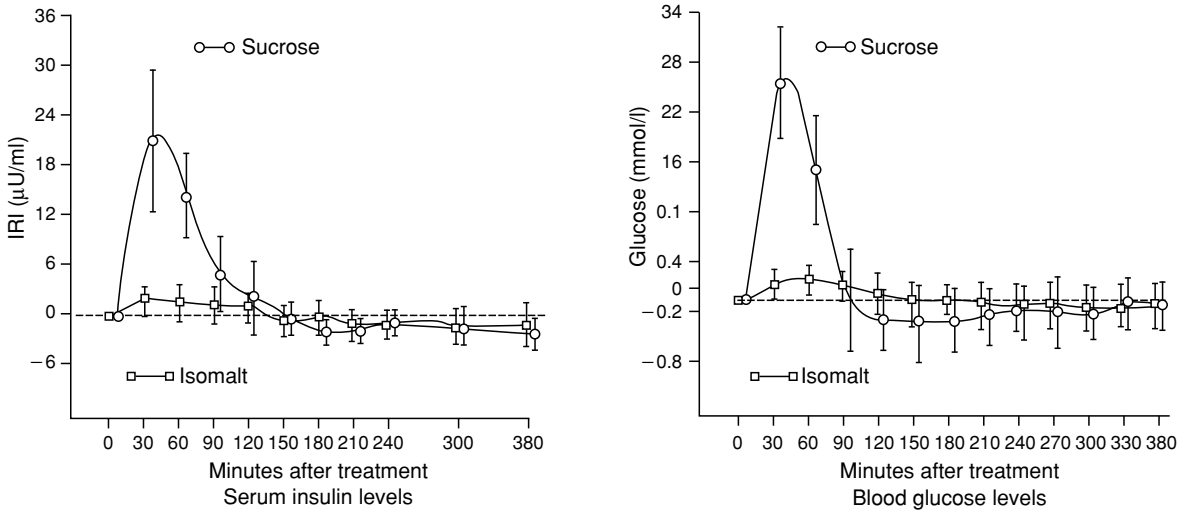

Figure 5 Comparison of serum insulin and blood glucose levels after oral isomalt versus after oral sucrose.

Gastrointestinal Response

Isomalt, as well as all sugar replacers, are low-digestibility carbohydrates, as are high-fiber foods, beans, prunes, cabbage, and dried fruits. Any food that contains a relatively large amount of low-digestibility carbohydrates is, as the term 'low digestibility' indicates, only partially (or not at all) absorbed in the upper part of the gastrointestinal (GI) tract. Nonabsorbed carbohydrates create an osmotic effect, pulling water into the intestine, which may *sometimes* in *some* people result in softer or more watery stools. In the lower part of the GI tract, the nonabsorbed portion of such foods is partially or completely fermented by normal colonic bacteria. During fermentation, some bacteria change low-digestibility carbohydrates into short-chain fatty acids and gases, which can cause flatulence.

Most people can eat fairly large amounts of foods containing low-digestibility carbohydrates without feeling uncomfortable. Others know from experience that, because such foods may cause an unpleasant GI response, they should limit their consumption to smaller portions. Very few, but some, people are so sensitive to low-digestibility carbohydrates that even small amounts cause discomfort.

Setting an absolute level below which an individual can be assured of never having a GI response is impossible, because so many variables affect how each person's body handles low-digestibility carbohydrates. The effect depends on whether the carbohydrate is consumed as a food or a beverage, whether it is eaten as part of a meal or on an empty stomach, and whether the total amount is ingested all at once or spread over the day in small portions. The same person may have no response to a particular food on one day but may be sensitive to it on another day. Usually, the few people who might otherwise be sensitive to a low-digestibility carbohydrate food, adapt after a few days if they start with small portions and gradually increase their total consumption.

A GI study of nonadapted US children given 15, 25, and 35 g, eaten at one sitting, showed no statistically significant difference between isomalt and sucrose. In an adult study, intake up to 48 g, eaten throughout the day over a 12-week period, showed that amounts of isomalt up to that level were well tolerated.

To increase consumer awareness of this physiological aspect of sugar replacers, food legislation in some countries requires a statement on the label of products such as 'Excessive consumption may have a laxative effect.'

Safety Evaluation

The safety of isomalt was evaluated by the Joint Expert Committee on Food Additives (JECFA) of the WHO/FAO. In 1986, JECFA decided that it was not necessary to lay down a numerical acceptable daily intake, and therefore, an acceptable daily intake of 'not specified' was allocated. The most recent Codex Specification for isomalt was published in 1996. A number of national authorities have evaluated isomalt's safety during their approval process, and isomalt is now marketed in more than 80 countries worldwide.

Product Applications

A decade ago, sugar-free products, other than soft drinks and tabletop sweeteners, were consumed almost exclusively by people who avoided sugar

because they had diabetes. People who were not medically required to avoid sugar rarely purchased sugar-free products, because the taste was not sufficiently comparable with those sweetened with sugar. However, the isomalt-containing products available today taste so good that they are marketed to the growing number of people who simply want more healthful diets, as long as they do not have to sacrifice the pleasure of eating.

Isomalt has been used as a product ingredient in Europe and Asia since the early 1980s and in the USA since 1990. It is currently used in more than 80 countries worldwide in a wide range of confectionery, chocolate, baked goods, pharmaceutical, and functional products, including hard candies, lollipops, chewing gum, breath mints, cough drops, throat lozenges, chocolates, fudge, cookies, and wafers.

Several variants of isomalt are available. Standard isomalt ST is offered in different particle sizes, depending on the intended application. Isomalt GS is used in products that require optimum solubility such as coated products; this form also allows low process temperatures. Isomalt LM has a low-moisture content and is designed for premium no-added-sugar chocolate applications using standard processing and standard conching temperatures. Isomalt DC provides good compressibility, flowability, hardness, and stability for direct compressed tablets.

A wide range of deposited, stamped, or filled sugar-free sweets can be produced with isomalt. Existing processing equipment can be used for all applications, and usually, only minor formula and process parameter modifications are needed. The production is similar to that of candies based on sucrose/corn syrup, and only minor changes in parameters and processing are required. In either batch or continuous process production, differences from sucrose/corn syrup that must be taken into consideration include a lower solubility, higher boiling point, lower viscosity of the melt, and higher specific heat capacity.

Because of its low hygroscopicity, isomalt-based hard candies can be packed in flip-top carton boxes without secondary wrapping. When the water content in the finished product is below 2%, hard candies with a very good shelf-life are obtained. Pan-coated products also have a good shelf-life. All types of centers can be panned with isomalt. Depending on the consistency, however, it may be necessary to pregum centers to build up the coating over the smooth surface.

Because isomalt-containing pharmaceutical products, such as cough drops, have a very low moisture absorption, oxidation or hydration of active ingredients is reduced, and shelf-life is extended. Isomalt's minimal hygroscopicity makes it an especially useful ingredient when hydrolysis-proof ingredients are involved. Furthermore, isomalt's lower solubility allows a slower release of active ingredients when a pharmaceutical hard-candy-like product is sucked.

See also: **Carbohydrates**: Classification and Properties; Digestion, Absorption, and Metabolism; **Dental Disease**: Role of Diet; **Diabetes Mellitus**: Treatment and Management; **Energy**: Measurement of Food Energy; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Functional Foods**; **Glucose**: Glucose Tolerance and the Glycemic (Glycaemic) Index; **Sugar Alcohols**

Further Reading

American Diabetes Association (2002) Position statement: Nutrition recommendations and principles for people with diabetes mellitus. *Diabetes Care* 25(supplement 1): S43–S46.

American Dietetic Association (1998) Position paper: Use of nutritive and nonnutritive sweeteners. *Journal of The American Dietetic Association* 98: 580–587.

FASEB (1994) *The Evaluation of the Energy of Certain Sugar Alcohols Used as Food Ingredients*. Bethesda, MD: Federation of American Societies for Experimental Biology.

Featherstone JDB (1994) Effects of isomalt sweetener on the caries process: a review. *Journal of Clinical Dentistry* 5: 82–85.

Food and Drug Administration (1996) Health claims: dietary sugar alcohols and dental caries. *Federal Register*, 61(154): 43433–43445, August 23, 1996 and 21 *Code of Federal Regulations* §101.80.

Fritzsching BR (1995) Isomalt in hard candy applications. *Manufacturing Confectioner* November 65–73.

Imfeld TJ (1983) *Identification of Low Caries Risk Dietary Components* Basel: Karger.

Imfeld TJ, Hirsch RS and Mühlemann HR (1978) Telemetric recordings of interdental plaque pH during different meal patterns. *British Dental Journal* 144: 40–45.

McNutt K (1998) Sugar replacers: a new consumer education challenge. *British Nutrition Foundation Nutrition Bulletin* 23: 216–223.

McNutt K (2000a) What clients need to know about sugar replacers. *Journal of The American Dietetic Association* 100: 466–469.

McNutt K (2000b) Sugar-free ain't what it used to be: Here's why sugar replacers are improving your options. *Diabetes Interview* 9(8): 57–58.

McNutt K (2000c) Sugar replacers and the FDA noncariogenicity claim. *Journal of Dental Hygiene* 74: 36–40.

McNutt K and Sentko A (1996) Sugar replacers: A growing group of sweeteners in the United States. *Nutrition Today* 31: 255–261.

Paige DM, Bayless TM and Davies RL (1992) Palatinit® (isomalt) digestibility in children. *Nutrition Research* 12: 27–37.

Spengler M, Somogyi JC, Pletcher E and Boehme K (1987) Tolerability, acceptance and energetic conversion of isomalt (Palatinit®) in comparison with sucrose. *Aktuelle Ernährungsmedizin* 12: 210.

Thiébaud D, Jacot E, Schmitz H, Spengler M and Felber JP (1984) Comparative study of isomalt and sucrose by means of continuous indirect calorimetry. *Metabolism* 33: 808–813.

Warshaw HS and Powers MA (1999) A search for answers about foods with polyols (sugar alcohols). *The Diabetes Educator* 25: 307–321.

WHO/FAO (1996) Isomalt. *Compendium of Food Additives Specifications. FAO Food and Nutrition Paper*, 52, Addendum 4, 79–83. Rome: FAO.

Wijers MC and Sträter PJ (2001) Isomalt. In: Nabors LO (ed.) *Alternative Sweeteners*, 3rd edn, revised and expanded, pp. 265–282. New York: Marcel Dekker.

IUFOST *See* **International Union of Food Science and Technology**

K

KETONE BODIES

C R Barnett and Y A Barnett, University of Ulster at Coleraine, Coleraine, County Londonderry, Northern Ireland

Introduction

The liver is the major organ associated with the processing, production, and export of glucose and fatty acids to support the energy requirements of the other tissues and organs in the body. Under certain physiological and pathophysiological conditions such as fasting and insulin-dependent diabetes mellitus (IDDM) respectively, the liver can produce significant quantities of acetoacetate and D-3-hydroxybutyrate (ketogenesis). These compounds enter the circulation where acetoacetate spontaneously decarboxylates, giving rise to acetone. These three substances are collectively known as ketone bodies (Figure 1).

Formation and Utilization of Ketone Bodies

Glucose is the primary source of respiratory fuel for the body. However, when circulating levels of glucose fall and cannot meet the energy requirements of the tissues, other energy stores are mobilized. Lipid provides the most important energy store within the body and can be mobilized through the action of lipase in the adipose tissues. The released fatty acids are transported to the liver and extrahepatic tissues where they are oxidized in the mitochondria to generate adenosine triphosphate. However, the increased supply of fatty acids to the liver results in an increased abundance of acetyl coenzyme A (CoA). As the concentration of acetyl CoA increases, ketone body production is initiated. The hydroxymethylglutaryl-CoA (HMG-CoA) pathway (Figure 2) is responsible for the production of acetoacetate and D-3-hydroxybutyrate. The initial step is the formation of acetoacetyl CoA from two molecules of acetyl-CoA catalyzed by acetoacetyl-CoA thiolase. The unfavorable equilibrium for this reaction is overcome by the acetoacetyl CoA being bound to the HMG-CoA synthase enzyme. A further molecule of acetyl CoA then reacts to form HMG-CoA which is the precursor for two pathways. HMG-CoA can be reduced through the activity of HMG-CoA reductase to form mevalonic acid, an intermediate in the pathway of steroid biosynthesis, or HMG-CoA can be cleaved by HMG-CoA lyase to form acetoacetate and acetyl CoA. An alternative route for the formation of acetoacetate involves the direct deacetylation of acetoacetyl CoA. However, the concentration of acetoacetyl-CoA in the liver is very low, whereas the Michaelis constant (K_m) of the acetoacetyl CoA deacetylase is high. D-3-hydroxybutyrate is produced from acetoacetate by the action of the enzyme D-3-hydroxybutyrate dehydrogenase (Figure 2), while acetone is produced from the spontaneous decarboxylation of acetoacetate in the blood. The concentrations of acetoacetate and D-3-hydroxybutyrate in the blood are in equilibrium with each other. The equilibrium is dependent on the mitochondrial redox state and can result in plasma ratios of D-3-hydroxybutyrate to acetoacetate concentrations of 1:1 to 10:1.

Many tissues can utilize ketone bodies, with the notable exception of the liver. Conversion of D-3-hydroxybutyrate and acetoacetate back to acetyl CoA involves the enzymes 3-oxoacid-CoA transferase and acetoacetyl CoA thiolase (Figure 3). The equilibrium of the transferase reaction is in favor of acetoacetate but that of the thiolase reaction is towards acetyl CoA. Thus the linking of the two reactions overcomes the unfavorable equilibrium of the transferase reaction. D-3-hydroxybutyrate is metabolized by conversion to acetoacetate catalyzed by D-3-hydroxybutyrate dehydrogenase and then via the reaction pathway outlined previously (Figure 3) to yield acetyl CoA.

Regulation of Ketogenesis

Two hormones are primarily responsible for the regulation of ketogenesis by the liver: glucagon and

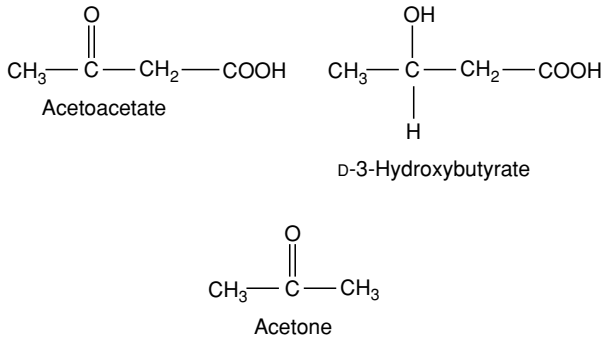

Figure 1 Chemical structures of the ketone bodies. Reproduced from Ketone Bodies, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

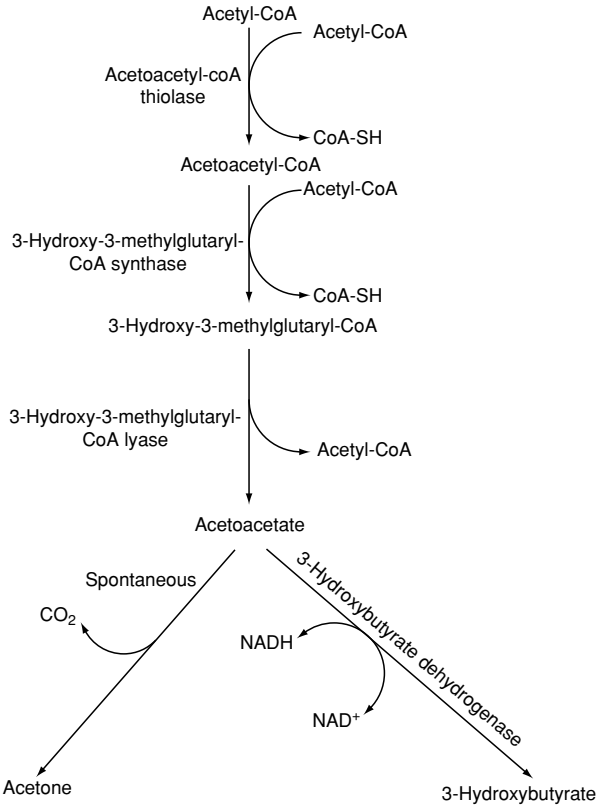

Figure 2 Major pathway for the formation of ketone bodies. NAD+/NADH, Nicotinamide dinucleotide (oxidized and reduced respectively); SH, Sulphydryl Group. Reproduced from Ketone Bodies, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

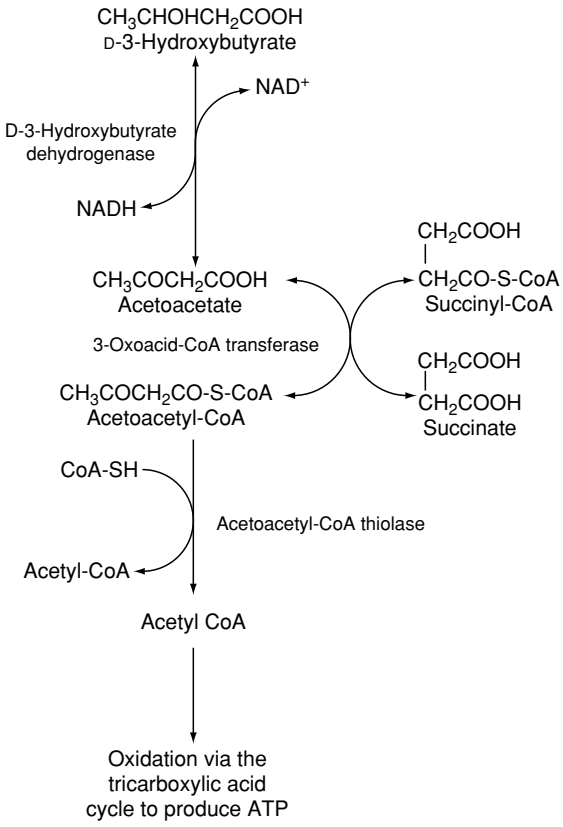

Figure 3 Major pathways of ketone body utilization by extrahepatic tissues. CoA, coenzyme A; ATP, adenosine triphosphate. NAD+/NADH, Nicotinamide dinucleotide (oxidized and reduced respectively); SH, Sulphydryl Group. Reproduced from Ketone Bodies, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

insulin. Both hormones are produced by the pancreatic islets of Langerhans which respond to the fluctuations in circulating nutrient concentrations by adjusting the secretion of glucagon (from the pancreatic A cells) or insulin (from the pancreatic B cells) into the circulation. The plasma ratio of these two hormones plays a central role in the regulation of glucose metabolism by the liver and ketogenesis (Figure 4). In the fed state, the ratio of insulin to glucagon is high and glucose is stored as glycogen and oxidized via glycolysis and the tricarboxylic acid cycle, producing a range of biosynthetic intermediates and adenosine triphosphate. The presence of the biosynthetic intermediates, in particular malonyl CoA (used for fatty acid biosynthesis), inhibits the transport of fatty acids into the mitochondria, thereby preventing fatty acid oxidation. As such, ketogenesis is prevented when there is a plentiful supply of plasma glucose for oxidation by the tissues (Figure 4).

During periods of food deprivation the ratio of circulating insulin to glucagon falls, causing activation of glycogenolysis, a reduction in the activity of glycolytic enzymes, and reduced activities of acetyl CoA carboxylase. The decreasing concentration of malonyl CoA results in activation of the

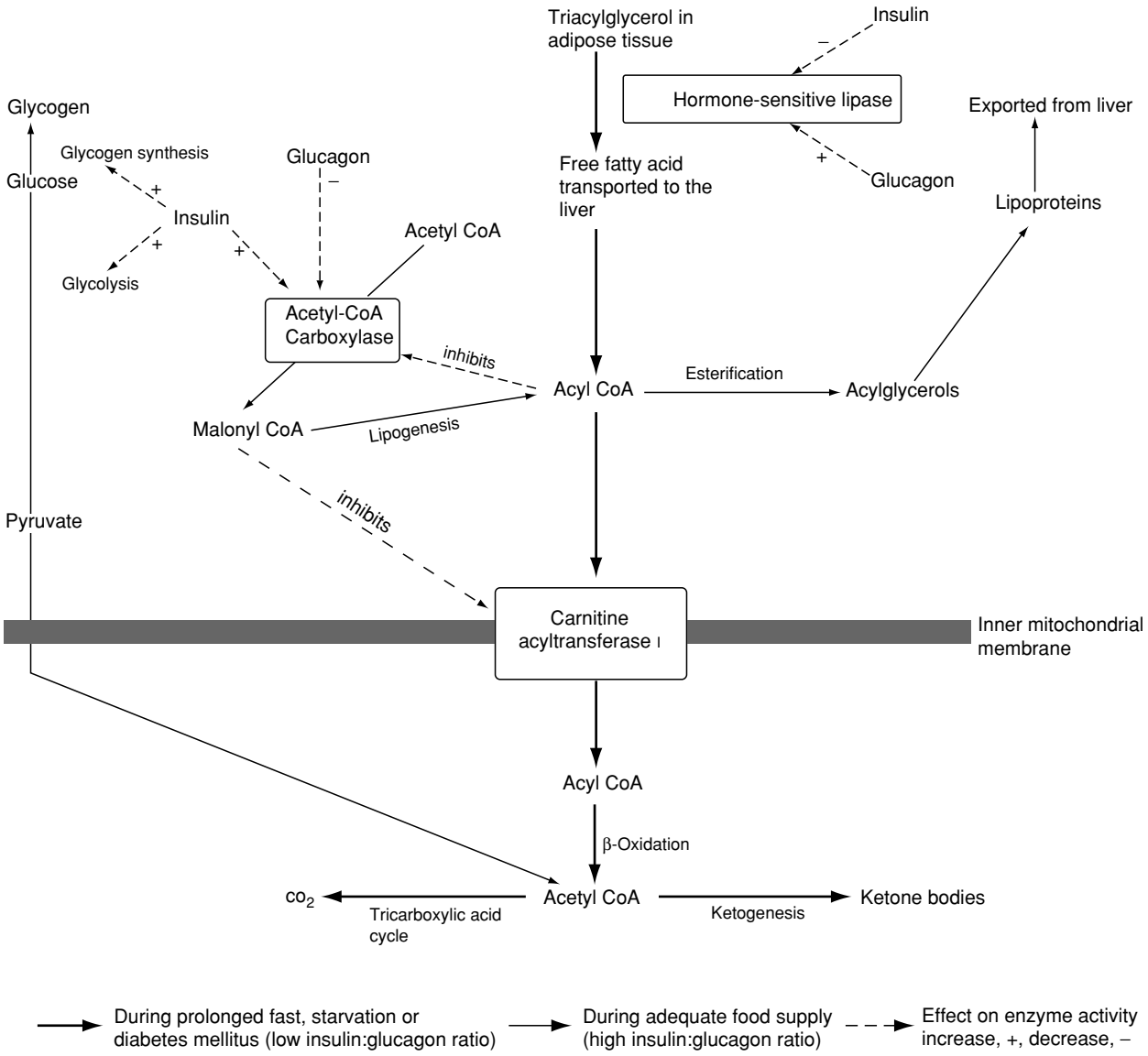

Figure 4 Regulation of ketogenesis in the liver. CoA, coenzyme A. Reproduced from Ketone Bodies, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

mitochondrial transport system, allowing free fatty acids to be transported into the mitochondria for oxidation. As the concentration of acetyl CoA in the mitochondria of the liver rises, ketogenesis is stimulated. In the adipose tissue, glucagon stimulates adenylate cyclase which in turn activates lipases that convert the stored triglyceride into free fatty acids. The increased delivery of these fatty acids to the liver results in increased ketogenesis, the rate of which is primarily governed by the delivery of the fatty acids (Figure 4).

When a fast is terminated, the increased concentrations of nutrients in the plasma cause insulin release from the endocrine pancreas. The rising insulin-to-glucagon ratio causes cessation of adipose tissue lipolysis, and plasma free fatty acid concentrations rapidly fall. Insulin also acts on the tissues, causing a reversion to glucose utilization. In the liver, glucose output is reduced, glycogen stores are repleted, lipogenesis is reinstated, and ketogenesis inhibited.

Physiological Ketosis

A normal balanced diet provides sufficient nutrient content to satisfy the glucose requirements of the brain and other tissues. Under these conditions ketogenesis by the liver is negligible, with plasma ketone

body concentrations generally less than 0.5 mmol l^{-1}. However, the ability of an individual to survive periods of food deprivation normally relies on a complex hormonal and biochemical interaction that insures the appropriate interorgan redirection and utilization of major fuels.

In the early stages of a fast there is a shift in liver metabolism from glucose storage to glucose production, coupled with other metabolic alterations that conserve the glucose released for the maintenance of brain and central nervous system function. Following an overnight fast, most of the glucose produced by the liver is utilized by the tissues with an obligatory requirement for glucose as the major fuel. At this stage, tissues such as skeletal muscle are primarily using nonesterified fatty acids (NEFA) as their major fuel supply. Muscle NEFA oxidation causes a decrease in glucose uptake and subsequent glucose metabolism by muscle cells and glucose transport across the muscle cell membrane is decreased.

Reduction in the plasma glucose concentration as a result of a prolonged fast or food deprivation will result in the production of glucose from noncarbohydrate precursors. Fatty acids are mobilized from the adipose tissue as the fast takes place and gluconeogenic and ketogenic pathways are activated in the liver. The plasma glucose concentration may fall significantly with the concomitant increase in plasma fatty acids and ketone bodies to about five fold and 20-fold, respectively. As more than 99% of plasma fatty acid is complexed with albumin, very little is free to diffuse into the interstitial spaces, limiting their utilization by tissues. However, ketone bodies are soluble in aqueous medium and the concentration in the interstitial fluid will be comparable with that in the plasma (2–3 mmol l^{-1}) during prolonged fasting. As the concentration of glucose in the interstitial fluid under these conditions may be around 4–5 mmol l^{-1}, the ketone bodies are oxidized in preference to glucose and free fatty acids and sparing glucose for use by the brain and central nervous system.

The pancreatic B cells remain responsive to the plasma nutrient concentration and facilitate the maintenance of plasma fatty acid concentrations such that the plasma ketone body concentration does not usually rise to such an extent that ketone bodies are excreted in the urine.

Pathological Ketosis

IDDM is probably the best-known disease associated with life-threatening ketosis if not treated by insulin therapy. IDDM is associated with autoimmune destruction of the pancreatic islets of Langerhans. This results in very low or undetectable levels of circulating insulin in the plasma and is often also associated with increased plasma concentrations of glucagon. As described previously, the alteration of the insulin-to-glucagon ratio has significant effects on the metabolism of the liver and other tissues. The increased relative or absolute levels of glucagon increase the rate of lipolysis in the adipose tissue, resulting in increased circulating levels of free fatty acids. The liver responds to this alteration in the insulin-to-glucagon ratio by increasing glucose synthesis and release. As insulin-stimulated glucose uptake by extrahepatic tissues is decreased, the circulating glucose concentration increases and hyperglycemia occurs. Lipid and glycogen synthesis are inhibited and there is increased oxidation of the free fatty acids being delivered to the liver. In physiological ketosis there is a regulated release of fatty acids from the adipose tissue allowing tight regulation of the plasma ketone body concentrations. However, in IDDM this regulation is nonfunctional due to the autoimmune destruction of the pancreatic B cells. As a result, hyperlipidemia and hyperketonemia occur and, as the concentration of ketone bodies continues to increase above a threshold level of around 7 mmol l^{-1}, the kidneys begin to excrete the excess ketone bodies into the urine.

Metabolic Acidosis

Metabolic acidosis is caused by a decrease in the concentration of bicarbonate ion in the plasma with little or no change in the carbonic acid concentration. The condition is frequently associated with uncontrolled IDDM where it is caused by the presence of high plasma concentrations of ketone bodies. Acetoacetate and 3-hydroxybutyrate are relatively strong acids and dissociate in the plasma, giving rise to H^+ and the corresponding anion. The H^+ ions are effectively buffered by plasma bicarbonate, giving rise to H_2CO_3, which breaks down to yield CO_2 and H_2O. The increase in plasma CO_2 concentration stimulates the respiratory center, producing rapid and deep breathing, thereby excreting the excess CO_2 by way of the lungs. This is an attempt by the body to reestablish the normal CO_2 tension in the blood. As the kidney is the major organ responsible for the maintenance of the acid–base balance, it is also employed to help correct the imbalance caused by the presence of the ketone bodies. The anionic components of the dissociated ketone bodies are filtered in the kidney, each with a cation (principally a sodium ion), thus maintaining electrical neutrality. The renal tubular cells secrete H^+ into the filtrate while reabsorbing one sodium ion and one bicarbonate ion for each H^+ secreted. The energy-dependent H^+ carriers in

the tubular cells are capable of secreting H^+ against a concentration gradient until the urine becomes approximately 800 times more acidic than the plasma (approximately pH 4.5). At this point the concentration gradient becomes too great for the secretory process to continue. For further H^+ secretion to proceed, the majority of the secreted H^+ must be buffered in the tubular fluid. This is achieved in the first instance by phosphate. However, phosphate is present in the urine as a result of dietary excess and the buffering capacity of urinary phosphates is quickly exceeded in severe acidosis. Under these circumstances the tubular cells secrete ammonia (NH_3), which combines with the urinary H^+, giving rise to NH_4. In this way larger amounts of H^+ can be excreted into the urine before the concentration gradient becomes too large for the secretory mechanism to operate. However, when the acid load is very large, as in diabetic acidosis, this occurs and the filtered cations are lost with the ketone bodies. The concomitant electrolyte and water loses lead to dehydration, hypovolemia, hypotension, and subsequent death, if not treated.

Diabetes and Drug Metabolism

Numerous studies have now demonstrated that IDDM in animals and humans is associated with the alteration of drug metabolism by the liver. The cytochrome P450-dependent mixed-function oxidase system is the most important enzyme system involved in the metabolism of plethora of drugs and chemicals. IDDM is associated with alterations in the levels and activities of some of these cytochrome P450 enzymes and ketone bodies have been shown, at least in part, for the alterations observed. Ketone bodies appear selectively to induce some of the cytochrome P450 enzymes, and this has been shown to result in alteration of metabolism of some drugs in humans. Whether these alterations in clinically controlled IDDM individuals over prolonged periods have any significant therapeutic consequences has yet to be determined.

See also: **Diabetes Mellitus**: Etiology; **Liver**: Structure and Function

Further Reading

Ioannides C, Barnett CR and Flatt PR (1999) The effect of experimental diabetes on the cytochrome P450 system and other metabolic pathways. In: McNeill JH (ed.) *Experimental Models of Diabetes*, pp. 81–117. Boca Raton: CRC Press.

Sherwood L (1993) *Human Physiology: From Cells to Systems*, 2nd edn. New York: West.

Kidney *See* **Renal Function and Disorders**: Kidney: Structure and Function; Nutritional Management of Renal Disorders

KIWIFRUIT

A R Ferguson and R Stanley, HortResearch, Auckland, New Zealand

Introduction

Kiwifruit are enjoyable to eat. They are also a good source of minerals, of dietary fiber, and they contain a most effective laxative. Their outstanding nutritional quality is their very high content of vitamin C. There is very little, if any, loss of nutritional quality during long-term storage or ripening even if the overall appeal and quality of the fruit deteriorates. In particular, kiwifruit that are still fit to eat are likely to be an excellent source of vitamin C. Their main disadvantage is that they can cause allergic or inflammatory responses in some consumers. Kiwifruit account for perhaps only 1% of the total world production of fresh fruit. Nevertheless they are an important fruit crop. They have rapidly become popular amongst consumers, they are novel and new cultivars are now coming on the market, they are available in a range of colors and flavors, and they can mostly be stored for remarkably long periods under refrigerated and controlled-atmosphere (CA) conditions. (*See* **Allergens**; **Antioxidants**: Natural Antioxidants; **Ascorbic Acid**: Properties and

Determination; **Controlled-atmosphere Storage**: Effects on Fruit and Vegetables; **Dietary Fiber**: Properties and Sources; **Food Intolerance**: Food Allergies; **Fruits of Temperate Climates**: Commercial and Dietary Importance.)

What are Kiwifruit?

Kiwifruit are amongst the most recently domesticated of all fruits. Although the term 'kiwifruit' has traditionally been restricted to cultivars of *Actinidia deliciosa* (A. Chev.) C.F. Liang et A.R. Ferguson, the name is now being extended to include cultivars of the closely related species, *A. chinensis* Planch. Fruit of this latter species have recently appeared on the international market: they are still recognizably kiwifruit, even though they are much less hairy, the fruit shape can be different, the skin may be a different color, and the internal flesh can be yellow instead of green. Fruit of other *Actinidia* species, such as *A. arguta* (Sieb. et Zucc.) Planch. ex Miq. are also being commercialized. Although they are very different in appearance to the usual kiwifruit, being much smaller and with smooth green skins, it is likely that they too will eventually be commonly referred to as kiwifruit. We have therefore included any species within the genus *Actinidia* under the term 'kiwifruit.'

The genus is widespread throughout Asia, with its center of origin in China. Plants of *Actinidia* species were first grown outside Asia during the last decades of the nineteenth century. Commercial orchards of *A. deliciosa* were first established in New Zealand about 1930 and the very first commercial plantings of *A. chinensis* outside China have only been established during the past five years.

World Production

The principal areas of kiwifruit production are located approximately between the latitudes 25° and 45°. Kiwifruit are deciduous, temperate plants and they require a period of winter chilling for adequate budbreak and flowering. They are susceptible to damage from spring and autumn frosts and, depending on the species and cultivar, require a long frost-free growing period of at least 7 or 8 months (about 220 days). They require shelter from strong winds and have a high demand for water: irrigation must be sufficient to supplement natural rainfall to between 800 and 1200 mm of water throughout the growing season.

There are extensive natural resources of kiwifruit in China and appreciable quantities of fruit (perhaps 100 000 tonnes) are collected from the wild each year. China also has the greatest commercial plantings of kiwifruit – about 50 000 ha. Outside China, there are about 70 000 ha planted – approximately 20 000 ha in Italy, 12 000 ha in New Zealand and 8000 ha in Chile. Many of these plantings, particularly in China, are still not mature and yields are therefore increasing. Production in the various countries also varies with climatic conditions each year. In 2001, Italy was producing around 350 000 t of kiwifruit, New Zealand 250 000 t, and China and Chile each about 150 000 t per annum.

The kiwifruit industries of New Zealand, Chile, and Italy are dependent on exports and these industries expanded largely to meet external demand. Thus, 85–90% of the kiwifruit produced in New Zealand are exported each year and Chile and Italy each export about 75% of the fruit they produce. China is quite different and nearly all the kiwifruit it produces are consumed within the country: exports are negligible. However, it is predicted that Chinese exports of kiwifruit may soon be significant in world trade. (*See* **Fruits of Temperate Climates**: Commercial and Dietary Importance.)

Kiwifruit Cultivars

In most countries with commercial kiwifruit plantings, only *A. deliciosa* has been grown, and 'Hayward' is the only fruiting cultivar planted along with its corresponding males. (All *Actinidia* species are dioecious, with both male and female vines required for fruit production.) 'Hayward' has therefore become the standard cultivar of international commerce in kiwifruit, accounting for more than 95% of the kiwifruit traded. This reliance on a single cultivar in most parts of the world is most unusual. Such reliance has advantages in that it facilitates standardization amongst suppliers from different countries. There are, however, also disadvantages in that standardization can lead to greater competition in individual markets and makes branding more difficult. 'Hayward' is such a good cultivar, and responds so well to storage, that it is likely to remain important for many years yet.

Kiwifruit plantings in China are much more diverse – approximately 75% in cultivars of *A. deliciosa* and 25% in *A. chinensis*. The cultivar 'Qinmei' (*A. deliciosa*) accounts for about 40% of the total plantings and is therefore the second most widely planted kiwifruit cultivar in the world after 'Hayward'. Six cultivars of *A. deliciosa* and seven of *A. chinensis*

together make up about 85% of all Chinese kiwifruit plantings.

One cultivar of *A. chinensis*, 'Hort16A', bred in New Zealand from germplasm introduced from China, is now being planted in New Zealand (about 2000 ha in 2001) and in California, Italy, and Japan. It is the first cultivar of *A. chinensis* to have been grown commercially outside China but it is likely that other cultivars, selected in China itself, will soon be grown commercially in other countries.

The Structure of Kiwifruit

The kiwifruit is a berry: that is, it has a large number of seeds embedded in fleshy, edible tissue. The fruit develops after pollination from the lower parts of numerous carpels (usually more than 30) which have fused to form a syncarpous, superior ovary. Each carpel has two rows of 10–20 ovules attached to the central axis. If pollination is adequate, a fruit can thus contain 1000 or more seeds.

Fruit of large-fruited selections of *A. chinensis* and *A. deliciosa* typically weigh on average about 100 g but fruit of many of the other species are much smaller, e.g., those of *A. arguta* average between 10 and 20 g. The exterior surface of the fruit differs according to the species and the cultivar. Fruit of *A. deliciosa* (e.g., 'Hayward') are brown and densely covered with persistent, long, stiff hairs, although these are largely lost during commercial grading and packing. Fruit of *A. chinensis* (e.g., 'Hort16A') are light-brown or greenish and are covered with sparse, much softer hairs, rather like the down of a peach. Fruit of some species, e.g., *A. arguta*, lack obvious hairs and their edible skins are smooth and polished.

Internally, the fruit are very attractive with dark seed in a green to dark-green flesh (*A. deliciosa*) or flesh which ranges from green to lime-green to a clear golden-yellow (*A. chinensis*). In both species there may also be attractive red pigmentation in the inner pericarp. The external appearance and the fruit flesh of some species change color during ripening.

The bulk of the outer pericarp consists of thin-walled parenchyma cells, variable in size. The cells are loosely organized and there can be large intercellular spaces. Nevertheless, fruit are reasonably solid ('Hayward' density, *c.* 1.03 Kg m^{-3}). There are also many mucilage-containing cells with crystals of calcium oxalate. Chloroplasts are generally similar to those found in leaves. In the inner pericarp, the seeds are suspended in a mucilaginous matrix in the locules. Chloroplasts of the inner pericarp are very different in appearance and there are many raphide cells containing calcium oxalate. The central core of the mature fruit is formed of homogeneous, large parenchyma cells, although there is often a hard, woody spike at the stalk end.

Following pollination, there is a phase of rapid growth for the first 8 or 9 weeks, during which the fruit reaches half to two-thirds final volume, then about 3 weeks of much slower growth followed by another period of growth, which is initially rapid but then decreases. During the first stage of fruit growth, there is rapid cell division in the outer and inner pericarp and the central core. Cell division in the pericarp stops after 3–4 weeks and most of the subsequent increase in fruit size is due to cell enlargement as the parenchyma cells of the fruit are 10–15 times the size of those in the ovary wall. Cell division may continue for much longer in the central core, although at a much reduced rate.

The Chemical Composition of Kiwifruit

Only fruit of *A. deliciosa* have been systematically analyzed in any detail and most results are for the cultivar 'Hayward'. There is considerable variation in the published values and those given in Table 1 should be considered only as indicative. Individual fruit from a single vine can vary greatly in composition and many factors such as growing conditions or postharvest treatment can affect some fruit constituents. The ripeness of fruit will obviously also affect the proportions of the various carbohydrates. (*See* **Fruits of Temperate Climates**: Factors Affecting Quality; **Ripening of Fruit**.)

Table 1 Approximate chemical composition of 'Hayward' (*Actinidia deliciosa*) kiwifruit

Proximates	*g 100 g^{-1} fresh weight edible portion*
Total solids	15–20
Protein	1
Lipids	0.5
Carbohydrate	15
Energy	60 kcal
	250 kJ
Minerals	*mg 100 g^{-1} fresh weight*
Calcium	40
Chloride	35
Iron	0.4
Magnesium	25
Phosphorus	30
Potassium	300
Sodium	5
Vitamin C	80 mg 100 g^{-1} fresh weight

Proteins, Lipids, and Amino Acids

Kiwifruit, like most other fruit, contain insignificant amounts of proteins, lipids, and amino acids.

Carbohydrates

Ripe kiwifruit contain almost no starch as it has largely been converted during ripening into soluble sugars, mainly glucose, smaller quantities of fructose, and only minor amounts of sucrose. Carbohydrates in kiwifruit would normally meet only a very small part of the daily requirements. (*See* **Starch**: Structure, Properties, and Determination.)

Minerals

Kiwifruit are typical fruit, being a good source of potassium, with a high potassium-to-sodium ratio. They are also a useful source of magnesium, although a 100-g serving supplies less than 10% of the recommended dietary allowance (RDA) for an adult male. Other minerals are not sufficient to make a significant contribution to the diet. Refer to individual minerals.

Vitamin C

Vitamin C is currently recognized as the single most important nutrient in kiwifruit. 'Hayward' kiwifruit typically contain about 85 mg ascorbate $100\,g^{-1}$ fresh weight of edible portion. The vitamin C content can vary with fruit size, position on the vine, the season, and growing location. Many kiwifruit are eaten after having remained in cool store for long periods and it is therefore a great advantage that comparatively little vitamin C is lost from kiwifruit during storage or ripening. A 'Hayward' kiwifruit stored for 6 months at 0 °C and then ripened will still contain at least 90% of the vitamin C present in the fruit at harvest. Most other selections of *A. chinensis* or *A. deliciosa* likewise show little or no loss of vitamin C during storage over extended periods. It is therefore very likely that when consumers buy or actually eat kiwifruit, the fruit contain about as much vitamin C as they did at harvest. Assuming usual storage conditions and sensible handling, kiwifruit that are acceptable eating are also likely to be an excellent source of vitamin C.

Kiwifruit contain more vitamin C than almost all other fruit: on a fresh-weight basis they typically contain 50% more vitamin C than an orange, five or six times as much as a banana, and 10 times as much as an apple. Only a few readily available fruit, such as blackcurrants, are richer in vitamin C.

Recommended daily requirements (USA) for vitamin C range from 30 mg for a child to 90 mg for a lactating mother. These requirements can easily be met by two medium-sized peeled 'Hayward' kiwifruit. Any new selections of *A. chinensis* and *A. deliciosa* are likely to contain as much or even more, as 'Hayward' contains only relatively modest amounts of vitamin C compared to the rest of the genus *Actinidia*. 'Hort16A', at present the only commercially available fruit of *A. chinensis*, outside of China, usually contains 30–40% more vitamin C than 'Hayward' and many of the commonly grown cultivars in China contain at least twice as much. Even these levels are an order of magnitude less than those in fruit of *A. eriantha* Benth. or *A. latifolia* (Gardn. et Champ.) Merr. which can contain an astonishing 1 or 2% fresh weight of vitamin C. (*See* **Ascorbic Acid**: Properties and Determination; **Dietary Requirements of Adults**.)

Other Vitamins

One kiwifruit could provide 10–20% of the daily requirements for folic acid. Folic acid is readily available in a wide variety of foods and is especially rich in offal and raw leafy vegetables but easily destroyed by cooking. Fresh fruit are therefore a useful source. (*See* **Folic Acid**: Properties and Determination.)

Most foods contain vitamin E but it is seldom found in fruit or vegetables without significant fat content. It is likely, although not yet confirmed, that the vitamin E in kiwifruit is present in the seed and that these seeds will survive passage through the gut, thus making any vitamin E present unavailable. (*See* **Tocopherols**: Properties and Determination.)

The amounts of other vitamins are small in terms of human daily requirements and are similar to those found in many other fruit and vegetables. These vitamins are also available in a wide range of foods.

Pigments

Green-fleshed kiwifruit (e.g., *A. deliciosa* 'Hayward') are amongst the very few fruit that are still green when ripe. This color is due to the presence of chlorophyl and, although many other common fruit are green during the early stages of development, they lose their chlorophyl during subsequent maturation and ripening. The color of green kiwifruit is one of their most appealing characteristics. Chlorophyls, however, have very little direct nutritional value. (*See* **Chlorophyl**.)

Fruit of other kiwifruit cultivars or species may be yellow, orange, or red when ripe, and these colors are due to their content of carotenoids and anthocyanins. Carotenoids could contribute usefully to vitamin A requirements. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments.)

Oxalate

Kiwifruit contain appreciable but not exceptional amounts of oxalate. At least half of this oxalate is

complexed as highly insoluble calcium oxalate, mostly in the form of raphides (needle-like crystals), and this would render much of the calcium present in the fruit unavailable. The levels of oxalate in kiwifruit do not constitute a nutritional problem, assuming normal consumption, as some other common foods, e.g., spinach, contain much higher levels. Eating some processed 'Hayward' kiwifruit products such as nectars, dried slices, or fruit leathers, can cause irritation of the mucous membranes of the mouth. This is due, at least in part, to the mechanical irritation of the membranes by the oxalate raphides. In the fresh fruit, the raphides are embedded in mucilage and therefore do not cause any appreciable irritation. The shape of the raphides varies with *Actinidia* species and this could affect the amount of irritation caused.

Actinidin

Kiwifruit contain large amounts of the highly active proteolytic enzyme actinidin (EC 3.4.22.14) similar to the proteases found in other fruits such as pineapple, figs, or papaya. Actinidin is ideal as a meat tenderizer but can cause problems if fresh fruit are incorporated into gelatine-based jellies (which then do not set) or are mixed with dairy products.

Actinidin, at the levels found in 'Hayward' kiwifruit, does not seem to be a major health hazard for most people, but has been associated with allergic responses to kiwifruit. Actinidin can cause damage to the lips (especially at the corners of the mouth) but only if very large quantities of fruit are eaten. Peeling large numbers of fresh fruit can result in skin loss if hands are not protected. Other *Actinidia* cultivars need to be checked to insure that they do not contain appreciably higher levels of actinidin, capable of causing problems. There is no published evidence on the postulated beneficial role of actinidin in gut health.

Dietary Fiber

Kiwifruit contain about 2–3% dietary fiber due to pectins and other oligosaccharides and polysaccharides that are not broken down and absorbed in the small intestine. A 100-g serving of kiwifruit will therefore supply about 10% of the recommended daily requirement. The relative contributions of the various polysaccharide fractions in kiwifruit to dietary fiber effects such as stool-bulking and laxative action are not yet defined. (*See* **Dietary Fiber**: Properties and Sources.)

Laxatives

Kiwifruit are notorious for being laxative and individuals vary in their response. Consumption of large numbers of fresh fruit could therefore have an excessively vigorous purgative effect and the laxative effect can limit the number of kiwifruit some people are able to eat in any one day. Fresh kiwifruit or dried products are often used to maintain regularity in bowel movement, especially for older and sedentary people such as hospital patients. Although the laxative action is not in doubt, there is as yet no definitive information on the mode of action or the compounds responsible. (*See* **Elderly**: Nutritional Management of Geriatric Patients.)

Allergens

Kiwifruit contain allergens that can cause allergic responses in susceptible consumers. The response can include swelling and itching of the lips and tongue, swelling of the tongue and throat which can make breathing difficult, rashes, stomach pains and vomiting, itchy nose and eyes, or, at worst, anaphylactic shock requiring urgent medical attention. Fortunately, extreme allergic responses to kiwifruit are not frequent. Kiwifruit allergy is often cross-reactive with other common allergies such as to pollens, particularly birch pollen, latex, housedust mites, and some other tropical fruit. Actinidin, the most abundant protein in kiwifruit, has been postulated as a major cause of kiwifruit allergy but it is not known whether kiwifruit are the primary cause of the allergic reactions, or are secondary to exposure to other allergens. (*See* **Food Intolerance**: Food Allergies.)

Possible Medical Uses

Early Chinese pharmacopeia from the Tang Dynasty onwards list a variety of medicinal uses for kiwifruit or extracts from other parts of the plant. Various *Actinidia* species have long been used in traditional Chinese and Japanese medicine but their efficacy has yet to be fully confirmed. The anticancer and antimutagenic potential of the fruit are now being studied. The antioxidant capacity of kiwifruit likewise requires more study.

Myoinositol is a major carbohydrate in some *Actinidia* species and in ripe 'Hayward' fruit comprises about 6–7% of the free sugars. Myoinositol supplementation has been suggested for some medical conditions, but the supplementary amounts proposed would require consumption of 20 or more kiwifruit every day.

Handling, Storage, and Marketing

Actinidia chinensis and *A. deliciosa* flower in mid to late spring and the fruit are ready for harvest nearly 6 months later in late autumn. Harvest maturity is

taken as the stage at which fruit can be harvested from the vine and yet continue to develop to give an acceptable final eating quality which meets consumers' expectations. Fruit that are picked too early often have a poor color and flavor when ripened, a shorter storage life, and a shorter shelf-life when taken out of storage. However, fruit soften as they remain on the vine and if they are picked too late they may be not firm enough for handling, grading, and subsequent storage. Harvest maturity cannot be gaged by the external appearance of the fruit, and if there are no heavy frosts or bird attack, the fruit can hang on the vines until budbreak the following season.

Changes in the chemical composition of the fruit as they mature on the vine are therefore used to determine when fruit may be safely picked. Maturity values are taken on representative samples of fruit from vines of an individual maturity area, an area in which the vines are presumed to be similar through uniformity of age, management, and growing environment. The harvest maturity of 'Hayward' kiwifruit is usually assessed using the soluble solids content of the fruit as determined using a refractometer. As the fruit mature, the soluble solids content increases, largely as a result of the conversion of starch to sugars. In New Zealand, a minimum maturity index of 6.2 Brix is normally used following a standardized measurement procedure: in other countries, a minimum of 6.5 Brix has been set. Dry matter has also been used as an additional harvest index attribute. Other parameters may be more appropriate for different cultivars or different environments. The fruit of 'Hort16A', for example, are promoted in the marketplace for their yellow fruit flesh, but the chlorophyl in the pericarp is lost relatively late in the season and the harvest maturity indices currently recommended therefore include attributes such as soluble solids content (Brix), firmness, and hue angle.

Although kiwifruit are relatively firm when picked, they are still easily damaged if handled roughly. Fruit of 'Hort16A' are particularly prone to damage because their sharp 'beaks' can cause small wounds which allow the entry of fungal rots. Fruit are therefore picked by hand and grading equipment is designed to minimize injury.

Kiwifruit are seldom marketed immediately after harvest and most will be sold only after a period in storage. *Botrytis* stem-end rot was the cause of much of the losses of stored kiwifruit, but curing of the fruit can reduce the severity of rots. Fruit are not therefore cooled immediately after harvest but are left at ambient temperature for several days. Curing presumably allows the picking scar or wound to dry out, making entry of *Botrytis* less likely. Fruit may then be stored in bulk, either in conventional coolstores or in CA stores, or packed immediately into the consumer packs which are then stored until required. Traditionally, fruit were packed into wooden trays containing 3.6 kg of fruit, but a variety of different-sized packs is now common.

The aim in storage is to reduce fruit temperatures and respiration to slow the rate of ripening. This is achieved with kiwifruit by holding them at as low a temperature as possible ($0 \pm 0.5\,^{\circ}C$). Fruit will freeze at about $-1.5\,^{\circ}C$. The fruit may be stored for up to 6 or 7 months so to prevent the shriveling caused by water loss, fruit are held in their packs within plastic film liners and maintained at a relative humidity greater than 95%.

Kiwifruit are very sensitive to ethylene and even low levels of ethylene will promote ripening. It is therefore critical that kiwifruit not be exposed to ethylene from other fruits, e.g., apples, to ethylene from ripening kiwifruit, or ethylene produced in the exhausts of petrol-driven machinery. Cool stores are routinely flushed with fresh air and there is increasing use of ethylene scrubbers to remove any traces of ethylene produced. Although 'Hayward' kiwifruit survive lengthy periods in cool store, their storage life and their quality can be enhanced even further by CA storage and ethylene scrubbers are then essential. (*See* **Controlled-atmosphere Storage: Effects on Fruit and Vegetables**; **Ripening of Fruit**.)

Fruit ripened after a few weeks in storage seem to be of better quality than those ripened immediately after harvest, but during prolonged storage, the processes of ripening occur slowly over a long period: fruit soften and there may be some loss of eating quality. The intensity of the fruit flesh color may diminish and there is a loss of volatile flavor compounds.

Fruit that are newly harvested and stored for only a brief period may reach consumers before they are fully ripe, and an inexperienced customer trying to cope with a rockhard kiwifruit is unlikely to repurchase. About one-fifth of the New Zealand crop is therefore treated with ethylene either on board vessels or at the destination before being distributed to the markets. This insures that consumers are able to buy 'ready-to-eat' fruit. This does, of course, necessitate much more careful stock management.

'Hayward' (green) kiwifruit are now available in many markets throughout the year. For half the year, supplies will come from the big southern hemisphere producers, Chile or New Zealand, and for the other half of the year from northern hemisphere countries such as France, Greece, Italy, or the USA. New Zealand is so far the only large supplier of yellow kiwifruit to international trade, but as cultivation of

'Hort16A' or other *A. chinensis* cultivars starts in the northern hemisphere, they too will be available for much of the year.

'Hayward' kiwifruit store remarkably well. Most other cultivars of *A. chinensis or A. deliciosa* have shorter storage lives. Some other *Actinidia* species are being grown on a small scale, e.g., *A. arguta*, but their fruit generally have a much more restricted storage life.

Industrial Utilization

Processing of kiwifruit into high-quality products is a challenge, as much of the uniqueness of their subtle flavors is lost. Furthermore, if green kiwifruit, such as *A. deliciosa* 'Hayward', are processed, the bright-green color of the chlorophyl is usually modified to an unattractive, sludgy brown. Most juice or nectar products or canned kiwifruit products do not retain their green color unless frozen without heat treatment. Drying likewise usually results in a loss of flavor, green color, and vitamin C. It is therefore not surprising that in most countries kiwifruit are grown primarily for the fresh-fruit market and only fruit that do not reach export quality standards are processed. Extraction of chemical byproducts from waste fruit remains a possibility.

In China, processing of kiwifruit is relatively more important since storage and transport systems for fresh fruit are still being developed. Between 20 and 30% of all Chinese kiwifruit production is currently processed into products such as fruit juices, jams, wines, and spirits. Yellow-fleshed fruit of *A. chinensis* are often preferred for processing because they are sweeter and the color withstands processing better.

Kiwifruit in the Future

Kiwifruit are amongst the most recent of all domesticated fruits. *Actinidia deliciosa* has been cultivated for just on a century, *A. chinensis* for little more than 20 years. Although the kiwifruit currently available commercially throughout the world are promoted for their contribution to health, they probably do no more than set a minimum standard nutritionally for kiwifruit, even if these standards are higher than for most other fruits. The variation within the genus *Actinidia* indicates that it is realistic to predict the development of new kiwifruit cultivars with enhanced nutritional qualities but containing smaller amounts of the compounds having potentially adverse effects on human health.

See also: **Antioxidants**: Natural Antioxidants; **Ascorbic Acid**: Properties and Determination; **Chlorophyl**; **Controlled-atmosphere Storage**: Effects on Fruit and Vegetables; **Dietary Fiber**: Properties and Sources; **Dietary Requirements of Adults**; **Elderly**: Nutritional Management of Geriatric Patients; **Folic Acid**: Properties and Determination; **Food Intolerance**: Food Allergies; **Fruits of Temperate Climates**: Commercial and Dietary Importance; Fruits of the Ericacae; Factors Affecting Quality; **Ripening of Fruit**; **Starch**: Structure, Properties, and Determination; **Tocopherols**: Properties and Determination

Further Reading

Beever DJ and Hopkirk G (1990) Fruit development and fruit physiology. In: Warrington IJ and Weston GC (eds) *Kiwifruit: Science and Management*, pp. 97–126. Auckland, New Zealand: Ray Richards Publisher and NZ Society for Horticultural Science.

Ferguson AR (1984) Kiwifruit: a botanical review. *Horticultural Reviews* 6: 1–64.

Ferguson AR (1990) Kiwifruit (*Actinidia*) In: Moore JN and Ballington JR Jr (eds) *Genetic Resources of Temperate Fruit and Nut Crops*, pp. 603–653. Wageningen, The Netherlands: International Society for Horticultural Science.

Ferguson AR and MacRae EA (1991) Vitamin C in *Actinidia*. *Acta Horticulturae* 287: 481–487.

Huang HW, Gong JJ, Wang SM *et al.* (2000) Genetic diversity in genus *Actinidia*. In: Huang HW (ed.) *Advances in Actinidia Research*, pp. 65–79. Beijing, China: Science Press.

Perera CO, Young H and Beever DJ (1998) Kiwifruit. In: Shaw PE, Chan HT Jr and Nagy S (eds) *Tropical and Subtropical Fruits*, pp. 336–385. Auburndale, Florida, USA: AgScience.

United States Department of Agriculture (1999) *USDA Nutrient Database for Standard Reference*, no. 09405 for Kiwifruit (Chinese gooseberries), held in storage, raw. Release 13. http://www.purpleforest.net/Fruit/kiwistored.html.

Visser FR and Burrows JK (1983) *Composition of New Zealand Foods. 1. Characteristic Fruit and Vegetables*. DSIR Bulletin 235. Wellington, New Zealand: New Zealand Department of Scientific and Industrial Research.

Kosher *See* **Religious Customs and Nutrition**

Krebs Cycle *See* **Tricarboxylic Acid Cycle**; **Oxidative Phosphorylation**

Krill *See* **Shellfish**: Characteristics of Crustacea; Commercially Important Crustacea; Characteristics of Molluscs; Commercially Important Molluscs; Contamination and Spoilage of Molluscs and Crustaceans; Aquaculture of Commercially Important Molluscs and Crustaceans

KWASHIORKOR

C Leitzmann, Justus-Liebig University, Giessen, Germany

Background

Kwashiorkor is a nutritional disease, common in rural areas of developing countries, and it primarily affects children in the second year of life. This typical disease of the weaning period is associated with a diet based on starchy, low-protein staple foods, such as cassava, sweet potatoes, and bananas.

The term 'kwashiorkor' was used in the Gold Coast area of West Africa (present-day Ghana) and describes the 'disease of the weaned child when the next baby is born.' Cecily Williams introduced this term into modern medicine in 1933.

The realization of the close interactions between protein and energy deficiencies led to the introduction of the term 'protein-energy malnutrition' (PEM), which describes a range of clinical disorders in which marasmus and kwashiorkor are the two poles. Between these, there are many forms, with different clinical features, depending on the varying combination of deficiencies of protein, energy, and other nutrients, together with associated infections.

Etiology, Clinical Features, and Physiological Effects

Etiology

Although dietary protein deficiency plays an important role in the causation of kwashiorkor, this alone does not explain the etiology. Kwashiorkor is a complex multifactorial condition in which geographic, climatic, educational, infective, psychological, cultural, agronomical, nutritional, and other factors, frequently in combination, are involved. Kwashiorkor can be considered as the end point of a combination of these adverse factors, in which the body is not able to adapt to these changed conditions, and the characteristic clinical features develop. (*See* **Protein:** Deficiency.)

Clinical Features

The clinical picture can vary, depending on the degree of protein deficiency, accompanied vitamin or mineral deficiencies, age of onset and duration of deficiency, genetic and other factors. Figure 1 shows the typical signs of kwashiorkor.

Edema Edema is the cardinal sign of kwashiorkor and must be distinguished from the edema caused by other diseases such as hookworm infestation or beriberi. Edema causes tissue to swell and may be distributed over the whole body, but it usually appears first in the dorsum of the feet and lower legs. Later, it can also be detected in other parts of the body, including hands, forearms, back, upper extremities, and, in severe cases, in the face, especially the cheeks and around the eyes. Together with increased locally deposited fat in obese variants of kwashiorkor, the edema of the face results in the so-called 'moon-face.' The edema fluid generally represents 5–20% of the body weight. Ascites and other effusions into serous cavities are a late manifestation of edema and may be attributable to the presence of an infection.

Hair changes Kwashiorkor can be accompanied by striking dyspigmentation and changes in the texture of the hair, which can vary greatly from one part of the world to another. Thus, in African children, the dyspigmentation causes a change from black to red, blonde or light brown. In South American children, the 'flag sign,' an appearance of alternating streaks of

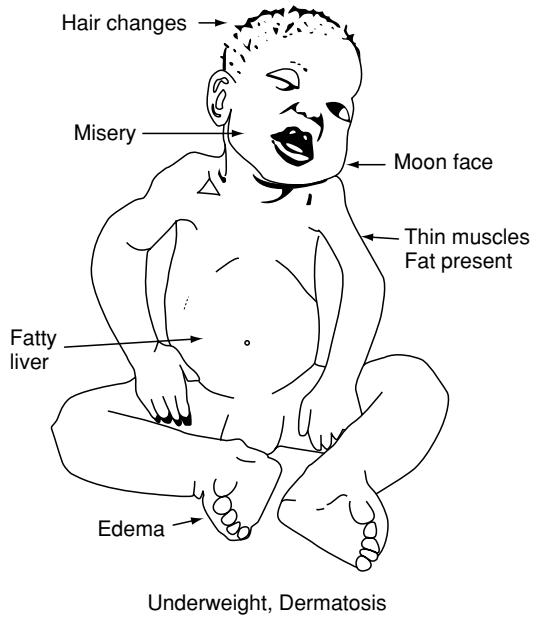

Figure 1 Clinical signs in kwashiorkor. (From Jelliffe DB and Jelliffe EF (1989) *Community Nutritional Assessment with Special Reference to Less Technically Developed Countries*, pp. 273–292. Oxford: Oxford University Press. modified.)

dyspigmented and normal black hair, prevails. The dyspigmented areas reflects periods when the child's nutrition was low in protein, 1–3 months before the color change appears. Nonnutritional environmental and genetic aspects that can also cause dyspigmentation should be excluded. Changes in the texture of the hair can be seen in African children. Curly hair becomes sparse and straight, and the bulbs of the hair roots are reduced in diameter, as a result of which, the hair is easily pluckable.

Dermatosis Dermatoses are characterized by areas of desquamation, ulceration and of both hypo- and hyperpigmentation of the skin. The lesions usually affect the lower limbs, buttocks, and perineum, but may be found anywhere on the body. Whereas the desquamated parts of the skin look as if there has been a burn and resemble 'crazy paving' thehyperpigmented areas resemble 'flaky paint.' Often dermatoses are complicated by associated skin infections.

Fatty liver In many, but not all, cases of kwashiorkor, the liver is considerably enlarged and can be palpated. The enlargement occurs mainly because of an infiltration with fat, especially triglycerides. This is the result of a decreased synthesis of β-lipoproteins, the specific transport agents for triglycerides. The fat accumulates in small droplets within the cells, first in the periphery of the lobules, and then spreads to the center of the lobules. In general, the liver functions are not altered. Severe liver failure is unusual, and cirrhosis only occurs in combination with any hepatic toxin, including aflatoxins. With recovery, the fatty liver is completely reversible. (*See* **Liver**: Nutritional Management of Liver and Biliary Disorders.)

Muscle wasting The muscles are often wasted, and the child regresses in their physical development. The infant becomes weak and is unable either to stand and walk or to hold up their head when lifted from a lying to a sitting position. In many cases, it is difficult to recognize the muscle dystrophy because of edema, which often conceals the wasting, but in the upper arms, where there is less edema, a reduction in circumference is evident.

Subcutaneous fat, which reflects the child's energy intake in the form of starchy food, and psychomotor changes are additional features. The child appears apathetic and miserable, with intermittent crying, and is often anorexic. In many cases, some degree of anemia is present and may be aggravated by malaria, hookworm infestation, or other parasites. Diarrhea is another common finding in kwashiorkor.

Physiological Effects

Kwashiorkor leads to changes in the organs and systems of the body and results in altered physiological functions.

Cardiovascular system Atrophy of the heart (which leads to reduced cardiac output), circulatory insufficiency (associated with prolonged circulation time) and bradycardia have been described. As a result, the extremities of kwashiorkor patients are cold and pale, and the pulse rate is low or even impalpable. Cardiac failure has been postulated as a possible cause of death in these children.

Endocrine organs and hormones The functions of the endocrine glands are not reduced in kwashiorkor. Secretion of human growth hormone from the pituitary appears to be normal or may be increased. Thyroid-stimulating hormone is also elevated. Owing to the prolonged half-life of cortisol, the plasma levels are elevated, with a catabolic effect on nutrients. Thyroid function seems to be normal, but reduced plasma concentrations of thyroxine-binding proteins and a reduced total plasma concentration of thyroxine have been observed. In

general, insulin secretion after oral glucose ingestion is impaired. (*See* **Hormones**: Thyroid Hormones.)

Digestive organs Protein depletion probably plays a predominant role in the pathogenesis of the mucosal changes. Cells of the intestinal mucosa and pancreas are atrophic. The mucous membranes of mouth and tongue are also affected. The synthesis and activity of digestive enzymes are reduced. Most affected are the disaccharidases, lactase, sucrase, and maltase. A reduced absorptive area together with diminished enzyme activities leads to impaired utilization of nutrients. In addition, lactose intolerance is particularly common.

Kidneys There is no specific structural or functional abnormality, but in a state of dehydration, the glomerular filtration rate is reduced. The impaired ability to concentrate urine may be attributable to accompanying electrolyte deficiencies, particularly to potassium. (*See* **Renal Function and Disorders**: Kidney: Structure and Function.)

Immunological system The immune system is still developing in the first 2 years of life and is very sensitive to nutrient deficiencies in early childhood. In kwashiorkor, the lymphoid tissue, primarily the thymus, is atrophic, and the immune response to infectious agents, particularly the cell-mediated immune response, is diminished. Thus, the child becomes more susceptible to infections such as measles or gastroenteritis. These infections themselves exhaust the nutrient stores.

Biochemical and Metabolic Disorders

There is a profound change in the body composition with the onset of kwashiorkor. Total body water increases from 60% to about 80% of the total bodyweight and may be comparable with that of a newborn infant. The high water content is mainly attributable to the reduced body fat stores and to muscle wasting.

The mineral content also changes. In particular, potassium decreases as a result of losses from diarrheal stools. Other minerals, such as sodium, calcium, phosphorus, and magnesium, are also diminished. individual minerals.

Total body protein is severely reduced by up to about 60% of the expected value in a normal child. This reduction is accompanied by a change of collagen protein from 27% in healthy children to about 40% of total body protein in children with PEM. Thus, the noncollagen protein is severely affected.

As a result of diets high in carbohydrates, children with kwashiorkor may still have subcutaneous and other fat stores.

General Metabolism

Protein metabolism In kwashiorkor, protein is adequately digested, and nitrogen retention is much more efficient, as can be seen in a fall of urinary nitrogen output. Whether or not total protein turnover is influenced by the level of protein intake, protein synthesis decreases in individual tissues such as muscle and skin, because they act as buffers in the adjustment of protein metabolism. (*See* **Protein**: Digestion and Absorption of Protein and Nitrogen Balance.)

The pattern of plasma amino acids shows characteristic changes. The concentration of essential amino acids, especially the branched-chain amino acids, is decreased. The plasma concentration of albumin is also decreased, since the rate of synthesis is very sensitive to reduced protein intake and falls immediately at the onset of dietary protein deficiency. The concentration of γ-globulins is usually normal and may even be elevated in the presence of infection. Other plasma proteins, such as transferrin and retinol-binding protein, are lowered and this explains the frequency of accompanying anemia and keratomalacia in kwashiorkor. Reduced concentrations of some plasma enzymes, including cholinesterase, alkaline phosphatase, amylase, and lipase, have also been reported.

Lipid metabolism As a result of decreased ability of the liver cells to synthesize lipoproteins, levels of plasma triglyceride and plasma cholesterol are low. Owing to increased lipolysis in the fatty tissue, the concentration of free fatty acids in the plasma tends to be high. (*See* **Fatty Acids**: Metabolism.)

Carbohydrate metabolism Hypoglycemia and glucose intolerance are frequent occurrences. Glucose intolerance is probably attributable to an impaired insulin release or a peripheral insensitivity to insulin and could predispose to diabetes in later life. (*See* **Carbohydrates**: Digestion, Absorption, and Metabolism; **Glucose**: Maintenance of Blood Glucose Level.)

Effect on Growth and Mental Development

Growth Failure

Compared with adults, young children need high nutrient supplies for rapid growth. Thus, growth failure can be understood as an early adaptation to

reduced dietary energy and protein intake. In contrast to mild forms of malnutrition, where stunted growth is characteristic, growth retardation in kwashiorkor is of less importance. In kwashiorkor, the restricted intake of nutrients together with acute infections, such as measles or gastroenteritis, leads to diminished body weight. Affected children have between 60 and 80% of the expected weight for their age. A decreased weight: height ratio, lowered mid-arm circumference and diminished skinfold thickness are also characteristic signs. (*See* **Growth and Development**.)

Mental Development

It is rather difficult to assess the effects of malnutrition in early childhood on brain and mental development because malnutrition frequently occurs in an unfavorable psychological environment, such as deprivation of parental and social care, with less mental stimulation of the children. Consequently, it becomes difficult to separate these effects from those of malnutrition or other factors.

The development of the human brain occurs mainly between midgestation and the second year of life. In this time, there is a rapid increase in the total cell number, which continues past the first year of postnatal life. The myelination also starts to increase before birth and continues until the second year of life. In this period, the brain development is vulnerable to nutrient deficiencies, which are demonstrated by a reduced number of brain cells in cerebrum, cerebellum, and brain stem, and reductions of total lipid, cholesterol, and phospholipid content. A smaller head circumference, reflecting reduced brain growth, has also been observed in undernourished children.

Although a correlation between anatomical changes of the brain and intellectual development is uncertain, many observations in developing countries have shown correlations between intelligence scores and malnutrition in the first 2 years of life. It seems that acute malnutrition such as kwashiorkor exerts less of an influence on intelligence than chronic malnutrition. The results are somewhat different, and the duration and time of malnutrition, as well as the psychological environment of the children, must also be considered.

Drug Metabolism

Malnutrition is frequently accompanied by infections and other complications, necessitating treatment with drugs such as antibiotics. Present knowledge about drug metabolism in undernourished organisms is limited. In kwashiorkor, the binding capacity for drugs such as salicylates, which are carried in the circulation bound to plasma proteins, is lowered, owing to decreased plasma albumin concentrations. Higher concentrations of free forms of the drug may be toxic. The ability of the liver to detoxicate drugs by oxidizing them may be impaired and can result in a prolonged half-life of the drug. For these reasons, all drugs should be used with caution.

Diagnosis, Treatment, and Prevention

Diagnosis

The diagnosis of kwashiorkor includes the above-mentioned clinical features and anthropometric measurements, such as weight between 60 and 80% of the standard, reduced mid-arm circumference, and triceps skin-fold thickness. A decreased serum albumin concentration below $35\,g\,l^{-1}$ represents the probability of protein deficiency. Other plasma proteins, such as retinol-binding protein, have been used in diagnosis, because they react more sensitively to protein deficiency. (*See* **Nutritional Assessment**: Anthropometry and Clinical Examination.)

Treatment

Children with severe kwashiorkor require treatment in hospital. Table 1 shows the major points of treatment. Since the main causes of death are dehydration and other complications, such as electrolyte disturbances, hypoglycemia, and various infections, the treatment has to start with fluid replacement and treatment of the various complications. After initial resuscitation, refeeding should be started with one- to two-hourly oral or nasogastric feeds of an electrolyte solution. On the second day, a three-hourly milk feed with added sugar (or any similar preparation) should be given. The amount of 90 ml per kilogram of body weight per day can be increased in the following few days. At the same time, the feeding frequency can be reduced to every 4 h. Once the child's appetite has improved, local staple food should be added to the meals to provide an extra source of energy, which can meet the need for catch-up growth.

Severely undernourished children are also likely to be deprived of vitamins (particularly vitamin A, D, folic acid, and some B-group vitamins) and trace minerals; supplements of these nutrients are needed. Refer to individual vitamins.

After the acute phase of illness, the children need several weeks of special feeding and some supervision until recovery is complete. For this period of

Table 1 Synopsis of the major points of treatment of severe PEM

Object of treatment	*Treatment*
Complications	
Dehydration	Plasma 20 ml kg^{-1} in hypovolemia. Half-strength Darrow's solution at 50–100 ml kg^{-1} over 6 h followed by 100–150 ml per kilogram per day for maintenance
Electrolyte disturbance	Potassium: 6 mmol per kilogram per day
	Calcium: 3 g per day
	Magnesium: 2–3 mmol per kilogram per day
Hypoglycemia	Dextrose, 50%, 1 ml kg^{-1} iv stat. and then 10% dextrose solution iv as maintenance
Infection	Antibiotics covering both Gram-negative and Gram-positive organisms, e.g., penicillin 25 000–50 000 units per kilogram per day in four divided doses, iv, im, or orally, and kanamycin 5–15 mg per kilogram per day in two to three divided doses, im
Dietary deficiencies	Initially frequent small feeds, e.g., half-strength Darrow's solution in a dose of 90 ml per kilogram per day in eight divided feeds for 24 h, then full strength milk (90 ml per kilogram per day) increasing to 150 ml per kilogram per day over 2–3 days
	Cereals and other foods are added once improvement occurs
	Vitamin A: 33 000 μg im on admission and then 5000 μg orally per day
	Vitamin K: 5 mg iv or im on admission
	Vitamin D: 10 μg per day
	Ferrous sulfate: 5 mg per kilogram per day once improvement in general condition has occurred
	Folic acid: 5 mg per day for 10 days

iv, intravenous; im, intramuscular.
Source: From Hansen JDL *et al.* (1982).

treatment, nutrition rehabilitation units, where recovery of children is combined with nutrition education for the mothers, are most suitable. Mothers are taught the use of local foods, cooking, feeding practice, and preparation of weaning diets that are more suitable for the children. Moreover, mothers receive instructions for dealing with simple illnesses such as diarrhea. This may help to prevent children who are cured from going back to the conditions that precipitated the malnutrition.

Prevention

In view of its multifactorial causation, the prevention of kwashiorkor cannot be achieved through medical or nutritional means alone. Political, economical, and agricultural measures are also necessary. From the nutritional point of view, the management of weaning is most important. Mothers need educational services in order to understand the special dietary needs of the child at weaning. They should be educated to prepare weaning diets with higher protein and energy contents. (*See* **Infants:** Weaning.)

Owing to lack of time, children are fed only once or twice a day with bulky foods, frequently late at night, when they are too tired to eat. Thus, it is impossible for children to receive their nutrient needs; it is necessary to feed children at least four times a day.

Prognosis

The prognosis of severe kwashiorkor depends on the particular living conditions, including hygiene, which can lead to severe infections, on complications such as cardiac failure, and on the facilities available to treat the sick child. In addition, the conditions to which the child must return after treatment are very important. Prognosis is often unfavorable, with mortality rates between 10 and 25% for severe kwashiorkor cases. (*See* **Malnutrition:** The Problem of Malnutrition; **Marasmus**.)

See also: **Carbohydrates**: Digestion, Absorption, and Metabolism; **Fatty Acids**: Metabolism; **Glucose**: Maintenance of Blood Glucose Level; **Growth and Development**; **Hormones**: Thyroid Hormones; **Infants**: Weaning; **Liver**: Nutritional Management of Liver and Biliary Disorders; **Malnutrition**: The Problem of Malnutrition; Malnutrition in Developed Countries; **Marasmus**; **Nutritional Assessment**: Anthropometry and Clinical Examination; **Protein**: Digestion and Absorption of Protein and Nitrogen Balance; Deficiency; **Renal Function and Disorders**: Kidney: Structure and Function

Further Reading

Albrecht R and Pélissier MA (1995) About the oxidative stress status in children with kwashiorkor. *Food Chemistry and Toxicology* 33: 1081–1083.

Elmadfa I and Leitzmann C (1998) *Ernährung des Menschen*, 3rd edn. Stuttgart: Verlag Ulmer.

Hansen JDL, Buchanan N and Pettifor JM (1982) Protein energy malnutrition (PEM). In: McLaren DS and Burman D (eds) *Textbook of Paediatric Nutrition*,

2nd edn, pp. 114–142. Edinburgh: Churchill Livingstone.

Jackson AA (1990) The aetiology of kwashiorkor. In: Harrison GA and Waterlow JC (eds) *Diet and Disease in Traditional Societies*, pp. 76–113. Cambridge: Cambridge University Press.

Jelliffe DB and Jelliffe EF (1989) *Community Nutritional Assessment with Special Reference to Less Technically Developed Countries*, pp. 273–292. Oxford: Oxford University Press.

King FS and Burgess A (1993) *Nutrition for Developing Countries*, 2nd edn. Oxford: Oxford University Press.

McLaren DS (1982) Protein energy malnutrition (PEM). In: McLaren DS and Burmann D (eds) *Textbook of Paediatric Nutrition*, 2nd edn, pp. 103–113. Edinburgh: Churchill Livingstone.

Passmore R and Eastwood MA (1986) *Human Nutrition and Dietetics*, 8th edn, pp. 279–291, 516–545. Edinburgh: Churchill Livingstone.

Poskitt EME (1988) *Practical Paediatric Nutrition*, pp. 96–113. Stoneham, MA: Butterworth.

Suskind RM and Lewinter-Suskind L (1990) *The Malnourished Child, Nestlé Nutrition Workshop Series 19*. New York: Raven Press.

Waterlow JC (1992) *Protein Energy Malnutrition*, 2nd edn. London: Edward Arnold.

L

Labeling (Labelling) *See* **Food Labeling (Labelling)**: Applications

LABORATORY MANAGEMENT

Contents

Chemical Safety

G Collier, University of Hull, Hull, UK

Background

There are several obvious prerequisites for safe working. These include well-designed buildings, equipment suited to the intended purposes, and carefully planned operating systems that are clearly defined and effectively administered. Moreover, the choice and use of buildings, equipment and systems of work should be strongly influenced by a clear, progressive, and well-publicized safety policy. In addition to defining aims and objectives, the policy statement should specify the persons to whom particular safety responsibilities are delegated; the extent of those responsibilities; the arrangements for advice and instruction on safety matters; and the procedures for reporting defects and making complaints.

It is widely recognized that the same principles that have been adopted within many organizations with respect to total quality management (TQM) are equally applicable to the management of health and safety. Organizations that have applied this approach to their health and safety management systems, often as part of a TQM philosophy, generally achieve high standards of health and safety performance.

The TQM approach seeks to promote continuous improvement in all aspects of an organization's activities. In health and safety terms, the ultimate goal is a harm-free environment. Such an objective is clearly desirable, even in financial terms. Since the staff of any organization is invariably its most valuable asset, ill health and accidents have appreciable financial and operational consequences. An effective system for managing health and safety will reduce what is known, in quality terms, as the cost of nonconformance.

The following paragraphs highlight certain aspects of chemical safety that are of particular relevance.

The Control of Substances Hazardous to Health (COSHH)

Against a background of increasing awareness of a need to provide a general legal foundation for good occupational hygiene practices, the European Commission adopted a Framework Directive (80/1107/EEC) on the protection of workers against chemical, physical and biological hazards. In the UK, this led to the Control of Substances Hazardous to Health (COSHH) Regulations, which may be considered to represent best current practice.

Under these Regulations, no employer may carry out any work that is liable to expose employees to any substance hazardous to health unless he or she has made a suitable and sufficient assessment of the risks created by that work to the health of those employees,

and of the steps that need to be taken to prevent or control exposure. The Regulations are wide-ranging, not least in their definitions of substances hazardous to health.

In practice, there is no single simple way of defining what is, or is not, a hazardous substance. The hazard is the potential of the substance to cause harm to health. Whether that potential is realized, i.e., whether, in other words, there is a risk or danger to health, will depend upon a number of factors: the amount of substance used; its physical state; the likelihood of exposure and the method of entry into the body. All these circumstances need to be taken into account. The hazard, or potential for harm, arises from the intrinsic properties of the substance and, depending upon what those properties are, there may be hazard of a high or low order.

The COSHH Regulations define several categories of hazardous substance as follows:

1. Substances that have been formally designated as very toxic, toxic, harmful, corrosive, or irritant.
2. Substances to which an occupational exposure standard (OES) or maximum exposure limit (MEL) has been assigned. An OES is the concentration of an airborne substance, averaged over a reference period, at which, according to current knowledge, there is no evidence that it is likely to be injurious to personnel if they are exposed by inhalation, day after day, to that concentration. A MEL, however, is a strictly applied limit. It is the maximum concentration, averaged over a reference period, to which personnel may be legally exposed under any circumstances. Furthermore, employers have a duty to ensure that exposure is kept as far below the MEL as is reasonably practicable. So far, MELs have been assigned only to substances that have been generally perceived to present a high level of hazard, including a number of carcinogens, mutagens, and teratogens.
3. Biological agents.
4. Dust, of any kind, in substantial concentrations (see below).
5. Other substances with similar hazards to those in the previous categories.

The risk assessment that is required must be undertaken by a competent person. It involves consideration of all the types of substances to which an employee is likely to be exposed and the effects that such substances can have on the body. The nature and extent of exposure must also be assessed, taking due account of reasonably foreseeable deterioration in, or failure of, control measures. Although inhalation is the main route of entry into the body, some substances may be readily absorbed through the skin, and this needs to be taken into account when designing control measures.

There is a firm requirement to ensure that exposure is *prevented* wherever reasonably practicable. Only where this is not reasonably practicable (i.e., when it cannot be achieved without completely disproportionate expense or trouble) is adequate control an acceptable alternative. Nevertheless, the opportunities for prevention of exposure are often limited, requiring either elimination of the use of the substance altogether or substitution by a less hazardous material.

In most circumstances, effective control is the only practicable option, and this is mainly achieved by enclosure, total or partial, and local exhaust ventilation. The use of personal protective equipment is only acceptable in a minority of situations where adequate control cannot be achieved by other means.

As soon as control measures are introduced, a system of regular maintenance, examination, and testing must be instituted, and the results must be recorded, together with details of any repairs or other remedial work. Defects must be promptly reported by employees, and remedial action must be taken as soon as possible.

Once effective and well-maintained control measures are in place, that is by no means the end of the matter. The TQM approach requires constant feedback to improve performance, and three further lines of action are appropriate.

Where it is necessary to ensure the maintenance of adequate control, a procedure for monitoring for hazardous substances is required. The use of certain specified substances renders this mandatory. Usually, monitoring is not just a matter of atmospheric sampling but requires repeated or continuous measurements in an individual's breathing zone.

Health surveillance may also be called for in certain circumstances, and the results, like those of monitoring exercises, must be recorded. When an individual leaves an employer's service, their health surveillance record should be passed to their new employer.

Above all, there is a requirement to provide continual information, instruction, and training to employees. Success in meeting this requirement is clearly crucial to the success of all other measures. In this connection, one of the notable developments of the last few years has been the ever more stringent requirement for the inclusion of chemical hazard information on labels and packaging.

Sample Handling

Samples of material are presented to the laboratory in many different forms. When submitted for quality assurance or other routine testing, the nature of the

material may be known in some detail. In all other circumstances, the material should be treated as potentially harmful, and appropriate measures should be adopted to control exposure of personnel.

Samples must be adequately labeled, and laboratory workers should always ensure that they have the maximum amount of relevant information before handling the material. Many of the comments made in the following paragraphs in respect of dusts, solvents, and hazardous reagents are equally applicable to materials submitted to the laboratory as samples for examination or other experimental work.

Dusts

Finely powdered materials and dusts can be hazardous to health when present in the environment in substantial concentrations. That which can be said to constitute a substantial concentration is, to some extent, dependent on particle size. This, in turn, governs the aerodynamic diameter and drag, the degree of penetration into the respiratory system, and the region of deposition.

Total inhalable dust (i.e., airborne material that enters the nose and mouth during breathing) includes particles up to 50 μm in diameter. The larger particles are commonly called 'nuisance dusts,' and the body usually expels them by coughing or sneezing. Total inhalable dust of any kind, present at concentrations of 10 mg m^{-3} or greater as an 8-h time-weighted average, is regarded as hazardous to health.

The term 'respirable dusts' is applicable to smaller particles (of average diameter 7 μm or less), which are much more harmful because they can penetrate deep into the respiratory system, reaching the gas-exchange regions of the lungs.

As a general guide:

- particles of diameter > 7 μm will deposit in the mouth and throat;
- particles of diameter 4.7–7 μm will deposit in the pharynx;
- particles of diameter 3.3–4.7 μm will deposit in the trachea and bronchi;
- particles of diameter 1.1–2.1 μm will deposit in the terminal bronchi;
- particles of diameter about 1 μm and less will deposit in the alveoli.

Respirable dusts include grain dust, which may be defined as dust arising from the harvesting, drying, handling, storage, or processing of barley, wheat, oats, maize, and rye, including contaminants. In the UK, grain dust has been assigned a MEL of 10 mg m^{-3} as an 8-h time-weighted average respirable dust of any kind, present at concentrations of 4 mg m^{-3} or greater as an 8-h time-weighted average, is regarded as hazardous to health (see COSHH, 1999). Atmospheric concentrations are determined gravimetrically.

Occupational exposure to respirable dusts can cause asthma and other allergic lung conditions as well as a variety of recognized dust diseases. Such dusts should therefore be considered as respiratory sensitizers, which, when inhaled, can trigger irreversible allergic reactions in the respiratory system. Once this has occurred, subsequent exposure, even to minute amounts, may cause respiratory illness including asthma, rhinitis and extrinsic allergic alveolitis. If exposure is allowed to continue, symptoms are likely to become progressively worse and may result in chronic disease, which may become so severe as to threaten life. The earlier a sensitized person is removed from exposure, the greater the probability of avoiding serious damage to health.

The highest priority should be accorded to prevention of exposure. If this is not reasonably practicable, then effective control measures, such as partial enclosure and local exhaust ventilation, must be instituted. Dust respirators or masks should be used only for short-term tasks where satisfactory dust control cannot be achieved by other means.

Solvents and Toxic Reagents

Many organic solvents present a fire and explosion hazard because of their low flash points. The magnitude of the hazard is dependent upon the quantity of the solvent involved and on other physicochemical properties such as autoignition temperature, explosive limits, vapor density and ability to accumulate electrostatic charge.

Fire or explosion may occur when flammable vapors are mixed with oxygen or air in proportions that lie between certain critical values known as the upper and lower explosive limits (UEL and LEL). For many common solvents, the LEL is only a few per cent in air, and good ventilation is essential to prevent the formation of an explosive atmosphere when significant quantities are used. As most organic solvent vapors are denser than air, low-level ventilation may be required for larger-scale operations in order to avoid dangerous accumulation of vapors at floor level.

In this connection, it is helpful to have some appreciation of the percentage loss of a solvent, by evaporation, when it is used in various laboratory operations. This may range from a loss of about 1% or less of the liquid for a simple decantation, to around 20% or more for hot filtration, but the precise values vary considerably from one solvent to another. Nevertheless, the larger the volume of vapor

liberated, the greater the chance of ignition by a remote heat source.

Ignition may also occur *without* the obvious presence of a high-temperature source. Liquids flowing in pipes can readily generate static electricity, which may result in a spark. It is therefore advisable to dispense large volumes via metal pipes that have been doubly earthed.

Solvents with a low autoignition temperature (e.g., carbon disulfide (100 °C) and diethyl ether (180 °C)) are particularly hazardous. In such cases, the presence of a hot plate in contact with a vapor/air mixture is quite capable of causing violent ignition.

Storage of highly flammable liquids must be in accordance with statutory requirements, which generally limit the total volume to about 50 l within any given working area. The capacity of individual bench bottles must not exceed 500 ml. Many disastrous fires have resulted from inadvertent spillage, and places in which substantial volumes of flammable liquids are used should have clearly defined emergency procedures to ensure rapid evacuation and ventilation of the affected area.

Certain solvents (alcohols and, especially, ethers) have a propensity for peroxide formation on storage, and these may detonate on distillation or on evaporation to low bulk. Such solvents should not be stored for extended periods, once opened. Bottled ether waste should have 100 ml of saturated ferrous sulfate solution added to each 2.5 l before disposal.

Owing to their volatility and the large volumes in common use, many solvents present significant health risks in addition to any fire and explosion hazard. Long-term effects of exposure to high concentrations may include dermatitis, damage to the central nervous system, the liver, the kidneys or the blood, and even cancers. Smoking in the vicinity of chlorinated solvent vapors is particularly hazardous, not because of any fire hazard but because of toxic breakdown products that can be produced by pyrolysis.

Many laboratory reagents in common use are also toxic to some degree and therefore hazardous. The term 'toxic' does not have an absolute quality of the kind that tends to be assumed in common parlance. Almost all chemicals can cause harm in certain circumstances, and, conversely, even highly toxic chemicals can be used safely if adequate precautions and control measures are taken. Most toxicological studies have been undertaken with laboratory animals, and not every species shows the same degree or type of response. Extrapolation of data to the human species is therefore fraught with difficulty. Nevertheless, toxicological data from studies of other mammals can normally be taken as a useful indication of the likely degree of toxic hazard to human beings, and reference works such as the Registry of Toxic Effects of Chemical Substances (NIOSH) can be valuable sources of information.

In toxicology experiments, increasing doses of the test substance are administered to laboratory animals, and the response is observed. The dose level that results in the death of half of the group of animals receiving it is known as the LD_{50} value (a lethal dose for 50%) and is usually expressed as milligrams of the substance per kilogram of body weight of the animal ($mg\,kg^{-1}$). As a rough guide, for oral ingestion by the rat, substances with LD_{50} values lying between 200 and 2000 $mg\,kg^{-1}$ should be regarded as harmful, and those with values between about 25 and 200 $mg\,kg^{-1}$, as toxic. Any substance with an LD_{50} of less than 25 $mg\,kg^{-1}$ should be treated as very toxic. For substances on which extensive toxicological data already exist (by far the minority), formal assignments to one of these categories may already have been made.

Consideration of the toxicity levels of the substances involved, together with the respective amounts of the substances to be used, will be an important aspect of the risk assessment for any operation involving hazardous substances. Such considerations will influence the conditions under which the operation is carried out, whether partial containment in a fume cupboard is necessary, and what degree of personal protection is required.

Many operations are carried out in solution, and in such cases, the solvent may present a greater hazard than the reagents. Very often, however, the fact that the reagent is present in the solution state, together with a solvent that may greatly facilitate absorption through the skin, considerably increases the risks associated with its use.

Waste Disposal

All producers of waste, of whatever kind, have a duty of care to ensure that it is safely disposed of, without risk to others and, as far as possible, without detriment to the environment. The welter of environmental protection legislation throughout the developed world is intended to ensure that this duty is responsibly discharged.

Laboratory waste is regarded as especially hazardous and is usually subject to special controls. Until relatively recently, it was quite common practice to dispose of all aqueous waste down the drains, but this is no longer acceptable, as it can have adverse consequences for the public sewers to which most laboratory drains are connected. In order to avoid causing problems with sewage treatment, no effluent outside the pH range of 5–9 should be discharged into the

public drainage system. It is therefore necessary to neutralize or massively dilute any acid or alkaline wastes before they can be disposed of by such a route.

Similar dilution is also required before even small quantities of water-miscible solvents such as acetone or ethanol can be disposed of down the drain. Accumulation of solvent vapors in sewers would pose a considerable risk to maintenance workers and, indeed, to the system itself. Consequently, only very limited use can be made of the drains for disposal of liquid waste, and it is necessary to ensure that any waste so consigned does not contain significant concentrations of heavy metals or other toxic substances.

In considering discharge to the local environment, discharge of vapors and fumes from local exhaust ventilation and fume cupboard outlets must not be overlooked. Adequate dilution must occur before the fumes are discharged, and, if possible, noxious materials should be trapped at source rather than simply vented to the atmosphere.

Most laboratory waste is now destined either for high-temperature incineration or for burial at a duly licensed landfill site, and the producer of the waste is deemed to be responsible for it right up to the moment of its ultimate disposal. This gives rise to the so-called 'cradle to grave' view of the producer's responsibility.

The onus upon the producer to ensure proper disposal is, in any case, a heavy one. Suitable and separate containers should be available in the laboratory for the collection and segregation of various types of waste. In particular, flammable solvents should be segregated from nonflammables. Halogenated solvents should be collected separately, as far as possible, and must not be mixed with ketones. In general, care must be taken not to bulk waste, especially solvent waste, in such a manner as to create additional hazards by the mixing of incompatible materials.

Producers must ensure that laboratory waste is clearly and adequately labeled as to its nature and hazardous properties, securely packaged, and disposed of via an authorized contractor. Moreover, they should satisfy themselves as to the methods to be employed by the disposer, and this may involve visits to the treatment plant or landfill site. Some types of waste may require separate secondary containment before packaging for transportation, especially highly reactive materials, oxidizing agents and foul-smelling substances. Others may need to be segregated simply because of a high toxicity hazard. This certainly applies to carcinogens, and conspicuous labeling is essential in such cases.

Finally, detailed written records of all waste disposals should be made and retained for at least two years.

See also: **Effluents from Food Processing**: Disposal of Waste Water; **European Union**: European Food Law Harmonization; **Quality Assurance and Quality Control**

Further Reading

COSHH (1999) *Control of Substances Hazardous to Health (Approved Codes of Practice) – Control of Substances Hazardous to Health Regulations 1999*. London: HMSO.

Health and Safety Executive (1997) *Successful Health and Safety Management*, 2nd edn. London: HSE Books.

National Institute for Occupational Health and Safety (2001) *Registry of Toxic Effects of Chemical Substances*. NIOSH: Cincinnati, OH.

Royal Society of Chemistry (1992) In: Luxon SG (ed.) *Hazards in the Chemical Laboratory*, 5th edn. London: Royal Society of Chemistry.

Royal Society of Chemistry (1989) *Safe Practices in Chemical Laboratories*. London: Royal Society of Chemistry.

Microbiological Safety

V N Scott and I Walls, National Food Processors Association, Washington DC, USA

Introduction

Food laboratories may handle potentially infectious materials, e.g., foods contaminated with pathogenic microoganisms such as *Salmonella* spp., *Escherichia coli* O157:H7, *Listeria monocytogenes*, *Shigella*, and even *Clostridium botulinum*. Numerous published reports have documented laboratory transmission of infectious agents, including those from *E. coli* O157:H7, *Salmonella*, and *Shigella*, although infection occurs primarily in clinical laboratories. The incidence of laboratory transmission of foodborne pathogens is not known but is believed to be very low. Nevertheless, the risk of exposure exists; this risk can be reduced to a very low level by use of good laboratory practices and containment techniques (biosafety level 2), as described in this chapter. Interestingly, laboratory-acquired infections with *Cryptosporidium parvum* have been reported in almost every laboratory working with this parasite; therefore, it is important to adhere strictly to all precautions to prevent transmission of this agent. Good laboratory practices are particularly important when it is recognized that fewer than 20% of

laboratory-associated infections are attributed to a known accident.

The most frequently recognized causes of laboratory infections are oral aspiration due to mouth pipetting, inoculation with syringes and needles, animal bites, spray from syringes, and centrifuge accidents. Other recognized causes include cuts from contaminated glassware and the spilling of pathogenic cultures. Exposure to infectious aerosols is considered to be the most likely source for cases not associated with a known accident.

Biosafety practices relating to sample handling, containment of biological hazards, disinfection and decontamination in the laboratory and personnel practices will be covered in this article. A list summarizing microbiological safety practices can be found at the end. These guidelines should be considered minimal guidelines. They should be customized for individual laboratories after consulting additional references.

Biosafety Levels

There are four biosafety levels for working with microorganisms in the laboratory. These consist of a combination of laboratory practices and techniques, safety equipment, and laboratory facilities that provide increased protection to personnel, the environment, and the community as the level increases. Biosafety level 1 is suitable for work involving well-characterized agents not known to cause illness consistently in healthy persons; work is generally conducted on open benchtops using standard microbiological practices. Biosafety level 2 is suitable for work involving agents of moderate potential hazard to personnel and the environment. The primary differences between biosafety level 1 and 2 is that, with the latter, extreme precautions are taken with contaminated sharp items and certain procedures that may create aerosols are conducted in biological safety cabinets or other physical containment equipment. Biosafety level 3 is applicable to laboratories working with agents that may cause serious disease via the inhalation route, and biosafety level 4 is required for work with dangerous agents that pose a high risk of aerosol-transmitted laboratory infections. Currently all microorganisms expected to be encountered in a food laboratory can be handled with biosafety level 2 in the USA. Some countries may require biosafety level 3 precautions when working with *E. coli* O157:H7. In addition, biosafety level 3 precautions are indicated for *Clostridium botulinum* when activities pose a high potential for aerosol or droplet production and for those activities involving large quantities of toxin.

Sample Handling – Hazards

Laboratory personnel can be exposed to infectious agents through a number of routes (Table 1).

Inhalation Hazards

The generation of droplets and aerosols is considered to be the primary means of laboratory exposure to infectious agents. All laboratory procedures should be performed carefully to prevent the creation of aerosols and droplets. In most instances use of proper containers and equipment will minimize the chance for aerosolization of infectious materials. However, even trivial laboratory operations generate droplets and aerosols: a drop falling off the end of a pipette, streaking the surface of a rough agar plate with a loop, inserting a hot loop into a culture, pulling a cotton plug from a flask or a tube, and many others.

Pipettes should be drained gently with the tip against the inner wall of the receiving tube, bottle, or flask. The liquid should not be expelled forcibly from the pipette. Material should be centrifuged in sealed tubes or using centrifuges with sealed buckets, safety trunnion cups, or sealed heads to prevent the escape of aerosols and material being centrifuged. Only sealed tubes should be used for vortex mixing. Blenders should have leak-proof bearings and tight-fitting gasketed lids. Sonication or homogenization of infectious materials should be done inside a safety cabinet.

Molds pose a potential inhalation hazard, as spores may cause irritation. Moldy food should be sampled with care, to minimize the potential for laboratory contamination. When molds are grown on agar, plates should not be inverted, as mold spores may fall on to the lid of the Petri dish, and if it is opened these can contaminate the laboratory.

Table 1 Routes of exposure of laboratory personnel to hazardous agents

Route	*Procedures resulting in exposure*
Contact	Decanting liquids, pipetting, removal of screw caps, vortex mixing, unsealed containers (spills)
Oral	Mouth pipetting, not washing hands after working with infectious materials
Ocular	Procedures resulting in splashing, aerosol generation (sonication, centrifugation, homogenization)
Inoculation	Use of needles, syringes
Inhalation	Procedures generating aerosols (sonication, centrifugation, blowing out pipettes, heating inoculation loops; streaking inoculum on agar)

Contact Exposure and Inoculation Hazards

Infectious materials should be handled very carefully, especially during sample transfers and decanting of supernatants, to avoid spills. Safety glasses should be used to protect against splashing infectious material into the eyes.

Because of the hazard of self-inoculation, the use of needles and syringes should be restricted to procedures for which there are no alternatives. Needle-locking syringes or disposable syringe-needle units in which the needle is integral to the syringe should be used. Needles should not be replaced in the guard following use, nor should they be clipped.

Care should be taken to avoid cuts when handling sharp objects such as scalpels or broken glass. Broken glass should not be picked up by hand; tongs, forceps, or a brush and dustpan should be used. Dispose of sharp objects in proper receptacles. In laboratories working with animals, training should be given in animal-handling techniques to avoid bites and scratches. Where appropriate, gloves should be worn.

Ingestion Hazards

Mouth pipetting poses such a high risk of exposure that good laboratory practices prohibit this procedure. Only mechanical pipetting devices should be used in the laboratory.

Additional ingestion hazards result from hand contamination and subsequent transfer to the mouth; this can be avoided by appropriate personnel practices such as proper hand washing and wearing gloves (see section on personnel practices and personal hygiene).

Containment – The Biological Safety Cabinet

The most important element of containment is strict adherence to good microbiological practices and techniques. However, appropriate laboratory design and safety equipment are also required for proper containment of microbiological hazards.

The use of proper containment equipment helps minimize the exposure of laboratory personnel to infectious agents. Equipment designed to avoid laboratory hazards includes biological safety cabinets, mechanical pipetting devices, and centrifuges and blenders. Containment is also achieved by use of appropriate closures on laboratory glassware such as test tubes, flasks, and bottles.

The biological safety cabinet is the principal device to provide containment of aerosols. There are three types – class I, II, and III. Class I and II cabinets are generally open-fronted and depend on air flow to protect laboratory personnel.

The class I cabinet (**Figure 1a**) is designed to protect the user; air enters the front work opening and exits at the top through a high-efficiency particulate air (HEPA) filter. This provides no protection against contamination of the work materials, since unfiltered ambient room air flows into the cabinet. A class I cabinet should not be confused with a horizontal laminar flow cabinet in which HEPA-filtered air is blown across the work surface toward the researcher to protect the research/materials against contamination. A horizontal laminar flow cabinet should never be used with infectious agents due to potential worker exposure.

Class II biological safety cabinets (**Figure 1b**) protect the research materials from contamination, as well as the worker, since both the supply and the exhaust air are HEPA-filtered. Room air is drawn through the front work opening into a plenum below the work area by a fan. Air exiting the fan is passed upwards through a channel in the rear into a space between two HEPA filters. A portion of the air is exhausted through one HEPA filter while the rest of the air passes through the second HEPA filter and is directed downward as clean, laminar flow air to the work surfaces. In subclass IIA cabinets 30% of the air is exhausted and 70% recirculated. In subclass IIB cabinets a smaller portion of air (30%) is recirculated or there is total exhaust (no recirculation). In the latter case room air entering through the front opening to protect the researcher against escaping pathogens is drawn into the plenum and is exhausted.

The class III cabinet is a totally enclosed, airtight workspace equipped with protective gloves, and hence is often referred to as a glove box. Such cabinets are used for high-risk biological agents and are not necessary in food microbiology laboratories, although they are sometimes used for work with *Clostridium botulinum*.

Biosafety cabinets should be located in a low-traffic-area free from drafts. The cabinet must be tested and certified at installation, after moving, and annually thereafter. It is desirable that safety cabinets exhaust outdoors since materials such as formaldehyde (used for decontamination), radioisotopes, or chemical carcinogens should not be exhausted indoors. When cabinets are exhausted to the roof the exhaust stack height should insure that discharge occurs above the heads of maintenance personnel. The exhaust stack should not exhaust in close proximity to the air intake. HEPA filters should be disposed of as contaminated biological waste, preferably by incineration.

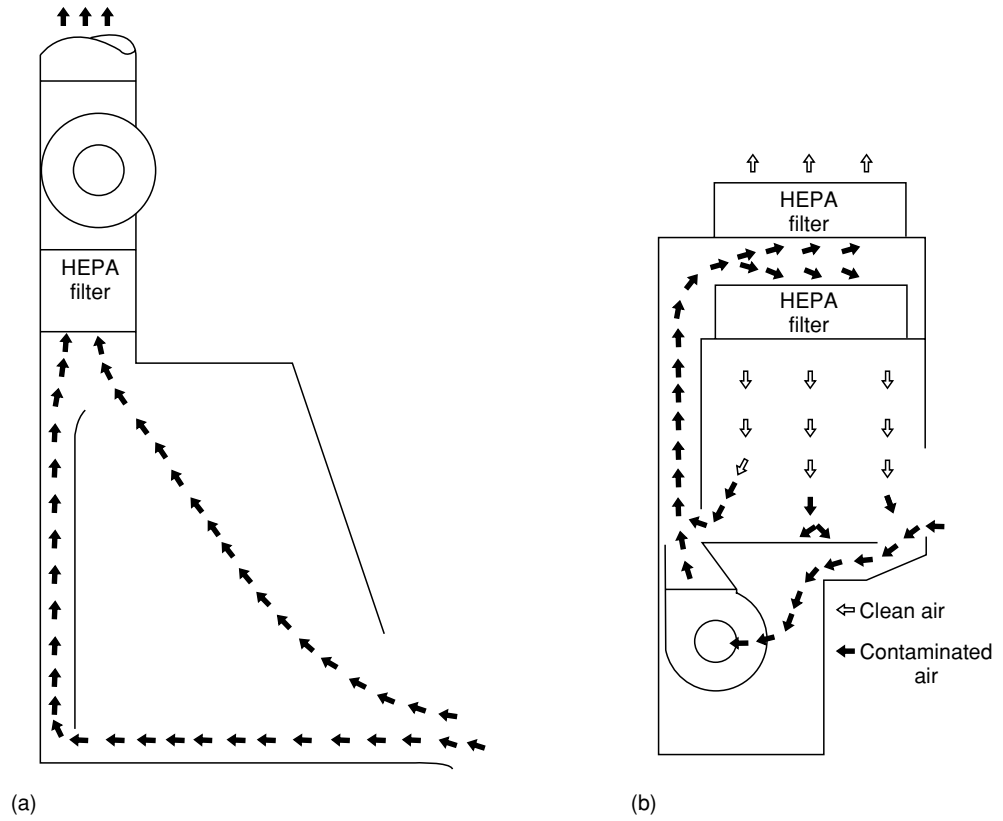

Figure 1 Biosafety cabinets: (a) class I and (b) class II. HEPA, high-efficiency particulate air. Reproduced from Furr AK (ed.) (1995) *CRC Handbook of Laboratory Safety*, 4th edn. Boca Raton, FL: CRC Press, with permission.

Disinfection/Decontamination

Laboratory Waste

Food laboratories generate potentially hazardous infectious wastes during the analysis of samples contaminated by pathogens and during the propagation of pathogenic microorganisms. Biological waste containing viable microorganisms not known to be hazardous to humans is not considered infectious laboratory waste. Nevertheless, it is good laboratory practice to decontaminate all waste containing viable microorganisms and to disinfect work surfaces as if working with infectious agents.

Containment of waste Materials to be disinfected or decontaminated prior to disposal or cleaning should be placed in a centralized area designated for biohazardous waste. A cart clearly labeled 'To Be Autoclaved' is recommended for all laboratories in which infectious organisms are handled. Tubes, bottles, and flasks should be closed. Petri dishes should be placed in trays or biohazard or autoclave bags. Pipettes should be placed in pans or jars with sufficient disinfectant to fill the pipette and cover it completely. Materials to be disinfected and discarded should be separated from those which are to be cleaned and reused. Materials to be discarded can be placed in autoclave bags which are sealed, autoclaved, and then discarded.

Infectious waste (liquid and solid) should be decontaminated at the laboratory facility prior to transport to a disposal site. If this is not possible, adequate physical containment measures (sealed, leak-proof containers) must be taken to prevent subsequent exposure to those handling the waste.

Physical hazards Items with sharp points or edges, such as needles, disposable pipettes, scalpel blades, and broken glass, pose a special problem in decontamination and disposal. These should be placed in puncture-resistant containers either before or after autoclaving. Needles and syringes should be placed directly in the disposal container without replacing the guard on the needle, since replacing the guard is a frequent cause of inoculation of laboratory personnel. After decontamination, special containers that provide a means to cut through both the needle and the plastic syringe should be used to prevent reuse of disposable needles and syringes.

Disinfectants

Disinfection/decontamination in a food laboratory generally occurs by means of chemicals or by heat (usually autoclaving). The type of treatment will depend on the nature of the material to be treated, the level and type of contamination, and the amount of disinfection/decontamination required.

Chemical Chemicals most frequently used for disinfection include quaternary ammonium compounds (0.1–2%), chlorine compounds (0.01–5%), iodophor compounds (0.5–1%), and alcohol (70–95%). Biosafety cabinets are sometimes decontaminated with formaldehyde. Commercial disinfectants should be used according to the manufacturer's directions. Disinfection will depend on selection of the appropriate chemical for the intended use, contact time, concentration, and the presence of interfering substances. For example, although 200 p.p.m. available chlorine is generally adequate for disinfection, organic materials may be present which will combine with chlorine and reduce its effectiveness. Thus it is recommended that discard jars contain 1000 p.p.m. available chlorine; 5% bleach (50 000 p.p.m. chlorine) is recommended for cleaning spills.

Heat Autoclaving is the method of choice for decontaminating cultures, glassware, and pipettes. Autoclaving for 60–90 min at 121 °C (15 psi), depending on the loading conditions, to achieve a waste temperature of at least 115 °C for 20 min is recommended. Effective autoclaving will depend on time, temperature, and steam penetration. Tubes, bottles, flasks, and bags should be loosely closed to allow steam penetration. Tubes should be autoclaved in racks. Glassware should be arranged to allow steam to circulate around the individual containers. The autoclave should never be overloaded. Contaminated glassware should contain liquids to facilitate the sterilization process. Care should be taken when adding water to avoid generating aerosols containing infectious microorganisms. Stainless-steel pans, glass, and heat-resistant plastic can be autoclaved. Metal containers enhance transfer of heat to the waste load, whereas plastic retards steam penetration. Thus with certain materials or with larger volumes of liquids it may be necessary to decrease the size of the load or to increase time to compensate for poor heat transfer/steam penetration. Autoclaved waste can subsequently be disposed of as general waste.

Dry heat may be used to decontaminate some materials, including glassware and instruments, and is recommended for anhydrous materials such as oils, greases, and powders. These should be heated at 160–180 °C for 180–240 min.

Large volumes of infectious wastes, animal carcasses, and contaminated bedding materials should be incinerated. Since most incinerators are off-site, solid waste must be contained in sturdy bags or boxes and liquids in leak-proof containers for transport. Trained personnel should transport the containers to the incinerator.

Inoculating loops are sterilized by heat. Although electric loop sterilizers are available, loops are generally sterilized by holding in a flame until the loop is red-hot. Inoculating loops should be flamed by heating the shaft until the sample has heat-dried before flaming the loop itself. (Alternatively, flaming can be avoided by use of sterile disposable plastic loops.)

Laboratory Decontamination

Frequency Chemical disinfection is used for benchtops, incubators, furniture, and floors which may become contaminated. Work surfaces should be decontaminated at least daily and preferably before and after each use. Biosafety cabinet surfaces should be disinfected before and after each use. Other areas should be cleaned and disinfected at least weekly. Immediate cleaning and disinfection are required when spills occur. Chemicals are also used in disinfecting items placed in discard jars; chemical solutions should be replaced as often as necessary to maintain an effective concentration of disinfectant.

Spills Spills of biohazardous materials require special considerations. Spills can be contained by using enamel or stainless-steel trays as a work surface. If a spill occurs, the entire tray can be autoclaved. Small spills can be cleaned up by pouring a concentrated disinfectant around the edges of the spill and then covering the spill with disinfectant-soaked paper towels. Additional paper towels are used to wipe up the area. Contaminated paper towels are then transferred to a biohazardous waste container for autoclaving. Larger spills, however, can generate significant aerosols which should be allowed to settle for approximately 30 min prior to decontamination. The laboratory should be evacuated and closed during this period. A laboratory worker wearing a lab coat, gloves, and respiratory protection should then clean up the spill with disinfectant and paper towels, sponges, or mops as appropriate. Clean-up materials should preferably be disposable. When there is a large spill in a biosafety cabinet, special disinfection measures may be needed to decontaminate the fan, filters, and air-flow plenums. When infectious agents are spilled on clothing, contaminated garments (lab coats, shoes, etc.) should be removed and decontaminated.

Centrifuge Accidents

If breakage occurs during centrifugation the centrifuge should be turned off and allowed to stand closed for 30 min. If breakage is discovered after the centrifuge is opened, the lid should be replaced and the centrifuge left closed for 30 min. Broken tubes, glass fragments, buckets, trunnions, and the rotor should be placed in an appropriate noncorrosive disinfectant and held for an appropriate time (generally 1–24 h) or autoclaved.

Personnel Practices and Personal Hygiene

The most important aspects of microbiological safety in the laboratory rely on good laboratory practices by the laboratory worker. Persons working with potentially infectious materials should be made aware of the hazards and properly trained in the proper methods for safely handling such materials. Insuring that laboratory personnel are properly trained and informed of potential hazards is the responsibility of the laboratory director.

Access to the laboratory should be limited, especially when working with infectious agents, and doors should be kept closed when experiments are in progress. As noted previously, mouth pipetting should be prohibited. Eating, drinking, smoking, handling contact lenses, and applying cosmetics should not be permitted in the work area. Food should be stored in cabinets or refrigerators designated for this purpose only and food storage areas should be located outside the work area. Drinking-water fountains should also be located outside the work area. Workers should wear laboratory coats or other protective covers. In some cases protective gloves should also be worn. Open-toed shoes or sandals should not be worn in the laboratory. Safety glasses should be worn when filling syringes or when handling certain chemicals such as strong acids. Lab coats and gloves should be removed when leaving the laboratory. Workers should wash their hands following all laboratory activities, after removing laboratory coats and gloves, and immediately following contact with any infectious materials. Workers should develop the habit of keeping their hands away from their face, especially the mouth, nose, and eyes, to prevent self-contamination and infection. This is particularly important when working with agents of low infectious dose, such as *E. coli* O157:H7, where fewer than 10 organisms have been shown to cause illness.

Some employees may have personal health conditions that place them at increased risk in the laboratory, especially when working with certain organisms. For example a pregnant woman may be at increased risk from *Listeria monocytogenes*. Thus most food laboratories do not allow pregnant women to work in a laboratory where *L. monocytogenes* is being handled. Persons who are immunocompromised may be at increased risk not only from *L. monocytogenes*, but also from opportunistic pathogens that may be encountered in foods. Careful assessment of the risk of exposure should be made. If the exposure cannot be eliminated or reduced, consideration should be given to changing jobs. The employee must be informed of the risk and involved in the decision-making process.

Workers with open sores or cuts should not be allowed to handle infectious materials that could potentially contaminate the wound. If sores or cuts are on the hands, bandages and gloves must be worn to protect the worker.

For some infectious agents vaccines that provide an extra level of protection may be available. However for most foodborne disease organisms, vaccines are either not available (e.g., *Shigella* spp., *L. monocytogenes, Campylobacter* spp., most *Salmonella* spp.) or only provide partial protection of short duration (e.g., *Vibrio cholerae*). A pentavalent (ABCDE) botulinum toxoid is available through the US Centers for Disease Control and Prevention as an investigational new drug and is recommended for personnel working with cultures of *C. botulinum* or its toxins. Individuals receiving the toxoid may experience minor side-effects, including redness and swelling at the site of the inoculation; in rare cases more generalized reactions have occurred. Hepatitis A vaccine is available; although laboratory-acquired infection with this virus is not considered a significant risk, the vaccine is recommended for persons who work in laboratories conducting research on this virus.

Other Considerations

All samples and cultures containing infectious materials should be labeled with the universal biohazard symbol and the hazard identified on the label.

Foods for analysis that may be contaminated with pathogens should be assumed to be contaminated and handled appropriately. For example, swollen cans may be contaminated with *C. botulinum* and should be handled, using appropriate safety handling practices, by an individual who is immunized against the toxins. Consideration should be given to the need to open the cans in a safety cabinet.

Laboratories working with infectious agents should have posted on the door the universal biohazard symbol identifying the infectious agent. The sign should also list the name and phone number of the

laboratory supervisor or other responsible party and any special requirements for entering the laboratory. Laboratories dealing with infectious materials should be restricted to authorized persons only.

Hazardous materials should be stored in locked cabinets, refrigerators, or other storage areas and access should be limited to authorized persons only.

Food microbiology laboratories should have safety guidelines posted or a safety manual located in an accessible location. Safety practices appropriate for the specific microbiological hazards that may be encountered should be included, as well as information on chemical and physical hazards.

Summary of Microbiological Safety Practices

1. Avoid mouth pipetting.
2. Handle samples in a manner that minimizes aerosol generation; for procedures known to create aerosols, take steps to contain aerosols.
3. Avoid the use of needles and syringes wherever possible.
4. Decontaminate work surfaces daily and after spills.
5. Keep laboratory doors closed when experiments are in progress.
6. Do not eat, drink, smoke, handle contact lenses, or apply cosmetics in the laboratory.
7. Wear a lab coat when working in the laboratory and remove it when leaving the lab.
8. Wash hands when leaving the laboratory and upon contact with infectious materials.

See also: **Cleaning Procedures in the Factory**: Types of Disinfectant; **Hazard Analysis Critical Control Point**; **Laboratory Management**: Chemical Safety; **Microbiology**: Classification of Microorganisms

Further Reading

Fleming DO, Richardson JH, Tulis JJ and Vesley D (eds) (1995) *Laboratory Safety: Principles and Practices*, 2nd edn. Washington, DC: ASM Press.

Furr AK (ed.) (1995) *CRC Handbook of Laboratory Safety*, 4th edn. Boca Raton, FL: CRC Press.

National Research Council (US) Committee on Hazardous Biological Substances in the Laboratory (1989) *Biosafety in the Laboratory*. Washington, DC: National Academy Press.

Richmond JY and McKinney RW (eds) (1999) *Biosafety in Microbiological and Biomedical Laboratories*, 4th edn. US Public Health Service, Centers for Disease Control and Prevention and National Institutes of Health, HHS publication no. (CDC) 0093–8395. Washington, DC: US Government Printing Office.

LACTATION

Contents

Human Milk: Composition and Nutritional Value

C M Donangelo and N M F Trugo, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil

Introduction

It is well recognized that human milk is the most appropriate source of nutrition for infants during the first 4–6 months of life. Human milk provides a balanced composition of nutrients as well as many types of bioactive components which have an important role in infant growth and development. Benefits of breast-feeding for the infant include adequate nutrition, protection against disease, development of gastrointestinal and immune functions, and psychological well-being. It has been recognized recently that breast-fed infants may have an advantage over formula-fed infants as regards long-term cognitive development and lower rates of neurological disorders.

Human milk is a very complex biological fluid that contains numerous constituents distributed into several physicochemical compartments: aqueous phase (mineral ions, lactose, citrate, amino acids, whey proteins, water-soluble vitamins), colloidal dispersion

(casein micelles), emulsion (fat globules, membrane-bound compounds, fat-soluble vitamins), and cells (macrophages, neutrophils, lymphocytes, epithelial cells). This compartmentation of human milk constituents affects binding properties and bioavailability to the infant. Interactions among components in the compartments may influence the accessibility and flow of nutrients to the infant as well as transmission of biochemical messages to the infant via milk.

Human milk composition shows marked inter- and intraindividual variation. Several factors are known to affect composition (Table 1). Stage of lactation is responsible for the largest variation. Remarkable changes in composition occur as lactation progresses from the first few days (colostrum) to over 3 weeks (mature milk) after childbirth. Typical values of selected nutrients in colostrum and mature human milk are shown in Table 2. Nutrients such as protein, sodium, and zinc are present at higher concentrations in colostrum while others, such as lactose and fat, are present in higher concentrations in mature milk. After the first month postpartum, milk composition generally stabilizes and further changes are less dramatic. It has been proposed that alterations of nutrient supply as lactation progresses reflect the changing nutrient requirements of the infant, with a marked decrease in growth rate over time, and also the development of the infant's capability to digest and absorb nutrients.

Maternal diet is an important factor affecting the nutrient composition of milk but the influence varies in magnitude between nutrients. Concentrations of individual fatty acids and of fat-soluble and water-soluble vitamins generally reflect maternal intake. Concentrations of lactose and minerals such as sodium, calcium, phosphorous, magnesium, and iron do not vary with maternal diet. Maternal hormonal adjustments and nutritional stores may contribute to prevent fluctuations depending on the nutrient so that moderate dietary deficiencies or excesses are not always reflected in milk composition. Severe maternal malnutrition may, however, impair milk volume and milk composition.

Several components of human milk, either nutrients or nonnutrients, exert nonnutrient actions which benefit the health and nutrition of the recipient infant.

Table 1 Factors affecting human milk composition

Stage of lactation
Changes during nursing
Diurnal rhythm
Gestational age at birth (preterm versus term)
Maternal diet
Maternal weight and body composition
Maternal disease (infections, metabolic disorders)
Maternal age
Parity
Geographic area
Ethnicity ?
Individuality

Table 2 Nutrient concentrations in human colostrum and mature milk

Nutrient	*Colostrum (1–5 days postpartum)*	*Mature milk (1–6 months postpartum)*
Macronutrients		
Protein ($g l^{-1}$)	11–32[a]	8–13[a]
	19–54[g]	8–19[g]
Lactose ($g l^{-1}$)	20–58[a]	68–73[a]
	42–64[g]	50–84[g]
Fat ($g l^{-1}$)	10–28[a]	36–43[a]
		27–40[g]
Major minerals		
Sodium ($mmol l^{-1}$)	17–22[b]	4.7–8.0[b]
	7–12[g]	5.2–10.4[g]
Potassium ($mmol l^{-1}$)	17.5–18.5[b]	10.4–13.9[b]
	7.2–15.1[g]	9.5–15.9[g]
Calcium ($mmol l^{-1}$)	5.4–8.0[b]	4.7–7.5[b]
	5.1–11.6[g]	3.5–9.0[g]
Magnesium ($mmol l^{-1}$)	1.0–2.2[b]	1.2–1.7[b]
	1.0–1.8[g]	1.2–1.5[g]
Phosphorus ($mmol l^{-1}$)	3.5–4.7[b]	4.4–4.7[b]
		2.8–6.8[g]
Trace minerals		
Iron ($mg l^{-1}$)	0.6–1.0[c]	0.3–0.9[c]
	0.4–1.2[g]	0.3–2.2[g]
Zinc ($mg l^{-1}$)	8–12[c]	1–3[c]
	2.2–10.4[g]	0.8–4.1[g]
Copper ($mg l^{-1}$)	0.5–0.8[c]	0.2–0.4[c]
	0.1–0.5[g]	0.2–0.5[g]
Manganese ($\mu g l^{-1}$)	5–12[c]	3–6[c]
Selenium ($\mu g l^{-1}$)	14–83[c]	6–28[c]
		10–21[g]
Fat-soluble vitamins		
Vitamin A (retinol $mg l^{-1}$)	0.34–5.0[g]	0.19–1.12[d]
		0.24–1.18[g]
Vitamin K ($\mu g l^{-1}$)	1.8–5.2[d]	1.2–9.2[d]
Vitamin E (α-tocopherol $mg l^{-1}$)	8–22[e]	1.8–3.4[e]
Water-soluble vitamins		
Ascorbic acid ($mg l^{-1}$)	44–60[g]	19–41[g]
		20–105[f]
Thiamin ($\mu g l^{-1}$)	56–200[g]	100–160[g]
		160–238[f]
Riboflavin ($\mu g l^{-1}$)	140–300[g]	190–280[g]
		475–580[f]
Vitamin B_6 ($mg l^{-1}$)	0.02[g]	0.06–0.12[g]
		0.07–0.31[f]
Folate ($\mu g l^{-1}$)	10–40[g]	2–137[g]
		79–133[f]
Vitamin B_{12} ($ng l^{-1}$)	260–1200[g]	78–170[g]
		160–970[f]

Data from: [a]Kunz *et al*. (1999); and [b]Atkinson *et al*., [c]Casey *et al*., [d]Canfield *et al*., [e]Lammi-Keefe, [f]Picciano, [g]Prentice (data from outside Europe and the USA) In: Jensen (1995) *Handbook of Milk Composition*, San Diego: Academic Press.

They can be classified in several categories, according to their actions. Since many of these bioactive substances present multiple roles and act synergistically with each other, a particular component may be present in more than one category. A broad category is that of the defense agents, comprising cellular elements (leukocytes) and a variety of heterogeneous soluble compounds, which may act directly or indirectly against pathogens (antiinfectious agents) and/or alter manifestations of infection due to their antiinflammatory properties (antiinflammatory factors), and modulation of the immune responses (immunomodulators). Other categories of beneficial compounds present in human milk are growth factors and hormones, which promote growth and maturation of organ systems; binding proteins for hormones, vitamins, and minerals, which protect and increase the bioavailability of the ligands; and several enzymes, which ensure a better utilization of nutrients and act as defense agents.

Nutritional Components of Human Milk

Protein

Proteins are present in human milk as whey proteins (50–90% of total), casein (10–50%) and milk fat globule membrane proteins (1–3%). Whey proteins include α-lactalbumin, the major protein in this fraction (10–20% of total milk protein), serum albumin, immunoglobulins, and lactoferrin. Casein is mainly composed of two proteins: β-casein (85%), with different phosphorylated forms, and κ-casein (15%), a highly glycolysylated protein containing 40–60% carbohydrate residues (hexoses, hexosamines, and sialic acid). Casein micelles are formed by aggregation of caseins, calcium phosphate, and other ions, depending on both hydrophobic interactions and electrostatic binding. Human milk casein micelles are considerably smaller (30–75 nm in diameter) than bovine casein micelles (about 50–600 nm, mean 120 nm).

The total protein concentration in milk is high during early lactation (colostrum) and gradually declines to a low level in mature milk (Table 2). Protein concentration during early lactation is usually higher in preterm than in term milk. There are pronounced changes in the protein composition of milk as lactation progresses. Whey proteins decrease while casein increases in milk during the first few days of lactation. The ratio of whey proteins to casein is high (about 90:10) in colostrum and declines to 50:50 in mature milk. The pattern in caseins is different for premature milk, colostrum, and mature milk. In general, κ-casein is detectable in human milk only 3–4 days postpartum and may not be detectable in preterm colostrum. β-Casein concentration is higher in colostrum and mature milk of mothers with term compared to preterm infants.

Milk proteins are a source of peptides, amino acids, and nitrogen for the growing infant. In addition, milk proteins serve several nonnutritive roles. Proteins in human milk are not completely nutritionally available to the infant. Proteins with nonnutrient functions must remain intact in the infant gastrointestinal tract and may be largely unavailable for nitrogen metabolism, particularly during early lactation. However, at 6 weeks postpartum, 95% of breast milk protein is considered nutritionally available in term infants. Casein in human milk is an easily digested protein which provides amino acids, calcium, and phosphorus to the newborn. Casein phosphopeptides in human milk have been shown to keep calcium (and probably also bound iron, zinc, copper, and manganese) in soluble form in the infant intestinal tract, thus favoring absorption. α-Lactalbumin in human milk has a very high nutritional value (protein quality) and its amino acid composition is similar to the estimated amino acid requirements of young infants.

Nonprotein nitrogen

The nonprotein nitrogen (NPN) fraction of human milk comprises 20–25% of the total nitrogen and remains relatively constant during lactation. This fraction consists of more than 200 compounds, including free amino acids, urea, uric acid, ammonia, carnitine, choline, taurine, amino sugars, amino alcohols, nucleic acids, nucleotides, and polyamines. Many of these components, such as taurine, choline and nucleotides, are considered conditionally essential for the young infant. For example, human milk is a source of purine and pyrimidine bases as well as preformed nucleotides that are required by young infants due to their high need of nucleic acid formation but limited capacity for *de novo* synthesis of nucleotides.

Components of human milk NPN that contribute to infant protein metabolism include some peptides, free amino acids and, mainly, urea. Urea nitrogen is a significant part (30–50%) of total NPN in human milk, and it may contribute to infant protein synthesis. Free amino acids represent 3–5% of the total amino acids in human milk with higher levels in colostrum than in mature milk. Glutamate/glutamine and the nonprotein sulfonic acid, taurine, are the predominant free amino acids in mature human milk. Levels of urea and free amino acids in human milk seem to reflect protein quality and protein intake in the maternal diet. Increased urea and free amino acid levels have been observed in women receiving

high-versus low-protein diets. In mothers consuming diets low in lysine, methionine, and/or tryptophane, milk levels of these amino acids in the free form are also low.

Carbohydrates

Lactose is the major carbohydrate and, after water, is the second major constituent in human milk (*c.* 70 g l^{-1} in mature milk). Other carbohydrates in human milk include nucleotide sugars, glycolipids, glycoproteins, and oligosaccharides, many of which have several nonnutrient functions. Total oligosaccharide content in human milk is 5–8 g l^{-1}. Low levels of free galactose and minor levels of glucose are also found in human milk (*c.* 3 and 0.3 g l^{-1} in mature milk, respectively).

Lactose is one of the most stable constituents of human milk, with variation in levels reflecting mainly heterogeneity of individuals. Levels in milk increase (slightly) during the first few days postpartum but remain quite stable in mature milk (Table 2). Its concentration decreases during nursing but no variation occurs if levels are corrected for the increase in fat. Lactose levels may be low in preterm colostrum and in colostrum of women with insulin-dependent diabetes mellitus. Lactose in human milk is readily available to the infant and has been reported to stimulate the absorption of calcium, magnesium, and manganese.

Lipids

Human milk lipids are the major energy source for the infant, as well as a source of essential fatty acids (linoleic, 18:2 *n*-6 and α-linolenic, 18:3 *n*-3), and their long-chain polyunsaturated fatty acid metabolites such as arachidonic acid (20:4 *n*-6), eicosapentaenoic acid (20:5 *n*-3), and docosahexaenoic acid (22:6 *n*-3).

Lipids are present in human milk in the form of globules with a hydrophobic core consisting of triacylglycerols, cholesteryl esters, and retinyl esters and other neutral lipids, coated with bipolar compounds, phospholipids, proteins, cholesterol, and enzymes, formed into a loose layer called the milk lipid globule membrane. The diameter of the globules ranges from 1 to 10 μm, with most of the globules at 1 μm, giving a large surface area for efficient binding to lipolytic enzymes, thus facilitating lipid digestion and absorption in the infant gastrointestinal tract.

Triacylglycerols account for at least 98% of the milk lipids, followed by phospholipids (0.7%), cholesterol (0.5%), and minor amounts of free fatty acids, and mono- and diacylglycerols. The major fatty acids in triacylglycerols are 18:1 and 16:0 (*c.* 30% and 20% of total, respectively). The structure of the triacylglycerols in human milk is unique with most of the 16:0 at *sn*-2, 12:0 at *sn*-3, 18:0 at *sn*-1, and 18:1 and 18:2 at *sn*-1 and *sn*-3. Structure is an important factor controlling products formed by the action of lipases and their intestinal absorption in the infant.

Lipids are the most variable constituents of human milk. Total milk fat increases with the stage of lactation (Table 2) and during nursing; it decreases with parity, maternal infection, and long-term maternal undernutrition; it shows diurnal rhythm, and may vary with maternal diet and body composition (adiposity). With the increase in total milk fat during the first month of lactation, there is an increase in the average size of the milk fat globules and a reduced ratio of phospholipids and cholesterol (membrane lipids) to triacylglycerols (core lipids).

A major proportion of human milk fat is formed from circulating lipids derived from maternal diet, and from maternal body stores. A small proportion is synthesized *de novo* in the mammary gland from glucose, primarily as saturated fatty acids with medium-chain length (10–14 carbon atoms). Both maternal diet and maternal adipose tissue affect the composition of human milk triacylglycerols.

Maternal diet has a considerable influence on the fatty acid composition of human milk. Long-chain polyunsaturated fatty acids (such as 20:4 *n*-6, 20:5 *n*-3 and 22:6 *n*-3) which are synthesized by the maternal organism from their essential precursors (18:2 *n*-6 and 18:3 *n*-3), respond to changes in the diet. For example, levels of eicosapentaenoic acid (20:5 *n*-3) and docosahexaenoic acid (22:6 *n*-3) in human milk are increased by consumption of fish such as herring, mackerel, and salmon. The contents of *trans* isomers of fatty acids in milk reflect maternal intake of partially hydrogenated food fats and oils, deep-fried foods, and dairy products.

Typical western and nonwestern diets affect the fatty acid composition of human milk (Table 3). On high-fat (western) diets, fatty acid composition reflects dietary fatty acid pattern with high levels of unsaturated fatty acids related to consumption of vegetable oils. In this condition, only about 10% of the milk fatty acids are synthesized by the mammary gland. On high-carbohydrate low-fat (nonwestern) diets, the proportion of fatty acids originating in the mammary gland increases, resulting in higher levels of medium-chain fatty acids in milk. On low-energy diets, much of the milk fat is derived from adipose tissue and the composition of milk triacylglycerols reflects maternal depot fat. Maternal diet does not affect total cholesterol and phospholipids levels of milk.

Vitamins

All nutritionally essential vitamins are found in human milk. Fat-soluble vitamins are present in the core of milk fat globules (vitamin A, vitamin K) or bound to membrane or specific binding proteins (vitamin E, vitamin D). Water-soluble vitamins such as folate and vitamin B_{12} are bound to specific whey proteins.

Typical levels of vitamins in colostrum and mature human milk are presented in Table 2. Stage of lactation and particularly maternal nutritional status are the most important factors affecting the vitamin content of human milk (Table 4). In general, when maternal vitamin intakes are low, milk levels are also low and respond to supplementation; when they are high, milk levels approach a plateau and are less responsive to maternal supplementation. For many vitamins, such as folate and vitamin E, levels in milk are largely preserved through maternal regulation even, if maternal status is poor. For others, such as vitamin A, vitamin D, and vitamin B_{12}, levels in milk may be very low with maternal deficiency, and in this situation, inadequate for infant nutrition. Premature delivery is another factor affecting vitamin levels in human milk. Preterm milk has a high content of ascorbic acid, vitamin B_{12}, and pantothenic acid but may not be sufficient to meet the needs of the preterm infant for thiamin, vitamin B_6, vitamin E, and vitamin D.

Fat-soluble vitamins Vitamin A is present in human milk mainly as retinyl esters (95%) and as free retinol. Vitamin A activity is also provided as β-carotene which accounts for up to 30% of total carotenoids in milk. Other carotenoids present in human milk are lycopene, lutein, and cryptoxanthin, which have no provitamin A activity. The total level of vitamin A shows a rapid decrease over the first month postpartum. Vitamin A milk level is responsive to increased maternal intake, particularly in mothers with poor vitamin A status. The retinol:retinyl esters ratio in mature milk does not change after loading with vitamin A in well-nourished mothers but it may rise up to 30% in vitamin A-deficient mothers. Poorly nourished mothers have a low level of vitamin A in their milk and are thus at risk of providing insufficient amounts to the infant. Clinical evidence of vitamin A deficiency has been found after weaning in infants reared on breast milk containing 30 μg dl^{-1} vitamin A or lower.

Vitamin E is present in human milk mainly as α-tocopherol (*c.* 80%); the β-, δ-, and γ-isomers are also present in small quantities. The level is very high in colostrum and quite stable in mature milk; generally, it does not respond to maternal supplementation,

Table 3 Fatty acid composition of human milk lipids with different diets

	Western diets		*Nonwestern diets*	
Fatty acid	*wt %*	*range*	*wt %*	*range*
Saturated fatty acid				
8:0	0.17	0.02–0.37	0.37	0.10–0.59
10:0	1.0	0.06–2.4	1.6	0.5–3.4
12:0	4.9	1.7–12.3	8.1	2.4–16.5
14:0	5.6	2.0–11.8	9.6	5.3–15.9
16:0	20.3	19.3–25.1	21.5	14.1–25.8
18:0	7.5	5.8–9.7	5.6	0.80–8.2
Monounsaturated fatty acids				
14:1 *n*-5	0.31	0.22–0.49	0.68	0.19–5.00
16:1 *n*-7	3.3	1.8–5.0	0.9	0.6–2.2
18:1 *n*-7	3.4	3.2–3.8	2.9	2.3–3.8
18:1 *n*-9 (*cis*)	31.0	22.6–38.7	30.5	17.9–47.0
18:1 *n*-9 (*trans*)	3.6	3.1–4.7	0.9	0.5–4.9
Polyunsaturated fatty acids				
18:2 *n*-6	12.6	9.6–16.8	13.8	8.8–23.8
18:3 *n*-6	0.4	0.09–1.03	0.1	0.07–0.3
20:2 *n*-6	0.3	0.2–0.5	0.4	0.2–0.8
20:4 *n*-6	0.5	0.4–0.7	0.5	0.1–0.7
18:3 *n*-3	0.7	0.3–1.9	0.5	0.1–1.0
20:5 *n*-3	0.07	0.00–0.16	0.24	0.05–1.10
22:5 *n*-3	0.05	0.01–0.25	0.24	0.10–0.80
22:6 *n*-3	0.14	0.00–0.53	0.57	0.10–1.40

Data from: Jensen (1995) *Handbook of Milk Composition*, San Diego: Academic Press and Rodriguez-Palmero *et al*. (1999) Nutritional and biochemical properties of human milk. Part II: Lipids, Micronutrients, and bioactive factors. *Clinics in Perinatology* 26: 335–359.

Table 4 Effects of stage of lactation and maternal diet on vitamin content of human milk

Vitamin	*Advancing lactation*[a]	*Maternal diet*[b]
Fat-soluble vitamins		
Vitamin A	↓	+
Vitamin D	=	+
Vitamin E	↓	=
Vitamin K	=	=
Water-soluble vitamins		
Ascorbic acid	↑	+
Thiamin	↑	+
Riboflavin	↑	++
Niacin	↑	+
Folate	↑	+
Vitamin B_6	↑	++
Vitamin B_{12}	↓	+
Pantothenic acid	↑	+
Biotin	↑	+

[a]Mature milk levels (> 1 month postpartum) compared to colostrum (1–5 days postpartum): higher (↑), lower (↓), or no change (=).
[b]+ indicates that maternal intake of the vitamin influences milk vitamin content primarily in deficient women; ++ indicates that maternal intake influences milk content even when vitamin status of mother is adequate; = indicates no influence of maternal intake on milk content.
Adapted from: Picciano MF Human milk: nutritional aspects of a dynamic food. *Biology of the Neonate* 74: 84–93. In: Jensen (1995) *Handbook of Milk Composition*, San Diego: Academic Press.

and is similar between populations with different vitamin E status, suggesting that maternal vitamin E stores are mobilized during lactation to insure vitamin E supply to the neonate.

Vitamin D-active compounds in human milk are the 25-hydroxylated forms, cholecalciferol and ergocalciferol, present in the nonlipid fraction of milk at 1.5–6% of their concentration in maternal plasma. The level of vitamin D in milk responds to maternal intake and exposure to sunlight. The level may be very low in women with poor vitamin D intake (e.g., strict vegetarians) and in those who live at northern latitudes. Exclusively breast-fed infants in areas of poor sunlight exposure may be at risk of vitamin D deficiency.

Vitamin K in human milk occurs mainly as phylloquinone. Levels in milk do not change consistently with stage of lactation and do not relate with maternal dietary intake. The level can be increased by maternal supplementation of vitamin K at very large doses. In general, it appears that the content of vitamin K in human milk is low and may not be sufficient to meet the needs of all newborns. In many countries, it is recommended that newborns receive vitamin K after birth.

Water-soluble vitamins Except for vitamin B_{12}, concentrations of water-soluble vitamins are lower in colostrum than in mature milk. In general, levels of water-soluble vitamins in mature human milk are several-fold greater than those in maternal plasma. For example, levels of ascorbic acid, thiamin, vitamin B_6, and folate in milk are 7–10 times the corresponding plasma levels. This suggests that the mammary gland actively transports and metabolizes these vitamins and regulates secretion, but mechanisms remain largely unknown.

The impact of maternal nutrition on the concentration of water-soluble vitamins in milk depends on the specific vitamin, the maternal status, and stage of lactation. Milk levels of vitamin B_{12} are very low in strict vegetarian women. The level of folate in milk is maintained with the concomitant depletion of maternal folate. Maternal supplementation of vitamins such as niacin, riboflavin, and vitamin B_6 affects levels in milk only when maternal status is inadequate. In well-nourished mothers, maternal supplementation at high doses has either no effect (i.e, ascorbic acid, folate, riboflavin) or little effect (other water-soluble vitamins) on levels in milk. The content of vitamin B_6 in human milk responds to maternal supplementation during established lactation but not during early stages. The level of vitamin B_6 in milk is greatly reduced in women with long-term use of oral contraceptive.

Minerals

Major minerals in human milk are the monovalent ions – sodium, potassium, and chloride – and the divalent species – calcium, magnesium, phosphate and sulfate – followed by the trace minerals, all distributed among the different physicochemical compartments of milk in a highly specific manner. Several studies have demonstrated a high bioavailability of most minerals in human milk, related to compartmentation and specific binding, as well as their favorable interrelationship with other milk nutrients for efficient utilization by the infant.

In general, mineral constituents in human milk do not correlate with maternal diet or maternal blood levels. Exceptions are the levels in milk of anionic elements such as iodine, fluoride, and selenium, which strongly reflect maternal status and intake, and also show marked geographic differences. Stage of lactation affects levels of all minerals in milk, with most rapid changes occurring during the first few days postpartum. Typical levels of minerals in colostrum and mature human milk are presented in Table 2.

Major minerals Electrolytes in milk (sodium, potassium, and chloride ions) are at much lower concentrations than those in maternal plasma, and collectively contribute to 10% of the total osmolarity of human milk. Variation in levels of electrolytes from colostrum to mature milk are inversely related to changes in lactose concentration. Levels are not influenced by nutritional factors or by maternal systemic diseases such as diabetes, but increase two- to three fold during mastitis.

Calcium is found in human milk mainly in the aqueous phase (80%), both as a free ion and associated with low-molecular-weight species such as citrate and phosphate, and bound to casein (20%). The concentration of ionized calcium in human milk is quite constant ($3\,\text{mmol}\,l^{-1}$) throughout lactation. Total calcium levels in milk increase markedly during the first days postpartum, stabilize in mature milk, and decline gradually in extended lactation. Magnesium is present mainly in the aqueous fraction of human milk (90%) and in the lipid globule membrane (10%). The total concentration of magnesium in human milk falls gradually during lactation. Maternal diet has no effect on calcium, magnesium, phosphate, and citrate levels in milk. Preterm milk may not be sufficient to meet the infant needs for calcium and phosphorus.

Trace minerals Iron in human milk is distributed in various compartments, about one-third associated

with low-molecular-weight compounds in the aqueous fraction, one-third with the milk fat, mainly the outer fat globule membrane, and about 10% in the casein fraction. At least some of the 20–30% in the whey protein fraction is bound to lactoferrin, a glycoprotein with two high-affinity binding sites for ferric iron, but only 3–5% of the total iron-binding capacity is used. Lactoferrin has been frequently proposed to account for the high bioavailability of iron in human milk.

The total concentration of iron in human milk is highest during the first few days after birth, with a 30% decrease during the first month of lactation followed by more stable levels. Neither iron supplementation of nursing women with adequate iron status nor poor maternal iron status has been shown to affect the level of iron in milk. There are reported geographic differences in human milk iron levels not related to maternal diet or nutritional status. The concentration of lactoferrin is higher in colostrum (5–6 mg ml^{-1}) than in mature milk (1.5 mg ml^{-1}). Lactoferrin content may be elevated in milk of women with very high iron intakes, and is reduced in generally malnourished women.

Zinc in human milk is distributed between the aqueous (40–70%), lipid (10–30%), and casein (10–45%) fractions. Serum albumin and citrate, followed by glutamate and picolinate, are the major zinc-binding ligands in the whey fraction; alkaline phosphatase, present in the fat globule membrane, appears as the major zinc-binding protein of human milk fat; and phosphoserine residues are the primary zinc-binding sites in caseins. Zinc in human milk is utilized very efficiently by infants. Breast-fed infants were shown to maintain high plasma zinc values compared to formula-fed infants even if the concentration of zinc in the formula was about three times that in breast milk. Human milk has therapeutic value in treating infants suffering from acrodermatitis enteropathica, a hereditary zinc-deficiency disease.

The concentration of zinc is very high in human colostrum, then declines throughout lactation. The most marked decrease (three- to fivefold) occurs during the first 2 weeks postpartum. Levels do not appear to be directly related to maternal nutritional status. Values reported from developed and developing countries are very similar. The effect of maternal zinc supplementation on milk zinc levels may depend, however, on maternal zinc status. For example, in lactating women with marginal zinc intakes it was shown that the decline in zinc concentration in milk over 1–9 months postpartum was significantly less in zinc-supplemented compared to nonsupplemented women.

Most (75%) of the copper in human milk is in the whey fraction, mainly associated to serum albumin and citrate; 15–20% is in the fat phase and the remainder (5%) is bound to casein. Citrate and free amino acids are believed to contribute to the efficient absorption of copper from human milk. The concentration of copper is generally high in colostrum and falls mainly during the first month but less markedly than that of zinc. The concentration of copper in human milk does not appear to be affected by maternal dietary intake or nutritional status of copper. The copper content of human milk may not be sufficient to meet the needs of the preterm infant.

Bioactive Components of Human Milk

Defense agents

The defense agents in human milk provide protection for the infant not only against infections, but also against diseases such as insulin-dependent diabetes mellitus, Crohn's disease, and atopic and allergic diseases. The mucosal surfaces of the intestinal and respiratory tracts of the infant are the major targets for their action, but some may also have a systemic effect after being taken up into the circulation. In general, the concentrations of the defense agents in human milk decrease with the progression of lactation. They are thus higher in a period when the defense system of the infant is immature and when there is a low production of immune agents in mucosal secretions. In order to play their roles in the infant's intestinal tract, these defense agents have to survive digestion, at least partially. Compartmentation and the presence of antiproteinases in milk, resistance to digestion, and low secretion of hydrochloride and pancreatic enzymes are mechanisms that allow survival of the defense factors in the gastrointestinal tract of the infant.

Leukocytes Leukocyte levels are high in colostrum and decrease in mature milk (less than 10% of the colostrum value). Neutrophils and macrophages represent about 90%, and lymphocytes about 10%, of the leukocytes in human milk. Leukocytes in human milk may transmit immunocompetence and promote maturation of the gastrointestinal tract of the infant through the secretion of cytokines, immunoglobulin A (IgA), and growth factors. They may also protect the breast against infection.

Soluble antiinfectious agents Many are effective mainly as antibacterial factors but some may be antiviral, antifungus, and antiprotozoan factors

also. They may act by direct inhibition of the growth of pathogens, inhibition of binding of bacterial pathogens and toxins to mucosal surfaces, and promotion of the growth of protective microorganisms. The main antiinfectious agents in human milk are:

1. Lactoferrin: most of it (> 90%) is present in milk in the form of apolactoferrin, which competes with siderophilic bacteria for ferric iron, thus impairing the proliferation of pathogenic microorganisms. However, lactoferrin may have a more complex antimicrobial effect since it can sensitize microorganisms to antibiotics and to the attack of lysozyme by the release of lipopolysaccharides from bacterial cell walls.
2. Digestive products of milk lipids: products of human milk lipid digestion, especially medium-chain, mono- and polyunsaturated fatty acids and monoacylglycerols, exert antiviral, antibacterial, and antifungal activity *in vitro*. The bile salt-stimulated lipase of human milk appears to be important in the production of these antimicrobial lipids.
3. Enzymes: peroxidase, which is produced by milk leukocytes, catalyzes the oxidation of thiocyanate ions to products with bacteriostactic activity. Lysozyme cleaves peptidoglycans in the cell wall of susceptible bacteria, causing their lysis.
4. Secretory IgA and other immunoglobulins (IgG, IgM, and IgD): secretory IgA (sIgA) is the principal immunoglobulin in human milk, secreted as dimeric IgA bound to a secetory component. Its concentration is higher in colostrum than in mature milk. sIgA protects the infant against mucosal pathogens in the maternal environment, and its antibody specifity reflects the maternal immunological experience. It causes agglutination of bacteria, blocks adhesion of several bacteria, viruses, and parasites to mucosal surfaces in the intestinal and respiratory tracts, neutralizes toxins, and prevents translocation of gut bacteria through the epithelial barrier.
5. κ-Casein and mucin: like other glycoproteins, including sIgA, they interact with pathogenic bacteria and viruses, possibly through their carbohydrate side chains, inhibiting their adhesion to epithelial cells, thus preventing microbial colonization and invasion of intestinal tissue.
6. Oligosaccharides: human milk contains several complex oligosaccharides, which are monomers of glucose, galactose, *N*-acetylglycosamine, fucose, and neuraminic acid. Almost all carry lactose at the reducing end. Oligosaccharides act as soluble receptor analogs of epithelial cell surface carbohydrates and are therefore potent inhibitors of bacterial adhesion to epithelial surfaces. Their biological functions are closely related to their conformation. The nitrogen-containing oligosaccharides also favor the proliferation of intestinal bifidobacteria and lactobacilli, resulting in an acidic intestinal environment, which aids in suppressing the growth of enteric pathogens. Glycoproteins, such as lactoferrin and κ-casein, exert a similar effect, ascribed to their oligosaccharide moiety. Glycoconjugates, such as monosialogangliosides, are receptor analogs for heat-labile toxins produced by bacteria, thus inhibiting binding of toxins.

Anti-inflammatory factors They protect the infant's intestinal tract against inflammatory damage and modulate inflammatory process. Several antiinfectious agents in human milk also have antiinflammatory activity. For instance, lactoferrin inhibits the formation of oxygen radicals by scavenging free iron, inhibits the complement system, and suppresses cytokine release from macrophages stimulated with bacterial products. Lysozyme interferes with the production of superoxide anion. Enzymes such as catalase and glutathione peroxidase act by degrading and removing hydrogen peroxide. Somatostatin-like immunoreactivity has been detected in human milk and may act as an antiinflammatory factor due to its immunosuppressive properties. Other antiinflammatory factors present in human milk are cortisol, epithelial growth factor, polyamines, antioxidants (ascorbate-like compound, uric acid, β-carotene, tocopherol), and inhibitors of nonoxidative inflammatory systems (prostaglandins and platelet-activating factor).

Immunomodulators Human milk components, besides providing passive protection for the neonate, may also directly modulate the immunological development of the infant. The immunomodulators in human milk include bioactive substances which have other primary biological functions, such as hormones and growth factors, substances that present an overlapping of their antiinfectious or antiinflammatory properties, and those that are actual immunological factors, such as cytokines and chemokines. Some examples of the major immunomodulators in human milk are prolactin, which may modulate differentiation and function of gut-associated lymphoid tissue (e.g., intraepithelial lymphocytes); antiidiotypic sIgA, which can provide natural immunization to the infant by sensitizing it against foreign antigens and microbial agents; lactoferrin, which suppresses *in vitro* antibody production and T-lymphocyte proliferative

response to antigens; nucleotides, which can act locally at the respiratory and alimentary tract, or systemically, enhancing T-cell maturation and function, and providing partial resistance to bacterial infection; β-casomorphins, which are derived from casein digestion and have opioid-like activity, and may be able to regulate intestinal motility and immunostimulating functions.

Several cytokines (interleukin (IL)-1β, IL-6, IL-10, interferon-γ, tumor necrosis factor-α), chemokines (IL-8, growth-related peptide, monocyte chemotactic protein) and granulocyte- and macrophage-colony-stimulating factors are present in human milk. These substances are protein hormones which mediate both natural and specific immunity and direct the development and functions of the immune system. The potential role of the immunological factors present in human milk, which would have to retain their biological activity in the gastrointestinal tract and to be transferred to the circulation of the infant, could compensate for the developmental delay of the neonate immune system, both at the mucosal and systemic sites.

Growth factors and hormones

Several nonpeptide hormones and hormonally active peptides are present in human milk. Among the former are the thyroid hormones, cortisol, progesterone, and estrogens. Examples of peptide hormones and growth factors, which promote the development and maturation of organ systems, are insulin, epithelial growth factor (EGF) and insulin-like growth factors (IGF-1 and 2). Both EGF and IGFs stimulate the growth of intestinal tissue. Other substances in human milk, such as lactoferrin and nucleotides, also have similar epithelial growth-promoting activities of intestinal cells *in vitro* or in animal models. The growth and maturation of the gut epithelial barrier would benefit the infant by limiting the penetration of foreign antigens in the intestinal mucosa. Gastrointestinal regulatory peptides (gastrin, gastric inhibitory peptide, gastrin-releasing peptide, neurotensin, somatostatin, etc.) and hypothalamic and hypophyseal hormones (prolactin, growth hormone, thyrotropin-releasing hormone, thyroid-stimulating hormone, gonadotrophin-releasing hormone, etc.) are also present in human milk. Leptin, a peptide hormone secreted by adipocytes and involved in the regulation of energy metabolism and food intake, is also present in human milk, as a result of the possible transfer from maternal plasma and also of the secretion by mammary epithelial cells.

Binding proteins for vitamins, minerals, and hormones

The binding proteins in human milk may improve solubilization and protect the ligands throughout their passage in the gastrointestinal tract of the infant and may increase the bioavailability of the vitamins and minerals attached to them. Binding proteins for vitamin D, thyroxine, and corticosteroids are present in human milk. Cobalamin- and folate-binding proteins in human milk show antimicrobial effects *in vitro* and enhance folate and cobalamin absorption in the neonatal period through specific receptors in animal and in intestinal membrane models. However, their role in cobalamin and folate absorption in infants is still not known. The role of lactoferrin from human milk in enhancing iron absorption by the infant is still controversial, but it might act as a regulator of iron absorption when iron stores are adequate and as an enhancer of iron absorption in deficiency states.

Enzymes

Several enzymes present in human milk are important in neonatal development. Proteinases, which could aid in hydrolysis of milk proteins, thus improving their bioavailability, may have their action limited by the antiproteinases also present. The antiproteinases would nevertheless have a beneficial effect by protecting bioactive milk proteins against hydrolysis in milk and in the infant's intestine. α-Amylase facilitates the digestion of polysaccharides by the infant, and its concentration in milk increases with the stage of lactation, reaching a plateau at 6 months. This is important, since salivary and pancreatic α-amylase only reaches adequate levels in the infant by 2 years after birth. Lysozyme, peroxidase, and glutathione peroxidase are all defense agents in human milk.

In addition to the lipoprotein lipase, human milk contains another lipase that is bile salt-dependent. It is called milk digestive lipase or bile salt-stimulated lipase and it is the most extensively studied enzyme in human milk. It facilitates the hydrolysis of milk fat in the intestine in the neonatal period, aiding the incipient endogenous lipid digestive function of the newborn, especially in premature infants. It acts on lipid substrates and also on water-soluble esters and its concentration is lower in colostrum than in mature milk. The milk digestive lipase completely hydrolyzes triglycerides, since it lacks positional or fatty acid specificity. The contribution of milk lipases to the newborn's endogenous lipases results in an enhancement of antibacterial and antiviral effects due to the production of free fatty acids and monoglycerides that have antiinfective properties.

See also: **Ascorbic Acid**: Properties and Determination; **Calcium**: Properties and Determination; **Copper**: Properties and Determination; **Enzymes**: Functions and Characteristics; **Fatty Acids**: Properties; **Growth and**

Development; **Iron**: Properties and Determination; **Lactation**: Physiology; **Lactose**; **Pregnancy**: Maternal Diet, Vitamins, and Neural Tube Defects; **Retinol**: Properties and Determination; **Tocopherols**: Properties and Determination; **Vitamins**: Overview; **Zinc**: Deficiency

Further Reading

Fomon SJ (ed.) (1993) *Nutrition of Normal Infants*. St. Louis: Mosby.

Jensen RG (ed.) (1995) *Handbook of Milk Composition*, San Diego: Academic Press.

Picciano MF (1998) Human milk: nutritional aspects of a dynamic food. *Biology of the Neonate* 74: 84–93.

Kunz C, Rodriguez-Palmero M, Koletzko B and Jensen R (1999) Nutritional and biochemical properties of human milk. Part I: General aspects, proteins, and carbohydrates. *Clinics in Perinatology* 26: 307–333.

Rodriguez-Palmero M, Koletzko B, Kunz C and Jensen R (1999) Nutritional and biochemical properties of human milk. Part II: Lipids, micronutrients, and bioactive factors. *Clinics in Perinatology* 26: 335–359.

Physiology

M Neville, UCHSC, Denver, CO, USA

Background

The provision of a specialized maternal body fluid, milk, for neonatal nutrition frees the mother from the necessity of providing a specific environment for rearing of the young. It allows birth to occur at an early stage of development and provides a time of intense maternal interaction with the newborn during early behavioral development. In addition, the nutritional reserves of the mother may be able to sustain the suckling through a period of famine. For all these reasons, lactation allows mammals to adapt to a wide variety of environments. It is becoming increasingly clear that breast milk is the most appropriate source of nutrition for human infants at least up to the age of 6 months. Many components of human milk, including, but not-limited to, the protein lactoferrin, growth factors, long-chain polyunsaturated fatty acids (LCPUFA), bile salt stimulated lipase, and anti-infectious oligosaccharides and glycoconjugates are not duplicated in formula. These components may be particularly important for the challenged infant (e.g., the preterm infant, infants with feeding problems, and infants in homes lacking adequate sanitation) but are increasingly being shown to be beneficial to healthy infants in well-protected environments as well.

After reviewing the macronutrient composition of milk, we will summarize the anatomical structure of the mammary gland at both the gross and microscopic levels to provide a basis for discussion of the processes of milk secretion and ejection. The stages of mammary development will be summarized briefly. The process of lactogenesis, the onset of copious milk secretion around parturition, will be discussed as it relates to milk composition, the provision of colostrum to the infant, and the relative roles of hormones and milk removal in initiating milk secretion. We will examine the regulatory mechanisms that help adapt the milk supply both to maternal food supply and to the requirements of the infant. The effects of dietary composition and maternal metabolic status on milk composition and secretion will be summarized, followed by a brief discussion of the involution of the mammary epithelium after the termination of suckling.

Milk Composition

After the period of colostrum formation, the composition of human milk is relatively constant, although some nutrients vary significantly within the feed and with duration of lactation. Breast-feeding can provide adequate nutrition for at least 4 months after birth; whether human milk alone provides sufficient nutrients after this period is currently a matter of controversy. The major macronutrients in milk are the sugar, lactose, (a disaccharide unique to milk), oligosaccharides, milk fat (mainly in the form of triglyceride), proteins, including casein, lactoferrin, secretory immunoglobulin A (sIgA), α-lactalbumin and many others present at lower concentrations, and minerals including sodium, potassium, chloride, calcium, and magnesium. The secretion mechanisms for most of these milk components are not as well understood. There are many minor components, including enzymes, vitamins, trace elements, and growth factors whose function and secretion are less well understood. The concentrations of the major components of human milk are compared with those of bovine milk in **Table 1**.

The major differences between human and bovine milk can be related to the specific needs of the young of the two species. For example, the concentration of lactose is higher and that of the mono- and divalent ions lower in human than in cow's milk. The high concentration of lactose provides a large amount of 'free water' – that is, water that does not need to be obligatorily excreted by the kidneys with salts, providing a reserve for temperature regulation via

Table 1 Comparison of the macronutrient contents of human and bovine milk

Component	*Human milk*	*Bovine milk*
Carbohydrates (g dl^{-1})		
Lactose	7.3	4.0
Oligosaccharides	1.2	0.1
Proteins (g dl^{-1})		
Caseins	0.2	2.6
α-Lactalbumin	0.2	0.2
Lactoferrin	0.2	Trace
Secretory IgA	0.2	Trace
β-Lactoglobulin	0	0.5
Milk lipids (%)		
Triglycerides	4.0	4.0
Phospholipids	0.04	0.04
Minerals and other ionic constituents (mM)		
Sodium	5.0	15
Potassium	15.0	43
Chloride	15.0	24
Calcium	7.5	30
Magnesium	1.4	5
Phosphate	1.8	11
Bicarbonate	6.0	5

sweating in human infants. The high concentration of casein in bovine milk provides protein and associated calcium and phosphate to support the very rapid growth of the young calf. In general, casein is removed when bovine milk is used to produce infant formula designed for the much more slowly growing human infant. There are a number of agents in human milk that protect against gastrointestinal and respiratory infections, including the oligosaccharides that recently have been shown to interact specifically with pathogen receptors, lactoferrin, and secretory IgA. These compounds are present at very small concentrations in bovine milk and infant formula.

The Anatomy of the Breast

The secretory apparatus of the breast consists of 15–25 ducts extending from the nipple and coursing through the mammary fat pad to terminate in grape-like clusters of alveoli (Figure 1). Each duct serves a specific lobule. The lobules are separated and supported by thick connective tissue septa and, in the nonpregnant nonlactating breast, by large amounts of adipose tissue. Blood vessels, nerves, and lymphatics run in the septa, which merge imperceptibly with the fascia at the anterior thoracic wall. The nipple, which serves as the termination point for the lactiferous ducts, is surrounded by an area of pigmented skin, the *areola*, containing sebaceous glands and sweat glands. The areola serves as the termination point for the fourth intercostal nerve, which carries sensory information about suckling to the spinal cord and brain. This is extremely important in the regulation of oxytocin secretion from the posterior pituitary and prolactin from the anterior pituitary. The mammary ducts expand slightly to form sinuses beneath the areola.

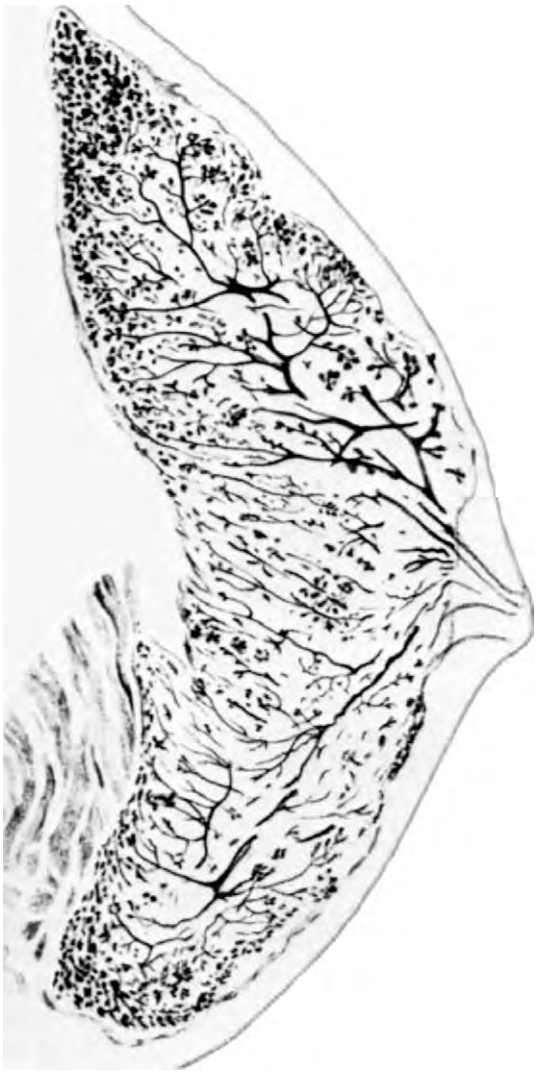

Figure 1 Camera lucida drawing of a sagital section through the breast of a 19-year-old nulligravida. From Dabelow A (1941) Die postnatale Entwicklund der menschlischen Milchdruse und ihr Korrelationen. *Morphology Journal* 85: 361–416, with permission.

Histologically, the secretory compartment of the breast consists of ducts and alveoli. The larger ducts appear to have two layers of cuboidal epithelium. Smaller ductules and alveoli have a single luminal epithelial layer surrounded by a discontinuous layer of contractile cells, the *myoepithelium* (Figure 2). This latter network of cells forms a basket-like framework around the alveoli and runs longitudinally along the ducts. Myoepithelial cells contract in response to oxytocin, ejecting milk from the alveoli, where it is stored after secretion by the luminal

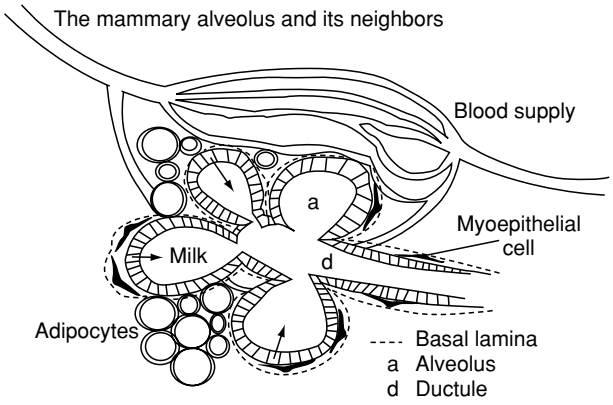

Figure 2 Cartoon showing the relation of the mammary epithelial compartment composed of mammary ducts and alveoli to the other anatomical compartments of the gland: the myoepithelial cells, the stroma composed of adipocytes and fibroblasts, and the extensive blood supply. Milk secretion by the epithelial cells into the alveolar lumen is indicated by arrows.

epithelial cells. Adipocytes are intimately associated with the glandular elements, as is a prominent blood supply.

Although the location and external form of the mammary glands differ from one species to another, the mechanisms of milk production are remarkably similar. Milk is produced and stored in alveolar units like that shown in Figure 2. Removal of the milk is accomplished by a process called *milk ejection*, during which, milk is forced from the alveoli by contraction of surrounding myoepithelial cells. Milk ejection is often called the 'let-down reflex,' and the milk exits through ductules into ducts draining several clusters of alveoli. In the human, the small ducts coalesce into 15–25 main ducts that drain sectors of the gland. The main ducts dilate into small sinuses as they near the areola, where they open directly on the nipple. For milk removal, the entire areola with the underlying milk sinuses is drawn into the infant's mouth, and the milk is withdrawn by the stripping action of the tongue.

In comparison with related dermal glands such as the salivary and sweat glands, the rate of milk secretion is slow, about 1.5 ml of milk per gram of tissue per day, and the secretory product is stored in the alveolar spaces until it is forced out by myoepithelial cell contraction. The larger ducts play a passive role in milk secretion, merely transferring the milk from the alveolar stores to the subareolar sinuses, where it is available to the suckling infant. Because the composition of the aqueous phase of milk changes very little during a feed or milking, it is unlikely that reabsorptive processes, like those important in the formation of saliva or sweat, play a significant role in determining milk composition.

Although the mammary epithelial cells are ultimately responsible for converting most precursors into milk constituents and transporting them to the mammary lumen, other cell types are also intimately involved in milk production, as illustrated in Figure 2. We have already mentioned the myoepithelial cells responsible for milk ejection from the breast. The mammary ducts and alveoli are embedded in a stroma that contains fibroblasts, adipocytes, plasma cells, and blood vessels. Blood flow is greatly expanded during lactation to make available the large amounts of substrate required for milk synthesis. During lactation, B lymphocytes 'home' in to the mammary gland, where they become plasma cells and settle in the interstitial space producing the immunoglobulins that ultimately find their way into milk. Stromal cells synthesize the growth factor IGF-1, which may promote survival of mammary epithelial cells. The mammary epithelium should, therefore, be viewed as an integrator of the activities of many cells and organs that contribute in a coordinated fashion to the synthesis and secretion of milk and its ejection from the gland.

Mechanisms of Milk Secretion

Figure 3 shows a single mammary epithelial cell in a lactating mammary gland. It is positioned on a basement membrane, and processes of myoepithelial cells course along its basal surface. Tight junctions join this cell to its neighbors, and in lactation, they prevent the direct transfer of interstitial fluid

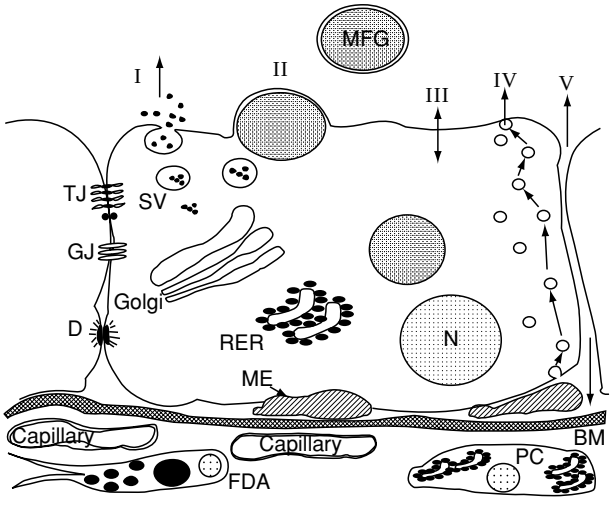

Figure 3 Mammary epithelial cell showing the pathways for the secretion of milk components. See text for further information. SV, secretory vesicles; RER, rough endoplasmic reticulum; ME, myoepithelial cell process; TJ, tight junction; GJ, gap junction; D, desmosome; BM, basement membrane; FDA, fat-depleted adipocyte; PC, plasma cell.

components to the alveolar lumen. Four major *transcellular* pathways are responsible for the secretion of milk components during lactation. The *paracellular* pathway is open during pregnancy and, under certain other conditions, allowing movement of interstitial fluid components into the milk space and the flux of milk components in the other direction.

Pathway I: Exocytosis

The exocytotic pathway secretes milk components by a mechanism similar to exocytotic secretion in all endocrine and exocrine glands. Casein and α-lactalbumin are synthesized and secreted via this largely constitutive secretory pathway, which is also responsible for the secretion of lactose, calcium, phosphate, citrate, and most other components of the aqueous phase of milk. Lactose is synthesized in the terminal portion of the Golgi vesicles from glucose and UDP (uridine diphosphate)-galactose by the enzyme galactosyl transferase using α-lactalbumin as an essential coenzyme. Lactose contributes two-thirds of the osmolarity of human milk. Various glycosyl transferases responsible for the formation of the wide variety of complex oligosaccharides present in human milk are also found in the terminal Golgi.

Pathway II: Lipid Secretion

The mechanism for the secretion of lipid is unique to the mammary epithelium. Triacyglycerols are synthesized in the mammary alveolar cell from glycerol and free fatty acids. The free fatty acids are derived from the plasma lipid using the enzyme lipoprotein lipase, from plasma free fatty acids carried there on albumin or by synthesis within the alveolar cell from glucose. Free fatty acids synthesized in the mammary alveolar cell are medium-chain fatty acids with 12 or 14 carbons rather than the long-chain fatty acids derived from the plasma. Once synthesized, the triacylglycerols form droplets, merge, and gradually move to the apical membrane. When they reach the apical surface of the cell, they press on the membrane, gradually becoming enveloped, and eventually pinch off as membrane-bound milk fat globules. The membrane of the milk fat globule prevents coalescence of the fat droplets and serves to deliver both triglycerides and phospholipids to the infant.

The composition of milk triglycerides depends on the diet. On the usual high-fat diet consumed by the American woman, the fatty acid composition reflects dietary lipid with long-chain fatty acids predominating and a fair proportion of unsaturated fatty acids reflecting the consumption of vegetable oils (Table 2). In this condition, only about 10% of the milk fatty acids are synthesized in the mammary gland itself as medium-chain fatty acids (C8–C14). On high-carbohydrate, low-fat diets, the proportion of fatty acid synthesized in the mammary gland from carbohydrate increases, leading to an increase in medium-chain fatty acids, as can be seen from the data from the milks of Nigerian women in Table 2. The milks of the Nigerian women also contained a higher proportion of *n*-3 LCPUFA. On reduced- or low-calorie diets, much of the milk fat is derived from adipose tissue; in this case, the composition of the triglycerides resembles depot fat.

Table 2 Major fatty acids of human milks (wt%)

Fatty acid		*Human milk*[a]	*Human milk*[b]
Saturated fatty acids			
Intermediate- and medium-chain (formed in mammary gland)			
8:0	Octanoic acid	0.46	
10:0	Decanoic acid	1.03	0.54
12:0	Lauric acid	4.40	8.34
14:0	Myristic acid	6.27	9.57
Long chain			
16:0	Palmitic acid	22.00	23.35
18:0	Stearic acid	8.06	10.15
Monounsaturated fatty acids			
16:1 *n*-7 (*cis*)	Palmitoleic acid	3.29	0.91
18:1 *n*-9 (*cis*)	Oleic acid	31.30	18.52
18:1 *n*-9 (*trans*)		2.67	0.86
Polyunsaturated fatty acids (PUFA) (essential fatty acids)			
18:2 *n*-6	Linoleic acid	10.76	11.06
18:3 *n*-3	Linolenic acid	0.81	1.41
Long-chain PUFA (n-6)			
18:3 *n*-6	γ-Linolenic acid	0.16	0.12
20:2 *n*-6		0.34	0.26
20:3 *n*-6	Dihomo-γ-linolenic acid	0.26	0.49
20:4 *n*-6	Arachidonic acid	0.36	0.82
Long-chain PUFA (n-3)			
20:5 *n*-3	Eicosapentaenoic acid	0.04	0.48
22:5 *n*-3		0.17	0.39
22:6 *n*-3	Docosahexaenoic acid	0.22	0.93

[a]Data from Jensen RG (1995) *Handbook of Milk Composition*, San Diego, CA: *Academic Press*.
[b]Data from Nigerian women who also had a diet high in fish, as reflected in the *n*-3 LCPUFA.

Pathway III: Transmembrane Transport

There are transport mechanisms for only a few milk components in the apical membrane of the mammary alveolar cell. This membrane is known to be permeable to monovalent ions such as sodium, potassium, and chloride, as well as to glucose. It is impermeable to divalent cations and disaccharides such as lactose. The mechanisms that control the substantial concentration gradients of monovalent cations across this membrane are not understood.

Pathway IV: The Transcytotic Pathway

Proteinaceous substances from the interstitial space find their way into milk via *transcytotic* pathways.

One of these pathways, that for the secretion of IgA, is well understood. An IgA receptor on the basolateral membrane of the cell interacts with dimeric IgA made by plasma cells residing in the interstitial spaces of the mammary gland. The IgA is endocytosed and transported in a vesicle to the apical membrane, where it is secreted with a piece of its transporter (called 'secretory component') into the milk. The secretory component renders the protein resistant to protease, and the IgA thus has a reasonable survival time in the infant's intestine, where it may contribute to protection from infectious agents. Many other proteins that are not synthesized in the mammary alveolar cell enter milk via transcytosis, e.g., hormones such as insulin and prolactin, growth factors such as IGF-1 and even serum albumin.

Pathway V: Paracellular Transport

During lactation, tight junctions (Figure 3) form a seal between alveolar cells, closing the paracellular pathway and completely isolating the alveolar lumina from the interstitial spaces. In pregnancy, after involution and during mastitis, these junctions open by mechanisms that are not well understood. At these times, even large proteins can be transferred between the milk space and interstitial space. Under these conditions, plasma components like sodium, chloride, and albumin enter the milk space directly from the interstitial space, milk components like lactose enter the blood stream, and the concentrations of sodium and chloride in milk are high.

Cells from the immune system pass into milk via the paracellular pathway. They appear to be able to squeeze even through the tight junctions of the lactating gland.

Mammary Development

The mammary gland is one of the few organs that undergoes almost the entire cycle of development, differentiation, function, and involution in the adult animal. Each of these stages has different control mechanisms, only some of which are understood despite decades of study. The developmental cycle of the mammary gland is usually divided into four stages: (I) mammogenesis, or development, (II) sensory differentiation (III) secretory initiation, and (IV) lactation and (V) involution. Mammogenesis takes place in the embryo, and at puberty. Sensory differentiation may begin at late puberty and is completed during pregnancy. Secretory initiation begins at parturition and is completed during the first 3–5 days after birth in the human. Lactation lasts as long as milk continues to be removed from the breast after birth of the infant. Involution is the regression of the gland after suckling has been discontinued. Each phase has its own unique control mechanisms.

Embryogenesis

The mammary gland begins to develop in the fourth week of fetal life as a thickening of the epidermis that gradually begins to invade the underlying parenchyma forming an epithelial bud. This bud branches and canalizes forming about 15–20 rudimentary ducts consisting of a lumen surrounded by a layer of ductal epithelium with an underlying layer of myoepithelial cells. The entire process requires about 28 weeks, during which time, a fully committed epithelial rudiment and an underlying fat pad are formed, as well as the nipple. It is becoming clear that interactions between the epithelium and the underlying stroma are critical in determining the morphology of the structures formed. Towards the end of pregnancy, the mammary epithelium of the infant may become secretory under the influence of the maternal steroid hormones, and at birth, many infants secrete a small amount of milky fluid called 'witches' milk.

Postnatal Development

During childhood, the mammary gland does little more than keep pace with the general growth of the body.

Pubertal Development

One of the earliest signs of puberty in girls is enlargement of the breast. This enlargement involves both the epithelium and the adipose stroma, *an obligatory substratum for epithelial growth*. Transplant experiments in rodents have shown that the mammary epithelium grows only in white adipose tissue fat pads and best in the mammary fat pad. Under the influence of *estrogen* from the developing ovary, the ducts begin to elongate. The most direct evidence that estrogen is essential for ductal development comes from experiments in which time-release pellets were used to deliver drugs directly into the developing gland. Implantation of estrogen pellets enhances ductile development, whereas implantation of pellets that release antiestrogens into the local tissue environment inhibits ductal growth. Estrogen receptors are found in the stroma as well as in scattered epithelial cells.

Pregnancy

During pregnancy, the increasing concentrations of sex steroids along with prolactin and placental lactogen bring about maximal development of the breast. Lobular development intensifies until the fat pad is completely filled with epithelial structures. By midpregnancy, the alveolar cells have differentiated

sufficiently to be capable of secreting milk, a point in time called lactogenesis, Stage I. However, milk secretion is held in check by the high concentrations of progesterone from the placenta.

Secretory Initiation, Lactation, and Involution

With the birth of the young, the hormonal support for mammary growth and inhibition of milk secretion are withdrawn, and secretory initiation begins. This process involves the terminal differentiation of the mammary cells through a carefully programmed series of events that result in copious secretion of milk 4 days after birth. Milk secretion is maintained until regular removal of milk from the gland ceases, at which point, the mammary epithelium begins to involute. Many of the alveolar cells undergo apoptosis, and the gland returns to its prepregnant state.

Secretory Initiation

Although the mammary epithelium becomes fully competent to secrete milk sometime in midpregnancy, the term secretory initiation is used to describe the *onset of copious milk secretion in the first few days after parturition* and will be so used in the rest of this discussion. Withdrawal of progesterone and estrogen in the presence of sustained prolactin secretion brings about a complex series of well-programmed events that lead to the secretory activity of differentiated mammary cells. Immediately after birth, the tight junctions between the cells begin to close, and the paracellular pathway is no longer available for the exchange of components between the milk space and the interstitial space. The concentrations of sodium and chloride in the milk fall, and the milk lactose concentration increases (Figure 4). Next, secretion of IgA oligosaccharides and lactoferrin increases, adding large quantities of these protective substances along with high concentrations of lymphoid cells to the mammary secretion, now called *colostrum*. Beginning about 40 h postpartum, the alveolar cells increase their secretion of lactose, casein, and the other components of mature milk, and milk volume production goes from about 100 ml per day on the second day postpartum to about 500 ml per day on day 5. The corresponding feeling of fullness of the breast is often referred to as the 'milk coming in.' Suckling by the infant is not required to initiate the process; however, milk secretion cannot be maintained if milk is not removed shortly after the secretory activity increases.

Lactation

At the end of secretory initiation, the breast enters the stage of lactation, sometimes referred to as

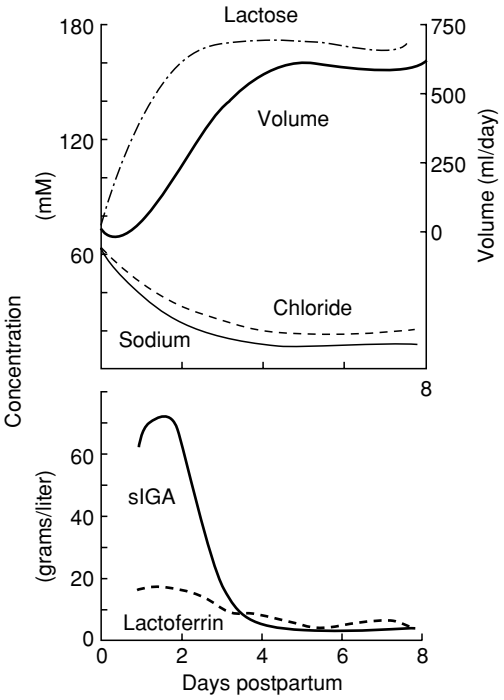

Figure 4 Changes in milk composition and volume during secretory activation in the woman.

galactopoiesis. This stage persists as long as the infant continues to suckle at least twice a day. The daily milk volume transferred to the infant increases on average from 500 ml on day 5 to about 650 ml at 1 month and about 750 ml at 3 months. Milk volume is largely regulated by infant demand. Infants who are larger at a given age tend to take more milk; if breast milk is supplemented with other foods or formula, milk volume decreases. In most women, the breast appears to have the capacity to secrete much more milk than is needed for nourishment of a single infant. In fact, wet nurses who removed milk from their breasts using a breast pump have been recorded as producing up to 3500 ml per day, and some women are able to breast-feed triplets!

Regulation of Milk Secretion

Prolactin is necessary for milk secretion, and suckling promotes its secretion. However, the volume of milk secretion is not directly regulated by the concentration of prolactin in the blood. Rather, local mechanisms within the mammary gland related to the amount of milk removed by the infant are responsible for the day-to-day regulation of milk volume. A protein factor called *feedback inhibitor of lactation* is secreted with other milk components into the alveolar lumen. If milk is not removed from the gland, this substance, whose identity is not yet entirely clear,

interacts with the mammary alveolar cell and inhibits milk secretion possibly by altering the sensitivity of the cells to prolactin.

Regulation of Milk Ejection

Suckling initiates a neuroendocrine reflex essential for removal of milk from the breast. Afferent impulses travel via sensory neurons from the areola to the hypothalamus, where they stimulate magnocellular neurons to fire, sending an impulse down their axons into the posterior pituitary where the hormone *oxytocin* is released. This hormone travels through the blood stream to the breast, where it causes contraction of the myoepithelial cells, forcing the milk from the alveoli into the mammary ducts and sinuses whence it can be removed by the sucking infant. Without an adequate let-down reflex, complete removal of milk from the breast is not possible. The activity of the magnocellular neurons can be profoundly influenced by higher brain centers. For this reason, emotional distress can inhibit let-down. Conversely, let-down is subject to conditioning, so that a woman often releases oxytocin at the sound of her (or someone else's) infant's cry or in response to a picture of her infant.

Interactions of Lactation with Maternal Conditions

The lactating breast is largely an autonomous organ, depending only on permissive levels of prolactin and other hormones such as insulin and hydrocortisone necessary for continued milk secretion. The amount of milk produced is under local control and depends on the rate of milk removal from the breast, as discussed above. The gland removes nutrients from the plasma according to its needs for substrate to synthesize milk components and is usually only marginally affected by acute changes in maternal dietary intake and other maternal factors such as fluid intake, body composition, illness, and exercise. The limits to which lactation is maintained in starvation are unknown, since experimental observations in this area are difficult. A single case report suggests that 7 days of total fasting does not impact milk secretion. The composition of the diet does, however, affect some milk components, most notably certain vitamins such as vitamin C, vitamin A, vitamin B_{12}, vitamin D, and possibly others. Vitamin B_{12} deficiency has been reported in infants of mothers consuming a vegetarian diet. The milk levels of some trace elements, for example, selenium are also affected by maternal dietary status. The effects of the lipid composition of the diet on milk fat composition have been discussed above. Fluid intake is not directly linked to milk production, as women who take no water during the day during Ramadan produce normal volumes of slightly hypertonic milk. It is likely that hydration severe enough to threaten maternal metabolism would also threaten milk production. The total fat content of milk does not appear to be affected by dietary intake but is inversely proportional to body fat. Diabetics may have difficulty initiating lactation, but once lactation is established, they have no problem nursing an infant. Mastitis alters milk composition by opening the tight junctions, so that the sodium and chloride content of the milk is increased and the protein content decreased. Pregnancy anecdotally reduces milk volume, although there are no good data on this point. Exercise has not been observed to alter either milk volume or composition.

The impact of lactation on maternal nutritional and reproductive status is more notable. Exclusive breast-feeding inhibits the release of gonadotropin-releasing hormone from the hypothalamus, postponing the return of reproductive function after birth of the infant. However, even full breast-feeding does not necessarily offer protection against a subsequent pregnancy, particularly in well-nourished women. Bone calcium has been shown to decline during lactation in several studies, probably owing to the lack of estrogen during the period of postpartum amenorrhea, but appears to recover to normal values once the menses have resumed. Women with marginal protein nutrition may suffer some deterioration of their protein status. During lactation, the folate stores of women with marginal folate status may be depleted of the vitamin, a situation that may have a negative impact on a subsequent pregnancy. However, compared with species like rodents or dairy animals, whose body lipid reserves are depleted within a short time after lactogenesis by their very copious milk production, metabolic adaptations to lactation in the woman are relatively small, because feeding a single infant does not place an enormous metabolic load on the lactating women compared with rodents or dairy animals.

See also: **Infants**: Nutritional Requirements; Breast- and Bottle-feeding; **Lactation**: Human Milk: Composition and Nutritional Value; **Pregnancy**: Metabolic Adaptations and Nutritional Requirements

Further Reading

Dewey KG and McCrory MA (1994) Effects of dieting and physical activity on pregnancy and lactation. *American Journal of Clinical Nutrition* 59(2 supplement): 446S–452S.

Jensen RG (1995) *Handbook of Milk Composition*. San Diego, CA: Academic Press.

Lawrence RA (1989) Breastfeeding and medical disease. *Medical Clinics of North America* 73: 583–603.
McNeilly AS, Tay CCK and Glasier A (1994) Physiological mechanisms underlying lactational amenorrhea. *Journal of Mammary Gland Biology and Neoplasia* 709: 145–155.
Neville MC (1989) Regulation of milk fat synthesis. *Journal of Pediatric Gastroenterology and Nutrition* 8: 426–429.
Neville MC (2001) Anatomy and physiology of lactation. *Pediatric Clinics of North America* 48: 13–34.
Neville MC and Neifert MR (1983) *Lactation: Physiology, Nutrition and Breastfeeding*. New York: Plenum Press.
Neville MC and Picciano MF (1997) Regulation of milk lipid secretion and composition. *Annual Review of Nutrition* 17: 159–184.
Newburg DS (1996) Oligosaccharides and glycoconjugates inhuman milk: Their role in host defense. *Journal of Mammary Gland Biology and Neoplasia* 1: 271–284.
O'Connor DL (1994) Folate status during pregnancy and lactation. *Advances in Experimental Medicine & Biology* 352: 157–172.
Peaker M and Wilde CJ (1996) Feedback control of milk secretion from milk. *Journal of Mammary Gland Biology and Neoplasia* 1: 307–316.
Prentice A (1994) Calcium intakes and bone densities of lactating women and breast-fed infants in The Gambia. *Advances in Experimental Medicine & Biology* 352: 243–255.
Russo J and Russo IH (1987) Development of the human mammary gland. In: Neville MC and Daniel CW (eds) *The Mammary Gland*, pp. 67–96. New York: Plenum Press.
Sanz AM and Diaz RC (1995) Selenium in human lactation. *Nutrition Reviews* 53(6): 159–166.
Specker BL (1994) Nutritional concerns of lactating women consuming vegetarian diets. *American Journal of Clinical Nutrition* 59(5 supplement): 1182S–1186S.
Telemo E and Hanson LA (1996) Antibodies in milk. *Journal of Mammary Gland Biology and Neoplasia* 1: 243–250.

LACTIC ACID BACTERIA

J A Narvhus, Agricultural University of Norway, Aas, Norway
L Axelsson, MATFORSK, Norwegian Food Research Institute, Aas, Norway

Introduction

Lactic acid bacteria (LAB) are responsible for a great diversification in the flavor and texture of food products due to their fermentation of food raw materials. However, in some circumstances they can be responsible for food spoilage. Food fermentation technology, along with drying and salting, is one of the most ancient food-preserving techniques known to humans. Fermented foods are less perishable than the original raw materials, their nutritional value may be enhanced, and the safety of these foods may be improved due to the inhibition of pathogenic bacteria by the low pH and the presence of organic acids and antimicrobial compounds. The ubiquity of LAB in nature often results in opportunistic inoculation of food raw materials. Humans have, since prerecorded times, aided such natural events by purposely adding a portion of previously fermented food to the new batch of raw materials (so-called 'back-slopping') in order to promote successful fermentation. In addition, specific traditional technologies often lead to the dominance of particular types of LAB, which gives the product special characteristics. (*See* **Fermented Foods**: Origins and Applications; **Traditional Food Technology**.)

Today, many traditional fermentation technologies have been industrialized and form large commercial enterprises. Examples of these are the production of cheese and other fermented dairy products, fermented sausages such as salami, and fermented cabbage – sauerkraut. LAB are only weakly proteolytic and lipolytic, which is also important since breakdown products from protein and fat are often organoleptically unpleasant in high concentrations. LAB also inhabit the mucous membranes of the intestinal and reproductive tracts of humans and animals and incorporation of selected strains into certain fermented foods has gained increasing interest in recent years. LAB have a reputation of being 'generally recognized as safe' (GRAS). However, some species are pathogenic and the recent developments in the field of probiotics have brought safety aspects of LAB into focus. Aside from the known pathogens, LAB are rarely isolated from disease and in almost all cases are associated with underlying illness which increases the susceptibility of the host.

LAB were amongst the very first bacteria studied. In 1873, Joseph Lister isolated the first bacterial pure culture which he called *Bacterium lactis*. This lactic

acid bacterium is now called *Lactococcus lactis* and is used for fermenting milk to produce hundreds of different dairy products. Early work on LAB was mainly concentrated on those associated with dairy products and, with time, many commercial starter cultures were developed. However, LAB were soon discovered in other habitats and a broader view of the group developed. From concentrating on the reactions of the organisms in milk, interest was diverted into more diverse attributes and the taxonomy of the group was gradually mapped. (*See* **Cheeses**: Starter Cultures Employed in Cheese-making; **Fermented Milks**: Types of Fermented Milks; **Yogurt**: The Product and its Manufacture.)

However, the relationships between different LAB genera have seen many changes. Orla Jensen described seven genera in 1919, but only the name of one of them, *Streptococcus*, remains valid today. The LAB group now comprises 16 genera, 12 of which are associated with food (Table 1). All LAB have a G+C% under 50. Morphologically, they are cocci, coccobacilli, or rods and, with the exception of the tetrad-forming genera *Aerococcus, Pediococcus*, and *Tetragenococcus*, chain formation is common. LAB are Gram-positive, nonsporing, and most species are nonmotile. They have complex nutritional requirements and are dependent on the presence of a fermentable carbohydrate for active growth. As an end product of this fermentation, LAB produce copious amounts of lactic acid either homofermentatively (> 95% lactic acid from glucose) or heterofermentatively (producing acetic acid, ethanol, and carbon dioxide in addition to lactic acid). Under conditions of limiting carbohydrate, fermentation by homofermentative strains may become mixed. The group lacks the ability to synthesize porphyrin groups and they thus have a nonrespiring (cytochrome-negative) metabolism and are catalase-negative in standard laboratory media. It should be mentioned here that the genus *Bifidobacterium* is often regarded as being a member of the LAB group as it fits the general description above. In addition, certain strains of bifidobacteria are used in probiotic applications, much in the same way as some LAB. However, bifidobacteria are phylogenetically unrelated to LAB and also differ physiologically in that they utilize a sugar fermentation pathway unique to the genus. (*See* ***Bifidobacteria* in Foods**.)

The separation of the various genera of LAB from each other using conventional biochemical tests has always been problematic, since a certain overlapping of properties exists (Table 1). Nevertheless, the allocation of most LAB isolates to the correct genus should be possible by the use of the characterization tests shown in the table.

Within each genus, classical division of LABs into species relies heavily on carbohydrate fermentation profiles and some other simple physiological tests such as growth at particular temperatures and tolerance of salt and extremes of pH. However, the development of genotyping techniques has to a certain extent revolutionized our understanding of the relationships within the group. Some of these techniques can also be used to identify and characterize isolates down to species and even strain level. Figure 1 shows the changes in nomenclature within LAB important in foods, from 1980 to 2000. Figure 2 shows a schematic, unrooted phylogenetic tree of the lactic acid

Table 1 Differential characteristics of lactic acid bacteria[a]

Character	*Rods*		*Cocci*							
	Carnobacterium	*Lactobacillus*	*Aerococcus*	*Enterococcus*	*Lactococcus Vagococcus*	*Leuconostoc Oenococcus*	*Pediococcus*	*Streptococcus*	*Tetragenococcus*	*Weissella*[b]
Tetrad formation	−	−	+	−	−	−	+	−	+	−
CO_2 from glucose[c]	−[e]	±	−	−	−	+	−	−	−	+
Growth at 10 °C	+	±	+	+	+	+	±	−	+	+
Growth at 45 °C	−	±	−	+	−	−	±	±	−	−
Growth in 6.5% NaCl	ND[f]	±	+	+	−	±	±	−	+	±
Growth in 18% NaCl	−	−	−	−	−	−	−	−	+	−
Growth at pH 4.4	ND	±	−	+	±	±	+	−	−	±
Growth at pH 9.6	−	−	+	+	−	−	−	−	+	−
Lactic acid[d]	L	D, L, DL[g]	L	L	L	D	L, DL[g]	L	L	D, DL[g]

[a] +, positive; −, negative; ±, response varies between species; ND, not determined.
[b] *Weissella* strains may also be rod-shaped.
[c] Test for homo-or heterofermentation of glucose; negative and positive denotes homofermentative and heterofermentative, respectively.
[d] Configuration of lactic acid produced from glucose.
[e] Small amounts of CO_2 can be produced, depending on media.
[f] No growth in 8% NaCl has been reported.
[g] Production of D-, L-, or DL-lactic acid varies between species.

Reproduced from Axelsson L (1998) Lactic acid bacteria: classification and physiology. In: Salminen S and von Wright A (eds) *Lactic Acid Bacteria. Microbiology and Functional Aspects*, 2nd edn. New York: Marcel Dekker with permission from Marcel Dekker.

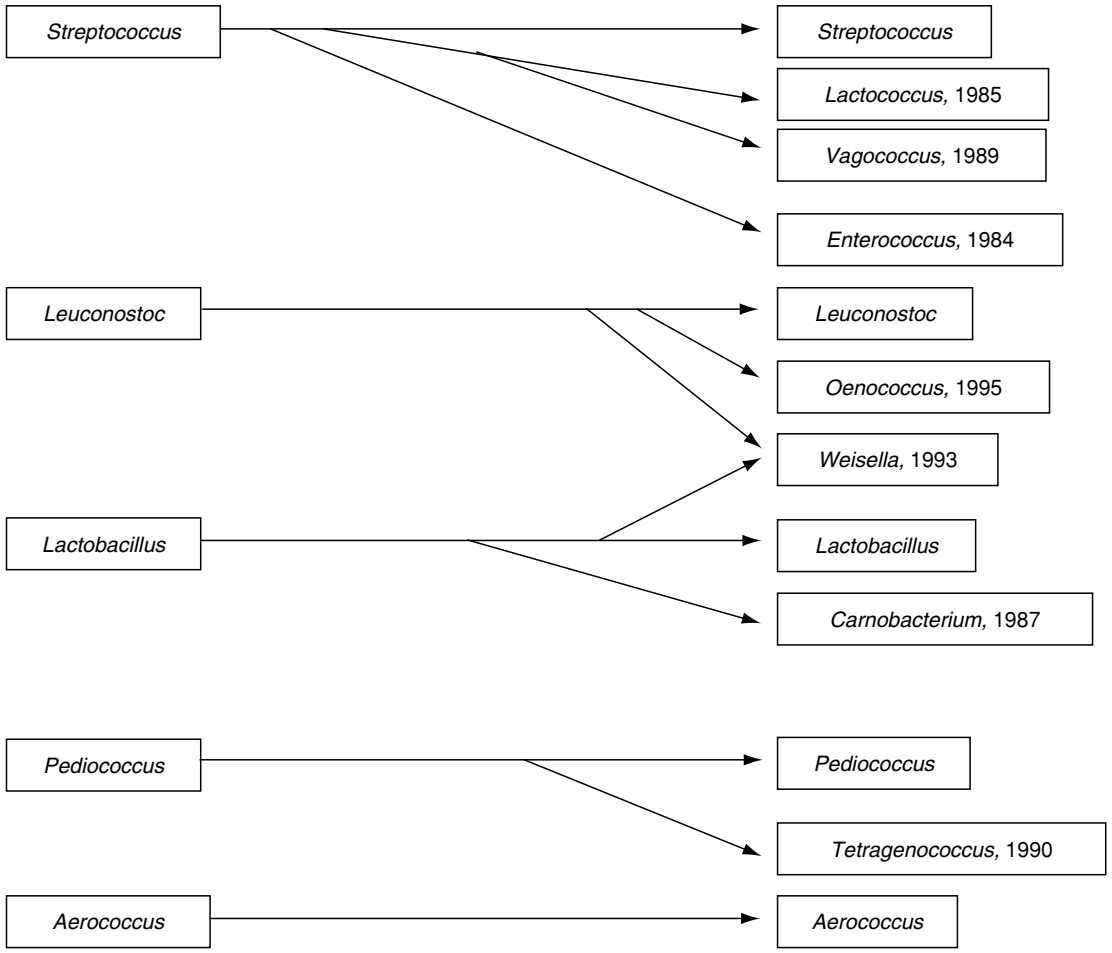

Figure 1 Genera of lactic acid bacteria important in food, showing changes in nomenclature from 1980 to 2000.

bacteria. Phylogenetically, LAB belong to a cluster of Gram-positive bacteria which also includes aerobic and facultative anaerobic genera such as *Bacillus* and *Staphylococcus*. So far, not one genotyping method has achieved universal acceptance but these methods are gradually partially replacing the traditional characterization tests. However, molecular biological techniques alone have the disadvantage that they give no information concerning the properties of the organisms under study. For this reason, studies of the metabolic properties of LAB, especially with relation to food product characteristics, are also receiving increasing attention.

Descriptions of Selected Genera

The Genus *Streptococcus*

The genus *Streptococcus* is one of the original genera of LAB and was heterogeneous. Taxonomic studies in the 1980s, however, supported the unofficial subgrouping of the genus and concluded that the fecal and lactic species were sufficiently different to deserve allocation to their own genera of *Enterococcus* and *Lactococcus*. However, most streptococci are not food organisms; most are commensals of humans and animals and, indeed, many are pathogens or opportunistic pathogens. The notable exception to this is *Streptococcus thermophilus*. This species is characterized by its rapid growth at high temperatures and active fermentation of milk. It is used, together with *Lactobacillus delbrueckii* ssp. *bulgaricus*, for the production of yogurt and is found in many indigenous fermented milks, where it may occur along with a variety of other LAB and yeasts. It is also used as a starter culture for a variety of Italian cheeses, including Parmesan. No other species of *Streptococcus* is reported to be important for either the manufacture or spoilage of food stuffs. (*See* **Yogurt**: The Product and its Manufacture.)

The Genus *Lactococcus*

Some species of the genus *Lactococcus*, previously grouped within *Streptococcus* and having the group N

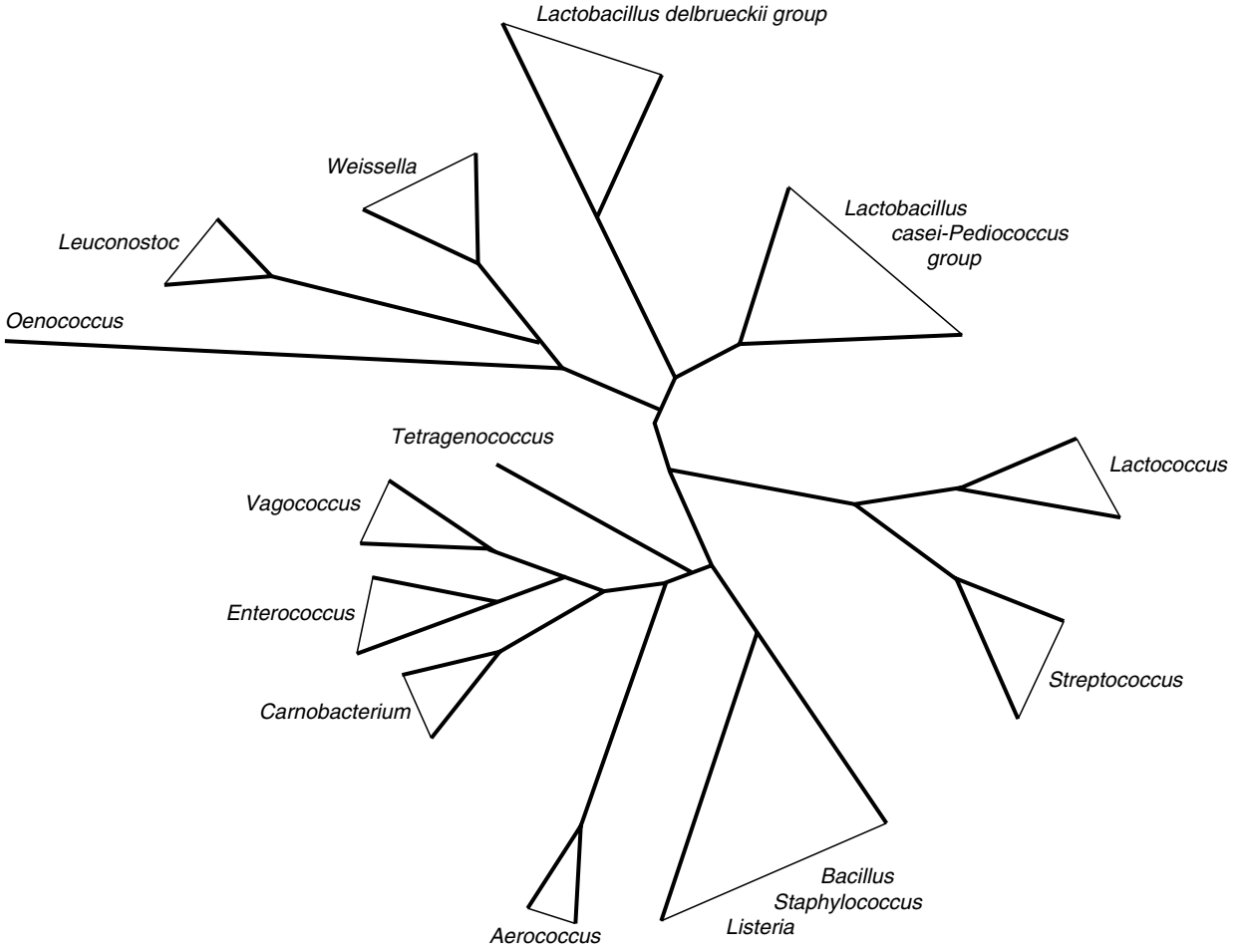

Figure 2 Schematic, unrooted phylogenetic tree of the lactic acid bacteria, including some aerobic and facultative anaerobic Gram-positives of the low G+C subdivision. Note: evolutionary distances are approximate. Reproduced from Axelsson L (1998) Lactic acid bacteria: classification and physiology. In: Salminen S and von Wright A (eds) *Lactic Acid Bacteria. Microbiology and Functional Aspects*, 2nd edn. New York: Marcel Dekker with permission from Marcel Dekker.

antigen, are extremely important in the dairy industry. They can be isolated from spontaneously fermented milks and, more importantly, are an essential part of commercial dairy starter cultures used for the manufacture of fermented milks, sour cream, fermented butter, unripened cheeses (such as cottage and cream cheeses), and also for a multitude of vastly varying matured cheeses. Lactococci are often used in coculture with *Leuconostoc* strains. *L. lactis* is the most commercially important *Lactococcus*. It has two subspecies – *cremoris* and *lactis* – and the latter subspecies has a biovariant – *diacetylactis*. The two subspecies differ from each other in a variety of biochemical characterisation tests; the subspecies *cremoris* is markedly less reactive than *lactis* but it has the reputation of being a better flavor producer in cheese. In the nondairy environment, strains of *lactis* are isolated more frequently than *cremoris*. Some strains of *cremoris* form extracellular polysaccharides and are found in traditional Scandinavian ropy milks. (*See* **Cheeses:** Starter Cultures Employed in Cheese-making; **Fermented Milks:** Types of Fermented Milks; **Starter Cultures.**)

The biovariant *diacetylactis* deserves a special mention since its one distinguishing feature – the ability to metabolize citrate – results in the formation of diacetyl in fresh fermented milk products and in the production of carbon dioxide which forms eyes in Gouda-type cheeses. Other species of *Lactococcus* are not considered important in the food industry.

The Genus *Vagococcus*

The genus *Vagococcus* is phenotypically indistinguishable from lactococci and also possesses the group N antigen. However, vagococci are often motile; they have been isolated from chicken feces, water, and diseased fish.

The Genus *Enterococcus*

Cocci which are Gram-positive, catalase-negative, possess the group D antigen, and are able to grow in

6.5% salt and in the presence of 0.4% sodium azide, can be identified presumptively as enterococci. This genus was previously grouped within *Streptococcus*. Certainly all the classical species of *Enterococcus* fall into this category, but some newly described species have some aberrant properties which has resulted in the unofficial division into five subgroups. Several species of *Enterococcus* are naturally found in the intestinal tract of humans and animals. The role of enterococci in food is unclear and much debated. These organisms have been isolated from many indigenous fermented foods, but their possible positive contribution has not been elucidated since such products commonly contain several species of LAB. The tolerance of members of the genus for extremes of pH, salt, and temperature makes their survival in fermented foods unproblematic. The fecal original of many species, such as *E. faecalis* and *E. faecium*, however, means that their presence in foods is regarded as being an indication of poor hygiene. Further complications are the opportunistic pathogenicity of some species of *Enterococcus*, their ability to decarboxylate amino acids and thereby produce amines in foods, and finally their propensity to be resistant to antibiotics and to transfer such traits by means of mobile genetic elements. A new area of debate concerns the advisability of using these organisms as probiotics.

The Genus *Leuconostoc*

All leuconostocs are heterofermentative cocci. Their main habitat is on plants and, to a lesser extent, in meat, milk, and dairy products. As in other genera of LAB, the member species in this genus are continually being revised, but this is a comparatively small genus, with about 11 species. The most common species are *Leuconostoc mesenteroides*, which has three subspecies: *mesenteroides, cremoris*, and *dextranicum*, and *Leuc. lactis*. Several other species at present placed in the genus *Leuconostoc* show poor genotypic homology with the above species, and also with each other. Leuconostocs play an important role in the early stages of the fermentation of vegetable products such as dill pickles, olives, and sauerkraut. They take over from the initial, mainly Gram-negative flora and are important for the development of the characteristic flavor of the products. Later in the fermentation other more aciduric LABs dominate.

Several leuconostocs produce extracellular polysaccharides which can be used as food-thickening agents or stabilizers. However, these dextrans cause problems in sugar factories and also in fermented sausages, where the production of slime is undesirable. Some mesophilic dairy starter cultures contain leuconostocs. They produce flavor compounds (diacetyl and acetate) from the metabolism of citrate and reduce acetaldehyde to ethanol. In addition, production of CO_2 from both citrate and carbohydrates is important for the eyes in Gouda-type cheeses and the slight effervescence in cultured buttermilk.

The Genus *Oenococcus*

Oenococcus oeni (previously *Leuconostoc oenos*) is the only member of this new genus. It is found in wine and is responsible for the malo-lactic fermentation which occurs during the maturation of wine. *O. oeni* converts L-malic acid to L-lactic acid, thus increasing the pH by 0.1–0.3 pH units, and rendering the wine microbiologically stable. Traditionally present as a natural contaminant in wine, this organism may be purposely introduced to some wines to insure the desired maturation.

The Genus *Lactobacillus*

Members of the genus *Lactobacillus* are rod-shaped and number over 60 species. Despite the fact that several species have recently been removed to two new genera (*Weissella* and *Carnobacterium*), the genus is still heterogeneous. Lactobacilli are found in a variety of habitats and vary considerably in their characteristics. In 1919, Orla-Jensen grouped the lactobacilli into three subgenera: *Thermobacteria, Streptobacteria*, and *Betabacteria*. These subgenera are currently not taxonomically valid, but this subdivision is still useful to divide the genus into three primary phenotypic groups: the obligately homofermentative; the facultatively heterofermentative; and the obligately heterofermentative.

Lactobacilli are naturally found in nutritive and carbohydrate-rich habitats such as plants or material of plant origin, meat, milk, and dairy products. They are also found in manure, silage, and sewage. The mucous membranes (intestinal, vaginal, and oral) of humans and animals are often dominated by lactobacilli, where they are thought to have a probiotic function. (*See* **Microflora of the Intestine: Probiotics.**)

In the food industry, lactobacilli may be responsible for the desired fermentation of some products but also for the spoilage of others. In dairy foods, *Lb. delbrueckii* ssp. *bulgaricus* is used as a starter culture for yogurt and a variety of cheeses, particularly those manufactured using a high cooking temperature. During the ripening of most cheeses, so-called 'non-starter' lactobacilli (NSLAB) develop and are important for the development of the cheese flavor. Present research focuses on identifying useful strains with the prospect of greater control over cheese ripening. Heterofermentative lactobacilli may cause unwanted openness in some cheeses due to CO_2 production from residual carbohydrate.

Lactobacilli play an important role in the fermentation of many cereal products, such as sourdough, Tanzanian *togwa* and Nigerian *ogi*, where they produce certain flavor compounds and reduce the pH to as low as 3.5. Several species of lactobacilli are naturally found in fermented sausages such as salami, and commercial starter cultures for such products commonly contain *Lb. plantarum*. However, they may also cause the spoilage of refrigerated fresh and cooked meats kept in a CO_2-enriched atmosphere. Although this may inhibit the growth of other spoilage organisms and also pathogens, defects such as slime, green color, gas, and acidic taste may develop. Various fermented vegetables also contain lactobacilli. The early stages are dominated by leuconostocs and/or heterofermentative lactobacilli, but the final product is usually dominated by more acid-tolerant, homofermentative species.

The Genus *Carnobacterium*

The genus *Carnobacterium* is comparatively new, comprising some former lactobacilli which do not grow on acetate agar or at pH 4.5. Phylogenetically, however, they are most closely related to enterococci. They have so far been isolated from five habitats: meat (and meat products), cheese, fish, poultry, and sea water. Most reports are of isolates from meat that is stored vacuum-packed or in a CO_2/nitrogen atmosphere at refrigeration temperatures.

The Genus *Weissella*

Weissella was created when some aberrant species of *Leuconostoc* and heterofermentative *Lactobacillus* were found to be genotypically related. Species within this genus have been isolated from habitats such as meat and dairy products, sewage, and plants.

The Genera *Pediococcus, Aerococcus*, and *Tetragenococcus*

Pediococci, along with aerococci and tetragenococci, divide in two perpendicular directions (often wrongly described as 'dividing in two planes'), and thereby form tetrads of cells rather than chains. *Tetragenococcus* contains strains formerly regarded as *P. halophilus*. This species is characterized by its extreme tolerance to salt (>18%) and is important in the fermentation of high-salt-containing foods such as soy sauce. Aerococci are of minor interest in food technology and will not be further described here. Pediococci are found, along with other LAB in fermented vegetables, meat and fish. *P. damnosus* (also called *P. cerevisciae*) is of considerable nuisance in the brewing industry where it can cause 'ropy' beer due to capsular material and flavor defect due to diacetyl production. Pediococci do not grow well in milk but can be isolated from milk and dairy products. They may possibly be significant in the maturation of cheese where they can be found as part of the NSLAB community.

Commercial Applications of LAB in the Food Industry

The occurrence of specific LAB in various fermented foods has been outlined above. It is important to emphasize that there is a great difference in the degree of sophistication in the use of these bacteria. As quality control regulations for all foodstuffs become more stringent, the demand for control of fermented foods is also increasing. Whilst in the dairy industry commercial starter cultures have been available for many decades, their introduction into other fermented food products is either relatively new or is not yet in evidence. (*See* **Starter Cultures**.) This applies in particular to the countless varieties of fermented foods produced at artisan level in developing countries. Many of these products have not yet been scientifically described and the microorganisms responsible for the fermentation processes have not been identified. Despite considerable research into the metabolism of LAB, details of their various important metabolic activities in well-known products are only partially mapped.

The use of starter cultures in the dairy industry is a special case since milk pasteurization results in the destruction of the indigenous LAB. To achieve a successful fermentation it is therefore necessary to add suitable cultures of LAB. The development of the commercial starter culture industry has led to greatly increased quality and stability of dairy products and also for the few other fermented foods where they are utilized.

There is currently considerable research on so-called 'indigenous' fermented foods. These products are produced by natural fermentation and usually the responsible organisms create a complex mixture of various LAB, sometimes also interacting with other bacteria and/or fungi. Present studies are gradually elucidating the LAB present in these products, their interactions, and attempting to identify them. Although many of the strains isolated from these products resemble known species of LAB, they often have special properties which may make them a useful source of gene material for modifying known commercial cultures.

Bacteriophage Infection of Cultures

In 1935, fermentation problems in the dairy industry were diagnosed as being due to virus particles –

bacteriophages – attacking and destroying the starter culture. This has remained one of the most persistent problems in the dairy industry, despite the introduction of many innovative technological procedures to prevent their introduction to the process, and despite the wealth of research into phage–host interactions and phage resistance systems. The mode of action of phages, at a molecular level, is now known and the future will probably see the development of commercially available cultures that have been engineered to be resistant to phage attack. This methodology will also be of use in the future development of new starter cultures for controlled fermentations of other products where the availability of a selection of strains showing identical fermentation properties, but different phage susceptibility, may be problematic.

Instability of Starter Cultures

Variation in the properties of starter cultures has been a problem since their introduction, and has two main causes.

Natural, heterogeneous starter cultures, often produced by commercialized back-slopping techniques, show a certain variation in their fermentation properties from production to production. This is due to the variation in the balance of strains in the mixture, which may be caused by differences in food raw materials or culture media, in fermentation conditions or by interreactions between the composite strains. Since the different strains present in the culture may well have specific roles to play in product characteristics, an imbalance between the strains will result in unstable product quality.

Many LAB bear some of their genetic material on extracellular DNA particles called plasmids. Plasmids in lactococci have been extensively studied and they have also been found in many other LAB. Although the function of many plasmids is cryptic, in lactococci they code for technologically important properties such as bacteriophage resistance mechanisms, proteinase activity, and the ability to break down lactose and citrate. Loss of plasmids may occur during cell division, and this naturally results in the loss of the specific gene and the culture becomes commercially useless. In the commercial preparation of starter cultures the propagation conditions are carefully controlled so as to minimize the risk of plasmid loss.

Bacteriocins from LAB

Strains from almost all genera of LAB have been shown to produce many different proteinaceous antimicrobial substances, collectively called bacteriocins. Most of these are cationic, amphipathic peptides and will insert into the cell membrane of closely related bacteria, thereby leading to inhibition of growth by pore formation. The best known bacteriocin is nisin, and this is used as a food preservative in more than 50 countries. (*See* **Nisin**.) The 1990s have seen considerable research in this field, aimed at finding bacteriocins active against foodborne pathogens. Many bacteriocins have been thoroughly characterized and the genetics of their biosynthesis has been elucidated. However, commercial application of these compounds – or the bacteria producing them – has, apart from nisin, so far not materialized to any large extent.

Probiotic LAB

Since Metchnikoff's work in 1908, the LAB normal microbial flora of the body has been believed to have a beneficial effect on the well-being of the host. There is increasing interest worldwide for developing food products which contain strains of LAB which have been isolated from the human gut – so-called probiotic products. The majority of products on the market are fermented dairy products, and some fermented cereals have been developed. The commercial potential of such products is dependent upon the availability of convincing scientific documentation of the positive effect of ingesting these organisms. When based on well-designed double-blind experiments, there have been many recent reports showing significant effect on the balance of the gut microflora and also some that indicate that probiotic bacteria can promote a general stimulation of the immune system. Most countries demand documented proof of the efficacy of such products before claims can be made in promotion advertising. (*See* ***Acidophilus*** **Milk**; ***Bifidobacteria*** **in Foods**; **Probiotics**.)

Genetic Modification of LAB

Most LAB can now be genetically modified through recombinant DNA technology, although the degree of advanced modification that can be done varies between species and strains. For *Lactococcus lactis*, the most studied and probably the most economically important species, the available techniques are very advanced indeed. Essentially any protein of choice can be highly expressed in a controlled way, and this can be done by chromosomal integration in a so-called food-grade manner (i.e., no antibiotic resistance markers or other unwanted genetic elements are used in the procedure). It is now possible to change the properties of starter strains, for instance aroma formation capability or enhanced antimicrobial activity against pathogens. The introduction of

such strains into the market depends more on legislation and consumer acceptance than the construction of the strains. It should be noted that some properties can be transferred between strains by classical genetic methods, e.g., conjugation, and that this is still an important tool in starter development since these techniques are unproblematic from a legislative and consumer point of view.

See also: ***Acidophilus* Milk**; ***Bifidobacteria* in Foods**; **Cheeses**: Starter Cultures Employed in Cheese-making; **Fermented Foods**: Origins and Applications; **Fermented Milks**: Types of Fermented Milks; **Microflora of the Intestine**: Probiotics; **Nisin**; **Probiotics**; **Starter Cultures**; **Traditional Food Technology**; **Yogurt**: The Product and its Manufacture

Further Reading

Axelsson L (1998) Lactic acid bacteria: classification and physiology. In: Salminen S and von Wright A (eds) *Lactic Acid Bacteria. Microbiology and Functional Aspects*, 2nd edn. New York: Marcel Dekker.

Board RG, Jones D and Jarvis B (eds) (1995) *Microbial Fermentations: Beverages, Foods and Feeds*. Oxford, UK: Blackwell Science.

Ross RP, Galvin M, McAuliffe O *et al*. (1999) Developing applications for lactococcal bacteriocins. *Antonie van Leeuwenhoek International Journal of General and Molecular Microbiology* 76: 337–346.

Sanders ME and Huis in't Veld J (1999) Bringing a probiotic-containing functional food to the market: microbiological, product, regulatory and labelling issues. *Antonie van Leeuwenhoek International Journal of General and Molecular Microbiology* 76: 293–316.

Stiles ME and Holzapfel WH (1997) Lactic acid bacteria of foods and their taxonomy. *International Journal of Food Microbiology* 36: 1–29.

Wood BJB (ed.) (1998) *Microbiology of Fermented Foods*, 2nd edn. London: Blackie Academic & Professional.

Wood BJB and Holzapfel WH (eds) (1995) *The Genera of Lactic Acid Bacteria*. Glasgow, UK: Blackie Academic and Professional.

LACTOSE

J M Johnson and F D Conforti, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

Introduction

Lactose may be found as a naturally occurring compound of foods containing dairy products, or it may be isolated from such products and added for specific functional properties. This article discusses the recovery of lactose and the importance of lactose as a food ingredient. Many of the properties of lactose make it a useful food ingredient, but some may be at a disadvantage and require special control. The sensory and dietary considerations of lactose are also discussed.

Sources

Lactose is one of the main constituents of human and animal milk (**Table 1**). As a disaccharide consisting of glucose and galactose, lactose acts as an energy-carrier in milk. Due to its physiological and functional characteristics, industrially manufactured lactose is used today in a large number of foodstuffs as well as in the pharmaceutical industry. Lactose is produced from whey, a byproduct of cheesemaking and casein production, by crystallizing an oversaturated solution of whey concentrate. Global demand for lactose has grown appreciably over the last 10 years, the lactose industry having adapted accordingly, especially in the USA and Europe. The world production amounts to approximately 500 000 tons, the tendency still being upwards.

Physical and Chemical Properties

Lactose is a disaccharide practically unique to milk. It consists of a D-glucose and a D-galactose molecule joined in a β-1,4-glycoside linkage. Its systemic name is β-*O*-D-galactopyranosyl-(1-4)-α-D-glucopyranose (α-lactose) or β-*O*-D-galactopyranosyl-(1-4)-β-D-glucopyranose (β-lactose). Lactose has two isomeric forms, α- and β-lactose, which differ in the steric configuration of the hydroxyl group of C-1 moiety of glucose.

Table 1 Lactose content of selected dairy foods

Food item	*Amount (g)*	*Lactose (g)*
Milk (skimmed, low-fat, whole)	245	8.5–12.8
Lactose reduced milk	244	2.5–3.5
Nonfat dry milk	30	11.6
Cheese (cured, natural)	28	0.4–0.8
Processed cheese	28	0.4–4.1
Yogurt	227	4.3–17.6
Butter	15	0.1
Icecream/icemilk	99	5.5–14.4

Lactose may appear in different crystal polymorphs, depending on the crystallization conditions. Each polymorph has its specific properties. As an example, α-lactose (which crystallizes as a hydrate containing equimolar amounts of lactose and water) has very hard crystals and is not hygroscopic, as with β-lactose. However, a hydrous β-lactose (which crystallizes at temperatures above 93.5 °C) forms small crystals with a solubility up to 10 times higher than that of α-lactose. Because of this higher solubility, β-lactose is experienced sweeter than α-lactose.

The solubility of α-lactose is more temperature-dependent than that of β-lactose, and the solubility curves intersect at 93.5 °C. A solution at 60 °C contains approximately 59 g of lactose per 100 g of water. Therefore, if a 50% solution of lactose (approx. 30 g of β- and 20 g of α-) at 60 °C is cooled to 15 °C, then at this temperature, the solution can contain only 7 g of α-lactose or a total of 18.2 g per 100 g of water at equilibrium. Therefore, α-lactose will crystallize very slowly out of solution as irregularly sized crystals, which give rise to a sandy, gritty texture.

α- and β-lactose differ considerably in solubility and in the temperature dependence of solubility. If α-lactose is brought in water, much less dissolves at the outset than later. This is because of mutarotation: α-lactose is converted to β, hence the α-concentration diminishes, and more α can dissolve. If β-lactose is brought in water, more dissolves at the outset than later (at least below 70 °C): on mutarotation, more α-lactose forms than can stay dissolved, and α-lactose starts to crystallize.

The solution of lactose is temperature-dependent, and solutions are capable of being highly supersaturated before spontaneous crystallization occurs, and even then, crystallization may be slow. In general, supersolubility at any temperature equals the saturation (solubility) value at a temperature of 30 °C higher. The low solubility of lactose, coupled with its capacity to form supersaturated solutions, is of considerable practical importance in the manufacture of concentrated milk products. This is of prime importance in sweetened condensed milk, and crystal size must be controlled if a product with a desirable texture is to be produced. (*See* **Carbohydrates**: Classification and Properties.)

Usually, α-lactose crystallizes as a hydrate containing one molecule water of crystallization. The crystals are very hard, slightly hygroscopic, often fairly large, and dissolve slowly. The water of crystallization is very strongly bound. Above 93.5 °C, anhydrous β-lactose crystallizes from an aqueous solution. β-lactose is not very hydroscopic, and it dissolves quickly; its solubility is good. Therefore, dehydrating α-hydrate is difficult. It may cause problems when determining the dry matter of milk and milk products; this determination implies evaporation of water at elevated temperatures. Maintaining the temperature > 93.5 °C during the assay is paramount to prevent the formation of α-lactose hydrate crystals.

Amorphous lactose is formed during rapid drying, as in a spray drier. It is present in the glassy state, which means that many properties, including hardness, density, and specific heat, are similar to those of crystalline sugar but that the packing of the molecules does not show perfect order. Amorphous lactose contains at least a few percent of water and can quickly dissolve on addition of water. But then, α-lactose hydrate may start to crystallize. If the water content of the amorphous lactose is low, say 5%, crystallization is postponed. However, the product attracts water from moist air, and when moisture content rises about 8%, α-lactose hydrate starts to crystallize (at room temperature). The postponed crystallization is an important factor in relation to spray-dried powders made from skim milk or whey because it leads to hard lumps in the powder, and eventually, the whole mass of powder turns into one solid cake.

This caking is caused by the crystallization of lactose, as it causes the powder particles, largely consisting of lactose, to grow together (to sinter). Since water is needed for crystallization of α-lactose, caking does not occur at a low a_w, below 0.4. At a higher temperature, crystallization can occur far more readily, a_w being higher; moreover, the viscosity of the highly concentrated lactose solution (essentially the continuous phase of the powder particles) is lower, leading to a more rapid nucleation, hence crystallization. The susceptibility of caking, especially higher in whey powder, is considerably reduced if most of the lactose is crystallized before the drying (in the concentrate). Such precrystallized powder is usually called 'nonhygroscopic,' which may be a misnomer because the powder concerned does not attract less water (this is determined by its a_w in relation to that of the air), but the effects differ.

Sensory Properties and Digestibility

Lactose is a disaccharide that yields D-glucose and D-galactose on hydrolysis. Lactose occurs in both α- and β-forms, which differ in the steric configuration on the H and OH around the C-1 glucose. Lactose is not as sweet as sucrose or fructose. For example, 1.08, 2.2, and 4.91% sucrose are of equivalent in sweetness to 3.75, 7.5, and 15% equilibrium lactose solutions. β-lactose is 1.05–1.22 times as sweet as α-lactose, but this small difference in sweetness is of no practical value because equilibration of the anomers occurs rapidly in solution.

The low sweetness of lactose compared with other saccharides may be due to a weaker binding of the molecule to the tongue papillae when compared with fructose, sucrose, or glucose. To take advantage of the fact that the β-form is sweeter than the α-form, methods have been developed to convert the monohydrate α-form to the β-form by treatment with methanol and sodium hydroxide. However, when the β-form is in solution, mutarotation occurs quickly, and the solution becomes a mixture of α- and β-anomers and must be utilized in quick dissolving applications if only the β form is specified.

The sweetness of lactose syrups is increased with partial hydrolysis to form a solution of galactose, glucose, and lactose. The three sugars have a synergistic effect on sweetness when mixed in solution.

Hydrolysis of lactose to glucose/galactose markedly changes the two properties of greatest commercial importance in sugar, i.e., sweetness and solubility. The hydrolyzed lactose syrup is expected to find applications in nonalcoholic beverages, ice cream, yogurt, salad dressings, jams, pectin jellies, toffees, fudge, and boiled sweets. The technology exists for the commercial conversion of glucose (from starch hydrolysis) into fructose *via* immobilized glucose isomerase. Thus, using two linked immobilized enzyme reactors, it is possible to produce sweet glucose–fructose–galactose syrup from lactose:

$$\text{Lactose} \xrightarrow{\beta\text{-galactosidase}} \text{glucose} + \text{galactose} \xrightarrow{\text{glucose isomerase}} \text{glucose} + \text{fructose} + \text{galactose}$$

Low levels of intestinal lactase activity occur in many otherwise healthy adults and children in various population groups. Current evidence indicates that these low levels are normal for most populations of the world, notable exceptions being peoples of northern European extraction: approximately two-thirds of the world's adult population are believed to be lactose malabsorbers. Individuals who have genetically acquired low lactase levels as adults were able to drink milk without any ill effects as infants.

Clinically identifiable lactose malabsorption evolves over time and may be classified into three distinct periods: the initial period is characterized by a decline in lactase activity through childhood; the second phase in this continuum is noted when the decline in lactase activity is sufficient to result in a level of undigested lactose quantitatively high enough to cause symptoms associated with the ingestion of a lactose load; and the third phase occurs when the level of lactase declines to the point that it is no longer sufficient to digest even the modest levels of lactose usually consumed.

Diarrhea, intestinal cramps, and abdominal swelling are the symptoms experienced by individuals with lactose malabsorption when foods containing milk products are consumed. The unhydrolyzed lactose passes to the large intestine and is fermented to lactic acid by microflora, contributing to diarrhea. People experiencing lactose intolerance should limit their consumption of milk and milk products or select lactose-modified milk. (*See* **Food Intolerance**: Lactose Intolerance.)

Lactose-hydrolyzed milk is a viable product of enzyme technology, but its manufacture must be conducted optimally in terms of adaption to milk processing and lactase efficiency. Possibilities for the widespread production of lactose-modified dairy products have been opened up by the development of commercial sources of the enzyme lactase. Lactose-modified nonfat dry milk may be manufactured for the use as a food ingredient, and cultured products such as yogurt may be prepared successfully from lactase treated milk. Lactose-modified whey may be condensed to high-solid noncrystallizing syrups for use in soft drinks, confections, and ice cream.

A rare, and possibly fatal, condition of lactose intolerance is galactosemia, in which an infant is found to have a deficiency of galactose-1-phosphate uridyltransferase, which converts galactose to glucose for absorption. Infants with this condition must be diagnosed early and given substitute milk products that have fructose or sucrose added for energy and sweetening properties.

Isolation and Purification

The principal raw material for lactose production is cheese whey, although casein may be used. With the application of new membrane processing techniques in dairy technology, ultrafiltration permeate from the production of cheese and whey protein is now becoming a major dairy waste in the USA, Canada, Europe, and New Zealand. In membrane processing operations, the problem of waste is reduced almost entirely to milk salts and lactose. Expansion in the use of ultrafiltration will depend largely on the profitable use of permeate. A general solution to the problem of utilizing permeates rests with the human food industry.

The whey protein byproduct of lactose serves as a functional ingredient for other foods. The recovery of lactose and protein from whey resolves two problems: (1) an economical source of a functional ingredient and (2) the utilization of a product (whey) in abundance from cheese production that otherwise is a disposal problem. (*See* **Whey and Whey Powders**: Protein Concentrates and Fractions.)

The manufacture of lactose is similar to that of sucrose. It is basically a five-step process:

concentration, coagulation and removal of whey, further concentration, crystallization, and recovery of crystals (usually by centrifugation). The crystals are usually washed and may be dissolved and recrystallized to yield a product of high purity. The normal form of commercial lactose is α-monohydrate, but there is a limited market for β-lactose that is produced by crystallizing above 93.5 °C. Commercial lactose is available in a number of grades.

The addition of β-galactosidase to milk for cheese manufacture shortens the time needed to produce cheese. The hydrolysis of lactose by the enzyme stimulates the action of the microorganisms by converting the monosaccharide to glucose and galactose, which are more readily available for fermentation. However, this limits the suitability of whey for lactose recovery and poses a problem of disposal of whey that cannot be used for lactose recovery. (*See* **Cheeses**: Chemistry and Microbiology of Maturation.)

Food Applications

One of the oldest uses of milk sugar is as a constituent of baby food to adapt cow's milk to the composition of human milk. Due to its flavor-enhancing properties and low sweetening powers, lactose is primarily used now in numerous foodstuffs. Particularly worth mentioning here are: baby food, cakes and biscuits, chocolate and chocolate products, sugar confectionery, soups, and sauces.

The presence of lactose can affect a number of properties of confectionery systems that can have implications for the processibility of the system and for the overall product characteristics, such as flavor, mouth feel, texture, color, and stability. The changes observed when lactose is used in confectionery formulations would be expected to be a function of (1) the concentration, type and physical characteristics of the lactose, (2) the function and physical characteristics of the ingredient that is substituted by lactose, and (3) the processing parameters used in the products' manufacture.

Typically, at least 10% lactose can be used in most systems as a part-substitute for sucrose or skimmed milk powder without altering the textural acceptability of a product to an unsatisfactory degree, whereas up to 20% substitution is possible in certain cases. The tendency for lactose to crystallize and grain the confection appears to be a factor limiting the use of higher levels in many cases, whereas the difference in viscosity behavior compared with sucrose also inhibits the preparation of textural attributes similar to products containing sucrose. These observations, however, suggest that modifications to the manufacturing procedure to account for these properties could enable even higher levels to be incorporated successfully.

Human milk contains ~7.0% of lactose (bovine *sim*5.0%) and 1.0% of protein (bovine *sim*3.5%). Most baby formulas are essentially 'humanized' cow's milk, i.e., the lactose content is increased and the protein content reduced. Most of the lactose used in baby food preparations is added in the form of demineralized whey powder. Lactose may be used as an additive in powdered foods as a free-flowing agent. The hygroscopicity of lactose glass may be exploited to advantage as a means of absorbing free moisture from low moisture foods and fixing it as water of crystallization of lactose. For example, food particles may be coated in a solution of flavoring and/or coloring materials and then the coated particles tumbled in a powdered lactose glass that absorbs surface moisture and forms a capsule around the food particles. Lactose glass is used in 'instantizing' or increasing the dispersibility of certain foods. Products are prepared containing 5–50% lactose, spray-dried, and then 'instantized' by moistening and redrying. Alternatively, powdered lactose glass is added to the powdered food, which is then wetted slightly, causing the lactose to crystallize and form agglomerates that entrap the other food components. Such products are free-flowing and capable of dispersing rapidly, similar to instant milk powders.

Lactose is used extensively in the formulation of various types of candy, fudges, and caramels. It has certain advantages in these products, e.g., less sweetness, better color binding and better mouth-feel particles, but it also has the disadvantage of tending to crystallize and causing grittiness. Lactose, along with sucrose, may be used to sugar-coat chocolate buttons and hazelnuts: the lactose suppresses sucrose crystallization, thereby allowing the coating to be effected at relatively low temperatures, and reduces the sweetness of the coating.

Lactose is difficult to crystallize and can be a disadvantage in food applications. α-Lactose has a very low solubility and cannot be used in partially prepared food products that are rehydrated without heating such as instant puddings. Large crystals may form in the manufacture of ice cream when the mix is cooled to a low temperature. The result is a sandy texture or crystals large enough to be felt by the tongue and palate. The problem can be overcome by inhibiting crystal formation by the addition of gums or by seeding with additional crystal nuclei to keep the crystal size small. (*See* **Ice cream**: Methods of Manufacture.)

The concentration of lactose in sweetened condensed milk is critical in protecting the texture of the milk product. Condensed milk has 60% of the water removed and sucrose added to produce a product with a total carbohydrate concentration of 56%.

Too much lactose results in a product that is grainy, and too little lactose produces a product that is slimy in texture. (*See* **Condensed Milk**.)

Lactose is used in the bakery and confectionary industries, where it reacts with protein amino groups to give Maillard browning, for color and flavor. Lactose enhances the emulsifying and creaming properties of shortenings, which improves product quality, facilitates baking operations, gives increased loaf volume and external appearance score, and extends the shelf-life freshness by 50–100% compared with the shelf-life of standard formulations. Lactose may also be used in sugar coating and as a humectant in bakery products. If proper precautions are taken, lactose is a very suitable icing sugar for many purposes: it gives a reduced sweetness and better body with less chipping and cracking.

The vulnerability of lactose in the Maillard browning reaction is a problem in the manufacture and storage of dried milk and dried whey powder. Both products brown when stored at ambient temperature, making them undesirable in terms of solubility for rehydration and nutritional value. Dried milk that has browned is difficult to reconstitute, and browned milk is deficient in available lysine, the ε-amino group of which takes part in the Maillard reaction. Dried milk and whey solids should be stored under dry, cool conditions and rotated when in stock to avoid browning during storage. (*See* **Browning**: Nonenzymatic.)

See also: **Browning**: Nonenzymatic; **Carbohydrates**: Classification and Properties; **Cheeses**: Chemistry and Microbiology of Maturation; **Condensed Milk**; **Ice Cream**: Methods of Manufacture; **Whey and Whey Powders**: Protein Concentrates and Fractions

Further Reading

Aquilar CA, Hollender R and Ziegler GR (1994) Sensory characteristics of milk chocolate with lactose from spray-dried milk powder. *Journal of Food Science* 59: 1239–1243.

Aquilar CA and Ziegler GR (1995) Viscosity of malted milk chocolate with lactose from spray-dried whole milk powders. *Journal of Food Science* 60: 120–124.

Arvanitoyannis I and Blanshard JMV (1994) Rates of crystallization of dried lactose–sucrose mixtures. *Journal of Food Science* 59: 197–205.

Delaney RAM, Donnelly JK and O'Sullivan AC (1973) Lactose and reverse osmosis. In: Birch GG and Green LF (eds) *Molecular Structure and Function of Food Carbohydrate*, pp. 155–179. New York: Wiley.

Fox PF and McSweeney PLH (1998) *Dairy Chemistry and Biochemistry*, pp. 21–66. London: Blackie Academic and Professional.

Hartel RW and Shastry AV (1991) Sugar crystallization in food products. *Critical Reviews in Food Science and Nutrition* 1: 49–112.

Kedword CJ, MacNaughton W, Blanshard JMV and Mitchell JR (1998) Crystallization kinetics of lactose and sucrose based on isothermal differential scanning calorimeter. *Journal of Food Science* 63: 192–197.

Morrissey PA (1985) Lactose: chemical and physicochemical properties. In: Fox PF (ed.) *Developments in Dairy Chemistry – 3*, pp. 1–35. London: Elsevier Applied Science.

Nickerson TA (1977) Lactose sources and recovery. In: Birch GG and Shallenberger RS (eds) *Developments in Food Carbohydrate*, 1st edn, pp. 77–90. London: Applied Science.

Thelwall LAW (1980) Lactose. In: Lee CL (ed.) *Developments in Food Carbohydrate*, 2nd edn, pp. 275–326. London: Applied Science.

Walstra P, Guerts TJ, Noomen A, Jellema A and van Boekel MAJS (1999) *Dairy Technology: Principles of Milk Properties and Processes*, pp. 33–36. New York: Marcel Dekker.

Whitaker R (1993) Some factors influencing the crystallization of lactose in ice cream. *Journal of Dairy Science* 16: 177.

Yang ST and Silva EM (1995) Novel products and new technologies for use of a familiar carbohydrate, milk lactose. *Journal of Dairy Science* 78: 2541–2562.

Lactose Intolerance *See* **Food Intolerance**: Types; Food Allergies; Milk Allergy; Lactose Intolerance; Elimination Diets

Lager *See* **Beers**: History and Types; Raw Materials; Wort Production; Biochemistry of Fermentation; Chemistry of Brewing; Microbreweries

Lamb *See* **Sheep**: Meat; Milk

LEAD

Contents
Properties and Determination
Toxicology

Properties and Determination

A R Flegal and G M Scelfo, University of California at Santa Cruz, Santa Cruz, CA, USA

Background

Lead (Pb) is a highly toxic trace element with no recognized biological requirement in organisms and no established threshold concentration for sublethal toxicity. Moreover, lead concentrations in the biosphere have been substantially elevated by industrial releases of lead over the past five millennia. These contaminant lead emissions have been difficult to quantify accurately, and many measurements of lead in the biosphere have been biased by undetected contamination during sampling, storage, and analysis.

Occurrence in Foods

The relative magnitude of environmental lead contamination is unparalleled among the elements. Approximately 350 million tonnes of lead have been produced since the advent of cupellation (a smelting process that separates precious metals from lead ores) five millennia ago (Figure 1), and most (> 50%) of it has been dispersed as an environmental contaminant on a global scale. Production rates of ~3.4 million tonnes of lead per year were estimated to release ~1.6 million tonnes of contaminant lead into the environment annually during the preceding decade. These inputs included atmospheric emissions of industrial lead aerosols (3×10^8 kg year^{-1}), which were nearly two orders of magnitude (10^2) greater than the sum of all natural atmospheric emissions of lead (4.5×10^5 kg year^{-1}). These industrial emissions increased average lead concentrations in the biosphere by approximately

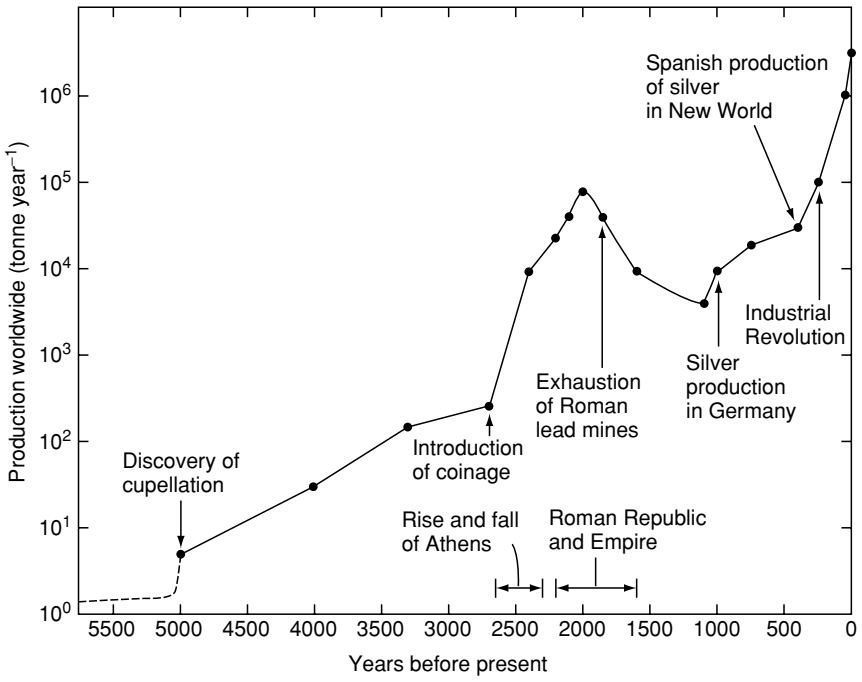

Figure 1 Industrial production of lead over the past five millennia. From Settle DM and Patterson CC (1980) Lead in albacore: Guide to lead pollution in Americans. *Science* 207: 1167–1176, with permission.

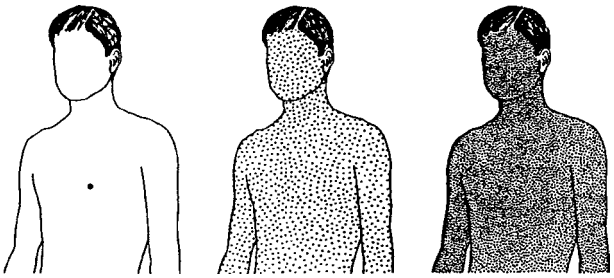

Figure 2 Relative amounts of lead contamination in adult humans. Each dot represents 40 μg Pb per 70-kg person. The figure on the left with one dot represents the natural concentration of lead in humans (40 μg per 70 kg). The figure in the center represents the current average lead concentration of adults in urban environments (40 000 μg per 70 kg) in previous decades; recent studies of ambient blood lead concentrations indicate that average lead concentrations in children and young adults are now substantially lower (Pirkle *et al.*, 1998). The figure on the right represents the minimum concentration that will actuate classical lead poisoning in a significant fraction of the population (160 000 μg per 70 kg). From Patterson CC (1980) An alternative perspective – lead pollution in the human environment: Origin, extent and significance. In: National Research Council. *Lead in the Human Environment*. Washington, DC: National Academy of Sciences.

one order of magnitude above natural levels. However, there have been measurable reductions in environmental lead concentrations that closely correlate with reductions in industrial lead emissions over the past three decades.

One consequence of industrial lead emissions into the biosphere has been a 500–1000-fold increase in the lead body burdens of older (>40 years) adults of industrialized countries, since most (>90%) of that lead is in bone, where its biological residence time is relatively long (Figure 2). Contemporary baseline lead concentrations of those individuals are estimated to be still approximately 40 000 μg of Pb per 70-kg person, which is ~1000-fold (10^3) higher than natural levels in prehistoric individuals (40 μg of Pb per 70-kg person), and only fourfold lower than levels in individuals with overt, clinical lead poisoning (160 000 μg Pb per 70-kg person). The body burdens of younger adults and children are now estimated to be much lower (~10 000 μg of Pb per 70-kg person), since they have had less exposure to industrial lead.

Presumably, there have been comparable decreases in the lead concentrations of other components of the biosphere, including foods in human diets. Yet, contaminated foods are now a principal source of elevated lead concentrations in humans living in regions where the use of leaded gasolines has sharply declined (Table 1). The dietary intake of lead by adults in those areas is estimated to account for approximately two-thirds of their baseline daily consumption of lead (~50 μg day^{-1}).

Table 1 Recent trends in total baseline human exposure to lead (units in $\mu g\ day^{-1}$)

Age/sex category	*1982*	*1983*	*1984*	*1985*
2-year-old children				
Food and beverage Pb	28.4	21.6	21.7	13.1
Air Pb	2.4	1.8	1.7	1.1
Dust Pb	20.1	20.1	20.1	20.1
Total food, air, and dust Pb	50.9	43.5	43.5	34.3
Teenage females (14–16 years)				
Food and beverage Pb	37.6	25.7	27.6	14.7
Air Pb	4.8	3.5	3.4	2.2
Dust Pb	4.5	4.5	4.5	4.5
Total food, air, and dust Pb	46.9	33.7	35.5	21.4
Teenage males (14–16 years)				
Food and beverage Pb	53.8	36.1	41.1	20.0
Air Pb	4.8	3.5	3.4	2.2
Dust Pb	4.5	4.5	4.5	4.5
Total food, air, and dust Pb	63.1	44.1	49.0	26.7
Adult females (25–30 years)				
Food and beverage Pb	37.9	27.2	28.0	14.5
Air Pb	4.8	3.5	3.4	2.2
Dust Pb	4.5	4.5	4.5	4.5
Total food, air, and dust Pb	47.2	35.2	35.9	21.2
Adult males (25–30 years)				
Food and beverage Pb	55.2	36.3	42.5	19.8
Air Pb	4.8	3.5	3.4	2.2
Dust Pb	4.5	4.5	4.5	4.5
Total food, air, and dust Pb	64.5	44.3	50.4	26.5
Senior adult females (60–65 years)				
Food and beverage Pb	36.7	28.6	28.3	15.5
Air Pb	4.8	3.5	3.4	2.2
Dust Pb	4.5	4.5	4.5	4.5
Total food, air, and dust Pb	46.0	36.6	36.2	22.2
Senior adult males (60–65 years)				
Food and beverage Pb	45.8	34.4	36.1	18.8
Air Pb	4.8	3.5	3.4	2.2
Dust Pb	4.5	4.5	4.5	4.5
Total food, air, and dust Pb	55.1	42.4	44.0	25.5

From Flegal AR *et al*. (1990) Lead contamination in food. In: Nriagu JO and Simmons MS (eds) *Environmental Food Contamination, Advances in Environmental Science and Technology*. New York: John Wiley.

Background lead concentrations in basic food crops and meats range from 2 to 45 p.p.b. ($ng\ g^{-1}$) fresh weight (FW) (Table 2). Most (50 to >99%) of that lead is derived from direct and indirect atmospheric inputs of industrial lead aerosols. The rest (<1 to 50%) is derived from soils (primary minerals, humic substances, and soil moisture), which have natural lead concentrations of <8 to 25 p.p.m. ($pg\ g^{-1}$). Consequently, contemporary background lead concentrations in foods are two- to 100-fold above natural concentrations, which are estimated to range from 0.1 to 20 $ng\ g^{-1}$ (FW).

Leafy crops are most susceptible to contamination from atmospheric deposition of industrial lead (Table 2). These crops include wheat, rice, spinach,

Table 2 Background lead concentrations in basic food crops and meat (units in $\mu g\ g^{-1}$ fw)

Crop	Natural	Indirect atmospheric	Direct atmospheric	Total
Wheat	0.0015	0.0015	0.034	0.037
Potatoes	0.0045	0.0045		0.009
Field corn	0.0015	0.0015	0.019	0.022
Sweet corn	0.0015	0.0015		0.003
Soyabeans	0.021	0.021		0.042
Peanuts	0.005	0.005		0.010
Onions	0.0023	0.0023		0.0046
Rice	0.0015	0.0015	0.004	0.007
Carrots	0.0045	0.0045		0.009
Tomatoes	0.001	0.001		0.002
Spinach	0.0015	0.0015	0.042	0.045
Lettuce	0.0015	0.0015	0.010	0.013
Beef (muscle)	0.0002	0.002	0.02	0.02
Pork (muscle)	0.0002	0.002	0.06	0.06

From US Environmental Protection Agency (1986) Air quality criteria for lead, vol. II. US EPA Environmental Criteria and Assessment Office, Research Triangle Park, NC. EPA/600/8-83/0286F.

Table 3 Environmental lead concentrations in remote/rural and urban areas (units in $\mu g\ g^{-1}$, except those for 'Air,' which are $\mu g\ m^{-3}$)

	Remote/rural	Urban[a]
Air	0.05	2.3
Fresh water	1.7×10^{-5}	0.005–0.030
Soil	10–30	150–300
Plants	0.18[b]	950[c]
Herbivores	2.0[c]	38[c]
Omnivores	1.3[c]	67[c]
Carnivores	1.4[c]	193[c]

[a]Values can be highly variable, depending on organism and habitat location.
[b]Fresh-weight basis.
[c]Dry-weight basis.
From Flegal AR *et al*. (1990) Lead contamination in food. In: Nriagu JO and Simmons MS (eds) *Environmental Food Contamination, Advances in Environmental Science and Technology*. New York: John Wiley.

and lettuce, whose contemporary baseline lead concentrations are five- to 100-fold above natural concentrations. Contaminant lead concentrations in foods not directly exposed to atmospheric depositions (e.g., root crops, nuts, beans, and some fruits) are approximately twofold lower. Atmospheric industrial leads are also a primary (>90%) source of lead contamination of some livestock and poultry. This occurs initially through the ingestion of contaminated forage, feed, and soils, and subsequently during the processing of meat products.

Current baseline lead concentrations of foods are several orders of magnitude lower than lead concentrations of foods in some industrial areas (Table 3). For example, lead levels in some plants are at sub-part-per-million concentrations in remote, rural environments ($0.2\ \mu g\ g^{-1}$) and part-per-thousand concentrations in some highly contaminated urban areas ($950\ \mu g\ g^{-1}$) as a consequence of local emissions of industrial lead aerosols (e.g., automobile exhausts of leaded fuels). Lead concentrations of consumer organisms (herbivores, omnivores, and carnivores) may also be increased by one to two orders of magnitude above baseline concentrations by the ingestion of contaminated foods and soils.

Numerous other sources of lead contribute to the lead ingested by humans. For example, lead solder may contribute approximately 40% of the dietary lead accumulated by adults through the contamination of both drinking water and foods. It is estimated that 42 million people in the USA are exposed to drinking water with elevated lead levels ($>20\ \mu g\ l^{-1}$) due to the lead solder in pipes. Lead solder also accounts for >99% of the lead in some foods packaged in lead-soldered cans. Lead-glazed ceramics may increase lead concentrations in foods to parts-per-million concentrations, and are a notable source of acute lead toxicity. Leaded crystal, pewter, potions, and sealants are other sources of lead contamination to foods. Finally, dirt may be ingested inadvertently as a food contaminant or directly through pica. A child that ingests as little as 1 g of soil and 200 mg of lead paint chips a day may be consuming 2600 µg of lead per day and absorbing 550 µg of lead per day.

Chemical Properties

Lead has the most stable divalent (+2) oxidation state of the group IVB elements, though it also occurs in a tetravalent (+4) oxidation state. Coordination numbers for its divalent compounds range from 2 to 7. Those for its tetravalent compounds range from 4 to 8. The stereochemistry of lead is usually octahedral or tetrahedral. The divalent oxidation state dominates the inorganic chemistry of lead, while its organic chemistry is dominated by the tetravalent (+4) oxidation state.

Divalent lead, Pb^{2+}, is a B-type or soft sphere metal cation. Lead's classification, which is based on the number of electrons in its outer shell, indicates that Pb^{2+} ions tend to form covalent bonds, in sharp contrast to the mainly ionic binding of hard sphere ions, such as Ca^{2+}. This is evidenced by the tendency of Pb^{2+} to form strong bonds with groups containing nitrogen and sulfur (e.g., CN, R-S, -SH, and imidazol) in biological systems. It is also apparent in mineral deposits, where galena (PbS) is the most abundant lead ore and where anglesite ($PbSO_4$) and cerrusite ($PbCO_3$) are less common minerals.

Despite differences in their chemistries, lead is introduced and cycled in biological systems as an analog of calcium. Most (70 to >90%) of the lead

in vertebrates is associated with calcareous tissues, where lead can substitute for calcium in apatite matrices. In addition, lead competes with Ca^{2+} for sites in cellular and subcellular systems. These include synaptosomes, mitochonrdria, membrane vesicles, protein kinase C, calmodulin and other Ca^{2+}-binding proteins. This ionic substitution is associated with some of the primary mechanisms of lead toxicity, which are attributed to the alteration of calcium-mediated cellular processes and the mimicry of Ca^{2+} binding to regulatory proteins.

Since lead is a biogeochemical analog of calcium, lead concentrations are normalized to calcium concentrations (atomic ratios) to quantify distributions and fluxes of lead within organisms and food chains. This normalization indicates a systematic decrease in the atomic ratio of lead to calcium at higher trophic levels, a process referred to as biopurification. The biopurification of calcium relative to lead is attributed to the selective transport of Ca^{2+} across cell membranes, the selective retention of Ca^{2+} in metabolic processes, and the selective removal of Pb^{2+} in detoxifying processes. These biopurification processes may be circumvented by the direct assimilation of contaminant lead by organisms, which occurs in addition to the ingestion of lead within food. As a result, atomic ratios of lead to calcium in humans may exceed the ratios in their diets.

Isotopic Composition

Lead (atomic number 82) has an atomic weight of 207.19, based on the relative abundance of its four stable isotopes: ^{204}Pb (1.5%), ^{206}Pb (23.6%), ^{207}Pb (22.6%), and ^{208}Pb (52.3%). The latter three isotopes (^{206}Pb, ^{207}Pb, and ^{208}Pb) are formed by the radioactive decay of the long-lived natural isotopes ^{238}U, ^{235}U, and ^{232}Th, respectively, while ^{204}Pb has no radioactive progenitor. Since each parent isotope has a different half-life, the relative natural abundance of stable lead isotopes in a geological deposit varies systematically with the age and evolution of the formation. Different geological formations may, thus, be distinguished by their characteristic (but not necessarily unique) stable lead isotopic ratios.

Ratios of stable lead isotopes may also be utilized to identify and trace different sources of lead in the environment. Natural differences in stable isotopic compositions of lead, which are characteristic of different geological formations, persist in the environment because there is no measurable physical, chemical, or biological fractionation of lead isotopes within the biosphere. This allows sources of industrial lead inputs into the atmosphere, water, soils, and organisms to be distinguished by their isotopic composition. Additionally, stable lead isotopic tracer techniques may be utilized to investigate lead exposure and uptake by organisms, as well as the cycling and therapeutic manipulation of accumulated lead.

However, most lead tracer studies, to date, have used man-made radiolead isotopes. One radioisotope of lead, ^{210}Pb (with a half-life ($t_{½}$) of 22 years), is commonly used to characterize and date relatively short (< 100 years) geological events. It is also employed as a radiotracer in biological systems, as are three other radioisotopes with shorter half-lives: ^{211}Pb ($t_{½}$ = 36.1 min), ^{212}Pb ($t_{½}$ = 10.64 h), and ^{213}Pb ($t_{½}$ = 26.8 min). The applicability of of those radioisotopes is constrained by their half-lives, which often are not suitable for analyses of many metabolic processes, the relative insensitivity of radiometric measurements, and increasing controls on the use of radioactivity in human health studies. Conversely, the applicability of stable lead isotopic tracers has been rapidly increasing with the development of high-resolution inductively coupled plasma mass spectrometers.

Analysis – Isolation and Concentration

Analyses of lead concentrations in environmental matrices have been plagued by problems of contamination during sampling, storage, and analysis. Unrecognized or poorly documented contamination have invalidated most environmental lead data reported prior to 1980 and indicated that many subsequent data are highly questionable. This has been demonstrated by numerous comparisons of data from traditional laboratories and those using ultraclean analytical techniques. Perhaps the most celebrated of those comparisons were analyses of lead in tuna, which demonstrated that the lead concentration of fresh tuna (0.3 $\mu g\,g^{-1}$ FW) was three orders of magnitude lower than the concentration (400 $\mu g\,g^{-1}$ FW) measured with previously accepted techniques.

Numerous procedures have been developed to isolate and concentrate lead for analysis. One of the most common procedures is liquid/liquid extraction, which typically involves a water-immiscible solvent, an aqueous solution, and a complexing agent that forms a nonionic or neutral chelate with lead. Another common procedure is an ion-exchange extraction, which sequesters lead ions on a chelating functional group immobilized on a solid substrate. A third is electrodeposition, which concentrates lead on to an electrode for direct electrochemical analyses or to preconcentrate it for other instrumental analyses.

Total lead concentration measurements usually require a rigorous acid digestion prior to these isolation and concentration procedures. They are required for

most of the following analytical methods, because they do not have the requisite selectivity, sensitivity, or freedom from matrix interferences to accurately and precisely measure trace concentrations of lead in complex matrices. Measurements of lead concentrations in different fractions or phases typically have the same constraints.

Colorimetric Methods

Historically, lead concentrations were measured by colorimetric methods, including spectrometry. The primary spectrometric method for measuring lead now is with the reagent dithizone (diphenylthiocarbazone). It is a relatively simple and inexpensive technique, with a relatively low (pg) level of sensitivity. The accuracy and precision of this method is highly dependent upon the skills of the analyst and the utilization of rigorous trace metal clean techniques. The same qualification applies to the following techniques, which involve much more elaborate instrumentation.

Spectrometric Methods

The most common method of measuring lead concentrations in a variety of matrices is atomic absorption spectrometry (AAS). This method measures lead concentrations by the absorbance of lead spectra emitted from a source lamp by vaporized lead atoms. While flame AAS techniques may still be appropriate for measurements of relatively high ($pg\,g^{-1}$) concentrations of lead, most AAS analyses of lead are now by graphite furnace AAS. Evolving AAS methodologies include the development of alternatives to graphite furnaces (e.g., heated quartz tubes) and couplings with chromatographic separations for species composition measurements. (*See* **Spectroscopy**: Atomic Emission and Absorption.)

Atomic emission spectrometry (AES) is similar to AAS, except that lead is measured by the light emitted by excited ions rather than by the light they absorb. The primary advantage of AES is that numerous elemental concentrations may be measured simultaneously in a relatively small sample. This methodology is being improved with the development of inductively coupled plasma (ICP) systems, which increase sensitivity and decrease analytical interferences. Current ICP–AES systems now provide the sensitivity required for analyses of lead in most biological matrices.

The most accurate method of measuring lead concentrations in biological matrices is with isotope dilution thermal ionization mass spectrometry (TIMS). This definitive method is yield-independent, extremely sensitive, and precise. It is, also, the most accurate and precise method for measuring stable lead isotopic compositions.

However, other rapidly evolving mass-spectrometric methods provide comparable or complementary measurements of lead concentrations and isotopic compositions in different matrices. All of these mass spectrometric methods involve the electromagnetic and physical separation of lead ions within an instrument flight tube. The most prominent of these is inductively coupled plasma mass spectrometry (ICP-MS), which includes both the quadrapole ICP-MS and the high-resolution magnetic sector ICP-MS. The latter's sensitivity, accuracy, and precision now rival those of TIMS. Additional mass-spectrometric methods which have evidenced potential for analyses of lead in biological matrices include secondary ion mass spectrometry, glow-discharge mass spectrometry, and laser microprobe mass analysis.

Nuclear magnetic resonance (NMR) techniques are being developed for diagnostic analyses of lead. These focus on the resonance frequencies of ^{207}Pb, which are influenced by the molecular environment around the isotope. NMR provides a measure of lead speciation in both environmental and clinical settings. For example, NMR has been used to simultaneously measure intracellular free calcium and lead concentrations in erythrocytes. (*See* **Spectroscopy**: Nuclear Magnetic Resonance.)

Electrochemistry

Electrochemical methods are used to measure both lead concentrations and its organic complexation in aqueous systems. These measurements are based on the current produced as lead is reduced or oxidized by varying the potential of a working electrode in an electrochemical cell containing the sample. The sensitivity of total lead concentration measurements, which are determined after digesting the sample, can be increased by a preconcentration step. Organic complexation is determined by titrating the sample with lead and measuring the current produced after each lead addition.

The two principal electrochemical methods for determining lead in environmental samples are differential pulse polarography and anodic stripping voltammetry. The first method measures the faradaic current produced by the reduction of Pb(II) to Pb(0), while the potential on the working electrode is scanned to negative potentials. In the second method, lead is preconcentrated on a mercury electrode by the reduction of Pb(II) to Pb(0) during a reducing step. The potential is then scanned positive, and the oxidation current is measured.

Chromatography

Chromatography separation techniques can detect a wide range of molecular species containing metals or metalloids. This method involves isolation of the species of interest from the sample in a sufficiently volatile, thermally stable, and physicochemically resistant form for separation by gas–liquid or high-performance liquid chromatography. These methods are coupled with elemental detectors, principally flameless atomic absorption spectrometers, which provide low (< 50 pg) detection limits. Other detectors include flame atomic absorption spectrometers, thermal conductivity detectors, flame ionization detectors, electron-capture detectors, thermionic specific detectors, and atomic emission spectrometers. (*See* **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography.)

There have been numerous applications of chromatography methodologies to investigate the speciation of lead in the environment, because of its widespread dispersion, diverse chemical forms, and toxicity. Gas chromatography has been used to detect organolead compounds, including trialkyllead chlorides, in the environment. Gas–liquid chromatography was initially used to measure tetraalkyl derivatives of lead in the environment, and it has since been used to investigate the biomethylation of inorganic and ionic alkyl lead by microorganisms. Additional speciation measurements may be obtained with high-performance liquid chromatography coupled with GFAAS, mass spectrometry, and NMR.

X-ray Fluorescence

A rapidly evolving methodology for measuring lead concentrations *in situ* is with X-ray fluorescence (XRF). Sufficient incident radiation from an X-ray generator or a radioactive source is provided to excite an inner shell electron, with the resultant emission of a fluorescent X-ray. The emitted energy is characteristic of the element that absorbed the original X-ray, and the amount of energy released is proportional to the mass of element present.

Two types of XRF are applicable for noninvasive analyses of lead concentrations. L-line techniques require 10.5 keV to remove an L-shell electron, and K-line techniques require 88 keV to remove a K-shell electron. These XRF techniques are most appropriate for analyses of bone lead concentrations of individuals. This is significant because bone is the major reservoir of lead in humans and accounts for > 90% of their total lead content.

See also: **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **Mass Spectrometry**: Principles and Instrumentation; **Spectroscopy**: Atomic Emission and Absorption; Visible Spectroscopy and Colorimetry

Further Reading

Flegal AR, Smith DR and Elias RW (1990) Lead contamination in food. In: Nriagu JO and Simmons MS (eds) *Environmental Food Contamination, Advances in Environmental Science and Technology*. New York: John Wiley.

Halliday AN, Lee DC, Christensen JN *et al*. (1998) Applications of multiple collector-ICPMS to cosmochemistry, geochemistry, and paleoceanography. *Geochimica et Cosmochimica Acta* 62: 919–940.

Hamester M, Wiederin D, Willis J, Kerl W and Douthitt CB (1999) Strategies for isotope ratio measurements with a double focusing sector field ICP-MS. *Fresnius Journal of Analytical Chemistry* 354: 495–497.

National Research Council (1993) *Measuring Lead Exposure in Infants, Children and Other Sensitive Populations*. Washington, DC: National Academy of Sciences.

Patterson CC (1980) An alternative perspective – lead pollution in the human environment: Origin, extent and significance. In: National Research Council. *Lead in the Human Environment*. Washington, DC: National Academy of Sciences.

Pirkle JL, Kaufman RB, Brody DJ *et al*. (1998) Exposure of the U.S. population to lead, 1991–1994. *Environmental Health Perspectives* 106: 745–750.

Scelfo GM and Flegal AR (2000) Lead in calcium supplements. *Environmental Health Perspectives* 108: 309–313.

Settle DM and Patterson CC (1980) Lead in albacore: Guide to lead pollution in Americans. *Science* 207: 1167–1176.

Smith DR, Calacsan C, Woolard D *et al*. (1999) Succimer and the urinary excretion of essential elements in a primate model of childhood lead exposure. *Toxicology Science* 54: 473–480.

Steding DJ, Dunlap C and Flegal AR (2000) New isotopic evidence for chronic lead contamination in San Francisco Bay: Implications for the persistence of past industrial lead emissions in the biosphere. *Proceedings of the National Academy of Sciences* 97: 11181–11186.

Toxicology

A L Ehle, University of Chicago, Chicago, IL, USA

Introduction

Lead occurs widely throughout the environment. Naturally occurring forms and lead contributed from human activities constitute the potential source for human exposure. Lead was one of the first metals to be used by humans. Almost half of the world's production of lead is consumed by the storage battery industry. Lead pigments, chemicals, and metals

account for the remainder. Of importance to the food industry has been the past use of lead for water pipes, as a lining for containers, and in solder used in the manufacture and sealing of steel cans. Most inorganic lead compounds, with the exception of lead acetate and lead chloride, are not readily soluble in water. Solubility of lead compounds is enhanced in acidic solutions.

Absorption and Distribution

Human studies indicate that, in adults, approximately 10% of ingested lead in the diet is absorbed from the gastrointestinal tract. The fraction absorbed is greater if a solution of lead salts is ingested in a fasting state, compared to the same amount of lead ingested with food. In children, evidence suggests that the fraction absorbed may be as high as 40–50%. Absorption appears to be increased if the diet is deficient in protein, calcium, or iron. In addition to gastrointestinal absorption, significant lead absorption can occur through the respiratory system from inhalation of lead particles.

Absorbed lead is transported by blood, predominantly in red blood cells. It is initially rapidly distributed throughout the various soft tissues and organs. It is then slowly redistributed into two pools, a rapidly exchangeable pool in blood and soft tissue, and a less mobile pool in bone. In humans, about 90% of the total body burden is in bone. Placental transfer of lead occurs and there is a relatively good correlation between blood lead concentrations of mothers and newborn infants; the levels in infants are 5–10% lower than those in the mothers.

Excretion of absorbed lead is primarily through the kidneys; about 75% of the loss is through this route. The mechanism of urinary excretion appears to be essentially a filtration process. Gastrointestinal losses account for about 16% of the loss. The remainder of the loss is through hair, skin, and sweat.

The body half-life of lead is approximately 10–30 years. The half-life in the blood and soft-tissue compartment is about 20 days. The bone pool has a much longer half-life, 10–30 years, and accounts for the long total body half-life.

Lead in blood is generally accepted as the best indicator of exposure and body burden for soft tissues and organ systems. Average blood lead levels in the general population of industrial countries are typically found to be between 5 and 15 $\mu g\,dl^{-1}$.

Toxic Effects

Lead has toxic effects on several organ systems. The blood cell-forming system is the first to be affected by lead exposure. Frank anemia begins to occur in adults with blood lead above 60 $\mu g\,dl^{-1}$. Children are more sensitive to lead effects and begin to show anemia with blood levels above 30 $\mu g\,dl^{-1}$. At levels as low as 10 $\mu g\,dl^{-1}$, reversible biochemical changes in the hemoglobin synthetic pathway occur that are without clear adverse effects. This early change consists of progressive inhibition of δ-aminolevulinic acid dehydrase activity. Abdominal pain or 'lead colic' occurs at somewhat higher levels of 80–100 $\mu g\,dl^{-1}$ in adults.

Acute central nervous system toxicity can occur when blood levels are above 100 $\mu g\,dl^{-1}$ in adults, and are above 80–90 $\mu g\,dl^{-1}$ in children. This syndrome is an acute encephalopathy, characterized by coma, convulsions, and death or irreversible loss of function. Lead encephalopathy is much more common in children than adults. There is evidence that a chronic intellectual impairment may occur at lower blood levels, with children again being significantly more sensitive to these effects. The extent of intellectual impairments and the blood lead levels at which they occur are still a matter of controversy. In the USA blood lead levels above 10 $\mu g\,dl^{-1}$ in children are felt to be potentially harmful by public health authorities. Concern over this potential effect has been the prime motivator behind environmental regulatory efforts to reduce lead exposure for the general population.

Lead can also produce a peripheral neuropathy, resulting in significant weakness in some individuals. This effect is usually seen at relatively high blood levels. Electrophysiological changes in peripheral nerve function in the absence of any clinical symptoms have been reported at blood lead levels below 70 $\mu g\,dl^{-1}$ in some studies of lead-exposed workers. The significance of these findings is controversial in terms of indicating a clear toxic effect. Lead encephalopathy and neuropathy may be reversible if adequate treatment is instituted early in the evolution of these symptoms.

A prolonged high-level exposure is associated with kidney disease in some individuals. This takes the form of an interstitial fibrosis and a loss of renal function, which appears to be irreversible. Reversible change in renal function is more common and is characterized by mild elevation of blood urea nitrogen or creatinine, blood tests that indicate altered kidney function. There have been reports that serum concentration of 1,25-dihydroxy vitamin D, the active form of vitamin D, is depressed in children with moderate blood lead levels. This metabolite is formed in the kidneys. Recent studies suggest that the effect may only be seen against a background of nutritional deficiency, or low vitamin D intake.

In a given individual the manifestations of these toxic syndromes are variable at any given level of lead in the body. The factors responsible for the variability are poorly understood. Children, in general, are more sensitive to the toxic effects of lead.

Other possible toxic effects of lead in humans are less well defined and remain unsubstantiated. Although long-term high-level lead exposure in rodents can produce kidney tumors, there is no conclusive evidence that lead is carcinogenic in humans. In some studies an association between increasing blood lead levels and small increases in blood pressure were noted. Other investigators believe that confounding factors may account for the blood pressure changes. Although there are experiments indicating that high doses of lead are toxic to the reproductive functions of both male and female laboratory animals, data in humans are limited and inconsistent. There is no clear association of lead with the frequency of congenital anomalies in humans. There has been concern that *in utero* lead exposure may have harmful effects on brain development, in view of the toxic syndrome seen in children.

Lead toxicity is treated by identification and elimination of the source of exposure if symptoms are mild. In more severe toxicity, treatment with a chelating agent such as calcium ethylenediaminetetraacetic acid, D-penicillamine, or succimer is used to increase lead excretion.

Lead in the Food Chain

Lead is ubiquitous throughout the food chain. Table 1 is a compilation from a World Health Organization (WHO) document illustrating the typical lead content of various foods. Plants subject to the effects of air pollution potentially have the highest lead concentration, followed by root and tuberous plants. The lowest lead concentrations are found in plants in which the edible portion is above ground and shielded from airborne lead deposits. In animal food sources, muscle tissue has the lowest concentration of lead; the concentration in visceral organs, such as liver or kidney, is considerably higher. Concentrations of lead in all raw foodstuffs will vary, depending on the environment in which they were produced. Tetraethyl lead, where it is still in use as a gasoline additive, can be a significant source of airborne lead.

Food processing and preparation can contribute significantly to the lead content of foodstuffs. As noted in Table 1, lead concentrations in canned foods are noticeably higher in this study, primarily as a result of the lead contamination from solder used in steel cans. Alternative packaging can eliminate this. Significant reduction in dietary lead intake in the USA is attributed to increased consumption of fresh and frozen foods and elimination of lead-soldered cans for foods. Lead contamination of food may occur, beyond lead migration from soldered cans, from other sources during food processing. Water with a high lead content that is used in processing would be one such source of contamination. Contact with lead or lead-containing substances would be another potential source of contamination. The likelihood of this type of contamination is increased if the food in contact with the lead source is acidic. A common source for this form of lead contamination has been the use of pottery dishes or containers with improperly fired glazes that contain lead. Acidic food can leach large amounts of lead from such glazes, and

Table 1 Lead content in foods ($\mu g\, g^{-1}$ fresh weight)

	Uncanned	*Canned*
Dairy products and eggs		
Milk	0.02	0.10–0.13
Butter	0.07	
Icecream	0.01	
Cheese	0.05	
Eggs	0.17	
Meat and poultry		
Beef, pork, lamb, veal	0.06	0.24
Hamburger	0.25	
Beef liver	0.09	
Poultry	0.12	0.24
Fish and shellfish		
Salmon	0.39	0.72
Mackerel	0.40	0.99
Tuna		0.45
Cod	0.06	
Flounder	0.10	
Oysters	0.17	
Clams	0.21	
Cereal, nut, and sugar products		
Flour (white)	0.05	
Bread (white)	0.08	
Cereals (breakfast)	0.11	
Peanut butter	0.06	
Sugar (refined)	0.03	
Vegetables		
Potatoes	0.05	0.12
Cabbage	0.1–0.04	0.08
Lettuce	0.12–0.15	0.39
Beans	0.01–0.04	0.16–0.32
Peas	0.03	0.27
Carrots	0.14	0.13
Onions	0.18	0.32
Tomatoes	0.05–0.08	0.30–0.37
Cucumbers	0.02	
Fruit		
Citrus (oranges, lemons)	0.01	0.39
Apples	0.02	0.22
Cherries	0.02	0.39
Pears	0.02	0.18–0.19

Compiled in IMO/FAO/UNESCO/WHO/IAEA/UN/UNEP Joint Group of Experts on the Scientific Aspects of Marine Pollution Report (GESAMP) and Studies no. 22 (1985). Geneva: World Health Organization.

cases of clinical lead poisoning continue to be reported from this source of contamination. Despite awareness of this potential problem improperly fired pottery with lead-containing glazes continue to be produced, predominantly in developing countries.

Total food-related lead intake can be highly variable, depending on the dietary pattern. Estimates of average daily lead ingestion from food in western cultures where studies have been performed have typically ranged from 25 to 200 $\mu g\,day^{-1}$, although intakes as high as 400 $\mu g\,day^{-1}$ have been reported for some populations. The current US recommended daily intake from all sources is 75 $\mu g\,day^{-1}$ or less for adults and 6 $\mu g\,day^{-1}$ or less for children under 6 years of age. These amounts are one-tenth the daily intake estimate to produce blood lead levels of 30 $\mu g\,dl^{-1}$ in an adult and 10 $\mu g\,dl^{-1}$ in a child.

Dietary lead intake accounts for approximately 70% of the total daily lead absorption from all sources in most studies. More recent studies, from the US Food and Drug Administration, have found significant reductions from these figures that are attributed to almost complete elimination of lead-soldered food cans and reduced airborne lead contamination from leaded petrol. In these studies, dietary lead was found to be less than half of the daily lead intake.

Monitoring and Surveillance

Monitoring of the lead content of foods is not routinely performed. Avoidance of lead contamination in food preparation and food processing should be sought as a part of an integrated public health program to minimize lead exposure of the population. While lead exposure from the diet is normally far below that associated with toxic symptoms, it should be remembered that dietary lead is a major source of an individual's total lead intake. Although it is not clear at what level harmful effects first occur in humans, it is clear that lead is not an essential dietary component and it is prudent to avoid the addition of lead to food during its processing or preparation. This begins with the obvious exclusion of food contact with any lead-containing alloy in food processing and preparation. It also means attention to more subtle details, such as the exclusion of lead-based solder from the repair of food-handling equipment and being aware of the potential risk of lead contamination from improperly glazed pottery containers.

See also: **Contamination of Food**; **Renal Function and Disorders**: Kidney: Structure and Function

Further Reading

Carrington CD and Bolger PM (1992) An assessment of the hazards of lead in food. *Regulatory Toxicology and Pharmacology* 16: 265–272.

Ehle AL and McKee DC (1990) Neuropsychological effects of lead in occupationally exposed workers: a critical review. *Critical Reviews in Toxicology* 20: 237–255.

Mahaffey KR (ed.) (1985) *Dietary and Environment Lead: Human Health Effects*. Amsterdam: Elsevier Science.

Smith MD, Grant LD and Sors AI (eds) (1989) *Lead Exposure and Child Development: An International Assessment*. Dordrecht: Kluwer.

Tsuchiya K (1986) Lead. In: Friberg L, Nordberg GF and Vouk VB (eds) *Handbook of the Toxicology of Metals*, vol. II. Amsterdam: Elsevier Science.

US EPA (1986) *Air Quality Criteria for Lead*. Research Triangle Park, North Carolina: US Environment Protection Agency.

LEAVENING AGENTS

J L Vetter, Technical Assistance Services, Leawood, KS, USA

History of Chemical Leavening

Bakery foods that depend on chemical leavening systems for their appearance, texture, and taste are relatively recent developments. Whereas yeast-leavened products have nourished humans for over 4000 years, chemically leavened products have been available for less than 200 years. We have come to enjoy a wide variety of chemically leavened products as part of our daily menu. Some are sweet, e.g., cakes, doughnuts, and biscuits, whereas others are more bland, e.g., baking powder biscuits and many types of crackers (see **Figures 1** and **2**).

Leavening with Yeast

The word 'leaven' comes from the Latin word '*levare*' or '*levo*,' which means 'to raise' or 'to make light by aeration.' Leavened or aerated baked foods can be

Figure 1 Leavening agents: machine-cut doughnut production. (Courtesy of Rhone-Poulenc, New Jersey.)

Figure 2 Examples of chemically leavened baked goods.

traced as far back as 2000 BC, when Egyptians used yeast to leaven bread. (*See* **Bread**: Breadmaking Processes; **Yeasts**.)

Chemical Leavening

Yeast was the primary source of leavening gas to raise baked foods until late in the eighteenth century, when it was discovered that sour milk or buttermilk would release carbon dioxide gas from sodium bicarbonate (baking soda) during baking of a dough or batter to which these ingredients had been added. Carbon dioxide from baking soda is the same gas as that released by yeast during fermentation of a dough.

Sour milk and buttermilk were replaced in many recipes with cream of tartar after about 1835. Cream of tartar (potassium bitartrate) was obtained as a byproduct from wine-making. There were early attempts to combine the baking soda and cream of tartar into a more convenient blended product, and this led to the development of baking powders. A US chemist by the name of Hoagland, living in Fort Wayne, Indiana, originated the cream of tartar baking powder in 1850 and founded the Royal Baking Powder Company.

Sour milk, buttermilk, and cream of tartar all had one thing in common: they released the carbon dioxide from baking soda very rapidly, and much of the leavening gas was actually lost during the mixing of a dough or batter, resulting in a baked product with less volume than desired. There was a need for a material that would not react so quickly and allow sufficient time for mixing and forming products before the leavening gas was released. This led to the development in the mid-1800s of monocalcium phosphate (MCP; also known as acid calcium phosphate, ACP) as an alternative to cream of tartar. Although MCP is considered a 'fast-acting' leavening acid, it is a little slower than cream of tartar. A patent was issued in 1864 for such an acidic phosphate leavening agent. However, better agents for controlling or delaying leavening activity were still required, and this need led to the development of a number of alternatives to cream of tartar and MCP for use with baking soda. Sodium aluminum sulfate (SAS) was developed in the early 1890s, sodium acid pyrophosphates (SAPPs) in the early 1900s, and sodium aluminum phosphates (SALPs) around 1960.

Design of Leavening Systems

There are three primary factors to consider in designing a chemical leavening system for a specific product. These are:

1. How much leavening gas is needed to produce the desired finished product?
2. What leavening acid is needed to achieve the desired product characteristics?
3. What other effects might the leavening acid have on the dough, batter, or finished product?

This article will discuss these three factors and describe the characteristics of commercially available chemical leavening agents and their application in specific bakery foods.

Mechanisms of Chemical Leavening

Air and Water as Leaveners

Although air and water are 'chemicals' in the truest sense of the word, they are generally not considered 'chemical leavening agents.' Nevertheless, air and water do provide leavening activity in a number of bakery products. Air, which is incorporated into a dough or batter as small gas cells during mixing, will expand during baking and cause some leavening of the product. In fact, air is the primary leavening agent in a fat-free sponge cake. During baking, water contained in the dough or batter will

be vaporized by the heat of the oven and cause a leavening – or raising – of the product. Water is the primary leavening agent for saltine crackers, contributing to the tender, flaky texture of the product. The contributions of air and water to the leavening and characteristics of baked products should not be overlooked in designing a leavening system based on more traditional 'chemical' leavening agents.

Leavening by Decomposition

There are two primary chemical leavening agents that can release carbon dioxide gas by decomposition during baking. These are sodium bicarbonate (baking soda) and ammonium bicarbonate. Heat is required for these materials to decompose, and the temperature of baking is generally adequate for this purpose. Sodium bicarbonate decomposes when dissolved and heated to release carbon dioxide according to eqn (1).

$$2NaHCO_3 \xrightarrow{\text{Heat}} H_2O + CO_2 + Na_2CO_3. \quad (1)$$

Ammonium bicarbonate decomposes when dissolved and heated according to eqn (2).

$$NH_4HCO_3 \xrightarrow{\text{Heat}} NH_3 + H_2O + CO_2. \quad (2)$$

Both ammonia and carbon dioxide are gases and contribute to the leavening of the baked product.

Leavening by Chemical Reaction

The vast majority of chemical leavening systems are based on the reaction of an acid with sodium bicarbonate to release chemically the carbon dioxide from the soda, and it is with these systems that the majority of this article is concerned. A general equation for the reaction is shown in eqn (3).

$$HX + NaHCO_3 \xrightarrow[\text{Heat}]{\text{Moisture}} \underset{\text{(Neutral sodium salt)}}{NaX} + H_2O + CO_2. \quad (3)$$

In addition to the carbon dioxide leavening gas, water is released, and a sodium salt of the leavening acid is formed. The nature of the neutral sodium salt depends on the leavening acid being used.

There are a number of leavening acids from which to choose in formulating a chemical leavening system. They differ in the amount required to release completely all of the carbon dioxide in the soda, the speed with which they release this leavening gas, and their effects on doughs, batters, and finished product characteristics.

Neutralizing value The neutralizing value (NV) is a term used to describe the amount of leavening acid needed to react completely with the amount of baking soda used in formulating a bakery product. If all of the soda reacts with the leavening acid, the finished product should be close to neutral in pH, which is desired in most products. Higher (alkaline) or lower (acidic) pHs can be achieved, if desired, by adjusting the amount of leavening acid and/or soda from the amounts needed for complete reaction.

The NV is defined as the weight of soda neutralized by 100 parts of leavening acid. Since most formulae for bakery products begin with a given weight of soda (the weight needed to give the desired amount of leavening gas), the NV is most commonly used to determine the amount of leavening acid required for the amount of soda used according to the following equation:

$$\%\ \text{Leavening acid} = \frac{\%\ \text{Soda} \times 100}{\text{NV}}. \quad (4)$$

The NVs for commonly used leavening acids are listed in Table 1. It should be noted that larger amounts of leavening acids with low NVs are required for a given amount of soda than for those leavening acids with higher NVs. This could be a factor in the economics of developing a leavening system.

Rate of reaction The speed with which a leavening acid reacts with baking soda to release carbon dioxide leavening gas is very important in controlling the characteristics of the finished baked product. If the acid reacts too rapidly with the soda, all of the leavening gas could be released during mixing and, therefore, would not be available to raise the product during baking. In this case, the finished product would be low in volume and dense in texture. However, if the acid reacts with the soda too late in the baking process, the structure of the product will be 'set' by the heat of baking, and the leavening gas cannot raise the product without causing cracks or splits.

The release of some leavening gas, or incorporation of some air, during mixing is desirable for the formation of small air or gas cells in the dough or batter. This process, called 'nucleation,' is responsible for the development of a fine and uniform grain in the final

Table 1 NVs of commonly used leavening acids

Leavening acid	*NV*
Monocalcium phosphate monohydrate	80
Anhydrous monocalcium phosphate	83
Sodium acid pyrophosphate	74
Sodium aluminum phosphate	100
Dicalcium phosphate dihydrate	33
Dimagnesium phosphate	40
Sodium aluminum sulfate	104
Cream of tartar	45
Glucono-δ-lactone	45

product. However, most of the leavening gas should be reserved for release at the proper time during baking in order to achieve the total volume desired.

A number of factors affect the proper time for release of gas during baking. Product size is certainly important. A small product will heat, and the structure will set, faster than a larger product. Oven temperature obviously affects the rate at which a product's temperature rises to the point of setting the structure. Other ingredients in the formula can affect the temperature at which a product sets. The amount of time to mix fully a dough or batter and the time required to form or deposit the product to be baked will influence the time at which the leavening acid should react with the baking soda.

The manufacturers of leavening acids have attempted to meet all the diverse needs of the household, institutional and large-scale wholesale baker by developing and marketing products with a wide range of 'reaction rates.'

Some materials have essentially no delay in their reaction with baking soda. They react as soon as they dissolve in the water present in the formula; the soda has probably already dissolved. These materials include the lactic acid in sour milk or buttermilk, cream of tartar, and MCP monohydrate.

Anhydrous monocalcium phosphate (AMCP) has a time delay, which, although relatively short, does allow time for preparing doughs and batters for baking. A major use for AMCP has been self-raising flour, which is often used to prepare baking powder-type biscuits. The time delay of AMCP allows sufficient time for mixing, rolling, and cutting the dough prior to baking.

Manufacturers of SAPPs have been able to vary the amount of time delay of these leavening acids from a few minutes to as much as an hour or more. This is done by controlling the conditions under which the SAPP is manufactured.

SAS, SALP, and dicalcium phosphate dihydrate (DCP) are basically heat-triggered leavening acids. They have essentially no reaction with soda until the temperature of the product is raised by the heat of baking. In some baked products, this delay may be too long, and SAS, SALP, and DCP perform better when used in combination with a faster-acting leavening agent such as MCP or AMCP. Of course, all leavening agents are essentially 'heat triggered,' since they dissolve and react with soda faster as the temperature of the dough or batter increases. Nevertheless, the term 'heat triggered' is usually reserved for those materials that have an almost indefinite time delay until the temperature rises to a certain point, in contrast to those leavening agents that are both time and temperature triggered.

Table 2 Rates of reaction with baking soda for commonly used leavening acids

Leavening acid	*Rate of reaction*
Cream of tartar	Rapid
Monocalcium phosphate monohydrate	Intermediate
Anhydrous monocalcium phosphate	Intermediate (slower)
Sodium acid pyrophosphates	Slow
Sodium aluminum phosphate	Very slow
Sodium aluminum sulphate	Very slow
Dicalcium phosphate dihydrate	Very slow
Dimagnesium phosphate	Slow
Glucono-δ-lactone	Slow

Glucono-δ-lactone (GDL) does not fit any of the above categories of no delay, time delay, and temperature triggered. GDL reacts continuously, but slowly, with baking soda. The speed of reaction, of course, increases with temperature and becomes quite rapid during baking. The steady release of carbon dioxide in a dough containing GDL is very similar to the release of carbon dioxide by yeast in a fermenting dough. Therefore, GDL has been popular in the development of chemically leavened bread.

The rates of reaction of commonly used leavening acids are listed in Table 2.

Ionic effects of leaveners Although the primary function of chemical leavening systems is to leaven, or raise, a product, leavening agents can also affect characteristics of doughs, batters, and the finished baked product. These effects are due to reactions between the leavening agents and components (particularly starch and protein) of the dough or batter. These are called 'ionic' effects, since the leavening agents dissolve in the water of the dough or batter and dissociate into positive cations and negative anions. In some cases, it is the cation (such as calcium ions) that is responsible for the observed effect. In other cases, it is the anion (such as pyrophosphate) that causes a change in the dough, batter, or finished product.

The calcium and aluminum cations in phosphate-based leavening acids provide more resiliency to cake products than the sodium cation in other phosphate-based leavening agents. Therefore, if resiliency is a property desired in a cake product, the formulator may choose a leavening system containing calcium and/or aluminum ions. Calcium ions tend to stiffen doughs and thicken batters; they also tend to dry out moist or slightly sticky doughs. These properties of doughs and batters are important considerations in handling and processing, particularly in high-speed, automated commercial bakeries. (*See* **Bread**: Dough Mixing and Testing Operations.)

The pyrophosphate anion in SAPP leavener interacts with proteins from other ingredients in the formula. This tends to result in a moister texture, which is usually desirable. Unfortunately, the pyrophosphate ion also imparts a slightly bitter aftertaste when used at higher levels. Therefore, the product formulator must try to balance such desirable and undesirable effects in developing a leavening system for a specific product.

Final product pH can affect the color of a baked product, and the pH can be controlled by the leavening system. In white products such as white layer cakes, a pH that is slightly below neutral accentuates whiteness. This lower pH can be achieved by adjusting the balance between leavening acid and baking soda (more acid, less soda) or by selection of leavening acid. Pyrophosphate leavening acids tend to buffer pH in the range 7.3–7.5 and it is somewhat difficult to lower the pH by adjusting the pyrophosphate and soda balance. Calcium phosphates, however, do not have the pH buffering capacity of the pyrophosphates, and pH can be manipulated more easily by adjusting the acid/soda balance. The color and flavor of chocolate products are enhanced by higher pH. Again, this can be achieved by selection of leavening acid and/or adjustment of the acid/soda balance (less acid, more soda).

The development of a suitable leavening system is a highly complex activity. The formulator has a number of objectives to meet, and fortunately, there is available a number of leavening agents with varying properties to help meet these objectives.

Characteristics of Leavening Agents

Sodium Bicarbonate (Baking Soda)

As mentioned previously, the heat of baking can cause baking soda to decompose, giving off carbon dioxide leavening gas without reacting with a leavening acid. The sodium carbonate formed by the decomposition of baking soda is very alkaline and will tend to give the baked product a high pH. This is desirable in some applications, because the high pH enhances certain flavors, such as chocolate. Many recipes for chocolate chip cookies call for leavening only with baking soda.

Ammonium Bicarbonate

Ammonium bicarbonate will also decompose during baking and release ammonia and carbon dioxide leavening gases without reacting with a leavening acid. Unlike sodium bicarbonate, which leaves a residue of alkaline sodium carbonate, ammonium bicarbonate leaves no residue when it decomposes by heat. It, therefore, has no effect on the pH of the baked product. If there is more than about 5% moisture in the baked product, however, the ammonia gas will dissolve in this water and impart an ammoniacal flavor to the product. For this reason, ammonium bicarbonate is used only in low-moisture products such as crackers.

Cream of Tartar

Cream of tartar (potassium acid tartrate) reacts very rapidly with soda in a dough or batter, releasing 70–80% of the leavening gas within 2 min of completion of mixing. For this reason, it has little application in commercial production of bakery foods, since there is insufficient time to process the dough prior to baking.

Sodium Aluminum Sulfate

SAS is a very slow-reacting leavening agent and does not release leavening gas until the product is in the oven, and its temperature rises. SAS is most often used in combination with a faster-acting leavening acid such as MCP. This combination (SAS and MCP) is the most popular composition of household baking powders. SAS is not used extensively in commercial baking.

Monocalcium Phosphates

Monocalcium phosphate leavening acids are available in two forms: the monohydrate (MCP) and the anhydrous product (AMCP). The monohydrate is fast-acting, although slightly slower than cream of tartar. It is usually used in combination with slower-acting leavening agents. MCP will release leavening gas from baking soda during mixing, which is desirable in achieving a fine, uniform grain in the baked product as long as a slower-acting leavener is present to provide leavening gas during baking. AMCP is coated with a slowly dissolving phosphate material that delays its reaction with baking soda. This delay is relatively short, but sufficient for some products that are baked in the home. These products include biscuits, pancakes, and waffles.

Sodium Acid Pyrophosphates

SAPPs are available with a range of reaction rates from slow to very slow. SAPPs are used almost exclusively in some applications such as cake doughnuts and canned, refrigerated biscuit doughs. They are one of several options available to the formulator of cakes, biscuits, pancakes, etc. The aftertaste, to which many people are sensitive, somewhat limits the use of

SAPPs, particularly since the development of SALP leavening agents.

Sodium Aluminum Phosphates

SALPs are the most recently developed chemical leavening agents. They are very slow-reacting with baking soda and are often marketed as blends with a faster-acting leavener such as MCP or AMCP. SALPs have become quite popular with product formulators, because of their high NV, relatively low cost, lack of undesirable flavors, and ability to be blended with other leaveners to achieve optimized properties for specific applications.

Dicalcium Phosphate Dihydrate

DCP is not technically a leavening 'acid,' because it is an alkaline phosphate salt. However, when incorporated into a dough or batter and subjected to the heat of baking, DCP decomposes to MCP and tricalcium phosphate (TCP). The TCP is insoluble, precipitates out of the system, and leaves MCP – a fast-acting leavening agent – to react with baking soda. The temperature at which DCP decomposes is relatively high (55–60 °C), so that decomposition occurs rather late in the baking process. For this reason, DCP is classified as a very slow-acting leavening agent and is commonly used in combination with a fast-acting leavener.

Dimagnesium phosphate

Magnesium phosphate is a new leavening acid that is slow-acting and is activated by heat (40 °C). It has a clean flavor profile, provides a uniform cell structure, and imparts moistness to baked product crumbs. It is particularly useful in frozen and fresh baked cakes, muffins, biscuits, and pizza crusts.

Glucono-δ-lactone

GDL is an organic compound that forms gluconic acid when dissolved in water. It reacts slowly, but steadily with baking soda. The rate of reaction, of course, increases as the temperature of the dough or batter increases during baking. GDL has a low NV of about 45, requiring about two parts of GDL for every one part of soda. It is, therefore, relatively expensive to use and finds limited application in specialized products.

Summary

The technology of chemical leavening is quite complex, because of the varying requirements of different products and processes. It is made even more complex by the availability of a variety of chemical leavening agents, each with its own set of properties, advantages and disadvantages.

An understanding of the technology of chemical leavening should lead to the production of even better, cost-effective, and consumer-satisfying bakery foods.

See also: **Bread**: Dough Mixing and Testing Operations; Breadmaking Processes; **Yeasts**

Further Reading

Dubios DK (1981) Chemical leavening. *Technical Bulletin*, vol. 3, No. 9. Manhattan: American Institute of Baking.

Kichline TP and Conn JF (1970) Some fundamental aspects of leavening agents. *The Bakers Digest* August: 36–40.

LaBaw GD (1982) Chemical leavening agents and their use in bakery products. *The Bakers Digest* February: 16–21.

Reiman HM (1977) Chemical leavening systems. *The Bakers Digest* August: 33–36, 42.

Lecithin *See* **Phospholipids**: Properties and Occurrence; Determination; Physiology

Leeks *See* **Onions and Related Crops**

LEGISLATION

Contents

History

P Berry Ottaway, Berry Ottaway & Associates Ltd, Hereford, UK

Background

The history of food legislation can be traced back over millennia. Laws have been introduced into all civilized society for one or more of four main reasons. The first is food safety and protection of the population; the second is fiscal, for the collection of taxes; the third covers various aspects of the adulteration of food; and the fourth is concerned with weights and measures.

Food Safety

Although food safety now predominates in the legislation of our modern society, it was not until the nineteenth century that food laws protecting public health became important in western Europe and North America. However, laws based on early experience of food safety problems go back many thousands of years. The detailed Jewish dietary laws (kashrut) can now be rationalized with our knowledge of food safety, microbiology, and parasitology. The dietary laws are mainly written down in Deuteronomy and Leviticus in the Old Testament in the Bible.

In these works, the Israelites were given the rules on permitted and forbidden foods. The only animal (mammal) meat allowed was from those species that chewed the cud, parted the hoof, and were cloven-footed. This meant that sheep and deer were allowed as food but the pig was strictly forbidden. It had already been deduced that the meat from grazing herbivores was likely to be safer than that from a scavenging omnivore.

Similar explanations can now be applied to the rest of the detail of the early Jewish food law. All birds of prey, such as vultures, eagles, seagulls, and owls, had to be excluded from the diet. Of the lower orders of animals, none which were creeping and go on all-fours, or winged, swarming animals were allowed, except if the legs of the swarming beasts were jointed so that they could jump. Within the insects these definitions separated the disease-carrying flies and cockroaches from the grasshoppers and locusts; locusts are still a common food in many parts of the world.

Among the aquatic animals only those fish that possess both scales and fins were permitted. Crustaceans such as prawns, lobsters, and crabs, and all shellfish may not be eaten. Again, the scavengers with their higher risk of disease are excluded.

Even with the permitted animal foods there are further rules. No animal that has died naturally or from the result of an attack from another animal must be eaten and complex laws were developed on the separation of meat and dairy products and the use of utensils. These strict dietary laws originated thousands of years ago and many are still strictly observed today by followers of both Judaism and Islam.

It was not until the second half of the nineteenth century that legislation relating to the public health and safety aspects of food began to be generally established in Europe.

In the UK the 1860 Adulteration of Food and Drink Act prohibited the sale of impure food and food to which materials or ingredients injurious to health had been added. This was strengthened in 1872 by the Adulteration of Food and Drugs Act which required the appointment of public analysts and empowered local enforcement officers to take samples and bring prosecutions.

The Sale of Food and Drugs Act, passed in 1875, increased the scope of the previous legislation by making it an offence to sell, to the prejudice of the customer, food which was not of the nature, quality, or substance demanded by the consumer. In the same year, the Public Health Act 1875 permitted the local enforcement authorities to take preemptive action by inspecting food. That which was considered to be 'diseased or unsound, or unwholesome or unfit for

the food of man' could be officially seized and taken before a justice of the peace.

It was over 50 years later, in 1928, that significant changes were made to the 1875 Acts, and in 1938 a consolidation took place and the appropriate sections of the Sale of Food and Drugs Act and the Public Health Act were brought together in the Foods Act 1938. The 1938 Act introduced the offence of applying false or misleading labels to food and also contained the powers to make regulations to govern the preparation, storage, and sale of food. This Act brought most of the British food safety legislation into place. The 1938 Act was further consolidated in 1955 (Food and Drugs Act), by the Food Act 1984, and the Food Safety Act 1990.

During the second half of the nineteenth century and first half of the twentieth century, the developments in food safety legislation in the UK had a very significant influence on the introduction of similar legislation into the British dominions and colonies at the time.

Fiscal Laws

The idea of levying taxes on foods goes back more than two millennia in Europe. Import and export duties (*portoria*) on goods, including food commodities, formed an important part of the well-organized taxation system of the Roman Empire.

In the UK the earliest written reference to customs dues being levied is to be found in a charter dated 743. In a document from 979 there is evidence of duty to be paid on cheese, butter, and eggs coming into England through the port of Billingsgate. Until the beginning of the thirteenth century, there were specific grants giving permission to individual ports to charge duties on imported foods. Introduction of a nationally organized English customs service started in the reign of King John (1166–1216). In 1203 a duty of one-fifteenth of the value was placed on all imports and exports, including food. This was called the quindecima. It was decreed that the duty raised be paid direct to the treasury and removed the permission for it to be retained by the ports. This established a basic centralized system of customs control at the ports, which has survived in the UK with very little change to the present day.

For a number of centuries, most of the food-related laws in what is now the UK were based on the need to raise taxes and at one time or another a wide variety of food and drink has been subject to tax. It is interesting to reflect that a tax imposed on tea by the English in 1773 initiated the American War of Independence and the American colonies breaking away from the control of Europe, eventually to become the USA.

The Adulteration of Food

Sadly, history shows that where food is traded there is the temptation to adulterate or misdescribe the food to make illicit profits. It would appear that in all ages and societies the temptation has been irresistible. An Indian law dated to about 300 BC prohibited the adulteration of grain, scents, and medicines.

Documents in the UK show that food adulteration has been a problem for nearly a thousand years. For centuries the methods of adulteration, and methods of detection, were very crude and adulterated foods often caused illness and even death.

In England during the Middle Ages, the trade guilds took the first positive actions against adulteration of the foods for which their members were responsible. This was voluntary regulation by the tradesmen in the guilds to insure their reputation and also to insure a trading advantage. As time went on, the voluntary controls began to be replaced or supplemented by statutes.

In each age to the present day it has been the case that the more expensive items of food and drink are at risk of adulteration and manipulation. Thus, sugar, spices, tea, coffee, wine, beer, and spirits have all been the targets of malpractice. Milk was commonly diluted with water and a proportion of butterfat removed from the top of the milk. Bread was also commonly adulterated and records show that in a number of towns in medieval England bakers were pilloried. In the more serious cases, where widespread illness or death resulted from the malpractice, the tradesmen were sent to the gallows.

During the eighteenth century, legislation was introduced (in 1718) to prevent the adulteration of coffee. This was followed by a similar law on tea in 1730 and one on bread in 1758. The weakness of these laws at the time was due to the lack of effective analytical methods. Most methods that had been developed for the detection of adulteration were based on taste, smell, or texture.

It was not until 1860 that the Adulteration of Food and Drink Act was passed in England. This was more broad-based than the previous commodity-related legislation and it created an offence of knowingly selling food containing injurious material. Between 1860 and 1870, similar laws were passed in Germany and Sweden.

In the USA, Professor GW Wigner won a Board of Trade prize in 1880 for a paper recommending legislation to combat food adulteration. The US Congress failed to pass the first act drafted on the basis of the paper, but between 1880 and 1890 several states passed laws based on Wigner's work.

Weights and Measures

From the earliest days of human communities bartering or trading food, there has been a requirement for a mutually acceptable system of weights and measures.

The oldest weights identified by archeologists are believed to be between 9000 and 10 000 years old. There is also evidence that trade in weighed products was used by the Indus civilization and in Sameria about 5000 years ago. A little later, weighed trade was adopted by the Babylonians and then by the Egyptians about 4500 years ago.

There are also references in the Bible (Old Testament) to the importance of weights and measures in a structured society. They appear where Moses is told the laws for the Israelites in Leviticus: 'Ye shall do no unrighteousness in judgment, in metyard, in weight or in measure. Just balances, just weights, a just ephah and a just hin, shall ye have' (ephah and hin were measuring equipment of that time). Also the Book of Proverbs states: 'A just weight and balance are the Lord's; all the weights of the bag are his work.'

The Romans introduced an organized system of weights and measures into all the countries which became part of the Roman Empire. The Roman libra was almost the same weight as a standard measure used by the Egyptians and another used by the Babylonians. For a number of centuries after the collapse of the Roman Empire, practically every country in Europe had a weight with its name derived from either the word *libra* or *pondus*, the latter being the origin of the word pound. The pound was the principal unit of measure in UK law until it began to be phased out in the mid-1980s in favor of metrication and as part of the UK harmonization with the rest of the European Community. It was finally abolished as a measure of food in 2000. The pound is still retained in the USA.

In the Dark Ages that followed the collapse of the Roman Empire, the system of weights and measures deteriorated and in northern Europe there was a gradual return to the primitive ways of measuring, for example, with the human body. The situation developed where almost every main town had its own standards which could differ from those of its neighbors. This resulted in considerable confusion in trade.

In England, the earliest law to introduce and enforce consistent standards of weights and measures was introduced during the reign of King Edgar (AD 959–975). To quote from a translation: 'And let one money pass through the King's Dominions and that let no man refuse; and let one measure and one weight pass, such as is observed at London and Winchester.'

A major change in the weighing and measuring of food began in France during the Napoleonic period, with the introduction of the metric system. The principal unit of weight became the gram, and that for volume became the milliliter.

The metric system originated in 1791 when a committee of the French Academy made a report to the National Assembly on proposals for a new scientifically based method of measurement. These proposals were subsequently adopted in France. The report defined the standard of length as the meter (metre), which was one ten-millionth of the earth's meridian quadrant at sea level.

The unit of weight, the gram (g), became a secondary standard based on the centimeter (one-hundredth of a meter), and the density of water at a given temperature. Minor inaccuracies were encountered in the measurement of length and in 1875 the International Bureau of Weights and Measures was established at Sèvres in France. The meter was redefined as 1 650 763.73 wavelengths of the orange-red light emitted when a gas consisting of pure krypton-86 is excited in an electrical discharge. A standard kilogram (1000 g) was produced from a platinum-iridium alloy and is carefully preserved at Sèvres.

Through the twentieth century, the metric system was adopted by more and more countries and has now become the main system of measurement for science and also for trade. One of the main areas still retaining the imperial measurements of pounds, ounces, and gallons is the USA.

Early Development of Food Law in the USA

During the late nineteenth and early twentieth centuries, the development of food legislation in the USA was complex. Congress did not have constitutional authority over the regulation of manufacturing processes or the provision of pure and safe food. The federal laws were primarily concerned with interstate and international trade. Early federal food legislation was mainly concerned with the export and import of meats such as a law in 1890 requiring the inspection of salted pork and bacon. This was introduced to facilitate exports as a number of European countries were concerned about contamination with trichinae.

Towards the end of the nineteenth century, there was an intermingling of state and federal laws. Among laws passed by individual states was the California pure food and drug law of 1850; the Massachusetts law prohibiting milk adulteration in 1856; the New York food and drug law in 1881; and the District of Columbia food and drug law of 1891.

Chiefly as a result of a campaign by Dr Harvey W Wiley, chemist for the US Department of Agriculture, against adulterated and misbranded foods, President Theodore Roosevelt signed the Food and Drug Act in 1906. This was after 12 similar bills had failed over the previous 20 years.

In 1927 a separate enforcement agency covering food was constituted as the Food, Drug and Insecticide Administration. This became the Food and Drug Administration in 1931.

In 1939 the Federal Food, Drug and Cosmetic Act was passed which, whilst retaining most of the provisions of the 1906 Act, added new provisions such as the definition of food, food standards of identity, and making truthful labeling mandatory.

The Hale Amendment in 1954 allowed for a more simplified method of promulgating food standards, and the Food Additive Amendment of 1958 required that any substance added to food must be proved to be safe by the manufacturer before being offered for sale. It also permitted the use of food chemicals for technological purposes where they were safe at the intended level of use. The Delaney proviso prohibited the approval of any food additive shown to induce cancer in humans.

Following the enactment of the Food Additives Amendment, the Food and Drug Administration published the first list of substances generally recognized as safe (GRAS) in the Federal Register. This list contained nearly 200 substances.

In 1960 the Color Additive Amendment was enacted requiring manufacturers to establish the safety of all color additives intended for use in foods, drugs, and cosmetics. This was also subject to the Delaney proviso prohibiting the approval of any color additive shown to induce cancer in humans or animals.

With the establishment of the Environmental Protection Agency in 1970, the program for setting limits and tolerances for pesticides was transferred from the Food and Drug Administration to the new agency.

Food Law of the European Union

The most ambitious and comprehensive program to develop new food law was that undertaken in the European Union in the second half of the twentieth century.

From the signing of the Treaty of Rome, which brought the European Economic Community into existence in March 1957, it was realized that in order to insure freedom of trade in foodstuffs it was essential to establish common rules on the presentation, control, and in some cases the composition of the foods.

Of particular concern in the first years of the Community were the diverse national laws regarding the use of additives in foods and the first directive in the food sector was that on food colors which was adopted in 1962. During the early 1960s, work was in progress on directives to control other groups of food additives.

The first real step towards the harmonization of the food legislation was a European Commission recommendation published in 1965 which required prior notification from the member states of any intended new legislation or amendments to existing legislation on foodstuffs.

Four years later, in 1969, the European Council issued a resolution which included the drawing-up of a program for the elimination of technical barriers to trade in foodstuffs. The program that emerged as a result of this resolution was ambitious as it was divided into priority groups covering 70 subjects with a time schedule built around five phases, with completion in 1971, 2 years after the publication of the program. As very little of the program had been accomplished by the 1971 deadline, a new resolution in 1973 revised the program and extended the deadline to 1978. Again, progress was slower than anticipated and a further revised program was adopted in 1977.

One of the main objectives of the harmonization program during the 1970s was the development of compositional or 'recipe' legislation, covering specific groups of foodstuffs. This 'vertical' legislation, as it came to be known, was considered essential for free movement of foods and a large number of directives were envisaged, each concerned with the composition and control of a product category. By the mid-1970s, over 30 categories of foodstuffs had been identified as requiring recipe or vertical legislation.

The first of these directives was that on cocoa and chocolate products, which was still in force in 2000, having been amended 12 times. Although work commenced on over 25 vertical directives during the first half of the 1970s, very little of this proposed legislation had been adopted by 1977. Most of the delays were caused by lack of agreement between the member states as to what was important for the control of the product category. One of the problems associated with the vertical legislation was that, to a considerable degree, it restricted product development and the introduction of technological innovation.

A major turning point in the progress towards harmonization and an important influence on the revised programs for the 1980s was the European Court of Justice judgment on the Cassis de Dijon case. This concerned a German prohibition on the import of the blackcurrant liqueur Cassis de Dijon

from France as the alcohol content was lower than that permitted by German law. The Court ruled that the barrier created by the German regulation was incompatible with the Treaty of Rome and that quantitative restrictions on imports should be prohibited. This judgment provided a general principle for the trade in foods between member states which was that, provided a foodstuff complies with the regulations in force in the member state in which it is manufactured, it should be permitted on the market of any other member state.

The Cassis de Dijon judgment is now regarded as being pivotal in the major change of approach by the European Council and Commission to the harmonization program. There was a shift of emphasis away from the vertical directives and it was determined that future Community legislation on foodstuffs should be limited to provisions which are justified by the need to protect public health, provide consumers with necessary information about foods, provide for necessary public controls, and insure fair trade.

The new approach to harmonization was given an additional impetus in mid-1985 with the publication of a white paper from the Commission to the Council entitled *Completing the Internal Market*. The white paper addressed the fact that there has been a significant imbalance between the 'horizontal' food legislation (e.g., labeling and additives) where considerable progress had been made and the 'vertical' legislation covering composition in which there had been relatively little progress.

The white paper, and a Commission Communication which followed it, defined a new approach in which the vertical system was to be replaced by a horizontal one. The procedures were also simplified to reduce the work involved and to speed up the harmonization process.

The new program was to be centered on five framework directives that were considered crucial to the single market in foodstuffs. These were:

1. Additives
2. Food labeling and presentation
3. Materials and articles in contact with food
4. Official control of foodstuffs
5. Foods for particular nutritional uses

The framework directives were required to lay down the general principles for control in the relevant areas and specific technical directives were to be developed as adjuncts to these framework directives.

As some of the five key areas had already been covered by earlier directives, priority was to be given to revising them and bringing them into line with the new requirements. However, even with the more simplified and streamlined program, progress was slow and much of the essential proposed legislation was not adopted by the introduction of the single market in January 1993.

At the end of 1999, the European Commission published a white paper on food safety which included proposals to set up a centralized European Food Authority and which also set timetables for the completion of over 80 outstanding items of food legislation.

In anticipation of their eventual incorporation into the European Union, a number of states in the former Soviet Union, particularly Hungary, Poland, and the Czech Republic, have been revising their food laws in line with those of the European Union.

Codex Alimentarius

Since the mid-1960s, the Codex Alimentarius Commission has played an increasing role in shaping the future of international food law.

Codex was instituted as a joint Food and Agriculture Organization (FAO) and World Health Organization (WHO) initiative to develop internationally accepted standards and codes of practice for food commodities. All member nations and associate members of the FAO and WHO can become members of Codex. By 2000, the number of member countries had reached 165, representing 97% of the world's population. Over the years a wide range of standards covering a number of food categories have been agreed, and in some countries these standards form the basis for national laws.

The Codex Alimentarius has gained a greater significance since the formation of the World Trade Organization (WTO). The Agreement on the Technical Barriers to Trade (TBT) that was introduced following the Tokyo Round on World Trade in 1979 had a substantial impact on the establishment of policies on food control. The TBT Agreement did not specifically mention Codex but dealt with aspects of food not directly related to safety, such as labeling, quality, and packaging, and thus impinged on Codex. The WTO, however, has recognized Codex as the preferred international organization for the arbitration and settlement of disputes related to food trade.

See also: **Adulteration of Foods**: History and Occurrence; **Food Additives**: Safety; **Food Safety**; **Legislation**: Additives; Codex

Further Reading

Furia TE (ed.) (1972) *Handbook of Food Additives*. Boca Raton, FL: CRS Press.

Marsh SB and Soulsby J (1982) *Outlines of English Law*, 3rd edn. London: McGraw-Hill.

Rowell R (ed.) (1999) *Butterworth's Law on Food and Drugs*. London: Butterworths.
Smith G (1980) *Something to Declare: 1000 Years of Customs and Excise*. London: Harrap.
Food and Agriculture Organization/World Health Organization (1999) *Understanding the Codex Alimentarius*. Rome: FAO/WHO.

International Standards

J A Smith, Leatherhead Food Research Association, Leatherhead, Surrey, UK

Standards

The initial object of standardization was to provide a means of communication between the supplier and the user. Standards are, in principle, voluntary instruments but, as legislation began to deal with more complex technological developments, legislators referred to standards as a means of achieving the desired objectives of legislation. Before the advent of detailed food labeling, food standards provided the only way to guarantee the consumer that the desired quality in terms of product characteristics was maintained, and they thus became frequently referred to in legislation or laid down under mandatory acts.

Food standards generally lay down requirements for the composition of a food in terms of percentage and quality of ingredients and may include provisions on food additives. They may also specify physical properties, such as water content, particle size, density, or processing methods.

The Codex Alimentarius Standards of the Food and Agriculture Organization/World Health Organization (FAO/WHO) are, in effect, model laws for use in international trade; by granting full acceptance of a Codex standard, a country or trading bloc engages to accept conforming products on its market.

Food standards have generally been established for more basic food commodities, such as flour, bread, and meat products, in order to insure minimum standards are met, but have also been widely applied to products such as soft drinks and chocolate, which are consumed by a mass market. In insuring that basic standards are met, food standards have partly a social function in guaranteeing the quality of basic necessities. Composite products, the appeal of which depends on diversity, are rarely the subject of standards. Standards are usually tied to a product denomination or name and, since they reflect national tradition and characteristics, there can be different standards in different countries for the same product descriptor.

Nowhere has this been more evident than in the European Union (EU) with its variety of culinary traditions. Examples include use of durum wheat only in pasta in France and Italy, differences in composition of sausages in the UK and Germany, and the use of vegetable fat in chocolate throughout the EU.

Attempts during the 1980s to harmonize compositional food standards were unsuccessful due to the unwillingness of the then European Community (EC) member states to broaden the scope of their own standards and come to a common level with others. It is for this reason that the establishment of compositional standards *per se* is no longer actively pursued at EC level; the emphasis now is on public safety, removal of barriers to trade, and prescribed food labeling in order to allow the consumer to make an informed choice.

This trend away from prescribed standards of composition has become international in recent years, as such standards are seen as restrictive in terms of product development and obstructive to trade. However, a large number of food standards are still in force worldwide and are still a significant factor when considering the scope of technical food legislation.

In this article, to illustrate the role typically played by food standards, the situation is described first under EC law, in the UK as an EU member state, and then in the USA, with is totally different system of food legislation.

The EC

A number of food standards were set by so-called vertical directives (i.e., directives applying to a specific product sector) in the 1970s and 1980s; products such as jams, coffee and chicory extracts, fruit juices, cocoa and chocolate products, honey, preserved milks, sugars, caseins and caseinates were included. In order to reach agreement on some of these, it was necessary to allow certain national derogations or exemptions to be maintained, the most famous being the use of vegetable fat in chocolate (UK, Denmark, and Ireland at the time allowed vegetable fat to be used up to 5% of the total weight of the product in chocolate manufacture, whereas other member states were of the opinion that the use of vegetable fat rendered the product a 'chocolate substitute'). Some issues were left open to the member states to make a decision, for example, the addition of vitamins for nutritional purposes to fruit juices. These directives included provisions on additives and labeling as well as minimum levels of required ingredients.

Key to the change in emphasis in EC food law was the 'cassis de Dijon' case, which was a case brought under Article 30 of the EEC Treaty concerning the composition of a blackcurrant liqueur. Under this Article, no EC member state may refuse a product legally made in another member state provided its exclusion cannot be justified in terms of public health or provision of information. In the absence of compositional standards at EC level for the majority of foods manufactured the impact of this was highly significant, as the concept of mutual recognition of another EU member state's products began to be widely accepted. Certain of the member states began writing mutual recognition clauses into any new legislation on compositional standards to insure they did not fall foul of EC law.

Today EC directives on these foods are still in existence, although work has been continuing for some years to simplify them and bring them more into line with trends in modern food legislation. Examples of this are the removal of provisions on additives (now covered by horizontal (i.e., applying to all foods) additives directives) and the bringing of labeling requirements into line with those required by the general food-labeling directive, 2000/13/EEC. Revised directives have been agreed on coffee and chicory extracts and fruit juices and similar products; after many years' discussion, agreement has finally been reached on a revised cocoa and chocolate products directive. The key issues of vegetable fat and the naming of milk chocolate made with a lower percentage of cocoa solids have largely been resolved by requiring more detailed labeling, as well as restricting the types of vegetable fat that can be used.

However, variations on compositional standards are still being produced under EC law and are deemed necessary in order to insure the market in key agricultural sectors. Marketing standards have been produced for tuna, poultry, and a number of fruit and vegetables. Of other EC regulations, the one which is probably most familiar to many in the food industry is that on spreadable fats. This standard details the product descriptions that can be used for dairy, animal, and mixed-fat spreadable fat products with differing fat contents between 10 and 90%. Another important regulation is that on drinking milk, which lays down fat contents and descriptions for the different types of milk. As EC regulations, and not directives, such legislation is binding on the individual member states.

A new type of legislation has been introduced during the 1990s. Although not generic standards in the true sense, regulations on certificates of specific character for agricultural products and foodstuffs and on the protection of geographical indications and designations of origin for agricultural products and foodstuffs requires that names of the registered products may only be used when the product in question complies with approved specifications. Only products that are traditional or are produced in a traditional way will be able to be registered and protected by their specific character; mozzarella cheese is thus protected. PGIs/PDOs are intended to insure that a product's name is linked with its qualities and characteristics that may be due to environmental factors. Generic names (such as Cheddar cheese) may not be registered; however, a range of meat products (including Parma ham), olive oils, cheeses (including West Country farmhouse Cheddar cheese and Parmesan cheese) and other foods have been registered. Following a European Court ruling in March 1999, the name 'feta' has been deleted from the register, but the name will remain protected at a national level in the meantime.

The UK

Over the last 10 or so years, the number of food standards still in force under UK law has been significantly reduced. In some cases (for example, mayonnaise and soft drinks) revocation has been total; in other cases (for example, cheese, icecream) key definitions have been included in revised regulations on food labeling to insure that a minimum quality of product is still maintained when a product descriptor is used. To illustrate this, the minimum fat content of icecream and milk fat content of dairy icecream are still laid down, together with fat contents for named cheeses. Standards still remaining include those on bread and flour and meat and spreadable fish products, as well as on basic foods such as drinking milk, jams, honey, and other EC-derived standards.

Typically, a standard includes a definition of terms used in the regulations in order to aid interpretation; any exemptions that may apply; minimum compositional requirements for named foods; specific labeling requirements that are relevant for named products; penalties and enforcement, and any revocations of existing standards. An explanatory note finishes the text, summarizing the main points of the regulations. The requirements detailed in regulations may be backed up by codes of practice or guidelines, or by decisions reached by case law, which can aid in the interpretation of particular points contained within the regulations, or by advice issued to enforcement officers.

Are other EU member states similar to the UK? Certain have undergone a similar extensive deregulation exercise, most noticeably The Netherlands;

detailed guidelines are still used as indicators of consumer expectation in Germany. Spain still has in existence a large number of compositional standards across a wide range of food products, which contrasts with the Scandinavian countries, which tend to have fewer standards. But throughout the EU the concept of mutual recognition in terms of compositional requirements insures that national standards of the future can only be justified on the grounds of consumer information or public health/safety needs.

The USA

Establishment of Food Standards

Mandate for the production of standards for food comes from the Federal Food, Drug and Cosmetic Act (FFDCA), as amended. Section 401 of this Act, concerning definitions and standards for food, states that wherever, in the judgment of the Secretary of Health and Human Services, such action will promote honesty and fair dealing in the interests of consumers, regulations shall be made fixing and establishing for any food, under its common or usual name so far as practicable, a reasonable definition and standard of identity, a reasonable standard of quality, and/or reasonable standards of fill of a container.

Section 401 states that no definition and standard of identity, and no standard of quality, may be established for fresh or dried fruits, fresh or dried vegetables, or butter, with the exception that definitions and standards of identity may be established for avocados, cantaloupes, citrus fruits, and melons; however, these may relate only to maturity and the effects of freezing.

In prescribing any standard of fill of container, due consideration must be given to the natural shrinkage in storage and in transit of fresh natural food and the need for necessary packing and protective material. In the prescribing of any standard of quality for any canned fruit or canned vegetable, consideration must be given and due allowance made for the different characteristics of the varieties of that fruit or vegetable. When prescribing a definition and standard of identity for any food or class of food in which optional ingredients are permitted, the optional ingredients that must be named on the label will be designated, for the purpose of promoting honesty and fair dealing in the interests of consumers.

Agencies Responsible for Food Standards

The major federal agencies responsible for food standards in the USA include the Food and Drug Administration (FDA), the US Department of Agriculture (USDA) and the National Oceanic and Atmospheric Administration (NOAA), which is directed by the US Department of Commerce.

Food and Drug Administration The FDA, by delegation from the Secretary of the Department of Health and Human Services, regulates food other than red meat, poultry, and eggs. These standards of identity, established by the FDA for these products in Title 21 of the Code of Federal Regulations (21CFR), must be followed in every respect.

US Department of Agriculture The USDA is responsible for the control of red meat, poultry, and eggs and also for the quality assurance of other foods. Within the USDA, specific responsibilities are delegated to other agencies. Those involved in establishing standards for food include the following:

- Food Safety and Inspection Service (FSIS)
- Agricultural and Marketing Service (AMS)
- Federal Grain Inspection Service (FGIS)

The AMS is primarily responsible for standardizing the quality and grading of meat, poultry, dairy products, horticultural products, and certain other commodities. The grade standards established for these commodities are published in 7CFR. The majority of the functions of the AMS are voluntary, as are the grade standards. Federal law does not require that a food processor or distributor use the grade standards; however, they are widely used in wholesale and retail sale to enable the purchaser to select the required grade or level of quality.

The Meat and Poultry Inspection regulations, established by the FSIS and published in 9CFR, include compositional standards and labeling requirements for meat and poultry products. Definitions and standards of identity or composition established in 9CFR, 319 and 381 subpart P, for meat and poultry products, respectively, are mandatory and the producers covered must meet the provisions laid down to be legally manufactured and sold.

National Oceanic and Atmospheric Administration The National Marine Fisheries Service (NMFS) is one of the major services administered by the NOAA and is involved in the development of specifications and standards for the quality, grade, and identity for fishery products and the establishment of grade standards for processed fishery products.

Procedure for Establishing Standards

The procedure for establishing a food standard, referred to in section 401 of the FFDCA, is governed by the FDA in 21CFR. Any petition for a food standard

must show that the proposal, if adopted, would promote honesty and fair dealing in the interests of consumers. A petition includes an application or other document requesting the Commissioner of Food and Drugs to establish, amend, or revoke a regulation or order, or to take, or not, any other form of administrative action under the laws administered by the FDA.

Review of Codex Standards All food standards adopted by the Codex Alimentarius Commission are reviewed by the FDA. These Codex standards will be accepted without change, accepted with change, or not accepted. This review is accomplished in one of several ways, according to the procedures detailed in 21CFR. The review is officially published in the *Federal Register* for comments. After reviewing these comments the FDA makes a decision on the standard. A partial adoption of a standard makes it legally enforceable in the USA.

The FDA encourages all interested parties to confer with different interest groups such as consumers, industry, academia, and professional organizations in formulating petitions or comments.

Rule making Legal and scientific knowledge is used by the FDA to formulate regulations on compositional and processing requirements for food. There is considerable interaction between the FDA and the food industry during the establishment of regulations and the FDA is often dependent on industry for adequate scientific data on which to base regulatory decisions. Petitions from industry may initiate regulations. When federal agencies intend to issue a new regulation, or amend an existing one, they publish a notice of intent in the *Federal Register*; this stage is often omitted by the FDA. This rule-making notice gives the text of the proposed regulation, describes the issues involved, and invites comment within a specific period of time. If, after comments have been received, the change is favorably viewed, a proposed rule will be published, inviting comments, upon which the final rule is judged. When the final rule is published the background to the change and a summary of the comments is given.

Type and Scope of Standards

The FDA issues three types of standards, as prescribed by the FFDCA, namely standards of identity, quality, and fill of container, for a number of food products.

Standards of identity are the basic food standards issued by the FDA and their purpose is to protect consumers from being cheated by inferior products or confused by misleading labels. These standards establish or define the food product, e.g., by specifically stating the ingredients that the food may contain. They specify the correct name of the food and other mandatory label information, including the type and amount of certain vitamins that must be present in foods labeled 'enriched.' Required amounts of certain ingredients and maximum limits of others are specified. The standards are based on the assumption that food represented as, or purported to be, a food for which a standard of identity has been established must comply with the specification of the standard in every respect. However, temporary permits may be issued for interstate shipment of experimental packs of food varying from the requirements of definitions and standards of identity.

Certain of the standards of identity are supplemented by standards of quality and fill of container. Standards of quality establish specifications for quality requirements or factors for canned fruits and vegetables and are minimum standards. Standards of quality must not be confused with standards for grades that are published by the USDA for agricultural products and by the NMFS for fishery products. The standards for grades may classify the products from average to excellent in quality.

Standards of fill of the container designate the quantity in terms of solid or liquid components, or both for products such as canned fruits and vegetables, and certain fish products. If a food for which a standard of quality or fill has been established falls below that standard, it must bear a label statement in the prescribed size and style of type, stating that the product is substandard.

Comparison of certain food standards Table 1 is intended to illustrate some of the current differences in composition between certain food products in different countries. The product categories chosen (icecream and chocolate products) have been selected to show the contrast in these products. Similar product designations are used, but the composition is quite different. The tables include information concerning the major mandatory compositional requirements for the products covered; they are not intended to cover complete compositional data for these products.

The Future of Food Standards

The EC

New standards of composition in the traditional format are no longer relevant in the legislative arena of the twenty-first century. However, reference has been made to EC marketing standards that have an impact on all of the member states. The introduction of legislation requiring the declaration of the quantity

Table 1 Comparison of certain UK and USA standardized foods

	UK	*USA*
Icecream	Minimum 5% fat, minimum 2.5% milk protein	Minimum 1.6 lb total solids/gallon, weighs not less than 4.5 lb to the gallon
	Minimum 5% milk fat for 'dairy icecream'	Minimum 10% milk fat (milk solids, not fat content, may vary depending on the milk fat content)
	Food Labelling Regulations 1996, as amended	Title 21 CFR Part 135.110
	Milk fat only in dairy icecream, other than that present as a result of using egg, flavoring, emulsifier, stabilizer, or vegetable fat in nondairy icecream	Milk-derived fats only, except those natural components of flavoring or incidental ingredients
Chocolate (EC) Sweet chocolate (USA)	Minimum 35% total dry cocoa solids, including minimum 14% dry nonfat cocoa solids and minimum 18% cocoa butter	Minimum 15% chocolate liquor, less than 12% milk solids
	EC Directive 73/241/EEC, as amended	Title 21 CFR Part 163

CFR, Code of Federal Regulations.

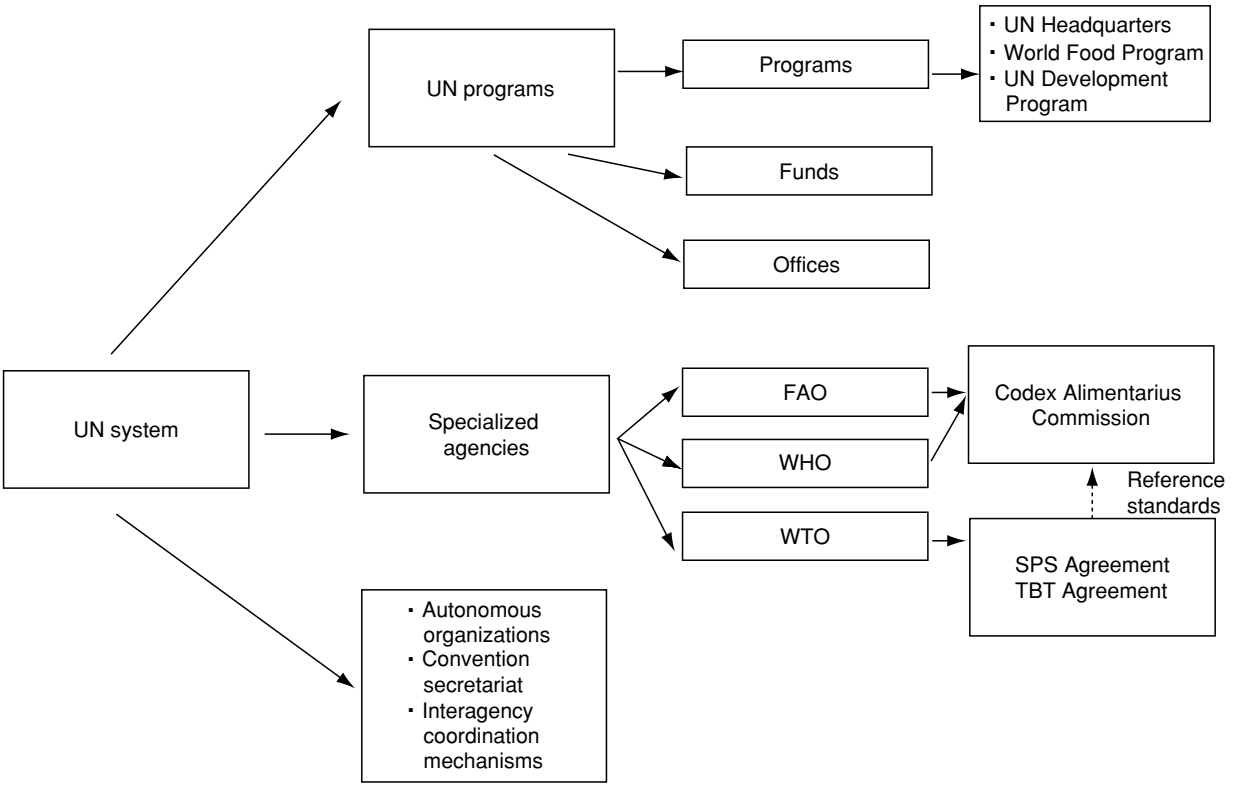

Figure 1 United Nations system of organizations. FAO, Food and Agriculture Organization; WHO, World Health Organization; WTO, World Trade Organization; SPS, Sanitary and Phytosanitary Measures; TBT, Technical Barriers to Trade.

of the characterizing ingredient(s) of foods (so-called QUID legislation) has made compositional standards irrelevant to a large degree; by requiring the content of the key ingredient to be declared on a label there is no need to set a minimum level as consumers can make their own choice. The QUID labeling requirement will largely determine the future format of revised legislation on meat products and spreadable fish products in the UK. However, it is clear that the revised EC standards on coffee, cocoa and chocolate, preserved milks, and other named foods will be here to stay for some time, particularly after the intensity of debate in reaching agreement.

USA

In October 1995, the FDA published its policy on the development and use of standards with respect to international harmonization of regulatory

requirements and guidelines. The FDA established that its main goals in standard-setting activities is to preserve and to enhance its public health mission, increase its efforts in harmonizing regulatory requirements with foreign governments, including setting new standards that better serve public health, and respond to laws and policies that encourage agencies to use international standards that provide the desired degree of protection.

The FDA policy lays down the criteria it uses for participating in standards development. For example, the expected standard would be required to stress product safety and effectiveness, thereby contributing to safe, effective, and high-quality products.

Codex Standards

In 1995, two agreements of the World Trade Organization gave a new significance to the work of the Codex Alimentarius Commission (CAC). These were the Agreement on the Application of the Sanitary and Phytosanitary Measures (SPS) and the Agreement on Technical Barriers to Trade (TBT). Both agreements formally recognized that international standards, guidelines, and recommendations, including the Codex Alimentarius, are reference points that can be used to facilitate international trade and resolve trade disputes in international law. The SPS agreement goes a step further and specifically identifies Codex texts as scientifically recognized and as a means of harmonization.

The adoption of Codex texts as scientifically justified standards for the SPS and TBT agreements is of immense significance (Figure 1). Codex texts are now an integral part of the legal framework within which international trade is being facilitated through harmonization. They can be used as benchmarks in international trade disputes and to evaluate the relevance of national measures. This formal recognition has awakened considerable interest in CAC activities and attendance at Codex meetings has increased markedly in recent years.

Until now, the individual commodity standards, which dealt with the quality aspects of foodstuffs, were the most important part of the Codex Committee's work. However, interest in these recipe standards is shifting to a more horizontal approach. It can already be seen that work from the so-called horizontal committees, such as labeling and additives, are attracting more interest and gaining in importance and profile. Moreover, topical issues, such as those concerning the use of biotechnology, are attracting the interest of both consumer groups and the food industry. The CAC is already examining consumer protection against health hazards and food safety in this key area. These provide an insight into the direction that the CAC activities are likely to take in the future.

Codex texts have formed the basis for food standards legislation in certain countries, for example the Middle Eastern countries. They have often been used as a guide in many other countries throughout the world. This role is still significant and, in the light of the SPS and TBT agreements, is likely to increase further as Codex texts are used as reference in international trade disputes.

See also: **European Union**: European Food Law Harmonization; **Legislation**: History; Additives; Contaminants and Adulterants; Codex; **World Health Organization**; **World Trade Organization**

Further Reading

Federal Register Volume 62 (129), 7 July 1997, 36243–36248. Washington, DC: Government Printing Office.

Navarro S and Wood R (2001) Codex Deciphered. Leatherhead: Leatherhead Publishing/LFRA.

Smith JS, Smith JA and Douek NL *EC Food Legislation*, 3rd edn. Leatherhead: Leatherhead Publishing/LFRA.

UK Food Labelling Regulations (1996) SI 1996 no. 1499, as amended, London: HMSO.

Vetter JL (1999) *Food Laws and Regulations* 41–66. Kansas, USA: American Institute of Baking.

Additives

P Berry Ottaway, Berry Ottaway & Associates Ltd, Hereford, UK

Introduction

The addition of chemical substances to food for technological purposes such as preservation or for organoleptic purposes such as color and flavor can be traced back over millennia. Salt was valued in early civilizations not only as a condiment but also for its ability to increase the stored life of meat and fish. There is evidence that the Egyptians used colorants in their foods nearly 4000 years ago.

The use of chemical additives to foods increased considerably during the twentieth century, partly due to the globalization of trade in food and partly to the increasing knowledge of food technology.

Many hundreds of substances have been researched for use as food additives for either technological or organoleptic purposes, often not without

controversy, and one of the challenges to international organizations such as the Codex Alimentarius Commission (Codex) is to try and achieve harmonization in the use of food additives throughout the world. Codex is an organization, set up under the auspices of the Food and Agriculture Organization (FAO) and the World Health Organization (WHO), which is charged with developing a food code to be used on a truly international basis.

Definition of a Food Additive

The definition of a food additive is, by necessity, closely linked to a definition of food. Unfortunately, there is no agreement between the food laws of a number of the developed countries on the actual definition of food.

In 1966, in the early days of the Codex, there was an attempt to define food, as follows:

> Food means any substance, whether processed, semi-processed or raw, which is intended for human consumption and includes drink, chewing gum, and any substance which has been used in the manufacture, preparation or treatment of food, but does not include cosmetics or tobacco or substances used only as drugs.

This definition can be found in many modified forms in the legislation of different countries. Even though the European Union (EU) has been developing its harmonized food legislation since the mid-1960s, it has not yet been able to obtain agreement between its member states on the precise definition of 'food' or 'foodstuff,' even though both terms are used extensively in EU legislation. It is likely that an agreed definition will not be dissimilar to that of Codex.

The Codex defines a food additive as:

> Any substance not normally consumed as a food by itself and not normally used as a typical ingredient of the food, whether or not it has nutritive value, the intentional addition of which to food for a technological (including organoleptic) purpose in the manufacture, processing, preparation, treatment, packing, packaging, transport or holding of such food results, or may reasonably be expected to result (directly or indirectly) in it or its by-products becoming a component of or otherwise affecting the characteristics of such food. The term does not include contaminants or substances added to food for maintaining or improving nutritional qualities.

In the legislation of the EU, the definition of a food additive is contained in European Council Directive 89/107/EEC. This definition is similar to the Codex definition and is given as:

> For the purpose of this directive 'Food Additive' means any substance not normally consumed as a food itself and not normally used as a characteristic ingredient of food whether or not it has nutritive value, the intentional addition of which to food for a technological purpose in the manufacture, processing, preparation, treatment, packaging, transport or storage of such food results, or may be reasonably expected to result, in its by-products becoming directly or indirectly a component of such foods.

The directive specifically excludes:

1. processing aids;
2. substances used in the protection of plants and plant products in conformity with Community rules relating to plant health;
3. flavorings for use in foodstuffs, falling within the scope of Council Directive 88/388/EEC;
4. substances added to foodstuffs as nutrients (e.g., minerals, trace minerals, or vitamins).

Extraction solvents are also not considered to be food additives in the EU and are subject to specific legislation on both their use and residual levels.

A processing aid is further defined in European law as:

> any substance not consumed as a food ingredient by itself, intentionally used in the processing of raw materials, foods or their ingredients, to fulfill a certain technological purpose during treatment or processing and which may result in the unintentional but technically unavoidable presence of residues of the substance or its derivatives in the final product, provided that these residues do not present any health risk and do not have any technological effects on the finished product.

In European law one of the key differences between a food additive and a processing aid is that the latter must not have any technological effect in the finished product. For example, residues of a mold-release agent for a confectionery product are unlikely to have a technological effect on the final product, whereas an antioxidant or preservative added to protect ingredients during processing could still exert a technological function if carried over into the finished product. The former would be considered a processing aid whilst the latter could be considered a technological additive.

In the USA, the Food Additives Amendment 1958 to the Federal Food, Drug and Cosmetics Act contains a more complex definition:

> The term food additive means any substance the intended use of which results or may reasonably be expected to result, directly or indirectly, in its becoming a component or otherwise affecting the characteristics of any food (including any substance intended for use in producing, manufacturing, packing, processing, preparing, treating, packaging, transporting, or holding food, and including any source of radiation intended for any such use), if such substance is not generally recognized,

among experts qualified by scientific training and experience to evaluate its safety, as having been adequately shown through scientific procedures (or, in the case of a substance used in food prior to January 1, 1958, through either scientific procedures or experience based on common use in food) to be safe under the condition of intended use; except that such a term does not include

1. a pesticide chemical in or on a raw agricultural commodity; or
2. a pesticide chemical to the extent that it is intended for use or is used in the production, storage, or transportation of any raw agricultural commodity; or
3. a color additive; or
4. any substance used in accordance with a sanction or approval granted prior to the enactment of this paragraph pursuant to this act, the Poultry Products Inspection Act (21 U.S.C. 451 and the following) or the Meat Inspection Act of March 4, 1907 (34 Stat. 1260), as amended and extended (21 U.S.C. 71 and the following).

This amendment categorized food chemicals as: (1) those generally recognized as safe (GRAS); (2) those with prior sanction; and (3) food additives. Pesticides on raw agricultural products and food color additives were excluded from the legal definition as they were covered by other legislation.

Joint FAO/WHO Expert Committee on Food Additives (JECFA)

Since it was established in 1955 the Joint FAO/WHO Expert Committee on Food Additives (JECFA) has played an international role in providing a review and opinion on food additives, particularly on safety in use.

JECFA was originally set up to consider the chemical, toxicological, and other aspects of contaminants and residues of veterinary drugs found in foods intended for human consumption.

The committee is composed of an international group of independent scientists appointed in their own right for their expertise in food additives, veterinary drug residues, and contaminants. The members are not appointed as government representatives. The committee is administered jointly by the FAO and the WHO. JECFA provides scientific advice directly to member countries of FAO and WHO and also to two committees of Codex; the Codex Committee on Food Additives and Contaminants (CCFAC) and the Codex Committee on Residues of Veterinary Drugs in Foods (CCVDF).

In relation to Codex, the Codex committees identify the additives, contaminants, and residues that should receive priority evaluation and refer them to JECFA for assessment before considering them for incorporation into Codex Standards.

The JECFA reports on their toxicological evaluations are published so that the information becomes widely available.

In its interaction with the Codex Committees, JECFA is independent of Codex and is able to provide impartial and independent advice without political influence.

With the advent of the World Trade Organization (WTO) in 1995, Codex has taken on a greater international significance as the WTO recognizes Codex as the preferred international organization for the arbitration and settlement of disputes related to food trade.

As a consequence of the WTO and also the involvement of Codex in setting international standards for foods and commodities, Codex now has a major influence on food additive legislation worldwide.

Approval of Additives

As part of the assessment of an additive, the JECFA examines the available toxicological data and chemical specifications of the additive and establishes an acceptable daily intake (ADI). The ADI is the amount that can be taken daily in a human diet for a lifetime without risk. Unless stated to the contrary, ADIs are expressed as mg kg^{-1} body weight, one of the exceptions being flour treatment agents that have been given as parts per million.

The ADI is normally derived from the 'no observed adverse effect' level (NOAEL) determined from long-term animal (*in vivo*) studies. The ADI is calculated by applying a safety or uncertainty factor, which is commonly 100, to the NOAEL obtained from the most sensitive test species. The 100-fold safety factor is based on the need to take into account both the differences in species and differences in toxicokinetics and toxicodynamics. Although most commonly used, the factor of 100 may be varied depending on the nature of the toxic effect and the availability of relevant toxicity data.

Approval for the use of an additive is given on the understanding that the additive will be kept under surveillance and reevaluated as necessary in light of changing conditions of use or new scientific evidence. The ADI value may be qualified in five different ways:

1. *Not limited*, which means that there is no explicit indication of an upper limit: this is assigned to substances of very low toxicity (this category has now been superseded by 'not specified').
2. *Not specified*, which means that on the basis of available data the total daily intake required to

achieve the desired additive effect does not represent a hazard to health. For this reason the establishment of an ADI is not deemed necessary. However, the usage of an additive in this category must conform to good manufacturing practice (GMP).

3. *Unconditional status* is allocated to those substances for which the biological data include favorable results from appropriate long- and short-term toxicological studies and/or biochemical and metabolic studies.
4. *Conditional* is allocated when the committee has considered that the quality of the available data is inadequate for an unconditional ADI and further work is required, or if there are other reasons such as those arising from dietary requirements. The reasons and the restrictions imposed on the use of the substance are stated in the evaluation.
5. *Temporary acceptance* means that there are insufficient data to establish whether or not the substance is toxic and further evidence must be submitted within a stated period of time. Details of the requirements are included in the evaluation.

After the JECFA has completed its assessment of an additive and issued its opinion, Codex is responsible for the formal approval or disapproval.

When Codex gives approval for the inclusion of the substance in the list of additives, in many cases the approval is limited, as far as possible, to specific foods for specific purposes at the lowest effective level of use. Any judgments are determined on the basis of the ADI of the additive and its probable intake from all dietary sources.

Classification of Additives

Food additives can perform a number of technological functions during food processing and storage and in a few cases the same substance may have more than one function.

A comprehensive, but not exhaustive, list of additive functions is given in the Codex Standard on Food Labeling (Table 1). This list illustrates the diversity of functions covered by additives and in most countries it forms the basis for the classification of approved additives.

Legislative Controls on Additives

Most countries, and particularly those which are members of Codex, have introduced legislation for the control of food additives. Whilst the details of the legislation differ from country to country, the basic requirements of only using approved additives tend to be consistent. There is, however, no complete consistency between the lists of different countries.

Table 1 Codex Alimentarius additive classes for labeling. Codex standard 1–1985 (revised 1–1991)

Acidity regulator	Flavor enhancer
Acids	Flour treatment agent
Anticaking agent	Foaming agent
Antifoaming agent	Gelling agent
Antioxidant	Glazing agent
Bulking agent	Humectant
Color	Preservative
Color retention agent	Propellant
Emulsifier	Raising agent
Emulsifying salt	Stabilizer
Firming agent	Sweetener
	Thickener
The following class titles may also be used for food additives falling into the respective classes:	
Flavor(s) and flavoring(s)	Modified starch(es)

Table 2 Categories of food additives in the European Union (Directive 89/107/EEC)

Color	Preservative
Antioxidant	Emulsifier
Emulsifying salt	Thickener
Gelling agent	Stabilizer
Flavor enhancer	Acid
Acidity regulator	Anticaking agent
Modified starch	Sweetener
Raising agent	Antifoaming agent
Glazing agent	Flour treatment agent
Firming agent	Humectant
Sequestrant	Enzyme
Bulking agent	Propellant gas and packaging gas

A significant group of countries now working to one of the most complex laws on food additives are those in the EU. A complete review and revision of the European legislation on food additives took place from the late 1980s to the mid-1990s.

Within the EU the framework legislation is found in Directive 89/107/EEC which lists 24 categories of food additives (Table 2). Whilst essentially similar, the EU list differs in detail from that of Codex (Table 1) in that the EU list includes sequestrants, antifoaming agents and enzymes, whilst the Codex list has color-retention agents that are not in the EU list. These categories are further defined in the specific directives, such as those on colors, sweeteners, and the directives on additives other than colors and sweeteners (Directive 95/2/EC).

Also detailed in the legislation (Directive 95/2/EC, Article 1,5) are substances or groups of substances not considered to be food additives under European food law (Table 2). This includes edible gelatin, caseinates, inulin, and chewing gum bases. Since the

middle of the 1990s, the detailed European legislation on food additives has centered on three directives:

- Directive 94/35/EC on sweeteners for use in foods (as amended);
- Directive 94/36/EC on colors for use in foods (as amended);
- Directive 95/2/EC on food additives other than colors and sweeteners (as amended).

The interpretation of all three directives has become complex, as in many cases the additives are not only restricted to defined categories of foods, but can also be subject to varying upper limits on their use depending upon the food category in which their use is intended. Thus a color, sweetener, or technological additive may appear in a list of permitted additives, but its use may be severely restricted to a small number of foods and its usage levels laid down as g or mg of the additive per kg or liter of the product as consumed. An example is given in Table 3. This table clearly illustrates the way the controls are imposed. The maximum levels of each additive are related to both technological need and the degree of likely exposure of the consumers to the additive in terms of the relative proportions of the food consumed as part of a diet. Thus, in case of propane-1,2 diol alginate, the additive is restricted to $100\,mg\,l^{-1}$ in beer, $300\,mg\,l^{-1}$ in nonalcoholic flavored drinks, but allowed up to $10\,g\,l^{-1}$ in an emulsified liqueur.

The introduction of these directives consolidated and harmonized the legislation on food additives across the EU and all food and drink products on

Table 3 Example of European Union food additive legislation showing food categories and maximum levels

EC no.	*Name*	*Food*	*Maximum level*
E355	Adipic acid	Fillings and toppings for fine bakery wares	$2\,g\,kg^{-1}$
E356	Sodium adipate	Dry powdered dessert mixes	$1\,g\,kg^{-1}$
E357	Potassium adipate	Gel-like desserts	$6\,g\,kg^{-1}$
		Fruit-flavored desserts	$1\,g\,kg^{-1}$
		Powders for home preparation of drinks	$10\,g\,l^{-1}$
			Expressed as adipic acid
E363	Succinic acid	Desserts	$6\,g\,kg^{-1}$
		Soups and broths	$5\,g\,kg^{-1}$
		Powders for home preparation of drinks	$3\,g\,l^{-1}$
E385	Calcium disodium ethylene diamine tetraacetate (calcium disodium EDTA)	Emulsified sauces	$75\,mg\,kg^{-1}$
		Canned and bottled pulses, legumes, mushrooms, and artichokes	$250\,mg\,kg^{-1}$
		Canned and bottled crustaceans and molluscs	$75\,mg\,kg^{-1}$
		Canned and bottled fish	$75\,mg\,kg^{-1}$
		Minarine	$100\,mg\,kg^{-1}$
		Frozen and deep-frozen crustaceans	$75\,mg\,kg^{-1}$
E405	Propane-1,2-diol alginate	Fat emulsions	$3\,g\,kg^{-1}$
		Fine bakery wares	$2\,g\,kg^{-1}$
		Fillings, toppings, and coatings for fine bakery wares and desserts	$5\,g\,kg^{-1}$
		Sugar confectionery	$1.5\,g\,kg^{-1}$
		Water-based edible ices	$3\,g\,kg^{-1}$
		Cereal- and potato-based snacks	$3\,g\,kg^{-1}$
		Sauces	$8\,g\,kg^{-1}$
		Beer	$100\,mg\,l^{-1}$
		Chewing gum	$5\,g\,kg^{-1}$
		Fruit and vegetable preparations	$5\,g\,kg^{-1}$
		Nonalcoholic flavored drinks	$300\,mg\,l^{-1}$
		Emulsified liqueur	$10\,g\,l^{-1}$
		Dietetic foods intended for special medical purposes-dietetic formulae for weight control intended to replace total daily food intake or an individual meal	$1.2\,g\,kg^{-1}$
		Dietary food supplements	$1\,g\,kg^{-1}$
E416	Karaya gum	Cereal- and potato-based snacks	$5\,g\,kg^{-1}$
		Nut coatings	$10\,g\,kg^{-1}$
		Fillings, toppings, and coatings for fine bakery wares	$5\,g\,kg^{-1}$
		Desserts	$6\,g\,kg^{-1}$
		Emulsified sauces	$10\,g\,kg^{-1}$
		Egg-based liqueurs	$10\,g\,l^{-1}$
		Dietary food supplements	*quantum satis*
		Chewing gum	$5\,g\,kg^{-1}$

sale in the EU must now comply with the requirements of this legislation. The introduction of the new laws between June 1996 and March 1997 resulted in a number of products having to be reformulated as the new requirements in some cases differed significantly from the internal legislation of some of the member states of the EU. In Germany, for example, the options open to manufacturers for the selection and use of a number of additives increased, whereas in the UK, manufacturers were confronted with more restrictions, particularly in the use of colors and sweeteners.

During the development of the directives on additives, there was considerable discussion about which foods should be allowed to contain which additives. The categories of foods and drinks were modified a number of times during the debates, with the end result that there was no pattern to the product categories finally agreed. Thus, in the section dealing with food preservatives, there are general groupings such as 'nonalcoholic flavored drinks' or 'dried fruit,' and very product-specific rules such as those for gnocchi or polenta.

All applications for the extension of the conditions of use to allow other categories of foods to contain additives, or applications for new additives, are referred by the European Commission to the Scientific Committee on Food of the EU (SCF). The SCF has published guidelines on the data required for applications for new additives and this was updated in July 2001. One of the conditions is that a new additive must exhibit a technological function or a significant advantage (i.e., considerably improved function or organoleptic properties) over those already approved for the technological purpose. Extensive toxicity studies are required to establish the safety of the additive. Since the EU directives on colors, sweeteners, and miscellaneous technological additives were adopted in the mid-1990s, there have been relatively few additions to the lists.

Labeling of Food Additives

The legal requirements for the labeling of food additives can vary from country to country.

The Codex standard on the Labeling of Prepackaged Foods lays down guidelines for the labeling of additives. These guidelines are closely followed in the EU directive on Food Labeling (2000/13/EC). Codex requires that for food additives falling into the respective classes (Table 1), the class titles given in the table should be used together with the specific or recognized name or a recognized numerical identification of the additive. The EU already has a recognized numerical classification of food additives commonly described as the 'E number' system. Thus, in the EU sulfur dioxide can be labeled as either: preservative: sulfur dioxide or preservative: E220.

Codex also requires that a food additive that is carried over into a food in a significant quantity or in an amount sufficient to perform a technological function in a food as a result of the use of raw materials or ingredients containing the additive shall also be included in the list of ingredients on the label. This means, for example, that if a significant ingredient in a food product contains an antioxidant or preservative, which is in sufficient quantity in the final product to continue to play its additive role, it must be added to the ingredients list. Excluded from this requirement are processing aids that do not have a technological function in the final product.

In the USA there are a number of specific rules for the labeling of additives. Substances that perform a functional purpose (i.e., technological additives) may be declared in the ingredients list using a collective or generic term followed by the function in parentheses. Alternatively, the function can be listed followed by the substance(s) performing that function, e.g., soy lecithin (emulsifier) or emulsifier (soy lecithin).

When a chemical preservative or antioxidant is used, American law requires both the common or usual name of the preservative and the function, using terms such as 'preservative,' 'to retard spoilage,' or 'to preserve (or retain) freshness.'

There is also a specific requirement in the USA for the listing of colors. Food and Drug Administration (FDA)-certified colors and their lakes must be listed specifically by their complete or abbreviated name for all food products except butter, cheese, and icecream.

Mixtures of colorings may be listed as 'artificial color' followed by a sublisting of the certified colors in parentheses, e.g., artificial color (FD and C yellow FD and C red 40).

Other additives used for coloring (termed noncertified), such as caramel, may be listed by their specific, usual or common name, e.g., caramel coloring or colored with caramel.

The differences in approach between the USA and EU can lead to complications with transatlantic trade in food products.

For countries other than the USA and those in the EU, there are a number of instances where there are requirements for the labeling of additives which are specific to that country. For many countries, however, the general principles of the Codex standard on labeling apply.

The Future

The complexities of international control on food additives together with the high costs of the toxicological

studies have been instrumental in reducing the number of food additives coming on to the market.

The rate of food additive developments appears to have peaked in the mid-1970s and by the early 1980s there were consumer reactions and food scares related to additives, particularly intolerances to the azo food colors, preservatives, and antioxidants.

Future developments will, in general, be focusing on improvements to the efficiency or functionality of additives and on reducing the overall intake of those found to have relatively low ADIs.

It is likely to be some years before there is truly international harmonization on the additives that are permitted and their maximum levels of use in foods.

See also: **Food Additives**: Safety; **Food and Agriculture Organization of the United Nations**; **Food Safety**; **Legislation**: Codex

Further Reading

Codex Alimentarius Commission. Available at: http://www.fao.org/es*/esn/codex.

Furia TE (ed.) *Handbook of Food Additives*. Florida: CRS Press.

Joint FAO/WHO Expert Committee on Food Additives (JECFA) (2001) *Compendium of Food Additives Specifications* (FNP 52 plus addenda 1–7). Available at: http://www.codexalimentarius.net/CONTACT/Htframe.htm.

Scientific Committee on Food of the European Commission (2001) *Guidance on Submissions for Food Additive Evaluations*. SCF/CS/ADD/GEN/26 Final of 12 July 2001. Brussels, Belgium: Scientific Committee on Food of the European Commission.

Watson D (ed.) *Food Chemical Safety*, vol. 2 – *Additives*. Cambridge: Woodhead Publishing.

Contaminants and Adulterants

C M Stansfield, Leatherhead Food International, Leatherhead, Surrey, UK

Introduction

Despite the fact that we now have a more tightly controlled, and therefore safer, food supply than ever before, there is a noticeable increase in consumer awareness and interest in food quality and safety, and one cannot exclude from this context adulterants, contaminants, and residues. This article outlines the way in which such areas of major public concern are regulated and controlled, including pesticide and heavy metal residues, microbiological contamination, foreign substances, and others.

General Overview

In addition to specific legislative acts on food safety issues, all modern food law embodies the fundamental concept that food should be of the nature, quality, and substance demanded by the consumer. This implies that, even where a country lacks specific legislation on a particular area, it is able to take action against adulteration or contamination of food.

To achieve the objective of protecting the consumer, each country has its own way of regulating the occurrence of contaminants and adulterants in foodstuffs. Some countries control food contamination at commodity level, i.e., levels for heavy metals and microbiological criteria or other contaminants are laid down in food standards of composition. In contrast, other countries insure that food quality and safety are achieved throughout the food process, as opposed to being concerned with the examination and control of the final product only.

In recent years, new technological developments, consumer awareness, and demand for a healtheir diet have resulted in an increase in the amount of new legislation, either already in force or in circulation as proposals, and much of this is related to food safety. In addition to these factors, the increase in new legislation in Europe is mainly a result of the need to harmonize the area of food safety at European Union (EU) level.

Food Adulteration

In Europe, an adulterant is a substance added to food in order to defraud the consumer, whereas in the USA, food containing any substance that is not supposed to be in the food is said to be adulterated. Deliberate adulterants are frequently inferior, cheaper ingredients that are used for economic advantage, e.g., water in milk, and are not dangerous, but constitute fraud.

The addition of mineral oil to cooking oil in Spain in 1979 was a dangerous adulterant that caused injury to some hundreds of people, and this incident incited the EU to set up a rapid alert system between the control authorities of member states. The task of such alert systems is to insure that affected food is identified and removed from the market and to return the trading system to normal as soon as possible. Some countries have powers to seize and destroy adulterated food but others operate by warning the trade and the public. Since lack of immediate action by the trade could not only lay them open to prosecution but also lead them to disrepute, it is extremely unusual for them not to cooperate with the authorities.

The presence of foreign bodies such as stone, glass, metal, rubber, wood, human hair, and parts of

clothing usually arise from lack of care, and are dealt with under general legal provisions on food. Unfortunately, there have been cases of deliberate addition of foreign bodies to food, usually by tampering with food on display, with intent to inflict economic damage, either for personal gain or as a threat against the manufacturer or trade, and such action is dealt with under criminal law. Food alert systems have again proved their worth in limiting risk and economic damage.

Contaminants

Contaminants are defined by the World Health Organization/Food and Agriculture Organization (WHO/FAO) Codex Alimentarius as follows:

> 'Contaminant' means any substance not intentionally added to food, which is present in such food as a result of the production (including operations carried out in crop husbandry, animal husbandry and veterinary medicine), manufacture, processing, preparation, treatment, packing, packaging, transport or holding of such food or as a result of environmental contamination. The term does not include insect fragments, rodent hairs and other extraneous matter.

It is interesting to note that the consumer would certainly regard the first two exclusions as contaminants. Legislation on contaminants has been developed in response to specific concerns and is therefore largely incoherent and in many cases outdated. Substances that accumulate in the environment, polluting air, farmland, and sea, eventually become present in foodstuffs. Specifying levels for the intake of environmental contaminants in foods must create a balance between encouraging the reduction of their presence through good manufacturing practices and consideration of conditions beyond the control of the producer that may determine the achievable level; for example, aflatoxin levels in cereal-based foods reflect climatic conditions during maturation and storage. However, within the EU, harmonization of certain areas is a continuing process. The Codex definition of contaminant has been adopted by the EU in an overriding framework regulation, Council Regulation (EEC) no. 315/93, which lays down a general prohibition on the marketing of foods containing contaminants in unacceptable amounts. In this way, levels are controlled through good manufacturing practice; however, there is also provision for the establishment of maximum tolerances, analytical methods and detection limits and references to sampling. There are separate committees that consider industrial contaminants, such as heavy metals and dioxins, and agricultural contaminants, such as nitrates, mycotoxins, and inherent plant toxins. This has resulted in the setting of maximum levels for certain contaminants in foodstuffs. However, existing limits on contaminants in each member state apply where no level is laid down by the EU.

Heavy Metals

Heavy metals represent an important group of inorganic contaminants that may be present in food products as a result of (1) their natural occurrence in food products in the raw materials used; (2) contamination of raw materials; or (3) undesirable transfer during the processes of manufacturing and packaging. The heavy metals most commonly covered by restrictive legislation are arsenic, copper, lead, mercury, zinc, and tin.

Regulations on heavy metals are frequently product-specific in nature, owing to susceptibility of certain foodstuffs to contamination by particular metals. For example, maximum levels for mercury in fish and fish products are often laid down because these products have been identified as principal sources of this element in the human diet. Levels of 0.5 p.p.m., or 1 p.p.m., depending on fish species, have been established by Codex Alimentarius and the EU. Similarly, levels of tin are of primary concern in canned foods, for obvious reasons. Tolerances for tin in canned foods normally fall in the range of 100–250 mg kg^{-1} (Table 1).

Most commonly, legislation on heavy metals is covered by regulations specifying maximum tolerances for named metals in a range of foodstuffs when the presence of a substance reaches a level known to present significant risk; this is the case for Canada, the UK, Australia, and South Africa. In some cases, a general tolerance is also laid down for foodstuffs not specifically listed, as in the UK, and Poland.

The limits for mercury in fish, and for lead and cadmium in a range of food categories, are the only heavy-metal limits set, at present, by the EU.

Some member states have approaches that differ from the one outlined above. Certain countries, including Spain, Austria, and Greece, lay down provisions in individual standards of composition, as does Israel.

Recommendations of the Codex Alimentarius Commission are often considered in the formulation of legislation on heavy metals but, as Table 2

Table 1 Permitted levels of tin (p.p.m.) in canned solid foods

UK	*Finland*	*Norway*	*Hungary*	*Canada*	*Malaysia*
200	150	250	200	250	250

Data from Legislation Unit (2000, updated March 2001) *Contaminants in Foodstuffs – An International Review of Maximum Limits Part 2*. Leatherhead, UK: Leatherhead Food RA.

demonstrates for wine, there may still be variations in levels set worldwide.

The situation for lead in wine has been rectified partly by action by the EU and the US Food and Drug Administration (FDA), which have banned the use of lead capsules on wine bottles, the major source of lead in wine. Amongst efforts to reduce their presence in other foods has been a prohibition on or general phasing out of the use of lead solder in food cans.

Heavy metals may also be indigenous in food, and their presence in water can derive partly from the rocks of the aquifer. In this case they are referred to as undesirable substances and are regulated, in the EU, the USA, and elsewhere, by legal provisions on the quality of drinking water and bottled mineral water, for which a Codex standard also exists.

Polychlorinated Biphenyls and Dioxin

Various inert organic compounds with industrial applications may pose a threat to food safety. Polychlorinated biphenyls (PCBs) are examples of such compounds; their presence is now widespread and uncontrolled in the environment. These have a tendency to accumulate in animal feeds, particularly those of animal and marine origin, and are subsequently introduced into food.

At EU level, there are no limits at present for PCBs in foods and, in general, the regulations of most countries do not make specific provisions for the control of PCBs in foodstuffs. However, certain countries (e.g., Germany and the USA) have set maximum tolerances for PCBs in a range of foodstuffs. In some other countries, The Netherlands and France included, maximum tolerances are laid down for fish and fish products, which are particularly susceptible to PCB contamination (Table 3). Concern over PCB contamination led the UK and the USA to discontinue their manufacture in 1977. In some countries, the use of PCBs in industrial applications is restricted where they may constitute a threat to food safety. For example, in Denmark, the use of PCBs in heating installations for foodstuffs is prohibited.

Dioxin, a contaminant from chemical manufacture, has caused concern in recent years, as it can accumulate in the fatty tissue of animals and humans. In 1999 contamination from dioxin resulted in widespread infection of poultry and eggs in various European countries. This was believed to have originated from an oils- and fats-processing company in Belgium, which supplied feed manufacturers. Protective measures were established by the EU, and in the longer term they are evaluating the risk of dioxins and PCBs. 'It has emerged that more than 90% of human dioxin exposure derives from foodstuffs, with a large portion of this derived from foodstuffs of animal origin. Council Regulation (EC) No. 2375/2001 lays down maximum limits for dioxins in products from meat, milk, fish and oils and fats. Levels for the types of PCBs that exhibit toxicological properties to dioxins have not been set due to limited data available on their presence. The European Commission intends to review situation by the end of 2004.'

Nitrates

Concern over the harmful properties developed from the reduction of nitrate to nitrite has led the EU to set levels for nitrates in water, and for spinach and lettuce, depending on the season in which they are harvested. Currently, levels for fresh spinach range from 2500 to 3000 mg kg^{-1}, for processed spinach 2000 mg kg^{-1}, and a range of 2500–4500 mg kg^{-1} is set for lettuce. It has been estimated that, in many European countries, vegetables contribute most of the nitrate exposure, and some member states, such

Table 2 Permitted levels of lead (p.p.m.) in certain food commodities

Commodity	*Codex*	*EU*	*Sweden*	*Canada*	*Malaysia*	*Japan*	*Brazil*
Wine	0.2	0.2	0.2	0.2	0.5	–	0.5
Soft drinks	–	–	0.05	0.2	0.2	Not detectable	0.2

Data from Legislation Unit (1999, updated) *Contaminants in Foodstuffs – An International Review of Maximum Limits Part 1*. Leatherhead, UK: Leatherhead Food RA. Legislation Unit (2000, updated March 2001) *Contaminants in Foodstuffs – An International Review of Maximum Limits Part 2*. Leatherhead, UK: Leatherhead Food RA.

Table 3 Maximum tolerances for polychlorinated biphenyls (PCBs) in fresh-water fish

	France	*Germany*	*Netherlands*	*USA*	*Japan*
PCB content in fresh-water fish (mg kg^{-1})	2.0	0.2 or 0.3[a]	0.04–0.12[a]	2[b]	0.5–3.0

[a]Levels set for individual PCBs.
[b]Temporary tolerance.
Data from Legislation Unit (2000, updated March 2001) *Contaminants in Foodstuffs – An International Review of Maximum Limits Part 2*. Leatherhead, UK: Leatherhead Food RA.

as Austria and The Netherlands, have introduced upper limits for other types of vegetables. In Austria, guideline levels are laid down for a range of other vegetables, including carrots, cabbage, and parsley. Switzerland has levels for nitrates in cheese and infant foods as well as vegetables. Outside Europe, Argentina sets levels for nitrates in various milk products.

Radioactivity

Natural radioactivity occurs in all foods, arising from the radioactive isotopes carbon-14 and potassium-40. The specific activity of carbon-14 is relatively low; however, potassium-40 gives rise to specific activities of 40–50 Bq kg^{-1} in cows' milk, 400–500 Bq kg^{-1} in milk powder, 600–800 Bq kg^{-1} in concentrated fruit juice, and over 1000 Bq kg^{-1} in instant coffee. There is no legislation governing this radioactivity since it cannot be eliminated, and consumption of such foods does not significantly alter the potassium content, and hence the potassium radioactivity of the body.

Human exposure to radiation is laid down in law and national measures are usually based on recommendations of the International Committee for Radiological Protection (ICRP). Drinking water contributes a very small radiation dose to total population exposure and is largely due to naturally occurring radionuclides in the uranium and thorium decay series. WHO has established guideline levels for radioactivity in drinking water, based on ICRP recommendations and available data on risk exposure. Some countries lay down limits for radionuclides in bottled waters, for example in the USA, Sweden, Israel, and Switzerland.

Before the accident at Chernobyl nuclear power station in 1986, very few countries had radioactivity limits for food since the policy was, and still is, to control emission and waste at source, with local incidents being handled by monitoring of food and population. Contamination of food on a continental scale arising from Chernobyl required immediate action and resulted in the EU setting limits for radioactive contamination of imports from outside its member states. Since then, contamination levels have decreased to a negligible risk of health, and later rules exclude a number of products from the scope of the original regulation. There is a maximum tolerance of 370 Bq kg^{-1} for combined cesium-134 and cesium-137 in milk, milk products, and special infant foods, and a single level of 600 Bq kg^{-1} for other food products concerned. In 1999, the EU established procedures for carrying out checks on the radiocesium content of such foods. A number of countries followed the EU post-Chernobyl regulation. Infant foods, milk, and milk products, and animal foods are among the most widely regulated commodities in this regard. The EU has also established a regulation setting maximum permitted levels for radioactive contamination of foodstuffs, which could be applied in the event of a future nuclear accident or other radiological emergency potentially involving radionuclides other than cesium.

Mycotoxins

The concern over the carcinogenic hazard of mycotoxins has also increased in recent years. Mycotoxin contamination is most often related to the susceptibility of certain crops, especially grains and nuts, during storage, to the attack by toxin-producing fungi, in particular *Aspergillus flavus*, which produces a group of highly toxic metabolites called aflatoxins.

Specific mycotoxin legislation was developed in several countries as a result of the discovery of aflatoxin in the early 1960s; it was later extended to other mycotoxins, including patulin, ergot, ochratoxin A, and pomphsin. Currently, levels for some mycotoxins are regulated at EU level, namely aflatoxins B_1, B_2, G_1, G_2 in nuts, dried fruit, cereal products, and spices and a level of aflatoxin M_1 in milk, and limits for ochratoxin A in cereals and vine fruit. Limits for mycotoxins in infant foods and patulin in fruit juices have been proposed, and further categories for ochratoxin A are under review at EU level. Certain countries may control the presence of aflatoxins in other foods, as is the case in Germany, where limits are laid down for enzyme preparations, dietetic foods for babies and infants, milk, and other foodstuffs. Regulations on aflatoxins in non-EU countries, where they exist, usually determine levels for peanuts and peanut products and milk, or lay down a general level for all foodstuffs. Table 4 gives examples of maximum levels of aflatoxins in foodstuffs, especially nuts and nut products, in major countries worldwide.

Inherent Toxic Substances

If present in excessive amounts, certain inherent toxins pose a threat to human health. Whilst such toxins are not widely regulated, some countries have laid down product-specific tolerances in certain instances. For example, Finland has levels for solanidine glycosides in potatoes and for indole in crustaceans; however, in general, natural toxins can be controlled through regulations that establish species and strains that are for food use.

Levels for shellfish toxins have been established by a number of countries, including Japan and the USA, as well as in the EU. The adverse health implications associated with erucic acid content in some rapeseed

Table 4 Examples of maximum levels of aflatoxins in nuts and nut products, in major countries

Country	*Food*	*Aflatoxin*	*Maximum level ($\mu g\,kg^{-1}$)*
European Union member states	Groundnuts, nuts, and dried fruit and their processed products, intended for direct human consumption or as an ingredient in foodstuffs	$B_1 + B_2 + G_1 + G_2$	4[a]
		B_1	2[a]
	Groundnuts to be subjected to sorting, or other physical treatment, before human consumption or use as an ingredient in foodstuffs	$B_1 + B_2 + G_1 + G_2$	15[a]
		B_1	8[a]
	Nuts and dried fruit to be subjected to sorting, or other physical treatment, before human consumption or use as an ingredient in foodstuffs	$B_1 + B_2 + G_1 + G_2$	10[a]
		B_1	5[a]
USA	Peanut and peanut products, Brazil nuts and pistachio nuts, other foodstuffs (except milk)	Aflatoxins	20
Australia	Peanut and tree nuts and the nut proportion of products containing nuts	Aflatoxins	15
Japan	Peanut and peanut products, pistachio nut, almond, Brazil nut, cashew nut, hazel nut, macedemia nut, walnut, and giant corn	B_1	10
South Africa	Grains, nuts, and their products and all other foodstuffs	Aflatoxins	10 of which M_1 is maximum 5

[a]Maximum limits apply to the edible part of groundnuts, nuts, and dried fruits. If nuts 'in shell' are analyzed, it is assumed when calculating the aflatoxin content that all contamination is on the edible part.
Data from Legislation Unit (2000, updated March) *Contaminants in Foodstuffs – An International Review of Maximum Limits Part 2*. Leatherhead, UK: Leatherhead Food RA.

and mustard oils led the EU to set maximum levels for erucic acid in oils and fats intended for human consumption and for foodstuffs containing the same. EU regulations on flavorings include a list of maximum limits for certain potentially toxic substances, such as thujone and coumarin, which may be present in a foodstuff as a result of the use of flavorings. 3-monochloropropane-1, 2-diol is created during food processing under certain conditions. At EU level maximum levels have been set for its presence in hydrolyzed vegetable protein and soy sauce.

Microbiological Contamination

The contamination of foodstuffs by pathogenic and spoilage microorganisms has always been one of the major concerns of consumers and enforcement authorities, owing to the great risk that this poses to human health.

Countries may differ in the way in which they approach this problem; however, most governments have the single philosophy of seeking to insure protection of public health. This is usually achieved by provisions of food acts that exist in most countries to insure food safety and quality by making it an offence to sell food that is injurious to health and does not conform with the quality demanded by the purchaser. This means that in most countries it is an offence to sell foodstuffs contaminated with pathogenic organisms (*Salmonella, Staphylococcus aureus*, sulfite- reducing *Clostridium*, *Listeria monocytogenes*, and others) or spoilage microflora. Through the enforcement of food acts, powers may be given to the relevant authorities to impose emergency control measures in order to prevent the circulation and sale of contaminated food at any time should they consider that the public health is at risk. The UK Food Safety Act 1990 illustrates this clearly in that powers are given to ministers and other relevant authorities to act immediately upon cases that endanger the public health, e.g., to issue emergency orders, to close down food premises, etc.

In addition, microbiological criteria may be set which can be achieved throughout food processing by means of the application of good manufacturing practice. Microbiological criteria are generally expressed as three main categories:

1. Microbiological standards: these are applicable to the final product and have legal force; they are mandatory.
2. Microbiological guidelines: these are guides intended to help manufacturers to insure good hygiene practice; although guidelines are optional, it is strongly recommended that they are followed in the absence of microbiological standards.
3. Microbiological end-product specifications: these are generally contractual agreements between manufacturers and purchasers to check that the goods concerned are of the quality required.

Where a microbiological standard exists, sampling and analytical methods are required in order to monitor, in qualitative and quantitative terms, the criteria laid down.

Microbiological Criteria

The development and application of microbiological criteria for foods are carried out by international bodies and national authorities, including the WHO, the International Commission on Microbiological Specifications for Foodstuffs (ICMSF – a voluntary specific body, which provides basic information on the microbiology of foods based on extensive studies) and the EU.

Certain countries place more emphasis on the microbiological standard for the final product; in other words, microbiological criteria are laid down in food standards of composition. Examples include Spain, The Netherlands, Australia, Sweden, and Argentina, amongst others. In contrast, other countries follow a more systematic approach, where food quality and safety are achieved by the identification of hazard and assessment of risk throughout the processing, from the raw material stage to the final product. This is usually carried out by means of the Hazard Analysis Critical Control Point (HACCP) system or equivalent process, which is used in conjunction with good manufacturing process. The UK is an example of the latter trend, where there are very few specific microbiological standards; instead, the approach to food hygiene is that of prevention backed up by codes of practice and microbiological guidelines. In the USA, HACCP is mandatory in meat plants, seafood and juices used for beverages. Good manufacturing process regulations are laid down for foodstuffs in general, dietary supplements, low-acid canned foods, and acidified foods, and there are pathogen-reduction performance standards for *Salmonella* in slaughter plants and raw ground products.

The EU general hygiene directive was established in 1993 with the aim of insuring confidence in the standard of safety of all foods in circulation, particularly in respect of hygiene, through all stages from production to sale to the final consumer. While this does not establish criteria, the principles of HACCP are required and the directive makes provisions for future criteria concerning microbiological criteria and temperature control, requirements for food business operators to insure appropriate training for food handlers, and general hygiene requirements for premises. EU microbiological criteria are laid down for mineral water, milk and milk products, mincemeat, and meat preparations. The European Commission is revising its food hygiene rules to streamline and harmonize existing requirements. The new rules would apply to all food and all food operators throughout the food chain, and requires HACCP systems to be implemented by all food business operators.

Residues

Residues of pesticides and veterinary pharmaceuticals in food are not legally regarded as contaminants but are permitted to safe levels defined at the time that each product is authorized.

Pesticides

The use of pesticides in agriculture has contributed to the worldwide increase in food production. As well as improving the efficiency or productivity of agricultural practices, their use has also raised concern over their side-effects on animal and human health.

In most countries the use of pesticides is strictly controlled by regulatory bodies, whose responsibility is to insure that their use is safe and that any residues present in foodstuffs do not constitute any risk to public health. The WHO/FAO Joint Meeting on Pesticides Residues (JMPR) plays a major role at international level in assessing the safety of pesticides and by setting standards, and its recommendations and rules are taken into account by most countries.

Legislation on pesticides at EU level is extensive and the increasing tendency to use pesticides on certain crops, and the continuous formulation and commercialization of new substances for pesticide use, have resulted in a frequent revision of their safety and a subsequent need to review the legal provisions. Existing directives set mandatory maximum limits for pesticide residues in and on fruits and vegetables, cereals, and foods of animal origin, dried processed and composite products, infant formulae, and baby foods. Another directive prohibits the placing on the market and use of plant protection products containing certain active substances, including ethylene oxide. Methods of analysis for the detection and monitoring of fixed maximum limits are required to be established by these directives.

In comparison with the EU, non-EU countries may have a similar approach to the way in which they control residues of pesticides in foodstuffs, in that their legislation is based on lists laying down maximum levels for certain pesticide residues in and on specific food commodities. For example, in the USA, tolerances are set for a wide range of foods, processed foods, and animal feeds. Changes in the overall framework of their regulation mean that all pesticide tolerances must be reevaluated by 2007, establishing safe tolerance levels for groups of pesticides with similar mechanisms.

The overall trend observed in legislation on pesticides is that they are drawn up on a very restrictive basis; only those pesticides in the lists may be used and maximum levels should not exceed those laid down, unless special authorization is granted.

Veterinary Medicines

'Veterinary medicines' is a generic term embracing all substances that have an effect on the functioning of the organisms of the animal, whether administered by injection or through animal feed. Concerns over human health relate to possible immunological effects of repeated exposures to antibiotic residues, the possible development of antibiotic-resistant strains of pathogens, and the possible adverse effect of animal hormones on humans. The trend observed across major countries worldwide is for the prohibition or restriction of residues of veterinary medicines in foodstuffs. This may be made under the power of food acts, general health regulations, or under specific food commodity standards. For example, the EU, USA, Canada, and South Africa have regulations setting maximum residue levels of antibiotics for a wide range of foodstuffs of animal origin.

In addition to the indirect presence of antibiotic residues in foodstuffs, certain substances having an antibiotic effect may be intentionally added to foods as preservatives. Examples of the most widely regulated and used are natamycin and nisin, which are permitted in processed cheese in a great number of countries worldwide.

The concern over use of growth promoters in the treatment of animals intended for human consumption results from possible carcinogenic and teratogenic effects. There is a great deal of legislation in the EU in this regard, including directives on the following:

1. The trade in animals treated with certain substances having hormonal action and their meat.
2. Methods to detect residues of substances having a hormonal and thyrostatic action.
3. Prohibiting the use in livestock farming of certain substances having a hormonal action.

The legislation of non-EU countries also tends to concentrate on the control of the use of hormones in animal feed and meat for human consumption.

See also: **Adulteration of Foods**: History and Occurrence; **Consumer Protection Legislation in the UK**; **Contamination of Food**; **European Union**: European Food Law Harmonization; **Hazard Analysis Critical Control Point**; **Heavy Metal Toxicology**; **Metals Used in the Food Industry**; **Mycotoxins**: Occurrence and Determination; **Pesticides and Herbicides**: Types, Uses, and Determination of Herbicides; Toxicology; **Radioactivity in Food**; **Spoilage**: Bacterial Spoilage; Molds in Spoilage; Yeasts in Spoilage; **Toxins in Food – Naturally Occurring**

Further Reading

Egmond HP (1999) Natural toxins I. Mycotoxins. In: van der Heijden K, Younes M, Fishbein L and Miller S (eds) *International Food Safety Handbook: Science, International Regulation, and Control*, pp. 341–355. New York: Marcel Dekker.

Elmadfa I (1999) *Expertgutachten zur Lebensmittelsicherheit Nitrat*. Vienna: Institut für Ernährunugswissenschaften der Universität Wien.

Gray PS (1991) Agriculture and trade. Radioactivity in foods. In: Segerstähl B (ed.) *Chernobyl, A Policy Response Study*. Berlin: Springer Verlag.

Gray PS (1999) The regulation of food safety – general aspects. In: van der Heijden K, Younes M, Fishbein L and Miller S (eds) *International Food Safety Handbook: Science, International Regulation, and Control*, pp. 597–615. New York: Marcel Dekker.

Legislation Unit (1999, updated) *Contaminants in Foodstuffs – An International Review of Maximum Limits Part 1*. Leatherhead, UK: Leatherhead Food RA.

Legislation Unit (2000, updated March 2001) *Contaminants in Foodstuffs – An International Review of Maximum Limits Part 2*. Leatherhead, UK: Leatherhead Food RA.

Richmond M (1990) The Microbiological Safety of Food – Report of the Committee on the Microbiological Safety of Food to the Secretary of State for Health, the Minister of Agriculture, Fisheries and Food, and the Secretaries for Wales, Scotland and Northern Ireland. London: Her Majesty's Stationery Office.

Smith JS, Smith JA and Douek NL (eds) (1993, updated April 2002) *EC Food Legislation*, vol. 1, 3rd edn. Leatherhead, UK: LFRA.

Codex

P Berry Ottaway, Berry Ottaway & Associates Ltd, Hereford, UK

Introduction

From the mid-1960s, the Codex Alimentarius (Codex) has grown in importance as the forum for the development of international standards and codes of practice for food commodities and products. Codex now has an enhanced role in the international trade in food as it is the preferred organization for the arbitration of disputes under the World Trade Organization (WTO) agreements.

Background

The Codex Alimentarius Commission was inaugurated in 1963 as a subsidiary body of the World Health Organization (WHO) and the Food and Agriculture Organization of the United Nations (FAO). The name Codex Alimentarius was derived from the Latin and can be translated as 'food code.' In the early 1960s, when the Commission was established, there was already a perceived need to facilitate world trade in food and food commodities and it was believed that this could be helped by the development of internationally accepted standards.

Although most of the developed countries had detailed food legislation, sometimes for centuries, there was a diversity of approach to food control. Barriers to international trade in foods were developing at a time when improvements in food preservation techniques and transport meant that international trade in food was no longer being limited by either the perishability of the food or the means to transport it safely.

The Codex Alimentarius, or Codex system, as it is now more commonly known, was set up to facilitate trade by establishing internationally acceptable standards and codes of practice for the production and handling of food. The objectives were to provide added protection for consumer health and to promote fair practices in the food trade. To accomplish these objectives, Codex has to try and harmonize the best features of ideas and concepts devised by its various national members.

Membership of the Codex Alimentarius Commission

All member nations and associate members of the FAO and WHO can become members of Codex. The membership has increased over the years and 165 countries were Codex members in 2000, representing 97% of the world's population. The activities are funded by the regular budgets of both the FAO and WHO. There are no other financial obligations on members except that an important contribution, mainly in terms of services, is made by those member states who act as host countries for the meetings of the Codex Committees. As there is no requirement for a financial contribution from the individual members, it means that the poorer nations can participate in Codex activities on equal terms with the wealthier ones.

Codex and the World Trade Organization

The beginning of 1995 marked the start of a new era in world trade with the establishment of the WTO. The aims of WTO are 'to provide the common institutional framework for the conduct of trade relations among its members in matters related to the agreements and associated legal instruments.' The actual functions of the WTO are to facilitate the implementation and operation of the Multilateral Trade Agreements; to provide a forum for trade negotiations concerning multilateral trade; to administer trade disputes; to monitor national trade policies; and to cooperate with the International Monetary Fund and the World Bank.

The Codex Alimentarius has gained a greater significance since the formation of the WTO. The Agreement on the Technical Barriers to Trade (TBT) that was introduced following the Tokyo Round on World Trade in 1979 had a substantial impact on the establishment of policies on food control. The TBT agreement does not specifically mention Codex but dealt with aspects of food not directly related to safety such as labeling, quality, and packaging and thus impinged on Codex. The WTO, however, has recognized Codex as the preferred international organization for the arbitration and settlement of disputes related to food trade.

The Codex Alimentarius

The Codex Alimentarius used to consist of 25 volumes, of which 17 consisted of the agreed Codex standards and the remainder of the recommended codes of practice and guidelines. In the 1990s these were revised into 13 volumes (Table 1). The published volumes of the Codex are available in English, French, and Spanish. Individual standards

Table 1 Structure of the Codex Alimentarius

Volume	*Subject matter*
1A	General requirements
1B	General requirements – food hygiene
2A	Pesticide residues in foods (general texts)
2B	Pesticide residues in foods (maximum residue limits)
3	Residues of veterinary drugs in foods
4	Foods for special dietary uses (including foods for infants and children)
5A	Processed and quick-frozen fruits and vegetables
5B	Fresh fruits and vegetables
6	Fruit juices
7	Cereals, pulses (legumes), and derived products and vegetable proteins
8	Fats and oils and related products
9	Fish and fishery products
10	Meat and meat products; soups and broths
11	Sugars, cocoa products, and chocolate and miscellaneous products
12	Milk and milk products
13	Methods of analysis and sampling

are being made available on the worldwide web and CD-ROM.

Codex Standards

The purpose of a Codex standard is to set out the required qualities of a food or commodity as sold. The standard has to be set out in objective terms and the required features should be able to be verified. Where necessary, procedures for the verification of a given parameter are included. When a standard has been adopted, the Codex Alimentarius Commission will recommend it to the governments of its member states and invite them to accept the standard and embody it in their national law. The governments are also asked for their reactions to the standard and all their responses are collated and published as part of the Codex Alimentarius.

Codex standards are by nature quantitative in that they have to give limits for most of the parameters such as the minimum required or the maximum allowed and, in some cases, both. If such limits are to have any validity, it must be possible to check the product's compliance by agreed measurements or analysis. A test must be capable of being carried out to similar degrees of accuracy by different laboratories across the world. In the development of testing methods there are always a number of issues surrounding the validity of tests, including the sampling procedures and the problem that food products are not always sufficiently homogenous. These problems have to be considered and resolved by the Codex Committee on Methods of Analysis and Sampling (CCMAS).

The CCMAS has established methods that may be used as agreed procedures by producers, processors, quality examiners, enforcement agencies, and trades worldwide. In summary, these methods:

- define what is being measured and the units in which the answers are to be expressed;
- describe in detail and with precision how measurements are to be made;
- specify how the results should be interpreted (i.e., to determine whether or not the sample complies with the requirement of the standard).

The methods are normally devised from those developed in various countries and they have to be validated to international satisfaction through practical evaluation in a number of different laboratories. It is important that all methods become internationally accepted procedures and there is the advantage that, by developing them through the Codex system, the high cost of development can be shared by a number of countries. This also has a direct benefit for the smaller member countries with limited resources.

Once a standard together with its methods of analysis and sampling has been agreed, the next step in the Codex system is to get the standard accepted by the governments of the member countries. There are various levels of acceptance and each type, together with its attendant obligations, is detailed in the Codex Procedural Manual.

The acceptances fall into three sets of conditions: full acceptance; acceptance with specified deviations; and free distribution. Under the conditions of full acceptance, the products to which the Codex standard and criteria apply, both home-produced and imported, should comply with all relevant requirements. A country may also accept the standard with specified deviation. In such cases statements giving the reasons for the deviations from the standard or required limits are necessary. A country may also determine that a product conforming to the standard and defined limits may be allowed unrestricted distribution.

The conditions of acceptance are slightly different for certain types of standards. There are differences between those for general standards such as nutrition labeling and those for specific commodity standards. For some, such as the Codex Maximum Limits for Residues of Pesticides and Veterinary Drugs in Food, acceptance with specified deviations is not allowed.

Governments are formally requested to indicate which form of acceptance they are adopting and their responses form part of the Codex Alimentarius. This section is regularly updated and permits those involved in international trade to determine the current state of the various national regulations for those foods that are covered by Codex standards.

As the current voluntary adoption procedure with its options does not appear compatible with the WTO system, and as the Codex Alimentarius is the officially recognized reference for international standards in terms of food safety, changes to the acceptance procedure are being discussed.

Codex Codes of Practice and Guidelines

Whilst the development and use of Codex standards effectively produces product specifications, it is also important that issues such as hygiene and good manufacturing practice (GMP) are also addressed, preferably in the context of the Codex Alimentarius. A series of Codex codes of practice and guidelines have been compiled from the collective knowledge and experience of members and observers. Such compilations include the general principles of food hygiene and recommended international codes of practice for a wide range of foods, including fishery products, low-acid canned foods, egg products, frozen fruits and vegetables, and dried milk.

The objectives of the codes of practice and guidelines are fourfold. They provide checklists on GMP for the established processors and also guidance for new or expanding processors in the setting-up of production operations in terms of premises, equipment, manufacturing methods, and control.

The codes and guidelines also assist the food control and monitoring agencies by providing guidance on internationally recommended methods and provide the food industry with grounds for supporting recommended practices and for rebuttal of criticism. As they are codes or guidelines, as opposed to clearly defined standards, they can allow for adjustment to meet particular local circumstances but still provide the essential requirements to support GMP.

Before a code of practice and guideline can be adopted within Codex, representatives of the member governments must give their assent. This level of national recognition should encourage the authorities within the countries to develop their criteria in an internationally accepted rather than an arbitrary manner.

The Codex Organizations

The development of Codex standards, codes of practice, and guidelines are coordinated by the Codex Alimentarius Commission and its Executive Committee. The Commission is also supported by the Secretariat of the Joint FAO/WHO Food Standards Programme in Rome, Italy.

The detailed work on Codex is divided between a number of committees. The committees have been set up so that some focus on 'horizontal' issues such as food labeling, whilst others are concerned with 'vertical' issues such as specific requirements for foods for special dietary uses. During the development of a standard it will be considered by both types of committee as necessary.

In addition, there are five regional coordinating committees covering Africa, Asia, Europe, Latin America and the Caribbean, and North America and South-West Pacific. The Commission and its committees can receive advice from FAO and WHO Expert Committees such as the Joint FAO/WHO Expert Committee on Food Additives.

There are also three other intergovernmental bodies dealing with specific groups of commodities which report to the Codex Alimentarius Commission. These are a Joint FAO and WHO Committee of Government Experts on the Code of Principles Concerning Milk and Milk Products, a Joint United Nations Economic Commission for Europe (UNECE)/Codex Group of Experts on Fruit Juices, and a Joint UNECE/Codex Group of Experts on Quick Frozen Foods.

The Codex committees are grouped as either specific or general subject committees, of which there are currently nine dealing with horizontal subjects such as food additives, labeling and pesticide residues. The other committees are responsible for specific groups of commodities such as fats and oils or vegetable proteins. There have been 15 Codex Commodity Committees, some of which have been adjourned, leaving seven active in 2002 (**Figure 1**).

As part of Codex policy, each committee has a host country. The governments of the host countries are responsible for hosting the meetings of their allocated committee and for providing the administrative infrastructure for the meetings. As the host country also takes financial responsibility for the meetings held in its territory, the current host countries are mostly in North America or western Europe. New Zealand hosted the Meat Hygiene Committee from 1972 until its adjournment in 1983 and Hungary took over the Methods of Analysis and Sampling Committee from Germany in 1972, and the Committee on Tropical Fresh Fruits and Vegetables, which was formed in 1988, is hosted by Mexico.

Due to rapid developments in biotechnology, and particularly genetic engineering, Codex has formed a task force on biotechnology. This task force held its first meeting in March 2000 during which it agreed to develop principles for the risk analysis of foods obtained through modern biotechnology and to develop guidelines for the safety assessment of foods obtained by such processes.

Development of Codex Standards

The present procedures for the development and adoption of Codex standards have evolved from a number of years of practical experience. The basis on which the mechanism relies is the requirement for objectivity at all stages of the development of a standard. The procedure must also be sufficiently robust to accommodate a dichotomy of views and approaches and to bring them together to an agreed standard. Due to the often very large attendance, the committee meetings have to be structured to allow the business to be conducted efficiently. The procedure has been broken down into eight discrete steps:

Step 1

A decision is taken by the Commission to elaborate a standard on a particular subject. The requirement may be either horizontal (e.g., requirements for the nutrition labeling of foods), or vertical (e.g., requirements for the composition of a specific category of food products). After the decision is made by the Commission, the work is assigned to the appropriate

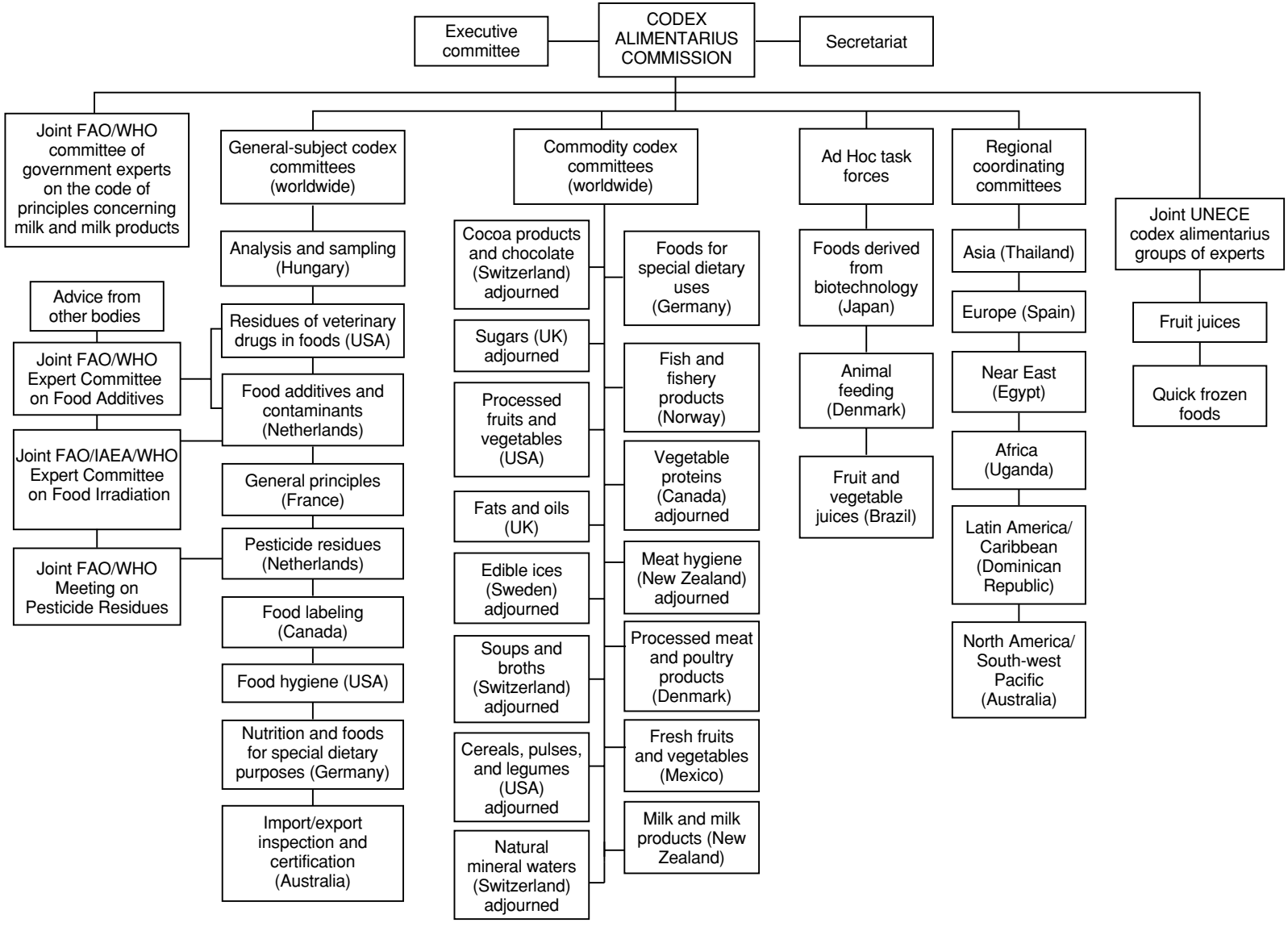

Figure 1 Organization of the Codex Alimentarius Commission. FAO, Food and Agriculture Organization; WHO, World Health Organization; IAEA, International Atomic Energy Agency; UNECE, United Nations Economic Commission for Europe.

committee. In some circumstances the decision may also be taken by the committee and endorsed by the Commission.

Step 2

The Secretariat arranges for a proposed draft standard to be drawn up. This will include consultation with the Committee and the relevant experts and interests in the particular field.

Step 3

The proposed draft standard is sent to all the member governments and international organizations for comments.

Step 4

The Secretariat collates the comments and forwards then to the Committee.

Step 5

Once agreed, the proposed draft standard is sent through the Secretariat to the Commission for adoption as a draft standard. The Commission adopts the proposal as a draft standard on the basis of a majority of the votes cast.

Step 6

The adopted draft standard is sent to the governments and international organizations for comment.

Step 7

The Secretariat collates and forwards the comments to the Committee.

Step 8

When finally agreed, the draft standard is returned to the Commission for adoption as a Codex standard, which is then sent to the governments for their acceptance.

When a proposal is adopted by the Commission as a draft standard (step 5), in certain circumstances the Commission may, on the basis of a two-thirds majority of the votes cast, decide on a simplified procedure which takes the draft directly to Step 8. A decision to go to the simplified procedure can be taken in the case of urgent problems related to trade or public health; the revision or updating of existing Codex standards; or in the case of matters concerning new scientific information or technology.

The proposed draft standard and draft standards are considered in detail at the biannual committee meetings to which the member governments are encouraged to send delegates. The meetings are run in a democratic way and the national delegates have an equal say regardless of the size of their country. Unlike the European Union, where the development of legislation is governed by a weighted voting system whereby the larger countries have more votes than the smaller members of the Community, consensus in Codex committees is reached with all member countries having an equal say.

For each Codex or Commission meeting the agenda is effectively set by the status of the draft standard following the previous meeting. In preparation for each meeting the Secretariat will send out, some months in advance, circular letters to the governments and interested organizations with notification of the specific items in the draft which require an opinion or comment. Those replies which are received in time are made available before the meeting so that all members can assess the range and nature of the opinions.

At the meetings verbal argument can supplement or discuss the written opinions and it is not unusual for divergent views to be presented during the first stages of the development of a standard. These are often due to serious differences of interest. A section dealing with consumer interest in one country may have a social or economic impact upon another. A good example of this would be where a proposal put forward by an importing country for unrealistically low levels of pesticides in a food could lead to serious economic problems for a producing country. The advantage of the meetings is that it provides a forum in which the various views can be aired and where views on both the risks and benefits of proposals can be discussed and objective solutions agreed.

Before each meeting adjourns, a draft report on the meeting and its conclusions is discussed, amended, and adopted. The agreed report is then circulated.

The procedure for the development of a Codex standard normally takes between 7 and 10 years to accomplish. Although a few Codex committees meet annually, the Commission and most of the committees only meet biannually. This has the limitation in that the time scale can make it difficult for the Codex to keep pace with a rapidly changing scientific and political environment. As a consequence, additional working groups may have to be convened to consider urgent matters reflecting a changing environment, such as international concerns over the trade in genetically modified foods.

Codex and Food Safety

There are a number of areas of consumer protection and health where Codex has already played a significant role. One of these areas is food safety. Food

safety is multifaceted and each aspect may need a particular approach.

For example, poor food hygiene is a major cause of human illness worldwide and exceeds by manyfold the effects on the consumer from other aspects of food safety. Most food hygiene problems are due to faulty handling of the food, often through ignorance. Whilst establishing good hygiene practices is relatively easy in principle, the realities of life are that it is far more difficult to maintain the standards.

Codex has, therefore, developed a number of international codes of hygiene and/or technological practice. These cover the requirements for premises, equipment, and handling procedures. In addition to the general principles of food hygiene, Codex has specific codes relating to commodity groups such as fish and fishery products, egg products, meat and poultry products, and dried milk. These Codes are complemented by the Codex General Principles for the Establishment of Microbiological Criteria in Food. These principles have been developed with their practical application in mind and include clear advice on how microbiological testing can be used to check ingredients and processes and to determine the necessary actions or changes before problems develop. It is also indicated that the testing can be adapted so assurance can be obtained that the provisions in the codes have been observed. This latter aspect has relevance to imported foods. In addition, the principles are laid down by which to determine those instances or foods where the presence of specific microorganisms in excessive numbers means that the food should not reach the consumer. These specific provisions were developed as it was recognized that in the absence of such guidelines it was possible that the control authorities in some countries might place excessive importance on numerical data for some microorganisms, which might lead to needless rejection of a food without any increase in consumer protection.

Other areas of food safety where Codex is directly involved are contaminants and additives. The term contaminants is used in this context to denote the presence in food of small quantities of undesirable substances such as dioxins or polychlorinated biphenyls (PCBs), and the elements that can cause health hazards such as lead, mercury, and cadmium. The Codex Committee on Food Additives has the responsibility of setting achievable limits for these contaminants.

Although the use of agricultural chemicals such as pesticides has become essential to modern agricultural and horticultural practices, their residues in food have consistently given rise to concerns. Lists of pesticides which should be authorized for use have been drawn up by the Codex Committee on Pesticide Residues. These lists have been developed from the work of the Committee of Experts of the FAO/WHO Joint Meeting on Pesticide Residues. As a result of this work, over 2000 maximum residue limits have been defined based on advice from the Expert Committee on Toxicity and also the procedures of good agricultural practice. In a similar way, committees are considering the complex problems of veterinary drug and hormone residues in foods.

Food additives play an important part in the quality and safety of much of the processed food. The Codex Committee on Food Additives (CCFA) has responsibility of implementing the reports of the Joint Expert Committee on Food Additives (JECFA). JECFA is a group of independent and internationally recognized experts which is appointed jointly by WHO and FAO to review and report on the safety of additives.

The CCFA has, as a primary task, to consider the technological need for an additive on the one hand and the JEFCA safety evaluation on the other. To accomplish this, it has to elicit from the Codex Commodity committees both the degree of usage of the additive and its importance to the various sectors of the industry. It is the responsibility of both the CCFA and JECFA to treat all additives on an equal and objective basis and to give equal consideration to those of both natural and synthetic origin.

In dealing with additives it is important for the governments, enforcement agencies, and the public that any of the additives listed as permitted in foods have been considered by a group of independent and respected experts who have provided an impartial consensus opinion that can be accepted with confidence. One of the strengths of Codex is that it provides the structure by which this can be accomplished.

The Future of Codex

The recognition by the WTO and the greater involvement of the developing countries are contributing to the strength of Codex. The Codex Commission is aware of the need to keep pace with changing food quality and safety concepts and developments such as the application of genetic engineering to the food supply. The strength of Codex has been and will continue to be its truly international base.

See also: **Food Additives**: Safety; **Food and Agriculture Organization of the United Nations**; **Food Safety**; **Legislation**: History; International Standards; Additives; Contaminants and Adulterants; **World Health Organization**

Further Reading

Codex Alimentarius Commission (1999) *Rules of Procedure of the Codex Alimentarius Commission*. Available on line at HTTP:www.fao.org. Directory: waicent/faoinfo/ economic/esn/codex/Manual/Files:rules.htm.

Food and Agriculture Organization of the United Nations, World Health Organization (1989) *Introducing Codex Alimentarius*. Rome: FAO/WHO Joint Secretariat.

Food and Agriculture Organization of the United Nations, World Health Organization (1999) *Understanding the Codex Alimentarius*. Rome: FAO/WHO Joint Secretariat.

World Trade Organization (1994) *Agreement Establishing the World Trade Organization*. Geneva: WTO.

World Trade Organization (1994) *Understanding on Rules and Procedures Governing the Settlement of Disputes*. Geneva: WTO.

LEGUMES

Contents

Legumes in the Diet

M A Uebersax and L G Occeña, Michigan State University, East Lansing, MI, USA

Introduction

Legumes are a vital food resource which contribute to the nutritional well-being of diverse human diets. They provide essential nutrients and high levels of protein with moderate levels of energy and dietary fiber. This article discusses the structural features, chemical composition, handling, storage, processing, and utilization of legumes.

General Features

The term legume is derived from a Latin word *legumen*, meaning seeds harvested in pods. Alternative terms for the edible seeds include grain legumes and pulse, from the Latin *puls*, meaning pottage. Food legumes refer to those species of the plant family Leguminosae (pea or bean family) that are consumed directly as mature dry seeds, as immature green seeds, or even as green pods with the immature seeds enclosed. The plants vary in size from tiny wild vetches to large trees. Most of the important legumes in the human diet are mainly herbaceous, often annuals, and grow throughout the world. **Table 1** lists some of the more important food legumes with their scientific and common names. Two categories of grain legumes are: (1) oilseeds such as soy beans and groundnuts, grown for their protein and oil content; and (2) pulses, or grain legumes, including common beans and peas, which are grown principally as a protein source and used either as boiled whole grain in different preparations or as split seeds.

Only about 20 out of approximately 13 000 species of legumes are commonly consumed, with most of the consumption occurring in Central and South America, Africa, and Asia. World consumption (percentage of total) of common beans has been characterized by regions as follows: Latin America 46.6%; Africa 24.3%; western Europe 6.3%; eastern Europe 7.0%; Asia 4.5%; Middle East 4.0%; and North America 7.5%. The pattern of consumption and preference for one type of legume over another depends upon the availability in the area. The annual per capita supply for 1998 is highest for Central America (13.7 kg), followed by South America (10.3), sub-Sahara (8.8 kg), Africa (8.2 kg), North America (7.2 kg), Asia (5.4 kg), Oceania (4.8 kg), and Europe (2.7 kg). However, the annual per capita consumption of legumes of some Mediterranean countries (except Lebanon 14.5 kg; Turkey 13.5 kg) is more than twice as much as the rest of Europe.

World production of major food legumes exceeds 59×10^6 t per annum **Table 2**. India is the world's leading producer, as well as the world's leading importer of pulses. Canada is the leading world exporter of pulses; Australia, Myanmar, China, and the USA are also world exporters outside the European Community (EC). India, China, and the USA are the major peanut-producing countries, with India contributing nearly 30% of total world production.

Table 1 Important food legumes with their scientific and common names

Scientific name	*Common name*
Arachis hypogea L.	Arachide, groudnut, peanut, earthnut, goober pea, Philippine pink, mani
Cajanus cajan L. *Millsp.*	Pigeon pea, Congo pea, gandul, guandu, red gram, Angola pea, yellow dhal, catjang pea
Cicer arietinum L.	Chickpea, gram pea, Egyptian pea, pois chiche, Bengal gram, gram, chana, garbanzo, alter, shibu
Cyamopsis tetragonolaba L.	Guar gum
Dolichos uniflorus	Horse gram, horse grain, Madras gram, kallu
Glycine max (L.) *Merr.*	Soybean, soya bean, soya, haba soya, preta
Lablab purpureus (L.) *Sweet*	Hyacinth bean, Egyptian bean, val
Lathyrus sativus L.	Khesari, chickling vetch, grasspea
Lens culinaris Medik	Lentil, masur
Lupinus albus L.	White lupine
Lupinus angustifolius L.	Blue lupine, New Zealand blue lupine
Lupinus luteus L.	European yellow lupine
Macrotyloma uniflorum (Lam.) Verdc.	Horse gram, Madras gram, kulthi
Phaseolus lunatus L.	Lima bean, butter bean
Phaseolus vulgaris L.	Common bean (field bean, haricot, pinto, navy, kidney, dry bean), garden bean, frijol, caraota, poroto
Pisum arvense sativum L.	Pea, pois, arveja, Alaska pea, common or garden pea, dry peas, guisante, muttar
Psophocarpus tetragonolobus (L.) *DC*	Winged bean, Goa bean, four-angled bean, Manila bean, princess bean
Vicia faba L.	Broad bean, faba bean, horse bean
Vigna aconitifolia (Jacq.) Marechal	Moth bean, mat bean
Vigna angularis Wild	Adzuki bean
Vigna aureus	Green gram, golden gram, mung bean, Oregon pea, chickasano pea, mungbean
Vigna mungo (L.) *Hepper*	Urd, black gram
Vigna radiata (L.) *Wilczek*	Green gram, golden gram, mung bean
Vigna umbellata (Thunb.) Ohwi and Ohashi	Rice bean, mambi bean
Vigna unguiculata (L.) *Walp. ssp. Unguiculata*	Cowpea, black-eyed pea, crowder pea
Voandzeia subterranea (L.) *Thouars*	Bambarra groundnut

Structural and Anatomical Features of Legume Seeds

The seeds of leguminous plants differ in color, size, shape, and seed coat thickness. The seed coat (testa) is the outermost layer of the seed that protects the embryonic structure from damage caused by water absorption and microbial contamination. Figure 1 illustrates the anatomical features of a legume seed. The hilum (large, oval attachment scar) and the micropyle (minute opening in the seed coat) regulate seed water imbibition. The internal structure of the seed includes the shoot (leaf tissue), consisting of two cotyledons and the embryonic axes, including the epicotyl (stem tip), hypocotyl (stem), and the radicle (root). Seed coat thickness, seed volume, hilum size, and protein and starch content all influence water absorption.

The seed coat comprises between 6% and 8% of the total seed weight. The major components of the seed coat of legumes include the waxy cuticle layer (hydrophobic barrier), intermediate layer (palisade and hourglass cells) and the thick-cell-walled parenchyma cells that appear spongy and swell dramatically during hydration.

The cotyledon, which contains reserve starch and protein, is the largest component of the dry bean by both weight and volume. Processed texture and nutrient availability of beans are both influenced by the dimension and arrangement of the cotyledon cells. The outermost layers consist of epidermal cells, containing granular protein structures with limited starch, and the hypodermis, which has larger, elliptically shaped cells. The parenchyma cells comprise the largest portion of the cotyledon. These cells are bound by a distinct cell wall and middle lamella with a few vascular bundles. Thick walls, in the longitudinal section, give rigidity to the cotyledon. The secondary walls, found only in mature parenchyma cells, are very thick and contain pits that facilitate the diffusion of water during soaking. Within each parenchyma cell, starch granules are embedded within a protein matrix; between adjacent cells, the middle lamella, composed mainly of pectic substances, adds potential rigidity to the tissues.

Comparative Chemical Composition

Carbohydrates

The total carbohydrates of legumes range from 24.0%, on a dry-weight basis (dwb) in winged beans to 68% in cowpeas. Oligosaccharides of the raffinose family (raffinose, stachyose, verbascose, and ajugose)

Table 2 World production of important food legumes

	Year	World production (Mt)	Production (Mt)	
			Developing countries	Developed countries
Soybeans	1996	130 206 297	61 229 828	68 976 469
	1997	144 411 997	66 023 152	78 388 845
	1998	159 954 599	79 718 137	80 236 462
	1999	156 812 413	79 044 937	77 767 476
	2000	161 042 126	80 445 242	80 596 884
Groundnuts (in shell)	1996	31 231 836	29 231 442	2 000 394
	1997	29 416 106	27 532 194	1 883 912
	1998	34 029 781	32 005 254	2 024 527
	1999	31 928 137	29 908 338	2 019 799
	2000	34 522 077	32 742 008	1 780 069
Beans, dry	1996	16 451 242	14 261 463	2 189 779
	1997	16 989 109	14 591 274	2 397 835
	1998	16 740 216	14 332 655	2 407 561
	1999	19 069 325	16 462 744	2 606 581
	2000	18 825 653	16 504 738	2 320 915
Peas, dry	1996	10 841 662	2 327 998	8 513 664
	1997	12 040 571	2 239 786	9 800 785
	1998	12 172 357	2 312 513	9 859 844
	1999	11 105 548	2 281 597	8 823 951
	2000	10 928 256	2 180 017	8 748 239
Chickpeas	1996	8 006 206	7 604 771	401 435
	1997	8 254 562	7 943 335	311 227
	1998	8 769 939	8 471 886	298 053
	1999	9 326 676	8 897 155	429 521
	2000	8 801 590	8 339 013	462 577
Broad beans, dry	1996	3 501 080	3 090 063	411 017
	1997	3 215 712	2 789 143	426 569
	1998	3 604 467	3 156 529	447 938
	1999	3 300 685	2 810 239	490 446
	2000	3 183 427	2 719 268	464 159
Cow peas, dry	1996	2 583 941	2 528 590	55 351
	1997	2 551 747	2 501 447	50 300
	1998	3 328 615	3 276 415	52 200
	1999	3 273 714	3 223 864	49 850
	2000	3 303 590	3 253 740	49 850
Pulses, total	1996	53 071 052	37 827 833	15 243 219
	1997	55 153 122	38 411 608	16 741 514
	1998	55 924 233	39 080 996	16 843 237
	1999	58 114 070	41 789 204	16 324 866
	2000	55 738 452	40 065 145	15 673 307

Data adapted from FAO (1996–2000). (http://apps.fao.org/page/form?collection = Production. Crops. Primary&Domain = Production&servlet = 1&language = EN&hostname = apps.fao.org&version = default)

account for 31.1–76.0% (dwb) of the total sugars. Starch makes up a larger portion, ranging from 24% to 56% (dwb) of total carbohydrate. These variations in reported starch contents result in part from differences in cultivars and analytical procedures. The physiocochemical properties and internal molecular structures of bean starch differ depending on the original source, maturation, and environmental factors. The majority of legume starches have molecular weights higher than 2×10^4, and over 90% of them have molecular weights about 4×10^4. (*See* **Carbohydrates**: Classification and Properties.)

The amount of amylose, ranging from 10% to 66%, may influence starch solubility, lipid binding, and other functional properties. The variability between amyloses from different legumes may be the result of: (1) maturity of seed; (2) genetic control of amylose synthesis; (3) cultivar differences; and/or (4) seed storage history. A degree of amylose polymerization may confer structural stability on the granule but may also be partially responsible for its resistance toward *in vitro* α-amylolysis.

Most legume starches have gelatinization temperatures of 60–90 °C and are characterized by no distinctive pasting peak, but rather a very high viscosity which remains constant or increases during cooling. The restricted and single-stage swelling patterns of legume starches are explained by the crystalline and

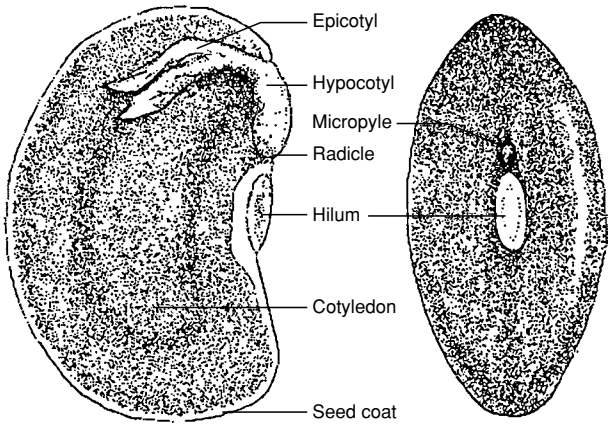

Figure 1 Anatomical structure of a legume seed.

amorphous regions of the starch granules and the presence of strong binding forces which relax within one temperature range.

Large variation in nonstarch polysaccharides (NSP), including fiber content (1.2–13.5%), has been observed in selected types of legumes. A significant proportion (80–93%) of the fiber is localized in the seed coat. Beans possess relatively high levels of soluble and insoluble fiber, and significant levels of pectin. Cellulose is the major component of fiber in smooth and wrinkled peas, red kidney, and navy beans, while in other legumes (lupins, lentils, broad beans, etc.), hemicellulose predominates. Cotyledon cell walls contain higher levels of pectin than cellulose. The seed coats are primarily composed of cellulose (29–41%) with small amounts of lignin (1.2–1.7%). Total dietary fiber is higher in cooked beans than in raw beans. Water-soluble fractions can be lost with cooking water, even though the beans are intact, while long boiling may cause modifications in certain constituents of other fiber fractions. (*See* **Dietary Fiber**: Properties and Sources.)

Proteins

Legumes are dense sources of plant protein that ranges from 20% to 40% (dwb). The bean storage (reserve) proteins are predominant in the globulin fractions, while the metabolic (enzymatic) proteins are primarily found in the albumin fraction. The storage proteins have high affinity for water and contribute dramatically to the nutritional and functional properties of the cooked seed. Bean proteins are relatively high in essential amino acids, in particular lysine, threonine, isoleucine, leucine, phenylalanine and valine. However, they are deficient in sulfur-containing amino acids, in particular methionine and cystine. This explains their association with cereal grains that are deficient in lysine, but are a substantial source of methionine. (*See* **Protein**: Food Sources.)

The globulin fraction can be subdivided into globulin I (GI: high salt solubility) and globulin II (GII: low salt solubility). Phytohemagglutinin (PHA), the lectin of dry beans, is identified as GII with two general types of subunits: an erythrocyte reactive subunit, E, and a lymphocyte-stimulating subunit, L. PHA has relatively high levels of aspartic acid and serine, but possesses very little cystine and methionine. It has the ability to agglutinate red blood cells, disrupt the internal mucosa, and reduce nutrient absorption. Leguminosa lectins will bind both the storage protein and the glycosidase enzymes. It has been suggested that lectins act during the maturation of the plant to contribute to an orderly arrangement of the storage proteins in the protein bodies.

Total biological utilization of the legume protein is relatively low, with digestibility less than 76%. Protein digestibility may be impaired by the presence of numerous antinutritional compounds that must be removed or destroyed during processing. There is considerable variability in protein digestibility among different commercial classes of *Phaseolus vulgaris*, and also among cultivars within a class. Heating for longer than 30 min at 121 °C results in lowered protein quality and decreased available lysine in *P. vulgaris*. The protein efficiency ratio (PER) of raw and cooked legumes is approximately 0 and 1.2, respectively.

Minor Constituents

Ash and minerals The total ash content of *P. vulgaris* ranges from 3.5 to 4.1% (dwb). Beans are generally considered to be substantial sources of calcium and iron. They also contain significant amounts of phosphorus, potassium, zinc, and magnesium. Differences among bean classes and environmental factors influence variability. Ash content decreases after cooking because of leaching, with losses ranging from 10% to 70%. The wide range of reported losses could be due to different soaking and cooking methods. (*See* **Minerals – Dietary Importance**.)

Vitamins Dry edible beans provide several water-soluble vitamins (thiamin, riboflavin, nicotinic acid, and folic acid) but very little ascorbic acid. However, variability of vitamin content is high. Commercial methods of preparation of canned beans cause a significant loss of water-soluble vitamins. (*See* **Vitamins**: Overview.)

Lipids Neutral lipids are the predominant class and account for 60% of the total lipid content. Phospholipids comprise 24–35%, while glycolipids account

for up to 10% of the total lipid content of legume seeds. Most dry beans possess relatively low total fat content, generally 1–2%. The fatty acid composition of legumes demonstrates a significant amount of variability; however, legume lipids are generally highly unsaturated (55–85%), with linoleic acid present in the highest concentration. Palmitic acid is the predominant saturated fatty acid. (*See* **Fats**: Classification.)

Tannins and polyphenols The procyanidin and condensed tannin content of dry beans ranges from 0.4% to 1.0%. These compounds are localized in the bean seed coat, with low or negligible amounts present in the cotyledon. The hydroxyl groups of the phenol ring enable the tannins to form cross-links with proteins that may be implicated in postharvest seed hardening or decreased digestibility. During cooking, polyphenolic compounds are partially leached into the cooking medium and may interact with cotyledon constituents, yet complete destruction of the tannins does not occur during heat processing. (*See* **Tannins and Polyphenols**.)

Grading, Handling, and Storage

Various genetic, cultural, and environmental factors, including variety, seed source, agronomic conditions, handling, and storage of the dry product, critically affect the final quality of legume food products. Seed germination and vitality are extremely vulnerable to damage during handling and storage.

Beans are not ready for harvest until the moisture level drops to about 20%. Conditions best suited for threshing (removal of seed from pod and separation of plant residue) are seed moisture of 16–20% (on a wet-weight basis) and pod moisture at least 4% lower than seed moisture. Moisture content is also an important factor influencing postharvest storage stability. Accumulation of moisture will result in high seed respiration rates (heat) and subsequent mold development can be a storage problem. Aeration of seeds during storage provides moisture equilibration and ability to inhibit localized 'pockets' of moisture. The seed is vulnerable to mechanical damage, which can result in split seed coats, so design of handling systems (conveyors and augers) must minimize seed dropping.

Bean moisture control, storage temperature, and relative humidity exert a major influence on dry bean quality and on the final product. High moisture and high temperature during storage will result in brown, discolored, and off-flavored beans ('bin burn'). Beans subjected to adverse storage conditions may undergo cooked-seed textural defects which severely limit their food use. Legume seeds that will not soften sufficiently because of a failure to imbibe water during soaking are termed 'hard-shell.' Legume seeds that absorb adequate water during soaking, but will not soften sufficiently during a reasonable cooking time, are designated as hard-to-cook (HTC).

The most noticeable structural changes caused by HTC defect is a failure of cotyledon cells to separate during cooking. Disintegration of cytoplasmic organelles, with inclusions and loosening of attachments between the cell wall and the plasmalemma, are observed in HTC beans. Loss of plasmalemma integrity is considered to be responsible for increases in electrolyte leakage observed in HTC beans. Much research has been directed toward understanding and controlling these reactions.

Processing and Uses

Food legumes are an economical source of supplementary protein for many populations lacking animal protein. Legumes are currently of interest to the consuming public because they provide high nutrient levels without supplying excessive calories, and the perception of their food values has dramatically shifted from the 'poor man's steak' label.

The major physical constraints limiting the use of dry beans are as follows:

1. Preparation: time, convenience, expense
2. Postharvest storage changes: seed hardening – hard shell and HTC; moisture absorption; mold growth; seed discoloration; flavor and odor changes.

The major biological constraints limiting utilization of dry beans are as follows:

1. Antinutrients: protease inhibitors and trypsin inhibitors, lectins and PHA; amylase inhibitors; phytic acid (mineral chelation and malabsorption).
2. Digestibility: procyanidin interactions (tannins and condensed polyphenols, binding of proteins); protein (limited proteolysis, amino acid balance); starch (resistant granules, limited amylolytic degradation).
3. Flatulence factors: oligosaccharides; nonstarch polysaccharides; proteins; retrograded starch.

Figure 2 summarizes the processing and edible byproducts of dry beans.

Soaking

Soaking of legume seeds may be used to moisten and soften the seed to reduce cooking time or aid in seed coat removal. Soaking also reduces the toxin content and surface contamination. The time needed for

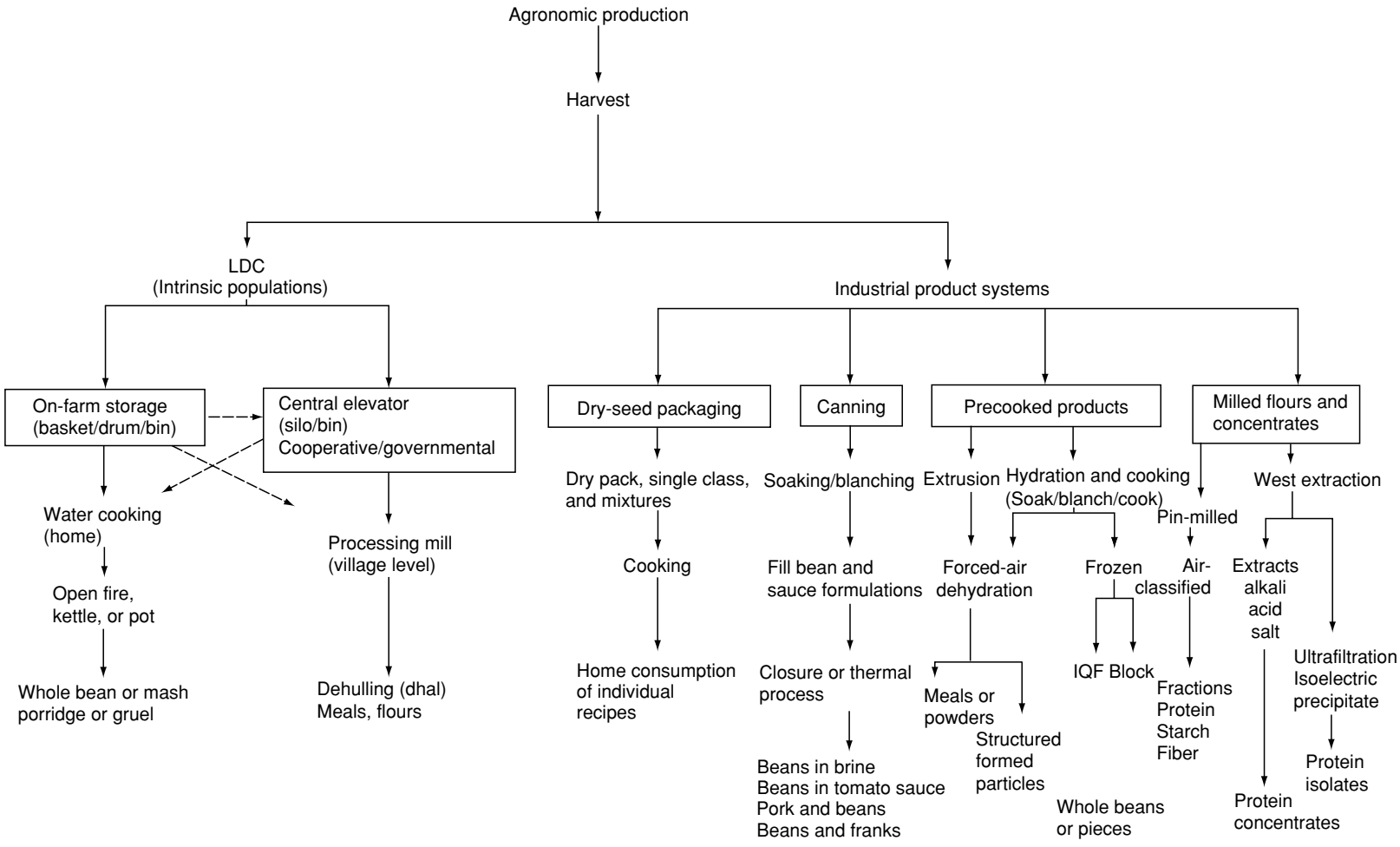

Figure 2 Outline of primary dry-bean processing and end products. LDC, less developed countries; IQF, individually quick-frozen.

soaking varies with variety and species, and with length and conditions of storage. Traditionally, dry beans have been soaked overnight (8–16 h) in cold water. High-temperature soaking accelerates hydration. Experimental methods designed to accelerate seed hydration and tenderization include vacuum, ultrasonic sound, the use of soak-water additive (phosphates) and gamma-irradiation. Cracking or scouring seeds with tough seed coats before soaking assists moisture penetration. The degree and rate of hydration of the starch–protein matrix influences the cooking rate and final texture of cooked beans. The content of total soluble sugars (reducing and non-reducing) and soluble minerals in legumes decreases during soaking because of leaching losses. In many less developed countries (LDC), dry beans are cooked without soaking because of water quality limitations.

Germination, Sprouting, and Fermentation

Soaking legume seeds and holding at ambient temperatures will facilitate germination. Germination has profound effects on the physicochemical composition changes and improves the nutritional quality of legumes by increasing vitamins such as ascorbic acid, riboflavin, choline, thiamin, tocopherol, and pantothenic acid. The sprouted grains may be consumed fresh or, if followed by seed coat removal, can be roasted or ground for use in soups or side dishes.

In Asian countries, soybeans have been fermented to produce a variety of nutritious food products. Some important soybean-based foods are tofu (soybean curd), natto (Japanese product), miso (soy paste), tempeh (Indonesian product), shoyu (soy sauce) and tamari. Fermented foods that are traditional in India include idly (steamed) and dosai (fried) from black gram, and dhokla (steamed) and khaman from Bengal gram. Groundnuts are also fermented with a mold to make ontjom, an Indonesian product similar to tempeh.

Dehulling and Roasting

Dehulling and splitting the cotyledons reduces cooking time considerably. The ease of removing the hull depends on the thickness of the seed coat and can be achieved via wet or dry methods. The wet method involves soaking the grain in water for a few hours, draining, drying, milling, and winnowing to remove the seed coat. In a large-scale dry method, oil is mixed with the seeds by passing through emery-coated rollers to abrade the surface. This is common practice for legumes with particularly tough seed coats, including velvet, hyacinth, and winged beans.

Dehulling and splitting the cotyledons into halves is a process traditionally practiced in Asia and Africa. In India, nearly 70% of food legumes are milled to make dehulled split cotyledons (referred to as dhal), which is prepared by cooking until soft, mashing, mixing with water, then boiling again to give a porridge of uniform consistency. In addition, dhals are cooked with vegetables, spice, and condiments. Roasting and puffing involve the application of dry heat to legume seeds using a hot drier at a temperature of 100–200 °C for a short time (1–5 min) or by direct heating with hot sand. Dhal or whole seeds are roasted to obtain an appealing flavor and maintain maximum protein quality.

Pigeon peas are usually consumed as dehulled splits or dhal. The dhals are cooked until soft and served as a sauce or boiled with rice. They are also incorporated in cereal legume mixtures to feed infants and young children in LDC. Mature seeds can also be soaked before boiling or pounding, then steamed and fried. However, the green seeds, pods, or sprouts may be eaten directly or canned.

Thermal Processing

Heating increases tenderization of the cotyledon, improves flavor, and inactivates enzymes and intrinsic toxic factors of legume seeds.

Cooking Whole legumes may be cooked either with or without the seed coat, in boiling water or steam. Traditionally, cold-water-soaked beans are simmered at subboiling temperatures for several hours or until adequately tender for eating. Cultivar variability, hard-shell, and HTC storage defects may dramatically alter cooking time. The use of 'crock-pot' cookers has also generated interest for preparing beans, although this low-temperature cooking procedure may not adequately inactivate PHA. There is loss of solids from beans during cooking, with the amount depending on species, variety, and cooking method. Although cooking is essential in order to render the beans palatable and inactivate toxic substances, excessive cooking results in loss of protein biological value in the form of lysine and methionine.

Canning Industrial canning overcomes the long period of time and high fuel costs involved in home preparation of legumes. Hot-water hydration (final moisture 53–57%) can be accompanied via serial rotator blanchers with differential temperatures, or continuous pipe blanchers. Hydrated beans may be processed in a standard rotary water blancher for 3–8 min, then continuously washed with cold-water sprays, utilizing rod washers or vibrating sieves to facilitate removal of broken beans and skins. Beans are generally processed in water or brine, tomato sauce, or molasses. The beans will further hydrate during the thermal process and equilibrate with the

sauce or brine to a final moisture content of 65–70% within 2–4 weeks.

Food Ingredient Preparation: Bean Meals, Legume Flours, Concentrates, and Powders

Flours and concentrates made from legumes other than soybeans have had little commercial success. Initial efforts to produce bean ingredients were focused to minimize cell rupture in order to retain the same texture, appearance, and taste of conventionally prepared beans. The compartmentalization of protein and starch enables size reductions and fractionation of dry beans into coarse and fine fractions containing high starch and high protein, respectively. Examples of high-protein food ingredient products are flour processed by pin milling and air classification, and concentrates and isolates processed by alkali, salt, and acid extraction together with isoelectric precipitation or ultrafiltration. Bean protein concentrates containing 72–81% protein, protein isolates containing about 90% protein, and fractions containing 50% starch have all been produced using water extraction techniques. Instant precooked bean powders have been prepared by soaking, cooking, slurrying, and drum- or spray-drying. Recent product development has included the use of spray-dried powders from pea, imitation milk replacers from pea and lentil, and protein curd similar to tofu from pea. Ingredients from dry beans, peas, and lentils have been used in pasta, noodles, and spaghetti. Legume meals and concentrates have been blended into ground-beef preparations. In addition, the high-fiber or high-protein bean flour fractions have been successfully substituted (20%) for wheat flour to produce quality biscuits, doughnuts, quick breads, and leavened doughs.

When raw legumes are ground without pretreatments, they develop undesirable odors and flavors. Lipoxygenase has been held responsible for the appearance of off-flavors by catalyzing formation of hydrogen peroxides from unsaturated fatty acids. However, treatment with dry heat for 6–8 min at 105 °C completely inactivates this enzyme. It is therefore advantageous to apply a rapid heat treatment to beans prior to bean flour production. The attractions of conventional air heating are ease of handling and lack of product contamination, but air does not provide a high rate of heat transfer. Instant navy bean powders, obtained by roasting dry beans in a bed of heated ceramic beads, followed by grinding, resulted in functional food ingredients which showed: (1) stable gel-forming capacity; (2) low trypsin inhibitor and hemagglutinin activities; (3) increased water-holding capacity and cold paste viscosities; (4) no changes in available lysine; (5) limited degree of starch damage; and (6) reduction of water-soluble nitrogen content. Air classification of roasted beans produced high fiber, starch, and protein fractions that are suitable as food ingredients in a variety of food products.

Groundnuts or peanuts are eaten raw, roasted, or pounded into peanut butter, or processed as salted peanuts, sweets, and snacks. They are also incorporated in soups and stews. In Africa, India, China, and South-east Asia, the peanut crop is converted to oil and cake. Edible peanut flour is prepared by grinding peanut kernels in a hammermill or bar-cracking machine. The ground seeds are conditioned to 10% moisture and cooked in a stack cooker for 45–60 min to attain an exit temperature of 116 °C. An expeller is used to reduce the oil content of the cooked meal to 8–12%. The meal is then ground, conditioned to 10% moisture, flaked, and extracted with hexane. The defatted flakes are desolventized, then cooled and ground to either flour or grits. Nutritious, low-cost, peanut-flour-enriched wheat noodles have been developed in Thailand.

Processing of full-fat, partially defatted or fat-free flakes from peanut kernels includes grinding low-moisture (6%) kernels to a flour consistency, mixing with water to form an emulsion suspension, heating this to 93–116 °C to reduce lipoxygenase activity, and drum-drying to form the peanut flakes. Peanut protein concentrate is prepared using defatted peanut flour or meal with protein content between 45% and 60%. The processes or techniques which can be used to isolate peanut protein concentrate include immobilization of the major protein fraction, air or solvent classification, and wet milling.

Bambarra groundnut seeds have lower oil content compared to groundnut, but are richer in protein and calories. They are consumed fresh, grilled, or broiled, or made into flour to form cakes. Immature shelled or unshelled nuts are pounded and boiled into a stiff porridge. They are also served as appetizers and relishes.

Convenience Bean Products

Preparation and cooking time is a major constraint to bean utilization, but numerous chemical and physical processing techniques have been initiated to alleviate this problem.

Quick-cook beans Bean products which cook in less than 15 min have been developed using vacuum infiltration of a solution containing sodium chloride (NaCl), pentasodium tripolyphosphate ($Na_5P_3O_{10}$), sodium bicarbonate ($NaHCO_3$), and sodium carbonate (Na_2CO_3). Further soaking of the beans in this solution is followed by rinsing, drying and cooking or

freezing, depending on the intended use. These products have not been commercially developed, in part because of the elevated sodium level associated with the process.

Frozen or dehydrated beans Numerous precooking and dehydration procedures have been developed which enable production of rapidly reconstituted beans, bean pieces, or bean powders. Fully cooked, dehydrated bean flours, bean 'crumbles' and refried bean mixes are prepared for use in institutional food service systems. Main-dish retail products with a wide variety of styles and seasonings (barbecue, Cajun, *al presto*) have recently entered the market.

Partially or fully cooked beans marketed as individually quick-frozen (IQF) or as a block-frozen product are suitable as institutional food ingredients or for direct retail distribution. Appropriate hydration and softening under low-temperature cooking conditions prior to freezing provide products that are readily suitable for use in recipe-based foods, such as chilli con carne and acidified three-bean salad dishes.

Extrusion Extrusion utilizes continuous high-pressure and high-temperature cooking of starch or flour from legumes and cereals. The use of this technology and proper ingredient formulation may provide precooked or quick-cook legume products. Starch gelatinization occurs primarily during the later stages of cooking of the beans. The amount of water available is critical to the gelatinization process. When the cooking and extrusion processes are combined, the increased pressure differential will facilitate expansion of the final product. This formed product, of open porosity, may require additional air-drying to assure shelf-stability. The dry formed 'bean' will readily hydrate and soften when reconstituted. Die configuration can be designed to produce a variety of intricate product shapes that can be further optimized to produce simulated beans. Appropriate binders and coating agents can be added to improve the appearance of precooked or quick-cook bean products, resulting in promising prototypes of extrusion-cooked formed 'beans.' (*See* **Extrusion Cooking**: Principles and Practice.)

Microwave heating/cooking Several frozen bean and refried bean products, as well as instant bean flour mixtures for sauces and salsa, are commercially available and can be prepared using a microwave oven. Microwave treatment of whole soybeans after soaking to prepare soy flour or soy grits without a burnt odor was reported completely to inhibit trypsin activity. Proper microwave treatment also reportedly increased the *in vivo* digestibility and the metabolizable nitrogen of soybeans.

Legumes are prepared, processed, and consumed using a variety of locally and culturally adapted methods or styles, and are important in subsistence and industrialized economies. Research directed to improve utilization of legumes has included removal of constraints such as postharvest storage defects, cookability, digestibility, and inactivation of antinutritional and flatulent factors. New processes and products may enhance the utility and acceptability of these valuable food resources.

See also: **Beans**; **Canning**: Principles; **Carbohydrates**: Classification and Properties; **Extrusion Cooking**: Principles and Practice; **Fats**: Classification; **Flour**: Dietary Importance; **Legumes**: Dietary Importance; **Peanuts**; **Protein**: Food Sources; **Pulses**; **Tannins and Polyphenols**

Further Reading

Aykroyd WR and Doughty J (1982) *Legumes in Human Nutrition*, 2nd edn. Rome: Food and Agriculture Organization.

FAO (1998) *Production Yearbook*, vol. 52. Rome: Food and Agriculture Organization.

Liu KS (1997) *Soybeans: Chemistry, Technology and Utilization*. New York: Chapman and Hall.

Matthews RH (ed.) (1989) *Legumes: Chemistry, Technology and Human Nutrition*, pp. 187–196. New York: Marcel Dekker.

Salunkhe DK, Kadam SS and Chavan JK (1985) *Postharvest Biotechnology of Food Legumes*. Boca Raton, Florida: CRC Press.

Sinha SK (1977) *Food Legumes: Distribution, Adaptability and Biology of Yield*. Rome: Food and Agriculture Organization.

Uebersax MA and Occeña LG (1993) Suitability of dry edible beans for world food relief programs: applications and nutritive value. *Michigan Dry Bean Digest* 18: 2–13.

Uebersax MA and Ruengsakulrach S (1989) Utilization of dry field beans, peas and lentils. In: Applewhite TH (ed.) *Vegetable Protein Utilization in Human Foods and Animal Feedstuffs*, pp. 123–130. Champaign, Illinois: American Oil Chemists' Society.

Uebersax MA and Ruengsakulrach S (1989) Structural and compositional changes during processing of dry beans (*Phaseolus vulgaris*). ACS Symposium Series 405. In: Jen JJ (ed.) *Quality Factors of Fruits and Vegetables: Chemistry and Technology*, pp. 111–124. Washington, DC: American Chemical Society.

Uebersax MA, Ruengsakulrach S and Hosfield GL (1989) Uses of common dry field beans. In: Lucas EW and Erickson DR (eds) *Food Uses of Whole Oil and Protein Seeds*, pp. 231–252. Champaign, Illinois: American Oil Chemists' Society.

Dietary Importance

Y Lam, Bio-Rad Laboratories, Hercules, CA, USA
B O de Lumen, University of California, Berkeley, CA, USA

Introduction

Since man settled down to practice agriculture, the human diet has consisted mainly of cereals or starchy roots and fruits, which are poor sources of protein. With few exceptions, these classes of food crops have been the basis of the human diet throughout recorded history, and this is still true today. As shown in Table 1, legume cultivation and consumption are widespread throughout many countries and cultures, suggesting that legumes might have evolved to fulfill human needs for proteins and other nutrients, complementing cereals.

North America

Common bean, lima bean, soybean, peanut, pea, and lentil are the main legumes used in this region. Common bean and lima bean are typically processed commercially and canned or cooked at home. Peanut is used roasted as whole seeds or as ingredients in various products such as peanut butter. Pea is sold commercially as green immature seeds in frozen form or as dry mature seed. Lentil is usually used as dry seeds and incorporated into a number of dishes. Soybean is a major cash crop in the USA, where the soymeal that remains after oil extraction is used mainly for animal feed. The USA and Brazil are the biggest producers and exporters of soybean. Soybean in the USA is a major source of cooking oil, protein concentrate, and protein isolates, which find uses in many different products including meats. The market for soy products is increasing at a rapid rate owing to the health benefits attributed to consumption of soy products such as reduction of cardiovascular disease and prevention of cancer. Tofu (soybean curd), soy sauce, miso, and nato are consumed by an increasing number of the US population in addition to the Asian American population already consuming them in large quantities.

The kidney bean is the principal legume of Mexico, occurring in several varieties. The dietary patterns of the Native American population in Arizona and New Mexico are similar to those of Mexico and Central America. The average daily intake of the Pima Indians in Arizona is 70 g per person, about four times the average US daily intake, providing a high percentage of the total protein. Kidney bean is again the principal legume, but other beans such as lima are also eaten.

Latin America

The common bean is the major legume in this region, making up about 90% of the total legume production, excluding soybean, which is produced in large quantities in Brazil as an export cash crop. In South and Central America, cowpea, pea, pigeon pea, and lupine are also grown. Common bean is typically used

Table 1 Cultivation of economically important legumes in different regions of the world

Common names	*Latin name*	*Areas of cultivation*
Peanut, groundnut, monkey nut, pigeon pea	*Arachis hypogaea*	Tropics and subtropics
	Cajanus cajan	Tropics and subtropics in Asia and Africa; widely cultivated in India and Pakistan; introduced into tropical America, Pacific Islands, and Australia
Chick pea, Bengal gram	*Cicer arietinum*	Tropics and subtropics in Asia and Africa; widely cultivated in India and Pakistan mainly as a cold weather crop; also cultivated in Mediterranean countries and South America
Soybean	*Glycine max*	South and East Asia, particularly in China and Japan (but not in India) and USA; other countries only on a limited, experimental scale
Lentil, split pea	*Lens esculenta*	Near east, North Africa, India, Burma, central and southern Europe
Lima bean, butter bean	*Phaseolus lunatus*	Tropical America, West Indies, tropical Africa, Madagascar, New Guinea, and north Australia
Golden gram, green gram	*Phaseolus aureus*	South and East Asia, particularly India, East Africa
Black gram, mung bean	*Phaseolus mungo*	India, Africa, West Indies, South and East Asia
Kidney bean, French bean, haricot bean	*Phaseolus vulgaris*	Throughout the world, particularly in the Americas
Pea, garden pea	*Pisum sativum*	Throughout temperate zone; in warm countries as a cold weather crop or at high altitudes
Broad bean, horse bean	*Vicia faba*	Temperate zone, particularly Mediterranean region, highlands in Asia, Africa, and Central and South Americas
Cowpea, blackeye pea	*Vigna unguiculata*	Tropical Asia and Africa, West Indies, southern Europe, and the Americas
Bambara, groundnut, earth pea	*Voandzeia subterranea*	Tropical Africa

as a dry bean, cooked at home, and included in a number of dishes.

Tropical Africa

In addition to the dry seeds, other forms of legumes are also consumed, such as the tender green shoots, leaves, and immature green pods. The three major legumes that are consumed locally are cowpea, common bean, and pigeon pea. Peanut is also grown but mostly exported.

Asia

Legumes are the only high-protein sources of the average diet in India, where they are eaten mainly with rice, wheat, and other minor cereals. Chickpea and pigeon pea make up 60% of the total legume production in India, followed by pea and khesari (*Lathyrus sativus*). The term 'gram' is applied to dry legume seeds with the husk, whereas the split, dehusked seeds are called 'dhal.' The most common way of preparing dhal is to cook until soft, after which they are mashed, mixed with water, and boiled again to give a soup or gruel of uniform consistency. These are eaten as part of meals composed of the staple cereal of the locality, which is usually boiled rice or chapatties made from wheat, sorghum, or millet.

Soybean, peanut, common bean, adzuki bean, black gram, horse gram, green gram, fava bean, and lima bean are grown and consumed to different degrees in China, Taiwan, Japan, Korea, Myanmar, Vietnam, Philippines, Malaysia, Indonesia, and Sri Lanka. The legumes are consumed as dry seeds, immature pods, sprouted and fermented. The winged bean (*Psophocarpus tetragonolobus* L) is mainly eaten as a green vegetable in the Philippines, Malaysia, Indonesia, Vietnam, and Sri Lanka. Because its chemical composition is similar to that of soybean, the winged bean has been promoted in its dry form as a rich source of protein and oil.

Soybean is the principal legume of East Asia (China, Japan, and Korea) and is also widely used in South-east Asia. It is consumed in forms that have been used for centuries such as green, dry seed, fermented, and germinated. Green immature soybean has an attractive flavor, and the pods are usually boiled in salted water, so that the beans can be easily removed. Immature beans are used in salads or cooked in vegetables dishes or in soups. The popularity of green soybean in the USA is increasing, together with other soy products.

Sprouted soybeans are among the most popular of Chinese vegetable dishes, which are either eaten raw as a salad or included in cooked dishes. Tofu is the most popular form of soy products in Asia. Typically, the whole beans are soaked overnight, ground into an emulsion, and strained to remove coarse particles, and the mash is boiled or steamed and passed through cloth filters. The curd is then precipitated from the milky fluid, allowed to settle, transferred to boxes lined with cloth filters, pressed with weights, washed, and dried to various extents. The final product, with a white color and delicate texture, contains 6–17% moisture and is sold in slices or slabs. The curd is included in a variety of dishes and is widely used as food for children.

A number of fermented soybean products are widely used in Asia. Soy sauce (shoyu) is produced in large quantities in Asia and the USA. It is made with a long, elaborate fermentation process with various fungi, bacteria, and yeast. Soybean paste (miso) is made from fermenting a mixture of soaked and cooked soybean plus steamed rice inoculated with *Aspergillus oryzae*. The final product, with a paste-like consistency and containing about 10% protein, is made into soups or served with rice and other foods as a side dish or condiment. Natto is a Japanese product made by fermenting whole soybeans through the action of *Bacillus subtilis* using an ancient method said to have been invented by Buddhist monks. The finished product, which is eaten with rice, has a musty flavor and gray color. A similar fermented product is hamanatto, which originally came from Korea. Here, the fermenting agent is a mold. Tempeh, an Indonesian product, is made by treating soybeans with the fungus *Rhizopus oryzae*. This highly nutritious product contains about 25% protein and is rich in amino acids produced by the fermenting fungus. (*See* **Fermented Foods**: Origins and Applications.)

Soy beverage is prepared as a white emulsion, which has the appearance of milk. Typically, the soybeans are soaked and passed through a mill, emerging as an emulsion. The emulsion is filtered, transferred to a boiler, and boiled for 20 min after adding supplements such as vitamins, sugar, salt, calcium, and flavorings. The boiled mixture is finally passed through a special emulsifier. The milk contains about 3.2% protein, 3.4% fat, and 4.5% carbohydrates, with contents of vitamins and calcium according to the amounts added. In Asia, soy milk is used as a beverage for adults but not for infants. Various proprietary formulations of soy milk are produced in the USA mainly for use in feeding infants and young children allergic to cows' milk.

Middle East, North Africa, and Mediterranean Countries

Populations in these regions consume legumes such as fava bean, pea, lentil, lupine, and some species of *Lathyrus*. In Africa, groundnut, mainly cultivated as

an oil-yielding cash crop, is also used directly as human food. Cowpea is the next most widely used legume in Africa. The common beans, particularly kidney beans, are also grown extensively in hilly regions with a good rainfall. Stewed broad beans are commonly eaten as the first meal of the day in the Middle East.

Europe

Pea and lentil are produced and used extensively in European countries. Common bean is also used in all of Europe but is usually imported from Canada and the USA. Lentil is also often imported.

Macronutrients

The energy value and proximate composition of 10 species of legume seeds are shown in Table 2. The largest component by weight in legume is carbohydrate averaging 53%. The second largest component in legumes is protein, varying from 19.5 to 39.7% by weight and contributing significantly to the seed energy value. Lipids comprise on average 7% of the seed weight. Most legumes are low in lipids, with the exception of peanut and soybean, which are low in carbohydrate but high in protein and lipids.

Protein

Methionine is the first limiting essential amino acid in legumes because the major storage proteins, the globulins, are low in this amino acid (Figure 1). Cysteine, although not an essential amino acid, is included with methionine, because it has a sparing effect on methionine when added to the diet. This limitation, together with its high lysine content, makes legumes nutritionally complementary with cereals, which is first limiting in lysine and relatively high in methionine. This makes the overall protein quality of cereal–legume mixtures better than that of either protein source alone. (*See* **Protein**: Food Sources.)

Supplementation of legumes with free methionine improves their value in supporting animal growth and increases the efficiency of utilization of dietary protein. In the feeds of poultry and swine, where soybean is the main protein source, methionine is used as a supplement. Little practical use of supplementation in human diets has been made, despite clinical studies with humans demonstrating their effectiveness. (*See* **Amino Acids**: Metabolism.)

Bioengineering (genetic engineering) is another strategy to enhance the protein quality of legumes. The sulfur amino acid requirements of humans vary with age, with infants requiring the most, adults requiring the least, and growing pigs requiring more than human infants (Figure 1). The FAO recommends using the essential amino acid requirement pattern of 2–5-year-old children to evaluate protein quality for all ages, except infants. The methionine + cysteine content of legumes has to be increased to approximately 25 mg per gram of protein for humans except infants, to 42 mg per gram of protein for human infants and to about 48 mg per gram of protein for swine. The digestibility of legume protein, which varies from 72 to 98%, also needs to be considered.

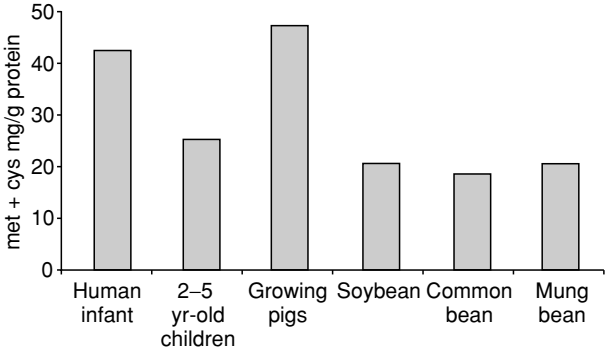

Figure 1 Methionine and cysteine contents of some legumes compared with methionine and cysteine requirement for human infants, 2–5 year-old children and growing pigs.

Table 2 Macronutrients contents per 100 g of edible legume seeds

Common names	*Energy (kJ)*	*Water (g)*	*Protein (g)*	*Lipid (g)*	*Crude fiber (g)*	*Ash (g)*	*Carbohydrate (g)*
Soybean	1735	9.0	34.3	18.7	3.8	5.1	31.6
Peanut	2306	6.5	27.6	52.1	2.0	2.4	13.3
Pea	1453	10.5	24.5	1.1	4.0	2.5	62.0
Lentil	1453	10.5	27.7	1.0	4.1	2.6	61.2
Cowpea	1453	11.7	22.0	1.3	4.5	3.3	63.4
Chickpea	1546	10.7	19.5	5.7	4.0	2.7	61.7
Pigeon pea	1449	9.9	19.5	1.3	1.3	3.9	65.5
Fava bean	1470	10.6	24.8	1.4	7.0	3.3	60.4
Lupine	1596	10.7	39.7	9.0	12.5	3.5	35.0
Winged bean	1701	9.7	32.8	17.0		4.1	36.5
Mean	1537	10.0	25.8	7.0	6.0	3.4	52.3

Taking soybean as an example with an average methionine+cysteine content of 22 mg per gram of protein, this value has to be increased by 14% (for humans), by 91% (for infants), and by 118% (for swine). Figures would have to be higher if protein digestibility is taken into consideration and for other beans (Figure 1), whose methionine+cysteine contents are generally lower than those of soybean. In any case, animal feeding trials need to be carried out to establish the actual nutritional value of transgenic seeds.

Soybean is a primary target for genetic engineering, since it is the single largest source of protein in animal feed. In feed formulations where corn is added to soybean meal to provide methionine, increasing the methionine content of corn would be desirable, although its protein quality would not be significantly improved. Other legumes that are lower in methionine content than soybean should also be targeted.

There are basically four types of bioengineering strategies to increase the methionine content in legumes: (1) introduce methionine residues or methionine-rich peptides into nonconserved, and presumably noncritical, regions of storage protein; (2) transfer genes coding for methionine-rich protein (MRP) from other species, i.e., heterologous genes into legume of interest; (3) increase the level of nonabundant, endogenous MRP in the target plant; and (4) manipulate the key enzymes in the biosynthetic pathways of essential amino acids. The transfer of heterologous MRP genes into legumes is the most widely used and has been successful in increasing the methionine content of soybean by 26% and lupine by 94%. The transgenic soybean containing the Brazil nut MRP is allergenic to individuals who are allergic to Brazil nut, demonstrating that the Brazil nut MRP is a major allergen in Brazil nut and that allergens can be transferred from one food source to another through bioengineering. In feeding trials with rats, the transgenic lupine seeds have statistically significant increases in live weight gain, true protein digestibility, biological value, and net protein utilization compared with wild-type seeds. The strategy of cloning nonabundant, MRP in soybean led to the discovery of lunasin, a unique soybean peptide that has been shown to have anticancer properties (see section Bioactive Phytochemicals).

Digestibility of Legume Proteins

While the amino acid pattern of legumes is the most important index of protein quality, digestibility and bioavailability of the constituent amino acids are the next important factors. Differences in protein digestibility may be due to the nature of the protein, the presence of nonprotein components (dietary fiber, tannins, phytates), the presence of antinutritional factors, and processing conditions. The true digestibility of legumes varies from 73% (canned pinto beans) to 98% (soy protein isolate). Recognizing the importance of digestibility in evaluating protein quality, the recommended method is to combine amino acid score with digestibility data in what is called protein digestibility-corrected amino acid score. (*See* **Protein:** Quality.)

Carbohydrates

The total carbohydrate content of legumes is typically 60%, and these carbohydrates are generally well absorbed and utilized. Starch accounts for the major portion of legume carbohydrates, ranging from 24 to 56%. The oilseeds have lower carbohydrates content, with values such as 14% for peanuts and 32% for soybean. Oligosaccharides of the raffinose family (raffinose, stachyose, verbascose, and ajucose) account for 31–76% of the total sugars. The resistance of the raffinose oligosaccharides to human and monogastric animal digestion presents a problem to certain individuals sensitive to flatulence. It is estimated that reducing the raffinose oligosaccharide content of soybean used for livestock feed by 1% would result in an additional total value of $101 million per year (1990 value) owing to improved nutrition and reduced flatulence. (*See* **Carbohydrates:** Classification and Properties.)

Lipids

The lipid content of legumes is generally 7%. Exceptions are the oilseeds peanut and soybean, which contain about 52 and 20% oil, respectively. Soybean oil is the most widely used cooking oil. Winged bean is potentially a rich source of oil because of its similarity with soybean in chemical composition, but has not been exploited commercially. Peanut and soybean oils provide essential fatty acids to the human diet. Peanut contains 22% linoleic, whereas soybean has 54% linoleic and 8% linolenic. (*See* **Fats:** Classification.)

Micronutrients

Selected mineral and vitamin contents of 10 species of legume seeds are shown in Table 3.

Vitamins

Legume seeds are excellent sources of folate, which, in addition to being an essential nutrient, is thought to reduce the risk of neural tube defects. One serving of beans provides more than half of the current recommended dietary allowance (RDA) for folate. Beans are also good sources of thiamin and pantothenic

Table 3 Micronutrients contents per 100 g of edible legume seeds

Common names	Ca (mg)	Fe (mg)	Zn (mg)	Vitamin A (IU)	Thiamin (mg)	Folate (μg)	Vitamin C (mg)
Soybean	277	15.7	4.89	24	0.874	375.1	6
Peanut	92	4.58	3.27	0	0.64	239.8	0
Pea	43	2.08	0.27	145	0.15	41.7	60
Lentil	51	9.02	3.61	39	0.475	432.8	6.2
Cowpea	110	8.27	3.37	50	0.853	632.6	1.5
Chickpea	105	6.24	3.43	67	0.477	556.6	4
Pigeon pea	130	5.23	2.76	28	0.643	456	0
Fava bean	37	1.55	1.00	333	0.133	148	3.7
Lupine	176	4.36	4.75	23	0.64	355	4.8
Winged bean	440	13.44	4.48	0	1.03	44.6	0

acids. On average, 100 g of legume seeds provides 23% of the nicotinic acid, 50% of the thiamin, 15% of the riboflavin, 20% of the vitamin B_6, 195% of the folate, and 30% of the pantothenic acid requirement of an adult (based on US RDAs). Among the legumes, peanut is particularly rich in B vitamins. In general, legume seeds are poor sources of fat-soluble vitamins and vitamin C, but the leaves and immature seeds of some legumes contain high amounts of carotene, vitamin C, and vitamin E. For example, immature soybean seed is a good source of vitamin C, whereas mature soybean seed contains virtually no vitamin C. (*See* **Vitamins**: Overview.)

Minerals

Legumes are high in iron, with one serving of legume providing around 2 mg of iron. This compares favorably with the iron RDAs of 10 and 15 mg for adult men and premonopausal women, respectively. However, iron bioavailability from legumes is poor, and thus their value as a source of iron is diminished. In contrast, zinc bioavailability from legumes is relatively good at around 25%. Also, many beans are good sources of calcium, providing on average 50 mg of Ca per serving. Calcium bioavailability from soyfoods is essentially equivalent to that from milk, but the balance of calcium to phosphorus is inadequate, since the ratio is 1:2, whereas the ideal is 2:1. Legume seeds contain very low amounts of sodium. (*See* **Minerals – Dietary Importance**.)

Bioactive Phytochemicals

The correlation between diets rich in soybean and reduced incidence of certain types of cancer and cardiovascular disease has led to studies on identifying the responsible substances.

Effects on Heart Disease

Animal studies have shown clearly that soy protein lowers serum cholesterol and triglycerides, but the results in humans have not been consistent. However, a metaanalysis of a number of human studies using strict criteria, concluded that soy protein reduces serum cholesterol and triglycerides. The US FDA approved a health claim that diets low in saturated fat and cholesterol that include 25 g of soy protein may reduce the risk of heart disease. The mechanism of action remains to be elucidated.

Effects on Cancer

A number of soybean compounds are being studied for their anticancer properties. These include phytosterols, phytates, saponins, protease inhibitors, isoflavones, and a recently discovered unique peptide termed lunasin.

Phytic acid may play a role in reducing cancer risk, possibly because of its antioxidant effects. Specifically, it has been suggested that phytic acid may lower the risk of colon cancer and perhaps breast cancer.

Saponins have been shown to have anticancer properties, as suggested by a recent rodent study that a saponin-containing diet inhibited by about two-thirds the development of preneoplastic lesions in the colon. However, given that the human intake of saponins is much less than the experimental amount, it is not clear to what extent these results in rodents are relevant to humans.

Bowman–Birk inhibitor (BBI) is a protease inhibitor found in soybean and many other types of legumes. BBI in the form of BBI concentrate (BBIC) has shown the greatest suppression of carcinogenesis in animal models so far. BBIC is now being tested in human clinical trials. (*See* **Trypsin Inhibitors**.)

Isoflavones, which are found in plants but are especially high in soybean, belong to the class of phytoestrogens. Isoflavones offer potential alternative therapies for a range of hormone-dependent conditions that include cancer, menopausal symptoms, cardiovascular disease, and osteoporosis. Epidemiological evidence and animal experiments strongly suggest beneficial effects of isoflavones on human health. Large-scale human clinical trials are needed to confirm limited but promising pilot studies.

Lunasin is a 43-amino acid peptide with a polyaspartic acid carboxyl end, a cell adhesion motif (-RGD- or arg–gly–asp), and a predicted conserved helical region. The peptide was isolated and sequenced 14 years ago, and a few biological effects were proposed, but none was proven until the lunasin gene was isolated. Transfection of the *lunasin* gene into mammalian cells leads to the arrest of cell division, enlargement of the cells, and eventually cell death, resulting in cell lysis and chromosome fragmentation. The expressed lunasin peptide binds preferentially to chromatin. This suggests the use of the *lunasin* gene as a chemotherapeutic agent in conjunction with a specific delivery system that targets cancer cells. While the *lunasin* gene is potentially a chemotherapeutic agent, the lunasin peptide has been shown in *in vitro* experiments using mammalian cells to suppress transformation of normal cells to cancer cells caused by chemical carcinogens and oncogenes. The first *in vivo* experiment using a mouse skin cancer model shows that topical application of lunasin peptide suppresses skin tumorigenesis.

Anti-nutritional Factors

Amylase and Protease Inhibitors

Protease inhibitors from legumes can interfere with protein digestion, leading to growth retardation and, in some species of animals, can cause pancreatic enlargement and enhance chemically induced pancreatic tumors. Two types of trypsin and chymotrypsin inhibitors have been identified in soybean: Kunitz trypsin inhibitor and BBI. Boiling dry beans generally reduces the protease inhibitor content by 80–90%, and there is little reason to think that the amount of trypsin inhibitors obtained by eating commonly consumed beans would exert any adverse effects in humans. (*See* **Plant Antinutritional Factors:** Characteristics.)

Lectins

Lectins, also known as hemagglutinins, are found in many legume seeds. Lectins are toxic to animals by causing lesions in the intestinal mucosa, malabsorption of nutrients, weight loss, and eventually death. Symptoms of lectin toxicity in humans include nausea, vomiting, diarrhea, and abdominal pain. Proper cooking of beans until a soft texture is achieved eliminates lectin toxicity. (*See* **Hemagglutinins (Haemagglutinins)**.)

Goitrogens

Goitrogenic compounds are found in soybean and peanut. Consumption of unheated soybean or peanut can lead to enlargement of the thyroid gland in rat and chicken. Iodine administration and heat treatment are common ways of eliminating goitrogens.

Cyanogenic Glycosides

Cyanide poisoning from lima beans in humans has been reported. However, once lima beans are ground and cooked, the enzymes responsible for the liberation of the cyanogenic compounds are generally denatured.

See also: **Bioavailability of Nutrients**; **Cancer**: Diet in Cancer Prevention; Diet in Cancer Treatment; **Coronary Heart Disease**: Prevention; **Ethnic Foods**; **Fermented Foods**: Soy (Soya) Sauce; **Peas and Lentils**; **Plant Antinutritional Factors**: Characteristics; **Protein**: Quality; **Soy (Soya) Beans**: Dietary Importance; **Toxins in Food – Naturally Occurring**

Further Reading

Galvez AF and de Lumen BO (1999) A soybean cDNA encoding a chromatin-binding peptide inhibits mitosis of mammalian cells. *Nature Biotechnology* 17: 495–500.

Galvez AF, Chen N, Macasieb J and de Lumen BO (2001) Chempreventive property of a soybean peptide that binds to deacetylated histones and inhibits histone acetylation. *Cancer Research* 61: 7473–7478.

Kelloff GJ, Boone CW, Crowell JA *et al.* (1996) New agents for cancer chemoprevention. *Journal of Cellular Biochemistry* 26S: 1–28.

Kelloff GJ, Sigman CC and Greenwald P (1996) Cancer chemoprevention: Progress and promise. *European Journal of Cancer* 35: 1755–1766.

Kennedy AR (1995) The evidence for soybean products as cancer preventive agents. *Journal of Nutrition* 125: 733S–743S.

Kennedy AR (1998a) Chemopreventive agents: protease inhibitors. *Pharmacology and Therapeutics* 78: 167–209.

Kennedy AR (1998b) The Bowman–Birk inhibitor from soybean as an anticarcinogenic agent. *American Journal of Clinical Nutrition* 68: 1406S–1412S.

Messina MJ (1999) Legumes and soybeans: overview of their nutritional profiles and health effects. *American Journal of Clinical Nutrition* 70 (Supplement): 439S–450S.

Messina M, Persky V, Setchell KDR and Barnes S (1994) Soy intake and cancer risk: A review of *in vitro* and *in vivo* data. *Nutrition and Cancer* 21: 113–131.

Moyad MA (1999) Soy, disease prevention, and prostate cancer. *Seminars in Urology and Oncology* 17(2): 97–102.

Setchell KDR and Cassidy A (1999) Dietary isoflavones: Biological effects and relevance to human health. *Journal of Nutrition* 129: 758–767.

Lemons *See* **Citrus Fruits**: Types on the Market; Composition and Characterization; Oranges; Processed and Derived Products of Oranges; Lemons; Grapefruits; Limes

Lentils *See* **Peas and Lentils**

Lettuce *See* **Salad Crops**: Leaf-types; Other Types of Salad Crops; Root, Bulb, and Tuber Crops

Leukotrienes *See* **Prostaglandins and Leukotrienes**

Levulose (Laevulose) *See* **Fructose**

LIGNIN

M T Holtzapple, Texas A&M University, College Station, TX, USA

Introduction

Lignin is the world's most abundant noncarbohydrate biological material. An estimated 35 000 million tonnes is produced annually, a daily production of about 16 kg per person.

Lignin is derived from the Latin *lignum*, meaning wood. It is an important component of wood and other higher plants, serving as nature's 'glue.' Not only does it hold plant cell walls together, its hydrophobic properties prevent water loss from plant vascular systems. Lignin is highly resistant to enzymatic degradation, so it protects plants from invading insects and microbial attack.

Because lignin is nondigestible, it is an important component of dietary fiber. Lignin resists decay and thus forms a large fraction of soil humus. When subjected to geological forces, lignin and other plant components are transformed into lignite and, ultimately, bituminous and anthracite coals. (*See* **Dietary Fiber:** Properties and Sources.)

Pure lignin does not occur in nature because it always accompanies plant polysaccharides as a composite material called 'lignocellulose.' Lignin removal is required to manufacture paper, so tree lignin – particularly softwood lignin – is well studied. (*See* **Carbohydrates:** Classification and Properties; **Cellulose**; **Hemicelluloses**.)

Structure

Monomers

The three phenylpropylene lignin monomers – *trans*-coniferyl alcohol, *trans*-sinapyl alcohol and *trans-p*-coumaryl alcohol – are shown in Figure 1. The monomers are essentially the same, differing only in the number of methoxyl groups. In softwoods, coniferyl alcohol is the dominant monomer (94% coniferyl, 1% sinapyl, 5% coumaryl alcohols). In hardwoods, both coniferyl and sinapyl alcohols are

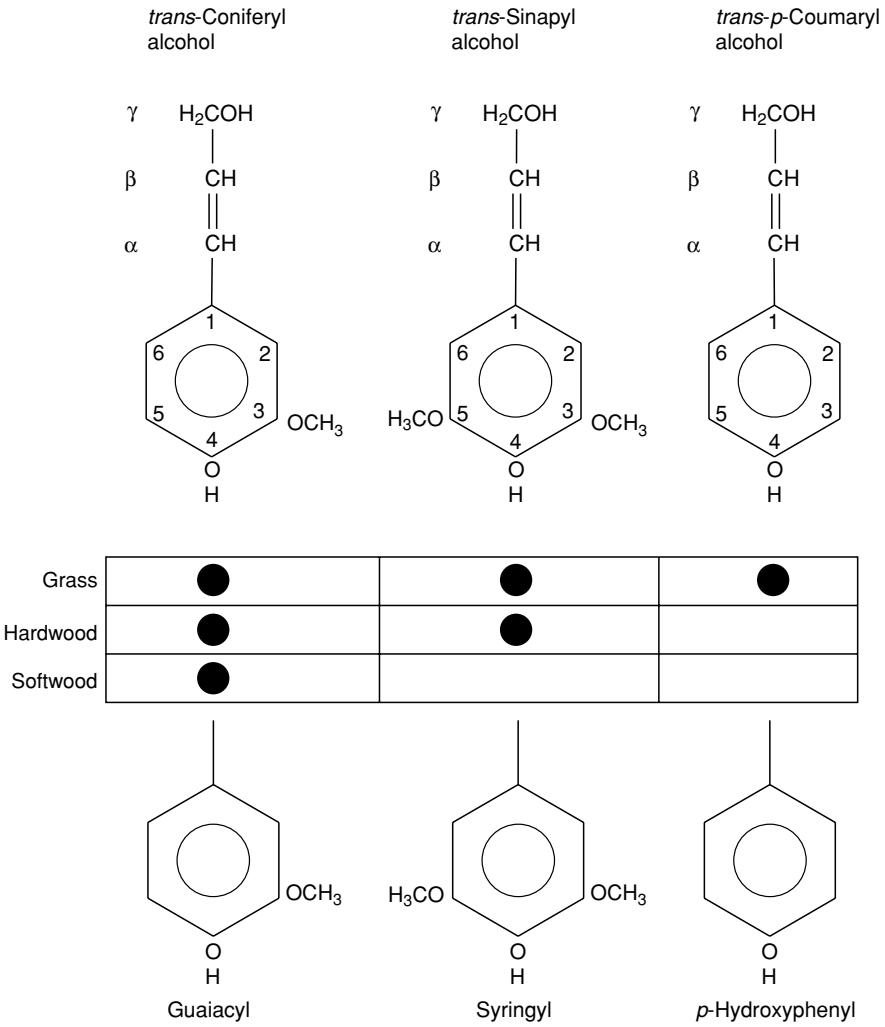

Figure 1 Lignin monomers. Reproduced from Lignin. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

present in significant amounts, whereas grasses have large quantities of all three phenylpropylenes. The guaiacyl, syringyl and *p*-hydroxy-phenyl aromatic groups of the three monomers are also described in Figure 1. The aromatic carbons are numbered 1 through 6, whereas the propylene carbons are lettered α, β, and γ, starting from the phenyl group. Plants produce these phenylpropylene monomers by deaminating tyrosine and phenylalanine, two aromatic amino acids. (*See* **Amino Acids**: Properties and Occurrence.)

Polymer Structure

The phenylpropylene monomers are extremely reactive because they contain two hydroxyl groups and one double bond. Plant enzymes oxidatively dehydrogenate the phenylpropylene, causing it to polymerize randomly into the three-dimensional structure schematically shown in Figure 2. During polymerization, the reactive double bonds are lost, so phenylpropane is the repeating unit. The polymer contains the ether and carbon–carbon linkages shown in Table 1. (Although this structure has been determined for softwood lignin, it is probably representative of other lignins as well.) The β-O-4 ether linkage is the most common, so it is assigned a relative frequency of 1.00, whereas other linkages have a lower relative frequency. Because of the reactive double bond, the α and β carbons are the primary linkage sites between monomers. The γ carbon is rarely linked, except in the circular (or multiple) bonds. About 70% of the phenolic oxygens are involved in ether linkages, whereas the remaining 30% are free. (*See* **Phenolic Compounds**.)

The propane unit contains a variety of functionalities, including hydroxyl groups, ketones, aldehydes, double bonds, and ether linkages. The β carbon is frequently involved in an ether linkage. The adjacent

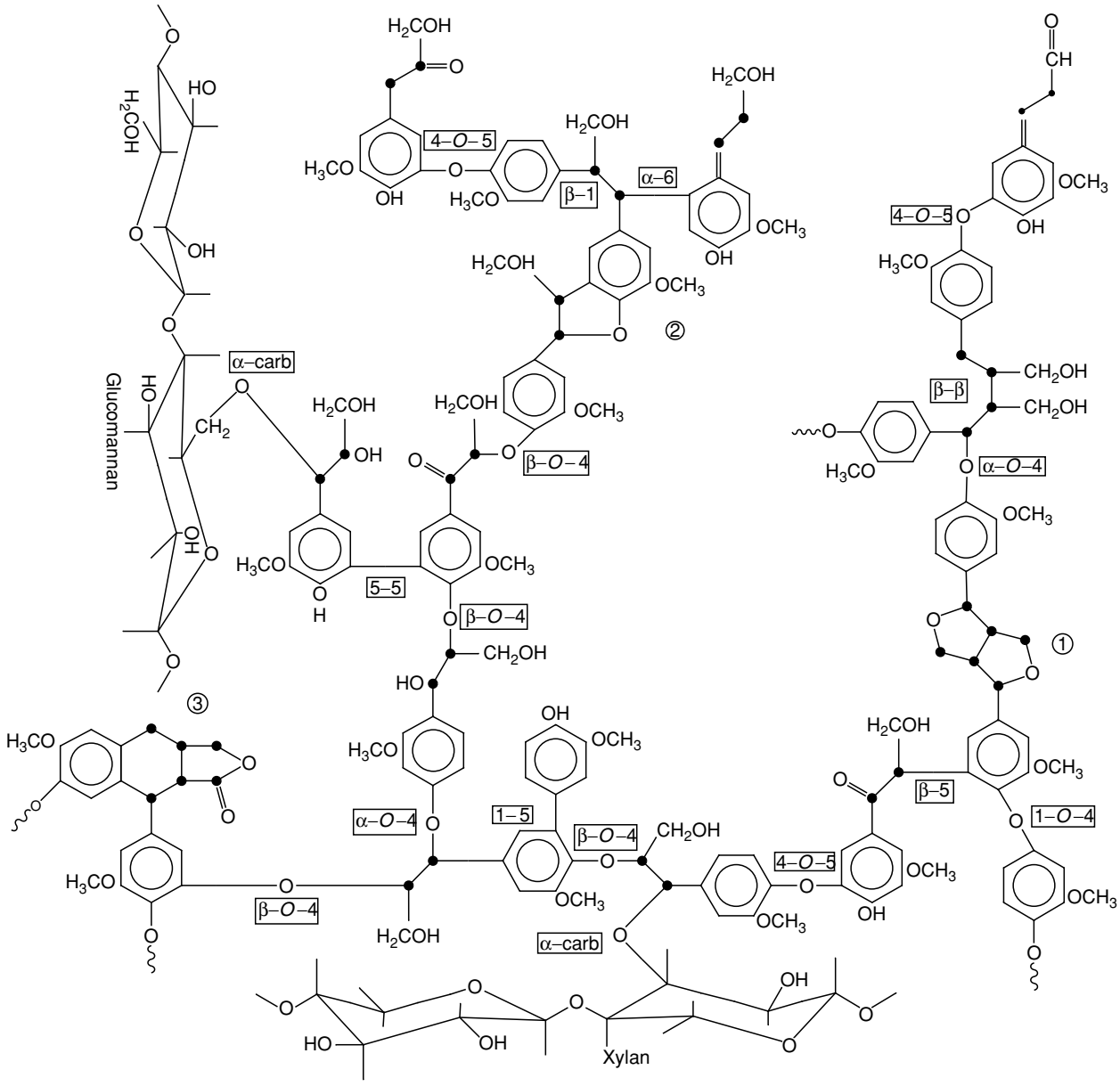

Figure 2 Schematic of softwood lignin structure. Reproduced from Lignin. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 1 Lignin bonds

Ether	*Relative frequency*	*Carbon–carbon*	*Relative frequency*	*Circular*	*Relative frequency*
β-O-4	1·00	*β–β*	0·38	① *β–β*, *α-O-γ*, *α-O-γ*	0·29
α-O-4	0·18	*β–1*	0·33	② *α-O-4*, *β–5*	0·22
α-carb	—	*β–5*	0·31	③ α–6, *β–β*, *γ-O-γ*	—
4-*O*-5	—	5–5	*sim*0·27		
1-*O*-4	—	α–6	—		
		1–5	—		

α carbon generally has a hydroxyl group, or another ether linkage. The γ carbon normally retains its hydroxyl functionality, although it is sometimes converted to an aldehyde.

The phenylpropylene monomers not only cross-link with other phenylpropylenes, but also covalently bond to hemicellulose, primarily through the α carbon. The exact nature of the carbohydrate bond

is still being studied. It is thought that carbohydrate linkages infrequently occur as periodic 'spot welds' among the intertwined hemicellulose/lignin matrix.

Softwood lignin has a repeating nine-carbon phenylpropane subunit with an approximate formula of $C_9H_{7.95}O_{2.4}$ $(OCH_3)_{0.92}$ and a molecular weight of about 184. The polymer is believed to have an approximate molecular weight of over 10 000, which corresponds to a degree of polymerization (DP) of 54. It is difficult to know accurately the molecular weight of native lignin because it is impossible to isolate it in an unaltered state.

Hardwood lignin has a repeating nine-carbon phenylpropane subunit with an approximate formula of $C_9H_{7.49}O_{2.53}(OCH_3)_{1.39}$ and a molecular weight of about 200. It has more methoxyl groups than softwood lignin because sinapyl alcohol is a more prevalent monomer in hardwoods. The polymer probably has a molecular weight under 5000 (DP about 25).

Cellular Structure

Plant cell walls are composed of microfibrils deposited in lamellar layers spaced approximately 7–10 nm apart (Figure 3). According to Fengel's model of the microfibril, four 4×4 arrays of elementary fibrils are embedded in a hemicellulose matrix. The elementary fibrils are composed of cellulose which is largely crystalline, except for the paracrystalline edges and amorphous regions interspersed along the length. These rigid cellulose elementary fibrils are primarily responsible for structural load bearing. The hemicellulose is tightly associated with the cellulose, particularly in the paracrystalline and amorphous regions. Hemicellulose is noncrystalline, but is largely oriented in the same direction as the cellulose fibers. The microfibril is encrusted with a lignin layer about 2 nm thick spanned by about two to four phenylpropane units. Although the lignin may be partially oriented, it is mainly amorphous. In effect, the lignin acts as a 'net' holding the fibrous structure together. This net-like form of lignin is dispersed throughout the secondary plant cell wall and accounts for about 20% of its mass.

In mature plant cells, the interstitial space between the cells (i.e., middle lamella) is highly lignified. The lignin layer is about 100 nm thick and forms a completely amorphous three-dimensional array. Although the lignin concentration is very high in the middle lamella, it is very thin and accounts for only about 20% of the total lignin.

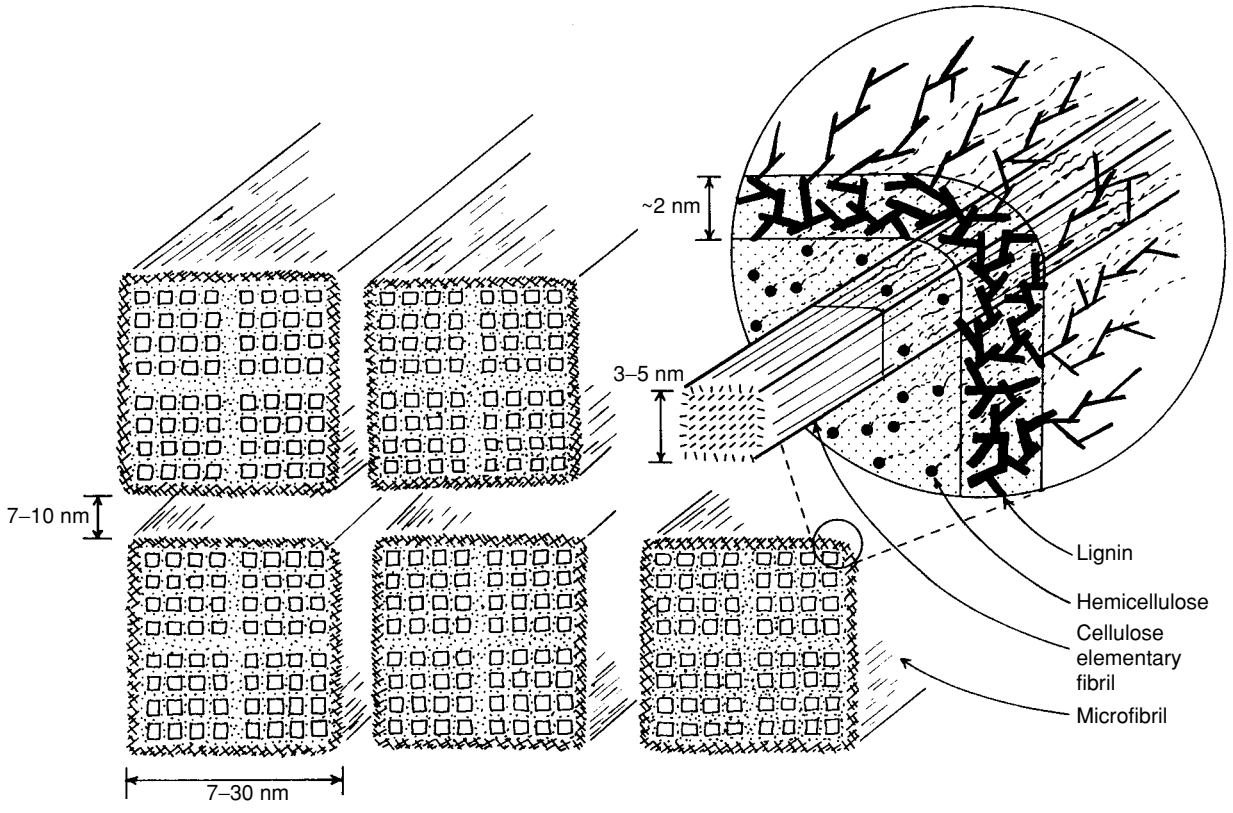

Figure 3 Lignin, cellulose, and hemicellulose composite structure. Reproduced from Lignin. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

The lignification process occurs late in plant cell life after cell growth has ceased. The lignin is first deposited within the primary wall at the cell corners. Then lignification extends along the middle lamella (completely covering the cell exterior) and into the thick secondary wall. In oats, immature plants have a very low lignin content (2% tops, 6% roots) during their first 42 days. Then, they enter a period of rapid lignification for about 20 days in which the lignin content dramatically increases (8% tops, 10% roots). At the end of 112 days, the lignin content reaches its highest level (11% tops, 12% roots).

Properties

Physical

Lignin is a noncrystalline solid with a density of about 1.3–1.4 g cm^{-3} and a refractive index of 1.6. Commercially available lignins, isolated from paper pulp manufacture, are brown powders. The brown color probably comes from degradation reactions because native lignin in some woods is very light in color.

When heated, lignin does not melt, but softens at its glass transition temperature and transforms from a glassy solid to a rubbery plastic. Higher glass transition temperatures (about 193 °C dry; 128 °C moist) occur with high-molecular-weight (85 000) lignin, whereas lower glass transition temperatures (about 127 °C dry; 72 °C moist) occur with low-molecular-weight (4300) lignin. Minor lignin components begin to decompose at about 200 °C, with major decomposition occurring at about 300–400 °C. The heat of combustion is about 29.5 MJ kg^{-1}, which is 67% greater than that of cellulose and hemicellulose. In softwoods, lignin is only about 28% of the mass, but produces approximately 40% of the combustion heat.

No known solvents are capable of dissolving significant amounts of lignin from intact plants as it is so intimately associated with other cell wall components. However, lignin isolated by extensive ball milling is soluble in a number of solvents, including pyridine, acetone, dioxane/water (9:1, v/v), 1,2-dichloroethane/ethanol (2:1, v/v), 50% aqueous sodium thiocyanate, and benzyl alcohol/dimethyl formamide. Lignin which has been subjected to acidic treatment conditions is soluble in aqueous sodium hydroxide.

Chemical

Lignin bonds can be broken by a number of chemical reagents, including acids, bases, halogens, nitrates, oxidants, and reductants.

Acidic conditions are often employed in wood pulping. Even without acid addition, steaming spontaneously generates mild acidic conditions (pH 3.6–4.0), primarily from acetic acid released from hemicellulose acetyl groups. α Ether linkages are more labile than β ether linkages. In acid, not only are bonds broken, allowing the lignin to be solubilized, but new bonds are also formed in undesirable condensation reactions. A classic laboratory method for identifying lignin is to reflux plants with ethanol/hydrochloric acid and assay for the degradation products known as 'Hibbert's ketones.'

Basic conditions are also employed in some pulping reactions. Ether linkages are broken with the simultaneous formation of phenolic hydroxyl groups. In addition, new bonds are again formed in undesirable condensation reactions. Lignin methoxyl groups generally resist alkaline cleavage at temperatures below 200 °C.

Oxidation of lignin readily occurs using peracetic acid, chlorine, nitric acid, chlorine dioxide, sodium hypochlorite, sodium chlorite, hydrogen peroxide, and ozone. These oxidants selectively attack lignin (primarily the aromatic ring) while leaving carbohydrates largely unaffected. Thus, they are useful as bleaching agents in the production of high-quality pulp, or in the laboratory preparation of holocellulose (i.e., cellulose + hemicellulose). Hypochlorite oxidation is thought to be initiated at a free phenolic hydroxyl group and adjacent molecules are then attacked in a 'peeling' reaction analogous to alkaline degradation of cellulose and hemicellulose. Biological enzymatic cleavage of lignin occurs by oxidation. Irradiation at short wavelengths (< 385 nm) photooxidizes lignin, resulting in darkening, a phenomenon readily exhibited when a newspaper is left in the sun.

Reduction of lignin occurs with hydrogen gas using Raney nickel catalyst in a mildly acidic or basic aqueous solvent at 160–170 °C. The product is 'hydrol lignin,' a mixture of low-molecular-weight lignin fragments. The yield of hydrol lignin is higher under alkaline conditions.

Occurrence

The lignin content in vegetables and fruits is low (generally < 5%), in agricultural residues it is medium (7–13%), in hardwoods it is high (16–24%), and in softwood it is extremely high (26–30%). 'Compression wood' (i.e., from the underside of a horizontal branch) is higher in lignin, whereas 'tension wood' (i.e., from the topside) is lower in lignin.

Chemical Pulping

Chemical pulping removes lignin from lignocellulose. The resulting cellulose fibers may be used for paper production, or 'chemical cellulose' for manufacturing

cellulose derivatives such as carboxymethyl cellulose. Pulping conditions tend to be so harsh that hemicellulose is hydrolyzed and removed from the fiber. Softwoods are generally preferred for paper manufacturing because of their long fibers, but hardwoods and agricultural residues (e.g., bagasse) may also be pulped. Pulping necessarily requires that the lignin be cleaved into smaller fragments which may be washed away from the fiber. Because lignin is insoluble in water it must be converted into water-soluble derivatives. Advanced organosolv pulping uses an organic solvent in which the lignin is soluble. Pulping is unable to remove all lignin, so high-quality pulp is subsequently bleached with oxidants to remove residual lignin and color. The two major chemical pulping technologies are the kraft and sulfite processes, with kraft dominating about 96% of the market.

Kraft pulping is an alkaline process wherein lignin is removed by the action of sodium hydroxide and sodium sulfide. The pulping digester operates at about 175 °C for 2–5 h. Kraft pulping has replaced the older soda process, in which sodium hydroxide was the primary pulping chemical. The addition of sodium sulfide in kraft pulping promotes ether cleavage and inhibits undesirable condensation reactions. The simplified chemistry can be represented as shown in eqn (1).

$$R_{lig}-O-R_{lig} + H_2O \xrightarrow[NaOH/Na_2S]{} 2\ R_{lig}-OH$$

$$R_{lig}-OH + NaOH \longrightarrow \underset{\text{Sodium lignate}}{R_{lig}-O^- Na^+} + H_2O \qquad (1)$$

Some sulfur is also incorporated into the sodium lignate. The number-average molecular weight is about 1600 (DP about 8) for softwood sodium lignates and 1050 (DP about 5) for hardwood sodium lignate. The solubilized sodium lignate and spent chemicals are present in aqueous solution as 'black liquor.' The black liquor is concentrated and burned to regenerate the pulping chemicals. Because unrecovered sodium and sulfur are replenished by adding sodium sulfate, the process is often called 'sulfate pulping.' Sulfate is not the active pulping chemical, however, because organic carbon reduces it to sodium sulfide when the black liquor is burned. The kraft process produces high yields of strong fibers, hence the name 'kraft,' meaning strong in German.

Sulfite pulping is an acidic process that produces lower yields than the kraft process, and the fibers are also weaker. Its advantage is that a greater percentage of lignin is removed, making the resulting fibers more suitable for high-quality paper and 'chemical cellulose.' The wood is treated with magnesium bisulfite and excess sulfur dioxide for 6–12 h at about 175 °C. The simplified chemistry can be represented as shown in eqn (2).

$$R_{lig}-O-R_{lig} + Mg(HSO_3)_2 \longrightarrow \underset{\text{Magnesium lignosulfonate}}{2\ R_{lig}SO_3^-\ Mg^{2+}} + H_2O \qquad (2)$$

The magnesium lignosulfonate number-average molecular weight is typically 10 000–40 000 (DP 45–180), although some fractions may be as high as 22 000 000 (DP 100 000). The magnesium lignosulfonate and spent chemicals are dissolved in an aqueous 'weak red liquor' which contains about 2–3% sugar from the degraded hemicellulose. These sugars may be fed to *Candida utilis* (also known as *Torulopsis utilis*) to make single-cell protein. The spent liquor is concentrated and burned to produce magnesium oxide and sulfur dioxide, which are reacted together to remake magnesium bisulfite.

Organosolv pulping heats lignocellulose with water/solvent/catalyst for about 2 h. Many solvents have been tried, including mono- and polyalcohols, aldehydes, ketones, thio compounds, organic acids and bases, and dimethyl sulfoxide. Triethylene glycol is a superior solvent, but its high boiling point (190 °C) makes it difficult to recover. Phenol and butanol are excellent solvents, but ethanol is preferred because of its low cost and low toxicity. Various catalysts have been employed, including hydrochloric acid, sulfuric acid, ferric chloride, ammonia, aluminum salts, and urea. If the temperature is raised above 180 °C, self-generated acids serve as a catalyst. The simplified reaction shown in eqn (3) occurs.

$$R_{lig}-O-R_{lig} + C_2H_5OH \xrightarrow{\text{Catalyst}} R_{lig}=O + HO + R_{lig} + R_{lig}-O-C_2H_5 \qquad (3)$$

Note that some ethanol solvent is incorporated into the solubilized lignin. Double-bonded carbons have also been detected in the propane side chain. The organosolv lignin number-average molecular weight is about 1000 (DP about 5). Of the three pulping technologies, the organosolv process produces lignins most similar to native lignin. Organosolv pulping has been practiced on an industrial scale in Canada.

Uses

In the USA, less than about 2% of the lignin produced by pulping is used for nonfuel applications. Thus, it represents an enormous, underutilized resource. Over 3700 literature references (mostly patents) have explored potential uses, as described in Table 2. The

Table 2 Potential uses for lignin

Use	*Kraft sodium lignate*	*Sulfite ligno-sulfonate*	*Organosolv lignin*
Polymer applications			
Phenol–formaldehyde resin extender for wood adhesives	+	+	+
Flame-retardant wood glue	+	+	+
Asphalt extender	+	+	+
Engineering plastics	+	−	+
Electrically conducting polymers	−	+	−
Elastic polyurethanes	+	−	+
Carbon black extender in rubber	+	−	+
Slow release of chemicals (herbicides, pesticides, pharmaceuticals)	+	+	+
Animal fodder pelletizing	−	+	−
Storage battery plate	+	+	−
Road dust suppression	−	+	−
Ore and mold binder	+	+	+
Cement additive	+	+	−
Liquid additive			
Dispersant	+	+	−
Surfactant	−	+	+
Tertiary oil recovery	−	+	−
Flocculant	+	−	−
Sequestering agent	+	+	−
Corrosion inhibitor	−	+	−
Diesel fuel additive	−	−	+
Drilling muds	+	+	−
Chemical production			
Activated carbon	+	−	−
Vanillin	−	+	−
Dimethyl sulfide	+	+	−
Low-molecular-weight chemicals (e.g., carbon monoxide, hydrogen, methane, acetylene, ethylene, phenol)	+	+	+

+, reference found; −, no reference found.

applications most relevant to food technology are binding agents in animal feed pellets and vanillin production.

Vanillin is one of the most popular food flavors. It is found throughout the plant world, but generally in low concentrations. The orchid vanilla bean has an unusually high concentration (< 5%) and serves as a natural source. Vanillin can also be manufactured from lignosulfonate by alkaline hydrolysis or oxidation. The yields are low (< 10%), so other uses must be found for the unconverted lignin. (*See* **Flavor (Flavour) Compounds**: Production Methods.)

Dietary fiber contains lignin because it is very nondigestible. Lignin adsorbs bile salts and reduces their intestinal uptake, thus possibly lowering blood cholesterol levels.

Lignin Isolation

Milled wood lignin is isolated from plants by breaking the lignin–carbohydrate complex through extensive ball milling or vibratory milling for 2–28 days. The lignin is extracted by dioxane/water (9:1, v/v) with subsequent purifications in solvents. Approximately half the lignin can be isolated by this process. Its properties are very similar to native lignin.

Enzyme lignin is prepared by ball milling the sample for 5–8 h to expose the carbohydrates to enzymatic attack by extracellular enzyme preparations, such as cellulase/hemicellulase isolated from *Trichoderma reesei*. The residue consists primarily of lignin with some unreacted carbohydrates, which can be removed with further ball milling and water extraction. Because the enzymes selectively dissolve polysaccharides, the lignin is similar to native lignin.

Brauns lignin is prepared by extracting finely ground plants with 95% ethanol followed by subsequent solvent purification steps. The chemistry of Brauns lignin is very similar to that of native lignin; however, the molecular weight is lower (generally < 1000). Less than 10% of plant lignin can be isolated by this technique.

Enzymatically liberated lignin results when pure cultures of brown rot fungi selectively digest plant carbohydrates, leaving residual lignin that may be subsequently extracted with 95% ethanol. It is similar to Brauns lignin, but the yields are substantially higher.

Klason lignin is produced when plant matter is contacted with 72% sulfuric acid to dissolve polysaccharides. The residue contains most of the original lignin, so this procedure is often used to quantify lignin. However, the lignin has been highly modified by condensation reactions, making it unsuitable for studying lignin chemistry.

Lignin Analysis

Gravimetric Methods

Two approaches to gravimetric lignin measurement are taken: (1) the polysaccharides are removed and the residual lignin weighed, or (2) the lignin is removed and the residual polysaccharides weighed.

The first approach is used when preparing Klason lignin. Raw plant matter or acid detergent fiber (ADF) is soaked in 72% sulfuric acid, which hydrolyzes the polysaccharides. The residue is primarily lignin which is then filtered, dried, and weighed. Details may be found in the Technical Association of the Pulp and Paper Institute (TAPPI) methods T13 or T222, or the American Society for Testing and Materials (ASTM) method D1106. Approximately 0.2–0.5% of softwood lignin and 3–5% of hardwood lignin is soluble in concentrated acid, so this method (and all other methods) does not indicate 'true' lignin content.

The second approach uses either potassium permanganate or acidified triethylene glycol to remove lignin from ADF or uses chlorine compounds to prepare hollocellulose.

The most popular approaches for determining lignin content are the Klason and permanganate methods mentioned above. These two methods were compared using 75 different materials with lignin contents ranging from 1 to 50%. Although the two measurements were well correlated, the Klason lignin was about 19% less than the permanganate lignin. Sample handling can also affect the measured lignin content because heating may produce 'artifact lignin' by promoting Maillard reactions with proteins and sugars. Thus, it is difficult to measure the 'true' lignin content, so researchers must be satisfied with an operational definition of lignin. Some feel that the permanganate assay is the most accurate. (*See* **Browning:** Nonenzymatic.)

Lignin Indices

Some assays for lignin measurement do not give the lignin content directly, but use arbitrary measures which correlate with lignin. Examples are the κ number (TAPPI standard T236), which measures potassium permanganate depletion, the chlorine number (TAPPI standard T202), which measures chlorine depletion, and the methanol number which measures lignin methoxyl groups.

See also: **Amino Acids**: Properties and Occurrence; **Browning**: Nonenzymatic; **Carbohydrates**: Classification and Properties; **Cellulose**; **Dietary Fiber**: Properties and Sources; **Flavor (Flavour) Compounds**: Production Methods; **Phenolic Compounds**

Further Reading

Austin GT (1984) *Shreeve's Chemical Process Industries*, 5th edn, pp. 613–632. New York: McGraw-Hill.

Dintzis FR (1982) Dietary fiber analysis – concepts and problems. In: Lineback DR and Inglett GE (eds) *Food Carbohydrates*, pp. 312–332. Westport: AVI.

Glasser WG and Sarkanen S (1989) *Lignin: Properties and Materials*. ACS Symposium Series no. 397. Washington, DC: American Chemical Society.

Goheen DW and Hoyt CH (1981) Lignin. In: *Kirk-Othmer Encyclopedia of Chemical Technology*, 3rd edn, vol. 14. New York: Wiley.

Ni Y, Kubes GJ and Van Heiningen ARP (1990) Methanol number: a fast method to determine lignin content of pulp. *Journal of Pulp and Paper Science* 16: J83–J86.

Sarkanen KV and Ludwig CH (1971) *Lignins: Occurrence, Formation, Structure and Reactions*. New York: Wiley-Interscience.

Van Soest PJ and Wine RH (1968) Determination of lignin and cellulose in acid-detergent fiber with permanganate. *Journal of the AOAC* 51: 780–785.

Limes *See* **Citrus Fruits**: Types on the Market; Composition and Characterization; Oranges; Processed and Derived Products of Oranges; Lemons; Grapefruits; Limes

Lipids *See* **Fats**: Production of Animal Fats; Uses in the Food Industry; Digestion, Absorption, and Transport; Requirements; Fat Replacers; Classification; Occurrence

LIPOPROTEINS

J M Ordovas, Tufts University, Boston, MA, USA

Physiology

In the USA, cardiovascular diseases (CVD) affect about 62 million people. In 1999, almost one million subjects in the USA died as a consequence of CVD, which represents 1 of every 2.5 deaths, placing CVD as the major cause of morbidity and mortality. This is also becoming true at the global level, especially with the recent westernization of lifestyles in developing countries, which has dramatically increased the prevalence of CVD risk factors in those populations. Multiple risk factors and biochemical pathways have been identified with the progression of these diseases. These include tobacco smoke, dyslipidemias, physical inactivity, overweight and obesity, diabetes mellitus, and genetic factors. Several of these risk factors can be ameliorated by adequate nutritional behavior. In this article we concentrate on the physiology of plasma lipoproteins and the impact of nutritional factors on homeostasis and on the prevention of dyslipidemias and of CVD.

Lipids (cholesterol, triacylglycerols, and phospholipids) are transported in blood as part of lipoprotein particles. Lipoproteins are generally spherical particles, with a shell composed of unsterified cholesterol and phospholipid with the fatty acids oriented toward the core of the particle. Included in this shell are specific proteins known as apolipoproteins. The core of the lipoprotein particles is made up of cholesteryl ester and triacylglycerol molecules.

The nomenclature and classification of serum lipoproteins were originally determined by the laboratory technique used for their separation or isolation. These included mainly electrophoretic, ultracentrifuge, and immunological techniques. Accordingly, lipoproteins have been classified based on their electrophoretic mobility, hydrated density, and apolipoprotein composition (Table 1). Here, we present a description of the different lipoprotein particles, their metabolism, and the consequences of the alterations of their homeostasis.

Classification of Lipoproteins

Classification of Serum Lipoproteins According to their Electrophoretic Mobilities

With the development of techniques to separate proteins according to their electrophoretic characteristics, it was shown that most of the lipid present in serum was associated with proteins migrating with α_1- and β-globulin mobilities. This resulted in the first classification of lipoproteins according to their colocalization with other plasma proteins as α_1- and β-lipoproteins. Development of more advanced electrophoretic techniques resulted in further discrimination among the lipoprotein classes into β-, pre-β- and α-lipoproteins. Careful observation of thousands of electropherograms from normal and dyslipidemic subjects provided the basis for the first structured

Table 1 Classification and characteristics of human plasma lipoproteins

Lipoprotein	*Diameter (nm)*	*Density (g ml^{-1})*	*Electrophoretic mobility*	*Major lipids*	*Major apolipoproteins*
Chylomicrons	80–500	< 0.95	Origin	Dietary triacylglycerols and cholesteryl esters	ApoA-I, ApoA-IV, ApoB-48, ApoC-I, ApoC-II, ApoC-III, ApoE
Quilomicron remnants	> 30	< 1.006	Origin	Dietary cholesteryl esters and triacylglycerols	ApoB-48, ApoE, ApoC-III
VLDL	30–80	< 1.006	Pre-beta	Endogenous triacylglycerols	ApoB-100, ApoC-I, ApoC-II, ApoC-III, ApoE
IDL	25–35	1.006–1.019	Pre-beta and beta	Endogenous cholesteryl ester and triacylglycerosls	ApoB-100, ApoE
LDL	18–28	1.019–1.063	Beta	Cholesteryl esters	ApoB-100
Lp(a)				Cholesteryl esters	Apo(a), ApoB-100
HDL2	9–12	1.063–1.125	Alpha	Cholesteryl esters, phospholipids	ApoA-I, ApoA-II
HDL3	5–9	1.125–1.210	Alpha	Cholesteryl esters, phospholipids	ApoA-I, ApoA-II

VLDL, very-low-density lipoprotein; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; Lp(a), lipoprotein(a); HDL, high-density lipoprotein.

Table 2 Classical classification of hyperlipidemias according to Fredrickson

	Total cholesterol	*Triacylglycerol*	*Lipoprotein(s) fraction affected*	*Atherosclerosis risk*	*Genetic defect*
I	Normal to elevated	Very elevated	Chylomicrons	No	Familial LPL deficiency Apo C-II deficiency
IIa	Elevated	Normal	LDL	High	Familial hypercholesterolemia Familial combined hyperlipidemia Polygenic hypercholesterolemia
IIb	Elevated	Elevated	LDL and VLDL	High	Familial hypercholesterolemia Familial combined hyperlipidemia
III	Elevated	Very elevated	IDL	High	Familial dysbetalipoproteinemia
IV	Normal or elevated	Elevated	VLDL	Moderate	Familial hypertriglyceridemia Familial combined hyperlipidemia
V	Normal or elevated	Very elevated	VLDL and chylomicrons	Moderate	Familial hypertriglyceridemia

LDL, low-density lipoprotein; VLDL, very-low-density lipoprotein; IDL, intermediate-density lipoprotein; Apo C-II, apolipoprotein C-II.

classification of lipoprotein disorders, which is known as Fredrickson classification (Table 2). This classification played a pivotal role in the clinical aspects of lipoprotein metabolism, genetics, and coronary heart disease (CHD). Recently, a more functional approach to the classification of familial lipoprotein disorders has been made possible thanks to the advances in the elucidation of the specific gene loci responsible for some of these disorders.

Several electrophoresis media have been used for the separation of plasma lipoproteins. These include liquid, paper, cellulose acetate, agarose, and polyacrylamide. These techniques were used well into the 1980s by scientists and clinicians as a relatively quick tool for the assessment of lipoprotein patterns and dyslipidemias. These techniques, especially agarose gel electrophoresis, were used for identifying the presence of what is know as a broad beta band in the diagnosis of type III hyperlipidemia, a rare genetic disorder resulting from the presence of specific mutant forms of apolipoprotein E, primarily Apo E2.

Although, the classification of lipoproteins based on their electrophoretic behavior has been relegated to a historical value, the use of electrophoresis as a separating tool with potential clinical applications has seen a rebirth in recent years resulting from the realization that specific lipoprotein subclasses may be more informative in terms of CHD risk prediction than the traditional plasma lipid measures. At this regard, gradient agarose-polyacrylamide gel electrophoresis under nondenaturing conditions has been an essential tool to analyze the heterogeneity of low-density- (LDL) and high-density lipoprotein (HDL) subclasses. LDL subclasses can be resolved by nondenaturing polyacrylamide gradient gel (2–16%) in up to seven LDL subfractions with densities ranging from 1.020 to 1.063 g ml^{-1} and diameters ranging from 22.0 to 28.5 nm. The individual profiles usually show a major subclass and several minor ones (one to four). A predominance of smaller, denser LDL, versus larger, more buoyant LDL particles in plasma has been associated with increased CHD risk. The presence of specific subclasses in serum is determined by a combination of genetic and nongenetic factors, including age, gender, and diet. A significant correlation has been demonstrated between high triglyceride concentrations and the presence of small LDL subclasses. HDL subclasses have been resolved using similar techniques, with a polyacrylamide gradient ranging from 4 to 30%, into five subclasses (HDL3c, HDL3b, HDL3a, HDL2a, and HDL2b). Improved electrophoresis techniques have allowed the resolution of up to 14 HDL subclasses. The clinical significance of these subfractions needs further clarification. However, the current notion is that the presence of large HDL particles appears to be associated with protection, whereas small HDL particles are either neutral or even atherogenic.

Classification of Serum Lipoproteins According to their Ultracentrifugal Characteristics

The presence of lipids within the lipoprotein particles confers these macromolecular complexes with a lower density compared with other serum proteins. The development of ultracentrifugation in the 1940s allowed their separation from the other proteins based on this flotation property. Initially separated just a single discrete peak, it was not long before it was possible to separate a wide spectrum of particle sizes and densities (d) ranging from 0.92 to 1.21 g ml^{-1}.

Lipoproteins were classically separated into four major subclasses designated as chylomicrons (exogenous triacyglycerol-rich particles of $d < 0.94$ g ml^{-1}), very-low-density lipoproteins (VLDL, endogenous triacylglycerol-rich particles of $d = 0.94$–1.006 g ml^{-1}), LDL (cholesteryl ester-enriched particles of $d = 1.006$–1.063 g ml^{-1}), and HDL particles containing approximately 50% protein of

$d = 1.063–1.21\ g\ ml^{-1}$) (Table 1). Continued improvement of ultracentrifugation equipment and techniques permitted the resolution of these major families into several density subclasses such as HDL2a ($d = 1.10–1.125\ g\ ml^{-1}$), HDL2b ($d = 1.063–1.10\ g\ ml^{-1}$), and HDL3 ($d = 1.125–1.21\ g\ ml^{-1}$) and several LDL and VLDL subfractions.

Ultracentrifugation has been essential for the development of this field; however, this technique is very labor-intensive, thus seriously limiting its clinical and epidemiological applications. Moreover, lipoproteins are subjected during long periods of time to a harsh environment consisting both of high salt concentration and **g** forces. These problems have been mitigated by the development of small vertical and near-vertical rotors, which allow shorter runs, while achieving similar separations.

Classification of Serum Lipoproteins According to their Apolipoprotein Content

The characterization of plasma apolipoproteins (Tables 1 and 3) and the elucidation of their physiological function reinforced the notion that the measurement of the apolipoproteins rather than the lipids in lipoprotein fractions could provide more precise biochemical prediction of CHD risk. However, so far, apolipoprotein measurement has not provided a clear practical advantage over the measurement of more classical risk factors (total cholesterol (TC), HDL cholesterol (HDL-C), and LDL cholesterol (LDL-C)). However, this concept gave a thrust to the development of methods of separation based on apolipoprotein composition, namely those based on immunoaffinity chromatography. By using columns containing antibodies against specific apolipoproteins, several VLDL and HDL subpopulations can be resolved. More effort has been placed on the characterization of the HDL subfractions.

Lipoproteins containing apoA-I can be separated into two major species: those containing both apoA-I and apoA-II, known as LpAI:AII, and those containing apoA-I but not apoA-II (LpAI). It has been hypothesized that the protective role of HDL for CHD is relegated to the LpAI particles, whereas the LpAI:AII fraction may not be protective and could even be atherogenic. Likewise, lipoproteins containing apoB consist of four lipoprotein families. Lipoproteins containing apoB only (LpB) are cholesteryl ester-rich and are found primarily within the LDL density range. Particles containing both apoB and apoC (LpB:C), apoB and apoE (LpB:E), and all three apolipoprotein groups (LpB:E:C) are tryacylglycerol-rich lipoproteins (TRL) and are found within the VLDL and IDL density range.

The power of affinity chromatography has been utilized to separate another lipoprotein fraction known as lipoprotein(a) (Lp(a)). Lp(a) is a lipoprotein containing apoB100 as well as an antigenically unique apolipoprotein (apo(a)). Its separation from other plasma lipoproteins can be readily accomplished

Table 3 Classification and properties of apolipoproteins

Apolipoprotein	*Amino acids*		*Chromosomal localization*	*Main function*
ApoA-I	243	Liver, intestine	11q23	HDL structure, LCAT activator, ligand for HDL binding, reverse cholesterol transport
ApoA-II	77	Liver	1q21–1q23	HDL structure, LPL and HTGL activity
A-IV	377	Intestine	11q23	LCAT activator, lipid absorption (?) appetite regulator (?), regulator of LPL activity
A-V	366		11q23	TRL metabolism
Apo(a)	Variable	Liver	6q27	Lp(a) structure, plasminogen inhibitor (?)
Apo B-48	2152	Intestine	2p24	Chylomicron structure Secretion of chylomicrons
ApoB-100	4536	Liver	2p24	VLDL, IDL, LDL structure, LDL receptor ligand
ApoC-I	57	Liver, intestine	19q13.2	LPL and LCAT activator, inhibitor of LRP (?)
ApoC-II	79	Liver, intestine	19q13.2	LPL activator
ApoC-III	79	Liver, intestine	11q23	LPL inhibitor,
ApoC-IV	127		19q13.2	(?)
ApoD	169	Multiple tissues	3q26.2-qter	Closely associated with LCAT (?), radical scavenger (?)
ApoE	299	Liver macrophages	19q13.2	Ligand for most lipoprotein receptors, reverse cholesterol transport
APOF	326		12	Lipid transfer inhibitor, CETP inhibitor
ApoH	345		17q23-qter	Platelet function (?)
ApoJ	416		8p21-p12	(?)

HDL, high-density lipoprotein; LCAT, lecithin cholesterol acyl transferase; LPL, lipoprotein lipase; HTGL, hepatic triacylglycerol lipase; TRLs triacylglycerol-rich lipoproteins; CETP, cholesterol ester transfer protein; Lp(a), lipoprotein(a); VLDL, very-low-density lipoprotein; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein.

using immunoaffinity chromatography or alternatively taking advantage of its specific affinity for lectin. The latter has been used to develop a new technique to measure the concentrations of this atherogenic lipoprotein in clinical laboratories.

Measurement of Lipoprotein Subfractions using Nuclear Magnetic Resonance (NMR) Spectroscopy

More recently, NMR spectroscopy is being used to determine the concentrations of VLDL, LDL, and HDL subfractions. NMR distinguishes among the subfractions on the basis of slight differences in the spectral properties of the lipids carried within the particles, which vary according to the diameter of the phospholipid shell. Each profile displays the concentrations of six VLDL, one IDL, three LDL, and five HDL subclasses and the weighted-average particle size of VLDL, LDL, and HDL. The 10 lipoprotein subclass categories used were the following: large VLDL and remnants (80–220 nm), intermediate VLDL (35–80 nm), small VLDL (27–35 nm), large LDL (21.3–27.0 nm), intermediate LDL (19.8–21.2 nm), small LDL (18.3–19.7 nm), large HDL (8.8–13.0 nm), intermediate HDL (7.8–8.8 nm), and small HDL (7.3–7.7 nm). Levels of VLDL subclasses are expressed in units of triglyceride (mg dl^{-1}), and those of LDL and HDL subclasses in units of cholesterol (mg dl^{-1}). LDL and HDL subclasses distributions determined by gradient gel electrophoresis and NMR have also been shown to be closely correlated. The major advantages of NMR spectroscopy to measure lipoprotein subclasses are the low sample requirements, the speed and high throughput of the technique, and the relatively low cost.

Metabolism of Serum Lipoproteins (Figure 1)

Metabolism of Lipoproteins Carrying Exogenous Lipids

Although most of our knowledge about the relation between plasma lipid concentrations and cardiovascular disease comes from studies in the fasting state, we should keep in mind that, in affluent countries, humans are postprandial animals. Considering that the gastrointestinal system is busy with the processing of an ingested meal for at least 6–8 h and the average person takes three meals per day, the conclusion is that for most of the 24 h of a day the intestine is working. The digestive system works efficiently and dynamically with the dietary fat, with virtually no remains of the ingested fat being excreted. This places the gut as a central organ in lipid metabolism, and

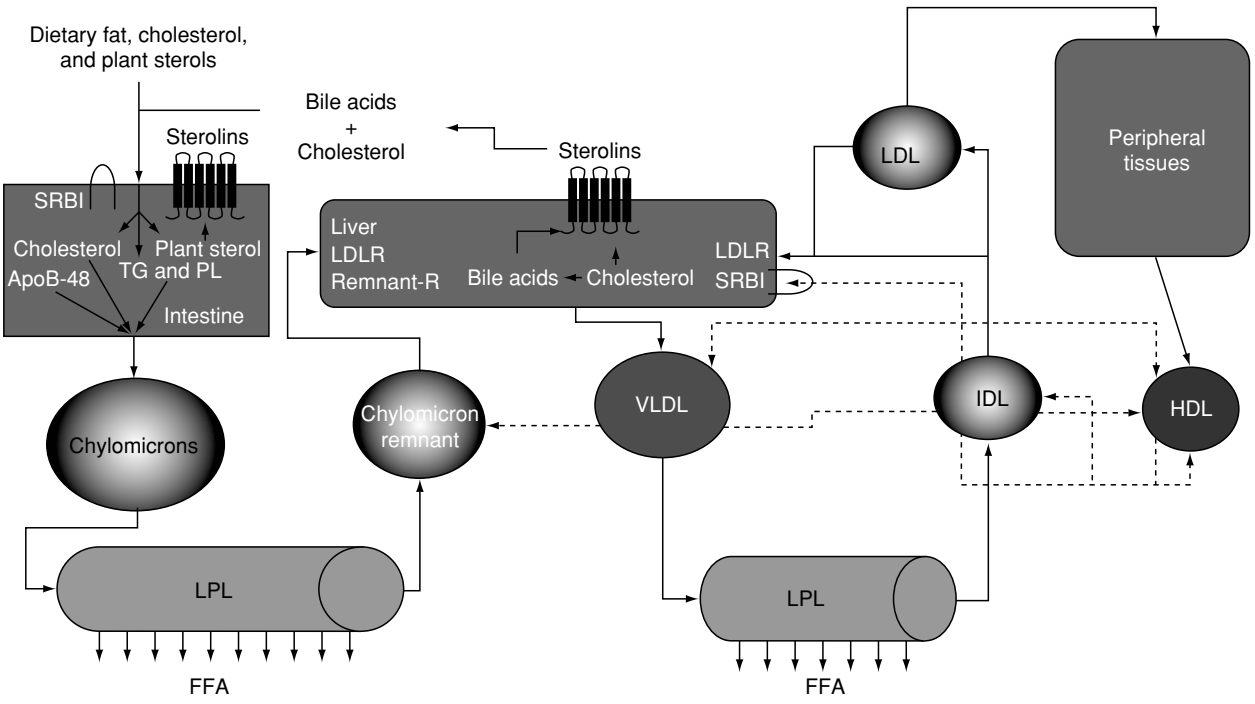

Figure 1 Depiction of the exogenous and endogenous lipoprotein metabolism pathways. SRBI, scavenger receptor, class B; TG, triglycerides; PL, phospholipids; LDLR, low-density lipoprotein receptor; LDL, low-density lipoproteins; VLDL, very-low-density lipoproteins; IDL, intermediate-density lipoproteins; HDL, high-density lipoproteins; LPL, lipoprotein lipase; FFA, free fatty acids.

underlines the importance of postprandial lipidemia. Postprandial lipid metabolism has received considerable attention since it was shown that postprandial TRL are involved in the development of atherosclerosis. Many studies comparing CHD patients with controls have demonstrated differences in postprandial TG after an oral fat load test and that the postprandial TG concentration is an independent predictor of CHD in multivariate analysis. A delayed clearance of retinyl palmitate (RP), used to study the metabolism of TRL of intestinal origin, discriminates between patients with CHD and controls, even after adjustment for fasting TG or HDL-C in normolipidemic men. The relation of postprandial lipid metabolism to atherosclerosis makes this evaluation a priority in atherosclerosis research. It is no longer possible to ignore its importance and to try to explain the relation between lipids and atherosclerosis exclusively on the basis of the fasting lipid levels. Therefore, it is of paramount importance to gain a better understanding about the mechanisms involved in the metabolism of lipoproteins carrying exogenous (dietary) lipids, primarily triacylglycerols and cholesterol. It is important to emphasize that, whereas the absorption of fat is highly efficient and uniform in most humans, the absorption of cholesterol is highly variable. Therefore, the understanding of this variability could be applied to the generation of new therapeutic targets for the reduction of CHD.

Cholesterol has for decades received a bad press and for most people is just a 'fat' in blood that increases the risk of heart disease. However, cholesterol is essential for the growth and viability of cells in higher organisms. We have about 100 g of cholesterol; of those 100 g, about 7 g is found in the blood circulating as part of lipoproteins, whereas the remaining 93 g is located in the body cells. The physical characteristics of cholesterol, with its insolubility in water and its rigidity, provides cell membranes with their structural integrity and modulates their fluidity, thus maintaining an optimal communication between the cell and its environment, including the transport of nutrients and the proper maintenance of energy resources. Another major use of cholesterol is on the synthesis of bile acids. These are synthesized in the liver from cholesterol and are secreted in the bile. They are essential for the intestinal absorption of fat. Without cholesterol we could not absorb the essential fatty acids and fat-soluble vitamins A, D, E, and K from the food. In lesser amounts, cholesterol is used to synthesize steroid hormones, including the sex hormones estrogen, progesterone, and testosterone as well as the corticosteroids. Cholesterol is also the precursor from which the body synthesizes vitamin D.

Cholesterol homeostasis is crucial for optimal performance of multiple biochemical pathways, and it is maintained by a delicate equilibrium between dietary cholesterol absorption, *de novo* synthesis, and fecal excretion. In terms of the *de novo* synthesis, higher organisms have developed a complex biosynthetic pathway for the synthesis of cholesterol, which has been precisely defined, requiring approximately 30 steps. Conversely, the process of intestinal cholesterol absorption from foods has been an area of controversy. Many believed this to be a process of passive diffusion across the brush-border membrane from the lumen of the gut into the enterocyte. However, evidence exists that points to a highly regulated carrier-mediated process. The scavenger receptor SR-BI and the 'sterolins' have been implicated in this process and various apolipoproteins can affect the rate of cholesterol transfer *in vitro*.

We synthesize approximately 900 mg of cholesterol per day. In addition, we consume on average $\sim$300 mg of cholesterol per day with the diet. Under optimal homeostatic conditions, the same amount (1200 mg day^{-1}) is excreted as fecal sterols. In terms of the intestinal cholesterol absorption, the exogenous cholesterol is hydrolyzed to free cholesterol (FC) by pancreatic lipases secreted into the intestine. Bile acids solubilize dietary and biliary cholesterol in micelles in the proximity of the brush border of enterocytes, where it diffuses through the epithelial membrane to the interior of the cell. Our current knowledge indicates that some of the exogenous cholesterol and most of the plant sterols entering the enterocyte are secreted back to the intestinal lumen via the recently characterized ABC transporter G5 (ABCG5) and ABC transporter G8 (ABCG8), also known as 'sterolins.'

In the smooth endoplasmic reticulum, exogenous cholesterol is converted to cholesteryl ester (CE) by acyl CoA:cholesterol acyltransferase (ACAT), and then used by microsomal triglyceride transfer protein (MTP) to synthesize chylomicrons. These chylomicrons are pumped through the thoracic lymph duct into the blood, where lipoprotein lipase (LPL) hydrolyzes them into chylomicron remnants.

Chylomicrons and their remnants are heterogeneous particles that vary in size, composition, and metabolism. They transport dietary lipids and are considered 'exogenous' lipoproteins, in contrast to 'endogenous' VLDL. Chylomicrons are synthesized continuously, and may increase in size and cholesterol content when dietary fat and cholesterol increase. They have a higher turnover rate than that of LDL particles and may contribute more cholesterol to tissues in a 24-h period. Most chylomicron remnants are rapidly catabolized in the liver after their uptake

by specific receptors. In addition, large chylomicron particles may be metabolized to small remnants (45 nm) that are similar in size to atherogenic LDL remnants and may readily penetrate arterial cells via transcytotic vesicles. Data from animal and human studies strongly support the concept that chylomicron remnants may have a dramatic impact on the risk of atherosclerosis and CHD.

There has been a long and heated debate surrounding the use of low-fat, high-carbohydrate diets as the preferred recommendation to lower CVD risk in the general population. On one hand, these diets have been shown to lower plasma TC and LDL-C; however, on the other hand, in populations leading a westernized lifestyle, these carbohydrate-rich diets can simultaneously raise the levels of the (also atherogenic), TRL and lower HDL-C concentrations. This leads to the concern that, at least in some individuals, the beneficial LDL-lowering effect may be offset by adverse changes in other circulating lipoprotein fractions. This concept is specially relevant following the demonstration of a significant correlation between TG levels and the presence of the highly atherogenic small dense LDL subfraction.

Endogenous Lipoprotein Metabolism

Liver cells can metabolize CE into bile acids or directly secrete it into the bile. Bile acids aid in the formation of micelles, and the secreted cholesterol is excreted via the feces. Hepatic cells also synthesize endogenous cholesterol and release VLDL and LDL-C into the circulation. These molecules are delivered to peripheral organs for either utilization or storage. Therefore, the liver is involved in both the uptake of newly synthesized cholesterol and the export of the metabolized sterol into the feces through either bile acid or neutral sterol formation. Consequently, the liver must respond to variations that occur in both input and output pathways.

Lowering LDL-C by reducing the intake of saturated fat is one of the cornerstones of the current dietary recommendations to reduce CHD risk. In most cases the reduction in saturated fat is accompanied by a parallel reduction in total fat intake. The reduction of LDL-C following a reduction in saturated fat appears to be mediated by increases in LDL receptor activity, resulting in enhanced liver LDL clearance. However, the downside of these recommendations, as previously indicated, lays in the effects on triglyceride levels and possibly on LDL subfractions.

Two major LDL subclasses have been defined: type A and type B. Type A is larger, more buoyant and apparently less atherogenic. Conversely, type B is defined as those particles that are smaller, denser, and more atherogenic. In view of these qualitative differences in CHD risk, it is very important to examine the diet-induced changes on each of those subfractions. The current data support the notion that in asymptomatic subjects carrying the pattern B phenotype, switching from a high-fat to a low-fat diet results in significant LDL-C reduction, which is larger than for those subjects carrying the pattern A phenotype. Moreover, as expected, the LDL-C reduction is primarily due to lowering of the small LDL class and this should result in significant CHD risk reduction. Conversely, in asymptomatic subjects carrying the pattern A phenotype, a change from a high-fat to a low-fat diet results in a shift from buoyant to denser LDL, with very small changes in particle numbers. This shift may not be associated with a reduction in CHD risk and may even increase the atherogenic risk.

Reverse Cholesterol Transport

The recent characterization of SRBI as an HDL receptor and the identification of ABCA1 as the membrane transporter that mediates the efflux of cholesterol from the cells have pushed forward our understanding of HDL metabolism and reverse cholesterol transport.

HDL is synthesized by both the liver and the intestine. Its precursor form is discoidal in shape and matures in circulation as it picks up unesterified cholesterol, and probably phospholipids, from cell membranes by a mechanism mediated by the ABCA1 transporter. Moreover, HDL can also obtain lipids and apolipoproteins from TRL as these particles undergo lipolysis. The cholesterol is esterified by the action of the lecithin cholesterol acyltransferase (LCAT) and the small HDL3 particle becomes a larger HDL2 particle. The esterified cholesterol is either delivered to the liver or to organs synthesizing steroid hormones, such as the adrenals, via SRBI or other receptor-independent mechanisms, or transferred by the action of cholesteryl ester transfer protein (CETP) to other lipoproteins (such as chylomicron, VLDL remnants, or LDL), in exchange for triacylglycerols. This cholesterol may then be taken up by the liver via receptors specific for these lipoproteins (LDLR, VLDLR, remnant receptor), or it can be delivered again to the peripheral tissues. The triacylglycerol received by HDL2 is hydrolyzed by hepatic lipase and the particle is converted back to HDL3, completing the HDL cycle in plasma. In the liver, cholesterol can be excreted directly into bile, converted to bile acids, or reutilized in lipoprotein production.

Diets high in simple carbohydrates reduce HDL cholesterol concentrations. This effect appears to be mediated by increases in the catabolism of apoA-I; however, it has also been proposed that decreases in apoA-I production contribute to this effect.

Disorders of Lipoprotein Metabolism

Many of the major advances in our knowledge of lipoprotein metabolism and physiology have been driven by the study of abnormalities. As indicated above, familial lipoprotein disorders were first classified according to their different electrophoretic profile. This classification was successfully used for decades; however, it had shortcomings. One of them was that it included under the same denomination diseases with different molecular basis. Another one was that some important phenotypes (i.e., familial hypoalphalipoproteinemia) were not part of the original classification. The elucidation of the molecular mechanisms responsible for several of these syndromes has allowed a more functional classification, based on the genetic loci influencing the phenotypes. Below, a brief summary of the most common or relevant familial lipoprotein disorders is presented.

Familial Hypercholesterolemia

Familial hypercholesterolemia (FH), also known as FHC, and according to the original Fredrickson classification as hyperlipoproteinemia type IIa, is an autosomal dominant disorder characterized by dramatic elevations of plasma LDL cholesterol concentrations. The locus responsible for this disorder is the low-density lipoprotein receptor (LDLR) in chromosome 19. The LDLR gene family consists of cell surface proteins involved in receptor-mediated endocytosis of specific ligands. LDL is normally bound at the cell membrane and taken into the cell, ending up in lysosomes where the protein is degraded and the cholesterol is made available for repression of microsomal enzyme 3-hydroxy-3-methylglutaryl coenzyme A (HMG CoA) reductase, the rate-limiting step in cholesterol synthesis. At the same time, a reciprocal stimulation of cholesterol ester synthesis takes place. Hundreds of mutations have been identified at the LDLR locus, resulting in the FH phenotype. The global frequency of FH is about 1 in 500 in the heterozygous state and 1 in a million in the homozygous state. However, this varies considerably and there are places like the province of Quebec where the frequency of FH is much higher.

The ranges of LDL-C concentrations in the plasma of heterozygous FH are 200–400 mg dl^{-1}, whereas in homozygotes the LDL-C concentrations range between 400 and 1000 mg dl^{-1}. The clinical characteristics observed in heterozygotes include tendinous xanthomas, corneal arcus, and CHD in the 40s and 50s. In homozygotes, there is also the presence of planar xanthomas and the clinical manifestation of the disease is more premature.

Inhibitors of HMG CoA, a key enzyme in the synthesis of cholesterol, are the drugs of choice in the treatment of heterozygotes FH. Most pharmacological therapies are ineffective in the homozygous state. These subjects may be treated with LDL apheresis, liver transplant, and, based on some preliminary results, in the future, these subjects could be excellent candidates for gene therapy. Diet and lifestyle modification are also recommended and in some cases may be highly effective. This is supported by the fact that the phenotypic expression of some of the LDLR mutations is not expressed in populations living a traditional, healthy lifestyle, but it is expressed in those subjects living in societies with a more westernized lifestyle. The impact of modernization on the clinical expression of the disease has also been demonstrated by studying mortality during the last 200 years in a large pedigree with a well-characterized LDLR abnormality. In these subjects the presence of the mutation did not manifest as premature death until the middle of the twentieth century, suggesting that lifestyle could play a relevant role in the clinical expression of this phenotype. Moreover, we are starting to identify some other genetic factors that can modulate the biochemical and clinical manifestation of the phenotype: gene-by-gene interactions.

There are other forms of hypercholesterolemia that are not due to defects on the LDLR. One of them produces similar biochemical phenotypes and it has been named familial defective apoB-100. This disease is also autosomal dominant. The frequency of this disorder has not been well elucidated and it varies considerably depending on the ethnicity of the population studied, as it appears to have a central European origin. The locus responsible for this phenotype is the apolipoprotein B (Apo B) in chromosome 2 and the specific defect is a point mutation that changes the amino acid 3500 of the mature apoB (ARG3500GLN). It has been shown that the arginine-3500 interacts with tryptophan-4369 in the carboxyl tail and facilitates the conformation of ApoB-100, required for normal receptor binding of LDL. Disruption of this conformation prevents the normal binding of LDL to its receptor, resulting in the observed hypercholesterolemia.

The genetic defects associated with the common forms of hypercholesterolemia present in subjects with cholesterol levels between 250 and 300 mg dl^{-1} have not been elucidated. These are, most probably, due to a combination of predisposing alleles (i.e., the

presence of the E4 allele at the Apo E locus) with environmental factors (i.e., diet, physical inactivity, smoking, obesity). This 'garden variety' of mildly severe hypercholesterolemia is also known as polygenic hypercholesterolemia.

Familial Dysbetalipoproteinemia

This disorder is characterized by simultaneous increases in plasma cholesterol and tryglyceride concentrations. The locus responsible for this familial disorders is the Apo E gene in chromosome 19. The most common genetic defect is the presence of the E2 (arg158-to-cys) allele. Other less common mutations include the E2 (lys146-to-gln), E2 (arg145-to-cys) and E2-Christchurch (arg136-to-ser). In most patients, the complete expression of the clinical phenotype needs additional interactions with age, obesity, and diabetes, as well as other genetic loci. In addition to the accumulation in plasma of VLDL and chylomicron remnants, other clinical features of this disorder are tuboeruptive and, in some cases, planar xanthomas. Therapies include diet and hypolipidemic agents such as fibrates, statins, or nicotinic acid. In most cases, diagnosis can be carried out first by lipoprotein electrophoresis, followed by the molecular elucidation of the Apo E defect.

Familial Combined Hyperlipidemia

Familial combined hyperlipidemia (FCHL) was initially described as the combination of hypercholesterolemia and hypertriglyceridemia within the same kindred, and with kindred members having one of these abnormalities or both. Moreover, most subjects with FCHL have HDL-C concentrations below the 10th percentile. In addition, predominance of small dense LDL particles and elevated apolipoprotein B levels is commonly found in members of FCHL families. This disorder has a frequency of approximately 10% in survivors of premature myocardial infarction (less than 60 years of age) and about 14% in kindred with CHD.

Given the high frequency and the clinical implications of this disorder, there has been an intense search for the molecular and genetic defect(s) responsible for this phenoype. It has been reported that affected subjects have overproduction of ApoB-100, and several chromosomal regions have been associated with this defect, including 2p, 11p, 16q, and 19q, as well as mutations at several candidate genes (LPL, ApoA1-ApoC3-ApoA4). The expression of this disorder may be triggered by other factors such as overweight, hypertension, diabetes, and gout. The treatment should include diet and physical activity and, if necessary, statins or fibrates, depending on the major lipid present in excess.

Another common biochemical syndrome associated with increased CHD risk is familial hyperapobetalipoproteinemia, which has been suggested to be a variant of FCHL. Hyperapobetalipoproteinemia is characterized by ApoB values above the 90th percentile in the absence of other lipid abnormalities. Therefore, there will be an increased number of small, dense LDL particles. The gene(s) responsible for this disease has not been elucidated; however, there is evidence showing the Mendelian inheritance and the etiologic heterogeneity of the hyperapoB phenotype.

Familial Hypoalphalipoproteinemia

Severe HDL deficiency, characterized by HDL-C concentrations $< 10\,\mathrm{mg\,dl^{-1}}$, is rare and may be due to Tangier disease (TD), ApoAI, and LCAT deficiencies. The ApoAI deficiency states are due to rare deletions, rearrangements, or point mutations within the ApoA1 locus. Familial hypoalphalipoproteinemia (FHA) is relatively common and is characterized by HDL-C concentrations below the 10th percentile of normal. These subjects have been reported to have either decreased HDL production or increased HDL ApoAI catabolism. This phenotype is present in about 4% of kindred with premature CHD.

A major recent breakthrough in relation to this phenotype has been the identification of ABC1 as the gene for TD and for some cases of FHA. This finding opens enormous opportunities for the lipoprotein field and for CHD therapy. The metabolic pathways determined by the product of the *ABCA1* gene will contribute to our understanding of reverse cholesterol transport and its regulation by dietary fat and cholesterol. It remains to be elucidated whether common mutations at this locus are associated with altered HDL-C levels in the population. This information will determine the full impact of the *ABCA1* gene in controlling cholesterol efflux and HDL metabolism in the general population, beyond those affected by TD or FHA. This information will be applied to the genetic screening of subjects at high risk for CHD and it could provide us with new therapeutic targets, thus increasing our ability to prevent and treat CHD.

Familial Hypertriglyceridemias

Several familial disorders share the phenotype characterized by high TRLs. The initial classification distinguished three distinct categories: type I, IV, and V. The genetic basis has been elucidated only for type I hyperlipidemia, also known as chylomicronemia, familial LPL deficiency, or ApoC-II deficiency. As its name suggests, chylomicronemia is characterized by

greatly elevated concentrations of exogenous triacylglycerols and it is the result of impaired lipolysis of chylomicrons due to a deficiency of LPL of its activator, the ApoC-II. Several genetic mutations at the LPL and *APOC2* genes have been reported. These are autosomal recessive traits. In the heterozygous state, subjects have normal to slightly elevated plasma triglycerides, whereas homozygotes have triacylglycerol levels that may exceed 1000 mg dl^{-1} in the fasting state. The diagnosis of the homozygous state takes place during the first years of life from the presence of recurrent abdominal pain and pancreatitis. Eruptive xanthomas and lipemia retinalis also occur.

The recommended treatment includes a diet low in simple carbohydrates and with a fat content below 20% of total energy. The use of medium-chain triglycerides (MCT) has also been reported to be efficacious. Body weight should be maintained within normal limits and alcohol consumption should be avoided.

Fasting chylomicronemia has not been clearly associated with increased risk for atherosclerosis, but homozygous subjects are at very high risk for pancreatitis. However, there is considerable evidence supporting the atherogenic properties of chylomicron remnants.

Familial type IV and type V hypertriglyceridemias may have some overlapping phenotypes. In type IV or familial endogenous hypertriglyceridemia, triacylglycerol, specifically VLDL concentrations, is increased, even on a regular diet, and HDL is usually decreased. This disorder appears to be autosomal dominant and relatively frequent in populations consuming high-fat diets. The precise molecular defect(s) has not been identified; however, the increase in triacylglycerol is associated with overproduction of triacylglycerol by the liver and often with the subsequent reduced clearance. Precocious atherosclerosis, abnormal glucose tolerance, and atheroeruptive xanthoma may occur. The disorder is undoubtedly heterogeneous and the phenotype is strongly influenced by environmental factors, particularly carbohydrate and ethanol consumption.

Other conditions causing hyperlipoproteinemia IV are uremia, hypopituitarism, contraceptive steroids, and glycogen storage disease I. Diet should be the first step in therapy, followed if necessary by pharmacotherapy using fibrates.

Type V hyperlipidemia is a much rarer disorder associated with increased susceptibility to atherosclerosis. Usually the first signs of this abnormality are abdominal pain or pancreatitis. VLDL concentrations are high and chylomicrons are present in the fasting states. This abnormality has not been linked to any specific molecular defect.

Familial Lipoprotein(a) Excess

Lipoprotein(a) (Lp(a)) is an LDL particle with one molecule of Apo(a) attached to it. Elevated levels of Lp(a) (> 35–40 mg dl^{-1} or 90th percentile) have been associated with premature CHD. This increased risk appears to result from two different mechanisms: cholesterol deposition in the arterial wall and inhibition of fibrinolysis.

Lp(a) concentrations are highly variable; however, they are relatively constant during a person's lifetime. Between 80 and 90% of the variability appears to be of genetic origin, owing, for the most part, to variations at the structural Apo(a) gene locus. Lp(a) concentrations are inversely associated with a size polymorphism of Apo(a). This polymorphism is due to differences in the number of a multiple repeat of a protein domain highly homologous to the kringle 4 domain of plasminogen. Diets and medications used to lower LDL-C do not appear to have a significant effect on Lp(a) concentrations. However, niacin has been reported to decrease Lp(a) concentrations. There is some evidence suggesting that diets high in *trans* fatty acids have a significant raising effect on Lp(a) concentrations, whereas *n*-3 fatty acids and estrogen replacement therapy lower Lp(a) concentrations in the general population and in postmenopausal women, respectively.

Current Guidelines for Dietary Prevention of CHD

The cardiovascular benefit from lowering LDL-C with diet or drug therapies in the general population and in patients with hyperlipidemia or CHD has been shown in many population and clinical studies. In general, dietary therapy includes using diet restricted in total fat, but more specifically saturated fat (< 7% of calories) and cholesterol (< 200 mg day^{-1}) (Table 4). Pharmacological therapies include primarily statins and fibric acid derivatives. Clinical studies have shown that both families of drugs lower CHD and total mortality. However, there is great heterogeneity in the way humans respond to any therapeutic approach, often requiring empirical strategies to find the appropriate diet, behavioral or drug class and dose for each patient. Nutrigenetics and pharmacogenetics provide the experimental basis to understand the current unpredictability in response to diet and drugs as a function of the genetic variability. Some of the evidence for the genetic basis of this phenomenon came from the seminal observations showing differences in drug responses and drug metabolism by subjects from different ethnic groups, suggesting that genetic factors, underlying ethnicity, were responsible for those effects. The formal emergence of pharmacogenetics as a new field of experimental science took

Table 4 Guidelines for the prevention of coronary heart disease

Nutrient	*Recommended intake*	
	American Heart Association recommendations	*International Task Force recommendations*
Saturated fat[a]	Less than 7% of total calories	No more than 7–10% of energy
Polyunsaturated fat	Up to 10% of total calories	No more than 7–8% of energy
Monounsaturated fat	Up to 20% of total calories	No more than 10–15% of energy
Total fat	25–35% of total calories	No more than 30% of energy
Carbohydrate[b]	50–60% of total calories	> 50% of energy
Fiber	20–30 g day^{-1}	More than 25 g dietary fiber per day
Plant stanols/sterols	2 g day^{-1}	
Protein	About 15% of total calories	Less than 300 mg day^{-1}
Cholesterol	Less than 200 mg day^{-1}	
Total calories (energy)[c]	Balance energy intake and expenditure to maintain desirable body weight/prevent weight gain	

[a]*Trans* fatty acids are another low-density lipoprotein-raising fat that should be kept at a low intake.
[b]Carbohydrate should be derived predominantly from foods rich in complex carbohydrates, including grains, especially whole grains, fruits, and vegetables.
[c]Daily energy expenditure should include at least moderate physical activity (contributing approximately 200 kcal day^{-1}).

place during the 1960s and over the past 40 years there has been great progress in understanding the molecular basis of drug action and in elucidating genetic determinants of disease pathogenesis and drug response. More recently, with a better understanding of the human genome and the technical developments in molecular biology as well as in statistics, we are moving towards larger and more complex studies, which may involve multiple genes, gene-by-gene and gene-by-environment interactions. This knowledge should attain the much searched goal of providing each subject at risk with the most effective dietary, behavioral, or pharmacological therapy to reduce the risk of disease.

Acknowledgments

Supported by NIH/NHLBI grant no. HL54776, NIH/NHLBI contract no. 1-38038, and contracts 53-K06-5-10 and 58-1950-9-001 from the US Department of Agriculture Research Service.

See also: **Cholesterol**: Properties and Determination; Absorption, Function, and Metabolism; Factors Determining Blood Cholesterol Levels; Role of Cholesterol in Heart Disease; **Coronary Heart Disease**: Prevention; **Spectroscopy**: Nuclear Magnetic Resonance

Further Reading

Chisolm GM and Steinberg D (2000) The oxidative modification hypothesis of atherogenesis: an overview. *Free Radical Biology & Medicine* 28: 1815–1826.

Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). (2001) *JAMA* 285: 2486–297.

Fielding CJ and Fielding PE (2001) Cellular cholesterol efflux. *Biochemica et Biophysica Acta* 1533: 175–189.

Griffin BA and Fielding BA (2001) Postprandial lipid handling. *Current Opinion in Clinical Nutrition & Metabolic Care* 4: 93–98.

Krauss RM (2001) Dietary and genetic effects on low-density lipoprotein heterogeneity. *Annual Review of Nutrition* 21: 283–295.

Lu K, Lee MH and Patel SB (2001) Dietary cholesterol absorption; more than just bile. *Trends in Endocrinology & Metabolism* 12: 314–320.

Mann J and Chisholm A (1999) Nutritional determinants of serum triglycerides. *Nutrition Metabolism & Cardiovascular Diseases* 9: 1–8.

Nestel PJ (2000) Fish oil and cardiovascular disease: lipids and arterial function. *American Journal of Clinical Nutrition* 71: 228S–231S.

Ordovas JM (2001) Gene–diet interaction and plasma lipid response to dietary intervention. *Current Atherosclerosis Reports* 3: 200–208.

von Eckardstein A, Nofer JR and Assmann G (2001) High density lipoproteins and arteriosclerosis. Role of cholesterol efflux and reverse cholesterol transport. *Arteriosclerosis, Thrombosis & Vascular Biology* 21: 13–27.

LIQUEURS

Contents

Composition

M L González-Sanjosé and S Pérez-Magariño,
University of Burgos, Burgos, Spain

Background

Back in the fourth century, ancient alchemists discovered that the medicinal qualities of some of the remedial herbs, spices, fruits, and barks were preserved longer when infused with alcohol.

In the twelfth century, an alchemist located in France wrote of flavored alcohols that had restorative and life-giving powers. He talked of the distillation of wine into 'aqua vitae' and the subsequent flavoring of these spirits with various herbs and spices. It was believed that these beverages were vital and life-restoring. At first, they were used only as alchemy potions but later, around the fourteenth century, were used as a pleasurable drink in the court of 'Catherine de Medici' of France.

Between the fourteenth and seventeenth century, alchemists and monks did most of the production of these liqueurs. Benedictine monks are attributed with the production of Benedictine and Chartreuse, an 'Elixir de longue Vie' (elixir of long life). By the end of the seventeenth century, several distilleries had been set up, which produced commercial quantities of liqueurs.

The word 'liqueur' is derived from the Latin 'liquefacere,' which means 'to melt, or dissolve' and refers to the methods of flavoring the alcohol, forming the base of the liqueur.

Definition

The alcoholic beverages, which have levels of alcohol less than 15–16 ° Alc, are usually named low alcoholic beverages, and their alcohol content is produced by fermentation of their original musts. However, spirits are alcoholic beverages with levels of alcohol over 15–20 ° Alc, from the distillation of a fermenting medium. Liqueurs are, by definition, sweetened spirits that can be both colored and colorless. Most of the older classical liqueurs have their origin in the monasteries and have been produced for several hundred years. Until fairly recently, liqueurs were produced by traditional methods without legal controls. However, nowadays, their classification and legislation are well defined, especially in Europe and USA.

Under European Council Regulations (EEC 1576/1989), a liqueur is defined as:

> a spirit drink having a minimum sugar content of $100\,g\,l^{-1}$ (as invert) produced by flavoring ethyl alcohol of agricultural origin, or a distillate of agricultural origin or one or more spirit drinks, sweetened, and possibly with the addition of products of agricultural origin such as cream, milk, or other milk products, fruit, wine or flavored wine.

Liqueurs are a subclassification of 'spirit drinks', so they must contain at least 15 ° Alc. In practice, they are produced with strengths ranging from 15 to 60 ° Alc, although most of the traditional liqueurs contain 35–45 ° Alc. Owing to consumer demands, many of the newer brands are weaker.

In the USA, the manufacture of liqueurs is controlled by the Bureau of Alcohol, Tobacco, and Firearms, and liqueurs are defined as:

> products obtained by mixing or re-distilling distilled spirits with or over fruits, flowers, plants or pure juices therefrom, or other natural flavoring materials, or with extracts derived from infusions, percolation or maceration of such materials and containing sugar, dextrose or levulose, or a combination thereof in an amount not less than 2.5% by weight of the finished product.

Therefore, the liqueur definitions in Europe and the USA are significantly different, especially in terms of minimum sugar content. In any case, both definitions are not very specific and allow a wide range of liqueur types.

Although liqueurs and spirits are both alcoholic distillates, the terms are not usually interchangeable. The distinction lies in the way in which the various flavors are obtained. Liqueurs are spirit-based drinks to which flavors are added, or extracted from vegetal material by maceration or infusion, and are often enhanced by sweetening.

Sometimes, generic terms such as aquavits, schnapps, eau de vies, or vodkas are used in different countries for both spirits and liqueurs. However, they are in fact two different products, not only in their sugar content but also in their elaboration process. Both are fermented and distilled, but liqueurs are obtained by steeping vegetal materials in an alcohol that has already been fermented and distilled. Some special liqueurs in the British Commonwealth nations, especially the UK and Australia, are called cordials, which are prepared by steeping fruit pulps or juices in sweetened alcohol.

Raw Material and Composition

All liqueurs are produced from a common list of ingredients such as alcohol, sugar, flavoring, colorant or pigment, and water, but some liqueurs may contain other components such as cream, fruit juices, emulsifiers, plant materials, and some additives, which are subject to legislation. The elaboration of a good liqueur implies a good selection of the raw materials and exhaustive control of the substances added to obtain equilibrate and stable beverages.

Alcohol Base

The alcohol base (ethanol) may be originated from any fermentable agricultural material. Synthetic alcohol from petroleum is not allowed for use in alcoholic beverages.

Plants contain simple carbohydrates (sugars) or complex carbohydrates (starch or cellulose), from which it is possible to obtain alcohol by fermentation. This general principle is the basis for the production of the agricultural alcohol for alcoholic beverages. Sugars present in certain vegetable products, simple (most fruits) or complex (cereals) must be extracted, and, if necessary, they must also be hydrolyzed (as starch). Then, the simple sugars can be used as a substrate to obtain alcohol (ethanol) by the action of yeast.

The main steps in this transformation, which may vary slightly depending on the type of material, are as follows:

- Pretreatment of the biomass. This process implies the transformation of the raw material to favor sugar extraction. Raw materials are crushed, milled, grinned, cut, etc., to reduce the particle size and increase the surface of extraction, favoring the extraction of sugars and some other compounds. These processes are needed, especially in the case of cereals, potatoes and other root crops, as well as for other plants such as agaves. Extraction of the juices or dissolution–extraction of the sugars from the solid parts is usually carried out by pressing, boiling, mashing, and extraction by diffusion.
- The next step is the hydrolysis or transformation of the complex carbohydrates into simple sugars in aqueous medium. These sugars are used in fermentation. In the case of alcoholic beverages, this process always takes place by enzymatic activity, generally taking advantage of the activity of the substrates themselves, or resorting to the use of exogenous permitted enzymes.
- Alcoholic fermentation is then carried out, transforming sugars into ethanol by the action of yeasts. Few species of yeast are capable of maintaining themselves in a state of active fermentation once high levels of alcohol have been reached. Several products obtained in this way, with some further modifications, make up the largest group of drinks produced by fermentation with a relatively low alcohol content, such as cider, beer, and wine.
- In order to obtain products with a higher alcohol content, distillation must be carried out. Distillation involves the separation and concentration of ethanol from the fermented medium, to obtain alcoholic distillates of different strengths and sensorial characteristics, depending not only on the original substrate but also on the type of distillation equipment employed. Such products obtained in this way are sometimes called neutral spirits, and they are the base for the manufacture of different spirit beverages such as brandy, whisky, gin, vodka, and liquors, amongst others.

A direct distillation process produces distillates with alcohol strength of no more than 86 ° Alc. To obtain a product with a higher content of alcohol, redistillation is necessary.

The most common process for obtaining high-quality distillates uses a copper still, in which successive distillations are carried out to yield highly flavored products and with an alcoholic strength of about 72 ° Alc. However, the distillation can be also carried out in distillation towers, in which the distillate can achieve higher alcoholic grade, up to 86 ° Alc. In general, these distillates are less flavored than those obtained in stills. Columns can be made of copper or stainless steel.

The distillation process generates several fractions:

1. Heads, which contain the most volatile constituents, are often detrimental to the quality of the spirit. Methanol is one such undesirable product.
2. Heart, which contains the ethanol and other important volatile compounds responsible for the aroma, which evoke the origin of the distillate.

3. Tails, which are rich in alcohols with a high number of carbons and with a high potential to reduce the quality of the spirit. However, this fraction contains some interesting less volatile components that can be important; these can be recirculated to recover these products without spoiling the quality of the spirit.

In order to guarantee the purity of the alcohol, the first fraction (heads) must be correctly rejected, preventing methanol from appearing in the final product, and only the heart is used later to improve the final spirit. The different methods of carrying out the distillation (type of stills, number of successive distillation, recycling of tails, etc.) confer different characteristics to the spirits, and have a direct effect on their quality.

The alcohol base used for liqueurs is usually 'neutral,' i.e., clear and tasteless, so flavor compounds of distillates do not contribute to the final flavor characteristics of liqueurs, which are obtained by adding specific products. However, for special liqueurs, flavored alcohols, rich in odor compounds isolated together with ethanol during distillation, are preferred. For example, in 'juice liqueurs' or 'fruit liqueurs' flavored alcohols obtained from fermentation and distillation of the same juice fruit are used to intensify the flavor of the final product.

Water

Water with a high level of purity is required to avoid any stability problems caused by inorganic ions. The salt content needs to be controlled, especially in calcium, which can produce precipitates. These precipitates produce cloudiness in the liqueurs and, when precipitated, can sweep along with them other components of the liqueurs such as pigment or flavoring.

However, chlorine levels should be controlled, since chlorine imparts undesirable flavors. Thus soft water with a hardness below 8 must be used, and for this purpose, demineralized water can be used.

Sugars

Sugars in liqueurs are a very important material and can be obtained from many sources, depending on the availability, price, legality, and final characteristics required.

The most commonly used sugar is sucrose from sugar cane or beet, although other sweet substances such as liquid glucose syrups, high-fructose corn syrup, honey, or rectified concentrated grape must can also be used. In addition, several other sweet compounds may be used, such as polyalcohols and synthetic sweeteners, which are added in order to increase the sweetness and reduce the levels of sugars added, thus reducing the calorific value of some liqueurs.

The best form in which to add sugars, irrespective of their source, is as a liquid solution. In this way, sugars are already dissolved and can mix better with the hydroalcoholic medium. These solutions are free of impurities (having been eliminated previously during their manufacture), thus reducing the risk of cloudiness or haze. In addition, liquid solutions produce a better taste and mouthfeel probably owing to the inversion of sucrose to glucose and fructose.

Sugars can also be added to liqueurs as a thickening or coloring agent. Starch syrup produces thicker liqueurs, with a higher density, without overly increasing the sweetness. Other carbohydrates may be used as thickening agents also. However, when hydrocolloids are applied, their solubility and stability in hydroalcoholic solutions must be tested beforehand.

Liquid caramels and some syrups obtained by heat treatment of crystalline sucrose or other sugar sources can act as colorants. In such cases, temperature must be carefully controlled during the heat treatment in order to avoid the formation of high quantities of furfural and derivatives, the levels of which are legally limited in liqueurs.

Flavoring and color

Flavoring components can be extracted from natural plants such as herbs (barks, roots, seeds, flowers), fruits, spices, and beans such as coffee and cocoa. These vegetal materials add color, taste, and aroma (flavor) to liqueurs owing to the extraction of different compounds present within them. Also, flavoring compounds can be obtained from essential oils, natural-flavoring extracts, or any combination of these sources with different permitted flavoring and coloring food additives. Usually, synthetic flavors are added in cheaper brands of liqueurs.

The color of liqueurs may result directly from the use of plants such as fruits, seeds, leaves, etc. Plain caramel is often used, as it is the most stable color for spirits and liqueurs. For other colors, synthetic colors can be used, where legislation allows.

Fruits can be used whole or peeled, and they should be employed in the optimum grade of maturation in order to achieve their maximum flavor, color, and sugar contribution.

Generally, a large number of brands use synthetic flavors and colorants to adjust or intensify the 'natural' characteristics of their products. It is important to note that natural extracted pigments are more susceptible to degradation than some synthetic colorant additives, which are very useful to maintain the attractive color of liqueurs.

Different methods to obtain the flavor and color from fruits, herbs, spices, and beans can be used. The methods used most often are as follows:

1. Maceration: This procedure involves steeping the aromatic/flavoring agent, which has usually been bruised in a hydroalcoholic solution for a period of time in order to extract an essence from it. This essence is then added to the base as the flavoring agent. This is the traditional method used in higher-quality liqueurs. The herbs used can be fresh or dry. Their extraction is be carried out using alcohol of different strengths. The fresh herbs can be extracted quickly and with a high yield in high alcoholic solutions between 70 and 96%. However, dry herbs should be extracted with an alcoholic solution of lower ethanol concentration (40–60%), because they must be rehydrated to facilitate the extraction of internal pigments and flavored compounds.
2. Infusion: This is similar to the maceration process in which the flavoring agent is steeped in a base alcohol, but in this case, the alcohol is heated. The heat can be maintained for several days, and the result is a more flavorful and less expensive product.
3. Percolation: This method can be done with either a hot or a cold solution. The alcoholic base is allowed to drip though the raw material (or flavoring agent). The alcoholic base can also be heated, and then the formed steam passed through the agent prior to being collected and recondensed to form the base product (extract).
4. Distillation: Usually, this is steam distillation of the desired flavoring agent, which has previously been extracted by maceration. This is often repeated many times with large amounts of the flavoring agent and reducing the volume to a relatively small amount of liquid. This method produces a very strong essence that can be added to the bulk of the alcohol base.
5. During the 1990s, liquid CO_2 extraction was largely improved, and nowadays, this technique has many applications in the food industry. One such application is the manufacture of essential oil extracts, especially from herbs and species, which are applied in the elaboration of liqueurs. The CO_2 extracts are usually better than extracts obtained by steam distillation in terms of quality, yield, and energy savings.

The choice of method depends on the source from which the flavor is to be extracted and on the particular flavor required from the flavoring agent. Some flavoring will yield different flavors, depending on the type of extraction used.

The final products used to be alcoholic bases, which contain the flavor of the vegetal material. Owing to the fact that potentially toxic substances, present in vegetal material, can be extracted together with pigment and flavoring compounds, these extracts must be analyzed and controlled adequately before they are used in liqueur manufacture. Some of these undesirable compounds include thujones mainly extracted from herbs, cianhydric acid and its derivatives, mainly extracted from almonds and drupes, and methanol, mainly present in extracts from fruits with a high pectin content or when pectinases have been applied in fruit treatment, and so on.

Innovations in liqueur manufacture, mainly in response to consumer demands, have become important in fruit-flavored liqueurs, the development of products with flavors that combine well with mixers, increased sophistication and diversity of flavorings, and greater use of natural flavorings.

Elaboration Processes

There are two main types of processes for producing liqueurs, maceration systems and blending systems, which are usually automated with feedback controls. Both systems yield products with similar characteristics, although the final quality in each case is different.

Maceration Systems

The maceration system is the most traditional procedure, which involves a process very similar to the maceration/infusion processes previously described to extract pigments and flavoring compounds from fruits and plants (Figure 1). After the extraction process, liqueurs acquire their flavor and color characteristics, which evoke the raw material or their origin.

Different studies have been carried out in order to improve the extraction of pigments and flavor compounds during maceration. These include optimization of the strength of hydroalcoholic extracting solutions according to the different vegetal raw material, optimization of the relation between the weight of the raw material and the extracting solution volume, and the contact time or duration of the maceration process. In addition, the use of diverse commercial enzymes, especially clarifying pectolytic and color-extracting preparations, have been described. These types of enzyme preparations can improve the final color of extracts and liqueurs and can prevent haze formation, because they hydrolyze, and reduce the levels of pectins in the final products.

Vegetal materials are usually separated from the liquid extract by decantation or filtration, depending on the type and size of the raw material used (whole fruits, whole or powder herbs, etc.). Different types

of equipment can be used, such as plate and frame filters, leaf filters, or sieve systems. The liquid extract is racked off to a blending tank, where sugars or syrup are added; caramel and other ingredients can also be added. The tanks usually have agitation systems to improve the mixture. If necessary, the alcoholic grade can also be adjusted. In addition, several flavor and colorant substances can be added to adjust the final color and flavor.

The final step is bottling. Prior to bottling, liqueurs are maintained in repose. During this period, different settling processes are developed, which can be improved by cooling. Then, liqueurs are filtered to eliminate any precipitates and to yield a product with a high purity and brightness.

Haze formation after bottling is one of the most undesirable effects, as cloudiness greatly reduces the quality of the final product. This haze formation depends on many factors such as the presence of proteins, polyphenols (mainly ellagic tannins), and pectins, which are present in vegetal materials and extracts. Thus, in recent years, different studies on the use of different clarificants such as polyvinylpyrrolidone and adsorbed resins have been carried out. These products are injected directly into alcoholic extracts or liqueurs, and then, the centrifugation or filtration processes are carried out. However, the effect of the clarification agent on the retention of pigments and flavor compounds must be tested beforehand, in order to prevent any decoloration or undesirable losses of flavor.

Continuous Blending Systems

Continuous blending systems are carried out in steel blending tanks connected to other secondary tanks containing the different components of the liqueur, such as water, alcohol, syrup, liquid starch, solution of colorant, and flavor compounds, etc. (Figure 2). Automated valve systems control the volume of every component of the liqueur pumped to the blending tank. Usually, the blending tank has agitation systems to improve the mixture of the ingredients. Sometimes, the blending is stabilized by the homogenization process. This step is necessary, especially when the final product has a high level of essential oils or other compounds with a low water solubility. In addition, juice liqueurs and other cloudy liqueurs can be homogenized, in order to stabilize their turbidity.

After blending, with the exception of cloudy liqueurs, the final steps are similar to those previously described for traditional methods. Because the initial constituents are liquid solutions that have been

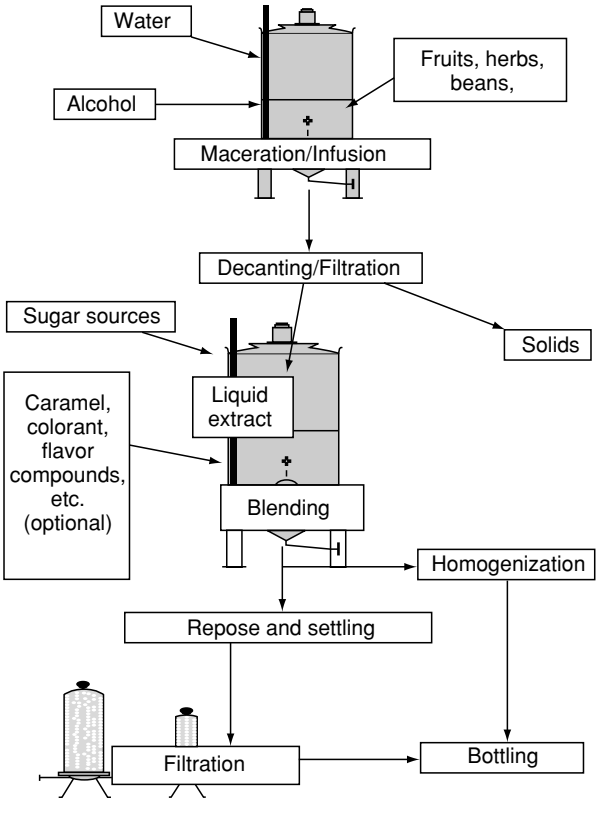

Figure 1 Diagram of traditional (maceration) liqueur-making process.

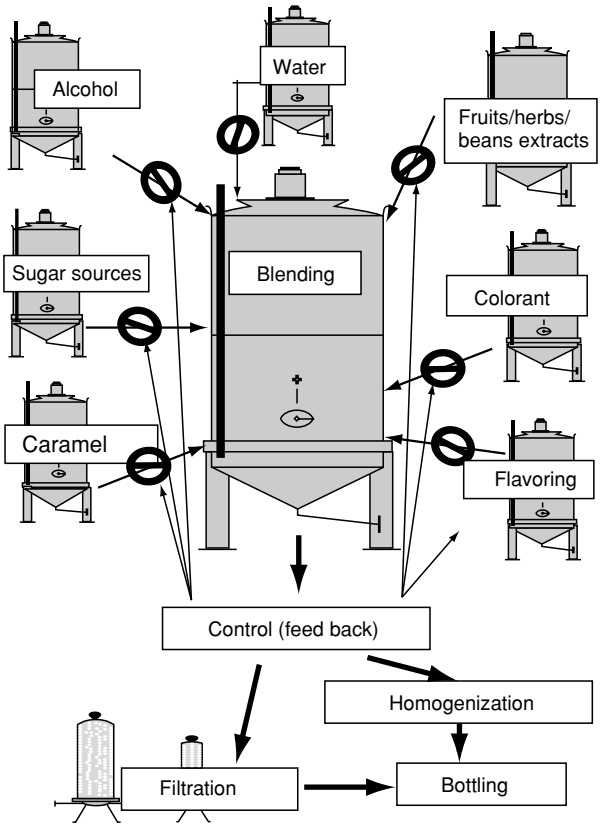

Figure 2 Diagram of continuous blending systems.

adequately treated and filtrated, usually, a few new precipitates are formed during blending. Then, a repose period is optional, and membrane filters are usually employed to increase the brightness of liqueurs.

Modern systems are feedback control systems. From the outlet of the blending tank, an aliquot of the product is sent to the analyzer system, which can be connected directly to the interior of the blending tank (in line control systems). The analysis center confirms whether the product complies with fixed quality parameters such as color, pH, °Brix or sugar content, density, etc. If the product shows any deviation from the allowed limits of any parameter, control systems send a signal to the respective valve to add the component necessary to correct the detected deviation. Only when the liqueur complies with the specifications is the valve of discharge (outlet) of the blending tank opened, and the product is then conveyed to the stabilization process and bottled.

Different software packages, for a wide range of technical calculations in liqueur manufacture, have been developed in the last 5–6 years, and some are used in the aforementioned automatic modern systems.

Types of Liqueurs

The range of liqueurs is enormous; many companies produce a wide range of brands, and different countries produce different national products, probably because of the different raw materials found in different countries. Thus, there are different possibilities to classify liqueurs, such as by strength, sugar content, origin (fruit, beans, or herbs, and type), color (color and colorless), appearance (clear or cloudy), manufacturing process (maceration or blending), etc. It is useful at this point to note the differences between similar products such as 'Cream' and 'Crème' liqueur.

'Cream liqueurs' are thick liqueurs made with cream, milk, or some other dairy product, and their strength is usually lower than that of other liqueurs. Owing to their 'special' composition, their shelf-life is shorter than that of other liqueurs, and after they have been opened, they should be kept refrigerated. 'Crème' (pronounced 'Krem') is a term that is followed by the name of a specific fruit or other vegetal material and is reserved only for liqueurs with a minimum sugar content of 250 g l^{-1}. These liqueurs contain enough sugar to become thick and creamy in texture, but there is no actual cream in them.

According to European legislation, liqueurs that specify fruits or plant materials (including pictorial representation) must contain at least 90% of the cited materials, and only 100% natural products can be used in their elaboration. If the term 'natural' source of flavor is not specified in the label, artificial (synthetic) flavorings may be used. However, USA legislation allows the use of the term 'boosted natural' flavors, which may contain up to 0.1% of artificial flavor components. Some regulations of other countries may also specify different specifications, e.g., Spanish legislation determines that 'juice liqueurs' must contain at least 20% natural juice from the cited fruit.

Recently, 'gelled liqueurs' and 'alcopops' have appeared on the market. 'Gelled liqueurs' are characterized by their gel texture, and an alcohol-free beverage base is mixed with hydrocolloids (e.g., gellan and gelatin) at high temperature, followed by turbulent mixing with an alcohol/water mixture and homogenization to produce a solid gel. This gel is stable at ambient temperature and may be used for direct consumption or as a filling in wafers, chocolates, and other foods. The mouthfeel of the gelled product depends on the composition of the alcoholic beverages and the gelling agents used.

'Alcopops' are liqueurs, usually fruit liqueurs, of low alcoholic content, and have recently appeared on the UK market, being marketed primarily at young people.

Owing to the specific characteristics of certain liqueurs or their manufacturing processes, some liqueurs are protected under specific quality denominations, similar to spirits such as whiskies, brandies, and gins. The most common indication, especially for products elaborated in European countries, is PDO (Protected Designation of Origin).

However, because liqueurs are not a static medium, their composition and characteristics do not remain constant. Their minor constituents, especially pigments and flavoring, undergo a large number of changes, not only during manufacture but also during storage. These are mainly degradation reactions catalyzed by aqueous acidic medium, high temperature, and light (UV irradiation).

Such changes make it difficult to determine the authenticity of these products, and so, in the last decade, many papers on the determination of authentication parameters of different liqueurs have been published. In general, aroma compounds have been studied and described as useful parameters to determine the authentication of liqueurs. Isomers and isotopes studies are very useful, because their presence and relative amounts in natural flavor extracts and synthetic preparations are usually different.

Phenolic compounds and other pigments such as carotenoids have also been evaluated as characteristic parameters of natural liqueurs.

See also: **Alcohol**: Properties and Determination; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Liqueurs**: Cream Liqueurs

Further Reading

Adam L, Bartels W, Christoph N and Stempfl W (1995) *Brennereianalytik. Qualitaetskontrolle und Analytik in Fachlabor*, vols. I and II. Hamburg: B. Behr's Verlag.

Bryne RM (1993) *How to Make Liqueurs*. Kenthurst, Australia: Kangaroo Press.

Clutton DW (1995) Speciality products. In: Lea AGH and Piggott JR (eds), *Fermented Beverage Production*. Glasgow: Blackie, pp. 301–324.

Durkan A (1998) *Spirits and Liqueurs*. Lincolnwood: NTC Contemporary Publishing.

Hunt J (2000) Alcoholic fruit drinks grow up. *Food Review* 27: 21–22.

Jeffers HP (2000) *High Spirits: a Celebration of Scotch, Bourbon, Cognac and More*. Collingdale: DIANE Publishing.

Long C and Kibbey H (1996) *Classic Liqueurs: The Art of Making and Cooking with Liqueurs*. Lake Oswego, OR: Culinary Arts.

Meilach DZ (1996) *The Best 50 Homemade Liqueurs*. San Leandro, CA: Bristol Publishing Enterprises.

United States Code (1994) *Intoxication Liquors and Title 28; Judiciary and Judicial Procedure*. Washington, DC: United States Government Printing Office.

Walton S (1999) *Complete Guide to Spirits and Liqueurs*. New York: Anness Publishing.

Walton S and Miller N (1997) *Spirits and Liqueurs Cookbook*. New York: Lorenz Books.

Walton S and Miller N (2000) *Encyclopedia of Spirits and Liqueurs and How to Cook With Them*. London: Southwater Books.

Wustenfeld H and Haeseler G (1996) *Trinkbranntweine und Likoere*. Berlin: Blackwell Wissenschafts Verlag.

Cream Liqueurs

B T O'Kennedy and W J Donnelly, National Dairy Products Research Centre, County Cork, Ireland,

Definition of Product

Cream liqueur is an alcoholic beverage containing emulsified milk fat. The first of the new generation of cream liqueurs which now dominate that market is manufactured in Ireland and marketed under the name Bailey's Irish Cream. Similar products are now manufactured in several countries. A standard definition of cream liqueur does not exist internationally and since there can be wide variations in the composition of liqueurs from different manufacturers only the broadest technical definition is possible. For marketing and regulatory purposes more precise definitions exist or are contemplated in several countries. An example is the proposed standard for Irish cream liqueur currently under discussion among manufacturers in Ireland, which specifies among other things that the product must contain 12% (w/v) minimum butterfat, 15% v/v minimum alcohol, $100\,g\,l^{-1}$ minimum sugar, and have an 18 months' shelf-life.

Table 1 Composition of a typical cream liqueur

Ingredient	*Percent*
Milk fat	16%
Sugar	20%
Sodium caseinate	3%
Nonfat milk solids	1%
Total solids	40%
Ethanol	14%
Trisodium citrate	$10\,mmol\,l^{-1}$
Water	46%

Percentage values are on a w/w basis.

A typical cream liqueur composition is shown in **Table 1**. Most of the major liqueurs approximate the typical composition.

Basis of Formulation

High-quality cream is a prerequisite for the production of consistent quality and stability. For the most part, the fat in the cream acts as an inert diluent, imparting a creamy texture, the absence of which can be quite noticeable when the fat is reduced. Homogenization reduces fat globule size such that creaming is prevented on storage. Most commercial cream liqueurs incorporate sodium caseinate with the purpose of stabilizing the newly formed fat emulsion. The use of small amounts of low-molecular-weight surfactants such as glycerol monostearate (GMS) is also widespread in commercial practice. While sodium caseinate would be recognized as the prime emulsifier, the effect of GMS can be significant in the subsequent behavior of the product. The presence of 14% by weight of ethanol has a marked effect on sensory properties and is an effective control on microbial spoilage. The presence of large quantities of sucrose in the formulation is central to the taste and organoleptic properties of the product. Besides its obvious importance to the sensory characteristics of the product, sucrose can also have direct effects on the behavior of the other components and on such physical properties as refractive index or viscosity of the aqueous phase. (*See* **Casein and Caseinates**: Uses

in the Food Industry; **Emulsifiers**: Uses in Processed Foods; **Sucrose**: Properties and Determination.)

During the production of cream liqueurs, the concentration of some of the components in the system may exist at a much higher level than is apparent in the final product. This is due to the order of addition and the water or alcohol content at various stages during formulation. Because multicomponent interactions are seen as the basis for consistency of product, regulation of the water activity at any given step in the production process, within the constraints of practicality, can be a means of altering final product characteristics.

Manufacturing Protocol

A typical manufacturing process for cream liqueur is shown in Figure 1. This involves the blending of sodium caseinate solution with a cream/sugar solution, followed by homogenization and alcohol addition. Considerable variation is likely in the sequence of different steps and additions and in the precise composition of the final mixture. Some of these variations may influence the characteristics of the final product. Use of citrate as stabilizer has a dramatic effect on shelf-life and is to be recommended. Use of an emulsifier is optional since adequate emulsion stability can be provided by caseinate alone, provided sufficiently high pressures are used in the emulsification step. In practice, moderate pressures are preferred and hence emulsifiers are desirable. The homogenization step may be carried out before or after alcohol addition. Homogenization with alcohol present gives more efficient emulsification, but also imparts a different mouth feel. Depending on the homogenization conditions, the stage of emulsifier addition may also be critical. (*See* **Stabilizers**: Types and Function.)

Characteristics Expected in End Products

Following the manufacturing processes outlined, it is expected that the cream liqueur will be stable to excess creaming, gelation, or deposit formation for a period in excess of 18 months. However, most exporters will take precautions to insure that the product is not subjected to extremes in temperature. Exposure to temperatures in excess of 40 °C for any period of time may result in a dramatic thickening of the product. The viscosity of the product following manufacture is one of the main textural characteristics of the product, typically being in the region of 30 mPa at moderate shear rates and 20 °C. Cream liqueur is very subject to shear thinning, having a high viscosity at very low shear rates but becoming instantly pourable on disturbance. It may therefore be regarded as having the structure of a weak particle gel.

The mouth feel and textural parameters of the product are modified by fat globule distribution. Different products generally show a difference in size

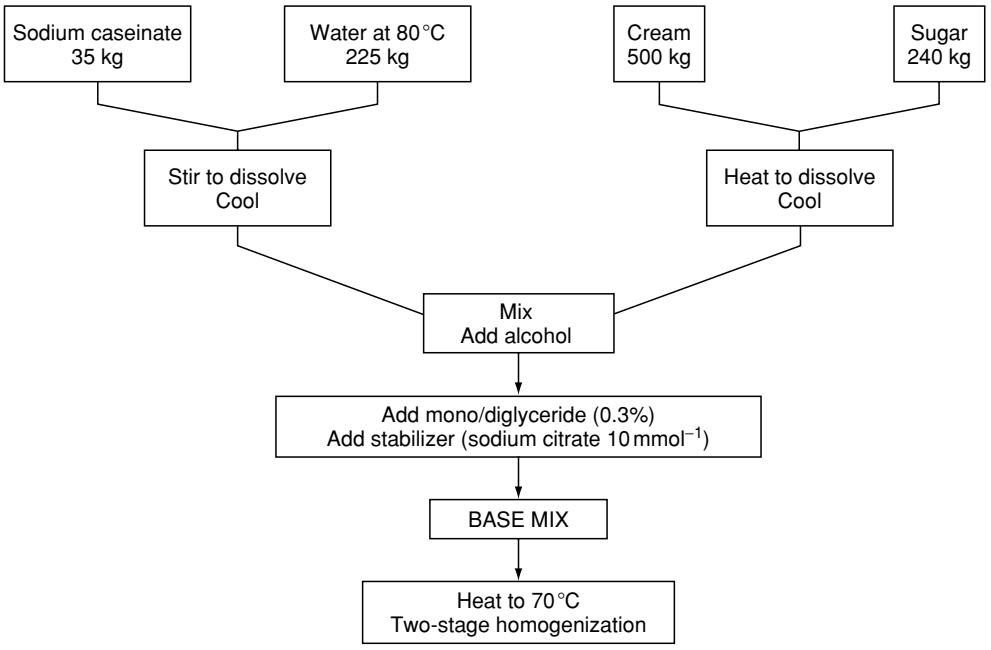

Figure 1 A flow diagram for a typical cream liqueur manufacturing process. Reproduced from Liqueurs: Cream Liqueurs, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

distribution, most probably due to differences in processing and homogenization regimes. The opacity and glass-holding ability are also directly related to fat globule size distribution; the smaller the fat globules, the more translucent or less opaque the product can become.

A characteristic of cream liqueurs to a greater or lesser extent, which is very noticeable to the consumer, is the dramatic whitening effect that occurs on dilution with neutral mixers, or on addition of ice. This is due to a change in the refractive index of the aqueous phase, emphasizing the light-scattering properties of the fat globules. The choice of sodium caseinate as the proteinaceous emulsifier makes the liqueur unstable to acid. This characteristic is of some importance as the addition of acid mixers to the system may lead to coagulation and consumer rejection. This can be overcome by omitting the caseinate and relying on low-molecular-weight surfactants as a means of stabilizing the emulsion. However, a loss in product quality is probable.

Storage/Stability Problems

There are two main problems of instability associated with cream liqueurs. These are creaming and age thickening. Creaming arises because of inadequate fat dispersion and is controlled by varying homogenization conditions during manufacture. The important variables are homogenization pressure, temperature, and number of homogenization stages. Normally, manufacturers employ a two-stage process, the first stage being at intermediate pressure (about 20 MPa) and the second stage at low pressure (about 40 MPa). Alternatively, multiple-stage high-pressure (about 30 MPa) homogenization may be employed. When homogenization is carried out after alcohol addition, fat dispersion is improved. Since there is a relationship between homogenization conditions and liqueur characteristics such as mouth feel, opacity, and glass adhesion, the precise conditions used will represent a balance between stability and sensory requirements.

A much more complex instability problem is age thickening. This is generally accompanied by phase separation and emulsion breakdown. Age thickening is best viewed as protein aggregation, involving complex interactions between casein in the continuous phase and at the fat/water interphase (recombined fat globule membrane). Aggregation is promoted by ionic calcium and by alcohol. An important additional destabilizing factor is the process of homogenization itself, which, in some manner not yet understood, enhances the susceptibility of casein to aggregation. The severity of the homogenization effect increases with fat content, and liqueurs with lower than normal fat are less likely to show instability problems. The simplest remedy for problems of protein instability is addition of a calcium complexing agent such as sodium citrate. Liqueur pH may also be a determining factor and small pH adjustments may prove beneficial for individual manufacturers.

Recently, it has been shown that the apparent viscosity of cream liqueurs on storage at 45 °C was dependent on the sodium caseinate source and suggested that electrostatic and sulfhydryl interactions were involved. This was further substantiated by information indicating that certain fractions of the sodium caseinate ingredient were far more susceptible to alcohol-induced aggregation reactions. The alcohol-susceptible fraction contained sulfhydryl groups.

Other stability problems encountered from time to time are deposit formation, lactose crystallization, or citrate precipitation. These are formulation-dependent and in most instances the remedy is slight adjustment of the concentrations of individual components.

See also: **Casein and Caseinates**: Uses in the Food Industry; **Emulsifiers**: Uses in Processed Foods; **Stabilizers**: Types and Function; **Sucrose**: Properties and Determination

Further Reading

Banks W, Muir DD and Wilson AG (1981) Extension of the shelf-life of cream-based liqueurs at high ambient temperatures. *Journal of Food Technology* 16: 587–595.

Banks W, Muir DD and Wilson AG (1982) Formulation of cream-based liqueurs: a comparison of sucrose and sorbitol as the carbohydrate component. *Journal of the Society of Dairy Technology* 35(2): 41–43.

Donnelly WJ (1987) Ethanol stability of casein solutions as related to storage stability of dairy-based alcoholic beverages. *Journal of Food Science* 52(2): 389–393.

Lynch AG and Mulvihill DM (1997) Effects of sodium caseinate on the stability of cream liqueurs. *International Journal of Dairy Technology* 50(1): 1–7.

Liquid Properties of Food *See* **Rheological Properties of Food Materials**

LISTERIA

Contents

Properties and Occurrence

M W Griffiths, Canadian Research Institute for Food Safety, University of Guelph, Guelph, ON, Canada

Introduction

Although an organism, later to be identified as *Listeria monocytogenes*, was isolated from cerebrospinal fluid as early as 1918, characterization of the bacterium which we now know as *L. monocytogenes* did not take place until 1926. It is only since the 1980s that the importance of food in the transmission of listeriosis has been acknowledged. There was a suspicion that the fecal–oral route was involved in the spread of listeriosis as an enlightened Russian scientist had written as early as 1953 that 'The most important way of the infection of people is the food-borne one.' Nevertheless, as late as 1976, *L. monocytogenes* was not listed among bacteria that could cause foodborne disease by a World Health Organization/Food and Agriculture Organization expert committee.

Listeria species are short, Gram-positive, nonspore-forming, facultatively anaerobic rods. *L. monocytogenes* is motile at 20–25 °C but nonmotile at 37 °C because the gene encoding flagellin is not transcribed at this higher temperature. It can grow over a wide range of temperatures, at low pH and at low oxygen tensions. Thus, it is able to grow in many environments.

The genus *Listeria* contains six species (*L. monocytogenes, L. innocua, L. ivanovii, L. seeligeri, L. welshimeri*, and *L. grayi*), of which only *L. monocytogenes* can cause infections in humans. Analysis of the genome of *L. monocytogenes* indicates that the bacterium exists in at least three evolutionary lines which are strongly conserved. However, most isolates are distributed between two of these lineages.

Occurrence in Foods and the Environment

Environmental Isolation of *Listeria* spp.

L. monocytogenes is widely distributed in the environment. It has been commonly isolated from surface waters, decaying vegetation, soils, sewage, and silage. It has been found on all continents except Antarctica.

Listeria can grow in sterile soil samples, but does not compete well in the presence of natural soil microflora. In several instances it has been demonstrated that *Listeria* spp. can survive in soil for several years. It is likely that the principal source of the bacterium in soil is from animal feces. Listeriosis is recognized most frequently in ruminants, but *L. monocytogenes* has been isolated from more than 37 species of mammals and 17 species of birds. The occurrence of the disease in ruminants has been strongly associated with silage feeding as the organism is rarely detected on growing grass and vegetables. The incidence of *L. monocytogenes* in silage increases with the use of 'big bale' silage production and in silage contaminated with mold. It has also been postulated that birds are a possible source of listerial contamination of silage. Fecal samples from gulls feeding at sewage sites were reported to have a higher carriage rate for *Listeria* spp. than those from birds feeding elsewhere.

The excretion of *L. monocytogenes* by farm animals is linked to their diet. Animals fed entirely on hay or manufactured diets do not excrete detectable levels of *Listeria*, but animals fed on silage commonly excrete the organism.

The practice of applying sewage to agricultural land may also create a risk of listerial contamination. It has been calculated that 40% of the total sludge produced at sewage works in England and Wales may be applied to agricultural land. Sewage sludge can contain large numbers of *L. monocytogenes*, with counts ranging from 800 to $>18\,000\,l^{-1}$. The numbers of the bacterium showed no detectable decrease 8 weeks after spraying. Isolation of *Listeria* spp. from nonagricultural, residential, suburban communities indicates that the organism could be associated with areas other than those involved in agriculture.

L. monocytogenes has been isolated from effluent, not only from sewage treatment plants, but also those of poultry-packing factories, abattoirs, and cattle markets. Surface waters receiving these discharges

may be a good route for recycling these bacteria. *L. monocytogenes* was isolated from 62% of fresh or low-salinity waters in tributaries draining into an estuary impacted by humans and domesticated and wild animals. One *L. monocytogenes* serogroup appeared to predominate in fresh water when domesticated animals were nearby, but a greater variety of strains with no predominance by a single serogroup was observed in areas with no direct animal influence. The majority of strains found on decaying plants are not animal pathogens and this suggests that the predominance of certain serotypes of *L. monocytogenes* may depend on environmental factors.

Presence in Food

L. monocytogenes has been isolated from a wide variety of foods of animal and plant origin (Table 1). The incidence of the bacterium in foods will be described on a commodity basis.

Occurrence in Milk and Dairy Products

Because of the well-documented foodborne outbreaks of listeriosis associated with the consumption of dairy products, these foodstuffs have attracted the attention of many researchers. Several studies have been carried out across the world to assess the incidence of *L. monocytogenes* in raw milk. Five large studies were carried out in Pennsylvania, USA (2511 samples), England and Wales (2009 samples), Ontario, Canada (1720 samples), France (1409 samples), and Transvaal, South Africa (961 samples). These studies revealed the incidence of *L. monocytogenes* to be 3.1, 5.1, 2.7, 6.0, and 5.2%, respectively. Collectively these investigations found that 363 of 8610 samples or 4.2% of raw milks were positive for *L. monocytogenes*. The level of the organism in raw milk is usually less than 1 cfu ml^{-1} but counts may reach 64 cfu ml^{-1}. The source of *L. monocytogenes* in the milk supply is still in some doubt. *L. monocytogenes* is a causative agent of mastitis, albeit only rarely. The organism has been recovered from milk of mastitic cows at counts ranging from 10^3 to 10^4 cfu ml^{-1}. Prolonged excretion of *L. monocytogenes* in milk and feces of infected animals and asymptomatic carriers has been described. This would lead to the build-up of an endogenous pool of *Listeria* in the environment which could be augmented from other sources such as the use of poor-quality silage and feces of feral animals and birds. In a large case-control study involving 128 dairy farms in France, poor-quality silage (pH > 4), inadequate frequency of cleaning the exercise area, poor cow cleanliness, insufficient lighting of milking barns and parlors, and incorrect disinfection of towels between milkings were associated with contamination of milk by *L. monocytogenes*.

Numerous other dairy products have been found to contain *L. monocytogenes*. The bacterium has been found in pasteurized milk and an outbreak of listeriosis in Massachusetts in 1983 was associated with the consumption of this product. One report from Spain suggested that 21% of pasteurized milk samples tested

Table 1 Occurrence of *Listeria monocytogenes* in some foods

Food type	*Samples contaminated with* L. monocytogenes *(%)*
Dairy	
Raw milk	0–45.3
Pasteurized milk	0–21.4
Flavored milks	0.8–3.3
Creams	0
Soft and semisoft cheeses	0–61
Hard cheeses	0–9
Icecream	0.3–6.1
Butter	0
Yogurt	0
Powder	0
Sheep and goats' milk	0.1–2.9
Meat and poultry	
Raw beef	0–30.9
Raw lamb	0–60
Raw pork	0–68
Ground meats	11.7–92
Ham	11.8–33.8
Sausages	15.7–88
Fermented, dried sausages	0–56.7
Frankfurters	7.5
Processed meats	0–72.2
Pâté	2.76–10.7
Raw poultry	3–58.6
Cooked poultry	0–27.2
Eggs	0
Seafood	
Raw fish	0–33.3
Smoked salmon	0–44.4
Uncooked shellfish	12.1–25.4
Cooked shellfish	0–28.1
Vegetables	
Beansprouts	12.1–85.7
Cabbage	1.1–2.1
Cucumber	2.2–80
Potatoes	21.2–27.1
Prepacked salads	0–14.3
Radishes	14.4–36.8
Salad vegetables	2.3–44
Spring rolls	62.5
Tomatoes	13.3
Vegetables	6.2–12.2
Frozen vegetables	2.2
Other	
Ready-prepared meals	2.9–13.2
Frozen meals	11.9
Vegetarian meals	31.8
Mayonnaise	4.2–5.2

were positive for the organism. However, these milks showed evidence of gross contamination with other bacteria, indicative of poor hygiene conditions in the processing plant and of postpasteurization contamination, e.g., at bottling/packing. More reliable estimates put the level of contamination of pasteurized milk with *L. monocytogenes* at <1%. Other fluid milk products, such as chocolate milk, can have *L. monocytogenes* present. The isolation of *L. monocytogenes* from pasteurized milk prompted investigators to suggest that, because of its intracellular location, it was able to survive pasteurization. However, many studies have now confirmed that this is not the case (**Figure 1**).

Dairy products most at risk of contamination with *L. monocytogenes* are soft cheeses such as Camembert and Brie. Again, numerous surveys have been carried out worldwide to determine the incidence of listerial contamination of these products. These studies have revealed a widespread distribution of the organism with *L. monocytogenes* counts in the range 1×10^2 to $1 \times 10^6\,g^{-1}$. In a study carried out in the Netherlands, all positive cheeses were manufactured from raw milk, but other work has shown no difference in the incidence of *L. monocytogenes* in cheese produced from raw or pasteurized milk. The contamination of Swiss cheese with *L. monocytogenes* was primarily due to the presence of the organism on shelves used for maturation of the cheeses. On rare occasions, hard cheeses, including Cheddar, have become contaminated with the bacterium.

Among other dairy products shown to harbor *L. monocytogenes* are icecream and butter, although there have been no documented cases of listeriosis where icecream was the vehicle of infection.

Since the recognition of the link between listeriosis and consumption of contaminated milk products, research has focused on the source of this contamination. In one such study the incidence of *Listeria* contamination of processing equipment was compared with that of the general dairy-processing environment. Positive equipment samples were obtained from six of the 21 (28.6%) plants, whereas 19 of the 21 (90.5%) plants had positive environmental sites. *L. monocytogenes* was isolated from 11 of the 215 (5.1%) samples taken from equipment, and these included three holding tanks, two table tops, three conveyor-chain systems, a pasta filata wheel, a pint milk filler and a brine prefilter machine. Twenty-four of the 163 (14.7%) samples from environmental sites were positive for *L. monocytogenes*. These results suggest that environmental contamination with *Listeria* does not necessarily translate into contamination of equipment within the same dairy plant. Sites that are typically positive for *L. monocytogenes* include drains, conveyor belts, floor mats and foot baths, coolers, freezers, areas underneath equipment, areas associated with case washers, areas where raw milk is handled, cheese-ripening rooms, and cheese-washing brushes.

In addition, the processing environments of 30 dairies were surveyed for the presence of *Listeria* spp. Nine of these dairies had a dairy farm in close proximity to the processing facilities, and these plants were more likely to be contaminated (9/9) than the plants without on-site dairy farms (17/21). Plants producing dairy ingredients, frozen dairy products, or fluid milk had significantly higher incidence rates than expected, whereas plants producing fermented dairy products, or a combination of fermented dairy products and fluid milk, had significantly lower incidence rates than expected. There was no significant difference in contamination rates between areas within the plants.

Occurrence in Meat and Meat Products

L. monocytogenes has been isolated from beef, lamb, pork, turkey, and chicken with high frequency (Table 1). There is considerable evidence that the contamination of meat arises from environmental sources which can occur at all points along the processing and retail chains. Common sources of contamination are the environment of the chilling and cutting rooms, hands, and equipment. Cross-contamination may also occur between raw and cooked product, and from contact of cooked products with soiled surfaces due to inadequate cleaning and disinfection.

The source of contamination can vary depending on meat species. For beef and lamb, the hide seems to be a more important source of carcass contamination by the bacterium than feces, but for pork, fecal contamination is more important. Poultry carcasses can become contaminated from the environment of processing plants or due to carriage of the organism in the gut of healthy birds. *L. monocytogenes* can contaminate poultry at several points during production, including at evisceration due to spillage of gut contents, from handling, from equipment such as the rubber fingers used for plucking, and from the scald tank water. *L. monocytogenes* can persist in the environment of poultry-processing plants for up to 1 year.

Levels of *L. monocytogenes* found in cooked poultry products can be high ($>10^4$ cfu g^{-1}) and this may be the result of inadequate cooking or growth of the bacterium arising from contamination after the cooking step. However, most studies have shown that when *Listeria* is present in meats it is usually found in low numbers. The level of contamination is generally

higher for poultry than pork, which, in turn, is higher than in beef. There is evidence that *L. monocytogenes* apparent numbers are affected by the microbiological methods applied, especially where cells might be injured.

L. monocytogenes is widely distributed in meat-processing plants and the conditions present in these locations are ideal to promote the growth of the organism. This makes it difficult, but not impossible, to control contamination of the meat.

Occurrence in Fish and Seafood

Because of the many routes by which *Listeria* can contaminate water, it is not surprising that *L. monocytogenes* has been isolated from fish and seafood at high prevalence rates. The organism can become widely disseminated throughout smokehouses during production of smoked salmon and *L. monocytogenes* can be isolated from many sites in a raw-shrimp-processing plant. These include processing water, equipment, food handlers, doors, walls, floors, and drains. Thus, it is essential to institute a regular cleaning and disinfection program within these operations.

A study in the UK which examined 4435 seafood samples found that *L. monocytogenes* could be isolated from 128 (3%) of these samples. The organism was most frequently isolated from smoked fish and fish pâté. The extent of contamination was generally low, but in eight samples *L. monocytogenes* counts exceeded 10^2 cfu g^{-1}.

Occurrence in vegetables

As has been previously stated, *L. monocytogenes* is found in many habitats, including soil, water, manure, decaying vegetation, and sewage effluents. This provides many opportunities for the contamination of vegetables by this bacterium. Despite this potential for contamination, surprisingly few types of vegetables are consistently found to contain *L. monocytogenes*. These are mainly potatoes and radishes. The incidence of the organism on fresh-cut vegetables varies between 0 and 19%, and when *L. monocytogenes* is detected it is present in low numbers ($< 2 \times 10^2$ cfu g^{-1}).

Occurrence in Processing Plants and Kitchens

Listeria spp. are widely distributed in the environment of food-processing plants (Table 2). Common sources of these bacteria include drains, floors, standing water, and food residues. The absence of *Listeria* spp. from dry culinary food units prompted one group of researchers to conclude that dry conditions and the absence of food residues contribute to the control of *L. monocytogenes*. Codes of practice for the food industry have been set out to help eliminate problems caused by *L. monocytogenes*. These entail a rigid program encompassing personal hygiene, separation of raw material and product, and effective clean-up and disinfection procedures. These, coupled with the adoption of a hazard analysis critical control point (HACCP) management system operated by trained and knowledgeable staff, should help to insure the microbiological safety of foods.

Even when the product entering the home is free from contamination, there is considerable scope for mishandling foods in domestic kitchens. One survey has reported that 20% of domestic environments are contaminated with *Listeria* spp., with dishcloths being the main culprits. *L. monocytogenes* was also isolated from a domestic refrigerator.

Table 2 Occurrence of *Listeria* spp. and *L. monocytogenes* in environmental samples taken from food-processing plants[a]

Sampling point	*Samples positive (%) for*	
	Listeria *spp.*	L. monocytogenes
Drains	21	7
Floors	43	4
Condensate/stagnant water	11	1
Residues	25	5
Processing equipment/ food contact surfaces	30	19
Conveyors	53	26
Effluent	69	47
Miscellaneous[b]	16	3
All samples	18	6

[a]Environmental samples taken from milk-processing factories (including fluid milk, cheese, icecream, cultured products, and powder plants), frozen food, meat (including poultry)-processing and potato-processing sites.
[b]Miscellaneous sampling sites included wooden pallets, soil, and cleaning materials.

Fate During Processing

Temperature

The identification of pasteurized milk as a vehicle in the transmission of listeriosis led to the assumption that *L. monocytogenes* was unusually heat-resistant. It was thought that this resistance was mediated by its intracellular location within the leukocytes found in milk. This provoked much interest in the thermostability of the organism. It is now commonly accepted by most researchers that it is unlikely that *L. monocytogenes* can survive commercial pasteurization conditions in appreciable numbers and there is little difference in the heat resistance of free and intracellularly located cells.

Results of several researchers who examined the effect of heat on *L. monocytogenes* have been analyzed and, in general, the heat stability of *L. monocytogenes* can be described by the equation:

$$\log D = 13.64 - 0.19\ T \qquad (1)$$

where D is the time in seconds to achieve a 10-fold reduction in bacterial number, and T is the heating temperature in degrees Celsius. Representative D-values at a variety of temperatures and in different foods are given in Table 3.

The heat resistance of this bacterium is increased after sublethal heat shock. This increase in resistance due to sublethal injury can also be triggered by prior exposure to other stressors, such as low pH. As well, the heat resistance of *L. monocytogenes* may be increased severalfold on cured meats, and this increase may not be entirely due to changes in water activity (a_w) in the presence of curing salts.

The method used to recover heat-stressed cells can also have an impact on the apparent heat-resistance of the bacterium. For example, the use of anaerobic conditions leads to greater recovery of heat-stressed cells than culture in the presence of oxygen.

Table 3 Heat resistance of *Listeria monocytogenes* in various products

Product	*Temperature (°C)*	D-*value (min)*	z-*value (°C)*
Ground meat	60	3.12	5.3
Fermented sausage	60	9.2–16.7	4.6
Roast beef	60	3.5–4.5	7.2
Beef	60	3.8	
Beef	70	0.14	
Beef homogenate	60	6.3–8.3	
Naturally contaminated beef	60	1.6	
Beef steak	60	6.3–8.3	6.0
Beef steak	70	0.1–0.2	
Weiner batter	60	2.3	
Chicken leg	60	5.6	6.7
Chicken leg	70	0.11	
Chicken breast	60	8.7	6.3
Chicken breast	70	0.13	
Chicken	60	5–5.3	6.7–7.4
Chicken	70	0.16–0.2	
Chicken homogenate	60	5.0–5.3	
Raw liquid whole egg	60	1.5	6.6
Crabmeat	60	2.6	8.4
Mussels	60	5.5	4.3
Crawfish tail meat	60	2.0	5.5
Salmon fillets	60	4.2–4.5	5.6–6.7
Cod fillets	60	2.0	5.7–6.1
Lobster	60	2.4	5.0
Carrot homogenate	60	5.0–7.8	
Milk and cream	52.2	24.1–52.8	5.5
Milk and cream	57.8	4.0–8.2	
Milk and cream	63.3	0.2–0.6	
Milk and cream	66.1	0.1–0.3	

The organism is resistant to freezing. The viable populations of four strains of *L. monocytogenes* decreased by only 3–6% after 14 days' storage at −18 °C. In another study, the effects of freezing and subsequent frozen storage at −18 °C on the viability of a mixed inoculum of seven *L. monocytogenes* strains in ground beef, ground turkey, frankfurters, canned corn, icecream mix, and tomato soup was determined. *Listeria* responses were related to the pH of the food. Where the pH of the food was >5.8, *L. monocytogenes* survived freezing and storage well. However, in tomato soup, where the pH was 4.74, the organism showed a decline in viability during prolonged storage and could not be quantitatively recovered on selective media. *L. monocytogenes* can also survive in seafood during storage at −20 °C for 3 months. Under these conditions populations on the frozen products decreased by less than 1 log after 3 months.

Sanitizers

A number of approved sanitizers, including acid anionic detergents, quaternary ammonium compounds, carrier-bound iodine, and chlorine, all reduced *L. monocytogenes* populations by a minimum of 99.999% at the recommended doses. Products showing a high efficacy contained iodine, peroxide, or quaternary ammonium as the active agent. However, the efficiency of sanitizers can be reduced in the presence of organic material. There is also evidence that repeated exposure to quaternary ammonium compounds can result in the acquisition of resistance to the sanitizer. The mechanism of action appears to be by the use of efflux pumps which can reduce intracellular concentrations of the sanitizer.

In addition, *Listeria* spp. can attach to processing surfaces where they may form biofilms. Cells present in biofilms have increased resistance to desiccation and the effects of disinfectants. Resistance of adherent cells to sanitizers is dependent upon the surface to which they are attached. Complete biofilm removal and/or inactivation was obtained in many cases where the surface was first cleaned prior to exposure to sanitizer.

Survival in Foods

L. monocytogenes has the ability to survive in a wide variety of foodstuffs (Table 4). The main factors affecting survival appear to be pH, salt concentration, and a_w, although the resistance of the organism to drying is also affected by temperature and relative humidity, as well as the nature of the suspending menstruum. Survival times are generally longer at lower temperatures. The organism can survive pH

Table 4 Survival of *Listeria monocytogenes* in environmental and food samples

Product	*Approximate inoculum size (cfu ml^{-1} or g^{-1})*	*Storage temperature (°C)*	*Final pH*	*Change in cell numbers*[a]	*Survival period (days)*
Environmental samples					
Cooling water	10^4	4		+/–	> 70
Tiled surface	10^8	?		?	> 30
Waxed cardboard/ plastic cartons	–	−0.8/6.6		–	2–14
Dairy products					
Raw milk[b]	10^4	4	–	+++	> 16
Fermented milk	10^3–10^4	4	4.2–4.7	– – –	14–49
Buttermilk	10^4	4	4.4	– –	18–26
Evaporated milk	10^4–10^7	7/21	–	+++	28–56
Sweetened condensed milk	10^4–10^7	7/21	–	–	> 42
Yogurt	10^2–10^7	4/5	4.1	– – –	1–27
Nonfat dried milk	10^5–10^6	25	–	– – –	28–70
Butter	10^4–10^5	−18	–	–	> 70
		4/6	–	+	> 70
		13	–	++	> 70
Icecream	10^0	−12	–	+/–	> 42
Cheeses					
Blue	10^5	10	6–6.5	+	> 168
Brie	10^5	4	5.5–6	– –	> 120
Camembert	10^5	6	7	+++	> 56
Cheddar	10^5	6/13	5.1	– –	154–434
Colby	10^5	4	5.1	–	> 140
Cold pack	10^5	4	5–5.5	+/–	> 182
Cottage	10^5	3	5.2–5.3	– – –	3–28
Feta	10^4	4	4.3–4.6	–	> 90
Goats' milk	10^6	12	5.5–6.5	+/–	> 126
Gouda	10^3	13	6	+/–	> 42
Maasdam	10^3	4	5.5	+/–	> 42
Manchego type	10^3–10^5	15	5.1–5.8	+/–	> 60
Red smear	10^2	5	7.4	+++	> 91
White brined	10^2	10/12	4.5–5.1	+/–	> 42
White mold ripened soft	10^2	5	7.7	+++	> 91
Meat and meat products					
Beef					
High-moisture	10^4	5	–	+++	> 42
Low-moisture	10^4	5	–	– –	> 42
Vacuum-packed	10^3–10^5	0/10	–	++	> 50
		−2	–	+/–	> 60
Carbon dioxide-packed	10^3–10^5	5/10	–	+++	> 21
		0/2	–	– – –	NS[c]
Spray-chilled	10^6	5	–	–	> 16
Ground	10^5–10^7	4	5.6–5.9	+/–	> 30
Roast	10^{-2}–10^2	4.4	4.6–5.3	+/–	> 70
Liver	10^5	4	–	+/–	> 30
Chicken, sliced	10^0–10^3	4.4	6	+++	> 28
Ham	10^{-1}–10^4	4.4	5.1–6.2	+++	> 42
Lamb	?	−10/−23	–	– – –	> 20
Turkey, sliced	10^{-2}–10^3	4.4	5–5.8	+++	> 42
Sausage					
Bologna	10^{-2}–10^2	4.4	5.1–6.2	++	> 42
Bratwurst	10^{-2}–10^3	4.4	5.4	+++	> 42
Finnish fermented	10^6	10	4.6	– – –	> 21
Norwegian fermented, dry	10^5–10^6	4	4.9	–	> 170
		20	4.9	– – –	< 170
Pepperoni	10^5	4	4.5	– – –	> 56
Salami	10^3–10^4	4	4.3–4.5	– – –	> 84
Summer	10^{-2}–10^2	4.4	4.6–5	+/–	> 84
Wieners	10^{-2}–10^3	4.4	4.4–5.8	+++	> 63
Ravioli	10^5	5	–	–	> 14
Egg, whole pasteurized	10^3	4/10	–	+++	> 42

Continued

Table 4 Continued

Product	*Approximate inoculum size (cfu ml^{-1} or g^{-1})*	*Storage temperature (°C)*	*Final pH*	*Change in cell numbers*[a]	*Survival period (days)*
Seafood					
Shrimp, raw	10^5	−20	–	?	> 90
Fruit and vegetables					
Apple cider, unpasteurized	10^5–10^6	4	3.3–3.5	– – –	> 21
		10		– – –	< 7
Asparagus	10^4	4	–	+/–	> 14
Broccoli	10^4	4	–	+/–	> 14
Cabbage	10^4	5	–	+++	> 64
Cauliflower	10^3	4	–	+/–	> 14
Lettuce	10^4–10^5	5/25	–	+	> 14

[a]Symbols represent change in *L. monocytogenes* counts during storage: > 3 $\log_{10}$ increase (+++) or decrease (– – –) in count; 1–3 $\log_{10}$ increase (++) or decrease (– –) in count; 0–1 $\log_{10}$ increase (+) or decrease (–) in count; no change in count (+/–).
[b]Naturally contaminated raw milk.
[c]NS; no survival.

values of 3.6 for up to 25 days if kept at refrigeration temperatures. However, survival is dependent on the acidulant used. There are also reports indicating that some strains of *L monocytogenes* can survive salt concentrations as high as 25.5% for at least 32 days when stored at 4 °C.

Because it is essential to minimize recontamination of processed foods from the environment, processors must be aware that once *L. monocytogenes* is present in a food-processing plant it can persist for long periods of time (> 136 days), especially if the temperature is low and the bacterial cells are protected by residual food components.

Irradiation

In studies carried out using chicken breast meat, *L. monocytogenes* was found to be the most resistant to irradiation when compared to six other common foodborne pathogens. However, no significant differences were observed in the radiation resistance of *L. monocytogenes* in different meats. The D_{10} value at 5 °C was approximately 0.5 kGy for all the meats tested. Results obtained for *L. monocytogenes* were the same regardless of whether irradiation was delivered by electron beam or by gamma-rays from a ^{60}Co source. It is generally accepted that the irradiation dose currently approved by the Food and Drug Administration for use in meats is sufficient to control *L. monocytogenes*.

Other Nonthermal Processing Techniques

In an effort to provide consumers with minimally processed foods that have retained their nutritional status, a variety of nonthermal processing techniques are in various stages of development. These include high-pressure, pulsed electric fields, oscillating magnetic waves, high-intensity light, and others. The effects of a number of these on *L. monocytogenes* have been investigated. The ability of the organism to survive high-pressure processing of a variety of foods, including milk, cheese, beef, pork, and icecream mix, has been evaluated. In one study, early-stationary-phase cells in 1% peptone solution were pressurized at 345 MPa for 5 min at 25 °C or for 5, 10, or 15 min at 50 °C. Viability losses (in decimal reductions) following pressurization at 25 °C ranged from 0.9 to 3.5 among nine *L. monocytogenes* strains. When two strains showing the greatest difference in pressure resistance were treated at 345 MPa and 50 °C, both strains were reduced by greater than eight decimal reductions within 5 min.

In raw milk, pressures greater than 350 MPa did not permit enumeration of survivors of *L. monocytogenes* Scott A, due to the rapid fall in numbers. Although *L. monocytogenes* had a higher instantaneous pressure kill value, it was more resistant during holding under pressure than natural microflora.

Effects of high hydrostatic pressure on two *Listeria monocytogenes* strains (NCTC 11994 and a poultry strain) were examined in different foods at approximately 18 °C. Samples of ultra heat-treated (UHT) milk, raw and cooked chicken mince, and raw and cooked beef mince were inoculated with *L. monocytogenes* and exposed to a pressure of 375 MPa. Generally, for both strains, increasing the time of exposure led to a decrease in the number of survivors; however, resistance did vary depending on the nature of the food substrate. Survival was significantly greater ($P < 0.01$) in UHT milk compared to that in meat, and survival of both strains was higher in cooked meats compared to raw meats. Complete inactivation of strain NCTC 11994 occurred in milk treated with a pressure of 375 MPa for 15 min at 45 °C.

In order to investigate the effect of pulsed electric field (PEF), pasteurized whole, 2% and skim milk were inoculated with *L. monocytogenes* Scott A and the effects of milk composition (fat content) and PEF parameters (electric field strength, treatment time, and treatment temperature) on the inactivation of the bacterium were studied. No significant differences were observed in the inactivation of *L. monocytogenes* Scott A in the three types of milk by PEF treatment. However, the lethal effect of PEF was a function of field strength and treatment time, with higher field strength or longer treatment time resulting in a greater reduction of viable cells. A 4-log reduction of the bacterium was obtained by increasing the treatment temperature to 50 °C. The use of PEF is a promising technology for inactivation of *L. monocytogenes*, especially when combined with the addition of nisin.

Factors Affecting Growth

Many factors affect the growth of *L. monocytogenes* and, because they interact, they should not be considered in isolation. Many mathematical models have been developed that predict the response of the organism when grown under a variety of conditions. Some of the important variables governing the growth of this bacterium are discussed below and limiting parameters for growth are shown in Table 5.

Temperature

One of the major reasons why *L. monocytogenes* has posed such a threat to the food industry is its ability to grow at refrigeration temperatures. Some strains of the organism are capable of growth at −0.1 to −0.4 °C in UHT milk and chicken broth, but a study of 100 strains of *L. monocytogenes* found a mean minimum growth temperature of −1.1 °C. At refrigeration temperatures (4–5 °C) the generation time for *L. monocytogenes* varies between about 20 and 60 h, depending on the food in which it is present. At abusive temperatures (10 °C) the organism will grow with doubling times of 10–15 h.

The previous growth temperature of the inoculum can have a marked effect on both rate and extent of growth at low temperatures. Prior acclimatization of the bacterium at low temperatures reduces the lag time for growth.

There is evidence that virulence of *L. monocytogenes* is enhanced by growth at low temperatures.

Table 5 Limiting growth conditions for *Listeria monocytogenes*[a]

Factor	*Minimum*	*Optimum*	*Maximum*
Temperature (°C)	−0.4–0.1	37	44
pH	4.3–5.7	7	8
Water activity	0.9–0.96	–	–

[a]Values vary depending on strain, medium, humectant, acidulant, and interactive effects between growth conditions.

pH

Listeria spp. grow at a wide range of pH (5–9.5) which encompasses the pH of most foods. The minimum pH at which growth occurs is inversely proportional to the growth temperature and growth at pH values below 5 has been observed at temperatures near the optimum for growth (30 °C). The nature of the acidulant also affects growth. In terms of concentration of undissociated acid required to inhibit growth, lactic and citric acids are the most effective inhibitors of the organism.

Other factors, such as a_w, alter the response of *Listeria* spp. to pH changes. The inhibitory effects of salt and pH appear to be additive.

Water Activity and Osmotic Pressure

Growth of *L. monocytogenes* is inhibited in media with an a_w of less than 0.942 (achieved by the addition of sucrose or sodium chloride) or 0.932 (by the addition of glycol). In one study involving the addition of sucrose to broth, growth of *L. monocytogenes* was observed at an a_w of 0.92. Proliferation of the organism was observed in 12% sodium chloride at pH 7 and 30 °C, but the concentration of sodium chloride supporting growth increased with a decrease in temperature.

L. monocytogenes can grow in high-osmotic-strength environments. A number of compounds, including glycine betaine, and carnitine, act as osmoprotectants, and their addition to defined medium causing osmotic stress resulted in an increase in growth rate of *L. monocytogenes* 10403S and Scott A. These same compounds also act as cryoprotectants, as, when added to media, they increased the growth rate of *L. monocytogenes* at 5 °C.

Substrate

The rate of growth of *L. monocytogenes* is remarkably similar in a variety of foodstuffs (Table 6). This may be expected, considering the pH and a_w values for these foods are similar. When slices of cooked beef, pork, chicken, or turkey were inoculated with *L. monocytogenes* strain Murray B, vacuum-packaged, and stored at 5 °C, decreases in pH (6.9 to 5.9) and a_w (0.993 to 0.960; adjusted with NaCl) of the cooked meats increased the lag time and reduced the growth rate of the bacterium, but the type of meat had no effect on the growth of the organism after allowance was made for pH. Thus, slower

Table 6 Typical generation times for *Listeria monocytogenes* growing in various foods at 7 °C

Food	*Generation time (h)*
Raw milk	
Naturally infected	16
Experimentally infected	19
Pasteurized milk	13
Ultra heat-treated (UHT) milk	12
Broccoli	13
Cabbage	24
Carrots	14
Radish	13
Beef	15
Lamb	9
Pork	13
Chicken	11
Chicken broth	10
Eggs, pasteurized whole	17
Crabmeat	12
Shrimp	14
Surimi	12
White fish	15
Tryptic soy broth	13

growth could be expected in fermented foods, although other factors may also contribute to the rate of *Listeria* replication in foods. For example, when growth of *L. monocytogenes* on raw and cooked beef, chicken meat, catfish, and shrimp at 4 °C was compared, *L. monocytogenes* grew faster and reached a higher population on raw and cooked catfish and shrimp than on beef or chicken. Some difference in growth rate was due to inherent pH differences of fresh tissues, but on cooked tissues, growth rate on acidified shrimp was the same as that on shrimp with the normal pH of 7.6. Growth rates of *Listeria* on cooked beef at either pH were similar to those noted on uncooked tissue. This indicates that additional factors in shrimp exert inhibitory effects at reduced pH levels, and that these may be heat-related.

The fat content of milk may dictate growth rates and population levels for *L. monocytogenes* strains of serotype 4b. However, other work failed to confirm this enhanced growth with increasing fat levels. Better growth of *L. monocytogenes* on fatty meat rather than lean tissue has also been found, and this was reportedly due to a shorter lag phase of growth on the fat.

L. monocytogenes was able to grow to greater cell numbers in chocolate milk than in skim or whole milk. The growth of the organism in milk was also enhanced by the addition of sugar, cocoa, and carrageenan stabilizer. The major substrate supporting the growth of *L. monocytogenes* in milk appears to be glucose.

The characteristics of *L. monocytogenes* isolated from different foods appear to be different.

Modified Atmospheres

In a study on cooked chicken loaf stored under a modified atmosphere containing 50% carbon dioxide and 10% oxygen, growth of *L. monocytogenes* was retarded at 3 and 7 °C but not at 11 °C. Modification to obtain an atmosphere of 80% carbon dioxide and 0% oxygen resulted in decreased growth at all three storage temperatures. Similar results have been reported for beef, pork, lamb, and turkey packed under carbon dioxide when growth was observed at 10 °C, but was inhibited at temperatures between 0 and 5 °C. In general, for *L. monocytogenes* present in modified-atmosphere-packaged raw meats, growth is decreased with increasing carbon dioxide concentrations and decreasing temperature. Growth of *L. monocytogenes* occurs in beef and cooked chicken packaged under vacuum and in oxygen-permeable or -impermeable films.

To investigate the effect of modified-atmosphere packaging on processed meats, samples of sliced frankfurters were inoculated with *L. monocytogenes* and packaged under atmospheres containing between 20 and 80% carbon dioxide (the remainder being nitrogen). Only the 80% carbon dioxide atmosphere fully inhibited *L. monocytogenes* at 4, 7, and 10 °C, but this carbon dioxide concentration caused unacceptable souring of the sample.

Modified-atmosphere packaging also has little effect on the ability of *L. monocytogenes* to grow in fish and seafood or in fresh-cut vegetables such as lettuce. However, there is evidence to suggest that bacteriocin-forming lactic acid bacteria or bacteriocin isolates may be useful as biopreservatives for modified-atmosphere-packaged vegetables.

Inhibitors

A number of naturally occurring compounds and food additives are inhibitory to *L. monocytogenes*. These include the enzymes lactoperoxidase (found in milk) and lysozyme (milk and egg white). Lactoperoxidase enhances thermal destruction of the bacterium. Milk lactoferrins also inhibit the organism and it is speculated that an animal tissue bacteriostatic agent active against *L. monocytogenes* exists.

Fatty acids, naturally occurring in milk from several mammalian species, are inhibitory for *L. monocytogenes*, with lauric (C12:0), linoleic (C18:2), and linolenic (C18:3) acids exerting the strongest bactericidal activity. These fatty acids also decreased the ability of *L. monocytogenes* to invade mammalian cells.

Antibiotic compounds, bacteriocins, produced by lactic acid bacteria are effective against *Listeria* spp. Among these are nisin produced by *Lactococcus*

lactis, pediocins of *Pediococcus acidilacti* and *P. pentosaceus*, as well as several other compounds secreted by other lactic acid bacteria.

Some preservatives commonly used in the food industry inhibit and/or inactivate *L. monocytogenes*. These include sorbic acid, sodium benzoate, and sodium propionate. The levels of nitrite (120 p.p.m.) and salt (3%) used in the meat industry reduced numbers of the organism during production of sausage lasting 21 days, but surviving cells were detectable. Populations of *L. monocytogenes* can be reduced to below detection limits by a variety of liquid smoke products.

Among other compounds that inhibit this bacterium are lauric (C12:0), linoleic (C18:2), and linolenic (C18:3) acids, monolaurin, organic acids, and high-molecular-weight sodium polyphosphates. Extracts of numerous herbs and spices also have the ability to inhibit growth of *L. monocytogenes*, including rosemary extract and cinnamon extract.

It is noteworthy that the efficacy of these antilisterial compounds in food is dependent on many factors, such as the food matrix, pH, and a_w.

Interaction with Other Microorganisms

Growth of *L. monocytogenes* is inhibited in the presence of lactic acid bacteria, including *Lactococcus lactis, L. cremoris, Leuconostoc gelidum*, and *Lactobacillus bulgaricus*. The inhibition is partly due to production of bacteriocins by the lactic acid bacteria but there is also a substantial pH effect. In most instances growth of the pathogen is completely inhibited when the pH drops below 4.75.

Interactions between *Listeria monocytogenes* and other organisms have been described. Coinoculation of milk or broth with *L. monocytogenes* and *Pseudomonas* spp. resulted in either slight retardation of growth of the pathogen or no effect, depending on strains and temperature of incubation. However, a significant stimulation of growth of *L. monocytogenes* in milks that had previously supported growth of *Pseudomonas* spp. has been reported. This effect on growth is due to the production of proteases by the pseudomonads. Accelerated growth of *L. monocytogenes* in whey cultured with *Penicillium camembertii* compared with uncultured whey has been reported.

Foodborne Outbreaks of Listeriosis

Several major outbreaks of listeriosis have been described where food was shown to be, or was strongly implicated as, the source of infection (Table 7). Several other sporadic outbreaks have occurred where food has been the suspected vehicle for infection. In most foodborne cases, the foodstuff was either unprocessed or inadequately processed. Some of the major foodborne outbreaks of listeriosis are discussed.

Table 7 Notable outbreaks of listeriosis associated with foods

Year	*Location/country*	*Approximate no. of cases (deaths)*	*Possible vehicle for transmission*	Listeria monocytogenes *serotype*
1979	Boston, USA	20 (5)	Raw celery, tomatoes, lettuce	4b
1980	New Zealand	29 (9)	Shellfish, raw fish	1/2b
1981	Maritime Provinces, Canada	41 (17)	Coleslaw	4b
1983	Massachusetts, USA	49 (14)	Pasteurized milk	4b
1985	California, USA	142 (48)	Jalisco cheese	4b
1983–87	Switzerland	122 (31)	Raw milk cheese	4b
1986–87	Philadelphia, USA	36 (16)	Icecream, salami	Several
1989	Connecticut, USA	9 (1)	Shrimp	4b
1987–89	UK	355 (94)	Pâté	4b
1990	West Australia	11 (6)	Pâté	1/2a
1991	Tasmania, Australia	4 (0)	Smoked mussels	1/2
1992	New Zealand	4 (2)	Smoked mussels	1/2
1992	France	279 (83)	Pork tongue in jelly	4b
1993	France	39 (12)	Pork rillettes (pâté)	4b
1993	Italy	23 (0)	Rice salad	1/2b
1993–94	Texas, USA	8 (1)	Frozen mixed vegetables	4b
1994	USA	45 (0)	Chocolate milk	1/2b
1994–95	Sweden	6 (1)	Cold-smoked rainbow trout	4b
1995	France	20 (4)	Raw milk soft cheese	4b
1998–99	USA	100 (21)	Hot dog sausages	4b
1999–2000	France	6 (2)	Rillettes, langoustines	4b
1999–2000	France	32 (10)	Pork tongue in jelly	4b
2000	USA	29 (4)	Processed turkey	?

Maritime Provinces, Canada, 1981

This outbreak was traced to contaminated coleslaw. *L. monocytogenes* of the same serotypes were isolated from patients, coleslaw from the refrigerator of an infected person, and unopened packs of the product from the same manufacturer. The coleslaw had been made from cabbages grown on land fertilized by sheep manure from a flock known to be infected with listeriosis.

An outbreak of the illness at a hospital in Boston in 1979 may also have been due to the consumption of contaminated raw vegetables.

Massachusetts, 1983

A case-control study showed that this outbreak was strongly associated with the consumption of pasteurized milk. The organism was not isolated from the implicated milk, but cases of bovine listeriosis had occurred in herds supplying milk to the creamery involved. This led to the suggestion that *L. monocytogenes* could survive pasteurization due to its intracellular location in somatic cells. Plant inspection revealed no obvious faults that would have resulted in inadequate heat treatment or postpasteurization contamination. However, the fact that the pasteurizer was not located adjacent to the packaging room could have been a contributing factor in the contamination. A few facts are difficult to explain. For instance, the strain isolated from the raw milk was a different phage type from the epidemic isolate. Also, only consumption of 2% fat and whole milk was associated with infection and not skim milk, which was processed on the same day, with the same equipment and milk supply.

California, 1985

An outbreak of listeriosis in California between January and August 1985 resulted in 48 deaths. The most likely source of infection was Mexican-style cheese from one manufacturer. *L. monocytogenes* serotype 4b was isolated from the cheese and processing plant. Phosphatase tests on the cheese showed that pasteurization was inadequate or the product was contaminated with raw milk. Inspection of the plant revealed several defects. There were holes in the pasteurizer plates, a cleaning-in-place line connected raw milk supply to the cheese vat, and the capacity of the pasteurizer was exceeded. This led to the conclusion that raw milk was the cause of the outbreak.

Vaud, Switzerland, 1983–87

Only sporadic cases of listeriosis had been encountered in the Swiss canton of Vaud prior to 1983. The number of cases increased dramatically in the winter of that year. In the following years a similar seasonal peak in cases was recorded. Cheeses were found to harbor *L. monocytogenes*, including a particular type of soft cheese exclusively produced in the winter and mainly eaten in the region of the listeriosis outbreak. The sero- and phage type of the strains isolated from this cheese, Vacherin Mont d'Or, were the same as the epidemic isolates. The cheese was made from raw milk, although *L. monocytogenes* was rarely isolated from the milk supply (< 1% of samples positive). In 1987 the product was recalled and production stopped. After introduction of these measures. no cases of listeriosis were recorded in the area in 1988.

France, 1992

An outbreak of listeriosis in France in 1992 was reported to be responsible for 279 cases, including 22 abortions and 63 deaths. It was attributed to the consumption of pork tongue in aspic, but some other cases may have been caused by other delicatessen products that had become cross-contaminated. The epidemic strain was isolated from 220 food products at 30 retail stores where the pork tongue was sold.

France, 1995

An outbreak of listeriosis in France in 1995 was traced to the consumption of Brie de Meaux cheese made from unpasteurized milk. The outbreak involved 20 cases and four deaths. Further cases were prevented by instigating disinfection and other control measures at the cheese plant.

USA 1998–2000

Between 2 August 1998 and 8 February 1999, the Centers for Disease Control and Prevention (CDC) reported an outbreak of listeriosis involving a rare strain of *L. monocytogenes*, serotype 4b, involving approximately 100 illnesses in 22 states covering almost every region of the contiguous USA. This outbreak was responsible for 21 deaths (15 adults and six miscarriages/stillbirths). CDC and state and local health departments identified the vehicle for transmission as hot dogs and possibly deli meats produced under many brand names by one manufacturer, Bil Mar Foods. On 22 December 1998, the manufacturer issued a voluntarily recall of specific production lots of hot dogs and deli meats that might be contaminated and the number of illnesses caused by the outbreak strain decreased dramatically following the recall. The direct cost of recalling the meats was $76 million, but that did not include the resulting loss of sales. Meat sales dropped about $200 million in the 6 months after the recall. The outbreak strain of *L. monocytogenes* was isolated from an opened and

previously unopened package of hot dogs manufactured at the company's plant in Zeeland, Michigan.

In 1997, the plant was shut down for several days for violating standard sanitation procedures, and in July 1998 evidence of *Listeria* was found in the plant, but not along the food production line. The plant was supposedly aggressively sanitized and cleaned after the July testing. Federal investigators suspect dust kicked up during a summer maintenance project at the meat-processing plant may have been the source of the outbreak. The US Department of Agriculture, which regulates the quality, sanitation, and safety of meat products in the USA, found a dramatic increase in the amount of *Listeria* levels on floors and equipment in the plant in the 6 weeks after the construction work, during which time the isolation rate for '*Listeria*-like indicator organisms' on production equipment rose from 25 to 92%. Maintenance on the air-conditioning system performed in one of the rooms at the plant could have caused widespread contamination of meat products throughout the plant.

In another outbreak involving processed meats, 29 illnesses caused by a strain of *L. monocytogenes* were identified in 10 states between May and November 2000. The isolates from these cases were indistinguishable when subjected to genotyping by pulsed-field gel electrophoresis and ribotyping. The outbreak was linked to the consumption of contaminated turkey and chicken deli meats produced at two establishments; one of these establishments produced turkey meat for the second establishment.

Sporadic Cases

Sporadic cases of listeriosis have arisen from the consumption of meat and poultry products, dairy products, vegetable products, and seafood. *L. monocytogenes* counts of $3–5 \times 10^7$ were found on analyzing samples of goats' milk cheese implicated in one case, and subsequent investigations suggested that counts may have increased to these levels during storage in the display cabinet of the retailer. This led one researcher to propose that the increase in consumption of relatively young soft cheese, coupled with extended refrigeration of foods as family shopping is carried out at weekly, or even monthly, intervals, may be important factors contributing to listeriosis cases reported in the UK.

See also: **Hazard Analysis Critical Control Point**; **Lactic Acid Bacteria**; ***Listeria***: Detection; Listeriosis

Further Reading

Farber JM and Peterkin PI (2000) *Listeria monocytogenes*. In: Lund BM, Baird-Parker AC and Gould GW (eds) *The Microbiological Safety and Quality of Food*, pp. 507–534. Gaithersburg, MD: Aspen.

Griffiths MW (1989) *Listeria monocytogenes*: its importance in the dairy industry. *Journal of the Science of Food and Agriculture* 47: 133–158.

Johnson JL, Doyle MP and Cassens RG (1990) *Listeria monocytogenes* and other *Listeria* spp. in meat and meat products. A review. *Journal of Food Protection* 53: 81–91.

Lovett J (1989) *Listeria monocytogenes*. In: Doyle MP (ed.) *Foodborne Bacterial Pathogens*, pp. 283–310. New York: Marcel Dekker.

Ryser ET and Marth EH (1999) *Listeria, Listeriosis, and Food Safety*, 2nd edn. New York: Marcel Dekker.

Tompkin RB *et al*. (1999) Guidelines to prevent post-processing contamination from *Listeria monocytogenes*. *Dairy, Food and Environmental Sanitarian* 19(8): 551–562.

Various authors (1996) Special issue on *Listeria monocytogenes*. *Food Control* 7: 181–258.

Detection

M W Griffiths, Canadian Research Institute for Food Safety, University of Guelph, Guelph, Ontario, Canada

Background

One of the primary goals of a food microbiologist is to detect the presence of specific pathogens in foods in the shortest possible time. This is not an easy task as often, the organism of concern is present in low numbers, and more often than not, it is found in foods in the presence of large populations of other bacteria. Detection of pathogens is primarily achieved using a cultural approach involving essentially three steps: (1) a nonselective and/or selective enrichment procedure to increase the numbers of the target organism to detectable levels and provide an opportunity for injured cells to recover, (2) a selective step when the organism is presumptively identified and (3) confirmation of the identity of the isolates obtained by the first two phases.

Detection

The type of food being analyzed is an important factor to consider when evaluating methods for recovery and enumeration of specific bacteria, and *L. monocytogenes* is no exception. The physical nature of the food may affect distribution of the inoculum on the medium or mask color reactions important for colony differentiation. More importantly, foods differ with respect to

their indigenous microflora. Thus, a medium selective for *Listeria* may successfully inhibit contaminants from one type of foodstuff, but not from another. For this reason, there is no one universally acceptable method for the isolation of *L. monocytogenes* from foods, and each regulatory jurisdiction has its own preferred method.

Enrichment

One of the most useful methods for enrichment of *L. monocytogenes* involves incubation of samples at 4 °C for up to 12 weeks in a nonselective medium. Unfortunately, the length of this procedure precludes its general use in the food industry.

Experiments performed to optimize basal broth medium for the growth of *L. monocytogenes* gave rise to the formulation of an enrichment broth containing proteose peptone (0.5%), phytone peptone (0.5%), sodium chloride (0.5%), dextrose (0.1%), yeast extract (0.1%), sodium pyruvate (0.1%), and disodium phosphate (0.25%). Under optimum conditions, the generation time of *L. monocytogenes* was at least 6 min shorter in this broth than in tryptose broth.

A number of selective agents have been incorporated into enrichment broth. Most enrichment broths for *L. monocytogenes* have contained nalidixic acid to prevent growth of Gram-negative bacteria and acriflavin to provide additional selectivity with regard to Gram-positive organisms. A commonly used enrichment broth, developed by the US Food and Drug Administration, contains acriflavin, nalidixic acid, and cycloheximide. Formulations based on these inhibitors may not be entirely satisfactory due to the growth of interfering species. Of special concern are highly resistant strains of *Enterococcus faecalis* and other enterococci. Reports of resistant strains of *Staphylococcus aureus* and *Kurthia* spp. have also appeared in the literature. An improved enrichment formulation has been described by researchers at the US Department of Agriculture (USDA). They claim that buffered tryptose phosphate broth containing acriflavin and claforan may be an effective enrichment system, but still expressed some reservations about its ability to control the growth of all enterococci.

It has been claimed that two-stage enrichment procedures are more effective for amplifying the growth of *L. monocytogenes* from food samples. In one such procedure, UVM *Listeria* enrichment broth (containing acriflavin and nalidixic acid) was used initially. After 24 h, samples were transferred to a modified broth containing lithium chloride to inhibit enterococcal growth. The second enrichment broth also contained esculin and ferric ammonium citrate so that presumptive growth of *L. monocytogenes* could be detected by the appearance of a black precipitate due to the formation of iron phenolic compounds resulting from esculin hydrolysis. No false negatives were found when this technique was used on a variety of dairy and environmental samples. Unfortunately, a high rate of false positives was detected, which was again primarily due to growth of enterococci. The strongly buffered UVM broth has advantages when isolating *L. monocytogenes* from soft cheese as it combats the rapid drop in pH that results when sampling this product.

Where there is likely to be heavy contamination with interfering organisms, the more selective enrichment broth, L-PALCAMY, may give better results. However, the use of highly selective media may be problematic as the high concentration of inhibitors in the media can slow the growth of sublethally injured or sensitive strains of *Listeria*. To overcome problems associated with the recovery of sublethally injured cells, a nonselective medium, such as buffered peptone water, has been used as the primary enrichment broth followed by secondary enrichment in a more selective medium. A nonselective universal enrichment broth has been described for the combined enrichment of *Listeria* and *Salmonella*. It has also been shown that strict anaerobic incubation aids in the recovery of injured *L. monocytogenes*.

A two-stage enrichment of particular value for meats involves primary enrichment for 24 h at 35 °C in proteose peptone–yeast extract broth followed by a secondary enrichment under the same conditions in the same broth, but this time modified to contain the selective agents lithium chloride, glycine anhydride, acriflavin, Tween 80, polymyxin B, ceftazidime, phosphomycin, and nalidixic acid.

The resuscitation of *L. monocytogenes* depends on the media, incubation conditions, and sample matrix. For example, the recovery of *L. monocytogenes* from cheese was affected by the method of homogenization of the sample, the type of diluent used, and the temperature of the diluent. The best results were obtained when cheese samples were homogenized in a Waring blender for 2 min and diluted in trisodium citrate (2% w/v) pretempered to 40 °C. However, due to safety concerns, the authors recommended that blending be replaced by stomaching for 3 min. When enrichment was performed at 30 °C, there was generally no advantage in prolonging the incubation time from 24 to 48 h. Optimum growth of *L. monocytogenes* during enrichment is obtained if the initial pH of the broth was between 6.8 and 7. Improved isolation of *L. monocytogenes* was reported when the enrichment was carried out under microaerophilic conditions

(5% oxygen; 10% carbon dioxide; 85% nitrogen). There was a decreased lag period noted for growth of *L. monocytogenes* under microaerophilic conditions with increased carbon dioxide concentrations. Dilution of the sample after enrichment with potassium hydroxide is often recommended in protocols but is of limited value, having no effect on the number of recoveries or the background microflora. Enrichment broths used for the detection of *Listeria* are shown in Table 1.

Selection

After amplification of the number of *L. monocytogenes* cells by enrichment, a presumptive isolation is carried out. This usually entails plating on to selective or differential media. Several such plating media for detection of *L. monocytogenes* have been described (Table 2).

One formulation that has been used widely is McBride's *Listeria* agar. Phenylethanol, lithium chloride, and glycine serve as the selective agents, and blood is incorporated into the medium to identify hemolytic strains. Overgrowth of plates with resistant staphylococci and streptoccoci can occur with this medium. Modifications to this medium include the incorporation of cycloheximide to prevent the growth of yeasts and molds, and substitution of glycine with glycine anhydride (a more effective selective agent). Cephalosporin antibiotics such as moxalactam have also been added to McBride's agar to produce a highly selective medium, although periodic problems due to growth of resistant isolates of enterococci, *Staphylococcus* spp. and *Kurthia* spp. can arise. Another cephalosporin, ceftazidime, may be more effective than moxalactam. A medium containing ceftazidime and acriflavine as selective agents inhibited all organisms, other than *L. monocytogenes*

Table 1 Composition of enrichment broths used for the detection of *Listeria*

Stage of enrichment/enrichment broth	*Acriflavine (mg l^{-1})*	*Nalidixic acid (mg l^{-1})*	*NaCl (g l^{-1})*	*Buffer*	*Other additions*
Primary enrichment					
Buffered peptone water	0	0	5	Phosphate	None
IDF standard	10	40	5	Phosphate	Cycloheximide
Levinthal	15	40	5	None	None
Lovett	15	40	5	Phosphate	Cycloheximide
Modified Lovett K	10	40	5	Phosphate	Cysteine; N_2
Modified Lovett MOPS	15	40	5	Phosphate/MOPS	None
L-PALCAMY	5	0	5	None	Ceftazidime; polymyxin; LiCl; egg yolk
UVMI	12	20	20	Phosphate	None
Universal	0	0	5	Phosphate	$MgSO_4$; pyruvate
Secondary enrichment					
Fraser	25	20	20	Phosphate	LiCl
Lovett	15	40	5	Phosphate	Cycloheximide
UVM II	25	20	20	Phosphate	None
Combined					
Listeria repair broth	37 239	5	0	Phosphate/MOPS	Cycloheximide; glucose; yeast extract; $MgSO_4$; $FeSO_4$; pyruvate

Table 2 Composition of *Listeria* selective agars

Agar	*Acriflavine (mg l^{-1})*	*Polymyxin (P) or Colistin (C) (mg l^{-1})*	*LiCl (g l^{-1})*	*Other additions (mg l^{-1})*
EHA	5	10 (P)	10	Ceftazidime (30)
LCA	0	0	5	Ceftazidime (50)
LPM	0	0	5	Moxolactam (20); glycine anhydride (10); phenylethanol (2500)
LSA	5	10 (P)	15	Ceftazidime (20); tellurite (200)
MC	0	10 (C)	12	Ceftazidime (20)
MVJ	0	0	5	Bacitracin (20); nalidixic acid (50); moxolactam (5); glycine (10); tellurite (105)
Oxford	5	20 (C)	15	Fosfomycin (10); cefotetan (2); cycloheximide (400)
PALCAM	5	10 (P)	10	Ceftazidime (30)

with the exception of a few enterococci, members of the Enterobacteriaceae, yeasts, and molds.

Differentiation of *Listeria* colonies on many media is achieved by their characteristic blue to blue–gray color when the plates are illuminated with obliquely reflected light (Henry illumination). This technique has been criticized because it is often difficult to perform effectively, is subjective, and is not easily quantifiable. Therefore, research has been concentrated on the development of media that rely on other means of differentiation.

A series of differential selective media have been developed by researchers in The Netherlands. The first, RAPAMY agar, used a Columbia blood agar base with the addition of ferric ammonium citrate and egg yolk emulsion. Selectivity was provided by nalidixic acid, acriflavine, phenylethanol, and incubation under microaerobic conditions. The presence of esculin and ferric ammonium citrate together with D-mannitol and Phenol Red results in presumptive positive colonies that are dark gray with a black center surrounded by a dense black halo (due to esculin hydrolysis) against a cherry red background (caused by the inability of *Listeria* to ferment mannitol). However, RAPAMY medium failed to isolate *Listeria* from food in which *Enterococcus* spp. exceeded *Listeria* by a factor of about 10^3. A modification of the medium to replace nalidixic acid with lithium chloride to prevent growth of enterococci gave encouraging results with a variety of foods, including raw chicken, raw milk, raw vegetables and raw milk soft cheeses. *Enterococcus* spp. and *Staphylococcus* spp. that were able to grow on this medium (ALPAMY) could be differentiated from *Listeria* spp. on the basis of their colony morphology. However, sublethally injured cells of *L. monocytogenes* could not be recovered reliably on ALPAMY. Further modification of the medium resulted after addition of polymyxin B and ceftazidime and supplementation of the Columbia agar base with glucose. The modified medium (PALCAM agar) recovered strains of *L. monocytogenes* and other *Listeria* spp. quantitatively and suppressed growth of most other bacteria commonly found in food. The occasional *Enterococcus* or *Staphylococcus* isolates that developed on the medium gave gray colonies with a brown–green halo or yellow colonies with a yellow halo.

Another differential agar that has found widespread use is Oxford agar. Again, Columbia agar forms the basis of this medium, but the selective agents used consist of lithium chloride, acriflavine, colistin, cycloheximide, cefotetan, and fosfomycin. Differentiation of *Listeria* spp. is obtained by the incorporation of esculin and a ferric salt, resulting in black zones around suspect colonies. Good inhibition of interfering organisms, including enterococci, has been reported using this medium.

Good results have been obtained for the isolation of *Listeria* spp. from meats, poultry, and seafood using a modified Vogel Johnson agar. The medium includes glycine, elevated levels of lithium chloride, nalidixic acid, bacitracin, and moxalactam as selective agents. The medium also includes mannitol as a differential agent and potassium tellurite as both a differential and selective agent. On this medium, *Listeria* spp. are entirely black on a red background, whereas resistant strains of enterococci and staphylococci are surrounded by a yellow zone due to mannitol fermentation. The medium surrounding resistant *Kurthia* strains turns maroon due to their strong alkaline reaction.

Media for the presumptive identification of *L. monocytogenes* based on the formation of the enzyme phosphatidylinositol-specific phospholipase C have been described. CHROMagar *Listeria* helps to easily differentiate *Listeria monocytogenes* from other *Listeria* directly at the isolation step. *Listeria monocytogenes* colonies are blue and are surrounded by a white halo due to a specific phospholipase activity. Formation of an opaque halo surrounding *L. monocytogenes* colonies on Otaviani & Agosti Listeria Agar (ALOA) permits identification of *L. monocytogenes* even in mixed bacterial populations. When ALOA was compared with Oxford medium for the detection of *L. monocytogenes* in a variety of foods, including meat, meat products, sea foods, egg products, dairy products and foods of plant origin, the proportion positive for *Listeria* spp. was 34.1% for ALOA compared with 30.4% for Oxford medium. The proportion positive for *L. monocytogenes* was 18.3% on ALOA and 3.6% on Oxford medium. The chromogenic (5-bromo-4-chloro-3-indoxyl-myo-inositol-1-phosphate) or fluorogenic (4-methylumbelliferyl-myo-inositol-1-phosphate) substrates have been added to media in order to identify the presence of the phospholipase.

The temperature at which the selective medium is incubated may be important as the minimal inhibitory concentration of *Listeria* spp. to four antibiotics commonly used in selective plating media decreased with increasing temperature. Strains tended to be more susceptible to inhibition by the antibiotics at 37 °C rather than at 30 °C.

It is unlikely that one medium will prove suitable for the isolation of *L. monocytogenes* from all types of food. In studies on direct plating of *L. monocytogenes* from foods, noncultured dairy products such as pasteurized milk and chocolate icecream mix presented the fewest problems. However, Brie cheese

and raw cabbage contained high levels of contaminants that grew on isolation media. The main contaminants associated with Brie were molds, yeasts, and Gram-positive cocci. High levels of Gram-positive, rod-shaped contaminants were isolated with *Listeria* spp. from cabbage, although Gram-negative rods could often be detected on isolation media. Meats also contained high populations of contaminating microorganisms, principally Gram-positive cocci. Of the food examined, raw oysters posed the biggest problem because of the high numbers of indigenous Gram-positive, rod-shaped bacteria present. In addition, pseudomonads were often recovered from test media plated with raw oysters.

Confirmation

Confirmation of the identity of presumptive isolates on selective agar is usually achieved by a study of morphological and biochemical characteristics.

L. monocytogenes is a Gram-positive, small, short rod-shaped organism. These bacteria are flagellated and exhibit a characteristic tumbling motility at 20 °C. At higher temperatures (30–37 °C), production of flagella is suppressed with a concomitant loss in motility. On solid medium, *L. monocytogenes* colonies are translucent but take on a blue–green coloration when viewed by transmitted light. On blood agar, colonies produce a narrow zone of β-hemolysis.

Listeriae are aerobic to microaerophilic, and growth is enhanced in atmospheres with reduced oxygen and 5–10% carbon dioxide. They possess a heme-containing catalase and are oxidase-negative. *L. monocytogenes* is positive for the Voges–Proskauer and tellurite tests. The bacterium ferments glucose with the production of mainly L(+)-lactic acid, but no gas. All members of the genus hydrolyze esculin and maltose. Rhamnose and xylose reactions are important in differentiating *Listeria* spp. The CAMP test is also used to differentiate *L. monocytogenes* from other members of the genus. Both *L. seeligeri* and *L. monocytogenes* may not produce enough hemolysin to be noticeably hemolytic. However, when grown in an area of *Staphylococcus aureus* growth, the hemolytic reaction of these *Listeria* spp. is enhanced. The hemolysis is unreactive to metabolites of *Rhodococcus equi*. Reactions used for differentiation of *Listeria* spp. are summarized in Table 3.

Confirmation by biochemical tests can be made simpler by the use of ready-to-use test strips and automated identification systems such as the Vitek Junior. The Vitek Gram-positive identification card correctly identified 229 of 236 *Listeria*-positive isolates, and 95% of these were confirmed within 6–8 h. Other miniaturized test kits for the identification of *L. monocytogenes* based on phenotypical biochemical reactions include an arylimidase test on paper disks, API *Listeria* miniaturized test kit, Biolog Micro-ID *Listeria* and Biolog Phenotype microarrays for *Listeria*, and the Rosco system. When the Micro-ID *Listeria* System, API *Listeria*, and the Rosco system were compared with conventional identification tests, all these rapid biochemical test kits gave fast, reliable identification of *Listeria* species when compared with classical techniques. However, the API *Listeria* system identified the test strains without a complementary CAMP test.

Detection in Processed Foods

The main difference between the detection of *L. monocytogenes* in raw and processed foods is that the latter may have been subjected to conditions that sublethally stress the organisms of concern as well as reducing their numbers. It is known that *L. monocytogenes* can be sublethally stressed by heat, acidulants, and, to a lesser extent, freezing. Injured organisms need a period of recovery to repair cellular damage. The resistance of sublethally injured cells to selective agents used in their detection may be substantially reduced. In general, the greater the selectivity of the medium, the less likely it will be that stressed cells will grow.

Injured *Listeria* cells are unable to grow in the presence of 2% sodium chloride, and the proportion of stressed cells may be determined by comparing

Table 3 Biochemical tests for the differentiation of *Listeria* spp

Test	*Listeria spp.*				
	monocytogenes	*innocua*	*seeligeri*	*ivanovii*	*welshimeri*
Hemolysis	+	−	+	+	−
CAMP – *S. aureus*	+	−	+	−	−
CAMP – *R. equi*	−	−	−	+	−
Xylose	−	−	+	+	+
Rhamnose	+	V[a]	−	−	V

[a]Variable reaction.

counts on media with and without sodium chloride addition. Using this technique, the effect of various agents on the recovery of sublethally injured *L. monocytogenes* has been determined. Selective agar used for isolation of *Listeria* spp. generally did not allow resuscitation of heat- or acid-injured cells. However, freezing did not appreciably affect the ability of test strains to grow on selective agar. The inability of a medium to support colony formation by thermally injured *L. monocytogenes* could be correlated with the inclusion of phenylethanol, tellurite, polymyxin B, sodium thiosulfate and oxgall, as well as sodium chloride. The addition of Tween 80, bovine fetal serum or egg yolk emulsion to modified Vogel–Johnson medium increased the ability to detect heat injured cells of the organism 100-fold without any loss of selectivity of the medium.

The universal response of cells undergoing injury appears to be the accumulation of hydrogen peroxide, and injured cells have an increased sensitivity to the toxic effects of hydrogen peroxide. Removal of hydrogen peroxide by the addition of catalase, or the nonenzymatic peroxide decomposer sodium pyruvate, has a beneficial effect on the recovery of injured cells of several microorganisms, including *Staphylococcus aureus* and *Salmonella typhimurium*. Effective repair of stressed *L. monocytogenes* cells on solid medium containing catalase has been described. Also, pyruvate addition to modified McBride agar enhanced recovery of heat-injured *L. monocytogenes*, but it was necessary to incubate plates at 22–25 °C for 7 days in order to achieve beneficial effects. However, the situation is confused as another study failed to show enhanced recovery of *L. monocytogenes* Scott A strain after addition of pyruvate to modified McBride agar. The addition of reducing agents to the medium has been shown to have a mixed effect on the recovery of heat-injured cells of *L. monocytogenes*, but recovery is increased if cells are incubated anaerobically. Enrichment in broth containing oxygen scavengers, such as oxyrase, enabled recovery of heat-injured listeriae within 6 h.

When the effects of incubation temperature, oxygen availability and supplementing the culture medium with magnesium chloride, magnesium sulphate, D-glucose, L-cysteine, catalase, or lithium chloride on the recovery of salt- or acid-damaged *Listeria monocytogenes* serovar 1/2c was studied on a solid repair medium, results showed that conditions promoting resuscitation of acid- or salt-injured cells are stress-specific, and differ in part from those described for heat-stressed *Listeria*.

It is strongly recommended that a nonselective enrichment is included when processed foods are analyzed for *L. monocytogenes*. An incubation time of 6–9 h at temperatures in the range of 20–40 °C is required for complete recovery of injured cells. The temperature optimum appears to be 25 °C. The organism does not appear to be able to repair at low temperatures. Prevention of large pH changes during the enrichment by buffering of the medium also promotes recovery of heat-stressed cells.

Detection in Suspected Food Poisoning

Confirmation of the involvement of *L. monocytogenes* in a case of illness suspected to be of food origin relies on isolation of identical strains of the organism from the patient and the suspected food. In practice, this is not feasible in the majority of cases, and epidemiological case-control studies have proved crucial in implicating a specific food product in an outbreak. The epidemiological surveillance of listeriosis has been hampered by the lack of adequate typing schemes. A serotyping system is available, but is of limited value as a high proportion of strains causing disease belong to only a small number of serovars. Of all *L. monocytogenes* strains isolated in Britain from human cases of listeriosis, 91% comprised only three serovars: 4b (59%), 1/2a (18%), and 1/2b (14%). A phage-typing scheme is also used, but a high proportion of strains (about 36% in Britain) are not typable by this system. An international multicenter typing study of *L. monocytogenes* was initiated by the WHO (Food Safety Unit, Geneva) in order to evaluate the usefulness of various phenotypic and genotypic typing methods for *L. monocytogenes* and to select and standardize the most appropriate methods. Several genotyping methods were compared, including pulsed field gel electrophoresis (PFGE) coupled with the use of the restriction enzymes, *Apa*I and *Sma*I, random amplification of polymorphic DNA (RAPD), high frequency restriction endonuclease analysis (REA), and restriction-fragment length polymorphism (RFLP). In general, this study reconfirmed that PFGE is a very accurate and reproducible method for fine structure comparison and molecular typing of *L. monocytogenes*. An automated genotyping system, the RiboPrinter, has proved useful for the analysis of *L. monocytogenes* and was used successfully to determine the source of the outbreak linked to hot dogs produced at the Bil Mar plant of Sara Lee. *L. monocytogenes* has been classified into three genetic lineages using ribotyping and PCR-RFLP typing of *hlyA* and *octA* genes. A higher proportion of human isolates (69.1%) than industrial isolates (36.8%) were classified as lineage I, which contains human sporadic isolates and all epidemic isolates. All other industrial isolates (63.2%) were classified as lineage II, which contains only human sporadic isolates.

Other Methods of Detection

As well as the classical cultural approach to the detection of *Listeria* spp., more sophisticated techniques are available commercially. These have advantages in their ease of use and in rapidity.

Impedance Monitoring

When microorganisms grow in a medium, large-molecular-weight substrates are catabolized to more highly charged, low-molecular-weight end products. These modify the electrical properties of the medium, and the changes can be detected as changes in conductance or capacitance. A selective medium has been described that contains acriflavin, ceftazidime, and nalidixic acid for use with impedance monitors such as Bactometer, Malthus, and RABIT. The medium also contains esculin and a ferric salt. The presence of *Listeria* spp. is confirmed by detection of growth and black coloration of the medium. In addition, growth of *Listeria* spp. gives a characteristic response curve when the capacitance signal is monitored. The medium gave good results with fermented dairy products, but there appears to be a high rate of false positives when soft cheeses are analyzed. A medium (LED medium) based on the selective agents used in Oxford agar has also been described for the electrical detection of *Listeria* spp. The medium relied upon the ability of *Listeria* spp. to induce a greater than 30% change in capacitance within 30 h. When compared with a traditional Fraser broth enrichment, the method gave far fewer false-positive results when used to analyze environmental swabs and products, such as meat, cream cakes, cheeses, and milks.

An indirect conductimetric assay that detects carbon dioxide evolved during growth of microorganisms is available. Such methods allow the incorporation of high concentrations of salts, such as lithium chloride, into media to aid selective growth of *Listeria* spp.

Immunological Methods

Monoclonal antibodies specific for *Listeria* spp. have been produced in a number of laboratories. There are several commercially available immunoassays for *Listeria* and *L. monocytogenes* (Table 4). A commercially available system applicable to foods involves the use of two monoclonal antibodies with an enzyme-linked immunosorbent assay (ELISA). Following enrichment in broth for 40–48 h at 30 °C, the ELISA procedure is performed. One antibody is used for antigen capture, and one for enzyme conjugation with horseradish peroxidase. For the ELISA to be effective, the enrichment broth must contain at least 10^6 *Listeria* cells per milliliter. The method was able to identify all milk samples containing *L. monocytogenes*. However, the detection rate with raw vegetable samples was only 68%. This was probably due to inadequacies in the enrichment procedure caused by low levels of *Listeria* spp. in the presence of mixed microflora. The method lacks species specificity within the genus *Listeria*. A monoclonal antibody specific for strains of *L. monocytogenes*, *L. innocua*, and *L. welshimeri* has been isolated.

A direct immunofluorescence test has been used to detect successfully *L. monocytogenes* in naturally contaminated soft cheese samples when numbers exceeded 10^2 per gram. The test could be carried out in 2 h. An immunomagnetic capture assay has been developed that is capable of detecting less than 2×10^2 *Listeria* cells per milliliter in enriched foods.

An automated instrument, Vidas, has been developed consisting of a solid phase receptacle (SPR) and pipetting device. The SPR is coated with antibodies against the target antigen, and the necessary reagents to perform an enzyme-linked fluorescent assay are contained in a ready-to-use reagent strip. The fluorescence measured is directly proportional to the quantity of antigen present in the sample. Two tests are available: the VIDAS LIS method detects *Listeria* spp., and the VIDAS LMO method is claimed to be specific for *L. monocytogenes*. The VIDAS LIS method was compared with two conventional culture methods for detection of *Listeria* spp. in naturally contaminated or inoculated ice cream, cheese, green beans, fish, roast beef, and turkey mince. The agreement between the VIDAS LIS and culture methods for all samples tested was 86%. The VIDAS LIS method has been approved by AOAC International for detection of *Listeria* in foods. When used to detect *L. monocytogenes* in cheese, the Vidas assay gave positive results in 97.5% of all contaminated cheeses after 48 h of enrichment. The detection limit was 6.3×10^5 to 2.0×10^6 cfu ml^{-1}.

A comparison of commercially available immunological kits for detection of *Listeria* spp. (Listeria Rapid Test Clearview, Pathalert Listeria, VIDAS Listeria LIS, VIP Listeria) and *Listeria monocytogenes* (Pathalert *Listeria monocytogenes*, VIDAS *Listeria monocytogenes* LMO) in food showed that there were no significant differences between the tests, but all methods had difficulties with the detection of *L. monocytogenes*.

Gene Probes and Related Techniques

A nonisotopic, colorimetric nucleic acid hybridization assay based on 16S rRNA sequences for the detection of *Listeria* spp. in dairy and environmental samples has been described. After a two-stage

Table 4 Commercial immunoassays for *Listeria* spp. and *L. monocytogenes*

Company	*Test name*	*Specificity*	*Time to perform*	*Principle*
BioMérieux, Inc.	Vidas Lis	*Listeria* spp.	45 min (excluding enrichment)	Automated enzyme-linked fluorescent assay
	Vidas LMO	*L. monocytogenes*	45 min (excluding enrichment)	
TECRA International Pty Ltd	*Listeria* Via	*Listeria* spp.	48 h (including enrichment)	Visual immunoassay
	Listeria Unique	*Listeria* spp.	32 h (including enrichment)	Antibody-coated dipstick
BioControl	VIP for *Listeria*	*Listeria* spp.		Single-use, visual immunoprecipitate assay
	Assurance *Listeria*	*L. monocytogenes* and related species	30 min (excluding enrichment)	Polyclonal enzyme immunoassay, microplate with breakaway wells
Neogen	Reveal for *Listeria*	Five of six *Listeria* spp. including *L. monocytogenes*	43 h (including enrichment)	Antibody against specific flagellar antigen; ELISA sandwich immunoassay combined with chromatography in a single-use device
Hardy Diagnostics	Listeria Rapid Test kit			
Vicam	Listertest	*Listeria* spp.	24 h	Immunomagnetic beads coupled with selective media and confirmation with membrane lift immunoassay
Diffchamb	Transia Plate *Listeria*	*Listeria* spp.	24–48 h	Based on a sandwich-type, immunoenzymatic reaction using a mixture of antibodies
	Transia Plate *Listeria monocytogenes*	*Listeria monocytogenes*		Based on three-step, sandwich-type, immunoenzymatic reaction; solid support of the reaction is a microtiter plate coated with antibodies specific for the extracellular P60 protein excreted by *Listeria monocytogenes*

enrichment procedure, the hybridization assay is performed, permitting detection of the organism in about 60 h. The procedure was able to identify 98% of *Listeria*-positive milk samples but only 45% of positive raw vegetable samples. The latter results were thought to be due to inadequate growth of the organism in the enrichment broth used.

A DNA probe specific for β-hemolytic *L. monocytogenes* has been applied to colonies obtained on selective media inoculated with food samples. The method was rapid and accurate for detecting the pathogen in raw milk, semisoft cheese, and ricotta cheese. The method was quantitative, provided that the contamination level in the food was greater than 10 cells per gram. Enrichment was necessary if numbers of *L. monocytogenes* were below this level. A probe, specific for *L. monocytogenes* has been obtained from two DNA fragments encoding part of the listeriolysin O molecule.

Two DNA hybridization assays for *L. monocytogenes* have been commercialized (Table 5). One, the Gene-Trak *Listeria monocytogenes* Assay, has been compared with the IDF Standard 143A:1995 using 107 cheese, fish, and meat samples naturally contaminated with *Listeria*. The Gene-Trak assay gave a significantly higher number of positive results. Examination of a further 75 cheese and meat samples with both methods and, additionally, with three immunological tests (Listeria Rapid Test Clearview, VIDAS Listeria LIS, and a noncommercial ELISA) again confirmed that the Gene-Trak assay gave a greater number of positive samples than all the other methods. The differences, however, were not significant. Another commercially available gene probe

Table 5 Commercial nucleic-acid based kits for the detection of *Listeria* and *L. monocytogenes*

Company	*Test name*	*Specificity*	*Time to perform*	*Principle*
Gen-Probe, Inc.	Accuprobe *L. monocytogenes*	*L. monocytogenes*	45 min (from culture)	Probe method
Neogen	Gene-Trak *Listeria* and *Listeria* DLP assays	*Listeria* spp.	27–48 h (including enrichment)	Probe to 16S rRNA sequence; capture probe on dipstick; colorimetric reaction; sensitivity of 10^6 cells per milliliter
DuPont Qualicon	Genus *Listeria* Automated Bax system	*Listeria* spp	< 4 h (excluding enrichment)	Amplification of rDNA spacer polymorphisms; tableted PCR reagents; the automated BAX system allows detection of multiple targets in the same run; up to 96 samples per batch
	Listeria monocytogenes Automated Bax system	*L. monocytogenes*	< 4 h (excluding enrichment)	
Bio-Rad Laboratories	Probelia *Listeria monocytogenes*	*L. monocytogenes*	30 h (including enrichment)	PCR amplification; unique lysis reagent for DNA extraction; single-pipette dose and single test tube; ready-to-use amplification buffer that contains an enzymatic decontamination system; semiautomatic; colorimetric detection of amplified products on microplates

assay for *L. monocytogenes* (AccuProbe) correctly identified samples containing the organism. However, it has been determined that some enrichment media (University of Vermont broth and Fraser broth) interfered with the AccuProbe test, due to prevention of cell lysis by the high salt concentration in these media. The detection limit of the AccuProbe test was also affected by enrichment and selective broths.

The polymerase chain reaction (PCR), using thermostable DNA polymerase, permits the rapid amplification of specific DNA sequences by a factor of up to 10^7. The portion of DNA amplified is determined by oligonucleotide sequences (primers). Primers have been synthesized that result in PCR products specific for *L. monocytogenes*. PCR protocols based on amplification of the listeriolysin O gene (*hlyA*), the internalin gene (*inlA*), a gene encoding a metalloproteinase (*mpl*), and the gene encoding an invasive-associated protein (*iap*) are among those that have been developed. The products can be detected by agarose gel electrophoresis, but more recently, real-time PCR methods have been developed that allow the detection of amplified products using fluorescence.

To overcome problems due to inhibition of the PCR by components of food, several methods have been investigated such as the use of filters to entrap and lyse bacterial cells, the use of immunomagnetic separation, and magnetic capture hybridization.

Commercial kits to detect *L. monocytogenes* based on PCR have become available (Table 5). The Bax system has been semiautomated and has been shown to be an effective method of screening for *L. monocytogenes* in smoked fish and many different types of fresh produce.

Undoubtedly, with the further development of real-time PCR methods and the advent of DNA array technology, molecular methods will be increasingly used for routine testing of foods for *L. monocytogenes*.

See also: **Food Poisoning**: Classification; Tracing Origins and Testing; **Immunoassays**: Radioimmunoassay and Enzyme Immunoassay; ***Listeria***: Properties and Occurrence; Listeriosis; ***Salmonella***: Salmonellosis; ***Staphylococcus***: Food Poisoning

Further Reading

Batt CA (1997) Molecular diagnostics for dairy-borne pathogens. *Journal of Dairy Science* 80: 220–229.

Baylis C (2000) *The Catalogue of Rapid Microbiological Methods. Campden & Chorleywood Food Research Association Review*, No. 1, Ed. 4. Camden and Chorleywood Food Research Association.

Bhunia AK (1997) Antibodies to *Listeria monocytogenes*. *Critical Reviews in Microbiology* 23(2): 77–107, 179.

Dever FP, Schaffner DW and Slade PJ (1993) Methods for the detection of foodborne *Listeria monocytogenes* in the US. *Journal of Food Safety* 13(4): 263–292.

Farber JM and Peterkin PI (2000) *Listeria monocytogenes*. In: Lund BM, Baird-Parker AC and Gould GW (eds) *The Microbiological Safety and Quality of Food*, pp. 507–534. Gaithersburg, MD: Aspen.

Ryser ET and Marth EH (1999) *Listeria, Listeriosis, & Food Safety*, 2nd edn. New York: Marcel Dekker.

Tsen HY and Chen TR (2001) Development of DNA probes and PCR primers for the specific detection of food pathogens. *Food Science and Agricultural Chemistry* 3: 1–7.

Listeriosis

F J Pagotto and J M Farber, Bureau of Microbial Hazards, Ottawa, Ontario, Canada

Background

The bacterium *Listeria monocytogenes* was identified in 1940. However, it can be dated back further in time, as a bacterial sample from a meningitis case dating to 1921 was the source of the bacterium described by Murray and colleagues as *Bacterium monocytogenes*. *B. monocytogenes* was so named because it caused mononuclear leucocytosis in rabbits. In 1940, Pirie isolated an organism, which he named *Listerella hepatolytica*, in honour of Lister. Listerellosis was thus used to describe the disease caused by this bacterium. Over time, it was discovered that *B. monocytogenes* and *L. hepatolytica* were in fact the same organism. Because *Listerella* was already chosen to describe a group of slime molds, Pirie proposed that the organism be called *Listeria monocytogenes*.

Listeriosis, an infection caused by *L. monocytogenes*, is now considered an important emerging foodborne infection. The importance of this organism can be seen not only from a human disease point of view but also from an economic perspective. Listeriosis is addressed here from a number of aspects: food and clinical diagnosis, antimicrobial chemotherapy, prevention from infection, and mechanisms of pathogenesis.

Listeriosis

Listeriosis is a serious infection caused by the bacterium *L. monocytogenes*. Listeriosis can affect all individuals, but pregnant women, the elderly, and immunocompromised individuals are especially at risk. Although expectant mothers may only present mild flu-like clinical symptoms, they can suffer from more serious complications such as premature delivery and infection of the newborn. In some instances, stillbirth may occur. Because *L. monocytogenes* is able to cross the three protective barriers (gastrointestinal, blood–brain, and placental), the symptoms of listeriosis are diverse, ranging from minor aches to death (**Table 1**).

In nonpregnant adults, *L. monocytogenes* has been shown to cause meningitis, encephalitis, and septicemia. In rare cases, contact with infected animals has caused focal infections in veterinarians. Although skin lesions are the major symptoms usually observed, endocarditis, arthritis, osteomyelitis, intra-abdominal abscess, and pleuropulmonary infections have also been described, albeit on rare occasions.

Listeriosis can be sporadic or epidemic, with food being the primary vehicle of transmission. While the reported incidence of listeriosis ranges from 0.2 to 8.3 cases per million population in industrialized countries, it is widely accepted that the true rate of the human disease is underestimated. Interestingly, the incubation period of listeriosis can range from 24 h to 91 days. The reason for this is not fully understood, but factors such as immune status, bacterial load, genetic predisposition, and other reasons may play a role. Another unresolved issue is the tropism of *L. monocytogenes* for the central nervous system.

Human listeriosis can be caused by all serotypes of *L. monocytogenes*. However, the majority of

Table 1 Symptoms ascribed to infections from *Listeria monocytogenes*

Category	*Symptoms*
Foodborne	Abdominal pain
	Chills
	Diarrhea
	Febrile gastroenteritis
	Fever
	Myalgia
	Nausea
	Vomiting
Pregnancy-related	Bacteremia
	Flu-like symptoms
	Intrauterine or cervical infections
	Neonatal meningitis
Other reported clinical manifestations	Cutaneous listeriosis
	Conjunctivitis
	Convulsions
	Hepatitis
	Meningoencephalitis
	Native/prosthetic valve endocarditis
	Osteomyelitis
	Rhomboencephalitis
	Spontaneous bacterial peritonitis
	Upper respiratory tract symptoms

listeriosis cases have been shown to be caused by serotypes 1/2a, 1/2b, and 4b. Further analyses has revealed that serotype 4b is prevalent in many European countries. In North America, serotypes 1/2a (Canada) and 4b (USA) are most often recovered in cases of listeriosis. Again, the reason for the differences observed in serotype and geographical distribution is unclear. Equally, it is not known at present why only three serotypes cause most of the illness.

Food and Clinical Diagnosis

The food microbiologist has the task of analyzing a food product for the presence of *L. monocytogenes*. They must be able to achieve this goal as rapidly as possible, as the food implicated may have already been consumed prior to the identification protocol being completed. Because listeriae are found throughout nature and in a wide variety of foods, there is no universal protocol at the moment. The majority of identification methods for the detection and enumeration of *L. monocytogenes* are cultural in nature. Other methods include the use of DNA probes, amplification-based methods, antibody-based technologies, and other rapid methods.

In Canada, the Bureau of Microbial Hazards (part of the Health Products and Food Branch) uses a detection method developed for all food and environmental samples (Figure 1). Presumptive colonies are then subjected to biochemical testing, using the criteria listed in Table 2 to differentiate

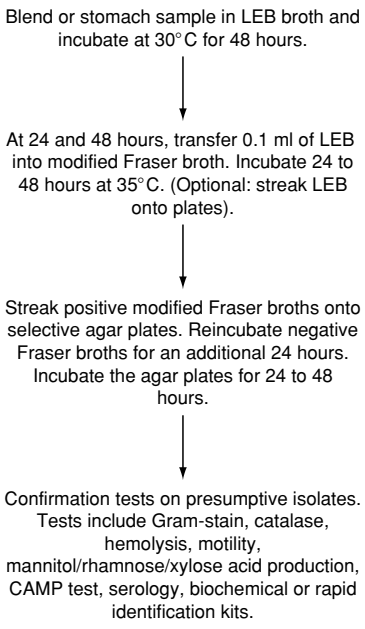

Figure 1 Flow chart for the isolation of *L. monocytogenes* from environmental and food samples.

L. monocytogenes from other species. Enumeration would involve macerating the sample with peptone water and plating on to any of two LPM, Oxford, MOX, or Palcam agar plates, followed by incubation at 30 °C (LPM) or 35 °C (Oxford, MOX, Palcam).

The cultural method utilizes selective agents with the hope of allowing cells of *L. monocytogenes* to grow and competing flora to be inhibited. LEB, also known as University of Vermont broth, contains nalidixic acid and acriflavin as selective agents. LPM medium contains the Gram-positive and Gram-negative inhibitory compounds, lithium chloride, phenylethanol, and moxalactam. Fraser broth contains lithium chloride, nalidixic acid, and acriflavin. The addition of esculin and ferric ammonium citrate causes a darkening of the broth or agar in the presence of *Listeria* species (*Listeria* species have the ability to hydrolyze esculin, and hydrolysis of esculin turns the medium black, denoting a presumptive positive test). Oxford agar contains lithium chloride, cycloheximide, colistin, acriflavin, fosfomycin, and cefotetan as selective agents, in addition to esculin and ferric ammonium citrate. Palcam agar and broth contain, in addition to esculin and ferric ammonium citrate, lithium chloride, polymixin-B-sulfate, ceftazidime, and acriflavin. Interestingly, it would appear that Palcam is preferred in Europe, while LPM, Oxford, and modified Oxford are used predominantly in North America. Many laboratories across the world are currently working on novel formulations that one day may be superior to currently used media.

Other methods have been reported in the literature for the detection and enumeration of *L. monocytogenes* in foods. The first report of a monoclonal antibody specific for the genus *Listeria* was reported in 1987. Since then, antibody-based detection methods have been described. These are based on the enzyme-linked immunosorbent assay. Many companies offer commercially available kits that use the antibody–antigen–antibody sandwich technique as the basis for detection. Antibodies have also been used in conjunction with magnetic beads. The beads are covered with *Listeria*-specific antibodies and mixed with the sample to be tested. A magnetic field is applied to separate the magnetic beads on which the *Listeria* cells are bound. After a washing step, other methods (for example, cultural methods) may be applied. Monoclonal antibodies have been used to serotype the genus *Listeria*, supplemented by hemolysis of blood and *in vivo* mouse assays (Table 3).

There are many protocols described that have, as their basis, hybridization of genetic material (DNA). Many probes have been described and used in colony hybridization methods, as have polymerase chain

Table 2 Characteristics differentiating the species of the genus *Listeria*

Characteristics	L. monocytogenes	L. innocua	L. seeligeri	L. welshimeri	L. ivanovii	L. gravi	L. murrayi
Gram stain[a]	+[b]	+	+	+	+	+	+
β-hemolysis[a]	+[c]	−	+	−	+[d]	−	−
Mannitol (acid production)[a]	−	−	−	−	−	+	+
L-*Rhamnose (acid production)*[a]	+	d	−	d	−	−	d
D-*Xylose (acid production)*[a]	−	−	+	+	+	−	−
CAMP test (S. aureus)[a]	+[e]	−	+	−	−	−	−
CAMP test (R. equi)	−	−	−	−	+	−	−
Acid production from:							
L-Arabinose	−	−	−	−	−	−	−
Dextrin	d	−			−	+	+
Galactose	d	−			d	+	+
Glycogen	−	−	−	−	−	−	−
Lactose	d	+			+	+	+
D-Lyxose	−	−	−			+	+
Melezitose	d	d			d	−	−
Melibiose	−	−	−	−	−	−	−
α-Methyl-D-glucoside	+	+			+	+	+
α-Methyl-D-mannoside	+	+	−[f]	+	−		
Sorbitol	d	−			−	−	−
Soluble starch	−	−			−	+	+
Sucrose	−	d			d	−	−
Voges–Proskauer	+	+	+	+	+	+	+
Hydrolysis of:							
Cellulose	−	−	−	−	−	−	−
Hippurate	+	+			+	−	−
Starch	d	d			−	−	−
Lecithinase	d	d			+	−	−
Phosphatase	+	+	+	+	+	+	+
NO_3 to NO_2 reduction	−	−			−	−	+
Pathogenicity for mice	+	−	−	−	+	−	−

[a]Tests most often used in differentiating the listeriae.
[b]Standard symbols: +, positive; −, negative; +, > or equal to 90% positive; −, > or equal to 90% negative; d, 11–89% of strains are positive.
[c]Not all strains of *L. monocytogenes* exhibit β-hemolysis – the type strain ATCC 15313 is a nonhemolytic mutant on horse, sheep, and bovine blood.
[d]A very wide zone or multiple zones of hemolysis are usually exhibited by *L. ivanovii* strains.
[e]Of 30 strains, ATCC 15313, the type strain, did not give a positive reaction.
[f]Of 10 strains tested, one gave a positive reaction.

Table 3 Serology, hemolytic activity, and mouse virulence for *Listeria* species

Species	*Serotype*	*Hemolysis of 7% horse blood agar*[a]	*Mouse virulence*
L. monocytogenes	1/2a, 1/2b, 1/2c, 3a, 3b, 3c, 4a, 4ab, 4b, 4b(x), 4c, 4d, 4e, 7	+[b]	+
L. ivanovii	5	+	+
L. innocua	4ab, 6a, 6b, un[c]	−[d]	−
L. welshimeri	6a, 6b	−	−
L. seeligeri	1/2b, 4c, 4d, 6b, un[c]	+	−

[a]Portion of colony is stabbed into the blood plate.
[b]Positive reaction.
[c]Undefined.
[d]Negative reaction.

reaction protocols. These methods offer the advantage of reducing the time required to obtain a positive results. However, a possible disadvantage of these methods is that isolation of *L. monocytogenes* is not required. A consequence of this would be the lack of pure isolates to do molecular typing and build on currently existing databases. Some of the existing alternative methods are summarized in Table 4.

In the clinical setting, the diagnosis of listeriosis is dependent on the isolation of the organism from normally sterile sites. These include the blood, placenta, amniotic, or cerebrospinal fluid (CSF). A positive diagnosis may also be ascertained through analysis of stool samples. Samples obtained from normally sterile environments are streaked on to blood agar. On sheep blood agar, narrow zones of β-hemolysis

Table 4 Alternative methods for the identification and/or detection of *Listeria monocytogenes*[a]

Name	*Type or format (target)*	*Company/Supplier*
Accuprobe *L. monocytogenes*	DNA based (probe)	GenProbe Inc.
BAX system	DNA based (PCR)	Qualicon
GENE-Trak	DNA based (probe)	Gene-Trak Systems
PROBELIA	DNA based (PCR)	Sanofi Diagnostics Pasteur
VIDAS	Immunoassay based (enzyme linked fluorescent assay)	bioMerieux Vitek
Listeria VIA	Immunoassay based (ELISA)	Tecra Diagnostics
Reveal (Listeria)	Immunoassay based (immunochromatography)	Neogen
ClearView	Immunoassay based (immunochromatography)	Oxoid
Assurance EIA (*Listeria*)	Immunoassay based (ELISA)	BioControl
EIAFOSS *Listeria*	Immunoassay based (ELISA)	Foss Electric
API	Biochemical	bioMerieux
Microbact	Biochemical	Microgen
Malthus	Biochemical (conductance)	Malthus
Microlog	Biochemical (carbon oxidation)	Biolog
MIS	Biochemical (fatty acid)	Microbial-ID

[a]This is a partial list only and not meant as an endorsement by the authors.

can be seen. *L. seeligeri* has a similar appearance to *L. monocytogenes*, except that a weaker zone of hemolysis is observed. A wider zone of hemolysis has been noted with *L. ivanovii*. As all strains of *Listeria* are motile, tumbling motility (hanging drop preparations from primary cultures isolated) can be used to distinguish them from *Erysipelothrix* and *Corynebacterium* species. One must remember that *Listeria* is motile at room temperature (25 °C), and much less so or not at all at 37 °C.

It is difficult to diagnose listeriosis infections in the absence of culture isolation. While there are many monoclonal antibodies available for the genus *Listeria*, clinical use of these has been hampered owing to their cross-reactivity with other bacteria. Serological studies have shown that cross-reactivity to staphylococci, enterococci, and corynebacteria can occur. Newer antibodies, targeting the virulence factors, for example, may be more useful in the clinical setting. Stool sample carriage rates can range from 1 to 5%. As such, isolation of *L. monocytogenes* from the stool may be helpful in identifying *Listeria* gastroenteritis.

Culture of *L. monocytogenes* from blood or CSF does not require selective media, thus explaining why blood agar plates are the medium of choice. However, the isolation of specimens from food or environmental sources, stools, and vaginal secretions requires selective media (as described in Figure 1). (*See* **Microbiology**: Detection of Foodborne Pathogens and their Toxins.)

Antimicrobial Chemotherapy

Because *Listeria* infections are often associated with high mortality rates, it is imperative that effective antibiotics be used for treatment. The current treatment for listeriosis involves a combination of ampicillin and gentamicin. Therapy for 2–3 weeks appears to be sufficient to prevent relapses. Most β-lactam antibiotics may be used. However, cephalosporins are not effective against *L. monocytogenes* and should not be used to treat serious infections such as neonatal meningitis. Other chemotherapeutic agents, considered secondary treatments, include erythromycin, fluoroquinolones, and vancomycin.

There have been reports of plasmid-encoded resistance to antibiotics such as chloramphenicol, erythromycin, streptomycin, and tetracycline. Tetracycline resistance has been reported in 2–5% of strains of *L. monocytogenes*. These resistant strains have been shown to contain the *tetM* chromosomal determinant (within a transposon). The *tetM* gene has also been reported in *L. innocua*. Plasmid-mediated resistance to tetracycline is due to a *tetL* (minocycline sensitive) or *tetS* (minocycline resistant) gene. Interestingly, the *tetS* gene, while first noted in the genus *Listeria*, has been detected in *Enterobacter faecalis*.

Prevention from Infection

There have been many outbreaks of listeriosis involving a wide variety of foods. Coleslaw, cold-smoked salmon, delicatessen meats, hot dogs, pâtés, soft cheese, and even pasteurized milk have been implicated in major outbreaks. In addition, because *L. monocytogenes* is ubiquitous in nature, it may not be possible to eliminate it completely from the food chain.

Many initiatives have been introduced by government agencies in an attempt to ensure that foods are safe. Newer technologies are also currently being investigated as means of preventing the growth of *L. monocytogenes* in foods. Irradiation is one

example that has been approved for certain foods, including meats, fruits, vegetables, and spices. (*See* **Irradiation of Foods: Basic Principles; Applications.**)

Numerous recommendations have been put forth by many scientists and government agencies regarding ways to reduce the exposure to *L. monocytogenes* and the risk of foodborne listeriosis. These are summarized in Table 5. The incidence of *L. monocytogenes* in ready-to-eat foods ranges from 1 to 10%. It is certain, therefore, that most healthy people around the world consume foods containing low levels of this organism, without getting sick. Although the minimum infectious dose (MID) for *L. monocytogenes* is still unknown, if it were low, one would expect many more cases of listeriosis to be reported than are currently observed. Most countries do not have active human listeriosis surveillance programs in place. It appears that the incidence of human listeriosis is similar in countries with and without a "zero tolerance" policy in place.

Mechanisms of Pathogenesis

It is presumed that most infections of *L. monocytogenes* occur in the gastrointestinal (GI) tract, where invasion is thought to occur. While it is believed to be a resident of the GI tract for a short period of time, there have been reports describing patients with a long incubation time (up to 90 days) after eating contaminated food. In this instance, the ability to remain associated with the GI tract for extended periods is still not clear or defined and is currently an area of active research.

There is much speculation on why most clinical isolates belong to serovars 1/2a, 1/2b, and 4b; however, factual data in the form of animal studies or cell-culture experiments are lacking. In 1979, Seeliger and Hohne described a serotyping scheme for the genus *Listeria*. According to this designation, the letters a, b, c, etc. refer to the flagellar antigens. The numerical aspect (i.e., 1/2, 3, 4, etc.) describes the somatic antigens, being mostly techoic acid-associated (Table 3). While much work has been done in the area of pathogenesis, it is important to remember that these studies have utilized the serotype 1/2a (e.g., strains 10403S, EGD, NCTC7973, and Mack) or those belonging to serotype 1/2c (LO28). Since serotype 4b is responsible for a large number of reported outbreaks, it is quite probable that significant differences exist when compared to those serotypes currently being studied. For example, several multilocus gel electrophoresis-based studies have classified *L. monocytogenes* into two genetic groups: one that includes 'a' and 'c' serotypes (1/2a, 1/2c, 3a, and 3c) and another including the 'b' serotypes (1/2b, 3b, and 4b). Thus, while currently known virulence factors seem to be a common denominator in the pathogenesis of *L. monocytogenes*, certain class-specific host–pathogen interactions may explain the tropisms observed related to serotype.

L. monocytogenes has proved to be an extremely useful tool for examining protective immunity and immune responses during systemic infections. Upon ingestion of food contaminated with *L. monocytogenes*, the organism is believed to invade the intestinal epithelium and/or Peyer's patches. From here, the bacteria enter the draining lymph node and disseminate via the bloodstream to the liver and spleen. In most cases, infections of *L. moncytogenes* are limited, with clinical symptoms appearing in the immunocompromised, pregnant women, and neonates. However, it is equally important to note that listeriosis has one of the highest case fatality rates of all foodborne bacterial infections.

Upon injection of the organism into the bloodstream of mice, *L. monocytogenes* is usually cleared. Studies have shown that over 60% of organisms are taken up by the liver within 10 min of intravenus inoculation, with hepatocytes being the major site of replication in the liver. Hepatocytes contain over 90% of the organism after 6 h of infection.

Table 5 How to lower the risk of acquiring foodborne listeriosis

Group of interest	*Food item*	*Recommendations*
'Normal' individuals	Meat, fish, poultry	Cook all food of animal origin thoroughly
	All foods	Separate uncooked from cooked
	Milk and milk products	Avoid unpasteurized milk
	Nonfood items	Properly wash hands, knives, and kitchen utensils
	Fruits and vegetables	Wash all fruits and raw vegetables prior to consumption
High-risk (elderly, pregnant, or immunocompromised) individuals	Dairy products	Do not consume soft cheeses such as Brie, Camembert or feta (hard cheese, cottage, processed and yogurt are OK)
	Leftovers	Heat thoroughly to 70°C for a minimum of 2 minutes
	Hot dogs/wieners	Heat to 70°C for a minimum of 2 minutes
	Deli meats	Avoid unless thoroughly heated
	Pâté	Should be avoided

The innate component of the immune system is responsible for controlling the replication of the organism during the first 3 days of infection, based on studies performed in mice. This response will either eliminate the organism or prevent sepsis. The T-cells, part of the cell-mediated immune system, mediate a protective primary as well as a memory response to *L. monocytogenes*. Within the liver, macrophages known as Kupffer cells are believed to trap the majority of *Listeria* that enter. However, it has been shown that over 90% of liver bacteria are associated with hepatocytes, and not Kupffer cells. Currently, it is believed that there is a physical interaction between the Kupffer cells and neutrophils that is mediated via intercellular adhesion molecule-1, thereby promoting resistance against *L. monocytogenes*. Chemokines and other adhesion molecules cause the efflux of neutrophils and monocytes from the bone marrow to the peripheral blood, probably via interleukin-6 and macrophage colony-stimulating factor stimulation, respectively.

L. monocytogenes is considered an intracellular pathogen because it is able to enter a host cell, move across the cell cytoplasm, and enter neighboring cells without ever becoming extracellular (Figure 2). The infection process begins with the internalization of the organism into cells that may or may not be phagocytic. In the case of phagocytes such as macrophages, the phagocytosis process itself causes the organism to be taken within. For nonphagocytic cells, induced phagocytosis (known as invasion) occurs. The plasma membrane envelops the bacteria upon contact. The bacteria reside within membrane-bound vacuoles prior to lysing of the vacuolar membrane. Once free within the cell cytoplasm, *Listeria* are able to replicate and become covered with host actin filaments. The host actin filaments are rearranged to a polar tail, consisting of short actin filaments and other host proteins. The formation of the actin 'tail' at one end of the organism 'propels' it to move through the cytoplasm. As the organism nears the cell membrane, it protrudes and neighboring cells internalize them upon contact. In this manner, *L. monocytogenes* is able to move from one cell into another without being exposed to the 'outside' environment. Once in a neighboring cell, the bacterium is contained in a double plasma membrane, one from the previous cell (while protruding outward from the membrane) and one from the new cell (as it takes in the organism via phagocytosis).

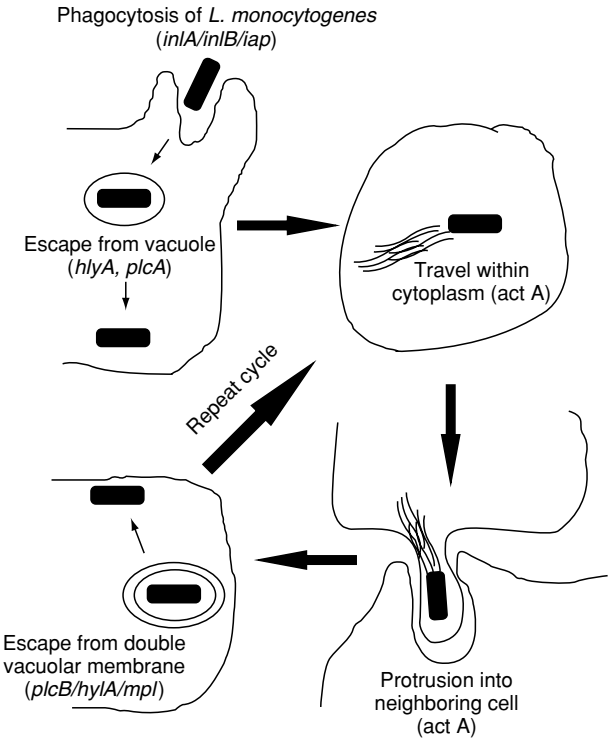

Figure 2 Proposed mechanism of entry into mammalian cells by *L. monocytogenes*. The genes implicated are listed.

The bacterium is able to internalize itself through the action of at least two proteins, internalin A (InlA) and internalin B (InlB). Currently, two pathways have been proposed for entry into mammalian cells. The first is the internalin-E-cadherin pathway, in which the invasin from *L. monocytogenes* interacts with the β-1-integrin adhesion molecule thought to be connected to the cytoskeleton. The second pathway is InlB-dependent, sharing similarities with the growth-factor-mediated signaling pathways. InlB has recently been shown to bind the extracellular domain of MET (also known as the hepatocyte growth factor (HGF)), a receptor tyrosine kinase. Upon binding to MET, InlB is proposed to cause a 'scattering' of epithelial cells. The protein p60, encoded by the gene *iap*, has been shown in experiments to be essential in the invasion process. Deletion of the *iap* gene is lethal. Believed to be a mureine hydrolase, its role is still unclear. What is known, however, is that cells producing reduced levels of p60 show a 'rough' morphology and exhibit reduced adherence, invasiveness, and virulence.

The bacterium is able to escape the vacuolar membrane through the action of listeriolysin O, a pore-forming toxin encoded by the gene *hylA*. Phosphatidylinositol-phospholipase C, a member of the pore-forming family of thiol-activated cytolysins, also contributes to the organism's ability to escape the vacuole. ActA is a surface protein with a membrane anchor at its carboxyl-terminal end. Its amino-terminal has been shown to be essential for actin

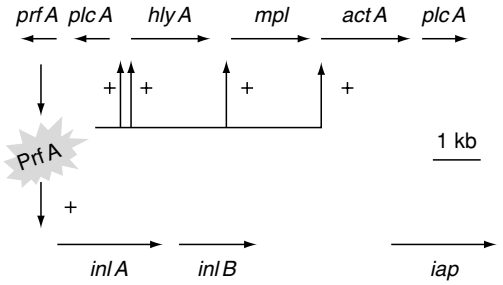

Figure 3 PrfA, the master regulator of most virulence genes found on the chromosome of *L. monocytogenes*.

assembly. ActA is responsible for the intracellular actin-based motility. It is believed that ActA is able to recruit the host actin filaments and make them assemble at one pole of the bacterium. This actin tail formation would then be the driving force in the intracellular movement observed using microscopy technology. The lecithinase PlcB acts on the double-membrane vacuole that the bacteria faces when it reaches its neighboring cell. PlcB is synthesized as a precursor and is cleaved by a zinc metalloprotease, shown to be encoded by the *mpl* gene. Listeriolysin is also implicated in escape from the second vacuole.

Regulation of the virulence genes that allow for intracellular movement, and thus escape from the immune surveillance system, all fall under the pleiotropic activator, PrfA (Figure 3). Encoded by the gene *prfA*, PrfA is able to regulate the expression of many of the virulence genes described, including *plcA*, *hlyA*, *mpl*, *actA*, and *inlAB*. A 27-kDa protein, PrfA, has been shown to activate the transcription of these genes and has been shown to be a DNA-binding protein (Figure 3).

The regulation of *prfA* itself is complex. It has been shown that at temperatures below 30 °C, PrfA-dependent genes are not transcribed, since there is a lack of *prfA* transcription. Shifting the temperature to 37 °C causes expression (transcription, translation) of the *prfA* gene, which in turn activates the virulence gene cluster. Addition of activated charcoal to a culture medium also has the same effect. Presumably, the charcoal somehow sequesters a signal molecule, which permits the transcription of *prfA* and, thus, the virulence genes. Increasing glucose concentrations in the medium turn off virulence gene expression, by directly influencing the expression of *prfA*. While the *prfA* gene product appears to auoregulate its own expression and synthesis, there is certainly more to the complexity of regulation of virulence genes and their prime activator, PrfA. A member of the cyclic AMP (cAMP) receptor protein (CRP)/FNR (regulator of fumarate and nitrate reduction) family of bacterial regulators, it is currently believed that PrfA is a regulatory protein that binds its symmetric target sequence and enhances (or, alternatively, inhibits) the binding of RNA polymerase, thereby influencing transcription. The Crp (or catabolite activator protein) is a regulatory protein that modulates the expression of genes involved with catabolite repession, whereas the Fnr family has been shown to regulate the cellular response to anaerobic growth.

See also: **Irradiation of Foods**: Basic Principles; Applications; **Microbiology**: Detection of Foodborne Pathogens and their Toxins

Further Reading

Charpentier E and Courvalin P (1999) Antibiotic resistance in *Listeria* spp. *Antimicrobial Agents and Chemotherapy* 43: 2103–2108.

Cousens LP and Wing EJ (2000) Innate defenses in the liver during *Listeria* infection. *Immunological Reviews* 174: 150–159.

Farber JM and Peterkin PI (1991) *Listeria monocytogenes*, a food-borne pathogen. *Microbiological Reviews* 55: 476–511.

Goebel W and Kuhn M (2000) Bacterial replication in the host cell cytosol. *Current Opinion in Microbiology* 3: 49–53.

Janda JM and Abbot SL (1999) Unusual food-borne pathogens: *Listeria monocytogenes*, *Aeromonas*, *Plesiomonas*, and *Edwardsiella* species. *Clinics in Laboratory Medicine* 19: 553–582.

Tage AJ (1999) Listeriosis: recognizing it, treating it, preventing it. *Cleveland Clinic Journal of Medicine* 55: 375–380.

Temple ME and Hahata MC (2000) Treatment of listeriosis. *The Annals of Pharmacotherapy* 34: 656–661.

LITHIUM

M Anke, U Schäfer and W Arnhold, Friedrich-Schiller-Universitat, Jena, Germany

History, Nature, and Function

Lithium (Li) is the lightest alkali metal (atomic weight 6.94). It shows physicochemical anomalies which result from the unexpectedly great difference between the electronegativity of this element and that of sodium, potassium, and rubidium. These differences are seen in both the properties and the reactions of this silvery white metal which was discovered by Arfvedson (1818) in Berzelius's laboratories. The new alkali element was detected when a fossil known as petalit was analyzed. It was called lithion (stone) because it was first found in the mineral kingdom.

The Earth's crust contains 50–65 mg lithium kg^{-1}. The geological origin of the soil is reflected in its lithium content. The plants of slate, keuper, and gneiss weathering soils are richer in lithium than those of diluvial and alluvial formations (Table 1).

Therapy and Toxicity

In 1843, lithium carbonate was introduced into the materia medica as a new solvent for stones in the bladder by the surgeon Ure. In 1859, the internist Garrod recommended a therapy with lithium salts for a wide range of diseases and complaints, especially gout, urinary calculi, rheumatism, mania, depression, and headache. All of them were grouped under the general heading of the uric acid diathesis which became a major unifying medical principle for almost a century. In 1941, however, this hypothesis was declared to be ill-founded.

The fascinating discovery of the specific antimanic effect of the lithium cation by the psychiatrist Cade in 1949 initiated the career of this chemically simple drug as a very potent substance against symptoms of manic-depressive illness. After numerous hindrances due to the lack of knowledge of its biochemical and pharmacokinetic properties had been overcome, lithium developed into a safely used psychopharmacological agent. Improvements in its monitoring, especially by the introduction of the lithium ion-selective electrode, as well as in patient compliance with medication, were also decisive. It was possible to extend the classical antimanic, antidepressive and recurrent-prophylactic action profile of lithium by an antipsychotic, antiaggressive, antisuicidal, and anti-neurotic component. Recently, topical lithium has found employment in dermatological disorders, e.g., seborrheic dermatitis and herpesvirus infections. It is promising that further applications of lithium as an antiinflammatory, antiviral, antifungal, antitumor, and immunomodulating agent, e.g., in the treatment of acquired immune deficiency syndrome (AIDS) and cancer, may become established in the future. The characteristic pharmacologically rare properties of the drug lithium are that it does not lose efficacy and does not induce addiction and dependence. Thus, a 'mechanical switch-on and -off function' in its biochemical mechanism is discussed.

However, lithium therapy (lithium exposure) also induces side-effects in the form of thyroid changes, nephrogenic diabetes insipidus, cardiovascular disturbance, epithelial damage, changes in carbohydrate metabolism, and skeletal damage in the form of bone mineralization. Patients treated with lithium usually tend to gain weight. It must be emphasized that the lithium dosage of 24–48 mmol or 167–333 mg day^{-1} (0.5–1.5 mg day^{-1} serum lithium) now used in humans is relatively high compared with the effects registered in animals. In rats, chickens, hens, pigs, and bulls, 50–100 mg lithium kg^{-1} ration dry matter reduced the feed consumption significantly. All lithium-exposed pigs took in considerably more water than control animals. The same is apparently true for humans treated with lithium, who meet these high fluid requirements with high-caloric beverages and thus consume more energy.

Table 1 The site-specific lithium content of indicator plants as a percentage of the site richest in lithium

Site	*Relative number*
Sandstone weathering soils of chalk formation	100
Slate	94
Keuper	94
Gneiss	92
Muschelkalk	83
New red sandstone (Bunte)	83
Lower strata of new red sandstone (Rot-bigende)	70
Diluvial sands	72
Granite	72
Loess	71
Syenite	68
Phyllite	65
Moor, peat	58
Boulder clay	54
Alluvial riverside soils	51

All species with >100 mg lithium kg^{-1} feed dry matter reduced life weight gain and egg production. Ruminants seem to be equally sensitive to lithium exposure. Lithium applications from 10 mg kg^{-1} feed dry matter upwards led to a significantly higher suet content.

The lower sexual aggressiveness of bulls receiving 100 and 200 mg lithium kg^{-1} ration should also be mentioned. Rats with lithium-rich ration gave birth to lighter offspring which developed more slowly.

All pigs receiving 1000 mg lithium kg^{-1} feed dry matter died within 92 days. Cases of death also occurred in animals on 500 mg lithium kg^{-1} feed dry matter.

Essentiality

Plants There are considerable differences in the tolerance of various plant species to lithium and in their ability to take up this element. Solanaceae are known to have the highest tolerance to lithium. Some members of this family accumulate more than 1000 p.p.m. lithium. On the other hand, citrus trees are the most susceptible to injury by an excess of lithium, which is reported to be toxic at a concentration of 140–220 p.p.m. in the leaves. In Russia, the beneficial effects of lithium are still used in agriculture. Lithium-containing fertilizers lead to increased starch levels and biomass in potatoes. The observed lithium-induced growth stimulation seems to allow the evaluation of lithium as an agrochemical. However, it is still not proven that lithium is an essential nutrient or an active substance for plants because they can complete their life cycle without this element. Lithium must rather be considered a toxic substance, with marked species-dependent differences in sensitivity. There is no proof yet that lithium is essential for plants. Lithium must rather be regarded as potentially phytotoxic.

Animals

Whether lithium is essential for fauna has been investigated systematically since 1976 in goats, sheeps, and rats in Germany, in Hungary, in the USA, and in Japan. The influence of a lithium-poor ration on the pre- and postnatal growth of goats was tested in 13 experiments over 13 years. It was shown that the kids of lithium-deficient goats had a 9% lower birth weight than those of control goats. In lithium-deficient rats the litter was 20–30% smaller in comparison to control animals. Moreover, the lithium-poor nutrition caused reproductive disorders in goats and rats. Lithium-poor rations of female goats did not have an effect on the intensity of estrus behavior. The first mating, however, produced a significantly worse rate of conception in these animals. Repeated services at the following ovulations improved the conception rate of lithium-deficient goats. The difference between the groups remained significant.

The increased abortion rate of lithium-deficient goats is also important. Miscarriages occurred at a rate of 14% between the third and fifth month of gravidity. The effect of lithium deficiency on sex ratio was most surprising. As a rule, hornless goats give birth to more male than female kids. Lithium-deficient goats, however, gave birth to significantly more female kids. The 5-year average showed that lithium-deficient goats produced 20% less milk than control animals.

Long-term lithium deficiency experiments with female goats allowed the analysis of the influence of the lithium-poor nutrition on the life expectancy of the animals. The investigations showed that 41% of lithium-deficient goats and 7% of control animals died during the first year of the experiments. Skin lesions occurred in individual animals with lithium-poor rations. In the meantime, it was shown, that lithium was also effective in the treatment of herpesvirus infections and of seborrheic dermatitis in humans.

Lithium deficiency did not affect the biochemical blood profile, but it changed the activity of several serum enzymes. As a rule, lithium deficiency reduced the serum enzyme activity, mainly the enzymes of the citrate cycle (ICDH, isocitric acid dehydrogenase; MDH, malate dehydrogenase), of glycolysis (ALD), and of nitrogen metabolism (GLDH, glutamine dehydrogenase), for which significant differences between control and lithium-deficient goats occurred. Only creatine kinase, a stress indicator enzyme, was significantly increased in lithium-deficient goats.

Owing to the particular role of monoamine oxidase (MAO) in manic-depressive diseases, chronic schizophrenia, and unipolar depression, this enzyme was also investigated in the liver of control and lithium-deficient goats. MAO activity in the hepatic tissue of the latter group was reduced by 28%.

Abnormal behavior was observed in lithium-deficient rats, and this disappeared after lithium supplementation. This assessment is interesting in so far as the MAO hypothesis is discussed as a biochemical mechanism of lithium effect in humans. In brief, the results of a lifelong lithium-poor nutrition over >10 generations of goats show that lithium may be essential to fauna, and thus, to humans as well.

Lithium in the Diet

The lithium intake of adult humans with mixed diets was systematically investigated in 10 test populations

of 7 women and 7 men, each from different regions of Germany and aged between 20 and 69 years (Table 2).

Men and women consumed the same amount of lithium-rich and lithium-poor foods and beverages. The lithium intake of both sexes doubled after the reunification of Germany and worldwide trade. Since men consume 24% more dry matter than women, they take in more lithium than women. The distribution of the individual lithium intake does not follow the Gaussian curve but it is shifted to the left (Figure 1). The lithium consumption was characterized by an extreme variation range, which may be caused by water and beverages.

Adults of both sexes with mixed diets consumed significantly more lithium with increasing age. On average, elderly test persons consumed 30–46% more lithium. They clearly preferred lithium-rich foodstuffs. The lithium intake per day on individual sites varies between 200 and 2400 μg day^{-1} on average. The postulated potential lithium requirement of 100 μg day^{-1} is met on the average of the population. Ten percent of the men tested consumed < 100 μg lithium day^{-1} as a weekly average. Worldwide, a lithium intake for adults between 660 and 3420 μg day^{-1} is calculated.

The Lithium Content of Foodstuffs and Beverages

The lithium content of water seems most likely as an explanation for the different lithium intake. Water from swamp and peat sites contained eight times more lithium than that from diluvial sands (Table 3). In addition, a relationship between high lithium contents in drinking water on the one hand and low rates of crimes, suicides, and arrests on the other was controversially discussed.

In line with the low lithium content of cereals, flour and other cereal products contained little lithium. Honey and sugar are also extremely poor in lithium. Ready-to-serve soups with meat and eggs were richer in lithium, whereas various puddings, macaroni, and vermicelli usually contained < 1 mg lithium kg^{-1} dry matter (Table 4).

Bread, cake, and pastries are usually poor sources of lithium. On average, they contained less lithium than wheat flour. The addition of sugar apparently leads to a further reduction of the lithium content in bread, cake, and pastries.

All vegetables and potatoes contain > 1.0 mg lithium kg^{-1} dry matter. Peeling potatoes decreases their lithium content, as potato peel stores more lithium

Table 2 Lithium intake of German adults with mixed diets depending on time (μg day^{-1})

Year (n;n)	*Women*		*Men*		*Fp*	*%*[a]
	SD	*x*	*x*	SD		
1988 (196; 196)	444	374	419	432	< 0.01	112
1992 (294; 294)	724	713	990	1069		139
Fp		< 0.001			–	
%[b]		191		236		

[a]Women = 100%, men = *x*%;
[b]1988 = 100%, 1992 = *x*%.
x, arithmetical mean; SD, standard deviation; *n*, number; Fp, significance level in one or multiple factorial variance analysis.

Table 3 The lithium concentration (μg l^{-1}) of drinking water from different geological sources

Source of water	*Mean*	SD
New red sandstone (Bunte)	4.2	3.4
Gneiss	4.3	4.8
Phyllite	6.9	3.4
Diluvial sands	7.7	5.8
Slate	11	10
Boulder clay	19	22
Syenite	22	18
loess	27	23
Muschelkalk, keuper	34	26
Lowland swamps	60	20

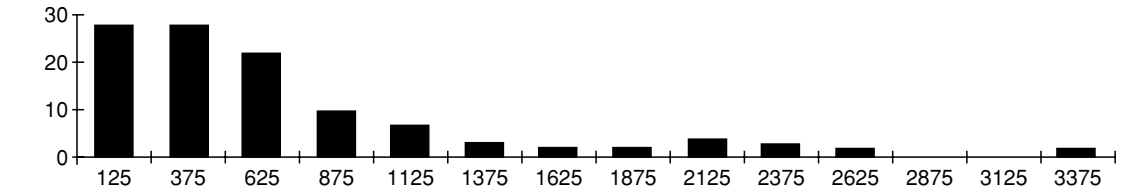

n	12	20	15	7	5	2	1	1	3	2	1	0	0	1

Figure 1 The mean weekly lithium intake of men in μg day^{-1}.

Table 4 The lithium content of several groups of foodstuffs

Group of foodstuffs	*μg/kg resp. μg/l*	*Variation*
Beverages	192	18–329
Sugar, honey	363	199–527
Bread, cake	612	317–935
Cereal products	888	538–1391
Luxuries	1678	372–3737
Spices	1850	1046–2316
Fruits	2846	383–6707
Meat	3217	2379–3844
Milk, dairy products	3626	1070–7533
Eggs	7373	SD 6500

than the inner part of the potato that is commonly eaten.

As a rule, fruits contain less lithium than vegetative parts of plants (vegetables). Lemons and apples contained significantly more lithium, with about 1.4 mg kg^{-1} dry matter, than peas and beans, which, like the different kinds of cereals grains, are extremely lithium-poor as seeds.

Owing to the small amounts used in their application, spices do not contribute much lithium to the diet. It is surprising that mustard is relatively lithium-rich, with 3.4 mg kg^{-1} dry matter, whereas mustard seed contains extremely little lithium.

Various semiluxury foods usually provide only modest amounts of lithium. The total amount in tea and coffee, not their water-soluble fraction in the beverage, was registered. Their low lithium content indicates that insignificant amounts of lithium enter the diet via these beverages. The lithium content of chocolates, chocolate candies, and sweets amounted to about 0.5 mg kg^{-1} dry matter. Cocoa is somewhat richer in lithium. The addition of sugar in chocolates reduces their lithium content.

On average, eggs, meat, sausage, and fish deliver significantly more lithium per kg of dry matter than most cereal foodstuffs. Eggs, liver, and kidneys of cattle had a mean lithium content of 5 mg kg^{-1}. Beef and mutton contain more lithium than poultry meat. Green fodder and silage consumed by cattle and sheep are much richer in lithium than the cereals largely fed to poultry. Sausage and fish contain similar amounts of lithium to meat. There were significant differences between the lithium content of milk: milk contains a high lithium concentration of 10 mg kg^{-1} dry matter.

In contrast to milk, curd cheese and other cheeses only retain 20–55% of lithium in the original material available for human nutrition. The main fraction of lithium certainly leaves cheese and curd cheese via the whey. Like margarine made of plant fat, butter is also lithium-poor, with 1.2 mg kg^{-1} dry matter.

The fact that lithium is easily soluble may also be the reason for the high lithium content of beverages. Cola and beer deliver considerable amounts of lithium for humans, and this must be taken into consideration when calculating the lithium balance of humans.

Identification of Lithium Status

As already demonstrated in the literature, blood serum best reflected lithium status, followed by hair. Milk reflects the lithium status well. Lungs, spleen, and carpal bones also reflected the different lithium supply. Other parts of the body indicated the lithium status only insignificantly and the cardiac muscle not at all.

Determination of Lithium

Lithium may be determined in foods and biological samples with the same techniques employed for sodium and potassium. However, the much lower levels of lithium compared with these other alkali metals, mean that techniques such as flame photometry often do not show adequate sensitivity. Flame (standard addition procedure) or electrothermal atomic absorption spectrophotometry are the most widely used techniques after wet or dry ashing of the sample. Corrections may have to be made for background/matrix interferences. Inductively coupled plasma atomic emission spectrometry is not very sensitive for this very low-atomic-weight element.

See also: **Sodium**: Properties and Determination

Further Reading

Anderson CE (1990) Lithium in plants. In: Bach RO and Gallicchio VS (eds) *Lithium and Cell Physiology*, pp. 25–46. New York: Springer-Verlag.

Anke M (1997) Lithium. In: O'Dell BL and Sunde RA (eds) *Handbook of Nutritionally Essential Mineral Elements*, pp. 465–477. New York: Marcel Dekker.

Anke M, Arnhold W, Glei M *et al*. (1995) Essentiality and toxicity of lithium. In: Kosla T (ed.) *Lithium in the Trophic Chain, Soil–Plant–Animal–Man*, pp. 17–42. Proceedings of International Symposium. Warsaw: Agricultural University.

Anke M, Arnhold W, Glei M *et al*. (1997) Lithium intake of adults as a function of sex, time, location, season, age and time. In: Ermidou-Pollet S (ed.) *International Symposium on Trace Elements in Human: New Perspectives*, pp. 181–198. Athens, Greece: Marogianni.

Anke M, Arnhold W, Groppel B and Krause U (1991) The biological importance of lithium. In: Schrauzer GN and K-F Klippel (eds) *Lithium in Biology and Medicine*, pp. 148–167. New York: VCH Weinheim.

Anke M, Groppel B, Grün M, Kronemann H and Riedel E (1981) Effects of Li-poor rations in ruminants. In: Szentimihályi S (ed.) *Proceedings of International*

Conference on Feed Additives, pp. 245–248. Budapest, Hungary: Hungarian Society of Agricultural Sciences.
Anke M, Groppel B, Kronemann H and Grün M (1983) Evidence for the essentiality of lithium in goats. In: Anke M, Baumann W, Bräunlich H and Brückner C (eds) *Spurenelement-Symposium, Lithium*, pp. 58–65. Universität Leipzig, Universität Jena. Oberlungwitz: VEB Kongreß- und Werbedruck.
Arnhold W and Anke M (1988) Möglichkeiten der Identifizierung des Lithiumstatus. *Zentralblatt für Pharmasie Pharmacotherapeutik und Laberdiagnostic* 127: 243–244.
Horrobin DF (1991) Lithium effects on fatty acid metabolism and their role in therapy of seborrhoeic dermatitis and herpes infections. In: Schrauzer GN and Klippel K-F (eds) *Lithium in Biology and Medicine*, pp. 65–79. Weinheim: VCH.
Kabata-Pendias A and Pendias H (1992) Lithium. In: Kabata-Pendias A and Pendias H (eds) *Trace Elements in Soils and Plants*, 2nd edn, pp. 89–92. Boca Raton: CRC Press.
Ono T, Wada O and Yamamoto M (1992) Study on the essentiality of lithium. *Biomedical Research and Trace Elements* 3: 41–47.
Pickett EE and O'Dell BL (1992) Evidence for dietary essentiality of lithium in the rat. *Biological Trace Element Research* 34: 299–319.
Schäfer U (1997) Essentiality and toxicity of lithium. *Journal of Trace and Microprobe Techniques* 15: 341–349.
Schäfer U (1998) Past and present conceptions concerning the use of lithium in medicine. *Journal of Trace Microprobe Techniques* 16: 535–556.
Schrauzer GN and Klippel K-F (1991) *Lithium in Biology and Medicine*. Weinheim: VCH.
Szilágy M, Anke M and Szentmihályi S (1985) The effects of lithium deficiency on the metabolism of goats. In: Pais I (ed.) *New Results in the Research of Hardly Known Trace Elements*. Proceedings of International Symposium 1984, pp. 167–172, Budapest: University of Horticulture.

LIVER

Contents

Structure and Function

D G Cramp, South Wonston, Winchester, UK

Gross Structure

The liver is the largest solid organ in the body, weighing between 1200 and 1500 g. It lies in the right hypochondrium and is roughly wedge-shaped, with its apex to the left side and base to the right. Grossly, it appears to be divided into a larger right and a smaller left lobe by the falciform ligament. However, this has no apparent functional significance, as the blood supply to the two lobes is almost equal. This blood supply is derived from two sources: the hepatic artery, usually a branch of the coeliac axis, and the portal vein, formed from the mesenteric vein and the splenic vein.

The total hepatic blood flow is some 1300 ml min^{-1}, approximately 25% of the resting cardiac output, with the hepatic artery supplying about 25–30% of blood by volume and the low-pressure portal vein the remainder. Both of these vessels run in the free edge of the lesser omentum to the liver hilum before dividing into major right and left branches. Blood leaves the liver through several hepatic veins that drain directly into the inferior vena cava, which runs in a groove between the two lobes. Three such hepatic veins enter just below the diaphragm, with a number of smaller veins draining the posterior aspect of the right lobe directly into the inferior vena cava.

The biliary system collects, stores, and concentrates bile in the gall bladder, prior to its delivery to the duodenum as required. The main right and left hepatic ducts join in the liver hilum to form the common hepatic duct. The gall bladder itself lies on the underside of the right lobe of the liver, with the cystic duct running backwards and upwards in order to join the hepatic duct; the resulting common bile duct runs with the hepatic artery and portal vein and then passes behind the first part of the duodenum

and the head of the pancreas before it enters the duodenum.

Fine Structure

At the tissue level, the liver is comprised of a large series of channels lined with endothelial cells, the sinusoids, which run between layers or plates of hepatocytes. The endothelial cells possess fenestrations, which allow access of circulating blood to the sinusoidal surface of the hepatocytes. Some of the cells lining the sinusoids, Kuppfer cells, possess phagocytic activity. The space of Disse, i.e., the space between the endothelial cells and the hepatocytes, contains interstitial fluid. The membrane transport of nutrients and waste material between the space of Disse and the parenchymal cells is facilitated by the numerous microvilli, which extend the sinusoidal surface of these cells. From and around the liver branches of the hepatic artery, portal vein and bile duct are carried portal tracts, the smallest of which contain terminal branches supplying groups of sinusoids that form a functional unit, the acinus. Blood from each acinus passes into a number of efferent veins. It is the radially orientated sinusoids centered on these efferent veins that give the liver its classical lobular architecture.

Histochemical studies have demonstrated a marked heterogeneity of hepatocytes, with cells at different locations differing both structurally and functionally. This differentiation probably reflects the many specific metabolic functions that the liver has to meet. The overall functional capacity of the liver (liver functional mass) is highly dependent on the number and activity of these hepatocytes and their contact with circulating blood (effective liver perfusion).

Functions of the Liver

The anatomical site and cellular architecture of the liver enable it to perform many diverse metabolic functions that are essential to the body. Its position, between the digestive tract and the general circulation, ensures that substances ingested orally and absorbed by the intestine must normally pass to the liver. As a result, the liver has developed the ability not only to receive, process, and store these substances but also to release them or their metabolites in a highly regulated fashion by monitoring demand and responding to multiple hormonal and neural stimuli.

A cardinal role of the liver is to regulate the supply and utilization of energy-containing substrates through its participation in the integrated control of carbohydrate, protein, and lipid metabolism. In the body, there is a fine balance between anabolic and catabolic processes, between supply and demand. This state of metabolic homeostasis is essential for the maintenance of an adequate and continual supply of fuel to the tissues, especially cerebral tissue, as the body adapts to periods of relative plenty alternating with periods of fasting. This is facilitated by three processes: providing fuel from ingested food to meet immediate energy demands; storing fuel reserve as glycogen in the liver and muscle, at the same time replenishing tissue protein lost since the last meal; diverting excess energy-rich substrate into triglyceride for transport to adipose tissue for storage. As for any tissue, these metabolic processes in the liver are dependent on several factors, including the circulating concentration of substrate, blood flow to tissue, cell permeability, the mechanism of entry to the metabolic pathways and their hormonal regulation. (*See* **Amino Acids**: Metabolism; **Carbohydrates**: Digestion, Absorption and Metabolism; **Fatty Acids**: Metabolism; **Glycogen**.)

The liver is the major site of plasma protein synthesis, including albumin, the α- and β-globulins, clotting factors, and transport proteins. These proteins have important roles in body homeostasis: regulation of blood coagulation and hemostasis; the transport and optimization of circulating concentrations of calcium and magnesium ions, bilirubin, fatty acids, and hormones. Also, albumin contributes to the maintenance of the plasma colloidal oncotic pressure and hence the control of body fluid distribution. There is a close relationship between the liver, gut, and muscles with respect to amino acid metabolism. The average adult turns over about 2% of total body protein, approximately 250 g per day, with muscle the major source, contributing 140 g, and dietary intake accounting for 90 g. After a meal, approximately 25% of the amino acid nitrogen absorbed into the portal bloodstream reaches peripheral tissues because the liver has the ability to regulate the rate of gluconeogenesis and transamination, thereby modifying the circulating plasma amino acid composition. The aromatic amino acids phenyalanine, tyrosine, and methionine are preferentially metabolized to urea, while the branched-chain amino acids valine, leucine, and isoleucine are selectively able to reach peripheral muscle for reincorporation into protein. A major synthetic function of the liver is the formation of bile acids from cholesterol and their secretion into the intestine, thereby generating bile flow and facilitating dietary fat emulsification and absorption. (*See* **Bile**.)

Substances circulating in the systemic bloodstream have access to the liver by way of the hepatic artery and the mesenteric artery. The liver is the main site of metabolic transformation and detoxification of

endogenous compounds, including bilirubin, ammonia, and thyroid, steroid, and other hormones, as well as exogenous foreign compounds.

Role of Liver in Postabsorptive Events

Over a 24-h period, a subject of average body mass utilizes about 7560–8400 kJ (1800–2000 kcal), derived from 75 g of protein (mainly muscle) and some 160 g of triglyceride from adipose tissue. During that time, the liver releases about 180 g of glucose, 80% of which is totally oxidized by nervous tissue, mainly the brain. The rest is metabolized by those tissues for which glucose is the preferred energy substrate, namely red blood cells, hemopoietic tissue, and the renal medulla. The endproducts of this glycolysis, pyruvate and lactate, are recycled to the liver and converted to glucose. This carbon shuttle, the Cori cycle, provides a means of sparing gluconeogenesis from protein. (*See* **Glucose**: Function and Metabolism; Maintenance of Blood Glucose Level.)

Following a large meal with significant carbohydrate content, the liver takes up glucose in response to increased insulin secretion and suppression of glucagon. This glucose is converted into glycogen and stored. If the meal has a relatively low carbohydrate content, the liver will continue to produce glucose in response to raised glucagon levels, notwithstanding a normal insulin response to food. Insulin and glucagon regulate the intracellular concentration of cyclic adenosine monophosphate (cAMP), which in turn initiates complex enzyme changes that modulate glycogen synthesis and breakdown as well as gluconeogenesis. In a normal subject, following the ingestion of a 100-g glucose load, the liver retains approximately 60–80%, the rest escaping into the systemic circulation. The liver contains about 10% by weight of glycogen (approximately 150 g), less than the daily body glucose requirement of some 180–200 g; thus, it is only able to meet short-term requirements.

The conventional model suggests that glycogenolysis is able to maintain glucose homeostasis for some 18 h into a fast and is then superseded by gluconeogenesis. The switch to gluconeogenesis is initiated by two processes: a raised glucagon: insulin ratio, which increases the hepatic cAMP concentration, and increased peripheral tissue lipolysis, leading to a higher intrahepatic concentration of free fatty acids, the oxidation of which yields an increase of gluconeogenic precursors such as acetylcoenzyme A (acetyl-CoA).

It is the action of insulin and its counter-regulatory hormones that integrate intermediary metabolism in both the liver and peripheral tissues. The systemic regulation of protein and lipid metabolism, through a variety of actions that affect the flux of protein and lipid substrates to the liver, complements glucose regulatory mechanisms.

The hepatic actions of insulin lead to fuel conservation, by optimizing glucose availability, reducing fatty acid oxidation – and thus ketone body production – and diverting the fatty acids into triglyceride for transport to peripheral tissues where they are stored.

In the fasting state, insulin modulates hepatic glucose production by restraining glycogenolysis and increasing gluconeogenesis. In the fed state, insulin levels rise, and there is virtually total inhibition of both these processes so that hepatic glucose production drops to zero. This is achieved by inhibition of phosphorylase and the key gluconeogenic enzymes pyruvate carboxylase, phosphoenolpyruvate carboxykinase and fructose 1,6-diphosphate, probably through inhibition of cAMP-dependent protein kinases. Insulin is also essential to the disposition of incoming glucose through its enhancement of glucokinase activity and glucose phosphorylation. The glucose 6-phosphate formed can follow several alternative metabolic routes:

1. Glycogen formation following the activation of glycogen synthase: this step is subject to the limitations on the capacity of hepatocytes to store glycogen. Excess glucose 6-phosphate is converted to pyruvate, which, through activation of pyruvate dehydrogenase and acetyl-CoA carboxylase, is diverted into fatty acid synthesis.
2. Increased flux of glucose 6-phosphate into the pentose phosphate shunt increases the availability of the reduced form of nicotinamide adenine dinucleotide phosphate for fatty acid synthesis. The fatty acids are esterified with glycerol 3-phosphate also formed during glycolysis. The resulting triglycerides are incorporated into very-low-density lipoproteins (VLDLs) and secreted by the liver to extrahepatic sites for storage. (*See* **Lipoproteins**.)
3. As fatty acid synthesis is accelerated, fatty acid oxidation and ketogenesis are inhibited. The first committed step of fatty acid synthesis is malonyl-CoA formation. Increased levels inhibit the enzyme carnitine acyltransferase I, which facilitates the transport of fatty acids, as fatty acyl-CoA, into mitochondria for oxidation. If fatty acids cannot be oxidized, they become available for esterification; furthermore, the blocking of fatty acid oxidation also prevents ketogenesis, owing to the lack of precursor acetyl-CoA.

Glucose Homeostasis

During a brief fast, a precipitate fall in the blood glucose concentration is avoided by drawing upon

hepatic glycogen stores. During long-term fasting, i.e., starvation, gluconeogenesis plays the most significant role in maintaining blood glucose concentrations. If the pathways activated during a brief fast persisted through a period of prolonged food deprivation, the loss of body protein to provide glucose would be extremely rapid. However, with prolonged starvation the brain is progressively able to adapt to utilizing lipid-derived ketone bodies as a substitute for glucose. The reduction in protein breakdown, as reflected by a fall in urinary nitrogen, and the concomitant decreased gluconeogenesis result from the increased utilization of ketone bodies. They not only decrease the requirement of the brain for glucose derived from protein but also have a direct effect on reducing peripheral tissue glucose utilization and the flux of alanine from muscle.

Cholesterol and Bile Acid Metabolism

The liver has a primary role in cholesterol metabolism within the body because it is the site of most of the endogenous cholesterol synthesis, it handles exogenous cholesterol from the diet and extrahepatic tissues received by way of plasma lipoproteins, and it is the only organ with the capacity to excrete cholesterol from the body.

Hepatocytes have the capacity both to synthesize and to take up cholesterol from lipoprotein remnants. Negative feedback control of the cholesterol synthesis pathway and control of the cell surface receptor population ensure that intracellular accumulation of cholesterol does not occur. Some cholesterol may be secreted from hepatic cells in VLDL, the vehicle for triglyceride transport to adipose tissue. On removal of triglyceride, the lipoproteins become smaller, cholesterol-rich particles, low-density lipoproteins (LDLs), which are recognized by LDL receptors on hepatocytes. Following internalization, cholesterol is released from the disrupted lipoprotein.

The dietary cholesterol intake is about 600–800 mg daily and is mainly derived from dairy products, eggs, and meat. Cholesterol derived from dietary fat, together with that from the bile, is taken up by mucosal cells of the intestinal wall with about 30–60% of the total presented being absorbed or reabsorbed. The efficiency of absorption increases with the saturated fat content of the diet. Absorbed cholesterol reaches the liver in chylomicron remnants and is taken up by chylomicron receptors.

The primary bile acids, cholic and chenodeoxycholic acid, are synthesized in the liver from cholesterol, conjugated with the amino acids taurine or glycine to form bile salts, and then excreted by an active transport mechanism into the bile canaliculi and, eventually, into the duodenum. In the distal ileum, there is an active reverse transport mechanism, whereby bile salts are conserved by reabsorption after fulfilling their digestive function. The enterohepatic circulation is extremely efficient: the entire body pool of bile salts is cycled through the gut roughly twice during the course of a meal, with so little being lost that only about 5% needs replacement by *de-novo* synthesis. The small amount that reaches the stools (about 4%) is acted upon by bacteria in the gut lumen to yield, by 7-α hydroxylation, the secondary bile acids deoxycholic and lithocholic acids. All bile salts are powerful detergents, solubilizing lipids by enclosing them in bile salt aggregates to form micelles. These are structured so that hydrophobic groups are orientated towards the lipid-rich interior of the micelle, while the hydrophilic hydroxyl (OH) and carboxyl (COOH) groups are on the outside.

Cholesterol and lecithin are excreted into the bile, but if their concentration is too high, relative to that of bile salts, the bile becomes supersaturated, and precipitation of cholesterol occurs, particularly if a nidus for crystal formation is present. This is the pathophysiological basis of gallstone formation. The active transport of bile salts is the main driver of bile formation and secretion. Failure leads to intrahepatic cholestasis. A bile-salt-independent contribution to bile secretion is the result of active sodium transport in the distal canaliculus, with bicarbonate being secreted into the small bile ductules under the stimulus of the hormone secretin. Cholecystokinin stimulates gall bladder contraction.

See also: **Amino Acids**: Metabolism; **Bile**; **Carbohydrates**: Digestion, Absorption, and Metabolism; **Fatty Acids**: Metabolism; **Glucose**: Function and Metabolism; Maintenance of Blood Glucose Level; **Glycogen**; **Lipoproteins**

Further Reading

Arias IM, Popper H, Schachter D and Shafritz DA (eds) (1982) *The Liver Biology and Pathobiology*. New York: Raven Press.

Cramp DG (ed.) (1982) *Quantitative Approaches to Metabolism*. Chichester, UK: John Wiley.

Cramp DG and Carson ER (eds) (1990) *Liver Function*. London: Chapman & Hall.

Enterohepatic Circulation

A Lanzini, University of Brescia, Brescia, Italy

Definition

The enterohepatic circulation is an anatomical and physiological entity, only part of which is vascular. It involves passage of substances that have been absorbed from the intestine via the portal vein to the liver, and their subsequent passage via the biliary system back into the intestine. Bile acid molecules are unique in that they are restricted to the enterohepatic circulation (Figure 1). They are secreted in bile as glycine and taurine conjugates, and are mainly absorbed by active transport in the distal ileum. Many other endogenous and exogenous compounds undergo some degree of enterohepatic circulation, but this usually represents a minor component of their tissue distribution. By contrast with bile acids, these substances are excreted in bile as glucuronate or sulfate conjugates, and deconjugation by colonic bacteria is essential for these molecules to reenter the enterohepatic circulation by passive colonic absorption following biliary excretion.

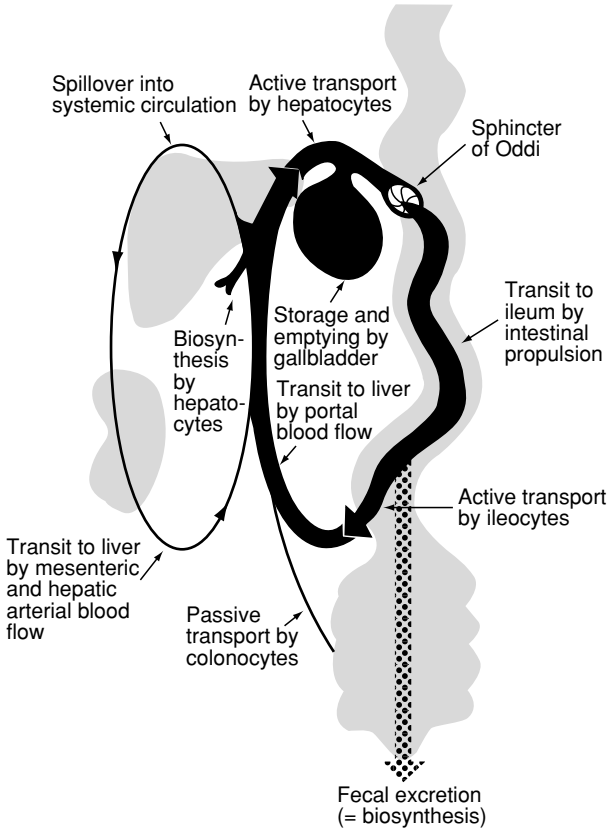

Figure 1 Schematic representation of the enterohepatic circulation of bile acids: see text for details. Reproduced from Hofmann AF (1992) Bile acids in liver and biliary disease. In: Millward-Sadler GH, Wright R, and Arthur MJP (eds) *Wright's Liver and Biliary Disease*, 3rd edn, pp. 289–316. London, UK: WB Saunders, with permission.

Bile Acids

The driving force for the enterohepatic circulation of bile acids comprises chemical and mechanical pumps, both affecting bile acid kinetics to a different extent.

Kinetics

All aspects of bile acid kinetics are measurable under physiological conditions in humans using the isotope dilution technique developed by Lindstedt. According to this technique, bile acid pool size is the mass of bile acid that dilutes a trace amount of an isotopically labeled bile acid. The fractional turnover rate is the fraction of the pool lost from the enterohepatic circulation and replaced by *de novo* bile acid synthesis each day. Chenodeoxycholic acid and cholic acid are the only two bile acids synthesized by the human liver, and are thus called 'primary' bile acids (*See* **Bile.**) These two bile acids are biotransformed in the intestine to form 'secondary' bile acids, lithocholic and deoxycholic acid. Further intestinal and hepatic biotransformation of primary bile acids results in the formation of 'tertiary' bile acids (ursodeoxycholic acid and ursocholic acid). Input rate is the amount of bile acid entering the pool each day, and represents synthesis in the case of the primary bile acids.

According to the Lindstedt's technique, the specific activity (SA = disintegrations per minute/bile acid mass) of the radiolabeled bile acid injected intravenously is measured in gallbladder bile samples collected on 4–5 successive days by nasoduodenal intubation during intravenous infusion of cholecystokinin to contract the gallbladder. The reduction of the SA over time is exponential (Figure 2) because a constant fraction rather than a constant amount of the pool is lost each day (first-order kinetics). This exponential decrease of SA is described by the following equation:

$$SA_t = SA_0\, e^{-Kt}$$

were SA_t and SA_0 are the SAs at times t and zero respectively. The extrapolated value for SA_0 on logarithmic scale gives the specific activity of the pool assuming instantaneous mixing of the injected tracer with the pool, and the slope of the line ($-K$) gives the turnover rate. Total bile acid pool size is derived by the following formula:

$$\text{Bile acid pool} = SA_{inj}/SA_0$$

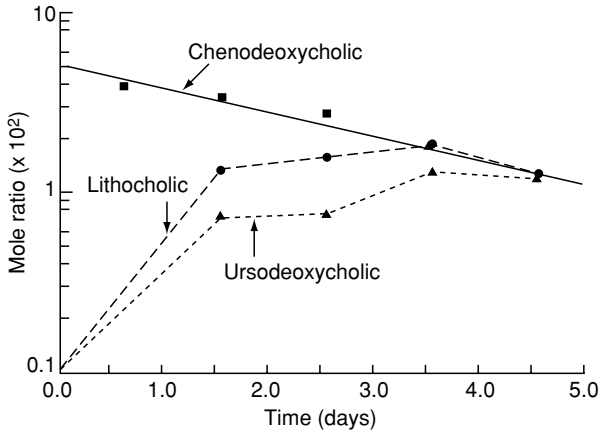

Figure 2 Decline of specific activity of a 'primary' bile acid (chenodeoxycholic) and input of 'secondary' (lithocholic acid) and 'tertiary' (ursodeoxycholic acid) bile acids in gallbladder bile samples collected over a 5-day period. Reproduced from Hofmann AF (1989) Enterohepatic circulation of bile acids. In: *Handbook of Physiology*, section 6, vol. III, pp. 567–596. American Physiological Society, with permission.

Table 1 Kinetics of individual bile acids in humans

Bile acids	*Pool size (g)*	*FTR (days^{-1})*	*Synthesis (g)*	*Input (g)*
Cholic	0.5–1.5	0.2–0.5	0.2–0.4	
Deoxycholic	0.2–1.0	0.2–0.3		0.04–0.2
Chenodeoxycholic	0.5–1.4	0.2–0.3	0.1–0.25	
Lithocholic	0.05–0.1	1.0		0.04–1.0
Total	1.3–4.0		0.28–0.61	0.08–0.3

FTR, fractional turnover rate.
From Carey MC and Duane WC (1994) Enterohepatic circulation. In: Arias IM, Boyer JL, Fausto N, Jakoby WB, Schachter D and Shafritz DA (eds) *The Liver: Biology and Pathobiology*, 3rd edn, pp. 769–786. New York: Raven Press, with permission.

where SA_{inj} is the specific activity of the injected tracer. Daily synthesis rate is derived by the following formula:

$$\text{Synthesis rate} = \text{pool size} \times -K$$

If bile acid secretion rate is measured by means of duodenal perfusion techniques simultaneously with bile acid pool size, the recycling frequency of the pool can be derived from the following formula:

$$\text{Recycling frequency} = \text{24-h bile acid secretion/bile acid pool size}$$

The kinetics of the enterohepatic circulation vary for individual bile acids, and values for each individual bile acid are reported in Table 1. Given a secretion rate of 11–40 g day^{-1} and a pool size of 1.3–4.0 g, the recycling frequency of the pool varies between 4 and 15 per day in humans.

Chemical Pumps

Intestinal absorption Two chemical pumps are involved in the enterohepatic circulation, intestinal absorption and hepatic uptake. The mechanisms involved in intestinal bile acid absorption vary with the anatomical site and consist of passive nonionic diffusion in the whole intestine, and of active absorption in the distal ileum.

The ileal transport system satisfies the criteria for active and saturable transport, and the rate of active ileal bile acid transport is described in terms of maximal transport velocity (V_{max}) and Michaelis–Menten constant (K_m). In humans, V_{max} is higher for taurocholic acid than for chenodeoxycholic and glycochenodeoxycholic acid by a factor of 4:2:1 respectively. Values for K_m are similar for conjugated and unconjugated bile acids. It appears, therefore, that active ileal transport is best suited for taurine-conjugated bile acids, i.e., for bile acids with negligible passive absorption in the proximal small intestine.

The active uptake of bile acid on the enterocyte apical membrane is mediated by the ileal Na^+-taurocholate cotransporter (iNTCP), a glycoprotein containing 348 aminoacids with predicted molecular mass of 38 kDa. This glycoprotein presumably contains seven transmembrane domains and belongs to a superfamily of Na^+-solute transporters. The expression of the sodium-bile acid cotransporter is inducible by taurocholate, suggesting that this mechanism may protect against ileal bile acid loss and subsequent formation of excess secondary bile acids in the colon. This uptake mechanism is absent in the duodenum and jejunum.

During transmembrane transport, bile acid orientation is though to be 'head ahead,' i.e., with the steroid nucleus 'head' crossing first the apical membrane followed by the side-chain 'tail.' Bile acid molecules are transported within the ileocyte by a cytosolic bile acid-binding protein constituted of 128 amino acids and with putative molecular weight of 14 kDa. Following intracellular translocation, bile acids leaves the ileocyte by a sodium-independent anion exchange mechanism. The overall efficiency of the ileal transport system in humans is very high, in the region of 90% per meal, or 70–80% per day.

By contrast with the distal ileum, bile acid absorption in the jejunum and in the colon occurs by passive nonionic diffusion alone. At the physiological pH of the upper small intestinal content (pH 5.5–6.5), about 50% of unconjugated bile acids is undissociated by comparison with only about 25% of glycine conjugates. At this pH no taurine-conjugated bile acid is protonated. These observations suggest that, on a purely physicochemical basis, passive diffusion

in the proximal small intestine is best suited for glycine-conjugated bile acids, i.e., for bile acids less efficiently absorbed by the active ileal transport system. At the slightly acidic pH of colonic content unconjugated bile acids resulting from bacterial deconjugation of bile acids escaping ileal absorption are also absorbed by passive nonionic diffusion.

Hepatic uptake Following absorption, bile acids are transported to the liver in the portal blood, and to a negligible extent in the lymph. In blood, bile acids are strongly bound to albumin, a phenomenon preventing dispersion of bile acids outside the enterohepatic circulation. The binding affinity of bile acids for serum albumin is greater for unconjugated than for conjugated bile acids, and for mono- than for di- and trihydroxylated bile acids.

Given a portal blood flow of $500\,ml\,min^{-1}$ and a bile acid concentration of $20\,\mu mol\,l^{-1}$, about $600\,\mu mol\,h^{-1}$ bile acid reaches the liver by the portal system. The process of hepatic uptake is extremely efficient, and consists of a saturable carrier-mediated process involving a 54-kDa carrier protein. Hepatic first-pass clearance of bile acids depends upon both portal blood flow and hepatic first-pass extraction of bile acids. This latter is relatively independent of flow within the physiological range of portal flows, and is mainly influenced by bile acid structure. First-pass extraction of total (conjugated plus unconjugated) bile acids ranges between 80% and 90% for cholic acid, and is about 70% for both chenodeoxycholic and deoxycholic acid.

The transport system responsible for hepatic bile acid uptake is located on the sinusoidal membrane of the hepatocyte and consists of the liver NTCP (Figure 3), a glycoprotein with 40% structural identity to the ileal iNTCP described above. Other transport systems for bile acids have been described in addition to NTCP. In particular an organic anion-transporting polypetide has been cloned from rat (oatp) and human (OATP) liver. OATP is a polyspecific Na^+-independent transport system involved in the hepatic uptake of a large number of structurally unrelated anphipatic compounds, including bromosulfophthalein, bile acids, and steroids (*See* **Bile**.)

Portal bile acids escaping first-pass hepatic extraction account for the $1–5\,\mu mol\,l^{-1}$ bile acid concentration in peripheral blood. Intravenous administration of radiolabeled bile acids is followed within a few minutes by clearance of radioactivity from plasma, suggesting that bile acids in peripheral blood are rapidly taken up by the liver, thus reentering the enterohepatic circulation.

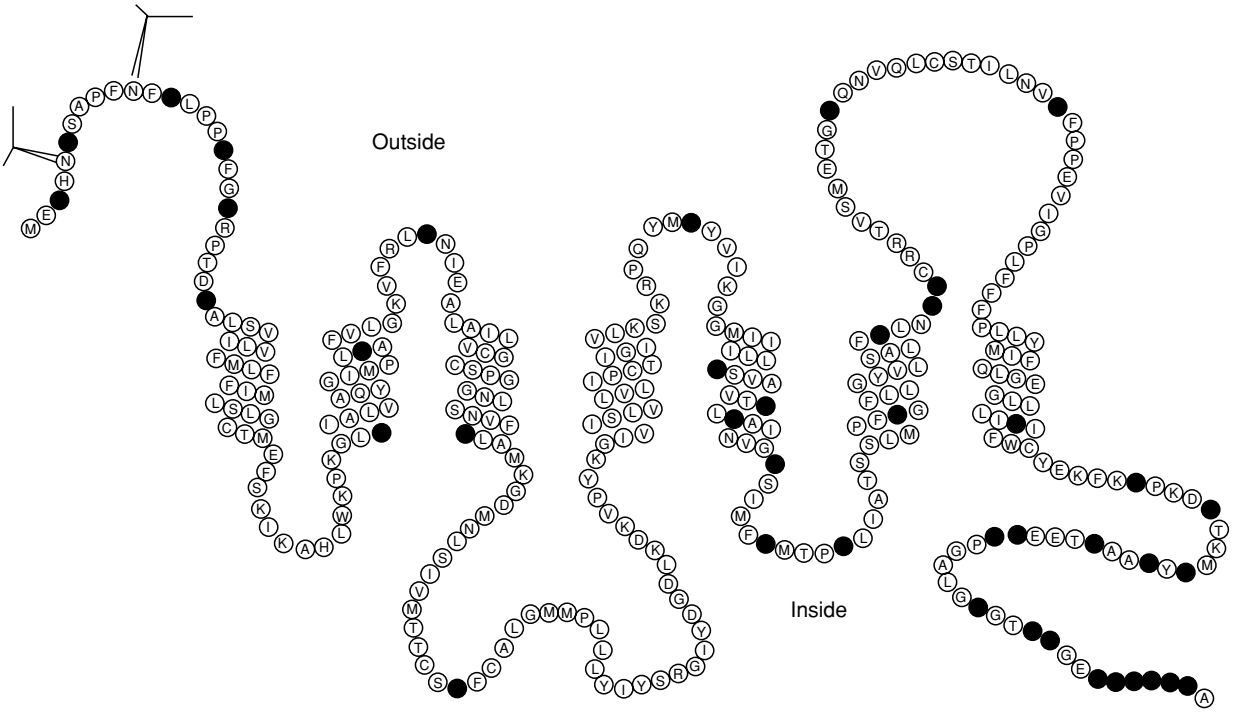

Figure 3 Model of the hepatic Na^+-taurocholate cotransporter. Reproduced from Mejer PJ, Hagenbuch B, and Stieger B (1994) Properties of two-cloned basolateral bile acid uptake proteins of rat and human liver. In: Hofmann AF, Paumgartner G and Stiehl A (eds) *Bile Acids in Gastroenterology, Basic and Clinical Advances*, pp. 139–146. Lancaster, UK: Kluwer Academic Publishers, with permission.

Mechanical Pumps

Gallbladder and intestinal motility provide the main propulsive forces for bile acids to reach the site of intestinal absorption, thus acting as mechanical pumps for the enterohepatic circulation. By contrast with the very rapidly acting chemical pumps, mechanical pumps have storage capacity for the bile acid pool and therefore comprise the slow limb of the enterohepatic circulation (Figure 4).

There is a striking inverse relationship between the recycling frequency of the bile acid pool and bile acid pool size (Figure 5). If one assumes that bile acid pool size is the dependent variable and that the chemical pumps are not saturated under physiological conditions, then gallbladder and small intestinal motor function represent the main factors affecting the kinetics of the enterohepatic circulation. Two main lines of evidence support this concept. First, bile acid pool size has been reported to vary inversely with gallbladder emptying rate in healthy male subjects. Furthermore, the smaller pool size in subjects with a faster gallbladder emptying rate is associated with a greater fractional turnover rate for primary bile acids, and with no change in their synthesis rate. These observations indirectly suggest an effect of gallbladder emptying on the recycling frequency of the bile acid pool.

There is also a negative relationship between small intestinal transit rate and total bile acid pool size. By contrast with gallbladder emptying, it is synthesis rate for both primary bile acids, and not their fractional turnover rate, that is significant associated with small intestinal transit rate. Taken together, these observations suggest that, whereas gallbladder emptying is an important factor in affecting the recycling frequency of the pool, small intestinal transit rate may be more important as a regulating factor for bile acid synthesis, probably by affecting the rate of bile acid return to the liver. This hypothesis is supported by the finding that hepatic microsomal cholesterol 7α-hydroxylase, the rate-limiting enzyme of a major synthetic bile acid pathway, is subject to feedback repression by hydrophobic bile acids returning to the liver via the enterohepatic circulation.

A second line of evidence supporting a role for the mechanical pumps as regulatory factors in bile acid kinetics is provided by studies involving manipulation of gallbladder emptying and small intestinal transit rate. Reduction in postprandial gallbladder emptying by means of dietary measures (sugar diet) results in increased bile acid pool size. This effect is accompanied by a decreased fractional turnover rate for the primary bile acids, with no change in synthesis rate. Similar results are achieved by prolonging small intestinal transit time by chronic administration of anticholinergic drugs. Opposite effects occur with chronic reduction in small intestinal transit time with sorbitol administration, or with regular injections of cholecystokinin octapeptide.

Bile Acid Biotransformation

Bile acid biotransformation occurs at two sites in the enterohepatic circulation – the intestine and the liver. During enterohepatic cycling, bile acids are exposed

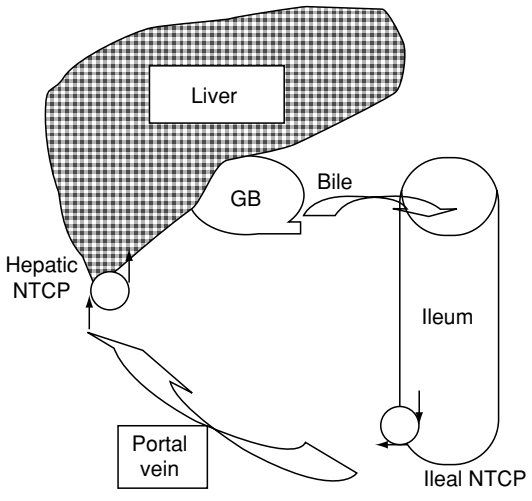

Figure 4 Schematic illustration of the main mechanical (GB, gallbladder, and ileum) and chemical pumps (ileal and hepatic Na^+-taurocholate cotransporter (NTCP)) of the enterohepatic circulation.

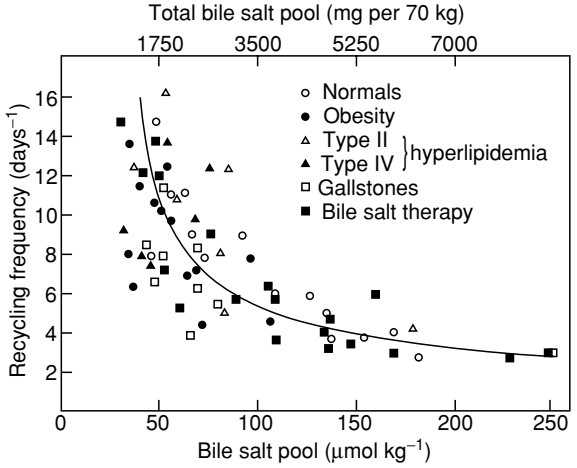

Figure 5 Inverse relationship between bile acid pool size and recycling frequency of the pool in various pathophysiological conditions. Reproduced from Carey MC and Duane WC (1994) Enterohepatic circulation. In: Arias IM, Boyer JL, Fausto N, Jakoby WB, Schachter D and Shafritz DA (eds) *The Liver: Biology and Pathobiology*, 3rd edn, pp. 769–786. New York: Raven Press, with permission.

to an increasing concentration of anaerobic bacteria from the mid-ileum down to the rectum. These bacteria are capable of a variety of metabolic biotransformations of bile acids, and the three most important are deconjugation, 7-dehydroxylation, and 3- (or 7)-oxidation. A small proportion of conjugated bile acids is deconjugated in the mid-ileum and absorbed by passive nonironic diffusion. Bile acids reaching the distal ileum after each meal are absorbed by active transport. Only about 10% of conjugated bile acids reaching the distal ileum after each meal escape absorption and are deconjugated in the colon by cholylglycine hydrolase, an enzyme produced by a great variety of bacterial species. The two primary bile acids, cholic and chenodeoxycholic acid, undergo further biotransformation by bacterial 7-alpha dehydroxylase with formation of deoxycholic and lithocholic acid respectively. By contrast with deconjugation, dehydroxylation of bile acids at C-7 is mediated solely by a relatively small number of anaerobic bacterial species. A total of 20–50% of the secondary bile acids is reabsorbed by the colonic mucosa, returned back to the liver, conjugated with glycine or taurine and secreted in bile as secondary bile acids. The human liver has the capacity further to conjugate glyco- and taurolithocholic acid with a sulfate group, a phenomenon favoring fecal excretion of this potentially toxic bile acid. As a result of efficient fecal excretion lithocholic acid is less than 5% of total bile acid pool in all species examined. By contrast, the proportion of deoxycholic acid varies between 0 and 60% in humans in health and in disease depending on variable input of newly formed deoxycholic acid from the colon.

As a result of bacterial dehydrogenase activity, 3- and 7-oxo bile acids are formed in the intestine. These compounds are absorbed in the colon and reduced stereospecifically in the liver. About 20% of the 7-oxo bile acids are reduced in the liver with formation of the beta-epymers and are called 'tertiary' bile acids. This mechanism accounts for the small proportion of ursodeoxycholic acid and ursocholic acid, the 7-beta-epymer of chenodeoxycholic and of cholic acid respectively, present in human bile.

Physiological Effects

The distribution of the bile acid pool changes during the day, as illustrated by the diurnal variation of bile acid concentration in blood, bile, and intestine. These two latter phenomena have important effects on the physicochemical state of bile and on the digestion and absorption of dietary lipids.

Hepatic bile acid secretion rate declines progressively in humans from 10 to 30 $\mu mol\,kg^{-1}\,h^{-1}$ during the first hour following an evening meal, to less than 5 $\mu mol\,kg^{-1}\,h^{-1}$ before the following breakfast. These changes are accompanied by about 50% net storage in the gallbladder of the hepatic bile during nocturnal fasting, indicating 50% interruption of the enterohepatic circulation. These diurnal changes in bile acid secretion rate are accompanied by changes in the relative proportion of cholesterol and phospholipid in hepatic bile, especially at low bile acid secretion rates. This phenomenon is accounted for by the hyperbolic relationships between bile acid and cholesterol secretion and between bile acid and phospholipid secretion (Figure 6). These relationships explain why the cholesterol/phospholipid ratio is relatively constant in all animal species at high bile acid secretion rate, and why it sharply increases at bile acid secretion lower than about 12 $\mu mol\,kg^{-1}\,min^{-1}$. This phenomenon also explains why hepatic bile is supersaturated with cholesterol at bile acid secretion rates < 20 $\mu mol\,kg^{-1}\,h^{-1}$ in healthy subjects.

The diurnal variation in cholesterol saturation of hepatic bile which accompanies the diurnal fasting–eating pattern is mirrored by a similar but smaller diurnal variation in cholesterol saturation of gallbladder bile. The diurnal variation in bile acid concentration entering the intestinal limb of the enterohepatic circulation also has important physiological effects on the digestion and solubilization of dietary lipids. As a result of dietary stimuli to gallbladder contraction, intraduodenal bile acid concentration increases above the critical value of 4–8 mmol required for normal lipid digestion and absorption. Disease conditions causing detective bile acid secretion secondary to cholestasis or to interrupted enterohepatic circulation (ileal diseases or resection) result in lipid malabsorption and steatorrhea.

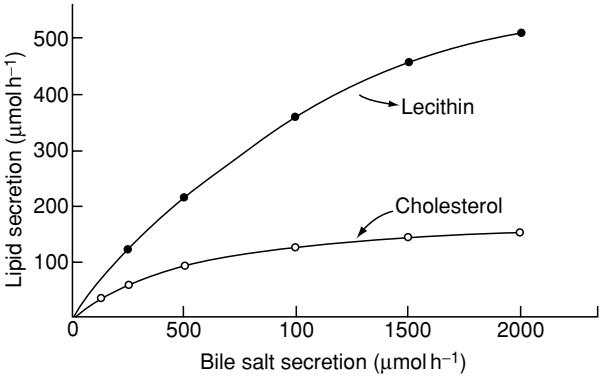

Figure 6 Relationship between bile acid and cholesterol, and between bile acid and phospholipid hepatic secretion rate. Reproduced from Carey MC and Duane WC (1994) Enterohepatic circulation. In: Arias IM, Boyer JL, Fausto N, Jakoby WB, Schachter D and Shafritz DA (eds) *The Liver: Biology and Pathobiology*, 3rd edn, pp. 769–786. New York: Raven Press, with permission.

Other Endogenous Compounds

Apart from bile acids, no other major organic compounds excreted in bile undergo an enterohepatic circulation. Although both biliary cholesterol and phospholipids are absorbed in the small intestine together with their dietary sources, their biliary secretion originates from a preformed hepatic pool. Daily bilirubin secretion equals its daily production, a phenomenon indicating no enterohepatic recirculation for this pigment under physiological conditions. This is because mono- and diglucuronated bilirubin molecules are too large and polar to be absorbed in the small intestine. In the colon, bacterial β-glucuronidase hydrolyzes bilirubin conjugates and unconjugated bilirubin either precipitates as an insoluble calcium salt or is further catabolized to urobilinogen. This latter may be partly absorbed in the colon and undergo a limited enterohepatic circulation.

Studies in rodents suggest that bilirubin absorption may occur under conditions of increased bile acid concentration in the colon resulting from ileal bile acid malabsorption. Under these circumstances excess colonic bile acid concentration may solubilize unconjugated bilirubin, thus favoring passive intestinal absorption and reentry in the enterohepatic circulation of this pigment. This phenomenon may explain the incidence of pigment gallstones in patients with ileal disease and the phenomenon of acquired gallstone calcification in a minority of patients with cholesterol gallstones treated with ursodeoxycholic acid, a bile acid causing malabsorption of endogenous bile acid.

Other endogenous compounds present in very low concentrations in human bile undergo an enterohepatic circulation. This is the case for some vitamins, such as 25-hydroxyvitamin D_3, 1,25-dihydroxyvitamin D_3, vitamin B_{12}, and folic acid. Enterohepatic circulation has also been reported to contribute to the conservation of several hormones, such as corticosterone, hydrocortisone, androsterone, estrone, and thyroid hormones. These endogenous compounds are mainly excreted in bile as their glucuronide or sulfate conjugates, and deconjugation by colonic bacteria is essential for them to be passively reabsorbed by the colonic mucosa.

Exogenous Compounds

Many environmental pollutors, toxins, and drugs undergo an enterohepatic circulation. For the majority of these substances the evidence supporting an enterohepatic circulation is based on animal studies, and the large interspecies variation in biliary excretion of xenobiotics indicates that extrapolation of results to humans must be taken with caution.

Among environmental pollutants, heavy metals such as mercury, arsenic, manganese, and copper undergo an enterohepatic circulation. Methylmercury poisoning has been reported to occur in Iraqi and Japanese populations exposed to this pollutant. Oral administration of a polythiol-binding resin interrupts the enterohepatic circulation of methylmercury, and has been reported to benefit exposed populations. An enterohepatic circulation has been reported for many insecticides and for polycyclic aromatic hydrocarbons. Mutagenic metabolites of benzo-[*a*]-pyrene, an environmental carcinogen, are excreted in rat and rabbit bile and undergo an enterohepatic circulation, 1-nitropyrene, a combustion product of diesel engines, is also a potent mutagenic substance which is mainly excreted in bile and reabsorbed in the gut in rat.

Certain fungal toxins exert their effect by accumulating in the body via the enterohepatic circulation. This is the case for ochratoxin, a mycotoxin that may contaminate food, and for α-amanitin, a toxin produced by mushrooms of the *Amanita phalloides* species. *Amanita* poisoning in humans is due to enterohepatic recirculation of α-amanitin for up to 48 following ingestion of *Amanita* mushrooms. This toxin is harmless in some herbivorous animals lacking an enterohepatic circulation of α-amanitin.

Many drugs undergo enterohepatic circulation, including the following examples:

- Demeclocycline
- Rifampicin
- Digoxin
- Digitoxin
- Imipramine
- Sulindac
- Eterophine hydrocloride
- Estradiol
- Mycophenolate mofetil
- Verapamil
- Colchicine
- Oxazepam
- Phenytoin
- Spironolactone
- Indometacin
- Diclofenac
- Valproic acid
- Morphine
- Warfarin
- Lormetazepam
- Adriamycin

For some of these drugs, conservation via the enterohepatic circulation is beneficial because it prolongs

their therapeutic efficacy. Thus, for example, the enterohepatic circulation of morphine and of other opiates may prolong their analgesic effect. On the other hand, accumulation of the anticoagulant warfarin, or of the tranquillizers benzodiazepines, or of digitoxin may expose patients to the risk of drug overdose. For most of these drugs, administration of activated charcoal or ion-exchange resins (cholestyramine, for example) is of proven value in the treatment of overdose by interrupting their enterohepatic circulation, thus favoring fecal excretion.

Colonic Bacteria and the Enterohepatic Circulation

Both endogenous and exogenous substances are excreted in bile in conjugated forms. These are mainly glucuronate and sulfate conjugates, with the noticeable exception of bile acids which are conjugated almost exclusively with glycine and taurine in humans. Conjugation, on the one hand, makes these substances sufficiently hydrophilic to be excreted in bile and, on the other hand, makes them too hydrophilic to be absorbed by passive diffusion in the gut. Bile acids are unique in being actively absorbed in the distal ileum in the conjugated form. For all other substances, deconjugation is a necessary step to increase their lipophilicity prior to passive colonic diffusion The human colonic flora is capable of producing glucuronidases and sulfatases, and passive diffusion of the resulting hydrolyzed compounds is usually quantitatively small, because of the relatively small surface-to-volume ratio in the colon, and because binding to bacteria and to nonabsorbable dietary components favors fecal excretion. This explains why the enterohepatic circulation is quantitatively small for all substances relying on a colonic deconjugation step to enter the enterohepatic circulation following biliary excretion.

See also: **Bile**; **Gallbladder**; **Liver**: Structure and Function; **Mycotoxins**: Classifications; Toxicology

Further Reading

Angelin B, Eriksson M and Rudling M (1999) Bile acids and lipoprotein metabolism: a renaissance for bile acids in the post-statin era? *Current Opinion in Lipidology* 10: 269–274.

Bahar RJ and Stolz A (1999) Bile acid transport. *Gastroenterology Clinics of North America* 28: 27–58.

Carey MC and Duane WC (1994) Enterohepatic circulation. In: Arias IM, Boyer JL, Fausto N, Jakoby WB, Schachter D and Shafritz DA (eds) *The Liver: Biology and Pathobiology*, 3rd edn, pp. 769–786. New York: Raven Press.

Cooper AD (1999) Role of the enterohepatic circulation of bile salts in lipoprotein metabolism. *Gastroenterology Clinics of North America* 28: 211–229.

Hofmann AF (1989) Enterohepatic circulation of bile acids. In: *Handbook of Physiology*, section 6, vol. 111, pp. 567–596. American Physiological Society.

Hofmann AF (1992) Bile acids in liver and biliary disease. In: Millward-Sadler GH, Wright R and Arthur MJP (eds) *Wright's Liver and Biliary Disease*, 3rd edn, pp. 289–316. London, UK: WB Saunders.

Hylemon PB and Harder J (1998) Biotransformation of monoterpenes, bile acids, and other isoprenoids in anaerobic ecosystems. *FEMS Microbiology Reviews* 22: 475–488.

Lanzini A and Northfield TC (1991) Biliary lipid secretion in man. *European Journal of Clinical Investigation* 21: 259–272.

Love MW and Dawson PA (1998) New insights into bile acid transport. *Current Opinion in Lipidology* 9: 225–229.

Mejer PJ, Hagenbuch B and Stieger B (1994) Properties of two cloned basolateral bile acid uptake proteins of rat and human liver. In: Hofmann AF, Paumgartner G and Stiehl A (eds) *Bile Acids in Gastroenterology. Basic and Clinical Advances*, pp. 139–146. Lancaster, UK: Kluwer Academic.

Nutritional Management of Liver and Biliary Disorders

E Medici, King's College Hospital, London, UK

Background

Nutrition intervention in liver disease is both challenging and well researched. Intervention will depend on the individual patient needs and is related to symptoms rather than disease state. Hence, patients with liver cirrhosis will suffer much the same range of problems irrespective of the etiology. Preserving the patient's nutritional status is central to the treatment of all liver patients. Liver transplantation is now well established as the treatment for end-stage liver disease and restores the patient's nutritional status; however, patients often develop different nutritional complications, predominantly excess weight gain and hyperlipidemias. This section will aim to give a brief understanding of the current evidence-based nutrition practice in liver disease and transplantation.

Type of Liver Disease

Acute or fulminant hepatic failure (FHF) refers to massive hepatocellular necrosis resulting in hepatic

encephalopathy and coma within 8 weeks from the onset of symptoms and in the absence of previous liver disease. Most common forms of FHF in the UK are due to exogenous hepatotoxins such as paracetamol (acetaminophen). FHF carries a mortality rate of 50–80%. Nutrition intervention is short-term until the patient recovers, is transplanted, or dies. Diet manipulation is rarely required, and standard ICU feeding protocols will apply.

Chronic liver failure refers to the progressive damage and cirrhosis of the hepatocytes over a prolonged period. Complications are stage-dependent and often reflect degree of damage to the liver. A patient with cirrhosis can be asymptomatic for up to 40 years. Normally, the patient will alternate between an asymptomatic or controlled symptoms state (compensated liver disease) and a symptomatic state (decompensated liver disease). A decompensated state may result in any one or a combination of symptoms such as ascites, variceal bleeding, diabetes, steatorrhea, and/or encephalopathy. All of these symptoms will require dietary manipulation.

Protein Energy Malnutrition

As the liver is central in the metabolism, assimilation, and synthesis of protein, it is not surprising that protein-energy malnutrition (PEM) is common in patients with chronic liver disease. PEM refers to the increased catabolism of muscle and other visceral proteins for gluconeogenesis. PEM in chronic liver disease has been well documented with a prevalence of 27–100%. The severity of PEM is dependent on etiology and length of disease, with alcoholic liver disease patients being worst effected.

PEM in chronic liver disease has been demonstrated as an independent risk factor to clinical outcome and, therefore, is always prioritized in the treatment of cirrhotic patients. Protein energy malnutrition has been associated with:

- worsening encephalopathy;
- poor control of ascites;
- increased infection rate;
- increased hospital admissions;
- increased mortality.

The reasons for this gross PEM are multifactorial; however, the imbalance of catabolic and anabolic hormones plays a pivotal role. Factors associated with PEM in chronic liver disease include:

- increased energy and protein requirements;
- increased levels of catabolic hormones, predominantly catecholamines, but also corticosteriods and glucagon;
- resistance to the anabolic hormones, insulin and growth hormone;
- reduction of the anabolic hormones, growth factor 1 and its carrier proteins;
- alteration in nutrient utilization;
- reduced glycogenolysis;
- increased gluconeogenesis;
- increased lipolysis

PEM is exacerbated in cirrhotic patients by accompanying anorexia and malabsorption. Aggressive nutrition support is often needed to meet increased energy and protein requirements in order to attenuate catabolism and PEM.

Dietary intervention commences with appropriate assessment of the patient's muscle status and not simply body weight. The latter may be affected by ascites and/or edema, and does not differentiate between fat and muscle stores. Normal weight or obesity may mask PEM, as normal or excessive fat stores can be present in such patients. The patient's requirements will need to be calculated and met, orally, nasoenterally or intravenously.

Elevated Energy and Protein Requirements

Increased requirements for energy and protein have been demonstrated in compensated as well as decompensated liver disease, acute liver failure, and post-liver transplantation by a number of authors. Basal metabolic rate (BMR) has been found to be 40% higher than in healthy individuals, and protein requirements have been estimated at 1–1.8 $g\,kg^{-1}\,day^{-1}$ (healthy individual requirements 0.8 $g\,kg^{-1}\,day^{-1}$). Energy and protein requirements have been found to be closely associated with lean body mass; hence, standard calculations should use dry body weight, and adjustments have to be made for excess fat stores, fluid overload, and ascites (see Table 1). It has been suggested, that patients with a dry body mass index (BMI) of $\geq 30\,kg\,m^{-2}$ and $\geq 40\,kg\,m^{-2}$

Table 1 Estimating fluid weight

Estimated ascitic weight (kg)	
Mild	2.2
Moderate	6.0
Severe	14.0
Estimated peripheral edema weight (kg)	
Mild	1.0
Moderate	5.0
Severe	10.0

From Wicks C and Madden A (1994) *A Practical Guide to Nutrition in Liver Disease*, 2nd edn. Birmingham, UK: The Liver Interest Group of the British Dietetic Association.

(see Table 2) should have their requirements calculated at 75 and 65%, respectively, of their actual body weight.

Some authors strongly recommend that resting energy expenditure should be measured and not predicted. However, as indirect calorimetry carts are not readily available and may be impractical, the use of predictive equations and stress factors is inevitable. Two validated equations for calculating requirements are used in the UK and are shown with examples in Tables 3–5.

Table 2 Calculating body mass index

$$\text{BMI kg m}^{-2} = \frac{\text{Weight in kilograms}}{\text{Height in meters}^2}$$

Interpretation of BMI values (kg m^{-2}):
≤ 15 = critically underweight – Increase weight
15.9–19.9 = underweight – increase weight
20–24.9 = ideal weight – grade 0 obesity – maintain weight
25–29.9 = overweight – grade I obesity – maintain or lose weight
30–39.9 = obese – grade II obesity – must reduce weight by at least 10%
≥ 40 = grossly obese – grade III obesity – must lose weight by at least 10%

Table 3 ESPEN guidelines recommendations

	Protein g kg^{-1} day^{-1}	*Energy kcal kg^{-1} day^{-1} (kJ kg^{-1} day^{-1})*
Compensated liver disease	1.0–1.2	25–30 (105–125)
Decompensated liver disease	1.0–1.5	25 (if ventilated) – 40 (105–167)

Diabetes

The incidence of diabetes is higher in patients with liver disease resulting from hepatitis C virus and alcohol. The prevalence has been quoted at 27–39%. It is essential to maintain good glycemic control (i.e., 4–8 mmol l^{-1}). Hyperglycemia exacerbates muscle catabolism and can increase infection risk. A standard diabetic diet should not be prescribed, as it encourages a low intake of protein and total energy and will compromise the patient's nutritional status. A modified diet encouraging the avoidance of simple carbohydrate whilst increasing fat and protein dense foods should be followed. This will allow for normal glycemic control whilst still meeting energy and protein requirements. Hypoglycemic agents (oral or insulin) will often be required in conjunction with diet.

If the patient is anorectic and not meeting requirements, oral sip feeds and/or nasogastric feeding will be indicated. As these feeds will contain maltodextrins, hypoglycemic agent therapy will need alteration to compensate for the expected elevation in blood sugar.

Ascites

This is the excessive accumulation of extracellular fluid in the peritoneal cavity, usually as a late complication of cirrhosis. The development of ascites is a complex multifactorial process of which the precise mechanism is still unknown. It is probably due to sodium retention (enhanced by secondary hyperaldosteronism), compounded by hypoalbuminemia and portal hypertension. A diagrammatic representation

Table 4 Estimating energy requirements using basal metabolic rate with activity and stress factors (estimated daily requirements = BMR × (% stress + % activity) + BMR)

	BMR kcal (kJ)	
Age	*Men*	*Women*
15–18 years	17.6W + 656 (74W + 2742)	13.3W + 690 (56W + 2884)
18–30 years	15.0W + 690 (63W + 2884)	14.8W + 485 (62W + 2027)
30–60 years	11.4W + 870 (48W + 3637)	8.1W + 842 (34W + 3520)
> 60 years	11.7W + 585 (49W + 2445)	9.0W + 656 (38W + 2742)
Stress factors		
Compensated liver disease	0–20%	
Acute (fulminant hepatic failure)/ventilated/transplant	20–30%	
Decompensated liver disease	30–40%	
Activity factor		
Bed-bound/immobile	10%	
Bed-bound mobile/sitting	15–20%	
Mobile on ward	25%	
Protein requirements		
Compensated liver disease	0.8–1.2 g kg^{-1} day^{-1}	
Decompensated liver disease/transplant	1.2–1.8 g kg^{-1} day^{-1}	

W, dry lean body weight.

Table 5 Example for calculating requirements using both methods

Subject: 36-year-old female with moderate ascites and hepatic encephalopathy secondary to chronic liver disease; the patient is bed-bound
Height = 1.6 m
Weight = 86 kg
(1) Estimate dry body weight: 86 kg − 6 kg = 80 kg
(2) Calculate BMI: $80/(1.6)^2 = 31\ kg\ m^{-2}$
(3) Estimate lean body weight: 75% of 80 kg = 60 kg (the weight to be used for requirement calculations is 60 kg)
Estimated daily requirements using the ESPEN guidelines:
1800–2400 kcal (7524–10032 kJ) per day
60–90 g of protein per day
Estimating daily requirements using BMR and stress factors
BMR = (8.1 × 60) + 842 = 1328 kcal (5551 kJ)
Daily energy requirements = 1328 × (10% activity + 30% stress) + BMR = 1860 kcal (7771 kJ)
Daily protein requirements = 70–108 g

is shown in Figure 1. Ascites usually develops when the patient has advanced 'decompensated' disease. Ascites causes discomfort and shortness of breath, reduced appetite, increased risk of infection (spontaneous bacterial peritonitis), and increased mortality. In addition, ascitic patients will also display even higher energy and protein requirements than nonascitic liver patients.

Ascites is normally managed by a combination of dietary sodium restriction and diuretic therapy. The aim should be to achieve a net fluid loss of 500 ml day^{-1} (equivalent to a weight loss of 0.5 kg), until the ascites clears. Serum urea, creatinine, and electrolyte concentration should be checked at least once a week at the start of treatment. A rising serum urea or creatinine and falling serum sodium concentration indicate impending renal failure and should lead to

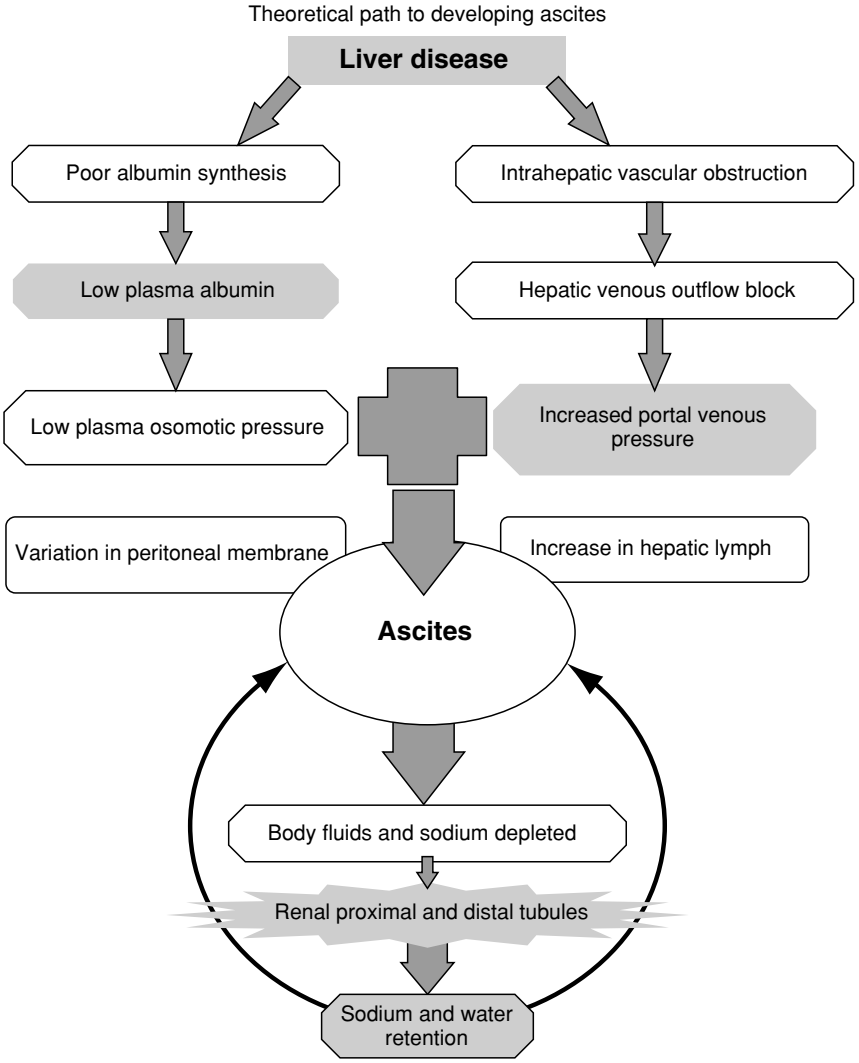

Figure 1 Theoretical pathway to the development of ascites.

a reduction in diuretic treatment. A combination of frusemide and spironolactone diuretic therapy is often used. The latter is potassium-sparing and may result in hyperkalemia after prolonged use. Other interventions include paracentesis (removal of large volumes of ascitic fluid via a cannula).

If a low-sodium/no-added-salt diet and diuretic therapy do not work, the ascites is classified as refractory/resistant, and transplantation or surgical procedures will be required to alleviate the symptoms.

Sodium Restriction

These patients will be extremely catabolic, and display anorexia. Meeting their nutritional requirements is paramount and also extremely difficult without supplementary feeding. Reducing salt intake will automatically result in reduced palatability and possibly further reduce the patient's intake. During hospitalization, it is much easier to impose a sodium restriction on patients whilst ensuring that energy and protein requirements are met. Once discharged, this level of restriction is very difficult to maintain, and a 'realistic' diet of no added salt must be prescribed. An example of a no-added-salt diet is shown in Table 6. This will depend on the patient's socio-economic background, motivation, mental acuity and prognosis. Cutting down salt intake will also help with adherence to the fluid restriction, which is often imposed at 750–1500 ml day^{-1}.

Oral supplements and nasogastric feeds should be chosen carefully. Feeds low in sodium and volume (i.e., 2 kcal ml^{-1}) will be ideal.

Confusion often arises over serum sodium levels. Irrespective of the serum levels (patients are normally hyponatremic), no-added-salt diets should always be prescribed/encouraged for all patients with ascites.

Varices

These are veins that have become grossly enlarged as a consequence of portal hypertension. They are caused by the development of venous back flow as a result of portal vein obstruction by the cirrhotic liver. They are more often seen in the lower 5 cm of the esophagus and fundus of the stomach. The most significant complication of varices is hemorrhage, which is fatal for 50% of patients.

Table 6 Low-sodium alternatives

Low-sodium energy and protein-dense foods	*High-sodium foods – to be avoided*
Meat, fish, poultry	
Fresh or frozen meat, fish, poultry (roast, fry, cook with unsalted butter or oil), tinned fish in oil or water (not brine)	Sausages, tinned meat, tinned fish in brine, corned beef, bacon, ham, gammon, burgers, pies, ready meals, smoked/pickled meat/fish, fish in batter, pâté, sandwich spreads, salami, etc.
Dairy products	
Full-cream milk, yogurt, fromage frais, all creams, cream cheese, mozzarella, cottage cheese, ricotta cheese, eggs	Lassi, most cheeses not mentioned in the 'USE' list, all processed cheeses, e.g., cheese spreads/slices
Fats and oils	
All oils, unsalted butter	Dressings, bottled sauces, mayonnaise
Cereals and breads	
Shredded Wheat, Raisin Splitz, Sugar Puffs, puffed wheat, porridge, unsalted muesli, all flours, bread, rolls, pasta	Most other breakfast cereals, rich breads, e.g., brioche, focaccia, ciabatta, croissant, pot noodles
Fruit and vegetables	
Fresh, frozen, tinned or dried fruit, fresh or frozen vegetables, tinned tomatoes, tomato purée, tinned vegetables in water only, all fruit juices	Dried, pickled, and most tinned vegetables, olives, gherkins, etc.
Savory snacks	
Unsalted nuts, peanuts and raisins, Smiths 'Shake + Salt' crisps (throw away the salt sachet), all butter popcorn	Salted nuts, crisps, Twiglets, salted popcorn and other salted snacks
Biscuits	
Most biscuits – especially Jaffa Cakes, wafers, Kit-kat, Malted milk, Bourbon creams, Nice, plain Hobnobs, Matzo crackers	Digestives, shortbread cheese biscuits, Ryvita
Sweet snacks/condiments	
All milky puddings, custard, cakes, gateaux, chocolate, especially Smarties, Walnut whips, plain chocolate, Fry's chocolate cream, After Eights, fruit sweets – boiled sweets, jelly babies, pastilles, Opal Fruits, marmalade, jam, honey, chocolate spread	Fudge, toffees, Marmite, peanut butter, Vegemite
Drinks and miscellaneous	
Whole milk, hot chocolate, milk shake, yogurt shakes, Lucozade, cordials	Soup, stock cubes, Bovril, Marmite, Build Up Soup

Diet

The presence of varices is not necessarily a contraindication to nasogatric feeding. Two randomized controlled trials investigating the benefit of aggressive nutrition support via nasogastric feeding to full requirements compared with no feeding found no increase in variceal bleeds between the study and control groups. Nasogastric feeding is often not appropriate when a patient is actively bleeding or within 48 h of variceal treatment (banding/sclerotherapy).

No dietary modifications are required for these patients, except to reassure and encourage a normal consistency diet. Many patients will self-regulate their diet to soft/sloppy consistency, because of a misconception that solid foods will rupture their varices. This would only be encouraged if the patient finds swallowing painful, because of variceal treatment such as sclerotherapy or banding.

Hepatic Encephalopathy

Hepatic Encephalopathy (HE) in patients with chronic liver disease (CLD) is a common complication, with a prevalence of 50–70%. HE as a result of FHF is accompanied by intracerebral edema, and intervention relies on the reduction of the cerebral pressure and/or liver transplantation.

HE is a neuropsychiatric condition as a result of portal vein shunting away from the cirrhotic liver. Toxins present in the portal vein shunted away from the liver and hepatocytes without detoxification are thought to be able to cross the blood–brain barrier, with resultant alteration in the patient's consciousness, personality, intellectual capacity, and speech. HE in chronic liver disease can be reversed with medical and nutritional intervention.

The pathogenesis for HE remains unknown but is thought to be multifactorial. Several theories have been proposed and researched. The more popular theories to HE are those relating to the elevated ammonia levels and disturbance in the branched chain amino acid (BCAA) to aromatic amino acid (AAA) ratio in the blood.

The onset of HE in patients with CLD is often precipitated by a number of factors, which, once treated, alleviate the encephalopathy.

Such factors include:

- constipation;
- sepsis;
- gastrointestinal bleed;
- alcohol binge;
- dehydration;
- enthusiastic use of diuretics;
- paracentesis;
- electrolyte imbalance;
- hypokalemia (increased renal NH_4 production);
- protein-rich diet.

Ammonia Theory

Elevated circulating ammonia and ammoniagenic substances are probably the result of portal vein shunting and the inability of ammonia from the gastrointestinal tract to reach the hepatocytes for urea production. Once ammonia diffuses across the blood–brain barrier, neurotoxicity results, as the brain does not have a urea cycle. Toxicity on the central nervous system may be direct or indirect. The latter has been suggested, as ammonia serum levels correlate very poorly with the presence or level of encephalopathy. It is now well established that the action on luminal content by colonic gut flora and glutaminase in the ileum are major contributors of ammonia in the portal vein. The small and large intestine are therefore targeted for treatment.

This theory encouraged the use of low-protein diets to try and reduce ammoniagenic compounds in the portal vein. However, this practice has been proven to be ineffective and, in many cases, detrimental to the patient's well-being.

False Neurotransmission Theory

The false neurotransmission theory in the 1970s was based on the altered ratio of BCAA to AAA in the serum seen in most cirrhotic patients. In healthy individuals, BCAA (leucine, isoleucine, and valine) serum levels are three to four times greater than AAA (free tryptophan, phenylalanine, and tysosine). The higher BCAA concentration results in a greater affinity for the receptor sites of the single carrier system and transport across the blood–brain barrier for normal neurotransmission. In cirrhotic patients and especially encephalopathic patients, this level is altered to 1:1.

AAAs are predominantly utilized by the liver, whereas BCAAs are predominantly utilized by peripheral muscle. During liver disease, BCAAs continue to be metabolized by the muscle for energy, but AAA metabolism by the cirrhotic liver is heavily reduced and results in the reduced ratio.

With this altered amino acid plasma ratio in the blood, AAA can compete with more affinity to diffuse across the blood–brain barrier. Elevated serum ammonia indirectly favors the uptake of AAA into the brain (see Figure 2).

The use of altered amino acid feeds with encephalopathic patients was therefore encouraged. It was hypothesized that altering the amino acid composition of these patients would lead to an improved

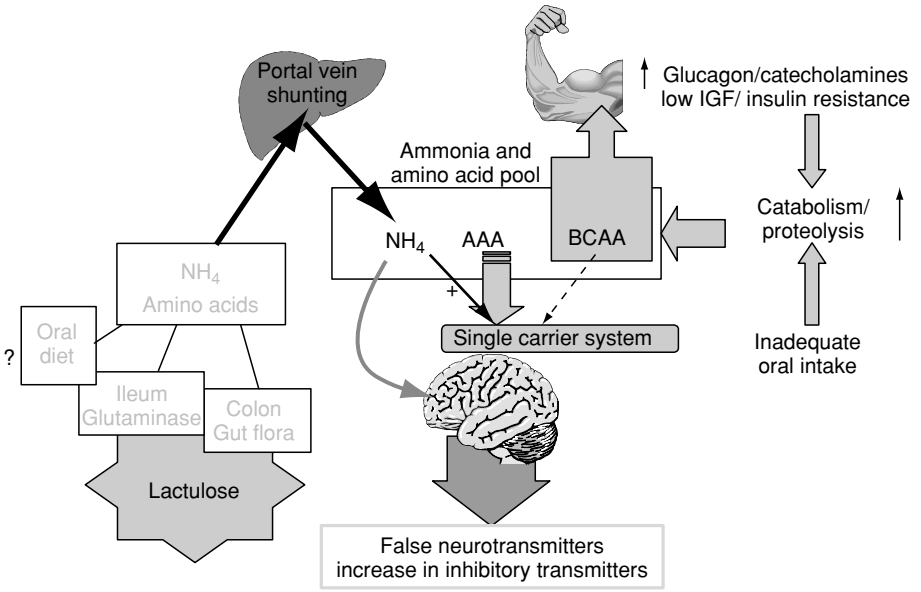

Figure 2 Route of ammonia and amino acids during portal shunting. Ammonia and amino acids from gastrointestinal metabolism enter the portal vein. As the portal vein is shunted away from the cirrhotic liver, ammonia and amino acids are in the systemic circulation. BCAAs are rapidly taken up and utilized by peripheral muscle for energy and serum levels become reduced. AAAs cannot be cleared from the circulation, as they require the urea cycle in the liver. In addition, gluconeogenesis and muscle catabolism increase the release of ammonia and AAA in the circulation. Ammonia may diffuse across the blood–brain barrier, resulting in neurotoxicity, and also increases the affinity of the single carrier system for AAA to cross into the brain and result in the production of false neurotransmitters.

BCAA:AAA serum ratio and improved encephalopathy. Unfortunately, the trials have not been conclusive on the benefit of these expensive and less palatable supplements. Improvement in HE has not always been accompanied by an improvement in the amino acid ratio.

In the 1980s and 1990s, many metabolic and prospective randomized trials investigating the effect of dietary protein intake and use of BCAA feeds in encephalopathy have concurred that:

- Treatment and elimination of precipitating factors often lead to the improvement and elimination of HE without manipulation of the diet.
- Spreading the protein loads throughout the day results in more effective ammonia clearance.
- Endogenous ammonia production from muscle catabolism often exceeds oral protein intake. This is due to the imbalance of anabolic and catabolic hormones and push toward gluconeogenesis in cirrhotic patients. This picture is often exacerbated by the patient's inability to consume adequate nutrition (owing to confusion and/or drowsiness).
- Protein requirements are elevated at 1.2–1.8 g kg^{-1} day^{-1}.
- Positive nitrogen balance is directly linked to improvement and/or elimination of encephalopathy.
- Negative nitrogen balance has been directly linked to the presence and worsening of HE.
- Positive nitrogen balance is achieved by meeting protein requirements. Supplementary feeding with oral sip feeds, nasoenteral, or intravenous nutrition is often required. Similar outcomes have been achieved from standard feeds or BCAA feeds.
- Peripheral muscle is responsible for clearing 50% of the circulating ammonia. Therefore, the more muscle-wasted a patient is, the greater the circulating ammonia.
- Peripheral muscle utilizes BCAA for energy and preservation of muscle mass, and so BCAA feeds have improved nitrogen balance quicker than standard feeds.
- BCAA (predominantly leucine) in cirrhotic patients have been found to:
  - reduce the flux of AAA from muscle cells;
  - reduce proteolysis;
  - increase protein synthesis;
  - increase uptake of AAA by the liver;
  - increase uptake and metabolism of ammonia in muscle cells.

Current Practice for Hepatic Encephalopathy

Diet

Any precipitating factors must be dealt with before dietary effectiveness can be of benefit.

Meeting energy and protein requirements, which often requires supplementation (oral and/or nasoenteral), will normally resolve HE. The use of BCAA feeds in HE is not established in the UK but may benefit patients with particularly persistent encephalopathy. Low-protein diets are contraindicated and should not be used. However, some patients with persistent HE, good muscle mass who are exceeding their protein requirements, may benefit from reducing dietary protein intake to recommendations and possibly using BCAA supplements.

Lactulose/lactitol

The effectiveness of lactulose/lactitol is undisputed and is one of the first-line treatments. This is not surprising, as constipation is cited by many authors as a common precipitating factor for HE. As mentioned earlier, colonic flora and glutaminase in the ileum are responsible for the significant production of free ammonia carried in the portal vein.

It is desirable for a patient with a history of hepatic encephalopathy to open their bowels two to four times per day and the stools to be soft and acidic (pH < 6). Lactulose/lactitol remove both dietary and endogenous ammonia from the gut lumen by osmotic cathartic action.

Antibiotics – Neomycin and Metronidazole

Antibiotics work similarly to lactulose but can only be used for a short time, owing to their ototoxicity and nephrotoxicity.

Lactobacillus acidophilus

Lactobacillus acidophilus populates the colonic lumen with nonurease-producing bacteria, and the research has been extremely controversial.

Ornithine Aspartate Treatment

Ornithine is essential for converting NH_4 to urea, whereas aspartate is essential for the conversion of NH_4 to glutamine in the urea cycle. Ornithine aspartate via enteral or parenteral infusion has been found to be effective in resolving mild encephalopathy in small trials.

Zinc Supplementation

It has been well documented that patients with chronic liver disease have low serum zinc levels as a result of increased urinary losses. In addition, two out of the five urea cycle enzymes are zinc-dependent. Trials investigating zinc supplementation to increase the activity of the urea cycle have been inconclusive.

Steatorrhea

Steatorrhea is a common complication as a result of a mechanical block in the biliary tract or by intrahepatic lesions. When bile salt entry into the duodenum is hindered, the emulsification of dietary fats is incomplete. Steatorrhea is more frequently seen in patients with cholestatic disease, e.g., primary biliary cirrhosis (PBC) and primary sclerosing cholangitis (PSC). The reduced absorption of fat results in nausea, indigestion, and steatorrhea, with accompanying calcium and fat-soluble vitamin malabsorption. The patient's nutrition status will obviously be compromised if steatorrhea is profuse. However, steatorrhea varies from patient to patient.

The prescription of low-fat diets should only be prescribed if the steatorrhea is socially unacceptable for the patient or is causing muscle loss, or the patient reports nausea and/or indigestion when consuming fats. Blanket fat restrictions will compromise a patient's intake of energy and make nutrition support difficult. Therefore, fat restrictions should never be applied to patients not experiencing any of the above symptoms.

Malabsorption may also be caused or exacerbated by concomitant inflammatory bowel disease. There is a high prevalence of ulcerative colitis in patients suffering from PSC, whereas celiac disease has been associated with PBC patients. It is therefore important for bowel motions to be assessed carefully.

If fat restrictions are indicated, energy intake will need to be manipulated and monitored. The level and type of fat restriction required are extremely variable from patient to patient. Increased use of carbohydrates and carbohydrate-only supplements is often beneficial. There is also a vast range of fat-free oral supplements. If energy requirements cannot be met with increased carbohydrate alone, then medium-chain triglyceride (MCT) emulsions and oils can be utilized. MCTs are water-soluble and do not require emulsification by bile salts. The emulsion can be taken on its own at a maximum dose of 45 ml qds or added to drinks such as milkshakes and low-fat milk. The oil can be used in various recipes but has a low smoking point, often making it unsuitable for frying. If supplementary feeding is required, oral and enteral feeds are available, containing MCT as its predominant fat source.

Plasma levels of fat-soluble vitamins A, D, E, and K should be routinely checked and supplementation given accordingly. Blanket supplementation should not be encouraged, as excessive intake is toxic.

Prevalence of osteoporosis is much higher in patients suffering from PSC and PBC. It is therefore

advisable to use a calcium supplement with active vitamin D.

Patients suffering from celiac disease need further adaptation to their diet to exclude gluten.

Liver Transplantation

Immediately Post-transplant

Aggressive nutrition support immediately post-transplant is essential to ensure quick and rapid recovery. Initially, nasogastric feeding may not be tolerated, as gastric stasis is common. In such instances, nasojejunal feeding should be used. These patients' requirements have been estimated at 140% BMR and protein up to $2\,g\,kg^{-1}\,day^{-1}$. Aggressive nutrition support has been shown to reduce the patients' ICU stay and time on the ventilator.

Medication interaction Patients are placed on immunosuppressive drugs (Tacrolimus or Neoral), steroids, and antibiotics. Immunosuppression drugs are gradually reduced, and a minimal dose must be taken daily for the rest of the patient's life. The steroids are used to reduce inflammation and are stopped within a few months. The antibiotic therapy is prophylactic and lasts for a few weeks only.

Diet Once the patient is on the ward, elevated protein and energy requirements should continue to be met. The patient's appetite (usually) returns to normal as a result of the steroids, the cessation of restrictive diets, and a general feeling of well-being. Food-safety advice is necessary for up to the first four months, because of the initial high dose of immunosuppression medication. The patient should also be advised to avoid grapefruit and grapefruit juice, as these have been shown to interfere with the absorption of Tacrolimus and neoral from the gut.

Glucose intolerance and diabetes are also common posttransplant. They may be short-lived until the patient stops prednisolone therapy. However, the incidence of diabetes post liver transplantation is high and thought to be related to immunosuppression therapy. Hyperkalemia is also a common side-effect seen with immunosuppression therapy but normally resolves with a reduction in the dose. Low-dietary-potassium advice may be required.

Long-term Post-transplant

Unfortunately, up to 65% of transplant patients, at one year post-transplant, become obese with hypercholesterolemia and an increased risk of coronary heart disease. This is more common in patients transplanted for chronic liver disease. Healthy eating advice is therefore instigated once a normal BMI of $20–25\,kgm^{-2}$ is achieved. The exact reason for obesity is not known. However, it has been suggested that the patient's inactive lifestyle, increased appetite, enjoyment of foods, and immunosuppression therapy all play a role. Premorbid obesity has been found to correlate with postransplant obesity. Therefore, such patients should be encouraged to follow a healthy eating and exercise plan as soon as possible post transplantation.

Diabetes Diabetes in post liver transplant patients is common, despite cessation of steroids from their therapy. Once a patient achieves a good appetite and is eating regularly, standard healthy eating diabetic advice should be given.

See also: **Amino Acids**: Properties and Occurrence; **Diabetes Mellitus**: Etiology; **Energy**: Intake and Energy Requirements; **Lactic Acid Bacteria**; **Malnutrition**: The Problem of Malnutrition; Malnutrition in Developed Countries; **Metabolic Rate**; **Protein**: Requirements

Further Reading

Briony T (1996) *Manual of Dietetic Practice. Liver and Biliary Disease*, 2nd edn. pp. 396–407. Oxford: Blackwell Science.

Fischer and Baldessarini (1971) *False neurotransmitters and hepatic failure. The Lancet* 2: 75–80.

Hasse J (1991) *Role of the dietitian in the nutrition management of adults after liver transplantation. Journal of the American Dietetic Association* 91: 473–476.

Hussaini SH *et al.* (1998) *Effects of orthotopic liver transplantation on body composition. Liver* 18: 173–179.

Kondrup J and Müller MJ (1997) *Energy and protein requirements of patients with chronic liver disease – a review. Journal of Hepatology* 27: 239–247.

Mizock B (1999) *Nutritional support in hepatic encephalopathy. Nutrition* 15: 220–228.

Morgan MY and Heaton KW (2000) Human nutrition and dietetics. In: Garrow JS, James WPT and Ralph A (eds) *Nutrition, the Liver and Gallstones*, 10th edn. pp. 575–599. Edinburgh: Churchill Livingstone.

O'Keefe SJ *et al.* (1981) Protein turnover in acute and chronic liver disease. *Acta Chirurgica Scandinavica Supplementum* 507: 91–101.

Plauth M, Merli M, Kondrup J *et al.* (1997) Consensus statement: ESPEN guidelines for nutrition in liver disease and transplantation. *Clinical Nutrition* 16: 43–55.

Runyon BA (1998) Management of adult patients with ascites caused by cirrhosis. *Hepatology* 27: 264–272.

Selberg O, Böttcher J, Tusch G *et al.* (1997) Nutrition and metabolic assessment of candidates for liver transplantation. *Hepatology* 25: 652.

Sherlock S (1993) *Diseases of the Liver and Biliary System*, 9th edn. Oxford: Blackwell Scientific.

Wicks C and Madden A (1994) *A Practical Guide to Nutrition in Liver Disease*, 2nd edn. Birmingham, UK: The Liver Interest Group of the British Dietetic Association.

Zein NN, Abdulkarim AS, Weisner RH *et al.* (2000) Prevalence of DM in patients with endstage liver cirrhosis due to hepatitis C virus, alcoholic liver disease and cholastasis. *Journal of Hepatology* 32: 209–217.

Lobsters *See* **Shellfish**: Characteristics of Crustacea; Commercially Important Crustacea; Characteristics of Molluscs; Commercially Important Molluscs; Contamination and Spoilage of Molluscs and Crustaceans; Aquaculture of Commercially Important Molluscs and Crustaceans

LOW-FAT FOODS

Contents

Types and Manufacture

A J Rosenthal, Oxford Brookes University, Headington, Oxford, UK

Introduction

In a weight-conscious society where obesity is a common cause of death, and where icons of beauty are portrayed by a thin stature, there are social pressures to believe that we are empowered to control our weight through selecting low-fat foods. Inspired by this idea, the food industry has invested in developing processes and ingredients which enable the production of traditional products but with a fraction of the fat and therefore a fraction of the energy content.

The idea of reduced-fat foods has led to a variety of paradoxes in the public's perception, such as being abstemious with the selection of low-fat foods may be rewarded by high-fat treats (such as chocolate). Clearly this kind of logic is flawed. Furthermore there is evidence that the energy deficit resulting in the selection of low-fat foods is made-up by additional food consumption.

Figure 1 charts some high-energy foods in relation to their energy contribution from fat.

Strategies for Fat Reduction

Fats are often defined as food components which are soluble in organic solvents. Triglycerides are the predominant chemical species in food fats.

For a limited number of foods the fat may be removed without any adverse effect on the product. Such a simple approach has been successful with some dairy items in which whole milk has been replaced by skimmed. However, changes can occur in the functional properties of foods when fat is removed and generally one of three strategies is employed.

1. Fat may be wholly or partially replaced with one or more ingredients which have a lower energy content, but which contribute the essential functional and sensory attributes which were previously offered by fat. Reformulation is implicit in this approach. Alternative ingredients rarely behave identically to fat and often more than one new functional ingredient needs to be incorporated in order that the sensory and functional properties of the traditional food are maintained.

   A fat substitute is a compound which, like triglycerides, is soluble in an organic solvent, but whose chemical structure is neither metabolized nor adsorbed in the human digestive process. Chemists have rationally assessed the structure of triglycerides and produced analogs which do not occur in nature, for example by esterifying sugars with fatty acids one can produce compounds which have many similarities to triglycerides (Olestra is an example of such a product). In contrast to fat substitutes, a fat mimetic merely mimics the effects of fats but is predominantly water-soluble. Fat mimetics include water-soluble hydrocolloids (such as β-glucan from oat or

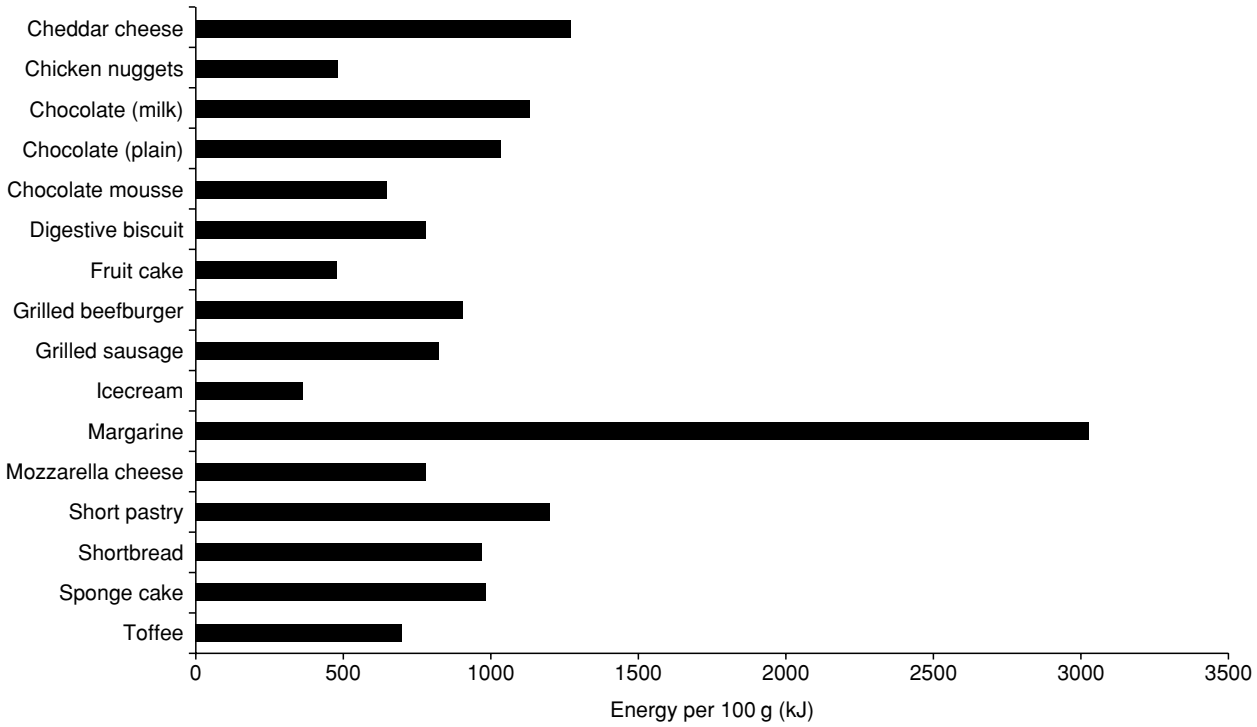

Figure 1 Energy contribution from fat.

barley) as well as microparticulate proteins (Simplesse is an example of such a product).

Many of the fat mimetics create the same mouth feel as fats by offering a viscous yet homogeneous texture. For example, tapioca dextrin can be hydrated to form a gel which can be used to replace < 50% oil in products like salad cream, providing viscous smooth mouth feel reminiscent of fat.

2. Fat may be diluted by the addition of air or water. In the past cynics have criticized the food industry for adding water to certain products (e.g., glaze on frozen foods). In contrast, the addition of air to whipped desserts, icecream, and mousses is often seen as desirable, producing a lighter, fluffier texture – this is despite the fact that these products are normally sold on a volume basis!

   Food emulsions consist of two immiscible phases – a high-energy oil phase and the lower-energy aqueous phase. The stability of emulsions may be controlled by increasing the viscosity of the continuous phase. In the case of oil-in-water emulsions, one often adds hydrocolloid stabilizing agents to the aqueous continuous phase, thus maintaining a high apparent viscosity. Since the gelation of such emulsions depends largely on the hydrocolloid concentration, one can effectively reduce the fat content while maintaining the texture. The texture of many hydrocolloids is smooth and creamy, thus furthering the illusion, without the energy from the fat.

   Another approach is to produce a three-phase water-in-oil-in-water emulsion in which oil droplets within the aqueous phase are further thinned out by a suspension of smaller water droplets inside. Multiple emulsions are normally achieved by two-stage homogenization. First, fine droplets of water are dispersed in oil emulsion using a lipophilic emulsifier. This primary emulsion is then dispersed in an aqueous phase using low concentrations of a hydrophilic emulsifier and a relatively gentle homogenizer pressure.

3. A limited number of products may be produced in the conventional way and subsequently have the fat removed by solvent extraction to produce reduced-fat foods. Solvent extraction has long been used in the production of cooking oils by extracting them from seed cake. One drawback of solvent extraction is the potential loss of desirable fat-soluble flavor compounds; this is relevant, as much of the flavor of volatiles in fried foods is produced during frying.

   Supercritical carbon dioxide has also been used to produce low-fat foods by extraction of fractions from conventional products such as potato crisps. A solid food like a flaked almond, whose shape and structure are defined, can have almost half the

fat removed without any appreciable loss in shape. The result is a lower-energy, lower-bulk-density nut with a relatively high protein content.

Considerations when Reducing Fat

To achieve successful reduction in the fat content of foods one needs to understand the roles which it plays during processing and subsequent storage, for example, its effect on the rheological and thermal properties of the food, its influence on the chemical and physical stability during storage, and, most importantly, its impact on sensory properties.

It has been suggested that one deciding factor in the enjoyment of fat is the smooth, creamy, homogeneous viscous mouth feel. Fats tend to be plastic in behavior and, when they melt, lubricating oils are produced. This is the desirable characteristic of cocoa butter with its unusually sharp melting point of 34 °C which happens to be just below internal temperature of the mouth, hence cocoa butter in chocolate elicits the sensation of melting away to a smooth creamy texture in the mouth.

The resolving power of the tongue and palate to small particles is in the region of 3 μm. Particulate materials with a diameter which is less than the resolving power of the tongue are therefore perceived as smooth homogeneous materials and can be used to mimic the smooth creamy mouth feel of fats. Microparticulate materials which have been used include microcrystalline cellulose, milk, and egg proteins.

Since fats are made up of mixtures of triglycerides, each with its own melting point, fats exhibit varying physical properties with temperature. When cold they tend to be solid and hard; as the temperature rises they soften as some of the triglycerides melt. Consequently melting occurs over a broad temperature range. Melting behavior is further complicated by the ability of fats to crystallize into different polymorphic forms, each with its own structure and melting point. For example, cocoa butter has six polymorphic forms (denoted I–VI). Type V is the most desirable, giving commercial chocolate its characteristic appearance; however if the chocolate melts during storage and then cools slowly, the more stable type VI can form with its less attractive appearance. The tempering step during chocolate manufacture is intended to form the desired polymorph preferentially.

In bakery shortenings, the β′ polymorph is reputed to be necessary to whip air into batters. This stabilization has been attributed to the lower level of crystal–crystal interaction compared with the β polymorph; this allows the β′ polymorph to adsorb as a dense layer at the interface, raising the interfacial viscosity.

Within emulsions, fats are of structural importance, providing one of the bulk phases. A high viscosity in the continuous phase helps to achieve emulsion stability. In the case of water-in-oil emulsions fat crystals present within the oil phase are found to adsorb on to the interface of the aqueous droplets and can help to stabilize the emulsion. Moreover as the crystals grow they form a solid matrix around the aqueous-phase droplets.

People sometimes ask ‘why do fried foods taste so good?’ In addition to the residual frying fats on the surface, with its lubricating properties, chemical changes such as the Maillard reaction occur during frying, which leads to the development of flavor, brittle, crisp textures, and characteristic golden-brown colors.

Many flavor volatiles are fat-soluble and may be introduced to the food in fatty ingredients. Replacing fats with substitutes or mimetics can significantly reduce the overall flavor.

Technological and Sensory Aspects of Selected Low-Fat Foods

Food Emulsions

Numerous food emulsions exist; the small size of the dispersed-phase droplets gives rise to the desirable sensation of creaminess. The high-calorie oil phase of emulsion-based foods makes them prime candidates for calorie replacement. Emulsion-based foods include margarine and butter substitutes, icecream, mousses and frozen desserts, salad cream, sandwich spread, and mayonnaise. Emulsions require stability against phase separation and attempts to replace their fat must take this into account. Emulsions also occur as intermediate products in the manufacture of other items, such as cakes, and emulsion stability during processing is important for product success.

Frozen desserts such as icecream are often high-fat products. A variety of fat replacers have been tried in frozen milk systems; temperature is a key parameter in terms of both product melting and ingredient solubility. It has been shown that carbohydrate-based fat mimickers are not able to achieve the high overruns found in conventional icecreams; moreover their consistency adversely affects the sensory properties of the products (making them coarser, more watery, and less creamy).

Intensity of flavor volatiles is reduced at low temperatures and clearly oil-soluble flavors are affected by replacing fat with microparticulate or hydrocolloid-based substitutes. Changes in flavor intensity become more apparent with storage and compensation in the product formulation is necessary.

Low-fat alternatives for butter are a burgeoning market. Several approaches have been taken in creating these products, for example, the formation of multiple-phase emulsions which dilute the dispersed fat droplets by filling them with smaller droplets of water. Another approach has been the incorporation of fat replacers into the formulation.

Melting of characteristics of low-fat spreads is important in producing the right mouth feel. Some starch-based low-fat spreads have long melting profiles on heating, which contrast with the melt-in-the mouth behavior of butter. Another important characteristic is rheological behavior. Spreads should behave as plastic solids, that is to say, they must resist small stresses and then start to flow when a small yield stress has been applied.

When dealing with products that contain raw egg, the protective effects of hydrocolloids must be considered if pathogens are to be destroyed during pasteurization. Such effects are well documented and rates of destruction of pathogens have been studied. It has been found that the destruction of *Listeria monocytogenes* is directly correlated with aqueous-phase concentration, thus reduced-calorie mayonnaise dressings offer greater protection than traditional mayonnaise.

Dairy Products

Though its composition changes with season, breed, and feed, the fat content of bovine milk is in the region of 4%. Removing the fat results in a change in the viscosity and color. Moreover the appearance markedly affects the perceived thickness and residual mouth coating of the milks. Attempts to enhance the color of skimmed milk have included the addition of whiteners such as titanium dioxide.

Low-fat, lite or light cheeses have had their fat content reduced by 50%, while reduced-fat cheeses have a fat content which is at least 25% less than the original reference fat content. Reduced-fat cheeses have been produced since the mid-1950s, though, as with other low-calorie products, demand has increased in recent years. Most of the work on fat replacement in cheese has tended to focus on a small number of cheeses, notably cheddar and mozzarella. Two basic approaches have been taken to achieve low-fat cheese and these are based on either modifications to the manufacturing procedure or the incorporation of fat replacers.

As far as modifying the manufacturing is concerned, increasing the moisture content can provide some of the creamy mouth feel offered by fat. Water content of the final cheese can be increased in a variety of procedures such as lowering the cook temperature, shortening the cook time, or curd washing. Flavor of low-fat cheese can be a problem. Rapid growth of the starter culture can lead to accelerated acid production which results in bitter flavors. It is better to have a small inoculum of starter culture to limit acid production and possibly to incorporate adjunct cultures which produce specific flavor compounds through increased proteolysis. Some adjunct cultures are deliberately attenuated by heat treatment to achieve particular flavor nuances. Commonly used adjunct cultures include *Lactobacillus* spp. and the growth of conditions in cheddar cheese requires a salt concentration of 1.8%. Considering that the market for low-fat cheese is likely to be a health-conscious one, concern about salt concentrations have prompted product development of low-fat, low-salt cheese replacing salt with potassium chloride.

A variety of fat replacers, including sucrose polyesters, microparticulate proteins, and hydrocolloid products, such as pectin or gelatin, have been used to produce low-fat cheeses. While fat substitutes successfully produce a similar mouth feel to full-fat cheese, they do not necessarily provide the flavor characteristics of the traditional product and some recipe compensation may be required. In practice, combining the two approaches of modifying the manufacturing process and incorporating fat replacers may give products of optimum quality. (*See* **Cheeses**: Types of Cheese.)

The essential properties of a cheese have to reflect the use to which it is being made, for example, when it is an ingredient used as a pizza topping, its meltability and rheological properties are fundamental. Microparticulate fat replacers interact with the casein from the milk, becoming embedded in the curd and forming links between adjacent bits of curd. Relatively large microparticulate fat replacers therefore create wide serum channels and a high degree of openness in the cheese. When melted, fat-replaced cheeses have a similar viscosity to conventional cheese.

Milk fat is often used as a food ingredient and consequently when fat replacement is an issue, the melting characteristics and consistency at different temperatures become important. If sucrose polyesters are used to replace milk fat in foods, then care must be exercised in the choice of the fatty acid esters used, as the chain length affects the hardness and melting characteristics. (*See* **Milk**: Physical and Chemical Properties.)

Reformed Meat Products

The main approach in fat reduction within meat products is to increase the water-holding capacity of the meat. A certain amount of water is lost from meats as a result of postmortem drip; further losses

can occur if the meat is cured because the curing salts draw water and soluble proteins out of the flesh by osmosis. With the advent of mechanization in food processing, attempts were made to speed the curing process and to limit the amount of water lost during curing. One solution is to incorporate water-binding salts, such as phosphates, into the brining solution. A similar approach has been taken in enhancing the water-holding capacity of reformed meat products, by adding phosphates. In the case of frankfurters, the addition of acid phosphate allows the incorporation of more connective tissue into the meat emulsion, thus adding to the value of the product and further diluting the fat content. The arguments for using phosphates in restructured meat products is not just one of economy, for, in addition to the ability to reduce the proportion of fat, added phosphates and water improve on some of the textural characteristics such as juiciness, tenderness, and overall palatability in meat patties.

Water binding of restructured meats has also been achieved by incorporating grain or bean flours, starches, as well as purified hydrocolloid gums like carrageenans or cellulose esters. These materials bind water but tend to modify the texture of the product. By way of reiterating the need to reflect on the desired product characteristics when selecting a fat replacer, consider the use of λ-carrageenan which has been investigated as a water-binding agent in a variety of low-fat meat products such as beef patties, restructured pork nuggets, and bologna sausage. This gum appears to soften products such that beef patties are juicier and more tender, but a firm product like bologna is also softer than the normal high-fat version.

Addition of low-fat protein extenders can also serve as a means of reducing the relative fat content of comminuted meat products. Lean meats such as goat can be incorporated into patties replacing 60% of the beef without consumers detecting any difference. Other extenders which have been used to extend the protein and lower the fat include connective tissue and soya flour.

Flour Confectionery

Butter has traditionally been used as the fat in most flour confectionery. Industrialization brought with it butter substitutes such as shortenings which, unlike butter, are not emulsions and contain little water. Butter, margarines, and shortenings are used in many bakery products such as cakes, cookies, crackers, and pastries. It has been observed that the addition of certain surfactants to biscuit dough and cake batters can help to maintain the texture of reduced-fat biscuits. Many commercial bakery shortenings actually have high levels of emulsifier. (*See* **Emulsifiers:** Phosphates as Meat Emulsion Stabilizers.)

In puff pastry the fat is laminated into the dough and forms an impervious layer which traps steam during subsequent baking, resulting in the formation of a light flaky characteristic texture. Various approaches have been taken to reduce the calorie content of the laminating material. Carbohydrate-based fat replacers are unsuitable as their hydrophilic properties prevent them from being able to form an impervious layer which means that steam is not trapped during baking. Instead it is necessary to use an oil-soluble, restructured lipid to form a water-insoluble barrier.

A variety of sweet biscuits and cookies have been the subject of fat replacement. A simple removal of fat without the addition of any substitute leads to a hard biscuit, with a reduction in spread during baking and a pale appearance. The incorporation of emulsifiers helps to soften the dough, enabling some spread during baking and resulting in a softer biscuit. Since reducing the fat content of biscuits generally makes a harder product, it might be assumed that reducing the fat of soft dough-type cookies is more difficult to achieve. Oat fiber may be used to replace 30% of the fat in soft-type cookies. The increased water content required to handle the dough insures a reduction in calories by one-third and also assists them to rise during baking. Care has to be exercised to avoid large particles of oat fiber which might result in a gritty mouth feel.

High levels of fat replacement in biscuits using microparticulates can lead to an increased mouth coating, a reduction in the fracturability and sensory firmness, as well as a reduction in the flavors. Being water-based, carbohydrate fat replacers lead to an increase in the moisture content which in turn results in denser and tougher biscuit. Overall it can be seen that studies on the replacement of fat in biscuits have not been altogether successful; only limited reductions in the fat content have been achieved and replacement tends to result in artifacts compared to the conventional product.

In cake batters the fat has several functions: it acts as a vehicle for some of the oil-soluble volatile flavor compounds; it traps air during mixing, and then during baking it becomes integrated into the protein matrix, adding to the tenderness of the final crumb. Complete substitution of fat in high-ratio cakes has been achieved using a variety of hydrocolloid-based fat replacers; however this is accompanied by a reduction in batter overrun, leading to a reduced baked volume.

Bakery fillings and coating mixtures are conventionally based on a fatty formulation. Usually such

materials have a continuous phase of fat into which sugar, milk solids, and cocoa are dispersed. When the coating is in the mouth, the fat melts, releasing the flavors and the characteristic creamy mouth feel. If the continuous phase is replaced with a water-soluble hydrocolloid, such as microcrystalline cellulose, then the sugar present forms a concentrated solution which may be pumped to the enrobing machines in the factory. When the mixture is cooled the sucrose solution becomes saturated, the solids present dominate, and the mixture sets. The higher temperature in the mouth causes the coating to liquefy, releasing the flavor and the same characteristic creamy mouth feel of a fat.

Frying Oils and Batters

The high-fat content, yet desirable characteristic flavors and textures of fried foods, make them important targets for fat replacement. The high temperatures involved in frying make most of the hydrocolloid-based and microparticulate fat mimetics unsuitable for the operation. High-boiling, noncalorific fat replacers such as sorbitol esters or polyorganosiloxanes may be used as alternatives to frying oils.

A different approach to reducing fat uptake during deep-fat frying is the incorporation of materials which provide barrier properties to the batter or breaded layers of the food. For example, an edible film of hydroxypropyl methylcellulose can reduce the fat uptake of chicken pieces by one-third.

Fried food analogs such as 'oven-ready' foods often have coatings or batter mixtures which contain fats with the intention that during cooking they yield some of the flavors and textures of their fried counterparts.

Some low-fat fried snack foods are produced in the conventional way and then the fat is extracted using supercritical carbon dioxide.

Conclusions

The promise of low-fat foods has enticed our society and captivated food technologists. Simple reduction in the fatty ingredients can result in a change in the sensory and functional properties of the food. Three approaches to reducing fat have been taken: (1) substitution of the fat with one or more functional ingredients; (2) dilution of the fat with water or air; and (3) extraction of the high-calorie constituents after conventional manufacture.

See also: **Colloids and Emulsions**; **Meat**: Sausages and Comminuted Products; **Sensory Evaluation**: Texture; **Sweets and Candies**: Sugar Confectionery

Further Reading

de Bruijne DW and Bot A (1999) Fabricated fat based foods. In: Rosenthal AJ (ed.) *Food Texture: Measurement and Perception*, Gaithersburg: Aspen Publishers. pp. 185–227.

Rosenthal AJ (1998) Technological factors in the development of low-calorie foods. In: Henry CJK and Heppell NJ (eds) *Nutritional Implications of New Food Ingredients and Processes*, Gaithersburg: Aspen Publishers. pp. 24–44.

Low-fat Spreads

J Dostálová, Institute of Chemical Technology, Prague, Czech Republic

Introduction

It is now generally recognized that a high-calorie diet with inadequate physical activity results in overweight or obesity, which in turn could lead to diseases such as noninsulin-dependent diabetes, hypertension, cardiovascular diseases, endometrial cancer, and gallstones. The apparent connection between diet and health has increased awareness for 'healthier' eating; the reduced-fat diet is especially in vogue. This trend is reflected, in particular in the USA and in Europe, in the increased market for low-fat products, including low-fat spreads. Most of the market surveys carried out in recent years indicate that the demand for 'healthier' food will increase, but not at the cost of sensory quality and at too high a price. The sensory characteristics (flavor, taste, color, and texture) of low-calorie products have to be comparable with those of traditional products. These requirements present the industry with a challenge and an opportunity to develop new technologies and new and novel additives and food ingredients. A great deal of these targets were realized with the development of low-fat and very-low-fat spreads or fat-free spread products.

Brief History

Margarine, which was developed in France in 1869, must contain a minimum of 80% fat. Health concerns have led to the development of many low-fat spread products, which have lower energy content. Some interesting data from the history of margarine with lower fat content are as follows:

- Diet margarine, containing half the calories of regular margarine, was introduced in 1964 in the USA. The US Food and Drug Administration (FDA) questioned the legality of the product but lost an ensuing court case, which confirmed diet margarine as a table-spread product. (*See* **Margarine**: Types and Properties; Methods of Manufacture.)
- A product containing 60% fat was introduced in 1975 that could not be identified as either margarine, as it did not contain 80% fat, or diet margarine, which must contain 40% fat or half the calories of a regular margarine.
- In 1981 margarine butter blends containing as much as 40% to as little as 5% butter were introduced by several margarine and butter processors. Some were in the low-fat form. These premium products, priced between butter and margarine, had a more characteristic butter flavor, but improved spreadability and the health benefits of margarine.
- Lower-fat-content spreads of less than 20% fat were introduced in 1989.
- In the 1990s, low-fat spreads containing plant sterols which decrease cholesterol level in human blood serum were developed.
- Spreads, which ranged from zero to 68% fat, and were free of *trans* fatty acids, were introduced in 1997. (*See* **Emulsifiers**: Organic Emulsifiers; Phosphates as Meat Emulsion Stabilizers; **Fatty Acids**: Properties.)

Statistical data in the western countries identify the shift in consumer table-spread popularity to low-fat spreads. The spread market share increased, for example, in the USA from less than 5% in 1976 to more than 74% in 1995. In the European Union, 75–80% of commercial spreads in 1985 contained 80% fat. By 1993, 75–80% of spreads marketed contained 40% fat.

Description, Terminology, and Classification

The food category referred to as table spreads consists of several types of products, including butter and margarine. Commercial table spreads now exist that contain fat levels ranging from a high of 80% all the way down to 0%. The standards in most countries require that butter and margarine contain not less than 80% by weight of fat. Products resembling margarine containing less than 80% fat are usually called spreads. After regulations in some countries, only products containing less than 80%, but more than 40%, fat, 40–70% fat, 62–80% or less than 75% fat are labeled as spreads. Products with 60–80% fat or with 41–60% are reduced-fat spreads, and products containing less than 40% fat are referred to as low-fat spreads. The term very-low-fat spread is used for spreads of 5–15% fat and even less. The spreads with extremely low fat content are sometimes called ultra-low-fat products.

The terminology and classification are not unified and they are laid down by national legislation in accordance with the needs of each individual country. Therefore different terms can be found for low-fat spreads, e.g., halvarine, minarine, half-fat margarine, margarine 40, margarine with low fat content, light or lite margarines, etc.

The sixth edition of the *Dictionary of Nutrition and Food Industry* defines halvarine as 'a name sometimes given to low-fat spreads with less than the statutory amount of fat in a margarine.' The same definition is used for minarine in this encyclopedia.

Codex Alimentarius defines minarine as follows: 'minarine is a food in the form of spreadable emulsion, which is mainly of the type water/oil, produced principally from water and edible fats and oils which are not solely derived from milk, and in which the fat content is not less than 39% m/m and not more than 41% m/m.'

The term 'halvarine' is sometimes used for a product with the same fat content as minarine, defined by Codex Alimentarius, i.e., a fat product with fat content of not less than 39% m/m and not more than 41% m/m.

The above-mentioned definitions for minarine and halvarine are more precise than the definitions used in the *Dictionary of Nutrition and Food Industry.*

It is possible to find other terms for low-fat spreads with a fat content of not less than 39% m/m and not more than 41% m/m in national legislations as well. For example, half-fat margarine in Germany and Austria, margarine 40 in Denmark, margarine with low-fat content in the Czech Republic.

Fat products with a fat content less than 39% m/m and more than 41% m/m but less than 80% m/m are called spread fat x% m/m in the above-mentioned national legislation, with the exception of products with a fat content not less than 60% m/m and not more 62%. Those products are called three-quarter-fat margarines in Germany and Austria, margarine 60 in Denmark, margarine with reduced fat content in Czech Republic, and elsewhere mellarines.

In some literature and legislative sources, products with 40% fat or less are identified as 'light' or 'lite' margarines. The 'light' product must contain no more than one-half the fat and one-half the energy of the reference product.

There are no published regulations for some types of low-fat spreads in some countries but labels on such products must list the total percentages of fat, and in some countries, sources of fats in weight order as well.

Products based only on all-vegetable oil and fats blends are mostly defined as low-fat spreads. From the nutritional and culinary points of view, spreads with low fat content based on milk fat (milk-fat product) or blends of vegetable oils and fats with milk fat (mixed-fat products) are also considered to be low-fat spreads. The terminology of these products is analogous to that of vegetable low-fat spreads. Half-fat butter or half-fat blend are terms for products with a fat content of not less than 39% m/m and not more than 41% m/m. Fat products with a fat content less than 39% m/m and more than 41% m/m but less than 80% m/m are called dairy spread fat x% m/m or blended spread fat x% m/m, with the exception of products with fat content not less than 60% m/m and not more than 62%. These products are called three-quarter-fat butter or three-quarter-fat-blend. The terms low-fat butter (blend) or 'light' butter are used as well.

Low-fat mayonnaise with a thick texture can serve as fat spread on bread or biscuits as well. Hence, we can also rank low-fat mayonnaise (fat content can be less than 30%) among low-fat spreads. The composition of an oil phase of low-fat mayonnaise resembles the oil phase of low-fat spreads, but low-fat mayonnaise differs significantly from low-fat spreads in that it is an 'oil-in-water' emulsion. Various food ingredients (vegetables, spices, etc.) are often added to those products.

Classification systems used for global national statistical purposes do not differentiate spreads according to their fat content in most cases. For example, the list of the industrial production Prodcom (product of the European Community), which is obligatory in the states of the European Union, only differentiates the group 'spreads with reduced or low content of fat (less than 80% fat).'

The Procome (Eurostat (Statistical Office of the European Communities) version) used for household budget survey data only groups together 'margarine and other vegetable fats.' The new classification system Eurocode 2 was proposed, therefore, by the Cost Action 99 project. This system classifies spreads as follows:

- fat spread > 65% fat, > 25% saturated fatty acids
- fat spread > 65% fat, < 25% saturated fatty acids
- fat spread 45–65% fat, > 25% saturated fatty acids
- fat spread 45–65% fat, > 25% saturated fatty acids
- fat spread 30–45% fat
- fat spread < 30% fat

Formulation and Processing

Low-fat spreads designed for spreading on bread and biscuits from a refrigerated store are now typically based on all-vegetable-oil blends in the form of water-in-oil emulsions. In the margarine, fat or oil is the continuous phase and the majority phase; water is the discontinuous phase and the minority phase of the emulsion. In low-fat spreads the oil-continuous phase is still a major factor (in ultra-low-fat products water is the continuous phase). The oil phase must still have a structure which is determined by the size, shape, and arrangement of fat crystals. This network of crystals determines the distribution of melt point types and provides the form and texture of the finished spread. The amount of these crystals, expressed as the solid fat index or solid fat content, is important. In addition to the amount of solid fat, the crystalline form is also very important.

Two considerations – formation and stability of the emulsion, and the structure of the aqueous phase – take on greater importance in low-fat spreads in comparison with full-fat products.

The formation of an emulsion requires the presence of surface-active agents (called emulsifiers or surfactants) to facilitate the formation of the dispersion of one immiscible liquid (water) into another (oil). The emulsifier reduces the interfacial tension between the aqueous phase and the oil phase, and stabilizes the spread during its storage by preventing the aqueous phase from coalescing. The stabilization of emulsion becomes more important as the percentage of fat phase is lowered and the percentage of the aqueous phase is increased. Therefore, low-fat spread should contain larger amounts or more efficient emulsifiers. Water is still the dispersed phase of the emulsion but becomes the majority phase of the emulsion. The volume of water fraction is much larger and can create a far greater degree of instability in low-fat spreads.

Structuring of the aqueous phase is the second focus to obtain long-term shelf stability. Low-fat spreads can be stabilized in the following ways:

- Adding bulk ingredients (e.g., maltodextrin) to take up or hold the increased level of added water. Some of these ingredients may also increase the viscosity and thickness of the aqueous phase and contribute to better sensory quality similar to that of 'full-fat' products.
- Using various gelling agents which reduce syneresis, such as alginate, pectin, carrageenan, starch, whey protein, soy protein, casein, and gelatin.
- Adding gelled microparticles which have a spherical or globular shape which act as fat mimetics.

Very-low-fat spreads are water-continuous systems (oil-in-water emulsions). It is not difficult to disperse oil in water but it is difficult to provide a good sensory quality of a spread, comparable with that of full-fat products. Several systems have been developed for this purpose, mostly on the basis of proteins, polysaccharides, or blends of them.

The composition of the fat phase in low-fat spreads is very important. Low-fat spreads are usually formulated from the same fat and oil blends as those used for regular margarines, but certain basic features must be considered in terms of the solid fat content. The solid fat index profile determines spreadability, defines how the fat blend melts in the mouth, and determines the thickness on the palate, and speed of flavor release. Processors modify and blend oils in order to achieve optimal characteristics. It is now becoming increasingly appropriate to define oil blends for spreads by their triacylglycerol types. Nutritional aspects as well as shelf-life must be considered.

There are varieties of blends used, adjusted according to the fat level and sensory and nutritional characteristic targets. Selectively and nonselectively hydrogenated oils, interesterifed fractions, vegetable fats, or their selected fractions and blends are utilized as hardstock.

Milk fat is also used in some products. Margarine products usually contain milk fat maximum 3% of total fat and mixed-fat products minimum 15%, maximum 80% of total fat. Restrictions with respect to the proportion of milk fat to other types of fat may be imposed in accordance with national or other relevant legislations.

The use of other ingredients in low-fat spreads differs form their use in full-fat products in the following points:

- Most spread formulations are milk- and milk protein-free, because milk protein acts as an oil-water emulsifier and consequently, the use of milk, casein, or caseinates can result in a phase reversion.
- Emulsifier levels are increased to improve physical characteristics of the emulsion and its stability.
- Lecithin may decrease the emulsion stability and increase the tendency to oil off. However, it also functions to slow down emulsion breakdown in the mouth. Therefore, the use of lecithin and its level must be evaluated for each formulation.
- Hydrocolloids, such as gelatin, pectin, carrageenans, agar, xanthan gum, starch, alginates, or methyl cellulose derivatives, are used in some spreads as thickening agents to improve the body.
- Flavor content and types must be defined to produce oral response similar to full-fat products because the higher emulsifier levels used for spreads can produce tighter emulsions, and gelling or thickening agents can affect the rate and how flavor is perceived.
- Preservatives are more important in spreads than in regular margarines because of higher moisture content.
- The light reflection of a spread is different from that of a regular margarine because of the increased number of water droplets present. Therefore, it is necessary to add about twice the amount of color used in normal margarine to obtain the same color intensity in low-fat products.

During processing of low-fat spreads the preparation of an emulsion becomes of greater importance. Chilling and crystallization may be similar to those used for margarine.

Finished low-fat products have to meet all sensory requirements as for full-fat products – clean and full taste and smell, uniform and sheen color, even and smooth surface and pleasing melt in the mouth. Spreadability from the refrigerator and at room temperature must be good, without syneresis and separation.

Composition

Essential Composition of Minarine according to Codex Standard

- Raw materials: water and/or milk and/or milk products; edible fats and/or oils, or a mixture of these, whether or not they have been subjected to a process of modification.
- Fat content: not less than 39% m/m and not more than 41% m/m.
- Water content: not less than 50%, as determined by loss on drying.
- Optional ingredients: the following substances may be added: vitamins (vitamin A and its esters, vitamin D, vitamin E and its esters, other vitamins), egg yolk, sodium chloride, sugars, suitable edible proteins, gelatin, natural starches.

Food Additives for Minarine According to Codex Standard

- Colors: the following colors are permitted: beta-carotene, annatto extracts, curcumin or turmeric, beta-apo-8′-carotenal, methyl and ethyl esters of beta-apo-8′ carotenoic acid.
- Flavors: natural flavors and their identical synthetic equivalents, except those which are known to represent a toxic hazard, and other synthetic flavors approved by the Codex Alimentarius Commission are permitted.

- Antioxidants: propyl gallate, butylated hydroxytoluene (BHT), butylated hydroxyanisole (BHA), tertiary butyl hydroquinone (TBHQ), any combination of propyl gallate, BHA, BHT and/or TBHQ, natural and synthetic tocopherols, ascorbyl palmitate, ascorbyl stearate, dilauryl thiodipropionate.
- Antioxidant synergists: citric acid, sodium citrate.
- Thickening/stabilizing agents: agar, ammonium alginate, carboxy methyl cellulose, carrageenan (including furcelleran), guar gum, methyl cellulose, pectin (amidated and nonamidated), alginate (K, Ca, Na salts), locust bean gum, propylene glycol alginate, sodium carboxy methyl cellulose, xanthan gum.
- Emulsifiers: lecithin, mono- and diacylglycerols, polyglycerol esters of interesterified ricinoleic acid, polyglycerol esters of fatty acids, polyoxyethylene (20) sorbitan mono-oleate, monopalmitate, monostearate, tristearate.
- Preservatives: sorbic acid and its sodium, potassium and calcium salts, benzoic acid and its potassium and sodium salts.
- pH-correcting agents: lactic acid and its sodium and potassium salts, citric acid and its sodium salt, sodium hydroxide, sodium carbonate, sodium phosphate, sodium tartrate.

The composition of fatty acids in the product depends on the type and ratio of oils and fats used. Table 1 presents the fatty acid composition of 10 low-fat spreads produced in Europe and the USA. From a nutritional point of view, the composition of individual low-fat spreads does not differ too much. All low-fat spreads have a relatively high content of unsaturated fatty acids.

Table 1 Contents of saturated, monosaturated, and poly un-saturated fatty acids in 10 low-fat spreads (40% fat) produced in Europe and the USA (% of total fatty acids)

Spread	*Saturated fatty acids*	*Monounsaturated fatty acids*	*Polyunsaturated fatty acids*
1	30.4	43.3	26.3
2	21.3	36.4	42.3
3	32.7	22.3	45.0
4	32.3	23.6	44.1
5	17.2	39.0	43.8
6	19.5	49.7	30.8
7	22.4	38.4	39.2
8	27.3	38.1	34.6
9	17.6	45.1	37.3
10	20.7	42.2	37.1

1–4, low-fat spreads produced in Europe (Czech Nutrient Database).
5–10, low-fat spreads produced in the USA (USDA Nutrient Database for Standard Reference).

Dietary Importance and Use

Nutrition value, including energy value, of low-fat spreads is in good agreement with health authority recommendations for the populations of developed industrial countries to reduce their intake of energy, total fat, saturated fat, and cholesterol. The amount of energy in low-fat spreads containing 40% fat is reduced to 50% in comparison with classical margarine or butter. Average daily intake of fat spread used for spreading on bread or biscuits is 20–25 g, and this represents 16–20 g fat in the case of classical margarine. The same amount of low-fat spread represents only 8–10 g fat. The reduced intake of fat (about 15% average recommended daily fat intake) is not negligible. As has been discussed elsewhere, the decrease of fat in low-fat spreads is partially compensated by an increase in the consumption of complex carbohydrates. Hence, the total energy intake is not much lower but nutrition balance is better because fat intake is lower and complex carbohydrate intake is increased. Low-fat spreads help consumers to meet their needs in terms of low intake of saturated fat and high intake of polyunsaturated fatty acids.

Research over recent years has shown that *trans* fatty acids have adverse affects on health with regard to coronary heart disease, and therefore companies are now offering some low-fat spreads free of *trans* fatty acids.

Low-fat spreads are free of cholesterol (except trace amounts, which are insignificant from a nutritional point of view). In contrast, they contain serum cholesterol-lowering plant sterols (phytosterols) and therefore they have a significant role in decreasing the risk of heart disease. Clinical studies indicate that an intake of 2–3 g day^{-1} of these plant sterols or stanols reduces serum cholesterol level. Therefore plant sterol-enriched margarines or low-fat spreads have been developed. These contain 8 g plant sterols or stanols per 100 g, so that 25 g of margarine or low-fat spread ensures a recommended daily intake of 2–3 g plant sterols. This dose will reduce total cholesterol by around 10% and low-density lipoprotein-cholesterol by 10–15%, with minimal change in high-density lipoprotein-cholesterol and triacylglycerols. (*See* **Cholesterol**: Properties and Determination; Absorption Function, and Metabolism; Factors Determining Blood Cholesterol Levels.)

Low-fat spreads are formulated to supply amounts of fat-soluble vitamins (A, D, E) which are similar to or higher than those supplied by butter, and absorption is unlikely to be impaired. Low-fat spreads might thus be viewed positively as a vehicle for lipophilic nutrients without excessive lipid, and can contribute to fulfillment of the recommended daily intake of A, D and E vitamins.

Low-fat spreads are intended solely as spreads on bread or biscuits. They should not used for cooking or baking and definitely not for frying, on account of their high water content and in some cases (low-fat butter, low-fat blends) of their high protein content.

Safety

The safety of low-fat spreads is assured by special regulations. The use of raw material, optional ingredients, and food additives are regulated by national standards. Contents of contaminants are regulated as well. The products have to be prepared in accordance with the general principles of food hygiene. A safety issue, which is not commonly discussed in relation to light products, is that relating to microbiological stability and hazard. Where traditional foods are replaced by light foods with a higher water activity, the consumer may not recognize that they must be stored differently from traditional foods unless they contain preservatives. In traditional butter and margarine, salt is present not only for sensory reasons. The amount of 2% salt in whole butter is equivalent to 13–14% in the aqueous phase and exerts a significant preservative effect. Low-fat substitutes of butter with a higher water content are not preserved in this way and should be stored in a refrigerator. Instructions on special storage requirements need to be displayed prominently on the label so that the risks of spoilage and pathogenic mold growth are avoided.

Low-fat spreads are generally filled in plastic tubs, which are designed not only to give mechanical protection but also to protect the product from light and oxygen. Evaporation of water from the surface, which leads to an unsightly dark-yellow coloration, is prevented by use of a sealed aluminum foil.

Nutritional demands have led to the manufacture of blends containing a high content of polyunsaturated fatty acids. These products are more susceptible to oxidative deterioration and damage of texture, and therefore require storage and transport at low temperatures, preferably at less than 5 °C; these conditions reduce the risk of oiling out, and maintain flavor and textural qualities. Oxidative changes may be diminished by the use of antioxidants.

Conclusions

Low-fat spreads belong to progressive food products. They meet the demands of consumers for 'healthy eating' very well. Now, many types of low-fat spreads with varying fat content are available on the market. They have a high content of polyunsaturated fatty acids, and many of them are *trans* fatty acids-free. The sensory characteristics of low-fat spreads are comparable to those of full-fat products. The main trends in the development of low-fat spreads are towards tailor-made products to provide health benefits, e.g., low-fat spreads with plant sterols or stanols, *n*-3 fatty acids of fish origin as added ingredient.

See also: **Dressings and Mayonnaise**: The Products and Their Manufacture; **Emulsifiers**: Uses in Processed Foods; **Fats**: Uses in the Food Industry; **Fatty Acids**: Dietary Importance; **Food Safety**; **Gums**: Food Uses; **Legislation**: Codex; **Low-fat Foods**: Types and Manufacture; **Margarine**: Types and Properties; Methods of Manufacture; Composition and Analysis; Dietary Importance; **Storage Stability**: Parameters Affecting Storage Stability; **Vegetable Oils**: Types and Properties

Further Reading

Bylund G (1995) *Dairy Processing Handbook*. Lund: Tetra Pak.

FAO/WHO (1993) *Codex Alimentarius*, vol. 8, *Fats, Oils and Related Products*. Rome: FAO/WHO.

Hamm W and Hamilton RJ (eds) (1998) *Edible Oil Processing*. Sheffield: Academic Press.

Haumann FB (1998) Consumer table spreads. *INFORM* 9: 6–13.

Krawczyk GR, Buliga GS, Bertrand DT and Humphreys WM (1996) Reviewing the technology of low-fat spreads. *INFORM* 7: 635–639.

Leathwood PD, Louis-Sylvestre J and Mareschi JP (eds) (1995) *Light Food*. Brussels: ILSI Europe.

O'Brien RD (1998) *Fats and Oils: Formulating and Processing for Applications*. Lancaster: Technomic.

Verordnung des Rates nr. 2991/94 vom 5 Dezember (1994) mit Normen für Streichfette. Brussels: Der Rat der Europäischen Union.

LUPIN

L C Trugo, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
D von Baer, Universidade de Concepción, Concepción, Chile
E von Baer, CAMPEX Semillas Baer, Temuco, Chile

Introduction

Lupin is a legume plant from which the seeds have been used as food and feed from ancient times. The green plant has been used extensively in some countries as forage for animal feeding and also as organic material for soil enrichment or in crop rotation. Despite its qualities and potential, it has not become a major world crop. This article will review the main characteristics of this plant, the nutritive value of its seeds, and the uses and potential for more extensive production.

The Plant and its Origin

Lupin or lupine are trivial names for plants of the genus *Lupinus* belonging to the Leguminosae family, subfamily Papilonoideae. This genus is very diverse and contains several known species. It is an annual plant, usually 0.3–2 m in height, with a highly branched stem, digitated leaves, and a shrubby growth pattern. The flowers formed high above the leaves, may be of several different colors (white, yellow, purple, blue), and exude a honey-like aroma. The roots are relatively long, usually nodular, and may grow down to 3 m. The lupin plant has the ability to fix nitrogen and mobilize phosphorus and other elements by the exudation of citrate to the soil. This is very important for soil enrichment in infertile areas. The pods are normally flat and with a hairy aspect on the outside. They vary in length (4–10 cm) and, depending on the species, have different types of seeds that may vary in size, shape, or color. However, the endosperm is usually yellowish, and the most common seeds are flattened (*Lupinus albus*) or ovoid (*Lupinus mutabilis*) with diameters in the range 2–15 mm. It is a plant well adapted to poor sandy and acid soils and also to high altitudes, and to temperate climates, as found in Europe. The most important species that are produced on a commercial scale are *L. albus*, *L. angustifolius*, *L. luteus*, and *L. mutabilis* (see color plate 101).

The lupin has its origin in the Old World in the Mediterranean region, and also in the New World, in North America and in the Andean highlands. Some representative examples of the former are *L. albus*, *L. angustifolius*, *L. luteus*, and *L. consentinii*. Tarwi is the indigenous name of *L. mutabilis*, which is the lupin found in the Andes (Table 1). However, there are probably more than 200 wild species, most of them in the New World. *L. albus* was cultivated as a food by the ancient Romans, Egyptians, and Greeks; however, due to its bitter taste, the mature seed had to be submitted to prolonged washing before consumption. A similar treatment was also used by the ancient inhabitants of the Andes. The seeds of tarwi were boiled, thoroughly washed in a river for several days, and then properly cooked for immediate consumption or dried for future use. Presently, several different species of lupins are grown in Europe, some parts of Africa, South America, Australia, and New Zealand, including a number of sweet low-alkaloid varieties that are now commercially available.

Nutritive Value

Lupin seeds show similar nutritive attributes to soya beans, particularly with respect to protein and fat content. Some species may have protein contents up to 50 g per 100 g and up to 22 g per 100 g of fat in the seeds. The proximate composition of seeds of some relevant lupin species are presented in Table 2. The protein and fat content may be further increased by dehulling the seeds.

Protein

The lupin seed is one of the richest sources of plant protein. The protein content varies largely between species, with *L. mutabilis* and *L. luteus* presenting unusually high values (near 50 g per 100 g), surpassing most soya bean cultivars, whereas *L. albus* presents an intermediate protein content and *L. angustifolius* lower values. Lupin seed protein provides all the

Table 1 Taxonomic and common names of some commercial lupin species

Species	*Common names*
L. albus[a]	White lupin, Egyptian lupin, tremoo[b], altramuz[c]
L. angustifolius[a]	Blue lupin, narrow-leafed lupin
L. luteus[a]	Yellow lupin, tremosilla[c]
L. mutabilis[d]	Tarwi, tauri, tarhui, chocho[c], Andean lupin

[a]Mediterranean origin.
[b]Portuguese name.
[c]Spanish names: altramuz may be used also to designate other species.
[d]Andean origin; tarwi, tauri, and tarhul are from native languages.

essential amino acids, but some of them are not found in sufficient amounts for all sectors of the population. When lupin seed is used for animal feeding, such as ruminants or single-stomach animals, the relatively low level of methionine may be complemented by adequate formulation of the ration or by supplementation with the synthetic amino acid. Table 3 shows the amino acid composition of seeds of some relevant lupin species and the amino acid requirements for infants and adults established by a Food and Agriculture Organization/World Health Organization/United Nations University (FAO/WHO/UNU) joint committee (1985). When comparing the lupin amino acid profiles with the requirements for adults, all lupin species show ranges of essential amino acids with adequate amounts, and this will be observed also for most pulses. However, if the requirements for infants are used for comparison, it is apparent that lupin protein will show many essential amino acid deficiencies, with tryptophan being the most limiting, having a chemical score (CS) of 47–59%, followed by methionine plus cystine. The chemical score (CS) is the ratio of the content of an individual essential amino acids in food protein to the content of the same amino acid in a reference pattern expressed as a percentage. It must be taken into account that, if requirements for infants are used for comparison, even beef protein will be deficient in leucine, lysine, methionine+cystine, tryptophan, and valine, and egg protein will be deficient in leucine and lysine. However, if the previous FAO/73 amino acid pattern is used as a reference, methionine+cystine will be the limiting amino acids for lupin (56–71% CS), as in most pulses (soya bean 83% CS). Consequently, it is strongly recommended that the intended nutritional application of lupin seeds or lupin products be defined to allow a more realistic assessment of the protein quality. From Table 3, it can be seen that the amino acid profile varies considerably with lupin species. Large variations are also observed in biological indicators for protein-quality assessment. Nevertheless, protein utilization, as measured by the net protein utilization (NPU) index, does not correspond exactly to the amino acid CS of different lupin species. For example, *L. angustifolius* shows a relatively high CS (60%, FAO/85 for infants) but presents the lowest NPU (43%). Despite its amino acid composition, active biological components appear to have a negative influence on the NPU of *L. angustifolius* protein in rats. *L. albus* shows the highest NPU (70%) followed by *L. mutabilis* (64%). These figures approach the NPU of the soya bean (69%). Overall, lupin protein should be considered a good plant source of protein for

Table 2 Major components of lupin seeds[a]

Species	*TCW (g)*[b,c]	*Protein*[c]	*Oil*[c]	*Ash*	*Fiber*[d]
L. albus	370	38.2	9.5	3.1	14.0
L. angustifolius	165	28.9	6.6	3.2	17.6
L. consentinii	165	32.2	2.6	3.7	
L. hispanicus	71	48.2	5.8	2.9	19.6
L. luteus	164	49.2	4.8	5.3	17.6
L. mutabilis	211	46.6	15.8	3.6	9.0

[a]Results expressed in grams per 100 g of dry matter (except TCW).
[b]Thousand corn weight.
[c]Data adapted from Trugo LC, Almeida DCF and Gross R (1988) Oligosaccharide contents in the seeds of cultivated lupins. *Journal of the Science of Food and Agriculture* 45: 21–24, with permission.
[d]Adapted from Muzquiz M, Burbano C, Bouthelier V *et al*. (1982) Estudio de los elementos esenciales de destintas variedades de cinco especies del *Genus Lupins* cultivadas y espontaneas de la Peninsula Iberica. In: *Proceedings of the 2nd International Lupin Conference*, pp. 173–181.

Table 3 Amino acid composition of lupin seeds

Amino acid	*Concentration (mg per gram of protein)*				*FAO recommendations*[c]	
	L. albus[a]	*L. angustifolius*[a]	*L. luteus*[a]	*L. mutabilis*[b]	*Infants*	*Adults*
Histidine	22	27	29	26	26	16
Isoleucine	44	40	37	57	46	13
Leucine	75	71	79	71	93	19
Lysine	47	46	49	58	66	16
Mehionine	7	7	7	9		
Cystine	16	18	24	11		
Methionine + cystine	23	25	31	20	42	17
Phenylalanine	35	37	39	38		
Tyrosine	46	34	27	40		
Phenylalanine + tyrosine	81	71	66	78	72	19
Threonine	36	34	32	38	43	9
Tryptophan	8	8	10	8	17	5
Valine	39	35	32	41	55	13

[a]Adapted from Gross R (1988) Lupins in human nutrition. *Proceedings of the 5th International Lupin Conference*, pp. 51–63, with permission.
[b]Adapted from Gross R, Koch F, Malaga I *et al*. (1989) Chemical composition and protein quality of some local Andean food sources. *Food Chemistry* 34: 25–34, with permission.
[c]FAO/WHO/UNU (1985) *Necessidades de Energia y Proteinas*. Geneva: World Health Organization.

animal and human nutrition. However, the appropriate application and the variation encountered between species must be considered for assessing lupin protein quality. (*See* **Amino Acids**: Metabolism; **Protein**: Quality.)

Fat

The fat content of lupin seeds varies considerably between species. In some species, the lipid fraction contributes substantially to the total energy value, whereas in others, it is present in relatively small amounts (Table 2). *L. mutabilis* has the highest oil content (up to 22 g per 100 g) and is considered to be a potential material for oil production. *L. albus* has a combination of high protein and an intermediate fat content, which are important attributes for a food, especially if it is to be used in developing countries where protein and energy are scarce. The fatty acid distribution in the oil fraction is also variable between species. However, low levels of saturated fatty acids are normally found. A very high content of linolenic acid is present, particularly in *L. albus* and *L. luteus*. As in other seeds, high levels of oleic acid are common for all species, particularly *L. albus* and *L. mutabilis* (Table 4). The ratio of polyunsaturated fatty acids (PUFA) to saturated fatty acids (SFA) in lupin seeds is in the range 1.3–2.9:1, which is considerably higher than beef but lower than the soya bean (4.0). Higher rates are important, since it has been strongly recommended that a reduction of SFA in favor of an increase of PUFA in the diet should be achieved, to assist coronary heart disease prevention. Linoleic acid, which is an essential fatty acid, is found in significant amounts in lupin seed, with *L. luteus* being the richest source (44%). Linolenic acid, which is the most important source of ω-3 fatty acids, is found at very high levels in *L. albus* and *L. luteus*, but at relatively low levels in *L. angustifolius* and *L. mutabilis* (Table 4). *L. mutabilis* may be commercially used for oil production, and its low linolenic acid content, although not so nutritionally important as in other sources, may be advantageous to improve its conservation in comparison with other vegetable oils. Considering the fatty acid profile of lupin seeds, it is noted that *L. mutabilis* and *L. albus* are more similar to peanut oil, whereas *L. angustifolius* and *L. luteus* resemble maize oils, although they differ in flavor. (*See* **Essential Fatty Acids**; **Fatty Acids**: Properties; Gamma-linolenic Acid; **Peanuts**.)

Carbohydrates

In contrast with many legumes, lupin seeds are practically devoid of starch, and the major carbohydrates in mature seeds are oligosaccharides and nonstarchy polysaccharides, particularly from the cell wall structure. The oligosaccharides found in the cotyledons are sucrose and nondigestible galactosides of the raffinose family. The main galactosides present in lupin seeds are raffinose (degree of polymerization, DP, of 3), stachyose (DP of 4) and verbascose (DP of 5) (Table 5). These oligosaccharides are derived from sucrose, with increasing numbers of α-galactosyl units linked to C6 of glucose. They are present in most legumes and are not digested by monogastric animals and hence contribute to the flatulent property of legumes. More recently, nondigestible sugars have been used as agents for fermentation in the intestines, thus helping to maintain and regulate the microbial intestinal flora. In this aspect, lupin may be an important row material for production of oligosaccharides of the raffinose family to be used as functional food ingredients. The polysaccharides are present as major components of the cell wall surrounding the cytosol. Only a small number (1%) of polysaccharides are water-soluble, and, together with oligosaccharides, they are important as a source of energy for seed germination. The amount of

Table 4 Fatty acid composition of lupin seeds in comparison with soya beans

Legume	*Percentage of total fatty acids*[a]													
	14:0	*16:0*	*16:1*	*18:0*	*18:1*	*18:2*	*18:3*	*20:0*	*22:0*	*22:1*	*24:0*	*SFA*[b]	*MUFA*[b]	*PUFA*[b]
L. albus	0.2	7.2	0.5	2.1	48.6	19.9	13.0	0.8	4.4	2.2	1.2	16	51	33
L. angustifolius	0.3	11.0	0.2	6.2	33.5	39.4	4.6	0.7	2.6	0.2	0.5	21	34	44
L. luteus	0.3	5.5	0.2	2.8	24.3	43.9	10.6	2.4	7.0	1.1	1.0	19	26	55
L. mutabilis	0.2	11.5	0.5	9.9	45.7	27.3	2.3	0.9	1.2	tr	0.3	24	46	30
Glycine max (soya bean)[c]	nd	10.0	0.2	4.0	25.0	52.0	7.4	nd	tr	tr	tr	10	26	64

[a]Fatty acids: 14:0, myristic; 16:0, palmitic; 16:1, palmitoleic: 18:0, stearic; 18:1, oleic; 18:2, linoleic; 18:3, linolenic; 20:0, arachidic; 22:0, behenic; 22:1, erucic; 24:0, lignoceric.

[b]SFA, total saturated fatty acids; MUFA, total monounsaturated fatty acids; PUFA, total polyunsaturated fatty acids; nd, not determined.

[c]Adapted from Ayorinde FO, Garvin K and Saeed K (2000) Determination of the fatty acid composition of saponified vegetable oils using MALDI$_1$-TOFMS. *Rapid Communications in Mass Spectrometry* 14: 608–615. Other data adapted from Muzquiz M, Burbano C, Bouthelier V *et al*. (1982) Estudio de los elementos essenciales de destinadas variedades de cinco especies del Genus Lupinus cultivadas y espontaneas de la Peninsula Iberica. In: Proceedings of the 2nd International Lupin Conference, pp. 173–181. (see Table 2).

tr, traces.

Table 5 Oligosaccharide composition of seeds of different lupin species[a]

Species	*Sucrose*	*Raffinose*	*Stachyose*	*Verbascose*	*Gelactosides*
L. albus	2.9	1.0	6.6	1.1	8.7
L. angustifolilus	3.4	1.5	5.2	2.0	8.7
L. consentinii	2.6	0.9	4.9	0.9	6.7
L. hispanicus	0.7	0.9	6.6	1.8	9.3
L. huteus	1.7	1.2	4.9	4.1	10.2
L. mutabilis	2.4	2.5	8.5	1.1	12.1

[a]Results expressed in grams per 100 g of dry matter.

polysaccharides in lupin seeds is dependent on the species, with high contents found in *L. albus* and *L. angustifolius* and the lowest in *L. mutabilis*, due to its thinner seed hull. Due to the polysaccharide composition, lupins are extraordinarily rich sources of fiber. Crude fiber data are presented in Table 2, but even higher figures will be encountered if dietary fiber is considered. Lupin carbohydrates are important sources of energy only for ruminants, but due to their characteristics, they may provide useful material to produce a special high-fiber food for humans. (*See* **Carbohydrates**: Classification and Properties; **Dietary Fiber**: Properties and Sources.)

Vitamins and Minerals

The vitamin pattern of lupin seeds is somewhat similar to other legumes. Lupin seeds are good sources of vitamins of the B group, particularly niacin, with levels above 4 mg per 100 g. The thiamin content is about 0.5 mg per 100 g, and that of riboflavin about 0.4 mg per 100 g. Carotenoids and tocopherols are present, with the former being mainly responsible for the color of the oil fraction. The mineral composition is similar to other legumes in relation to the major elements, except for calcium, which is low in all species. However, some differences are found in respect to trace elements. Manganese is found in lupins in unusually high amounts, especially in *L. albus*. In some varieties of this species, it may reach values of 83.5–143 mg per 100 g, and this must be taken into account when it is used in daily animal or human diet. Nevertheless, common figures are in the range 2.5–38 mg per 100 g. Lupin seed is also a good source of iron and zinc, with values in the range 2.5–14 and 3–18 mg per 100 g, respectively. Refer to individual vitamins and minerals.

Bioactive compounds

Some components found in natural products may present biological activity but no nutritional properties. In fact, some of them may be undesirable for human and animal nutrition. In lupin, some compounds with these characteristics have been reported. Compounds sometimes called 'antinutritional factors' are present in lupin seeds in the same range as in other legume seeds, except for trypsin inhibitor, which is virtually absent in the lupin. Allergenic proteins that may cause allergenic food reactions sometimes observed in children are less intense with lupins than with other legume seeds and cow's milk. However, because peanut allergies are quite common, further studies are needed to check some sort of peanut–lupin cross-allergy when lupin is used as an ingredient of mixed flour. Some relevant bioactive compounds found in lupin are the alkaloids, phytates, saponins, tannins, and flavonoids. (*See* **Allergens**; **Alkaloids**: Properties and Determination; **Phytic Acid**: Properties and Determination; **Plant Antinutritional Factors**: Characteristics; **Saponins**; **Trypsin Inhibitors**.)

Alkaloids

The major alkaloids present in lupin are from the quinolizidine family, although some gramine alkaloids may also be found in *L. luteus*. Quinolizidine alkaloids have received much attention, because they have a strong bitter taste and may be toxic in high doses. Lupanine, sparteine, lupinine, and some forms of hydroxylated lupanine are some relevant examples. They have a sedative effect on the central nervous system, with sparteine producing the strongest effect. The total alkaloid content in bitter lupins is usually in the range 1.0–4.5 g per 100 g. Intensive breeding work carried out, mainly by Sengbush in Germany, Gladstones in Australia, and Baer in Chile, have led to the establishment of new lupin varieties practically devoid of alkaloids. Presently, sweet varieties of *L. luteus, L. angustifolius, L. albus, L. mutabilis, L. consentinii*, and *L. atlanticus* are defined, of which the first three are commercially productive. However, bitter lupins are still largely used in some regions where the new sweet varieties are not well adapted. In those cases, the preliminary soaking and washing procedure before consumption is essential. (*See* **Alkaloids**: Properties and Determination.)

Phytate

Phytate is sometimes considered an antinutritional factor because it is implicated with the impaired absorption of minerals. It is found in lupin seeds roughly

in the same concentration range as in other pulses, but it is usually lower than that found in soya beans. The average amount found in lupin seed is around 0.8 g per 100 g. The phytate:zinc molar ratio, which may be an indication of zinc bioavailability, is generally lower than in other legume seeds, making lupin a better dietary source of this mineral. The enzymatic degradation of phytate in the digestive tract or in food products submitted to special processing may lead to the formation of inositol phosphates with different degrees of phosphorylation. Some of these derived components may be detected as phytates by nonspecific methods, although they may not present mineral chelation activity. Consequently, more comprehensive information on inositol phosphate composition of different food items is becoming more important to assess the real role of food phytate in the daily diet. More recently, some attention has been directed to the desirable properties of vegetable phytates, particularly due to its antioxidative characteristics, which may be beneficial to counteract free-radical activity. Other biological properties may be derived from the lower inositol phosphates since some of them are involved in cell-signaling mechanisms.

Saponins

Saponins are compounds formed by triterpenoids or steroidal aglycones and a carbohydrate moiety by ester or ether linkages. They are present in different classes of plants, particularly in legumes, roots, and some medicinal herbs. Their presence in food products has been considered to be deleterious if consumed frequently. They are toxic to fish and promote retardation of growth in livestock and laboratory animals. They may also produce erythrocyte lysis *in vitro* and may alter intestinal epithelium function making the mucosa more permeable. Consequently, their continue use in the diet may jeopardize the process of nutrient absorption. Conversely, it has been claimed that they can also be beneficial since they show the ability to lower plasma cholesterol, they have anticancer activity, and they may act as an inhibitor of viral replication. It is not yet clear, though, whether the net effect in the diet would be negative. Some lupin species such as *L. luteus, L. mutabilis*, and *L. angustifolius* may present a saponin content from 57 to 470 mg kg^{-1}, but they are not present in *L. albus*. However, these figures are still low compared with soyabeans, which present values in the range 2000–5000 mg kg^{-1}.

Tannins

Tannins are complex polyphenolic substances found in plants, particularly pulses, with the property to precipitate proteins in aqueous medium. They interact with one or more protein molecules forming large cross-linked complexes that are insoluble in water. This property makes food tannins undesirable since they will make part of the dietetic protein indigestible. There is a wide variation on the content of tannins in legume seeds with higher values being found in faba bean and in peas. Lupin seeds present relatively low values of tannins in the range 0.2–0.5 g%. No correlation has been found between the tannin content and the bitter taste of some lupin seeds, and the impact of tannin from lupin seed in the diet has not yet been demonstrated both in humans or animals. However, the level of tannin in *L. angustifolius* has been considered to be low enough for use in pig diets without any kind of problem.

Flavonoids

Flavonoids are a class of phenolic compounds widely distributed in plants. Quercetin and rutin are among the most largely found flavonoids in a great variety of fruits and vegetables, including tea, coffee, and other grains. As it has been observed with other biological active nonnutrient components, flavonoids may promote desirable and nondesirable physiological effects in humans. The property of flavonoids to induce goitre has been suggested by studies using peanuts and millet as foods with millet flavonoids presenting a strong inhibition effect on thyroperoxidase activity. The healthy properties of flavonoids may be derived from their antioxidative characteristics as free-radical neutralizers. However, some more specific functions have been reported, including their effect on cancer prevention, antiinflammatory and antiviral activities, and their positive effect on capillary fragility and vascular protection. Data on the presence of flavonoids in lupin are scarce, and although it is not clear whether they are encountered in significant amounts in lupin seeds, different flavonoids have been reported recently in the plants of *L. luteus* and *L. albus*, including those with the aglycones apinegin, genistein, and kaempferol. (*See* **Phenolic Compounds**.)

Crop Diseases

Fungal and virus diseases are the most common forms of lupin crop contamination such as anthracnose, fusariose, and phomopsis. The first is the main threat for lupin cultivation in Central and South America, whereas the latter is the cause of a disease called lupinosis in animals fed from the infected plant. As it has been demonstrated in Australia, this problem may be overcome with an appropriate crop management and plant breeding for resistance. Necrotic and nonnecrotic strains of mosaic viruses may also be responsible for crop damage, but to a lesser extent than fungal diseases.

Food Uses

At the present time, the main applications of lupins are their utilization as green manure and as animal feed. Due to the ability of lupin plants to fix nitrogen and to make insoluble phosphorus available in the soil, they have been used increasingly in the crop-rotation system as an efficient and less expensive means of fertilization. For animal feeding, the whole lupin plant has been used as pasture or forage, or the dry seeds have been used as a direct soya bean substitute, or in feeding formulae for pigs, sheep, chickens, dairy cattle, and other livestock. It has been recommended that no more than 10–15% of bitter lupins should be used in pig rations as pigs seem to be more sensitive to alkaloids than chickens. As mentioned earlier, lupin seeds have long been an available protein source for human nutrition. It was a common procedure to submit the grains to prolonged washing to remove the bitter taste, and this washing process is still used, particularly in some regions of South America, where bitter lupin seeds are mainly used. Although the washing process may increase the total cost of lupin products, the water treatment will in fact wash out not only the bitter alkaloids but also the oligosaccharides responsible for the flatulence of legumes. In addition, the end product will show increased relative amounts of protein and oil, on a dry matter basis, due to differences in component solubilities. It is a common procedure in the Andean regions to use leached and cooked bitter lupin seeds directly for consumption as snacks, in soups, salads, and stews. Alternatively, they may be dried and/or milled for future use in several dishes, including baked products. Lupin flour may be used for cereal protein enrichment since lupins and cereals have complementary amino acid compositions. Bread or other bakery products have been made successfully using lupin flour as an additive in the proportion of 10–20%. Many attempts have been made to use lupin seeds as direct substitutes for soya beans, e.g., in fermented products, soya milk, or soya bean protein isolate. However, lupin protein does not have exactly the same functional properties as soya bean protein, being less stable to heat and forming a milk-like water suspension. Nevertheless, lupin protein may be added to soya products without major modifications to the end product. It is apparent that investigation is still needed to develop a lupin-specific technology and new genuine lupin products for human nutrition.

The Potential of Lupin

The world lupin seed productions is at present mainly concentrated in Australia, which has produced 1.38 million tonnes in 1993 and around 2 million tonnes in the year 2000. This represents 2–4% of the world pulse production. Other lupin producers are the countries of the former USSR, Poland and Germany in Europe, Chile, Bolivia, and Peru in South America, and also South Africa. However, the lupin seed shows a real potential for the grain world market due to its nutritional and agricultural properties. The high alkaloid content that used to be one of the major limitations for a wider application of the lupin seed has been gradually overcome by the introduction of commercial sweet varieties, mainly of *L. albus* and *L. angustifolius*. Although a new sweet variety of *L. mutabilis* is already available, work is still needed to establish a productive line with increased seed size and protein content. Besides the improvement in lupin nutritional qualities, research in specific areas such as plant breeding and biotechnology, crop protection and food lupin technology are vital for lupin to gain a stronger foothold in the international market, with a substantial increase in Europe and South America production. To consolidate this status, an appropriate international price policy together with a strong marketing strategy is mandatory. (*See* **Vegetable Oils:** Dietary Importance.)

See also: **Alkaloids**: Properties and Determination; **Allergens**; **Carbohydrates**: Classification and Properties; **Dietary Fiber**: Properties and Sources; **Essential Fatty Acids**; **Fatty Acids**: Properties; **Food Intolerance**: Food Allergies; **Peanuts**; **Phenolic Compounds**; **Phosphorus**: Properties and Determination; **Phytic Acid**: Properties and Determination; **Plant Antinutritional Factors**: Characteristics; **Protein**: Food Sources; **Saponins**; **Trypsin Inhibitors**

Further Reading

Cowling WA, Buirchel BJ and Tapia ME (1998) *Lupin. Lupinus L. Promoting the Conservation and Use of Underutilized and Neglected Crops. 23*. Rome: Institute of Plant Genetics and Crop Plant Genetic Research, Gatersleben/International Plant Genetic Resources Institute.

DuPont MS, Muzquiz M, Estrella I, Fenwick GR and Price KR (1994) Relationship between the sensory properties of lupin seed with alkaloid and tannin content. *Journal of the Science of Food and Agriculture* 65: 95–100.

Fletwood JG and Hudson BJF (1982) Lupinseed – a new source of edible oil. *Journal of Food Technology* 17: 11–17.

Franski R, Bednarek P, Wojtaszek P and Stobiecki M (1999) Identification of flavonoid diglycosides in yellow lupin with mass spectrometric techniques. *Journal of Mass Spectrometry* 34: 486–495.

Gladstones JS, Atkins CA and Hamblin J (eds) (1998) *Lupins as Crop Plants: Biology, Production and Utilization*. USA: Oxford University Press.

Gross R (1982) *El Cultivo y la Utilization del Tarwi. Estudio FAO, No. 36*. Rome: Food and Agriculture Organization.

Hill GD (1977) The composition and nutritive value of lupin seed. *Nutrition Abstracts and Reviews, B: Livestock Feeds and Feeding* 47: 511–529.

Trugo LC, Farah A and Trugo NMF (1993) Germination and debittering of lupin seeds reduce α-galactosides and intestinal carbohydrate fermentation in humans. *Journal of Food Science* 58: 627–630.

von Baer E (1994) Lupin development policy in South America. In: Neves Martins JM and Beirão da Costa ML (eds) *Advances in Lupin Research*, pp. 577–580. Lisbon: ISA Press.

van Barneveld RJ (1999) Understanding the nutritional chemistry of lupin seed to improve livestock production efficiency. *Nutrition Research Reviews* 12: 203–230.

Lychees *See* **Fruits of Tropical Climates**: Commercial and Dietary Importance; Fruits of the Sapindaceae; Fruits of the Sapotaceae; Lesser-known Fruits of Africa; Fruits of Central and South America; Lesser-known Fruits of Asia

M

Mackerel *See* **Fish**: Introduction; Catching and Handling; Fish as Food; Demersal Species of Temperate Climates; Pelagic Species of Temperate Climates; Tuna and Tuna-like Fish of Tropical Climates; Demersal Species of Tropical Climates; Pelagic Species of Tropical Climates; Important Elasmobranch Species; Processing; Miscellaneous Fish Products; Spoilage of Seafood; Dietary Importance of Fish and Shellfish; **Fish Farming**; **Fish Meal**

MACROBIOTIC DIETS

J Dwyer, Jean Mayer US Department of Agriculture, Human Nutrition Research Center on Aging at Tufts University, Boston, MA, USA

Definition

Macrobiotics is a philosophical system that was founded by a Japanese named Georges Ohsawa and popularized in the USA by Mishio Kushi. Macrobiotics advocates a diet that is virtually devoid of animal foods because of the belief that foods biologically far away from humans are better for them, and that these foods should form the basis of the diet to prolong life and health. Animal foods, if they are used at all, are used in very small amounts as condiments rather than in main course portions. Ohsawa believed that all disease could be cured by a proper therapy that consisted of balancing the forces of Yin and Yang. The optimal macrobiotic diet is viewed by practitioners as one that balances Yin and Yang foods, as defined by their activity characteristics in the universe, supposedly opposing and complementing each other. This dualistic Yin–Yang classification does not correspond to food grouping systems based on nutrient composition. Extensive chewing of food is recommended to enhance the strengthening Yang properties of the diet. It is important to recognize that macrobiotics is a life view and philosophy that involves far more than diet alone. The philosophy views disease as resulting from Yin–Yang imbalances that can be restored by an ideal diet.

History

In the 1960s and early 1970s, an extreme form of the macrobiotic diet, the Zen macrobiotic diet, gained some popularity. The final stage of an elaborate 10-stage dietary regimen consisted of brown rice, salt, and fluids such as herbal teas, and was viewed as bringing not only health but also spiritual enlightenment. Michio Kushi revised the Ohsawa diet and emphasized less extreme patterns in the 1970s. The macrobiotic diet, as currently practiced, although it may have nutritional deficiencies, is less restrictive. Nevertheless, for individuals who are ill, especially with diseases such as cancers of the gastrointestinal tract, such a high bulk diet may cause, rather than alleviate, symptoms. Macrobiotic leaders often complain that there is a bias in the scientific literature against macrobiotics. However, evidence that the macrobiotic diet has a place in the treatment of cancers is not available. Moreover, the proponents of macrobiotic diets claim that the negative health effects that were associated with the diet in the past are no longer present. However, population-based studies are available that indicate ill effects in infants and young children, and similar studies refuting them are not available, testifying that this is no longer the case. This is surprising, since several individuals trained at the doctoral level in nutritional sciences are adherents to these regimens.

Composition of Diet

The macrobiotic diet is similar nutritionally to a vegan–vegetarian regime with additional prohibitions against sugar, alcohol, and processed foods. Most of the energy comes from cereal staples, including rice, soy, and aduke and other beans. This basic diet is supplemented by soybeans, vegetables, some fruits, nuts, seeds, and very small amounts of fish. The diets prescribed in writings by the various authorities and that are actually eaten probably vary a good

deal from eater to eater, however. (*See* **Vegetarian Diets**.)

Dietary planning and menu selection are based on combinations of foods that are thought to be philosophically appropriate according to macrobiotic criteria but may not be so from the nutritional standpoint. Certain food groups, such as dairy and meats, are avoided. Milk and milk products are major sources in American diets of calcium, phosphorus, magnesium, vitamin D, riboflavin, and high-quality protein. Meat, fish, poultry, eggs, and beans are good sources of high-quality protein, ion, zinc, copper, and many of the B-complex vitamins. Without careful planning, when these food groups are avoided, it is difficult to meet recommended levels of energy, vitamins B_{12} and D, iron, calcium, zinc, and riboflavin. Dietary deficiencies result when appropriate substitutions of other foods rich in these nutrients are not made. The extent to which dietary deficiencies arise varies by life stage, type of diet adopted, and the eater's health status.

Lifestyles of Macrobiotic Practitioners

Macrobiotic beliefs affect ways of living that influence health and go beyond diet. Macrobiotics practitioners have some lifestyle habits that promote health, such as abstinence or moderation in the use of alcohol, nonuse of illicit drugs, and a physically active life with adequate rest. However, some macrobiotics smoke, and others eschew anticipatory guidance from conventional health professionals, avoid or delay the use of preventive measures such as immunizations, and refuse to use vitamin mineral supplements and prescription drugs. Some delay seeking medical treatment for some illnesses until they are quite ill, thus increasing the risks of secondary malnutrition. These attitudes are at variance with vegetarians of other types. (*See* **Malnutrition**: Malnutrition in Developed Countries; **Vegetarian Diets**.)

Health Effects of Macrobiotic Diets over the Life Cycle

Acceptable and appropriate diets must fulfill the dietary reference intakes and the recommendations of other expert groups on dietary balance, variety, and moderation. Finally, they must be developmentally appropriate. Table 1 presents a summary of the risks and benefits of macrobiotic diets. Generally, macrobiotics breast-feed their children through the first year of life and introduce some complementary feeding after 6 months. However, macrobiotic diets are very bulky for weanlings. In addition, a source of vitamin D should be provided after 6 months of age. It is difficult to make the diet nutritionally adequate and follow the pronouncements of macrobiotic leaders. Use of at least 20–25 g of dietary fat, and at least 100–150 g of fatty fishes, which are good sources of vitamin D and also provide some vitamin B_{12}, improve the diets of weanlings in more favorable directions. At least 150–250 g of milk and milk products per day are advised to assure adequate intakes of calcium in early childhood. Alternatively, soy milk that is fortified with vitamins B_{12} and D and calcium, highly fortified cereals, or vitamin mineral supplements may be used as sources of nutrients rich in the milk group. During later infancy and weaning, iron intakes may also be unsatisfactory, and iron-deficiency anemia may result. Iron-fortified cereals, plant foods high in iron such as iron-fortified soy formula, and iron-fortified cereals eaten with ascorbic acid-rich foods enhance iron bioavailabiliy and decrease the risks of iron-deficiency anemia. The macrobiotic diet also increases the risks of dietary deficiencies at other times of life when nutrient needs are particularly high. Similar adaptations of the diet to increase intakes of essential nutrients are also prudent during pregnancy and lactation, at adolescence, and during recovery from illness. During pregnancy, vegan-vegetarians and some macrobiotics who eat virtually no animal foods at all may have very low stores of DHA and EPA, and of vitamin B_{12}. Weight gain during pregnancy may not be satisfactory either, increasing the risks of low birth weight.

During the past 25 years, adherents to the macrobiotic philosophy have had a higher prevalence of dietary deficiency disease than other vegetarians. The reasons for this are probably several, and include the restrictions on consumption and sources of nutrients (prohibitions against vitamin mineral supplements and fortified foods, other processed foods) over and above those involving a vegan diet, attitudes toward health and use of conventional medical care, and the difficulties of fitting recommendations based on philosophical considerations with nutrient intakes based on modern knowledge of nutrition.

Most of the positive aspects of macrobiotic diets associated with health benefits are because they are not excessive. Among them are modest caloric intakes, very low intakes of saturated fat, cholesterol, alcohol, and caffeine, relatively low intakes of refined carbohydrates, and high intakes of dietary fiber. The health advantages associated with the macrobiotic diet are due only in part to dietary composition. Other positive lifestyle characteristics of the macrobiotic lifestyle are also health-promoting. Macrobiotics generally are lean rather than obese, tend to eschew alcohol and illicit drugs, and lead orderly lives. The macrobiotic dietary pattern provides no unique

Table 1 Risks and benefits of macrobiotic diets

	Infancy < 6 months	*Infancy > 6 months*	*Weaning*	*Preschool*	*School*	*Adolescent*	*Adult*	*Pregnancy*	*Lactation*	*Illness*
Risks										
Emaciation, protein-calorie malnutrition	×	×	×							
Slowed growth	×	×	×	×	×	×	×	×	×	×
Low stores/intakes of DHA, EPA	×	×	×					×	×	
Iron-deficiency anemia	×	×	×	×	×	×	×	×	×	×
Anemia due to vitamin B_{12} deficiency	×	×	×	×	×	×	×	×	×	×
Low intakes of folic acid							×	×		
Vitamin D deficiency rickets		×	×	×	×					
Osteomalacia						×		×		
Low dietary calcium intake	×	×	×	×	×	×	×	×	×	×
Low intakes of zinc, B_6	×	×	×	×	×	×	×	×	×	×
Benefits										
Long lactation and breast-feeding	×	×	×						×	
Decreased obesity					×	×	×	×	×	
Decreased alcohol abuse					×	×	×	×	×	×
Low serum lipids					×	×	×			
Low blood pressure					×	×	×			
Good laxation		×	×	×	×	×	×			
Decreased incidence of type 2 diabetes				×	×	×	×	×		
Decreased incidence of diverticular disease					×	×	×			
Decreased incidence colon cancer						×	×			
Decreased incidence of breast and prostate cancer						×	×			
Decreased incidence of endometrial cancer						×	×			
Decreased incidence of cervical cancer										

health advantages over those accruing to other vegetarian or vegan patterns, however.

Environmental factors are the major reasons why the nutritional status of macrobiotics differs from that of other vegetarians or omnivores. However, macrobiotics are not exempt from genetic disorders that require nutritional treatment either, such as vitamin D-dependent rickets, or polymorphisms in the vitamin D receptor that are associated with differences in bone mineral density and perhaps calcium absorption. They are also affected by more subtle variations in genetic endowment that contribute to the variability in nutrient requirements. For example, a genetic polymorphism has been found that adversely affects homocysteine concentrations in individuals with a relatively poor float status. Macrobiotics with such polymorphisms may be the individuals who fall ill first when diets are inadequate (See Table 1).

Nutritional Value of Macrobiotic Diets

Energy

Macrobiotics urges moderation in consumption. The regimen is low in fat, high in bulk, and high in fiber, and thus it tends to be low in energy. Macrobiotic practitioners are usually lean, in part because their energy intakes are low and in part because they lead physically active lives. Some malabsorption may occur. The risks of low energy intakes are most evident among macrobiotic infants who are not breast-fed weanlings, and young children. The practice of feeding a very-low-fat-cereal-based infant food to infants, called 'kohkoh', is no longer common; it was associated with frank protein calorie malnutrition in many infants. Macrobiotic children are usually smaller, lighter, and leaner than their peers. Some catch-up growth usually occurs in the preschool

and school years. (*See* **Energy**: Measurement of Food Energy.)

Protein

Human beings require nine essential amino acids that they cannot synthesize, the nonessential amino acids, and additional nitrogen to fulfill their protein needs. The amino acid composition of protein, the amount of protein, the bioavailability of protein and the availability of the protein for building tissue instead of being burned for energy are therefore all important. (*See* **Amino Acids**: Metabolism; **Protein**: Requirements; Quality.)

A major concern is that the individual be in energy balance rather than in a hypocaloric state, since, under such circumstances, protein will be used for energy. The macrobiotic regimen potentially poses both protein quality and quantity problems. On all but the most severe macrobiotic regimes, enough protein is present to be adequate if sufficient energy is provided. With regard to protein quality, legumes are moderately high in protein quality. When they are consumed along with rice and other plant proteins, the biological value of the protein is adequate. Animal proteins lack fewer amino acids than do plant proteins, but macrobiotic diets are virtually devoid of animal flesh, eggs, milk, and poultry, all of which provide relatively complete proteins and all the amino acids needed for building bodily protein. Fish and seafood are consumed by macrobiotics in small amounts, and these proteins are relatively complete. Plant tissues are somewhat more variable in the extent to which the nitrogen present in them is found in protein, ranging from a high of about 95% in seeds down to 50% in roots such as potatoes, carrots, and cassava. The nonprotein nitrogen is in the form of peptides and free amino acids. The amino acid composition of different plant proteins can complement each other, making up their various deficits. Thus, the aggregate of individual plant proteins provides a mixture that matches body needs. The lack of most animal protein does not necessarily mean that protein nutritional status must be compromised. Macrobiotics consume legumes such as soya beans and aduke beans, which are high in the essential amino acids lysine and threonine, and soya beans are also high in tryptophan. However legumes are limited in methionine, cystine, and (except for soya beans) tryptophan. Grain proteins are high in methionine, cystine and tryptophan (with the exception of cornmeal and rye, which are low in tryptophan). Moreover, grains are low in the essential amino acids lysine and isoleucine, although these are high in legumes. Thus, combinations of grains and legumes complement each other and together provide for the amino acid needs of the body if eaten in sufficient amounts. Mixtures of grains, milk or eggs, and legumes with seeds or nuts also complement each other well. Most nuts and seeds are high in cystine, methionine, and tryptophan, and low in lysine and isoleucine. The only exception is peanuts, which are low in both methionine and threonine. Other plant proteins are low in these amino acids, whereas they are relatively high in lysine and tryptophan. Thus, nuts, seeds, and legumes complement other plant proteins well, as do grains and legumes. Small amounts of animal foods such as seafood, milk products, and eggs also complement the amino acid profile of plant proteins well. Therefore, when a variety of complementary plant proteins are consumed in sufficient quantity, and energy is provided in sufficient amounts, there is little reason for concern. However, on the most rigorous macrobiotic diets, at least in the past, energy intakes were sometimes inadequate, and when only one or two plant protein sources were relied upon exclusively, protein calorie malnutrition sometimes ensued. These problems were reported more commonly in the 1960s and 1970s than later. Practices may have changed, but recent rigorous surveys of the nutritional status of macrobiotics are not available. Caution is warranted for the ill, who often consume hypocaloric diets, for young infants who are growing rapidly and have been weaned from the breast who may not be able to consume sufficient energy on exclusively plant diets, and for pregnant women and others who have especially high protein needs for growth. (*See* **Protein**: Digestion and Absorption of Protein and Nitrogen Balance; Deficiency.)

Fat and Cholesterol

Macrobiotic diets are low in fat (e.g., < 25% of energy, and often as little as 15%), and especially saturated fat and cholesterol. Those who consume such diets often have low total and low-density lipoprotein cholesterol levels, with decreased dietary risks of coronary artery disease. Macrobiotics' high intakes of soluble fiber and their relatively low weights also cause low serum cholesterol levels and lower the risk of coronary artery disease.

When diets are below approximately 15% of food energy, special care must be taken to ensure that intakes of essential fatty acids are met. At least 3% of energy should be from ω-6 fatty acids and 2% from ω-3 fatty acids. Both of the essential fatty acids, linoleic acid, an ω-6 fatty acid, and α-linolenic acid, an ω-3 fatty acid, are required by human beings. Linoleic acid is found in seeds, nuts and grains. α-Linolenic acid is found in the green leaves of plants, and in phytoplankton and algae, in certain seeds, nuts and legumes, such as flax, canola, walnuts, hazelnuts, and

soy. These can be converted into more highly unsaturated fatty acids, with linoleic acid producing arachidonic acid, and α-linolenic acid being converted to eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA). Arachidonic acid is found in animal foods such as meat, poultry, and eggs. EPA and DHA are largely found in fish and seafood. Arachidonic acid and EPA serve as precursors for the eicosanoids. Recommended intakes for these polyunsaturated fatty acids range from 3 to 10% of total energy intakes. Vegan vegetarians have no direct sources of long-chain ω-3 fatty acids EPA and DHA in their diets, and thus must convert α-linolenic acid to them. There is concern that pregnant women who are vegan-vegetarians, or macrobiotics who consume little or no fish or other animal foods, may not obtain enough of these fatty acids, especially during pregnancy and in early infancy, especially if the infants are premature, and their conversion capacity of α-linolenic acid to DHA is very limited. Such individuals may need DHA supplements, either from fish oils or from cultured microalgae. However, they should only be dispensed under a physician's direction, since they are also anticoagulants. ω-3 fatty acids are rich in flaxseeds and flaxseed oil, hemp seed oil, purslane, a dark green leafy vegetable, canola, walnut, and soy oil. Flaxseed oil is also rich in α-linolenic acid.

Carbohydrate and Fiber

Raw fruits, especially berries, vegetables, legumes such as beans and peas, nuts, seeds, whole-grain products, and seaweed provide relatively high amounts of dietary fiber in the macrobiotic diet, helping to promote laxation and perhaps lessening symptoms of asymptomatic diverticular disease. It is possible, but not yet proven, that high intakes of dietary fiber reduce the risks of colon cancer; there is no evidence that macrobiotic diets are especially efficacious in these respects. The macrobiotic diet is high in soluble sources of dietary fiber, such as legumes, whole grain rice, barley, oats, and raw fruits and vegetables. Soluble fiber may enhance glycemic control and also lower serum cholesterol, contributing the low total and low-density-lipoprotein cholesterol levels often evident in macrobiotics. Fiber increases the dietary bulk and lowers the energy density per gram of food consumed. The ill effects of dietary fiber arise if very large amounts are fed to young infants and weanlings under 2 years of age. Macrobiotics sieve cereals, pulses, and vegetables that are fed to infants to increase their digestibility, but fiber intakes and bulk may still be rather high, and the nutrient density of such feeds may still be quite low. The partial replacement of whole-grain cereals with more highly refined cereals that are lower in fiber increases energy intakes and decreases bulk in feeding infants.

Macrobiotics' intakes of refined sugar are very low, but other fermentable carbohydrates, such as sugars found in fruits and starches, are consumed, so the risk of dental caries is still present, especially if these are consumed in forms that are retained in the oral cavity, and oral hygiene is poor. (*See* **Carbohydrates**: Requirements and Dietary Importance; **Glucose**: Glucose Tolerance and the Glycemic (Glycaemic) Index.)

Vitamins

Vitamins B_{12} and D may be in particularly short supply in macrobiotic diets fed to rapidly growing infants and children.

Vitamin B_{12} Reliable plant sources of vitamin B_{12} are vitamin B_{12}-fortified soyamilk and formulas, yeast grown on, and mixed with, vitamin B_{12}-enriched media, highly fortified breakfast cereals, and vitamin B_{12} supplements. Vitamin B_{12}-containing foods that are sometimes acceptable to those who practice macrobiotics include fish such as sardines, pilchard, mackerel, herring, salmon, and tuna. In amounts of about an ounce a day, they may be sufficient to avoid deficiencies in young children.

Animal foods have a high content of vitamin B_{12} because they ingest vitamin B_{12}-containing microorganisms or because the microflora in their guts synthesize the vitamin. The richest sources of vitamin B_{12} provide over 19 μg per 100 g. They are lamb or beef organ meats such as kidney and liver, bivalves such as clams and oysters, which extract many microorganisms containing vitamin B_{12} from sea water, soyamilk that is fortified with vitamin B_{12}, and vitamin supplements. Other types of red meats, eggs, and milk products are also good sources. Macrobiotics rarely eat large amounts of any of these products on a regular basis. Therefore, inadequate intakes of vitamin B_{12} may be a problem. Moderately good sources, providing 3–10 μg per 100 g of vitamin B_{12} include seafood, such as lobster, scallops, haddock, tuna, swordfish, and flounder, and egg yolks, fermented cheeses, such as Camembert or Limburger, and nonfat milk solids. Most other red meats and fish provide less than this amount. Grains, fruits, and vegetables are virtually devoid of B_{12}, except for tiny amounts present in root nodules of some legumes. Seaweed is sometimes contaminated by plankton rich in vitamin B_{12}, but these are unreliable sources of the vitamin. 'Sea vegetables' such as wakame, arame, and kombu and a seaweed known as spirulina contain very few corrinoids that are available to humans. Other plant foods such as certain fermented soya products like miso (which is often served as a soup) and tempeh

have not been tested for the bioavailablity of their corrinoids to humans, although they may test positive in microbiological assays for the vitamin. They cannot be relied upon to provide sufficient amounts of vitamin B_{12}. Small amounts of vitamin B_{12} may also be present in some plant foods, owing to contamination of food with insects or feces.

The major risks of vitamin B_{12} deficiency are in pregnant women, growing infants, and children of mothers who have followed a macrobiotic diet for many years. Infants may be born with low stores of the vitamin, are reported to exhibit biochemical indices of vitamin B_{12} deficiency, even when they are suckling, and are at risk of deficiency after they have been weaned from the breast to vegan-like diets. (*See* **Cobalamins**: Physiology; **Food Fortification**.)

Vitamin D Humans can synthesize vitamin D endogenously when the skin is exposed to ultraviolet irradiation, but those who live in northern latitudes, use sunscreen, or are confined indoors by illness may not receive enough stimulation to meet their needs. Supplies of the vitamin may pose problems for macrobiotics. The naturally occurring food sources rich in vitamin D are human milk, which provides low but adequate amounts of the vitamin for human infants in the first few months of life, egg yolks, liver, and some fatty fish such as eel, herring, sardines, and salmon. In the USA, milk and milk products are fortified with vitamin D at approximately $0.5\,\mathrm{IU\,g^{-1}}$ (0–0.01 $\mu\mathrm{g\,g^{-1}}$) wet weight, and margarine at about $5\,\mathrm{IU\,g^{-1}}$ (or $0.10\,\mu\mathrm{g\,g^{-1}}$), so that food products containing them will also contain some vitamin D. Vitamin D-fortified milk products are also available in many other Western countries. In the USA, highly fortified cereals are sometimes fortified with vitamin D, often at levels of $1.2\,\mathrm{IU\,g^{-1}}$ (0.02–0.04 μg) dry weight, as packaged. Cod liver oil and menhaden oil macrobiotics refuse to eat liver, egg yolk, butter, fluid milk, and dried skim milk powder. The natural sources of vitamin D they may be willing to eat are margarine and fatty fish, and vitamin D-fortified soyamilk and other plant foods. Vitamin D supplements are also available without prescription in the USA at levels not exceeding 5 μg (the current adequate intake) per capsule, or at higher levels by prescription.

Minerals

Certain mineral elements in the macrobiotic diet are also sometimes limited.

Calcium Macrobiotics rarely include foods from the dairy group in their usual diets. Some plants including kale, collards, dried peas, and beans provide considerable amounts of calcium. However, absorption of calcium from spinach, kale, collards, and Swiss chard is inhibited by the presence of oxalates that bind with calcium during digestion to form insoluble calcium oxalate, which cannot be absorbed. Whole grains contain phytates, substances that interfere with calcium and zinc. However, some of the calcium is available. Foods that may be acceptable to macrobiotics that contain calcium include bony fish such as sardines and foods fortified with calcium (fortified soy milk, fortified rice beverage, fortified juices, highly fortified breakfast cereals, and tofu precipitated with calcium sulfate), and these provide considerable amounts of calcium.

Iron Macrobiotics exclude all heme iron sources, because they do not eat meat. Even lacto-ovovegetarians have a very much lower iron absorption than do omnivores. The bioavailabiliy is probably about 10% compared with about twice that in omnivores, so this is a potential risk factor. However, many other factors affect iron absorption, and in many studies, iron status does not seem to be adversely affected. It is possible to estimate bioavailabiliy in the diet of dietary iron. The amount of ascorbic acid and presence of other iron absorption enhancers as well as the presence of inhibitors must be measured. Foods high in iron include peas, beans, and highly iron-fortified breakfast cereals. During pregnancy, iron supplements are required, because the iron needs are so high.

Zinc The macrobiotic diet is low in meat and low in other foods that are high in zinc. Milk products are a good source of zinc for those who consume them. Zinc absorption from plant foods is poor when the foods contain phytate and oxalate inhibitors. Macrobiotic diets are high in foods high in the zinc inhibitor phytic acid (e.g., soy products, beans, lentils, peas, nuts, and whole grains). Good plant sources of zinc are zinc-fortified cereals, and yeast-fermented foods that reduce the phytic acid content of whole-grain breads, rather than refined white bread. Zinc supplements are also good sources of zinc that may be acceptable to some macrobiotics. Fermented soyfoods, like tempeh and miso, and sprouted legumes may have more bioavailable zinc than unfermented foods like tofu and soymilk.

Conclusion

Many of the aspects of the macrobiotic diet that convey risks to health can be easily corrected without losing the positive aspects of the regimen or violating philosophical principles. Vegetarian or vegan diets that are carefully planned to include modern scientific

knowledge of nutrition are associated with good health. This is also possible for macrobiotic diets, although, in reality, it does not always occur.

See also: **Amino Acids**: Metabolism; **Carbohydrates**: Requirements and Dietary Importance; **Cobalamins**: Physiology; **Energy**: Measurement of Food Energy; **Food Fortification**; **Glucose**: Glucose Tolerance and the Glycemic (Glycaemic) Index; **Malnutrition**: Malnutrition in Developed Countries; **Protein**: Requirements; Quality; Digestion and Absorption of Protein and Nitrogen Balance; Deficiency; **Vegetarian Diets**

Further Reading

Abrams HL (2000) Vegetarianism: Another view. In: Kiple KF and Connee K (eds) *The Cambridge World History of Food*, vol. 2, pp. 1564–1573. Cambridge: Cambridge University Press.

Alexander DB and Mann J (1994) Nutrient intake and haematological status of vegetarians and age sex matched omnivores. *European Journal of Clinical Nutrition* 48: 538–546.

Anderson B, Gibson RS and Sabry JH (1981) The iron and zinc status of long-term vegetarian women. *American Journal of Clinical Nutrition* 34: 1042–1048.

Coulston A (1999) The role of dietary fats in plant based diets. *American Journal of Clinical Nutrition* 70: 512S–515S.

Crane M, Register D and Lukens R (1999) Cobalamin studies on 2 total vegetarian (vegan) families. *American Journal of Clinical Nutrition* 70: 626S–627S.

Dagnelie P (1991) Vitamin B-12 from algae appears not to be bioavailable. *American Journal of Clinical Nutrition* 53: 695–697.

Dagnelie P (1994) Macrobiotic nutrition and child health: results of a population based, mixed longitudinal cohort study in The Netherlands. *American Journal of Clinical Nutrition* 59: 1187S–1196S.

Dagnelie P, Vergote FJVRA, Van Staveren WA *et al*. (1990) High prevalence of rickets in infants on macrobiotic diets. *American Journal of Clinical Nutrition* 51: 202–208.

Donovan U and Givson RS (1995) Iron and zinc status of young women aged 14 to 19 years consuming vegetarian and omnivorous diets. *Journal of the American College of Nutrition* 14: 463–472.

Dwyer J and Jacobs C (1988) Vegetarian children: appropriate and inappropriate diets. *American Journal of Clinical Nutrition* 48: 811S–818S.

Dwyer J and Lowe F (1994) Nutritional risks of vegan diets to women and children: are they preventable? *Journal of Agricultural and Environmental Ethics* 7: 87–109.

Dwyer JT (1999) Convergence of plant-rich and plant-only diets. *American Journal of Clinical Nutrition* 70: 620S–622S.

Estes JW (2000) Food as medicine. In: Kiple KF and Connee K (eds) *The Cambridge World History of Food*, vol. 2, pp. 1534–1553. Cambridge: Cambridge University Press.

Haddad E, Berk LS, Kettering JD, Hubbard RW and Peters WR (1999) Dietary intake and biochemical, hematologic, and immune status of vegans compared with nonvegetarians. *American Journal of Clinical Nutrition* 70: 586S–593S.

Hallberg L and Huthen L (2000) Prediction of dietary iron absorption: an algorithm for calculating absorption and bioavailability of dietary iron. *American Journal of Clinical Nutrition* 71: 1147–1160.

Harland B and Oberleas D (1987) Phytate in foods. *World Review Nutrition and Dietetics* 52: 235–259.

Hunt J, Gallagher SK, Johnson LK and Lykken GI (1995) High versus low meat diets: effects on zinc absorption, iron status, and calcium, copper, iron, magnesium, manganese, nitrogen, phosphorus, and zinc balance in postmenopausal women. *American Journal of Clinical Nutrition* 62: 621–632.

Hun J, Matthys LA and Johnson LK (1998) Zinc absorption, mineral balance, and blood lipids in women consuming controlled lacto-ovovegetarian and omnivorous diets for 8 wk. *American Journal of Clinical Nutrition* 67: 421–430.

Hunt J and Roughhead ZK (1999) Nonheme iron absorption, fecal ferritin excretion, and blood indexes of iron status in women consuming controlled lactoovovegetarian diets for 8 wk. *American Journal of Clinical Nutrition* 69: 944–952.

Jacques P, Bostom AG, Williams RR *et al*. (1996) Relation between folate status, a common mutation in methylenetetrahydrofolate reductase, and plasma homocystine concentrations. *Circulation* 93: 7–9.

Key T, Fraser GE, Thorogood M *et al*. (1999) Mortality in vegetarians and nonvegetarians: detailed findings from a collaborative analysis of 5 prospective studies. *American Journal of Clinical Nutrition* 70: 516S–524S.

Kushi L, Meyer KA and Jacobs DR (1999) Cereals, legumes and chronic disease risk reduction: evidence from epidemiologic studies. *American Journal of Clinical Nutrition* 70: 451S–458S.

Ladizesky M, Lu Z, Oliveri B *et al*. (1995) Solar ultraviolet B radiation and photoproduction of vitamin D_3 in central and southern areas of Argentina. *Journal of Bone and Research* 10: 545–549.

Latta D and Liebman M (1984) The iron and zinc status of longterm vegetarian women. *American Journal of Clinical Nutrition* 34: 1042–1048.

Mann J, Appleby PN, Key TJ and Thorogood M (1997) Dietary determinants of ischaemic heart disease in health conscious individuals. *Heart* 78: 450–455.

Matsuoka L, Ide L, Wortsman J, MacLaughlin JA and Holick MF (1987) Sunscreens suppress cutaneous vitamin D_3 synthesis. *Journal of Clinical Endocrinology and Metabolism* 64: 1165–1168.

Messina M and Messina V (1996) *The Dietitian's Guide to Vegetarian Diets: Issues and Applications*. Gaithersburg, MD: Aspen.

Messina V and Burke KL (1997) Position of the American Dietetic Association: vegetarian diets. *Journal of the American Dietetic Association* 97: 1317–1321.

Miller D, Specker BL, Ho ML and Norman EJ (1990) Vitamin B_{12} status in a macrobiotic community. *American Journal of Clinical Nutrition* 53: 524–529.

Resnicow K, Barone J, Engle A *et al*. (1991) Diet and serum lipids in vegan vegetarians: a model for risk reduction. *Journal of the American Dietetic Association* 91: 447–453.

Sanders T (1999) Essential fatty acid requirements of vegetarians in pregnancy, lactation, and infancy. *American Journal of Clinical Nutrition* 70: 555S–559S.

Sanders T (1992) The influence of a vegetarian diet on the fatty acid composition of human milk and the essential fatty acid status of the infant. *Journal of Pediatrics* 120: S71–S77.

Sandstrom B, Arvidsson B, Cederblad A and Bjorn-Rasmussen E (1980) Zinc absorption from composite meals I. The significance of wheat extraction rate, zinc, calcium, and protein content in meals based on bread. *American Journal of Clinical Nutrition* 33: 739–745.

Simopoulos A (1999) Essential fatty acids in health and chronic disease. *American Journal of Clinical Nutrition* 70: 560S–569S.

Specker B, Miller D, Norman EJ, Greene H and Hayes KC (1988) Increased urinary methylmalonic acid excretion in breast fed infants of vegetarian mothers and identification of an acceptable dietary source of vitamin B_{12}. *American Journal of Clinical Nutrition* 47: 89–92.

Van den Berg I (1991) Vitamin B-12 content and bioavailability of spirulina and nor. *Journal of Nutritional Biochemistry* 2: 314–318.

Weaver C, Proulx WR and Heaney R (1999) Choices for achieving adequate dietary calcium with a vegetarian diet. *American Journal of Clinical Nutrition* 70: 543S–548S.

Webb A, Kline L and Holick MF (1988) Influence of season and latitude on the cutaneous synthesis of vitamin D_3: exposure to winter sunlight in Boston and Edmonton will not promote vitamin D_3 synthesis in human skin. *Journal of Clinical Endocrinology and Metabolism* 67: 373–378.

Whorton J (2000) Vegetarianism. In Kiple KF and Connee K (eds) *The Cambridge World History of Food*, vol. 2, pp. 1553–1565. Cambridge: Cambridge University Press.

Wood R and Fleet JC (1998) The genetics of osteoporosis: Vitamin D receptor polymorphsims. *Annual Review of Nutrition* 18: 233–258.

Worthingon-Roberts B, Breskin MW and Monsen ER (1988) Iron status of premenopausal women in a university community and its relationship to habitual dietary sources of protein. *American Journal of Clinical Nutrition* 47: 275–279.

MADEIRA AND RELATED PRODUCTS

R W Goswell, Stoke Bishop, Bristol, UK

Definition

Madeira is the name given to a number of related types of dessert wine originally developed in the islands of Madeira and Porto Santo. It owes much of its characteristic flavor to a period of maturation at a relatively elevated temperature. Similar products are produced elsewhere under other names. Nowadays, the designation 'Madeira' is rarely used for wines not produced on the island.

Legal Definition

The Madeira wine region is one of the oldest demarcated areas of Portugal. Since April 1979, the method of production of Madeira has been controlled by the newly created Madeira Wine Institute.

Vineyards

The soil is volcanic, of acid pH, and with a high organic content. Vineyards are typically small (2 ha or less), at altitudes of between 100 and 700 m, and southerly aspects are preferred.

The climate is temperate, humid, and often misty.

Planting, Training, and Pruning

Vines are trained on very low trellises or low wires. Planting density is high (7000 vines per ha). A variety of virus-free rootstocks, including R99, 1103P, and R110, are available to suit the various soils and climatic conditions.

The Guyot pruning system is most often used, leaving many buds and cropping heavily to give 10–15 t ha^{-1}.

Grape Varieties

Historically, the important (noble) varieties in Madeira were Sercial, Verdelho, Boal, and Malvasia; their origin is not known for certain, but except for Malvasia, they probably originated in Portugal. There seem to have been several different types of Malvasia; the desirable version is the Malvasia Candida, which is thought to be of Greek origin. Less important 'noble' varieties, also capable of producing superior wine, are Terrantez, Bastardo, and Listrao. Because of the humid climate, the viticulture was

severely affected by *Oidium* and *Plasmopora* when these fungus diseases arrived in the nineteenth century, and also by *Phylloxera*, a bug which attacks the vine roots. Most farmers, searching for a cure, replaced the sensitive classic varieties by direct producing hybrids, such as Jacquez, Cunningham, or Isabella, or at best by more robust vinifera varieties, such as Malvasia Complexa and Tinta Negra Mole. The viniferas were of course grafted on to American root stocks. Since Portugal joined the European Community, growers have come under pressure to eliminate the direct producing hybrids, and vineyards are being converted by grafting to vinefera varieties. Good clones of the classic varieties Malvasia Candida, Verdelho, Boal, Sercial, Terrantez, Bastardo, and Listrao are available from the official service, but the grapes are not yet widely available for processing.

Vintage

Near sea level, the vintage begins in August but continues into October on the highest ground. The steep slopes and difficult terrain mean that the grapes must be picked by hand and often carried considerable distances on the pickers' backs.

Wine-making

Practices vary: roller crushers are used with or without destalking; red grapes are normally fermented on skins; white wines may or may not be pressed before fermentation. Depending on the scale, fermentation may be in stainless steel tanks, concrete tanks, Algerian-type autofermenters, or wooden casks. The same casks are used from year to year. There seems to be no temperature control, so that in the larger vessels, 35 °C may be reached. In contrast to other areas, such temperatures are not considered deleterious. Grape concentrate is sometimes added to the fermentation to increase alcohol production. Cultured yeast is rarely required. Traditionally, all wines were fermented dry, and sweetening wine was added later as in sherry production, but nowadays, some producers, particularly of wines intended for the sweeter styles, prefer to stop the fermentation at the required sweetness by adding brandy, as port producers do. (*See* **Port**: The Product and its Manufacture; **Sherry**: The Product and its Manufacture.)

Fortification

Some producers fortify immediately; this prevents the malolactic fermentation and inhibits acetic acid bacteria. Other producers delay fortification until January, accepting the slight increase in volatile acidity, which they regard as traditional and a contribution to the complexity of flavor. Fortification is to between 17 and 18% alcohol by volume: grape alcohol obtained from the Instituto da Vinha e do Vinha must be used. Cheaper wines are sometimes not fortified until after the heating process, 'estufagem,' (see below) to avoid heavy losses of alcohol during the heating process.

Maturation

As soon as a suitable vessel becomes available, estufagem takes place, and it is largely this process which gives Madeira wines their characteristic flavor. Estufagem must take place during the first 3 years and involves heating in a concrete tank at 40–50 °C for a minimum of 3 months. Lower temperatures and a longer heating period are conducive to quality.

Some wines are not heated in concrete tanks but matured in wooden casks stored in a warm part of the warehouse. Sometimes, they are transferred between hot and cool areas alternately each 15 days.

Before estufagem, the wine is sweetened to between 1 and 5 ° Baumé according to style. (For the purposes of comparison, it can be assumed that 1 ° Baumé ≡ 1.79 ° Brix ≡ 1.79 g of sucrose per 100 g of solution at 20 °C.) Because of its relatively high total acidity, Madeira is not normally consumed completely unsweetened. Three sweetening materials are in use: surdo, which is essentially a mistela, i.e., fortified grape juice; caldo, which is essentially hydrolyzed cane sugar syrup; and rectified concentrated grape must. Caldo has probably the longest history and gives the authentic Madeira style, but there is a prejudice against nongrape sugar in wines in some markets.

After estufagem, the wine requires some clarification, and albumen and bentonite are used. Some, but not all, producers regard a treatment with charcoal immediately after heating as highly beneficial. Whatever treatment is given, a further extended maturation in oak casks at normal temperatures is required.

Commercial Styles of Madeira

Vintage Madeira, while a year-old wine, is selected as being of superior quality. Once selected, the casks are sealed and monitored by the Madeira Wine Institute. Storage must last a minimum of 20 years in wood, and at least 2 years in glass. Rebottling, under supervision, every few years is not unusual.

Superior Madeira is a wine of outstanding quality produced from the noble varieties. The denomination is usually used for single variety wines, e.g., Superior Bual.

Reserva Velha or Muito Velho requires a minimum of 10 years' maturation, Reserva or Velho requires a minimum of 5 years' maturation, and Selecionado requires a minimum of 3 years' maturation.

Solera is matured in a fractional blending system, and a number of restrictions apply, e.g., no wine can be withdrawn in the first 5 years and only 10% per annum thereafter. The year when the system was established may be given.

Wines may be labeled according to sweetness as follows:

- Sercial: 1–1.2 ° Baumé
- Verdelho: 2–2.2 ° Baumé
- Boal: 3.5 ° Baumé
- Malmsey: 4.5–5 ° Baumé.

This is a traditional way of indicating sweetness, except that, in the case of vintage and superior wines, it is not an indication that the corresponding noble varieties have been used.

California Baked Sherry

This wine, which originated in the Central Valley in the nineteenth century, is produced by a technique very similar to that used for Madeira. Although the amount manufactured has declined significantly since 1960, it is still substantially greater than that of Madeira wine, and, together with the Madeira wines of Russia, probably represents the most significant volume of heat-matured dessert wines.

Vineyards

The wine is manufactured from grapes grown in the high-yielding Central Valley of California. The Palomino, the traditional grape of the sherry district of Spain, has been preferred, but historically, there was a surplus of Thompson Seedless, a high-yielding, neutral grape that has proved adequate for standard-quality wine.

Acid grapes (pH 3.2) are preferred, and fermentation at 25–30 °C is carried out by cultured yeasts, the Montrachet strain of *Saccharomyces cerevisiae* or similar, until more or less dry. The wine is fortified with neutral grape spirit to 17–18% alcohol by volume.

An acidity of 0.4–0.5 (g per 100 ml of tartaric acid) is preferred. Most producers sweeten the wine to about 2% before baking.

Baking

Temperatures of 55–60 °C are maintained for 9–20 weeks. The tendency is for lower temperatures and shorter times to be used than in the past. Some producers also expose the wine to contact with oak chips to increase the complexity of flavor. Relatively little further ageing after cooling is used. The wine is stabilized usually with charcoal, bentonite, and refrigeration.

Commercial Styles

Dry sherry is characterized by 1–2.2% reducing sugar, medium sherry by 2.7–3.5% reducing sugar, and sweet (cream) by 7.5–10% reducing sugar.

Baked Sherry of the Eastern USA

Most types of baked sherry are produced by a method patented by Tressler in 1939. The base wine is a fortified wine made from the non-*Vinifera labrusca* varieties grown locally (Concord, Catawba, etc.). The heating is for approximately 6 weeks at 60 °C. Oxygen is injected through fine diffusers. The oxygen leaving is passed through a trap of cold wine, which retains some of the alcohols and esters that would otherwise be lost.

Heated Dessert Wines made in the UK

There is a substantial production in the UK of alcoholic beverages made by the fermentation of reconstituted concentrated grape musts imported from Mediterranean countries. Selected yeasts are used to produce highly alcoholic wines (15–17% alcohol) without fortification with distilled spirit. The best of the wines are subjected to an extended period of maturation at elevated temperatures to improve the flavor. Specific information on the times and temperatures used is not available. In some cases, oak chips may be exposed to increase the complexity of the flavor.

See also: **Port**: The Product and its Manufacture; **Sherry**: The Product and its Manufacture

Further Reading

Amerine MA, Berg HW, Kunkee RE *et al.* (1980) *The Technology of Wine Making*, 4th edn, pp. 391–403. Westport, CT: AVI.

Cossart N (1981) *Madeira: The Island Vineyard*. London: Christies Wine Publications.

Goswell RW and Kunkee RE (1977) *Fortified Wines, Economic Microbiology*, vol. 1, pp. 477–535. London: Academic Press.

Nogueira JM and Nascimento AM (1999) Analytical characterization of Madeira wine. *Journal of Agriculture Food Chemistry* 47(2): 566–575.

MAGNESIUM

I J Griffin, Baylor College of Medicine, Houston, TX, USA

Introduction

Magnesium, a soft, silvery-gray metal, is the eighth most common element in the earth core, with an abundance of 2.5%. Magnesium is found in many geological minerals, including asbestos, brucite, magnesite, dolomite, talc, and meerschaum. It has an atomic number of 12, and a mass of 24.31. Magnesium is a member of the alkaline earth metals (Periodic Table IIa), which include calcium, barium, beryllium, and strontium. It exists as three natural isotopes, the most common being magnesium-24 (79%), with magnesium-25 and magnesium-26 each comprising 10–11% of naturally occurring magnesium.

Magnesium was known to the ancient Romans as *magnesia alba* in acknowledgment of the white color of its carbonate salts found in Magnesia, Greece. Sir Humphry Davy in England first described the element in 1808, when he mixed magnesia, water, and mercuric oxide, heated them, and found that a previously unknown element, magnesium, was left after evaporation. Antoine Bussy further purified and isolated the metal in 1828. Magnesium was first detected in the serum of healthy adults by Denis in 1915. In 1926, LeRoy demonstrated that it was essential for the health and well-being of animals.

Magnesium is relatively reactive, and is rarely found in its metallic form in nature. When exposed to air, magnesium is rapidly oxidized, producing a dull-gray layer of magnesium oxide that protects the rest of the metal from oxidation. Magnesium is best known as a component of Epsom salts (magnesium sulfate) and magnesia or milk of magnesia (magnesium oxide). Sea water, which contains about 0.13% magnesium, is the most important commercial source of metallic magnesium. Eighty percent of commercial magnesium is extracted from sea water by precipitation with calcium hydroxide, chloride conversion with hydrochloric acid, and electrolysis. The remaining 20% of commercial production is achieved by thermal extraction from the magnesium-rich mineral, dolcite.

Magnesium has proven useful in numerous industrial applications. Its reactivity has been employed in photographic flashes, fireworks, and incendiary devices, which take advantage of the rapid, explosive oxidation of finely divided magnesium powder in air. Because of its light weight, magnesium has found uses in the automobile, aviation, and space industries. Magnesium has poor intrinsic structural strength, but can be bolstered by alloying it with other metals (particularly aluminum, zinc, and manganese).

In recent years, much has been learned about its biological importance, but many areas of uncertainty remain, especially the effects of marginal magnesium status on human health.

Biological Role of Magnesium

Magnesium is the second most common cation within living cells, exceeded only by potassium. The body of a typical adult contains 20–28 g of magnesium, the majority within skeletal and cardiac muscle ($\cong 60\%$) and bone ($\cong 30\%$). Only 1–2% of magnesium is found in extracellular tissues, and less than 0.5% is found in plasma. The concentration of magnesium in the plasma is about 0.75–1.25 mmol l^{-1} (1.5–2.5 mEq l^{-1}), and is regulated within a relatively narrow range by poorly understood mechanisms. Hypomagnesemia is generally diagnosed when serum magnesium levels fall below 0.75 mmol l^{-1}, and hypermagnesemia when the levels exceed 1.25 mmol l^{-1}. Red blood cells contain about 2–3 mmol l^{-1} magnesium, while skeletal and cardiac muscle contains 5–10 mmol l^{-1} magnesium.

Magnesium plays a number of vital physiological and biochemical roles in humans. Its highest concentration within the cell is in the mitochondria, where it is an essential cofactor for carboxylase and coenzyme Q. It is vital for energy metabolism and the conversion of adenosine diphosphate (ADP) to adenosine triphosphate (ATP). It is important in maintaining normal chromatin structure, hence, normal gene transcription. It is present in a variety of metalloenzymes, is a component of teeth and bones, and takes part in complex interactions (both antagonistic and synergistic) with calcium metabolism. Magnesium is required for normal vitamin D metabolism; patients with hypocalcemia may not respond to treatment with calcium if coexisting magnesium depletion is not recognized and treated. Magnesium is required for cyclic adenosine monophosphate (cAMP) production and associated intracellular signaling. Intracellular ionized magnesium is essential for muscle contraction and nerve function. Many magnesium-containing enzymes have been identified, including enolases, phosphates, transphosphates, choline oxidase, and pyruvate oxidase. Magnesium is required for purine and pyrimidine biosynthesis, and therefore, synthesis

of RNA and DNA. Finally, magnesium in the form of magnesium-porphyrin is present in plant chlorophyl, and is therefore essential for fixing the solar energy upon which all life on earth ultimately relies.

Magnesium Metabolism and Homeostasis

Absorption

Magnesium is absorbed predominantly in the small intestine, mainly in the ileum. A small amount of magnesium may be absorbed in the colon and a number of prebiotic sugars, such as inulin and oligofructose, may enhance colonic magnesium absorption, either through a change in colonic flora, or, less likely, via a trophic effect on the gut. Other factors are known to inhibit or enhance magnesium absorption (Table 1).

Typically, 20–50% of dietary magnesium is absorbed, but much more can be absorbed when dietary magnesium intakes are low. Studies in humans have shown that up to 75% of dietary magnesium can be absorbed when intakes are low, a figure that falls to 10–15% when intakes are high. The kinetics of magnesium absorption are consistent with a relatively low-capacity, saturable, active transport mechanism, and a higher-capacity, nonsaturable, passive transport mechanism. Although the molecular basis of magnesium absorption is not clear, congenital cases of isolated magnesium malabsorption have been described.

Distribution

Magnesium is transported in the plasma, where about 45% is bound to proteins and other ligands, and about 55% is in the ionized form. Magnesium uptake into the cell is an active process, and is probably coupled to sodium efflux. It is modified by a number of hormones and drugs, including insulin, growth hormone, and β-agonists.

Excretion

Some magnesium is secreted into the gut in digestive secretions, but most of this is absorbed in the enterohepatic circulation. Magnesium is also excreted into the urine. Approximately 70% is filtered into the glomerulus, and most is subsequently reabsorbed. Between 30 and 40% is reabsorbed in the proximal convoluted tubule, and most of what remains is reabsorbed in the ascending loop of Henle. This is a major source of magnesium homeostasis. Reabsorption in the proximal convoluted tubule appears to be passive, but reabsorption in the loop of Henle is active, although the hormone or hormones regulating this are not clear. Renal losses increase with alcohol consumption and diuretic use, both of which may predispose to magnesium deficiency. Under normal circumstances, at least 90% of filtered magnesium is reabsorbed in the kidney. During periods of severe magnesium restriction, however, urinary magnesium excretion can fall to very low levels within 7 days. High levels of magnesium supplementation, in contrast, increase the filtered magnesium load at the kidney, and urinary magnesium excretion increases significantly. Because of the high percentage of magnesium that is reabsorbed in the kidney, medications that interfere with this (such as some diuretics) may lead to magnesium deficiency and hypomagnesemia.

Table 1 Enhancers and inhibitors of magnesium absorption

Enhancers of magnesium absorption	*Inhibitors of magnesium absorption*
Lactose	Calcium
Vitamin D	Phosphate
Parathyroid hormone	Phytate
	High-fiber diets
	Oxalic acid
	Some fats

Homeostasis

The molecular control of magnesium homeostasis is not known. Homeostasis seems to be achieved by a combination of the regulation of absorption and urinary excretion. Endogenous fecal excretion does not appear to be important in magnesium homeostasis, as most of this magnesium is reabsorbed by the enterohepatic circulation. The kidney is remarkably adept at maintaining magnesium homeostasis; it is only in the latter stages of chronic renal failure that magnesium concentrations begin to rise in the plasma.

Assessing Magnesium Status

It is difficult to assess magnesium status accurately. Studies in humans have been limited, as the radioactive isotope magnesium-28, and the two stable isotopes, magnesium-25 and magnesium-26, are poorly suited to metabolic studies. The radioisotope is too short-lived to be useful, and the stable isotopes are insufficiently rare to be easily used as tracers. It is possible to measure the magnesium concentration in serum, either total or ionized. The former can vary with albumin concentration and serum pH, but the serum magnesium concentration does fall during experimental magnesium depletion in humans. It has been suggested that ionized magnesium may be a more useful measurement of magnesium status. In practice, however, the two correlate well, and there is

little evidence to suggest the superiority of ionized magnesium concentrations.

Magnesium can be measured in red blood cells, white blood cells, and hair. It has been suggested that the magnesium concentration in mononuclear cells is a better index of magnesium status than the serum magnesium concentration, although definite evidence is lacking.

The urinary excretion of magnesium after an intravenous load has been used as a measure of magnesium status, but its usefulness as an assessment of magnesium status is limited.

Magnesium kinetics can be measured using stable isotopes and compartmental models, but this remains an evasive research technique. This model can identify one pool corresponding to plasma magnesium, another approximating to magnesium in muscle, and a third approximating magnesium in bone. The use of such models has demonstrated a reduction in these pool sizes in rats placed on magnesium-deficient diets. However, these measures are complex and expensive, and very few centers are able to carry them out. They are unlikely to become a clinical measure of magnesium status, although they may prove of use in research settings. Similarly, intracellular magnesium concentration can be measured by nuclear magnetic resonance spectroscopy, but the clinical utility of this method is unclear.

Greater understanding of the importance of marginal magnesium status in human disease is severely limited by the lack of a reproducible, sensitive, and specific measure of magnesium status. Although many measures have been proposed, there is little conclusive evidence to support their use.

Dietary Sources of Magnesium

The highest concentration of magnesium is found in nuts, vegetables, and legumes. Cashew nuts are the richest source, containing 270 mg magnesium g^{-1}, mostly in the outer layers; a significant proportion is lost during processing. Green vegetables contain significant amounts of magnesium. Meat, fish, and diary products contain relatively little magnesium. Drinking water contains variable amounts of magnesium, but particularly high levels are found in so-called hard water.

The average intake of magnesium has steadily declined during most of the twentieth century as the consumption of processed foods has increased and the intake of unprocessed whole foods has decreased.

In 1994, the US Department of Agriculture's Continuing Survey of Food Intake in Individuals (CSFII) estimated that magnesium intake in males aged 9 years or older averaged 323 mg, but only 228 mg in females in the same age range. Ethnic differences have also been found in magnesium intake. The third National Health and Nutrition Examination Survey, known as NHANES III, showed that African-Americans consumed significantly less magnesium than Caucasians or Hispanics.

Magnesium Deficiency

Overt magnesium deficiency is rare in otherwise healthy humans. However, it is seen in association with a number of other diseases (Table 2). These are usually associated with excessive losses (e.g., renal disease, aldosteronism) or poor absorption (gastrointestinal causes).

After starting on a magnesium-free diet, symptoms of deficiency can appear with a month, but in some individuals, they may not be seen for more than 3 months. Symptoms include nausea, muscle weakness, anorexia, ataxia, tetany, convulsions, mental confusion, and irritability. Often, magnesium deficiency will coexist with deficiencies of vitamins, calcium, and potassium, especially if the underlying cause is malnutrition. Magnesium deficiency is characterized by a fall in urinary magnesium and a decrease in loss of magnesium into the gastrointestinal tract. These are presumably adaptive mechanisms designed to conserve magnesium and ameliorate magnesium deficiency. Magnesium absorption may be very low during magnesium deficiency if poor absorption is the underlying cause, or increased as an adaptation to magnesium deficiency from other causes.

Table 2 Conditions associated with magnesium deficiency

Causes	*Examples*
Renal dysfunction	Renal tubular acidosis
	Diuretic use
	Chronic glomerulonephritis
	Bartter's syndrome
Endocrine causes	Hyperthyroidism
	Primary aldosteronism
	Secondary aldosteronism
Gastrointestinal causes	Malabsorption
	Steatorrrhea (e.g., cystic fibrosis)
	Short-gut syndrome
	Prolonged diarrhea
	Celiac disease
	Tropical sprue
	Inflammatory bowel disease
	Excessive losses
	Prolonged vomiting
	Prolonged diarrhea
Inadequate intake	Malnutrition
	Iatrogenic (e.g., TPN with Mg^{2+})
Others	Alcoholism
	Primary idiopathic hypomagnesemia

TPN, total parenteral nutrition.

One study has produced experimental magnesium deficiency by feeding magnesium-deficient formulas to humans recovering from surgery for oral cancer. When subjects were switched from a magnesium-adequate formula to a magnesium-deficient one, urinary and fecal magnesium losses fell sharply, but were not sufficient to prevent a steady decline in serum magnesium. In these subjects, hypocalcemia (low serum calcium) and hypokalemia (low serum potassium) were also seen, even though the magnesium-deficient formula provided adequate calcium and potassium. It would therefore seem that the hypocalcemia and hypokalemia seen in magnesium-deficient individuals are not solely due to low intakes, as magnesium appears to have a profound effect on potassium and calcium homeostasis.

Magnesium Excess

Magnesium toxicity is rare. This condition is characterized by smooth-muscle relaxation, nausea and vomiting, mental confusion, cardiac arrhythmias, and, ultimately, coma and cardiac arrest. Hypotension is one of the earliest signs, occurring by the time magnesium serum levels are doubled. Death does not usually occur until serum levels exceed 7–8 mmol l^{-1}.

It is unusual for excessive dietary intakes of magnesium to cause toxicity, but some cases have been reported following excessive use of medical magnesium salts such as Epsom salts (magnesium sulfate, used as a laxative) or milk of magnesia (magnesium hydroxide, an antacid and laxative). In humans, the most likely cause of magnesium toxicity is renal failure and an inability to excrete magnesium in the urine. It is possible for magnesium toxicity to occur during therapeutic use of intravenous magnesium to treat preterm labor or pregnancy-induced hypertension.

Recommended Dietary Intakes of Magnesium

Due to the lack of a reliable measurements of magnesium status, current recommendations for magnesium intake are largely based on metabolic balance studies and estimates of the magnesium required for 'normal' growth.

The Food and Nutrition Board of the Institute of Medicine recently published a comprehensive report of dietary reference intake for magnesium. Its findings are summarized in Table 3. A number of different definitions were used:

- The estimated average requirement (EAR) is the intake that will meet the needs of half of a given population. At this intake, therefore, half of the population will have an inadequate intake.
- The recommended dietary allowance (RDA) is the intake that will meet the needs of most (97%) of a given population.
- The adequate intake is given if insufficient data are available to define the RDA. At this intake, the needs of essentially all the population are met.

In addition, it was recommended that the EAR for pregnant women should be 35 mg higher than for nonpregnant women of the same age. No increase was deemed necessary for lactating women.

Table 3 Dietary reference intakes for magnesium

Age	*Gender*	*AI*	*EAR*	*RDA*
0–6 months	M/F	30 mg	•	•
7–12 months	M/F	75 mg	•	•
1–3 years	M/F	•	65 mg	80 mg
4–8 years	M/F	•	110 mg	130 mg
9–13 years	M/F	•	200 mg	240 mg
14–18 years	M	•	340 mg	410 mg
	F	•	300 mg	360 mg
19–30 years	M	•	330 mg	400 mg
	F	•	255 mg	310 mg
> 31 years	M	•	350 mg	420 mg
	F	•	265 mg	320 mg

AI, adequate intake; EAR, estimated average requirement; RDA, recommended dietary allowance. •, if insufficient data were available to determine the RDA and EAR, an AI was recommended instead.
Data from The Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (1997) *Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D and Fluoride*. Washington, DC: National Academy Press.

Magnesium and Human Disease

Magnesium deficiency or low magnesium intakes have been suggested as risk factors for a number of important human diseases, particularly hypertension and coronary heart disease.

Hypertension

Magnesium plays an important intracellular role in decreasing the excitability of smooth muscle, leading to a reduction in smooth-muscle tone. In addition, intravenous magnesium therapy is a useful treatment of hypertensive emergencies in pregnant women (pregnancy-induced hypertension). Animal studies have shown a consistently inverse relationship between magnesium intake and blood pressure, and an inverse correlation between intracellular magnesium concentration and blood pressure. For these reasons, there has been considerable study of the effect of magnesium nutrition on blood pressure.

The results of a number of observational studies have suggested that low serum or urinary magnesium

concentrations may be risk factors for hypertension. But these studies are contradictory, and negative results have been reported in many other studies. Some interventional studies have evaluated the effect of magnesium supplements on hypertensive individuals. Preliminary evidence suggests that, in hypertensive individuals treated with thiazide diuretics, magnesium supplementation can lower blood pressure. This is physiologically plausible, as thiazides inhibit magnesium reabsorption in the kidney and may lead to magnesium deficiency. However, a similar effect has been demonstrated with potassium supplementation, with no additional benefit from combined magnesium and potassium supplementation. Further, magnesium supplementation of hypertensive individuals with normal serum magnesium levels has not been of any benefit. Clearly, the current literature is confusing, and further work to determine potential interactions among hypertension, diuretic use, and magnesium supplementation is urgently needed.

Cardiovascular Disease

There is increasing evidence that populations residing in areas supplied with hard water (which has higher mineral content, including magnesium) have a lower risk of cardiovascular disease than people who live in similar regions supplied with soft water. Early studies, however, suggested that higher levels of serum magnesium can be detected in serum immediately after myocardial infarcts. More recent studies have suggested that this is a result of magnesium released from dying heart cells, rather than an indication that an existing high serum magnesium concentration is a risk factor for myocardial infarct *per se*. Some intervention studies have shown that intravenous magnesium may be a useful adjuvant therapy in acute myocardial infarcts.

Diabetes Mellitus

Diabetics have increased urinary magnesium losses. Such losses are especially high in individuals with poorly controlled diabetes, those with high amounts of glucosuria, and those with diabetic complications. It is unclear whether the correlation between low serum magnesium concentrations and diabetic complications is a cause and effect relationship, or whether both are independent effects of poor glycemic control. However, magnesium supplementation has been reported to improve insulin sensitivity and glycemic control in adults with noninsulin-dependent diabetes mellitus. In 1992, the American Diabetic Association (www.diabetes.org) convened a consensus panel to evaluate the possibility of a relationship between magnesium deficiency and diabetes mellitus. The panel members felt that there was evidence to suggest that magnesium deficiency might be related to some of the underlying causes of diabetes, such as insulin resistance and carbohydrate intolerance. However, they recognized the difficulty of diagnosing magnesium deficiency, and did not recommend universal magnesium supplementation for diabetics.

Osteoporosis

Thirty percent of total body magnesium is in bone. Magnesium is important for normal bone growth and mineralization. Some studies have suggested that magnesium supplementation can increase bone mineral content in subjects on relatively low magnesium intakes, but such benefits were not seen in women on higher magnesium intakes. Two studies have shown that magnesium supplementation may have beneficial effects on bone mineral content in women with established osteoporosis.

Magnesium as a Pharmacological Agent

Magnesium is unusual in that it has physiological effects as a nutrient, and in much higher doses, has a pharmacological effect as well.

One of the important biological functions of magnesium is to alter the excitability of nerves and muscles, either directly or through an interaction with calcium. Large doses of intravenous magnesium have been used to treat medical conditions involving excessive excitability of nervous tissues. The classic example is the use of intravenous magnesium to treat pregnancy-induced hypertension. This condition, also called preeclampsia, is characterized by proteinuria, edema, and hypertension. Untreated, preeclampsia leads to central nervous system excitability, convulsions and, ultimately, death. It is the most common cause of maternal morbidity in western countries. Intravenous magnesium has been widely used for many years to treat this condition; it has been shown to decrease the risk of seizures and improve maternal and fetal outcome.

Magnesium is also used to delay the progression of preterm labor, again by relaxing excitable cells, in this case, uterine muscle cells. It is also used as a treatment for primary pulmonary hypertension of the newborn (or persisting fetal circulation). This latter condition is caused by abnormal contraction of the blood vessels entering the lung, and leads to profound hypoxia and, ultimately, death. Magnesium acts to relax these pulmonary blood vessels and improve oxygenation. Unfortunately, it may also relax other blood vessels and lead to hypotension, counteracting its beneficial effect on pulmonary vessels.

There are conflicting data on the benefit of intravenous magnesium following acute myocardial

infarcts. The LIMIT-2 study suggested that intravenous magnesium reduced mortality from 10.3% to 7.8%, and also significantly decreased the risk of congestive heart failure. The much larger ISIS-4 study found no such benefit in a much larger sample size (≅ 60 000 subjects), and a possible increase in mortality in subjects with heart failure who were treated with magnesium.

Once again, magnesium offers possible benefits for treating human disease, but the evidence (particularly for the treatment of heart attacks) is confusing, and further studies are needed before magnesium can be recommended in these conditions.

Conclusions

Magnesium is a vital component of the plant protein chlorophyl, which is responsible for fixing solar energy into chemical energy in green plants. In this respect, most life on earth is dependent on magnesium.

In humans, magnesium is an essential nutrient, and is widely available in the diet. Severe magnesium deficiency is very rare in otherwise healthy individuals, but is seen in a number of disease states, particularly renal disease, malabsorptive syndromes, and in alcoholics. There is increasing concern about the effect of marginal magnesium status, or low dietary intake, on chronic diseases, including osteoporosis, myocardial infarct, and stroke. Over the next 5–10 years, more information is likely to accumulate to help us understand the effect of dietary magnesium on these conditions, and the potential use of magnesium as a therapy for these conditions when they occur.

Acknowledgment

Supported in part with federal funds from the USDA/ARS under Cooperative agreement No. 58-6250-6-001. This work is a publication of the US Department of Agriculture (USDA)/Agricultural Research Service (ARS) Children's Nutrition Research Center, Department of Pediatrics, Baylor College of Medicine and Texas Children's Hospital, Houston, Texas. Contents of this publication do not necessarily reflect the views or policies of the USDA, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government.

See also: **Alcohol**: Alcohol Consumption; **Chlorophyl**; **Coronary Heart Disease**: Etiology and Risk Factor; **Diabetes Mellitus**: Etiology; **Hypertension**: Hypertension and Diet; **Osteoporosis**

Further Reading

Aikawa JK (1981) *Magnesium: Its Biological Significance*. Boca Raton: CRC Press.

American Diabetic Association(1992) Magnesium supplementation in the treatment of diabetes. *Diabetes Care* 15(8): 1065–1067.

Committee on Diet and Health, Food and Nutrition Board, Commission on Life Sciences, National Research Council (1989) *Diet and Health, Implications for Reducing Chronic Disease Risk*. Washington, DC: National Academy Press.

Czajka-Narins DM (1996) Minerals. In: Mahan LK and Escott-Stump S (eds) *Krause's Food, Nutrition and Diet Therapy*, 9th edn. Philadelphia: WB Saunders, pp. 109–140.

Elmadfa I and Koenig JS (1993) Magnesium. In: Macrae R, Robinson RK and Sadler MJ (eds) *Encyclopedia of Food Science, Food Technology and Nutrition*. London: Academic Press, pp. 2822–2824.

Ensminger AH, Ensminger ME, Konlande JE and Robson JRK (eds) (1994) Magnesium. In: *Food and Nutrition Encyclopedia*, 2nd edn. Boca Raton: CRC Press, pp. 1338–1341.

Goulding A and Robinson M (1998) Calcium and magnesium. In: Mann J and Truswell AS (eds) *Essentials of Human Nutrition*. Oxford: Oxford University Press, pp. 123–136.

Sauberlich HE (1999) Magnesium. In: *Laboratory Tests for the Assessment of Nutritional Status*. Boca Raton: CRC Press, pp. 331–342.

Shils ME (1969) Experimental production of magnesium deficiency. *Annals of the New York Academy of Science* 162: 847–855.

Shils ME (1996) Magnesium. In: Ziegler EE and Filer Jr. LJ (eds) *Present Knowledge in Nutrition*, 7th edn. Washington, DC: ILSI Press, pp. 256–264.

Shils ME (1999) Magnesium. In: Shils ME, Olson JA, Shike M and Ross AC (eds) *Modern Nutrition in Health and Disease*, 9th edn. Baltimore: Williams & Wilkins, pp. 169–192.

The Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (1997) Magnesium. In: *Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D and Fluoride*. Washington, DC: National Academy Press, pp. 190–249.

Widdowson E and Dickerson JWT (1964) Chemical composition of the body. In: Comar CL and Bronner F (eds) *Mineral Metabolism*. Vol. II: *The Elements*. New York: Academic Press, pp. 2–207.

Maillard Reaction *See* **Browning**: Nonenzymatic; Enzymatic – Biochemical Aspects; Toxicology of Nonenzymatic Browning; Enzymatic – Technical Aspects and Assays

MAIZE

S R Eckhoff and M R Paulsen, University of Illinois, Urbana, IL, USA
S C Yang, Providence University, Taiwan, Republic of China

Introduction

Maize, *Zea mays* L. (corn), is the second most abundantly produced cereal in the world, exceeded only by rice. Maize is grown in every continent except Antarctica. North America grows over 43% of the world's maize crop, with 90% of the US total grown in the Corn Belt consisting of 12 midwestern and north central states. Other world leaders in maize production include China (18%), the EU (7%), Brazil (6%), the Balkan area (3%), and Mexico (where maize is generally agreed to have originated) (3%), Argentina, India, and South Africa (2% each). The ease with which maize can be dried, stored, and transported has made it a major source of energy for animal food and a stable raw material for the production of starch and starch-based industrial products. Maize is also a primary food source in many areas of the world, including South America, Central America, and Africa, where it is converted directly into food products via grinding, alkali processing, boiling/cooking, or fermentation.

This article reviews the structure and composition of maize, the genetic diversity of maize, the issues related to genetically modified maize, the drying, storage, handling, and grading of maize, and the major fractionation/processing methods.

Morphology/Anatomy

The maize kernel is the reproductive seed of the maize plant. As such, its structure and composition exist for the purposes of reproduction rather than processing. Figure 1 shows that the kernel is composed of four main parts expressed as a dry weight percentage of the whole kernel: germ, 12%, endosperm (horny and floury), 82%, pericarp, 5.2%, and tip cap, 0.8%. Each of these sections has distinct compositional characteristics that are important for utilization of the maize kernel (Table 1).

The *germ*, the primary focus of maize kernel from a reproductive standpoint, contains all the essential enzymes, nutrients, and genetic material to produce a maize plant. If the germ is removed from the kernel, planted, and supplied with the necessary chemicals for growth, it will produce a new plant. The remainder of the kernel's purpose is to protect the germ and to provide the energy and structural components necessary for the early growth of the plant.

The germ contains approximately 34% oil, 19% protein, and 28% solubles (sugars, water soluble proteins, mineral, and vitamins), and the balance insoluble materials. Although germ is relatively high in protein, the main interest to processors is the oil it contains. Maize oil is a highly desired food vegetable oil due to its relatively high level of linolenic fatty acid (polyunsaturated oil) and excellent flavor. The germ contains over 65% of the kernel's albumins (water-soluble proteins) and globulins (salt soluble proteins), 25% of the total maize protein, and a high percentage of the kernel's soluble sugars, vitamins, and minerals.

The *tip cap* is the point of the kernel attaching to the cob and the passageway for the movement of nutrients to the developing kernel. Near the end of maturation of the kernel, a black layer (hilar layer) forms across the tip cap region in order to serve as a diffusional barrier and to provide some protection for the kernel from invading insects and microorganisms. This black layer can be seen in Figure 1.

Enveloping the entire kernel (except the tip cap region) is the *pericarp*, which is comprised entirely of dead, empty cells. These cells are high in cellulose and hemicellulose and are arranged in several distinct layers. The epidermis, the outer one or two cell layers, is thick walled and covered with cutin, a waxy substance that restricts the entry of water, water vapor, and many other gases and liquids. Below the

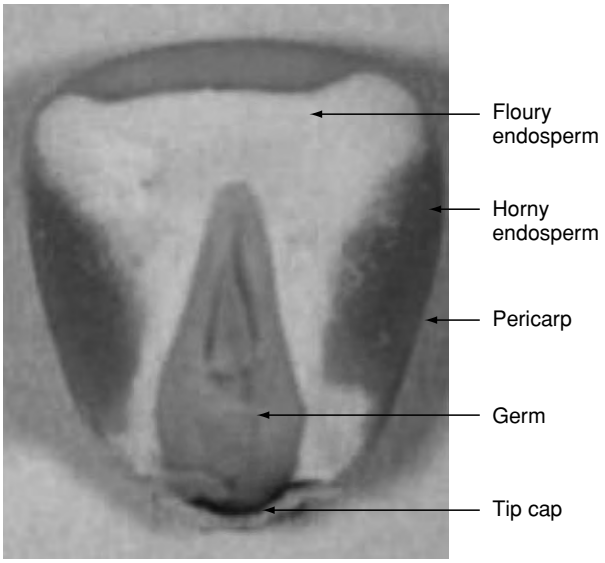

Figure 1 (see color plate 102) Maize kernel.

Table 1 Maize kernel composition by parts for nine midwestern US dent maize hybrids compared to flint and flour hybrids (*on a percentage dry-weight basis*)

	Proportion of whole kernel	*Starch*	*Protein*	*Oil*	*Ash*	*Sugar*
Kernel						
Mean of dents	100	71.5	10.3	4.8	1.4	2.0
Range of dents		67.8–74.0	8.1–11.5	3.9–5.8	1.3–1.5	1.6–2.2
Flint		67.8	13.6	5.6	1.5	2.0
Flour		66.8	13.2	5.6	1.5	2.0
Endosperm						
Mean of dents	81.9	86.4	9.4	0.8	0.3	0.6
Range of dents	80.3–83.5	83.9–88.9	6.7–11.1	0.7–1.11	0.2–0.5	0.5–0.8
Flint	80.6	84.7	12.8	0.8	0.4	0.7
Flour	79.7	84.7	12.8	0.8	0.3	0.8
Germ						
Mean of dents	11.9	8.2	18.8	34.5	10.1	10.8
Range of dents	10.5–13.1	5.1–10.0	17.3–20.0	31–38.9	9.4–11.3	10–12.5
Flint	13.5	8.0	20.2	35.3	9.5	10.0
Flour	14.1	8.0	19.9	35.0	9.2	9.9
Pericarp						
Mean of dents	5.3	7.3	3.7	1.0	0.8	0.3
Range of dents	4.4–6.2	3.5–10.4	2.9–3.9	0.7–1.2	0.3–1.0	0.2–0.5
Fling	5.1	7.0	4.6	1.0	1.2	0.3
Flour	5.5	7.4	4.6	0.9	2.0	0.3
Tip cap						
Means of all samples	0.8	5.3	9.7	3.8	1.6	1.6
Range of all samples	0.8–1.1		9.1–10.7	3.7–3.8	1.4–2.0	

Adapted from Earle FR, Curtis JJ and Hubbard JE (1946) Composition of the component parts of the corn kernel. *Cereal Chemistry* 23(9): 504–511.

epidermis is the mesocarp, which is composed of 17–25 layers of elongated cells, up to 1 mm long, connected by openings through which liquids and gases can diffuse. The next four to six layers of cells are large open cells, called the cross and tube cells, which allow for a high flux diffusion of gases or liquids. Through these cross and tube cells, water, which enters the tip cap region, can be rapidly transported around the kernel to allow for diffusion into the germ and endosperm.

Below the pericarp is a semipermeable membrane called the seed coat, which resists the movement of macromolecules into and out of the kernel. Just interior to the seed coat is the outermost endosperm layer, the aleurone, composed of large, thick-walled cells of high protein content. This aleurone layer may also impart semipermeable characteristics due to the compact arrangement of the cells in this layer. Damage to the pericarp and/or the seed coat increases the rate of water or chemical absorption.

The *endosperm* comprises hard (translucent, horny, and vitreous) and soft (opaque, floury, and corneous) endosperm. Both comprise a protein matrix, which encapsulates granules of starch. The translucent hard endosperm is composed of densely packed starch granules surrounded by a thick protein matrix, while the less dense soft endosperm has somewhat larger round starch granules surrounded by a thin protein matrix. As moisture escapes from the ear during the in-field drying of the dent type maize, the thin protein matrix ruptures, and the loosely packed soft endosperm structure collapses, leaving many microfissures which interrupt the structural continuity and give the kernel an opaque appearance. The external result of the soft endosperm structural collapse is that the crown region is pulled inward giving a dented appearance. During drying, the thicker protein matrix in the hard endosperm area shrinks but does not rupture. The pressure forces the starch granules into angular shapes, and the density increases. During processing, the hard endosperm structure resists disruption or modification, while the soft endosperm is easily disrupted.

The protein matrix in the endosperm is composed primarily of glutelin, a protein soluble in concentrated alkali, and zein, an ethanol-soluble protein. The glutelin is highly cross-linked with disulfide bonds and provides continuity and the structural characteristics of the endosperm. Zein does not add significantly to the endosperm structure because it primarily exists in spherical balls embedded in the protein matrix.

Starch granules are approximately 99.8% starch. Each granule in dent and flint maize starch is a mixture of two polysaccharides, amylose and amylopectin. Amylose is a linear polymer of glucose (several thousand anhydroglucose units per molecule), and amylopectin is a large multibranched polymer

(50 000–200 000 anhydroglucose units per module). The spherical starch granule consists of a radial arrangement of the amylose with the amylopectin interdispersed. Dent maize starch contains ~ 25% amylose and ~ 75% amylopectin, while waxy maize starch contains ~ 100% amylopectin, and high amylose maize starch contains 65–85% amylose. (*See* **Starch**: Structure, Properties, and Determination.)

Genetic Diversity

Zea mays is a member of the grass family and is one of the most diverse species of plants. Maize plants contain both male and female reproductive structures and reproduce by both cross-pollination and self-pollination. In most commercially viable maize genotypes, the female structure (the ear) projects outward from a central stalk, while the male structure (the tassel) projects out the top of the stalk. Pollen from the tassel is carried by the wind to other maize plants, where fertilization of the individual kernels on the ear occurs. The ears of maize may range in size from about 2.5 to over 45 cm long. The kernel size, shape, and color also vary widely.

Although there are hundreds of 'races' of maize, most of the commercially grown hybrids emanated from only a few major races. For the purposes of discussing the commercial importance of maize, maize types can be subdivided into four categories not related to race.

Dent maize is the primary type of maize grown in the US Corn Belt as well as in Europe, South Africa, and China. Dent varieties have been adapted through hybridization and selection to provide a wide range of agronomic and kernel characteristics. Special dent varieties have been produced with unique starch characteristics. High-amylose (linear starch) and waxy (branched starch) maize genotypes have been grown commercially for many years. Other unique genotypes such as high-oil maize and high-lysine maize are also being produced.

Flint maize is genetically different in ancestry from dent maize and is characterized by hard, round kernels. Flint maize endosperms consist predominantly of hard or vitreous endosperm. The hard endosperm allows flint maize to withstand greater impact forces before being damaged, which has advantages in commerical merchandizing. The agronomic characteristics of flint maize differ from dent maize, primarily due to the unique needs of the historical growing regions.

Popcorn is a flint-type maize that has been genetically selected for its ability to expand or 'pop' when heated. Popping occurs when the kernels are rapidly heated to ~240 °C. The dense endosperm restricts water vapor diffusion that causes pressure to build within the kernel until it explodes. It is the starch granules that explode and in the process stretch the protein matrix. As the protein matrix cools, it becomes rigid. The white fluffy part of popped popcorn is gelatinized starch splattered on the surface of the expanded cell protein matrix.

Sweetcorn is dent-type maize that is harvested while still immature for canning, freezing, and direct consumption as a vegetable. Sweetcorn hybrids contain a gene that retards the conversion of glucose to starch in the endosperm. There are generally three to four times more short-chain polysaccharides accumulated in sweetcorn than in other maize varieties.

Genetically Modified Maize

Biotechnology has made it possible to transfer genetic material within a species as well as between species. As a result of this development, the first genetically modified (GM) maize seed to resist pests, known as Bt (*Bacillus thuringiensis*) maize, was released in 1997. Bt maize has saved farmers millions of dollars annually due to the reduction of pesticides needed to reduce corn-borer infestations. It has been estimated that 25% of the maize area planted in the USA in 2000 was GM.

While biotechnology offers many opportunities to enhance maize production yields, functionality and millability, genetic engineering of crops has been met with resistance. Scientific data support the safety of genetically engineered crops. However, there are still a variety of nontechnical sociological issues surrounding the acceptance of these crops. Concerns vary from a general distrust of new technology to ethical concerns. Resistance to GM crops has been exacerbated by inflammatory press releases and action by some who view the issue as a means of passive tariffs. It appears that the controversy will not be settled in the near future. In the meantime, marketing systems for keeping GM grain separate from conventional grain are developing to provide customers with a choice.

Drying, Storing, Handling, and Grading

Maize is generally mechanically harvested at moisture contents ranging from 18 to 30% by weight and dried to a safe storage moisture. If the maize is harvested on the ear, it is harvested at moisture contents not exceeding 21% and dried in cribs or wire bins by using natural ventilation. Once dried, the ears are shelled by a stationary sheller. Nearly all maize in the major maize-growing regions is now harvested using mechanical picker-shellers, which separate the

kernels from the cob while in the field. The shelled maize can then be dried using one of a variety of drying procedures including high-temperature continuous driers, batch-in-bin high-temperature driers and low-temperature in-bin driers. The selection of a drying procedure depends upon the availability of equipment, energy cost, harvesting rate, and harvesting/drying weather.

Drying kernels at temperatures above 55 °C can decrease product yields and cause processing problems, as will be discussed in the following sections. Rapid drying causes internal stress fractures that reduce the mechanical strength of the kernel and results in more broken kernels during handling. An increase in the amount of broken maize during handling is the primary factor leading to grade reduction.

Most grades and standards for the merchandizing of maize contain the similar criteria of test weight (bulk density), broken maize, mold-damaged kernels, heat-damaged kernels, and foreign material (nonmaize material). In any grading standard, specific levels of the grade factors are assigned for any designated grade. Nongrade determining factors such as moisture content, mycotoxin levels, etc. can be added to the grade certificate depending upon contract specifications.

For long-term safe storage of maize, a moisture content of 13% is required; however, for storage up to one year in the temperate climates of the US Corn Belt, maize can be stored at 14% moisture content. The actual maximum moisture allowed for safe storage varies with the temperature and relative humidity of the air in the interstitial spaces surrounding the kernels. Fluctuations in temperature and atmospheric moisture can result in moisture migration within a mass of maize, which can result in quality reduction due to growth of fungi and insect infestation.

Maize may be handled as many as 20 times between harvesting in the field and final use. Each handling provides a series of impacts that can cause an increase in breakage. Breakage is measured in the USA as the percentage of the maize mass that passes through a 12/64 inch (4.76 mm) rounded hole sieve. This value may be 0.5–1.0% at the farm gate, 2.0–3.0% at in-land grain terminals, 2.5–4.0% at export terminals, and 5–15% at the ultimate international user. Breakage accumulates at each handling and depends upon the breakage susceptibility of the maize kernels and the impact experienced by the kernel during handling. The drying temperature and rate of drying have the greatest effect upon breakage susceptibility. Variety differences are significant, with hard endosperm maize having a lower breakage during handling as compared to softer dent hybrids. Breakage also increases with decreasing grain temperature. To reduce breakage during maize movement, gentle handling is needed. Long spout drops should be reduced so that the kernel velocity does not exceed $12\,m^{-1}$ (less than a 12-m free-fall drop). Velocity reduction equipment and proper facility design can greatly help reduce maize damage.

Processing and Food Uses

Although the major portion ($\sim 80\%$) of maize grown worldwide is used directly as animal food, processing maize for human food or chemicals is growing rapidly. In the USA, the amount of maize used for these purposes has grown from 7.2×10^6 t in 1960 to over 4.98×10^7 t in 2000. Worldwide growth in maize processing is expected to continue as the demand for processed food increases.

There are three major processes utilized in the production of maize for human food usage: dry milling, wet milling, and alkali processing. All three processes have developed specific products and markets.

Dry Milling

Dry milling is the simplest method of producing maize products for human consumption. Grinding whole kernel corn in a grind stone or roller mill to produce flour or meal is a simple method used worldwide when the ground products are to be consumed shortly after processing. The storability of such products is limited due to the presence of crushed germ in the flour and/or meal. Oil from broken germ cells is easily oxidized, producing a rancid odor and flavor. Whole ground maize can also be used as the substrate for beverages or fuel-ethanol production. (*See* **Milling**: Principles of Milling.)

To overcome any storage problems, most larger dry milling facilities remove germ as the first step in processing. Maize is tempered with water to a moisture level of 18–24% to toughen the germ and induce differential swelling of the germ, endosperm, and pericarp for ease of separation. A degerminator mechanically impacts the kernel to break loose the germ and pericarp from the endosperm. The resulting material is sieved, aspirated, gravity-separated, and roller-milled to separate the germ, pericarp, and endosperm pieces of different sizes.

The large grits (flaking grits) are used in the production of flaked breakfast cereals. Smaller grits are also used in some breakfast cereals, extruded maize snacks, and alcoholic beverage brewing. Meal fractions are used in extruded snack foods, food mixes, and brewing. Flour is used in food mixes, food coatings, and breadings. The low cost of maize flour relative to wheat flour makes it a good choice for many food applications where the strength of wheat

Table 2 Comparison of product yield for laboratory dry-milled medium-hard endosperm and medium soft endosperm dent maize dried at two temperatures

Drying temperature (°C)	*Product yield (percentage dry-weight basis)*			
	Medium-hard endosperm		*Medium-soft endosperm*	
	30	*105*	*30*	*105*
Grits[a]				
−3.5, +5	30.3	11.0	20.0	5.1
−5, +7	19.5	19.4	21.1	18.7
−7, +10	4.7	14.5	5.7	15.8
Total grits	54.5	45.8	46.7	39.6
Meal and flour	15.6	20.9	26.5	31.5
Germ fraction	20.5	23.1	26.5	31.5

[a]Standard US mesh.

gluten is not important. Maize dry-milled products also have a number of nonfood uses, including the manufacture of gypsum board, plywood, enzymes, chemical binders, and biodegradable plastic fillers and in intermediate chemical and biochemical fermentations.

Dry-milled germ can be pressed or solvent-extracted to recover the valuable oil. The defatted germ meal is most often combined with the pericarp fraction to produce an animal food product know as hominy food. Because the mechanical separation processes used in dry milling are imprecise, the germ and pericarp fractions often have significant amounts of attached endosperm material. The maize hybrid, drying temperature, tempering conditions, and degerminator adjustment affect the amount of attached endosperm in the pericarp and germ fraction, and the amount of broken germ in the endosperm products.

Table 2 lists representative dry milling yields of degerminated maize products. When properly degerminated, the larger grits will be low (0.45–0.55%) in oil. As the degerminated product size decreases, the amount of oil in the fraction increases to the content of 1.5–2.5% oil in the flour. Low-fat maize flour can be derived by the roller milling of larger low-fat grits or meal. Dry milling provides a variety of ingredients for use in food products, but the heterogeneous nature of the dry-milled products makes it difficult to find new markets. Maize dry milling in the USA only grew at an annual rate of *sim* 1.5% per year during the last several decades, from 2.9×10^6 t in 1960 to 3.5×10^6 t in 1998. The major advantages of maize dry milling are the lower use of energy in fractionation and lower capital costs as compared to wet milling.

Wet Milling

In the USA, the growth area in the utilization of maize has been wet milling, which has grown from 3.9×10^6 t in 1960 to over 43.2×10^6 t in 1998 due to the increasing demand for high-fructose corn syrup and ethanol. The relatively clean separation of maize into its functional components offers wet milling many opportunities for future growth. Maize yields several large molecules (starch, protein, and oil), which can be used as feed stocks for conversion to a wide variety of new food and industrial chemicals, including intermediate chemicals, amino acids, citric acid and other organic acids, synthetic fibers, plastic and styrofoam filler, and low-calorie fat. Environmental issues and regulations concerning petrochemical and crude oil transportation and utilization will likely allow these and other maize-based products to become more economically competitive in the future.

Wet milling is the use of biochemically and chemically induced changes in the maize kernel to facilitate mechanical separation of components. The kernel is first steeped for 20–48 h in a counter-current manner, which results in microaerophilic lactic acid fermentation during hydration of the kernel, followed by exposure to increasing levels of sulfurous acid (sulfur dioxide in water, 1000–2000 p.p.m.). Steeping toughens the germ, removes solubles from maize, which increases the oil content of the germ, loosens the interfacial forces between the endosperm germ and pericarp, and disrupts the protein matrix, in which starch granules are embedded by reducing disulfide bonding and enzymatic hydrolysis of the matrix protein, and hydrates the kernels to 45–50% (by weight) moisture. Concentrated steepwater, containing ~7% of the total solids, 3–5% protein and most of the soluble sugars, vitamins, and minerals, is often used as nutrient source for biochemical fermentations.

The kernels are then lightly milled to cause the germ to separate from the pericarp and endosperm. Germ recovery is accomplished in cyclones utilizing the density difference between the light germ and the heavier endosperm. The germ containing 45–50% oil is dry-pressed or solvent-extracted to recover the oil. Extracted germ meal is sold as a 20% protein animal food or added to the gluten feed fraction.

Nongerm material is further milled to mechanically disrupt the protein and starch in the endosperm. This fine grinding allows for the separation of cellular fiber from the starch and protein. The pericarp and cellular fiber are recovered using a series of bent pressure fed screens (4–7 units). The fiber is dewatered in a press, dried, and combined with concentrated steepwater to produce corn gluten feed (21%) protein. The underflow from the screens is thickened and centrifuged, resulting in a primary starch–protein separation.

The lighter protein fraction is further centrifuged to concentrate and purify. Belt filters are used to dewater the protein to 35–45% (by weight) moisture and then dried to produce corn gluten meal. The underflow of the primary centrifuge is starch containing ~ 1% protein. This starch slurry is pumped through a series of small cyclones (10–14 units) and washed with fresh water. The resulting starch, with a protein content of 0.25–0.50%, is dried in a flash drier. The starch can be chemically modified to enhance its functional characteristics for food or industrial use, chemically or enzymatically hydrolyzed for use in the production of fructose syrups, dextrose syrups, dextrins, or acid- or enzyme-thinned starch, or left unmodified for use in a variety of food and industrial applications.

Wet milling of specialty maize hybrids (high-amylose, waxy, high-oil, and high-lysine), produces starch or other components with enhanced characteristics. The steeping parameters of these specialty hybrids are often different from those of ordinary yellow dent maize. Currently, there is considerable interest in the USA in the development of specific maize hybrids which will have unique starch characteristics or which could increase wet milling profitability by enhancing millibility.

Flint and flint-type maize is less desirable than yellow dent maize for wet milling. In general, the flint maize takes longer to steep (50–60 h) and yields slightly less starch. Softer endosperm maize is more desirable for wet milling. Recent tests with all soft endosperm maize has shown that steep times can be reduced to 8–12 h, with no loss in starch yield and good starch quality.

The most deleterious factor affecting wet milling is the temperature at which the maize was dried when harvested (Table 3). When the drying temperature is about 50 °C, there is potential for starch gelatinization, protein denaturation, and endogenous protease inactivation. Processing of high-temperature dried maize results in a reduced starch yield, difficulty in protein (gluten) dewatering, increased starch in the fiber, and lower hydration of the maize kernels during steeping. Most US millers purchase US No. 3 or better maize because of the volume of maize purchased, although there has long been interest in being able to delineate high-temperature-dried maize from low-temperature-dried maize in the marketplace. International millers have observed major differences in the way in which US No. 3 maize mills is compared to natural-dried South African or Chinese maize. Natural air-dried or low-temperature-dried maize, as in South Africa or China, steeps easier, is less variable, has lower amounts of broken maize, and creates less dust. During wet milling, the fiber dewaters easier, gluten filters easier, and the starch yield is greater.

Alkali Processing

Alkali processing is the traditional lime cooking method employed by the Native Latin American Indians and Mexicans to produce tortillas. Tortillas are the major source of energy and nutrition in many Central American countries. In the USA, alkali-processed ethnic food and snack food is increasing at a faster rate than most other food uses of maize.

In alkali processing of maize, the maize is mixed with water and lime in a ratio of 1:1.2–3.0:0.0005 (maize:water:lime) and cooked at 94 °C for 50 min. The cooked maize is then steeped for 14 h in the lime solution before being washed with fresh water to remove loosened pericarp and residual alkali from the maize. The washed maize is milled to a gritty textured product, called *masa*. This masa is rolled into flat cakes and baked in an oven for 1–2 min to produce the traditional tortilla. Masa can also be deep-fried to produce tortilla chips or maize chips, or dried and finely milled to produce masa flour. The specific conditions and procedures for alkali processing maize vary with each processor.

Table 3 Wet milling yields *(percentage dry-weight basis)* for several types of maize and the effect of drying temperatures on wet milling yields

	Laboratory milling			*Commercial*		
	LTD Waxy	*HTD High amylose*	*LTD High amylose*	*LTD Medium-soft dent*	*HTD Medium-soft dent*	*Dent*[a]
Starch	59.0	29.8	45.0	65.2	62.5	67.5
Gluten	14.9	37.3	26.6	10.2	12.6	5.8
Germ	7.2	7.4	7.4	7.0	7.0	7.5
Fiber	10.2	18.1	11.3	9.3	10.8	11.5
Steepwater	5.0	4.5	5.3	4.4	3.9	7.5
Filtrate	3.2	2.5	4.0	2.6	2.5	

HT, high-temperature-dried; LT = low-temperature-dried.

[a]From Anderson RA and Watson SA (1982) The corn milling industry. In: Wolff IA (ed.) *CRC Handbook of Processing and Utilization in Agriculture, Vol. II: Part 1 Plant Products* Boca Raton, FL: CRC Press.

The quality of maize used in the alkali process is very important. The largest US processor desires hard endosperm yellow dent maize in order to provide a uniform endosperm texture for cooking. Mechanical damage to the kernel results in increased rates of water absorption and thus increased rates of cooking. Damaged or stress-cracked kernels tend to be overcooked, and they leach out more starch into the nejayote (steep water). Drying temperature is also important as high-temperature drying increases stress-crack damage and damages the starch prior to cooking. The use of hard-endosperm maize is not required for masa production, and quality masa products can be produced from high-quality soft endosperm or average dent maize with appropriate changes in processing conditions.

High-lysine hard-endosperm maize, called quality protein maize (QPM), is being developed to increase the nutritive value of alkali-cooked products. Maize protein as a total protein source is deficient in the amino acids tryptophan and lysine. The QPM is ~ 20–35% higher in these amino acids, and alkali processing of QPM results in masa products with improved levels of both amino acids. (*See* **Amino Acids**: Properties and Occurrence.)

See also: **Ethnic Foods**; **Milling**: Principles of Milling; Types of Mill and Their Uses; Characteristics of Milled Products

Further Reading

Anonymous, World of Corn – 2001, National Corn Growers Association, http://www.ncga.com/03world/main/biotechnology.html

Anderson RA and Watson SA (1982) The corn milling industry. In: Wolff IA (ed.) *CRC Handbook of Processing and Utilization in Agriculture, Vol. II: Part 1 Plant Products*, Boca Raton, FL: CRC Press.

Berkhout F (1976) Maize starch. In: Radley JA (ed.) *Starch Production Technology*, pp. 109–134. London: Applied Science.

Bridges A (2001) The impact of GMO grains. *Milling Journal* July/August/September: 52–54.

Christensen CM (ed.) (1982) *Storage of Cereal Grains and their Products*, 3rd edn. St. Paul, MN: American Association of Cereal Chemists.

Earle FR, Curtis JJ and Hubbard JE (1946) Composition of the component parts of the corn kernel. *Cereal Chemistry* 23(9): 504–511.

Eckhoff SR and Paulsen MR (1996) Maize. In: Henry RJ and Kettlewell PS (eds) *Cereal Grain Quality*, pp. 77–112. London: Chapman & Hall.

Grooms L (2001) Integrated technologies being applied to 'value preservation'. *Seed and Crops Digest* November: 34–38.

Johnson LA (1991) Corn: production, processing, and utilization. In: Lorenz KJ and Kulp K (eds) *Handbook of Cereal Science and Technology*, pp. 55–132. New York: Marcel Dekker.

Johnson LA, Hardy CL, Baumel CP and White PJ (2001) Identifying valuable corn quality traits for starch production. *Cereal Foods World* 46(9): 417–423.

Johnson LA, Hardy CL, Baumel CP, Yu T-H and Sell JL (2001) Identifying valuable corn quality traits for livestock feed. *Cereal Food World* 46(10): 472–474, 476–478, 480–481.

Knight JW (1969) *The Starch Industry*. Oxford: Pergamon Press.

Sprague GF and Dudley JW (eds) (1988) *Corn and Corn Improvement*, 3rd edn. Madison, WI: American Society of Agronomy.

Watson SA and Ramstad PE (eds) (1987) *Corn Chemistry and Technology*. St. Paul, MN: American Association of Cereal Chemists.

Whistler RL, BeMIller JN and Paschall EF (eds) (1984) *Starch: Chemistry and Technology*, 2nd edn. Orlando, FL: Academic Press.

Wolf MJ, Buzan CL, MacMasters MM and Rist CE (1952) Structure of the mature corn kernel (I-IV). *Cereal Chemistry* 29(5): 323–382.

MALABSORPTION SYNDROME

J F T Glasgow, Queen's University of Belfast, Belfast, Northern Ireland, UK

Introduction

Malabsorption syndromes are characterized by the clinical triad of chronic diarrhea, abdominal distention, and failure to thrive or grow. Until about 40 years ago the principal recognized disorders were celiac disease (gluten-sensitive enteropathy), tropical sprue, and cystic fibrosis, conditions in which fat malabsorption was the predominant feature. Since then, advances in basic concepts, such as lipolysis at the dietary lipid interface, micellar solubilization, sodium-coupled solute transport, together with the widespread use of diagnostic intestinal biopsy and an appreciation of the brush border as a digestive

organelle, have allowed the generic syndrome (of malabsorption) to be subdivided into disorders affecting specific stages in the physiological sequence of nutrient assimilation. The term malabsorption syndrome has therefore come to include a wide variety of clinical disorders of the following:

1. Intraluminal digestion (strictly speaking, maldigestion syndromes).
2. Intestinal absorption.
3. Carbohydrate absorption.
4. A number of specific defects in intestinal uptake or transport have also been described. These concern elements (zinc, magnesium, copper), amino acids (cystinuria, Hartnup disease, blue diaper, oasthouse, Lowe's and Joseph's syndromes; lysinuric protein intolerance, etc.), and hematinics (vitamin B_{12}, folic acid), but are not usually associated with stool abnormality.
5. There is also a miscellaneous group of disorders in which the cause of the diarrhea is uncertain. Here the only condition of clinical importance is so-called toddler's diarrhea. This group also includes very rare malabsorption states such as that caused by neuroendocrine-secreting tumors (e.g., ganglioneuroma). (*See* **Celiac (Coeliac) Disease; Cystic Fibrosis.**)

Disorders of Intraluminal Digestion (DID) (Table 1)

Diarrhea caused by exocrine pancreatic insufficiency (EPI) is the principal cause of DID. The physical characteristics of the stools are striking. They tend

Table 1 Disorders of intraluminal digestion

Pathophysiology	*Diagnosis*	*Supporting evidence*
Maldigestion of all macronutrients		
Exocrine pancreatic insufficiency	Cystic fibrosis	Sweat chloride > 65 mmol l^{-1}
	Pancreatic hypoplasia (lipomatosis) (Shwachman-Diamond syndrome)	Neutropenia, metaphyseal dysplasia
	(Johanson-Blizzard syndrome) (fibrosis)	Ectodermal defects
	(Pearson syndrome)	Sideroblastic anemia
	Severe malnutrition	Clinical circumstances
	Pancreatic trauma (pseudocyst)	History of abdominal trauma
	Chronic pancreatitis	Clinical history, plasma amylase
	Multiple endocrine deficiency, autoimmune candidiasis syndrome	Hypoparathyroidism or other endocrinopathies
	Cystinosis	Tubular dysfunction
	Advanced celiac disease	Table 2
Inactivation of pancreatic enzymes (Zollinger-Ellison syndrome; ZES)		Peptic ulcers, fasting serum gastrin > 100 pg ml^{-1}
Maldigestion of fat alone		
Pancreatic secretion	Congenital lipase or colipase deficiency	Assay in duodenal contents
Bile salts		
Synthesis-secretion	Extrahepatic biliary atresia	Cholestasis, liver biopsy
	Intrahepatic bile duct hypoplasia	Pulmonary artery stenoses, embryotoxon
	Various syndromes of infantile cholestasis	
	Hepatic cirrhosis	
	Preterm or mature newborn	Liver biopsy
	Impaired bile acid synthesis	
		Assay bile salts (ABS) duodenum, blood, stool
Deconjugation	Bacterial overgrowth proximal small bowel (obstruction, anastomoses, disturbed motility, blind loops)	Clinical history, H_2 breath test, trial of lincomycin
Enterohepatic circulation broken	Ileal resection(s) fistulae, Crohn's disease	Clinical history, radiology, tissue biopsy
	Congenital malabsorption of bile salts	ABS duodenum, blood, stool
Precipitation, etc.	ZES	See above
	Drugs: neomycin, laxatives, cholestyramine, sulphonamides	Clinical history
Maldigestion of proteins alone	Congenital enterokinase deficiency	Assay in duodenal mucosa
	Congential trypsinogen defiency	Assay in duodenal contents

to be loose and homogenous, pale and pasty with an offensive, penetrating, cheesy odor. These are often greasy and patients describe undigested dietary fat oozing like oil from the passed stool or floating on the water surface in the toilet. Under the microscope, fat droplets and partially digested meat fibers are seen. Balance studies reveal a more bulky stool than normal, containing large amounts of fat, nitrogen, and volatile fatty acids. The coefficient of fat absorption varies from 40% to 70%. Huge nitrogen losses of up to 3–4 g day^{-1} (normal < 0.5 g day^{-1}) are also recorded. Volatile fatty acids – acetic, butyric, and propionic – are usually > 30 mmol day^{-1} (normal < 10 mmol day^{-1}), although only trace or small amounts of lactic acid are present. These large fecal losses are typical of EPI, the commonest cause of which is cystic fibrosis. If the sweat chloride level is normal, however, the cause is likely to be congenital hypoplasia of the pancreas, several forms of which have been described, each with distinct extraalimentary features. (*See* **Fats**: Digestion, Absorption, and Transport; **Gallbladder.**)

Disorders of intraluminal digestion may occasionally affect only one class of nutrient. Congenital absence of pancreatic lipase results in a massive steatorrhea, which may be confirmed by enzymatic assay of lipase activity in aspirated duodenal contents. In isolated colipase deficiency steatorrhea is less severe.

Where the jejunal bile salts level falls below the critical physiological concentration, impaired micellar solubilization results in severe malabsorption of dietary fat and fat-soluble vitamins. Absorption of nitrogen and carbohydrate is normal. This is most usually seen in children with obstructive jaundice, such as extrahepatic biliary atresia or intrahepatic cholestasis (e.g., Alagille's syndrome). When cholestasis is complete or hepatocellular failure is present, fat excretion tends to be greater and may reach 20 g day^{-1} (normal up to 4.3 g day^{-1}). Fatty acids with one or more double bonds are malabsorbed, as well as saturated fats. In extensive Crohn's disease or surgical resection of the terminal ileum, there may be failure to absorb sufficient bile salts to maintain their enterohepatic circulation (and pool size), which is said to be broken. Inborn errors in both the synthesis and the absorption of bile acids have also been reported in occasional families. The Zollinger–Ellison syndrome (ZES), by markedly lowering jejunal pH and also the action of certain drugs (e.g., neomycin), interferes with the complex of events which affect intraluminal digestion. (*See* **Inborn Errors of Metabolism**: Overview.)

Isolated, massive proteolytic insufficiency has been described in the rare inherited disorder congenital enterokinase deficiency. This can be confirmed by finding reduced or absent proteolytic activity in duodenal fluid and assay of small-bowel mucosal enterokinase activity.

Disorders of Intestinal Absorption (DIA) (Table 2)

Although usually not fatty-looking, the stool is loose or watery, often with an acidic smell. Balance studies show that stool volume is more variable than in EPI and steatorrhea, and stool nitrogen losses are moderate (coefficient of fat absorption, 70–85%; nitrogen excretion, 1–2 g day^{-1}, respectively). Stool volatile fatty acids and lactic acid are modestly increased. Sheaves of fatty acid crystals may be appreciated microscopically.

Small-bowel capsule biopsy may reveal a nonspecific enteropathy with either a flat jejunal mucosa (subtotal or total villous atrophy) or a degree of partial villous atrophy (PVA). The degree of atrophy of the mucosa is but one feature of histological abnormalities which include the presence of enterocyte damage, crypt hyperplasia, cellular infiltration, and edema of mucosa and lamina propria. A flat jejunal mucosa in a child greater than 9–12 months of age is highly suggestive of celiac disease. This is confirmed by observing the clinical and histological response to exclusion of dietary wheat or rye gluten. The European Society for Paediatric Gastroenterology and Nutrition (ESPGAN) diagnostic criteria for celiac disease are as follows: (1) small-bowel mucosa usually flat; (2) clinical and histological improvement on a gluten-free diet; (3) histological, and sometimes clinical, relapse following dietary gluten challenge.

Partial villous atrophy is found in various disorders: under 6 months of age it is most likely to be caused by food protein intolerance(s) – most often cows' milk protein, less often to soya, wheat, rice, or multiple food proteins. Other causes include postinfective enteritis syndrome or *Giardia lamblia* infestation, where motile trophozoites may be seen in aspirated intestinal juice and histologically in the jejunal crypts. PVA has also been reported in various immunodeficiency states. Small-bowel bacterial overgrowth might be responsible, possibly linked to deconjugation of intraluminal bile salts. A very similar bowel lesion is found in infants with the syndrome of protracted diarrhea in which a combination of factors may be etiological – primary bowel infection, food intolerance, subadequate nutrition, a depressed immune status, bacterial overgrowth, etc.

A specific lesion on small-bowel biopsy is much more rarely found. In abetalipoproteinemia, mucosal architecture is generally normal but there is a dense accumulation of triglyceride droplets, especially

within mature enterocytes. This may be confirmed by the presence of peripheral blood acanthocytes and absent very-low-density lipoprotein (VLDL), low-density lipoprotein (LDL), and apolipoprotein B in plasma. There is severe impairment, if not virtual absence, of absorption of fat-soluble vitamins. Anderson's disease is a genetically distinct but similar disorder of chylomicron release. In each of these disorders early clinical features are similar to celiac disease. In intestinal lymphangiectasia, dysmorphic lacteals distort the normal villous contours and there is mild steatorrhea with evidence of a protein-losing enteropathy and lymphopenia. Infiltration by clones of abnormal or neoplastic cells (α-chain disease, histiocytosis X, tumor) has occasionally been reported. Crohn's disease may also involve the proximal small bowel. (*See* **Lipoproteins**.)

Disorders of Carbohydrate Absorption (DCA) (Table 3)

The DCA form a heterogeneous group of disorders, the common feature of which is carbohydrate malabsorption. The stool contains large amounts of volatile fatty acids. It is the presence of these small, osmotically active molecules (stool osmolality is greater than

Table 2 Disorders of intestinal absorption

Pathophysiology	*Diagnosis*	*Supporting evidence*
Intestinal biopsy: nonspecific inflammatory lesion		
Flat mucosa (total or subtotal villous atrophy)	Celiac disease	ESPGAN criteria (see text)
Partial villous atrophy (in severe cases villous atrophy may be subtotal)	Food protein intolerance (various – CMP, wheat, rice, etc.)	Response exclusion diet and challenge (see Table 3)
	Postenteritis syndrome	Clinical history
	Dermatitis herpetiformis	Dermal IgA deposit
	Giardia lamblia infestation	Trophozoites in jejunal juice/biopsy
	Immune deficiency states (hypogamma; SCIDS)	Plasma Ig, etc.
	Radiation damage, drugs (e.g., methotrexate, neomycin)	Clinical history
	Eosinophilic gastroenteropathy	Anemia, eosinophilia, jejunal biopsy
	Graft-versus-host disease	Clinical history
	Intractable diarrhea	Clinical history
Intestinal biopsy: specific lesions		
Fat-filled enterocytes	Abetalipoproteinemia	Acanthocytes, absence plasma VLDL. LDL, apo-B
	Anderson's disease	Decreased plasma VLDL, LDL, apo-B
Villi distorted by ectatic lymphatics	Lymphangiectasia	Protein-losing state, lymphopenia
Mucosal infiltration	α-Chain disease	Monoclonal abnormal plasma IgA
	Histiocytosis X	CT scan brain, tissue biopsy
	Crohn's disease, lymphoma, tumor	Clinical history, tissue biopsy
Intestinal biopsy: normal	Lysinuric protein intolerance	Dibasic aminoaciduria, osteoporosis
	Congenital chloride-losing diarrhea	Hydramnios, stool electrolytes
	Specific disaccharidase deficiencies	Table 3
	Transport defects: amino acids; zinc, copper; vitamin B_{12}, folic acid	Not part of malabsorption syndrome triad

CMP, cows' milk protein; SCIDS, severe combined immune-deficiency syndrome; IgA, immunoglobulin A; VLDL, very-low-density lipoprotein; LDL, low-density lipoprotein; apo-B, apolipoprotein B; CT, computed tomography.

Table 3 Disorders of carbohydrate absorption

Pathophysiology	*Diagnosis*	*Supporting evidence*
Intestinal biopsy: normal or almost normal	Congenital sucrase–isomaltase, congenital lactase, late-onset lactase, and congenital trehalase deficiencies	Assay disaccharidases jejunal mucosa; one specific activity affected
	Congenital glucose–galactose deficiency	In Ussing chamber, absence of glucose-induced short-circuit current
Intestinal biopsy: nonspecific inflammatory lesion	Acquired carbohydrate intolerances, especially celiac disease, cows' milk protein intolerance, postenteritis syndrome	All mucosal disaccharidases affected; Table 2

$2[Na^+ + K^+]$) that determines stool volume, which is therefore largely proportional to the amount of carbohydrate ingested. Stool lactic acid is increased and, accordingly, its pH is < 5.5. Large amounts of hydrogen, generated during the process of colonic fermentation and passed as flatus, form the basis of the breath hydrogen test. Stool fat and nitrogen tend to be normal. Reducing substances (sugars) can be detected by adding equal volumes of stool and tap water to a Clinitest tablet. A color change toward orange indicates the presence of up to 2% reducing sugar. (Sucrose is not a reducing sugar and must first be hydrolysed by $0.1\,mol\,l^{-1}$ hydrochloric acid.) (*See* **Carbohydrates: Digestion, Absorption, and Metabolism.**)

If, as is more likely, a nonspecific inflammatory lesion is present, causation has been discussed (Table 2). Disaccharidase and peptidase activities are reduced in the microvilli. Such secondary changes in celiac disease may be the main factor responsible for stool volume. A variety of specific inherited disorders of sugar absorption have also been described in the presence of microscopically normal jejunal mucosa. Specific enzymatic assay of mucosal disaccharidases, or more sophisticated studies, will confirm the diagnosis.

In the rare condition of congenital microvillous atrophy, electron microscopy reveals the presence of abortive microvilli as intracellular enterocyte inclusions with extreme stunting or absence of microvilli projecting into the lumen. As would be expected, carbohydrate intolerance, secretory diarrhea, and severe malabsorption result. In the majority of patients the outcome is fatal.

Clinical Features

The principal clinical expression of malabsorption is diarrhea. When this becomes a chronic symptom, i.e., of more than 14 days' duration, malnutrition and failure to thrive may follow. Depending upon etiology, nutritional failure may be compounded by symptoms such as poor appetite, anorexia, abdominal pain, or vomiting.

A careful clinical history will focus first upon stool number, size, color, smell, and fluidity. The pale, cheesy, homogeneous stools of cystic fibrosis, the loose, bulky motion of celiac disease, and the watery stool passed with much flatus of the child with sugar malabsorption have already been referred to.

The mode of onset of diarrhea is of special importance and may provide the most important pointer to diagnosis. For example, an onset at birth might suggest a congenital absorptive anomaly, such as chloridorrhea. Watery diarrhea which persists on treatment with intravenous fluid suggests a secretory diarrhea, such as congenital microvillous atrophy. Symptoms may be related to dietary changes: that from breast to cows' milk (cows' milk protein intolerance), introduction of cane sugar (sucrase–isomaltase deficiency), or weaning on to wheat- or rye-containing foods (celiac disease). There may be a close family history of similar patients. Persisting diarrhea may follow acute gastroenteritis and suggest postinfective enteritis syndrome with accompanying dietary sugar or protein intolerances.

Associated symptoms can be important: for example, weakness, anorexia, or vomiting may suggest celiac disease; increased appetite or pulmonary symptoms cystic fibrosis; dyspepsia, abdominal cramps, perianal pain, or discomfort may suggest Crohn's disease; bloating, abdominal cramps, or discomfort may suggest sugar malabsorption. Extraintestinal complaints are especially important in cystic fibrosis, Crohn's disease, and immune deficiencies.

The aims of the physical examination are essentially: (1) to determine clinical pointers toward a specific diagnosis and (2) to assess the degree of malnutrition. The general appearance of the child is possibly of greatest benefit, together with comparison of height (or length) and weight with charts of normal growth. Parental height is relevant in short children, together with any previous known weight or height measurements (calculate growth velocity) which may show age-related evolution of the condition. The Tanner stages of puberty may indicate delay in sexual maturation.

A measure of malnutrition is provided by the degree of weight loss, best seen as wasting of subcutaneous fat and/or muscle bulk around the groins, buttocks, shoulder girdles and neck. Malnourished children under 2 years of age often show delayed psychomotor development. Abdominal distention should be assessed in profile, allowing for the toddler's exaggerated lumbar lordosis. The presence of dehydration, edema of face or extremities, skin or conjunctival pallor, hair quality, and length of eyelashes should be assessed. The presence of chest deformity, finger clubbing, perianal abnormalities (e.g., excoriation, superficial ulceration, edema, fissures, fistulae, tags), skin rashes (e.g., erythema nodosum) and arthritis may be of great diagnostic assistance. The oropharynx should be carefully examined. Cheilosis and aphthous ulceration are seen in celiac disease, while in Crohn's disease there may be shallow ulcers, furrowed tag-like nodules, or swollen, inflamed, fissured lips.

In spite of much evidence of protein and energy malnutrition in the malabsorption syndromes, evidence of vitamin deficiencies is surprisingly uncommon, with the exception that in prolonged

obstructive jaundice clinical stigmata of fat-soluble vitamin deficiency are often found, e.g., defective dark adaptation, osteomalacia, bone pain, fractures, and a coagulopathy. Neurological features – loss of deep tendon reflexes and proprioception with ataxia, nystagmus, ophthalmoplegia, and retinitis pigmentosa – have been seen as a result of protracted deficiencies of vitamins E and A. Similar features have been reported in abetalipoproteinemia and, less commonly, in cystic fibrosis. In chronic small-bowel disorders (celiac disease, Crohn's disease), there may be angular stomatitis and other evidence of vitamin B complex deficiencies.

Two very common and confusing disorders are toddler's diarrhea (nonspecific chronic diarrhea) and the postenteritis syndrome. The former is characterized by persistent bouts of loose, mucusy, foul-smelling stools which often contain undigested vegetable material, punctuated by periods of normal or infrequent stools. Absorption and nutrition, well-being, and growth are normal. Postenteritis syndrome follows mucosal damage by enteric viruses or bacteria, and is in part caused by diminished enzyme activity in the damaged enterocytes. It persists until a new generation of normal enterocytes has been developed, a process which takes much longer in the malnourished patient.

Unless a sweat test or a jejunal biopsy reveals a diagnosis of cystic fibrosis or celiac disease, more detailed studies will be required. Careful 3-day balance studies of net intake and excretion, although used less frequently than in former times, may be necessary to identify malabsorption together with specific investigations set out in Tables 1–3.

The Consequent Malnutrition

Malnutrition varies a great deal and will depend on both the child's age and the diagnosis. The greater the age-related demands of growth, the more severe will be the effects upon nutrition, particularly in the first 2–3 years of life. In general, macronutrient malabsorption is more severe when the pathophysiological abnormality occurs in the intraluminal phase of the assimilation sequence. Thus EPI is associated with greater fecal losses of fat and protein than are celiac disease or disorders of lymphatic transport. The same cannot be said of micronutrient malabsorption, which tends to be relatively more severe in DIA, e.g., in celiac disease or extensive small-bowel Crohn's disease. Fat-soluble and probably water-soluble vitamins, minerals (iron, calcium), folic acid, vitamin B_{12}, and trace elements (zinc, magnesium) are poorly absorbed. In a specific, inherited condition such as abetalipoproteinemia there is profound, refractory deficiency of all fat-soluble vitamins (see above). (*See* **Malnutrition**: The Problem of Malnutrition *and also* individual nutrients.)

It should be pointed out, however, that important factors other than maldigestion or malabsorption contribute significantly to the degree of nutritional failure:

1. Poor dietary intake is probably the most important factor and is the result of poor appetite, altered taste, dyspepsia, vomiting, and abdominal discomfort or frank abdominal pain. These symptoms are relatively common in celiac disease and small-bowel Crohn's disease. In DCA a poor diet probably accounts for virtually all of the malnutrition.
2. Cystic fibrosis is an example of a disorder in which, in early life, the child may compensate for the considerably increased fecal losses by an increase in dietary intake. With the development of bronchopulmonary infections, however, appetite wanes, growth velocity diminishes, weight is lost, and this is compounded by increased work of breathing and, perhaps, elevation of resting energy expenditure.
3. Protein-losing states are associated with a threefold increase in the fractional catabolic rate of the plasma albumin pool which is not compensated by increased hepatic synthesis of up to 24%. This may be a factor in, for example, up to one-third of children with celiac disease; it is also common in Crohn's disease, intestinal lymphangiectasia, and eosinophilic gastroenteropathy. (*See* **Protein**: Synthesis and Turnover.)
4. Chronic inflammation has been shown in Crohn's disease to increase protein turnover with a consequent increase in resting metabolism.
5. Small-bowel bacterial overgrowth leads to further nitrogen loss and hypoalbuminemia, and results in sugar intolerance and steatorrhea.

In addition to malabsorption, each of these developments result in protein malnutrition with consequent lowering of plasma levels of insulin-like growth factor 1 (IGF-1, somatomedin C, basic somatomedin). This possibly explains why the growth rate is usually reduced, even though secretion of growth hormone, thyroid, and adrenal hormones is unimpaired. Where malnutrition is severe or long-lasting, IGF-2 may eventually be depressed also. The precise means by which nutrition influences IGF-1 production in liver and other tissues is unclear, although intake of protein and energy appears to be important. None the less, IGF-1 seems to be more sensitive than other commonly used indices of malnutrition, such as prealbumin, retinol-binding protein, or transferrin.

Likewise, when nutrition is corrected, IGF-1 increases more rapidly and to a greater degree than the other proteins. Clinically, serial measurement of IGF-1 has been proposed as a means of following nutritional progress in gastrointestinal disorders, such as chronic inflammatory bowel disease. (*See* **Colon**: Diseases and Disorders.)

Weight loss (wasting) of course occurs before reduction in height or length velocity results in stunting. The effects are seen most obviously in the infant and toddler. In the past, children occasionally presented with celiac disease in a state of crisis (vomiting, hypoalbuminemia, hypovolemia) which had a significant mortality. In older, untreated children, the effects on growth are less dramatic, although there tends to be a progressive falling away from the normal range of weight- and height-for-age. In adolescents this is associated with delayed pubertal development, psychological, and peer-group maladjustment and, ultimately, a reduced adult height. Reference has already been made to the specific effects of vitamin and mineral deficiencies.

Nutritional Management

Nutritional management should begin with diagnosis and thence a rational approach to management. Where there is deficiency, this is replaced, as for example in cystic fibrosis (pancreatic enzyme supplements must match the load of macronutrient being ingested, whereas in congenital enterokinase deficiency a small amount of enzyme (trypsin) is sufficient to activate pancreatic proenzymes secreted into the duodenum). Alternatively, the substrate is avoided or altered, as in congenital sucrase–isomaltase deficiency (dietary sucrose omitted); in congenital glucose–galactose malabsorption, dietary glucose and lactose are replaced by fructose. In celiac disease, a gluten-free diet is prescribed and it must be both strict and lifelong, provided that the diagnosis has been made according to ESPGAN criteria. (*See* **Carbohydrates**: Metabolism of Sugars.)

In many malabsorption disorders, however, management is nonspecific, and consists of dietary manipulations which ameliorate the pathophysiology and correct malnutrition by minimizing the degree of malabsorption. For example, in disorders in which there is sub-optimal small-bowel bile salt concentration (Table 1), a large proportion of dietary long-chain triglycerides (LCTs) are substituted by medium-chain triglyceride (MCT) which fully corrects the steatorrhea. In practice, this is simpler in infants, where a variety of commercially available formulae provide a feed with up to 80% of the lipid as MCT. It is important, however, to insure that the diet contains adequate amounts of essential fatty acids, either in the form of LCT or as a supplement of linoleic acid (e.g., safflower oil). Generous oral supplements of fat-soluble vitamins should be prescribed, but if this fails to produce normal plasma levels, parenteral preparations, especially of vitamins A and E, should be administered.

Where there has been extensive ileal resection or pathology which interferes significantly with the enterohepatic circulation of bile salts, cholestyramine may be required in addition to MCT, in order to prevent the production of disabling cholerrheic diarrhea; this results when large quantities of bile salts pass into the colon and inhibit electrolyte and water absorption. When cholestyramine is being used, supplements of fat-soluble vitamins and folic acid are required since these will be unavailable for absorption as they become bound to the anionic resin; supplements should be administered at times other than when the drug is being taken.

In the small, severely malnourished infant or toddler, prior to, or immediately after, diagnosis, there is often a need to replace enteral nutrition with a balanced intravenous supplementation mixture (amino acids, glucose, fat, electrolytes, minerals, vitamins, and trace elements). The classic example is the baby with intractable diarrhea (Table 2) for whom survival may depend upon this approach. Supplementing the oral intake using a glucose polymer, usually with a source of lipid (often rich in MCT), vitamins, and minerals, is reserved for those who are less ill and able to tolerate the extra oral intake (e.g., those with cystic fibrosis or celiac disease). Nasogastric feeding may be necessary if poor appetite, nausea, or other problems interfere with achieving a satisfactory nutrient intake. When longer-term supplementation is required, especially in older patients, a percutaneous, endoscopically positioned gastrostomy may be employed.

Such nutritional additives need to be used judiciously, starting with smaller volumes in order not to increase the diarrhea already present – as when excessive amounts of glucose polymer are used. Caution is needed in infants with advanced liver disease, or those with hypoalbuminemia, lest the feed volume per se precipitate fluid retention. Supplements of intravenous salt-poor human albumin may be necessary.

Finally, knowledge of a specific pathophysiological abnormality may suggest an appropriate therapy. For example, patients with abetalipoproteinemia, Anderson's disease and congenital lymphangiectasia respond well to oral MCT substituted for LCT. This relates to the fact that MCT is absorbed directly into the portal venous system rather than via the small-bowel

lacteals. In addition, supplements of fat-soluble vitamins are required both to prevent nutritional, particularly neurological sequelae and to treat those who are already showing clinical deficiency.

In disorders of carbohydrate absorption, avoidance of the offending dietary sugar is usually sufficient, although rehydration may also be necessary, together with the general management of the severely malnourished child.

See also: **Carbohydrates**: Digestion, Absorption, and Metabolism; Metabolism of Sugars; **Celiac (Coeliac) Disease**; **Colon**: Diseases and Disorders; **Cystic Fibrosis**; **Fats**: Digestion, Absorption, and Transport; **Gallbladder**; **Inborn Errors of Metabolism**: Overview; **Lipoproteins**; **Malnutrition**: The Problem of Malnutrition; **Protein**: Synthesis and Turnover

Further Reading

Anderson CM, Burke V and Gracey M (eds) (1987) *Paediatric Gastroenterology*, 2nd edn. Melbourne: Blackwell Scientific Publications.

Lebenthal E (ed.) (1981) *Textbook of Gastroenterology and Nutrition in Infancy*. New York: Raven Press.

Walker WA, Durie PR, Hamilton JR, Walker-Smith JA and Watkins JB (1991) *Pediatric Gastroenterology Disease – Pathophysiology, Diagnosis, Management*. Philadelphia: BC Decker.

MALNUTRITION

Contents

The Problem of Malnutrition

K K Sinha, T.M. Bhagalpur University, Bhagalpur, India

Background

Malnutrition literally means 'bad nutrition' and refers to the imbalance between the body's supply and demands for nutrition. Thus, malnutrition can be defined as 'an impairment of health resulting from a deficiency, excess, or imbalance of nutrients.'

It can be classified further into 'undernutrition' and 'overnutrition.' Undernutrition deals with the deficiency of nutrients and their consequences, while overnutrition relates with cases of consumption of excess nutrients and their effects. In this communication, the problem of malnutrition has been discussed under various heads, viz. causes and signs of malnutrition as well as the important diseases induced by the deficiency of essential nutrients.

Causes of Malnutrition

Overnutrition

Problems resulting from overnutrition are known to be comparatively less common than the problems associated with undernutrition. So many persons consume diets that are excessively high in calories, saturated fat, cholesterol, and sugars and that are also excessively refined. Each of these excesses is supposed to be one of the major factors for increasing the risk of chronic diseases.

Excessive caloric intake leads to obesity, which is highly prevalent in urban population. Obesity can also be seen in persons who sit in their offices for longer duration. In fact, these persons may not take food in excess of their requirements, but the poor physical activities fail to consume the energy they receive daily. In these circumstances, the normal diet may prove to be an overdiet, and the extra energy made available goes for deposition. Obesity may lead to many chronic diseases like diabetes mellitus, gall bladder disease, gout, cardiovascular diseases, etc. Excessive intake of fats and sugars may also induce some of the chronic diseases described earlier.

Consumption of excess amount of saturated fats and cholesterol are believed to be among the important risk factors in the incidence of cardiovascular and cerebrovascular diseases, while excessive intake of salt may lead to hypertension. An increase in the incidence of gastrointestinal disorders like diverticulosis, irritable colon, and possibly colon cancer has also been reported to be caused by the consumption of excessive refined foods.

Vitamins A and D are essential for human health, but consumption of excessive amounts of these vitamins may lead to toxicosis, which is characterized by dry itchy skin, lethargy, fatigue, general weakness, nausea, vomiting, headache, etc.

Undernutrition

Many problems have been related to cases of undernutrition or nutritional deficiencies. Primary malnutrition is the result of an insufficient intake of essential nutrients, because of the unavailability of a food supply, ignorance, or poverty. The regional unavailability of food and crop failure are basic causes of undernutrition in most parts of the world. Likewise, poor practices in processing, marketing, and storage can result in a serious reduction of the available food supply as well as losses in the nutritive values of foods consumed. Some of the important causes for inadequate consumption of food that can be prevented by sufficient motivation, knowledge, and income are as follows:

- lack of knowledge about kinds and amounts of food to be taken at specific growth periods;
- lack of means to procure foods essential for good nutrition;
- poor and faulty environment at the time of food consumption;
- refusal to take food due to poor appetite caused by prolonged illness or allergy to some specific foods;
- lack of teeth and inability to chew certain nutritive foods; and
- loss of desire to eat good foods owing to loneliness.

Several causes of malnutrition in individuals are not necessarily the result of a deficient dietary intake but are caused by physiologic failure beyond the ingestion stage. For example, a nutritional deficiency may develop as a result of a failure to absorb normally an essential nutrient that is supplied by the diet in adequate quantities. Similarly, nutritional deficiency may be brought about by impaired digestion, abnormal intermediate metabolism, excessive excretion, and increased biologic requirements. Such secondary or conditioned deficiencies may occur in the case of gastrointestinal diseases with rapid emptying time, vomiting, and diarrhea; in pancreatic or biliary diseases that interfere with normal digestion and absorption; in liver diseases with impaired hepatic function or when rapid growth, closely spaced pregnancies, lactation, surgery, injuries, extensive burns, or febrile or metabolic diseases increase the body's requirement for one or more nutrients.

The conditions that predispose to nutritional deficiency are usually multiple and simultaneous in operation. In most of the underdeveloped areas of the world, widespread malnutrition is not just the result of one factor, poverty; ignorance, religious or cultural taboos, and intercurrent chronic parasitic infection also play an important compounding role. If more than one predisposing condition is working naturally, more than one nutritional factor is in deficient supply. Underweight preschool children are prevalent in different parts of the world (Table 1).

Thus, the individual who suffers from a well-defined deficiency of a single essential nutrient is the exception. As a rule, a number of deficiencies involving more than one dietary factor are at work concurrently, bringing about an impaired nutritional status and multifaceted disease picture simultaneously or sequentially. The occurrence of simultaneous multiple deficiencies complicates the nutritional rehabilitation of the malnourished individual considerably. The

Table 1 Estimated prevalence and number of underweight preschool children, 1995–2005

UN regions and subregions	*Prevalence of underweight (%)*			*Number of underweight (million)*		
	1995	*2000*	*2005*	*1995*	*2000*	*2005*
Africa	27.9	28.5	29.1	34.03	38.32	42.45
Eastern	33.2	35.9	38.7	13.42	16.47	19.48
Northern	14.8	14.0	13.2	3.11	3.08	2.99
Western	34.9	36.5	38.1	13.34	15.41	17.66
Asia	32.8	29.0	25.3	121.03	107.91	93.16
South Central	47.3	43.6	40.0	82.40	78.49	73.48
South-east	32.6	28.9	25.3	18.56	16.68	14.27
Latin America and the Caribbean	8.3	6.3	4.3	4.48	3.40	2.35
Caribbean	14.4	11.5	8.7	0.54	0.43	0.32
Central America	15.3	15.4	15.4	2.46	2.52	2.51
South America	5.7	3.2	2.3	1.96	1.08	0.80
Oceania	na	na	na	na	na	na
All developing countries	29.2	26.7	24.3	159.55	149.63	137.95

na, not available.
Source: ACC/SCN (2000) Fourth Report on the World Nutrition Situation. Geneva: WHO.

treatment of a single deficiency in a chronically malnourished patient is as effective in bringing about an improvement as the granting of a loan to settle a single debt of a business burdened by a long-standing accumulation of financial obligations.

Signs of Undernutrition

As discussed earlier, undernutrition leads to several classic deficiency diseases, which are more prevalent in the developing and underdeveloped countries and can be diagnosed easily on the basis of their specific symptoms. However, there are some common physical, behavioral, and pathological signs that can be observed in people affected by undernutrition. In fact, these are not the symptoms of any specific deficiency disease, but anybody can recognize the disease based on their personal observations. Those signs can be summarized as follows:

- frequent absence from school and work;
- lower rate of weight gain;
- poor appetite;
- dull, dry, and thin hair, which can be pulled out easily;
- dry-looking face with inflammation in the nose, lips, and mouth; fissuring at the corners of the lips;
- red and inflammed eyes with congested eyelids;
- loss of enamel and decay in teeth;
- pale and rough skin with pigmented patches and scaling around the nose and ears;
- muscular weakness and flabby muscles;
- mental depression with apathetic behavior, often showing irritability, restlessness, nervousness, or sleeplessness;
- high susceptibility to infections as well as poor wound-healing.

Pathogenesis of Deficiency Diseases

It would be difficult to separate and pinpoint the individual stages, but the pathogenesis of a deficiency disease may be a continuous, progressive development, which, if not interrupted at some point by remedial action, may advance to the final full-blown clinical picture characterizing the specific disease.

This development starts with nutritional inadequacy because of primary or secondary deficiencies; eventually, the body resources of the nutrients in inadequate supply become depleted. When tissue depletion has reached a critical point, it interferes with the normal biochemical reactions in selected tissues or in the body at large. This will initially result in functional changes such as increased fatigue or abnormal gastrointestinal or neurologic function. As the nutritional deficiency continues, minor or eventually major anatomic lesions develop, and gross clinical signs and symptoms become manifest.

Problem of Undernutrition or Deficiency Diseases

Deficiency diseases can be grouped broadly into two categories: one caused by vitamin deficiency and another caused by a deficiency of other nutrients.

Disorders resulting from vitamin deficiency

Several important diseases in humans are caused by the deficiency of some common vitamins, viz. vitamins A, C, D, thiamin, riboflavin, niacin, etc. The names of the diseases, along with the symptoms and methods of treatment, are detailed below:

Beriberi Beriberi is principally a thiamin-deficiency disease that is quite common when pregnancy and lactation raise the individual's thiamin requirement, and breast-fed infants may acquire the symptoms of this disease, since the milky, malnourished mother may be deficient in thiamin.

Important symptoms of beriberi include a loss of appetite and disturbances in the gastrointestinal tract. Affected individuals feel extreme weakness, become lean and thin, develop cardiac trouble and edema, and show mental depression, loss of memory, anxiety, etc.

Because beriberi is a complex vitamin-deficiency disease, patients make the greatest improvement when B-complex vitamins, rather than thiamin alone, are prescribed. The persons may also consume a diet rich in thiamin or high in protein and calories.

Cheilosis and glossitis These are riboflavin-deficiency diseases (*ariboflavinosis*) that are common in persons who consume a marginal diet devoid of dairy products or other animal protein sources and leafy vegetables.

Cheilosis is characterized by the appearance of crack at the corners of the lips as well as bifurcation at certain points. Sometimes, blood also oozes out at those points. In the case of glossitis, small pimple-like eruptions arise on the tongue that become red and sometimes inflamed.

These diseases are treated by administration of riboflavin coupled with an improvement in the diet. Riboflavin supplementation should continue at the rate of 5–10 mg daily until complete recovery. A well-balanced diet rich in milk and milk products, meats, and leafy vegetables should accompany the specific vitamin therapy.

Keratomalacia, xerophthalmia, night blindness, and Bitot's spot All of these diseases are caused by a

deficiency of vitamin A (*avitaminosis A*) or its precursors, the carotenoids in the diet. These are the most prevalent vitamin-deficiency diseases in the world.

One of the earliest signs of vitamin A deficiency is the secretion of watery fluid from the eyes and, if unchecked or not treated in good time, leads to night blindness, or *nyctalopia*. The affected person is unable to see well in dim light, especially when coming into darkness from a brighter environment.

The term 'xerophthalmia' means 'dryness of the eye,' i.e., of the cornea and conjunctiva. This is prevalent mostly in young children. The full-blown syndrome of xerophthalmia is characterized by a wrinkled and lusterless conjunctiva and haziness and dryness of the cornea with small erosions or punctate superficial infiltrations. Subsequently, perforation occurs, and there may be a prolapse of the iris, or the lens may be expelled. Inflammation resembling conjunctivitis may also develop.

The above symptoms may lead to 'keratomalacia,' which is characterized by dryness and ulceration of the cornea as well as softening and destruction of the eyeball and blindness.

In some cases, a grayish shiny spot (Bitot's spot) appears on the conjunctiva, which also leads to xerosis of the cornea.

Vitamin A deficiency diseases can be prevented if carotene-rich foods are included in the diet. Treatment is rapidly effective with large doses of vitamin A, provided that the eye conditions have not become irreversible. Massive doses of vitamin A every 3–6 months given under medical supervision have been found to be successful in preventing high incidences of blindness.

Scurvy Scurvy results from a deficiency of vitamin C in the diet. It is common in infants and individuals who consume a diet devoid of fruits and vegetables, the main sources of vitamin C.

Pain, tenderness and swelling of the thighs and legs are frequent symptoms of infantile scurvy, accompanied by weight loss, fever, diarrhea, and vomiting. If the teeth have erupted, the gums are likely to be swollen, tender, and hemorrhagic. Scurvy in adults results after several months of a diet devoid of ascorbic acid. The symptoms include swelling, infection, bleeding of the gums ('*gingivitis*'), tenderness of the legs, anemia, and petechial hemorrhages. The teeth may become loose and eventually may be lost. There is degeneration of the muscle structure and of the cartilage, with impaired wound healing.

Scurvy patients respond dramatically and specifically to the administration of vitamin C. The patients can receive ascorbic acid either in synthetic form or in the form of fresh, unheated processed orange juice of standardized vitamin C content or in the form of noncitrus juices enriched with vitamin C.

Rickets and osteomalacia These are diseases resulting from a deficiency of vitamin D. In fact, a deficiency of vitamin D leads to inadequate absorption of calcium and phosphorus from the intestinal tract and to faulty mineralization of bone and tooth structures. The inability of the soft bones to withstand the stress of weight results in skeletal deformities.

Premature infants are more susceptible to rickets than full-term infants, since the growth rate and calcification of the skeleton impose additional demands for vitamin D. In early rickets, there may be few physical findings except irritability and restlessness during sleep. Closure of the cranial fontanelles is delayed, and the affected infant crawls or walks late. Softening of the skull ('craniotabes') and bulging or bossing of the forehead give the head a box-like appearance. Bones become soft and fragile, leading to a widening of the ends of the long bones, bowing of the legs, and enlargement of the costochondral junction with rows of knobs or beads ('*Rachitic rosary*'). In rachitic infants and children, there may be delayed dentition and malformation of the teeth.

Frequently referred to as 'adult rickets,' osteomalacia literally means 'bone softening.' In this disease, softening of the bones becomes so severe that the bones of the legs, spine, thorax, and pelvis bend into deformities. Rheumatic-type pain in the bones of the legs is also common.

As long as the dietary calcium and phosphorus intake is adequate, uncomplicated rickets may be treated solely by administration of sufficient doses of vitamin D. The first sign of improvement is a rise in serum phosphorus, which is commonly observed by the second week. This is usually followed by bone mineralization within 2–4 weeks of initiation of treatment.

Vitamin D deficiency diseases may also be prevented by adequate exposure of the infants to sunlight or by giving vitamin D in a variety of forms, such as cod-liver oil, other fish-liver oils, or vitamin D milk.

Pellagra 'Pellagra' is a term of Italian origin, which means 'rough skin.' The cause of pellagra is a severe deficiency of niacin and its precursor, the amino acid tryptophan. It is one of the important public health problems in many parts of the world.

Pellagra involves the gastrointestinal tract, the skin, and the nervous system. Primary symptoms include a sore tongue, mouth, and throat, with glossitis extending throughout the gastrointestinal tract. The tongue and lips become abnormally red in color. The mouth becomes so sore that it is difficult to eat and

swallow. Nausea and vomiting are followed by severe diarrhea. A characteristically symmetric dermatitis, especially on the exposed surfaces of the body (hands, forearms, elbows, feet, legs, knees and neck), appears. The dermatitis is sharply separated from the surrounding normal skin. At first, the skin becomes red, somewhat swollen, and tender, resembling a mild sunburn. If the condition is untreated, the skin becomes rough, cracked, and scaly, and may become ulcerated. Sunshine and exposure to heat aggravate the dermatitis. Some neurologic symptoms like confusion, dizziness, poor memory, and irritability, leading to hallucinations, delusions of persecution, and dementia, are also noted as the severity increases. The classic 'Ds' are the final stages of the disease – dermatitis, diarrhea, dementia, and death.

The patient should be confined to bed until convalescence is well under way. Specific therapy consists of the oral administration of a sufficient amount of niacinamide daily. In conjunction with specific niacinamide therapy and supportive multivitamin (particularly vitamin B complex) supplementation, a well-balanced diet is given, containing 100–150 g of high-quality protein (preferably from meat, milk, and eggs) and provides about 3500 cal. In case of gastrointestinal symptomatology, the diet initially should be a full liquid diet devoid of roughage. As the patient improves, a general diet high in good-quality proteins is given.

Disorders Resulting from Deficiency of other Nutrients

In addition to vitamins, there are many other nutrients whose deficiency in the diet can lead to several important human diseases. Some of these nutrients include iron, folic acid, iodine, calcium, phosphorus, protein, calorie, etc. The diseases resulting from a deficiency of these nutrients, along with the symptoms and methods of treatment, are as follows:

Anemia Anemia is a deficiency in the circulating hemoglobin and results from inadequate synthesis of hemoglobin or of red blood cells or both. It is caused by an inadequate supply of one or more nutrients such as iron, vitamin B_{12}, folic acid, vitamin C, and trace elements, or by many non-dietary factors such as blood loss, excessive blood destruction, or defects in blood formation.

Iron-deficiency anemia is the most common disease in different parts of the world and occurs more frequently in infants, preschool children, adolescent boys and girls, and women in childbearing years (Figures 1 and 2). Insufficient iron in the diet, faulty absorption of iron from the small intestine, and blood loss are causes of iron-deficiency anemia.

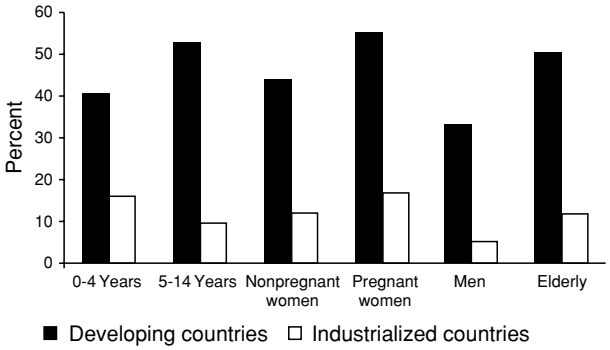

Figure 1 Prevalence of anemia by age group in industrialized and developing countries, 1998. Source: ACC/SCN (2000) Fourth Report on the World Nutrition Situation. Geneva: WHO.

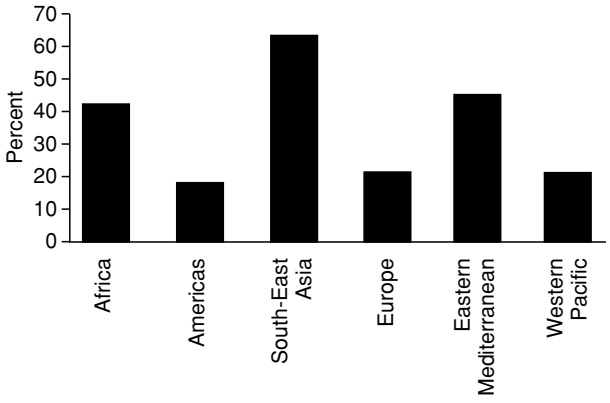

Figure 2 Prevalence of anemia in children 0–5 years old by WHO region, 1998. Source: ACC/SCN (2000) Fourth Report on the World Nutrition Situation. Geneva: WHO.

The symptoms of nutritional deficiency disease include weakness, fatique, pallor, dyspensia on exertion, and a constant feeling of tiredness. The onset is usually insidious. There may be vague gastrointestinal symptoms, and achlorhydria is not uncommon. The skin, mucous membrane, and nails are pale in proportion to the reduction in the circulating hemoglobin. Many adults with hypochromic anemia exhibit atrophy of the papillae of the tongue. There may be a slight cardiac enlargement, with eventual failure if the anemia becomes severe. Nails may be brittle, longitudinally ridged, or even concave and spoon-shaped. Individual blood cells are pale (hypochromic) and smaller than normal (microcytic).

Many instances of nondeficiency anemia can be prevented by appropriate diet. Once anemia is present, treatment requires iron salts such as ferrous sulfate, although attention should also be focused on correcting a faulty diet. Good sources of iron include meats, especially offal, whole-grain breads and cereals, as well as dark green leafy vegetables.

Anemia in infancy can be easily prevented by using an iron-fortified formula throughout the first year or introducing iron-rich foods, such as cereals, early and continuing their use throughout childhood. The requirements of pregnant women can be met only by using an iron supplement.

Kwashiorkor and marasmus These are diseases of protein-calorie malnutrition, which includes a wide spectrum of disturbances ranging from mild deficiency to the severe forms of kwashiorkor and marasmus. The disease is widely prevalent in the tropical and subtropical underdeveloped parts of the world and is a major health problem of young children among the poor strata of the population of those areas.

The term 'kwashiorkor' means 'the evil spirit that infects the first child when the second child is born.' In the developing and underdeveloped countries, babies are usually breast-fed until 18–24 months of age, and occasionally longer. When another baby is born, the older child generally subsists on the foods typical of the household – usually a high-carbohydrate, low-protein diet, and this results in the symptoms of kwashiorkor.

The symptoms start to appear after a few months of weaning. The increase in stature and tissue development is retarded, and the muscles become flabby and lose their tone. The severe protein levels are reduced so that the tissues hold water, especially evident by the swollen face and legs, and the large potbelly. The hair often changes color, appears lifeless, and can be easily plucked out. Diarrhea is frequently observed probably as a result of poor sanitation in the home. The child with diarrhea advances rapidly from mild to severe kwashiorkor.

Marasmus is another protein-calorie-deficient disease prevalent mostly in the urban areas of developing countries, where infants are usually weaned from the breast at a much earlier age (before 1 year) than in rural areas. The result is severe growth failure and emaciation in the first year of life. Muscle wasting is extreme, and subcutaneous fat is absent, so the ribs become visible. Retarted mental development is likely in children suffering from marasmus, as deprivation of protein occurs at a time when the division of brain cells occurs most rapidly.

Treatment of the above diseases is based on the administration of proteins of high biological value and correction of any existing dehydration and electrolyte imbalance. A diet rich in milk protein has been the treatment of choice. Fruit juice, eggs, meat, vegetables, and cereals are gradually added to the diet throughout the period of recovery. Several inexpensive vegetable and animal protein mixtures have also been developed on a regional basis for human feeding that increase the quantity and improve the quality of dietary protein.

Endemic goiter Endemic goiter is an iodine-deficiency disease (IDD) that occurs in those areas where the iodine content of the soil is so low that insufficient iodine is obtained through food and water and occurs when no provision is made for supplying iodized salt.

A lack of iodine leads to an increase in the size and number of epithelial cells in the thyroid gland and thus an enlargement of the gland. This condition is known as simple or endemic goiter and presents no other abnormal physical findings. The deficiency is more prevalent in females than males and is more frequent during adolescence and pregnancy. Tables 2 and 3 list the number of countries affected and the current magnitude of IDD.

Consumption of iodized salt is the most common treatment of this disease. Some natural sources of iodine like sea foods (fish, crabs, etc.) and green vegetables grown in marine soil can also provide sufficient amounts of iodine in the body.

Table 2 Number of countries affected by IDD, 1999

	Total number of countries in region	*Countries where IDD is a public health problem*	*Countries where IDD has been eliminated[a] or is nonexistent*	*Countries with insufficient data[b]*
Region				
Africa	46	44	1	1
Americas	35	19	3	13
South-east Asia (including India)	10	9	0	1
Eastern Mediterranean	22	17	1	4
Europe	51	32	13	6
Western Pacific (including China)	27	9	2	16
Total	191	130	20	41

[a]IDD elimination is defined as a total goitre rate of < 5% in school-age children.
[b]Data are insufficient to categorize countries.
Source: ACC/SCN (2000) Fourth Report on the World Nutrition Situation. Geneva: WHO.

Table 3 Current magnitude of IDD, 1999

Region	*Population*[a] *(million)*	*Population affected by goitre*		*At-risk population*	
		Million	*Percentage of regional population*	*Million*	*Percentage of regional population*
Africa	612	124	20	295	48
Americas	788	39	5	196	25
South-east Asia	1477	172	12	599	41
Eastern Mediterranean	473	152	32	348	74
Europe	869	130	15	275	32
Western Pacific	1639	124	8	513	31
Total	5857	740	13	2225	38

[a]Based on UN Population Division 1997 estimates.
Source: ACC/SCN (2000) Fourth Report on the World Nutrition Situation. Geneva: WHO.

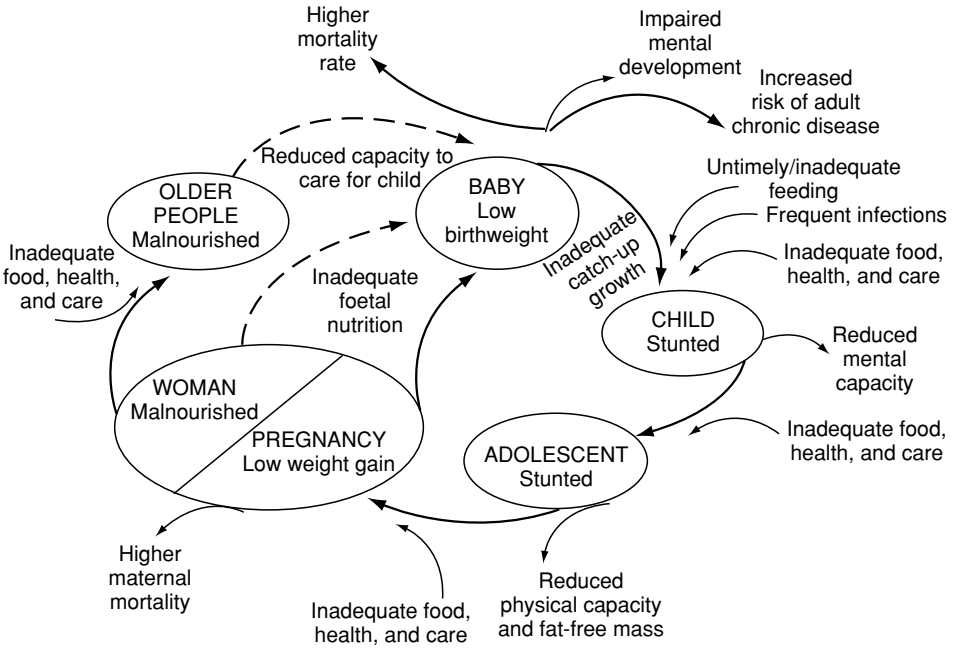

Figure 3 Nutrition throughout the life cycle. Source: ACC/SCN (2000) Fourth Report on the World Nutrition Situation. Geneva: WHO.

It is thus clear that nutrition challenges continue throughout the life cycle, as shown in Figure 3. Poor nutrition often starts *in utero* and extends, particularly for girls and women, well into adolescent and adult life. It also spans generations. Undernutrition occurring during childhood, adolescence, and pregnancy has an additive negative impact on the birthweight of infants. Low-birthweight infants who have suffered intrauterine growth retardation as fetuses are born undernourished and are at a far higher risk of dying in the neonatal period or later infancy. If they survive, they are unlikely to catch up significantly on this lost growth late and are more likely to experience a variety of developmental deficits. A low-birthweight infant is thus more likely to be underweight or stunted in early life.

The consequences of being born undernourished extend into adulthood. Epidemiological evidence from both developing and industrialized countries now suggests a link between fetal undernutrition and an increased risk of various adult chronic diseases – the 'fetal origins of disease hypothesis.'

See also: **Anemia (Anaemia)**: Iron-deficiency Anemia; **Ascorbic Acid**: Properties and Determination; **Carotenoids**: Occurrence, Properties, and Determination; **Cholecalciferol**: Properties and Determination; **Iodine**: Iodine-deficiency Disorders; **Malnutrition**: The Problem of Malnutrition; **Obesity**: Etiology and Diagnosis; **Riboflavin**: Properties and Determination; **Scurvy**; **Thiamin**: Properties and Determination; **Vitamins**: Overview

Further Reading

ACC/SCN (2000) *Fourth Report on the World Nutrition Situation*. Geneva: ACC/SCN.

Agarwal P (1998) *Human Nutrition, Meal Planning, Health Care and Common Diseases*, p. 359. Calcutta: Premier Publishing House.

Burton BT (1978) *Human Nutrition*, p. 530. New Delhi: Tata McGraw Hill.

Delange F, de Benoist B and Alnwick D (1999) Risks of iodine-induced hyperthyroidism after correction of iodine deficiency by iodized salt. *Thyroid* 9: 545–556.

Hibbert V (1973) The five possible causes of all nutrient deficiency. Illustrated by deficiencies of vitamin B_{12} and folic acid. *American Journal of Clinical Nutrition* 26: 77–88.

Pelletier DL, Frongillo EA and Habicht JP (1993) Epidemiological evidence for a potentiating effect of malnutrition on child mortality. *American Journal of Public Health* 83: 1130–1133.

Robinson CH (1978) *Fundamentals of Normal Nutrition*, p. 596. New York: MacMillan.

Suharno D and Muhilal (1996) Vitamin A and nutritional anemia. *Food and Nutrition Bulletin* 17: 7–10.

Tomkins A (2000) Malnutrition, morbidity and mortality in children and their mothers. *Procedures of the Nutrition Society* 59: 135–146.

Underwood BA (1998) Micronutrient malnutrition. Is it being eliminated? *Nutrition Today* 33: 121–129.

Malnutrition in Developed Countries

M Astier Dumas, Centre de Recherches Foch, Paris, France

Introduction

Although food is available to excess in developed countries, malnutrition is probably not infrequent, even in people free of disease. Malnutrition in developed countries is not the same as malnutrition in the developing world, where people are undernourished. In developing countries malnutrition is generally characterized by deficiencies, such as vitamin and mineral deficiencies.

Paradoxically, in developed countries, overnutrition leading to obesity, diabetes, or cardiovascular diseases can coexist with nutrient deficiencies, and fad diets can lead to an inadequate intake of nutrients. Whether these deficiences can really lead to diseases is not certain, but it is generally accepted that they are undesirable.

In developed countries it may be more accurate to speak of groups at risk of malnutrition than of malnutrition itself, and it is necessary to include overnutrition and undernutrition, depending on the groups concerned.

Correlations between fat and energy intake and obesity, diabetes, cardiovascular diseases, and cancer have been thoroughly investigated. Beyond all the discussions made on this point, there is general agreement that excess calories, whatever their source, can be a cause of disease ranging from cardiovascular disease to cancer and diabetes. Changes in the composition of food eaten, towards processed foods rich in fat, with a high energy density and poor nutritional content, may be an important factor, as vitamins and minerals are sometimes described as protective factors (mainly vitamins A, C, and E, β-carotene and selenium in studies of cancer and cardiovascular diseases). (*See* **Cancer**: Diet in Cancer Prevention.)

There are few data on the nutritional status of the obese subject compared to the nonobese, but in a French study red blood cell transketolase activity – a measure of thiamin status – was lower in obese subjects of both sexes. (*See* **Obesity**: Etiology and Diagnosis.)

Apart from diseases that produce malnutrition by altering either digestive processes or the metabolic use of nutrients (e.g., traumas, surgery, cancer, and burns), malnutrition in developed countries can affect poor people (often immigrants), old people, those on weight-reducing diets, youngsters, those who are unable to buy food or have difficulty eating it, those on special diets, or those who eat for fun rather than from hunger. **Table 1** gives an estimation of the

Table 1 Groups at risk of poor vitamin status in the US population (estimation)

Group	*Percentage of population group at risk*	*Nutrients*
Aged 10–19 years	14.6	Folates, vitamin A
Alcohol users	22	Vitamins B_1, B_6, A and D, carotene and folates
Smokers	28.6	Vitamins B_6, E and C, carotene and folates
Diabetics	4.6	Vitamins B_2, C and D
Dieters	20	Potentially any or all
Elderly people (aged over 65 years)	12.1	Vitamins C and D, folates
Oral contraceptive users	10.4	Vitamin B_6, folates, carotene
Pregnant women	3.9	All, especially folates
Strict vegetarians	> 0.9	Vitamins B_{12} and D

Reproduced from Gaby S *et al*. (1991) *Vitamin Intake and Health: A Scientific Review*. New York: Marcel Dekker, with permission.

percentage of population groups at risk of vitamin deficiency in the US population.

Evidence from Consumption Data

Many surveys have been made on food consumption levels, comparing the results with dietary reference values (DRVs) or recommended dietary allowances (RDAs) in the various countries. However, the fact that the RDA is not attained by a small fraction of the population does not mean that these people are malnourished. The RDAs are supposed to cover the needs of almost all the population, and a level of intake that is inferior by 20%, or even by 40%, can still be compatible with adequate nutrition. It is necessary to know the physiological status of these people in order to understand the meaning of an inadequate intake. For example, a serum level of a nutrient that is inferior to the reference level (biochemical index) is not automatically a sign of disease, but its significance is more reliable than the level of food consumption. In some instances the physiological signs of deficiency are difficult to obtain on a large scale in surveys. For example, identification of good bone health – the bone mass – is not easy to study, and calcium intake is therefore used to identify the population at risk. The percentage of the population with intakes of calcium well below the RDA is sometimes important: in Belgium, inadequate intakes were found in 67% of 13–19-year-olds, 73% of 9–12-year-olds, and 85% of 5–8-year-olds. It has been shown that the proportion of girls aged 11–19 years consuming less than 1200 mg of calcium per day is 91% in Portugal, 85% in the USA and 78% in Canada. Between 1980 and 1985 in the UK the mean daily intake of calcium fell from 960 mg per day to 850 mg per day; 33% of boys (10–11 years) and 53% of girls (same age), and 25% of boys and 57% of girls (14–15 years) were found to have intakes of calcium below 800 mg per day. Evidence suggests that the bone mass of an adult is largely influenced by the calcium intake during adolescence, indicating that these population groups are at risk of bone disorders later in life. (*See* **Dietary Surveys**: Measurement of Food Intake; Surveys of National Food Intake; **Nutritional Assessment**: Anthropometry and Clinical Examination; **Nutritional Surveillance**: In Industrialized Countries; **Dietary Reference Values**; **Dietary Requirements of Adults**.)

Evidence from Biochemical Data

Biochemical evidence given by serum levels is available for many nutrients, mainly vitamins but also for iron. (*See* **Nutritional Assessment**: Biochemical Tests for Vitamins and Minerals.)

Elderly People

The reasons for malnutrition in the elderly are numerous and can coexist. They include the following: lack of money to buy food; poor dentition; depression and loneliness; inadequate exposure to sunlight in housebound elderly people, leading to poor vitamin D status; lack of appetite; effects of drugs; loss of dexterity of hands, hindering food preparation; and, sometimes, disease. Some elderly people are on self- or medically prescribed diets unsuited to their needs. (*See* **Elderly**: Nutritional Status; Nutritionally Related Problems.)

Food intake is generally lower in institutionalized elderly people than in free-living elderly people. Drugs are more commonly taken and depression is more frequent in institutionalized elderly people. Results of biochemical studies in institutions show generally lower serum vitamin levels than for free-living elderly people. Nevertheless, some general deficiencies have been described in elderly people.

In the Dutch elderly population study published in 1990, anemia according to the criteria of the World Health Organization (WHO) was found in 4% of men and 1% of women. Frank nutritional deficiencies were not observed, but compared to younger adults a substantial percentage of the elderly had relatively lower values of folate (men, 30% below $5\,\text{nmol}\,l^{-1}$; women, 17% below $5\,\text{nmol}\,l^{-1}$), pyridoxal 5′ phosphate (men, 43% below $19\,\text{nmol}\,l^{-1}$; women, 13% below $19\,\text{nmol}\,l^{-1}$), total carotenoids (men, 22% below $1\,\mu\text{mol}\,l^{-1}$; women, 13% below $1\,\mu\text{mol}\,l^{-1}$), and 25-hydroxyvitamin D. It was underlined that 'since the health consequences of these low levels at old age are unknown, no conclusion can be drawn whether this indicates marginal status among these elderly.' Elderly people following a vegetarian diet were more at risk of iron, zinc, and vitamin B_{12} deficiencies.

In one study in France of 70 institutionalized elderly people, 24% of subjects had biochemical signs of malnutrition (low serum proteins and hemoglobin values). The energy consumption of these subjects was below the RDA, as it was for iron, calcium, and vitamin C, but not for protein. In another study of 204 institutionalized patients (convalescent elderly), 65 were found to have low serum albumin levels, 27 being severely malnourished. In 59 disease-free subjects in an institution it was found that 20% had serum folate deficiency. Later, 235 institutionalized subjects were studied and 72.8% found with folate deficiency and 11.4% with vitamin B_{12} deficiency.

In Belgium, the vitamin D status of 240 elderly subjects was analyzed and it was concluded that the deficiency of both the vitamin D substrate and hormone was frequent in the elderly population in Belgium.

In Italy, the results of a nationwide study on nutritional status of the elderly was published. Only 47% of subjects were found to be free from vitamin deficiencies, and three out of 10 had multiple vitamin deficiencies. Vitamin B_1 and vitamin B_2 deficiencies were most prevalent (25% and 20%), followed by folate (15%) and vitamin C (14%). Vitamin A and vitamin E deficiencies were negligible (6% and 2%). In Naples, low intakes of calcium, iron, and nicotinic acid, and a low status of vitamin B_1 for 20% of men and women, and low riboflavin status for 40% of men and 20% of women were reported.

In the USA, a biochemical study of 70 women found 14% of all subjects deficient or low in vitamin A, 17% low in vitamin B_1, and 15% low in vitamin B_2. Serum ascorbic acid levels were acceptable, but low or deficient levels of total serum protein and albumin were observed for 36% and 20% of the subjects; low hemoglobin and elevated total iron-binding capacity values were recorded for 19% and 40% of subjects. In other surveys in the USA, Sweden and Netherlands, a prevalence of iron-deficiency anemia, instances of potassium deficiency, and possible vitamin D and/or calcium deficiency – with or without osteoporosis – have been observed.

Obviously the situation can be worse in poor elderly people. Folacin, iron status, and hematological parameters of 193 people from urban low-income households in Florida were reported and, of these subjects, 60% could be classified as high risk for folacin levels (under $140\,ng\,ml^{-1}$) and 11% at medium risk ($< 160\,ng\,l^{-1}$). An incidence of 14% of anemia and 32% of leukopenia, with no iron deficiency, was found.

Adults

In one recent study in France, in a sample of 1039 subjects aged from 6 to 97 years, from a defined area, no major deficiency in vitamin A was reported – less than 3% of serum levels were in the range indicating high risk of anomalies, but approximately 20% of women showed moderate risk. For vitamin E, slight deficiencies are more frequent in young people. For vitamin C, around 10% of men and less than 5% of women can be considered at high risk, with three times more at moderate risk. For vitamin B_1, erythrocyte transketolase activity corresponding to a high risk of deficiency was found in 22% of adults. For vitamin B_2 the risk was judged high in up to 31% of women and 22% of men. For vitamin B_6, high risk was present in up to 16% of men and 25% of women. Only 60% of the population showed no biochemical defect related to any vitamin deficiency.

The authors of this report underlined the correlation between tobacco smoking, alcohol consumption, use of oral contraceptives, and vitamin status. (*See* **Alcohol**: Metabolism, Beneficial Effects and Toxicology; **Contraceptives: Nutritional Aspects**; **Smoking, Diet, and Health**.)

The most documented nutrient is iron. In Germany diminished iron body stores in a considerable proportion of blood donors were found. An incidence of 9% of completely exhausted iron stores in male blood donors was reported – a figure also observed in Scandinavia. An even higher rate was found in female blood donors (25%).

Women of Child-Bearing Age and Infants

Studies in North America and Scandinavia show that 1–6% of women of child-bearing age (for men in the same age range only 1–2%) are anemic. In France, in one group studied, 2–3% of women and up to 25% of pregnant women were anemic. In Germany, it was reported that 10% of pregnant women had iron deficiency during the first trimester and 50% had iron deficiency at the end of pregnancy.

Iron deficiency is also a problem in infants. A 10% incidence of anemia and a 40% borderline iron status was found in 10–12-month-old infants in France.

Rickets still exists in developed countries. Despite supplementation of infant formulae with vitamin D, and the drops prescribed daily to infants when this supplementation is not made, some groups at risk have been identified. These include the following: native Indians in Canada because of poor lifestyle, lack of exposure to the sun, and lack of vitamin supplementation; the Asian population in the UK; Turkish and Moroccan infants in the Netherlands; and, in France, infants with dark skin coming from Africa. (*See* **Infants**: Nutritional Requirements.)

Another population group at risk is that eating a macrobiotic diet. Fifty-three Caucasian infants aged 10–20 months were compared to 57 matched omnivorous infants in the USA. Of the children in the former group, 55% had rickets, and the 25-hydroxy-vitamin D level was $34\,nmol\,l^{-1}$ (SD 15.3) against $49.7\,nmol\,l^{-1}$ (SD 21.9) in the control group. (*See* **Macrobiotic Diets**.)

Children

Although some real cases of severe malnutrition (kwashiorkor) have been described in North America or Europe, these are limited to instances in which a deliberate deviation from normal infant practices is made, sometimes for pseudophilosophical purposes

(e.g., by sects). (*See* **Children**: Nutritional Requirements.)

Dutch children fed macrobiotic diets were compared to omnivorous children (controls) for iron, vitamin B_{12}, and folate status. Iron deficiency was observed in 15% of the macrobiotic group (none in the control group: $P = 0.003$). Plasma vitamin B_{12} concentrations in the macrobiotic group were far below those of the control group ($P < 0.001$), and plasma folate concentrations were higher in the macrobiotic group.

The main nutritional problem in children is dental decay. High consumption of sucrose in young children is a characteristic feature of developed countries, particularly in Anglo-Saxon countries, and less so in Latin countries. In a communication in Edinburgh in the UK, it was reported that in some children aged 2–4 years the proportion of energy derived from sugar was as high as 30%. The main source is fruit juices, generally considered as 'healthy' drinks.

A decrease of dental problems in children has been observed over the past few years and this is believed to result from: (1) supplementation of water or salt (depending on the country) with fluoride; (2) improved hygiene education; and (3) use of sugar substitutes. However, dental decay remains an important problem. (*See* **Dental Disease**: Role of Diet.)

An excess of sugars in children's diets can also lead to nutrient deficiencies, and one study found that 8% of children had a high risk of vitamin B_1 deficiency.

Biochemical data about the nutritional status of healthy children are infrequently obtained, but consumption studies are numerous. The pattern of consumption is rather similar in different countries, with a high percentage of sugar and fat, and a relatively low intake of calcium and iron.

The consequences of these inadequate intakes are not well known, although iron deficiency for pregnant and lactating women and their newborn babies can be a consequence of insufficient intakes during adolescence. Osteoporosis can be linked with development of insufficient bone mass during adolescence. It is also suggested that excessive fat intake during childhood and adolescence may be instrumental in the development of arteriosclerosis. (*See* **Atherosclerosis; Osteoporosis**.)

These studies suggest that malnutrition in the developed world is mainly the result of poor food selection in affluent societies. The shift of consumption from traditional basic foods towards convenience, value-added foods can lead to higher intakes of fat and sugar, and lower intakes of vitamins and minerals.

Information and nutrition education would seem to be the obvious approach to preventing these problems of malnutrition in developed countries. Refer to individual nutrient deficiencies.

See also: **Children**: Nutritional Requirements; **Dental Disease**: Role of Diet; **Dietary Reference Values**; **Dietary Requirements of Adults**; **Dietary Surveys**: Measurement of Food Intake; Surveys of National Food Intake; **Elderly**: Nutritional Status; Nutritionally Related Problems; **Infants**: Nutritional Requirements; **Macrobiotic Diets**; **Nutritional Assessment**: Anthropometry and Clinical Examination; Biochemical Tests for Vitamins and Minerals; **Nutritional Surveillance**: In Industrialized Countries; **Obesity**: Etiology and Diagnosis

Further Reading

Astier Dumas M and Lagrange Hanrez V (1982) L'alimentation des personnes agées: à propos d'une enquête dans les hopitaux de long séjour. *Médecine et Nutrition* XVIII(6): 359–376.

Bouillon RA (1987) Vitamin D status in the elderly: seasonal substrate deficiency causes 1,25-dihydroxycholecalciferol deficiency. *American Journal of Clinical Nutrition* 45: 755–763.

Dagnelie PC *et al*. (1989) Increased risk of vitamin B_{12} and iron deficiency in infants on macrobiotic diets. *American Journal of Clinical Nutrition* 50(4): 818–824.

ESVITAF (1986) Vitamin status in three groups of French adults. *Annals of Nutrition and Metabolism* 30: S1.

Gaby S *et al*. (1991) *Vitamin Intake and Health: A Scientific Review*. New York: Marcel Dekker.

Harrill I and Cervone N (1977) Vitamin status of older women. *American Journal of Clinical Nutrition* 30: 431–440.

Hercberg S (1991) Iron and folate deficiency anaemias. *International Child Health* II(2): 44–60.

INSERM (1983) Groups with high risk of iron deficiency in industrialized countries. INSERM Meeting, Paris, 27–28 May 1983. Institut National de la Santé et de la Recherche Médicale en France.

Lowik MR *et al*. (1990) *Nutritional Risk Assessment among Dutch Elderly People*. Report to the Netherlands Ministry of Welfare.

Myers Nutrition Symposium (1986) *Nutrition and Aging*. The 5th Bristol Myers Nutrition Symposium. New York: Academic Press.

MALT

Contents
Malt Types and Products
Chemistry of Malting

Malt Types and Products

M J Edney and M S Izydorczyk, Canadian Grain Commission, Winnipeg, Manitoba, Canada

Introduction

Malt is a natural food product produced by germinating grains and pulses. The three steps of malting, steeping, germinating, and kilning, can be altered to produce different types of malt. Barley malt is by far the most common malt, and its major use is in the brewing of beer. Barley malts are also used in the production of spirits, vinegar, bread, breakfast cereals, and several other minor products (Table 1). In many parts of the world, sprouted cereals and pulses are eaten directly for the improved availability of their nutrients.

Malting

Malting is the controlled germination of grain or pulses. Proper control of its three major steps, steeping, germination, and kilning, is essential if the desired end product is to be produced without excess losses. A range of malt products, with different constituents and, as a result, different values for end-users, can be produced by altering the conditions of the steps.

The steeping step hydrates the seed, which allows germination to commence. Germination allows the embryo to grow under controlled conditions, allowing the seed to develop enzymes that degrade or modify endosperm structure. Other enzymes hydrolyze and solubilize the reserve starch and protein of the endosperm. The amount of solubilized material, often referred to as 'extract' is an important quality parameter that is used to predict the amount of beer or spirits a malt can produce. Kilning is the final stage of malting, and it halts growth of the embryo by removing moisture. Temperatures and timelines used in kilning are especially important because they determine the final enzyme contents of a malt, the development of desirable flavors, aromas, and colors, as well as the concentration of undesirable flavors and aromas (Table 2).

Barley is the seed most often malted, be it for brewing, distilling, or the food industry. There are several reasons for barley's popularity in addition to simple tradition. The presence of a hull on barley protects the embryo during processing, strengthens the texture of steeped kernels, and serves as a filtration medium during brewing. Barley produces high levels of enzymes during germination, and the endosperm is easily degraded. Finally, the relatively low gelatinization temperature (*c.* 59–63 °C) of barley malt starch allows the heat-labile, starch-degrading enzymes from malt to remain somewhat active and hydrolyze the gelatinized starch. The result is a greater release of fermentable sugars compared with malts of grains with higher gelatinization temperatures. However, several other grains, including wheat, sorghum, rye, triticale, and oats, continue to be malted commercially for both traditional and practical reasons.

Table 1 Major uses of malted grains and pulses and contributions of the malts to the end-uses

Use	*Major contributions of malt*
Brewing beer	1. Source of fermentable sugars 2. Source of enzymes 3. Source of flavor and color
Malt distilling	1. Source of fermentable sugars 2. Enzymes 3. Peat flavor
Grain distilling	1. Source of enzymes 2. Source of fermentable sugars
Baked products	1. Source of enzymes 2. Source of flavor and color 3. Source of fermentable sugars
Breakfast cereals	1. Source of flavor, including sweetness, and color
Vinegar	1. Source of fermentable sugars 2. Source of enzymes
Germinated seeds and pulses for food	1. Source of readily available nutrients (amino acids and vitamins)

Malt Uses

Brewing

The brewing industry is the largest user of malt, and malt is arguably the most important ingredient in

Table 2 Malting steps and conditions, and their effects, in producing different types of malts

Malting steps	*Conditions*	*Options*	*Effects*
Steeping	• Barley placed under water • Periodic air rests to allow the grain to breathe	• Moisture content at end of steeping can vary from 42 to 48% (c. 24–48 h)	• The higher the moisture, the greater the potential to modify • Barley begins to germinate
Germination	• Barley grows at controlled temperatures with a good flow of moist air	• Germination can last from 4 to 7 days • Temperatures can vary from 12 to 20 °C • Temperatures can be raised and air flows reduced towards the end of germination	• Longer germination time gives better modification • Higher temperatures give better modification but greater malt losses • Raising temperatures and reducing air flows at the end of germination can increase levels of free amino acids and simple sugars
Withering (Kilning: Phase I)	• Green malt is dried to a moisture level of 10%	• Temperatures can vary from 45 to 65 °C • Rate of air flow can vary • Amount of recirculated air can vary	• Lower temperatures increase enzyme activity • Fast air flows with no recirculation dry the malt quickly, restricting any additional modification • Slow flow rates with recirculation allows the malt to stew producing free amino acids and simple sugars • Sealed containers with warm temperatures lead to starch conversion
Curing (Kilning: Phase II)	• Moisture content is further reduced to less than 6%	• Temperatures can vary from 65 to 250 °C	• It is at this stage that simple sugars and amino acids or small proteins combine in the Maillard reaction • Products provide color, flavor, and reducing power • Hotter temperatures result in greater production of Maillard products

beer. Malt is the major supplier of the fermentable sugars that are converted to alcohol by the yeast. The malt provides not only a sugar source, starch, but also the enzymes (i.e., α-amylase, β-amylase and limit dextrinase) to degrade the starch into fermentable sugars (i.e., glucose, maltose, and maltotriose). Yeasts do not produce these enzymes. The malt enzymes are even more important in adjunct brewing, where they hydrolyze not only the malt starch but also the starch of adjunct sources.

Malt supplies other essential nutrients for yeast growth. Yeasts require a source of nitrogen, but they lack the enzymes necessary for degrading proteins into products that they can assimilate. Therefore, barley protein needs to be degraded, during malting and brewing, into amino acids and peptides that the yeast can use. Vigorous yeast growth is also dependent on the supply of vitamins from malt including biotin, *myo*-inositol, pantothenic acid, vitamin E, folic acid, nicotinic acid, and thiamin. (*See* **Beers**: Biochemistry of Fermentation.)

Malt provides several components that affect beer quality. The mouth feel or body of beer is an important characteristic that is dependent, to a large extent, on the presence of dextrins in the beer. Dextrins are small fragments of starch that have not been degraded into fermentable sugars. Many beer types require a higher concentration of dextrins in order to develop a characteristic fullness. In contrast, many light, refreshing beer types require less body, and therefore, adjuncts are used that can result in a beer with fewer dextrins. Adjunct brewing can also be used to lower the cost of beer production. Low-calorie beers also require limited amounts of unfermentable dextrins in order to reduce the calorie content. Beers featuring less body require malts with higher levels of starch-degrading enzymes.

Several other properties of beer are also dependent on malt constituents. Malt supplies specific proteins, such as Z-protein, that support a good, stable foam or head in the final product. Reducing agents, such as Maillard products, scavenge oxygen in the beer, which increases the flavor stability. Flavor and color compounds, mainly Maillard products produced during kilning, are essential for developing the unique aroma, taste, and color of beer. (*See* **Browning**: Non-enzymatic.)

Some malt constituents have undesirable effects on beer or its processing. Certain proteins and polyphenols can cause hazes in the final product. Increased levels of nonstarch polysaccharides, for example β-glucan, can result in problems with beer filtration and hazes in the final product. Other components of malts, such as dimethyl sulfide, are sometimes

considered to be off-flavors. Many of these problems can be avoided through the use of appropriate malting conditions.

Distilling

The distilling industry uses malt to provide fermentable sugars and other nutrients for the yeast. In malt distilling, where only malt is added to the mash, all the fermentable sugars come from the malt. This is in contrast to grain distillers, where adjuncts, such as unmalted corn and wheat, are added to the mash as a cheaper source of starch. The malt provides the enzymes that convert the starch, whether from the malt or the cooked cereal, into fermentable sugars. In both types of distilling, malt provides other essential nutrients for the yeast such as free amino acids and vitamins. These must be in higher concentrations for grain distilling because the adjuncts provide only starch, which dilutes the concentration of other nutrients from the malt. In contrast to brewing, distilling malt provides no mouth feel or flavor to the final product, and in fact, a malt with a low concentration of unfermentable dextrins and limited flavor is desired. The only exception is the peat flavor that some malt distillers require in their malt.

Some malt continues to be used in the production of neutral spirits, such as gins, aqua vitae, or vodka, where it provides enzymes for degrading starch in a manner similar to grain distilling. Malt, usually in the form of extract, is the source of fermentable sugars in wine coolers, a recently developed alcoholic beverage in North America.

Food

Baked goods Fermented bakery foods are one of the largest users of food malt. Malt leads to better pan flow, increased loaf volume, as well as better crumb structure and crust color in baked goods. The malt supplies simple sugars, amino acids, vitamins, and enzymes, all of which help promote yeast growth during baking. Darker colored malts can also be added for increased flavor or color.

Malt is often included in cracker formulations to improve fermentation and to aid in sheeting and laminating. Malt also enhances the crust color of crackers allowing for reduced baking times and oven temperatures.

Breakfast cereals Barley malt is the most prevalent flavor material used in breakfast cereals. Most flavors come directly from the malt, but some are formed, during the processing of breakfast cereals, from the simple sugars, peptides and amino acids supplied by the malt. The simple sugars also provide sweetness to the cereals. In contrast to most other uses of malt, enzymes are undesirable in breakfast cereals. Their presence causes flavor instability due to their breakdown of starches during processing. As a result, malts used in breakfast cereals have very low enzyme levels with elevated levels of flavors and flavor precursors.

Vinegar Malt is used as a source of nutrients, including fermentable sugars, for yeast in the production of some vinegars. The object is to produce the maximum amount of alcohol, which requires a good complement of starch degrading enzymes, especially where adjuncts are used. In some cases, the mash is even supplemented with microbial enzymes. Bacteria are eventually used to convert the alcohol into acetic acid.

Other

Malt is also used in a range of other products. It is added to the diet formulation of both infants and the elderly because of the increased availability of nutrients in malt, such as amino acids, simple sugars, and vitamins. Malt enzymes can also increase the availability of nutrients from other ingredients in the diet. Malt is used as a carrier in pharmaceutical products where it increases palatability. Roasted malts are used as the basis for various hot drinks including the production of instant drinks from spray-dried malt extracts. The price of malt restricts its use in feed formulations, but limited amounts are used to increase palatability of pig starter diets and pet formulations.

Malt Types and Products

Pale Lager Malt

Pale lager malts are the most common type of malt. They are the major ingredient in the light lager beers that are so popular around the world. These beers require a malt with maximum extract and good enzyme potential but with limited development of color and flavor (Table 3). Pale lager malts are generally made from two-rowed barley with average levels of protein. In markets with high adjunct use, such as the USA, high protein, six-rowed barley is malted in order to ensure good enzyme potential and adequate levels of amino acids for yeast nutrition. Cheaper pale lager malts can be made from six-rowed, European, winter barleys for price sensitive markets.

Pale lager malts are well-modified and kilned gently. The gentle kilning conditions, with an upper temperature of 85 °C, produce a malt with high levels of enzymes and a low potential for color development.

Table 3 Quality characteristics and primary functions of malt types used in brewing

Malt type	*Type of beer*	*Primary use*	*Fine extract (%)*	*Protein (%)*	*Color (°Lovibond)*	*Diastatic power (°Lintner)*
Pale lager malt	Lager	Fermentable sugars and enzymes	81	10.5	1.5	85
Pale ale malt	Ale	Fermentable sugars and flavors	82	10.0	3.0	45
Vienna malt	Lagers, bocks, various	Fermentable sugars and color	81	11.0	3.5	70
Munich malt	Dark lager, various	Color, flavor and fermentable sugars	79	11.8	8	45
Crystal malt	Ales, light, alcohol-reduced, various	Fuller body, aroma, color, foam retention, flavor stability	75	12.0	50	na[a]
Carapils	Lagers, light, alcohol-reduced	Fuller body, foam retention, flavor stability	78	12.0	12	na
Amber malt	Ales, stouts	Source of color and flavor	75	na	20	na
Chocolate malt	Ales, stouts	Source of color and flavor	75	na	450	na
Black malt	Stouts	Source of color and flavor	73	na	550	na
Acidic malt	Lagers, wheat	Reduce pH in wort	73	10.5	2.0	na
Wheat malt	Wheat, various	Fermentable sugars, flavor, foam retention	85	13.3	2.0	110
Diastatic malt	High adjunct lagers, grain distilling	Enzymes	76	10.9	na	160

[a]na, not available.
Data adapted from Briggs (1998), Narziß (1999), and Anonymous (2001).

Pale Ale Malt

Pale ale malts are used to produce traditional, top fermented ales. In the past, this market had been restricted to the UK, but with the advent of micro-breweries, demand for pale ale malts has expanded to other parts of the world. The major difference between pale ale and pale lager malts is the darker colors and stronger flavors of the ale malts. This is achieved by using a more modified malt and higher final kilning temperatures (90 °C). The higher temperatures result in a stronger malty flavor and a deeper color, but they also reduce enzyme levels, resulting in more unfermentable dextrins and thus greater body in the final beer. Mild ale malts are less modified and are kilned at higher temperatures than pale ale malts. The result is lower levels of extract and enzymes but a greater production of color and flavor compounds. Beers made from mild ale malts are sweeter with more color and flavor.

Vienna Malt

Vienna malt is a lager type malt characterized by a red/copper color and a faint toffee flavor. This malt is the basic ingredient of the more highly colored lager beers, such as Märzen. It is also used, in minor amounts, to correct the color of beers made from excessively pale malts. The malt is produced by completing the kilning of a typical pale larger malt at a raised temperature (90 °C). Also, the use of recirculated air is increased at lower kiln temperatures in order to help produce the amino acids and free sugars required for color development at the higher final kilning temperature.

Munich Malt

Munich malt is characterized by a rich aroma and dark color. By tradition, it is used to make the dark, aromatic, full-bodied lager beers of Bavaria. The malt brings a more intense flavor and color to the beer than does a Vienna malt. Munich malt is produced from two-rowed barley with above average protein. The malt should be slightly overmodified during germination and then allowed to stew during the first stage of kilning. Stewing is intensified germination that results from a greater use of moist, recirculated air in conjunction with somewhat elevated temperatures. This releases the amino acids and free sugars that develop, during higher final kilning temperatures (100–105 °C), into Maillard products that are known for their intense color and flavor. Maillard products, such as melanoidins, also help to improve the flavor stability of beers. Brewers strive for flavor stability by limiting oxygen levels in their beer. The increased reducing power of Maillard products helps scavenge oxygen in the beer. Reasonable enzyme activity is preserved in Munich malt, despite extreme kiln temperatures, by drying the malt slowly.

Crystal and Carapils Malts

Crystal (caramel malt) and carapils malts are unlike other colored malts in that fully modified green malts are roasted in a sealed roaster (65–70 °C) prior to drying. As a result, the starches of crystal malts are degraded to simple sugars during this early stewing stage of kilning, in contrast to standard malts where the majority of starch conversion occurs later in the brewhouse. After this liquefaction of the endosperm,

crystal malts are dried at higher temperature (150–180 °C) in order to develop colors and flavors. The times and temperatures of the drying are adjusted to meet the color specifications of the customer. Crystal malts give a sweet, caramel flavor to beer. In addition, they produce a golden-colored beer with increased fullness and good foam retention. The significant increase in Maillard products leads to a greater reducing power and the ability to improve flavor stability by as much as 50%.

Carapils is a crystal malt but with less color due to the use of lower temperatures during final drying. This malt, which is sweet but without the caramel notes, is often added in small portions to lager-type beers to improve body, foam retention and flavor stability without changing the color or flavor of the final product. Carapils is used at higher concentrations in light and alcohol-free beers in order to increase the body of these beers without increasing fermentability. Crystal and carapils malts contain no active enzyme activity.

Brown and Amber Malts

Brown and amber malts are specialty dark-colored malts that have little or no enzyme activity. They are used in small proportions to increase beer color and add a biscuit-like flavor to stouts and top fermented ales. Brown malts are made by roasting low-moisture green malts. However, in order to prevent starch conversion, as in crystal malts, brown malts are partially dried at low temperatures with good air flows before roasting at temperatures of 130 °C. In contrast, amber malts begin with a pale ale-type malt that has been previously kilned. The kilned malt is roasted at temperatures of 100–150 °C, depending on the color desired. Amber malt is the more popular of the two types, but demand for both types has decreased in recent years.

Chocolate and Black Malts

Chocolate and black roasted malts are highly colored malts that give strong colors and particular flavors to beer. They are customarily used in stouts where they bring a flavor described as dry, burnt, astringent but with the hint of a rich, sweet taste. The malts start with an undermodified malt that has been kilned at low temperatures to a moisture of 4–6%. The dried malt is transferred to a roaster where temperatures are raised to 215 or 220 °C for chocolate and black malts, respectively. The major differences between the two malts is the length of roasting with the darker-colored black malts requiring longer times. In both cases, the high temperatures destroy all enzymes.

Other Brewing Malts

Several other types of barley malt with specific characteristics are available. Acidic malts are used to improve beer quality by lowering the pH of wort, especially in breweries using water with a high carbonate hardness. During mashing, the lower pH increases enzyme activity, especially proteases, a fact that sometimes results in the name 'proteolytic malt.' Acidic malt produces a beer with a paler color and a more mellow taste. The malt is produced by adding lactic acid or lactic acid bacteria to the barley during malting. Acidic malt can be used to avoid the need for additives such as sulfuric acid.

Chit or short malts are used as a cheaper source of fermentable sugars in areas where adjuncts are illegal. These malts are produced by severely limiting germination and kilning in order to produce an inexpensive product. The undermodification of chit malt, though, can lead to separation and filtration problems in the brewery, and therefore, it is used in limited proportions.

Brumalts are dark-colored and rich in Maillard products due to a high-temperature, anaerobic rest during the final day of germination. They are used to bolster the malt aroma of dark German beers such as Alt beer. Smoked malts are sometimes required in specialty beers such as Rauch beer. The malt is produced by allowing smoke to pass through the malt during the kilning process.

Wheat Malt

Wheat is the second most commonly malted cereal grain. The malt is a major ingredient in traditional wheat beers such as those of Belgium and Germany. Small portions of wheat malt are also added to standard ales and lagers in order to improve foam retention and yeast nutrition, which results from the different types and levels of protein in wheat versus barley malt. As a result of the missing hull, wheat malt produces higher levels of extract and lower levels of tannins, the latter can cause hazes in beer. A small but significant amount of wheat malt also goes into foods, where it produces a sweet luscious flavor. (*See* **Wheat**: Grain Structure of Wheat and Wheat-based Products.)

Wheat can be difficult to malt because the missing hull and smaller kernel size cause kernels to take water up more quickly. Overmodification must be avoided as it leads to excess solubilization of protein and the potential for hazes in the final beer. The smaller size of wheat kernels can also lead to processing problems in the malthouse, such as dense packing. Dark wheat malts are sometimes produced for German wheat beers, and for the food industry, by

malting wheat in a manner similar to that used for Munich malts.

Sorghum Malt

In Africa, sorghum is routinely malted on a commercial basis. The malt is used predominantly to produce traditional opaque beers such as kaffir as well as breakfast cereals. However, with a growing demand for European-type beers and with low barley production and high tariffs on barley malt, sorghum malts are also being used in Africa to make European-type beers.

It is difficult to produce a sorghum malt with appropriate quality for brewing. Germinating sorghum does not produce an abundance of enzymes, especially cell-wall-degrading enzymes. As a result, there is poor modification of the endosperm. Exaggerated malting conditions, such as long germination times and elevated temperatures, tend to increase the breakdown of starch and protein rather than cell walls. These conditions result in increased malt losses, sometimes as high as 20%, and excess levels of soluble protein in beer. Further problems are encountered in the brewery where high starch gelatinization temperatures combined with the heat lability of enzymes lead to incomplete breakdown of starch and thus poor fermentability. The residual starch, combined with high levels of β-glucan, can lead to filtration problems. The problems of malting sorghum are being addressed through the development of new varieties with better enzyme potential and through changes to the brewing process such as the use of microbial enzymes and of mash filters. (*See* **Sorghum**.)

Malt of Other Cereals

Rye, triticale, and oats are the only other cereals that are malted to any extent commercially. Malted rye is used traditionally in the rye whiskey industry, and some malted rye is used to add a 'spicy' flavor when brewing beer. Rye malt can produce extremely high levels of extract and enzymes, but its high levels of nonstarch polysaccharides cause increased viscosities, which result in problems with wort separation and beer filtration. These problems, along with a tendency to form hazes in the finished beer, have limited the use of rye in brewing. Malted triticale has seen limited use in the production of specialty beers, where it produces intense flavors. Beer made from triticale malt also has problems with high viscosity and excess levels of soluble protein, which cause the beer hazes.

The use of malted oats for brewing continues to decrease. The thick husk of oats causes reduced levels of extract, and a high level of lipid can lead to off-flavors in the final beer. Low levels of malted oats are still used in traditional oat stouts.

Malt for Distilling and Vinegar Production

Malt requirements differ between malt and grain distillers who use malt or malt plus adjunct in their mashes, respectively. Malt distillers require a malt that produces a maximum amount of fermentable sugars from its starch. There is no need for unfermentable dextrins. The malt, therefore, must have good levels of starch and starch-degrading enzymes. The malting conditions needed to produce this type of malt are similar to those used for pale lager malts. A plump barley is selected and well modified during germination. Milder kilning conditions are used to conserve enzyme activity. There is no need for color or flavor development in distillers malt, which also strengthens the use of low kilning temperatures. The one unusual practice with this malt can be the burning of peat during kilning. Varying levels of peat smoke can be passed through the bed during kilning in order to impart a lightly or heavily peated flavor, as demanded by distillers.

Grain distillers are primarily concerned with the levels of starch-degrading enzymes in their malt. The malt enzymes must degrade both the malt starch and the adjunct starch into fermentable sugars with, once again, no need for unfermentable dextrins. As with adjunct brewing, grain distiller's malt must have high levels of amino acids in order to support yeast growth. Maltsters meet these specifications by starting with a thinner, higher-protein barley. The barley is well modified and then killed at low temperatures to conserve enzyme activity. Grain distiller's malt is sometimes referred to as 'diastatic malt' because of extremely high levels of starch degrading enzymes (i.e., diastatic power). Diastatic malts are also used in brewing and in food applications when high levels of enzyme activity are required.

Vinegar processors require malt of a quality similar to that demanded by distillers. Processors that use all-malt mashes require a malt with high levels of both enzymes and starches, a profile similar to that demanded by malt distillers. In contrast, vinegar processors that use adjuncts require a malt with high enzyme levels and an adequate supply of amino acids, a quality profile similar to that of grain distillers. Vinegar processors are not interested in the presence of dextrins or the development of flavor and color in their malts.

Malt Extracts and Syrups

Malts are often processed into extracts or syrups in order to facilitate further processing. Extracts are

made by following the first few steps of brewing. A malt is mashed in order to complete the solubilization and conversion of starch into fermentable sugars as well as solubilization of protein. The resulting extract is separated from the spent grains and concentrated through evaporation under vacuum. Extracts destined for brewpubs, brew-on-premise, or home brewing are often supplemented with hops prior to concentration.

Syrups are produced in a manner similar to extracts except that adjuncts replace some of the malt primarily to reduce costs. The quality of syrups differs from that of extracts in that reducing sugars are increased, whereas soluble protein levels are lower, resulting in a lighter color and a distinctly mellower and sweeter flavor. Final levels of enzymes, flavors, and colors, in both extracts and syrups, can be altered by starting with different malts, by changing the mashing conditions or by changing the conditions used for concentrating the extracts and syrups.

Dry extracts and syrups can be formed by spray drying the liquid extracts and syrups. The color, flavor, and sweetness are preserved, but the heat inactivates the enzymes. Dry diastatic malts with a high enzyme activity, therefore, are produced by mixing milled malt with wheat flour. The final product contains enzymes from the malt and potential extract from the wheat but has a limited ability to produce color or flavor.

Malts for Food Products

Food industries require an array of malt types with different levels of flavor, color, and enzymes. They depend on brewing type malts that have been custom-made to meet their specifications. Extracts, syrups, and their dried products are often used, provided they have appropriate specifications, by the food industry as they can simplify processing.

See also: **Barley**; **Beers**: History and Types; Raw Materials; Chemistry of Brewing; Microbreweries; **Bread**: Dough Mixing and Testing Operations; **Browning**: Nonenzymatic; **Cereals**: Breakfast Cereals; **Enzymes**: Uses in Food Processing; **Malt**: Chemistry of Malting; **Sorghum**; **Syrups**; **Vinegar**; **Wheat**: Grain Structure of Wheat and Wheat-based Products; **Whisky, Whiskey, and Bourbon**: Products and Manufacture

Further Reading

Anonymous (2001) *Pauls Malt Limited*. http://www.paulsmalt.co.uk.

Bemment DW (1985) *Specialty malts*. *The Brewer* 71: 457–460.

Briggs DE (1998) *Malts and Malting*. London: Blackie Academic & Professional.

Briggs DE, Hough JS, Stevens R and Young TW (1981) *Malting and Brewing Science: Volume 1 Malt and Sweet Wort*, 2nd edn. London: Chapman & Hall.

Finney PL (1982) Effects of germination on cereal and legume nutrient changes and food or feed value: A comprehensive review. In: Nozzolillo C, Lea PJFA and Loewus PA (eds) *Recent Advances in Phytochemistry. Volume 17: Mobilization of Reserves in Germination*. pp. 229–305. New York: Plenum Press.

Henry RJ and Kettlewell PS (eds) (1996) *Cereal Grain Quality*. London: Chapman & Hall.

Hickenbottom JW (1996) Processing, types, and uses of barley malt extracts and syrups. *Cereal Foods World* 41: 788–790.

MacGregor AW and Bhatty RS (eds) (1993) *Barley Chemistry and Technology*. St. Paul, MN: American Association of Cereal Chemists.

Narziß L (1991) *Special Malt for Greater Beer Type Variety*, pp. 284–292. *Brauwelt International* 9: 284–292.

Narziß L (1999) *Die Technologie der Maltzbereitung*, 7th edn. Stuttgart: Ferdinand Enke.

Palmer GH (ed.) (1989) *Cereal Science and Technology*. Aberdeen: Aberdeen University Press.

Piggott JR, Sharp R and Duncan REB (1989) *The Science and Technology of Whiskies*. Harlow, UK: Longman Scientific and Technical.

Simpson JP (1995) Special malts. *The Brewer* 81: 455–458.

Chemistry of Malting

M S Izydorczyk and M J Edney, Canadian Grain Commission, Winnipeg, Manitoba, Canada

History

Beer, one of the most popular alcoholic beverages, is made from four basic ingredients: malted grain, hops, yeast (*Saccharomyces* spp.), and water. There is evidence that as long as 4000 years ago, the Egyptians and Mesopotamians enjoyed malted barley drinks that were the predecessors of our modern beer. Malt is a natural product of a process, called malting, whereby the grain is allowed to germinate under controlled conditions of moisture, temperature, and air. Early malsters copied the natural process of germination by steeping the grain in containers and then spreading it out in a thin layer across a floor. Once germination was completed, the 'green malt' was dried over open fire. The early maltsters were true craftsman passing down the skills of malting by word of mouth from generation to generation. Nowadays, malting is a large-scale commercial process yielding

value-added, modified grain product used primarily for the production of beer and distilled products such as whiskey, and, to a lesser extent, as an ingredient in certain food stuffs. Although many cereals have been processed into malt, barley (*Hordeum vulgare*) is considered the best malting grain, so the chemical processes described below are those occurring in barley grain during malting.

Technologically, malting is divided into three stages: steeping, germination, and kilning. However, the complex biochemical and chemical reactions occurring during malting are interrelated and often not confined to one stage of malting. Those reactions involve both major and minor constituents of barley grain, which are interconverted under the influence of various enzymes synthesized and/or activated during malting. The main purpose of malting, from a technological point of view, is to render the barley kernel more soluble in hot water and susceptible to fermentation by yeast. Extensive changes that occur in the grain during malting include complete destruction of the endosperm cell walls, breakdown of the protein matrix, controlled solubilization of barley proteins, and formation of hydrolytic enzymes, which carry out degradation of starch granules at subsequent stages of the brewing process. (*See* **Beers**: History and Types.)

Composition of Barley Grain

The husk and pericarp, the two most outer and protective tissues of the barley grain, consist primarily of hollocellulose, hemicellulose, lignin, and lignans (Figure 1). The testa and the nucellar cuticle, the pigment strand, and the micropylar structures make up the semipermeable system that divides the interior from the exterior of the grain and limits the ingress of water and dissolved substances. The husk and pericarp are not extensively altered during malting, but they do play a role in the process by limiting the rate of infusion of oxygen and water in respiring grain. Moreover, the husk protects the acrospire and the entire grain against physical damage. The husk also has major effects on the character, flavor, and stability of beer. Removal of husk from the grain leads to a blander and more insipid beer, whereas grain enriched in husk produces astringent beers with a tendency to form a haze.

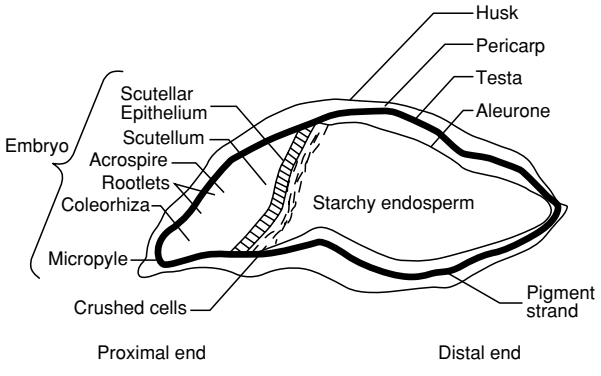

Figure 1 Longitudinal section of a mature barley grain showing component tissues.

The grain embryo contains little starch but is rich in protein (34%), lipids (14–17%), and ash (5–10%) and extremely rich in sugars (sucrose 5–10%; raffinose, 5–10%; fructans). During germination, the embryo synthesizes and releases gibberellin hormones and a complex mixture of hydrolytic enzymes, which are involved in mobilizing the reserves in the grain and supporting the embryo's metabolism and growth.

The rest of the mature grain consists of nucellar material, aleurone and endosperm. The thick walls of the aleurone layer (3–5 μm) contain mainly arabinoxylans (67–71%) and smaller amounts of β-glucan (26%). The tissue contains three layers of cells and no starch but abounds in proteins (17–20%), triacyl glycerol (20%), minerals, phytin, and sugars (including sucrose, raffinose, stachyose, verbascose, and fructans). Aleurone cells have bodies called aleurone grains. These consist of globoids, which contain phytin, and crystalloids, which contain protein and polysaccharides. The cells of the aleurone layer, like those of the embryo, are living. In hydrated grain, they are capable of synthesizing and secreting a variety of enzymes needed for breakdown of the reserves stored in the starchy endosperm.

The starchy endosperm makes up about 75% of the barley grain and is the major reserve tissue. Starch, confined to small and large granules, constitutes about 80% of the endosperm. In addition, the starchy endosperm contains a relatively large amount of proteins ($\sim$9% of the tissue) and smaller amounts of lipids, minerals, and nucleic acids. The cell walls (2 μm) are mainly built up from mixed linkage β-(1 $\rightarrow$ 3, 1 $\rightarrow$ 4)-glucans (70%), arabinoxylans (20%), and small amounts of proteins, β-(1 $\rightarrow$ 3) glucans and other polysaccharides containing galactose, mannose, and uronic acids. In mature grain, the cells of the starchy endosperm are nonliving and function solely as reserves of carbohydrates, proteins, lipids, and nucleic acids awaiting dissolution during germination. (*See* **Barley**.)

Regulation of the Germination Process

Barley germination commences at relatively low levels of grain moisture, but rapid and uniform germination for malting purposes requires a grain

moisture content of 40–45%. Most of the water that penetrates the barley kernel does so near the embryo, probably through the micropylar region. The biochemical processes of germination also require oxygen. The entry of oxygen into the interior of the grain is hindered by the diffusion barriers of the husk and pericarp, and by the water film on the surface of the grain. The proper hydration and oxygen uptake of barley grain for malting purposes are accomplished by multiple submersions in water alternated with air rest during steeping, the first stage of the malting process.

At the onset of germination, the embryo uses its internal reserves of sugar, lipids, proteins, and minerals to initiate development of roots and a shoot. A mixture of hydrolytic enzymes is released from the scutellar epithelium into the endosperm. The gibberellins, a family of plant hormones, are formed in the germinating embryo and migrate to the aleurone layer, where they trigger enzyme production and/or enzyme release.

The mechanisms by which gibberellins trigger the responses of aleurone layers are not fully understood, but there is some evidence for gibberellin receptors on the surface of protoplasts prepared from the aleurone layer. Exactly which gibberellins are present in malting barley also remains unclear. The main gibberellin appears to be GA_1, with smaller amounts of gibberellic acid GA_3 also present (Figure 2). The production and release of gibberellins seem to be affected by the concentration of sugars in the embryo. The process probably involves a feedback control loop, i.e., sugar depletion in scutellum → gibberellin formation and release → gibberellin-triggered enzyme synthesis and release from aleurone layer → endosperm breakdown and sugar release → sugar diffusion into scutellum → increased sugar level in embryo → cessation of gibberellin formation and release. There is no clear evidence that, in germinating barley, other nongibberellin hormones influence the formation of hydrolytic enzymes. However, some hormones such as indolacetic acid, cytokinins, abscisic acid, and ethylene, when added to isolated aleurone, influence the response of the tissue. Abscisic acid, in particular, is a potent inhibitor of gibberellin-induced enzymes synthesis in aleurone cells.

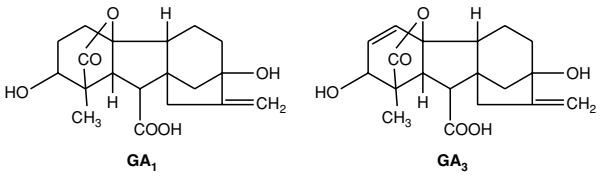

Figure 2 Chemical structure of gibberellins GA_1 and GA_3.

The scutellar epithelium also plays a role in producing hydrolytic enzymes. It has been suggested that the mixture of enzymes released from the scutellar epithelium differs from that produced by the aleurone layer. Embryos do not need an external supply of gibberellins to produce enzymes, but when supplied with amino acids, their productivity increases. It is agreed that some isoenzymes, e.g., carboxypeptidase I, (1 → 3, 1 → 4)-β-glucanase I, and some α-amylases, are generated in the scutellum. Biochemical studies now suggest that the scutellum accounts for approximately 1–10% of α-amylase production; the remaining α-amylase is produced by the aleurone layer.

The enzymes released by 'triggered aleurone' layers include α-amylase, (1 → 3, 1 → 4)-β-glucanase, (1 → 3)-β-glucanase, proteases, peroxidase, and enzymes able to cleave DNA, RNA, various disaccharides, glycosides, phosphates, and peptides.

Although aleurone responses are primarily triggered by gibberellins, it is also possible that other substances act as modulators in this process. It has been shown that tannins can act as gibberellin antagonists. Also, high concentration of sugars and other compounds, which create a high osmotic pressure, reduce enzymes formation by the aleurone layer. However, calcium ions promote the secretion of α-amylase as well as the formation of β-glucanase.

Chemical Reactions During Malting

The migration of hydrolytic enzymes, actively secreted from the aleurone and the scutellar epithelium into the starchy endosperm, where they depolymerize cell walls, storage proteins, starch, and residual nucleic acids, is brought about by a rather slow water-assisted diffusion process. Endosperm dissolution begins in the region immediately adjacent to the scutellum and progresses toward the distal end of the grain. Before they can reach their substrates in the starchy endosperm, the migrating enzymes must first overcome two physical barriers: the walls of the secretory cells themselves and the walls of starchy endosperm cells.

Degradation of Cell-wall Polymers

β-Glucans The β-glucans are linear polymers composed of D-glucopyranosyl residues (Glc*p*) linked via a mixture of β-(1 → 3) (~30%) and β-(1 → 4) (~70%) linkages. The linkage arrangement is not completely irregular; consecutive blocks of β-(1 → 4) linkages, mostly two or three, but occasionally up to 20, are separated at random by single β-(1 → 3) linkage (Figure 3). The β-glucans are high-molecular-weight polymers only partially soluble in water. As no evidence for covalent cross-linking has been found,

it seems likely that less soluble β-glucans are physically entrapped with insoluble proteins, arabinoxylans, or cellulose.

The breakdown of partially soluble β-glucans may have to be initiated by the release and solubilization of these polymers from the cell walls. Enzymes, such as carboxypeptidase, protease, phospholipase, cellulase, and/or esterase may be involved in these initiating processes. The actual depolymerization of β-glucan occurs by the action of two major hydrolyzing enzymes, endo-(1 → 3, 1 → 4)-β-glucanases (isoenzymes EI and EII; EC 3.2.1.73), which are developed in almost equal proportions in germinating barley (Table 1). The enzymes catalyze the hydrolysis of β-(1 → 4) linkages within the polymeric chains.

Table 1 Properties of barley endo-β-glucanases[a]

Property	*Isoenzyme EI*	*Isoenzyme EII*
Amino acids	306	306
Apparent molecular weight (SDS-PAGE)	30 000	32 000
Isoelectric point	8.5	10.6
pH optimum	4.7	4.7
Carbohydrate content (%)	Trace	~4
Thermal stability	Low	Greater
Secretion site	Scutellum, aleurone	Aleurone

[a]Data compiled from Fincher GB (1992) Cell wall metabolism in barley. In: Shewry PR (ed.) *Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology*, pp. 413–437. Wallingford, UK: CAB International; and Fincher GB and Stone BA (1993) Physiology and biochemistry of germination in barley. In: MacGregor AW and Bhatty RS (eds) *Barley Chemistry and Technology*, pp. 247–295. St. Paul, MN: American Association of Cereal Chemists.

These require adjacent β-(1 → 3)- and β-(1 → 4)-linked glucose residues next to the β-(1 → 4) linkage being hydrolysed (Figure 3).

The endo-β-glucanases produce mainly tri- and tetrasaccharides, which may be further degraded to glucose by exo-β-glucanase (EC 3.2.1.74). Three exo-β-glucanases of molecular weight 67 000–70 000 have been purified from grain and shown to hydrolyze both the β-glucans and the oligosaccharides released by the action of endo-(1 → 3, 1 → 4)-β-glucanases to glucose.

During malting, the total amount of β-glucans declines considerably; the content of β-glucans in barley may vary from 3 to 6%, and after malting, it should range from 0.2 to 1%. High levels of β-glucans in malt reflect incomplete cell-wall degradation, affect the yield of malt extract, and produce viscous extracts, causing filtration difficulties in the brewery and contributing to formation of hazes and precipitates in beer.

Arabinoxylans Arabinoxylans are side-chain branched heteroglycans built from pentose sugars, arabinose and xylose. They consist of a linear xylan backbone of β-(1 → 4)-linked D-xylopyranosyl residues (Xyl*p*) to which α-L-arabinofuranosyl units (Ara*f*) are attached mostly as single substituents (Figure 4). Some arabinose side-chains carry esterified residues of ferulic acid. As a consequence, arabinoxylan chains in cell walls may be cross-linked with each other via ferulic acid bridges and/or cross-linked with other cell-wall polymers via ether or ester linkages.

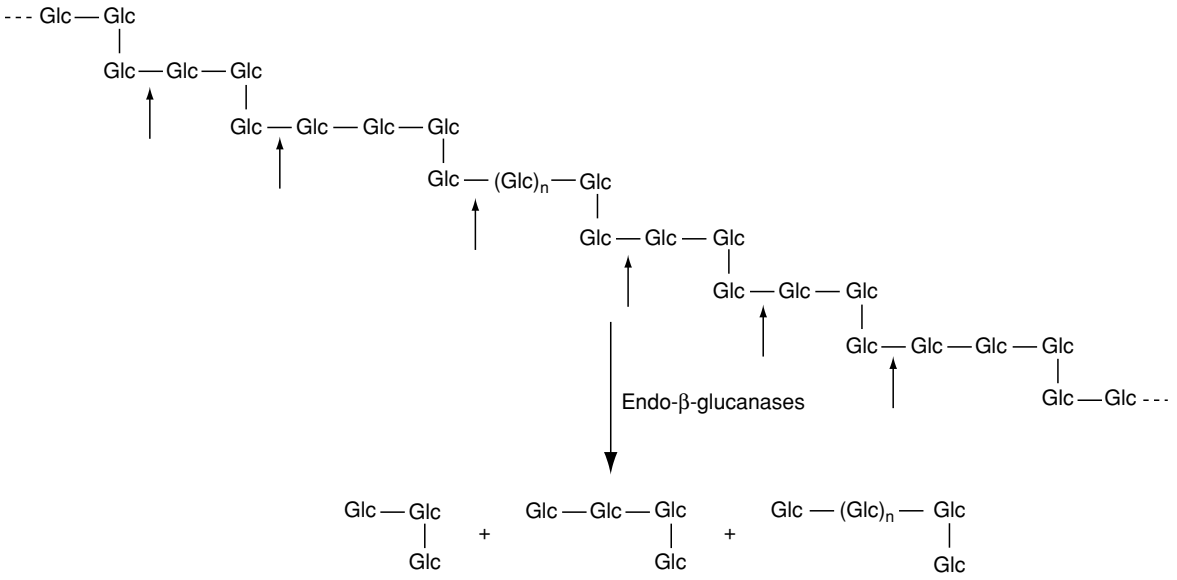

Figure 3 Generalized structure of barley β-glucans. Glc, D-glucopyranose residue; — and |, β-(1→4) and β-(1→3) linkages, respectively, between the glucose residues; n = 1–20; ↑, sites of hydrolysis by endo-β-glucanase.

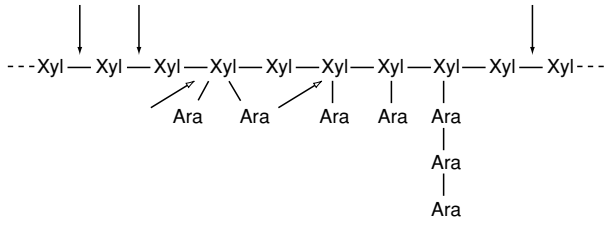

Figure 4 Generalized structure of barley arabinoxylans. Xyl, β-D-xylopyranose residues; Ara, α-L-arabinofuranose residues; ↓ and ↗, sites of hydrolysis by β-xylanase and α-arabinofuranosidase, respectively.

The routes for hydrolytic breakdown of barley arabinoxylans are less known. Their depolymerization to constituent monosaccharides is mediated by the combined action of endo-(1 → 4)-β-xylanases (EC 3.2.1.8), β-xylopyranosidase (EC 3.2.1.37), and α-L-arabinofuranosidase (EC 3.2.1.55). The amount of all these enzymes increases during malting. It is believed that, initially, α-L-arabinofuranosidase catalyzes the removal of arabinose residues from arabinoxylan chains or arabinoxylan oligosaccharides. When a sufficient number of arabinose substituents have been removed, the xylan chain is hydrolyzed by (1 → 4)-β-xylanases to smaller oligosaccharides. Three endoxylanase isozymes have been isolated from barley grain, but their hydrolytic actions have not been fully investigated. β-Xylopyranosidase catalyzes the hydrolysis of xylobiose to xylose and possibly removes the unsubstituted xylose residues from the end of xylan chains. As free pentose sugars do not accumulate in malting grain, they must be rapidly utilized by the living tissues. While β-glucans are degraded extensively during malting, arabinoxylans are not; the majority of barley arabinoxylans remain present in malt. (*See* **Carbohydrates**: Classification and Properties.)

Starch Degradation

The insoluble starch granules stored in the starchy endosperm cells contain two polymers: amylose (20–30%) and amylopectin (70–80%). Although built up from the same glucopyranose units, these polymers differ in the type of glycosidic linkages and require different enzymes for their degradation. Amylose consists predominantly of linear chains of about 2000–5000 of α-(1 → 4) linked D-glucopyranosyl residues. In granules, amylose may be associated strongly with lipids, and the resulting amylose–lipid complexes may be less susceptible to enzymic degradation. Amylopectin consists of very-high-molecular-weight molecules, estimated to be 10^6–10^8 Da. It differs from amylose in that its α-(1 → 4)-linked chains of glucose are highly branched through α-(1 → 6) linkages, and therefore, amylopectin is characterized by an extensively branched structure.

During malting, starch is only partially degraded (15–18%), and the composition of the residual starch may be slightly altered. The amylose content may increase slightly, and the molecular weight of both polymers is decreased. There is some evidence that small starch granules are more susceptible to degradation than the large starch granules. Also, close to the embryo, starch granules are almost completely degraded, whereas those further away show evidence of limited enzymic attack in the form of either roughened surfaces or 'pinholes' (Figure 5). Most important, all enzymes necessary for total degradation of starch, which occurs at subsequent stages of brewing, are synthesized and/or activated during malting. In the germinating grain, a considerable number of enzymes are involved in the degradation of starch.

α-Amylases The α-amylases are able to hydrolyze intact starch granules with the formation of soluble products. They are responsible for the initial degradation of starch granules during malting. α-Amylases (EC 3.2.1.1) are endoenzymes that cleave internal α-(1 → 4) glucosyl linkages of amylose and amylopectin

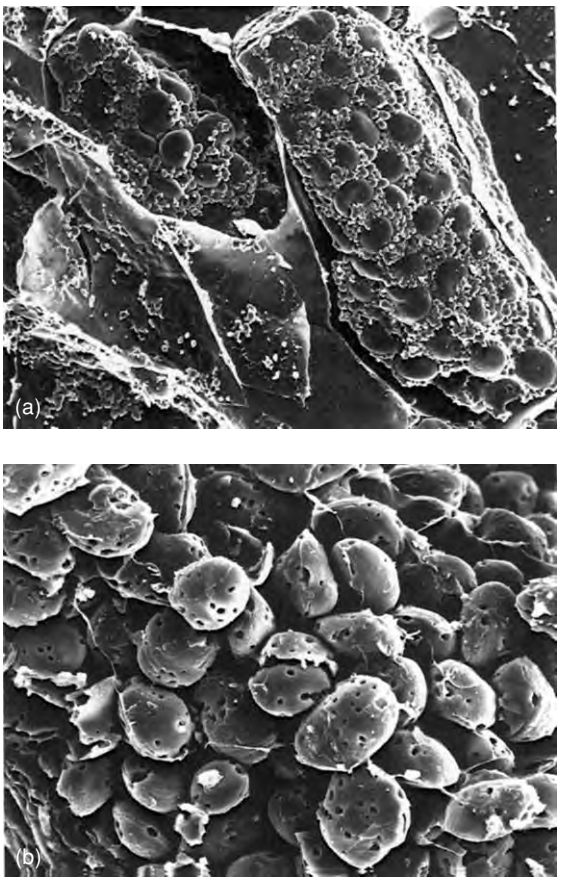

Figure 5 Scanning electron micrograph of the starchy endosperm showing the large and small starch granules (a) before and (b) after germination.

in a random fashion. They are unable to hydrolyze the α-(1 → 6) glucosyl linkages at the branch points in amylopectin. Their hydrolytic action is slower on short chains, and they vary in their ability to hydrolyze the α-(1 → 4) linkages in close proximity to the branch points. α-Amylases, acting on their own, are able to degrade amylose to a mixture of shorter linear α-glucan chains (linear α-dextrins), oligosaccharides, maltose, and glucose. Amylopectin, however, is degraded to a lesser extent, yielding a mixture of still large branched fragments (branched α-dextrins) as well as some linear oligosaccharides, maltose and glucose (Figure 6). Two isoenzymes of α-amylases with the same apparent substrate specificities and action patterns, but different properties, are synthesized in germinating barley (Table 2).

Table 2 Properties of α-amylase I and II from germinating barley[a]

Property	*α-Amylase I*	*α-Amylase II*
Molecular weight	45 342	45 005
Isoelectric point	4.8–5.3	5.9–6.6
pH optimum	3.0–5.5	5.0–5.4
Stability at pH 3.6	Relatively stable	Unstable
Reaction with endogenous inhibitor[b]	No inhibition	Strong inhibition
Ca^{2+} binding	Very strong	Strong
Major secretion site	Aleurone	Aleurone
Proportion in germinating grain	Minor (up to 10%)	Major (up to 90%)

[a]Data compiled from Briggs DE (1992) Barley germination: biochemical changes and hormonal control. In: Shewry PR (ed.) *Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology*, pp. 369–401. Wallingford, UK: CAB International; Hill RD and MacGregor AW (1988) Cereal α-amylases in grain research and technology. In: Pomeranz Y (ed.) *Advances in Cereal Science and Technology*, vol. IX, pp. 217–261. St. Paul, MN: American Association of Cereal Chemists; and MacGregor EA and MacGregor AW (1987) Studies of cereal α-amylase using cloned DNA. *CRC Critical Reviews in Biotechnology* 5: 129–142.
[b]Barley α-amylase/subtilisin inhibitor.

β-Amylases The β-Amylases (EC 3.2.1.2) are exoenzymes that cleave the penultimate α-(1 → 4) linkage from the nonreducing end of the polymeric chains and release the disaccharide maltose. β-Amylase, acting alone, can degrade amylose completely to maltose. This enzyme will not, however, attack the α-(1 → 6) linkages or α-(1 → 4) linkages close to the α-(1 → 6) links; therefore, it degrades only the outer chains of amylopectin and leaves a large portion (called β-limit dextrins), in which the outer chains have been degraded to stubs of two or three glucose residues adjacent to the α-(1 → 6) linkages (Figure 6). β-Amylase is not synthesized during germination. In barley, virtually all β-amylases occur in the starchy endosperm. The four major isoenzymes occur in free or protein-bound forms. During malting, the level of free β-amylase may initially fall but subsequently increases, as nearly all enzymes are liberated by the action of proteolytic enzymes.

Limit dextrinase Limit dextrinase (EC 3.2.1.10) is a debranching enzyme that hydrolyzes specifically α-(1 → 6) linkages in amylopectin or in branched dextrins derived by the actions of α- or β-amylases (Figure 6). The hydrolytic action of this enzyme results in the formation of linear α-(1 → 4)-linked chains that can be extensively depolymerized to glucose and maltose by the combined actions of α- and β-amylases. The limit dextrinase in germinating barley hydrolyzes the α-(1 → 6) linkages of branched dextrins faster than those in large amylopectin molecules. Mature barley contains low levels of the enzyme but during germination, the activity of the enzyme increases due to *de novo* synthesis in the aleurone. However, barley contains a heat-stable protein that inhibits the enzyme. This inhibitor decreases in amount during malting, but there is a sufficient amount left in malt to inhibit most of the limit

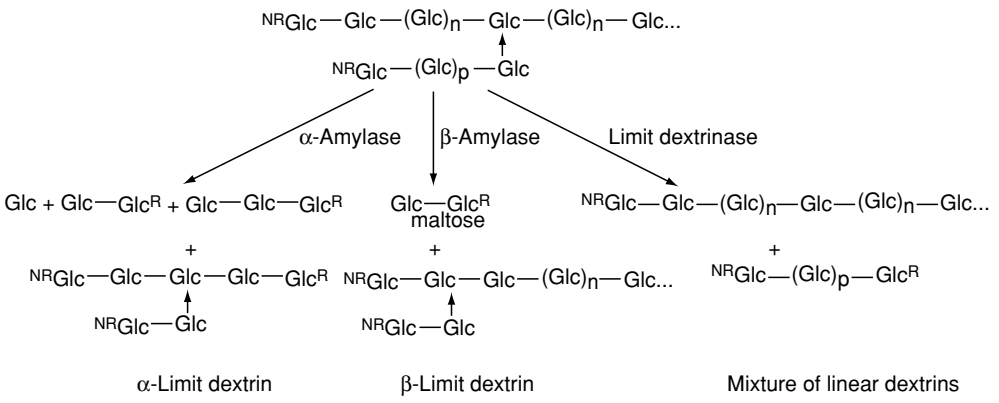

Figure 6 Summary of the modes of action of some enzymes important in degradation of starch polymers. Glc, D-glucopyranose residue; —, α-(1 → 4)-link; ↑, α-(1 → 6)-link; NR, nonreducing end; R, reducing end.

dextrinase. Although purified limit dextrinase has no action on starch granules, it enhances the rate of granule dissolution when combined with the actions of α- or β-amylases.

α-Glucosidase α-Glucosidase (EC 3.2.1.20) releases glucose residues from the nonreducing end of a variety of α-glucosides and, therefore, participates in the final conversion of maltose and other small dextrins to glucose. After germination is initiated, rapid synthesis of two or more isoenzymes of α-glucosidase is observed in both the aleurone layer and the embryo. (*See* **Starch**: Functional Properties.)

Chemical Changes in Proteins During Malting

Proteins account for about 8–15% of the dry weight of the mature barley grain. Albumins and globulins – the water- and salt-soluble proteins, respectively – function as enzymes, metabolic regulators, structural or storage proteins, or inhibitors of particular enzymes. Hordeins are insoluble in salt solution but dissolve in warm aqueous alcohol, whereas the residual insoluble material constitutes the fourth group of barley protein, glutelin. Hordeins belong to the group of cereal proteins known as prolamines because of their high proline and glutamine content. They are the major storage proteins of barley. The glutelin fraction contains storage and structural proteins. The quantities and relative amounts of the fractions change substantially during malting. Following hydration, the protein reserves in the scutellum and aleurone layer are degraded, and at least some of the liberated amino acids are used to synthesize hydrolytic enzymes, which are released into the starchy endosperm. The level of hydrolytic enzymes increases during malting and declines during kilning. The endogenous inhibitors of many enzymes decline during malting. Reserve proteins in the endosperm are substantially degraded to soluble peptides and amino acids by the action of many proteolytic enzymes. The simple peptides, amides, and amino acids diffuse to the embryo, where they may be completely degraded, interconverted, and utilized by the growing embryo to form new proteins in the growing rootlets and acrospire. In a well-modified malt, 40–45% of the total protein is water-soluble.

Degradation of proteins in germinating barley is accomplished by a range of proteolytic enzymes, including endopeptidases (proteases) and exopeptidases (Figure 7). Four groups of endopeptidases (EC 3.4.21), cysteine, serine, aspartic, and metalloproteases, are produced by germinating barley. The cysteine proteases are dominant, accounting for up to 90% of the total protease activity. The endopeptidases can attack proteins and large polypeptides, hydrolyzing peptide bonds from the interior of the chain and thus rapidly diminishing the molecular weight of proteins. During germination, protease activity in the starchy endosperm increases several-fold.

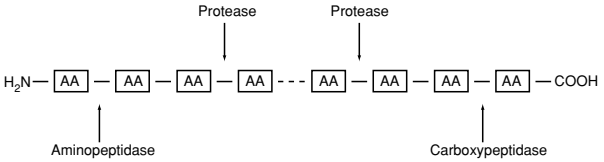

Figure 7 Summary of the modes of action of some enzymes important in degradation of proteins during malting.

Exopeptidases release one amino acid at a time from either the carboxyl or the amino end of the peptide chain. Five carboxypeptidases (EC 3.4.16.1) have been identified in germinating barley. Each enzyme attacks peptides exclusively from the COOH-termini of the chain and has a distinct preference for particular amino acids. However, their specificities are complementary, and together, they effect an extensive degradation of barley proteins. In addition to carboxypeptidase, several other exopeptidases, including four neutral aminopeptidases, leucine aminopeptidase, and dipeptidase, have been identified in germinating barley. The aminopeptidase and dipeptidase are active mainly in the scutellum and are involved in the absorption of amino acids by the growing embryo. The carboxypeptidases, however, are found both in the starchy endosperm and in the scutellum, and can release amino acids from hordeins as well as from smaller peptides. They might also be involved in the solubilization of β-glucans during the early stages of germination.

Chemical Changes in Other Grain Reserves

The total lipid content of barley may vary from 2 to 4.4%. Barley lipids comprise 67–78% of nonpolar lipids (mostly triacylglycerides (TG), some steryl esters and free fatty acids (FFF)), 8–13% of glycolipids, and 14–21% of phospholipids. Lipids associated with starch granules are almost exclusively lysophospholipids and free fatty acids. It is generally accepted that during germination, storage lipids are used as a source of metabolic energy and fatty acids for synthesis of new membrane and structural lipids. The overall decrease of total lipids by 10–50% may occur during malting, but the changes in the composition of lipid classes are minor and/or not clearly elucidated. The lipolytic activities in germinating barley increase. Enzymes such as lipases and phospholipases act in concert with lipoxygenase. When the barley tissue is stimulated by gibberellins, the aleurone bound lipase (EC 3.1.1.3) is activated and responsible for hydrolyzing the ester bonds in TG

and for releasing long fatty acids. Several different phospholipases (EC 3.1.1.32) are responsible for hydrolyzing four different ester bonds in phospholipids. The polyunsaturated fatty acids released from TG and phospholipids are preferentially oxidized by lipoxygenase. Lipoxygenase (EC 1.13.11.12) utilizes atmospheric oxygen to oxidize the *cis*-1, *cis*-4-pentadiene part of (n-6) and (n-3) polyunsaturated fatty acids (linoleate and linolenate) to hydroperoxides and hydroperoxyradicals. Lipoxygenase from germinating barley is more active toward methyl linoleate and trilinolein. It is thought that during steeping, the activity of lipoxygenase decreases and remains low during germination but rises sharply in the middle of kilning only to decrease again to a very low level at the end of kilning.

Inorganic materials also undergo significant changes in barley during malting. Best elucidated have been the changes pertaining to phosphates. The majority of the free phosphate raise occurring during malting originates from degradation of aleurone phytins, potassium and magnesium salts of phytic acid (*myo*-inositol haxaphosphate), by the action of phosphatase (EC 3.1.3.26), the enzyme able to hydrolyze phytic acid. In addition, phosphate is released from phospholipids, DNA, and RNA by appropriate hydrolytic enzymes.

Among the numerous and complex phenolic compounds that have been identified in barley, the most extensively studied have been phenolic acids, flavanols (such as (+)-catechin and gallocatechin), pigments, and condensed tannins derived from catechin. Phenol compounds are important in the brewing process because of their antioxidant properties, their astringency, and their tendency to form hazes in beer. Maltsters sometimes manipulate their content in malt by application of alkaline steeping, which removes a portion of phenolics from the grain. Various oxidase (EC 1.14.18.1) and peroxidase (EC 1.11.1.7) enzymes are able to act on phenolic substances, but little is known about their action during malting. Green malt probably contains oxidases that catalyze the oxidative polymerization of polyphenols to tannins.

Chemical Aspects of Kilning

When the desired enzyme levels and endosperm modification have been achieved, the germination is terminated through kilning, a process in which dry air of increasing temperature is passed through the green malt to produce a dry and friable grain product. Kilning eventually arrests all biochemical reactions, although appropriate manipulations of time, temperature, and air flow parameters allow survival of many hydrolytic enzymes needed in the later stages of the brewing process. Kilning also facilitates formation of a myriad of color, flavor, and aroma compounds that make malt such a unique product. Responsible for these phenomena are primarily the complex Maillard reactions. The malt reducing sugars and amino compounds (amino acids and compounds containing free amino groups), when exposed to higher temperatures, react with each other to form N-substituted glycosylamine, which undergoes a further reaction called the Amadori rearrangement to 1-amino-1-deoxy-2-ketose (Figure 8). Chemical reactions continue taking several possible routes. Enolization of 1-amino-1-deoxy-2-ketose leads to the formation of furfurals and hydroxymethylfurfurals, which polymerize quickly to melanoidins,

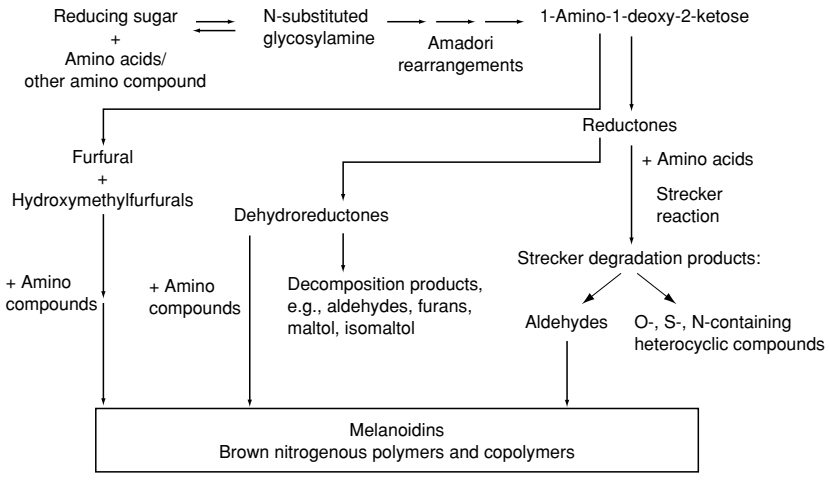

Figure 8 Simplified scheme of chemical reactions involved in the formation of flavor and color during the Maillard reactions. Information compiled from Briggs DE (1998) *Malts and Malting*. London: Blackie Academic & Professional; and Eskin NA (1990) *Biochemistry of Foods* London: Academic Press with permission.

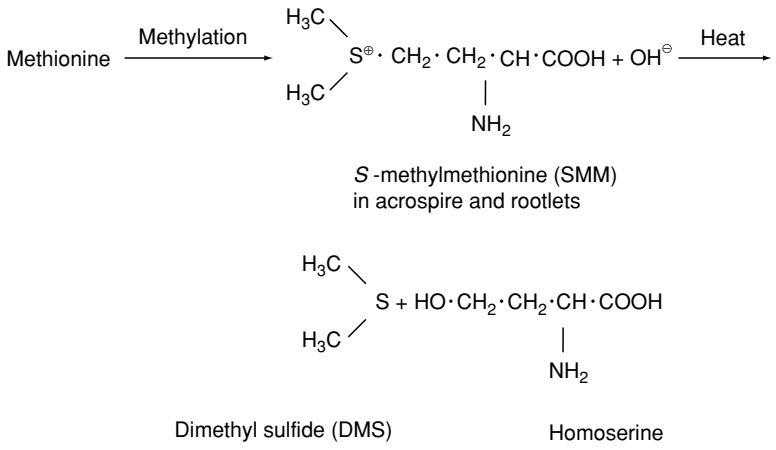

Figure 9 Formation of dimethyl sulphide during kilning.

dark-colored insoluble material containing nitrogen. Enolization may also lead to the formation of reductones, which, upon decomposition, yield some flavor compounds such as maltol, isomaltol, acetyl furan, and various aldehydes. Degradation of amino acids in the presence of reductones (the Strecker degradation) leads eventually to the formation of the major aroma compounds, the heterocyclic pyrrole, pyridine, pyrazine, thiazole, and oxazole.

In addition to the Maillard processes, several other reactions occurring during kilning may be responsible for the color and flavor of malt. The polyunsaturated fatty acids may be partially oxidized, via reactions catalyzed by lipoxygenase, to various volatile and nonvolatile aldehydes, alcohols, lactones, and acids, responsible for usually undesirable odors. Polyphenols may be converted to colored phlobaphenes. In specialty malts, still other chemical processes may be required. For instance, the caramelization reactions occurring when malts are kilned at very high temperatures result in the dark color and unique flavors of some food malts. The characteristic flavor and aroma of some lager-type beers are attributed to a high content of dimethyl sulfide (DMS) formed during careful kilning of malts rich in nitrogen; DMS is a degradation product of the sulfonium compound *S*-methylmethionine (Figure 9).

See also: **Barley**; **Beers**: History and Types; Biochemistry of Fermentation; Chemistry of Brewing; **Browning**: Nonenzymatic; **Starch**: Structure, Properties, and Determination; Sources and Processing

Further Reading

Briggs DE (1992) Barley germination: biochemical changes and hormonal control. In: Shewry PR (ed.) *Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology*, pp. 369–401. Wallingford, UK: CAB International.

Briggs DE (1998) *Malts and Malting*. London: Blackie Academic & Professional.

Briggs DE, Hough JS, Stevens T and Young TW (1981) *Malting and Brewing Science: Volume 1 Malt and Sweet Wort*, 2nd edn. London: Chapman & Hall.

Enari TM and Sopanen T (1986) Mobilisation of endospermal reserves during the germination of barley. *Journal of the Institute of Brewing* 92: 25–31.

Eskin NA (1990) *Biochemistry of Foods*. London: Academic Press.

Fincher GB (1992) Cell wall metabolism in barley. In: Shewry PR (ed.) *Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology*, pp. 413–437.Wallingford, UK: CAB International.

Fincher GB and Stone BA (1993) Physiology and biochemistry of germination in barley. In: MacGregor AW and Bhatty RS (eds) *Barley Chemistry and Technology*, pp. 247–295. St. Paul, MN: American Association of Cereal Chemists.

Hill RD and MacGregor AW (1988) Cereal α-amylases in grain research and technology. In: Pomeranz Y (ed.) *Advances in Cereal Science and Technology*, vol. IX, pp. 217–261. St. Paul, MN: American Association of Cereal Chemists.

Hudson OP (1986) Malting technology. *Journal of the Institute of Brewing* 92: 115–122.

MacGregor EA and MacGregor AW (1987) Studies of cereal α-amylase using cloned DNA. *CRC Critical Reviews in Biotechnology* 5: 129–142.

Palmer GH (1989) Cereals in malting and brewing. In: Palmer GH (ed.) *Cereal Science and Technology*, pp. 61–242. Aberdeen, UK: Aberdeen University Press.

Palmer GH and Bathgate GN (1976) Malting and brewing. In: Pomeranz Y (ed.) *Advances in Cereal Science and Technology*, vol. I, pp. 237–324. St. Paul, MN: American Association of Cereal Chemists.

Sebree BR (1997) Biochemistry of malting. *Technical Quarterly, Association of the Americas Master Brewers* 34: 148–151.

MANGANESE

C L Keen and S Zidenberg-Cherr, University of California, Davis, CA, USA

Background

The essentiality of manganese was established in 1931 when it was demonstrated that a deficit of it resulted in poor growth and impaired reproduction in rodents. It has long been appreciated that a deficiency of manganese can be a practical problem in the pig and poultry industry, and it has been suggested that manganese deficiency may be a problem in some human populations. Here, the recent literature related to manganese nutrition, metabolism, and metabolic function is briefly reviewed. For additional information, the reader is referred to more comprehensive reviews in the bibliography.

Chemical and Physical Properties

Manganese is widely distributed in the biosphere; it constitutes approximately 0.1% of the earth's crust making it the twelfth most abundant element. Manganese is a component of numerous complex minerals including pyroluosite, rhodochrosite, rhodanite, braunite, pyrochroite, and manganite. Chemical forms of manganese in their natural deposits include oxides, sulfides, carbonates, and silicates. Concentrations of manganese in groundwater normally range between 1 and 100 $\mu g\,l^{-1}$, with most values being below 10 $\mu g\,l^{-1}$. Typical airborne levels of manganese (in the absence of excessive pollution) range from 0.05 to 0.10 $\mu g\,m^{-3}$.

Manganese is a transition element located in column 7a on the periodic table. It can exist in 11 oxidation states, ranging from -3 to $+7$, with the most common valences being $+2$, $+4$, and $+7$. The $+2$ valence is the predominant form in biological systems and is the form that is thought to be maximally absorbed. The $+4$ valence occurs in MnO_2, and the $+7$ valence is found in permanganate.

The solution chemistry of manganese is relatively simple. The aquo-ion is resistant to oxidation in acidic or neutral solutions. It does not begin to hydrolyze until pH 10, and therefore, free Mn^{2+} can be present in neutral solutions at relatively high concentrations. Divalent manganese is a d^5 ion and typically forms high-spin complexes lacking crystal field stabilization energies. The above properties as well as the large ionic radius and small charge-to-radius ratio result in manganese tending to form weak complexes compared to other first-row divalent ions such as Ni^{2+} and Cu^{2+}. Free Mn^{2+} has a strong isotropic EPR signal that can be used to determine its concentration in the low micromolar range. Mn^{3+} is also critical in biological systems. For example, Mn^{3+} is the oxidative state of manganese in superoxide dismutase, is the form in which transferrin binds manganese, and is probably the form of manganese that interacts with Fe^{3+}. Given its smaller ionic radius, the chelation of Mn^{3+} in biological systems would be predicted to be more avid than that of Mn^{2+}. Cycling between Mn^{3+} and Mn^{2+} has been suggested to be potentially deleterious to biological systems since it can generate free radicals.

Occurrence and Speciation in Foods

Manganese concentrations in typical food products range from 0.4 $\mu g\,g^{-1}$ (meat, poultry, fish) to 20 $\mu g\,g^{-1}$ (nuts, cereals, dried fruit). Teas, which are often listed as an excellent source of manganese, can contain anywhere from 300 to 900 $\mu g\,g^{-1}$ of the element. An important consideration with respect to food sources of manganese, however, is the extent to which the manganese is available for absorption. For example, while tea contains high amounts of the element, the high content of tannin found in tea binds manganese and prevents its absorption from the gastrointestinal tract. Similarly, while the concentration of manganese in cereal grains is significant, the high content of phytates and fiber constituents may bind manganese, limiting its absorption. Thus, while a calculation of intake based on nutrient composition of foods may show that manganese intake is high, the actual amount of manganese absorbed will vary among food sources. Similarly, although meat products contain low concentrations of manganese, absorption and retention of manganese from them are high, making them good sources of the element in the diet. (*See* **Bioavailability of Nutrients**; **Cereals**: Dietary Importance; **Meat**: Nutritional Value; **Tea**: Chemistry.)

Analysis

Although manganese is widely distributed in the biosphere, it occurs in only trace amounts in animal tissues. Tissue concentrations of 4–8 μM are considered high, while serum concentrations can be

as low as 5 nM. Owing to the high environmental levels of manganese relative to its concentration in animal tissues, considerable effort must be made to minimize contamination of samples during their collection and handling. As a general precaution, all glassware and plasticware used in the storage or processing of samples should be washed in 20% nitric acid for at least 48 h and then rinsed exhaustively with distilled double-deionized water. Washed glassware and plasticware should be stored under dust-free conditions prior to use. To reduce contamination, tissues containing very low manganese concentrations should be dissected with plastic or quartz knives rather than steel knives.

The most common analytical methods that can sensitively measure manganese (18 nmol l^{-1}) include neutron-activation analysis, X-ray fluorescence, proton-induced X-ray emission, inductively coupled plasma emission, electron paramagnetic resonance (EPR), and flameless atomic-absorption spectrophotometry (AAS). Currently, the most common method employed is flameless AAS. All of these methods, with the exception of EPR, measure the total concentration of manganese in the samples. EPR allows selective measurement of bound vs free manganese.

Physiological Role

Tissue Concentrations

The average human body contains between 200 and 400 μmol of manganese, which is fairly uniform in distribution throughout the body. There is relatively little variation among species with regard to tissue manganese concentrations. Manganese tends to be highest in tissues rich in mitochondria; its concentration in mitochondria is higher than in cytoplasm or other cell organelles. Hair can accumulate high concentrations of manganese, and it has been suggested that hair manganese concentrations may reflect manganese status. High concentrations of manganese are normally found in pigmented structures, such as retina, dark skin, and melanin granules. Bone, liver, pancreas, and kidney tend to have higher concentrations of manganese (20–50 nmol g^{-1}) than do other tissues. Concentrations of manganese in brain, heart, lung, and muscle are typically < 20 nmol g^{-1}; blood and serum concentrations are about 200 and 20 nmol l^{-1}, respectively. Typical milk concentrations are in the order of 1 μmol l^{-1}. Bone can account for up to 25% of total body manganese because of its mass. Bone manganese concentrations can be raised or lowered by substantially varying dietary manganese intake, but bone manganese is not thought to be a readily mobilizable pool. In contrast to the situation for several other essential trace elements, the fetus does not accumulate liver manganese before birth, and fetal concentrations are significantly less than adult concentrations. This lack of fetal storage can be attributed to the apparent lack of storage proteins and the fact that most manganese enzymes are not expressed prenatally.

Absorption and Transport

Absorption of manganese is thought to occur throughout the small intestine. The efficiency of manganese absorption is relatively low and is not thought to be under homeostatic control. For adult humans, manganese absorption has been reported to range from 2 to 15% when ^{54}Mn-labeled test meals are used and 25% when balance studies are conducted. Manganese absorption and retention are higher in neonates than in adults. Data from balance studies indicate that manganese retention from human milk and infant formula is elevated during infancy, suggesting that neonates may be particularly susceptible to manganese toxicosis.

The higher retention of manganese in young animals in relation to adults may reflect an immaturity of manganese excretory pathways. The avid retention of the small amount of manganese in milk, and the postnatal changes in its excretory pattern, underscore the fact that important changes in manganese metabolism occur during the neonatal period.

In experimental animals, high amounts of dietary calcium, phosphorus, fiber, and phytate have been shown to increase the requirements for manganese; such interactions presumably occur via the formation of insoluble manganese complexes in the intestinal tract with a concomitant reduction in the soluble fraction available for absorption. The significance of these dietary factors with regard to human manganese requirements remains to be clarified. Studies in avians have demonstrated that high dietary phosphorus reduced manganese deposition in bone by approximately 50%. Given that the average diet of many individuals may be considered marginal in manganese, and high in phosphorus, this antagonism may have important implications for human health. The low fractional absorption of manganese from soy formula has been related to its relatively high phytate content. The mechanism underlying this effect of soy protein on manganese absorption/retention has not been fully delineated. (*See* **Calcium**: Physiology; **Dietary Fiber**: Effects of Fiber on Absorption; **Phytic Acid**: Nutritional Impact.)

An interaction between iron and manganese has been demonstrated in several species. Manganese absorption increases under conditions of iron deficiency

in both experimental animals and humans, whereas high amounts of dietary iron can accelerate the development of manganese deficiency in some species. The mechanisms underlying the interactions between iron and manganese have not been identified; however, they probably involve either a transport site or a ligand. (*See* **Iron**: Physiology.)

Manganese entering the portal blood from the gastrointestinal tract may either remain free or become associated with α_2-macroglobulin, which is subsequently taken up by the liver. A small fraction enters the systemic circulation, where it may become oxidized to Mn^{3+} and bound to transferrin. Studies *in-vivo* suggest that the Mn^{3+} complex forms very quickly in blood, in contrast to the slow oxidation of Mn^{2+}-transferrin complex *in vitro*. Manganese uptake by the liver has been reported to occur by a unidirectional, saturable process with the properties of passive mediated transport. Once entering the liver, manganese enters one of at least five metabolic pools. One pool represents manganese taken up by the lysosomes, from which it is thought to be transferred subsequently to the bile canaliculus. The regulation of manganese is thought to be maintained in part through biliary excretion of the element; up to 50% of manganese injected intravenously can be recovered in the feces within 24 h. A second pool of manganese is associated with the mitochondria. Mitrochondria have a large capacity for manganese uptake, and it is thought that mitochondrial uptake and release of manganese may be related. A third pool of manganese is found in the nuclear fraction of the cell; the roles of nuclear manganese have not been delineated. A fourth manganese pool is incorporated into newly synthesized manganese proteins; biological half-lives for these proteins have not been agreed upon. The fifth identified intracellular pool of manganese is free Mn^{2+}. It is thought that fluctuations in the free manganese pool may be an important regulator of cellular metabolic control in a manner analogous to those for free Ca^{2+} and Mg^{2+}. Consistent with this idea, several studies in pancreatic islets have shown that manganese blocks glucose-induced insulin release by altering cellular calcium fluxes.

The mechanisms by which manganese is transported to and taken up by extrahepatic tissues have not been identified. Transferrin is the major manganese-binding protein in plasma; however, it is not known to what extent transferrin facilitates the uptake of manganese by extrahepatic tissue. Manganese uptake by extrahepatic tissue does not appear to be increased under conditions of manganese deficiency, suggesting a lack of inducible manganese-transport proteins.

Currently, there is limited information concerning the hormonal regulation of manganese metabolism. Fluxes in the concentrations of adrenal, pancreatic, and pituitary–gonadal axis hormones affect tissue manganese concentrations; however, it is not clear to what extent hormone-induced changes in tissue manganese concentrations are due to alterations in cellular uptake of manganese-activated enzymes or metalloenzymes.

Biochemical Functions

Manganese functions as a constituent of metalloenzymes and as an enzyme activator. Manganese-containing enzymes include arginase, pyruvate carboxylase, and manganese-superoxide dismutase (MnSOD). Arginase, the cytosolic enzyme responsible for urea formation, contains 4 mol Mn^{2+} per mol of enzyme. Although the activity of this enzyme can be lower in managanese-deficient animals than in controls, the functional significance of the reduction has not been defined. With experimental diabetes, liver and kidney manganese concentrations and arginase activity can be markedly elevated. This manganese effect on arginase has been suggested to be due to an effect of Mn^{2+} on the conformational properties of the enzyme with a resultant modification of arginase activity. Whether this finding implies an increased manganese requirement for diabetics has not been determined.

Pyruvate carboxylase, the enzyme that catalyzes the first step of carbohydrate synthesis from pyruvate, contains 4 mol Mn^{2+} per mol enzyme. Although the activity of this enzyme can be lower in manganese-deficient animals than in controls, gluconeogenesis is not markedly inhibited in deficient animals.

MnSOD catalyzes the disproportionation of O_2^- to H_2O_2 and O_2. The activity of MnSOD in tissues of manganese-deficient rats can be significantly lower than in controls. That this reduction is functionally significant is suggested by the observation of higher than normal levels of hepatic mitochondrial lipid peroxidation in deficient rats. Tissue MnSOD activity can be increased by several diverse stressors including alcohol, ozone, interleukin-1, and tumor necrosis factor-α, presumably as a consequence of stressor-associated increases in cellular free radical (or oxidized target(s)) concentrations. The increase in MnSOD activity can be attenuated in manganese-deficient animals, potentially increasing their sensitivity to these insults.

For manganese-activated reactions, the metal can act by binding either to the substrate (such as ATP) or directly to the protein, resulting in conformational changes. In contrast to the relatively few manganese metalloenzymes, there are a large number of manganese-activated enzymes, including hydrolases, kinases, decarboxylases, and transferases. Many of

these metal activations are nonspecific in that other metal ions, particularly Mg^{2+}, can replace Mn^{2+}. An exception to the nonspecific manganese activation of enzymes is the manganese-specific activation of glycosyltransferases. Several manganese deficiency-induced pathologies have been attributed to a low activity of this enzyme class. A second example of an enzyme that may be specifically activated by manganese is phosphoenolpyruvate carboxykinase; low activities of this enzyme can occur in manganese-deficient animals.

A third example of a manganese-activated enzyme is glutamine synthetase. This enzyme, found in high concentrations in the brain, catalyzes the reaction $NH_3 + \text{glutamate} + ATP \rightarrow \text{glutamine} + ADP + Pi$. It can be speculated that a manganese deficiency-induced reduction in its activity could help to explain some of the behavioral defects associated with manganese deficiency. It should be noted that this enzyme can be inactivated by oxygen radicals; thus, a manganese deficiency-induced reduction in MnSOD activity theoretically could act to further depress the activity of glutamine synthetase.

Manganese Deficiency

Manganese deficiency has been demonstrated in several species, including rats, mice, pigs, and cattle. Signs of manganese deficiency include impaired growth, skeletal abnormalities, impaired reproductive performance, ataxia, and defects in lipid and carbohydrate metabolism.

The effects of manganese deficiency on bone development have been studied extensively. In most species, manganese deficiency can result in shortened and thickened limbs, curvature of the spine, and swollen and enlarged joints. The basic biochemical defect underlying the development of these bone defects is a reduction in the activities of glycosyltransferases; these enzymes are necessary for the synthesis of the chondroitin sulfate side-chains of proteoglycan molecules. In addition, manganese deficiency in adult rats can result in an inhibition of both osteoblast and osteoclast activity. The implications of this observation for human bone disease need to be ascertained.

One of the most striking effects of manganese deficiency occurs during pregnancy. When pregnant animals are deficient in manganese, their offspring exhibit a congenital, irreversible ataxia characterized by incoordination, lack of equilibrium, and retraction of the head. This condition is the result of impaired development of the otoliths, the calcified structures in the inner ear responsible for normal body righting reflexes. The block in otolith development is secondary to depressed proteoglycan synthesis due to the low activity of manganese-requiring glycosyltransferases.

Defects in carbohydrate metabolism, in addition to those described above, have been shown in manganese-deficient rats and guinea-pigs. In the guinea-pig, perinatal manganese deficiency results in pancreatic pathology, with animals exhibiting aplasia or marked hypoplasia of all cellular components. Manganese-deficient guinea-pigs and rats given a glucose challenge respond with a diabetic-type glucose tolerance curve. In addition to its effect on pancreatic tissue integrity, manganese deficiency can directly impair pancreatic insulin synthesis and secretion. The mechanism(s) underlying the effects of manganese on pancreatic insulin metabolism have not been fully delineated; however, it is known that insulin mRNA levels are reduced in the deficient animal. In addition, insulin sensitivity in adipose tissue is reduced in manganese-deficient rats, a phenomenon that may be related to fewer insulin receptors per adipose cell in the deficient animals. Finally, the effect of manganese deficiency on insulin production may also be due to the destruction of pancreatic β-cells. It is worth noting that constitutive pancreatic MnSOD activity is lower than in most tissues; this, coupled with the observation that most diabetogenic agents function via the production of free radicals with subsequent tissue damage, suggests that an additional mechanism underlying pancreatic dysfunction in the manganese-deficient animals may be free-radical mediated. (*See* **Carbohydrates**: Digestion, Absorption, and Metabolism; **Glucose**: Glucose Tolerance and the Glycemic (Glycaemic) Index.)

Although the majority of studies concerning the influence of manganese deficiency on carbohydrate metabolism have been conducted with experimental animals, there is one report in the literature of an insulin-resistant diabetic patient who responded to oral doses of manganese with decreasing blood glucose concentrations. While this is an intriguing case report, others have reported a lack of an effect of oral manganese supplements in diabetic subjects, and low blood manganese concentrations have not been found to be a characteristic of diabetics.

Abnormal lipid metabolism is also characteristic of manganese deficiency; a lipotrophic effect of manganese has been suggested. Severely manganese-deficient animals can be characterized by high liver fat, hypocholesterolemia and low high-density lipoprotein concentrations. As stated above, tissue lipid peroxidation rates can be increased in manganese-deficient animals, possibly as a result of low tissue MnSOD activity. (*See* **Fatty Acids**: Metabolism; **Lipoproteins**.)

There is considerable debate as to the extent to which manganese deficiency affects humans under free living conditions. It has been shown that manganese deficiency can be induced in humans under highly controlled experimental conditions. In this study, manganese deficiency was induced in adult male subjects by feeding a manganese-deficient diet for 39 days. The subjects developed a temporary dermatitis, and increased serum calcium and phosphorus concentrations and increased alkaline phosphatase activity, suggestive of bone resorption.

During the last decade, several diseases have been reported to be characterized in part, by low blood manganese concentrations. These diseases include epilepsy, mseleni disease, maple syrup urine disease and phenylketonuria, Down's syndrome, osteoporosis, and Perthes disease. The finding of low blood manganese levels in subsets of individuals with the above diseases is significant since blood manganese levels can reflect soft-tissue manganese concentrations. The reports of low blood manganese concentrations in individuals with epilepsy are particularly intriguing, given the observations that manganese-deficient animals can show an increased susceptibility to drug and electroshock-induced seizures and a genetic model for epilepsy in rats (the GEPR rat) is characterized by low blood manganese concentrations. Given that Mn^{2+} is implicated in activation of glutamine synthetase, a Mn^{2+}-specific brain ATPase, production of cyclic-AMP, altered synaptosomal uptake of noradrenalin and serotonin, glutamate, GABA and choline metabolism and biosynthesis of acetylcholine receptors, it is evident that a deficiency of manganese may contribute to the pathology of epilepsy at multiple points.

Although it is evident from the above that the role of altered manganese metabolism in several disease states needs to be clarified, evidence of widespread manganese deficiency in human populations is still lacking. Typically, manganese intakes are within the USA estimated safe and adequate daily dietary intakes, which are as follows: 0.3–1.0 mg d^{-1} for infants, 1.0–3.0 mg d^{-1} for children, and 2.0–5.0 mg d^{-1} for older children, adolescents, and adults.

Manganese Toxicity

In domestic animals, the major reported lesion associated with chronic manganese toxicity is iron deficiency, resulting from an inhibitory effect of manganese on iron absorption. Additional signs of manganese toxicity in domestic animals can include depressed growth, depressed appetite, and altered brain function.

In humans, manganese toxicity represents a serious health hazard, resulting in severe pathologies of the central nervous system. In its most severe form, the toxicosis is manifested by a permanent crippling neurological disorder of the extrapyramidal system, which is similar to Parkinson's disease. In its milder form, the toxicity is expressed by hyperirritability, violent acts, hallucinations, disturbances of libido, and incoordination. The above symptoms, once established, tend to persist even after the manganese body burden returns to normal. While the majority of reported cases of manganese toxicity occur in individuals exposed to high concentrations of airborne manganese (> 5 mg m^{-3}), subtle signs of manganese toxicity including delayed reaction time, impaired motor coordination, and impaired memory have been observed in workers exposed to airborne manganese concentrations lower than 1 mg m^{-3}. Manganese toxicity has been reported in an individual who consumed high amounts of manganese supplements for several years and in individuals who have consumed water containing high levels of manganese. There has been concern recently that the risk for manganese toxicity may be increasing in some areas because of the use of methylcyclopentadenyl manganese tricarbonyl in gasoline as an antiknock agent; however, this is an issue of active debate.

In additional to neural damage, reproductive and immune system dysfunction, nephritis, testicular damage, pancreatitis, lung disease, and hepatic damage can occur with manganese toxicity, though the frequency of these disorders is unknown. Similarly to the cases in humans, chronic manganese toxicity in rhesus monkeys is characterized by muscular weakness, rigidity of the lower limbs, and neuron damage in the substantia nigra. Neural toxicity is a consistent finding in rats exposed to chronic manganese toxicity. The mechanisms underlying the toxicity of manganese have not been agreed upon but probably involve both endocrinological dysfunction and excessive tissue oxidative damage.

To date, cases of manganese toxicity in humans have only been reported for adults; however, infants may be at a high risk for manganese toxicity owing to a high absorptive capacity for the element and/or an immature excretory pathway for it. If manganese is taken up by extrahepatic tissues via the manganese–transferrin complex, the developing brain may be particularly sensitive to manganese toxicity owing to the high number of transferrin receptors elaborated by neuronal cells during development, coupled with the putative need by neural cells for transferrin for their differentiation and proliferation. Studies aimed at evaluating the relative sensitivity of the developing brain to manganese toxicity are needed.

Biomarkers for Manganese Status

At present, reliable biomarkers for the assessment of manganese status have not been identified. Whole-blood manganese concentrations have been reported to be reflective of soft-tissue manganese levels in rats; however, it is not known whether a similar relationship holds for humans. Plasma manganese concentrations have been shown to decrease in individuals fed manganese-deficient diets, and to be slightly higher than normal in individuals consuming manganese supplements. Lymphocyte MnSOD activity has been reported to be increased in individuals who consume manganese supplements; however, its value as a biomarker for manganese status may be complicated owing to the number of cytokines and disease states, which may also increase its expression.

See also: **Bioavailability of Nutrients**; **Calcium**: Physiology; **Carbohydrates**: Digestion, Absorption, and Metabolism; **Cereals**: Dietary Importance; **Dietary Fiber**: Effects of Fiber on Absorption; **Fatty Acids**: Metabolism; **Glucose**: Glucose Tolerance and the Glycemic (Glycaemic) Index; **Iron**: Physiology; **Lipoproteins**; **Meat**: Nutritional Value; **Phytic Acid**: Nutritional Impact; **Tea**: Chemistry

Further Reading

Baly DL, Schneiderman JS and Garcia-Welsh AL (1990) Effect of manganese deficiency on insulin binding, glucose transport and metabolism in rat adipocytes. *Journal of Nutrition* 120: 1075–1079.

Brandt M and Schramm VL (1986) Manganese. In: Schramm VL and Wedler FC (eds) *Mammalian Manganese Metabolism and Manganese Uptake and Distribution in Rat Hepatocytes. Manganese in Metabolism and Enzyme Function*, pp. 3–16. New York: Academic Press.

Cooper WC (1984) The health implications of increased manganese in the environment resulting from the combustion of fuel additives: a review of the literature. *Journal of Toxicology and Environmental Health* 14: 23–46.

Davidsson L, Cederblad A, Lonnderdal B and Sandstrom B (1991) The effect of individual dietary components on manganese absorption in humans. *American Journal of Chemical Nutrition* 54: 1065–1070.

Davis CD and Greger JL (1992) Longitudinal changes of manganese-dependent superoxide dismutase and other indexes of manganese and iron status in women. *American Journal of Clinical Nutrition* 55: 747–752.

Friedman BJ, Freeland-Graves JH, Bales CW *et al.* (1987) Manganese balance and clinical observations in young men fed a manganese-deficient diet. *Journal of Nutrition* 17: 133–143.

Keen CL and Leach RM (1987) Manganese. In: Seiler HG and Sigel H (eds) *Handbook on Toxicity of Inorganic Compounds*, pp. 405–415. New York: Marcel Dekker.

Keen CL, Lonnerdal B and Hurley LS (1984) Manganese. In: Frieden E (ed.) *Biochemistry of the Essential Ultratrace Elements*, pp. 89–132. New York: Plenum.

Keen CL and Zidenberg-Cherr S (1990) Manganese. In: Brown ML (ed.) *Present Knowledge in Nutrition*, 6th edn, pp. 279–286. Washington, DC: ILSI.

Kondakis XG, Makris N, Leotsinidis M, Prinou M and Papapetropoulos T (1989) Possible effects of high manganese concentrations in drinking water. *Archives of Environmental Health* 44: 175–178.

Ono M, Sekiya C, Ohhira M *et al.* (1991) Elevated level of serum manganese-superoxide dismutase in patients with primary biliary cirrhosis: Possible involvement of free radicals in the pathogenesis in primary biliary cirrhosis. *Journal of Laboratory and Clinical Medicine* 118: 476–483.

MANGOES

J F Morton†, University of Miami, Coral Gables, FL, USA

Introduction

One of the most celebrated and heavily consumed of tropical fruits, the mango is often represented in ancient Indian artwork and is said to have been seen by Alexander the Great and his soldiers in the Indus Valley in 327 BC. In the past two centuries, it has spread to most parts of the tropics, and some subtropical areas as well; it has entered international commerce, and is in increasing demand, especially by relocated people now outside the warm climates, as well as by gourmets in temperate regions. It has therefore earned the attention of food scientists everywhere. (*See* **Fruits of Tropical Climates**: Commercial and Dietary Importance.)

Description

The mango is borne by a handsome tree which, under favorable conditions, grows up to 40 m or more in height and, with age, reaches an equal canopy spread.

†deceased.

It is densely foliaged, with rosettes of oblong-lanceolate, evergreen leaves, conspicuously red or yellow when immature. Its spectacular, reddish or yellowish pyramidal flower sprays stand out from the foliage in the cool season and the young fruits develop in late spring. The harvesting season extends from early summer through autumn because there are early, medium, and late cultivars.

The mango fruit, hanging in clusters on long stems, is extremely variable; it may be round, oval, egg- or kidney-shaped, or long-elliptical, often oblique at the base, and ranging in weight from 150 to 850 g. The skin is smooth or leathery, and varies from light- or dark-green, or all-yellow, to yellow flushed with pink, rose, bright- or dark-red, or nearly purple, generally with a multitude of minute yellow dots (Figures 1 and 2).

Ripe fruits are agreeably aromatic or, to some degree, redolent of turpentine. The flesh is yellow or orange, juicy, and of fine, melting texture, more or less fibrous ('stringy') in unimproved types. The somewhat peachlike flavor varies from subacid to sweet, is rich and mellow, and usually slightly but agreeably resinous. Some seedlings (now rare) have a pronounced turpentine tang. A 'beard' of fiber clings to the relatively large seed, flattish and firm-shelled. Some types with fiber-free flesh are classed as 'freestone' because the seed separates easily from the flesh. In most mangoes, however, the flesh must be cut from the seed. That which remains clinging to the seedcoat fibers is the most flavorful, and any real mango fancier will relish gnawing it off. In fact, there is a special mango fork to hold the seed for this purpose. The seed kernel is firm but tender, starchy and edible.

Origin and Distribution

It is believed that the mango is native to an area extending from the foothills of the Himalayas of

Figure 1 An assortment of red and yellow mangoes from home gardens in Florida: top row (left to right), Fascell, Brooks, Gibbons, Cecil; second row, Hodges, Yellow Brooks, Borsha, Kent; three in front, Sandersha, Peach, Haden; in saucer, Cecil. Reproduced from Mangoes, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Figure 2 Madame Francis, the leading mango of Haiti, yellow when ripe, is polyembryonic, reproduces true to type, and is grown from seed for export. It comes into season a month or two earlier than most mangoes grown in Florida. Reproduced from Mangoes, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

India and Burma, southward into parts of Malaysia, and ranges from sea level to 1500 m. However, more than 4000 years of domestication have obliterated its natural boundaries. It is cultivated from southern China to the Philippines, in Madagascar, in mild parts of South, Central and even northern Africa, the Near East, Mediterranean Europe, and the Canary Islands, throughout the islands of the Pacific, the Bahamas, the Caribbean Islands, the lowlands of Central America, northern South America, Mexico, and in a few sheltered spots of southern California. It is a very common dooryard fruit tree and a commercial crop in southern Florida and the Florida Keys.

Varieties

There are several species of *Mangifera* in Malaysia, but only *M. indica* has favorable fruit qualities for exploitation. Many of the natural forms of mangoes in India, Cambodia, and the Philippines are excellent but have never been widely promoted in the western world because they lack the red coloration that marketers feel is an essential appeal. Similarly, the Edward mango in Florida, of the highest quality, is modestly (but beautifully) pastel and has the additional handicap of being a shy and irregular bearer. Much selection and breeding in Florida has led to the naming and commercialization of certain high-colored, fairly regular-bearing, and good-shipping cultivars that are now being adopted for practical purposes in Latin America and elsewhere. The three leading cultivars at present are Tommy Atkins, Keitt, and Kent, the last two requiring ethylene treatment to enhance the color. In India, more than 1000 cultivars have resulted from asexual propagation of superior chance seedlings over the past 80 years. Hawaii and many other mango-growing centers of the world have ongoing mango selection and improvement programs aimed at combining the best economic and esthetic factors in a reliably bearing and disease-resistant cultivar, for a particular region. An important goal should be the development of dwarf trees (like the Indian cultivar, Neelum) not only to save space but

also to facilitate harvesting and all other operations. In northern Colombia there is a miniature (egg-size) mango, Azucar, with colorful skin and small seed, which ought to be adopted for gift-box packing.

Propagation

Improved cultivars are propagated vegetatively by various techniques of budding or grafting in order to reproduce the exact characters of the parent. They are not grown from seed because seedlings grown from the usual monoembryonic seeds are noted for their variation. The homeowner may plant a mango seed and be content with the result, but would be better advised to purchase a grafted tree which will bear sooner and provide a crop of predictable quality. Only polyembyronic mangoes are propagated by seed for commercial purposes. The majority of these reproduce 'true' to the parent. Of course, plantations of monoembryonic seedlings are necessary for research and selection in breeding programs.

Soil and Culture

Mango trees are not particularly sensitive to soil type and they will prosper even on oolitic limestone, but they respond well to appropriate fertilization, irrigation, and spraying to control insects and fungus diseases. In humid climates, the mango is subject to anthracnose, which attacks the flowers and considerably reduces production. A much more serious problem is the fungus causing powdery mildew on mango trees from Florida to India. There are a number of scale insects which attack the tree and its fruit. Mangoes to be exported or imported may be subject to fumigation, hot water, or other treatments to avoid the spread of fruit flies and the mango seed weevil. Mango trees may suffer from lack of adequate boron, copper, iron, manganese, or magnesium, and need regular attention to these deficiencies.

Yield and Handling

Mango trees are long-lived and the yield varies with the cultivar but always increases with age. A well-cared-for, fruitful cultivar will bear 200–300 fruits per year during the first 10–20 years, and twice that amount in the next 20–40 years. In the Far East, older trees of special cultivars have borne up to 5000 fruits or more in a single season. In Puerto Rico, some Florida cultivars have yielded roughly 30 000–60 000 kg of fruits per hectare. However, after its first 10 years, the mango is noted for being an irregular bearer, providing a heavy crop one year and a light crop the next. Certain branches may bear one year and other branches the following year, and bearing is strongly influenced by weather conditions. In some tropical climates, mango trees may be partially in bloom and partially in fruit throughout the year. There have been various efforts to enhance pollination, and promote better fruit-setting by smudging, girdling, deblossoming, or by spraying with chemicals to retard vegetative growth, but results are unstable and discouraging.

Mangoes are picked when fully mature but slightly unripe for domestic use, and allowed to soften at room temperature. For marketing, they are harvested earlier and some cultivars are treated with ethylene to achieve uniform color. In any case, the fruit should be washed immediately after clipping to remove the gummy sap that exudes from the stem. Otherwise, it will 'burn' the skin, resulting in black stains which lead to decay.

Fully ripe mangoes keep for several days under home refrigeration. They can be frozen whole for future use but must be eaten immediately after thawing. Sliced mango flesh in sugar syrup, with added lime juice to prevent discoloration, is easily frozen in leakproof, sealed plastic bags.

Keeping quality varies with the cultivar. The best storage treatment for each must be worked out by experimentation. But all mangoes are subject to chilling injury, which generally occurs at storage temperatures below 15 °C. Trials in India with the high-quality Alphonso cultivar demonstrated that the freshly picked fruit can endure 2 weeks of storage at 7 °C and subsequent ripening at 30 °C, but longer refrigeration brings on 'low-temperature breakdown'. Ripening of Tommy Atkins and Kent mangoes has been successfully delayed for 3 weeks at 13 °C, 90–100% relative humidity, and atmospheric pressure of 76–152 mmHg, and the fruits have thereafter ripened satisfactorily. (*See* **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Ripening of Fruit**; **Storage Stability**: Mechanisms of Degradation.)

In general, spoilage has been retarded or reduced by prestorage dipping in hot water, or a solution of maleic hydrazide, benomyl, or calcium chloride. Plastic wraps reduce weight loss but do not retard decay. Waxing is beneficial in delaying ripening, but it has the undesirable effect of preventing full coloring.

Preparation for Serving and Various Food Uses

As mentioned earlier, 'freestone' mangoes can be cut through to the stone and then the two halves are held in the palms of the hands and gently twisted to

Table 1 Food value of mangoes (raw)[a]

Nutrients and units	*Mean amount per kg, edible portion*
Proximate (g)	
Water	817.10
Food energy (kcal)	650
(kJ)	2730
Protein (N × 6.25)	5.10
Total lipid (fat)	2.70
Carbohydrate, total	170.0
Fiber[b]	8.40
Ash	5.00
Minerals (mg)	
Calcium	100
Iron	1.30
Magnesium	90
Phosphorus	110
Potassium	1560
Sodium	20
Zinc	0.40
Copper	1.10
Manganese	0.27
Vitamins (mg)	
Ascorbic acid	277.0
α-Tocopherol	11.2
Thiamin	0.58
Riboflavin	0.57
Niacin	5.84
Pantothenic acid	1.60
Vitamin B_6	1.34
Folacin (μg)	
Vitamin B_{12} (μg)	0
Vitamin A (RE)	389
(IU)	3894
Lipids (g)	
Fatty acids	
Saturated, total	0.660
4:0	
6:0	
8:0	
10:0	
12:0	0.010
14:0	0.090
16:0	0.520
18:0	0.030
Monounsaturated, total	1.010
16:1	0.480
18:1	0.540
20:1	
22:1	
Polyunsaturated, total	0.510
18:2	0.140
18:3	0.370
18:4	
20:4	
20:5	
22:5	
22:6	
Cholesterol (mg)	0
Phytosterols (mg)	
Amino acids (g)	
Tryptophan	0.080
Threonine	0.190
Isoleucine	0.180
Leucine	0.310
Lysine	0.410
Methionine	0.50
Cystine	
Phenylalanine	0.170
Tyrosine	0.100
Valine	0.260
Arginine	0.190
Histidine	0.120
Alanine	0.510
Aspartic acid	0.420
Glutamic acid	0.600
Glycine	0.210
Proline	0.180
Serine	0.220

[a]With refuse = 300 g.
[b]Insoluble dietary fiber as determined by the neutral detergent fiber method (1.08 g per 100 g).
RE, retinol equivalents; IU, international units.
Reproduced from Gebhardt SE, Cutrufelli R and Matthews RH (1982) *Composition of Foods: Fruit and Fruit Juices, Raw, Processed, Prepared*. Agriculture Handbook no. 8–9, p. 168. Washington, DC: Consumer Nutrition Center, United States Department of Agriculture, Human Nutrition Information Service.

separate them from the stone. Ordinary, somewhat fibrous mangoes, can be served 'in the half-shell' by cutting off both plump sides. The center slab is then peeled and the flesh cut from the stone. Otherwise, the whole fruit may be peeled and the flesh sliced for whatever use: eating as dessert, adding to icecream, milkshake, using as pie filling, stewing as sauce, or making into preserves. Of course, if the flesh is fibrous, it should be pressed through a sieve.

In the tropics, there are many seedling mangoes so fibrous that they cannot be sliced. Children and adults will bruise the fruit by massaging it between the hands, then make a hole in the stem end and suck out the juice. This should not be done by anyone who is sensitive to the irritant in the mango peel.

Many immature mangoes are blown down by spring winds. These are usually salvaged by peeling, slicing, and cooking as a substitute for apple in pie. They may also be cooked as a sauce, or used with tamarind pulp and spices in preparing green mango chutney. In India, thin slices of unripe mango are sprinkled with turmeric, then dried and powdered for use as seasoning.

Mango nectars and peeled, seeded, ripe mangoes, halved or sliced, are commonly canned in Latin America and South-east Asia and these products find a good market abroad. However, more attention must be given to the thorough coating of the inside of the metal can, to avoid off-flavors and spoilage.

In India, the juicy pulp strained from ultrafibrous ripe mangoes has been dehydrated and utilized as 'cereal flakes.' Mango juice powder was developed in the Philippines years ago and technologists in Puerto Rico, Hawaii, and elsewhere are developing uses for the great numbers of mangoes that are out of

reach of markets or do not meet fresh-market or export specifications. Canadian scientists have proved osmotic dehydration feasible. Solar drying is giving good results in underdeveloped tropical countries. (*See* **Drying**: Drying Using Natural Radiation.)

Utilization of Mango Waste

In India, mango peel has been found to be a useful source of pectin equal to that derived from the apple. The seed kernel contains 11% fat usable in food industries, and the residual starch has been used to extend conventional flour stocks during times of food shortage in South-east Asia.

Food Values

The chemical constituents of the ripe mango, especially sugars (sucrose, glucose, and fructose) and acids, vary considerably with the cultivar (Table 1). Total sugars may be 11.20–16.80%. Ascorbic acid, according to analyses in Central America, ranges from 41.8 to 172.0 mg per 100 g. In India, ascorbic acid values may be as low as 13 mg per 100 g. Carotene ranges from 0.283 to 1.872 mg per 100 g. The mango is a poor source of calcium, phosphorus, and iron. Refer to individual nutrients.

Toxicity

It should be kept in mind that the mango is a member of the plant family Anacardiaceae, to which belong poison ivy, poison sumac, the marking nut, and other well-known sources of severe dermatitis. Mango propagators, making cuttings of budwood or other grafting material and rootstocks, may suffer chronic skin reactions, and some have to give up such work or wear protective plastic disposable gloves, or barrier creams. Those who wash and pack mangoes for shipping should also be gloved.

The sap of the tree is irritant, as is that of the fruit stem and, to a lesser extent, the peel of the fruit, especially before complete ripening. The fruit flesh can be contaminated by being sliced with the same knife that has been used to peel the fruit. Sensitive people (especially those who have been previously sensitized by the tree's sap or by contact with poison ivy or some other member of the family) may be able to eat the flesh of the mango if someone else does the peeling, provided that a fresh, clean knife is utilized to cut the flesh. Excessive contact with the peel, as in peeling a number of mangoes for making preserves, can induce a rash and blisters on the hands of someone not normally affected by handling whole mangoes.

When the mango tree is in bloom it does not disperse airborne pollen, but the flowers on the tree exhale an irritant chemical into the atmosphere which is a common source of itching ears, swelling of the eyelids, and respiratory irritation.

See also: **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Drying**: Drying Using Natural Radiation; **Fruits of Tropical Climates**: Commercial and Dietary Importance; **Ripening of Fruit**; **Storage Stability**: Mechanisms of Degradation

Further Reading

Gebhardt SE, Cutrufelli R and Matthews RH (1982) *Composition of Foods: Fruit and Fruit Juices, Raw, Processed, Prepared*. Agriculture Handbook no. 8–9, p. 168. Washington, DC: Consumer Nutrition Center, United States Department of Agriculture, Human Nutrition Information Service.

Gonzalez-Agailar GA, Wang CY and Buta JG (2000) Maintaining quality of fresh-cut mangoes using antibrowning agents and modified atmosphere packaging. *Journal of Agricultural Food Chemistry* 48(9): 4204–4208.

Krishnamurthy S and Joshi SS (1989) Studies on low temperature storage of Alphonso mango. *Journal of Food Science and Technology* 26(4): 177–180.

Kulkarni VJ and Rameshwar A (1989) Effect of deblossoming and defruiting on off-season flowering and fruiting in mango (*Mangifera indicia* L.). *Scientia Horticulturae* 39: 143–184.

Morton JF (1987) *Fruits of Warm Climates*. Distributed by Creative Resource Systems, Winterville, North Carolina.

Rankins J, Hopkinson S and Diop M (1989) Palatability and nutritional significance of solar dried mangoes for Senegal. *Ecology of Food and Nutrition* 23: 131–140.

Singh LB (1960) *The Mango – Botany, Cultivation and Utilization*. London: Leonard Hill (Books).

Manihot *See* **Cassava**: The Nature of the Tuber

Manufacturing *See* **Biscuits, Cookies, and Crackers**: Methods of Manufacture; **Brandy and Cognac**: Armagnac, Brandy, and Cognac and their Manufacture; **Butter**: The Product and its Manufacture; **Cakes**: Methods of Manufacture; **Canning**: Cans and their Manufacture; **Caramel**: Methods of Manufacture; **Casein and Caseinates**: Methods of Manufacture; **Cheeses**: Manufacture of Extra-hard Cheeses; Manufacture of Hard and Semi-hard Varieties of Cheese; **Cider (Cyder; Hard Cider)**: The Product and its Manufacture; **Dressings and Mayonnaise**: The Products and Their Manufacture; **Gin**: The Product and its Manufacture; **Ice Cream**: Methods of Manufacture; **Jams and Preserves**: Methods of Manufacture; Chemistry of Manufacture; **Low-fat Foods**: Types and Manufacture; **Margarine**: Methods of Manufacture; **Pasta and Macaroni**: Methods of Manufacture; **Port**: The Product and its Manufacture; **Sherry**: The Product and its Manufacture; **Whisky, Whiskey, and Bourbon**: Products and Manufacture; **Yogurt**: The Product and its Manufacture

Maples *See* **Sugar**: Palms and Maples

MARASMUS

P Codoñer-Franch and V Valls Bellés, Universidad de Valencia, Valencia, Spain

Introduction

Marasmus is the most prevalent severe form of malnutrition and the most important nutritional disease in developing countries. Deficiency of energy-giving foods predominates over protein deficiency. Its origin may be primary, due to insufficient food intake, often associated with infectious or parasitic diseases, or secondary, due to other conditions that interfere with the absorption or assimilation of nutrients. It contributes significantly to high rates of morbidity and mortality in childhood. It is now recognized as a public health problem, and action to reduce its extent is the responsibility of government nutrition policies and health personnel.

Malnutrition or undesirable physical or disease conditions related to nutrition can be caused by eating too little, too much, or by an unbalanced diet that does not contain all nutrients necessary for good nutritional status. Undernutrition is a lack of adequate energy, protein, and micronutrients to meet basic requirements for body maintenance, growth, and development. Today, the more comprehensive term of protein-energy (or protein-calorie) malnutrition (PEM) is universally accepted for describing undernutrition. PEM is really a term describing different types of clinical disorders resulting from several combinations and degrees of protein and energy deficiency. It includes a wide spectrum of clinical manifestations conditioned by the relative intensity and duration of protein and energy deficits. Its severity ranges from weight loss or growth retardation to distinct clinical syndromes. It is the most important nutritional disease in the developing countries; this accounts for its high prevalence and its relationship with child mortality rates, impaired physical growth, and inadequate social and economic development.

Dietary energy and protein deficiencies usually occur together, and are frequently associated with deficiencies of minerals and vitamins. Sometimes one predominates and, if severe enough, may lead to the major forms of PEM: marasmus, in which the deficiency is primarily of energy-giving foods, and kwashiorkor, with predominant protein deficiency. Frequently there are mixed pictures, with a combination of energy deficiency and chronic or acute protein deficit – marasmic kwashiorkor.

These conditions are diagnosed mainly on the basis of anthropometric measurements, especially weight and height. The two severe forms of malnutrition present very different appearences and clinical features; the overriding feature of nutritional marasmus is severe underweight. The degree of malnutrition is almost always based on the child's percentage of standard weight-for-age, and the weight in marasmus is $\leq 60\%$ of that expected for a child of that age (Table 1).

Today a cut-off point using standard deviations (SD) with respect to the mean weight-for-height is considered more appropriate, and expert clinicians

Table 1 Wellcome classification of types of protein-energy malnutrition (PEM)

Percentage of standard weight for age	*Edema*	
	Present	*Absent*
60–80	Kwashiorkor	Undernourishment
< 60	Marasmic kwashiorkor	Nutritional marasmus

diagnose marasmus using a constellation of clinical findings (visible severe wasting), and very low weight-for-height *Z*-score (≤ 3). Low weight-for-height is identified as the indicator of choice for screening severely malnourished children who are at increased risk of dying.

Epidemiology

The global magnitude of PEM is difficult to estimate with precision because many patients had not been recorded or had not received medical attention.

PEM young children is the most important nutritional problem in nonindustrialized countries. The World Declaration on Nutrition produced by the Food and Agriculture Organization (FAO) and World Health Organization (WHO) International Conference on Nutrition (ICN) held in Rome in 1992 states that about 780 million people in developing countries still do not have access to enough food to meet their basic daily needs for nutritional well-being. An estimated 50 million preschool children were wasting in 1995. Wasting rates change rapidly, however, especially in situations of emergency food shortage and population displacement.

It is especially distressing that there is such a high prevalence and increasing numbers of malnourished children under 5 years of age in different parts of Africa, Asia, Latin America, and the Caribbean. Malnutrition causes more than half of the mortality in children of this age; thus, 55% (or more than 6 million) preschool children died annually as a result of preventable diseases (12 million) which had a direct or indirect relationship with undernutrition. It has a potentiating effect on mortality, rather than acting in a simple additive fashion, and multipled the numbers of deaths caused by infectious diseases (Table 2).

It was recognized that nutritional marasmus is more prevalent than kwashiorkor and that these two serious clinical forms of PEM constitute the small tip of an iceberg of all malnourished children. They are relatively easy to detect by their clinical manifestations, and children with moderate or mild malnutrition can remain undiagnosed because often they do not have clear manifestations of malnutrition; rather, they are shorter (stunted) and/or thinner than would be expected for their age. Thus, energy deficiency plays a more important part than protein deficiency in global malnutrition.

Table 2 Prevalence and number of wasted preschool children, 1998

UN regions and subregions	*Prevalence of wasting < 5 years (%)*	*Population < 5 years (thousands)*	*Number wasted < 5 years*
Sub-Saharan Africa	9	100 036	9 003 240
Northern Africa	8	42 576	3 406 080
Meridional Asia	18	163 209	29 377 620
Latin America and the Caribbean	3	54 841	1 645 230
Central/West Europe	5		
Baltic States		31 712	1 585 600
Developing countries	11	536 905	58 971 150
World	**11**	**603 449**	**66 379 390**

Data from UNICEF (2000) *The State of the World's Children*. New York: Oxford University Press.

Causes

The etiology can be complex and three factors contribute principally to undernutrition:

1. Inadequate intake of food, with chronic insufficiency of nutrients.
2. Nutrient losses because of improper digestion and absorption, or infections that also hinder nutrient absorption and utilization.
3. Incremental requirements in physiologic circumstances (growth, gestation, lactation).

The origin of PEM can be primary, when it is the result of inadequate food intake and is almost always linked to conditions such as poverty, ignorance, infectious diseases, and low food availability. The PEM may be secondary, when it is the result of other diseases that lead to low food ingestion because of poor appetite or an inadequate digestion, absorption, or utilization of nutrients or conditions that result in increased nutritional requirements. This form of malnutrition is seen in malignancies, acquired immune deficiency syndrome (AIDS) or tuberculosis, and eating disorders such as anorexia nervosa. Primary malnutrition is most prevalent in developing countries and in disadvantaged groups (immigrants, refugees) who live in similar conditions in the developed world.

Therefore inadequate dietary intake and disease are immediate causes of malnutrition but the factors that contribute are related to the host, the agent (diet), and the environment. Availability of food, however, is just part of the picture. It is now recognized that malnutrition is only the overt sign, or symptom, of much deeper problems in society.

PEM is not confined to children but is much more prevalent during early childhood, mainly affecting infants and preschool children because of the young child's high needs for both energy and protein per kilogram and the impossibility for children to obtain food by their own means. Most descriptions of the condition concentrate on this age group.

The chronic intake of insufficient food can result in marasmus that develops gradually over several months or years. This process allows a series of metabolic adjustments that result in decreased nutrient demands and a nutritional balance compatible with a lower level of cellular nutrient availability. In marasmus there is a metabolically dynamic state change to which the affected person adapts to survive in a compensated manner. The cost of this adaptation includes functional limitations and metabolic adjustments are more stable when the situation develops slowly.

Hormonal and Metabolic Changes in Severe Malnutrition (Figure 1)

Hormones play important roles in the adaptive processes of energy and protein metabolism in severe PEM. Metabolic adaptation in wasting is associated with a decline in insulin and a rise in glucagon concentrations which results in increased rates of gluconeogenesis. During early starvation, protein catabolism is increased to provide amino acids for gluconeogenesis; decreased plasma insulin concentrations also allow for a decrease in glucose use by peripheral tissues. If the situation is prolonged, metabolic adjustments are made to spare muscle protein such as increased reliance by the central nervous system on ketone bodies as an oxidative fuel, thereby decreasing the need for increased rates of gluconeogenesis. Insulin decrease enhances lipolysis and results in increased conversion of free fatty acids (FFA) to ketone bodies.

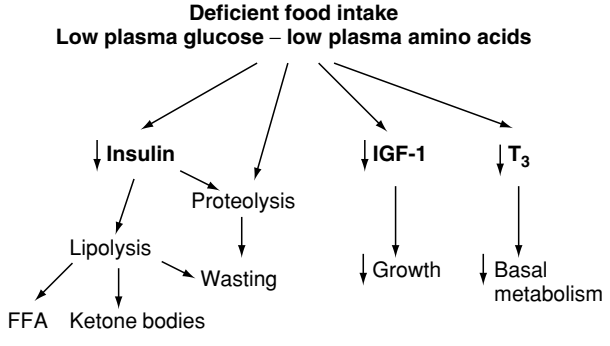

Figure 1 Endocrine changes in marasmus. IGF-1, insulin-like growth factor-1; T_3, triiodothyronine; FFA, free fatty acids.

Malnutrition frequently coexists with infection. Infection can stimulate the production of glucocorticoids that increase the release of gluconeogenic precursors, that is, amino acids from peripheral tissues, and accounts for increased proteolysis in this situation.

In the liver the synthesis of insulin-like growth factors (IGF) is decreased. This results in an alteration in the feedback for growth hormone (GH) and the increase of this hormone that contributes to lipolysis, mediated by direct catabolic effects on glucose and lipid metabolism.

The active thyroid hormone triiodothyronine (T_3) is decreased owing to deficient capturing of iodine by the gland and alteration in thyroxine (T_4) conversion. This produces a reduced thermogenesis and a spare in oxygen consumption. This metabolic adaptation leads to reduced metabolic cost.

Clinical Features

Growth Failure

A child's body responds to malnutrition with a deceleration or cessation of growth. In all cases the child fails to grow properly and presents a marked retardation in length or height-for-age. Maintenance of a satisfactory growth rate in the child depends on an adequate supply of energy. Subnormal intakes lead to a decline in growth velocity. Somatomedin C (IGF-1) is synthesized by the liver and other tissues in response to stimulation by GH. One of the major effects of GH is the promotion of linear growth and skeletal maturation. Somatomedin C is responsible for these effects and for increased synthesis of DNA, RNA, and protein in fibroblasts and chondrocytes. It also increases the rates of protein synthesis and cell proliferation in other tissues. Both adequate dietary protein and total energy contents are necessary for maintenance of somatomedin concentrations.

Undernutrition retards growth and this is true for all different parts of the body. Some parts are affected more than others; adipose tissue loses its fat, and muscle and liver lose some of their protein. If the undernutrition is severe, the adipocytes become unrecognizable as cells and the spaces between them are occupied by extracellular material. There is less change in the composition of the kidneys and heart, and the brain is the least affected of all.

If undernutrition is imposed early in life, when growth is faster, then complete recovery in size may never be possible. If, at the stage of development when the cells are dividing rapidly, cell division is hindered, the cell divisions lost during the period of undernutrtion will never be regained, even after

rehabilitation. The result is a child shorter than normal (stunting).

In severe malnutrition there is a relative increase in body water and decrease in body potassium. The extracellular fluid volume is better preserved than cell mass; the shrunken cells retain their fluid covering. The noncollagen proteins are reduced to about one-half and the collagen protein do not change significantly. Skin and bone account for about 70% of total body collagen. The child appears to be 'skin and bones.'

Wasting

Loss of tissue mass during starvation involves oxidation of endogenous fuel stores in order to generate energy required for life-sustaining metabolic activities. A decrease in energy intake is quickly followed by a decrease in energy expenditure. When the decrease in energy expenditure cannot compensate for the insufficient intake, body fat is mobilized with a decrease in adiposity and weight loss. Lean body mass diminishes at a slower rate, mainly as a consequence of muscle protein catabolism with major flow of amino acids, primarily alanine, that contributes to energy sources. As the cumulative energy deficit becomes more severe, subcutaneous fat is markedly reduced, and protein catabolism leads to muscular wasting.

Clinically, the muscles are always extremely wasted and loss of subcutaneous fat makes the ribs prominent. The loss of flesh is obvious, with little or no remaining subcutaneous fat. In contrast to the rest of the body, the belly is sometimes relatively protuberant. The skin, particularly over the buttocks, hangs in wrinkles. There may be pressure sores over bony prominences. These features and the sunken cheeks caused by disappearance of the Bichat fat pads give the marasmic child's face the appearance of a monkey or an old person (Figure 2).

Hair Changes

The hair usually shows changes. There is more frequently a change of texture and it becomes sparse, thin, and dry, without its normal sheen. It is easily pulled out without causing pain.

Appetite

Some children are anorexic while others may be ravenously hungry; they often violently suck their hands or make sucking noies, but they seldom tolerate large amounts of food and vomit easily.

Mental and Physical State

The child's mood may be anxious rather than apathetic. Marasmic children have a rather wide-awake appearance and cry easily, with an expression of misery and sadness (Figure 3). There is a marked weakness and the child frequently cannot stand without help.

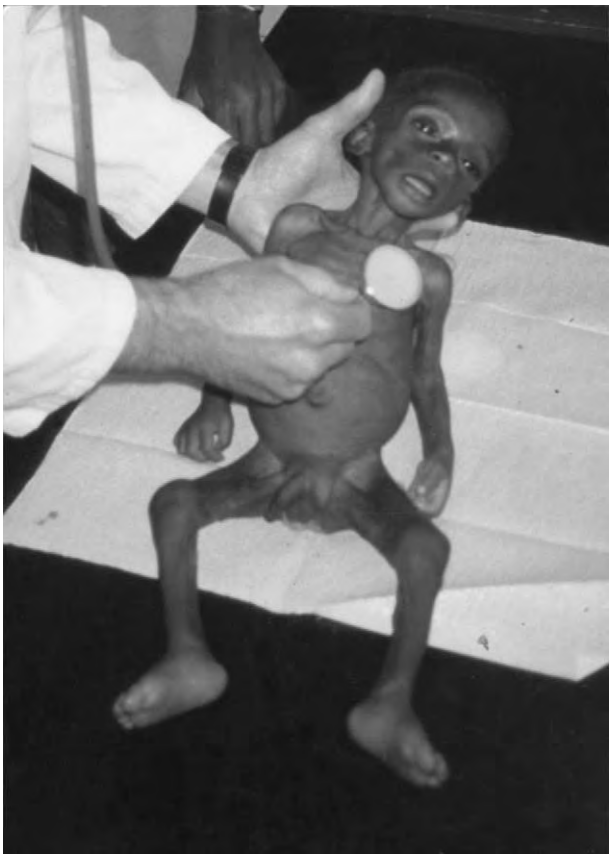

Figure 2 Marasmic infant, Rwanda 1997. Courtesy of Dr M Gracia.

Vital Constants

Heart rate, blood pressure, and body temperature may be low.

Biochemical and Metabolic Features

In severe PEM all the processes of the body show a reductive adaptation to current nutritional conditions that is better if harmonic changes in the metabolism of proteins, energy, and other nutrients occur.

Hemoglobin and hematocrit are usually moderately low. The serum iron may be normal, but iron body stores are usually depleted. Following nutritional rehabilitation there may be an additional decrease in iron stores because of the increased requirements for tissue synthesis and the greater needs for hematopoietic factors. If iron is not available, the patient is prone to develop severe anemia. Plasma ferritin is increased because of concurrent infections.

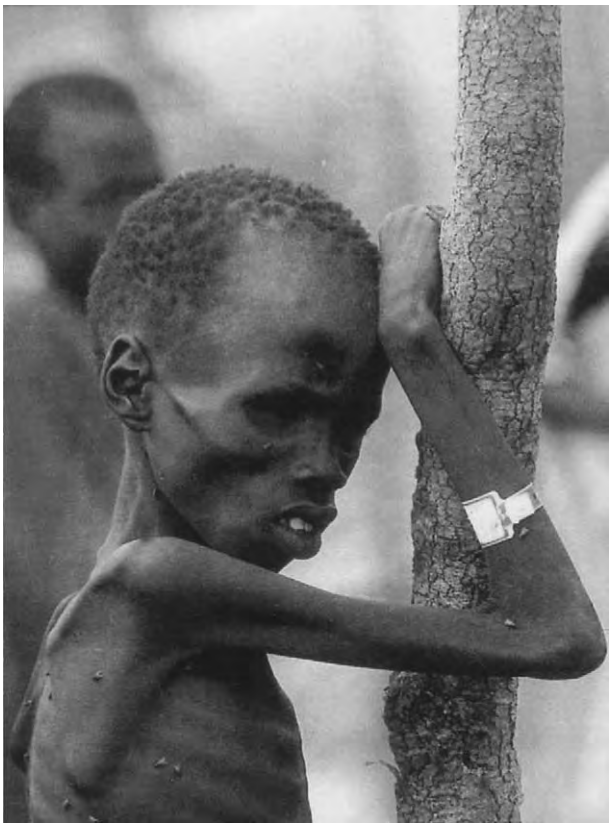

Figure 3 (see color plate 103) Expression of sadness in a marasmic person. Courtesy of Cruz Roja España.

Serum concentrations of total proteins are normal or moderately low, and the ratio of nonessential to essential amino acids in plasma is usually normal.

Excretion of hydroxyproline, 3-methylhistidine, creatinine, and urea nitrogen in urine is reduced.

Prealbumin and retinol-binding protein, two proteins with a short mean life (< 2 days) whose synthesis is energy-dependent, are decreased.

Blood glucose concentration remains normal, mainly at the expense of gluconeogenic amino acids and fats. Hypoglycemia is present especially after prolonged fasting or when malnutrition is complicated by serious infections.

Concentrations of cholesterol and triglycerides are low.

There is a deficiency in essential fatty acids and a imbalance between the *n*-3 and *n*-6 families that accounts for adverse effects (phospholipid deficiency in the retina).

Fluid and Electrolyte Disturbances

There is water and electrolyte characteristics in severe PEM, namely:

1. Intracellular potassium depletion, as body potassium is almost entirely stored within the lean tissues and muscle makes up most of the lean tissue.
2. Mild to moderate metabolic acidosis: due to rising circulating levels of ketone bodies in response to a change from a glucose-based to a lipid-based (FFA and ketone bodies) energy supply to minimize protein catabolism in skeletal muscle by reducing the need for amino acid-derived gluconeogenesis.
3. Decreased body magnesium, with or without hypomagnesemia.

Changes in Body Systems and Organs

In the gastrointestinal tract it is common that villi are flattened and atrophic, and the amount of almost all digestive enzymes is reduced. This condition leads to poor digestion and absorption and may contribute to diarrhea. Severely malnourished children are prone to diarrhea because of these alterations and possibly also because of irregular intestinal motility and gastrointestinal bacterial overgrowth. There is an alteration in pancreatic function, with a decrease of lipase and consequent steatorrhea.

Myocardial changes produced by atrophy of the cardiac muscle may lead to reduced cardiac output and electrocardiographic changes. Cardiac work decreases as does the function.

Renal function may be abnormal. Renal plasma flow and glomerular filtration rates may be reduced as a consequence of the decreased cardiac output, but water clearence and the ability to concentrate and acidify urine appear unimpaired.

Effect of Malnutrition on Immune Responses

Deficiencies in the immune system explain the high susceptibility of severely malnourished patients to infections and a greater predisposition to severe complications of otherwise less important infectious diseases. There is a marked impairment of both humoral and cell-mediated immune functions.

Severe thymic atrophy, tonsils, and other lymph tissues smaller in size than normal account for a T-cell deficiency that is particularly common in undernourished children. A depression of T-helper cells and possibly an increase in T-suppressor cell can also occur. There is decreased killer-cell activity because of reduced leukocytosis, lowered phagocytic activity of neutrophils, poor antibody formation, and decreased production of interferon and complement.

Reduced levels of secretory immunoglobulin A in secretions could be responsible for the compromised resistance to organisms that cause respiratory or

gastrointestinal infections. Atrophy of gut-associated lymphoid tissue contributes to diarrheal diseases.

Concomitant infections may lead to an increased concentration of serum immunoglobulins.

Although marasmus is a macronutrient deficiency due to insufficient food intake, and energy deficiency is more important and frequent than protein deficiency, it is usually complicated by multiple simultaneous nutrient deficiencies (vitamin and minerals) that also contribute to the effects and all the features can also be worsened by infection, thus establishing a vicious circle.

Treatment

Treatment strategies can be divided into three stages;

1. Resolving serious complications
2. Nutrient repletion
3. Ensuring nutritional rehabilitation

Resolving serious complications

Nutritional rehabilitation can be delayed until life-threatening conditions are solved. The most frequent conditions are described below.

Dehydration and electrolyte disturbance These children need careful attention and can usually be treated by appropriate oral fluids rather than by intravenous therapy. Rehydration can be achieved using the WHO standard oral rehydration solution (ORS) by mouth or by nasogastric tube. Intravenous fluids must be used in severe dehydration with hypovolemia, impending shock, frequent vomiting, and persistent abdominal distention. Careful attention must be displayed because cardiac failure may develop during or after administration of intravenous fluids. These may be the result of impaired cardiac function, sudden expansion of the intravascular fluid volume, severe anemia, or impaired membrane function.

Infections Infections are so common in severely malnourished children that antibiotics are often routinely recommended when an infection is suspected. Appropriate antibiotic therapy must be started, even before obtaining the results of microbiologic cultures. A broad-spectrum antibiotic or a combination, such as ampicillin and gentamicin, is usually given intravenously.

Treatment for intestinal parasites is rarely urgent and can be deferred until nutritional rehabilitation is under way.

Severely malnourished children not infrequently have tuberculosis or AIDS and should be examined. If the disease is found to be present, specific treatment is needed. In areas where malaria is prevalent, an antimalarial is desirable.

Hypothermia/hypoglycemia The seriously malnourished child has difficulty in maintaining normal body temperature. This may be due to impaired thermoregulatory mechanisms, reduced fuel substrate availability, or severe infection.

Untreated hypothermia is a common cause of death in malnourished children. If the child's temperature is below 36 °C, efforts must be made to warm the child. The child's temperature should be checked frequently as the patient may rapidly become hyperthermic.

Body temperature usually rises in the hypothermic patient shortly after the introduction of glucose-containing diets.

Asymptomatic hypoglycemia can be prevented or treated by the frequent feeding of small volumes of glucose-containing diets or solutions. Severe symptomatic hypoglycemia must be treated intravenously.

Nutrient Repletion

As recovery begins, restoration of nutritional status rapidly and safely as possible becomes the principal consideration. Children with severe nutritional marasmus may consume very high amounts of energy, and weight gain may be quite rapid.

It is best to begin with a liquid formula fed orally or by nasogastric tube, divided equally into 5–12 feedings per day, depending on the patient's age and general condition. If the child is still being breast-fed, breast-feeding should continue. Initially a mixture providing 120 kcal and 3 g protein per kg body weight should be provided.

In many hospitals in developing countries an aqueous mixture containing dried skimmed milk, vegetable oil, casein, and sugar is used. A marasmic child during the early part of recovery may be capable of consuming and utilizing 150–200 kcal and 4–5 g of protein per kilogram of body weight per day. As the appetite improves and no signs of intolerance or complications are evident, the energy intake may be increased to 250 kcal kg^{-1} body weight.

The various regimens must provide a diet that meets daily maintenance requirements, compensates for continued losses, and meets the requirements for growth. Adequate amounts of vitamins and minerals must be assured. Some evidence now suggests that zinc supplementation may be beneficial. Extra vitamin A and iron is often advisable since deficiency can develop as a result of increased demands when adequate protein and energy feeding begins.

In developed countries hydrolyzed milk formulas are preferred because of their digestibility and better absorption.

Ensuring Nutritional Rehabilitation

In all cases it is particularly important that the patient is not discharged from treatment until follow-up is assured and adequate feeding is continued. A mixed diet should gradually be introduced, aimed at providing the energy, protein, minerals, and vitamins needed by the child. Emotional and physical stimulation must be provided.

As a general guideline, vigorous dietary therapy should continue for 15 days after the marasmic patientq reaches an adequate weight-for-height and his or her clinical and overall performance are good. If he or she continues growing at a normal rate and has no functional impairments, treatment can be terminated after 1 month of adequate dietary intake. Specific treatment of other nutritional problems (iron deficiency) must sometimes be prolonged.

Instruction about the causes of PEM and nutrition education about the rational and nutritious use of foods, personal and environmental hygiene, appropriate immunizations, and early treatment of diarrhea and other diseases are vital for the person who will be responsible for the child.

Prevention

It is now understood that good nutrition lies in adequate levels of three factors:

1. Household food security
2. Appropriate care and feeding practices
3. Adequate health care

The prevention of PEM poses a huge challenge. Various strategies include appropriate nutrition policies with improving food security and care practices to insure good nutrition. These measures should be accompanied by promoting healthy lifestyles, suitable immunizations, growth monitoring, and early treatment for common diseases as well as prompt rehydration therapy for diarrhea and regular deworming. Severe malnutrition is more often associated with deficient food intake than with deficient protein intake. Cereal-based diets meet energy needs and also meet protein needs. Its necessary to pay attention to increasing total food intake, to protect and promote breast-feeding and sound weaning, and to encourage higher consumption of oil and fats that reduce bulk and increase the energy density of foods.

Severe malnutrition may manifest as a health problem but health professionals alone cannot solve it. This problem requires the contribution of agricultural professionals, educators, and professionals in economics, social development, and politics.

See also: **Kwashiorkor**; **Malnutrition**: The Problem of Malnutrition; Malnutrition in Developed Countries; **Nutritional Assessment**: Importance of Measuring Nutritional Status; Anthropometry and Clinical Examination; Biochemical Tests for Vitamins and Minerals; Functional Tests

Further Reading

Allen LH (1994) Nutritional influences on linear growth: a general review. *European Journal of Clinical Nutrition* 48: S75–S89.

De Onis M and Blössner M (1997) *WHO Global Database on Child Growth and Malnutrition*. Geneva: WHO.

Gracey M (1996) Diarrhea and malnutrition: a challenge for pediatricians. *Journal of Pediatric Gastroenterology and Nutrition* 22: 17–37.

Grunfeld C and Feingold KR (1992) Metabolic disturbances and wasting in the acquired immunodeficiency syndrome. *New England Journal of Medicine* 327: 329–337.

Keusch GT (1993) Malnutrition and immune response. In: Suskind RM and Lewinter-Suskind L (eds) *Textbook of Pediatric Nutrition*, 2th edn, pp. 161–171. New York: Raven Press.

Latham MC (1990) Protein-energy Malnutrition. In: Brown ML (ed.) *Present Knowledge in Nutrition*, 6th edn, pp. 39–46. Washington, DC: International Life Sciences Institute (ILSI) Nutrition Foundation.

Latham MC (ed.) (1997) *Food and Nutrition Series No. 29. Human Nutrition in the Developing World*, pp. 127–145. Rome: FAO.

Rutstein SO (2000) Factors associated with trends in infant and child mortality in developing countries during 1990s. *WHO Bulletin* 78: 1256–1270.

Schofield C and Ashworth A (1996) Why have mortality rates for severe malnutrition remained so high? *WHO Bulletin* 74: 223–229.

Shetty PS (1999) Adaptation to low energy intakes: the responses and limits to low intakes in infants, children and adults. *European Journal Clinical of Nutrition* 53 (suppl): 14–33.

Tenore A and Vargas A (1993) Endocrine changes in malnutrition. In: Suskind RM and Lewinter-Suskind L (eds) *Textbook of Pediatric Nutrition*, 2nd edn, pp. 151–161. New York: Raven Press.

Torun B and Chew F (1994) Protein-energy malnutrition. In: Shils ME, Olson JA and Shike M (eds) *Modern Nutrition in Health and Disease*, 8th edn, pp. 950–976. New York: Lea & Febiger.

UNICEF (2000) New York: *The State of the World's Children*. Oxford University Press.

WHO (1999) *Management of Severe Malnutrition: A Manual for Physicians and other Senior Health Workers*. Geneva: WHO.

World Health Organization Sub-Committee on Nutrition (ACC/SCN) (2000) *Fourth Report on the World Nutrition Situation. Nutrition Throughout the Life Cycle*. Geneva: WHO.

MARGARINE

Contents

Types and Properties

M Vaisey-Genser, The University of Manitoba, Winnipeg, Manitoba, Canada

Background

Margarine was invented in France by Hippolyte Mèges-Mouries in response to Napoleon III's call for a cheap alternative to butter for French workers and for his armies in the Franco-Prussian war. The first margarine, consisting of beef tallow churned with milk, was patented in 1869. Dutch entrepreneurs at Jurgens & Co., an established butter trader, purchased the margarine patent and popularized its use. Margarine production was limited by the availability of beef tallow until 1902 when Wilhelm Normann in Germany patented a process to harden oils by hydrogenation. This greatly expanded the market opportunities for vegetable oils and the availability of margarine. (*See* **Vegetable Oils**: Oil Production and Processing.)

By its centennial in 1969, margarine production in parts of Europe and in the USA began to rival that of butter. By 2000, margarine had become the table spread of choice for many people for reasons of either health or economics. Where this is the case, it also serves as a versatile fat in the home and in the food-service industry for preparing pan-fried foods, sauces, and bakery goods. This review considers the patterns of margarine consumption, the types of margarine available at the beginning of the twenty-first century, the physical structure of margarine, and its functionality, that is, its physical and sensory properties and their stability.

Margarine Consumption Patterns

The rate of margarine adoption has varied among countries, reflecting the relative strength of consumer groups and the oilseed industry in overcoming restrictive legislation achieved by dairy lobbies. Margarine was welcomed throughout Europe from its earliest introduction. However, the reverse was true in North America. In Canada, dairy farmers succeeded in imposing, in 1886, a countrywide ban on the manufacture, sale, and importation of margarine. Its prohibition endured until 1948, when the responsibility for margarine was transferred to each province. However, a federal excise tax was imposed on manufacturers until 1967. It took until the late 1990s for margarine to be sold legally in all 10 provinces, and even then, several imposed a color restriction on margarine to distinguish it from butter. In 1950, the USA abolished federal restrictive margarine taxation opening the door for liberal marketing. Nevertheless, some states resisted the legal sale of margarine for several years.

Contrary to Napoleon III's original expectation, the tablespread competition between margarine and butter has not centered solely on price. Some margarines have the advantage of being easy to spread, even at 4 °C, fresh from the refrigerator. This performance feature, coupled with a perceived nutritional advantage for vegetable oil-based spreads, led to early increases in margarine acceptance. In the USA, per-capita consumption of margarine surpassed that of butter in 1965 and more than doubled it 10 years later, a dominance which has been sustained into the last years of the twentieth century (Table 1). In both Canada and the UK, the shift from butter to margarine has been slower. Consumer research in Finland reported equal use of the two spreads in 1992 for a total of 15 kg per person per year. In France, butter apparently continues to be dominant.

In North America, there has been a gradual decrease in total table spread consumption over the last four decades of the twentieth century. In Canada, for example, the sum of margarine plus butter disappearance shown in Table 1 dropped from 13 kg per capita in 1955 to less than 8 kg per capita in 1995. In the USA, a decrease was also evident; total table spread disappearance dropped from almost 8 kg per capita in 1965 to only 5.8 kg per capita in 1995. Total table spread use in the UK also moderated between

Table 1 Examples of changes over time in margarine and butter consumption (kilograms per capita per annum, retail weight)

Year	USA[a]		Canada[b]		UK[c]	
	Margarine	Butter	Margarine	Butter	Margarine	Butter
1955			3.67	9.34		8.57
1965	4.75	3.12	3.95	8.44		
1975	5.01	2.14	5.18	5.18	7.10	8.92
1985	4.89	2.21	6.33	3.99	6.60	4.64
1995	4.17	2.04	5.17	2.76		
1997	3.90	1.99	4.11	2.62		

[a]Adjusted from *Oil Crops Situation and Outlook Yearbook, 1966–1998*. US Department of Agriculture, Economic Research Service.
[b]Adapted from *Food Consumption in Canada Part II, 32-230-XB, Annual*, Appendix A, p. 46. Statistics Canada 1998.
[c]Adapted from *Agricultural Outlook*. Paris: Organization for Economic Cooperation and Development. Aggregation of these data was subsequently discontinued.

1975 and 1985. Reductions in table spread consumption appear to be a positive response to nutrition educators' emphasis on the importance of lowering dietary fat to 35% of total energy intake to reduce the risk of cardiovascular disease. Health-conscious seniors and maturing baby boomers are paying attention, at least in their consumption of table spreads.

Market Form: Classification

The worldwide standard for margarine has been set by the Codex Alimentarius Commission for products containing a minimum of 80% fat and a maximum of 16% water (No. 32-1981, Rev.1-1989). It defines margarine as a 'food in the form of a plastic or fluid emulsion, which is mainly composed of the type water/oil, produced principally from edible fats and oils, which are not mainly derived from milk.' The standard lists permitted additives, which include vitamins, flavorings, colorants, emulsifiers, and preservatives.

By Firmness

Margarine table spreads are marketed in forms which are broadly classed as either hard or soft. The distinguishing feature is the degree of fluidity at the time of packing. Hard margarines are firm enough to be molded into a stick, print, or brick form. Soft margarines, which contain more oil than hardened fat, are too fluid to hold their shape and so require packaging in plastic or coated paperboard tubs. Stick-type margarines vary widely in their degree of hardness as the liquid oil in the formulation may range from as low as 5–10% to an upper limit of 60–65%. In contrast, the liquid oil in soft products may range from 60–65% to 80–85%. To enhance the spreadability, either type may be whipped with nitrogen or air, usually to a 33% volume increase. Liquid margarine was introduced in Europe and the USA in the 1960s. Its use continues to be modest.

In 1980, stick forms accounted for about two-thirds of the nondairy table spread market in the USA and continue to be popular. However, in Canada, where brick margarines have typically been firmer than those in the USA, the soft tub types had captured about 75% of market sales by 1995. With continuing emphasis on decreasing the intake of fat, particularly saturated fat, it is likely that soft margarines, which 'go farther' on spreading, and their low-fat counterparts will continue to retain a substantial proportion of the consumer market for table spreads in developed nations.

By Fat Combinations

Blends of margarine with butter, known in some areas as butterines, are popular in a number of countries within the EC and Scandinavia. For example, Sweden introduced the butter spread Bregott in 1969. The acceptance of this product, in which 20% of the milk fat was replaced with liquid oil, was credited with successfully reversing a decline in Swedish dairy sales. However, in some parts of the world, butter–margarine blends are still viewed as adulteration of dairy products. In Canada, for example, by 1999, blends were permitted in only five of 10 provinces, where either 50:50 or 20:80 blends of butter to margarine, respectively, could be marketed. Limiting the replacement of butter by margarine to only 20%, as in Sweden's Bregott may be the better choice as far as flavor is concerned. Sensory tests carried out in 1982 and summarized in Table 2 showed that both margarine and a blend of 80% margarine with 20% butter had significantly less buttery flavor than the butter prototype. However 'half-and-half' mixtures could be mistaken for either margarine or butter.

By Fat Content

'Reduced fat' margarines are those with less than 80% fat. They have been introduced in several parts

Table 2 Buttery flavor in margarine, butter, and blends[a]

Table spread (brick form)	Mean relative intensity of buttery flavor[b,c]
Butter	1.21[d]
Blend, 50 butter:50 margarine	0.80[de]
Blend, 20 butter:80 margarine	−0.29[e]
Margarine	−1.00[e]

[a]Adapted from Fyfe B and Vaisey-Genser M (1983) Salty and buttery flavors from tablespreads. *Canadian Institute of Food Science and Technology Journal* 16: 206–211.
[b]Mean score of 18 judgments/spread (six judges × three pairs).
[c,d,e]Values with the same letter are not significantly different ($P < 0.05$).

of North America and Europe, with the goal of decreasing consumers' total fat intake to lower the risk of heart disease. The amount of fat reduction varies. Where the fat content is approximately half of the prototype, i.e., ~40%, the spread may be referred to as minarine, halvarine, calorie-reduced (Canada), 'lite' or light margarine (USA), Mini-Light (Poland).

Structure

Margarine is a water-in-oil emulsion, in which the oil phase consists of both liquid oil and crystalline fat at room temperature. The semisolid structure is achieved by a three-dimensional, continuous, sheet-like matrix of fat crystals or crystal aggregates which entraps tiny water droplets suspended in oil. Structural integrity is due to irreversible primary chemical bonds and to reversible secondary bonds formed through weak London–van der Waals forces. (*See* **Colloids and Emulsions**.)

The number and size of the fat crystals in a margarine vary with the chemical composition of the oil source(s) and with its processing. Formulations are controlled to encourage the formation of many small β^1 crystals, which are each about 1 μm in length, and to prevent their polymorphic transition, that is, a shift to fewer but larger and more stable β crystals, which are 20–30 μm in length. At a high solids content, a network of β crystals will result in a brittle, hard margarine, while at a low solids content, the large crystals are less able to form a continuous matrix, and the product may become oily. Ideally, the solid crystal particles are so small that the force of gravity is exceeded by the adhesion of the particles, and the spaces between them are small enough to preclude the seepage of the liquid phase. Such a structure confers upon margarine the desirable property of plasticity.

Scanning electron microscopy has shown that the water droplets in a margarine emulsion may be as small as 1 μm in diameter, so that there may be as many as $5–10 \times 10^9$ per milliliter. Emulsifiers such as monoacylglycerols, diacylglycerols, lecithin, and milk proteins are located at the water–oil interface. They assist in dispersing the water as droplets in the oil during the formation of the emulsion and in keeping it dispersed due to their dual, but unbalanced, affinity for both phases. A dominance of nonpolar groups on the emulsifier succeeds in lowering the interfacial tension among oil molecules more than is effected by its polar groups on water, so that oil becomes the continuous phase. The presence of both lipophilic and hydrophilic groups in the emulsifier at the water/oil interface is necessary for emulsion stability. (*See* **Emulsifiers**: Organic Emulsifiers.)

Physical and Sensory Properties

Appearance

Color and gloss are the main determinants of margarine appearance. Both properties may be monitored by instruments such as the Hunter colorimeter or the Lovibond tintometer. β-Carotene, annatto, and tumeric are additives permitted in Canada to match the color intensity and hue of margarine to that of butter, a practice which was often banned in earlier years. In fact, punitive color requirements were imposed on margarine in several Canadian provinces for many years and in one, Ontario, until 1995.

Gloss or sheen, as opposed to a dull, matte surface, is a function of light reflectance from subsurface layers of oil droplets, which is influenced by the tightness of the emulsion. Excessive glossiness develops if the fat crystal network is not sufficiently fine or firm to enmesh the liquid oil. Soft margarines tend to be glossier than the hard type, due to the higher content of liquid oil. In the same context, Finnish scientists have reported that sensory panels perceived significantly less glossiness in reduced fat spreads than in those spreads with normal fat levels. (*See* **Sensory Evaluation**: Sensory Characteristics of Human Foods; Appearance.)

Texture

Spreadability and rate of mouth melt are two textural properties of table spreads which depend upon the relative amount of solid fat present and its crystalline structure. While both textural characteristics reflect a collective response to temperature and work, their criteria for success are somewhat in opposition. Good spreadability requires that the product retain its plasticity, as it is molded with modest pressure over temperatures ranging from refrigerator to room (4–22 °C). Desirable mouth melt, however, requires rapid melting at mouth temperature (35–37 °C) for prompt flavor release and clearance.

The traditional quality control tool for estimating spreadability has been the solid fat index (SFI) using dilatometry to measure volume changes as solid margarine converts to liquid when equilibrated at three temperatures: 10, 21, and 33.5 °C. Figure 1 illustrates the SFIs for butter and three different margarines. Sensory testing has shown that SFI values between 10 and 20, at the temperature of use, predict good spreadability, while values of 35 and 7 are, respectively, too hard or too soft/oily to spread. Of the SFI curves in Figure 1, the flatter patterns for the soft stick and soft tub margarines illustrate desirable spreadability, even at a temperature of 10 °C, whereas butter

is not spreadable until it reaches 20 °C. More recently, nuclear magnetic resonance (NMR), which depends on differences in the magnetic environment of protons in the solid and liquid phases, is being used to assess the spreadability of table fats. The solid fat content (SFC) values from NMR evaluations are considered more reliable than SFIs, particularly for samples at high solids content at lower temperatures. (*See* **Sensory Evaluation**: Texture; **Spectroscopy**: Nuclear Magnetic Resonance.)

Rapid mouth melt of table spreads provides a cooling sensation due to absorption of the heat of crystallization in melting and is accompanied by prompt flavor release. Figure 1 shows that butter has a steeper SFI slope and therefore more rapid mouth melt than any of the margarines. This provides it with an advantage which is difficult for the processor to duplicate without sacrificing spreadability. Sensory panel measurements of mouth melt time for 1 ml samples of Canadian table spreads on the 1982 market confirmed that butter melted significantly faster than brick margarines whether the starting temperature was 4 °C (18 s vs. 22 s, respectively) or 21 °C (11 s vs. 16 s, respectively). To our knowledge, the extent of the differences in mouth cooling sensations among table spreads as they melt has yet to be documented.

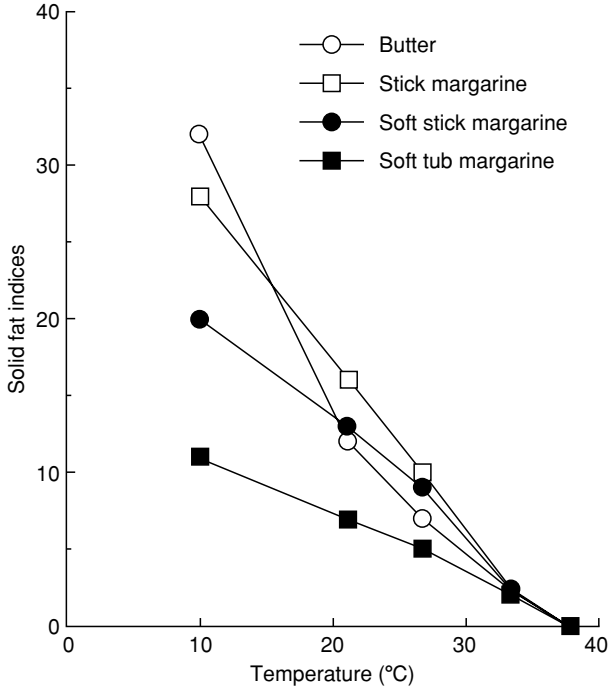

Figure 1 Effect of temperature on table spread percentage solids. Drawn from data cited by Chrysam M (1985) Table spreads and shortenings. In: Applewhite TH (ed.) *Bailey's Industrial Oil and Fat Products*, vol. 3, pp. 41–127. New York: John Wiley.

Flavor

Although North Americans have used as much or more margarine than butter since the mid-1970s, a 1986 survey of 4000 US consumers by the dairy industry found that the majority still preferred butter. Europeans echo the same sentiment, even though North Americans usually use salted and sweet cream table spreads, while Europeans favor cultured butter and unsalted spreads. For example, sensory studies in Poland and Finland reported in 1989 that margarines often lacked buttery flavor. In the absence of reports to the contrary, flavor apparently remains a challenge for the margarine processor.

Margarine flavor is a combined response to odorants soluble in the fat phase and tastants dissolved in the water droplets. Permitted flavor additives include compounds which have been identified as important in butter aroma such as diacetyl (1–4 p.p.m.), dimethyl sulfide, lactones, ethyl esters of short-chain fatty acids, ketones, and aldehydes. When ordinary salt is added to table spreads, it is the main tastant, although the lactose from milk solids will add a modest amount of sweetness. In a number of spreads, salt levels have been reduced from 2.0–3.0% by weight to 1.0–1.5%, in the interests of lowering the intake of dietary sodium. Sensory tests on a range of salt concentrations in butter have shown that reductions of this order of magnitude have little effect on perceived saltiness intensity. Sensory tests comparing the perception of salt in water and in butter show in Figure 2 that the high fat content of a water-in-oil

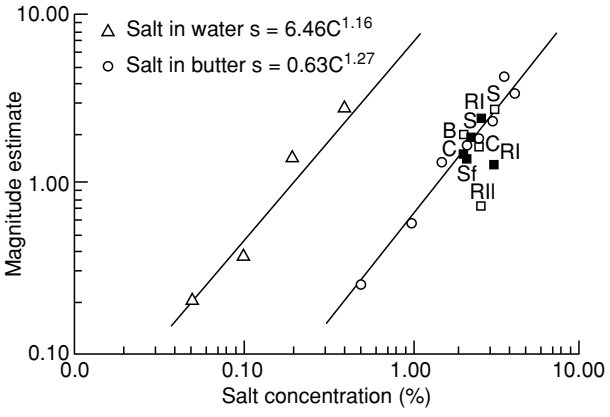

Figure 2 Psychophysical relationships of salty intensity with salt concentration in water and in butter; RI, RII, C, S, and Sf are various retail margarines, and B is butter. From Vaisey-Genser M and Vane BK (1995) Sensory evaluation of margarine. In: Warner K and Eskin NAM (eds) *Methods to Assess Quality and Stability of Oils and Fat-containing Foods*, pp. 76–106. Champaign, IL: AOCS Press, with permission. Originally published as Fyfe B and Vaisey-Genser M (1983) Salty and battery flavours from tablespreads. *Canadian Institute of Food Science and Technology Journal* 16: 206–211.

emulsion like a table spread interferes with the perception of a water-soluble tastant like salt as much as 10-fold. However, altering the salt content may affect the overall flavor balance of margarine because of influences on the partition coefficients of volatile components. (*See* **Sensory Evaluation**: Taste.)

The flavor advantage of rapid mouth melt has been demonstrated with simulated margarines consisting of 0.01% diacetyl and other dairy flavors combined with liquid soybean oil melted with either palm stearin or palm olein and recrystallized at 0 °C. Figure 3 shows the flavor intensity patterns described by trained panelists in time-intensity studies of 2.5-ml samples of these two products. Total flavor intensity peaked after 12 s in the mouth in both products, but flavor was consistently more intense in the more rapidly melting palm olein product. These sensory findings were confirmed by mass-spectrophotometric headspace measurement of diacetyl using the same simulated margarine products agitated at 35 °C with glass beads and artificial saliva.

Stability

Standard 80% fat margarine normally has a shelf life of 6–12 months if the product is refrigerated during marketing, distribution, and home storage. This period is shorter for low-fat, high-moisture products and for those which are salt-free.

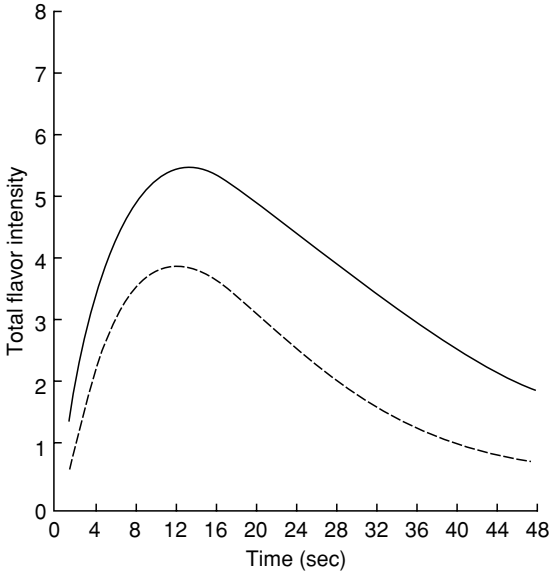

Figure 3 Effect of margarine melting point on the time and intensity of sensory flavor release; — palm olein; --- palm stearin; 0, none; 5, moderate; 10, a great deal. From Lee WE (1986) A suggested instrumental technique for studying dynamic flavor release from food products. *Journal of Food Science* 31: 249–250, with permission.

Microbiological spoilage

The growth of molds and yeasts is limited by controlling the water droplet size and by the addition of salt and other preservatives. Where droplets are smaller than 10 μm and the amount of milk solids is limited, it can be calculated that, per water droplet, there is not enough nutrient to sustain microbial growth. Accordingly, emulsion stability fosters keeping quality. Salt levels of 1.0–1.5% in margarine are considered necessary to avoid microbial hazard. At the level of 1.0% by formula weight, there would be a salt concentration of 6.25% in the serum, assuming 16% water in the margarine formula. Preservatives such as sorbic acid and benzoic acid or their respective sodium or calcium salts add protection by lowering the pH. A pH of 5–6 is recommended for salted margarine and a more acidic pH of 4–5 for unsalted products. (*See* **Preservation of Food; Spoilage**: Bacterial Spoilage.)

Graininess

Margarines prone to poymorphism may exhibit textural instability during storage at refrigerator temperature (4 °C), which is perceived as graininess and is due to crystal growth. This phenomenon is exaggerated by fluctuating storage temperatures. In Figure 4, sensory estimates of the presence or absence of graininess in a margarine during 6 months' refrigerated storage have been plotted against measurements of crystal size from photographs taken under polarized light. These data show a sensory detection threshold of crystal size as 22 μm. It remains likely that the sensation of graininess is a function of both the size

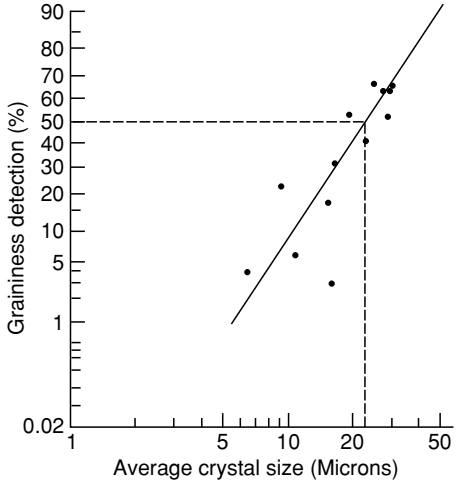

Figure 4 Relationship between fat crystal size in margarine and the percentage of panelists detecting graininess. Adapted from Vaisey-Genser M, Vane BK and Johnson S (1989) Graininess, Crystal size and firmness of stored canola margarines. *Journal of Texture Studies* 20: 347–361.

and the number of crystals. In cases of exaggerated crystal growth, there is the hazard of increased susceptibility to microbial spoilage. The development of extremely coarse fat crystals may squeeze liquid oil from the product and allow partial coalescence of the aqueous phase. Free moisture on the surface may result in mold growth. (*See* **Spoilage**: Molds in Spoilage.)

Flavor

The stability of margarine flavor is enhanced by refrigerated storage to retard autoxidation of the oil and by lightproof packaging to avoid photooxidation. Changes by either mechanism produce off-flavors in oils which are variously described as beany, painty, or fishy. The trend to greater use of soft margarines with higher proportions of liquid oil increases the risk of autoxidation, particularly when the oils used are high in unsaturated fatty acids. The addition of antioxidants is permitted to reduce this hazard. Photochemical oxidation of liquid oils has been shown to be greater at short wavelengths, with a maximum below 455 nm. As fluorescent light in supermarkets may transmit wavelengths between 350 and 750 nm, packaging impervious to low-wavelength light is highly recommended for maximum margarine flavor protection. (*See* **Antioxidants**: Natural Antioxidants, Synthetic Antioxidants.)

See also: **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; **Colloids and Emulsions**; **Emulsifiers**: Organic Emulsifiers; **Preservation of Food**; **Sensory Evaluation**: Sensory Characteristics of Human Foods; Appearance; Texture; Taste; **Spectroscopy**: Nuclear Magnetic Resonance; **Spoilage**: Bacterial Spoilage; Molds in Spoilage; **Vegetable Oils**: Oil Production and Processing

Further Reading

Chrysam M (1985) Table spreads and shortenings. In: Applewhite TH (ed.) *Bailey's Industrial Oil and Fat Products*, vol. 3, pp. 41–127. New York: John Wiley.

deMan JM, Dobbs JE and Sherman P (1979) In: Sherman P (ed.) *Food Texture and Rheology*, pp. 43–54. London: Academic Press.

Heick WH (1991) *A Propensity to Protect: Butter, Margarine and the Rise of Urban Culture in Canada*, pp. 1–163. Waterloo, Canada: Wilfred Laurier Press.

Hellemann U, Tuorila H, Lampi A-M and Matuszewska I (1990) Hedonic responses and attitudes related to fats used as spreads on bread. *Food Quality and Preference* 2: 29–38.

Juriaanse AC and Heertje I (1988) Microstructure of shortenings, margarine and butter – a review. *Food Microstructure* 7: 181–188.

Lee WE (1986) A suggested instrumental technique for studying dynamic flavor release from food products. *Journal of Food Science* 31: 249–250.

Mela DJ and Marshall RJ (1992) Sensory properties and perceptions of fats. In: Mela DJ (ed.) *Dietary Fats: Determinants of Preference, Selection and Consumption*, pp. 43–47. London: Elsevier.

Rohm H and Raaber S (1991) Hedonic spreadability optima of selected edible fats. *Journal of Sensory Studies* 6: 81–88.

Vaisey-Genser M, Vane BK and Johnson S (1989) Graininess, crystal size and firmness of stored canola margarines. *Journal of Texture Studies* 20: 347–361.

Vaisey-Genser M and Vane BK (1995) Sensory evaluation of margarine. In: Warner K and Eskin NAM (eds) *Methods to Assess Quality and Stability of Oils and Fat-containing Foods*, pp. 76–106. Champaign, IL: AOCS Press.

van Alphen J (1969) Hippolyte Mège-Mouries. In: van Stuyvenberg JH (ed.) *Margarine – An Economic, Social and Scientific History 1869–1969*, pp. 6–8. Liverpool, UK: Liverpool University Press.

Methods of Manufacture

R A Carr, Carr & Associates, Burlington, Ontario, Canada
M Vaisey-Genser, The University of Manitoba, Manitoba, Winnipeg, Canada

Background

Consumers have grown to appreciate the margarine products that spread easily at refrigerator temperatures and are enjoyed because of their flavor and texture. To understand the dynamics of the manufacture of these margarines, it is essential to appreciate the relationships between the processing effects, the specific physical properties of the products, and the composition of their oil blends and other components. A major factor in the formulation and control of margarine oils and margarine is the development of desirable solids-to-liquid ratios, achieved by hydrogenation and blending. The types of oils selected for margarine oil formulations are also important in obtaining a stable, uniform crystal structure from specific conditions for processing the oils, as well as the finished margarine products. As a result, margarines as semisolid plastic products in soft tub and wrapped stick forms have been designed to meet consumer demands for health and dietary benefits, as well as performance and convenience.

Physical Properties

A margarine structure is a network of small fat crystals, which acts as a matrix to contain trapped oil and water globules. The degree of finished product spreadability depends considerably upon the proportion and properties of the liquid oil and crystalline fat in the margarine oil component. Butter and stick margarines, for example, are brittle when cold from the refrigerator. They gradually become more spreadable as they approach room temperature, because some of the fat crystals melt as the margarine becomes warmer. Consumers desiring spreadability from the refrigerator may select soft tub margarines, which are designed with lower solids and higher liquid oil contents. Various combinations of hydrogenation, fractionation, interesterification, and blending of diverse oil types and solids contents allow the development of margarines with differing consistencies and properties. (*See* **Colloids and Emulsions**.)

Margarine consumers fall into two main classifications, each with specific quality requirements. The domestic cook requires a product for table use as well as baking and frying. The professional baker demands products that will give a consistent performance for cakes, icings, etc. For the home, a margarine must provide an enjoyable flavor that releases fully and quickly, must melt rapidly in the mouth leaving no gummy or gritty sensation, have sufficient body at room temperature to maintain its form, and must be easy to spread over a broad range of temperatures. Creamability, which is the ability to take up air when beaten, and plasticity are the key priorities required for satisfactory baking. For frying, the fat component must remain stable at high temperatures and contain virtually no free fatty acids, which create a smoking problem during drying.

The physical properties of various margarines depend primarily upon (1) the melting point of the oil component triglycerides, (2) the total solids content present at any given temperature, (3) the distribution of these solid fats over a broad temperature range, and (4) the polymorphic modification of crystal habit of the fat composition. Their combined response to temperature and work applied by external forces is the most characteristic. Margarine is designed to meet flavor, plasticity, and creamability requirements by combining a properly designed margarine basestock with an aqueous phase blend so that the water globules are finely dispersed, but combined loosely enough for the emulsion to break easily upon melting. The aqueous phase generally consists of reconstituted milk powder, brine, and water. Vitamins, coloring, and flavoring agents are included, and emulsifiers, such as monoglycerides and lecithin, are added to the oil phase. (*See* **Triglycerides**: Structures and Properties.)

Formulation

To ensure the development of a proper margarine emulsion prior to the crystallization process, two distinct phases, aqueous and oil, must be prepared individually, prior to blending (**Figure 1**). The composition of the phases is designed for the three basic types of margarines: (1) stick or regular margarine, (2) soft or tub margarine, and (3) diet or calorie-reduced margarine.

Diet margarines contain half as much margarine oil as the stick or tub margarines. Because of their higher water content, they are unsatisfactory for cooking purposes, but are used for spreading on bread by consumers interested in their lower energy content.

Aqueous Phase

The major component in the margarine aqueous phase is either sweet skim milk or water plus reconstituted skim milk powder, or even water without milk if required. Care is taken that the milk is properly pasteurized by rapid heating to 75 °C with sufficient contact time. If desired, cultured milk, which has had some of the lactose content converted to lactic acid, may be used to provide a distinct flavor and acidity to the milk.

Salt, or brine, is then added to the aqueous phase to accentuate the flavor, act as a microbial inhibitor, and reduce splattering during pan frying. Minor components in the aqueous phase include citric acid, ethylenediaminetetraacetic acid (EDTA) and a water-soluble dairy flavor. Citric acid reduces the pH of the liquid phase to approximately 5.3, to enhance the performance of the microbial inhibitor. EDTA acts as a chelating agent by tying up any metal ions picked up either in the equipment or from other components added to the aqueous phase. (*See* **Acids**: Natural Acids and Acidulants.)

Oil Phase

The main component in the oil phase is a margarine oil basestock specifically developed through the refining process to produce a final margarine with the appropriate flavor, keeping quality and melting characteristics necessary to produce a specific margarine with distinctive characteristics demanded by the consumer. Oil basestocks originate from crude oils such as soyabean, palm, corn, cottonseed, sunflower, canola, and rapeseed. Initial processing through refining and bleaching operations removes impurities from the oil. Refined–bleached oils can be modified

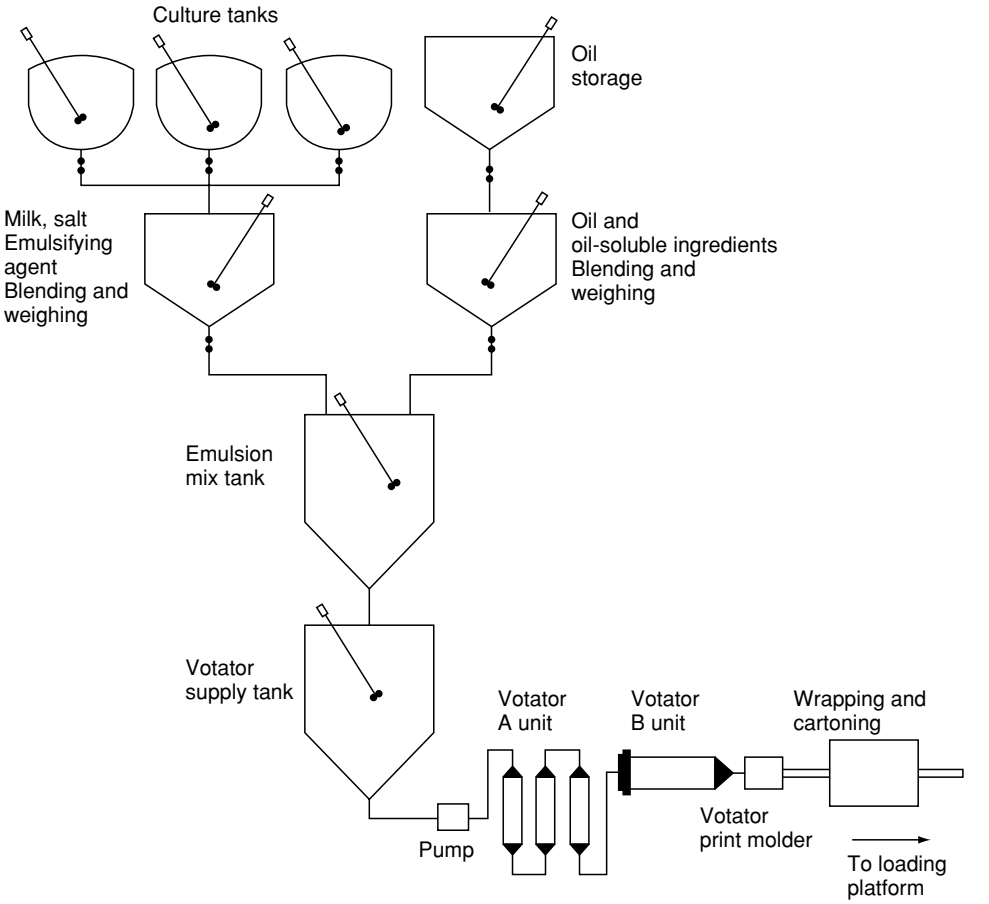

Figure 1 Simplified flow diagram for continuous margarine solidification. Courtesy of the Girdler Co. Reproduced from Margarine: Methods of Manufacture, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press, with permission.

by a hydrogenation process to alter their physical properties, such as melting characteristics and stability. Additional changes in the physical properties can be obtained by blending different oils and interesterification. (*See* **Ground Nut Oil**; **Palm Oil**; **Vegetable Oils**: Oil Production and Processing.)

Changes to the physical properties, or hardness, of a fat can be measured by melting points, dilatometry, and nuclear magnetic resonance (NMR). Dilatometry is based on the difference in specific volume for liquid and solid fats at a specific temperature. Dilatometric curves may be plotted for solid fat indices (SFIs) against the temperature of determination. NMR provides a different approach to the determination of solids in fats and their blends. The results are determined in absolute solids terms and are utilized to provide a relatively simple technique for determining SFI curves, which are somewhat comparable with the dilatometric method. (*See* **Spectroscopy**: Nuclear Magnetic Resonance.)

Most vegetable oils are mainly liquid at room temperature. Therefore, hydrogenation plays a key role in the preparation of basestocks with various SFI melting curves that can be blended for the preparation of margarine basestocks. Other techniques such as fractionation, interesterification, directed interesterification, and corandomization, are also available, but are normally used for specific situations related to oil availability and for products requiring a distinctive melting property, such as bakery margarines.

Liquid unsaturated oils have a relatively high number of double bonds, melt at low temperatures, are more unstable, and are likely to deteriorate over time in terms of odor and flavor. Treatment of the oil in hydrogenation convertors with a nickel catalyst at high temperature with violent agitation, in contact with a flow of tiny hydrogen bubbles, will insert hydrogen into some of the double bonds. Approximately 0.01% catalyst and 200 °C temperature are required for margarine oil basestocks. As the process proceeds, the melting point characteristics of the oil are increased. They are measured by a refractometer in the plant and confirmed precisely in the laboratory by dilatometric or NMR analyses. When the proper

end points are obtained, the oil in the convertor is cooled and then carefully filtered to remove all traces of catalyst.

Hydrogenation is the most flexible single process for producing margarine basestocks. Depending upon the operating conditions selected, the SFI curve can be steepened or flattened by adjustment of hydrogenation temperature, pressure, agitation, catalyst type, and catalyst concentration. For example, selective hydrogenation conditions using a high temperature and lower pressure are utilized to promote high *trans* fatty acid development with a minimum drop in iodine value, to produce steep solids curves. Recently, some nutritionists have expressed concern over the effect of *trans* fatty acids on health. If ultimately confirmed. Other techniques, such as enzymatic reactions, can be utilized to provide basestocks with steep melting curves and low *trans* fatty acid content.

Differing from shortenings, margarine oil blends are formulated to have a sufficient solids content at room temperature to perform as a stick margarine or tub margarine on the table, without slumping. At refrigerator temperature, the solid content should be kept at a minimum to ensure smooth, easy spreadability on bread when used directly from the refrigerator. In addition, margarines should melt very rapidly at body temperature (37 °C) to ensure a 'quick get away' in the mouth with minimum gumminess. This rapid melting characteristic is attained from the development of *trans* fatty acids during hydrogenation, which start to melt rapidly as the temperature is increased above about 25 °C. Opposite to the sharp melting curve required for a margarine, shortenings are formulated to produce flat melting curves, with very little change in solid content, from room temperature to temperatures well above body temperature. (*See* **Fatty Acids**: Properties.)

In addition to the solids content curve, a second important factor in margarine oil blends is their palmitic acid content ($C_{16:0}$), which has a definite effect on the crystal stability of the final margarine. Oil blends with insufficient palmitic acid content tend to revert from the β' crystal state to an undesirable β crystal state during storage of the packaged product. β' crystals produce both margarines and shortenings with a smooth uniform texture, because of the small size of the crystals, from 1 to 3 μm. However, β crystals are larger than 20 μm in size and produce obvious grainy texture and brittleness in margarines and shortenings. β crystal margarines are readily detected by the tongue as an undesirable gritty feel. Oils/fats that are β-tending include canola, cocoa butter, coconut, corn, lard, olive, palm kernel, peanut, safflower, sesame, soya bean and sunflower. β' Types are cotton seed, herring, milk fat, modified lard, palm and tallow. It is normal practice to blend approximately 5–15% of cotton seed or palm oil in margarine oil blends to minimize all possibilities of crystal conversion to the β form. Some margarine manufacturers use a larger number of different hardstock fractions also to minimize the possibilities of β crystal development. A crystal inhibitor such as sorbitan tristearate can add crystal stability insurance more effectively than diglycerides.

Final margarine oil blends are carefully formulated to produce the desired solids melting curve. For example, stick margarine solids contents are approximately 27% at 10 °C, 14.2% at 21.1 °C, and 2.5% at 33.3 °C. Soft or tub margarines, with improved spreadability characteristics, are approximately 13% (SFI) at 10 °C, 7% at 21.1 °C, and 2.3% at 33.3 °C. The stick margarine SFI profile provides sufficient solids at 10 °C to allow satisfactory parchment overwrapping during packaging, satisfactory spreadability at both room temperature and out of the refrigerator, plus a good 'get away' in the mouth. The margarine oil blend is then deodorized at about 250 °C under a vacuum of 6 mmHg for 40 min to remove flavor and odor components, prior to its use in the margarine oil phase.

The oil phase is made up separately from the aqueous phase. Sufficient oil is pumped into the blend tank to ensure that the final product will have an oil content of 80% in the final stick or tub margarine, or 40% for low-energy diet spreads. Approximately 0.2% lecithin is metered into the oil as an emulsifier and antispattering agent during pan frying. Food coloring and oil-soluble butter flavor (0.05% of each) are added, and vitamins A and D are included for nutritional purposes. Margarine flavor release can be affected by the 'tightness' of the final oil/aqueous emulsion. Tight emulsions reduce the flavor impact as compared with loose emulsions. Therefore, phase blending and crystallization practices need to be carefully controlled to ensure uniform flavor release characteristics. (*See* **Cholecalciferol**: Properties and Determination; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; Properties and Determinants of Synthetic Pigments; **Emulsifiers**: Organic Emulsifiers; Uses in Processed Foods; **Retinol**: Properties and Determination.)

Crystallization

Immediately prior to the packaging of margarine, the liquid and oil phases are mixed together in a 1:4 ratio, with gentle agitation, and maintained at about 40 °C. The temperature is selected to develop a stable emulsion and to prevent any precrystallization, which would occur at temperatures below 37 °C. A

proper emulsion will have the water phase droplets finely dispersed in the oil phase, but loosely enough for the emulsion to break easily on melting.

Margarine emulsions can be made by either a batch or continuous system. The batch system has been used for many years and comprises an agitated, temperature-controlled mixing tank to receive the water and oil phases. After the batch has been mixed to form a stable emulsion at the required temperature, the emulsion is then pumped to a scraped-surface heat exchanger for supercooling. The batch system is usually used for operations required to produce many different types of margarines with production runs of relatively short duration.

In lieu of a mixing tank, the continuous system may use a three-headed proportioning pump to meter and mix the oil phase continuously and the water phase simultaneously in the proper proportions into an agitated, temperature-controlled holding tank and then directly to the scraped-surface heat exchanger. The continuous system is designed for plants that run large quantities of similar types of margarines over extended time periods.

A scraped-surface heat exchanger is a unit that adds heat or removes heat from a substance. In the margarine industry, it is used as a closed chilling machine that induces partial crystallization of the fat in the margarine emulsion, all taking place in stainless steel, externally refrigerated cylinders through which the fat is pumped continuously. The cylinders are equipped with fast-revolving scraper blades that work the fat to achieve efficient, even heat transfer to the total mass in the heat exchange tubes. The process, if desired, can be utilized to add air or gas to a margarine, for the production of whipped margarines. In the industry, the process is often called 'votation,' because the original units were produced by the Votator Division of the Chemetron Corporation, Louisville, Kentucky.

The objective of votation is to develop and seed β' crystals throughout the margarine, to ensure their progression to the dominant crystal form in the finished product. If successful, the margarine will then have a smooth textural feel on the tongue, as well as contributing to a smooth-spreading performance on bread. For long-term uniform production, attention must be paid to the sharpness of votator blades and the smooth condition of the chilling cylinders. For margarine, the votator should rapidly reduce the incoming emulsion temperature from about 40 to 7 °C. The votator, also called an 'A unit,' is designed for direct-expansion refrigerants such as ammonia, freon, and propane. Advantage is taken of its high rate of heat transfer due to 'surface boiling.' A considerable proportion of the liquid vaporizes upon contact with the heat-transfer tube, and the velocity of the gas carries a relatively high percentage of liquid ammonia back to the surge drop, thus assuring complete flooding of the heat-transfer surface at all times. As a result, a constant uniform chilling effect is achieved for margarine passing through the chilling cylinder.

The emulsion is worked through the chilling cylinder over a period of about 18 s. Supercooled product leaving the A unit is then pumped to the 'B unit' cylinder, which is considerably smaller in diameter and longer than the B unit used for shortenings. In addition, the margarine B unit is not agitated, so that solidification of the supercooled margarine emulsion takes place under almost static conditions. If desired for whipped margarine, air or preferably an inert gas such as nitrogen can be drawn into the emulsion at the suction side of the product pump in precise amounts regulated by flow meters as it is fed to the chilling unit.

Limiting the amount of work given to the product in the B unit (1) produces a product that is not too soft to be handled in automatic print forming and wrapping equipment, (2) prevents the aqueous phase from being dispersed in a too fine state of suspension, and (3) induces the growth of β' seed crystals from the supercooled mass. Agitation in the B unit would require a long 'dwell period' for the product to become firm enough for packaging. In addition, the resultant tight emulsion and incorrect crystal structure would delay melting of the product in the mouth, producing a waxy impression with the user. Tight emulsions also fail to yield the desirable milk and salt flavors before the product has been swallowed, and contribute to brittleness or hardness, which is reflected in the lack of spreadability at refrigerator temperatures. While the product is in the B unit, the temperature rises by approximately 5 °C, primarily due to the latent heat of crystallization. The supercooled mass solidifies as it is slowly forced through the B unit by the pressure of the feed pump.

After leaving the B unit, the finished stick margarine product is extruded in a rectangular form, shaped, and wrapped. For many print margarines, a combination of vegetable parchment paper and aluminum foil is often used to make a final print product in a unit such as the Benhil. Popular soft margarines are continuously packed in tubs usually made from polyvinylchloride. The filled product is then automatically packed into cases, sealed, and sent to a tempering room. Packaged margarines should be stored at 5–10 °C for 48 h before shipment from the warehouse. This will ensure that the desired β' crystals have sufficient time to multiply and reach a stable state.

Quality Assurance

During production, it is important to carry out an ongoing routine inspection of processing conditions and finished product. The former includes items such as the A unit outlet temperature, which must be frequently checked, adjusted, and recorded by the operator. Product evaluations are usually carried out by the quality control inspectors and laboratory technicians. (*See* **Quality Assurance and Quality Control.**)

Samples from the packing lines are taken every 15–30 min and checked for weight performance versus limits for each product type. Line samples are then checked for package quality, including: carton or container denting or crushing; dirt or oil spots; parchment wrap defects for stick margarine, such as poor folds of the parchment or foil; inadequate gluing of carton inflaps; improper placement of lable.

Packaging materials are then removed, and the product's external appearance is evaluated. Prints are sliced in two or three places for internal examination, and a cone can be cut out of the center of the tub products. The outer texture is checked for slack or overfill, sloshing, product adhering to the lid or wrap, blistered or grainy texture, discoloration, oil separation, dull sheen, etc. After cutting, the inner texture is checked for oil or water separation and streaked, grainy, vacillated, cheesy, or porous texture.

Margarines undergo sensory evaluation to detect any off-quality characteristics, such as lack of flavor, or oil off-flavors such as oxidized, rancid, faulty oil processing, sour, plastic, fruity, artificial, and aged characteristics. If required, quality-rating systems can be established with defect scores assigned by the importance of the defect for the product. For example, a 100-point scale could be used, with 10 points assigned to packaging and 30 points each for outer texture, inner texture, and flavour. The assigned defect points are then subtracted from 100. A grade below an arbitrary level such as 70 would indicate that the related product should be placed on hold for further evaluation and disposition. (*See* **Sensory Evaluation**: Sensory Characteristics of Human Foods.)

See also: **Acids**: Natural Acids and Acidulants; **Cholecalciferol**: Properties and Determination; **Colloids and Emulsions**; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; Properties and Determinants of Synthetic Pigments; **Emulsifiers**: Organic Emulsifiers; Uses in Processed Foods; **Fatty Acids**: Properties; **Quality Assurance and Quality Control**; **Retinol**: Properties and Determination; **Spectroscopy**: Nuclear Magnetic Resonance; **Triglycerides**: Structures and Properties; **Vegetable Oils**: Oil Production and Processing; **Sensory Evaluation**: Sensory Characteristics of Human Foods; **Ground Nut Oil**

Further Reading

Carr R (1986) Oilseeds, fat and oils. *National Research Council Training Program, Chatham, Ontario*, pp. 10–14.

Dritschel M (1970) How to evaluate margarine quality. *Food Engineering*, October: 90–93.

Krytz K (1989) Formulation and packaging margarines and shortenings. Annual Meeting of the American Oil Chemists' Society, Cincinnali, Ohio, May 3–6, 1989. *Journal of the American Oil Chemists Society* 66(4): 459–460.

Moore E (1974) Margarine and cooking fats. *Unilever Educational Booklet Series*, No. 4, pp. 7–14.

Ward J (1988) Processing canola oil products. *Journal of the American Oil Chemists Society* 65: 1731–1734.

Wiederman L (1978) Margarine and margarine oil, formulation and control. *Journal of the American Oil Chemists Society* 55(11): 823–829.

Composition and Analysis

M Vaisey-Genser, The University of Manitoba, Winnipeg, Manitoba, Canada

Introduction

Margarine is the earliest of the designed food staples. As such, it can be tailored to deliver nutritional and functional advantages in terms of fatty acid composition, the addition of essential micronutrients, spreadability, and performance as a bakery fat. Accurate methods are necessary to guide margarine design and to monitor the legitimacy of claims on margarine labels. This article will describe the required composition of margarine, provide examples of the fatty acid profiles of contemporary products, and comment on the methods for margarine analysis and quality control.

Margarine Composition

The margarine of each country is subject to local standards of identity and regulations, which normally complement the international standard of the Joint FAO/WHO Food Standards Programme as set by the Codex Alimentarius Commission. The existing standard for margarine as set in 1989 (CODEX STAN 32-1981, Rev. 1-1989) is currently under review. Ultimately, it is to be replaced by a general 'Standard for Fat Spreads and Blended Spreads,' a proposal being drafted to address reduced-fat as well as full-fat spreads. The goal is to assure food safety concurrent with the globalization of food trade. As of July 1999, the draft proposal was at Step 3 of what can be

an eight-step Codex approval process. In the interim, the 1989 standard for margarine prevails. The 'essential composition and quality factors' of margarine and its endorsed 'food additives' are set out in Sections 3 and 4, respectively, of the 1989 Codex Standard, as follows:

Essential Composition and Quality Factors

Raw materials (1) Edible fats and/or oils, or mixtures of these, whether or not they have been subjected to a process of modification; (2) water and/or milk and/or milk products.

Minimum fat content 80% m/m of the product.

Maximum water content 16% m/m of the product.

Additions The following substances may be added to margarine: (1) vitamins – vitamin A and its esters, vitamin D, vitamin E and its esters, other vitamins (maximum and minimum levels for vitamins A, D, and E and other vitamins should be laid down by national legislation in accordance with the needs of each individual country, including, where appropriate, the prohibition of the use of particular vitamins); (2) sodium chloride; (3) sugars; (4) suitable edible proteins. (*See* **Cholecalciferol**: Properties and Determination; **Retinol**: Properties and Determination; **Tocopherols**: Properties and Determination.)

Food Additives

Colors (1) β-carotene, 25 mg kg^{-1} (2) annatto extracts, 20 mg kg^{-1} calculated as total bixin or norbixin; (3) curcumin or turmeric, 5 mg kg^{-1} calculated as total curcumin; (4) β-apo-8′-carotenal, 25 mg kg^{-1}; (5) methyl and ethyl esters of β-apo-8′-carotenoic acid, 25 mg kg^{-1}. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments; Properties and Determinants of Synthetic Pigments.)

Flavors Natural flavors and their identical synthetic equivalents, except those that are known to represent a toxic hazard, and other synthetic flavors approved by the Codex Alimentarius Commission are permitted for the purpose of restoring natural flavors lost in processing or for the purpose of standardizing flavor, as long as the added flavor does not deceive or mislead the consumer by concealing damage or inferiority or by making the product appear to be of greater actual value. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Emulsifiers (1) Mono- and diacylglycerols of fatty acids (limited by good manufacturing practice (GMP)) (2) mono- and diacylglycerols of fatty acids esterified with the following fatty acids: acetic, acetyltartaric, citric, lactic, tartaric, and their sodium and calcium salts (10 g kg^{-1} maximum); (3) lecithins and components of commercial lecithin (limited by GMP); (4) polyglycerol esters of fatty acids (5 g kg^{-1} maximum); (5) 1,2 propylene glycol esters of fatty acids (20 g kg^{-1} maximum); (6) esters of fatty acids with polyalcohols other than glycerol: sorbitan monopalmitate, sorbitan monostearate, sorbitan tristearate (10 g kg^{-1} maximum); (7) sucrose esters of fatty acids (including sucroglycerols) (10 g kg^{-1} maximum). (*See* **Emulsifiers**: Organic Emulsifiers; Uses in Processed Foods.)

Preservatives Sorbic acid and its sodium, potassium, and calcium salts and benzoic acid and its sodium and potassium salts (1000 mg kg^{-1} maximum individually or in combination expressed as the acids). (*See* **Preservation of Food**.)

Antioxidants (1) Propyl gallate, 100 mg kg^{-1}; (2) butylated hydroxytoluene (BHT), 75 mg kg^{-1}; (3) butylated hydroxyanisole (BHA), 175 mg kg^{-1}; (4) any combination of propyl gallate, BHA, BHT, 200 mg kg^{-1}, but limits in 4.5.1–4.5.3 not to be exceeded; (5) natural and synthetic tocopherols, 500 mg kg^{-1}; (6) ascorbyl palmitate and (7) ascorbyl stearate, 500 mg kg^{-1} individually or in combination; (8) dilauryl thiopropionate, 200 mg kg^{-1}. (*See* **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants.)

Antioxidant Synergists (1) Citric acid (limited by GMP); (2) sodium citrate (limited by GMP); (3) isopropyl citrate mixture; (4) phosphoric acid and (5) monoacylglycerol citrate (the maximum for 3, 4, and 5 is 100 mg kg^{-1} individually or in combination).

Acidity Regulators (1) Citric acid and lactic acids and their sodium or potassium salts (GMP); (2) L-tartaric acid and its sodium and sodium/potassium salts (GMP).

Antifoaming agents Dimethylpolysiloxane (dimethyl silicone) singly or in combination with silicon dioxide, 10 mg kg^{-1}.

Vitamin Fortification

As a manufactured food staple that is reasonably homogeneous, margarine is a logical vehicle for fortification with essential fat-soluble vitamins. The addition of preformed vitamin A to margarine is mandatory in several countries including the UK, the USA, and Canada, where the levels specified are

similar. In Canada, not less than 1000 μg of retinol (3300 international units (IU) of vitamin A) per 100 g of margarine are required. This recognizes the necessity for the nutritional equivalence of margarine with butter in respect of dietary vitamin A. Whether or not the vitamin A potential of the carotenoids that may be added to margarine as colorants may be counted as a contribution to the mandatory level of fortification varies among countries.

Although the addition of vitamin A to milk has been known to cause floral and hay-like off-flavors due to photooxidation in the absence of light-proof packaging, no such problems have been reported for vitamin A-fortified margarines. The analytical variability in monitoring the amount of vitamin A present is typically 10–15%. (*See* **Food Fortification**.)

Vitamin D is added to margarine to prevent the deficiency of a nutrient, which is not naturally available in food and is essential to prevent rickets, particularly in climates that limit the exposure of the skin to sunshine. In Canada, it is mandatory to add vitamin D to margarine at a level that will assure not less than 13.25 μg (530 IU) per 100 g in the finished product. Like vitamin A, vitamin D added to margarine is relatively stable; however, the typical analytical variability in monitoring the amount present has been reported as in the order of 30%.

Canadian legislation permits the addition of vitamin E to margarine 'if added in an amount that results in not less than 0.6 mg (0.6 IU) of α-tocopherol per gram of linoleic acid present.' However, the permission has not been widely exercised. Tocopherol isomers, particularly γ and δ, may be added as antioxidants and confer some biological value, though less than the α isomer. The tocopherols occur naturally in vegetable oils but are subject to heat degradation. Table 1 illustrates the content of tocopherol isomers in some margarines on the US market in 1985. It may be noted that the ratio of α-tocopherol equivalents to C18:2 ω-6 is lower in the tub margarines than in the stick form due to the former's higher content of linoleic acid.

Table 1 Tocopherols and linoleic acid in some US margarines

Oil source	*Tocopherols (mg per 100 g)*			*Vitamin E (α-tocopherol equivalents)*	*C18:2 ω-6*[a]
	α	*γ*	*δ*		
Stick margarines					
C/CH	15.2	49.7	0.5	20.2	33.8
SBH	5.4	29.1	8.2	8.4	7.6
SBH/SB/CSH	5.8	44.0	13.4	10.2	24.8
Tub margarines					
C/CH	5.1	46.0	ND	9.7	43.6
SBH	5.2	34.1	10.0	8.6	30.4
SBH/CSH	4.8	3.1	1.8	5.1	29.7

[a]Normalized wt%.

C, corn oil; CH, hydrogenated corn oil; SBH, hydrogenated soya bean oil; SB, soya bean oil; CSH, hydrogenated cottonseed oil; nd, not determined. Data adapted from Slover HT *et al*. (1985).

Enrichment with Plant Sterols

Plant sterols, and their hydrogenated counterparts, stanols, have been shown to lower blood cholesterol, which is an advantage in reducing the risk of cardiovascular disease. This prompted the introduction in 1995 of a Finnish margarine (Benecol) enriched with a fat-soluble stanol ester (9%) followed in 1998 by a stanol-enriched low-fat spread (~40% fat). In 1999, the Unilever Research Centre in The Netherlands announced its introduction of a new 'heart-health spread' (Take Control/pro.activ) in both the USA and Europe. This is a reduced-fat spread with 8% plant sterols, so that regular consumption of 20 g per day of the spread would contribute about 1.6 g of plant sterols daily.

Lipid component The total fat in margarine is routinely controlled through metering according to formula specifications and may be monitored in inspection by standard methods. Of most interest, from the perspectives of function and nutrition, is the nature of the triacylglycerols (TAGs) that make up the oils and fats of a margarine basestock. Since each TAG molecule is made up of one molecule of glycerol and three fatty acids, the structure of the constituent fatty acids and their arrangement on the glycerol stem affect its physical properties. These features are dictated by the origin of the oils and fats used in margarine formulation and by their processing prior to blending. The lipid component of margarine is most frequently characterized by its fatty acid composition. Local crops often are dominant constituents such as canola and soybean oils in North American products.

Individual fatty acids, which consist of a carboxyl group and a hydrocarbon chain, are distinguished from one another by chain length and by degree of saturation with hydrogen. Shorter fatty acid chains and greater unsaturation (i.e., less hydrogen, therefore more double bonds) within a TAG mixture favor fluidity and *vice versa*. Unsaturation has the greater softening effect provided that the double bonds are in the natural or *cis* form. This geometric configuration causes the fatty acid chain to double back on itself, resulting in a bend that discourages the secondary association among adjacent molecules, which leads to hardening. Certain processing conditions, such as hydrogenation, convert the natural *cis* isomer to a straighter-chain *trans* isomer that favors hardening. (*See* **Fatty Acids**: Properties.)

Table 2 summarizes the fatty acid composition of some margarines on the Canadian and European Markets between 1995 and 1998. There were some distinct regional differences. Danish margarines had higher levels of saturated fatty acids than those from the United Kingdom and Canada reflecting differences in the sources of the fats and oils which made up their respective basestocks. Canadian spreads were significantly higher in *trans* fatty acids than their European counterparts, likely due to differences in the processing used for hardening.

The data on Canadian spreads shown in Table 2 are drawn from a study by Health Canada of 109 margarine brands (30 print and 79 tub), selected on the basis of their importance in the Canadian market share. The relative amounts (grams per 100 g of total fatty acids) of total *trans* fatty acids (TFA), *t*-18:1, *trans,cis/cis,-trans*-18:2 (*ct*-18:2), *trans,trans*-18:2 (*tt*-18:2), *trans*-18:3 (*t*-18:3), total saturated fatty acids (SFA), *c*-18:1 (all isomers), *cis*-18:2*n*-6 (linoleic acid) and *cis*-18:3*n*-3 (α-linolenic acid), were published in 1998. The trend over time is a positive decrease in TFA from a 'most frequent occurrence' of 30–35% in 1979–1983, 20–25% in 1990, to 15–20% in the 1995 margarines. These changes reflect advances in processing technology and increased use of liquid oil in blends with partially hydrogenated oils in margarine basestocks. However, Table 3 shows that Canadian margarines, like those of Austria and the USA, have average TFA levels substantially higher than those of several European countries and Australia. This is a nutritional disadvantage, considering that dietary TFAs, like saturated fatty acids, adversely affect blood lipid patterns, increasing the risk of cardiovascular disease.

Table 2 Fatty acid composition of some Canadian and European margarines (grams per 100 g of fatty acids)

Source/market form	Total saturates	Total trans	Unsaturates		
			C18:1	C18:2$^{\omega-6}$	C18:3$^{\omega-3}$
Canadian print[a]	16.5	35.4	35.8	9.1	1.9
Canadian tub[a]	15.5	13.7	38.9	24.1	5.7
UK[b]	22.0	0.6	29.0	45.4	2.2
Denmark[c]					
Hard (frying)	35.0	4.2	39.3	14.0	4.4
Semisoft	34.6	1.2	32.4	31.5	4.1
Soft	27.4	1.3	21.1	49.2	0.8

[a]Means of 10 products with the greatest market share in 1994. Adapted from Ratnayake WMN, Pelletier G, Hollywood R, Backer S and Leyte D (1998) *Trans* fatty acids in Canadian margarines: recent trends. *Journal of the American Oil Chemists Society* 75: 1587–1594.
[b]Van den Bergh Foods. Adapted from Hendricks HFJ, Westrate JA, van Vliet T and Meijer GW (1999) Spreads enriched with three different levels of vegetable oil sterols and the degree of cholesterol lowering in normocholesterolaemic and mildly hypercholesterolaemic subjects. *European Journal of Clinical Nutrition* 53: 319–327.
[c]Means of 20, 8, and 6 of hard, semisoft and soft margarines, respectively. From Oveson L, Leth T and Hansen K (1996) Fatty acid composition of Danish margarines and shortenings, with special emphasis on *trans* fatty acids. *Lipids* 31: 971–975.

Table 3 *Trans* fatty acid (TFA) levels in margarines by country[a]

Country	Mean TFA level (percentage of total fatty acids)	Report source and date
Australia	13.1	Mansour and Sinclair, *Asia-Pacific Journal of Clinical Nutrition 3* (1993) 155
Austria	22.9	Ulberth and Henniger, *Journal of the American Oil Chemists Society 69* (1992) 829
Canada	23.6	Ratnayake *et al*. *Journal of the American Oil Chemists Society 75* (1998) 1587
Denmark	3.0	Oveson *et al*. *Lipids 31* (1996) 971
France	3.8	Bayard and Wolff, *Journal of the American Oil Chemists Society 72* (1995) 1485
Greece	10.0	Kafatos *et al*. *International Journal of Food Science and Nutrition 45* (1994) 107
USA	22.6	Emken, *American Journal of Clinical Nutrition 62* (1995) 669S

[a]As reported by Ratnayake WMN, Pelletier G, Hollywood R, Backer S and Leyte D (1998) *Trans* fatty acids in Canadian margarines: recent trends. *Journal of the American Oil Chemists Society* 75: 1587–1594.

Analytical methods for the fatty acid compositions shown in Table 2 varied among investigators. For the Canadian spreads, a 1 g central plug was taken for each of six samples of the same brand and homogenized. Rather than extract the fat, samples of about 200 mg of the homegenate were converted directly to methyl esters by treatment with 1.5 ml of NaOH in methanol at 100 °C for 5 min followed by mixing with 2 ml of 14% BF_3-methanol at 100 °C for 30 min. The fatty acid methyl esters were recovered with hexane and analyzed using an SP-2560 capillary column in a Hewlett Packard 5890 Series II gas chromatograph (GC) equipped with a HP GC Chemstation. The column oven temperature was held at 165 °C for 75 min, then programmed at 7 °C min^{-1} to a final temperature of 210 °C, which was held for 30 min. While this single-step GC analysis was suitable for samples with less than 5% *trans* 18:1, those with more required the additional step of total *trans* measurement by IR and appropriate corrections. Fatty acids of the UK margarine were analyzed by similar methods in the Netherlands. In the earlier Danish study, fatty acids were extracted from the margarine with isooctane following boiling with methanol and boron trifluoride. The temperature program and column for GC analyses were chosen to provide good separation between cis and trans C18:1. Certified reference materials were used by all

investigators. (*See* **Chromatography**: Thin-layer Chromatography; **Fatty Acids**: Analysis; **Spectroscopy**: Infrared and Raman.)

Hardening

The demand for table fats with the spreadable properties that characterize margarine could not be satisfied without the technology to convert liquid oil into a semisolid product. The two commercial processes used to achieve oil hardening are hydrogenation and interesterification. While hydrogenation is the more common process in North America, interesterification, or transesterification as it may be referred to, is widely used in parts of Europe. (*See* **Vegetable Oils**: Oil Production and Processing.)

The hydrogenation process adds hydrogen at the double bonds of unsaturated fatty acids under conditions of high temperature and high pressure in the presence of a suitable catalyst, usually nickel. As well as increasing the hydrogen content of the unsaturated fatty acids of TAGs, hydrogenation conditions favor a shift in the position of the double bonds along the fatty acid chain and/or a change in geometry to a straighter-chain *trans* form. The *trans* isomers of unsaturates, depending on their frequency, can account substantially for the hardening that results from hydrogenation. However, hydrogenation is a flexible process that may be more or less selective in its effects. By using higher temperatures with lower pressure (e.g., 200 °C, 6 psig) the polyunsaturates are selectively hydrogenated before the monounsaturates but, at the same time, isomerization is encouraged. In contrast, more random hydrogenation occurs under nonselective conditions, which use lower temperatures and higher pressures (e.g., 135 °C, 60 psig), resulting in more saturates and fewer *trans* isomers.

A comparison of the fatty acid profiles in Table 2 for the print and tub forms of Canadian spreads from similar oil sources illustrates the influence of saturates and total *trans* unsaturates on margarine firmness. As the content of total saturates was similar in the two market forms, the greater firmness of the print margarines must be attributed to their higher content of *trans* unsaturates. Because of their novel configuration, *trans* unsaturates cannot be expected to have the unique biological properties of their original *cis* form. For this reason, *trans* fatty acid values are not included in the values for monounsaturated and polyunsaturated fatty acids cited on food labels in North America, although they are included in the total fat values. (*See* **Vegetable Oils**: Composition and Analysis.)

Hardening by interesterification commonly involves combining highly saturated fats and unsaturated oils with a catalyst precursor such as sodium methoxide, alkali metals, or alkyl metal alkylates (about 0.1%) at temperatures ranging from 5 to 135 °C. Hardening depends upon the concentration of saturates because *trans* isomerization does not occur under these relatively mild processing conditions. It is likely that the low level of TFAs shown in Table 2 for the UK spread reflects processing by interesterification. Margarines made by this process have the further advantage of crystal stability. Interesterification conditions cause the fatty acids to trade places within and among the TAGs. As a result of these rearrangements, new and more random TAGs are formed that are less inclined to undergo the degree of TAG packing, which leads to large crystals. Accordingly, interesterification fosters the development of the stable fine β^1 crystals that are desirable for margarine smoothness.

Quality-control procedures

The control of quality in margarines begins with ingredient selection and continues with the monitoring of microbiological and engineering processing conditions, including packaging. The finished product quality is assessed from the packing line for compliance with in-house standards, regulatory requirements, labeling claims, and, where appropriate, trade standards. (*See* **Quality Assurance and Quality Control**.)

Margarine sheen and color are usually judged visually by comparison with a series of color standards using either the Lovibond Tintometer or sets of plastic or enamelled sticks arranged in color gradations from light to dark. Alternatively, the three color dimensions of hue, value, and chroma, may be described by visual comparison with chips in the matte or glossy edition of the '*Munsell Book of Colors*.' In this instance, hue is the color identity, value the lightness or brightness, and chroma the vividness or saturation. The American Society for Testing and Materials has published standards to guide such evaluations. It is important that inspectors who do these assessments be screened for normal color vision, be adapted to daylight and examine the samples when they are illuminated by a controlled source against a gray to white background.

The physical properties of firmness and spreadability are most commonly monitored by the solid fat content (SFC) at temperatures of 10, 21.1, and 33.3 °C, using low-resolution nuclear magnetic resonance under protocols defined by the International Union of Pure and Applied Chemistry (IUPAC). The IUPAC method for SFC provides a convenient and more accurate alternative to the traditional dilatometric determination of the solid fat index. Nuclear magnetic resonance provides the advantage

of accurate SFC readings on margarine 'as processed.' Recent studies have been aimed at relating the macroscopic rheological properties of fat crystal networks to the various levels of structure within the network. This involves not only the TAG structures and the polymorphic and polytypic levels of structure but also the morphology and spatial distribution of the microstructure. For example, fractal scaling theory appears promising to relate spatial distribution within the fat crystal network to the mechanical property of shear elastic modulus.

Sensory properties of spreadability, graininess, flavor release rate, and flavor blend call for analysis by sensory panels that have been trained to recognize typical margarine characteristics and defects as opposed to naïve consumer panels. References that could be useful in training for margarine assessment can be adapted from some of those developed for judging oxidized and sour flavor defects in butter, and for rancid and fruity off-flavors in cheese. It is critical that the preparation and presentation of margarine samples be well controlled to limit changes due to physical handling and temperature. Sampling the spread as cubes, balls, or cylindrical cores is less disruptive to texture than loading it into a container. Temperature changes may be minimized by the use of polystyrene foam blocks as sample holders. (*See* **Sensory Evaluation:** Sensory Characteristics of Human Foods.)

Chemical measurements to control margarine quality normally require monitoring of vitamin additions and fatty acid analyses, depending on the intended label claims. In Canada, '... no statements or representations are permitted for any individual fatty acid other than linoleic acid or for any group of fatty acids other than polyunsaturates, monounsaturates, or saturates. If a statement or claim for linoleic acid is made, a declaration of total fat, the above-mentioned group of fatty acids, and cholesterol is required in addition...' The amounts of these fat components present in a 10 g (two teaspoon) serving must be cited. Currently, negotiations are underway to harmonize North American nutrition recommendations as well as the information that is required and permitted on food labels.

See also: **Cholecalciferol**: Properties and Determination; **Emulsifiers**: Organic Emulsifiers; Uses in Processed Foods; **Fats**: Classification; **Fatty Acids**: Properties; Analysis; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Food Fortification**; **Preservation of Food**; **Quality Assurance and Quality Control**; **Retinol**: Properties and Determination; **Sensory Evaluation**: Sensory Characteristics of Human Foods; **Tocopherols**: Properties and Determination; **Vegetable Oils**: Oil Production and Processing; Composition and Analysis

Further Reading

Campbell I (1995) Food labeling in Canada. In: Przybylski R and McDonald BE (eds) *Development and Processing of Vegetable Oils for Human Nutrition*, pp. 114–120. Champaign, IL: AOCS Press.

Hendricks HFJ, Westrate JA, van Vliet T and Meijer GW (1999) Spreads enriched with three different levels of vegetable oil sterols and the degree of cholesterol lowering in normocholesterolaemic and mildly hypercholesterolaemic subjects. *European Journal of Clinical Nutrition* 53: 319–327.

Johnson B (1999) Bakery margarines. *Cereal Foods World* 44: SR9–SR12.

Narine SS and Marangoni AG (1999) Relating fat crystal networks to mechanical properties: a review. *Food Research International* 32: 227–248.

Ratnayake WMN, Hollywood R and O'Grady E (1991) Fatty acids in Canadian margarines. *Canadian Institute of Food Science and Technology Journal* 24(1/2): 81–86.

Ratnayake WMN, Pelletier G, Hollywood R, Backer S and Leyte D (1998) *Trans* fatty acids in Canadian margarines: recent trends. *Journal of the American Oil Chemists Society* 75: 1587–1594.

Rousseau D and Marangoni AG (1998) The effects of interesterification on physical and sensory attributes of butterfat and butterfat–canola oil spreads. *Food Research International* 31: 381–388.

Slover HT, Thompson RH, Davis CS and Merola GV (1985) Lipids in margarines and margarine-like foods. *Journal of the American Oil Chemists Society* 62: 775–786.

Vaisey-Genser M and Vane BK (1995) Sensory evaluation of margarine. In: Warner K and Eskin NAM (eds) *Methods to Assess Quality and Stability of Oils and Fat-containing Foods*, pp. 76–106. Champaign, IL: AOCS Press.

Yano J and Sato K (1999) FT-IR studies on polymorphism of fats: molecular structures and interactions. *Food Research International* 32: 249–259.

Dietary Importance

D H Morris, Mainstream Nutrition, Toronto, Ontario, Canada
M Vaisey-Genser, University of Manitoba, Winnipeg, Manitoba, Canada

Historical Background

Emperor Napoleon III prompted the invention of margarine in 1869 by calling for a cheap alternate to butter to feed his armies as they marched across Europe. A French chemist, Hippolyte Mège Mouriès, responded with a formula in which an extract from beef tallow

replaced butterfat. Between 1871 and 1874, Mège Mouriès sold his patent to Dutch, British, US, and Prussian companies. Present-day companies based in The Netherlands continue to be world leaders in margarine design and production. This article considers the importance of margarine in the diet from the perspective of its economic, sensory, and nutritional characteristics. In recent years, margarine has emerged as an important functional food. Although both standard butter and margarine are 80%-fat products and hence high in energy, margarine, being a formulated food, can be designed to have nutritional and sensory features that differ from its prototype.

Economic Issues

Price

Margarine is now accepted as a food staple in most countries where table-spread use is habitual. Initially margarine's advantage over butter was chiefly one of price. By the beginning of the twenty-first century, margarines had become available within a range of prices, from lower-cost 'regular' brick- or stick-type products to higher-cost soft 'tub' margarines that could claim a nutritional advantage based on fatty acid composition. This versatility is illustrated in Table 1, which lists the prices of several margarines available in a Canadian superstore in October 2000, together with a summary of the nutrition information presented on the package. The price of Becel margarine, which qualified for the most detailed label claim, was comparable to that of butter. The generic brick margarine was the least expensive, being about a third of the price of butter. Accordingly, consumers may reduce food costs by purchasing their spreads primarily on the basis of price. The irony of the price-based choice among Canadian margarines is that it favors the health prospects of the more affluent. Lower-income consumers, who as a group might have less nutrition knowledge, would be expected to choose the cheaper product. The cost to their families is to miss the opportunity of establishing a healthier eating pattern.

However, economic considerations by the thrifty may explain why a 1982 survey of over 500 western Canadians indicated that margarine was the most commonly used fat for home-baking products such as cookies and cakes. Consumer interest in using margarine as an alternate to shortening is recognized in parts of Europe (e.g., in Switzerland) where 100%-fat 'margarines' are available in grocery stores.

Patterns of Use

North American table-spread consumption patterns, shown in Figure 1, indicate that margarine use in Canada increased to match that of butter by 1975 and has exceeded it ever since. In the USA, the dominance of margarine began more than a decade earlier, following the passage of the Margarine Act enacted by President Truman in 1950 to lift restrictive taxes and laws on margarine production and sales. Concurrent with the increased dominance of margarine, the consumption of table spreads in total has shown a steady decline in both North American countries. This is likely a response to promotional campaigns of health authorities to reduce fat intake in the interest of decreasing the risks of obesity and cardiovascular disease.

Sensory Characteristics

From the sensory perspective, the presence of fat helps make food taste good – smoother, creamier, richer, and more tender. This is so whether the fat is a structural ingredient as in meat or an avocado, as an ingredient in sauces or baked goods, or as a spread on toast or in sandwiches. Consumers surveyed in

Table 1 Examples of table-spread prices (Canada, October, 2000)

Brand	*Market form*	*Price of a 1-lb (0.45 kg) container (Cdn $)*	*Label claim*
Margarine			
Generic	Brick	$1.09	% RDI for vitamins A, D
Generic	Soft	$1.49	% RDI for vitamins A, D; amount of calories, saturated, monounsaturated, and polyunsaturated fatty acids in 10 g (2 tbsp)
Supermarket brand	Soft	$1.69	% RDI for vitamins A, D
Imperial	Soft	$2.29	% RDI for vitamins A, D
Becel	Soft	$2.89	% RDI for vitamins A, D, E; 0 *trans* fatty acids; amount of cholesterol, saturated, monounsaturated and polyunsaturated fatty acids in 10 g (2 tsp); low in saturated fat, nonhydrogenated
Butter			
Supermarket brand	Brick	$2.99	
Beatrice	Brick	$3.99	

RDI, recommended daily intake.

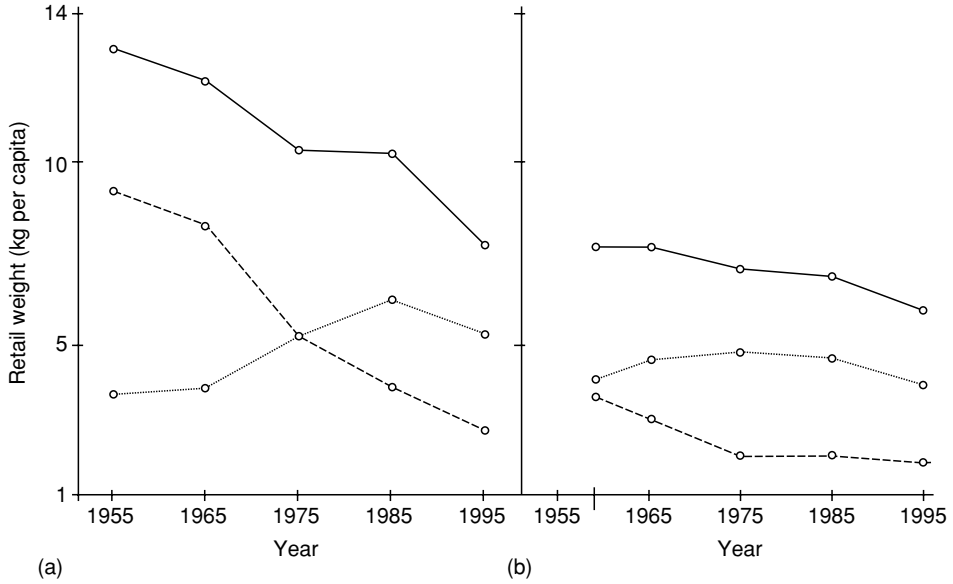

Figure 1 North American patterns of table-spread consumption. (a) Canada; (b) USA. Continuous line, total butter and margarine; dotted line, margarine; dashed line, butter. Data sources: Anonymous (1977) Margarine US per capita consumption may fall this year. *Journal of the American Oil Chemists Society* 54: 840A–841A; Statistics Canada (1998) 32–230-XB; US Department of Agriculture (1966–1998), *Oil Crops Situation and Outlook*, Washington, DC: USDA.

Canada between 1989 and 1997 cited taste as the primary reason for choosing a particular food, one which they were reluctant to forgo, even though 44% agreed that people should be cautious when serving food with fat. Similarly, in a survey of the dietary practices of some 200 aboriginal adults in the northern USA published in 1999, 'taste' was the most frequently mentioned barrier to cutting the intake of high-fat foods. In these cases the term 'taste' apparently was synonymous with how much the respondents liked the food; that is, it expressed their hedonic response to the overall sensory character of the food.

With respect to margarine, experienced sensory consultants, who were enlisted by *Consumer Reports* in 1982 to evaluate 17 different US margarines, described an excellent margarine as one with 'a pale yellow color, a slight hint of dairy flavor and aroma, an appropriate texture for its type, melts readily in the mouth, and has no more than a hint of the parent oil's flavor or aroma.' A later study of US consumer opinion of margarine attributes judged 'good taste' and 'good texture' as the most important features. In this instance, taste might have been more appropriately termed flavor, which embodies both tastes and odors.

Texture

Textural characteristics of margarines include firmness, spreadability, smoothness, and mouthmelt. With each of these, it is critical that the temperature of the margarine samples be rigorously controlled before and during testing. Of the four textural parameters cited above, firmness is the quality that best lends itself to measurement with standard laboratory instruments. For example, 'resistance to shear' can be measured physically with an Instron or other shearing device. What is measured is either the force required to obtain a given deformation or the deformation achieved under a given force. Such instrumental readings can be related to sensory estimates of firmness by the hand pressure required to cut with a knife, again provided that there is strict temperature control of the sample and related implements.

Spreadability represents the ease with which a spread can be applied in a thin, even layer to a controlled surface such as a slice of bread. It is a complex sensory characteristic reflecting the relative amount of solid fat present together with its crystalline structure. For a product to be spreadable, it must be plastic or moldable at the temperature of use, which will normally be between 4 and 22 °C. Fat plasticity depends upon the triacylglycerols involved in the margarine formula and the fineness of the fat crystal matrix which accounts for its structure. The advantage of margarine is that it 'worksoftens' more easily than butter does, even when their initial hardness values are equal. That is why margarines can be spreadable when taken directly from the refrigerator, whereas butter cannot. Easy spreading has a greater advantage than simply the work involved. The thinner layer of spread achievable with margarine contributes to the public health goal in developed countries to decrease fat intake.

Smoothness to the tongue is expected of a good table spread. It reflects a structure of crystals that are small enough to be imperceptible in the mouth, a size that has been estimated to be in the order of 22 μ through controlled storage studies where crystal size was measured under a polarized light microscope and related to sensory tests of smoothness. Margarines with larger crystals or clusters of smaller ones packed together will be perceived as the defect of sandiness or graininess.

Mouthmelt sensation from a table spread is important to both its texture and release of flavor. Margarines will be criticized as waxy or gummy if they fail to melt rapidly and completely in the mouth. Prompt melting has another advantage; it generates a pleasant cooling effect due to absorption of heat of crystallization. Sensory testing using stopwatches has confirmed that butter melts in the mouth faster than the same amount of margarine does, but the degree of temperature change realized in the mouth has not yet been published.

Flavor

Flavor is a combined sensory response by receptors in the mouth and nose to taste and odor stimuli. It is a sensation complicated by individual variations in dentition, breathing rate, and salivary flow.

Of the four basic tastes – sweet, salty, sour, and bitter – saltiness is the most relevant to table spreads, and then only where the use of salted spreads is the norm. In the latter case, salt is added for its value as a preservative as well as for taste, and even then at relatively low levels, usually 1–3%. The high fat content of standard table spreads appears to mask saltiness perception. Sensory tests of butter prepared with a range of salt levels showed that salt intensity in the spread was only one-tenth the magnitude of the same concentrations in water, although the growth of the salty sensation over increasing concentration was similar in both media. Apparently the oil in which the salt/water solution is suspended limited the intensity of salt perception. The eight-member trained panel used in this experiment considered that a level of 1% salt in butter was the most pleasant of eight concentrations ranging from 0.5 to 4.0%.

Odor perception as well as taste can be affected by the composition of the dispersion medium. For example, experimental margarines, containing diacetyl and other dairy flavors dispersed in a lipid mixture of soybean oil melted and recrystallized with either palm stearin or palm olein, differed in perceived flavor intensity. Although the time for flavor intensity to peak for both spreads was similar, total flavor was significantly greater in the spread containing the faster-melting palm olein. Head-space measurement of diacetyl confirmed the results of sensory tests.

Gas chromatography combined with mass spectrophotometry has been used since the 1950s to document volatile compounds in foods using techniques of simultaneous distillation extraction. However, identifying which volatiles in margarine and other foods are crucial to perceived flavor continues to be hampered by several factors. These include the number of compounds released at mouth temperature, the partitioning of flavor compounds between the oil and water phases and between the aqueous phase in the mouth and the gas phase in the nose, and the complexity of their odor impact. Novel tools such as neural networks used with the results of increasingly sophisticated methods for sample head space and oral breath analysis appear promising.

Popular opinion is that butter, in comparison with margarine, still has the better 'buttery' flavor according to studies conducted during the latter part of the 1980s and 1990s in both North America and Europe. Where this is so, it may be a function of the physical characteristics of the margarine emulsion. It has been speculated that 'if the aqueous droplets are uniformly small and heavily stabilized with emulsifiers, the flavor... release will be delayed.'

Nutritional Aspects

Margarine use increased in the 1970s and 1980s when scientists determined that a diet high in fat, especially saturated fat, increased blood levels of low-density lipoprotein (LDL) cholesterol – the so-called 'bad' cholesterol – and decreased blood levels of high-density lipoprotein (HDL) cholesterol – the 'good' cholesterol. High blood levels of total and LDL cholesterol are risk factors for coronary heart disease (CHD). The margarine industry responded to the consumers' demand for 'heart-healthy' food products by formulating margarines that were lower in total fat, saturated fat, and *trans* fatty acids. In some countries, margarine is also a source of one or more essential fat-soluble vitamins.

Vitamin Enrichment

Margarine is a convenient vehicle for the addition of the fat-soluble vitamins A and D. The international standard for margarine issued by the Codex Alimentarius Commission of the Food and Agricultural Organization (FAO) in Rome recommends their addition. In Canada and the UK, the addition of vitamins A and D is mandatory. In the USA, the addition of vitamin A is mandatory, while the addition of vitamin D is optional. Such enrichment gives margarine the edge over butter, in which the vitamin A content varies with the diet of the dairy cow and vitamin D is essentially absent. Vitamin D, which assists in the

absorption of calcium, is naturally produced from the sun's activation of precursors in the skin. However, in cold climates where skin exposure to the sun is precluded by outerwear for many months of the year, foods fortified with vitamin D, including margarines, are an important public health measure.

The addition of vitamin E to margarine is permitted in Canada, if added in an amount that results in not less than 0.6 IU of α-tocopherol per gram of linoleic acid present. However, most Canadian margarine manufacturers do not exercise this option. Fortification of margarine with vitamin E is not permitted in the USA, although tocopherols may be added as antioxidants.

Total Fat, Saturated Fat, and Cholesterol

Dietary patterns low in total fat, saturated fat, and cholesterol are recommended to maintain a healthy weight and reduce CHD risk. Margarine, which contains no cholesterol and is low in saturated fat, is perceived as a healthy alternative to butter. The compositional differences among various margarine products and butter are shown in Table 2. Standard butter and standard margarine both provide the same amount of energy (about 425 kJ or 101 kcal per tbsp or 15-g serving) and fat (11–12 g per tbsp or 15-g serving); reduced-fat margarines provide about half the energy and fat per serving as regular margarine. Butter is a source of dietary cholesterol, whereas margarine is not.

Diets low in saturated fat are considered desirable because saturated fat is the main dietary determinant of blood LDL cholesterol concentrations. Diets providing less than 10% of energy as saturated fat are recommended for the general population; individuals with diagnosed cardiovascular disease and elevated blood LDL cholesterol levels are advised to consume a diet containing less than 7% saturated fats. For an adult consuming 8372 kJ or 2000 kcal day^{-1}, these recommendations amount to a daily intake of only 16 or 22 g of saturated fat from all dietary sources.

One tbsp (15 g) of butter provides about 7.2 g of saturated fat, whereas 1 tbsp (15 g) of hard, brick margarine provides about 2.1 g – a saving of about 5 g day^{-1} of the recommended saturated fat allowance. Soft tub margarines, reduced-fat margarines, and liquid margarines, which are generally recommended over butter and brick margarines, all provide less than 2 g saturated fat per tbsp, allowing for an even greater reduction in saturated fat intake. Thus, substituting soft and liquid margarines for butter and other saturated fats in the daily dietary pattern helps reduce the saturated fat intake but not necessarily the total fat intake.

Trans Fatty Acids

Trans fatty acids are found naturally in butter, milk, beef, and lamb fats and in commercially prepared, partially hydrogenated margarines and solid cooking fats like shortenings. *Trans* fatty acids are produced in the margarine manufacturing process when liquid vegetable oils are hydrogenated to form a semisolid or solid fat with many of the same properties as animal fats and, in some cases, to increase the oxidative and thermal stability of the fat or oil. Indeed, partially hydrogenated vegetable oils were developed in part to take the place of highly saturated animal fats used in spreads and for frying and baking.

During the hydrogenation process, some of the naturally occurring *cis* double bonds in the vegetable oil fatty acids are changed to *trans* double bonds, so that the resulting *trans* fatty acids resemble saturated fats more than unsaturated fats. In Canada, nearly 70% of the vegetable oils consumed by humans are partially hydrogenated.

Table 2 Fatty acid composition of butter and margarines (serving size = 1 tbsp)

Product	*Energy*		*Cholesterol (mg)*	*Total fat (g)*	*Saturated fatty acids (g)*	*Monounsaturated fatty acids (g)*	*Polyunsaturated fatty acids (g)*	Trans *fatty acids (g)*
	kJ	*kcal*						
Butter	426	102	31.1	12	7.2	3.3	0.4	0.3
Margarine, regular, hard, corn and soybean oils	423	101	0.0	11	2.1	5.2	3.5	1.9
Margarine, soft, soybean (hydrogenated and regular)	423	101	0.0	11	1.9	5.1	3.8	0.8
Margarine-like spread (40% fat), soybean (hydrogenated)	208	50	0.0	6	0.9	2.4	2.0	0–2.5
Liquid margarine	428	102	0.0	11	1.9	4.0	5.1	0

One tablespoon = 3 teaspoons or about 15 g. Values for saturated, monounsaturated, and polyunsaturated fatty acids were obtained from the US Department of Agriculture, Agriculture Research Service, USDA Nutrient Database for Standard Reference, Release 13 (1999), Nutrient Data Laboratory Home Page at www.nal.usda.gov/fnic/foodcomp. Values for *trans* fatty acids for all products except liquid margarine were obtained from Jonnalagadda SS, Mustad VA, Champagne C and Kris-Etherton PM (1995) Margarine and plasma cholesterol – a perspective for dietitions on *trans* fatty acids in the diet. *Perspectives in Applied Nutrition* 2: 9–16. The value for *trans* fatty acids in liquid margarine was obtained from the National Association of Margarine Manufacturers (1999), available at www.margarine.org.

The US Department of Agriculture has estimated that biohydrogenation in ruminants accounts for up to 20% of the *trans* fatty acids in the US diet, with the remaining 80% being derived from food products formulated with partially hydrogenated fats like margarine, shortening, shortened baked goods (cake, cookies, pastry) and deep-fried snacks (corn and potato chips, French fries, doughnuts, extruded products). In the USA, the average per capita intake of *trans* fatty acids has been estimated at about 12.5 g day^{-1} or between about 5 and 8% of the total fat in the diet. In Canada, hydrogenated margarine is the main source of *trans* fatty acids, but other foods, particularly cookies, crackers, doughnuts, potato chips, and French fries, also contribute substantially to *trans* fatty acid intake because they are made with hydrogenated shortening.

The primary concern about *trans* fatty acids in the diet is their adverse effects on blood lipids. The results of clinical trials indicate that moderate and high levels of dietary *trans* fatty acids raise blood total and LDL cholesterol; their effect on blood HDL cholesterol concentrations has been variable. Several comprehensive reviews of the literature report an emerging consensus: when substituted for polyunsaturated fats in the diet, *trans* fatty acids tend to act like saturated fatty acids in their effects on blood lipids and thus are likely to increase the risk of CHD. However, when substituted for animal fats and vegetable oils high in saturated fats, *trans* fatty acids have been shown to lower blood total and LDL cholesterol concentrations. Because *trans* fatty acids are typically found in high-fat foods, a prudent approach to reducing CHD risk is to reduce total fat intake in general and saturated fat intake in particular. By choosing foods low in fat and saturated fat, consumers will also reduce their *trans* fatty acid intake.

Plant Sterols and Stanols

It has been known since the 1950s that plant sterols lower blood cholesterol levels. Plant sterols are analogs of cholesterol, differing structurally from cholesterol in their side-chain configuration. Sitosterol, campesterol and stigmasterol are the most important plant sterols from a nutritional standpoint and comprise up to 98% of the total sterols in certain seed and vegetable oils. Plant sterols can be hydrogenated to form plant stanol esters (e.g., sitostanol, campestanol, stigmastanol), a process that improves their solubility in margarine.

In studies conducted in the 1980s and 1990s, margarine was chosen as the vehicle for supplementing the diet with plant sterols because it is widely consumed and easily incorporated into usual dietary patterns. Recent studies show that daily consumption of a margarine enriched with 1–3 g plant sterols or plant stanols is sufficient to lower blood total cholesterol levels by 10% and LDL cholesterol levels by 10–15%. Blood concentrations of HDL cholesterol and triacylglycerols are not affected by plant sterol or plant stanol esters. Consumption of more than about 3 g plant stanol esters per day does not appear to increase its effectiveness.

The strength of these findings led the US Food and Drug Administration to authorize a claim on food labels and in food labeling that plant sterol/stanol esters reduce risk of CHD. The health claim, which became effective in September 2000, was designed to inform the general public that plant sterol/stanol esters should be consumed as part of a diet low in saturated fat and cholesterol. This followed the lead of some European countries, where plant sterol/stanol-enriched margarines were already available. The Raisio Group in Finland, for example, introduced Benecol in 1995; the multinational company Unilever/Lipton introduced a similar product, Take Control, into the US market in 1999.

One potential threat to the widespread adoption of margarines enriched with plant sterols and stanols is the reported adverse effect of these compounds on the absorption of nutrients such as fat-soluble vitamins and carotenoids. In some studies, consumption of about 1–3 g plant sterols daily for 3.5 weeks resulted in 20% reductions in plasma α- and β-carotene, lycopene, and α-tocopherol concentrations. It remains to be determined whether this adverse effect is clinically significant and whether plant sterol- or stanol-enriched margarines should or could be modified to compensate for the effect.

Emerging Health Issues

Although margarine consumption appears to benefit patients with hypercholesterolemia, who are advised to reduce their dietary intake of saturated fat and cholesterol, and may help the general public achieve healthy dietary patterns to reduce the risk of CHD, questions about a possible link between margarine consumption and risk of certain types of cancer have begun to emerge. For example, in a study of 472 women diagnosed with early-stage breast cancer in 1982–84, the reported baseline total consumption of butter, margarine, and lard (measured as times per day) was associated with an increased risk of recurrence of the cancer. Consumption of butter, margarine, and lard appeared to increase the risk of recurrence by 30% for each time they were consumed per day. Unfortunately, the study analysis did not attempt to separate the effects of margarine from butter and lard. In a study of 320 men with

histologically confirmed prostate cancer and 246 men without prostate cancer or a history of the disease, added lipids such as butter on bread, butter for cooking, margarine on bread and for cooking, and seed oils were positively associated with increased risk of prostate cancer. When adjustments were made in the statistical analysis for other food groups, butter, margarine, and seed oils were all positively associated with prostate cancer risk, although the effect was only significant for seed oils.

These findings are suggestive but far from conclusive. Both breast and prostate cancer are hormone-sensitive cancers, meaning that in the early stages of tumor development, tumor growth is influenced by the sex hormones. Thus, hormonal status and lifetime exposure to phytoestrogens (plant compounds that exhibit weak hormone-like activity in the human body) are confounding factors in these studies. Other factors such as family history, age, and diet, particularly the intake of total fat, animal fat, polyunsaturated fatty acids, vitamins, and antioxidants, are known to influence cancer processes. Indeed, the type of dietary pattern adopted over the course of a lifetime appears to be an indicator of cancer risk. Western-type diets high in fat, red meat, protein, fast food, refined grains and sugar, and low in fruits and vegetables are linked with increased risk of cancer, while so-called prudent diets high in fruits, vegetables, and whole-grain cereals and low in fat are associated with low cancer risk. For colon cancer risk, consumers who substitute margarine for butter, whole grains for refined grains, low-fat dairy products for high-fat dairy products, and poultry for red meat have a lower risk than those who eat a typical western diet. In the final analysis, dietary fats have significant effects on cancer processes through their biologic roles in regulating cell growth and immune responses, and a prudent approach to reducing cancer risk is to reduce intake of all fats.

Role of Margarine in the Diet

Margarine is recognized by health professionals and consumers alike as a healthy table spread and a cheaper alternative to butter for use in cooking and in food preparation. Margarine makes an important nutritional contribution to the diet by being a source of one or more essential fat-soluble vitamins such as vitamins A and/or D and by being low in saturated fat. Margarine is cholesterol-free. Newer margarine products are also low in *trans* fatty acids and in some cases are reduced in fat and energy. Some margarines can also be classified as a functional food, because they are enriched with plant sterol or stanol esters that have been shown to lower blood cholesterol levels. Its price, taste, spreadability, and convenience have made margarine a dietary staple in many countries around the world.

See also: **Butter**: Properties and Analysis; **Cancer**: Diet in Cancer Prevention; **Cholesterol**: Factors Determining Blood Cholesterol Levels; Role of Cholesterol in Heart Disease; **Coronary Heart Disease**: Etiology and Risk Factor; **Margarine**: Types and Properties; Methods of Manufacture; Composition and Analysis; **Sensory Evaluation**: Appearance; Texture; Aroma; Taste; **Fatty Acids**: *Trans*-fatty Acids: Health Effects

Further Reading

Chrysam M (1985) Tablespreads and shortenings. In: Applewhite TH (ed.) *Bailey's Industrial Oils and Fats*, vol. 3, pp. 41–127. New York: Wiley.

Food and Drug Administration (2000) Interim final rule: food labeling: health claims; plant sterol/stanol esters and coronary heart disease. *Federal Register* 65: 54686–54739 available at www.access.gpo.gov/su_docs/aces/aces140.html.

Harvard Center for Cancer Prevention (1996) Harvard report on cancer prevention. *Cancer Causes and Control*, vol. 7, available at www.hsph.harvard.edu/cancer.

Institute of Shortening and Edible Oils (1999) *Food Fats and Oils*, 8th edn. Washington, DC: ISEO available at www.iseo.org.

Jonnalagadda SS, Mustad VA, Champagne C and Kris-Etherton PM (1995) Margarine and plasma cholesterol – a perspective for dietitians on *trans* fatty acids in the diet. *Perspectives in Applied Nutrition* 2: 9–16.

Krauss RM, Eckel RH, Howard B *et al.* (2000) AHA dietary guidelines: revision 2000: a statement for healthcare professionals from the Nutrition Committee of the American Heart Association. *Circulation* 102: 2296–2311.

Ling WH and Jones PJH (1995) Dietary phytosterols: a review of metabolism, benefits and side effects. *Life Sciences* 57: 195–206.

Piggott JR (2000) Dynamism in flavour science and sensory methodology. *Food Research International* 33: 191–197.

Ratnayake WMN, Hollywood R, O'Grady E and Pelletier G (1993) Fatty acids in some common food items in Canada. *Journal of the American College of Nutrition* 12: 651–660.

Tuorila H, Matuszewska I, Helleman U and Lampi A-M (1989) Sensory and chemical characterization of fats used as spreads on bread. *Food Quality and Preference* 1: 157–162.

Vaisey-Genser M and Vane BK (1995) Sensory evaluation of margarine. In: Warner K and Eskin NAM (eds) *Methods to Assess Quality and Stability of Oils and Fat-Containing Foods*, pp. 76–106. Champaign, IL: AOCS Press.

Van Alphen J (1969) Hippolyte Mège Mouriès. In: van Stuyvenberg JH (ed.) *Margarine, an Economic, Social and Scientific History, 1869–1969*, pp. 5–7. Liverpool, UK: University Press.

MARINE FOODS

Contents

Production and Uses of Marine Algae

P A Cordero, Jr., Sultan Qaboos University, Al-Khod, Omen

Background

The introduction and subsequent improvement of mariculture (i.e., managed cultivation for farming of seaweeds) techniques explains the increased production of certain useful green, brown, and red seaweed species. However, the commercial production of seaweeds greatly depends on the demand for certain products and on the pre- and postharvesting and processing costs. The international market obtains its supply of seaweed materials both from natural vegetation and from sites cultivated by humans.

In addition to consuming seaweeds directly as food, humans have found far-reaching uses for its extracts as food additives and as raw materials for making various household products and other biomedical items.

Species of Importance as Food

Edible seaweeds, or sea vegetables, are mainly composed of green (Chlorophyta), brown (Phaeophyta) and red (Rhodophyta) algae. The predominance of pigments manifested externally has been the basis for their conventional classification. However, in the case of the red alga, e.g., *Eucheuma* growing in the shallower intertidal zone, overlying green coloration is visible owing to the destruction of the characteristic red pigment, phycoerythrin.

Some sea vegetables simulate the gross morphology of terrestrial plants in having leaf-like, stem-like, and root-like parts. The latter part, however, is mainly for anchorage, while the functions of nutrition extraction and the assimilation and conversion of inorganic substances into organic ones are performed by the leaf-like fronds or blades.

The species composition of seaweeds eaten as food varies according to the consumer groups. Thus, the orientals such as the Chinese, Japanese, Malays, and Filipinos include the following species among the common sea vegetables: the green algae, *Enteromorpha*, *Ulva*, *Monostroma*, *Codium*, and *Caulerpa*; the brown algae, *Hydroclathrus*, *Hizikia*, *Cladosiphon*, *Sargassum*, and *Laminaria*; and the red algae, *Porphyra*, *Gelidiella*, *Gracilaria*, *Halymenia*, *Hypnea*, and *Laurencia*. Popular among the occidentals are the green alga, *Ulva*, the brown algae, *Macrocystis*, *Ascophyllum*, *Laminaria*, and *Undaria*, and the red algae, *Porphyra*, *Gelidium*, *Palmaria palmata* (*Rhodymenia palmata*), and *Chondrus crispus*.

In an attempt to popularize the eating of sea vegetables, several seaweed recipes were developed, catering to the South-east Asian cuisine. Of the 80 sea vegetable recipes developed by Cordero and his research team, 63 proved to be acceptable to selected food tasters using the Hedonic Rating Scale, a measurement to determine the acceptability of food, in which a rating of 9 means highly acceptable and 1 least acceptable.

Harvesting and Use of Natural Floras

The methods of harvesting seaweeds have been revolutionized through the years. The more primitive method involved grabbing the seaweed with bare hands, or cutting it with sharp knife or sickle while wading in the shallow intertidal zone. This method, while still considered economical, defies the principle of conservation as the standing crop of wild seaweeds in a given area is easily depleted. An improved version finds a gatherer on-board a paddle-driven banca (boat), later to be substituted with motorized ones, from which the gatherer uses a 'grabber' pole with curved knife on one end to cut the seaweeds from the stem.

In northern Luzon, the Philippines, natural stands of *Porphyra marcosii* and *P. suborbiculata* are scraped from the rough surfaces of infratidal rocky ledges.

Two of the more recent methods of harvesting naturally grown seaweeds are an air-lift vacuum cleaner system, and a mechanized harvester device used in Europe and the USA. The former system was

developed to collect drift *Chondrus crispus* materials, and the latter for harvesting *Macrocystis pyrifera* off the coast of California. The latter device is mounted on a specially crafted boat so that the machine mechanically cuts, gathers, and loads the seaweed materials into the vessel.

Farming Techniques

The history of seaweed farming or mariculture is well recorded among oriental farmers. The technology has advanced considerably from the use of 'Hibi' (see below) to the present culture nets made popular for *Porphyra*, *Monostroma*, and *Ulva*. For *Eucheuma* farming, the use of the monoline system, in which ropes are tied to poles on both ends, has been adopted in Asia.

The use of Hibi was introduced by the Japanese fishermen long before any other farming methods were known. It utilized an artificial substrate composed of cut bundles of bamboo or brushwood, affixed in shallow areas where spores of wild *Porphyra*, once released, became attached to the Hibi. After sometime, the Hibi setups were gathered and transported to a 0.9–1.5 m deep growing area where the spores germinated into sporelings and then into harvestable plants. An innovation of the Hibi method is the use of rope or synthetic twine as an artificial substrate. A rope or synthetic twine is knitted to form a culture net with a 15×15 cm mesh size. In autumn, nets measuring 20–40 m by 1–2 m are set up in the natural habitats of *Porphyra*. The nets are suspended afloat by ropes moored on the sea bottom, or the ends of the nets are tied to wooden poles. *Porphyra* spores, once released, settle on the nets and are allowed to grow until harvest time.

The Taiwanese experience of mariculture adopts the polyculture system, whereby seaweed cuttings, e.g., *Eucheuma* and *Gracilaria*, are stocked in floating and moored cages together with other fish and shellfish stocks.

Other conventional, usually broadcast methods are actually modified systems of vegetative propagation. They can involve tying stocks of *Eucheuma*, *Gracilaria*, or *Gelidiella* to pieces of rock and then randomly broadcasting them in the rocky or corally bottom of the intertidal zone. Similarly, *Caulerpa* cuttings are scattered inside abandoned fishponds and allowed to grow under the prevailing natural conditions as seen in Mactan Island, Central Phillippines.

Marketing and/or Preservation

Seaweeds are marketed either in semiprocessed or processed forms. Fresh materials, after being cleaned of dirt, epiphytes, and invertebrates adhered to or mixed with them, are sun- or kiln-dried, or pickled for export or for household consumption. Those intended for export may be dried, or they may be processed in the form of algal extracts, such as carrageenan, agar, alginate, and algin. These and other phycocolloidal extracts are used as components in the manufacture of industrial products from dairy, pharmaceutical, and surgical items to plastic wares and paints, among others. The papery *Porphyra* and *Monostroma* are dried on bamboo slats so that the seaweeds can be molded into circular and rectangular shapes as sold in Philippine markets. In Japan, the same dried seaweeds are mechanically cut into desired strips and marketed in polythene packs.

In the Americas and parts of Europe, live *Palmaria* and *Chondrus* are impounded for days in specially designed tanks allowing water movement, sufficient nutrient exchange and entrance of sunlight. This method of preservation and storage helps to prolong the palatability of the sea vegetables. These are then dried using a mechanically operated dehydrating device and marketed. Indeed, the seaweed industry, a potential dollar-earning enterprise, faces a bright future. The world market needs a steady supply of seaweeds for its wide spectrum of industrial products. The Third World nations located within the Pacific Basin are favored with a rich variety of tropical seaweed species, which, if judiciously tapped, could plough back much-needed money to improve their economy.

The recently relaxed seaweed importation scheme by the USA, specially for *Eucheuma*, is a boost for the Asian sea-farmers. The Philippines ranks seaweeds such as *Eucheuma* among its top five fishery export items. The high production of the carrageenan-yielding *Eucheuma* in the Philippines is attributed to the simplified and less capital-intensive mariculture method introduced to the Filipino sea-farmers living along the coastal areas of the Visayas and Mindanao Islands. Moreover, this has given birth to the small-scale family business in *Eucheuma* farming. Similarly, Cordero (1982) has found the mariculture of *Porphyra* to be financially feasible for the Philippines, and this author has projected a modest increase in net income of 38% per hectare per year for the first 5 years using *P. marcosii* stock.

Role of Some Species in the Diet of Selected Societies

Edible seaweeds come from the green, brown, and red groups of algae. These protein- and iodine-rich marine plants, found in both cold and warm seas, have been known for their contemporary food

preparations as salad, soup, pickles, jam, etc. The Chinese and Japanese are among the prolific seaweed-eating people, and are accredited with recognizing its food value. The Hawaiians were known to keep 'Limu' gardens containing choice edible seaweeds for the consumption of their noble families.

Green Sea Vegetables

Species of *Caulerpa* and *Codium*, known in the Philippines as 'lato' (Visayan) and 'pok-poklo' (Ilokano), respectively, are the most common edible green algae among the South-east Asians. These, like the other brown and red seaweeds, are simply prepared as a salad with vinegar and a dash of salt. Some people add slices of green unripe mango and crushed tomatoes to enhance the salad preparation.

The low incidence of goiter cases among the coastal dwellers in Asia is attributed to the high iodine intake derived from the edible seaweeds.

Brown Sea Vegetables

Representing the more popular edible brown seaweeds are *Hizikia*, *Hydroclathrus*, *Laminaria*, and *Undaria*. Aside from their food use, the Japanese have made and marketed 'Kombu-Cha' – a kind of tea made out of processed *Laminaria*. The tender tops of *Sargassum* and *Hormophysa* are cooked with vegetables; *Cladosiphon* and *Laminaria* are relished as soup preparations.

Red Sea Vegetables

The red sea vegetables include the greatest number of known edible species. The orientals prefer *Porphyra*, *Gracilaria*, and *Eucheuma*; the occidentals have *Palmaria* or 'Dulse,' *Chondrus* or 'Irish Moss,' and *Porphyra* or 'Laver,' to cite the more popular ones. Their food preparations vary among European and American consumers, although salads, soups, pickles, jellies, etc. are among the favored ones.

It is no wonder that sea vegetable is now known as the poor man's source of cheap protein and iodine, and has an export potential for tropical countries.

See also: **Iodine**: Physiology; **Single-cell Protein**: Algae

Further Reading

Cordero JM (1982) *A Feasibility Study on Sea-farming of the Seaweed* Porphyra. MBA thesis, Philippine Christian University, Graduate School of Business Administration.

Cordero Jr. PA (1980) *Morphological Study, Recipe Preparation and Sensory Evaluation of the Sea Vegetables from the Philippines*. Terminal Report of the National Research Council of the Philippines, Project I-E 98.

Madlener JC (1977) *The Sea Vegetable Book*. New York: Clark N. Potter Inc. Publishers.

Marine Colloids of FMC Corporation (1977) *Carrageenan. Marine Colloids of FMC Corporation. Monograph No. 1*. Rockland, ME: FMC Corporation.

Robertson DL, Smith GJ and Alberte RS (2001) Glutamine synthetase in marine algae: New surprises from an old enzyme. *Journal of Phycology* 37(5): 793–795.

Viejo RM and Aberg P (2001) Effects of density on the vital rates of a modular seaweed. *Marine Ecology-Progress Series* 221: 105–115.

Edible Animals Found in the Sea

R D M Nash and J P Thorpe, University of Liverpool, Isle of Man, UK

Current consumption

In western society the main marine invertebrates consumed are those groups conventionally termed shellfish; that is, various crustaceans (mainly decapods) and bivalve and gastropod molluscs. However other marine invertebrates are widely, if not extensively, consumed in Europe and North America, whilst elsewhere in the world patterns of consumption vary considerably. In various areas, particularly in the tropics, artisanal fisheries for unusual invertebrates can be locally important and in the Far East there is a great demand for food species not commonly eaten in the west. One group in particular, the cephalopod molluscs (octopus, squid, and cuttlefish), are currently important in terms of world catches and also offer potentially highly important fisheries if the technology to exploit them can be developed.

Cephalopods

Of the two extant subclasses of cephalopods (phylum Mollusca, class Cephalopoda), only the Coleoidea are important as food for humans. The other subclass, the Nautiloidea, contains a small number of species, all in the single genus *Nautilus* and all confined to the tropical Indo-Pacific. These support local artisanal food fisheries only, although the very attractive shells are more widely traded. The Coleoidea include the three orders Sepioidea (cuttlefish), Teuthoidea (squids), and Octopoda (octopuses), all of which include species eaten by humans.

Sepioidea (Cuttlefish)

Cuttlefish are almost absent from the Americas, although a few species of *Semirossia* occur along

the west Atlantic coast and are fished off Argentina and in the Gulf of Mexico. Elsewhere both large (mainly *Sepia*) and small (Sepiolidae) cuttlefish are fished in the coastal waters of most temperate and tropical latitudes. Total world annual catch is around 200 000 tonnes, with the most important areas being the eastern central Atlantic (Spain, Morocco), north-west Pacific (Korea, Taiwan, Japan), and western central Pacific (Thailand, Vietnam, Malaysia). Cuttlefish are caught in most areas as a bycatch of other fisheries, but in some areas they are specifically fished with trawls or other nets, dredges, pots, or jigs (lures). The major species fished are *Sepia officionalis* (which occurs down the eastern Atlantic from northern Europe to South Africa and throughout the Mediterranean and Baltic seas), *S. pharoensis* (found from the Red Sea to Japan and south to Australia), *Sepiella inermis* (which occurs from the Persian Gulf to eastern Asia, fished mainly in India and Sri Lanka) and *S. japonica* (Japan and China), but numerous other species are caught in smaller quantities. The tiny *Sepiola rondeleti*, which occurs from northern Europe to west Africa, is extensively eaten in many Mediterranean countries.

In the larger species the muscular mantle and tentacles are mainly eaten, but in many areas sepiolids are consumed whole. Major markets, even for Atlantic- and Mediterranean-caught cuttlefish, are in Japan and Korea and large quantities are exported frozen. In Japan larger cuttlefish are commonly eaten raw, cut into thin strips (sashimi) and smaller animals are packed and frozen ready for cooking. Some species are mainly eaten dried (surume).

Teuthoidea (Squids)

There are two suborders of squids, the Myopsida which are inshore, mainly demersal, squids, generally confined to shallower waters near shores or on continental shelves, and the Oegopsida, which are mainly pelagic offshore squids of deep oceanic waters, although these may come on to the continental shelves to spawn.

All the myopsid squids currently eaten are confined to the family Loliginidae and most are from the very large genus *Loligo*, species of which are found on continental shelves worldwide. Inaccurate and inadequate statistical data make it difficult to assess annual world catches, but the total for all loliginid species may be around 200 000 tonnes. Major fishery species are *Loligo bleekeri* (fished in the Sea of Japan), *L. chinensis* (which occurs from China south to eastern Australia), *L. duvauceli* (fished extensively in coastal waters throughout most of Asia and north-east Africa), *L. edulis* (China to northern Australia), *L. forbesi* (coastal waters from Sweden to Senegal, Mediterranean, Red Sea, south-east Africa), *L. gahi* (southern South America, Falkland Islands), *L. japonica* (Japan, China), *L. opalescens* (Pacific coast of North America), *L. pealii* (Newfoundland to northern Brazil), *L. reynaudi* (South Africa), *L. vulgaris* (northern Europe to South Africa, Mediterranean), *Sepioteuthis lessoniana* (Red Sea to Japan and Australia), *Alloteuthis media* (occurs from the UK to North Africa and Mediterranean, trawled in western Mediterranean). Capture methods for loliginid squids are mainly with nets (purse seines, otter trawls, pair trawls) and at night with lights. Some occur as bycatches in finfish fisheries.

Oceanic (oegopsid) squids form well over half of the total world cephalopod catch, with an annual total of perhaps 1 500 000 tonnes. A few species are demersal at least at some stage of the life cycle, but most are found in the water column, although this may be at great depths. For obvious reasons the species presently exploited commercially are generally only those occurring comparatively close to the surface. Currently only about nine (Enoploteuthidae, Octopoteuthidae, Onychoteuthidae, Gonatidae, Psychroteuthidae, Lepidoteuthidae, Histioteuthidae, Ommastrephidae, Thysanoteuthidae) of 32 recognized families have any species fished by humans, although many others have potential as food. Exploited species of major commercial significance are confined to only four families:

1. Enoploteuthidae, *Watasenia scintillans* (China, Japan).
2. Onychoteuthidae, *Onychoteuthis borealijaponica* (occurs in the northern Pacific from Japan and USA to about 55 °N, fished off Japan and north-west USA).
3. Gonatidae, *Berryteuthis magister* (found in the northern Pacific, from Japan to Canada, fished off Japan and north-eastern Russia).
4. Ommastrephidae *Illex argentinus* (south-eastern South America, Falkland Islands), *I. coindetii* (eastern Atlantic, Sweden to South Africa, and Mediterranean: fished in the western Mediterranean and Atlantic, off Spain and North Africa), *I. illecebrosus* (north Atlantic from USA to northern Europe, north to Arctic: fished north-west Atlantic), *Todaropsis eblanae* (continental shelves Europe, Africa, Asia to Australia: fished Mediterranean and north-west Africa), *Todarodes pacificus* (northern Pacific, China to Alaska, fished by Japanese), *T. sagittatus* (eastern half of Atlantic north of 20 °S to Greenland, Mediterranean, Arctic Ocean from Greenland east to northern central Russia: fished Norway, Italy), *Nototodarus gouldi* (Australia), *N. sloani* (New Zealand),

Ommastrephes bartrami (circumglobal, subtropical to temperate, northern and southern hemispheres, fished northern and southern Pacific), *Dosidicus gigas* (eastern Pacific, Chile to Mexico).

The exploitation of oegopsid squid is largely to satisfy the demands of the Japanese and, to a lesser extent, the Taiwanese market. Thus several of the major fishery species are those occurring in the northern Pacific, whilst others (for example *Nototodarus gouldi* off Australia, *N. sloani* off New Zealand and *I. argentinus* off the Falkland Islands) are exploited largely or almost entirely by Japanese, Korean, and Taiwanese vessels. Capture of oceanic squids is mainly by special ships equipped for jigging; squid are attracted towards the surface at night by the use of bright lights and are caught on barbless hooks attached to lures. The whole process of capture, unhooking, and processing and freezing of the catch is automated. In a few fisheries some of the catch is taken by drift nets in a manner similar to the gill netting of fish.

In Japan the way in which squids are sold varies between species, but the main methods are fresh and raw as sashimi, cooked and processed as saliika, dried as surume, canned, and frozen.

Octopoda (Octopuses)

Only octopods of the suborder Incirrata are fished commercially. Of these, only one (Octopodidae) of the eight families is at all important, although small numbers of the circumglobally distributed *Argonauta argo* (Argonautidae) are occasionally marketed from bycatches. Total world annual landings of octopods probably vary around 200 000 tonnes. All the main species caught are in the widely distributed genus *Octopus: O. briareus* (Caribbean and northern South America), *O. conispadiceus* (Japan), *O. cyaneus* (widespread from East Africa to India and Australia, throughout the Indian Ocean and western and central Pacific), *O. dofleini* (coastal from China to south-western USA), *O. globosus* (India), *O. maya* (Gulf of Mexico), *O. membranaceus* (coastal waters from India to Japan and Australia, important fishery species in Japan and China), *O. variabilis* (Japan, China), *O. vulgaris* (coastal distribution from northern Europe around Africa, including Mediterranean and Red Sea, and Asia east to Japan, also western Atlantic from USA to Brazil: fished over most of this huge range; of major importance in Japan, north-west Africa).

As with other cephalopods, much of the world market for octopus species is in Japan, where they are highly esteemed and are marketed fresh, frozen, dried, and/or salted.

Barnacles (Crustacea, Cirripedia)

There are a variety of barnacles exploited and eaten throughout the world. The large stalked barnacle *Polliceps polliceps* is eaten particularly by the Basque people, but is also consumed throughout Iberia and the Atlantic Island groups (Canaries, Madeira, and Azores). The acorn barnacles (*Megabalanus* spp.) are consumed in Chile (*M. psittacus*) and in the Azores (*M. tintabulum*). The goose barnacle *Mytella mytella* is eaten in Japan.

Jellyfish (Scyphozoa)

At least three species of jellyfish, of which *Rhopilema esculenta* is the most important, are consumed in Japan. These are usually served raw in a vinaigrette or similar dressing.

Sea Anemones (Anthozoa)

A variety of species are eaten in various countries, including Japan, Samoa, and France. The intertidal *Actinia equina* is eaten in Italy. All are usually eaten fried.

Tunicates (Urochordata, Ascidiacea)

The Japanese eat at least two species of sea squirts (*Halocynthis roretzi* and *H. aurantium*) and the French primarily utilize just one species (*Microcosmus claudicans*). All three species are from the same family (Cynthiidae).

Sea Urchins and Sea Cucumbers (Echinodermata)

Sea urchins are increasingly becoming subject to commercial fisheries, primarily to satisfy the Japanese market. Two species (*Paracentrotus lividus* and *Echinus esculentus*) are traditionally eaten in some Mediterranean countries and on the west coast of Ireland. Similar species are eaten in the Far East, particularly Japan. Sea urchins are also eaten by Maoris in New Zealand, who eat the complete contents out of the shell. Some North Atlantic species are exported to Japan and Japanese boats fish sea urchins commercially in Australian waters. In general only the roe (gonad) is eaten and this is consumed raw. In Japan the roe is only considered top-grade if it is a pale yellow color and the skeins are unbroken. Darker brown or damaged roes do not command such a high price. For prime roes prices are very high, around £30 per kg.

The market for sea cucumbers is also mainly in Japan. The very large black species *Stichopus japonicus* is taken locally, whilst several other species are

exported to Japan from various Pacific Islands (where they are locally known as bêche de mer).

Bristle Worms (Polychaeta)

The main type of polychete consumed is what are known as palolo worms, several species of which occur in various parts of the world. A common feature of these worms is that they swarm to the surface in very large numbers at predictable times of the year. The large swarms are primarily for reproduction, and are of a stage in the life cycle known as a sexual epitoke. A number of species are eaten: the Pacific palolo (*Eunice viridis*) from the areas surrounding Fiji and Samoa, wawo (*Lysidice oele*) from the Ambonia area, and the Atlantic palolo (*Eunice schemacephala*) in the Dry Tortugas. The Japanese palolo (*Tylorrhynchus heterochaetus*) is known by a number of different names in Japan.

Other Exploited Marine Invertebrates

Many more species and other marine invertebrate groups are, or are likely to be, eaten or otherwise exploited on a limited scale somewhere in the world, but to find and list them all is beyond the scope of this article. Those outlined above give some indication of the diversity of marine invertebrates consumed by humans.

However, one further area which should be mentioned is the recent and expanding exploitation of species of certain phyla (e.g., sponges, bryozoans, coelenterates, ascidians, nemerteans) which contain allelopathic (usually antipredator or anticompetitor) or otherwise 'useful' biologically active chemicals of value to the pharmaceutical industry. Many of these biologically active natural products have been studied and a number are likely to become medically and commercially important. One such product is what has been termed bryostatin, a biologically active compound, which has been found to have a potent action against various types of cancer cells. Bryostatin is obtained from the North American bryozoan *Bugula neritina*, but huge quantities of the bryozoan are needed; for example, one group studying this compound based its synthesis on a sample of 10 000 gallons (about 45 000 kg) of the bryozoan. With collection on such a large scale, any small invertebrate in pharmaceutical demand may become overexploited unless aquaculture can take over.

Possibilities for Future Exploitation of Marine Invertebrates as Food

Many of the world's conventional shellfish fisheries, like those for finfish, are being overfished and are in decline. It is unlikely therefore that there is much potential for future expansion of fisheries for decapod crustaceans or shelled molluscs, but as fish stocks, which provide the great bulk of wild caught marine animals, appear certain to continue to decline, other resources are needed.

Future demand for other invertebrate food species will presumably depend partly on the success of aquaculture in providing fish and shellfish and whether untapped sources (e.g., deep-sea species) can be exploited. Marine invertebrates of many groups are eaten little or not at all in most parts of the world. Thus there is apparent scope for the development or expansion of fisheries. However, most of the species within these groups are either inedible or present in only relatively small biomasses and therefore have little potential. Abundant marine invertebrates for future exploitation could be Antarctic krill (mainly *Euphausia superba*) and oceanic squid, particularly some of the deeper-living species.

Krill stocks are thought to be unnaturally high, because the near elimination of their major predators, the large whales, and the huge biomasses present in the Antarctic Ocean offer a potentially very great fishery resource, although of low value and palatability. They have been used for human consumption in Russia, but their commercial exploitation to date has been mainly as fish meal. Large-scale exploitation of krill could lead to ecological problems because they are generally considered the 'cornerstone' species for the whole vast Antarctic marine ecosystem and hence major exploitation could be serious, and overfishing a disaster.

Currently cephalopods and in particular squids make a useful contribution (about 2–3%) to fisheries on a global scale and in some areas (e.g., Japan) they are of great importance. Squids collectively rate as the sixth most commercially important fish species in world fishery statistics. Their significance as food is, for various reasons, greater than catch figures would suggest. They have little skeletal tissue and are highly muscular and hence, compared to most fish species, a relatively greater proportion of the catch is edible. Several of the fish species making the greatest contribution to world catches are used for fish meal (for fertilizer or animal feeds) and thus are not directly eaten by humans. Consequently the amount of fish eaten is considerably less than that caught. However, from the human viewpoint, the most important feature of cephalopods is probably their potential to increase world food supply.

Currently total world landings of fish are approximately static, at about 80 million metric tonnes (Mmt), despite continual increases in fishing efficiency and effort. This is occurring because many

fish stocks are being overfished and are declining, and overall the prospects for improvement do not look good. Since about 1970 world catches of cephalopods (mainly squids) have increased annually by an average of about 6% and even among coastal octopus species (probably the group with the least capacity for expansion) there is considered to be the potential to increase world catches greatly. Among oceanic squids the potential is enormous. Estimates of stock sizes are necessarily very approximate because there has been little or no relevant scientific study of the stocks of most squid species. Indeed, it is likely that there are still many species which are as yet unknown.

Evidence pointing to massive untapped squid populations comes from several sources. For a few species there are adequate scientific data for stocks to be estimated. For example, the large (up to 150 kg) cranchid squid *Mesonychoteuthis hamiltoni* is abundant in Antarctic waters and the flesh is considered to be of excellent quality and flavor. This species is not caught commercially at all, but the stock is estimated at about 90 Mmt (i.e., larger than the total annual catch of fish by humans). A second example is the very large species *Architeuthis* (Architeuthidae) which is thought to be distributed throughout all the oceans apart from continental shelves and close to the poles. Very few have been caught by humans, but its frequent consumption by whales (identified mainly from the beaks, which remain undigested in whale stomachs) and occasional stranding in shallow water suggest it is not rare. As individuals weigh up to about 1000 kg, the total biomass is likely to be huge. Current direct estimates by fishery biologists of total world squid stocks vary considerably, up to about 500 Mmt, but indirect estimates from other sources indicate that even this figure may be conservative.

Studies of the squid consumed by various predators suggest very large populations. In the Antarctic, penguins and elephant seals, for example, are thought to take annually about 13 and 4.5 Mmt respectively (neither of these, incidentally, eats any of the estimated 90 Mmt of *Mesonychoteuthis hamiltoni*). Sperm whales eat mainly squid found on or near the bottom in depths restricted by their diving capabilities. Sperm whale populations have been seriously depleted by overexploitation, but even so they probably eat annually a weight of squid equivalent to about twice the entire world catch of fish by humans.

From recent research it is also becoming clear that there are great biomasses of largely unknown squid at mesopelagic and bathypelagic depths throughout much of the world's oceans. Over 70% of the world surface consists of sea deeper then 2000 m and by volume more than 98% of the sea is off the continental shelves, with an average depth of several thousand meters. Over many years the work of research ships towing midwater trawls at these depths had indicated that throughout most of this vast volume of water, animal life was rather scarce. However, recent studies using television cameras attached to the net frames reveal considerable numbers of often quite large (up to tens of kilos) squids present at a wide range of depths and over huge areas. These are not generally caught simply because, as large and highly mobile visual predators, they easily avoid nets. Since they are distributed through such huge volumes of the deep ocean the biomass of these squid is likely to be immense.

Thus various sources of evidence indicate that oceanic squids may be present in biomasses which dwarf known stocks of exploited fish species. Since most squid are thought to be annual and semelparous (they grow for a year, spawn, and die) they 'turn over' faster than most fish species and thus, if well managed, should be able to sustain higher relative rates of exploitation. If fished in the later stages of the annual life cycle and as long as enough adults survive to spawn, the entire stock can be exploited because there are no juveniles which need to be preserved for future breeding. However, for many of the oceanic squids, the depths at which they live present major barriers to the development of commercial exploitation. Our knowledge of their biology and behavior is at best minimal and technological advances will be necessary before most can be utilized.

Marine Reptiles as Food

Marine reptiles (turtles, saltwater crocodiles, iguanas, and sea snakes) occur in many areas of the tropics and subtropics and various species are exploited on a small scale by local human populations. Turtles were formerly exported on some scale for consumption, particularly in Europe, but many of the wild populations are now endangered and their import is banned in most countries. Most species are, however, good to eat and are still netted or harpooned on a small scale by native fishermen. Turtle eggs are also considered a delicacy in many parts of the world. Saltwater crocodiles can grow to a length of about 10 m and are the world's largest living reptiles. These are occasionally killed in the Far East, although their food value is considerably less than that of the hide.

There are six widespread species of marine turtle:, loggerhead (*Caretta caretta*), leatherback (*Dermochelys coriacea*), hawksbill (*Eretmochelys imbricata*), Kemp's Ridley (*Lepidochelys kempii*), olive Ridley (*Lepidochelys olivacea*), and the green turtle (*Chelonia mydas*). The only one currently commercially exploited on a large scale for food is the green turtle. This is a medium-sized species growing up to about 1 m and

occurs naturally in the tropical Atlantic, Caribbean, and elsewhere. The species has been severely overexploited and is now scarce and heavily protected in the wild, but is the subject of a large aquaculture enterprise in the Caribbean. Captive broodstock adults, kept in large saltwater ponds, lay their eggs 'naturally' on an artificial sandy beach and the eggs are then dug up and incubated artificially. This method is highly successful, with nearly 90% survival of the newly hatched young (compared with about 0.2% in the wild). The young turtles are then reared in large tanks of sea water for 3 years before harvesting. The meat is highly prized and has almost no fat content, but more protein than prime beef or chicken.

See also: **Shellfish**: Characteristics of Crustacea; Characteristics of Molluscs

Further Reading

Caddy JF (1983) *Advances in Assessment of World Cephalopod Resources*. FAO Fisheries Technical paper no. 231. Rome: FAO.

Clarke MR (1987) Cephalopod biomass – estimation from predation. In: Boyle PR (ed.) *Cephalopod Life Cycles*, pp. 221–237. London: Academic Press.

FAO (1988) *Yearbook of Fisheries Statistics*. Rome: FAO.

FAO (1990) *Review of the State of World Fishery Resources*. FAO Fishery circular no. 710 (revision 7). Rome: FAO, Fishery Resources and Environment Division, Marine Resources Service.

Lipinski MR (1998) Cephalopod life cycles: patterns and exceptions. *South African Journal of Marine Science* 20: 439–447.

Lipinski MR *et al.* (1998) Cephalopod fisheries. *South African Journal of Marine Science* 20: 463–469.

Roper CFE, Sweeney MJ and Nauen CE (1984) Cephalopods of the world. An annotated and illustrated catalogue of species of interest to fisheries. *FAO Fishery Synopsis* 125(3): 1–277.

Marine Mammals as Meat Sources

D Bloch, Museum of Natural History, Faroe Islands
M Dam, Food and Environmental Agency, Faroe Islands
M Hanusardóttir, Home Rule Government, Faroe Islands

Introduction

From the earliest times, humans have, wherever possible, included a wide range of wild terrestrial and aquatic animals in their diet as sources of animal protein and fat. Seals and whales have long constituted a natural part of the diet of the inhabitants in those regions of the world where these marine resources occur.

For example, in the Faroe Islands, a tiny archipelago in the eastern north Atlantic, the economy is dependent almost exclusively on marine resources. The long-finned pilot whale, *Globicephala melas*, and other whales have been caught for food in the Faroes ever since people first settled in the islands (around the year 675). Detailed statistical records of whale hunts date back to 1584, and from 1709 to the present these are unbroken. For this reason the following description of marine mammals as a food source is based largely on the example of the long-finned pilot whale off the Faroe Islands. In addition to these historical records, from 1986 to 1988 an international research project examined the long-finned pilot whale off the Faroe Islands in detail, and this research resulted in studies of, among other things, the percentage of whales used for human consumption, as well as their content of nutrients and contaminants. An earlier study examined the significance of the pilot whale in the everyday diet of Faroe Islanders.

The use of seals is also significant for human consumption in many coastal communities.

Classification of Whales

Whales, which are cetaceans, are divided into two main groups – baleen whales (Mysticeti), and toothed whales (Odontoceti).

According to the North Atlantic Marine Mammal Commission (NAMMCO) and the International Whaling Commission (IWC), the following baleen whales are currently used for human consumption: bowhead, gray, fin, humpback, and minke whales. Of the toothed whales, the following are used for human consumption: Baird's beaked whale, narwhal, white whale, false killer whale, long-finned pilot whale, short-finned pilot whale, Atlantic white-sided dolphin, bottlenose dolphin, pantropical spotted dolphin, striped dolphin, common dolphin, Risso's dolphin, harbour porpoise, and Dall's porpoise.

Annual catches are listed by NAMMCO and the IWC and published in their annual reports; the number taken in 1997, directly and indirectly as bycatches, was around 33 500 whales worldwide.

Hunting Methods and the Use of Whales

The ban on the use of the cold grenade (i.e., nonexplosive grenade) harpoon in commercial whaling was

extended to include minke whales from the beginning of the 1982–83 pelagic and 1983 coastal seasons.

In the fishery of small cetaceans, harpoons, rifles, nets, and driving methods are used. In the drive fishery in the Faroe Islands, entire schools are driven into an authorized whaling bay and killed by severing the jugulars and arteries supporting the brain with special knives. All stages in this organized hunt are regulated by law. It is worth mentioning that there are no professional whalers or whaleboats involved in this whale drive, and the shares are distributed free to all local inhabitants. No factory processing of the catch occurs. For indepth descriptions of the various stages of driving, beaching, killing, and butchering, see Further Reading.

The size of the driven schools varies and it can be seen from the statistics that pilot whales occur in Faroese waters with a cyclic periodicity of about 120 years. In the period from 1709 to 1999, the school size was on average 139.8 whales (range 1–1200), with an average of 7.1 schools beached a year (range 0–23), giving an annual average catch of 850 whales (range 0–4448).

Regulations state that pilot whales must be butchered and the shares removed within 1 h after the shares have been allotted. According to the regulations, the shares are given to the participants in the drive and to local households in the district.

The whales are butchered on the quay by those receiving shares. The section of the whale which is used is cut behind the head to the anus. First the blubber is peeled off in transversal strips and placed on the quay with the skin down so as not to soil the blubber side. After that, the meat is removed and placed on top of the blubber, and when the whole whale has been butchered in this way, the shares are then divided between those who have a share in the same whale.

Food Uses and Products: Significance in National Diets

The different parts of the whale used for human consumption in different parts of the world are listed in Table 1. In general, the meat is used by all the whaling countries, either raw or, more commonly, prepared in some way. Some of these countries, where edible fat has been in short supply, have also developed national dishes using the blubber in different ways. Finally, especially in Japan, national dishes have been developed for nearly all parts of the whales, except the backbone.

In the Faroes both the meat and blubber of pilot whales are utilized for human consumption. On average, 54% in weight of the animals is used, with the amount of meat and blubber used being nearly equal. On the beach, kidneys are taken, free, as anyone's share. The kidney is boiled fresh in salted water and eaten at once. The share of meat and blubber is prepared in different ways in the home. The traditional and still the most common method for conserving the household share is by salting or wind-drying. In the last 25 years this method has been supplemented by deep-freezing the products fresh or after some wind-drying. Normally, however, very little of the meat and blubber is frozen fresh, partly owing to limited freezer space, and partly because the traditional methods are thought to provide the most delicious result.

The different methods used to prepare the whale products are discussed in the following two sections.

Table 1 Preparation methods of the various parts of whales used today for human consumption

Preparation method	*Part of whale*	*Country*
Raw	Meat	Japan[a]
	Gum meat	Japan
	Lung	Japan
	Skin with blubber	Greenland[b]
Boiled or fried in oil or water	Meat	Worldwide use
	Blubber	Japan, Faroe Islands[c], Greenland
	Tongue	Japan
	Ventral groove	Japan
	Kidney	Faroe Islands
	Internal organs	Japan
	Lung	Japan
Smoked	Ventral groove	Japan
	Meat	Japan
Salted dry or pickled	Meat	Japan, Faroe Islands
	Blubber	Faroe Islands
	Fin	Japan
	Ventral groove	Japan
Dried by sun/wind	Meat	Japan, Faroe Islands
Lactic acid	Meat	Iceland[d]
	Ventral groove	Iceland
Wine-preserved	Cartilage	Japan

[a]Data from Small-Type Coastal Whaling in Japan (1988) *Report of an International Workshop*. Occasional publication number 27. Alberta, Canada: Japan Social Sciences Association of Canada Fund to Promote International Educational Exchange and Boreal Institute for Northern Studies.

[b]Data from Hansen C (1963) *Nerissagssiornermik ilitsers«tit kalatdlit-nunnut t°ngatitat*. Den Grønlandske Handel. (Greenlandic cooking book.) Nakitertitai.

[c]Data from Skaale and Johannesen (1974) and Henriksen (1987): Skaale Ó and Johannesen M (1974) *Matur og Matger*ð. Grønalíð. Tórshavn. (Faroese cookbook.). Henriksen J (1987) *Gœða rÃ*ð. Góða Ráð. Klaksvík. (Faroese cookbook.).

[d]Data from Kristjánson (1986) Kristjánsson L (1986) *N»ting Hvalafangs*. In: *ëslenskir SjÂvarhættir*, vol. 5, pp.70–79. Reykjavík: Bókaútgáfa Menningarsjóðs.

Whale Meat

In general, most of the whale meat is eaten fresh (worldwide), salted, wind-dried (Faroe Islands), or sun-dried after being thinly sliced (Japan). In Iceland, the ventral grooves and meat of baleen whales were conserved in containers with lactic acid. (*See* **Drying**: Drying Using Natural Radiation.)

In the salting process, pieces of meat or blubber are placed in dark containers, pressed together, and covered in a brine of boiled water with a salt solution in which a potato can float. After salting for at least 1 month, the meat is washed in water for 24–48 h before being boiled in water, and the blubber is eaten boiled or raw, unwashed (Iceland and the Faroe Islands).

When drying, whale meat undergoes a process of fermentation combined with drying (the Faroe Islands). The pieces of meat are sprinkled with salt (12–24 h) after being cut into thinner pieces of about 50 cm in length and 10 cm in diameter. The surface is smoothed, resulting in a quick stiffening of the surface in the wind which prevents flies from laying eggs. The pieces are hung up outside on the northern side of the house, and are ready for boiling in salted water after 3–4 weeks. Alternatively, when dried quite hard after about 3 months it is eaten raw in thin slices. (*See* **Preservation of Food**.)

The above-mentioned means of food preservation is only possible in the north Atlantic where the climatic conditions are cold and windy. Moreover, in the Faroes it is a means still commonly used for preserving local sheep, fish, and geese.

Whale meat and blubber are also smoked in both Iceland and Japan.

Blubber

Very little blubber is frozen fresh. Nearly everything is salted, either dry or in brine, as described above (the Faroe Islands, Iceland) or wind-dried (Iceland). After lying in salt for at least 1 month, preferably up to 3 months, the blubber is edible and conserved for years. The skin and outermost layer of blubber is eaten raw (Greenland), and elsewhere the blubber is boiled or fried (Japan, the Faroe Islands, Greenland).

Significance in National Diet

For many years, pilot whaling has been essential to the survival of the Faroese community. In modern Faroese society an annual catch of approximately 2500 whales still represents 50% of the Faroese meat production (equal to sheep and cow production) and 25% of the Faroese meat consumption.

In former times, and still today, salted whale meat and blubber has been used as food on board ships, especially on smaller boats without freezers (the Faroe Islands, Japan).

Furthermore, marine mammals are virtually the only reserve for some Inuit communities in the Arctic.

Chemical Composition and Nutritional Value

Whale meat and blubber, most commonly used for human consumption (Table 1), have a very high nutritional value. The meat is an excellent source of protein (Table 2), with a low fat content. Thus it agrees well with western health recommendations. (Refer to individual nutrients.)

Table 2 The chemical composition of whale meat, expressed per 100 g of meat

Component[a]	*n (number of samples)*	*Mean value*	*Standard deviation*	*Species*
Protein (g)	20	25.0	0.7	Pilot whale
Fat (g)	20	1.0	1.4	Pilot whale
Carbohydrate (g)	20	0.5	0.6	Pilot whale
Sodium (g)	10	330	160	Pilot whale
Potassium (mg)	10	1630	350	Pilot whale
Calcium (mg)	10	10	0	Pilot whale
Magnesium (mg)	10	110	40	Pilot whale
Phosphorus (mg)	10	930	100	Pilot whale
Iron (mg)	10	51.9	19.4	Pilot whale
Zinc (mg)	10	6.7	1.0	Pilot whale
Copper (mg)	10	0.14	0.05	Pilot whale
Selenium (mg)	3	0.19	0.06	Pilot whale
Vitamin B_1, thiamin (μg)		200		Narwhal
Vitamin B_2, riboflavin (μg)		300		Narwhal
Vitamin A, retinol (μg)		100		Narwhal
Vitamin C, ascorbic acid (μg)		1		Narwhal

[a]Data for protein-selenium from Poulsen (1995). Data for vitamins B and C from Jepsen (1983): Jepsen O (1983) *Kostvurderinger over Grønlandske fødevarer*. FDB's centrallaboratorium. Copenhagen: Forenede Danske Brugsbutikker's Centrallab.

Whale Meat

Compared with other mammals and various fish species, whale meat has a very high content of practically all types of macro- and microminerals. Of particular note is the high content of selenium and iron (Table 2). Selenium has an important role because it seems to prevent carcinogenic reactions, and iron is required to meet the needs for growth, maintenance, and repair of tissues. The estimated safe and adequate daily dietary intake of selenium for adults is 0.05–0.2 mg, and the recommended dietary allowances (RDA) for iron for adults is 18 mg. Thus 100 g of pilot whale meat a day will provide enough of these two minerals. (*See* **Fish Oils**: Dietary Importance; **Fish**: Dietary Importance of Fish and Shellfish; **Meat**: Nutritional Value; **Offal**: Types of Offal.)

Little information is available about the content of different vitamins in pilot whales, but the meat from narwhals is a good source of vitamins B_1, B_2, and A.

Blubber

Blubber is composed of practically pure fat (70–80%) and water. However, the fatty acid composition is very good (Table 3). Blubber has a high ratio of polyunsaturated to saturated fatty acids (P:S value), and the amount of eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) is high, compared with various fish species, which are the main natural sources of these fatty acids. EPA and DHA are probably the most important fatty acids in preventing coronary heart disease. (*See* **Fatty Acids**: Dietary Importance.)

Table 3 Fatty acid composition in blubber from long-finned pilot whales off the Faroe Islands

	Fatty acid content (%)
Saturated fatty acids (S)[a]	16.34
Monounsaturated fatty acids	60.45
Polyunsaturated fatty acids (P)[a]	14.90
Polyunsaturated fatty acid fraction	
Eicosapentaenoic acid (EPA)	1.20
Docosahexaenoic acid (DHA)	5.41

[a]P/S is 0.91.

Data from Poulsen (1995): Poulsen M (1995) øFroyskar Fødslutalvur (Faroese Food Composition Tables.) Heilsufrøðiliga Starvstovan, pp.1–25. Tœrshavn: Food and Environmental Agency.

Chemical and Microbiological Hazards

Despite the fact that whales are of extremely high nutritional value, there are reasons for caution. The meat can contain large amounts of heavy metals, such as mercury and cadmium, and the blubber can be contaminated with persistent, lipid-soluble synthetic compounds such as pesticides and polychlorinated biphenyls (PCBs) and halogenated aromatic hydrocarbons in general. It is sometimes found that individuals carry a muscle mercury concentration of $6\,mg\,kg^{-1}$. The normal level is however one-third of this – about $2\,mg\,kg^{-1}$ (Table 4). The proportion of the total mercury which is present as the most potent toxic species methylmercury is highly variable among the individuals in the school, varying from 24% to 86% in one school in 1978; the average methylmercury component in samples from two schools were however approximately 50%. Also cadmium may pose a threat to both individual whales and also to consumers of pilot-whale livers and kidneys. The concentration of the metals in the liver are severalfold higher than in the muscle where in individual instances concentrations approaching $250\,mg\,kg^{-1}$ of cadmium and $600\,mg\,kg^{-1}$ of mercury have been detected. The content of mercury in different parts of the pilot whale (Table 4) and examinations of Faroe Islanders indicate that the estimated long-term intake of mercury and cadmium from eating whales may reach or exceed the recognized critical level for negative or toxic effects in humans. (*See* **Cadmium**: Toxicology; **Mercury**: Toxicology.)

Inspecting the mean values of total dichlorodiphenyltrichloroethane (DDT) and PCB in blubber from pilot whale schools in the Faroes in 1987 and 1997 (Table 5) reveals the frequently observed

Table 4 The total mercury concentration in various tissues of long-finned pilot whales off the Faroe Islands

Year of sampling	*Tissue*	*Number of samples*	*Mean value (mg kg^{-1})*	*Standard deviation*	*Reference (see Further Reading)*
1977	Meat	20	2.1	1.2	Bloch *et al*. (1987)
1987	Meat	31	1.35	1.78	Caurant *et al*. (1993)
1997	Meat	28[a]	1.7		Food and Environmental Agency (1976, 1978)
1987	Kidney	31	4.91	3.81	Caurant *et al*. (1993)
1987	Liver	40	84.1	92	Caurant *et al*. (1993)
1978	Blubber	20	0.69		Food and Environmental Agency (1976, 1978)

[a]Three (in one instance four) pools from nine schools representing samples from a total of 417 individuals were analyzed. The mean values have been adjusted to represent an 'average' school with respect to number of females and males, juveniles and adults.

Table 5 Organochlorine concentrations in the blubber of long-finned pilot whales off the Faroe Islands

Year of sampling	*Sex*	*Body length (cm)*	*Standard deviation*	n *whales*	n *pods*	*tDDT*[a] *mg/kg lipid*		*PCB as Aroclor 1254:1260 or 1260 (mg kg^{-1} lipid)*		*CB 153 (mg kg^{-1} lipid)*	
						Mean	*Standard deviation: range*	*Mean*	*Standard deviation: range*	*Mean*	*Range*
1987	im ♀♀	304	50	44	NA	37.2	26.9	47.8	29.5		
1987	im ♂♂	362	68	15	NA	49.1	32.5	61.4	27.3		
1987	ad ♀♀	431	28	113	NA	9.2	9.1	15.5	12.4		
1987	ad ♂♂	546	48	11	NA	50.9	20.7	57.4	24.6		
1997	im ♂♂♀♀	356	85	173	10	15.9	10.6–23.5	35.1	18.4–45.0	3.7	2.0–4.7
1997	ad ♀♀	430	22	193	9	6.3	3.3–11.7	16.0	11.0–29.0	1.8	1.2–3.1
1997	ad ♂♂	547	25	54	9	17.7	10.6–25.4	38.2	24.5–52.0	4.1	2.6–5.7

[a]tDDt = *p,p′*-DDT + *p,p′*-DDE + *p,p′*-DDD + *o,p′*-DDT.

Polychlorinated biphenyls (PCB) are given with reference to the commercial product; the 1987 data were quantified with a 1:1 mixture of Aroclor 1254 and 1260. The 1997 data were quantified as single congeners; however, for reasons of comparision with the 1987 data, these were expressed as Aroclor 1260 units. The calculation of Aroclor 1260 units is based on regression analysis in human samples which had been analyzed for PCBs both with an Aroclor standard and with a single congener standard. For reasons of

feature of mature females to excrete persistent organic pollutants and thus carry a lower body burden than immature whales and mature males. A recent study of the group of polybrominated diethylethers (PBDE), also known as flame retardants, in blubber samples revealed a widespread distribution of these compounds also in pilot whales, and the same pattern of elevated concentrations in the young as is seen with PCBs and pesticides in general. The concentrations of PBDE as well as PCBs, both the nonortho congeners which are also known as the coplanar ones, as well as the bulk of PCBs, are high (Table 6). In addition, the concentration of polychlorinated dibenzo-*p*-dioxines and -*p*-furanes, often referred to simply as dioxins, are given for the adult whales sampled in Vestmanna in 1996. Due to the scarcity of data, the distribution of contaminants between the groups of adults in this particular school should not be considered as typical. What can cause changes in the relative distribution is that the adult males are more transient members of the school, and the males and females need therefore not have common history regarding food and foraging area, which are both important parameters in determining pollutant body burden. What may be compared, regardless of the uncertainty between the sexes, is the distribution of pollutants within the groups. We note that the common PCBs, although each one less toxic than the coplanar ones and the dioxins in general, are still responsible for the highest toxic load, expressed in toxic equivalents, and this is due to the much higher concentrations. The toxic equivalents are defined with reference to the most toxic dioxin compound, 2,3,7,8-TCDD. It is apparent that although this seveso-dioxin is assigned a toxic equivalent value of 1, the 2,3,7,8-TCDD is only accounting for less than 10% of the toxic effect

Table 6 Concentrations of polybrominated diethylethers (BDE), polychlorinated biphenyls (PCB), also of the non-ortho congeners, polychlorodibenzo-*para*-dioxin (PCDD), and polychlorodibenzo-*para*-furanes (PBDF) in pilot whale blubber

Lipid %	*ad ♀♀ 79 % (ng g^{-1} lipid)*	*ad ♂♂ 83 % (ng g^{-1} lipid)*
Te-BDE #47	529	862
Pe-BDE #99	209	292
Sum BDE[a]	1.048	1.610
CB 153	1.500	2.700
Sum PCB[b]	10.672	21.785
	(pg g^{-1} lipid)	*(pg g^{-1} lipid)*
PCB 77	770	420
PCB 126	340	310
PCB 169	620	530
TEQ sum coplanar PCB[c]	40.6	36.5
TEQ sum PCB[b]	180	390
TEQ sum PCDD/PCDF	12.9	10.0
TEQ, 2,3,7,8-TCDD	0.86	0.58

[a]Sum BDE is the sum of 19 individual tetra, penta and hexa bromodiethylethers.

[b]Sum PCB is the sum of 61 congeners quantified using single congener standards.

[c]Sum of the three nonortho-substituted PCB congeners CB 77, 126, and 169.

The samples were taken in Vestmanna 27 June 1996; the analyses were done on pooled samples of adult females (ad ♀♀) and males (ad ♂♂), each representing 19 and 8 animals respectively. The results are given in units of ng g^{-1} lipid and pg g^{-1} lipid, depending on the range of concentrations detected. The chlorinated compounds are also presented in toxic equivalents (TEQs), using the conversion factors of the World Health Organization.

of the dioxins. (*See* **Heavy Metal Toxicology; Pesticides and Herbicides**: Toxicology.)

Based on different research work and the recommended acceptable daily intake, Faroese health authorities issued the following recommended intake levels for the Faroese population in 1989:

1. On average, adults should eat no more than 150–200 g of whale meat per week. Pregnant women should eat much less or, preferably, not eat whale meat at all.
2. Adults should eat no more than 100–200 g of blubber per month.
3. People should not eat entrails (liver, kidney, etc.) from pilot whales.

In 1998 the dietary advice was revised according to new data on toxicity and contamination load and the recommendations given were:

Diet Recommendation Concerning Pilot Whale Meat and Blubber – Faroe Islands August 1998

Blubber High PCBs contents in blubber lead us to recommend that adults eat pilot whale blubber at most once to twice a month.

However, the best way to protect fetuses against the potential harmful effects of PCBs is for pregnant women not to eat blubber until they have given birth.

Meat The mercury content of pilot whale meat is high and is one of our main mercury sources. Therefore we recommend that adults eat no more than one to two pilot whale meals a month.

Women who plan to become pregnant within 3 months, pregnant women, and nursing women should abstain from eating pilot whale meat.

Organs Pilot whale liver and kidneys should not be eaten at all.

The above recommendations are considered the most advisable for the present. When new information is acquired, this diet recommendation will be revised accordingly.

This recommendation is given by the following Faroese authorities: Food and Environmental Agency, Department of Occupational and Public Health, and Chief Medical Officer.

After the killing and before the butchering of the carcasses, it is important that the beach is clean and away from sewer outlets or the like. This is important to avoid contamination by pathogenic bacteria. Most often the meat is heated before eating, except for wind-dried meat which is eaten raw (see above). When eaten raw it is of utmost importance that the meat does not contain pathogenic bacteria. The drying circumstances must be suitable to avoid the growth of such bacteria. (*See* **Spoilage**: Bacterial Spoilage.)

Waste Product Utilization

According to Faroese law, all remains after a whale drive must be removed within 24 h after the killing, and the quay cleaned again. The remaining 46% of pilot whales in the Faroes not used for consumption consists of the skeleton, head, tail, and intestines. Before the introduction of electric lighting in homes, the heads were boiled down for lamp oil, and the bones were used as fertilizer in the fields, a method which is still used in Japan. In the Faroe Islands today, the intestines are used by some as bait in line fishery, and production of fertilizer from whale remains is being considered.

See also: **Cadmium**: Toxicology; **Drying**: Drying Using Natural Radiation; **Fatty Acids**: Dietary Importance; **Fish**: Dietary Importance of Fish and Shellfish; **Fish Oils**: Dietary Importance; **Heavy Metal Toxicology**; **Meat**: Nutritional Value; **Mercury**: Toxicology; **Offal**: Types of Offal; **Pesticides and Herbicides**: Toxicology; **Preservation of Food**; **Spoilage**: Bacterial Spoilage

Further Reading

Bloch D, Hanusardóttir M, Davidsen Á and Kraul I (1987) *The Contamination of Mercury and Persistent Organochlorines in the Marine Environment of the Faroe Islands*.Torshavn: IWC document SC/39/O 26.

Bloch D, Desportes G, Hoydal K and Jean P (1990) Pilot whaling in the Faroe Islands July 1986–July 1988. *North Atlantic Studies* 2: 36–44.

Borrell A and Aguilar A (1993) DDT and PCB pollution in blubber and muscle of long-finned pilot whales from the Faroe Islands. *Report of the International Whaling Commission* (special issue)14: 351–358.

Caurant F, Amiard-Triquet C and Amiard J-C (1993) Factors influencing the accumulation of metals in pilot whales (*Globicephala melas*) off the Faroe Islands. *Report of the International Whaling Commission* (special issue) 14: 369–390.

Food and Environmental Agency (1976) *Yearly Report*. Faroe Islands: Tórshavn.

Food and Environmental Agency (1978) *Yearly Report*. Faroe Islands: Tórshavn.

Hansen C (1963) *Nerissagssiornermik ilitsersûtit kalatdlitnunänut túngatitat*. Den Grønlandske Handel. (Greenlandic cooking book.) Nakitertitai.

Henriksen J (1987) *Góða ráð*. Góða Ráð. Klaksvík. (Faroese cookbook.)

Jepsen O (1983) *Kostvurderinger over Grønlandske fødevarer*. FDB's centrallaboratorium. Copenhagen: Forenede Danske Brugsbutikker's Centrallab.

Kristjánsson L (1986) *Nýting Hvalafangs*. In: *Íslenskir Sjávarhættir*, vol. 5, pp. 70–79. Reykjavík: Bókaútgáfa Menningarsjóðs.

Poulsen M (1995) Føroyskar Føðslutalvur (Faroese Food Composition Tables.) Heilsufrøðiliga Starvstovan, pp. 1–25. Tórshavn: Food and Environmental Agency.

Skaale Ó and Johannesen M (1974) *Matur og Matgerð*. Grønalíð. Tórshavn. (Faroese cookbook.)

Small-Type Coastal Whaling in Japan (1988) *Report of an International Workshop*. Occasional publication number 27. Alberta, Canada: Japan Social Sciences Association of Canada Fund to Promote International Educational Exchange and Boreal Institute for Northern Studies.

Market Research *See* **Food Acceptability**: Affective Methods; Market Research Methods

MASS SPECTROMETRY

Contents

Principles and Instrumentation

F A Mellon, AFRC Institute of Food Research, Norwich, UK

Principles

Mass spectrometry is an extremely sensitive and specific analytical technique, capable of providing qualitative and quantitative analytical data on nanomolar to attomolar amounts of analyte. It has been applied successfully to a very wide range of analytical problems in the food and nutrition sciences. Mass spectrometers have become far more accessible, cheaper to purchase and operate, and easier to use, especially over the last decade. This is mainly a consequence of the advent of compact, bench-top instruments, usually coupled to gas or liquid chromatographs. The following article describes the basic theory of mass spectrometry and the principal types of mass spectrometric equipment relevant to food and nutrition analysis and research. Applications of mass spectrometry and techniques of combined chromatography mass spectrometry are covered in separate articles.

Mass Spectrometer

All mass spectrometers comprise four main components. First, an ion source, where sample molecules may be ionized by a variety of means, second, a mass analyzer that separates ions according to their mass-to-charge ratio, *m/z*, third, a detector that measures the abundances of the separated ions as an electrical signal, and fourth, a recording device that converts the detector signal into a form suitable for further study and processing.

Ionization Techniques – Organic Mass Spectrometry

Ionization of organic molecules may be accomplished by a variety of techniques. The most common practical methods are described below.

Formation of Ions

Organic mass spectrometry

Electron ionization (EI) The analyte is vaporized into the EI source and bombarded with 70-eV electrons. The ion source chamber is maintained at a vacuum of 10^{-3} Pa. A proportion of the energy of the electron beam is transferred to the molecule, ejecting an electron to generate a *molecular ion* ($M^{+\bullet}$), the most important ion in the mass spectrum (eqn (**1**)).

$$M + e^- \rightarrow M^{+\bullet} + 2e. \qquad (1)$$

Excess electronic energy is also transferred to the molecule during the ionization process and is rapidly converted to internal vibrational energy. This energy may be sufficient to induce bond cleavage in some ions, generating fragment ions that may also decompose. This is shown in the sequence of decompositions of a hypothetical molecule ABCD (eqn (**2**)).

$$ABCD^{+\bullet} \rightarrow ABC^+ + D^\bullet \rightarrow AB^+ + C \rightarrow A^+ + B. \qquad (2)$$

Rearrangement processes may also occur (eqn (**3**)).

$$ABCD^{+\bullet} \rightarrow AD^{+\bullet} + CB. \qquad (3)$$

Most ions formed by EI have a single positive charge ($z = 1$), and m/z is then equivalent to the mass of the ion. The way in which a molecule fragment is determined by a number of factors that will not be discussed here. Semiempirical rules for predicting or rationalizing bond cleavages and rearrangements can be found in several textbooks. The most important consequence of the fragmentation process is that it generates a characteristic, reproducible mass spectrum that serves as a spectrometric 'fingerprint' of the analyte.

Mass spectra are generally recorded in the form of a line diagram (bar graph). Ion intensities are normalized so that the largest peak in the mass spectrum, the *base peak*, has an intensity of 100 units, and all other ion abundances are expressed as a percentage of this value. The low-resolution mass spectrum of caffeine is shown in Figure 1.

A molecular ion can be seen at m/z 194, and a number of abundant fragment ions are also present. In this example, the molecular ion is also the base peak, although this is not always the case; strongly favored fragmentations may yield abundant fragment ions that are more intense than the molecular ion.

Some molecules are difficult to analyze by EI mass spectrometry for either or both of two main reasons. They may fragment very readily, so that the molecular ion is of very low abundance and therefore (in the case of an unknown) difficult to assign. Alternatively, the molecules may be involatile or thermally labile and difficult or impossible to vaporize without inducing thermolytic decomposition. Several 'soft' ionization methods are available for analyzing molecules that fall into this category. These techniques are used to transfer analytes from the liquid or solid phase to the gas phase and ionize them by mechanisms that impart minimal amounts of energy.

Chemical ionization (CI) CI is based on the reactions of gaseous analyte molecules with an excess of ions formed by electron bombardment of a reagent gas. These reactions take place at high (relative to EI) ion source pressures, typically in the range 10–150 Pa. Typical reagent gases include methane and ammonia: these generate CH_5^+ and NH_4^+ ions, respectively under CI conditions. The reagent ions are present in large excess (10^4:1) relative to analyte molecules. The reagent ions generally react with analyte molecules M by protonation (eqn (**4**)),

$$M + NH_4^+ \rightarrow MH^+ + NH_3. \qquad (4)$$

Negative ion chemical ionization is also possible. Negative ion reagent gases include water. This generates OH^- ions and reacts by proton abstraction (eqn (**5**)),

$$M + OH^- \rightarrow [M - H]^- + H_2O. \qquad (5)$$

A special case of negative ionization, electron capture negative chemical ionization (ECNCI) is not strictly CI. It is an electron capture process where low-energy electrons generate radical anions (eqn (**6**)),

$$M + e(\text{thermal}) \rightarrow M^{-\bullet}. \qquad (6)$$

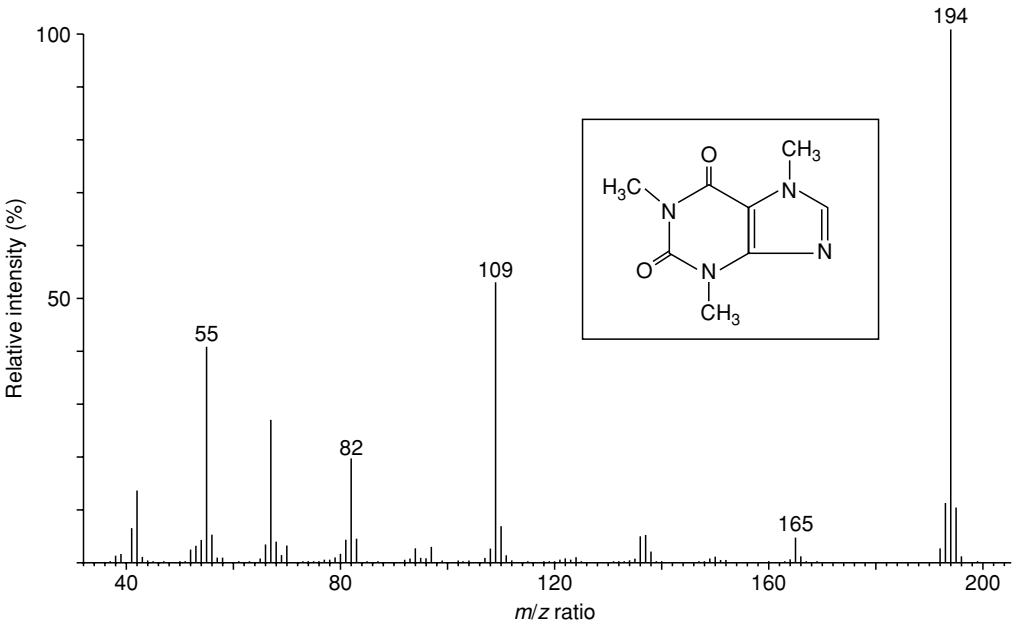

Figure 1 Low-resolution electron ionization mass spectrum of caffeine in bar graph format.

In this case, the 'reagent' gas acts as a moderator, slowing the electrons to the thermal energies where the electron capture process become efficient. ECNCI is particularly useful for determining analytes with strong electron capturing properties, for example, halogenated molecules such as dioxins. ECNCI can yield much greater sensitivities and lower detection limits in these cases.

Positive and negative reagent ions may also react in other ways. A range of gases are available for specific applications. CI spectra usually contain intense protonated or negatively charged molecules and fewer fragment ions than EI.

Desorption EI and CI In conventional EI and CI, analyte molecules are typically introduced into the ion source via a heated probe or GC inlet. Desorption EI and desorption CI provide techniques for obtaining mass spectra of less volatile, thermally labile molecules by rapid heating whilst the samples are very close to the electron beam or are immersed in the CI plasma. These techniques are still used occasionally but have largely been superseded by more robust and efficient methods for analyzing involatile molecules, as described below.

Fast atom bombardment (FAB) and liquid secondary ionization mass spectrometry FAB and liquid secondary ionization mass spectrometry essentially describe the same technique, and the terms can be used interchangeably. FAB was the first ionization technique to enjoy widespread success in analyzing polar, labile, and large molecules, including peptides and polysaccharides. The analyte in a viscous, low-boiling liquid matrix, typically glycerol that is bombarded with a stream of atoms or ions, typically Xe atoms or Cs^+ ions that have average translational energies in the keV range. Abundant protonated or cationized molecules, for example $[M+H]^+$ or $[M+Na]^+$ and deprotonated ions, $[M-H]^-$, are formed and detected in positive and negative ion modes, respectively. Characteristic fragment ions may also be formed.

FAB enjoyed great and deserved popularity in the 1980s and helped to establish mass spectrometry very strongly in the biological sciences. However, it has generally been superseded by the more powerful, versatile techniques of electrospray and matrix-assisted laser desorption ionization.

Electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) In ESI, the analyte is dissolved in a suitable solvent, for example a 50:50 mixture of acetonitrile and water, and is pumped through a narrow diameter stainless steel capillary tube that is maintained at a voltage in the range of 1–4 kV. This process generates a plume of charged liquid droplets in atmosphere (the Taylor cone). A heated (usually nitrogen) gas aids evaporation of the charged droplets and helps break up clusters of analyte and solvent ions. Analyte ions are generated from the charged droplets by an evaporation mechanism. The ions are conducted into the mass spectrometer vacuum via a pumped nozzle-skimmer system (**Figure 2**).

A curtain of nitrogen gas, which also aids evaporation of the charged droplets, prevents cluster ion formation. The ions are sampled through an orifice and enter the mass analyzer. ESI has revolutionized mass spectrometry by enabling mass spectra to be obtained routinely on polar, involatile molecules. Even high-molecular-weight biomolecules can be analyzed by ESI. Proteins, for example, yield a series of multiply charged ions well within the mass range of conventional mass spectrometers such as quadrupoles or magnetic sector instruments. These signals can be transformed mathematically to yield the molecular weight of the sample (**Figure 3**).

APCI is closely related to ESI and often employs an almost identical ion source. The only changes are the addition of a discharge electrode and replacement of the ESI spray probe with an APCI probe. In APCI, flowing liquid, typically an HPLC eluent, that contains the dissolved analyte(s) is conducted to a pneumatic nebulizer, where it forms a spray. As with ESI, the spray is generated in atmosphere and is directed towards a discharge electrode that is maintained at 1–4 kV. This is placed close to a small-diameter orifice that leads to the high-vacuum mass analyzer region. A pumped nozzle and skimmer arrangement removes excess solvent molecules, and desolvated ions are admitted to the mass analyzer, in a similar manner to an ESI source. Reagent ions are generated from solvent molecules in the electrical discharge near the corona pin. These react with analyte molecules to generate protonated or deprotonated molecules ($[M+H]^+$ or $[M-H]^-$), in a similar manner to conventional CI.

APCI is more robust (being less prone to solvent or solvent modifier effects) than ESI but is usually too vigorous and energetic to yield useful data on fragile biomolecules. However, it is very useful for determining molecules of low to intermediate polarity and is capable of producing useful data on some quite polar molecules, for example monoglycosides.

Matrix-assisted laser desorption Ionization (MALDI) MALDI is one of the best techniques for obtaining analytical data, including molecular weight information, on medium and large biomolecules. The sample, in low concentration, is mixed

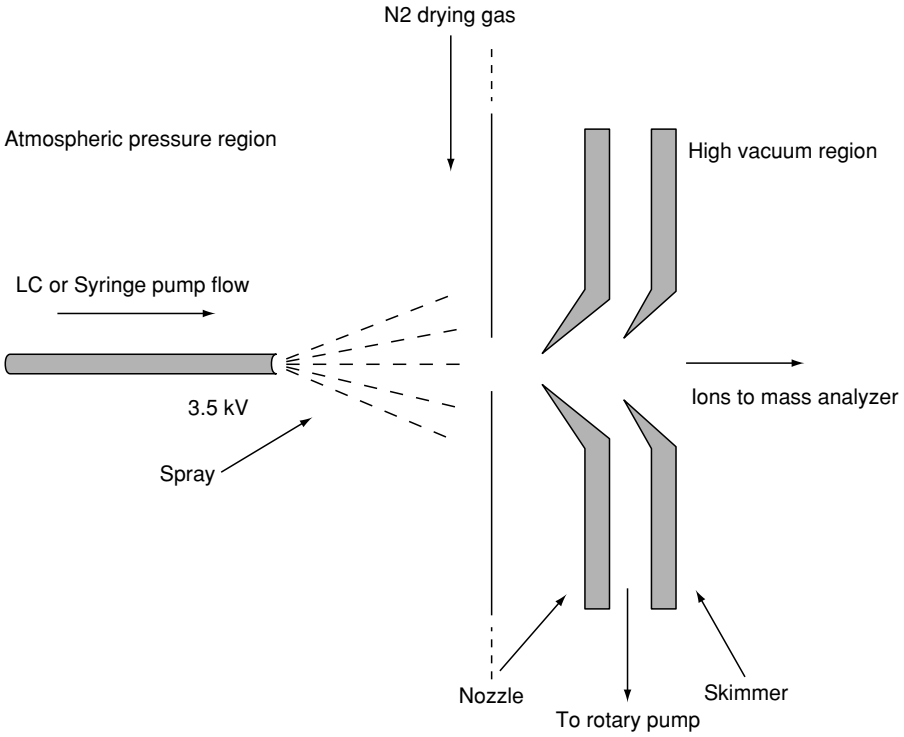

Figure 2 Schematic diagram of an electrospray ionization source.

with a concentrated solution of, for example, sinapinic acid, on a metal target. The solution is dried, and the target is introduced into the mass spectrometer ion source vacuum. The sample plus matrix is irradiated with a pulsed laser (typically a 327-nm nitrogen laser). The matrix chosen should be sublimable and be able to absorb and transfer energy at the laser wavelength. The mechanism of desorption and ionization is complex. It is thought that the matrix then transfers sufficient transverse vibrational energy to the sample molecules to desorb some of them with very little excess energy for fragmentation. The technique is particularly useful for mass measuring large and medium-sized biomolecules, especially peptides, proteins, and DNA fragments, and generates $[M+H]^+$, $[M-H]^-$ or adduct ions, $[M+Na]^+$.

Other ionization methods Several other methods have been used to ionize organic molecules in the past. For example, field ionization, field desorption, and ^{252}Cf plasma desorption were once popular techniques for analyzing otherwise intractable molecules, such as biopolymers. These techniques are now mainly of historical interest and have generally been superseded by more practical, reliable and robust methods, particularly ESI and MALDI.

Inductively coupled plasma ion source (ICP-MS) Inorganic mass spectrometry has increased in importance in food and nutrition science through applications in authentication, nutrient mineral metabolism and toxicology. One of the principle techniques of inorganic mass spectrometry, ICP-MS, is based on the production of ions in a high-temperature (6000–8000 K) atmospheric-pressure argon plasma formed by a self-sustaining electrical discharge that is induced by a high-frequency induction coil. This plasma is capable of ionizing most elements in the periodic table very efficiently. Samples are introduced into the ion source in nebulized solutions flowing at rates of 0.1 –1 ml min^{-1}. They are then ionized efficiently by the high-temperature plasma. Alternative sample introduction techniques, including electrothermal vaporization, laser desorption, hydride generation, electrospray, and combined chromatography/mass spectrometry, are also available.

The ionization efficiency is generally very high for most nutrient and toxic elements. The plasma forms a supersonic jet in the differentially pumped region behind the sampling cone. Ions are introduced into the vacuum of the mass analyzer via a series of differentially pumped stages. A schematic diagram of an ICP-MS ion source is shown in Figure 4.

Most ICP-MS instruments are based on quadrupole mass analyzers. However, high-resolution magnetic sector instruments are also available. These are capable of resolving polyatomic interferences generated in the argon plasma that may have the same mass

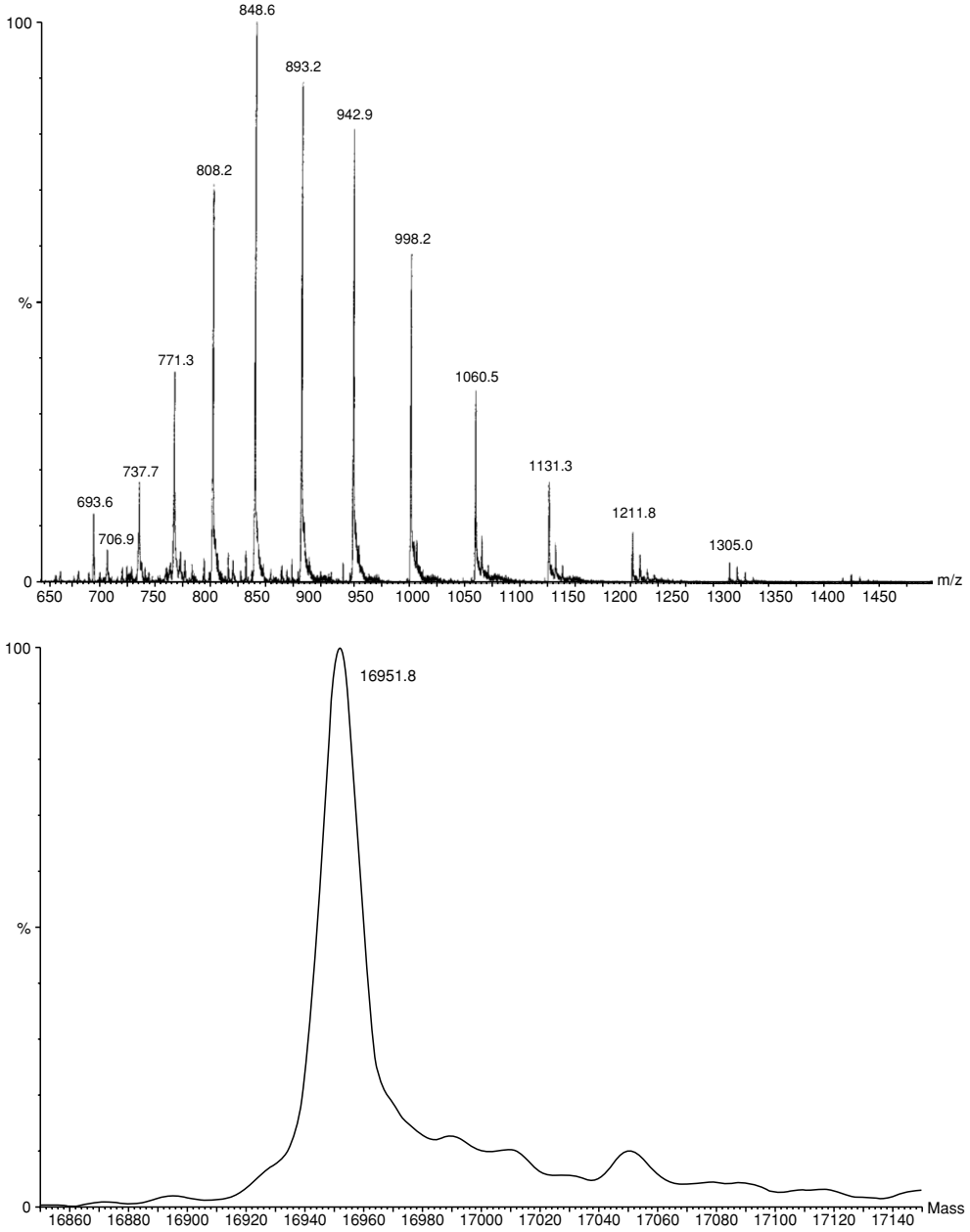

Figure 3 Electrospray mass spectrum of horse heart myoglobin, showing the raw, multiply charged data (upper trace) and the transformed spectrum (lower trace) showing the measured molecular weight of the sample.

as isotopes of interest, for example $^{40}Ar\ ^{16}O^{+}$ and $^{56}Fe^{+}$. High-precision isotope ratio measurements may also be conducted if multicollector magnetic sector mass analyzers are employed.

A major step forward has been the introduction of collision cell technology to eliminate polyatomic interference ions. These cells have been used with quadrupole and magnetic sector instruments. The important nutritional element selenium suffers from an abundant interference peak from $^{40}Ar_2^{+}$ in the ^{80}Se mass channel. This interference can be attenuated considerably by charge exchange with H_2 in the collision cell.

ICP-MS is a rapid and useful multielement analysis technique that is also capable of determining isotope ratios in studies of trace element metabolism in humans.

Thermal ionization mass spectrometry (TIMS) TIMS is a well-established technique in inorganic mass

spectrometry with a much longer pedigree than ICP-MS. It is based on the production of positive or negative atomic or molecular ions at the surface of a resistively heated metal filament. It has been used to conduct stable studies of nutrient mineral metabolism and is a 'gold standard' analytical technique for quantitative analysis of nutrients and toxicants in foods by isotope dilution mass spectrometry, in which measurements are conducted by adding a predetermined quantity (the 'spike') of an enriched stable isotope (of known isotopic composition) to the target analyte. Isotope ratio measurement of the mineral of interest then yields the amount of analyte precisely and accurately.

Mass analyzer The mass analyzer of the mass spectrometer separates ionized atoms and molecules according to their mass-to-charge ratio, m/z. Several different types of mass analyzer are available. The main types of relevance to food and nutrition research are described below.

The maximum practical resolving power, R, of a mass spectrometer is an important instrumental characteristic. It is effectively a measure of the instrument's ability to separate ions. Several different methods can be used to define R, but the two most common are the '10% valley' definition (eqn (7)), which is based on the separation between two peaks of equal height and mass m and $m + \Delta m$,

$$R = \frac{m}{m + \Delta m}, \qquad (7)$$

when the peaks are separated by a valley of 10% peak height, and the full width at half height definition (eqn (8)),

$$R = \frac{m}{\Delta m}, \qquad (8)$$

where Δm is the width of a peak of mass m at 50% of its height.

Some instruments, magnetic sector-based mass spectrometers, for example, yield constant resolution over the entire mass range (i.e., $m/\Delta m$ is constant). In contrast, quadrupole mass analyzers yield a constant Δm over the entire mass range. In this case, resolution is mass-dependent, usually some multiple of the mass, e.g., twice the mass.

Magnetic sector mass spectrometers Ions formed in the ion source region are accelerated through a potential and are injected into a sector magnetic field (*See* **Figure 8**). The motion of the ions in the magnetic field is described by eqn (**9**),

$$m/z = \frac{B^2 r^2 e}{2V}, \qquad (9)$$

where z is the number of charges on an ion of mass m, e is the electronic charge, B is the magnetic field strength, and r is the radius of the ion path. The mass scale is generally scanned by varying, B rather than V, as this yields the most sensitive and reproducible data. Voltage scanning is used to perform selected ion monitoring measurements, in which a small number of mass channels are monitored consecutively and repetitively. This mode is used to record quantitative data with high sensitivity (selected ion monitoring measurements may also be performed using other types of mass analyzer).

Single-focusing magnetic sector mass spectrometers (now something of a rarity) are capable of a maximum resolution of about 5000. Increased resolving power is attainable by adding an energy resolving sector, a pair of curved plates held at a defined electrical potential, to yield a 'double-focusing' instrument. These help to increase resolving power by

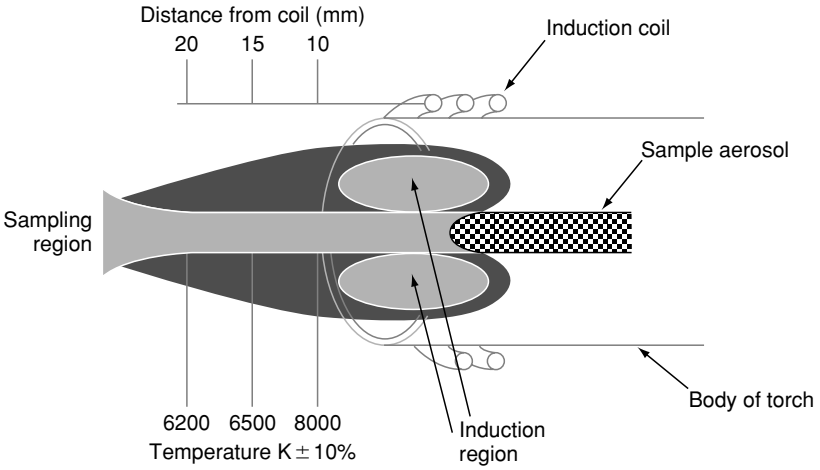

Figure 4 Schematic diagram of an ICP-MS ion source. Redrawn from a diagram supplied by VG Elemental, with permission.

correcting for the spread in ion velocities that most ionization techniques produce. The electric sector can be placed before or after the magnetic sector. Figure 5 shows a design of 'Nier–Johnson' geometry, in which an electric sector deflects the ion beam through $\pi/2$ radians, before it enters a magnetic sector with a deflection of $\pi/3$ radians.

Double focusing mass spectrometers are available in a number of designs and can attain maximum resolving powers from 10 000 to over 200 000, depending on their design.

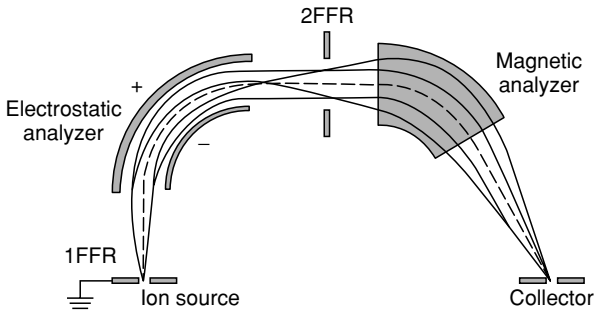

Figure 5 Schematic diagram of double-focusing magnetic sector ion optics of Nier–Johnson geometry. FFR, first field-free region. From Chapman JR (1985) *Practical Organic Mass Spectrometry*, 2nd edn. Chichester, UK: Wiley-Interscience with permission.

High-resolution mass spectrometry is commonly used to measure 'accurate masses', i.e., masses determined to three or four decimal places. These data may then be converted into the empirical formulae of ions in the mass spectrum. High-resolution mass spectrometry may also be used to increase the specificity of quantitative measurements, or to resolve isotopic clusters of multiply charged peaks formed by electrospray so that the charge state can be measured unambiguously.

Isotope ratio mass spectrometers Although isotope ratios may be measured successfully using conventional mass spectrometers, some applications much require higher precision measurements than can be attained by using scanning instruments. In such cases, specialized high precision isotope ratio mass spectrometers are available. An example of this type of instrument is the gas isotope ratio mass spectrometer (GIRMS), used to determine the isotope ratios of gases such as CO_2, N_2, SO_2, and H_2 to a very high degree of precision and accuracy.

The analyte is combusted or reduced to generate the gases that are ionized in a gas tight EI source. A magnetic sector separates the ion beam, with the difference (from a conventional scanning instrument) that each mass channel is collected simultaneously, i.e., the mass spectrometer uses a *multicollector* ion detection system. This yields high-precision isotope ratios because any fluctuations in ion beam

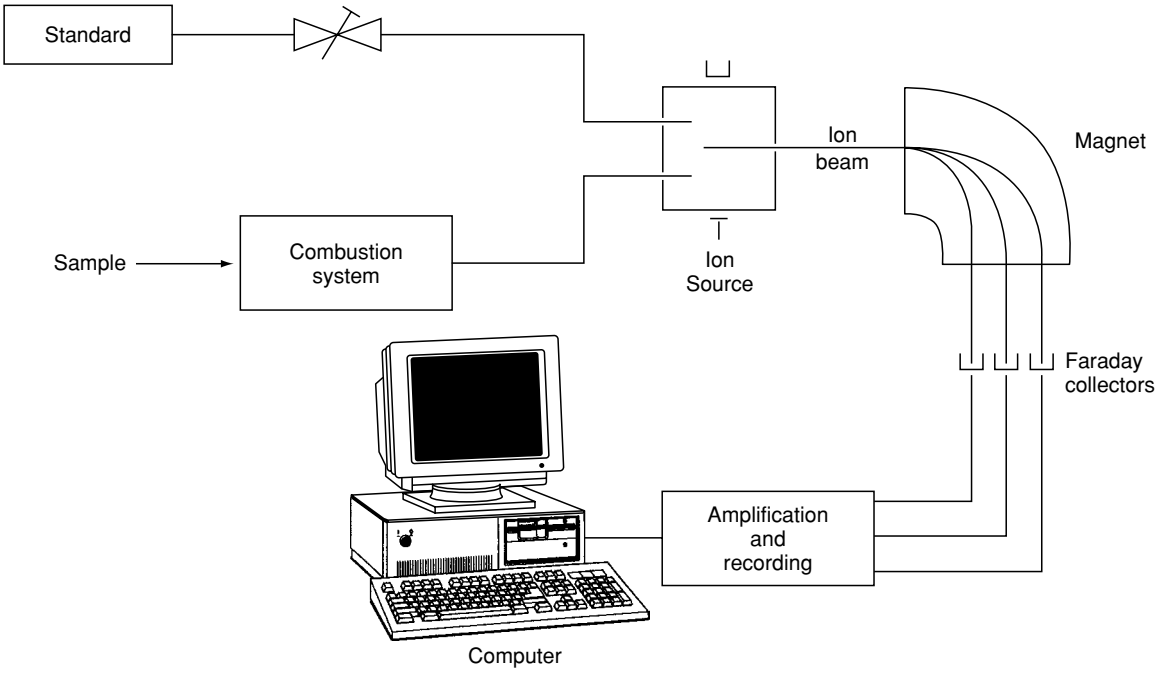

Figure 6 Schematic diagram of a gas isotope ratio mass spectrometer. From Mellon F, Self R and Startin JR (2000) *Mass Spectrometry of Natural Substances in Food*. London: The Royal Society of Chemistry with permission.

intensity occur simultaneously at each detector. A diagram of a GIRMS instrument is shown in Figure 6.

The main uses of GIRMS in food and nutrition science are in determining the authenticity of foods and in human studies of nutrient metabolism. High-precision multicollector mass spectrometers are also available for use with TIMS or ICP-MS.

Quadrupole mass analyzers Quadrupole mass analyzers comprise four parallel rods of circular or hyperbolic cross-section. These are connected to radio-frequency (V_0 *cos* ωt) and direct-current (U) power supplies, as shown in Figure 7.

The equations of motion governing the path of ions between the rods describe a complex series of oscillations and will not be discussed here. These oscillations are stable for particular values of the equations and unstable for other values. By operating the quadrupole in a stable region, the ions will be constrained to follow a path between the rods until they reach the detector. The mass spectrum is scanned by varying V_0 and U so that the ratio U/V_0 remains constant. The quadrupole mass analyzer acts as a mass filter, allowing one mass channel at a time to reach the detector as the mass range is scanned. Quadrupole mass spectrometers are intrinsically low-resolving-power instruments but are relatively cheap and robust.

Ion traps The Paul ion trap is related to quadrupole mass analyzers. The term 'Ion Trap' is, strictly speaking, generic and can encompass the Penning ion traps used in Fourier transform ion cyclotron resonance (FTICR) instruments (see below). However, ion traps are generally used in the scientific literature to refer to Paul ion traps, thanks to the great success of these instruments in compact, bench-top designs of mass spectrometer.

An ion trap may be described as a quadrupole that has undergone a solid of rotation. A typical ion trap comprises two endcap electrodes and a ring electrode, all of hyperbolic or hemispherical cross-section (Figure 8).

The end-cap electrodes contain small-diameter holes for allowing ions to enter and leave the trap. Ions are confined inside the trap by a radio-frequency field of constant frequency but variable power. The ions may be detected, according to their m/z ratio, by applying voltages sufficient to eject them from the trapping field. Ion traps, despite their simplicity and cheapness of construction, can be used to perform sophisticated tandem mass spectrometry experiments in addition to fulfilling a role as conventional scanning mass spectrometers.

FTICR mass spectrometry FTICR (sometimes simply known as Fourier transform mass spectrometry) is a Penning trap in which ions confined by a strong magnetic field, B, move in circular orbits of characteristic frequency, ω (Figure 9).

The motion of the ions is governed by the cyclotron equation (eqn. (**10**)),

$$\omega = z\frac{B}{m}. \qquad (10)$$

A broad-band 'chirp' of electromagnetic radiation is applied to excite all ions in the trapping cell to their cyclotron frequency. The orbiting ions induce an alternating 'image current' in the walls of the cell

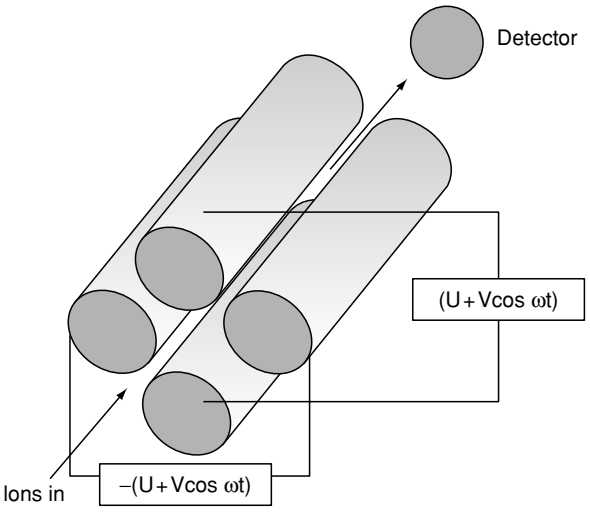

Figure 7 Schematic diagram of a quadrupole mass spectrometer.

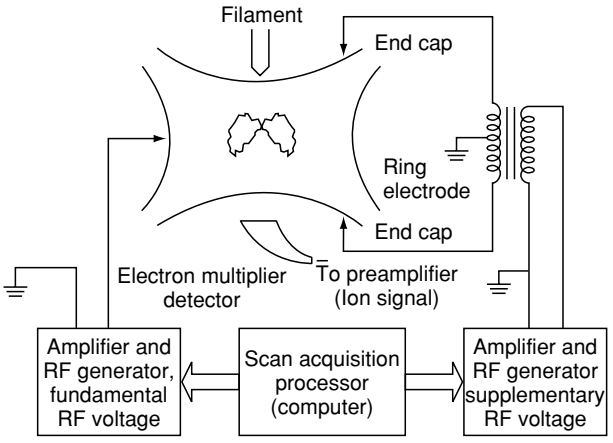

Figure 8 Schematic diagram of an ion trap (Paul trap) mass spectrometer. From Andrews DL (1990) *Perspectives in Modern Chemical Spectroscopy*. Berlin: Springer-Verlag with permission.

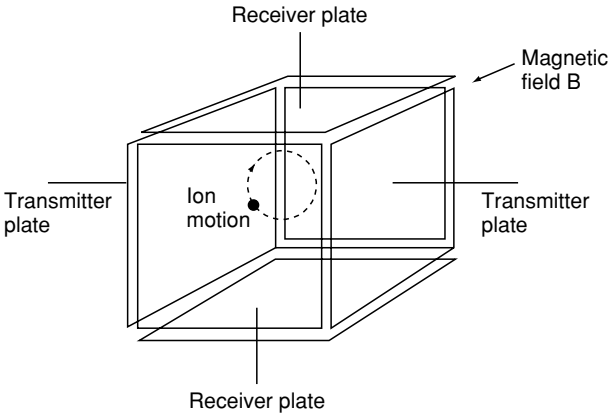

Figure 9 Schematic diagram of a Fourier transform ion cyclotron resonance mass spectrometer. From Andrews DL (1990) *Perspectives in Modern Chemical Spectroscopy*. Berlin: Springer-Verlag with permission.

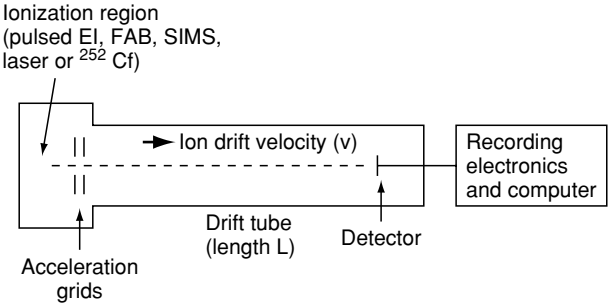

Figure 10 Schematic diagram of a ToF mass spectrometer. From Andrews DL (1990) *Perspectives in Modern Chemical Spectroscopy*. Berlin: Springer-Verlag with permission.

receiver plates. This signal is complex because each *m/z* value generates its own characteristic frequency. A Fourier transform of this complex signal yields the masses of all ions in the spectrum. Frequencies can be measured with great accuracy, and so very high resolving powers, over 1 000 000 in some cases, are attainable. However, resolution does fall off with increasing *m/z*.

FTICR instruments can be used to perform sequential tandem mass spectrometry experiments and are compatible with pulse ionization techniques. Although they are expensive to purchase, and running costs are high, they have undergone something of a renaissance in recent years. This is partly due to their compatibility with newer pulse ionisation techniques like MALDI. Perhaps more importantly, the multiple charging observed in ESI of proteins yields a mass envelope in the region where full advantage can be taken of FTICR's high-resolution performance (i.e., 600–2000 Th). This allows unambiguous determination of the charge state of product ions formed in tandem mass spectrometry experiments.

Time of flight (ToF) mass analyzers The ToF mass analyzer is now becoming widespread for several reasons. First, it is compatible with MALDI, a pulsed ionization technique that is a perfect match for ToF. Second, improvements in design and performance have yielded major advances in performance, particularly resolution and sensitivity. Third, ToF mass analyzers have, theoretically, an unlimited mass range.

ToF analyzers are based on the property that the drift velocities of ions are dependent on their *m/z* ratio. Ions formed in the source region are accelerated through a potential, *V*, and then enter a field-free drift region (Figure 10).

According to eqn (**11**), the ions will all have the same kinetic energy.

$$\frac{1}{2}mv^2 = zeV, \qquad (11)$$

where v is the ion velocity. The ToF, t, of the ions is governed by eqn (**12**),

$$t = \left(\frac{m}{2zeV}\right)^{\frac{1}{2}} L. \qquad (12)$$

Early ToF instruments had limited mass resolution, but 'delayed extraction' techniques and the introduction of electrostatic mirrors (reflectrons) have increased their resolution dramatically. ToF instruments are capable of a very high sensitivity because they record all the ions generated in a single ionization pulse.

Ion detection and recording The most common detector in mass spectrometry is the *electron multiplier*; this amplifies the weak ion beam signal greatly, yielding gains of up to the order of 10^6. Multipliers can be either a *discrete dynode* type or a *continuous dynode*, generally known as a *channel electron multiplier* or *channeltron*. The discrete dynode electron multiplier comprises a series of Be–Cu alloy dynodes arranged as shown in Figure 11.

The first dynode is held at a negative potential of several kilovolts relative to the (positive ion) beam. The beam strikes the first dynode and generates a shower of secondary electrons. These electrons are then attracted to the second dynode, generating more electrons, and so on, in a cascade through the device. Negative ions may be detected by using a conversion dynode held at a high positive potential. Channeltron detectors operate according to similar

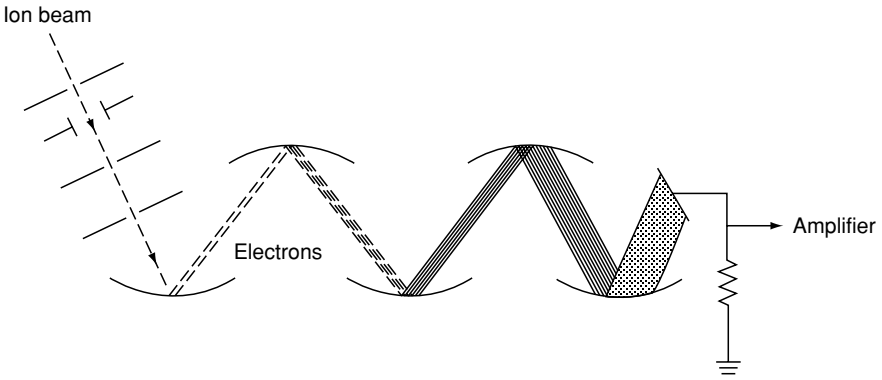

Figure 11 Schematic of a discrete dynode electron multiplier ion detector. From McFadden WH (1973) *Techniques of Combined Gas Chromatography/Mass Spectrometry*. Chichester, UK: Wiley-Interscience with permission.

principles as discrete dynode electron multipliers. The main difference is that individual dynodes are replaced by a tapered, coiled tube, and electrons bounce from along the tube towards the earthed terminus. *Array detectors* comprise a series of miniaturized channeltrons that are arranged adjacent to each other. They can be used to detect ions simultaneously over a range of masses and are used to increase sensitivity in some types of magnetic sector mass spectrometer.

The *scintillation detector* is an alternative ion beam amplification device in quite widespread use. The ion beam strikes a surface that emits electrons towards a phosphor screen, generating photons. The photons are amplified using a device akin to the image intensifiers used in night-vision binoculars and similar devices. Scintillation-based photomultiplier arrays are also available.

Faraday cup detectors are used mainly in isotope ratio mass spectrometers. The Faraday detector is a very simple device, comprising a conducting 'bucket' that collects the ion beam, yielding an electrical signal. The minute electrical signal generated is then amplified and recorded. Faraday cups are used in applications where accurate measurement is more important than sensitivity, for example in high-precision isotope ratio measurements conducted by GIRMS, ICP-MS, or TIMS.

Tandem mass spectrometry (mass spectrometry/mass spectrometry (MS/MS)) Tandem mass spectrometry, or mass spectrometry/mass spectrometry (MS/MS), experiments are conducted classically by combining two mass analyzers in a single instrument. For example, two double focusing magnetic sector or two quadrupole mass analyzers may be combined in series. Tandem quadrupole mass spectrometers are generally known as 'triple quadrupoles.' However, only two of the quadrupoles are involved in mass analysis: the third (middle) quadrupole transmits all ions and is used to 'activate,' i.e., transfer energy to, the ions of interest.

MS/MS has been described as 'taking the mass spectrum of an ion in a mass spectrum.' The principle of MS/MS is that a sample is ionized and mass-analyzed in the first mass analyzer. A particular *m/z* value of interest is selected from the mass spectrum and is directed into a *collision cell* that contains a neutral gas (argon, for example). The ion is vibrationally excited by collision with the target gas, a process known as *collision induced dissociation*. This generates fragment ions that are separated and recorded by the second mass analyzer. The process is shown schematically in Figure 12.

The example given is of a *product ion scan* and is used as a general technique for analyzing target components in a mixture, or for generating characteristic fragment ions from a sample that yields stable molecular ions under the ionization conditions used. Other scan modes are available, including *precursor ion scans, constant neutral loss*, and *selected reaction monitoring*.

Tandem mass spectrometers are available in several different configurations. 'Hybrid' instruments, for example magnetic sector/quadrupole or quadrupole ToF, are also found in many laboratories. ToF mass spectrometers may also be used in MS/MS mode. MS/MS experiments may also be performed using ion trap or FTICR instruments. In this case, MS/MS is conducted by temporal, rather than spatial, separation of ions. For example, an ensemble of ions in an ion trap is 'swept' using a scan function that expels all ions from the trap except the *m/z* value of interest. These ions are activated by collision with a gas admitted to the ion trap, and the ionic products of

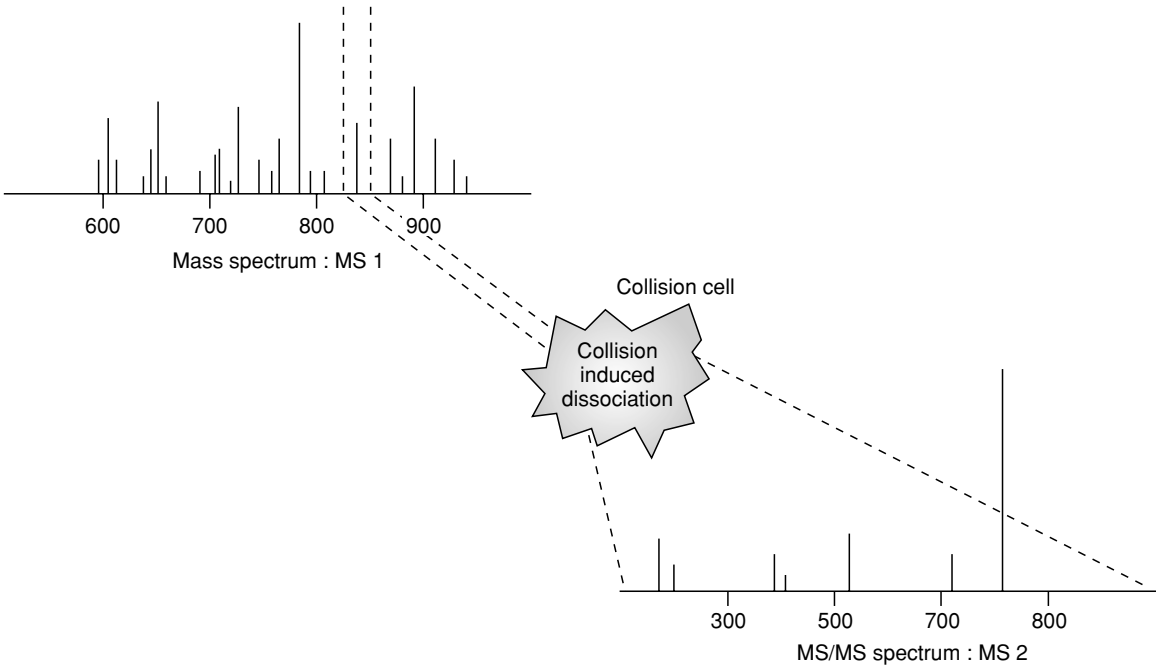

Figure 12 Schematic representation of tandem mass spectrometry (product ion analysis).

this activation are measured by a conventional mass scan. This process of isolation can be repeated, i.e., a particular product ion can, in turn, be isolated and activated, and so on in a sequence of reactions. This allows MS/MS/MS or MS^n experiments to be carried out.

MS/MS has many analytical applications. These include the analysis of complex mixtures without any chromatographic separation, generation of characteristic fragment ions from molecular ions, and improving the selectivity of quantitative measurements.

Data acquisition and processing Computers, typically PCs, are now used routinely both to control mass spectrometers and to acquire and process mass-spectrometric data. Parameters such as focusing, resolution, and scan speed may be adjusted and set via a computer. The same computer may be used to control mass spectrometer inlet systems such as gas chromatography, high-performance liquid chromatography, or automated direct insertion probes. The computer is also the final link in the signal-recording chain. It forms part of the system used to digitize the mass spectrum, calibrate the mass scale, enhance the quality of data, and, ultimately, to aid in the interpretation of data. All these functions may be carried out by a single computer coupled to a sophisticated data digitization device.

Library searching of a large database of the mass spectra of known compounds is now routine as an aid to interpreting EI spectra and identifying components in (for example) a complex gas chromatography – mass spectrometry chromatogram. More sophisticated, compound class-specific algorithms are available for particular types of analysis, such as identification of proteins in proteomics experiments. Computer systems are now an integral component of mass spectrometers and have greatly simplified the process of instrument control, data acquisition, and interpretation.

See also: **Chromatography**: Gas Chromatography; **Mass Spectrometry**: Applications

Further Reading

Adams F, Gijbels R and van Grieken R (1988) *Inorganic Mass Spectrometry*. Chichester, UK: Wiley-Interscience.

Chapman JR (1995) *Practical Organic Mass Spectrometry*, 2nd edn. Chichester, UK: Wiley-Interscience.

McLafferty FW and Turecek F (1993) *Interpretation of Mass Spectra*, 4th edn. Sausalito, USA: University Science Books.

Rose ME and Johnstone RAW (1996) *Mass Spectrometry for Chemists and Biochemists*, 2nd edn. Cambridge: Cambridge University Press.

Applications

F A Mellon, AFRC Institute of Food Research, Norwich, UK

Background

Food science, food technology, and nutrition present the scientist with a very wide range of analytical problems. Consequently, many different mass spectrometric techniques, alone or in combination with chromatographic or other methods, have been used to address specific analytical problems in food and nutrition research. It is impossible to give a comprehensive account of applications of mass spectrometry in food science in a short article. Instead, some representative examples are chosen from a range of topics that illustrate the diversity of techniques and applications in selected areas.

Major Food Components

Carbohydrates

Carbohydrate analysis has a long history in the food sciences because of the importance of carbohydrates as macronutrients, as major constituents of dietary fiber, and as food structure components and food additives. A classic, still widely used, method for determining the composition and linkages of polysaccharides involves gas chromatography/mass spectrometry (GC/MS) analysis of partially methylated alditol acetates (PMAAs). The polysaccharides are methylated, hydrolyzed, and reduced to *O*-methyl alditols. These are acetylated and analyzed (both qualitatively and quantitatively) by GC/MS. Partially methylated alditol acetates yield characteristic mass spectra that allow determination of the nature and linkage position of the monosaccharides comprising the polysaccharide. This technique has been used to analyze cell wall polysaccharides in many fruits, vegetables, and cereals.

The molecular weights of large polysaccharides can be measured with the aid of electrospray ionization (ESI) (in liquid chromatography/mass spectrometry (LC/MS) or direct injection modes) and matrix-assisted laser desorption ionization (MALDI) mass spectrometry (MS). Both techniques may also be used to obtain partial or (for smaller polysaccharides) complete sequence information in MS/MS mode. Pectins, for example, are widely used in the food industry as stabilizers or to form gels. Nanospray MS/MS of an unseparated digest of partially methyl-esterified (DE (degrees of esterification) 31%) pectin has been used to obtain useful structural information. The tandem mass spectrum of the doubly deprotonated decamer with two methyl esterified residues is shown in Figure 1. ^{18}O labeling is used to distinguish the reducing and nonreducing ends of the molecule.

Complete Foods

Many mass-spectrometric methods used in food analysis and research focus on the determination of specific components or of mixtures of structurally related compounds. However, pyrolysis mass spectrometry (Py/MS) is capable of yielding analytical information on microgram or submicrogram amounts of whole foodstuffs. Py/MS is based on the rapid, controlled thermal degradation of samples in an inert atmosphere. This produces mixtures of volatile compounds whose electron ionization (EI) or chemical ionization (CI) mass spectra are recorded. This generates a complex 'fingerprint' of the sample under investigation. Py/MS data are generally processed by chemometric methods to obtain optimal information from the analyte. The technique has been used in the quality assurance of foods and drinks, in the study of food plants, and in the identification of food pathogens.

Flavors and Taints

Because most flavor components are highly volatile, combined GC/MS (using EI and/or CI) has been a primary technique in flavor analysis and research since the early 1960s. For example, almost 1000 discrete compounds have been identified in coffee volatiles.

Direct injection of headspace samples on to a GC column reflects flavor composition more accurately than extraction or trapping methods and is less likely to generate artefacts. However, the ease with which solid-phase microextraction GC/MS studies may be conducted has made this an important and useful technique for sampling flavors and taints. GC/MS has been supplemented by LC/MS techniques for studying involatile flavor precursors (glucosinolates, for example) or semivolatile or involatile food components that have important flavor characteristics.

The availability of a large knowledge base of flavor profiles (largely defined by GC/MS analysis) and recent instrumental advances have resulted in a recent shift in emphasis of mass spectrometric applications. Instead of characterizing complex mixtures of volatiles, several

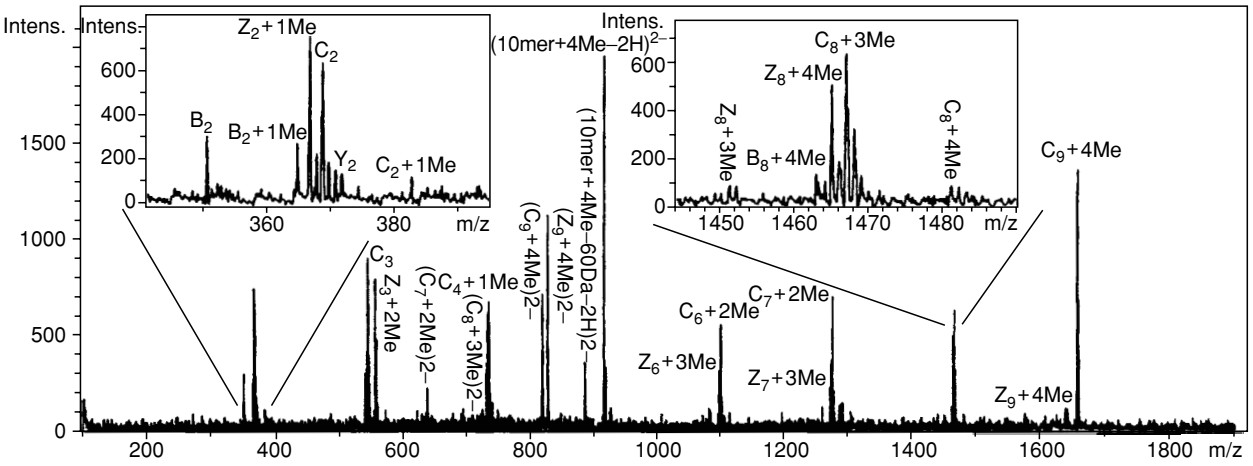

Figure 1 Negative ion MS/MS spectrum of the doubly deprotonated decamer with two methyl esterified residues and ^{18}O-labeled reducing end from an unseparated PG II digest of 31% methyl esterified pectin. From Korner R, Limberg G, Christensen TMIE, Mikkelsen JD and Roepstorff P (1999) Sequencing of partially methyl-esterified oligogalacturonates by tandem mass spectrometry and its use to determine pectinase specificities. *Analytical Chemistry* **71**: 1421 with permission.

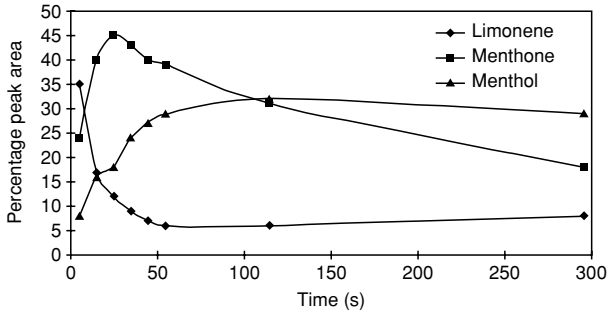

Figure 2 Average volatile release curves for menthol, menthone, and limonene released while eating mint-flavored sweets. From Linforth RST, Ingham KE and Taylor AJ (1996) *Flavour Science: Recent Developments*, pp. 361–368. Cambridge: Royal Society of Chemistry with permission.

researchers are now focusing on flavor release and, more specifically, on sampling volatiles released into the mouth and nose. By using a special atmospheric-pressure chemical-ionization (APCI) probe coupled to a mouth or nose piece, it is possible to conduct dynamic, breath-by-breath analyses of air expired during eating. An example of the type of data that can be acquired is illustrated in Figure 2, which shows volatile release curves for menthol, menthone, and limonene released while eating a mint-flavored sweet. The APCI technique represents the 'state of the art' in modern flavor-release studies.

Mass-spectrometric methods are also useful for authenticating flavor components. Flavor concentrates from natural sources are usually sold at a premium price compared with their synthetic analogs. Mass spectrometry has an important role in detecting fraudulent representation of synthetic flavor compounds as natural. Microcombustion of flavor samples yields CO_2 that is then swept into the ion source of an isotope ratio mass spectrometer where $^{13}CO_2/^{12}CO_2$ ratios are measured to a very high precision. These isotope ratios are characteristic of the source of the sample and can be used to authenticate the material. The technique can be used directly on the sample without prior pretreatment. If a more sophisticated analysis is required, GC/combustion/isotope ratio mass spectrometry will yield accurate isotope ratios on the individual components of a flavor sample.

Lipids

Although complex mixtures of lipids cannot be characterized by any single analytical technique, mass spectrometry is very useful in the qualitative and quantitative analysis of many different lipid types, including long-chain fatty acids and their esters, mono-, di-, and triacylglycerols, and phospholipids.

Branched-chain fatty acids and their esters exhibit characteristic bond scissions adjacent to the branching point. The degree of unsaturation of fatty acids esters can be determined readily from EI mass spectra. However, double-bond *location* presents considerable difficulties because long-chain unsaturated compounds rearrange extensively under EI, and consequently, fragment ions do not generally supply reliable information about double-bond position. A number of strategies have been devised to overcome these problems: one of the most popular is derivatization, followed by GC/MS. Several different procedures are available, including oxidation followed by silylation, epoxidation and the analysis of methoxy or dimethyldisulfide derivatives. An alternative strategy is to generate chemical derivatives that localize

charge, preventing charge-induced double-bond migration. Pyrrolidide derivatives have been used for this purpose but are insufficiently volatile for GC/MS. A useful alternative that generates derivatives suitable for GC/MS is provided by 2-substituted 4,4-dimethyloxazoline formed by condensation of the fatty acid with 2-amino-2-methylpropanol. The 'charge-localizing' derivatives yield clusters of ions 14 mass units apart, reducing to 12 mass units when a double bond interrupts the chain. MS/MS techniques for locating double bonds include 'charge-remote' fragmentation of molecules ionized by negative ion CI. If high collision energies are used, free fatty acids yield ions characteristic of the double-bond position.

Capillary GC/MS is useful for identifying the carbon and unsaturation number of acylglycerols. However, APCI LC/MS has the ability to analyze less volatile acylglycerols (particularly-triacylglycerols) that may be unsuitable for GC/MS. It is fast becoming the technique of choice for qualitative and quantitative analysis of acylglycerols and, when combined with MS/MS, is capable of distinguishing fatty acid chains in the *sn*-2 position from those in the *sn*-1/3 positions.

The sensitivity, selectivity, and convenience of positive and negative ion ES LC/MS make this the current benchmark technique for analyzing phospholipid mixtures. Additional structural information can be obtained by MS/MS. Where these advanced techniques are unavailable, a more laborious approach of chromatographic class separation, chemical degradation, and GC/MS analysis of fatty acid can be used to generate useful structural information.

Proteins, Oligopeptides, and Amino Acids

Amino acids, peptides, and proteins are of interest in several areas of food and nutrition science, and only a small sample of current applications can be described here. For example, GC/MS, and LC/MS are particularly useful for analyzing the products of amino acid/sugar reactions (Maillard or Amadori products) that have important flavor characteristics. Free amino acids themselves may be analyzed by electrospray ionization (ESI) LC/MS. However, alternative (and cheaper) techniques to mass spectrometry are available, and mass spectrometric methods are generally confined to specialist applications (for example, metabolic studies or the identification of unusual amino acids).

The mass-spectrometric study of proteins and oligopeptides was once largely conducted by fast-atom bombardment (FAB) MS, but this methodology has now been supplanted by the more robust and sensitive methods of ESI and MALDI time of flight (ToF). An instructive example is provided by a rapid screening system, based on MALDI-ToF, developed to determine the presence of gliadins in food samples. Direct observation of the characteristic gliadin mass pattern in processed and unprocessed gluten-containing foods has yielded an nonimmunological alternative for the quantitative estimation of gluten gliadins in foods. Figure 3 shows the MALDI-ToF spectrum of an unfractionated ethanol extract of wheat gliadins.

Tandem mass spectrometry of oligopeptides, usually following ESI ionization, yields characteristic ions that allow full or partial amino acid sequencing of the analyte molecule. These techniques can be used to confirm the structures of individual oligopeptides, for example the food preservative nisin and genetically engineered analogs of this molecule.

One of the most powerful applications of mass-spectrometric peptide sequencing methodology is in the rapidly developing field of proteomics. Two main mass-spectrometric techniques have been developed for proteomic studies. The first involves enzymatic digestion of individual spots excised from a two-dimensional gel. The resulting peptide mixture is analyzed, without prior separation, by MALDI-ToF mass spectrometry. This yields a unique pattern of molecular weights that are searchable against protein databases. If this fails to yield a positive identification, a backup technique, involving ultrasensitive MS/MS sequencing of peptides generated by specific digestion of a protein two-dimentional spot, yields a 'sequence tag,' again searchable against a database to identify the unknown protein. Finally, a third level in this hierarchy of mass spectrometric techniques may be used to identify posttranslational modifications of proteins, where required. Proteomics has considerable potential in food and nutrition research, for example in the investigation of environmental responses in food-poisoning microorganisms, in the study of diet–health relationships, and in the assessment of genetically modified food safety.

Nutritional Studies

Stable Isotope Methods

Interest in the use of stable-isotope mass spectrometry for studying the nutritional value of foods and diets has increased considerably. Enriched stable isotopes, in contrast to radioisotope methods, provide safe, scientifically rigorous techniques for determining nutrient bioavailability and metabolism in all human population groups.

The only major drawbacks of stable isotope studies are associated with the presence of endogenous isotopes of the elements under investigation. Sufficient label must be administered to generate a measurable increase in isotope ratio above natural

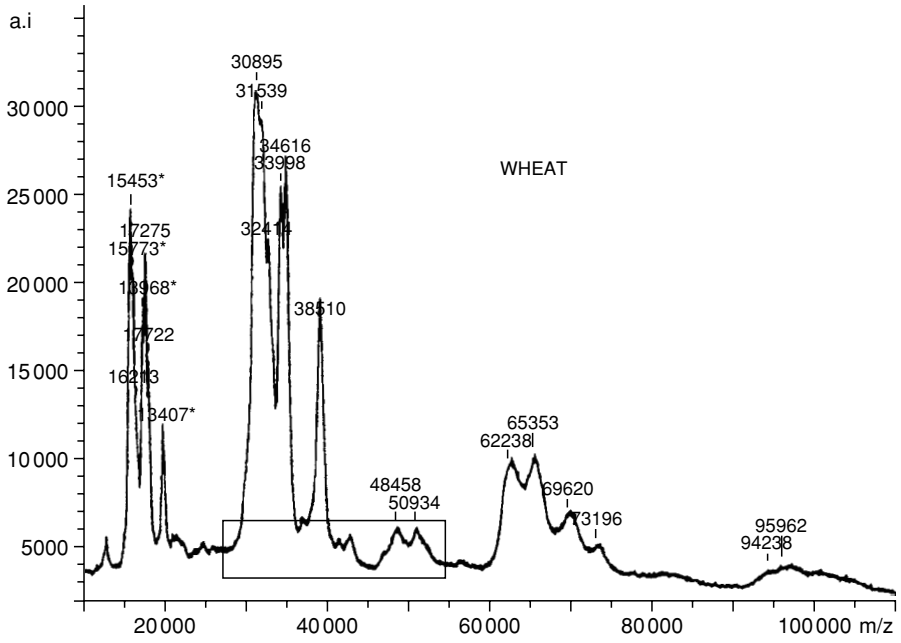

Figure 3 MALDI-ToF spectrum of the unfractionated 70% ethanol extract of wheat endosperm. A box indicates the characteristic gliadin region. Asterisked peaks are doubly charged ions. From Mendez E, Camfieta E, Sansebastian J *et al*. (1995) Direct identification of wheat gliadins and related cereal prolamins by matrix assisted laser desorption ionization mass spectrometry. *Journal of Mass Spectrometry and Rapid Communications in Mass Spectrometry (Combined Issue)* S123 with permission.

abundance. However, care is needed to prevent administration of excessive amounts of label, as this would yield unrepresentative data and might even have toxic effects. However, administration of small quantities of label requires very accurate measurement of isotope ratios because the enriched material is diluted by the endogenous nutrients already present in the body. In the case of mineral metabolism studies, additional precautions are necessary when studying trace elements because of the possibility of contamination with environmental minerals during sample processing.

Stable isotope nutritional studies usually involve administration of the enriched stable isotope of an element or isotopically labeled compound in, or with, a meal. The method of labeling depends on the type of study undertaken. For elements such as selenium, where absorption and metabolism are highly dependent on chemical form, an intrinsic label (i.e., one incorporated biosynthetically in the food) should be used in most cases. However, minerals believed to form a 'common pool' in the digestive system may be mixed directly with the food and administered following equilibration (extrinsic labeling). In some cases, a second isotope is injected or infused intravenously to correct for endogenous losses. Samples of breath, blood, urine, saliva, or feces are then collected for an appropriate period and subjected to isotopic analysis. Isotope ratio measurement often requires specialized instrumentation such as high-precision gas isotope ratio mass spectrometers or specialized inorganic mass spectrometers (e.g., thermal ionization mass spectrometry (TIMS)) or inductively coupled plasma mass spectrometry (ICP-MS).

Mineral Nutrients

Although GC/MS, FAB TIMS, and ICP-MS have all been used to measure enriched stable isotopes of nutrient minerals (in the case of GC/MS, after derivatization to generate volatile metal–chelate complexes), the dedicated inorganic methods of TIMS and, especially, ICP-MS are now used almost exclusively. ICP-MS has distinct advantages over alternative techniques because it is rapid and sensitive, and sample preparation can be minimized. Furthermore, because ICP-MS can be used with continuous flow sample introduction methods, it is suitable for coupling directly to separation techniques such as size-exclusion chromatography or HPLC. This means that speciation studies, i.e., determination of the chemical form of particular elements (oxidation state, protein binding, etc.) and the effect of this on metabolism may also be studied. The two major drawbacks of ICP-MS, low precision relative to TIMS and interference effects from polyatomic ions in the argon plasma, have largely been overcome by new generations of instruments equipped with multiple collectors and collision cells, respectively.

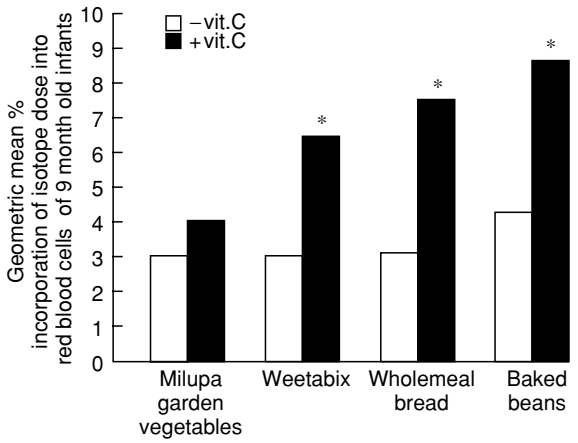

Figure 4 Bioavailability (percentage incorporation into red blood cells, measured by TIMS mass spectrometry) of iron in selected weaning foods given to 9-month-old infants, with and without a vitamin C drink. From Fairweather-Tait SJ, Fox T, Wharf SG and Eagles J (1995) The bioavailability of iron in different weaning foods and the enhancing effect of a fruit drink containing ascorbic acid. *Pediatric Research* 37: 389 with permission.

A wide range of human studies have been conducted, using enriched stable isotopes of nutrient minerals, including determination of the absorption and metabolism of iron, zinc, calcium, copper, selenium, and molybdenum. An example of the type of information that can be obtained is provided by a study of iron absorption from different weaning foods, and the effects of vitamin C on iron absorption (Figure 4). These measurements, conducted using enriched stable isotopes of ^{57}Fe and ^{58}Fe, demonstrated a doubling of iron absorption when a drink containing 50 mg of vitamin C was administered with the food.

A recent development, especially useful for conducting very long-term studies of the effect of diet on metabolism, is to administer extremely low levels of long-lived radioisotopes, e.g., ^{41}Ca, that are then measured by the ultrasensitive, highly specific technique of accelerator mass spectrometry. Because the activity of the radioisotope is extremely low (typically $< 10^{-6}$ of annual background radiation dose), it is safe for human use. Accelerator mass spectrometry is required to determine these radioisotopes because the activity is too low to measure by counting decay particles.

Vitamins

Mass spectrometry has always had considerable, though until recently unrealized, potential in the study of vitamins and their metabolism. The newer LC/MS techniques of APCI and ESI are particularly useful, as many vitamins are too thermally labile or unstable for conventional (EI and CI) mass-spectrometric measurement. The current range of mass-spectrometric techniques can be used to determine all vitamins in foods, biological tissues and fluids. Mass spectrometry is not generally a cost-effective technique for determining vitamins, as alternative, cheaper techniques are usually available. However, mass spectrometry can be used as a 'gold standard' method to validate alternative techniques and is also the method of choice for measuring isotope ratios in human studies of vitamin absorption and metabolism.

Vitamins D_2 and D_3 and their major metabolites have been studied extensively by GC/MS of volatile derivatives; these studies include quantitative determination by isotope dilution. More recently, ESI and particle-beam LC/MS and LC/MS/MS have been used in the qualitative and quantitative measurement of vitamin D and vitamin D analogs.

Recent trends include studies of the metabolism of isotopically labeled folates and carotenoids by GC/MS and LC/MS. There is evidence that folates are associated with a reduced risk of chronic disease (particularly cardiovascular disease) and with prevention of neural-tube defects. Improved mass-spectrometric methods determining these B-group vitamins are desirable, and there is evidence that ESI LC/MS fulfills this requirement, without the necessity for preparing volatile derivatives (a necessity in GC/MS studies). Figure 5 shows the negative ion ESI spectrum of 5-methyltetrahydrofolic acid at low and high cone voltage.

Carotenoid metabolism is also undergoing study by LC/MS techniques because of epidemiological evidence of their putative role in cancer prevention. Although ESI LC/MS has shown promise in preliminary mass-spectrometric studies, APCI LC/MS now appears to be the method of choice because of its robustness and tolerance of a wider range of organic solvents.

Trace Components of Foods

Food Additives

Food additives comprise a wide variety of compounds that are generally monitored by techniques other than mass spectrometry; however, occasional quantitative applications do occur, exemplified by GC/MS determination of antioxidants in stored products.

Biologically Active Naturally Occurring Nonnutrients

Biologically active (bioactive) nonnutrients in foods encompass natural substances that may have deleterious or beneficial effects. This wide-ranging category of food components has not been studied as much as

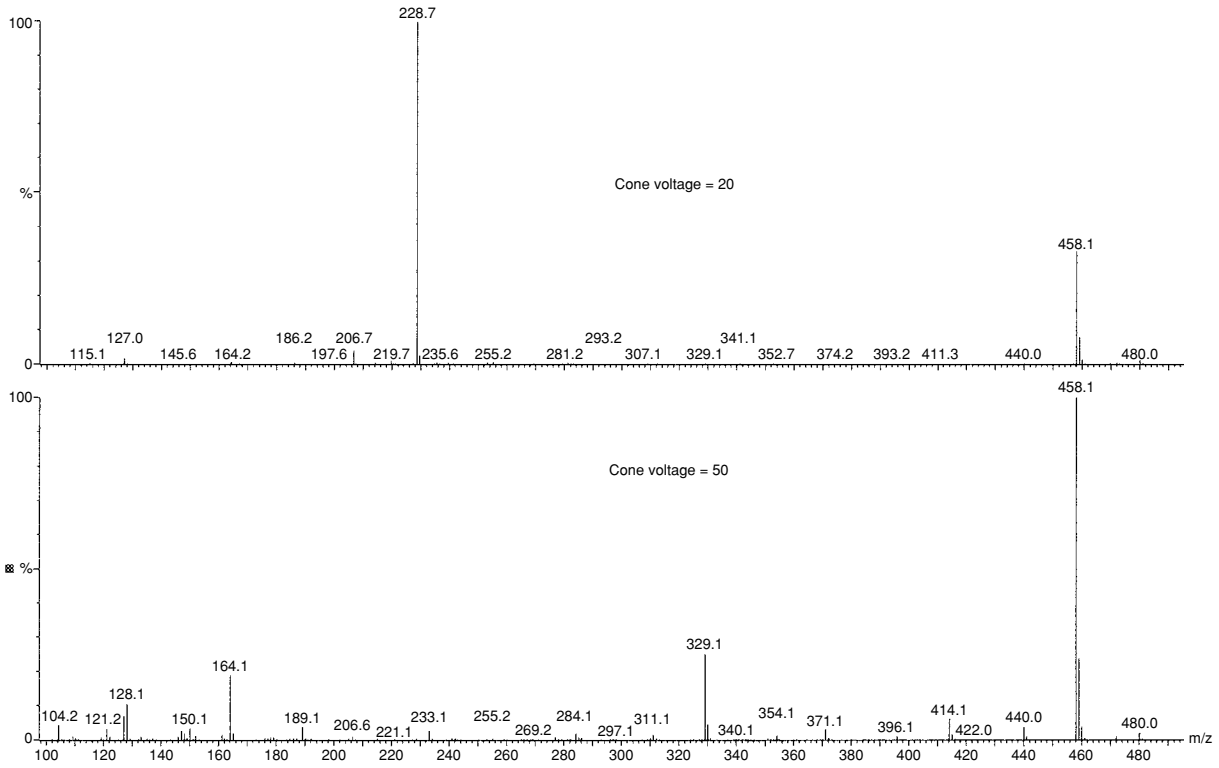

Figure 5 Negative ion ESI spectrum of 5-methyl tetrahydrofolic acid at low (upper trace) and high (lower trace) cone voltages. Fragment ions (e.g., *m/z* 329 and 164) are more intense, and the doubly charged ion at 228.7 is considerably reduced at the higher cone voltage (copyright Institute of Food Research, redrawn with permission).

anthropogenic compounds (e.g., pesticide or veterinary drug residues). However, this situation is now changing for a variety of reasons. These include the amounts present in the diet (e.g., natural pesticides in foods may be found at concentrations 10 000 times higher than synthetic pesticide residues) and the potential health benefits or toxicity of these substances. Because such a wide range of bioactive natural substances are found in foods, it is only possible here to discuss mass-spectrometric analysis of a small range of these compounds.

Glucosinolates (Figure 6) and their biologically active breakdown products are found in many plant foods. More than 100 different types of glucosinolate have been isolated from plants. Although some glucosinolates may have toxic (e.g., goitrogenic) properties, they are known to be potent inducers of Phase II enzymes that protect against carcinogens and other potentially toxic electrophiles.

The structural variability is in the aglycone, R, that may comprise linear or branched alkyl and alkenyl side-chains, alcohols, methylthioalkyl, methylsulfinyl, aralkyl or heterocyclic groups. GC/MS is the most useful mass-spectrometric technique for analyzing *volatile* glucosinolate breakdown products. Negative ion ESI LC/MS is now the method of choice for

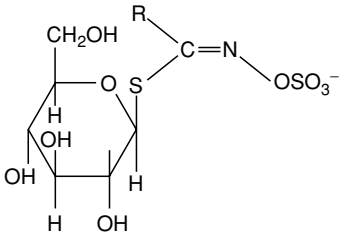

Figure 6 General structure of the glucosinolates.

determining intact glucosinolates and positive ion APCI LC/MS for analyzing the more thermally labile breakdown products (e.g., sulforaphane). ESI LC/MS and MS/MS are also useful for conducting metabolic studies, for example, by detecting and determining glutathione conjugates of isothiocyanate breakdown products of glucosinolates. Figure 7 shows a positive ion ESI mass spectrum of the glutathione conjugate of allyl isothiocyanate, a metabolite of the glucosinolate sinigrin, under low and high cone voltage and MS/MS conditions. These data are useful in developing qualitative and quantitative methods for analyzing glucosinolate metabolites *in vivo* and *in vitro*.

Phenolics are a distinctive feature of all plant tissues and are of interest because they can affect

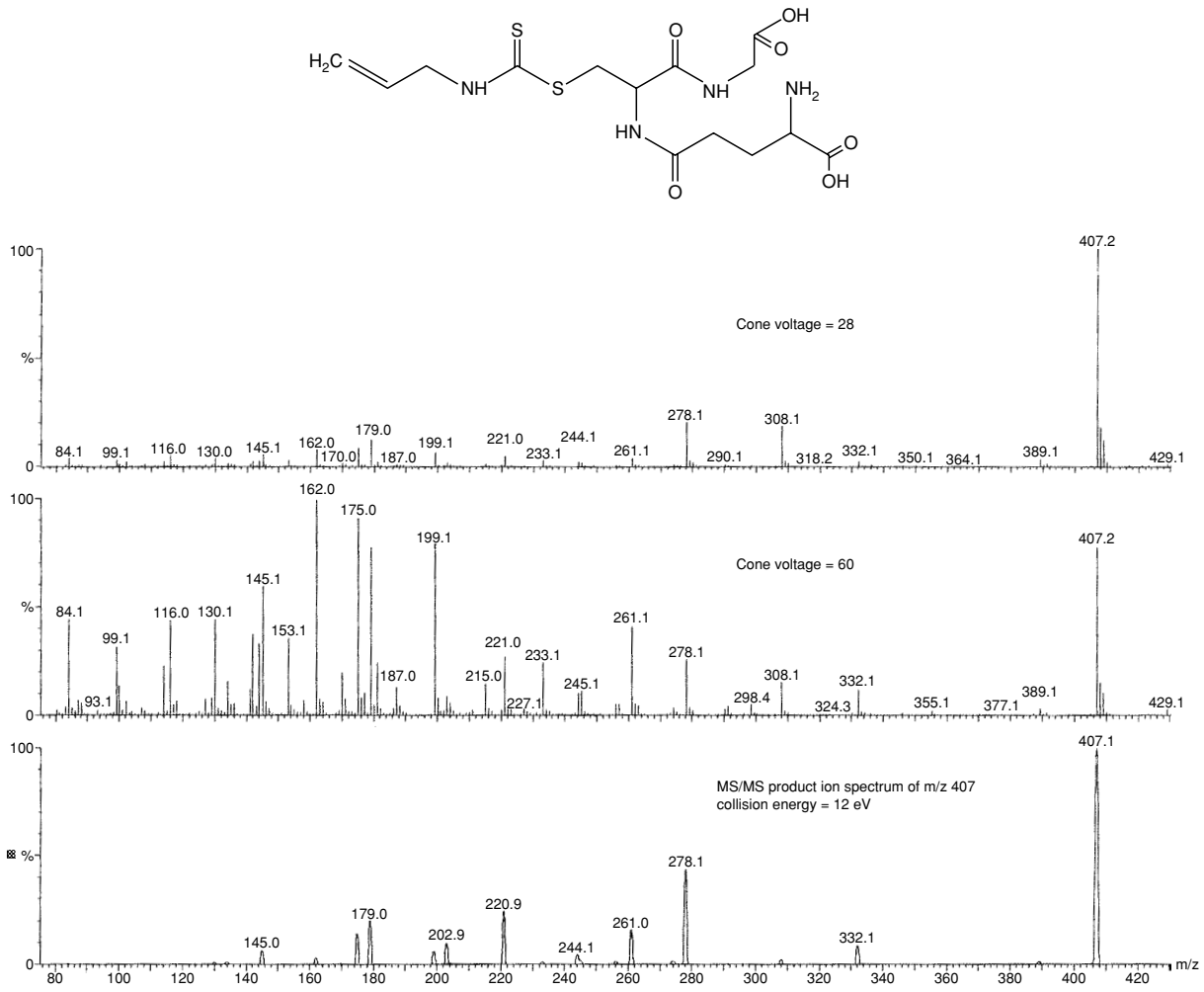

Figure 7 Positive ion ESI spectrum at low cone voltage (upper trace), high cone voltage (middle trace), and the MS/MS product ion spectrum of the $[M+H]^+$ ion (*m/z* 407) of the glutathione conjugate of allyl isothiocyanate, a metabolite of the breakdown products of the glucosinolate sinigrin (copyright Institute of Food Research, redrawn with permission).

the palatability, taste, nutritional value, and particularly the health properties of foods. For example, the flavonoids and isoflavonoids occur very widely in plants, and there is great interest in their role as protective factors in the diet. More recently, the modern methods of LC/MS analysis APCI and electrospray have been applied to the detection and quantification of flavonoids and isoflavonoids. Although early mass-spectrometric techniques for analyzing these molecules focused on GC/MS methods (after appropriate derivatization), the newer LC/MS methods are now coming to the fore. For example, isoflavones and their conjugates have been determined in soy foods (including soy beans and soy flour) by positive and negative ion APCI LC/MS. These methods also offer great potential for conducting metabolic studies.

Mycotoxic secondary metabolites produced by *Aspergillus* (aflatoxins) and *Fusarium* (tricothecenes, etc.) food-spoilage molds comprise two of the most notorious groups of natural toxicants. Many different mass-spectrometric techniques, including EI, positive and negative ion CI, GC/MS, LC/MS, supercritical fluid MS, and MS/MS have been devised to monitor the levels of mycotoxins in foods, body fluids, and tissues. The most recent developments include ESI LC/MS determination of aflatoxins down to low picogram levels. Additional selectivity (at the cost of loss of sensitivity) is provided by MS/MS selected reaction monitoring.

Anthropogenic Toxicants

Many examples of mass-spectrometric methods for determining toxic or potentially deleterious anthropogenic compounds in foods can be found in the scientific literature. The range of compounds analyzed is wide and includes dioxins, polyaromatic hydrocarbons, pesticide and veterinary drug residues,

plasticizers from packaging materials, and environmental contaminants. GC/MS and GC/MS/MS techniques are used widely for determining these compounds down to p.p.b. or even p.p.t. levels.

Isotope dilution mass spectrometry is an accurate and sensitive technique for determining toxic trace elements in food matrices. Lead, cadmium, and thallium have been analyzed down to very low levels by TIMS and ICP-MS. The latter technique is particularly useful for simultaneous measurement of a wide range of elements. Because the toxicity of an element can be highly dependent on its chemical form (for example, organomercury compounds are more toxic than inorganic mercury), ICP-MS is also useful in the speciation of toxic minerals in foodstuffs by combination with HPLC or size-exclusion chromatography.

See also: **Amino Acids**: Determination; **Carbohydrates**: Determination; **Fatty Acids**: Analysis; **Mass Spectrometry**: Principles and Instrumentation; **Peptides**; **Protein**: Determination and Characterization; **Sensory Evaluation**: Texture; Aroma; Taste; **Vitamins**: Determination

Further Reading

Gilbert J (ed.) (1987) *Applications of Mass Spectrometry in Food Science*. Amsterdam: Elsevier.

Mellon FA and Sandström B (1996) *Stable Isotopes in Human Nutrition: Inorganic Nutrient Metabolism*. London: Academic Press.

Mellon FA, Self R and Startin JR (2000) *Mass Spectrometry of Natural Substances in Food*. London: Royal Society of Chemistry.

Mayonnaise *See* **Dressings and Mayonnaise**: The Products and Their Manufacture; Chemistry of the Products

MEAT

Contents

Sources

A J Kempster, Formerly of Meat and Livestock Commission, Milton Keynes, UK

Meat Species

Meat is defined as the flesh of animals used as food. In practice this definition is restricted to a few dozen of the 3000 mammalian species, but it is often widened to include, as well as the flesh, organs such as liver and kidney, brains, and other edible tissues. A major part of the meat consumed in the world is derived from pigs and cattle (**Table 1**). Mutton and lamb and goat meat also make a significant contribution. Poultry meat is important and growing rapidly, at the rate of 4% per annum. Refer to individual sources.

In addition, substantial quantities of horse meat, buffalo meat, and venison are consumed; and various other mammalian species are eaten in different parts of the world according to their availability or because of local custom. Thus, for example, the seal and polar bear are important in the diet of Eskimos, and a wide

Table 1 World meat production (1989)

	Million head	*Million tonnes*	*Change in weight produced over 1980 (%)*
Beef and veal	237	49.0	+12
Buffalo meat	11	1.5	+68
Mutton and lamb	443	6.5	+13
Goat meat	204	2.4	+31
Pig meat	882	67.2	+30
Horse meat	NA	0.5	−4
Poultry meat	NA	37.8	+43

Reproduced from Food and Agriculture Organization of the United Nations (1990) *Production Yearbook*, vol. 43. Rome: FAO, with permission.
NA, information not available.

range of species in that of tribes of Central Africa: the kangaroo is eaten by aborigines and the whale is eaten in Norway and Japan. (*See* **Marine Foods**: Marine Mammals as Meat Sources.)

The vast majority of meat derives from domesticated animals. Domestication of cattle followed the establishment of settled agriculture about 5000 BC. Domesticated hump-backed cattle (*Bos indicus* Zebu) existed in Mesopotamia by 4500 BC. Major developments of breeds have taken place, with important differences in characteristics between parts of the world.

Pigs were not domesticated before the permanent settlements of Neolithic agriculture. They are descendants of wild pigs, of which the European representative is *Sus scrofa* and the eastern Asiatic representative is *S. vittatus*. There is definite evidence for their domesticity by about 2500 BC in Europe.

Domestic sheep belong to the group *Ovis aries* and appear to have originated in eastern Asia. By 3000 BC several breeds of domestic sheep were well established in Mesopotamia. The domestic goat derives from the wild goat, *Capra aegagrus*, and was probably the first ruminant to be domesticated, some time before 7000 BC in Iran.

Several other bovine species and bison have been domesticated and make a small contribution, together with camels, llamas and alpacas, rabbits, reindeer, and some rodents.

In order to classify meat further, consideration has to be given to the carcass from which it comes. This has a major impact on the size and composition of prepared cuts. The next sections will, therefore, be concerned with carcass composition and evaluation.

Selection for Meat Production

Over the centuries, the domesticated species have been selected to produce carcasses with high meat content, using visual appearance and tactile assessments. Historically, carcass evaluation has been dominated by considerations of size and shape. Carcass shape has been considered very important, its influence stemming from the esthetic appeal of certain types of stock. For example, in the last century, the traditional British beef breeds, Hereford, Aberdeen Angus, and beef shorthorn, were all altered in shape to fit a human idea of what a 'good' beef animal should look like. In making this change to blocky, rectangular animals, breed producers and meat traders have come to believe that these shapes are associated with more meat and better-quality meat. More recently, selection has focused on producing larger, leaner animals to meet consumer demand. This has been particularly successful with the pig in producing a dramatic increase in leanness in the European and North American pig populations.

Detailed compositional studies of carcass composition, involving tissue separation, are a relatively new phenomenon. About 70 years ago, Hammond and coworkers at Cambridge University began to examine the influence of stage of growth and level of feeding on proportions of lean (muscle), fat, and bone in the carcass. Some 40 years ago, Butterfield in Australia demonstrated that differences in shape between breeds are not reflected in important differences in the way total muscle weight (of a particular species) is distributed between different joints of the carcass.

Today, the principal trend in the western world is to reduce the proportion of fat in the carcass because of concern about the effects of high fat consumption, especially of saturated fat on health. However, there is a conflict here with the need to maintain sufficient fat in the meat to insure good eating quality, and a minimum fatness level is often required.

Carcass Structure and Composition

The type of meat from a carcass is influenced by the carcass from which it comes, in particular its (1) weight, (2) proportions of the main tissues (muscle, fat, and bone), (3) distribution of these tissues through the carcass, (4) muscle thickness, (5) chemical composition, and (6) meat quality.

Weight

The weight and size of a carcass have a major influence, not only on the quantity of the various tissues, but also on the size of the muscles exposed on cutting and of the individual joints prepared from it. This is of importance particularly in relation to a retailer's ability to provide cuts of suitable size for customer requirements. Over generations, the meat industries in different countries and the different regions of the same country have become accustomed to handle

certain weight ranges of carcasses; abattoir practices and cutting methods have been developed accordingly. Most wholesalers state desired weight ranges in buying schedules for their producers and apply discounts to carcasses falling outside these ranges. For example, in the UK pork has traditionally been produced from pigs of relatively light carcass weights; in the USA, lamb has been produced from relatively heavy carcasses.

Proportions of the Tissues

Among carcasses of similar weight, the percentage formed by each tissue varies considerably depending on breed type and growth rate. The proportion of lean meat in the carcass is of major importance since this is the principal determinant of meat yield and commercial value. Certainly, among western consumers, leanness is the criterion by which most consumers judge quality and value for money. Taken as a generalized ideal, the best carcasses should have a maximum lean content, an optimum level of fatness and minimum bone. The meat from the carcass should follow the same ideal but with most bone removed before sale. There are many different commercial techniques for assessing carcass composition.

Distribution of Tissues

The distribution of tissues through the carcass is potentially important because there are wide differences in the quality and retail value of meat from different regions of the carcass, largely influenced by the degree of tenderness and the type of cooking normally required. For example, beef fillet steak (the psoas muscle) is worth nearly three times as much as stewing steak at present in the UK. However, muscle weight distribution is a fairly constant characteristic and there is little variation to exploit commercially. On the other hand, the distribution of fat between different depots in the carcass varies considerably. The general distribution of fat is important because it influences the overall efficiency of meat production: fat in the body cavity or excess fat trimmed during retail preparation is of little commercial value in comparison with that sold as part of retail cuts. The position of fat in the carcass is also important because subcutaneous fat can be trimmed more easily than intermuscular fat and is, therefore, preferable in carcasses containing fat in excess of consumer requirements. Indeed, excess intermuscular fat cannot be trimmed from some joints, especially from lamb carcasses, without mutilating them. The evenness of fat distribution is also important because wedges or bulges of fat in some joints can lead to excessive trimming or devaluation. Intramuscular fat (that within the muscles) is important in insuring that the meat is juicy and tender.

Muscle Thickness

While this characteristic appears to vary considerably from carcass to carcass, much of this variation can be attributed to weight and fat content variations. Among carcasses of similar weight and fatness, 'blockier' carcasses will tend to have thicker muscles, but this remaining variation has little impact on retail realization values. Retailers tend to favor carcasses with good meat thickness, however, as this is thought to be associated with a higher yield of saleable meat and to improve the appearance of joints. There may also be advantages in terms of tenderness and reduced weight losses in the preparation and cooking of cuts, but the commercial significance of these advantages is not clear.

Carcass Classification

The above characteristics influence the commercial value of carcasses and national and international carcass classification schemes have been developed to describe carcasses. These have been used for grading purposes (encouraging the production of meat most suited to consumer requirements) and to facilitate trade at a distance. Techniques range from simple visual assessments of fat cover and shape to much more sophisticated objective measurement devices.

Carcass Anatomical Structure and Different Cuts of Meat

The carcass is composed of hard and soft tissues. Bones, and to some degree cartilage, form the hard tissues, while muscle, fat, and connective tissue form the soft tissues. Important differences exist between the main meat-producing species in size and bone structure which influence the types of cut produced.

The largest component of the carcass is muscle. Excluding the head, there are over 100 different muscles, which can be grouped in a variety of ways and in different degrees of detail. The distribution and structure of muscles have a major impact on the cuts of meat produced. Approaches to commercial cutting differ markedly between countries. In many countries, and especially the English-speaking world, particular joints are made by cutting across muscles, so joints as sold will consist of a mixture of muscles differing markedly in expected eating quality. In other countries, notably France and Belgium, commercial cutting is more anatomically based, the muscles are trimmed of fat and often individually

separated, allowing cuts as sold to consist of more homogeneous groupings of muscles.

In commercial practice, muscle or lean is usually sold with some adhering fat. This mixture of lean and fat is referred to as 'saleable meat' or 'trimmed deboned cuts,' although, depending on the species, cut, and country, saleable meat may contain some bone. For example, lamb cuts in the UK are commonly sold bone-in. In developed countries, there is a general trend for meat to be sold deboned as prepared retail cuts or in more processed forms.

The ratio of lean to fat in the meat sold depends on the fat content of the carcass from which the cuts are derived, the nature of the cut, and the preferences of retailers and their customers. When customers self-select their purchases from retail display cabinets, there is a worldwide tendency for those cuts with high lean-to-fat ratios to be preferred, other things being equal; but there is still much variation in the fat content of the carcasses that butchers select and in the amount of fat they trim in the preparation of retail cuts. Consumer attitude studies show that many consumers in developed countries do not like fat and that the preference for lean meat and products is growing. Fat requires extra labor for trimming and, in any case, fat is extremely expensive to produce. There can be little doubt that market forces will filter back to producers and economics will favor the production of leaner-meat animals.

In less-developed countries where meat consumption is low, all parts of the animal are often considered to be of equal value (certainly as far as boneless meat is concerned), and the degree of processing is relatively low. There is also a greater tolerance for fat.

See also: **Marine Foods**: Marine Mammals as Meat Sources

Further Reading

Butterfield RM (1963) Relative growth of the musculature of the ox. In: Tribe DE (ed.) *Carcase Composition and Appraisal of Meat Animals*. CSIRO: Melbourne.

Food and Agriculture Organization of the United Nations (1990) *Production Yearbook 1989*, vol. 43. Rome: FAO.

Kempster AJ (1990) Marketing procedures to change carcase composition. In: Wood JD and Fisher AV (eds) *Reducing Fat in Meat Animals*, pp. 437–458. London: Elsevier Applied Science.

Kempster AJ, Cuthbertson A and Harrington G (1982) *Carcase Evaluation in Livestock Breeding, Production and Marketing*. Granada: London.

Mason IL (ed.) (1984) *Evolution of Domesticated Animals*. Longman: London.

Sissons S and Grossman JD (1975) *The Anatomy of the Domestic Animals*. London: WB Saunders.

Woodward J and Wheelock V (1990) Consumer attitudes to fat in meat. In: Wood JD and Fisher AV (eds) *Reducing Fat in Meat Animals*, pp. 66–100. London: Elsevier Applied Science.

Structure

C R Calkins and K M Killinger, University of Nebraska Lincoln, Lincoln, NE, USA

Introduction

Meat is the flesh of animals used for food. The primary components of meat include muscle, fat, and connective tissue. Of these, muscle is the major food component. The unique composition and structure of muscle allow it to perform a variety of physiological functions and give rise to the characteristic properties of meat.

Composition

Moisture (water), protein, and lipid (fats) are the primary constituents of muscle. Their proportions can be quite variable, as are their functions. Additional trace minerals, carbohydrates, and inorganic compounds are located within the tissue, but contribute little to the gross structure.

Water makes up 68–80% of muscle weight; much of this is located within the individual muscle cell. There it functions as the liquid medium (sarcoplasm), as a coolant, and as a transporter of nutrients and waste products. Water-soluble compounds essential for life are distributed throughout the sarcoplasm.

Proteins contribute 15–22% of muscle weight. The numerous and varied types of protein provide critical features to muscle. Proteins are frequently classified on the basis of solubility. Sarcoplasmic proteins are water-soluble. Myofibrillar proteins are salt-soluble and contribute to the contractile machinery (located within the myofibril). In contrast, stromal proteins create a fibrous, structural element within the tissue and are relatively insoluble. Another nomenclature reflects the specific function of proteins. Under this system, proteins are classified as contractile, regulatory, or cytoskeletal (**Table 1**).

Lipid content is the most variable of all meat constituents, making up 0.5–20% (or more) of the weight. Lipids are located within cell membranes, adipose

Table 1 Location and function of muscle proteins

Protein	*Location in sarcomere*	*Proposed function*[a]
Major contractile proteins		
Myosin	Thick filaments	Contraction
Actin	Thin filaments	Contraction
Regulatory proteins		
Tropomyosin	Thin filaments	Regulates contraction
Troponin	Thin filaments	Regulates contraction
α-Actinin	Z-line	Anchors thin filaments from opposing sarcomeres in the Z-line
Cytoskeletal proteins		
Titin	Extends from M-line to Z-line	Protects sarcomere from overstretch
C-protein	Thick filaments	Inhibits skeletal muscle myosin adenosine triphosphatase (ATPase)
H-protein	Thick filaments	Inhibits skeletal muscle myosin ATPase
F-protein	Thick filaments	Binds to thick filaments, inhibited by C-protein
I-protein	Thick filaments	Inhibits Mg^{2+}-ATPase activity of actomyosin in the absence of Ca^{2+}
Nebulin	Thin filaments	Controls thin filament length
Tropomodulin	Thin filaments	Blocks depolymerization of actin filaments
β-Actinin/CapZ	Thin filaments	Anchors thin filaments in Z-line
M-protein	M-line	Forms the transverse portion of the M-line in fast-twitch fibers
Creatine kinase	M-line	Binds to M-protein
Myomesin	M-line	Anchors titin in the M-line and binds myosin
Skelemin	M-line	Attaches M-lines of adjacent myofibrils
Cypher	Z-line	Binds at Z-line
Telethonin/T-cap	Z-line	Binds titin in Z-line
Desmin	Z-line periphery	Transversely links myofibrils at Z-line
Filamin	Z-line periphery	Transversely links myofibrils at Z-line
Vimentin	Z-line periphery	Transversely links myofibrils at Z-line
Synemin	Z-line periphery	Links desmin to Z-line, modulates assembly/disassembly of intermediate filaments
Dystrophin	Near the sarcolemma	Stabilizes sarcolemma and binds actin
Utrophin	Neuromuscular junction	Immobilizes acetylcholine receptor clusters
Dystrophin-associated proteins	Sarcolemma	Family of proteins involved in anchorage of dystrophin in the sarcolemma
Vinculin	Near the sarcolemma	Attachment of actin to the sarcolemma
Talin	Near the sarcolemma	Attachment of actin to the sarcolemma
Paxillin	Near the sarcolemma	Adapter molecule for signal transduction
Tensin	Near the sarcolemma	Anchors actin and maintains tension
Integrin	Sarcolemma	Family of mechanoreceptors, anchors some cytoskeletal proteins in the sarcolemma
Spectrin	Near the sarcolemma	Binds to actin and stabilizes the sarcolemma
Ankyrin	Near the sarcolemma	Attachment of spectrin to the sarcolemma

[a]Some proteins listed have several functions, while some functions have not been fully characterized.
Adapted from Pearson AM and Young RB (1989) *Meat and Muscle Biochemistry*. San Diego, CA: Academic Press.

cells, or the sarcoplasm of muscle cells. Lipids are a source of energy in living muscle and also serve a vital role in cell membrane structure.

Muscle Structure

Contraction must be translated from the cellular level to gross muscle movement. This is accomplished through the integration of skeletal muscle contraction with a connective tissue harness that traverses the entire muscle down to the individual muscle cells (Figure 1). The connective tissue harness is primarily composed of the protein collagen. The epimysium, the outermost connective tissue layer, surrounds the individual muscles themselves. Individual muscles contain large groups of muscle cells (called bundles) which are surrounded by the perimysium, the middle connective tissue layer. Finally, the individual muscle cells or myofibers are surrounded by the endomysium, the innermost layer of connective tissue. Myofibers, which are usually several centimeters in length, rarely transverse the entire length of a muscle.

Muscle cells contain the organelles typically associated with eukaryotic cells; however, some muscle cell structures merit attention. The prefix 'sarco' is derived from the Greek word for flesh and is often used to denote certain muscle cell structures, including the sarcolemma (cell membrane), sarcoplasm

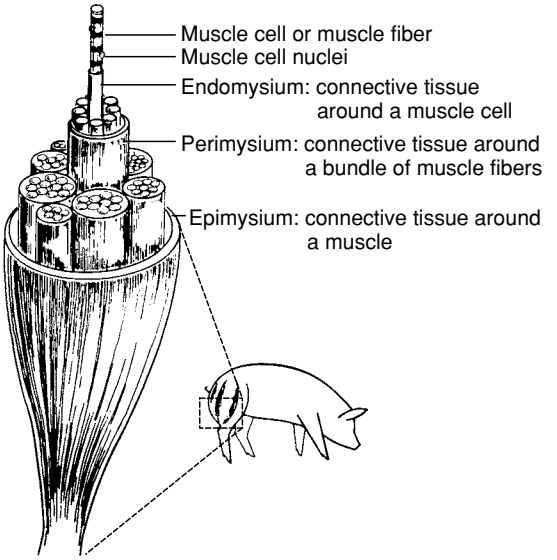

Figure 1 Connective tissue organization of muscle. Courtesy of Dr JE Novakofski, University of Illinois Urbana-Champaign.

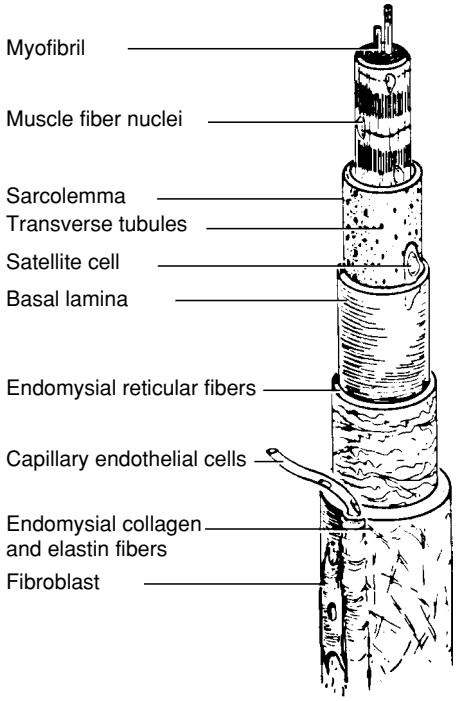

Figure 2 Organization of a muscle cell. Courtesy of Dr JE Novakofski, University of Illinois Urbana-Champaign.

(cytoplasm), and sarcoplasmic reticulum (endoplasmic reticulum).

The structure of the sarcolemma is unique in that it has two sturdy outer layers, the reticular lamina and the basal lamina, in addition to the lipid bilayer (Figure 2). The sarcolemma also has invaginations which reach deep into the muscle cell. These invaginations are known as the transverse tubules or t-tubules. In mammalian, striated skeletal muscle, the t-tubules are located at the junction of the dark and light bands, thus each sarcomere is associated with two t-tubules. The terminal cisternae of the sarcoplasmic reticulum is connected to the t-tubules via protein bridges called ryanodine receptors (Figure 3). The sarcoplasmic reticulum regulates the dynamic pool of calcium within the muscle cell.

Skeletal muscles are unique because they are multinucleated, with the nuclei typically located between the sarcolemma and the underlying myofibrils. Nuclei contain the genetic material that regulates protein synthesis within the muscle. Numerous mitochondria are also found in muscle cells and are located just beneath the sarcolemma as well as between myofibrils. Mitochondria, the powerhouse of the cell, synthesize energy compounds necessary for cellular functions. Muscle cells that perform more aerobic metabolism tend to contain more mitochondria. The sarcoplasm and the myofibrils make up most of the muscle cell volume.

When viewed under a light microscope, skeletal muscle exhibits a characteristic pattern of alternating dark and light bands that are parallel to the long axis of the myofiber. This pattern gives rise to the term 'striated, skeletal muscle.' The dark bands, called A-bands, are dense, birefringent, and anisotropic, which means they do not allow the passage of light. The light bands, called I-bands, are less dense and are isotropic (allow the passage of light). The I-bands are bisected by a dark line perpendicular to the long axis of the myofiber, which is called the Z-line.

Within a myofiber, the aggregation of contractile proteins into long filaments creates myofibrils (Figure 2), which contain sarcomeres, the functional contractile unit of the muscle cell. Each sarcomere is approximately 2–2.5 μm and spans from Z-line to Z-line (Figure 4). The sarcomere is composed of a series of overlapping protein filaments. The thick myofilaments, located within the A-band, contain myosin as the predominant protein. The thin myofilaments, predominantly located within the I-bands, are composed mostly of actin. When a myofiber is at rest, the thin and thick myofilaments overlap slightly in the sarcomere. The H-zone spans the central portion of the sarcomere where the thin and thick myofilaments do not overlap. Within the H-zone, the pseudo-H-zone lies over the portion of the thick myofilament which does not contain the globular heads of myosin. Additional structural and regulatory proteins are also found in the center of the A-band in the centrally located M-line, which serves to maintain the three-dimensional alignment of the thick myofilaments.

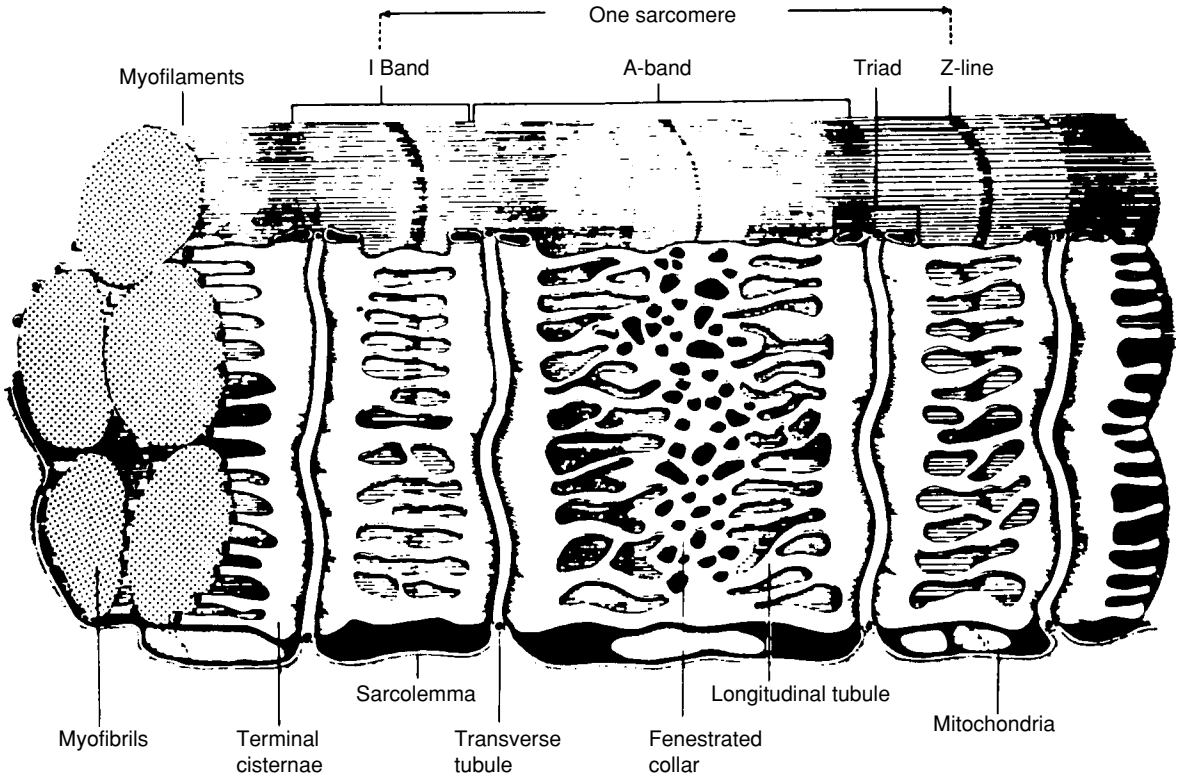

Figure 3 Diagram of t-tubules and sarcoplasmic reticulum in longitudinal view of skeletal muscle. From Pearson AM and Young RB (1989) *Meat and Muscle Biochemistry*. San Diego, CA: Academic Press, with permission.

Myofibril cross-sections have different appearances depending on the location within a sarcomere (Figure 4). In the I-band, the thin myofilaments (smaller in diameter) form off-set rows, with every other row being aligned. In the H-zone, the thick myofilaments (larger in diameter) form a similar pattern of off-set rows, with every other row being aligned. Within the area where the A- and I-bands overlap, the thin myofilaments surround the thick myofilaments. Each thick myofilament has access to six thin myofilaments, and each thin myofilament is accessible to three thick myofilaments.

Function

The proteins associated with the M-line are believed to provide support to the thick myofilaments and to help maintain the structural integrity of the sarcomere. The proteins in the Z-line interior are believed to help anchor the thin myofilaments, while the proteins at the periphery of the Z-line are believed to play a role in attaching adjacent myofibrils.

The costameres are rib-like structures encircling the myofiber. They overlie the I-bands on each side of the Z-line (Figure 5). The costamere structure is comprised of several proteins (like vinculin, dystrophin, and spectrin), some of which are closely associated with the sarcolemma and others which are filamentous strands that appear to associate with the Z-line. Some of the proteins found in costameres also appear to connect the M-line to the sarcolemma, although this connection has not been fully characterized. Costameres are integral in the attachment of the sarcolemma to underlying myofibrils. In addition, they help transmit the force of contraction to the extracellular matrix and laterally along the myofiber.

The thin myofilament is predominantly composed of actin. Actin monomers, called globular actin or G-actin, are almost spherical in shape and polymerize to form two strands of actin that wrap around each other in a helical fashion to form the bulk of the thin myofilament. The strands of polymerized actin are called filamentous actin or F-actin. Each G-actin molecule contains a myosin-binding site that is critical for contraction to occur. A rod-shaped protein, tropomyosin, winds around the strands of actin and blocks the myosin-binding site on the actin molecules in muscle at rest. Troponin, a second regulatory protein, is located at regular intervals along the thin myofilament. Troponin is composed of 3 subunits. Troponin T (TnT) binds to tropomyosin and to

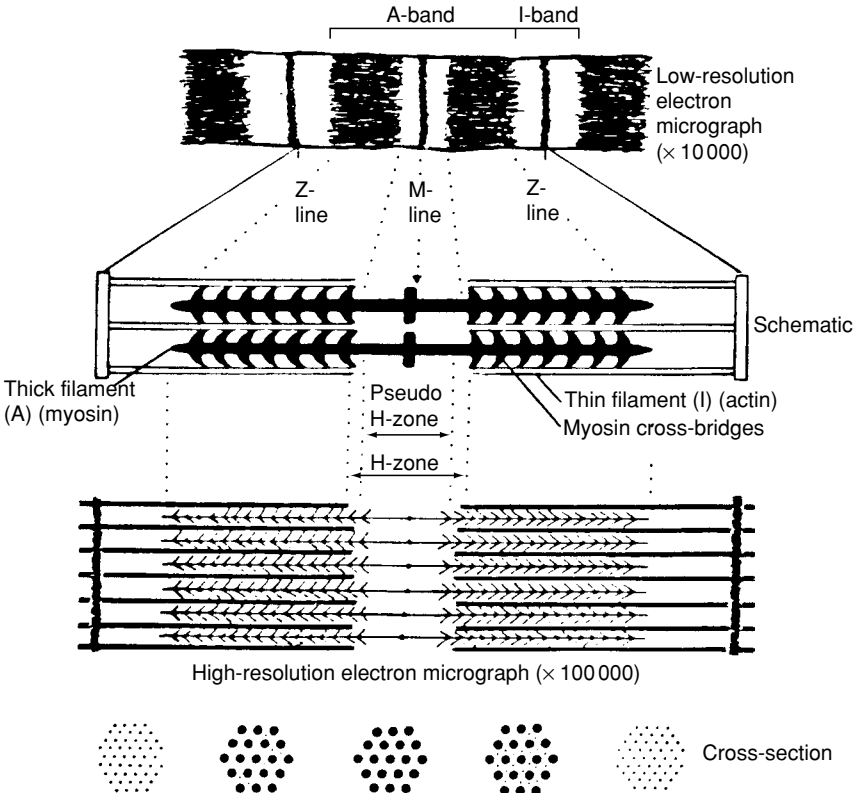

Figure 4 Fine structure of a sarcomere along with portions of two adjacent sarcomeres. From Pearson AM and Young RB (1989) *Meat and Muscle Biochemistry*. San Diego, CA: Academic Press, with permission.

troponin C (TnC), which is the calcium-binding subunit. Troponin I (TnI), the inhibitory subunit, connects TnC to the thin myofilament. At sufficient levels of calcium, the interaction between the troponin, tropomyosin, and actin exposes the myosin-binding site of actin.

In the thick myofilament, the best-characterized protein is myosin. Myosin has been called a mechanoenzyme as it not only plays a structural role in the thick myofilament, but also acts as an adenosine triphosphatase (ATPase). Myosin's globular head has both enzymatic and structural roles. It breaks down adenosine triphosphate (ATP) to provide the energy necessary for contraction, while it also interacts with actin to form a cross-bridge between the two proteins. The tail portion of the myosin molecule plays a structural role, and is involved in the formation of the thick myofilament. Other proteins in the thick myofilament help maintain the structure of the myofiber and also play regulatory roles.

During contraction, the globular head of myosin breaks down ATP to adenosine diphosphate (ADP) and inorganic phosphate, binds to actin, and subsequently changes the angle of the actomyosin cross-bridge from 90° to 45°. When another molecule of ATP binds to actomyosin, myosin detaches from actin. This process is cyclical and results in the thin myofilament sliding inward past the thick myofilament. Contraction ceases when calcium ions are sequestered back into the sarcoplasmic reticulum. Contraction of an antagonist muscle is required to return the original muscle to resting length.

The signal for muscle to contract begins as a wave of depolarization, or action potential, that sweeps down a nerve cell. When the action potential reaches the neuromuscular junction (the connection between a nerve cell and a muscle fiber), acetylcholine is released from the nerve cell, crosses the synapse (the gap between the nerve and muscle cell), and attaches to receptors located on the sarcolemma. The sarcolemma depolarizes in a similar manner to the nerve cell, and the message is carried into the t-tubule system in order to carry the message deep into the myofiber. The sarcoplasmic reticulum is stimulated by the t-tubules to release calcium which interacts with regulatory proteins to initiate contraction.

All myofibrils within a muscle cell contract in concert. By moderating the number of muscle cells recruited for contraction, the strength of a particular muscle can be controlled. Thus, fine motor movements and considerable physical strength can both be demonstrated by the same muscle.

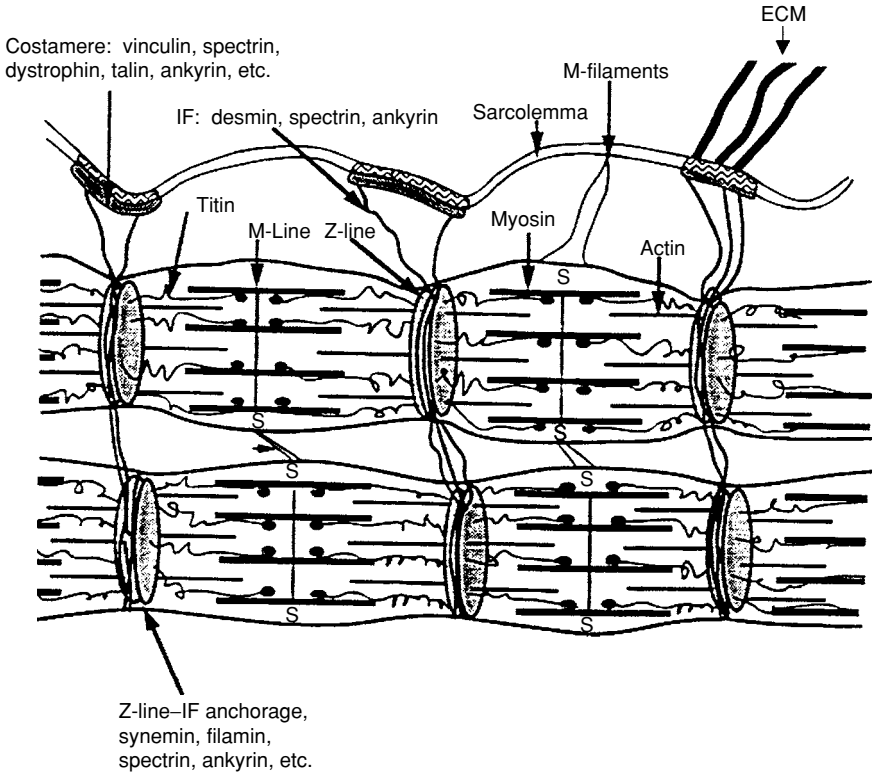

Figure 5 Structural arrangement of costameres, intermediate filaments, and myofibrils. ECM, extracellular matrix; IF, intermediate filaments; S, skelemins. Adapted from Price MG (1991) Striated muscle endosarcomeric and exosarcomeric lattices. In: Malhotra SK (ed) *Advances in Structural Biology*, vol. 1, pp. 175–207. Greenwich, CT: JAI Press.

Postmortem Changes in Meat

In the conversion of muscle to meat, several changes occur that impact meat quality. During slaughter and processing, the muscle cell environment undergoes changes like: declining energy levels, pH, and temperature, increasing ionic strength, and changes in redox potential. The pH declines as lactic acid is produced as a byproduct of anaerobic metabolism. Lower muscle pH results in reduced water-holding capacity and influences meat color.

An important contributor to the increase in ionic strength is calcium. As energy sources are depleted, the sarcoplasmic reticulum loses its ability to sequester calcium. With increased calcium levels in the sarcoplasm, conditions for contraction exist (as long as sufficient energy is available). Eventually, the cross-links between the thick and thin myofilaments (myosin and actin) are sustained. These attachments (rigor bonds) cause muscle rigidity (rigor mortis) and reduce meat tenderness. Shorter sarcomeres have a greater number of rigor bonds and create less tender meat.

Meat is usually held at refrigeration temperatures for a time postmortem to modify meat palatability (flavor and tenderness). This storage period is called aging. Meat undergoes several structural changes during this period which lead to an increased fragility of the myofibrils (as measured by an increase in meat tenderness). The decrease in the intensity of the Z-line and subsequent loss of Z-line integrity is the most notable change in structure. Additionally, the sarcolemma separates from the underlying myofibrils. Adjacent myofibrils also separate from each other, and some researchers have observed discontinuities in the I-band.

Despite considerable research, a unified theory on the mechanism of postmortem tenderization in meat has not been embraced. Tenderness increases most rapidly and to the greatest extent during the first 3–4 days postmortem. Changes in tenderness are believed to be associated with the degradation of structural proteins in the myofibers. Early researchers attributed postmortem tenderization to the cathepsins, a group of hydrolytic proteins found in the lysosomes, while other researchers have suggested that the change in ionic strength within the cell is the cause of postmortem tenderization. More recently, the calpain system has been implicated as a link in postmortem tenderization. This group of proteases

requires calcium for activation. They maintain activity in the cellular environment that predominates in meat. There are three calpain isoforms: m-calpain, μ-calpain and skm-calpain. All three forms have been identified in skeletal muscle; however, skm-calpain (also called p94) is only found in skeletal muscle. skm-calpain is difficult to purify, and therefore has not been fully characterized. The other two isoforms are activated at different calcium levels, and the typical conditions during postmortem aging favor μ-calpain. The calpain system is subject to control by a powerful inhibitor, calpastatin. It is believed that one molecule of calpastatin can inhibit multiple calpain molecules. The calpains have been shown to degrade several proteins like titin, nebulin, desmin, and TnT, which may influence postmortem tenderization.

Collagen is another contributor to meat tenderness. Collagen molecules are cross-linked in order to stabilize the molecule and to impart the strength associated with connective tissue. In young animals, collagen cross-links are reducible, and therefore less stable. Collagen containing reducible cross-links will hydrolyze to form gelatin if heated in the presence of moisture. The conversion to gelatin, a softer material, results in more tender meat. However, in older animals, the collagen cross-links become heat-stable and nonreducible, thus contributing to tougher meat. Pyridinoline appears to be one of the more significant nonreducible cross-links.

The flavor of meat is determined by two major components: the meaty flavor is associated with protein, while the species flavor is associated with fat. During aging, proteins are altered to impart the characteristic flavor of aged meat. The breakdown of protein and fat produces flavor compounds, including hypoxanthine, xanthine, hydrogen sulfide, ammonia, acetylaldehyde, and acetone. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics.)

The structure of living muscle is inexorably linked to the structure and characteristics of meat; therefore, an appreciation of all three is required to understand meat-based food products.

See also: **Fats**: Production of Animal Fats; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Meat**: Eating Quality; Analysis; **Protein**: Food Sources

Further Reading

Bechtel PJ (1986) *Muscle as Food*. Orlando, FL: Academic Press.

Huff-Lonergan E and Lonergan SM (1999) Postmortem mechanisms of meat tenderization. In: Xiong YL, Chi-Tang H and Shahidi F (eds) *Quality Attributes of Muscle Foods*, pp. 229–251. New York: Kluwer Academic.

Judge MD, Aberle ED, Forrest JC, Hedrick HB and Merkel RA (1989) *Principles of Meat Science*, 3rd edn. Dubuque, IA: Kendall/Hunt.

Pearson AM and Young RB (1989) *Muscle and Meat Biochemistry*. San Diego, CA: Academic Press.

Price MG (1991) Striated muscle endosarcomeric and exosarcomeric lattices. In: Malhotra SK (ed.) *Advances in Structural Biology*, pp. 175–207, vol. 1, Greenwich, CT: JAI Press.

Swatland HJ (1994) *Structure and Development of Meat Animals*. Englewood Cliffs, NJ: Prentice-Hall.

Taylor RG, Geesink GH, Thompson VF, Koohmaraie M and Goll DE (1995) Is Z-disk degradation responsible for postmortem tenderization? *Journal of Animal Science* 73: 1351–1367.

Slaughter

K Troeger, Bundesanstak f■r Fleischforschung, Kulmbach, Germany

Introduction

Meat consumers, especially young people in largely urbanized regions of developed countries, are increasingly demanding that animals be reared, handled, transported, and slaughtered using humane practices. At slaughter, this means that stunning should render animals insensible to pain to insure that they do not suffer needlessly during slaughter. Because of animal welfare considerations in European law, all animals must be stunned before they are slaughtered. The second aspect of slaughtering is the quality of the final product. This article describes technical parameters of the slaughter process in relation to both meat quality and aspects of animal welfare.

Animal Handling

The moving, handling, and unloading of animals within a slaughterhouse should be carried out with care. Good handling procedures will reduce pale, soft, exudative (PSE) meat, bloodsplash, and injury losses. Animals should be unloaded as soon as possible after arrival at the slaughterhouse. The unloading facilities (ramp, bridge, gangway) should be fitted with siderails and the slope should not exceed 20°.

While cattle should if possible be slaughtered immediately after unloading to avoid poorer meat quality (dark-cutting beef or DCB), lairage times for pigs of at least 2 h and at most 4 h have proved to be successful. The shape and construction of the pens

should permit most of the animals to lie against a wall. Holding pens should therefore be long and narrow rather than square. The following conditions of lairaging have proved to work well:

- Group penning: groups of animals should be kept together, not mixed with 'strangers'
- Number of animals per pen: each animal must have enough space to lie down and turn around
- Sprinkling: showering, at least during hot weather, is particularly recommended for pigs
- Water: drinking water must always be available

Cattle and pigs are driven to the stunning place along passageways and races. Sticks and electric prods are used commonly as guiding instruments. According to European law, personnel involved in preslaughter handling must be trained and certified. Suitable passageways allow the use of gates on rollers or guiding boards, so a more gentle driving of pigs is possible. Races should be short, straight, horizontal, and wide enough to enable two animals to walk side by side for as long as possible. Shadows and sparkling reflections should be eliminated. Animals tend to move from a darker area towards a lighter one, but they will avoid blinding light. Animal welfare-relevant problems mostly occur if it is necessary to line up pigs in a single race. A new Danish carbon dioxide (CO_2) stunning system (Backloader) makes it possible to handle pigs in a group into and through the stunning unit. The principle of the new concept is that 15 pigs are driven forward using a series of push-hoist gates or, alternatively, manually driven to the area in groups of 15. There are three main elements to this concept: first, there is an area where the group of 15 hogs is divided into smaller groups (e.g., five); second, there is automatic transfer of these smaller groups to and through the stunning equipment; and third, there is a system for presenting the stunned hogs for shackling and sticking. As the hogs are stunned in groups, the concept assumes that CO_2 stunning will be used and this has several advantages. Hogs do not like to be isolated away from one another and stunning in groups eliminates this problem. Moreover, hogs find restraint stressful and in modern CO_2 systems hogs are not restrained at all during the stunning itself (Figure 1).

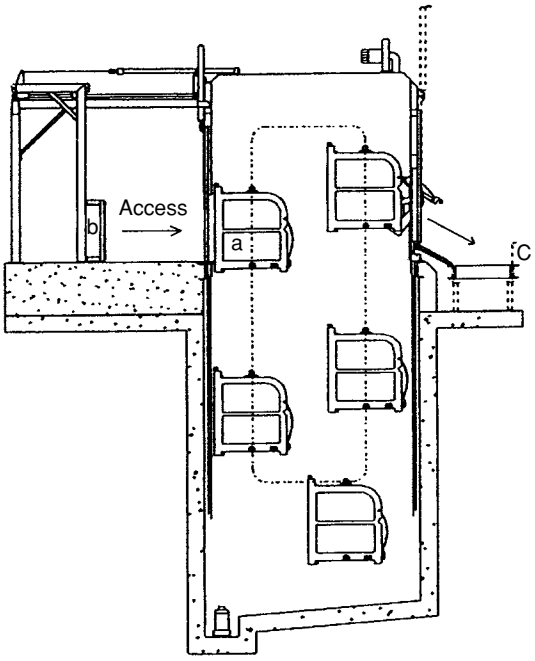

Figure 1 Backloader CO_2 stunning plant for group stunning of hogs (Butina, Denmark). a, Stunning box; b, pushing gate; c, conveyor belt.

Stunning

General Aspects

Stunning aims to bring the animal immediately to a state of insensibility that lasts until it is dead. Another aim is to produce sufficient immobility to facilitate exsanguination. These two aims are not necessarily complementary. Some methods of stunning result in violent reactions by the animals after they have been rendered insensible. Conversely, certain procedures, such as low-voltage application, may produce total-body immobility without causing insensibility. In the light of present scientific and practical knowledge, acceptable stunning methods, if applied properly, are electrical stunning, CO_2 stunning or stunning with alternative gas mixtures (e.g., argon), and percussive stunning (penetrative; nonpenetrative with restriction). Irreversible stunning methods, such as cardiac arrest stunning, induced by an electrical current or longer exposure to gas mixtures, is preferred from the animal welfare point of view as well as for meat quality reasons.

Electrical Stunning

Electrical stunning is important first for slaughtering pigs and poultry; sheep and cattle are also stunned electrically in some countries. The electrodes must be placed so that they span the brain. Effective electrical stunning results in immediate insensibility, which is exhibited as an epileptiform seizure. The effectiveness of electrical stunning depends on a minimum current level passing through the brain. Using a constant stunning current of 1.3 A, for pigs effective stun can be induced within a minimum current flow time of 0.3 s. This time requirement seems to be appropriate

for animal welfare-conforming electrical stunning. There are two main types of electrical stunning: head-only and head-to-body. In head-only systems electrodes are applied only to the head, whereas in head-to-body stunners an (additional) electrode is placed elsewhere on the animal's body (back, breast) so that the flow of current passing through the body should additionally cause cardiac ventricular fibrillation. With this method, all species will be rendered immediately and permanently insensible.

Besides manual stunning with a pair of tongs or single-handed prods, fully automated electrical stunners are available for use on both sheep and pigs. With older systems, animals are fixed and conveyed in a v-conveyor (restrainer) to the point of stunning (bilateral electrode contact on the head). The stunning voltages used in these systems are about 450–1000 V (high-voltage systems). Recent developments have led to improvements in automatic electrical stunning devices. The Midas system (Stork, Netherlands) incorporated a new conveyor belt transport system; the pig rides on this belt with its feet hanging down on both sides and becomes relatively relaxed in this position. This is a good condition for an effective stun. The system has the capacity for 200–600 hogs per hour. It is not a high-voltage system. The induction of cardiac ventricular fibrillation is carried out by an automatic breast electrode, so the animals show little or no convulsions. After a current flow through the head, the third electrode is contacted pneumatically with the left side of the breast and an additional current flows head-to-body.

The current flow through the brain leads to depolarization of neurons, with subsequent repolarization, which causes uncoordinated motor activity (tonic–clonic spasms). These muscle contractions stimulate glycolysis, with production of lactic acid in the musculature. The consequence is an accelerated pH drop postmortem, which favors the incidence of PSE meat. Fewer muscle spasms occur after cardiac arrest stunning. Other adverse effects of electrical stunning can be fractures of bones or bloodsplash (intramuscular hemorrhages) and speckling (petechial hemorrhages in subcutaneous tissues). Factors important for a careful slaughter process including electrical stunning are summarized in Table 1.

Gas Stunning

Gas stunning with high concentrations of CO_2 in air is used primarily for pigs; turkeys and chickens are also sometimes stunned or killed with CO_2. CO_2 is colorless and heavier than air, so stunning can be done in a pit. The gas atmosphere for stunning consists of about 70–95% CO_2 and air. CO_2 has the general effect of depressing neuronal function by both hypoxia and lowering tissue pH < 7.0. There are two main types of CO_2 stunning or killing systems: a dip-lift system, where pigs are lowered continuously into the gas, and the paternoster system, where pigs are lowered successively into the gas with stops as pigs enter or leave the equipment. The paternoster system is most commonly used. The latest development is an automatic Danish system for group CO_2 stunning of up to 800 hogs per hour. This system has the advantage that lining up and restraint of animals are no longer necessary. Up to five pigs enter the stunning box simultaneously. Adequate stunning can be achieved using high CO_2 concentrations (> 90%) and a long exposure time (> 150 s).

The welfare implications of CO_2 stunning are still under debate. Some studies have shown that most pigs will avoid an atmosphere of high concentrations of CO_2. This aversion was found to be greater than the motivation to obtain a reward in the CO_2 atmosphere, even after 24 h fasting. On the other hand, exposure to CO_2 was less aversive than electrical shocks. Scientists have also interpreted the increase in rate and depth of respiration that occurs during inhalation of CO_2 as respiratory discomfort. In contrast to stunning of pigs with CO_2, the animals show no aversive reactions when exposed to gas mixtures of either 90% argon in air or 60% argon and 30%

Table 1 Electrical stunning: important factors for animal welfare and meat quality

Factors	*Criteria*
Minimum of excitement and physical strain	Internal muscle temperature (ham) 45 min post mortem $\leq$ 40.5 °C
Immediate unconsciousness	No vocalization (squealing) if the stunning current is interrupted after 0.3 s
Two-step cardiac arrest system:	Little or no convulsions
1. Constant current ($\geq$ 1.3 A) through the head, followed by	Death within 60 s (no eye reactions, dilatation of the pupils)
2. Head-to-body current	Insignificant bloodsplash
Horizontal bleeding table	Conveying till muscles are relaxed (1.5–2 min)
Short stun-to-stick interval	Horizontal bleeding < 10 s after stunning
Effective sticking	Amount of blood > 3 l per pig

CO_2 in air. The phase before unconsciousness happens without stress. The practicability of these gas mixtures under commercial conditions was tested in the UK.

CO_2 anesthesia results in fewer carcass defects than electrical stunning. The incidence of bloodsplashes, fractures, and PSE meat is lower.

Percussive Stunning

Percussive stunning methods can be divided into two categories: penetrative and nonpenetrative. The first category includes free bullets, captive bolts, and high-pressure water jets. Nonpenetrative methods deliver a controlled obtuse blow to the head (mushroom head, manual blow). In all cases kinetic energy is transferred to the brain, which causes neuronal dysfunction and subsequent unconsciousness. While penetrative methods, which produce a trauma of brain tissue, cause permanent insensibility, unconsciousness after the use of nonpenetrative methods is only temporary.

For penetrative stunning – in general of cattle, sheep, goats, and horses – captive bolt apparatus and pistols are mostly used. These consist of a steel rod with a collar and piston (bolt) at one end contained in a chamber. On activation the bolt is propelled forward by the energy of a cartridge or by compressed air and penetrates the head of the animal. Water jet stunning is the use of a high-pressure water jet with 2500–4000 bar and an injection time of 20–100 ms for stunning or killing slaughter pigs. This method is in an experimental phase. The only method of nonpenetrative percussive stunning commonly used is the 'mushroom head' fitted to a captive bolt and used almost exclusively on cattle, especially in cases of ritual slaughter. The 'mushroom' is a metal convex disk approximately 4 cm in diameter and the impact is applied to the frontal region of the head.

After the correct use of a captive bolt, corneal reflexes cannot be evoked and respiration ceases. The effect of a nonpenetrative percussive technique depends on the size of impulse delivered. Studies have shown that only about 80% of animals are being stunned effectively by such methods.

Percussive stunning, like other stunning methods, causes an increase in blood pressure, violent muscular contractions, and elevated levels of catecholamines. These factors can result in bloodsplash and the occurrence of PSE in pigs. In some European countries, where cases of bovine spongioform encephalopathy (BSE) have occurred, the use of captive bolts for stunning cattle is under debate. There is a clear risk, because of the destruction of brain tissue, including blood vessels, with this method, that in the case of BSE infectious brain tissue might be dispersed into the blood stream, and especially to the lungs.

Killing

Killing of slaughter animals is usually done by exsanguination (bleeding). An incision (sticking) which severs the major arteries of the neck or anterior thoracic cavity causes a rapid loss of blood, resulting in a lack of oxygen to the brain. Irreversible cellular changes occur and the animal dies. Because most stunning methods are reversible as a rule, prompt and accurate (effective) sticking is of high importance from the animal welfare point of view.

Bleeding can be done with the animal in a vertical or horizontal position. The most common technique is to shackle a hind leg and to elevate the stunned animal on the bleeding rail, where sticking is done in a hanging position. There is a significant positive effect on pork quality (reduction of PSE), if prone bleeding is practiced. Sticking is done immediately after electrical stunning on a movable conveyor. For a bleeding time of about 2 min, the pigs are conveyed in a horizontal position. As a result, muscle spasms have a less detrimental effect on meat quality, and internal ham bruising and bloodsplashing are obviated.

Severing of the common carotid arteries or the vessels from which they arise during sticking can be done in different ways. A transverse incision of the neck is a common method of bleeding sheep. This method is also used for ritual slaughtering of sheep and cattle.

Ritual slaughter is performed by a throat cut, which severs all soft tissues of the throat (muscles, blood vessels, esophagus, trachea, nerves) without preslaughter stunning. It is necessary to distinguish between kosher (Jewish) slaughter and halal (Muslim) slaughter. For kosher slaughter, there exist exact rules for the slaughter process, the instruments used, and the qualifications of the slaughterman. A straight, razor-sharp knife (chalaf) that is twice the width of the throat is required, and the cut must be made in a single continuous motion. For halal slaughter, no special knife design is required.

Another sticking method, lateral stab incision of the neck, avoids incision of the trachea and esophagus, because this may result in aspiration of blood and pollution of the blood by stomach contents. Finally, thoracic sticking is commonly used for cattle and pigs. The knife is inserted in the midline, cranial of the sternum bone, so that the bicarotid trunk and anterior vena cava are incised.

Skinning, Scalding/Dehairing, Evisceration, and Splitting

In industrial slaughterhouses skinning, including scalding and dehairing (hogs), evisceration and splitting of carcasses occurs continuously within a slaughter conveyor (Figure 2). The capacities of slaughter lines in Europe may be up to about 80 cattle and 800 pigs per hour. There is a great deal of mechanization, automatization, and robotization in slaughterhouses worldwide.

Skinning (Cattle, Sheep)

After bleeding, the free hanging leg is skinned; this leg is then used temporarily for hanging, so that the slaughter chain can be taken away from the other leg. This leg is skinned too. The gambrel is provided and the belly, chest, and neck part are skinned. Further skinning is done mechanically by means of a skinning machine. For this, the front legs are fixed and the skin is drawn off from the lower to the upper side. This is the traditional method of depelting sheep. A lower carcass microbial level can be achieved if the sheep are depelted from the shoulder to the hind leg; this process is most easily done by suspending the animal from the front legs (inverted system). For cattle, with a roller-mounted hide stripper, skinning is possible either from top to bottom or from bottom to top. If the head remains on the carcass and is skinned too, the carcass has to be stabilized by an electrical current (electrical stimulation with about 150 V, electrodes on the back). A disadvantage of this method is that the force needed for dehiding is so high that overstretching of the hide can happen, which lowers its value for processing.

Scalding, Dehairing, Singeing, Polishing (Pigs)

The conventional process of hair removal from slaughtered hogs utilizes hot water for scalding and a gas flame for singeing. The traditional scalding method is vat scalding. Using a water temperature of 58–62 °C, the scalding process requires 4–6 min. On economic grounds, many hogs are scalded in the same hot water; this results in rapid pollution of the scalding water. From a bacteriological point of view, therefore, it is important that the scalding temperature exceeds 60 °C, but it should not exceed 62 °C because of the coagulating effect on pig skin, which results in damage to the skin by the dehairing machine. A disadvantage of vat scalding is the possibility that scalding water can enter the lungs and, via the sticking wound, into deeper tissues of the carcass. Investigations have revealed that the risk of bacterial internal contamination of lung, liver, heart, and muscles depends on the scalding and dehairing technology. From a hygienic point of view, vat scalding with simultaneous dehairing (mechanical treatment by dehairing implements with pressing and pumping effects on the thorax) was the most dubious method.

An alternative to vat scalding is the scalding tunnel. The pigs are brought through the tunnel by means of

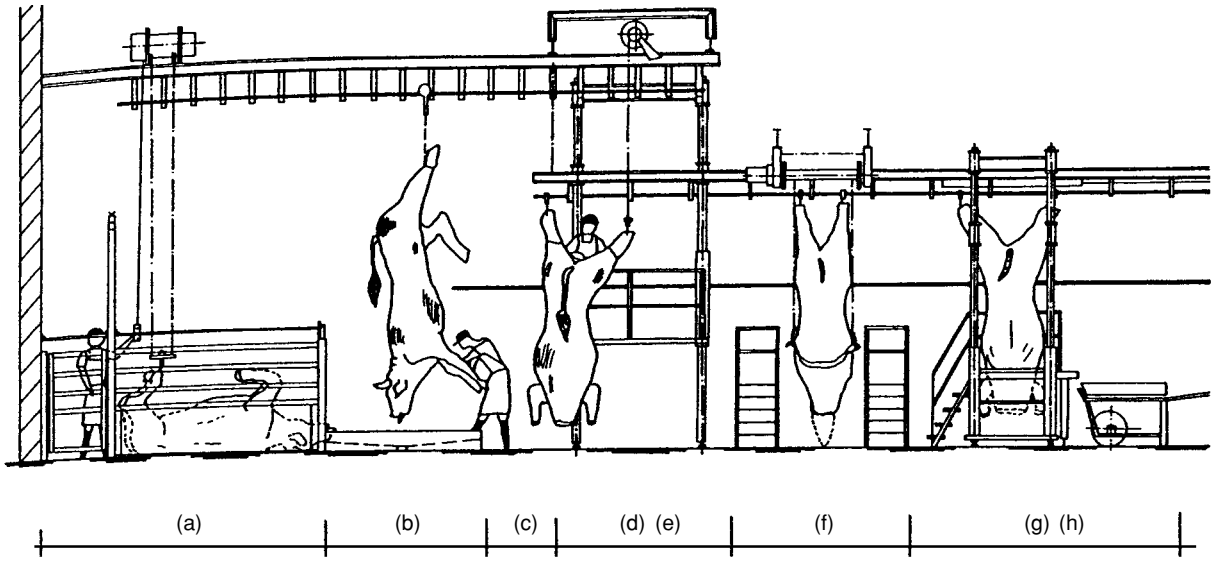

Figure 2 Slaughter line for cattle with hanging up to bleed. (a) Stunning; (b) sticking; (c) decapitation; (d), (e) transfer to slaughter line, removal of hind legs, and preliminary skinning; (f) removal of remaining skin and opening of breast bone by sawing; (g), (h) removal of paunch and breast organs, and cleaving. Reproduced from Meat: Slaughter. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

a chain-conveyor in a hanging position. Scalding is done either by spraying carcasses with hot water that is pumped around the system or by condensed steam. To obtain a low bacterial count on the carcass surface, it is still necessary that the pigs pass through a prewashing machine prior to the scalding tunnel.

After scalding, the next process in a pig slaughter line is dehairing. Dehairing (scraping) machines are constructed from heavy U-shaped bars, a steel frame, and one or two shafts to which belt scrapers are attached. The hogs are sprayed with water as they pass lying and rotating through the dehairer. If the water in the machine circulates, the water temperature should be at least 55–58 °C for hygienic reasons. After dehairing, the pigs are conveyed to the prepolisher. The prepolish machine is needed to remove any remaining hair and to dry the surface of the pig. In this way the effect of the singeing furnace will be optimal and gas consumption is reduced. The singeing furnace is intended to reduce surface bacteria and to singe hair stubble. The surface bacterial count can be reduced by 1–2 powers of 10 by the heat effect. This is not the case if the carcasses are still very moist after singeing. This is followed by the polishing machine, which cleans the singed pigs by brushing. Here the mechanical treatment of the pig skin comes to an end. The washing and polishing process usually causes recontamination of the carcass surface. This is due to the whips inside the washing machine, which are hard to clean.

Evisceration

This follows the 'clean' side of the slaughter line where, when working hygienically, the surface bacterial count of the carcasses is no longer greatly affected. Opening and evisceration of the slaughtered animals take place on the slaughter conveyor manually by means of knife, axe, or saw. First the aitch-bone is split at its midpoint. The next step is loosening the anus and the reproductive tract. For this step, mechanical rectum looseners are frequently used. After splitting the sternum and opening the belly cavity, the gut set and the stomach are taken out once the esophagus has been cut approximately 2 cm from the stomach (in pigs). In the USA and the Far East, the guts and the pluck (contents of the thoracic cavity) are taken out together.

To avoid the hygiene risk of an outflow of rumen content in cattle, the esophagus should not be cut but rather it should be separated from the trachea and shut with a special instrument (rodding). Removing the pluck is the last step of evisceration.

Hog opening can also be done by a robotic system that adapts itself to the physical shape of the hog. The slaughtering robot cuts loose the rectum, splits the aitch-bone, and separates the ham. The front unit cleaves the sternum and opens the abdominal wall of the hog.

Splitting

Splitting of carcasses (cattle, pigs, horses) can be executed manually by axe or saw or in automatic carcass-splitting systems. The automatic systems work within the continuous dressing conveyor. The carcasses are split by a circular saw. Backbone guides and holding jaws on each side stabilize the carcass and thus prevent miscuts. During the cutting operation the saw blade is cooled with cold water. After splitting, the saw blade as well as the holding jaws are sterilized with hot water (82 °C). For pig carcasses, automatic chopper machines are also in use.

Waste Product Handling

Waste products are those parts of a slaughter animal that are not suitable for direct utilization. This includes condemned meat and offals, obligatorily confiscated tissues, e.g., meat of the sticking cut, eyes, tonsils and gut, stomach and rumen contents, glands, hair, inedible fat and blood. As a result of the BSE epidemic in some European countries, in the European Union the so-called specific risk material (SRM) has to be extirpated. This includes the following tissues of cattle, sheep, and goats greater than 12 months old: skull, including brain and eyes, tonsils, and spinal cord; the spleen of sheep and goats and the ileum (part of the small intestine) of cattle greater than 12 months old. In comparison to live weight, waste material accounts for about 20% for sheep (live waste weight 45 kg), 14% for bulls (live weight 450 kg) and 10% for pigs (live weight 90 kg).

With modern technologies, nearly all the slaughter waste can be economically transferred to useful products (Table 2). (*See* **Offal**: Types of Offal.)

Precleaning of waste water is frequently done within the plants. With mechanical methods such as rotating filter tubes or paternoster filter grids, up to 60% of sedimentable particles can be separated. More complete cleaning can be achieved by a flotation process, whereby suspended substances (colloids) are coupled with microgas bubbles so that they come to the surface by flotation and can be removed. (*See* **Effluents from Food Processing**: On-Site Processing of Waste; Composition and Analysis.

Effect on Meat Quality

There is a close relationship between meat quality and the treatment of animals before slaughter. Rough handling of pigs and fighting between animals

Table 2 Waste product handling

Waste material	*Processing*	*Final product*
Condemned meat, condemned offal		Liquefied, sterilized material
Obligatorily confiscated tissue (eyes, reproductive tract, etc.)	Rendering	Meat meals, animal feeds, inedible fats (oils, soaps, fatty acids for cleaners, cosmetics, etc.)
Inedible blood		
Inedible fat		
Blood, not for human consumption	Drying	Blood meal, animal feeds
Stomach and gut contents, inedible blood	Fermentation (anaerobic)	Biogas (methane)
Rumen contents	Fermentation (aerobic)	Fertilizer
Horns, toe nails	Drying, chopping	Horn chippings (fertilizer)
Hair	Physical/chemical treatment	Brushes, upholstery, animal feeds
Endocrine glands	Extraction	Hormones (e.g., insulin), enzymes (e.g., heparin), bile acids (e.g., chenodeoxycholic acid)

prior to stunning both increase the incidence of PSE meat. When the glycogen level in a stressed pig is high, much lactic acid is produced after killing, while muscle temperature is still high (about 40 °C). The combination of high lactic acid level and high temperature leads to denaturation of muscle proteins and thus produces PSE meat. The next severe stress factor is stunning. Many stunning methods cause vigorous clonic spasms (with stimulation of glycolysis and accelerated pH drop) and high release of stress hormones such as epinephrine (adrenaline) and norepinephrine (noradrenaline). By activating enzyme systems, epinephrine causes increased breakdown of muscle glycogen to lactic acid. If sticking of the stunned animal is delayed, the glycolytic effect of stress hormones is more pronounced and the incidence of PSE meat is higher.

The treatment of the carcass can influence the postmortem biochemical reactions in the musculature. It has been found that pigs that had been dehided instead of scalded had musculature that was darker, with better water-binding properties and greater tenderness. The reason is that scalding and singeing of pig carcasses produce muscle temperatures that are about 1 °C above those that apply during dehiding. This results in a quicker pH drop postmortem. Also, the intensive mechanical treatment of the carcasses from hair scrapers and other mechanical machines may contribute to the production of PSE meat.

Long times of stress, long fasting times (transport, lairage), and mixing animals from different farms produce dark, firm, dry (DFD) meat. In these cases, at the time of slaughter most of the muscle glycogen reserves are already exhausted. The remaining glycogen is therefore at a much lower level, so that the amount of lactic acid, which is produced via glycolysis, is also low. Consequently, the ultimate pH value in DFD musculature is relatively high (> 6.2). Refer to individual meats.

See also: **Beef**; **Meat**: Eating Quality; Hygiene; **Pork**; **Sheep**: Meat

Further Reading

Blackmore DK and Delany MW (1988) *Slaughter of Stock*. Publication no. 118, Veterinary Continuing Education. Palmerston North, New Zealand: Massey University.

Eikelenboom G (1983) *Stunning of Animals for Slaughter*. The Hague: Martinus Nijhoff.

Prändl O, Fischer A, Schmidhofer T and Sinell H-J (1988) *Fleisch, Technologie und Hygiene der Gewinnung und Verarbeitung*. Stuttgart: Verlag Eugen Ulmer.

Romanus JR, Jones KW, Costello WJ, Carlson CW and Ziegler PT (1985) *The Meat we Eat*, 12th edn. Danville, Illinois: Interstate.

Troeger K (1999) *Slaughter Method and Animal Welfare*. Proceedings of the 45th International Congress of Meat Science and Technology, vol. I, 40–48, August 1–6, 1999. Yokohama, Japan.

Preservation

D A Ledward, University of Reading, Whiteknights, Reading, UK

Introduction

The principles of meat preservation are mainly associated with preventing or delaying microbial spoilage and chemical action, and avoiding as far as possible weight loss and any change in taste or texture. Methods of preservation are based on control by low or high temperature, reducing available water (water activity), or by the application of chemical additives. In developed countries the shelf-life of meat and meat products is, in the main, controlled by refrigeration or by a combination of refrigeration and

other methods of preservation, such as modified-atmosphere packaging, curing or high pressure. Whereas other techniques result in some changes to color, taste, and texture, refrigeration and high pressure are distinct in that their effect on appearance and taste is minimal. (*See* **Legislation**: Additives; **Preservation of Food**; **Spoilage**: Chemical and Enzymatic Spoilage; **Spoilage**: Bacterial Spoilage; Molds in Spoilage; Yeasts in Spoilage.)

Refrigeration

Product Cooling

Refrigerated meats are generally classified as chilled or frozen. However, there may be several steps involved in reducing a hot product to chill or frozen storage temperature, with several additional links before final consumption. The design and operation of equipment to perform these functions require an understanding of the thermal properties of foods and an appreciation of their complexity and that of prevailing legislation.

Immediately after heat treatment or cooking, it will be necessary to commence cooling under controlled conditions. Filtered ambient air or mains water are suitable media for cooling the product temperature to about 35°C. Naturally, the latter would only be used where there is a hermetically sealed skin packaging. These represent cheaper energy sources than mechanical refrigeration or expendable refrigerants, which are required for achieving the statutory temperatures, which lie below 10°C.

Chilling

The rate of evaporation of moisture from an exposed surface is a function of the temperature difference between the surface and the air and is influenced, to a lesser extent, by its relative humidity. The air velocity over the surface has a marked effect, at least in the initial stages of cooling. A similar set of conditions prevails if the cooling medium is a gas other than air, e.g., carbon dioxide or nitrogen.

By lowering the applied temperature and increasing the velocity, the initially high rate of evaporation is offset by a reduced surface cooling time. This results, overall, in reduced weight loss. Hence different modes of refrigerated chilling may be employed:

1. Rapid chilling in which chilling is applied at the maximum rate possible (without freezing the surface of the product) with the objective of minimizing evaporative loss. Where air is the cooling medium, the applied temperature would be below 0 °C at a velocity of 2.0–3.5 m s^{-1}. However, rapid chilling introduces the risk of toughening due to 'cold shortening' of carcasses chilled to < 10 °C before the pH has decreased to below 6.0. In beef carcasses this would normally be in less than 10 h.
2. Conventional chilling at a rate sufficient to meet production requirements without undue regard to evaporative weight loss. In this case the applied temperature and velocity would be greater than 0 °C and in the range 0.25–0.50 m s^{-1}, respectively. Even at this rate of chilling there is a risk of 'cold shortening' causing irreversible toughening in areas close to the surface.

Freezing

In exceptional circumstances, the product may be frozen from an initial temperature of about 35 °C ('hot' freezing), but is more likely to have already been chilled to 4 °C for pork and lamb or to 7 °C for beef carcasses. Since lean meat begins to freeze at a temperature of −1 °C, there is initially a sensible heat load until this temperature is reached, followed by the removal of latent heat. Conventions for freezing time are not clearly defined as freezing may not be entirely complete, even at temperatures below −20 °C, owing to the depression of freezing point by remaining concentrated solutes. One convention quotes the 'nominal freezing time' as the time required to achieve a temperature 10 °C below that at which freezing commences. This is distinct from the 'effective freezing time', being the elapsed time interval from product initial temperature to low-temperature storage. (*See* **Freezing**: Principles.)

Tempering

The heat capacity of foods containing water is virtually halved once the temperature is reduced below −2.0 °C. At the same time, the thermal conductivity is doubled. Hence, meat which has been taken through most of the latent heat phase will respond much more rapidly to changes in external temperature, e.g., if storage is to be at −25 °C a final core temperature of −15 °C may be satisfactory for the freezing stage, allowing the product to equalize and stabilize at the holding temperature in the cold store.

Cold Storage

The low-temperature store will be designed to accommodate the maximum quantity of a product within a given volume. The store will maintain a temperature of about 1 °C if the product is to be kept chilled or less than −20 °C if frozen. In the former case it will be more critical that the product has been chilled to the design storage temperature, whereas for frozen products a small proportion of residual heat, as already

outlined, may be removed during the holding period. Packaging assists in preventing weight loss, freezer burn, and the onset of rancidity. (*See* **Freezing**: Storage of Frozen Foods.)

Refrigerated Transport

In theory there should be few problems in the storage or transport of chilled or frozen products, especially in the case of meat or meat-based products, since:

1. Cooling, chilling and, where applicable, freezing and tempering have been completed so that refrigerated transport equipment has merely to insulate the product from sources of ambient heat
2. The product is insulated by primary and secondary packaging
3. There is no respiratory heat load such as occurs with some fruits and vegetables.

Retail Display

With the introduction of the European Community (EC) Food Hygiene Regulations, the temperature of chilled foods in the EC has to be kept at or below 5 °C. This requires much more stringent control than bulk storage since the displayed products have to be accessible and hence are potentially exposed to a warm atmosphere. A high intensity of lighting can lead to a further rise in temperature due to a 'greenhouse' effect inside transparent packaging.

Similar conditions apply to frozen foods, but these are more likely to be cartoned in opaque material with entrapped air, which acts as an insulator. It is also a fact that it is much more difficult to check and hence legislate on the temperature of frozen foods.

Domestic Storage

As with retail display cabinets, domestic refrigerators have minimal reserve capacity for chilling, freezing, or tempering. Therefore the addition of substantial amounts of a product which is above the set storage temperature will affect that of the existing contents. An awareness of the necessity to maintain the cold chain from the retailer to domestic refrigerator is paramount. While it has been demonstrated that meat which has been repeatedly frozen and thawed under controlled conditions suffers minimal deterioration, such a practice cannot be recommended wherever precise control cannot be guaranteed.

Refrigeration Plant Selection

The first priority in plant selection is to quantify the refrigeration duty. To do so requires a knowledge of the thermal properties of the product under consideration, such as: (1) initial and final temperatures; (2) heat capacity above and below freezing; (3) latent heat of freezing; and (4) effective chilling or freezing time between specified initial and final temperatures. (*See* **Freezing**: Operations.)

Recognition has to be taken of the fact that the maximum rate of product heat output (which occurs in the early stage of chilling or freezing) may be two or three times the average. To the load is added the heat due to: (1) leakage through insulation into the refrigerated space; (2) leakage of ambient air through door openings; (3) electrical input such as lights, fans, or fork-lift trucks; (4) metabolic heat from personnel.

Having estimated the total refrigeration load, a decision is taken on the type of refrigeration equipment to be installed. In the majority of cases this will be mechanical refrigeration using ammonia or a fluorocarbon as the refrigerant. Where production is seasonal, it may be more economical to use an expendable refrigerant such as liquid nitrogen or carbon dioxide.

Moisture Control

Freezing *per se* is an effective means of preservation as it makes moisture unavailable for microbial growth by its conversion to ice. However, other means of moisture control have been devised that enable products to be stored at temperatures above the freezing point of water. These include curing, several methods of dehydration, including freeze-drying, intermediate-moisture technology, and high-pressure processing.

Curing

Salting as a means of meat preservation has been practiced for 3000 years and contamination of salt by potassium nitrate probably led to the development of modern cured meats, since it led to characteristic and acceptable effects on color and flavor. Thus, although heavily salted products are not now available, having been supplanted by more acceptable preservation procedures, cured pork and to some extent poultry, beef, and fish products have established niches in the marketplace, though the relatively low levels of salt now present (usually < 4%) are more for flavor than preservation.

To cure meat, salt (sodium chloride) and nitrite must be present and other ingredients may be used during processing to accelerate curing, stabilize color (e.g., ascorbic acid), modify flavor (e.g., sugars) and texture, and reduce shrinkage (e.g., polyphosphates).

Several methods may be used to incorporate the curing ingredients into the meat, the object being to achieve uniform distribution. In comminuted

products, such as sausages, the cure ingredients (in the dry form or as a concentrated solution) are incorporated into the product during mixing or comminution.

For whole sides or muscles the oldest technique is dry-curing. Here, the dry-curing agents (salts) were rubbed over the surface of the meat, the meat then being placed on a shelf or in a box to cure. For large muscles or cuts the process needed to be repeated several times and it may have taken up to 80 days to get uniform distribution in, for example, the top of the hind leg.

Another method which is still practiced for a few specialty products involves immersing the cuts or sides in concentrated brine, usually at 3–7 °C, and allowing the brine slowly to diffuse through the meat. This method is also very time-consuming and care must be taken to insure microbial spoilage does not occur during curing.

The curing procedure can be considerably shortened by injection of brine into the meat to achieve rapid, uniform distribution. Three such methods are: (1) artery pumping, in which brine is pumped directly into the intact vascular system; (2) stitch pumping, in which brine is injected (through a hollow needle) into various parts of the meat, especially the thickest parts and those near to joints; and (3) multiple injection, in which brine is injected simultaneously and automatically through a series of hollow needles which have a number of evenly spaced holes along their length. Another modern method is the 'tumbling' or 'massaging' of meat pieces in rotating drums. In the presence of about 0–6% of their weight of salt the physical interactions of the meat pieces draw out salt-soluble protein which, on subsequent heating, gels to bind the pieces together in 'reformed' hams. Tumbling can shorten the curing period to 24 h or less. Tumbling may be carried out under vacuum to eliminate problems caused by dissolved air.

A typical traditional brine contains 25–30% sodium chloride and 3–4% sodium or potassium nitrate and specific salt-tolerant bacteria capable of reducing the nitrate to nitrite. In modern cures, nitrite may be added directly at a concentration of 200 p.p.m. or less. In most countries there are strict limits on the level of nitrite permitted in meats (usually 200 p.p.m.) and addition of nitrite directly, rather than relying on nitrate reduction *in situ*, insures such limits are not breached.

Dehydration

That dehydration can preserve meat has been known for over 5000 years and simple drying procedures are still used today in some parts of the world. In these processes, strips of lean meat are exposed to the sun, e.g. pemmican manufactured by North American Indians, or a combination of light salting and drying is used as in the preparation of charqui (South America) or biltong (South Africa). (*See* **Drying**: Theory of Air-drying; **Drying**: Drying Using Natural Radiation.)

Hot-air drying has been used in the past to prepare dehydrated meat for the armed forces, but has not been manufactured successfully for domestic consumption. Due to case hardening, raw meat cannot be successfully dehydrated and the need for a high surface-area-to-volume ratio has limited its use to cooked, minced meat. A typical procedure involves cooking the meat in thin slices, mincing (0.3–0.8 cm in diameter) and drying under controlled conditions. The temperatures that can be tolerated without causing unacceptable damage to the meat depend on both the residual water and fat contents. A typical regime that lowers the moisture content to about 5% in about 5 h involves an air speed of 3 m s^{-1}, a temperature of 60 °C, a relative humidity of 40%, and a tray loading of about 10 kg m^{-2}. Vacuum drying has been unsuccessfully tried to improve the quality of dehydrated meat. A successful technique for both raw and cooked meat is freeze dehydration. In this process the product remains frozen throughout the drying cycle and thus the water (ice) is sublimed from the meat. At a vacuum pressure of 1.0–1.5 mmHg, temperatures up to 43 °C can be tolerated with no thawing. The rate of freeze-drying is limited by the rate of heat transfer to the product for sublimation and several methods have been developed to increase the rate, including the use of microwave or radiant (infrared) energy. Freeze-drying is an expensive processing procedure and freeze-dried meats (containing about 2% moisture) are usually only used in such foods as dehydrated soup mixes and specialty products for campers, the military, and space agencies. (*See* **Freeze-drying**: The Basic Process.)

Dehydrated meats need to be rehydrated prior to consumption and, to address this step, with all its attendant problems, and evolve a technology suitable for developing countries, intermediate-moisture meats have been formulated. With intermediate-moisture foods the object is to lower the water activity to a level at which bacteria will not grow, but without lowering the water content to a point where the product becomes unpalatable. This involves mixing the meat with suitable humectants, such as sugars, salts, polyols, and amino acids, to lower the water activity to about 0.85 whilst keeping the moisture content at 20–50%. Under these conditions bacteria will be inhibited, but molds and yeasts may proliferate and thus an antimycotic such as sorbate or propionate is often incorporated at a concentration

of 0.1–0.3%. (*See* **Intermediate-moisture Foods; Water Activity.**)

For minced products, such as semimoist pet foods and Chinese pork, the equilibration is easily achieved by batch-mixing. In the production of semimoist pet food, the mixture is subsequently shaped and formed in a cooker-extruder. For whole meats two techniques are available: (1) desorption processing, in which the meat pieces are soaked in an infusing solution of high osmotic pressure so that, after equilibration, the meat has the desired water activity (equilibration is normally aided by heating, which has the additional benefit of inactivating enzymes); and (2) absorption processing, in which freeze-dried meat is equilibrated with the infusing solution to the appropriate water activity. Though this latter method allows close control of water activity, the cost is prohibitive for all but the most specialized uses (e.g., space flights).

Other Preservation Techniques

Although temperature reduction and moisture control are the most commercially important methods of meat preservation, other techniques may have important niches in the marketplace. For example, fermented meat products, in which specific bacteria, such as selected *Lactobacillus* strains, are encouraged to grow and reduce the pH of the meat to levels inhibitory to the growth of spoilage organisms, are well established in many parts of the world. This technology is often combined with partial dehydration to enhance shelf-life. Though not commonly made in the UK, European varieties include French saucisson sec, German Dauerwurst and Rohwurst and Italian crudi stagionati. In modern processes a starter culture is used to initiate two fermentations – the conversion of nitrate to nitrite (by micrococci) and of carbohydrates to lactic acid by lactobacilli. Fermentation ceases after 3–7 days under high (e.g., 26 °C) temperatures and humidities (e.g., 95%). The product is then dried at lower temperatures and humidities to moisture contents in the range 30–35%. This takes several weeks (usually 3–7) and the product is then matured for several months. The combination of low pH (about 4.5), lactic acid, low water activity, and residual nitrite gives a shelf-life from 6 months to over 2 years, provided the whole process has been controlled and managed correctly. (*See* **Lactic Acid Bacteria.**)

The use of chemical preservatives in meat is not widely practiced (except for cured/fermented products), but sulfur dioxide is permitted in some countries for certain products. Thus 'English'-style sausages may contain up to 450 p.p.m. of this compound, usually added in the form of a salt such as metabisulfite.

High-temperature preservation (canning) of meat and meat products is also well established and, although certain cured products may only require a pasteurization regime, others are heated under the more severe regimes necessary for low-acid foods. (*See* **Canning:** Principles.)

In the 1940s, ionizing radiation was advocated as a method of meat preservation and after very extensive trials it was permitted in most countries under certain conditions. Thus, in 1991 in the EC pasteurizing doses of γ-rays up to 5 kGy were permitted for poultry. (*See* **Irradiation of Foods**: Basic Principles; Legal and Consumer Aspects.) More recently, high pressures (≥ 400 MPa) have been used to extend the shelf-life of cured meats. However, this technology has less potential for uncured products, although the technology is developing rapidly.

Chemical, Physical, and Microbiological Implications of Preservation

Chilling and Freezing

The physical dimensions of postrigor meat are unaffected by both chilling and freezing but fluid (drip) is exuded from both chilled and frozen products. The amount expressed is far greater in frozen-thawed meat and some evidence suggests the amount increases as the freezing rate decreases.

In frozen meat, bacterial growth usually ceases, but yeasts and molds may continue to grow down to about $-5\,^{\circ}C$. In chilled meat, psychrotrophic bacteria become the dominant species, mainly *Pseudomonas, Achromobacter, Micrococcus, Lactobacillus, Streptococcus, Leuconostoc, Pediococcus, Flavobacterium*, and *Proteus* spp. Of these, during storage in air, *Pseudomonas* spp. are the most important since they give rise to slime and off-odor formation at levels of 10^7–$10^8\,cm^{-2}$. The time for this to occur depends on several factors (especially temperature, relative humidity, and initial load) but meat kept in retail packs at a relative humidity of 99.3% usually spoils after 5–10 days at 1 °C. If meat is stored under elevated levels of carbon dioxide the growth of *Pseudomonas* spp. is inhibited and *Lactobacillus* spp. become dominant. Such meat can be kept for several weeks at chill temperatures with little evidence of spoilage.

Chemical changes take place during storage of fresh meat and some, such as those due to enzymatic proteolysis, may be favorable, leading as they do to increased tenderness. The major deteriorative changes are oxidative in nature, leading to the formation of the undesirable brown methemoproteins and rancidity. The formation of methemoproteins is

important in both chilled and frozen meat but lipid oxidation leading to rancidity is usually only important in frozen meats. This oxidation is responsible for recommending maximum shelf-lives of 4 months for pork and poultry and 6 months for beef and lamb at −18 °C. (*See* **Oxidation of Food Components**.)

Cured Products

Nitrite is essential for the development of the pink-red color of cured meats and has the further important attribute of being a very effective inhibitor of *Clostridium botulinum*. The pink-red color is due to nitric oxide, from nitrite, binding to the iron atom of hemoprotein (myoglobin and hemoglobin) at the sixth coordination site; there is still controversy regarding the mechanism involved in the formation of this nitrosyl pigment. In the presence of light and oxygen (air), nitric oxide can be liberated from the hematin and the color fades as the oxidized methemoprotein forms. Thus, cured meats should be stored in the absence of oxygen and/or light if an extended shelf-life is required. Although cured meats are eaten uncooked (e.g., Parma Iberian ham), many are cooked prior to consumption, either by the consumer, e.g., bacon, or by the manufacturer to be sold as cooked, ready-to-eat meats (e.g., boiled or roast ham). The red pigment formed on heating is relatively stable in the presence of light and oxygen and is believed to be a complex of free hematin and either one (pentacoordinated) or, less likely, two (hexacoordinated) molecules of nitric oxide.

Dehydrated Meats

Hot-air-dried meats are invariably shriveled and are virtually impossible to rehydrate to typical 'full-moisture' products. Freeze-dried meats retain the shape and color of the original meat pieces and the porous structure does rehydrate more satisfactorily than hot-air-dried products. Intermediate-moisture meats have the physical characteristics of their fresh-meat counterparts.

Unless partial rehydration inadvertently occurs, dehydrated meats are not susceptible to microbial spoilage and can be kept for several weeks at elevated temperatures. The factors limiting their shelf-life are chemical in nature, mainly lipid (or sorbate) oxidation, and/or nonenzymatic browning. With freeze-dried meats, nonenzymatic browning involving glucose and glucose-6-phosphate (with protein) occurs, as also, in the presence of oxygen, does oxidation of the hemoproteins to yellow bile pigments. Some fat oxidation due to residual lipolytic activity may also occur. In intermediate-moisture meats, browning involving reactions between sorbate or lipid oxidation products and protein occurs, as also does oxidation of the hemoproteins. However, a reaction occurring in intermediate-moisture meats, but not fully dried meats, is hydrolytic degradation of collagen. Anaerobic storage of intermediate-moisture meats inhibits all these reactions, except for the breakdown (hydrolysis) of collagen. (*See* **Browning: Nonenzymatic**.)

See also: **Drying**: Theory of Air-drying; Drying Using Natural Radiation; **Freeze-drying**: The Basic Process; **Freezing**: Principles; Operations; Storage of Frozen Foods; **Irradiation of Foods**: Basic Principles; Legal and Consumer Aspects; **Lactic Acid Bacteria**; **Legislation**: Additives; **Oxidation of Food Components**; **Preservation of Food**; **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage; Molds in Spoilage; Yeasts in Spoilage

Further Reading

Diestre A and Monfort JM (1998) *Meat Science*, 49, supplementary issue.

Lawrie RA (1998) *Lawrie's Meat Science*, 6th edn. Cambridge: Woodhead.

Ledward DA (1981) Intermediate moisture meats. In: Lawrie RA (ed.) *Development in Meat Science*, vol. 2, pp. 159–194. London: Applied Science.

Ledward DA (1998) High pressure processing of meat and fish. In: Autio K (ed.) *Fresh Novel Foods by High Pressure*, pp. 165–175. Helsinki, Finland: VTT Biotechnology & Food Research.

Ranken MD (2000) *Handbook of Meat Product Technology*. Oxford: Blackwell Science.

Wilson NRP (1981) *Meat and Meat Products – Factors Affecting Quality Control*, pp. 138–163. London: Applied Science.

Eating Quality

I Lebert, S Rousset, A Lebert and R Talon, National Institute of Agronomic Research, France

Background

Meat eating quality refers to microbiological safety, shelf-life, sensory properties, nutritive values, and healthiness. Safety is the priority, but sensory qualities must also satisfy the consumer. The acceptability of the final product is often influenced by the processing of the raw material. The hygienic and sensory

Table 1 Beef chilling: effect of increasing time, temperature, air velocity, and humidity on quality

	Time	*Temperature*	*Air velocity*	*Humidity*
Microbiological quality	−	−	+	−
Tenderness	+	+	0	0
Appearance	−	−	−	+
Weight loss	+	+	+	−

+, increase; −, decrease; 0, no effect.

qualities (texture, juiciness, flavor, and color) are particularly important. The control of the quality must therefore be implemented from the entry of the live animal into the meat chain. Initial carcass chilling is particularly important, and its effects on beef quality are listed in Table 1.

Consumers are becoming more sensitive to issues of safety, hygienic, and sensory qualities, and these issues are discussed in this article.

Microbiology of Meat

The microbiological quality of meat carcasses is highly dependent on the processing conditions: most important are the temperature and time of storage and distribution, which largely determine the number and spread of contamination during slaughter and processing (Figure 1).

During slaughter and dressing, the three major sources of contamination are: (1) the skin of the animal, which is mainly contaminated by spoilage flora such as staphylococci, micrococci, pseudomonads, yeasts, and molds; (2) fecal contamination during evisceration, which, if little care is taken, could happen by bursting of the intestinal tract; then *Salmonella, Escherichia coli*, enterococci, and *Clostridium perfringens* can be detected in meat; (3) contamination by direct contact or cross-contamination between dirty and clean surfaces. Staphylococci are the main concern, since they are widely distributed, and contamination may arise from both animal and human sources. Airborne particles and aerosols also contribute to the dissemination of microorganisms.

Under good hygienic conditions, dressed carcasses may have typical surface aerobic bacterial counts of 10^2–10^4 cm^{-2} for beef, 10^2–10^5 bacteria cm^{-2} for sheep, and 10^4–10^5 bacteria cm^{-2} for pig carcasses.

Once the temperature at the surface of the carcass has fallen below 10 °C, growth of mesophilic bacteria no longer occurs, and a psychrotrophic flora (*Pseudomonas, Acinetobacter, Psychrobacter, Lactobacillus, Brochothrix thermosphacta, Shewanella putrefaciens*, some genera in the family Enterobacteriaceae) gradually displaces the micrococci and coliforms, which are commonly found on freshly dressed carcasses.

Two bacterial flora can be found on meat products: a pathogenic flora (*Salmonella* spp., *Clostridium* spp., *Yersinia enterocolitica, Staphylococcus aureus, Listeria monocytogenes*) that causes foodborne diseases and a spoilage flora (*Pseudomonas* spp., lactobacilli, staphylococci, micrococci, *B. thermosphacta*) that limits the storage of the product.

Pathogenic Flora

Epidemiology Public health authorities need to be informed about the nature and magnitude of foodborne illnesses and their epidemiology for the early detection of outbreaks, planning, evaluation of any disease, and policy implementation. Foodborne disease cases are significantly underestimated because of the weakness in foodborne disease surveillance and of the variation in reporting systems between countries. As a consequence, it has been estimated that the reported incidence of foodborne diseases represents less than 10%, perhaps even less than 1%, of the actual incidence.

Foodborne outbreaks occur in most, if not all, countries (Table 2). *Salmonella* spp. cause 84.5% of all outbreaks (*Salmonella enteritidis*, 50.9%); *S. aureus*, 3.5%; *Clostridium botulinum*, 1.1%; *Bacillus cereus*, 1.0%; and unknown agents, 9.5%. Sporadic cases have been associated with *Escherichia coli, Campylobacter jejuni, L. monocytogenes, Y. enterocolitica, Aeromonas hydrophila, Shigella, Vibrio parahaemolyticus*. The main factors affecting the growth of these pathogenic bacteria are the temperature, pH, and water activity (Table 3).

The most frequently identified foods associated with outbreaks are eggs and egg products (25.4%), meat and meat products (23.4%), and confectionery, cakes, pastries, puddings, desserts, and icecream (17%). Fish and shellfish, milk and dairy products, and poultry and poultry products were each implicated in less than 5% of the outbreaks.

Main pathogenic bacteria As mentioned above, *Salmonella* spp. cause 84.5% of all outbreaks, and many foods of animal origin, particularly beef and pork, have often been identified as vehicles in outbreaks of salmonellosis. *Salmonella* are transferred to meat surfaces from the intestinal contents and from feces contaminating hair, skin, and feet of animals. Considerable contamination of pig carcasses occurs during scalding and dehairing. Later, bacteria can be transferred from carcass to carcass and from contaminated carcasses to knives, work surfaces, and the hands of meat handlers. Minced pork and sausages, as well as

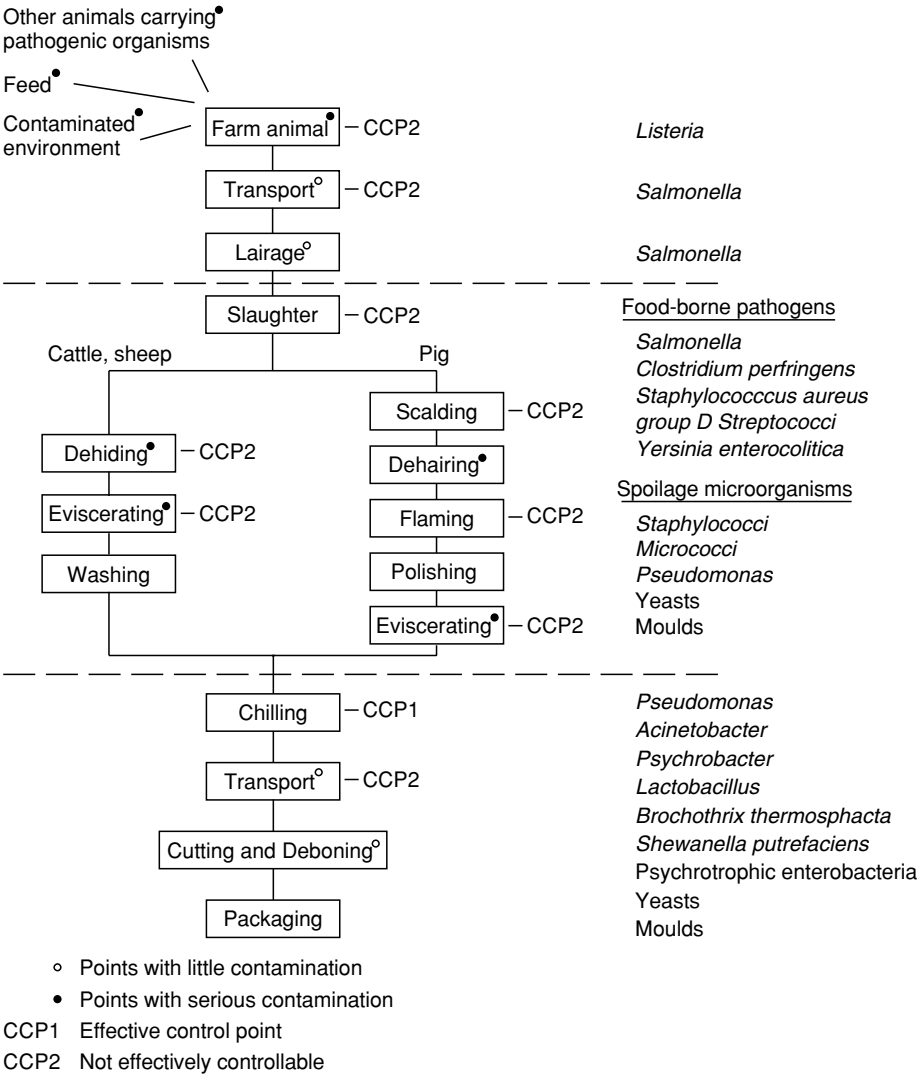

Figure 1 Causes of contamination from live animal to store.

Table 2 Cases of foodborne diseases in 1997 and 1998

Date	*Locations*	*Number of cases*	*Bacteria*	*Associated foods*
1997	Australia	650 (three deaths)	*Salmonella*	Different food in many foodborne outbreaks
	Russia	12 (one death)	*C. botulinum*	Fish
	Finland	200	*Salmonella*	Pig
	USA	745 (one death)	*Salmonella*	Stuffed ham
1998	Argentina	9	*C. botulinum*	Matambre (meat specialty from Argentina)
	Japan	1142	*Salmonella*	Unknown origin
	USA	316	*V. parahaemolyticus*	Oysters
	Algeria	1400 (17 deaths)	*C. botulinum*	Spoiled chicken
	USA	100 (20 deaths)	*L. monocytogenes*	Hot dogs
	USA	165	*Shigella*	Interhuman?
	Canada	130	*Salmonella*	Turkey meat
	France	300	*Salmonella*	Egg/egg product 52%, meat/meat product/chicken 24%, other foods 24%
	France	68	*C. perfringens*	meat/meat product/chicken 40%, other foods 60%
	France	155	*S. aureus*	Meat/meat product/chicken 15%, milk/milk product 24%, other foods 61%
	France	139	*Unidentified*	Meat/meat product/chicken 14%, other foods 86%

Table 3 Minimal and maximal temperature (T), pH and water activity (a_w) values for the growth of some pathogenic bacteria (ICMSF, 1996)

Organisms	*T (°C)*	*pH*	a_w
Bacillus cereus	4–55	5–8.8	> 0.93
Vibrio parahaemolyticus	5–43	4.8–11	> 0.94
Escherichia coli	7–45	4.4–9.0	> 0.95
Campylobacter jejuni	32–45	4.9–9	> 0.987
Yersinia enterocolitica	1–42	4.2–9;6	> 0.97
Listeria monocytogenes	0–45	4.4–9.4	> 0.92
Aeromonas hydrophila	0–45	4.5–NR	> 0.97
Salmonella spp.	5.2–46.2	3.8–9.5	> 0.94
Shigella spp.	7.9–45	5–9.2	> 0.978
Staphylococcus aureus	7–48	4–10	> 0.83
Clostridium perfringens	12–50	5.5–9	> 0.93
Clostridium botulinum			
A & B	10–50	4.7–9	> 0.93
Non proteolytic, C & D	5–NR	4.7–9	> 0.95
E	3.3–15–30	4.7–9	> 0.97
F	4–NR	4.7–9	> 0.95

NR: not reported.

inadequately cooked meat, are frequently implicated in salmonellosis. *Salmonella* are widespread in poultry, especially turkey and chicken carcasses, and are transferred to the meat from the intestinal tract or from fecal material on feet and feathers. Cross-contamination is a particular problem in which the critical steps include defeathering, evisceration, and chilling. The major recent change in the epidemiology of *Salmonella* concerns the emergence of *S. enteritidis* and *S. typhimurium*. These bacteria, first studied in the USA and UK, are now reported in many industrialized countries. The factors that may contribute to their emergence are the microbial adaptation and dissemination of the pathogen through increasing global distribution of food and international travel, the increasing susceptible populations as well as lifestyle changes leading to an increase in food-service establishments and food outlets combined with inadequate knowledge of food-handling.

S. aureus is ubiquitous and occurs on the skin of most warm-blooded animals, including all food animals. The organism is resistant to drying and may colonize food-processing equipment that may be wet and difficult to clean. Meat and meat products are frequently incriminated in staphylococcal food poisoning, but cases have arisen in egg products, sandwich fillings, milk, and dairy products. Foods that require considerable handling during preparation and that are kept at slightly elevated temperatures after preparation (20–40 °C) are frequently involved in staphylococcal food poisoning. Human intoxication is caused by ingesting the enterotoxins produced in the food by some strains of *S. aureus*, usually because the food has not been kept hot enough (60 °C or above) or cold enough (7.2 °C or below). The toxin is a highly heat-stable protein. The toxin is produced when *S. aureus* populations exceed 10^5 per gram. A toxin dose of about 1.0 μm produces symptoms of intoxication.

The incidence of *C. botulinum* in meats appears to be low, but its presence has been demonstrated in sufficient large surveys of pork. Botulism in man results most often from eating improperly stored foods in which *C. botulinum* has multiplied and produced toxin.

C. perfringens is part of the normal flora of the intestinal tract of animals, including man, and is found in a large variety of foods. In raw meat, a few vegetative cells may be found deep in muscle tissue after slaughter, and more cells are found in meat from fatigued animals. The presence of spores on the surface of recently slaughtered carcasses is the result of fecal contamination. Care is needed with foods made of several ingredients such as meat pies, cooked meats, and poultry products, because spores may survive cooking.

Other pathogenic bacteria causing sporadic cases

Undercooked or raw ground beef has been implicated in nearly all documented outbreaks of *E. coli* O157:H7 and in other sporadic cases. *E. coli* is a normal inhabitant of the intestine of all animals, including humans. Currently, there are four recognized classes of enterovirulent *E. coli* that cause gastroenteritis in humans. The enterohemorrhagic strain, designated *E. coli* serotype O157:H7, is a rare variety of *E. coli* that produces large quantities of one or two toxins that cause severe damage to the lining of the intestine. These toxins are closely related to the toxin produced by *Shigella dysenteriae*.

C. jejuni frequently contaminates raw chicken. Surveys show that 20–100% of retail chickens are contaminated. This is not entirely surprising, since many healthy chickens have these bacteria in their intestinal tracts. Raw milk is also a source of infections. The bacteria are often carried by healthy cattle and by flies on farms. However, properly cooking chicken or pasteurizing milk kills the bacteria. Campylobacters can be isolated from freshly slaughtered red-meat carcasses, but in smaller numbers than on poultry. This bacterium is recognized as an important enteric pathogen. Recent surveys have shown that *C. jejuni* is the leading cause of bacterial diarrhea in the USA, causing more illness than *Shigella* spp. and *Salmonella* spp. combined.

L. monocytogenes has been associated with foods such as raw milk, cheeses (particularly soft-ripened varieties), raw vegetables, but also fermented raw-meat sausages, raw and cooked poultry, all types of

raw meats, and raw and smoked fish. Its ability to grow at temperatures as low as 3 °C permits multiplication in refrigerated foods. The contamination of meat and meat products can be due to fecal contamination during slaughter, presence on clean and unclean sections in slaughterhouses, and contaminated ground and processed meats: 10–80% of contaminated samples contain less than 10–100 CFU g^{-1}. *L. monocytogenes* is a ubiquitous bacteria found in soil, silage, and other environmental sources, and is present in the intestines of 1–10% of humans. *L. monocytogenes* is quite hardy and resists the deleterious effects of freezing, drying, and heat.

Y. enterocolitica has been recovered from a wide variety of animals, foods, and water. Pigs seem to be the principal reservoir of bioserotypes pathogenic to humans, but the exact cause of the food contamination is unknown.

Aeromonas spp. are ubiquitous and are also associated with foods of animal origin (raw meats, poultry, and milk). *A. hydrophila* grows rapidly in a refrigerated environment and can increase its number 10–1000-fold in meat and fish samples over 1 week of refrigerated storage.

Among several environments (Table 4), the home is where the pathogens are frequently identified (13%), with 46% of the outbreaks occurring in people eating at home largely due to mishandling of food products (Table 4). The consumer must take care when handling food at home, and recommendations have been given by The National Advisory Committee on Microbiological Criteria for Foods to prevent the contamination of food products by foodborne pathogens (Table 5).

Spoilage Flora

Without exception, there is a loss in microbial quality in meat products following slaughter and manufacture (Table 6). Microorganisms cause deterioration in appearance, texture, flavor, odor, or safety of the product. The first detectable sign of spoilage is usually an off-odor, which is slightly sour or putrid. The odors are caused by bacterial metabolic end products, as they accumulate, the color of meat may change to a gray or green color, while the texture changes due to the destruction of proteins and lipids. Spoilage may occur slowly in the refrigerator and more rapidly at higher temperatures.

On meat held under aerobic conditions, spoilage is often related to the growth of *Pseudomonas* and occurs when the *Pseudomonas* counts exceed about 10^6 CFU cm^{-2}. Because of their psychotrophic characteristics, *Pseudomonas* species have a marked advantage in growth rate compared with other genera, and this advantage tends to increase with decreasing temperature. *Pseudomonas* is less tolerant to reduced water activity than lactobacilli and *B. thermosphacta*, so these organisms may become dominant in the drier parts of the carcass. *Acinetobacter* may become dominant in high pH meat.

Vacuum packaging is used to prolong the storage life of fresh meat (for 21–28 days at 0–1 °C) and

Table 4 Results of foodborne disease surveillance

Place of contamination or mishandling	*Identified outbreaks when people eat food products*	*Factors contributing to outbreaks*
Entering the food chain at the farm (50%)	Homes (46%)	Temperature abuse, inadequate cooling, and improper cooking (44%)
	Restaurants/hotels (15%)	
Mishandling	Catered events (8%)	Contaminated or toxic raw products (16%)
Restaurants (22%)	Medical-care facilities (6%)	Contamination by personnel or equipment (15%)
Homes (13%)	Canteens (6%)	Lack of hygiene in processing, preparing, and handling (10%)
Catering establishments (7%)	Schools (5%)	Cross-contamination (4%)

Table 5 Recommendations of safe food preparation

Wash hands and utensils before handling food, especially after handling raw foods
Reheat all foods thoroughly (above an internal temperature of 74 °C)
Keep hot food hot (above 63 °C)
Keep cold foods cold (below 4 °C)
Thoroughly cook meat, poultry, and seafood, and adequately heat frozen or refrigerated foods
Chill foods rapidly in shallow containers
Keep raw and cooked foods separate, especially when shopping, preparing, cooking, and storing these products
Wrap and cover foods in the refrigerator
Keep the refrigerator temperature between 1 and 4 °C

Table 6 Type of spoilage caused by spoilage bacteria in meat

Frequency	*Bacteria*	*Spoilage*
High	*Pseudomonas*	Green color, slimy
	Acinetobacter	Putrid
	Enterobacteria	Gas, green color (*Proteus*)
	Flavobacterium	Putrid
	Brochothrix	Green color, cheesy odor
	Lactobacilli	Green color, sour
Medium	*Bacillus*	Putrid odor, brown color
	Alcaligenes	Putrid
	Streptococcus	Sour
	Aeromonas	Putrid, gas
	Clostridium	Putrid, gas, green color
	Shewanella	Putrid, green color
Rare	*Pediococcus*	Green color, sour
	Kurthia	Putrid

restricts the growth of *Pseudomonas* so that *Lactobacillus* or *B. thermosphacta* becomes the dominant flora. In most cases, *Lactobacillus* can grow more rapidly than *B. thermosphacta*, but *B. thermosphacta* dominates in fatty meat with a high pH.

The deterioration of the food occurs progressively during the storage and results from the activity of the microorganism in the food. To avoid these changes, preservative techniques can be applied to inactivate the microorganism: by heat or irradiation; and to inhibit growth: by chilling or freezing, drying, curing with salt, vacuum or gas packaging, acidifying, or by adding preservatives. Consequently, in dried cured meat product, the pathogenic and spoilage flora are inhibited by drying and salting, whereas in dried sausage, the main inhibiting factors are drying and acidification. Eventually, the cumulative effects of the changes reach a point of consumer rejection due to unacceptable quality expectations and perception.

Acceptability and Sensory Properties of Meat

In order to determine the characteristics of food products and their acceptability, food manufacturers need two kinds of information: how much people like their products overall and the sensory characteristics of the products. Hedonic liking tests are carried out by consumer panels and usually ask a question of overall acceptability and ask for comments about the characteristics which consumers like or dislike. At least 60 consumers are generally included in such surveys because of the large range of their hedonic responses. Retailed sensory characteristics are not asked of consumers because they usually use only a limited vocabulary to describe the food. Qualitative and quantitative sensory descriptions are reserved for a trained panel composed of 12 assessors. Their work is to generate sensory attributes, to give accurate definition of these terms, and to assess the intensity of each attribute for each product. Sensory and hedonic guidelines which are reviewed regularly and published by standards organizations include: ISO (International Organization for Standardization), AFNOR (Association Française de Normalisation), AMSA (American Meat Science Association), DIN (Deutsches Institut für Normung), BSI (British Standards Institution), UNI (Ente Nazionale Italiano di Unificazione), AENOR (Asociación Española de Normalización y Certificación), and ASTM (American Society for Testing and Materials). Data are analyzed using multivariate statistical techniques, such as preference mapping, which are used to relate sensory and acceptability data in order to understand the acceptance or rejection criteria.

Factors Affecting Consumer Acceptability for Meat

Many experiments have shown that acceptability varies widely among consumers. A large number of factors can influence meat acceptability (Table 7), which is related to the product sensory characteristics: flavor, toughness, and appearance. For most consumers, food flavor seems to be the most important acceptability criterion before texture and appearance. However, in beef, tenderness is probably the most studied characteristic because its variation is wider than that in flavor. If beef is tough, other sensory properties become less important, and as meat tenderness increases, the overall acceptability to the consumer generally increases. For beef, tenderness is a major criterion of acceptability and toughness a criterion of reject. However, for poultry, firm meat may be preferred. Appreciation of the right level of tenderness thus depends on the product and on the individual or groups of consumers. This fact suggests that the sensory expectancies are variable for various meat products.

Table 7 Factors influencing the acceptability of meat products

Sensory characteristics of meat	*Type of meat products*	*Sensitivity of consumer*	*Food habits of consumer*
Flavor	Species	Olfactory	Gender
Texture	Cut	Gustatory	Age
Appearance	Fat		Custom

Appearance of meat also determines acceptance. Many consumers do not want to eat meat with too much visible fat because they consider it harmful to health that dominates over any contribution to flavor. Exudates or off-colors serve as indicators of unsafe products to consumers. Consumers also use color to determine when meat has been cooked sufficiently, essentially as a predictor of flavor and texture.

Acceptability depends on the product and the context, and consumers' preferences may be based on different sensory characteristics: for example, some for its tenderness and some for its grilled meat aroma. Some individuals are very much taste- and smell-oriented, while others divide their attention among multiple sensory inputs. Moreover, what one person likes, another may dislike. Some consumers prefer higher-salt/higher-fat restructured steaks, and others prefer and actively seek out lower-salt/lower-fat steaks. Olfactory and gustatory sensitivity can also play a role in interindividual differences in liking.

Meat acceptability depends strongly on the food habits of consumers. Consumers generally like what they know. For example, in New Zealand, consumers eat much lamb and find the specific flavor of lamb attractive, whereas American people appear not to because they are more used to consuming pork and beef. In an inter-European country study on consumer preferences for dry-cured hams, most of the French and Italian consumers preferred Italian home-style Parma and Bayonne hams, whereas Spanish people preferred Iberian hams. These differences in liking are attributed to the different sensory traits among the different types of ham. Parma and Bayonne have a pink red color, salty taste, and fresh meat flavor, whereas the Iberian hams are generally soft, purplish, slightly sweet, and bitter with a strong, rancid, blue cheese and mushroom aroma.

Most consumers eat meat because it has desirable sensory properties. Many of the sensory attributes depend on the species and cut of meat, which determines to a large extent the cooking method.

Main Factors Affecting the Sensory Characteristics of Meat

Sensory properties have been used to study meats from different origins (species, age, sex, rearing conditions of animals), to monitor processes (aging duration, electrical stimulation, mincing), and to compare products. Such qualitative and quantitative descriptions, without any hedonic connotation, are carried out by a panel of trained panelists.

Texture Sensory profiles of texture describe in detail the different stages of its perception (Table 8). The surface texture is evaluated, followed by an impression of the first bite or mastication stage and then the characteristics of the residual or swallowing stage. Surface properties are smoothness/roughness, surface moisture, or a granular appearance. During the mastication stage, the degree to which the sample returns to its original form (springiness/elasticity), the tenderness/hardness, the fatty perception, and the moisture release (juiciness) can be evaluated.

Table 8 Texture attributes at different stages of mastication

Stage of perception	*Attribute*
First stage: surface perception	Granular
	Rough
	Moist
Mid stage: during mastication	Tender/hard
	Springy/elastic
	Cohesive
	Moisture release
	Moisture absorption
	Juicy
	Gristle
	Stringy
	Fibrous
	Particle size and shape
	Sticky
	Greasy
Final stage: residual perception	Number of chews
	Mouth coating
	Ease of swallowing
	Residual particles

Source: Bett KB (1993) Measuring sensory properties of meat in the laboratory. *Food Technology* 47: 121–127.

Tenderness is the opposite of toughness or resistance to rupture during chewing. Meat tenderness depends largely on the amount and status of the connective tissue and myofibers. The connective tissue is composed mainly by collagen, of which its quantity, quality, and distribution in muscle contribute to the specific character of meat tenderness. The extent of aging (chill storage) increases, whilst a contracted state of the myofibrils decreases tenderness. Collagen content imposes the cooking method of muscles. When the collagen content is low, the muscle can be grilled quickly at high temperature to avoid myofibrillar toughness. However, if meat is rich in collagen, a long cooking in a wet atmosphere and at a moderate temperature is essential to solubilize the collagen and to reduce mechanical resistance.

Juiciness is related to the perception of water released from meat during chewing and to the intramuscle fat content, which has a stimulant effect on salivation.

The residual texture stage depends on the number of chews to prepare the sample for swallowing, the fibrousness of the connective tissue, the denseness, the bolus formation, the ease of swallowing, and how much and what is remaining in the mouth

(Table 8). It includes the description of remaining particles and any residual coating. The essential components of the qualitative description of texture are the same for all meats from different species.

Flavor Flavor refers to the taste and olfactory properties that are perceived by the gustatory receptors of tongue and the olfactory mucous membrane in the nasal cavity. The four basic tastes of sweet, salty, acid, and bitter can be found at low to moderate intensities in the various muscle foods. Sweet and salty have naturally low intensities in fresh meat but can be more intense in cured meats. Astringent and metallic perceptions, which result from the presence of metal ions, are two mouth feel descriptors commonly found in meats. Umami is another taste common to meats; it is given by the presence of monosodium glutamate and nucleotides.

Meat flavor has been shown to be affected by several *ante-mortem* (age, sex, species, nutritional status, stress level, muscle type, fat level and composition) and *post-mortem* (slaughter, carcass handing, aging, cooking, and storage after cooking) factors. Animal age is an important factor; flavor intensity is higher in meats from older animals. Meat contains nonvolatile flavor precursors such as lipids, fatty acids, sugars, organic acids, nucleotides, peptides, vitamins, and volatile components such as products of reactions of oxidation. The flavor of cooked meat is developed by the application of heat that induces Maillard reactions and the degradation of lipids. The Maillard reaction products are formed by reactions between sugars and amino acids or peptides. Heterocyclic compounds produced by the Maillard reaction, especially those containing sulfur, are important and provide meaty, roast, and boiled aroma. Most cooked meats have characteristics brothy, browned, and meat flavor. The accurate determination of flavor is cumbersome because of the multitude of sensory terms and has not been fully elucidated.

The flavor characteristics of meat from different species are listed in Table 9. The lipid fraction probably contains the specific odorant tracers of each species, while the aqueous fraction holds the common nonvolatile meat taste. The main beef flavor attributes developed are the serum/bloody aroma associated with raw beef lean, the brothy/meaty aroma associated with cooked lean meat, the aroma associated with cooked beef fat, and the browned aroma associated with the outside of grilled or broiled beef. The beef flavor note is due to the presence of many compounds: methyl-branched aldehydes, ketones, nitrogenated heterocycles, and other compounds. Lamb exhibits the strongest and the most variable flavor: animal, cabbage, rancid odor, and meaty flavors. The branched fatty acids (4-methyloctanoic and 4-ethyloctanoic acids) are found in abundant quantities in lamb fat and are responsible for the mutton flavor. Abnormal flavors androstenone (boar taints) or other metabolites (skatole) contained in the fatty tissue of male pigs. Cured pork meats have more aroma than fresh meat. The fermentation process used in dry sausage with bacterial starters or the drying process of dry-cured ham leads to the formation of flavor such as buttery, vinegar, nutty, and cheese aroma. Poultry skin contains fat and sulfur compounds responsible for the characteristic chicken aroma. Moreover, inosinic acid, inosine, and guanylic acid may also react with other components of meat

Table 9 Flavor attributes for various species of meat

Beef	*Lamb*	*Pork*	*Poultry*
Cooked beef lean	Gamey/muttony	Porky	Boiled chicken
Brothy/meaty	Fatty	Sex taint/boar odor	Meaty
Cooked beef fat	Bloody/serum	Fatty	Brothy
Serum/bloody	Browned/caramel	Bloody	Browned
Browned	Livery/organ meat	Browned	Liver/organ-like
Grainy/cow-like	Meaty		
Livery/organ meat	Sheep meat		
	Boiled meat		
	Poultry		
	Animal		
	Rancid		
	Oily		
	Buttery		
	Cabbage		
	Roast		
	Rubbery		

Sources: Bett KB (1993) Measuring sensory properties of meat in the laboratory. *Food Technology* 47: 121–127. Rousset S, Young OA and Berdague JL (1997) Diet and growth effects in panel assessment of sheepmeat odour and flavour. *Meat Science* 45(2): 169–181.

during the Maillard reaction and add complexity to the chicken flavor.

Conclusion

Good manufacturing practices, developed in the past 20 years, have introduced new hygienic rules that limit contamination by pathogenic and spoilage bacteria. Consumers are more aware of the quality and storage of meat and of the safety measures taken during its preparation. Consumer preferences are considered by the food industry in order to provide meat products with the desired sensory and nutritional qualities. However, several criteria other than eating quality play an important role in meat acceptability, e.g., price of meat, attitudes and knowledge of consumers towards meat, and convenience in preparing meat. Multidisciplinary studies should be useful in understanding the complexity of eating behavior.

See also: ***Aeromonas***; ***Clostridium***: Botulism; ***Escherichia coli***: Food Poisoning; ***Listeria***: Listeriosis; **Meat**: Hygiene; ***Salmonella***: Salmonellosis; **Sensory Evaluation**: Texture; Taste; **Spoilage**: Molds in Spoilage

Further Reading

AFNOR (2000) *Sensory Analysis – Methodology – General Guidance for Conducting Hedonic Tests in Sensory Analysis Laboratory or in Controlled Conditions Room Involving Consumers. XP V09-500.*

AMSA (1995) *Research Guidelines for Cookery, Sensory Evaluation and Instrumental Tenderness Measurements of Fresh Meat*. Chicago, IL: National Live Stock and Meat Board.

Bett KB (1993) Measuring sensory properties of meat in the laboratory. *Food Technology* 47: 121–127.

Chambers E IV and Bowers JR (1993) Consumer perception of sensory qualities in muscle foods. *Food Technology* 47: 116–120.

Civille GV and Lyon BG (1996) *Aroma and Flavor Lexicon for Sensory Evaluation Terms, Definitions, References and Examples. ASTM Data Series Publication DS 66.*

ICMSF (1996) *International Commission on Microbiological Specifications for Foods, Microorganisms in Foods 5. Characteristics of Microbial Pathogens*. London: Blackie Academic & Professional.

Mottram DS (1998) Flavour formation in meat and meat products: a review. *Food Chemistry* 62: 415–424.

Moskowitz H and Krieger B (1993) What sensory characteristics drive product quality? An assessment of individual differences. *Journal of Sensory Studies* 8: 271–282.

Rousset S, Martin JF, Bayle MC and Touraille C (1996) *Cross-cultural Studies on Food Preference and Consumer Behaviour, AIR-CAT Measurements of Consumer Attitudes and their Influence on Food Choice and Acceptability, Lillehammer, February 3–6.*

Rousset S, Young OA and Berdague JL (1997) Diet and growth effects in panel assessment of sheepmeat odour and flavour. *Meat Science* 45(2): 169–181.

WHO (1997) *World Health Organization, World Health Statistics Quarterly, 50*, 154, no. 1/2. Geneva: World Health Organization.

Young OA, Berdagué JL, Viallon C, Rousset S and Terriez M (1997) Fat-borne volatiles and sheepmeat odour. *Meat Science* 45(2): 183–200.

Sausages and Comminuted Products

J G Sebranek, Iowa State University, Ames, IA, USA

General Information

Sausages and comminuted (ground, chopped, or otherwise subdivided) meats are processed-meat products which offer consumers a wide range of alternatives to the standard fare of steaks, chops, and roasts. These products not only offer meat with different flavors and textures but also allow processors to improve the keepability and safety of meat products. Spices and other ingredients are used to create unique flavors while processes such as curing and drying improve product color and textural properties. The use of ingredients, such as salt and nitrite, and processes, such as cooking, together serve to kill microorganisms and consequently improve keepability and safety of sausage products.

Comminuted meats can be as simple as ground beef, an immensely popular consumer product, or as complex as fermented, smoked, dried hard salami. The flavor, color, texture, and storage stability of ground beef compared with hard salami provide a good example of the extremes which can be achieved with this group of food products. Ground beef is an inexpensive product, popular by itself but also frequently used in combination with other ingredients (vegetables, mushrooms, cheese, pasta, etc.) to increase menu flexibility and eating variety. The hard salami, in contrast, is a relatively expensive, gourmet product that provides a unique gustatory experience by itself. Sausage products also offer consumers a high degree of convenience since many are cooked and may be eaten cold or simply reheated using a microwave. There is little or no waste from these products.

Sausages and comminuted meats are good examples of 'value-added' processing by the food industry whereby relatively low-value meat cuts and

trimmings are dramatically improved for eating enjoyment. Chopping and grinding, for example, can greatly improve the tenderness of tough meat cuts; spices, smoking, and cooking can create unique textures, flavors, and aromas which increase eating pleasure. Each processing step can be achieved in a number of different ways (different comminution methods, spice blends, smoking and cooking processes, etc.) which means there are hundreds of varieties of sausages and comminuted meats available. In addition, the meat industry is constantly striving to create new products in order to remain competitive with the rest of the food industry. This means that consumers can expect to see more new varieties of sausages and comminuted products in the future.

The preparation and processing of sausage products was once considered exclusively an art, with secrets handed down through generations of sausage makers. However, scientific research in the twentieth century has provided the understanding necessary to make this field a sophisticated science.

Historical Background

Sausage products are believed to have been developed well before written history. The word 'sausage' has been described as originating from the Latin word *salsus*, which means salted. This would imply that sausage processing, in the form of adding salt to meat, existed at least several hundred years BC.

It is likely that most sausage-processing techniques were discovered by accident. People who used salt for flavoring purposes eventually noticed that salted meat did not spoil as quickly as unsalted meat. Likewise, smoking was probably discovered as a side-effect of cooking over an open fire, and drying for preservation may have resulted from observing meat exposed to the sun during warm, dry days. The discovery of spices resulted in a tremendous increase in meat product variety as people experimented with different spice combinations and blends. The need for convenience in handling led to the use of sausage casings derived from the gastrointestinal tract of animals slaughtered for meat. Thus, allied industries developed to supply meat processors with spices, casings, and equipment necessary to manufacture sausage products.

The development of cured sausage probably originated with salt-preserved meat where the salt used contained saltpeter. Saltpeter (potassium nitrate) is a relatively common contaminant in crude salt which will result in a unique pink-red cured color when used in meat. Additional effects from saltpeter include a marked resistance to spoilage. Such observations would have greatly encouraged use of this compound for cured meat processing. By the late 1800s, scientific studies had shown that the nitrate ion from saltpeter was the source of meat-curing reactions and, by the 1920s, scientists determined that nitrite rather than nitrate was the active curing agent. Consequently, direct addition of sodium nitrite is the most common means today of producing cured meats. Cured products include not only many sausages and comminuted meats but also intact muscle products such as hams, bacon, and corned beef.

Many of the sausage products common in the USA can be traced to their origins in Europe and other parts of the world. In most cases, the type of product is related to environmental conditions within the geographical region in which they originated. Dry sausage, for example, was largely developed in southern Europe where such products are unlikely to spoil in the warm climate. On the other hand, northern Europe gave rise to cooked sausage where a cold climate permitted longer storage. It has been suggested that the development of highly spiced foods, often typical of warm climates, was encouraged by the ability of spices to prevent and mask off-flavors which are likely to develop at warm temperatures. Many meat products also have names which suggest origin from a specific region or city; examples include Frankfurt (frankfurters) and Bologna (bologna). The development of mechanical refrigeration, modern packaging, and rapid transportation systems has allowed widespread distribution of traditional sausage products all over the world.

Classification of Sausages and Comminuted Meats

Because sausages and comminuted meats are characterized both by processing procedures and by ingredients used, classification of these products largely reflects their processing history. Of course, this also means that classification becomes somewhat arbitrary in cases where several processing techniques may be combined to produce a specific product. Nevertheless, most sausages and comminuted meats can be placed in one of seven general classes (**Table 1**).

Close review of **Table 1** shows that products can be distinguished by five fundamental treatments: seasoning, cooking, curing, smoking, and drying. For example, fresh comminuted meats such as ground beef or restructured nuggets are not cured (do not contain nitrite), usually contain little seasoning, and are not smoked. While 'fresh' may be defined as 'raw, uncooked' meat, it is more specifically understood in the sausage industry to mean 'uncured.' This category of products is becoming increasingly popular as cooked products ready for quick reheating. This popularity has been facilitated by use of batters, breadings,

Table 1 Classification of sausages and comminuted meats

Product group	*Examples*	*General characteristics*
Fresh comminuted meats	Ground beef Ground pork Ground poultry Nuggets Restructured patties, steaks, chops	Uncured, typically uncooked and unseasoned but may include limited seasoning and/or binders; nuggets and restructured products may also be coated with batter or breading. Increasingly, this category of products is found as fully cooked and ready for microwave reheating
Fresh sausage	Pork sausage Bratwurst Italian sausage	Uncured, well-seasoned with salt and spices, usually stuffed into casings but not smoked or cooked; may be found as patties or in bulk quantities besides the traditional sausage links
Uncooked, smoked sausage	Kielbasa Mettwurst	Cured or uncured, seasoned, stuffed into casings and cold-smoked; must be fully cooked before serving
Cooked, smoked sausage	Frankfurters Bologna	Cured, seasoned, stuffed into casings, smoked and fully cooked; may be served cold or reheated. Casings are usually removed prior to sale
Cooked sausage	Braunschweiger Liver sausage	Cured or uncured, seasoned, stuffed into casings and cooked; may or may not include smoke flavorings
Cooked luncheon meats and loaves	Pickle and pimiento loaf Ham and cheese loaf Sandwich spread Scrapple	Cured or uncured, seasoned, cooked, not smoked. Products in this category show wide variety of nonmeat (cheese, olives, pickles, honey, etc.) ingredient use
Dry and semidry sausage	Pepperoni Hard salami Summer sausage Lebanon bologna	Cured, seasoned, stuffed into casings; usually fermented; usually smoked and cooked. Controlled air-drying utilized for dry sausage, with most products characterized by final moisture-to-protein ratio

and other coatings which add new flavors, improve mouth feel, and increase retention of meat flavor and juiciness in these products. Fresh sausage, in contrast to fresh comminuted meats, is well-seasoned with salt and spices and is usually stuffed into casings or formed into patties. Uncooked, smoked sausage, which is relatively uncommon in the USA, is smoked for flavor enhancement but remains uncooked.

Cooked and smoked sausage makes up by far the largest sausage category in the USA and is typically cured, seasoned, smoked, and fully cooked. Because these products are considered ready-to-eat (RTE), they are very convenient and easy to use. This group makes up about 40% of all sausage products in the US, with the most common being frankfurters and bologna. The cooked-sausage category includes products which may be cured or uncured but will usually be stuffed into a cylindrical casing. Because smoke is not usually applied to these products, cooking, when done, is often carried out in hot water instead of hot air, as is typically done for combined smoking/cooking processes. Cooked loaves and luncheon meats is a category of processed meats which includes extensive use of unique nonmeat ingredients such as pickles, pimientos, cheese, olives, honey, and corn meal. These nonmeat ingredients offer a tremendous variety of flavors and appearance characteristics for these products. The dry and semidry sausages may be the most distinctive group of processed meats because, in addition to curing, seasoning, and stuffing into a casing, these products are usually fermented with a microbial culture to acidify the product and facilitate subsequent drying. Semidry products such as summer sausage are not further dried and are ready for distribution after fermentation and cooking. With an appropriate pH (acidity) from fermentation, and adequate vacuum packaging, these products will be shelf-stable and may not require refrigeration. Dry sausage, such as pepperoni, is dried after the fermentation/cooking process to reach a characteristic moisture:protein (M:P) ratio as required by regulation. In the USA, for example, the required M:P ratio for pepperoni is 1.6:1 or less; for Genoa salami, it is 2.3:1, and for beef jerky, it is 0.75:1. For comparison, fresh meat has an M:P ratio of about 3.7:1. The overall result of drying processes for dry and semidry sausage is a product with distinctive flavor, unique texture, and very high resistance to spoilage. The high resistance to microbial growth and spoilage which characterizes this category of sausage products also contributes to improved product safety.

Ingredients Used for Sausages and Comminuted Meats

The ingredients chosen to manufacture sausages and comminuted meats not only determine product identity but are also critical to the quality of the finished

product. Ingredients include both meat ingredients and nonmeat ingredients, both of which play a major role in creating product identity and eating quality.

Meat Ingredients

The selection of meat ingredients to be used for manufacturing sausages and comminuted meats is one of the first, and most important, decisions made by meat processors. The first requirement is that the meat be of high quality and as fresh as possible. It is sometimes assumed that meat quality is not critical to final product quality because spices and other ingredients will overwhelm and mask small quality differences in meat ingredients. It has been clearly shown by research, however, that the subtle chemical changes which occur in meat before processing, such as the beginning of lipid oxidation, can have major effects on finished product quality.

The second requirement for meat ingredients to be used for sausages and comminuted meats is information on composition and color of each ingredient. The various meat ingredients available are highly variable in color, composition, and ability to retain moisture and fat when cooked. Consequently, sausage processors must choose and combine meat ingredients in the proportions that will create the finished product qualities desired. The meat ingredients are not only the source of the fat, water, and protein content of the finished product, but also affect the texture, juiciness, flavor, and appearance of the product. When the desired product qualities have been achieved, it becomes important to reproduce the same product consistently every time it is manufactured.

The differences in meat ingredients are the basis for the use of least-cost computer formulations by the meat industry. For these computerized calculations, all meat ingredients are first chemically analyzed for input data to be used in the program. The computer program then determines the proportions of each ingredient needed to achieve the desired product characteristics at the lowest possible cost. Because there are dozens of different meat ingredients available and each will vary from day to day, computerized least-cost formulation programs are very useful. An additional advantage to cost control, however, is the improvement in product consistency and uniformity that is achieved at the same time.

Nonmeat Ingredients

After selecting the meat ingredients, a sausage manufacturer is faced with decisions on nonmeat ingredients. Many of the nonmeat ingredients are essential to product characteristics; salt for most sausages is one such example and the only decision to be made concerns the amount of salt to be used. Other nonmeat ingredients such as spices offer the opportunity to create unique products by using different blends and combinations.

Table 2 includes the most important nonmeat ingredients with the major functions and typical use levels of each. In many cases, nonmeat ingredients are restricted by regulations which may vary in different countries of the world. When trying to determine the limits of nonmeat ingredient use, it is important to check the regulations appropriate to where the product is being manufactured and/or sold. Awareness of the specific functions of each nonmeat ingredient is important because each contributes specific product characteristics that often cannot be produced any other way.

Salt, for example, is responsible for several different characteristics essential to sausage products and is much more than a simple flavoring ingredient. While the flavor from salt is important, salt is even more critical to microbial control (both for spoilage and food safety), product texture, and binding of fat and water. Because there has been concern about human dietary sodium intake and the potential relationship of sodium to hypertension, considerable research has been conducted on the reduction of salt levels in processed meats. While the research has shown that salt can often be reduced to some degree, it has also become clear that microbial problems and cooking losses quickly increase as salt is reduced, especially below 2%. Salt is also necessary for typical texture and binding qualities of meat because the meat proteins responsible for these characteristics are solubilized by salt. The proteins must be at least partially solubilized to be effective. Protein solubility and binding is one of the main reasons a small (0.5–1.0%) amount of salt is used in restructured meats and nuggets. Recent developments with enzyme technology, such as the use of transglutaminase to achieve meat protein cross-linking, may offer an alternative to the use of salt for binding in restructured meats.

While salt is used for every sausage product and most comminuted meats (except ground beef, pork, lamb, or poultry), one ingredient that differentiates product classes very clearly is sodium (or potassium) nitrite, which separates cured and fresh products. Nitrite is responsible for cured meat color and flavor, as well as being a strong microbial inhibitor. Sodium (or potassium) nitrate is often described as a curing ingredient but is ineffective by itself. The only function of nitrate in processed meats is to form nitrite which then contributes all the cured meat characteristics. The use of nitrite (and nitrate) is strictly regulated. The most common limitation, and that which applies in the USA, is a maximum of 156 mg sodium nitrite per kilogram (156 p.p.m.) of meat. One exception to

Table 2 Nonmeat ingredients used for sausages and comminuted meats

Ingredient	*Function*	*Level of use and applications*
Salt (sodium chloride)	Flavoring, microbial inhibition, protein solubilization (which controls fat and water retention in cooking), product texture, and product adhesion	0–5% in all sausages and many comminuted meats Dry sausage may be as high as 5%, other sausages 1.5–2.5%, restructured comminuted meats 0.5–1.0% and ground beef 0%. There are no regulatory limits
Sodium or potassium nitrite	Curing agent responsible for cured color, flavor, and microbial control, particularly for *Clostridium botulinum*	Limited to 156 p.p.m. in most products. Used in all cured meats. Maximum levels established by regulation
Sodium ascorbate or erythorbate	Accelerates the curing reaction of nitrite, contributes antioxidant (protection from off-flavor) function	Limited by regulation to 550 p.p.m. when used Usually included with nitrite in meat curing
Phosphates (sodium tripolyphosphates; sodium polyphosphate, glassy; tetrasodium pyrophosphate; sodium hexametaphosphate)	Increase protein solubility (with related fat and water binding) and product adhesion, increase meat pH, chelate metal ions that encourage rancidity, contribute to microbial inhibition	0.2–0.5% in both cured and fresh products Maximum level of 0.5% is controlled by regulation
Sweeteners (sucrose, dextrose, lactose, corn syrups, sorbitol)	Provide sweetness, improve water retention, substrate for fermentation	Usually used at 0.5–1.0%. Sucrose and dextrose are unlimited but corn syrups and sorbitol are restricted to 2.0% by regulation
Starter cultures	Produce lactic acid from sugars and lower product pH	Limited to 0.5% (with water) for fermented sausage
Binders and extenders (milk proteins, soy proteins, cereal products, starches, carrageenans, hydrocolliods)	Improve water binding, product slicing, cooking yields, reduce product costs	May be used in any product. Generally limited (in USA) to 3.5% except soy isolate (2%), with some exceptions
Antioxidants (BHA, BHT, TBHQ, propyl gallate) and synergists (citric acid, sodium citrate, monoisopropyl citrate, monoglyceride citrate)	Inhibit rancid flavors; synergists are used to chelate metal catalysts, making the antioxidants more effective	Used in fresh sausage and dry sausage. Limited by regulations (in USA) to 0.01% of the fat in fresh sausage (0.02% in combination) and to 0.003% of product weight in dry sausage (0.006% in combination)
Antimicrobial agents (sodium or potassium lactate, sodium diacetate, potassium benzoate, potassium sorbate, microbial fermentation products such as pediocins)	Inhibit microbial growth, especially pathogenic bacteria	Limited by regulation with amounts ranging from 0.1 to 3.5% depending on product and specific use

BHA, butylated hydroxyanisole; BHT, butylated hydroxy toluene; TBHQ, tertiary butylhydroquinone.

this is pumped bacon which is limited to 120 p.p.m. Nitrite is a remarkable ingredient in that it performs its functions at extremely low concentrations and no other ingredient has been found that will produce the same effects. Nitrite is also important because, in addition to preventing spoilage, it is a highly effective inhibitor of *Clostridium botulinum*, thus providing food safety assurance. Cured, cooked meat in vacuum-sealed cans or packages, an environment where *C. botulinum* is most likely to grow, is very well protected against the likelihood of causing botulism. Most modern curing processes also include sodium ascorbate or erythorbate as accelerators to speed up the nitrite curing reactions and facilitate more rapid production systems. These two ingredients differ only in that ascorbate is biologically active as vitamin C while erythorbate is not. Both are equally effective when used to accelerate meat curing reactions. Interestingly, there is a persistent rumor which circulates among consumers that erythorbate means earthworms and products with erythorbate are to be avoided. Because this unfounded rumor has persisted, some meat-processing companies have begun indicating on product labels the true source of erythorbate, that is, 'made from sugar.'

Other multifunctional nonmeat ingredients include the polyphosphates. These compounds, of which there are several, interact strongly with salt for protein solubilization. Consequently, the phosphates are major contributors to product texture and binding characteristics. The importance of the phosphates becomes more marked in meat products with reduced

salt content and, in practice, one of the roles of phosphates is to permit lower concentrations of salt to be used. Restructured, comminuted meat products and nuggets which usually contain only 0.5–1.0% salt depend greatly on the presence of phosphates to achieve protein solubility and binding. Most of the approved phosphates are good chelating agents and will bind (chelate) contaminants such as metal ions which commonly initiate rancid flavor production. Thus, phosphates also function as flavor protectors.

Spices are ingredients traditionally used for flavoring of sausage products. More recently, spices have been found to have other important functions as well. For example, pepper, ginger, cloves, sage, and rosemary are effective antioxidants and serve to protect product flavor. Other spices have been noted to contribute to inhibitory effects on bacterial growth. However, because spices are dried plant products, they are often a source of microbial contamination which can be a significant problem in processed meats. Spices are usually treated with ethylene oxide or irradiation to reduce the microbial contamination. Spice extractives or liquid spices may also be used to avoid the microbial problem. Flavoring agents such as nucleotides, monosodium glutamate, and hydrolyzed proteins are sometimes used to increase meat and/or spice flavor intensity.

Sweeteners, which are most often sugars, might also be considered flavoring agents but, in addition to contributing sweetness, these ingredients also help bind water because they are hygroscopic. Sucrose or cane sugar may be used but dextrose or corn sugar is more common. Dextrose is also utilized for fermented sausage as the carbohydrate of choice for fermentation. The amount of lactic acid produced by the fermentation and the product tanginess it produces can be controlled by the amounts of dextrose used. Dextrose is also well-recognized as a reducing sugar which will result in caramel-colored surfaces when heated in the presence of protein, a color desired for some meat products.

Starter cultures are another nonmeat ingredient developed for use in fermented sausage. The use of starter cultures has given fermented sausage processors immensely improved control over fermentation processes. By using pure strains of desired bacteria, processors have avoided many of the problems which were caused by 'wild' cultures inherent to the meat ingredients used. There are several types of bacteria used for sausage starter cultures, with specific strains of *Pediococcus, Lactobacillus*, and *Micrococcus* being most common in the USA. In addition to lactic acid, starter cultures may also produce a variety of other fermentation end products, some of which contribute to flavor. Thus, the specific culture used may create a unique flavor and product as a result of the fermentation process.

A wide variety of extenders or binders may be used to improve water and fat retention, slicing or cooking yields of processed meats. These compounds differ greatly and range from milk and plant proteins to starches and hydrocolloids. This category of nonmeat ingredients is closely regulated, with limits for most being 3.5% or less.

A unique group of nonmeat ingredients permitted only in fresh sausage or dry sausage is the antioxidants and synergists. These compounds are used to retard the development of rancidity and thus protect flavor. Included are butylated hydroxyanisole (BHA), butylated hydroxy toluene (BHT), tertiary butylhydroquinone (TBHQ), and propyl gallate. These compounds are frequently used in combinations to gain improved effectiveness. While very effective in meat systems, these antioxidants are considered synthetics and are not normally found in meat products unless they are added. Some questions have been raised concerning the use of these antioxidants and natural antioxidants have been sought as alternatives. Most natural antioxidants have been found to be less effective, though some show enough antioxidant activity to be useful. One such compound is rosemary extract, which has significant antioxidant activity and is widely accepted because it is derived from the spice.

Sausages and Comminuted Meats in the Human Diet

Consumers are sometimes critical of the dietary value of sausages and comminuted meats despite the fact that these products provide high-quality protein, B vitamins, and several minerals such as iron. Dietary concerns are usually based on fat and caloric content, salt (sodium) concentrations, and presence of additives such as nitrite and synthetic antioxidants.

Concern for fat and caloric content has been most significant for traditional sausage products which have been formulated with 30% or more fat. A limitation to fat reduction in the past has been the tough rubbery texture that occurs in low-fat, high-protein products if other adjustments are not made. Recent technology developments with nonmeat ingredients have permitted meat processors to offer reduced-fat and fat-free products with very acceptable textural characteristics. Thus, consumers are given a choice and can seek out low-fat products by reviewing the product labels. Salt (sodium) content has also been reduced in many products, though sausage products will always contain higher sodium levels than fresh meat. Sodium content of processed meats may remain

a concern to individuals who are hypersensitive to dietary sodium. Product labels provide information on sodium content so that consumers can make educated choices. Nitrite and synthetic antioxidants have both been controversial because of concern that these compounds may be potential carcinogens. Nitrite, however, is not, by itself, a carcinogen, though it can react with other components in meat systems to form carcinogenic nitrosamines. Past research has improved curing procedures to the point where virtually all nitrite is utilized in meat curing reactions with none available for formation of carcinogens. Recent research has shown that residual nitrite in commercial processed meats has declined dramatically from that found a few years ago. Total nitrite consumption from processed meat is presently less than 5% of the total dietary intake. Concerns for the synthetic antioxidants have not been investigated as thoroughly as for nitrite. However, in this case, other antioxidants may be used to replace the synthetics if necessary – an alternative that does not exist for nitrite.

See also: **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; **Beef**; **Browning**: Nonenzymatic; **Convenience Foods**; **Curing**; **Fermented Foods**: Fermented Meat Products; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Food Safety**; **Meat**: Analysis; **Plant Design**: Process Control and Automation; **Pork**; **Poultry**: Chicken; Ducks and Geese; Turkey; **Spoilage**: Bacterial Spoilage

Further Reading

Cassens RG (1994) *Meat Preservation: Preventing Losses and Assuring Safety*. Trumbull, CT: Food and Nitrition Press.

Claus JR, Colby J-W and Flick GJ (1994) Processed meats/poultry/seafood. In: Kinsman DM, Kotula AW and Breidenstein BC (eds) *Muscle Foods*, pp. 106–162. New York, New York: Chapman and Hall.

Girard JP (1992) Smoking. In: Girard JP (ed.) *Technology of Meat and Meat Products*, pp. 165–205. Chichester, West Sussex: Ellis Horwood.

Girard JP and Bucharles C (1992) Acid fermentation. In: Girard JP (ed.) *Technology of Meat and Meat Products*, pp. 138–164. Chichester, West Sussex: Ellis Horwood.

Goutefonzea R (1992) Salting and curing. In: Girard JP (ed.) *Technology of Meat and Meat Products*, pp. 115–137. Chichester, West Sussex: Ellis Horwood.

Hammes WP and Hertel C (1998) *New Developments in Meat Starter Cultures. Meat Science*, vol. 49, pp. 5125–5138. Exeter: Elsevier Science.

Hedrick HB, Aberle ED, Forrest JC, Judge MD and Merkel RA (1994) *Principles of Meat Science*, 3rd edn. Dubuque, IA: Kendall Hunt.

Hugas M (1998) *Bacteriocinogenic Lactic Acid Bacteria for the Biopreservation of Meat and Meat Products. Meat Science*, vol. 49, pp. 5139–5150. Exeter: Elsevier Science.

Ince K (1998) *Dry Fermented Sausages. Meat Science*, vol. 49, pp. 5169–5178. Exeter: Elsevier Science.

Knutson J (1999) *American Meat Institute – Meat Facts*. Mt Morris, IL: Watt.

Romans JR, Costello WJ, Carlson CW, Greaser ML and Jones KW (1994) *The Meat We Eat*, 13th edn. Danville, IL: Interstate.

Rust RE (1987) Sausage products. In: Price JF and Schweigert BS (eds) *The Science of Meat and Meat Products*, pp. 457–486. Westport, CT: Food and Nutrition Press.

Schmidt GR (1987) Functional behavior of meat components in processing. In: Price JF and Schweigert BS (eds) *The Science of Meat and Meat Products*, pp. 413–430. Westport, CT: Food and Nutrition Press.

Taubes G (1998) *The (Political) Science of Salt. Science*, vol. 281, pp. 898–907. Washington, DC: American Association for the Advancement of Science.

Analysis

B D McLean, Chemistry and Biochemistry Dept, Chipping Campden, UK

Introduction

Meat consists essentially of water, protein, fat, and mineral salts (ash) and it is these proximate analyses that are carried out routinely in many laboratories. This is often to identify the data which can be used for labeling purposes, or to calculate the apparent meat content. This in turn can be used to calculate the added-water content of products such as ham and bacon that contain added curing solutions. Meat products such as bacon, sausages, and burgers often also contain ingredients such as wheat rusk, sugar, and soya and hence also contain carbohydrates. The level of carbohydrates needs to be known when calculating energy content and possibly meat content and is routinely estimated 'by difference' using the formula in **Table 1**. This carbohydrate 'by difference' value is usually sufficient for routine work; however, if necessary the carbohydrate can be analyzed directly from the analysis of sugar, starch, and dietary fiber which makes up the bulk of the carbohydrate in meat products. **Table 1** also lists the most common analyses carried out on meat and meat products.

Sampling

Meat, being a natural product, is very variable in its composition. Raw meat consists essentially of muscle

Table 1 Common analyses on meat and meat products

Analyte or test	*Commonly used methods*
Nitrogen	Kjeldahl or Dumas method
Fat	Acid hydrolysis and ether extraction or Soxhlet-type extraction of 'free fat'
Ash	Incineration at 450–550 °C
Water/moisture	Oven drying at 101–105 °C (with or without use of sand)
Carbohydrate	Calculation by difference from analytes above using the formula Carbohydrate (by difference) = 100 − (% fat + % water + % protein + % ash)
Qualitative identification of meat, cereal, soya, onion, heart, kidney, MRM, etc.	Microscopic examination
Soya protein	Soya immunoassay
Connective tissue content (collagen)	Hydroxyproline determination by colorimetric method
Indication of previous freezing	β-HADH enzyme test
Species of meat present	Species immunoassays and DNA techniques

β-HADH, β-hydroxyacyl-coenzyme A-dehydrogenase; MRM, mechanically recovered meat.

and fatty tissues held together by tough connective tissue. It is somewhat difficult therefore to obtain a homogeneous sample from such a mixture that is suitable for analysis. Prior to the analysis of meat and meat products, it is almost always necessary to chop and blend the samples. It is also necessary to consider the amount of sample to take, for example, the number of entities such as pies and sausages to include, and what part is to be included (e.g., should the rind be removed from bacon? Should the pastry be separated from pies?). The presence of pieces of rind in the final sample makes it difficult to blend evenly and it may be preferred to remove the rind prior to blending. The more uniform the final mix, the easier it will be to obtain good-quality repeatable analytical results. Once blended, the samples must be stored in airtight containers with the minimum amount of air space to avoid loss of moisture. If they are not to be analyzed immediately (the preferred option) then they must be deep-frozen to prevent chemical changes to the proteins and fat.

Quality of Results

To show that analytical results are under control it is necessary to carry out quality control procedures. These may involve the analysis of reference materials (prepared in-house or purchased with precertification), replicate analysis, or participation in proficiency testing schemes. One proficiency scheme which has a meat test series called 'nutritional components' is the UK Food Analysis Performance Assessment Scheme (FAPAS). This includes the analysis of fat, water, ash, protein, hydroxyproline, and salt (sodium and chloride) in meat and meat products and it is a useful way to show the quality of these analyses. Other series are available which are suitable for meat analyses including those covering the composition of fats, presence of veterinary medicines, and levels of trace elements.

Nitrogen and Protein

This is an important analysis in determining the composition and meat content of meat and meat products. For many years the protein content has been determined by the Kjeldahl procedure, often in one of its semi-automated forms. This assumes that the proportion of non-protein nitrogen in a food material is too small to be significant and that a determination of total nitrogen (excluding that present as nitrate and nitrite, which is not measured by Kjeldahl anyway) will therefore be an accurate reflection of the total protein content. In the Kjeldahl procedure, the entire food sample is digested at high temperatures (420 °C) with boiling sulfuric acid under reflux. Essentially, the protein is converted to ammonium ions by digestion with superheated sulfuric acid. The digestion includes potassium sulfate and a catalyst of copper, mercury, or titanium, although the use of mercury is frowned upon these days due to its environmental pollution problems. The nitrogen content is then determined by making the digest alkaline and distillation of the resulting ammonia into boric acid or standardized acid solution. The ammonia is then titrated with standard acid or alkali respectively. The nitrogen content is converted into a protein content using a factor of 6.25, which assumes 16% of nitrogen is present in meat proteins.

There is now increasing use of the Dumas analysis technique to determine the nitrogen content of foods. In this procedure, the nitrogen-containing constituents of the sample are combusted at high temperature in the presence of oxygen and helium. This produces oxides of nitrogen which are then reduced over copper to gaseous nitrogen and measured by thermal conductivity. The

Table 2 Comparison of the nitrogen content of chicken as determined by the Dumas and Kjeldahl methods

Sample	*% Nitrogen*	
	Kjeldahl	*Dumas*
1	3.14	3.28
2	3.18	3.26
3	3.19	3.27
4	3.22	3.27
5	3.25	3.28
Mean	3.196	3.271
Calculated protein	19.98%	20.44%

level of nitrogen is then converted to protein content by the conventional factors mentioned above. Modern developments in instrumentation have allowed the Dumas procedure to be automated to a far greater extent than the digestion and steam distillation stages of the Kjeldahl approach.

It should be noted that the Dumas technique does not necessarily measure the same substances in meat as the more traditional Kjeldahl technique. Table 2 shows some details of the results obtained on five samples of chicken by both of these methods. It has been shown that the means of the Dumas results are statistically significantly higher than the Kjeldahl method and would give rise to a calculated apparent meat content approximately 2% higher than the latter method. This difference may not normally be a problem but it may be of significance in calculations of added water content.

The other meat species tested in the work were not significantly different. It must also be noted that the official nitrogen factors used to convert nitrogen levels into a meat content have so far been determined only by the Kjeldahl procedure. The British Standard quotes a level of 0.1 g per 100 g for the repeatability of the nitrogen determination using the Kjeldahl procedure. In practice, a repeatability of 0.07% absolute (95% confidence) is achievable.

Ash

The ash content of foods is the least important value and is relatively easy to determine. The method for ash is well-documented, with the food normally being incinerated at a temperature of between 525 and 575 °C. The ash content of meat is normally about 1%, but when salt is added (e.g., cured meats), then this level increases proportionally. The ash consists of minerals such as sodium (especially if salt is present), calcium, potassium, magnesium and iron and zinc, with chloride and phosphates. The presence of bone particles in meat and meat products will obviously also increase the ash content.

Where the ash is to be used for the determination of the salt content or trace metals analysis, then the lower end of the quoted temperature range for ashing is recommended as chloride and some metals can volatilize above 550 °C. In general, the repeatability (95% confidence) of the ash determination should be 0.1% absolute or less.

Fat

The fat present in meat may be considered as consisting of free fat and bound fat. The free fat can be extracted by less polar solvents such as petroleum spirit or diethyl ether. The bound lipids need to be released from the protein matrix by the use of acid or alkaline hydrolysis. This allows all of the fat present in the food to be extracted and measured gravimetrically.

Problems are sometimes encountered when laboratories achieve low-fat-content results due to the use of analytical methods that do not remove all of the bound fat. These are normally based on modifications of the traditional Soxhlet solvent extraction process determined on commercially available equipment. When such fat results are declared on the label of the food, the data have been found to be low by enforcement laboratories that frequently use acid hydrolysis methods to remove all of the fat present from the food. On some occasions, laboratories have used 'correction factors' to increase the free fat to the expected level of total fat. The use of such correction factors is not good practice.

The Werner-Schmid and Weibbull-Stoldt methods are particularly good methods for the determination of fat in meat as they involve an acid hydrolysis step that releases the bound fat prior to extraction. The repeatability level quoted in the British Standard of 0.5% is achievable at low level but is very difficult for laboratories to achieve consistently at levels above 30% fat.

Moisture

Moisture or water is normally determined in meat by drying at temperatures between 101 and 105 °C. This is achieved by weighing a few grams of the blended food with a small amount of sand in a dish. The inclusion of a rod assists in mixing the meat with the sand. The meat–sand mixture in the dish is then partly dried on a water bath and mixed further to form a friable mass before further drying in an oven at around 103 °C. Using this procedure it should be noted that any substance that is volatile at 103 °C would be measured as water. Alcohol or acetic acid, which may be present in some meat products (e.g.,

pâté with spirits), would also be lost using this procedure. The repeatability (95% confidence) of this method is generally of the order of 0.6% absolute, and laboratories are normally able to achieve this level of precision.

It is becoming increasingly common to use microwave drying to determine the moisture content of foods. It is essential when using this procedure to insure that the microwave heating does not burn the food and the settings on the microwave must be determined with care. This can be achieved by comparing the results of microwave drying with the more traditional oven-drying reference method and adjusting the drying time accordingly.

For rapid results, instruments are available which consist of a balance which has an infrared heater above it. Samples are weighed on to the balance and then immediately dried using the heating element to a constant weight.

Connective Tissue Content

Enforcement authorities are increasingly interested in the connective tissue content (rind, skin, gristle, etc.) of meat and meat products. Connective tissue content is also used as a measure of the quality of exported meat products within the EC. The amino acid hydroxyproline is present in animal protein collagen at a relatively high and consistent level compared with muscle proteins. The determination of hydroxyproline content of meat and meat products therefore allows an assessment of the level of connective tissue. This in turn allows a decision to be made on whether the amount of connective tissue in a product is acceptable.

The method of analysis used for the determination of hydroxyproline is essentially a colorimetric procedure. After hydrolysis of a test portion of the meat in sulfuric acid at 105 °C, the hydrolysate is filtered and diluted. This is followed by an oxidation step with chloramine T.

The resulting solution is then treated with *p*-dimethylaminobenzaldehyde, which produces a red compound, the amount of which is measured using a spectrophotometer.

In order to calculate the amount of connective tissue that is present in foods, a number of factors have been derived (Table 3). The procedure documented by Lord and Swan is commonly used by enforcement authorities to determine the connective tissue content of meat and meat products.

Table 3 Factors used to convert hydroxyproline data into useful parameters

1. Hydroxyproline × 37 = % Wet fat-free connective tissue
2. Hydroxyproline × 1.28 = % Connective-tissue nitrogen
3. Hydroxyproline × 8 = % Collagen

Calculation of Meat Content and Added Water Content

Analytical data can be used to calculate the apparent total meat content of meat and meat products. This is achieved by measuring the amount of nitrogen present and applying standard factors to determine the apparent fat-free meat content and, with a further allowance for fat, the apparent total meat content. This is the basic method described by Stubbs and More, which has changed little since its publication in 1919. This approach to calculating apparent total meat content can be modified to allow for the presence of the more complex ingredients now used – such as soya – which contribute protein (and so nitrogen), but not meat. Also, the lean meat content can be estimated by further calculation allowing for the level of fat deemed to be appropriate for lean meat.

The principle of the Stubbs and More method involves the estimation of the apparent fat-free meat content, from the amount of nitrogen originating from the meat. Factors are used for the conversion of the nitrogen content to meat content. An allowance is then made for the amount of meat fat to determine the apparent total meat content. 'Nitrogen factors' are available for the conversion of the nitrogen content to raw meat equivalent for many meats. The average nitrogen factors used for meat content calculation have been the subject of much debate; however, the Nitrogen Factors Sub-Committee of the Analytical Methods Committee of the Royal Society of Chemistry publish the most widely accepted values (Table 4). It should be remembered that a nitrogen factor is the average nitrogen content of the fat-free meat and represents a range of values and not a single exact value. A worked example, applying this approach to roast lamb, and including the relationship between nitrogen content and protein content, is given in Table 5.

The above basically holds true for meat and meat products which consist essentially of meat. However, for more complex products (e.g., sausages, burgers), further corrections must be applied for any sources of non-meat nitrogen. If these corrections are not applied, the calculation would give an artificially elevated meat content result.

Lean meat content is often referred to by legislation. It is not possible to determine the lean meat content of a food due to the variability of the fat content of lean meat (e.g., the lean meat from pork can naturally contain from less than 5% to

Table 4 Sources of official nitrogen factors and hydroxyproline data for meat

Type of meat	*Nitrogen factor (NF)*	*Reference*	*Comment*
Pork–general	3.5	Analyst 1991, 116, 761[a]	Also Analyst 1986, 111, 969[a]
Pork–leg	3.49	Analyst 1991, 116, 761[a]	Also Analyst 1986, 111, 969[a]
Pork–neck	3.38	Analyst 1991, 116, 761[a]	Also Analyst 1986, 111, 969[a]
Pork–hand	3.42	Analyst 1991, 116, 761[a]	Also Analyst 1986, 111, 969[a]
Pork–loin	3.66	Analyst 1991, 116, 761[a]	Also Analyst 1986, 111, 969[a]
Pork–belly	3.5	Analyst 1991, 116, 761[a]	Also Analyst 1986, 111, 969[a]
Beef	3.65	Analyst 1993, 118, 1217[a]	
Turkey–whole	3.65	Analyst 2002, 127, 859–869	
Turkey–dark meat	3.45	Analyst 2002, 127, 859–869	
Turkey–breast	3.9	Analyst 2002, 127, 859–869	
Chicken (broiler) – whole, skin off	3.55	Analyst 2000, 125, 1359–1366	Contains data for skin, breast, leg, thigh and dark meat of hen and broilers
Chicken (broiler) – dark meat, skin off	3.3	Analyst 2000, 125, 1359–1366	Contains data for skin, breast, leg, thigh and dark meat of hen and broilers
Chicken (broiler) – breast, skin off	3.85	Analyst 2000, 125, 1359–1366	Contains data for skin, breast, leg, thigh and dark meat of hen and broilers
Mutton	3.47	Analyst 1995, 120, 1823[a]	
Lamb	3.5	Analyst 1996, 121, 889[a]	
Kidney	2.7	Analyst 1966, 91, 538	
Tongue – ox or pig	3.0	Analyst 1967, 92, 326	
Veal	3.35	Analyst 1965, 90, 256	
Ox liver	3.45	Analyst 1964, 89, 630	
Blood	3.2	Analyst 1968, 93, 478	
Pigs, liver	3.65	Analyst 1964, 89, 620	

[a]Includes hydroxyproline data.

Table 5 Example calculation to determine the meat content of roast lamb from analytical data

Fat	17.9%
Water	55.3%
Ash	0.7%
Nitrogen	4.18% (equivalent to 26.1% protein (4.18 × 6.25))
Carbohydrate	0
The appropriate nitrogen factor for lamb is 3.50	
Apparent fat-free meat content $=\frac{4.18}{3.50}\times 100 = 119\%$	
and adding the fat content gives:	
Apparent (or raw equivalent) total meat content $= 119.4 + 17.9\%$ (fat) $= 137\%$	

Note: Meat contents can exceed 100% because on cooking the moisture reduces and the nitrogen increases above the level naturally present in meat. This gives a raw equivalent meat content of over 100%.

over 12% fat). However, it is usual to estimate the lean meat content by adding 10% of the fat-free meat on to the fat-free meat content. This assumes that lean meat contains 10% fat. Therefore, a food with an apparent fat-free meat content of 67% would have an apparent lean meat content of 74% (viz. $67 \times 100/90$). Where the actual fat content of the food is less than the 10% allowance being added, then it is usual to add all of the fat content of the food on to the apparent fat-free meat content to give an estimation of the apparent lean meat content. This will also equal the apparent total meat content.

Added water is present in some whole-meat products such as ham and bacon due to the use of curing solutions. Such added water is often subject to maximum limits or must be declared in the labeling.

The amount of added water can be estimated, by calculation, in various ways. The most widely used formula for calculating the added-water content of cured meat products is based on the Stubbs and More approach. Using this method, the added water is calculated by subtracting the total meat content and all of the added ingredients from 100 (Table 6). This is therefore an indirect method and the result would be in error if any of the analytical parameters or the meat content were incorrectly determined.

Species Testing

It is often necessary to check the species of meat present in meat mixtures and, in some instances, of individual pieces of meat. Over the last 30 years

Table 6

Added water (%) = 100 − (% Apparent total meat content + % salt + % carbohydrate + % other ingredients)
As an example, a sweet cured bacon containing salt (3%), sugar (2%), and with an apparent total meat content of 88%, would contain, for example:
Added water (%) = 100 − (88 + 3 + 2) = 7%

methods have progressed from assessment of the composition of the fat present in the meat, and electrophoretic methods of protein separation and identification, through to the use of serological techniques. Current procedures are based on immunological techniques in the form of enzyme-linked immunosorbent assays and dipstick tests to the newer polymerase chain reaction methods which use the sample's DNA to identify the species present. As yet, the methods do not offer sufficient quantitative precision for it to be possible to determine precisely (with certainty) the amount of the individual species present.

The immunological test takes the form of kits for both raw and cooked meats. As an example, the kit for cooked meat is a noncompetitive sandwich-type enzyme immunoassay that uses a biotin-avidin enhancement process. The antibody is designed to recognize antigens in the cooked meat. Raw or uncooked meat may be tested by heating prior to the extraction of the target antigen. This kit is widely used within the meat industry for routine meat species analysis. However, it does not differentiate between poultry species and hence results can only be reported as poultry. A raw species kit, which is also available, will differentiate between chicken and turkey meat.

The relatively new methods use the analysis of the DNA from the species concerned. These are gaining popularity and are currently the subject of extensive method development. In the cleaved amplified polymorphic sequences (CAPS) method, DNA is first extracted from the meat sample. A specific region of the mitochondrial cytochrome *b* gene is then amplified using the polymerase chain reaction and the amplified product is subject to digestion using a range of restriction endonucleases. The products generated by the restriction endonucleases are subject to electrophoresis on an agarose gel and stained with ethidium bromide dye. The gel is visualized under ultraviolet light and a photographic record of the resulting patterns is obtained.

Spoilage

During spoilage, the proteins present in meat break down and the fat can become rancid. The assessment of the freshness of meat can normally be easily assessed from the odor of the food. It also often includes a microbiological assessment and measures to assess the breakdown of proteins and fat. With respect to the microbiological aspects of meat spoilage, there are wide ranges of microorganisms that can grow on meats and result in spoilage, including *Pseudomonas*, Enterobacteriaceae and lactic acid bacteria. The particular group of organisms responsible for meat spoilage will vary from product to product, depending on the chemical characteristics such as pH, water activity, and storage conditions used, i.e., temperature and atmosphere.

The protein breakdown can be assessed by using the traditional total volatile nitrogen determination, otherwise known as total volatile bases. This measures the protein breakdown products and is carried out by distillation of a portion of the meat from slightly alkaline conditions and titration of the ammoniacal distillate with standard acid. As an example, beef is normally considered to be acceptable if the total volatile nitrogen value does not exceed 16.5 mg nitrogen per 100 g food.

During breakdown of meat, the fat may become rancid and the measures of this rancidity which are often used are those of free fatty acid content and peroxide value of the fat. The free fatty acids are derived from the breakdown of the triglycerides, and the results of the free fatty acid determination is generally expressed as oleic acid. The fat is extracted from the food, and the acidity of a measured amount of the fat is determined by titration. In order to carry out this titration, the fat is dissolved in neutral alcohol and the solution is titrated with standard sodium hydroxide solution using phenolphthalein indicator. A preferred option, to avoid the need to evaporate down the fat extract and hence incur further fat breakdown, is to extract the fat from the meat with chloroform and carry out the titration directly on this extract. A portion of the filtrate can be evaporated to determine the amount of fat present. It is generally accepted that most fresh beef samples would not have a free fatty acid content (calculated as oleic acid) above 1.2%. The peroxide value is a measure of the peroxides contained in the fat. During storage, peroxide formation is slow at first during an induction period, which may vary from a few weeks to several months, depending on the fat involved. It is usually determined volumetrically from the reaction of the bound oxygen with potassium iodide in acid

Table 7 Changes in chicken meat stored in a fridge at 4 °C

Number of days after purchase	*Appearance and odor*	*Total volatile bases (mg·100 g^{-1})*	*Acidity of fat (%) (oleic)*
0	Normal	12.8	0.54
2[a]	Normal	12.0	0.51
5	Some off-odor; appearance normal	30.1	1.5
7	Off-odor; green/yellow discoloration	31.0	1.3
9	Increasing odor and color; slimy	39.7	1.3
12	Unpleasant odor, green color and slimy consistency	46.4	1.7

[a]Best before date.

solutions. The liberated iodine is then titrated with sodium thiosulfate. Fresh oils normally have a peroxide value of well below 10 mmol kg^{-1}.

As rancidity is a complex matter, it is advisable to carry out as many tests as possible on doubtful samples. Other tests that are used to evaluate the oxidation of lipids include the Kreis test, the anisidine value, and the thiobarbituric acid number.

Data on the total volatile nitrogen and free fatty acid values of individual chicken legs removed from storage in a domestic refrigerator are shown in Table 7. The literature available on the normal values for these parameters is inadequate and often data have to be derived from small surveys undertaken when data are needed.

Differentiation of Fresh and Frozen-Thawed Meat

Legislation prohibits the sale of previously frozen meat as 'fresh' and thawing and refreezing are improper treatments of meat. Such treatment can enhance the possibility of the increase of the populations of microorganisms, such as *Salmonella*, and thus pose a serious health risk. Furthermore, the consumer may be disadvantaged because fresh meat normally commands a higher price than frozen meat and it is therefore important that it can be differentiated. The usual means of determining whether meat has been previously frozen and thawed is from the use of the β-hydroxyl-coenzyme A dehydrogenase (β-HADH) enzyme test. This test involves measuring the activity of mitochondrial enzymes that are released when the meat is frozen and thawed. The freezing and thawing of meat damages the muscle mitochondria, which results in the partial release of the mitochondrial enzymes into the sarcoplasm or muscle juices. The enzyme that has most commonly been used is β-HADH. The biochemical method was developed by Gottesmann and Hamm in meat press juice by means of a relatively simple spectrophotometric enzyme assay and has been the subject of research and validation. The activity of β-HADH is measured by determining the rate at which nicotinamide-adenine-dinucleotide (reduced NADH) is converted to nicotinamide-adenine-dinucleotide (NAD^+). This conversion is dependent on the level of β-HADH activity and can be measured by the decrease of absorption of the reaction solution at 340 nm. Work carried out on chicken indicates that the test will give an indication that it has been frozen where the meat has been stored below − 12 °C. The method has also been collaboratively tested.

See also: **Fats**: Occurrence; **Immunoassays**: Radioimmunoassay and Enzyme Immunoassay; **Meat**: Sources; **Protein**: Chemistry; Food Sources; Determination and Characterization; **Spoilage**: Bacterial Spoilage

Further Reading

Allen JC and Hamilton RJ (1994) *Rancidity in Foods*. 3rd edition.

Billington M, Bowie H, Scotter S, Walker H and Wood R (1992) The differentiation of frozen and frozen-thawed poultry meat by the determination of the beta – hydroxyacyl-CoA-Dehydrogenase (HADH) activity of chicken breast press juice: collaborative trial. *JAPA* 28: 103–116.

BS 4401 Part 1: (1998) ISO 936:1998, Methods of Test for Meat and Meat Products, Determination of Ash content.

BS 4401 Part 2: (1980) ISO 937:1978, Methods of Test for Meat and Meat Products. Part 2 Determination of Nitrogen content (reference method).

BS 4401 Part 3: (1997) ISO 1442, Methods of test for Meat and Meat Products, Determination of Moisture Content (reference method).

BS 4401 Part 4: (1970) Methods of Test for Meat and Meat Products. Part 4 Determination of Total Fat Content.

BS 4401 Part 5: (1996) Methods of Test for Meat and Meat Products, Determination of Free Fat Content.

BS 4401 Part 11: (1995) ISO 3496:1994, Methods of test for Meat and Meat Products Determination of Hydroxyproline content.

Davies AR (1994) Spoilage of raw meat. *PHLS Microbiology Digest* v11(2): 108–110.

Eburne RC and Prentice G (1994) Modified-atmosphere-packed ready-to-cook and ready-to-eat meat products.

In: Man CMD and Jones AA (eds) *Shelf life evaluation of foods*. Book Number 5571. Shelf F1. pp. 156–178.
Egan H, Kirk RS and Sawyer R (1981) *Pearson's Chemical Analysis of Foods* 8th edition.
Food Analysis Performance Assessment Scheme (FAPAS) Sand Hutton, York YO41 1LZ UK: Central Science Laboratory.
Gottesman P and Hamm R (1983) *Fleischwirtsch* 63(2): 219–221.
Her Majesty's Customs and Excise Integrated Tariff of the United Kingdom (2002).
Huis in't Veld JHJ (1996) *International Journal of Food Microbiology* 32(3).
Jackson TC, Acuff GR and Dickson JS (1997) Meat, poultry, and seafood. In: Doyle MP, Beuchat LR and Montville TJ (eds) *Food Microbiology. Fundamentals and Frontiers*. ASM Press, Washington DC.
Lord and Swan (1984) *Journal of the Association of Public Analysts* 22: 131–140.
McLean BD (1999) CCFRA Guideline NO. 22 Meat and Meat Products: The Calculation of meat content, added water and connective tissue from analytical data 1999.
Stubbs G (1919) The estimation of the approximate quantity of meat in sausages and meat pastes (1919) Geo. Stubbs, OBE., FIC & A. More. ARCS, FIC. *The Analyst* 44: 125.

Nutritional Value

J Higgs and J Pratt, Meat & Livestock Commission, Milton Keynes, UK

Background

This article considers the different animal species available for human consumption and the significance of the nutritional value of meat as a whole. Increased awareness of animals as sentient beings, and the option to elect to avoid eating meat in affluent societies, influences discussion of the role of meat in the modern human diet.

In theory, it is possible to construct both omnivorous and vegetarian diets that meet dietary recommendations, although allowing for the differences in bioavailability of some micronutrients and practical issues relevant to modern society (preparation time and cooking); this task is easier for an omnivorous diet. It is unknown whether or not there would be any significant differences between two such diets that would influence public health advice.

Of all foods, meat provides one of the widest ranges of nutrients in useful amounts. This may be one of the reasons for its importance throughout history as the central focus of the diet, included particularly within celebratory meals. It is perhaps the fact that meat has been eaten as much for enjoyment as for its nutritional qualities that allows any offending aspect of meat production to affect its consumption by affluent nations.

Types of Meat

The types of meat commonly consumed in different countries are dependent on eating habits and the ability to rear the animals successfully, which is influenced by local climate, geography, and economy. Sheep and goat meat are more popular in developing countries, while native llama, buffalo, and antelope are important parts of the local ecosystems in many areas, especially Bolivia, Peru, Ecuador, Asia, and Africa.

Beef, lamb, pork, and chicken are the major meats consumed in Western societies, with beef predominating. Although the total meat consumption in the UK has remained fairly steady since the 1980s, consumption of chicken, pork, and meat products has increased, while that of beef and lamb has declined. Major influences include increasing demand for convenience, a trend towards eating out of the home, and a superior health image for poultry during this period. Veal, horse meat, hare, rabbit, and game birds such as duck and pheasant also contribute to meat intake. With the establishment of farmed deer in recent years, venison has become more popular and affordable, promoted for its low fat content. Commercial rearing of ostrich meat is new to Europe and, like emu, has grown rapidly in the USA. Both ostrich and emu meat are promoted for their favorable fat profile, which is similar to that of poultry and pork, although the iron content of ostrich is closer to that of beef and lamb.

Meat includes carcass meat and edible offals (organ meats, e.g., liver, kidney, tongue, heart, sweetbreads (thymus and pancreas), and tripe (stomach)). Meat products is a term that includes the result of any type of processing ranging from composite dishes such as pies and ready-made meals to the addition of other ingredients such as water to increase succulence, curing, fermentation, and preparations such as brawn, black pudding, meat bars (pemmican), etc.

Consumption

Meat consumption increases with affluence. Japanese intake has increased to half that in the USA, and in China, meat intake is also increasing dramatically (**Table 1**).

Table 1 Meat consumption 1980s and 1990s (g per person per day)

Country	*1980s*	*1990s*
USA	310	231
Australia	296	290
UK	201	204
France	290	305
Germany	269	261
Japan	100	123
China	26	108

The values quoted are from the OECD for the years 1982–84 and 1992, respectively. The exception is China; these values are taken from *Asia Pacific Food Industry* (1995) 7(11): 14, using the years 1979 and 1994. The figures include bone weight.
Reproduced from Meat, Poultry and Meat Products: Nutritional Value, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Macronutrient Content

Protein

Meat is a significant source of protein, which is of high biological value. Within developed nations, meat provides around 60% of protein intake compared with only 15% in poorer nations. The protein content of different meats and cuts varies inversely with fat content. Meat is a rich source of the amino acid taurine; this is essential in newborn infants as they are unable to synthesize this amino acid from cysteine in sufficient quantities. The significance of lower levels of taurine in the breast milk of vegan mothers is unknown.

Fat

The total fat content varies with species and across the carcass. Although the fat content of lean muscle from most species is around 2–5%, the total fat content depends on rearing practices, feeding regimes, and age. The level of fat trim sold with the meat provides opportunities to affect the fat content as-consumed further, through cooking and trimming by the end user. This is illustrated by looking at the changes seen in the fat content of pork since the 1970s (Figure 1). Trimming visible fat (subcutaneous and intermuscular) has the most influence on the final fat content, since the invisible (intramuscular) fat makes a relatively small contribution to the total. Table 2 provides UK data showing the range of trimmable fat on specific red meat cuts.

Fat is lost on cooking meat, assuming that it is allowed to escape, for example. The amount lost depends on the initial fat content, and fully trimmed lean meat loses little fat. The concurrent loss of moisture on cooking meat obscures fat losses, as viewed in composition tables.

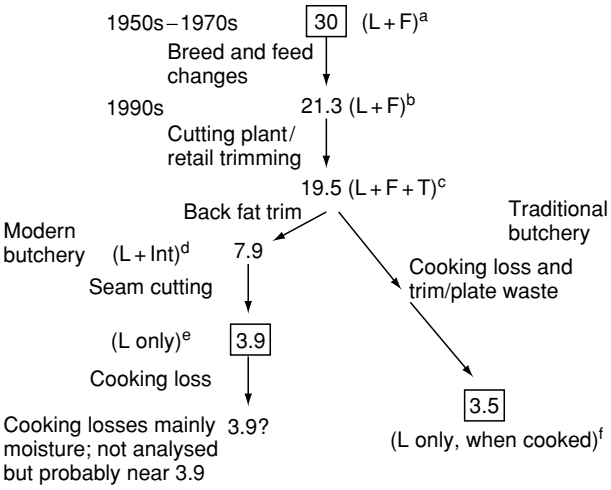

Figure 1 Change in fat content of pork loin for 100 g of raw edible tissue. Key: (L + F), lean + fat, untrimmed; (L + F + T), lean + fat, some trimming; (L + Int), lean + intermuscular fat; (L only), lean only. Source of data: Untrimmed, UK Database 1970s; [b]untrimmed including rind, UK Database 1990s; [c]at retail, UK Database 1990s (79% lean, 21% fat); [d]trimmed of all back fat, UK Database 1990s; [e]fully trimmed, UK Database 1990s; [f]100 g raw as-purchased, cooked, and fully trimmed. [a]From Holland B, Welsh AA, Unwin ID, Buss DH, Paul AA and Southgate D (1991) *McCance & Widdowson's, The Composition of Food*. Royal Society of Chemistry, with permission. 5th edn. Campridge: From Analytical Methods Committee (1991) Nitrogen Factors for Pork: A Re-assessment. *Analyst* **116**: 761–766, with permission. [f]From Chan W, Brown J, Lee FM and Buss DH (1995), Meat, Poultry and Game. Supplement to *McCance & Widdowson's, The Composition of Food*. Cambridge: Royal Society of Chemistry/London: Ministry of Agriculture, Fisheries and Food, with permission. Reproduced from Meat, Poultry and Meat Products: Nutritional Value, *Encyclopedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 2 Proportion of trimmable fat in some raw cuts of beef, lamb and pork

Meat	*Means and ranges (%)*
Beef	
Braising steak	7 (0–18)
Fore-rib	21 (10–35)
Sirloin steak	18 (6–34)
Lamb	
Chump chops	21 (11–32)
Breast	33 (20–48)
Rack of lamb	22 (18–25)
Pork	
Leg	18 (12–28)
Loin chops	25 (16–34)
Steaks	11 (3–35)

From Chan W, Brown J, Lee SM and Buss DH (1995) Meat, poultry and game. Supplement to *McCance and Widdowson's The Composition of Foods*. Cambridge: Royal Society of Chemistry and the Controller of Her Majesty's Stationery Office.

During the years following World War II, the emphasis for meat production was on maximizing its energy (fat) content and minimizing cost in an attempt to produce enough nutrient-dense food to improve the nutritional well-being of the whole population. Thus, in the 1970s, red meat and meat products contributed 27.4% to the total daily UK fat intake, and little of this was trimmed away. In the early 1980s, diet and health reports identified meat as a major contributor to fat intake, and so started the negative health association with red meat. Poultry meat was recognized as being significantly lower in fat and was recommended in preference by health experts. The red meat industry responded by reducing the fat content of red meat once again. Selective breeding and feeding practices, carcass classification changes to favor leaner production, as well as butchery techniques (seaming out whole muscles and trimming away all intermuscular fat) have meant that the meat on sale in the 1990s is much leaner than it was in the 1970s. The fat content of the carcass has been reduced in the UK by over 30% for pork (making British pork virtually the leanest in the world), by 15% for beef, and by 10% for lamb. Further reductions are anticipated for beef and lamb in the future. Overall, little change has occurred in poultry meat; most of the fat is in the skin, which is 48% fat.

Minced meat, predominantly beef, has become more popular in recent years, reflecting a move to convenience and a more cosmopolitan approach to meals. Although viewed as a fattier product, fat levels have dropped to typically 15%, with some extra-lean mince as low as 5%.

Since food-composition tables throughout the world have not kept pace with these changes, the health image of red meat has suffered disproportionately to the nutritional benefits it offers. Food-composition tables from different countries (e.g., the UK, USA, and Germany) vary due to the analysis of different cuts, different consumer demands, and how upto date the tables are. Data on the composition of red meat are shown in Table 3, and a 1995 fat audit for the UK to trace all fat in the human food chain provides useful information to illustrate the reduction in fat intake since 1982 from meat, compared with other foods (Table 4).

The fat content of meat products can vary considerably, depending on the proportion of lean and fat from meat as well as the other ingredients. Traditional types such as sausages, pastry-covered pies, and salami were high in fat, up to 50%, but modern products include ready-made meals and prepared meats, which can be low in fat (5%). The trend downwards in fat of red meat is reflected in the reduced fat content of a number of meat products, such as hams and sausages. Reduced fat versions of popular hamburgers and sausages continue to enter the market, and utilization of fat replacers will enable this trend to continue (Table 5).

Fatty Acids

As with other foods, meat contains a mixture of fatty acids. As the fat content of red meat has reduced, so the fatty acid mix has changed with the percentage of saturated fat than in the past. Poultry fat is 30% saturated, while pork, beef, and lamb are less than 50% saturated. The contribution made by meat to dietary saturated fat is given for the UK (Table 6). The principal saturated fatty acids are palmitic acid (C16:0 and stearic acid C18:0, which are thought not to raise blood cholesterol levels. There are only minor amounts of myristic acid, the most atherogenic.

Meat is one of the main contributors to the monounsaturated fat content of the British diet, principally oleic acid (C18:1) from red meat. Around 40% of the fat in meat is monounsaturated. Ruminant meats also contain *trans* fatty acids, contributing around 18% of British intakes.

Recent international analyses of beef, lamb, and pork demonstrate useful levels of long-chain polyunsaturated fatty acids in pasture- and grain-fed animals, specifically eicosatrienoic acid (20:3*n*-6), arachidonic acid (20:4*n*-6), eicosapentaenoic acid (20:5*n*-3), and docosapentaenoic acid (22:5*n*-3). Significant losses of some fatty acids, particularly the more highly unsaturated fatty acids (C20 and C22) that are more susceptible to oxidation, occur in cooking. The small amount of docosahexaenoic acid (22:6*n*-3) present in raw samples is lost on cooking. The richest dietary sources of 20:4*n*-6 include liver, kidney, and turkey, while smaller quantities are found in beef, lamb, pork, and chicken (Table 7). Grass-fed animals have higher concentrations of α-linolenic acid compared with grain-fed animals, and beef and lamb have favorable *n*-6:*n*-3 ratios. Meat, along with fish, provides the only significant dietary sources of C20 and C22 *n*-3 fatty acids, and for countries like the UK, where fish intakes remain low, the contribution to the diet from meat is significant, despite absolute levels in meat being low relative to those in fish (Table 8).

It is possible to manipulate the fatty acid profile of meat to satisfy human nutritional concerns more easily in monogastrics (pigs and poultry) than in ruminants (cattle and sheep), where polyunsaturated fatty acids are hydrogenated in the rumen. The fatty acid profiles of meats will continue to change as knowledge of their significance to human health improves, such that food composition data for meat can only represent a single point in time.

Table 3 Nutrient composition of different meats per 100 g

Nutrient	*Beef (raw, with fat)*	*Bison (raw)*	*Buffalo (raw)*	*Chicken (raw, skin/fat)*	*Duck (raw, with fat)*	*Goat (raw)*	*Kangaroo (raw)*	*Lamb (raw, with fat)*	*Ostrich meat (cooked)*	*Pork (raw, with fat)*	*Rabbit (raw meat)*	*Turkey (raw, skin/fat)*	*Veal (raw)*	*Venison (raw)*
Energy (kJ)	945	456	610	840	1622	673	493	1024	548	924	573	556	523	431
Protein (g)	21.7	21.6	20.8	19.1	13.1	19.5	24.2	18.5	24.8	19.2	21.9	21.6	21.5	22.2
Fat (g)	15.5	1.8	7.1	13.8	37.3	7.9	2.3	19	2.8	16.1	5.5	5.2	4.4	1.6
SFA (g)	7.1	0.7	3.24	3.8	2			9.35	1.06	5.8	2.1	1.7	1.8	0.8
MFA (g)	7	0.7	3.13	6.4	3.2			7.23	0.89	6.6	1.3	1.9	1.9	0.4
PFA (g)	0.6	0.2	0.3	2.6	1			0.87	0.59	2.7	1.8	1.2	0.4	0.4
Cholesterol (mg)	61	62		100	110			78	81	65	53	78	57	50
Iron (mg)	1.56	2.6	2.7	0.7	1.3	1.95	6	1.23	2.98	0.63	1	0.6	0.7	3.3
Zinc (mg)	3.5	2.8		1.1	1		4.1	2.71		1.75	1.4	1.8	2.7	2.4
Calcium (mg)		6	7.5	7	8	9.5	3		1.69		22	6	6.5	5
Magnesium (mg)				22	14						25	23	tr	25
Copper (mg)	0.02	0.09		0.05						0.05	0.06	0.03	8	0.21
Selenium (μg)	5.93			12				0.07		11.6		12	8	
Sodium (mg)	55	54	111	70	73		63	62	71	59	67	70	71	55
Vitamins														
B_1 (mg)	0.09		0.06	0.11	0.14	0.15		0.08		0.8	0.1	0.07	0.09	0.16
B_2 (mg)	0.19		0.19	0.25	0.51	0.28		0.18		0.21	0.19	0.2	0.19	0.7
Niacin equivalents (mg)	8.24		6.3	10	5.9	4.9		7.8		9.5	12.5	11.5	9.8	6.7
B_6 (mg)	0.46		0.44	0.3	0.33	0.3		0.25		0.45	0.5	0.55	0.57	0.65
B_{12} (μg)	1.61		1.28	tr	2			1.76		1	10	1	2	1
Folate (μg)				9.0	7			5.51		2.76	5		19	6
D (μg)	0.39			0.6				0.42		0.68		0.4	1.2	N
E (mg)	0.12			0.15				0.1		0.04	0.13	0.01	0.22	N

MFA, monounsaturated fatty acids; N, nutrient present but no reliable data on amount; PFA, polyunsaturated fatty acids; SFA, saturated fatty acids; tr, trace.
Data taken from relevant international food composition databases published between 1986 and 1995.

Micronutrient Content

The micronutrient content of meat can make a significant positive contribution to nutritional intakes, since it is a relatively concentrated source of a large number of vitamins, minerals, and trace elements, and the bioavailability of several such as iron, copper, and zinc is greater from meat than from plant foods.

Iron

Meat, especially red meat and offal, is a rich source of iron, a 100-g serving of beef rump steak providing 3 mg. Red meat contains 50–60% of its iron in the heme form (from hemoglobin and myoglobin), and this is absorbed by a more efficient mechanism than nonheme iron, the sole source of iron in plant foods. Moreover, heme iron is unaffected by the numerous inhibitors of iron absorption such as phytate. Thus, the absorption of iron from meat is typically 15–25%, twice that from plant sources. Meat also plays a valuable role as an enhancer of iron absorption from plant foods, so the presence of meat in a meal can double the amount of iron absorbed from the other components of the meal. The exact mechanism for this, although not confirmed, is believed to be due to the iron-binding capacity of cysteine within peptides following proteolysis of meat muscle. Although the absolute iron content of white meats, e.g., chicken, and meat products may appear low compared with some plant food sources, this enhancing function and better iron availability make these foods useful contributors to iron intakes. Cooking in iron or steel utensils can increase the iron content of the meal owing to this enhancing effect of meat. Meat is thus a particularly important influence on iron intake and iron status. Serum ferritin levels, which indicate iron stores in the body, are strongly correlated with heme iron intake. Meat can have a major impact in groups vulnerable to iron deficiency, specifically toddlers, adolescents, and women of child-bearing age.

Table 4 Total UK dietary fat[a], per person, contributed from all foods groups in 1982 and 1992

Food group	*Dietary fat (kg per person per year)*	
	1982	*1992*
Beef and veal	2.09	1.93
Lamb	0.66	0.51
Pork and products (1981, 1992)	6.54	4.36
Poultry meat	9.29	6.80
Fats and oils	17.40	21.98
Dairy fat (1983, 1992)	12.91	11.04
Fish and fish oil	4.40	2.20
Eggs	1.56	1.09
Chocolate	1.54	1.92
Cereals	0.92	1.05
Nuts	0.27	0.34

[a]Contribution of fat from primary food source rather than the end product, e.g., 'Fats and oils' includes fat used ultimately in such products as bakery goods, meat products, and confectionery.

Zinc

The best sources of zinc are meat, especially beef, lamb, pork, poultry, and seafood. As with nonheme iron, the bioavailability of zinc is reduced by inhibitors in the diet such as phytate and oxalate, and is enhanced when consumed with animal protein. In metabolic studies, zinc absorption and retention are greater on diets with a high meat content compared with those on a low meat content. Lower plasma zinc levels in vegetarians and vegans, despite higher intakes, suggest that meat has a major influence on zinc status. Approximately 20–40% of zinc is absorbed from meat, which is the major contributor to zinc intakes in Western countries (Table 6).

Table 5 Nutrient content for a range[a] of meat products per 100 g

Product	*Fat content (g; range)*	*Energy (kJ; range)*	*Energy (kcal; range)*	*Iron (mg; range)*
Bacon, all types	11.7–26.6	727–1388	174–332	0.4–1.1
Beefburgers, raw	6.2–25.4	472–1258	113–301	1.7–2.7
Chicken nuggets/fingers	1.5–17.1	552–1212	132–290	0.6–1.1
Cornish pasties	16.3–24.3	1066–1496	255–358	1.0–1.1
Frankfurters, cooked	20.0–31.0	1041–1400	249–335	1.1
Ham, all types	1.5–20.0	372–1003	89–240	0.6–0.8
Pastrami	3.2–7.2	477–736	114–176	
Pate	11.9–33.0	811–1484	194–355	1.5–7.4
Pork pies	19.5–33.5	1329–1827	318–437	0.9–1.2
Pork sausages, raw	8.8–30.4	740–1446	177–346	0.8–1.0
Salami	28.3–52.3	1442–2249	345–538	1.3

[a]Where available, ranges include low-fat versions of the product.

Data taken from: MLC (1996) *Meat Products Composition Survey*; Chan W, Brown J, Church SM and Buss DH (1996) Meat products and meat dishes. (Available on request from MLC). Supplement to *McCance and Widdowson's The Composition of Foods*. Cambridge: Royal Society of Chemistry and the Controller of Her Majesty's Stationery Office.

Table 6 Contribution made by meat to nutrient intake

Nutrient	*Percentage of total daily intake contributed by meat for UK*
Energy	12.9
Protein	38.0
Fat	20.0
Saturated fatty acids	19.2
Monounsaturated fatty acids	23.8
Polyunsaturated fatty acids	15.0
Iron	15.0
Zinc	33.3
Retinol	32.2
Vitamin D	18.5
B_1	34.3
B_2	18.6
Niacin equivalents	42.9
B_6	25.0
B_{12}	37.5
Folate	5.4
Vitamin E	1.2
Vitamin C	0.4

Per capita consumption figures for beef and veal, mutton and lamb, pork, bacon and poultry for 1995 in the UK.
Reproduced from Meat, Poultry and Meat Products: Nutritional Value, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 7 Fatty acid composition of muscle from English beef, lamb, and pork

Fatty acid	*Percentage by weight of total fatty acids*		
	Beef	*Lamb*	*Pork*
12:0 lauric	0.08 ± 0.03	0.31 ± 0.18	0.12 ± 0.05
14:0 myristic	2.66 ± 0.54	3.30 ± 1.07	1.33 ± 0.20
16:0 palmitic	25.0 ± 1.77	22.2 ± 1.56	23.2 ± 1.46
16:1 *cis*	4.54 ± 0.81	2.20 ± 0.26	2.71 ± 0.45
18:0 stearic	13.4 ± 1.84	18.1 ± 2.80	12.2 ± 1.11
18:1 *trans*	2.75 ± 1.28	4.67 ± 1.67	nd
18:1*n*-9 oleic	36.1 ± 2.87	32.5 ± 3.25	32.8 ± 3.91
18:1*n*-7 vaccenic	2.33 ± 0.40	1.45 ± 0.26	3.99 ± 0.59
18:2*n*-6 linoleic	2.42 ± 0.63	2.70 ± 0.86	14.2 ± 4.09
18:3*n*-6 γ-linolenic	nd	nd	0.06 ± 0.03
18:3*n*-3 α-linolenic	0.70 ± 0.18	1.37 ± 0.48	0.95 ± 0.33
20:2*n*-6	nd	nd	0.42 ± 0.11
20:3*n*-6	0.21 ± 0.06	0.05 ± 0.04	0.34 ± 0.09
20:3*n*-3	0.007 ± 0.011	nd	0.12 ± 0.05
20:4*n*-6 arachidonic	0.63 ± 0.21	0.64 ± 0.23	2.21 ± 0.73
20:4*n*-3	0.08 ± 0.03	nd	0.009 ± 0.022
20:5*n*-3 EPA	0.28 ± 0.11	0.45 ± 0.13	0.31 ± 0.15
22:4*n*-6	0.04 ± 0.02	nd	0.23 ± 0.07
22:5*n*-3	0.45 ± 0.14	0.52 ± 0.14	0.62 ± 0.20
22:6*n*-3 DHA	0.05 ± 0.02	0.15 ± 0.05	0.39 ± 0.23

DHA, docosahexaenoic acid; EPA, eicosapentaenoic acid; nd, not detected.
From Enser M, Hallett K, Hewitt B, Fursey G and Wood J (1996) Fatty acid content and composition of English beef, lamb and pork at retail. *Meat Science* 42: 443–456.

Table 8 Contribution to UK diet of *n-3* and *n-6* polyunsaturated fatty acids from meat and fish (g per person per year)

Food	n-*3*	n-*6*
Beef	17.7	39.1
Lamb	13.0	17.5
Pork	55.7	451.6
Total – red meat	86.3	508.2
Total – fish	12.9	1.5

From Ulbricht (1995) *Fat in the Food Chain*. A report to the Ministry of Agriculture, Fisheries and Food. Available from MAFF library.

Selenium

Selenium is added to animal feeds to prevent deficiency. Concerns of excess build-up of selenium in soil and water supplies as a consequence of this are difficult to substantiate. The bioavailability of selenium from plant foods was thought to be greater than that from animal foods, but recent data demonstrate that meat, both raw and cooked, provides selenium in a form which is as bioavailable as that in cereals. Meat provides on average 10 μg of selenium per 100 g and contributes around a quarter of daily needs.

Vitamins

Meat can make a significant contribution to intakes of all B vitamins except folate and biotin. Pork and pork products, such as bacon and ham, are particularly rich sources of thiamin, typical servings providing the daily requirement. Liver and kidney provide more than the daily requirement of riboflavin in less than 100 g. Meat also provides the richest source of niacin (half of which is derived from the amino acid tryptophan) and vitamin B_6. Per 100 g portion, veal liver provides half the daily B_6 needs, and other meats provide around a third. Organ meats are particularly rich sources of folate and vitamin A (Table 9). Currently in the UK, liver is not recommended during pregnancy, despite being an excellent provider of iron and folate, owing to the very high levels of retinol it contains, which may potentially cause malformation of the developing fetus. Vitamin A in its active form, retinol, is found only in foods of animal origin, so meat provides a preformed source of this vitamin.

Vitamin B_{12} merits particular attention, since foods of animal origin provide the only significant dietary source. It is recommended that where meat and dairy products are avoided, supplements should be taken. In the past, some B_{12} was provided from the soil on poorly cleaned foods, which may in part explain the apparent absence of deficiency in some vegan groups.

Pantothenic acid is universal in all living matter, and rich sources include liver and kidney. Most of the vitamin is leached into the drip loss associated with frozen meat.

Table 9 Nutrient content of a range of offal meats

	Energy (kJ)	Fat (g)	Iron (mg)	Zinc (mg)	Vitamin A (μg)	Vitamin D (μg)	Vitamin B_{12} (μg)	Cholesterol (mg)	Folate (μg)
Heart									
Lamb, raw	500	5.6	3.6	2.0	tr	N	8.0	140	2
Kidney									
Lamb, raw	382	2.6	5.5	2.5	96	N	17.0	315	8
Liver									
Calf, raw	437	3.4	11.5	14.2	18 800	0.3	68	370	155
Lamb, raw	575	6.2	7.5	4.0	17 300	0.5	54	430	205
Pig, raw	475	3.1	13.9	7.0	17 400	1.1	23	260	295
Brains									
Lamb, boiled	529	8.8	1.4	1.4	tr	tr	8	2 200	6
Tongue									
Ox, stewed	1021	18.3	2.5	3.8	tr	N	4.0	N	11

N, nutrient present in significant quantities but no reliable data on amount; tr, trace.
From Chan W, Brown J, Lee SM and Buss DH (1995) Meat, poultry and game. Supplement to *McCance and Widdowson's The Composition of Foods*. Cambridge: Royal Society of Chemistry/London: Ministry of Agriculture, Fisheries and Food.

Vitamin D

Meat, apart from liver, was thought to be a poor source of vitamin D. New analytical data for meat and liver include significant amounts of 25-hydroxycholecalciferol, assumed to have a biological activity five times that of cholecalciferol. Previous work on the British diet and nutrition surveys using updated food composition data of adults and young children has shown that, rather than contributing as little as 4% of dietary vitamin D intakes, meat and meat products appear to provide 21 and 18%, respectively, becoming the major contributor of natural dietary vitamin D intakes.

Role of Meat in the Diet

Meat has played a significant role in human evolution. Early herbivorous hominids did not survive, whereas those that developed the ability to hunt became our ancestors (*Homo erectus*).

Anthropological research shows that the length of the gut in primates and humans became shorter with the introduction of animal-derived food. Smaller quantities of food of high digestibility required relatively smaller guts, characterized by simple stomachs and proportionately long small intestines, emphasizing absorption. Furthermore, it has recently been suggested that selection for relatively large brains in humans over the last 2 million years could not have been achieved without a move to a high-quality diet, based on animal products, since the energy saved with a shorter gut allowed an increase in brain mass. It is believed that better nutrition, specifically increased animal protein and fat intake together with improved public health, has allowed us to move towards attaining our genetic optimum.

Meat has traditionally been considered an essential component of the diet to ensure optimal growth and development. With a limited range of foods available in primitive societies throughout history, meat was important as a concentrated source of a wide range of nutrients. With the range and abundance of foods now available to developed societies, the nutritional significance of any one food is reduced. Where meat is excluded, the missing nutrients can be supplied from a combination of other foods, and within traditional vegetarian cultures, this appears at least adequate. Concern arises when meat is excluded with insufficient attention given to selecting appropriate combinations of foods to ensure that nutrients are supplied in adequate quantities. There is no doubt that in modern Western societies where less time is devoted to planning and preparing meals, and an increasing proportion of daily food intake is consumed outside the home as snacks and quick meals, the inclusion of meat in the diet minimizes the risk of micronutrient deficiencies, especially iron, albeit subclinical.

Meat and Health: Addressing the Concerns

Traditionally, meat intake was associated with improved health, but as a consequence of its association with dietary fat, this role has been marginalized. To assess the role played by meat, numerous studies have compared the health status and mortality of vegetarians with those of omnivores. Simplistic conclusions from such work can be flawed, since interpretation of the results must take into account the many other influential differences between such groups. In particular, lifestyle factors must be

considered: vegetarians traditionally smoke less, consume less alcohol, tea, and coffee, and, being generally more health conscious, tend to exercise more. Typically, the meat-eating groups tend to be dominated by sedentary, physically inactive subjects with little interest in diet. Since the 1950s, omnivorous diets have been relatively high in animal protein and fat with insufficient dietary fiber, fruit and vegetables, such that meat intake may have acted as a marker for a generally 'unhealthy' diet.

Regular consumption of red meat is associated, epidemiologically, with increased risk of coronary heart disease (CHD), owing to its fat composition. There is, however, a growing bank of evidence that a healthy diet which includes lean red meat can produce positive changes in lipid biochemistry. Blood cholesterol levels are increased by inclusion of beef fat, not by lean beef in an otherwise low-fat diet. Equal amounts of lean beef, chicken, and fish added to low-fat, low-saturated-fat diets similarly reduce plasma cholesterol and low-density lipoprotein cholesterol levels in hypercholesterolemic and normocholesterolemic men and women.

The association between the arachidonic acid content of meat and increased thrombotic tendencies has recently been challenged. The presence of large amounts of linoleic acid in current diets results in plasma increases of linoleic and arachidonic acids only, but in the absence of linoleic acid, the long-chain *n*-6 and *n*-3 polyunsaturated fatty acids present in lean meat can influence the plasma pool, increasing plasma 20:3*n*-6, 20:4*n*-6, and 20:5*n*-3 fatty acids and probably reducing thrombotic tendencies.

Meat has been associated with various cancers, particularly colon cancer. However, meat-eating is strongly inversely correlated with gastric cancer, which makes any public-health recommendations difficult to determine, based on meat-eating alone. Epidemiologically it is difficult to distinguish the influence of animal fat, protein, and meat.

There does appear to be an independent role for protein as a source of cooked food mutagens. Heterocyclic amines produced on overcooking meat have been shown to be carcinogenic in rats, although at normal cooking temperatures and average cooking times, the level of these compounds produced is not excessive, compared with the amounts required for carcinogenicity.

Nitrates are added to meat products to prevent microbial proliferation. However, the levels used now are much reduced such that the nitrosocompounds (NOC) produced from these additives are present at very low levels in meat products. Contrary to public perception, their carcinogenic potential is likely to be minor. In humans, high-protein (meat) diets increase fecal NOC levels, although the significance of this for normal healthy diets remains unclear.

Table 10 Paradoxical associations between meat and health

- Generally, in countries where meat consumption is high, intake of protective fruit and vegetables tends to be low, and vice versa.
- There is a lower risk of colon cancer among South Asian immigrants to the UK, than in the general population, and this is equally true of vegetarian and meat-eating Asians.
- Mormons who abstain from alcohol, tea, coffee, and smoking but consume meat have death rates from cancer and CHD similar to those of vegetarian groups, indicating that meat is not the major determinant. Blood pressure is reduced on switching from a meat-based diet to a vegetarian diet, but adding meat to a vegetarian diet is not associated with an increase in blood pressure, suggesting that the potassium-to-sodium ratio is more influential than the presence of meat *per se*.
- Consumption of meat has almost doubled in Japan since the 1970s, yet CHD morbidity has reduced from an already low level. Meat consumption in the USA remains higher than in the UK, yet, unlike the UK, morbidity from CHD has been falling steadily since the 1970s.

Daily consumption of green and yellow vegetables with meat is associated with a reduced risk of cancer at many sites, whereas daily meat consumption with less frequent vegetable consumption has been associated with increased risk. The significance of meat's role as a risk factor in cancer causation remains the subject of much debate, and results of more informative long-term cohort studies are required to elucidate the full picture.

There are many paradoxes concerning the apparent associations between meat consumption and both CHD and cancer. Some of these are listed in Table 10. They serve to demonstrate that the diet and health story, at least as far as meat is concerned, is both complex and incomplete. Two reasons for this are, first, that meat is but one influential aspect of the diet. Its effects are modified by other foods (e.g., fruit and vegetables) such that the net effect may be quite different. Second, the changing lipid composition of meat, in response to increased knowledge of lipids in relation to health, will continue to alter the role played by meat in this context. Fortunately, since most of the valuable nutrients in meat are within the lean component, reducing the visible fat has little bearing on its contribution to micronutrient status (Table 11).

Hygiene and Safety Aspects

The wholesomeness of meat and meat products is dependent on a variety of factors, some of which act in the period prior to arrival of the animal or bird at the slaughterhouse. The health status of the herd or

Table 11 Nutrient variability across the carcass – values given for lowest and highest fat cuts for different species

Cut of meat	*Nutrients per 100 g weight*				
	Energy (kJ)	*Fat (g)*	*Vitamin D (μg)*	*Iron (mg)*	*Zinc (mg)*
Veal escalopes, raw	445	1.7	1.3	0.6	2.4
Beef flank, lean raw	735	9.3	0.5	1.8	3.6
Beef, stewing steak, lean, raw	512	3.5	0.8	2.1	5.7
Chicken, light meat, raw	445	4.4	0.2	0.5	0.7
Chicken, leg quarter with skin, raw	811	13.3	0.5	0.8	1.4
Lamb, best end neck cutlets, 34% fat, raw	1327	27.9	0.4	1.2	3.5
Lamb leg, lean, raw	617	4.7		1.5	3.1
Pork, leg joint lean, raw	449	2.2	0.5	0.8	2.2
Pork, belly joint/slices, raw	1084	20.0	0.8	0.6	1.7

From Chan W, Brown J, Lee SM and Buss DH (1995) Meat, poultry and game. Supplement to *McCance and Widdowson's The Composition of Foods*. Cambridge: Royal Society of Chemistry/London: Ministry of Agriculture, Fisheries and Foods.

flock is in turn influenced by the method of husbandry practiced on the farm, the health prevention measures applied, the conscientiousness of monitoring, and the level of prompt and necessary action taken, e.g., the isolation and treatment of veal calves infected with one or more species of salmonellae. However, calves may be carriers of infections such as *Escherichia coli* and *Listeria monocytogenes*, and display no symptoms of disease but may prove a hazard to public health when meat and offal are prepared from the carcasses. Similarly, campylobacteria may be present in a mild form in a batch of broilers which appear perfectly healthy in life but may prove a hazard to the consumer. Close monitoring of farm health records and scrutiny by the herd/flock veterinarian and collaboration with the supervising official veterinary surgeon in the slaughterhouse and poultry plant will help prevent subsequent potential consumer health hazards.

The raising of stock in dirty conditions allows the hides, fleeces and feathers to contaminate the environment during the slaughter and dressing of carcasses. Cross-contamination of the carcasses from feces and soil in addition to the soiling of plant, equipment, and the protective clothing of slaughterhouse operatives can markedly reduce the general hygiene in the meat plant with levels of bacteria, bacterial spores, parasites, viruses, and fungi exceeding acceptable levels. *Staphylococcus aureus* is a normal commensal of skin and hair, and can produce toxins in food. A highly successful recent campaign has highlighted the necessity for only clean livestock to be presented for slaughter.

Transport of animals, especially loading, unloading, and lairaging, can lead to stress conditions, particularly if groups are mixed. In some countries, two-tier lorries are used to transport cattle, and three-tier sheep transporters are common worldwide, resulting in gross contamination with urine and feces. Again, stress from loading and unloading can occur. Stressed animals, particularly carriers of disease, are known to demonstrate clinical disease and excrete pathogenic organisms. Welfare and hygiene regulations governing animals during transport minimize such occurrences.

The slaughterhouse itself may not be free from external vectors of disease – insects, vermin, and birds must be excluded from the curtilage of the meat plant. Measures are in place and reviewed regularly so that the environment of the meat plant suffers minimum contamination from the animals consigned to the plant, from the lorries and staff transporting the stock, and from all other potential vectors of pathogenic and spoilage agents into and throughout the plant.

Procedures within the meat or poultry plant itself dictate the standard of hygiene and safety of the product. The training and continuing professional development of staff in handling of and caring for the animals, in stunning and slaughter techniques, and in the dressing procedures of the carcasses are significant contributors to the overall standards pertaining through the plant. The operative has the potential to contaminate the carcass, and vice versa, but training and legislation obviate this from occurring.

Management must be aware of the likely points of entry of pathogenic agents and where they may have maximum impact on the meat and meat products. Recognition of these critical points where hazards are potentially most dangerous is now enshrined in the Hazard Analysis and Critical Control Point principle. Checking regularly and acting on findings at these areas in the meat plant allow a structured approach to the maintenance of hygiene and safety aspects of meat and poultry and their products. Increasing rates of throughput on the slaughterline must be accompanied by even more vigilant monitoring of hygiene standards.

The proper dressing of the carcass is fundamental to good hygiene. The hide, fleece, or feathers constitute a dirty coat to the sterile muscle and organs. The alimentary tract extending from the mouth to the anus is also a source of microbes and parasites. Therefore, the exterior coat and the complete intestinal tract must both be removed carefully so that the absolute minimum of contamination occurs. The surface of meat or muscle is an excellent medium on which pathogens and spoilage organisms grow, so the separation of edible and inedible tissues must be carried out expertly and completely. For this reason, the movement of these two elements follows an agreed flow chart implicitly within the plant so that 'clean' and 'dirty' areas are entirely separate. No crossover of staff, equipment, and tools is allowed; however, in the case of supervising officers moving between areas, a strict regime of changing protective clothing must be followed.

Water used within the meat plant must be of a quality that does not present a risk to the meat or meat product. Potable water is routinely checked in the slaughterhouse and meat processing plant so that pathogenic and spoilage organisms are not allowed entry. The monitoring is supervised by the official veterinary surgeon of the plant, as are the antemortem examinations, post-mortem inspections of the animals or birds and carcasses to check for pathological change, residues of veterinary medicines, and other substances, and the storage temperature control for meat and meat products. In many countries, veterinarians are helped in their duties by assistants/meat-hygiene inspectors to ensure that only healthy carcasses and their offals are deemed fit for human consumption.

These duties are delineated in national meat hygiene and inspection legislation, which is harmonized throughout the European Community. Other meat-producing countries, e.g., the USA, Australia, and New Zealand, adhere to their own similar legislative controls.

Meat products may undergo little processing, e.g., steak tartare, when no cooking takes place, or undercooking may unwittingly allow pathogenic agents to survive, e.g., lightly barbecued beef burgers being the source of an outbreak of hemolytic *Escherichia coli* O157. The increase in the number of chilled ready-made meals supplementing some frozen products may also be conducive to poisoning with organisms or toxins surviving in the meat if proper procedures in storage, preparation, and cooking are neglected.

The integrity of properly prepared meat and poultry and their products can be – and often is – undermined by contamination by a food handler who has neglected proper personal hygiene procedures or who is a carrier of a food-poisoning organism. The mishandler or carrier could, of course, be the consumer in their own home.

In conclusion, therefore, the arrival of meat and poultry and their products on the table from the farm can be hazardous if neglect occurs at any of the steps of the journey. Legislation, guidelines, and well-recognized proven routines ensure safe and healthy products, provided these are properly understood and implemented. However, one significant area of concern to the consumer, in the UK in particular, is the catastrophic bovine spongiform encephalopathy (BSE) epidemic that has devastated the livestock and meat industry.

Control measures introduced early on in the epidemic reduced the incidence of BSE markedly after its peak in 1992 and 1993, although the long incubation period (4–5 years) meant that any preventative legislation took the same time for its effects to be realized. Much research has been carried out in the UK and around the world on BSE, with consequent benefits accuring to other transmissible spongiform encephalopathies, including scrapie in sheep and Creutzfeldt–Jakob disease (CJD) in humans.

Some young consumers of beef contaminated by the BSE agent have contracted a new variant of CJD. However, this contamination occurred prior to the controls introduced in the late 1980s. It will be some years before it is known how many victims will succumb to this new form of CJD on account of the unknown length of the incubation period in humans.

See also: **Anemia (Anaemia)**: Iron-deficiency Anemia; **Bioavailability of Nutrients**; **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; **Coronary Heart Disease**: Etiology and Risk Factor; Prevention; **Fatty Acids**: *Trans*-fatty Acids: Health Effects; **Iron**: Physiology; **Protein**: Food Sources; Quality; **Selenium**: Physiology; **Vegetarian Diets**; **Zinc**: Physiology

Further Reading

Aiello LC and Wheeler P (1995) The expensive-tissue hypothesis. The brain and the digestive system in human and primate evolution. *Current Anthropology* 36: 199–221.

Analytical Methods Committee of the Royal Society of Chemistry and Meat and Livestock Commission (MLC) (1991) Nitrogen factors for pork: a reassessment. *Analyst* 116: 761–766.

Analytical Methods Committee of the Royal Society of Chemistry and Meat and Livestock Commission (MLC) (1993) Nitrogen factors for beef: a reassessment. *Analyst* 118: 1217–1226.

Analytical Methods Committee of the Royal Society of Chemistry and Meat and Livestock Commission

(MLC) (1996) Nitrogen factors for lamb. *Analyst* 121: 889–896.
Chan W, Brown J, Church SM and Buss DH (1996) Meat products and meat dishes. Supplement to *McCance and Widdowson's, The Composition of Food*. Cambridge: Royal Society of Chemistry/London: Ministry of Agriculture, Fisheries and Food.
Chan W, Brown J, Lee SM and Buss DH (1995) Meat, poultry and game. Supplement to *McCance and Widdowson's, The Composition of Food*. Cambridge: Royal Society of Chemistry/London: Ministry of Agriculture, Fisheries and Food.
Denke M (1994) Role of beef and beef tallow, an enriched source of stearic acid, in a cholesterol lowering diet. *American Journal of Clinical Nutrition* 60(supplement): 1044S–1049S.
Gibson SA and Ashwell M (1997) New vitamin D values for meat and their implications for vitamin D intake in British adults. *Proceedings of the Nutrition Society* 56.
Keenan JM and Morris DH (1995) Hypercholesterolemia. Dietary advice for patients regarding meat. *Postgraduate Medicine* 98: 111–121.
Nutrition Society (1994) Symposium on 'The role of meat in the human diet'. *Proceedings of the Nutrition Society* 53: 263–333.
Scott L, Dunn JK, Pownall HJ *et al.* (1994) Effects of beef and chicken consumption on plasma lipid levels in hypercholesterolemic men. *Archives of Internal Medicine* 154: 1261–1267.
Shi B and Spallholz JE (1994) Selenium from beef is highly bioavailable as assessed by liver glutathione peroxidase EC 1.11.1.9) activity and tissue selenium. *British Journal of Nutrition* 72: 873–881.
Sinclair AJ, Johnson L, O'Dea E and Holman R (1994) Diets rich in lean beef increase arachidonic acid and long-chain n-3 polyunsaturated fatty acid levels in plasma phospholipids. *Lipids* 29: 337–343.
Ulbricht TLV (1995) *Fat in the Food Chain. A Report to the Ministry of Agriculture, Fisheries and Food*. Available from MAFF Library.

Hygiene

J F Gracey, Brookvale Hornsby, Carlisle, Cumbria, UK

Introduction

Traditional meat inspection procedures have remained virtually unchanged for over 100 years in many countries. They were mainly based on the work of the German veterinarian, Robert von Ostertag, who in 1872 published his textbook *Handbuch der Fleischbeschau fr Tierärzte, Arzte und Richter*, and on the pathogenesis of bovine tuberculosis and are thus solely concerned with postmortem gross pathology.

However, with the elimination of bovine tuberculosis and other notifiable diseases in many areas, the change in disease type to the so-called diseases of production, many of which are associated with outbreaks of food poisoning, and the new systems of intensive husbandry, especially of pigs and poultry, attention is now beginning to be focused on the live animal. Consumer concern about intensivism, animal welfare, and residues in foods of all forms, has also served to influence authorities about the safety of the food we eat.

In addition, there has been concern about some of the techniques used in meat inspection, notably the number of routine incisions made in meat and offal, for each incision results in varying degrees of contamination. Indeed, it has been advocated by some experts that the incision of lymph nodes should be abandoned and replaced by inspection only. The cost-effectiveness of traditional methods has also come under adverse criticism, but the complete absence of proper veterinary antemortem (AM) examination and laboratory procedures in many quarters has probably been the main reasons for a more scientific approach to the inspection of meat.

Although public health is the main consideration, much can also be achieved with regard to the promotion of animal health and welfare, which also demand attention when cost-effectiveness is under inquiry.

Demands of the Consumer

The consumer demands and is entitled to the following safeguards:

1. Meat free from all zoonotic bacterial and viral diseases and their toxins, e.g., tuberculosis, salmonellosis, listeriosis, campylobacteriosis, yersiniosis, giardiasis, vibriosis, colibacillosis, botulism, and ornithosis, and also free from parasitic diseases such as cysticercosis, trichinosis, hydatidosis. (*See* ***Campylobacter***: Campylobacteriosis; ***Listeria***: Listeriosis; **Parasites**: Illness and Treatment; **Viruses**; **Zoonoses**.)
2. Meat free from residues of drugs, hormones, antibiotics, heavy metals, pesticides, and tranquilizers. (*See* **Heavy Metal Toxicology**; **Pesticides and Herbicides**: Toxicology.)
3. Meat produced in a hygienic manner from livestock handled humanely.
4. Meat which is wholesome, palatable in flavor and tenderness and of the nature, substance, and quality demanded.

Efficient Meat Inspection

In order to achieve the foregoing desiderata, three important investigations are required:

1. AM examination of the live animal or bird with reference to its previous husbandry.
2. Postmortem (PM) examination of the carcass and its viscera.
3. Microbiological, pathological, and chemical tests, where required.

Should any of these fundamental requirements be omitted or performed at suboptimal levels, efficient meat inspection can be said to be lacking and the consumer possibly put at risk. While PM examination is still required to eliminate lesions of gross pathology, modern meat inspection must be concerned with the nontangible agents of *Salmonella, Listeria*, and *Campylobacter* which are related to the increasingly common human foodborne diseases. The vital importance of laboratory procedures is illustrated in the diagnosis of bovine spongiform encephalopathy (BSE). First identified in England in 1986, the condition, because of its long incubation period, only affects adult bovines, has a long clinical course with neurological and behavioral symptoms, and is invariably fatal. Diagnosis can only be made by demonstrating spongiform changes and the presence of scrapie-associated fibrils (SAF) in fixed brain tissue by light microscopy. (*See* **Bovine Spongiform Encephalopathy (BSE)**.)

It is necessary to remember that many conditions can only be recognized or suspected in the live animal; PM examination may reveal nothing or at the most only minor lesions which are easily overlooked. Among these are tetanus and neurological diseases such as listeriosis, BSE, Aujeszky's disease and rabies, and many subclinical conditions. The bovine which shows only slight loss of condition and mild diarrhea may well be a case of salmonellosis, but in the absence of significant PM findings may be passed as fit for human consumption and in addition contaminate other meat as well as personnel.

Legislative Authority

In England and Wales, legislative authority is contained in the Fresh Meat Export (Hygiene and Inspection) Regulations 1987 for export purposes; AM inspection is not required for the home market under the 1987 Meat Inspection Regulations. The equivalent European Community (EC) law is the 91/497/EC Directive which amends and consolidates Directive 64/433/EC on health problems affecting intra-EC trade in fresh meat to extend it to the production and marketing of fresh meat.

General provisions are made concerning the time of inspection – day of arrival at abattoir with reinspection if held overnight; facilities to be provided; need for adherence to animal welfare rules; resting for 24 h of tired and agitated animals; prohibition of slaughter of animals whose meat is likely to be unfit; and a detailed PM examination of certain categories of diseased animals.

Less than one page is devoted to AM, compared with four on PM. A further anomaly in EC law relating to AM inspection is given in the Council Regulation COM(89) 668 laying down health rules for the production and placing on the market of fresh poultry meat, which indicates that AM inspection can be carried out by lay assistants. One of the cardinal rules in the diagnosis of disease, in bird or beast, is that it be accurate, otherwise both public and animal health may well be jeopardized. The diagnosis of poultry disease is no less difficult and important than that in the larger farm animals.

Legal Controls in Milk and Meat Production

Apart from the Zoonoses Order 1989, which requires the State Veterinary Service (SVS) to be notified when *Salmonella* or *Brucella* organisms are isolated from samples taken from farm animals and their carcasses, products, and surroundings, and similar legislation relating to residues, the law associated with meat production is sparse compared with that for milk.

In England and Wales the Milk and Dairies (General) Regulations 1959 stipulate the following requirements: premises to be of registered basic structural standards; a safe and adequate water supply; effective hygiene and milking routines; prohibition on the sale of milk from diseased cows, with periodic inspections of premises and stock. In addition, the Milk (Special Designation) Regulations 1989 require the registration of licenced processing dairies (pasteurization; sterilization and ultrahigh-temperature (UHT) milk); registration of heat treatment plants with special reference to temperature recorders and flow diversion valves; microbiological standards for raw and heat-treated milk; prescribed heat-treatment processes and special conditions of sale for untreated milk. (*See* **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; **Pasteurization**: Principles; **Sterilization of Foods**.)

Now that the new Food Safety Act of 1990 has defined 'food source' as 'any growing crop or live animal, bird or fish from which food is intended to be derived (whether by harvesting, slaughtering, milking, collecting eggs or otherwise),' it is to be

expected that comparable legislation will be enacted for meat production.

Purposes of Antemortem Inspection

Since AM and PM procedures are intimately linked, some of the following also apply to the latter:

1. Protection of the public health against bacterial, viral, parasitic and chemical hazards.
2. Protection of livestock against the spread of infectious diseases, especially notifiable diseases.
3. Protection of animal and food handlers against zoonoses.
4. Reduction of certain conditions and injuries causing unnecessary losses in livestock through disease data retrieval and feedback to producers.
5. Assurance of the meat industry and the public against meat of inferior quality, with avoidance of wastage of a valuable commodity.
6. Prevention of contamination of meat, premises, equipment, and personnel by dirty animals.
7. Efficient handling of livestock before and at slaughter to insure high standards of animal welfare and subsequent good meat quality and shelf-life.
8. The area of PM examination rendered more straightforward with fewer condemnations.
9. 'One of the most important functions of antemortem inspection is to ensure that animals are rested sufficiently so that signs important to inspection disposition are not masked. It also ensures that signs that are important to inspection disposition but that may be less readily observed (or not evident) at postmortem examination can be taken into account in reaching a decision as to safety and wholesomeness of meat' (Codex Alimentarius 1991).

Preslaughter Livestock Handling

Although AM inspection is usually taken to represent the clinical examination immediately before slaughter (AM immediatum), its full scope embraces preslaughter livestock handling and extends to the husbandry of the animal on the farm.

Factors such as the length of journey, time of year, environmental temperature, climatic conditions, food and water availability, species/age/sex of animal and its source, e.g., market or farm, influence the lairage holding time. It is important to keep farm social groups intact, to segregate horned and aggressive animals, to isolate the female in estrus, and the sick or injured beast (for further examination). Young bulls and spring lambs are best slaughtered early, as is the cow in good condition in cold conditions since she may succumb to hypomagnesemic tetany. Stress is the evil which must be controlled at all costs through efficient handling by trained stockmen. Much can also be achieved by providing straw and fine tepid-water sprays for pigs in lairage pens to prevent fighting.

Preslaughter lairage periods can therefore vary. But excessively long periods are contraindicated since they can result in deaths, injuries, loss of condition, cross contamination, difficulties in lairage cleansing, and problems with carcass dressing, e.g., excess ruminal/intestinal fluid in cattle.

To be humane to animals is to handle them in a manner least distressing to them and with full consideration for their normal species-specific behavior requirements.

Decisions on Antemortem Inspection

UK and EC law in the main deal with the system of AM inspection. The only legislation in the UK governing the admission of animals to meat plants is contained in the Slaughterhouses Hygiene Regulations 1977 (Amendment 1987), which deal only with casualty animals, i.e., those which are affected with certain diseases or injuries or represent culls from herds due to infertility or old age.

Since the aim of meat inspection is to produce a safe and wholesome product for the consumer, the conditions governing the admission of animals for slaughter must be extremely stringent in order that only apparently normal livestock and genuine casualty animals are accepted.

The following categories of decisions constitute the ideal rather than that specified by law:

1. Refusal or admission under special control.
2. Release for slaughter of all apparently normal animals.
3. Delayed slaughter, e.g., because of inadequate drug withdrawal period or temporary illness.
4. Slaughter under special circumstances, e.g., casualty animals at the end of the kill or in emergency slaughter unit.
5. Total condemnation: (a) dead, dying and disabled animals; (b) pigs with temperatures of 41°C and higher (other species > 40.6°C); (c) conditions justifying total condemnation on PM examination; (d) animals showing symptoms of toxic, metabolic, nervous, or circulatory disturbance; (e) nutritional imbalances and parasitic diseases – listeriosis, leptospirosis, Aujeszky's disease, tetanus, rabies, anaplasmosis, ketosis, parturient paresis, scrapie, BSE, grass tetany, transport tetany, strangles, acute inflammatory lameness, extensive fistula,

glanders, farcy, swine fever, anthrax, and animals which have been vaccinated with anthrax vaccine within the previous 6 weeks.

Category 5 is considered to be a most important one since these are all casualty animals, many of which have been certified as such by practicing veterinary surgeons under the Slaughterhouses Hygiene Regulations. Although in existence in the USA and other countries, this particular category does not yet apply in the UK. It is, however, a most important sector and one which should be embodied in all meat inspection legislation. The fact that up to 50% of such cases may ultimately be totally condemned is not only a reflection of some standards of veterinary certification but can also be regarded as a waste of the time of meat plant staff, whose health on occasions is also put in jeopardy.

The information gleaned at AM inspection must be recorded and passed to the meat inspectorate. Should a zoonosis, e.g., ringworm or contagious pustular dermatitis (orf), be detected, all staff – especially the lairage personnel – must be alerted. While orf usually occurs as a benign skin lesion on the hands and arms of humans, there is evidence that a serious systemic form occurs and is becoming more prevalent. Extreme vigilance is necessary, therefore, especially with casualty animals. It is not unknown for an anthrax case to be consigned for slaughter!

Antemortem Completum Inspection

It was indicated earlier that a proper examination of the live animal must extend beyond the immediate preslaughter clinical inspection if guarantees of food safety and wholesomeness are to be made.

> The health status of the farm of origin and the husbandry of slaughter animals have a significant effect on the safety and wholesomeness of meat. In this respect, all efforts should be made to collect and evaluate information that might have influence on antemortem and postmortem (Codex Alimentarius October 1991).

The farm is, in fact, the most important hazard analysis and critical control point (HACCP) in all the meat hygiene operations which begin with animal husbandry and end in the consumer's home. (*See* **Hazard Analysis Critical Control Point**.)

Several countries, notably The Netherlands (pigs), New Zealand (sheep), Northern Ireland (cattle), USA, and Canada, which are concerned about current meat inspection systems, are enacting schemes referred to as integrated quality control (IQC) based on abattoir and farm disease data in conjunction with codes of practice for livestock producers (Figure 1). In some cases AM inspection is carried out on the farm, and is repeated at the point of slaughter. Especially for poultry flocks, this would seem to be an ideal which could be easily performed to great advantage.

The on-farm information which is of importance for meat inspection purposes relates to morbidity and mortality rates, animal welfare levels, the use of antibiotics, hormones, and veterinary drugs, results of laboratory tests (microbiological, serological, and chemical for residues), results of tests for tuberculosis and brucellosis and type of feeding. Since one of the main concerns of the meat hygienist is the contamination of carcass meat from filthy hides and fleeces, the standard of farm hygiene, especially of the loose housing of cattle, feeding of sheep on root crops and stock during the winter months, are of vital concern. Commenting on this important issue, which is both a

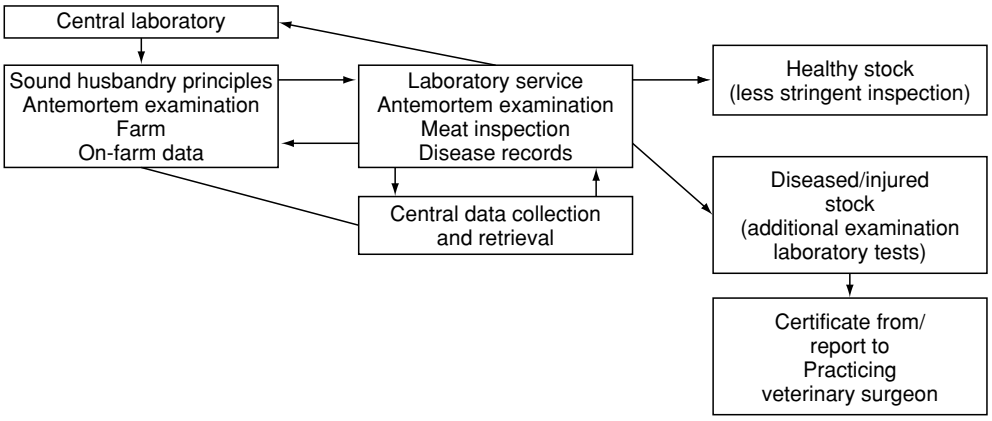

Figure 1 Centralized meat hygiene service. After Skovgaard N (1991) Ante mortem and post mortem meat inspection. Need for improvement. In: Hannan J and Collins JD (eds) *The Scientific Basis for Harmonising Trade in Red Meat*. Dublin: University College, with permission.

hygiene and welfare matter because of the damage to skin, the Richmond Committee on the Microbiological Safety of Food (1991) stated that 'farmers can, therefore, contribute to food safety by producing healthy, clean and unstressed animals for slaughter.' The abattoir data which are transmitted back to the producer for control purposes consist of the disease and injury findings encountered in the stock at slaughter. The practicing veterinary surgeon who supplies certificates for casualty animals should receive a report in respect of such animals.

The first Farm Quality Assurance Scheme for Beef in the UK is already under way in Northern Ireland. At present a voluntary arrangement, it should provide a basis for eventual legislation governing the husbandry of beef cattle. Organized by the Northern Ireland Livestock and Meat Commission (NILMC), it has the following Code of Practice:

> Commitment by farmers to produce a wholesome product with due regard to animal welfare and disease control; 3 months' residence on farm for all accepted stock; feeding to be grass and vegetable protein (root crops, brassicas, molasses, etc., may be approved); no unlicensed growth promoters or feed additives to be used; Department of Agriculture Code of Recommendations for Welfare of Livestock to be adopted; animals to be kept clean and in good condition; dirty animals ineligible for scheme; farm investigators by NILMC may be necessary; only healthy animals, naturally produced, to be eligible; declaration of freedom from BSE necessary; veterinary record and programme of disease control to be kept; strict use of veterinary drugs and observance of withdrawal periods; segregation of sick, injured and fallen animals; stress-free handling and transport; bruised carcasses ineligible for scheme.

Future Developments

AM and PM inspection procedures must always be appropriate to the incidence and type of diseases prevalent in the various species of the food animals. However, a supply of clean, healthy, and residue-free animals for slaughter can only be achieved by careful on-farm and preslaughter inspection, due regard having been paid to proper husbandry methods.

The detection of certain parasitic diseases, e.g., cysticercosis, trichinosis, fascioliasis, hydatidosis, thysanosomiasis, now only possible at PM, may well be carried out at AM by enzyme-linked immunosorbent assay (ELISA) examination of blood at AM, thus obviating the need for the current mutilating and contaminating incisions of meat and offal.

Strict veterinary certification of healthy and casualty stock will also assume a dominant role in the future by providing data for the meat inspectorate to assist in carcass/offal judgments. Animal health will also be promoted through feedback from PM inspection to the farmer and the practicing veterinary surgeon, the latter in relation to casualty and cull animals. Refer to individual meats.

See also: **Bovine Spongiform Encephalopathy (BSE)**; ***Campylobacter***: Campylobacteriosis; **Hazard Analysis Critical Control Point**; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; **Heavy Metal Toxicology**; **Legislation**: Codex; ***Listeria***: Listeriosis; **Parasites**: Illness and Treatment; **Pasteurization**: Principles; **Pesticides and Herbicides**: Toxicology; **Sterilization of Foods**; **Viruses**; **Zoonoses**

Further Reading

Hathaway SC and McKenzie AI (1989) Impact of ovine inspection systems on processing and production costs. *Veterinary Record* 124: 189–193.

Hathaway SC, McKenzie AI and Royal WA (1987) Cost-effective meat inspection. *Veterinary Record* 120: 78.

Kilgour R (1978) The application of animal behaviour and the humane care of animals. *Journal of Animal Science* 46: 1478–1486.

Report of the Committee on the Microbiological Safety of Food (1991) Part 11, 1991. London: HMSO.

Shuttleworth VS (1988) Systemic contagious pustular dermatitis or orf. *British Homeopathic Journal* 77: 12.

Skovgaard N (1991) Ante mortem and post mortem meat inspection. Need for improvement. In: Hannan J and Collins JD (eds) *The Scientific Basis for Harmonising Trade in Red Meat*. Dublin: University College.

Extracts

A M Pearson, Oregon State University, Corvallis, OR, USA

Introduction

Because of its distinctive and desirable flavor, meat extract has been a highly prized and useful item of commerce since the eighteenth century. As its name suggests, meat extract is obtained upon liquid extraction and concentration of the water-soluble fraction from meat. It is not the same as meat juice (drip) or press fluid since it contains a number of different constituents and others are generally present in higher concentrations. Nevertheless, the term 'meat extract' is not very specific but refers to various extracts made from meat, bones, and liver, which differ in their composition and flavor as a result of differences in

their origin and the components found in various raw materials.

Production of Meat Extract

The famous German chemist Justus von Liebig reported that meat extract could be produced by mixing 10 lb (4.5 kg) of lean meat with 5 lb (2.25 kg) of water and then pressing and collecting the soluble fluid; the process was repeated three times. The combined extracts were filtered, heated, skimmed, reheated, and concentrated under vacuum to give the first commercial product, which was produced and marketed in 1865 by Liebig's Extract of Meat Company Limited. The early procedure included removal of the meat from the bones, trimming off any visible fat, and aerating the remaining tissue for at least 20 h. The aerated lean was then hashed into a series of round-bottom kettles arranged so that the most concentrated meat stock always came in contact with freshly hashed meat and the previously extracted meat was reextracted with clear water. The extraction temperature was maintained at 90 °C or below while concentrating the meat extract. If higher temperatures are used, the alcohol-soluble material decreases and the amount of gelatin extracted increases.

Meat used for making the extract can be pressed or extracted with water or cooked to help remove the constituents responsible for the meaty flavor. The extract is then concentrated by removal of most of the water. Raw bones, preferably without removal of any firmly attached lean meat, and liver can also be the source of raw materials for making two products similar to meat extract, i.e., bone extract and liver extract, respectively.

Argentina is the largest producer of meat extract, which is a byproduct from canning beef and corned beef production. The meat to be canned is given a rapid cook for about 30 min at 95–100 °C. Then the watery stock is filtered to remove the fines and coagulated protein. The excess moisture is initially removed under vacuum and then further concentrated by heating in an open pan. Coarsely ground bones can be used to produce beef stock by adding water, and cooking at 115 °C or above in a pressurized kettle. The fat is allowed to separate and solidified by cooling in order to remove it. The beef stock is then concentrated under vacuum followed by pan evaporation to help develop the flavor.

Meat cooked as described above loses about 40% by weight, with most of the losses being recovered in the cook water. The cooked meat is rinsed with cold water, which is then added to the cook water. This results in about 2.2 l of stock per kilogram of raw meat. Generally, the stock is concentrated by cooking up to six batches of meat in the same stock before it is concentrated further under vacuum. Since the final stock from the combined cook water is more concentrated, there is a saving in the amount of energy needed for production. Figure 1 demonstrates the effect of the number of cooks on the concentration of solids in grams per liter for both the beef extract raw stock and the beef extract heated and filtered stock. The plot demonstrates that each batch increases the amount of solids in a fairly linear manner. It also shows that heated and filtered stock always has less solids than the raw stock due to the removal of the fines and scum or coagulated proteins upon filtering. The duration of boiling also increases the concentration of solids in the extract, although the maximum yield per unit of meat is reached at about 1 h. The increase in yield drops off sharply upon boiling for longer periods.

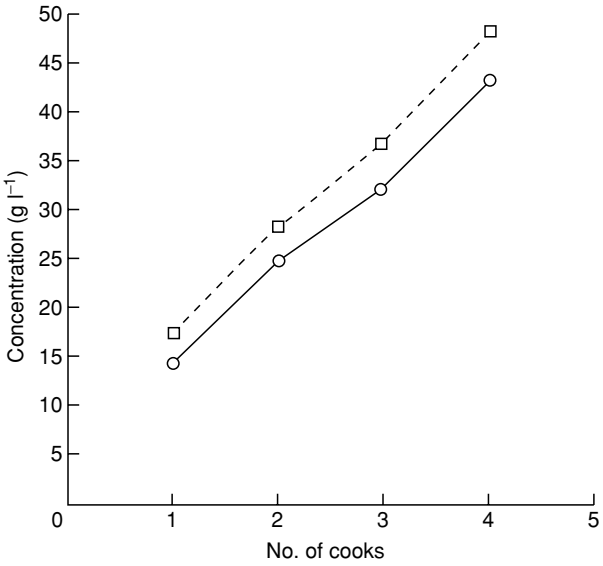

Figure 1 Influence of the number of cooks and filtering on the concentration of beef extract to 16% moisture content. Squares, 16% beef extract raw stock; circles, 16% beef extract heated and filtered stock. Reproduced from Ockerman HW and Pellegrino JM (1988), with permission.

The holding temperature of the stock also has a marked effect upon the concentration, with higher temperatures increasing the solids per unit volume. On the other hand, lower temperatures decrease the rate of concentration.

Types and Grades of Meat Extract

A number of types of meat extract and similar products are available on the market, and include the following: direct extract (essence of meat), no. 1 extract, fluid extract of meat, broth or stock, and liver

extract. However, there are no strict standards of identity, which results in a great diversity and variability between and within products.

Direct extract or essence of meat is manufactured by extracting minced beef muscle with water for a sufficiently long time to produce a gel upon cooling. Gel or jelly formation is due to extraction of collagen by heat so that it solidifies upon cooling. Approximately 13–14 kg of lean meat is required to yield about 1 kg of extract, or essence of meat that contains around 34% moisture.

No. 1 extract, sometimes simply called meat extract, is produced as a byproduct of corned beef. Its production involves immersion of chipped beef into boiling water for a short period of time. Little gelatin is extracted by this procedure. About 50 kg of meat is needed to produce 1 kg of extract containing approximately 17% moisture. The US Department of Agriculture Code of Federal Regulations of 1986 states that meat extract should contain not more than 25% moisture. Thus, it is considered to be a preserved product because it contains only 20% moisture or less and a minimum of 5% sodium chloride and 7% creatinine. The fat content should not exceed 0.6% and the nitrogen content should be at least 8%. The nitrogenous components should consist of at least 40% meat bases and 10% creatine plus creatinine.

Fluid extract of meat is made by cooking fresh meat to produce a stock containing 3–4% of solids. The stock is concentrated and evaporated under vacuum to yield a product with 60% solids. Salt is then added to the water fraction at 25.5% (w/w). This results in an extract containing 9.3% of added salt (10.2 kg of added salt per 100 kg of concentrated extract).

Broth (beef or pork) or stock are terms used interchangeably to describe the product obtained by simmering meat and/or bones (crushed) in water along with seasoning ingredients. The product(s) must have a moisture-to-protein ratio of 135:1. A concentrate of broth or stock can be made by reducing the moisture-to-protein ratio to 67:1.

Liver extract is produced by extracting raw ground livers with slightly acidified hot water at pH 5. The stock is then concentrated to a paste under vacuum at low temperature to a moisture content of 35%. Thus, liver extract contains more moisture than meat extract (35 versus 25%). Its high moisture content makes liver extract highly perishable so more salt needs to be added or it must be refrigerated or heat-sterilized. In contrast to meat extract, which is primarily used as a flavor base for soups, sauces, and other foods, liver extract is used by the pharmaceutical industry as a nutritional supplement.

Although individual companies producing meat extract may have their own grading sytems, there is no standard grading. Some companies divide their products into no. 1 and no. 2 grades on the basis of their creatinine content, with the former containing 7% and over and the latter grade containing less than 7% creatinine. Some other processors produce three grades of meat extract, with International Packers Limited calling their products Supreme, Premium, and Select. Supreme contains 16.75–17.25% moisture, a minimum of 44% soluble organic material and 7% creatinine, and a maximum of 1.5%water-insoluble material, 25% ash, and 4% salt. Premium contains 19–20% moisture, a minimum of 44% soluble organic material and 7% creatinine and a maximum of 1.5% hot-water-insoluble material, 25% ash, and 4% salt. Select contains 23–24% moisture, a minimum of 40.5% soluble organic material and 6.4% creatinine and a maximum of 1.4% hot-water-insoluble material, 23% ash, and 3.75% salt.

Composition of Meat Extract

The chemical compositions of beef extract, mutton extract and liver extract are shown in Table 1. The data provided give the moisture, salt, creatinine, and ash contents as well as the percentages of soluble organic material for beef and mutton extracts. For liver extract, however, the compositional data are presented in terms of moisture content, solids, and ash. Although data are given for a number of B-vitamins for all three extracts, the list of vitamins is more complete for liver extract, reflecting its greater nutritional contribution. Particularly noticeable are the higher values for vitamin B_{12} (15.4 times greater), riboflavin (3.5-fold greater), and pantothenic acid (20-fold higher) in liver extract. (*See* **Beef** and refer to individual nutrients.

Table 2 provides information on the nonvolatile constituents present in meat extract. Particularly note worthy is the presence and relatively large amounts of creatine and creatinine, which are important quality parameters. Although yeast extract is frequently used as a substitute for meat extract, it does not contain either creatine or creatinine. These two guanidine compounds combined should not comprise less than 10% of the total nitrogen on a dry basis in high-quality meat extracts.

Another interesting characteristic of meat extract is the considerable amount of inosine and inosinic acid, which have been shown to be meat flavor enhancers. Thus, inosine and inosinic acid probably make a major contribution to the desirable flavor and/or aroma of meat extract. The relatively high concentration of amino acids may also contribute to the Maillard reaction during heating of the extract

Table 1 Chemical composition of beef, mutton, and liver extract

	Minimum (%)	*Maximum (%)*	*Average (%)*
Beef extract			
Moisture	18.80	19.95	19.41
Salt	4.00	4.97	4.76
Creatinine	7.00	7.10	7.02
Soluble organics	43.00	44.50	43.70
Insoluble in water	1.50	2.00	1.70
Ash	23.00	25.00	24.00
Thiamin			10 μg g^{-1}
Riboflavin			35 μg g^{-1}
Niacin			1200 μg g^{-1}
Pyridoxine			5 μg g^{-1}
Pantothenic acid			25 μg g^{-1}
Vitamin B_{12}			0.52 μg g^{-1}
Mutton extract			
Moisture			17.90%
Salt			3.97%
Creatinine			5.90%
Organics soluble in alcohol (80 °C)			46.55%
Insoluble in water			1.90%
Ash			23.35%
Protein			57.83%
Liver extract			
Moisture			35 $\pm$ 1%
Solids			65 $\pm$ 1%
Ash			11 $\pm$ 1%
Vitamin B_{12}			Not less than 8 μg g^{-1}
Thiamin			Not less than 10 μg g^{-1}
Riboflavin			Not less than 200 μg g^{-1}
Folic acid			Not less than 15 μg g^{-1}
Nicotinic acid			Not less than 1000 μg g^{-1}
Pantothenic acid			Not less than 500 μg g^{-1}
Choline			Not less than 15 000 μg g^{-1}
Insoluble: the product gives a clear solution in water			

Reproduced from Ockerman HW and Pellegrino JM (1988) Meat extractives. *Advances in Meat Research* 5: 303–339, with permission.

for concentration and play a role in the flavor of meat extracts. (*See* **Browning**: Nonenzymatic; **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Microbiology of Meat Extract

Although commercial meat extract usually contains bacteria and mold spores, pathogenic organisms are seldom present. Boiling of the stock to concentrate it results in a product that is nearly sterile, but subsequent handling and use of contaminated equipment result in recontamination. As meat extract provides an excellent medium for the growth of bacteria and temperatures used during concentration by vacuum are favorable for growth of microorganisms, the finished meat extract contains viable bacteria and molds. The best procedures for controlling the growth of microorganisms are to decrease the moisture content and increase the salt content. Neither procedure, however, can be relied upon to control the growth of microbes alone, although together they provide some degree of safety from spoilage. Combined with a low storage temperature, a low moisture and high salt content are synergistic and help to control microbial growth. (*See* **Spoilage**: Bacterial Spoilage; Molds in Spoilage.)

At a moisture content of 16%, meat extract stored at 20 °C has been shown to be relatively stable during 48 days of storage. At 20% moisture, however, meat extract needs to be protected from microbial spoilage by addition of salt. Since extracts made from liver and bones are hard to dehydrate below 25% moisture, they are more susceptible to bacterial and mold growth so must be either held at low temperatures (> 10 °C) or have a greater amount of added salt or both.

Table 3 presents some average microbial counts for meat extract. The data demonstrate that there is considerable variation in both the kinds and numbers of organisms present. The great amount of variability is no doubt associated with sanitation, processing, and

Table 2 Chemical analysis showing nonvolatile components in meat extract

(a) Composition of no. 1 meat extract (17–28% moisture and 13.21% protein)

Component	*Dry weight*	
	Amount (% of total)	*N (% of total wt)*
Amino acids		
3-Methylhistidine	Tr	
α-Alanine	1.32	
Serine	0.10	
Methionine	0.01	
Isoleucine	0.08	
Leucine	0.08	
Histidine	0.03	
Taurine	0.32	
Citrulline	0.26	
Total	2.20	0.22
Peptides		
Carnosine	3.70	0.92
Anserine	0.75	0.16
Iminazole peptide	1.88	0.47
Guanidines		
Creatine	4.80	1.54
Creatinine	5.51	2.04
Methylguanidine	(~0.1)	
Guanidine	(~0.1)	
Purines, etc.		
Hypoxanthine	1.90	0.78
Inosine	0.70	0.15
Inosinic acid	Tr	
Protein (14.9% N, 0.1% ash)	10.92	1.65
Organic acids		
Lactic acid	14.60	
Glycollic acid	0.98	
Succinic acid	1.26	
β-Hydroxybutyric acid	(>1)	
Carnitine	3.3	0.29
Choline	Tr	
Urea	0.11	
Ammonia	0.42	
Inorganic matter (8.95% K; 7.3% P_2O_5)		
Coloring matters	18.30	4.23
Total	100.73	12.85

(b) Composition of volatile components of beef extract

Component	*Peak size*
Hydrogen sulfide	Large
Methyl mercaptan	Large
Ethyl mercaptan	Small
Dimethyl sulfide	Small
Acetaldehyde	Large
Propionaldehyde	Small
Isobutyraldehyde	Small
Acetone	Medium
Isovaleraldehyde	Small
Methyl ethyl ketone	Small
Methanol	Small and only in some samples
Ethanol	Small and only in some samples

Tr, trace.

Reproduced from Bender AE and Ballance PE (1961) as modified by Ockerman HW and Pellegrino JM (1988) Meat extractives. *Advances in Meat Research* 5: 303–339, with permission.

Table 3 Average microbiological count for meat extract with values for products from Argentina and the USA

Argentine product	
Soup in evaporator before going to evaporator	$32\ g^{-1}$ TPC
Soup going to evaporator	$1300\ g^{-1}$ TPC
Concentrated soup out of evaporator	$2000–30\,000\ g^{-1}$ TPC
Thermophiles	$500\ g^{-1}$
Anaerobes	$100\ g^{-1}$
Coliforms	0
Clostridium perfringens	Negative in 0.1 g
Meat extract out of can (30 °C)	$100\,000\ g^{-1}$ TPC
Meat extract out of can (55 °C)	$237\,500\ g^{-1}$ TPC
Putrefactive anaerobes	$927\ g^{-1}$
Indole producers	0 (sometimes present)
H_2S producers	0 (sometimes present)
Staphylococcus group	0
Salmonella–Shigella group	0
US product	
Total plate count (TPC)	$<1000\ g^{-1}$
Coliforms	$<10\ g^{-1}$
Escherichia coli	Negative g^{-1}
Staphylococcus aureus	$<10\ g^{-1}$
Salmonella spp.	Negative per 25 g
Flat sour spores (FSS)	<50 per 10 g
Total thermophilic spores (TAS)	<125 per 10 g

Reproduced from Ockerman HW and Pellegrino JM (1988) Meat extractives. *Advances in Meat Research* 5: 303–339, with permission.

storage conditions, as well as the final contents of moisture and salt. Although any great increase in the number of thermophilic bacteria would not be expected to occur in a meat extract containing 16% moisture, a potential danger may exist upon adding it to other foods having a higher moisture content. Good sanitation and clean-up and avoiding contaminated tanks, pipes, and evaporators can help to reduce recontamination of meat extracts. Low-temperature storage is also an important step in controlling bacteria and molds.

Uses for Meat Extract

Meat extract is a common flavoring additive for soups, stews, sauces, casseroles, canned meat items, pot pies, bouillon and bouillon cubes, gravies, and other items where meat flavoring improves the products. The flavor of meat extract makes it a desirable additive to a variety of products. The percentage of meat extract needed for flavoring varies widely, depending upon the food to which it is being added, but generally falls in the range 5–25%. Its desirable flavor and aroma may enhance the flavor of some products at even lower concentrations. Although yeast extract is used as a substitute, chefs and food connoisseurs prefer the characteristic flavor and aroma that meat extract imparts to other foods.

See also: **Beef**; **Browning**: Nonenzymatic; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Spoilage**: Bacterial Spoilage; Molds in Spoilage

Further Reading

Bender AE and Ballance PE (1961) A preliminary examination of the flavor of meat extract. *Journal of the Science of Food and Agriculture* 12: 683–687.

Dvorak Z (1981) Creatine as an indicator of net muscle proteins. *Journal of the Science of Food and Agriculture* 32: 1033–1036.

Liebig J (1847) Ueber die Bestaudtbeile der Fliisigbeetin des Fleisches. *Aundleu der Chemie und Pharmacia* 62: 257–298.

LJ Minor Corporation (1984) *Quality Food Bases – An Essential Culinary Tool*. Cleveland, OH: LJ Minor.

Ockerman HW and Pellegrino JM (1988) Meat extractives. *Advances in Meat Research* 5: 303–339.

Wood T and Bender AE (1957) Analysis of tissue constituents. Commercial ox muscle extract. *Biochemical Journal* 67: 366–373.

Megaloblastic Anemia *See* **Anemia (Anaemia)**: Iron-deficiency Anemia; Megaloblastic Anemias; Other Nutritional Causes

MELONS, SQUASHES, AND GOURDS

A A Schaffer, Agricultural Research Organization, Bet Dagan, Israel
H S Paris, Agricultural Research Organization, Ramat Yishay, Israel

Background

The Cucurbitaceae family encompasses over 800 species of plants known collectively as gourds or cucurbits. These include cucumbers, melons, watermelons, pumpkins, squash, and many others. Perhaps more than 20 genera are used for culinary purposes, which usually includes consumption of the mature fruit flesh, whole immature fruits, and/or seeds. Three genera, *Cucumis* (cucumbers, melons), *Cucurbita* (pumpkins, squash), and *Citrullus* (watermelons), rank among the top 10 in economic importance among the vegetable crops of the world, and several others have regional importance. Cucurbits are also grown for use as ornaments and containers, and some are used for medicinal applications and other purposes. Some wild cucurbits have potential economic value. This article reviews the botany, horticulture, and food value of cucurbits, with emphasis on the three leading genera.

Botany

Most cucurbits are herbaceous, tendril-bearing vines that are adapted to warm climates and are sensitive to frost. The natural distribution of most species is in the tropics, but some genera (*Bryonia*, *Cucurbita*, *Ecballium*, *Echinocystis*) contain species that are distributed in temperate regions. In cultivation, cucurbits are distributed in almost all arable regions worldwide. Those cucurbits that are not adapted to cool temperatures and that are grown for their mature fruits and seeds are usually not successful in regions with short, cool summers. Others, most notably cucumber (*Cucumis sativus*) and summer squash (*Cucurbita pepo*), are grown for their immature fruits and are more tolerant of cool temperatures than most cucurbits, and therefore have an especially wide distribution in cultivation.

Most cucurbit species are mesophytes and have large palmate leaves, fibrous roots, and prominent fruits. Under favorable conditions, the plants grow and spread quickly, as stems can elongate 30 cm or more within 24 h. The foliage can be glabrous, softly or stiffly hairy, or spiculate. The plants usually are monoecious, that is, have separate staminate (male) and pistillate (female) flowers on the same plant. Most often, the flowers are nectar-producing and foraged by bees, with green calices consisting of five sepals fused at the base and yellow corollas consisting of five petals fused at the base. Staminate flowers differentiate at lower nodes and are produced in greater numbers than pistillate flowers. However, once differentiated, the pistillate flowers can develop at a faster rate than the staminate flowers. The staminate flowers have thinner, longer pedicels than the

pistillate flowers. The pistillate flowers have inferior ovaries that usually are round or oval in shape. In cross-section, the triloculate ovary can be seen to contain 10 subepidermal main veins, five of which connect the pedicel with the central vein of each sepal and five of which connect the pedicel with the central vein of each petal. In addition to the main vasculature system of the cucurbit ovary, there is also a highly anastomosing vasculature system in the fruit, which contributes to some of the genetic variability of fruit structure in the family. For example, the fibrous flesh of the spaghetti squash (*Cucurbita pepo*) and the fibrous net of the dried sponge gourd (*Luffa cylindrica*) are products of this anastomosing vasculature.

After fertilization, the ovary develops into a fleshy fruit possessing exocarp, mesocarp, and endocarp. Generally, the fruits of wild cucurbits are small, round, and green, and the flesh often extremely bitter due to alkaloid compounds known as cucurbitacins. A fruit can contain as many as several hundred seeds, which, at maturity, contain two cotyledons each but no endosperm. The seeds are not bitter and usually are flat, but there is a greater than 10-fold range in seed length among cucurbit genera, from several millimeters to a few centimeters.

Domestication of some of the Cucurbitaceae is ancient, predating that for some major grain crops. There is evidence that *Cucurbita pepo* (pumpkin, squash) was first domesticated at least 10 000 years ago. Several other cucurbits have also been domesticated for thousands of years. In the Bible (Numbers 11:5), the Israelites longed for the *avattihim* (watermelons) and *qishu'im* (melons, var. *chate*) of Egypt during their wanderings in the Sinai Desert.

Following the thousands of years of domestication by man, the cultivated cucurbits differ from their wild counterparts, most obviously in their larger and fewer vegetative and reproductive parts. Cultivated forms have larger leaves, thicker stems, fewer branches, and larger and fewer fruits and seeds. In some cultivated forms, fruit shape is not round but distinctly elongate. Also, different fruit colors and color patterns occur among cultivated forms. Cultivated forms also differ sharply from their wild cousins by having fruits that are more palatable, being nonbitter and less coarsely fibrous, but higher in starch, sugar, and carotenoid contents. A wealth of variability for fruit size, shape, and color occurs in the Cucurbitaceae family.

Quality traits desired in mature cucurbit fruits in some cases are the same as those desired in immature cucurbit fruits but differ in other cases. Some characteristics, such as sweetness and flesh color, are fully expressed in mature fruits but hardly at all in young fruits. Other characteristics, such as size of the seed cavity, can be important in determining the palatability of immature fruits but are less important in mature fruits. Many of the cultivated cucurbit species have been grown for the culinary use of their immature fruits, and some for their mature fruits. A few, most notably pumpkins and squash (*Cucurbita* spp.) and melons (*Cucumis melo*) have been grown for the dual usage of their immature fruits as well as their mature fruits, or for the use of immature fruits in some regions but for the use of mature fruits in other regions, and as a consequence possess a great deal of intraspecific variation in fruit characteristics.

Among the cucurbits are some of the largest and fastest growing fruits in the plant kingdom. For this reason, some of the larger cucurbits have served as experimental material to study growth rates, carbohydrate import, and translocation physiology in developing fruit. Large pumpkins may show dry weight gains of $1.71\,g\,h^{-1}$, whereas the smaller fruited cucumbers and melons may increase in dry weight by $100–500\,mg\,h^{-1}$. The large differences in fruit size and shape in the cucurbits are already determined at the early stages of ovary development, prior to and immediately after anthesis. These traits are under strict genetic control, although nongenetic factors such as environmental conditions and intraplant competition for photoassimilates play strong roles in determining fruit size. Assimilate supply in the form of photosynthetic products can often be a limiting factor of fruit growth and final fruit size.

Melons

Melons (*Cucumis melo* L.) are prostrate, highly branched, softly hairy vines possessing rounded, heart-shaped leaves and unbranched tendrils. They are most commonly andromonoecious, producing a hermaphroditic flower at the first one or two nodes of the branches and one or more male flowers at all other nodes. Less commonly, they are monoecious, with pistillate flowers being produced instead of hermaphroditic flowers. Rare, recessive mutants produce perfect flowers or pistillate flowers at every node. The flowers are bright yellow but small, approximately 3 cm in diameter. *C. melo* is diploid, with 12 pairs of chromosomes ($2n = 24$), but some members of the genus *Cucumis* are polyploids.

Melons originated from tropical Africa north of the equator, and their cultivation spread northward and eastward thousands of years ago. Much diversity of melons is found in Iran, Turkey, Spain, India, China, and the central Asian republics of the former Soviet Union. Apparently, the first culinary use of melons by man was of the immature fruits, much in the manner

of cucumbers. Immature melons are not sweet but instead bland, similar to cucumbers. Sweet melons are a relatively recent development, perhaps dating back only to the Middle Ages.

Botanically, melons are considered to consist of two subspecies, *C. melo* ssp. *melo* and *C. melo* ssp. *agrestis* Jeffrey. Melons contain a wide diversity of fruit sizes, shapes, and colors (see Figure 1). This diversity has been the subject of many attempts at classification. Probably the most comprehensive and comprehensible classification is that of Pitrat *et al.*, who have recognized 16 cultivar-groups. Their classification is summarized in Table 1. The *inodorus* group includes the winter or casaba melons, as well as the honeydew, Jaune Canary and Piel de Sapo melons. The major netted muskmelons, especially those grown in the USA as 'cantaloupes' are in the *reticulatus* group. These include the heavily netted varieties such as PMR45, Hale's Best Jumbo, Top Mark and also the lightly netted, nonribbed varieties such as the Galia and Persian melons. The French Charentais-type melon, which is ribbed but generally smooth skinned, is classified in the *cantalupensis* group. However, the distinction between groups may become blurred with increasing breeding efforts, since all the groups of *Cucumis melo* are easily hybridized with one another. For example, while, in the past, the smooth cream-skinned honeydew melons were green-fleshed, there are now varieties of honeydew melons with the orange flesh of the *reticulatus* group. In the USA, melons are often classified as either cantaloupes, honeydews or 'specialty melons,' which include all the other types.

Figure 1 (see color plate 104) Melons.

According to FAO statistics, approximately 19 million tonnes of melons were produced in 2000,

Table 1 Classification of melons, *Cucumis melo*, to subspecies and cultivar-groups[a]

Subspecies Group	*Maturity at consumption*	*Description*
agrestis (*ovary with short hairs*)		
conomon	Immature, similar to cucumber, in Eastern Asia	Elongate
makuwa	Mature in Eastern Asia	White flesh, little aroma, flat/round/pyriform
chinensis	Mature in Eastern Asia	Colored flesh, little aroma, pyriform
momordica	Immature in India	Skin bursting when ripe
acidulus		Cooked in India, firm with crisp flesh, no sweetness or aroma
melo (*ovary with long hairs*)		
cantalupensis	Mature in Europe	Sweet flesh, aromatic, smooth or warted, climacteric
reticulatus	Mature in North America, Japan, Western Asia, and Europe	Sweet flesh, aromatic, netted, climacteric
adana	Mature in Turkey	Thin flesh, low in sugar, climacteric
chandalak	Mature in Central Asia	Sweet, climacteric, slightly wrinkled, low aroma
ameri	Mature in Central Asia	Sweet, climacteric, low aroma, netted
inodorous	Mature in Central Asia, Europe, and North America	Sweet, no aroma, usually wrinkled, not climacteric, long-keeping
flexuosus	Immature in North Africa	Very long fruit, climacteric
chate	Immature in North Africa	Elongate
tibish	Immature in Sudan	Small, oval, no sugar, no aroma
dudaim	Mature	Used as an ornamental or aromatic, small, round, no taste, highly aromatic, climacteric
chito		Very small, round, smooth yellow fruits

[a]From Pitrat M, Hanelt P and Hammer K (2000) Some comments on infraspecific classification of cultivars of melon. In: Katzir N and Paris HS (eds) *Proceedings of Cucurbitaceae 2000: the 7th Eucarpia Meeting on Cucurbit Genetics and Breeding*. *Acta Horticulturae* 510: 29–36.

one-third of that in China. Other important producing countries are Turkey, USA, Spain, Romania, Morocco, India, Mexico, and Egypt. The mature fruits of the orange-fleshed melons rank high in ascorbic acid and carotenoid content.

Cucumbers

Cucumbers (*Cucumis sativus* L.) are prostrate, branched, stiffly hairy vines possessing sharply five-cornered leaves and unbranched tendrils. They are placed botanically in the same genus as melons only because the Swedish botanist Linné had done so. They are in fact a distantly related species, with cucumber having a diploid chromosome number of 14 ($2n = 14$). Cucumbers are monoecious, but many newer cultivars are gynoecious and several old cultivars are andromonoecious. The flowers are bright yellow and fairly small, approximately 4 cm in diameter.

Whilst the melons and their relatives are of African origin, the cucumber is native to the foothills of the Himalayas of Nepal, where the wild, bitter-fruited form, *C. sativus* var. *hardwickii* (Royle) Alefeld, can be found. Cucumbers have been in cultivation several thousand years and spread from their native habitat to neighboring eastern and southern Asia, and later to Europe and Africa. Cucumbers are almost invariably used when immature, and eaten raw in salads, but in eastern Asia, they are also cooked. Cucumber contains a large amount of diversity in fruit characteristics when compared with other species, but relatively little when compared with melons, pumpkins, and squash. There is some variation in mature fruit color, immature fruit color, uniformity of color, size, and number of fruit spines, and color of spines. Much variation can be found for the length but not width of mature fruits; thus, fruit size is essentially a function of fruit length. The fruits can range in length from approximately 10 cm in some forms to 60 cm or more in some Chinese cultivars. Mature fruit range in width from 4 to 8 cm.

Cucumbers can be classified into six main groups on the basis of variation in fruit characteristics and geographic origin. The French *cornichons* are short fruits with small, numerous spines and are usually grown for pickling. The European greenhouse cucumbers are long, smooth fruits that are grown for slicing into salads. The American pickling cucumbers are short fruits with protruding warts and large spines. The American slicing cucumbers are medium-long fruits having a thick skin and protruding warts that bear large spines. The Middle Eastern cucumbers are medium-short, spineless, and wartless, and are used for both slicing and pickling. The Far Eastern cucumbers are very long with protruding warts and spines. Interestingly, short-fruited kinds tend to be used for pickling, and long-fruited kinds are used fresh or, in the Orient, for cooking. Two additional groups of cultivars have been reported, one from southwestern China and the other from Nepal and high elevations of India. An interesting novelty is the 'Lemon' cultivar, which bears a striking resemblance to its namesake.

Cucumbers are an important greenhouse crop in Europe, the Middle East, and the Far East. Greenhouse cucumber cultigens differ from those grown in the field by being parthenocarpic, that is, they have the genetic ability to set fruit without pollination, and by being completely gynoecious, that is, they produce female flowers exclusively.

According to FAO statistics, world cucumber production exceeds 29 million tonnes, with China producing over half of that. Other important producing countries are Iran, Turkey, USA, Japan, and Russia. Cucumbers are an important part of the diet in these and other countries, but they are low in nutritional value.

Watermelons

Watermelons (*Citrullus lanatus* (Thunberg) Matsumura & Nakai) are prostrate, basally branched, softly hairy vines possessing deeply and roundly indented, blue–gray–green leaves and branched tendrils. They are monoecious, producing one flower per node. At most nodes, a male flower is produced but a female flower appears at every seventh or eighth node. The flowers are small, approximately 3 cm in diameter, and light yellow. *C. lanatus* is diploid, with 11 pairs of chromosomes ($2n = 22$).

Watermelons originate from tropical Africa south of the equator and are well adapted to warm, seasonally dry habitats. Watermelon culture reached Egypt no later than 3500 years ago. Watermelons require a long, warm growing season and are usually grown for the consumption of their fresh, edible fruit flesh. Unlike melons and pumpkins, the part consumed is the placental region, or endocarp. In some regions, consumption of the seeds is important, too.

There is a great variation in fruit size in watermelons, from approximately 2 kg to over 20 kg. Fruit shape can range from almost spherical to obovate. Fruit color is almost always green, but some rare forms have yellow fruits or bicolor, yellow and green, fruits. The watermelon fruit rind can be of a single green shade, such as black–green, dark green, or light green, or have two green shades as striping. However, the width of the dark and light stripes can vary with respect to one another, that is, the dark stripes can be

much broader, broader, about the same as, narrower, or much narrower than the light stripes. Fruit interior (placental) color can vary from light green to white to light yellow, intense yellow, orange, pink, red, or intense red. Most cultivars have red flesh. 'Crimson Sweet,' a cultivar bred at Kansas State University in the USA that was commercially released in 1966, is a leading cultivar possessing medium-size (8–11 kg), spherical-oval fruits having an attractive rind of broad, dark stripes and high-quality, very sweet, fine-grained red flesh with relatively few small brown seeds. The idea of triploid ($3n = 33$) seedless hybrids was proposed by Japanese and American geneticists in the mid-twentieth century. Seedless cultivars related to 'Crimson Sweet' have been developed, and these are becoming increasingly popular.

In some countries, most notably China, consumption of watermelon seeds is popular. Watermelon seeds can be black, brown, tan, white, red, or green, or a combination or pattern of these. Seed size and number per fruit also vary considerably. Cultivars grown for seed consumption usually have a distinctive seed-coat color (such as red or green) or pattern (such as tan ringed with black) and small fruits that are packed with many medium- to large-size seeds.

Watermelons have the greatest world production of any cucurbit, exceeding 63 million tonnes according to the FAO. Not surprisingly, China is by far the largest producer of watermelons, with over 38 million tonnes. However, this tonnage would not indicate relative economic importance, as other cucurbits, such as melons, summer squash, and cucumbers, have several times the monetary value per unit weight than do watermelons. Most of the other important producers are countries that have regions with hot, dry summers and include Turkey, Iran, USA, Egypt, Mexico, Kazakhstan, Spain, and Saudi Arabia. Watermelons are not high in nutritional content, but their flesh does contain a large amount of carotenoids.

Squash and Pumpkins

Squash and pumpkins belong to the genus *Cucurbita* and encompass four cultivated species. Plants are large, usually viney and branched, with branched tendrils. All are monoecious, bearing showy, large, 10–16-cm-diameter, bright yellow to orange flowers. Usually, one flower is produced per node, with the first nodes producing male flowers exclusively. The species of *Cucurbita* are diploids, all with 20 pairs of chromosomes ($2n = 40$). Fruit size, shape, and color vary greatly among cultigens. The species differ from one another in the aspect of the foliage and shape of the leaves. Edible fruits that are more or less round in shape are usually referred to as pumpkins, and those that are not round in shape are usually referred to as squash. Unpalatable or inedible fruits are referred to as gourds.

The most widely grown species is *Cucurbita pepo* L. This species is native to warm-temperate to temperate regions of North America and was one of the first American plants to have been described and depicted by Europeans. The plants of this species can be viney or bushy, and leaf shape, while generally pentagonal, can vary from not incised to deeply incised. The foliage is spiculate, a characteristic serving to distinguish *C. pepo* from other species of the genus. *C. pepo* is well adapted to all temperate regions and is grown mostly for the use of its immature fruits, also known as summer squash. In cool regions, this species is also grown extensively for the culinary use of the flesh of its mature fruits, for consumption of the seeds or seed oil, and for ornaments. *C. pepo* is perhaps the most diverse species for fruit characteristics in the entire plant kingdom. Fruits range in size from less than 100 g to over 20 kg, ranging in shape from round to extremely long to flat, can be with or without ribs, grooves, furrows, wrinkles or warts, and can be green, yellow, or orange ranging in intensity and shading from almost black to almost white and possess patterns of striping, which is longitudinal, or bicolor, which is latitudinal (see Figure 2). This species is considered to consist of three subspecies and 10 cultivar-groups, as summarized in Table 2. The familiar Halloween and pie pumpkins, the oil-seed pumpkins, the zucchini, cocozelle, vegetable marrow, scallop, crookneck, straightneck, and acorn squash, the unique

Figure 2 (see color plate 105) Squash and pumpkins.

Table 2 Classification of *Cucurbita pepo* (squash, pumpkin, gourd) to subspecies and cultivar-groups[a]

Subspecies Group	*Description*
pepo *(cultivated forms, wild forms unknown)*	
Pumpkin	Spherical, globose, oblate, oval, obovate
Vegetable marrow	Short, tapered cylindrical
Cocozelle	Long, bulbous cylindrical
Zucchini	Uniformly cylindrical
Pepo gourd	Round, smooth or warted
texana *(Scheele) Filov (wild and cultivated forms)*	
Acorn	Top-shaped, furrowed
Scallop	Flat with scalloped edge
Crookneck	Long, narrow neck
Straightneck	Short, thick neck
Ovifera gourd	Oviform or pyriform, smooth
fraterna *(Bailey) Andres (wild forms only)*	

[a]From Paris HS (2000) History of the cultivar-groups of *Cucurbita pepo*. *Horticultural Reviews* 25: 71–170.

'Delicata' and 'Rolet,' as well as the small, fancy ornamental gourds are all members of *C. pepo*.

Another widely grown species is *Cucurbita maxima* Duchesne. Like *C. pepo*, this species is well adapted to warm-temperate and temperate regions. *C. maxima* is native to warm-temperate regions of South America, and the plants are usually viney, infrequently bushy, and with round stems. The leaf blades are more or less pentagonal but are also rounded. The foliage is stiffly hairy. In South America, *C. maxima* is grown for its immature and mature fruits, but elsewhere, it is grown almost exclusively for the consumption of the mature fruit flesh and, in parts of Europe, for consumption of the seeds. There is also a great deal of variation in fruit characteristics. *C. maxima* is well named, as its fruits are the largest of the plant kingdom, and new records for size are established every several years. *C. maxima* pumpkins exceeding 450 kg have been grown. Many of the small-fruited kinds are rated high in quality. Among the familiar named cultigens of *C. maxima* are 'Atlantic Giant,' 'Buttercup,' 'Delica,' 'Delicious,' 'Gold Nugget,' 'Hubbard,' 'King of the Mammoths,' 'Lumina,' 'Queensland Blue,' 'Rouge Vif d'Etampes,' and 'Turk's Turban.'

Yet another widely grown species is *Cucurbita moschata* Duchesne. Unlike *C. pepo* and *C. maxima, C. moschata* is best adapted to tropical climates. The nativity for this species has not yet been confirmed, but apparently, it is in tropical regions of northern South America. The plants of *C. moschata* are large, spreading vines, with angular stems. The leaf blades are pentagonal and cornered, but not incised. The foliage is softly hairy. *C. moschata* is usually grown for consumption of its mature fruits. This species is an important source of nutrition, especially of provitamin A, to peoples of the tropics, particularly in northern South America, Central America, and the Caribbean Islands. A great deal of variation in fruit characteristics can be found in northern South America. Although the fruits of *C. moschata* are generally referred to as tropical pumpkins, this species does include the 'Cheese' pumpkin and 'Butternut' squash, which are grown in temperate climates.

The fourth species of squash and pumpkins is *Cucurbita argyrosperma* Huber, a species that is native to warm lowland regions of Mexico. Closely related to *C. moschata*, this species also has large, spreading vines, angular stems and cornered leaf blades but tends to be less hairy. *C. argyrosperma* is not as widely cultivated as *C. moschata*, but it is important in its native country, more for consumption of its seeds than for its fruit flesh. One cultigen, 'Green Striped Cushaw,' has been grown for centuries in the USA for consumption of its mature fruit flesh.

A fifth cultivated species of the genus is *Cucurbita ficifolia* Bouché. This species is commonly known as the fig-leaf gourd rather than as a squash or pumpkin and has less variation in fruit characteristics than the other cultivated species of the genus. Its fruits usually weigh several kilograms, are round to oval or nearly so, and are mottled green and white. The white, fibrous fruit flesh is most often used for making sweets or preserves. *C. ficifolia* also has large, spreading vines with leaves that usually are similar in shape to those of fig leaves and is grown in the cool, highland regions in low latitudes of the Americas. Although adapted to cooler temperatures, it is day-length-sensitive and does not flower in the long days of summer of the mid-latitude regions.

FAO statistics list pumpkins, squash, and gourds together. As gourds can include members of *Cucurbita* as well as other genera, most notably the bottle gourd (*Lagenaria*), it is difficult to give an accurate estimate of production of *Cucurbita*. The production figures from some large producers of pumpkins and squash, notably the USA, Austria, and Guatemala, are not included in the FAO statistics. According to the FAO, world production of pumpkins, squash, and gourds was over 15 million tonnes in 2000. The leading producers, with approximately one-fifth of the total for each, are China and India, followed by Ukraine, Egypt, Iran, Italy, Mexico, Argentina, Turkey, South Africa, and Spain. Pumpkins and winter squash provide an important source of sustenance in some of the poorer regions of the world. Of the cucurbits, and indeed of vegetable crops generally, pumpkins and squash rank high overall in nutritional value, most familiarly in the carotenoid content of the fruit flesh.

Other Cucurbits

Bottle gourd, *Lagenaria siceraria* (Molina) Standley, has large vines bearing large (approximately 10 cm wide), white flowers that open at night. Their leaves are softly hairy, and large and rounded. This species contains a tremendous variety of fruit shapes and sizes, the result of human selection in different cultures of various regions for a variety of purposes. Young fruits are hairy and usually light green but become glabrous and beige or brown at maturity. The moist, white flesh of the immature fruit dries out as the fruit matures. The young fruits are used for culinary purposes, much like summer squash, in Italy (where they are known as cocuzzi), China, and India. They are also known as calabash gourd. The mature bottle gourds have a variety of other uses, too, first and foremost as containers.

Wax gourd, *Benincasa hispida* (Thunberg) Cogniaux, is important in China and other countries of eastern and southern Asia. It bears some similarities to *Cucurbita* by having long, sturdy vines, large leaves, large yellow flowers, and large fruits. However, the stamens of its flowers are free rather than fused, and the mature fruits are covered with a waxy cuticle. The young fruits are hairy. Fruit color is green, with shapes ranging from round, pumpkin-like to cylindrical, and the flesh is white and crisp. Both the immature and mature fruits are used for culinary purposes. The young fruits are usually cooked or pickled but sometimes eaten raw, and are sometimes referred to as Chinese squash. The mature fruit is sliced or cut into chunks for cooking in soup and is also known as Chinese winter melon.

Sponge gourd, *Luffa cylindrica* (L.) M.J. Roemer is a fairly large vine bearing large, 8–10 cm wide, bright yellow flowers. The cylindrical, green fruits are harvested when young, approximately 10 cm long, and usually boiled, but can achieve a length of 40 cm or so at maturation. They are also called smooth loofah or dishcloth gourd. The endocarp dries out as the fruit matures, forming a sponge that has a wide variety of uses. This fibrous net is the dried remains of the highly anastomosing vasculature of the fruit mesocarp. A related species, the angled loofah (*Luffa acutangula*) is similarly used and referred to as Chinese okra.

Bitter melon, *Momordica charantia* L., also known as Balsam pear and bitter gourd, is a relatively small, glabrous or nearly glabrous vine. The elongate fruits average 15–20 cm in length and several centimeters in diameter and are eaten prior to maturity, as the mature fruits are bitter. The seeds are covered with sweet, red arils, but the seeds themselves are poisonous. This species is grown mostly in southern Asia. It has had considerable medical interest as a treatment for a variety of ailments, including HIV.

Chayote, *Sechium edule* (Jacquin) Swartz, is a perennial vine native to Mexico and Guatemala. Chayote is grown almost entirely in subtropical and tropical areas and does not flower in the long days of summer of the middle latitudes. The green fruits are round to pyriform, averaging 14 cm long, and contain a single seed. The fruits are harvested when immature and cooked.

Snake gourd, *Trichosanthes cucumerina* L., is a hairy vine bearing distinct fringed white flowers that open at night. The mature fruits are up to 150 cm long, very slender, and curved. The immature fruit has distinct broad intense green and narrow pale stripes. As the mature fruit is fibrous and can be bitter, usually only the immature fruits are used for culinary purposes.

Food Value

As described in the previous sections, cucurbit fruits are often consumed when young and immature. Cucumbers, for example, are usually consumed no more than one week after anthesis. Immature cucurbit fruits are generally very high in water content; young cucumbers may contain 96% water and therefore are considered to have little nutritional value on a volume basis (Table 3).

The mature fruits of many cultivated cucurbits are inedible or palatably undesirable for a variety of reasons, including the accumulation of bitter compounds in the mature fruit flesh. However, the mature fruits of some cucurbit species are highly appreciated and have immense worldwide economic value. The cucurbit fruits that are consumed mature can be divided into two categories: those that are enjoyed uncooked for their sweet flesh (melons, watermelons) and those that are enjoyed after being cooked in some manner, such as boiling, stewing, roasting, or baking (pumpkins, winter squash, wax gourd). The water content of the mature fruit is lower than that of the immature fruit, and the different cucurbit fruit are specialized in various types of metabolite accumulation during maturity. For example, melon and watermelon fruit are characterized by soluble sugar accumulation during the final stages of fruit development. In the case of melon, which has been studied most extensively, this sugar accumulation is due almost exclusively to the storage of sucrose, which is practically absent in the immature fruit and can increase dramatically during the final weeks of fruit development to 5% on a fresh weight basis, or 60% on a dry weight basis, of the fruit flesh. This storage of sucrose is determined by developmental changes in

Table 3 Nutrient composition of cucurbits in 100-g edible raw portion

Crop	Water (%)	Calories (kcal)	Protein (g)	Fat (g)	Carbohydrate (g)	Fiber (g)
Melon						
Cantaloupe	90	35	0.9	0.3	8.4	0.8
Casaba	92	26	0.9	0.1	6.2	0.8
Honeydew	90	35	0.5	0.1	9.2	0.6
Winter squash						
Acorn	88	40	0.8	0.1	10.4	1.5
Butternut	86	45	1.0	0.1	11.7	
Hubbard	88	40	2.0	0.5	8.7	
Spaghetti	92	31	0.6	0.6	6.9	
Summer squash	94	20	1.2	0.2	4.4	1.9
Cucumber	96	13	0.7	0.1	2.8	0.8
Watermelon	92	32	0.6	0.4	7.2	0.5
Bitter gourd (Momordica)	94	17	1.0	0.2	3.7	2.8
Wax gourd (Benincasa)	96	13	0.4	0.2	3.0	2.9
Luffa gourd (Luffa)	94	20	1.2	0.2	4.4	
Calabash gourd (Lagenaria)	96	14	0.6	0.02	3.4	
Chayote	94	19	0.8	0.1	4.5	1.7
Pumpkin and squash, seeds	7	541	24.5	45.9	17.8	3.9
Watermelon, seeds	5	557	28.3	47.4	15.3	

Data obtained from the US Department of Agriculture, Agricultural Research Service (2001) *USDA Nutrient Database for Standard Reference, Release 14*. Nutrient Data Laboratory Home (www.nal.usda.gov/fnic/foodcomp).

carbohydrate metabolism in the fruit mesocarp, which is under strict genetic control. There is a large amount of genetic variability among melon cultigens for the trait of sucrose accumulation, and most primitive varieties do not accumulate high levels of sugar. The genetic trait of sweetness in melon is recessive and was selected for during the later domestication of the species.

Mature pumpkins and squash generally store significant amounts of starch, which accumulates during the later stages of fruit development. The starch content is correlated with fruit sensory textural attributes, and the soluble sugar levels in the mature fruit determine the fruit sweetness. Therefore, the balance between starch and soluble sugar plays a major role in determining fruit quality. During storage of the mature fruit, starch is metabolized, and the soluble sugar levels often increase, contributing to the flavor of the cooked fruit.

In addition to the soluble sugar and starch components, the organic acid levels of cucurbit fruit can also contribute to fruit taste, although this subject has not attracted much attention to date. Cucumber fruit accumulate high levels (approximately 1% on a fresh weight basis) of citric acid during the later stages of fruit development, but acid levels are low in immature cucumber fruit. Melon fruit have low acid levels, and their sensory quality is therefore determined largely by their soluble sugar content. However, the fruit of many of the nonsweet melon types accumulate high levels of organic acids during their development, and this genetic trait can be combined with the genetic trait of high sugar content, leading to melon varieties with novel tastes.

With respect to nutritional value, cucurbit fruits are generally not ranked high, as they are fleshy with a high water content. The immature fruits have lower levels of nutritional components than the mature fruits. However, some mature cucurbit fruits contain high levels of vitamins C and A. Of the cucurbit fruit, the melons have the highest vitamin C levels and a 1-cup serving of sweet melon (236 g) can meet the recommended daily requirement for vitamin C. The orange-fleshed melons as well as the orange-fleshed pumpkins and squash have high carotenoid levels and a high vitamin A activity. However, the green-fleshed honeydew melons have low carotenoid levels and, hence, a lower vitamin A activity. The major carotenoids of the orange-flesh fruit are α- and β-carotene and the xanthophyls, lutein and violaxanthin. The red color of the watermelon fruit is due to the carotenoid lycopene, which is also responsible for the characteristic red color of tomato fruit. A daily serving of the orange-fleshed melon meets the recommended daily requirement for provitamin A, making melon the only significantly consumed fruit in the USA that can provide the RDA of both provitamin A and vitamin C. There exists a large amount of genetically determined variation for carotenoid levels in the cucurbit fruit. In the case of *Cucurbita pepo*, which has been studied most extensively, over a dozen genes have been characterized, which affect fruit color ranging from pale cream to intense orange.

Besides provitamin A, vitamin C and carotenoids, cucurbit fruits can contain low levels of additional antioxidants, such as flavonoids, but the total antioxidant activity of the various cucurbit fruit remains to be studied. Cucurbit fruit also contain numerous bioactive compounds, including vitamin E, biotin, and folic acid, and here too, there is no doubt a tremendous amount of genetic variability in the family waiting to be discovered. *Momordica* fruit has high levels of both vitamin C and folic acid.

The cucurbits, especially wild germplasm, frequently contain high levels of the cucurbitacin terpenoids. These are highly bitter compounds, which may also be toxic, and they play roles in herbal and folk medicine. They are found at high concentrations in the colocynth gourd (*Citrullus colocynthis* (L.) Schrader), which is likely the poisonous fruit mentioned in Kings 2:4 of the Bible. Bitterness in the cucurbits is genetically controlled and was selected against early in plant domestication. Nevertheless, the problem of undesirable bitter fruit in melon, squash, and cucumber still occurs.

The fruit of exotic cucurbits may be sources of important pharmacological and bioactive compounds, and these are attracting increasing research attention. For example, the mature fruit of *Siraitia grosvenorii* (Swingle) Lu & Zhang (common name luo-han-guo) contains an unusual terpenoid glycoside, mogroside, which is over 100 times sweeter than sucrose. Bitter gourd (*Momordica charantia*) and snake melon (*Trichosanthes cucumerina*) are just some of the cucurbits that are being tested for potential medical benefits.

Seeds of many cucurbits, such as pumpkin, squash, and watermelon, are consumed largely as snack foods. The seeds generally have high levels of oils and proteins, and may be considered a highly nutritious food. The oils contain mainly unsaturated fatty acids, primarily oleic and linoleic and, here too, there is much genetic variation in fatty acid composition within the family. For example, the seeds of the white-seeded melon have primarily linolenic acid (18:3), and lines of buffalo gourd (*Cucurbita foetidissima* HBK) have been developed, with seeds containing over 80% linoleic acid (18:2). One of the interesting advances in cucurbit seed improvement has been the utilization of the naked seed, or hull-less, mutant in *Cucurbita pepo*. In this mutant, discovered a little over 100 years ago, the seed coat does not develop, thus allowing for the efficient extraction and processing of the oils. Oil extracted from *C. pepo* seeds is highly prized for use on salads and in cooking in much of central and eastern Europe.

Other Uses

In addition to their uses in edible consumption, many cucurbit fruits have also served other utilitarian purposes throughout history. Collectively described as gourds, the dried lignified fruit rinds of numerous cucurbit species have been used primarily as containers of various sizes, and also as fish-net floats, masks, musical instruments, birdhouses and other uses of anthropological interest. Species grown for their dried fruit are primarily from the *Lagenaria* and *Cucurbita* genera. As previously mentioned, *Luffa* gourds are grown for the fibrous interior of the dried fruit, which can serve as a sponge.

Fruit Ripening

Melon cultigens can be divided into two major groups based on their ripening physiology. The *reticulatus* and *cantalupensis* groups are climacteric, meaning that there is a characteristic sharp and rapid peak of respiratory activity during ripening. This is accompanied by the development of an abscission layer between the pedicel and the fruit (termed 'full-slip'). The nonclimacteric melon fruit of the *inodorus* group, such as honeydew and casaba melons, do not show a sharp respiratory peak, and their development generally extends over a longer period of time. They do not develop an abscision layer, and the appropriate stage of ripeness is determined by changes in skin color, from white–green to creamy yellow. In all melons, sugar is accumulated until the fruit naturally abcisses or is removed from the vine, and therefore the tastiest melons are those that are not prematurely harvested. Since no storage starch accumulates in melon fruit, in contrast to the winter squash and pumpkins, there is no postharvest increase in sugar content or sweetness.

The netted climacteric melons have a shorter storage life than the nonclimacteric smooth-skinned melons. However, the traits of rind netting and respiration physiology are unlinked, and recently, new varieties of netted melons with nonclimacteric respiration have been developed, revolutionizing the harvesting and shipping technologies of netted melons due to their long storage life. The climacteric trait in melons is controlled by only two genes, and the breeding of the new types of melon is therefore relatively straightforward.

The cucurbit fruit that are consumed while immature, such as cucumber and summer squash, have very short storage lives due to their undeveloped skin and rind. Ideally, storage of immature fruit should be under a relatively high relative humidity, 85–95%, to reduce fruit water loss. At the other extreme with

respect to storage, pumpkins and winter squash may be stored for a long time, frequently many months, and during this period, the fruit undergoes 'curing,' part of which is the transition of the starch in the stored fruit to sugar. For the long-term storage of winter squash and pumpkins, the optimal relative humidity is generally 70–75%.

Future Research and Development

The tremendous genetic variability for fruit quality components in the cucurbits remains to be characterized and harnessed for the breeding and production of improved melon, squash, cucumber, pumpkin, watermelon, and other cucurbit fruit. Within each species of cucurbit, there exist genotypes, at times primitive landraces, that are extraordinary in their levels of a particular quality component or bioactive compound. This is true for sugar content, acid content, volatile productions, nutritional content, ripening physiology and other metabolically controlled processes in the developing fruit. Future research will survey and uncover these metabolic variations, and their genetic control will be determined, allowing for their efficient utilization. This will set the stage for the production of new and novel varieties with beneficial characterisitics, be it improved taste or nutrition. This effort will be aided by advances in understanding the molecular-genetic control of these metabolic processes. The unfolding of the cucurbit genome, including the cataloging of genes involved in plant metabolism, will help determine the potential for the genetic improvement of each of the cucurbit crops.

See also: **Antioxidants**: Natural Antioxidants; **Ascorbic Acid**: Properties and Determination; **Carotenoids**: Occurrence, Properties, and Determination; **Ripening of Fruit**; **Starch**: Structure, Properties, and Determination; **Storage Stability**: Mechanisms of Degradation

Further Reading

Corrigan VK, Irving DE and Potter JF (2000) Sugars and sweetness in buttercup squash. *Food Quality and Preference* 11: 313–322.

Heiser CB (1979) *The Gourd Book*. Norman, OK: University of Oklahoma Press.

Kirkbride JH, Jr. (1993) *Biosystematic Monograph of the Genus* Cucumis *(Cucurbitaceae)*. Boone, NC: Parkway.

Lester G (1997) Melon (*Cucumis melo* L.) fruit nutritional quality and health functionality. *HortTechnology* 7: 222–227.

Paris HS (2000) History of the cultivar-groups of *Cucurbita pepo*. *Horticultural Reviews* 25: 71–170.

Pitrat M, Hanelt P and Hammer K (2000) Some comments on infraspecific classification of cultivars of melon. In: Katzir N and Paris HS (eds) *Proceedings of Cucurbitaceae 2000: the 7th Eucarpia Meeting on Cucurbit Genetics and Breeding. Acta Horticulturae* 510: 29–36.

Robinson RW and Decker-Walters DS (1997) *Cucurbits*. Wallingford, UK: CAB International.

Schaffer AA, Pharr DM and Madore M (1996) Cucurbits. In: Zamski E and Schaffer AA (eds) *Photoassimilate Distribution in Plants and Crops: Source-Sink Relationships*, pp. 729–758. New York: Marcel Dekker.

Seymour GB and McGlasson WB (1993) Melons. In: Seymour GB, Taylor JE and Tucker GA (eds) *Biochemistry of Fruit Ripening*, pp. 274–290. London: Chapman & Hall.

Wang H, Cao G and Prior RL (1996) Total antioxidant capacity of fruits. *Journal of Agricultural and Food Chemistry* 44: 701–705.

Whitaker TW and Davis GN (1962) *Cucurbits*. New York: Interscience.

Yamaguchi M (1983) Cucurbits. In: *World Vegetables*, pp. 312–349. New York: AVI.

MEMBRANE TECHNIQUES

Contents

Principles of Reverse Osmosis

H N Lazarides and E Katsanidis, Aristotelian University of Thessaloniki, Hellas, Thessaloniki, Greece

Introduction

Vast amounts of liquid food are industrially concentrated for preservation purposes or in order to reduce storage, packaging, handling, and transportation costs. Vacuum evaporation is the predominant method used by the food industry to produce liquid food concentrates, despite serious drawbacks such as poor product quality due to thermal degradation and high energy demand for the evaporation process. Reverse osmosis was first suggested by Sourirajan in 1959 as a method for water desalination. Since then, significant developments in both material science and membrane technology have allowed reverse osmosis to be used in the food industry, mainly for concentrating liquid foods without the disadvantages of vacuum evaporation.

Theory

A membrane is a permeable or semipermeable material (polymer, inorganic, or metal), which restricts the motion of certain species through itself. The membrane, acting as a barrier, controls the relative rates of through transport of various species and thus gives one product depleted in certain components and a second product concentrated in these components.

Membranes are used in many filtration and separation processes. Some representative processes and relative particle sizes are shown in Table 1.

When a membrane separates two solutions of different solute concentrations, water moves through the membrane towards the solution with the high concentration of solutes tending to equalize the solute concentration on the two sides of the membrane. The driving force for the movement of water is the difference in osmotic pressures of the two solutions. This phenomenon is called osmosis.

The osmotic pressure of a dilute solution can be calculated by Van't Hoff's equation:

$$\Pi = cRT/\mathrm{M}$$

where Π is the osmotic pressure (Pa), c is the solute concentration (Kg/m^3), R is the gas constant (J/mol.K), T is the absolute temperature (K), and M is the molecular weight of the solute. For a wider range of concentrations, Gibb's equation has been found to be more suitable for the calculation of the osmotic pressure:

$$\Pi = -(RT \ln x_\alpha)/V_m$$

where x_α is the mole fraction of pure liquid and V_m is the molar volume of pure liquid. When external (hydrostatic) pressure greater than the osmotic pressure is applied, water is expelled from the concentrated solution and we observe a reverse flow, that is a flow against the osmotic pressure gradient. This process is called reverse osmosis. Reverse osmosis involves the

Table 1 Membrane separation processes

Separation process	*Particle size (microns)*	*Applications*
Microfiltration	0.1–10	Clarification, sterile filtration
Ultrafiltration	0.01–0.1	Separation of macromolecular solutions
Nanofiltration	0.001–0.01	Separation of small organic compounds and selected salts from solutions
Reverse osmosis (hyperfiltration)	0.0001–0.001	Separation of microsolutes and salts from solutions
Dialysis	0.0001–0.001	Separation of microsolutes and salts from macromolecular solutions
Electrodialysis	0.0001–0.001	Separation of ions from water and nonionic solutes
Gas permeation	0.00001–0.0001	Separation of gas mixtures

most 'tight' membranes, which are capable of separating the smallest solutes. The pores of the membrane are so small that, in some materials, they can hardly be resolved, even by scanning electron microscopy.

The basis for separation is molecular size, but other factors such as molecular shape and charge can also play an important role.

Membranes currently used in commercial reverse osmosis installations are asymmetric, flat sheet membranes of cellulose acetate (CA) or cellulose triacetate, fine hollow fibers of aromatic polyamides (PA) or cellulose triacetate and thin-film composites (TFC). In TFC membranes an extremely fine layer of highly hydrophilic polymer has been placed on a microporous support usually made from polysulfone. Figure 1 depicts the structure of a typical CA membrane and shows the chemical affinity (selectivity) of the membrane for water molecules. The chemically induced negative sorption (repulsion) of the ions in solution and the preferential sorption of pure water to the membrane result in rejection of ions, thus preventing them from permeating the membrane. It should be noted that very-low-molecular-weight organics and uncharged solutes may pass through the membrane. In general, reverse osmosis membrane efficiency is expressed as a percentage rejection of NaCl, and the typical range for such membranes is 96–99%.

Operational Requirements

In conventional filtration processes, all the material to be filtered flows perpendicularly through the filter where the separation occurs (dead-end flow). Suspended solids captured by the filter tend to build up, slowing down the filtration rate and requiring frequent cleaning.

In reverse osmosis, cross-flow systems are used almost exclusively. In a cross-flow system, the feed

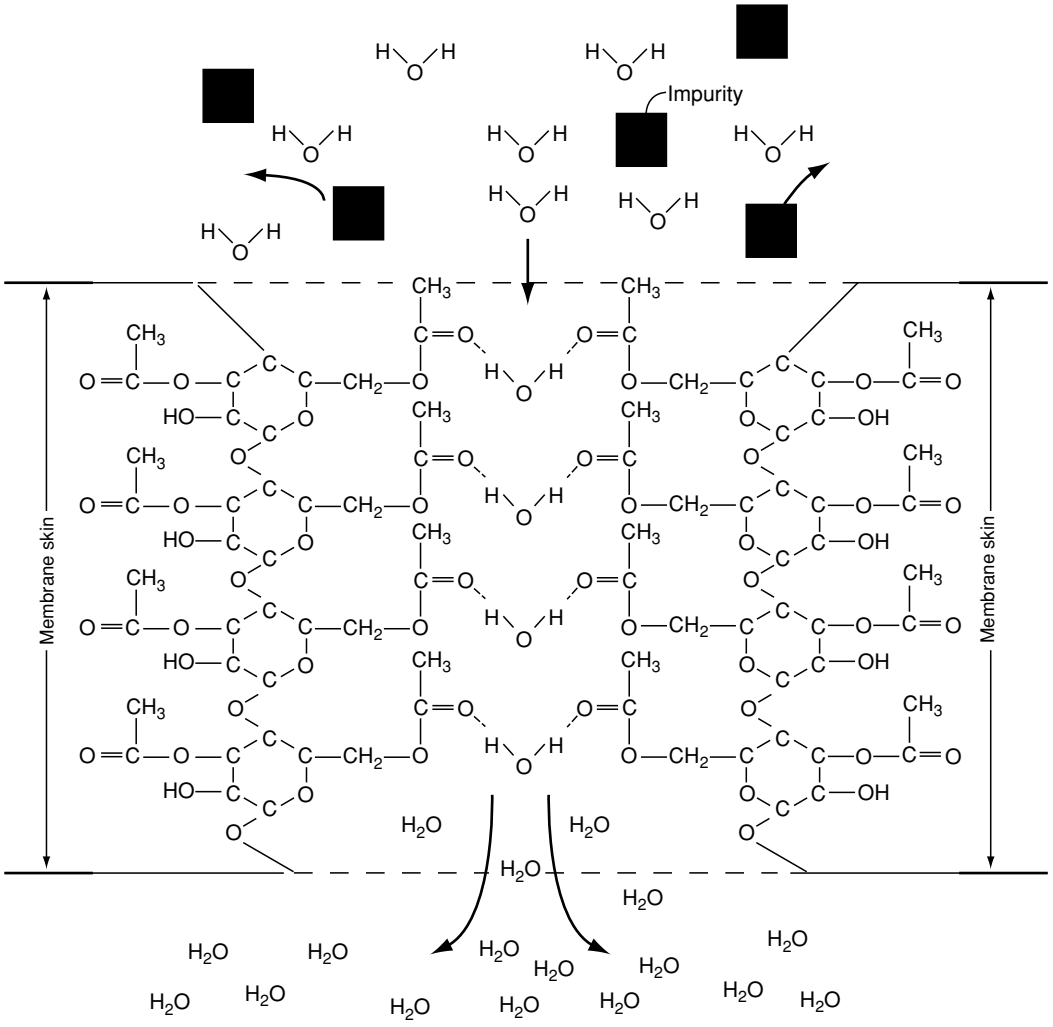

Figure 1 Movement of water through a cellulose acetate membrane. Adapted from Lacey (1972).

stream (influent) flows parallel (tangentially) to the membrane and only a portion of the stream passes through the membrane as permeate or filtrate. The remainder of the stream is called retentate or concentrate. The main advantage of the cross-flow system is that the feed stream constantly removes the accumulated solids (cake) or solutes from the membrane surface, reducing build-up, which allows for less frequent clean-up of the membrane system. The two types of flow and their representative flux and cake build-up are depicted in Figure 2.

In reverse osmosis, the external pressure has to exceed the osmotic pressure of the solution so that solvent (usually water) may be expelled. For example, the osmotic pressure of sea water is approximately 2.3 MPa and a desalination plant utilizing reverse osmosis may operate at pressures of up to 7 MPa. This makes it clear that the reverse osmosis systems must be very robust from a physical standpoint and the membranes must be able to resist mechanical compression (compaction), which can deform their morphology and negatively affect their performance.

Resistance to mechanical pressure is only one of the properties that a good membrane must have. Membranes need to function in a variety of environments, where the operating pressure, temperature, pH, and chemical compatibility need to be taken into account. Cellulose acetate membranes are usually used at temperatures up to 40 °C, but they may be operated at temperatures up to 65 °C.

The typical pH range for CA membranes is 2–8 and for PA 3–11. Membranes may be used outside the above ranges, but their useful life will be shorter.

A good reverse osmosis membrane needs to withstand chemical agents such as chlorine, iodine, other oxidizing agents, oils, and solvents. CA has good chlorine resistance (for sanitation purposes) but TFC PA has poor chlorine resistance.

Several reverse osmosis membranes used in commercial applications and their properties are presented in Table 2.

Membrane Configurations

There are several configurations for the containment and support of the membranes. These configurations need to address all the requirements discussed previously and also need to allow for easy access to the membranes for replacement. There are four major configurations that have been used in the food industry.

Spiral Wound

This module consists of a flat sheet membrane, a support system, and a spacer system, all wrapped around a perforated premeate-collection tube. The feed stream flows across the membrane through the spacer and the permeate is collected in the central permeate-collection tube. Specially designed spacer systems enhance turbulent flow, which helps reduce fouling of the membrane. The main advantages of this

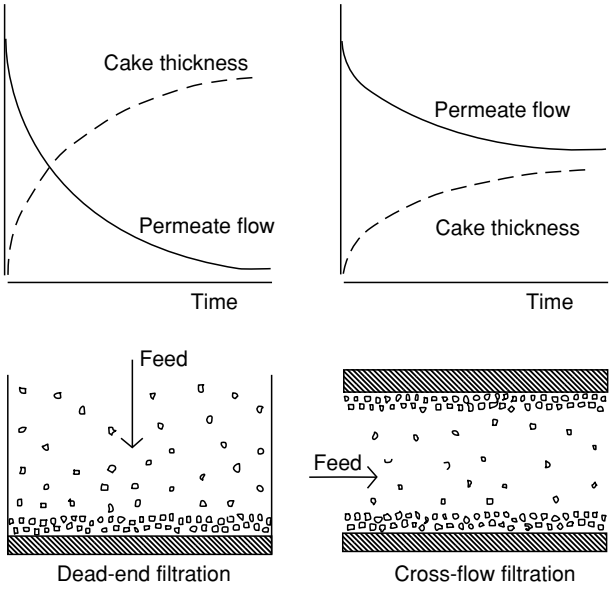

Figure 2 Dead-end vs cross-flow filtration. Adapted from Mannapperuma JD (1997) Design and performance evaluation of membrane systems. In: Rotstein E, Singh RP and Valentas KJ (eds) *Handbook of Food Engineering Practice*, pp. 168–208. Boca Raton, FL: CRC Press LLC.

Table 2 Properties of some commercial reverse osmosis membranes

Membrane type	*Module type*	*pH*	*Chlorine tolerance*	*Oxidation tolerance*
Cellulose acetate blend	Spiral wound	3–8	Fair	Fair
Cellulose triacetate	Capillary fiber	4–9	Fair	Good
Aromatic polyamide	Hollow fiber	4–11	Poor	Fair
Cross-linked polyether TFC	Spiral wound	1–12	Poor	Fair
Aryl-alkyl polyetherurea TFC	Spiral wound	3.5–12	Poor	Fair
Cross-linked fully aromatic polyamide TFC	Spiral wound	1–12	Poor	Fair

TFC, thin-film composites.
Adapted from Fell CJD (1995) Reverse osmosis. In: Noble RD and Stern SA (eds) *Membrane Separations Technology*, pp. 113–144. Amsterdam, The Netherlands: Elsevier.

module are the high packing density, the high flux, and the fact that the flat sheet membrane can withstand higher pressures than the hollow fiber module.

Capillary (Hollow Fiber)

In this module, the feed stream is pumped into a vessel that contains a bundle of thin hollow fibers (about 40 μm internal diameter). The permeate goes through the thin fibers and is separated from the concentrate. The advantage of this system is that it provides a very large membrane area for the total size of the system (high packing density). The disadvantages are the low flux and the need for pretreatment of the feed solution (i.e., filtration, centrifugation, microfitration, even ultrafiltration).

Plate and Frame (Flat Sheet)

This module consists of flat membrane sheets placed between support frames that create flow channels (about 0.5 mm) for the feed stream, the concentrate, and the permeate. The membrane and its supports are stacked together in large numbers (up to 360 membrane sheets) to create one unit. The main advantages of this module are that it is possible to remove and clean the membrane sheets and that the system is reusable (once the membranes have been cleaned or replaced). The disadvantages are the high labor and time investment required to disassemble and reassemble the module and the low packing density of the unit.

Tubular

This module is relatively uncommon in reverse osmosis. The feed and concentrate streams flow through the inside of the tube (up to 12 mm internal diameter) and the filtrate passes through the membrane and is collected in the housing around the tube. The advantages of this system are the ease of cleaning and the capability to handle high viscosities. The disadvantages are the relatively small membrane area for the size of the system and the high feed rates that are required for the system to operate efficiently (high energy/pumping cost).

The spiral wound and the hollow fiber are the two modules that are mostly used in the food industry. Figure 3 illustrates these module designs. In industrial systems, usually several units are assembled together in series and/or in parallel configuration to achieve the desired results. These systems may be single-pass systems or they may include recirculation loops and pumps so that part of the feed solution passes through the membrane multiple times.

Performance Comparisons

Reverse osmosis is a very efficient process, allowing for the simultaneous concentration, fractionation, and purification of the product and the accomplishment of multiple tasks in a single unit operation. It does not impart pH or chemical changes in the product and, since no significant heating is required, there is no heat degradation of the product. Thus, reverse osmosis has a minimal effect on the quality characteristics and nutritional value of the finished product, especially when compared to evaporative concentration where inevitably there is heat degradation, as well as flavor and nutritional losses.

Since reverse osmosis systems do not require steam, evaporators, and condensers, they require much less floor space and equipment than evaporative systems. Due to their relative simplicity, reverse osmosis systems have shorter come-up and shutdown time. They are also more flexible and easier to modify or upgrade than conventional evaporative systems.

Reverse osmosis is very energy-efficient, because it typically operates at ambient temperature (no heating or cooling required); and, most importantly, there is no phase change requirement for water removal, as in evaporative processes. Overall, reverse osmosis systems require less energy than evaporative systems per unit of water removed from the product. Reverse osmosis systems require almost exclusively electric energy for pumping and recirculating, whereas evaporative systems require steam in addition to electric energy for pumping.

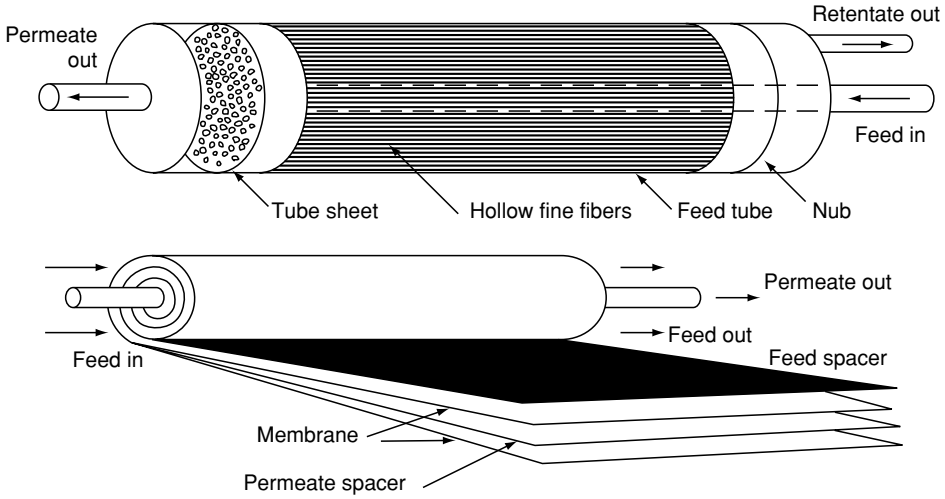

Figure 3 Hollow fine fiber (top) and spiral wound (bottom) module designs. Adapted from Mannapperuma (1997) Design and performance evaluation of membrane systems. In: Rotstein E, Singh RP and Valentas KJ (eds) *Handbook of Food Engineering Practice*, pp. 168–208. Boca Raton, FL: CRC Press LLC.

Table 3 Comparison of energy requirements of reverse osmosis and evaporation with either thermal vapor recompression (TVR) or mechanical vapor recompression (MVR)

Type of plant	*Direct energy input*		*Equivalent energy input (kWh kg$^{-1}H_2O$)*
	Steam (kg kg$^{-1}H_2O$)	*Power (kWh kg$^{-1}H_2O$)*	
Evaporation TVR 3 effects	0.20–0.25	0.003	0.17–0.20
Evaporation TVR 5 effects	0.13–0.17	0.003	0.11–0.14
Evaporation TVR 7 effects	0.077–0.13	0.003	0.07–0.10
Evaporation MVR	0.02	0.01–0.03	0.046–0.11
Reverse osmosis		0.005–0.012	0.015–0.048

In this table a heat-to-power ratio of 3 has been used.
Adapted from Hallström B (1988) Preconcentration: new developments. In: Bruin S (ed.) *Preconcentration and Drying of Food Materials*, pp. 46–47. New York, NY: Elsevier.

Thermal energy (steam) can be generated from oil at very high efficiency (80–90%), but electrical/mechanical energy is generated from oil at much lower efficiency (about 30%), yielding a heat-to-power ratio of 3.0. Table 3 compares the energy requirements of reverse osmosis to evaporation with either thermal vapor recompression or mechanical vapor recompression. The relative cost of electric energy to oil or natural gas energy varies a lot in different areas of the world, thus making it hard for generalized comparisons of energy costs of the two processes. However, it seems that in countries such as the USA, France, and Germany, reverse osmosis is much less costly than any multistage evaporative process, with either mechanical or thermal vapor recompression.

In large desalination plants, energy can be recovered from the rejected brine by running it through turbines to generate electricity that can be used in the plant. Such a recovery system can lead to energy savings of up to 30%.

Operational Issues

In membrane systems a reduction of flux over time is often encountered, while all operating parameters remain the same. This is typically due to membrane compaction, concentration polarization, and fouling.

Compaction

As discussed previously, the high operating pressures in reverse osmosis result in increased membrane density, decreased thickness, and deformed (compressed) membrane pores, which reduce the permeate flux through the membrane.

Concentration Polarization

In all cross-flow processes, as soon as the flow begins, a build-up of a polarized layer or a gel-like layer of solutes is observed at the membrane surface. This phenomenon can result in locally increased osmotic pressure in the vicinity of the membrane surface that

will negatively affect the flux through the membrane. Over time, the accumulated solutes can affect the affinity of the membrane for certain solutes or solvents. Concentration polarization is a reversible phenomenon that can be reduced by periodical changes in the velocity of the feed solution (tantalizing), by increasing the velocity, or by decreasing the pressure and concentration of the feed solution.

Fouling

Fouling refers to the irreversible changes that occur on the membrane and result in reduced flux over time. Many factors contribute to membrane fouling. Pore plugging and solute adsorption in the membrane pores can result in a decrease of the pore diameter and reduced flux. Solutes accumulated on the membrane surface due to concentration polarization may undergo irreversible changes over time and form a layer that will cause hydrodynamic resistance to flow. This layer formation depends on the membrane surface chemistry, solute–membrane and solute–solute interactions.

All the above make it clear that pretreatment of the feed stream is crucial in extending the useful membrane life. The objective of the pretreatment process would be to remove as many suspended solids and contaminants as possible before reverse osmosis. The pretreatment process may be a simple filtration or centrifugation step or it may be a much more thorough micro- or even ultrafiltration step. Pretreatment requirements will depend on the type of membrane used, the membrane configuration, and the composition and flow rate of the feed solution.

Cleaning

Mechanical cleaning can be applied in large-diameter tubular modules using oversized sponge balls to clean the tubes.

Hydraulic cleaning can be achieved by back-washing (reversing the flow), by back-shock treatment (back-flushing for very short time), or by the pulsation of flow (increasing and decreasing the pressure of the feed solution). Back-washing and back-shock treatment cannot be applied to TFC membranes because of their tendency to delaminate.

Electric cleaning uses a pulsed electric field resulting in the movement of charged particles away from the membrane. Electric cleaning can be applied without interrupting the process, but requires electricity-conducting membranes and special module design.

Chemical cleaning is the most important method for controlling fouling phenomena. Chemical compatibility between the cleaning agent and the membrane is the most important factor in selecting the proper cleaning agent. Acids (nitric, phosphoric, citric) are usually used for inorganic contaminants and basic (caustic) detergents or chlorine are typically used for organic contaminants (i.e., fat, protein). Other chemicals used may be complexing agents (ethylenediaminetetraacetic acid; (EDTA), enzymes, and disinfectants (H_2O_2, NaOCl).

Applications

Water desalination and/or purification is one of the major applications of reverse osmosis. Brackish water or sea water is treated by reverse osmosis to produce potable water. Also, ultrapure water for medical or industrial applications can be produced.

Reverse osmosis is used in the dairy industry for milk, whey, or lactose concentration. Concentration of whey, produced as a byproduct from the cheese-making process, is the most widely used application in the food industry. Spiral module configurations with CA or TFC membranes are most often used in the dairy industry.

Tubular and spiral module systems equipped with TFC membranes are used for juice concentration. Reverse osmosis is economically efficient for concentrations up to 24% solids, but it could be used to concentrate juices up to 60% solids, if economically feasible. Membranes developed for reverse osmosis have also been used in direct osmosis applications for the preconcentration of liquid foods (i.e., juices) with the use of high osmotic pressure brines on one side of the membrane. In these applications the driving force is not hydrostatic pressure (pressure are rather low; < 500 kPa) but the osmotic pressure difference on the two sides of the membrane.

A thorough discussion of reverse osmosis applications can be found in a following section.

See also: **Effluents from Food Processing**: On-Site Processing of Waste; **Evaporation**: Basic Principles; Uses in the Food Industry; **Filtration of Liquids**; **Membrane Techniques**: Applications of Reverse Osmosis; Principles of Ultrafiltration

Further Reading

Dziezak JD (1990) Membrane separation technology offers processors unlimited potential. *Food Technology* 44: 108–113.

Fell CJD (1995) Reverse osmosis. In: Noble RD and Stern SA (eds) *Membrane Separations Technology*, pp. 113–144. Amsterdam, The Netherlands: Elsevier.

Hallström B (1986) Energy consumption in membrane processing of foods. In: Singh PR (ed.) *Energy in Food Processing*, pp. 242–243. New York, NY: Elsevier.

Hallström B (1988) Preconcentration: new developments. In: Bruin S (ed.) *Preconcentration and Drying of Food Materials*, pp. 46–47. New York, NY: Elsevier.

Mannapperuma JD (1997) Design and performance evaluation of membrane systems. In: Rotstein E, Singh RP and Valentas KJ (eds) *Handbook of Food Engineering Practice*, pp. 168–208. Boca Raton, FL: CRC Press LLC.

Mulder MHV (1995) Polarization phenomena and membrane fouling. In: Noble RD and Stern SA (eds) *Membrane Separations Technology*, pp. 45–84. Amsterdam, The Netherlands: Elsevier.

Paulson DJ, Wilson RL and Spatz DD (1984) Crossflow membrane technology and its applications. *Food Technology* 38: 77–111.

Petrotos KB and Lazarides HN (2001) Osmotic concentration of liquid foods. *Journal of Food Engineering* 49: 201–206.

Scott K (1998) *Handbook of Industrial Membranes*, 2nd edn, pp. 3–17. Oxford, UK: Elsevier.

Singh RP and Heldman DR (1993) *Introduction to Food Engineering*, 2nd edn, pp. 385–407. San Diego, CA: Academic Press.

Applications of Reverse Osmosis

M H Abd El-Salam, National Research Centre, Cairo, Egypt

Introduction

Reverse osmosis (RO) is essentially a pressure-driven membrane diffusion process. In practice, RO membranes retain 95–99% of the dissolved solutes (organic and inorganic) from the feed stream into the concentrate, while the permeate can be considered as high-quality water. Therefore, RO is classified as a concentration process. RO has several advantages compared to other concentration technologies. RO is an energy-saving process as removal of the solvent does not require change in phases. Compared to other competing processes, RO is more economic in concentrating diluted solutions and for medium concentrations. Also, RO concentrated fluids are not subjected to any heat damage or losses in aroma compounds, as has occurred in other concentration processes. This is of utmost importance in concentrating liquid foods. Like other membrane processes, RO offers flexibility in different applications and scales, RO concentration is best explained as the mechanism of preferential sorption capillary flow. According to this mechanism, permeation occurs due to the preferential sorption of constituents from fluid mixture and their permeation through the porous membrane. For RO to take place, it is essential that a membrane has the right chemical nature (polar and nonpolar effects) and that its pores are of appropriate size and number (steric effect).

Although RO is mainly applied in the production of potable water from sea and brackish water, the use of RO in food processing is growing steadily. Until recently, the application of RO in the food industry has been based on the use of cellulose acetate membranes. During the 1980s a wide variation of thin-film composite membranes appeared on the market. This new generation of membranes gave food processors the opportunity to apply more rigorous cleaning routines and greater plant sanitation. Also, these membranes exhibit better recovery of valuable compounds in processed liquids, e.g., aroma compounds.

Reverse Osmosis in the Dairy Industry

RO was introduced into the dairy industry in the 1970s and rapidly expanded with time. It is estimated that about 75 000 m^2 surface area of RO membranes has been installed in the dairy industry until 1996 and is still increasing. Most of this area (76%) is used in concentrating whey, followed by permeate (20%) and milk (3%). A smaller surface area of RO membrane has been used to recover milk solids from rinse water in the dairy plant.

As lactose and minerals are almost completely retained by RO membranes, the osmotic pressure of the processed dairy liquid limits both the permeate flux and the maximum concentration achievable. Milk, whey, and permeate have nearly the same osmotic pressure, which is around 7 bar, a pressure that must be exceeded in RO processing of these fluids. On the other hand, milk and whey concentrated by RO to 25–28% total solids (maximum concentration) has an osmotic pressure of 27–35 bar. However, for practical and economic considerations, twofold RO concentration of whey and milk is usually carried out.

Concentration of whey by RO has several advantages:

- decrease in costs for transport and storage of whey
- as a preconcentration step in the manufacture of whey powder, demineralized whey powder, and whey cheeses
- improving the ultrafiltration performance of whey for the preparation of whey protein concentrates.

Figure 1 shows the use of RO in the manufacture of different whey products. One of the major concerns of dairy processors is the microbiological quality of the concentrate obtained when unpasteurized whey is processed. The microbiological quality of RO whey concentrates is controlled by the microbial load of

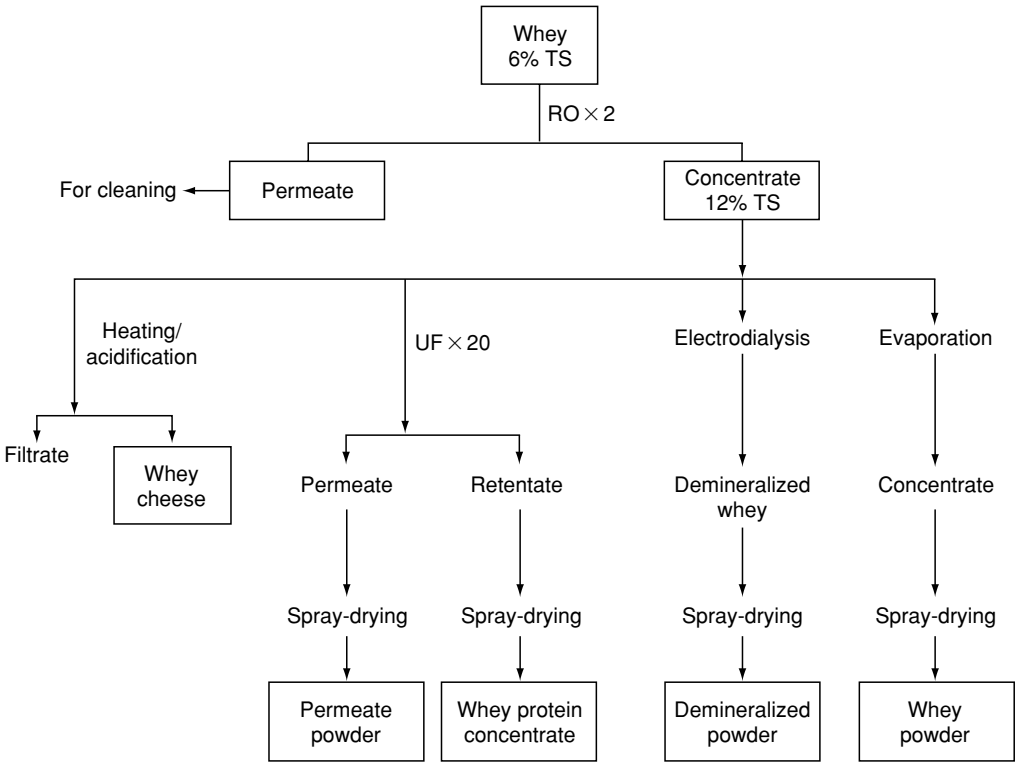

Figure 1 Application of reverse osmosis (RO) in the manufacture of different whey products. TS, total solids; UF, ultrafiltration.

the processed whey, RO operating temperature, and residence time in the RO equipment. However, the number of bacteria during RO of whey seldom multiplies by more than the concentration factor. When RO cellulose acetate membranes are used, an operating temperature of 10 °C has been recommended for control of the microbiological quality of RO whey concentrate, but when concentration is carried out with RO membranes from other synthetic polymers, higher temperatures can be used. However, RO concentration of whey must be carried out at a temperature < 55 °C to avoid precipitation of calcium phosphate and fouling of the membrane. (*See* **Membrane Techniques**: Principles of Reverse Osmosis.)

Concentrating permeate from the ultrafiltration of whey and milk by RO is widely practiced as the first step in the manufacture of lactose. RO cannot completely replace evaporation in lactose manufacture, as the concentrated permeate must contain 60% total solids in order to separate lactose crystals. This concentration cannot be achieved by RO processing alone. However, RO can be used for the preconcentration of permeate as a very economic way of removing the large bulk of the water from permeate. The concentration of ultrafiltration permeate to 15–18% total solids followed by evaporation up to 60% is considered an optimum combination for the manufacture of lactose. Even in already existing lactose manufacturing plants using evaporation only, the inclusion of RO can improve the efficiency of the plant.

The feasibility of concentrating milk by RO has been established in several studies. However, the industrial application of RO in concentrating milk is still limited. Whole milk can easily be concentrated by RO up to twofold. The presence of milk fat has a very limited effect on the permeation rate of whole milk compared to skim milk. The main objective of concentrating milk by RO is to reduce the storage and transport costs, which represent a significant part of the processing costs in some countries, such as Australia. Also, concentrating milk by RO increases the efficiency of processing facilities in the dairy plant and as a preconcentration step in the manufacture of dried milk. Generally all dairy products can be made from milk concentrated by RO with no marked effect on the quality of obtained products.

RO can also be used in the dairy industry for the recovery of waste solids from the dairy rinse water. This process can achieve two purposes: first, a decrease in the biochemical oxygen demand (BOD) of the plant effluent, which reduces the costs of waste water treatment, and second, recovery of valuable solids that can

be dried and used for feeding purposes. Both objectives increase the profitability of the dairy plant.

The RO permeate from the processing of different dairy fluids has a low BOD, in the range of 30–300 mg l^{-1}. Therefore it can be used as a good-quality supply of water for cleaning purposes.

Fruit and Vegetable Juices

Juices of most fruits and vegetables are characterized by their high osmotic pressure. However, they can be concentrated by RO to relatively high solids contents. The resulting concentrates are of premium quality as they are subjected to minimum heat damage during processing. However, early studies on RO concentration of juices were concerned with two factors:

1. The rapid fouling of membranes due to the presence of fibers and pectin in the juices. Therefore, pretreatment may be needed before concentrating juice by RO. The development of optimum RO operational conditions, and the selection of type of membrane and module configuration were established for several juices. In this respect a study showed that hollow fiber modules performed better than spiral-wound modules because of their low concentration polarization effects.
2. The variable retention of the different aroma volatiles during RO concentration. The balance in the concentration of these compounds is critical to obtain a concentrate with similar characteristic flavor of the fresh juice. In this connection, thin-film composite RO membranes show better retention for aroma volatiles than cellulose acetate membranes.

Different studies have shown the technical feasibility of concentrating the juice and purée of different fruits and vegetables, including orange, grapefruit, pineapple, apple, pear, passionfruit, apricot, cherry, strawberry, sugar beet, tomato, roasted coffee beans, and green tea extract. However, industrial processing by RO has been limited to the juices described below.

Tomato Juice

Tomato juice is a very variable material depending on the fruit variety, soil, and climatic conditions and processing conditions. However, the quality of the tomato juice has little effect on its behavior during RO processing. Early attempts at concentrating tomato juice by RO were based on removal of the suspended material (mainly dietary fiber) by centrifugation or ultrafiltration, concentrating the clear serum by RO, followed by mixing the concentrate with the insoluble precipitate or concentrate. However, this method has not been applied commercially.

The advent of the new generation of RO thin-film composite membranes enabled food processors to concentrate tomato juice using RO on a commercial scale. The different parameters for RO of tomato juice have been optimized. It is of interest to note that changing feed velocity has a small affect on flux. This unusual behavior has been attributed to the rheological properties of tomato juice. Several plants are now concentrating tomato juice by RO from 4.5–5.0 to 8 Brix. The different tomato juice concentrates – sauces (8–12 Brix), passata (16 Brix), and pastes (27–28 or 37–38 Brix) – can then be made from RO concentrate by evaporation. The rejection of the tomato juice solids is little affected by the quality of the fruit and the operating conditions. However, a rejection of 99% or more of tomato juice solids has been recorded. The RO tomato juice concentrate is characterized by improved color and flavor.

Citrus Juices

Unlike tomato juice, the retention of aroma and flavor is much more critical for the quality of the concentrate of most fruit juices. It is estimated that the various volatiles which give the characteristic flavor of fresh orange juice are removed with the first 15–20% of water removed by evaporation. These volatiles are partially recovered by appropriate condensation and stripping and added back to the concentrate prepared by evaporation. However, the flavor of the concentrate is different from that of the fresh juice. Orange juice can be directly concentrated by RO up to 30% solids at acceptable flux and with good-tasting product. An ultrafiltration/RO system has been developed for the preparation of orange juice and grapefruit juice concentrate that avoids the limitations of the traditional method for concentrating the juice. In this method, the fresh juice is first fractionated by ultrafiltration into concentrate and ultrafiltration permeate. The ultrafiltration concentrate is concentrated by evaporation and sterilized, while the ultrafiltration permeate is concentrated by RO. The RO concentrate is then mixed with evaporated ultrafiltration concentrate before aseptic packaging. When the prepared concentrate (double strength) is diluted, it is of similar quality to the fresh juice.

When lemon juice was concentrated by RO, 79.9% of volatiles were recovered. Loss of the volatile compounds has been attributed to adsorption in the polymeric matrix of the membrane.

Apple Juice

The presence of pectin in apple juice affects the behavior of the juice during RO concentration and the

quality of the concentrate. Therefore, apple juice is usually depectinized either by enzymatic treatment or by removal of pectin by microfiltration. Depectinized apple can be concentrated by cellulose acetate RO membranes from 10 to 20–25 Brix at 20 °C, with good retention of volatile flavor compounds. Using noncellulosic membranes, depectinized apple juice can be economically concentrated to 20–25 Brix at 30–50 °C, with better retention of soluble solids and flavor compounds. Economic evaluation showed that a RO unit used to concentrate apple juice to reduce shipment costs would pay for itself in 4 months.

The retention of aroma volatiles during RO concentration of apple juice varied between different compounds, with *trans*-2-hexenal showing the highest and hexyl acetate the lowest retention.

Coffee

Roasted coffee extract contains about 13% total solids. This extract can be concentrated by RO up to 36% solids content without appreciable loss of solids or aroma compounds. The RO concentrate of coffee extract can be used as a base for the preparation of instant coffee. (*See* **Coffee**: Roast and Ground.)

Alcoholic Beverages

Wine

Grape must with a low sugar content of < 17% makes table wine of inferior quality and insufficient alcohol content to meet legal requirements. The traditional method of combatting this problem is called chaptelization, in which sucrose or glucose is added to the must before fermentation. This method can be replaced by RO concentration of must. The wine results from RO concentrated must have improved flavor without any adverse effects on general quality.

Removal of bitterness has been reported as another application of RO in the wine industry. Red wine of poor quality can be subjected to RO concentration using cellulose acetate membrane, with an approximate molecular cut-off of 300 Da. The bitterness-causing compounds (probably ketones) pass in the permeate while the concentrated wine has improved quality. This wine concentrate has much shorter cold stabilization time to remove tartrate. (*See* **Wines**: Production of Table Wines.)

Low-Alcohol Beer

The demand for beer with low alcohol content has increased recently. This can be achieved either by controlling the fermentation conditions to develop less alcohol or the partial removal of alcohol from normal beer. The first trend has the disadvantage of impairing the development of flavor compounds in addition to alcohol.

With the use of special RO membranes, it is possible to remove the alcohol partially from beer. This has been applied commercially in many breweries in different parts of the world. A thin composite RO membrane with a molecular cut-off of 100 Da has been used to remove 25–30% of alcohol content in beer. The beer is processed by RO while diluting the retentate with water to a constant volume under carbon dioxide. Therefore, the alcohol content of the beer can be changed to the desired level. The quality of water used for diluting the concentrated beer determines the quality of the final product.

Corn Wet Milling

Corn wet milling is a process that gives starch as the main product output in addition to several other products, namely, oil, protein, and fiber. This process is a water-intensive technology as 1.5 m^3 of fresh water per ton of corn is needed in modern corn wet milling. RO has been used to recover water from the light middlings (the overflow from the hydrocyclone starch system). The recovered water is then used to wash starch. Application of this process in a corn wet milling plant reduced the load of the steep water evaporator to 30% of the original value. The use of RO in corn wet milling had no adverse effect on the quality of the produced starch.

Concentration of Dextrose Syrup and Sweet Water

In the manufacture of dextrose syrup, the syrup is usually concentrated by evaporation from 30–34% to 40–45% total solids before decolorization with carbon. Evaporation can be successfully replaced by RO concentration, with marked savings in energy.

Also, during the manufacture of the high-fructose syrup, an effluent called sweet water results from washing the ion exchange and carbon columns and storage and transportation facilities. The sweet water contain 3–8% sugar content. This effluent can be preconcentrated more economically by RO compared to evaporators.

Removal of Solvent from Micella

The edible oil industry is based on oil extraction from seeds with suitable solvent, degumming and bleaching of the extract, followed by removal of the solvent from the micella by evaporation.

Using special RO membranes, it is possible to remove large amounts of the solvent from the micella, and removal of the residual solvent can be completed by evaporation. However, most of the material used in RO modules does not resist the destructive action of the solvent.

See also: **Alcohol**: Properties and Determination; **Apples**; **Beers**: Biochemistry of Fermentation; **Citrus Fruits**: Oranges; Processed and Derived Products of Oranges; Lemons; Grapefruits; Limes; **Coffee**: Roast and Ground; **Milk**: Processing of Liquid Milk; **Milling**: Principles of Milling; **Syrups**; **Tomatoes**; **Vegetable Oils**: Oil Production and Processing; **Whey and Whey Powders**: Production and Uses

Further Reading

Darnes P and Jordanian P (1987) How el-goods reverse osmosis to brew low alcohol beer. *Brewing and Distilling International* 17: 47–49.

Gestoli C, Brandini S, di Francesca R and Zerdie G (1995) Concentrating fruit juice by reverse osmosis. The low retention–high retention method. *Fruit Processing* 5: 183–187.

Glover FA (1985) *Ultrafiltration and Reverse Osmosis for the Dairy Industry*. Technical Bulletin 5. Reading, UK: NIRD.

Kane L, Braddock RJ, Sims CA and Matthews RF (1995) Lemon juice aroma concentration by reverse osmosis. *Journal of Food Science* 60: 190–194.

Kollacks WA and Rekers CJN (1988) Five years of experience with the application of reverse osmosis on light middlings in a corn wet milling plant. *Starch/Starke* 40: 88–94.

Koseoglu SS and Egelgau DE (1990) Membrane applications and research in the edible oil industry: an assessment. *Journal of American Official Chemists Society* 67: 239–249.

Koseoglu SS, Lawhon JT and Lusas EW (1990) The use of membranes in citrus juice processing. *Food Technology* 44: 90–97.

Kosikowski FV (1986) Membrane separation in food processing. In: McGregor WC (ed.) *Membrane Separations in Biotechnology*, pp. 201–253. New York: Marcel Dekker.

Paulson DJ and Wilson RL (1987) Cross flow membrane technology: its use in food industry: recent innovations. In: Kroger M and Shapiro R (eds) *Changing Food Technology*, pp. 85–123. Lancester, PA: Technomic.

Rao MA and Vitali AA (1999) Fruit juice concentration and preservation. In: Shafier Rahman MS (ed.) *Hand Book of Food Preservation*, pp. 217–258. New York: Marcel Dekker.

Robe L (1983) Hyperfiltration methods for preconcentrating juice save evaporation energy. *Food Processing* 44: 100–108.

Singh N and Cheryan M (1998) Membrane technology in corn refining and bioproduct processing. *Starch/Starke* 50: 16–23.

Principles of Ultrafiltration

D Pal, National Dairy Research Institute, Karnal, Haryana, India

Introduction

Though very rapid progress has been made in all the membrane processes during the last two to three decades, ultrafiltration, because of its unique nature and diverse uses, has received more attention worldwide. This article discusses the principle of the ultrafiltration process, the membrane materials, and their configurations. Some major limitations and operating conditions are also included.

What is Ultrafiltration?

A pressure-driven, cross-flow membrane system that can simultaneously purify, concentrate, and fractionate organic molecules of a feed stream is known as ultrafiltration. The process is very simple (**Figure 1**). The pressurized influent stream is passed across a semipermeable membrane which separates it into two effluent streams known as permeate and retentate (concentrate). Permeate is that fraction which has passed through the semipermeable membrane and contains some small dissolved molecules. Retentate is that stream which has been enriched in the solutes or suspended solids not passed through the membrane. The physical and chemical nature of the membrane itself controls which components permeate and which are retained.

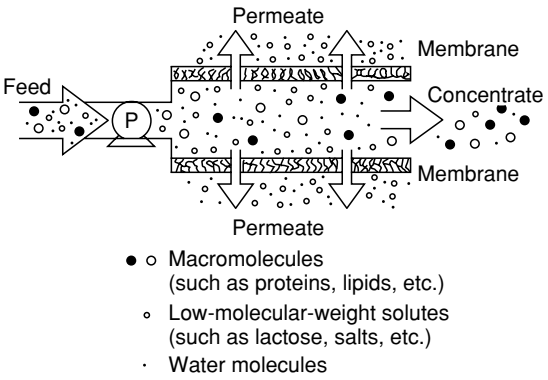

Figure 1 Principle of ultrafiltration process. P, pump. Reproduced from Membrane Techniques, Principles of Ultrafiltration, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 1 Molecular weight and relative diameters of some milk constituents

Milk constituent	*Molecular weight (Da)*	*Diameter (nm)*
Water	18	0.3
Chloride ion	35	0.4
Calcium ion	40	0.4
Lactose	342	0.8
α-Lactalbumin	14 500	3.0
β-Lactoglobulin	36 000	4.0
Blood serum albumin	69 000	5.0
Casein micelles	10^7–10^9	10–600
Fat globules	–	200–10 000

Data from Kessler HG (1981) *Food Engineering and Technology*. Freising: A Kessler.

Principle of Ultrafiltration

The mechanism of separation in ultrafiltration is essentially a sieving process in which constituents of a feed stream are separated according to their molecular weight. The ability of the membrane to retain the majority of the defined macromolecules of known molecular weight is generally used to specify the porosity of the membrane. The term used is molecular weight cut-off (MWCO), which should be the molecular weight of the smallest test macromolecule that is largely rejected by the membrane. Most ultrafiltration membranes reject the constituents having a molecular weight in the range 1–1000 kDa. Table 1 shows the molecular weight and relative diameter of major milk constituents. Since the majority of ultrafiltration membranes have pore sizes ranging from 1 to 50 nm, essentially all milk constituents except water, lactose, ions, and some water-soluble vitamins are rejected. Ultrafiltration membranes are completely impermeable to lipids and proteins and substances bound to them.

Ultrafiltration

Nearly all ultrafiltration membranes are asymmetric in morphology. It means they have a thin dense layer or skin (about 0.2 μm) on top of the membrane which defines the degrees of separation affected, and a spongy support layer (about 100 μm thick) underneath. These membranes are usually homogenous in material, in that they consist of the same polymer or copolymer throughout their structure. The following materials have been used successfully in the preparation of ultrafiltration membranes.

Cellulose Acetate (CA)

These were perhaps the first commercially developed ultrafiltration membranes. The CA membranes, however, are not used so widely now in ultrafiltration equipment because of the following limitations:

1. poor thermal (up to 35 °C) and chemical (pH range 3–7) stability;
2. poor tolerance to chlorine;
3. more susceptible to microbial spoilage than other types;
4. undergo creep or compaction to a relatively greater extent.

Polysulfone (PS)

Ultrafiltration membranes made of the tough, relatively inert polysulfone polymer are perhaps the most widely used currently in the food industry. In fact, polysulfone is considered a breakthrough for ultrafiltration principally due to its following advantages:

1. It can be operated at higher temperatures (up to 75 °C), with some manufacturers claiming that their PS membranes can withstand temperatures up to 125 °C.
2. It is compatible with a much wider pH range (0.5–13), and thus their cleaning is easy.
3. It is available in a wide range of pore sizes, ranging from 0.001 to 0.02 μm.
4. It has a better resistance to chlorine.

The only limitation of the PS membranes is that they can only be operated in a very narrow pressure range (170–700 kPa).

Mineral or Ceramic Membranes

The latest membranes on the market for ultrafiltration applications are mineral membranes. These are formed by the deposition of alumina, zirconia, carbon, or silver on to a microporous support of the same material. Since these are made up of inorganic materials, they are free from the disadvantages of polymeric membranes, such as narrow range limitations for temperature, pH, and operating pressure. The manufacturers claim that the ceramic membranes can withstand pressures up to 2.0 MPa without creep or compaction, and tolerate the whole pH range and temperatures up to 400 °C. The cleaning of these membranes is very easy. They can be sterilized by autoclaving. The limitations of ceramic membranes include their availability only in tubular configuration. The large pore sizes restrict their applications. Also, these membranes are very expensive.

Ultrafiltration Membrane Configuration

The commercial ultrafiltration plants are available in four designs, namely tubular, hollow fiber, plate and frame, and spiral wound. The general characteristics of each module are discussed here. Their advantages and drawbacks are listed in Table 2.

Tubular Module

Tubular modules were perhaps the earliest design of industrial-scale ultrafiltration equipment using synthetic membranes. In this module, the membrane itself may be cast directly on to a fiberglass tube, or they may be cast on to a separate paper tube which is then inserted into a perforated stainless-steel tube (Figure 2). Bundles of several individual membrane tubes are housed together in a stainless-steel shell. These modules have relatively large open-feed channels with internal diameters of 12–25 mm and lengths of 0.6–6.4 m. The feed solution, under turbulent conditions, flows inside each tube, whose inner wall contains the membrane. The permeate passes through the membrane and support material and is collected in the housing.

Hollow Fibers

These are in the form of self-supporting tubes with the dense 'skin' layer on the inside of the tube, the diameter of which generally ranges from 0.5 to 1.1 mm (Figure 3). Hundreds of these fibers, depending on fiber diameter and size of the housing, are sealed into a cartridge in a 'shell-and-tube' arrangement and bonded at each end into an epoxy tube sheet. The feedstream flows through the inside of the fibers and the permeate is collected on the outside. Each industrial cartridge may contain from 0.7 to 2.8 m^2 of membrane area. These are operated in laminar flow using high shear to control concentration polarization.

Plate and Frame

In this configuration, flat sheets of the membrane are placed between support plates which form flow channels over the membrane with heights in the range 0.5–2.5 mm. This arrangement is very similar to a conventional plate and frame filter press. The membrane and their supports are sandwiched together in

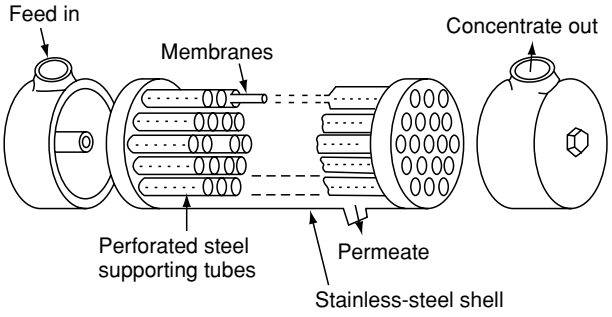

Figure 2 Tubular ultrafiltration module. Reproduced from Membrane Techniques, Principles of Ultrafiltration, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 2 Advantages and limitations of different ultrafiltration configurations

Configuration	*Advantages*	*Limitations*
Tubular	Handles suspended solids with larger particles Able to predict membrane performance using simple fluid dynamics Possible to replace individual membrane on site, so replacement cost is less Cleaning easy	Highest energy consumption per unit volume of permeate High pressure drop Lowest surface area-to-volume ratio, so needs maximum floor space Hold-up volume per unit area high
Hollow fiber	Energy consumption lowest among the modules Highest surface area-to-volume ratio, lowest hold-up volume Only module that allows backflushing	Since no support for the hollow fibers, operates in narrow pressure range (170–270 kPa) Fibers are susceptible to plugging Handling of large particles and suspended solids problematic Complete cartridge needs to be replaced in case of leakage, so replacement costs high
Plate and frame	Energy consumption moderate; less than tubular In case of leakage the particular membrane is replaced; so replacement cost is lowest Surface area-to-volume ratio and hold-up volume intermediate between tubular and spiral wound	Cleaning of membrane more difficult Initial capital costs relatively high
Spiral wound	Allows very high applied pressure without damage to the membrane Very economical in terms of energy consumption and membrane replacement Capital costs very low Surface area-to-volume ratio very high Low hold-up volume	Relatively difficult to process fluids having high suspended solids or fibrous matter Large particles may hang up in the mesh spacer, so causing cleaning problems High pressure drop

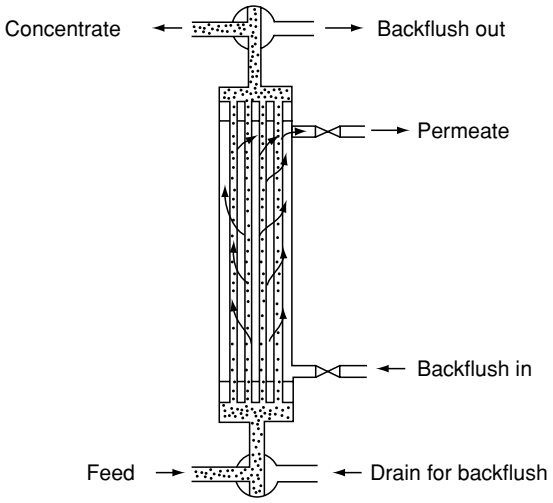

Figure 3 Romicon hollow fiber ultrafiltration module. Reproduced from Membrane Techniques, Principles of Ultrafiltration, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

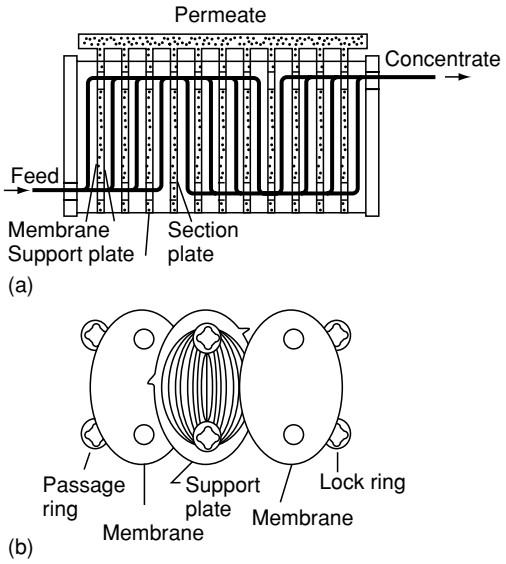

Figure 4 Plate and frame ultrafiltration module: (a) internal flow of product; (b) oval-shaped plate membranes and support plate. Reproduced from Membrane Techniques, Principles of Ultrafiltration, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

larger numbers (e.g., 180 spacer plates and 360 membrane sheets totalling 27 m^2 of active membrane area) to form one module. The plates containing membrane sandwiches on either side are generally horizontally stacked together in ultrafiltration plants, and the feed is pumped between the plates (Figure 4). The permeate may be collected either from each sandwich or together from the whole stack. Most plate units tend to be operated in laminar flow, although some wide-channel configurations could be under turbulent flow.

Spiral Wound

Like plate and frame, the spiral wound module also makes use of economical, flat sheet membranes. Two membrane sheets with a mesh-type spacer in between, to form the permeate channel, are pasted along three sides (appearing like an envelope). The fourth side is pasted to a perforated tube through which the permeate passes. Another mesh-type spacer is placed on top of the envelope and the whole arrangement is rolled around the perforated central permeate tube (Figure 5). The feed channel height is controlled by the thickness of the mesh-type spacer placed between two envelopes; spacers of 0.75–1.55 mm are common. The feed, under laminar flow, is pumped lengthwise along the unit, while the permeate is forced through the membrane sheets into the permeate channel and spirals its way towards the perforated central tube.

Operating Conditions

The processing parameters during membrane processes are optimized with a view to attaining maximum possible flux, which is the volumetric rate of flow of the permeate through the membrane. Four vital parameters that affect the flux during ultrafiltration processing are discussed here.

Pressure

For a particular membrane, the flux is directly proportional to pressure of the feed and inversely proportional to viscosity (the Hagen–Poiseuille's pore theory) and can be expressed as

$$J = A(\Delta P_t - \Delta\pi) \qquad (1)$$

where J is the flux ($l\,m^{-2}\,h^{-1}$), A is the diffusivity constant, ΔP_t is $P_f - P_p$ (P is the hydraulic pressure) and $\Delta\pi$ is the osmotic pressure. The subscript f refers to feed and p to the permeate. Since the osmotic pressure of the retained macromolecules in ultrafiltration is negligible, the flux is directly proportional to applied transmembrane pressure, i.e., $J = A\,\Delta P_t$. This relationship, however, holds true under ideal conditions only. When concentration polarization, associated boundary layer, and/or fouling takes place the flux becomes pressure-independent. The membrane itself also provides some resistance to the flux.

Feed Concentration

In the mass transfer-controlled region, the flux can be expressed employing the film theory as:

$$J = k\,\log_e(c_g/c_b) \qquad (2)$$

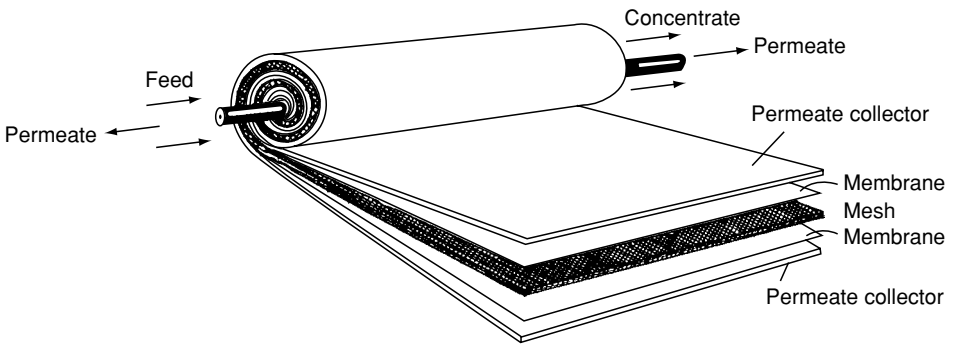

Figure 5 Spiral wound UF module. Reproduced from Membrane Techniques, Principles of Ultrafiltration, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

where k is the mass transfer coefficient, c_g is the concentration of gel at the membrane surface, and c_b is the concentration of solutes in the bulk. According to this theory, the flux should decrease with increasing feed concentration, and $J = 0$ when the concentration of solutes at the membrane and in the bulk are the same.

Temperature

A higher operating temperature reduces the viscosities of the feed and permeate and so increases the diffusivity. Thus it helps increasing the flux in both pressure-controlled and mass transfer-controlled regions. In general, it is best to operate at the highest possible temperature compatible with the feed and the membrane.

Feed Velocity

Higher turbulence has a major beneficial effect on flux in the mass transfer region. The higher the feed velocity (flow rate) across the membrane, the faster is the 'sweeping' of the built-up solutes away from the membrane surface, thereby reducing the thickness of the concentration boundary layer. Turbulence during the operation of ultrafiltration can be generated by stirring or by pumping of the feedstream.

Flux-Limiting Factors

The performance of membrane systems is generally gauged by the flux behavior. During ultrafiltration processing, the flux is lower than that of pure water. Concentration polarization and fouling are perhaps the main factors for this decline.

Concentration Polarization

During ultrafiltration, the solution is brought to the membrane surface by hydraulic transport. Because of the semipermeable nature of membrane, a portion of solvent, with or without some solutes, permeates through the membrane. This leads to a higher concentration of solutes at the membrane surface compared to the bulk. The building-up of the solutes on the membrane surface, known as concentration polarization, causes a steep concentration gradient within the boundary layer and results in a backdiffusion of solutes into the bulk. Eventually, a steady state is reached when the movement of the solutes towards the membrane by convective transport and their return to the bulk by diffusive transport balance each other. An increase in applied pressure will not help in this situation.

The solutes rejected by the membrane are usually deposited on the membrane surface in the form of a fairly viscous and gelatinous layer. This gel layer is considered to be mainly responsible for the lower flux, due to hydrodynamic resistance. When the concentration of solutes in the gel layer reaches the supersaturation point, fouling of the membrane begins.

Formation of the gel layer can be minimized by proper fluid management, such as (1) lowering the pressure, (2) lowering the feed concentration, and/or (3) increasing the feed velocity.

Membrane Fouling

The fouling of membranes is another major limiting factor in ultrafiltration processing. In contrast to concentration polarization, which is considered a time-independent, reversible process, fouling is an irreversible phenomenon. During fouling, flux drops with operating time, usually rapidly in the initial stages and slowly in later stages. Fouling is generally attributed to:

1. the accumulation or absorption of macromolecules or colloidal particles, such as proteins, lipids, microorganisms and/or inorganic salts, on the membrane surface;
2. the precipitation of permeable solutes, such as sugars and salts, due to overcrowding within the membrane pores.

The fouling of membranes not only reduces the flux but also makes the cleaning operation more difficult and expensive. The rejection properties of the membrane are also altered due to the formation of this secondary layer.

The fouling problem can be alleviated by proper: (1) selection of the membrane and plant design; (2) pretreatment of the feed; and (3) cleaning of the membrane. In fact, the pretreatment of the feed plays a most important role in reducing fouling problems, and differs from feed to feed.

In addition to concentration polarization and fouling, changes in membrane properties, either due to compaction or chemical deterioration, are also responsible for a decline of flux. The compaction of membranes generally takes place due to faulty operating conditions (high pressure and temperature), and the use of strong cleaning solutions is responsible for chemical deterioration.

Cleaning and Maintenance of Ultrafiltration Membranes

The most important consideration in cleaning and maintenance is never to allow the feed constituents to dry out on the membrane surface. While deciding on the cleaning schedule for an ultrafiltration membrane, the nature of the feed constituents that foul the membrane, and the recommendations of the membrane and detergent manufactures should be kept in mind. The following procedure can be adopted for routine cleaning of a membrane:

1. After the processing is complete, the feed stream is flushed out from the whole plant with water. The highest temperature of wash water compatible with the membrane should be used. Flushing is stopped when clear water starts coming out of the system.
2. Then a commercial detergent solution is circulated through the plant. For PS membranes, alkali (1%) and acid (0.5%) detergents are circulated one after the other. The addition of a sequestering agent, e.g., hexametaphosphate, to cleaning detergents is very beneficial. In the case of CA membranes and excessive fouling of PS membranes, the use of a proteolytic enzyme is essential.
3. After each circulation of detergent, the whole system is thoroughly rinsed with hot water.
4. To insure proper cleaning, the flux rate with water under standard conditions is tested. If the water flux is lower than the initial flux, step (2) is repeated.
5. The membrane is sanitized and stored in 0.5% formaldehyde solution. It is never allowed to remain dry. (*See* **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant.)

See also: **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant

Further Reading

Cheryan M (1986) *Ultrafiltration Handbook*. Lancaster, PA: Technomic Publishing.

Cheryan M (1998) *Ultrafiltration and Microfiltration Handbook*. Lancaster, PA: Technomic Publishing.

Glover FA (1985) *Ultrafiltration and Reverse Osmosis for the Dairy Industry*. Technical bulletin no. 5. Reading: NIRD.

Makardij A, Chen XD and Farid MM (1999) Microfiltration and ultrafiltration of milk: some aspects of fouling and cleaning. *Food and Bioproducts Processing* 77(C2): 107–113.

Marshall AD and Daufin G (1996) Physico-chemical aspects of membrane fouling by dairy fluids. *Bulletin of International Dairy Federation* 311: 38.

Nwuha VO (1996) Influence of cleaners on polysulphone membrane used for milk ultrafiltration. *International Journal of Food Science and Technology* 31(1): 27–36.

Pal D and Cheryan M (1987) Membrane technology in dairy industry. Part II. Ultrafiltration. *Indian Dairyman* 39(8): 373–392.

Paulson DJ, Wilson RL and Spatz DD (1984) Crossflow membrane technology and its applications. *Food Technology* 38(12): 77–87.

Renner E and Abd El-Salam MH (1991) *Application of Ultrafiltration in Dairy Industry*. Barking: Elsevier.

Roberts JJ (1994) Membrane technology: basic principles, latest developments and applications. *Food Industries* 47(4): 24–29.

Applications of Ultrafiltration

D Pal, National Dairy Research Institute, Karnal, Haryana, India

Introduction

The applicability of ultrafiltration has widened considerably because of substantial decrease in cost of membranes, availability of better designed units and decrease in fouling problem of membrane. It is now being used in diverse fields such as water treatment, chemical processing, food processing, and biotechnology. Dairy applications probably account for the largest share of installed membrane capacity. Current estimates suggest that over 300 000 m^2 of membranes are installed in the dairy industry worldwide. This article focuses on the use of ultrafiltration in the processing of milk and whey, clarification of fruit

juices and beverages, and some biotechnological applications.

Processing of Milk

The ultrafiltration process alters the chemical composition and physicochemical properties of milk. These changes are discussed here, with particular reference to cheesemaking.

Effect of Ultrafiltration on Chemical Composition

Ultrafiltration membranes are completely impermeable (about 100% rejection) to lipids and proteins and to all those substances associated with them. Hence their concentration in the retentate increases in proportion to the concentration factor. Since the rejection of lactose by ultrafiltration membrane is virtually zero, almost all of it passes into the permeate. This results in either the same or a lower concentration of lactose in the retenate than in the milk.

Most of the water-soluble vitamins pass through the membrane and their concentrations do not increase in the retentate. Vitamin B_{12} and folic acid bound with proteins and vitamins A, D, E, and K bound with fat become increasingly concentrated.

The rejection of minerals by ultrafiltration membranes depends on their nature and state in milk. Typical rejections can range from about 25% to over 90%. Calcium, magnesium, zinc, iron, copper, and phosphate are partly associated with the casein micelles and partly present in solution. Only the bound form of these minerals is concentrated in the retentate, while those in solution pass through the membrane. The net result is that the concentration of minerals in the ultrafiltration retentate increases, but not in proportion to the concentration factor. The nonprotein nitrogen (NPN) fraction, which consists of urea, amino acids, and ammonia, is not concentrated in ultrafiltration retentate. Table 1 shows the distribution of the major milk constituents in an ultrafiltration permeate and retentate. (*See* **Milk**: Dietary Importance.)

Factors like membrane composition, pasteurization, and/or homogenization of milk do not affect the rejection of components. However, treatments like acidification or changes in pH may alter the distribution of minerals, particularly calcium phosphate, between retentate and permeate. Diafiltration (that is, adding water to the retentate and its reultrafiltration) can reduce the concentration of both lactose and minerals in the retentate.

Change in Viscosity

The viscosity of the retentate increases during ultrafiltration in proportion to the increase in protein concentration, and a significant non-Newtonian behavior can be observed. Viscosity may increase up to 10 times by increasing the protein from 3 to 18% at 50 °C, and as high as 100 times at 15 °C. Hence higher operating temperatures are preferred during ultrafiltration. The increase in viscosity is higher for whole-milk retentate than for the skim-milk equivalent. As a consequence of high viscosity, the ultrafiltration retentate cannot be cooled quickly, and this may cause rapid growth of microbial contaminants. Another problem associated with high viscosity is that the entrapped air bubbles in the retentate are not released quickly, resulting in a spongy texture in cheese produced subsequently.

Influence on Fat Globules

Recirculation of pressurized whole milk during ultrafiltration processing damages the fat globule membrane; this leads to fat separation in the retentate during storage. The extent of fat separation depends on the quality of the raw milk and the characteristics of the plant. Such destabilization of fat globule membrane affects the textural properties of certain cheese varieties and other dairy products. Use of skim milk

Table 1 Distribution of major milk constituents between retentate and permeate during ultrafiltration

Constituents	*Retentate*			*Permeate*		
(%)	*1 ×*	*3 ×*	*5 ×*	*1 ×*	*3 ×*	*5 ×*
Total solids	12.9	28.6	43.3	5.7	6.1	6.7
Fat	3.9	12.6	21.8	0	0	0
Protein	3.1	9.8	16.1	0	0.06	0.49
Nonprotein nitrogen	0.18	0.18	0.18	0.18	0.19	0.19
Lactose	4.7	4.1	3.2	4.8	5.1	5.2
Ash	0.77	1.3	1.9	0.53	0.53	0.54

Data from Glover FA (1985) *Ultrafiltration and Reverse Osmosis for the Dairy Industry*, Technical Bulletin no. 5. Reading: NIRD.

for ultrafiltration and subsequently adding cream to the retentate may overcome this problem.

Effect on Proteins

Under normal operating conditions, casein and whey proteins are not normally affected by ultrafiltration process. Air incorporation into the retentate may however lead to denaturation of whey proteins at the air–water interface. Though low-temperature processing does not affect the whey proteins, ultrafiltration processing at high temperature (> 55 °C) even for a short period like 2 h may cause complexing of β-lactoglobulin with casein. Such complexing increases with residence time, concentration level, temperature, and amount of air entrapped. Diafiltration of retentate may also contribute to whey protein denaturation.

Effect on Starter Activity

Milk proteins and insoluble salts of calcium and phosphate, responsible for buffering action, are concentrated by the ultrafiltration process. Increased buffering capacity causes technological problems in the manufacture of cheese from highly concentrated retentate because a relatively large amount of lactic acid must be produced to obtain the desired pH. Consequently, the use of very active starter culture is necessary to produce the desired level of lactic acid; failure to reduce the pH would enhance the risk of growth of undesirable microorganisms in the cheese. (*See* **Starter Cultures**.)

Effect on Rennet Coagulability

The renneting time, in general, decreases as the concentration of ultrafiltration increases. The faster and more effective collision of micelles, due to the increase in protein content or increased calcium concentration, may result in more secondary-phase interactions and a decrease in clotting time.

Protein Standardization by the Ultrafiltration Technique

Protein standardization has now become an important international issue mainly because of the following reasons:

1. Seasonal and regional differences in protein contents of milk are inherent features which affect the quality and composition of products made therefrom, hence the need for equalizing protein.
2. Fortification of milk with total protein to increase the nutritional value.
3. Manufacture of total milk protein concentrates for use in special and dietetic foods.

Ultrafiltration can be used with great success for upward standardization of milk proteins in all the above-mentioned situations.

Cheesemaking with the Ultrafiltration Process

Several varieties of cheese are commercially produced employing the ultrafiltration process. Ultrafiltration helps to obtain a retentate with a composition identical to that of the desired cheese. Ultrafiltration technology offers the following advantages:

1. The increase in cheese yield is the most attractive benefit. In the traditional cheesemaking process, all the whey proteins and some casein fines, which account for nearly 20% of total milk proteins and some fat particles, are lost into whey. Ultrafiltration processing helps to retain all these constituents in the cheese. Though the actual increase in the yield depends on the level of concentration achieved during ultrafiltration and the type of cheese made, about 8% higher yield for hard cheese (e.g., Cheddar cheese) and up to 30% for semihard and soft varieties of cheeses is commercially obtainable.
2. Mechanization and automation of cheesemaking are possible.
3. The requirements for starter culture and rennet are reduced.
4. The amount of whey to be disposed of is either substantially reduced or completely eliminated, and the byproduct of the ultrafiltration process, the permeate, has many useful applications. Thus the biochemical oxygen demand (BOD) of the cheese waste steam is considerably reduced. (*See* **Effluents from Food Processing**: Composition and Analysis.)
5. The milk-handling capacity of existing cheesemaking equipment can be increased without increasing the vat size or floor area.
6. The overall energy requirements are low for the concentration of milk, as well as for heating and cooling the retentate.

Classification of Ultrafiltration Methods for Cheesemaking

Based on the levels of concentration achieved, the manufacture of cheese by the ultrafiltration process can be broadly divided into three categories.

Low-concentration factor (LCF) method Milk is usually concentrated up to twofold by ultrafiltration followed by cheesemaking employing traditional processes and equipment. Alternatively, milk is concentrated up to fivefold, and then used to supplement

whole/skim milk intended for cheesemaking to the extent that the concentration of solids in the supplemented milk does not exceed twofold the original level. The method can be used for most varieties of cheese without difficulty. The quality of cheeses produced from LCF retentate is generally identical to the traditional cheese. Better utilization of existing cheese vats and production of cheese of identical protein content throughout the year are two of the advantages of the LCF process. Since the drainage of whey occurs just as in the traditional cheesemaking process, the benefit of higher yield is not achieved. (*See* **Cheeses**: Chemistry of Gel Formation.)

High-concentration factor (HCF) method In the HCF process, milk is concentrated between three- and sixfold. Beyond sixfold it is not possible to concentrate milk with the present generation of membranes. Since many changes in composition and other properties of the retentate take place during high concentration, modified process and equipment are required for cheesemaking. This method is used for the manufacture of some hard and semihard varieties where partial drainage of whey is essential to achieve the desired total solids and/or for texturizing the product. The most successful commercial application of this process is in the manufacture of Cheddar cheese. Based on the fivefold concentration of milk, the CSIRO Dairy Research Laboratory, Australia and APV Ltd have jointly invented a commercial process for making Cheddar cheese (known as APV Siro curd process). Cheddar cheese made by this method is identical to traditional cheese.

Precheese method Skim milk or whole milk is concentrated up to the solids equivalent desired in the final cheese. There is usually no production of whey, which enables 100% utilization of whey proteins and casein in the final product. This process is based on completely new cheese technology and new equipment. A true precheese ultrafiltration process is most successful for the preparation of soft and semihard varieties, i.e., those cheeses containing more than 45% moisture, e.g., cream cheese, quarg, ricotta, feta, mozzarella, and camembert. (*See* **Cheeses**: Types of Cheese.)

Other Uses of Ultrafiltration Retentate

The ultrafiltration retentate can also be used in the manufacture of some other fermented dairy products like yogurt, shrikhand, and ymer. The ultrafiltration retentate can be spray-dried in the flush reason and utilized in the lean season for manufacturing cheese. Ultrafiltration in combination with diafiltration helps to produce highly purified food-grade proteins with low lactose content, and having highly improved functional properties such as emulsion capacity and foaming stability (Table 2). Such functional proteins are finding increasing applications in the food industry. (*See* **Drying**: Spray Drying.)

Table 2 Composition and functional properties of skim milk, ultrafiltered (UF) retentate (5 ×) and whole-milk protein isolates (WMPI) of cow and buffalo milk

Components	*Skim milk*		*UF retentate*		*WMPI*	
	Cow	*Buffalo*	*Cow*	*Buffalo*	*Cow*	*Buffalo*
Total solids (%)	8.85	9.67	24.33	24.24	18.46	19.69
Total nitrogen (%)	0.534	0.642	2.85	2.93	2.61	2.85
Total protein (%)	3.20	3.66	18.00	18.55	16.44	18.06
Total protein (DM) (%)	37.30	39.92	73.95	76.53	89.06	91.72
Casein (DM) (%)	32.40	34.95	64.88	66.79	77.31	79.23
Whey protein (DM) (%)	4.90	4.95	9.12	9.74	11.75	12.49
Casein/WP	6.60	7.06	7.11	8.86	6.58	6.34
Lactose (DM) (%)	56.50	48.60	18.10	17.90	6.70	8.70
Ca^{2+} (DM) (%)	1.66	2.07	2.67	3.65	2.26	3.09
Alcohol stability (%)	82	75	75.7	62.5	52.3	41.8
Solubility (%)			92.7	93.5	86.7	88.6
Clotting time (min)	4.9	1.47	2.1	0.67	3.1	2.55
Curd firmness (K20) (min)	2.1	0.93	0.9	0.40	1.8	1.0
Emulsion capacity ($m^2\,g^{-1}$)			61.2	102.8	59.74	114.23
Foaming capacity (%)			120	110	120	110
Foam stability						
30 min			31.0	58.8	32.95	58.8
60 min			24.0	41.8	25.75	41.0
90 min			19.0	33.3	19.0	34.0

Data from El-Shibny S, Shahein NM and Sheikh M (1996) Preparation and functional properties of ultrafiltered milk protein isolates. *Bulletin of International Dairy Federation* 311: 28–30.

Treatment of Whey

Whey is a byproduct of cheese and casein manufacture, and of the 6.5% solids present in whey, the proportions of protein, lactose, and minerals are 12, 75, and 9%, respectively. Small amounts of fat and some organic acids are also present. Whey is a heavy pollutant with a BOD as high as 50 000 p.p.m. Until the advent of membrane technology, whey had to be either drained into the sewers or given to livestock growers. Because of its very high water content, the processing of whey employing conventional evaporators and driers is highly uneconomical. The high lactose-to-protein ratio is another factor limiting the proper utilization of concentrated and dried whey. (*See* **Whey and Whey Powders**: Production and Uses.)

The development of the ultrafiltration process has undoubtedly proved a boon for the cheesemakers in the treatment of whey. The use of ultrafiltration to fractionate and concentrate whey proteins, followed by evaporation and spray drying, is now a well-established commercial process for the manufacture of whey protein concentrate (WPCs). WPCs prepared by ultrafiltration technology, having 35–80% protein, are now available commercially. WPC can be further separated into β-lactoglobulin and α-lactalbumin fractions, or used for the manufacture of casein macropeptide, a compound that may have pharmatherapeutic value. The main attraction of ultrafiltration for processing whey includes an almost 100% recovery of whey proteins for edible and other useful applications, with a simultaneous reduction in lactose content. It is possible to concentrate protein up to 20 times with a 95% reduction in volume; further increases in protein and reductions in lactose content are possible employing the diafiltration technique. The energy requirements for removing water from whey using the ultrafiltration process are minimal compared with any other dewatering techniques. Since ultrafiltration is carried out at low temperatures (about 50 °C), the native and functional properties of whey proteins are kept intact. The pollution load of whey is remarkably reduced and, in combination with reverse osmosis, the BOD value of whey can be reduced by about 98%.

Decline in the flux rate due to fouling of the membranes is the main limitation in processing of whey by ultrafiltration. In fact, whey is one of the most notorious foulants of all the fluid milk products. Some remedial steps to alleviate fouling problem are given below:

1. Whey should be separated and clarified before processing in order to remove suspended casein fines and lipid constituents.
2. In general, the flux is lowest at the isoelectric point of protein and becomes higher as the pH deviates from this point. A combined treatment of heating to 80–85 °C for 15 s followed by pH adjustment to about 3 or 7 is very useful for acid whey (casein whey); pH adjustment is not necessary for sweet whey (e.g., Cheddar cheese whey).
3. Complexing of proteins (e.g., casein micelles with β-lactoglobulin) by heat treatment or solubilization of proteins by either enzymatic treatment or the use of carboxymethyl cellulose is reported to increase the flux.
4. Salts, particularly calcium phosphate, exaggerate the fouling problem. Addition of sequestering agent, e.g., ethylenediaminetetraacetic acid (EDTA) or sodium hexamataphosphate, to an acid whey has a very beneficial effect on the flux. Partial replacement of calcium with sodium or its removal by a demineralization process may also be adopted.

Clarification of Fruit Juices and Beverages

Many juices and beverages contain colloidal and suspended solids which need to be removed before the product is marketed. It is economically feasible to use an ultrafiltration process for the commercial production of many 'single-strength' beverages. The principle behind the ultrafiltration clarification process is that the beverage flows over the membranes, and while the clarified beverage passes through the membrane in the form of permeate, the colloidal and suspended solids are rejected into the retentate stream. Hence, the permeate is the desired stream and the retentate is discarded.

Fruit Juices

Traditionally, press juice is subjected to pasteurization followed by enzymatic hydrolysis of pectin. Then a fining agent (gelatin) is added to the juice, and the suspension is held for about 20–30 h. After decanting, the juice is passed through a filter aid. The application of ultrafiltration to fruit juice clarification helps replace the holding, decantation, and filtration steps. All the undesirable components, such as pectin, large carbohydrates, and tannin protein complexes, which are responsible for cloudiness and sedimentation, are retained by the ultrafiltration membrane. Polyphenol oxidase, which causes browning of fruit juices, is also not allowed to permeate along with the clarified juice.

So far, the largest application of ultrafiltration in the juice industry has been in the production of apple juice; a flow diagram of the process in given in

Figure 1. Since microorganisms are larger than the pore size of the membranes, these are retained in the residue and the clarified juice is essentially sterile. The aseptic handling of this juice can either completely eliminate or substantially reduce the subsequent heat treatment prior to storage and bottling. It is also possible to minimize the losses of sugar and low-molecular-weight solutes by diafiltration of the retentate and blending of the permeate, after concentration, into the clarified juice. Other fruit juices that can be clarified by ultrafiltration include citrus, grape, cranberry, and pear.

Some of the potential benefits of ultrafiltration over the traditional processing of fruit juices are:

1. a superior-quality juice in terms of organoleptic properties and clarity can be produced;
2. it is possible to develop a continuous and automated process, thereby lowering production cost;
3. a marginally higher recovery of juice in the ultrafiltration process (95–98%) is possible in comparison with traditional processing (90–93%);
4. enzyme requirements can be reduced by more than half (as compared with the traditional process) because of partial depectinization;
5. waste disposal can also be minimized.

A major drawback of the ultrafiltration system is that it can only handle clear juices. Other problems such as reduction in flux, primarily due to fouling, is less in ultrafiltration than in microfiltration, which can be rectified by selecting the correct membrane material and configuration, and pretreating the feed.

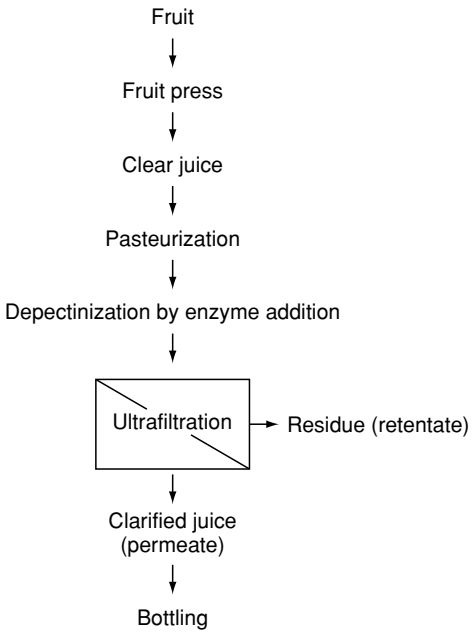

Figure 1 Flow diagram for clarification of apple juice.

Some vegetable juices, such as tomato, carrot, celery, and cucumber, can also be clarified by the ultrafiltration process with most of the advantages discussed above for fruit juices.

Wine and Beer

The application of ultrafiltration improves the organoleptic properties and stability of wine and beer. In wine production, a clear quality product is insured by treating either the must or the finished wine with ultrafiltration so as to remove undesirable microorganisms, stabilize proteins, and reduce the decolorizing and astringent-tasting polyphenol compounds, such as tannins. Other undesirable components, which otherwise require more expensive and time-consuming steps in conventional methods, are also removed by ultrafiltration membranes. (*See* **Beers**: Chemistry of Brewing; **Wines**: Production of Table Wines; Production of Sparkling Wines.)

Vinegar is produced by fermenting wine into acetic acid. After fermentation vinegar shows a slight turbidity due to low levels of polysaccharide and protein. Ultrafiltration is an ideal process for producing clarified vinegar with yield more than 97%. (*See* **Vinegar**.)

Biotechnology-Oriented Applications

Lately, membrane technology has been applied to the biotechnological processes. Ultrafiltration can be used to harvest microorganisms from fermentation broths with much higher recoveries than traditional filtration and centrifugation processes. For example, hollow-fiber ultrafilters can give 100% rejection of *Lactobacillus bulgaricus* cells in whey permeate fermentation broth. The fouling problem with ultrafiltration membranes during cell harvesting is less than with the microfiltration process. Ultrafiltration can be used for numerous applications related to the preparation of enzymatic hydrolysates from milk proteins, including:

1. the removal of enzyme and unhydrolyzed proteins from a reaction mixture;
2. the adjustment of the molecular-weight distribution profile of peptide mixture;
3. other compositional adjustments related to salts or free amino acids;
4. the removal of bitterness and antigenic molecules from casein and whey protein hydrolysate; and
5. the selective concentration of highly functional peptides.

Enzyme extracted from plant and animal tissues can also be purified by the ultrafiltration process. Ultrafiltration membranes containing immobilized enzymes or whole microbial cells function as continuous bioreactors. These bioreactors have been

reported to provide an opportunity for greatly improving the performance and productivity of fermentation. Ultrafiltration is also used for the production of high-quality water suitable for pharmaceutical and biotechnology purposes. The most efficient removal of pyrogens, bacteria, and particulates at low energy are the main advantages of ultrafiltration over other water treatment techniques. (*See* **Water Supplies**: Water Treatment.)

Other Applications

Other useful applications of ultrafiltration in the food industry include production of concentrates and isolates from soya, sunflower, and cotton seeds; removal of glucose from egg white and its partial concentration prior to drying; refining of sugar solutions; fractionation and concentration of gelatin, and many more.

See also: **Beers**: Chemistry of Brewing; **Cheeses**: Types of Cheese; Chemistry of Gel Formation; **Drying**: Spray Drying; **Effluents from Food Processing**: Composition and Analysis; **Milk**: Dietary Importance; **Starter Cultures**; **Vinegar**; **Water Supplies**: Water Treatment; **Whey and Whey Powders**: Production and Uses; **Wines**: Production of Table Wines; Production of Sparkling Wines

Further Reading

Cheryan M (1986) *Ultrafiltration Handbook*. Lancaster, PA, Technomic.

Cheryan M (1998) *Ultra-filtration and Micro-filtration Handbook*. Lancaster, PA: Technomic.

EI-Shibiny S, Shahein NM and Sheikh M (1996) Preparation and functional properties of ultrafiltered milk protein isolates. *Bulletin of International Dairy Federation* 311: 28–30.

Fukumoto, LR, Delaquis P and Girard B (1998) Microfiltration and ultrafiltration ceramic membrane for apple juice clarification. *Journal of Food Science* 63: 845–850.

Gauthier SF and Pouliot Y (1996) Use of ultra-filtration for the preparation of enzymatic hydrolysates from milk proteins. *Bulletin of International Dairy Federation* 311: 31–32.

Glover FA (1985) *Ultrafiltration and Reverse Osmosis for the Dairy Industry*. Technical Bulletin no. 5. Reading, NIRD.

Horton BS (1997) Whatever happened to the ultrafiltration of milk? *Australian Journal of Dairy Technology* 52(1): 47–49.

Kosikowski FV (1986) New cheese making process utilising ultra-filtration. *Food Technology* 40(6): 71–77.

Lelievre J and Lawrence RC (1988) Manufacture of cheese from milk concentrated by ultra-filtration: a review. *Journal of Dairy Research* 55: 373–392.

Novak A (1996) Application of membrane filtration in the production of milk protein concentrates. *Bulletin of International Dairy Federation* 311: 26–27.

Pal D and Cheryan M (1987) Membrane technology in dairy industry. Part II. Ultrafiltration. *Indian Dairyman* 39(8): 373–392.

Puhan Z (1996) Protein standardization. *Bulletin of International Dairy Federation* 311: 4–6.

Renner E and Abd El-Salam MH (1991) *Application of Ultra-filtration in Dairy Industry*. Barking: Elsevier.

Vasiljevic T and Jelen P (1999) Temperature effect on behaviour of minerals during ultrafiltration of skimmilk and acid whey. *Milchwissenchaft* 54(5): 243–246.

Menaquinone *See* **Vitamin K**: Properties and Determination; Physiology

MENSTRUAL CYCLE: NUTRITIONAL ASPECTS

F Ford, University of Sheffield, Sheffield, UK

Background

The effects of dietary intake on the menstrual cycle have been the subjects of a great deal of research, but many aspects are still poorly understood. The relationship between nutritional status and menarche will be reviewed, together with the influences of: body weight, body fat, dietary restraint, dietary fat and energy intake, lignans, isoflavones, vegetarianism, and PCB residues on the menstrual cycle.

Nutritional Status and Menarche

Some research shows that girls who experience early menstruation reach their mature height sooner than girls who experience late menarche (the onset of

menstrual periods). In one study, girls who ate primarily protein foods were shown to reach menarche sooner, and girls who eat primarily carbohydrates reached menarche later. However, in other studies of all the dietary variables analyzed, only energy intake was related to age at menarche. In a longitudinal study, girls who consumed more (energy-adjusted) animal protein and less vegetable protein at ages 3–5 years had an earlier menarche, and girls aged 1–2 years with higher dietary fat intakes and girls aged 6–8 years with higher animal protein intakes became adolescents with earlier peak growth. Controlling for body size, girls who consumed more calories and animal protein 2 years before peak growth had a higher peak growth velocity. These findings may have implications regarding adult diseases whose risks are associated with adolescent growth and development factors.

Body Weight and the Menstrual Cycle

It is well established that starvation and emaciation are almost invariably associated with amenorrhea (lack of menstrual periods), the most profound disturbance of the menstrual cycle. The work of Frisch in the USA identified the link between body composition and ovulation in the human female. She suggested that fat must comprise at least 22% of body weight for the maintenance of ovulatory cycles and also observed that in normal postpubertal women, fat is about 28% of body weight. It is well recognized that the relationship is to the fat content of the body rather than absolute body weight. Trained athletes of average or above average body weight may have a very low body fat content and may be oligo- or amenorrheic. Frisch observed that amongst trained athletes who became fit after a normal menarche, 60% continued regular cycles, but 40% had irregular cycles and presumably associated subfertility.

Eating disorders such as anorexia nervosa and bulimia nervosa are also causes of oligo- or amenorrhea. In anorexia nervosa, amenorrhea and failure to maintain a body weight within 15% of that expected are both diagnostic criteria. However, many women engage in pathologic dieting behaviors without meeting the current diagnostic criteria for anorexia or bulimia nervosa. Clinical eating disorders are only the most extreme form of pathologic eating attitudes and behaviors that are present in many young women. Specific food choices and nutrient intakes may be associated with altered gonadal hormone status of these dieters.

Extreme weight loss is not a prerequisite for menstrual cycle disturbances because dieting can induce missed cycles before substantial weight loss occurs. Even a very short-term (i.e., 4-day) acute energy shortage can interfere with luteinizing hormone pulsatility and thereby affect cycle function. In humans, dieting with minor or moderate weight loss has been shown to cause menstrual cycle disturbances or may be a risk factor for the development of reproductive dysfunction in normal weight healthy women. Cognitive factors may also be associated with the stability of the menstrual cycle. One such factor is cognitive dietary restraint, the perception that food intake is constantly being limited in an effort to control body weight. Menstrual differences between women with high and low restraint scores were detected in a number of studies. One study found that women with high restraint scores had significantly shorter cycle lengths, shorter luteal-phase lengths, and lower mean luteal-phase progesterone concentrations. In a group of women with a wide range of physical activity levels who were initially confirmed to ovulate normally, it was found that the luteal-phase length was shorter, without alteration of cycle length, in women with high restraint scores.

Cycle disturbances are associated with obesity as well as with energy shortages. A high prevalence of obesity among amenorrheic women was reported many years ago, and anovulatory cycles appear to be more common in obese women. Weight loss in obese women results in improved ovulation and pregnancy rates. It has been suggested that high androstenedione concentrations observed in obese women may activate the conversion of estradiol to estrone in adipose tissue. Estrone in turn may trigger higher luteinizing hormone concentrations, leading to ovarian hyperstimulation, thus increasing testosterone concentrations, resulting in anovulatory cycles.

The Nurses Health Study in the USA has also produced some interesting information about the relative risk of menstrual cycle irregularity not only in the underweight but also in the overweight woman. For women with a body mass index (BMI) below 20, at the age of 18 years, ovulatory infertility was found with a relative risk of about 1.2 compared to women with a BMI of 20–25. Interestingly, however, the relative risk of ovulatory infertility was 1.5 in those with a BMI of 28 and more than 2 in the obese group with a BMI above 30. About half of the risk is associated with polycystic ovarian syndrome, in which ovulatory infertility and obesity coexist, but there is still a doubling in relative risk of ovulatory infertility in women with a BMI above 30 who do not have ultrasonically detectable polycystic ovaries.

Energy Intake and the Menstrual Cycle

The biological regulation of appetite is currently an important topic in nutrition, because of its relationship with the increasing burden of obesity. Cyclical fluctuations in food intake occur in women across the menstrual cycle, with a periovulatory nadir and a peak in the luteal phase. These alterations in food intake, in response to ovarian steroid hormone changes, may be more than 2.5 MJ per day, with the mean reported changes shown in 19 separate studies of 1.0 MJ per day. Hormonally induced fluctuations in food intake could, therefore, contribute to energy imbalance and consequent weight gain. The regulation of food intake by menstrual cycle hormones suggests that it is essential to consider the phase of the menstrual cycle in studies of nutrient intake performed in women.

Vegetarian Women and the Menstrual Cycle

The question of whether menstrual disturbances are more common in vegetarian than in nonvegetarian women in complex. Three general mechanisms that could contribute to menstrual disturbances that may differ between vegetarians and nonvegetarians include energy imbalances associated with body-weight disturbances or exercise, psychosocial and cognitive factors, and dietary components. Although results from several cross-sectional studies suggest that clinical menstrual disturbances may be more common in vegetarians, a prospective study that controlled for many potential confounders found that subclinical disturbances were less common in weight-stable, healthy vegetarian women. Because the sample studied may not be representative of all vegetarian women, however, these results cannot be generalized. Population studies are needed to draw definitive conclusions.

Effects of Exercise on the Menstrual Cycle

Several prospective studies have shown that menstrual function does not change with exercise, provided that increases in activity are gradual, and body weight is maintained. This suggests that adequate energy availability may be a key factor in maintaining normal hormonal function with exercise.

A study was conducted to determine whether rigorous exercise training adversely affects ovarian hormone levels and bone health in cyclically menstruating trained runners. Results indicated that lower estrogen production, especially during the early follicular phase, but not progesterone, was associated with a lower whole body calcium per kilogram of soft lean tissue and probably bone mineral density.

Dietary Components and the Menstrual Cycle

Research suggests that concentrations of ovarian hormones or their metabolites may be lower at various points in the menstrual cycle in women consuming diets high in fiber, low in fat, or both. This is consistent with cross-sectional studies reporting inverse associations between hormone concentrations and fiber intake, direct associations between hormone concentrations and fat intake, and lower serum estrogen concentrations with a faster intestinal transit. It also corroborates the lower estradiol and progesterone concentrations observed in women 2 years after being randomly assigned to a low-fat, high-carbohydrate diet for a study in breast cancer risk reduction. However, the data must be interpreted cautiously because many experimental studies were conducted over only one or two menstrual cycles, and other data suggest that changes observed acutely may not persist over time. In addition, some subjects were free-living and others were housed in metabolic wards, which could affect the menstrual cycle differently.

Studies have been conducted to explore the effects of two classes of phytoestrogens, lignans and isoflavones. One study examined the effects of flaxseed powder, which is high in lignan precursors, in 18 women with normal ovulatory cycles using a randomized, crossover design. Each women followed her usual omnivorous, low-fiber diet for three cycles and then followed her usual diet supplemented with flaxseed powder for another three cycles. The last two cycles of each dietary period were compared. During flaxseed supplementation, luteal-phase lengths were significantly longer, and the ratio of progesterone to estradiol during the luteal phase was higher. Moreover, no anovulatory cycles occurred during flaxseed supplementation, but three anovulatory cycles occurred during the control diet period.

In another study, the effects of isoflavones were examined in six women with regular ovulatory cycles. It was found that diets containing 60 g of soy protein per day (with 45 mg of isoflavones) significantly increased follicular-phase length compared with control diets. Subsequently, they showed that a soybean product from which the isoflavones had been extracted had no effect on the cycle. Thus, lignans and isoflavones, two different classes of phytochemicals, appeared to have different effects: flaxseed lignans increased luteal-phase length and did not affect follicular-phase length, whereas soybean isoflavones

increased follicular-phase length but did not influence luteal-phase length.

Previous studies have suggested that high soybean intakes are associated with lower concentrations of serum cholesterol and may be related to decreased rates of coronary artery disease, cancer, and osteoporosis. One of the postulated mechanisms of these effects is that soy products, particularly the isoflavones in soy, act through changes in endogenous hormonal balance. Isoflavones significantly improved the lipid profile across the menstrual cycle in normocholesterolemic, premenopausal women. Although of small magnitude, these effects could contribute to a lower risk of developing coronary heart disease in healthy women who consume soy over many years.

Relationship Between Diet, Menstrual Hormones, and Breast Cancer

There is compelling evidence linking ovarian hormonal activity to breast cancer risk. Since the mid-1980s, dietary fat intervention studies have been conducted to investigate the effect of fat intake on endogenous estrogen levels. A meta-analysis of dietary fat intervention studies that investigated serum estradiol levels reviewed the nature of the evidence provided by prospective analytic studies of fat consumption and breast cancer risk. Results indicated that dietary fat reduction can result in a lowering of serum estradiol levels and suggest that a low-fat high-carbohydrate diet may reduce the risk of breast cancer by reducing exposure to ovarian hormones that are a stimulus to cell division in the breast.

Intake of soybean protein is associated with a reduced risk of breast cancer in a case-control study. It has also been demonstrated to increase menstrual cycle length in an experimental setting.

Most epidemiological studies of the relationship between alcohol consumption and breast cancer risk over the past decade have shown that persons who consume a moderate amount of alcohol are at 40–100% greater risk of breast cancer than those who do not consume alcohol. Dose–response effects have been observed, but no causal relationship has been established. However, it has also been shown that alcohol consumption in premenopausal women is associated with increases in total estrogen levels and amount of bioavailable estrogens. This may help to explain the relationship between the risk of breast cancer and alcohol intake.

Dietary Contaminants and the Menstrual Cycle

Highly contaminated Lake Ontario sport fish represent an important human dietary exposure to polychlorinated biphenyls (PCBs) and other toxic contaminants that may disrupt endocrine pathways. In one study, women anglers were interviewed by telephone to determine menstrual cycle length and fish consumption in order to calculate a PCB exposure index. Multiple regression analyses identified significant cycle-length reductions with consumption of more than one fish meal per month and a moderate/high estimated PCB index. Women who consumed contaminated fish for 7 years or more also had shorter cycles.

Menstrual Cycle and Iron-deficiency Anemia

Menorrhagia (heavy menstrual bleeding) is a benign yet debilitating social and health condition. The widely accepted clinical definition of menorrhagia is blood loss of 80 ml or more per period. This figure is derived from population studies that have shown that the average blood loss is between 30 and 40 ml, and 90% of women have blood losses of less than 80 ml. Excessive menstrual bleeding is the commonest cause of iron deficiency in the UK, affecting 20–25% of the fertile female population. Menorrhagia is a common problem accounting for 12% of all gynecological referral in the UK.

Conclusion

The evidence presented above appears primarily to show associations between dietary factors and the menstrual cycle rather than causation. What does, however, seem clear is that the relationship between normal ovarian hormone function and diet is complex and needs further evaluation from prospective rather than epidemiological research studies.

See also: **Anemia (Anaemia)**: Iron-deficiency Anemia; **Anorexia Nervosa**; **Bulimia Nervosa**; **Exercise**: Metabolic Requirements; **Obesity**: Etiology and Diagnosis; **Vegetarian Diets**

Further Reading

Barnard ND, Scialli AR, Bertron P, Hurlock D, Edmonds K and Talev L (2000) Effectiveness of a low-fat vegetarian diet in altering serum lipids in healthy premenopausal women. Randomized Controlled Trial. *American Journal of Cardiology* 85(8): 969–972.

Barr SI (1999) Vegetarianism and menstrual cycle disturbances: is there an association? *American Journal of Clinical Nutrition* 70 (3 supplement): 549S–554S.

Boyd NF, Lockwood GA, Greenberg CV, Martin LJ and Tritchler DL (1997) Effects of a low-fat high-carbohydrate diet on plasma sex hormones in premenopausal women:

results from a randomized controlled trial. Canadian Diet and Breast Cancer Prevention Study Group. *British Journal of Cancer* 76(1): 127–135.

Bruinsma K and Taren DL (1999) Chocolate: food or drug? *Journal of the American Dietetic Association* 99(10): 1249–1256.

Campbell WS, Franz C, Kahle L and Taylor PR (1996) Relation of energy, fat, and fiber intakes to plasma concentrations of estrogens and androgens in premenopausal women. *American Journal of Clinical Nutrition* 64(1): 25–31.

Doll H, Brown S, Thurston A and Vessey M (1989) Pyridoxine (vitamin B_6) and the premenstrual syndrome: a randomized crossover trial. *Journal of the Royal College of General Practitioners* 39(326): 364–368.

Eck LH, Klesges RC, Meyers AW, Slawson DL and Winders SA (1997) Changes in food consumption and body weight associated with smoking cessation across menstrual cycle phase. *Addictive Behaviors* 22(6): 775–782.

Frisch RE (1994) The right weight: body fat, menarche and fertility. *Proceedings of the Nutrition Society* 53(1): 113–129.

Gendall KA, Bulik CM, Joyce PR, McIntosh VV and Carter FA (2000) Menstrual cycle irregularity in bulimia nervosa. Associated factors and changes with treatment. *Journal of Psychosomatic Research* 49(6): 409–415.

Jones DY, Judd JT, Taylor PR, Campbell WS and Nair PP (1987) Influence of dietary fat on menstrual cycle and menses length. *Human Nutrition – Clinical Nutrition* 41(5): 341–345.

Kato I, Toniolo P, Koenig KL *et al*. (1999) Epidemiologic correlates with menstrual cycle length in middle aged women. *European Journal of Epidemiology* 15(9): 809–814.

Lu LJ, Anderson KE, Grady JJ, Kohen F and Nagamani M (2000) Decreased ovarian hormones during a soya diet: implications for breast cancer prevention. *Cancer Research* 60(15): 4112–4121.

McLean JA, Barr SI and Prior JC (2001) Cognitive dietary restraint is associated with higher urinary cortisol excretion in healthy premenopausal women. *American Journal of Clinical Nutrition* 73(1): 7–12.

Mendola P, Buck GM, Sever LE, Zielezny M and Vena JE (1997) Consumption of PCB-contaminated freshwater fish and shortened menstrual cycle length. *American Journal of Epidemiology* 146(11): 955–960.

Norman RJ and Clark AM (1998) Obesity and reproductive disorders: a review. *Reproduction, Fertility, & Development* 10(1): 55–63.

Thys-Jacobs S (2000) Micronutrients and the premenstrual syndrome: the case for calcium. *Journal of the American College of Nutrition* 19(2): 220–227.

MERCURY

Contents

Properties and Determination

C J Cappon, Cappon Associates, Henrietta, NY, USA

Introduction

Mercury is probably the most ubiquitous heavy metal in the environment, resulting from natural geological activity and industrial pollution. Although it is extremely useful, it is also highly toxic, depending on its specific chemical form. The hazards of mercury and its compounds have been long known and of much concern. Its history as an occupational hazard has been linked to its elemental and divalent ionic – inorganic – forms, mostly due to volatility and skin contact. However, there has been much recent concern for widespread environmental contamination from organic mercury species, namely the alkylmercuries – methyl, ethyl-, and phenylmercury. Alkylmercury poisoning has been almost exclusively linked to methylmercury pollution of waterways and aquatic food sources, specifically in Japan, Sweden, and Canada. This has initiated several comprehensive worldwide analytical surveys of foods to assess human mercury intake levels and potential adverse health effects. It has also highlighted the critical need for more accurate and reliable mercury analytical methods for a variety of materials. Most modern methods involve mercury speciation separation schemes. This article discusses the environmental sources and toxicological impact of mercury in food, with special emphasis on analytical speciation

methodology – sample preparation procedures and instrumental techniques.

Uses of Mercury

Mercury has been utilized since antiquity. Although a comparatively rare element – ranking 16th from the bottom of natural element abundance – almost 3500 applications for mercury compounds are currently recognized. The major uses of mercury are summarized in Table 1.

Environmental Input and Speciation

Mercury in food can come from natural or human activities. Mercury from natural activities can enter air, soil, and water via weathering, dissolution, vaporization, and biological processes. Three oxidation states are possible in nature: metallic (Hg^0), mercurous (Hg_2^{2+}), and mercuric (Hg^{2+}). The two main categories of human mercury release are: (1) agriculture – run-off of excess pesticide application and absorption by vegetation roots and foliage; and (2) industry – waste-water discharge, disposal of solid waste containing mercury residues (thermometers, fluorescent lamps, labware, paints), burning of coal and petroleum-based materials.

Table 2 lists the specific mercury species that have been identified in air, soil, and water. Water is the most significant and direct source of food mercury contamination. Of special importance are the factors which favor highly toxic methylmercury formation in soil and water (including sediment), resulting in subsequent uptake by edible vegetation and animal food sources. In addition, biotransformation of methylmercury from inorganic mercury (and vice versa) can occur in specific marine and terrestrial food animal species.

Table 1 Uses of mercury

Dissipative	Paints, agriculture, dental, catalyst, pharmaceuticals, pulp and paper, metallurgy and mining
Recyclable	Chlorine and caustic soda production, electrical, measurement, laboratory

Table 2 Environmental mercury contamination: input sources

Agriculture	Fungicides and seed preservatives
Industrial	Chlorine and caustic soda production; amalgamation, pulp and paper preservative; catalyst for plastic production, pharmaceuticals (drugs, antiseptics); prints (fungicide), electrical and measurement instruments (thermometers, switches, neon lamps); power production (coal burning)
Geological	Weathering of rocks, volcanic activity

Occurrence and Speciation in Food

Mercury occurs in a wide variety of foods, most commonly fish and seafood, where the typical total mercury concentration range is 0.05–3.0 p.p.m. (fresh weight). Grain and dairy products, animal meats, and certain vegetable crops usually have a lower and narrower range (0.01–0.05 p.p.m.). There is no reported evidence of external mercury contamination of food during processing, packaging, or storage, so that food mercury content results solely from the natural and human sources previously identified.

In the mid-1960s, methylmercury was identified as the major mercury species in fresh-water and marine fish. Since the early 1970s, several comprehensive analytical surveys of total, inorganic, and methylmercury content of fish and other types of seafood have been undertaken. Methylmercury levels in the edible fillet of most fish species is typically 70–95% of the total mercury content. Methylmercury contents below 50% have been reported for a limited number of fish samples. A notable example is Pacific blue marlin, the only pelagic species having a large amount (> 80%) of total muscle mercury present as inorganic compounds. Inorganic mercury also predominates in shellfish (molluscs, crustaceans) and other nonfish species (e.g., octopus, squid). Other common organomercurials – ethylmercury and phenylmercury – are generally not found in seafood. (*See* **Fish**: Spoilage of Seafood; **Shellfish**: Contamination and Spoilage of Molluscs and Crustaceans.)

Mercury content and speciation have also been reported for other food sources. Some examples include marine mammals (seal, whale), terrestrial wild and domestic animals (chicken, otter, beaver, deer), and vegetable crops. Regarding edible vegetation, mercury uptake and methylmercury percentage are enhanced in the tuberous and leafy regions of crops grown on sludge- and compost-treated soil. Compared to the relatively low mercury concentrations in edible vegetation, mushrooms and related fungi accumulate higher amounts of total mercury and methylmercury.

Table 3 summarizes mercury levels and speciation in a variety of foods.

Impact of Mercury in the Environment

Knowledge of mercury speciation is very critical when evaluating the impact of environmental mercury pollution and consumption of mercury-contaminated food, since different chemical forms of

Table 3 The distribution of mercury in various foods

Food	*Total mercury concentration ($\mu g\,kg^{-1}$ fresh weight)*	*Methylmercury content (%)*
Vegetables and fruits[a]		
Cabbage	4.1–9.0	25–29
Beans	0.4–1.0	ND
Tomatoes	0.1–8.0	13
Peppers	1.7	13
Potatoes	0.1–1.7	ND
Onion	0.5–5.1	2.4
Lettuce	0.1–12.2	9–17
Squash	0.7–1.6	1–7
Cucumber	0.1–2.0	2
Broccoli	1.8–5.1	6–8
Carrot	0.6–5.0	15
Garlic	2.1–4.8	24
Spinach	0.5–20.0	7–11
Grains		
Wheat	< 0.1–46	NA
Oats	1.0–30	NA
Other		
Mushrooms	1–16	< 1–8

[a]Edible parts.
ND, not determined; NA, not available.

mercury have different metabolic and biotransformation properties which influence their overall toxicity.

Historically, mercury poisoning has been a well-documented occupational hazard. However, it was only well into the twentieth century that specific outbreaks of environmental mercury poisoning demonstrated that specific mercury species could exert serious health effects on exposed human populations. Examples of major outbreaks are: (1) Japan (Minamata, Niigata), from the mid-1950s to the early 1960s from ingestion of contaminated fish and shellfish from waters polluted by industrial discharges; (2) Sweden and Canada, in the mid- to late 1960s, from similar but milder cases of industrial pollution; and (3) Iraq in the early 1970s, from ingestion of grain and grain products originating from seed treated with a mercury fungicide. In each case, methylmercury was the causative agent.

Environmental mercury – whether from natural or human activities – will continue to enter the human food chain. Presently, the aquatic food chain is more susceptible than the terrestrial food chain. However, the ability of mercury to undergo biotransformation to more toxic organic species in various environmental compartments (air, soil, aquatic sediment, water, animal tissue) justifies the essential need to evaluate mercury speciation when assessing mercury contamination of foodstuffs. Therefore, any analytical methodology for mercury must be capable of qualitative and quantitative determination of specific mercury species found in food.

Properties and Effects of Food Preparation

Introduction

Environmental mercury species that eventually become incorporated in food may be classified as inorganic or organic. This section considers the specific physical and chemical properties of the various forms that would also allow the element to enter the food chain. The effect of food-processing methods on mercury content and speciation will also be discussed.

Properties of Mercury Species

Inorganic The two inorganic forms are elemental and ionic, the latter commonly occurring as the divalent species Hg^{2+}. The high vapor pressure of elemental mercury renders it quite volatile from water and soil surfaces. However, it is readily soluble in these media in the whole range. It is readily oxidized to Hg^{2+} by bacterial action or chemical oxidizing agents. Divalent inorganic mercury can form insoluble salts with the ions O^{2-}, Cl^{-}, I^{-}, and S^{2-}, rendering it less mobile and bioavailable to food-chain organisms. However, its solubility and bioavailability can be enhanced through complex formation with Cl^{-}, CO_3^{2-}, and OH^{-} ions, with ammonia and chelation with humic substances. An acidic pH environment will also enhance its solubility. Divalent mercury can be reduced to elemental mercury by interaction with bacteria, specific inorganic reducing agents, and sunlight. It also binds strongly to sulfhydryl (– SH) ligands on plant and animal tissue proteins. Monovalent mercury (Hg_2^{2+}) can originate in the environment from Hg^{2+} disproportionation, but the predominance of oxidizing conditions in water and soil insures its transformation to Hg^{2+} species.

Organic Of all the possible organomercurials, the monomethylmercury ion (CH_3Hg^{+}) is the most common environmental species. It originates from Hg^{2+} through bacterial synthesis, organic oxidants, and chelation. Its formation from Hg^{2+} in natural waters is optimized at low salinity. It also has a high affinity for sulfhydryl ligands on tissue proteins, but these protein complexes rapidly break down at very low pH. Its solubility – and bioavailability – in soil and water is enhanced by its ability to

complex with anions (mostly OH^-, Cl^-) and humic materials. Under conditions of very low (< 2.0) or very high (> 11.0) pH, the methylmercury ion is converted to inorganic mercury via bacteria or by sunlight. Dimethylmercury is the only divalent organic form commonly reported. Its higher volatility allows it to exist mostly in air. However, acidic conditions cause it to convert readily to monomethylmercury.

Impact of Food Preparation

Little is known about the influence of food processing (dehydration, steaming, preservation, and other additives), storage (packing medium and container material, temperature, time period), and cooking (backing, freezing, smoking) techniques on mercury content and speciation in food. Infact, little is known about the exact chemical nature of inorganic mercury and methylmercury in food that may influence their relative bioavailability.

Processing Research from the author's laboratory concerning the distribution of inorganic mercury and methylmercury in processed and nonprocessed fish and seafood revealed different mercury distribution patterns. Freshwater and older processed (canned, > 5 years' storage) marine fish had the majority (55–90%) of total mercury tissue content in a water-extractable form. Only 22–47% was extractable for nonprocessed and newly processed (< 2 years' storage) samples. Water-extractable mercury is of interest because it represents mercury bound to lower-molecular-weight tissue proteins or unbound (neutral, ionic) species. Such species may be more readily absorbed from the intestine into the blood stream, thus having higher bioavailability. These specific mercury distribution patterns possibly result from differences in the mercury source and form in marine and fresh-water food chains.

Storage The above marine fish mercury distribution pattern suggests that food storage time may have a greater impact on tissue mercury form and distribution than food processing. Although processing may initiate a change in mercury distribution, further changes can occur over an extended storage period. Samples of canned tuna and seafood showed no apparent effect of packing medium (water, oil) on the percentage of water-extractable mercury. Both types of packing media contained mostly inorganic mercury, but this represented a very small fraction (generally < 2%) of the corresponding total mercury sample content.

Cooking Broiling, baking, or smoking fish meat does not alter the total and methylmercury content. However, smoking increased the methylmercury percentage in brown trout fillets. (*See* **Smoked Foods: Principles.**)

Analysis: Sample Preparation

Introduction

Sample preparation is a key component in any analytical procedure, yet it is often the most time-consuming to conduct and optimize. Ideally, it should be simple, rapid, reproducible, provide quantitative analyte recovery, and effectively remove all interfering substances prior to instrumental assay. The specific sample preparation procedure used depends on: (1) complexity of the sample matrix; (2) physical and chemical properties of the desired mercury species; and (3) the instrumental technique. For food analysis, the sample matrix may be relatively simple (water, beverage) or quite complex (vegetation, soft tissue). For total mercury analysis of food, sample preparation is greatly simplified by the use of mineral acid (nitric, perchloric, sulfuric). However, mercury speciation usually necessitates more elaborate solvent extraction, clean-up, and analyte isolation steps.

Sample Preservation and Storage

Preservation of total mercury concentration in aqueous samples is critical, especially at microgram per kilogram levels. The major mercury loss occurs by conversion of Hg^{2+} (and Hg_2^{2+}) to elemental vapor and subsequent loss to air. Another source of loss is from adsorption of mercury on the container walls. Several factors affect mercury preservation, and include mercury concentration, container type, method of container cleaning, and types of oxidizing/complexing agents used as preservatives. Generally, low pH, high ionic strength, acidic (nitric or sulfuric acid) dichromate as an oxidizer/preservative insures optimum mercury preservation. Acid-washed containers of Teflon, polypropylene, polyethylene, Pyrex, and soft glass are satisfactory for mercury storage and preservation. Methylmercury preservation in aqueous solutions requires storage in acid-washed polytetrafluoroethylene (PTFE) containers under slightly acidic pH as optimum conditions for long-term preservation.

The need for mercury preservation in solid food samples is less demanding since nearly all of the mercury – organic and inorganic – is tightly bound to tissue proteins, which minimizes absorption and

volatilization losses. However, to minimize mercury biotransformation – especially organomercury breakdown – by bacterial activity, all food and beverage samples should be kept frozen (−20 °C) until analysis preparation.

Sampling

Sample homogenization is a convenient way to insure that a more representative subsample will be obtained for analysis. This is commonly achieved using a blender and/or homogenizer, after which the sample is sieved using 30–40 mesh. Deionized, organic-free water is then added and the homogenate is further mixed to a more uniform texture. However, care must be taken not to contaminate the food sample through contact with the container walls. A more recent sampling procedure which further enhances sample homogeneity is digestion with a mixture of 40–45% sodium hydroxide (50%), 1% cysteine, and 1% sodium chloride. This is the basis for preliminary sample preparation in several (vapor atomic absorption spectroscopy (VAAS) and gas–liquid chromatography (GLC)) extraction procedures. Alternatively, for samples having mercury levels approaching the lower detection limit of a given method, lyophilization provides a dry powder more concentrated in mercury.

Sample Extraction

Total mercury Most food sample preparation procedures for total mercury involve various degrees of chemically treating the sample (i.e., wet ashing). This is mandatory since inorganic and organic mercury are essentially volatilized by dry ashing. Wet ashing with mineral acids usually destroys organic matter, which can sometimes produce interferences, depending on the instrumental technique being used. Acids and acid mixtures that have been used successfully and produce negligible mercury loss are: perchloric, nitric, sulfuric; nitric/sulfuric/perchloric, nitric/perchloric, sulfuric/nitric. With the acid mixtures oxidants such as hydrogen peroxide, permanganate, and vanadium oxide have also been used. Loss of mercury can be minimized by refluxing/condensing the acidic digest vapors.

Mercury speciation Since organomercury species are readily degraded to inorganic mercury under the drastic acidic and oxidizing conditions for total mercury, much milder digestion conditions are required. In food samples, any protein-bound inorganic mercury and organomercury requires an acidic medium (usually hydrochloric, hydrobromic, and sulfuric acids are employed) for liberation prior to solvent extraction. Depending on the complexity of the original sample, one or more organic solvent cleaning steps are usually required. Aqueous or ethanolic thiosulfate is the most widely used cleanup reagent because of its rapid quantitative extraction of organomercurials. The organic mercury species are usually isolated either as a volatile halide derivative (chloride, bromide, iodide for GLC assay) or as a water-soluble thiol complex (for high-performance liquid chromatography (HPLC) assay). Typical overall organic mercury recoveries are 70–100%.

Inorganic mercury assay by GLC requires prior derivatization of the nonvolatile inorganic species into a volatile organic form. Reagents that have been used for this purpose are listed in Table 4. Cold VAAS (CVAAS) employs specific reagents – tin chloride and cadmium chloride – to convert separately inorganic mercury and both inorganic and organic mercury, respectively, to elemental mercury.

All of the above wet-ashing procedures can be partially or fully automated. This has been successfully implemented for CVAAS. Semiautomated headspace analysis has been used for GLC analysis of methylmercury in biological samples.

Table 4 Modern mercury speciation techniques

Technique	*MDL*
Gas-liquid chromatography (GLC)	
Detector	
Electron capture (ECD)	5–200 pg
Atomic absorption (AA)	0.1 ng
Microwave-induced plasma (MIP)	0.2–0.6 pg
Direct-current plasma (DCP)	0.5 ng
High-performance liquid chromatography (HPLC)	
Detector	
Electrochemical (EC)	40 pg
Atomic absorption (AA)	0.1 ng
Inductively coupled plasma (ICP)	0.3–0.7 ng
Ultraviolet (UV)	0.5 ng
Atomic absorption spectroscopy (AAS)	
Mode	
Cold vapor	0.02 ng
Graphite furnace	0.05 ng

MDL, minimum detection limit.

Analysis: Instrumental Techniques

Because of environmental and toxicological considerations, most modern mercury analysis methods for food emphasize speciation instead of total element assay. Therefore, this review will highlight modern mercury speciation techniques and summarize classical – total mercury – techniques.

Total Mercury Techniques

Numerous analytical methods for total mercury in biological media are summarized in Table 5.

Table 5 Early mercury analytical techniques

Total mercury	*Chromatographic speciation*
Titrimetric (indicator, thermometric, amperometric, potentiometric)	Ion exchange
Neutron activation analysis	Thin-layer
Spectrophotometry	Paper
Catalysis	Column
X-ray fluorescence	Radiochemical
Polarography (alternate current, anodic stripping voltametry)	
Atomic spectroscopy	
Absorption (flame, furnace, cold vapor)	
Emission (inductively coupled plasma, fluorescence)	

While many of these methods are still reliable, they suffer from major limitations which render them unsuitable for routine analysis of food samples: (1) they cannot accurately detect or measure mercury at low levels (micrograms per kilogram), which are normally background levels in foods; (2) they cannot be adopted for speciation; and (3) most are time-consuming and are subject to numerous inorganic and organic interferences from the sample matrix. In addition, some techniques – X-ray fluorescence (XRF), neutron activation – require the use of expensive specialized equipment and skilled operators.

Mercury Speciation Techniques

Chromatography Chromatographic methods provide specific identification of the form in which mercury is present. Early methodology in this area was primarily intended for sample purification prior to mercury determination by other methods. These early techniques are listed in Table 5. The microgram detection limits and time requirements have greatly restricted their routine application. Modern chromatographic mercury speciation techniques for food analysis are GLC and HPLC. Both are highly powerful and sophisticated techniques which demand more technical attention and skill for optimization and routine operation. However, they can be highly automated in terms of sample preparation and introduction, operation, and data processing. Each technique is summarized in Table 4.

GLC has been the most widely used chromatographic method for routine mercury speciation. It was initially used for this purpose in the mid-1960s, being applied to water, fish, and related seafood samples. The major focus was on methylmercury, the most toxic chemical form. Eventually, ethyl-, phenyl-, and inorganic mercury were also included. Currently, conventional (i.e., packed) column chromatography coupled with electron capture detection (ECD) is still the most commonly used GLC system. However, recent significant improvements in GLC mercury speciation involve the use of capillary columns and element-specific detectors. Capillary columns exhibit enhanced performance over packed columns – greater separation efficiency and shorter analysis times. Wide-bore (0.53 mm internal diameter) based silica capillary columns are the most popular compared to narrower-bore and borosilicate glass columns. During the 1980s atomic emission and absorption spectroscopic detectors began to be used more for GLC mercury speciation. These detectors provide better elemental selectivity and sensitivity compared to ECD, along with freedom from contaminant interferences. (*See* **Chromatography**: Gas Chromatography.)

HPLC mercury speciation application to food analysis has been more recent. However, its use has increased significantly since the late 1970s. Key reasons for this are due to the advantages of HPLC over GLC: (1) no analytical restrictions on sample volatility and thermal stability; (2) more versatility in optimizing instrument operation parameters because of specific separation modes (normal-phase, reversed-phase, ion exchange), wide selection of mobile phases, and use of solvent and flow gradient programming; (3) sample derivatization can be performed using precolumn, on-column (*in situ*) or post-column techniques. This greatly simplifies sample preparation; and (4) larger sample extract volumes can be injected, typically, up to 250 μl against 10 μl and 1 μl, respectively, for packed and capillary column GLC. Recently, the availability and application of more sensitive and element-specific detectors have allowed HPLC to be a more viable mercury speciation alternative to GLC. These include electrochemical (EC) and atomic spectroscopic (AAS, inductively coupled plasma (ICP), and microwave-induced plasma (MIP)) detection. Because of cost, simplicity, and commercial availability, the EC detector is the most widely used. Currently, reversed-phase chromatography (RPC) using conventional analytical columns with 5 or 10 μm C_{18} packings is the most commonly used HPLC separation mode for mercury speciation. The polar aqueous-based mobile phase is compatible with the newer selective detectors and offers greater flexibility in optimizing analyte resolution. (*See* **Chromatography**: High-performance Liquid Chromatography.)

Atomic Spectroscopy

Two modes of this technique – CVAAS and graphite furnace (GFAAS) – have been successfully applied to

routine assay of biological materials containing low levels (microgram per kilogram) of organic and inorganic mercury. While neither mode can identify the exact form of organic mercury, they are still useful if one considers that methylmercury is by far the predominant organomercury species in foods and most other biological materials. CVAAS is the more commonly used technique. It selectively reduces inorganic mercury to mercury vapor via tin chloride, while both inorganic and organic mercury (i.e., total mercury) are reduced via cadmium chloride with tin chloride. Organic mercury is calculated by difference. The total sample aliquot as its acid or alkaline digest is normally used. In GFAAS, mercury (organic/inorganic) is released as mercury vapor during absorption or thermal decomposition of the total sample aliquot. However, it requires prior separation of both mercury forms using extraction procedures similar to those of GLC, and the use of special matrix modifiers to facilitate volatilization removal of chemical interferences.

Both techniques have distinct advantages over GLC and HPLC in the area of mercury speciation: (1) the entire sample/aliquot can generally be used; (2) they are less subject to matrix interferences, thus simplifying sample handling and eliminating the need for solvent clean-up. They can also be more extensively automated regarding sample handling, mixing, and injection. Increased sample throughput is also a major advantage.

Recently, isotope dilution inductively coupled plasma–mass spectroscopy (ICPMS) using flow injection analysis has been successfully applied to determining organic mercury in seafood. This technique offers lower detection limits, freedom from spectral interferences, and rapid sample throughput. As with GFAAS, it does require prior isolation of RHgCl from the sample. (*See* **Mass Spectrometry**: Principles and Instrumentation.)

Mercury's environmental ubiquity and specific chemical properties allow the element ease of entry to the human food chain in several chemical forms. The resulting long-term health and economic impact man-dates the use for mercury analysis of modern instrumental analysis techniques which are capable of qualitative and quantitative mercury speciation at submicrogram per kilogram levels. Currently, such methods include GLC, HPLC, and flameless AAS (CVAAS, GFAAS).

See also: **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **Mass Spectrometry**: Principles and Instrumentation; **Shellfish**: Contamination and Spoilage of Molluscs and Crustaceans; **Smoked Foods**: Principles

Further Reading

Cappon CJ (1987) GLC speciation of selected trace elements. *LC-GC* 5: 400–418.

Cappon CJ (1988) HPLC speciation of selected trace elements. *LG-GC* 6: 584–599.

Cappon CJ (1990) Speciation of selected trace elements in edible seafood. In: Hriagu JO and Simmons MS (eds) *Food Contamination from Environmental Sources*, pp. 145–193. New York: Wiley.

Fishbein L (1971) Chromatographic and biological aspects of inorganic mercury. *Chromatographic Reviews* 15: 195–238.

Krenkel PA (1974) Mercury: environmental considerations. Part II. *CRC Critical Reviews in Environmental Control* 4: 252–314.

Shrivastava AK and Tandon SG (1982) The determination of mercury: a mini review. *Toxicological and Environmental Chemistry Reviews* 5: 311–329.

Uthe JF and Armstrong FAJ (1974) The microdetermination of mercury and organomercury compounds in environmental materials. *Toxicological and Environmental Chemistry Reviews* 2: 45–77.

VanLoon JC (1985) Selected methods of trace analysis: biological and environmental samples. In: Elving PJ, Wine-fordner JD and Kothoff IM (eds) *Chemical Analysis*, vol. 80. New York: Wiley.

Toxicology

G L Klein and W R Snodgrass, The University of Texas Medical Branch, Galveston, TX, USA

Introduction

This section will cover the absorption and bioavailability of mercury, its distribution within and excretion from the body, the mechanisms involved in its harmful effects, toxic levels in humans, and accepted treatment.

Absorption and Bioavailability

Since mercury is a toxic heavy metal with no known role in normal metabolism, absorption must be interpreted in the broadest possible terms. Vapor inhalation, enteric ingestion, and skin contact are the chief routes by which mercury enters the body. A more recent concern has been mercury contamination of hepatitis B immune globulin as well as the adjuvant for the hepatitis B vaccine. Inhalation of mercury

vapor may come from its release from 'silver' dental amalgams, which are approximately 50% mercury, and from which vapors are released into the mouth by chewing or brushing the teeth. Humans have been documented to absorb a mean of 10 μg day^{-1} from this source, with a range of 1.2–27 μg day^{-1}. Metallic mercury, as a contaminant of latex paints, batteries, or electrical industrial equipment, vaporizes and, either in its metallic or inorganic form, is almost totally absorbed across the pulmonary alveolar membrane, perhaps owing to its high lipid solubility. (*See* **Heavy Metal Toxicology.**)

Toxicology

The mechanism of mercury absorption across the skin from topical application of various fungicidal preparations is uncertain. Absolute absorption rate increases directly with increasing topical mercury concentration, until a plateau is reached. (*See* **Fungicides.**)

A major organic salt of mercury, methylmercury, is approximately 100% absorbed in the intestines of humans and mice, while inorganic mercury salts (mercurous and/or mercuric) are only about 15% absorbed in human volunteers. Mercury is primarily concentrated in the aquatic food chain as a result of release of inorganic mercurials as industrial contaminants into either salt or fresh water. The inorganic mercury compounds are acted upon by methanogenic bacteria, producing methyl mercury. As such, it is taken up by various forms of sea life and enters the human food chain by direct consumption of fish and other seafood. It has been suggested that high selenium concentrations in fish can be protective against methyl mercury toxicity. There is intracellular colocalization of mercury and selenium and both have been reported to be bound in a complex with a plasma protein. Methylmercury can also enter the human food chain indirectly by means of human consumption of poultry and hens' eggs when the poultry have been raised on mercury-contaminated fish meal. (*See* **Fish Meal.**)

Distribution and Excretion

It is thought that the broadest exposure to mercury now comes from mercury vapors from dental amalgams. Studies in sheep and cynomolgus monkeys have demonstrated large concentrations of mercury in the mandible, kidneys, and liver. Significant concentrations have also been found in brain tissue. In pregnant sheep, fetal tissues begin to accumulate mercury within days of amalgam placement in the mothers and fetal tissue accumulation increased progressively during gestation. In monkeys a significant quantity of mercury has also been found in bile and in the colon. With methylmercury the distribution is similar but more uniform. There is a significant difference between methylmercury and inorganic mercury with regard to the red cell-to-plasma ratio (i.e., the distribution between red blood cells and plasma), with methylmercury in erythrocytes approximately 10-fold higher than inorganic mercury. This difference in concentration may affect the interpretation of urinary mercury excretion since it may affect the amount of mercury available for filtering through the kidneys. Furthermore, there is a great deal of interspecies variability in the red cell-to-plasma concentration of methylmercury, with the rat having the greatest and humans the least. Interestingly, the rat has the highest blood–brain mercury ratio, while humans have the lowest.

While mercury is excreted primarily by the urine, its appearance in the bile and colon after ingestion of dental amalgam vapor by monkeys suggests that the intestine is another source of mercury elimination. The mercury body burden varies directly with the excretory half-time. Thus, humans, with an excretory half-time of 70 days, develop steady-state tissue levels of mercury 10 times that of the mouse, which has an excretory half-life of only 7 days. However, this is merely an illustrative point because, based on autopsy data of brain mercury concentrations in two human beings after 15–25 years of Minamata disease, the true mercury excretory half-life in humans may be much longer than 70 days.

The excretion of mercury has been studied in rats, and three steps have been identified. The initial rapid phase involves approximately one-third of the total dose, lasts a few days, and is associated with high concentrations of mercury in liver and feces. The two slower phases involve the remaining two-thirds of the mercury intake and have a combined excretory half-life of approximately 130 days. These slower excretory phases are primarily associated with urinary excretion of mercury. Mercury is probably also excreted by the renal tubules and not the glomeruli. Mercury may bind to sulfhydryl groups within the tubular cells and, during excretion, damage to the tubular cells could trigger the exfoliation of mercury-containing tubular cells with a resultant increase in urinary mercury excretion.

Binding to metallothionein may play a protective role in both liver and kidney. In liver, biliary mercury excretion is reduced with a high-zinc diet, suggesting that zinc induces metallothionein to bind mercury and the biliary excretion of mercury is primarily related to nonmetallothionein-bound mercury. In kidney, administration of mercury to rats is

associated with an increase in urinary metallothionein. (*See* **Zinc**: Physiology.)

Mechanisms of Toxic Effects

Once absorbed from the digestive and respiratory tracts, or absorbed transcutaneously, mercury produces a variety of symptoms affecting the marrow, kidneys, skin, and blood cells. In children it produces a syndrome complex called acrodynia, or 'pink' disease. Acrodynia is characterized by redness of the lips and pharynx, loss of teeth, a strawberry tongue, swelling, redness, and desquamation of the skin, with pink or red fingertips, palms, and soles. Redness of the conjunctival membranes, photophobia, cervical lymph node enlargement, joint pain, loss of appetite, and vascular thromboses are also observed. Neurological manifestations include irritability, apathy, withdrawn behavior, proximal muscle weakness and hypotonia, and diminished reflexes.

Methylmercury ingestion, which has occurred twice in Japan and in Iraq on epidemic scales, produces acute and sustained neurotoxicity. In exposed infants, cerebral palsy-like symptoms developed, including psychomotor retardation, microcephaly, spastic or flaccid paralysis, ataxia, athetotic movements of the hands, constriction of the visual fields, and generalized tonic convulsions. *In utero* exposure can result in distinct toxicity to the embryo, including low birth weight, dysplasias of the cerebral and cerebellar cortices, and an abnormal migratory pattern of neurons. Gross findings, such as severe mental retardation, have been reported in these epidemics, as have more subtle findings, such as failure to achieve developmental milestones, observed on follow-up studies.

Occupational exposure to mercury vapour has led to acute cases of respiratory distress, renal failure requiring dialysis, and severe oropharyngeal inflammation and a 'flu-like' syndrome. With more prolonged exposure, proteinuria develops. The hypothesized mechanism for proteinuria is stimulation by mercury of T lymphocytes to produce an antibody to the glomerular basement membrane. This may be mediated in part by mercury binding to nucleoprotein of T lymphocytes. However, it has recently been shown that methylmercury induces T-cell apoptosis (programmed cell death), possibly by damaging mitochondria and inducing oxidative stress that activates pathways for apoptosis.

Neurological symptoms in workers exposed to mercury 20–35 years previously have included reduced muscle strength and coordination, increased tremor, decreased sensation, and increased prevalence of abnormal reflexes. It is hypothesized that progressive neuron loss with aging may unmask previous exposure to mercury. Furthermore, mercury has been demonstrated to persist in neurons for many years after exposure has ceased. A survey of workers previously exposed to mercury failed to reveal any impairment of reproductive function.

Also of interest, although not directly observed in humans, *in vitro* experiments with human platelets have shown that mercury can produce platelet aggregation and increase production of eicosanoids by platelets. Increased platelet aggregation may contribute to the thrombus formation seen in acrodynia as well as some of the neurological structural abnormalities produced by hemorrhages in the brain secondary to mercury exposure of animals. Despite the considerable exposure to mercury from dental amalgams, no toxicity has yet been reported in humans resulting from this exposure.

In animals, toxicity is of a similar nature to that observed in humans. Rats given mercuric chloride intravenously develop complete proximal tubular necrosis. If given a high-protein diet 48 h prior to mercury ingestion, they develop only partial proximal tubular necrosis, and blood urea nitrogen and creatinine concentration are normal. Mechanisms for renal injury are, at least in part, felt to be autoimmune. Mercuric chloride administration to the brown Norway rat not only resulted in production of antiglomerular basement membrane antibodies but also induced a rise in CD4+ splenocytes, CD4+ and CD8+ (helper and suppressor) T cells, and B lymphocytes in peripheral lymphoid organs. In glomeruli there was an increase in CD8+ (suppressor) T lymphocytes. These findings suggested that the mechanism for autoimmunity is T lymphocyte-dependent. Glomerular basement membrane autoantibodies are both immunoglobulin G (IgG) and IgM, detectable 7 days after mercury injection. The degree of proteinuria seen in this group of rats did not correlate with autoantibody levels. Although glomerular histology was unremarkable in these rats, by day 18 after injection there were vascular changes, including endothelial cell swelling in arteries and arterioles. Antiglomerular basement membrane IgG appeared by day 9 after injection, peaked by day 15, and then declined. Proximal tubular basement membrane IgG appeared on day 9 and peritubular capillary basement membrane IgG appeared on day 15. These antibodies persisted for at least 30 days. Urinary metallothionein increased after mercury administration, probably reflecting mercury induction of tissue metallothionein as a means of self-protection.

Significant neuropathology has also been produced by mercury *in vivo*, both in the fetus and in more mature animals. In pregnant rats given an intraperitoneal

injection of methylmercuric chloride effects were seen in the fetal nervous system within 2 h. These early changes included mitochondrial degeneration of the endothelium of cerebral capillaries leading to subsequent hemorrhage, which interferes with normal neuronal migration and normal development of cortical cytoarchitecture. It is of interest that alterations in vascularity leading to thrombus formation are seen in the kidney, as mentioned above. Methylmercury also disrupts neuronal microtubular assembly. Mercury is thought to reduce glutathione peroxidase activity in cells, to interact with the sulfhydryl groups of natural tubular and membrane proteins, and to cause peroxidative injury. Mercury has been found in organelles, primarily in lysosomes, bound to selenium. When injected intraperitoneally into the rat, methylmercuric chloride distribution was uniform in cerebellar Purkinje cells and Golgi cells, as well as in cerebral cortical layers III, IV, and VI. In rats exposed to inorganic mercury, only scattered mercury was isolated in level VI. Once in the brain, methylmercury is slowly converted into inorganic mercury. Both methylmercury and inorganic mercury accumulated in the anterior horn motor neurons of the spinal cord.

Functionally, subcutaneous administration of methyl mercuric chloride to neonatal rats resulted in movement and postural disorders developing during the fourth week of life. These occurred in association with a selective degeneration of cortical interneurons, which use γ-aminobutyric acid (GABA) as a neurotransmitter. In the putamen and caudate nucleus, GABAergic and somatostatin immunoreactive interneurons showed degenerative changes. Sodium-dependent high-affinity uptake of tritiated choline in cerebral cortex was also reduced, suggesting that both cholinergic and GABAergic neuronal function may be disturbed. Finally, *in vitro* incubation of rat cerebellar granular cells with methylmercury caused a delayed but increased phosphorylation of selected cellular proteins. The time course from exposure to phosphorylation was 12–24 h, a time period consistent with mercury alteration of gene expression.

Long-term neurotoxic effects from methylmercury exposure have also been demonstrated in cynomolgus monkeys. In one group of monkeys given methylmercury orally from birth to 7 years of age, whole blood concentrations of mercury fell almost immediately after the cessation of oral administration. However, monkeys assessed neurologically 6 years later exhibited insensitivity to touch and pinprick, clumsiness, and slowness. Other monkeys were exposed to methylmercury *in utero* by administration of methylmercury to their pregnant mothers and continuing postpartum administration to the infants for 4–5 years. There were more profound visual defects in the monkeys with prenatal and postnatal exposure compared to those exposed after birth only.

Toxic Levels in Humans

The most accessible sampling sites in humans are urine, hair, and whole blood. Urinary concentration of mercury is the monitoring test which is most often reported and is considered to be the analysis of choice for chronic mercury exposure. Normal values for urinary mercury concentration are less than $100\,\text{nmol}\,\text{l}^{-1}$ ($20\,\mu\text{g}\,\text{l}^{-1}$). These reference values were obtained as a result of a World Health Organization (WHO) study in 1961. Urinary mercury concentration was determined in urine samples from 1107 adults in 15 countries, and 85% of the subjects had urine mercury concentrations within the abovementioned range. No reference values exist for children, although acrodynia has been reported in children with urinary mercury concentrations as low as $249\,\text{nmol}\,\text{l}^{-1}$ ($50\,\mu\text{g}\,\text{l}^{-1}$). Urinary mercury concentrations do not correlate with ambient mercury levels due to the long half-life of mercury in the body (approximately 40–70 days). In addition, mercury is eliminated by other routes, such as sweat and bile. Young children generally have higher urinary mercury concentrations than adults because they crawl on floors, where the highest concentrations of mercury vapor tend to accumulate; toddlers also tend to become dirty and may mouth objects that have been lying in mercury-contaminated soil.

Workers occupationally exposed to mercury 20–35 years prior to being tested had neurological symptoms, including weakness, tremors, decreased sensation, and abnormal Babinski reflex (upgoing toes), with peak urinary mercury levels of $0.6\,\text{mg}\,\text{l}^{-1}$ ($3.0\,\mu\text{mol}\,\text{l}^{-1}$). Initial serum mercury concentration in two patients with acute intoxication from occupational mercury exposure were 550 and $490\,\mu\text{g}\,\text{dl}^{-1}$ (27 and $24\,\mu\text{mol}\,\text{l}^{-1}$). Serum or whole blood mercury concentration is a sensitive indicator of acute mercury loading, especially with inorganic mercury, owing to the near-equal distribution between red blood cells and plasma, with normal serum mercury concentration being $< 50\,\text{nmol}\,\text{l}^{-1}$ ($1\,\mu\text{g}\,\text{dl}^{-1}$). Hair mercury content is reported to be a reliable indicator of methylmercury load. When methylmercury has undergone distribution to various body parts, the blood and hair are in equilibrium, with hair concentration being approximately 250 times that of whole blood. Normal hair content of mercury is less than $7.5\,\mu\text{g}\,\text{g}^{-1}$ hair.

Ranges of Dietary Intakes

Mercury levels in the diet, especially of methylmercury, are of great importance, but are often poorly documented. Recently, data have been reported on the mercury content of poultry and dairy products from hens and broilers raised on fish meal from a number of countries in Europe and North America. Table 1 lists these values and, notably, most products fall within the acceptable tolerance limits of 50 p.p.b. wet weight set by the Food and Agriculture Organization (FAO) and the WHO of the United Nations. However, these FAO/WHO limits were established for foods other than fish and other seafood. Acidification of lakes by acid rain has increased the levels of mercury in the edible tissues of fish and has resulted in the prohibition of sport fishing in many American and Canadian lakes and rivers. In addition, the sale of shark and swordfish has been restricted, owing to their high mercury content.

Inhalation of mercury vapor is also a significant source of mercury intake as inorganic mercury is volatile at room temperature, and absorption through the respiratory tract occurs readily. Mercury can be detected in human volunteers within 30 min following exposure. Current recommendations by the American Conference of Governmental Industrial Hygienists are for a threshold limit value of mercury of 25 $\mu g\ m^{-3}$ (125 $nmol\ m^{-3}$) for a 40-h work week, while the US National Institute of Occupational Safety and Health (NIOSH) for workplace mercury vapor exposure for 10 $h\ day^{-1}$ is 250 $nmol\ m^{-3}$ (50 $\mu g\ m^{-3}$). For mercury concentration in household air, the US Agency for Toxic Substances and Disease Registry mandates an acceptable level of only 2.5 $nmol\ m^{-3}$ (0.5 $\mu g\ m^{-3}$). In the ambient outdoor air in a major US city such as Detroit, mercury concentration is reported to be as low as 0.015 $nmol\ m^{-3}$ (0.003 $\mu g\ m^{-3}$). In contrast, indoor mercury concentrations in a house just painted with mercury-containing latex paint can be as high as 847 $nmol\ m^{-3}$ (170 $\mu g\ m^{-3}$), falling to 8 $nmol\ m^{-3}$ by 7 days after the painting is completed. However, this latter amount is still three times greater than the recommended indoor ambient concentration.

Mercury tends to accumulate near the floor. Thus, in the room of a child in the UK who developed acrodynia from mercury accumulation in a vacuum cleaner, ambient mercury concentration was 1.0–1.2 $\mu g\ m^{-3}$ (5–6 $nmol\ m^{-3}$), while at the floor level it was 30.0 $\mu g\ m^{-3}$ (150 $nmol\ m^{-3}$). Similar findings have been reported in the USA, with a range of ambient mercury reported from 0.5 to 50.0 $nmol\ m^{-3}$ (0.1–10 $\mu g\ m^{-3}$).

Table 1 Mercury content (methylmercury and total mercury) of poultry products and eggs reported from Europe and North America

Country	*Hen meat (p.p.b.)*	*Broiler meat (p.p.b.)*	*Eggs (p.p.b.)*
Methylmercury			
Greece	0.5–3.1[a]	0.1–3.0[a]	0.4–28.9[d]
	0.4–2.9[b]	0–2.7[b]	0.1–0.6[e]
	1.7–16.5[c]	0.5–7.1[c]	
USA			40
Sweden	17–37[f]		6–22[g]
			1–2[h]
Total mercury			
Germany		< 1–141	
Italy	< 1–15		
Canada			1–23
USA			20–40

[a]Breast; [b]leg; [c]liver; [d]free-range; [e]commercial range; [f]after prohibition of methylmercury in agriculture; [g]egg white; [h]egg yolk.

Incidences of Mercury Toxicity and Scares

The most notable occurrences of epidemic mercury exposure were described in a group of cities bordering Minamata Bay in Japan in 1956 and 1965, and in Iraq in 1971–72. In the Japanese epidemics, an acetaldehyde plant utilizing mercury released the mercury in its industrial waste into Minamata Bay, thus entering the aquatic food chain. In the case of Iraq, wheat contaminated with a mercury-containing fungicide was used to make bread. Follow-up studies in 1986–87 of 35 victims of Minamata disease, diagnosed between 1968 and 1978, revealed not only persistent and progressive neurological dysfunction but newly appeared lesions of spontaneous nystagmus in nine cases (25%). Fourteen affected Japanese patients from the Niigata area demonstrated temporal lobe gyrus degeneration, loss of small myelinated fibers, and a reduction in large neurons in the cochlear nerve. These findings suggested that hearing impairment in these victims may have been caused by progressive or irreversible degeneration of neurons or nerve fibers. It is from these epidemics that we have learned most about methylmercury poisoning.

Treatment

Both dimercaprol (BAL) and D-penicillamine have been used to treat victims of mercury poisoning. Use of BAL has been complicated by an increase in brain concentration of mercury. Dimercaptosuccinic acid (DMSA or succimer – Chemet®), is currently considered to be the chelator of choice to reduce the body burden of mercury.

See also: **Fish Meal**; **Fungicides**; **Heavy Metal Toxicology**; **Immunology of Food**; **Renal Function and Disorders**: Kidney: Structure and Function; **Selenium**: Physiology; **Zinc**: Physiology

Further Reading

Agocs MM, Etzel RA, Parrish RG *et al.* (1990) Mercury exposure from interior latex paint. *New England Journal of Medicine* 323: 1096–1101.

Clarkson TW (1972) Pharmacology of mercury compounds. *Annual Review of Pharmacology* 12: 375–405.

Clarkson TW (1990) Mercury: An element of mystery. *New England Journal of Medicine* 323: 1137–1139.

Garza-Ocanas L, Torres-Alanis O and Pineyro-Lopez A (1997) Urinary mercury in twelve cases of cutaneous mercurous chloride (calomel) exposure: effect of sodium 2,3-dimercaptopropane-1-sulfonate (DMPS) therapy. *Journal of Toxicology and Clinical Toxicology* 35: 653–655.

Goyer RA (1995) Nutrition and metal toxicity. *American Journal of Clinical Nutrition* 61 (suppl): 646S–650S.

Hahn LJ, Kleiber R, Leininger R, Vimy MJ and Lorschneider F (1990) Whole body imaging of the distribution of mercury released from dental fillings into monkey tissues. *Journal of the Federation of American Societies for Experimental Biology* 4: 3256–3260.

Lowell JA, Burgess S, Shenoy S, Curci JA, Peters M and Howard TK (1996) Mercury poisoning associated with high-dose hepatitis B immune globulin administration after liver transplantation for hepatitis B. *Liver Transplantation and Surgery* 2: 475–478.

Myers GJ and Davidson PW (1998) Prenatal methylmercury exposure and children: neurologic, developmental and behavioral research. *Environmental Health Perspectives* 106 (suppl 3): 841–847.

Nakagawa R, Yumita Y and Hiromoto M (1997) Total mercury intake from fish and shellfish by Japanese people. *Chemosphere* 35: 2909–2913.

Oyanagi K, Ohama E and Ikuta F (1989) The auditory system in methylmercurial intoxication: a neuropathological investigation in 14 autopsy cases in Niigata, Japan. *Acta Neuropathologica* 77: 561–568.

Renzoni A, Zino F and Franchi E (1998) Mercury levels along the food chain and risk for exposed populations. *Environmental Research* 77: 68–72.

Yoneda S and Suzuki KT (1997) Detoxification of mercury by selenium by binding of equimolar Hg–Se complex to a specific plasma protein. *Toxicology and Applied Pharmacology* 143: 274–280.

METABOLIC RATE

W P T James, IASO/IOTF Offices, London, UK
A Ferro-Luzzi, Istituto Nazionale della Nutrizione, Rome, Italy

Introduction

The metabolic rate of the body is the overall rate of tissue oxidation of fuels by all the body's organs. The dietary fuels are the carbohydrate, fat, protein, alcohol, and minor dietary components that are oxidized in the tissues, oxygen being taken up by the lungs and the combusted end products (carbon dioxide, water, and urea) being excreted by the lungs, urine, and skin. The total rate of body metabolism is assessed by monitoring the rate of oxygen uptake by the lungs. The sources of fuel can then be estimated from the proportion of carbon dioxide produced and the rate of urea production. The equations for calculating these are set out below. The rates of utilization of body stores of carbohydrate (c, expressed in terms of monosaccharide units) and fat (f), in g h^{-1}, were calculated from the VO_2 and VCO_2 (1 h^{-1}) and the rate of leucine oxidation (L, mmol h^{-1}) using formulae derived by Garlick (1987). C and F are the rates of dietary intake of carbohydrate and fat. In the fasted state, the rates were as follows:

$$c = (4.574 \times V_{CO_2}) - (3.260 \times V_{O_2}) - 0.864L$$

$$f = (1.673 \times V_{O_2}) - (1.682 \times V_{CO_2}) - 0.434L.$$

In the fed state, the rates were as follows:

$$c = (4.574 \times V_{CO_2}) - (3.260 \times V_{O_2}) - 0.850L - 0.077F - C$$

$$f = (1.673 \times V_{O_2}) - (1.682 \times V_{CO_2}) - 0.383L - 0.947F.$$

The energy equivalence of oxygen varies, depending on the precise nature of the fuels being oxidized, but a value of 20 kJ per liter of oxygen is taken as an appropriate average. (*See* **Energy**: Measurement of Food Energy.)

Factors Affecting Metabolic Rate

The process of oxidation involves a series of enzymatically controlled biochemical reactions leading

eventually to the combination of oxygen with the carbon and hydrogen components of the body's fuels to yield the carbon dioxide and metabolically derived water. The incompletely oxidized nitrogen is excreted as urea, which is synthesized by the liver and excreted by the kidneys. The intermediate steps in the metabolism of the body's fuels are linked biochemically to drive the generation of phosphate-containing organic molecules, such as adenosine triphosphate (ATP), which in turn serve as the direct energy sources for all the body's cell activities, including the synthesis of complex molecules, the maintenance of tightly controlled ionic gradients in the cell, and the excretion of ions and molecules outside the cell. Thus, the oxygen being taken up by the lungs reflects the tissue metabolism of the fuels needed to regenerate the ATP used up in either biochemical 'internal' work or mechanical external work undertaken by the body's muscles. The rate at which the body burns its own stored fuels in the fasted, resting and relaxed state, i.e., in the basal state in a warm room, is called the basal metabolic rate (BMR). This varies with the age of the individual, mainly because of the varying sizes of metabolically very different organs at different ages. Thus, a child has a relatively large brain, liver, and intestine with a higher metabolic rate per kilogram of body weight than a more muscular adult. Body fat cells are metabolically active but contain a substantial amount of inert fat, so that the larger fat mass of a woman has a lower BMR per unit weight than a man, but if the oxygen uptake is calculated in terms of the metabolically active fat-free mass, then her metabolic rate is the same. As men and women age, they tend to lose lean tissue and store extra fat, so that the BMR on a weight basis falls with age. Again, the BMR remains the same when expressed in terms of the fat-free mass of the body.

Figure 1 depicts the changes in BMR of girls and women as they grow and age. Equations can now be used to estimate a group's BMR from their sex, age and bodyweight (see Table 1), but there is a range of BMR amounting to $\pm$ 20% of the mean value at each weight. Thus, in a young 25-year-old woman of 55 kg, the anticipated mean BMR is 1305 kJ per day but could vary under normal conditions from 1063 to 1547 kJ per day. The kilojoule is the standard measure, and 4.184 kJ corresponds to 1 kcal, which was originally defined in energy terms as that required to increase the temperature of 1 g of pure water by 1 °C from 14.5 to 15.5 °C. The variation in BMR at a constant weight in part reflects differences in the fat content of individuals of the same weight. Thus, the BMR per unit fat-free mass varies by 12–15% rather than by $\pm$ 20% as for weight. About 40% of the BMR may be explained by differences in the size of the body's organs, e.g., liver, intestine and muscle, but there is a residual difference between individuals, which seems to be explicable only in terms of differences in the rate at which every organ of the body metabolizes its fuel. This in turn is controlled principally by the circulating concentration of thyroid hormones. Adults with a normal but above-average level of thyroid hormones in their blood tend to have a BMR in the upper part of the normal range. Smokers also have a BMR that is about 5% above the normal, but whether or not this relates to changes in thyroid hormones is unknown. Young women who have normal menstrual cycles show a swing in BMR that is at its lowest in the late follicular phase, just before ovulation. On ovulation, the basal body temperature rises rapidly by about 0.5 °C. The BMR is also increased but rises further to a peak in the later luteal phase, immediately before menstruation. This metabolic cycle with swings of $\pm$5% of the mean is independent of changes in food intake, but the recognized

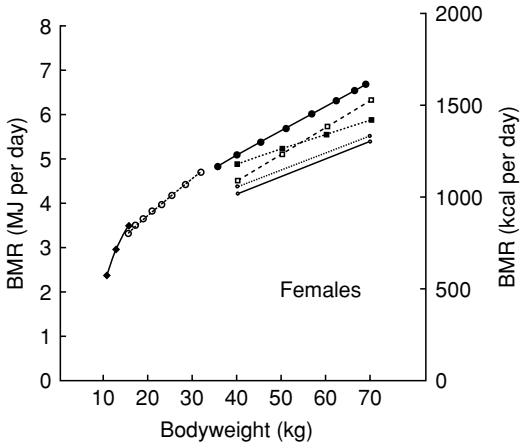

Figure 1 Relationship of basal metabolic rate (BMR) with age and weight in females: —◆— 1–3 years; --○-- 3–10 years; —●— 10–18 years; --□-- 18–30 years; --■-- 30–60 years; ○--○ 60–74 years; ○——○ over 74 years (data based on equations in Table 1). Reproduced from Metabolic Rate, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 1 Equations for prediction of basal metabolic rate (used in Figure 1)

Age range (years)	*Males*	*Females*
0–3	0.255W − 0.226	0.225W − 0.214
3–10	0.0949W + 2.07	0.0941W + 2.09
10–17	0.074W + 2.754	0.056W + 2.898
18–29	0.063W + 2.896	0.062W + 2.036
30–59	0.048W + 3.653	0.034W + 3.538
60–74	0.0499W + 2.930	0.0386W + 2.875
75+	0.0350W + 3.434	0.0410W + 2.610

5–10% fall in intake during the follicular phase with a similar rise in the luteal phase may accentuate the hormonally dependent swing. The effects of contraceptives that inhibit ovulation and the subsequent rise in basal temperature are unknown. The previous day's food intake does not affect the BMR unless there has been substantial overeating. However, the mixture of fuels combusted during fasting is influenced by the proportion of the previous 3–4 days' intakes, which is derived from carbohydrate, much of the glucose from glycogen being metabolized in the fasting state if carbohydrate intake was previously high. When the carbohydrate store of glycogen in the liver is nearing exhaustion, the body's output of carbon dioxide falls as the body switches to using body stores of fat. The oxygen uptake for combusting the fatty acids continues since the demand for regenerating ATP is unaffected by the change in fuel supply, but a carbohydrate-rich diet tends to induce a slightly higher fasting metabolic rate than an energy-equivalent, fat-enriched diet, probably because of a slight induction of thyroid metabolism by dietary carbohydrates. (*See* **Hormones**: Thyroid Hormones.)

The BMR falls by between 2 and 5% when individuals transfer to live in a tropical warm environment, and in uninsulated houses, seasonal changes in the BMR were readily seen with a 5–10% rise from summer to winter being observed in the Japanese before World War II. The BMR formulae shown in Table 1 ignore any temperature effects. The BMR of some people living in the tropics may be as much as 10% below the values shown, but these studies have been conducted on children and adults who are, or were, undernourished. Poor nutrition may have both an immediate and a long-term effect in lowering the basal metabolic rate. Semistarvation leads to a fall in BMR from about day 4, and within 2 weeks, the BMR can fall by 15% as thyroid metabolism changes and the body's organs become more efficient. More prolonged or severe semistarvation leads to a progressive loss of the body lean tissues as well as fat, and the BMR therefore continues to decline in proportion to the loss of lean tissues. Body weight can eventually stabilize at a new low level and, if physical activity is also reduced, semistarved volunteers can come back into energy balance on 50% of their initial intake. However, this requires a 40% loss of weight and marked lethargy if energy balance is to be preserved on such a low intake.

The Components of Metabolic Rate

Traditionally, the metabolic rate is divided into three components: BMR, postprandial thermogenesis, and physical activity. The BMR usually amounts to 50–60% of an individual's total energy expenditure, and postprandial thermogenesis to 10% used for the metabolic cost of processing, i.e., eating, absorbing, transporting, and storing food. The remaining energy is used for physical activity.

Postprandial Thermogenesis

The surge in oxygen uptake after a meal, known as postprandial thermogenesis, has been variously described as the specific dynamic action of food, dietary-induced thermogenesis or the thermic effect of feeding. The last term is particularly favored by animal nutritionists. It is difficult to measure it accurately because after ingesting the food with minimum physical effort, an individual has to lie at complete rest while the oxygen uptake and carbon dioxide production are monitored for many hours until the metabolic rate has returned to the basal rate. This may take more than 10 h, which explains why BMR is measured after a 14-h fast. Separate feeding of different fuels shows that the maximum effect on oxygen uptake occurs after protein. This response is equivalent to about 30% of the protein's energy; glucose induces a 5–10% effect, fat only a 2–5% effect, consistent with its slow absorption by the lymphatic tissue, and alcohol a 0–8% effect. Certain dietary components also increase metabolic rate, e.g., a caffeine equivalent to two cups of tea increases metabolic rate by 1–3%, and spices, such as those found in an Indian curry, increase it by 25% compared with a nonspiced meal. Moderate exercise amplifies the metabolic response to a standard meal, so that the combined effect of exercise and food is greater than the sum of the response to each stimulus given separately. The effect, however, is small and amounts to 2% of the total energy expenditure. (*See* **Thermogenesis**.)

Differences in postprandial energy expenditure have been sought as an explanation for the propensity of some individuals and animals to obesity. Results are often conflicting because in any person, the response tends to vary from day to day and is readily influenced by changes in gastric emptying. A proportion of obese subjects have a reduced metabolic response to a meal; this effect may prove to depend on the degree of abdominal insulation since the response is reduced if volunteers are swathed in insulation to reduce the abdominal heat loss, thereby increasing the temperature of the blood entering and leaving the liver. This seems to reduce the stimulus to body metabolism. Lactating mothers (and pregnant women) have a lower postprandial thermogenesis that returns to normal after they have stopped breast-feeding. Smoking and postprandial thermogenesis interact synergistically so the thermic output

after a meal is enhanced. The small response during lactation is consistent with that observed in many species of animal in which brown adipose tissue is used as the organ for modulating heat production as a mechanism to maintain body temperature. However, this organ is not very active in humans.

Prolonged underfeeding and overfeeding lead to changes not only in postprandial thermogenesis but also in BMR. The effects of semistarvation (see above) are modest, but overfeeding can produce a much greater response, provided that the intensity of overfeeding (especially with carbohydrate) is high. Thus, progressive overconsumption of 6.3 MJ (1500 cal) per day leads to a 33% increase in daily energy expenditure. This composite response to more prolonged overfeeding is usually classified as dietary-induced thermogenesis. Nevertheless, this apparent mechanism for dissipating excess energy is limited because, at most, 27% of the excess intake is metabolized, the remaining 73% being stored, two-thirds of it as fat and about a third as lean tissue. The majority of the excess response in metabolism is accounted for by the theoretical cost of fat synthesis from carbohydrate, although the human capacity to transform carbohydrate into fat is limited, preference being given to the selective storage of the fat component of the ingested energy.

Physical Activity

The energy cost of physical activity is predictable, but in order to obtain a reasonable estimate of energy expenditure on a daily basis, an analysis of activity patterns on a minute-by-minute basis is required. Children can be very active, making a detailed analysis difficult because the type of activity needs to be specified if an energy cost is to be assigned to each type of activity. Weight-bearing movement and anti-gravitational moves, e.g., walking up a hill with a load, are particularly expensive. The simplest way of estimating individual costs is to use the extensive tables now collected on the energy cost of different movements in children and adults. For simplicity, these can be expressed as a ratio of the BMR, since, in this way, differences between the sexes and individuals of varying size are removed. Table 2 illustrates how this is achieved for a female domestic worker. The physical activity ratios (PARs), i.e., energy costs in relation to BMR, can be assigned on the basis of extensive measures of similar activities. If the average energy requirement of the individual is to be estimated, account must be taken of the different types of work involved throughout the year. Thus, the individual shown, working in a Mediterranean environment, worked both as a domestic help and as an agricultural seasonal laborer. Each type of day is compartmentalized on a minute-by-minute basis, a division being made for convenience between occupational and other work. Maintenance of the individual's household varies depending on the day of the week and the season. Socially desirable activities encompass a variety of activities, including, for example, visiting friends, walking, churchgoing or dancing. The assumption is also made that some intensive exercise, amounting, on average, to 15 min daily, is undertaken; this is desirable if muscle tone and physical fitness are to be preserved. After integrating all

Table 2 Energy requirements of a female domestic helper (age, 25 years; weight, 60 kg; BMR, 5.8 MJ day^{-1})

	Integrated Energy Index[b]	*Day type 1*[a] *Housewife*		*Day type 2*[a] *Domestic labor*		*Day type 3*[a] *Agricultural labor*	
		h	*MJ*	*h*	*MJ*	*h*	*MJ*
In bed	1.0	8	1.93	8	1.93	8	1.93
Occupational activities							
Household work	2.7	2	1.31	1	0.65	1	0.65
Domestic labor	2.8			8	5.41		
Tomato harvesting	3.0					8	5.80
Discretionary activities							
Household maintenance	2.0	2	1.00				
Socially desirable	1.7	4	1.64	2	0.82	2	0.82
Cardiovascular and muscular maintenance	6.0	0.25	0.36				
Rest of day	1.4	7.45	2.62	5	1.69	5	1.69
Daily total PAL[c] *(MJ)*							
Integrated PAL (MJ) is 1.69 (9.78)		1.53	(8.86)	1.81	(10.50)	1.88	(10.89)

[a]Day type 1, 4 out of 7 days for 10 months (48% of the year); day type 2, 3 out of 7 days for 10 months plus 1 out of 7 days for 2 months (38% of the year); day type 3, 6 out of 7 days for 2 months (14% of the year).
[b]Integrated Energy Index (IEI) is the energy cost of the activity including pauses for rest, expressed on a minute or hourly basis and calculated as a ratio of BMR.
[c]PAL, physical activity level.

these activities, the ratio of the total energy expenditure to the BMR is found to average 1.69; the ratio is designated the physical activity level (PAL). Table 3 provides a listing of PAL values for adults of all ages according to their activity patterns. Thus, by knowing the sex, age, and bodyweight of individuals, it is possible to estimate their BMR (see Table 1). Given this BMR figure in MJ per day, multiply by the PAL value shown in Table 3, and the energy needs can be estimated. Note that individuals vary in their energy needs by $\pm 15\%$, so that an individual's needs cannot be predicted very accurately unless his or her BMR is measured. Nevertheless, the total energy expenditure of an individual child or adult is remarkably consistent from day to day, varying by only 1–2% provided that food intake and physical activity are meticulously standardized, and account is taken in women of the stage of the menstrual cycle. Thus, the energy needs of individuals are surprisingly stable, and once this BMR is measured, the total energy needs of the child or adult can be readily estimated from a knowledge of their activity patterns. The factors modulating BMR or the metabolic response to food are many, but their effect is small, so that it is not surprising that energy expenditure is very predictable; the human body is therefore a finely tuned and well-regulated machine. Abnormalities of regulation, e.g., in obesity, only arise because of a consistent discrepancy between the physiological controlled intake and expenditure, which, if discrepant by 1–2%, produces a 2–5 kg weight change in a year. (*See* **Exercise**: Muscle; Metabolic Requirements.)

New measures of total energy expenditure estimated over 2–3 weeks can be obtained by the use of the double-labelled water technique, which relies on the difference in labelling of urine or saliva with the two heavy isotopes of water, deuterium and ^{18}O. The differential dilution of deuterium and ^{18}O in urinary water is monitored over a 2–3-week period following a single oral dose of $D_2{}^{18}O$. The ^{18}O content is diluted more rapidly than the deuterium because the oxygen in water exchanges rapidly with the body's bicarbonate pool, which is turning over rapidly as carbon dioxide is produced by tissue metabolism. Thus, the difference in dilution rates of ^{18}O and deuterium provides a measure of the rate of carbon dioxide production. The technique is expensive and difficult to perform analytically but very convenient for the subject being studied, since only single daily or occasional urine or saliva specimens are needed over the period of observation. This method is likely to be increasingly used, but classical methods have allowed current estimates of energy needs to be calculated. These are accurate for groups of people but not for individuals who vary not only in their physical activity patterns but also in their metabolic rate at rest. Figure 2 gives an indication of the decline in energy needs during adult life. This results from the atrophy of the lean tissues, which may be related to the fall in physical activity. Lack of exercise is therefore a handicap because it directly reduces energy expenditure, and it may also lead to a slow shrinkage of tissues, such as muscle, thereby producing a long-term fall in

Table 3 Development of the International Labour Organization's classification to derive simple estimates of PAL values[a] for each occupational group separated into averages for developed countries (DC) and less developed countries (LDC)

	Occupational PAL values			
	Males		*Females*	
Occupational group	*DC*	*LDC*	*DC*	*LDC*
Professional, technical, and related workers	1.55	1.61	1.56	1.58
Administrative and managerial	1.55	1.61	1.56	1.58
Clerical and related workers	1.55	1.61	1.56	1.58
Sales workers	1.67	1.78	1.60	1.64
Service workers	1.67	1.78	1.60	1.64
Agricultural, animal husbandry, forestry, fishing, and hunting	1.78	1.86		1.69
Production and related transport equipment operators and laborers	1.78	1.86	1.64[b]	1.69
Housewives			1.56	1.64
Students	1.55	1.61	1.56	1.58
Unemployed	1.55	1.61	1.56	1.58
Subsistence farmers	1.78	1.86	1.64	1.69
Domestic helpers		1.78	1.60	1.64
Elderly (>65 years)	1.51	1.51	1.56	1.56

[a]PAL, physical activity level. These values are based on different proportions of the population being assumed to be engaged in light, medium, and heavy activities. The assumption is made that in less developed countries, there will be a greater demand for physical activity because the mechanization of transport, work, home, and leisure time activities will be appreciably less than in developed countries.
[b]Excludes female laborers.

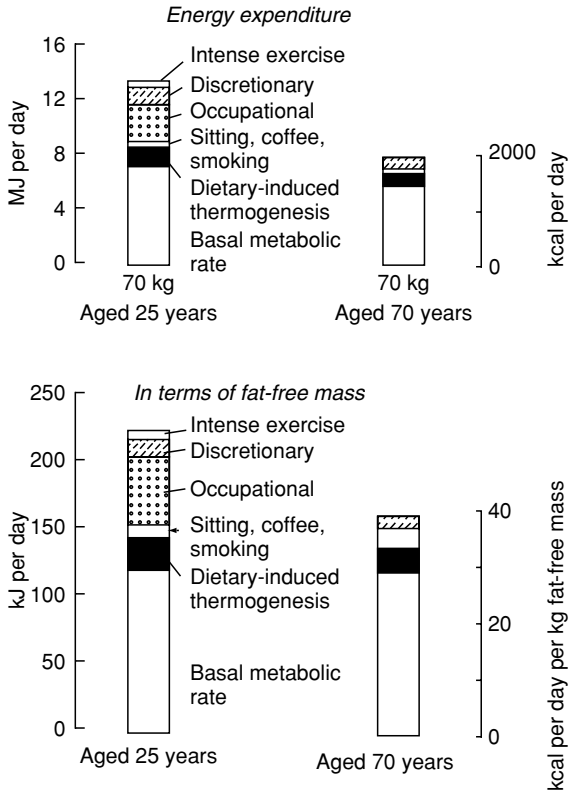

Figure 2 Different components of energy expenditure in a young man and the impact of aging. The energy expenditure falls mainly because of a decline in physical activity. There is also a fall in the basal metabolic rate, but this is in line with the reduction in the lean tissues, i.e., fat-free mass, in the elderly. Reproduced from Metabolic Rate, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

metabolism at rest. Unless people adapt their intake extraordinarily well to this progressive decline in energy output, energy storage, weight gain, and obesity are inevitable.

Extra Energy Costs

The cost of growth amounts to 20 kJ per gram of new tissue deposited; the value can be higher if fat with little lean tissue is laid down. A newborn has a high energy requirement of about $480\,kJ\,kg^{-1}$, but by 1 year of age, this has fallen to about $400\,kJ\,day^{-1}$ as growth slows. Without sufficient energy, a child will fail to grow, but the causes of growth failure usually relate to a deficiency of other nutrients or to infection, rather than to a lack of dietary energy. Adolescents, particularly boys, who are physically very active, may have a high demand for energy. However, the actual cost of even rapid growth rates at this age is modest. (*See* **Adolescents; Children:** Nutritional Requirements.)

Traditionally, pregnancy is considered a time of great demand for food. Good nutrition is extremely important, but the need for additional energy as such is small. The total cost of the 9 months of pregnancy has been estimated as 293 MJ (70 000 kcal) once the cost of tissue deposition (167 MJ) and increase in BMR (126 MJ) have been allowed for. In practice, however, physical activity declines, particularly in late pregnancy, and some enhanced metabolic efficiency, with an early fall in BMR in early pregnancy, seems to occur. The overall extra demand is very small, amounting to $420\,kJ\,day^{-1}$ in the last third of pregnancy only. (*See* **Pregnancy:** Safe Diet.)

Lactation imposes a greater demand on mothers, since their milk contains $1.9\,MJ\,day^{-1}$ after birth, rising to about $2.3\,MJ\,day^{-1}$ on exclusive breast-feeding at 3 months. Extra energy is involved in making this milk, and the total extra energy demand is 2.4–$2.9\,MJ\,day^{-1}$. Part of the energy for this comes from the extra fat stored by the mother during pregnancy, so that the effective extra need amounts to 1.9 MJ at birth and 2.4 MJ in the third to sixth month of exclusive breast-feeding. This explains why mothers are more hungry when nursing their child than when pregnant. (*See* **Lactation:** Physiology.)

Convalescent patients who need to put on weight need extra food, but the cost of this weight gain amounts to between 20 and 40 MJ per kilogram weight gain. If 1 kg is gained per month, the extra food needed therefore amounts to about $1\,MJ\,day^{-1}$.

See also: **Adolescents**; **Children**: Nutritional Requirements; **Energy**: Measurement of Food Energy; **Exercise**: Muscle; Metabolic Requirements; **Hormones**: Thyroid Hormones; **Lactation**: Physiology; **Obesity**: Etiology and Diagnosis; **Pregnancy**: Safe Diet; **Thermogenesis**

Further Reading

Erlichman J, Kerbey AL and James WPT (2002) Physical activity and its impact on health outcomes. Paper 2: prevention of unhealthy weight gain and obesity by physical activity: an analysis of the evidence. *Obesity Reviews* 3: (in press).

FAO/WHO/UNU (1985) Energy and protein requirements. Report of a Joint Expert Consultation. *Technical Report Series, No. 724*. Geneva: WHO.

James WPT and Schofield EC (1990) *Human Energy Requirements*. Oxford: Oxford Medical Publications.

Schofield WN, Schofield C and James WPT (1985) Basal metabolic rate: Review and prediction. *Human Nutrition: Clinical Nutrition* 39C: 1–96.

METALS USED IN THE FOOD INDUSTRY

Ph Decléty, Chemin de la Treille, Vernouillet, France

Background

When talking about materials in the food industry, it is practically impossible to avoid use of the term 'stainless steel,' as so much of this material is used in the fabrication of storage, processing, handling, and other plant and equipment essential to the industry. After a brief recall of the definition and characteristics of the stainless steels, reference will be made to their properties relevant to food industry engineering, and the main applications where these properties are utilized will then be surveyed.

Definition and Classification of Stainless Steel

Stainless steels make up a family of corrosion- and heat-resistant iron-based alloys containing a minimum of about 12% chromium. The corrosion resistance is improved by increasing the chromium content. Both corrosion resistance and fabricating characteristics are further improved by addition of nickel. Other important alloy additions include molybdenum, titanium, and nitrogen.

Stainless steels may be divided into three groups according to composition and metallurgical characteristics:

1. *Austenitic stainless steels* contain chromium and nickel, frequently with additions of other elements. These steels are normally nonmagnetic and can be hardened only by cold working. The commonly used grades containing 18% chromium and 8–10% nickel account for about two-thirds of stainless steel produced. They are in general use in food industry engineering. Steels containing up to about 2.5% molybdenum are used when the corrosive environment is particularly severe. For specialized applications involving the most corrosive environments, higher-alloy grades containing up to 20% chromium, 25% nickel, and 6% molybdenum are available.
2. *Ferritic stainless steels* contain chromium as the principal alloying element. They are magnetic and can be hardened only by cold working. The most commonly used ferritic grade contains 17% chromium with a carbon content of about 0.05%. The ferritic steels are for general use where the corrosive environment is not too severe.
3. *Martensitic stainless steels* are iron–chromium alloys containing a significant level of carbon, which allows the steel to be hardened by heat treatment. Other alloying elements are sometimes added. The most commonly used steels of this family contain 12–13% chromium. Mechanical parts subject to heavy wear and knife blades are their principal applications. The martensitic steels are magnetic. They are not so resistant to corrosion as the ferritic and austenitic grades.

To these three main groups must be added a fourth class, the duplex stainless steels, which have a mixed austenitic/ferritic structure. These steels have a higher chromium content and a lower nickel content than the austenitic stainless steels and may have additions of molybdenum, nitrogen, and copper. They are characterized by good corrosion resistance and high strength. Use of these relatively new materials may be expected to increase.

Table 1 gives the compositions of the main stainless steels available for food industry applications.

Stainless Steel Properties

Physical and Mechanical Properties

Tables 2 and 3 list the physical and mechanical properties of the stainless steels most commonly used in the food industry. The following points should be noted:

1. the relatively low thermal conductivity and high coefficient of thermal expansion of austenitic steels;
2. the high strength of the stainless steels relative to carbon steel;
3. the high tensile ductility of the austenitic steels, which, coupled with good weldability, contributes to the excellent fabricability of these steels.

It should be noted that the austenitic stainless steels offer mechanical strength and ductility also at high temperatures.

Corrosion Resistance

The corrosion resistance of the stainless steels results from the formation of an invisible protective surface oxide film, which is able to repair itself by contact with an oxygen-containing environment if accidentally destroyed. A minimum of 12% chromium is

Table 1 Main stainless steel grades used in food industry engineering[a]

Grade designation			*Composition (%)*					
AISI	*BS*	*Euronorm*	*C (Max.)*	*Cr*	*Ni*	*Mo*	*Other*	*Typical uses*
Austenitic steels								
304	304S16	X6CrNi8–10	0.06	17–19	8–11			General use
304L	304S11	X3CrNi18–10	0.03	17–19	9–12			General use Extra low-carbon grade
321	321S31	X6CrNiTi18–10	0.08	17–19	9–12		Titanium content is over five times the carbon content	General use Ti-stabilized grade
316	316S31	X6CrNiMo17–12–2	0.06	16–18.5	10.5–13.5	2–2.5		Severe environments
316L	316S11	X3CrNiMo18–12–2	0.03	16–18.5	11–14	2–2.5		Severe environments Extra low-carbon grade
316Ti	320S31	X6CrNiMoTi17–12–2	0.08	16–18.5	10.5–13.5	2–2.5	Titanium content is over five times the carbon content	Severe environments Ti-stabilized grade
Ferritic steels								
409	409S19	X8CrTi12	0.08	10.5–12.5			Titanium content is over six times the carbon content	For painted welded structures
430	430S17	X8Cr17	0.08	16–18				Claddings, not very severe environments
Martensitic steels								
420	420S37	X20Cr13	0.20	12–14				Mechanical parts

[a]Other stainless steels are available for specialized applications and enhanced resistance to the most aggressive environments.
AISI, American Iron and Steel Institute.

Table 2 Physical properties of main stainless steel grades used in the food industry

Stainless steel types	*Density (kg dm^{-3})*	*Coefficients of thermal expansion (20–100 °C) ($\times 10^{-6}$ °C^{-1})*	*Thermal conductivity at 20 °C (W m^{-1} °C^{-1})*
Austenitic steels	7.9	16–16.5	15
Ferritic steels	7.7	10–12	26
Martensitic steels	7.7	10	26

necessary to form this 'passive film,' and, as indicated earlier, its protective properties are improved by an increase in the chromium content and by addition of other alloying elements such as nickel and molybdenum. (*See* **Corrosion Chemistry**.)

Ferritic stainless steels are used for food industry equipment where service is not particularly onerous, but, in general, nickel-containing austenitic stainless steels must be selected. In the presence of most foodstuffs, AISI type 304 grade offers a good resistance to general corrosion and pitting corrosion, and it is only for some particularly aggressive products or processes that recourse to more resistant materials containing molybdenum, notably AISI type 316, is necessary.

In welded constructions, a particular type of corrosion can occur in the zone affected by heat during welding. This type of corrosion, termed 'intergranular corrosion,' is due to precipitation of chromium carbide at the grain boundaries, which results in chromium depletion of the adjacent material. This chromium-depleted zone is subject to corrosive attack if exposed to some aggresive media. To avoid the risk of this form of corrosion, it is necessary to prevent the carbide precipitation, and this is achieved by lowering the carbon content or by adding a stabilizing element (titanium or niobium). In practice, extra low-carbon grades (types 304L and 316L) or stabilized grades (types 321 and 316Ti) should be used for welded structures that are to be exposed to aggressive media.

Stress corrosion cracking can occur if a structure has residual surface tensile stresses (resulting from welding or a deep-drawing operation or, in service, from differential expansion effects or vibration) and is exposed to particular chloride corrosive environments at temperatures above about 60 °C. In particular, care is needed to remove any traces of chlorine-containing cleaning agents (which can result readily in chloride contamination) after a cleaning operation. Enhanced resistance to chloride stress

corrosion cracking is shown by the relatively new high-strength duplex stainless steels (e.g., 22% chromium, 5.5% nickel, 3% molybdenum, nitrogen), and this grade has been used for, *inter alia*, sugar industry applications in which stress cracking had been a problem. The higher alloy austenitic grades (e.g., 20% chromium, 25% nickel, 4% molybdenum, copper, and 20% chromium, 18% nickel, 6% molybdenum, copper, nitrogen) also offer enhanced resistance to chloride cracking, and their use for critical applications is expected to develop.

Crevice corrosion can occur in stainless steels when a chlorine-containing solution is trapped under stagnant conditions within a confined crevice, as may be formed under a bolt head or under deposits on a stainless steel surface. Laps and cold shuts in poor-quality welds, inadequately dressed, can form typical sites for crevice corrosion. Design and manufacture of plant and equipment intended to work with chlorine-containing solutions must avoid such crevices, and if deposits are likely to occur during operation of the plant, arrangements should be made to remove these on a regular basis. To avoid risk of crevice corrosion, particular care should be taken to eliminate chlorine-containing solutions after a cleaning operation. In general, it is good practice to avoid prolonged contact with stagnant chlorine-containing solutions.

Stainless Steel Products

Stainless steel is available in all forms known for carbon steel: hot-rolled and cold-rolled flat products; seamless and welded round tubes; square and rectangular welded tubes; flat, round, square, and hexagonal bars; wires; forgings; castings; hot-rolled or extruded sections.

As far as flat products are concerned, cold-rolled sheets and strip thicknesses range usually from 0.25 to 6 mm. For thicknesses over 6 mm, sheets are delivered in the hot-rolled condition.

Thin wall sections can be obtained from sheets and strips by brake or roll forming.

Plates, Sheets, and Strip Surface Finish

The standard mill finishes available are described in Table 4.

Special finishes are available from mills, steel service centers, or independent finishing specialists. Such finishes range from highly reflective electropolished and mirror-polished surfaces to textured and colored surfaces, and epoxy and tern-coated surfaces.

In selecting a surface finish, it is important to recognize that a bright finish offers, generally, superior corrosion resistance and superior cleanability. It is also important when using oriented polished finishes

Table 3 Mechanical properties of main stainless steel grades used in the food industry (cold-rolled sheets and strip)

Grade designation			*Condition*	*Tensile strength* R_m *(N mm*^{2}*)*	*Proof stress* $R_p = 0.2$ *(min) (N mm*$^{-2}$*)*	*Elongation A (min) (%)*
AISI	*BS*	*Euronorm*				
Austenitic steels						
304	304S16	X6CrNil8–10	Softened	500–680	230	48
304L	304S11	X3CrNil8–10	Softened	480–630	230	44
321	321S31	X6CrNiTi18–10	Softened	540–690	245	35
316	316S31	X6CrNiMo17–12–2	Softened	510–700	240	43
316L	316S11	X3CrNiMo18–12–2	Softened	490–670	240	43
316Ti	320S31	X6CrNiMoTi17–12–2	Softened	510–765	270	37
Ferritic steels						
409	409S19	X8CrTi12	Softened	350–520	200	20
430	430S17	X8Cr17	Softened	430–590	240	20

Table 4 Standard mill finishes of stainless steel flat products

Finish designation		*Description*
AISI	*BS*	
1	1	A rough dull surface produced by hot rolling, followed by annealing and descaling
2D	2D	A dull finish produced by cold rolling, followed by annealing and descaling
2B	2B	A bright finish produced by cold rolling, followed by annealing, descaling and final light cold rolling on polished rolls
BA (bright annealed)	2A	A bright, cold-rolled, highly reflective finish obtained by annealing under a controlled atmosphere
3	3A	A cold-rolled polished surface obtained by finishing with an approximately 100 grit abrasive
4	4A	A cold-rolled polished surface obtained by finishing with an approximately 150 grit abrasive or finer

to arrange for the direction of polishing to be vertical if this is possible. It is also important to be aware that polishing is a costly operation and that, from an economic point of view, it is preferable to select a bright, cold-rolled mill finish whenever possible.

Finish 1 (Table 4) is a rough finish, specific to hot-rolled plates of thicknesses over 6 mm. It is not the best finish from the cleanability point of view and would require extensive grinding and polishing to achieve a satisfactory finish. Where a heavy plate (over 6 mm thick) is necessary, the most economic solution is often to use carbon steel plate clad with thin-gauge stainless steel rather than a solid stainless steel plate.

Properties Particularly Relevant to the Food Industry

Chemical Neutrality Towards Foodstuffs

Stainless steels of the appropriate grade are resistant to attack by a very wide range of foodstuffs and can be regarded as chemically neutral. Specifically, provided the correct grade is selected, there is no significant transfer of metallic elements, and the taste and color of the product are not affected.

Cleanability and Retention of Bacteria

Soil contamination on a surface can harbor bacteria and also hinder removal of bacteria that may be attached directly to the surface. A surface with desirable hygiene characteristics must, therefore, readily release soil contamination on cleaning and also readily release bacteria attached to the surface. Furthermore, the surface must be resistant to any disinfectants used to kill residual bacteria.

The stainless steels meet these requirements to a high degree. Specifically, the smooth surface of the steels facilitates removal of contamination, and, with the hardness and impact resistance of stainless steels, this surface characteristic is little affected during service. The high resistance to corrosion of the stainless steels allows more aggressive cleaning and disinfectant solutions to be used to ensure effective disinfection.

The cleanability of stainless steel as a food contact surface has been studied by several authors. One study concluded that stainless steel was not as cleanable as glass or china, but was more cleanable than aluminum and four other plastics tested. More recently, it has been shown that stainless steel domestic sinks retain 10 times fewer bacteria than other sinks (in enameled steel, polycarbonate plastic and mineral–resin composite) after standardized simulated wear, contamination, and spray wash treatments,

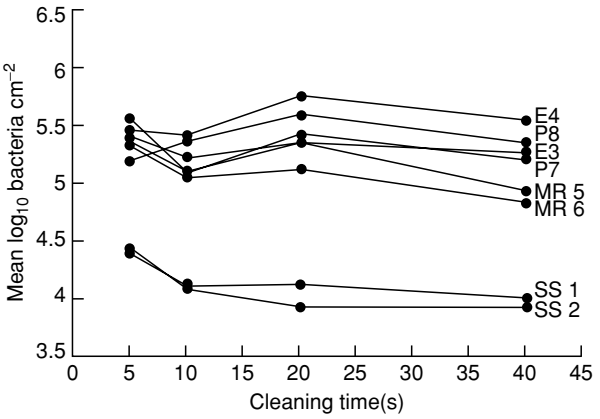

Figure 1 Comparative cleanability of abraded sink material after spray treatment. Mean logarithmic bacteria retained per square centimeter (15 trials) versus cleaning time for a range of abraded sink materials. SS, stainless steel; E, enamelled steel; P, polycarbonate; MR, mineral resin. The numbers refer to the sink number. Holah JT and Thorpe RH (1990) Cleanability in relation to bacterial retention on unused and abroaded domestic sink materials. *Journal of Applied Bacteriology* 69: 599–608 with permission.

and that stainless steel with its high resistance to abrasion and impact damage is more likely to retain its hygienic characteristics throughout a working life (see Figure 1).

The choice of cleaning and disinfectant products for use with stainless steels depends on the nature of the contamination. Some products containing chloride, iodine, or peracetic acid require care in their use. They may cause crevice corrosion if allowed to remain in contact with surfaces under stagnant conditions, for example at seals. Chlorine-containing solutions can lead to stress corrosion cracking if allowed to remain in contact with stressed surfaces at elevated temperatures. There are three precautions to be observed when using such products: the temperature should be kept low and not higher than 40 °C, exposure time should not exceed 30 min, and the plant or equipment should be thoroughly rinsed with clean potable water after the cleaning operation. (*See* **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant.)

Mechanical Properties

The high strength, toughness, and wear resistance, particularly of the austenitic stainless steels, offer important advantages for applications in the food industry.

The high strength coupled with the absence of need for corrosion allowances permits the use of thinner gauges and sections when compared with carbon steel, for example. This brings about savings in weight and costs. The good impact and wear resistance

coupled with the high resistance to corrosion and ease of cleaning results in low maintenance costs and long life. Overall, stainless steel will often prove to be the best material choice on a lifecycle cost basis.

Main Stainless Steel Applications

Dairy Industry

Stainless steel is used for almost all equipment along the chain of milk production, milk processing and milk distribution. (*See* **Milk**: Processing of Liquid Milk.)

At the farm, milking machines as well as piping for transporting the milk up to the refrigerated storage tank are in type 304 stainless steel. The inside of the storage tank is always made from type 304 stainless steel, but type 430 ferritic grades can be used for the outside cladding.

For collecting milk at the farm, stainless steel tankers of suitable size are used. They are fitted with stainless steel accessories: piping and refrigerating circuits, pumps, automatic cleaning equipment, etc.

At the dairy processing plant, all equipment is in stainless steel: storage tanks, pasteurizing plate heat exchangers, piping, pumps, cleaning circuits, etc. Type 304 stainless steel is of general use, but type 316 molybdenum-containing stainless steel is sometimes used for heat exchanger plates to prevent any risk of stress corrosion cracking when aggressive solutions are employed for disinfection.

Equipment for butter processing is also made from type 304 stainless steel, but type 316 molybdenum-containing grades must be selected for equipment used for processing salted cheese because of the better resistance to chloride corrosion. It is also worth noting the use of stainless steel wire for cheese trays.

Generally speaking, surface finish in the dairy industry must be of a high quality in order to facilitate the rigorous cleaning and disinfection required.

Wine Industry

In the wine industry, it is of prime importance to prevent the deterioration of taste and color of products, which can occur through loss of flavoring esters, and through contamination, notably by iron and copper. (*See* **Contamination of Food**.)

Stainless steels meet these requirements with their chemical neutrality when in contact with a wine: they do not modify the pH of the wine, they do not transfer any metallic taste or induce 'ferric breaking' to white wines, and they do not interfere with the biochemical fermentation process.

Many tests have been conducted on stainless steel behavior in wines of various qualities. Their results, confirmed by a long experience, can be summarized as follows:

1. For the storage of red wines in full tanks, for the wine-processing tanks and for wine tankers, austenitic chromium nickel type 304 stainless steel without molybdenum can be used.
2. For the storage of sulfited white wines, when the tanks are not completely full, molybdenum-containing stainless steel type 316 must be used because of the risk of corrosion by sulfur dioxide condensates above the wine surface. In order to reduce cost, it is current practice to limit the use of type 316 to the upper parts which may be exposed above the wine surface; the lower submerged parts are made with type 304 stainless steel. (*See* **Wines**: Production of Table Wines; Production of Sparkling Wines.)

The surface finishes currently used on cold-rolled sheets in the wine industry are 2B finish and bright annealed finish. Weld beads must be ground and polished in order to eliminate oxides, laps, and similar defects.

It should be noted that stainless steel offers the possibility of setting up storage tanks outside protective buildings. The atmospheric corrosion resistance of type 304 stainless steel is generally quite satisfactory, but the molybdenum-containing type 316 steel would be necessary if the tanks are to be exposed to marine environments.

An important advantage of stainless steel over nonmetallic wine tanks is the ease of fitting an external double wall where a refrigerating fluid is circulated for the control of the fermentation temperature. Such control is increasingly used for the production of quality wines.

Vinegar obtained from wine is more aggressive than vinegar obtained from alcohol. It contains elements originating from the wine, such as iron, copper and manganese salts, and, sometimes, chlorides and fluorides, which activate the corrosive action of acetic acid. Type 316 molybdenum-containing stainless steel must be used for the processing and storage equipments of wine vinegar. (*See* **Vinegar.**)

Brandy and cognac are particularly sensitive to contamination from the organoleptic point of view and may be spoiled by a very small quantity of iron. For this reason, the use of type 316 stainless steel is advisable for production and storage equipment (Figure 2). (*See* **Brandy and Cognac**: Armagnac, Brandy, and Cognac and their Manufacture.)

Beer-making Industry

Today, stainless steel has replaced copper, which was the traditional material used for beer-making plant

Figure 2 A battery of wine storage tanks made from chromium nickel molybdenum stainless steel. Addition of molybdenum allows the tanks to be used for white as well as red wine. (Courtesy of UGINE S.A., France.)

and equipment. (*See* **Beers**: Biochemistry of Fermentation; Chemistry of Brewing.)

All equipment is usually in type 304 stainless steel, the corrosion resistance of which is well suited to the product and to the solutions used as cleaning agents. As mentioned previously, some precautions must nevertheless be taken when using chlorine-containing cleaning solutions, and the cleaning instructions given by equipment fabricators must be strictly observed. Type 316 stainless steel is used for brewing equipment when the chloride content of available water is particularly high (Figure 3).

Practically all new beer barrels are now fabricated in stainless steel in substitution of aluminum barrels which, as a consequence, play a declining part in beer storage and transportation. Stainless steel barrels, fabricated from two deep-drawn half-shells of type 304 stainless steel welded together, are price-competitive and bring decisive advantages: high strength, easy cleaning and disinfection and limited maintenance while maintaining a similar weight (Figure 4). (*See* **Barrels**: Wines, Spirits, and Other Beverages; Beer Making.)

Mineral Water, Soft Drinks, and Fruit Juice

Stainless steel is now in general use for equipment used in the collection and treatment of mineral waters as well as for the processing of soft drinks. According to the type of water and its temperature and, for soft drinks, the nature of additions, type 304 or type 316 stainless steel is used.

For fruit juices and fruit juice-based drinks, use of type 316 molybdenum-containing stainless steel is recommended to avoid any risk of contamination by iron or copper, which could spoil the taste or destroy vitamins.

Fruits and Vegetables

There is a great diversity of processes for the preparation and treatment of fruit and vegetables and, as a consequence, a great diversity of processing equipment. Generally speaking, stainless steel type 304 is used for all surfaces in contact with the product at any stage of its preparation. Nevertheless, for some products and, in particular, for treatments at elevated temperature, it is necessary to call for molybdenum-containing stainless steel type 316. In the processing cycle of tomato concentrate, for instance, evaporators are made from type 316 stainless steel.

For the storage of virgin olive oil, EC regulations now specify the use exclusively of type 304 stainless steel for the fabrication of storage tanks. This is to avoid any risk of quality alteration during storage. Owing to the aggressiveness of hot fatty acids, edible oil refining equipment, operated at elevated temperature, needs the use of type 316 stainless steel. (*See* **Vegetable Oils**: Oil Production and Processing.)

In the beet sugar industry, the use of thin-wall welded tubes of ferritic or austenitic stainless steel for the fabrication of evaporators is expanding, in substitution for carbon steel. It appears the most economic solution on a lifecycle cost basis. (*See* **Sugar**: Refining of Sugarbeet and Sugarcane.)

Meat Industry

Stainless steels are used for the fabrication of virtually all equipment for slaughtering, preparation, and

Figure 3 (see color plate 106) Stainless steel fermentation vessels and associated pipework in a modern brewery. (Courtesy of Courage Ltd and British Steel plc.)

handling of meat. They provide a guarantee of cleanability and hygiene. (*See* **Meat**: Slaughter.)

Type 304 stainless steel is generally satisfactory, and its applications range from conveyors, cutting machines, and work surfaces to easy-clean wall covering. For the collection and treatment of blood, however, type 316 molybdenum-containing stainless steel must be selected for this more aggressive use. The same grade must also be used for meat salting plants.

Miscellaneous Food Industry Equipments

Stainless steel is used in many food-processing industries other than those reviewed above. These include confectionery, chocolate and biscuit manufacture, industrial bakery, food deep freezing, fish preparation, and canning. As a general rule, stainless steel is used at least for parts in contact with foodstuffs but increasingly for the whole equipment of the plant. A recent tendency is to move to stainless steel for the

Figure 4 (see color plate 107) Stainless steel's hygienic and corrosion resistance properties are vital in the brewing industry. Today, all new beer barrels are made with type 304 stainless steel. (Courtesy of British Steel plc.)

framework of equipment traditionally made from painted carbon steel; the advantages of this substitution are a better cleanability and a lessening of maintenance works, which compensate quickly for the higher initial cost. Refer to individual processes.

Some equipment is used by several food industries, for example bottling and packing machinery; stainless steels are used for parts in direct contact with the beverages and foodstuffs handled. These include band or chain conveyors, tanks of filling machines, cladding of machinery framework and, sometimes, as mentioned above, also for the supporting framework itself.

At the end of the food-preparation chain, the preparation of ready cooked dishes and their preservation by chilling, deep freezing, vacuum packing or canning, and the catering industry with its kitchen and serving facilities must be mentioned. Stainless steel is extensively used in equipment for these industries where cleanability and hygiene are a prime necessity.

See also: **Barrels**: Wines, Spirits, and Other Beverages; **Beers**: Chemistry of Brewing; Biochemistry of Fermentation; **Brandy and Cognac**: Armagnac, Brandy, and Cognac and their Manufacture; **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant; **Contamination of Food**; **Corrosion Chemistry**; **Meat**: Slaughter; **Milk**: Processing of Liquid Milk; **Sugar**: Refining of Sugarbeet and Sugarcane; **Vegetable Oils**: Oil Production and Processing; **Vinegar**; **Wines**: Production of Table Wines; Production of Sparkling Wines

Further Reading

Crolet JL, Hooper RAE, Lennartz G and Tytgat G (1975) Contribution à l'élaboration d'une réglementation des matériaux destinés à être mis en contact avec les denrées et substances alimentaires: cas des aciers inoxydables. *Industries Alimentaires et Agricoles* 6: 619–827.

Daufin G and Kerherve L (1990) Le nettoyage des aciers inoxydables dans l'industrie agroalimentaire. In: Lacombe P, Baroux B and Beranger G (eds) *Les Aciers Inoxydables*, Ch. 23, pp. 839–851. Les Vulis: Les Editions de Physique.

Di Caprio G (1976) Le applicazioni nell' industria alimentare. *Gli Acciai Inossidabile*, Ch. 12, pp. 362–390. Milan: Ulrico Hoepli.

Henry P, Blouin J, Fourton S, Guimberteau G and Peynaud E (1972) Sur l'utilisation des aciers inoxydables dans les installations vinicoles. *Connaissance de la Vigne et du Vin* 177–212.

Holah JT and Thorpe RH (1990) Cleanability in relation to bacterial retention on unused and abraded domestic sink materials. *Journal of Applied Bacteriology* 69: 599–608.

Ridenour GM and Armbruster EH (1953) Bacterial cleanability of various types of eating surfaces. *American Journal of Public Health* 43: 138–149.

MICROBIOLOGY

Contents

Classification of Microorganisms

A A A Kwaasi, King Faisal Specialist Hospital & Research Centre, Riyadh, Saudi Arabia

Background

Microorganisms comprise prokaryotes as well as some eukaryotic organisms, and these include bacteria, fungi, viruses, protozoa, and algae. Food can be defined as anything that, when taken into the body, serves to nourish or build up the tissues, or to supply body heat: essentially, an aliment or nutrient. This definition of food encompasses all items that are consumed by man and other animals. Water, beverages, and all edible solid material are food. Food sources are plankton, seaweed, fish, crustacea, mushrooms, molluscs, meat, poultry, fruits, vegetables, edible algae and fungi, as well as grains.

Microbes are in contact with food whether it is on the farm or in the forest, garden, lake, pond, river, or sea. Food in storage, industrial processing machines, supermarkets, warehouses, brewery holdings, restaurants, hotels and public dining places all have, at one stage or another, harbored or been in contact with microorganisms.

A number of microorganisms are in themselves food or can be exploited and used as food. Some are contaminants on food and food products, and a number can also cause food spoilage or produce toxins in foods leading to foodborne illnesses in man and animals. Some microorganisms are also capable of transforming a food's properties in a beneficial way such as the various food fermentation processes. Since there are a large number of different microorganisms with diverse habitats and different attributes, food microorganisms will be classified here in terms of the way in which they are associated with foods in different habitats, cause food spoilage, contaminate food, or harm the consumer by the production of noxious metabolites. Microorganisms are ubiquitous and can survive and grow at different temperatures at which food is either kept or processed. They exist from the coldest of salty ponds in the frozen wasters of the polar regions to the boiling water of hot springs. Actively growing bacteria may occur at temperatures in excess of 100 °C in thermal volcanic vents at the bottom of the deeper parts of the oceans, in acid wastes from mines, or in alkaline waters of soda lakes. Microbes have been isolated from black anaerobic silts of estuarine mud and also have been found in the purest waters of the biologically unproductive or oligotrophic lakes. This means that, irrespective of the environment, microorganisms can find a way of adapting, surviving, and multiplying at different temperatures, and they exist in a similar manner on food.

Microorganisms on Food Crops in the Farm or Garden

Crops are established by sowing seed or planting vegetatively propagated materials such as tubers, corms, stem cuttings, or suckers. These vegetative materials are rich in food reserves and, given suitable conditions of temperature, moisture, and light, they are hydrolyzed into simple sugars, which are then utilized in the early growth of the rudimentary young plant or seedling. Under certain conditions, however, these sugars are utilized by soil-inhabiting fungi with the result that the seed or propagative material rots before it can germinate to produce a seedling, or else causes the rotting of the young seedling at soil level. An example of such a microorganism is the saprophytic fungus *Pythium ultimum*, which causes the diseases, referred to as pre- and post-emergence damping-off. *Rhizoctonia solani* causes damping-off on a wide range of cultivated plants – cereals, potato, root and fodder crops, legumes, and vegetables. On lettuce, this fungus is often associated with some *Pythium* spp. and pathotypes of *Botrytis cinerea*. In most cases, the fungus remains quiescent on the consumable material and only grows actively when conditions are favorable. Onions are commonly attacked by *Botrytis alli*, and beet is often colonized by *Pleospora bjoelingii* (*Phoma betae* in pycnidial stage).

Viruses are no exception in their association with foods, for several viral 'mosaics,' 'crinkles,' and

'yellows' are caused by the virus called arabis mosaic virus, which mostly affects strawberries, raspberries, and grapes.

Plant pathologists and agriculturists often refer to microorganisms as being field or storage microorganisms. Microbial flora of cereals undergo pronounced and significant changes in the field. Grains may become colonized by field fungi, and these include species of *Cladosporium, Alternaria, Helminthosporium*, and *Chaetomium*. After harvest, however, the numbers of these field flora decrease, giving room to storage fungi, the commonest of which are *Penicillium, Aspergillus*, and some nonpathogenic species of *Fusarium*. When leaves unfold, they act as spore traps for airborne microbial propagules, some of which grow and form the phylloplane, or leaf surface flora. Close observation has revealed that these propagules can be found on leaves that animals graze and man uses in cooking or as raw components of salads. The majority of these are fungi of the *Ascomycotina* and the commonest recorded are *Cladosporium* spp., *Epicoccum nigrum, Aureobasidium pullulans*, and *Alternata tenuis*. There are frequently true yeasts of the genera *Sporobolomyces* and Bullera, the so-called mirror yeasts on plant leaf surfaces.

An interesting and richer association of yeast in the wild is to be found with nectaries of flowers and the surfaces of fruits. The presence of some of these fungi is a prerequisite in the spontaneous fermentation of fruit juices, such as grape in the production of wine. Palm wine obtained from the oil palm (*Elaeis guineensis*), date palm (*Phoenix dactylifera*), and the raffia palm (*Raphia vinifera*) in parts of the world are also fermented in a similar manner by wild epiphytic and airborne yeasts.

Food crops are commonly colonized by Gram-negative rods such as *Erwinia, Pseudomonas*, and *Xanthomonas*, but several Gram-positive bacteria such as *Lactobacillus* and *Leuconostoc* are present, which are important in the production of fermented vegetable products such as sauerkraut (fermented pickled cabbage) (Table 1). When leafy vegetables that are consumed fresh are colonized, it is almost impossible to eliminate these microflora completely by ordinary washing, but in most cases they are harmless to the consumer.

A rather unusual association of some microorganisms and food plants can be found in the rhizoplane. Root nodules of the Leguminosae are formed as a result of a symbiotic association of a bacterium *Rhizobium melioloti* with root nodules of alfalfa (*Medicago sativa*). This symbiotic relationship greatly increases the nitrogen content not only of the soil but also of the plant crop at maturity. Some mycorrhizal associations between fungi and roots of some food plants not only offer protection to the roots but also contribute indirectly to their overall crop yield. Similarly, some microorganisms can establish associations with members of the botanical family Fabaceae (Papilionaceae and Leguminosae), which provide man and other animals with the most important plant proteins, the pulses. In spite of their low water content after drying, they are usually attacked by lipolytic molds of the genera *Aspergillus* (such as *A. niger, A. tamarii), Penicillium* and *Paecilomyces*.

The ability of some soil microorganisms to degrade complex organic materials makes these microbes potent spoilage agents, if they are present on foods. At times, the soil can be a harsh and competitive environment, and many soil microorganisms adapt to these sudden changes by producing resistant structures such as endospores (by the bacterial genera *Bacillus* and *Clostridium*), chlamydospores, and sclerotia of many fungi (such as *Claviceps purpurea* or ergot). The subsequent germination of these structures is responsible for the occurrence of food spoilage and food poisoning caused by the clostridia and bacilli. The spore-forming anaerobe *Clostridium* can be predominantly proteolytic or saccharolytic. Both of these activities are accompanied by gas production and cause some infected cans on supermarket shelves to swell, resulting in the 'blown can' effect.

Microorganisms on Food in Transit or in Storage

Food in transit is often damaged by a number of microorganisms if conditions are ideal for their growth. Considerable losses of vegetables such as carrots, potatoes, onions, and celery are caused by bacteria causing soft rot and the species belonging to the Genus *Erwinia*, such as *E. carotovora, E. atroseptica* (black leg), *E. aroideae*, and *E. chrysnthemi*. This situation is usually worsened when there are bruises and lesions, but at high temperatures and high humidity, *Bacillus polymxya* is capable of reducing potato in storage to a yellow sticky mass with a distinct fruity odor. *Pseudomonas phaseolicola* perennates on dwarf bean (*Phaseolus vulgaris*), and this same fungus causes the diseases leaf spot and halo blight.

Fungal survival, or their contamination of foods, in transit is a common phenomenon, and several genera are involved. Species of the fungus *Rhizopus* are capable of causing considerable damage to peaches, grapes, sweet potato, crucifers, cucurbits, tomatoes and eggplant (*Solanum melongena*) and produce soft rot very rapidly at temperatures above 10 °C. Potatoes are commonly attacked by *Fusarium caeruleum*, causing dry rot of the tuber. Ergot of cereals is a disease caused by members of the genus *Claviceps*.

Table 1 Foodborne pathogenic bacteria

Organism	*Food source*	*Illness*
Salmonella spp.	Poultry, eggs, milk, swine, water, shrimps, raw meats	Typhoid and typhoid-like fever
Clostridium botulinum	Any food with pH above 4.6; all canned meats, honey, vegetables	Foodborne botulism (intoxication)
Staphylococcus aureus	Most foods not kept hot or cold enough	Staphylococcus food poisoning (staphyloenterotoxicosis)
Yersinia enterolitica	Meats, icecreams, milk	Yersiniosis, bacteremia, arthritis
Campylobacter jejuni	Water, unpasteurized milk, meat, poultry	Enteritis, gastroenteritis, campylobacteriosis
Listeria monocytogenes	Milk, cheese, icecream, vegetables, fermented raw meats, all meats raw or smoked	Listeria meningitis
Vibrio cholerae	Contaminated water (poor sanitation)	Cholera
Vibrio parahaemolyticus	Raw improperly cooked or contaminated fish and shellfish	Gastroenteritis, diarrhea, cramps, nausea
Vibrio vulnificus	Oysters, clams, crabs (raw or contaminated)	Gastroenteritis, wound infection, septicemia
Clostridium perfringens	Improperly prepared food: meats, meat products, gravy	Food poisoning, pig belly disease
Bacillus cereus	Meats, milk, vegetables, fish (vomiting is associated with rice, potato, pasta, cheese, and food mixtures)	Diarrhea-type, food poisoning, vomiting
Aeromonas hydrophilia	Beef, pork, lamb, poultry, shellfish, fish	Gastroenteritis to septicemia
Plesiomonas shigelloides	Contaminated water or raw shellfish	Gastroenteritis (self-limiting), fever, chills
Shigella spp, *S. sonnei*, *S. boydii*, *S. flexneri*, *S. dysenteriae*	Salads, raw vegetables, milky dairy products, poultry	Bacillary dysentery
Miscellaneous Gram-negatives: *Klebsiella*, *Enterobacter*, *Proteus*, *Citrobacter*, *Aerobacter*, *Providencia*, and *Serratia*	Dairy products, raw shellfish, fresh vegetables, water	Gastroenteritis
Group A *Streptococcus* spp: *S. pyogenes*	Group A: milk, icecream, eggs, steamed lobster, ground ham, potato salad, egg salad, custard, rice pudding, shrimp salad	Tonsillitis, high fever, septic sore throat
Group D *Streptococcus* spp. *S. faecalis*, *S. faecium*, *S. durans*, *S. avium*, *S. bovis*	Dairy products, raw shellfish, fresh raw vegetables	Scarlet fever, diarrhea, abdominal cramps, vomiting, septicemia
Enterotoxigenic, *Escherichia coli*	Contaminated water, dairy products, e.g., semisoft cheeses	Gastroenteritis, travellers' diarrhea
Enteropathogenic *E. coli*	Raw beef, chicken, and any food exposed to fecal contamination	Infantile diarrhea
Enteroinvasive *E. coli*	Hamburger, meat, unpasteurized milk	Mild dysentery

It is widely distributed throughout the world, and the most economically important is ergot of rye caused by *Claviceps pupurea*. This fungus is important not so much for the effect it has on the growth of crop but for the characteristic fungal fruiting structures called ergot (sclerotia), which can contaminate cereals during harvest and cause poisoning in man and animals. Feeding livestock with cleanings from contaminated grain or grazing pastures with infected grass heads can lead to abortions and peripheral gangrene.

Bread can be infected with the bread mold at several locations, even in the fridge, and it has been known to cause infections due to colonization by *B. cereus*. There have been a number of instances where survival of pathogens or their toxins have caused problems in products such as dried herbs, rice, dried milk, chocolate, pasta, and eggs, and generally, *Salmonella* and *Staphylococcus aureus* have been associated with these outbreaks. Xerosporic (nonwater wettable) microorganisms can survive almost indefinitely on dried products.

Bananas in transit are subject to several fungal rots, and the most troublesome and commonest of these, which is considered largely responsible for the decline in the quality of Jamaican bananas imported into England, is attributed to rotting by *Gleosporium musarum*. Stone and pome fruits are particularly prone to brown rot, and the three main fungi involved belong to the genus *Sclerotinia*. *Sclerotinia fructigena* and *Sclerotinia laxa* occur in Europe, and *Sclerotinia fructicola* occurs in the USA. Apples infected by *Sclerotinia fructigena* are initially covered with conidial cushions, and the infected fruits gradually shrivel and become mummified. A species of this fungus *Sclerotinia sclerotiorum* is particularly dangerous, because it can continue to cause rotting of vegetables at temperatures near freezing point. Fungi also cause considerable damage to citrus fruits in transit: *Penicillium italicum* and *Penicillium digitatum* cause the diseases called blue and green mold, respectively, and *Penicillium expansum* causes a soft rot in apples. Other common diseases of apples and pears include the black spot or scab caused by the Ascomycotina fungus called *Venturia inaequalis* (anamorph *Spilocaea pomi* = *Fusicladium dendriticum*). Lemons are transported under relatively high temperatures, and this condition permits the development of *Phytophthora* brown rot.

An especially widespread mold on fruit and vegetables is the gray mold *Botrytis cinerea*, which is an imperfect stage of the Ascomycete, *Botryotinia fuckeliana* (= *Sclerotinia fuckeliana*). In canned fruits, the low pH and heat treatment means that most molds are killed, but some members of the Eurotiales are sufficiently heat-resistant and survive. The best known are species of *Byssochlamys*, but recent spoilage of some exotic fruits in canneries has been due to *Neosartorya fischeri* (anamorph *Aspergillus fischerianus*) and *Talaromyces flavus* var. *macrosporus* (anamorph *Penicillium* sp.).

Survival of remnants of fungal pathogens on seeds is very common: it is usual to find a reservoir of the loose smut disease agent of wheat and barley *Ustilago nuda* in the scutellum. Similarly, the hyphae of the fungus that causes loose smut of oats, *Ustilago avenae*, are common on the embryo, seed, or fruit coat of oats, and *Tilletia caries* and *Septoria apiicola* can be commonly detected on the seeds of wheat and celery, respectively. Species of the fungus *Aspergillus* can cause considerable deterioration of stored grain. The best studied is *A. flavus*, which is also capable of producing several toxic materials, collectively known as aflatoxin. This toxin, which is carcinogenic, has been found in most stored products such as peanuts, dates, and an array of foods.

Relatively few viruses survive on, and are transmitted by, seeds, and lettuce mosaic virus is a common example. When crops are harvested, they may harbor several microbes. As a rule, field microbes give way to the colonization of storage microorganisms, often resulting in fruit and fleshy storage organs rotting, and dry cereal grains becoming moldy. The pathogens responsible for this type of spoilage are frequently those that do not develop at all, or develop only to a limited extent on the host in the field. An example of such an organism is *Fusarium solani* var. *coeruleum*, which causes dry rot of potato tubers. Similarly, potato blight (*Phytophthora infestans*) can be initiated from tubers. The air of farmyard barns may contain many millions of spores of Actinomycetes per cubic meter. Spores of some species such as *Thermoactinomyces vulgaris* and *Micropolyspora faeni* cause lung allergy in farmers, and some geosmin-producing strains of *Streptomyces* may be responsible for earthy odors and off-flavors in potable water; also, geosmin can impart earthy taints to food such as shellfish.

Microorganisms in Water

Aquatic environments represent the largest part of the biosphere, and many species of microorganisms are found in both marine and freshwater environments. Open oceans contain salt and many marine bacteria, referred to as oligotrophic psychrophiles, which have a requirement for salt and can grow at relatively low temperatures. The surfaces of fish caught from cold water in the open sea contain psychrophilic and psychrotrophic bacterial flora that are capable of breaking down proteins, polysaccharides, and lipids, and can double in numbers at refrigeration temperatures of 0–7 °C in as short a time as 10 h. The implication of this is that a single bacterium on freshly caught fish can multiply to 2^{22} or between 10^7 and 10^8 in just 9 days, and in this situation, cold-stored fresh fish starts giving off off-odors.

The sea has been used by man as a dumping ground for sewage and other waste products. Unfortunately, many shellfish used for food grow in these polluted coastal waters and actually feed by filtering out particles from large volumes of seawater. Therefore, there is always a high risk of contamination by enteric organisms from infected individuals, and the microbes concentrate in numbers within the organism as a result of the filter-feeding habits of shellfish. Severe diseases such as hepatitis and typhoid fever, and milder illnesses such as gastroenteritis, have been caused by eating contaminated oysters and mussels that look sound in appearance and taste normal. In warmer seas with unpolluted water, high numbers of the bacterium *Vibrio parahaemolyticus* may also be concentrated through the filter feeding of shellfish.

This bacterium then becomes part of the stable commensal enteric flora and has been responsible for episodes of food poisoning associated with seafood.

Fresh water from lakes and rivers has a complex flora, which includes genuinely aquatic microbial species as well as those introduced from terrestrial, animal, and plant sources. Like other water sources, fresh water may act as a repository and vehicle for bacteria, protozoa, and viruses of fecal origin, which can also cause diseases.

Fungi are also found in marine and fresh waters, but they are not as important in their association with food as the groups of microorganisms described above. There are true aquatic fungi, and these include some that are pathogens of molluscs and fish. The most well-adapted aquatic fungi belong to all the major groups of terrestrial fungi, namely: *Ascomycotina, Basidiomycotina, Deuteromycotina*, and *Zygomycotina*. There is the possibility that some of these could be responsible for spoilage of some food commodities, including salad crops cultivated with overhead irrigation from rivers or lakes. The aquatic photosynthetic microorganisms, the cyanobacteria (or blue–green algae), and the eukaryotic dinoflagellates are groups of microorganisms that may also become concentrated in shellfish and can cause severe illness such as paralytic shellfish poisoning.

A significant global problem is the viral contamination of water. In the past, about half of the foodborne disease outbreaks through water were unrecognized, primarily as a result of inadequate diagnostic methods and sampling. Recently, with improved diagnostic methods, it has been established that most of these food and waterborne outbreaks are caused by caliciviruses and are a major cause of seafood associated gastroenteritis.

Bacterial strains of the genus *Aeromonas* are well-recognized enteropathogens of man. The main source of infection by this organism seems to be untreated water, because these microorganisms can be found in virtually all aquatic environments. Furthermore, some enterotoxigenic pathotypes of *Aeromonas* are capable of rapid growth and toxin production at temperatures as low as 5 °C. This means that vegetable products irrigated with contaminated water may reach critical *Aeromonas* levels during refrigeration and could represent a serious public health hazard when these vegetables are consumed as uncooked salads.

Microorganisms in Milk

Raw milk becomes laden with microorganisms from three sources: udder interior, teat exterior and its surrounding, and the milking and milk handling equipment. Milk expressed aseptically from a healthy cow may be sterile or contain low numbers of microorganisms, usually about 10^2–10^3 cfu ml^{-1}. The most frequently isolated organisms are micrococci, streptococci, and *Corynebacterium bovis*. Higher microbial counts are often seen if the animals suffer from mastitis, and under these circumstances, many types of bacteria are seen. The most important of these are the potential human pathogens *Staphylococcus aureus*, *Escherichia coli*, *Streptococcus agalactiae*, *Streptococcus uberis*, *Pseudomonas aeruginosa*, and *Corynebacterium pyogenes*. Occasionally other human pathogens, such as *Salmonella*, *Listeria monocytogenes*, *Mycobacterium bovis*, and *Mycobacterium tuberculosis*, are isolated.

The udder exterior, the teats, and the cow's immediate environment such as bedding and manure can also be a source of heavy contamination (10^5 cfu ml^{-1}) of the milk. In these situations, the microorganisms involved are mostly human pathogens such as *E. coli*, *Helicobacter* (*Campylobacter*) from bedding and manure, *Salmonella* and *Bacillus* spp. from soil, and *Clostridium butyricum* and *Clostridium tyrobutyricum* from silage. These *Clostridium* spp. are the cause of a form of spoilage called 'late blowing cheese.' In some parts of the world, a potent contaminant of milk and meats is caused by *Brucella* spp.

In most developed countries, milk is chilled soon after it is issued from the cow, and under these circumstances, only psychrotrophic organisms are found in raw milk. The most common are Gram-negative rods of the genera *Pseudomonas*, *Acinetobacter*, *Alcaligenes*, and *Flavobacterium*, the psychrotrophic coliform *Aerobacter*, and Gram-positive *Bacillus* spp. In order to reduce the numbers of harmful organisms, milk is either pasteurized or ultraheat-treated. Raw milk after pasteurization may contain organisms, referred to as thermodurics, that can survive this process. These organisms are generally Gram-positive, such as the spore-forming bacteria and members of the genera *Microbacterium*, *Micrococcus*, *Enterococcus*, and *Lactobacillus*, but it is usual to find some strains of the Gram-negative bacterium *Alcaligenes tolerans*. Spoilage of pasteurized milk is due to the growth in milk of the Gram-negative rods *Pseudomonas*, *Alcaligenes*, *Acinetobacter* and *Psychrobacter*, which may be inadvertently introduced after pasteurization. *Salmonella* has been isolated from icecream that has been stored at −23 °C for 7 years.

Microorganisms on Meat

In theory, freshly harvested meat should be relatively free from microorganisms, but microbial numbers

detected in aseptically sampled tissues contain some microbes, usually less than 10 cfu kg^{-1}. There is evidence, however, that these numbers can increase under conditions of stress and if the animal is suffering from an infection, as in the case in *Brucella*-infected animals. Most meat contamination is from the heavily colonized areas of the animal, such as the skin (fleece) and the gastrointestinal tract, and the type and numbers found will reflect both the animal's indigenous microflora and its environment. The animal hide carries mixed microbial population of micrococci, staphylococci, pseudomonads, yeasts, and molds as well as organisms derived from soil and feces. After dressing and chilling, the surface microbial numbers are typically of the order of 10^2–10^4 cfu cm^{-2} and are usually higher in sheep carcasses than in beef, and even higher still in pigs. In some parts of the world, the fur on sheep, cattle and pigs is thoroughly burned soon after slaughtering the animal. Traditionally, the method is used as a way of imparting flavor to the meat, but it should also be lauded and seen as a first step in reducing the microbial numbers on the skin that could otherwise contaminate the flesh. Likewise, after defeathering, in some traditions, birds are flamed before the flesh is cut, and this treatment also greatly reduces the potential contamination of the poultry meat by microbes on the skin.

Species of the genus *Brucella* are known human pathogens that can cause undulant fever, and are associated with a particular animal host: *B. abortus* in cattle, *B. melitensis* in sheep and goats, *B. suis* in pigs, and *B. canis* in dogs. Even though the diseases are contracted from close contact with infected animals, they can also be contracted by the consumption of milk, milk products, and meat from these animals.

The processing of poultry is different from red meat, and this has microbiological implications: an active processing plant that can handle 12 000 birds per hour leaves little chance for effective sanitation and favors the spread of microbes between carcasses. After scalding, birds are mechanically defeathered, and a number of studies have suggested that this method actively passes organisms like *Salmonella* from one carcass to the other. The intestinal tract of poultry also contains high numbers of the human pathogens *Salmonella* and *Helicobacter* (*Campylobacter*); these are easily passed on and are the two major causes of foodborne illness in the UK. Interestingly, *Campylobacter coli* and *C. jejuni* do not grow at temperatures below 30 °C and only become pathogenic when food contaminated by these bacteria is consumed and the microbes start thriving under the higher human body temperature. Being microaerophillic, they are also able to survive the low-oxygen conditions of modern storage processes.

The first indication of spoilage of meat is the production of off-odors, which become apparent when the microbial numbers in meat reach 10^7 cfu cm^{-2}. Aerobic storage of chilled red meats, whether covered or uncovered, produces a high redox potential at the meat surface that favors the growth of psychrotrophic aerobes. Nonfermentative Gram-negative rods grow most rapidly, causing spoilage, and the principal genera are *Pseudomonas*, *Acinetobacter*, *Psychrobacter*, *Pseudomonas fragi*, and *Pseudomonas lundensis*. Other microbes that form a minor component of the spoilage microflora are members of the Enterobacteriaceae, such as *Serratia liquefasciens* and *Enterobacter agglomerans*, lactic acid bacteria and the Gram-positive bacterium *Brochothrix thermosphacta*.

In vacuum-packed meat, the microflora of the meat changes as a result of the accumulation of CO_2 and lack of oxygen. This restricts the growth of *Pseudomonas*, and colonization is dominated by Gram-positives, lactic acid bacteria of the genera *Lactobacillus*, *Carnobacterium*, and *Leuconostoc*.

The perception of spoilage is subject to a number of influences, particularly social: foods acceptable in some cultures are unacceptable in others. Matured cheeses and game birds (e.g., pheasants) that have been hung for several weeks are seen in some cultures as objectionable, but these same products are treasured and sold for large amounts of money to the affluent in other societies.

Microorganisms on Fish (Teleosts and Elasmobranchs), Crustaceans, and Molluscs

Like meat, the muscle and internal organs of freshly caught healthy fish are usually sterile, but the skin, gills, and alimentary track carry large numbers of bacteria. These can be as high as 10^7 on the skin and up to 10^9 in the gills and the gut, and these are mainly Gram-negatives of the genera *Pseudomonas*, *Shewanella*, *Psychrobacter*, *Vibrio*, *Flavobacterium*, and *Cytophaga*, as well as a few Gram-positive micrococci and coryneforms.

Fish can be easily contaminated by human pathogens during handling on board, at the dock, and at markets when they are being offered for sale. The common foodborne illness pathogens associated with fish are *Vibrio cholerae*, *Vibrio parahaemolyticus*, *Vibrio vulnificus*, *Clostridium botulinum* Type E, enteric viruses, and those that cause Scombroid fish poisoning and paralytic shellfish poisoning.

Spoilage of chilled fish is mainly due to the activity of the psychrotrophic Gram-negative rods *Shewanella putrefaciens* and some *Pseudomonas* species.

Shewanella putrefaciens can grow under a variety of conditions, producing trimethylamine and hydrogen sulfide to spoil even vacuum-packed fish. Work in Denmark has demonstrated that marine Vibrios like *Photobacterium phosphorous* cause nonsulfurous spoilage of fish produce.

Foodborne Viruses

Viruses are obligate intracellular parasites and can only multiply in a susceptible host cell. Recently, viruses have been recognized as an important source of foodborne diseases (Table 2), and over one hundred human enteric viruses have been implicated. These viruses are spread mainly by the fecal–oral route, making foods for human consumption a potential route for transmission. However, the genus enterovirus includes the polio and the hepatitis A and E viruses, all of which are known to be foodborne. Hepatitis A is transmitted by the fecal–oral route, primarily spreads by person-to-person contact, but food- and waterborne outbreaks do occur. Foods such as milk, strawberries, raspberries, lettuce, and shellfish are common food vehicles.

The causative agent of enterically transmitted nonA nonB virus, which is now designated as hepatitis E virus, is also transmitted fecal–orally. Other viruses have been implicated in gastroenteritis by their sheer large numbers (10^8–10^{10}) and are also known to cause diarrhea and other foodborne gastroenteritis. Most of these enteric viruses may be introduced into foods either as primary contamination at source of food production (harvest) or as secondary contamination during handling, preparation, or serving. Fruits and vegetables that are fertilized with human excrement are thought to be the sole source of contamination. Shellfish are grown commonly in shallow, inshore, coastal waters and are often contaminated with viruses. The problem is compounded by the fact that some shellfish such as oysters are consumed with little or no cooking (as in most Mediterranean countries). In the UK, strict guidelines aimed at inactivating the hepatitis A virus exist for the cooking of cockles.

Spongiform encephalopathies (SEs), Creutzfeldt–Jakob disease, and other scrapies are degenerative disorders of the brain caused by the so-called 'slow viruses,' which are contracted by consumption of meat from infected animals. Lassa fever, which occurs in Nigeria, Guinea, Liberia, and Sierra Leone, is caused by an arenavirus in infected rats, the main reservoir of infection.

Foodborne Protozoa

Only a few protozoa are associated with food or are foodborne (Table 3); the flagellate *Giardia*, the amoeboid *Entamoeba* and three sporozoids of the phylum *Apicomplexa*, which include the genera *Toxoplasma*, *Sarcocystis*, and *Cryptosporidium*, are some of the commonly encountered protozoa.

Giardia lamblia (also called *G. intestinalis* or *Lamblia intestinalis*) survives on food and water as cysts. The cysts, which are aided by gastric juices to release the infective flagellate protozoa, have been found on salad vegetables such as lettuce, fruits such as strawberries, and any food that is washed with contaminated water or handled by infected persons. The cysts are resistant to chlorination but can be disabled if the food is well cooked. *Entamoeba histolytica*, which causes amoeboid dysentery, is widespread and almost endemic in many poor communities. It is estimated that an actively infected person can shed up to 50 million cysts per day, and the potential of these to cause fecal–oral infections cannot be overemphasized. It has been suggested from epidemiological studies that foods such as raw sausage are a risk factor in the sporozoid infection cryptosporidiosis. The organism *Cryptosporidium parvum* can go through its entire life cycle in a single host, and the host could be human, cattle, or sheep. Two species of *Sarcocystis*, however, are obligate parasites: *S. hominis*, which

Table 2 Foodborne pathogenic viruses

Organism	*Food source*	*Illness*
Hepatitis A	Contaminated water and food, sandwiches, fruit and fruit juices, milk and milk products, vegetables, salads, shellfish and iced drinks	Mild fever, malaise, nausea, anorexia, abdominal discomfort and jaundice
Hepatitis E	Water	Malaise, anorexia, abdominal pain, arthralgia, fever
Rotavirus	Salads, fruits, hors d'ouvres	Viral gastroenteritis, infantile diarrhea, winter diarrhea
Norwalk viruses	Water and foods	Nausea, vomiting, diarrhea, abdominal pain
Other gastroenteritis viruses	Food and water	Nonbacterial viral gastroenteritis
Parvo-like virus	Shellfish	Viral gastroenteritis

Table 3 Foodborne pathogenic protozoa

Organism	*Food source*	*Illness*
Giardia lamblia	Contaminated water, food contaminated by food-handlers, contaminated vegetables	Giardiasis, nonbacterial diarrhea
Entamoeba histolytica	Drinking water and foods	Amebiasis
Cryptosporidium parvum	Raw milk and sausage, and any food touched by infected food-handlers	Intestinal, tracheal and pulmonary cryptosporidiosis
Cyclospora cayetanensis, *Acanthamoeba* spp. and *Naegleria fowleri*	Food, water, fresh fruits and vegetables (basil, berries, mesclun lettuce)	Cyclosporiasis infection of small intestines, diarrhea, weight loss, nausea, stomach cramps, muscle ache, vomiting, fatigue

infects cattle, and *S. suihominis* from pigs cause human infections if beef and pork have been undercooked. Even though domesticated or wild cats are hosts to *Toxoplasma gondii* this organism can cause foodborne infections in humans through consumption of undercooked pork or mutton.

Prokaryotic cyanobacteria and blue–green algae are microorganisms, as are the benthic and the planktonic algae. These are used as food and are very expensive in restaurants in different parts of the world. However, some of these can produce very toxic compounds that may be transported to mussels, clams, and small herbivorous fish. Shellfish are not harmed or affected by these high levels of toxins, but a number of distinct illnesses from these sources are now recognized as a result. The illnesses they cause include paralytic shellfish poisoning, neurotoxic shellfish poisoning, diarrheal shellfish poisoning, amnestic shellfish poisoning, and ciguatera fish poisoning. The dinoflagellates *Gonyaulax catenella* and *G. tamarensis* (now *Alexandrium*) cause paralytic shellfish poisoning. Some cyanobacteria, including *Microcystis*, *Anabaena*, and *Aphanizomenon*, are main constituents of algal blooms in lakes, ponds, and reservoirs, and can easily contaminate water drunk by animals and man. It is worth noting here that protozoa are important in assisting in the breakdown of organic matter in ruminant nutrition.

Microorganisms Exploited by Man for Food

The fact that microorganisms can transform organic matter was not known until the middle of the nineteenth century. However, man had used certain microbial metabolic processes, since prehistoric times, for the preparation of food and drink, e.g., in the production of beers and wine, the leavening of bread, and the manufacture of vinegar, cheese, butter, and, more recently, yogurt. The advent of microbiology has led to the development of entirely new industries based on the use of microorganisms not previously exploited by man. In wine production, the grapes are pressed, and the mulch yeast, molds, and bacteria derived from the surface of the ripe grapes, including the so-called true wine yeast, *Saccharomyces cerevisiae* var. *ellipsoids*, start the fermentation process.

The soil has a rich reservoir of microorganisms, and many of these microbial strains have been exploited for use in industrial production of enzymes, amino acids, antibiotics, enzymes, and vitamins that are used in the food industry.

Modern yogurt processing involves inoculating milk with *Streptococcus thermophilus* and *Lactobacillus bulgaricus*. For the production of Russian kefir, milk is inoculated with kefir grain, which is essentially symbiotically growing *Lactobacillus* and yeasts. Indian Idhi (Dosai) used in south India is essentially fermented by *Leuconostoc mesenteroides*, but *Streptococcus faecalis* is also present and contributes to its acidic content.

In West Africa, the use of naturally fermented acidic porridges and dough forms the basis of their staple 'ethnic foods'. Gari (a fermented and fried cassava grit) is prepared from an acid-fermented ground cassava. The principal fermenter is *Corynebacterium manihot*, which hydrolyzes the starch, producing lactic acid and formic acid (evolving heat). At the optimum fermentation temperature of 35 °C, any cyanide-containing sugars present in the cassava (*Manihot utilissima*) are hydrolyzed, removing the cyanide. As the product becomes more acidic (around pH 4.25), a yeast-like fungus *Geotrichum candidum* (also found in Camembert cheese) develops. Other acid porridges are made from millet and maize (ogi and koko), maize and wheat (mahewu), and sorghum. Production of ogi and mahewu proceeds better at temperatures of about 50 °C, which also favors the rapid development of *Lactobacillus delbrueckii*.

Edible mushrooms and some types of seaweed can be considered as foods that are themselves microorganisms. Mushrooms are the fruiting bodies of the fungi and are mainly the stalk surmounted by the cap, which in reality is a mass of closely packed hyphae.

Many lactic acid bacteria occur normally in milk and are responsible for souring. They produce large amounts of acid, which inhibits the subsequent development of other microorganisms. The manufacture of cheese involves two main steps: curding and ripening, and this ripening step is carried out by the action of various bacteria and fungi. Ripening of the curd is a complex process: Cheddar and ordinary American cheese is ripened by lactic acid bacteria in the curd. As these organisms die, proteolytic and fat-splitting enzymes are released from the cells, slowly breaking down the milk fat and proteins with the formation of materials that impart the characteristic cheese flavor. The so-called mold-ripened cheeses such as blue cheese and Camembert are produced by inoculating the curd with special kinds of fungi, which develop either throughout the curd, as in blue cheese, or over its surface, as in Camembert. Swiss cheese is ripened by propionic acid bacteria, which ferment the lactic acid present in the curd to propionic acid, acetic acid, and carbon dioxide. The 'holes' in Swiss cheese are produced by carbon dioxide and the characteristic flavor by propionic acid.

Butter manufacture is also in part a microbiological process, since the microbial souring of the cream is desirable for a good subsequent separation of the butterfat during the churning process. Certain lactic acid bacteria found on plant materials are responsible for souring processes that occur in the preparation of pickles. Some lactic acid bacteria belonging to the genus *Leuconostoc* produce large amounts of the extracellular polysaccharide dextran from sucrose.

In the Far East and Pacific, seaweed has been used as a source of food for millennia. Seaweed's diversified composition makes it far superior to higher plants nutritionally being rich in polysaccharides, minerals (especially iodine), and vitamins, and containing 33–35% total fiber. The green alga *Ulva lactuca* (sea lettuce) is occasionally used fresh in salads, and members of the family Fucaceae like *Fucus* spp. (also called rockweed or bladderwrack) and *Ascophyllum* spp. are used in animal feeds. Rockweeds have been used in recipes like clambakes, as flavorings, and as teas. *Laminaria longicruris* (oarweed or kelp) is a useful source of algin in oriental markets, sold as 'kombu' in health food stores, or may be cooked as vegetables or added to soup. *Laminaria digitata* is used as a vegetable and as flavoring in baked beans. Red algae of the *Porphyra* spp. have seen various uses as a seasoning, in soups, and as a constituent of leven bread.

Microorganisms have been used to produce certain foods, beverages, condiments, and animal feeds, and recently, several new commercial microbial processes have been developed. These include the production of single-cell proteins from microbes to supplement animal feeds, mushrooms (*Agaricus campestris*) for human foods from agricultural wastes (from beet and cane sugar molasses), microbial rennet for cheese making, meat-like flavorings using Chinese soy sauce and Japanese miso processes (employing *Aspergillus soyae* and *Aspergillus oryzae*) xantha, and some vitamins. Algae have also been used as a source of single-cell protein, and the genera *Chlorella* and *Scenedesmus* have been grown for food in Japan. *Spirulina* species have been eaten for many years by the inhabitants of the northern shores of Lake Chad in Africa and by the Aztec Indians in Mexico.

See also: ***Bacillus***: Occurrence; Food Poisoning; ***Clostridium***: Occurrence of *Clostridium perfringens*; Detection of *Clostridium perfringens*; Food Poisoning by *Clostridium perfringens*; Occurrence of *Clostridium botulinum*; **Contamination of Food**; ***Escherichia coli***: Food Poisoning; Food Poisoning by Species other than *Escherichia coli*; **Food Poisoning**: Classification; ***Listeria***: Listeriosis; **Microbiology**: Classification of Microorganisms; Detection of Foodborne Pathogens and their Toxins; **Mycotoxins**: Classifications; ***Salmonella***: Salmonellosis

Further Reading

Adams MR and Moss MO (2000) *Food Microbiology*. Cambridge: The Royal Society of Chemistry.

Anonymous (2001) *The Bad Bug Book: Foodborne Pathogenic Microorganisms and Natural Toxins Handbook*. US Food & Drug Administration Center for Food Safety and Applied Nutrition. College Park, MD, USA: US Government Printing Office.

Chan EC and McManus EA (1965) Distribution, characterisation and nutrition of marine microorganisms from the algae *Polysiphonia lanosa* and *Ascophyllum nodosum*. *Canadian Journal of Microbiology* 5: 409–420.

Garrity GM (ed.) (2001) *Bergey's Manual of Systematic Bacteriology*. New York: Springer-Verlag.

Lucas JA (1998) *Plant Pathology and Plant Pathogens*. Oxford: Blackwell Science.

Microbial Processes: Promising Technologies for Developing Countries (1979) Washington, DC: National Academy of Sciences.

Philips CA (1997) Preserving food: a chance in the atmosphere. *Biologist* 44(2): 301–303.

Stanier RY, Doudoroff M and Adelberg EA (1972) *General Microbiology*. London: Macmillan.

Stephen J and Pietrowski RA (1981) *Bacterial Toxins*. Walton-on-Thames, UK: Thomas Nelson.

Wheeler BEJ (1969) *An Introduction to Plant Diseases*. London: John Wiley.

White S and Keleshian M (1994) *A Field Guide to Economically Important Seaweeds of Northern New England*. University of Maine/University of New Hampshire Sea Grant Marine Advisory Program.

Detection of Foodborne Pathogens and their Toxins

J W Austin and F J Pagotto, Bureau of Microbial Hazards, Ottawa, Canada

Background

Microbiologists often need to enumerate bacteria in food samples or determine the presence or absence of low numbers of specific foodborne pathogens or their toxins. Detection and quantification of foodborne pathogens and their toxins are necessary for several reasons. The most obvious application is for inspection of consumer foods to determine their quality and/or safety. Pathogen detection and enumeration are also necessary for determination of possible sources of contamination during development of a Hazard Analysis and Critical Control Point plan. Enumeration of bacteria, or detection of specific toxins, is necessary for assessing the growth potential of pathogens in foods.

Traditional techniques have relied upon either direct detection using microscopy or, more commonly, cultural methods that involve growth of microorganisms in selective and/or differential culture media. Traditional techniques are still used but are gradually being replaced by more rapid methods employing detection of specific antigens (i.e., immunoassays) or DNA sequences (i.e., nucleic acid hybridization or polymerase chain reaction (PCR)). These latter methods are very specific and allow detection of specific pathogens within minutes or hours, compared with days using traditional cultural techniques. While these newer methods do not require microbial growth and allow for more rapid detection, they typically detect both viable and nonviable microorganisms. When it is desirable to know if viable microorganisms are present, a culture method is preferred.

Direct Detection Using Light Microscopy

Light microscopy allows useful magnifications up to 1000 times and allows examination of food samples for the presence of bacteria and the morphology of the contaminants. Direct microscopic counts involve smearing a known volume of specimen over a defined area of a microscope slide or counting chamber, fixing and staining the specimen, and counting bacterial cells. This method is useful only for samples with at least 10^6 bacteria per milliliter. The best-known example of direct microscopic counts is the Breed smear, which is used to enumerate bacteria in milk. The numbers of bacterial cells in several fields are counted and averaged to obtain the count of bacteria in the original sample.

The direct epifluorescent filter technique involves concentration of microorganisms on membrane filters to increase the limit of detection, followed by detection of specific microorganisms using a fluorescent label. Depending upon the fluorochrome used, the technique can be used for several applications. When stained with acridine orange, for example, viable cells generally fluoresce orange–red, whereas nonviable cells fluoresce green. Fluorescently tagged antibodies specific for bacterial surface antigens can be used to identify specific bacteria.

Cultural Techniques

Cultural techniques usually involve several steps, including enrichment for the target organism, isolation using selective and differential media, and finally identification of the isolate using phenotypic or genotypic analysis. Cultural methods such as the most probable number (MPN) technique and plate counts are used to isolate and enumerate microorganisms in foods.

Enrichment

Preenrichment media are designed to increase the numbers of the target organism, without suppressing the growth of other organisms. Preenrichment may allow repair of damaged organisms and encourage multiplication of the target organism from a few cells to 10^5 or more per milliliter over a specified period of time. Selective enrichment media are then employed to increase the numbers of the target organism, relative to the other bacteria in the specimen, by using chemical, physical, or biological methods that favor the growth or survival, or effect the physical separation of the target organism. Physical enrichment methods may make use of conditions such as growth temperature, heat treatment, or size characteristics of specific organisms. Chemical enrichment methods may employ antibiotics or specific toxic agents to suppress the growth of unwanted bacteria, while allowing the growth of the target organism. Selective agents include antibiotics, dyes, detergents, and various organic and inorganic chemicals.

Differential Media

Once the numbers of the target organism are increased during various enrichment steps, isolation is attempted using differential plating media, often incorporating selective agents. Differential media allow

the target microorganism to be distinguished from other microorganisms based on a recognizable reaction produced by the growth of the target microorganism. There are many media that include pH indicators to allow for a visual detection of changes in pH as a result of the metabolic reactions that take place as the bacteria grow on the media. Commonly used indicators include bromophenol blue (red in acid; blue in alkaline), methyl red (red in acid; yellow in alkaline), bromocresol green (yellow in acid; blue in alkaline), and phenolphthalein (colorless in acid; red in alkaline). Bacterial colonies that exhibit the expected reaction are streaked on nonselective media to ascertain purity. Pure cultures are then characterized by genotypic or phenotypic methods to obtain a positive identification.

MPN

The MPN technique involves making replicate dilutions of a sample in an appropriate liquid growth medium and incubating for growth. The sample is diluted until aliquots are estimated to contain one viable cell each. Some aliquots will contain a single viable cell, whereas others will not contain a viable cell. Theoretically, a single cell will grow and cause turbidity or will be recorded via a biochemical reaction giving recognizable changes in the medium. The tubes showing no growth have presumably not received a single cell capable of growth. By counting the number of positive and negative tubes at each dilution, and referring to statistical tables, the MPN can be determined. The MPN technique is useful when the organism cannot be grown on solid medium, if it is easily overgrown by other bacteria, or if it produces a unique substance that can be assayed in broth. An example is determination of numbers of *Clostridium botulinum* in a sample by dilution, followed by detection of botulinum neurotoxin in the cultures using the mouse bioassay.

Direct Plate Counts

Direct plating on solid agar media allows isolation of single clones of bacterial cells, in addition to determination of numbers of bacteria in a sample. Dilutions of food samples are evenly spread on to solid growth media and incubated to allow formation of discrete colonies. Alternatively, dilutions of samples can be mixed with melted agar and poured into Petri plates. Such 'pour plates' give rise to colonies within the agar that can be counted to enumerate the bacterial cells in the original sample. As a rule, to obtain accurate results, the lower number of colonies per plate is often specified as 20 (lower numbers give less statistically significant results), whereas the upper limit is usually 200 colonies per plate (higher numbers are difficult to count, and colonies may coalesce).

Physiological Tests

The presence of a particular microorganism may be indicated by detection of a metabolic product or toxin unique to that organism. Proteolytic strains of *Clostridium botulinum* are phenotypically identical to *C. sporogenes*, with the exception that the former produced botulinum neurotoxin. Detection of botulinum neurotoxin with the mouse bioassay is an example of a physiological test to indicate the presence of a specific microorganism.

Dye Reduction

Dye-reduction assays use dyes to estimate the number of viable cells that are present in certain food products. The most commonly used dyes are methylene blue and resazurin. Supernatants of food products are prepared and are added to predetermined solutions of methylene blue or resazurin. Dye reduction time, inversely proportional to the number of cells in the sample, causes a color change from blue to white (methylene blue) or slate blue/pink to white (resazurin). In foods that do not contain many inherent reductive compounds, the dye reduction assay is comparable with the standard plate count method. The most common food analyzed is milk.

Bioluminescence

Bioluminescence-based methods for the detection of bacteria can be divided into two: adenosine triphosphate (ATP) measurement, and *lux* gene technology. Extraction and measurement of bacterial ATP can be indicative of the number of organisms present in a sample. The firefly luciferin–luciferase system is the most commonly used method. In the presence of ATP, the enzyme luciferase causes light emission, which can be readily measured using a luminometer. The amount of light emitted is directly proportional to the amount of ATP in the sample. Specific inhibitors are often used to destroy nonbacterial sources of ATP in food samples. The assay is attractive in determination of microbial load owing to its rapidity in obtaining results (usually within 10 min).

Some marine bacteria, such as *Vibrio fischeri* and *V. harveyi*, are able to luminesce, and the genes encoding this capacity have been characterized. Coding for the bacterial luciferase, the *luxA* and *luxB* genes have been inserted in species-specific bacteriophages. The bacteriophages are not metabolically active and do not emit light. However, when the bacteriophages attach themselves to their target (for example, a *Salmonella typhimurium*-specific bacteriophage would only attach itself to *S. typhymurium* if it were present

in the sample), the phage genetic material becomes inserted into the bacterial cell, where it can replicate and cause the host cell to luminesce. The reaction between luciferase and a long-chain aliphatic aldehyde substrate (such as dodecanal) provided exogenously in the sample results in the production of light.

Immunoassays

Agglutination

Agglutination reactions involve particulate antigens capable of binding antibody molecules. Since antibody molecules are multivalent, suspended particulate antigens form large clumps or aggregates, easily visible without magnification, when exposed to specific antibodies. Antibodies that cause this reaction are referred to as agglutinins. Agglutination assays can be used to determine concentrations of specific antibodies in a patient's immune sera. A constant amount of a suspended particulate antigen is added to a series of tubes containing a twofold dilution of patient's immune serum, and the titer of antibody in the serum is the reciprocal of the highest serum dilution showing agglutination of the particulate antigen. Agglutination reactions are routinely used for identification and serotyping of a wide range of bacterial foodborne pathogens.

Passive Agglutination

Whereas agglutination allows the use of particulate antigens to determine concentrations of antibodies in sera, passive Agglutination allows the determination of the presence and concentration of soluble antigens. In passive particle (indirect) agglutination, soluble antigens are coupled to large particles such as erythrocytes or latex spheres. When specific antibodies are added to coated erythrocytes or latex spheres, antibody bridges are formed between the particles, and agglutination occurs. By addition of a constant amount of coated particles to various dilutions of antisera, the titer of antibodies to the antigen used to coat the particles can be determined.

Alternatively, to determine the concentration of a soluble antigen, particles can be coated with antibody specific to the soluble antigen and added to dilutions of sample containing the soluble antigen. Agglutination of the antibody-coated particles indicates the presence of the soluble antigen.

Enzyme-linked immunosorbent assay (ELISA)

ELISA uses antibodies linked to enzymes to detect nanogram to picogram amounts of antigen. ELISAs rely on the fact that antigens or antibodies can be bound to a solid support, and antibodies can be coupled to enzymes without the enzyme losing activity or the antibody losing binding activity. Enzymes coupled to antibodies include alkaline phosphatase, horse-radish peroxidase, and β-galactosidase. Bound antibody–enzyme conjugates are detected by the formation of a colored reaction product produced by the enzyme. Two basic ELISA methods are used: the double antibody sandwich ELISA (or direct ELISA, Figure 1), for detection of soluble antigen, and the indirect immunosorbent ELISA, used for measuring presence and concentration of antibody in sera.

The double antibody sandwich, or direct ELISA, employs a capture antibody immobilized to a microtiter plate. The capture antibody 'captures' specific soluble antigens and immobilizes them on to the microtiter plate. A reporter antibody binds to the captured antigen to complete the 'sandwich.' This reporter antibody is conjugated to an enzyme (e.g., alkaline phosphatase or peroxidase), allowing detection of the reporter antibody. If the specific soluble antigen is not present, the sandwich of the capture antibody–antigen–reporter antibody does not form, and the reporter antibody is not available to produce a colored reaction product.

Immunofluorescence

Antibodies can also be coupled to fluorescent dyes, allowing detection of bound antibodies by fluorescence. Immunofluorescence can be used as a rapid procedure to detect a specific agent in a specimen containing a mixture of microorganisms. In the

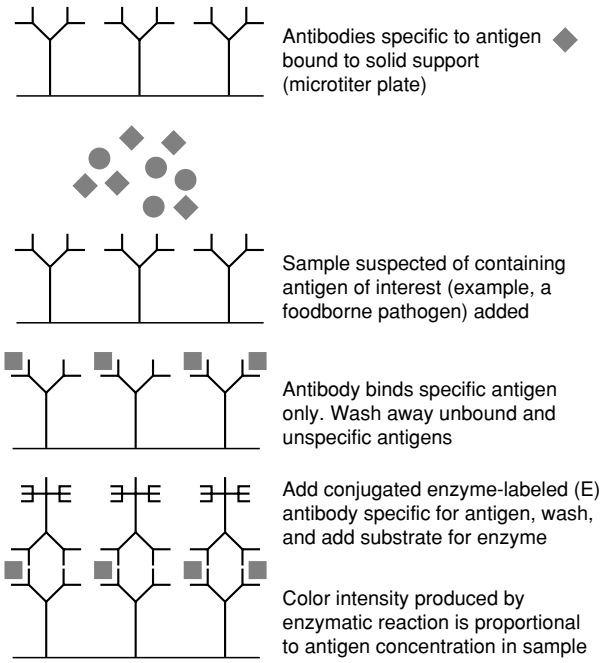

Figure 1 Direct ELISA assay.

direct fluorescent antibody technique, a smear of the sample is incubated with a solution containing specific antibody that is directly labeled with a fluorescent dye. After allowing the fluorescent antibodies to bind to the microorganisms, unbound fluorescent antibodies are removed by washing, and fluorescent bacteria are observed by epifluorescence light microscopy. The indirect fluorescent antibody technique is similar, but the smear is first incubated with unlabeled antibody. Bound antibody is then detected by incubation with a second fluorescent-labeled antibody against the immunoglobulin of the animal species used for the preparation of the initial antibody.

Immunomagnetic Separation

Immunomagnetic separation involves coupling of biological macromolecules, such as specific antibodies, to superparamagnetic iron oxide (Fe_3O_4) particles. Superparamagnetic particles exhibit magnetic properties when placed within a magnetic field but have no residual magnetism when removed from the magnetic field. This technology has been incorporated to make uniform porous polystyrene spheres, approximately 2–5 μm in diameter, with an even dispersion of magnetic Fe_3O_4 throughout the bead. These magnetic beads are coated with a thin polystyrene shell that encases the magnetic material and provides a defined chemical surface area for the adsorption of coupling of molecules such as antibodies and streptavidin.

The magnetic particles are added to a heterogeneous suspension to bind to the desired target (bacterial cells, viruses, proteins, nucleic acids, etc.) and form a complex composed of the magnetic particle and target. A magnet is used to immobilize the magnetic particles complexed with the target against the vessel wall, and the remainder of the material is removed. Washing steps to remove food material and other microorganisms are easily performed while the particle–target complex is retained. The target then can be detected using conventional immunoassays, pipetting or streaking on to agar culture media, or by using nucleic acid-based methodologies, as the magnetic particles do not interfere with other methods of detection.

Nucleic Acid-based Detection and Typing

DNA Probes

Under the proper conditions, a single-stranded DNA molecule will hybridize selectively with a complementary sequence of DNA, forming double-stranded DNA. If the reformed double helix is composed of DNA strands from different sources it is referred to as a 'hybrid.' The use of a labeled single-stranded DNA molecule (i.e., a DNA probe) allows detection of specific, complementary, nucleic acid sequences. DNA probes can be designed at varying levels of specificity to detect to species level, or beyond species level to particular pathogenic strains. Hybridization is a very specific method and has been used to detect several specific foodborne pathogens.

For hybridization to occur between a DNA probe and the target sequence, the double-stranded target DNA must be denatured into two separate strands by an increase in temperature. When the temperature is lowered, the strands will reform a double helix if strands contain similar sequences. The temperature and salt concentration used for hybridization are critical. If the temperature is too low, hybrids can be formed by strands that are not exactly complementary (low stringency). If the temperature is too high, strands that are exactly complementary may not hybridize, resulting in a negative reaction. The salt concentration is also adjusted to increase or decrease the stringency of the hybridization. Together, temperature and salt concentrations are used to optimize the hybridization conditions so that only the probe and a filter-bound nucleic acid that is highly homologous to that probe will bind to each other.

PCR and Other Amplification-based Methods

PCR (Figure 2a) is used for *in vitro* amplification of specific sequences of DNA. Specific sequences can be detected by amplification up to several million times. The amplified DNA can be visualized by electrophoresis on an agarose gel, followed by staining with ethidium bromide. The PCR reaction requires two oligonucleotide primers, complementary to sequences on opposite DNA strands. The two strands of the target DNA are separated by an increase in temperature (94 °C), and the primers are allowed to anneal to the complementary sequences in the denatured target DNA at a reduced temperature (50–70 °C). A thermally stable DNA polymerase from the thermophilic bacterium *Thermus aquaticus* (*Taq* polymerase, although other polymerases may be used) is used to synthesize the second strands. Two double-stranded DNA molecules have now been created from a single double-stranded molecule. The process is repeated, with each amplification step doubling the number of strands of target DNA.

Immuno-PCR combines the powerful amplification of PCR with antibodies to detect specifically low levels of antigens. This method is similar to an ELISA, but the reporter antibody is linked to a DNA fragment that can be amplified by PCR. Immuno-PCR has been shown to be 10^2–10^5-fold

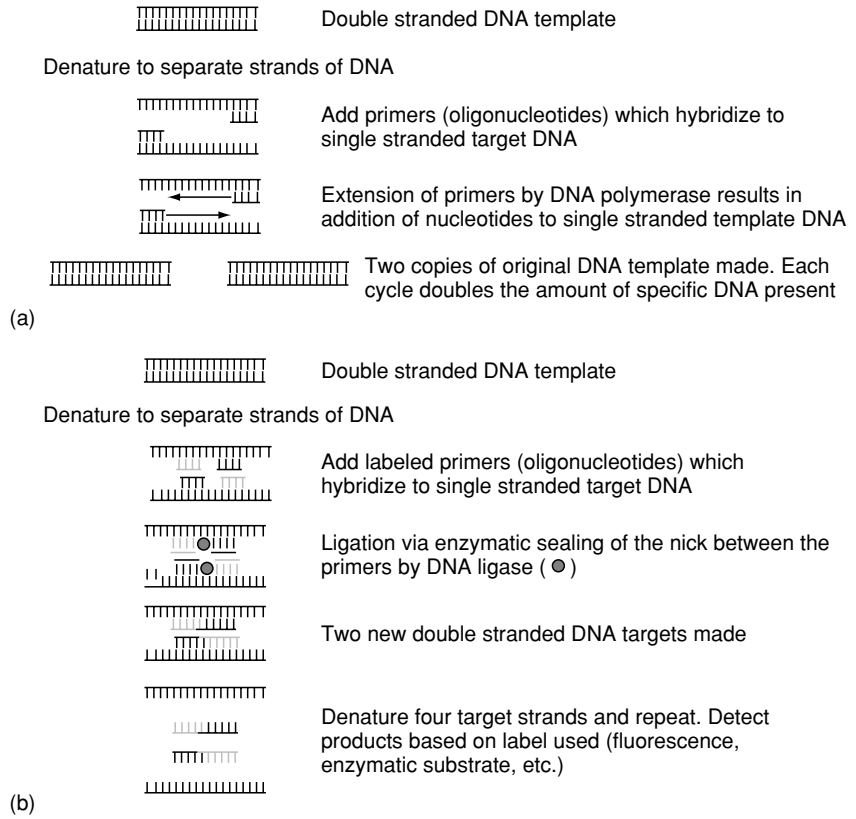

Figure 2 Diagrammatic differences between (a) PCR and (b) LCR. PCR generates a new target via synthesis of new DNA, whereas LCR joins primers that become the new target DNA molecules.

more sensitive than standard ELISA methods and has been used to detect antigens such as hormones, tumor markers, botulinum neurotoxin, and viral antigens.

Several techniques that employ PCR-based DNA fingerprinting of microorganisms have been developed in the last decade. PCR-based DNA fingerprinting is based on binding of primers to regions of DNA followed by the generation of species or strain-specific PCR amplification products. Randomly amplified polymorphic DNA assay (RAPD), also referred to as arbitrarily primed PCR, employs a single short (typically 10 base pairs) primer that is not targeted to any specific bacterial DNA sequence. At low annealing temperatures, the primer hybridizes at multiple random chromosomal locations and initiates DNA synthesis. Owing to its arbitrary nature, RAPD-PCR is susceptible to technical variation which may cause problems in reproducibility.

The ligase chain reaction (LCR, Figure 2b) uses two contiguous oligonucleotides that are joined by DNA ligase upon perfect hybridization to their target. The ligated probes are then amplified via thermal cycling with complementary oligonucleotides. LCR is better suited than PCR for diagnostic purposes, although there still remain few data comparing the two techniques in clinical settings. LCR makes use of known sequences for the sole purpose of detection, and this aspect makes it attractive for detection of foodborne pathogens. PCR, however, generates new DNA or RNA molecules that may require further verification or characterization (such as RFLP or direct sequencing). Although no report exists describing LCR as a detection system for foodborne pathogens, a major advantage would be the high specificity of the assay. Automation of any amplification-based method makes such technologies attractive for diagnostic laboratories.

Nucleic acid sequence-based amplification, or NASBA, is a technique that amplifies RNA and DNA targets as antisense, single-stranded RNA by the concurrent activity of reverse transcriptase, RNase H, T7 RNA polymerase, and two primers. NASBA is commonly used to determine HIV load in clinical samples. However, the technology should be applicable to food microbiology. Recently, a method for detection of viable oocysts of *Cryptosporidium parvum* in environmental samples using the NASBA technique was described. Another report described the use of NASBA to detect *Escherichia coli* isolates, which included enterohemorrhagic serotype

0157:H7. A 140-fg amount of total RNA could be amplified reproducibly to give a clear, detectable signal. In addition, a single *E. coli* cell in a sample could be detected in less than 14 h.

Pulsed-field Gel Electrophoresis (PFGE)

When bacterial DNA is digested with restriction enzymes that cut at a small number of sites within the genome (usually five to 20), high-molecular-weight fragments of DNA are obtained. These large fragments of DNA can be separated in a pulsed electric field, where the direction of the electric field is alternated between forward and reverse directions, with the forward pulse being longer in duration than the reverse pulse. Alternatively, contour-clamped homogeneous electric field gel electrophoresis, where a hexagonal arrangement of six electrodes are used to generate uniform electric fields at 120° to each other, can be used to separate DNA molecules as large as 3 million bases. When the agarose gel is stained with ethidium bromide, a restriction map is obtained with the pattern specific to individual bacterial isolates.

PFGE is highly discriminatory, capable of distinguishing between isolates within the same species or serotype. The major limitations of PFGE include the requirements for technical skill, sophisticated equipment, and the long duration until completion of the analysis. Not all strains of bacteria can be typed using PFGE because of degradation of DNA as a result of extracellular DNase production or the resistance of cell walls of some bacteria to lysis. Equally, some organisms may not yield sufficient discriminatory information using PFGE.

Ribotyping

Ribotyping makes use of ribosomal RNA gene restriction pattern analysis to discriminate between bacterial isolates. Ribosomal RNA (rRNA) is present in ribosomes of all bacterial cells and is composed of molecules of three different sizes: 23S, 16S, and 5S. The DNA encoding for rRNA is highly conserved among closely related bacteria, making ribotying less discriminatory than other DNA-typing procedures. Ribotyping involves isolation of total bacterial DNA followed by digestion of the DNA with specific restriction enzymes. The digested DNA is separated by electrophoresis on an agarose gel and transferred on to a nylon or nitrocellulose membrane. The DNA fragments on the membrane are then probed with labeled DNA fragments complementary to the rRNA gene sequences. After probing, each fragment of bacterial DNA containing a ribosomal RNA gene will be highlighted, creating a fingerprint pattern. An automated ribotyping system is available from DuPont Qualicon that allows identification of bacterial isolates, beyond the species level, in approximately 8 h.

DNA Sequencing

Automated DNA sequencing procedures now allow rapid sequencing of extended DNA sequences. This has facilitated sequencing of entire genomes, including the human genome. Perhaps the ultimate bacterial identification procedure of the future will be genome sequencing. Certainly, current technologies allow for rapid amplification and sequencing of the genes, such as the 16S rRNA of bacterial isolates. Because 16S rRNA genes are conserved among isolates of the same species, yet vary between species, phylogenetic trees describing their evolutionary relationships have been described. Indeed, the universal phylogenetic tree describing the phylogeny of the living world has been based on 16S or 18S rRNA comparative sequence analysis.

See also: ***Clostridium***: Occurrence of *Clostridium botulinum*; Botulism; **Immunoassays**: Principles; Radioimmunoassay and Enzyme Immunoassay; **Microbiology**: Classification of Microorganisms; **Microscopy**: Light Microscopy and Histochemical Methods; **Nucleic Acids**: Properties and Determination; Physiology; **Vibrios**: *Vibrio cholerae*; *Vibrio parahaemolyticus*; *Vibrio vulnificus*

Further Reading

Barbour WM and Tice G (1997) Genetic and immunologic techniques for detecting foodborne pathogens and toxins. In: Dolye MP, Beuchat LR and Montville TJ (eds) *Food Microbiology. Fundamentals and Frontiers*. Washington, DC: American Society for Microbiology.

Farber JM (1996) An introduction to the hows and whys of molecular typing. *Journal of Food Protection* 59: 1091–1101.

Hill WE and Jinneman KC (2000) Principles and applications of genetic techniques for detection, identification, and subtyping of food-associated pathogenic microorganisms. In: Lund BM, Baird-Parker TC and Gould GW (eds) *The Microbiological Safety and Quality of Food*. Gaithersburg, MD: Aspen.

Holbrook R (2000) Detection of microorganisms in foods – Principles of culture methods. In: Lund BM, Baird-Parker TC and Gould GW (eds) *The Microbiological Safety and Quality of Food*. Gaithersburg, MD: Aspen.

Koch AL (1981) Growth measurement. In: Gerhardt P, Murray RGE, Costilow RN *et al.* (eds) *Manual of Methods for General Bacteriology*. Washington, DC: American Society for Microbiology.

Krieg NR (1981) Enrichment and isolation. In: Gerhardt P, Murray RGE, Costilow RN *et al.* (eds) *Manual of Methods for General Bacteriology.* Washington, DC: American Society for Microbiology.

Schweitzer B and Kingsmore S (2001) Combining nucleic acid amplification and detection. *Current Opinions in Biotechnology* 12: 21–27.

Sharpe AN (2000) Detection of microorganisms in foods: Principles of physical methods for separation and associated chemical and enzymological methods of detection. In: Lund BM, Baird-Parker TC and Gould GW (eds) *The Microbiological Safety and Quality of Food.* Gaithersburg, MD: Aspen.

MICROCAPSULES

C Thies, Washington University, St. Louis, MO, USA

Background

Microencapsulation is the coating of small solid particles, liquid droplets, or gas bubbles with a thin film of coating or shell material. Although no officially approved definition of a microcapsule exists, most workers use the term microcapsule to describe particles with diameters between 1 and 1000 μm that contain a desired ingredient of some sort. Particles smaller than 1 μm are called nanoparticles; particles greater than 1000 μm can be called microgranules or macrocapsules.

Many terms have been used to describe the contents of a microcapsule: active agent, actives, core material, fill, internal phase, nucleus, and payload. Many terms have also been used to describe the material from which capsules are formed: carrier, coating, membrane, shell, or wall. For the sake of consistency, in this article, the material being encapsulated is called the core material, and the material from which the capsule is formed is called the shell material.

Several reviews discuss the general features of microencapsulation technology, including encapsulation techniques and applications not discussed here. A number of authors have reviewed the preparation of microcapsules that contain food components and the application of microcapsules to food products.

An understanding of microencapsulation technology and the potential contribution that microcapsules can make to food products rests on a knowledge of mass transport phenomena, properties of coating materials, and an understanding of processes by which small particles are produced. Since the primary purpose of microencapsulation is to control in some manner mass transport behavior, the shell of a microcapsule must control diffusion of material either from a microcapsule or into a microcapsule (**Figure 1**). The shell can provide protection of sensitive food components such as flavors, vitamins, or salts from oxygen, water, light and heat, convert difficult-to-handle liquids into free-flowing powders readily incorporated into various foods or isolate specific food components during storage.

In most cases, food components concentrated inside microcapsules are released as the food is consumed or during a food preparation step. Release is achieved by destroying the integrity of the microcapsule shell. This is done by dissolving it in water, melting it, or mechanically rupturing it. There are cases where no release of core material is desired until after the food has been ingested and is present in the digestive system. Perhaps, the intent is for the core material carried by a microcapsule never to be released. In these latter situations, the capsule shell must remain intact throughout the food preparation and ingestion steps.

In order to develop microcapsules with shells that function as intended, it is essential to understand the fundamentals of mass transport through shell materials from which microcapsules are formed, especially under use conditions. In many cases, the shell is thin, perhaps having a thickness of a few micrometers or less. Thus, the morphology or structure of a capsule shell has a significant impact on microcapsule performance. Overall, capsule morphology also affects capsule performance. The geometry of a microcapsule sample can vary significantly and often is a function of the process by which the microcapsule was formed. **Figure 2** illustrates two typical capsule structures, with **Figure 2a** representing a continuous core/shell capsule in which a single continuous shell of uniform thickness surrounds a continuous spherical region of core material, and **Figure 2b** representing a multinuclear capsule in which a number of small domains of core material are distributed uniformly throughout a matrix of shell material.

During the development of microcapsules for a specific food application, it is tacitly assumed that a food-grade microcapsule shell material with suitable

barrier and fabrication properties is available at an economically viable price. Proper selection of such materials requires an appreciation of the properties of candidate coating materials and requires an understanding of materials science. Significantly, few, if any, candidate capsule shell materials are able to provide all desired functions economically, so compromises must consistently be made when selecting specific shell materials for a given application.

Table 1 lists representative examples of capsule shell materials currently used to produce microcapsules for commerical food products, and the chemical class to which the shell material belongs, the encapsulation process typically used to produce microcapsules with each shell material and frequent food applications. Although gelatin–gum arabic complex coacervate capsules treated with glutaraldehyde are approved for limited consumption of selected food flavors, they are not approved for general food use (see Figure 3). Shell material costs vary greatly. As expected, the food industry favors the cheapest acceptable shell materials that are capable of providing desired performance, are available in commercial quantities, and are approved by the FDA.

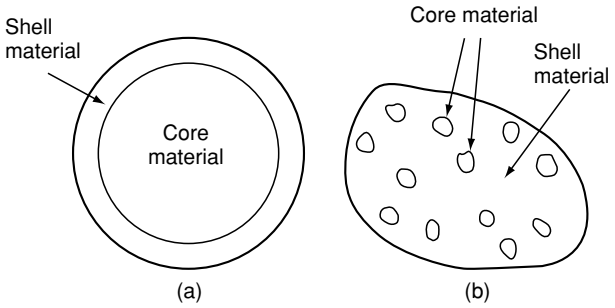

Figure 1 Two structures characteristic of many commercial microcapsules: (a) continuous core/shell structure; (b) multinuclear.

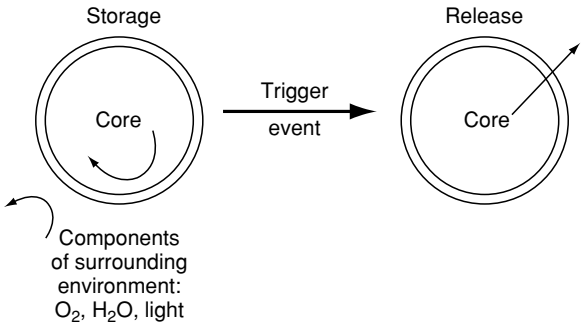

Figure 2 Ideal storage and release behavior properties desired for many capsules used by the food industry. Courtesy Thies Technology.

Microencapsulation Processes

General Comments

Many microencapsulation processes exist. Some are based exclusively on physical phenomena. Some utilize polymerization reactions to produce a capsule shell. Others combine physical and chemical phenomena. Because there are so many encapsulation techniques, it is logical to make an effort to attempt to categorize or classify them in some manner, thereby providing a means of identifying the concepts on which various encapsulation technologies are based. Many authors do this by identifying encapsulation processes as either chemical or mechanical processes. This author prefers to classify them as Type A or Type B processes, since so-called mechanical or physical processes actually may involve a chemical reaction, and so-called chemical processes may rely exclusively on physical phenomena. Table 2 lists representative examples of Type A and B processes. Type A processes are processes in which microcapsules are immersed in a liquid-filled stirred tank or tubular reactor throughout the encapsulation procedure. In a Type B process, a gas phase is involved at some stage of the encapsulation process. Microcapsules are formed by spraying droplets of coating material on a core material being encapsulated, solidifying liquid droplets sprayed or ejected into a gas phase, gelling droplets sprayed or ejected into a liquid bath, or

Table 1 Shell materials used to produce commercially significant microcapsules

Shell material	*Chemical class*	*Encapsulation process*	*Applications*
Gum arabic	Polysaccharide	Spray drying	Food flavors
Derivatized starch	Polysaccharide	Spray drying	Food flavors
Gelatin	Protein	Spray drying	Vitamins
Whey protein	Protein	Spray drying	Fats
Maltodextrins	Low-molecular-weight carbohydrates	Spray drying and desolvation	Food flavors
Hydrogenated vegetable oils	Glycerides	Fluidized bed	Assorted food ingredients
Complex coacervation	Protein–polysaccharide complex	Complex coacervation	Assorted flavors

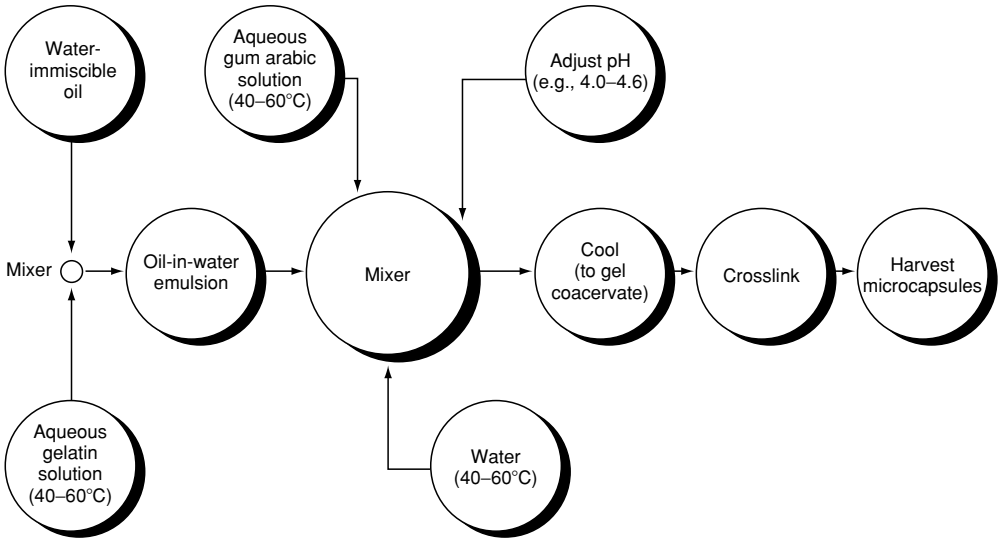

Figure 3 Sequence of steps of a complex coacervation encapsulation process. Courtesy Thies Technology.

Table 2 Representative examples of two broad classes of encapsulation processes

Type A processes	*Type B processes*
Complex coacervation	Spray drying and spray chilling
Polymer/polymer incompatibility	Fluidized bed
Interfacial polymerization at liquid/liquid and solid/liquid interfaces	Interfacial polymerization at solid/gas of Liquid/gas interfaces
In-situ polymerization	Centrifugal extrusion
Solvent evaporation or in-liquid drying	Extrusion or spraying into a desolvation bath
Submerged nozzle extrusion	Rotational suspension separation (spinning disk)
	Melt spinning

by carrying out a polymerization reaction at solid/gas or liquid/gas interfaces.

No encapsulation process developed to date is able to produce the full range of capsules desired by potential capsule users. Some processes readily produce small, liquid-filled capsules, whereas others produce relatively large capsules with a solid core material, capsules with water-soluble shells, or capsules with water-insoluble shells. In order to provide insight into which process is most appropriate for a specific food application, features of Type A and B encapsulation processes used by the food industry are summarized in the following sections. Processes that produce capsules not approved for food applications are discussed elsewhere.

Type A Encapsulation Processes

Complex coacervation Complex coacervation is a phenomenon in which cationic and anionic water-soluble polymers interact in water to form a liquid, polymer-rich phase called a complex coacervate. This coacervate is used to form a microcapsule shell (see Figure 3). Gelatin is normally the cationic polymer used. A variety of natural and synthetic anionic water-soluble polymers interact with gelatin to form a complex coacervate, but the food industry appears to use gum arabic exclusively. When the complex coacervate forms, it is in equilibrium with a dilute solution called the supernatant. The supernatant acts as the continuous phase in which the complex coacervate is dispersed. If a water-insoluble core material such as a flavor or fat is dispersed in the system, each droplet or particle of dispersed core material is spontaneously coated with a thin film of coacervate provided that the coacervate adsorbs on the surface of the dispersed oil droplets. When this liquid film is gelled, capsules are formed. The 'wet' complex coacervate gel is a very rubbery shell that deforms extensively without rupturing. In order to increase the strength of the water-swollen shell and create a gel structure that is not thermally reversible, complex coacervate capsules shells are typically cross-linked with glutaraldehyde.

Complex coacervation encapsulation processes can encapsulate many water-immiscible oils such as vitamins, plant oils and flavors as well as water-insoluble solids. Such processes routinely produce single capsules 20–800 μm in diameter that contain 80–90 wt.% core material. Capsules outside these size and core

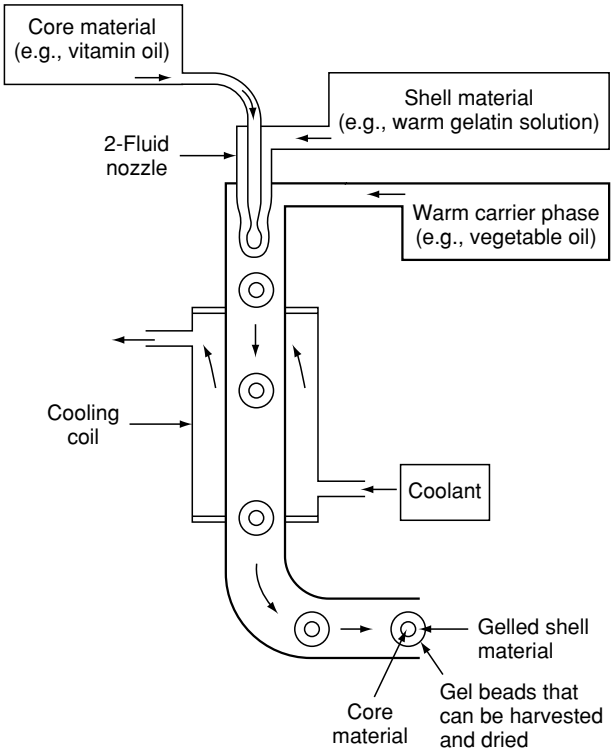

Figure 4 Stationary two-fluid nozzle used to produce capsules.

content ranges can be produced, although considerable experience may be required. Most coacervate capsules approach the continuous core/shell structure shown in Figure 2a, but the shell may not have a uniform thickness. Complex coacervation processes are adversely affected by active agents that have a finite water solubility, are surface active, or are detrimentally affected by pH conditions used to produce the microcapsules. The shell of dry complex coacervate capsules is sensitive to variations in atmospheric moisture content and becomes plasticized at a relative humidity of 70% or higher. Thus, the release properties of a complex coacervate capsule can be affected significantly by environmental conditions to which the capsules are exposed during storage.

Submerged centrifugal force and nozzle processes
Several Type A processes use submerged centrifugal force or nozzles to form microcapsules (see Figure 4). The nozzles can be rotating, stationary, or vibrating. They may be single or multifluid nozzles. An example of the former is the spinning cup encapsulation apparatus developed by the Atlantic Coast Fisheries Company. This consists of a spinning cup that contains a row of small holes. When the cup is immersed in a container filled with mineral oil or vegetable oil, an oil-in-water (o/w) emulsion fed into the interior of this spinning cup is extruded dropwise by centrifugal force through the openings into the oil phase that surrounds it. This creates a suspension of o/w emulsion droplets suspended in an oil phase. The continuous phase of these droplets is an aqueous solution of a polymer that gels on cooling (e.g., gelatin), whereas the dispersed phase is a water-immiscible liquid like a vitamin or fish oil. As the extruded droplets sink in the oil phase, the aqueous phase of the emulsion gels thereby creating gel beads throughout which an oil such as a vitamin or fish oil is dispersed. The beads are harvested from the oil phase by a solvent wash and dried. The dried beads have the structure shown in Figure 2b (see Figure 5).

Various submerged stationary and vibrating nozzles are used to produce microcapsules. A number of different nozzle configurations exist, and these produce capsules with a variety of structures. Sophisticated multifluid vibrating nozzles have been incorporated into capsule production units. Such units typically produce microcapsules by coextruding an aqueous polymer solution that gels on cooling and an oil to be encapsulated into a third or suspending fluid. The suspending fluid is a cooled oil phase in which the aqueous polymer solution gels, thereby forming gel beads with a continuous core/shell structure, as shown in Figure 2a. Such beads are isolated from the suspending fluid, washed in order to remove residual oil from the suspending medium, and subsequently dried.

To date, submerged Type A centrifugal force and various nozzle encapsulation processes are used primarily for the encapsulation of a variety of water-immiscible oils. The capsules tend to be larger than 250–500 μm but are typically formed from food-grade shell materials that can be gelled thermally.

Type B Encapsulation Processes

General comments Type B encapsulation processes utilize centrifugal force, extrusion, coextrusion, and spray technology in order to produce microcapsules. They predate Type A encapsulation processes, since spray drying encapsulation was developed in the 1930s. This class of encapsulation processes generally utilizes self-contained encapsulation units marketed by equipment manufacturers. In contrast, equipment for Type A encapsulation processes is typically custom-designed and manufactured for a specific capsule producer. It is the author's opinion that Type B processes often cannot produce microcapsules < 100 μm economically, whereas many Type A processes can. A notable exception is spray drying. Other exceptions undoubtedly exist, but there is clearly a question of how well many current Type B processes function when an effort is made to use them to produce economically viable amounts of small (e.g., < 100 μm) microcapsules.

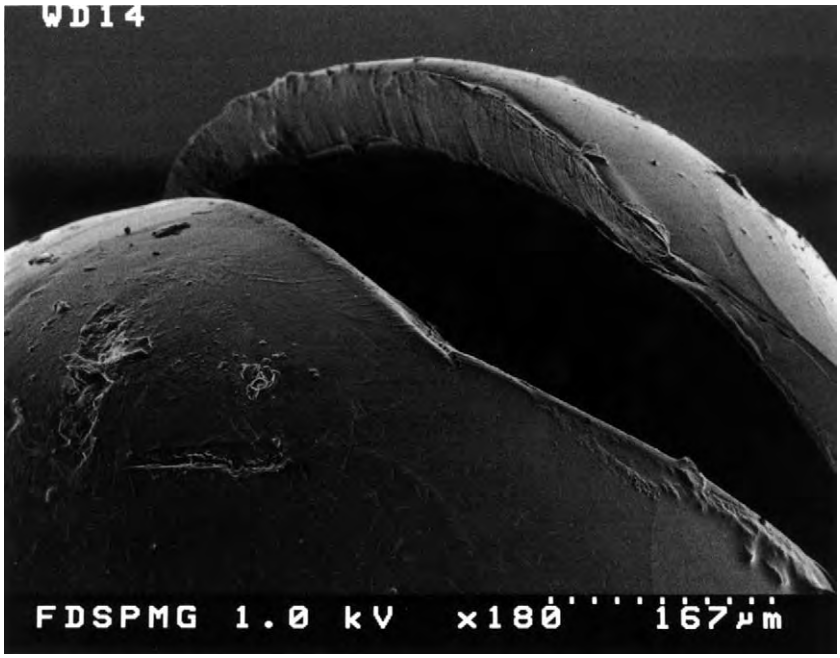

Figure 5 Scanning electron photomicrograph of a cut flavor capsule produced by a submerged nozzle encapsulation apparatus. Magnification: 180 ×. Courtesy Thies Technology.

Spray drying Spray drying is used to produce a variety of encapsulated food components. The first step is to emulsify the core material, typically a water-immiscible flavor, vitamin, fish or plant oil, in a concentrated (35–60 wt.%) aqueous solution of shell material until 1–3 µm oil droplets are obtained. The shell material normally is a water-soluble material like gum arabic, maltodextrin (hydrolyzed starch), modified starch, hydrolyzed gelatin, sugars, or various mixtures of these materials. All are approved for food use. They are water-soluble and not chemically cross-linked. Capsules prepared from them dissolve in water and release core material without leaving any residual capsule shell debris. Figure 6 illustrates the steps involved in the process. (*See* **Drying**: Spray Drying.)

Once a suitable core material/shell material dispersion has been prepared, the resulting emulsion is fed as droplets into the heated chamber of a spray drier. The droplets are rapidly dehydrated, thereby producing dry capsules that fall to the bottom of the drying chamber where they are harvested. Capsules produced in this manner are typically 10–300 µm in diameter. They may have an irregular geometry and be aggregates of a number of small particles. Spray-dried capsules contain a number of small droplets of core material dispersed throughout it, as illustrated in Figure 2b.

Spray-dry encapsulation has a number of advantages:

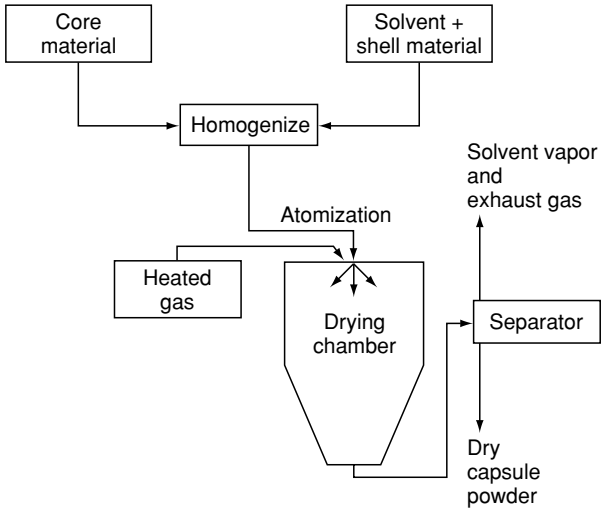

Figure 6 Steps involved in a spray-drying encapsulation process.

1. It is a well-established technology.
2. It utilizes readily available equipment.
3. It can produce large amounts of capsules economically.

It also has several limitations. For example, the list of candidate water-soluble shell materials is limited to candidate materials that form concentrated (30–60 wt.% solids) aqueous solutions of sufficient low viscosity that they can be pumped. Another limitation

is the 20–30% core loading carried by most spray-dried capsules. Spray-drying protocols that allow core loading up to 50–60% have been reported, but current spray-dried capsules carry a lower loading. A persistent problem with spray-dried capsules is free or surface core material. Because water evaporation from a capsule in the chamber of a spray drier occurs rapidly, it is not uncommon to harvest spray-dried capsules that have a significant amount of free or unencapsulated oil. This free oil exists because of discontinuities in the capsule shell, some of which form as a result rapid evaporation of water from the droplets being dried, thereby forming microcapsules. The higher the core loading, the more pronounced this problem can become. Free or surface oil is undesirable, since it is susceptible to oxidation and development of an off-odor or off-taste when the encapsulated core material is a fragrance or a flavor, respectively. It has been found that maximizing core loading and minimizing free core content involve a judicious choice of coating material, emulsifying agent, and spray-dryer operating conditions. Finally, it is important to stress that low-boiling-point compounds with a finite degree of water solubility have posed a persistent problem to spray-drying encapsulation. Such compounds volatilize from the capsules in the spray-drying chamber.

In summary, spray drying is a viable commercial method of forming microcapsules for the food industry. It is an established, comparatively low-cost encapsulation technology that continues to develop. To date, spray drying for food products has primarily been used to encapsulate flavors, various fish and plant oils as well as liquid vitamins.

Spray chilling Spray chilling, cooling, or congealing are variations of conventional spray-drying. In these cases, chilled air is used to solidify molten capsule shell material formulations rather than volatilize a solvent. Various fats, waxes, fatty alcohols, fatty acids, or combinations of these materials are the shell materials used. In such encapsulation procedures, the active agent is dispersed in a molten shell material with the aid of an emulsifier, if necessary. The resulting dispersion is atomized through heated nozzles into a cooling chamber analogous to that used in a spray drier. The shell material is solidified by cooling, thereby producing solid particles.

Particles produced by this method have water-insoluble shells. The processing temperature has an influence on shell material polymorphism, a phenomenon that is characteristic of many fats. If polymorphic changes in a fat coating occur on storage, such changes can have a significant negative effect on capsule storage and release properties.

Fluidized-bed coaters Fluidized-bed coaters function by suspending a bed or column of solid particles in a moving gas stream, usually air. Three types of fluidized beds are available: top-spray, tangential-spray, and bottom-spray (see Figure 7). These units differ in location of the nozzle or nozzles used to apply the coating formulation. Nevertheless, in each case, a liquid coating formulation is sprayed on to the individual suspended particles, and the freshly coated particles are cycled into a zone where the coating formulation is solidified either by solvent evaporation or by cooling. This coating and drying sequence may be repeated until a desired coating thickness has been applied. (*See* **Drying:** Fluidized-bed Drying.)

In a bottom- or tangential-spray unit, droplets of liquid coating material leaving a spray nozzle move

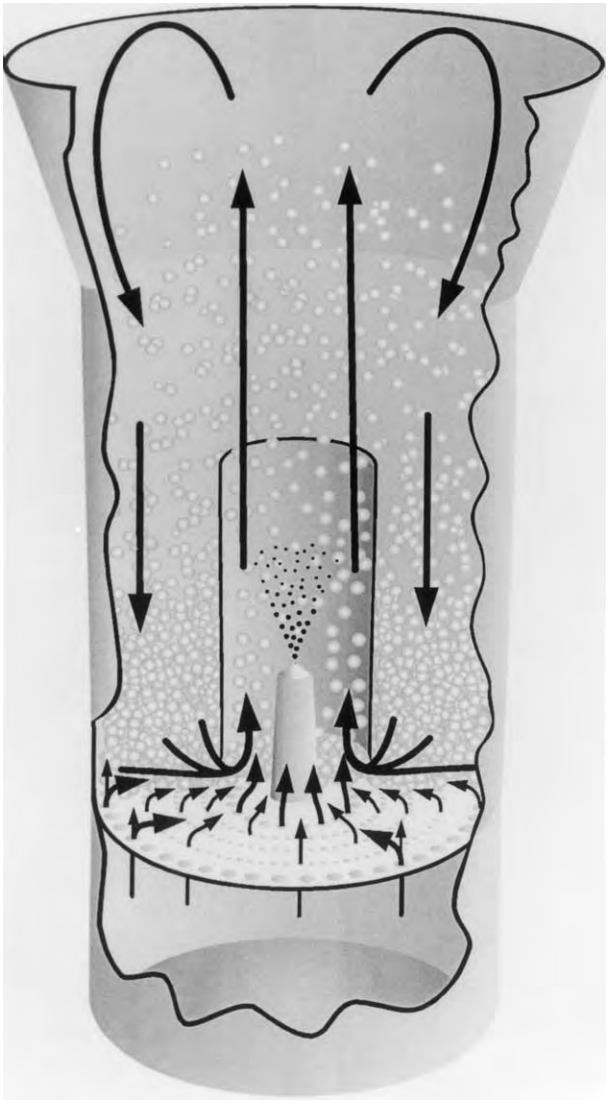

Figure 7 Bottom spray or Wurster fluidized-bed coating apparatus. Courtesy Coating Place, Verona, WI.

concurrent with the gas stream used to suspend the bed of solid particles being coated. In a top-spray unit, the coating formulation is sprayed down on and into the top of the fluidized bed. The droplets of spray leaving the nozzle move countercurrent to the gas stream until they impact the particles being coated. In the former case, a uniform film of coating material can be deposited when the coating formulation contains a volatile solvent. In the latter case, volatile solvent(s) in the coating formulation can evaporate from the spray droplets, thereby increasing the solids content, perhaps to such a degree that they cannot spread on to the particles being coated. This spray-drying effect has been used to explain why solvent-based coating formulations applied in top-spray fluidized bed coaters often yield coated particles with a degree of internal void volume and porous coatings. Enteric coatings applied as aqueous latex dispersions are an exception. Such coatings applied in a top-spray fluidized-bed coater form a continuous coating analogous to that obtained with tangential- and bottom-spray units. Molten fats also can be applied by top-spray units, since, in such cases, there is no solvent evaporation. Although top-spray units have limitations, they are claimed to be simpler to operate and have a higher production capacity than the other types of units.

Application of the capsule shell material in a series of steps, as is typically done in a fluidized-bed coating process, is both an advantage and a disadvantage. The advantage is that sequential deposition of the shell formulation in a number of different cycles covers up or heals defects found in any single application, thereby yielding relatively defect-free coatings. It is also possible to coat layers of different shell materials on a particle surface. The disadvantage is that it takes time to apply the capsule shell in a series of cyclical steps, and this increases the operating costs.

Fluidized-bed coaters can encapsulate only solid particles or perhaps porous particles into which a liquid has been absorbed. However, no other encapsulation technology can apply as broad a range of shell materials: hot melts, aqueous latex dispersions, organic solvent solutions, or aqueous solutions. Several commercial organizations use fluidized-bed coaters to coat various solid food ingredients with fats (see Figure 8). This appears to be the primary use of fluidized-bed coaters by the food industry.

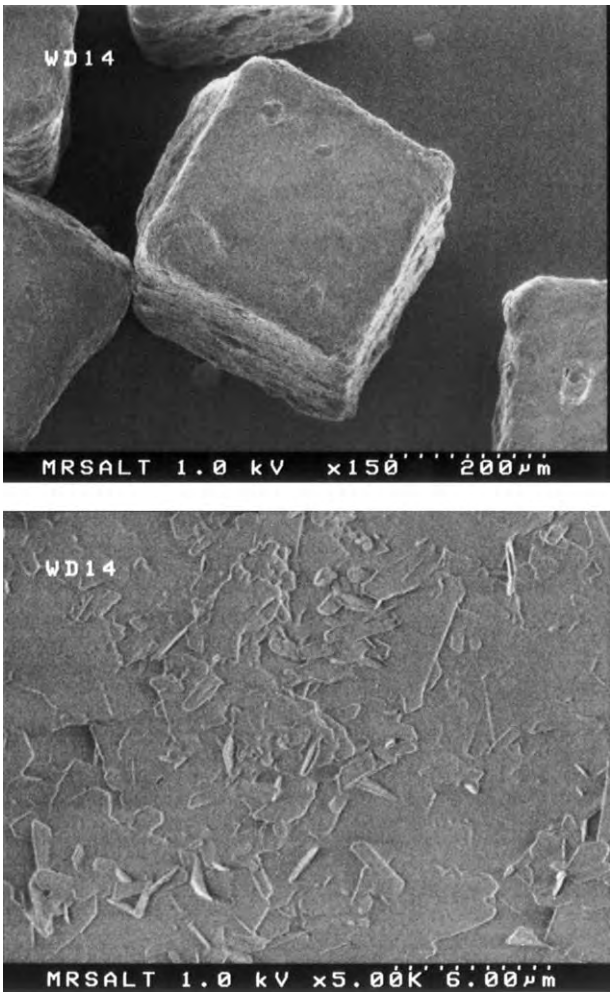

Figure 8 Scanning electron photomicrographs of a single NaCl crystal coated with a fat by a fluidized-bed coating process: top, 150 × magnification; bottom, 5000 × magnification. Courtesy Thies Technology.

Desolvation or liquid extraction encapsulation technology Desolvation encapsulation processes consist of dissolving a shell formulation in a finite amount of water, thereby forming a concentrated solution (e.g., 40 wt.%) of shell material. Core material is dispersed in this solution with the aid of a surfactant. The dispersion produced can be extruded or atomized directly into a desolvation bath (Figure 9) that solidifies the shell formulation by extracting the solvent used to dissolve the shell material from the spray droplets. Alternately, the core material/shell dispersion can be extruded or atomized into a vessel that contains excess spinning solvent. The spinning bath produces droplets of core/shell formulation that are subsequently solidified in a desolvation bath. Capsules with a relatively low core loading (e.g., 15–20%) are produced, although more recent patents indicate that a core loading up to 40% is feasible.

The core material is generally a water-immiscible liquid such as a flavor. The shell material can be various combinations of carbohydrate(s) or protein(s). Spinning solvents (e.g., mineral or vegetable oil) are immiscible with the shell material solvent (water).

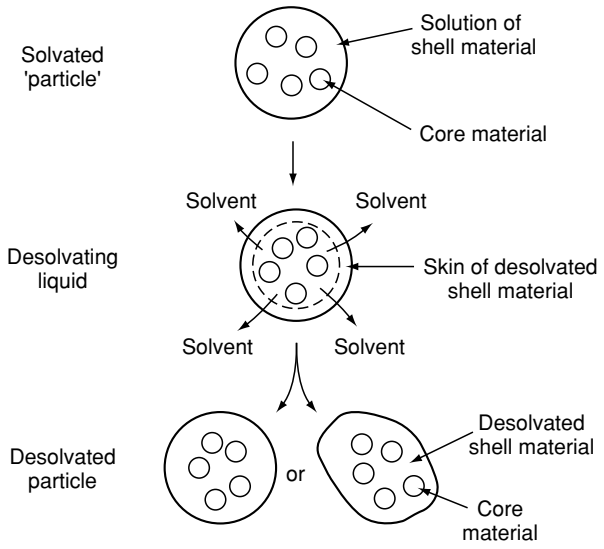

Figure 9 Sequence of steps that occur in a desolvation bath.

Water-miscible desolvation solvents (e.g., isopropanol, ethanol) are nonsolvents for the shell material (s). The desolvation solvent, often cooled, extracts water from the shell formulation, thereby producing solid particles or rods. Residual desolvation solvent is removed from solidified particles by drying. Rods or strands of solidified extrudate can be broken up into small particles by agitation in the extraction vessel or after isolation.

For this encapsulation technology to be successful, the desolvation solvent must be able to extract solvent (e.g., water) from the dispersed core/shell formulation with minimal simultaneous leaching of core material from the formulation. This occurs if the desolvation solvent rapidly solidifies the outer surface region of core/shell formulation droplets or rods to form an outer barrier that prevents diffusion of core material into the extraction solvent, but allows diffusion of water from the interior of the droplets to the extraction medium.

Desolvation encapsulation technology can produce water-soluble food-grade capsules loaded with a range of flavors. Typical shell materials include edible water-soluble materials like gum arabic, maltodextrins, or modified starches (i.e., the same materials used in spray drying). This technology can be carried out in an inert atmosphere and at reduced temperatures, thereby minimizing loss or degradation of sensitive components. It can incorporate water soluble materials in water-soluble shells and produce stable capsules free of unencapsulated core material, since the desolvating liquid extracts actives from defective capsules.

Problems with this technology include the use of solvents other than water that are more costly than heated air, are flammable, must be recycled, and must be essentially eliminated from the final product. Desolvating liquids also can break core/shell material emulsions, thereby preventing suitable particle formation. The formation of stable emulsions in concentrated aqueous shell solutions can pose serious problems, especially if they are heated.

Melt extrusion A melt extrusion encapsulation process (Figure 10) involves dispersing a core material in a molten shell formulation at 85–125 °C with or without the aid of a surfactant. Once formed, the dispersion is extruded as filaments into a relatively cool environment that solidifies the extruded mass. The receiving environment can be a gas phase or a tank that contains a suitable solvent. If extruded into a gas phase, the cooled mass is simply broken up into particles and used. If extruded into a solvent (e.g., −20 °C isopropanol), the solvent simultaneously cools and removes unencapsulated or free core material from the filaments. The solidified product is subsequently dried and broken up to yield particles with the multinuclear structure shown in Figure 11. The particles are glass matrices loaded with dispersed core material. Flavor-loaded particles produced in this manner have an excellent resistance to oxidation during storage. An anticaking agent like pyrogenic silica can be added to enhance particle flow properties. (*See* **Extrusion Cooking**: Principles and Practice.)

The patent literature discloses a number of melt extrusion methods and combinations of shell materials used to produce particles loaded with a core material. Claimed core loading in the final product ranges from 5 to 40% of final particle weight, although a loading of 5–30% is most common. Preferred core materials are flavoring oils. Both oil- and water-miscible flavors have been entrapped. The core/shell formulation typically has a temperature of $\leq 125\,^{\circ}C$ at the time of extrusion. An antioxidant

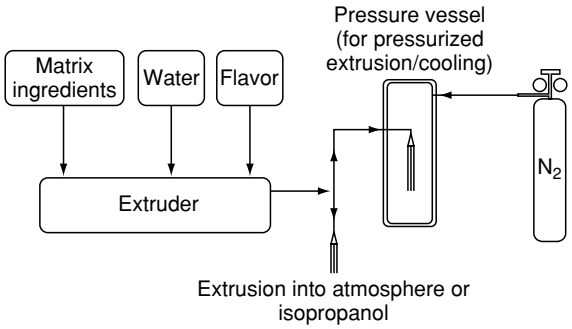

Figure 10 Melt extrusion encapsulation process. Courtesy Thies Technology.

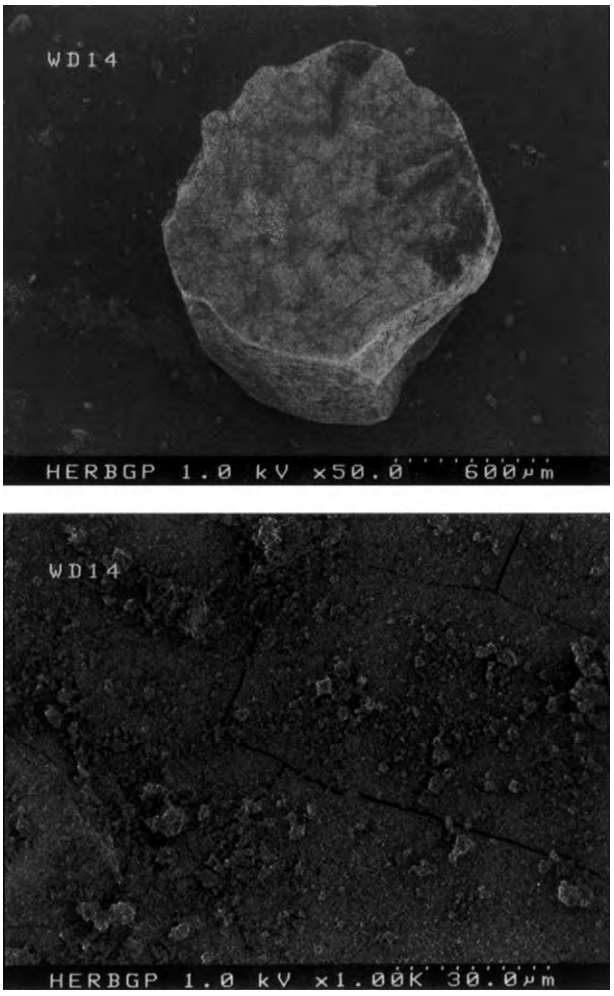

Figure 11 Scanning electron photomicrographs of the cross-sectional surface of a commercial flavor capsule produced by melt extrusion. Top, 50 × magnification; bottom, 1000 × magnification. Courtesy Thies Technology.

can be added to heat-sensitive oils. Preferred shell materials are various combinations of sugars, corn-syrup solids, maltodextrins, modified starches, and gum arabic, although other shell materials such as proteins like gluten have been used. In all cases, the water present in the system at the time of extrusion is limited to that required to reduce melt viscosity of the core/shell formulation to a point where extrusion occurs freely. This is usually no more than 10% of the mass of material extruded. Extrusion is carried out under pressure through spinnerets or screw extruders (single and double) with zoned heating. That is, different parts of the extruder screw and die-head are at different temperatures.

The shell formulation in the original extrusion process was corn-syrup solids containing 3–8.5% water. Improved properties were obtained by using a corn-syrup solids shell formulation plasticized with glycerol, and a further improvement in the process was achieved by using a mixture of sucrose and maltodextrins with a dextrose equivalent of less than 20 (preferably 10–15). This mixture produced particles with an improved stability towards oxygen and humidity variations.

Since the processing temperatures involved in melt extrusion are relatively high, loss of volatile components and oxidative degradation of heat-sensitive materials are problems that must be addressed. Accordingly, particle formation and solidification can be carried out in an extrusion system pressurized with an inert gas like nitrogen or carbon dioxide. The final particles are isolated under pressure in order to avoid loss of volatiles resulting from venting or the formation of so-called puffed particles with flaws that cause loss of volatile core material.

Suspended nozzles A number of workers have produced microcapsules by ejecting droplets that contain a core and shell material from suspended nozzles into a gas phase, usually air. The droplets pass through a gas phase until the capsule shell is solidified by cooling, or they fall into a curing bath where they are gelled and subsequently harvested. The particles produced can have a structure like that shown in Figure 2a or b. In all cases, the rate of droplet production and, hence, rate of capsule production is increased by increasing the number of nozzles, using centrifugal force, or a combination of these approaches.

Many types of nozzles have been described: single-fluid, two-fluid, or even three-fluid nozzles. Some are stationary, whereas others are vibrating or spinning. They can be single-fluid tubes through which an emulsion of core material in an aqueous shell solution flows by gravity extrusion or under an applied external pressure. Improved control of droplet, and ultimately capsule, size distribution is accomplished by vibrating the extrusion tubes or orifices. Single-fluid tubes usually produce capsules with a structure like that shown in Figure 2b, since the extruded droplets are dispersions of core material in a solution of shell material.

When droplets are formed by using multifluid nozzles, each droplet formed contains a continuous core region surrounded by a liquid shell or by placing the nozzles in a spinning head through which the droplets are ejected with the aid of centrifugal force. The nature of the shell material determines how ejected droplets are converted into capsules (See Figure 2b). If the shell material is a relatively low-viscosity hot melt that crystallizes rapidly on cooling (e.g., a wax or wax toughened with a polymer), the droplets are converted into solid particles as they fall

away from the nozzle and are cooled. Suitable core materials are typically polar liquids like water or aqueous solutions, since they are immiscible with a range of hot-melt shell materials like waxes.

Alternatively, droplets emerging from the spinning nozzle may have a shell that is an aqueous solution of a gelable polymer. In this case, the compound droplets fall into a gelling bath where they are converted into gel beads. A specific example is the gelation of an aqueous sodium alginate shell by an aqueous $CaCl_2$ gelling bath. The calcium alginate gel beads produced can be dried to give free-flow capsules. In one previously documented encapsulation apparatus, compound droplets produced by a vibrating two-fluid nozzle pass through an aqueous $CaCl_2$ (20%) mist, thereby causing a degree of prehardening before the droplets fall into the actual $CaCl_2$ hardening bath. Prehardening minimizes droplet breakage on impact with the surface of the curing bath.

An understanding of the physical phenomena involved in the formation of microcapsules by nozzle devices requires an understanding of the principles of fluid mechanics. Capsules prepared by centrifugal extrusion tend to be large, with diameters typically ranging from over 250 μm up to several millimeters. Capsules produced by stationary vibrating nozzles can be smaller (e.g., 100 μm in diameter) and very uniform in size. Atomization can produce much smaller capsules, although the particle size distribution can be large, and the capsules may not have a perfectly spherical geometry.

Spinning or rotating disc In spinning disc encapsulation processes, core material dispersed or emulsified in a liquid shell formulation is fed onto the center of a rotating disc. Centrifugal force generated by the rotating disc forces the dispersed core material and liquid coating formulation across the surface of the disc to the outer edge. There, the dispersion is thrown off the disc into a gas phase (usually air) as discrete droplets. If the core material is in particle form, the droplets are discrete particles covered with a thin film of liquid shell formulation. If the core material is an emulsion or a dispersion of particles in a shell formulation, droplets filled with dispersed core material are produced. If the shell is a hot-melt formulation, it cools and solidifies as the particles fall through the gas phase, thereby producing capsules that can be harvested and used. If the shell formulation is an aqueous polymer solution that can be gelled by ions or a combination of ions and cooling, the coated particles fall into a liquid-filled curing bath that gels the shell formulation. The particles with a gelled coating can be used as gel particles or, in principle, can be dried to a free-flowing powder.

Disc geometry, diameter, and speed of rotation, as well as volume flow rate of liquid across the disc are parameters that affect the size of the capsules produced. In the case of gel beads, surface tension of the aqueous phase being dispersed in droplet form affects gel bead geometry. In addition to capsules, one patented spinning disc process produces particles of shell material free of core material. In this process, the disc is operated in such a manner that the particles of shell material collect in a zone or region distinctly separated from the capsules and can be recycled. A variety of hot-melt shell materials can be applied to dispersed solid particles, but a melt viscosity below 5000 centipoises is favored. Capsule shell formulations that do not solidify rapidly pose problems.

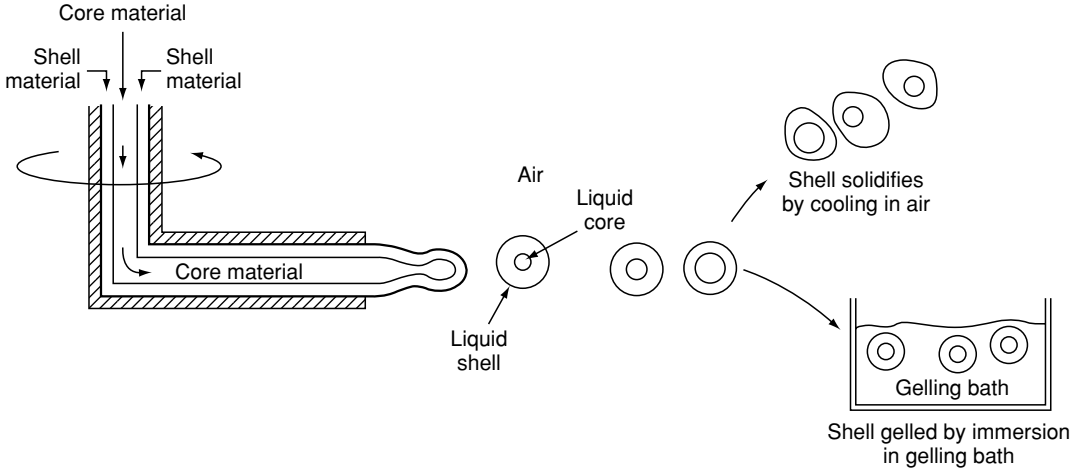

Figure 12 Spinning two-fluid drop-forming device used to produce microcapsules.

Spinning disc processes are claimed to be a fast and efficient way of producing large amounts of capsules economically. Claimed sizes of capsules produced in this manner range from below 100 μm to over 2 mm. Both solid and liquid core formulations have been encapsulated. In the case of solid core formulations, the core material must approach spherical geometry.

Microencapsulated Food Ingredients

A number of food ingredients and additives have been encapsulated and are available commercially. Microencapsulation holds much hidden potential and promise for the food industry in the future.

Solid ingredients encapsulated are typically water-soluble compounds. They are encapsulated with a hydrophobic or hydrophilic coating material usually applied by the Wurster process. Preferred hydrophobic coating materials are partially hydrogenated vegetable oils of varying melting points, monoglycerides, and diglycerides. Hydrogenated vegetable oils used include cottonseed, soybean, and palm. Hydrophilic coating materials tend to be maltodextrins and occasionally gum arabic. Both types of coating materials are well-accepted food-grade products.

Acidulants like citric and lactic acid encapsulated in partially hydrogenated vegetable oil are used in meat processing, where they provide direct acidification and shortened processing times. Sodium acid pyrophosphate encapsulated in hydrogenated vegetable oil is used in frozen cake batters in order to aid mixing and reduce gas release during batter make-up. In both types of applications, release of core material occurs during a heating cycle that melts the shell formulation and releases the core material.

Acidulants like citric, lactic, and fumaric acids encapsulated in a water-soluble maltodextrin shell formulation are used in dry-mix beverages and desserts as well as prepared premixes for the baking and dairy industries. The maltodextrin coating is designed to minimize hygroscopicity, reduce dusting, and minimize reactions with incompatible ingredients. It dissolves in the presence of liquid water to rapidly release the contents of the capsules during a mixing cycle.

Ferrous sulfate, and vitamin C (ascorbic acid) encapsulated in hydrogenated vegetable oil or maltodextrin are used to fortify a nutritional product of some sort. The capsules provide taste-masking, possibly a degree of prolonged release, stabilization of the core material against oxidation, and minimization of reaction with other ingredients in the final product.

Calcium proprionate, and sodium bicarbonate encapsulated in hydrogenated vegetable oil are used in chemically leavened products. Release typically occurs during the baking cycle as a result of melting of the hydrogenated shell material. Sodium bicarbonate encapsulated with maltodextrin is used in dry mix baking and other chemically leavened products. In this case, release occurs during the mixing step. Sodium chloride encapsulated in a hydrogenated vegetable shell is used in various meat products, yeast-containing mixes, and assorted types of dough. The capsules are designed to minimize inhibition of yeast activity, rancidity, and excessive salt binding during product storage.

Liquid food ingredients encapsulated are typically oil-soluble flavors, spices, vitamins, food oils, and fats. These core materials are often encapsulated with a water-soluble shell material applied by spray drying from water, but fat shell formulations are used occasionally. Preferred water-soluble shell materials are gum arabic, modified starch, or blends of these polymers with maltodextrins and sugars. Vitamins are encapsulated with zero bloom strength gelatin by spray drying.

A range of spray-dried flavor-filled capsules primarily with water-soluble shell formulations are used in various dry beverage mixes and other dry food products. Flavors containing ethyl acetate and other low boiling point components have been found to pose problems for successful spray-dry encapsulation. Such components are lost either during the initial emulsification process or during the actual dewatering step as a result of azeotrope formation. Another problem with spray-dry encapsulation is the formation of free surface oil. The rapid desolvation that occurs in the drying chamber can produce blow holes in the capsules, thereby essentially leaving a small amount of flavor oil of free or surface oil that oxidizes on storage and reduces product quality.

Melt extrusion under pressure is an alternate approach to food flavor encapsulation. Rod-like particles loaded with a flavor are often produced in this manner. An advantage of this technology is that it yields essentially defect-free particles with superb shelf-life storage stability. A commercial product (FlavorCell™) has been produced in this manner. Release data have revealed that the FlavorCell™ particles, when placed in ambient-temperature water, release their flavor contents as rapidly as spray-dried capsules that had the same shell formulation. Other approaches to the encapsulation of food flavors include complex coacervation and spray chilling.

Summary

The food industry is currently incorporating a number of different microcapsules into a range of food products. Nevertheless, the volume of capsules

being sold remains small relative to the huge potential that exists. Many groups continue to develop encapsulation technology and microcapsule-based products for the food industry. For this reason, it is anticipated that the volume of microcapsules sold for food use will experience a steady growth for the foreseeable future.

See also: **Chilled Storage**: Principles; **Drying**: Spray Drying

Further Reading

Achanta AS, Adusumilli PS, James KW and Rhodes CT (1997) Development of hot melt coating methods. *Drug Devel. Ind. Pharm* 23: 441–449.

Balassa LL (1970) Encapsulation of aromas and flavors, US Patent 3,495,988.

Balassa LL and Fanger GO (1971) Microencapsulation in the food industry. *CRC Critical Reviews in Food Technology* 2: 245–265.

Beck EE (1972) Essential oil composition and method of preparing the same, US Patent 3,704,137.

Bhandari BR and D'Arcy BR (1996) Microencapsulation of flavour compounds. *Food Australia* 48: 547.

Boccio JR, Zubillaga MB, Caro RA *et al*. (1996) Bioavailability and stability of microencapsulated ferrous sulfate in fluid milk: studies in mice. *Journal of Nutritional Science and Vitaminology* 42: 233–239.

DeZarn TJ (1995) Food ingredient encapsulation. In: Risch SJ and Reineccius GA (eds) *Encapsulation and Controlled Release of Food Ingredients, ACS Symposium Series 590*. Washington, DC: American Chemical Society.

Franjione J and Vasishtha N (1995) The art and science of microencapslation. *Technology Today*.

Gibbs BF, Kermasha S, Alli I and Mulligan CN (1999) Encapsulation in the food industry. *International Journal of Food Science and Nutrition* 50: 213–224.

Hagenbart S (1993) Encapsulated ingredients keep problems covered. Food Product Design 3: 28–38.

Hincal AA and Kas HS (1994) Preparation of micropellets by spray congealing. In: Ghebre-Sellassie I (ed.) *Multiparticulate Oral Drug Delivery*. New York: Marcel Dekker.

Jackson LS and Lee K (1991) Microencapsulated iron for food fortification. *Journal of Food Science* 56: 1047–1050.

Janovsky C (1993) Encapsulated ingredients for the baking industry. *Cereal Foods World* 38: 85–87.

Jones DM and Percel PJ (1994) Coating of multiparticulates using molten materials: Formulation and process considerations. In: Ghebre-Sellassie I (ed.) *Multiparticulate Oral Drug Delivery*. New York: Marcel Dekker.

Lehmann K (1992) Fluidized-bed spray coating. In: Doubrow M (ed.) *Microcapsules and Nanoparticles in Medicine and Pharmacy*, pp. 73–97. Boca Raton, FL: CRC Press.

Masters K (1991) *Spray Drying Handbook*, 5th edn. New York: Halstad Press.

Matsumoto S, Kobayashi H and Takashima Y (1986) Production of monodispersed capsules. *Journal of Microencapsulation* 3: 25–31.

Ogbonna JC, Matsumura M, Yamagata T, Sakuma H and Kataoka H (1989) Production of micro-gel beads by a rotating disk atomizer. *Journal of Fermentation and Bioengineering* 68: 40–48.

Popplewell LM, Black JM, Norris LM and Porzio M (1995) Encapsulation system for flavors and color. *Food Technology* 49: 76–78, 80, 82.

Reineccius G (1995) Controlled release techniques in the food industry. In: Risch SJ and Reineccius GA (eds) *Encapsulation and Controlled Release of Food Ingredients, ACS Symposium Series 590*. Washington, DC: American Chemical Society.

Schlameus W (1995) Centrifugal extrusion encapsulation. In: Risch SJ and Reineccius GA (eds) *Encapsulation and Controlled Release of Food Ingredients, ACS Symposium Series 590*. Washington, DC: American Chemical Society.

Shahidi F and Han X-Q (1993) Encapsulation of food ingredients. *Critical Reviews in Food Science and Nutrition* 33(6): 501–547.

Sparks RE, Jacobs IC and Mason NS (1999) Microencapsulation.In: Avis KE, Shukla AJ and Chang R-K (eds) *Pharmaceutical Unit Operations: Coating*, pp. 177–222. Buffalo Grove, IL: Interpharm Press.

Thies C (1995) *Microencapsulation, Kirk-Othmer Encyclopedia of Chemical Technology*, vol. 16, 4th edn, pp. 628–652. New York: John Wiley.

MICROFLORA OF THE INTESTINE

Contents
Role and Effects
Probiotics

Role and Effects

B R Goldin, Tufts University School of Medicine, Boston, USA

Description of the Microflora of the Normal Adult Gastrointestinal Tract

The gastrointestinal microflora of humans is a complex ecosystem. More than 400 bacterial species have been identified in feces from a single person. Ninety per cent of the total microflora is represented by 30–40 different species. Anaerobic bacteria are the predominant microorganisms in the gastrointestinal tract, outnumbering aerobes by a factor of 10^3. The distribution and composition by genera of the most prevalent microflora in the gastrointestinal tract are outlined in Table 1.

The saliva washes the bacteria in the oral cavity deposited on teeth, soft tissue surfaces, and food into the stomach where most of the bacteria are killed by the low pH of the gastric juice. Consequently, only 10–100 organisms are isolated per milliliter of gastric juice. The most commonly isolated bacteria in the stomach are Gram-positive facultative forms such as *Streptococcus* and *Lactobacillus*.

The bacterial population in the small intestine increases from the duodenum to the distal ileum. The bacterial composition of the proximal small bowel is similar to the stomach and the number of organisms per milliliter of contents increases from 10^2 to 10^4 in the distal jejunum. The bacterial counts and complexity of the microflora increase dramatically in the distal ileum. The total concentration of bacteria is between 10^6 and 10^8 per milliliter of contents, and anaerobic bacteria such as *Bacteroides*, *Bifidobacterium*, and *Clostridium* are found in substantial numbers. Distal to the ileocecal sphincter, the number of bacterias increase dramatically. In the colon, the total bacterial count approaches 10^{12} per milliliter of fecal material, and one-third of the fecal dry weight consists of viable bacteria. The complexity of the microflora also increases in the colon, and a number of different strict anaerobes (see Table 1) become predominant components of the flora.

Table 1 Distribution of the bacterial flora in human gastrointestinal tract

Region of the GI tract	*Bacteria*	*Number per milliliter of contents*
Stomach	*Streptococcus*	$10^1 - 10^2$
	Lactobacillus	$10^1 - 10^2$
Small intestine		
Duodenum and jejunum	*Streptococcus*	$10^2 - 10^4$
	Lactobacillus	$10^2 - 10^4$
	Veillonella	$10^2 - 10^4$
Ileum–cecum	*Bacteroides*	$10^4 - 10^8$
	Clostridium	
	Streptococcus	
	Lactobacillus	
	Enterobacteria	
	Bifidobacterium	
Colon and feces	*Bacteroides*	10^{11}
	Bifidobacterium	$10^{10} - 10^{11}$
	Enterobacter	10^{10}
	Peptococcus	10^{10}
	Clostridium	10^{10}
	Streptococcus	$10^9 - 10^{10}$
	Fusobacterium	$10^9 - 10^{10}$
	Lactobacillus	$10^6 - 10^8$
	Veillonella	$10^6 - 10^8$

Development of the Human Intestinal Microflora

The human fetus exists in a sterile environment until birth. Upon passage through the vaginal canal and exposure to the outside environment, the newborn gastrointestinal tract is rapidly colonized. In breastfed and bottlefed infants, by the third day after birth, *Bacteroides, Bifidobacterium, Clostridium*, and *Enterobacter* are isolated from the feces. For breastfed infants, by day 7, Bifidobacterium is the predominant organism accounting for more than 90% of the bacterial population. For bottlefed infants, by day 7, *Enterobacterium* accounts for the majority of the fecal population. At 1 month of age, *Bifidobacterium* is predominant in both breast and bottlefed infants, but breastfed infants have approximately 10-fold higher *Bifidobacterium* counts. With the introduction of a varied diet for bottlefed infants and weaning for breastfed infants, the microflora

becomes more complex, and the numbers of *Bacteroides* and anaerobic Gram-positive cocci increase over a period of a few months until their numbers equal those of *Bifidobacterium*. By the second year of life, the child's intestinal microflora come to resemble those of an adult.

Intestinal Microflora of the Elderly

Studies in the elderly have indicated that there are some changes that occur in the flora of humans in the seventh through the ninth decade of life. Some of the more significant changes are shown in Table 2. The numbers of *Lactobacillus, Streptococcus, Enterobacter*, and *Clostridium perfringes* increase, and *Bifidobacterium* decreases.

Interaction of the Microflora with the Intestinal Mucosa

Owing to the difficulty of sampling human colonic mucosa, there is limited information on the type (and number) of bacteria that are firmly attached to the colonic epithelial mucosa. The current information available indicates that Gram-positive rods and cocci and spirochetes are located on the mucosal layer. There is a maximum of 10^8 bacteria adhering per gram of colonic mucosal epithelial tissue, and *Bacteroides* is the most common isolated organism. Antibiotics can reduce by 5 logs the number of bacteria on the colonic surface.

Alteration of Normal Gastrointestinal Microflora

The composition of the gastrointestinal microflora is relatively stable. New bacteria introduced into the gastrointestinal tract have difficulty permanently implanting into an established flora. Diet can affect alterations in the microflora, but comparative studies on human populations eating radically different diets have revealed only selective changes (see below) and not a general shift in the flora. In animals, it has been demonstrated that limiting nutrient intake can cause major changes in the composition of the microflora. In humans, it is not known if self-imposed severe dieting or anorexia will result in substantial changes in the composition or distribution of the microflora.

Two known perturbants of human gastrointestinal ecosystem: antibiotics and chemotherapeutic agents can drastically alter the microflora, and this is often reflected in diarrhea and, in some cases, an enteritis. A specific example is the development of *Clostridium difficile*-induced pseudomembraneous colitis after antibiotic treatment. The antibiotic alteration of the intestinal flora interferes with the normal suppression of sporulation of *Clostridium difficile* with subsequent pathogenic toxin production.

Table 2 Comparison of counts per gram of feces for selected bacteria of the fecal flora

Bacterial genera	*Log of counts*	
	Adults	*Elderly*
Lactobacillus	5.8	7.5
Enterobacter	7.8	8.2
Streptococcus	7.9	7.4
Clostridium perfringens	4.4	6.6
Bifidobacterium	10.0	9.4

Diet and the Gastrointestinal Microflora

There have been a number of studies in humans investigating the ability of diet to alter specific components of the adult flora. One study reported no change in the predominant organisms in the fecal flora of individuals changing from an omnivorous to a vegetarian diet. The most detailed study involved a comparison of subjects eating a Western diet with Japanese eating a traditional diet and vegetarian and nonvegetarian Seventh-Day Adventists. The subjects eating a Japanese diet had higher fecal counts of *Streptococcus faecalis, Eubacterium lentum*, and *Eubacterium contortium*, and lower counts of *Bacteroides*.

In summation, dietary studies indicate that different eating patterns influence specific bacteria in the intestine, but overall, changes in the composition of the flora are hardly dramatic.

Effect and Role of the Gastrointestinal Microflora

Metabolic Activities of Microflora

In order to discuss the effects and role the microflora has on the host, a discussion of the metabolic activity of the flora should be considered. The more than 10^{14} organisms that occupy the intestinal tract at any given time can perform a large number of different reactions, which can impact on the host. Any compound taken orally or any substance entering the intestine through the biliary tract or the blood or by secretion directly into the lumen can potentially be a substrate for bacterial conversion. Table 3 lists some of the major reactions performed by the intestinal bacteria. Most bacterial reactions can be classified as reductive, hydrolytic, or removal of various functional groups. A brief description of these reactions is presented below. The potential impact on the health and physiology of the host is discussed later in this article.

Table 3 Reactions performed by the gastrointestinal microflora

Reaction	*Example of substrate or products*
Hydrolysis	Estrogen glucuronide
Glycosides	Lactose
Sulfamates	Cyclamate
Amides	Methotrexate
Esters	Acetyldtigoxin
Nitrates	Pentaerythritol trinitrate
Dehydroxylation	Bile acids
C-hydroxy groups	*N*-hydroxyfluorenyl-acetamide
N-hydroxy groups	
Decarboxylation	Amino acids
Demethylation	Biochanin A
Deamination	Amino acids
Dehydrogenation	Cholesterol, bile acids
Dehalogenation	DDT
Reductive	*p*-nitrobenzoic acid
Nitro groups	Polysaturated unsaturated
Double bonds	Fatty acids
Azo bonds	Azo dyes
Aldehyeles	Benzaldehdes
Alcohols	Benzyl alcohols
N-Oxides	4-Nitroquinoline-1-oxide
Nitrosamines	Dimethylnitrosamine
Gas production	Products CO_2, H_2
Fermentation of sugars	Products lactate, butyrate, propionate, acetate, formate
Fiber, inulin, oligosacchrides, starch	Disaccharides, monosacchides, short-chain fatty acids
Fermentation of amino acids	Products H_2, CO_2, CH_4, amines, phenols, indoles, NH_3, organic acids
Aromatization	Quinic acid
Aceylation	Histamine
Esterification	Galic acid

Role and Effect of the Intestinal Flora

Metabolic Role

The hydrolysis of the glycoside linkage is one of the most important reactions performed by the intestinal microflora. Among the reactions performed by the bacteria are α- and β-glucosidase and galactosidase, β-glucuronidase, xylosidase, and cellobiase, and α-mannosidase and fucosidase. These reactions are responsible for the hydrolysis of oligo- and polysaccharides, as well as a wide range of exogenous and endogenous compounds conjugated by a glyosidic bond with glucuronic acid or other sugar moities. The bacterial glycosidases can remove terminal sugars as well as hydrolyzed bonds in the middle of a polysaccharide. The hydrolysis of amides, glucuronides, esters, and nitrites, and a number of other reactions listed in Table 3 relate to conversion of drugs, pesticides, and various endogenous compounds, such as cholesterol, estrogens, bile acids, and their conjugates produced in the liver. Decarboxylation and deamination are reactions related to the bacterial fermentation of amino acid, and the bacterial reactions of the Embden–Meyerhof–Parnas glycolytic pathway result in fermentation of hexoses to short-chain fatty acids.

Glycosides

Glycosides enter the intestine from two major sources: from the diet or via the biliary tract. The diet contains fiber, resistant starch, and phytochemicals, such as the glycosides of flavonoids and isoflavonoids. These compounds are not normally absorbed in the upper gastrointestinal tract and, therefore, serve as substrates for the vast majority of microflora that reside in the distal ileum, colon, and rectum. Glycosides coming from the liver include endogenous compounds, an example being estrogens and other biological molecules, and drugs or environmental agents that are conjugated by the liver via glycoside formation primarily with glucuronic acid subsequently secreted into the bile, which is deposited in the duodenum. The intestinal flora then hydrolyze the glycoside bond, leading to the release of potentially biologically active agycones (nonsugar moiety).

There are a number of proven examples of the importance of the action of bacterial glycosidases on the host. Cycasin is a compound that is found in the cycad plant, also referred to as tropical fern. The compound is a glycoside of methylazoxymethanol, which can cause intestinal tumors in conventional and germ-free mice and rats. Germ-free animals lack an intestinal microflora and do not develop tumors when given cycasin; in contrast, conventional animals when fed cycasin develop intestinal tumors. These data suggest that the bacterial glycosidase in the intestine is necessary for the carcinogenic activity of cycasin. This has been confirmed when germ-free animals are monocontaminated with bacteria having different levels of glycosidase. The carcinogenicity after feeding cycasin has been found to be positively correlated with the specific implanted bacterial strain.

The action of the intestinal microflora on the plant glycoside rutin is another example of the potential for generating carcinogens. Rutin is not mutagenic prior to the hydrolysis of the glycoside linkage, but the hydrolysate is mutagenic.

There are a number of drugs whose action and/or pharmacokinetics are influenced by bacterial glycosidases. The metabolism of the cardiac glycoside digoxin is an example of the effect of bacterial action. To produce the drug's pharmacological action, the

bacterial flora must remove a trisaccharide from the parent compound, releasing diagoxigenin. The fecal flora of 36% of the residents of New York given digoxin has been found to reduce the double bond in the lactone ring, resulting in the formation of dihydrodigoxigenin. This metabolite is not pharmacologically active. The extent of this bacterial modification is reflected in the observation that 14% of New Yorkers had high levels of digoxin metabolites in their feces. This would result in lower-than-predicted serum levels in these individuals. In this example, the intestinal microflora inhibit the action of an orally administered drug. The opposite is true for the cathartic agent, Cascara Sagrada. This agent is a mixture of glycosides that are not pharmacologically active until bacterial glycosidases in the intestine release the active aglylone.

An important physiological implication of the bacterial hydrolytic glycosidic reaction is the cleavage of disaccharides. Lactose, sucrose, and maltose are disacchrides that are normally absorbed in the upper small intestine after hydrolysis by tissue enzymes located in the intestinal mucosal brush border. Therefore, contact with the majority of intestinal microflora does not occur. However, because of either a genetic defect or an acquired disorder, individuals who have few or no disaccharidases do not absorb the disaccharides in their small intestine. The sugars are transported to the ileum and large bowel, where bacterial sucrase, lactase, and maltase hydrolyze the disaccharidases. The final bacterial end products are short-chain fatty acids. An osmotic imbalance results from the increased concentration of short-chain fatty acids, and water flows into the lumen of the bowel, resulting in diarrhea.

The most common disorder is lactose intolerance, which is common in adults residing in temperate and warm climates. Less frequently encountered is intolerance towards sucrose or maltose.

Amino Acids and Other Amines

To some extent, food-derived proteins can remain intact until they reach the large bowel. In addition, proteins from the epithelial cells, and digestive secretions and mucins all contain protein. Bacteria deaminate amino acids by five different mechanisms. Amino acids also undergo bacterially derived decarboxylation. These reactions result in the formation of a wide range of different amines and short-chain organic acids that can be absorbed from the colon. This process helps the host reclaim energy. The overproduction of, or lack of, absorption of hydroxy short-chain fatty acids derived from amino acid fermentation can cause diarrhea.

Fermentation of Carbohydrates

In this section, a more general description will be given of the fermentation of carbohydrates by the intestinal microflora. The major source of fermentable carbohydrates in the human colon is plant cell-wall polysaccharides, such as pectins, cellulose, and hemicellulose. Other sources of carbohydrate in the colon are resistant starch and intestinal mucus. It is estimated that between 20 and 70% of carbohydrates are fermented per day. The end products of carbohydrate fermentation in the feces are acetate, propionate, and butrarate. This fermentation also leads to the production of hydrogen, carbon dioxide, methane, and water.

Bacterial Lipid Metabolism

In humans, most of the fatty acids derived from dietary lipids are absorbed in the small intestine. The human colonic anaerobic microflora can hydrogenate and hydrate unsaturated fatty acids. Evidence for these reactions comes from the isolation of 10-hydroxystearic acid in the feces. The metabolism of lipids by the intestinal flora has only a minor role in humans.

Bacterial Conversion of Cholesterol, Bile Acids, and Bile Pigments

Cholesterol is a precursor of bile acids, and both classes of compounds are structurally similar. These compounds are endogeously synthesized in the liver initially from two carbon units.

Cholesterol is also a component of the diet. Between 75 and 200 mg of cholesterol and its metabolites are excreted in the feces daily. There are two principal fecal metabolites of cholesterol, coprostanone and coprostanol, and one minor product, cholestenone. The formation of coprostanol requires a steroid nuclear hydrogenation of the 5,6 double bond. The conversion of cholesterol to coprostanone results from the reduction of the 4,5 bond and a C3 oxido-reductase converting the hydroxyl to a keto group. Coprostanol accounts for approximately 50% of the total fecal neutral sterols, coprostanone 10–15%, and unmetabolized cholesterol the remainder. In germ-free animals, only cholesterol is recovered in the feces, providing evidence that the intestinal microflora is solely responsible for the metabolism of cholesterol in the intestine. The precentage of intestinal cholesterol conversion varies from 70 to 85% in Americans and Western Europeans to 55 to 65% in Africans and Asians.

Bile acids synthesized in the liver are conjugated via an amide bond to glycine or taurine. The conjugated

bile acids are secreted in bile and are deposited in the upper small intestine. The principal intestinal bacterial reactions involve bile acids, primarily occuring in the distal ileum and colon. The bacterial conversion of bile acids includes: the hydrolysis of the amide bond to release free bile acid from the glycine and taurine conjugates; an oxidoreduction of the hydroxyl groups at C3, C7, and C12 to form either oxo bile acids (hydroxyl to keto groups) or α-hydroxyl groups after the reduction of the keto groups (inversion products); and dehydroxylation at C7, and to smaller extent at the C3 and C12 positions. The result of these reactions is the conversion of primary bile acids to secondary bile acids and the reabsorption of free bile acids from the terminal ileum and, to a lesser extent, from the large bowel.

Bile pigments derive from the breakdown of hemoglobin. Between 200 and 300 mg of bilirubin, the product derived from heme, are excreted daily in the bile. Approximately 90% of bilirubin is conjugated to glucuronic acid. The bilirubin conjugate is hydrolyzed by the intestinal microflora releasing the free compound, which can be reabsorbed in the intestine. Alternatively, bilirubin can be reduced by the microflora to urobilonogen. Approximately 20% of urobilinogen is reabsorbed from the intestine of humans. The major components of bile, namely, cholesterol, bile acids, and bile pigments, are subject to bacterial action in the intestine. Additional compounds secreted in the bile such as, androgens and estrogens (discussed below), are also subject to the action of the intestinal microflora.

Additional Intestinal Bacterial Reactions

Vitamins

The bacteria residing in the human intestinal tract can synthesize homologs of vitamin K_2 (menoquinone 7). The bacterial conversion, in part, occurs in the ileum where menoquinone can be absorbed. Human adults given low vitamin K diets for several weeks did not exhibit a deficiency. Treatment of subjects with antibiotics at the same time as they were fed a low vitamin K diet had a significant decrease in plasma prothrombin levels. The synthesis of the peptides in the prothrombin blood-clotting complex requires menoquinone. These human studies demonstrate the importance of intestinal bacterial metabolism with respect to vitamin K synthesis.

Humans depend solely on this action of intestinal microflora for their vitamin B_{12} (cyanocobalamin) requirement indirectly via the consumption of meat from ruminants that derive B_{12} from synthesis by their microflora. In addition, the human intestinal microflora contributes significant amounts of vitamin B_{12}, and approximately 5 μg is excreted in feces daily. At least two organisms, *Pseudomonas* and *Klebsiella*, have been shown to synthesize vitamin B_{12} in the small intestine.

Biotin is synthesized to a major extent by the intestinal flora. The administration of high doses of antibiotics can cause biotin deficiency in laboratory animals. The administration of antibiotics can lower biotin urinary levels.

Several other B complex vitamins, namely folic acid and thiamin, are synthesized by the human intestinal microflora. The amount synthesized by bacteria is not sufficient to eliminate the requirement for dietary sources of folate and thiamin.

Estrogens and Androgens

Estrogens and androgens are modified by the intestinal microflora. Estrone, estradiol, and estriol are three major circulating estrogens excreted in the bile, conjugated to glucuronic acid and/or sulfate. Bacteria found in the feces can hydrolyze these conjugates, releasing the free estrogens. The nonconjugated estrogens are then subject to further bacterial action. A major reaction involves oxidoreduction of the C17 position. Bacteria can convert estrone to estradiol, and the fecal flora can also convert 16 α-hydroxyestrone to estriol.

Several intestinal bacterial conversions of androgens are also catalyzed by the human fecal flora. The intestinal microflora can reversibly oxidize and reduce the 3-hydroxy group, reducing steroid nuclear double bonds at the 1- and 4-positions and resulting in a complex interconversion of androgens. Evidence for an introduction of a double bond into the androgen nucleus has also been described. However, this reaction only occurs under aerobic conditions and therefore would be unlikely to occur in the highly anaerobic environment of the colon.

Role of the Intestinal Flora in Health

Colon Cancer

Colon cancer is a common disease in Western Societies. Epidemiological studies have shown that the incidence of colon cancer is higher in North America and Western Europe than in Africa, Asia, and South America. A number of studies have shown a positive correlation between consumption of beef, fat, and protein, and a negative correlation with fiber, and the incidence of colon cancer. It has been proposed that the dietary influence on the etiology of colon cancer results, in part, from alteration of the metabolic activity of the

intestinal microflora, more specifically, the ability of intestinal bacteria to convert procarcinogens to proximate carcinogens. This conversion would be greatest in the large bowel where the bacterial population is highest, and the transit time for the fecal stream is the slowest. Several bacterial enzymes have been implicated in the formation of mutagens, carcinogens, and tumor promoters. Among the enzymes that have the potential to increase the incidence of colon cancer are: β-glucosidase, β-glucuronidase, β-galactosidase, azoreductase, nitroreductase, 7-α-steroid dehydrogenase, and 7-α-hydroxy-steroid dehydroxylase.

There have been at least five different classes of mutagens isolated in the feces that have the potential to cause colorectal cancer. The final step in the synthesis of fecapentenes (dodecapentaenyloxy-1,2 propanediol), one class of mutagen, involves the action of the intestinal flora.

Bile acids, particularly the secondary bile acids, can act as colon tumor promoters, and, as discussed previously, the fecal microflora are responsible for the generation of secondary bile acids in the colon.

Heterocyclic aromatic amines are also procarcinogens, which can be acted on by bacterial and intestinal tissue enzymes to generate proximal carcinogens. Nitrosamines, which are also present in the colon, are often direct-acting and spontaneously breakdown to electrophiles, which react with DNA and do not require activation by bacteria or tissue enzymes. The final broad class of known potential colon carcinogens are polycyclic aromatic hydrocarbons, which are generally activated by tissue microsomal enzymes. From this brief discussion, it is apparent that the intestinal flora can be an important factor in the etiology of colon cancer.

Bacterial Overgrowth

Bacterial overgrowth occurs when there is a significant increase in the bacterial population in the small intestine and stomach. The proliferation of bacteria in the upper gastrointestinal tract can be caused by anatomic or physiologic derangements. One common cause of bacterial overgrowth is the underproduction of gastric acid. Gastric acid limits bacterial growth in the stomach and upper small intestine. There are a number of different causes for decreased gastric acid production, such as atrophic gastritis, pernicious anemia, surgical resections, and drug therapy. The consequences of bacterial overgrowth are numerous, including steatorrhea, vitamin deficiencies, and carbohydrate malabsorption.

More than 20 different species of bacteria have been identified in the upper small intestine of patients with bacterial overgrowth. The most common clinical manifestation of bacterial overgrowth is malabsorption of fat and fat-soluble vitamins. Small bowel bacterial overgrowth can also cause megaloblastic anemia, which results from vitamin B_{12} deficiency. Binding of vitamin B_{12} to bacteria in the upper small intestine can prevent absorption of the vitamin in the distal ileum. *In-vitro* studies have demonstrated that *Bacteroides*, among the most common of intestinal bacteria, bind the intrinsic factor–cobalamin complex. Patients with bacterial overgrowth can have high levels of B_{12} in the lumen of the small intestine, from nonabsorbed dietary sources and from local bacterial synthesis. Despite this, patients can still suffer from vitamin B_{12} deficiency because of the lack of absorption from the intestine.

Other consequences of bacterial overgrowth include impaired amino acid and carbohydrate absorption. Fecal nitrogen content is increased, serum proteins are low, and this could result in a clinical protein-calorie malnutrition.

Diarrheal Diseases

Disruption of the intestinal flora by antibiotics or other drugs can lead to diarrhea. A disruption in the balance of the intestinal microflora can lead to the overgrowth of pathogens in the large bowel. An example is antibiotic-induced *Clostridium difficile* colitis. This pathogen resides as a spore in the colon of a significant number of healthy individuals. Patients receiving antibiotics for various infections have, as a result of their treatment, a decrease and shift in the population of the normal flora. In some instances, this leads to situations in which sporulation of *Clostridium difficile* takes places with a concomitant production of toxin, resulting in severe diarrhea and, in some cases, a pseudomembraneous colitis. The use of antibiotics in general, in a hospital setting, results in diarrheal disease in 30–50% of the antibiotic-treated patients.

The etiological agents in most of the cases are unknown. Another common cause of diarrhea is the introduction of pathogens from food, water, or other orally ingested material. The enterobacteriaceae groups of Gram-negative rods include a number of diarrheal pathogens, including specific strains of *Escherichia, Shigella, Salmonella*, and *Yersinia*. Other pathogens are *Vibrio cholerae* and *Mycobacterium*. Most of these pathogens are not normal residents of the intestinal tract; however, these agents can disrupt the normal intestinal flora.

Health Implications of Probiotics

Just as bacterial pathogens introduced into the lumen of the intestine can cause disease, the feeding of

Table 4 Probiotics in use for humans and animals

Genera	*Lactobacillus*	*Bifidobacterium*	*Streptococcus*	*Additional bacteria*	*Yeast and molds*
Species	*acidophilus* *casei* *rhamnosus* *reuterii* *delbrueckii* ssp. *bulgaricus* *brevis* *fermentum* *lactis* *cellobiosus* *plantarum*	*infantis* *longum* *adolescents* *animalis* *bifidum* *thermophilus*	*cremoris* *lactis* *salivarius* ssp. *thermophilus* *intermedius*	*Leuconostoc* *Pediococcus* *Bacillus* *Propionibacterium* *Enterococcus* *E. faecium*	*Aspergillus niger* *A. cerevisiae* *A. oryzae* *Candida pintolopesii*

Table 5 Some of the health claims made for probiotics

Intestinal problems	*Other disorders and diseases*	*Other uses*
Constipation	Cystic fibrosis	Adjuvant for vaccines
Colitis	Vaginitis	Recolonization of bowel after antibiotic use
Lactose intolerance	Alcohol-induced liver disease	
Salmonella *and* Shigella *infections*	Cancer	
Flatulence	Hypercholesterolemia	
Crohn's disease	Food allergies	
Diarrhea		
Antibiotic-induced		
Infantile		
Travelers		

specific types of bacteria can be beneficial to the host. The term probiotic has been given to bacteria that, when fed, confer health benefits to the host.

Table 4 lists some of the microorganisms currently being fed to humans and animals. A large number of health benefits have been attributed to probiotics. A partial list of these therapeutic claims for probiotics is given in Table 5. A brief discussion of some of these claims is included below.

Diarrheal Diseases

An initial report with five patients treated with *Lactobacillus rhamnosum* strain GG (LGG) for relapsing diarrhea caused by *Clostridium difficile* toxin showed that this probiotic could stop the diarrheal symptoms. An additional 32 subjects were subsequently studied. Of these 32 patients, 84% were cured by initial treatment with LGG daily for 2 weeks. Three additional patients were cured upon retreatment with LGG, resulting in an overall cure rate of 94%.

LGG was also shown to lower the rate of diarrhea in Finnish travelers to Turkey. A study has also been performed with American travelers to developing countries fed either LGG or a placebo in a double-blind study design. The subjects taking the placebo had a 7.4% per day risk of developing diarrhea, and travelers who ingested LGG had a 3.9% per day risk. These data indicate that LGG provides a 47% protection rate against traveler's diarrhea.

LGG has also been shown to be effective for treating children hospitalized for severe diarrhea. Children given LGG had a decrease in the duration of their diarrhea from 2.5 to 1.1 days.

These studies indicate that a probiotic can be helpful in preventing antibiotic-induced, travelers', and young childhood diarrhea.

Lactose Intolerance

Ingestion of yogurt when compared with feeding milk containing equal amounts of lactose resulted in one-third the amount of hydrogen in the breath. Breath hydrogen is an indicator of the passage of intact lactose to the lower intestine. The reason for this observation rests on the fact that the *Lactobacillus* and *Streptococcus* in yogurt hydrolyze the lactose in the yogurt, as well as in the upper small intestine.

Food Allergy

Casein in milk can cause the first allergic reaction in infants. *Lactobacillus* can degrade milk proteins to smaller peptides and amino acids. One of the symptoms of milk allergy is atopic eczema. Several studies have been performed showing that the feeding of LGG to infants significantly decreased clinical allergic symptoms when compared with a placebo group. These results suggest that probiotics down-regulate intestinal inflammation and hypersensitivity

reactions in infants with food allergies and resulting atopic eczema.

Liver Disease

Studies in rats chronically fed alcohol have shown that feeding LGG reduced liver pathology and plasma endotoxin levels. These findings suggest that probiotics may be useful in preventing alcohol-induced liver damage, a common human medical problem.

Colon Cancer

There have been a number of studies showing that feeding *Lactobacillus acidophilus* or LGG lowers colon tumor incidence in rats administered the carcinogen dimethylhydrazine. There are no human clinical intervention trials that have been performed to determine if probiotics can prevent colon cancer.

See also: **Amino Acids**: Properties and Occurrence; ***Bifidobacteria* in Foods**; **Bile**; **Biotin**: Properties and Determination; Physiology; **Cholesterol**: Absorption, Function, and Metabolism; **Colon**: Structure and Function; Diseases and Disorders; Cancer of the Colon; **Fats**: Digestion, Absorption, and Transport; **Food Intolerance**: Food Allergies; Lactose Intolerance; **Hormones**: Steroid Hormones; **Lactic Acid Bacteria**; **Probiotics**; **Vitamin K**: Properties and Determination

Further Reading

De Kok TMCM and Van Maanen JMS (2000) Evaluation of fecal mutagenicity and colorectal cancer risk. *Mutation Research* 463: 53–101.

Gibson GR and Macfarlane GT (1995) *Human Colonic Bacteria*. Boca Raton, FL: CRC Press.

Goldin BR (1986) *In situ* bacterial metabolism and colon mutagens. *Annual Review of Microbiology* 40: 367–393.

Goldin BR (1998) Health benefits of probiotics. *British Journal of Nutrition* 80 (supplement 2): S203–S207.

Goldin BR and Gorbach SL (1989) *Impact of Anaerobic Bowel Flora on Metabolism of Endogenous and Exogenous Compounds*. In: *Anaerobic Infections in Humans*. New York: Academic Press.

Goldin BR, Lichtenstein AH and Gorbach SL (1994) Nutritional and metabolic roles of intestinal flora. In: *Modern Nutrition in Health and Disease*, 8th edn. Philadelphia, PA: Lea & Febiger.

Hill MJ (1986) *Microbial Metabolism in the Digestive Tract*. Boca Raton, FL: CRC Press.

Mitsuoka T (1992) *The Human Gastrointestinal Tract. The Lactic Acid Bacteria in Health and Disease*. London: Elsevier Applied Science.

Salminen S (1996) *Lactobacillus* GG. *Nutrition Today* 31(No. 6 supplement 1): 1S–52S.

Schrezenmeir J, de Vrese M and Heller K (2001) International Symposium on Probiotics and Prebiotics. *American Journal of Clinical Nutrition* 73(2S): 361S–498S.

Probiotics

J L Rašić, Jasenova 11/II (Lerak), Belgrade, Yugoslavia

Introduction

The word 'probiotic' derives from the Greek words meaning 'for life.' It refers to live microbial dietary supplements which contribute to the intestinal microbial balance, thus benefiting host health; they can also beneficially affect the properties of the indigenous flora. The use of probiotics in the diet stems from Metchnikoff's theory (1907) that the consumption of large amounts of yogurt containing lactobacilli prolongs life. Many reports since then have shown that gut organisms, such as bifidobacteria, and other lactobacilli also beneficially affect the host, including human hosts.

The indigenous gut microorganisms have their own protective function against invading pathogens. Various factors, including diet, can affect the protective gut flora. Special ingredients selectively stimulate the growth and/or activity of a small number of beneficial organisms in the colon. The potential benefits of ingested bifidobacteria and lactobacilli include various healthful aspects. Selection of microbial strains is important in the development of improved probiotics. There is evidence that probiotics ingested can influence the gut flora.

Overall Concept of the Importance of the Gut Flora

The gut microorganisms colonize available habitats in the gastrointestinal tract and, although not essential for life, they live in a stable relationship with their host. Their presence is manifested by their effects on the development of resistance factors, including defense systems of the host, the protective effect of the flora, and the production of microbial products.

Defense Systems and the Protective Flora

Experiments with gnotobiotes (animals which are either germfree or have been inoculated with one or more known species of microorganisms) have shown that the lymph nodes are larger and more numerous, and the number of lymphocytes within the nodes is greater in conventional animals (those which retain their normal gut flora) than in germfree animals. Lymphocytes are the major constituents of the lymphatic tissues, and are associated with immunity. Specific lymphocytes participate, via plasma cells, in the production of immunoglobulins.

The levels of plasma cells and immunoglobulins are much lower in germfree than in conventional animals, but when germfree animals are exposed to microorganisms, the lymphatic tissues rapidly enlarge and there is a gradual increase of plasma cells and immunoglobulins. Consequently, microorganisms are the major source of antigens in normal animals, including humans, with dietary constituents making a secondary contribution. The macrophages from germfree animals digest bacteria, but more slowly than those from conventional animals.

The gut microorganisms have their own protective function. They prevent colonization of an area in the gut by invading pathogens by competition for essential nutrients and for attachment sites on the epithelium. They inhibit the growth of invading pathogens by the production of organic acids, particularly volatile fatty acids, by the deconjugation of bile salts, and by the production of antibacterial substances other than acids. The protective effect of the indigenous gut flora against pathogens has been proved experimentally. For example, *Shigella* spp. are lethal to germfree guinea-pigs, whereas they are unable to establish in conventional guinea-pigs. Similarly, mice treated with streptomycin show increased sensitivity to *Salmonella*. The indigenous gut flora normally acts synergistically with its host's immunological system in protecting against infections by intestinal pathogens.

Microbial Metabolic Products

The gut microorganisms break down undigested or unabsorbed dietary constituents, body secretions, and ingested foreign compounds; they also produce certain vitamins. Consequently, numerous metabolic products occur in the gut. These include a variety of organic acids, protein degradation products, dehydroxylated bile acids, cholesterol metabolites, and others. Many such products are potentially harmful to humans, but most of them are detoxified in the liver before excretion. High levels of some products (e.g., dehydroxylated bile acids, cholesterol metabolites) found in the gut, due to the effect of long-term changes in diet on the metabolic activities of the indigenous flora, may be implicated in the etiology of some degenerative diseases and cancerous conditions. The type of diet influences the levels of harmful microbial products. (*See* **Bile**; **Protein**: Synthesis and Turnover.)

Influence of Diet on the Gut Flora

Effects of Composition of the Flora

It is difficult to induce major changes in the indigenous gut flora by alterations in the diet of healthy adults. Dietary effects on the composition of the fecal flora are demonstrably slow and their long-term effects may be variable. Subsequent studies suggest that the type of diet may be the major factor causing the variation in fecal flora and enzyme activity among various populations.

One study examined the fecal flora of nine rural healthy Japanese people and eight urban healthy Canadians, who were consuming typical Japanese and western diets, respectively. The number of eubacteria, bifidobacteria, lactobacilli, and veillonelae, and the frequency of occurrence of bifidobacteria were higher in the Japanese than in the Canadians, but larger numbers of *Bacteroides* and lecithinase-negative *Clostridium* were found in the Canadians. In another study, the fecal flora of 15 healthy elderly persons in a rural area in Japan were compared with the flora of persons living in urban areas. The diet of elderly persons was characterized by a high intake of dietary fiber. Significantly larger numbers of bifidobacteria, but not of clostridia, were found in the group of elderly persons. Dietary fiber composed of cellulose, noncellulosic polysaccharides (hemicellulose, pectin, inulin, guar, and plant gums) and lignin is resistant to digestion by the human intestinal tract. Some of the fiber is degraded in the colon by the flora, and the main products of digestion are volatile fatty acids, gases, and energy. (*See* **Dietary Fiber**: Properties and Sources.)

Effects on Microbial Metabolism

The metabolic activities of the gut flora may be affected by alterations in the diet, and extreme diets may induce great changes in the bacterial activities. For example:

1. Bacteria isolated from persons living on a high-fat diet were more active in the dehydroxylation of bile acids than were similar bacteria isolated from persons living on a low-fat diet.
2. Bacteria isolated from persons living on a diet rich in dietary fiber may be more active in digesting cellulose and noncellulosic polysaccharides than are similar bacteria isolated from persons living on a low-fiber diet, because the enzymes catalyzing polysaccharide degradation are inducible.
3. A diet rich in meat products but low in fiber indirectly provides the flora with urea, amino acids, bile acids, and neutral steroids. As a result, increasing concentrations of ammonia, amines, indole, dehydroxylated bile acids, and cholesterol metabolites occur in the gut, along with increasing activities of beta-glucoronidase, azoreductase, and nitroreductase enzymes. (*See* **Dietary Fiber**: Physiological Effects.)

Influence of Special Ingredients (Prebiotics)

Special ingredients refer mainly to nondigestible oligosaccharides, that benefit the host by selectively stimulating the growth and/or activity of one or a limited number of beneficial bacteria (e.g., bifidobacteria and lactobacilli) in the colon. These ingredients, called prebiotics, are not hydrolyzed in the small intestine; they are not absorbed and reach the colon where they are selectively utilized by the bacterial flora.

The best-known prebiotics are: fructooligosaccharides such as oligofructose and inulin, lactulose, maltooligosaccharides, galacto-, xylo-, gentio-, palatinose-oligosaccharides, etc. Among many nondigestible ingredients, lactulose has been most extensively studied for its growth-promoting effect *in vivo* for bifidobacteria and lactobacilli. It is not present in a free state in milk but, upon heating milk, a part of the lactose is converted to lactulose, and in sterilized liquid formula feeds it amounts to 2–8% of the total carbohydrate content. High levels of lactulose have laxative effects, leading to diarrhea. (*See* **Carbohydrates**: Metabolism of Sugars.)

Lactulose is used for babies in some formula feeds, and for adults to promote the growth of bifidobacteria and lactobacilli. It is also used in the management of liver disease, and in the treatment of chronic constipation.

Other nondigestible ingredients occur naturally in many foods (e.g., Jerusalem artichoke, onion, wheat), and some are isolated or synthesized. They are commercially available under various trade names.

Potential of Ingested Bifidobacteria and Lactobacilli

Selected strains of bifidobacteria and lactobacilli are commonly used in probiotics to suppress harmful microbial populations in the gut and encourage beneficial organisms.

The major potential benefits of ingested bifidobacteria and lactobacilli are claimed to be:

1. modification of the gut flora in infants (bifidobacteria)
2. prophylactic and therapeutic functions
3. improved lactose utilization (lactobacilli)
4. possible anticarcinogenic activity
5. control of serum cholesterol
6. stimulation of the immune system

Modification of the Gut Flora in Infants

It may be possible to modify the gut flora of formula-fed infants if large numbers of *Bifidobacterium bifidum* (10^8–10^9 cells day^{-1}) are fed, if they survive gastric transit, and if a fermentable carbohydrate is available to the cells in the large intestine. Since lactose of formula feeds does not reach the large intestine in sufficient amounts to promote the growth of bifidobacteria, the addition of bifidogenic substances could provide a fermentable carbohydrate. The improved formulation of milk is also an important factor. (*See* **Infants**: Breast- and Bottle-feeding.)

Prophylactic and Therapeutic Functions

There is a clear evidence that bifidobacteria and lactobacilli are antagonistic to various harmful bacteria in the gut. Bifidobacteria (*B. bifidum, B. longum, B. infantis*, etc.) produce acetic and lactic acids, with small amounts of formic acid; and some strains produce antimicrobial substances other than organic acids. The antagonistic effect of acetic acid against Gram-negative bacteria is stronger than that of lactic acid, and the former is produced in greater amounts by bifidobacteria. Lactobacilli, such as *Lactobacillus acidophilus* and *L. casei*, create lactic acid and small amounts of hydrogen peroxide to suppress harmful bacteria; some strains of *L. acidophilus* produce antibiotic-like substances, including acidophilin, lactocidin, acidolin, and possibly others. Deconjugation of bile salts to free bile acids by both bifidobacteria and lactobacilli may also function in the control of the bacterial flora. Large numbers of bifidobacteria and lactobacilli in the gut may involve competitive antagonism against invading pathogens.

The protective effect of bifidobacteria and lactobacilli against enteric infections and the side-effects of oral antibiotic therapy was shown in human studies. Comparative feeding studies of two groups of infants, one receiving formula feeds containing viable *B. bifidum* and lactulose, and the other receiving a buttermilk preparation, showed that enteric infections were eight times more frequent in the buttermilk-fed group than in the formula-fed group. The beneficial effect of *L. acidophilus* or *L. casei* in the prevention of side-effects of oral antibiotic therapy (e.g., moniliasis, staphylococcal enteritis) in children was also shown. It has been suggested that daily ingestion of large numbers of the lactobacilli (e.g., *L. acidophilus*) may protect consumers from diarrhea caused by intestinal pathogens (e.g., traveler's diarrhea). Dietary supplements of bifidobacteria and lactobacilli have beneficial effects in children with enteric infections. For example, the ingestion of a freeze-dried culture of *B. bifidum* in conjunction with lactulose was shown to eradicate enteropathogenic *Escherichia coli* strains in more than 80% of cases. Many reports indicated the beneficial roles of bifidobacteria in the management of chronic liver

disease. The ingestion of *B. bifidum* together with lactulose was shown to assist in reestablishing the balance of the gut flora, which is usually disturbed in liver cirrhosis, and this was accompanied by a decrease in fecal pH and by a reduction of ammonia and free phenols in the blood. (*See* **Liver**: Nutritional Management of Liver and Biliary Disorders.)

The beneficial effects of ingested bifidobacteria and/or lactobacilli may be obtained if: (1) large numbers of viable cells (10^8–10^9 cells day^{-1}) are introduced; (2) they survive gastric transit and, preferably, can adhere to epithelial surfaces and grow; (3) a fermentable carbohydrate is available in the gut; and (4) they have a strong antagonistic effect against harmful microorganisms.

Improved Lactose Utilization

It was shown that lactase-deficient persons can digest lactose in yogurt better than the same amount of lactose in unfermented milk. Yogurt bacteria, *Streptococcus salivarius* subsp. *thermophilus* and *L. delbrueckii* sp. *bulgaricus*, produce the enzyme, lactose (beta-galactosidase), which hydrolyzes lactose to glucose and galactose. It is possible that they supply preformed lactase to the gut, thus allowing digestion of lactose. *L. acidophilus* and *L. casei* also allow digestion of lactose by producing lactase. These organisms survive in the gut better than yogurt bacteria, thus suggesting that lactase from the bacterial cells remains in the small intestine for a longer time.

Possible Anticarcinogenic Activity

The potential anticarcinogenic activity of some lactobacilli and bifidobacteria has been shown in many studies. The effect of *L. acidophilus* on Ehrlich ascites tumor was studied using mice as an animal model. Feeding acidophilus milk resulted in a smaller number of tumor cells in mice than in those not receiving *L. acidophilus*. Similar results were obtained with mice, transplanted with Meth-A ascites tumor cells, that received a suspension of *B. infantis*, repeated four to six times. (*See* **Cancer**: Diet in Cancer Prevention.)

Other studies with mice indicated a possible role of macrophages in suppressing the growth of tumor cells. Intraperitoneal injection of *L. casei* in mice was shown to increase the phagocytic activity of peritoneal macrophages and their acid phosphatase activity. In addition, the consumption by mice of milk containing *L. casei* resulted in activation of the macrophages, measured by increased levels of lactic dehydrogenase activity.

Some of these effects may be due to substances produced by the organism during growth, and/or to the stimulation of the host immune response. It is also possible that some lactobacilli and bifidobacteria inhibit the growth of organisms which may convert procarcinogens into carcinogens in the gut. The oral administration of *L. acidophilus* (10^{10} cells day^{-1}) to meat-fed rats was shown to reduce the activity of azoreductase, beta-glucuronidase, and nitroreductase. These fecal enzymes can catalyze procarcinogen conversion to a proximal carcinogen. In a study with human subjects it was found that feeding 500 ml of milk containing 2×10^6 cells of *L. acidophilus* ml^{-1} day^{-1} for 4 weeks significantly reduced beta-glucuronidase, azoreductase, and nitroreductase activities. When feeding with lactobacilli had ceased, fecal enzyme levels returned to normal after 4 weeks.

Another mechanism for the anticarcinogenic effect refers to nitrosamines (potent carcinogens) which are synthesized *in vivo* from amines and nitrite. The synthesis of nitrosamines may be reduced by enzymatic degradation carried out by some strains of lactobacilli and bifidobacteria. (*See* **Nitrosamines**.)

Apparently, various mechanisms may be involved in the potentially anticarcinogenic effects of lactobacilli and bifidobacteria.

Control of Serum Cholesterol

The influence of gut bacteria on sterol metabolism may be of significance in human nutrition in preventing the accumulation of cholesterol. The feeding of a milk formula supplemented with *L. acidophilus* to infants was shown to result in lower levels of blood cholesterol than those recorded in infants receiving the milk without *L. acidophilus*. Similar results were reported on rats supplementary-fed skim milk fermented with *L. acidophilus*, compared to those fed unfermented skim milk.

Laboratory studies have shown that strains of *L. acidophilus* originating from humans actively assimilate cholesterol, but there are considerable differences in the activity of strains. These results are supported by experiments with pigs using pig isolates of *L. acidophilus*. It was shown that supplementary-feeding *L. acidophilus* to pigs resulted in lower blood cholesterol than in those not receiving *L. acidophilus*.

Strains of *L. casei* and *B. bifidum* were also shown in laboratory studies to assimilate cholesterol to varying degrees. In a study with human subjects it was found that the ingestion of large numbers of *B. bifidum* daily during 6 weeks, significantly reduced the levels of serum cholesterol and serum triglycerides in subjects suffering from elevated serum lipids. This treatment was without effect in subjects with normal values of serum lipids. (*See* **Bile**.)

Stimulation of the Immune System

There are indications that the ingestion of lactic acid bacteria and/or bifidobacteria stimulates the immune response of the host. A significant increase in the number of B lymphocytes and T lymphocytes, was observed, as well as an augmentation of some immunoglobulins in mice given yogurt or other lactic acid bacteria. Also, phagocytic activity of macrophages was shown to be enhanced in mice supplementary-fed with lactic acid-producing bacteria (e.g., *L. acidophilus, L. casei, B. longum*).

Studies with human subjects also showed the immunogenical effects of lactic acid bacteria. For example, (1) a significant increase of phagocytic activity of leukocytes in peripheral blood, and an increase of immunoglobulin A (IgA) were found in subjects given fermented milk containing *L. acidophilus* strain Lal for 3 weeks. This effect was observed at least 6 weeks after termination of the consumption fermented milk; (2) a significant increase in numbers of B lymphocytes and natural killer cells, the augmentation of IgG, and an increase of the serum levels of gamma-interferon were observed in subjects given large numbers of yogurt bacteria for 28 days.

The immunogenic effects of lactic acid bacteria and/or bifidobacteria may be influenced by: (1) the bacterial strain characteristics; (2) the numbers of culture bacteria ingested; and (3) a duration of ingesting culture bacteria. Further research is needed.

Selection of Microbial Strains

Development of improved microbial supplements involves selection of proper strains of the organism. If the ingested bacteria are to survive, they must be resistant to all antimicrobial factors which exist in the gut. Gastric acidity is an important barrier to gut colonization, but some protection is afforded by the buffering effect of the food and food boli will tend to raise the pH of the stomach. Various strains of lactobacilli and bifidobacteria were shown to differ significantly in acid tolerance. Other factors affecting bacterial survival in the gut include bile acids, lysozyme, and organic acids. The ability of bacteria to adhere to the gut epithelium is also an important factor. Adhesion is host-specific, e.g., only lactobacilli and bifidobacteria originating from the human gut have the potential to attach to human epithelial cells. The adhesion ability varies between bacterial species, and even different strains of the same species show variations. It was also shown that growth in milk enhances the ability of lactobacilli to adhere to epithelial cells.

Various strains of lactobacilli and bifidobacteria were shown to differ significantly in the antagonistic effect against harmful bacteria. Other desirable properties include immunogenic effects, assimilation of cholesterol, and hydrolysis of lactose.

The selection of probiotic strains may include the following criteria: (1) not pathogenic; (2) of human intestinal origin; (3) survival of intestinal passage; (4) adhesion to epithelial cells; (5) clinically tested for the specific therapeutic effect; and (6) applicable for manufacturing food products.

Evidence of Influence of Ingested Bacteria

The maximum effect of ingested probiotic bacteria is obtained when the protective indigenous flora have been changed by diet, antibiotic therapy, disease, or some other factors, as the following examples illustrate.

1. A disturbed balance of the gut flora in 28 patients with leukemia was shown to be improved by ingesting 2×10^9 bifidobacteria and 2×10^9 *L. acidophilus* in 200 ml of milk daily for 3 months. The counts of *Klebsiella* spp., *Proteus* spp., *Candida* spp., and *Pseudomonas* spp. in the feces decreased, and the levels of urine indican and blood endotoxin reduced, compared to patients not receiving supplements.
2. Fifteen patients (mean 2.5 years) received antibiotics, such as cephema, penicillins, and aminoglycosides, for curing septicemia and respiratory tract infections. During treatment, diarrhea occurred and lasted for 1–10 weeks. The counts of bifidobacteria decreased in the feces, and those of *Candida* and *Enterococcus* increased. The ingestion of bifidobacteria and *L. acidophilus* improved the stool frequency within 3–7 days, along with the restoration of the gut flora.

The beneficial effects of probiotics may be enhanced by including special ingredients (prebiotics), and thus improving host health.

Probiotic bacteria can affect the composition of the gut flora and/or its metabolic activity during the period of ingestion and for some time after. For example:

1. Milk fermented with *L. acidophilus* fed to human volunteers reduced the *E. coli* count and increased the lactobacillus count, but when the supplement ceased, the bacterial counts returned to normal after 9 days.
2. Five healthy volunteers ingested 3×10^9 *B. longum* daily for 5 weeks; during feeding, counts of bifidobacteria increased and those of clostridia decreased; in addition, ammonia levels and beta-glucuronidase activity decreased in the feces and serum.

3. As already mentioned, feeding *L. acidophilus* to humans reduced beta-glucoronidase, azoreductase, and nitroreductase activities, but when the supplement ceased, enzyme levels returned to normal after 4 weeks.

It is difficult to achieve a permanent colonization of ingested probiotic bacteria. However, their ingestion continuously or at regular intervals may insure the desired benefits.

See also: **Bile**; **Cancer**: Diet in Cancer Prevention; **Carbohydrates**: Classification and Properties; Metabolism of Sugars; **Cholesterol**: Factors Determining Blood Cholesterol Levels; **Dietary Fiber**: Properties and Sources; Physiological Effects; **Immunology of Food**; **Infants**: Breast- and Bottle-feeding; **Liver**: Nutritional Management of Liver and Biliary Disorders; **Nitrosamines**; **Protein**: Synthesis and Turnover

Further Reading

Alm L (1991) The therapeutic effects of various cultures – an overview. In: Robinson RK (ed.) *Therapeutic Properties of Fermented Milks*, pp. 45–64. Essex: Elsevier Applied Science.

Fuller R (1992) *Probiotics. The Scientific Basis*. London: Chapman & Hall.

Gilliland SE (1989) Acidophilus milk products: a review of potential benefits to consumers. *Journal of Dairy Science* 72: 2483–2494.

Havenaar R and Huis in't Veld JHJ (1992) Probiotcs: a general view. In: Wood BJB (ed.) *The Lactic Acid Bacteria: Vol. 1. The Lactic Acid Bacteria in Health and Disease*, pp. 151–170. Essex: Elsevier Applied Science.

Mitsuoka T (1989) Bifidobacterium microecology. In: *Les Laits Fermentés*, pp. 41–48. Paris: John Libbey.

Mollet B, Donnet A, Neeser JR *et al.* (1997) Effects of probiotics on the immune system. In: *Probiotics – Facts and Opinions*, pp. 35–40. Bern: Verlag Schweizerische Milchkommission.

Playne MJ and Crittenden R (1996) Commercially available oligosaccharides. In: *Bulletin IDF* 313, pp. 10–22. Brussels: International Dairy Federation.

Rašić JLJ and Kurmann JA (1983) *Bifidobacteria and their Role*, p. 87 Basel: Birkhäuser Verlag.

Savage DC (1977) Interaction between the host and its microbes. In: Clarke BTJ, Bauchop T (eds) *Microbial Ecology of the Gut*, pp. 277–310. London: Academic Press.

Sellars RL (1991) Acidophilus products. In: Robinson RK (ed) *Therapeutic Properties of Fermented Milks*, pp. 81–116. Essex: Elsevier Applied Science.

Simone CD, Bianchi Salvatori B, Jirilo E, Baldinelli L, Fabio SD and Vesely R (1989) Yogurt and the immune response. In: *Les Laits Fermentés*, pp. 63–67. Paris: John Libbey.

Microorganisms *See* ***Aeromonas***; **Antibiotic-resistant Bacteria**; ***Bacillus***: Occurrence; Detection; Food Poisoning; ***Bifidobacteria* in Foods**; ***Campylobacter***: Properties and Occurrence; Detection; Campylobacteriosis; ***Clostridium***: Occurrence of *Clostridium perfringens*; Detection of *Clostridium perfringens*; Food Poisoning by *Clostridium perfringens*; Occurrence of *Clostridium botulinum*; Botulism; ***Escherichia coli***: Occurrence; Detection; Food Poisoning; Occurrence and Epidemiology of Species other than *Escherichia coli*; Food Poisoning by Species other than *Escherichia coli*; **Lactic Acid Bacteria**; **Microbiology**: Classification of Microorganisms; Detection of Foodborne Pathogens and their Toxins; ***Mycobacteria***; ***Salmonella***: Properties and Occurrence; Detection; Salmonellosis; ***Shigella***; ***Staphylococcus***: Properties and Occurrence; Detection; Food Poisoning; **Vibrios**: *Vibrio cholerae*; *Vibrio parahaemolyticus*; *Vibrio vulnificus*; **Viruses**; **Yeasts**; ***Yersinia enterocolitica***: Properties and Occurrence; Detection and Treatment; **Zoonoses**

MICROSCOPY

Contents

Light Microscopy and Histochemical Methods

J H Holgate and J Webb, Reading Scientific Services Ltd, Whiteknights, Reading, UK

Background

Light microscopy has been in use in the food industry since the first instruments were introduced. In the early years, its primary application was for detecting adulteration of foodstuffs. Identification was dependent primarily upon the recognition of structural features combined with the reaction to selected microchemical tests. (*See* **Adulteration of Foods**: Detection.) Today, a wide range of routine and research microscopes and microscopy techniques are available, providing the food scientist with microstructural information on the nature and form of individual food components and on their distribution within a product. This allows the role and interactions of food ingredients to be examined and provides information valuable to the understanding of the properties of finished products, cooking and manufacturing processes, and the influence of both the type and form of ingredients. Electron microscopy techniques complement this approach, allowing the microstructure of the same types of food materials to be examined at greater magnifications and with different types of preparation procedures. (*See* **Analysis of Food**.) The light microscope is probably one of the most versatile laboratory instruments available to the food scientist. A basic instrument, available today for the same price as a good-quality balance, provides a wide range of magnifications and other facilities to provide structural information on almost any form of food sample.

Principles

The light microscope is an instrument for visualizing fine detail of an object. It does this by creating a magnified image through the use of a series of glass lenses, which first focus a beam of light onto or through an object, and convex objective lenses to enlarge the image formed. In the majority of light microscopes, the image is viewed directly through binocular eyepieces that act as a secondary lens in the form of a magnifying glass to observe the projected image. Such instruments are termed 'compound microscopes,' and the total magnification is the sum of the objective magnification and the eyepiece magnification. The magnification range extends from $\times 10$ to $\times 1000$, with a resolving power of the order of $0.2\,\mu m$, depending on the type and numerical aperture (area available for passage of light) of the objective lenses. A number of books are available, providing comprehensive details on the theory of the light microscope and guidance to the practical use of the instrument, including methods of image enhancement and instrument care. The reader is referred to these, in particular an extensive series of handbooks published by the Royal Microscopical Society, for further information.

Microscopy Techniques

Microscopy techniques for use in the food industry are frequently a combination of those used in biological and materials sciences. Samples are viewed using either transmitted or incident (reflected) illumination. Most frequently, transmitted light is used where a beam of light is allowed to pass through a relatively thin object. The microstructural detail becomes visible due to the absorption of a portion of the light. Incident light is used mainly for the examination of solid, opaque objects, although it now forms the basis for all modern fluorescent microscopes, referred to as 'epifluorescence.' (*See* **Spectroscopy**: Fluorescence.)

The light source is often a single or multicoil tungsten lamp or quartz halogen lamp. However, xenon arc or mercury vapor lamps provide more intense beams, as required for fluorescence microscopy.

Contrasting Techniques

Though many structures are visible by standard (bright field) transmitted light, there are components

that are nonabsorbent and appear transparent or lack any contrast. This can be overcome, or additional information may be gained by using one of several contrasting techniques available that will introduce or enhance image contrast, frequently with minimal disruption of the specimen. The most frequently used are phase contrast, differential interference contrast, and polarized light.

When using phase contrast or differential interference contrast (Nomarski) optics, for example, the phase of part of the light is altered, and this is then recombined with light that has passed through the specimen and thus yields improved differentiation within a specimen. Phase contrast images are characterized by enhanced contrast and visibility of unstained materials and are used routinely for examination of microorganisms isolated from food. Interference contrast provides a distinct relief appearance, with a shallow depth of focus often giving the appearance of a three-dimensional image. It has proved particularly valuable for examining small particles and emulsion droplets (Figure 1).

A different type of contrast is generated in certain samples when they are viewed under crossed polars. Plane polarized light (light vibrating in only a single plane) is generated using a polarized filter and allowed to impinge upon the specimen. The light then travels through a second polarization filter (the analyzer), which is orientated at 90° to the first, effectively blocking the uninterrupted plane polarized light. Any material that is anisotropic or birefringent, e.g., crystals, is capable of rotating the light plane so that the polarization of emerging light will be altered and partially extinguished and will have an orientation that will pass through the analyzer filter. The resultant image is of bright features against a black background.

The use of polarized light has many applications in the study of food. Food starches and many crystalline food ingredients, e.g., sugar, have characteristic sizes, shapes, etc., that can be observed using polarized light. Intact starch granules are strongly birefringent but lose this property during gelatinization (e.g., cooking), and polarized light microscopy is frequently used to follow this process during, for example, the baking of bread doughs. (*See* **Starch**: Structure, Properties, and Determination.)

Fluorescence

Irradiation of samples with light of specific short wavelengths can result in reemission of this energy as light of longer wavelengths. This is referred to as 'fluorescence.' The emitted light, different in color from the excitation light, may be separated by the use of specific filters. Some components of food, for example collagen and lignin, fluoresce naturally after excitation with light (autofluorescence). However, such fluorescence is often very weak, and it is more common for selective, strongly fluorescent dyes (fluorochromes) to be used to produce a high contrast within a sample (Figure 2). (*See* **Lignin**.)

Almost all modern fluorescence microscopes use epifluorescence, in which the specimen is illuminated by a high-density mercury vapor source (via a beam splitter through the objective) to excite only the surface layers of the sample.

The technique has particular value in food microscopy as it can be used to detect substances at low concentrations and for rapid sample screening at low magnification where the particles of interest are in localized regions within the sample. Dyes in common use in food microscopy are Acridine orange for staining bacteria and milk proteins and Nile blue for fats. However, the availability of fluorescent dyes

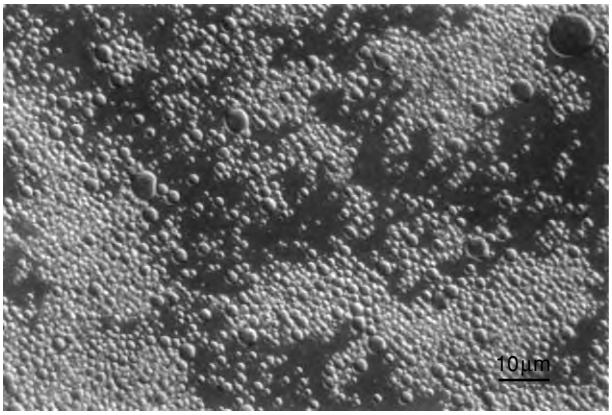

Figure 1 Emulsified droplets of fat in a salad cream. Preparation viewed using Nomarski interference contrast.

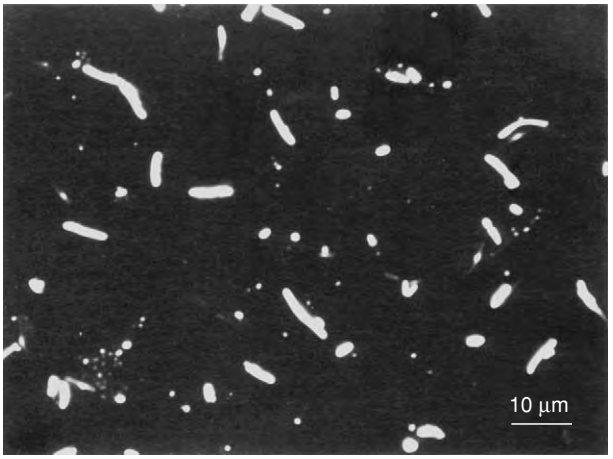

Figure 2 Bacteria isolated from milk. Preparation is stained with Acridine orange and viewed using epifluorescence.

is steadily increasing, especially in combination with highly specific labeling agents such as antibodies (immunolabeling) and lectins, and is likely to be one of the main growth areas within food light microscopy over the coming years.

Sample Preparation

The majority of food materials and their ingredients require some kind of preparation before the microstructure can be examined by light microscopy. The choice of sample preparation will depend on the material being examined and the kind of information required. It may range from simple whole mounts and smears, to more complex, sectioned preparations.

Often, food samples or their ingredients, e.g., milk powders, can be examined simply by direct viewing on a glass microscope slide; alternatively, further valuable information can be gained after supporting the sample in a transparent mounting medium. Such media are frequently inert as well as transparent and include water, liquid paraffin, glycerol, chloral hydrate (used to 'clear' plant cells), and many commercial mountants. The choice will depend on the nature of the material to ensure that little or no change, for example dissolution or swelling, occurs prior to examination.

Liquid or semisolid foods, for example viscous pastes, fats and emulsions, are often viewed as 'smears' across a slide. However, it is important to be aware of the potential disruption of the microstructure during this type of preparation and the differences in detail provided by samples of different thickness. Contrast is often enhanced by phase or interference contrast. The more traditional preparation for light microscopy, and one frequently used for food samples, is the preparation of thin slices or sections through the material. Such sections are mounted on to slides, their contrast often enhanced by multiple staining before examination by bright-field illumination.

Sectioning

Sectioning techniques for light microscopy range from the relatively simple freehand sectioning with a knife or razor blade to the precision cutting of embedded material using sophisticated bench microtomes or ultramicrotomes. A few foods can be sectioned with little or no prepreparation, and it is still commonplace in a food laboratory to see initial freehand sectioning. A bench microtome, however, enables ribbons of sections of more uniform size and thickness (5–50 μm) to be cut. Stainless-steel knives are used most frequently, although more resilient materials, such as tungsten carbide, are often used to cut hard materials and provide longer-lasting blades. Alternatively, freshly cleaved glass knives, used routinely for ultrathin sectioning for transmission electron microscopy, are used to cut thinner sections (< 5 μm) for light microscopy. Food plant material, particularly dried tissue and some solid composite foods, can be sectioned successfully with no prior preparation or treatment. However, the majority of foods and their ingredients require some preparation prior to sectioning to ensure that they are rigid enough to enable sections to be cut with minimum distortion. Wax or resin embedding and low-temperature 'cryofixation' are the most frequently used preparative techniques and have been used in food microscopy for many years.

Fixation/embedding

Whatever the embedding media used, the food material needs to be chemically preserved or 'fixed' prior to infiltration and embedding with the wax or polymer. Successful fixation depends on the type of food material being prepared, in particular the ease of penetration of the fixative into the material. The use of traditional aldehyde fixatives (e.g., formaldehyde or glutaraldehyde) is common, as is the use of osmium tetroxide, in liquid or vapor form, particularly with fats present. However, many different fixation procedures have now been adapted and developed for specific foods in order to achieve optimum preservation of individual components. These include a number developed more specifically for electron microscopy. After fixation, the specimen is usually dehydrated with solvent prior to being infiltrated and then embedded in a wax or polymer. Solidification or polymerization of the embedding media produces a hard, rigid preparation from which sections can be cut. Wax embedding remains one of the cheapest preparation procedures for light microscopy; the procedure can be readily automated for routine use, and considerable improvements have been made to the waxes available. However, it is considered an awkward, somewhat time-consuming and messy procedure and is increasingly being replaced with resins and other polymeric materials for preparing food samples. A wide range of different types of resins and polymers are now available, providing high flexibility in viscosity and conditions for polymerization, e.g., temperature and time and including some acrylic polymers that provide low-temperature hardening (often using ultraviolet light polymerization). The use of the newer resins for embedding samples generally offers a significantly

improved safety in the form of the reduced hazards of the solvents used compared with wax embedding and in the handling of the resins compared with the previously used epoxy resins, etc.

Cryostat Sectioning

Cryofixation of samples, followed by 'cryostat' low-temperature sectioning, has the significant advantage of speed. Samples are supported (using a gel or commercial preparation such as Tissue-Tek) on metal blocks (often after an initial chemical fixation), and then fast-frozen, generally in liquid nitrogen (−196 °C) to preserve or 'cryofix' their structure. The frozen block is sectioned at approximately −20 to −30 °C using a conventional bench microtome housed in a refrigerated unit (cryostat). The technique obviously has particular application for frozen foods such as icecream, but is also used frequently for foods that have high water contents or that are difficult to handle in any other way. However, some food products are unsuitable for sectioning when frozen, for example, if they remain soft or become brittle. Also, the initial capital expenditure can be high and, although the sections can be preserved and permanently mounted, the frozen block can only be kept if stored under liquid nitrogen. Sections are collected on to glass slides, sometimes aided by a thin coating/layer of adhesive such as egg albumen, and allowed to warm to room temperature. They are often further chemically fixed at this stage prior to staining.

Staining Methods

Other than using the optical contrasting techniques described above, the use of dyes to stain components specifically is the most common way of increasing contrast for light microscopy. Staining techniques rely on the interaction or absorption of the dyes with the sample components. Many different stains are now available, and procedures have been developed for food and food ingredients in order to color specific individual components and thus produce enhanced contrast and enable identification of protein, fat, starch, and sugar, for example, (*See* **Protein**:, Determination and Characterization; **Starch**: Structure, Properties, and Determination.)

Frequently, two or more stains can be used in a single preparation. Within food microscopy, many different stains are now readily available for the differentiation of different types of protein, carbohydrate, and fats. Some of these are available in vapor form, e.g., iodine for starch and osmium tetroxide for fats, which has a particular advantage in causing little disruption to the specimen. The selection of stains and the procedures used will vary, depending not only on the type of food material under examination but also on the sample preparation technique employed and the exact information required. However, stains are still often classed according to the nature of their binding to the specimen. Acidic or basic dyes are used for attaching to positively or negatively charged sites, respectively, and thus can be used, for example, to distinguish between different types of proteins. Solubility dyes rely on the dye being more soluble in one component than another. The Sudan family of stains use this property to stain fats in food.

Stains are described as 'histochemical' when a specific chemical reaction is used to attach the chromophore to a particular chemical grouping in the specimen. The most frequently used is the periodic acid Schiff (PAS) reaction used to demonstrate the presence of different polysaccharides by the formation of aldehyde groups following oxidation.

Applications of Light Microscopy

The versatility of the preparation procedures and techniques for light microscopy makes them applicable to a wide variety of different foods and ingredients. Selected samples are presented below, although many more are presented in other parts of the Encyclopedia as well as numerous food microscopy publications.

Powders

A wide range of food materials, in particular raw ingredients, exist in powdered form. Common examples include ground spices, spray-dried flavors and milk powders, and freeze-dried beverages and proteins. Light microscopy provides a way of examining the size, shape, and often the internal structure of individual powder particles. Liquid paraffin is frequently used as a simple mountant, although glycerol is more successful when fat is present. Examination under polarized light shows the presence of crystalline or other birefringent structures, enabling their size and location to be determined. The presence of lactose crystals, for example, is often demonstrated by their characteristic tomahawk shape when viewed under polarized light (Figure 3). In more complex powder mixtures, individual components can also be discriminated by selective staining. The use of an iodine/potassium iodide solution or simply iodine vapor will stain starch a distinctive blue/black color, whereas fat appears pale brown, and proteins appear yellow.

Animal-based Products

Many meat and comminuted meat products are readily prepared using cryostat sectioning. A variety of staining procedures are available for identifying

individual components including fats, muscle, connective tissue, and bone. These range from single-step stain mountants, for example, Toluidine blue, to more complex sequential staining with three or four different solutions (e.g., the Picro–Mallory technique). Polarized light can further complement the use of stains in order to locate crystalline fatty regions and connective tissues and to differentiate fresh and processed muscle tissue.

Fruit and Vegetables

Many cellular structures are visible using phase contrast, but more detailed information on changes to cell wall and cell contents can be obtained following staining for cellulose, pectins, etc. The understanding of the effects of processing of fruit and vegetables by brining, freezing, cooking, etc. has been greatly facilitated by microscopical studies on samples taken at different stages of such processes (Figure 4).

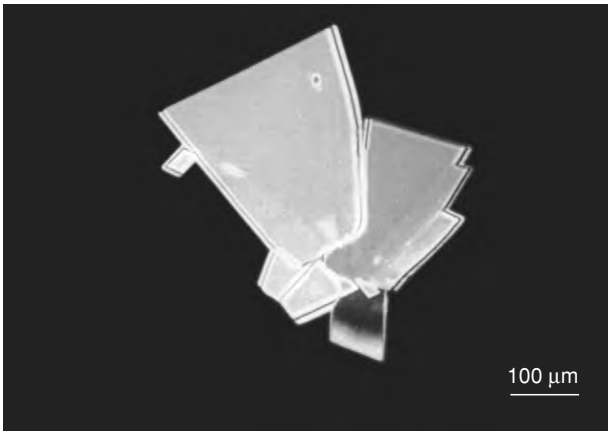

Figure 3 Characteristic tomahawk crystals of lactose viewed under crossed polars.

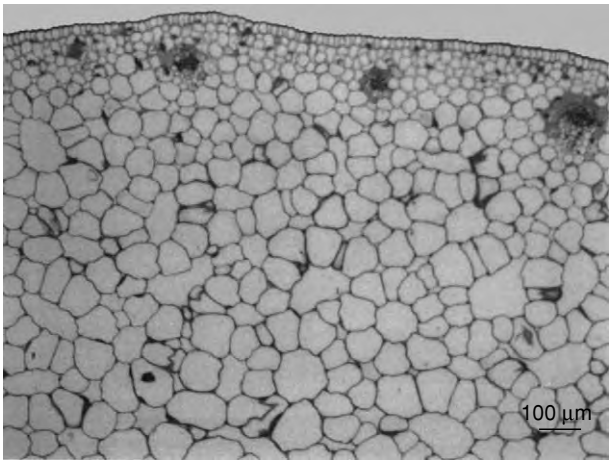

Figure 4 Section through a fresh onion shoot; Toluidine blue staining.

Bakery Products

A combination of preparation and observation techniques has been important in examining the complex microstructure of many bakery products, e.g., bread, cakes, and biscuits. Such techniques have enabled the changes in ingredients, especially starch and protein, to be followed and correlations to be made between the microstructure and the physical and sensory (e.g., texture) properties of the finished products. One of the key changes in baking is the gelatinization of starch, which can be followed successfully through examination of preparations using polarized light.

Photomicrography

The recording of images obtained by light microscopy is an important area in its own right and thus, unfortunately, beyond the scope of this article. The advent of high-resolution digital cameras has facilitated image acquisition, enabling rapid recording of images, and, where the investment has been made, even short 'videos' can be made, e.g., for studying the dissolution of particles. As always, the food microscopist still needs to experiment to ensure that high-quality images observed through binocular eyepieces are permanently recorded for future reference.

Developing Areas within Light Microscopy

The two areas of most recent development in the field of light microscopy have been the use of specific labeling techniques, in particular immunolabeling, and the availability of confocal instruments. In the former, labeled antibodies are used as specific reagents for the detection/staining of specific specimen components. Labeling is most frequently achieved by fluorescent dyes. Although antibodies have been raised for many biological and biomedical applications, they are currently limited for foods mainly because the antigenic sites are modified as a result of food processing and manufacture.

In a confocal microscope, a confocal image is built up point by point by scanning a sample with a focused beam of laser light. It is a noninvasive technique enabling optical sectioning of a sample and producing images with well-defined focal planes but without the problems of out-of-focus blur. Combined with current developments in image-processing techniques, confocal microscopy is being applied widely in non-food areas and is beginning to be applied to food, and has considerable potential for the future.

Light microscopy offers versatility in the number of optical techniques available for examining food materials, combined with ease of preparation and

operation at a relatively low cost. Although a decline in its routine use was evident after the introduction of electron microscopes, it has reemerged as an invaluable instrument for investigations into the science of food and food processing.

See also: **Adulteration of Foods**: Detection; **Analysis of Food**; **Lignin**; **Protein**: Determination and Characterization; **Spectroscopy**: Fluorescence; **Starch**: Structure, Properties, and Determination

Further Reading

Bradbury S (1982) An introduction to the optical microscope. *Royal Microscopical Society Microscopy Handbook No. 01*. Oxford: Oxford University Press.

Cook HC (1974) *Manual of Histochemical Demonstration Techniques*. Oxford: Butterworths.

Flint O (1994) Food microscopy. *Royal Microscopical Society Microscopy Handbook No. 30*. Oxford: BIOS Scientific Publishers.

Hartley WG (1979) *Hartley's Microscopy*. Oxford: Senecio.

Heathcock JF (1988) The use of light microscopy and electron microscopy in the food industry. *Microscopy and Analysis* 2: 81–90.

Heathcock JF and Holgate JH (1990) *The Continued Importance of Light Microscopy as a Complementary Technique to EM in the Elucidation of Food Microstructure*. Proceedings of RMS Conference Micro 90, 20, pp. 713–718.

Ploem JS and Tanke HJ (1986) Introduction to fluorescence microscopy. *Royal Microscopical Society Microscopy Handbook No. 10*. Oxford: Oxford University Press.

Vaughan JG (ed.) (1979) *Food Microscopy*. London: Academic Press.

Winton AL and Winton KB (1935) *The Structure and Composition of Foods*, vols. 1 and 2. New York: John Wiley.

Scanning Electron Microscopy

J Webb and J H Holgate, Reading Scientific Services Ltd, Whiteknights, Reading, UK

Introduction

A scanning electron microscope uses a finely focused beam of electrons to reveal the detailed surface characteristics of a specimen and provide information relating to its three-dimensional structure. It also has a particular advantage of providing great depth of field. With the introduction of the first commercial instruments in the mid-1960s the scanning electron microscope provided the link between the magnification ranges offered by light microscopy and the higher resolving capability of the transmission electron microscope (TEM). Nowadays, following development of electron guns, electromagnetic lens systems and vacuum systems, scanning electron microscopes can have a resolution of 1 nm. Therefore, there is now considerable overlap between the resolving power of the two forms of electron microscope available, although the techniques are complementary and different types of microstructural information are obtained in each case.

Food and its ingredients have been examined by scanning electron microscopy (SEM) ever since the first instrument was available. In the early days the main applications were in studies on the microstructural characterization of dry foods, in particular powders. However, with the development of a range of different preparation techniques able to handle many different types of foods, almost every food material has now been examined in some way by SEM. (*See* **Analysis of Food**.)

Principles

When a fine beam of electrons is focused on to the surface of a specimen, different interactions occur, including the emission of secondary and backscattered primary electrons. If these are collected and amplified they can be used to create an image corresponding to the surface topography of the specimen. The electron beam is scanned across the specimen repeatedly in a raster pattern, which is synchronized with the scan of a cathode ray tube such that the image is presented in a digitized form built up on a TV monitor. Magnification is achieved through the electron beam scanning an increasing smaller area of the same specimen and most modern scanning electron microscopes have a magnification range of $\times$ 7 20, up to more than $\times$ 300 000, and resolution in the order of 2–4 nm or better. The great depth of focus of the microscope (> 500 $\times$ that of light microscopy) is achieved by the convergence angle of the primary electron beam and the relatively long working distance between the final lens and the specimen. This depth of focus enables production of images that appear to be three-dimensional.

As with transmission electron microscopes, conventional instruments must be operated under high vacuum in order that the electron beam can travel to the specimen surface without collision with gas molecules. However, the introduction of specialized vacuum systems with pressure-limiting apertures brings a new perspective to SEM and enables samples

to be examined at higher pressures. In conventional SEMs, the source of electrons is generally a heated tungsten filament, which provides ease of operation, low cost, and simplicity. More intense sources include lanthanum hexaboride (LaB6) tips and field emission guns and these are used where greater brightness is required. Accelerating voltages used typically range from less than 1 kV up to 40 kV. The higher voltages can provide greater resolving power, but at the risk of excessive penetration into many specimens and damage by the electron beam. Lower voltages are used to image fine structure on sample surfaces.

The generation of an image from secondary electrons remains the most frequently used form of imaging of biological samples within a conventional SEM (CSEM). However, the interaction of a primary electron beam with a specimen surface can also provide other information, for example, from backscattered or Auger electrons and X-rays.

Specimen Preparation

In its simplest form, preparation for SEM involves securing a specimen on to a metal support 'stub' and, if the sample material is nonconducting, coating the surface with a conducting thin layer of metal. The requirement for conductivity is to prevent the build-up of electrical charge on the specimen surface once it is bombarded with electrons within the instrument and also to enhance the secondary image. Many different procedures have been developed for securing dry food specimens on SEM stubs. These vary depending on the size, shape, and composition of the specimen and the type of information required. One of the most frequently used support media employs a colloidal silver cement (silver DAG), which not only holds the specimen firmly but improves conductivity. However, others include carbon cement (carbon mixed with an acrylic carrier), glues, gels, and adhesive tape.

Metal Coating

In the early days of SEM, surface metal coats were prepared by resistive evaporation in a high-vacuum coating unit. However, this procedure often caused localized heating of the specimen surface and, as the metal was applied unidirectionally, it did not always cover all the crevices in the specimen. It was soon superseded by the introduction of diode sputter coating, which is carried out under a lower vacuum. The coating material forms the cathode and the specimen the anode. Sputtering is carried out in an argon plasma generating a multidirectional spray of metal particles that carry into crevices and around corners. Relatively inexpensive sputter coaters are now available, which provide continuous, uniformly thin metal coats of controllable thickness, often in automated units. Such instruments deposit a metallic layer with a thickness typically of 10 nm and a grain size of approximately 2–5 nm to achieve a continuous film. Thicker films, up to 20 nm, are used for specimens with particularly irregular surfaces.

For conventional SEM, the most frequently used metal has been gold, although others used include silver, platinum, palladium, tungsten, chromium, or mixtures of two of these. Chromium is now being used increasingly with the requirement for thinner and finer layers which do not mask the fine detail of the specimen and provide ultrahigh-resolution images. It is deposited under a higher vacuum to achieve a finer grain size (0.5 nm) and a continuous, thin (< 5 nm) layer.

Drying Techniques

Many food products and ingredients contain relatively high levels of water. In the virgin state, such products are unsuitable for examination in a conventional SEM, as the high vacuum system causes rapid desiccation and consequential damage to the fine structure of the sample. For these types of specimens, examination by conventional SEM has been carried out successfully following some form of drying step. In many cases, preparation has involved a brief chemical fixation to stabilize the components of the sample *in situ* (for example, with glutaraldehyde, which cross-links proteins) before solvent dehydration and drying using critical-point drying. Alternatively freeze-drying has been used.

Water needs to be removed with great care in order to minimize shrinkage or other artifacts and a great deal of work has been undertaken to optimize the procedures. Critical-point drying involves the replacement of the organic dehydration fluid in the specimen with liquid carbon dioxide and subsequent conversion of the CO_2 into gas in a chamber pressurized to the critical point and temperature where liquid and gaseous CO_2 are in equilibrium. After gaseous CO_2 is released, the specimen is dry without having passed through any phase boundary. It has been shown to be a particularly useful technique for single-cell preparations (**Figure 1**) and food with high tissue contents such as meat, fish, and vegetables. However, the technique has now been shown to cause shrinkage of greater than 30% in some samples and to generate many surface distortions and is now losing favor as a routine SEM procedure. Furthermore, any solvent-extractable components are lost during the dehydration process.

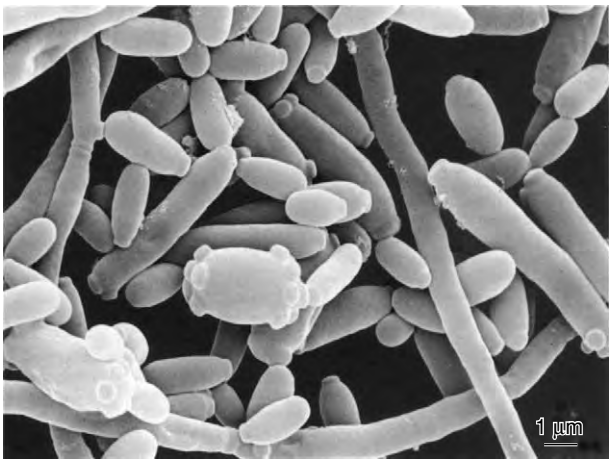

Figure 1 Spoilage yeast (*Brettanomyces* sp.) isolated from a soft drink and prepared using critical-point drying. Bar = 1 µm. Reproduced from Microscopy: Scanning Electron Microscopy. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

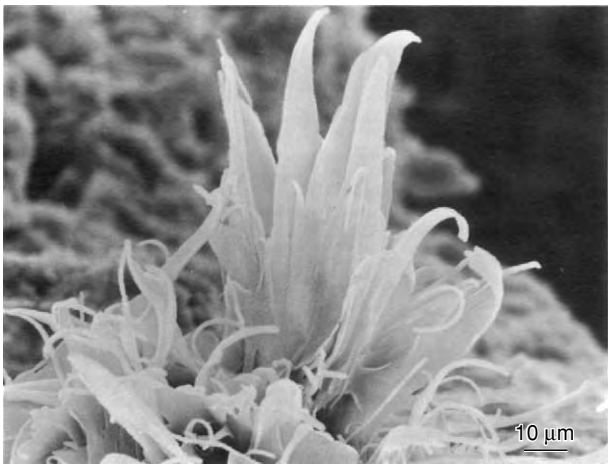

Figure 2 Fat bloom crystals (form VI) on the surface of a plain chocolate bar examined using low-temperature scanning electron microscopy. Bar = 10 µm. Reproduced from Microscopy: Scanning Electron Microscopy. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

In freeze-drying, wet samples are fast-frozen, transferred to a freeze-drying unit where a low temperature is maintained (typically $-80\,^{\circ}C$ to $-90\,^{\circ}C$) and the ice is sublimed slowly under low vacuum. Careful control over the different stages, including initial freezing, temperature, and vacuum of drying and, in particular, of the rate of rewarming, has proved to be important in retaining the microstructural features of food materials.

Unfortunately, these drying techniques are not readily adaptable to fatty foods, because solvent dehydration will remove the fat and freeze-drying will not adequately stabilize it. Many investigators, however, have found the techniques successful, for example, in the study of foods containing carbohydrate or protein gels and 'fully biological' tissues, including meat, vegetable, and fruit.

Low-temperature SEM (Cryo-SEM)

One way to overcome the problems of conventional drying techniques is to make use of low-temperature preparation procedures to examine frozen, fully hydrated food materials. If the temperature of the sample is sufficiently low (i.e., $< -130\,^{\circ}C$), there will be negligible freeze-drying in the vacuum system of the SEM; typically, samples are held at *c.* $-170\,^{\circ}C$ during imaging. No chemical fixation or solvent dehydration needs to be involved as the sample is 'cryofixed' (physically preserved). The technique has been used widely since the early 1980s in various biological applications, including in the examination of food.

The procedure involves rapid freezing of the specimen before transfer on to the precooled stage of a cryopreparation unit, which may be either dedicated (attached permanently to the SEM) or nondedicated (free-standing) vacuum equipment. The sample is often freeze-fractured in the unit to reveal internal structure. When particularly high levels of water are present, fractured and original surfaces can be 'etched' by raising the specimen temperature while it is still under vacuum so that a thin layer of surface ice is sublimed away to reveal more of the topographical detail.

As with low-temperature preparations for TEM, one of the most crucial stages is in the initial freezing of the specimen. It is desirable that the water present is solidified in a vitreous, noncrystalline state and the remaining components are preserved in the same form as they were originally. However, the relatively large size of samples typically used for SEM prevents vitrification of all but a very thin layer at the surface. In practice, the aim of the freezing process is to render the size of the resulting ice crystals as small as is practicable. Low-temperature SEM is particularly suitable for examination of liquid or semisolid materials such as margarines, low-fat spreads and dressings, as well as foods with high levels of crystalline fat such as chocolate (Figure 2) and frozen foods such as icecream.

Auxiliary Techniques

X-ray Microanalysis

The principle of X-ray microanalysis is related to the emission of X-rays from a specimen through

Figure 3 (see color plate 108) Digital X-ray map showing calcium distribution (of calcium phosphate) in frozen hydrated cheddar cheese.

interaction of the electron beam. When any material with an atomic number of four or more is examined in an electron microscope, X-rays are generated from just below the surface. As each element has a unique X-ray spectrum, the elemental composition can be determined on the basis of detected X-rays. X-ray microanalysis systems are of two types: energy-dispersive X-ray analysis (EDX) and wavelength-dispersive spectroscopy (WDS). EDX is the more frequently used technique for biological and food-type samples. This technique measures the energies of X-rays and provides analysis of all detectable elements simultaneously. By contrast, WDS measures the wavelengths of X-rays with individual crystals, each of which focuses on one element at a time. WDS has a better resolution than EDX but is less easy to use.

In the past, routine EDX detectors have been protected by beryllium windows that have excluded transmission of X-rays with an energy of less than *c.* 1 kV, i.e., the X-ray energies of elements with an atomic number below 11 (sodium) and, traditionally, for the detection of lighter elements, WDS has been used. However, improved detector design and the use of plastic windows has enabled the full range of elements to be detected by routine EDX systems and WDS is used for specialized analyses. Through computer interfacing, the X-ray microanalytical technique can be used to generate quantitative data and produce elemental distribution 'maps' of food materials. X-ray mapping techniques have already provided valuable information relating to chemical changes in food that have been associated with their processing, storage, and packaging (Figure 3).

Cathodoluminescence

Certain materials emit long-wavelength photons in the ultraviolet and visible regions of the spectrum when they are irradiated with a beam of high-energy electrons and this phenomenon is termed 'cathodoluminescence.' Although this technique is widely used for geological applications, with recent developments in detector design it has the potential to be applicable to a wider range of materials, possibly including foods. Novel treatments could include labeling specific structures with reagents that exhibit cathodoluminescence, much in the same way as fluorescent labeling is used for light microscopy.

Image Analysis

SEM produces electronic/digitized images and as such they are available for computer manipulation, for example, in image processing or analysis. Image processing involves computer filtration and enhancements to improve the final image quality from the microscope and this type of facility is often a standard feature on modern instruments. Image analysis is used to obtain quantitative data directly from the images, for example, relating to size and size distribution and shape of features as well as relative areas and coincidence measurements. Many purpose-built image analyzers are available commercially with which analysis of any image source, e.g., objects, photographs, light microscopy, and TEM images, as well as SEM images, can be undertaken. The basis for discrimination of features/areas of black and white images, such as SEM images, is the difference in gray levels (differences in black to white) across the image. SEM images often cause problems because of the very wide range of gray levels within any one image, although use of high-contrasting techniques, in particular backscattered imaging, can help eliminate the problem.

Applications of SEM

Many food laboratories have their own scanning electron microscope or have direct access to one and the use of the technique is commonplace in the majority of investigations on food. Applications are many and diverse, covering practically all of the different foodstuffs and their ingredients. The strength of SEM in food studies is its wide magnification range, large

depth of focus, and its ease of use. The relatively straightforward preparation of dry materials has led to the extensive use of SEM for the characterization of powders, for example, spray-dried ingredients (Figure 4), flour and sugars, as well as dried foods such as coffees and teas. Comparative information is obtained on the gross size and shape of the particles and, by breaking them open, details of the internal microstructure can be determined, for example the form and interaction of individual components, including crystalline inclusions and air bubbles.

SEM images often provide information that can be directly related to textural properties of food in a unique way and can also be related to processing information as well as sensory panel data. Examination of different types of bakery products, for example, can reveal the form of the starch, in particular the intactness of starch granules together with the levels of association of the protein strands (Figure 5). Such detail can often be related to the way in which the product is likely to break down in the mouth and explain terms such as 'crumbly' or 'brittle.' (*See* **Starch**: Structure, Properties, and Determination.)

Many food-manufacturing processes involve production of a number of relatively unstable, liquid, or semisolid intermediate products prior to the finished commercial sample. SEM, frequently in combination with low-temperature preparation techniques, provides a way of monitoring the changes taking place during the development of the final microstructure and helps to investigate the effects of different processing conditions and/or ingredients. Its use has been particularly valuable, for example, following crystallization of key food components such as sugar or fat in different foods, including changes in size and shape of individual crystalline regions.

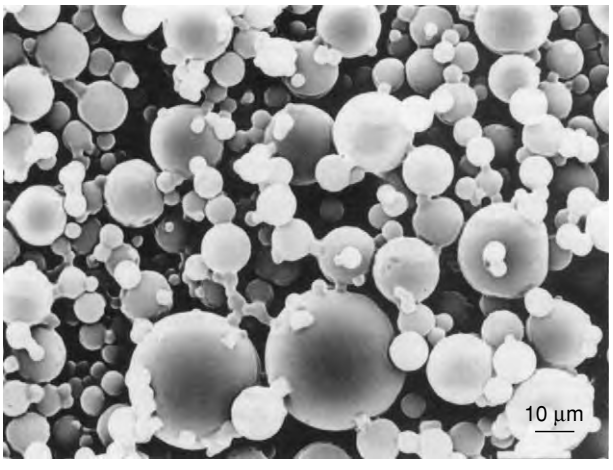

Figure 4 Spray-dried particles of polydextrose. Bar = 10 μm. Reproduced from Microscopy: Scanning Electron Microscopy. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

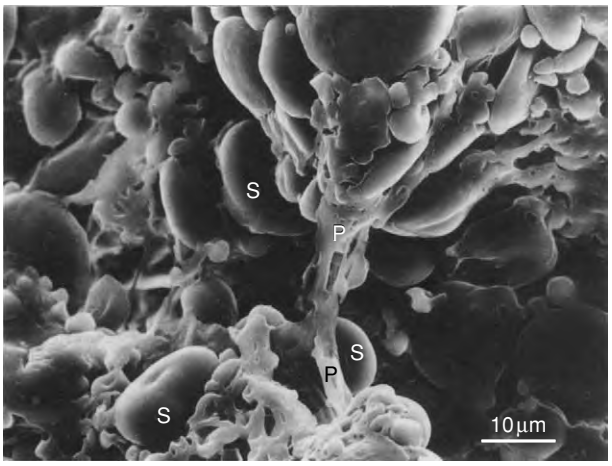

Figure 5 Fracture through the center of a semisweet biscuit showing intact starch granules (S) and fine strands of protein (P). Bar = 10 μm. Reproduced from Microscopy: Scanning Electron Microscopy. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

A significant area of food microscopy is concerned with the identification of contaminants in food, for example, fragments of glass, metal, paints or insects, which are introduced accidentally or deliberately during and after production. The use of SEM, in combination with EDX, provides a rapid way of identifying the composition of contaminants, which can be quite small (< 1 mm). The technique is noninvasive and thus the evidence is not destroyed. (*See* **Contamination of Food**.)

Recent Developments and Specialized SEM Techniques

Low-vacuum and Environmental SEM

Although conventional SEMs offering the best resolution operate at high vacuum, they suffer from limitations in that the specimens intended for examination have to be rendered stable, so that they are able to withstand the vacuum, and electrically conductive. Accomplishing this without altering the structure is difficult and, clearly, delicate hydrated or volatile samples often present in food systems are not vacuum-tolerant. Operation of an SEM at higher pressures opens up the possibilities of being able to examine a wide range of vacuum-intolerant specimens without prior preparation. Not

only can hydrated and semisolid samples be examined without prior preparation, the gas ionization that occurs with increasing pressure in the sample chamber effectively neutralizes charge that accumulates on the sample and eliminates charging artifact. The consequence of this is that a conductive layer of metal does not need to be applied to electrically nonconductive samples when they are examined at higher pressures.

SEMs that operate at higher pressures are now commercially available and are termed low-vacuum SEMs (LV-SEMs) or variable-pressure SEMs (VP-SEMs). A further category can provide (in addition to variable pressures) a saturated water vapor environment and this is termed environmental SEMs (ESEMs). All of these achieve their ability to produce a source of illuminating electrons from normal electron gun systems whilst maintaining a sample chamber at low (i.e., poor) vacuum by separating different regions of the instrument by small, pressure-limiting apertures.

Because they lend themselves to the examination of nonconductive samples without the need for coating, LV-SEMs, VP-SEMs, and ESEMs are now increasingly being used to examine different food systems. Examples of specimens suitable for examination with this type of instrument include finished products and samples taken from different stages of food processing as well as for the investigation of complex structures. Novel approaches to the preparation of samples can yield further information, e.g., exposing a lipid-containing matrix, such as a biscuit, to osmium vapor which will bind to the lipid phase preferentially and can then be contrasted by backscattered electron imaging. Further applications include samples immunolabeled with gold or silver probes and also dynamic experiments, such as following the crystallization of products in an ESEM.

Demonstration of elemental composition of unfixed, uncoated samples by X-ray microanalysis is possible, although (because of the higher pressure) the spatial resolution of the X-ray signal is reduced. However, mathematical corrections have been developed by some workers and at least one major manufacturer of X-ray microanalysis systems offers such algorithms in its software.

High-resolution and Low-voltage SEM

Resolution is largely dependent on the diameter of the scanning spot of electrons used to examine the specimen. However, with decreasing spot size there is decreasing signal and, with conventional SEMs, the signal-to-noise ratio becomes unacceptably poor at very high magnifications. With decreasing accelerating voltage there is a concomitant decrease in signal so, traditionally, relatively high (> 20 kV) accelerating voltages have been used for SEM examination of specimens. However, reducing the voltage has a number of advantages. There is a significant reduction in the build-up of charge and, with reduced beam penetration, fine detail on the surface of samples is revealed.

With the increasing availability of instruments equipped with a field emission gun (FEG) and advanced electron optics, the opportunity now exists for the examination of specimens with high-intensity electron beams with small spot sizes over a wide range of accelerating voltages (including very low voltage).

High-resolution cryo-stages have been developed to complement this range of FEG-SEMs so that frozen samples can be examined at high resolution. There is now considerable overlap between the information that can be obtained from fractured frozen samples examined by high-resolution SEM and that obtained from traditional freeze-fracture studies made by TEM. Often, larger areas of a sample are available for examination by SEM, when compared with complementary TEM freeze-fracture studies, and interpretation of the results can be less difficult. However, in both cases, the quality of freezing is paramount and normally it is only a small region close to the edge of a sample that will have been frozen sufficiently rapidly to be comparatively free of ice crystal artifact.

Applications that are particularly suitable for examination with such instrumentation include emulsions, foams, liposomes, microorganisms, and food systems where interfaces are of interest.

There is an additional category of instruments (ESEM-FEG) that exists that combines the ability to operate at high resolution and high gun intensity with the capability of operating at higher pressures and relative humidity. These instruments offer an enormous range of different operating modes to suit different sample types. However, this flexibility comes at a price and the principal limitation is the cost of purchase.

Specimen Preparation

Significant improvements have been made in both instrumentation and electronics in SEM over recent years. In applying the technique more widely to food studies, however, the limitations probably still lie in the area of presentation/preservation of the specimen to insure that the microstructure is retained in its original form when examined by SEM. It is

this area of the technique where most developments are likely, particularly in the use of environmental or LV-SEM or VP-SEM, low-temperature preparation procedures, and also in high-resolution cryogenic techniques.

When the first scanning electron microscopes were available commercially, the appeal of a three-dimensional image led to the rapid production of numerous micrographs of food materials with perhaps only limited scientific value. Now that this period is over, an increasing number of workers are using the fuller potential of SEM and it has emerged as an important research tool both in its own right and as a complementary technique to other microscopical techniques in microstructural investigations of food and food processing.

See also: **Analysis of Food**; **Contamination of Food**; **Starch**: Structure, Properties, and Determination

Further Reading

Cohen SH (1984) Scanning electron microscopy in the analysis of food microstructure: a review. In: Charalambous G (ed) *Analysis of Foods and Beverages: Modern Techniques*, pp. 421–449. Orlando: Academic Press.

Goldstein JI, Newbury DE, Echlin P, Joy DC, Fiori C and Lifshin E (eds) (1992) *Scanning Electron Microscopy and X-ray Microanalysis*. New York: Plenum.

Heathcock JF(1988) Cryotechniques in the examination of food. *Institute of Physics Conference Series* no. 93 3(2): 55–57.

Holcomb DN and Kalab M (eds) (1981) *Studies of Food Microstructure*. Scanning Electron Microscopy. Chicago, USA: AMF O'Hare.

Kalab M (1983) Electron microscopy of foods. In: Peleg M and Bagley EB (eds) *Physical Properties of Food*, pp. 43–94. Westport: AVI.

Sargent JA (1988) Low temperature scanning electron microscopy: advantages and disadvantages. *Scanning Microscopy* 2: 835–949.

Transmission Electron Microscopy

J H Holgate and J Webb, Reading Scientific Services Ltd, Whiteknights, Reading, UK

Introduction

Many investigators have studied the textural behavior of numerous products using physical measurements and showed good correlation with sensory evaluations of texture and mouth feel. Microstructural analysis now seems to be playing a key role in explaining the type of data produced. It is being used increasingly in the study of a wide range of food products and ingredients, with the data obtained being related to a sample's textural properties and in some cases to its perceived flavor. Such studies can help in the overall understanding of the effects of food processing. (*See* **Analysis of Food**; **Sensory Evaluation**: Sensory Characteristics of Human Foods.)

The first commercial transmission electron microscope was available in the 1940s when it was recognized that significant improvements in resolution could be made over the light microscope through the use of an electron beam rather than a light beam. However, the instrument could only operate under vacuum and required ultrathin specimens for the beam to pass through. Thus, specialized preparation techniques had to be developed to insure that microstructural and ultrastructural details of specimens were adequately preserved during each of the preparative stages.

In the early years, the application of transmission electron microscopy (TEM) to the food industry lagged behind that in metallurgical and biological fields. This was partly due to problems in specimen preparation and handling that were specific to food materials, arising from the lack of homogeneity and often the presence of high levels of water, sugars, or fats.

Principles

A hairpin tungsten filament of an electron gun, encased in a column at high vacuum, is heated to generate a cloud of electrons, which is accelerated through a hole in an anode to produce a narrow beam that is directed towards the specimen. By using electromagnetic lenses to focus the resultant beam, an image can be formed by projecting the electron beam through a thin specimen. The first instruments produced magnifications up to $\times$ 10 000 but subsequent development of electromagnetic lens systems and guns capable of accelerating voltages exceeding 100 kV and up to 400 kV on standard TEMs has extended the magnification range up to $\times$ 1 000 000 or more, with resolutions of better than 0.2 nm possible on modern instruments. As the human eye is not sensitive to electrons, the final image is visualized by being focused on to a fluorescent screen or directly on to a photographic plate or film.

TEM operates under high vacuum ($< 10^{-5}$ mm Hg or $< 10^{-4}$ Pa) in order to minimize the scattering of the electron beam by gas molecules and allow it to

travel through the microscope without contamination. This brings with it the physical problems of having a specimen which must withstand exposure to high-vacuum as well as high-voltage electron beams.

Preparation Techniques

A number of different techniques are available for the preparation of specimens for TEM. The choice depends very much on the nature of the food, for example, whether it is liquid or solid or whether it contains high levels of water, fats, or sugars, which may make it difficult to handle during the various preparative stages. In all cases, however, the basic requirement is to produce an ultrathin preparation that allows an electron beam to pass through it, whilst at the same time being dry and resistant to the high vacuum and potential high temperature of a high-voltage beam.

Food often presents particular difficulties because of its heterogeneity and, in many cases, the presence of relatively high levels of fats and sugars, often in combination.

The most frequently used TEM preparation techniques that have been applied fairly successfully to different food materials are negative staining, thin sectioning, and freeze-fracture replication, with the latter two being used most often.

Negative Staining

Negative staining represents one of the easier and more rapid techniques in terms of specimen preparation but is limited in food applications to studies of dilute suspensions or dispersions, for example, food proteins, liposomes, cellular fragments, liquid crystals, or microorganisms. It cannot be used for large or more complex food systems.

In this technique the particulate or colloidal components are examined directly on a TEM specimen support grid after they have been surrounded or embedded in an electron-dense 'stain.' The technique relies on the metallic stain producing an outline of the structures, rather than reacting positively with them, thus providing information on overall size and shape and detail of features such as bacterial flagella. Phosphotungstic acid (PTA) is one of the negative stains used most commonly (**Figure 1**), although some success has been achieved with uranyl acetate, particularly for the examination of food protein macromolecules. A third reagent, ammonium molybdate, has been used, but with less predictable results, and has been shown to be particularly sensitive to localized pH changes.

There was a renewed interest in the use of negative staining in the mid 1980s as it became possible to use it in combination with low-temperature TEM preparation techniques. Fully hydrated, negatively stained preparations are fast-frozen and then examined by TEM on a low-temperature (below $-150\,^{\circ}C$) stage. This approach has been particularly successful in the examination of viruses and single-cell preparations, but has been of limited use to date in studying food components.

Thin Sectioning

Probably the most frequently used preparation procedure for TEM is thin sectioning, where the specimen is initially preserved chemically or 'fixed,' solvent-dehydrated, and then infiltrated and embedded in a resin or other polymeric material. On polymerization (usually by high-temperature curing), the resin hardens so that ultrathin slices or sections (< 100 nm) of the embedded material can be cut using an ultramicrotome.

Among the various chemical fixatives currently in use, glutaraldehyde, followed by osmium tetroxide (O_SO_4), remains one of the most effective. However, a number of different fixative regimes have been developed for different food systems, often involving alternative aldehydes or mixtures of aldehydes and using temperatures, osmolarities, pH values, and buffers appropriate to the food. Osmium textroxide has the advantage of being uniquely suitable for fats and thus has been used particularly for foods containing high levels of fat (e.g., oil seeds, emulsions, chocolate). It also functions in its vapor form and, therefore, is useful as a primary fixative to stabilize the specimen prior to further treatment. As well as fixing the sample, chemical bonding of the osmium to the unsaturated fats produces electron-dense 'stained' regions in the preparations. Extended fixation times are often employed for fatty foods, although these are often a compromise, as they have sometimes been shown to produce artifacts in the nonfat areas of the specimen.

Many different types of polymers have been used as embedding materials over recent years. Epoxy resins and methacrylates are perhaps used most frequently for food materials, particularly lower-viscosity polymers. A number of different resins were developed during the 1970s and 1980s. Initially they offered particular advantages of very low viscosity, and thus more rapid and efficient infiltration, and were thought to be safer to handle. Unfortunately, a number of the components of the resins were still found to be unsafe, having carcinogenic properties in particular. However, a range of new acrylates became available in the 1990s, which are not only considered safer to handle

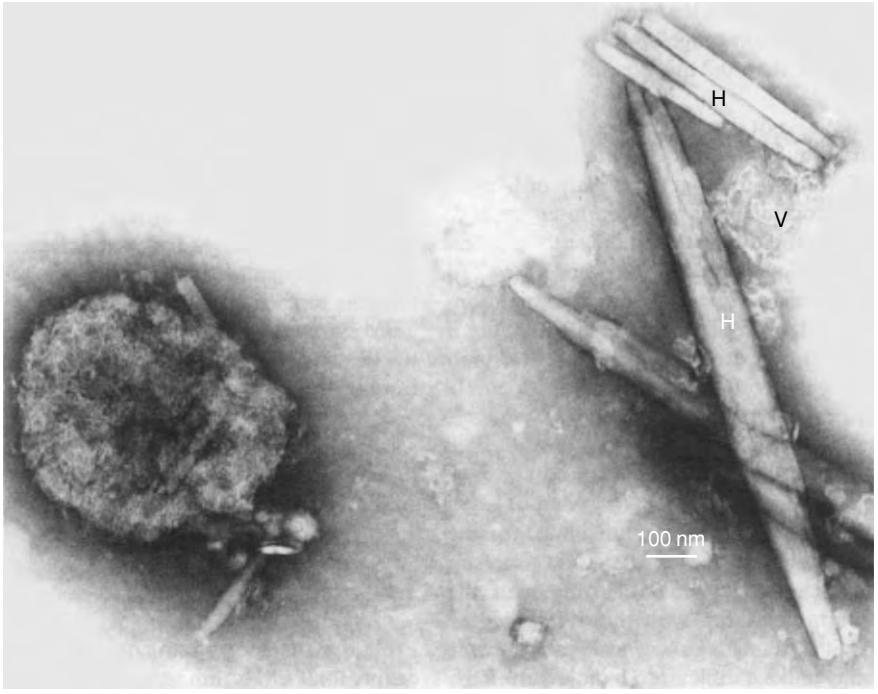

Figure 1 Negatively stained preparation of material isolated from a comminuted fruit drink showing the presence of hesperidin crystals (H), chromoplasts (C), and membraneous vesicles (V). Bar = 100 nm. Reproduced from Microscopy: Transmission Electron Microscopy, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

but have relatively low viscosities. Some of these can be polymerized at low temperatures by ultraviolet curing, though care has to be taken to prevent localized heat build-up as the accelerator is activated. Thus, the problems of exposing materials to high-temperature polymerization conditions are avoided. Other polymeric materials available are miscible with water and thus eliminate the solvent dehydration stages. However, these have been of only limited use as they often produce fairly brittle embedded blocks, which are difficult to section.

When food samples are in liquid or semisolid form, there is obviously a special requirement to keep the sample intact during the various preparative steps. This is achieved by some means of encapsulation and a number of workers have used a modification of the Salyaev technique with considerable success in preparing milks, fruit juices, cell suspensions, and other liquid food materials. The principle is to prepare a capillary tube of agar that is filled with the sample and sealed at both ends with further agar, before being put into liquid fixative and processed as for other materials.

The embedded materials, usually in the shape of small tubes or blocks, are sectioned using an ultramicrotome to produce ribbons of ultrathin sections measuring between 50 and 100 nm. Traditionally, freshly cleaved glass knives are used as they have very sharp cutting faces. However, diamond knives have increased in popularity over recent years, as they remain sharp for extended periods of time and thus are suited to hard materials, e.g., woody tissue, bone, and crystalline components. Their main disadvantage is price, including the high cost of resharpening.

Cut sections are collected on meshed support grids designed to fit the TEM holders and then examined in the microscope, usually after staining to enhance the contrast of the sample and specifically identify individual components. Contrast in a TEM image is a result of differential electron opacity and therefore the most effective stains are based on reaction with electron-dense materials, in particular those with the highest atomic numbers, e.g., inorganic stains containing heavy metals.

Traditionally, staining has involved an initial reaction with uranyl acetate followed by lead citrate to provide a fairly nonspecific staining of the whole sample. This approach is still used routinely for many food specimens, with subsequent identification of components based mainly on their characteristic morphological features rather than their reaction to

stain. However, many different staining regimes have been developed for specific components, e.g. proteins, carbohydrates, and polysaccharides, in a wide range of different foods. Frequently these have been based on traditional biochemical reactions, but for TEM the staining compound includes a heavy-metal salt rather than a chromophore (coloured moiety). (*See* **Carbohydrates**: Determination; **Protein**: Determination and Characterization.)

Low-Temperature Preparation Techniques

For specimens which contain high levels of water or which are liable to change significantly during fixation or embedding procedures, preservation of the ultrastructure can be achieved with a cryofixation step, i.e., rapid freezing of small samples. This preserves the natural three-dimensional structure and solidifies the material so that it can subsequently be sectioned or fractured to reveal its internal structure.

Cryosectioning

Cryosectioning (sectioning at low temperatures) of frozen materials can be used for biological and biomedical tissues. It involves the controlled sectioning of frozen material using an ultramicrotome at low temperatures. Sections can be examined directly on a low-temperature TEM stage or after freeze-drying. However, there are few published data on the successful application of this technique to foods because of the difficulty of cutting ultrathin frozen sections of food materials and handling them prior to examination.

Freeze-Fracture/Freeze-Etch Replication

The alternative to sectioning frozen material is to fracture the sample and reveal its internal structure. Combining freeze-fracture with metal shadowing/carbon replication techniques has proved a particularly useful procedure for preparing many different foods for TEM, including emulsions, foams, and fatty confectionery, as well as frozen products such as icecream (Figure 2). This procedure is referred to as freeze-fracture or freeze-etch replication. One of the critical stages in the procedure is achieving rapid and efficient freezing of the sample to prevent the formation of visible ice crystals or other artifacts and considerable efforts have been made to optimize this process. Nitrogen 'slush,' i.e., a mix of liquid and solid nitrogen ($-210\,^{\circ}C$), is probably the

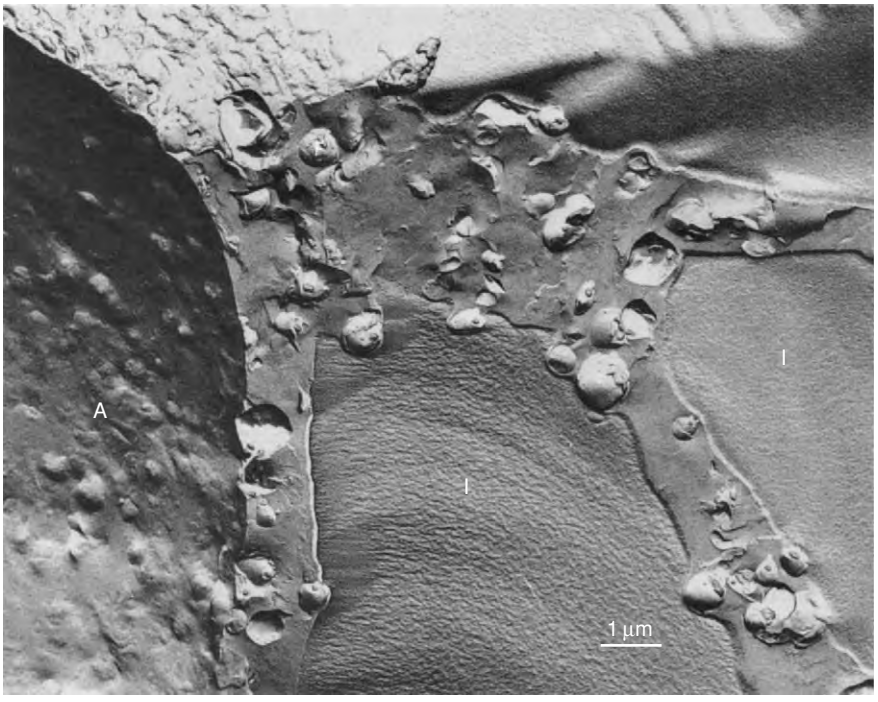

Figure 2 Freeze-etch preparation of icecream showing the presence of ice crystals (I), air bubbles (A), and fat droplets (F) within an aqueous sugar matrix. Bar = 1 µm. Reproduced from Microscopy: Transmission Electron Microscopy, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

most frequently used cryogen, as it is relatively easy to prepare and use. However, faster freezing rates are possible with liquid propane and hexane and these are now being used increasingly, particularly in combination with purpose-built freezing apparatus that are designed to enable safe handling of these flammable materials. Commercial developments in this field include specialized apparatus for cold metal block freezing ('slamming') and spray jet freezing. However, these are relatively expensive and have been of limited application in food microscopy.

Some success has been achieved with the use of cryoprotectants, for example glycerol, which can be used to infiltrate TEM specimens and inhibit the development of ice crystals on freezing. However, the interaction of the chemical has been shown to change the fine structure of samples and is now used more as a complementary technique in low-temperature preparations. Certain foods have been shown to offer an advantage in achieving fast freezing as they provide their own cryoprotection with the presence of high levels of sugars or salts.

Once the frozen samples are transferred on to the cold stage of a freeze-fracture/etch unit, they are fractured under vacuum and a carbon/platinum replica is prepared of the frozen fractured surface. Replication is achieved by the evaporation of a thin supporting layer of carbon followed by platinum deposited at an angle to produce a 'shadow' and hence reveal the topographical details of the surface.

For samples with high levels of water present, 'freeze-etched' replicas are made by the sublimation or 'etching' of a thin layer of surface ice before replication. After replication, samples are removed from the unit and the replicas are then cleaned in detergent and/or solvents to remove the food substrate before TEM examination. The choice of cleaning regime is important in order to clean the replica adequately whilst at the same time preserving its integrity. Strong caustic solutions have proved useful for many food materials, often combined with detergent or organic solvent washes to remove the fatty components.

Freeze Substitution

Freeze substitution involves the replacement of water (ice) in a frozen preparation with an organic solvent and subsequently with resin. There has been renewed interest in the technique recently because of its suitability for preparing biological tissues for immunolabeling. However, it is still of only limited application to food materials.

TEM Attachments

Scanning Unit

Many modern TEMs have the facility for attachment of a scanning unit to the main instrument. The unit allows the beam to be scanned across a selected area of the specimen and with the addition of suitable detectors the electron microscope can be used in a surface imaging mode or in a scanned transmission mode (STEM). In surface imaging mode, high-resolution images of the surface of samples can be obtained, although the small sample chamber restricts the size of sample that can be examined. The primary application of STEM is in imaging thicker samples. Since the images produced are electronic/digital, they have the advantage that they can be subjected to image processing and analysis to enhance the information further.

X-ray Microanalysis

Interaction of an electron beam with a specimen produces X-rays which are characteristic of its elemental composition. These X-rays can be collected to provide qualitative and quantitative chemical data from a sample. The technique is applicable to both transmission and scanning electron microscopes and is dealt with in more detail in the preceding article **Microscopy**: Scanning Electron Microscopy. In TEM its main application is in determining the distribution and changes in distribution of specific elements within thin-sectioned preparations. It has been of limited value for food samples to date, as the fixation and embedding techniques required for adequate preservation of microstructure frequently result in redistribution of elements within the material.

Applications of TEM

Many different food materials have now been examined using TEM. However, its main use has been in terms of characterizing the ultrastructure of different foods and relating this information to process changes and to textural and other physicochemical properties. Early studies involved the use of thin-sectioning techniques on foods with relatively high levels of plant or animal tissue, for example, meat, fruit, and vegetables. These studies were able to demonstrate the cellular changes associated with processing and cooking fresh tissue and included the use of specific staining regimes to monitor chemical changes such as the integrity of plant cell walls, meat fibrils, etc. More recent studies tend to use a combination of microscopy techniques to study the changes in more detail and to characterize more complex, composite materials, such as many confectionery products (Figure 3).

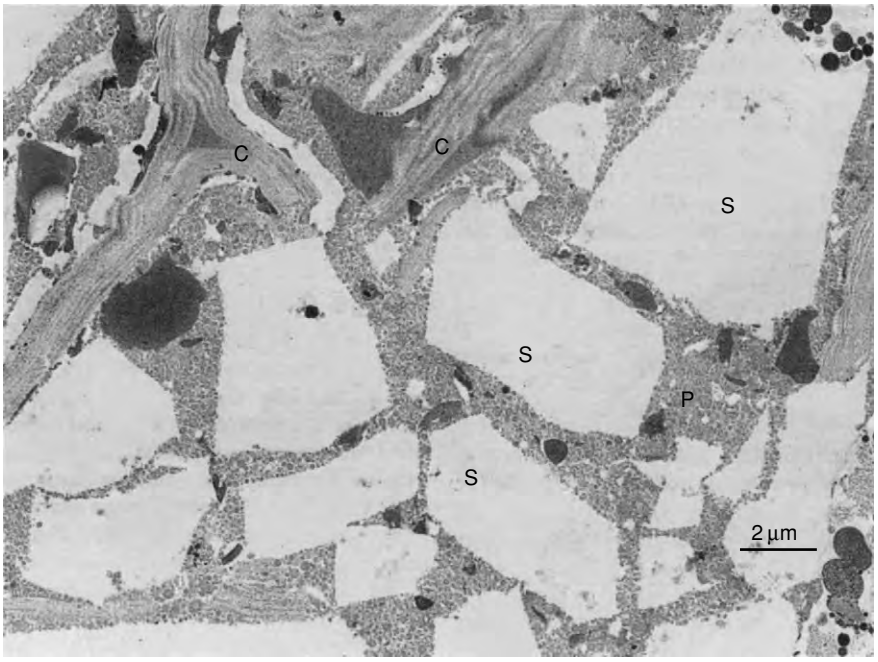

Figure 3 Thin section of milk chocolate showing the close association of milk protein (P), sugar (S), and cocoa tissue (C) within a chocolate particle. Bar = 2 μm. Reproduced from Microscopy: Transmission Electron Microscopy, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Studies on liquid or semisolid foods, for example, milk, emulsions, and fruit beverages, have increased significantly with the introduction of low-temperature techniques for handling and preparing these types of material. It has been possible, therefore, to identify and characterize the individual components present, for example, size and shape of fat droplets, form and level of aggregation of proteins (Figure 4), and to evaluate changes arising as a result of processes (heat, shear, pH changes, etc.), ingredients (different fats, protein source, etc.) or recipes.

The use of TEM for examining dry materials, for example milk powders, has been mainly superseded by the introduction of scanning electron microscopy techniques. However, it has been found that the use of thin sectioning, and more especially carbon replication techniques for TEM, has produced valuable information on the ultrastructural detail of these materials, in particular, the location and appearance of protein, which was not possible in any other way.

Future Developments in TEM of Food Materials

Immunolabeling

Colloidal gold immunolabeling is now probably the most highly acclaimed technique for electron immunocytochemistry. Gold spheres, bound to selected antibodies, are incubated with the relevant receptor to effect high-resolution and highly specific labeling. The technique is now enhanced by application of multiple labeling of different antigens using different-sized gold particles and quantitative techniques. However, it has yet to be used extensively for food materials. This is due to the problems of raising and isolating antibodies specific to food components, particularly after the foods have been processed, combined with the overall difficulties in handling food for TEM.

Low-Temperature Preparation Techniques

The use of low-temperature preparation procedures has increased significantly over the last few years. However, considerable advances have been made in faster freezing methods, alternative ways of handling frozen materials, e.g., sectioning, complementary fracturing techniques, and combinations of fracturing/sectioning/labeling, which have mainly been applied to nonfood materials. It is likely that these will be applied increasingly to studies on food.

Scanning Probe Microscopy (SPM)

This is a developing field and the techniques relevant to the microscopy of foods are centered on using scanning tunneling microscopes (STM) and atomic force microscopes (AFM). They provide high-resolution visualization of surface morphology and

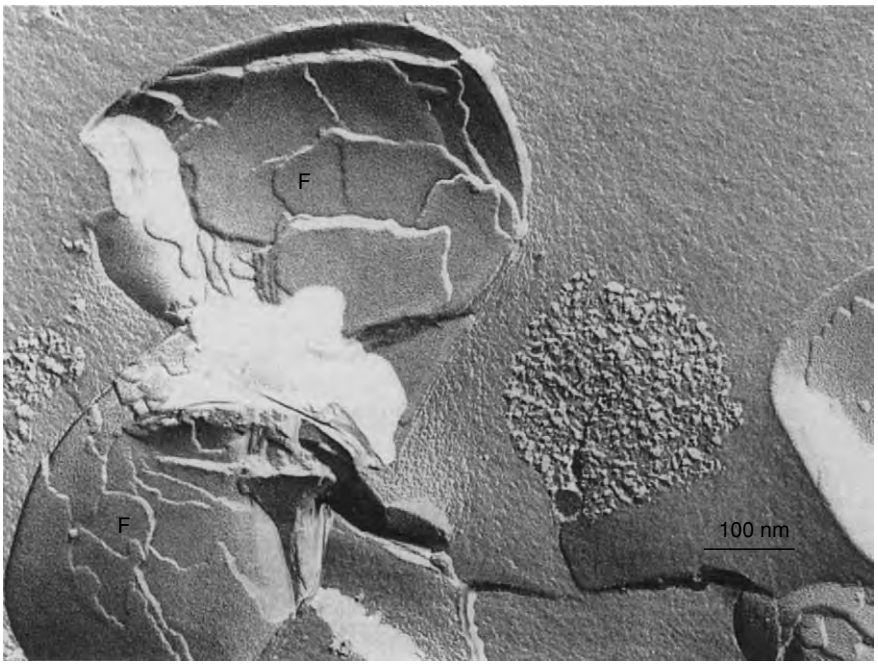

Figure 4 Freeze-fracture replica of a confectionery emulsion showing the fine emulsified fat droplets (F) and aggregated casein protein (P). Bar = 100 nm. Reproduced from Microscopy: Transmission Electron Microscopy, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

the applications of the different microscopes in this family extend to provide information on physical as well as structural properties, including information on topography, hardness, capacitance, stiffness/adhesion, and microcalorimetry. For some sample types the resolutions quoted are to < 1 nm and applications have been demonstrated in biology, pharmaceutical, and metallurgical industries. The techniques have potential for applications for foods and have already been used to demonstrate the structures of proteins in some food gels and emulsions.

See also: **Analysis of Food**; **Carbohydrates**: Determination; **Microscopy**: Scanning Electron Microscopy; **Protein**: Determination and Characterization; **Sensory Evaluation**: Sensory Characteristics of Human Foods

Further Reading

Aguilera JM (1990) *Microstructural Principles of Food Processing and Engineering 1. Food Science*. Barking: Elsevier.

Chescoe D and Goodhew P (1982) *The Operation of the Transmission Electron Microscope. Royal Microscopical Society, Microscopy Handbooks* No. 02. Oxford: BIOS Scientific Publishers.

Hayat MA (1989) *Principles and Techniques of Electron Microscopy; Biological Applications*. London: Macmillan.

Jewell GG and Heathcock JF (1988) Structured fat systems. In: Blanshard R and Mitchell J (eds) *Food Structure: Its Creation and Evaluation*, pp. 279–296. London: Butterworth.

Kalab M (1983) Electron microscopy of foods. In: Peleg M and Bagley EB (eds) *Physical Properties of Foods*, pp. 43–104. Westport: AVI Publishing Co.

Polak J and Van Noordan S (1984) *An Introduction to Immunocytochemistry: Current Techniques and Problems. Royal Microscopical Society, Microscopy Handbooks* No. 11. Oxford: BIOS Scientific Publishers.

Salyaev RK (1968) A method of fixation and embedding of liquid and fragile materials in agar microcapsules. *Proceedings of the Fourth European Conference of Electron Microscopy* 11: 37–38.

Vaughan JG (ed.) (1979) *Food Microscopy*. London: Academic Press.

Image Analysis

A Flook, Bedford, UK

Introduction

Image analysis is a computer-based process of extracting quantitative information from images. The process begins with the input of an image and

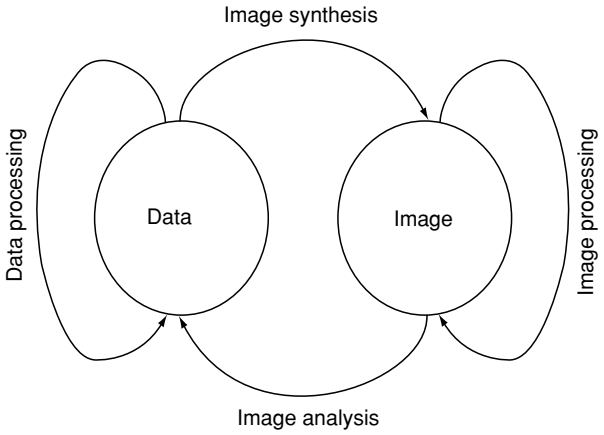

Figure 1 Conversion of information between data and image format.

ends with the output of numerical data (Figure 1). This distinguishes it from image processing where both input and output are in the form of an image.

Image processing is the means by which the input image is modified by one or more mathematical algorithms to generate an output image that is enhanced in some way. For example, edges may be enhanced or the noise reduced. Image processing is often used to prepare images prior to analysis.

Image analysis requires specialized computer equipment fitted with an imaging device, such as a television camera, coupled to a microscope or a macroviewer.

The first commercial image-analyzing computers were constructed in the early 1960s for grading steel by measuring nonmetallic inclusions. The value of using an automated method for the rapid, objective measurement of images was soon recognized for many different applications in life and materials sciences. Image analyzers with greater versatility have been developed but initially the application of the technique was limited by the high cost of the instrumentation. However, the price of computer hardware has steadily decreased over the years and now many manufacturers offer relatively inexpensive systems using a personal computer as the host processor. With the advent of more affordable equipment, applications have proliferated in many widely differing fields, including food science and technology.

Terminology and Nomenclature

Image-processing terminology is mainly derived from the mathematical functions used (e.g., Laplacian and Gaussian filters are used for image enhancement). These terms may not always be easy to understand but are used universally and can, in the main, be used without ambiguity. This cannot be said for image analysis. There are no internationally agreed standards and the terminology used can be specific to the manufacturer of the equipment. Indeed, in the early years of image analysis it seemed almost obligatory for each new analyzer on the market to come with its own set of terminology. For example, one manufacturer uses the term segmentation for the process of extracting a binary image by gray-level thresholding. Another calls this process detection and reserves the term segmentation for the process of separating touching regions. Also, the binary regions to be measured can be referred to as regions, objects, features, or even blobs, depending on the equipment used. This makes it difficult for newcomers to read research papers and relate the functions named to those installed on their own equipment.

This situation, although regrettable, is unlikely to change. The terms used in this article are those most commonly used but the reader should be aware that they are not unique.

The Hardware

There is a wide range of image analyzers commercially available. They all differ in detail but are broadly similar in design. A typical system consists of the following components.

Microscope

This is usually fitted with a trinocular head to allow direct viewing of the specimen and for the mounting of a television camera. (A macroviewer is an alternative optical input device for photographs or larger specimens.)

Television Camera

Most image analyzers use standard-format charge-coupled device (CCD) TV cameras, although an increasing number of systems are being fitted with color and high-definition cameras.

Analog-to-Digital Converter (ADC)

This is a device that converts the continuous voltage output signal from the camera into a sequence of numbers representing the voltage at timed sampling intervals. The lighter the region of the picture at the sampling point, the higher the voltage and hence the larger the numerical value. Most systems are fitted with an 8-bit converter, allowing $2^8 = 256$ different gray-levels. Three independent channels are used for color images – one each for the red, green, and blue images that constitute a conventional real-color image.

Image Store

The sampling interval of the ADC is chosen to give a discrete number of pixels (an abbreviated term for picture elements) for each scan line of the television signal. These are read into an image store. The format for the image store is a matrix of pixels corresponding to the CCD format of the camera. The store is usually large enough to accommodate several digitized images at once. Systems based on personal computers use buffers within the random access memory.

Monitor

This displays the images and provides a user interface for programming. Most image analyzers use a scripting language to build programs that control the processing of the images and the analysis of the data.

Image Processor

Digitized images are very large arrays of numerical data. Early instruments had specialized image processors to manipulate the images but modern personal computers have processors powerful enough to cope with the demands of image processing without using special processors.

Binary Image Store

In most applications the image will ultimately be reduced to a binary form where each pixel of the original image is converted to either a 1, representing a pixel within an area occupied by one of the objects to be measured, or a 0 representing a pixel in the background. Binary images are normally displayed either as bright pixels on a dark background or by the use of two distinct colors representing the 0s and 1s.

Early instruments had specialized memory stores for this type of data but these are now obsolete.

The Host Processor

This is often, but not necessarily, a personal computer. On older machines it is used to manage the data flow within the computer and issue instructions to the specialized processors. Now all the processing is usually done using the standard computer processor. The host processor also supervises peripheral devices such as printers and the automatic focus and x,y stage controls on the microscope (where fitted). Finally, it provides the means to process the data and provide any statistical, graphical, or other utility program that may be required for a particular application. The host processor will usually be fitted with disk drives for the storage of data and programs.

Output Device

This will commonly be a printer or plotter that will enable the results to be converted into hard copy in graphical or tabular form.

The Principles of Image Analysis

Image analysis often requires a greater degree of care in preparing the sample and setting up the microscope than is necessary for visual inspection. The human brain is well adapted to the recognition of objects of interest in an image and to ignore extraneous information. The computer, on the other hand, can only cope with relatively simple images. For this reason the image analysis process should be considered to start from the sample preparation onwards.

Measurements can only be made on a binary image. The ultimate goal of the image prepossessing is to simplify the image in such a way that will make it easy to convert it to this form. In the binary image the bright pixels should correspond to those areas in the original image that represent the objects or structures that are to be measured.

A typical sequence will include the following steps.

Sample Preparation

Small variations in the preparation protocol can determine whether or not a project is amenable to image analysis. The best images for analysis are those that show the objects that are to be measured in greatest contrast with respect to the rest of the image. Color contrast may be used with those instruments fitted with a color camera. Intensity contrast is used for monochrome images.

Image Capture

The analysis proceeds by setting up the input device to deliver the best possible image to the camera. When a suitable image is obtained, it is read into the computer memory via the ADC.

Image Processing

There are many methods of mathematically processing an image so that the output image is enhanced in a manner that renders it more suitable for image analysis. For example, the edges of structures may be sharpened or the image may be given enhanced contrast or reduced noise.

It is not always necessary to preprocess the image in this way, but shading correction is almost always used. Shading correction is the process by which the computer attempts to compensate for any uneven illumination of the specimen. This is usually achieved by storing the pattern of the variation in intensity

obtained from a blank image field. This image, called the shading pattern, is used to correct subsequent images.

Image processing can never increase the information content of an image. Any enhancement of one image attribute is obtained at the expense of another. For example, edge enhancement algorithms tend to increase the noise content of an image. Many algorithms have been developed over the years to enhance images with the minimum of unwelcome side-effects. There are, for example, methods of reducing noise without smoothing out the edges of objects. However, heavy processing can still introduce unwanted artifacts in the image. For this reason sample preparation and lighting conditions are best chosen to reduce the need for image enhancement to a minimum.

Segmentation

This is the process by which the digitized, gray, or color image is converted into a binary image representing the objects to be measured. This is the most important, and often the most difficult, step in image analysis. The computer needs a binary image to make measurements. There are several ways of defining the regions to be segmented. The simplest form is by setting a gray-level value below which (or above, for light objects) a pixel will be counted as being part of the object. All other pixels will be counted as belonging to the background. This form of segmentation is also referred to as thresholding or detection. Another method allows the operator to draw round or point to a typical region; the computer will sample the intensities of the pixels in the region and generate a binary image of all regions containing pixels with matching intensities. There are also methods that specifically detect the edges of objects.

Color discrimination provides an alternative method of segmentation for true color images. There are two main color models used in image analysis–the red, green, blue (RGB) model and the hue, luminance, saturation (HLS) model. Color images are separated into their three components (RGB or HLS) according to which color model is to be used for segmentation. Each component is held in memory as a separate image. These component images can be viewed separately as monochrome images but are normally displayed combined as a color image (Figure 2). Segmentation methods, similar to those used on monochrome images, are applied to each of these components and the resultant binary images combined. The resultant binary image will represent those regions in the original image that have a similar color composition.

Binary Image Processing

Often the segmentation process does not produce a perfect binary image of the objects of interest and many binary operations have been devised to amend the defects in the image. For example, there are operations for filling in holes and others that dilate or erode the binary image by adding or subtracting pixels from the perimeter of objects. Erosion and dilation can often be used to good effect by combining them to form new operations known as closing and opening. Closing is the process of dilation followed by an equal amount of erosion and is useful for joining up fragmented regions of an object. Opening is the converse operation and is often used to clean up ragged edges and remove small or thin artifacts from the image. Opening and closing operations leave the overall size of the objects relatively unchanged compared with erosion or dilation used on their own.

There are other more complex binary operations which can be used, for example, to separate touching objects, reduce objects to a skeletal representation, and erode objects to the point just before they would otherwise disappear. These tools allow the image analyst to automate much of the editing of binary images that would otherwise have to be done by time-consuming manual methods. They form an important part of image analysis and can make the difference between being able or otherwise to perform an analysis in a reasonable time. The theory behind these operations is known as mathematical morphology.

Editing

The binary image can be edited manually by using a hand-held device, such as a mouse or a light pen, which is interactive with the computer. By this means sections or lines can be cut out or drawn in at will. Although it is often necessary to do a little manual editing, this should always be regarded as a method of last resort since it is a slow operation and introduces a subjective element into the analysis.

Measurement

There are two distinct modes of measurement that can be made in image analysis: field-specific and object-specific. Before discussing these it is important to distinguish between the image and measurement frames. The total area of the specimen scanned by the TV camera and displayed on the monitor forms the image frame. The measurement frame is a subframe within the image frame over which the measurements are made. The area within the image frame that lies outside the measurement frame forms the guard region.

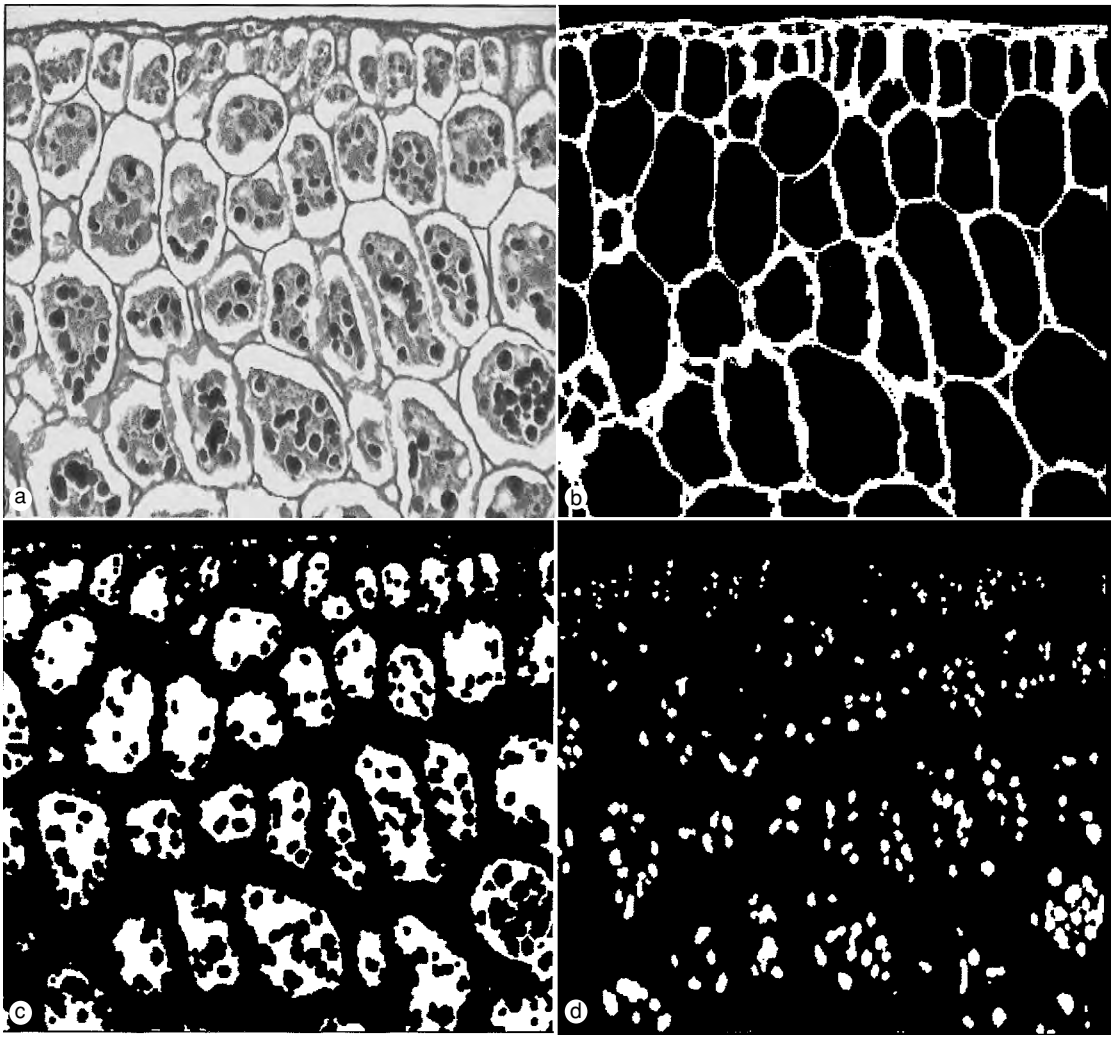

Figure 2 (see color plate 109) Color segmentation of a section from a pea cooked from frozen. (a) Original image; (b) cell walls; (c) cytoplasm; (d) gelatinized starch granules.

Field-Specific measurements Field-specific measurements are summed over all the objects and parts of objects lying within the measurement frame. In this mode, if an object is intersected by the edge of the measurement frame, only those parts of the object that project into the frame are measured. Field-specific measurements are used for stereological calculations of global parameters such as phase volume or count per unit volume.

Object-specific measurements Object-specific measurements are made on each individual object in the measurement field. In this mode, each object is associated with a unique object location reference point that is determined by the counting logic of the particular image analysis equipment used. The whole of an object is measured if its reference point is inside the measurement frame, even if much of it spills over the edge of the frame. Conversely, the object will not be measured at all if the reference point falls outside the frame. Any uniquely definable point may be chosen as the reference but, for practical reasons, the location of the last pixel visited within the object as the scan progresses down the screen is most commonly chosen. The measurement frame size is chosen so that the guard region, which surrounds it, is wide enough to insure that any object whose location point is within the measurement frame is wholly within the image frame.

Some primary measurement parameters for objects are defined below. Most of these measurements can also be made in field-specific mode:

- Area: the total number of pixels within an object
- Perimeter: the length of the boundary between an object and the background

- Horizontal intercept: the number of intersections of the TV scan line with an object
- Vertical intercept: the equivalent of the horizontal intercept but in a vertical direction
- Feret's diameters: These are equivalent to caliper diameters, i.e., the perpendicular distance between the parallel jaws of a caliper just touching the object. The magnitude of these measurements is dependent on orientation and the more advanced image analyzers are able to measure Feret's diameters at a large number of small angular increments.
- Integrated brightness: The gray-level values summed over the region corresponding to the binary image of the object. With suitable calibration, these measurements can be converted to optical density values.
- Count: In field-specific mode this gives the total number of objects in the field. It may also be possible, on some machines, to count objects within objects (e.g., the number of vacuoles within a cell membrane).

These primary measurements are often combined to create additional measurement parameters such as:

- Equivalent circle diameter: The diameter of a circle having the same area as the binary image of the object. Used as a size measurement that is insensitive to the orientation of the object.
- Form factors: The dimensionless ratios of primary measurements whose magnitude gives an indication of the morphology of an object. Dimensionless ratios are used to give values that are independent of size, e.g.:

$$\text{Circularity} = 4\pi \text{ area}/\text{perimeter}^2$$

A perfect circle will have a circularity value of one. All others shapes will give a lower value.

- Length: Usually taken to be the longest Feret's diameter.
- Breadth: This can be defined in a number of ways. For example, the shortest Feret's diameter may be used or it may be estimated by dividing the area by the largest Feret's diameter. There are other ways of defining breadth; the choice will depend on the application.

Three-Dimensional Imaging

One of the restrictions of image analysis has been the need to work on two-dimensional sections. Volumetric properties have to be derived from stereological computations. Whilst this approach is sufficient for most work, it does have significant drawbacks: care must be taken with anisotropic materials to avoid biasing the data and the sectioning process may mechanically distort the material. Furthermore, it is necessary to fix and embed soft or friable material specimens in order to section them. This may lead to further structural changes in the material, thus rendering any measurements of dubious relevance to its original state.

There is a number of methods for imaging beneath the surface of structures. Confocal microscopy, magnetic resonance imaging, and microfocus X-ray tomography can be used to provide series of two-dimensional, virtual sections. Many image analysis packages have software specifically designed to reconstruct three-dimensional images from a series of images representing two-dimensional slices generated by these methods (Figure 3). The reconstructed images can be used to create further virtual sections in any direction or to make direct volumetric measurements.

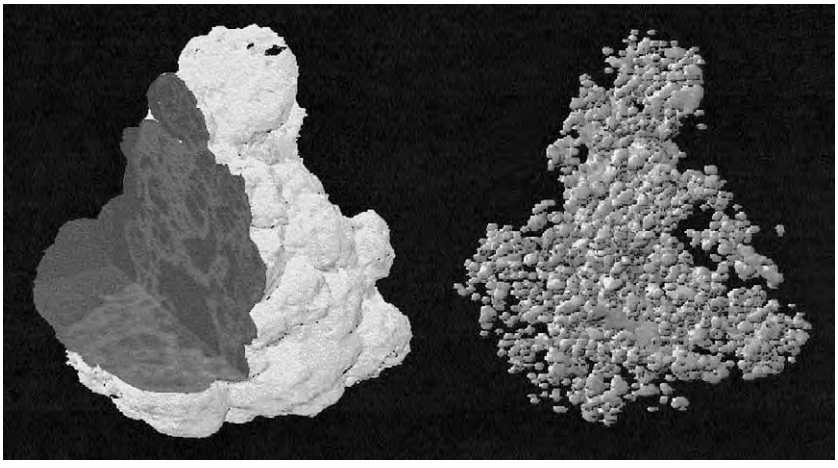

Figure 3 (see color plate 110) Reconstruction of puffed wheat from X-ray tomographic sections. Left, reconstruction with cut-out to show the internal structure; right, reconstruction of the voids inside the puffed wheat. Courtesy of Dr Paul Jenneson, Surrey University, UK.

Applications

The potential for applications of image analysis is enormous – anywhere where rapid, objective, quantitative measurements are required. Particle size and shape analysis can be used to monitor comminution or agglomeration processes. The technique has also been used to measure the air bubble size distribution in aerated products such as icecream, chocolate, and batters. Where objects have well-defined morphologies, shape discrimination can be used to count several components simultaneously, e.g., counting the population densities of lactobacilli and streptococci in yogurt.

The Future

The development of high-resolution scanners for computer document archiving and the advent of high-definition TV technology has provided a number of high-resolution image input devices. This, coupled with the relentless pursuit of even cheaper and faster computer memory chips, will provide the image analyst with access to much higher-resolution images; 1024×1024 pixel image matrices are already commonplace and larger formats will follow.

The latest image analyzers use standard computer components. The only specialized hardware is the frame-grabber board. New computer interfaces are being developed that will allow the fast transfer of data directly into the computer from a digital camera. This will reduce the cost of the equipment further and will make image analysis more available to cost-sensitive applications.

Three-dimensional analysis is already an option on many image analyzers. Its use is restricted due to the huge amount of memory required and the high demands made on the processor. These restrictions will become less relevant and this type of analysis will become more common in the future.

Most electron microscopes and all of the newer forms of microscopy (e.g., scanning tunneling, atomic force, and confocal) are controlled via an interface to a personal computer. These provide an ideal platform for image processing and analysis.

See also: **Microscopy**: Scanning Electron Microscopy; Transmission Electron Microscopy

Further Reading

Gonzalez RC and Wintz P (1977) *Digital Image Processing*. Reading, MA: Addison Wesley.

Russ JC (1990) *Computer Assisted Microscopy*. New York: Plenum Press.

Serra J (1982) *Image Analysis and Mathematical Morphology*. London: Academic Press.

Microwave Cooking *See* **Cooking**: Domestic Techniques; Domestic Use of Microwave Ovens

MIGRAINE AND DIET

M R Costa and M B A Glória, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

Migraine Headache

'Migraine' is a French derivative of the Greek word for hemicrania, or half a head. The problem of migraine is of long standing. Aretaeus of Cappadocia, in the second century AD, defined migraine as 'a paroxysmal headache disorder that occurred on one side of the head and then the other, recurred at regular intervals, associated with vomiting and photophobia, and ameliorated by dark surroungings.'

Throughout the ages, migraine has been a problem to patients and an enigma to physicians. Migraine is a public health problem of enormous scope that has an impact on both individual sufferers and society. Several studies have indicated that 82–85% of individuals have some headache-related disability. About one-third of them are severely disabled or need bed rest. In addition, many live in fear, knowing that at any time an attack could disrupt their ability to work, care for their families, or meet social obligations. With regard to the impact on society, the costs in

human terms are overwhelming. The totals, in terms of lost production, medication, doctors' and hospital fees are huge.

Migraine may begin in childhood and usually starts before the age of 40. Its prevalence in childhood is about equal among girls and boys, but at puberty the two groups diverge sharply and females begin to outnumber males. Frequency tends to decrease with advancing age, and up to three-fourths of women note relief after menopause.

It is believed that sufferers from migraines may have an inherited predisposition for this disorder. It has been reported that 50–70% of patients with migraine have a positive family history. The precise mode of inheritance is not known, but current thought is that this predisposition is an autosomal dominant trait characterized by incomplete penetrance.

Symptoms of a Migraine Headache

The symptoms most frequently associated with migraine headache are indicated in Table 1. Migraine is characterized by throbbing head pain. Most of the time, it is unilateral in adults; pain may occur on either side of the head but rarely changes sides during an attack. In children, migraine is ordinarily bilateral. It is often accompanied by nausea and sensitivity to light and/or sound. Less frequently noted are abdominal bloating and diarrhea, pallor, sweating, chills, and cold extremities. Every individual is different and may experience some symptoms and not others, or in various combinations. Intensity varies from one attack to another.

The frequency of migraine ranges from every few days to less than once a year. However, most patients experience one to four attacks a month. These generally last from 4 h to 2 days, but severe episodes of prolonged migraine may persist for weeks. Proper management can significantly reduce the impact of this debilitating chronic disorder and improve both functioning and quality of life.

Table 1 Symptoms associated with migraine headache

Symptoms of migraine headache
Moderate to severe headache pain
Duration of 4–48 h, sometimes longer
Throbbing pain located on one side of the head
Nausea
Vomiting
Sensitivity to light, sound, or odors
Pain becomes worse when moving

Pathophysiology of Migraine

Migraine is a vascular headache believed to be caused by blood flow changes and certain chemical changes in the brain leading to a cascade of events, including constriction of arteries supplying blood to the brain and the release of certain brain chemicals.

Migraine has a number of forms, some of which may be confused with other diseases. It can be classified (Table 2) as migraine with aura (classic) and without aura (common).

The exact cause of migraine is uncertain, although various theories are being studied. The neurogenic theory emphasizes the significance of active substances acting upon the central nervous system and promoting intracerebral vasoconstriction and extracranial vasodilatation. According to many researchers, migraine is due to the vulnerability of the nervous system to sudden changes in either the body or the surrounding environment. They believe that migraine sufferers have inherited a more sensitive nervous system response than those without migraine. During a migraine attack, changes in brain activity produce inflamed blood vessels and nerves around the brain.

Another theory is that migraine represents a disorder of cerebrovascular regulation. In the typical migraine sequence, a trigger precipitates a cascade of events that includes release, activation, and production of active substances. When these substances

Table 2 Symptoms and signs of different forms of migraine

Types of migraine	*Characteristics*
Migraine with aura (classic)	Neurologic symptoms occur 15–60 min before headache: scotomas or blind spots are common; photopsia and teichopsia or fortification spectra occur frequently; diplopia, macropsia, and metamorphopsia are less common Headache begins as neurologic symptoms diminish
Migraine without aura (common or nonclassic)	Occurs in 85% of all patients No neurologic aura precedes the headache Premonitory disturbances of mood or appetite (craving for sweets, euphoria, fatigue, yawning) can happen 1–3 days before an attack Bilateral or generalized pain

concentrate in the tissue, vascular and nervous phenomena are enhanced. A fact of paramount importance is that some of these substances potentiate one another. In patients with regular attacks, each episode of migraine is followed by a phase of immunity and then by one of increasing reactivity, which can be divided into two stages. The first is called adrenergic, and is characterized by relatively mild symptoms, such as no pain, vasoconstriction, slight excitation, and sensory manifestations. In the second one, there is intense pain, marked vasodilatation, and pronounced psychic depression. During the reactive phase, migraine episodes may be triggered by stimuli that promote vasoconstriction (e.g., cold, anxiety) as well as vasodilatation (e.g., heat, alcohol, nitrites).

Triggers or Provokers of Migraine

Certain factors can provoke or trigger attacks more frequently in those susceptible to migraine headache. Some possible triggers include biochemical or environmental factors; hormonal changes in women; too much or too little sleep; stress (feeling overly tired); pharmacological precipitants; and diet. According to many studies (Table 3), the most commonly cited triggers are stress, change in sleep patterns, menstrual cycle, hunger, and diet or specific foods. Not all migraines have the same provoking factors, nor do all these factors necessarily provoke an attack. Migraine can also arise from a number of factors acting in unison.

Biochemical Factors

It has been postulated that there is a biochemical defect in migraine sufferers, especially with respect to their ability to metabolize amines and phenols when compared to controls. The monoamine oxidases (MAO) act as a barrier against the absorption of vasoactive substances contained in foods. MAO defect can prevent amine metabolism and, therefore, cause accumulation of monoamines. There could also be a deficiency of phenolsulfotransferase P, which is specific for the detoxification of dietary phenols.

Table 3 Major precipitating factors in migraine

Precipitating factors	*Frequency (%)*
Heredity	66
Stress, anxiety, and worry	43–88
Change in sleep patterns	77
Menstrual cycle	39–65
Oral contraceptive	31
Bright light	30
Diet or specific foods	13–74
Fasting	7–77

Environmental Triggers

Environmental triggers of migraine headache include weather or temperature changes, exposure to bright lights (glare from snow or fluorescent lights, computer screens), loud noises, and strong odors. High altitudes or the atmosphere of a pressurized cabin in a commercial airliner may also lead to migraine attacks or similar vascular headaches.

Menstrual Cycle

Many women with migraine have attacks linked to their menstrual cycles. Menstrual migraine occurs mainly between 2 days prior and 3 days following the onset of flow. It is usually related to changing levels of menstrual hormones during menstruation, puerperium, hysterectomy, and pill-taking. Migraine prevalence peaks in women in their early 40s and then declines steadily with age. Migraine may worsen in early pregnancy, but it generally improves in later pregnancy. Although some women find that their migraines worsen when taking oral contraceptives, the majority experience no change or improvement in their headache pattern.

Pharmacological Agents

Some pharmacological agents used to treat hypertension and coronary artery disease may increase migraine frequency. Reserpine may increase migraine frequency and exacerbate depression in some patients. Amyl nitrite, used for treatment of angina pectoris, may adversely affect the frequency of migraine. Migraine was observed after beginning 25 000 IU daily supplementation of vitamin A.

Headache potential is also increased in patients treated with MAO inhibitor drugs, such as phenylzine, isocarboxazid, pargyline and tranylcypromine for depression; iproniazid and isoniazid for tuberculosis; procarbazine hydrochloric for cancer; and furazolidone for antimicrobial use.

Emotional Stress and Anxiety

Stress is the most common migraine trigger. Emotional stress or response to life's daily pressure can trigger migraine attack in some individuals. Stress can cause alteration in adrenal hormone level. Stress, fear, and anger may liberate vasoactive substances from tissue stores.

Depression has been associated with migraine in both prospective and retrospective epidemiological studies, and psychological distress has been shown to be responsive to cognitive-behavioral headache interventions.

Sleep Patterns

An extra 1–2 h sleep can be enough to initiate a headache. The mechanism involved is not understood

but carbon dioxide retention with vasodilatation or delayed breakfast provide two possibilities. However, lack of sleep for social or work reasons, as well as psychological factors, can be a migraine precipitant.

Hunger and Hypoglycemia

Blood glucose levels have important effects on the tone of cerebral blood vessels. Marked cerebral vasodilatation may occur if blood glucose levels are reduced by insulin or other causes. Therefore, hunger headache can occur during fasting or food deprivation. It is provoked after stopping food intake for 5 h during the day or 13 h overnight. This condition may also be responsible for the weekend headache that develops when a person sleeps late and delays the morning meal. Since blood glucose levels do not fall below normal levels, it cannot be called hypoglycemia. However, hypoglycemia has also been associated with headache. It is a prominent symptom in a patient entering insulin shock. In addition, the reactive hypoglycemia that occurs after ingestion of a large simple-carbohydrate meal may also precipitate a similar headache in susceptible persons. Therefore, in order to reduce the number and severity of migraines, a patient should not skip meals, should have three or more well balanced meals a day and should avoid overabundance of carbohydrates at any single meal.

Cold Drinks or Icecream

Swallowing cold drinks or icecream can quickly cause a brief frontal icecream headache lasting a few minutes. The pain occurs either in the palate and throat or the forehead, temple, or ear. The cause is believed to be the sudden cooling of the oral pharynx. Icecream headache is an excessive vascular reaction to cold and a manifestation of erratic vasomotor regulation in patients with migraine.

Allergy

There is a very long medical history which attempts to relate migraine to allergy. However, the view of many researchers in the field of migraine appears to be that dietary migraine is not an allergic phenomenon. Recently, it has become evident that the immune system might be affected in migraine. Foods such as milk, wheat, eggs, and meat are usually involved. In patients with migraine, these foods induce formation of immune complexes which can release active substances. They may also react with mast cells, releasing histamine and serotonin.

Influence of Diet

Studies on the influence of diet on migraine are difficult to evaluate. A diet trial cannot be performed in a double-blind manner. Moreover, it is hard to reduce selectively one or a couple of components in the diet and yet preserve the original proportions of nutrients. Diet trials are also difficult to carry out with regard to compliance and practical inconvenience for the patients.

Although the relationship between diet and migraine has been a source of controversy for many years, certain relationships have become clear. Some specific foods, alcoholic beverages, and some food additives are possible triggers of diet-precipitated migraine, even though sensitivity varies with each patient.

Chocolate, cheese, and citrus fruits are the most frequently cited provoking factors to which patients ascribe their dietary migraine (Table 4). Fatty fried foods, vegetables, tea and coffee, pork, and seafood have also been cited. In 490 migraine patients attending the Princess Margaret Migraine Clinic at Charing Cross Hospital in 1984, 19% cited chocolate, 18% cheese, 11% citrus fruit, 29% reported alcohol, 7% coffee, 3% pork and dairy products, and 2% eggs as precipitants of their headaches. In another study with 429 migraine patients attending the same clinic from 1989 to 1991, 17% reported that headaches could be precipitated by cheese and chocolate, and nearly always by both; 18% reported sensitivity to all alcoholic drinks, while another 12% were sensitive to red wine but not to white wine; and 28% reported that beer would precipitate headaches.

Food-induced headaches are mediated by chemical constituents common to these foods (Table 5). Some are naturally occurring components, including tyramine, phenylethylamine, serotonin and other vasoactive amines, alcohol, flavonoid, caffeine, and copper. Additives added to foods, such as sodium nitrite and monosodium glutamate (MSG), may also elicit migraine headache. Some other food or food

Table 4 Major precipitating foods in migraine

Precipitating factors	*Patients with dietary migraine (%)*
Chocolate	17–75
Alcohol	11–67
Red wine	12
Beer	28
Cheese	17–50
Citrus fruits	11–30
Nuts	12–17
Fatty fried foods	18–19
Coffee	7–15
Ham	4
Pork	3–15
Dairy products	3–9
Seafood	6–10
Onion	18

Table 5 Causes of diet-precipitated headache or migraine and foods involved

Substances	*Foods*
Food components	
Vasoactive amines	
Tyramine	Aged cheeses (cheddar, Swiss, Gouda, Roquefort, Stilton, parmesan), sausages, pickled herring, red wine, beer
Phenylethylamine	Chocolate, cocoa
Serotonin	Banana, tomato, walnuts, tryptophan-rich protein
Ethanol	Wine, beer
Caffeine	Coffee, tea, cola, and candy withdrawal
Copper	Chocolate, cocoa, nuts, shellfish, distillates produced in copper stills, wheat germ
Additives	
Nitrites	Aged, canned, and cured meat such as hot dog, bacon, ham, salami, cheese (read labels)
Monosodium glutamate	Chinese food, soy sauce, instant soup, bouillon cubes, meat tenderizer, seasonings, ready-to-eat dishes, snacks (read labels)

components have been associated with migraine, such as food colorings (erytrosin), food flavorings, licorice extract, pesticides, boron, aspartame, vitamins (A, D, niacin), minerals (calcium, manganese), herbal tea, and seafood. However the prevalence is low, and further research is needed to ascertain their role in migraine.

Vasoactive Amines

A variety of amines have been implicated in the development of headache, most commonly the vasoactive monoamines tyramine, tryptamine, norepinephrine (noradrenaline), epinephrine (adrenaline), dopamine, beta-phenylethylamine, serotonin, and synephrine.

Ingested monoamines are normally inactivated by MAO, found in high amounts in the gut and which prevent monoamines from the diet to enter the general circulation in significant amounts. However, if large amounts of monoamines are present in foods or if the enzymes are inhibited by genetically deficient metabolism or by MAO inhibitor drugs, the metabolism is prevented. A MAO defect would act in precipitating migraine through the mechanism of increased absorption of vasoactive substances. Headaches have also been described in patients under treatment with MAO inhibitor drugs.

The monoamines that are not metabolized are free to act directly on blood vessels. They are also known to cause release of endogenous active substances, creating the potential for migraine and also for hypertensive crisis.

Tyramine Tyramine is a vasoactive amine that promotes blood pressure elevation, resulting in pain. Tyramine leads to cerebral vasoconstriction and subsequent rebound vasodilatation that causes a migraine attack in susceptible persons. Episodes can be accompanied by nausea and visual abnormalities. Symptoms are evident 1–12 h after ingestion of food containing tyramine. A dose of 10 mg tyramine has been associated with migraine onset; however, levels of 6 mg can cause migraine in patients under treatment with MAO inhibitors. Alcohol has been observed to facilitate tyramine absorption.

Tyramine is found in a number of foodstuffs, most notably aged and fermented foods and beverages. Cheeses (especially Camembert, Cheddar, Parmesan, and Emmental), overripe bananas, avocado, canned figs, peanuts, pickled herring, dried and fermented meat products and alcoholic beverages (wine, beer) are known to contain tyramine.

Phenylethylamine Similarly to tyramine, phenylethylamine causes headaches after 12 h of ingestion of 3 mg by susceptible individuals. Chocolate, a commonly cited trigger of headache, is specially rich in phenylethylamine. An ordinary chocolate bar may have at least 3 mg phenylethylamine per 56 g. Dark chocolate has been found to be more potent than milk chocolate in causing migraine.

Serotonin Chemically, serotonin is 5-hydroxytryptamine. It is a neurotransmitter that acts as a powerful constrictor of arteries, reducing blood supply to the brain, causing migraine. Because serotonin released from platelets has been suggested to initiate migraine, a decreased platelet serotonin content attained by a reduced intake of serotonin and the serotonin precursor tryptophan might be beneficial. Serotonin-rich foods include bananas, tomatoes, and walnuts. A protein-tryptophan-reduced diet decreased headache frequency and skin symptoms in patients with migraine-like headache. Therefore, the intake of foods with a high tryptophan-to-protein ratio should be restricted.

Other amines Dopamine is found in the pod of Italian broad green (fava) bean and bananas. Synefrine and octopamine are found in citrus fruits, including oranges, lemons, and grapefruit. Port, Chianti, and red wines, and some other alcoholic drinks contain histamine, which is also a vasoactive amine.

Ethanol

Alcoholic beverages may provoke migraine. Ethanol, which has both central and direct vasodilator properties, is a well-known potential trigger for migraine. The vasodilatation is assumed to be the result of depression or alteration of the central vasomotor centers, since the direct action of alcohol on blood vessels is insignificant. Ethanol can inhibit MAO, thus preventing amine metabolism and enhancing its activity. Ethanol can induce migraine in some persons regardless of the amount consumed. It generally occurs 7 h after drinking, usually starting in the following morning. Clinically, excessive dietary salt intake also appears to potentiate alcoholic migraine.

Some believe that congeners present in the majority of the alcoholic beverages are even more potent contributors to migraine than ethanol itself. Three troublesome chemicals associated with wine are histamine, tyramine, and phenylethylamine. Furthermore, compounds in the flavonoid fraction of wine have been observed to cause release of active substances.

In addition to the usual alcoholic beverages (wine, beer), many cough preparations, tonics, mouthwashes, liquid vitamins, and nonprescription drugs contain high concentrations of ethanol.

Copper

Foods that are high in copper have been responsible for the precipitation of migraine. Copper is involved in metabolic and enzymatic activities such as hemoglobin synthesis, bone and elastic tissue development, and normal function of the central nervous system. Copper is known to be necessary for the metabolism of vasoactive amines such as serotonin, tyramine, and catecholamines. Abnormal copper metabolism following ingestion of copper-containing foods would lead to decreased inactivation of amines and therefore the onset of migraine.

There are substances that can facilitate copper absorption causing migraine, among them citrate, glutamate, and histamine. Glutamate from MSG acts to transport significant amounts of copper to tissues, facilitating copper absorption. Histamine is known to bind and transport copper to tissues.

Foods containing high levels of copper include liver, cocoa, chocolate, nuts, shellfish, mushrooms, wheat germ, and certain cereals. Whiskies and distilled beverages that are produced in copper stills can also be a source of copper.

Caffeine

Caffeine is an ingredient of coffee, cola, tea, chocolate, and many over-the-counter and prescription analgesic compounds. It is a vasoconstrictor; however it does not cause headache as a primary manifestation. But rebound vasodilatation may occur, causing a migraine-like withdrawal headache when excessive caffeine consumption is suddenly interrupted. For example, a person who consumes large quantities of caffeine-containing beverages during the week may experience a headache on weekends if caffeine consumption is decreased. The headache is relieved by taking another dose.

Theobromine, a methylxanthine found in cocoa that is similar in chemical structure to caffeine, may act as the headache trigger.

Food Additives

Nitrates An aching headache often accompanied by facial flushing occurs in sensitive persons exposed to nitrites added to cured meats. As little as 5–10 mg sodium nitrite may provoke headache in a migraine-prone person. Headaches are noted to occur 35 min after sodium nitrite ingestion, and last approximately 1–2 h.

Up to 0.02% of nitrate and nitrite compounds are employed in meat products to create a pink color, to impact a cured flavor, and to prevent botulism. These foods include bologna, salami, sausage, pepperoni, bacon, frankfurters, corned beef, beef jerky, lunch meats, canned ham, and sausages. In some countries, nitrate can also be added to cheese to prevent late blowing. Migraine sufferers are advised to eliminate cured meats from their diets.

Monosodium glutamate MSG is another chemical that migraine patients are often advised to restrict. The amount of MSG added to food has increased enormously since its first introduction. It is a commonly used food additive that enhances the taste of meat in particular and is used in appreciable quantities in the preparation of food in Chinese restaurants. MSG is now found in varying amounts in many food products, among them bacon bits, baking mixtures, bouillon cubes, bread stuffing, breaded foods, canned meats, cheese dips, corn chips, dry roasted peanuts, frozen dinners and pizza, gelatins, oriental foods, potato chips, pot pies, salad dressings, salt substitutes, seasonings, soups, soy sauce, and meat tenderizers.

As little as 3–10 mg of this compound may provoke headache in a migraine-prone person. The Chinese restaurant syndrome is well documented and includes

Table 6 Dietary suggestions for migraine patients[a]

No alcoholic beverage, particularly red wines
No aged or strong cheese, particularly cheddar cheese
Avoid chicken livers, pickled herring, pods of broad beans
Use monosodium glutamate (MSG) sparingly. Avoid foods containing MSG
Avoid cured meats such as hot dogs, bacon, ham, and salami, if these can be demonstrated to evoke migraine. Avoid foods containing nitrate and nitrite
Eat three well-balanced meals per day. Avoid skipping meals, prolonged fasting, or excessive ingestion of carbohydrates at any single sitting
Avoid chocolate

[a]All of these foods have been demonstrated to contain vasoactive materials of varying types and all have, in specific circumstances, evoked migraine in predisposed patients.

symptoms of tightening of the muscles of the face and throat, constriction of muscles of the back and neck, paresthesia of the mucous membranes of the mouth and palate, nausea, giddiness and dizziness, flushing and sweating of the face, severe pain in jaws, and a band-like headache. The syndrome occurs within 20–30 min after eating foods with a significant amount of MSG.

Treating Migraine Headache

When having severe headaches, it is important to see a doctor for evaluation and diagnosis. Several therapeutic methods, both nonpharmacologic and pharmacologic, may prove useful. The approach to pharmacologic treatment is either abortive (acute) or prophylactic (preventive). The ergotamines remain the agents of choice for abortive therapy of migraine. For greatest benefit, these preparations must be taken either during the aura or as soon as possible after the headache starts. If a patient experiences more than two migraines a month or abortive agents are ineffective or contraindicated, prophylaxis should be considered. A variety of agents have been shown to be useful, but proptanolol is the safest and most effective. Patients on preventive medication still require acute treatment for breakthrough attacks.

Nonpharmacologic techniques are oriented primarily to reducing the frequency as well as the severity of migraine attacks. The first step in prevention is the elimination of triggering factors, which should be reviewed thoroughly with each patient. A detailed diary of when you have a headache (date, time, severity, possible triggering factors, time and amount of exposure, treatment used, the outcome) may help determine what may influence your headaches and could give you an idea of what triggers you may want to eliminate or avoid. It is important that time be taken to identify those sensitivities and eliminate only those foods related to the patient's migraine. Many times patients are given lists admonishing them to avoid all foods that have ever been implicated in migraine causation. This can create apprehension about eating, and the resultant avoidance of a variety of foods can lead to malnutrition and cachexia in the conscientious patient. An overconcern with diet can also be a psychological stressor that may further aggravate the headache problem. A diary can also help you see how well your medication and lifestyle changes are working.

Most experts in headache suggest that foods that have vasoactive properties should be avoided (Table 6). It is also important to eat a healthy diet, at regularly scheduled times, not skipping a meal. In addition, patients should be consistent in their sleeping patterns. There should be a regular bedtime and patients should wake up at the same time each morning. Stress and anxiety management strategies, such as regular exercise, relaxation techniques, biofeedback, or cognitive-behavioral therapy may also have a role in treatment.

See also: **Alcohol**: Properties and Determination; **Amines**; **Caffeine**; **Copper**: Properties and Determination; Physiology; **Nitrates and Nitrites**; **Premenstrual Syndrome: Nutritional Aspects**; **Taste Enhancers**

Further Reading

Diamond S, Prager J and Freitag FG (1986) Diet and headache. Is there a link? *Headache* 79(4): 279–285.

Goadsby PJ, Lipton RB and Ferrari MD (2002) Drug therapy: migraine – current understanding and treatment. *New England Journal of Medicine* 346(4): 257–270.

Harrison DR (1986) Copper as a factor in the dietary precipitation of migraine. *Headache* 26: 248–250.

Hasselmark L, Malmgren R and Hannerz J (1987) Effects of a carbohydrate-rich diet, low in protein-tryptophan, in classic and common migraine. *Cephalalgia* 7: 87–92.

Journal of the American Medical Association Patient page. Available at www.ama-assn.org/consumer.htm.

National Headache Foundation Available at www.headaches.org.

National Institute of Neurological Disorders and Stroke Headache. Available at www.ninds.nih.gov.

Peatfield RC (1995) Relationship between food, wine, and beer-precipitated migrainous headaches. *Headache* 35(6): 355–357.

Peatfield RC, Glover V, Littlewood JT, Sandler M and Clifford Rose F (1984) The prevalence of diet-induced migraine. *Cephalalgia* 4(3): 179–183.

Perkin JE and Hartje J (1983) Diet and migraine: a review of the literature. *Perspective in Practice* 83(4): 459–463.

Raskin NH and Knittle SC (1976) Ice-cream headache and orthostatic symptoms in patients with migraine. *Headache* 16(5): 22–25.

MILK

Contents

Liquid Milk for the Consumer

J Manners, University of Melbourne, Parkville, Victoria, Australia
H Craven, Food Science Australia, Werribee, Victoria, Australia

Introduction

Historically, milk from cows and other mammals has been an important source of nutrients, especially protein for humans. Traditionally, villages and individual families tended one or two animals such as cows, goats, or buffalo, which provided milk for immediate consumption and then meat when the animal was slaughtered. Over time, the number of animals in a milking herd has increased from the one-animal situation to herds of tens and hundreds of animals. With this increase in size of herds, there have been developments in how the animals are milked and how the milk is handled from milking to receival at the milk factory.

Composition of Milk

Milk is a complex biological fluid consisting of fats, proteins, minerals, vitamins, enzymes, and sugar. The composition of milk varies according to the breed of the animal, the genetic background of the animal, the stage of lactation, the nutritional quality of the animal's feed, the frequency of milking, the incidence of disease such as mastitis, and general environmental conditions.

Animals from different species and breeds produce milk of varying fat, protein, and lactose content and volume. For example, Friesian or Holstein animals are more likely to produce larger quantities of milk with a lower percentage of fat compared to milk produced by a Jersey animal. Buffaloes produce milk with a higher fat and protein content than cows, while goats' milk has a lower lactose content compared to cows' milk. Within breed types, animals are largely selected for breeding on the basis of their performance as producers of milk measured in terms of fat, protein, and volume.

The greatest changes in composition occur during early lactation. Colostrum, the first milk after calving, has a high concentration of fat and protein, especially immunoglobulins, and a low concentration of lactose. Over the first weeks of lactation, the fat and protein levels decrease while the lactose levels increase. **Table 1** compares the composition of colostrum with that of milk produced 10 days later. The volume of milk produced by the cow increases during the early stages of lactation and after peaking slowly declines towards the latter stages of lactation. This decline usually coincides with the development of pregnancy in the animal.

Clinical or subclinical mastitis in the udder results in changed composition of the milk with decreased levels of lactose and potassium and increased levels of sodium, chloride, and serum proteins. In addition, a large number of somatic cells and bacteria are present. The higher number of somatic cells results in an increased concentration of the enzyme plasmin in the milk. Plasmin is a protease which is normally found in the blood stream. The industry commonly uses the

Table 1 Typical composition of milk at the start of lactation and 10 days later

Component	*Start of lactation concentration (g per 100 ml)*	*10 days after commencement of lactation: concentration (g per 100 ml)*
Fat	6.5	3.8
Protein		
Casein	5.0	2.6
Globulins	11.1	0.7
Lactose	2.9	4.9
Ash	1.2	0.8
Total solids	26.7	12.6

levels of somatic cells in milk as a screening test for the incidence of mastitis. A count of under 200 000 somatic cells per milliliter in bulk herd milk generally indicates an absence of mastitis, while levels above 400 000 per milliliter indicate a high incidence of mastitis in the herd.

The impact of feed consumed by the cow on the composition of milk is complex. Overall, animals on nutritionally superior feed will produce more milk with a higher level of fat and protein than animals on feed of lower nutritional quality. The quantity and type of fat and the composition of the nonlipid compounds in the feedstuffs affect the fat content and composition.

In cold climates, animals are generally housed indoors and the feed is bought to them. The volume of production and composition of milk from cows housed indoors is likely to be different to that produced by animals in temperate or tropical climates which are fed by range-grazing. The difference in composition and volume may also be related to the nutritional quality of the feed.

Milking

Historically, milk has been extracted from the udder by manual stimulation and expression of milk from the teats. The milk is collected in a small storage vessel such as a bucket or can. Manual milking is commonly carried out where the numbers of cows are small and there is plenty of available labor. As herds have grown in size, mechanical means of extracting milk have been introduced. These are generally called milking machines. The milk is extracted from the teat and udder by a pulsating vacuum in rubber liners called teat cups which fit around the teat. In most situations the milk is carried from the cow through pipelines to the milk storage vessel. These vessels may range from a bucket for milk from an individual cow to vats or silos of large capacity. The silos may be used for the storage of the milk from the several milkings.

Herds are generally milked twice a day during most of the lactation period: the frequency is reduced to once a day as the milk yield declines. Research work has demonstrated that more frequent milking, such as three times a day, can increase the total volume of milk, but this is not a common practice amongst dairy farmers.

Control of Microbial Contamination

Milk from the udder of a healthy cow contains very few microorganisms. However, addition of microorganisms during the milking process and growth of contaminating microorganisms will reduce the time required for milk to spoil. To minimize the addition of microbes to the milk, the farmer needs to prevent contamination of milk through the process of milking and transport to storage vessels.

The exterior of the teat and udder must be carefully washed to remove residues of manure and dirt, which contain large numbers of microorganisms. It has been demonstrated that the optimal treatment to minimize addition of bacteria to milk is to dry the udder carefully with a clean cloth after washing. In some situations such as large herds, strategic washing of udders is carried out. This involves a judgment about the level of visual contamination on the udder.

Another important source of microbial contamination is the hygienic condition of the rubbers and pipelines of the milking equipment. The materials used for the transport and storage of raw milk must be easy to clean and sanitize. Stainless steel and plastic are easy to clean and are more resistant to mechanical damage than rubber, wood, and other materials. Fat and protein deposits in the pipelines and storage vessels will prevent sanitizer solutions acting on bacteria. In addition, cracked rubberware is difficult to clean and sanitize effectively.

Other sources of contamination are water supplies used for washing udders and equipment, exposure of the milk to the air, dirty cloths used for cleaning udders, and the presence of mastitis in the udder. The total microbial count of milk can increase dramatically when an udder is infected with mastitis.

Care must be taken when milking to minimize the possible transfer of mastitis infection from one animal to another. Precautions used by farmers include dipping the rubber teat cups in a chemical sanitizer solution such as iodophor, dipping the teats after milking with a disinfectant such as iodophor, use of good-quality running water to wash udders, and single-use towels to dry udders after washing. The incidence of mastitis infection can be increased by the use of malfunctioning milking machines. Many

mastitis control programs include regular maintenance checks of rubberware and mechanical parts.

After milking, the major concern is to minimize the growth of microorganisms during the storage and transport of milk from the farm to the factory. The temperature of the milk from the cow is about 37 °C, or body temperature. To minimize the growth of microorganisms, it is necessary to reduce the temperature to less than 5 °C in a short period. In many situations, the temperature is lowered on the farm using a water cooling system based on plate or tubular heat exchange before the milk enters the milk storage vats.

Storage

Many storage vats or silos are now equipped with refrigeration units which are able to reduce the temperature to less than 5 °C. In some areas, the milk may be stored in cans in a large refrigerator for 1–2 days before dispatch to the factory. In areas where refrigeration is not common, the milk is collected once or twice per day from the farm or may be taken by the farmer to a milk collection depot. The milk is generally cooled at the milk collection depot before dispatch to the factory. The microbiological quality of milk received at the factory is dependent upon the initial contamination of the milk during the milking process and subsequent transport to storage vessels, the temperature of the milk during storage, and the length of time of storage. Table 2 shows the number of microorganisms in milk of varying quality stored at different temperatures.

Milk stored at low temperatures for extended periods is more likely to have large numbers of psychrotrophic bacteria present. Some of these bacteria may produce heat-stable enzymes (lipases and/or proteases) which will survive pasteurization and could therefore cause flavor defects in the finished product. Research work has demonstrated that enzymatic activity is unlikely if the numbers of microorganisms are less than 1 million per milliliter. Conversely, if milk is not cooled until hours after milking, the number of microorganisms may increase to in excess of 1 million per milliliter. Cooling of the milk after microbial numbers have increased significantly will increase the shelf-life only marginally.

Table 2 Influence of storing milk for 24 h at different temperatures on the bacterial content of three milk samples with varying numbers of bacteria

Time of storage (h)	*Storage temperature (°C)*	*Standard plate count at 30 °C (CFU ml^{-1})*		
		Sample A	*Sample B*	*Sample C*
0		1500	45 000	260 000
24	5	1600	46 000	660 000
24	10	10 500	126 000	8 900 000
24	15	220 000	860 000	32 000 000
24	20	580 000	1 570 000	128 000 000

It has been suggested that an alternative means for extending the shelf-life in regions without adequate refrigeration is to use the antibacterial activity of the lactoperoxidase enzyme system. Small quantities of hydrogen peroxide are added to milk to activate the lactoperoxidase system. The technique has been trialled successfully in Kenya, Sri Lanka, and Mexico.

Key factors in the successful use of the system to reduce the frequency of pick-up to once per day include relatively low numbers of microbes and reduction of the storage conditions for milk to less than 15 °C. Other techniques that have been demonstrated to extend the shelf-life of raw milk are thermization and addition of carbon dioxide.

One means of extending the shelf-life of milk on the farm is to separate the cream from the skim milk. The skim milk may be used to feed animals such as pigs while the cream is transported to the factory once or twice per week.

In countries that experience severe winters, steps must be taken to minimize the likelihood of the milk freezing. Freezing and thawing of milk in uncontrolled conditions may result in disruption of the fat globules and coagulation of protein, which will cause problems with subsequent processing and manufacture.

Transport of Milk from Farm to Milk Collection Depot

Milk sold by the farmer for manufacturing and processing is transported from the farm to the milk collection and receival depot in a number of ways. In countries with many small herds, the farmer or householder carries the milk to a milk collection center. Here the milk is graded for organoleptic quality, weighed, and then cooled before transport to the larger milk manufacturing factory. The number of milk collection centers will depend upon the density of the herds and the relative ease of transport from the farm. The milk collection center may be cooperatively owned by the farmers or owned by the milk factory. In some countries, the milk is purchased by a middleman from the farmers and transported to the factory.

In areas where farms produce large volumes of milk and the means of transport is more advanced, the factory may collect the milk from the farm on a twice-daily, daily, or alternate days basis. The frequency of collection will depend upon the storage conditions for the milk on the farm, the volume of

milk, and the distance from the factory. If the milk is not cooled on the farm it should be cooled immediately on arrival at the factory before storage. A road tanker of varying size may be used to pick the milk up or a tray truck able to carry cans or churns. Insulated tankers will be used for the transport of refrigerated milk. On arrival at the farm the driver usually measures the volume of milk and assesses whether the milk is organoleptically acceptable for pick-up. Samples of milk may also be taken for testing at the factory for composition and quality parameters.

Testing for Payment and Quality

A range of tests are carried out on raw milk supplies. A farmer may be paid for milk according to the volume of milk supplied, the quantity of fat, and other components such as protein and whether the milk is contaminated by added water, excessive numbers of bacteria, residues of antibiotic and other chemicals, somatic cells, and extraneous matter such as dirt. The payment systems vary around the world. In many countries, regulatory bodies or factories may establish minimum quality levels or standards for milk. The analytical techniques used to test for quality parameters will depend upon the quality standard, the number of suppliers, the accuracy required for results, and available technology. The use of automated instruments for measurement of milk composition, somatic cells, and bacterial numbers enables faster feedback of results to the farmer.

Payment

Historically, volume and fat content have been used for payment purposes. Where milk is used for consumption in a liquid form, the farmer is normally paid on volume basis provided the milk contains a minimum content of fat and solids-nonfat (SNF) and meets other defined quality standards. In some countries, where milk is used for manufacture of product, it is purchased on a fat content basis while other countries have introduced compositional payment based upon fat and protein content. Simple tests to determine the fat content of milk have been available for many years but development of instrumentation such as infrared equipment to determine SNF and protein has enabled companies to test for fat, protein, or SNF and volume on a daily basis. The frequency of testing milk for payment purposes depends upon the local situation.

Quality

Over the years, the quality parameters used to determine whether milk is fit for sale by a farmer have increased from simple sensory evaluations using sight and taste to complex analyses for the presence of chemical residues.

In developed dairy industries, where herds are large and the milk is used primarily for manufacturing purposes, the tests used to determine the quality of milk – apart from compositional analysis – include total bacterial numbers, presence of antibiotic residues, presence of added water, numbers of thermoduric bacteria, coliforms and other bacterial groups, presence of extraneous matter or sediment, number of somatic cells, residues of iodide, pesticides, and other chemicals. Table 3 summarizes typical standards for the different quality parameters.

Table 3 Typical standards for different quality parameters of raw milk

Plate count	*50 000–300 000 colonies ml^{-1}*
Freezing point	−0.548 to 0.520 °C
Antibiotic residues	0.0025–0.1 μg ml^{-1}[a]
Methylene blue reduction	> 5 h
Somatic cells	200 000–400 000 ml^{-1}
Thermoduric bacteria	1000–30 000 ml^{-1}

[a]Penicillin G-equivalent.

In other countries, milk quality tests, such as alcohol stability, methylene blue reduction, or resazurin, are used to determine the extent of microbial contamination while specific gravity determination may detect adulteration by water and other substances. Generally, failure to comply with quality standards results in a penalty such as rejection of the milk for a specified period or a payment discounted from the normal payment level.

Potential Public Health Risks

The consumption of unpasteurized milk by humans poses potential risks from the following factors: transmission of disease from the animal, transmission of disease from a human milker or handler of milk before consumption, or growth of undesirable organisms in the milk. Pathogenic bacteria which may be excreted in the milk by diseased animals include *Mycobacterium tuberculosis, Brucella abortus, Staphylococcus aureus, Streptococcus agalactiae, Escherichia coli, Listeria monoctyogenes, Leptospira* spp. and *Coxiella burnettii*. In many countries, stringent veterinary surveillance of herds has significantly reduced the risk of tuberculosis and brucellosis transmitted to humans by milk. With the exception of *C. burnettii*, which requires a higher heat treatment than normal pasteurization conditions, all other organisms are killed by effective pasteurization. A number of cases of food-poisoning

outbreaks caused by *Campylobacter* strains and *Listeria monoctyogenes* have been traced to the consumption of raw or improperly pasteurized milk.

Human handlers of milk may cause problems if they are clinically or subclinically infected with *Salmonella* spp., particularly *Salmonella typhi*, or *Streptococcus pyogenes*, and contaminate the milk during the milking, storage, and transport process. Heat treatment of the milk equivalent to pasteurization will eliminate milk as a likely vehicle for food poisoning by these organisms. Many health authorities recommend the scalding of raw milk by bringing milk to the boil before consumption in the home. This form of heat treatment kills microbes with the potential of causing disease and also increases the shelf-life by killing many bacteria which will spoil the milk.

Care needs to taken with the quality of feedstuffs, including grains, to prevent the possibility of contamination of the milk by aflatoxins.

Direct Sales from the Farm to the Public

The practice of direct sales of nonheat-treated milk from the farm to the public will depend upon the country. In some countries direct sales are prohibited while in other countries the practice may be permitted if the farmer has the herd regularly checked by a veterinarian for absence of specific diseases and carries out strict hygienic practices during milking, storage, and packaging. Some farmers may heat-treat the milk using a batch or high-temperature, short-time (HTST) pasteurization system, while other farmers sell raw or unpasteurized milk. The degree of surveillance by health authorities of farms that sell raw or heat-treated milk for human consumption varies from regular to no inspections or surveillance.

See also: **Aflatoxins**; **Buffalo**: Milk; ***Campylobacter***: Properties and Occurrence; ***Escherichia coli***: Occurrence; ***Listeria***: Properties and Occurrence; **Milk**: Processing of Liquid Milk; Physical and Chemical Properties; Analysis; Dietary Importance; ***Mycobacteria***; **Pasteurization**: Principles; **Sheep**: Milk; **Spoilage**: Bacterial Spoilage; ***Staphylococcus***: Properties and Occurrence

Further Reading

Banks W and Dalgleish DG (1990) Dairy Microbiology. In: Robinson, RK (ed.) *The Microbiology of Milk*, 2nd edn. pp. 1–36. London: Elsevier.

Bramley AJ and McKinnon CH (1990) The Microbiology of Raw Milk. In: Robinson RK (ed.) *The Microbiology of Milk*, 2nd edn, pp. 163–208. London: Elsevier.

Fox PF and McSweeny PLH (1998) *Dairy Chemistry and Biochemistry*. London: Blackie Academic and Professional.

Harding F (ed.) (1995) *Milk Quality*. London: Blackie Academic and Professional.

IDF (1990) *Milk Collection in Warm Developing Countries*. Brussels: International Dairy Federation.

IDF (1994) *The Significance of Pathogenic Microorganisms in Raw Milk*. Brussels: International Dairy Federation.

IDF (1995) *Milk Payment System for ex Farm Milk, Antibiotic Testing, Mastitis Control*. Bulletin 305. Brussels: International Dairy Federation.

IDF (1996) *Bacteriological Quality of Raw Milk*. Brussels: International Dairy Federation.

IDF (1997) *Residues and Contaminants in Milk and Milk Products*. Brussels: International Dairy Federation.

La Fondation De Technologie Laitière du Quebec (1985) *Dairy Science and Technology*. Quebec: Les Presses de l'Université Laval.

Marth EH and Steele JL (eds) (1998) *Applied Dairy Microbiology*, pp. 1–64. New York: Marcel Dekker.

Varnam AH and Sutherland JP (1994) *Milk and Milk Products Technology, Chemistry and Microbiology*. London: Chapman and Hall.

Walstra P, Geurts TJ, Noomen A *et al*. (1999) *Dairy Technology Principles of Milk Properties and Processes*, pp. 1–187. New York: Marcel Dekker.

Welch RAS, Burns SR, Davis SR *et al*. (eds) (1997) *Milk Composition, Production and Biotechnology*. New York: CAB International.

Wong NP (ed.) (1988) *Fundamentals of Dairy Chemistry*, 3rd edn. New York: Van Nostrand.

Processing of Liquid Milk

J Manners, University of Melbourne, Parkville, Victoria, Australia
H Craven, Food Science Australia, Werribee, Victoria, Australia

Background

Advances in technology and the demand by larger communities of people for a regular supply of various types of milk have led to the development and application of a range of processes for liquid milk. These include a defined heat treatment to minimize changes to the product while eliminating pathogenic bacteria and extending shelf-life, addition of supplements to improve nutritional value or flavor, and modification of composition to standardize consistency or improve dietary content.

Types of Products

Traditionally, the principal product sold as liquid milk was whole milk, heat-treated and packed in a

returnable container such as glass bottles. In recent years, changes in customer demands and technological advances have increased the range of liquid milk products available. Major types of liquid milk range from skim or nonfat milk, reduced- or low-fat milk, high-fat milk to reduced-fat, high-protein products. These may be sold as natural-flavored milk or with added flavors such as chocolate, coffee, and strawberry. The products may be heat-treated using conventional pasteurization or ultrahigh temperature (UHT) processing. Flavored milk products usually contain added sugar, coloring and flavoring agents and additives such as stabilizers. The range of flavors added to milk depend on the cultural background of the market. Chocolate-flavored milk is a popular product in many countries. Milks mixed with fruit juice or fruit pulp have been marketed successfully in some countries, while the consumption of drinking yogurt is growing. The typical compositions for different milk products are compared in Table 1. (*See* **Fermented Milks**: Types of Fermented Milks; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; **Stabilizers**: Types and Function.)

Milks may be sold with added vitamins, such as vitamins A and D, to increase the dietary intake of essential nutrients. In recent years, with increased awareness of the risks of osteoporosis, calcium has been added to some milk products. Low-lactose milk made by the addition of the enzyme, lactase, to milk before or after heat treatment is gaining a market share in markets with a large number of lactose-intolerant consumers. (*See* **Calcium**: Properties and Determination; **Cholecalciferol**: Physiology; **Dairy Products – Nutritional Contribution**; **Food Intolerance**: Lactose Intolerance; Milk Allergy; **Lactose**; **Osteoporosis**; **Retinol**: Physiology.)

Other new products incorporate *Lactobacillus acidophilus* and *Bifidobacterium* spp. in liquid milk to assist those with sensitive gastrointestinal tracts. In Australia, sales of pasteurized ultrafiltered low-fat milk have increased steadily since the initial market launch. (*See* ***Acidophilus*** **Milk**; ***Bifidobacteri*** **in Foods**; **Lactic Acid Bacteria**; **Membrane Techniques**: Principles of Ultrafiltration; **Probiotics**.)

In many countries, the new types of milk products have steadily eroded the market share of whole milk over time to the situation in Canada and The Netherlands, where reduced-fat milk accounts for more than half of the total milk sales.

Table 1 Typical chemical composition (g per 100 g of product) for different milk products[a]

Product	*Fat*	*Solids-nonfat*	*Total solids*
Whole milk	3.8–4.0	> 8.5	> 12.3
Skim milk	0.1	> 8.5	> 8.6
High-fat milk	5.0–6.9	> 8.5	> 13.5
Reduced-fat milk	1.0–2.0	> 8.5	> 9.5
Reduced-fat, high-protein milk	1.0–2.0	11.3	12.3–13.3

[a]Flavored milks may be made by adding sucrose, color, and flavors to the typical formulations described.

Collection and Storage of Raw Milk

As many milk-processing dairies are located near the market in urban areas, the milk is transported in large-capacity tankers to the processing dairy from milk collection depots in the farming areas. The milk is usually cooled at the depot before dispatch to the processing dairy. Some processing dairies may receive milk directly from farms.

On arrival at the milk-processing dairy, the milk is inspected for quality attributes such as temperature, flavor, smell, and presence of extraneous matter and volume. Samples of milk may be taken to determine fat and solids-nonfat content, numbers of bacteria, freezing point and presence of antibiotic residues. The actual tests performed on incoming supplies of milk depend on a company's quality program for inspection of raw materials and the requirements of regulatory authorities. (*See* **Milk**: Liquid Milk for the Consumer.)

After the tanker load has been accepted for use, the milk is pumped into a storage vessel designated for raw milk. The milk may be filtered during the pumping process to reduce the level of extraneous matter. It may be necessary to cool the incoming milk to 5 °C to minimize the growth of contaminating microorganisms. Plate heat exchangers are commonly used to cool the milk before storage. In many countries, the temperature of milk during storage and the allowed time for storage before pasteurization are limited by regulatory requirements. However, storage conditions of less than 5 °C and 48 h maximum are generally accepted as good manufacturing practice by the industry across the world. Storage of milk above 5 °C and extended times of storage encourage the growth of psychrotrophic bacteria, which may cause organoleptic defects.

The milk in the storage vessels must be agitated slowly to prevent the separation of fat. Mechanical agitation or passage of compressed air through the milk is commonly used by the dairy industry. Care must be taken during agitation to minimize the possibility of lipolysis caused by disruption of the milk fat globules.

Preparation

Before heat treatment and packaging, it may be necessary to prepare the raw milk for processing. Regulatory authorities in some countries permit the milk depot to standardize the milk to desired fat and solids level before dispatch to the processing dairies. In other countries, this form of preparation is not permitted. For skim milk, the fat is removed through the separation process while low-fat, high-solids products may require separation and the addition of milk solids in the form of skim milk powder or skim milk concentrate. In the preparation of flavored milk, sugar, flavors, colors, and other additives are added to the milk in the desired quantities. Often, these ingredients are purchased in a dry, premixed form, and blended with a small quantity of milk or water before mixing with the raw milk. Sometimes, the syrup mixture may be injected into the milk after pasteurization.

Recombination

Where the supply of raw milk fluctuates or is insufficient to meet the market needs, milk fat in the form of butter oil and milk solids in the form of full cream or skim milk powder may be used to prepare reconstituted or recombined milk. Recombination may also be used where the available milk supply has a high fat content. The butter, unsalted butter, and milk powders must be selected carefully and stored to ensure proper mixing of the materials and an absence of undesirable flavors in the finished product. The added water must be of a drinkable quality. In some countries, vegetable fats may be used instead of milk fat to prepare recombined milk. (*See* **Recombined and Filled Milks**.)

Heat Treatment

Milk is heat-treated to kill many of the microorganisms present in raw milk. The heat treatment makes the milk safe to drink and extends the potential shelf-life. Several different forms of heat treatment are used. These are pasteurization, UHT, and *in-container* sterilization. Pasteurization is the most commonly used form of heat treatment. The shelf-life of the milk depends on the heat treatment used in the process. (*See* **Heat Transfer Methods**; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; **Pasteurization**: Pasteurization of Liquid Products; **Sterilization of Foods**.)

Pasteurization

This form of heat treatment evolved from Pasteur's discovery that heating wines to 50–60 °C destroyed many of the organisms responsible for spoilage. With milk, a temperature–time combination sufficient to kill the organism, *Mycobacterium tuberculosis*, without adversely affecting the flavor, color, and nutritional value is used. (*See* ***Mycobacteria***.)

The two principal forms of pasteurization now used by industry throughout the world are *batch* pasteurization, where milk is held at a determined temperature for a long time, and *high-temperature, short-time* (HTST) pasteurization using a higher temperature for a relatively short period. Typical temperature–time combinations are 62–63 °C for 30–35 min for batch pasteurization and 72 °C for 15 s for HTST pasteurization. The temperature–time combination may need to be increased for effective pasteurization of products such as flavored milk with higher levels of solids, or reduced for skim milk to minimize damage to flavor and color. After heating, the milk is cooled immediately to 5 °C or lower.

The batch pasteurization process is carried out using slow agitation of milk in a jacketed vat, which can be heated and cooled. The cost of the equipment is relatively low, but the labor and energy costs are higher compared with the HTST system.

In the modern processing dairy with a large throughput, the HTST pasteurization process is favored, as the throughput can be as high as 50 000 l h^{-1}, and energy costs are reduced by using the heated milk to warm incoming milk in the regeneration section of the pasteurizer. The heated milk is cooled after pasteurization in the regeneration section.

The plate heat exchanger has historically been used for HTST pasteurization (Figure 1). The plate design offers a large surface area for heat transfer in a relatively compact area. Once the milk reaches the desired temperature of 72 °C, it enters a holding tube designed to ensure that all milk particles reside in the tube for at least 15 s. A temperature probe is located at the end of the holding tube. If the temperature falls below the desired temperature, a flow diversion valve is opened, which directs underprocessed milk back into the balance tank. The pasteurizers need to be checked regularly to ensure that there is no risk of raw milk contaminating the pasteurized product.

Heat exchangers designed using a tubular system have become increasingly popular in recent years. The heat is transferred from heated outer tubes to product in inner tubes. The tubular system is considered to be cheaper to maintain and can handle products with a higher viscosity compared with a plate heat exchanger.

In recent years, the design of pasteurizers and control systems used on pasteurizers has been improved in response to public health problems thought to be caused by inadequate pasteurization of milk.

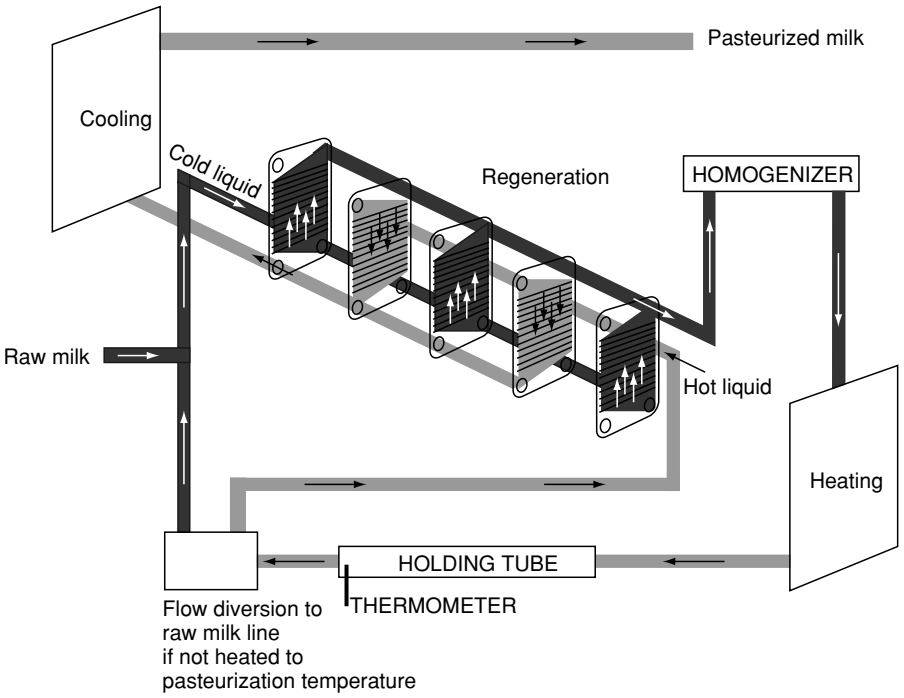

Figure 1 Schematic diagram of a plate heat exhanger pasteurizer showing milk flow.

Some processors use microfiltration in combination with pasteurization to extend milk shelf-life. The skim milk fraction is filtered to remove bacteria, sterile cream is added to adjust the fat content, and then the product is pasteurized. Microfiltration reduces levels of spoilage bacteria while avoiding flavor defects caused by high heat treatment of skim milk. Cream is not as susceptible to this problem. Canadian experience has demonstrated that the shelf-life of pasteurized milk can be extended from 12–18 to 32 days by using microfiltration.

UHT

The UHT treatment was developed to minimize damage to milk components caused by *in-container* sterilization while killing or inactivating all microorganisms, so there is little likelihood of microorganisms spoiling the product during storage and transport. The product is said to be 'commercially sterile.' The term 'UHT' may stand for ultraheat treatment or ultrahigh temperature.

UHT processing is being used increasingly for heat-treating milk, as the shelf-life is extended from days to months, and UHT milk can be stored and transported without refrigeration. The use of UHT treatment varies around the world. In Germany, France, and Spain, more than 50% of liquid milk is consumed in UHT form, whereas the process is rarely used in the USA.

The taste and appearance of UHT-treated milk are different from those of pasteurized milk. The milk may taste *cooked* or burnt to some consumers, because of the development of a sulfurous flavor during the heat treatment. The form of heat treatment used influences the intensity of the sulfurous flavor. However, the flavor has generally dissipated by the time of consumption, and the milk has a slightly brown color.

In some countries, the temperature and time conditions for UHT treatment are specified by regulation. Typical temperature–time conditions for UHT treatment of milk are 130–150 °C for 1–3 s. Products with increased solids will require a higher temperature or longer time to ensure effective heat treatment.

UHT treatment can be carried out using a number of different systems. These include direct means of heat treatment such as injection of steam into milk and infusion of milk into a steam chamber or indirect means using heat exchangers where the milk is separated from the heating medium. Scraped-surface heat exchangers may be necessary for UHT treatment of viscous products or products containing particulate matter. When steam is used for direct heat treatment, care must be taken to ensure that there is no dilution of the milk. The steam must also be suitable for contact with food products. All UHT equipment must be sterile to prevent recontamination of the milk with microorganisms. Products that have been UHT-treated must be packed aseptically to ensure optimum shelf-life.

The raw milk to be used for UHT treatment must be selected carefully to ensure that there is minimal

chance of heat-stable indigenous or microbial enzymes present causing gelation in the stored products.

Some dairies use equipment designed for UHT treatment to heat the product to 120–130 °C without a holding time. This is claimed to double the shelf-life of product packed under nonaseptic conditions compared with normal pasteurized milk.

In-container sterilization

The consumption of *in-container* sterilized milk has declined with the introduction of UHT processing and improved storage and distribution systems for pasteurized milk. The term 'sterilized' is used to describe milk that has been packed in airtight containers and subjected to temperatures in excess of 100 °C. The severe heat treatment of the milk results in a caramelized, burnt flavor and a brown color. The product is usually packed before heat treatment. If it is packed after heat treatment, the containers must be sterilized before use and then filled under aseptic conditions. Typical heating parameters for packaged sterlized milk are 115–120 °C for 10–20 min. The temperature–time conditions applied depend on the composition of the product, the type and dimensions of the package, and the desired degree of caramelization.

Homogenization

Many liquid milks need to be homogenized to prevent the separation of fat during storage and distribution. The large fat globules are broken up to form small globules that do not aggregate and rise to the surface. Homogenization also improves the mouth-feel of many products.

The homogenization step normally takes place after the preheat treatment in the pasteurizer. The normal temperature of milk required for effective homogenization is 60–70 °C. Heating the milk prior to homogenization helps to break up the fat globules and inactivates milk lipase that may act on the disrupted fat. The homogenization process is normally in two stages with 14 000–20 000 kPa pressure for the first stage to reduce the size of the fat globules and 3500 kPa for the second stage to disrupt and prevent the reformation of fat globules into large clumps. UHT milk may be homogenized before or after heating, although it is said that homogenization after UHT treatment gives a more stable product. (*See* **Homogenization**.)

Packaging

After heat treatment, liquid milk products are packed with minimal time delay. This helps reduce the likelihood of postpasteurization contamination that may affect the shelf-life.

Historically, returnable glass bottles were the containers commonly used, but technological advances have seen the introduction and regular use of cardboard cartons, plastic bottles and sachets, and laminates of foil, plastic, and cardboard. Many of these containers are *single-trip* and thus do not require specialized cleaning and sterilizing equipment in the dairy.

Selection of the type of package and packaging method requires consideration of many factors. The principal objective of packaging a product is to maintain the nutritional properties and quality of the product during storage, transport, and distribution to the consumer. Other factors to be considered when selecting packaging material include the safety of the packaging materials, compatibility of the product and package, hygiene, accuracy and mechanical efficiency of the packaging process, hygienic design of packaging equipment, marketing and promotion functions, consumer acceptance, energy consumption, environmental impact, and economics.

The sterilization of the packaging material and filling equipment for pasteurized products extends the shelf-life of the pasteurized milk product compared with the shelf-life of pasteurized milk packed under normal conditions.

When packaging UHT products, the equipment and packaging material must be sterilized to ensure that no microorganisms enter the product. This requires stringent cleaning and sanitation procedures and quality control checks. (*See* **Packaging**: Packaging of Liquids; **Plant Design**: Designing for Hygienic Operation.)

Storage and Distribution

This step is very important in the chain of events from raw materials to consumers, but its importance is often overlooked. In many countries, regulations may prescribe conditions for the temperature of the product during storage and distribution and for hygienic conditions in the storerooms and transport vehicles.

A major factor influencing the shelf-life of pasteurized milk is the temperature of the product after pasteurization. Many milk containers act as insulators; therefore, the temperature of the milk as it leaves the pasteurizer is critical to ensure that the pasteurized milk reaches the maximum shelf-life. The effective temperature of cool storage rooms is dependent upon factors such as the temperature of the external environment, the temperature of incoming product, the refrigeration capacity, and the isolation or separation of the cool room from the external environment. It is generally agreed that the temperature of milk should remain below 5 °C to ensure the maximum shelf-life. This

requires an efficient cold chain during the transport of milk from the pasteurizer to the customer.

In many countries, storage of milk at temperatures below 0 °C experienced during winter may result in freezing and subsequent damage to the milk during thawing. Steps must be taken to minimize the possible uncontrolled freezing and thawing of the product.

Care must be taken during storage and distribution to prevent possible contamination of milk with off-flavors caused by close contact with chemicals or other foods. Mechanical damage to the container may permit contamination by microbes and other material such as dust.

Factors that Influence Shelf-life

The shelf-life of milk products is defined as the period between manufacture or processing and when the consumer considers the product unsuitable for use. The product may be considered to be unsuitable for consumption, because of the presence of flavor defects and/or changes in physical appearance. (*See* **Storage Stability**: Shelf-life Testing.)

Factors determining the shelf-life of pasteurized milk include:

1. The presence of thermoduric bacteria in raw milk that are able to grow under refrigeration, e.g., *Bacillus cereus* and *B. circulans*. Processors using the extended shelf-life packaging technique often encounter problems with shelf-life owing to the presence of low numbers of *Bacillus* spp. in the raw milk. (*See* ***Bacillus***: Occurrence.)
2. The presence of psychrotrophic bacteria in the plant after the heat treatment step. These organisms grow under refrigeration and eventually cause spoilage as a result of their biochemical activity.
3. The temperature of the product after heat treatment. Ideally, milk should be stored between 1 and 2 °C to minimize the growth of psychrotrophic bacteria present in the milk from postpasteurization contamination. However, in many production situations, 1–2 °C is not practical, and the recommended temperature is 4–5 °C, with some countries permitting higher storage temperatures. Elevated temperatures allow bacteria to grow more quickly, thus reducing the time required for spoilage. (*See* **Spoilage**: Bacterial Spoilage.)
4. The presence of heat-stable enzymes produced by some species of psychrotrophic bacteria. Pasteurization and UHT treatments easily kill the bacteria, but the enzymes survive and are able to break down protein or fat. The problem is more pronounced in UHT products. Generally, a large number of bacteria must be present to produce the enzyme, and so low numbers of bacteria in raw milk will minimize this risk. (*See* **Spoilage**: Chemical and Enzymatic Spoilage.)
5. The action of lipase on damaged fat globules in raw milk results in rancid flavors in the finished products. Fat globules may be damaged during pumping and agitation. Maintaining the temperature of raw milk below 4 °C minimizes the lipolytic activity. The lipase is inactivated during pasteurization.
6. The general microbial growth in raw milk may result in damage to the milk solids as a result of metabolic activity. Subsequent handling of the milk may result in precipitation of the milk proteins or off-flavors. Numbers in excess of 5 million per milliliter are generally required for this to be a major concern.
7. Excessive temperatures (> 75 °C) during pasteurization are likely to induce the germination of microbial spores present in the raw milk; these microbes are then able to grow under low temperatures.
8. The action of heat-stable plasmin on proteins in milk. This causes flavor defects and age gelation in UHT products. Plasmin concentration increases with time during the lactation period, the degree of mastitic infection and the age of the cow.

Many consumer complaints relate to the presence of *off-flavors* in the milk. Frequently, these off-flavors may be caused by exposure of milk to sunlight or artificial light. One form of off-flavor, called *activated* flavor, occurs quickly on exposure to sunlight. The intensity of the off-flavor decreases over time. The light-induced oxidized flavor is very strong and objectionable, and does not decrease. Many dairies now use opaque packaging material to reduce the frequency of complaints from this problem.

Quality Assurance and Quality Control

Effective programs for the manufacture of safe, high-quality milk products are based on preventative systems for the control of microbial and other contamination. Many dairies now implement Hazard Analysis Critical Control Point-based programs to identify and manage points in the process where hazards can occur, including where pathogenic organisms can survive and proliferate. Key areas of hygienic control for the safe production of milk include raw material quality, adequate heat treatment, avoidance of postpasteurization contamination, and limiting the growth of microorganisms during storage prior to consumption. These practices will not only ensure the microbial safety of milk products but also assist in the prevention of milk spoilage.

Programs to ensure that the consumer receives milk with acceptable properties, including safety and shelf-life, include the following key elements:

1. Inspection of incoming raw materials for organoleptic and compositional properties, plus the presence of contaminants such as added water, extraneous matter, residues, and bacteria.
2. Monitoring of milk temperature during storage and heat treatment using chart recorders.
3. Monitoring of effectiveness of pasteurization heat treatment using the alkaline phosphatase test and operation of the diversion flow valve.
4. Monitoring of effectiveness of the homogenization process.
5. Testing of product to ensure compliance to specifications.
6. Monitoring the effectiveness of cleaning and sanitation programs. This may involve the application of tests to ensure that detergents and sanitizers are used at the correct strength, temperature, and time period, visual checks or swabs for the presence of milk residues, and tests for bacterial indicators such as Gram-negative psychrotrophs to determine the presence of postpasteurization contamination. (*See* **Sanitization**.)
7. Monitoring the volume of packed containers.
8. Monitoring the effectiveness of formation and sealing of containers, including the presentation of the pack.
9. Monitoring the control of temperature and housekeeping practices during the storage and distribution of the finished product.

A complete quality-management system will depend upon the processing requirements of individual dairies. The selection of process control tests should recognize the need for quick results to enable an operative to take corrective action, rather than for analytical precision. (*See* **Hazard Analysis Critical Control Point; Quality Assurance and Quality Control**.)

See also: **Hazard Analysis Critical Control Point**; **Heat Transfer Methods**; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; **Homogenization**; **Milk**: Liquid Milk for the Consumer; **Packaging**: Packaging of Liquids; **Pasteurization**: Pasteurization of Liquid Products; **Plant Design**: Designing for Hygienic Operation; **Quality Assurance and Quality Control**; **Recombined and Filled Milks**; **Sanitization**; **Sterilization of Foods**; **Storage Stability**: Shelf-life Testing

Further Reading

Burton H (1988) *Ultra-high-temperature Processing of Milk and Milk Products*. London: Elsevier.

Champagne CP, Laing RR, Roy D, Mafu AA and Griffiths MW (1994) Psychrotrophs in dairy products: their effects and their control. *Critical Reviews in Food Science and Nutrition* 34: 1–30.

Fox PF (ed.) (1995) *Heat-induced Changes in Milk*, 2nd edn. Brussels: International Dairy Federation.

International Dairy Federation (1986) *Monograph on Pasteurized Milk, Document 200*. Brussels: International Dairy Federation.

International Dairy Federation (1987) *Factors Affecting the Keeping Quality of Heat Treated Milk, Document 130*. Brussels: International Dairy Federation.

International Dairy Federation (1988) *Recombination of Milk and Milk Products*. Brussels: International Dairy Federation.

International Dairy Federation (1994) *Recommendations for the Hygienic Manufacture of Milk and Milk-based Products, Document 292*. Brussels: International Dairy Federation.

International Dairy Federation (1995) *Technical Guide for the Packaging of Milk and Milk Products, Document 300*. Brussels: International Dairy Federation.

International Dairy Federation (1996) *Heat Treatment and Alternative Methods*. Brussels: International Dairy Federation.

Lewis MJ (1994) *Heat Treatment of Milk*. In: Robinson RK (ed.) *Modern Dairy Technology, vol. 1 Advances in Milk Processing*, 2nd edn., pp. 1–60. London: Chapman & Hall.

Marth EH and Steele JL (eds.) (1998) *Applied Dairy Microbiology*, pp. 405–460. New York: Marcel Dekker.

Society of Dairy Technology (1983) *Pasteurizing Plant Manual*. Huntingdon, UK: Society of Dairy Technology.

von Bockelmann B (undated) *Long Life Products from Raw Materials to Finished Products*. Lund: Tetra Pak International.

Walstra P, Geurts TJ, Noomen A, Jellema A and van Boekel MASS (1999) *Dairy Technology, Principles of Milk Properties and Processes*. New York: Marcel Dekker.

Physical and Chemical Properties

D Otter, New Zealand Dairy Research Institute, Palmerston North, New Zealand

Introduction

Milk comes from a wide variety of animal sources, e.g., human, cow, buffalo, goat, and sheep, and is an important source of nutritional and physiological components. The gross composition of milk from all these sources has some similarities and this article will discuss only bovine milk. It has a very complex

Table 1 Chemical composition of bovine milk

Component	*Concentration (gl^{-1})*
Lactose	36–55
Fat	
Triacylglycerols	36–38
Diacylglycerols	0.1–0.23
Monoacylglycerols	0.006–0.015
Sterols	0.09–0.16
Sterol esters	Trace
Unesterified fatty acids	0.04–0.17
Hydrocarbons	Trace
Phospholipids	0.08–0.39
Protein	30–35
Caseins	24–28
α_{s1}-Casein	12–15
α_{s2}-Casein	3–4
β-Casein	9–11
κ-Casein	2–4
Whey	5–7
β-Lactoglobulin	2–4
α-Lactalbumin	0.6–1.7
Bovine serum albumin	0.2–0.4
Immunoglobulins	0.5–1.8
Casein fragments	
γ-Casein	1–2
Proteose-peptones	0.6–1.8
Milk fat globule membrane	0.4
Salt	0.7–0.8
Calcium	1.1–1.3
Chloride	0.9–1.1
Iron	0.3–0.6
Magnesium	0.09–0.14
Phosphorus	0.9–1.0
Sodium	0.35–0.9
Potassium	1.1–1.7

composition, comprised principally of water (88%), lactose (4.8%), fat (3.9%), protein (3.4%), and minerals (0.8%). These values however, can vary considerably depending on cow breed, and time of lactation, and between individual cows (Table 1).

Composition of Bovine Milk

Milk Fat

The fat (or lipid) in milk appears as an emulsion of globules dispersed in the milk serum that is held together by a very thin membrane derived from the plasma membrane of the mammary gland. The main functions of the fat are as an energy source for the neonate and a source of essential fatty acids (especially linoleic acid, $C_{18:2}$) and fat-soluble vitamins (A, D, E, K). Triglycerides typically make up approximately 98% of the total milk fat; the remainder consists of di- and monoglycerides, fatty acids, sterols, carotenoids, vitamins, and phospholipids. The membrane, in contrast, contains phospholipids, lipoproteins, cerebrosides, proteins, nucleic acids, enzymes, and trace elements (metals).

More than 400 distinct fatty acids have been detected in milk, with $C_{14:0}$ (myristic acid), $C_{16:0}$ (palmitic acid), $C_{18:0}$ (stearic acid), and $C_{18:1}$ (oleic acid) being the most abundant. The first three saturated fatty acids are solid at room temperature while the unsaturated oleic acid is liquid. The ratios of these fatty acids can therefore affect the hardness of the fat and any fat-based products. The fatty acids either come from the plasma lipids or are synthesized in the mammary gland. They therefore display marked patterns based on seasonal and dietary variations.

Esterification of fatty acids is not random, with the *sn*-1 and *sn*-2 positions on the glycerol comprising mainly myristic, palmitic, stearic, or oleic acid and the *sn*-3 principally butanoic ($C_{4:0}$), hexanoic ($C_{6:0}$) or oleic. The relative concentrations of C_4 and C_{18} fatty acids are important to keep the fat liquid at body temperature.

Proteins

Milk contains hundreds of different proteins, although most occur only in very small amounts. They are generally classified according to their chemical properties into caseins which is defined as those proteins, which precipitate from solution when the pH of milk is adjusted to 4.6, and whey proteins, which remain soluble under these conditions. In addition, milk contains the milk fat globule membrane proteins, which are found on the surface of the fat globules and are only released by mechanical action such as churning cream into butter; and a group of minor proteins/enzymes or protein-like material, i.e., the proteose-peptone and the nonprotein nitrogen (NPN) fraction. This last group is generally included in the whey proteins and includes casein fragments produced by posttranslational proteolysis or during casein processing.

Casein The caseins represent approximately 80% of the milk protein and consist of four subgroups: α_{s1}-, α_{s2}-, β-, and κ-casein. Each is heterogeneous and includes between two and eight genetic variants that differ from each other by one to 14 amino acids. The frequency of genetic polymorphism is breed-specific and can affect the processing properties of the milk proteins, e.g., cheese-making properties and the concentration of protein in milk. All the caseins are phosphorylated to varying degrees (between one and 13 phosphoserines) and κ-casein has up to nine glycosylated forms. α_{s2}- and κ-casein each contains two cysteines per molecule which form intermolecular disulfide bonds such that α_{s2}-casein exists as a dimer and κ-casein can exist as dimers through to decamers.

The α_{s1}-, α_{s2}-, β-, and κ-caseins contain 199, 207, 209, and 169 amino acids (mol. wt ~23.6 kDa, 25.4 kDa, 24 kDa, and 19 kDa, respectively), the composition of which is unusual in the following ways:

1. They contain high amounts of the apolar amino acids valine, leucine, isoleucine, phenylalanine, tyrosine, and proline (35–45%). The hydrophobic nature of these apolar amino acids is counteracted by the high content of phosphate and low content of sulfur-containing amino acids (methionine and cysteine) such that the caseins are reasonably soluble in water.
2. They all contain a high proline level that contributes to the low amounts of α-helix and β-sheet in their secondary structure and this makes them more vulnerable to proteolysis.
3. They have a very low sulfur amino acid content that lowers their nutritional and biological value to approximately 80% of the value for egg albumen.
4. They are rich in the essential amino acid lysine, and can therefore be used to supplement lysine-deficient food sources such as the cereal proteins and other plants. The high cationic lysine levels also make the caseins susceptible to nonenzymatic Maillard browning when exposed to heat and reducing sugars.

The high levels of polar and apolar amino acids in caseins manifest themselves as clusters in the primary structure, resulting in both hydrophobic and hydrophilic regions. This detergent-like arrangement, together with a high surface hydrophobicity and amphipathicity, gives the caseins good emulsification properties and helps with micelle stabilization.

The low amounts of α-helix and β-sheet give the caseins very little secondary or tertiary structure, resulting in 'an open, flexible, mobile conformation'. This lack of structure may give them stability against heat and other denaturing agents. It also renders them vulnerable to proteolytic attack and gives them high digestibility and hence nutritional value. They readily adsorb at air–water and oil–water interfaces, giving them good foaming and emulsifying properties that are used in the food industry.

There are also a number of minor caseins, the γ-caseins, which are produced from the C-terminal end of β-casein by proteolysis by plasmin, an indigenous proteinase. They occur at approximately 3% of the casein level but can be up to 10% depending on the stage of lactation and health of the cow. The remaining N-terminal segments of the β-casein comprise a large proportion of the proteose-peptone fraction. α_{s2}-Casein in milk and isolated α_{s1}-casein can also be hydrolyzed to a lesser degree by plasmin, although the degradation products have not been positively identified.

Micelle In milk the caseins interact with each other and with Ca^{2+} to form large colloidal complexes known as casein micelles. The micelle contains 63% water with a dry-matter composition of approximately 94% protein and 6% colloidal calcium phosphate (calcium, magnesium, phosphate, and citrate). Electron micrographs show that the micelles are roughly spherical, with a diameter between 50 and 500 nm, and evidence suggests that the micelles are made up of a large number of submicelles with diameters of less than 20 nm. The colloidal nature of the micelles results in light scattering that gives milk its whitish-blue tinge.

Whilst the definitive structure for the casein micelle has yet to be elucidated, there are a number of phenomena which help explain the stability of the micelle. Overall, calcium phosphate and hydrophobic interactions between the submicelles help form the micelles. The various caseins are distributed heterogeneously throughout the micelles, with κ-casein localized predominantly on the surface of the micelles. The solubility of the calcium salt of κ-casein thus helps to stabilize the calcium-sensitive α_{s1}-, α_{s2}-, and β-caseins. Micelle stability is further enhanced by the hydrophilic C-terminal of κ-casein that is situated outside the micelle complex, giving them a 'hairy' look under electron microscopy, and which contains the protein's carbohydrate moiety. The calcium ion concentration dictates the micelle size, with removal of colloidal calcium phosphate resulting in disintegration of the micelles.

The micelles are physically stable and can withstand all the processes to which milk is routinely subjected, such as heating (up to 140 °C for 15–20 min), compaction, homogenization, and high calcium ion concentrations (200 mmol l^{-1} for temperatures up to 50 °C). They are destabilized by alcohol (40% ethanol at pH 6.7) and freezing and aggregate and precipitate when the pH is reduced to 4.6 (the isoelectric point of the caseins), when the colloidal calcium phosphate concentration is reduced or when proteinases hydrolyze the C-terminal end of κ-casein.

Precipitation/Coagulation The precipitation of milk principally involves the precipitation of the casein proteins. It can be caused by a number of different mechanisms and is dependent on whether the casein is present in either the micellar or nonmicellar form.

The coagulation of micellar casein can be induced by lowering the natural pH of the milk using either an acid or by allowing the growth of an acid-producing

bacteria. As the pH is lowered to 4.9, the colloidal calcium phosphate in the casein micelle becomes soluble and forms calcium ions that can leave the micelle. As the pH is lowered further and reaches the isoelectric point of the casein proteins (approximately 4.6), the micelles aggregate and enlarge, creating a dense coagulum. This aggregation and precipitation is temperature-dependent.

Coagulation can also be caused by the hydrolysis of κ-casein by the acid proteinase, chymosin. This is the first step in the manufacture of most cheese. Chymosin is traditionally found in rennet that has been prepared from calf stomachs, although the calf chymosin gene has also been cloned into a number of microbial vectors for use in vegetarian cheeses. Chymosin specifically cleaves the peptide bond between the amino acids 105 (phenylalanine) and 106 (methionine), producing para-κ-casein (κ-CN f1-105) and glycomacropeptides (f106–169). Glycomacropeptides, which are soluble, hydrophilic, and contain the different carbohydrate moieties, diffuse into the whey, creating imbalances in the intermolecular forces in the micelles that then aggregate, forming a coagulum. Hydrophobic interactions then occur that lead, in the presence of calcium ions, to water expulsion from, and shrinkage of, the coagulum (syneresis).

The initial peptide bond cleavage is known as the primary phase of rennet action, coagulation is the secondary phase, and subsequent rennet action on the casein components during cheese-ripening is called the tertiary phase. The extent of the first two phases is governed by the pH and temperature and the secondary phase is calcium ion-dependent.

Micelle destabilization and coagulation also occur in the presence of approximately 40% ethanol at pH 6.7 and at lower ethanol concentrations when the pH is lowered.

Whey Proteins Whilst the caseins can be considered as a related family of proteins, the whey proteins are more diverse and encompass any protein or protein fragment, e.g., from casein degradation, not removed from skim milk by precipitation at pH 4.6 (i.e., not a casein). These soluble proteins (or milk serum proteins) have a well-ordered, globular structure and are generally not precipitated at their isoelectric points but are more heat-labile than the caseins.

β-Lactoglobulin, the major bovine whey protein, contains 162 amino acids with an approximate mol. wt of 18 kDa and is only found in ungulates, although a closely related protein occurs in other mammals. More than four genetic variants have been observed with the A and B variants being the most prevalent. β-Lactoglobulin is rich in sulfur amino acids, containing five cysteine residues, and has a high biological value of 110. At physiological pH it generally exists as a dimer of mol. wt approximately 36 kDa linked by up to three disulfide bonds.

At temperatures above 60 °C denaturation of β-lactoglobulin can lead to the exposure of the reactive sulfur amino acids that can then form disulfide bridges with κ-casein and/or α-lactalbumin. As the temperature is increased, other sulfurous compounds such as hydrogen sulfide are produced, leading to the 'cooked' flavor of heat-treated milk. β-Lactoglobulin also has a large number of lysine amino acids that can undergo lactosylation, which alters the physicochemical properties of β-lactoglobulin.

The principal biological function of β-lactoglobulin is presently unknown but it can act as a carrier for vitamin A (retinol). It can also stimulate lipolysis by binding the free fatty acids which inhibit lipases.

α-Lactalbumin is found in the milk of most mammals; it is the principal protein in human milk, and is thus considered to be the 'typical' whey protein. It is a compact globular protein of 123 residues (mol. wt ~14 kDa) and has three genetic variants, although the milk of western cattle contain only α-lactalbumin B. It is relatively rich in tryptophan and the sulfur amino acids cysteine and methionine. It contains four intramolecular disulfides, exists as a monomer and has 54 of its amino acids identical to the corresponding residues in lysozyme. Up to 15% of α-lactalbumin has been shown to be glycosylated to varying degrees. It is also a metalloprotein, binding one Ca^{2+} per mole (this increases its heat-stability).

α-Lactalbumin has an essential physiological role in lactose synthesis in the udder. Although it is not an enzyme, it acts as a 'specifier protein' with uridine diphosphate-galactosyl transferase to form lactose synthetase which catalyzes the transfer of galactose to glucose to form lactose.

The immunoglobulins are a very complex group of proteins that provide various types of immunity to the cow and its offspring. Immunoglobulin levels in colostrum (milk from the first 2–3 days postpartum) can be up to $100\,g\,l^{-1}$ as the colostrum provides passive immunity to the neonate prior to its own immune system functioning fully. This level falls rapidly over the first week postpartum to approximately $0.6–1\,g\,l^{-1}$ during normal lactation.

Bovine serum albumin is a large molecule (mol. wt ~ kDa) which is presumed to occur in milk as a result of leakage from the blood in the mammary gland during milk production. Although it can bind metals and fatty acids, it probably has little functional significance in bovine milk.

Lactotransferrin (or lactoferrin) occurs in trace amounts in bovine milk, although in human milk it represents up to 20% of the total nitrogen. It is

similar to serum transferrin, is involved in iron transport, and has an antibacterial role in the neonate gut.

Membrane Proteins The membrane proteins are a diverse group of proteins that, together with lipids, form the milk fat globule membrane. Some are lipoproteins and all tend to have varying degrees of hydrophobicity and hence align themselves in a gradient from the fat surface to the water. This attracts other hydrophobic molecules such as phospholipids and lipolytic enzymes that are adsorbed within the membrane structure.

Biologically Active Proteins and Peptides Milk contains a wide variety of biologically active substances. These include: (1) proteins such as lactoperoxidase, lactotransferrin, immunoglobulins, and vitamin-binding proteins (e.g., folate-binding protein); (2) growth factors such as insulin-like growth factors (IGF-1 and IGF-2), transforming growth factors ($TGF_{\alpha 1}$, $TGF_{\alpha 2}$ and TGF_{β}), mammary-derived growth factors (MDGF I and MDGF II), fibroblast growth factors, platelet-derived growth factors (PDGF), bombesin and bifidus factor; and (3) peptides from milk protein hydrolysates, e.g., glycomacropeptides from κ-casein, phosphopeptides from caseins, caseinomorphins, immunomodulating peptides, platelet-modifying peptide, angiotensin-converting enzyme (ACE) inhibitor, calmodulin-binding peptides, and bactericidal peptides from lactotransferrin.

Enzymes Milk contains both indigenous and endogenous enzymes. As many as 60 indigenous enzymes have been reported, with the majority derived from the milk fat globule membrane. Peroxidase and alkaline phosphatase are indicators of efficient pasteurization; catalase occurs naturally and also results from mastitic infection; lipase releases free fatty acids from milk fat, resulting in rancidity; plasmin degrades caseins; and lysozyme and lactoperoxidase possess antimicrobial activity.

The endogenous enzymes derive from microorganisms present in the milk and can cause both undesirable, e.g. rancidity, bitterness, or off-flavors, and desirable, e.g., ripe cheese flavors, properties.

Lactose

Lactose is the principal carbohydrate in milk and varies with the breed of cow, individuality factors, udder infection, and stage of lactation. It plays an important role in osmotic pressure maintenance in the mammary system and has an inverse relationship with chloride in milk. Lactose is a disaccharide consisting of galactose and glucose linked by a β1–4 glycosidic bond, with the glucose moiety being potentially free (i.e., a reducing sugar) and existing as either an α- or β-anomer.

Vitamins

Milk is a good source of the daily vitamin requirement of an adult person. It contains the fat-soluble vitamins A, D, E, and K, and the water soluble B-group vitamins B_1, B_2, niacin, biotin, panthothenic acid, B_6, folate and B_{12}, and ascorbic acid (vitamin C). The amount of each vitamin varies with stage of lactation and the animal's diet and health.

Minerals and Salts

The soluble and colloidal phases of milk contain varying amounts of different salts with the composition influenced by breed, individual cow, stage of lactation, feed, mastitic infection, and season. The most important cations are calcium, sodium, potassium, and magnesium that occur as phosphates, chlorides, citrates, and caseinates. Individual salt concentrations are dependent on the requirements to: (1) maintain electric neutrality; (2) maintain milk isotonic with blood, through lactose, sodium, potassium, and chloride ion concentrations; and (3) form casein micelles which constrain pH and calcium salt levels and require calcium phosphate to complex with casein. Variation in total calcium is due to changes in citrate and casein during lactation.

Approximately 20 trace elements are found in milk, including copper, iron, silicon, zinc, and iodine. The mineral (ash) content of milk is approximately 0.7–0.8%. Some trace minerals are essential for health, e.g., iron, zinc copper, and manganese, whilst others can be toxic, e.g., aluminum, arsenic, cadmium, and lead. Deficiencies or excesses in these minerals can result in a variety of physiological symptoms.

Other Constituents

Somatic cells (white blood corpuscles or leukocytes) are always present at low levels (200 000 cells ml^{-1}) in milk but increase, and are therefore a useful marker, in the diseased udders of unhealthy cows.

The gases carbon dioxide, nitrogen, and oxygen make up 5–6% by volume in fresh milk and exist dissolved in the milk, bound and nonseparable from the milk, or dispersed in the milk. They can represent a problem during processing, causing burning to heated surfaces.

Effects of Storage

Three chemical processes – oxidation, lipolysis, and hydrolysis – can affect the fat and protein in milk and

milk products during storage, generating either off-flavors (as in milk or butter) or 'mature' flavors (as in cheese).

Oxidation of fat, principally the unsaturated fatty acids, is a major cause of deterioration in milk and milk products, leading to a metallic flavor in milk or an oily tallow taste in butter. This oxidative rancidity can be accelerated by iron or copper salts, high dissolved oxygen levels, or exposure to light. The presence of reducing agents, e.g., from microorganisms, pasteurization, or added antioxidants, can help offset oxidation. With advanced oxidation the fatty acids can ultimately be degraded into aldehydes and ketones, resulting in oxidation rancidity in fat-containing dairy products.

Oxidation of protein by exposure to light results in the conversion of the amino acid methionine to methional. This can then react with riboflavin (vitamin B_2) and ascorbic acid (vitamin C) to form 3-mercapto-methylpropionaldehyde, the main component of 'sunlight flavor'. The overall dynamics of the reactions are factors such as the light intensity and duration, the condition of the milk, and the packaging of the milk.

Lipolysis, or degradation of fat into glycerol and free fatty acids, causes a rancid smell and taste due to the release of butyric and caproic acids. The degradation is due to the action of lipase, an indigenous enzyme normally associated with the casein micelle. This enzyme attacks the triglyceride in fat globules when the milk fat globule membrane is damaged. High storage temperatures and routine dairy processing (e.g., homogenization, pumping, and stirring) all contribute to the breakdown of the milk fat globule membrane and subsequent lipolysis. The lipases can be inactivated by high-temperature pasteurization.

Hydrolysis of milk proteins during storage can be caused by either indigenous proteinases, such as plasmin or cathepsin D, or enzymes derived from introduced sources, such as rennet or bacteria. Plasmin is fairly heat-stable, appears to be associated with the casein micelle during rennet coagulation and cheese manufacture, and preferentially attacks the β-casein to yield the γ-caseins and proteose-peptones. It contributes to the primary proteolysis of caseins in high-cook cheese and may affect the age gelation of ultra heat-treated milk. Introduced proteinases, e.g., from starter bacteria in cheese, are normally used in controlled conditions to produce a wide range of desired flavors or other properties in various milk products.

Effects of Heat

Milk and milk products are routinely subjected to various heat treatments during processing to eliminate any potentially pathogenic microorganisms (e.g., pasteurization). Specific heating regimes of time and temperature are required to destroy these microorganisms – regimes that can also result in detrimental or advantageous changes to other constituents in the milk.

Pasteurization (heating at 70–80 °C for 15 s) results in the 'cream plug phenomenon', where some fat gobules are ruptured, allowing the free fat to bind other intact fat globules to form the plug. This can be alleviated by homogenization. At high temperatures (above 135 °C), the proteins in the milk fat globule membrane partially aggregate, creating a denser, less permeable membrane.

Whilst the caseins are considered heat-stable within the normal pH, salt, and protein concentration ranges, the whey proteins, being globular proteins, are much more susceptible to heat denaturation. Denaturation starts between 60 and 65 °C and is characterized by an uncoiling of the proteins, in particular β-lactoglobulin, followed by the interchange of disulfide bonds to form binding of the β-lactoglobulin to the κ-casein of the casein micelle. Total denaturation is achieved with heating between 90 and 95 °C for 5 min. Alteration of the κ-casein in this way inhibits the renneting ability of milk, resulting in a softer coagulum. Thus, milk for cheese manufacture should not be subjected to heating regimes harsher than 72 °C for 15–20 s. In cultured milk products, such as yogurt, however, the whey protein–casein interactions result in reduced syneresis and improved viscosity, which improves the final product.

Of the other whey proteins, α-lactalbumin is more heat-stable than β-lactoglobulin, undergoing reversible denaturation up to approximately 95 °C. The immunoglobulins are denatured at temperatures above approximately 70 °C, resulting in the loss of their immune properties. Most of the indigenous enzymes in milk are also denatured and lose their activity on high heat treatments.

Above 100 °C, lactose reacts with the milk proteins in the Maillard reaction to cause nonenzymatic browning. A chain of reactions between the amino groups (principally the ε-NH_2 group of lysine) of the proteins and the carbonyl group of lactose produces brown-colored melanoidin polymers, off-flavors, a loss of nutritional value (through loss of lysine), and loss of solubility in milk powders.

Physical Properties

Appearance

The white-yellow color of milk is dependent on the amount of the hydrocarbon carotene. Its opaque appearance is caused by suspended particles of fat and

casein micelles. The greenish color of milk serum and whey is due to the presence of riboflavin.

Density

The density of milk is dependent on its composition and temperature and is generally between 1.028 and 1.038 g cm^{-3}. At 15.5 °C, the following formula can be used:

$$\text{Density (15.5°C)} = \frac{100}{\frac{\text{F}}{0.93} + \frac{\text{SNF}}{1.608} + \text{water}}$$

where F = % fat, SNF = % solids nonfat, and water % = 100 − [F + SNF].

Freezing Point

The freezing point of milk can vary from −0.53 to −0.59 °C and is used to check for adulteration of milk with water. After high-temperature processing, e.g., ultra heat treatment or sterilization, the freezing point will rise due to precipitation of some phosphates.

Acidity

Milk is usually slightly acidic, with a pH value between 6.5 and 6.7. A decrease in the pH is often an indication of bacterial activity and spoilage.

Viscosity

Under moderate shear rates (> 10 s^{-1}), fat contents below 40% and at temperatures above 40 °C (i.e., the fat is a liquid), milk displays Newtonian rheological properties. The coefficient of viscosity at 20 °C is approximately 2.127 mPa s^{-1}. Processing, such as homogenization, heating, and cold agglutination, all alter the viscosity to varying degrees.

See also: **Butter**: The Product and its Manufacture; **Cheeses**: Types of Cheese; **Cream**: Types of Cream; **Dairy Products – Nutritional Contribution**; **Food Intolerance**: Milk Allergy; **Infant Foods**: Milk Formulas; **Milk**: Liquid Milk for the Consumer; Processing of Liquid Milk; Analysis; Dietary Importance; **Powdered Milk**: Milk Powders in the Marketplace; Characteristics of Milk Powders; **Recombined and Filled Milks**; **Whey and Whey Powders**: Production and Uses; **Yogurt**: The Product and its Manufacture

Further Reading

Bylund G (1995) *Dairy Processing Handbook*. Lund: Tetra Pak Processing Systems.

Clark WS Jr (ed.) (1998) Whey. *Proceedings of the Second International Whey Conference*. SI no. 9804. Brussels: IDF.

Fox PF and McSweeney PLH (1998) *Dairy Chemistry and Biochemistry*. London: Blackie Academic & Professional.

Jensen RG (ed.) (1995) *Handbook of Milk Composition*. San Diego: Academic Press.

Wong NP, Jenness R, Keeney M and Marth EH (eds) (1988) *Fundamentals of Dairy Chemistry*. New York: Van Nostrand Reinhold.

Analysis

R J Marshall, Formerly of Asherton, Newbury, Berkshire, UK

Introduction

Milk is complex and the type of analysis to be undertaken depends on the needs of the analyst. Regulatory authorities routinely test milk for protein, fat, adulteration by dilution with water, antibiotics, and bacterial contamination. Determination of milk composition and comparison with milk products enables yields and losses to be measured. The exact composition of milk varies with the breed of cow, the stage of lactation, diet and with the season.

The major proteins of milk may be divided into the caseins (α_{s1}, α_{s2}, β, and κ) (organized into macromolecular units called casein micelles) and the whey proteins (β-lactoglobulin, α-lactalbumin, immunoglobulins, and bovine serum albumin), which are globular proteins. The fat in milk is in the form of droplets surrounded by a membrane, the milk fat globule membrane. It consists mainly of triacylglycerols of which most are composed of fatty acids $C_{4:0}$ to $C_{18:3}$. About 60 different enzymes have been identified in milk. The efficiency of pasteurization may be determined by measuring the activity of alkaline phosphatase. Under adverse conditions, others cause rancidity and proteolysis. Calcium, magnesium, inorganic phosphate, and citrate are essential for the stability of the casein micelles. Milk contains 14 vitamins important in human nutrition and many essential trace elements. It may also contain a large number of contaminants ranging from microorganisms to pesticide residues to radionuclides. (*See* **Contamination of Food**; **Fatty Acids**: Properties; **Pasteurization**: Principles; **Protein**: Chemistry; **Trace Elements**; **Triglycerides**: Structures and Properties; **Vitamins**: Overview.)

There are, then, numerous properties which may be determined. This article reviews the main methods used to characterize milk and to measure important contaminants. The methods are mostly those which are recommended jointly by the International Dairy

Federation (IDF), International Organization for Standardization (ISO), and Association of Official Analytical Chemists (AOAC). In most cases there are several other methods available for determining the properties and composition. A few of these are included where they are based on different principles from the standard methods.

Physical Properties

Physical properties include viscosity, specific gravity, freezing point, thermal conductivity and heat capacity, vapor pressure, surface and interfacial tension, electrical conductivity, refractive index, light absorption, and light scattering. Most of these are usually measured only for research purposes.

Specific Gravity

The specific gravity of milk varies with its composition and is usually measured by either a standard hydrometer (lactometer) or with a pycnometer (density bottle). A correction table for specific gravity at different temperatures has been calculated. If the fat content is known, the total solids content can be calculated using Richmond's formula:

$$\text{Total solids}, T(\%) = 0.25\text{G} + 1.2\text{F} + 0.14$$

where G = lactometer reading at 60 °F (1505 °C) and F = fat%.

Freezing Point

The freezing point of milk is used to detect the addition of water either accidentally during milking and processing or fraudulently. The standard method accepted by the World Health Organization (WHO) uses a thermostatically controlled water bath cooled by electrical refrigeration and a thermistor probe in place of the mercury-in-glass thermometer of the classical Hortvet method. Two types of instrument are available. The first determines a plateau in the freezing curve and the other, used for routine screening, reads at a fixed time from the start of freezing.

In the standard method, supercooled milk is induced to crystallize by mechanical vibration which causes the temperature to rise to a plateau corresponding to the freezing point. Standard solutions of sodium chloride are used for calibration. The method gives highly reproducible results – differences between laboratories testing the same sample of milk should not be greater than 0.005 °C. In general, 1% of additional water in milk raises the freezing point by about $T/100$, where T is base freezing point of authenticated samples. Although dietary, daily, and seasonal variations in the composition of milk affect the freezing point, the osmotic pressure does not change very much, so that the freezing point normally remains within the range −0.530 to −0.570 °C (mean −0.540 °C).

Viscosity

At all but the lowest shear rates, milk may be considered to be a Newtonian fluid, that is, shear stress is directly proportional to shear rate and the extrapolated line passes through the origin. Three main types of viscometer are used: capillary tube (e.g., Ostwald), coaxial cylinder (e.g., Brookfield, Couette) and falling sphere (e.g., Hoeppler). For research, a sophisticated controlled stress rheometer such as a Carrimed or Bohlin may be used. Measured values are dependent on the composition of the milk, temperature, and pH. Simple processing, for example homogenization, has relatively little effect. Viscosity may be measured for research into the effects of processing or chemical additions on the behavior of milk components.

Chemical Analysis

Proximates: Total Solids, Fat, Protein, and Ash

Individual methods are available for the determination of the proximates of milk. However, many laboratories use an infrared analyzer (e.g., Foss Electric) which measures the protein, fat, and lactose content of the milk simultaneously.

For fat, infrared analyzers measure the radiation absorbed by the carbonyl groups of ester bonds in the acylglycerols (at $\sim 5.73\,\mu$m) or by the CH groups (at $\sim 3.48\,\mu$m). For protein, the absorption at 6.46 μm by secondary amide groups is measured, and for lactose the absorption at 9.61 μm by the hydroxyl groups. To reduce errors the milk is first homogenized.

Total Solids

The total solids (TS) of milk is the mass remaining after complete removal of water and volatiles. In the standard method, accepted by WHO, a measured quantity of milk is dried under controlled conditions to constant weight. The TS is expressed as a percentage by mass.

Fat

The chemical method most often used to measure the fat content of milk is the Gerber method. A measured quantity of milk is added to a standard amount of concentrated sulfuric acid in a special test bottle (Butyrometer). Isoamyl alcohol is added to extract the fat and then the bottle is centrifuged to separate the aqueous and fat phases. The fat phase separates into the graduated neck of the bottle and the

percentage of the total volume may be read off. (*See* **Fatty Acids**: Analysis.)

The Rose–Gottlieb method is another which is used quite often. This method gives excellent results, but requires very careful attention to detail to achieve the best results. The milk is treated with concentrated ammonia and then the fat is extracted using organic solvents. Finally, the extract is dried, weighed, and expressed as a percentage of the original weight of milk.

Protein

A number of different methods are commonly used to determine the protein content of milk. (*See* **Protein**: Determination and Characterization.)

In the classical Kjeldahl method, the proteins are 'digested' in sulfuric acid with a catalyst (selenium, mercury, or copper salts). An acid deposit of ammonium sulfate is formed, which is then dissolved in water. The solution is made alkaline with NaOH and heated to distil off ammonia into excess standard acid (sulfuric). The excess acid is back-titrated with standard NaOH to determine the amount of ammonia and hence nitrogen.

$$\text{Casein protein} = \text{Casein N} \times 6.38$$

The total protein, casein protein, and noncasein protein are determined simultaneously by measuring total nitrogen (TN) and noncasein nitrogen (NCN), after precipitation of the casein fraction at its isoelectric point by acetic acid/sodium acetate:

$$\text{Casein N} = \text{TN} - \text{NCN}$$

Dye-binding methods use the ability of dyes to bind to charged amino acid residues of the proteins. The proteins may be separated as previously described and then treated with the dye. Dyes used include amido black and orange G. When they bind to the protein they form an insoluble complex which may be removed by centrifugation. The amount of residual color in the supernatant is then measured by spectrophotometry.

Individual Components

Lactose

Several methods may be used to measure lactose in milk. It can be measured enzymatically by hydrolysis of the lactose in the presence of nicotinamide adenine dinucleotide (NAD), which is reduced to NADH (eqns (1) and (2)). The amount formed is determined spectrophotometrically at 340 nm.

$$\text{Lactose} \xrightarrow{\beta\text{-Galactosidase}} \text{Glucose} + \beta\text{-Galactose} \quad (1)$$

$$\beta\text{-Galactose} + \text{NAD}+ \xrightarrow{\beta\text{-Galactose dehydrogenase}} \text{Galactonic acid} + \text{NADH} + \text{H}^+ \quad (2)$$

NADH has an absorption coefficient of 6.3 mmol^{-1} cm^{-1} at 340 nm, so that the exact amount of anhydrous lactose can be calculated. (*See* **Lactose**.)

Polarimetry may also be used if the proteins are removed first using phosphotungstic acid. The optically active lactose rotates the plane of polarized light in proportion to its concentration.

Ash

The salt content and ash content of milk are not identical. Ash is the residue after incineration at $< 550\,^{\circ}\text{C}$ and comes from both the organic and inorganic constituents of the milk. Chlorides are easily volatilized if the ashing temperature is too high, and citrate is lost completely.

Acidity

This is usually measured in milk at the start of cheesemaking as the baseline for the development of acidity by the starter bacteria which produce lactic acid. Phenolphthalein indicator is added to a measured quantity of milk diluted to twice its volume with carbon dioxide-free water. This is then titrated with 0·1111 mol l^{-1} sodium hydroxide (mol l^{-1}/9) to a first persistent pink color. (1 ml NaOH = 0.0010 g lactic acid). (*See* **Cheeses**: Starter Cultures Employed in Cheese-making.)

Lactic Acid

In the IDF/AOAC standard method for lactic acid in milk, the proteins are removed by precipitation with phosphotungstic acid and filtration. The filtrate is acidified with sulfuric acid and lactic acid is extracted in a special extractor using ether. After adding water, the ether is evaporated off and the remaining solution neutralized. After removal of the ethanol used to wash the solution, it is treated with ferric chloride in hydrochloric acid and a blue color develops which may be quantified at approximately 365 nm using a spectrophotometer.

Gas–liquid chromatography (GLC) and enzymatic methods are also used. (*See* **Chromatography**: Gas Chromatography; **Enzymes**: Use in Analysis.)

Individual Proteins

These are measured by electrophoresis of extracts of milk on polyacrylamide gels (PAGE) using standard protein solutions for quantitation. The caseins are solubilized with urea and separated by electrophoresis on polyacrylamide slab gels containing urea to

prevent aggregation. However, this method is not very satisfactory for the caseins. The whey proteins (β-lactoglobulin and α-lactalbumin) are determined in casein-free extracts by polyacrylamide slab or disk electrophoresis. Proteins in milk may also be analyzed by fast protein liquid chromatography (FPLC), in which they are separated on a column of ion exchange resin, eluted, and detected by their absorption of ultraviolet light.

Minerals

Metallic cations (including calcium and heavy-metal contamination) may be measured by atomic absorption spectrophotometry (AAS), in which the sample is atomized in the gas flame and the absorption is measured at the characteristic wavelength of each element.

Calcium

In the standard method for determining total calcium, the proteins are precipitated with trichloroacetic acid and filtrate obtained containing the soluble calcium. This is converted to the oxalate by adding saturated ammonium oxalate and precipitated by adding acetic acid. The calcium oxalate is washed into water, acidified with sulfuric acid, and heated over a boiling water-bath to dissolve it. The hot solution (> 60 °C) is titrated with potassium permanganate to a first permanent pink color. (*See* **Calcium**: Properties and Determination.)

To measure soluble and colloidal calcium separately, the milk must be separated into a milk serum fraction and a whole casein fraction by, for example, ultrafiltration. The soluble calcium is further subdivided into protein-bound and ionic. Ionic calcium may be determined in protein-free extracts by AAS.

Chloride, phosphate, and other salts in milk may be measured by various titrimetric and colorimetric methods. (*See* **Spectroscopy**: Visible Spectroscopy and Colorimetry.)

Phosphatase Activity

The enzyme alkaline phosphatase is naturally present in milk and, because it is sensitive to heat, it is used to measure the effectiveness of pasteurization. Milk is diluted with a buffer (pH 10.6) containing disodium phenylphosphate which is hydrolyzed by the enzyme, releasing phenol. The phenol is reacted with 2,6-dibromoquinonechloroimide, producing a color which is measured spectrophotometrically at 610 nm.

Contaminants

Microorganisms

Microorganisms in milk (bacteria, yeasts, and molds) come from various sources: from the udder of the cow during milking, especially if the animal has mastitis, from contaminated equipment used to handle and transport the milk, or from personnel, whilst others come from the air. Hygiene control measures, such as use of antimicrobial washes and closed systems and keeping the milk cold, help to prevent microorganisms from entering the milk and growing. Heat treatment of processing equipment and of the milk reduces the levels of pathogenic organisms. The growth of most bacteria is reduced by the use of low temperatures, but certain types, called psychrotrophs will continue to grow at around 6 °C and cause problems when milk is stored for prolonged periods before use. (*See* **Adulteration of Foods**: Detection; **Pasteurization**: Principles.)

The same principle is used to measure most microbial contamination of milk. Dilutions of milk are carefully mixed into molten agar (40 °C) containing nutrients (purified meat extracts, salts, and a carbohydrate source such as glucose) – the growth medium. This is poured into sterile Petri dishes, cooled to set the agar into a gel, and incubated at a suitable temperature (e.g., 30 °C). Each microorganism which grows forms a small colony of cells in the agar. Colonies are counted by placing the dish over an indirect light source and marking each one on the outside of the dish with a felt-tip pen.

Total microorganisms are measured by culturing them in a complete growth medium that provides nutrients for all types. Specific types of microorganism are determined by using media which promote their growth but retard or inhibit others. For example, selective media are used to measure total coliforms (bacteria originating in the gut of animals) and fecal coliforms separately. To measure yeasts and molds, an antibiotic, usually chloramphenicol, is added to the growth medium to prevent bacteria growing. *Salmonella* species are measured by isolation in enrichment media before being grown, recognized, and confirmed.

With the need to prevent contaminated milk from being consumed or processed, rapid tests have been developed. The direct epifluorescence test (DEFT) is used routinely by some laboratories. A fluorescent marker is attained chemically to living cells. The marked microcolonies which grow in a short time fluoresce in ultraviolet-light and are counted automatically by an image analyzer.

Psychrotrophs

Psychrotrophs grow very slowly in milk at low temperatures. They are measured by using a growth medium similar to that described above but incubated at 6.5 °C for 10 days. A recent modification to the method allows the number of psychrotrophs to be

measured after 25 h at 21 °C, but other nonpsychrotrophs may interfere, reducing the accuracy of this method.

Antibiotics and Other Microbial Inhibitors

There are at least 40 antibiotics and sulphonamides which may occur in milk. They can be detected and identified by both microbial and chemical methods.

Microbial methods An inhibitory substance in the milk will prevent or retard the growth of test bacteria (*Bacillus stearothermophilus, Streptococcus thermophilus, B. subtilis, Sarcina lutea, Lactobacillus bulgaricus, B. megaterium*, etc.). In the disk-assay plate method, a disk of absorbent paper impregnated with the milk sample to be examined is placed on the surface of an agar plate preinoculated with a suitable test organism. When incubated, the growth of the bacteria causes the agar to become cloudy. The presence of inhibitory substances such as antibiotics which diffuse into the agar is indicated by a clear zone around the edge of the disk. The inhibitor may be identified by adding an agent that blocks its action. For example, the effect of penicillin is prevented by the addition of penicillinase; *p*-aminobenzoic acid blocks sulphonamides. With such substances in the growth medium, the antibiotic no longer prevents the test organism from growing.

Commercially available methods include Interest, Delvotest, and TTC.

On the farm, a Charm Farm Test kit may be used. The test organism is inoculated in tablet form into liquid medium in a testtube with a sample of the milk. An indicator in the buffer changes color if the bacteria are able to grow.

Chemical methods Inhibitory substances are extracted from the milk and then detected chemically by thin-layer chromatography (TLC), high-performance liquid chromatography (HPLC), fluorescence liquid chromatography, gas chromatography (GC), electrophoresis, bioautoradiography, enzyme-linked immunosorbent assay (ELISA), radioimmunoassay, enzymatic methods, and other forms of immunological methods. They are usually more specific than the microbiological methods. (*See* **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography; **Immunoassays**: Principles; Radioimmunoassay and Enzyme Immunoassay.)

Organic Residues

Organic residues occur in milk either through the diet or by contamination of the milk directly. Pesticide residues, polychlorinated biphenyls (PCBs), and dioxins are all toxic and tend to accumulate in the body. PCBs are used in a very wide range of industrial processes and products; dioxins are a byproduct of many chemical processes, including incineration of chlorine-containing materials at low temperatures.

Pesticide Residues

There is considerable difficulty in identifying pesticide residues in milk because they may be present in very low concentrations. To measure organochlorine (e.g., DDT) and organophosphorus (e.g., bromophos, malathion) residues, the fat is first extracted from the milk by a suitable solvent. The residues are then extracted from this into acetonitrile and diluted with water. The residues are extracted into petroleum ether, purified through a column of fluorosil, and eluted with petroleum ether and diethyl ether. The residues in the eluate are measured by GC and identified by a combination of GC and TLC. This is tending to be superseded by the more sensitive combination of GC with mass spectroscopy (GC-MS). (*See* **Pesticides and Herbicides**: Residue Determination.)

Polychlorinated Biphenyls

Two similar methods are used to measure PCBs in milk. In the first they are isolated by solvent extraction along with organochlorine pesticides and measured by GLC. In the second method they are extracted with solvents and purified on alumina or on fluorosil columns before determination by GC and GC-MS.

Dioxins

After treatment with ammonia and ethanol, the fat phase containing the dioxins is extracted into ether/pentane, then dried and concentrated. It is cleaned up by passing it through a series of columns of base silica, alumina, and activated carbon. The eluate is analyzed by high-resolution GC-MS. The method is calibrated by using standard dioxin labeled with carbon-13.

Radionuclides

Radionuclides which arise in milk are either naturally occurring or come from nuclear accidents such as the Chernobyl reactor disaster. Strontium-89 and strontium-90 (^{89}Sr, ^{90}Sr), cesium-137 (^{137}Cs), and iodine-131 (^{131}I) are probably the most important contaminants. Strontium is absorbed into bone, cesium mainly into muscle, and iodine only into the thyroid. These radionuclides are measured by scintillation counting of their radioactivity. Samples have to be enriched, unless the contamination is extremely high, to get counts which are significantly above background level. Scintillation counting consists of

placing a clarified sample into a special glass bottle with a reagent (scintillant) which absorbs the radiation and emits light in proportion. The light is detected by a sensitive photomultiplier and expressed as counts per unit time or time to standard count. (*See* **Radioactivity in Food**.)

Quaternary Ammonium Compounds (QACs)

These residues from cleaning agents and other fluids containing surfactants (e.g., cetylpyridinium chloraide (CPC) and cetyltrimethylammonium bromide (CTAB)) may have bacteriostatic effects or cause flavor taints. To determine QACs in milk, they are first extracted into solvent (1,1,2,2-tetrachloroethane) and then detected by the formation of a pink color with dioctyl sodium sulfosuccinate (Aerosol T) which is quantified by spectrophotometry.

Species Identification

Identification of the source of milk is generally used to detect the mixing of bovine milk into nonbovine milk and vice versa. Various methods have been tried, including TLC, GLC, and HPLC of the triacylglycerols. However, these methods are not very reliable because the identification of the species of milk present depends on measuring ratios of the fatty acids in the mixture. Improvements in HPLC methodology are helping to overcome some of these difficulties.

The proteins provide a better means of identifying the sources of the milks. The α_{s1}-casein in bovine milk has a higher electrophoretic mobility than the equivalent protein from other species. Therefore, in principle, simple electrophoresis on polyacrylamide gel should be sufficient to analyze the mixture. This is not very reliable, and better results can be obtained by more sophisticated electrophoretic methods such as isoelectric focusing and immunoelectrophoresis. The position and intensity of the electrophoretic bands gives a measure of the identity and amount of protein present. The availability of antisera to total milk proteins, and more recently the availability of monoclonal antibodies to specific proteins, means that exact identification of the origins of proteins in mixed milks is now relatively easy. Antibodies to particular proteins are allowed to react with the milk proteins. When the antibody binds to its antigen it normally forms a line or band of precipitation, the position and extent of which may be used to identify and quantify the protein present. Methods that combine electrophoresis with immunoreaction, such as rocket electrophoresis, give the most precise results.

See also: **Adulteration of Foods**: Detection; **Calcium**: Properties and Determination; **Cheeses**: Starter Cultures Employed in Cheese-making; **Contamination of Food**; **Enzymes**: Uses in Analysis; **Fatty Acids**: Properties; Analysis; **Immunoassays**: Principles; Radioimmunoassay and Enzyme Immunoassay; **Lactose**; **Pasteurization**: Principles; **Radioactivity in Food**; **Trace Elements**; **Triglycerides**: Structures and Properties; **Vitamins**: Overview

Further Reading

Andrews AT (1986) *Electrophoresis. Theory, Techniques and Biochemical and Clinical Applications*, 2nd edn. Oxford: Oxford University Press.

Association of Official Analytical Chemists (1990) *Official Methods of Analysis (OMA)*, 15th edn. Arlington, Virginia: Association of Official Analytical Chemists (AOAC).

International Dairy Federation (1990) *Inventory. Methods of Analysis for Milk and Milk Products*. Bulletin of the International Dairy Federation no. 248, 3rd edn. Brussels: IDF.

International Dairy Federation (1991) *Detection and Confirmation of Inhibitors in Milk and Milk Products*. Bulletin of the International Dairy Federation no. 258. Brussels: IDF.

Marth EH (ed.) (1978) *Standard Methods for the Examination of Dairy Products*, 14th edn. Washington, DC: American Public Health Association.

Ribadeau-Dumas B and Grappin R (1989) Milk protein analysis. *Le Lait* 69: 357–416.

Royal Society of Chemistry (1984) *Challenges to Contemporary Dairy Analytical Techniques*. Special publication no. 49. London: Royal Society of Chemistry.

Dietary Importance

J Buttriss, British Nutrition Foundation, London, UK

Introduction

Milk has been a component of the diet in many parts of the world for thousands of years. This article will review the nutritional value of milk, trends in milk consumption in the UK, and the role of milk in the diet of different population groups.

Nutrient Composition

Milk can be described as one of the most nutritionally complete foods. It provides a wide range of essential nutrients, in particular protein and a range of vitamins and minerals (**Table 1**). It is, however, a poor source of iron and vitamin D, and contains no starch

Table 1 Nutrient composition (per 100 g) of pasteurized milk in the UK

	Whole milk	*Skimmed milk*	*Semiskimmed milk*	*Channel Islands milk*
Energy				
(kcal)	66	33	46	78
(kJ)	275	140	195	327
Protein (g)	3.2	3.3	3.3	3.6
Carbohydrate (g)	4.6	4.8	4.8	4.6
Sugars (g)	4.6	4.8	4.8	4.6
Fat (g)	3.9	0.1	1.6	5.1
Saturates (g)	2.4	0.06	1.0	3.3
Monounsaturates (g)	1.1	Trace	0.5	1.3
Polyunsaturates (g)	0.1	Trace	Trace	0.1
Sodium (mg)	55	55	55	54
Dietary fiber (g)	Nil	Nil	Nil	Nil
Vitamin A (μg)	56	1	23	58
Thiamin (mg)	0.04	0.04	0.04	0.04
Riboflavin (mg)	0.17	0.18	0.18	0.19
Nicotinic acid (mg)	0.08	0.09	0.09	0.07
Potential nicotinic acid from tryptophan (mg)	0.75	0.78	0.78	0.85
Vitamin B_6 (mg)	0.06	0.06	0.06	0.06
Folic acid (mg)	6	6	6	6
Vitamin B_{12} (μg)	0.4	0.4	0.4	0.4
Pantothenic acid (mg)	0.35	0.32	0.32	0.36
Biotin (μg)	1.9	2.0	2.0	1.9
Vitamin C (mg)	1	1	1	1
Vitamin D (μg)	0.03	Trace	0.01	0.03
Vitamin E (mg)	0.09	Trace	0.03	0.11
Calcium (mg)	115	120	118	130
Chloride (mg)	100	100	100	100
Copper (mg)	Trace	Trace	Trace	Trace
Iodine (μg)	15	(15)	(15)	N
Iron (mg)	0.05	0.05	0.05	0.05
Magnesium (mg)	11	12	11	12
Phosphorus (mg)	92	95	95	100
Potassium (mg)	140	150	150	140
Selenium (μg)	1	(1)	(1)	(1)
Zinc (mg)	0.4	0.4	0.4	0.4

N, no reliable information available; values in parentheses, estimated value.
From Holland B, Unwin ID and Buss DH (1989) *Milk Products and Eggs*. McCance and Widdowson's *The Composition of Foods*, suppl, 4. London: Royal Society of Chemistry and Ministry of Agriculture, Fisheries and Food.

or dietary fiber. By volume, water is the major constituent of milk, comprising just over 87%. The remainder consists of milk fat and solids-nonfat (SNF).

Protein

The principal proteins found in milk are casein, lactalbumin, and lactoglobulin. Milk protein has a high biological value since it contains all of the eight essential amino acids, which need to be provided by diet. In addition, milk can improve the overall protein quality of a meal when consumed with foods of lower protein quality such as cereals and pulses. (*See* **Amino Acids: Metabolism.**)

Considerable interest has developed in a number of the peptides present in milk. It is now clear that cows' milk contains a number of biologically active peptide components that may prove to be of relevance to human health. Much of the research in this area has sought to identify and characterize components that influence the immune system.

Carbohydrate

Carbohydrate in milk is in the form of lactose, a disaccharide comprising a molecule of glucose and a molecule of galactose. This sugar is found naturally only in milk and is much less sweet than sucrose.

In the small intestine, lactose is digested by the enzyme lactase to its two component monosaccharides. This enzyme is present in babies and young children. In some adults, mainly noncaucasians, enzymatic activity can decrease, making milk in quantity less well tolerated. Such individuals are described as being lactose-intolerant. Most can tolerate small quantities of milk, and fermented milk products

appear to be better tolerated. In the UK, lactose intolerance is relatively rare in people of European descent but is more common in those of Asian, Far Eastern, and African descent, particularly first-generation members of these ethnic groups. (*See* **Carbohydrates**: Metabolism of Sugars; **Food Intolerance**: Lactose Intolerance.)

Fat

The fat in milk is in the form of minute droplets which rise to the top when milk is left to stand. The principal component of milk fat is triglyceride (triacylglycerol), three fatty acids joined to a glycerol backbone. All triglycerides contain mixtures of three types of fatty acids: saturated, monounsaturated, and polyunsaturated. The contribution of these three types of fatty acid to milk fat in the UK is 61% saturated, 28% monounsaturated, and 3% polyunsaturated. The percentages do not add up to 100% because milk fat is not composed totally of fatty acids. In the UK, whole cows' milk typically contains 3.9 g of fat in every 100 g of milk; semi-skimmed has 1.6 g per 100 g, and skimmed milk has less than 0.1 g per 100 g. The fatty acid profile remains the same. Milk fat contains small amounts of the two essential fatty acids, α-linoleic acid (1.4 g per 100 g fatty acids) and linolenic acid (1.5 g per 100 g fatty acids). (*See* **Fatty Acids**: Properties; **Triglycerides**: Structures and Properties.)

There is growing interest in a fatty acid found in milk known as conjugated linoleic acid (CLA). It is a polyunsaturated fatty acid with one *trans* bond, and is a product of rumen fermentation and so is present only in the meat and milk of ruminant animals. Work on rodents has provided evidence to support an inhibitory effect of CLA against mammary tumors, with some evidence for a similar effect against colon tumors. This work is supported by *in vitro* evidence using human tumor cell lines. CLA has also been shown to inhibit the accumulation of fat during growth in small and large animal models, and there is some evidence of a protective effect in cardiovascular disease. If the health benefits suggested in animal studies can be shown to apply to human populations, the consequences of advice to reduce fat intake may have important public health implications.

Vitamins

All of the known vitamins are present in whole milk (Table 1), although some are present in small quantities.

The fat-soluble vitamins – A, D, E, and K – are removed with the fat when milk is skimmed. Consequently, they are present in only trace amounts in skimmed milk and in reduced amounts in semi-skimmed milk. Whole milk is a good source of vitamin A, a pint (560 ml) providing 47% of the adult male and 55% of the adult female UK reference nutrient intake (RNI). Refer to individual nutrients.)

All of the three major types of cows' milk (whole, semi-skimmed, and skimmed) are good sources of riboflavin (vitamin B_2) and vitamin B_{12}. Where the overall diet is poor, milk can also provide useful amounts of thiamin (vitamin B_1), nicotinic acid, and ascorbic acid (vitamin C).

Heat treatment can have an effect on the vitamin content of milk. Of the heat-treated milks, fresh pasteurized milk has the highest vitamin levels. Milk which has undergone the ultraheat treatment (UHT) process keeps for longer but the higher temperature used in the processing results in lower levels of some vitamins, particularly vitamin B_6, vitamin C, and folate. The sterilization process used for milk has a somewhat greater effect and levels of riboflavin, vitamin B_{12}, and pantothenate will be lower than in milk heat-treated by the other processes. Vitamin C and folate are virtually absent from sterilized milk. (*See* **Heat Treatment**: Ultra-high Temperature (UHT) Treatments.)

Some loss of vitamins is inevitable when milk is stored. Milk exposed to bright sunlight on the doorstep for several hours can lose up to 70% of its riboflavin. Vitamin C levels also fall under these conditions, and measures should be taken to limit such exposure. There will also be gradual losses of folate and vitamin B_{12} from UHT and sterilized milks, even under ideal storage conditions, because of reactions with small amounts of oxygen remaining in the pack or bottle. Boiling milk also reduces its vitamin content, ranging from a 5% reduction in vitamin B_{12} to a 50% reduction in vitamin C.

Minerals

Milk makes a contribution to human needs for virtually all the minerals and trace elements known to be essential for health. These are often present in a form which is well absorbed and utilized by the body (high bioavailability), e.g., calcium and zinc. For most people in the western world, milk and milk products are a major source of calcium. In the UK, the National Food Survey, conducted by the Ministry of Agriculture, Fisheries and Food (now DEFRA), indicates that 55% of the calcium in the typical British diet is contributed by milk and milk products. Milk alone contributes about 40% of the total. Although the contribution to zinc requirements made by milk is relatively low compared with meat, the major contributor, the zinc is in a highly bioavailable form and there is evidence that the combination of milk (or meat) with vegetable foods, in which the zinc is less bioavailable, can enhance the bioavailability of zinc from the whole meal. (*See* **Bioavailability of Nutrients**.)

Seasonal and Breed Variations

Seasonal

The nutritional value of milk is influenced by the cow's diet. Milk produced in summer has a slightly different nutritional composition from that produced in winter. For most of the grazing season (spring), fresh grass can provide all the nutrients the cow needs. However, in some circumstances, e.g., where a cow has a high milk yield, additional food in the form of high-energy pellets is provided. When grazing, a cow may eat up to 70 kg (150 lb) of grass each day. The quality of grass is therefore an important factor in milk production.

The main diet during winter is grass conserved during the summer in the form of either hay (sun-dried grass) or silage ('pickled' grass). This is always less nutritious than fresh grass, so it is necessary to provide additional rations, sometimes in the form of concentrates, usually during milking. Alternative components of the winter diet include sugar beet pulp, maize gluten, or treated straw. Kale, cabbage, forage maize, or fodder beet are often specially grown for winter feeding.

As a result of the seasonal changes in diet, levels of vitamin A, β-carotene, and vitamin E are higher in summer milk. Iodine levels are higher in winter than in summer. Other nutrients remain the same.

The stage in the cow's lactation also has an effect on milk composition. Fifty days into the lactation, the lactose content of the milk starts to fall, and the fat and protein contents rise. However, since milk from individual cows at different stages of lactation is combined, these changes in composition are not reflected in seasonal changes in milk purchased.

Breed

Some breeds, such as Friesian and Holstein, are noted for producing large quantities of milk. Others, such as Guernsey and Jersey cows, are noted for the high fat content of their milk (5.1 g per 100 g compared with 3.9 g per 100 g). Other nutrients, including protein, calcium, phosphorus, carotene, and riboflavin, are also present in slightly higher concentrations in milk from Guernsey and Jersey cows, known as Channel Islands milk. Semiskimmed Channel Islands milk is also available in the UK and retains its slightly higher calcium content compared with ordinary whole milk.

As with milk from Friesian and Holstein cows, Channel Islands milk shows seasonal variations in retinol, carotene, and vitamin E content.

Health Implications of Raw Milk

The hygienic quality of raw (untreated) milk is very high in the UK. This is largely due to the introduction of a centralized testing scheme by the dairy industry in the early 1980s and a payment system to farmers based on the hygienic and compositional quality of raw milk produced on the farms.

Less than 1% of milk does not receive any form of heat treatment. In 1985, legislation reduced the availability of raw milk by prohibiting sales through shops, hotels, schools, and other catering establishments. Direct sales to the public are still allowed in England, Wales, and Northern Ireland. Raw milk has not been available in Scotland since 1983. It can be produced only by a farm holding a licence from OEFRA; it must be clearly labeled 'this milk has not been heat-treated and may therefore contain organisms harmful to health.' The address at which it was bottled or cartoned must appear on the container.

The Chief Medical Officer has advised that young children, pregnant women, elderly people, and those who are currently unwell or have a chronic illness should not consume raw cows' milk.

Trends in Milk Consumption

According to National Dairy Council figures, in 1990 weekly per capita consumption in the UK was 2.3 l (4.1 pints) per week, and had been falling steadily over the previous 10 years, as indicated in Table 2, which uses figures from a different source. The trends in the 1990s are shown in Figure 1.

Growth in Low-Fat Milks

There has been continued growth in purchases of low-fat milk in the UK. Most of this growth has arisen via increased sales of semiskimmed milk (Figure 2).

Table 2 Estimated liquid milk consumption (pints per person per week)

April to March	*England and Wales*	*Scotland*	*Northern Ireland*	*UK*
1959–1960	4.85	4.35	NA	NA
1964–1965	4.94	4.38	5.05	4.89
1969–1970	4.80	4.24	5.09	4.75
1974–1975	4.73	4.58	4.80	4.73
1978–1979	4.50	4.45	4.93	4.52
1979–1980	4.42	4.40	4.91	4.45
1980–1981	4.31	4.35	4.72	4.35
1981–1982	4.29	4.36	4.66	4.31
1982–1983	4.21	4.35	4.66	4.24
1983–1984	4.21	4.31	4.69	4.28
1984–1985	4.16	4.25	4.63	4.21
1985–1986	4.13	4.25	4.59	4.19
1986–1987	4.08	4.15	4.58	4.12
1987–1988	4.05	4.11	4.47	4.05
1988–1989	4.00	4.08	4.43	4.02

NA, not available.
From Milk Marketing Board. *Dairy Facts and Figures*. Thames Ditton, Surrey: Milk Marketing Board.

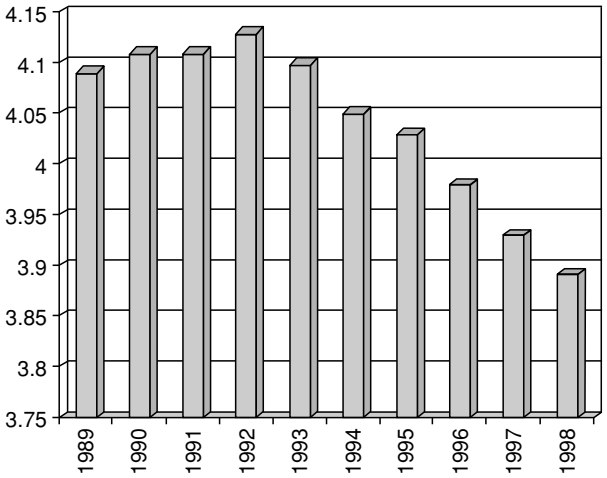

Figure 1 Estimated total liquid milk consumption in the UK from 1989 to 1998. Values given are pints per head per week. Reproduced from National Dairy Council (1999) *Dairy Facts and Figures*. London: NDC, with permission.

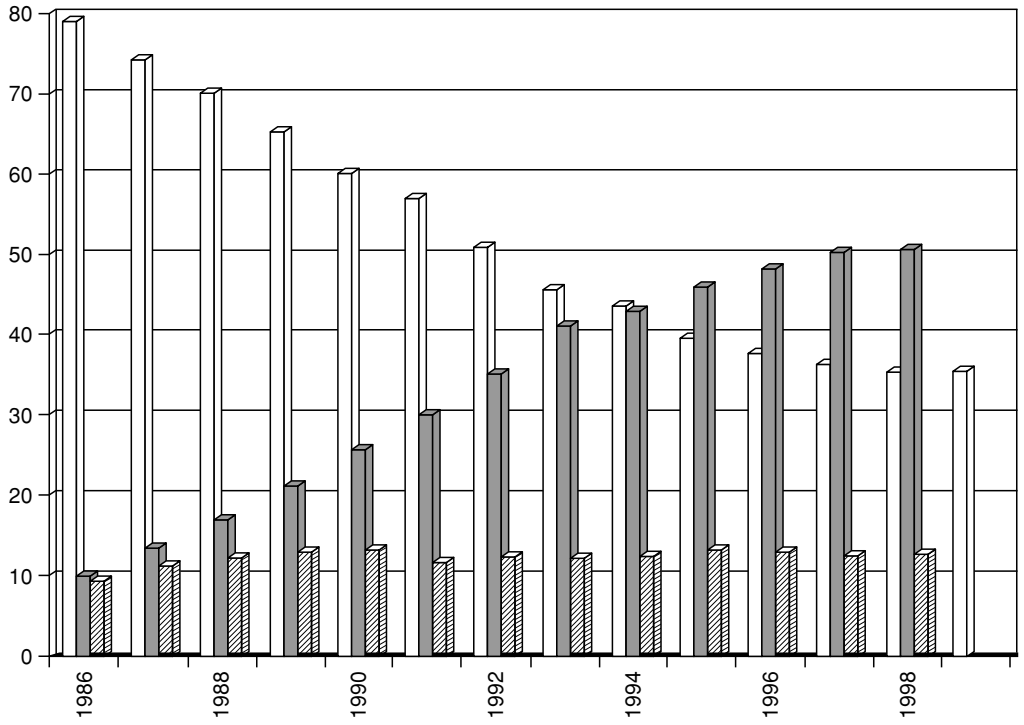

Figure 2 Percentage household milk purchases by fat content in England and Wales from 1986 to 1998. Open columns, whole milk; filled columns, semiskimmed milk; cross-hatched columns, skimmed milk. Reproduced from National Dairy Council (1999) *Dairy Facts and Figures*. London: NDC, with permission.

A similar pattern is seen in many other western countries.

Importance of Milk for Different Age Groups

As part of a balanced diet incorporating a variety of foods, milk can make a useful contribution to requirements for a number of nutrients. It is particularly valuable in the diets of certain groups of the population.

Infants and Preschool Children

In discussing milks suitable for children, the 1994 report, from the government's Committee on Medical Aspects of Food Policy, makes the following recommendations:

- There are advantages in continuing breast milk or infant formula during the first year. Cows' milk should not be introduced as the main drink during the first year of life. (*See* **Infants**: Breast- and Bottle-feeding.)
- However, from 6 months of age, whole cows' milk can be introduced in small quantities as part of the weaning diet, e.g., to mix cereals. It is important with all milks, but particularly with cows' milk, which is low in iron, that care is taken to insure that adequate iron-containing weaning foods are also provided. Suitable foods include those based on lean meat, fortified infant cereals, eggs, pulses, and green leafy vegetables. (*See* **Infants**: Weaning.)
- Skimmed milk is not suitable for children under 5 years, and semi-skimmed milk should not be given to children under 2 years of age because both milks are lower in energy and vitamin A.
- Semiskimmed milk can be introduced as the main drink at age 2 years, provided that the child has a good appetite and is eating a wide range of foods.
- A pint of milk per day is recommended by many health professionals for the under-fives. (*See* **Children**: Nutritional Requirements.)

Older Children

The report on Dietary Reference Values (1991), from the government's Committee on Medical Aspects of Food Policy, advises that children over 2 years and adults should make attempts to reduce their fat intake, especially intake of saturates, such that the proportion of food energy derived from fat is reduced from its current level of 42% to an average of 35% and intake of saturates is reduced to an average of 11% of food energy. At the same time, the reduction in fat intake should be compensated for by an increase in starchy, fiber-rich foods. Replacing whole milk with skimmed or semiskimmed milk is one of the ways in which fat intake can be reduced. Other ways include reduction in fried food, choosing leaner meat, using less spreading and cooking fat/oil, and cutting down on cakes, pastries, and fatty snack foods.

For children and teenagers, milk is a nutritious component of the diet and a major source of calcium, which is essential for the development of a strong skeleton.

School milk schemes exist in the UK and the rest of the European Community which enable primary schoolchildren to receive milk at a subsidized price whilst at school. A nursery school scheme also exists in the UK, providing free milk.

Pregnancy and Lactation

During pregnancy there is no increment in the advised daily calcium intake in the UK (700 mg) as the efficiency of absorption is known to increase. However, it is important that all pregnant women insure that they obtain an average of 700 mg per day (equivalent to a pint of milk). During lactation, calcium needs are covered by a daily intake of 1250 mg per day (1350 mg per day for pregnant teenagers). Milk and milk products are a useful way of obtaining the extra calcium. (*See* **Lactation**: Human Milk: Composition and Nutritional Value.)

Slimmers

Calcium intakes are frequently low among slimmers. This is avoided if skimmed milk, with its lower energy content, is included in the diet. A pint provides over 700 mg calcium, the reference daily intake for women.

Vegetarians

Milk is acceptable to most vegetarians and can make a useful contribution to protein quality. Whole milk is very important in the diet of young vegetarian children. Such diets can be bulky for the young child, and whole milk is a compact source of energy and nutrients in an easily absorbed form. Milk will also make a valuable contribution to vitamin B_{12} intake, which can be low in the diet of vegetarians who exclude eggs. (*See* **Vegetarian Diets**.)

Convalescents

During periods of illness and convalescence, appetite can be poor. Although energy requirement may be lower, essential nutrients are still required. Milk is a useful and versatile food at this time.

Elderly People

Milk is a valuable food for the elderly, who may also have a reduced appetite. Milk provides many essential nutrients in an easily assimilated form. Whole milk provides small amounts of vitamin D, which may be particularly beneficial for the housebound elderly who are unable to get out in the sunlight.

See also: **Amino Acids**: Metabolism; **Bioavailability of Nutrients**; **Carbohydrates**: Metabolism of Sugars; **Fatty Acids**: Properties; **Food Intolerance**: Lactose Intolerance; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; **Infants**: Nutritional Requirements; Breast- and Bottle-feeding; Weaning; **Lactation**: Human Milk: Composition and Nutritional Value; **Triglycerides**: Structures and Properties; **Vegetarian Diets**

Further Reading

Department of Health and Social Security (DHSS) (1988) *Present Day Practice in Infant Feeding*. Committee on

Medical Aspects of Food Policy. Report on Health and Social Subjects 32. London: Her Majesty's Stationery Office (HMSO).
Department of Health (1991) *Dietary Reference Values for Food Energy and Nutrients for the UK*. Report on Health and Social Subjects 41. London: Her Majesty's Stationery Office.
Department of Health (1994) *Weaning and the Weaning Diet*. Report on Health and Social Subjects 45. London: Stationery Office.
Gill HS, Rutherford KJ and Cross ML (2000) Bovine milk: a unique source of immunomodulatory ingredients for functional foods. In: Buttriss J and Saltmarsh M (eds) *Functional Foods II: Claims and Evidence*. Proceedings of a conference held in 1998, organised by the British Nutrition Foundation and the Royal Society of Chemistry, pp. 82–90. Saltmarsh. Cambridge: RSC.
Holland B, Unwin ID and Buss DH (1989) *Milk Products and Eggs. McCance and Widdowson's The Composition of Foods*, suppl. 4, 4th edn, London: Royal Society of Chemistry and Ministry of Agriculture, Fisheries and Food.
Kritchevsky D (2000) Conjugated linoleic acid. *BNF Nutrition Bulletin* 25(4): 25–28.
McBean L (1999) Lactose intolerance. *BNF Nutrition Bulletin* 24: 32–38.
Ministry of Agriculture, Fisheries and Food (1999) *Household Food Consumption and Expenditure 1998*. London: HMSO.
National Dairy Council (1988) *From Farm to Doorstep*. London: National Dairy Council.
National Dairy Council (1999) *Dairy Facts and Figures*. London: NDC.
Renner E (1983) *Milk and Dairy Products in Human Nutrition*. Munich: Volkswirtschaftlicher Verlag.
Renner E (1989) *Micronutrients in Milk and Milk-Based Food Products*. London: Elsevier Applied Science.
Williams C (2000) Commentary: is CLA an important dietary variable relevant to chronic diseases of humans? *BNF Nutrition Bulletin* 25(4): 29–30.

Milk Allergy *See* **Food Intolerance**: Types; Food Allergies; Milk Allergy; Lactose Intolerance; Elimination Diets

Milk Powders *See* **Powdered Milk**: Milk Powders in the Marketplace; Characteristics of Milk Powders

MILLETS

M I Gomez and S C Gupta, SDCC/ICRISAT, Bulawayo, Zimbabwe

Background

Millets comprise a number of small-grained, annual cereal grasses, which include several distinct botanical species. The most important types are pearl, finger, proso, and foxtail millets; other types of local significance include kodo, little, barnyard, and fonio millets, and teff. They differ in climatic and soil requirements, length of growing period, grain consistency, size, and taste. This article describes their common names, origin, geographical distribution, plant description, and crop utilization. The proximate composition of millets is summarized in **Table 1**. (*See* **Cereals**: Contribution to the Diet; **Cereals**: Dietary Importance.)

Pearl Millet

Pearl millet (*Pennisetum glaucum* (L.) R. Br.) is also known as bulrush, cattail, or spiked millet in English, bajra in Hindi, dukhn in Arabic, and mil à chandelles or petit mil in French, and as mhunga or mahango in parts of southern Africa.

Most botanists consider pearl millet to have originated in Africa and to have been subsequently introduced into India. Pearl millet originated in western Africa, where the grain developed by natural selection and acquired considerable resistance to a large number of diseases and insects.

Table 1 Proximate composition (% dwb) of millets

Millets[a]	*Protein (N × 6.25)*		*Lipid*		*Fiber*		*Ash*		*Carbohydrate*	
	Mean	*Range*	*Mean*	*Range*	*Mean*	*Range*	*Mean*	*Range*	*Mean*	*Range*
Pearl (621)	11.6	8.6–17.4	4.8	1.5–6.8	2.3	1.4–7.3	2.2	1.6–3.6	75.6	61.5–89.1
Finger (227)	8.7	6.0–10.9	1.8	1.0–4.6	3.4	3.0–7.5	2.8	2.3–3.9	82.3	73.5–87.5
Common (15)	13.5	10.9–18.6	3.7	2.3–4.9	5.5	0.7–9.0	3.3	2.8–3.7	68.9	60.6–80.1
Foxtail (14)	11.8	10–15.8	4.1	2.5–6.8	7.1	6.3–8.1	3.3	1.4–5.7	66.9	63.0–72.4
Little (9)	10.7	7.5–13.8	6.0	5.3–6.8	7.0	3.7–7.6	5.9	4.6–10.1	66.3	62.6–71.0
Japanese barnyard (3)	11.3	10.2–11.6	4.0	3.0–5.1	13.9		4.6		55.7	
Kodo (7)	10.2	6.6–12.1	3.9	1.5–6.6	8.4	6.2–10.5	3.6	2.9–4.1	73.5	72.5–74.0
Fonio (8)	8.7	5.1–10.4	3.5	2.1–5.2	8.5	4.6–11.3	3.8	1.8–6.0	73.6	62.7–80.0

[a]Figures in parentheses denote the number of samples.
dwb, dry-weight basis.
Data derived from several sources and different regions.

Pearl millet is the most widely cultivated of all millets and is one of the most drought-tolerant of all domesticated cereals. Soon after its domestication, it became widely distributed across the semiarid tropics of Africa and Asia. It is planted on some 15×10^6 ha in Africa and 12×10^6 ha in Asia. In Africa, major pearl millet-growing countries are Nigeria (3 133 000 t), Niger (1 227 000 t), Mali (695 000 t), Senegal (539 000 t), Burkina Faso (447 000 t), Sudan (355 000 t), Chad (161 000 t), and Tanzania (144 000 t). It is also grown in Cameroon, Ghana, Zimbabwe, Togo, Angola, Namibia, The Gambia, Ivory Coast, Zaire, Sierra Leone, Guinea Bissau, Kenya, Central African Republic, and Guinea where the production is above 10 000 t but less than 100 000 t per year. There are many other African countries, such as South Africa, Benin, Zambia, Ethiopia, Mozambique, Malawi, Libya, Mauritania, Morocco, and Botswana, where it is grown on small acreages. In Asia, it is the most important crop in India (5 625 000 t), followed by Pakistan (252 000 t) and Yemen (108 000 t). It is also grown in Syria, Myanmar, and Saudi Arabia. The average grain yield production is about 0.5 t ha^{-1}, but the improved varieties with good management under dry land conditions can produce up to 3 t ha^{-1}.

Pearl millet is a robust annual grass, usually 1.2–3.5 m tall. It is a staple cereal in India, and in some countries of west and southern Africa, such as Senegal and Namibia. The stems are 1–3 cm or more in diameter. The plant can tiller from few to as many as 20 culms based on spacing, management, and the cultivar. The tillering could be basal as well as nodal. Most of the cultivars have basal tillering. The internode is usually light green but sometimes purple. The nodes are often marked by a ring of long, white cilia pointing upward and bearing a ring of adventitious roots on the basal side. The leaves are long, scabrous, and rather slender; they may be smooth or have hairy surfaces, and they have hairy ligules. The blades are lanceolate, cordate, and sometimes 90–100 cm or more in length and 5–8 cm in width. The midrib color is white or green but sometimes brown. Leaves are upright or drooping, and the ligule is narrowly membranous with a fringe of hairs. The leaf sheaths are open and hairy, and the ligule is short, Leaves and culms may vary in color from light yellowish green to dark green and, sometimes, deep purple.

The ear head can be compact, semicompact or loose, cylindrical, conical, or spindle-shaped, 2–5 cm in thickness, and usually 15–45 cm long, although some Nigerian varieties can have ear head lengths over 100 cm. The rachis is straight, cylindrical, solid, often 8–9 mm thick, and unbranched. It extends the entire length of the inflorescence, tapering gradually from base to apex. The seeds are mostly gray but are sometimes grayish brown, grayish white, purple, yellowish brown, or white. Endosperm colors range from white to yellow and gray. The individual seed weight varies from 3 to 14 mg. The spikelets can be without bristles or with long bristles.

The aleurone layer is only one cell thick, and the endosperm consists of an outer hard or vitreous portion and an inner opaque region of soft endosperm. The hard endosperm consists of a tightly packed cellular structure containing no air spaces, whereas the soft endosperm is rather loosely packed and contains relatively large air spaces. The soft endosperm appears to contain no protein bodies, whereas large protein bodies constitute the protein matrix of the hard endosperm in which the starch is embedded. Proximate analysis of whole pearl millet grain shows it to have a relatively low protein content of about 10% but a high fat content of 3.8–4%, largely concentrated in the germ.

Processing and Food Uses

Pearl millets are used largely to prepare traditional, thick or thin, fermented or unfermented porridges in Africa. The second major use in Africa is malting for

the brewing of traditional beers and wines. In west African countries, e.g., Senegal, millet is used for making couscous, pap, and fritters. In Cameroon, pearl millet-based gruels and steamed cakes are prepared for feeding infants and preschool children. Malted pearl millet in combination with legumes has been used to prepare malted weaning foods. Pearl millet has also been used in composite flour with wheat for making bread. Up to 30% pearl millet was used successfully in making bread in Senegal. The nutritional advantages of pearl millet are its high fat content and a relatively high lysine content, comparable with that of high-lysine corn in some varieties. Antinutritional factors, however, have been reported in several studies. A thionamide-like substance has been identified that interferes with the formation of thyroid hormones, which in turn leads to undesirable goitrogenic effects. (*See* **Fermented Foods**: Beverages from Sorghum and Millet; **Goitrogens and Antithyroid Compounds**.)

Pearl millet is traditionally milled by hand-pounding in a wooden mortar, yielding a flour of about 85% extraction. Experimental milling systems, such as the Storamil process, have given 65–75% yields of fine flour. Pearling or debranning of pearl millet has also been achieved with a vertical cone polisher of the type normally used for polishing rice.

Pearl millet can be malted and used wholly or partially in place of sorghum malt in the traditional or industrial brewing of opaque beer. However, the small size of the grain is a disadvantage in large-scale industrial malting plants.

Pearl millet is also grown as a forage crop in the southeastern USA and in some parts of southern Africa.

Finger Millet

Finger millet (*Eleucine coracana* (L.) Gaertn, ssp. *coracana*) is commonly known as ragi in India and wimbe in East Africa; these are the major finger-millet-growing areas. Common English names are African millet, birdsfoot, and coracana.

Finger millet was developed in Africa from *E. coracana* ssp. *africana*, probably in the Ethiopian region. It was introduced to India perhaps more than 3000 years ago. It is a tropical crop grown from sea level to 3000 m above sea level.

The global production of finger millet is about 3.7×10^6 t of grain per year with an average of 0.6 t of grain per hectare. The improved varieties under good management can produce up to 4 t of grain per hectare. In Asia, the major finger-millet-producing country is India (2 613 000 t), followed by Nepal (122 000 t), China (79 000 t), and Afghanistan (38 000 t). In Africa, Uganda ranks first (451 000 t), followed by Ethiopia (198 000 t), Tanzania (95 000 t), and Zimbabwe (46 000 t). To a lesser extent, it is also grown in many other countries, such as Iran, Sri Lanka, Myanmar, Bhutan, Iraq, Pakistan, and Jordan, and (in Africa) Kenya, Burundi, Zambia, eastern Zaire, Sudan, Malawi, South Africa, and Rwanda.

Plants are annual, tufted, erect, or with geniculately ascending culms; they reach an average height of 103 cm and a maximum height of 165 cm, and they sometimes root from the lower nodes. Culms are commonly branched from the upper nodes to produce secondary inflorescences. Leaf blades are linear to linear-lanceolate, generally 35 cm but sometimes up to 70 cm in length, and 20 mm wide. Inflorescences are digitate, often with one or more racemes some distance below the main cluster of four to 19 branches. Inflorescence branches are slender, or incurved at the tip when robust, sometimes with secondary branches. Spikelets are six- to nine-flowered and 6–10 mm long, overlapping, and mostly arranged in two rows along one side of the rachis. Glumes are unequal and shorter than the spikelet. The inflorescence shape is variable. The digitately arranged branches may spread out and become reflexed, or they may be erect and incurved, often forming a fist-like structure.

Finger millet is largely a peasant crop. Seeds vary in color from light brown to dark brown or purple, and are small, up to 2 mm long, globose, and hard. The individual seed weight can vary from 1 to 3 mg. The small, flinty grains are highly resistant to both stored insect pests and molds, and thus have excellent storability. Some of the brown finger millet varieties contain high levels of tannin – 3–4% catechin equivalent. (*See* **Tannins and Polyphenols**.)

Processing and Food Use

Immature grain is roasted green and consumed in the dough stage. Traditionally, the grain is processed by wet grinding on stones. The resulting flour is dried and consumed as thick or thin porridge. Light-colored millets are particularly well suited for malting, the diastatic activity of millets (except pearl millet) being higher than that of sorghum. One of the major uses of finger millet in Africa is therefore in malting, prior to the brewing of traditional beers. To a limited extent, finger millet is reported to be grown in Japan and China, where its major use is the making of beer. In India, and Sri Lanka, finger millet flour is used in the preparation of unleavened bread (roti). (*See* **Sorghum**.)

Advantageous nutritional attributes of finger millet are its particularly high calcium and iron contents

(220–855 mg per 100 g and 64–15 mg per 100 g, respectively) and its high methionine content (83–226 mg per gram total N). In Nepal and India, farmers attach economic value to finger millet straw, since they use it as a livestock feed.

Processing of finger millet by dehusking or debranning was unsuccessful because of the small seed size, the relatively thick pericarp, and a highly friable endosperm. Roller milling of whole grain and sieving, however, produces a refined flour. Similarly, in traditional processing, finger millet is ground or pounded to a flour as whole grain, and particles of pericarp are sifted or winnowed out. (*See* **Flour**: Roller Milling Operations.)

In the preparation of unleavened breads such as roti or chapati, a cold-water dough does not hold together owing to the lack of gluten. Hot water is therefore used to pregelatinize the starch and to attain the required degree of cohesiveness for preparing the thin, flat, griddle breads. (*See* **Chapatis and Related Products**.)

Foxtail Millet

Foxtail millet (*Setaria italica* (L.) P. Beauv.) is also known as Italian millet and setaria. It is grown as a cereal in southern Europe, and in temperate, subtropical, and tropical Asia. It was domesticated in the highlands of central China around 5000 years ago, and spread to Europe and Asia soon after its domestication.

Total production of foxtail millet is around 5.7×10^6 t of grain per year, of which China alone produces 4.9×10^6 t, followed by India (500 000 t). It is also grown in Europe, particularly in Portugal, Turkey, Hungary, France and Spain, in Asia, primarily in Korea, Pakistan, Myanmar, Bhutan, Nepal, and South Africa. Each of these countries produces less than 10 000 t per year.

There are three morphologically distinct races: *moharia*, *maxima*, and *indica*. These differ in morphological characteristics, such as plant height and inflorescence, as well as in distribution. Race *moharia* is cultivated in southeastern Russia, Afghanistan, and Pakistan. Race *maxima* is cultivated in transcaucasian Russia, the Far East, and the USA for bird feed. Race *indica* is largely cultivated in India and southern Asia.

Cultivars from India are morphologically different from those of Europe and the Far East, and are typically robust with inflorescences bearing branches loosely arranged on the primary axis.

Variability has been reported for several characteristics such as plant height (20–175 cm), peduncle length (0–500 mm), inflorescence length (10–320 mm), exertion (10–360 mm), basal tillers (1–80), and in time to flowering (30–135 days) in germplasm accessions.

Uses

Foxtail millet is widely used as a food grain and as a forage crop in Asia, North Africa, and southeast Europe. It is used for brewing beer in Russia and is an important food crop in China, where it is used in gruels, for making cakes, and in beer and vinegar production. It is also used as a medicinal food in Chinese medicine, the beneficial effects being partly attributable to its high nutritional quality. The grain is also used as an animal feed, and the straw as forage.

Proso Millet

The millet with the botanical name of *Panicum miliaceum* (L.) is known as proso millet in Russia.

The common English names of this millet are hog millet, common millet, and broomcorn millet. The progenitor of proso millet is native to Manchuria. It was probably domesticated in central and eastern Asia, and was cultivated in Europe in Neolithic times. The species was introduced to Europe as a cereal at least 3000 years ago. This is essentially a crop of the temperature regions, but it is also grown in the subtropics and on high ground in tropical winters.

Proso ranks third after pearl millet and foxtail millet in terms of total global millet production, about 5×10^6 t of grain per year. The major proso-millet-producing countries are the Commonwealth of Independent States (2 330 000 t), China (1 600 000 t), India (500 000 t), Argentina (173 000 t), Myanmar (150 000 t), and the USA (110 000 t). It is also grown on small acreages in Korea, Australia, Turkey, Austria, Yugoslavia (Montenegro, Serbia), Pakistan, Greece, Bhutan, Nepal, Mongolia, and Japan.

Plant height varies from 25 to 133 cm. The inflorescence length varies from 2.2 to 40 cm. The flag leaf may cover as much as 5 cm of the inflorescence; however, the exertion of the inflorescence may be up to 32 cm. This is a highly tillering crop. It is early maturing, some of the accessions flower in 28 days, and a crop can be harvested in 50–55 days. However, most of the accessions flower in 30–35 days, and a few take up to 50 days. The individual seed weight can vary from 5 to 9 mg – similar to that of pearl millet. It is a self-pollinated crop, and hybridization can be achieved by hand emasculation.

Food Uses

In China, the grain is steamed or boiled like rice. It is also eaten flaked and sweetened with sugar. In the

USA, it is the largest millet crop and is largely used as a feed grain. Proso millet has been comilled in an experimental, semiindustrial process with wheat after individually conditioning the grains to 16% moisture. The flour can be used at up to 10% addition rates with no deleterious effects on the quality of bread.

Kodo Millet

Kodo millet (*Paspalum scorbiculatum* L.) is also known as ditch millet. It is grown only in India, although the wild grass is a widespread tropical weed that is harvested as a wild cereal in West Africa. The species was domesticated in India some 3000 years ago. It is grown in India from Kerala and Tamil Nadu in the south, to Rajasthan and Uttar Pradesh in the north, and West Bengal in the east, as a food grain.

The plant height can vary from 30 to 90 cm, with 10–48 basal tillers per plant. The length of the inflorescence varies from 2 to 12 cm. It is a relatively late-maturing crop compared to other small millets. The flowers are highly cleistogamous with a maximum opening of up to 50% only, i.e., this is a highly self-pollinated crop. Cytological studies have revealed the chromosome number as $2n = 40$.

The statistics for kodo, little, and barnyard millets are combined together, and these species are grown only in India, except barnyard millet, which is also grown in Japan (only 1000 t per year). The total production in India is around 386 000 t per year.

The crop is hardy and drought-resistant, and is capable of growing in marginal soils. The grain is enclosed in a hard, corneous husk, which makes debranning difficult. The grain is said to be poisonous after rain; this may be due to a fungal infection. Winnowed, clean, healthy grain seems to pose no health problem. Kodo millet is primarily cooked as rice.

Little Millet

Little millet (*Panicum sumatrense* Roth. ex Roem. & Schult.) is also known as miliare. It is widely cultivated as a cereal across India, Nepal, and western Myanmar. It is particularly important in the eastern ghats of India, where it forms an important part of tribal agriculture.

A comparative morphological study of accessions resulted in the recognition of two races. Race *nana* plants are small (60 cm) to large (170 cm) with decumbent culms that root at the lower nodes before becoming erect to produce flowering culms. Terminal inflorescences are 14–50 cm long, erect, open, and strongly branched, with the branches sometimes clumped at the time of maturity. Race *robusta* includes plants that are erect, or produce flowering culms from a shortly geniculate base. Flowering culms are 120–190 cm tall and robust. Terminal inflorescences are 20–46 cm long, open, or compact and strongly branched. Open inflorescences are essentially erect, and compact inflorescences become curved at maturity. It is largely a selfpollinated crop; natural cross-pollination can occur up to 3.5%.

The dehusked grain is cooked and consumed like rice, or is milled into flour. The crop thrives under conditions that will sustain no other edible plant and will mature in 2–5 months.

Barnyard Millet

Japanese barnyard millet (*Echinochloa crusgalli*) is native to temperate Eurasia and was domesticated in Japan some 4000 years ago. Barnyard (sawa) millet (*E. colona*) is widely distributed in the tropics and subtropics of the Old World, and was domesticated in India. The two species have different chromosome numbers ($2n = 54$ and $2n = 36$, respectively), and hybrids between them are sterile.

Cultivated plants of *E. colona* (L.) Link. are erect or geniculate ascending, often tufted, annual, and up to 242 cm tall. Culms are slender to robust. Slender plants are decumbent, with the culms strongly branched and rooting from the lower nodes. Stout plants are erect, with a few culm branches from the upper nodes. Inflorescences are variable; decumbent plants are characterized by short racemes that are appressed to the primary axis, whereas more robust plants have larger, pyramidal inflorescences with spreading lower racemes. Inflorescences are usually erect, rarely drooping, and can be up to 28 cm long. Spikelets are persistent at maturity, 2–4 mm long, acute but never awned. The grain is 2–3 mm long and 1–2 mm wide.

Cultivated plants of *E. crusgalli* (L.) P. Beauv. are erect, tufted, annual and up to 100 cm tall. Culms are mostly robust, simple, or branched from the upper nodes. Inflorescences are erect or slightly nodding, well exertion from the upper leaf sheath, with the spike-like racemes more or less erect and sometimes incurved at the tip. Spikelets are persistent at maturity, very crowded, 3–4 mm long, shortly cuspidate but rarely awned. The grain is 2–3 mm long and almost as wide.

Japanese barnyard millet is the fastest maturing of the millets and, under favorable temperature and moisture conditions, ripens within 45 days. The crop is grown for forage and food in India, and in the USA is said to produce two harvests within the cropping

season. *Echinochloa decompositum*, the Australian variety is used as food by aboriginal tribes.

Teff

Teff (*Eragrostis tef*) is the most important cereal crop in Ethiopia, particularly in the poorly drained, heavy soils that predominate in the Central Plateau. The total production is slightly above 1×10^6 t per year. Ethiopia is considered the center of origin of the crop, and domestication is thought to have started in the northern highlands. Teff grows in Australia, Kenya, and South Africa, but is cultivated as a major staple crop only in Ethiopia.

Teff is cultivated from sea level to 2800 m, on soils varying from waterlogged to well drained, and in areas that have less than 300 mm or more than 1000 mm seasonal rainfall.

Teff is a sexually propagated, self-pollinated annual tufted grass. It is an allotetraploid with $2n = 40$ chromosomes. The plants are short in stature (height varying from 45 to 150 cm) with loose, open panicles. The culm thickness varies from 1.2 to 3.1 cm. Teff can mature in 3–4 months, but a few early-maturing accessions can mature in 2 months. The average grain yield is 0.8 t ha^{-1}, but with improved management and cultivars, up to 2 t ha^{-1} can be obtained.

Two types of teff are found in Ethiopia – white- and brown-seeded varieties – and the white is preferred for food. The grain is very small, about 1–1.5 mm in length, and 1000 grains weigh approximately 0.3 g.

The flour of teff is used in Ethiopia for making a flat, fermented, griddle bread or pancake known as injera. It is also used for making porridge and the traditional alcoholic drinks, 'tella' and 'katikalla.'

Fonio

Fonio (*Digitaria exilis* Stapf) is also known as 'hungry millet,' 'hungry rice,' fundi, and acha. It is grown throughout the savanna areas of west Africa, notably in Guinea and Nigeria. The crop is not cultivated outside Africa and is capable of thriving in poor, rocky, and marginal soils. It is considered to be the oldest West African cereal, and its cultivation is thought to date back to 5000 BC. Black fonio (*D. iburua* Stapf) is grown by the Hausa of Nigeria (northern Nigeria), and in Guinea, Togo, and Benin, but the total production of black fonio is insignificant.

Total production of Fonio is 309 000 t per year, and about 70% of this is produced in Guinea (200 000 t). The other major producing areas are Nigeria (58 000 t) and Mali (36 000 t). To a lesser extent, it is also grown in Burkina Faso, Ivory Coast, Niger, and Gambia. It is grown on the plains as well as in the mountains, i.e., under annual rainfalls from 400 to 3000 mm.

Fonio is an annual plant with a plant height of 45–50 cm. The roots are well developed and attached firmly to the soil. The inflorescence is composed of two to six fingers of 5–12.5 cm length. The grains are small, and the individual grain weight is 0.5–0.6 mg. The grain yield varies from 0.2 to 0.5 t ha^{-1}. In low-rainfall areas, the crop matures in 2–4 months, whereas in high-rainfall areas, the maturity period varies from 4 to 5 months.

The seed is very small and is usually threshed and pounded to remove chaff; it is used alone or mixed with other cereals to make porridge. The grain can be added to stews, or boiled and eaten with palm oil. Fonio can also be used for brewing beer. In comparison with other cereals, fonio has a poor nutritional value in terms of total protein and amino-acid balance.

See also: **Cereals**: Contribution to the Diet; Dietary Importance; **Chapatis and Related Products**; **Fermented Foods**: Beverages from Sorghum and Millet; **Flour**: Roller Milling Operations; **Goitrogens and Antithyroid Compounds**; **Sorghum**; **Tannins and Polyphenols**

Further Reading

FAO (1990) *Structure and Characteristics of the World Millet Economy. CCP:GR 90/4, July 1990*. Rome: Food and Agriculture Organization.

Hulse JH, Laing EM and Pearson OE (1980) *Sorghum and the Millets: Their Composition and Nutritional Value*. London: Academic Press.

Krishna Kumari S and Thayumanavan B (1998) Characterization of starches of proso, foxtail, barnyard, kodo, and little millets. *Plant Foods for Human Nutrition* 53(1): 47–56.

Pathak P, Srivastava S and Groves S (2000) Development of food products based on millets, legumes and fenugreek seeds and their suitability in the diabetic diet. *International Journal of Food Science and Nutrition* 51(5): 409–414.

Prasada Rao KE, de Wet JMJ, Gopal Reddy V and Mengesha MH (1991) *Diversity in the Small Millets Collection at ICRISAT. Paper Presented at the Second International Small Millets Workshop, 8–12 April 1991, Bulawayo, Zimbabwe*.

Seetharam A, Riley KW and Harinarayana G (eds) (1989) *Small Millets in Global Agriculture. Proceedings of the First International Small Millets Workshop, Bangalore, India, October 29–November 2, 1986*. New Delhi: Oxford and IBH Publishing.

MILLING

Contents

Principles Of Milling

E S Posner, ESP International, Savyon, Israel

Introduction

The wheat seed which is used for the manufacture of flour is an elongated structure 7–12 mm long and 2–4 mm wide (Figure 1). A crease along the kernel extends nearly to the center. Structurally, the wheat kernel consists of three main parts. The pericarp is the outer coat and is divided into several layers. The wheat germ is divided into the embryo and the scutellum. The endosperm is an accumulation of cells that makes up to 83% of the kernel. The endosperm cells, which are the 'flour' when reduced to small particles, contain many starch granules embedded in a proteinaceous matrix. (*See* **Wheat**: The Crop.)

Wheat is a living organism which varies in its quality as a result of genetic characteristics as well as a result of growing and storage conditions. Wheat-processing characteristics are related to its millability, protein quality and quantity, and baking results of the end product. Three main kinds of wheat can be recognized: hard, soft, and durum. Hard wheats tend to be higher in protein content (10–15%), usually have harder structured endosperm, and find their principal use in bread flour. Flour produced from internally softer wheats, with lower protein content (8–11%), are used in cake and cookie production. Durum wheats are processed to a granular or floury endosperm product for pasta. Distinctively different milling systems are designed to accommodate the differences in wheat that affect their millability. Although in the past mills were usually constructed to mill the locally harvested wheat, modern mills in any part of the world can be designed to process any desired wheat for a special end product.

The wheat flour-milling operation consists of the following functions: receiving, analyzing, wheat storage, blending, cleaning, conditioning, milling, product storage, and shipping. Table 1 indicates the various products and byproducts of the wheat-milling operation.

Wheat Quality for Milling

The determination of wheat milling quality is very complex. No single test has yet been devised to establish conclusively the milling quality of a wheat. Fluctuations in wheat characteristics and quality reduce the efficiency of the fully automated, high-capacity milling unit of today. High moisture replaces the amount of dry material in the wheat and is a significant economical quality factor. Larger wheat kernels are correlated with potential flour yield. Smaller kernels have a larger ratio between pericarp and endosperm, resulting in a lower amount of available endosperm. The weight of a thousand sound, whole wheat kernels, corrected to a fixed moisture level, is also a good indication of the flour yield and millability. Hardness of the kernel, especially after it has been tempered for milling, is an important quality characteristic. Harder wheats require longer tempering stages and produce more granular material, which flows and sieves better. Harder wheats also tend to give better extraction levels as a result of better separation of endosperm from bran. All the above-mentioned kernel characteristics affect the milling quality of wheat and require the appropriate design of a milling system and its adjustment. Recent development of the single-kernel characterization system (SKCS) can be calibrated to measure single-kernel hardness, protein, starch, internal insect infestation, color, disease, size, and moisture.

Wheat Storage and Blending

Wheat storage involves a substantial investment in facilities, equipment, and the value of the stored wheat. Different means of preservation are used, such as aeration with ambient air of the right characteristics or cooled air, fumigation, and control systems to monitor any physical or chemical changes in the wheat. Uncontrolled storage conditions could cause wheat deterioration, starting in 'pockets' in the bin as a result of high moisture. Live insects would also flourish if the temperature, humidity and grain fractions were readily available. There is an equilibrium between moisture content of the wheat, air

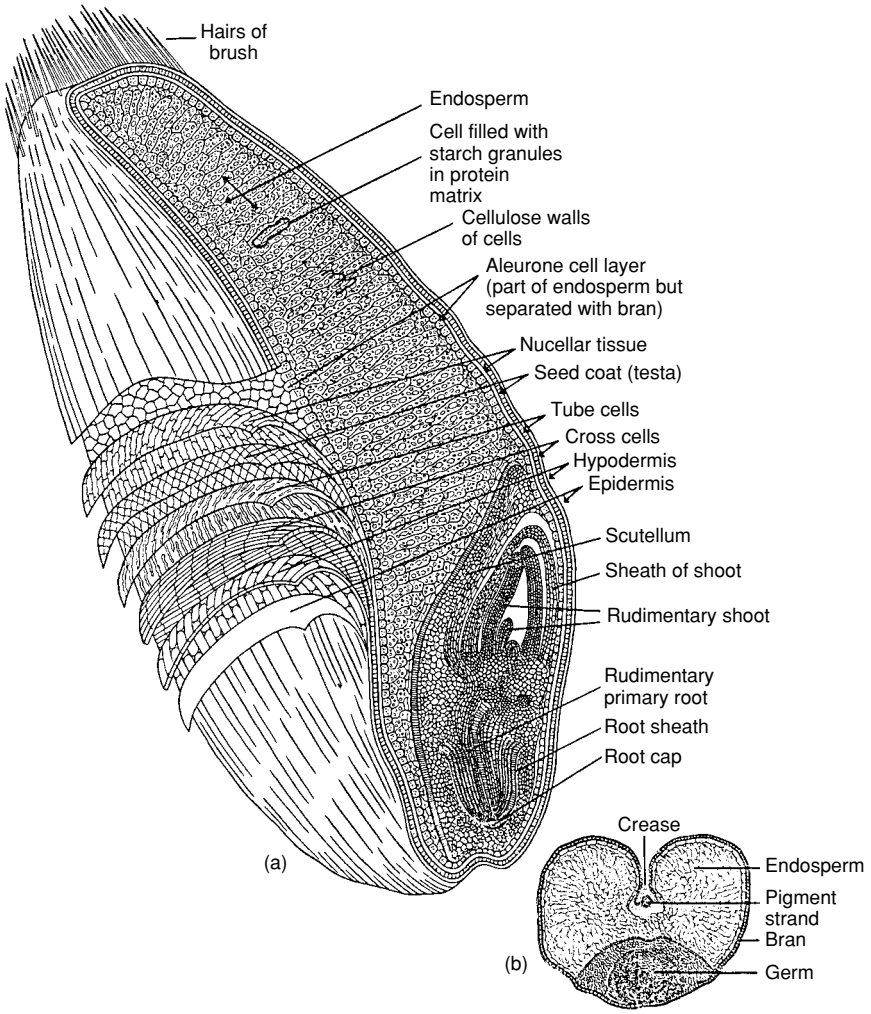

Figure 1 (a) Longitudinal section (enlarged about 33 times) and (b) cross-section of the wheat kernel. The endosperm makes up about 83% of the kernel, the bran about 14.5%, and the germ about 2.5%. Courtesy of the Wheat Flour Institute, Washington DC, USA.

Table 1 Products from a milling system[a]

Product	*Percentage of total*	*Primary end use*
Wheat screenings[b]	1–3	Feed
Flour I	60–65	Baked products
Flour II	10–17	Baking, industrial utilization
Farina	3–8	Breakfast foods
Semolina[c]	65–67	Pasta
Bran[d]	22–25	Feed, high-fiber
Germ	0.2–0.5	Food, cosmetics

[a]Based on wheat entering the mill.
[b]Screenings are impurities removed from wheat before grinding.
[c]In durum milling, only the rest (8–10%) of extracted endosperm is flour II.
[d]Bran is a distinct byproduct consisting of broad pericarp flakes, the aleurone layer, and some endosperm. Commercially, the term includes all offals from the process, such as the shorts (fine bran, germ, and endosperm particles) and red dog (very fine bran, germ, and endosperm particles).

temperature, and humidity. Accordingly, safe storage conditions include a moisture content below 12.5%, and temperature below 25 °C. Even under optimal conditions, stored wheat goes through a continuous change in its milling and baking characteristics by aging; this change is significant in the 3 months immediately after harvest. (*See* **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Insect Pests**: Problems Caused by Insects and Mites.)

One of the miller's main objectives is to be a supplier of uniform product to the customer. To accomplish this, there must be on hand wheats with such qualities that blending them will produce a mill-grist that is uniform in all its physical and chemical characteristics. To accommodate this need, it is important to evaluate wheat upon receipt and segregate it in bins according to quality.

Wheat Cleaning

The current methods of wheat harvesting make it impossible to obtain a pure product off the field. Foreign materials in the grain bulk have to be removed before milling, and this is accomplished in the wheat-cleaning section.

Separation of the unmillable fractions from wheat is based on differences in the following material characteristics: size, specific weight, shape, color, solubility, and response to magnetic forces. Many versions and combinations of those principles have been combined in cleaning machines. Modern equipment combines more than one principle so that, in many cases, the bulk of the grain in a certain range of size and specific weight is diverted to subsequent stages in the process, whereas only a small part of the load is cleaned intensively.

Wheat Conditioning

The clean wheat is subjected to the conditioning process, which is the addition of water followed by a time period of tempering it, allowing the water to be absorbed by the kernels. The objective in conditioning the wheat before milling is to reach a state at which the pericarp and germ parts are tough and plastic and do not splinter during the grinding process, because they absorb water very readily. On the other hand, the endosperm should be as friable as possible. Three variables effect the conditioning process, namely tempering time, temperature, and moisture.

Characteristics such as internal structure and level of pericarp damage of the wheat kernel exert different effects on the conditioning of wheat for milling. Wheat size significantly affects the water penetration rate because of its relationship to the specific surface area of the kernels. The relative best tempering time is the period that allows the bulk of the wheat to reach the optimum equilibrium of water penetration into all kernels. Uncontrolled moisture fluctuations in wheat fed to the fully automated mill might cause operational problems, as well as change in flour quality.

Wheat-Milling Process

The objective of the flour-milling process is initially to separate the major botanical parts of the wheat kernel – the endosperm, the pericarp, and the germ – from each other. Subsequently, it reduces the endosperm to finely ground flour, which passes through a sieve with an aperture of not larger than about 200 μm. From the technical aspect of flour milling, the efficiency is weighed mainly by the level of separation among those parts, and production of a high percentage of flour with a relatively low contamination with germ and pericarp parts. The miller's objective is to produce flour with good color, and products which have not deteriorated to an unexpected level during the milling process. The practical miller, for example, would count on about 1% reduction of protein between the whole wheat and the resulting flour, when using wheat in the range of 10–13% protein.

The current basic concept of the wheat-milling process was established at the end of the nineteenth century, when iron rollers were applied instead of grinding stones. The wheat is ground between two rollers which rotate against each other and are positioned horizontally in a rollermill machine. The commercial grinding rollers dimensions range from 180 to 350 mm in diameter and up to 1500 mm in length. From a performance point of view, smaller-diameter rollers are more fitted to those stages in which the objective is a shearing action to separate the endosperm from the bran. On the other hand, larger-diameter rollers create a longer grinding path that acts on the material, mainly through compression. Most modern wheat mills have rollers that are of the same length and diameter to achieve maximum uniformity and avoid keeping a large inventory of spare rollers. (*See* **Flour**: Roller Milling Operations.)

The outer layer of the cast-iron rollers, about 13 mm thick, is harder than the inner core; the casting process is responsible for this. Frequently, the rollers are called 'chills' because they undergo a cooling process. Fast cooling after casting will keep the carbon mixed evenly with the steel and create a hard surface. Slow cooling will allow the carbon in the steel to crystallize, which will result in a softer surface. To extend the duration of the corrugations' sharpness, the harder-surface rollers are used. In the smooth rollers, made out of softer steel, the carbon crystals that are dissipated during the wearing of the surface cause it to stay rough and grip the endosperm particles better, causing an effective reduction to flour.

Mill designers use a specific and cumulative roller length in the design of a certain milling process. As an example, a mill will be designed based on a total of 18 mm roller length per 100 kg wheat per 24 h. Recently, improved machine design has reduced substantially specific roller length. Automation and high-efficiency roll bearings contribute to the ability to run the rollermill machine at a speed higher than previously. The first grinding stages in the mill, known as 'breaks,' ideally open the kernel and scrape off the endosperm from the bran. Subsequently,

smooth 'sizing' and 'reduction' rollers are used to downsize endosperm granules to fine flour. The surface of the breaking rollers is corrugated along the axis. There are fewer corrugations per centimeter in the head breaks and they increase in number to the last break. The number of corrugations per centimeter ranges from four to 10 in the breaks. The corrugations are cut in spiral, relative to the axis of the roller, ranging from 8% to 18% (4°40′ to 10°20′). A large spiral is responsible for more flour, whereas a smaller spiral generates more granular endosperm particles. The spirals cross in a scissor-like fashion when the rollers are rotating against each other. A single corrugation has a front angle and a back angle, which creates a sharper edge and a duller edge respectively (Figure 2). By rotating the rolls at different speed (differential of 2.5:1) against each other, the main effect is a shearing action of the surfaces against each other. The front angle ('sharp' edge) or back angle ('dull' edge) of one roll acts against the 'sharp' or 'dull' edge of the other roller (Figure 3). There is also a compression effect, but it is very carefully controlled just to open the kernel and release the endosperm.

Between the reduction rollers, there is a differential of ratios 1.2:1 to 1.5:1. The lower differential is responsible for a larger compression effect on the material. Smooth rollers are set closer to reduce endosperm chunks to flour and, as a result, they require considerably more power than corrugated rollers. The miller controls the compression between the rollers to avoid extensive damage to the flour and prevent excessive flacking of the endosperm particles. Flacked endosperm might also cause flour extraction loss because the tailover from the top sieves in the reduction sieving sections end up in the tailover streams of the mill and, subsequently in the bran (stock that has gone over a sieve is known as 'overs').

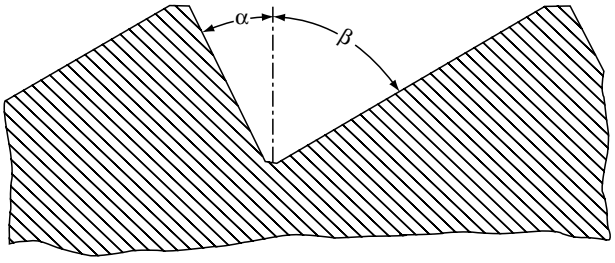

Figure 2 Roll corrugation: α front angle; β back angle.

The heterogeneous material after each stage of the grinding rollers is sent to a sieving machine called 'plansifter'. This is an enclosed unit in which up to 10 stacks or compartments, of up to 32 sieve frames, are placed. The exact number of frames depends on the materials sieved and on the manufacturer involved. The machine is centrifugally balanced and rotates at a certain prescribed number of rotations per minute and at a definite 'throw,' or radius. In a sifter compartment, sieves are arranged in groups which differ in their aperture size. Sieve aperture (measured in microns) is determined by the miller. The mill designer allocates a certain sieving surface area to each granulation of the material entering each sieving stack. Tailover materials from the sieves are directed to the next stage in the process for further grinding or purification. A certain percentage of flour is extracted from each of the sieving stages in the mill.

Purification is one of the three major processing principles in the flour milling process. The purifier is a machine in which air currents are drawn through one to three layers of sieves, and the sieves move simultaneously in a reciprocating motion (Figure 4). The material fed to the purifiers is made up of particles of about the same size range but differing in their content of endosperm and bran. The differences in relative density between pure endosperm (about 1.44 g cm^{-3}) and pure bran coat or pericarp (about 1.22 g cm^{-3}) are used in the purification process. Stratification with the help of the controlled air currents occurs between the heavier particles of relatively

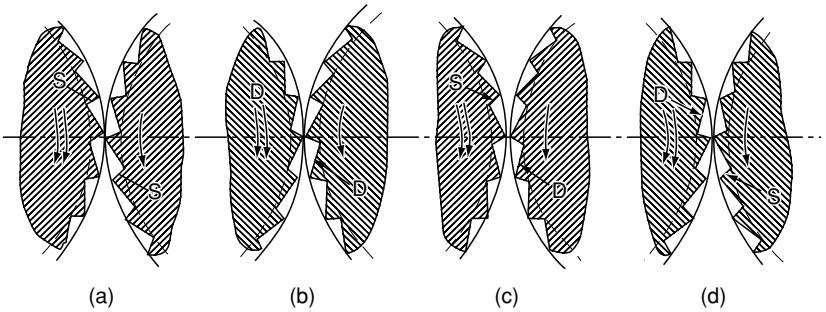

Figure 3 Roll disposition: S, sharp edge: D, dull edge of a corrugation. (a) sharp to sharp: (b) dull to dull; (c) sharp to dull; (d) dull to sharp.

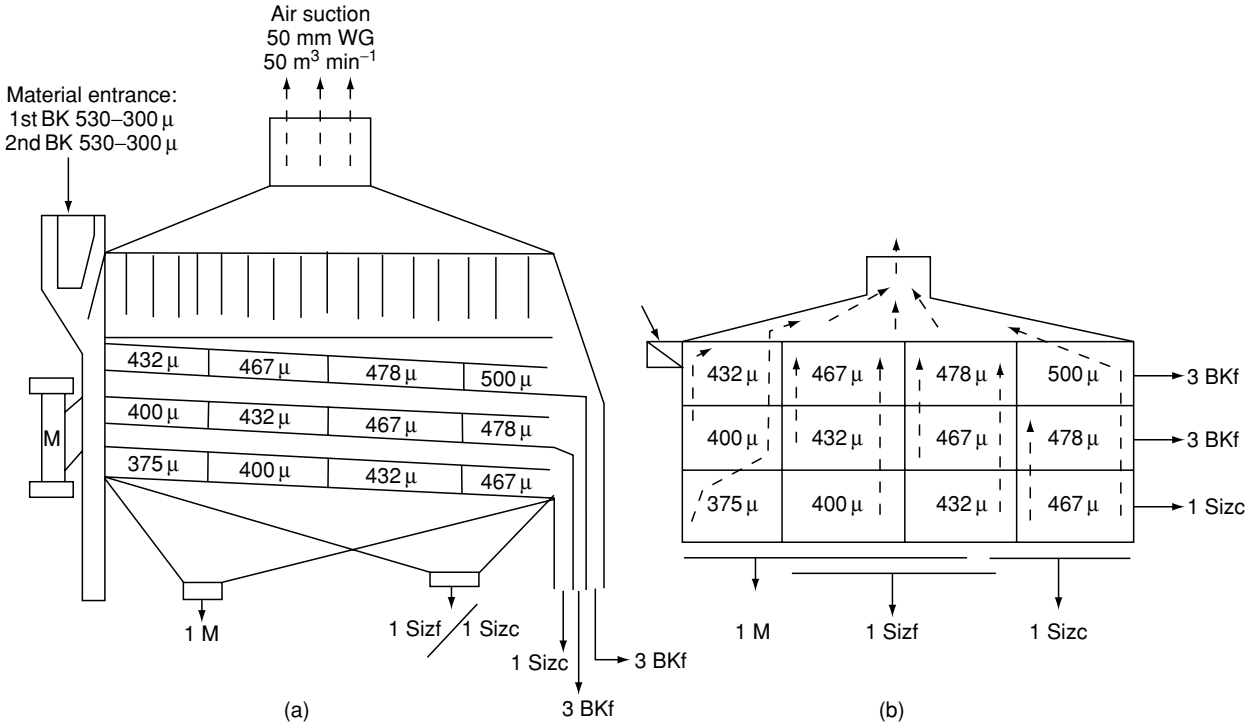

Figure 4 Schematic view of a purifier: (a) machine cross-section; (b) sieve arrangement. WG, water gauge; BK, break; M, middling; Sizc, coarse sizing; Sizf, fine sizing.

pure endosperm, which flow through the sieves, and the lighter ones of higher bran content, which tail over the sieves. The particles of pure endosperm will go through the sieves at the head-end of the machine, whereas the lighter ones will float over. Relatively pure endosperm particles of the head-end are sent to the reduction rollers for maximum production of flour. Tailover materials of the sieves are directed to sizing rollers or tail-end fine breaks.

Basically, the different stages of the process are grinding, sifting, separation, and regrinding, which are repeated again and again until the endosperm is reduced to an acceptable degree of fineness and no more flour can be separated from the bran (red dog fraction).

Each flour stream extracted under the sifter is of different quality, based on its origin within the kernel, equipment adjustment, and the stage in the milling process. The collected flours from under the sifter sections are mixed while transferred in screw conveyors (a screw conveyor uses a continuous helical attached to a drive shaft which runs the length of the conveyor) and sent to storage bins. Starting under the sifter, the miller is in position to blend those flour streams to the desired quality.

The miller also produces granular materials of relatively clean endosperm, which are coarser than flour. This type of product is called farina, when originated from hard or soft wheat, or semolina, when originated from durum wheat.

Experimental Milling

Experimental milling is an approach to determine milling quality of the wheat. The miller uses a small-scale system with which different settings can be tested to optimize the processing of the wheat in question. Results should reveal to the miller the relative expected performance of the wheat on the commercial mill. With the experimental milling system, the miller is in a position to learn about the separation of the bran, the disintegration of the endosperm, and characteristic qualities and quantities of the intermediate and final products from the wheat sample.

A milling system flowsheet, whether experimental or of a commercial unit, is the 'road map' of the process. The experimental milling flowsheet (Figure 5) describes the milling by indicating how, where, and what materials are distributed to the different machines and stages of the process. In a flowsheet, grinding rollers are designated by two adjacent circles. Some pairs are striped and some are clear, indicating corrugated and smooth rollers, respectively. A summary of roller characteristics in an experimental milling system is shown in Table 2. In the experimental milling process, after each

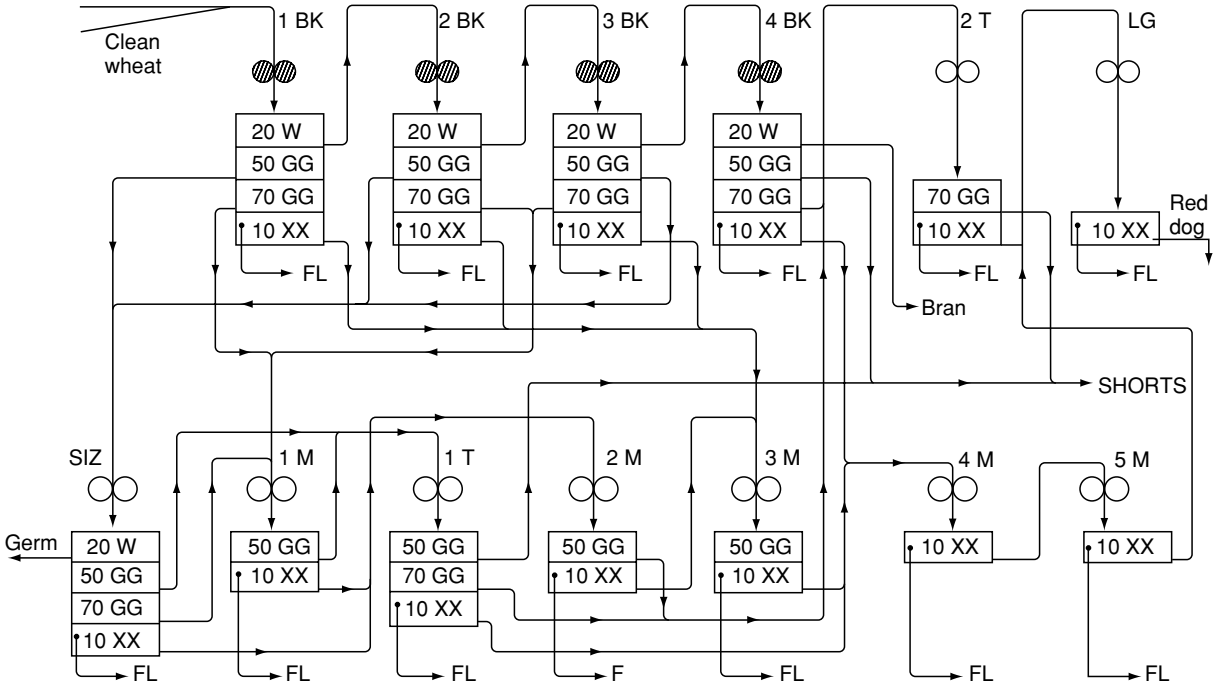

Figure 5 Experimental milling flowsheet. BK, break; SIZ, sizing; T, tailings; M, middlings or reductions; LG, low grade; FL, flour; W, metal wire sieving surface; GG, grits gauze sieving cloth; XX, double extra-quality sieving cloth.

Table 2 Experimental milling rolls surface characteristics

Stage	*Corrugations (cm)*	*Spiral (%)*	*Differential*	*Action*
1st Break	4	8	2.5:1	D:D
2nd Break	5	10	2.5:1	D:D
3rd Break	7	11	2.5:1	D:D
4th Break	8	12	2.5:1	D:D
5th Break	11	14	2.5:1	D:D
Reductions	Smooth		1.5:1	

D, dull edge of a corrugation.

grinding stage, the material is sifted as a batch on a small-size sifter which includes a stack of sieves of different apertures.

Flour Handling and Control

Flour quality control is a very important function of the milling process. Flour color, starch damage, gluten denaturation, and particle size distribution can be affected by an uncontrolled milling operation. The quality of flour, if it is not affected by the milling process, is largely dependent on the quality of wheat used for grinding. For example, softer-structured wheat kernels with a lower protein level will generate less starch damage in the flour than vitreous kernels with higher protein content. The mill laboratory routinely checks different characteristics of raw materials and finished products to determine whether or not first, they meet specifications and second, there is any need to take appropriate measures in wheat blending or mill adjustments. The final adjustment of flour quality to customer needs is usually performed by the miller by blending different flours before shipping. In modern operations, batch-type mixers blend 3–5 t of different flours with optional additives in a few minutes before they are pneumatically transported to the customer's bulk truck. (*See* **Flour**: Analysis of Wheat Flours.)

To measure mill efficiency and control performance, the miller uses a variable called ash. Ash is the completely burned remains from a flour sample, or any other material from the process burned completely and expressed in percentages of the original weight. The pericarp and the aleurone layer, which technically end up in the bran, contain about six times more minerals than the inner endosperm. The percentage of ash residue from an intermediate material in the mill or the final flour will indicate the efficiency in separating the bran and germ from the endosperm. A comparison of flours cannot be based on ash if the flours are not milled from the same wheat.

In contrast to ash, which is mainly a parameter of the mill's technical efficiency, indicating the amount of bran in flour, color is a quality parameter, which is significant to the flour end user. Accurate wheat selection, blending, cleaning, conditioning, and mill

adjustment all affect the resulting color and baking characteristics of the flour.

Flours and other mill byproducts can be mechanically modified after the milling process to accentuate certain characteristics or qualities. One of those systems is air classification of flour. In the case of flour particle sizes ranging between 1 and 200 μm, sieving can be accomplished with sieves as fine as 75 μm. Below that size, the efficiency of sieving with a sieve becomes too low. It is of interest to the miller, in some cases, to separate endosperm particles below the range of 75 μm. It was determined that broken-down endosperm would end up with the fraction in the range of 1–17 μm, which contains a major part of the protein matrix in the endosperm. Particles in the range of 17–37 μm will contain different parts but, predominantly, whole starch granules. The usage of air to segregate between particles of endosperm below 75 μm is used in the flour-milling industry for two purposes – protein shifting and narrowing the range of particle size. The principles used in air classification are based on the differences in size, shape, and weight of particles, and the characteristic terminal or critical air velocities to segregate between them. Regrinding of flours on pin mills, before air classifying, separates more starch granules from the protein matrix and improves the efficiency of the system. (*See* **Air Classification**: Principles of Operation; Uses in the Food Industry.)

Important Landmarks in the Industry

The ancient milling process was very simple. Wheat was crushed between two stones, and then the pericarp was sifted out to a limited extent using reels. The reel is a rotating cylinder sieve, and a counterrotating internal rotor throws the product against the sieve to force the fine fraction through, while the coarse fraction tails over in the end. The same principle is currently applied in high performance sieves for hard-to-sieve materials in modern flour-milling systems. However, stone mills and reels are still assembled into commercial wheat mills to accommodate specific quality and taste demands.

Currently, machines are being built at a much better technical level than during the first half of the twentieth century. Up to four pairs of rollers are installed in one machine. High-precision bearings, better adjustable mechanisms, and higher velocities currently characterize mill machinery. The result has been a significant reduction in total roll length in the mill. Water cooling of the rollers has given the miller more flexibility in manufacturing desired final products.

In the past, silk screens were used to cover the sieve frames. Advances in synthetic threads replaced the silk with synthetic material inert to environmental conditions and with more stable sieve apertures.

In the 1980s, mill automation became the main concern of the industry, for control of the process as well as online monitoring of mill products.

During the 1990s a new approach to wheat flour milling technology was established in commercial mills, abrading and polishing off the bran layers before the grinding process. This approach causes a significant reduction in the number of grinding stages and roller length in the mill. The debranning process is responsible for an improvement in flour extraction and quality. The changes in flour quality can be attributed to the heat generated during the debranning stage of conditioned wheat.

One of the important mechanical means used in the process of flour milling is air as a tool in the process. In the modern milling industry, almost 10 times more air than wheat is moved through the operation. Pressure differences created by fans or blowers in pipes between two points and the movement of an air quantity at a certain velocity make it possibile to convey materials by air. The mill, which includes a large number of machines assembled on several floors, is an industrial unit in which intermediate products flow by gravity or are conveyed by air. Air also has a positive effect on the efficiency of machine performance and sanitation. Water evaporation off surfaces of particles improves sieving of the intermediate stocks in the mill.

See also: **Air Classification**: Principles of Operation; Uses in the Food Industry; **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Flour**: Roller Milling Operations; Analysis of Wheat Flours; **Insect Pests**: Problems Caused by Insects and Mites; **Wheat**: The Crop

Further Reading

Kent NL (1966) *Technology of Cereals*. London: Pergamon Press.

Posner ES and Deyoe CW (1986) Changes in milling properties of newly harvested hard wheat during storage. *Cereal Chemistry* 63: 451–456.

Posner ES and Hibbs AN (1997) *Wheat Flour Milling*. St Paul, MN: American Association of Cereal Chemists.

Wingfield J (1989) *Dictionary of Milling Terms and Equipment*. Kansas City: Association of Operative Millers.

Types of Mill and Their Uses

F Meuser, Technical University of Berlin, Berlin, Germany

Milling and Grinding

The term 'milling' covers a wide range of processes and businesses, including primarily the methods of processing cereal grains to flours and any other processing of metal, cloth, etc., to end products of the most diverse nature in plants or factories. The plants are very often described as mills. In the context of this article, the term 'milling' will mainly be limited to milling operations in the food industry. Emphasis will be laid on the milling of cereal grains and the grinding procedures used for this purpose. For this reason, it is necessary to differentiate between milling and grinding. Milling includes all operations which lead to defined end products, and grinding covers those processes which bring about a particle size reduction.

The grinding of cereal grains can be traced back to ancient times and is linked with some of the most important advances made by humankind which came about through the development of tools for this process. It all started with the use of stones to crush cereal grains (Figure 1). This type of crushing led to the development of the pestle and mortar, and other stamping tools. The next stage of development was to grind the grains under pressure between two stone slabs. This technique was the starting point for all modern milling processes as it opened up the possibility of being able to grind the various components selectively of the cereal grain. It had, of course, always been possible to use a sieve to separate the ground material from a pestle and mortar or stamping tool into a fine fraction containing mainly endosperm and a coarse fraction consisting principally of husk and bran. However, it was possible to achieve separation far more effectively for material obtained from grinding stones. The main disadvantage of oscillatorily moved grinding stones was that they could only be operated manually. The invention of the centrally holed grinding stone made it possible for the oscillatory grinding motion to be replaced by a circular one. This development meant that (1) the manual grinding performance could be increased considerably, and (2) the required work effort could be obtained at first from animals and later from wind and water energy.

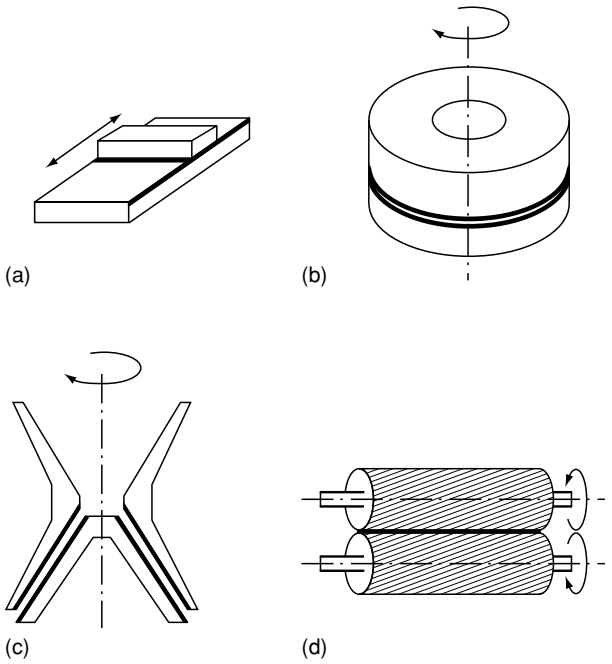

Figure 1 Historical development of grinding grain. (a) grinding slabs; (b) grinding stones; (c) grinding cones; (d) grinding rollers. Reproduced from Milling: Types of Mill and Their Uses, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

The advantages of this development were that grinding could take place at a constant speed owing to the rotary motion of the stones, and that the latter provided a larger active surface area when compared with that of grinding stones. From the rotating grinding stone, the grinding cone was evolved, enabling the size of the grinding surface to be increased in comparison with that of the rotating grinding stones. However, flat milling stones again took the place of cone mills after the advent of power transmission by cogwheels, as these enabled wind and water energy to be used to drive the grinding stones. The resulting increase in available driving power was used mainly to boost the speed at which the grinding stones turned, thus improving the milling performance.

Grinding stones driven by wind and water power remained the most important means of grinding until the invention of the steam engine enabled the performance to be increased once more. This meant that grinding rollers (rolls), which had been invented centuries previously, could now be used for milling grain. The steam engine enabled two rollers to turn in opposite directions at speeds which – when the gap between them was multiplied by the distance covered by the surface of the roller – resulted in an acting surface comparable with that of a grinding stone. Grinding rollers had the advantage that they could be set according to the fineness of grinding required. This was a development of considerable importance when compared with the rather random degree of fineness obtained by using grinding stones, as the

grinding process could then be subdivided into any number of stages by altering the size of the gap between the rollers. Thus it became possible to combine grinding and grading stages in such a way that the endosperm could be separated almost entirely from the bran and the germ.

The operations described above, by which flour, the end product, is obtained and which consists of grinding cereal grain step by step, with interposed grading and sifting stages, are referred to as milling operations. These have assumed various forms as milling, sieving, and sifting technology has developed over the past 100 years. As the methods of sieving and sifting are described elsewhere in the *Encyclopedia*, only grinding – the theory behind it and the most important types of grinders (mills) used in the food industry – will be dealt with here. (*See* **Flour**: Roller Milling Operations.)

Grinding

Purpose and Aim

The purpose of grinding is to reduce a mass to a number of parts by the action of mechanical forces (energy). In the simplest case, the sole aim of the grinding actions is to achieve a size reduction. This applies, for example, to the grinding of masses which are homogeneous with regard to material and structure. The aim is to reduce the diameter of the mass concerned and increase its surface area. Size reductions can be carried out on solids (cereal grain, sugar, cocoa, spices), liquids (oil droplets in water), and gases (gas bubbles in liquids).

Grinding solid masses which are inhomogeneous in material and structure can result in a number of particles and groups of particles, the composition of which may differ considerably from the average composition of the original mass. One of the most important aims of grinding is to obtain particles of different sizes which can then be separated selectively by means of grading, sedimentation, and sifting. Essentially solid materials (cereal grain, oil seeds, spices) are first ground in preparation for these processing stages. It is the grinding of solid materials which will be dealt with in this article. (*See* **Ground Nut Oil; Palm Oil; Spices and Flavoring (Flavouring) Crops**: Fruits and Seeds; Leaf and Floral Structures; Tubers and Roots; **Vegetable Oils**: Oil Production and Processing.)

Range of Materials

The range of solid materials which may be ground covers both hard and soft materials (Table 1). In the food industry, it is mainly products up to 4 on the hardness scale which are subjected to grinding.

Table 1 Grinding classifications according to the hardness of the material to be ground

Hardness (Mohs scale)	*Material*
6–10	Quartz, cement
3–4	Salts, coal
< 3	Cereal grains, foodstuffs
	Gypsum

Data from Zogg M (1987) *Einführung in die mechanische Verfahrenstechnik*, 2nd edn, pp. 13–76. Stuttgart: BG Teubner.

Table 2 Grinding classifications according to the particle size of the ground material

Type of grinding	*Particle size of product (mm)*	*Surface area (m^2) of a mass of spherical particles with a volume of 1 m^3*
Crushing		
Coarse	> 50	< 120
Fine	5–50	120–1200
Grinding		
Coarse	0.5–5	1200–1.2×10^4
Fine	0.05–0.5	1.2×10^4–1.2×10^5
Ultrafine	0.005–0.05	1.2×10^5–1.2×10^6
Colloidal	< 0.005	> 1.2×10^6

Data from Zogg M (1987) *Einführung in die mechanische Verfahrenstechnik*, 2nd edn, pp. 13–76. Stuttgart: BG Teubner.

Grinding may also be broken down roughly into classes according to the particle size of the ground material (Table 2). Thus it is possible to distinguish among coarse crushing, fine crushing, coarse grinding, fine grinding, ultrafine grinding, and colloidal grinding. In relation to the volume per unit weight, the surface area of the particles increases as their diameter decreases when ground.

Grinding can take place either in a gaseous medium (air, gas, steam) or a liquid one (water, oil), giving rise to the terms dry and wet grinding (or milling). The type of grinding is selected in accordance with the desired result. Dry grinding is always preferable where dry end products (flour, cocoa, sugar) are required and where the products have to remain practically unaffected by the medium. Where grinding aims to separate out intermediate products, dry grinding has the disadvantage that it is impossible to set very high standards for the purity of the individual particles and groups of particles. In this case, wet grinding can meet far more stringent requirements as the medium can be used to refine the particles. (*See* **Cocoa**: Production, Products, and Use; **Sugar**: Refining of Sugarbeet and Sugarcane.)

Actions on the Material

The grinding process is influenced not only by the hardness of the material but also by its toughness,

inner tension, and surface structure. In addition to this, the heat capacity and conductivity of the material, the manner in which they vary according to temperature, and the type of mechanical action also affect grinding.

The material may be ground by subjecting it to various types of action. According to Rumpf, there are eight types of action which can be divided into four different groups (Figure 2):

1. Actions occurring between solid surfaces (compression, friction, shear stress, cutting) which can be either the surfaces of the grinding tools or those of an adjacent body. In both cases, the extent of the actions is limited either by the maximum amount of force which can be applied, or by the smallest gap which can be set between the surfaces of the grinding tools.
2. Actions occurring on contact with a single solid surface, which could be either that of a tool or of another particle. In this case, the actions are limited by the kinetic energy of the movement of the surfaces in relation to each other.
3. Action of the surrounding medium which results from the physical state of the medium, or from changes which occur within it.
4. Application of nonmechanical energy (electrical discharges). This type of action will not be given further consideration below.

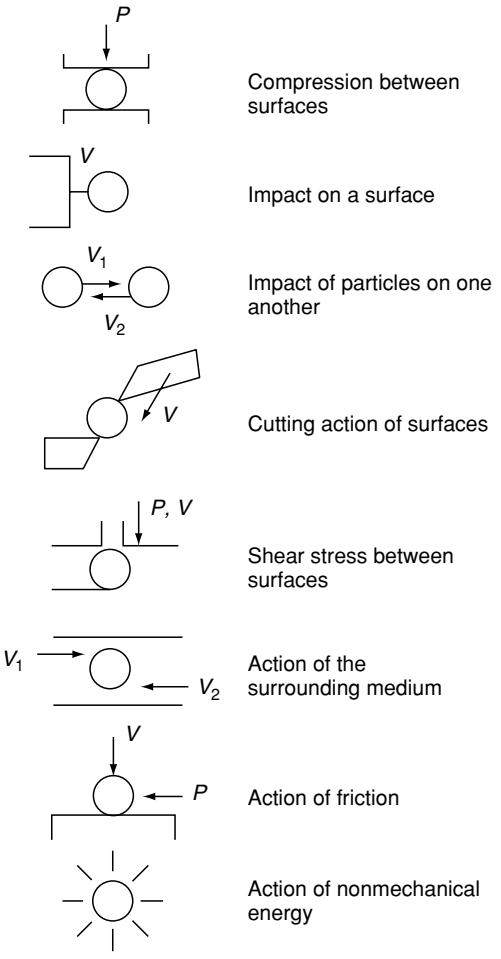

Figure 2 Types of grinding actions (according to Rumpf H (1954) Die Zerkleinerung unter besonderer Berücksichtigung lebensmitteltechnologischer Fragen I: Aufgabenstellungen. Zerkleinerungsaufgaben in der Lebensmittelindustrie. *Fette, Seifen, Anstrichmittel* 56: 404–408.

The various types of action are frequently combined during grinding operations carried out using grinding machinery. Many grinding machines are even designed to permit the combined application of several types of action. The best example of a machine of this type is the roller mill, in which compression and friction, as well as shear and cutting forces, are used in grinding.

The various types of action are combined in order to obtain the best possible grinding results. The aim is frequently to obtain a particular grinding result. Apart from the physical properties of the material and the type of action to which it is subjected, the degree of grinding also depends on the intensity of the action, which is limited by the type of grinding tool and the capacity of the material to absorb energy.

Grinding Result

The grinding result is a collection of particles which differ in size, shape, and surface structure. It is these distinctive features which provide the basis upon which the grinding result is assessed. They are determined by a variety of methods, of which those used to measure the particle size are of considerable importance. This may be determined as a function of the particular size range using various methods of analysis, for which special instruments exist (e.g., oscillating screen, air jet screen, sedimentation scales, optical particle spectrometer, laser diffraction spectrometer, image analyzer).

The methods of analysis yield different results for a single grinding result depending on the measuring technique used. Consequently, the assessment of the grinding result is affected by, amongst other things, the method by which particle size and particle size distribution are determined. Thus the choice of a particular method is always a compromise which must take account of the technical feasibility of the analyses on the one hand, and the conclusions to be drawn from the results on the other hand.

The means of measuring, e.g., in the case of the oscillating screen, are frequently smaller versions of actual sieving machines, so that it is possible to assess

the grinding results solely by analogy. In this case, the particle sizes are expressed in terms of a single linear dimension. The specific surface area of the material to be ground must, however, be determined where information on it is required for the purpose of calculating the energy needed for grinding, or in order to obtain data on those physical properties of the particles which are related to size and surface structure. It is usually obtained after first determining the particle size and is generally an approximate value. The calculation includes shape factors determined in separate tests.

Graphical Representation of Particle Size Distributions

Ground materials, such as flours, are frequently composed of particles of various sizes which cover a range of values extending over several orders of magnitude. The frequency with which particles of any one particular size occur as a result of grinding is random and follows no universally applicable distribution law. However, it is possible to describe the particle size distributions of many ground materials by means of empirically established function equations which provide a reasonable match with experimental size distribution data. A frequently used volume (mass) distribution function is the Rosin–Rammler–Sperling–Bennet (RRSB) function which describes skewed distributions of particle sizes, especially those of finely ground materials and dusts where significant quantities of particles exist above and below a predominant size.

The RRSB function may be expressed as follows:

$$F(d) = 1 - R(d) = 1 - \exp\left[-(d/d')^n\right] \qquad (1)$$

In this equation, $F(d)$ is the cumulative weight of particles passing size d, and $R(d)$ the cumulative weight of particles coarser than size d. The equation can be inverted as follows:

$$\begin{aligned}\log\log\{1/[1 - F(d)]\} &= n\log d - n\log d' + \log\log e \\ &= n\log d + \text{const.}\end{aligned} \qquad (2)$$

When this equation is applicable, a straight line, the RRSB line, results if $\log\log[1/(1-F)]$ is plotted versus $\log d$. The RRSB line is determined by the slope n (spread parameter) and the abscissa d' of the point at which $d = d'$. Where $d = d'$, $F(d) = 1 - e^{-1} = 0.632$ so that d' (position parameter) is the diameter of either the throughput of $0.632 = 63.2\%$ or the residue of $0.368 = 36.8\%$. The diameter is to be found at a point vertically below that at which the RRSB line intersects the parallel line at a distance $F = (0.632)$ from the abscissa. The slope of the RRSB line can be read from the specially devised scale at the side of the grid by parallel displacement of the line through the pole of the system of coordinates (Figure 3).

Determination of the Specific Surface Area of Particle Size Distributions

The graphical representation of particle size distributions, such as those according to the RRSB function, can also be used to obtain an approximate value of the specific surface area. To illustrate this, an approximation formula was developed for spherical particles:

$$S_m = (6.39/\rho\, d')\, \exp\,(1.795/n^2)\left(\mathrm{m^2\,kg^{-1}}\right) \qquad (3)$$

S_m ($\mathrm{m^2\ kg^{-1}}$) is the surface area as a function of the mass, and ρ ($\mathrm{kg\ m^{-3}}$) is the density of the mass.

As the specific surface area of a mass of spherical particles is, apart from the density, dependent only upon the spread parameter (n) which characterizes the RRSB line and the position parameter (d'), it is possible to place a scale at the side of the RRSB grid with which the specific surface area can be calculated as a function of the volume.

In Figure 3, the dimensionless characteristic value $S_v\, d'/\phi$ is given in the scale at the side of the grid. In this characteristic value, ϕ represents a form factor with which the deviation of the particles from the spherical form is taken into account. Nonspherical particles have larger surface areas than spherical ones and the form factors of these are therefore greater than 1. For example, the form factors for ground wheat, rye, barley, and oats are 1.2, 1.3, 1.5, and 1.8 respectively. (*See* **Barley**; **Oats**; **Rye**; **Wheat**: The Crop; Grain Structure of Wheat and Wheat-based Products.)

Using the outer scale at the side of the grid, the surface area is calculated by moving the RRSB line from the position parameter d' to the pole, keeping it parallel to its original position. The point at which the RRSB line intersects the outer scale is the characteristic value of the position and spread parameters. The specific surface area as a function of the volume S_v ($\mathrm{m^2\ m^{-3}}$) is determined by dividing the characteristic value obtained by the position parameter d' and then multiplying it by the form factor ϕ. The value thus obtained can be related to the mass by inserting the density (ρ):

$$S_m = S_v/\rho \quad \left(\mathrm{m^2\,kg^{-1}}\right) \qquad (4)$$

It should be pointed out that the data for the specific surface area thus obtained may frequently differ considerably from those determined by other methods

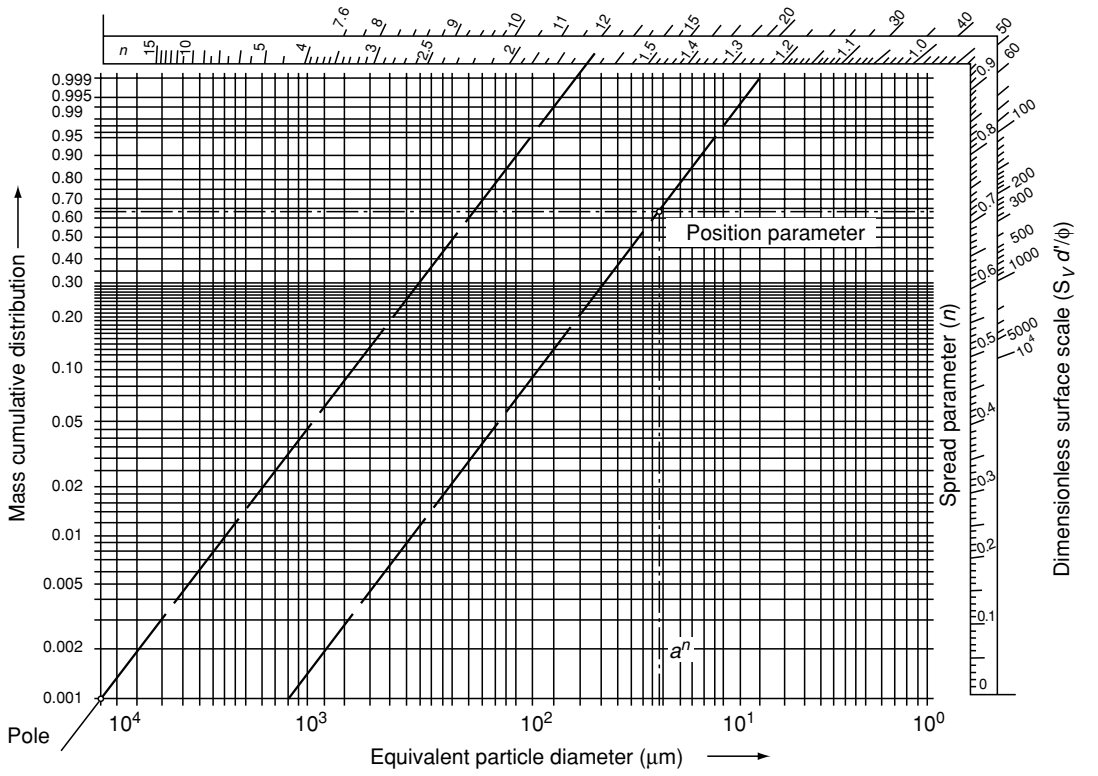

Figure 3 Graphical representation of particle size distribution, according to the Rosin–Rammler–Sperling–Bennet (RRSB) grid. Reproduced from Milling: Types of Mill and Their Uses, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

(e.g., adsorption and permeability methods). The values for specific areas obtained by calculation are therefore frequently of limited value only.

Energy Required for Grinding

Grinding requires energy which has to be applied to the material to be ground by means of tools. Part of the work applied is used to drive the tools and only the remaining, often much smaller part is used in the grinding process itself. Several equations have been developed to estimate the amount of energy required, based on different concepts of the grinding process. All formulae are based on model concepts in which it is assumed that separate particles are to be ground. They are therefore founded upon very rough simplifications compared with the actual grinding processes which take place in grinders. Thus the formulae yield results which are of value in certain areas of grinding only. For example, the energy required for grinding according to Rittinger's law is proportional to the sum of the newly created surfaces, irrespective of the type of mechanical action (tension, compression, shear stress).

Rittinger's law obeys the following function equation in which K_R is a proportionality factor (Rittinger's constant) which depends on the material and the grinder. Given the same material and grinder, it is independent of the degree of grinding. D_0 is the grain diameter prior to grinding and D_n the grain diameter after grinding. Rittinger's law is mainly used to estimate the energy A_R required to grind either fine or brittle materials.

$$A_R = K_R(1/D_n) - (1/D_0) \qquad (5)$$

By contrast, Kick's grinding law states that the energy A_K required for grinding consists entirely of the elastic and residual deformation of the particles to be ground, this energy being proportional to their volume. The function equation is written as follows:

$$A_K = K_K \log(D_0/D_n) \qquad (6)$$

K_K is also a proportionality factor (Kick's constant). Kick's law applies above all to coarse materials with fewer brittle and more elastic characteristics. In practice, however, it is impossible to obtain an exact result for the energy required for grinding by means of these formulae as it is influenced by too many factors, including the design of the grinder and its mode of operation. Neither is it possible to calculate the energy required from the particle size distributions before and after grinding or from

the mechanical strength of the material to be ground, as a considerable part of the energy applied is used not for grinding but for moving the material, or is absorbed by the material which has already been ground.

Grinders (or Mills) Used in the Food Industry

In view of the requirements which ground material has to meet and the large number of materials which must be ground to produce foodstuffs and similar goods, the differences in shape, size, composition, and strength characteristics have led to the development of grinders (or mills), the design of which is the result of the type of grinding operation to be performed. The grinders operate by applying compressive, shear, frictional, and cutting forces or, as is frequently the case, various combinations thereof can be applied to the material to be ground either simultaneously or consecutively, depending on the design of the grinder.

The application of mechanical energy to the material being ground can lead to the latter heating up considerably. The grinder parts are also heated by the energy conversion. Where the design and action of a grinder cannot contribute adequately both to dissipating the heat produced in the material in order to preserve its quality and to keeping the operating temperature of the machine within tolerable limits, it is necessary to cool the grinding tools, as a result of which the material is also cooled.

Coolant may therefore be pumped through the grinding tools and the surrounding sections of the housing. The rollers of roller mills or the rotor and stator of stirrer bead mills, for example, may be cooled in this way. In impact mills, the material being ground can be cooled by increasing the air flow through the impellers attached to the shaft of the mill.

Specially constructed impact mills, which are equipped with heat exchangers, enable the transfer of evaporable mass from the material being ground to the airstream. Thus a specific exchange of heat and mass occurs, during which the material being ground is dried. The machines in which grinding and drying take place simultaneously are referred to as drier-mills.

A few important examples of mills used in the food industry to process raw materials in order to obtain specific ground materials are described below. They have been selected in order to illustrate the most important principles of grinding and their applications by means of concrete examples. Hence the list of mills is not exhaustive, either with regard to the grinding principles or the design of the mills.

The majority of mills work continuously. The material to be ground is fed into them by means of conveyor equipment. The grinding of materials in air or gas is referred to as dry grinding (or milling) and that of suspensions as wet grinding (or milling). Some of the mills described below are designed to be operated in both dry and wet milling.

Roller Mills

Roller mills are used both for solely grinding and for selectively grinding and milling particles of granular shape. Their most important application is the selective grinding of cereal grain. In addition to this, roller mills are also used for fine grinding of cocoa and spices.

The great advantage of milling cereal grains by means of grinding rollers is that the process can be subdivided into many stages, the grinding forces acting on the material only once at each stage. A further advantage is that the mills can be set according to the material to be ground and the required grinding process, because not only can the distance between the rollers be freely selected, but also their diameter, surface structure, the speed at which they turn, and the differential speed.

Roller mills consist of a housing in which, as a general rule, several pairs of rollers in bearing rings are placed on either side on sliding bearings or on self-aligning roller bearings. Figure 4 shows a modern roller mill with four horizontal pairs of rollers which are used to grind cereal grains. The material to be ground is fed into the machine by two feed rollers which are located above the two upper pairs. The ground product passes, without sifting, directly from the upper pairs to the two lower pairs of rollers, where it is subjected to further coarse or fine grinding. In order to improve the supply of material to the grinding rollers, the machine is equipped with a fan to extract the air from the gap between the rollers. The latter can be set automatically, the automatic setting device being connected to a computer in which the optimum roller gap settings for all grinding phases of a grinding diagram are stored, thus insuring the maximum flour production for any given cereal or mixture of cereals.

The rollers of roller mills are driven by flat belts or, in the case of single rollers, by fan belts. The different roller speeds are produced by a cogwheel or chain drive. Driving the rollers by chains or stepped gears enables the distance between the axes of the pairs of rollers to be adjusted without changing the transmission wheels.

The rollers are made of centrifugally cast, chilled cast iron which is extremely dense and wear-resistant.

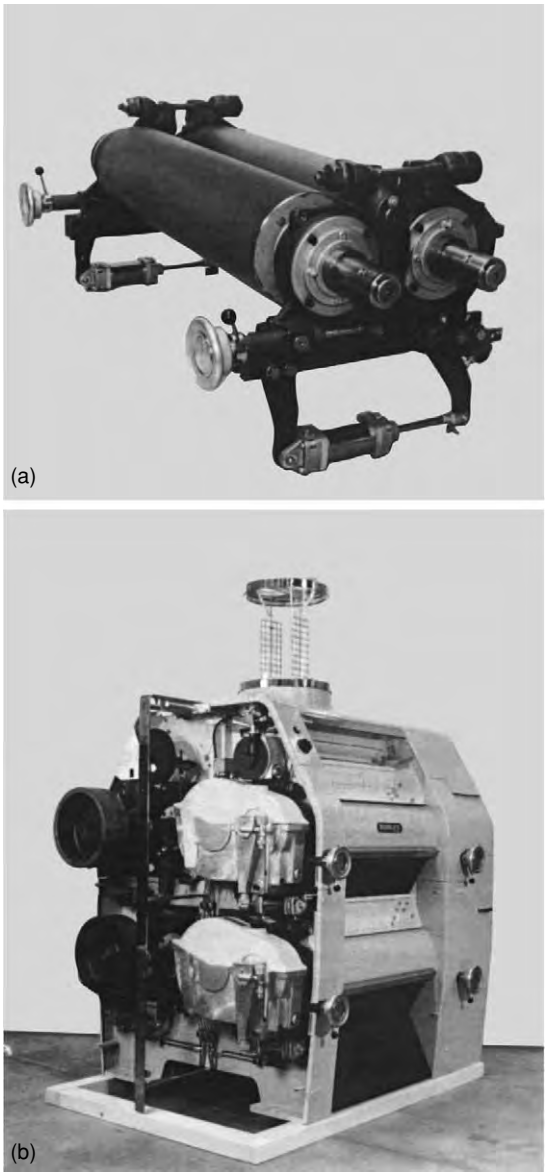

Figure 4 Roller mills: (a) pair of rollers; (b) eight-roller mill. Courtesy of Bühler AG, Uzwil, Switzerland.

Depending on the hardness of the cast iron, the rollers are either ground or corrugated by machine to produce either smooth or corrugated rollers.

The rollers, which are standardized, have a diameter of 220–315 mm, and a length of 315–1500 mm, and can be machined several times. They may also be made of porcelain. This type of roller has a working surface which comprises a porcelain sleeve with a wall thickness of approximately 50 mm. They are 220–350 mm in diameter and can be up to 1000 mm in length.

The grinding rollers can heat up considerably owing to the production of heat caused by friction, rendering it difficult to set the gap to a fixed distance, especially when subjected to a high degree of wear. As this can have a negative effect on the grinding result, water-cooled grinding rollers have been developed which are cooled by pumping water over the shaft through the hollow rollers. Cooling increases the specific grinding performance and leads to a more even grinding result.

As it is possible to change the working parameters, the material can be ground between the rollers by compressive, shear, and cutting forces. The length of time during which these forces act on the material depends on the radius of curvature and the absolute circumferential speed of the rollers as well as on the ratio between the gap between the rollers and the size of the material to be ground. In order for the material to be fed into the gap, the feed angle must be smaller than the angle of friction. This means that the ratio between the diameter of the rollers and that of the material must be large. The larger the diameter of the rollers, the slower the circumferential speed and the smaller the gap selected, the greater the forces acting on the material will be. The forces are limited by the hardness of the material being ground and its fracture.

Compressive forces only act on the material where the circumferential speed of the rollers is the same and the surface of the rollers is smooth. This results in the product being ground by being deformed to rolled plate-like particles, especially in the case of soft products such as cocoa. By comparison, different roller speeds give rise to shear forces which, combined with a corrugated surface and the finite distance between the rollers, results in a slicing action. This type of grinding occurs above all during the sharp-to-sharp grinding action of the corrugations, while the dull-to-dull grinding action grinds the product mainly by compression. Grinding predominantly follows the product's own fracture which can be deduced from its natural structure.

This is the case, for example, when grinding many types of cereal grains: the particle size distributions of the ground products exhibit characteristic differences which depend on the hardness of the material. Where the dull-to-dull grinding action of the corrugations is applied, it is possible to guide the grinding result towards the RRSB function. Combined with the slicing action of the rollers, it is also possible to impose a characteristic distribution deviating from the RRSB function on to the particle size of the ground product. The range of particle sizes and the shape of each particle are also influenced to a large extent by the design of the corrugations.

The corrugations are shown in section in Figure 5. The angle of the corrugations is composed of the cutting angle (α) and the rear angle (β). The slope of

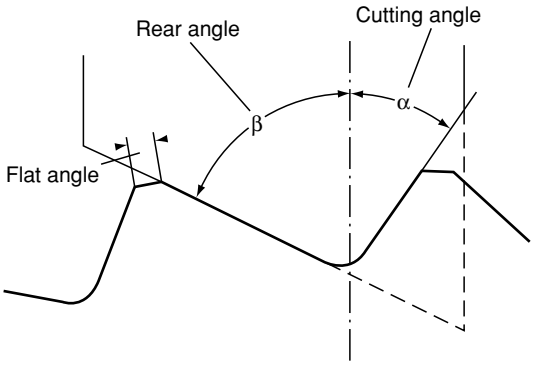

Figure 5 Corrugation of a grinding roller.

the corrugations, which runs parallel to the axis of the roller, is referred to as the twist. The cutting and rear angles, together with the position of the corrugations, and the absolute speed of the surface of the roller casing determine the efficiency of the grinding process. Where the cutting angle is small, the cutting effect of the corrugations predominates, favoring the production of grits. A larger cutting angle causes the material to be crushed instead, resulting in the production of flour.

The size of the cutting angle ranges from 15° to 45° and that of the rear angle from 60° to 75°, so that the angle of the corrugation usually lies in the region of 90–115°. The size of the slope of the corrugations, measured at the circumference, may range from 20 to 300 mm, given a twist of 5–20% with a roller length of between 400 and 1500 mm, for example. The flat angles of the corrugations have a lateral length of 0.05–0.30 mm, measured in the direction of the circumference of the roller. The depth of the corrugations (h) can be calculated using the formula given below, ignoring the fact that the root of the corrugation is rounded:

$$h = (T - S)/(\tan\alpha + \tan\beta) \qquad (7)$$

The efficiency of the roller mills depends on the amount of material fed on to the rollers, the type of corrugations, and the differential speed of the rollers as a function of the circumferential speed. The amount of material processed is limited by the air cushion which forms over the rollers owing to the parameters referred to above. The air cushion is therefore frequently reduced by means of special feed mechanisms in order to increase efficiency, either by compacting the material to be ground, for example, or by drawing off the air above the gap.

The differential speed can be as high as 1:15, depending on the product. For this ratio, the driven roller, which has a fixed position, is reduced to the circumferential speed of the movable roller. The upper limit of the circumferential speed is in the region of $6\,m\,s^{-1}$. For example, corrugated and smooth rollers with a standard diameter have a circumferential speed of $3.3–4.4\,m\,s^{-1}$ when grinding wheat and rye at a speed of 250–360 rpm.

Hammer Mills

Hammer mills are mainly used for coarse grinding of brittle and fibrous materials. The material is ground by beating, during which it may be partly broken down into its constituents. The disintegration of the material often serves to prepare it for subsequent processing stages, such as separating the constituents by sieving, squeezing, and extraction.

The mills consist of a rotor with suspended hammers which swing freely and which are then forced upwards and outwards radially by centrifugal force. The hammers strike a sieve casing fitted with impact ribs. The material, which is fed into the machine tangentially from above, is crushed by impact and rebound forces and expelled through the sieve casing. The fineness of grinding depends on the size of the sieve perforations, the circumferential speed, the number and width of the hammers, and the throughput.

Beater Mills

The structure of beater mills is similar to that of hammer mills. Their rotors are equipped with rigid ribs, plates, or arms, arranged in the shape of either a cross or a star. Such mills are suitable for finely grinding spices, for example. They may also operate as wet-milling colloidal mills, in which part of the energy used in grinding is applied by imploding cavitation bubbles which form on the backs of the rotor parts.

Slicing Mills

Slicing mills are mainly used to grind products which are resistant to compression, impact, and friction, such as fruits and vegetables. The rotors of slicing mills are fitted with removable blades with strike blades attached to the housing and a sieve casing. The knives slice the material which is fed into the mill from above, after which it passes through the slits in the sieve casing. The material is ground as a result of the shear stress which builds up between the stationary and rotating blades.

Toothed Disk Mills

Toothed disk mills are used for fine dry- or wet-grinding of soft, nonadhesive materials. They comprise

Figure 6 Toothed disk mill. Courtesy of Dorr-Oliver Deutschland GmbH, Grevenbroich, Germany.

a housing containing a stator and a rotor (Figure 6). The material to be ground is fed into the mill through an opening in the centre of the stator and falls on to the rotor. It is then transported by shear, compressive, and frictional forces through the teeth arranged in rows on the disk from the center to the outer rim, being ground in the process. The ground material passes into a collecting ring, from which it is transported into a radial outlet. The final fineness of the ground material is determined by the distance between the axes of the disks, the profile of the grinding tools, and the speed.

Pinned Disk Mills

Pinned disk mills are used to grind either oily materials or moderately hard to soft materials which are dry or contain little moisture. They comprise a housing containing a rotor and a stator, both of which are fitted with removable pins arranged in concentric circles such that they engage the opposite spaces. The material to be ground is fed into the mill through the center of the stator, falling on to the rotor. The radially increasing circumferential speed causes the rotor pins to press the material at high speed through the rows of pins, grinding it in the process by impact and rebound. The final fineness of the ground material depends on the distribution and shape of the pins, the circumferential speed of the rotors, the throughput, and the strength characteristics of the material.

Pinned disk mills may be used as either dry or wet mills. They are employed as wet mills in starch extraction, for example, where steeped maize is ground in order to remove the germ by disintegrating the maize kernels.

Underrunner Disk Huller

The underrunner disk huller is a corundum stone disk mill. It is used above all for dehulling rather than grinding, serving as the main huller in rice mills to separate the husk from the rice kernel. Other cereal grains and leguminous seeds, such as beans, peas, or lentils, may also be dehulled using this machine. (*See* **Legumes**: Legumes in the Diet; **Peas and Lentils**.)

The huller consists of a stator disk located above a rotor disk, both of which have abrasive coverings of corundum stone. The hulling process and its effectiveness are the result of the grains being rubbed between the disks and the distance between them. The grains describe a curved path between the disks, passing from the central opening to the outer rims. The disks are surrounded by a casing which is completely closed.

Stirrer Bead Mills

Stirrer bead mills are employed predominantly in the chocolate and confectionery industry, although they may also be used in the delicatessen industry. They are used for fine grinding (1–50 μm) of raw cocoa and chocolate, as well as in the production of peanut and hazelnut fillings and in the manufacture of mustard. (*See* **Mustard and Condiment Products**.)

Stirrer bead mills comprise a cylindrical or conical rotor body which moves in a corresponding stator body. In the example illustrated (Figure 7), a gap (2–25 mm) is formed between the conical rotor and the stator, and in this gap the beads move radially outwards from the centre. The material is thus ground uniformly, resulting in a close particle size distribution. The geometry of the grinding space insures that every single grain in the suspension passes

Figure 7 Coball mill. Courtesy of FRYMA-Maschinen AG, Rheinfelden, Switzerland.

along the prescribed path through the grinding /element. The suspension is fed into the mill through a central inlet and discharged through an outlet attachment near the shaft, the beads being separated from the ground material at the same time.

A large number of different settings is possible for this mill, depending on the fineness of grinding required, as not only the rotor speed may be freely selected but also the width of the gap between rotor and stator, the number of beads, the material of which they are made (glass, steel, ceramic, etc.), their diameter, and the rate of flow of the suspension. The rotor, stator, and cover of the mill can be cooled so that it is possible to maintain the suspension at a moderate temperature.

Drier-Mills

Among the mills which simultaneously grind and dry, particular attention is drawn here to one which is frequently used in the food industry, both as a grinder and as a drier-mill. It is known as the Ultra-Rotor mill and is used for fine grinding of materials such as starch derivatives, guar, and stearates. Furthermore, it is also employed in the selective grinding of various cereals which, when combined with subsequent sieving or sifting, serves to separate the endosperm fraction from the bran. As a drier-mill,

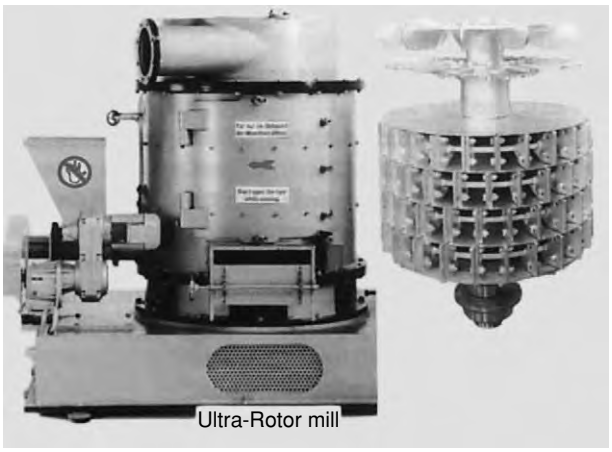

Figure 8 Drier-mill. Courtesy of Altenburger Maschinen, Jäckering GmbH, Hamm, Germany.

it is used above all in wheat starch plants to dry wheat gluten.

The mill consists of a specially designed rotor (Figure 8) mounted vertically in a housing which acts as a stator and is equipped with a removable corrugated inner casing. A fan disk is attached to each end of the rotor shaft. Disk-shaped grinding tools, arranged in a radial pattern, are fitted to plates between the fan disks. A second disk, on to which rod-shaped metal parts are screwed, is located below the upper fan disk. Between rotor and stator there is a narrow gap, through which the material to be ground passes at high speed.

The material is fed into the machine through an inlet in the wall of the housing at the level of the lowest grinding disk, being drawn into the grinding zone by the air stream produced there. The rotation of the rotor causes strong eddies of air to form in the chamberlike grinding tools which, together with the impact of the particles on the walls of the grinding tools and the stator casing, grind the material. Owing to the high throughput of air, a large part of the heat produced in the material during grinding is expelled with the air.

After leaving the grinding zone, the ground material passes into the space between the disk fitted with metal rods and the upper fan disk. The impact of the material on the rods produces a sifting effect. The rods cause the coarser particles to concentrate in the proximity of the wall of the housing, from which they are extracted through an adjustable opening placed at a tangent, and then recycled. The finer particles leave the mill as a finished ground product, being collected separately after passing through a sieve and subsequently through filters. The mill can be used to advantage to dry the ground material owing to the way in which the air flows and

the type of grinding. Only the air stream requires heating. This is best carried out by means of an interposed heat exchanger. The formation of eddies of air and the considerable increase in the surface area owing to grinding result in an intensive exchange of material and heat, so that the air outlet temperature does not need to exceed 70 °C even where the air inlet temperature is as high as 400 °C. The temperature of the product is normally in the range of 40–50 °C and the limiting dew point temperature is in the region of 38–45 °C. The thermal efficiency of the mill is approximately 90%. The average heat consumption for simultaneous grinding and drying is about 2800–3000 kJ kg^{-1} of evaporated water.

See also: **Barley**; **Cocoa**: Production, Products, and Use; **Flour**: Roller Milling Operations; **Legumes**: Legumes in the Diet; **Mustard and Condiment Products**; **Oats**; **Peas and Lentils**; **Rye**; **Spices and Flavoring (Flavouring) Crops**: Fruits and Seeds; Tubers and Roots; **Sugar**: Refining of Sugarbeet and Sugarcane; **Wheat**: The Crop; Grain Structure of Wheat and Wheat-based Products

Further Reading

Bernotat S and Schönert K (1988) Size reduction. In: Gerhartz W (ed.) *Ullmann's Encyclopaedia of Industrial Chemistry, Unit Operations I*, 5th edn, vol. B 2, pp. 5/1–5/39. Weinheim: Verlag Chemie.

Earle RL (1983) *Unit Operations in Food Processing*, 2nd edn, pp. 159–165. Oxford: Pergamon Press.

Leschonski K (1988) Particle size analysis and characterization of a classification process. In: Gerhartz W (ed.) *Ullmann's Encyclopedia of Industrial Chemistry, Unit Operations I*, 5th edn, vol. B 2, pp. 2/1–2/33. Weinheim: Verlag Chemie.

Loncin M (1969) *Die Grundlagen der Verfahrenstechnik in der Lebensmittelindustrie*, pp. 726–747. Aarau am Main: Verlag Sauerländer.

Meuser F, Reimers H and Fischer G (1977) Physikalische und chemische Untersuchungen an Mehlen aus unterschiedlichen vermahlenen Weizen im Hinblick auf deren Backqualität. *Deutsche Müller-Zeitung* 75: 145–154.

Pallmann H (1977) Zerkleinerungstechnik im Bereich der Nahrungs und Genussmittelindustrie. *Verfahrenstechnik* 11: 270–276.

Rumpf H (1954) Die Zerkleinerung unter besonderer Berücksichtigung lebensmitteltechnologischer Fragen I: Aufgabenstellungen. Zerkleinerungsaufgaben in der Lebensmittelindustrie. *Fette, Seifen, Anstrichmittel* 56: 404–408.

Vauck WRA and Müller HA (1988) *Grundoperationen chemischer Verfahrenstechnik*, 7th edn, pp. 254–303. Weinheim: Verlag Chemie.

Zogg M (1987) *Einführung in die mechanische Verfahrenstechnik*, 2nd edn, pp. 13–76. Stuttgart: BG Teubner.

Characteristics of Milled Products

E S Posner, ESP International, Savyon, Israel

Background

The quality and characteristics of commercial mill products are a matter of geography and are set in commercial mills based on end users' expected demands. In addition to quality characteristics, specifications may include notes stating that products are untreated with chemicals during production, free from foreign matters, rancidity and/or other alimentary substances, suitable for human consumption and free from dead or alive insects in any developing stages. Nomenclature of wheat products such as 'All-Purpose Flour,' 'Type 00,' or 'Type 812' creates confusion, especially in a global trading environment. An international effort is required to set guidelines for wheat mill product specifications. In many cases, the flours originate from mixes of different kinds of wheat or blends of flours after processing. Following the description of the different wheat products, quality variables of products and milling methods that control the characteristics are dealt with.

Mill Products Description

Figure 1 shows a schematic diagram exemplifying products from a flour mill and their proximate analysis. Products in this case are referred to in typical USA nomenclature.

Flour Types from Milling Operations

Straight-grade flour This denotes all the flour streams generated in the mill combined to one final product. Depending on the wheat kind and mill design, in general, straight-grade flour would have a protein content about 1% below that of the wheat.

All-purpose flour This is medium-protein wheat flour made from hard wheat or a combination of soft and hard wheat. All-purpose flour is used mainly for a wide range of household baked products. The color and gluten characteristics are the important variables.

Patent flour This is the blend of flour streams from the front head of the mill. It is low in ash and protein, with a good color that is considered highest in economical value. Various blends of flour are named patent such as 'top patent' or 'baker's patent,' etc. They differ in the amount of clear or tail end streams that they contain.

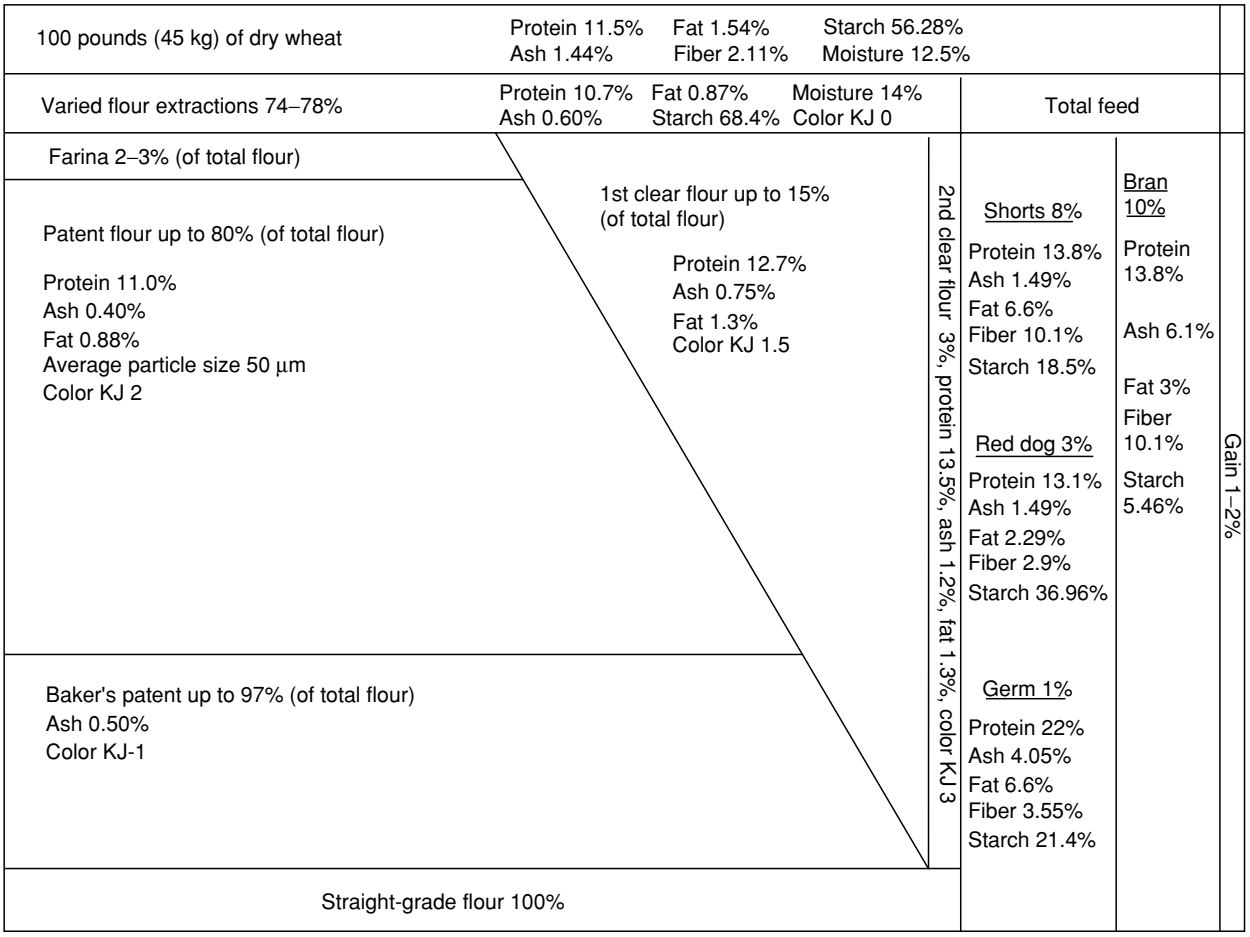

Figure 1 Products from a typical wheat-milling system. Adapted from Posner ES (2000) Wheat. In: Kulp K and Ponte JG, Jr. (eds) *Handbook of Cereal Science and Technology*. New York: Marcel Dekker, with permission.

Clear flour The portion of flour streams remaining after patent flour has been taken off. Clear flour normally contains a higher amount of bran particles, ash, and protein than patent and economically is secondary in value. The miller can produce more than one clear by further dividing the tail end flour streams.

Farina Regulations by the US Food and Drug Administration (FDA) define farina as a coarse granular separation of endosperm material extracted from wheat other than durum wheat and red durum wheat. Farina passes though a 841-μm (No. 20 US Standard) sieve, but no more than 3% passes through a 150-μm (No. 100 US Standard) sieve. It is separated from the bran coat, or bran coat and germ, to such an extent that the percentage of ash therein, calculated to a moisture free basis, is no more than 0.6%. Its moisture content is no more than 15%.

Semolina Semolina is defined in a similar way to farina, except that it is made only of durum wheat, and its maximum dry ash content is 0.92%. Durum wheat semolina is used for the production of pasta products. Coarse durum semolina, preferably between 550 and 1100 μm is used for the production of couscous, which is consumed mainly in North Africa. Semolina extraction can be up to 72% of the wheat kernel.

Durum flour Durum flour particles are such that 98% passes through a 212-μm (No. 70 USA) sieve. It is separated from the bran coat, or bran coat and germ, to such an extent that the percentage of ash therein, calculated on a moisture-free basis, is no more than 1.5%, and its moisture content is no more than 15%. 'Durum patent flour' is the re-ground semolina. 'Durum first clear flour' is Kaan, defined by the FDA as the through of a 150 μm (No. 100 USA) sieve. 'Durum second clear flour' is made during the milling of semolina, has a high amount of ash, is 'specky,' and has a dull color. It is used in the manufacture of pet foods. 'Durum granulars' is a combination of durum

semolina and durum flour containing no more than 15% durum flour. The ash content is normally about 0.72%. It is sometimes used as a substitute for durum semolina.

Atta flour This is high-extraction (92%) ground material from wheat, and is used mainly in the Indian subcontinent for baking chappati. The high degree of extraction leads to a darker shade and a maximum ash value of 2.0% (d.b.).

Wholewheat flour Wholewheat flour, also named 'Graham flour,' is a product that includes the endosperm, germ, and bran from cleaned kernels. The FDA indicates that 90% of ground wholewheat should pass through a 2.38-mm (No. 8 US Standard) sieve and no less than 50% through an 841-μm (No. 20 US Standard) sieve. Bran particles in the meal should pass through a 1.191-mm (3/64″) screen. However, meal granulation is subject to customer demand and depends on the grinding system. The product is used in some cases in different ratios with white flour. The bran particles in the meal interfere with the gluten structure of the dough causing a reduction in baked volume. Usually, to achieve an acceptable baked product from wholewheat flour, vital wheat gluten is added to improve the baking performance. The inclusion of the germ in the product limits its storability period.

Byproduct Characteristics

Mill products that do not coincide with flour are considered by the miller as byproducts and are usually diverted to feed. The total byproducts from the mill are named 'Mill Run' or 'Wheatfeed.' Different combinations of mill byproduct exist in the trade and vary in name and analysis. Proximate values of feed products are shown in Table 1.

Middlings These consist of fine particles of wheat bran, wheat shorts, wheat germ, wheat flour, and some of the offal from the tail of the mill.

Bran Bran consists of coarse outer layers of the wheat kernel. Large pieces of bran remain after the flour has been extracted from the wheat.

Shorts Shorts is a mixture of fine bran, endosperm, and some germ fragments that remain after flour extraction has been completed and used for animal feed.

Red dog This is a mixture of endosperm and bran powder, taken from the tail of the mill, that has a high ash and protein content, poor dress, and is dark in color. It is a product intermediate between a low-grade flour and feed.

Wheat germ products The germ of the wheat kernel consists of two parts, the scutellum and the embryo. The scutellum is softer (with a fat content of about 32%), and the embryo is relatively loosely attached to the scutellum and covered with the pericarp. In general, it is the embryo that is referred to in the trade as wheat germ. The 'wheat germ meal' consists of embryo together with some bran and shorts, and contains at least 25% protein. 'Pure wheat germ' is used for human consumption and contains at least 30% protein.

Effect of Wheat as the Raw Material

Flour quality depends about 75% on the quality of the raw wheat and about 25% on mill design procedures and adjustment. The different wheat kinds vary in the level of protein characteristics, which are mainly related to the quantity and quality of insoluble protein that forms the gluten. Wheat kernels vary in terms of the color of the pericarp. Accordingly, white wheat has an advantage over red wheat by allowing a higher flour extraction with a comparable color. Figure 2 shows the relation between wheat qualities of flours and their usage for different end products.

The objective of the milling process is to reduce the wheat kernel and deliver different products of consistent quality to bakers or other end users. To achieve this objective, the miller mixes wheat and blends final flours with particular specifications in mind. The reason is that wheat quality in commerce changes continuously as a result of variety, soil, and climate. Even in areas where only the local wheat is used by a mill, the miller adjusts the quality by mixing.

Table 1 Proximate analysis of mill feed products

Product	*Minimum protein (%)*	*Minimum fat (%)*	*Maximum fiber (%)*	*Maximum moisture (%)*	*Maximum ash (%)*	*kg m^{-3}*
Middlings	13.0	3.0	9.5	14.0	4.2	288–400
Bran	14.0	2.5	12.0	14.0	6.0	256
Shorts	13.0	3.0	7.0	14.0	4.1	160–256
Red dog	15.0	3.0	4.0	14.0	1.64–2.85	352–448

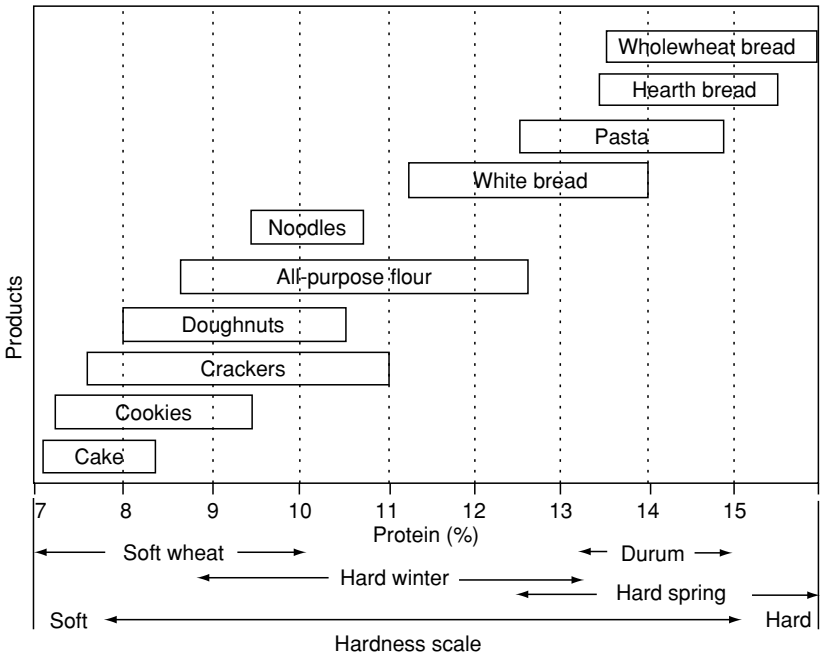

Figure 2 Schematic diagram of relationship between percentages of wheat protein, wheat type, hardness, and end-product utilization. Adapted from Posner ES (2000) Wheat. In: Kulp K and Ponte JG, Jr. (eds) *Handbook of Cereal Science and Technology*. New York: Marcel Dekker, with permission.

Definitions of Wheat Product Variables

The following explanation of milled wheat product variables that affect characteristics are arranged in alphabetical order. Different official methods and instruments exist to measure flour and other mill products. Flour analysis allows an approximation of the determined values. Flours with the same analysis have been found to show different baking characteristics or end-use performance. The interaction of a few or more characteristics, rather than one alone, affects the end-use qualities. (*See* **Wheat**: Grain Structure of Wheat and Wheat-based Products.)

Ash

Ash, or the mineral content of the wheat or its product, is determined by incinerating the sample and expressing the residue as a percentage of the original sample. The ash content of the wholewheat kernel depends on the wheat variety as well as the soil, irrigating water, and growing conditions. However, erroneously, flour ash is considered as a grading factor by some in the industry. A better expression of milling performance is where the whole kernel ash is considered. Accordingly, ash values of the wheat kernel products should not be taken as absolute numbers but expressed as a ratio of flour ash divided by wheat ash (FA/WA).

Where ash is considered as grading factor types 405, 550, 630, 812, and 1050, the ash range for each grade is quite wide. For example, flour could be considered as type 550 when the flour's ash content is between 0.490 and 0.580% (d.b.). Under regulations in the USA, the ash content added to flours as iron, or salts of irons, or calcium is excluded in calculating the ash of the shipped flour.

Baking Quality

The baking quality requirement of flour is a function of specific baking formulas, the baking process, and the final product requirements. Accordingly, millers adjust the flour characteristics to accommodate customers' specifications. The miller and the bakers test the baking qualities of flours in the laboratory. A scoring method should be developed in the laboratory to evaluate baked bread objectively.

Color

Flour color is one of the main production challenges to the miller, because it is directly related to the shade of the final baked product. Flour color evaluation is performed by the Kent Jones and Martin Color Grader, the comparison at a certain wavelength of light or the Tristimulus system and the image analysis of bran particles. All have their advantages and limitations. However, the miller usually uses the method that is acceptable to the majority of customers. The operative miller compares the flours every hour against an accepted standard. This test is the 'Pekar

test,' in which a flattened portion of flour is wet and dried to exaggerate the color differences.

Falling Number

The falling number value indicates the level of α-amylase activity in wheat and flours. Sprouting of wheat in a wet harvest period leads to a high α-amylase activity and, consequently, a low falling number. Depending on the amount of wheat germination, it weakens the gluten, shortens dough fermentation, and increases flour acidity and ash. Upon germination, the level of α-amylase enzyme activity increases many-fold. The enzyme breaks down large starch molecules and thereby reduces the starch viscosity. Germination increases the diastatic activity and sugar production during the baking process. Accordingly, flours that lack gassing power could show an improvement with some sprouted wheat in the mill mix. Testing instruments such as the falling number and amylograph are used to measure the relative viscosity of the starch in flour and, accordingly, the level of damage to the wheat.

Fragment Count

The actual number of particles of insect origin, regardless of size, in 50 g of flour or meal is determined by using a filth test method and by a microscopic count. FDA regulations for insect and rodent filth, consider as defective 75 or more insect fragments and one or more rodent hairs per 50 g of wheat flour.

Lipids

The lipid content of wheat flour depends on the rate of extraction. The lipid content ranges from about 0.88% in 60% flour extraction to about 2.5% in whole meal product. Lipids originate during the milling process mainly from germ particles that end up in the final flour. The oxidative deterioration of lipids is the factor limiting the prolonged storage of flour. (*See* **Fats**: Classification.)

Moisture

Moisture in mill products depends on variables such as the moisture content of the raw wheat, the conditioning process before milling, ambient conditions in the mill, the design of the milling process, and mill adjustment.

In general, there is an optimal range of moisture content in the wheat before milling to achieve optimal physical attributes such as kernel hardness for the process, milling performance, and final product moisture. The miller achieves this by conditioning of the wheat, which is also the main stage where physical and chemical changes occur in the raw material and, subsequently, the characteristics of end products.

All mill product values such as ash and protein content should be corrected and expressed on a fixed moisture basis that differs in different parts of the world. For example, in the USA, flour values are corrected to a 14% moisture basis, and in others, to a dry basis. Flour is corrected to that basis because at an average relative humidity of 70% the amount of 14% moisture in flour is in equilibrium. In addition, at about 14%, no significant chemical changes or proliferation of microflora would occur in stored flour.

Flour Particle-Size Distribution

The particle size of flour produced by a commercial mill ranges between 1 and 200 μm. Flour produced from soft wheat is finer, whereas that from hard wheat is coarser. The flour particle-size distribution is directly related to water absorption, during dough preparation, and baking performance. The coarsest fraction in flour is positively related to water absorption, loaf weight, and crumb quality. Finer fractions from any given flour absorb water more slowly, produce a sticky sponge, and show an inferior baking quality. It is the miller's objective to optimize the flour particle size distribution to accommodate end-use qualities.

Whereas the flour particle-size distribution is within a known range, durum semolina is a wheat product that is produced within significantly different particle size ranges. Depending on variables in the pasta production system, including semolina wetting, extrusion, and drying, an optimum particle size distribution is required.

pH

The pH levels of flour effect their baking performance, especially those from soft wheat used for cake baking. The optimum pH levels of some flours are as follows: bleached white bread flour, 5.7–5.9; cake flour, 4.9–5.8; cookie and pie, 4.9–5.8.

Protein

Mill product protein content is directly related to that of the wheat kernel protein. Accordingly, dividing the straight-grade flour into different products would result in less protein in the combination of head streams (patent) and more protein in the combination of tail-end flour streams (clear). The protein content of flour is the amount of nitrogen determined, multiplied by a factor of 5.7. A factor of 6.25 is used for wheat bran.

The gluten that amounts to about 85% of the total protein of white flour is made up of two components, glutenin and gliadin, which are normally present in roughly equal amounts. Gliadin is a soft, sticky substance that acts as a bond for the less coherent, but harder, glutenin. Gluten is insoluble in cold water and may be recovered from a weighted flour sample. One-third of the weight of wet gluten approximates the protein content of flour. Excessive temperatures during the wheat handling and grinding (above 55 °C) would affect the gluten characteristics and, as a result, the baking quality of the flour. Accordingly, millers are sensitive to any rise in grinding roll temperature and, in some designs, can cool them with internally flowing water.

Rheological Characteristics

Instruments that record rheological properties of dough made from a flour determine its development process, rising ability, elasticity, and the quality of the baked loaf of bread. The instruments generate curves that allow the quality to be interpreted from the shape of the curves.

Starch Damage

Mechanically damaged starch granules are susceptible to enzyme action. Starch damage at a desired level of flour is related to the effective water absorption and is reflected in an increase in the maltose value that is responsible for the gassing power during the baking process. To produce acceptable bread, an optimum relation should be maintained between the flour protein content and the level of starch damage. (*See* **Starch**: Functional Properties.)

Water Absorption

The weight of water per weight of flour, expressed as a percentage, is used to make dough with an optimal consistency. Various instruments are used to measure this characteristic of flour. Variation in the flour particle size distribution, protein, and starch damage content accounts positively for changes in the water absorption of flours. Damaged starch absorbs 100% of its weight in water. Undamaged starch absorbs only 30% of its weight in water.

Effect of the Mill Design and Adjustment on Product Characteristics

The design of the mill flow has a decisive effect on product characteristics. Mill design relates to sieve-cover aperture, the characteristics of grinding stages, and the direction of intermediate material flow. Long or a short mill diagram is characterized by the specific machine allocations for the milling process. It is based on a fixed quantity (100 kg of wheat) processed in 24 h. The longer mill flow that contains a larger number of grinding stages allows a higher degree of flour extraction and a relatively lower total amount of cumulative ash content in all flours. This is caused by a relatively less severe action at each of the grinding stages.

Starch Damage Effect

Setting of the grinding rolls, pressure exerted between them, the rate of feed to the rolls, the surface of the rolls, and the available sifting area in the mill affect the level of starch damage in the flour. In general, the smooth rolls in the mill that reduce the chunks of endosperm particles to fine flour are responsible for the main starch damage. Smooth rolls differ in their surface characteristics; they can be shiny smooth or matt. Higher starch damage levels can be achieved with matt surfaces and a greater differential between the grinding rolls. Rolls are more effective in generating damage to starch in flour that impact grinding.

Blending of Flour Stream in the Mill

The professional miller is familiar with the quantity and quality of the individual mill flour streams and blends them to generate the resulted final flours. Millers use different methods to achieve desired final commercial flours.

1. One method is to combine all the flour streams while grinding a certain wheat mix and subsequently make some adjustments in quality before shipment by blending flours from different wheat mixes.
2. A second method would be to divide the flour streams into two or three collecting conveyors. Those would be the final flours specified as patent, first clear, etc. Some blending would occur in the flour handling system before shipment.
3. The third method, named divide or split milling, would be to group flour streams into 'groups' of certain quality and blend those at the end of the process at different ratios, to generate required qualities.

A flour stream grouping in a typical spring wheat mill processing 13.7% protein wheat is shown in Table 2. The estimated percentage of each of the individual flour streams and their diversity vary from one mill to another.

The three main groups of flour streams are shown in Figure 3, which also shows the groups combined to produce the commercial grades.

Product Handling

Depending on the regulations and customer requirements, mill products are treated before shipment by the addition of different additives at specified levels, moisture adjustment, air classification, or modification by heat.

Table 2 Flour streams grouping in mill grinding 13.7% protein spring wheat

Stream	*Flour (%)*	*Ash (%)*	*Protein (%[a])*
1 M	35.0	0.37	11.9
2 M	12.0	0.39	11.8
3 M	8.0	0.39	11.6
Group I	**55.0**	**0.38**	**11.8**
2 Q	4.0	0.50	12.2
4 M	5.0	0.60	12.5
1 BK	5.0	0.58	14.6
2 BK	7.0	0.54	15.4
3 BK	4.5	0.68	18.0
Break redust	6.5	0.58	14.8
Group II	**32.0**	**0.58**	**14.7**
1 T	1.0	0.75	12.5
5 M	3.0	0.60	12.8
Exhaust	2.5	0.65	14.8
4 BK	2.5	0.95	18.5
Bran duster	1.0	1.20	18.7
Group III	**10.0**	**0.77**	**15.2**
6 M	1.0	1.30	13.2
7 M	0.5	1.50	13.0
Short duster	1.5	1.40	16.5
Low Grade	**3.0**	**1.38**	**14.8**

[a]Adapted from Panter A (1988) Divide milling of Canadian spring wheat flour. *Association of Operative Millers Technical Bulletin* December: 5347–5353, with permission.

Nutritional Enrichment

Since 1940, all white wheat flour or semolina in the USA has been supplemented with vitamins and essential minerals that are natural to the wholewheat but are lost during the milling process (Table 3). Good manufacturing practices should insure that the required levels of added vitamins and minerals are maintained throughout the expected shelf-life of the food under customary conditions of distribution and storage. Prepared premixes are available that contain all the vitamins with added requirements of folic acid, vitamin A, calcium, and others.

Flour Improvers

Flour aging Following the milling process, the flour goes through an 'aging' or maturation process. Aging enhances the gas-retention properties of the gluten by oxidizing, among others, the –SH groups. In countries where additives are not permitted, millers store flours up to 12 days before shipment to allow the process to take place. Where permitted, oxidizing agents such as ascorbic acid (vitamin C), potassium bromate, azodicarbonamide, and chlorine dioxide are used to improve the bread-making qualities (Table 3).

Flour bleaching The yellowish shade of freshly milled flour is the result of the carotene present. Bleaching agents are added, where permitted, in the

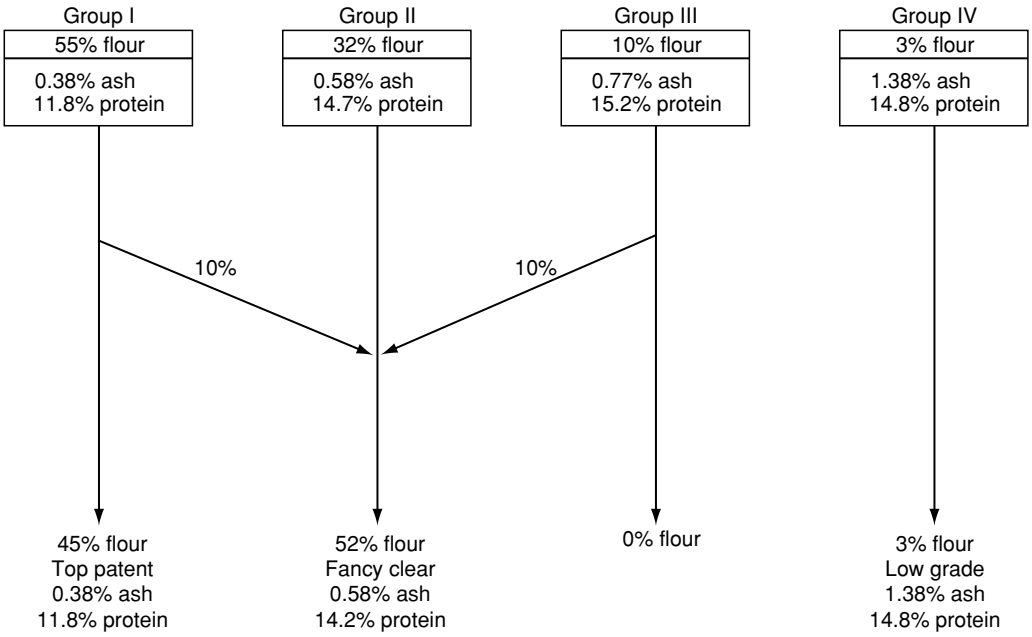

Figure 3 Schematic diagram of combining flour groups. Adapted from Panter A (1988) Divide milling of Canadian spring wheat flour. *Association of Operative Millers Technical Bulletin* December: 5347–5353, with permission.

Table 3 Additives to wheat flour

Flour additive	*Physical state*	*Rate (p.p.m.)*	*Effect*	*Maximum level (p.p.m.)*
Nutritional additives				
Thiamin	Premix	4.4–7.7	Prevents beriberi	Regulations
Riboflavin	Premix	3.7–4.8	Eye and mouth tissue health	Regulations
Niacin	Premix	35–64	Prevents pellagra	Regulations
Iron	Premix	29–43	Forms hemoglobin	Regulations
Folic acid	Premix	1–2	Prevents birth defects	Regulations
Calcium	Powder	2116	Counter phytic acid	Optional
Flour-improving additives				
Potassium bromate	Powder	10–20	Oxidation	50 p.p.m.
Azodicarbonamide	Powder	2–20	Aging	45 p.p.m.
Chlorine	Gas	1000–1400	Bleach/age	As required
Benzoyl peroxide	Powder	50	Bleach	150 p.p.m.
Ascorbic acid	Powder	70	Mix-time reduction	200 p.p.m.
L-Cysteine (hydrochloride)	Powder	30	Mix-time reduction	90 p.p.m.
Fungal α-amylase	Powder	30	Diastatic supplement	As required

mill to improve the flour color. One of the most commonly used bleaching agents is benzoylperoxide. Chlorine is another agent that is added as a gas mainly to soft wheat flour, where another objective is to control the flour pH level (Table 3).

Enzymatic activity To compensate for any natural deficiency of enzymes to convert starch to maltose, usually malted barley flour or enzymes are added. The treatment is approximately 0.25%, depending on the strength of the malt and the requirements of the flour treated.

Heat Treatment

Heat treatment of flour is used to change its physical and rheological properties. The heat reduces the elasticity and even denatures the gluten. The starch can be gelatinized and enzymatic activity reduced or eliminated. The heat also causes a reduction in the bacterial count of flour. In terms of amylograph values, a flour viscosity of about 300 Brabender units (BU) could be effected by heat treatment and reach 800 BU. The new characteristics improve the water-holding capacity of the treated flour. The heat-treated flour is used for special end uses such as soups, sausage filling material, infant foods, and other food and nonfood products.

New Developments and Their Effect on Product Characteristics

Debranning of Wheat Before Milling

The newly applied process of debranning the wheat outer layers from the kernels before milling yields products that are different from those produced by a conventional mill, in terms of quality and nutritional values. The finely adjusted debranning machine action on the wheat kernel allows the miller to remove by abrasion individually each of the seven bran layers before grinding the endosperm to flour. From the outer to inner layer, these are: cuticle, epicarp, endocarp, testa, nucellar layer, and aleurone layer. Each is significantly different in characteristics from the other. While the nutritous aleurone layer is high in protein content (28%), the epidermis, epicarp, and endocarp have a very high level of insoluble dietary fiber.

As mentioned earlier, the process of sprouting softens the wheat kernel, and accordingly, when milled on a conventional milling system, the bran is pulverized somewhat more than sound wheat, and the ash increases as a result of the fine bran particles. Wheat debranning before milling removes most of the kernel layers that contain high levels of α-amylase such that the resulting flours have a reduced enzymatic activity and are more suitable for bread-baking purposes.

See also: **Air Classification**: Uses in the Food Industry; **Bread**: Dough Mixing and Testing Operations; **Flour**: Roller Milling Operations; Analysis of Wheat Flours; Dietary Importance; **Milling**: Principles of Milling; Types of Mill and Their Uses; **Starch**: Structure, Properties, and Determination; **Wheat**: Grain Structure of Wheat and Wheat-based Products

Further Reading

Canadian International Grains Institute (1993) *Grain & Oilseeds Handling, Marketing, Processing*. 4th edn, vol. II.

Code of Federal Regulations (1991) *Whole Wheat Flour. No. 21, Part 137*. Washington, DC: Food and Drug Administration.

Code of Federal Regulations (1996) *Food and Drugs Standards of Identity. Parts 100–169, Sections 137.300 and 137.320*. Revised 1 April 1996. Washington, DC: Food and Drug Administration.

Current Good Manufacturing Practices (1986) *Federal Food and Drug Administration, Food, Drug, and Cosmetics, Act Sections 402(a)(3) and 402(a)(4)*. Washington, DC: Federal Food and Drug Administration.

Evers T and McRitchie F (2000) Ash who needs it? *World Grain* 18(6): 56–63.

Farrand EA (1972) The influence of particle size and starch damage on the characteristics of bread flours. *The Bakers Digest*, February, 22–26.

Gruber FJ (1964) Getting the most from a bread score. *American Society of Bakery Engineers Bulletin 174*. September, 702–708.

Manser J (1985) Feinheitsgrad von Durum-Mahlerzeugnissen aus der Sicht der Teigwarenindustrie. *Getreide Mehl & Brot* 39(4): 117–123.

Panter A (1988) Divide milling of Canadian spring wheat flour. *Association of Operative Millers Technical Bulletin* December: 5347–5353.

Posner ES (1991) Wheat and flour ash as a measure of millability. *Cereal Foods World* 36(8): 626–629.

Posner ES (2000) Wheat. In: Kulp K and Ponte JG, Jr. (eds) *Handbook of Cereal Science and Technology*. New York: Marcel Dekker.

Posner ES and Hibbs AN (1997) *Wheat Flour Milling*. St. Paul, MN: American Association of Cereal Chemists.

Pyler EJ (1982) *Baking Science and Technology*. Chicago, IL: Siebel.

Mincing *See* **Meat**: Sausages and Comminuted Products

MINERALS – DIETARY IMPORTANCE

J H Freeland-Graves, The University of Texas at Austin, Austin, TX, USA
P J Trotter, Augustana College, Rock Island, Illinois, USA

Introduction

Minerals are of critical importance in the diet even though they comprise only 4–6% of the human body. Major (macro-) minerals are those required in amounts greater than 100 mg day^{-1} and represent 1% or less of body weight (**Figure 1**). These include calcium, phosphorus, magnesium, sulfur, potassium, chloride, and sodium (**Table 1**). Trace (micro-) minerals are essential in much smaller amounts, less than 100 mg day^{-1}, and make up less than 0.01% of body weight. Essential trace elements are chromium, copper, fluoride, iodine, iron, manganese, silicon, and zinc (**Table 2**). Ultratrace minerals are a subcategory of trace minerals that are required in amounts less than 50 ng g^{-1} in the diets of animals. These include arsenic, boron, molybdenum, nickel, selenium, and vanadium (**Table 3**). Other inorganic elements which may contribute to biological processes, but which have not been established as essential, are barium, bromine, cadmium, lead, lithium, and tin. Nonnutritive metals such as aluminum, bismuth, gallium, gold, mercury, and silver may also be present in small amounts in foods. These metals have no known function and may contaminate wholesome food and create toxic symptoms. (*See* **Heavy Metal Toxicology**; **Trace Elements**; Refer to individual minerals.)

Food Sources

A mineral is an inorganic homogeneous substance. When food is burned, the organic portions oxidize; the ashes remaining are minerals. The ash can be analyzed to determine the specific quantity and type of minerals present. The level of some minerals, such as iodine and selenium, in foods is particularly dependent upon the amount of the mineral in the soil of the region where the food is produced. A variety of common food sources of minerals are listed in **Tables 1–3**.

The natural occurrence of minerals in food is not always desirable. In vegetable oils, miniscule amounts of iron and copper contribute to the development of rancidity. Sequestrants, such as ethylenediaminetetraacetic acid, are added to render them ineffective. Other foods are enriched by the addition of minerals. Some examples are the addition of iodine to salt, iron to flour and bread, calcium to orange juice, and iron and other minerals to breakfast cereals. (*See* **Food Fortification**.)

1 H																	2 He
3 Li	4 B											5 B	6 C	7 N	8 O	9 ? F	10 Ne
11 Na	12 Mg											13 A1	14 Si	15 P	16 S	17 Cl	18 Ar
19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe

Figure 1 A portion of the periodic table of elements. Elements thought to be essential are shaded; fluoride may be beneficial rather than essential.

Table 1 Food sources, physiological functions, deficiency symptoms, and requirement of major minerals

Mineral	*Food source*	*Physiological functions*	*Deficiency symptoms*	*Toxicity symptoms*	*Requirements*
Calcium	Milk, cheese, yogurt, turnip, greens	Bone calcification, blood clotting, muscle contraction, nerve transmission	Rickets, osteoporosis, osteomalacia, tetany	Hypercalcemia, kidney stones, constipation, depression, confusion UL 2.5 g day^{-1}[a]	1000 mg[b]
Chloride	Salt, processed foods	Electrolyte in fluid balance, gastric acidity, acid–base balance	Hypochloremic metabolic alkalosis, hypotension	Hypertension (in conjunction with excess sodium)	750 mg[c]
Magnesium	Spices, nuts, coffee, cocoa, vegetables	Cellular metabolism, muscle relaxation, nerve transmission	Nervous disorders, muscle weakness, tetany, arrhythmia, hypertension	Diarrhea, UL 350 mg nonfood magnesium per day[a]	310–420 mg[b]
Phosphorus	Cheese, meats, peanuts, soft drinks	Bone calcification, energy release, membrane structure, acid–base balance	Fatigue, anorexia, bone demineralization, muscle weakness	Metastatic calcification and skeletal porosity (in animals) UL 4 g day^{-1}[a]	700 mg[b]
Potassium	Fresh foods, molasses, milk, legumes, bananas	Electrolyte in fluid balance, nerve transmission, muscle contraction, blood pressure	Weakness, anorexia, cardiac arrhythmia, irrational behavior	Muscle weakness and paralysis, cardiac arrhythmia or arrest	2000 mg[c]
Sodium	Salt, cured meats, processed foods	Electrolyte in fluid balance, membrane potential of cells, active transport, blood pressure	Hyponatremia, nausea, anorexia, weakness, confusion, convulsions	Edema, hypertension, congestive heart failure in elderly	500 mg[c]
Sulfur	Meat, fish, eggs, cheese, legumes	Energy transfer (constituent of sulfur-containing amino acids, insulin and some vitamins)	Unknown	Depressed growth (animals)	

[a]Tolerable upper intake level (UL) recommended for adults from the Institute of Medicine (1999): see Further Reading.
[b]Daily recommendation for an adult man from the Institute of Medicine (1999): see Further Reading.
[c]Daily recommendation for an adult man from the National Research Council (1989): see Further Reading.

Table 2 Food sources, physiological functions, deficiency symptoms, and requirement of essential trace minerals

Mineral	*Food sources*	*Physiological functions*	*Deficiency symptoms*	*Toxicity symptoms*	*Requirements*[a]
Chromium	Mushrooms, yeast, prunes, nuts	Glucose metabolism, enhanced insulin sensitivity, nucleic acid stability	Glucose intolerance, neuropathy, elevated serum insulin and lipids	Cancer (hexavalent form)	35 μg
Copper	Nuts, shellfish, liver, raisins, grains, chocolate	Iron utilization, nervous system, pigmentation, immune defense, neovascularization	Neutropenia, anemia, decreased pigmentation, neurological, skeletal, cardiovascular, and cartilage abnormalities	Gastrointestinal discomfort ($>5\,mg\,day^{-1}$), liver damage, weakness, nausea	900 μg
Fluoride[b]	Fluoridated water, seafood, tea	Precipitates calcium and phosphorus in bone and teeth	Increased dental caries	Mottled tooth enamel, joint stiffness UL $10\,mg\,day^{-1}$[c]	3.0 mg[d]
Iodine	Saltwater fish, iodized salt, bakery goods	Thyroid hormones in basal metabolism	Goiter, stunted growth, mental retardation, myxedema, cretinism, hypothyroidism	Goiter, hypothyroidism	150 μg
Iron	Liver, meats, molasses, prunes, nuts	Hemoglobin and myoglobin formation for oxygen transport, cellular oxidation	Anemia, poor body temperature regulation, impaired psychomotor/ intellectual performance	Liver damage, diabetes, increased risk of heart disease due to lipid oxidation	8 mg[e]
Manganese	Tea, nuts, oatmeal, bran, pineapple	Cartilage and bone equity, brain function, lipid and carbohydrate metabolism	Rash, nervous disorders, hypocholesterolemia, skeletal/mitochondrial abnormalities	Brain or neurological disorders resembling Parkinson's disease, schizophrenia	2.3 mg
Silicon	Pectin, grains, beer, cereals	Bone calcification and cartilage formation, growth	Depressed growth and skeletal development (chick)	Kidney stones; generally nontoxic	40 mg
Zinc	Meats, shellfish, liver, legumes	Reproduction, growth, skin integrity, wound healing, taste acuity, immune response	Impaired sexual development/growth, skin lesions, hair loss, anorexia, behavioral disturbances	Poor immune response, lower high-density lipoprotein-cholesterol, copper deficiency (chronic $>0.50\,mg\,day^{-1}$)	11 mg

[a]Daily recommendations or usual dietary intakes for an adult male from the Institute of Medicine (2001) *Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc*. Washington, DC: National Academy Press.
[b]May be beneficial rather than essential.
[c]Tolerable upper intake level (UL) recommended for adults from the Institute of Medicine (1999) *Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride*. Washington, DC: National Academy Press.
[d]Daily recommendation for an adult male from the Institute of Medicine (1999) *Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride*. Washington, DC: National Academy Press.
[e]18 mg for adult female.

Bioavailability

The total amount of a mineral in a food does not necessarily reflect the amount that is available for absorption into the body. The presence of dietary fibers and mineral-binding ligands, such as phytates and oxalates, may substantially diminish the bioavailability of minerals. Phytates are abundant in bran, whole grains, and oil-seed legumes; oxalates are found in dark-green leafy vegetables, such as spinach. Diets rich in these compounds, found in strict vegetarian diets, may have the potential to affect mineral status adversely. (*See* **Bioavailability of Nutrients**; **Vegetarian Diets**.)

Polyphenolic compounds, formerly called tannins, are other dietary components that bind minerals. In tea, large concentrations of polyphenolic compounds can reduce the absorption of iron in a meal by as much as 87%. (*See* **Tannins and Polyphenols**.)

The amount of available mineral in a food may be further reduced by the large quantities of other minerals. For instance, the amount of oral manganese absorbed may be greatly decreased by the concomitant ingestion of calcium (Figure 2). Phosphates are known to diminish the absorption of zinc. Food sources that are high in fiber, phytates, oxalates, polyphenolic compounds, and other minerals should therefore be used in moderation in order to insure optimal mineral adequacy of the diet. (*See* **Dietary Fiber**: Effects of Fiber on Absorption; **Phytic Acid**: Nutritional Impact.)

In contrast, some dietary factors improve the bioavailability of minerals. Ascorbic acid, sugars, and amino acids enhance the absorption of iron and amino acids; citrate, phosphate, gluconate, and high

Table 3 Food sources, physiological functions, deficiency symptoms, and requirement of essential ultratrace minerals

Mineral	*Food sources*	*Physiological functions*	*Deficiency symptoms*	*Toxicity symptoms*	*Requirement*[a,b]
Arsenic	Fish, meat, poultry	Taurine and polyamine metabolism	Depressed growth, reproductive abnormalities, sudden death (animals), nervous disorders	Dermatosis, low hematopoiesis, sensory disturbances, skin cancer	2–29 μg
Boron	Milk and dairy products, fruits, nuts, vegetables	Energy utilization, development/ maintenance of bone	Elevated urinary calcium, aggravated arthritis (humans), poor bone development (chicks)	Nausea, anorexia, weight loss, decreased sexual activity and sperm count	1 mg
Molybdenum	Legumes, cereals, leafy green vegetables	Sulfur, pyrimidine, and purine metabolism	Low uric acid, tachycardia, tachypnea (humans), kidney stones (sheep)	Gout	45 μg
Nickel	Oatmeal, legumes, peas, nuts, chocolate	Production of hormones, membrane properties, oxidation/reduction	Low blood glucose, abnormal bone growth, poor iron absorption, altered calcium metabolism	Nasal and lung cancers, asthma, contact dermatitis, poor stress and immune response	< 100 μg
Selenium	Meats, fish, grains	Antioxidant, thyroid hormone metabolism, capillary integrity	Cardiomyopathy, liver damage, birth defects (wild fowl)	Skin lesions, brittle hair, nails, and hooves (animals) UL 400 μg	55 μg
Vanadium	Parsley, black pepper, mushroom, dill, shellfish	Iodine metabolism	Reduced growth, poor bone growth, impaired reproduction (animals), poor feathers (chicks)	Green tongue, intestinal upset, dehydration, breathing difficulty	6–18 μg

[a]Daily recommendations or usual dietary intakes for adults from the Institute of Medicine (2001) *Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc*. Washington, DC: National Academy Press.
[b]Nutrient information from the American Society of Nutritional Sciences: www.nutrition.org (2000).
UL, upper intake level.

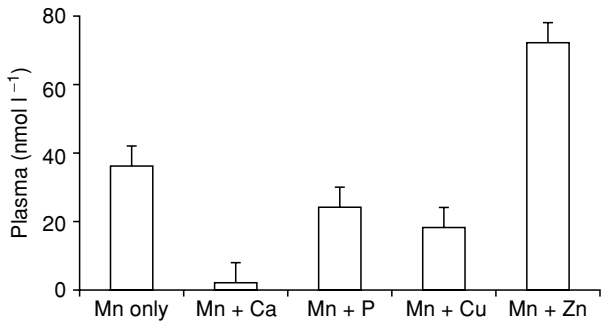

Figure 2 Plasma uptake of manganese as affected by oral loads of manganese (Mn), calcium (Ca), phosphorus (P), copper (Cu), and Zinc (Zn). Areas under the curve for response of plasma manganese when the following was administered: 40 mg Mn, 40 mg Mn plus 800 mg Ca, 40 mg Mn plus 800 mg P, 40 mg Mn plus 2 mg Cu, and 40 mg Mn plus 50 mg Zn.

dietary protein increase the absorption of copper. The milk sugar lactose increases the absorption of calcium by approximately 30%.

Requirements

The amount of a mineral required in the diet for optimal health and prevention of chronic disease is the basis for new dietary recommendations, dietary reference intakes (DRIs), that have been formulated in the USA in conjunction with the reference nutrient intakes (RNIs) from Canada. These have been set for major minerals (calcium, magnesium, and phosphorus) and trace elements (chromium, copper, fluoride, iodine, iron, manganese, molybdenum, selenium, and zinc). The DRIs consist of four standards:

1. Estimated average requirements (EAR), based on the median of a population;
2. Recommended dietary allowances (RDA), designed to meet nutrient needs of almost all (97.5%) healthy individuals;
3. Adequate intakes (AI) for minerals with insufficient data to establish an EAR;
4. Tolerable upper intake levels (UL) based on the maximum intake per day that appears to be safe for an individual over a long period of time (Figure 2).

No recommendations are made for other trace and ultratrace elements that participate in biological processes because of a lack of research.

In the UK, dietary reference values have been suggested for six minerals (calcium, phosphorus, magnesium, sodium, potassium, and chloride), and for five trace elements (iron, zinc, copper, iodine, and selenium). 'Safe' intakes have been set for molybdenum, manganese, chromium, and fluoride. (*See* **Dietary Reference Values**.)

Recommendations for minerals are more difficult to define than other nutrients as minerals can be

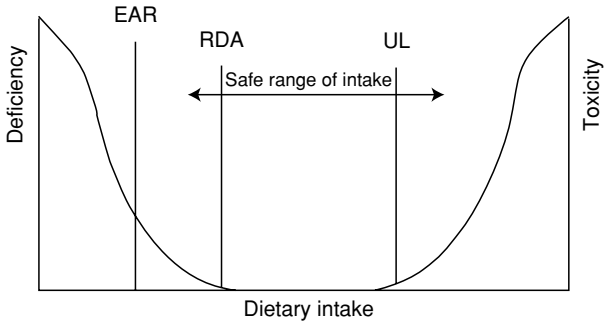

Figure 3 Dietary recommendations for minerals are located within a continuum between deficiency and toxicity. EAR, Estimated average requirements; RDA, recommended dietary allowances; UL, upper intake level.

both essential and toxic. The safe range of intake is somewhere between levels known to produce a deficiency and those that result in toxicity with continued exposure (Figure 3). Some minerals, such as zinc and copper, have a relatively large difference between the adequate and toxic doses, whereas others, such as iron and selenium, have a fairly narrow window of optimal intakes.

It should be emphasized that the recommendations mentioned above are for healthy persons as they consume their usual diets. The values may not be applicable to individuals receiving parenteral solutions, those with inborn errors that may influence absorption, transport, utilization, storage, or excretion, and during conditions of malnutrition, disease, and increased demand (pregnancy, lactation, growth, stress, injury).

Physiological Importance of Minerals

Minerals serve structural, catalytic, and regulatory functions in the body. The major minerals serve as structural components of tissues and function in cellular and basal metabolism, as well as water and acid–base balance (Table 1). Three of the major minerals – sodium, potassium and chloride – are called electrolytes. Although the functions of the trace and ultratrace minerals vary considerably, as shown in Tables 2 and 3, these minerals are primarily catalytic. These function as cofactors for enzymes or essential components of important biomolecules. For example, copper, manganese, and zinc are important components of enzymes, zinc is a component of some transcription factors (DNA-binding proteins) that regulate gene function, iron is an essential component of the heme portion of hemoglobin, iodine is an integral part of thyroid hormones, and cobalt is a necessary element in vitamin B_{12} (cobalamin). (*See* **Coenzymes**.)

It should be emphasized that the presence of a mineral in the body is not proof of essentiality. Metals that are toxic may enter the body via contaminated foods and nonfoods, inhalation of fumes, and absorption through the skin. They remain in the body since there is no mechanism for excretion. For example, strontium may be found in bones as a result of drinking contaminated milk; mercury may be deposited in the brain after inhalation of vapors; and lead may be present in the brain following ingestion of lead-containing paint.

The standard criteria of essentiality are as follows: (1) a deficiency or conditioned deficiency state is induced and prevented or cured by a mineral; (2) the mineral is required as a structural component by body tissues, regulators, or is a dietary essential; (3) biological fluids maintain a certain concentration of the mineral as a component electrolyte; and/or (4) a molecular process uses the mineral as a cofactor. For the ultratrace elements, another criteria for essentiality may be that (5) a deficiency state creates a suboptimal biological function, which can be circumvented or revoked by the addition of a usual dietary level of the mineral. (*See* **Coenzymes**.)

Minerals and their Interactions

Minerals act as cofactors in biochemical reactions since they have the ability to form stable chelates with nitrogen donor ligands in amino acids, peptides, and proteins. They can stabilize the structure and change the conformation (which may influence function), bind substrates or cofactors, activate an enzyme–substrate complex, and influence the enzyme's affinity for a substrate by the specificity of the metallic ion.

Binding to proteins can be beneficial as it limits the amount of free ions available. When some minerals (such as arsenic, chromium, copper, iron, nickel) are consumed in excess they may bind to DNA in a cross-link manner. This type of cross-link binding may create a dysfunctional or mutated DNA that may play a role in carcinogenesis (initiating cancer). Excess free ions of iron, for example, may stimulate the formation of free radicals which attack lipids or the free ions may bind to DNA in cells. Mutations in DNA may be related to the association between excess iron and colon cancer.

Some minerals, such as the cupric ion, bind strongly to proteins, and are not freed by dialysis. In contrast, zinc undergoes nonenzymatic chelation. Zinc ions are known to bind covalently with the nitrogen from the imidazole moiety of histidine and, possibly, glutamate. The binding of copper and zinc to metallothionein, a protein involved in the homeostasis of these minerals, confers resistance against the enzymatic degradation of this protein. Metallothionein chelates copper and zinc via the sulfhydryl groups (—SH) of its cysteine moieties. The affinity of this protein for copper is 105 times greater than for

zinc. (*See* **Protein**: Interactions and Reactions Involved in Food Processing.)

In calcium binding, coordination primarily involves oxygen atoms. A main-chain carbonyl (—COO—) is one ligand and each oxygen is the carboxyl group (—COOH) of glutamyl and aspartyl residues. Iron in nonheme proteins, such as transferrin, invariably binds to the sulfhydryl group of a cysteine residue or inorganic sulfur. In heme proteins, one ligand is the nitrogen of imidazole in histidine. A sulfhydryl group is a ligand for copper, zinc, cadmium, mercury, and other heavy metals. Some proteins have double metal sites (e.g., aspartate, glutamate) so that different metallic ions can have a site in common.

Mineral–Mineral Interactions

Minerals interact with each other according to their physicochemical properties, i.e., valence shell electronic structure, ionic radius, coordination number and geometric configuration, redox potential, spin transition state, and ligand exchange. Minerals with similarities in some of these properties can be predicted to develop antagonistic relationships. For example, both divalent iron and trivalent cobalt have a d^6 electronic configuration (six electrons of the *d* orbital), divalent copper (0.81) and divalent iron (0.83) have similar ionic radii, and vanadate and chromate have similar 3*d*-4*s* configurations. Thus antagonistic relationships can be predicted between iron and cobalt, copper and iron, and vanadate and chromate.

A calcium–phosphorus interaction occurs in the body when levels of serum calcium decline. Parathyroid hormone is secreted, which decreases the renal reabsorption and urinary excretion of phosphorus, resulting in lower serum phosphorus. This hormone also induces the active form of vitamin D which increases intestinal calcium absorption and renal phosphorus reabsorption, resulting in higher serum levels of calcium and phosphorus. This feedback mechanism maintains a homeostatic control of the serum levels of these minerals. During low dietary intake of calcium, zinc supplementation has an inhibitory effect. (*See* **Cholecalciferol**: Physiology.)

Magnesium also affects parathyroid hormone secretion and competes with calcium for intestinal absorption. At the cellular level, calcium transport by the sarcoplasmic reticulum is a magnesium-dependent process. When dietary magnesium is high, intestinal absorption of phosphorus is inhibited. It has been observed that both calcium and magnesium negatively influence fluoride uptake at the intestine.

Diets high in zinc have been shown to be detrimental to the status of both iron and copper. Excessive zinc results in iron-deficiency anemia via a direct effect on absorption. The opposite is true, i.e., high dietary iron reduces zinc absorption. The antagonism between zinc and copper is being exploited in the treatment of Wilson's disease, a genetic disease in which copper abnormally accumulates in the liver. The administration of 75 mg zinc day^{-1} blocks the intestinal absorption of copper and returns copper balance to normal quantities.

Iron-deficiency anemia is also produced by excessive quantities of manganese. During periods of low manganese intake, the absorption of iron greatly increases. (*See* **Anemia (Anaemia)**: Iron-deficiency Anemia.)

The significant interactions that can occur when excessive amounts of minerals are ingested suggest that mineral supplements be avoided under usual circumstances. High-dose supplements of one mineral can upset the delicate balance of other minerals in the body. Consumption of a varied, nutrient-dense diet is a better choice than supplements for achieving optimal mineral status.

Influence of Minerals on Food Processing and Food Quality

Minerals have the potential to affect the color, texture, flavor, pH, and nutritive value of foods and are often used as food additives.

Color

The brilliant colors of fruits and vegetables are due to a variety of plant pigments. Chlorophyll pigments, which contribute blue-green, yellow-green, and gray-green colors, react with both zinc and copper to produce a bright green color. The color change occurs when these minerals replace the central magnesium atom in the chlorophyll molecule. (*See* **Chlorophyl; Colorants (Colourants)**: Properties and Determination of Natural Pigments.)

Anthocyanins are red and blue pigments which turn red with acid, blue with an alkali, and are sometimes colorless with prolonged heating. When foods containing anthocyanins are processed in tin (actually tin-coated) cans, the tin must be lacquered to prevent the anthocyanins from forming greenish-blue pigments. A greenish-blue color is also created when rusted tin pans are used for fruit pies. Contact of a fruit filling, such as blueberry or raspberry, with the tin and iron salts may cause it to discolor.

Anthoxanthins (flavones) are pigments that are clear and white in acid, yellow in alkali, and pink with prolonged heating. Tin and aluminum react with these pigments and change the color to a bright yellow. This is illustrated by the bright-yellow cooking water produced when onions are cooked in an aluminum pan. When anthoxanthins react with

iron and copper, blue-black and reddish-brown colors appear, as seen in onions fried in such pans.

Sulfur prevents the darkening of foods which occurs when melanin is formed. This brown-black pigment is created via enzymatic browning in the presence of oxygen. Dried fruits, such as apricots and golden raisins, which might turn an unappetizing brown when drying, can be dipped in a sulfur solution or exposed to sulfur fumes to prevent discoloration. Pineapple juice is used as a dipping solution for cut fresh fruits, such as bananas, because its high sulfur content retards color changes. Cut lettuce for salad bars was once sprayed or dipped in a sulfite solution to retard browning, but this practice has been banned since over 100 people had allergic reactions (including one death) in response to the sulfur.

Chloride salts, such as chlorine dioxide and nitrosyl chloride, and chlorine are used to accelerate the natural aging and bleaching of flour.

Texture

When canned vegetables and fruits are cooked during processing, the tissue softens as cementing pectic substances in the cell walls of plant tissue degrade. This softening can be counteracted by the addition of calcium ions in the form of calcium hydroxide and calcium pectate. The calcium salts react with the pectic substances to form a firm material. Calcium salts are often added to canned tomatoes as a firming agent. The presence of phytates in vegetables, such as peas, decreases the firming effect of the calcium ions due to formation of a calcium–phytate complex.

In commercial baking, the texture and baking quality of bread dough are improved by the use of bromates and iodates which act as oxidizing agents. If the dough must wait for the oven, gas that is evolving may be lost before cooking and the baked goods will lose their characteristic light texture. To prevent the loss of gas, baking powders have been specially formulated to produce two reactions. In sodium aluminum sulfate–phosphate powder, monocalcium phosphate reacts first when moistened at room temperature to create a smooth, light batter; then sodium aluminum sulfate reacts when it is solubilized by hot water. In sodium acid pyrophosphate (SAPP)-baking powder, a pyrophosphate replaces the monocalcium phosphate because it has a slower reaction rate. An even slower reaction rate is seen with sodium acid aluminum phosphate (SALP) powder that contains sodium aluminum phosphate. SALP powder is used in cakes because it retains carbon dioxide until the gluten strands coagulate, thus preventing formation of tunnels. (*See* **Bread**: Chemistry of Baking.)

Mineral salts are also used as anticaking agents and flow conditioners for powdered foods that have a tendency to cake or form lumps, such as salt, confectioner's sugar, and baking powder. Some compounds used are tricalcium phosphate, silicon dioxide, calcium silicate, aluminum stearate, ferric ammonium citrate, and monocalcium phosphate.

Flavor and pH

The intensity of flavor and tartness in sherbets, carbonated beverages, and fruit drinks is enhanced by the addition of potassium citrate and phosphoric acid. The correct proportion of acidity and alkalinity is critical in controlling the correct flavor, texture, and keeping quality of several dairy products. Buffering agents such as sodium bicarbonate, calcium carbonate, hydrogen chloride, sodium citrate, sodium hydroxide, and calcium oxide may be used to control pH. (*See* **pH – Principles and Measurement**.)

Safety and Quality of Foods

Sulfur dioxide and sulfites are added to foods because of their ability to act as antioxidants, and to fermenting alcoholic beverages because it is more toxic to bacteria and molds than to yeast. Wine can have an exceptionally high sulfur concentration, as much as 200 mg per 0.5 l. This level far exceeds the acceptable daily intake of 0.7 mg kg^{-1}. In the USA, all foods containing detectable levels of 10 $\mu g\, g^{-1}$ sulfite or more are labeled. (*See* **Antioxidants**: Synthetic Antioxidants.)

Mold inhibitors, such as calcium and sodium propionate, monocalcium phosphate and sodium diacetate, are added to baked goods. These prevent ropiness in bread and increase the shelf-life. (*See* **Spoilage**: Molds in Spoilage.)

Sequestrants or chelating agents are added to foods to bind metals, such as calcium, iron, and copper. When the metals bound to the chelator are no longer in an ionized form, oxidative changes such as staleness, rancidity, and off-flavors are prevented from developing. These compounds are important in fruit juices, canned seafood, milk, and salad dressings. They are also used to clarify wine and other beverages of minerals.

The presence of minerals in foods may influence the cooking or processing time. Calcium ions, for example, have a firming effect, which prolongs the length of the cooking time. This may occur when hard water is used since it naturally contains calcium salts. If a long cooking period is desired for flavor development, as in preparing baked beans, calcium, as well as acids, can be added. In home cooking, molasses is used since it contains high concentrations of calcium as well as aconitic acid.

The use of soft water in food processing, however, may not always be desirable. In the formulation of

low-sodium foods, soft water derived from a water softener should be avoided because it may contain as much as 50 mg of sodium dl^{-1}. Sodium is also added when foods are brined, as in preparing pickles or preventing the discoloration of vegetables. When meats are processed according to kosher laws, large amounts of sodium are added as the blood is removed by salting.

See also: **Anemia (Anaemia)**: Iron-deficiency Anemia; **Antioxidants**: Synthetic Antioxidants; **Bioavailability of Nutrients**; **Cholecalciferol**: Physiology; **Coenzymes**; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Dietary Fiber**: Effects of Fiber on Absorption; **Dietary Reference Values**; **Food Fortification**; **Functional Foods**; **Heavy Metal Toxicology**; **Protein**: Interactions and Reactions Involved in Food Processing; **Spoilage**: Molds in Spoilage; **Trace Elements**; **Vegetarian Diets**

Further Reading

Department of Health (1991) *Dietary Reference Values for Food Energy and Nutrients for the United Kingdom*. Report of the Panel on Dietary Reference Values of the Committee on Medical Aspects of Food Policy. London: Her Majesty's Stationery Office.

Freeland-Graves J (1985) Mineral adequacy of vegetarian diets. *American Journal of Clinical Nutrition* 48: 859–862.

Freeland-Graves J (1999) Manganese: development of a dietary recommendation for an essential yet toxic trace element. In: Abdulla M, Bost M, Gamon S, Arnaud P and Chazot G (eds) *New Aspects of Trace Element Research*, pp. 22–27. London: Smith-Gordon.

Freeland-Graves J and Peckham G (1996) *Foundations of Food Preparation*, 6th edn. Englewood Cliffs, NJ: Prentice Hall.

Institute of Medicine (1999) *Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride*. Washington, DC: National Academy Press.

Institute of Medicine (2001) *Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc*. Washington, DC: National Academy Press.

National Research Council (1989) *Recommended Dietary Allowances*. Washington, DC: National Academy Press.

O'Dell BL and Sunde RA (eds) (1997) *Handbook of Nutritionally Essential Mineral Elements*. New York: Marcel Dekker.

Shils ME, Olson JA, Shike M and Ross AC (1999) *Modern Nutrition in Health and Disease*, 9th edn. Baltimore: Williams & Wilkins.

MINERAL WATER

Contents

Types of Mineral Water

M Abdulla, M Bost and S Gamon, Trace Element – Institute for UNESCO, Lyon, Cedex, France

Background

The term 'mineral water' is currently applied to several types of drinking water originating from underground sources that is packaged and marketed for human consumption. Mineral water may be defined as drinking water containing naturally or artifically supplied minerals, trace elements, and gases. Natural mineral water is microbiologically wholesome water, originating from an underground formation and emerging from a spring tapped at one or more natural or bore exits. Natural mineral water can be distinguished from ordinary drinking water by its natural and original state.

Formation of Mineral Water

It takes a significant number of years for mineral water to accumulate in an underground formation. When rain and snow fall on the land, over a period of many decades, it percolates through layers of soil, sand, and clay. Successive layers filter out the

impurities while adding dissolved minerals and trace elements. Some trace elements such as arsenic can be found in significant quantities in certain geographical areas of our planet, which may contaminate the water. Finally, perhaps 100 years or more after it first touched the land, this mineral water settles on bedrock deep below the surface.

Classification of Mineral Waters

The packaged (bottled/canned/in carton or foil) water market comprises two main segments: still and sparkling water. The main difference between the two categories is in the carbon dioxide content. Still waters have no carbon dioxide, are often sold as an alternative to tap water, and are used for cooking and mixing with coffee, powdered mixes, concentrated juices, soft drinks, etc. Sparkling waters contain carbon dioxide and are generally used as an alternative to soft drinks or alcoholic beverages. There are several types of packaging (bottled/canned/in cartons and polyethylene bags), depending upon the source of water. The various types of water that are commercially available at present are shown in Table 1.

Regulations

The definition of mineral waters may differ in different countries. The rules and regulations for marketing the mineral water in the country of origin and outside that country may also differ. These differences in the rules and regulations may hinder the free movement of packaged mineral water outside the boundaries of the countries producing special brands. Efforts are being made, especially in the European Community, to eliminate these differences in order to promote the free movement of the packaged natural mineral waters. The producers of the natural mineral water, however, must fulfill the basic laws and regulations concerning foodstuffs and water intended for human consumption. From a global point of view, there is a standard set by the Food and Agricultural Organization and the World Health Organization of the United Nations (FAO/WHO codex standard). In the countries of the European Communities, in response to the opinion of the European Parliament delivered in October 1995 on the approximation of the laws of the member states relating to the exploitation and marketing of natural mineral waters, certain amendments have been made. These amendments, however, do not differ from those of the FAO/WHO codex standard. The following conditions apply to natural mineral waters packaged for human consumption, particularly in affluent industrialized countries.

1. Natural mineral water, in its state at source, may not be the subject of any treatment or addition other than the addition or elimination of carbon dioxide. It must contain a certain amount of dissolved mineral salts. In many industrialized countries, this amount is around $500\,mg\,l^{-1}$.
2. Natural mineral water, in its state of source, must meet the specified microbiological requirements. It must be free from all pathological organisms. After packaging, mineral water must not have more than a specified total colony count. In many countries, especially in industrialized countries, the figure may not exceed 100 per ml at 20–22 °C in 72 h on agar or agar–gelatine mixture and 20 per ml at 37 °C in 24 h on agar. The total colony count shall be measured within 12 h following packaging, the water being maintained at 4 ± 1 °C

Table 1 Types of water that are commonly available at present

Type of water	*Characteristics*
Natural mineral water	Originates from a geologically and physically protected and approved underground source. It must have a high quality and constant composition. It must naturally contain a certain amount of dissolved mineral salts and trace elements, and should not undergo any treatment. The amount of dissolved mineral salts is at least $500\,mg\,l^{-1}$. Nothing should be added or removed other than carbon dioxide. It must be tapped from the original source and pumped to the surface through stainless steel pipes for packaging in bottles, cans, or other approved packaging materials.
Spring water	Originates from an underground formation and flows naturally to the surface of the earth. It must have a high quality and constant composition.
Well water	Originates from a hole bored, drilled, or otherwise constructed in the ground that taps the water of an aquifer.
Mineral water	Originates from a protected underground formation that has a high quality and constant composition. It should not be modified other than by adding mineral salts, aroma-producing compounds and carbon dioxide.
Other packaged water	This applies to all other types of bottled, canned, or packaged table waters. Names such as table water, carbonated table water, soda water or other accepted names belong to this category. They may be treated and prepared in the same way as commercial drinking water according to the local public health regulations. Mineral salts, essential trace elements, aroma-producing substances, and carbon dioxide may be added.

during this 12-h period. At source, these values should not normally exceed 20 ml^{-1} at 20–22 °C in 72 h and 5 ml^{-1} at 37 °C in 24 h, respectively, on the understanding that they are to be considered as guide figures and not as maximum permitted concentrations. Further, the natural mineral water may not contain any organoleptic defects.

3. All containers used for packaging natural mineral water shall be fitted with closure designed to protect against any adulteration or contamination.
4. The labeling of the containers used for the packaging of mineral water must include certain mandatory information. The name should specify the type of water such as 'natural mineral water,' 'naturally carbonated mineral water,' and 'carbonated natural mineral water.' The label should also give the place where the spring is exploited and a statement of the analytical composition giving its characteristic constituents. It is not permitted to market natural mineral water from the same spring under more than one trade description. It is also prohibited to mention both on the packaging or labels a characteristic that the water does not possess, especially with regard to treatment or cure of human illness.
5. Exploitation of a natural mineral water spring shall be subject to permission from responsible authority of the country where the water has been extracted.
6. Equipment for exploiting the water must be so installed as to avoid any possibility of contamination and to preserve the properties of the water at source.
7. When the natural mineral water at source becomes polluted and no longer meets the microbiological criterion for human consumption, all operations leading to the commercial sale should be suspended. It is the responsibility of the local authority to undertake periodic checks to control the original characteristics of the natural mineral water.
8. In developing countries, the regulations concerning the marketing of natural mineral water vary significantly from country to country. Even for the export of natural mineral water from industrialized to developing countries, the directive above may not apply.

The last report of the FAO/WHO codex standard in 1997 did not deviate significantly from the above directive.

History of Consumption

Water has always been of vital importance to man's health and well-being. Water originating or flowing from hot or cold springs with a strong odor or taste has always excited the imagination of humans. The human race has used such spring waters in a variety of ways throughout history. Mineral wells and spas are found all over the world. From time immemorial, healing powers have been attributed to water from mineral springs, wells, and spas. Even at present, water from such sources in many parts of the world, especially in developing countries, is accredited with miraculous properties. Many of the well-known health resorts in the world became famous mainly because of the mineral springs originating there. Millions of people visit such resorts every year hoping to gain relief from many chronic diseases, such as arthritis and skin diseases. Large numbers of patients have subjectively benefited from visits to such health resorts. Except for a few elements such as sulfur and fluorine, the concentrations of other minerals and trace elements in many spa waters are very low. Scientifically, there is little evidence suggesting any physiological or pharmacological action resulting from mineral or trace element contents in spas, except, in some cases, laxation.

From the Middle Ages came herbal brewing, another forerunner of modern packaged drinks. Such brews, flavored with nettles or dandelions, were made by boiling the locally available water. Such brews continued as part of the housewife's art for many decades, and in the 1820s, brewed ginger beer first became commercially available. In 1772, Joseph Priestly, a British scientist, published a method of producing artificially carbonated water. This development ultimately resulted in the commercial production of artificial mineral water during the latter part of the eighteenth century. From a public health point of view, the introduction of this artificial mineral water might have been a turning point from a bactericidal point of view. Originally, these waters were taken primarily for medical reasons prescribed by doctors for their patients. When it was possible to analyze the mineral composition of spa waters, artifically carbonated waters were fortified with the same mineral and trace element concentrations and sold as copies of the original spa water in various parts of Europe until after World War II. Later on, the use of these artificially carbonated waters became common owing to the pollution of urban water supplies. At present, such is the situation in the urban regions of most developing countries.

Populations in many European countries have consumed mineral waters from ancient times. The best-known waters come from France and Germany. Packaged waters from underground formations in Perrier, Contrexeville, Vichy, Appollinaris, and Evian are exported on a large scale. In North America,

the tradition of drinking mineral water is not as old as that in Europe. However, the packaged water industry in the USA and Canada has experienced radical growth owing to consumer awareness about the quality and taste of drinking water. In a recent survey conducted in the USA, more than 60% of consumers of packaged mineral water cited taste and harmful chemicals in the tap water as the primary reasons for buying packaged mineral water. The case is much the same with other affluent countries. The consumption of packaged mineral water has increased significantly in many European countries during the last few decades. This can be attributed to environmental, demographic, and aggressive marketing strategies. Consumers, especially in urban areas, have become increasingly concerned with the pollution of water supplies. Packaged mineral water, however, allows the consumer to control the quality, appearance, and taste of the water. Demographically, well-educated families have moved into high-income brackets with high levels of income. This factor, coupled with the modern trend in health consciousness, has positioned packaged mineral water at the forefront of the beverage category. Additionally, the prestige value of well-known brand names with connotations of quality has become a status symbol of high-income families. Finally, the aggressive marketing efforts by manufacturers of packaged mineral water have accelerated the growth in consumption of mineral water. In certain affluent countries such as Sweden, drinking of alcoholic beverages and driving is strictly controlled, and this in turn has increased the consumption of mineral waters on social occasions.

Health Implications

From a nutritional point of view, water is an essential constituent of diet. The dissolved substances in water include traces of essential and toxic elements. Since the concentration of these dissolved substances in drinking water varies from place to place, the total quantity of water consumed during a 24-h period affects the amount of these substances in the diet. In certain tropical regions of the globe, the water consumed during a 24-h period can be substantial. Although fasting on water to cure disease has diminished greatly in Western countries during the last few decades, a great deal has been written lately in the popular medical literature about the relationship between the presence of mineral elements such as calcium and magnesium in water and certain diseases in man, especially cardiovascular diseases.

Drinking water in many parts of the world is a good source of trace elements. It is also an important source with regards to the intake of certain unwanted components such as toxic metals, nitrates, and a number of organic and inorganic substances from industrial discharges and fertilizers. Any beneficial health effect that may be attributed to mineral water is likely to be due to its mineral content and freedom from pathogenic organisms. Table 2 shows the concentration of a number of trace elements that are allowed in natural mineral water according to the FAO/WHO Codex Alimentarious Commission.

Table 2 Levels of certain trace elements allowed in natural mineral water (FAO/WHO)

Element	*Concentration (mg l^{-1})*
Antimony	0.005
Arsenic (total)	0.05
Barium	1.00
Boron (as borate)	5.00
Cadmium	0.003
Chromium (total)	0.05
Copper	1.00
Cyanide	0.07
Fluoride[a]	1.00
Lead	0.01
Manganese	2.00
Mercury	0.001
Nickel	0.02
Nitrate	50.00
Nitrite	0.02
Selenium	0.05

[a]The levels of fluoride should be declared in the labeling by a general warning 'contains fluorides' when the levels are higher than 1 mg l^{-1}, and when it contains more than 2 mg l^{-1}, it should be stated that this water is not suitable for infants and children under the age of 7 years.

The essential mineral and trace element content in commercially available mineral waters in most parts of the world is so low that it will hardly influence the overall mineral and trace-element metabolism. In a limited number of studies, it has been found that mineral water from certain areas in Europe and North America contains levels higher than the permitted levels of a few trace elements, particularly that of iodine and fluorine. A few cases of osteofluorosis have been reported in individuals consuming mineral water containing higher than the permitted levels of fluorine. Such cases were observed when the fluorine levels were around 8 mg or more per liter. The calcium content may also vary considerably between different brands of still and carbonated mineral waters. Increased consumption of such mineral waters may not be ideal for individuals who are susceptible to stone formation in the urinary tract. In certain areas of the former Soviet Union, the consumption of highly mineralized water is known to be very high, and a correlation has been noted between the amount of water consumed and the concentration of crystal-forming substances in the urine of

adults and children. The same is the case with radioactive elements. An increased content of radon has been found in some of the wells in the USA and Europe. In some European countries where radon mineral water is used for treatment of diseases, an increased incidence of lung cancer has been reported in employees working in the spas.

In comparison with the amount of mineral water consumed in different parts of the world, the incidence of adverse effects described in literature, is, however, negligible. When one takes into consideration the ill health caused by waterborne diseases in many parts of the world, the introduction of mineral water that is free from pathogenic organisms can play a role in protecting societies from such diseases. At the same time, the considerable increase in mineral water sales in developed countries during the last few decades has compelled the public health authorities to set up very strict regulations concerning the quality of mineral water. (*See* **Calcium**: Physiology; **Heavy Metal Toxicology**; **Iodine**: Physiology; **Nitrosamines**; **Trace Elements**; **Water Supplies**: Water Treatment.)

Commercially available mineral waters have no calories or alcohol. These two qualities make mineral water a suitable beverage for low-calorie diets. It is not uncommon for doctors to prescribe mineral water to pregnant women and infants, especially in areas where the drinking water contains high levels of nitrates and fluorides. In recent years, attention has been focused on the health hazard arising from the ingestion of nitrates and nitrites. Nitrate, when reduced to nitrite, is a precursor of the carcinogenic nitrosamine or nitrosamide. The extensive use of fertilizers in recent years has resulted in the contamination of drinking water with nitrates in many parts of the world. The situation is much worse in most developing countries. In such areas, mineral water can be an excellent substitute for drinking water for risk groups such as pregnant women, infants, and young children. Drinking water in some parts of the world may also contain significant amounts of toxic elements, including arsenic, cadmium, mercury, and lead. Contamination of ground water with arsenic, for example, in Bangladesh and West Bengal is currently a serious health problem in Bangladesh and India. Several million people are affected by this contamination. Mineral water, however, has only negligible amounts of these elements (see Table 2). During the last few decades, a number of investigations have reported an association between the calcium–magnesium (water hardness) contents of drinking water and mortality from ischemic heart disease. Most brands of mineral water that are commercially available at present do not have a high concentration of calcium or magnesium. Since mineral water contains no alcohol, it is an ideal beverage to drink before driving. Mineral water has become very common and popular as a healthy beverage during social gatherings in affluent countries owing to the low level of alcohol (in some countries like Sweden, almost zero level) that is permitted to be consumed before driving.

In conclusion, mineral water *per se* has no medicinal property other than being free from pathogenic organisms. The mineral salts and trace elements present in many of the well-known brands of mineral water are in such low concentration as to be virtually devoid of physiological or pharmacological action. The possible exceptions are with regards to the content of fluorine, calcium, and some toxic metals and metalloids such as arsenic. Mineral waters are usually free from chlorine, and the taste is often superior to treated water.

Trends in Consumption

In recent years, sales of commercially available mineral waters have rapidly expanded, transforming mineral water from a market niche to a sizeable mainstream market segment, in many countries. The reasons for this growth have already been described. People enjoy still and carbonated waters all over the world. Sales of packaged waters in Europe and North America have risen considerably in recent years. The consumption of packaged water in some countries such as the UK and the Netherlands is somewhat modest when compared with that of the other central European countries. The demand, however, is growing steadily in most countries. The per capita consumption of packaged mineral water a decade ago in the central European region was around 80 l. The corresponding figures in the UK and USA were 23 and 28 l, respectively. Even in East European countries, the per capita consumption of mineral water was significant during the last couple of decades. In the former Yugoslavia, the per capita consumption during the 1990s was around 20 l. Although the consumption of mineral water, especially still water, is fairly common in the former Soviet Union, consumption data are not currently available. The same applies to many countries in Asia where the consumption of packaged mineral water is significant mainly owing to the poor quality of public water and the increase in tourism. Generally speaking, the per capita consumption of packaged mineral water in most affluent countries has more than doubled during the last decade. This trend is likely to increase in future years as a result of a heightening of the consumer awareness of health and increasing concern about the quality and taste of public drinking water.

See also: **Calcium**: Physiology; **Heavy Metal Toxicology**; **Iodine**: Physiology; **Nitrosamines**; **Trace Elements**; **Water Supplies**: Water Treatment

Further Reading

Abdulla M, Jagerstad M, Svensson S, Norden A and Dencker I (1979) Water, nutrition and old age. *Scandinavian Journal of Gastroenterology* 14 (Supplement 52): 203–207.

European Commission (1996) *Directive 96/70/CE of the European Parliament and of the Council of 28 October 1996 amending Council Directive 80/777/EEC on the Approximation of the Laws of the Member States Relating to the Exploitation and Marketing of Natural Mineral Waters*, O.J. No. L 229, 23.11.96, p. 26.

Golding AMB, Noah N and Stanwell-Smith R (eds) (1994) *Water and Public Health*. Nishimura, London: Smith-Gordon.

New York State Department of Health (1991) *Chapter 1, State Sanitary Code, Part 5: Drinking Water Supplies, Sub part 5–6: Bottled and Bulk Water Standard, May 1991.*

World Health Organization (1996) *Guidelines for Drinking-water Quality*, vol. 2, *Health Criteria and Other Supporting Information*, 2nd edn, Geneva: WHO.

Sources and Analysis

H Kussmaul, Institut Fresenius, Taunusstein, Germany

Background

Mineral water is a special form of groundwater, from which it differs in its composition and properties. The criteria for differentiation can lie in increased mineral content, elevated carbon dioxide content, elevated spring (source) temperature, or the presence of higher concentrations of trace substances such as iron, iodine, sulfur, fluoride, or radon. Such mineral springs have been preferred by people from ancient times, whether for drinking or bathing, because mineral waters often have, or are associated with, beneficial health properties.

There is no generally applicable definition of mineral waters. They are often bottled or canned and marketed as beverages on account of their pleasant taste and their claimed beneficial properties. Most countries, however, have regulations concerning such mineral waters that define what is required of the product.

The countries of the European Community, for example, have guidelines for the production of, and trade in, natural mineral waters that have been adopted as national law by the individual member states. Here, it is required that natural mineral water shall differ appreciably from drinking water. Mineral water is water that meets the following special criteria:

- It has its origin in underground water resources that are protected from contamination and obtained from one or more natural or artificial sources.
- It is of original purity.
- It possesses particular nutritional–physiological effects on account of its content of minerals, trace elements, or other components.
- Its composition, temperature, and other characteristics remain constant within defined limits of natural variation.

Some countries (e.g., Germany) lay down maximum permissible concentrations for toxic substances (**Table 1**). The European Commission is currently discussing uniform limits for such substances, in some cases the limits will be more stringent.

The use of the product description 'natural mineral water' for such a packaged beverage requires that compliance with the defining characteristics has been checked officially and that the product has been recognized officially as a natural mineral water.

The natural purity requirement presumes that there is no microbiological or chemical danger of anthropogenic or environmental origin.

Once a source has been officially recognized for natural mineral water, it must be exploited in such a manner that all possibility of contamination is avoided and that the characteristics of the water

Table 1 Permissible limits for natural mineral water (Germany)

Substance	*Limit (mg l^{-1})*	*Calculated as*
Arsenic	0.05/0.005[a]	As
Cadmium	0.005	Cd
Chromium, total	0.05	Cr
Mercury	0.001	Hg
Nickel	0.05	Ni
Lead	0.05	Pb
Antimony	0.01	Sb
Selenium, total	0.01	Se
Borate	30	BO_3^{3-}
Barium	1	Ba
Sodium	20[a]	Na
Nitrate	10[a]	NO_3
Nitrite	0.02[a]	NO_2
Sulfate	240[a]	SO_4
Fluoride	0.7[a]	F
Manganese	0.05[a]	Mn

[a]Declaration: Suitable for the preparation of baby food.

that justify its being a natural mineral water are maintained when it leaves the source.

It is forbidden to process the water or to add substances to it during the packaging, with the exceptions that unstable components such as iron or sulfur compounds may be removed, and carbon dioxide may be added to the water. In particular, it is forbidden to disinfect the water by any means or to add microbiological inhibitors. Any other treatment to change the microbial count of the natural mineral water is also forbidden.

In the USA, there are no uniform, special rules for mineral water. It is covered by the regulations of the bottled water standards. If the major components exceed the maximum contaminant level (MCL) for drinking water, this must be stated on the label, for instance; the following phrases are used in New York:

> This is a mineral water and should not be used as a sole source of drinking water

if bottled water products have received exemption from only the esthetic-related inorganic chemical MCL or

> This water contains levels of minerals in excess of standards for drinking water established by the New York State Commissioner of Health and, therefore, should not be used as a principal or sole source of drinking water.

if bottled water products have received exemption from the health-related MCL.

Furthermore, in the USA, all bottled water must be disinfected, which is forbidden in Europe. The standards in Florida allow, in this case, that imported water, which may not be disinfected in its land of origin, may be imported into Florida without being subjected to disinfection.

With certain exceptions, the rules for many other countries lie between these two extremes, of mineral water that differs appreciably from drinking water in the countries of the European Community and mineral water that is on the same footing as drinking water in the USA.

Sources

Natural mineral water is obtained either from free-flowing natural sources or from artificially constructed wells. Sources or wells are to be so constructed that microbiological or chemical contamination of the natural mineral water is prevented. The measures concerned involve securing the exit from free-flowing sources, and building a well-head and seal between bore and piping in the case of a well. The conveyance and transport of the water must be carried out using piping that does not contaminate the water and that is proof against corrosion by the mineral water, which is often aggressive against iron (not stainless steel) pipes.

Mineral and thermal waters often come from great depths. The temperature of the water increases with the depth from which it is drawn, the geothermal gradient normally amounts to 3 °C per 100-m depth. Water that is appreciably above average in temperature (for example, by more than 10 °C) is frequently referred to as 'thermal water.' The composition of the mineral water depends on the composition of the underground rocks through which it flows. Limestone and dolomite rocks produce calcium or calcium–magnesium–hydrogen carbonate (bicarbonate) waters; salt deposits yield sodium chloride water; and the rocks lying over salt deposits yield calcium sulfate waters. The amount of dissolved minerals is a function of the time spent underground and the solubility of the individual minerals. The solubility can be considerably increased by the influence of volcanic carbon dioxide or elevated temperatures. However, primary rocks, e.g., in Scandinavia, are very insoluble and yield very few minerals to the water obtained from them.

Since water often flows great distances through various rocky strata and remains underground for decades or centuries, the waters that are produced are frequently not pure simple water types, but mixtures of waters of various compositions that come together either underground or during extraction.

Analysis and Composition

The extent of a mineral water analysis depends on the purpose for which it is being made. Microbiological analysis and analysis of the major components and characteristic minor components are generally sufficient for the control of a mineral water. A basic investigation or an application for official recognition require a more comprehensive analysis that covers the inorganic and organic traces, as completely as possible, as well as including the microbiological status and the major components.

Sampling is simple when bottled mineral water is investigated. The product is sampled immediately after filling, or a sample is purchased from a retail outlet. However, proper sampling is essential if a representative analysis is to be obtained from the source. A range of measurements together with the variable parameters such as temperature, redox potential, oxygen content, carbon dioxide content, and radioactivity should be carried out on-site. Other constituents, e.g., iron(II) compounds, hydrogen sulfide, and cyanides, require stabilization in order to allow their determination in the laboratory without change.

Major Components

The major components of mineral waters are the cations of sodium, potassium, magnesium, and calcium, and the chloride, sulfate, and hydrogen carbonate (bicarbonate) anions. The amounts and proportions depend on the rocky strata that have been permeated. The composition of a well-known German mineral water is given as an example in Table 2. Refer to individual minerals.

The first column gives the named ion's mass concentration in milligrams per liter. The next column gives the equivalent concentration of the individual components in millimoles per liter. The sum of the equivalent concentrations of the cations must be identical to the sum of the equivalent concentrations of the anions. The last column reports the equivalent proportion of the individual ions with respect to the total number of equivalents. This provides a particularly clear indication of the characteristics of the water. In the present case, it is a sodium–magnesium–bicarbonate–chloride water. The slightly or completely nonionic undissociated components primarily consist of silicic acid, whose concentration depends on the temperature of the region from which the water is obtained, and boric acid. The sum of the mass concentrations of cations, anions, and undissociated components yields the total mineral content of the water.

In addition, it is important for the assessment of a mineral water to measure the content and composition of dissolved gases. Higher concentrations of carbon dioxide come primarily from water of volcanic origin. Groundwaters from great depths do not normally contain oxygen, as they are present in a reduced, anaerobic state. In addition to carbon dioxide, they contain nitrogen, increased concentrations of noble gases, and sometimes low-molecular-weight hydrocarbons such as methane and ethane. Apart from increased concentrations of carbon dioxide and radon, the composition of the gases in groundwaters from nearer the surface is broadly in agreement with the composition of air.

Inorganic Trace Substances

Deep waters with elevated mineral contents more frequently contain higher concentrations of trace substances than the groundwater from near the surface, which is often used to supply drinking water. However, these higher concentrations are natural and do not indicate anthropogenic contamination.

The EC guidelines are general and state only that heavy metals should be analyzed. The German mineral water regulation lays down which inorganic trace substances must be investigated. Some maximum permissible concentrations are laid down that are frequently equal to the limiting concentrations for drinking water (Table 1). (*See* **Heavy Metal Toxicology** and Refer to individual metals.)

Organic Compounds

The organic substances in mineral waters are characterized in summation by the coefficients of spectral absorbance at 254 and 436 nm, by the dissolved organic carbon, and by the amount oxidized with potassium permanganate. The extractable substances,

Table 2 Main components of 'Apollinaris' naturally sparkling mineral water

	Mass concentration (mg l^{-1})	*Equivalent concentration (meq l^{-1})*	*Equivalent percentage*
Cations			
Sodium (Na^+)	430	18.70	57.19
Potassium (K^+)	28	0.7161	2.19
Magnesium (Mg^{2+})	105	8.640	26.42
Calcium (Ca^{2+})	93	4.641	14.19
Barium (Ba^{2+})	0.031	0.0005	
Manganese (Mn^{2+})	0.002	0.0001	
Iron (Fe^+)	0.006	0.0002	
Total	656	32.70	100
Anions			
Fluoride (F^-)	0.65	0.0342	0.10
Chloride (Cl^-)	134	3.780	11.51
Nitrate (NO_3^-)	7.0	0.1129	0.34
Sulfate (SO_4^{2-})	106	2.207	6.72
Hydrogen carbonate (HCO_3^-)	1630	26.71	81.32
Total	1878	32.84	100
Undissociated substances			
Silicic acid (meta) (H_2SiO_3)	26.7		
Total	2561		

the organic nitrogen, the phenol index, and the concentration of surface-active substances, where raised levels can provide an indication of environmental pollution, are also determined. The same is also true of polycyclic aromatic hydrocarbons and phenols. Indicators of environmental pollution such as chlorinated solvents, chlorinated phenols, plant-protection agents and pesticides, high-boiling-point halogenated compounds and aromatics, plasticizers, and antioxidants are only determined with respect to the original purity. Their presence is of anthropogenic origin. In such a case, it is not possible to guarantee the original purity, which is a prerequisite for official recognition as a natural mineral water. (*See* **Contamination of Food; Pesticides and Herbicides**: Toxicology; **Phenolic Compounds; Polycyclic Aromatic Hydrocarbons**.)

Ratios of natural isotopes are frequently employed to determine the origin and age.

Microbiological Analysis

The legal regulations describe and lay down precisely the requirements and methods to be used in the microbiological investigations.

Methods of Chemical Analysis

Since the concentrations encountered are frequently higher than those occurring in drinking water, the methods used for the chemical analysis of mineral waters are often specially adapted and validated. The traditionally tested methods of water analysis still have their place, however. For example, gravimetry is the preferred method for high concentrations of alkaline earths (magnesium, calcium, strontium) and for sulfate concentrations; titrimetry is of importance for the determination of chloride and hydrogen carbonate. However, many of the modern instrumental methods of analysis must also be used. Atomic absorption spectroscopy, atomic-emission spectrometry with inductively coupled plasma, gas chromatography, gas chromatography–mass spectrometry, and high-performance liquid chromatography have established indispensable positions in the analysis of inorganic and organic trace substances. Ion chromatography finds its major application in the rapid determination of the major anions. (*See* **Chromatography**: Gas Chromatography; High-performance Liquid Chromatography; **Mass Spectrometry**: Principles and Instrumentation.)

See also: **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **Contamination of Food**; **Heavy Metal Toxicology**; **Mass Spectrometry**: Principles and Instrumentation; **Pesticides and Herbicides**: Toxicology; **Phenolic Compounds**; **Polycyclic Aromatic Hydrocarbons**

Further Reading

American Public Health Association (1998) *Standard Methods for the Examination of Water and Wastewater*, 20th edn. Washington, DC: American Public Health Association.

Association of Official Analytical Chemists (2002) *Official Methods of Analysis of the AOAC*.

EEC (1980) *Council Directive of 15 July 1980 on the Approximation of the Laws of the Member States Relating to the Exploitation and Marketing of Natural Mineral Waters* 80/777 *EEC. J.E.C. No. L 229/1*, Amended October *1996 (96/70/EC) J.E.C. L 330*.

Fachgruppe Wasserchemie in der GdCH (2002) *Deutsche Einheitsverfahren zur Wasser-, Abwasser- und Schlammuntersuchung*. Wiley: VCH Weinheim, no. 53.

Federal Register (USA) (1995) Part III Dept Health and Human Services: 21 CFR Part 103. Beverages Bottled Water; Final Rule, Nov. 13.

Fresenius W, Quentin KE and Schneider W (1988) *Water Analysis*. New York: Springer.

World Health Organization (1993) *Guidelines for Drinking water Quality*, vol. 1, 2nd edn. Geneva: WHO.

Bottling and Storage

H Kussmaul, Institut Fresenius, Taunusstein, Germany

Introduction

As in the previous article concerning the definition of mineral waters, when discussing the processing of mineral waters from their extraction at the source to their filling into bottles or other types of packaging, it is necessary to distinguish between the requirements placed specifically on natural mineral waters by the European Community (EC) guidelines and international Codex Alimentarius Standard 108 and those for the filling of drinking water (bottled water).

The EC guidelines and the Codex Standard require that the extraction, storage, and filling be carried out in such a manner that impurities are avoided and that the properties the water possesses as it leaves the source, and which justify its character as a natural mineral water, are preserved. In particular, it is necessary to insure that:

1. The source and the exit from the source are protected from danger of contamination.
2. Attachments, piping and containers are made of materials suitable for mineral water and are so constructed that they do not cause any detrimental chemical, physical, or microbiological changes to the water.
3. The technical facilities, in particular the bottle-filling and cleansing plant, meet the hygiene requirements.
4. The containers are so treated or manufactured that they do not change the microbiological or chemical characteristics of the water.

Mineral Water Extraction and Storage

In accordance with the EC guidelines on the exploitation of mineral water sources, mineral water is mostly extracted through a stainless-steel riser by means of a submerged pump and transported to the filling facility in suitable piping (hard polyvinyl chloride (PVC) or high-density polyethylene (HDPE)). The piping must be constructed so that microbiological and chemical contamination cannot reach the water.

In some regions of the EC, e.g., Italy, the water may not be pumped but must travel to the filling plant and the packing line under natural pressure.

There are often storage facilities to form a buffer supply. This means that the water can be pumped continuously from the source at a moderate rate even though the filling facility is only operated on working days for one, two, or three shifts.

In order to avoid deleterious effects on the water as it passes through the piping or stands in the storage tank, the surfaces must be smooth and no growth of microorganisms should occur, otherwise reduction or oxidation processes might take place in the water. Ammonia could be oxidized to undesirable nitrite, or nitrate reduced to nitrite, depending on the conditions. Hence, the piping and storage vessels must be cleaned regularly. This can be done by means of chemical, thermal, or mechanical cleansing and disinfection measures. (*See* **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant; Overall Approach.)

Water Preparation

In the USA all the water preparation measures applicable to drinking water may be used that are thought necessary. The standard for the state of New York reads:

> All bottled water facilities packaging water for distribution in New York State must provide satisfactory treatment of each water supply source used. Minimum treatment of each water supply source used shall be disinfection by chlorination, ozonation, ultraviolet radiation, or other disinfection methods as protective of the public health as the above.

Since drinking water in the USA is normally heavily chlorinated, ozone is frequently used to disinfect bottled water in order to avoid a chlorine flavor.

In the countries of the EC, on the other hand, mineral water may not be subjected to any processing at all; it is merely permitted to remove from natural mineral water unstable components which could alter the properties of the water as it is bottled or while in the bottle. These are primarily iron and sulfur compounds.

Iron often occurs in mineral waters that originate at great depth in reduced bivalent form. This is unstable in the presence of air and is oxidized to trivalent iron by oxygen in the air, causing the formation of yellowish or, in the presence of manganese, brownish floccules in the water. This is unattractive to the consumer. For this reason it is permissible to remove this naturally occurring iron from the water by oxidation and filtration.

Reduced mineral waters also often contain reduced sulfur compounds, usually hydrogen sulfide. As is well known, in such cases the water has a smell reminiscent of rotten eggs. This smell would also be unpleasant in the bottles and it is thus necessary to remove these naturally occurring sulfur compounds by aeration or oxidation, followed, if necessary, by filtration. In EC countries it is necessary to declare on the label that iron or sulfur has been removed.

Carbon dioxide may be removed from heavily carbonated waters if this is necessary for filling. There is also the possibility, frequently made use of in some countries, of adding carbon dioxide to the water, thus endowing it with a refreshing, tickling flavor.

Other treatment processes, such as the addition or removal of mineral substances, are forbidden for natural mineral water in the EC. The disinfection of the water is explicitly forbidden. Rather, the microbiological state of the water must always be satisfactory at the source, and deterioration must be avoided before the water is filled.

Filling and Packaging

In the countries of the EC, natural mineral water that is to be marketed must be filled at the source. Filling is carried out on high-speed filling machines, generally into bottles, which can be made of glass or plastic. Only limited quantities are filled into cans (aluminum

or tinned steel) and soft packagings (tetrapacks, combiblocks, etc.).

Disposable glass bottles are often used, which are purchased from glassmakers and are intended to be recycled to the glass works in a voluntary manner. Mineral water is mostly filled into reusable glass bottles in Germany. These are standard returnable bottles that can be refilled c. 40–70 times.

The plastic bottles that are usually or exclusively (e.g., in Arab countries) used in some countries, are primarily made from polyvinylchlorides (PVC), polyethylenes (PE), or polyethyleneterephthalats (PET); they are either bought-in ready for use or blown from blanks (preforms) on site. The plastic bottles in use at the present time are disposable bottles, while returnable PET plastic bottles are now in use for nonalcoholic beverages; these are still under test in the case of mineral water. However, for environmental reasons it will be impossible to avoid their use in so far as they allow unrestricted production of mineral water.

Complicated, multistage cleansing machines that supply acceptably cleaned bottles for filling are required for reusable glass or plastic bottles. The acceptable condition of the bottles must be monitored by an expensive bottle-checking system. The control of reusable plastic bottles has been found to be particularly expensive since it must also be insured that substances that may have been adsorbed on to the walls are completely removed. Expensive gas-sensitive detectors are necessary for this purpose.

The package used for filling of natural mineral water must be fitted with a cap that is suitable for preventing adulteration or contamination. This serves to protect the consumer, who can then establish whether the container is in its original state or whether it has been tampered with.

Labeling

For the labeling of all bottled waters in the USA, each label shall indicate:

- the type of source water: spring water, well water, public water supply, etc. (for distilled water, the type of source water does not need to be indicated)
- address and location of the bottling facility or corporate offices
- net contents and/or capacity of the container
- the certificate number

There can be individual differences from state to state. Mineral waters containing high concentrations of mineral substances must, if necessary, be labeled to point out that the drinking water standards are exceeded for certain substances.

In the countries of the EC there are special labeling rules for natural mineral water. The marketing name is always 'natural mineral water,' with a possible indication concerning the carbon dioxide content:

- 'naturally carbonated natural mineral water' means water whose carbon dioxide content from the spring after decanting, if any, and bottling is the same as at source, taking into account where appropriate the reintroduction of a quantity of carbon dioxide from the same water table or deposit equivalent to that released in the course of those operations and subject to the usual technical tolerances;
- 'natural mineral water fortified with gas from the spring' means water whose carbon dioxide content from the water table or deposit after decanting, if any, and bottling is greater than that established at source;
- 'carbonated natural mineral water' means water to which carbon dioxide has been added from a

Table 1 Labeling indications relating to the composition of mineral waters (EC)

Indications	*Stipulation*
Low mineral content	The inorganic constituents, calculated as dry residue, shall not be above 500 mg l^{-1}
Very low mineral content	The inorganic constituents, calculated as dry residue, shall not be above 50 mg l^{-1}
Rich in mineral salts	The inorganic constituents, calculated as dry residue, shall be above 1500 mg l^{-1}
Contains bicarbonate	The bicarbonate content shall be above 600 mg l^{-1}
Contains sulfate	The sulfate content shall be above 200 mg l^{-1}
Contains calcium	The calcium content shall be above 150 mg l^{-1}
Contains magnesium	The magnesium content shall be above 50 mg l^{-1}
Contains fluoride	The fluoride content shall be above 1 mg l^{-1}
Contains iron	The bivalent iron content shall be above 1 mg l^{-1}
Acidic	The free carbon dioxide content shall be above 250 mg l^{-1}
Contains sodium	The sodium content shall be above 200 mg l^{-1}
Suitable for the preparation of infant food	Different regulations apply in the various member states
Suitable for a sodium-reduced diet	The sodium content shall not be above 20 mg l^{-1}
May be laxative	–
May be diuretic	–

source other than the water table or deposit from which the water comes.

The label must also state the site and name of the source, the composition of the water, details of the mode of treatment (not every country), and minimum storage life. If the label provides information concerning particular components or special suitabilities, the requirements listed in Table 1 apply. For instance, statements concerning the diuretic or laxative properties of the water are not allowed in some EC countries, e.g., Germany, since such indications are reserved for medicinal mineral waters.

In general the various member states lay down their own regulations for labeling natural mineral waters that go beyond the EC requirements, and which cannot be reported in detail here. Sometimes the labels must be submitted for certification or authorization.

See also: **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant; Overall Approach; **Mineral Water**: Sources and Analysis

Further Reading

EEC (1980) Council Directive of 15 July 1980 on the Approximation of the Laws of the Member States relating to the Exploitation and Marketing of Natural Mineral Waters 80/777/EEC. J.E.C. No. L 229/1, amended 28. October 1996 (96/70/EC) J.E.C. L 330.

FAO/WHO Codex Alimentarius Committee (2001) Codex Standard for Natural Mineral Waters 108-1981, rev. 1-1997, amended 2001. Rome, Italy. (www.@codexalimentarius.net; http://www.@codexalimentarius.net.)

Kussmaul H (1988) Enteisenung von natürlichem Mineralwasser durch Belüftung. *Der Mineralbrunnen* 38: 4–7.

Statutory Instruments 1985, no. 71 (1985) *Food: Natural Mineral Waters Regulations 1985 of Great Britain*.

New York State Department of Health (1991) *State Sanitary Code*, Part 5: Drinking Water Supplies, subpart 56: Bottled and Bulk Water Standards.

MINIMALLY PROCESSED FOODS

T Ohlsson, SIK–The Swedish Institute for Food and Biotechnology, Gothenburg, Sweden

Introduction

Minimal processing of foods covers a wide range of technologies and methods of preserving food during their transport from the agricultural production to the consumer. Minimal processing methods have in common that the processing changes the inherent fresh-like quality attributes of the food as little as possible (minimally), while the processing gives the food product sufficient shelf-life for transport from processing to consumers. Sometimes the term 'invisible (to the consumers) processing' is used for some of these methods.

Consumer Demand for Minimal Processing

Diet and Health

Diet and health are firmly established as long-term lifestyle, where healthy food is one of the important components of well-being.

Convenience and Simplicity

Convenience and simplicity are still very important, and must also now include the increasing health aspects.

Uncertainty

Uncertainty and the need for guarantees are signs of the increasing concerns for the impact of environmental pollution. There is a feeling that few foods are safe, and this results in reluctance to trust the traditional food supply and especially 'new' foods.

Food is Part of the Good Life

Interest in food and cooking is rising, with an emphasis on natural ingredients and cooking from scratch, including ethnic foods.

Market Trends

The trends in consumer values are strongly influencing the trends on the food markets. 'Low' and 'light' products are a major trend, oriented at lowering the caloric content of the food, often combined with lowering the content of ingredients such as fat, sugar, and salt. Additives (with E numbers) are removed from the foods and sometimes replaced by

natural ingredients with similar properties, if such are available. This trend goes hand in hand with the push towards more natural, more fresh-like products, evident from reviews of new product introductions on the UK market. The most active market segment is the chilled, ready-to-eat or ready-to-heat foods; typically in single-serving packages, designed to meet the needs of convenience and simplicity. But many of these products have short shelf-lives, typically a few days only, limiting the geographical area where they can be marketed. Technologies that allow for a two- or threefold prolongation of the shelf-life are much sought after, and increasingly applied.

It is not only the extended time of distribution and storage that is attractive. Added product safety is much in demand by producers and distributors.

The fresh-like products are highly perishable, and actions that add to the safety factors are important, e.g., in the light of the new legislation on product liability, and due diligence.

On the other side of the shelf-life spectrum, traditional long-life ambient storage products are also affected by the consumer trends. Long shelf-life is no longer an important selling point, with market trends towards more fresh-like products. Therefore, preservation technologies that prolong shelf-life but do not act detrimentally on the product quality attributes are in favor on the market. The push towards shorter shelf-lives for ambient shelf-life products is also supported by the demand for more rapid turnover of stock. This is due to the need to decrease capital costs and to the more rapid market changes pushed by improved communication between market and production, through electronic data identification (EDI) and computerized checkout counters.

Finally, an important packaging trend is to downsize material thickness and choose 'mono' material packaging. With the resulting lower barriers to oxygen of the packages, the product shelf-life is reduced.

In conclusion, products with a very short shelf-life will need preservation methods that will prolong the shelf-life, while long-shelf-life products need methods that give a shorter shelf-life but with improved quality. These methods are to be found in the group of technologies and methods called minimal processing.

Minimal Processing Technologies

In minimally processed foods, storage, processing, packaging, and distribution are accomplished in highly integrated systems. Each step must be considered in conjunction with the other steps. Different minimal processing methods can be applied at the various steps in the food distribution chain, from storage of agricultural products to packaging/processing of the ready-to-eat product. In this introduction and overview, different minimal processing technologies and methods will be presented following the food distribution chain. Detailed accounts of these methods can be found in the literature. A list of the methods reviewed here is given in Table 1.

Modified Atmosphere Storage

By modifying the composition and sometimes the overall pressure of the atmosphere in the storage environment, quality-and safety-degrading biological reactions can be slowed down or inhibited. For storage of fruits and vegetables, increased carbon dioxide concentration (up to 10%) and reduced oxygen concentration (3–5%) will retard respiration and prolong the shelf-life. Relative humidity is also an important factor, as is control of ethylene emission from the respiring fruits and vegetables.

A reduced pressure during storage is also used, often in combination with modified CO_2 and O_2 concentration levels. This method is often called hypobaric storage, using pressures of some hundreds of mmHg.

Postharvest Treatments

In order to add convenience to vegetables and other agricultural products, centralized cleaning, peeling, and cutting is common. The resulting products are often less stable after the treatment, due to enzymatic activity of cut cell walls and bacteriological contamination from the handling. Various postharvest treatment methods are employed to add biological stability and extend shelf-life. Chlorinated cleaning water is used. Soaking in solutions of reducing agents such as ascorbic acid or sulfite or preservatives such as sorbate or benzoate are used. Also, divalent ions, Ca^{2+}, are used to strengthen the texture. In all these treatments low temperature and good processing hygiene are essential to achieve the desired shelf-life, as treatments such as cutting instead reduce the shelf-life.

Clean-Room Technologies

The objective in clean-room technology is to eliminate microbiological contamination from humans or from the environment. The handling of the food is automated as much as possible. Equipment and the processing environment are presterilized before production. Air curtain and a positive air pressure are maintained in the processing line using sterile-filtered air. Production personnel use extensively protective clothing to reduce air- and human-borne microorganisms. Clean-room technologies are primarily

Table 1 Applications of minimal processing methods

Process	*Applications*	*Mechanisms*
Controlled-atmosphere storage	Bulk-stored fresh fruit and vegetables	Antimicrobial effect (inhibition of aerobic and anaerobic organisms); altered respiration rates
Postharvest treatments	Fresh vegetables	Antimicrobial effect
Chlorinated water soaking		
Reducing agents		Oxidation prevention
Preservatives		Antimicrobial effect
Divalent ions		Improved texture
Clean-room technologies	Fresh meat and fish	Reduced levels of environmental microorganisms
Protective microbes (lactic acid bacteria)	Dairy products; sausages	Release by the protective microbes of bacteriocins that reduce pH and microorganisms
Nonthermal processing methods		
High-pressure treatment	Many products, currently especially fruit products	Microorganisms ruptured under high pressure
Gamma irradiation	Many products, particularly fresh fruits, poultry, and spices	Ability of microorganisms to reproduce eliminated
High-electric-field pulses	Many products, particularly fruits	Microbial cell rupture due to uneven distribution of electrical charge across cell
Thermosonication	Drinks	Synergistic effect of ultrasound and heat
New thermal processing methods	Many products, especially finished meals	Optimized heating regime reduces levels of microorganisms while minimizing thermally induced quality losses (e.g., impaired flavor)
Ohmic heating		
High-frequency heating		
Microwave heating		
Sous-vide technology		
New packaging technologies		
Modified-atmosphere packaging and active packaging	Fresh meat and fish, prepared foods, baked goods, fresh fruit, and vegetables	Antimicrobial effect (inhibition of aerobic or anaerobic microorganisms); altered respiration rates in fruits and vegetables
Edible films	Dry, frozen, and semimoist foods	Protection against oxygen ingress, moisture loss, and flavor loss

developed for fresh prepared food and for dairy products. Clean-room technology is often an expensive method, which often means that the application is limited to high-value-added products and to a limited part of the production line. Also the hygiene classes used in the food industry are higher than in other industries using clean-room technology.

Protective Microbiological Treatment

It is well known that many microorganisms produce antimicrobial agents. Some lactic acid-producing bacteria produce bacteriocins that are efficient in stopping the growth of, in particular, Gram-positive spoilage bacteria. By adding selected lactic acid-producing strains to the surface of foods, controlled growth can create an antimicrobial condition at the food surface. The application of these methods is still in development. (*See* **Microbiology**: Classification of Microorganisms.)

Combination Methods: Hurdle Technology

Reducing the levels of salt, sugar, or acid in foods in order to improve consumers' acceptance often means increased perishability of the food. To attain sufficient shelf-life, a combination of preservation methods is often applied. The 'hurdle concept' is a simplified illustration of the principle for combination processing. By fine-tuning a preservation system with a number of methods, more knowledge is built into the product – an important step in product development.

Nonthermal Processing

High-pressure By applying pressures in the range of some thousand atmospheres on biological material, some enzymes and vegetative microorganisms can be inactivated. Cell membranes are broken. Very high pressures in combination with elevated temperatures are needed for the inactivation of bacterial spores. High-pressure treatment also changes the texture of food, e.g., coagulate proteins and swell starches. The method is commercialized in Japan and Europe for treatment to prolong the shelf-life of low-pH fruit products. In Spain, ham products are treated, with a resulting prolonged shelf-life. In the USA, avocado paste is treated at very high pressure. Extensive

research and development is now taking place to investigate the possibilities and limitations of the method.

Irradiation Very extensive research into the method in the 1950s and 1960s, particularly the wholesomeness of the irradiated products, demonstrated that the method is an efficient and safe preservation method. In spite of this, the commercial use has been very limited, due to consumer scepticism coupled with legislative limitation. Use of the method is gradually increasing in Europe and the USA, with applications for fresh fruits, poultry, and spices. The method is expensive, with typical processing costs of 0.1–0.2 Euro kg^{-1} product. Of course, this limits the application to high-value products. (*See* **Irradiation of Foods**: Applications.)

High-electric-field pulses When cells are subjected to electrical pulses with field strength of 15–35 $kV\,cm^{-1}$, cell membranes are broken due to an uneven distribution of electrical changes on both sides of the cell membranes. The broken cell walls cause inactivation of microorganisms. The method is more efficient the larger the cell, e.g., for yeast. Thus, the most interesting future application is for products where yeast growth is limiting the shelf-life, e.g., fruit products and other drinks. A number of active research programs are found in Europe and the USA, but the method is not commercial yet. (*See* **Heat Treatment**: Electrical Process Heating.)

Thermosonication By combining ultrasound treatment at 20–40 kHz with heat treatment at moderate temperatures, the inactivation of microorganisms can be strongly enhanced. It will thus enable pasteurization of drinks at lower temperatures with less thermally induced quality changes. Thermosonication is also efficient for enhancing enzyme inactivation. The method is not used commercially in the food industry. (*See* **Preservation of Food**.)

Thermal Processing

Mild heating methods There is much interest in mild heat treatment methods, which avoid excessive temperatures, resulting in thermally induced quality losses. Often a reduction of only a few degrees can have a dramatic influence on liquid losses of meat or fish. Yet the heat treatment needs to give microbiological safety to the product. With the help of modern process optimization and control methods, mild heat treatments that combine these objectives have been developed. Among these, most interest is found in direct heating methods such as microwave heating, that can be used to raise the temperature quickly and shorten the processing time. For industrial applications, the microwave equipment is designed by computerized methods to control the heating uniformity, by controlling the overheating of the edges and corners of the foods. The direct electric heating methods are typically two to five times more expensive than traditional heating methods. Thus their use is limited to processes with benefits in terms of better production yield or product quality.

Sous-vide cooking In sous-vide cooking, fresh food is vacuum (sous-vide)-packed, under hygienic conditions. The packed products are cooked at fairly low temperatures in water to internal temperatures determined by culinary objectives. The long cooking times give some tenderization effects. Excessive temperatures are avoided, allowing for high moisture retention and juiciness. As the microbiological safety of the sous-vide cooking can give rise to some concerns, depending on the processing temperature, the shelf-life is often limited to 6–20 days at temperatures of $+3\,^{\circ}C$.

Packaging

Modified-atmosphere The methods described above for controlled-atmosphere storage are also applied for individually packed food products. Fresh meat and fish, prepared foods, and baked foods are packed in modified atmosphere, with high concentrations of CO_2. Permeability of the packaging material to CO_2 is important to control.

Carbon dioxide is bacteriostatic and fungistatic. Increased CO_2 concentration will thus inhibit the growth of microorganisms as long as a sufficient concentration of dissolved CO_2 is maintained in the surface of the food. Refrigerated storage is required for CO_2 to be effective. A wide range of fresh and prepared foods are distributed and stored in the atmosphere with high CO_2 concentration (50–10%). Often O_2 concentration is reduced or essentially nil, except for fresh meat, where it is held at 20% oxygen in order to maintain the red meat color.

For vegetables, where the respiration is continuing and the CO_2 concentration would increase to levels that adversely affect the quality, microperforations in the packaging film allow the respired CO_2 gradually to escape from the package. (*See* **Packaging**: Packaging of Solids.)

Active packaging This term covers packaging methods and agents that actively influence the shelf-life of the food during storage. The best-known example is the oxygen absorbers or scavengers, which reduce the head space and permeating oxygen levels. The scavengers come as small sachets or tablets to be introduced into the package. It is predominantly

Fe^{2+} ions that are used. The reduced level of oxygen in the head space prevents development of rancidity, and is also effective in reducing growth of certain types of microorganisms, e.g., molds.

Other types of active packaging systems are ethanol vapor generators. The ethanol absorbed on silicon dioxide powder and contained in paper sachets prevents the growth of molds. They are mainly used for bakery products.

Edible Coatings

The rapid development of biodegradable films for food packaging has helped to strengthen the development of edible coating applied directly on food. Coatings are made from films of proteins, starches, or waxes. The coating will protect against oxygen, aroma components, and moisture to the product, reducing the requirements on packaging. Most films are sensitive to moisture, which limits their application to dry, frozen, and semimoist foods.

Minimal Processing in the Future

The minimal processing technologies represent a means of meeting the well-established, long-term trends in consumer demands for convenience, variety, and fresh-like quality. Such technology will allow the food industry the possibility of producing high-quality, high-value-added products to meet future consumer demands. The technologies require varying degrees of capital investment. But, most importantly, the technologies require investment into product and process know-how, not only of the minimal processing technology itself, but also of the integrated chain of food distribution, from agricultural production to the consumer.

An important future area is the understanding of the antimicrobial effects of enzymes and other biochemical agents. In addition, nonthermal processing methods need to be further investigated in terms of the mechanisms of preservation as well as engineering aspects of the industrialization of the methods.

Finally, future research into shelf-life treatments must include integrated approaches involving many different aspects along the production chain, such as hygienic design, and automated handling of products and packaging.

See also: **Convenience Foods**; **Cooking**: Domestic Use of Microwave Ovens; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; Electrical Process Heating; **Irradiation of Foods**: Basic Principles; Applications; Processing Technology; Legal and Consumer Aspects; **Packaging**: Packaging of Liquids; Packaging of Solids; **Storage Stability**: Parameters Affecting Storage Stability

Further Reading

Ahvenainen R, Mattila-Sandholm T and Ohlsson T (eds) (1994) *Minimal Processing*. SIK and VTT Symposium in Majvik, Kirkkonummi, Finland, April 14–15, VTT Symposium Series no. 142. VII Information Service, PO Box 2000, FI-02044 VII, Finland.

Ahvenainen R, Mattila-Sandholm T and Ohlsson T (eds) (1995) *New Shelf Life Technologies and Safety Assessment*. SIK and VTT Symposium in Helsinki, Finland, May 18–19. VTT Symposium Series no. 148. VII Information Service, PO Box 2000, FI-02044 VII, Finland.

Francis GA, Thomas C and O'Beirne D (1999) The microbiological safety of minimally processed vegetables. *International Journal of Food Science and Technology* 34(1): 1–22.

Leistner L and Gorris LGM (1995) Food preservation by Hurdle technology. *Trends in Food Science and Technology* 6(2): 41–46.

Martens T and Luchetti A (1999) *Harmonization of Safety Criteria for Minimally Processed Foods. Rational and Harmonization Report*. FAIR Concerted Action CT 96–1020. Copies can be obtained from www.europe.eu.int.

Minimal Processing of Food: A Challenge to Quality and Safety. EFFoST conference in Cologne, Germany, November 6–9, 1997. Wageningen, The Netherlands: EFFoST, c/o ATO.

Ohlsson T, Ahvenainen R and Mattila-Sandholm (eds) (1996) *Minimal Processing and Ready Made Foods*. Symposium in Göteborg, Sweden. April 18–19. SIK Document 114. SIK, PO Box 5401, S-40229, Gothenburg, Sweeden.

Oliveira FAR (ed.) (1995, 1996 and 1997) *Process optimisation and minimal processing of foods*. EU Copernicus project. Proceedings of first, second and third meeting. Copies can be obtained from www.europe.eu.int.

Singh RP and Oliveira FAR (1994) *Minimal Processing of Foods and Process Optimization*. Boca Raton, FI: CRC Press.

Wiley RC (ed.) (1994) *Minimally Processed Refrigerated Fruits and Vegetables*. London: Chapman and Hall.

Mites *See* **Insect Pests**: Insects and Related Pests; Problems Caused by Insects and Mites

MIXING OF POWDERS

N Harnby, University of Bradford, Bradford, West Yorkshire, UK

Introduction

The mixing of dry or near-dry powders is a very common operation in the food industry. Powder mixing may be required prior to dispersing the powder into a dough, paste, or dispersion form, but an increasing number of products are now packaged and marketed directly to the customer in the dry form. In either case a poor understanding of the mixing process can result in a failure to mix consistently and effectively.

Many of the problems associated with powder mixing arise because it is assumed that powders behave in a similar way to the more familiar liquid and gaseous mixing processes. They do not! The major differences are that, in comparison to liquid and gaseous systems, the powder system has poor mobility and large primary particles. The poor mobility or flow leads to processing and packaging problems, while large particles lead to poor mixture quality. (*See* **Agitation**.)

The Influence of Powder Flow on the Mixing Process

Powder Flow Types

For all process fluids a high mobility is generally a desirable property. This speeds up the rate-controlling processes and facilitates both process control and the product-packaging process.

Visual examination of a variety of powders leads to a broad classification into free-flowing and cohesive categories. Free-flowing powders, such as granulated sugar, exhibit a smooth flow, an attractive nondusty appearance, and little adhesion to the container. Cohesive powders, such as flour, have an erratic stick–slip flow, are often very dusty, and stick to the walls of the container. Particle size is evidently an important determinant of flow type. If the mean particle size of a powder is greater than about 50 μm, it will tend to be free-flowing, whilst below that size it will tend to be cohesive. This boundary between flow types is dependent on more characteristics than particle size, but it is a very significant boundary in that it determines the philosophy of processing the powder and the type of mixer to be used.

Particularly for the large-tonnage industries, the attractions of a free-flowing powder from both a processing and a marketing point of view tend to outweigh the quality advantages of the cohesive powder. The attractions of free flow are often so strong that significant process costs will be incurred in order to aggregate an otherwise cohesive powder to achieve a free-flowing characteristic. Only when product quality becomes dominant does the finer texture of the cohesive powder become attractive. This is often the case for high-value-added applications, such as in the pharmaceutical and ceramics industries.

Free-flowing Powders

A single particle is the ultimate mobile element of a powder mixture and, in contrast to the uniform molecular nature of the gaseous and liquid systems, each of these elemental particles will have unique properties of size, shape, roughness, density, porosity, etc.

It is the uniqueness of individual particles that gives rise to the extremely dangerous process characteristic of segregation. If particles are subject to motion they will not randomize as in a molecular system, but will instead take preferred or segregating paths depending on their individual characteristics. This can lead to a gross loss of quality. If the constituent powder particles are chemically identical, then the segregation will only lead to physical variation between product packages. If there is also a chemical variation in a powder mixture then segregation can be much more dangerous, as the product will then have chemical as well as physical variation between packages.

Free-flowing powders are especially prone to segregation as individual particles are not subjected to any structuring restraints and have considerable freedom of movement.

All powder motion can potentially cause segregation, but the following are some of the more dangerous:

1. Projection. Coarser particles are generally projected farther than finer particles if projected at a constant velocity. Sources of projection in powder processing can be identified in pneumatic conveying, conveyor belt transport, chute flow, and higher-speed mixers.
2. Percolation. Smaller, denser particles have a greater ability and mobility to percolate through a loose mass of particles. One of the severest segregating mechanisms takes place when particles tumble or roll down an inclined plane.

When particles are poured into a heap the coarser particles roll to the edge of the heap and the fine particles percolate through the mobile face of the heap to form a central core. A rolling plane can also often be observed as hoppers, drums, sacks, and packages are being filled or emptied, or when a drum of powder is rolled. Another form of percolating segregation occurs when a loose powder is vibrated. Coarse particles tend to rise to the surface as fine particles percolate to the base of the container. Most product packages are vibrated in the transport operation between manufacture and receipt by the customer, and the product homogeneity can be destroyed in the process.

3. Elutriation. If a gas is passed through a loose powder mixture, the smaller particles will tend to be blown clear or elutriated from the mixture. This process commonly occurs when containers are filled. Powder displaces air as filling occurs, and the air elutriates out the finer particles, which are either blown clear or settle back on to the surface of the container.

As even small differences in particle size, shape, and density can produce significant segregation, the best way of eliminating or significantly reducing segregation is to build a structuring mechanism into the powder flow in some way. This can be done by reducing the mean particle size of the mixture, adding a fibrous or flake-like material, or adding some moisture to the system. This is to say, making the powder more cohesive. If this is not possible, then segregation has to be minimized by careful choice of mixer and mixture handling.

For a free-flowing mixture, the philosophy of mixer choice should be to minimize the opportunity of individual particles to choose their own path of motion and to impose a displacement on the particles by the mixer. A rolling, tumbling mixer gives individual particles freedom and is undesirable, whilst a ribbon or paddle mixer pushes and relocates groups of particles and is to be preferred (**Figure 1**). For this type of mixer, considerable thought should also be given to the method of emptying the mixer, as the emptying process can segregate an otherwise acceptable mixture as the mixture discharges.

Ideally, a free-flowing mixture should be packaged immediately on discharge from the mixer as any intermediate handling between discharge and packaging can potentially destroy the mixture quality.

Cohesive Powders

Cohesive powders have many unattractive properties. They flow badly and can disrupt a packaging line. They cannot easily be metered into the mixing process. They can be dusty and potentially hazardous. They are a potential source of contamination between batches. They have an unattractive appearance.

From a mixing point of view, however, cohesive powders have two very strong advantages. The cohesive structure inhibits segregation and the relatively small constituent particles provide a fine texture. Both these advantages can be overriding if the mixture quality and texture are the dominant requirements of the product.

Cohesive powders have poor mobility because individual particles are held in a structure with their neighbors and cannot easily move independently. If particles cannot move independently, they cannot segregate in the gross manner associated with free-flowing mixtures and the overall mixture quality is enhanced. Structuring is also promoted when the individual particles are small and lack the weight to fall free from the structure. Small particles also give a better statistical chance of achieving a high mixture quality.

A general comment would be that it is much more difficult to produce a bad mixture with a cohesive powder than with a free-flowing powder. With the reduced mobility associated with cohesivity, mixing will be a slower process but, unlike the free-flowing mixing process, it is irreversible.

The two main problems associated with cohesive powder mixing both relate to the structuring capabilities of the powder. It is very easy for cohesive powders to be held up in dead spaces in the mixer and be isolated from the mixing process, and efficient scraping or sweeping of the entire mixing volume is essential. The second problem concerns the strength of the structure between individual particles. If the mixer does not impose a breaking force larger than the structuring force, then aggregates of an ingredient can pass through the mixing process without being dispersed. The aggregates might be efficiently distributed throughout the bulk of the mixture, but on a localized scale they are not producing the same texture as individually dispersed particles.

Whilst the philosophy of mixing free-flowing powders was to suppress the free movement of individual particles, the philosophy for a cohesive powder must be to encourage individual particles repeatedly to break free from their structuring neighbors.

For many cohesive powders, the gentle rolling and breaking action of a tumbler mixer might be sufficient to cause restructuring. For strongly cohesive or aggregating systems it may be necessary to increase significantly the energy input to the process by the addition of high-speed impactors or intensifiers, or by creating a high-shear mixing zone within the mixer (**Figure 1**).

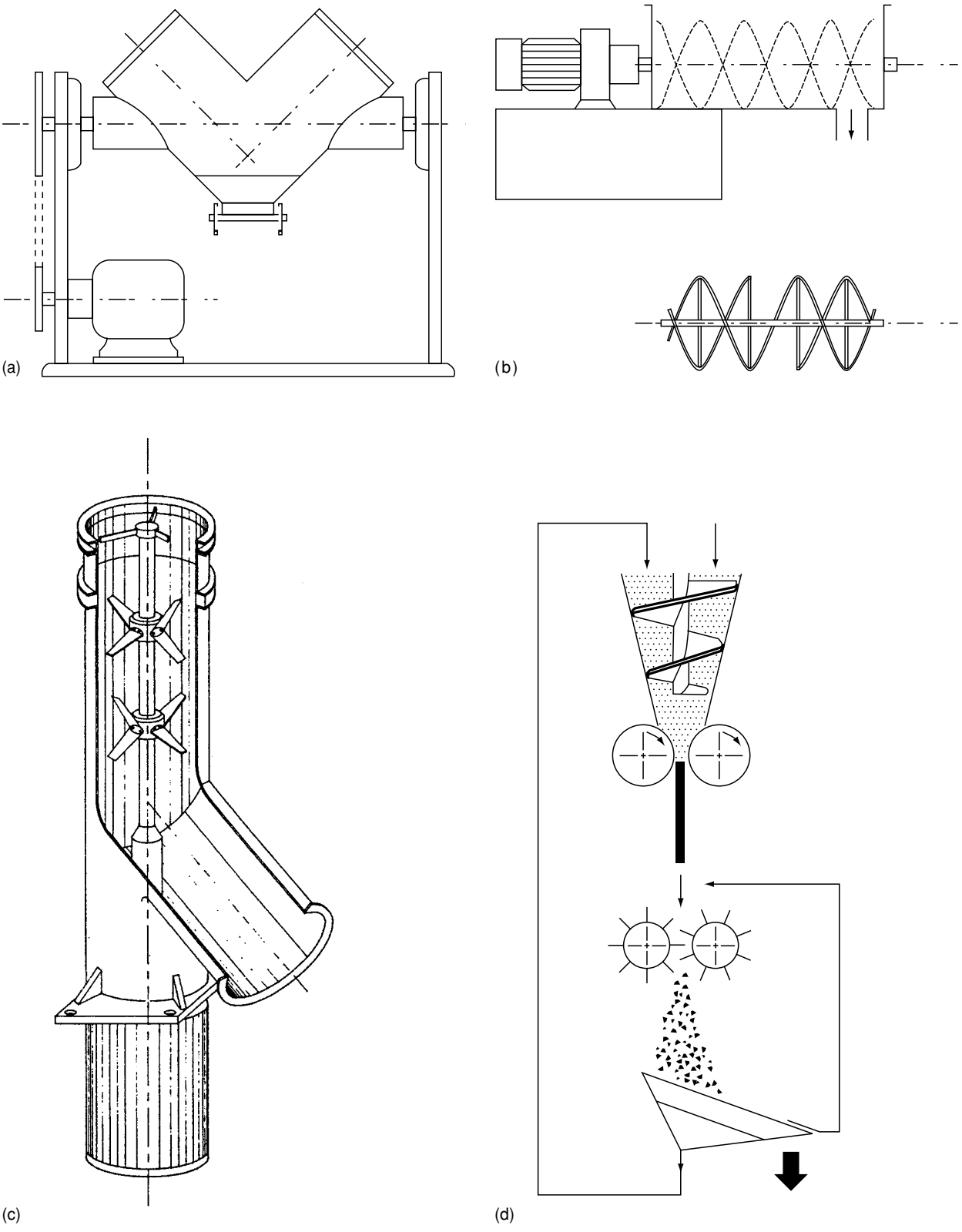

Figure 1 Typical powder mixers: (a) the tumbler mixer; (b) ribbon blender; (c) impactor; (d) high-shear rollers. Reproduced from Mixing of Powders. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

The Texture of Powder Mixtures

What is well mixed? Very often a powder-mixing process is designed and installed without properly defining the quality requirements of the product. Mixture quality can be assessed on the basis of the number of letters of complaint received from disappointed customers, but usually a better and earlier quantitative assessment is required. To make this assessment it is essential first to identify the weight of mixture which will be judged for quality by the customer, and then to assess both the possible and attained statistical quality achieved in the mixing process.

Scale of Scrutiny

It is essential to identify the least weight of a powder mixture used by a customer in a single 'application.' This weight is frequently referred to as the required 'scale of scrutiny' for the mixture. Mixture quality is sensitive to the scale of scrutiny and the smaller the scale, the greater will be the difficulty of obtaining a satisfactory mixture quality.

Identifying the scale of scrutiny is often quite straightforward. Thus, if a sachet of dried soup powder is completely emptied into a pan of boiling water, the quality of the soup is determined by the weight and quality of powder delivered in one sachet, and this is then the scale of scrutiny. If marketing requirements decree that a cup rather than a pan of soup should be the package content requirement then the scale of scrutiny has been reduced, and it is possible that the mixer will no longer be able to meet the higher quality requirement.

A significant problem arises if the weight of package delivered to the customer is larger than the customer's scale of scrutiny. It could be, for example, that a customer makes two small pans of soup from a single sachet or takes many bowls of muesli from a single packet. The manufacturer fills the sachet or packet with a satisfactory mixture but, in the subsequent transporting and pouring operations, the contents may well segregate and the customer receives an unsatisfactory mixture when the original quantity is subdivided. It is evidently desirable to deliver the mixture to the customer in package quantities identical to the customer's scale of scrutiny.

As the scale of scrutiny is the weight of mixture assessed by the customer to determine quality, it is also the sample weight removed from the process by the manufacturer for a statistical assessment of quality.

The Statistical Assessment of Quality

If samples are withdrawn from a series of points in a mixture and analyzed separately for composition, then the variance, S^2_{ex}, of these samples can be estimated. A large variance indicates a poor mixture. In the form of the standard deviation, this experimental statistical value can be used to predict the number of product packages which would be outside the required composition tolerance.

When related to mixing data states, the experimental variance can also be used to predict the minimum mixing times or the presence of segregation. The process in Figure 2 illustrates three possible mixing states for a binary mixture of equal-sized particles. The unmixed state has the largest variance value S^2_0, and in most circumstances the best attainable mixture will be that of the randomized mixture, S^2_R. If segregation occurs, the variance S^2_R will not be attained and the product will have some higher variance value.

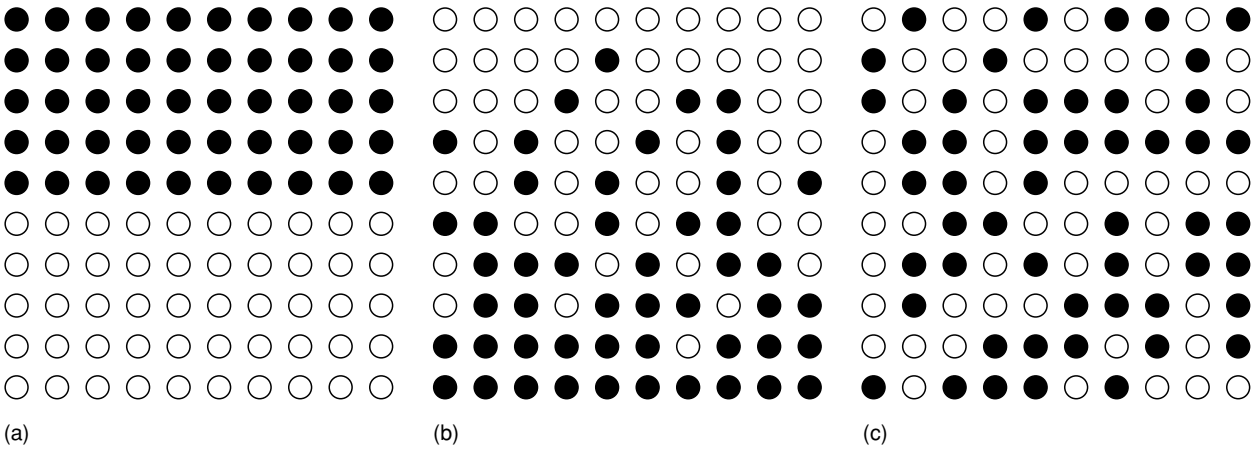

Figure 2 Possible mixing states for a binary mixture of equal-sized particles: (a) unmixed state; (b) segregated; (c) randomly mixed. Reproduced from Mixing of Powders. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

For the illustrated binary system of equal-sized particles, the limiting variance values can be predicted as

$$S_0^2 = pq \tag{1}$$

and

$$S_R^2 = pq/A \tag{2}$$

where p and q are the proportions of the two ingredients and A is the number of particles in a given sample size or scale of scrutiny.

Evidently, few 'real' mixtures fulfill the requirement of being binary systems of equal-sized particles, but the ideal form of eqns (1) and (2) enables us to derive some important mixing principles which will be applicable to multicomponent mixtures of multisized particles.

Eqn (2) illustrates the importance of the scale of scrutiny on mixture quality. A large scale of scrutiny gives a large value of A and a low potential variance, S_R^2. If the weight or the scale of scrutiny is fixed, then an alternative method of reducing the potential variance would be to reduce the particle size and hence increase the value of A. This is a quantitative assessment of the value of mixing cohesive particles as a means of improving texture. Whilst there is a definable relationship between the random mixing state and sample size, this relationship is not valid for other mixing states and, for completely unmixed systems, the variance value S_0^2 is independent of sample size. The inability to relate quality to sample size for intermediate mixing states makes it very difficult to predict the effect on the resultant mixture quality of scaling up or down a scale of scrutiny.

Eqns (1) and (2) can be used in conjunction with the experimentally determined variance S_{ex}^2, to give a mixing index (M). A common index is

$$M = \frac{(S_0^2 - S_{ex}^2)}{(S_0^2 - S_R^2)} = \frac{\text{How far the mixture has progressed}}{\text{How far the mixture could progress}}$$

In this case $M = 0$ for no mixing and $M = 1$ when the mixture is randomized. Evidently, a plot of M against time would indicate the minimum mixing time required.

Eqns (1) and (2) are available for multicomponent and multisized particulate systems. The equations are more complex but the principles are the same. When numbers are inserted for real systems some general conclusions become evident.

1. A few large particles in an ingredient have a significant effect on attainable mixture quality. This is a particularly important effect for minor ingredients and emphasizes the importance of keeping a tight control on any prior processes which affect the size distribution of mixture ingredients.
2. For a multicomponent mixture, the randomization of one ingredient does not insure that other ingredients will be randomly distributed. It is necessary to carry out separate statistical analyses for at least the key ingredients.
3. The derivatives of eqns (1) and (2) enable calculations to be carried out to insure that the mixing process is not attempting the impossible. When a product at its scale of scrutiny contains relatively few particles of one ingredient (say, less than 500), the spread of the number of particles per sample can be calculated and checked with the marketing requirement. Typical examples of this calculation would be the number of chocolate chips on a cookie, the number of sausages in a can of sausages and beans, or the number of mushrooms in a soup packet.

Given that a scale of scrutiny, the proportion by weight of the ingredients, the particle shape, density, and size distribution of the ingredients are known, then the possible attainable quality of the mixture can be calculated. This may not be reached because of segregation but it enables the potential quality of a range of formulations to be assessed and compared to a marketing requirement without carrying out any experimental testing.

The Testing of Mixers

If properly carried out, the testing of a mixer with a particular powder formulation is a long and expensive process and at best is to be avoided and at worst to be carried out with the most effective use of time and effort. One essential first step is to calculate the possible value of S_R^2 for a mixture formulation at a given scale of scrutiny, and check that the mixture quality required is attainable. A second essential is to test the quality of the mixture as it exits the mixer and is about to be packaged rather than in the mixer itself. This is especially important for free-flowing powder formulations. All testing procedures require that samples be taken from the mixture and subsequently assessed for the experimental variance value S_{ex}^2 or for some running statistical control. Especially for free-flowing powders, such sampling procedures are very prone to bias and, if this occurs, much expensive effort can be wasted.

Sampling of Mixtures

The quantity of sample removed from a point in the discharge of a mixture should be that of the scale of scrutiny. A larger sample will give an optimistic measure of quality and a smaller sample a pessimistic measure. It is important to retain the identity of

every point sample withdrawn from the mixture and to carry out a bulk analysis on each of the point samples taken. Many misunderstandings occur between production and quality assurance sections when the quantity of material required for analysis is less than the scale of scrutiny of the mixture. This reduces the effective scale of scrutiny, increases the variance between samples, and indicates a poor mixture quality. Ideally, analysis should be carried out on the entire point sample but, if this is not practically possible, then the sample should be riffled down to the required analytical quantity.

Poor decisions on where and how to take samples can result in significant bias of estimates of sample composition and quality of the mixture. This is especially so for free-flowing powders, which are prone to segregation.

If careful instructions are not given to the process operator on where to take a sample, then the most convenient sample will be taken. Typically this will be on the surface of a drum, adjacent to a valve, or nearest to the control room, and will almost certainly be nonrepresentative of the bulk material. A simple random selection using random numbers is statistically the best way of locating representative samples but, for a routine sampling procedure, it can be inconvenient. A systematic sampling process which locates samples at a fixed time interval is an easier selection process to apply, but care must be taken not to choose a sample frequency which coincides with a natural frequency of the process.

The thief or probe sampler is commonly used to retrieve samples from a selected location in a bulk powder. For a free-flowing powder, it can easily be demonstrated that this is a heavily biased retrieval. Rules to be followed to minimize the bias of sample retrieval are:

1. Retrieve the samples from a flowing stream and not from a bulk of powder. The mixture should ideally be sampled as it discharges from the mixer and not *in situ*.
2. Sample the entire section of flowing powder to allow for any segregation across the flowing stream.
3. Take as many samples at the required scale of scrutiny as possible, as this will improve the statistical precision of the quality estimates. The reader is referred to a standard text to determine the precision reward-to-sampling-effort ratio but, for an assessment of a batch mixture quality, there would be little point in taking fewer than 10 samples.

When samples are taken in a time sequence, considerable benefit can be obtained by plotting composition against time to produce a 'discharge profile' for the system. A mixture variance S_{ex}^2 can be calculated for the mixture but the discharge profile often reveals cyclical variations in the product, which can result in both process improvements and economy of sampling effort.

See also: **Agitation**; **Bread**: Dough Mixing and Testing Operations; **Cakes**: Methods of Manufacture; **Cereals**: Handling of Grain for Storage; **Cleaning Procedures in the Factory**: Overall Approach; **Marine Foods**: Production and Uses of Marine Algae; **Packaging**: Packaging of Solids; **Plant Design**: Basic Principles; **Quality Assurance and Quality Control**

Further Reading

Danckwerts PV (1953) The definition and measurement of some characteristics of mixtures. *Applied Scientific Research, Section A* 3: 279–296.

Harnby N, Edwards MF and Nienow AW (1997) *Mixing in the Process Industries*, 2nd edn. Oxford: Butterworth-Heinemann.

Lacey PMC (1954) Improvements in the theory of particle mixing. *Journal of Applied Chemistry* 4: 257–268.

Rhodes MJ (1990) *Principles of Powder Technology*. Chichester: John Wiley.

Modified-atmosphere Packaging *See* **Chill Foods**: Effect of Modified-atmosphere Packaging on Food Quality; **Chilled Storage**: Principles; Attainment of Chilled Conditions; Quality and Economic Considerations; Microbiological Considerations; Use of Modified-atmosphere Packaging; Packaging Under Vacuum

Mold-ripened Cheese *See* **Cheeses**: Types of Cheese; Starter Cultures Employed in Cheese-making; Chemistry and Microbiology of Maturation; Manufacture of Extra-hard Cheeses; Manufacture of Hard and Semi-hard Varieties of Cheese; Cheeses with 'Eyes'; Soft and Special Varieties; White Brined Varieties; Quarg and Fromage Frais; Processed Cheese; Dietary Importance; Mold-ripened Cheeses: Stilton and Related Varieties; Surface Mold-ripened Cheese Varieties; Dutch-type Cheeses

Molds *See* **Mycotoxins**: Classifications; Occurrence and Determination; Toxicology; **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage; Fungi in Food – An Overview; Molds in Spoilage; Yeasts in Spoilage

Molluscs *See* **Shellfish**: Characteristics of Crustacea; Commercially Important Crustacea; Characteristics of Molluscs; Commercially Important Molluscs; Contamination and Spoilage of Molluscs and Crustaceans; Aquaculture of Commercially Important Molluscs and Crustaceans

Monosodium Glutamate *See* **Taste Enhancers**

MULTIPLE SCLEROSIS – NUTRITIONAL MANAGEMENT

R Cichon and L Wadolowska, Warmia and Masuria University in Olsztyn, Olsztyn, Poland

Background

Multiple sclerosis (MS) belongs to the demyelinating diseases, which comprise a group of neurologic disorders. Demyelinating diseases have in common the pathologic feature of local or patchy destruction of myelin sheaths in the central nervous system accompanied by an inflammatory response. Destruction of the myelin sheaths leads to dysfunction of the central nervous system (CNS) in transmission of electrical nerve impulses.

The term 'sclerosis' is applied to the hardening of the CNS and is called 'multiple,' because multiple areas of optic nerves, brain and spinal cord undergo sclerosis, whereby myelin is replaced with sclera and/or scar tissue. The disease was first described by Cruveilhier in 1835 under the name 'sclerose en plaque,' in that the disease is marked by the appearance of hard patches throughout the brain and spinal cord. Its outstanding symptoms are weakness, muscular incoordination, jerky movements of the eyeballs, paralysis, and a curious, deliberate way of speaking, known commonly as 'scanning speech.' The condition has periodic remissions and on the whole tends to become worse, but is seldom an immediate threat to life. The disease most commonly strikes between the ages of 20 and 40 years, and females account for 60–70% of cases. The average age of onset of the first clinical episode of MS is in the third and fourth decades. It is rare for the disease to begin in childhood or beyond the sixth decade, but it is not unknown.

Features of Multiple Sclerosis

MS is pleomorphic in its presentation. The clinical picture is determined by the location of foci of demyelination within CNS. Classic features include impaired vision, nystagmus, dysarthria, decreased perception of vibration and position, ataxia and intension tremor, weakness or paralysis of one or more limbs, spasticity, and bladder problems. Diagnosis is based on the clinical presentation, neurological tests and magnetic resonance imaging (MRI, to reveal the presence of plaques). There are two main forms and two subtypes of the disease: relapsing-remitting MS (RRMS), chronic progressive MS (CPMS), and secondary progressive and benign MS. In RRMS, periods of acute worsening of symptoms are followed by spontaneous improvement. In CPMS, there are no clear-cut attacks, and the symptoms become steadily worse. This form of MS might progress rapidly,

causing severe disability and even premature death within a few years. Secondary progressive MS is described when RRMS becomes progressive, and benign MS when an initial attack is never followed by another. The criteria for the diagnosis of clinically definite MS include a reliable history of at least two episodes of neurologic deficit and objective clinical signs of lesions at more than one site within the CNS. Clinically probable MS is defined as either two attacks with clinical evidence of one lession or one attack with clinical evidence of two lesions. However, classifications of MS are based on clinical phenomenology, and the definitions of relapse, remission, and progression are still imprecise.

MS usually presents in the form of recurring attacks of focal or multifocal neurologic disfunction, reflecting lesions within the CNS. The disease begins most commonly in early adult life. The frequency of flare-ups is greatest during the first 3–4 years of the disease, but a first attack may be so mild as to escape medical attention, for example, blurred or double vision. Once the initial symptoms disappear, they may never recur, and a period of months or years may pass before other symptoms appear, such as weakness of certain muscles, unusual tiring of a limb, minor interference with walking, muscle stiffness, dizziness, loss of bladder control, and disturbances in the senses of touch, pain, and heat. Each symptom may appear and then disappear, only to be followed by another. The disease continues to progress, and eventually, the patient becomes crippled and even bedridden. However, the progression and remission of MS are different for every individual. Although occasional patients die within the first few years of disease onset, most do not, and the average survival from MS is more than 30 years after the onset of disease.

Pathology

In 1961, Swank proposed the hypothesis that a dysfunction of the blood–brain barrier is a primary damage defect in MS. Current opinion holds that autoimmunity or viral infection is likely to be implicated in their pathogenesis.

Many scattered, discrete areas of demyelination, termed plaques, are the pathologic hallmark of MS. Plaques vary in size and may be found anywhere in the white matter but typically occur in the paraventricular areas of the cerebrum and subpially, and within the brainstem and spinal cord. The peripheral nervous system is not affected. Many plaques are clinically silent. In fact, some autopsy studies indicate that 20% of MS cases are clinically silent during life.

Active MS lesions feature T-lymphocyte and monocyte–macrophage accumulations about venules and at plaque margins is being destroyed. The invasion of white matter by inflammatory cells is held responsible for the myelin breakdown. They also function as scavengers of myelin debris; fat-laden macrophages may persist for months, perhaps years, after the acute inflammatory response has subsided. Despite much research, viral inclusions have not been detected. Only limited regeneration of myelin occurs, and the reason is unclear. There are probably several mechanisms responsible for recovery from an MS attack. The pathologic features of MS fail to account for hour-to-hour and day-to-day waxings and wanings in function so characteristic of the disease. Conduction of impulses through demyelinated nerve is compromised and is further altered by transient changes in the internal milieu such as functuations in temperature and electrolyte balance or by stress.

Etiology

The causes of MS remain unknown. A role for immune-mediated or infectious factors has been proposed, but data to support these postulates are inconclusive. Epidemiological studies have implicated an interplay between genetic and environmental factors in the etiology of MS. It has been suggested that nutrition and food pattern, particularly high consumption of animal fat and low intake of fish products may play a role in the etiology of MS. Recently, studies have focused on long-chain polyunsaturated fatty acids (LCPUFA) and vitamin D.

Epidemiology

Epidemiologic studies have established several facts that ultimately will have to be incorporated into any coherent theory of the disease. Neuroepidemiology has played a fundamental role in the study of multiple sclerosis by providing some etiologic clues, although a definitive basis for the conclusive resolution of its enigma is still lacking. Epidemiologic and genetic studies have indicated that MS is probably caused by multiple factors, both genetic and environmental, none of which is individually sufficient, and appear to act before adolescence – or possibly later – in genetically susceptible individuals. This unifying hypothesis emphasizes the role of a genetic–racial susceptibility, the importance of environmental factors, a possible etiologic heterogeneity, and a lack of specificity of the unknown endogenous and exogenous agents. Situations or events with biological plausibility, such as childhood or adolescent infectious diseases, exposures to geographic and sociocultural factors,

nutritional habits, hypersensitivity, significant head and spinal trauma, and other factors may contribute, at different times, to the putative acquisition of MS, trigger its onset, and modify its subsequent course.

In general, the incidence in temperate climates exceeds that in tropical zones, but variations within regions with similar climates do exist; hence, the effect is not simply one of latitude or temperature. Geographically, MS describes three frequency zones. High-frequency areas (characterized by a prevalence of 30+ per 100 000) now comprise most of Europe, Israel, Canada, northern USA, southeastern Australia, New Zealand, and eastern Russia. Medium-frequency areas include the southern USA, most of Australia, South Africa, the southern Mediterranean basin, Russia into Siberia, the Ukraine and parts of Latin America. Prevalence rates of less than 5 per 100 000 are found in the rest of Asia, Africa, and northern South America. The incidence of MS in northern Europe, Canada, and the northern USA is approximately 10 new cases each year per 100 000 persons between the ages of 20 and 50, two to three times higher than that in Australia, New Zealand, and the southern USA. MS is rare in Japan, elsewhere in the Orient, and Africa. Some epidemiologic evidence also suggests that persons emigrating from high- to low-risk regions as children may be partially protected from MS. Immigrants from high- to lower-risk areas retain the MS risk of their birth place only if they are at least 15 years old at the time of emigration. Those from low- to high-risk areas increase their risk even beyond that of the natives, with susceptibility extending from about 11 to 45 years of age. MS occurred in epidemic form in the North Atlantic islands, particularly in the Faroe Islands, where the first symptom onset was in 1943, heralding the first of four successive epidemics at 13-year intervals. What was transmitted is thought to be a specific, widespread, persistent infection called primary multiple sclerosis affection (PMSA), which only rarely leads years later to clinical MS. The search for PMSA is best attempted on the Faroes, where there are regions still free of MS after 50 years. The data are consistent with the existence of an environmental factor, possibly a virus, and perhaps geographically restricted, that influences the development of MS. Immunologic, epidemiologic, and genetic data indicate that tissue injury in MS results from an abnormal immune response to one or more myelin antigens that develop in genetically susceptible individuals after exposure to an as-yet undefined causal agent.

There is some evidence for genetic factors in multiple sclerosis. The evidence comes from epidemiologic studies, racial predilection, risk in family members (sibs, half sibs, adoptees), and twin studies. MS is not a Mendelian inherited disease; only the susceptibility to the disease is inherited. MS seems to be an oligo- or multigenic disorder with an apparently similar phenotype for the different genes involved. A genetic component in MS is indicated by an increased relative risk to siblings compared with the general population and an increased concordance rate in monozygotic compared with dizygotic twins. Whole genome screens conducted in different populations have identified discrete chromosomal regions potentially harboring MS-susceptibility genes.

Nutrition and Multiple Sclerosis

Specific nutritional therapy has been advocated for a number of neurologic diseases of undetermined etiology. In some of the disorders, improper or inadequate nutrition is implicated. Therefore, dietary manipulation constitutes the main therapeutic mode. For instance, a low-fat diet has been recommended as an effective means of reducing the incidence of exacerbations in patients suffering from MS. However, objective evaluation of the efficacy of this treatment modality has not yet been possible. In MS, a combination of genetic and environmental factors, including dietary factors, underlies the symptoms and signs of neurologic dysfunction that can be ameliorated by changes in the diet and/or vitamin supplementation.

Epidemiologic studies relating MS to nutritional factors have revealed a possible link between the incidence of the disease, total fat intake, and percentage of calories (of animal origin) consumed. A long-term study carried out by Swank showed significantly less deterioration and much lower death rates among MS patients when they consumed less fat (20 g of fat per day).

The results of prospective case-control studies showed that some foods consumed at certain 'critical' ages could play a causal role in the oneset of MS. An association was suggested between MS and high consumption of bread and 'pasta,' butter and lard, legume soup, horse flesh, coffee and tea in the period from infancy to adolescence, and of eggs and wine during adulthood. A possible autoimmune demyelinating disease with lipid changes suggests a deficiency of PUFA.

MS tends to be more prevalent in countries where the use of animal fat is high. It has also been suggested that the administration of unsaturated fatty acids, such as linoleic or arachidonic acid, may reduce the number and severity of MS attacks. This idea is based on the observation that brain and spinal cord obtained from patients who have died of MS is deficient in unsaturated fatty acids and that linoleic and

arachidonic acids tend to inhibit the lymphocyte–antigen interaction, the cellular mechanism that may enhance demyelination. Sensitized lymphocytes probably interact with myelin components in the affected parts of the nervous system during an attack of MS. That the ingestion of unsaturated fatty acids, such as linoleic acid, affects the course of MS remains unproved.

Early epidemiological studies demonstrated a low incidence of atherosclerotic, inflammatory, and autoimmune diseases in Greenland Eskimos who consumed diets high in marine-derived (*n*-3) PUFA. Human studies have consistently demonstrated a decrease in production of pro-inflammatory cytokines when moderate to high levels of marine-derived (*n*-3) PUFA are taken orally. The decrease in production of pro-inflammatory cytokines and eicosanoids contributes to the beneficial effect of fish oil in reducing the pathogenesis of inflammatory and atherosclerotic diseases. Essential fatty acids (EFA) are necessary for normal immune function of humans and animals. EFA deficiency impairs B- and T-cell-mediated responses. These impairments can be ameriorated by the inclusion of EFA in the diet. The possibility that polyunsaturated fatty acids have an immunosuppressant effect has led to the evaluation of high-polyunsaturated fatty acids diets in MS, as neurologic disorders of immune-inflammatory etiology that affect CNS. α-Linolenic acid (C18:3 *n*-3) can be converted to EPA (C20:5 *n*-3), which can replace arachidonic acid in membrane phospholipids and can be preferentially used by cyclooxygenase. This results in the reduced production of prostaglandins (PG) of the 2 series and leukotrienes of the 4 series. Thus, the consumption of oils containing C18:3 *n*-3 might be associated with immunological changes. This opinion is supported by many studies that have shown immunological modulation following marine oil consumption.

Evidence on etiology in MS suggests that the prevalence depends on the interaction of two factors, diet and exposure to visible sunlight. The dietary features that may be beneficial include supplementation with fish oils, avoidance of saturated fats, and the associated intake of antioxidants with unsaturated fatty acids. Inhibition, by antioxidants, of the enzyme lipoxygenase inhibits leukotriene synthesis, and the presence of fish oils leads to the production of leukotrienes with less inflammatory properties. This is of particular importance in the retina, where leukotrienes might be the underlying cause of retrobulbar neuritis.

A significant protective effect has been observed with other nutrients, including vegetable protein, dietary fiber, vitamin C, thiamin, riboflavin, calcium, and potasium. With respect to specific foods (as opposed to nutrients), a higher intake of fruit juices was inversely associated with risk. A protective effect was also observed with cereal/bread intake; pork/hot dogs and sweets/candy were positively associated with risk. The study generally supports a protective role for components commonly found in plants (fruit/vegetable and grains) and increased risk with high-energy and animal food intake.

An epidemiologic study on the relationship between mortality rates from MS for the period 1983–1989 for 36 countries has shown that saturated fatty acids, animal fat, animal minus fish fat, and latitude correlated independently and positively with MS mortality. The ratio of polyunsaturated fatty acids to saturated fatty acids (P/S ratio) and the ratio of unsaturated to saturated fatty acids (U/S ratio) correlated independently and negatively with MS mortality. The results support the data, at least partially, on an association between certain dietary factors and MS.

For MS patients, good nutrition has the potential to enhance the quality of life and reduce the risk of lifestyle diseases and secondary conditions. Eating patterns in patients with MS showed the need for interventions to promote good nutrition among persons with MS. Usually, nutritional intake for MS patients is not optimal in many important areas.

Acute and chronic neurologic diseases of nonnutritional origin that impair consciousness of motor functions essential to adequate nutritional intake, such as paralysis or weakness of the facial muscles, the tongue, the pharynx, and the muscles of deglutition, frequently require the temporary or permanent use of a nasogastric tube or gastrostomy for the patient's well-being. These measures make possible proper enteral nutrition and may, in fact, prove to be life-saving. In most of these disorders, nutritional support does not affect the basic pathogenesis of the disease. The use of B-complex vitamin preparations for a variety of acute and chronic neurologic diseases of cryptic origin continues to be common medical practice. To date, however, there is no objective evidence that this practice has any effect on the speed of recovery of the afflicted nervous tissue.

Despite significant advances over previous decades in the field of nutrition, large gaps remain in our understanding of the role in the normal and abnormal activity of the nervous system. More sensitive and critical methods for the assessment of the nutritional status of the nervous system and for the evaluation of the dietary management of MS are essential for more rational and effective therapeutic manipulation (treatment).

Nutritional therapy consisting of vitamins and other nutrients continues to be used in neurologic

disorders of unknown cause, in spite of a total lack of scientific rationale. The popularity of administering large doses of vitamins (and other nutrients) as therapy for chronic neurologic diseases of uncertain etiology stems from the fact that experimentally induced deficiencies of this nutrients in animals frequently result in symptoms caused by reversible lesions of the CNS. In most instances, the far-advanced lesions show demyelination, a frequent finding in many human neurologic diseases. However, considering the large doses of vitamins consumed daily, for a variety of reasons, by an ever–increasing number of people, vitamin toxicity affecting the nervous system is relatively rare. Of all the vitamins, only two – vitamin A and pyridoxine – are known to produce adverse neurologic reactions when ingested in pharmacologic, rather than recommended, nutritional doses.

Patients who are being treated with drugs may incur a significant degree of vitamin deficiency (D, B_{12}, and folic acid). The prolonged administration of drugs may lead to a decrease in bone mineral content. The drugs interact with the vitamin D and its active forms, decreasing intestinal calcium absorption and redistribution. The severity of demineralization appears to be related to the dose of the drugs, a deficiency of physical activity, and a lack of exposure to sunlight. In general, the problem does not appear to be of sufficient clinical importance to warrant the routine vitamin D supplementation of patients with MS drugs; however, the possibility should be borne in mind, and vitamin D metabolites should be estimated periodically in long-term patients with poor intake or malabsorption.

Studies with animals, tissue cultures, as well as humans indicate that both the level and degree of saturation of dietary fats influence inflammatory and immunologic responses. Some of these immunologic changes are associated with beneficial clinical effects, for example, reduction in pro-inflammatory cytokine by marine-derived (*n*-3) PUFA. Others might not be desirable, for example, decreased T-cell-mediated function by marine-derived (*n*-3) PUFA in older people.

Recommendations to increase PUFA intake should be accompanied by appropiate recommendations for an increase in antioxidant intake, in particular vitamin E. Without adequate antioxidant protection, the substitution of membrane fatty acids with highly oxidizable (*n*-3) PUFA of fish oil, that is, EPA and DHA, may potentiate the peroxidation of cellular mambranes. Foods high in polyunsaturates should contain at least 0.6 mg of tocopherol equivalents per gram of polyunsaturated fatty acids. Higher levels may be necessary for fats rich in fatty acids containing more than two double bonds.

In general, for MS patients, recommendations to decrease dietary total fat intake, with the inclusion of moderate amounts of (*n*-6 and *n*-3) PUFA with adequate antioxidant nutrients should provide for a competent immune response. The *n*-6 and *n*-3 fatty acids have critical roles in the membrane structures and as precursors of eicosanoids, which are potent and highly reactive compounds. Various eicosanoids have widely divergent, and often opposing, effects on inflammatory processes and the immune system. Since they compete for the same enzymes and have different biological roles, the balance between the *n*-6 and *n*-3 fatty acids in the diet can be of considerable importance. The ratio of linoleic to α-linolenic acid in the diet should be between 5:1 and 10:1. Patients with a ratio in excess of 10:1 should be encouraged to consume more *n*-3 rich foods such as green leafy vegetables, legumes, fish, and other seafood.

Dietary Management of Multiple Sclerosis

Although there have not been any valid clinical trials supporting the efficacy of nutrition in delaying the progression of MS, it is imperative that the dietitian evaluates the nutritional health status and arranges a patient's diet. The nutritional management of patients with MS is primarily concerned with ensuring that their nutrient intake meets the requirements. The specific dietary modifications that are needed in some patients, mostly because of drug therapy, are of secondary importance. The principles of management have to take in consideration the individual needs, circumstances and difficulties. The current consensus is that extreme dietary change should have no place in the management of MS. Patients with MS should be encouraged to follow a varied, well-balanced diet based on good nutritional practice and principles for the population in general. In order to achieve the recommended dietary targets, considerable changes in the dietary habits of the MS patient will be required, usually necessitating:

- increased intake of micronutrient-rich foods such as fruits, vegetables, pulses, seeds, legumes, and cereals. Leafy green vegetables may contribute adequate amounts of antioxidative vitamins, e.g., vitamins E and C;
- use of low-fat alternatives for full-fat milk and dairy products;
- replacing saturated fat spreads and cooking fats with low-fat spreads and monounsaturated oils;
- avoiding sugar-rich/fat-rich food group, such as meat products, pastries, chips, snakes, cakes and biscuits;
- encouraging the consumption of all types of fish and seafood; two portions of fish or seafood per

week, including oily fish, are recommended to secure a moderate increase intake of *n*-6 and *n*-3 fatty acids, in particular LCPUFA.

An important consideration in the management of all patients is their hydration status. The consumption of at least six to eight glasses of fluid per day usually meets the adequate fluid intake of a patient.

Nutritional problems associated with MS vary according to the symptoms of illness and the degree of physical disability of patients. Dysphasia may occur as the disease progresses, and the diet consistency may need to be modified from solid to soft or puréed foods and even thick liquids. Other problems include impaired vision, dysarthria, and poor ambulation, thus making eating less enjoyable because meal preparation becomes a difficult task. In this situation, single-serving or convenience foods often permits independent preparation of meals. Given the chronic nature of this disease, patients may require enteral nutrition support.

Neurogenic bladder is common, causing urinary incontinence, urgency, and frequency. To minimize these problems, it is helpful to distribute fluids evenly throughout the waking hours and limit them before bed. Some patients limit fluid intake severely to decrease the frequency of urination, thus increasing the risk of urinary tract infection.

Neurogenic bowel can cause either constipation or diarrhea, and the incidence of fecal impaction is increased in MS. A diet that is high in fiber with additional prunes and adequate fluid intake can moderate these problems.

Supplement Usage

Patients may be taking combinations of different supplements in large doses, sometimes risking overdosage with fat-soluble vitamins. The types of supplements likely to be used include evening primrose oil, fish oil capsules, cod liver oil, and linseed oil. If a well-balanced diet containing dietary sources of *n*-3 and *n*-6 fatty acids is consumed, additional supplements of LC-PUFA sources, vitamins, and minerals are unlikely to be necessary or beneficial. Patients who wish to take supplements should be advised to restrict these to one multiple multivitamin/mineral preparation providing no more than the recommended dietary allowance, rather than taking several different types of supplement. In any case, it is best to do this after consultation with the dietitian.

See also: **Cholecalciferol**: Properties and Determination; **Essential Fatty Acids**; **Fish Oils**: Dietary Importance; **Vitamins**: Overview

Further Reading

Antel JP and Arnason BGW (1987) Demyelinating diseases. In: Braunwald E *et al.* (eds.) *Harrison's Principles of Internal Medicine*, 11th edn, pp. 1995–2000. New York: McGraw-Hill.

Briony T (ed.) (2001) *Manual of Dietetic Practice*, 3rd edn. Oxford: Blackwell Science.

FAO/WHO (1994) *Fats and Oils in Human Nutrition. FAO Food and Nutrition Paper 57*. Rome: FAO/WHO.

Hamazaki T and Okuyama H (ed.) (2001) Fatty acids and lipids – new findings. *World Review of Nutrition and Dietetics* 88.

Kahana E (2000) Epidemiologic studies of multiple sclerosis: A review. *Biomedicine and Pharmacotherapy* 54: 100–102.

Kurtzke JF (2000) Multiple sclerosis in time and space – geographic clues to cause. *Journal of Neurovirology* 6(supplement 2S): 134–140.

Mahan LK and Escott-Stump S (eds) (2000) *Krause's Food, Nutrition & Diet Therapy*, 10th edn. Philadelphia, PA: W.B. Saunders.

Oksenberg JR and Barcellos LF (2000) The complex genetic aetiology of multiple sclerosis. *Journal of Neurovirology* 6(supplement 2S): 10–14.

Muscle *See* **Exercise**: Muscle; Metabolic Requirements

MUSHROOMS AND TRUFFLES

Contents

Classification and Morphology

S Rajarathnam and M N Shashirekha, Central Food Technological Research Institute, Mysore, India

Background

Mushrooms constitute a specialized group of fungi and represent the fruiting bodies with a great diversity in shape, size, color, and texture, bearing the spores of sexual reproduction. In nature, they serve the function of production, maturation, and dissemination of spores, for the perpetuation of the species. Thus, these mushrooms as a group of fungi, are nonchlorophyllous organisms that cannot utilize solar energy to manufacture their own food as do green plants. However, mushrooms can produce a wide range of enzymes that can degrade the complex substrates on which they grow, following which they absorb the soluble substances for their own nutrition. This absorptive nutrition is a characteristic of fungi. Mushrooms as macrofungi can be either epigeous, as in 'morels' or 'field mushrooms,' or hypogeous, as in 'truffles,' and large enough to be seen with the naked eye and to be picked by hand. Mushrooms need not be aerial, or fleshy, or edible. Out of 69 000 described species of fungi, there are about 10 000 species of fleshy macrofungi, and only a handful of these are lethal. There are no simple ways of distinguishing between edible and poisonous mushrooms. This will be dealt with in a later section.

Classification

Mushrooms and truffles fall under the class Basidiomycetes and a few orders of Euascomycetes, respectively (Figure 1).

Euascomycetes

Euascomycetes include morels (species of *Morchella*) with epigeous fruiting bodies and truffles (species of *Tuber*) with hypogeous fruiting bodies and are characterized by the production of ascospores endogenously inside special club-shaped structures called asci.

Basidiomycetes

Basidiomycetes are considered the most highly evolved group of fungi and are characterized by the exogenous production of basidiospores borne over club-shaped structures called basidia.

Subclass: Hymenomycetes

This group includes all fungi composed of membranes, fleshy, woody, or gelatinous, whether growing on the ground or on wood. The spore-bearing surface, called the 'hymenium,' is external at an early stage in the life cycle. The spores are borne on basidia, and when the spores ripen, they fall to the ground or are carried by the wind to a host that presents all the conditions necessary for germination; there, they produce the thread-like structures called mycelia. Within these threads are small knots, which in time develop into full-grown mushrooms.

The subclass Hymenomycetes is generally classified in one large order, the Agaricales, which in turn is divided into six families:

1. Agaricaceae–hymenium with gills;
2. Polyporaceae–hymenium with pores;
3. Hydnaceae–hymenium with spines;
4. Thelephoraceae–hymenium horizontal and mostly on the under surface;
5. Clavariaceae–hymenium on a smooth club-shaped surface;
6. Tremellaceae–hymenium even and superior – gelatinous fungi.

Subclass: Gastromycetes

The word 'Gastromycetes' is derived from two Greek words, '*gaster*' (stomach) and '*mycetes*' (fungus). The hymenium is enclosed in the rind or peridium. The word 'peridium' comes from 'peridio' ('I wrap around'), because the peridium entirely envelopes the spore-bearing portion, which eventually sheds the enclosed spores formed inside the basidia and spicules. The cavity within the peridium consists of two parts: a threaded part, called the capillitium,

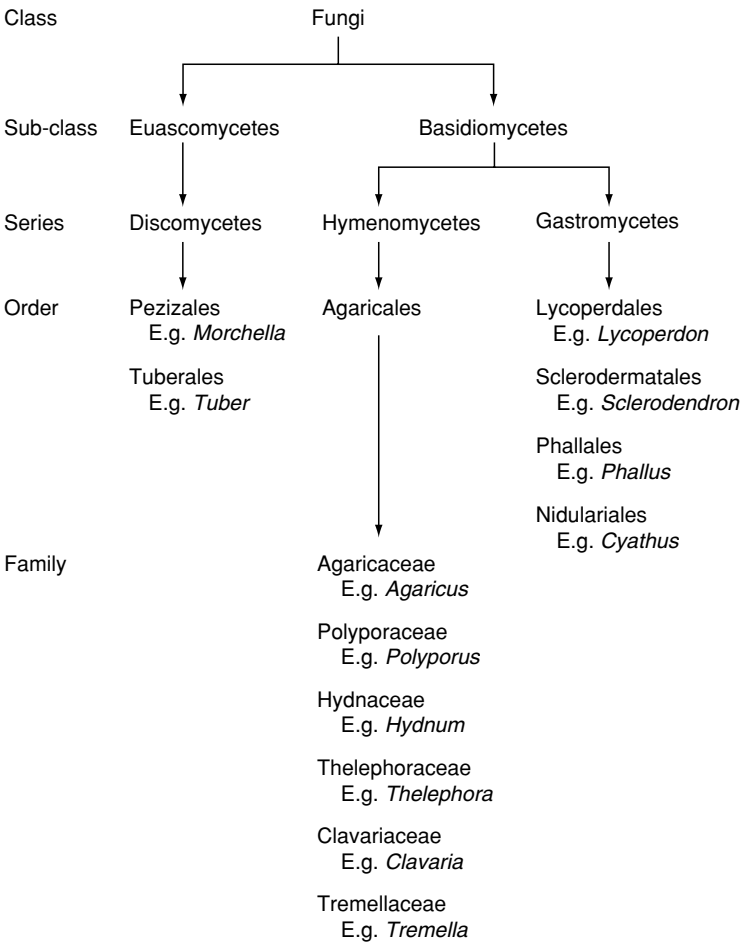

Figure 1 Location of mushrooms and truffles in the classification of fungi.

which can be seen in any dried puff ball, and a cellular part, called the gleba, which is the spore-bearing tissue, composed of minute chambers lined with the hymenium. The peridium breaks in various ways to permit the spores to escape. When children pinch a puff ball to 'see the smoke,' they are doing just what the puff ball would have them do, scattering its seeds to the winds.

Gastromycetes has four orders:

A. Glebal chambers usually not separating from the peridium or from each other;
B. Gleba powdery;
C. Hymenium present in early stages; spores mostly light-colored, small – *Lycoperdales*;
CC. Hymenium lacking or indistinct; spores mostly dark, large – *Sclerodermatales*;
BB. Gleba slimy and fetid; exposed on a receptacle – *Phallales*;
AA. Glebal chambers forming waxy peridioles, or entire gleba separating as a unit from the peridium – *Nidulariales*.

Distribution of Mushroom Species

Mushroom species are distributed throughout the world, some very specific to certain geographic areas, while others are universal and may occur in different seasons in the same region. Temperature plays a critical role in the distribution of mushroom species, and so their distribution can be categorized under tropical, subtropical, and temperate belts (Table 1). The reason for specific distribution of mushroom species in relation to temperature is to be

Table 1 Geographic distribution of various mushroom species

Climatic belt	*Mushroom species*
Tropical	*Volvariella* spp.
Subtropical	*Agaricus bitorquis*
	Pleurotus spp.
Temperate	*Agaricus bisporus*
	Boletus grevillei
	Tricholoma saponaceum
	Laccaria laccata
	Polyporus squamosus

viewed to the specific enzymes secreted and factors influencing these enzymes by the mushroom species under study, in relation to the geographic conditions of natural occurrence.

Morphology

Natural Habitat

Mushroom species can be found growing on dead and decaying plant material, humus soil, e.g., *Agaricus*, as saprophytes on the dead trunks of live trees, e.g., *Pleurotus*, as parasites on the living trees, e.g., *Armillaria*, or characteristically in symbiotic association with the roots of pine trees, e.g., *Morchella*. Species of *Termitomyces* grow on the ant hills, again an example of symbiosis. *Tuber* species grow underground in certain European climates and their flavour is often detectable by dogs.

Vegetative Phase

Mushrooms have two phases in their life cycle: a mycelium that is usually hidden and inconspicuous with a longer growth phase followed by short-lived fruiting bodies. The mycelium derives the nutrition for its growth (which, after a period of establishment, culminates in the production of fruiting bodies) under certain defined favorable conditions. The conditions for mycelial growth are not very specific, in contrast to the very specific growth conditions required by the fruiting bodies. The constraints and problems associated with the latter explain why several mushroom species have yet not been able to be cultivated artificially. Mycelia may be uninucleate (monokaryotic) or binucleate (dikaryotic) (Figure 2), depending on the species and stage in the life cycle. In nature, generally, the spore germinates under favorable conditions to produce the uninucleate mycelium.

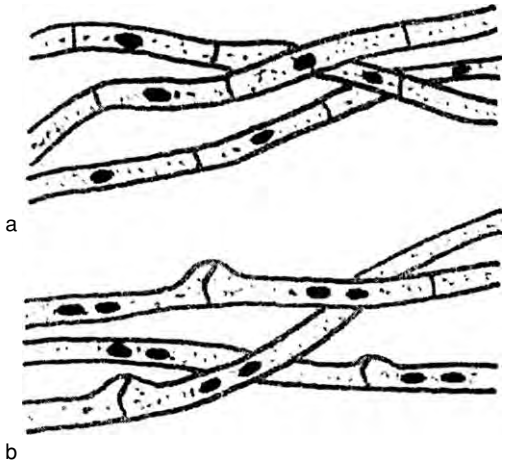

Figure 2 Mono- and dikaryotic mycelia. (a) septate without clamp connections; (b) septate with clamp connections.

Reproductive Phase

After a certain amount of vegetative growth depending on the availability of the compatible strain, the uninucleate mycelium reaches the dikaryotic phase, which in turn leads to the production of fruiting bodies, bearing spores. For clarity, the typical life cycles of three mushroom varieties are described below.

Primary homothallism Here, the basidiospore germinates to produce the mycelium, which in turn is self-fertile and ultimately can result in the production of fruiting bodies, which bear gills, basidia terminating with four basidiospores (per basidium), as a product of meiosis, e.g., *Volvariella volvacea* (Figure 3).

Secondary homothallism *Agaricus bisporus* is an example of this phenomenon, wherein the basidium by name itself bears only two spores, and each spore in turn is binucleate, carrying both the compatible factors in the same spore, so that the mycelium that results from germination of the basidium finally yields to the development of fruiting bodies (Figure 4).

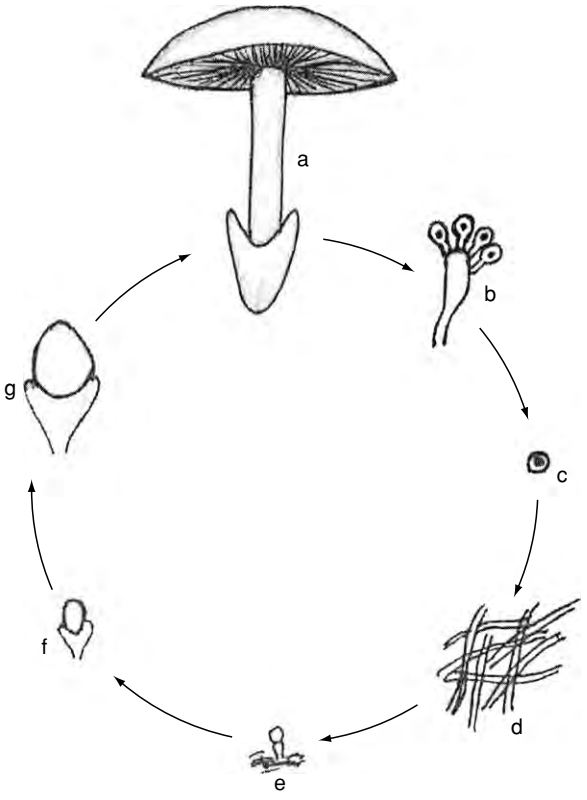

Figure 3 Primary homothallism in *Volvariella volvacea*. (a) fruiting body; (b) basidium; (c) basidiospore; (d) mycelium; (e) fruiting initial; (f and g) developing fruiting body.

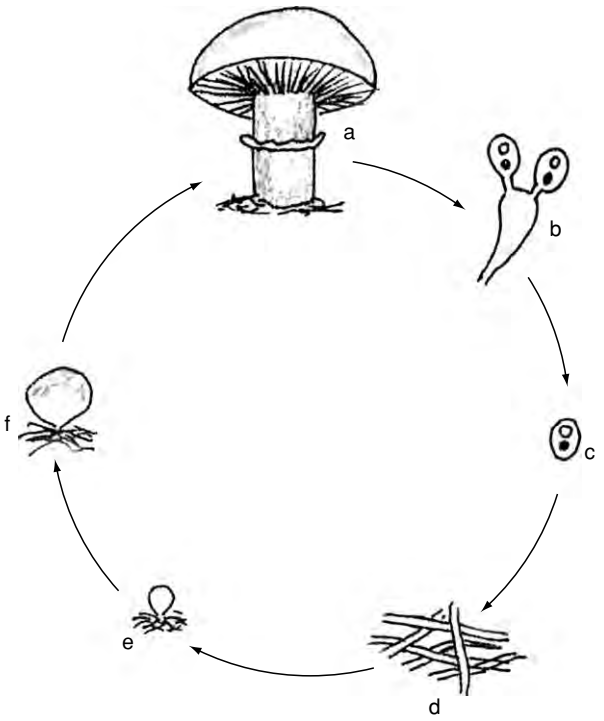

Figure 4 Secondary homothallism in *Agaricus bisporus*. (a) fruiting body; (b) basidium with two basidiospores; (c) basidiospore; (d) mycelium; (e and f) developing fruiting body.

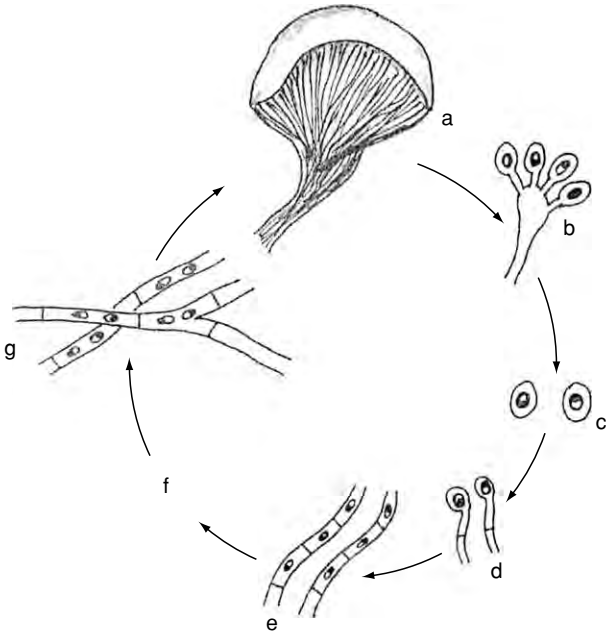

Figure 5 Heterothallic and tetrapolar life cycle in *Pleurotus sajor-caju*. (a) fruiting body; (b) basidium with four genetically different basidiospores; (c) basidiospores; (d) germinating basidiospores; (e) monokaryotic mycelia; (f) dikaryotization; (g) dikaryotic mycelia.

Heterothallism Species of *Pleurotus* are classic examples of heterothallism, in which each basidium bears four basidiospores, each genetically different from the other, so that the degree of inbreeding is reduced to 25%, with increased scope for outbreeding of 75%. This represents possibly the most highly evolved life-cycle system of the basidiomycetes and is strengthened further by the operation of multiple alleles at the two different loci (Figure 5).

Variations in Mushroom Morphology

The fruiting body is the net result of a certain minimum amount of vegetative growth, followed by the onset of conditions (physical, chemical, or biological) favoring reproduction. The fruiting body that develops typically has a stem/stalk called a 'stipe,' with or without a basal cup, called a 'volva,' and a ring on the stem known as an 'annulus' (Figure 6). The occurrence of an annulus/volva is dependent on the characteristic nature of development or morphogenesis of the fruiting body. The display of annulus and volva also varies depending on the species (Table 2). The stipe terminates with a cap, also called a 'pileus' varying in shape, size, texture, structure, and color, and constituting one of the most beautiful objects of nature on the earth. The great diversity in morphology of the mushroom species is depicted in Figure 7. The cap is variously colored, bright, or dull. However, the toxicity of a mushroom bears no relation to its coloration (Table 3). The cap consists of a fertile layer called a 'hymenium' (Figure 8) comprising gills, which in turn bear basidia with basidiospores at the distal end intercalated with sterile cysitidia. The fruiting body may collapse as a result of autodigestion, as in *Coprinus* (inky cap) or may even have latex-containing tissues as in *Lactarius* (milky mushroom).

Orientation of gills The morphology of gills contributes to valuable taxonomic criterion in *Agaricus*, when the gills are open from the genesis to the shedding of basidiospores. Their attachment to the stipe/stem varies and is specific to the species. Gills may be free and detached from the stipe. A mushroom is described as decurrent when the gills run down the stem (the usual situation in many funnel-shaped fruiting bodies), adnate when the gill is attached along its entire width, adnexed when attachment is along only a part of the total width, and sinuate when there is a notch where the gill joins the stem. The attachment of gills in the cap has a bearing on the confluency of the cap, and this interrelation is depicted in the Figure 9.

In Euascomycetes, ascospores are formed inside the club-shaped structures called 'asci,' whereas in

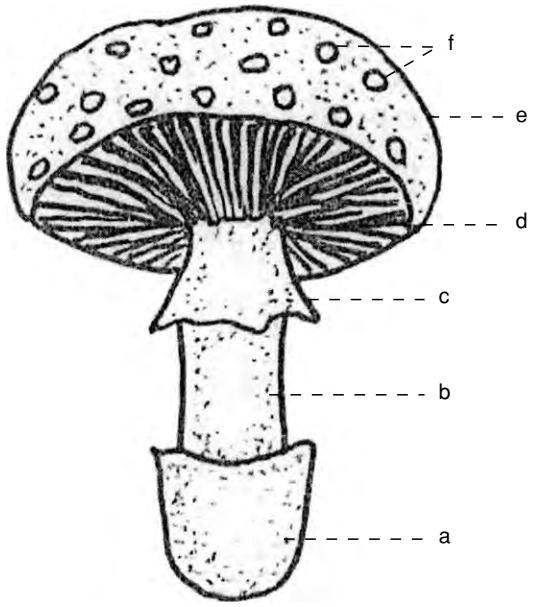

Figure 6 Structure of a typical mushroom. (a), volva; (b), stipe; (c), annulus; (d), gills; (e), pileus; (f), scales (remnants of universal veil).

Table 2 Variations in morphology

Characteristics	*Present (+)/ absent (−)*	*Mushroom species*	*Edible/ poisonous*
Annulus	−	*Volvariella volvacea*	Edible
		Amanita vaginata	Poisonous
Annulus	+	*Agaricus bisporus*	Edible
		Agaricus xanthodermus	Poisonous
Volva	−	*Agaricus bisporus*	Edible
		Scleroderma citrinum	Poisonous
Volva	+	*Volvariella volvacea*	Edible
		Amanita crocera	Inedible
Both	+	*Lepiota procera*	Edible
		Amanita porphyria	Poisonous
Both	−	*Russula virescens*	Edible
		Lactarius helvus	Poisonous
Stipe central		*Agaricus compestris*	
		Russula cyanoxantha	
Stipe excentric		*Pleurotus florida*	
		Polyporus squamosus	
Stipe absent		*Tremella mesenterica*	
		Hirneola auricula-judae	

Basidiomycetes, the basidiospores are formed exogenously on the basidia (Figure 10).

Ornamentation of basidiospores The color of the spore print immediately after collection of a species may help to identify a species. Basidiospores are of various colors, shapes, sizes, and arrangements (Figure 11). In fact, they are one of the most useful features in mushroom taxonomy. They may be hyaline and thin-walled, as in *Agaricus* and *Pleurotus*, set for immediate germination; they cannot carry out the function of surviving to the next season. In some cases, they are thick-walled and variously ornamented to survive to the next season and also result in effective dissemination. Peridioles of *Cyathus* species exemplify a special mode of dissemination, that is effected by a drop of rain water. In fact, basidiospores of various mushroom species are microscopic objects of ornamentation on the earth (Table 4).

Vegetative reproduction Several mushroom species multiply using structures other than asco- or basidiospores (Table 5). Oidia represent thin-walled asexual spores that participate in germination. Chlamydospores are thick-walled spores produced vegetatively from mycelial cells, that resist the unfavorable conditions and germinate under the onset of favorable conditions. Sclerotia are thick-walled, conspicuous mycelial tufts of different sizes. These provide an easier means of producing the fruiting bodies in *Morchella*. In *Pleurotus tuber-regium*, the sclerotium carries out the function of sustaining during unfavorable weather conditions. Thus, mushroom species, as the mostly highly evolved of the fungi, have reminiscent characteristics of primitive fungi, which have served to prolong the life cycle and create alternative means of perpetuating the species.

Importance of Identification and Classification

Identification of mushrooms is a very difficult job in mycology, and the specimens collected, because of their delicacy, should be handled with care. Marking the characteristics of the species *in situ* in nature for shape, size, color, flavor, and latex, if any, and collecting spore print are essential in the field/forest or zone/area of collection. Further identification rests on the spore morphology and, to some extent, on the characteristics of the mycelium. It may be easy in some cases to assign the correct class, order, and family to mushrooms collected in the field, but assigning a genus and species requires more knowledge. If it is a new species, more interest would be evinced in understanding the edibility, life cycle, growth substrate, periodicity of fructification, enzymes secreted, and possibility of artificial culturing and any special feature in reducing environmental pollution by biodegradation of toxic compounds. Thus, the identification and classification of mushroom species may be difficult but sometimes rewarding.

The Royal Botanical Gardens, in Kew, London maintains records of all species discovered to date. Another dimension of natural collection of any mushroom species, its identification and characterization,

(a)

(b)

(c)

(d)

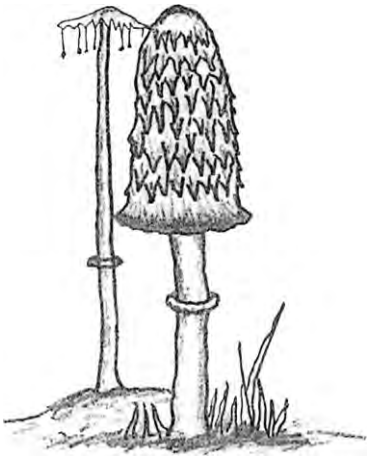

(e)

(f)

(g)

(h)

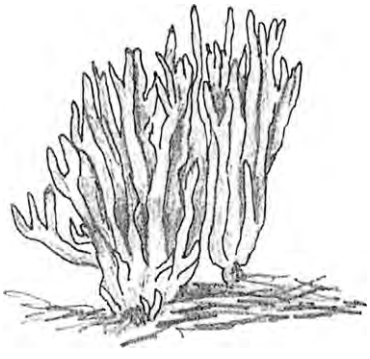

(i)

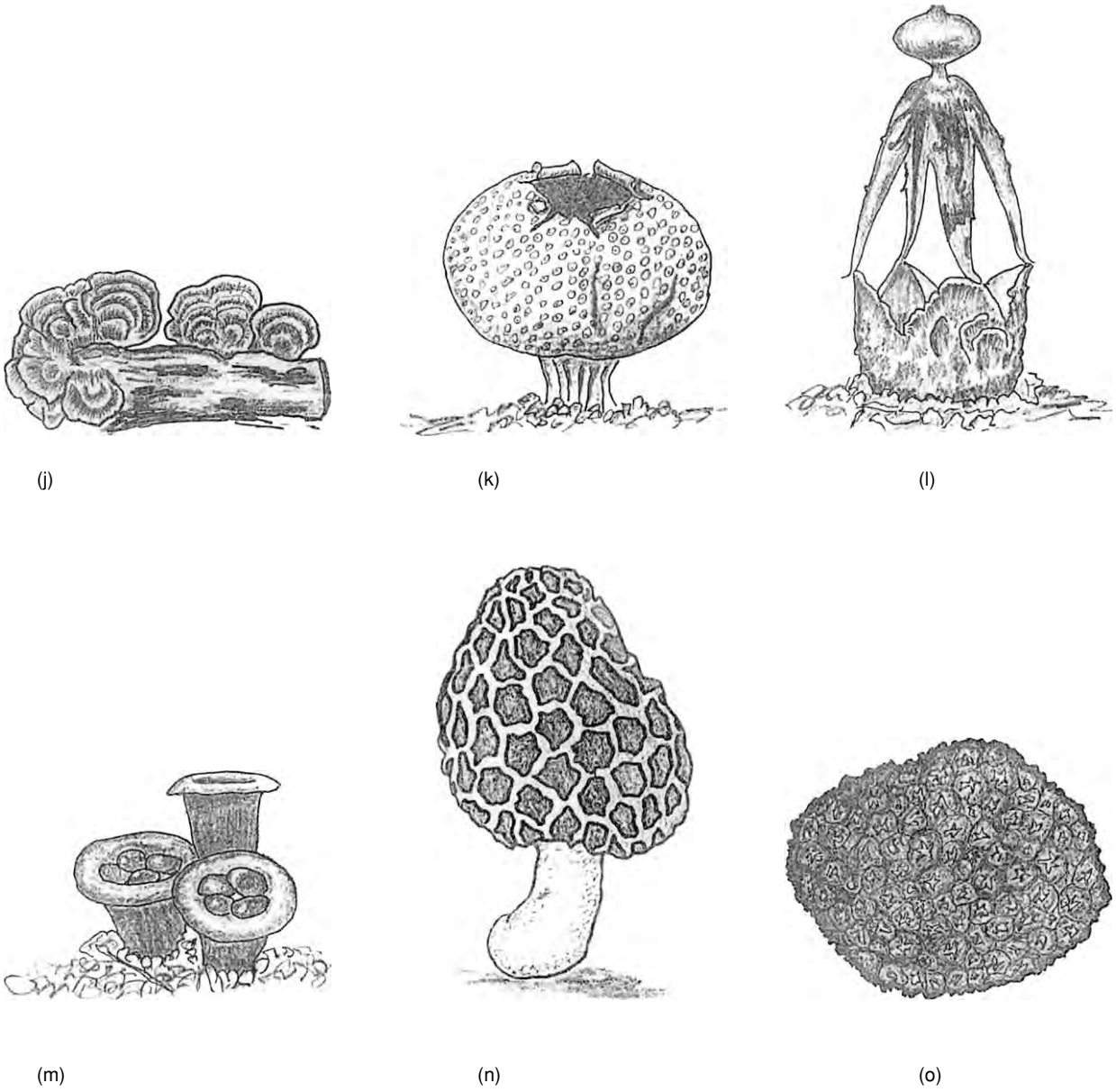

Figure 7 Morphological variations of mushrooms and truffles. (a) *Amanita muscaria*; (b) *Lactarius turpis*; (c) *Russula sardonia*; (d) *Mycena galopus*; (e) *Coprinus comatus*; (f) *Agaricus campestris*; (g) *Tricholoma gambosum*; (h) *Entoloma clypeatum*; (i) *Calocera viscosa*; (j) *Coriolus versicolor*; (k) *Scleroderma citrinum*; (l) *Geastrum fornicatum*; (m) *Cyathus olla*; (n) *Morchella esculenta*; (o) *Tuber melanosporum*.

involves the maintenance of its culture in viable but importantly vigorous. Often, the culture may be viable but cannot yield to fructification on growth. In the American Type Culture Collection in Rockville, USA, identified mushroom cultures can be deposited and given a code number. This has led to a centralized system in which mushroom cultures from any area of the world can be deposited and made available for study, which would not otherwise be possible if species/cultures were lodged only with collectors.

Poisonous Mushrooms

It is very difficult to state whether a mushroom is poisonous or edible. Traditionally, mushroom species have been classed as 'poisonous,' nonfatal but causing stomach upset, termed 'inedible,' and 'edible.'

At present, animal-feeding trials are the best way to identify the poisonous varieties, but some of the characteristics relating to poisonous species are:

- bright-spored, with both annulus and volva – *Amanita* spp.;

Table 3 Coloration of various mushroom species

Color	*Edible*	*Poisonous*
Red/brick red deep reddish brown	*Russula nigricans*	*Amanita muscaria*
Yellow/greenish yellow yellowish red	*Amanita citrina*	*Hypholoma fasiculare*
	Tremella mesenterica	
Light brown/brown/coffee brown	*Polyporus squamosus*	*Scleroderma citrinum*
	Hygrophorus hypothejus	*Paxillus involutus*
Grayish/grayish brown	*Amanita vaginata*	*Amanita porphyria*
	Boletus scaber	
Orange/orange red	*Boletus grevillei*	*Boletus satanas*
	Lactarius deliciosus	
Violet/purple blue	*Lepista nuda*	*Mycena pura*
	Russula cyanoxantha	
Green	*Russula virescens*	*Amanita phalloides*
White	*Clitocybe nebularis*	
	Agaricus bisporus	

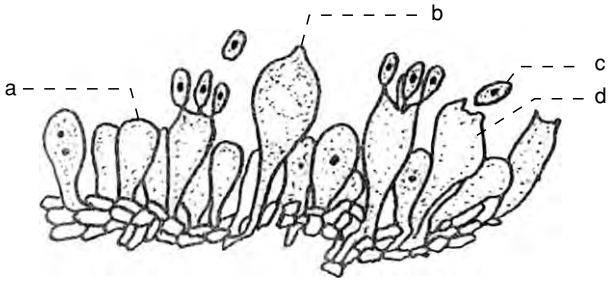

Figure 8 Structure of the hymenium. (a) paraphysis; (b) cystidium; (c) basidiospore; (d) basidium.

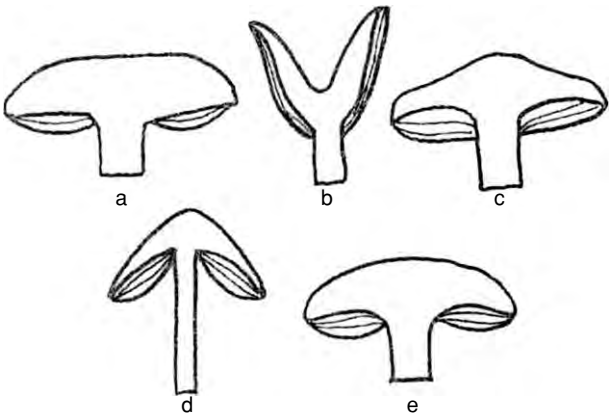

Figure 9 Orientation of gills in mushrooms. (a) free gills with shallow convex cap; (b) decurrent gills with funnel-shaped cap; (c) adnate gills with campanulate cap; (d) adnexed gills with conico-campanulate cap; (e) sinuate gills for cap with incurved margin.

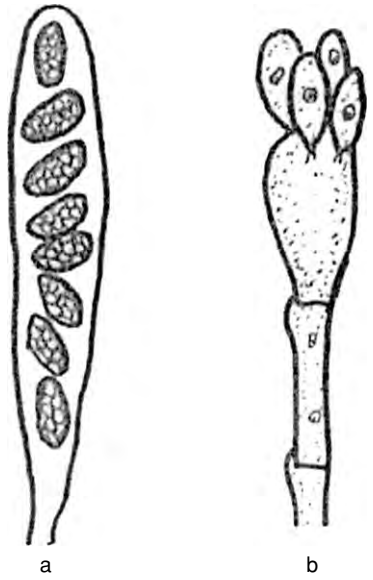

Figure 10 Ascus and basidium. (a) ascus with ascospores; (b) basidium with basidiospores.

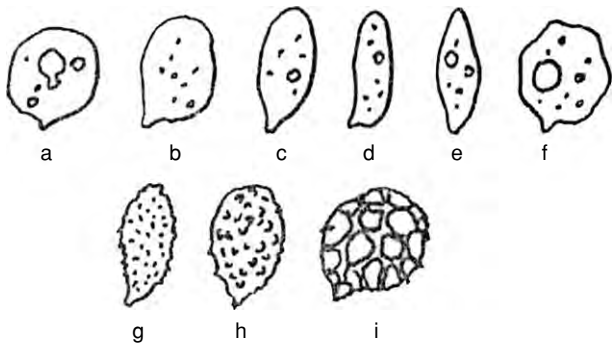

Figure 11 Shape and ornamentation of the basidiospores. (a) globular; (b) oval; (c) ellipsoid; (d) cylindrical; (e) fusoid; (f) angular; (g) echinulate; (h) warted; (i) reticulate.

Table 4 Morphology of spores of various mushroom species

Shape	*Mushroom species*
Globular	*Helvella elastica, Pseudohydnum gelatinosum*
Oval	*Paxillus involutus, Pholiota squamosa*
Ellipsoid	*Collybia dryophila*
Cylindrical	*Polyporus squamosus, Calocera viscosa*
Fusoid	*Inonotus hispidus*
Angular	*Entoloma sinuatum*
Echinulate	*Laccaria laccata, Cortinarius mucosus*
Warted	*Peziza badioconfusa, Ramaria aurea*
Reticulate	*Lactarius volemus, Lactarius helvus*

Table 5 Forms of vegetative reproduction

Types	*Mushroom species*
Oidia	*Pleurotus cystidiosus*
Chlamydospores	*Agaricus* spp.
	Volvariella spp.
Sclerotium	*Morchella esculenta*
	Pleurotus tuber-regium

- oozing yellow latex on rupture of the surface – *Psalliota xanthoderma*;
- spore print turning greenish (copper compounds) – *Lepiota morgani*;
- luminescent – *Clitocybe illudens*.

See also: **Mushrooms and Truffles**: Use of Wild Mushrooms; **Mycotoxins**: Classifications; Occurrence and Determination; Toxicology

Further Reading

Alexopoulos CJ (1961) *Introductory Mycology*. London: John Wiley.

Chang ST and Quimio TH (eds) (1982) *Tropical Mushrooms – Biological Nature and Cultivation Methods*. Hong Kong: Chinese University Press.

Chang ST, Buswell JA and Miles PG (eds) (1993) *Genetics and Breeding of Edible Mushrooms*. Philadelphia, USA: Gordon & Breach.

Christensen CM (1969) *Common Edible Mushrooms*. Minneapolis, USA: The University of Minnesota Press.

Elliott TJ (1985) The general biology of the mushroom. In: Flegg PB, Spencer DM and Wood DA (eds) *The Biology and Technology of the Cultivated Mushroom*. New York: John Wiley.

Hard ME (1976) *Mushrooms, Edible and Otherwise*. New York: Dover.

Largent DL (1973) *How to Identify Mushrooms to Genus. 1: Macroscopic Features*. Eureka, CA: Mad River Press.

Rajarathnam S and Zakia B (1987) *Pleurotus* mushrooms. Part IA. Morphology. life cycle, taxonomy, breeding, and cultivation. *CRC Critical Reviews in Food Science and Nutrition* 26(2): 157–223.

Rajarathnam S, Shashirekha MN and Zakia B (1992) Biopotentialities of the Basidiomacromycetes. *Advances in Applied Microbiology* 32: 234–361.

Ramsbottom J (1989) *Mushrooms and Toadstools*. London: Bloomsbury Books.

Singer R (1961) *Mushrooms and Truffles*. New York: Interscience.

Svrček M (1975) *Mushrooms and Fungi*. London: Octopus Books.

Wilkinson J and Buczacki S (1982) *Mushrooms and Toadstools*. Glasgow: Harper Collins.

Use of Wild Mushrooms

S Rajarathnam and M N Sashirekha, Central Food Technological Research Institute, Mysore, India

Background

Mushrooms were regarded by the early civilizations of Egypt and Rome as a special delicacy and were perceived by the Romans as the food of the gods. A clear indication of the high esteem in which mushrooms were held by different cultures is provided by the practice of restricting the use and consumption of many species to the rulers of the time. The medicinal and analeptic qualities of mushrooms, the scientific basis for which we are only just beginning to understand, have long been appreciated by the Chinese. Legend states that Chinese emperors consumed *Lentinus edodes* in large quantities to fend off old age. Other accounts tell us that the ancient Japanese courts valued these mushrooms for their aphrodisiac properties. 'The growing sites were well hidden and heavily guarded.' One species of mushroom, *Ganoderma*, is cultivated today strictly for its medicinal benefits. It is now realized that mushrooms have an important role to play in the recycling of organic wastes, thereby relieving environmental pollution.

Useful Chemical Constituents

Mushrooms are known for their characteristic biting texture and pleasant aroma. These properties have derived the meaning for mushrooms as items of food delicacy. In recent years, it has been well proven and documented in the world literature that mushrooms do provide definite nutrition and health benefits for humans. Significant compositional changes occur with a strain as a result of differences in age

or stage of development, lapse of time after harvest, study of different portions of a single fruiting body, the inaccuracies inherent in the methods of analysis, and the relative precision of the analyst. In order to insure universal comparison of the analytical values of different workers, it is worth expressing all the analytical data on dry weight of the fruiting bodies, since it is well known that mushrooms in general contain about 90% moisture in fresh form. Mushrooms on a dry-weight basis contain ~63% carbohydrates, 25% protein, 4% fat, and the remaining 8% is accounted by minerals represented by ash (Table 1). Most carbohydrates are in the polymeric form, glucan and hemicellulose types; starch as such is absent. Mannitol and trehalose represent the bulk of free sugars. Accordingly, the energy value is also relatively less. Mushrooms do contain considerable amounts of dietary fiber.

Protein represents the major fraction next to carbohydrates. The fraction 4.38 is used to convert nitrogen into mushroom protein (instead of the general factor 6.25), excluding nonprotein nitrogen in the form of chitin present in the cell walls. Albumins and globulins predominate in the protein fraction of *Boletus edulis* and *Cantharaellus cibarius*. Free amino acids account for nearly one-fifth of the total nitrogen. Protein is made up of most of the essential amino acids; however, some essential sulfur-containing and aromatic amino acids are lacking. Based on the amino acid analysis, several indices are used to predict the nutritional quality of the mushroom.

Mushrooms are relatively low in fat content. The fat includes representatives of all classes of lipid compounds, including free fatty acids, mono-, di-, and triglycerides, sterols, sterol esters, and phospholipids. There is a preponderance of unsaturated fatty acids, in particular linoleic acid (~78% of the total fatty acids). It is the degree of unsaturation of fatty acids in mushrooms that reflects on their safety in consumption without any danger of atherosclerosis.

The ash is mainly composed of calcium, potassium, phosphorus, and ferrous iron (Table 2). Some species also contain germanium, which is known for maintaining vitality in humans. Mushrooms have a special feature of bioaccumulation of minerals available in their growth medium. This property is useful in providing desired minerals in good quantities. When they accumulate toxic elements, they become dangerous for consumption.

Table 1 Composition[a] of mushrooms

Mushroom species	*Crude protein (N × 4.38)*	*Fat*	*Carbohydrate*	*Fiber*	*Ash*	*Energy value (kcal)*
Agaricus bisporus	26.3	1.8	59.9	10.4	12.0	328
Auricularia auricula-judae	8.1	1.5	81.0	6.9	9.4	356
Boletus edulis	29.7	3.1	59.7	8.0	7.5	362
Cantharellus cibarius	21.5	5.0	64.9	11.2	8.6	353
Coprinus comatus	25.4	3.3	58.8	7.3	12.5	366
Flammulina velutipes	17.6	1.9	73.1	3.7	7.4	378
Lentinus edodes	17.5	8.0	67.5	8.0	7.0	387
Lycoperdon lilacinum	46.0	7.5	38.8	12.3	7.7	358
Pholiota nameko	20.8	4.2	66.7	6.3	8.3	372
Pleurotus florida	18.9	1.7	58.0	11.5	9.3	265
Termitomyces microcarpus	27.4	4.3	54.2	2.2	14.1	364
Tricholoma species	16.7	3.1	71.9	12.9	8.3	342
Volvariella esculenta	34.4	20.6	31.7	11.2	13.3	396

[a]On a dry-weight basis.

Table 2 Mineral content[a] of mushrooms

Mushroom species	*Ca (mg 100 g^{-1})*	*P (mg 100 g^{-1})*	*K (mg 100 g^{-1})*	*Fe (ppm)*
Pleurotus florida	24	1850	4660	184
Agaricus campestris	23	1429	4762	186
Volvariella diplasia	58	1042	3333	177
Lentinus edodes	118	650	1246	30.3
Flammulina velutipes	19	278	2981	11
Pholiota nameko	42	771	2083	22
Auricularia polytricha	287	Trace	47	
Tricholoma matsutake	26	640	41	417

[a]On a dry-weight basis.

Table 3 Vitamin contents of mushrooms

Mushroom species	*Ascorbic acid*[a]	*Thiamin*[a]	*Niacin*[a]	*Riboflavin*[a]	*Pantothenic acid*[a]	*Folic acid*[b]
Pleurotus florida	113	1.36	72.9	7.88	29.4	1412
Agaricus bisporus	82	1.14	56.19	4.95	22.8	933
Auricularia auricula-judae		0.16	4.10	0.48		
Lentinus edodes		0.40	11.90	0.90		
Volvariella volvacea		0.32	59.5	2.73		
Flammulina velutipes	46.3	6.10		5.2	106.5	
Pholiota nameko		18.8	72.9	14.6		

[a]mg $100\,g^{-1}$ dry mushrooms.
[b]μg $100\,g^{-1}$ dry mushrooms.

Table 4 Vitamins contributed by 100 g fresh *Pleurotus* mushrooms as a percentage of the daily requirement recommended by the Food and Agriculture Organization/World Health Organization (FAO/WHO)[a] per person

	Ascorbic acid	*Thiamin*	*Niacin*	*Riboflavin*	*Folic acid*
Adult man	24–43	10–15	26–33	33–39	53–60
	(30)	(1.2)	(19.8)	(1.8)	(0.2)
Adult women	24–43	13–19	36–46	46–54	53–60
	(30)	(0.9)	(14.5)	(1.3)	(0.2)

[a]*Report of a Joint FAO/WHO Expert Group* (1974) Rome: FAO.
Note: Figures in parentheses are values in mg recommended intake by FAO/WHO per person per day.

Mushrooms are a good source of several vitamins. Species of *Pleurotus* are estimated to contain predominantly vitamins of B-complex and folic acid. Relative to *Auricularia, Lentinus*, and *Volvariella*, the thiamin, niacin, and riboflavin content of *Pleurotus* species were higher (Table 3). Considering the amount of vitamins contributed by 100 g fresh *Pleurotus* mushrooms as a percentage of the requirement level indicated by the Food and Agriculture Organization/World Health Organization (FAO/WHO) per person per day and allowing for differences in moisture contents, *Pleurotus* mushrooms are a particularly good source for meeting the human requirements for riboflavin and folic acid (Table 4).

Mushrooms are known for their B-complex vitamins (niacin, thiamin, and B_{12}) and folic acid. Their ability to contain these vitamins, even during their growth on lignocellulosic wastes, eventually substantiates their biosynthetic capacities. In fact, folate synthetase and B_{12} synthetase enzyme systems have been demonstrated in mushroom cells.

Food Value

Determination or prediction of the nutritional value of mushrooms (Table 5) has been made possible based on their content of essential amino acids.

The amino acid score is the amount of the most limiting amino acid in the food protein expressed as a percentage of that amino acid present in a reference protein.

$$\text{Amino acid score} = \frac{\text{mg of amino acid in 1 g test protein}}{\text{mg of amino acid in 1 g reference protein}} \times 100$$

The essential amino acid (EAA) index evaluates the quality of dietary protein in terms of the ratio of the EAA contained in a food, relative to the EAA content of the reference protein, mostly egg.

$$\text{EAA} = \sqrt[n]{\frac{\text{Lysine}_p - \text{tryptophan}_p - \text{histidine}_p}{\text{Lysine}_s - \text{tryptophan}_s - \text{histidine}_s}}$$

where p = food protein, s = standard (egg), and n = the number of amino acids.

The biological value is a measure of the nitrogen retained by the body after consuming the test protein.

$$\text{Biological value} = 10.9\ (\text{EAA index}) - 11.7$$

In an attempt to resolve the difficulties inherent in comparisons between those mushrooms containing small amounts of high-quality protein with those containing larger amounts of a protein of lesser nutritional quality, the use of a nutritional index has been proposed:

$$\text{Nutritional index} = \frac{\text{EAA index} \times \text{percentage protein}}{100}$$

Amino acid scores and EAA indices of the most nutritive mushrooms (highest values) rank in potential nutritive value with those of meat and milk, and are significantly higher than those for most legumes and vegetables. The least nutritive mushrooms rank appreciably lower but are still comparable to some common/nutritious vegetables.

As Flavor Ingredients

In general, mushroom flavor substances are derivatives of fatty acids and nucleotides. Flavor is the most

Table 5 Estimated nutritive values of mushrooms

Mushroom species	*Essential amino acid index*	*Biological value*	*Nutritional index*	*Amino acid score*[a]
Agaricus bisporus	55.8	49.1	17.0	36
Boletus edulis	76.6	71.8	9.3	37
Cantharellus cibarius	86.2	82.3	3.0	68
Lentinus edodes	55.8	49.1	9.8	40
Pleurotus florida	84.5	80.4	15.9	67
Russula vesca	88.9	85.2	6.0	70
Termitomyces microcarpus	74.7	69.7	20.5	45
Volvariella displasia	87.6	84.1	25.1	71

[a]Using egg as reference protein.

important inducement for past and present widespread consumption of wild and commercially grown edible mushrooms. As many as 150 volatile compounds have been identified in various mushroom species. A series of eight carbon (C_8) compounds are believed to be the most important volatile flavor compounds. It has been demonstrated that some of the C_8 and C_{10} compounds can be enzymatically formed from linoleic and linolenic acids, both of which are usually predominant in fruiting bodies. The gills were found to be the major compartment for the metabolic interconversion of fatty acids into aromatic compounds. 1-Octen-3-ol is the general fresh mushroom flavor, whereas 1-Octen-3-one represents the dried-mushroom flavor. Several alcohols, aldehydes, ketones, and phenols are known to be mushroom flavor components characteristic of each species.

Umami, the full, rounded taste (as defined by the Japanese) of a mushroom, is predominantly the result of the presence of monosodium glutamate, further improved with addition of minor amounts of nucleotides, such as guanidine monophosphate (GMP) and adenosine monophosphate (AMP). Tricholomic acid and ibotenic acid have been found in *Tricholoma muscarium* and *Amanita pantherina*, respectively. These amino acids have higher threshold values than that of glutamic acid in evoking the umami flavor. It is of great interest that these two amino acids are toxic to flies. A Japanese proverb states that the *Tricholoma matsutake* is noted for its aroma and *Lyophyllum shimeji* is noted for its taste. The aroma components of *T. matsutake* are matsutakeol and methyl cinnamate. A sulfur-containing compound, lenthinic acid, in fresh *Lentinus edodes* is flavorless, but it is converted by heat to lenthionin, which is responsible for the specific flavor, when *L. edodes* is dried.

As Source of Enzymes

Mushrooms are involved in the rotting of wood. Studies have been performed on many enzymes produced by cultured mushroom mycelia internally or externally. Mushrooms (basidiomycetes and a part of ascomycetes) are generally divided into lignin-hydrolyzing fungi (oxidoreductase producers) and cellulose-hydrolyzing fungi (hydrolase producers). In addition to phenol oxidase and cellulase, there are many reports on other mushroom enzymes, including protease, pectinase, chitinase, hemicellulase, amylase, and various glycosidases.

As Indicators of Toxic Elements

Mushrooms have a natural tendency to bioaccumulate the minerals from their growth substrates. *Agaricus campestris* finds use as a bioindicator species to trace the concentrations of silver. *Phanerochaetae chrysosporium* was most effective in sorption of lead compared with species of *Polyporus, Volvariella*, and *Pleurotus*. *Coprinus* species is shown to be able to methylate mercury after the transformation of mercury from the soil humus. The ability of several mushroom species to bioaccumulate cadmium, zinc, lead, copper, nickle, and manganese in their fruiting bodies is employed to monitor the intensity of environmental pollution. Several toxic metals are selectively sorbed from industrial waste waters by mushrooms. *Inonatus nikadol* is used to recover uranium selectively from the culture medium.

Biological Uses

Many mushrooms, such as *Ganoderma lucidum, Coriolus versicolor, Hericium erinaceus, Armillaria mellea, Marasmius androsaceus, Tremella fuciformis*, and *Lentinus edodes*, have traditionally been used as medicine and tonics in China, Korea, and Japan (Tables 6 and 7). Terpenoids, substances which inhibit growth of cancer cells such as He La cells *in vitro* (chemotherapeutic agents), have been found in *Ganoderma lucidum*. Steroids in *Coriolus versicolor* and *Agaricus blazei*, novel γ-pyrones, esinapyrones A and B, cytotoxic novel phenols hericenones A and B,

and novel fatty acid in *Hericium eriaceus*, as well as lampterol (illudin-S) in *Lampteromyces japonicus*, represent other inhibitors of growth of cancer cells. The germanium content of mushroom is involved in carcinostatic and analgesic functions. Many polysaccharides exhibiting remarkable antitumor activity *in vivo* have been isolated from various species of mushrooms belonging to Polyporaceae, Tricholomataceae, and Agaricaceae through screening studies using sarcoma-180 in mice (Table 8). These antitumor substances are regarded as a kind of biological response modifier (BRM) – immunotherapeutic agents, whose activities are based on the activation of immunological functions in a host.

Biological properties of several of the natural mushroom species due to the components present in them are presented in Table 9. This information reflects on the diverse manifestations of medicinal properties of these mushrooms.

'Mushroom nutraceuticals' represent the medicinal preparations from mushrooms. They are functional foods which are enriched or modified and consumed as normal diet to provide health-giving benefits. They are refined or partially defined extractives which are consumed in the form of capsules or tablets, as a dietary supplement (not a food) and which has potential therapeutic applications, increasing resistance to diseases and causing regression in disease state. Nutraceuticals are not to be confused with pharmaceuticals which are defined chemical preparations, specifications for which are listed in a pharmacopoeia, and are mainly used therapeutically; pharmaceuticals have high potency, so the chances of overdose and toxicity require regulation by the authorities. Mushrooms with little scope of toxicity and overdose do not require very strict regulation and can be sold as 'over-the-counter' medicines.

In the recent past, a variety of medicinal preparations in the form of tablets, capsules, and extracts from mushrooms have been produced and marketed. In 1991, the value of the world mushroom crop was estimated at $8.5 billion and in the same year $1.2 billion were estimated to have been generated from medicinal products of mushrooms. In 1987 PSK, a top selling anticancer drug from the mushroom *Coriolus versicolor*, had annual sales in Japan totalling $358 million and accounted for 25.5% of the country's total sales of anticancer drugs. It was not without reason that production of *Grifola, Lentinula*, and

Table 6 Use of mushrooms as drugs in China

Mushroom species	*Indication*
Agaricus bisporus	Stimulating digestion, curing hypertension
Auricularia polytricha	Strengthening health, helping blood circulation
Boletus edulis	Causing muscles and joints to relax
Coriolus versicolor	Curing chronic disease
Ganoderma lucidum	Rejuvenating effect, neurasthenia
Grifola frondosa	Strengthening health and resisting diseases
Hericium erinaceus	Beneficial to heart, kidneys, liver, lungs, spleen
Lentinus edodes	Strengthening health and resisting diseases
Pleurotus ostreatus	Causing muscles and joints to relax
Tremella fuciformis	Strength, health, and resisting diseases
Tricholoma mongolicum	Beneficial to stomach and intestine

Table 8 Antitumor activities of water extracts of mushroom species (against sarcoma – 180 in mice)

Mushroom species	*Tumor inhibition (%)*
Polyporaceae	
Ganoderma tsugae	77
Coriolus versicolor	77
Trametes gibbosa	49
Fomes fomentarios	5
Tricholomataceae	
Lentinus edodes	80
Flammulina velutipes	81
Pleurotus ostreatus	75
Tricholoma matsutake	91
Strophariaceae	
Pholiota nameko	86
Agaricaceae	
Agaricus bisporus	2
Auriculariaceae	
Auricularia auricula-judae	42

Table 7 Medicinal effects of mushrooms

Japanese name	*Mushroom*	*Medicinal effects*
Shiitake	*Lentinus edodes*	Antiinflammation, antitumorigenic, gastroenteropathy, headache, vertigo, hepatocirrhosis, arteriosclerosis
Fukurotake	*Volvariella volvacea*	Hypotensive, immunopotentiator, reduce cholesterol
Bukuryou	*Poria cocos*	Diuresis, edema, vertigo, dipsia, anxiety, neurosis, insomnia, antitumorigenic
Shirokikurage	*Tremella fuciformis*	Cough, sore throat, stethalgia, constipation, paramenia
Yamabushitake	*Hericium erinaceus*	Indigestion, neurasthenia, gastric ulcer, antitumorigenic
Tohtyukasoh	*Cordyceps sinensis*	Mild roborant, expectorant, antitussive, antiinflammation

Hypsizygus – three prominent medicinal mushrooms – registered a 102, 110, and 141% increase respectively during 1990–94.

Toxic Components in Poisonous Mushrooms

Mushroom poisoning accounts for approximately 70% of natural poisoning and often causes death. However, there are only 30–50 poisonous species among the thousands of species found on earth, and of these, no more than 10 are fatally poisonous. Nevertheless, many cases of mushroom poisoning are reported every year. There is no convenient way to distinguish edible mushrooms from poisonous ones. All the folklore about detecting poisonous mushrooms is not true. For example, poisonous mushrooms do not blacken silver spoons when cooked with mushrooms, as folklore claims. The best way to avoid risk is to learn the physical features of each species. If 10 of the most poisonous mushrooms were recognized by mushroom pickers, the number of cases of mushroom poisoning would decrease dramatically.

Amanita phalloides, the green death cap, is known as the most dangerous and poisonous mushroom. It is widely distributed in Europe and North America and is responsible for 90–95% of fatal mushroom poisonings. Cholera-like symptoms – nausea, vomiting, and diarrhea – of *Amanita* poisoning begin 10–20 h after ingestion; severe damage to the liver and kidney follows, and this damage finally leads to death. The cyclic peptides, the amatoxins, and the phallatoxins have been isolated as the toxic principles. *A. virosa* (Table 10), the 'destroying angel,' is as toxic as *A. phalloides* because it contains an amatoxin, amaninamide.

Amino acids structurally related to glutamic acid have been obtained as mushroom toxins. Some of them made significant contributions to neurobiological research because of the biological and clinical importance of glutamic acid. Acromelic acids isolated from *Clitocybe acromelalga* are neurotoxins showing highly potent activity as glutamate agonists. This mushroom, found only in Japan, has been known to induce unique symptoms. If ingested, it causes severe pain and reddish edema in the fingers and toes, which is similar to acromelalgia and erythromelalgia, and the symptoms persist for a month or so.

Amanita muscaria has a brilliant red cap flecked with white spots and is undoubtedly the best-known poisonous mushroom in the world. It is widely distributed throughout the world, except for tropical zones. The mushrooms belonging to the genus *Inocybe* contain far greater amounts of muscarine than do the *A. muscaria*, so that ingestion of *Inocybe* mushrooms induces typical symptoms of muscarine poisoning. Psilicybin and psilocin are well-known mushroom toxins with hallucinogenic effects detected in the majority of *Psilocybe* species. These alkaloids act on the central nervous system because the structure and activity are similar to those of the known hallucinogen, lysergic acid diethylamide (LSD).

Thus, mushrooms constitute a unique class of foods praised and priced because of their characteristic biting texture and pleasant aroma, and are aptly regarded as a food delicacy. They serve to complement and supplement the human diet with various chemical constituents not encountered in or deficient in food items of plant or animal origin. They have a tremendous potential to synthesize components and constituents not found in the nitrogen-poor, low-cost, lignocellulosic growth substrates. They can

Table 9 Biological properties of mushrooms

Mushroom species	*Compound*	*Biological property*
Agaricus blazei	Ergosterol derivatives	Uterus cervical cancer, He La and hepatoma cells
Armillariella tabescens	Armillarisia A	Increase secretion of bile
Coriolus versicolor	Ergosterol derivatives	Uterus cervical cancer, He La and hepatoma cells
Flammulina velutipes	Protein	Cardiotonic
Ganoderma lucidum	Adenosin, guanosin	Antithrombolytic
Ganoderma lucidum	Organic germanium	Pain reducer during final stages of cancer
Ganoderma lucidum	Two polysaccharide – protein complexes	Hypoglycemic activity
Ganoderma lucidum	Triterpenoids	Hypotensive
Ganoderma lucidum	Heterogalactan – protein complex	Antiinflammatory and antiallergic
Hericium erinaceus	Hericenones	Synthesis of nerve growth factor
Lentinus edodes	Adenine derivatives	Hypocholesterolemic
Lentinus edodes	RNA from the spores	Antiviral (influenza infection)
Marasmius androsaceus	Marasmic acid	Analgesic, sedative effect
Polyporus confluens	Polysaccharide	Peptic ulcers
Volvariella volvacea	Protein	Cardiotonic
Various mushroom species	Lectins	Agglutination of erythrocytes/precipitation of polysaccharide
Various mushroom species	Eritadenine, grifolin, neogrifolin	Reduction in cholesterol/neutral fat

Table 10 Mushroom poisoning

Mushroom species	*Compounds*	*Nature of compound*	*Symptoms*
Amanita phalloides	Phallotoxin		
Amanita phalloides	Amatoxin	Cyclopeptide	Fatal
Amanita virosa	Virotoxin		
Gyromitra esculenta	Gyromitrin		Affects autonomic nervous system
Coprinus atramentarius	Coprine	Amino acid	Drunken sickness
Clitocybe acromelalga	Clitidine		
	Acromelic acid	Novel amino acids	Neurotoxin
	Clithioneine		
Amanita muscaria	Muscarine	Amino acid	Affects parasympathetic, cholinergic nervous system
Inocybe patouillardii			
Clitocybe acromelalga			
Amanita pantherina	Mycoatropine		
Amanita muscaria		Alkaloids	Psychotropic poisoning
Amanita regalis			
Psilocybe species	Psilocybin		
Panaeolus species			
Panaeolina species			
Stropharia species			
Cortinarius orellanus	Orellanin	Polypeptide	Fatal
Ramaria formosa	Emodin		Laxative effects

be prepared in diverse recipes that can provide great satisfaction to human sensorials. Their unique chemical composition encourages the fight against diabetes, alimentary ulcers, tumors, blood cholesterol, and hypertension. As low-calorie foods, they are dietary items ideal for slimming. With the examples of mushroom species mentioned above, the chemical constituents described and their biological properties, it is clear that mushrooms, in addition to being an item of food delicacy, can claim value as 'health conditioner.' Probably within another decade, more research will further define more quantitatively the biomedicinal values of mushrooms.

See also: **Enzymes**: Functions and Characteristics; **Mushrooms and Truffles**: Classification and Morphology; **Mycotoxins**: Classifications; Occurrence and Determination; Toxicology; **Protein**: Food Sources; **Vitamins**: Overview

Further Reading

Chang H and Butt P (1986) *Pharmacology and Application of Chinese Materia Medica*, pp. 642–663. Singapore: World Scientific Publishing.

Chang ST and Hayes WA (eds) (1978) *The Biology and Cultivation of Edible Mushrooms*. London: Academic Press.

Chang ST and Quimio TH (eds) (1982) *Tropical Mushrooms – Biological Nature and Cultivation Methods*. Hong Kong: Chinese University Press.

Chang ST, Buswell JA and Miles PG (eds) (1993) *Genetics and Breeding of Edible Mushrooms*. USA: Gordon and Breach.

Chen AW and Miles PG (1996) Biomedical research and the application of mushroom nutriceuticals from Gandoderma lucidum. In: Royse DJ (ed.) *Mushroom Biology and Mushroom Products*, pp. 161–175. Pennsylvania: Penn State University.

Chihara G (1990) Lentinan and its related polysaccharides as host defence potentiator: their application to infectious diseases and cancer. In: Masihi KN and Lange W (eds) *Immunotherapeutic Prospects of Infectious Diseases*, pp. 9–18. Heidelberg: Springer Verlag.

Mizuno T (1995) Bioactive biomolecules of mushrooms: food function and medicinal effect of mushroom fungi. In: Teranishi R and Honstein I (eds) *Food Reviews International*, pp. 7–21. USA: Marcel Dekker.

Rajarathnam S and Zakia Bano (1988) *Pleurotus* mushrooms: part II. Chemical composition, nutritional value post-harvest physiology, preservation and role as human food. In: Clydesdale FM (ed.) *CRC Critical Reviews in Food Science and Nutrition*, Vol 27(2), 87–158. Boca Raton, FL: CRC Press.

Rajarathnam S and Zakia Bano (1991) Biological utilization of edible fruiting fungi. In: Arora DK, Mukerji KG and Marth EH (eds) *Handbook of Applied Mycology: Foods and Feeds*, vol. 3, pp. 241–292. New York: Marcel Dekker.

Rajarathnam S, Shashirekha MN and Zakia Bano (1992) Biopotentialities of the Basidiomacromycetes. In: *Advances in Applied Microbiology*, Vol. 37. USA: Academic Press.

Rajarathnam S, Shashirekha MN and Zakia Bano (1998) Biodegradative and biosynthetic capacities of mushrooms: present and future strategies. In: Stewart GF and Russell I (eds) *Critical Reviews in Biotechnology*, vol. 18, pp. 91–236. Boca Rator FL: CRC Press.

Rajarathnam S, Zakia Bano and Shashirekha MN (1999) Mushrooms: an item of food delicacy or health conditioner. In: Rajak RC (ed.) *Microbial Biotechnology for Sustainable Development and Productivity*, pp. 140–151. Jodhpur (India), Scientific Publishers.

Mussels *See* **Shellfish**: Characteristics of Crustacea; Commercially Important Crustacea; Characteristics of Molluscs; Commercially Important Molluscs; Contamination and Spoilage of Molluscs and Crustaceans; Aquaculture of Commercially Important Molluscs and Crustaceans

MUSTARD AND CONDIMENT PRODUCTS

J S Hemingway, Brundall, UK

Introduction

The term 'mustard' is believed to be derived from the use of seeds as condiment, the sweet 'must' of old wine being mixed with crushed seeds into a paste – 'hot must' or 'mustum ardens'. Mustard seed has been known as a spice since the earliest recorded times; it is described in Sanksrit and Sumerian texts back to 3000 BC, Egyptian to 2000 BC, and Chinese before 1000 BC; seed fragments exist from the Indus Valley civilizations around 2000 BC and Egyptian tombs at Thebes of the same period. Records of mustard being traded as a manufactured condiment extend back to the days of the Avignon popes and Shakespeare.

In the present day, it is by far the largest volume spice in international trade, totalling over 170 000 $kg \times 10^3$ per year. Other particular features which make it unusual as a spice crop are that, for condiment purposes, the entire world supply is grown in temperate regions, as the crop needs long days; hence, its production is virtually all concentrated in the Northern Hemisphere, where it has necessarily been developed by plant breeding to suit full mechanization of every operation of its cropping and the bulk handling of mustard seed into trade (a further unusual aspect of a spice material).

Cropping

For use as spice, the predominant production is in North America, in the Canadian prairie provinces of Saskatchewan, Alberta, and Manitoba, and just southwards into Montana and North Dakota, USA. Annual crops of the order of 200 000 $m \times 10^3$ in those areas produce total quantities in excess of 150 000 $kg \times 10^3$ of export value US\$40 $\times 10^6$. Other important production centers are the Hungarian plain and the East Anglian counties of the UK, also Poland, the Czech Republic, and Slovakia. Smaller-scale cropping occurs in Argentina and Australasia for local use, but world manufacture effectively depends on the single annual crop of the Northern Hemisphere. The main manufacturing usage is in the USA, Canada, France, Germany, Japan, and the UK.

Botanically, two principal species of the Cruciferae family are involved: 'White' (or 'yellow') mustard, *Sinapis alba*; 'brown' and 'oriental' mustards, *Brassica juncea*. In condiments, *S. alba* contributes a hot (in terms of mouth feel) nonvolatile principle and *B. juncea* a pungent (i.e., olfactory) volatile principle, each from the hydrolysis of their respective glycosides. Until the 1950s, seed of *B. nigra* was almost the sole source of pungency, but with perhaps unparalleled rapidity in any industry, let alone in an area as traditional as the spice trade, *B. juncea* replaced *B. nigra* over a single decade. This change met the requirement for full mechanical harvesting, as the older type needed to be cut by hand.

Mustard species are also of great importance as oilseed crops, particularly *B. juncea* in the Indian subcontinent, China, and the Sarepta area of the southern Ukraine. In recent years, there has developed an increasing export trade from Canada to the Indian subcontinent, principally Bangladesh, of the order of 50 000 $kg \times 10^3$ per year. In all these countries, the growth is closely similar to that of rapeseeed. Mustards also find wide usage as salads, green manure, and fodder crops and as leaf and stem vegetables in the Far East.

The principal centers of plant breeding and agronomic research on mustard crops have long been Norwich, UK, and Saskatoon, Canada, where successions of varieties of improved habit, yield, and quality aspects (e.g., seed size and shape, lower oil contents, and higher essential principles, kernel and seed-coat mucilage contents, all being features of manufacturing importance) have been produced.

Mustard crops for condiment are all grown as spring-sown annuals. They show three marked periods of growth:

1. A rapid vegetative development of leaves and the branching framework.

2. Three to four weeks of characteristic intensive yellow flowering during June.
3. Six to eight weeks of seed filling and ripening.

The crops reach a full height of 1.5–2.0 m, with each plant branching to bear several racemes of pods, all in a clear 'podding zone' that forms the upper part of the crop canopy. The main differences between the two principal species are visible during growth: *S. alba* foliage and stem are bright green in color, with surface hairs on the hollow stems; *B. juncea* is more glaucous, with smooth, waxy, pith-filled stems. The pods of *S. alba* are 2–3 cm in length, with a pronounced 'beak' at the tip and with hairy rough surfaces, containing up to eight seeds, whereas pods of *B. juncea* are up to 5 cm in length, smooth and tubular, with up to 20 seeds.

Mustard crops are both easy and reliable to grow, and perform well on land of reasonable fertility, on which they need only nitrogen as an added fertilizer. *Brassica juncea* has considerable resistance to drought and can crop with limited rainfall but *S. alba* flourishes best with more regular precipitation. As neither species is affected by plant diseases to economic levels, and both suffer to only a limited degree from insect pests, they do not require repeated treatments from pesticides. The requirement for clean seed necessitates adequate weed control practices, for which well-proven and simple systems are available. The crops can tolerate reasonable extremes of weather, such as spring frosts, hail, or strong winds, without serious losses, standing well on stiff, erect stems. In cropping systems in which rotational practice is important, mustards form an excellent 'break crop' between cereals, especially as a preparation for wheat. As mentioned earlier, modern varieties have been developed that suit farming machinery used for other crops, and do not clash with them in the timing of harvesting, so that there is no need for special investment or any hand-labor operations.

Mustards are prolific yielders and can increase over 2000-fold in a season in optimum conditions. Current mean yields in UK cropping are of the order of 900–1000 kg of seed per acre (2220–2500 kg per $m \times 10^3$), ranging up to 1500 kg (3700 kg); in the North American prairie, expectations are lower, at some 400–700 kg per acre (990–1730 kg per $m \times 10^3$) for *B. juncea* and 275–550 kg per acre (680–1360 kg per $m \times 10^3$) for *S. alba*, with substantial variation according to soil moisture status. Yields in eastern European countries are intermediate.

Seed Handling and Storage

After harvest in Canada, the seed is usually below 10.5% moisture, at which level, it is safe for long-term storage; indeed, if thoroughly dry and clean, there is never any risk of infestation of mustard by stored product pests. In other countries, the seed may need to be dried gently to reduce it to safe storage levels; in such situations, it is important to act quickly, as the seed can deteriorate very rapidly, and to restrict the temperature of the drying airflow so that the seed temperature never exceeds 52 °C; otherwise, damage to its enzymes will result.

Cleaning operations are quite straightforward with mustards' round seeds of uniform size. Indented cylinders, perforated screens, and spiral separators are all effective, and surface dust can be removed by aspirated brush-conveyors. The presence of green or weathered seeds and of specific inseparable weed seeds is the main cause of downgrading within the official grades of the Canadian Board of Grain Commissioners, under which system the preponderance of world trading is operated. Storage is normally in bulk bins, with pneumatic or mechanical conveying, and carriage in bulk lorries or rail grain-cars, then bulk shipping holds. Smaller quantities are shipped in standard freight containers, occasionally in 100-lb (45-kg) bags if the customer so requires.

Anatomy of the Seed

The seed of *S. alba* is larger than that of *B. juncea*, averaging 148 000 seeds per kg, mean diameter 2.22 mm, against 408 000 and 1.63 mm. Features of the fine structure of the outside of the testa and the degree and type of surface reticulation and pitting are characteristic and of importance in differentiating the seed from that of other brassicas, e.g., rapeseed.

The seed coat consists of four or five layers of cells, including an outer layer of large cells and an inner palisade layer. Inside this lies the true endosperm, with an outer layer of cells that possess thickened walls and contain fat globules and small aleurone grains, and an inner, very narrow layer of compressed cells with no evident structure. The center of the seed is filled with two hemispherical cotyledons with the embryo lying between. There is only a small hilum and micropyle. The testa of *S. alba* contains a large number of mucilage glands opening to the surface; *B. juncea* has many fewer.

In *B. juncea*, the two distinct seed types are 'brown' and 'oriental.' The former has a darkly colored endosperm of more substantial cells, but such color is absent from the latter, leaving a golden-yellow shade. In this 'oriental' type, the testa occupies only 12–15% of the seed by weight, against 19–22% in brown; hence, the oriental type has a higher kernel content, of substantial economic value, provided that the testa is sufficiently robust to stand up to mechanical handling without rupture.

Seed Composition

General specifications for mustard seed and standard analytical methods are detailed under ISO 1237–1981 (E) UDS 633.844.004.1. The gross composition is shown in Table 1, the variation in content of common fatty acids in the oils of *S. alba* and *B. juncea* is shown in Table 2, and the principal amino acids in the protein fractions of these two species are shown in Table 3. (*See* **Amino Acids**: Properties and Occurrence; **Fatty Acids**: Properties.)

The organoleptic properties of mustards result from their contents of the essential mustard oils, which are colorless, irritating liquids; that of *S. alba* has little aroma, as it is nonvolatile, but a hot taste, whereas *B. juncea* gives a volatile oil with a pungent aroma. Within the seed, they exist as glycosides within intact cells of the cotyledons, and it is only on the rupture of the cells and in the presence of adequate moisture that the enzyme myrosinase can catalyze the reactions to produce the isothiocyanate principles.

In *S. alba*, the storage glycoside is sinalbin, which on breakdown gives *p*-hydroxybenzyl isothiocyanate; in *B. juncea*, the glycoside sinigin gives allyl-isothiocyanate. (Certain strains of *B. juncea*, all originating in the Indian subcontinent, give rise to a mixture of allyl and 3-butenyl isothiocyanates, the latter being an undesirable flavor, at least to the Western palate.)

Table 1 Gross composition of mustard seed

	Percentage of seed (dry weight)
Neutral oil	24–45 (*Sinapis alba* 25–30, *Brassica juncea* 35–45)
Polar lipids	6–12
Protein	20–30
Carbohydrate	12–18
Glycoside	1–3
Phytins	2–3
Water[a]	8–12

[a]Normal commercial mustard seed as traded on an international basis.
Source: Vose JS (1972) *A Review of the Chemistry and Enzymology of Mustard Seeds*. Montreal: Reckitt & Colman (Canada) Ltd.

Table 2 Ranges of variation in content of common fatty acids in the oils within each species

	Percentage of oil	
	Sinapis alba	Brassica juncea
Palmitic	2–3	2–4
Oleic	16–28	7–22
Linoleic	7–10	12–24
Linolenic	9–12	10–15
Eicosenoic	6–11	6–14
Erucic	33–51	18–49

Variations under genetic control have been identified but are not presently utilized as they appear to affect product quality.
Source: Vose JS (1972) *A Review of the Chemistry and Enzymology of Mustard Seeds*. Montreal: Reckitt & Colman (Canada) Ltd.

Table 3 Principal amino acids in the protein fractions of *Sinapis alba* and *Brassica juncea* seeds

	Amino acid content (mg per gram of nitrogen)	
	Sinapis alba	Brassica juncea
Lysine	362	335
Methionine	97	104
Cystine	124	159
Isoleucine	207	236
Leucine	412	395
Phenylalanine	233	240
Tyrosine	206	167
Threonine	171	251

Processing and Food Uses

The availability of these distinct flavor effects, heat and pungency, from the two mustard species allows the manufacture of a wide range of condiment products and food ingredients. A variety of processing techniques are used:

- wet milling of whole seeds;
- wet milling with husk separation;
- dry milling with husk separation;
- dry grinding of whole seeds;
- dry grinding after oil pressing.

A French (e.g., 'Dijon') mustard uses only brown-colored *B. juncea*, which, because it is wet-milled, loses much of its initial pungency. The extent of removal of the husk fragments by screening after milling governs the final color of the paste. Such mustards are commonly blended with wines and vinegars, and whole or split mustard grains, herbs, or other spices may be added. (*See* **Milling**: Characteristics of Milled Products.)

In contrast, an American 'cream-salad' mustard is made by wet milling only *S. alba* seed with water in a colloid-type mill. The viscosity of the product is greatly aided by the presence of mucilage in the seed coats. Turmeric is usually added to overlay the natural pale color to a rich yellow. The product has a sweet, tangy flavor and can be used in large amounts, almost as a ketchup.

Traditional German mustard is similarly made from *S. alba* alone, and is thus nonpungent. The texture of the paste can be varied by the degree of grinding, and admixtures of herbs, spices, and water or vinegar are used according to the product.

In contrast, traditional English mustard is a blend of the dry-milled flours of both *S. alba* and *B. juncea* and is both hot and pungent. This product is sold as a

dry fine powder, with no reaction until blended with water for the table, when the characteristic flavors develop very rapidly. However, as the pungency is volatile, and the hotness of the essential principle from *S. alba* breaks down within a day, the made condiment loses strength quite quickly. Its strength can only be maintained in a bottled paste ('hot English mustard') by the inclusion of fruit acids in its manufacture from milled flours.

The dry milling process to produce flour is broadly similar to the milling of wheat, with the difference that mustard seeds are very small and have 30–40% of oil in their cells. This needs great care, as the oil causes enormous problems if released by excess pressure or temperature. *Sinapis alba* and *B. juncea* are milled separately as they have distinct seed sizes. The seeds are first thoroughly cleaned, graded for size and shape, blended, and then dried to a very low level of moisture content. Their husks are then moistened (so that the kernel is dry and friable but the husk slightly elastic and less brittle) and metered steadily on to the 'break rolls,' of fluted surface, on which the seed is nipped to crack open the husk. This is separated from the kernels by sieving and currents of air in the 'purifiers': a remarkably clean separation can be achieved with little kernel adhering to the husks, and minimal presence of fine husk particles, which would show as specks in the flours milled from the kernel fractions.

The kernels then pass on to a series of smooth 'reduction rolls,' which progressively reduce them to a powder. After each pass through a roll, the fine particles are sieved off to ultrafine by the use of swirling sieves, their frames covered with silk bolting cloths of differing meshes; the coarse particles return to a further reduction roll. The resultant flours – the *S. alba* a deep creamy color and *B. juncea* a rich golden-yellow shade – are blended in varying proportions according to the product desired. Some products may be blended with stabilized wheat flour in small proportions to attain a standard flavor strength, and coloring may be enriched by the curcumin present in ground turmeric.

In Japan, the production of mustard powder is usually by the fine grinding and sifting of whole seeds, but after the extraction of some of the seed-oil content by pressing. The Japanese taste, like the Anglo-Saxon, is for very strongly flavored mustard products.

The grinding of whole mustard seeds is usually to obtain a coarser, granular product for use as an ingredient in sauces, pickles, or chutneys. An increasing usage of ground *S. alba* is in the North American meat products industry, where it adds flavor and protein, and its mucilage helps to absorb moisture. Some *S. alba* seed is now being heat-treated to inactivate its myrosinase, then pin-milled to coarse particles of no strength of flavor, thus allowing a higher and wider usage in such meat products.

The byproducts of dry milling are principally the husks. That of *S. alba* can be ground into a coarse meal used to impart viscosity to paste mustards or sauce products, but the husk of *B. juncea* has little value.

Beyond the ingredient usages of mustard already mentioned, based on its range of flavors and applications of the mucilage content of *S. alba*, the major usage is the inclusion of mustard flour – predominantly of *S. alba* – in the formulation of salad creams and mayonnaises. It provides some flavor and natural color, has natural bacteriostatic and antioxidant properties, and, most importantly, acts as an emulsifying agent, helping to bind the oil and fat phases of the recipe. This was perhaps a fundamental virtue within the original popularity of mustard as a spice, technically to 'enhance flavor' – or disguise the taste of salted meats or fish – but also to 'aid digestion' by its use with fatty meats or oily fish, attributes praised by early medical writers or gourmands. (*See* **Antioxidants:** Natural Antioxidants; **Emulsifiers:** Uses in Processed Foods.)

See also: **Amino Acids**: Properties and Occurrence; **Antioxidants**: Natural Antioxidants; **Emulsifiers**: Uses in Processed Foods; **Fatty Acids**: Properties; **Milling**: Characteristics of Milled Products

Further Reading

Duffus C and Slaughter C (1980) *Seeds and their Uses*. Chichester, UK: Wiley.

Hemingway JS (1995) Mustards. In: Smartt J and Simmonds NW (eds) *Evolution of Crop Plants*. London: Longman.

Hemingway JS (1995) The mustard species: condiments and food ingredients, and potential as oilseed crops. In: Kimber DS and McGregor DI (eds) *Brassica Oilseeds, Production and Utilisation*. Wallingford, UK: CAB International.

Man R and Weir R (1988) *The Compleat Mustard*. London: Constable.

Rosengarten F (1969) *The Book of Spices*. Wynnewood, PA: Livingston.

Tsunoda S, Hirata K and Gomez-Campo C (1980) *Brassica Crops and Wild Allies*. Tokyo: Japan Scientific Societies Press.

Vaughan JG and Hemingway JS (1959) The utilisation of mustards. *Economic Botany* 13(3): 196–204.

Vose JS (1972) *A Review of the Chemistry and Enzymology of Mustard Seeds*. Montreal: Reckitt & Colman (Canada).

MUTAGENS

T J Schrader, Sir Frederick Banting Research Centre, Health Canada, Ottawa, Ontario, Canada

Background

Mutagens are chemical compounds or forms of radiation (such as ultraviolet (UV) light or X-rays) that cause irreversible and heritable changes (mutations) in the cellular genetic material, deoxyribonucleic acid (DNA). Mutagenic lesions persist when they escape detection by protective cellular DNA repair mechanisms, when mistakes occur in the repair process, or when repair mechanisms are overwhelmed by extensive damage. Upon subsequent cellular replication, these mutations become fixed in the genome and are inherited by all daughter cells. In this way, mutagenesis becomes a cumulative process, stretching over the lifetime of an organism.

The biological consequences of a mutation depend upon many critical factors such as the target loci, size of the mutation, timing during the cell cycle, and compounding effects of preexisting mutations. Thus, a mutagenic event occurring in a nonfunctional area of DNA will have no effect (*silent mutation*), whereas a similar change in an actively transcribed region may profoundly affect gene expression and phenotype or even lead to cell death (*lethal mutation*). The influence of mutations in human health is underscored by several human disease states caused by mutations that disrupt regulatory regions or gene coding sequences, resulting in altered gene expression and protein function. For example, mutations in genes that promote or inhibit growth and cellular replication (*protooncogenes* and *tumor suppressor genes*, respectively) or code for components of DNA repair pathways are important contributors to the multistage development of cancer. In addition, the accumulation of mutations over time, leading to gradually less efficient cellular repair capabilities, has been linked to the aging process and associated degenerative diseases. While the health of a particular individual can be affected by mutations in somatic cells, mutagenic events in germ cells (*germline mutations*) lead to the transmission of genetic diseases to subsequent generations. Genetic predispositions to certain breast and colon cancers, cystic fibrosis, and Huntington's disease are included among many examples of such inheritable diseases.

Although many dietary and environmental agents have been classified as mutagens, cells are constantly subjected to a barrage of spontaneous DNA damage. The spontaneous hydrolysis of DNA bases from the sugar–phosphate backbone, constant exposure of cellular DNA to oxygen-derived free radicals and low rates of miscopying during DNA replication all contribute towards the mutational burden. In addition to these sources, significant exposure to mutagenic compounds can occur through food and water, as well as through environmental and occupational sources. Many natural constituents of food are mutagenic and are produced by plants as defense agents. However, additional foodborne mutagens can be present as residues of compounds used during food production or leached from packaging materials. Mutagenic compounds can also be produced during food cooking and preparation. Foods also contain many compounds that can modulate the activity of mutagens. In order to minimize exposure to mutagens, the majority of developed countries have instituted regulatory protocols governing the introduction of new foods and food-associated chemicals before they are accepted into the marketplace. With advances in the understanding of mutagenic processes, existing foods and food contact items may also be reassessed for acceptability.

Molecular Mechanisms of Mutagenesis

DNA Damage

DNA is a double-helical structure composed of paired heterocyclic nitrogenous purine and pyrimidine bases attached to a backbone of deoxyribose and phosphoric acid. The purine bases (guanine and adenine) and pyrimidine bases (cytosine and thymine) form specific base pairs (guanine/cytosine, adenine/thymine) stabilized by hydrogen bonding. The specificity of base-pairing and sequence of bases along the DNA double helix forms the foundation of the genetic code, and any change can give rise to mutations, expressed as alterations in gene expression or in protein structure and function.

Spontaneous damage Instability of various parts of the DNA molecule in aqueous solution leads to spontaneous DNA damage and mutations. In addition to base deamination, the bonds between purines or pyrimidines and deoxyribose can spontaneously hydrolyze, creating apurinic/apyrimidinic (*AP*) sites, leaving the sugar–phosphate backbone susceptible to strand breakage. Also, the bases can exist in a number of slightly different (*tautomeric*) forms, several of

which lead to mispairing because of altered hydrogen bonding characteristics.

Chemical adducts DNA also comes under assault from a number of dietary and environmental factors that interact with several relatively reactive sites within the structure of DNA, leading to mutagenic events. Many chemical mutagens are electrophiles that are reactive, electron deficient species able to form covalent adducts with nucleophilic sites within DNA. Binding of chemical adducts to the purine/pyrimidine bases can stabilize alternative tautomeric structures or otherwise alter their structural and hydrogen bonding characteristics. These modifications not only alter the base pair coding properties and cause mistakes to be made during replication and transcription, but also can lead to increased base hydrolysis and AP site formation. In addition, some chemical agents produce inter- and intrastrand cross-linking of DNA, which prevents strand separation and causes particular difficulties for replication, transcription, and repair processes.

Oxidative damage Other mutagens can act through the induction of oxidative stress and cause oxygen or nitrogen radical formation. Radicals are highly reactive compounds that contain unpaired electrons and include hydroperoxides, hydroxyl, and superoxide radicals. Transition metal ions such as iron and copper can catalyze the formation of free radicals (*Fenton reaction*). The interaction of free radicals with DNA results in single- and double-strand breaks, hydroxylated derivatives of bases and several other lesions.

DNA intercalation Finally, some mutagens do not actually react with DNA but modify its structure by intercalating between the complementary strands of DNA, disrupting hydrogen bonding between base pairs. The presence of such compounds causes misreading during replication and transcription, leading to the generation of mutations.

Metabolic Activation

Some mutagenic compounds (*direct-acting*) are able to interact with DNA without any further modifications; however, many other compounds (*indirect-acting* mutagens), become mutagenic as unwanted byproducts of transformations carried out by cellular detoxification pathways. Several hydroxylation (*phase I*) and conjugation (*phase II*) pathways are available for detoxification to increase the water solubility of chemical compounds and aid in their excretion. Although the enzyme systems are generally most highly expressed in the liver, other tissue types maintain these capabilities to varying degrees. The phase I reactions are carried out by the cytochrome P450 enzyme system, which consists of a large family of related enzymes in which each isozyme shows specificity for particular chemical structures. During phase II reactions, more polar groups can be added by glutathione-*S*-transferase, glucuronide transferase, microsomal epoxide hydrase, or acetyltransferase. Generally, phase II reactions inactivate mutagens, but instances of the opposite outcome have been found. Several other enzymes, including flavin monooxygenase and prostaglandin H synthetase, are also known to activate mutagens.

DNA Repair

Several mechanisms have evolved to remove lesions in DNA and to restore the structural and coding properties of the base sequence in the damaged region.

Direct Reversal Direct reversal of damage is perhaps the simplest process, whereby a single gene product can repair a particular form of damage. In mammalian cells, mutagenic O^6-methyguanine residues are removed by a damage-reversal mechanism in which the O^6-alkyl group is removed by an O^6-methylguanine transferase in a 'suicide reaction,' resulting in inactivation of the enzyme. A second form of damage reversal is observed in the repair of simple strand breaks, which can be repaired by ligation in a process catalyzed by DNA ligase. Finally, a DNA purine insertase activity has been described by which apurinic sites in DNA are repaired by the direct insertion of purines.

Excision Repair The major and most versatile mechanism used by mammalian cells to remove altered bases or nucleotides within DNA is known as *excision repair*. Excision repair proceeds in several successive steps in a multiple enzyme process. Three major excision repair pathways using different sets of enzymes have been described: base excision repair, nucleotide excision repair and mismatch repair.

Classes of Mutations

Base excision repair removes misincorporated bases such as uracil as well as most methylated base damage. The majority of base damage is repaired by the replacement of a single damaged nucleotide with its normal counterpart, but base excision repair can also result in the synthesis of two to 10 nucleotide 'repair patches.' Initially, the damaged base is recognized and removed by a damage-specific glycosylase leaving an AP site in the DNA. For single nucleotide repair, the deoxyribose-phosphate residue in the DNA backbone is removed and a replacement nucleotide

inserted by activities associated with DNA polymerase β. Longer patches result when either polymerase β or another repair DNA polymerase synthesizes a stretch of DNA. In this process, a flap region of single stranded DNA is displaced and is then degraded by a deoxyribonuclease activity. Both types of base excision repair are completed when DNA ligase seals the final nick remaining in the DNA backbone.

The *nucleotide excision repair* pathway involves 20–30 gene products and is responsible for removing damage caused by agents that introduce bulky adducts into DNA. The initial recognition of DNA damage by a recognition complex recruits the entire multicomponent repair enzyme complex (*exinuclease*) to the lesion. The exinuclease nicks the damaged strand on either side of the damage site, releasing an oligonucleotide of approximately 30 base pairs in length. After the gap is filled in by a DNA polymerase using the opposite strand as a template, the nick at the end of the repair patch is sealed by ligase. Nucleotide excision repair has also been shown to be closely linked with transcription (*transcription-coupled repair*) by which actively transcribed genes are preferentially repaired compared to silent regions of the genome. A component of the nucleotide excision repair complex (TFIID) is shared with the complex involved in gene transcription and mediates the recruitment of DNA repair machinery to the damage site when RNA polymerase stalls at a lesion.

Mismatch repair is a form of excision repair that corrects mismatched base pairs resulting from replication errors and heteroduplexes formed during recombination. The interaction of various components involved in mismatch repair is less understood but results in repair patches of 100–1000 nucleotides. Mutations in mismatch repair genes have been linked to an increased risk for inherited forms of colon cancer.

Mutations can range in size from single base pair changes to the loss or duplication of entire chromosomes. Many factors, such as the mutagenic agent, particular region of DNA affected, preexisting damage, and timing during the cell cycle can influence the magnitude of a mutation. Mutations can be classified into four broad groups: extragenic, gene, chromosomal, and genomic mutations.

Extragenic Mutations Extragenic mutations occur outside of the gene coding sequence and are usually silent. However, if the mutation occurs in an important regulatory sequence, gene expression can be affected.

Gene Mutations Gene mutations occur within the coding sequence of a gene. *Point mutations* involve the substitution of bases, causing either transitions (the substitution of a purine for purine, or pyrimidine for pyrimidine) or transversions (the substitution of a purine for a pyrimidine or vice versa). If such a mutation occurs in the wobble position of a codon, the chance of a change occurring in the translated protein structure or function will be much reduced. However, a single base pair change elsewhere in the sequence can result in proteins with decreased or modified functions and can also affect protein synthesis and degradation. *Frameshift mutations* occur when bases are either lost or inserted, thus resulting in a shift of the reading frame by the transcriptional machinery. Frameshift mutations can result in the production of modified or inactive proteins and may be lethal to the cell. *Block mutations* occur when large sections of a gene are deleted, inverted, or duplicated, and can lead to defective proteins being produced.

Chromosomal Mutations Compared with gene mutations, chromosomal mutations represent a class of mutations that affect larger regions of DNA within chromosomes and may involve the deletion, translocation, duplication, or further amplification of whole genes. Chromosomal mutations may lead to either increased or decreased gene expression and can also result in the abnormal expression of normally silent parts of the genome.

Genomic Mutations Genomic mutations involve the duplication or loss of entire chromosomes. Such mutations can have far-reaching effects on gene expression and survival. Typically, the progression of a cancer is marked by cells that have developed progressively altered complements of chromosomes.

Mutagenicity Assays

Many assay systems have been developed to examine mutagenicity and induction of DNA damage (*See* **Carcinogens**: Carcinogenicity Tests). In many instances, mutagenicity assays are used as shorter term, more economical surrogates for the 2-year rat carcinogenesis bioassay. Although a high correlation exists between mutagenicity and carcinogenicity, neither this correlation nor that between results obtained with different mutagenicity tests is perfect. Therefore, test chemicals are generally subjected to a battery of three tests or more to gain a more balanced assessment. Widely used mutagenicity assays include both short-term bacterial, fungal, and mammalian cell-based (*in vitro*) and animal-based (*in vivo*) assay systems. Mutations within a target gene are detected as a loss of function (*forward mutations*) or as a reversal of the effect of a previous mutation (*reverse*

mutations). These assays generally rely upon a change in phenotype as an endpoint and therefore may miss some silent mutations. *In vitro* systems are generally conducted in the presence and absence of a source of metabolic activation, which can include rat liver homogenate (S9 microsomes) or liver cells. The application of molecular biology and biotechnology to many of these systems has led to further refinement and increased molecular characterization of the mutations induced.

The Ames *Salmonella* Test and the *E. coli* tryptophan assay are bacterial systems that measure reversions induced at genes responsible for the synthesis of histidine and tryptophan, respectively. Growth of the bacteria on minimal plates requires back mutations at the indicated loci, and therefore, test chemical mutagenicity is assessed according to colony formation. The Ames *Salmonella* Test has been widely accepted and validated as a bacterial assay with different strains available that respond to base pair and frameshift mutagens as well as oxidative damage. Various refinements to the Ames *Salmonella* Test have been developed, including a series of bacterial strains that are reverted at specific base pairs, allowing the spectrum of induced mutations to be assessed, as well as bacteria that express various mammalian Phase I and Phase II genes to examine the effects of particular types of activation. Several targets can be examined for the induction of forward mutations in mammalian cells, including gene loci encoding thymidine kinase in L51278Y mouse lymphoma cells, hypoxanthine guanine phosphoribosyl transferase (*HGPRT*) in Chinese hamster ovary or V-79 cells and Na^+/K^+ ATPase in V-79 cells. Mutagenesis at the HGPRT locus has also been examined in human lymphocytes and fibroblasts. Plasmid shuttle vectors have also been developed, which can be exposed to the test chemical, introduced into mammalian cells for repair, and then recovered and reintroduced into an indicator strain of *E. coli* for scoring of mutations. If needed, the plasmids can be recovered again and sequenced.

Germ-line mutations may be assessed in rodents using the specific locus test, dominant lethal assay, and heritable translocation test. Transgenic mice incorporating λ phage *LacI* or *LacZ* gene sequences have also been developed, which provide shorter-term, *in vivo* mutagenicity assays. DNA isolated from these animals is incorporated into λ-phage which is used to infect the proper *E. coli* host, forming either clear or colored plaques in the bacterial lawn. The transgenic assays allow mutations to be examined in all somatic tissues as well as germ-line tissues, and provide a well-characterized target for sequence analysis.

Chromosomal aberrations can be examined both *in vitro* using cell lines or *in vivo* using animal tissues, generally bone marrow. Stained metaphase chromosome spreads are examined microscopically for structural chromosomal aberrations. A newer variation of this technique, fluorescent *in situ* hybridization, uses fluorescent labeled DNA probes specific for particular chromosomal sequences to test for aberrations. Another chromosomal aberration assay is the micronucleus test using bone marrow and peripheral blood erythrocytes or lymphocytes. Micronuclei are small particles containing acentric fragments of chromosomes or entire chromosomes that are left behind in the cytoplasm after mutagen treatment. Micronuclei formation may be enumerated by microscopy or flow cytometry.

Sister chromatid exchange uses differential labeling techniques to measure the crossing over between each half of a replicated chromosome. The biological significance of sister chromatid exchange is not known, as the phenomenon does not appear to be related to the mechanisms involved in chromosomal aberration formation and is not directly related to mutagenesis. However, most alkylating agents and compounds that create adducts and single-strand breakage increase sister chromatid exchange. Therefore, the assay tends to be used as a complement to mutagenesis and chromosomal aberration assays.

Food-associated Mutagens

The investigation of mutagens in food is an important component of food-safety research. Mutagens found in food can be divided into three groups: naturally occurring compounds, those formed by heating or processing, and additives and contaminants, including pesticides.

Naturally Occurring Mutagens

Natural components of food account for over 99% of the carcinogenic and toxic chemicals to which humans are exposed. Most of these natural mutagens are produced by plants as natural pesticides. Many of these intrinsic chemicals are almost impossible to regulate or remove, although guidelines for acceptable exposure limits of some natural contaminants such as mycotoxins are enforced.

Flavonoids form a major group of natural plant compounds, and in a survey of mutagenesis using the Ames *Salmonella* Test, 33 out of 70 natural and synthetic flavonoids proved to be mutagenic, producing both base pair and frameshift mutations. Studies indicate that only the flavonol (3-hydroxyflavone) members of this group are mutagenic. Myricetin and quercetin are direct-acting mutagenic flavonoids, whereas kaempferol is indirect-acting. Norwogonin

and sexangularetin are two other indirect-acting flavonoids. The *pyrrolizidine alkaloids* represent another large group of widespread mutagenic plant compounds found in some herbal teas and medicines, and include compounds such as monocrotaline, dehydomonocrotaline, retrorsine, and isatidine. Pyrrolizidine alkaloids are indirect-acting mutagens, with some compounds inducing DNA-strand breakage, whereas others form DNA and DNA–protein cross-links. Another class of mutagenic compounds is the *glucosinolates*, which are found in cruciferous vegetables. Glucosinolates are broken down into isothiocyanates or nitriles, including the mutagenic allyl isothiocyanate glucoside (sinigrin). The mutagen aquilide A (ptaquiloside) from bracken fern (fiddleheads) is a related *glycoside* and can also be passed through cows' milk. *Catechol-type phenolics*, such as tannins and caffeic acid and its esters (chlorogenic and neochlorogenic acids), are more widespread than other natural pesticides and can induce chromosomal damage. Caffeic acid conjugates can be present in higher plants, with coffee, apple, and lettuce as major sources in the human diet. A class of chemicals known as *linear furanocoumarins*, found in parsley and celery, has several mutagenic members, including psoralen, 5-methoxypsoralen (bergapten), and 8-methoxypsoralen (xanthotoxin). Linear furanocoumarins intercalate into DNA and form mono- and di-adducts in the presence of long-wave UV light. A number of mutagenic *hydrazine* compounds such as hydrazinobenzoate are found in 30 species of mushrooms. Estragole and safrole are mutagenic *alkenylbenzenes* that are found in a number of spices such as black pepper, basil, fennel, and nutmeg. Other food-related mutagens, include ethyl acrylate from pineapple, sesamol from sesame seeds, and benzyl acetate from basil, jasmine tea, and honey.

Mycotoxins are an important source of mutagenic contamination resulting from mold growth in foodstuffs. Human exposure can occur through grain crops, meat products, milk and eggs, nuts and peanuts, fruits, and juice. *Aspergillus* spp. produce a family of aflatoxin mycotoxins (aflatoxins B1, B2, G1, G2), of which aflatoxin B1 is the most potent. Food products produced from livestock fed aflatoxin-contaminated feeds can also serve as a potential source of human exposure. Other mutagenic mycotoxins include vomitoxin and fusarin C produced by *Fusarium* sp., luteoskyrin (*Penicillium* sp.), sterigmatocystin (*Aspergillus* sp.), ochratoxin A, and patulin (*Aspergillus* and *Penicillium* spp.). *Alternaria* sp., of which *A. alternata* is primarily responsible for black spot on tomatoes and black head on grain, produce a number of mutagenic toxins such as altertoxins I, II, and III, and stemphyltoxin III (Figure 1).

Mutagens Formed During Processing, Preparation or Storage

Many very significant mutagens are formed during the processing and preparation of food. Mutagens that are formed as a result of heating and preparation are extremely difficult to avoid or regulate.

Grilling or charring of protein rich foods has been shown to produce some of the most potent mutagens described to date. *Polycyclic heterocyclic amines* are indirect mutagens formed during frying or grilling of meat and fish. This group includes several related amino imidazo-azaarene structures belonging to 3 major families: (1) imidazoquinolines (4-MeIQ), (2) imidazoquinoxalines (8-MeIQx, 4,8-DiMeIQx), and (3) imidazopyridines (PhIP). Grilling also produces a number of amino acid pyrolysis products (Trp-P-1, Trp-P-2, Lys-P-1, Phe-P-1, IQx) and oxygen containing heterocycles. Generally, PhIP is the predominant heterocyclic amine found and can be transferred to infants through breast milk.

Heterocyclic amines appear to be formed by a Maillard reaction between creatine or creatinine, an amino acid and a sugar such as glucose or fructose. The Maillard reaction is responsible for browning, which provides the improved appearance, flavoring, and aroma of heated foods. The reaction starts as a condensation between reducing sugars and the amino groups of amino acids, peptides, or proteins. The initial products may proceed through a very complex series of transformations resulting in the formation of thousands of compounds, including both mutagens and nonmutagens.

Polycyclic aromatic hydrocarbons are a second class of mutagenic compounds formed during the grilling or smoking of meat and fish. Polycyclic aromatic hydrocarbons are common pyrolysis products of most organic material, and include benzo[a]pyrene, 1-nitropyrene and 1, 6-dinitropyrene. The *furfurals*, another class of compounds, are produced during the heating of sugars. Hydroxymethylfurfural is a heat-induced decomposition product of hexoses that becomes mutagenic after sulfate conjugation.

Other processes can give rise to mutagens. *Urethane* (ethyl carbamate) is found in fermented foods and in beverages treated with the fungicide pyrocarbonate. *N-Nitrosodimethylamine* is a naturally occurring mutagen present in cheese, soybean oil, canned fruit, meat products, and alcoholic beverages, but is also formed from the use of elastic rubber nettings in cured pork products. Most nitrosamines cross the placental barrier or pass into the mother's milk, thereby leading to possible *in utero* and infant exposure.

Lipid peroxidation occurs with the storage of fats and oils, owing to the reaction of oxygen with sites of

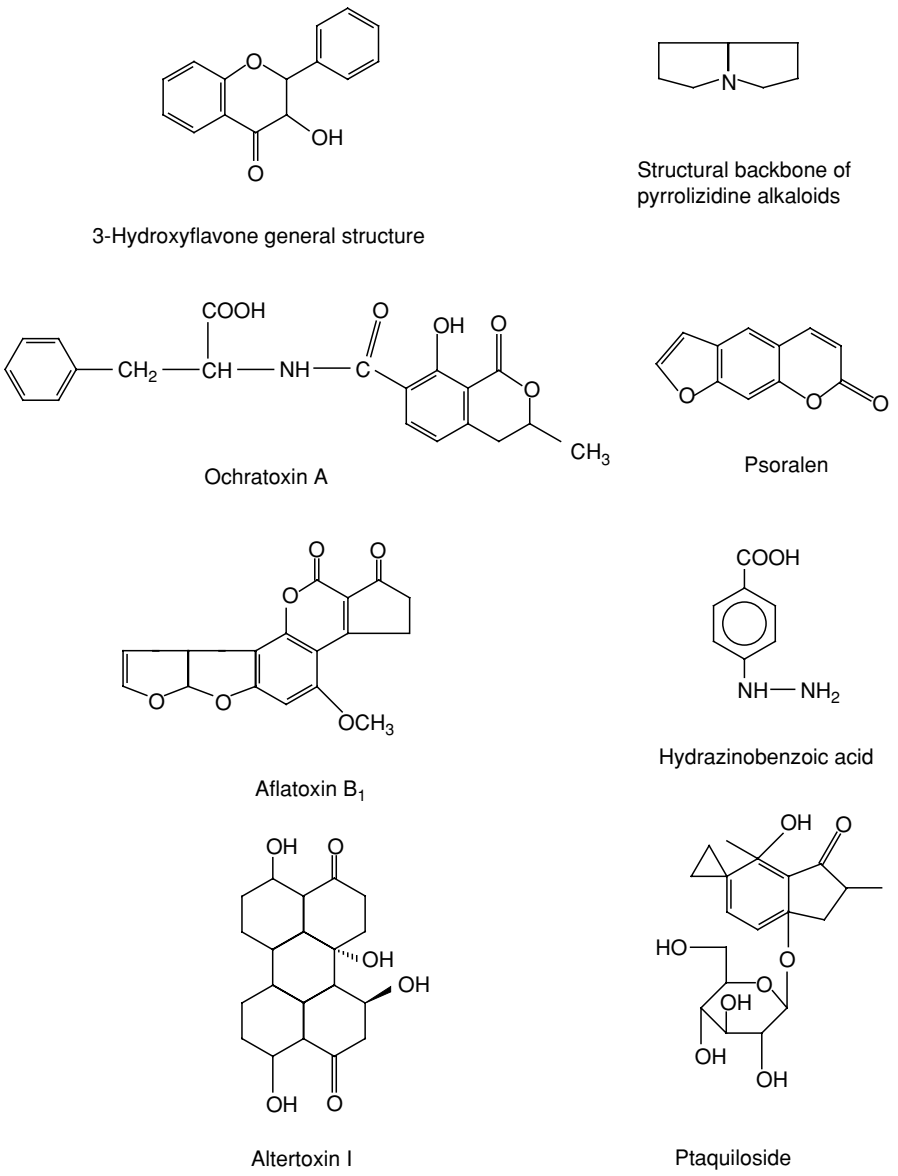

Figure 1 Structures of selected naturally occurring mutagens.

unsaturated bonds. The resulting lipid hydroperoxides, peroxy radicals, and hydroxyl radicals cause rancidity and are potential DNA-damaging agents. A variety of epoxides and aldehydes are also formed, including malondialdehyde, a potential mutagen (Figure 2).

Food Additives and Contaminants

Food additives and contaminants (including pesticide residues) are compounds that are added to food at some point during its production. Exposure to these substances can be controlled, and the introduction of new chemicals used in food manufacture falls under government regulation. Older additives or chemical contaminants are also reviewed periodically for exposure and health effects.

Use of the preservatives nitrite, butylated hydroxyanisole (BHA) and butylated hydroxytoluene (BHT) as food preservatives has been controversial. *Nitrite* enters the diet both through its use as a common food preservative and through gut microbe-mediated reduction of nitrate from ingested vegetables. Nitrite cycles from the gut to the saliva and can react at acidic pH with amines, amino acids, phenols, and mercaptans, frequently resulting in the formation of mutagenic (nitroso) compounds. Mutagens can also be found in the smoke of nitrite-treated meat during frying. The formation of 2-chloro-4-methylthiobutanoic

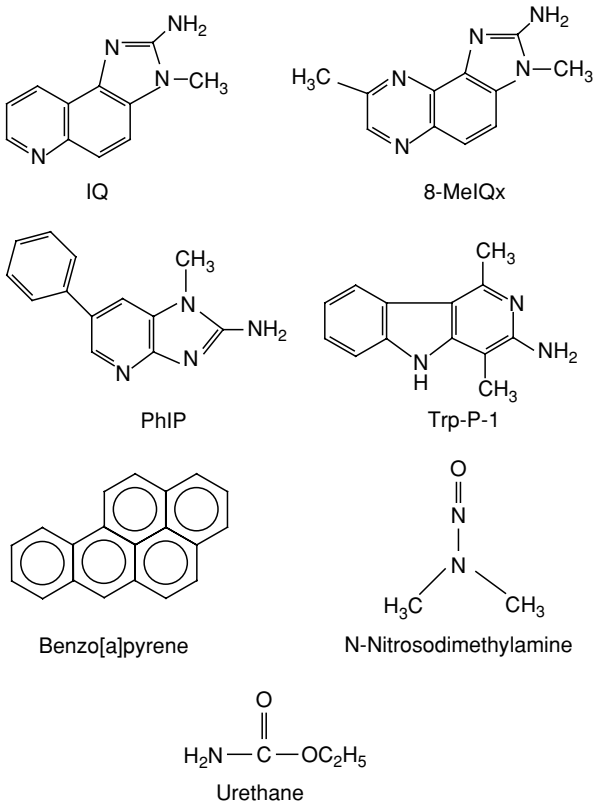

Figure 2 Structures of selected mutagens formed during grilling and processing.

acid, a mutagen in salted, pickled fish, has also been linked to nitrite. BHA and BHT have been added to foods and packaging as synthetic antioxidants. Previous concerns over BHA's carcinogenic risk to humans appear unwarranted and related to the large doses used in rodent-feeding studies. BHT increases the proliferation of pretumor liver cells in rodent models but has little effect on normal cells.

The artificial sweeteners saccharin and cyclamate have also been shown to have mutagenic properties, although saccharin-induced bladder carcinogenesis in rodents also appears to result from the large doses used. Also, the mutagenicity of some colors has led to their use being banned, although low levels of extractable mutagenic contaminants have been detected in others. Mutagenic heterocyclic amino acids have been found in some flavors.

Contaminants may occur in food as residues of pesticides and fungicides, solvents, or compounds, such as styrene leaching from packaging material. The persistent organochlorine pesticide toxaphene has been found in rural foods consumed by indigenous peoples in the Arctic region, having been carried and deposited through airborne transport from sites in the south.

Mutagens in Water

Drinking water provides another potential source of mutagenic compounds. Water disinfection by chlorination can produce many types of mutagenic chlorinated organic compounds, of which 3-chloro-4-(dichloromethyl)-5-hydroxy-2-(5H)-furanone (mucochloric acid, MX) is the most potent. Other mutagenic chlorinated butenoic acids, including EMX (the reduced form of MX), brominated trihalomethanes, and dichloroacetonitrile, are also produced. Dichloroacetonitrile is a direct-acting mutagen and induces DNA-strand breaks in cultured human lymphoblastic cells. Other less potent mutagenic disinfection byproducts include dichloroacetic acid, trichloroacetic acid, and chloral hydrate. Radioactive contamination can also contribute mutagenic activity in areas where water is contaminated with ^{40}K or uranium and thorium decay products (such as radon).

Mutagens in Feces

Human feces have been found to contain a class of mutagenic compounds known as fecapentaenes, the major forms of which are fecapentaene-12 and fecapentaene-14. These compounds appear to be synthesized by intestinal bacteria and are lipid-soluble, direct-acting base pair and frameshift mutagens. Fecapentaene-12 also causes increased sister chromatid exchanges, *HGPRT* mutagenesis, and oxidative damage. Bile salts have been shown to augment the ability of fecapentaenes-12 and -14 to induce DNA repair in bacteria, underscoring their possible role in colon carcinogenesis associated with high-fat, low-fiber diets.

Prevention and Modification of Mutagen Formation

A well-balanced diet also provides many compounds that can decrease exposure to mutagens. Studies consistently show that a greater consumption of fruits and vegetables is associated with a lowering of cancer incidence. These compounds can act at all levels of mutagenesis by: (1) preventing mutagen formation and acting as antioxidants, (2) deactivating mutagenic compounds, and (3) inducing detoxifying enzymes or inhibiting monoxygenase activity.

Several natural antioxidants, such as ascorbic acid, β-carotene, α-tocopherol, polyphenolic acids, flavonoids, and other plant phenolics have a direct mutagen-scavenging activity. For example, in addition to neutralizing oxygen radicals, many of these compounds inhibit the formation of nitrosamines by competing with the amine group. Mutagenic agents may

be inactivated through the formation of complexes with calcium, casein, chlorophyl and carbonyl compounds. Some evidence suggests that calcium can inhibit familial colon carcinogenesis through inhibition of lipid peroxidation. Both iron deficiency and excess have been associated with oxidative DNA damage, and zinc deficiency causes chromosome breaks. The activity of glutathione-dependent detoxification systems is increased by selenium and diallyl disulfide, from garlic, cysteamine, cysteine, as well as a variety of phenolic compounds. Monooxygenase activity and the formation of reactive compounds can be inhibited by phenolic acids, free fatty acids, flavonoids, and other polyphenols, methylxanthines, vitamins, and biogenic amines. Conjugated dienoic isomers of linoleic acid are antimutagenic compounds in extracts of fried ground beef and dairy products that appear to act through signal-transduction pathways and prostaglandin synthesis. Tannins or polyphenols in green tea have been found to be antimutagenic in some assays, but the effects appear to depend upon the dose and timing of exposure. Several other compounds have shown antimutagenic activity, including β-diketones, catechins, lignans, and isoflavones. Finally, dietary fiber can complex with mutagenic compounds and limit their uptake.

Mutagenic effects resulting from altered DNA metabolism can also be influenced by diet. Adequate intake of essential nutrients such as folate, vitamin B_{12} and vitamin B_6 aids nucleotide synthesis and prevents DNA-strand breakage. Niacin, methylxanthines, and some phenols have been attributed to promotion of DNA-damage repair (Figure 3).

Further Considerations

Much work remains to completely characterize the role of food-associated mutagens and antimutagens in human carcinogenesis. Despite the magnitude and frequency of exposure to these compounds, they

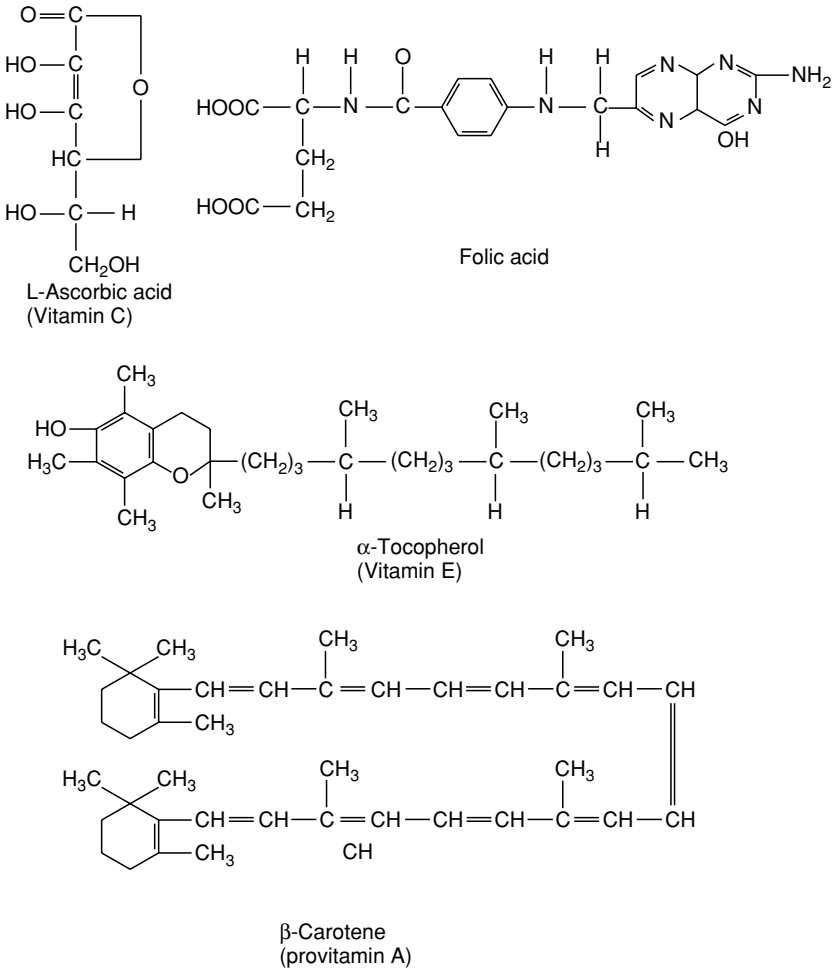

Figure 3 Structures of common dietary antioxidant antimutagens.

remain underrepresented in mutagenesis studies relative to synthetic chemicals. Additional characterization is needed of the human consequences of exposure to many antioxidants that appear to inhibit mutagenesis at low concentrations but have the opposite effect at high concentrations. Finally, more emphasis should be placed on the effects of chemical mixtures to study the effects of combinations of compounds encountered in the diet.

See also: **Aflatoxins**; **Amines**; **Carcinogens**: Carcinogenicity Tests; **Food Additives**: Safety; **Mycotoxins**: Classifications; Occurrence and Determination; Toxicology; **Polycyclic Aromatic Hydrocarbons**; **Saccharin**

Further Reading

Ames BN (1998) Micronutrients prevent cancer and delay aging. *Toxicology Letters* 102–103: 5–18.

Clayson DB (2001) *Toxicological Carcinogenesis*. Boca Raton, FL: Lewis Publishers.

Franks LM and Teich NM (eds) (1997) *Introduction to the Cellular and Molecular Biology of Cancer*, 3rd edn. Oxford: Oxford University Press.

Friedberg EC, Walker GC and Siede W (1995) *DNA Repair and Mutagenesis*. Washington, DC: ASM Press.

Hui YH, Kitts D and Stanfield PS (2001) *Foodborne Disease Handbook, Seafood and Environmental Toxins*, 2nd edn., vol. 4. New York: Marcel Dekker.

Hui YH, Smith RA and Spoerke DG, Jr. (eds) (2001) *Foodborne Disease Handbook, Plant Toxicants*, 2nd edn, vol. 3, New York: Marcel Dekker.

Kilbey BJ, Legator M, Nichols W and Ramel C (1984) *Handbook of Mutagenicity Test Procedures*. Amsterdam: Elsevier Science.

Kitchin KT (ed.) (1999) *Carcinogenicity: Testing, Predicting, and Interpreting Chemical Effects*. New York: Marcel Dekker.

Klug WS and Cummings MR (2000) *Concepts of Genetics*, 6th edn. Upper Saddle River, NJ: Prentice Hall.

McKinnell RG, Parchment RE, Perantoni AO and Pierce GB (1998) *The Biological Basis of Cancer*. Cambridge: Cambridge University Press.

National Research Council Monograph (1996) *Carcinogens and Anticarcinogens in the Human Diet*. Washington, DC: National Academy Press.

Pariza MW, Aeschbacher H-U, Felton JS and Sato S (eds) (1990) Mutagens and carcinogens in the diet. In: *Progress in Clinical and Biological Research*, vol. 347. New York: Wiley-Liss.

Pfeifer GP (ed.) (1996) *Technologies for Detection of DNA Damage and Mutations*. New York: Plenum Press.

Twyman RM (1998) *Advanced Molecular Biology: A Concise Reference*. Oxford: Bios Scientific.

Mutton *See* **Sheep**: Meat; Milk

MYCOBACTERIA

C H Collins and J M Grange, National Heart and Lung Institute, London, UK

Introduction

The genus *Mycobacterium* contains over 60 species; these are divided into rapid-growers, slow-growers, and the human leprosy bacillus which has not been convincingly cultivated *in vitro*. A few of the species are obligate parasites, but most of them are environmental saprophytes. The latter are widely distributed in nature and have been isolated from natural waters, wet soil, mud, compost, grasses, vegetables, unpasteurized milk, and butter. They have also been isolated from domestic water pipes from which they readily enter drinking water. Mycobacteria are characterized by the possession of very thick, waxy, lipid-rich hydrophobic cell walls. Being hydrophobic, they tend to grow as fungus-like pellicles on liquid culture media: hence the name *Mycobacterium* – 'fungus bacterium.'

Even the rapidly growing mycobacteria grow slowly in comparison with most other bacteria. Mycobacteria may play a role in the decomposition of organic material, particularly in sphagnum

marshes, but there are no reports of mycobacteria causing spoilage in foods. They do not produce appreciable amounts of toxic substances and do not cause food poisoning. Pathogenic strains owe their virulence to their ability to resist immune defense mechanisms; they cause chronic infections.

The obligate parasites are *Mycobacterium tuberculosis* (the human tubercle bacillus), *M. bovis* (the bovine tubercle bacillus), and *M. africanum* (a species first described in equatorial Africa and having rather variable properties that are intermediate between the other two). Variants of these species are occasionally encountered and include *M. microti* (the vole tubercle bacillus) and biochemically and genetically distinct isolates from goats termed the caprine genotype or *M. tuberculosis* subsp. *caprae*. All these species, often grouped as the tuberculosis complex, are very closely related and should be regarded as variants of a single species. Other pathogens thought to be obligate parasites are *M. leprae* (the causative organism of leprosy), *M. paratuberculosis* (the cause of Johne's disease or hypertrophic enteritis in cattle and other ruminants) and *M. lepraemurium* (the cause of rat leprosy). As *M. leprae* is uncultivable and other two are very difficult to cultivate *in vitro*, the possibility that they can replicate in the inanimate environment cannot be excluded.

Some of the saprophytic species occasionally cause opportunistic infections of animals and humans. The principal opportunist species are *M. avium*, *M. intracellulare*, *M. scrofulaceum*, *M. kansasii*, *M. xenopi*, *M. malmoense*, and the rapidly growing species *M. chelonae* and *M. fortuitum*. The first two species are closely related and are often grouped together as the *M. avium* complex (MAC). Although essentially saprophytic, MAC are the most pathogenic of the opportunist mycobacteria. Members of this complex are a common cause of tuberculosis in birds and they also cause limited lesions, particularly cervical lymphadenopathy, in mammals, including pigs, cattle, and deer. Closely related to the MAC are *M. lepraemurium*, *M. paratuberculosis* and the wood pigeon bacillus (*M. avium* subsp. *sylvaticum*) which resembles *M. paratuberculosis* in requiring mycobactin, an iron-binding lipid extracted from mycobacterial cell walls, for its *in vitro* cultivation. A few other nutritionally very fastidious mycobacteria, requiring enriched media for their cultivation, have recently been described. These include *M. genavense*, which was originally detected in acquired immunodeficiency syndrome (AIDS) patients by means of DNA amplification techniques. This species has also been isolated from pet birds and a dog and infected pets may thus pose a health hazard to severely immunosuppressed persons.

Mycobacterial Disease

Although the human mycobacterial diseases differ enormously in their clinical features they have certain characteristics in common. They all commence with a local lesion, which may or may not be clinically evident, at the site of implantation of the causative organism. They may be in the lung in the case of inhaled bacilli, in the skin following trauma, or in the pharynx or intestinal tract following ingestion of the bacilli in contaminated food or water. In some cases, notably in tuberculosis, the local lymph nodes are also involved in the primary infection and the lymphatic lesion may be much more extensive than that at the site of implantation. Thus, ingestion of milk contaminated with *M. bovis* often leads to a cryptic tonsillar or pharyngeal lesion and gross enlargement of one or more of the lymph nodes in the neck – a condition known as scrofula. Bacilli may enter the blood stream from the primary lesion and then cause serious forms of primary tuberculosis, notably tuberculous meningitis and disease of the kidneys, bones, and joints. Thus, in countries where there is (or was) a high incidence of human tuberculosis of bovine origin, scrofula and associated nonpulmonary manifestations of the disease are (or were) commonly seen in children.

Unless one of the serious sequelae of primary tuberculosis develops, the primary complex usually resolves, but the disease reactivates in about 5% of infected persons to cause postprimary tuberculosis, years or even decades later. (The risk of reactivation is much higher in immunosuppressed persons and human immunodeficiency virus (HIV) infection has emerged as the most prevalent predisposing factor for the development of tuberculosis worldwide. A person dually infected by HIV and *M. tuberculosis* has a 8–10% chance of developing active tuberculosis annually.) Cases of tuberculosis due to late reactivation of *M. bovis* infection still therefore occur in countries that have virtually eradicated tuberculosis from cattle. In Europe and Australia, *M. bovis* accounts for around 1% of cases of tuberculosis.

The way in which residual mycobacteria persist in tissues for long periods is unknown. The reactivation may occur at the site of the primary lesion (such as cervical lymph nodes in the case of tuberculosis of bovine origin), but more often it reactivates, for unknown reasons, in the upper lobes of the lung. This leads to open or infectious tuberculosis and there have been reports of cattle being infected by such patients. Also, and for unknown reasons, some cases (about 25%) of postprimary human tuberculosis due to *M. bovis* in Europe involve the genitourinary tract. This rather insidious form of tuberculosis often passes

undiagnosed for long periods and farm workers with this condition are known to have transmitted tuberculosis to cattle by urinating in cowsheds.

In humans, some environmental mycobacteria are able to cause lymphadenopathy, particularly in young children; skin lesions following injections of contaminated material, superficial or penetrating injuries, and surgery involving implantation of contaminated prostheses such as porcine heart valves; pulmonary disease, particularly in patients with pneumoconiosis or other predisposing lung diseases; and disseminated disease. The last, once rare, is now a common AIDS-related condition and, for unknown reasons, is usually caused by the MAC, although some cases are due to the more recently described *M. genavense*, referred to above.

In recent years there has also been speculation that ingested mycobacteria, possibly from milk, might be responsible for Crohn's disease, a granulomatous disease of the human intestine. Normal forms and spheroplasts of mycobacteria have been isolated from tissue excised from a few patients with Crohn's disease and DNA supposedly specific for *M. paratuberculosis*, the cause of Johne's disease in cattle, has been detected by the polymerase chain reaction (PCR) in biopsies from patients with Crohn's disease. In some studies such DNA was detected in the majority of specimens but in other studies no such DNA was detected. These conflicting findings may be attributable to methodological problems and technological advances may resolve the issue but at present no firm conclusions can be drawn. Likewise, attempts to detect antibodies to *M. paratuberculosis* in serum and antigen in tissues, as well as trials of antimycobacterial drugs, have yielded unconvincing results.

Sources of Infection

The human tubercle bacillus, *M. tuberculosis*, is usually spread from person to person via expectorated aerosols and causes pulmonary tuberculosis. Individuals with open tuberculosis who work in the food industry may therefore transmit the organisms to other people and could possibly, in theory, contaminate food. The bovine bacillus, *M. bovis*, causes pulmonary tuberculosis in farm workers who inhale the bacilli in aerosols expectorated by diseased cattle while town dwellers usually become infected by drinking contaminated milk and predominantly develop nonpulmonary tuberculosis.

Although persons working with other animals with disease due to *M. bovis*, including seals, elk, and a rhinoceros with tuberculous rhinitis, have become tuberculin-positive, overt disease has seldom been reported. A few strains of the caprine genotype have, however, been isolated from humans with tuberculosis, including one isolate from a veterinary surgeon with a recent history of working with tuberculous goats, and a seal trainer contracted tuberculosis due to an *M. bovis*-like strain from a captive seal in Australia.

Mycobacterial disease may result from the inoculation of the bacilli into the skin. In areas where tuberculosis of cattle is common, food handlers are at risk of developing tuberculous lesions of the hands as a result of contamination of minor cuts and abrasions, the resulting condition being known as butcher's wart (it is also known as prosector's wart, as anatomists and pathologists were likewise exposed to this occupational hazard).

Food and drink may serve as vectors for the transmission of pathogenic mycobacteria. By far the most likely vehicle, leading to disease of the alimentary tract, is unpasteurized milk. Meat and water are also possible sources, but are much less important. (*See* **Milk**: Processing of Liquid Milk.)

Unpasteurized Milk

There is a long history of human infection resulting from the consumption of milk that contained *M. bovis*. In developed countries such as the UK these infections are now rare for two reasons: the eradication of tuberculosis among cattle by tuberculin testing and slaughter; and the pasteurization of milk for human consumption. Both have been the subject of legislation. Milk-borne tuberculosis is still a hazard, however, in developing countries.

The history of milk-borne tuberculosis in the UK illustrates this. The disease in cattle (and other animals) had been recognized as an economic problem for centuries. By the turn of the century it was known that tuberculosis in cattle was caused by *M. bovis*, but Robert Koch, who had discovered the tubercle bacillus in 1882 and was regarded as a world authority on the subject, told the British Congress on Tuberculosis in 1901 that the bovine organism did not cause the disease in humans and that measures to counteract infection with that organism were unnecessary. This was disputed by certain British bacteriologists and veterinarians and it was not until after the publications of the findings of a Royal Commission in 1911 that the bovine bacillus was accepted as the causative organism of many cases of nonpulmonary tuberculosis, notably cervical lymphadenopathy, abdominal and skeletal tuberculosis, particularly in children.

Official action was slow, however, and in the 1930s it was known that about 40% of all cattle slaughtered in public abattoirs had tuberculous lesions and that about 0.5% of all dairy cows produced milk that

contained the bovine tubercle bacillus. The eradication scheme that followed was initially voluntary and did not become compulsory until after World War II. By 1960 all dairy herds in the UK had been tested and by 1979 the reactor rate had fallen dramatically and only 0.18% of herds were infected. This may well represent an irreducible minimum, because reinfection, from humans and other animals (e.g., badgers) may still occur and therefore there is still a risk of infection from unpasteurized milk although, at present, as a result of more legislation, very little of this is sold for immediate consumption in this country.

Goats may also be infected, and tubercle bacilli have been recovered from their milk. The consumption of raw goats' milk is therefore not without risk.

Little is known about the role of milk in other mycobacterial infections. The possible link between milk-borne *M. paratuberculosis* and Crohn's disease has been referred to above. Members of the MAC and other saprophytic mycobacteria have been isolated from milk, but no direct link has been established between the presence of these organisms in milk and human disease, although contaminated milk is a possible source of *M. avium* causing disease in AIDS patients. Although DNA specific for *M. avium*, including the *sylvaticum* subspecies, has been detected in a few biopsies from patients with Crohn's disease, their role as primary pathogens remains, as in the case of *M. paratuberculosis*, highly speculative.

Meat

In most parts of the world meat is cooked before it is eaten, but there is nevertheless the possibility that there is insufficient heat penetration and that tubercle bacilli may remain viable in parts of the meat. Cross-contamination from infected meat to food ready for consumption is also a possibility. As indicated above, some infected cattle still reach slaughterhouses and the meat of some may be tuberculous. Five cases of tuberculosis due to *M. bovis* occurred over a 2-year period in abattoir workers in South Australia: four of the five cases involved the lung, suggesting that infection followed inhalation of aerosols. In addition, a DNA fingerprinting study of strains of *M. bovis* isolated from patients in Australia between 1970 and 1994 revealed that the majority of Australian-born patients had worked in the meat and livestock industries and had disease due to strains similar to those found in Australian cattle.

The incidence of tuberculosis among sheep is low, as it is among pigs, although members of the MAC have been responsible for outbreaks and single cases of tuberculous disease in piggeries. Bovine tubercle bacilli have been isolated from rabbits, wild and farmed, and along with the avian bacillus, from a variety of poultry. In the UK and New Zealand farmed, but not wild, deer are now known to suffer from tuberculosis. In the Middle East and India there is evidence that some of the camels that are slaughtered for their meat are tuberculous.

Fortunately, in developed countries, the systems of meat inspection, required by law, prevent the sale for human consumption of bovine, ovine, and porcine carcasses, or parts of them, that may be infected. There is still a trade, however, in carcasses, e.g., of goats, pigs, rabbits, and poultry, that escapes an official inspection and as yet in the UK carcasses of farmed deer are exempt from official inspection. All these animals offer a degree of hazard to the food handler and the consumer.

Fish are not normally a source of mycobacterial disease, although there have been a few cases of disease due to *M. marinum* infection, manifesting as warty skin lesions, in fishermen and fishmongers, presumably as a result of infection of dermal abrasions. Multiple regression analysis has shown that regular consumption of raw or partly cooked fish and shellfish is associated with an increased risk of disseminated disease due to the MAC in AIDS patients.

Water

Several species of mycobacteria, some of which are opportunist pathogens, have been isolated from piped, bottled, and natural waters. Tubercle bacilli, however, have never been recovered from water supplies.

There is adequate evidence that infections due to some opportunist mycobacteria result from the inhalation of aerosols containing the organisms. In the great majority of cases ingestion of these species does no harm (overt abdominal infection associated with them is extremely rare), but they may temporarily colonize the alimentary and upper respiratory tracts. There is evidence that they may cross the epithelial barrier (translocation) and persist for some time in the draining lymph nodes: thus they have been isolated from excised tonsils and abdominal lymph nodes. Water has been postulated as a source of infection in AIDS patients with disseminated mycobacterial disease due to the *M. avium*, although epidemiological evidence is conflicting, but *M. genevense* was detected in tap water at a hospital with an unusually high incidence of HIV-related disseminated disease due to this species.

Isolation of Mycobacteria from Milk, Meat, and Water

Tubercle bacilli and other pathogenic mycobacteria are in Hazard Group 3 of the Advisory Committee on

Dangerous Pathogens and other similar authorities. Such organisms may be dispersed as aerosols during laboratory manipulations, and individuals who inhale them may become infected. Any bacteriological procedure involving them or material that may contain them must be carried out under Containment Level 3 conditions in microbiological safety cabinets.

In some parts of the world the conventional techniques for isolating mycobacteria on solid media are being replaced by more rapid radiometric methods and more recent nonradiometric automated systems and also by the PCR and its derivatives. Originally evaluated in the field of human disease, they are increasingly being applied to veterinary microbiology, food hygiene, and environmental work. There have, for example, been reports of the use of PCR to detect and identify waterborne mycobacteria, *M. paratuberculosis* in pasteurized milk, and *M. bovis* in tissues obtained postmortem from various animals. Likewise, DNA fingerprinting techniques suitable for epidemiological studies on *M. bovis* from animal and human sources are now available.

The traditional methods for the isolation of mycobacteria are described below, but the final identification of tubercle bacilli and other mycobacteria is too detailed to be included. Alternatively, cultures of suspected mycobacteria may be sent to a reference laboratory, a regional tuberculosis laboratory, or a veterinary laboratory. For information on automated systems and nucleic acid-based technology, the rapidly changing primary literature should be consulted.

Milk

Direct microscopic examination of milk for acid-fast bacilli is unrewarding as saprophytic mycobacteria are often present, as are artifacts that retain the fuchsin stain.

A suitable method for recovering tubercle bacilli (and other mycobacteria) from the milk of individual cows or from small groups of cows is given below.

Centrifuge at least 50 ml of the milk at 3000 rpm for 20 min. Remove the cream to a small tube, add 4 ml of 4% sodium hydroxide and stand for 15 min; add 6 ml of 14% monopotassium phosphate containing phenol red indicator to indicate neutralization; centrifuge at 3000 rpm for 15 min; discard supernatant fluid and culture the deposit.

Discard the remaining supernatant fluid from the original centrifugation and add 2 ml of 4% sodium hydroxide to the deposit and stand for 15 min; add 3 ml of 14% monopotassium phosphate (3 ml) and proceed as above.

Divide the deposits between several tubes of Löwenstein–Jensen medium or Middlebrook medium containing antibiotics. Incubate at 37 °C for 6–8 weeks. Check the identity of any organisms that grow by Gram and Ziehl–Neelsen stains.

Identify acid-fast bacilli by the methods described in standard textbooks or send cultures to a specialist laboratory.

Meat

For full details, the World Health Organization guidelines for speciation within the *Mycobacterium tuberculosis* complex should be consulted. The recommended methods are briefly summarized here.

The material should be selected by a veterinarian or meat inspector familiar with the macroscopic appearance of tuberculous lesions. The usual specimens are lymph nodes from the respiratory and gastrointestinal tracts, lung, and liver tissue and these should be placed in wide-mouthed, hermetically sealed plastic or glass pots.

Cut about 10 g of the sample into small pieces with sterile instruments and further homogenize the pieces in a Griffith tube or mechanical blender.

Remove about 3 ml of homogenized tissue with a wide-bore pipette and add to an equal quantity of autoclaved 4% sodium hydroxide in a screw-capped tube, mix well and stand for 15–20 min, with occasional shaking.

Neutralize with a 14% w/v solution of monopotassium phosphate containing phenol red indicator (about 40 mg l^{-1}) and allow to stand until the indicator turns red, indicating that neutralization is complete. Centrifuge fluid at 3000 rpm for 20 min, decant the supernatant fluid, and inoculate the deposit on two slopes of Löwenstein–Jensen media and two slopes of a related medium, such as Stonebrink's medium, that contains sodium pyruvate instead of glycerol and which favors the growth of *M. bovis*.

Incubate slopes at 35–37 °C for up to 8 weeks. Read weekly, examine colonies for acid-fastness, and arrange for identification by a specialist laboratory.

Water

The sample should consist of at least 1 l of water and be delivered to the laboratory as soon as possible.

Pass separate 100-ml volumes of the whole sample through membrane filters. Place the membrane filters in 4% sodium hydroxide solution for 5 min (e.g., in a Petri dish), then in 14% monopotassium phosphate solution containing phenol red for several minutes.

Cut the membranes into strips 5 mm wide and place each strip on the surface of Löwenstein–Jensen or Middlebrook antibiotic medium in screw-capped tubes. Incubate at 35–37 °C. Examine every few days as some mycobacteria grow rapidly.

Examine colonies for acid-fastness and arrange for identification, if necessary.

In developed countries, foodborne mycobacteria do not offer a serious health hazard and routine laboratory examinations are not justified. In special circumstances attempts may be made to culture the organisms by the methods described above. However, the essential and expensive equipment requirements for Containment Level 3 preclude identification in most laboratories. The advice and services of reference and other specialized laboratories should be sought.

See also: **Milk**: Processing of Liquid Milk

Further Reading

Collins CH and Grange JM (1983) A review: the bovine tubercle bacillus. *Journal of Applied Bacteriology 55*: 13–29.

Collins CH and Grange JM (1987) Zoonotic implications of *Mycobacterium bovis* infections. *Irish Veterinary Journal* 41: 363–366.

Collins CH, Grange JM and Yates MD (1984) A review: mycobacteria in water. *Journal of Applied Bacteriology* 57: 193–211.

Collins CH, Lyne PM and Grange JM (eds) (1995) *Collins and Lyne's Microbiological Methods*, 7th edn. Oxford: Butterworth Heinemann.

Collins CH, Grange JM and Yates MD (1997) *Tuberculosis Bacteriology Organization and Practice*, 2nd edn. Oxford: Butterworth Heinemann.

Cosivi O, Grange JM, Daborn CJ *et al*. (1998) Zoonotic tuberculosis due to *Mycobacterium bovis* in developing countries. *Emerging Infectious Diseases* 4: 59–70.

Cousins DV and Dawson DJ (1999) Tuberculosis due to *Mycobacterium bovis* in the Australian population: cases recorded during 1970–1994. *International Journal of Tuberculosis and Lung Disease* 3: 715–721.

Cousins DV, Williams SN and Dawson DJ (1999) Tuberculosis due to *Mycobacterium bovis* in the Australian population: DNA typing of isolates, 1970–1994. *International Journal of Tuberculosis and Lung Disease* 3: 722–731.

Covert TC, Rodgers MR, Reyes AL and Stelma GN (1999) Occurrence of nontuberculous mycobacteria in environmental samples. *Applied and Environmental Microbiology* 65: 2492–2496.

Fordham von Reyn C, Arbeit RD, Tosteson AN *et al*. (1996) The international epidemiology of disseminated *Mycobacterium avium* complex infection in AIDS. International MAC Study Group. *AIDS* 10: 1025–1032.

Grange JM and Collins CH (1987) Bovine tubercle bacilli and disease in animals and man. *Epidemiology and Infection* 92: 221–234.

Grange JM, Yates MD and de Kantor IN (1996) *Guidelines for Speciation within the* Mycobacterium tuberculosis *complex*, 2nd edn. WHO/EMC/ZOO/96.4. Geneva: World Health Organization.

Hubbard J and Surawicz CM (1999) Etiological role of *Mycobacterium* in Crohn's disease: an assessment of the literature. *Digestive Diseases* 17: 6–13.

Millar D, Ford J, Sanderson J *et al*. (1996) IS900 PCR to detect *Mycobacterium paratuberculosis* in retail supplies of whole pasteurized cows' milk in England and Wales. *Applied and Environmental Microbiology* 62: 3446–3452.

Moda G, Daborn CJ, Grange JM and Cosivi O (1996) The zoonotic importance of *Mycobacterium bovis*. *Tubercle and Lung Disease* 77: 103–108.

O'Reilly LM and Daborn CJ (1995) The epidemiology of *Mycobacterium bovis* infections in animals and man: a review. *Tubercle and Lung Disease* 76: 1–46.

Roberts TA (1986) A retrospective assessment of human health protection benefits from removal of tuberculous beef. *Journal of Food Protection* 49: 293–298.

Robinson P, Morris D and Antic R (1988) *Mycobacterium bovis* as an ocupational hazard in abattoir workers. *Australia and New Zealand Journal of Medicine* 18: 701–703.

Thoen CO and Steele JH (eds) (1995) Mycobacterium bovis *Infection in Humans and Animals*. Ames, Iowa: Iowa State University Press.

Wards BJ, Collins DM and de Lisle GW (1995) Detection of *Mycobacterium bovis* in tissues by polymerase chain reaction. *Veterinary Microbiology* 43: 227–240.

Van Kruiningen HJ (1999) Lack of support for a common etiology in Johne's disease of animals and Crohn's disease in humans. *Inflammatory Bowel Disease* 5: 183–191.

MYCOPROTEIN

M J Sadler, MJSR Associates, Ashford, UK

Background

Myco-protein is produced from microfungi, aerobic organisms that live in the soil and convert carbohydrate to protein. Research to produce food from microfungi began in the early 1960s. This was at a time when many projects were being set up to develop single-cell protein sources suitable for animal and human consumption in response to the predicted protein gap in developing countries. There are reports of three products produced form microfungi, but by far the most successful myco-protein product is sold under the trade name 'Quorn™.'

Myco-protein is the generic name of the major raw material used in the manufacture of Quorn™ products. It comprises the RNA-reduced biomass composed of the hyphae (cells) of the organism *Fusarium venenatum* A3/5 (deposited with the ATCC as PTA-2684) grown under axenic conditions in a continuous fermentation process. Quorn™ is the brand name of a range of meat-alternative products made from myco-protein. Quorn™ products include pieces and mince for use in home cooking in addition to a range of convenience products such as burgers, fillets, goujons, nuggets, and ready meals.

Quorn™ myco-protein is the first truly 'new' food product developed for humans. Introduced to the UK food market in 1984, Quorn™ products are the result of a development program, designed to produce single-cell protein, that was started in the 1960s. In addition to availability in all major UK and Irish food retailers, Quorn™ products are now available in an increasing number of retailers within Continental Europe, including Belgium, Luxembourg, Holland, Sweden, Switzerland and France, and in the USA.

This article covers the cultivation and harvesting of myco-protein, and its properties including acceptability, safety, and nutritional value. Known and other potential functional health effects of consuming myco-protein are also discussed.

It was predicted in the late 1950s that within three decades, there would be a shortage of protein across the world – the so-called 'protein gap.' In response to this, many projects were set up in the early 1960s to develop feeds or foods from single-cell protein sources that would be suitable for animal or human consumption.

Ranks Hovis McDougall (RHM) Research Centre in the UK undertook one such project, which started in 1964. The original aims were to convert starch into a protein-rich food that was highly nutritious, delicious, and safe to eat. The project focused on the potential of microfungi to produce a high-protein food (myco-protein). This resulted two decades later in the successful development of the Quorn™ (a registered trade mark of Marlow Foods Ltd, UK) range of products, which have been marketed as a human food for the past two decades.

It should be noted that the commercial development of myco-protein was a joint venture between RHM and ICI (Marlow Foods). In 1990, ICI assumed full control of the business. However, in 1993, ICI de-merged into ICI and Zeneca, and Marlow Foods became part of Zeneca. In 1998, Zeneca merged with Astra (Sweden) to become AstraZeneca.

Production

Cultivation

Cultivation of fungal mycelia has many factors in common with the production of yeast and bacteria, although because fungi are multicellular or coenzocytic, the term 'single-cell protein' cannot strictly be used to describe fungi produced by fermentation. The advantages of microfungi are that they can utilize a wide range of substrates, they have straightforward nutritional requirements, and they are relatively easy to harvest from culture broths on account of their particle size.

In addition, the whole microfungi can be consumed without any reported problems of toxicity and with only a very low level of adverse reactions. Fungal mycelia have a high nutritional value, desirable texture, and mild smell and flavour. Acceptability is less of a problem than for food from other lower organisms, since fungi already play a role in the human diet.

The microfungi *Fusarium venenatum* A3/5 (deposited with the ATCC as PTA-2684) was chosen by RHM for the production of the new myco-protein product, on account of its suitable organoleptic and nutritional properties.

Growth requirements and conditions Myco-protein is grown in an airlift or pressure-cycle fermenter in a liquid medium providing all the nutrients required for growth. In preference to a batch culture, a continuous-flow aerobic culture system is used in which the culture medium is continuously fed to the fermenter and the broth continuously drawn off for filtering such that the volume of the fermenter is displaced approximately every 5–6 h. This ensures that steady-state conditions are maintained and that optimal productivity is achieved. Optimizing the process conditions improves the yield, quality, and uniformity of the end product, and the efficiency of the process. The cell mass doubles every 4–5 h.

The carbon substrate used is glucose syrup, which is obtained from hydrolyzed corn. Ammonium is the source of nitrogen, and is also used to regulate the pH, which is maintained between 4.5 and 7.0. Salts of potassium, manganese, cobalt, calcium, magnesium, iron, copper, and biotin are supplied as a liquid feedstock from tanks adjacent to the fermenter. Compressed air is the source of oxygen, and its injection into the fermenter keeps the broth in continuous motion. All the ingredients of the culture medium are sterile and of food- or reagent quality. The temperature of the broth is maintained at 30 °C, and it is therefore necessary to cool the fermenter.

The fermentation must be conducted under monoseptic conditions, allowing only the specified culture to grow. To maintain the purity of the culture, industrial-scale continuous fermentations last up to a maximum of 6 weeks (1000 h).

Nucleic Acid Content

The nucleic acid content of myco-protein is mostly ribonucleic acid (RNA). There are potential clinical problems associated with the consumption of foods high in RNA, and the maximum intake recommended by the World Health Organization (WHO) is 2 g per day. If RNA consumption is too high, blood levels and urinary excretion of uric acid are increased, and uric acid crystals may precipitate out in the joints and soft tissues. This can lead to gout and stones in the urinary tract.

To remove the excess RNA in myco-protein (on average *c*. 10% w/w), the postfermentation broth (cell suspension) is subjected to a 'heat shock' treatment, a procedure specifically developed to reduce the RNA content. The broth is heated briefly to 64 °C to inactivate enzymes that convert cellular ribonucleic acids into monomer nucleotides, which can diffuse out of the cell. This process reduces the nucleic acid content to an acceptable level (< 2% w/w) and also causes the loss of some protein.

Harvesting

The cell suspension described above is then dewatered by centrifugation. The harvested cells, collectively known as myco-protein, are paste-like in consistency and contain around 75% moisture. This biomass has little, if any, intrinsic flavor or odor.

Quorn production process The harvested hyphae have a similar morphology to animal muscle cells, i.e., they are filamentous with a high length/diameter ratio. The product assembly process therefore seeks to reproduce in Quorn™ products the same structural organization as that in natural meats, where the muscle cells are held together by connective tissue. To establish a similar product texture in Quorn™ products, the hyphae are mixed with binders, flavorings, and other ingredients, depending on the final product format, and then heated. This causes the protein binder to gel and hence 'bind' the hyphae together.

The resulting structures are similar to those in meat products, and they also break down in the mouth during chewing in a manner similar to meat products, which accounts for their sensory textural properties.

Efficiency of Production

Animal flesh or fish is the most common source of high-quality protein for humans. However, animals are inefficient at converting proteinaceous vegetable material into meat in terms of both energy input and energy output. The organism described above converts 1 kg of glucose into 1 kg of wet cell mass, representing 136 g of pure protein. The efficiency of conversion is approximately 2:1 on a dry-weight basis, which is considerably higher than animal conversion rates. These may be as high as 10:1 for beef, calculated as dried feed intake against wet carcass weight. Only part of the animal carcass is eaten, whereas all of the myco-protein produced can be consumed.

Properties of Quorn™ Products

Acceptability

Since the microfilaments in Quorn™ myco-protein are almost identical in size to meat fibrils, Quorn™ products have a similar texture and eating quality to their meat equivalents. Table 1 details some of the physical characteristics of Quorn™ myco-protein.

Humans have a long tradition of consuming fungi. Truffles are an expensive delicacy in the Western world, and mushrooms and other cap fungi are eaten across the globe in a variety of forms. Foods such as Camembert cheese are mold-fermented, and in the Far East, microfungi are used to ferment soya beans in the production of miso and tempeh foods. In southern Africa, a variety of foodstuffs are produced from fermented cassava.

Safety

Myco-protein was safety tested according to the guidelines of the Protein Advisory Group of the United Nations. Purified diets were formulated containing myco-protein as the only source of protein at levels of up to 54% (w/w). Control diets contained casein in place of myco-protein, and a stock diet control was included in each trial.

The trials were started in 1970 and included teratology, carcinogenicity, toxicity, and multigeneration studies. Other studies looked at mineral balance, gastrointestinal effects, and estrogenic effects. There were no signs of toxicity or carcinogenicity from feeding myco-protein in any of these trials.

Table 1 Physical characteristics of Quorn™ myco-protein

Physical characteristics	*Typical dimensions*
Length	400–700 μm
Diameter	3–5 μm
Branch frequency	One per 250–300 μm

Data from Marlow Foods, Stokesley, UK.

Measurement of serum levels of uric acid was undertaken in trials conducted at the Massachusetts Institute of Technology in the USA. Eating 20 g of myco-protein per day did not significantly increase serum uric acid levels. Additional daily intake of nucleic acid from the myco-protein did not exceed 2 g. This confirmed that keeping within this limit maintains serum levels of uric acid within the acceptable range for normal subjects and minimizes the risk of deposition of uric acid crystals in kidneys, joints, and other tissues.

Low potential for allergenicity Adverse reactions to foods have been estimated to occur in 1–4% of the population. Feeding microbial proteins at a level of 10 g per day is associated with gastrointestinal symptoms and skin rashes. To test human tolerance to myco-protein, over 400 volunteers were involved in the initial trials conducted at Massachusetts Institute of Technology in the USA.

In a double-blind crossover trial lasting 67 days, the myco-protein was baked into experimental biscuits, each containing 5 g of myco-protein (2.7 g of protein). Identical biscuits were baked for the control group without myco-protein. One hundred volunteers ate two biscuits twice a day, the experimental biscuits providing 20 g myco-protein per day. After 30 days, subjects ate normal foods for 7 days, and then the alternative biscuits for the next 30 days.

No subjects had any gastrointestinal reactions or skin rashes that could be ascribed to ingesting the experimental material. Myco-protein was well tolerated, and the likelihood of adverse reactions was shown to be no greater than that to most common foods.

However, trading experience as the distribution of Quorn™ products increased suggested a need for further investigation. A small number of reactions had been reported, which equated to 1.5 complaints for every million Quorn™ products sold. Cross-reactivity studies showed that Quorn™ myco-protein shares common allergenic determinants with *Aspergillus fumigatus*, *Cladosporium herbarum*, and some with *Alternaria alternata*, suggesting the potential for some mold-allergic patients to react adversely to myco-protein when it is ingested or inhaled. Screening of 33 myco-protein production workers over a 2-year period, for whom any sensitization to myco-protein would occur via inhalation, suggested that two workers had increased RAST binding to myco-protein. However, specific IgE antibody to myco-protein was not significantly raised in any of the 10 consumers tested who had complained of reactions to Quorn™ products. Overall, these results are reassuring as they indicate that the potential for adverse reactions is extremely low. (*See* **Food Intolerance**: Food Allergies.)

Nutritional Value

On a dry-weight basis, myco-protein is 45% protein, 14% fat, and 26% dietary fiber, and on a wet-weight basis, it is 11% protein, 3% available carbohydrate, 6% fiber, 3% fat, 2% ash, and 75% water. Table 2 shows a comparison of the nutritional value of Quorn™ myco-protein with other protein-rich foods.

Quorn products are available as pieces and 'mince' for cooking, and also as a range of convenience products, in which the content of myco-protein may vary between 50 and 90% myco-protein.

Protein content The protein content of myco-protein is above average for fungal mycelia. However, myco-protein contains a significant proportion of nonamino-acid nitrogen, which is in the purine and pyrimidine bases of the nucleic acids and in the glucosamine and galactosamine residues of chitin, a structural component of the cell wall of the filaments and the form of fiber present in myco-protein. As this nitrogen is unavailable, it is more practical to express the protein content as amino-acid nitrogen $\times$ 6.25 rather than total nitrogen $\times$ 6.25. Nonamino-acid nitrogen accounts for about 20–30% of the total nitrogen. (*See* **Protein**: Food Sources.)

Net protein utilization (NPU) values before and after allowance for glucosamine nitrogen reflect the unavailable nitrogen present in chitin. The usual procedure for expressing amino acid content in grams per 16 g of total nitrogen underestimates the amino acid content of myco-protein. Amino acid values will increase after allowing for nonprotein nitrogen.

Although developed as a high-protein food initially, the protein content of Quorn™ myco-protein on a weight-for-weight basis is less than that of lean meat, fish, and poultry, but is comparable with egg (Table 2) and cottage cheese. The NPU is lower than these foods but higher than wheat and beans (Table 3). (*See* **Protein**: Quality.)

Myco-protein provides the complete range of essential amino acids (Table 4). The limiting amino acids are methionine and cystine, with a level of 2.9 g per 100 g of myco-protein. In the UK diet, for example, threonine is the limiting amino acid overall, of which Quorn™ myco-protein is a good source, in contrast to most cereals, which have a low content of lysine and threonine. The relatively high concentration of these amino acids in myco-protein indicates a potential value as a supplement to cereal-based diets.

Unusually, for a nonanimal food, the protein is of high biological value and compares favorably with that of casein. Studies on its biological availability

Table 2 Comparison of the nutritional value of Quorn™ myco-protein with other foods (g per 100 g)

Nutrient	*Myco-protein*[a]	*Semi-skimmed milk*	*Raw egg*	*Braised beef*	*Roast chicken*	*Cooked haricot beans*[b]	*Baked cod*
Energy (kJ)	357	195	617	937	622	406	403
Protein	11.0	3.3	12.5	30.9	24.8	6.6	21.4
Fat: total saturated	3.0	1.6	10.8	11.0	5.4	0.5	1.2
	0.7	1.0	3.4	3.1	1.6	0.1	0.5
Fiber	6.0	0.0	0	0	0	6.1	0

[a]Data from Marlow Foods, Stokesley, UK.
[b]Data from *Vegetables, Herbs and Spices – Fifth Supplement to McCance and Widdowson's The Composition of Foods*. Holland B, Unwin ID and Buss DH (eds), 4th edn. Cambridge: The Royal Society of Chemistry.
All other data from *McCance and Widdowson's The Composition of Foods* (1991) Holland B *et al*. (eds), 5th edn. Cambridge: The Royal Society of Chemistry.

Table 3 Net protein utilization (NPU) of Quorn™ myco-protein compared with other protein-rich foods

Food	*NPU*
Egg	100
Fish	83
Beef	80
Cows' milk	75
Myco-protein[a]	60
Wheat flour	52
Beans	47

[a]Data from Edelman J *et al*. (1983) Myco-protein – a new food. Nutrition Abstracts and Reviews in Clinical Nutrition 53: 1–9.

Table 4 Amino acid profile (g per 100 g) of myco-protein compared to beef

Amino acid	*FAO/WHO reference*[a]	*Myco-protein*[b]	*Beef*
Isoleucine	4.0	5.2	5.0
Leucine	7.0	8.6	7.7
Methionine and cystine	3.5	2.9	3.9
Phenylalanine and tyrosine	6.0	8.9	8.3
Threonine	4.0	5.5	4.3
Tryptophan	1.0	1.6	1.3
Valine	5.0	6.2	5.1
Lysine	5.5	8.3	8.8

[a]Using FAO/WHO (Food and Agriculture Organization, World Health Organization) (1973) scoring profile.
[b]Data from Marlow Foods, Stokesley, UK.
All other data from *McCance and Widdowson's The Composition of Foods* (1991) Holland B *et al*. (eds), 5th edn. Cambridge: The Royal Society of Chemistry.

were carried out early in the development stages of myco-protein.

One way of assessing protein quality is to use the protein digestibility-corrected amino acid (PDCAA) scoring method, which is required by the US Food and Drug Administration (FDA) to be used for most nutritional labeling purposes. This method compares the essential amino acid profile of a food, corrected for digestibility, to the FAO/WHO requirements for 2–5 years olds. This age group is chosen because it has the most demanding pattern for amino acid use other than infants. The PDCAA score for myco-protein is compared with that of other foods in Table 5.

Table 5 Protein digestibility-corrected amino acid score of myco-protein, compared with other food proteins

Protein source	*PDCAA score*
Myco-protein[a]	0.91
Quorn™ pieces[a]	1.0
Roasted chicken (white meat)[b]	1.0
Cooked lentils (canned)[c]	0.52

[a]Calculated from Marlow Foods data, Stokesley, UK.
[b]Calculated from amino acid data in USDA Nutrient Database for Standard Reference, March 12, 1998 (assumes a digestibility equivalent to beef of 94%).
[c]FAO/WHO Joint Report (1989).

Quorn products typically contain between 11 and 15 g protein per 100 g. Most of this protein is from the myco-protein content, but egg albumen and milk proteins, added in product manufacture, also contribute some additional protein.

Fat content The fat content of myco-protein is approximately 3%, which is generally lower than that of most meat and poultry. The cells of the microfungi do not store fat because of their rapid growth rate, and the fats present come mainly from the cell membranes.

The content of saturated fat (0.7%) is particularly low, and no cholesterol is present (Table 6). The ratio of polyunsaturated to saturated fats (3.5) is also favorable when compared with braised beef (0.1) and lean roast chicken meat (0.5). Myco-protein contains no *trans* fatty acids and no cholesterol. In Quorn™ products, small quantities of fat may be added to improve the taste and texture, and these generally contain between 1.5 and 11 g fat per 100 g. (*See* **Fats**: Classification.)

Energy content On account of its low fat content, myco-protein is low in calories. It has fewer calories

Table 6 Typical fatty acid composition of harvested Quorn™ myco-protein

Fatty acid	*Grams per 100 g of myco-protein*
Saturated (S)	
Palmitic acid (C16:0)	0.3
Stearic acid (C18:0)	0.1
Total S	0.4
Monounsaturated	
Oleic acid (C18:1)	0.3
Total	0.3
Polyunsaturated (P)	
Linoleic acid (C18:2)	1.0
α-Linoleic acid (C18:3)	0.4
Total P	1.4
P:S ratio	3.5
Total fatty acids	2.1

Data from Marlow Foods, Stokesley, UK.

Table 7 Vitamin and mineral content of myco-protein compared with chicken (units per 100 g)

Nutrient	*Myco-protein*[a]	*Chicken*[b]
Vitamins		
B_1 (mg)	0.01	0.10
B_2 (mg)	0.23	0.16
Nicotinic acid (mg)	0.35	11.6
B_6 (mg)	0.13	0.42
B_{12} (μg)	0.00	Trace
Pantothenic acid (mg)	0.25	1.20
Biotin (μg)	15.00	2.00
Folic acid (μg)	10.00	12.00
Minerals (mg)		
Calcium	45	10
Phosphorus	254	200
Potassium	100	320
Sodium	4.7	81
Magnesium	45	25
Iron	0.6	0.7
Zinc	10	1.1
Copper	0.5	0.2

[a]Data from Marlow Foods, Stokesley, UK.
[b]Data from *McCance and Widdowson's The Composition of Foods* (1991) Holland B *et al*. (eds), 5th edn. UK: The Royal Society of Chemistry.

than lean chicken (Table 2), making it a useful food to include in calorie-controlled diets and in weight-regulating diets. Quorn™ products generally have a significantly lower energy density than their equivalent meat products.

Fiber content In contrast to animal sources of protein, myco-protein contains nonstarch polysaccharides or dietary fiber (Table 2). The content of fiber (6%) is higher than that in most vegetables, and the amount per serving compares favorably with plant protein foods such as grains, beans, peas, lentils, nuts, and seeds. (*See* **Dietary Fiber**: Properties and Sources.)

The types of fiber present are chitin (35%) and β-glucan (65%), which are present in the cell walls. The fiber is 88% insoluble and 12% soluble.

Vitamins and minerals Myco-protein provides some minerals and some B vitamins, although it is lacking in B_{12} (Table 7). The content of iron is approximately 50% that of chicken and 25% that of beef. The iron is present in an inorganic form, and is therefore not as bioavailable as heme iron. Myco-protein is rich in zinc, containing higher levels than meat, making it a particularly useful source for vegetarians, who generally tend to have low zinc intakes.

Experimentation has shown that the type of fiber in myco-protein is unlikely to cause interference with mineral absorption since it does not contain phytic acid or phytic salts. Research conducted at the Dunn Nutrition Laboratory in Cambridge, England, demonstrated that myco-protein had no significant effect on the absorption of calcium, magnesium, zinc, iron or phosphorus when compared with a polysaccharide-free diet. (*See* **Vitamins**: Overview.)

Known and Potential Functional Health Effects

Lipid-lowering Effects

Studies have demonstrated, under both controlled and free-living conditions, that myco-protein significantly reduces serum total and low-density lipoprotein (LDL) cholesterol levels, and that it may raise high-density lipoprotein (HDL) cholesterol levels. This suggests that myco-protein consumption has a beneficial influence on serum lipid variables, in relation to protection against coronary heart disease. (*See* **Cholesterol**: Factors Determining Blood Cholesterol Levels; Role of Cholesterol in Heart Disease.)

In a metabolic study, nine subjects with mildly raised blood cholesterol levels ate myco-protein–based products in place of meat (control diet – eight subjects) for 3 weeks. The diets were similar in fatty acid and cholesterol content, and in polyunsaturated: saturated ratio. The only significant difference was a 29% higher dietary fiber content in the experimental diet, on account of the high fiber content of the myco-protein products. The experimental diet produced a 13% decrease in total cholesterol compared with no change in the controls. LDL cholesterol fell by 9% in the experimental group compared with a 12% increase in controls. HDL cholesterol rose by 11% in the experimental group compared with an 11% decrease in controls.

Favorable effects of myco-protein products on blood lipid levels in subjects with slightly raised

serum cholesterol concentrations have been confirmed in free-living subjects over an 8-week study period. The experimental group ate biscuits containing myco-protein, and the control group ate a nutrient-balanced biscuit without myco-protein. Total cholesterol was reduced by 0.46 mmol l^{-1} and LDL cholesterol by 0.35 mmol l^{-1} in the control group, compared with reductions of 0.95 mmol l^{-1} and 0.84 mmol l^{-1} in the myco-protein group, respectively.

This suggests that myco-protein has a beneficial effect on dyslipidemia and may be a useful dietary means of improving unfavorable blood lipid profiles. A potential mechanism for this effect is likely to involve the high dietary fiber content of myco-protein.

Blood Glucose Regulation

Myco-protein has also been investigated for potentially beneficial effects on blood glucose and insulin levels, which may benefit diabetics. A study consisting of two single-meals in a crossover design investigated the effects of myco-protein on acute glycemia and insulinemia in 19 healthy individuals. Glycemia was significantly reduced after the meal containing myco-protein (13% at 60 min) compared with the control meal. Insulinemia was also significantly reduced in the experimental group (19% at 30 min and 36% at 60 min) compared with the control.

These preliminary data suggest a potentially beneficial effect of myco-protein for diabetics, although further studies are needed before firm conclusions can be drawn. However the high fiber and low saturated fat content of myco-protein makes it a suitable component of the diabetic diet for other reasons.

Satiety

Several studies have suggested that myco-protein may help to induce feelings of satiety and to reduce hunger and calorie intake at subsequent meals following myco-protein consumption.

In a study (two 3-day study periods), 13 female subjects, all nonrestrained eaters, ate either an isoenergetic meal containing myco-protein or chicken. Energy intake was significantly reduced on the day of the study (by 24%) and the following day (by 16.5%) in those consuming the myco-protein meal compared with the chicken meal. In a further crossover design study, 18 male and female subjects ate an isocaloric lunch containing either myco-protein-based products or chicken, which were similar apart from their dietary fiber content (11 g vs. 3 g, respectively). At the evening *ad libitum* test meal, the energy intake was reduced by 18% in those who had eaten the myco-protein product-based lunch compared with the chicken-based lunch. These studies suggest that myco-protein may play a useful role in weight-reducing diets.

Other Potential Health Effects

Myco-protein has a number of other attributes that may produce beneficial health effects, and these warrant scientific investigation.

Provision of α-linolenic acid Myco-protein is a useful source of α-linolenic acid (ALA), an essential fatty acid belonging to the *n*-3 family. ALA is a precursor of the very-long-chain *n*-3 eicosapentaenoic (EPA) and docosahexaenoic acids (DHA) in the body. EPA and DHA are associated with protection against coronary heart disease through antithrombotic effects in addition to other potential mechanisms. Research suggests a protective effect of ALA itself against CHD, through lowering blood cholesterol concentrations, and beneficial effects on hemostatic factors and heart arrhythmias.

Provision of β-glucans The sources of dietary fiber in Quorn™ myco-protein are chitin (35%) and β-glucan (65%). β-Glucan occurs in three main forms, characterized by linkages that are designated as β 1–3, β 1–4 and β 1–6 linked. Myco-protein contains 16% β 1–3 and β 1–6 glucans on a dry-weight basis, such that a typical Quorn™ product delivers 2–3 g of β glucans per serving. β 1–4 Glucans are present in certain cereal products such as oats, and are associated with lowering blood lipids for which an FDA-approved health claim is permitted. There is less evidence for beneficial effects of β 1–3 and β 1–6 glucans, but reports to date have suggested possible immune-stimulating properties and a cholesterol-lowering effect from these forms of β-glucans present in yeast.

Provision of chitin Chitin is also being investigated for potential cholesterol-lowering effects. Chitin is a type of dietary fiber that is found in mushrooms, fungi, and yeast, as well as in the exoskeleton of insects and in the hard shells of shellfish. Myco-protein contains 8% chitin on a dry-weight basis with a typical serving of a Quorn™ product providing 1–1.5 g of chitin.

Provision of prebiotics β-Glucan and chitin are both being investigated as potential prebiotics, which can be defined as 'nondigestible food ingredients that beneficially affect the host by selectively stimulating the growth and activity of a limited number of

Table 8 The developing role of food

Time period	*Context*
Early to mid-1900s	Focus on adequate vitamin intakes to cure deficiency diseases
1960s–1990s	Eat less of what is bad for you
1990s	Eat more of what is nutritionally good for you
2000s	Foods that are positively good for health beyond their nutrient content

bacteria in the colon that have the potential to improve health.'

Provision of glucosamine Research suggests a potential role for glucosamine in the alleviation of joint pain arising from osteoarthritis. Free glucosamine is not present in any food sources, but polymeric glucosamine is present in chitin. Myco-protein therefore contains 8% glucosamine in polymeric form (chitin). There is preliminary evidence that polymeric glucosamine may also have a beneficial role, as it can potentially provide sustained release of glucosamine, but this requires further investigation.

Potential of Myco-protein

Though originally researched and developed as a high-protein food, the predicted protein gap did not materialize. However, the ever-changing climate of consumer demands for food products meant that the Quorn™ range of products could be successfully marketed in a climate where demand for acceptable meat-alternatives was growing. This was a response to an increasing emphasis on healthy eating, for which, in the 1980s, the context was on removing elements from the diet considered to be 'bad' for health (Table 8).

However, as we have entered the new millennium, and the role of food has developed further, there has been a shift in emphasis towards eating products that are positively good for health (Table 8). Within the context of this focus on the functional properties of foods beyond their nutritional value (so-called 'functional foods'), Quorn™ products would appear to potentially have a number of positive attributes that may be of functional value in this context, and these warrant further investigation.

See also: **Amino Acids**: Properties and Occurrence; **Fats**: Digestion, Absorption, and Transport; **Fatty Acids**: Properties; **Food Intolerance**: Food Allergies; **Functional Foods**; **Iron**: Properties and Determination; **Lipoproteins**; **Prebiotics**; **Probiotics**; **Protein**: Chemistry; **Single-cell Protein**: Algae; Yeasts and Bacteria; **Slimming**: Slimming Diets; Metabolic Conseqences of Slimming Diets and Weight Maintenance; **Zinc**: Properties and Determination

Further Reading

Anderson C (1974) The growth of micro-fungi on carbohydrates. In: Tanenbaum SR and Wang DIC (eds) *Single Cell Protein II*, pp. 314–329. Cambridge, MA: MIT Press.

Edelman J, Fewell A and Solomons GL (1983) Myco-protein – a new food. *Nutrition Abstracts and Reviews in Clinical Nutrition* 53: 1–9.

Tee RD, Gordan DJ, Welch JA and Newman Taylor AJ (1993) Investigation of possible adverse allergic reactions to myco-protein ('Quorn™'). *Clinical and Experimental Allergy* 23: 257–260.

Trinci APJ (1994) Evolution of the Quorn™ myco-protein fungus, *Fusarium graminearum* A3/5. *Microbiology* 140: 2181–2188.

Turnbull WH, Leeds AR and Edwards DG (1990) Effect of mycoprotein on blood lipids. *American Journal of Clinical Nutrition* 52: 646–650.

Turnbull WH, Leads AR and Edwards DG (1992) Mycoprotein reduces blood lipids in free-living subjects. *American Journal of Clinical Nutrition* 55: 415–419.

Turnbull WH and Ward T (1995) Mycoprotein reduces glycemia and insulinemia when taken with an oral-glucose-tolerance test. *American Journal of Clinical Nutrition* 61: 135–140.

Turnbull WH, Walton J and Leads AR (1993) Acute effects of myco-protein on subsequent energy intake and appetite variables. *American Journal of Clinical Nutrition* 58: 507–512.

MYCOTOXINS

Contents

Classifications

L B Bullerman, University of Nebraska-Lincoln, Lincoln, NE, USA

Introduction

Mycotoxins are a group of structurally diverse, naturally occurring chemical substances produced by molds (microfungi). The term 'mycotoxin' is derived from the Greek word '*Mykes*', which means fungus, and the Latin word '*toxicum*,' which means toxin or poison. Thus, mycotoxin literally means 'fungus toxin' or 'fungus poison,' and its use is restricted to describing the metabolites of microfungi, as opposed to the toxic principles of certain macrofungi, i.e., mushrooms.

Toxicity and Biological Effects of Mycotoxins

In general, mycotoxins produce a number of adverse effects in a range of biological systems, including microorganisms, plants, animals, and humans. The toxic effects of mycotoxins in humans and animals, depending upon dose, may include (1) acute toxicity and death as a result of exposure to high amounts of a mycotoxin, (2) reduced milk and egg production and lack of weight gain in food-producing animals from subchronic exposure, (3) impairment, stimulation, or suppression of immune functions and reduced resistance to infections from chronic exposures to low levels of toxins, and (4) tumor formation, cancers, and other chronic diseases from prolonged exposure to very low levels of a toxin. In addition, mycotoxins may be mutagenic, capable of inducing mutations in susceptible cells and organisms, and teratogenic, capable of causing deformities in developing embryos. Other manifestations of mycotoxins that can affect the food supply in economic terms are reduced growth rates and increased reproductive problems in food-producing animals and livestock. (*See* **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; **Mutagens**.)

Mycotoxin-producing Molds

Most molds that produce the mycotoxins of greater concern can be found in three main genera, *Aspergillus*, *Penicillium*, and *Fusarium*. Organisms of the *Aspergillus* and *Penicillium* genera tend to be saprophytic and often attack commodities while in storage, though some aspergilli can also invade in the field. The genus *Fusarium* contains a number of plant pathogenic species as well as saprophytic types. A number of factors affect mycotoxin production by molds, including moisture, relative humidity, temperature, substrate, pH, competitive and associate growth of other fungi and microorganisms, and stress on the plant such as drought and damage to seed coats from hail, insects, and mechanical harvesting equipment. The major commodities that are susceptible to contamination with mycotoxins include corn (maize), wheat, barley, peanuts, cottonseed, and some tree nuts (Table 1). Dairy products, primarily milk, can become contaminated as a result of feeding contaminated feed to dairy animals. Cheese made from contaminated milk will also be contaminated. (*See* **Cereals**: Dietary Importance; **Spoilage**: Molds in Spoilage.)

Specific Mycotoxins

Aflatoxins

Aflatoxins are produced primarily by some strains of *Aspergillus flavus* and most, if not all, strains of *A. parasiticus* and *A. nomius*. There are four main aflatoxins B_1, B_2, G_1, and G_2, plus two additional aflatoxins that are of significance, M_1 and M_2 (Figure 1). The M toxins were first isolated from the milk of lactating animals fed aflatoxin preparations; hence the M designation.

Aflatoxins are potent liver toxins, in all animals in which they have been tested, and carcinogens in some species. Aflatoxin B_1 is the most toxic of the group. Effects of aflatoxins in animal tests vary with dose, length of exposure, species, breed, and diet or nutritional status. These toxins may be lethal when consumed in large doses; sublethal doses produce a

Table 1 Commodities in which mycotoxin contamination has been found and the resulting effects on animals and humans

		Effects of mycotoxins	
Mycotoxin	*Commodities contaminated*	*Affected species*	*Pathological effects*
Aflatoxins (B_1, B_2, G_1, G_2, M_1, M_2)	Peanuts, corn (maize), wheat, rice, cottonseed, copra, nuts, various foods, milk, eggs, cheese	Birds: Duckling, turkey poult, pheasant chick, mature chicken, quail Mammals: Young pigs, pregnant sows, dog, calf, mature cattle, sheep, cat, monkey, human Fish Laboratory animals	Hepatotoxicity (liver damage) Bile-duct hyperplasia Hemorrhage: Intestinal tract Kidneys Carcinogenesis (liver tumors)
Citrinin	Cereal grains (wheat, barley, corn (maize), rice)	Swine, dog, laboratory animals	Nephrotoxicity (tubular necrosis of kidney) Porcine nephropathy
Cyclopiazonic acid	Corn (maize), peanuts, cheese, kodo millet	Chicken, turkey, swine, rat, guinea-pig, human (?)	Muscle necrosis Intestinal hemorrhage and edema Oral lesions
Fumonisins (FB_1, FB_2, FB_3)	Corn (maize)	Horses and other equidae	Equine leukoencephalomalacia (ELEM), death
		Swine	Porcine pulmonary edema (PPE) and hydrothorax, death
		Rats	Liver cancer
		Monkey	Atherosclerotia, hypercholesterolemia
		Human	Possible esophageal cancer
Moniliformin	Corn (maize), wheat, triticale, rye, oats	Chicken	Cardiotoxicity myocardial degeneration Muscular weakness Death
		Human	Possible myocardial impairment (Keshan disease)
		Plants (corn, tobacco)	Necrosis, chlorosis
Ochratoxin A	Cereal grains (wheat, barley, oats, corn (maize), dry beans, moldy peanuts, cheese, tissues of swine	Swine, dog, duckling, chicken, rat, human	Nephrotoxicity (tubular necrosis of kidney) Porcine nephropathy Mild liver damage Enteritis Teratogenesis Carcinogenesis (kidney tumors)
Patulin	Moldy feed, rotten apples, apple juice, wheat straw residue	Birds: Chicken, chicken embryo, quai Mammals: Cat, cattle, mouse, rabbit, rat Others: Brine shrimp, guppy, zebra fish larvae	Edema: Brain Lungs Hemorrhage: Lungs Capillary damage: Liver Spleen Kidney Paralysis of motor nerves Convulsions Carcinogenesis Antibiotic

Continued

Table 1 Continued

Mycotoxin	*Commodities contaminated*	*Effects of mycotoxins* *Affected species*	*Pathological effects*
Penicillic acid	Stored corn (maize), cereal grains, dried beans, moldy tobacco	Mouse, rat, chicken embryo, quail, brine shrimp	Liver damage (fatty liver, cell necrosis) Kidney damage Digitalis-like action on heart Dilated blood vessels Antidiuretic Edema in rabbit skin Carcinogenesis Antibiotic
Penitrem	Moldy cream cheese, English walnuts, hamburger bun, beer	Dog, mouse, human	Tremors, death, incoordination, bloody diarrhea
Sterigmatocystin	Green coffee, moldy wheat, Dutch cheeses	Mouse, rat	Carcinogenesis Hepatotoxin
Trichothecenes (T-2 toxin, diacetoxyscirpenol, neosolaniol, nivalenol, diacetylnivalenol, deoxynivalenol, HT-2 toxin, fusarenon X)	Corn (maize), wheat, commercial cattle feed, mixed feed	Swine, cattle, chicken, turkey, horse, rat, dog, mouse, cat, human	Digestive disorders (emesis, diarrhea, refusal to eat) Hemorrhage (stomach, heart, intestines, lungs, bladder, kidney) Edema Oral lesions Dermatitis Blood disorders (leucopenia) Food-borne illness (nausea, vomiting, abdominal pain, diarrhea) Immune disfunction (suppression and stimulation)
Zearalenone	Corn (maize), moldy hay, pelleted commercial feed	Swine, dairy cattle, chicken, turkey, lamb, rat, mouse, guinea-pig	Estrogenic effects (edema of vulva, prolapse of vagina, enlargement of uterus) Atrophy of testicles Atrophy of ovaries, enlargement of mammary glands Abortion

chronic toxicity, and low levels of chronic exposure result in cancers, primarily liver cancer in a number of animal species. In general, young animals of any species are more susceptible to the acute toxic effects of aflatoxins than are older animals of the same species. Susceptibility also varies between species. Of all the mycotoxins, the aflatoxins are of greatest concern because they are highly toxic and potently carcinogenic. Mold growth and aflatoxin production are favored by warm temperatures and high humidity, typical of tropical and subtropical regions. (*See* **Aflatoxins**.)

Sterigmatocystin

Sterigmatocystin is produced by several species of *Aspergillus*, *Penicillium luteum*, and a *Bipolaris* species. Chemically, sterigmatocystin resembles aflatoxins and is thought to be a precursor in the biosynthesis of aflatoxin. The acute toxicity of sterigmatocystin is low, and the main concern is that it is carcinogenic, about one-tenth as potent a carcinogen as aflatoxin B_1. Sterigmatocystin has been detected at lower levels in green coffee, moldy wheat, and rind of hard Dutch cheese.

Ochratoxins

Ochratoxins are a group of related compounds that are produced by *Aspergillus ochraceus* and related species, *Aspergillus niger*, as well as *Penicillium verrucosum* and certain other *Penicillium* species. The main mycotoxin in this group, ochratoxin A, is a potent toxin that causes kidney damage in rats, dogs and swine (Figure 2). Ochratoxin is thought to be involved in a disease of swine known as porcine nephropathy, which has been associated with the feeding of moldy barley. Ochratoxin has also been

reported to be teratogenic to mice, rats and chicken embryos, and has been suggested as a possible causative factor, though never proven, in a human disease known as 'Balkan endemic nephropathy'.

Citrinin

Citrinin is a yellow-colored compound that is produced by several *Penicillium* species as well as *Aspergillus* species (Figure 3). Like ochratoxin A, citrinin causes kidney damage in laboratory animals similar to swine nephropathy. Citrinin may be involved with ochratoxin A in cases of swine nephropathy in Denmark. However, the toxicity of citrinin is low compared with ochratoxin, although possible synergistic activity with ochratoxin A cannot be ruled out.

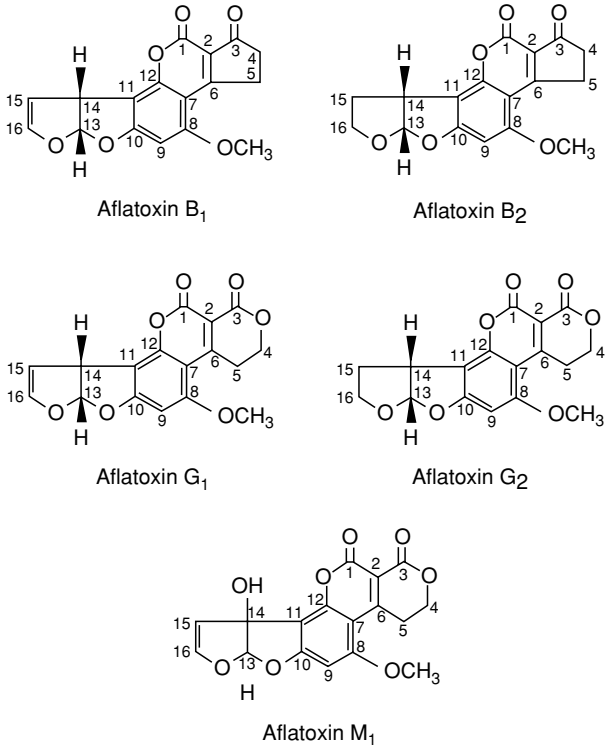

Figure 1 Chemical structures of aflatoxins.

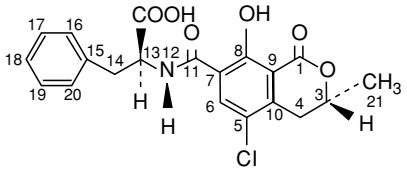

Figure 2 Chemical structure of ochratoxin A.

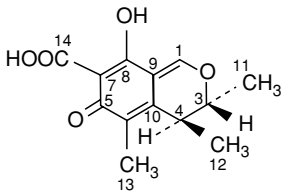

Figure 3 Chemical structure of citrinin.

Patulin

Patulin is toxic to many biological systems, including bacteria, mammalian cell cultures, higher plants, and animals, but its role in causing animal and human disease is unclear. Patulin has a lactone structure and is carcinogenic when injected intradermally into mice (Figure 4). Patulin is produced by numerous *Penicillium* and *Aspergillus* species and *Byssochlamys nivea*. *Penicillium expansum*, which commonly occurs in rotting apples, produces patulin. Patulin is of some public health concern because of its potential carcinogenic properties, and because it has been found in commercial apple juice and other apple products. Patulin appears to be unstable in grains, cured meats, and cheese. When administered orally to rats, patulin showed no toxicity or carcinogenicity.

Penicillic Acid

Penicillic acid has a low oral toxicity. The concern about penicillic acid in foods is based on the fact that the compound bears structural similarity to known carcinogens such as patulin, and has in fact been shown to be carcinogenic to rats when injected subcutaneously (Figure 5). However, the potencies of penicillic acid and patulin as carcinogens are much lower than aflatoxins. When given in lethal doses, penicillic acid caused fatty-liver degeneration in quail and liver-cell necrosis in mice. Mixtures of penicillic acid with ochratoxin A are synergistic and cause

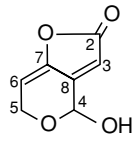

Figure 4 Chemical structure of patulin.

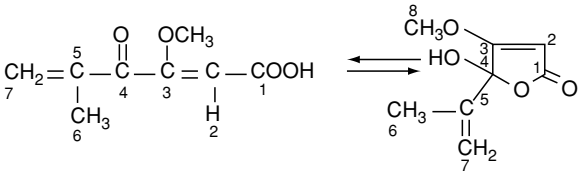

Figure 5 Chemical structure of penicillic acid as the straight-chain acid and the lactone configuration.

death in mice. Pharmacologically, penicillic acid dilates blood vessels and has antidiuretic effects. However, penicillic acid is similar to patulin in its rapid reaction with sulfhydryl-containing compounds in foods to form nontoxic products. Penicillic acid is produced by strains of *A. ochraceus* and related species and *Penicillium aurantiogriseum* and other species. Some strains of *A. ochraceus* are capable of producing penicillic acid along with ochratoxin A. Penicillic acid has been found in large quantities in corn with a high moisture content stored at low temperatures.

Cyclopiazonic Acid

Cyclopiazonic acid (CPA) is produced by several molds that commonly occur on agricultural commodities or are used in certain food fermentations (Figure 6). CPA has been reported to be produced by *A. flavus*, *A. versicolor*, and *A. tamarii* as well as several *Penicillium* species, some of which are used in the production of fermented sausages in Europe. Other molds used in the food fermentations that have been reported to produce CPA are *Penicillium camemberti*, used to produce Camembert cheese, and *Aspergillus oryzae*, used to produce fermented soy sauces. CPA has been reported naturally occurring in corn and peanuts, and a type of millet (kodo) that reportedly caused a human intoxication in India. It is also possible that CPA was involved along with aflatoxins in 'turkey X' disease in England in 1960, since some isolates of *A. flavus* produced both aflatoxins and CPA. CPA has been shown to affect rats, dogs, pigs, and chickens. Clinical signs of intoxication include anorexia, diarrhea, pyrexia, dehydration, weight loss, ataxia, immobility, and extensor spasm at time of death. Histopathological changes in CPA-exposed animals include alimentary tract hyperemia, hemorrhage, and focal ulceration. Focal necrosis can be found in liver, spleen, kidneys, pancreas, and myocardium. In broiler chicks given CPA, skeletal muscle degeneration characterized by myofibular swelling and fragmentation has been observed. About 50% of a dose of CPA given orally or intraperitoneally to rats or chickens is distributed to skeletal muscle within 3 h. CPA has the ability to chelate metal cations. Chelation of such cations as calcium, magnesium, and iron may be an important mechanism of toxicity of CPA.

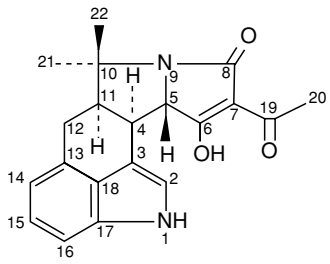

Figure 6 Chemical structure of cyclopiazonic acid.

Mycophenolic Acid, β-Nitropropionic Acid, Tremorgens (Penitrem), and Rubratoxin

Many toxic compounds have been obtained from mold cultures; however, not all have been shown to cause disease in humans or animals. Other mycotoxins such as mycophenolic acid, β-nitropropionic acid, and tremorgens have been reported but have not been studied extensively. Tremorgenic mycotoxins called penitrems have been reported to have caused poisonings of dogs that consumed moldy cream cheese, moldy walnuts, and other moldy debris (Figure 7). The toxins caused severe muscle tremors, uncoordinated movements, and generalized seizures and weakness in the dogs. The disease can also occur in cattle, where it is called staggers. Tremorgenic mycotoxins can be produced by fungi in the genera *Aspergillus*, *Penicillium*, *Claviceps*, and *Acremonium*. Mycophenolic acid (Figure 8) and β-nitropropionic acid have been associated with cheeses produced in Europe and are believed to be antibiotic substances of low oral toxicity. Rubratoxin B (Figure 9) produced hepatic degeneration, centrilobular necrosis, and hemorrhage of the liver and intestine when given to experimental animals. However, no natural occurrence of disease caused by this toxin has been documented, though it is suspected of causing a hepatotoxic, hemorrhagic disease of cattle and pigs fed

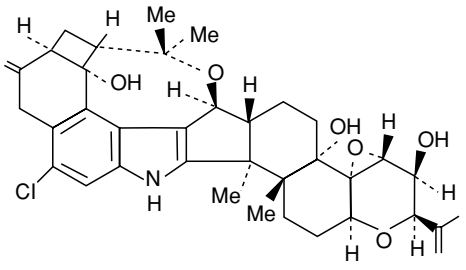

Figure 7 Chemical structure of penitrem A.

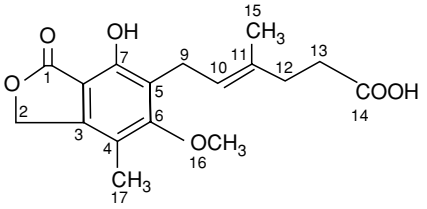

Figure 8 Chemical structure of mycophenolic acid.

moldy corn. Rubratoxin is produced by *Penicillium rubrum* and may exert a synergistic effect with aflatoxins.

Zearalenone

Zearalenone (Figure 10), an estrogenic compound also known as F-2 toxin, causes vulvovaginitis and estrogenic responses in swine. Zearalenone is produced by *Fusarium* species. Zearalenone has been found to occur naturally in corn with a high moisture content in late autumn and winter, primarily from the growth of *Fusarium graminearum* and *Fusarium culmorum*. Although the compound is not especially toxic, 1–5 p.p.m. is sufficient to cause physiological responses in swine. Zearalenone can be transmitted to piglets in sows' milk and cause estrogenism in the young pigs. Zearalenone has been found in moldy hay, high-moisture corn, corn infected before harvest, and pelleted feed rations. The involvement of zearalenone in human toxicoses has not been confirmed, but it is classified as an endocrine disrupter and is considered potentially hazardous for humans. The formation of zearalenone is favored by high humidity and fluctuating temperatures, followed by low temperatures. These conditions often occur in temperate regions during autumn harvest.

Trichothecenes

Trichothecenes (Figure 11) are a family of closely related compounds produced by several *Fusarium* species. There are more than 20 naturally occurring compounds produced by *Fusarium* species, which have similar structures, including T-2 toxin, diacetoxyscirpenol, neosolaniol, nivalenol, diacetylnivalenol, deoxynivalenol (DON, vomitoxin), HT-2 toxin, and fusarenon X. DON has been implicated in a disease known as moldy corn toxicosis of swine, symptoms of which include refusal to eat (refusal factor), lack of weight gain, digestive disorders, and diarrhea, ultimately leading to death. T-2 toxin is quite toxic to rats, trout, and calves. Large doses of T-2 toxin fed to chickens result in severe edema of the body cavity and hemorrhage of the large intestine, along with neurotoxic effects, oral lesions, and, finally, death. T-2 toxin is also thought to be one of the toxins involved in the human disease alimentary toxic aleukia (ATA) and has been implicated as a component of the so-called 'yellow rain' samples collected from South East Asia in the early 1980s. T-2 toxin causes severe dermal responses in rabbits, rats, and other animals, including humans, when applied to the skin. However, it is not thought to be carcinogenic. Although T-2 toxin is very toxic, its natural contamination in foods and feeds appears to be low. Deoxynivalenol causes vomiting in animals and has been suspected as a cause of foodborne illness in India in which nausea, vomiting, and diarrhea were the symptoms. Short-term feeding trials with animals suggest low acute toxicity, but other evidence indicates that deoxynivalenol may have a teratogenic potential. Unlike T-2 toxin, deoxynivalenol contamination in commodities, especially wheat and corn, is significant in certain years. Deoxynivalenol is produced by *F. graminearum* and *F. culmorum* and is the most commonly occurring trichothecene. The trichothecenes are also of concern because chronic exposure to low levels of these compounds may cause immunosuppression or immunostimulation-like illnesses in humans and animals.

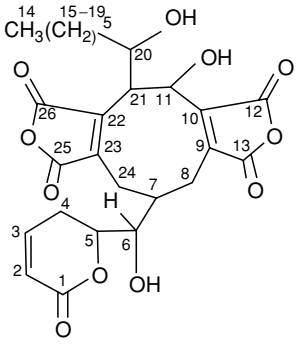

Figure 9 Chemical structure of rubratoxin B.

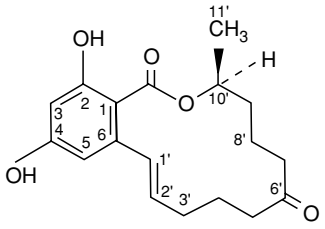

Figure 10 Chemical structure of zearalenone.

Fumonisins

Fumonisins (Figure 12) are a group of compounds produced primarily by *Fusarium verticillioides* (formerly known as *Fusarium moniliforme*) and *F. proliferatum*. In animals, fumonisins have been shown to cause several diseases such as equine leucoencephalomalacia, porcine pulmonary edema, and liver cancer in rats. While not as potent a liver carcinogen as aflatoxins, it has been suggested that the lifetime carcinogenic potential of fumonisin B_1 in the rat falls somewhere between carbon tetrachloride and dimethylnitrosamine. In terms of human disease, fumonisins are of concern because they have been linked to esophageal cancer. It has been shown that there is a significant correlation between the

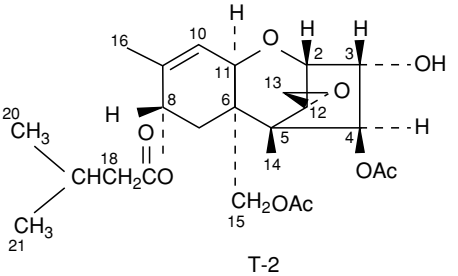

T-2

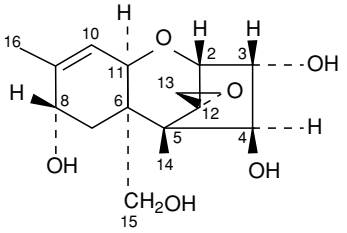

T-2 Tetraol

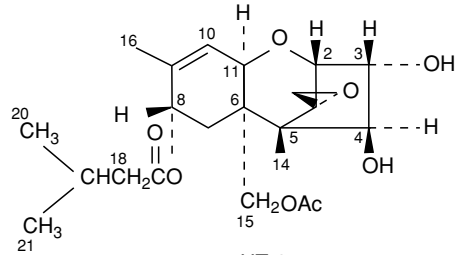

HT-2

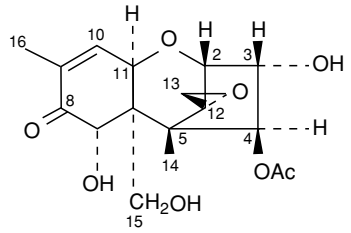

Fusarenon X

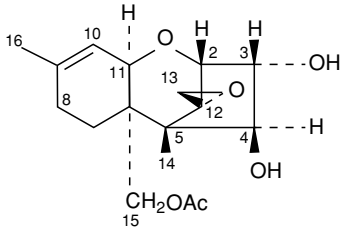

Monoacetoxyscirpenol

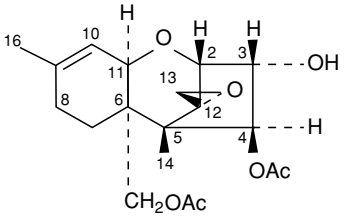

Diacetoxyscirpenol

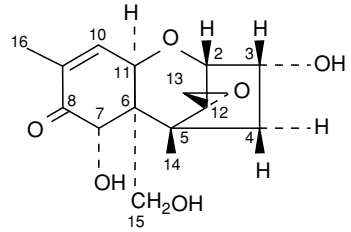

Deoxynivalenol

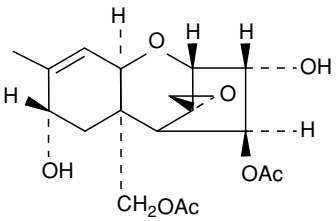

Neosolaniol

Figure 11 Chemical structure of several trichothecenes.

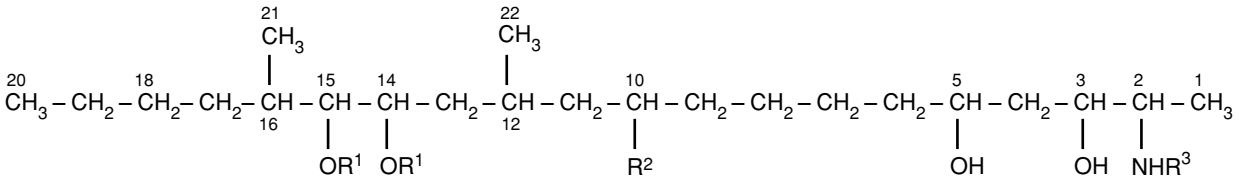

Fumonisin A_1: $R^1 = COCH_2CH(CO_2H)CH_2CO_2H$; $R^2 = OH$; $R^3 = C^{23}OC^{24}H_3$

Fumonisin A_2: $R^1 = COCH_2CH(CO_2H)CH_2H$; $R^2 = H$; $R^3 = COCH_3$

Fumonisin B_1: $R^1 = COCH_2CH(CO_2H)CH_2CO_2H$; $R^2 = OH$; $R^3 = H$

Fumonisin B_2: $R^1 = COCH_2CH(CO_2H)CH_2CO_2H$; $R^2 = R^3 = H$

Figure 12 Chemical structures of fumonisins.

consumption of corn contaminated with *F. verticillioides* and fumonisins and esophageal cancer in humans in the Transkei region of South Africa. Fumonisins and *F. verticillioides* have also been linked to corn (maize) consumption and a high incidence of esophageal cancer in certain provinces of northern China. *F. verticillioides* commonly occurs in corn-growing regions of the world and has been said to be found virtually everywhere that corn is grown. *F. verticillioides* invades the corn plant and can exist there as an endophyte, and may not be evident as a surface contaminant or infestation. Infection of the corn kernel by *F. verticillioides* may occur by invasion through the silk, fissures in the kernel pericarp and/or through systemic infection of the plant.

Moniliformin

Moniliformin (Figure 13) was first reported to be produced by *F. moniliforme* isolated from corn. While the toxin apparently was named after this organism, the name has turned out to be a misnomer, since subsequent work has shown that most strains of *F. moniliforme* (now called *F. verticillioides*) do not produce moniliformin, or are only weak producers. The toxin is produced by *F. proliferatum* and *F. subglutinans*, as well as other *Fusarium* species. Moniliformin is highly toxic when given orally to experimental animals, causing rapid death without severe cellular damage. One-day-old cockerels and mice have been shown to be quite sensitive to the toxin. Clinical lesions observed include acute degenerative lesions in the myocardium and other tissues. Moniliformin has also been suggested as a possible cause of Keshan disease, a human myocardial impairment occurring in rural areas of China and South Africa where there is a high consumption of corn (maize) that has been shown to be contaminated with moniliformin.

Potential Toxicity of *Penicillium roqueforti*

Penicillium roqueforti has been shown to produce several toxic compounds, including roquefortine, PR toxin, and festuclavine (Figure 14). Toxicities of PR toxin and roquefortine are low. Roquefortine is a neurotoxin that reportedly causes convulsive seizures, liver damage, and hemorrhage in the digestive tract in mice. However, repeated studies have failed to reproduce these results. Roquefortine has been recovered from blue cheese and was associated with the mold mycelia rather than the nonmoldy areas of the cheese. A toxic factor in the fat of Roquefort cheese that caused severe injury to the liver and other organs of rats has been reported. Atypical, wild strains of *P. roqueforti* have been shown to produce patulin and penicillic acid simultaneously, patulin alone, patulin plus citrinin, and mycophenolic acid. The significance of the various toxins produced by *P. roqueforti* to public health is not clear. Patulin, penicillic acid, and citrinin have been observed only in wild-type isolates of the organism and not in commercial strains, nor in any cheese produced by commercial strains. As such, the wild isolates represent no greater significance than any other toxinogenic isolates of other species. The significance of PR toxin, mycophenolic acid, the roquefortines, and related alkaloids to human health is likewise unclear,

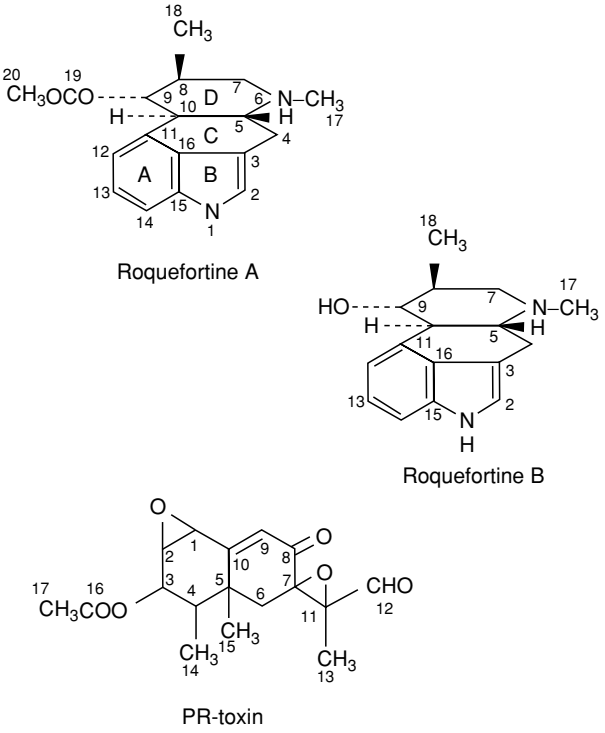

Figure 14 Chemical structures of some *P. roqueforti* toxins.

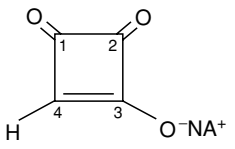

Figure 13 Chemical structure of moniliformin.

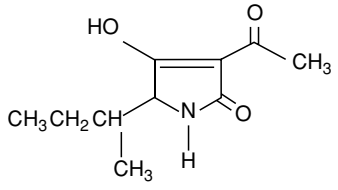

Figure 15 Chemical structure of tenuazonic acid.

particularly in view of the limited toxicological information available on these compounds. PR toxin apparently reacts with cheese components and is neutralized. The fact that blue-veined cheeses have been consumed for centuries without any apparent ill effect suggests that the hazard to human health is minimal or nonexistent. (*See* **Alkaloids**: Properties and Determination.)

Alternaria Toxins

Alternaria species, including *A. alternata, A. citrii, A. tenuissima*, and others produce several toxic compounds, alternariols, altenuene, tentoxin and tenuazonic acid (Figure 15). These organisms are common in many foods and grains. *Alternaria* species require high-moisture conditions and tend to be found in foods that are high in moisture such as grains prior to harvest and fruits and vegetables. The *Alternaria* toxins can be found in grains when drying in the field and harvest are delayed by rain, high humidity, or early frost. *Alternaria* species are most common in sorghum grain, but the toxins have only been found in heavily weathered sorghum. Postharvest occurrence of *Alternaria* species in fruits and vegetables is more common because the moisture content remains high after harvest. *Alternaria* infection of fruits and vegetables has been observed in apples, oranges, tomatoes, and bell peppers. *Alternaria* toxins have been detected in oranges, tomatoes, tomato paste, and commercial apple products. *Alternaria* toxins are toxic to *Bacillus mycoides* and HeLa cells. The toxins, however, are only weakly toxic to mice and do not appear to be toxic to rats or chicks when administered as single, purified compounds. There is some evidence that mixtures of the compounds may be more toxic. The *Alternaria* toxins are also phytotoxins which affect various plants. *Alternaria alternata* f. sp. *lycopersici*, which is a pathogen of tomatoes, produces a host specific phytotoxin known as AAL toxin which is nearly identical in structure to fumonisins.

Ergot

Ergot is a disease of plants, particularly small grains such as rye and barley and other grasses, which is caused by species of *Claviceps*, in particular *C. purpurea, C. paspalli*, and *C. fusiformis*. These fungi invade the female sex organs of the host plant and replace the ovary with a mass of fungal tissue known as a sclerotium. The sclerotia, also called ergots, are about the same size and density as the grain kernels and tend to go with the grain when harvested. The sclerotia contain alkaloids that are produced by the fungus. The alkaloids ergotamine (Figure 16), ergosine, and others are derivatives of

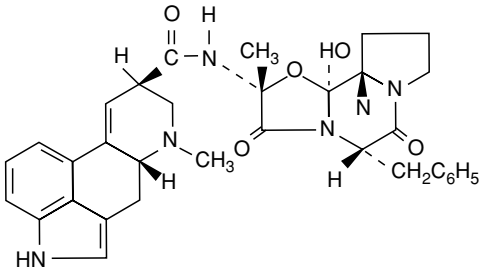

Figure 16 Chemical structure of ergotamine.

lysergic acid, and cause disease in animals and humans. The disease is manifested by a sensation of cold hands and feet followed by an intense burning sensation. As the disease progresses, the extremities may become gangrenous and necrotic, and in animals, sometimes the extremities are sloughed. In severe cases, death may ensue. Ergotism, also known as St. Anthony's fire, reached epidemic proportions during the Middle Ages. At that time, the cause of the disease was not known, but it was most probably associated with bread made from flours of rye and other grains that were infested with ergot sclerotia. In recent times, outbreaks involving humans have occurred in Africa and India. Outbreaks of animal poisonings still occur in areas where rye, barley, and other susceptible small grains and grasses are grown.

Stability of Mycotoxins in Foods

Most of the information on mycotoxin contamination of commodities concerns agricultural products such as grain (Table 1). Therefore, much of what is known about the stability of mycotoxins pertains to the effects of storage, food processing, and cooking. Aflatoxins tend to be stable to moderately stable during storage under most conditions and in most food processes. Various food processes can reduce the amount of aflatoxin by destruction or removal, but usually, some percentage of the toxin will persist and carry into the finished food. Aflatoxins are stable in peanut materials at room temperature, but are partially destroyed by roasting; however, destruction is not complete. In the wet milling of corn, aflatoxin, if present, will be found primarily (over 97%) in the steep water and fiber fractions, but little (1% or less) in the starch. Aflatoxins may be reduced in the baking process, but are not completely destroyed. The manufacture of corn flakes has been shown to reduce aflatoxin levels in flaking grits by 75–80%. Similarly, some sterigmatocystin will be lost during roasting of coffee beans, but some will also remain. Ochratoxin is not eliminated from grain by cleaning or milling, after which it appears equally divided between flour and bran fractions. Ochratoxin will persist through

the baking process, manufacture of corn flakes and through the cooking of mash for beer. However, roasting of green coffee beans has been shown to completely destroy ochratoxin. Patulin and penicillic acid appear to be moderately stable in processed apple products but are unstable in grains, grain-based foods, meats, and cheese. Patulin and penicillic acid react with sulfydryl and amino compounds to form nontoxic adducts. Patulin also appears to be destroyed by the alcoholic fermentation of apple juice and by wine-making. Deoxynivalenol (DON) in wheat milled into flour will be found primarily in the bran fraction but is not destroyed. Also, baking does not destroy DON, though there may be some reduction of DON levels. Likewise, extrusion processing of cereal grains may give some reduction of DON under conditions of high temperatures and pressures. In the wet milling of zearalenone-contaminated corn, like aflatoxins, zearalenone concentrates in the fractions used for animal feed and not in the starch. Likewise, in wet milling of T-2 toxin-contaminated corn, most of the T-2 toxin is found in the steep water, germ, gluten, and fiber, and only 6–8% is found in the starch. Extrusion processing of corn (maize) contaminated with zearalenone has been shown to reduce zearalenone concentrations by 65–83%. Ethanol fermentation of fumonisin-contaminated corn results in little degradation of fumonisin during fermentation, with most of the toxin being found in the distillers' grains, thin stillage and distillers' solubles fractions. None of the fumonisin is found in the distilled alcohol. Fumonisins are stable in baked or canned corn (maize) products, with little or no loss. Roasting at temperatures above 200 °C result in apparent loss of fumonisins, whereas the corn flake process and extrusion processing result in low to moderate reductions of fumonisins. However, if glucose is added, the reduction of fumonisins increases dramatically to more than 80% in both the corn flake process and extrusion processes. Thus, in most cases, mycotoxins are not completely destroyed by processing and cooking. Some destruction occurs, the extent of which depends on the process and the toxin present, though destruction usually is not complete. Also, mycotoxins may tend to concentrate in certain fractions during milling and fermentation, depending upon the type of mycotoxin and process used.

See also: **Aflatoxins**; **Alkaloids**: Properties and Determination; **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; **Cereals**: Dietary Importance; **Food Safety**; **Immunology of Food**; **Mutagens**; **Mycotoxins**: Occurrence and Determination; Toxicology; **Spoilage**: Molds in Spoilage

Further Reading

Betina V (ed.) (1984) *Mycotoxins: Production, Isolation, Separation and Purification*. New York: Elsevier.

Betina V (ed.) (1989) *Mycotoxins: Chemical, Biological and Environmental Aspects*. New York: Elsevier.

Bullerman LB (1986) *Mycotoxins and Food Safety. A Scientific Status Summary by the Institute of Food Technologists' Expert Panel on Food Safety and Nutrition*. Chicago, IL: Institute of Food Technologists.

Bullerman LB (2000 Fusaria and toxigenic molds other than aspergilli and penicillia. Chapter 23. In: Doyle MP, Beuchat LR and Montvile TJ (eds) *Food Microbiology: Fundamentals and Frontiers*. Washington, DC: ASM Press.

CAST (2000) *Mycotoxins. Risks and Impacts in Plant and Animal Systems*. Ames, IA: Council for Agricultural Science and Technology.

Eaton DL and Groopman JD (eds) (1994) *The Toxicology of Aflatoxins, Human Health, Veterinary and Agricultural Significance*. New York: Academic Press.

International Agency for Research on Cancer (1987) Aflatoxins. In: *IARC Monograph on the Evaluation of Carcinogenic Risks to Humans*, Supplement 7, pp. 83–87. Lyon: IARC.

Jackson LS and Bullerman LB (1999) Effect of processing on *Fusarium* mycotoxins. *Advances in Experimental Medicine and Biology* 459: 243–261.

LeBars J and Galtier P (eds) (1998) Mycotoxins in the food chain: processing and toxicological aspects. *Revue d Elevage et de Medecine Veterinaire des Pays Tropicaux* 149: 469–715.

Miller JD and Trenholm HL (eds) (1994) *Mycotoxins in Grain. Compounds Other Than Aflatoxins*. St. Paul, MN: Eagan Press.

Sharma RP and Salunkhe DK (eds) (1991) *Mycotoxins and Phytoalexins*. Boca Raton, FL: CRC Press.

Sinha KK and Bhatnagar D (eds) (1998) *Mycotoxins in Agriculture and Food Safety*. New York: Marcel Dekker.

Smith JE and Henderson RS (eds) (1991) *Mycotoxins and Animal Foods*. Boca Raton, FL: CRC Press.

Occurrence and Determination

A Chrevatidis, Eclipse Scientific Group, Chatteris, Cambs, UK

Occurrence

The name mycotoxin is a compound word derived from the Greek *mykes*, which means fungus (mold), and the Latin *toxicum*, which means poison. From early on, fungi were used in food preservation and processing, producing particular flavors and taste in products like cheese and soy sauce. They are also

producers of antibiotics and other useful chemicals. On the other hand, such species have been linked with a number of toxic outbreaks around the world, causing serious illnesses and even death. They have been discovered after the outbreak of 'St Anthony's fire,' when a large number of people died in Russia in the early 1900s after the consumption of rye contaminated with ergot alkaloids produced by the mold *Claviceps purpurea*. In their despair, due to a tremendous burning sensation and epileptic fits, many pilgrims travelled to St Anthony's shrine in France hoping for a cure.

During the last 40 years of research on mycotoxins, hundreds have been discovered but it is fortunate that a relatively small number has been proven to be harmful. Although the number is limited, it is important to bear in mind that a number of agricultural products are susceptible and the fungi infesting them are capable of producing amounts of toxins able to cause devastating effects.

Mycotoxins are toxic secondary metabolites (byproducts) occurring naturally by several fungal species under various climatic conditions on a large number of agricultural commodities. The reason for their existence is not quite clear; however, it has been suggested that they are produced under adverse conditions or that they may play a role in the survival of the fungus itself. Spores of these fungi exist in the air and soil throughout the world and can germinate when appropriate conditions arise. Mycotoxins can be formed in the field prior to and after harvest during inappropriate storage. More specifically, mycotoxin occurrence is due to several factors:

1. Biological: crop susceptibility and fungus compatibility;
2. Environmental: temperature, moisture, mechanical injury, insect damage, fungus;
3. Harvesting: crop maturity, temperature, moisture, detection/diversion;
4. Storage: temperature, moisture, detection/diversion;
5. Distribution/processing: detection/diversion.

Mycotoxins are a reason for concern because they affect the economy and health status of the world population. It has been estimated that approximately 25% of the world's grain is lost annually as a result of contamination by mycotoxins leading to produce being destroyed, dumped, or degraded, therefore, causing heavy financial burden on the exports of many developing and developed countries. Some Asian, African, and Latin American countries whose economy strictly relies on agricultural commodities can suffer dramatically when their consignments are denied export. These losses can vary from year to year depending on the environmental conditions, which entirely depend on nature. They are also a significant burden to the farming industry by reducing the growth rate and reproductive efficiency of animals as well as condemning meat and dairy products. The health implications associated with mycotoxins are numerous, and include acute toxicity outbreaks and chronic cases of disease. Acute effects such as death are caused by the ingestion of high amounts of the toxin in a given short time period; such cases have been recorded in Asia, India, and Africa, where quality control of food may be compromised. Chronic effects are a more serious matter, as is a type of exposure that can take place more frequently. It is caused by the long-term ingestion of low levels of the toxin; such exposure has been associated with increased risk of primary liver carcinoma in Asia. Further health effects caused by exposure to mycotoxins are allergies and infections.

Mycotoxins are very stable at high temperatures with little or no destruction occurring under the conditions of ordinary cooking or conventional processing. As a result they have repeatedly been detected in many final products. Consumption of contaminated foods is considered the main route of exposure to mycotoxins, and this must be avoided. Table 1 provides an overview of the most frequently occurring mycotoxins, their fungal source, and the foods susceptible to infestation.

Aflatoxins

A single event in the UK in 1960 established the significance of mycotoxins. During this outbreak thousands of turkeys, ducklings, and chicks died as a result of Turkey X disease. Chemical analysis by thin-layer chromatography (TLC) identified several ultraviolet (UV) fluorescent compounds in the feed (Brazil peanut meal) which were thought to be responsible for this outbreak. These compounds were named aflatoxins. Aflatoxins are produced primarily by the *Aspergillus* genera, and more specifically by *A. flavus, A. parasiticus*, and *A. nonius*, which are ubiquitous in the environment and infest a large variety of commodities. Within the aflatoxin family, aflatoxin B_1 (AFB_1) is the most acutely and chronically toxic, carcinogenic, and mutagenic of all. There is sufficient epidemiological evidence to support the carcinogenic properties of AFB_1 to humans and animals and therefore to lead to its classification as a class 1 carcinogen by the International Agency for Research on Cancer (IARC).

To date, a total of 20 compounds have been designated as aflatoxins with AFB_1, AFB_2, AFG_1, and AFG_2 the most significant and common food contaminants. Aflatoxin M_1 and M_2 are metabolites of AFB_1

Table 1 The most important mycotoxins produced by fungal species in various agricultural commodities

Mycotoxin	*Fungal species*	*Commodities*
Aflatoxins B_1, B_2, G_1, G_2	*Aspergillus flavus*	Nuts, cereals, pulses, spices, soybeans, dried fruit, cottonseed
	A. parasiticus	
	A. nomius	
Aflatoxins M_1 and M_2	Metabolic products of B_1 and B_2	Milk
Ochratoxin A	*A. ochraceus*	Grains, coffee, wine, beer, dried fruit
	Penicillium verrucosum	
Patulin	*P. expansum*	Apples, plums, peaches, pears, and vegetables
	A. clavatus	
	Byssochlamys nivea	
Trichothecenes	*Fusarium sporotrichoides*	Wheat, barley, corn, oats, rye
	F. graminearum	
	F. verticilloides	
	F. nivale	
	Stachybotrys	
	Trichoderma	
Zearalenone	*Fusarium culmorum*	Corn, wheat, barley, rice
	F. graminearum	
	F. crookwellense	
Fumonisins	*F. moniliforme*	Corn/maize, grains
	F. proliferatum	
Cyclopiazonic acid	*P. cyclopium*	Cheese, cereals, pulses, nuts
	A. flavus	
	A. versicolor	
Moniliformin	*F. moniforme*	Cereals
	F. sporotrichoides	
Sterigmatocystin	*A. nidulans*	Cereals, cheese
	A. versicolor	
Citrinin	*P. citrinum*	Wheat, barley, rice, rye, oats
	P. verrucosum	
	A. oryzae	
Penicillic acid	*P. cyclopium*	Corn, beans, apples
	P. puberculum	

and AFB_2 excreted in milk and urine of ruminants that have consumed contaminated feeding stuffs. Aflatoxins are produced when food is stored for long periods of time under high temperature and humidity. Seasonal variation has been found to affect mycotoxin levels, while regional variations due to the cultural habits of storing the food are usually considered as secondary factors. The occurrence of aflatoxins in foods and feeds has been reported in several countries and, although it is greatest in the warmest and humid parts of the world like Asia, Africa, and South America, it is also an annual problem to some degree in areas of the USA and Europe. Agricultural commodities rich in lipids and carbohydrates are susceptible to aflatoxin contamination and such raw ingredients include cereals, rice, pulses, oilseeds, dry fruits, spices, nuts (peanuts, Brazil, and pistachios) and animal-derived products (i.e., milk, eggs). Aflatoxins were also detected in Indian chilli powder, capsicum pepper, nutmeg, and dried chilli peppers. Recent evidence has revealed that aflatoxins have been detected in heroin and marijuana samples, compromising even further the health status of drug-dependent individuals. Since aflatoxins are very stable, food-processing techniques like pasteurization or sterilization may not be effective in removing the toxin from its natural substrate and as a consequence may be detected in the final product. Epidemiological evidence in Asian and African countries has revealed a positive correlation between exposure to aflatoxins and primary liver carcinoma in humans. In order to protect the consumer from such a potent food contaminant, the Food and Drug Administration (FDA) in the USA has set action levels for the presence of aflatoxins in foods ranging from 0.5 p.p.b. (for AFM_1 in milk) to 20 p.p.b. for total aflatoxins in other food. In Europe, the European Commission (EC) has set lower levels of 0.05 p.p.b. and 15 p.p.b. for milk and other foods respectively, while new legislation has set limits of 10 p.p.b. for total aflatoxins in spices.

Ochratoxin A

Ochratoxin A (OTA) is the most commonly encountered and important of the group. It is produced mainly by *Penicillium verrucosum* under temperate climatic conditions of Eastern and South Europe,

Canada, and South America. It is also of great concern to the Baltic states and Scandinavia as the local cereals grown (barley and oats) are particularly susceptible to high levels of ochratoxin contamination. *Aspergillus ochraceus* has been associated with the production of OTA in tropical areas of the world. The most frequently contaminated commodities are wheat, nuts, and corn, while a number of raw and processed commodities have been found to be contaminated by OTA; some of these are coffee, beer, wine, dried fruit, cocoa, and green coffee. Water activity of foods is important for the growth of fungi, therefore storage under dry conditions is critical. Similarly to aflatoxins, ochratoxin A is a moderately stable molecule and will survive most food processes, thus it is important that it is absent in the raw materials used. It is also important to highlight that ochratoxin is a regular contaminant of a large number of feeding stuffs, therefore contamination is transferred to animals such as swine and poultry. As a result humans are exposed directly by consumption of contaminated foods, as well as indirectly by consuming various animal tissues of an affected animal.

OTA is a potent toxin targeting the kidney and causing acute and chronic lesions, while associations have been made with Balkan endemic nephropathy on various parts of the world. As a result of its potency, it has been classified as a class 2B (potential) carcinogen by the IARC. To date, the EC has set legal limits for the presence of OTA in various food groups ranging from 3 to 10 p.p.b. for cereal products and dried vine fruit respectively.

Patulin

Patulin is a secondary metabolite of certain *Penicillium, Aspergillus*, and *Byssochlamys* species but, more specifically, *P. expansum* appears to be the fungus responsible for patulin production. It is a common contaminant of apples and apple products but commodities such as grains and other fruits and vegetables are also susceptible although to a lower extent. Even though the contamination incidence is high, it is fortunate that the levels are usually low. Refrigeration temperatures are shown to be ideal for the production of patulin by several fungi, thus long storage periods must be avoided. The presence of patulin in apple juice is evidence that rotten fruit were being used; however, it is fairly easy to avoid contamination as most of the toxin is located in the rotten part of the fruit and once it has been removed, the problem is minimized. Juice fermentation also results in a 99% destruction of the toxin. Patulin has been a reason for concern due to its carcinogenic properties; however, the IARC was unable to establish the severity of its effects on humans and as a consequence no classification exists at present. Regulation levels set in several countries vary from 30 to 50 $\mu g\,l^{-1}$. The EC proposed limits of 50 p.p.b. in apple juice and 25 p.p.b. in solid apple products.

Trichothecenes

Trichothecenes are a family of closely related chemical compounds. They are metabolites of the fungi of *Fusarium, Trichoderma*, and *Stachybotrys*. They have been reported in colder climates of Scandinavia, Canada, Japan, and a number of European countries, but they also occur worldwide and in high diverse geographic regions. The group was discovered in 1932 in the Soviet Union after an acute episode of alimentary toxic aleukia, which caused 60% mortality to the exposed population. Among the 50 known trichothecenes, a limited number appear to be of importance. According to their chemical characteristics, they have been divided into type A, including diacetoxyscirpenol, HT-2, neosolaniol and T-2 and type B trichothecenes including 3-acetyldeoxynivalenol, 15-acetyldeoxynivalenol, deoxynivalenol, fusarenone-X, and nivalenol. Commodities affected by *Fusarium* toxins are oats, rye, cassava, sorghum, bananas, groundnut, soybeans, mangoes, and many more. Some of the most commonly produced toxins are deoxynivalenol (DON), nivalenol (NIV), and diacetoxyscirpenol (DAS) produced by *F. graminearum, F. nivale*, and *F. sambucinum*, while HT-2 and T-2 are produced by *F. tricinctum* and *F. roseum* and are more rare and the most harmful of the group. DON has repeatedly been found in corn and wheat as well as other cereals and is considered the most ambudant trichothecene produced.

Trichothecenes are fairly stable compounds and can persist for long periods. Evidence has shown that 50–60% of the toxins are transmitted to the finished product as processing removes only a small amount, therefore, there are concerns in regard to dietary exposure through foods like breakfast cereals, breads, noodles, beer, and infant foods. Epidemiological data have associated DON ingestion with human mycotoxicoses in China, India, and Japan while trichothecenes have been associated with feed refusal, depression of immune response, nausea, occasional vomiting, and anemia in humans and animals. In addition to crops destined for human consumption, a large number of feeding stuffs have also been found to be contaminated by the same toxin; however, transfer to the human food chain appeared to be minimal. Infestation of crops by the mold appears red-pink colored, usually located at the head or tip of the crop. Formation of DON in the field can vary due to climate and geographical

region as well as variations from year to year. In view of consumer protection, the FDA has set advisory levels for DON of 1 p.p.m. in foods intended for human consumption. In Europe, the EC proposed action levels of 500 and 750 p.p.b. at the retail and raw material stage respectively.

Zearalenone

Zearalenone (ZEN) is a mycotoxin with estrogenic properties produced by the field fungi Fusaria. Similarly to the trichothecenes, ZEN is produced under the same conditions; however, a drop in the temperature during growth stimulates production of toxins. In addition to trichothecene production, *Fusarium graminearum* and *F. culmorum* are the primary fungi responsible for the production of ZEN; however, eight others can also produce the toxin at varying amounts. It is the second most commonly encountered *Fusarium* metabolite and is known to contaminate crops such as maize, wheat, barley, oats, and rye in Canada, the USA, and Japan. Although there is no epidemiological evidence to support esophageal cancer development from exposure to ZEN, an association in certain parts of Africa has been made. ZEN metabolites have been shown to transfer into milk; however, the evidence that they may pose potential risks to humans is limited. It is a compound that survives processing and fermentation and, as a result, it has been found not only in animal feed but products intended for human consumption. Examples are corn meal, breakfast cereals, beer, and fermented products.

The ZEN producing fungi usually grow in the field, causing pink discoloration of the kernel; however it is not always evidence of the presence of the metabolite. ZEN is only partly decomposed by heat treatment but cleaning a crop by removing its outer hull may result in 40–100% reduction of the toxin. The FDA has not issued an advisory level for ZEN but observations should continue. The EC has also urged member states to monitor ZEN levels in inflant foods and cereals.

Fumonisins

Similarly to aflatoxins and ochratoxins, fumonisins are a group of approximately 15 closely related mycotoxins. They were isolated in 1988 from cultures of *F. moniliforme*; however, several other *Fusarium* species, including *F. proliferatum* and *F. nygamoni*, are also producers. Among the six members of fumonisins, fumonisin B_1 (FB_1) is the most important as it is the most frequently occurring and potent of the group. They mostly occur in maize, with 90% of its amount produced in the field. Initially it was thought that their presence was limited to corn products, but they have also been found in lower concentrations in rice, sorghum, and navy beans. Fumonisins often occur together with other mycotoxins such as aflatoxins, ZEN, and deoxynivalenol and studies on their toxic interactions are currently under way. Crop infestation results in a disease called *Fusarium* ear rot and is the most common in the Midwest USA and around the world in both the northern and southern hemispheres. In terms of health implications, it has been associated with esophageal cancer in parts of South Africa, China, and Northeast Italy. Recently, the IARC has classified the toxins produced by *F. moniliforme* as class 2B (potential) carcinogens. No advisory level has been set in the USA nor Europe since there is insufficient evidence on its effects. However, the EC proposed a maximum daily intake of $2\,\mu g\,kg^{-1}$ body weight (based on rat studies).

Cyclopiazonic acid

Aspergillus flavus and various *Penicillium* species such as *P. cyclopium* are responsible for the production of cyclopiazonic acid, which has been detected in crops such as corn and peanuts and in several cheeses. Feeding stuffs are also commonly contaminated by cyclopiazonic acid and have been reported to coexist with aflatoxins in a range of substrates. It has been implicated in Kodua poisoning in India: this is a disease causing tremors, sleepiness, and giddiness. As a mycotoxin it is not well documented because the existing methods of analysis are not fully developed and therefore, this precludes a full assessment.

Moniliformin

Moniliformin is a mycotoxin which is produced by a number of *Fusarium* species, including *F. moniliforme, F. chlamydosporus, F. anthophilum*, and *F. proliferatum*. Only a limited number of surveys have shown that it is a frequent contaminant of maize and at low levels. It has also been reported to occur in wheat and rice and limited information exists about its stability during processing. In humans, a possible link between moniliformin ingestion and Keshan disease (a chronic heart condition) in certain areas of China has been suggested. Therefore, it has been recommended that food monitoring will continue to collect data on its frequency and levels of occurrence.

Sterigmatocystin

Sterigmatocystin was first isolated in 1957 from a mycelial mass of *A. versicolor* but *A. nidulans* is also known to produce it. AFB_1 has a very similar structure to sterigmatocystin and it is thought that it may be derived from it. With the same thought in mind, it is not accidental that sterigmatocystin has

been found in samples of pistachio nuts, pecans, wheat, and green coffee beans as well as reported in processed foods. Reports claim that its occurrence is not as frequent as aflatoxins, but this may be due to the lack of sensitive methods for its detection.

Citrinin

Citrinin was first recognized as a promising antibiotic but it was later found to cause kidney damage, retard growth, and eventually cause death in animals. Citrinin was isolated in the 1930s and produced by *Penicillium citrinum*; however, *P. verrucosum* is also known to produce the toxin. It is known as the yellow-rice toxin discovered in Japan in 1950–60, when hundreds of tonnes of rice imported from various parts of Asia were declared unfit for consumption. It was found in cereal products in parts of Europe, north America, and Canada and in maize samples from South-east Asia. A survey in Denmark detected both citrinin and OTA in barley samples and it is usually a cocontaminant with OTA, causing damage to the kidneys.

Penicillic acid

Penicillic acid was first isolated in 1913 from a moldy corn culture and has never been found in processed foods as it is an unstable compound. Penicillic acid is produced by *Penicillium* genera and is usually found in apples, corn, and beans at moderate incidence. Additional fungal species that occur during storage are able to produce substantial amounts.

Miscellaneous

There are many other mycotoxins that have been known to occur in foods and feeds; however, more work is required to establish their occurrence pattern and implications involved. Names of some mycotoxins are: alternaria toxins, aspergillic acid, kojic acid, ergotoxins, rubratoxins, the group of tremorgenic mycotoxins, and many more.

Determination

Once mycotoxins are formed, they are difficult to eliminate and being able to detect them is very important to ensure consumer protection. Their levels can vary from mg to pg levels, therefore, identification and quantitation at such low levels are challenges for the chemist. Since their discovery, a large number of methods have been developed by scientists and some have been used more extensively than others as their diversity varies depending on their physical and chemical properties. The methods used for the identification and quantitation of mycotoxins must be sensitive, specific, and as simple as possible. The presence of mycotoxins in their natural substrate is extremely unhomogeneous, therefore, it is critical that a representative sample is selected to avoid obtaining false results. Since sampling comprises the highest source of variation in the analytical procedures, a large number of random samples must be taken from a lot of commodity using the appropriate device. Although it is not always feasible and

Table 2 The most common methods used for the analysis of various mycotoxins

Mycotoxin	*Extraction solvents*	*Clean-up*	*Detection method*		*Approval status*
			Immunoassays	*Chemical*	
Aflatoxins	Me, AN, W, DiClMt, Ac	IAC	ELISA	HPLC	AOAC
Ochratoxin A	Me, AN, W, Cf, phosphoric acid	IAC	ELISA	HPLC	AOAC
Patulin	Ac, EtOAc,	Silica gel 60, fluorosil, celite	P	HPLC	AOAC
Fumonisins	Me, AN, W	IAC	HPLC (derivative)		AOAC
Trichothecenes	Me, AN, W, Cf, DiClMt, EtOAc	MycoSep	ELISA	GC/MS (derivative)	
Zearalenone	EtOAc, Me, AN, Cf, W	IAC	ELISA	HPLC	AOAC
Cyclopiazonic acid	Me, W, phosphoric acid, hydrochloric acid			HPLC	
Sterigmatocystin	Me, KCl, AN		P	HPLC	
Citrinin	Cf, phosphoric acid			HPLC	
Moniliformin	AN, W	C_{18} bond-elute		HPLC	
Penicillic acid	AN, W, KCl			HPLC	

AC, acetone, AN, acetonitrile, Cf, chloroform, DiClMt, dichloromethane, EtOAc, ethyl acetate, Me, methanol; IAC, immunoaffinity column; W, water; P, polyclonal antibody; ELISA, enzyme-linked immunosorbent assay; HPLC, high-performance liquid chromatography; GC/MS, gas chromatography–mass spectrometry; AOAC, Association of Official Analytical Chemists; KCl, potassium chloride.

economical to obtain a large number of samples, a balance between sampling variation and costs should be maintained.

Once appropriate sampling has taken place, the mycotoxins are isolated from their substrate by extraction. Several aqueous, organic solvents or mixtures of them with water and/or acid are used to assist the partition of the toxin in the solvent. Such solvents include methanol, acetonitrile, chloroform, dichloromethane, ethyl acetate and many others, shown in Table 2. The extract obtained is then cleaned up by various types of columns, which remove potential interfering compounds. This is a crucial step as the level of sensitivity depends on how clean the sample is. Immunoaffinity columns (IACs) are the most common clean-up and concentration systems used because they are simple and easy to use and the fact that the analysis is fully automated makes them amenable to a large number of samples. They are coupled with specific antibodies developed for the target mycotoxin and, when the extract is passed through the IAC, the toxin is bound and is then eluted with the appropriate solvents. The eluate is then ready for analysis either by a chemical or immuno-based method. Other clean up columns include MycoSep, fluorosil, C_{18}, and celite.

After clean-up, the sample will be subjected to analysis to identify and quantitate the mycotoxins present. The method used for this purpose depends on the chemical properties of the mycotoxin, as separation of the contaminants must take place prior to their quantitation. Thin-layer chromatography (TLC) was one of the first and most widely used chemical methods for mycotoxin analysis and relied on the fluorescent properties of the mycotoxin because the color produced and its intensity reflect the concentration of the toxin. Although TLC is a powerful tool in mycotoxin analysis, the data obtained are considered to be presumptive as they roughly estimate the quantity of the toxin and as a consequence further confirmation is required. High-performance liquid chromatography (HPLC), on the other hand, is considered a more sophisticated and sensitive technique which can provide information regarding the identity and quantity of a compound simultaneously. It is also possible to increase the fluorescence of a compound by derivatization, increasing the sensitivity of the method. Identification of the toxin is made by comparison of the retention time with a standard and its quantity determined by the absorbance value as well as the area of the elution peak. Therefore, HPLC has replaced TLC and is now the officially approved method of testing by the Association of Official and Analytical Chemists (AOAC). Even though HPLC is more extensively used, it suffers the disadvantages that it can only run one sample at a time and the samples to be tested have to be clean to achieve the required sensitivity. In addition to other methods, gas chromatography (GC) is commonly used for mycotoxins that do not have chromophores and fluorescent groups. It is considered one of the most effective methods for the analysis of trichothecenes and ZEN but the extract must be derivatized prior to analysis. Derivatization may differ depending on the detection method used, for example, acylation agents are used when flame ionization detection (FID) is used and trifluoroacetyl (TFA) esters are made when an electron capture detector (ECD) is used. Confirmation of the mycotoxin identified by GC is usually done by mass spectrometry (MS).

In order to overcome the problems encountered with the chemical methods, efforts to develop immunoassays have been made over the past 25 years. Antibody-based bioassays such as enzyme-linked immunosorbent assay (ELISA) and radioimmunoassay (RIA) are constantly under development and are being used for a number of mycotoxins. When compared with chemical methods, ELISA is more sensitive, requires smaller volumes of sample, and is more rapid, as a large number of samples can be processed on the same ELISA plate. A larger number of antibodies (monoclonal and polyclonal) exist and the efforts to develop new ones are ongoing. The only problem with ELISA assays is that the results obtained are usually semiquantitative and they can roughly estimate the quantity of the contaminant. Therefore, it must be used in conjunction with a quantitative method (such as HPLC) for confirmation. Table 2 gives a summary of the most commonly used reagents and methods for the analysis of mycotoxins.

See also: **Aflatoxins**; **Mycotoxins**: Classifications; Toxicology

Further Reading

Council for Agricultural Science and Technology (1989) *Mycotoxins: Economic and Health Risks*. Task Force Report 116.

Dhumal SS and Salunkhe DK (1992) Mycotoxins in foods. In: Tu AT (ed.) *Food Poisoning*, pp. 291–334. New York: Marcel Dekker.

Park DL, Ayala CE, Guzman-Perez SE, Lopez-Garcia R and Trujillo S (2001) Microbial toxins in foods: algal, fungal and bacterial. In: Helferich W and Winter CK (eds) *Food Toxicology*, pp. 94–136. Boca Raton: CRC Press.

Scudamore KA (1998) Mycotoxins. In: Watson DH (ed.) *Natural Toxicants in Food*, pp. 147–181. Sheffield: Academic Press.

Sharma RP and Salunkhe DK (1991) Introduction to mycotoxins. In: Sharma RP and Salunkhe DK (eds)

Mycotoxins and Phytoalexins, pp. 13–80. Boca Raton: CRC Press.
Smith JE, Lewis CW, Anderson JG and Solomons GL (1994) Mycotoxins in Human Nutrition and Health. European Commission, Directorate-General XII, Science Research and Development.
Trucksess MW and Wood GE (1994) Recent methods of analysis for aflatoxins in foods and feeds. In: Eaton DL and Groopman JD (eds) *The Toxicology of Aflatoxins: Human, Veterinary and Agricultural Significance*, pp. 409–432. San Diego: Academic Press.
USDA (1999) *Grain Fungal Diseases and Mycotoxin Reference*, pp. 1–55. Kansas City: GIPSA.

Toxicology

F S Chu, University of Wisconsin, Madison, WI, USA

Introduction

Developments in the last three decades have revealed many new mycotoxins that are potentially hazardous to human and animals. In considering their toxic effects, it is important to distinguish between 'mycotoxicosis' and 'mycosis.' Mycotoxicosis is used to describe the toxic action of mycotoxin(s) and is frequently mediated through a number of organs, notably the liver, kidney, and lungs, and the nervous, endocrine, and immune systems. Part of the name of the toxin or commodity involved is generally used to describe mycotoxicosis. For example, 'aflatoxicosis' generally refers to the toxicosis having typical clinical and pathological symptoms associated with the ingestion of aflatoxins. 'Mycosis,' however, refers to a generalized invasion of living tissue(s) by growing fungi.

Owing to their diverse chemical structures, mycotoxins may exhibit a number of biological effects, including both acute and chronic toxic effects as well as carcinogenic, mutagenic/genotoxic, teratogenic, and immunotoxic effects. Their toxic effects are often organ-specific. Once mycotoxins are ingested in the body, the toxins may either directly react with the target organs to exert toxic effects or be metabolized to an active intermediate (activated) and then exert their toxic effects. The toxins and their metabolites may also be detoxified and excreted. Thus, factors such as species, age, nutrition, and hormone status, etc. that affect the metabolism of mycotoxins, can also manifest mycotoxin toxicity. As will be shown below, the interaction of mycotoxins, either activated through metabolism or as intact molecules, with cellular macromolecules plays a dominant role in their toxic actions. Modulation of the immunosystem also plays an important role for the overall toxic effects, and mycotoxins can be either immunosuppressive (most often) or immunostimulatory. Recent studies on the effect of mycotoxin on apoptosis further revealed their mode of action at the cellular level. Owing to the large body of literature on mycotoxins that has accumulated over the years, this chapter will focus only on the toxic effects of selected mycotoxins (see **Table 1**) that are the most frequent contaminants in foods and feed. The mode of their toxic action will also be highlighted.

Historical Considerations

The mycotoxin problem is actually an old problem. Two classical examples, i.e., ergotism and mushroom poisons, have been known for centuries. Outbreaks of other types of mycotoxicoses have also been recorded. Because intoxications by ergots and poisonous mushrooms are a result of ingestion of a fungal body containing toxic metabolites, these two types of poisons are excluded from the most modern discussion on mycotoxins. Nevertheless, ergots and poisonous mushrooms still can be unintentionally introduced into food chains in the modern days. For a historical point of view, the toxicology of ergotism and one of the mycotoxicoses that occurred in World War II are briefly reviewed here.

Ergotism

Ergotism is a human disease that results from consumption of the ergot body in rye or other grains infected by a parasitic fungus of the genus *Claviceps*. Two types of ergotisms have been documented. In the 'convulsive' type, the affected persons have general convulsions and a tingling sensation in the muscles (such as the feeling of one's foot going to sleep), and sometimes, the entire body is racked by spasms. Epidemics occurred between 1581 and 1928 in European and other countries. Although it has been suggested that the consumption of ergots that cause convulsive ergotism may play a role in 'Salem Witchcraft' incidence, this is still a controversial issue. In the 'gangrenous' type, the affected parts became swollen and inflamed with violent, burning pains, hence, the term 'Fire of St. Anthony.' The affected area became numb first, turned black, then shrank, and finally became mummified and dry. Outbreaks occurred from the Middle Ages up to the nineteenth century. In some areas of France, grain contained as much as 25% ergots. Patients died as a result of consumption of

Table 1 Toxicological effects of selected commonly occurring mycotoxins[a]

Major mycotoxins	*Toxicological effect*[a]	*Mode of action*
Alternaria (AAL) mycotoxins	A, Hr, M	Inhibition of ceramide synthase
Aflatoxin B1 (AF) and other aflatoxins	A, C, H, I, M, T	Form DNA adducts and mutate p53 tumor suppressor gene
Citrinin (CT)	C(?) Nh, M	Inhibition of protein synthesis
Cyclopiazonic acid (CPA)	Cv, I, M, Nr,	Alteration of Ca^{2+} hemostasis and inhibition of Ca^{2+}-dependent ATPase
Deoxynivalenol (DON)	I, Nr	Inhibition of protein synthesis
Cyclochlorotine (CC)	A, C, H,	Binding with DNA
Fumonisins (FM)	A, C, Cv, H, Nr, R	Disruption of sphingolipid metabolism by inhibiting ceramide synthase
Luteoskyrin (LT)	A, C, H, M	Binding with DNA
Moniliformin (MN)	Cv, Nr	Binding with pyruvate dehydrogenase and affecting TCA cycle enzymes
Ochratoxin A (OTA)	A, I, Nh, T	Inhibition of protein synthesis and several enzymes; enhancement of lipid peroxidation; impaired cellular Ca^{2+} and cAMP homeostasis
Patulin (PT)	C(?), Nr, D, T	Inhibition of -SH enzymes
Penicillic acid (PA)	C(?), Nr, M	Inhibition of -SH enzymes
Rubratoxin B (RB)	A, H, T	Inhibition of protein synthesis
Sterigmatocystin (ST)	H, C, M	Same as aflatoxin
T-2	A, ATA, Cv, D, I, T	Inhibition of protein synthesis, binding with peptidyl transferase, membrane proteins and other enzymes
12–13, Epoxy-trichothecenes (TCTC) other than T-2 and DON	A, D, I, Nr	Same as T-2 toxin
Tenuazonic acid	Hr	Inhibition of protein synthesis
Tremorgenic toxins	Nr	Binding and inhibition of acetylcholinesterase
Zearalenone (ZE)	G, M	Binding with estrogen receptor

[a]Letters in parentheses are the abbreviations for the mycotoxin's name used throughout the text. A, apoptosis, ATA, alimentary toxic leukemia; C, carcinogenic; C(?), carcinogenic effect is still questionable; Cv, cardiovascular lesion; D, dermatoxin; G, genitotoxin and estrogenic effects; H, hepatotoxic; Hr, hemorrhagic; I, immunomodulation effects, M, mutagenic, Nh, nephrotoxin; Nr, neurotoxins; R, respiratory; T, teratogenic. From CAST (1989) *Mycotoxins, Economic and Health Risks*. Task Force Report No. 116. Ames, IA: Council of Agricultural Sciences and Technology and Chu (1998) Mycotoxins – Occurrence and toxic effects. In: Encyclopedia of Food and Nutrition – Food Contaminants, 2nd edn., pp. 858–869. New York: Academic Press.

100 g of ergot over a few days. In general, 2% ergots in the grain is sufficient to cause an epidemic. Between 1770 and 1771, about 8000 people died in one district alone in France. European and other countries have a regulatory limit of 0.1–0.2% ergots in flour. Ergotisms result from the intoxication of ergoline alkaloids that are produced by the fungus present in the sclerotia of *Claviceps purpurea*. The most active components are amides of D-lysergic acid, including both the cyclic-type peptide and nonpeptide amide of ergot alkaloids. These alkaloids cause smooth-muscle contraction, block neurohormones, have both vasoconstriction and vasodilation effects, and also affect the central nervous system. Thus, they have some therapeutic uses.

Alimentary Toxic Aleukia (ATA)

A well-documented example of mycotoxicosis in humans is the outbreak of ATA, a disease that resulted in more than 5000 deaths in humans in the Orenberg district of the former USSR during World War II. The cause of this outbreak was later determined to be trichothecene mycotoxins produced by fungi growing on grain allowed to stand in the field during winter. The toxicology of ATA will be discussed later.

Discovery of Aflatoxin

Although a number of antibiotics produced by fungi were found, in the middle of the 1940s, to be toxic to animals, renewed interest on toxic fungal metabolites only begun after the discovery of aflatoxin B1 in the early 1960s because of its highly potent carcinogenic effects to test animals, and the name of 'mycotoxin' was adopted thereafter.

Toxic Effects of Selected Mycotoxins

Aflatoxins

Since various aspects on aflatoxins have been dealt with elsewhere, the toxicology and mode of action on this groups of mycotoxins are reviewed only briefly herein.

Toxic effects Aflatoxin B1 (AFB1) is one of the most potent naturally occurring carcinogens and is the most toxic member of this group of mycotoxins. Other AFs are less toxic (B1 > G1 > B2 > G2). Although the main target organ is liver, AFB1 also affects other organs and tissues, including kidney and lungs, to a lesser degree. Typical symptoms for aflatoxicosis in animals include proliferation of the bile

duct, centrilobular necrosis and fatty infiltration of the liver, and hepatomas, in addition to generalized hepatic lesions. The susceptibility of animals to AFB1 varies considerably with species with a decrease in sensitivity (LD50: $mg\,kg^{-1}$) in the following order: rabbits (0.3), ducklings (0.34), mink (0.5–0.6), cats (0.55), pigs (6–7 kg size, 0.6), trout (0.8), dogs (1.0), guinea-pigs (1.4–2.0), sheep (2.0), monkeys (2.2), chickens (6.3), rats (5.5 for male, 17.9 female), mice (9.0), and hamsters (10.2). In addition to these acute (hepatotoxic) effects, carcinogenic effects of AFB1 is of most concern. A 10% incidence of hepatocarcinoma was observed in male Fisher rats fed a diet containing only 1 μg AFB1 per kilogram over a period of 2 years. Rats, rainbow trout, monkey, and ducks are most susceptible, but mice are relatively resistant to AFB1. Because of their presence in foods and strong evidence of their association with human carcinogenesis, aflatoxins are still a serious threat to human health, even after more than 40 years of research. Consumption of AFB1-contaminated feed by dairy cows results in the excretion of AFM1 in milk. AFM1, a hydroxylated metabolite of AFB1, is about 10 times less toxic than AFB1, but its presence in milk is of concern for human health.

Immunotoxic effects Animals fed with AFB1 are more susceptible to infection. While AFB1 primarily affects cell-mediated immunity (CMI) and phagocytic cells, it also reduces/suppresses delayed hypersensitivity (DH). For example, suppression of DH to keyhole limpet hemocyanin and a decrease in splenic CD4 cell (T-helper cells) and IL-2 were found in CD-1 and C57B/16 mice fed with AFB1.

Mode of action AFB1 must first be activated by mixed-function oxidases, more specifically cytochrome P450 1A2 and 3A4, to a putative short-lived AFB-8, 9-*exo*-epoxide before exerting its carcinogenic effects. The activated intermediate either can be converted to hydroxylated metabolites, conjugated to glutathione or glucuronic acid, etc. and then be excreted, or it can bind to DNA, RNA, and protein to exert its toxic, carcinogenic, and mutagenic effects. Adduct formation of this intermediate with DNA occurs thorough nucleophilic attack primarily at the N-7 guanine position, which then involves G–C to T–A or to A–T nucleotide substitutions and causes mutations.

Aflatoxin-induced G:C mutations, both G to T and G to A, have been implicated in the inactivation of the human *p53* tumor suppressor gene, and a high frequency of mutations has been found at the third position of codon 249 of *p53*. The prevalence of codon 249 mutation, i.e., AGG to AGT (Arg to Ser) was found to be around 50% in human primary hepatocellular carcinoma (PHC). In contrast, the prevalence of such mutations is low in nonPHC patients. The identification of mutations/inactivation of p53 at this site has been used as a biomarker for AFB-induced liver cancers in humans.

The role of the covalent interaction between AFB and specific enzymes/proteins is less clear. Several aflatoxin metabolites do interact with enzymes and proteins. Both aflatoxins B2a and G2a, which are very reactive with proteins through formation of Schiff's bases between the dialdehyde groups in AFB2a and the $-NH_2$ groups of proteins, have been shown to be very effective inhibitors of DNase *in vitro*. The activated AFB1 and AFG1 also react with albumin to form albumin adducts through the same Schiff-base mechanism, followed by Amadori rearrangement and subsequent condensation with another aldehyde group. The presence of this adduct in human serum has been used as an index of human exposure to AFs.

Aflatoxins B1 and G1 also bind with serum albumin, nucleic acid, and proteins such as lysine-rich histone and chromatin noncovalently. Aflatoxicol, a major AFB metabolite, interacts with estrogenic receptors. Although the role of such interaction was not clearly defined, NMR analysis of the binding of AFB1 and AFG1 with double-stranded $d(ATGCAT)_2$ or $d(GCATGC)_2$ and B-DNA and gel-shift analysis of the binding of both mycotoxins with plasmid pBR322 revealed that intercalation of aflatoxins between the base pair was involved. Intercalation of stable epoxide-AFB1 with these nucleotides has also been demonstrated. These data suggest that intercalation may be significant for subsequent covalent attachment of the carcinogen at the N7 position of guanine in DNA. The activity of several enzymes, such as deoxyribonuclease and RNA polymerase, is inhibited by AFB1, but these inhibitory effects are secondary. Recent data also showed that oxidative DNA damage is involved in the genotoxicity of AFB.

Apoptosis (APT) Induction of apoptosis by AFB has been observed in several cell systems. In the HL-60 human promyelocytic leukemia cells, the dose of AFB1 needed for induction of APT by AFB1 was considerably higher ($1–20\,\mu g\,ml^{-1}$) than those needed by trichothecene mycotoxins ($0.001–0.2\,\mu g\,ml^{-1}$). Such an effect was attributed to their inhibitory effect on RNA synthesis as a result of its binding with DNA and the inhibitory effect of DNA-dependent RNA polymerase.

Aflatoxin and human carcinogenesis Whereas AFB1 has been found to be a potent carcinogen in many

animal species, the role of AF in carcinogenesis in humans is complicated by hepatitis B virus (HBV) infections in humans. Epidemiological studies have shown a strong positive correlation between AF levels in the diet and PHC incidence in some parts of the world (China, Kenya, Mozambique, Philippines, Swaziland, Thailand and Transkei of South Africa). Aflatoxin–DNA and AF–albumin adducts, two of the major biomarkers for human exposure to AFB, as well as several AF metabolites, mainly AFM1, have been detected in serum, milk, and urine of humans in these regions. However, the prevalence of HBV infection is also correlated to liver cancer incidence in these regions and HBV in humans. Data on the enhancement of mutation of the *p53* gene suggest a synergistic effect of these two risk factors for PHC in humans. (*See* **Aflatoxins**; **Cancer**: Carcinogens in the Food Chain.)

Ochratoxins

Toxic effects Ochratoxin A (OA), the most toxic member (LD50 about 20–25 mg kg^{-1}) and also most commonly found toxin in this group, is a potent nephrotoxin causing kidney damage, including degeneration of the proximal tubule, in many animal species. Liver necrosis and enteritis were also observed. Other than acute toxic effects, OA also acts as an immunosuppressor and teratogen. Although OA has never been shown to be mutagenic, a weak genotoxic effect has been demonstrated in several systems. Ochratoxin A is only a weak nephrocarcinogen because a high level of toxin and an extended period of exposure are necessary to induce the tumors. A dose-related induction of renal tubular cell tumors was found in Fisher rats (F344/N) with a significant increase in renal tubular cell tumors at levels of 70 and 210 μg of OA per kg of body weight per day; but cancer incidence was not significantly different from the control at lower levels (21 μg of OA per kg of body weight per day). Recent studies showed that OA also causes liver cancer in rats.

Modulation of immune system Animals receiving OA show general immunotoxic signs, including lymphocytopenia and depletion of lymphoid cells, especially in the thymus, bursa of Fabricius and Peyer's patches. Increased counts of total leukocytes and neutrophils in the blood with reduced lymphocyte levels were observed in the weaner pigs receiving OA. A striking increase in the counts of eosinophils and of apoptotic phagocytes was also found. Depressed humoral immune responses, including lower serum IgM and IgG levels and suppressed antibody response to SRBC and other antigens in mice, have been observed in animals fed or injected with OA. DH was also suppressed. The immunosuppressive effects of OA were prevented by phenylalanine both *in vitro* and *in vivo*. OA also stimulates cytokine production with increase in IL-2 nd IL-5 in the EL4.IL-2 cell system.

Mode of action The mode of the toxic effect of OA is not as clearly defined as those of AFB. OA inhibits several enzymes, including phenylalanine hydroxylase, phenylalanine-tRNA synthetase and renal phosphoenolpyruvate carboxykinase, which may play a role in its toxic effects. The synthesis of the latter two enzymes was found to be inhibited by OA. In addition, it is also a strong inhibitor for carboxypeptidase A. Enhancement of lipid peroxidation is considered as one of the manifestations of cellular damage in OA toxicity. The induction of free radicals in a bacterial model system by OA was found to be regulated by Ca^{2+} ions. Superoxide dismutase and catalase have been shown to have some protective effect for OA-induced nephrotoxicity in rats. OA-induced lipid peroxidation in the Vero cell was also found to be decreased in the presence of these two enzymes, but the effects of aspartame and piroxcam were less pronounced. In the renal epithelial cells, OA interferes with hormonal Ca^{2+} signaling, leading to altered cell proliferation. Thus, OA may impair cellular Ca^{2+} cyclic AMP (cAMP) homeostasis.

Ochratoxin A binds strongly with serum albumin; thus, the toxin has a long half-life in animal (35 days with a single dose of OA) and human bodies. Such an interaction has not only played a significant role in the toxic effect of this mycotoxin because of its heterogenous body distribution, but also been used as a biomarker for human exposure of OA. Although OA has been found to bind to several proteins and enzymes noncovalently, administration of labeled OA to rats showed no covalent binding of OA to DNA of kidney and liver. Using the P-32 postlabeling technique, however, OA–DNA adducts were found in kidney, liver, and several other organs of mice and rats with OA. Such adducts may be due to the products derived from OA-mediated cytotoxicity. OA induces unscheduled DNA synthesis in cell cultures *in vitro*.

Apoptosis OA induces APT in lymphocytes and several other cell systems. The apoptotic effect was attributed to its inhibition of protein synthesis. Specific cell types with distinct members of the mitogen-activated protein kinase family were involved in the APT at OA concentrations at which no acute toxic effect can be observed. Induction of apoptosis via the c-jun amino-terminal-kinase pathway can explain some of the OA-induced changes in renal function

as well as part of its teratogenic action. Exposure of human proximal tubule-derived cells and renal epithelial cell lines to low OA concentrations can lead to direct or indirect caspase 3 activation and, subsequently, apoptosis. (*See* **Immunology of Food**.)

Ochratoxin A and human health Ochratoxin A has long been considered to be associated with the nephropathy of people residing in certain areas of Tunisia, Scandinavia, and the Balkan regions, where exposure may be 'endemic.' The pathological lesions of nephropathy in humans are similar to those observed in endemic porcine nephropathy, which is due to the consumption of feeds contaminated with OA. Since a high proportion of the patients suffering from endemic nephropathy in the Balkan region develop tumors of the renal pelvis and ureter, the possible involvement of nephropathy in tumor development was suggested. Ochratoxin A has been found in human serum in Tunisia and several European countries, including Bulgaria, Poland, Yugoslavia, and Germany. In certain areas of the Balkans and Tunisia, OA levels in the foods and in human serum in endemic regions are higher than in the nonendemic areas. OA also has been found in human milk and kidneys in some endemic regions.

Human exposure to OA could occur through consumption of OA-containing cereals or animal foods. The mean daily intake of OA by humans in European Union member countries, as calculated from data on OA levels in foods and in human bloods, was estimated to be 1.8 ng of OA kg^{-1} (range: 0.7–4.7) and 0.9 ng kg^{-1} (range: 0.2–2.4), respectively. The main dietary sources (55%) were from cereals and cereal products (0.2–1.6 μg kg^{-1}). Thus, the EU Committee for Food has recommended that the OA daily intake in humans should be below 5 ng kg^{-1}. Coffee (mean level 0.8 μg kg^{-1}), beer, pig meat, blood products, and pulses also contribute to the OA intake in humans. These findings reemphasize the possible involvement of OA in human carcinogenesis and the health hazard of exposure to OA in humans. Among 77 countries that have regulations for different mycotoxins, eight have specific regulations for OA, with limits ranging from 1 to 20 μg kg^{-1} in different foods.

Fumonisin (Fm)

Toxicologic effects Although FmB1 was originally found to be a potent cancer promoter, it is now recognized as a carcinogen. It is primarily a hepatotoxin and carcinogen. Feeding the toxin to rats has resulted in cirrhosis and hepatic nodules, adenofibrosis, hepatocellular carcinoma-ductular carcinoma, and cholangiocarcioma. In addition to liver, lesions were also observed in kidney, including renal carcinoma. Tubular nephrosis was found both in rats and in horses of field cases associated with equine leukoencephalomalacia (ELEM). The effective dose of FmB1 for cancer initiation in rat liver depended on both the level and duration of exposure. All of the three major Fms, i.e., FmB1, FmB2, and FmB3, show cancer initiation and promoting activities in rats. In cell-culture systems, FmB1 has been found to be a mitogen, cytotoxic to the cells, and can alter the expression of genes associated with the cell cycle. It induces *in vivo* genotoxicity, but has no mutagenic effect in the *Salmonella* system.

FmB1 was identified as an etiological agent responsible for ELEM in horses and other Equidae (donkeys and ponies), and for porcine pulmonary edema (PPE) in pigs. ELEM was originally found to be a seasonal disease occurring in late fall and early spring. The disease primarily causes neurotoxic effects in animals, including uncoordinated movements and apparent blindness. Liver damage has also been reported. Death sometimes occurs with no nervous symptoms. Clinically, the disease is characterized by edema in the cerebrum of the brain. The levels of FmB1 and FmB2 in feeds associated with confirmed cases of ELEM range from 1.3 to 27 p.p.m. In pigs, PPE occurs only at high FmB levels (175 p.p.m.), whereas liver damage occurs at much lower concentrations with a no observable adverse effect level (NOAEL) of < 12 p.p.m. Cardiovascular function is altered by FmB1. The FmB1-induced PPE is caused by left-sided heart failure, and not by altered endothelial permeability. In the cattle, renal injury, hepatic lesions, and alteration of sphingolipid in various organs were observed. The level causing disease in poultry is also high. Fms are also toxic to some plants: jimsonweed, black nightshade, duckweeds and tomatoes with the *asc/asc* genotype being most susceptible.

Mode of action The primary effect of FmB1 is due to its disruption of sphingolipid metabolism by inhibiting the enzyme sphinganine (sphingosine) *N*-acyltransferase (ceramide synthase). Because complex sphingolipids and ceramides are heavily involved in cellular regulation, including cell differentiation, mitogenesis, and apoptosis, such effects play an important role in the early phase of most of the diseases discussed above. The target organs may be very sensitive to the sphingolipid dysregulation. Because sphingolipids are natural inhibitors of Ca^{2+}-activated-protein kinase C (PKC), it has been suggested that FmB may affect PKC-regulated functions. A direct interaction of diacylglycerol binding site of PKC to phorbol esters, which are well-known tumor promotors, by FmB1 has been observed. The

alteration of sphingolipid metabolism results in a dramatic increase in the free sphingolipid bases, i.e., sphinganine (Sa) and sphingosine (So), in tissues and a decrease in complex sphingolipid levels. Significant increases in the So/Sa ratio were found in serum of pigs fed a diet containing as little as 5 p.p.m. of total FmB. Thus, Sa/So ratios in serum and urine are now being used in testing as an early biomarker of FmB exposure in animals and humans. FmB1 disrupts ω-6 fatty acid metabolic pathway and/or prostaglandin synthesis; such an effect is likely to be an important event in the mitoinhibitory effect of FmB1 on growth-factor responses. FmB1 was also found to inhibit the cytochrome P-450 enzyme selectively with most significant effect on CYP2C11. Such inhibition was considered to be due to a suppressed activity of protein kinase activity resulting from the inhibition of sphingolipid biosynthesis.

Apoptosis Whereas the role of FmB1 on immunosystem is not entirely clear, the effect of FmB1 on apoptosis in many cell lines has been extensively studied. Immune cells are one type of target cell. For example, an increase in nitric oxide and prostaglandin E-2 production was found when a murine macrophage cell line was grown in the presence of FmB1. The kidney, liver, and tomato cells, and other animal or plant cells treated with FmB or TA (one of the major AAL toxins), showed typical apoptotic characters with the degree of injury being related to the dose and time of exposure. The fundamental elements of apoptosis, as characterized in animals, are conserved in plants. Inhibition of the ceramide synthase by FmB and related mycotoxins are considered to play a key role in the apoptosis of the cells of the target organs/tissues. FmB1 diminished the cytotoxic effects of the lipids on esophageal tumor cells. The marked apoptosis induced by PGA2 and the PG2- and AA-induced p53 levels were lowered. It repressed specific isoforms of protein kinase C and cyclin-dependent kinase activity and induced expression of CDK inhibitors in monkey kidney cells (CV-1). Such effects lead to cell-cycle arrest and apoptosis. Eight genes induced by FmB1 have been identified. The ability of FmB1 to alter gene expression and signal transduction pathways are considered to be necessary for its carcinogenic and toxic effects. Since FmB1 is not genotoxic in bacterial mutagenesis screens and does not affect unscheduled DNA synthesis, the apoptotic necrosis, atrophy, and consequent regeneration may play an essential role in its carcinogenic effects. FmB1 may be the first example of an apparently nongenotoxic (nonDNA-reactive) agent producing tumors through such a mechanism.

Impact on human and animal health While Fms are commonly detected in corn-based foods and feeds, the impact of low levels of Fms in human foods is not known. Current data suggest that they may have more impact on the health of farm animals than on humans. Thus, the Mycotoxin Committee of the American Association of Veterinary Laboratory Diagnostics recommends that the FmB1 levels be limited to 5, 10, 50, and 50 p.p.m. for feeds to be used for horses, swine, beef cattle, and poultry, respectively. Several reports indicated that significantly higher levels of Fms, sometimes together with AFs, trichothecene mycotoxins, and carcinogenic nitrosamines, were present in corn samples collected from areas with high rates of human esophageal cancer. Thus, FmB1 in combination with several other etiological agents may play an important role in carcinogenesis in humans. Currently, only Switzerland regulates the Fm level (1.0 p.p.m.) in foods.

Selected Important Trichothecene (TCTC) Mycotoxins

General consideration of toxic effects The structural diversity of TCTC mycotoxins results in different toxic effects in animals and humans. Unlike AFB1, OA, and Fm, where the primary effects and clinical manifestations are well defined, TCTC mycotoxicoses are difficult to distinguish because they affect many organs, including the gastrointestinal tract, hematopoietic, nervous, immune, hepatobiliary, and cardiovascular systems. Ingestion of foods and feeds that contain TCTC mycotoxins cause many types of mycotoxicoses in humans and animals, including moldy corn toxicosis, scabby wheat toxicosis (red-mold, akakbi-byo disease, or scabby barley poisoning), feed refusal and emetic syndrome (swine), fusaritoxicoses, hemorrhagic syndrome, and alimentary toxic aleukia (ATA). More than 100 TCTC have been identified in the laboratory; only a few selected toxins that have been found in foods are described herein.

T-2 toxin and other type A TCTCs Although T-2 toxin occurs naturally in cereal grains, contamination with T-2 toxin is less frequent than with deoxynivalenol (DON). However, T-2 toxin (LD50 in mice: 2–4 mg kg^{-1}) is much more toxic to animals, perhaps also to humans, than DON (LD50 in mice: 50–70 mg kg^{-1}). Almost all the major TCTCs, including T-2 toxin, are cytotoxic and cause hemorrhage, edema, and necrosis of skin tissues. Inflammatory reactions near the nose and mouth of animals are similar to some lesions found in humans suffering from ATA disease. The severity of lesions is also related to chemical structure. Macrocyclic toxins such as verrucarin

A are most active, followed by group A toxins (T-2 toxin), and least with type B toxins, such as nivalenol. Neurologic dysfunctions, including emesis, tachycardia, diarrhea, refusal of feed/anorexia, and depression, were also observed. T-2 toxin and some TCTCs also induce major GI lesions, including perioral dermatitis, stomatitis, esophagitis, gastritis, radiomimetic lesions, and sometimes hemorrhage in the intestines. However, the major lesion of T-2 toxin is its devastating effect on the hematopoietic system in many mammals, including humans. Typically, there is a marked initial increase in the number of circulating white blood cells, especially lymphocytes, followed by a rapid decrease to 10–75% of normal values. Platelet counts are also reduced. There is also extensive cellular damage in the bone marrow, intestines, spleen, and lymph nodes, and in severe cases, complete atrophy of bone marrow and marked alteration of plasma coagulation factors. T-2 toxin and related TCTCs are the most potent immunosuppressants of the known mycotoxins and cause significant lesions in lymph nodes, spleens, thymus, and the bursa of Fabricus. The heart and pancreas are other target organs for T-2 toxin intoxication. Although urinary and hepatobiliary lesions have been observed for T-2 toxin and diacetoxyscirpenol (DAS), these effects are considered to be secondary.

Deoxynivalenol (DON) Deoxynivalenol, a major type B TCTC, causes feed refusal and emesis in swine; the name 'vomitoxin' is also used. Although DON is considerably less toxic than most other TCTC mycotoxins, the level of contamination of DON in corn and wheat is generally high (usually above 1 p.p.m., sometimes greater than 20 p.p.m.). Contamination of DON in the cereal grains is also world-wide. Toxicologically, DON induces anorexia and emesis in both humans and animals. Swine are most sensitive to feed contaminated with DON. Because of the frequent occurrence of high levels of DON in wheat and corn, its stability, and reported food poisoning outbreaks in humans, contamination of cereals with DON is a major concern of both the government and food and feed industries. Contamination of DON in wheat and corn may be associated with other toxic effects because other *Fusarium* toxins, including zearalenone and other TCTCs may be present also. Other type B TCTCs such as nivalenol and acetylated-DON are more toxic than DON to test animals.

Modulation of immune systems Because TCTC mycotoxins such as T-2 toxin exert a major toxic effect on the bone marrow, lymph nodes, thymus, and spleen in mammals, modulation of the immune system by this group of mycotoxins has been studied most extensively.

Effects on humoral immune response Trichothecenes have been shown to have a marked effect on humoral immunity. Administration of T-2 toxin and DAS to animals results in a reduction of their resistance to infection and a decrease in antibody formation. T-2 toxin may selectively affect subpopulations of T-suppressor cells or their precursors, and this suppression of antibody synthesis may be due to impairment of either antibody-forming or T-helper cell activities. Immunosuppressive effects have also been observed for other TCTCs, including the macrocyclic types. Mice fed a diet containing more than 10 p.p.m. of DON experienced atrophy of the thymus and other structural changes, decreased antibody formation, as well as suppressed B- and T-cell proliferation. However, DON is one TCTC that has both immune stimulation and suppression effects in experimental animals. Serum and saliva immunoglobulin (Ig) A was significantly increased in mice fed with high levels (25 p.p.m.) of DON, whereas serum IgM and IgG decreased. The increased IgA production is related to IgA-mediated nephropathy in mice; thus, it was postulated that DON might be one of the etiologic agents in IgA nephropathy, which is the most common glomerulonephritis in humans world-wide.

Cell-mediated immunity Lymphocyte proliferation is inhibited *in vitro* by T-2 toxin and DAS. Although the response of both T- and B-cells to mitogens was inhibited, the T-cell response was more sensitive. There is a stimulating effect due to a low concentration of T-2 toxin. Structure–activity studies showed that T-2, HT-2, and 3′-OH T-2 toxins were most effective in inhibiting proliferation of mitogen-stimulated human lymphocytes, whereas 3′-OH HT-2, T-2 triol, and T-2 tetraol toxins were 50–100 times less effective. The macrocyclic TCTCs, roridin, and verrucarin A, were 75–100 times more potent than the T-2 toxin. The differences in the immunotoxic effects of various TCTCs were attributed to the differences in both the uptake and metabolism of toxins by the cells. Interleukin-1 and 2 (IL-1, 2) production in spleen cell cultures was stimulated by trace amounts of TCTC. Hyperinduction of cytokines in T-helper cells by DON (most) as well as several other mycotoxins (cyclopiazonic acid, OA, and alpha-zearalenol) has been found, but patulin and T-2 toxin were inhibitory. Induction of expression of mRNA of ILs-2, 4, 5, 6 in a T-cell model EL4.IL-2 by DON was found at levels required for partial or maximal protein synthesis inhibition.

A single oral gavage with DON is sufficient to induce IL-1 beta, IL-2, IL-6, TNF α mRNA levels in Peyer's patches and spleen. The effect in IL-6-deficient mice was refractory to DON-induced dysregulation of IgA production and development of IgA nephropathy. TCTC-induced cytokine superinduction could lead to the terminal differentiation of immunoglobulin-secreting cells via T-cell-mediated polyclonal differentiation of B-cells or its precursors. The Peyer's patch might be particularly sensitive to such dysregulation.

Rather than suppression, an apparent increase in DH by T-2 toxin was found when the toxin was administered within a few days after the mice were sensitized with sheep red blood cells. The toxin treatment caused a marked decline in the cell population of the thymus. Thus, T-2 toxin may interfere with the generation of suppressor cells. Pretreatment of animals with either T-2 toxin or DAS was also found to prolong skin-graft survival time.

Mode of action

TCTCs are potent protein synthesis inhibitors Although a number of mycotoxins, including AFB1, OA, citrinin, PR-toxin, and tenuazonic acid (TzA), inhibit protein synthesis, most of these effects are secondary. For example, AFB1 inhibits protein synthesis at the transcriptional level. Inhibition of protein synthesis, however, is one of the early events in the manifestation of toxic effects of TCTC mycotoxins, which act at different steps in the translation process. The potency of inhibition varies considerably with the chemical structure of the side-chain. In general, T-2, HT-2, NIV, FS-x, DAS, verrucarin A, and roridin A affect at the initiation step, whereas verrucarol, trichotecin, and crotocin affect elongation. Binding of TCTC with ribosomes, specifically peptidyl transferase, is a key step leading to the inhibition of protein synthesis. Inhibition of the elongation step by DAS and fusarenon X has also been demonstrated. Some of the less well-known TCTCs such as trichodermol and trichodermone affect the termination step.

Binding with membrane proteins/enzymes TCTC may have a direct effect on cells through interaction with other cellular components such as cell membranes in addition to their effects on protein synthesis. Both T-2 toxin and fusarenon X interact with a number of '–SH enzymes,' thus inhibiting their activity.

Apoptosis Several TCTCs, including roridin A, T-2 toxin, nivalenol, and DON, induce APT in HL-60 human promyeloytic leukemia cells at significantly low levels with a minimum effective dose ranging from 0.001 to 0.2 $\mu g\,ml^{-1}$. Depending on the lymphocyte set, tissue source, and glucocorticoid induction, DON can either inhibit or enhance apoptosis in murine T, B, and IgA cells. Whereas suppression of protein synthesis was considered the mechanism for induction of APT by TCTC, a later study showed that the T-2-induced apoptosis was mediated by the elevation of Ca^{2+} ion levels. Epidermal cells developed apoptosis after T-2 toxin was applied to the dorsal skin of hypotrichotic WBN/ILA-Ht rats. Such an effect was found to be associated with the induction of c-fos and perhaps also c-jun mRNAs.

Impact of TCTC on human and animal health Because of their diverse toxic effects and also because of frequent contamination by toxins or toxigenic fungi in foods and feeds, TCTCs are also potentially hazardous to human and animal health. However, among the many types of TCTC mycotoxicoses mentioned earlier, only ATA and scabby wheat toxicosis have been demonstrated in human populations. The former, ATA, was attributed to the human consumption of overwintered cereal grains colonized by *Fusarium sporotrichioides* and *F. poae*; this caused the deaths of hundreds of people in the former USSR between 1942 and 1947. Later studies indicated that T-2 toxin and related TCTCs were the primary cause. The signs and symptoms of ATA disease, which include skin inflammation, vomiting, damage to hematopoietic tissues, leukocytosis, and leukopenia, are common in humans and animals, including cats, cattle, guinea-pigs, poultry, monkeys, and swine.

Deoxynivalenol has been found to be primarily responsible for outbreaks of scabby wheat toxicosis in humans that are quite common in several countries, but these toxicoses rarely cause death. For example, between 1961 and 1985, 35 outbreaks involving 7818 cases were found to be caused by consumption of foods made from either scabby wheat or moldy corn in China. The symptoms, which occurred within 15 min to 1 h after consumption of foods made from moldy corn meal, included: nausea (90%), emesis (61%), headache and drowsiness (78%), and 5–6% had abdominal pain, diarrhea, and a mild fever. People generally recovered 2–4 days after consumption of the foods. Analysis of the leftover moldy corn revealed that the samples had 0.34–93.8 p.p.m. of DON; no T-2 toxin or nivalenol (NIV) were found. Similar cases have been reported in people consuming scabby wheat flour.

The widespread natural occurrence of DON in wheat in Canada and the USA in the late 1970s and

early 1980s has alerted the general public to the potential hazard of this mycotoxin. The tolerable daily intakes of DON for adults and infants were estimated to be 3 and 1.5 $\mu g\,kg^{-1}$, respectively. A tolerance level of 1 p.p.m. for DON in grains for human consumption has been set by a number of countries, including the USA. In Canada, the guideline for DON in the uncleaned soft wheat used for nonstaple foods is 2 p.p.m., but 1 p.p.m. for the infant foods. Although inadequate storage may lead to the production of some TCTC mycotoxins, infestation of *Fusarium* spp. in wheat and corn in the field is of most concern for the DON problem.

TCTCs may also be involved in the so-called 'sick building syndrome' (SBS) in humans. *Stachybotrys atra* was isolated from a badly water-damaged Chicago suburban home, in which the occupants complained about headaches, sore throats, hair loss, flu symptoms, diarrhea, fatigue, dermatitis, and general malaise. Several TCTCs (verrucarins B and J, satratoxin H, trichoverrins A and B) were found in the *S. atra*-contaminated materials of this home. T-2 toxin, DAS, roridine A, and T-2 tetraol were isolated from the dust samples from the air ventilation system of another suspected sick building syndrome in three urban Montreal office buildings. *S. chartarum*, an indoor mold, has been associated with pulmonary hemorrhage cases in the Cleveland, Ohio, area. Although *Stachybotrys* spp. are consistently found in these buildings, other molds and mycotoxins might also be present. In a survey of buildings with moisture problems in Finland, scientists have identified sterigmatocystin (24%), TCTC (19%), and citrinin (4%) in 79 samples analyzed. The trichothecenes found included satratoxin G or H (five samples), DAS (five samples), and 3-acetyl-DON, DON, verrucarol, or T-2-tetraol in an additional five samples. *Aspergillus versicolor* was present in most sterigmatocystin-containing samples, and *Stachybotrys* spp. were present in the samples where satratoxins were found. Thus, it is important to identify both the mold and mycotoxins for the cause of SBS. A remediation program involving removal of all contaminated wallboard, paneling, and carpeting in the water-damaged areas of the home, as well as spraying a sodium hypochlorite to all surfaces during remediation, appeared to be effective. Air samples taken from post-remediation buildings showed no detectable levels of *S. chartarum* or related toxicity.

Other Selected Mycotoxins

Other than those mycotoxins discussed above, several naturally occurring mycotoxins with typical toxicoses may also be potentially hazardous to human and animal health.

Mycotoxins produced by *Aspergillus* other than AF and OA Sterigmatocystin (ST), structurally related to AFB1 and a precursor for biosynthesis of AFB1, is another naturally occurring hepatotoxic and carcinogenic mycotoxin. Although the carcinogenicity of ST is less than that of AFB1 in test animals (10–100 times less), this mycotoxin has been found to be mutagenic and genotoxic, and has been found in foods in certain high esophageal cancer incidence regions of China and Mozambique. Toxigenic fungi have been isolated from patients with esophageal cancer, and these strains are capable of producing ST in many commodities. Although the role of ST in human carcinogenesis appears to be indirect and inclusive, it acts mechanistically similar to AFB in carcinogenesis through the formation of DNA adducts and mutation of p53 suppressor gene. The carcinogenic effects of ST seem to be mediated by failure of p53-mediated G1 checkpoint because exposure of cells to ST resulted in failure of G1 arrest.

Aspergillus terreus and several other fungi (*A. flavus* and *A. fumigatus*) produce the tremorgenic toxins territrem A, B, and C, aflatrem, and fumitremorgin. Some of these, bound to acetylcholinesterase, cause irreversible inhibition of the enzyme. Gliotoxin, an epipolythiopiperazine-3,6-dione antibiotic containing sulfur produced by piperazines antibiotic produced by *A. terrus*, *A. fumigatus*, and *Trichoderma viride*, exerts immunosuppressive effects in animals. The toxic effect is due to its covalent binding with -SH enzymes. It induces apoptosis in spleen cells with a dose-related simultaneous increase in cAMP levels and markedly inhibits both perforin-dependent and Fas ligand-dependent cytotoxic T-lymphocyte (CTL)-mediated cytotoxicity. In addition, *A. flavus*, *A. wentii*, *A. oryzae*, and *Penicillium atraovenetum* are capable of producing nitropropionic acid (NPA), a mycotoxin causing apnea, convulsions, congestion in lungs and subcutaneous vessels, and liver damage in test animals. This toxin was also identified as an etiological agent for the 'deteriorated sugarcane poisoning (SP),' a fatal food poisoning that occurred in China. However, the fungi involved in the contamination in the sugarcane and NPA production were *Arthrinium sacchari*, *A. saccharicola* and *A. phaeospermum*.

Other mycotoxins produced by *Penicillium* Other than OA, *Penicillium* spp. produce many mycotoxins with diversified toxic effects. Cyclocholortine, luteoskyrin (LS), and rugulosin (RS) have long been considered to be possibly involved in the yellow rice disease during World War II. They are hepatotoxins and also produce hepatomas in test animals. However, incidences of food contamination with these

toxins have not been well documented. LS inhibits RNA polymerase through binding with the enzyme and also forms a three-component complex, i.e., DNA–Mg^{2+}–luteoskyrin, with DNA. LS and several related anthraquinone-type mycotoxins are capable of inducing free radicals leading to the formation of 8-hydroxydeoxyguanosine. Several other mycotoxins, including patulin (PT), penicillic acid (PA), citrinin (CT), cyclopiazonic acid (CPA), citreoviridin, and xanthomegnin, which are produced primarily by several species of *Penicillium*, have attracted some attention because of their frequent occurrence in foods. PT and PA are hepatotoxins and teratogens, but PT is not an immunotoxin. They inhibit dehydrogenases reversibly through interaction of the toxin with the -SH group in the enzyme active center. The α,β-unsaturated double bonds conjugated in lactone ring of the toxin were considered to be sites for the interaction with the –SH and $–NH_2$ groups of proteins through addition or substitution reactions. In the model systems, PT forms the same type of adducts with glutathione and *N*-acetyl-L-cysteine (NAC). However, free cysteine formed markedly different adducts, including mixed thiol/amino-type adducts involving the α $–NH_2$ group. PT–cysteine adduct(s) still have a teratogenic effect. Due to its highly reactive double bonds that readily react with –SH groups in foods, patulin is not very stable in foods containing these groups. Nevertheless, PT is considered a health hazard to humans. Many countries have regulatory limits, most commonly at a level of 50 μg kg^{-1}, for PT in various foods and juices.

Frequently associated with the natural occurrence of OA is CT, which is also a nephrotoxin. In addition to *Penicillium* and *Aspergillus* spp., a recent discovery of the ability of *Monocuus ruber* and *M. purpureus* (two molds frequently used in the preparation of certain oriental foods) to produce CT reiterated the potential health hazard of CT to human health. One of the mycotoxins closely associated with the natural occurrence of AF in peanuts is CPA, which causes hyperesthesia and convulsions as well as liver, spleen, pancreas, kidney, salivary gland, and myocardial damage. CPA alters Ca^{2+} hemostasis and induces charge alterations in plasma membranes and mitochondria. The reversible inhibition of reticulum Ca^{2+}-dependent ATPase is the primary cause of this effect. *Penicillium rubrum* and *P. purpurogenum* produce two highly toxic hepatotoxins (LD50, 3.0 mg kg^{-1} in mice, I.P.) called rubratoxins A (minor) and B (major), which are complex nonadrides fused with anhydrides and lactone rings. Rubratoxin B has been shown to have synergistic effects with AFB1. It induced apoptosis in several cell systems. A lack of extracellular Ca^{2+} ion accelerates RB-induced DNA fragmentation of HL-60 cells. Neither protein synthesis nor nucleic acid interaction was related to the RB-induced APT, but proteases are likely to be involved. p53 has also been found not to be involved in the RB-induced APT signal transduction. Secalonic acid D, a cleft palate-inducing mycotoxin produced by *P. oxalicum* and other fungi, reduces palatal cAMP levels. It inhibits the binding of the cAMP response elements to the binding protein, alters phosphorylation, and leads to altered expression of genes involved in cell proliferation, an event critical for normal palate development.

In addition to the above hepatotoxins and nephrotoxins, *Penicillium* spp. produce many mycotoxins with strong pharmacological effects on neurosystems. For example, the indoloditerpenes, penitrem A–F are a series of temorgenic mycotoxins. Penitrem A, the major toxin in this group, causes tremorgenic effects in mice at a level of 250 μg kg^{-1}. Roquefortines A–C (C is most toxic), produced by *P. roqueforti* and several other *Penicillium* spp., have neurotoxic effects in animals. Tremorgens in the paspalitrem group (paspalicine, paspalinine, paspalitrem A and B, paspaline and paxilline) are produced by *Claviceps paspali* and some *Penicillium* spp. Paxilline, a reversible, noncompetitive inhibitor of the cerebellar inositol 1,4,5-triphosphate (InsP-3) receptor, inhibits the InP-3-induced Ca^{2+} release.

Other mycotoxins produced by *Fusarium* Other than TCTCs and Fm, some *Fusarium* spp. also produce other mycotoxins. Zearalenone (ZE), also called F-2, is a phytoestrogen that has hyperestrogenic effects and causes reproductive problems in animals, especially swine. It interacts with estrogenic receptors and stimulates protein synthesis by mimicking the action of the hormone. Contamination with this mycotoxin, sometimes together with DON, in feed may result in a large economic loss in the swine industry. It has been postulated that ZE may contribute to the overall estrogen load of women because of its frequent natural occurrence. ZE stimulates cytokine production. It also stimulates the growth of estrogen receptor-positive human breast carcinoma cell line MCF-7 and functions as an antiapoptotic agent by increasing the survival of MCF-7 cell cultures undergoing apoptosis caused by serum withdrawal. The mitogen-activated protein kinase signaling cascade is required for ZE's effects on cell-cycle progression in MCF-7 cells. *Fusarium moniliforme* also produces several other mycotoxins, including fusarins A–F, moniliformin, fusarioic C, fusaric acid, and beauvericin, in addition to Fms. Although the impact of these mycotoxins on human health is still not known, fusarin C (FC) has been identified

as a potent mutagen. Moniliformin, which causes cardiomyopathy in test animals, may be involved in Keshan disease in humans in regions where dietary selenium deficiency is also a problem. It binds strongly with pyruvate dehydrogenase, thus affecting the enzymes in the TCA cycle. *Fusarium moniliforme* was also found to be most effective among a group of several fungi tested in reducing nitrates to form potent carcinogenic nitrosamines. Thus, the contamination of foods with this fungus could be one of the etiological factors involved in human carcinogenesis in certain regions of the world. Fusarochromanones, a group of mycotoxins produced by some *Fusarium equiseti* isolates, cause tibial dyschondroplasia in broiler chicks, turkeys, and ducks and were considered as the possible cause of Kashin-Beck disease in China.

Mycotoxins produced by other species *Alternaria alternata* and other *Alternaria* spp. species are capable of producing dibenzo-α-pyrone types of mycotoxins (alternariol, alternariol monomethyl ether (AME), altenuene, isoaltenuene and altenuisol), tetramic acid metabolites, tenuazonic acid (TzA) and related compounds, and perylene derivatives altertoxins (ATX) I, II (also called stemphyltoxin II), III and stemphyltoxin. Although most of those compounds are relatively nontoxic, AME has been shown to be positive in Ames tests at relatively high concentrations. TzA, a protein synthesis inhibitor that interacts with peptidyltransferase in ribosomes, is more toxic than others. It can chelate metal ions and is capable of formation of nitrosamines. It is also produced by *Phoma sorghina* and *Pyricularia oryzae*, and may be related to 'Onyalai,' a hematological disorder in humans in the south of Sahara in Africa. Although no extensive survey has been conducted to determine the occurrence of these mycotoxins in human foods, limited studies indicate that the incidence of some of these mycotoxins, such as AME and TzA, could be high in apple and tomato products. A recent discovery of the structural and functional similarity between fumonisins and AAL toxins, a series of phytotoxins produced by *Alternaria*, further shows the importance of mycotoxins produced by fungi in the *Alternaria* family.

Sporidesmine, a group of hepatotoxins produced by *Pithomyces chartarum* and *Sporidesmium chartarum*, causes facial eczema and liver damage in animals. Contamination of these mycotoxins in feeds can result in great economic loss to the sheep industry. Slaframine, an interesting mycotoxin produced by *Rhizoctonia leguminicola* (in infested legume forage crops), causes excessive salivation in ruminants as result of the blockage of acetylcholine receptor sites.

Role of Metabolism and Dietary Modification

Role of Metabolism

Because metabolism plays a key role in the activation and detoxification of mycotoxins, the biological effects of mycotoxins are greatly affected by the metabolism of the toxins in different animals. AFB1 is metabolically activated before the formation of the AFB–DNA adducts, which exerts carcinogenic effects. However, activated AFB1 can also be conjugated to proteins and interact with glutathione and then be excreted from the body. Thus, glutathione *S*-transferase serves as a key enzyme in the detoxification process for AFB1. Mice are resistant to the carcinogenic effects because they constitutively express an α-class glutathione *S*-transferase (mGSTA3-3). Rats do not constitutively express a GST with a high AFB1-epoxide conjugating activity and thus are sensitive to AFB1-induced hepatocarcinogenesis. Constitutively expressed human hepatic α-class GSTs have little or no AFB1-epoxide-detoxifying activity. In contrast to rodents, the v-class GSTs are responsible for most of the AFB–epoxide-conjugating activity in the liver. Other types of cytochrome P450 are also capable of metabolizing AFB1 to various hydroxylated metabolites, including AFM1 and AFQ1. Demethylase is capable of converting AFB1 to AFP1. The cytosolic steroid reductase can convert AFB1 reversibly to aflatoxicol, which serves as a reservoir for AFB1.

The importance of metabolism for other mycotoxins is generally related to the detoxification processes. For example, metabolic deacetylation and deepoxidation of trichothecence mycotoxins lead to the formation of hydroxylated derivatives and deepoxide-TCTCs that are less toxic. Hydrolysis of OA by proteolytic enzymes leads to the production of nontoxic metabolites. Hydroxylated OA, formed by the action of cytochrome p450, is also less toxic; no evidence of OA metabolism to form glutathione conjugate or other adducts was found. Hydroxylation of T-2 at the C-3′ and OA at the C-4 position by microsomal enzymes has also been demonstrated. Most of the hydroxylated mycotoxins and metabolites may be excreted directly or form conjugates of glucuronide or sulfate, which are then excreted. Thus, various factors affecting the kinetics of formation of adducts and detoxified metabolites greatly affect the toxicity of mycotoxins. A number of factors, including the sex

and species of the animal (genetics), environmental factors, nutritional status, and mycotoxin synergism, etc., can either directly or indirectly modulate the toxic effects.

Dietary Modifications

As discussed elsewhere in this encyclopedia, dietary modification greatly affects carcinogenesis in experimental animals and in humans. Without exception, the carcinogenic and other toxic effects of mycotoxins, especially aflatoxins, are also affected by nutritional factors, dietary additives, and anticarcinogenic substances. Since the manifestation of carcinogenic/toxic effects of AFB1 depends on its ability to be absorbed and metabolized to an active form and on its subsequent interaction with DNA to form AFB1–DNA adducts, dietary modifications on any or all of these steps would lead to an increase or decrease in the carcinogenic effects of aflatoxin. Mycotoxins have a high affinity for the hydrated sodium calcium aluminasilicate (or NovaSil) and other related products. Diets containing NovaSil and related absorbers are effective in preventing absorption of AFB1 and several other mycotoxins in animals. Chemoprotective agents and antioxidants such as ascorbic acid, BHA, BHT, ethoxyquin, oltipraz, pentaacetyl geniposide, Kolaviron biflavonoids, and even green tea have also been found to inhibit carcinogenesis caused by AFB1 in test animals. Dietary administration of the naturally occurring chemopreventive agents, ellagic acid, coumarin or α-angelicalactone has been shown to cause an increase in the activity of glutamate–cysteine ligase, a key enzyme for the synthesis of glutathione. The mechanism of such protective effects was found to be the shifting of metabolism to a detoxification route by formation of a AFB1–glutathione conjugate rather than the formation of AFB1–DNA adducts. Ebselen exerts a potent protective effect against aflatoxin B-1-induced cytotoxicity; such a protective effect may be due to its strong capability in inhibiting intracellular reactive oxygen species formation and preventing oxidative damage.

Other absorbents such as zeolite, bentonite, and superactive charcoal are also effective in decreasing the toxicity of mycotoxins such as T-2 toxin. The toxicity of OA in test animals is minimized when antioxidants such as vitamins C and E are added to the diet. Aspartame, which has been found to be partially effective in decreasing the nephrotoxic and genotoxic effects of OA, may compete with OA for binding to serum albumin. L-Phenylalanine was found to protect the toxic effects of OA; administration of this amino acid prevented the OA's inhibitory effect on some of the enzymes discussed earlier. Vitamin E was found to efficiently prevent cytotoxicity induced by FB1. (*See* **Aspartame**.)

Concluding Remarks

Mycotoxins can cause both acute and chronic effects in prokaryotic and eukaryotic organisms, including humans. Most of their effects are organ-specific, but some mycotoxins may affect on many organs. Induction of cancer by some mycotoxins through initiation and promotion processing, in animals and possibly in humans, is one of the major concerns for their chronic effects. Modulation of immune systems by some mycotoxins is another concern. Interaction with marcomolecules through noncovalent or covalent binding, or both, is the basis of their mode of action. With some mycotoxins, such as aflatoxins, metabolic activation is necessary for their binding with specific macromolecules, but with others, metabolic activation is not necessary. Nevertheless, metabolism plays a key role in modulating their toxic effects because metabolism can produce either activated or nontoxic metabolites for subsequent conjugation and excretions. Thus, factors affecting the metabolism of mycotoxins greatly affect the toxicity of mycotoxins. Recent studies have shown that many mycotoxins modulate apoptosis, and investigation of the mechanisms involved has become an effective tool for understanding the mode of action of mycotoxins.

Almost all of the mycotoxins are very stable; only a limited number of detoxification methods are currently available, and some of these are not economically feasible. From a toxicological point of view, dietary modifications can reduce the risk to some degree. Because most of these control measures are not very effective, rigorous programs for preventing human and animal exposure to the contaminated foods and feed are important, and the availability of effective methods for monitoring the toxin levels in the foods is essential. Although a risk assessment has been carried out for some mycotoxins, more epidemiological data for human exposure are needed for establishing toxicological parameters and safe doses in humans. The mycotoxin problem may be with us for a long time, and many mycotoxicoses are not well characterized. It is unrealistic to eliminate the problem. Only through multiple approaches can we minimize the problem and enhance human and animal health.

See also: **Aflatoxins**; **Alkaloids**: Properties and Determination; Toxicology; **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; **Mushrooms and Truffles**: Classification and Morphology; **Mycotoxins**: Classifications; Occurrence and Determination; **Nitrosamines**; **Spoilage**: Fungi in Food – An Overview

Further Reading

Beasley VR (1989) *Trichothecene Mycotoxicosis, Pathophysiologic Effects*, vols I and II. Boca Raton, FL: CRC Press.

Bhatnagar D, Lillehoj EB and Arora DK (eds) (1991) *Handbook of Applied Mycology*, vol. V, *Mycotoxins in Ecological Systems*. New York: Marcel Dekker.

Bondy GS and Pestka JJ (2000) Immunomodulation by fungal toxins. *Journal of Toxicology and Environmental Health B Critical Reviews* 3: 109–143.

Bucci TJ and Howard PC (1996) Effect of fumonisin mycotoxins in animals. *Journal of Toxicology (Toxin Reviews)* 15: 293–302.

Busby WF, Jr. and Wogan GN (1979) Food-borne mycotoxins and alimentary mycotoxicoses. In: Riemann H and Byran FL (eds) *Food-borne Infections and Intoxications*. New York: Academic Press.

CAST (1989) *Mycotoxins, Economic and Health Risks*. Task Force Report No. 116. Ames, IA: Council of Agricultural Science and Technology.

Chu FS (1997) Trichothecene mycotoxicosis. In: Dulbecco R (ed.) *Encyclopedia of Human Biology* (2nd edn. vol. 8), pp. 511–522. New York: Academic Press.

Chu FS (1998) Mycotoxins – occurrence and toxic effects. In: *Encyclopedia of Food and Nutrition – Food Contaminants* (2nd edn, pp. 858–869. New York: Academic Press.

Cole RJ and Cox RH (1981) *Handbook of Toxic Fungal Metabolites*. New York: Academic Press.

Dragan YP, Bidlack WR, Cohen SM *et al*. (2001) Implications of apoptosis for toxicity, carcinogenicity, and risk assessment: Fumonisin B-1 as an example. *Toxicologic Science* 61: 6–17.

Eaton DL and Gallagher EP (1994) Mechanisms of aflatoxin carcinogenesis. *Annual Review of Pharmacology and Toxicology* 34: 135–172.

Galvano F, Piva A, Ritieni A and Galvano G (2001) Dietary strategies to counteract the effects of mycotoxins: A review. *Journal of Food Protection* 64: 120–131.

Harris CC (1996) The 1995 Walter Hubert lecture – Molecular epidemiology of human cancer: Insights from the mutational analysis of the p53 tumour-suppressor gene. *British Journal of Cancer* 73: 261–269.

Jackson L, DeVries JW and Bullerman LB (eds) (1996) *Fumonisins in Food*. New York: Plenum.

Miller JD and Treholm HL (1994) *Mycotoxins in Grain: Compounds Other Than Aflatoxin*. St Paul, MN: Eagan Press.

Pitt JI (2000) Toxigenic fungi: which are important? *Medical Mycology* 38: 17–22.

Reiley RT, Voss KA, Yoo HS, Gelderblom WCA and Merrill AH, Jr. (1994) Mechanism of fumonisin toxicity and carcinogenesis. *Journal of Food Protection* 57: 528–535.

Sharma RP and Salunkhe DK (1991) *Mycotoxin and Phytoalexins in Human and Animal Health*. Boca Raton, FL: CRC Press.

Shier WT (2000) The fumonisin paradoxa review of research on oral bioavailability of fumonisin B-1, a mycotoxin produced by *Farium moniliforme*. *Journal of Toxicology – Toxin Reviews* 19: 161–187.

Smela ME, Currier SS, Bailey EA and Essigmann JM (2001) The chemistry and biology of aflatoxin B-1 from mutational spectrometry to carcinogenesis. *Carcinogenesis* 22: 535–545.

Ueno Y (1986) Toxicology of microbial toxins. *Pure and Applied Chemistry* 58: 339–350.

Van Egmond HP (1989) Current situation on regulations for mycotoxins: Overview of tolerances and status of standard methods of sampling and analysis *Food Additives and Contaminants* 6: 139–188.

Wild CP and Hall AJ (2000) Primary prevention of hepatocellular carcinoma in developing countries. *Mutation Research (Reviews in Mutation Research)* 462: 381–393.

Wogan GN (1992) Aflatoxins as risk factors for hepatocellular carcinoma in humans. *Cancer Research* (supplement) 52: 2114s–2118s.

N

NATAMYCIN

L V Thomas and J Delves-Broughton, Danisco Innovation, Beaminster, Dorset, UK

Introduction

Natamycin (also known as pimaricin) is a polyene macrolide antimycotic produced by *Streptomyces natalensis*. It is marketed commercially as the Natamax™ family of products by Danisco, and as the Delvocid® family of products by DSM. Natamycin is used worldwide as a food preservative; its primary application is surface treatment of cheese and processed meat (e.g., dry sausages) by dipping, spraying, or in emulsion coatings. It has several advantages as a preservative, including broad activity spectrum, efficacy at low concentrations, lack of resistance, and activity over a wide pH range. Due to its low solubility, natamycin does not migrate from the surface into the food, and thus does not affect the organoleptic properties. It has no effect on bacteria, including those used as starter cultures or to promote ripening. It is chemically stable and has prolonged effectiveness. Moreover, it is easy to apply and has a proven safety record. Its general characteristics are summarized in Table 1.

History

Natamycin was first discovered in 1955, when it was isolated from culture filtrates of an *Actinomycetes* bacterium isolated from soil near the town of Pietermaritzburg, in the Natal province of South Africa. The antimycotic was first called pimaricin, a name derived from this place of origin. Its current name natamycin, as well as the name of its producer organism *Streptomyces natalensis*, was similarly derived from Natal. Natamycin has also been produced from closely related *Streptomyces* strains isolated from other locations and has been given various names (e.g., tennecetin, antifungal A-5283) and trade names. Its potential as a food preservative was recognized shortly after its discovery, when its efficacy against molds and yeasts was investigated in foods including fruit juice, carbonated drinks, fresh strawberries and raspberries, dressed poultry, sausages, cottage cheese, and hard cheese.

Structure

Natamycin has the empirical formula $C_{33}H_{47}NO_{13}$ and a molecular weight of 665.7. The structure was first determined in 1958, and later revised. It belongs to the group of polyene macrolide antifungals, compounds characterized by possession of a macrocyclic ring of carbon atoms closed by lactonization, and four conjugated double bonds. This classes it as a tetraene antimycotic (Figure 1). The structure is closely related to other antimycotics such as nystatin, rimodicin, and amphotericin. The molecule is amphoteric due to one basic and one acid group and has an isoelectric point of 6.5.

Stability and Solubility

Natamycin is a white/cream-colored crystalline powder with no taste and little odor. It is usually found as the stable trihydrate formulation. No significant loss of activity occurs for several months if the powder is stored in the dark at room temperature. Aqueous suspensions are less stable, particularly if exposed to light, certain oxidants, and heavy metals, but remain sufficiently stable during practical use. Although solutions are more unstable in acid or alkaline conditions, the pH of food products is not normally at levels that would cause problems. For instance, natamycin stored at 30 °C for 21 days reportedly retained 100% activity at pH 5–7, retained *c.* 85% at pH 3.6, and 75% at pH 9.0. In acid solutions, the molecule undergoes hydrolysis at the glycosidic linkage, forming mycosamine and aponatamycin and subsequently other compounds. Aqueous suspensions at neutral pH are stable for 24 h at 50 °C and show little reduction in activity at this temperature for several days and for shorter periods at 100 °C.

Natamycin is poorly soluble in water (30–100 p.p.m. at room temperature) and almost insoluble

in nonpolar solvents, but shows good solubility in strongly polar organic solvents (Table 1). To be active, natamycin must be in solution. However, solubility is not normally a limiting factor as natamycin is usually effective at relatively low concentrations. In solution, natamycin has an ultraviolet absorption spectrum with minima at 250, 295.5, and 311 nm, and maxima at 220, 290, 303, and 318 nm.

Table 1 General characteristics of natamycin

Characteristic	*Description*
Names	Natamycin, pimaricin (pimafucin), tennecetin, antifungal A-5283 (trade names: Natacyn, Myprozine)
Commercial products	Natamax™, Delvocid®
Producer organism	*Streptomyces natalensis*
EU number	E235
Formula	$C_{33}H_{47}NO_{13}$
Molecular weight	665.7 Da
Structure	Polyene macrolide antimycotic compound with a macrocyclic ring of carbon atoms closed by lactonization, and four conjugated double bonds
Properties	Amphoteric. Isoelectric point: 6.5
Solubility in different solvents	Water: 30–100 p.p.m. *n*-Butanol: 50–120 p.p.m. Glycerol: 15 000 p.p.m. Methylpyrrolidone: 120 000 p.p.m. Glacial acetic acid: 185 000 p.p.m.
Absorption spectrum	Maxima: 290, 303, 318 nm Minima: 250, 295.5, 311 nm
Antimicrobial spectrum	Most molds and yeasts (MIC: < 5–20 p.p.m.)
JECFA ADI	0.3 mg kg^{-1} body weight

MIC, minimum inhibitory concentration; JECFA, Joint Food and Agriculture Organization/World Health Organization Expert Committee on Food Additives; ADI, acceptable daily intake.

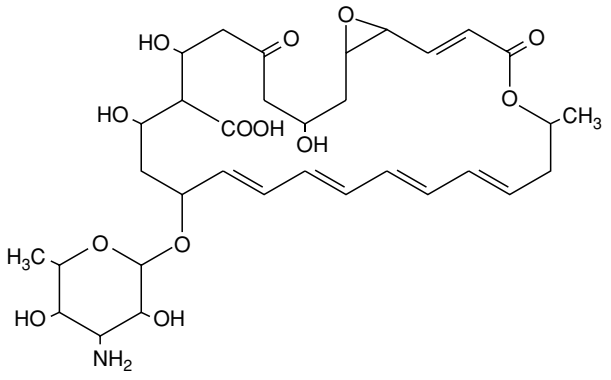

Figure 1 The structure of natamycin.

Mode of Action and Antimicrobial Effect

Mechanism of Action

Natamycin acts against yeasts and molds but is ineffective against bacteria. It combines with ergosterol and other sterols such as 24, 28-dehydroergosterol and cholesterol, compounds present in the cell membranes of yeast and molds but (with rare exceptions) not present in bacteria. The irreversible binding of natamycin to sterols, in particularl ergosterol, disrupts the cell membrane integrity and increases membrane permeability. This results in leakage of essential cellular constituents including cations, causing a rapid drop in intracellular pH and ultimately cell lysis. A study of laboratory-forced natamycin-resistant mutants of *Aspergillus flavus* has confirmed this theory. These either had reduced levels of ergosterol or none at all (and consequently much slower growth rates – critical to survival *in vivo*). A secondary mode of action is the inhibition of glycolysis and respiration.

Antimicrobial Spectrum

Natamycin is active against an extensive range of yeasts and molds (Table 2). The preservative is usually effective at concentrations between < 1 and 10 p.p.m. In general, yeasts are more sensitive than molds: the minimum inhibitory concentration (MIC) of yeasts is usually < 5 p.p.m., whereas that of molds can be at least 10 p.p.m. Examples of MICs are shown

Table 2 Antimicrobial spectrum: examples of yeast and molds sensitive to natamycin

Absidia	*Penicillium camemberti*
Alternaria	*P. chrysogenum*
Aspergillus chevalieri	*P. digitatum*
A. clavatus	*P. expansum*
A. flavus	*P. glabrum*
A. nidulans	*P. islandicum*
A. niger	*P. notatum*
A. ochraceus	*P. roqueforti* var. *punctatum*
A. oryzae	*Rhizopus oryzae*
A. penicilloides	*Rhodotorula gracilis*
A. roquefortii	*Saccharomyces bailii*
A. versicolor	*S. bayanus*
Botrytis cinerea	*S. cerevisiae*
Brettanomyces bruxellensis	*S. exiguus*
Candida albicans	*S. ludwigii*
C. guilliermondii	*S. rouxii*
C. vini	*S. sake*
Cladosporium cladosporioides	*Sclerotina fructicola*
Fusarium	*Scopulariopsis asperula*
Gloeosporium album	*Torulopsis candida*
Hansenula polymorpha	*T. lactis* var. *condensi*
Kloeckera apliculata	*Wallenis sebii*
Mucor mucedo	*Zygosaccharomyces barkeri*
M. racemosus	

Table 3 Examples of sensitivity of yeasts and molds to natamycin

Strain	*Minimal inhibitory concentration (p.p.m.)*
Aspergillus chevalieri 4298	0.1–2.5
Saccharomyces cerevisiae H	0.15
Penicillium chrysogenum	0.6–1.0
Aspergillus niger	1.0–1.8
Saccharomyces bailii	1.0
Candida albicans	1–2.5
Mucor mucedo	1.2–5.0
Penicillium notatum 4640	5.0
Saccharomyces rouxii 0562	5.0
Rhizopus oryzae 4758	10.0

in Table 3. Less susceptible species include *Verticillium cinnabarinum, Botrytis cinerea*, and *Penicillium discolour*, and also occur among the genera *Aspergillus, Fusarium, Penicillium*, and *Trichophyton*. The outcome of antifungal activity is usually cidal, in contrast to sorbate, which is fungistatic. Natamycin can inhibit aflatoxin synthesis, furthermore overall control of fungi leads to control of aflatoxins.

Resistance to polyene antimycotics such as natamycin does not seem to occur naturally and it has also proved difficult to generate resistant mutants in the laboratory. For example, in a survey of factories of natamycin-producers, no detectable difference in natamycin sensitivity was found for isolates compared to other factories. Similarly, in both a cheese warehouse and sausage factory where natamycin had been used for some time, no resistant strains were isolated. This lack of resistance may be partly explained by the fact that natamycin occurs as micelles in solution. Thus if a cell comes into contact with natamycin in solution, the antimycotic concentration is high and consequently lethal. Furthermore, the site of activity is an essential component of the cell membrane.

Factors Affecting Efficacy

The antimycotic activity of natamycin is affected by several factors, which may also affect its stability, such as pH, temperature, light, oxidants, and heavy metals. Natamycin is active over a wide pH range (pH 3–9); pH does not appear to affect its antimycotic activity but does affect stability. Compounds such as peroxides or chlorides, often used as cleaning/disinfecting agents, should be used with care in the proximity of natamycin.

Methods of Assay

The natamycin content of food products can be determined by microbiological, immunochemical, spectrophotometric, and liquid chromatographic (LC) procedures. The agar diffusion bioassay using *Saccharomyces cerevisiae* can also be used for quantitative assessment. Spectrophotometric and LC methods are commonly used for routine analysis. High-performance liquid chromatography (HPLC) with ultraviolet detection is considered one of the most sensitive and accurate methods, with a detection limit of 0.5 mg kg^{-1}. (*See* **Chromatography**: High-performance Liquid Chromatography.)

International Dairy Federation standard 140A: 1992 specifies the method for determining the natamycin contents of cheese rind and cheese adjacent to the rind. In this method a known quantity of sample is extracted with methanol, which is then diluted with water and cooled to between −15 and −20 °C to congeal and precipitate the majority of fat, followed by filtration. The natamycin content can then be determined in the filtrate by either a spectrometric or HPLC method. For the spectrometric measurement the spectrum of both a natamycin standard solution and the test sample is recorded in the range 300–340 nm. Absorbance is measured at the maximum of approximately 318 nm, at the minimum of approximately 311 nm, and then at 329 nm. For measurement by HPLC, ultraviolet detection should be at 308 nm and it is recommended that the mobile phase of the column should comprise methanol/water/acetic acid, 60 + 40 + 5 (v/v/v). Before analysis of test samples, a standard natamycin solution is first injected to determine retention time and for calibration.

More recently, a rapid method using derivative spectrophotometry has been described. In this method cheese is extracted with acidified aqueous acetonitrile, and natamycin content is directly quantified in filtered extracts on the basis of the depth of the trough at 322.6 nm after third-derivative processing of the normal ultraviolet spectrum. An enzyme immunoassay involving the use of a natamycin–protein conjugate has also been reported, in which natamycin coupled to horseradish peroxidase was used as the labeled ligand. Detection limits were 200–2000 pg ml^{-1}, which enabled the determination of concentrations as low as 0.005 mg dm^{-2} (0.1 mg kg^{-1}) in cheese rinds.

Toxicology and Legislation

Toxicology studies have been undertaken using mice, rats, rabbits, and guinea pigs. Natamycin was least toxic if administered orally (LD50 = 1500 mg kg^{-1} in rats and mice) or subcutaneously (LD50 = 5000 mg kg^{-1} in rats), and most toxic if administered intravenously (LD50 = 5–10 mg kg^{-1}). No natamycin was

absorbed from the human intestinal tract after 7 days' feeding of up to a maximum of 500 mg day^{-1}. Feeding studies have been conducted in rats, rabbits, and dogs. The acceptable daily intake was set at 0.3 mg kg^{-1} body weight per day in 1976.

Specification of natamycin in the USA (21 CFR 172.155) requires purity of the anhydrous compound to be 97% $\pm$ 2%, containing < 1 p.p.m. arsenic and not more than 20 p.p.m. heavy metals.

Natamycin is approved for use as an antimycotic in various cheeses and processed meats in 32 countries worldwide. A more general use as a food additive is allowed in a few countries such as South Africa (Table 4). In the European Union (EU) natamycin (designated E235) is permitted for surface treatment of hard, semihard, and semisoft cheese as well as dry sausages. The maximum surface coverage permitted in the EU is 1 mg dm^{-2}, and penetration is restricted to 5 mm from the surface. In the USA, natamycin is permitted only if the cheese standard allows the use of 'safe and suitable' antimycotics. Table 4 is a guide to food legislation. This list is not comprehensive – one should be aware that legislation for food additives is under constant review. The reader is advised to check the current legal situation with the appropriate authorities.

Table 4 Food legislation on the use of natamycin

Country	*Food in which natamycin is permitted*	*Maximum permitted level*
Algeria	Cheese rinds	Used in suspension at 2.5 g l^{-1}
Argentina	Surface treatment of hard and semihard paste cheeses	Limit of 1 mg dm^{-2}. Penetration limit of 2 mm
Australia	Surface treatment of cheese rind	Limit of 15 mg/kg
	Uncooked fermented manufactured meat products	Penetration limit of 3–5 mm
Bahrain	Permitted food preservative	
Brazil	Surface treatment of cheese	Limit of 2 mg dm^{-2}
Bulgaria	Cheese rind	500 mg kg^{-1}
Canada	Surface treatment of 47 listed cheeses	20 mg kg^{-1} based on total weight
	Grated/shredded cheese (0.5% sodium lauryl sulfate prohibited as dispersant)	10 mg kg^{-1}
Chile	Surface treatment of hard cheese (prohibited in wine)	
China	Surface treatment of cheese, processed meat products, moon cakes, baked goods, fruit juices, and processing utensils for easily moldy foods	Application by spraying or dipping in 200–300 mg kg^{-1}, to leave a residue of < 10 mg kg^{-1}
Colombia	Cheese	12.5 mg kg^{-1}
Czech Republic	Dairy and meat products – as EU regulations (contact authorities for further information)	Limit of 1 mg dm^{-2}. Penetration limit of 5 mm
Cyprus	Surface treatment of specified cheeses	
Egypt	Surface treatment of cooked cheese (dehydrated, semidehydrated, and semisoft cheese)	Limit of 2 mg 100 cm^{-2} (1 mg dm^{-2})
European Union (EU)	Surface treatment of specified cheese and sausage	Limit of 1 mg dm^{-2}. Penetration limit of 5 mm
Hungary	Surface treatment of hard and semihard cheese, dried, cured sausage	Limit of 1 mg dm^{-2}
Iceland	Surface treatment of ripened and whey cheese	Limit of 2 mg dm^{-2}
India	Surface treatment of hard cheese	Maximum application level: 2 mg dm^{-2}. Maximum residual level in finished cheese: 1 mg dm^{-2}. Penetration limit of 2 mm
Israel	Surface treatment of specified cheese	
Kuwait	Permitted food additive	
Lithuania	Surface treatment of hard, semihard, and semisoft cheese	Limit of 1 mg dm^{-2}. Penetration limit of 5 mm
Mauritius	Surface treatment of hard, semihard, and semisoft cheese and dried cured sausage	Limit of 1 mg dm^{-2}. Penetration limit of 5 mm

Continued

Table 4 Continued

Country	*Food in which natamycin is permitted*	*Maximum permitted level*
Mercosur	Surface treatment of cheese	Limit of 1 mg dm^{-2}. Maximum application of 5 mg kg^{-1}. Penetration limit of 2 mm
Mexico	Cheese surfaces	Limit of 0.002 %
Norway		Maximum level of suspension: 2 g kg^{-1}
	Surface treatment of hard, firm, and semifirm cheese, dried, cured sausage	Limit of 1 mg dm^{-2}. Penetration limit of 5 mm
Oman	Surface treatment of specified cheese	
Philippines	Surface treatment of specified cheese	
Poland	Permitted in colored and uncolored soft wax and polyvinyl acetate for application to skin of hard cheese	No limits
	Surface treatment of smoked, dried sausage	Limit of 1 mg dm^{-2}
Saudi Arabia	Permitted as a mold inhibitor in foodstuffs but controlled by standards of composition	
Slovakia	Surface treatment of cheeses and dried, cured sausage	Limit of 1 mg dm^{-2}
South Africa	Wine, alcoholic fruit beverages, and grape-based liquors	30 mg l^{-1}
	Fresh fruit pulp	5 mg kg^{-1}
	Fruit juice (blackcurrant, pineapple, etc.)	5 mg kg^{-1}
	Fish sausages	6 mg kg^{-1} to be applied to the outer inedible casing only
	Manufactured fish products, fish pastes, fish roe and spawn, with the exception of frozen fish, salted snoek, and canned fish products	6 mg kg^{-1}
	Lobsters (quick frozen)	6 mg kg^{-1}
	Edam, Gouda, Tilsiter, Limburger, Cheddar, Cheshire	2 mg kg^{-1} in rind without plastic coating; 500 mg kg^{-1} in a plastic coating; 10 mg kg^{-1} for application to the surface of the cheese only
	Cottage cheese, cream cheese	Limit of 10 mg kg^{-1}
	Process or blended cheese, including cheese spread, process cheese preparations, and soft cheese	10 mg kg^{-1} for application to the surface of the cheese only
	Yogurt	10 mg kg^{-1}
	Canned foods	6 mg kg^{-1}
	Manufactured meat products	Limit of 500 mg kg^{-1} on casing or 6 mg kg^{-1} in contents
	Canned chopped meat, canned corned beef, cooked cured luncheon meat, cooked cured pork shoulder, Biltong, frozen cooked-meat pie fillings	Limit of 6 mg kg^{-1}
Tunisia	Surface treatment of hard, semihard and semisoft cheese; dried, cured sausage	Limit of 1 mg dm^{-2}
Turkey	Surface treatment of hard and semihard cheese, dried cured sausages, salami, and hot dogs	Limit of 1 mg dm^{-2}. Penetration limit of 5 mm
Ukraine	Surface treatment of cheese	Limit of 1 mg dm^{-2}. Penetration limit of 5 mm
United Arab Emirates	Permitted food additive	
USA	Surface treatment of cuts and slices of cheese	Limit of 20 mg kg^{-1}
	Nonstandard of identity yogurt	Limit of 7 mg kg^{-1}
	Nonstandard of identity cream cheese	Limit of 7 mg kg^{-1}
	Cottage cheese	Limit of 7 mg kg^{-1}
	Sour cream	Limit of 7 mg kg^{-1}
	Soft tortillas	Limit of 20 mg kg^{-1}
	Nonstandard of identity salad dressing	Limit of 20 mg kg^{-1}
Venezuela	Surface treatment of specified cheeses and sausages	Maximum 0.5% suspension

Preservation of Foods with Natamycin

Types of Food Suitable for Natamycin Use

The principal use of natamycin is on the surface of cheese and dry sausages. Due to its low solubility, natamycin remains effective on the surface for extended periods. When first applied, only 30–50 p.p.m. will be present in solution on the surface – the remainder is present in the more stable crystal formation. The preservative then gradually dissolves, insuring a slow release and prolonged effectiveness. An additional advantage of surface treatment is that natamycin does not penetrate far into the food; a limit of penetration of 1–4 mm has been reported in cheese rind. This is particularly useful for the preservation of blue cheese, especially in comparison to sorbate. Sorbate migrates into the cheese matrix where it inhibits the desired blue mold development inside the cheese. Natamycin remains on the surface, acting only where it is needed to prevent the growth of surface spoilage molds.

Much of the early work on natamycin investigated its ability to inhibit fungal growth on fruit. Strawberries, raspberries, and cranberries sprayed with 50 p.p.m. natamycin in the field prior to harvesting showed reduced spoiling during storage. Treatment by immersion in 10–100 p.p.m. proved more effective. A lecithin coating containing natamycin can be used to treat harder fruit, such as apples and pears. An addition level of 1–10 p.p.m. natamycin may be effective for a variety of beverages, including juices, beer, wine, and beverages containing tea or milk solids. Natamycin (at 20 p.p.m.) has proved effective against *S. cerevisiae* spoilage of orange juice. Further potential applications include ready-to-eat frostings, salad dressings, and mayonnaise susceptible to yeast spoilage. Natamycin control of *Aspergillus niger* and *S. cerevisiae* in cottage cheese has been investigated. Levels of 20–100 p.p.m. added to wash water or 1–5 p.p.m. added to the cheese dressing inhibited fungal growth. A level of 5–10 p.p.m. can be effective in controlling yeast and mold growth in yogurt. Natamycin has also been shown to be effective against molds isolated from bakery products. A level of 100 p.p.m. has been shown to inhibit the growth of yeasts and molds in fillings and icings. (*See* **Yeasts**.)

Natamycin can also be used to treat the surface of cured-meat products such as raw hams and investigations have been conducted into its ability to inhibit the spoilage microflora on raw cut chicken. Dipping in a solution of 10 p.p.m. inhibited yeast counts for 12–15 days at refrigeration temperature.

Mode of Application

To treat the surface of cheese and sausages, the food can be immersed, sprayed, or coated in an aqueous suspension or the suspension can be used in a plastic coating. Penetration into the food then depends on the food type, being greater in soft cheese compared to hard cheese.

Aqueous dipping solutions for cheese usually contain 1250 p.p.m. natamycin, but concentrations vary depending on cheese type. For example, dips for blue cheese may require dip concentrations as high as 2500 p.p.m. The cheese should be dipped into the natamycin slurry for a few seconds, whilst the slurry is kept constantly agitated to maintain the natamycin in suspension. The concentration of natamycin in the dipping solution becomes reduced after each dipping, at a rate dependent on the ratio of surface area to weight of each piece of cheese.

A suspension of 1250 p.p.m. is recommended for spraying cheese, applied at a rate of 6 l t^{-1} to achieve 7–15 p.p.m. natamycin on the surface. It is important that the natamycin is evently applied as it binds rapidly and tightly on contact with the cheese surface, so that subsequent mixing after initial application will not achieve better distribution of the preservative. A tumbler of at least 1 m in length is recommended with a space of at least 0.3 m between the natamycin and flow agent application.

Lower levels of natamycin (100–750 p.p.m.) can also be added to an emulsion of a polymer in water, mostly polyvinyl acetate, which can be applied as a cheese coating. Alternatively, 100–20 000 p.p.m. of natamycin can be mixed with 0.5–50 g l^{-1} of a suitable thickener and approximately 20–250 g l^{-1} salt and used on the surface of cheese or sausages. This method of application overcomes the problem of uneven distribution of the fungicide caused by the heterogeneity of the fat content of the product. Coatings can be applied by 'painting' multiple layers on to the surface with a sponge or soft cloth, or by a commercially available coating machine.

Natamycin can also be used in natural and fibrous casings of dry fermented sausage products, preventing mold growth during the ripening and storage process. Commonly, casings can be prepared by immersion for 2 h in a 1000-p.p.m. suspension. Alternatively the sausages can be dipped or sprayed in a 1000–5000-p.p.m. natamycin suspension.

Fermentation and Production

Natamycin is produced by fermentation of *Streptococous natalensis* in an aqueous nutrient medium containing a carbon source (e.g., starch, molasses,

glycerol, glucose, lactose, maltose, sucrose, alcohols, organic acids), a fermentable organic and/or inorganic nitrogen source (e.g., corn steep liquor, casein, zein, lactalbumin, soya bean meal). The carbon source usually comprises 0.5–5% of the medium. Inorganic cations (potassium, sodium, or calcium) and anions (sulfate, phosphate, or chloride) may be needed as well as trace elements such as boron, molybdenum, or copper. Fermentation is aerobic and mechanical agitation and use of antifoaming agents can aid the process. The temperature is usually 26–30 °C and the pH range pH 6–8. The period of fermentation, varying with the medium, can be between 48 and 120 h.

Due to its low solubility in water, natamycin will accumulate mainly as crystals in the broth. Recovery commonly involves dissolving the natamycin using polar solvents with limited water miscibility such as butanol, methanol, and acetone and adjusting the medium to approximately pH 10. The fermentation broth is filtered to remove the mycelial biomass and impurities, and adjusted to a lower pH (*c*. pH 7) in order to crystallize the natamycin, which is then recovered and finally dried. An alternative process that does not use organic solvents involves the disintegration of the fermentation biomass by homogenization, high shear mixing, or ultrasonic techniques or treatment with heat, alkali, or enzymes.

See also: **Legislation**: History; **Mycotoxins**: Classifications; **Preservatives**: Classifications and Properties; Food Uses; Analysis; **Spoilage**: Molds in Spoilage; Yeasts in Spoilage; **Yeasts**

Further Reading

Anonymous (1992) Cheese and cheese rind determination of natamycin content method by molecular absorption spectrometry and by high-performance liquid chromatography. *IDF Standard 140A*. Brussels: International Dairy Federation.

Davidson PM and Doan CH (1993) Natamycin. In: Davidson PM and Branen AL (eds) *Antimicrobials in Foods*, pp. 395–407. New York: Marcel Dekker.

Hamilton-Miller JMT (1973) Chemistry and biology of the polyene macrolide antibiotics. *Bacteriological Reviews* 37: 166–196.

Raab WP (1972) *Natamycin (Pimaricin). Its Properties and Possibilities in Medicine*. Stuttgart: Georg Thieme.

Raghoenath D and Webbers JJP (1997) *Natamycin recovery*. AU patent application no. 1999717229 B2.

Stark J (1999) Permitted preservatives – natamycin. In: Robinson RK, Batt CA and Patel PD (eds) *Encyclopedia of Food Microbiology*, pp. 1776–1781. London: Academic Press.

Struyk AP and Waisvisz JM (1975) *Pimaricin and Process of Producing Same*. US patent 3892850.

Struyk AP, Hoette I, Drost G *et al*. (1957) Pimaricin, a new antifungal antibiotic. *Antibiotics Annual* 878–885.

Nectarines *See* **Peaches and Nectarines**

Neural Tube Defects *See* **Pregnancy**: Maternal Diet, Vitamins, and Neural Tube Defects; **Folic Acid**: Properties and Determination; Physiology

NIACIN

Contents
Properties and Determination
Physiology

Properties and Determination

K J Scott, Formerly of Institute of Food Research, Norwich, UK
P M Finglas, Institute of Food Research, Norwich, UK

Introduction

Niacin (nicotinic acid and nicotinamide) occurs widely in nature. Chemically, nicotinic acid is 3-pyridine carboxylic acid, and nicotinamide is 3-pyridine carboxylic acid amide (Figure 1). It should also be noted that in the USA niacin is used as a specific name for nicotinic acid, and niacinamide for nicotinamide.

Nicotinic acid (mol wt 123.11) occurs as colorless (white) nonhygroscopic needles or crystalline powder. It has a melting point of 235.5–236.5 °C and sublimes. It is soluble in water (1.67 g per 100 ml at 25 °C). It is also soluble in ethanol, acids, alkalis, and propylene glycol, but insoluble in ether. It has an absorption maximum at 261 nm, but the intensity is pH-dependent.

Nicotinamide (mol wt 122.11) also occurs as colorless (white) needles or crystalline powder but is slightly hygroscopic. It has a melting point of 128–131 °C and a distilling point of 150–160 °C at 5×10^{-4} mmHg. It is very soluble in water (100 g per 100 ml at *c.* 25 °C), ethanol (67 g per 100 ml) and glycerol (10 g per 100 ml) but is almost insoluble in ether. Like nicotinic acid, it has an absorption maximum at 261 nm. Both nicotinic acid and the amide are stable in the dry form and in neutral aqueous solutions. Nicotinamide can be converted to nicotinic acid by treatment with acids or alkalis. Nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP) are coenzyme forms. They combine with a wide variety of proteins and catalyze a large number of oxidation and reduction reactions and substrates in living organisms. (*See* **Coenzymes**.)

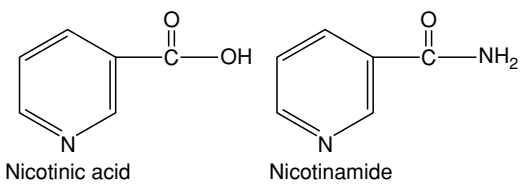

Figure 1 Structural formulae for nicotinic acid and nicotinamide. Reproduced from Niacin: Properties and Determination, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Salder MJ (eds), 1993, Academic Press.

The human requirement of nicotinic acid or nicotinamide is dependent on the tryptophan intake (60 mg of tryptophan is equivalent to 1 mg of nicotinic acid). Tryptophan is converted via a complex pathway. This pathway from tryptophan to NAD^+ does not directly produce nicotinic acid or nicotinamide; NAD^+ is transformed to nicotinamide, which is further *N*-methylated, and reenters the pyridine nucleotide cycle or may be deaminated to nicotinic acid, which is further converted to nicotinic acid mononucleotide.

Occurrence

Nicotinic acid and nicotinamide are found in a wide variety of foods: yeast, liver, heart, and muscle meats are good sources (Table 1). Because of the tryptophan contribution, milk, milk products, and eggs are also considered to be sources of niacin activity, although their actual content of niacin itself is relatively low. In natural products, niacin is usually in the bound form. Thus, in order to measure the total content it is necessary to include a hydrolytic stage. It is important to realize that this measure of total niacin must not necessarily be equated with what is biologically available. Most nicotinic acid in maize is in the form of nicotinyl esters, which are thought not to be hydrolyzed in the gut. In wheat bran it is bound to polysaccharides, peptides, and glycopeptides. Around 50% of niacin in unenriched wheat and wheat products has been reported to be in a bound form. The treatment of maize with lime water has been reported to liberate bound niacin, making it biologically available to pigs. It has also been claimed that, in areas where corn is routinely steeped in lime water, there is an absence of pellagra compared to areas with an equal consumption of unsteeped corn where pellagra was endemic. However, the various claims for bound forms of

Table 1 Nicotinic acid, potential nicotinic acid from tryptophan, and niacin equivalents in various foods (mg 100 g^{-1})[a]

Food	*Nicotinic acid*	*Tryptophan/60*	*Niacin equivalents*
Marmite	58	9	67
Liver			
Lamb (raw)	14.2	4.3	18.5
Lamb (fried)	19.9[b]	4.9[b]	24.8[b]
Ox (raw)	13.4	4.5	17.9
Ox (stewed)	10.3[b]	5.5[b]	15.8[b]
Heart			
Lamb (raw)	6.9	3.6	10.5
Ox (raw)	6.3	4.0	10.3
Ox (stewed)	4.7	6.7	11.4
Kidney			
Lamb (raw)	8.3	3.5	11.8
Lamb (fried)	9.6	5.3	14.9
Ox (raw)	6.0	3.4	9.4
Ox (stewed)	6.2[b]	5.5[c]	11.7[c]
Meats (lean, raw)			
Beef	5.0[b]	4.7[b]	9.7[b]
Lamb	5.4[b]	3.9[b]	9.3[b]
Pork	6.9[b]	4.5[b]	11.4[b]
Chicken	7.8	4.3[b]	12.1[b]
Milk (whole, average)	0.16[b]	0.6[b]	0.8[b]
Eggs			
Chicken (whole, raw)	0.1	3.7	3.8
Chicken (boiled)	0.1	3.7	3.8
Cheese			
Cheddar	0.1	6.8[b]	6.9[b]
Danish blue	0.63[b]	5.5[b]	6.1[b]

From: [a]McCance RA and Widdowson EM (1991) The Composition of Foods. In: Holland B, Welch AA and Unwin ID, Buss DH, Paul AA, Southgate DAT (eds) *The Composition of Foods*, 5th edn. Cambridge: Royal Society of Chemistry and Ministry of Agricultural, Fisheries and Food; [b]Food Standards Agency (2002) *McCance and Widdowson's The Composition of Foods*, Sixth Summary Edition. Cambridge; Royal Society of Chemistry. [c]Estimated.

niacin require thorough reinvestigation. (*See* **Bioavailability of Nutrients.**)

Biological Fluids (Blood) and Tissues

Nicotinamide is the primary circulating form of the vitamin. All tissues can incorporate nicotinamide into NAD. In the liver, nicotinic acid is incorporated into NAD, which is broken down into nicotinamide. The primary regulatory substance of the homeostasis in whole animals is nicotinamide itself. The levels in blood are buffered by the liver by being converted into a storage form of NAD.

Determination

Biological Assays

Historically, biological assays with animals such as the chick or the rat have been used as a baseline for other methods and to provide an evaluation of vitamin potency and availability of vitamins. However, biological assays are expensive to carry out and have an inherent lack of precision but, most importantly, are directly relevant only to the species used and conditions of the assay procedure. Thus, 'animal' assays for the determination of niacin are generally no larger used.

Other Assays

Unlike a biological system, which responds to the bioavailable vitamin in a food material, nonbiological assays are usually geared to measuring the total vitamin content. However, it is possible, using differential hydrolysis in conjunction with the microbiological assay, to make some assessment of the distinction between the 'free' and 'bound' vitamins, although this information must be treated with caution. Hydrolysis to release bound forms is normally carried out with sulfuric acid, although in some methods the use of alkali has been recommended, particularly for cereals.

Chemical assays This assay is based on the formation of colored complexes formed as a result of the reaction between niacin and cyanogen bromide. The reaction is as follows: an α-unsubstituted pyridine forms a pyridinium salt with cyanogen bromide, which reacts with an amine under ring opening to form a derivative of glutaconaldehyde. The color

produced can be measured spectrophotometrically. It has been reported that the colored products from the amide are not as stable or intense as from the acid. The amide is usually hydrolyzed to the acid. The Association of Official Analytical Chemists (AOAC) method employs sulfanilic acid as the chromogenic base in the manual and automated methods for nicotinic acid and nicotinamide in drugs, foods, feed, and cereal products, and barbituric acid in the analysis of the amide in multivitamin preparations. Foods and feeds are extracted with acid, and cereal products with calcium hydroxide. Factors which must be controlled for the method to produce reliable results include reaction temperature, pH, and preparation of blank corrections. However, this method lacks specificity, since all 3-pyridoxine derivatives react, and it requires the use of the highly toxic reagent cyanogen bromide. Thus, this method has been largely replaced by microbiological and high-performance liquid chromatography (HPLC) methods. (*See* **Spectroscopy**: Visible Spectroscopy and Colorimetry.)

Microbiological methods Although probably not suited for occasional use because of the specialist facilities and expertise required, the microbiological assay has been widely used for the determination of nicotinic acid in a variety of materials. It is more sensitive and specific than the chemical assay. The microbiological assay is based on the specific nutritional requirements in a defined medium for a particular vitamin(s) by a microorganism. The most widely used organism is *Lactobacillus plantarum* (ATCC 8014). In order to present the vitamin in a biologically active form to the organism the material to be assayed is normally subjected to acid hydrolysis. This procedure also converts any amide into nicotinic acid. This organism responds to nicotinic acid, nicotinamide, and nicotinuric acid (an inactive metabolite), and NAD, but not tryptophan, and it is also able to utilize bound nicotinic acid, present in cereals, to a considerable extent. It is specified in official methods for the determination of total niacin activity in food. The AOAC and AACC extraction procedures involve autoclaving the sample at 121–123 °C for 30 min with 1 mol l^{-1} sulfuric acid.

High-performance liquid chromatography HPLC offers an alternative to the chemical or microbiological assay, although the initial equipment cost and subsequent recurrent costs are relatively high. Analysis is commonly carried out after alkali, acid, or acid/enzyme hydrolysis, by separation on a reversed-phase column and ion pair reagents in the mobile phase and ultraviolet detection. A particular problem in the HPLC analysis of niacin in food materials is that because of its relatively low ultraviolet absorption, interference from other compounds can make peak identification and quantification difficult. The application of this technique to food products often requires clean-up procedures, like cartridge extractions and column switching. The use of fluorescence detection increases specificity and sensitivity, but requires postcolumn derivatization, because niacin is not natively fluorescent. (*See* **Chromatography**: High-performance Liquid Chromatography.)

Dietary Requirements

As stated earlier, the evaluation of data obtained from the analysis of foods is complicated by the uncertainties surrounding the bioavailability of bound forms of the vitamin and the contribution of tryptophan. These factors must be considered when considering the nutritional implications of niacin intake. (*See* **Dietary Requirements of Adults**.)

The presently accepted conversion rate of tryptophan to niacin is 60:1. Using this conversion rate and the tryptophan requirement to maintain nitrogen balance, it is possible to establish the niacin requirement.

Recommended daily amounts or reference nutrient intake values are based on niacin equivalents. The UK figures are calculated in terms of resting metabolism, that is, 2.7 mg of niacin equivalents per MJ or 11.3 mg per 1000 kcal. The daily level for men is about 17 mg, and about 13 mg for women, and is independent of activity. The recommended level does not fall with increasing age. An additional 2 mg is recommended during lactation. Recommended levels for infants and children are dependent on age, ranging from 3 to 5 mg for infants under 1 year, about 11 mg at 5 years and about 12 mg at 10 years of age. US recommendations are very similar.

See also: **Bioavailability of Nutrients**; **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; **Coenzymes**; **Dietary Reference Values**; **Dietary Requirements of Adults**; **Spectroscopy**: Visible Spectroscopy and Colorimetry; **Vitamins**: Overview; Determination

Further Reading

AACC (2000) Niacin-microbiological method. AACC methods 86-51 and 86-50A. In: Grami B (ed.) *American Association of Cereal Chemists Methods*, 10th edn. St. Paul, MN, USA: American Association of Cereal Chemists.

AOAC (2002) Niacin and niacinamide. Methods 985.34, 944.13 and 975.41. In: Horwitz W (ed.) *Official Methods of Analysis*, 17th edn. Gaithersburg, MD: AOAC International.

Ball GFM (ed.) (1998) Niacin and tryptophan. In: *Bioavailability and Analysis of Vitamins in Foods*, pp. 319–359.London: Chapman & Hall.

Eitenmiller RR and DeSouza S (1985) Niacin. In: Augustin J, Klein B, Becker D and Venugopal P (eds) *Methods of Vitamin Assay*, 4th edn, pp. 385–398. New York: Wiley.

Lahely S, Bergaentzle M and Hasselmann C (1999) Fluorimetric determination of niacin in foods by high-performance liquid chromatography with post-column derivatization. *Food Chemistry* 65: 129–133.

Report of the Committee on Medical Aspects of Food Policy (1991) *Dietary Reference Values for Food Energy and Nutrients for the UK. DHS Report on Health and Social Subjects*, no. 41. Report of the Committee on Medical Aspects of Food Policy. London: Her Majesty's Stationery Office.

Shibata K and Taguchi H (2000) Nicotinic acid and nicotinamide. In: De Leenheer A, Lambert WE and Van Bocxlaer JF (eds), *Modern Chromatographic Analysis of Vitamins*, 3rd edn. pp. 325–364. New York: Marcel Dekker.

Physiology

D A Bender, University College London, London, UK

Introduction

The vitamin niacin (nicotinic acid and nicotinamide) forms the functional moiety of the nicotinamide nucleotide coenzymes, nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP). These act as electron carriers in a wide variety of oxidation and reduction reactions. NAD is also the source of adenosine diphosphate (ADP)-ribose for the ADP-ribosylation of proteins, a mechanism of enzyme regulation, and the poly(ADP-ribosylation) of nucleoproteins controlling the deoxyribonucleic acid (DNA) repair mechanism. Nicotinic acid adenine dinucleotide and cyclic ADP-ribose both have a role in regulation of intracellular calcium concentration in response to hormone action.

Although niacin is generally regarded as a vitamin, it is not strictly a dietary essential, since the nicotinamide moiety of the coenzymes can be synthesized *in vivo* from the essential amino acid tryptophan. Under normal circumstances, the intake of tryptophan is probably adequate to meet niacin requirements without the need for any preformed niacin in the diet.

Metabolism of Niacin

Dietary Forms and Sources

Niacin is present in tissues, and therefore in foods, largely as NAD and NADP. The postmortem hydrolysis of NAD(P) is extremely rapid in animal tissues, so it is likely that much of the niacin of meat (a major dietary source of the vitamin) is free nicotinamide.

Coffee may provide significant amounts of nicotinic acid, formed as a result of the pyrolysis of trigonelline (*N*-methyl nicotinic acid) during roasting.

In the calculation of niacin intakes, the niacin content of cereals is normally ignored. Although chemical analysis reveals niacin in cereals (largely in the bran), this is mostly biologically unavailable, since it is bound as niacytin – nicotinoyl esters to a variety of polysaccharides and glycopeptides with molecular weights ranging between 1500 and 17 000. Although niacytin is not hydrolyzed by digestive enzymes, a small amount is hydrolyzed nonenzymically by gastric acid, and up to 10% of the niacytin in cereals may be biologically available.

Treatment of cereals with alkali (e.g., soaking overnight in calcium hydroxide solution, the traditional method for the preparation of tortillas in Mexico) or baking with alkaline baking powders releases much of the bound nicotinic acid. Roasting of whole-grain maize has a similar effect, since there is enough ammonia released from glutamine to form free nicotinamide by ammonolysis.

Digestion and Absorption

Nicotinamide nucleotides in the intestinal lumen are not absorbed as such, but undergo hydrolysis to free nicotinamide. A number of intestinal bacteria have nicotinamide deamidase activity, and a proportion of dietary nicotinamide may be deamidated in the intestinal lumen.

Both nicotinic acid and nicotinamide are absorbed from the small intestine by a sodium-dependent saturable process of active transport, although at unphysiologically high concentrations there is also passive diffusion across the intestinal mucosa.

Synthesis of the Nicotinamide Nucleotide Coenzymes

As shown in **Figure 1** NAD(P) can be synthesized from either of the niacin vitamers, or from quinolinic acid, which is a metabolite of the amino acid tryptophan.

In liver there is little utilization of preformed niacin for nucleotide synthesis. The enzymes for nicotinic acid and nicotinamide utilization are more or less saturated with their substrates at normal concentrations in the liver, and hence are unlikely to

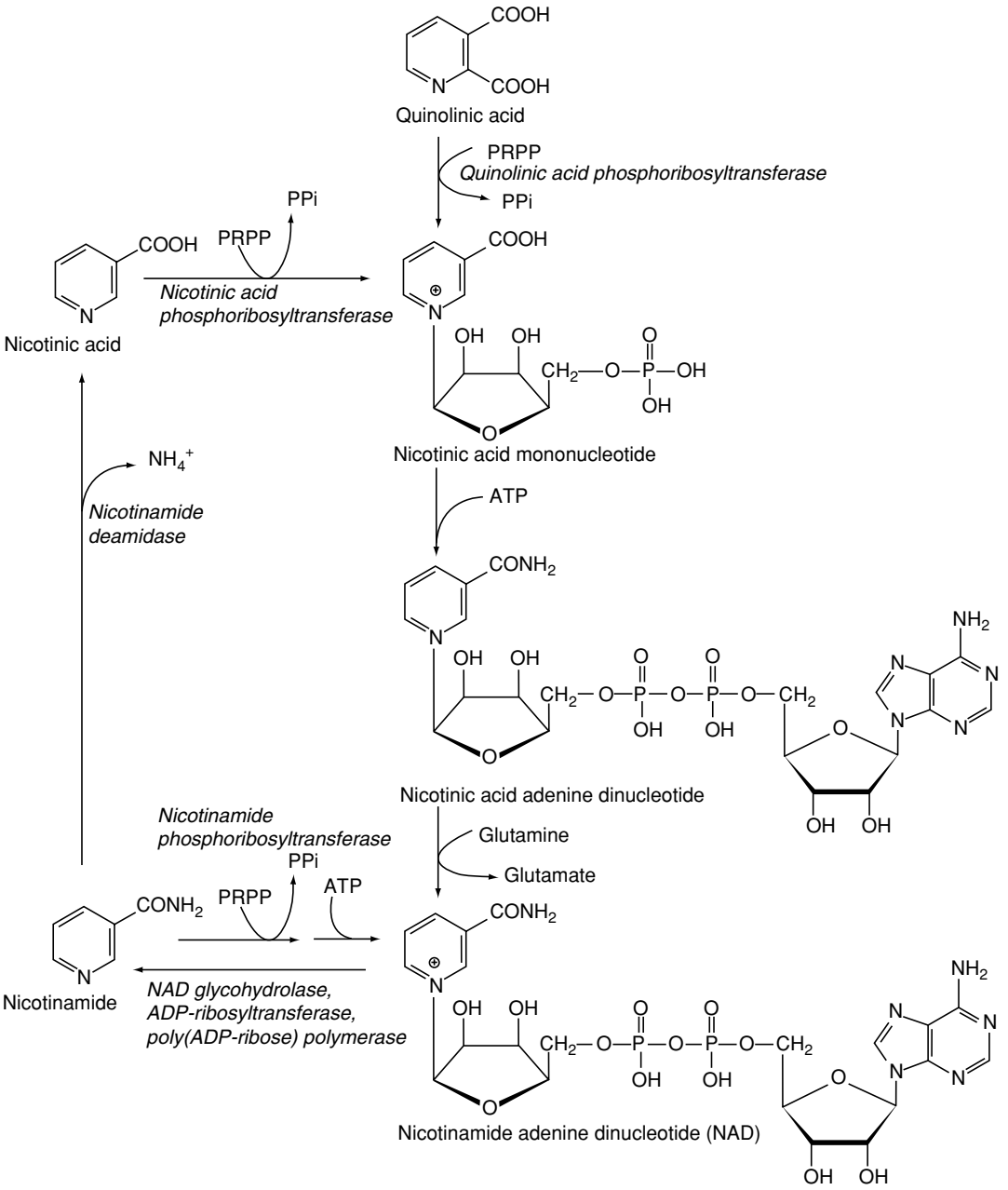

Figure 1 Synthesis of nicotinamide adenine dinucleotide (NAD) from nicotinamide, nicotinic acid, and quinolinic acid. Quinolinate phosphoribosyltransferase EC 2.4.2.19, nicotinic acid phosphoribosyltransferase EC 2.4.2.11, nicotinamide phosphoribosyltransferase EC 2.4.2.12, nicotinamide deamidase EC 3.5.1.19, NAD glycohydrolase EC 3.2.2.5, NAD pyrophosphatase EC 3.6.1.22, ADP-ribosyltransferases EC 2.4.2.31 and 2.4.2.36, poly(ADP-ribose) polymerase EC 2.4.2.30.

be able to use additional niacin for nucleotide synthesis. The liver synthesizes relatively large amounts of NAD(P) from tryptophan, followed by hydrolysis to release nicotinic acid and nicotinamide into the circulation for use by other tissues.

In most extrahepatic tissues, nicotinic acid is a better precursor of nucleotides than is nicotinamide. However, muscle and brain are able to take up nicotinamide from the blood stream effectively, and apparently utilize it without prior deamidation.

Synthesis of Nicotinamide Nucleotides from Tryptophan

The oxidative pathway of tryptophan metabolism is shown in Figure 2. Under normal conditions, almost all of the dietary intake of tryptophan, apart from the

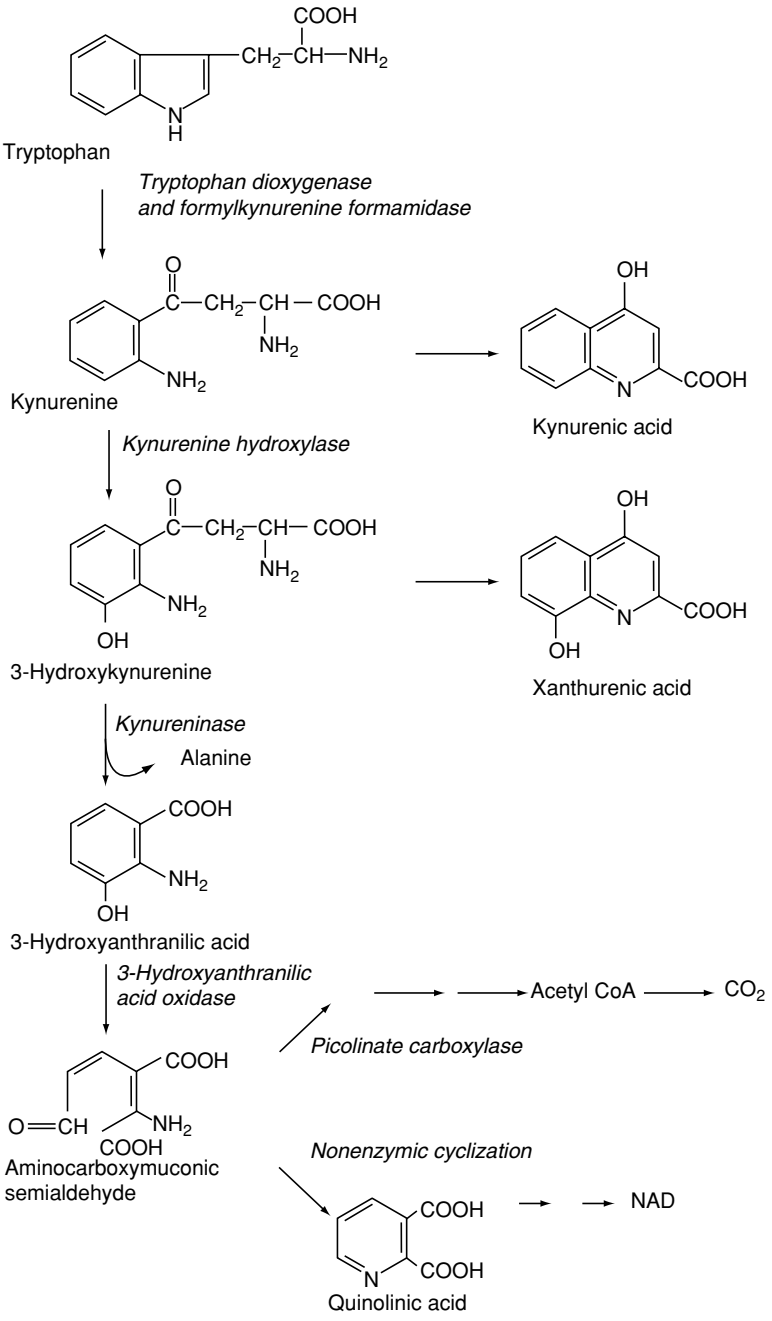

Figure 2 Tryptophan metabolism and the formation of quinolinic acid: tryptophan dioxygenase EC 1.13.11.11, formylkynurenine formamidase EC 3.5.1.9, kynurenine hydroxylase EC 1.14.13.9, kynureninase EC 3.7.1.3, 3-hydroxyanthranilate oxidase EC 1.10.3.5, picolinate carboxylase EC 4.1.1.45.

small amount that is used for net new protein synthesis, is metabolized by this pathway, and hence is potentially available for NAD synthesis. About 1% of tryptophan metabolism is by way of 5-hydroxylation and decarboxylation to form the neurotransmitter 5-hydroxytryptamine (serotonin).

The equivalence of dietary tryptophan and preformed niacin as precursors of the nicotinamide nucleotides has been assessed by determining the excretion of N^1-methylnicotinamide and methylpyridone carboxamide in response to test doses of the precursors, in subjects maintained on controlled diets. There is a considerable variation between subjects. It is generally accepted that in order to allow for individual variation it should be assumed that 60 mg of tryptophan is equivalent to 1 mg of preformed

niacin – an overestimate of the average requirement of tryptophan for NAD synthesis. This is the basis for expressing niacin requirements and intake in terms of niacin equivalents – the sum of preformed niacin and 1/60 of the tryptophan.

The synthesis of NAD from tryptophan involves the nonenzymic cyclization of aminocarboxymuconic semialdehyde to quinolinic acid. The alternative metabolic fate of aminocarboxymuconic semialdehyde is decarboxylation, catalyzed by picolinate carboxylase, leading to the formation of acetyl coenzyme A. There is thus competition between an enzyme-catalyzed reaction, which has hyperbolic, saturable kinetics, and a nonenzymic reaction which has linear, first-order kinetics.

At low rates of tryptophan metabolism, most will be by way of the enzyme-catalyzed pathway, leading to oxidation, and there will be little accumulation of aminocarboxymuconic semialdehyde to undergo nonenzymic cyclization. As the rate of formation of aminocarboxymuconic semialdehyde increases, and picolinate carboxylase becomes saturated, there will be an increasing amount available to undergo cyclization to quinolinic acid, and hence onward metabolism to NAD. There is thus not a simple stoichiometric relationship between tryptophan and niacin, and the equivalence of the two coenzyme precursors will vary as the amount of tryptophan to be metabolized and the rate of metabolism vary.

The activities of three enzymes, tryptophan dioxygenase, kynurenine hydroxylase, and kynureninase, may all affect the rate of formation of aminocarboxymuconic semialdehyde, as may the rate of uptake of tryptophan into the liver.

Tryptophan dioxygenase Tryptophan dioxygenase (also known as tryptophan oxygenase or tryptophan pyrrolase) has a short half-life *in vivo* (of the order of 2 h) and is subject to regulation by three mechanisms: stabilization by its heme cofactor, hormonal induction, and feedback inhibition and repression by high concentrations of NADP.

The holoenzyme is more stable than the apoenzyme, and in the presence of relatively large amounts of heme both the activity and the total amount of enzyme protein in the liver are increased. Induction of heme synthesis thus results in increased oxidative metabolism of tryptophan. This is not induction of tryptophan dioxygenase apoenzyme, but the result of reduced catabolism of the enzyme protein.

Tryptophan and a number of tryptophan analogs have a similar effect by promoting conjugation of the apoenzyme with hematin, and thus stabilizing the holoenzyme.

Tryptophan dioxygenase is induced by glucocorticoid hormones and glucagon; the mechanisms involved are different, and the effects are at least partially additive.

Glucocorticoid hormones cause induction of the new messenger ribonucleic acid (mRNA) and protein synthesis, unlike the increase in activity observed in the presence of higher than normal amounts of tryptophan or heme. In response to the administration of the synthetic glucocorticoid, dexamethasone, there is increased transcription of the rat liver tryptophan dioxygenase gene, resulting in a 10-fold increase in tryptophan dioxygenase mRNA in the liver.

Glucagon, mediated by cyclic adenosine monophosphate (cAMP), increases the synthesis of tryptophan dioxygenase following the administration of glucocorticoids, although it has little effect in unstimulated animals. The effect of glucagon appears to be the result of an increase in the rate of translation of mRNA rather than an increase in transcription, and is antagonized by insulin.

Kynurenine hydroxylase and kynureninase The activities of kynurenine hydroxylase and kynureninase are only slightly higher than that of tryptophan dioxygenase under basal conditions. Impairment of the activity of either enzyme may reduce the rate of tryptophan metabolism, and so reduce the accumulation of aminocarboxymuconic semialdehyde, and the synthesis of NAD.

Kynurenine hydroxylase is a flavoprotein, and in riboflavin-deficient rats its activity is only 30–50% of that in control animals. Riboflavin deficiency may thus be a contributory factor in the etiology of pellagra when intakes of tryptophan and niacin are marginal.

In a number of studies, sexually mature women show a higher ratio of urinary kynurenine to hydroxykynurenine than do children, postmenopausal women, or men, suggesting impairment of kynurenine hydroxylase by endogenous estrogens or their metabolites. In experimental animals the administration of estrogens results in a very considerable reduction in kynurenine hydroxylase activity. The mechanism of this effect is unclear, but it is physiologically important; in most areas where pellagra was common, about twice as many women as men were affected.

Kynureninase is a pyridoxal phosphate-dependent enzyme; impairment of its activity in vitamin B_6 deficiency leads to accumulation of kynurenine and hydroxykynurenine, and their transamination products, kynurenic and xanthurenic acids. This is the basis of the tryptophan load test for assessing vitamin B_6

nutritional status. It is also inhibited by estrogen metabolites, and both vitamin B_6 deficiency and inhibition reduce the rate of metabolic flux through the oxidative pathway, hence reducing the formation of quinolinic acid and NAD from tryptophan.

Catabolism of the Nicotinamide Nucleotide Coenzymes

There is no evidence of any specific storage of niacin or the nicotinamide nucleotide coenzymes in the body. Free NAD(P), which is not associated with enzymes, is rapidly hydrolyzed, and the resultant nicotinamide is either used for resynthesis of nucleotides or is methylated and excreted. The catabolism of NAD is catalyzed by four enzymes:

1. NAD glycohydrolase, which catalyzes the hydrolysis of $NAD(P)^+$ at the *N*-glycoside linkage to yield nicotinamide and either ADP-ribose or ADP-ribose phosphate. This enzyme catalyzes hydrolysis of both NAD^+ and $NADP^+$. It is also important in the formation of nicotinic acid adenine dinucleotide and cyclic ADP-ribose
2. NAD pyrophosphatase, which releases nicotinamide mononucleotide. This can either be hydrolyzed by NAD glycohydrolase to release nicotinamide, or be a substrate for nicotinamide mononucleotide pyrophosphorylase, to form NAD
3. ADP-ribosyltransferase(s)
4. Poly(ADP-ribose) polymerase

ADP-ribosyltransferases and poly(ADP-ribose) polymerase normally transfer ADP-ribose on to acceptor proteins (see below), although both are also able to catalyze simple hydrolysis of NAD^+ in the absence of an acceptor protein.

Urinary Excretion of Niacin and Metabolites

There is normally little or no urinary excretion of either nicotinamide or nicotinic acid, because both vitamers are actively resorbed from the glomerular filtrate. It is only when the plasma concentration is so high that the resorption mechanism is saturated that significant excretion occurs.

The metabolites of niacin are shown in Figure 3, and the use of their urinary excretion as an index of niacin nutritional status in Table 1. The principal metabolite of nicotinamide is N^1-methylnicotinamide, which is actively secreted into the urine by the proximal renal tubule. N^1-Methylnicotinamide can also be metabolized further, to yield methylpyridone carboxamide. The extent to which this oxidation occurs, and the relative proportions of the 2-pyridone and 4-pyridone, varies from one species to another and also shows considerable variation between different strains of the same species. Aldehyde oxidase catalyzes the formation of both pyridones, and some additional 2-pyridone arises from the activity of xanthine oxidase.

Nicotinamide can also undergo oxidation to nicotinamide *N*-oxide. This is normally a minor metabolite in humans, unless large amounts of nicotinamide are ingested. At high levels of nicotinamide intake, some 6-hydroxynicotinamide may also be excreted.

Nicotinic acid can be conjugated with glycine to form nicotinuric acid (nicotinoyl glycine), or may be methylated to trigonelline (N^1-methylnicotinic acid). Small amounts of 6-hydroxynicotinic acid may also be formed.

It is not clear to what extent urinary excretion of trigonelline reflects endogenous methylation of nicotinic acid. There are significant amounts of trigonelline in foods, and some is formed by the intestinal bacterial metabolism of niacytin. Trigonelline is absorbed, but cannot be utilized as a source of niacin, and is excreted unchanged.

Nicotinic acid can give rise to N^1-methylnicotinamide and methyl pyridone carboxamide as a result of its incorporation into NAD(P) and subsequent hydrolysis to release nicotinamide. Similarly, nicotinamide can give rise to nicotinic acid, and hence nicotinuric acid and trigonelline, as a result of deamidation. However, test doses of nicotinamide are excreted mainly as N^1-methylnicotinamide, pyridones, and nicotinamide *N*-oxide, while doses of nicotinic acid are excreted mainly as nicotinuric acid.

Metabolic Functions of Niacin

Redox Function of NAD(P)

The nicotinamide nucleotide coenzymes are involved as proton and electron carriers in a wide variety of oxidation and reduction reactions. Before their chemical structures were known, NAD and NADP were known as coenzymes I and II respectively. Later, when the chemical nature of the pyridine ring of nicotinamide was discovered, they were called diphosphopyridine nucleotide, or DPN (NAD), and triphosphopyridine nucleotide, or TPN (NADP). These names will still be found in the literature, and the nicotinamide nucleotide coenzymes are sometimes referred to as the pyridine nucleotide coenzymes.

As shown in Figure 4, the oxidized coenzymes have a formal positive charge, and are represented as NAD^+ and $NADP^+$, while the reduced forms, carrying two electrons and one proton (and associated with an additional proton) are represented as NADH and NADPH. The two-electron reduction of $NAD(P)^+$ proceeds by way of a hydride (H^-) ion transfer to carbon-4 of the nicotinamide ring.

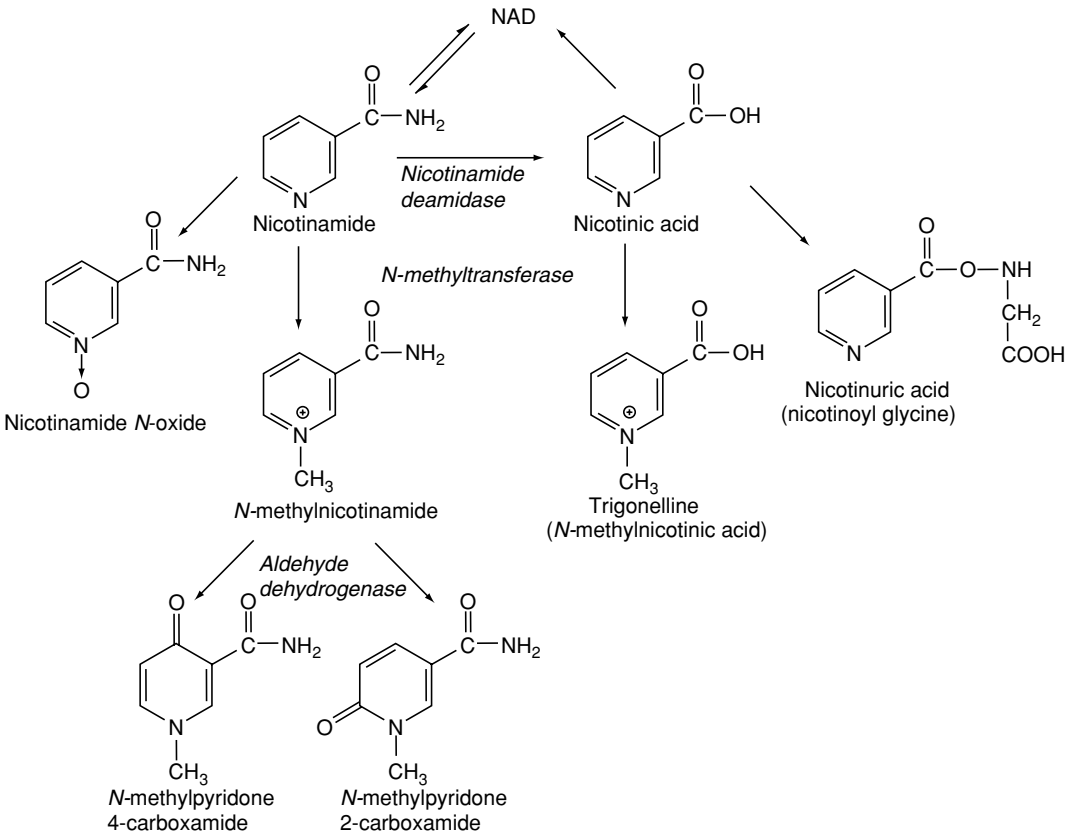

Figure 3 Metabolites of nicotinamide and nicotinic acid.

Table 1 Urinary excretion of niacin metabolites as an index of niacin nutritional status

	Elevated	*Adequate*	*Marginal*	*Deficient*
N^1-methylnicotinamide				
μmol $24\,h^{-1}$	> 48	17–47	5.8–17	< 5.8
mg g^{-1} creatinine	> 4.4	1.6–4.3	0.5–1.6	< 0.5
mmol mol^{-1} creatinine	> 4.0	1.3–3.9	0.4–1.3	< 0.4
Methylpyridone carboxamide				
μmol $24\,h^{-1}$		> 18.9	6.4–18.9	< 6.4
mg g^{-1} creatinine		> 4.0	2.0–3.9	< 2.0
mmol mol^{-1} creatinine		> 4.4	0.44–4.3	< 0.44
Ratio of methyl pyridone carboxamide to N^1-methylnicotinamide		1.3–4.0	1.0–1.3	< 1.0

In general, NAD^+ is involved as an electron acceptor in energy-yielding metabolism, being oxidized by the mitochondrial electron transport chain, while the major coenzyme for reductive synthetic reactions is NADPH. An exception here is the pentose phosphate pathway (hexose monophosphate shunt), which reduces $NADP^+$ to NADPH, and is the principal metabolic source of reductant for fatty acid synthesis.

Role of NAD in ADP-ribosylation of Proteins

ADP-ribosyltransferases catalyze the transfer of ADP-ribose from NAD^+ on to arginine, lysine, or asparagine residues in acceptor proteins, to form *N*-glycosides. In addition to endogenous ADP-ribosyltransferases, a number of bacterial toxins, including diphtheria and cholera toxins, *Escherichia coli* enterotoxin LT and *Pseudomonas aeruginosa* exotoxin A, also have ADP-ribosyltransferase activity, which is not subject to the same regulation as the enzymes in tissues.

ADP-ribosylation (Figure 5) is a reversible modification of proteins, and there are specific hydrolases which cleave the *N*-glycoside linkage. A variety of guanine nucleotide-binding protein (G protein)

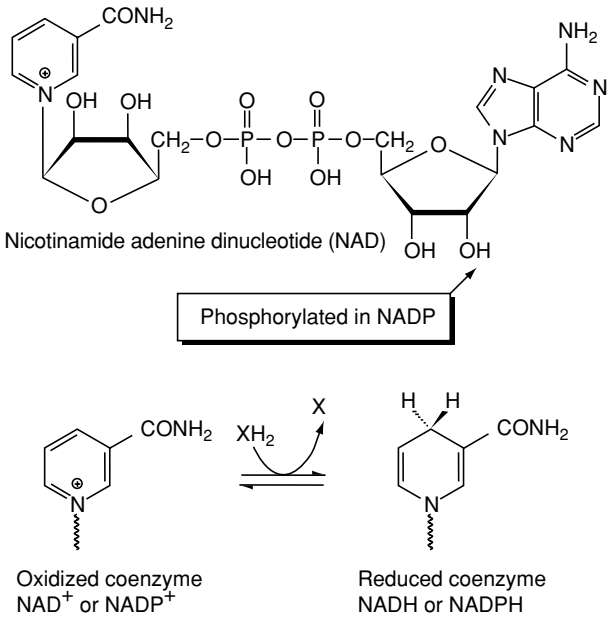

Figure 4 Oxidation and reduction of the nicotinamide coenzymes.

α-subunits involved with the regulation of adenylate cyclase activity are substrates for ADP-ribosylation, either activating the stimulatory G protein or inactivating the inhibitory G protein. The result of ADP-ribosylation by either mechanism is increased adenylate cyclase activity, and hence an increase in intracellular cAMP and the opening of membrane calcium channels. Other proteins that are regulated by ADP-ribosylation include the elongation factor(s) in protein synthesis and integrin and other cytoskeletal proteins.

Poly(ADP-ribose) polymerase (Figure 5) is primarily a nuclear enzyme, which is present in relatively large amounts – up to 1 mol per 1000 base pairs in DNA. It is a zinc finger protein that binds to a nick in DNA caused by strand breakage or excision of an incorrect base. When bound to DNA the enzyme undergoes an autocatalytic poly(ADP-ribosylation), adding branched chains of up to 200 ADP-ribose monomers at each of multiple sites on the enzyme. This poly(ADP-ribosylated) protein then interacts with the α-helical tail of histones, and displacing histones from DNA binding, so leaving the site clear for the DNA repair enzymes to act. There is then slow hydrolysis of the poly(ADP-ribose), permitting the histones to return to bind the repaired DNA.

There is some evidence that poly(ADP-ribose) polymerase may also be involved in DNA replication, since it copurifies with the replication fork and enzymes of DNA replication.

NAD Glycohydrolase and Intracellular Calcium Regulation

A cell membrane-bound NAD glycohydrolase catalyzes a base exchange between NAD^+ and nicotinic acid, yielding nicotinic acid adenine dinucleotide, as shown in Figure 6. The same enzyme also catalyzes exchange of the nicotinamide of NAD^+ with a variety of other compounds, including histamine; there is some evidence that this may provide a mechanism for rapid sequestration, and hence inactivation, of histamine.

The enzyme-bound ADP-ribose that is the intermediate in this base exchange reaction may also undergo either hydrolysis to yield ADP-ribose (so that the enzyme catalyzes overall hydrolysis of NAD^+ to nicotinamide and ADP-ribose), or internal cyclization to release cyclic ADP-ribose.

Both nicotinic acid adenine dinucleotide and cyclic ADP-ribose act to release intracellular stores of calcium. They appear to act as second messengers in transmembrane signal transduction, although the extracellular ligand has not been identified. There is evidence that all-*trans*-retinoic acid stimulates the cyclase activity of the enzyme, although it is not clear whether this is by a cell surface action or as a result of nuclear actions. In β-islet cells of the pancreas, glucose acts to stimulate the cyclase activity, and the resultant cyclic ADP-ribose causes a calmodulin-dependent mobilization of calcium and increased secretion of insulin.

Requirements and Recommendations

In view of the central role of the nicotinamide nucleotides in energy-yielding metabolism, and the fact that, at least under normal conditions, the nicotinamide released by ADP-ribosyltransferase and poly(ADP-ribose) polymerase is available to be reutilized for nucleotide synthesis, niacin requirements are usually expressed per unit of energy expenditure.

Depletion and repletion studies suggest that, on the basis of urinary excretion of N^1-methylnicotinamide, the average niacin requirement is 5.5 mg per 1000 kcal (1.3 mg per MJ). Allowing for individual variation, recommended dietary allowances (RDAs) in most countries are set at 6.6 mg niacin equivalents (preformed niacin plus 1/60 of the dietary tryptophan) per 1000 kcal (1.6 mg per MJ). When energy intakes are very low, it is assumed that expenditure will not fall below about 2000 kcal, and this is the basis for the calculation of RDAs for adults with low energy intakes.

There is little requirement for preformed niacin in the diet, since average intakes of protein (at least in developed countries) will provide enough tryptophan

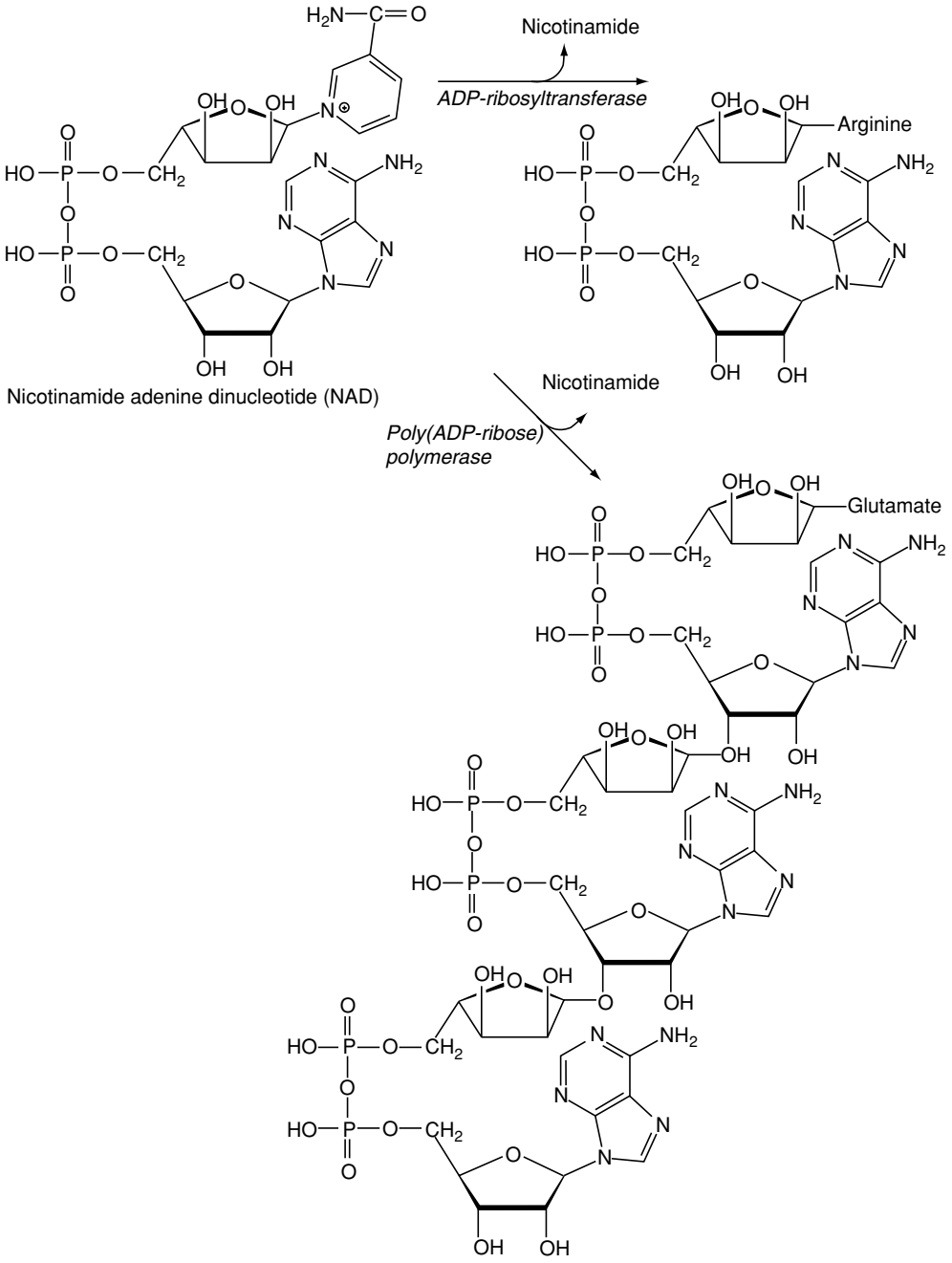

Figure 5 The reactions of adenosine diphosphate (ADP)-ribosyltransferase and poly(ADP-ribose) polymerase.

to meet requirements. Assuming that the diet provides some 15% of energy from protein, and this protein provides 14 g of tryptophan per kg, this implies an intake of 37.5 g of protein (525 mg of tryptophan) per 1000 kcal. Since 60 mg of tryptophan is equivalent to 1 mg of dietary niacin, this suggests that an average diet provides 8.75 mg niacin equivalents per 1000 kcal (2 mg per MJ) from tryptophan alone.

Groups at Risk of Niacin Deficiency (Pellagra)

The tryptophan–niacin deficiency disease pellagra has been a problem in areas of the world where maize or sorghum is the dietary staple, and intakes of both tryptophan and niacin are inadequate. The problem may be compounded by deficiency of riboflavin or vitamin B_6, both of which are required for the synthesis of NAD from tryptophan.

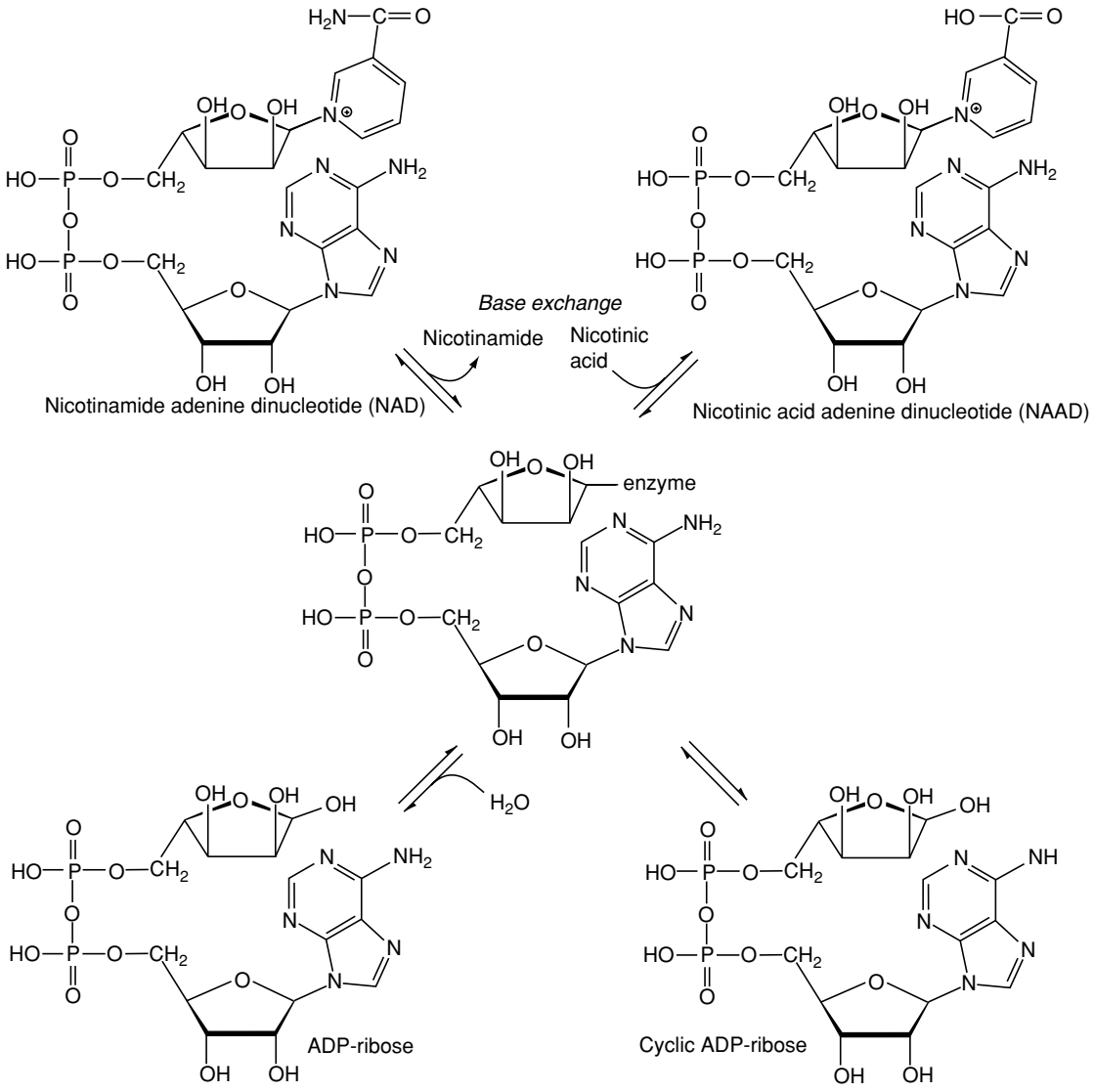

Figure 6 The reactions of nicotinanide adenine dinucleotide (NAD) glycohydrolase and formation of nicotinic acid adenine dinucleotide and cyclic adenosine diphosphate (ADP)-ribose.

A number of mycotoxins and chemotherapy agents used in the treatment of cancer activate poly(ADP-ribose) polymerase; exposure may result in significant depletion of NAD, and hence contribute to the development of pellagra.

Pellagra may arise as a result of drug-induced inhibition of tryptophan metabolism, e.g., by the anti-tuberculosis drug isoniazid (iso-nicotinic acid hydrazide) and the anti-parkinsonian drugs benserazide and carbidopa. Massively increased synthesis of 5-hydroxytryptamine, as occurs in the carcinoid syndrome, results in the development of pellagra, as a result of diversion of tryptophan away from NAD synthesis.

Pellagra also arises as a result of rare inborn errors of tryptophan metabolism or the impairment of tryptophan absorption (Hartnup disease).

Toxicity

Nicotinic acid has been used clinically in large doses (of the order of 1–3 g per day) as a hypolipidemic agent. It reduces both triglycerides and total cholesterol by about 20%, acting as an inhibitor of cholesterol synthesis. It has a more marked effect on cholesterol in low-density and very-low-density lipoproteins, and increases high-density lipoprotein cholesterol.

Nicotinic acid in such doses caused a marked vasodilatation, with flushing, burning, and itching of the skin; after a large dose there may be sufficient vasodilatation to cause hypotension. After the administration of 1–3 g of nicotinic acid daily for several days, the effect wears off to a considerable extent.

At intakes in excess of 1 g of niacin per day there are ultrastructural changes in the liver, and changes in liver

function tests, carbohydrate tolerance, and uric acid metabolism, which are reversible on withdrawal of niacin. Sustained-release preparations are associated with more severe liver damage than simple preparations, and may cause clinical liver failure, presumably because they permit more prolonged maintenance of high blood and tissue concentrations of the vitamin.

Supplements of several grams of tryptophan per day have been used with some success in the treatment of depressive diseases, apparently without ill effect. However, a potentially fatal eosinophilia–myalgia syndrome associated with the use of tryptophan supplements has been reported, with more than 1200 cases reported to the US Center for Disease Control in 1989. It seems most likely that the problem was due to a trace contaminant (ethylidene *bis*-tryptophan) in a single batch of tryptophan, rather than toxicity of tryptophan *per se*.

See also: **Cereals**: Dietary Importance; **Coffee**: Analysis of Coffee Products; **Enzymes**: Functions and Characteristics; **Pellagra**; **Vitamins**: Overview; Determination

Further Reading

Bender DA (1983) Biochemistry of tryptophan in health and disease. *Molecular Aspects of Medicine* 6: 101–197.

Bender DA (2003) *Nutritional Biochemistry of the Vitamins*, 2nd edition. New York: Cambridge University Press.

Bender DA and Bender AE (1986) Niacin and tryptophan metabolism: the biochemical basis of niacin requirements and recommendations. *Nutrition Abstracts and Reviews (Series A)* 56: 695–719.

De Murcia G and de Murcia JM (1994) Poly(ADP-ribose) polymerase: a molecular nick-sensor. *Trends in Biochemical Sciences* 19: 172–176.

Desnoyers S, Shah GM, Broche G, Hoflack JC, Verreault A and Poirier GG (1995) Biochemical properties and function of poly(ADP-ribose) glycohydrolase. *Biochimie* 77: 433–438.

Frei B and Richter C (1988) Mono-(ADP-ribosylation) in rat liver mitochondria. *Biochemistry* 27: 529–535.

Galione A and White A (1994) Ca^{2+} release induced by cyclic ADP-ribose. *Trends in Cell Biology* 4: 431–436.

Moss J and Vaughan M (1988) ADP-ribosylation of guanyl nucleotide binding regulatory proteins by bacterial toxins. *Advances in Enzymology* 61: 303–379.

Okamoto H, Takasawa S and Tohgo A (1995) New aspects of the physiological significance of NAD, poly ADP-ribose and cyclic ADP-ribose. *Biochimie* 77: 356–363.

Rose DP (1972) Aspects of tryptophan metabolism in health and disease. *Journal of Clinical Pathology* 25: 17–25.

Ueda K and Hayaishi O (1985) ADP-ribosylation. *Annual Review of Biochemistry* 54: 73–100.

Nicotinic Acid *See* **Niacin**: Properties and Determination; Physiology

NISIN

G C Williams and J Delves-Broughton, Danisco Innovation, Beaminster, Dorset, UK

Introduction

Nisin is a polypeptide bacteriocin that exhibits antibacterial activity against a wide range of Gram-positive bacteria and is particularly effective against bacterial spores. It shows little or no activity against Gram-negative bacteria, yeasts, and molds. Nisin is produced by certain strains of *Lactococcus lactis* subsp. *lactis*. Commercial preparations of nisin are widely used as food preservatives throughout the world. It is recognized as both a toxicologically safe and natural substance.

History

Nisin was discovered in England in 1928 when problems arose in cheese-making. Batches of milk became contaminated with a nisin-producing strain of *L. lactis*, and as a result of nisin's inhibitory properties, the growth of the cheese starter cultures was inhibited. Nisin was subsequently isolated and characterized. Its name was derived from Group *N* (*Streptococcus*) inhibitory substance. The discovery of nisin predated that of penicillin, so therefore it is not surprising that initial research focused on its potential use for

therapeutic purposes in medical and veterinary applications. It was found to be unsuitable for such purposes, mainly because of its limited antimicrobial spectrum in addition to its poor solubility and stability in body fluids. The first interest in its potential as a food preservative came in the 1950s when it was evaluated for the control of gas-producing *Clostridium* spp. in Swiss cheese. However, problems occurred due to the fact that the cheese starter cultures were inhibited, which adversely affected the cheese-ripening process. This problem has now been largely overcome by the development of nisin-resistant cheese starter cultures. The next milestone was its development as a preservative in processed cheese in the mid-1950s. During this period Aplin & Barrett Ltd were experiencing unacceptable spoilage problems due to clostridial growth in processed cheese spreads. This problem was eliminated by the inclusion of nisin-rich curd into the processed cheese mix. With this success the company went on to develop the commercial concentrated preparation of nisin termed Nisaplin. Nisaplin has a nisin content of 25 mg g^{-1} and is currently used worldwide as a food preservative. Recently Chr. Hansen from Denmark has introduced a similar product of the same potency, termed Chrisin.

Units of Activity

The first definition was that of a reading unit (later known as an international unit), which was defined as the amount of nisin necessary to inhibit one cell of *Streptococcus agalactiae* in 1 ml of broth. It was so termed in recognition that much of the early work on nisin was carried out at the National Institute for Research in Dairying, Shinfield (Reading), England. Since then an international reference preparation of nisin has been established by the World Health Organization Expert Committee in Biological Standardization. This reference preparation contains, like the commercial preparations Nisaplin and Chrisin, 25 mg (1 million international units) of pure nisin per gram. Throughout this article all levels of nisin are expressed as mg pure nisin per kilogram or liter.

Structure and Biosynthesis

Two natural nisin molecules exist, termed nisin A and nisin Z. The structure of the nisin A molecule was elucidated in 1971 and is presented in Figure 1. It is a 34-amino-acid polypeptide with amino and carboxyl endgroups, and five internal ring structures involving disulfide bridges. It possesses three unusual amino acids: dehydroalanine, lanthionine, and β-methyllanthionine. Lanthionine appears to be a common feature in a number of more recently characterized bacteriocins that are collectively known as lantibiotics. Nisin Z differs from nisin A by the substitution of asparagine for histidine at position 27. Nisin Z has a similar antimicrobial activity to nisin A, although nisin Z shows greater diffusion in agar gels. Japanese workers have successfully synthesized the nisin A molecule and confirmed its basic structure. Nisin A has a molecular weight of 3354 Da. There is evidence that nisin can exist as both dimers and tetramers.

A precursor molecule to nisin has been identified and its structural gene isolated and characterized. The

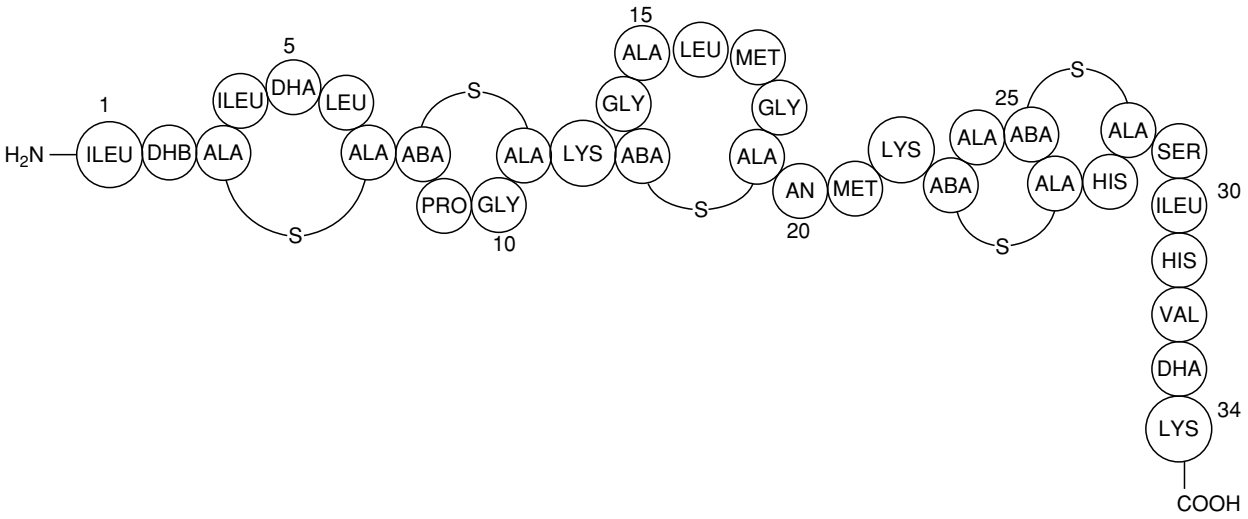

Figure 1 The structure of nisin A. ABA, aminobutyric acid; DHA, dehydroalanine; DHB, dehydrobutyrine (β-methyldehydroalanine); ALA-S-ALA, lanthionine; ABA-S-ALA, β-methyllanthionine.

nisin determinant of *L. lactis* is a component of a large transmissible gene block that also encodes for sucrose metabolism and nisin immunity. The gene block is located in the chromosome as opposed to being in a plasmid. The ability of *L. lactis* to synthesize nisin is a conjugally transmissible property that has been transferred to negative phenotype recipients.

Mode of Action and Antimicrobial Effect

Nisin works in a concentration-dependent manner both in terms of the amount of nisin applied and the number of vegetative cells or spores that need to be inhibited or killed. The primary site of action in sensitive Gram-positive cells is the cytoplasmic membrane where insertion of nisin into the membrane causes disruption of membrane function. An initial step is the docking of nisin on to lipid II, which is followed by the insertion of the nisin II lipid into the membrane where it forms pores. Through these pores, essential cell constituents leak from the inside to the outside of the cell and cause depletion of the proton motive force. Gram-negative bacteria are normally protected from nisin by their outer cell wall acting as an impermeable barrier to the nisin molecule. However, if the outer wall of Gram-negative cells is made permeable to nisin by partial or complete disruption, the cells can then become sensitive to nisin. This indicates that their cytoplasmic membranes are sensitive to nisin. Recent work has demonstrated that yeast protoplasts and spheroblasts are also sensitive to nisin. Gram-negative bacteria can be sensitized to nisin by sublethal heat treatment, freeze–thaw cycling, and exposure to chelating agents that remove divalent ions from their cell walls, thus making them more permeable to nisin. Good effects can be demonstrated against Gram-negative bacteria by the combined use of nisin and a chelating agent such as ethylenediaminetetraacetic acid (EDTA) in a simple buffer system. In food systems, however, such good results are not usually observed as the chelating agent will preferentially bind with the divalent ions that are more readily available in the food matrix.

The action of nisin against spores is predominantly sporostatic as opposed to sporicidal. Nisin affects the postgermination stages of spore development. It inhibits preemergent spore swelling, and thus the outgrowth and formation of vegetative cells. The active sites in spores are thought to be membrane-bound sulfhydryl groups. An important factor in relation to heat-processed foods is that progressive heat damaging of spores results in them becoming increasingly sensitive to nisin.

The nisin sensitivity of both cells and spores can vary between genera and even between strains of the same species. Nisin action against vegetative cells can either be bactericidal or bacteriostatic depending on a number of factors, such as nisin concentration, bacterial population size, physiological state of the bacteria, and the conditions of growth. Bactericidal effects against vegetative cells are enhanced under optimal growth conditions when the bacteria are in an energized state. In contrast, bacteriostatic effects are enhanced when nisin forms part of a multipreservation system in which growth conditions are nonoptimal and other inhibitory factors are exerted (hurdle technology). The fact that different conditions are required to insure either bacterial destruction or inhibition needs to be taken into consideration when applying nisin in a food preservation system. Nisin efficacy in foods is dependent on an effective level of nisin being maintained throughout the whole shelf-life of the food.

Solubility and Stability

Nisin is most soluble in acid substrates and becomes progressively less soluble as the pH increases. Thus at pH 2.2 the solubility is 56 000 mg l^{-1}, at pH 5 it is 3000 mg l^{-1} and at pH 7 it is 1000 mg l^{-1}. In practical food preservation situations the level of nisin treatment is unlikely to exceed 50 mg l^{-1} and thus solubility is never a problem. Commercial powder preparations of nisin contain residual solids from the fermentation process that are insoluble. This can produce cloudy suspensions in water but has no detrimental effect on the efficacy of nisin.

Commercial preparations of nisin are remarkably stable. They show no loss of activity over a 2-year period providing they are stored under dry conditions, in the dark, and at temperatures below 25 °C. Nisin solutions are most stable to autoclaving (121 °C for 15 min) in the pH range 3.0–3.5 (< 10% activity loss). Values below and above this pH range result in a marked loss of activity, especially those furthest removed from the range (> 90% activity loss at pH 1 or pH 7 in a buffer system). Losses at typical pasteurization temperatures used in foods are significantly less and food components can protect nisin during heat processing.

The stability of nisin in a food system is dependent on three factors: incubation temperature, length of storage, and pH. In a nonheat-processed food a fourth factor will be the possible presence of protease enzymes. Nisin retention at warm ambient temperatures above 25 °C will be far less than at cooler temperatures. On a practical level this means that higher nisin addition levels are required for the preservation of foods in warmer climates. Retention of nisin activity in an acidic food product will be better than in a

more neutral food stored under similar conditions. Nisin can be inactivated by many nonspecific proteases and by any proteolytic enzyme that can cleave the histidine-valine bond (residues 31–32) or the dehydroalanine bond (residues 33–34), e.g., thermolysin, chymotrypsin, subtilisin, ficin, papain, and bromelain. Elastase, pepsin, and leucine aminopeptidase have no action on nisin, and trypsin has a reversible action. A variety of bacteria can produce the enzyme 'nisinase' which specifically inactivates nisin. Some of the bacterial species reported as being able to produce nisinase include *Lactobacillus plantarum, Streptococcus thermophilus*, and *Bacillus cereus*.

The food additives, and related sulfiting agents sodium metabisulfite (an antioxidant, bleaching, and broad-spectrum antimicrobial agent) and titanium dioxide (a whitening agent) can also cause nisin degradation.

Methods of Assay

A number of bioassay methods have been devised. These include dye reduction methods using resazurin or methylene blue with a sensitive lactic acid bacterium in a milk-based medium, turbidometric growth measurement assay, horizontal agar plate diffusion assay, the bioluminescent measurement of released adenosine triphosphate (ATP) from *Lactobacillus casei*, and an enzyme-linked immunoabsorbent assay (ELISA). The method in most common use is the horizontal agar plate diffusion assay employing the test organism *Micrococcus luteus*. The lower limit of detection in this assay is about $0.025\,mg\,l^{-1}$. A recently described novel technique has been the development of a strain of *L. lactis* that can sense nisin and transduce the signal into bioluminescence with as little as $0.0125\,ng\,ml^{-1}$ nisin being detectable by this method.

Quantitative analysis of nisin can also be achieved by high-performance liquid chromatography (HPLC) analysis. A nisin-containing liquid is assayed chromatographically on a hydrophobic (C18) narrow-base HPLC column by gradient elution. Calculations are based on the peak height and quantification done by comparison with a standard nisin preparation. The lower limit of detection by HPLC is around $10\,mg\,l^{-1}$.

Toxicology and Legislation

Toxicity studies carried out in laboratory animals with levels of nisin far in excess of those used in foods have shown that nisin is nontoxic and is not carcinogenic. Nisin is rapidly inactivated in the intestine by digestive enzymes and cannot be detected in the saliva of human beings 10 min after consumption of liquid containing $5\,mg\,l^{-1}$ nisin. There is no evidence of sensitization (allergy problems) and microbiological studies have not shown any cross-resistance problems that may affect the efficacy of therapeutic antibiotics. It is important that nisin and other bacteriocins are not classified as antibiotics as this could hamper their future acceptance as food preservatives.

In 1969, the Joint Food and Agriculture Organization/World Health Organization (FAO/WHO) Expert Committee on Food Additives reviewed the toxicological data for nisin and recommended its use as a food preservative, with an acceptable daily intake (ADI) of $0.825\,mg\,kg^{-1}$ of body weight per day. At present its use is permitted in approximately 60 countries, including the US, the former USSR countries, and China. In the EU it has the food additive number E234.

Preservation of Foods using Nisin

Although nisin use as a food preservative originated with processed cheese products, other application areas have since been identified. Many of these are in products that, by their nature, are pasteurized during production but are not fully sterilized. In such foods Gram-negative bacteria, yeasts, and molds are destroyed by the heat treatment, and the surviving microflora consists of Gram-positive spore-forming bacteria that can be controlled by the use of nisin. The effectiveness of nisin in such products is only complete if postprocessing contamination is eliminated or minimized. Nisin is also used in canned vegetables to control thermophilic spoilage, in non-heated acidic products to control lactic acid bacteria, and in products that are at risk from the psychroduric food-poisoning bacterium, *Listeria monocytogenes*.

Table 1 summarizes the major categories of food in which nisin is used, and the typical spoilage or pathogenic bacteria in these products that are controlled by nisin.

Processed Cheese Products

Processed cheese products cover a wide range, including block cheese (approximately 44–46% moisture), slices (46–50% moisture), spreads (52–60% moisture), and sauces and dips (56–65% moisture). All are heat-processed and contain emulsifying salts. Product innovation in the industry is considerable and formulations can be of low fat or reduced sodium chloride content and may contain various flavor additives such as herbs, fish, shellfish, and meat. All these factors, along with bacterial quality of the raw ingredients, severity of the melt process, filling

Table 1 Typical addition levels of nisin (and Nisaplin) in examples of food applications

Type of food/application	*Addition level of nisin (mg kg^{-1} or mg l^{-1})*	*Addition level of Nisaplin (mg kg^{-1} or mg l^{-1})*	*Typical target organisms*
Processed cheese	5–15	200–600	*Bacillus* spp. *Clostridium* spp.
Pasteurized chilled dairy desserts	1.25–3.75	50–150	*Bacillus* spp.
Pasteurized milk and milk products	0.25–10.0	10–400	*Bacillus* spp. *Clostridium* spp.
Pasteurized liquid egg products	1.25–5.0	50–200	*Bacillus* spp., e.g., *B. cereus*
Pasteurized chilled soups	2.5–5.0	100–200	*Bacillus* spp.
Crumpets	3.75–6.25	150–250	*Bacillus cereus*
Canned foods (low acid)	2.5–5.0	100–200	*Bacillus stearothermophilus* *C. thermosaccharolyticum*
Canned foods (high acid)	1.25–2.5		*C. pasteurianum* *B. macerans* *B. coagulans*
Canned lobster	25.0	1000	*Listeria monocytogenes*
Ricotta cheese	5.0	200	*Listeria monocytogenes*
Continental-type cooked sausage:			
Added to mix	1.25–6.25	50–250	Lactic acid bacteria
Dipping	5.0–25.0	200–1000	
Salad dressings	1.25–5	50–200	Lactic acid bacteria
Beer:			
Pitching yeast wash	25–37.5	1000–1500	Lactic acid bacteria, e.g., *Lactobacillus* spp.
During fermentation	0.63–2.5	25–100	*Pediococcus* spp.
Postfermentation	0.25–1.25	10–50	

temperature, and shelf-life requirement can affect the microbial stability of processed cheese products and hence the requirement for nisin and the level at which it needs to be applied.

The ingredients used in the manufacture of these products are raw cheese, butter, skimmed milk powder, whey powder, phosphate, or citrate emulsifying salts and water. Spores of anaerobic clostridial species are often present in some of these ingredients, particularly the cheese, and are able to survive the heat process of 85–105 °C for 6–10 min commonly used in the heat process. The composition of processed cheese in terms of the relatively high pH (5.6–6.0) and moisture content combined with low redox potential (anaerobic conditions) can result in spore germination and growth, which may result in subsequent spoilage due to production of gas, off-odors, and digestion of the cheese. *Clostridium* spp. often associated with the spoilage of processed cheese are *C. sporogenes, C. butyricum*, and *C. tyrobutyricum*. Trials with processed cheese products have been carried out in the UK using a cocktail of spores of the aforementioned *Clostridium* spp. at inoculation levels of approximately 200 spores per gram. Spoilage was prevented during storage at 37 °C by 6.25 mg kg^{-1} nisin. Partial control was achieved with 2.5 mg kg^{-1} whilst control samples that did not contain nisin readily became spoiled.

The potential for growth and toxin production by *C. botulinum* in processed cheese products, particularly spreads, is of considerable significance. Trials in the US have indicated that, in processed cheese spreads, nisin is effective in delaying or preventing the growth and subsequent toxin production by inoculated spores of *C. botulinum* types A and B. These studies indicated that the use of nisin as an effective preservative in processed cheese spreads should form part of a multicomponent food preservation system. Levels of moisture, sodium chloride, phosphate emulsifier salt, and pH are all factors important in determining the necessary level of nisin to provide the required shelf-life. Facultative aerobic *Bacillus* spp. can also cause spoilage problems in processed cheese products and these organisms are also controllable using nisin. Levels used to prevent spoilage are 5–20 mg kg^{-1}, whereas levels used to provide protection against *C. botulinum* are 12.5 mg kg^{-1} and above.

Other Pasteurized Dairy Products

Other pasteurized dairy products, such as dairy desserts, cream, clotted cream, and mascarpone cheese, often cannot be subjected to full sterilization without damaging their organoleptic properties, texture, and/or appearance, and are thus sometimes preserved with nisin to extend their shelf-life. For example, tests on a chocolate dairy dessert resulted in a

20-day increase in shelf-life with 3.75 mg kg^{-1} nisin at 7 °C. The same nisin level gave a 30-day increase in shelf-life for a crème caramel dessert stored at 12 °C.

The addition of nisin to pasteurized milk is permitted in countries that may experience shelf-life problems due to high ambient temperatures, long-distance transport, and inadequate refrigeration. Trials at Reading University, UK, with nisin added at 1 mg l^{-1} before pasteurization at 72 °C for 15 s, 90 °C for 15 s, or 115 °C for 2 s all resulted in significant shelf-life extension of the milk when stored at 10 °C. Similar benefits to shelf-life have also been demonstrated with pasteurized flavored milk.

Pasteurized Liquid Egg Products

Pasteurized liquid egg products (whole, yellow, and white) and value-added egg products (e.g., omelets, scrambled eggs, pancake mixes) receive heat treatments designed to insure the destruction of *Salmonella*. In the UK, for instance, liquid whole egg must be pasteurized for at least 2.5 min at a temperature of 64.4 °C. However, such heat treatment is insufficient to kill bacterial spores and the more heat-resistant nonspore-forming Gram-positive bacteria such as *Enterococcus faecalis*. Many of these surviving bacteria are capable of growth at refrigerated temperatures and pasteurized egg products often have a short shelf-life. Application of nisin at levels of 2.5–5 mg l^{-1} has been shown to act as an effective preservative, giving significant increases in shelf-life and providing protection against the growth of the psychroduric food-poisoning bacteria *B. cereus* and *L. monocytogenes*.

Pasteurized Soups

A recent trend in soup manufacture has been a move towards the production of fresh pasteurized products with a relatively limited chilled shelf-life. Heat-resistant spores of *Bacillus* spp. are able to survive the pasteurization treatment and may be capable of growing and causing spoilage under conditions of chill abuse storage. Trials with nisin at levels of 2.5–5.0 mg l^{-1} have been found to be very effective at preventing or delaying the outgrowth of psychroduric spoilage *Bacillus* spp. during prolonged shelf-life storage of these pasteurized soup products.

High-moisture Hot Plate Products

Crumpets are high-moisture flour-based products that are popular in the UK, Australia, and New Zealand. Crumpets are produced on a hot plate from a flour batter and contain yeast, an aerating agent or both to give them a raised profile and open texture. They are toasted before eating. Crumpets have a non-acid pH (pH 6), high moisture (48–54%) and high water activity (0.95–0.97). The product is sold at ambient temperature and has a shelf-life of 5 days. There have been a number of food-poisoning outbreaks due to the growth of *B. cereus* in crumpets, particularly in Australasia. Flour used in the manufacture of crumpets invariably contains low numbers of *B. cereus* spores that are not killed during the hot plate cooking process. During the 3–5-day ambient shelf-life of the product, levels of *B. cereus* can increase from undetectable levels to $> 10^5$ cfu g^{-1} – sufficient bacteria to cause food poisoning. Addition of nisin to the batter mix at 3.75 mg kg^{-1} prevents the growth of *B. cereus* to these potentially dangerous levels. Such use of nisin has received regulatory approval in Australia and New Zealand.

Canned Foods

Nisin is used in canned foods mainly for the control of thermophilic spoilage. It is mandatory in most countries that low-acid canned foods (pH > 4.5) receive a minimum heat process of $F_0 = 3$ to insure the destruction of *C. botulinum* spores, i.e., the minimum botulinum cook. Low-acid foods processed at F_0 of 3 are susceptible to spoilage from surviving heat-resistant spores of thermophilic bacterial species of *B. stearothermophilus* (cause of flat sour spoilage) and *C. thermosaccharolyticum* (cause of can swells). Thus nisin addition can facilitate prolonged storage of canned vegetables at warm ambient temperatures by inhibiting spore outgrowth of these thermophilic spoilage organisms. The use of nisin can also allow a reduction in the F_0 process down to the minimum of 3 without increasing the potential risk of thermophilic spoilage. Other advantages are reduced heat damage to the foods as well as potential saving in energy consumption. Nisin usage levels in low-acid canned vegetables are 2.5–5.0 mg kg^{-1}. Residual nisin levels in canned foods after high temperature processing can be as low as 2% of the addition level. However, the fact that heat-resistant thermophilic spores are highly sensitive to nisin combined with the heat damage enhancing their sensitivity means that extremely low levels of residual nisin can still be effective in this application. Preacidification of the brine with citric acid improves nisin retention with minimal effect on the pH of the vegetables after processing.

Examples of use are canned peas, carrots, peppers, potatoes, mushrooms, okra, baby sweetcorn, and asparagus. Nisin is also used in canned dairy puddings containing semolina and tapioca.

Bacterial spoilage of canned high-acid foods (pH below 4.5) is restricted to nonpathogenic spoilage

species such as *C. pasteurianum*, *B. macerans*, and *B. coagulans*. Nisin addition levels of 1.25–2.50 mg kg^{-1} are used in high-acid tomato-based products.

Meat Products

Concern regarding the high levels of nitrite in cured meat has resulted in research investigating the use of nisin as a partial replacement for nitrite. Results indicated that only high (and uneconomic) levels of nisin achieved good control of *C. botulinum*. Further work is necessary before a case for such an application is demonstrated. More encouraging results have been obtained in vacuum-packed cooked continental-type sausages where lactic acid bacteria can cause spoilage by production of gas, off-odors, and slime. Addition of nisin into the sausage mix at levels of 1.25–6.25 mg kg^{-1} or dipping the cooked sausage into nisin solutions of 5.0–25.0 mg l^{-1} has proved effective in increasing shelf-life at storage temperatures of 6–12 °C. Investigations have shown that nisin is more inhibitory against lactic acid bacteria in sausages with lower fat levels. It was also shown that nisin had a greater effect in sausages containing diphosphate compared to those with orthophosphate.

Fish and Shellfish

Relatively few studies have been carried out using nisin in fresh fish mainly because the predominant flora tends to be Gram-negative bacteria. The potential hazard of botulism in both vacuum-packed and modified-atmosphere-packed fish has led to work at the Torry Research Station in the UK on the use of nisin as an antibotulinal agent. Application of nisin by spray to fillets of cod, herring, and smoked mackerel inoculated with *C. botulinum* type E spores resulted in a significant delay in toxin production at 10 and 26 °C. Another problem in smoked fish is the presence and growth of the psychroduric pathogen *L. monocytogenes*, especially in fresh and lightly preserved products. Nisin has been shown to be an effective antilisterial agent in smoked salmon, especially when packed in a carbon dioxide atmosphere.

L. monocytogenes can also be a problem in shellfish, particularly crabs and lobsters, as their meat can only be lightly heat-processed without significant product damage occurring. Nisin, in combination with a reduced heat process, that does not cause product damage of lobster meat, has been shown to achieve a *Listeria* kill that is significantly better than either heat or nisin used alone. An effective nisin level for this application is 25 mg kg^{-1}. Washing crabmeat with nisin has been shown to reduce levels of *L. monocytogenes*.

Salad Dressings

The development of cold blended salad dressings with reduced acidity can improve the flavor of many varieties that are considered to have an over-acid taste. Using reduced levels of acetic acid and raising the pH from 3.8 to 4.2 can make salad dressings prone to lactic acid bacterial spoilage during ambient storage. Such growth has been successfully controlled by the addition of nisin at 2.5–5.0 mg l^{-1}.

Natural Cheese

The first application of nisin was to prevent blowing problems in semihard ripe cheese such as Emmenthal and Gouda due to growth of *C. butyricum* and *C. tyrobutyricum*. Although promising results were obtained, a problem with inhibition of the starter cultures and consequent delay of the ripening process led to the work being discontinued. However, the increase in knowledge of lactic acid bacterial genetics and the need to devise methods for the control of *L. monocytogenes* has resulted in fresh interest into the use of nisin in natural cheese. To achieve success, nisin-resistant starters must be used in conjunction with nisin to insure successful development of the cheese. Natural nisin-producing, nisin-resistant strains tend to lack important properties associated with cheese starter cultures such as good flavor production, eye formation, acidifying characteristics, and bacteriophage resistance. Using the food-grade genetic transfer technique of conjugation it has been possible to develop nisin-producing, nisin-resistant starter cultures with the above-desired properties. Cheeses have been made with sufficient nisin content to provide protection against growth of *Clostridium* spp., *Staphylococcus aureus*, and *Listeria monocytogenes*.

Yogurt

The addition of nisin to stirred yogurt postproduction has an inhibitory effect on the starter culture (a mixture of *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Streptococcus thermophilus* strains), thereby preventing subsequent overacidification of the yogurt. Thus an increase in shelf-life is obtained by maintaining the flavor of the yogurt (less sour) and preventing syneresis. Typical addition levels for this application are 0.5–1.25 mg kg^{-1}.

Alcoholic Beverages

Acid-tolerant lactic acid bacteria of the genera *Lactobacillus, Pediococcus*, and *Leuconostoc* can spoil beer and wine due to growth along with production of off-flavors, off-odors, slime, or haze. At levels of 0.25–2.5 mg l^{-1}, nisin is effective in preventing such

spoilage. Yeasts are unaffected by nisin, thus the preservative can be added during the fermentation. Nisin can be added to fermenters to prevent or control contamination and can also be used to increase the shelf-life of unpasteurized and bottle-conditioned beers. Furthermore, nisin can be used in the pitching yeast wash as an alternative to acid washing for the control of lactic acid bacteria. Unlike nisin treatment, acid washing can have an adverse effect on yeast viability and performance. Typical levels for this application are 25.0–37.5 $mg\,l^{-1}$. Similar applications occur in the wine industry with the limitation that nisin cannot be used in wines that depend upon a desirable malolactic fermentation as the bacteria responsible are usually nisin-sensitive. This latter problem has been overcome by developing nisin-resistant strains of *Leuconostoc oenos* that can grow and maintain malolactic acid fermentation in the presence of nisin. In the production of distilled spirits, nisin can inhibit the lactic acid bacteria that compete with the yeast for substrate in the fermentation mash, thus resulting in increased alcohol yield in the final distillate. Alcohol yield has been increased by over 10% using this method.

Potential Future Applications

Research on nisin as a food additive continues as the demand increases for convenient long shelf-life and safe food preserved by safe preservatives with a natural connotation. Current research includes the use of nisin in combination with novel food preservation systems such as ultra high pressure, electroporation, pulsed electric fields, and nanothermosonication. It is evident that, like the classical use of heat preservation processes, bacterial spores are more resistant to these novel processes than bacterial vegetative cells and nisin in combination with these processes is synergistic against bacterial cells and surviving bacterial spores. Nisin in combination with other safe food additives that provide synergistic effects against Gram-positive bacteria or widen the antimicrobial spectrum to include Gram-negative bacteria, yeasts, and molds are also objectives of many research teams. Incorporation of nisin into packaging and edible films may also provide new areas of application. Finally, emerging food spoilage bacteria such as *Alicyclobacillus acidoterrestris* (an acid-tolerant, spore-forming bacteria that can grow at pH 2.5 and is a potential problem in pasteurized fruit juice) and *B. sporothermodurans* (a mesophilic, spore-forming bacteria whose spores can survive ultra heat-treated (UHT) processing) have been shown to be nisin-sensitive. New applications to control such bacteria are currently being realized.

See also: **Alcohol**: Properties and Determination; **Canning**: Principles; **Cheeses**: Chemistry and Microbiology of Maturation; Processed Cheese; ***Clostridium***: Occurrence of *Clostridium botulinum*; **Dressings and Mayonnaise**: The Products and Their Manufacture; Chemistry of the Products; **Eggs**: Use in the Food Industry; **Fish**: Processing; ***Listeria***: Properties and Occurrence; **Meat**: Sausages and Comminuted Products; **Milk**: Processing of Liquid Milk; **Preservation of Food**; **Shellfish**: Contamination and Spoilage of Molluscs and Crustaceans; **Spoilage**: Bacterial Spoilage

Further Reading

Delves-Broughton J (1990) Nisin and its uses as a food preservative. *Food Technology* 44: 100–117.

Delves-Broughton J and Gasson MJ (1994) Nisin. In: Dillon VM, Board RG (eds) *Natural Antimicrobial Systems and Food Preservation*, p. 99. Wallingford: CAB International.

De Vuyst L and Vandamme EJ (1994) Nisin, a lantibiotic produced by *Lactococcus lactis* subsp. *lactis:* properties, biosynthesis, fermentation and applications. In: De Vuyst, Vandamme EJ (eds) *Bacteriocins of Lactic A Bacteria: Microbiology, Genetics and Applications*, p. 151. London: Blackie Academic & Professional.

Hurst A (1991) Nisin. *Advances in Applied Microbiology* 27: 85–123.

Hurst A and Hoover DG (1993) Nisin. In: Davidson PM, Branen AL (eds) *Antimicrobials in Foods*, p. 369. New York: Marcel Dekker.

Thomas LV, Clarkson MR and Delves-Broughton J (2000) Nisin. In: Naidu AS (ed) *Natural Food Antimicrobial Systems*. CRC Press, USA, pp. 463–524.

NITRATES AND NITRITES

M J Dennis and L A Wilson, Central Science Laboratory, Sand Hutton, York, UK

Background

Nitrate is a normal component of plant tissues, whereas nitrite is not normally found in significant quantities unless microbiological spoilage occurs. Both nitrate and nitrite are permitted food additives for use in curing meat. Their original function was to confer microbiological safety, but they were also found to contribute useful color and flavor characteristics, which, with the advent of refrigeration, are now of greater technological significance. There have been considerable concerns over the safety of nitrate and nitrite in foods. Nitrate is largely unreactive but can be reduced to nitrite, which can then react with secondary amines to form nitrosamines (many of which are carcinogens). Equally, it is becoming apparent that oxides of nitrogen play an important role in human physiology. Hence, in this review, we will cover the occurrence of nitrate and nitrite in food, the legislation governing their use as food additives, and legislation governing their occurrence in vegetables, their human dietary intake, their potential health benefits, and concerns.

Occurrence in Foods

Nitrate is found as a naturally occurring compound in foods such as vegetables, fruit, cereals, fish, milk, and dairy products, and is also found in water as a consequence of agricultural practices such as the use of nitrogen-containing fertilizers and from animal waste. Low levels are generally found from these sources, except in the case of some vegetables. Nitrate and nitrite are also permitted as food additives in some foods, primarily as protection against botulism.

Vegetables and Fruit

Nitrate in the soil is taken up by plants for use as a nitrogen source in the formation of proteins. Protein production occurs as a result of photosynthesis, but when light levels fall, the rate of photosynthesis decreases, and nitrate accumulates in cell fluids and sap. The levels of nitrate in vegetables grown under low light conditions are thus correspondingly higher than those grown under bright light, as shown in Figure 1. Overall, nitrate accumulation in plants is determined by genotype, growing conditions, especially light levels and soil temperature, and nitrogen fertilization.

Nitrate content varies considerably according to species, with vegetables such as spinach and lettuce often containing up to 2500 mg kg^{-1}, whereas those such as asparagus have levels as low as 13 mg kg^{-1}. Table 1 lists the concentrations of nitrate found in some vegetables and fruits, and compares the levels found in different countries. The natural levels of nitrate in vegetables generally are high when compared with other food groups. It is estimated that 75–80% of the total daily intake comes from vegetables, compared with only about 5–10% from drinking water. Cooking has been shown to decrease the concentration of nitrate in foods, dependent on the cooking technique used (see Table 2).

Nitrite levels in vegetables and fruit are low, usually below 2 mg kg^{-1}, except where there has been damage or improper storage leading to the microbiological reduction of nitrate to nitrite.

Meat and Meat Products

Fresh meat normally contains low levels of nitrate and nitrite, estimated at < 4–7 mg kg^{-1} and < 0.4–0.5 mg kg^{-1}, respectively, in the UK in 1997, although higher concentrations of about 10–30 mg kg^{-1} of nitrate are found in cured meats such as ham and salami, where nitrite and nitrate have been incorporated as a permitted additive. Nitrite salts have

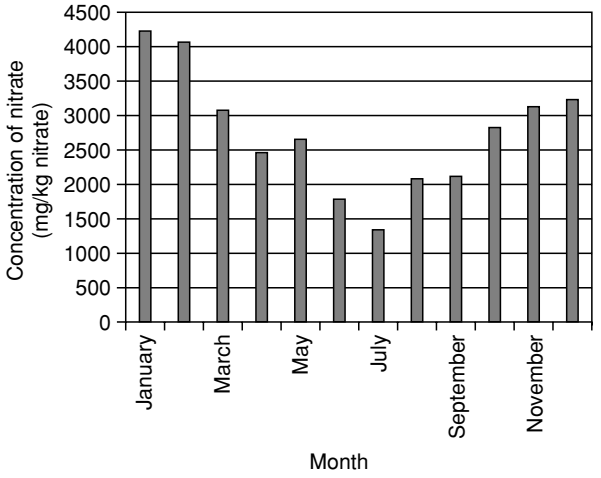

Figure 1 Seasonal variation in nitrate concentrations of lettuce (mg per kilogram of nitrate) (Denmark). From Petersen A and Stoltz S (1999) Nitrate and nitrite in vegetables on the Danish market: content and intake. *Food Additives and Contaminants* 16(7): 291–299, with permission.

Table 1 Comparison of average nitrate levels determined in vegetables and fruits (mg per kilogram of nitrate in fresh product) by different countries

Vegetable/fruit	*Australia*	*Denmark*[a]	*Italy*	*UK*	*USA*
Artichoke					16
Asparagus			13		60
Aubergine					370
Bean	265		450		466
Beetroot	2124	1490		1211	3288
Broccoli	310		400		1014
Brussels sprout	44		7–12	59	164
Cabbage	1062	342	240	338	712
Carrot	66		170–210	97	274
Cauliflower	221		37	86	658
Celery	1305				3151
Cucumber					151
Endive					1780
Kale					1096
Leek		308			700
Lettuce	943	2600		1051	2330
Melon					4932
Mushroom					219
Onion	22		80	48	235
Parsley	973				1380
Pea	66		57		40
Pepper (sweet)	88		78		165
Potato (sweet)	44				65
Potato	177	432	110	155	150
Pumpkin	177				550
Radish	1735				2600
Spinach (fresh)		1783	2100	1631	2470
Spinach (frozen)		680			
Tomato	44			17	80
Turnip (greens)					9040
Turnip (root)			970		535

[a]Average for 1993, 1994 and 1995/1996.

Data from Lyons DJ, Rayment GE, Nobbs PE and McCallum LE (1994) Nitrate and nitrite in fresh vegetables from Queensland. *Journal of the Science of Food and Agriculture* 64: 279–281; Petersen A and Stoltze S (1999) Nitrate and nitrite in vegetables on Danish market: content and intake. *Food Additives and Contaminants* 16(7): 291–299; Santamaria P (1997) Contributo degli ortaggi all'assunzione giornaliera di nitrato, nitrito e nitrosamine. *Industrie Alimentari* XXXVI (Novembre): 1329–1333; Ysort G, Miller P, Barrett G, Farrington D, Lawrance P and Harrison N (1999) Dietary exposures to nitrate in the UK. *Food Additives and Contaminants* 16(12): 521–532; Walker R (1990) Nitrates, nitrites and N-nitrosocompounds. A review of the occurrence in food and diet and the toxicological implications. *Food Additives and Contaminants* 7(6): 717–768. All data have been used with the author's permission.

Table 2 Effects of cooking on nitrate concentrations in vegetables (MAFF)

Vegetable	*Cooking method*	*Mean nitrate concentration (mg per kilogram of nitrate)*	*Percentage change*
Beetroot	Fresh	1211	
	Boiled	906	−25
Cabbage	Fresh	338	
	Boiled	114	−66
Carrot	Fresh	97	
	Boiled	71	−27
Cauliflower	Fresh	86	
	Boiled	46	−47
Onion	Fresh	48	
	Fried	44	−8
Potato	Fresh	167	
	Boiled	85	−49
	Fried	136	−19
	Baked	194	+16
Spinach	Fresh	1631	
	Boiled	468	−71
Sprout	Fresh	59	
	Boiled	24	−59
Swede	Fresh	118	
	Boiled	72	−39
Tomato	Fresh	17	
	Fried	27	+59

Data from: Ysart G, Miller P, Barrett G, Farrington D, Lawrance P and Harrison N (1999) Dietary exposures to nitrate in the UK. *Food Additives and Contaminants* 16(12): 521–532. All data used with the author's permission.

customarily been added to a range of meats for preservation purposes, especially as an antimicrobial agent with regard to *Clostridium botulinum*, to produce the pink color of cured meats via the formation of a nitric oxide–myoglobin complex and to give the traditional flavor expected from these products. Nitrite is preferred as a curing agent, as it reacts more quickly than nitrate, and less is required for color stabilization, but nitrate may be added to act as a reservoir in case the nitrite level is depleted during curing. Measures have been taken to reduce the amounts of nitrite used during meat curing processes owing to concern over the formation of nitrosamines. These restrictions have led to significant reductions in the nitrate and nitrite contents of cured meats. Nitrite is now frequently added to meat products at much lower levels, commonly 80–160 mg kg^{-1}, in conjunction with sodium ascorbate or sodium erythorbate. It is considered that the benefits of adding nitrite to products of this type far outweigh the potential risks. (*See* ***Clostridium***: Botulism.)

Cheese

Potassium nitrate is permitted as an additive in cheese manufacture in order to control microbiological contamination, which may produce early gassing, and thus to improve the quality of the finished product. Natural concentrations in cheese were found at 1–8 mg kg^{-1} of nitrate, with concentrations rising to 4–27 mg kg^{-1} of nitrate and < 0.6–1 mg kg^{-1} of nitrite when nitrate was used as an additive.

Fish and Fish Products

Nitrite is permitted in some countries as an additive to some smoked and cured fish and fish products, when it is used as a preservative and color fixative. Levels of both nitrate and nitrite are low in the fresh product, but a recent Japanese survey (1998) analyzed nitrite levels in different salted and processed

fish products and found that, although the mean concentrations were about 30–40% of the permitted limit, some products contained nitrite at levels of up to 32 mg kg^{-1}.

Beverages

Beverages make a significant contribution to nitrite intake due to high consumption. Although the concentrations found are low (< 0.4–0.8 mg l^{-1}), the mean nitrite intake in the UK from this source is 0.34 mg per person per day.

Water

Nitrate is ubiquitous in potable water. Its presence is due to a number of factors: significant sources are agricultural activity, for example, the use of nitrogenous fertilizers; sewage effluents and urban run-off and the geology of the area. It is formed when nitrogenous organic sources, such as urea and proteins, or fertilizers containing inorganic nitrogen are decomposed by microorganisms in water or soil to form ammonia, which is then oxidized to nitrate and nitrite. Up to 5 mg l^{-1} of nitrate have been found in rainwater in European urban areas, although levels are somewhat lower in rural areas. However, levels rise as a result of human activities, and groundwater with concentrations of up to 1500 mg l^{-1} have been found in an agricultural area in India. Nitrate is very soluble in water and easily moves through the soil into aquifers and into the drinking water supply. Nitrite is readily oxidized to nitrate and is consequently found in water at much lower levels, especially after water treatment involving chlorination. Water obtained from public water companies tends to contain lower levels of nitrate, usually less than 10 mg l^{-1} in most countries, but private wells throughout the world, especially in agricultural areas, may have concentrations often exceeding 50 mg l^{-1}. However, the exposure to water from the latter source is less common; for example, it is estimated that only about 2% of the European population draw their water from private wells.

Legislation

Maximum permitted levels of nitrate in potable water have been set at levels intended to prevent the occurrence of methaemoglobinemia in infants and other susceptible sections of the population. The second edition of the WHO Guidelines for drinking-water quality recommended a guideline value of 50 mg l^{-1} for nitrate and 3 mg l^{-1} for nitrite levels for water. The US Federal Government has set a similar maximum contaminant level of 45 mg l^{-1}.

Concern over the *in vivo* formation of nitrosamines from nitrites has led to a reduction in the levels permitted when they are used as food additives. In the UK, the current residual concentration of nitrite (as sodium nitrite) in cured bacon is limited to 175 mg kg^{-1} and 100 mg kg^{-1} for other cured meats. The maximum residual concentration of nitrate (as sodium nitrate) in all cured meats is 250 mg kg^{-1}. The use of potassium nitrate in cheese is limited to that giving a residual concentration of 50 mg kg^{-1}. Regulations in other countries are similar: Australia and New Zealand set a maximum of 125 mg kg^{-1} of sodium nitrite for cured meats and 50 mg kg^{-1} of total nitrate/nitrite in cheeses such as Gouda, Edam, and Havarti.

In Europe, Commission Regulation No. 194/97 sets the maximum levels for certain contaminants in foodstuffs, including nitrate in lettuce and spinach. By the use of specific measures intended to provide better control of the sources of nitrate and good agricultural practices, the levels in spinach should not exceed 2500 mg kg^{-1} (summer crop) and 3000 mg kg^{-1} (winter crop) and in lettuce, levels should not exceed 3500 mg kg^{-1} (summer crop except open-grown lettuce where the limit is 2500 mg kg^{-1} from May to September) and 4500 mg kg^{-1} (winter crop).

Dietary Intake of Nitrate and Nitrite

The acceptable daily intake (ADI) is defined as the maximum amount of a chemical that can be ingested daily over a lifetime with no appreciable health risk, and is based on the highest intake that does not give rise to observable adverse effects. The European Commission's Scientific Committee for Food set the ADI for nitrate at 0–3.65 mg per kilogram of bodyweight (equivalent to an intake of 219 mg day^{-1} for a 60-kg person) and 0–0.06 mg per kilogram of bodyweight (equivalent to an intake of 3.6 mg day^{-1} for a 60-kg person) for nitrite. These ADIs are similar to those set by the joint FAO/WHO experts committee. The ADIs apply to all sources of intake, although the use of nitrite as an additive in baby food intended for infants below the age of 3 months is not permitted. The nitrite ADI does not apply to infants under 3 months.

Dietary exposure to food components may be assessed by the use of several approaches, including the use of market basket studies, food diary records, *per capita* estimations using consumption data with the concentrations of nitrate and nitrite in foods, and urinary biomarkers. Intake studies have shown that the ADIs are not exceeded by the general population across the world, as shown in Table 3. The question of whether some subgroups of the population are consuming high levels of nitrate was addressed by the UK's Food Standards Agency, who carried out a study into the nitrate intake of vegetarians. This

Table 3 Comparison of daily dietary exposures to nitrate in different countries

Country	*Dietary exposure (mg day^{-1})*
UK	52
Finland	77
The Netherlands	52
The Netherlands	78–91
Basque Country, Spain	60
Belgium	154 (vegetables)
Egypt	296 (nitrate and nitrite)
USA	300
Poland	85 (duplicate hospital diets)
Poland	65 (estimate from hospital diets)
India	78
EU	18–131 (vegetables)

From Ysart G, Miller P, Barrett G *et al*. (1999) Dietary exposures to nitrate in the UK. *Food Additives and Contaminants* 16(12): 521–532.

showed that, although there was a slightly higher intake, dietary exposure was similar to other consumers and was well below the ADI.

In summary, food is the main source of nitrate, especially potatoes, green vegetables, and other vegetables. Water also makes a significant contribution. Estimates of nitrite intake suggest that 1–3 mg day^{-1} is typical and occurs mainly from beverages, due to high consumption.

Microbiological Nitrate Reduction

A detailed investigation of the reduction of nitrate to nitrite by microbiological contaminants of the brewing process has been undertaken. Concerns over the presence of nitrosamines in beer have led to detailed studies of the mechanism of formation of these compounds in order to prevent their formation as far as technologically feasible. Nitrosodimethylamine was formed in malt from nitrogen oxides present in kiln gases. Concentrations of this contaminant were greatly reduced by modifications to kilning conditions. Apparent total *N*-nitroso compounds (ATNC) provide an estimate of the total *N*-nitroso compounds present in a sample. The brewing industry sought to minimize ATNC and set a target of 20 μg (N-NO) kg^{-1}. It was established that the major source of nitrosating agent was from microbial reduction of nitrate. Hence, care in ensuring that microbiological contamination was minimized played an important role in meeting this target.

Microorganisms can be differentiated into assimilatory and dissimilatory nitrate reducers. The former reduce nitrate to nitrite and ammonia for use as their source of cell nitrogen. The latter use nitrate as a terminal electron acceptor for energy production (oxygen is the usual terminal electron acceptor under aerobic growth conditions). Nitrite, nitric oxide, nitrous oxide, and sometimes nitrogen may all be produced by dissimilatory nitrate reduction. A number of different bacteria and yeasts are capable of nitrate reduction. The effect is not merely to produce nitrite that subsequently undergoes the well-known acid-catalyzed chemical nitrosation reactions. Rather, the nitrosation reactions appear to be catalyzed by the nitrate and nitrite reductase enzymes. This mechanism enables nitrosation to occur at pH 6–9 which cannot happen in the chemical reaction at alkaline pH.

However, the significance of microbiological reduction of nitrate is wider than that caused by the microbiological contamination of food. Nitrate-reducing bacteria can be found in the mouth and the colon. Nitrate from the diet (or that produced metabolically) is transferred from the blood to the saliva in the salivary glands. It has been suggested that 25% of orally ingested nitrate may be secreted in saliva. Twenty per cent of this nitrate can then be reduced by oral bacteria to nitrite, which is swallowed. Hence, it is apparent that saliva makes the major contribution to dietary nitrite intake. Saliva nitrite levels rise at night and are higher in older people, especially men. Salivary nitrate and nitrite vary little from day to day, but vary more over a 5-year period.

Bacterial nitrosation may occur at any site in the body where there is infection, and these may be linked to cancers at these sites. However, the colon provides a site rich in microorganisms at all times. It is apparent from measurements of fecal ATNC that bacterial nitrosation reactions occur at this site. ATNC formation is influenced by the nature of the organisms present, the transit time of the fecal mass through the colon and to consumption of nitrate and (surprisingly) red meat. Since dietary nitrate is absorbed by the small intestine, it is difficult to appreciate how the nitrosation reaction may occur. It is possible that there may be diffusion of nitrate from the colonic epithelium. Alternatively, red meat might influence fecal ATNC by carrying nitric oxide into the colon as nitrosyl myoglobin, by favoring the growth of nitrate-reducing bacteria in the colon, or possibly by nitrosyl myoglobin acting as a nitrosating agent itself. These possibilities require further study.

Concerns over Nitrate and Health

A considerable body of evidence has been collected concerning the direct effects on health of the nitrate and nitrite ion. There is no evidence of carcinogenicity from nitrate consumption in experimental animals or through human epidemiological studies. Long-term animal studies on nitrite do not indicate carcinogenicity unless combined with amine intake

leading to the formation of carcinogenic nitrosamines. The latter studies usually involved nitrite concentrations far higher than those commonly encountered in the diet. Nitrite has demonstrated mutagenic activity in *in vitro* assays. There is considerable evidence of the safety of substantially higher exposures to nitrite during clinical use than those normally encountered from the diet.

The most important direct toxic effect of nitrite (or nitrate after its reduction) is the condition known as methemoglobinemia. Absorbed nitrite ion is rapidly oxidized to nitrate by reaction with oxyhemoglobin to produce methemoglobin. The latter can not carry oxygen around the body, and although the reaction can be reversed enzymatically, if large quantities of methemoglobin are produced, fatality can result. Infants are deficient in this repair enzyme and hence are especially at risk if they receive unusually high intakes of nitrate, for example from contaminated well water. This may be exacerbated by concurrent infections resulting in endogenous nitrate synthesis.

Another concern is that consumption of nitrite (or nitrate reduced by oral bacteria to nitrite) may result in the nitrosation of amines to produce carcinogenic nitrosamines in the acid conditions of the stomach. Such an effect has been found in a human study after consuming a meal consisting of vegetables high in nitrate and fish (to provide a source of amines). Similar studies in animals have been shown to induce cancer. Such nitrosation reactions may be enhanced by bacterial nitrosation in the achlorhydric stomach.

Volatile nitrosamines are not normally found in urine (unless a urinary infection is present) or in feces. However, feces do contain ATNC, although the toxicological significance of this measurement is not understood. The formation of ATNC appears to require an active microbiological flora in the colon capable of nitrate reduction. Hence, ATNC would appear to be, at least, a biomarker of nitrosation reactions occurring in the colon. It is possible for the colonic bacteria to make a more subtle contribution to colon cancer than just the production of potential carcinogens. For example, nitrate-reducing enteric fermentative bacteria lead to the production of more acetate and less butyrate from substrate. Butyrate is considered to have a protective effect against cancer, since it promotes apoptosis (programmed cell death), so any selection for bacteria producing acetate may favor colonic epithelial cell proliferation.

There are a number of medical conditions other than cancer that have been associated with nitrate/nitrite intake. In Finland, a significant correlation has been found between the occurrence of type 1 diabetes and the nitrite intake of both the children and their mothers. No association was found for nitrate intake. Since most nitrite intake arises from saliva, it seems likely that the association with nitrite is in fact a proxy for preformed dietary nitroso compounds, some of which (e.g., *N*-nitrosomethylurea) have well established toxic effects on pancreatic beta cells. Studies in Iceland and Sweden have also found similar correlations.

An association has been observed between maternal periconceptional exposure to nitrate and an increased risk of anencephalopathy (but not spina bifida). The lack of a quantitative association between nitrate intake and risk may indicate that something other than nitrate is the causative factor. It should be noted that not all studies have been able to demonstrate such a link, but nitrosocompounds have been shown to cause defects in the central nervous system of experimental animals.

Nitrate acts as a competitive inhibitor of iodine uptake by the thyroid and thyroid hypertrophy has been demonstrated experimentally in rats given drinking water containing large amounts of nitrate. There is some epidemiological evidence that populations exposed to high nitrate concentrations in drinking water show an average increase in thyroid volume, although it cannot be excluded that the effects are due to other compounds.

Potential Health Benefits of Nitrate

Nitric oxide plays a crucial role in human metabolism. It is synthesized from arginine by two related enzyme systems – one constitutive and one inducible. The constitutive form provides continuous vasodilation and inhibits platelet adhesion and aggregation; hence, it can help to prevent hypertension. The inducible isozyme is generated in response to bacterial lipopolysaccharide and hence is considered to play an important role in prevention of bacterial disease. However, NO production is also greatly increased in inflammatory bowel disease (e.g., Crohn's disease). NO synthesis also occurs in the central nervous system, but its physiological role in this tissue remains unexplained. These mechanisms for nitric oxide generation mean that man excretes more nitrate than is ingested.

Dietary nitrate is readily absorbed, transported in plasma, and concentrated 10-fold in the saliva by the salivary gland. Although nitrate is excreted in the urine, the kidney recovers some 80% of the nitrate. Hence, this renal salvage and salivary concentration suggest that nitrate is physiologically important. Nitric oxide has widespread antimicrobial activity, and hence, it is hypothesized that the salivary concentration of nitrate and its microbial conversion to nitrite are a symbiotic activity to generate quantities

of nitric oxide in the stomach for protection against pathogens. Inhibition of gastric NO synthesis does not have clinical relevance, but this could be due to the high microbiological standards of contemporary diets. Nitrate is also expressed in sweat where significant amounts of nitrite and bacterially generated nitric oxide can also be found and where a protective effect against pathogens has also been postulated.

Gastric juices provide a rich source of nitrosatable substrates, which, some consider, far exceeds those provided by food. Hence, since nitrate can be endogenously formed, carried in the blood, and excreted in saliva, it would appear unavoidable that some gastric nitrosation takes place. However, much more gastric nitric oxide is produced than might be expected from acid-catalyzed dissociation of nitrite to nitric oxide and nitrogen dioxide. Thus, it would appear that another reductant (e.g., ascorbic acid) is produced in gastric juice in order to facilitate the generation of NO. Therefore, it may be that man is metabolically adapted to the possibility of gastric nitrosation. This view is supported by the failure of a number of epidemiological studies to demonstrate a link between nitrate intake and cancer in man.

See also: **Curing**; ***Escherichia coli***: Occurrence and Epidemiology of Species other than *Escherichia coli*; **Legislation**: Contaminants and Adulterants; **Meat**: Preservation; **Nitrosamines**; **Smoked Foods**: Principles; Applications of Smoking; **Water Supplies**: Chemical Analysis

Further Reading

Benjamin N (2000) Nitrates in the human diet – good or bad? *Annales De Zootechnie* 49: 207–216.

Commission of the European Communities Commission Regulation (EC) No. 466/2001 (2001) Setting maximum levels for certain contaminants in foodstuffs. *Official Journal of the European Communities* No. L77: 1–13.

Croen LA, Todoroff K and Shaw GM (2001) Maternal exposure to nitrate from drinking water and diet and risk neural tube defects. *American Journal of Epidemiology* 153(4): 325–331.

Hill MJ (1996) Factors controlling endogenous N-nitrosation. *European Journal of Cancer Prevention* 5(supplement 1): 71–74.

Hughes R, Cross AJ, Pollock JRA and Bingham S (2001) Dose-dependent effect of dietary meat on endogenous colonic N-nitrosation. *Carcinogenesis* 22(1): 199–202.

Hunt J and Turner MK (1994) A survey of nitrite concentrations in retail fresh vegetables. *Food Additives and Contaminants* 11(3): 327–332.

Ishiwata H, Nishijima M, Fukasawa Y, Ito Y and Yamada T (1998) Evaluation of the inorganic food additive (nitrite, nitrate and sulfur dioxide) content of foods and estimation of daily intake based on the results of Official Inspection in Japan in the fiscal year 1994. *Journal of the Food Hygiene Society of Japan* 39(2): 78–88.

Lyons DJ, Rayment GE, Nobbs PE and McCallum LE (1994) Nitrate and nitrite in fresh vegetables from Queensland. *Journal of the Science of Food and Agriculture* 64: 279–281.

Ministry of Agriculture, Fisheries and Food (1998a) MAFF UK – Survey of nitrite and nitrate in bacon and cured meat products. *Food Surveillance Information Sheet No. 142*.

Ministry of Agriculture, Fisheries and Food (1998b) MAFF UK – 1997 Total Diet Study – Nitrate and nitrite. *Food Surveillance Information Sheet No. 163*.

Ministry of Agriculture, Fisheries and Food (1998c) MAFF UK – Duplicate diet study of vegetarians – nitrate analysis. *Food Surveillance Information Sheet No. 165*.

Mirvish SS, Reimers KJ, Kutler B *et al.* (2000) Nitrate and nitrite concentrations in human saliva for men and women at different ages and times of the day and their consistency over time. *European Journal of Cancer Prevention* 9: 335–342.

Parham NJ and Gibson GR (2000) Microbes involved in dissimilatory nitrate reduction in the human large intestine. *FEMS Microbiology Ecology* 31: 21–28.

Perner A and Rask-Madsen J (1999) Review article: the potential role of nitric oxide in chronic inflammatory bowel disorders. *Alimentary Pharmacology and Therapeutics* 13: 135–144.

Petersen A and Stoltze S (1999) Nitrate and nitrite in vegetables on the Danish market: content and intake. *Food Additives and Contaminants* 16(7): 291–299.

Rowland IR, Granli T, Bockman OC, Key PE and Massey RC (1991) Endogenous N-nitrosation in man assessed by measurement of apparent total N-nitroso compounds in faeces. *Carcinogenesis* 12(8): 1395–1401.

Santamaria P (1997) Contributo degli ortaggi all'assunzione giornaliera di nitrato, nitrito e nitrosamine. *Industrie Alimentari* XXXVI (Novembre): 1329–1334.

Smith NA (1994) Cambridge Prize Lecture: Nitrate reduction and N-nitrosation in brewing. *Journal of the Institute of Brewing* 100: 347–355.

van Maanen JMS, van Dijk A, Mulder K *et al.* (1994) Consumption of drinking water with high nitrate levels causes hypertrophy of the thyroid. *Toxicology Letters* 72: 365–374.

Virtanen SM, Jaakkola L, Rasanen L *et al.* (1994) Nitrate and nitrite intake and the risk for type 1 diabetes in Finnish children. *Diabetic Medicine* 11: 656–662.

Walker R (1990) Nitrates, nitrites and N-nitrosocompounds: a review of the occurrence in food and diet and the toxicological implications. *Food Additives and Contaminants* 7(6): 717–768.

World Health Organization (1998) Guidelines for drinking water quality, 2nd edn. *Addendum to vol. 2 Health criteria and other supporting information*, pp. 64–80.

Ysart G, Miller P, Barrett G, Farrington D, Lawrance P and Harrison N (1999) Dietary exposures to nitrate in the UK. *Food Additives and Contaminants* 16(12): 521–532.

NITROSAMINES

R A Scanlan, Oregon State University, Corvallis, OR, USA

Introduction

Investigations during the past 30 years on the occurrence of *N*-nitrosamines and the etiology of these compounds in human cancer began with two important discoveries. In 1956, the UK scientists David Barnes and Peter Magee reported *N*-nitrosodimethylamine (NDMA) as a carcinogen in experimental animals. In Norway in the early 1960s, NDMA proved to be the causative agent in the death of agricultural animals which had been fed nitrite-preserved fish meal. The nitrite had reacted with amines in the fish meal to produce lethal amounts of NDMA. Since nitrite is added to human foods, particularly cured meats, and amines occur commonly in most foods, this raised the question as to whether carcinogenic nitrosamines might also form in human foods. This article summarizes the results of the many investigations which have addressed this important issue.

Chemistry of *N*-Nitrosamine Formation

Basic Reactions

N-Nitrosamines are formed by a chemical reaction between a nitrosating agent and a secondary or a tertiary amine. Primary amines react with nitrosating agents to form unstable *N*-nitroso derivatives which degrade to olefins and alcohols. The structures of several nitrosamines which occur in foods are shown in Figure 1.

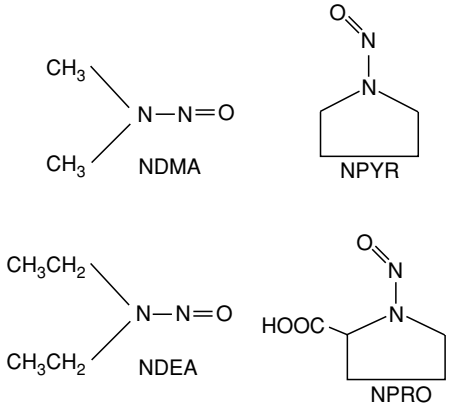

Figure 1 Structure of some nitrosamines which occur in foods. NDMA, *N*-nitrosodimethylamine; NPYR, *N*-nitrosopyrrolidine; NDEA, *N*-nitrosodiethylamine; NPRO, *N*-nitrosoproline.

Oxides of nitrogen (NO_x) in which the nitrogen is in a $+3$ or $+4$ oxidation state can serve as nitrosating agents. A much studied nitrosating agent which participates in nitrosamine formation in foods is nitrous anhydride (N_2O_3). Nitrous anhydride forms readily from nitrite in aqueous acidic solution, as shown in eqn (1).

$$2NO_2^- + 2H^+ \rightleftharpoons N_2O_3 + H_2O \qquad (1)$$

Nitrous anhydride combines with the unshared pair of electrons on unpronated secondary amines through a nucleophilic substitution reaction to form *N*-nitrosamines, as depicted in eqn (2).

$$R_2NH + N_2O_3 \rightarrow R_2N - N = O + HNO_2 \qquad (2)$$

The rate of nitrosation is pH-dependent and is governed by the concentrations of amine and nitrite, as shown in eqn (3):

$$\text{Rate} = k[\text{amine}] \times [\text{nitrite}]^2 \qquad (3)$$

Since the nitrous anhydride reacts with unprotonated amine, the rate of nitrosation for secondary amines is inversely proportional to amine basicity. The pH optimum for nitrosation of most secondary amines is between 2.5 and 3.5. This is due to the counteracting effects of acidity on nitrous anhydride concentration, which increases at low pH, and the concentration of unprotonated amine, which increases at high pH.

The pH optimum has a number of implications. Although most foods are less acidic than pH 2.5–3.5, many foods are sufficiently acidic to allow nitrosation, albeit at rates slower than maximal. Furthermore, the 2.5–3.5 pH range is sufficiently close to the acidity of the human stomach to allow nitrosamine formation, providing that amines and nitrosating agents are present.

Tertiary amines react with nitrosating agents in acidic aqueous solution through a mechanism called nitrosative dealkylation to form nitrosamines. Nitrosative dealkylation involves conversion of a tertiary amine to a secondary amine, which subsequently reacts with nitrosating agent to form a nitrosamine.

Nitrosation Inhibitors

Nitrosation can be influenced by a wide range of catalysts and inhibitors. Ascorbic acid, α-tocopheral, and sulfur dioxide are used to inhibit nitrosamine formation in foods. Ascorbic acid reduces nitrous anhydride to nitric oxide, as shown in eqn (4).

$$\text{Ascorbic acid} + N_2O_3 \rightarrow \text{Dehydroascorbic acid} + 2NO + H_2O \quad (4)$$

Ascorbic acid is not completely effective in blocking nitrosation, since under oxidative conditions the nitric oxide can be reconverted to nitrous anhydride. Presumably α-tocopheral reacts with nitrous anhydride through a similar redox process to inhibit nitrosation. Both ascorbic acid and α-tocopheral are added to cured meats, and sulfur dioxide is used in processing barley malt to inhibit nitrosation.

Carcinogenicity

Although *N*-nitroso compounds have been shown to be acutely toxic, mutagenic, and teratogenic, their carcinogenic properties are of grave concern and have been studied extensively. Over 300 *N*-nitroso compounds, including many nitrosamines, have been demonstrated to be carcinogens in experimental animals. More than 50 animal species, including higher primates, are susceptible to *N*-nitroso compound-induced carcinogenesis. It has been observed that tumors induced by nitrosamines in experimental animals show similar morphological properties to tumors in corresponding human organs. Few scientists doubt that nitrosamines are capable of inducing cancer in humans. However, the amount of a given nitrosamine which is required to induce cancer in a human population is unknown. (*See* **Cancer**: Carcinogens in the Food Chain.)

Analytical Methodology

For purposes of analysis, nitrosamines are classified as either volatile or nonvolatile. Volatile nitrosamines are relatively nonpolar, low-molecular-weight compounds, such as NDMA and *N*-nitrosopyrrolidine (NPYR). They possess sufficient vapor pressure to allow removal from a food matrix by distillation and subsequent separation by gas chromatography. Chemiluminescent detectors designed to be relatively specific for nitrosamines allow quantitative analysis of volatile nitrosamines in foods and other biological materials at submicrogram per kilogram levels. Confirmation of nitrosamine identity is accomplished by mass spectrometry. (*See* **Chromatography**: Gas Chromatography; **Mass Spectrometry**: Principles and Instrumentation.)

Nonvolatile nitrosamines such as *N*-nitrosoproline (NPRO) tend to be of higher molecular weight, or more polar than volatile nitrosamines, and hence possess relatively low vapor pressures. These properties have impeded development of analytical methodology. Although procedures are available for determination of some nonvolatile nitrosamines, mainly *N*-nitrosated amino acids and amino acid derivatives, analytical methodology generally is less well developed for this class of compounds.

A procedure usually referred to as apparent total *N*-nitroso compound determination has been developed and used for analyses of foods and body fluids. Presumably when using this procedure all *N*-nitroso compounds in the sample are measured as a single entity. A limitation is that information regarding the identity of individual *N*-nitroso compounds is not provided by this methology. It is therefore impossible to access the carcinogenicity of compounds detected by this procedure.

Formation and Occurrence

Exogenous Formation

Foods During the past 25 years many foods have been analyzed for volatile nitrosamines. In general, foods in most western diets have been examined more extensively than foods from Asia, Africa, and South America. Although approximately 20 volatile nitrosamines have been identified in a variety of foods and beverages, NDMA and NPYR have been found most commonly. Most of the other volatile nitrosamines which occur in foods do so infrequently. Table 1 presents a condensation of many reports of the occurrence of volatile nitrosamines in foods from around the world. The summary is not comprehensive; rather, it is intended to show the foods in which nitrosamines occur most commonly.

Nitrosamines form in foods because under certain circumstances the precursors, amines and nitrosating agents, occur in foods. Amines occur commonly in food, and are formed biosynthetically and by

Table 1 Volatile nitrosamines which occur most commonly in foods

Food	*Nitrosamine*	*Range*[a] *($\mu g\,kg^{-1}$)*	*Occurrence*
Fried bacon	NDMA, NPYR	1–100	Consistent
Cured meats	NDMA, NPYR, NPIP	ND–50	Sporadic
Beer	NDMA	ND–5	Sporadic
Cheese	NDMA	ND–5	Sporadic
Cooked fish	NDMA, NDEA	ND–50	Sporadic
Salt-dried fish	NDMA, NDEA	ND–1000	
Nonfat dry milk	NDMA	ND–1	Consistent in direct-fire dried product

[a]The range is intended to encompass data in most reports since 1980. In general, most positive samples contained nitrosamines at the lower end of the range. ND, not detectable.
NDMA, *N*-nitrosodimethylamine; NPYR, *N*-nitrosopyrrolidine; NDEA, *N*-nitrosodiethylamine; NPIP, *N*-nitrosopiperidine.

microbial activity. Nitrosating agents can be formed from certain compounds added to foods and as a result of specific processing conditions. In most instances, nitrosamines are found in foods in one of the three following categories.

Cured meats Nitrosamines are formed in cured meats because nitrite, and sometimes nitrate, are added to these products during processing. Nitrate is reduced to nitrite by the enzyme nitrate reductase, which occurs in a number of bacteria. As discussed earlier, nitrite is converted to nitrosating agents which subsequently react with amines in the meat during processing, storage, and cooking to form nitrosamines. (*See* **Curing**.)

Nitrite and nitrate have been added to cured meats for many years to prevent outgrowth and toxin formation by *Clostridium botulinum*. Nitrite, in combination with other curing ingredients such as sodium chloride, is particularly effective in inhibiting formation of the deadly botulism toxin. In addition, nitrite reacts with pigments in meat to impart the desirable pink color of cured meats and it prevents the development of off-flavors. (*See* **Nitrates and Nitrites**.)

Although not all cured meats have detectable amounts of nitrosamines, fried bacon has been shown consistently to contain these compounds. Typically, fried bacon contains 1–20 $\mu g\,kg^{-1}$ of NPYR and 1–3 $\mu g\,kg^{-1}$ of NDMA. The formation is related to the relatively high internal temperature of bacon during frying and the relatively low moisture content of bacon as compared to other cured meat products. When bacon is cooked by other methods, particularly in a microwave oven, considerably lower amounts of nitrosamines are found. The majority of evidence suggests that the free amino acid proline is first nitrosated and then decarboxylated to form NPYR during frying. Neither the precise chemical nature of the nitrosating agent nor the amine precursor for NDMA in bacon is known with certainty. However, the evidence suggests that the nitrosating agent is a reaction product of nitrite and lipids in the bacon.

Dried foods and ingredients A variety of processes and equipment are used to dry foods and food ingredients. During the direct-fire process, air used to dry the food is first heated by passing the air through the flames of burners. As a result, products of combustion, including oxides of nitrogen, are directly incorporated into the hot air used to dry the food. The oxides of nitrogen which include nitrosating agents such as nitrous anhydride can then react with amines in the food being dried to produce nitrosamines.

In 1979, scientists in Europe reported the occurrence of NDMA in beer. Soon after, reports from a number of countries confirmed that most beers contained 1–5 $\mu g\,kg^{-1}$ of NDMA. Although beer was found to contain NDMA, the NDMA was not formed during the brewing process. Investigation led to the discovery that direct-fire-dried malted barley, an ingredient used in the manufacture of beer, was the source of the NDMA.

A number of investigators have attempted to identify the amine precursors of NDMA in malted barley. The evidence suggests that dimethylamine and perhaps the tertiary amine alkaloids gramine and hordenine serve as precursors for NDMA in malted barley.

The discovery of nitrosamine formation in direct-fire-dried malted barley has led to the investigation of other dried foods. In the USA, certain dried dairy products, notably nonfat dry milk, are manufactured by the direct-fire drying process. Nonfat dry milk manufactured by this process consistently contains low amounts (less than 1 $\mu g\,kg^{-1}$) of NDMA. Since the direct-fire drying process is not used commonly in Europe for manufacture of nonfat dry milk, the product does not contain NDMA.

Seafood can contain nitrosamines as a result of either cooking or salt-drying. When fish is broiled with a gas flame, the oxides of nitrogen produced in the flame can cause nitrosamine formation in a manner analogous to nitrosamine formation in direct-fire drying. In certain areas of the world, particularly the Orient, fish is preserved by salt. Often sea salt is used which contains appreciable amounts of nitrate. The nitrate is reduced to nitrite with subsequent formation of nitrosating agents. In some cases, reaction with amines in the fish produces relatively large amounts of nitrosamines, principally NDMA. The potential for NDMA formation in fish is considerable since high levels of NDMA precursors, such as dimethylamine and trimethylamine, can occur in fish.

Migration from surfaces which contact foods Vulcanized rubber products, such as baby nursing nipples, have been shown to contain nitrosamines. It has been demonstrated that when baby bottle nipples are stored inverted in milk, the nitrosamines partially migrate from the nipples to the milk. Furthermore, studies have shown that nitrosamines migrate from rubber netting used to hold cured meats during the smoking process. In addition to rubber, nitrosamines in such substances as wax-treated wrapping paper and paperboard-based materials have been shown to migrate to foods.

Nonvolatile nitrosamines in foods The notion that nonvolatile nitrosamines might form in foods follows logically from the fact that the precursors,

nonvolatile amine and nitrosating agents, occur in foods. Due to limitations in analytical methodology for nonvolatile nitrosamines, less is known about the occurrence of nonvolatile than volatile nitrosamines in foods. Recently developed methods for nonvolatile nitrosamines have been limited to the detection of *N*-nitrosated amino acids and amino acid derivatives. Most analyses have been on cured meats and include reports of compounds such as NPRO, *N*-nitrososarcosine (NSAR), *N*-nitroso-4-hydroxyproline, *N*-nitrosothiazolidine-4-carboxylic acid, *N*-nitroso-2-methylthiazolidine-4-carboxylic acid, and *N*-nitroso-2-hydroxylmethylthiazolidine-4-carboxylic acid. Other than NSAR, which is a weak carcinogen, all the other *N*-nitrosated amino acids and amino acid derivatives which have been tested have not shown a carcinogenic response in animals.

Scientists conjecture that other nonvolatile nitrosamines such as nitrosated peptides and nitrosated amides occur in foods. Definitive information in this regard awaits further application of analytical methodology for nonvolatile nitrosamines to foods.

Reduction of nitrosamine formation Considerable efforts have been expended during the past 25 years to reduce nitrosamine formation in foods. Scientists have looked critically at the use of nitrite and nitrate in the manufacture of cured meats. Since nitrite is considered essential to guard against outgrowth and toxin production by *C. botulinum*, its use in the manufacture of cured meats has been retained. However, in many countries, permissible levels of nitrite have been reduced to the minimum necessary for control of botulism. Generally, nitrate is only permitted in a few fermented cured meat products where long-term inhibition of *C. botulinum* is required.

The nitrosation inhibitors ascorbic acid and α-tocopheral are either required or extensively used in processing cured meats. Alternatives and/or partial substitutes for nitrite in cured meats include use of lactic acid-producing organisms, potassium sorbate, sodium hypophosphite, fumarate esters, and ionizing radiation. To date, no adequate alternative for nitrite has been found. Consequently, nitrosamines continue to be found in cured meats, especially fried bacon, but generally at lower levels than occurred a number of years ago.

Efforts to reduce nitrosamine levels in barley malt, and hence in beer, have been very successful. This is largely due to conversion of direct-fired kilns to indirect-fired kilns for the manufacture of barley malt. With indirect-fired kilns, the products of combustion are not incorporated into the drying air and, therefore, nitrosamine formation in the barley malt is greatly reduced. Interestingly, use of an indirect-fired kiln does not completely inhibit nitrosation, probably because ambient air which is drawn into the kiln usually contains trace levels of oxides of nitrogen and, therefore, some nitrosation occurs. In addition to the use of indirect-fired kilns, sulfur dioxide, which has been shown to be a nitrosation inhibitor, is sometimes used with both direct-fired kilns and indirect-fired kilns to reduce nitrosamine formation.

Due to these changes in processing, the NDMA levels in beer have been markedly reduced. In a recent survey, Canadian and US beers were found to contain on average $0.07\,\mu g\,kg^{-1}$ of NDMA whereas, prior to 1980, beer commonly contained $1–5\,\mu g\,kg^{-1}$ of NDMA.

Similarly, nitrosamine formation in rubber products, including baby nursing nipples, has been reduced. This has been accomplished by substituting nitrosatable with nonnitrosatable vulcanization accelerator compounds in the manufacture of rubber products.

Products other than food A variety of industrial, agricultural, and consumer items have been shown to contain nitrosamines. These include cosmetics, tobacco products, industrial cutting fluids, vehicle tires, and pesticides. In each case, the formation can be traced to use of amines and contact with nitrosating agents during product formulation, manufacture, or use.

Tobacco products are especially important in terms of nitrosamine occurrence and subsequent human exposure. During growth, tobacco plants biosynthesize a variety of alkaloids such as nicotine and nornicotine. These compounds are either secondary or tertiary amines. During curing, fermentation, and aging the alkaloids react with nitrosating agents formed from nitrate to produce a group of compounds commonly referred to as tobacco-specific nitrosamines. Tobacco-specific nitrosamines have been shown to occur in a wide range of tobacco products such as cigarettes, cigars, pipe tobacco, chewing tobacco, snuff, masheri, zarda, and nass. It is noteworthy that the tobacco-specific nitrosamines frequently occur in tobacco products at several orders of magnitude higher than volatile nitrosamine occurrence in foods and beverages.

Endogenous Formation

Convincing evidence exists for endogenous formation of nitrosamines in humans. Based on our knowledge of nitrosation in acidic, aqueous media, it is not surprising that nitrosamine formation has been demonstrated to occur in the stomach. Estimation of endogenous nitrosation in humans and experimental animals has been accomplished by measuring urinary

NPRO following oral administration of proline and nitrate. Inhibition of nitrosamine formation has been accomplished by the administration of nitrosation inhibitors such as ascorbic acid. Gastric nitrosation proceeds through reaction between amines and nitrosating agents derived from the diet. Nitrate, which occurs in substantial amounts in certain vegetables, is partially reduced to nitrite by nitrate reductase-containing bacteria in the oral cavity.

Recently, the existence of endogenous nitrosation pathways other than gastric nitrosation has been recognized. Evidence exists for mammalian cellular enzymatic conversion of arginine to nitric oxide, which in turn affects nitrosation. Scientists believe that this process occurs in several cell types, including macrophages. The extent and relevance of cellular nitrosation are yet to be determined.

Estimates of Exposure to Nitrosamines from Foods

Several groups have made estimates of human exposure to volatile nitrosamines from food and beverage consumption. Most of the estimates relate to exposure from consumption of foods and beverages in western Europe. Volatile nitrosamine exposure was predominately from NDMA and NPYR.

The estimation of daily exposure to volatile nitrosamines in most reports ranged from 0.1 to 1.0 μg per person. In comparison, a US National Academy of Sciences report in 1981 estimated exposure to nitrosamines from cigarette smoking to be 17 μg per person per day.

Several recent reviews reported an estimated daily exposure to nitrosamines of 10–120 μg per person from diet. Estimates in this range include, in addition to volatile nitrosamines, *N*-nitrosated amino acids and amino acid derivatives and some include apparent total *N*-nitroso compounds as well. It should be recognized that most *N*-nitrosated amino acids and amino acid derivates have failed to elicit a carcinogenic response in animals and the identity and carcinogenicity of the apparent total *N*-nitroso compounds are unknown.

Some cautionary comments are needed regarding estimates of exposure to nitrosamines. First, estimates should be based on nitrosamines with known identity and known carcinogenicity in animals. Second, a number of the estimates were conducted a decade ago or are based on reports of the volatile nitrosamine content of foods from over a decade ago. As discussed previously, the volatile nitrosamine content of foods, e.g., NDMA in beer, has been dramatically reduced in recent years. Therefore, some of the estimates in the literature may be higher than reflected by current levels of volatile nitrosamines. Third, information currently available does not allow reliable estimation of exposure from currently undetected carcinogenic nonvolatile nitrosamines, of nitrosamines formed endogenously, and of other *N*-nitroso compounds formed exogenously and endogenously. A fourth limitation of estimates is that they are based on average food consumption for relatively large populations. Food consumption patterns for subgroups and individuals within populations vary widely and, therefore, so does exposure to nitrosamines in the diet.

Role of Nitrosamines and Other *N*-Nitroso Compounds in Human Cancer

Based on current information on carcinogenicity in experimental animals, and on other pertinent information, most scientists believe that nitrosamines and other *N*-nitroso compounds are capable of inducing cancer in humans. However, at the present time we do not know what exposure to the various nitrosamines is required to induce cancer in a human population.

Convincing evidence exists for the role of tobacco-specific nitrosamines in cancer induction in people who use tobacco products. Particularly compelling evidence exists for the causative role of tobacco-specific nitrosamines in cancer of the oral cavity for people who engage in snuff dipping and chewing betel quid.

In order to assess more fully the part nitrosamines and other *N*-nitroso compounds play in human cancer, progress will be needed in the following areas: improvement and application of analytical methodology will be required for estimation of exposure to nonvolatile nitrosamines and other *N*-nitroso compounds in foods and in other materials; second, a better understanding of formation and exposure to endogenously formed nitrosamines and other *N*-nitroso compounds is needed; and, finally, methodology will need to be improved in order to allow reliable estimation of human cancer incidence from exposure to relatively low levels of carcinogens, including nitrosamines.

See also: **Amines**; **Cancer**: Carcinogens in the Food Chain; **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; Carcinogenicity Tests; ***Clostridium***: Occurrence of *Clostridium botulinum*; Botulism; **Curing**; **Drying**: Equipment Used in Drying Foods; **Nitrates and Nitrites**; **Smoking, Diet, and Health**

Further Reading

Forman D and Shuker D (eds) (1989) *Cancer Surveys, Advances and Prospects in Clinical, Epidemiological and Laboratory Oncology. Nitrate, Nitrite and*

Nitrosocompounds in Human Cancer, vol. 8, no. 2. Oxford: Oxford University Press (published for the Imperial Cancer Research Fund).

Gloria MBA, Barbour JF and Scanlan RA (1997) *N*-Nitrosodimethylamine in Brazilian, US domestic, and US imported beers. *Journal of Agriculture and Food Chemistry* 45: 814–816.

Gloria MBA, Barbour JF and Scanlan RA (1997) Volatile nitrosamines in fried bacon. *Journal of Agriculture and Food Chemistry* 45: 1816–1818.

Hill MJ and Giacosa A (eds) (1996) Proceedings of the thirteenth annual ECP symposium *N*-Nitroso compounds in human cancer: current status and future trends. *European Journal of Cancer Prevention* 5 (supplement 1): 1–163.

Hotchkiss JH (1989) Relative exposure to nitrite, nitrate and *N*-nitroso compounds from endogenous and exogenous sources. In: Taylor SL and Scanlan RA (eds) *Food Toxicology, a Perspective on the Relative Risks. IFT Basic Symposium Series*. New York: Marcel Dekker.

O'Neill IK, Chen J and Bartsch H (eds) (1991) *Relevance to Human Cancer of N-Nitroso Compounds, Tobacco Smoke and Mycotoxins*. IARC Scientific Publication, no. 105. Lyon: International Agency for Research on Cancer.

Scanlan RA and Tannenbaum SR (eds) (1981) *N-Nitroso Compounds*. ACS Symposium Series, no. 174. Washington, DC: American Chemical Society.

Tricker AR (1997) *N*-nitroso compounds and man: sources of exposure, endogenous formation and occurrence in body fluids. *European Journal of Cancer Prevention* 6: 226–268.

Tricker AR and Preussmann R (1991) Carcinogenic *N*-nitrosamines in the diet: occurrence, formation, mechanisms and carcinogenic potential. *Mutation Research* 259: 277–289.

NMR Spectroscopy *See* **Spectroscopy**: Overview; Infrared and Raman; Near-infrared; Fluorescence; Atomic Emission and Absorption; Nuclear Magnetic Resonance; Visible Spectroscopy and Colorimetry

Nonstarch Polysaccharides *See* **Dietary Fiber**: Properties and Sources; Determination; Physiological Effects; Effects of Fiber on Absorption; Bran; Energy Value

NUCLEIC ACIDS

Contents

Properties and Determination

D W Gruenwedel, University of California at Davis, Davis, CA, USA

Presence of Nucleic Acids in Food

Nucleic acids are natural constituents of all foods derived from animal or plant sources, for they are intrinsic components of the cells making up food tissue. They occur usually in close association with basic proteins (histones, protamines), forming nucleoproteins. In general, their effect on texture, nutritive value, or sensory properties of foods is small in view of their low tissue concentrations (typically < 200 mg of nucleic acid phosphorus per 100 g of fresh tissue). These low tissue concentrations normally also preclude the occurrence of detrimental health effects. However, health consequences may arise with persons suffering from metabolic disorders such as impairment of purine excretion. In these instances, the dietary breakdown of the nucleic acids can generate

unacceptably high levels of uric acid in blood as well as tissues, producing ultimately hyperuricemia, acute and chronic arthritis, or other symptoms of gout.

Properties of Nucleic Acids

Chemical Properties

Nucleic acids are macromolecules of high molecular weight, made up of heterocyclic pyrimidine (Py) and purine (Pu) bases, a five-membered (furanose-type) sugar ring, and orthophosphate. Bases, sugar and phosphate occur in equimolar ratios, i.e., Pu(Py): sugar: phosphate of 1:1:1. Thus, schematically, their chemical composition may be denoted as

$$[\mathrm{Pu(Py)} - \mathrm{pentose} - \mathrm{PO_4^-}]n, \qquad (1)$$

with n being a large number. Among naturally occurring nucleic acids, n may range from 4.5×10^3 (polyoma virus) to 7.0×10^7 (*Drosophila melanogaster* chromosome). Two broad categories exist, deoxyribonucleic acids (DNAs) and ribonucleic acids (RNAs), depending on whether the sugar moiety is β-D-2′-deoxyribose or β-D-ribose (note: primed entities refer to sugar-binding sites). Their elemental composition is roughly 15% nitrogen, 10% phosphorus, 36% carbon, 4% hydrogen, and 35% oxygen. Cellular DNA is localized almost exclusively in the nucleus, while RNA is found predominantly in the cytoplasm.

The principal bases found associated with DNAs are the purine derivatives adenine (6-aminopurine, Ade) and guanine (2-amino-6-hydroxypurine, Gua), and the pyrimidine derivatives cytosine (2-hydroxy-4-amino-pyrimidine, Cyt) and thymine (5-methyluracil, Thy). Occasional base variations (e.g., 5-methylcytosine or 5-hydroxymethylcytosine) are encountered. In RNAs, the base Thy is replaced by uracil (2,4-dihydroxypyrimidine, Ura). Certain RNAs, e.g., (transfer) tRNAs, show considerable base variations: they may contain a wide variety of methylated bases, including Thy, bases with thio groups, inosine, etc.

The Pu bases are bound to the sugar in an N9 → C1′-glycosidic linkage; with Py bases, bonding occurs via Nl → Cl′. The resulting compounds are N-glycosides; they are called nucleosides. Some nucleosides are: adenosine (Ado or A), guanosine (Guo or G), cytidine (Cyd or C), uridine (Urd or U), and thymidine (dThd or dT). Attachment of orthophosphate in positions O3′ (or O5′) produces the corresponding 3′ (or 5′) (mono)nucleotides. 5′-Mononucleotides are the actual building blocks of DNAs and RNAs. Schematically, their polymerization takes place by 5′ → 3′ phosphodiester bond formation, e.g., the 5′-phosphate group of one nucleotide links up with the free 3′-hydroxyl group of another, thereby forming a dinucleotide. It is customary to use short-hand notations such as 5′-pApCpGpT-3′ or 5′-ApCpGpTp-3′ to indicate 5′-terminal and 3′-terminal phosphate, respectively.

In both DNA and RNA, *in-vivo* diester bond formation proceeds solely in the 5′ → 3′ direction, producing long-chain, unbranched structures (strands). Branching in RNA molecules can occur transiently during RNA splicing. Both DNA and RNA strands display polarity, meaning that their primary structure is to be read in the direction → O3′ → [PO_2 → O5′ → C5′ → C4′ → C3′ → O3′] → [PO_2 → O5′ → C5′ → C4′ → C3′ → O3′], etc. (the brackets define a DNA/RNA monomer subunit). The strand polarity is a consequence of furanose pucker as well as associated base stacking.

The chemical, i.e., base composition of nucleic acids is variable and differs from one species to another. Variation in base composition forms the basis of the genetic information content of DNAs. However, in DNAs, certain regularities apply (Chargaff's rules): (1) $\sum \mathrm{Pu} = \sum \mathrm{Py}$, (2) $\sum \mathrm{A} = \sum \mathrm{T}$ and $\sum \mathrm{G} = \sum \mathrm{C}$, (3) $\sum(\mathrm{A} + \mathrm{C}) = \sum(\mathrm{G} + \mathrm{T})$. On occasion, 5-methylcytosine needs to be taken into consideration, too, in order to satisfy points (2) and (3). These rules hold because of the double-stranded (complementary) structure of DNAs brought about, in part, by Watson–Crick hydrogen bonding between opposite bases (A $=$ T and G $\equiv$ C base pair formation; the symbols $=$ and $\equiv$ denote that base pairing between A and T involves two hydrogen bonds, while three hydrogen bonds are formed in the base pairing between G and C). Chargaff's rules do not apply to single-stranded nucleic acids, whether they are RNAs (usually) or DNAs (occasionally).

Physiochemical Properties

At physiological pH values, the orthophosphate groups of the nucleic acids are negatively charged. Outward charge neutralization occurs through positively charged counterions aligning themselves with the negative charges on the polymers. Counterions can be Na^+, Mg^{2+}, basic proteins, polyamines, etc. Nucleic acids are therefore polyelectrolytes, a property that determines their solution behavior: they are extensively hydrated and form gel-like structures if present in higher concentrations. Even in dilute solutions they display non-Newtonian behavior, as noted, for instance, by their shear-rate-dependent viscosity. Once hydrated, they are readily soluble in water. Isolation of nucleic acids from tissues usually yields their sodium salt. How they really exist *in vivo* is unknown, although it is assumed that Na^+ also serves there as the principal counterion.

DNAs, with very few exceptions, are double-stranded helices of opposite polarity. By contrast, RNAs are usually single-stranded, although double- (as well as triple-) stranded RNAs are known to exist. There are also duplex helices made up of DNA and RNA single strands; they form, for example, *in vivo* during transcription. DNA double-strandedness, brought about by interstrand hydrogen bonding and intrastrand base stacking, is termed DNA secondary structure; this structure is salient to the preservation of genetic information by DNA. Hydrophobic bonding and London dispersion forces are the major factors contributing to base stacking. DNA secondary structure is quite stable over a wide range of pH (3–12), ionic strength, I (0.001–5), and temperature (up to 100 °C). The ranges given are approximate: not only are the effects of I and temperature interdependent (rule-of-thumb: the higher I is, the higher the thermal stability), but base composition (GC-rich DNAs are more heat-stable than AT-rich DNAs) contributes heavily to the stability of DNA secondary structure as well. Lastly, at elevated levels of I, individual cation/anion effects, often destabilizing in nature, also become noticeable. Transforming the DNA double helix into two single strands of DNA, a process occurring cooperatively, is called 'denaturation.' Synonymous terms are 'helix-to-random coil transition' or 'order–disorder transition.' 'Denaturation' represents a change in DNA conformation (conformational changes are those that rupture hydrogen bonds and/or hydrophobic bonds but leave covalent bonds intact). It is evident from the above that variations in pH, I, and temperature can be used to bring about changes in DNA conformation experimentally. While much is known regarding the subtleties of DNA secondary structure, very little information is available with respect to the structural properties of the single-stranded states of DNA (or RNA).

Nucleic acids are optically active molecules, i.e., they rotate the plane of (linearly) polarized light and, hence, display chirality. The major contribution to optical activity comes from base stacking, for it imparts molecular asymmetry. The optical activity of the asymmetric carbon atoms of deoxyribose or ribose does not contribute much to the overall effect. Double-stranded DNAs are usually of right-handed screwness or helicity. This information has been obtained from X-ray diffraction analysis of DNA fibers. Depending on the sign and magnitude of the parameters defining the orientation of a base pair in DNA relative to the helix axis (e.g., rotational twist, t, tilt, θ_T, roll θ_R, propeller twist, θ_p, axial rise, h, pitch, width, and depth of major and minor grooves), one can distinguish between conformational families such as A-, B-, C-, and Z-type DNA. They exist among naturally occurring DNAs, i.e., DNAs with random base sequences. Synthetic DNAs (polynucleotides with nonrandom sequences) are characterized by additional families such as B′, C″, D, E, S, Z′. Z-type DNA is a left-handed double helix.

Until quite recently, the consensus was that double-stranded DNA is rather inflexible ('rod' or 'worm-like chain') and that, apart from its exciting base-pairing capabilities, the structure of the sugar–phosphate backbone, in view of its seeming monotony, is of little consequence to its biological function. This perception was due to the fact that X-ray diffraction studies executed on DNA fibers gave structural information only at low levels of resolution. However, experimental evidence has shown that double-stranded DNA is surprisingly flexible locally, giving rise to what has been termed DNA 'polymorphism.' DNA polymorphism refers to seemingly minor, quite localized alterations in secondary structure that may very well be of crucial importance in biological processes such as replication and transcription. That there is such a thing as DNA polymorphism has been demonstrated by X-ray diffraction analysis of oligonucleotide crystals. Oligonucleotide crystals (formerly not available) yield high-resolution X-ray data down to atomic dimensions. It was found that DNA polymorphism resides in the nonplanar ring structure of β-D-2′-deoxyribose (or β-D-ribose in RNA) and that this structure is conformationally quite flexible. Its two main conformations (called sugar pucker) are envelope E (four atoms in a plane, the fifth out of plane) and twist T (three atoms in a plane, the two others out of plane on opposite sides). Out-of-plane atoms on the same side of C5′ are denoted *endo*, while those on the opposite side are denoted *exo*. Transitions between E and T are facile, giving rise to a pseudorotation cycle of the furanose ring in nucleosides. The cycle contains 10 different (*endo*/*exo*) T and E forms. In addition, O5′ can assume a number of orientations about the C4′–C5′ bond axis; they are known as *gauche, gauche* (+ *synclinal*); *gauche, trans* (*antiperiplanar*); and *trans, gauche* (− *synclinal*). Lastly, relative to the sugar moiety, a Pu can adopt two major orientations along the N—Cl′ bond; *anti* (the bulk of the base turns away from the sugar) and *syn* (it is over or toward the sugar). All orientations influence each other, giving rise to a multitude of sugar conformations – and, hence, structural DNA families – with important implications for biological function. Thus, the binding of enzymes (e.g., DNase I, restriction endonucleases) or regulatory proteins such as TFIIIA (zinc fingers) to DNA is now known to be greatly influenced by the geometry of the backbone. In fact,

it has been said that while base pairing may be viewed as the 'brain' of DNA, the conformational flexibility of the sugar–phosphate backbone constitutes its 'heart.'

Determination and Characterization of Nucleic Acids

There exist literally hundreds of techniques that permit the detection, quantitative evaluation, structural, and biological characterization of nucleic acids. They are chemical, biochemical, physicochemical, or physical in nature; however, their description is beyond the scope of this article. 'Determination of nucleic acids' is therefore understood here to mean their detection (qualitatively as well as quantitatively) in cells and tissues in particular, and the determination of their chemical composition (average base composition) when in solution. Hence, techniques that serve their physical characterization in terms of structure, mass, or shape, or techniques used in DNA sequencing, in the enzymatic manipulation of DNA and RNA, in recombinant DNA research, in cloning, etc., are not presented. Techniques relating to their isolation from animal and plant tissues or from microbial sources are also not discussed here.

Detection of Nucleic Acids in Cells and Tissues

Histochemical approaches (*in-situ* staining) Most advantageous are histochemical staining techniques. Each constituent of DNA/RNA (cf. eqn (1)) can be identified histochemically: acidic phosphate groups are demonstrated with the aid of basic dyes, β-D-2′-deoxyribose with the help of a triphenylmethane dye, and Pu(Py) can be made visible via intercalation with fluorescent dyes. The specificity of the techniques can be improved further through judicious use of DNA (deoxyribonuclease) and RNA (ribonuclease) degrading enzymes prior to staining. This eliminates possible interference by the other polymer. Although staining techniques are usually used to obtain qualitative information, they can be adapted to furnish also quantitative data.

Techniques detecting DNA

Feulgen nucleal reaction DNA in tissue is depurinated by strong acid treatment. The resulting apurinic acid releases β-D-2′ deoxyribose, which in turn is converted to its open-chain aldehyde form. The sugar aldehyde reacts with the colorless triphenylmethane derivative leucofuchsin (Schiff's reagent) to give a magenta color (fuchsin). The chemistry of the reaction is known. Since the ribose of RNA cannot be changed to an open-chain aldehyde form, RNA will not be stained; hence, this technique is very specific.

Kurnick's methyl green method Methyl green, a triphenylmethane dye possessing two positively charged amine groups, adds on to double-stranded DNA. Details of the process remain obscure but appear to be related to the secondary structure of DNA: the two positively charged amine groups of the dye are said to have just the correct distance to add on to phosphate groups separated from one another by one turn of the DNA double helix. DNA stains green or green-blue.

Fluorescence flow cytometry A number of macrocyclic dyes (acridine orange (derivative of acridines), Hoechst 33342, Hoechst 33258 (derivatives of piperazine), ethidium, propidium (derivatives of phenanthridine), etc.) bind to DNA bases. Some of them (e.g., acridine orange) will also react with RNA. Their commonality resides in the fact that they are planar, highly aromatic molecules that fluoresce when irradiated. Because of their flatness, they can bind to DNA through intercalation, i.e., they can sandwich themselves between bases. Intercalation increases their basic fluorescence, which makes the nucleic acids visible. Excitation sources are mercury arc lamp, argon ion laser, or krypton ion laser. Since DNA–dye emission spectra differ usually from those of the RNA–dye complexes, distinction between the two types of nucleic acids is possible. Determination of DNA (or RNA) is quantitative. The technique is used most frequently in cell-culture work or in studies in which free-flowing cellular or subcellular particles are available (e.g., red blood cells).

Techniques detecting DNA and RNA

Methyl green–pyronin method If proper experimental conditions are maintained, nuclei (DNA) will be stained green or blue (methyl green), and nucleoli and cytoplasm (RNA) red (pyronin). Since pyronin, a xanthene derivative, is also a fluorescent dye, the technique can be adopted for flow-cytometric determination of RNA.

Acridine orange method The dye acridine orange binds to DNA and RNA. The two dye–polymer complexes exhibit different fluorescence emission spectra: when illuminated with purple light, DNA fluoresces green–yellow to bright yellow and RNA reddish brown to orange; when illuminated with ultraviolet light, DNA fluoresces greenish yellow and RNA crimson red. (*See* **Spectroscopy**: Fluorescence.)

Nonhistochemical Approaches

Radioactive labeling method Nucleic acids can be further identified qualitatively or quantitatively in cells or tissues by labeling them radioactively. Frequently used radionuclides are ^{31}P, ^{14}C, and ^{3}H.

They are administered experimentally in the form of labeled precursors of DNA or RNA (e.g., [^{3}H]dT for DNA, [^{3}H]U for RNA, [α-^{32}P]dATP for DNA, [α-^{32}P]ATP for RNA, etc.) and are incorporated into the polymers enzymatically during macromolecular synthesis. Labeling can be performed *in vitro* or *in vivo*. The presence of the radionuclides and, hence, of the nucleic acids is noted by the radioactive decay of the nuclides in the form of α or β particles or γ rays.

Burton assay for DNA This assay is a colorimetric procedure for measuring the deoxyribose moiety of diphenylamine DNA. Color is produced by a secondary amine (phenylamine) reacting with the sugar. The assay is relatively specific, although RNA in high concentrations, or sucrose, should be absent. DNA does not have to be present in purified form; thus, determinations can be undertaken in crude cellular extracts. DNA is usually extracted from tissue or cells through treatment with strong acids (perchloric acid, trichloroacetic acid), although alkaline extractions have also been used. Extraction is undertaken to remove interfering substances, particularly proteins. Color readings are taken at 595 and 650 nm. A standard curve is prepared by using samples of pure DNA (commercially available). A number of modifications of the basic procedure exist. (*See* **Spectroscopy**: Visible Spectroscopy and Colorimetry.)

3,5-Diaminobenzoic acid fluorescence assay DNA is extracted from cellular material with acid treatment (usually perchloric acid) and the sample exposed to hydrochloric acid (1 N) for about 45 min at 55–57 °C. This produces apurinic acid and free β-D-2′-deoxyribose (see section *Feulgen nucleal reaction*). The C1′ and C2′ carbons of the sugar then react with 3,5-diaminobenzoic acid to produce a strongly fluorescent compound. Excitation occurs at 410 nm; fluorescence is measured at 510 nm. Calibration with pure DNA is required. Submicrogram quantities of DNA can be determined.

Detection of Nucleic Acids in Solution

These techniques assume DNA as well as RNA to exist in solution in essentially pure form, i.e., extraction from tissues or cells and purification occurring prior to analysis.

Spectrophotometric methods

Spectrophotometry in the ultraviolet (concentration determinations) The presence of stacked Pu(Py) bases in DNA/RNA makes the polymers absorb radiation in the ultraviolet region. With duplex DNA, absorption starts at wavelengths below 300 nm and reaches a maximum near 260 nm and a minimum around 230 nm. The presence of a second peak around 200 nm has been suggested; however, its demonstration is not simple because most spectrophotometers cease functioning around 200 nm. Concentration determinations are usually performed at 260 nm by applying the Bouguer–Lambert–Beer law:

$$A^{260} = \varepsilon^{260} cd, \qquad (2)$$

where A^{260} is the absorbance of DNA at 260 nm, ε^{260} the DNA molar absorptivity (1 (mol phosphate)$^{-1}$ cm^{-1}), c its molar concentration in solution (mol phosphate l^{-1}), and d the optical pathlength (cm). By measuring A^{260}, one can readily compute c, the desired quantity, ε^{260}, and these are available from the literature for almost all DNAs. They are usually around 6.6×10^3 l (mol phosphate)$^{-1}$ cm^{-1}. Because of the 1:1:1 molar relationship between phosphate, sugar, and base, c can also be expressed in terms of moles of base per liter or moles of monomer per liter, etc. Since the average molecular weight of a nucleic acid monomer unit is 330, it is easily derived from eqn (2) that for duplex DNA:

$$1\, A^{260}\,\text{unit} = 50\,\mu\text{g DNA ml}^{-1}, \qquad (3)$$

while for single-stranded DNA or for RNA

$$1\, A^{260}\,\text{unit} = 40\,\mu\text{g RNA ml}^{-1}. \qquad (4)$$

Spectrophotometric methods in the ultraviolet (base composition determinations) The secondary structure of DNA can be used for analytical purposes. Heating duplex DNA at a given pH and ionic strength ultimately yields single-stranded DNA ('denaturation'). The process can be followed spectrophotometrically since single-stranded DNA has a higher absorbance (by about 25% at room temperature) at, say, 260 nm than double-stranded DNA. The increase in absorbance is called 'hyperchromicity' (H) and is defined as follows:

$$H^{260} = \left(A_t^{260} - A_0^{260}\right)/A_0^{260}, \qquad (5)$$

where A_t^{260} is the absorbance of DNA at 260 nm at a given temperature, t, while A_0 is its absorbance prior to heating (usually room temperature). H is thus the normalized absorbance increase of DNA due to heat-induced denaturation. Plotting H_{260} against temperature produces the so-called DNA 'melting curve.' The expression 'melting' refers to the fact that DNA 'helix-to-random coil transitions' resemble the infinitely sharp temperature-dependent phase transitions occurring in ice during melting. DNA 'melting curves' are of sigmoidal shape. The limiting value of A_t^{260} at

T_{max} is about 40% higher than A_0^{260}. The temperature at which H^{260} amounts to 50% of its limiting value H_{max}^{260} is called the 'melting temperature' (T_m in °C). T_m is linearly related to the (average) base composition of DNA. For the 'standard saline citrate' medium (0.15 M sodium chloride and 0.0015 M sodium citrate, pH 7.0):

$$\% \text{ GC} = 2.44(T_m - 69.3). \quad (6)$$

Since % GC = $100X_{GC}$, with X_{GC}, representing the 'mole fraction' of GC base pairs in DNA, and since $X_{AT} = 1 - X_{GC}$, eqn (6) furnishes equally well information on %AT. Equations similar to eqn (6) have been established for other solvents. In conclusion, heating native DNA under controlled conditions (constant pH and ionic strength, application of a linear temperature gradient, automatic recording of hyperchromicity) enables the rapid determination of its chemical (base) composition. It is best to work at $A_0^{260} \approx 1$ ($\approx 50\,\mu$g DNA ml^{-1}).

See also: **Spectroscopy**: Fluorescence; Visible Spectroscopy and Colorimetry

Further Reading

Adams RLP, Knowler JT and Leader DP (1986) *The Biochemistry of the Nucleic Acids*, 10th edn. London: Chapman & Hall.

Alberts B, Bray D, Lewis J *et al.* (1989) *Molecular Biology of the Cell*, 2nd edn. New York: Garland.

Ausubel FM, Brent R, Kingston RE *et al.* (1989) *Current Protocols in Molecular Biology*, vols 1 and 2. New York: Greene, Wiley Interscience.

Chayen J and Bitensky L (1991) *Practical Histochemistry*, 2nd edn. Chichester, UK: John Wiley.

Saenger W (1984) *Principles of Nucleic Acid Structure*. New York: Springer-Verlag.

Shapiro HM (1985) *Practical Flow Cytometry*. New York: Alan R. Liss.

Walker JM (ed.) (1984) *Methods in Molecular Biology, vol. 2. Nucleic Acids*. Clifton: Humana Press.

Physiology

H A Simmonds, Purine Research Unit, Guy's Hospital, London, UK

Physiology

Nucleic acids are essential components of all cells and consequently are found in many foods. The term 'nucleic acid' commemorates the first isolation of this vital cell constituent, 'nuclein,' from the spermatic fluid of Rhine salmon and the nuclei of pus cells, by Miescher in 1868. Some 20 years later, it was demonstrated that uric acid, which had been recognized by the Swedish chemist Scheele in 1776 as a constituent of human urine and kidney stones, arose from nucleic acid degradation. Fischer and his school in Germany at the end of the nineteenth century (1895–1899), established the first chemical structure for uric acid, as well as that of other purine and pyrimidine bases. It was demonstrated subsequently that nucleic acids consisted of chains of these purine and pyrimidine bases linked to a pentose sugar, esterified with phosphoric acid (**Figure 1b**). Two types of nucleic acid have been identified – deoxyribonucleic acid (DNA) and ribonucleic acid (RNA).

Physiological Roles

The purine and pyrimidine bases of DNA carry the genetic information of all prokaryotic and eukaryotic organisms, with the sugar and phosphate groups performing a structural role. The human genome is considered to contain between 50 000 and 100 000 genes, each of which is composed of a linear polymer of DNA of varying length. In viruses, genes are made of either DNA or RNA. The infinite variation in genetic information is achieved by the sequence of the four bases that constitute DNA (**Figure 1b**) – the purines, *adenine* and *guanine*, and the pyrimidines, *thymine* and *cytosine*. DNA is double-stranded, and each nucleotide of the chain in one strand is linked by hydrogen bonding to a complementary nucleotide in the other. Complementary pairs of nucleotides are adenine and thymine, and guanine and cytosine. DNA is principally found in the nucleus and is considered to be relatively stable in most cell types. (*See* **Viruses**.)

Ribonucleic acid is essential for the transmission of the genetic message in the form of protein synthesis and must first be synthesized from DNA. In the case of RNA, one of the four bases differs from that in DNA – *uracil* replaces the pyrimidine base thymine – and the molecule is single-stranded, except in some viruses. In contrast to DNA, most of the RNA is in the cytoplasm. Cells contain three types of RNA: *messenger* RNA (mRNA; 5% of total RNA) provides the template for protein synthesis and is relatively labile; *transfer* RNA (tRNA; 15%) carries the message in the form of activated amino acids to the ribosome for the synthesis of specific polypeptides, as determined by the particular mRNA template; and *ribosomal* RNA (rRNA), the major RNA component (80%), is metabolically stable. (*See* **Protein**: Synthesis and Turnover.)

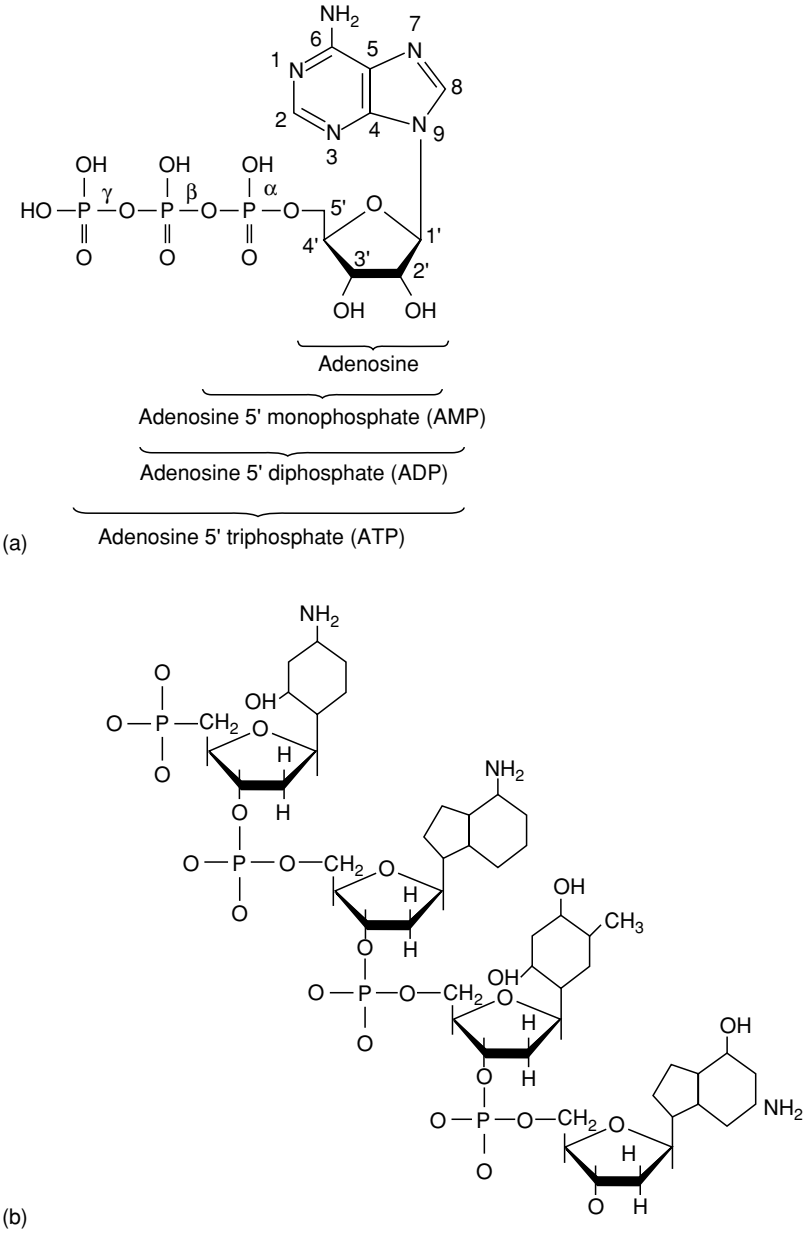

Figure 1 (a) Structural formula of adenosine 5′-triphosphate (ATP) indicating the numbering of the atoms on the ribose, as well as the purine ring, which consists of a six-membered pyrimidine ring fused to a five-membered imidazole ring. The position of attachment of the phosphate groupings for AMP, ADP, and ATP is also indicated. (b) Schematic representation of the structure of part of a single DNA strand made up of a sequence of the four bases, the pyrimidine cytosine, the purine adenine, the pyrimidine thymine, and purine guanine respectively, showing that the deoxyribose has an H group at the 2′ position on the pentose ring, instead of the OH group of ribose. These bases are linked via the 3′-OH group of the deoxyribose-phosphate moiety to the 5′-OH group of the next deoxyribose. Reproduced from Nucleic Acids: Physiology, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

The important physiological roles played by the metabolic pathways responsible for sustaining these different nucleic acid pools in humans is demonstrated by the clinical manifestations when different steps in the synthesis, degradation, and repair of the constituent mononucleotides are defective or absent.

Nucleic Acid Metabolism in Humans

The building blocks of the nucleic acids, the purine and pyrimidine ribonucleotides, are of central importance to virtually all biological processes. Whereas cellular purines are derived exclusively from endogenous sources, cellular pyrimidines are

not. As discussed in detail later, dietary purines are degraded by a battery of enzymes in the intestinal mucosa and enter the circulation as uric acid. By contrast, dietary pyrimidines can be absorbed in the nucleoside form and incorporated by salvage into the nucleotide pool, as evidenced by the lifelong treatment of patients with the pyrimidine *de-novo* salvage defect, hereditary oroticaciduria, with oral uridine.

Nucleotides in infants

Over the past two decades, there has been considerable interest in the role of dietary ribonucleotides in infant feeds, in which significant effects have been claimed. This debate stemmed from reports that ribonucleotides were present in human milk, but not in cows' milk or infant formulas. However, claims that nucleotide concentrations were much lower in cows' than in human milk were apparently related to the state of lactation, with concentrations higher during lactation. Infant formulas were confirmed as being devoid of nucleic acid.

These findings stimulated investigations to determine: (1) whether the nucleotides found in human milk result from degradation of nucleic acids, or are actively secreted as a response to a nutritional demand of the infant, and (2) the newborn infant's endogenous capacity to digest nucleic acids to absorbable products. Nucleotides were found in the greatest quantity in human milk followed by nucleic acids, nucleosides being a minor component. Importantly, the nucleotide/nucleosides were predominantly pyrimidines, purines being present only as uric acid. This specific profile must result from catalysis in the breast. Enzymes capable of degrading nucleotides, including xanthine dehydrogenase (XDH), are known to be present in human milk. Evaluation of the endogenous capability of infants to metabolize RNA and nucleotides confirmed that the intestine of infants, as in adults, digested RNA to cytidine, uridine, and uric acid *in vitro*. This study confirmed the extensive data derived from dietary purine loading in adults that dietary purine nucleotides (if not already degraded by gut bacteria) would be degraded to uric acid in the intestinal mucosa. However, uridine and cytidine would be absorbed (although the latter would be degraded rapidly to uridine by cells *in vivo*).

This research into the importance of human versus cows' milk, or formula feeding, was stimulated by animal studies that had indicated that dietary nucleotides may be required for maintenance of normal immune function in neonatal mice. A subsequent study in healthy term infants demonstrated that in those either breast-fed, or fed formula supplemented with nucleotides, some indices of immune function were indeed significantly higher (natural killer cell cytotoxicity and interleukin-2 production by stimulated mononuclear cells) compared with nonsupplemented formula groups. However, this was found only in the newborn. The rate of growth and incidence and severity of infections did not differ significantly among these different dietary groups at 2 months. It was concluded that nucleotides may be a component of human milk that contributes to the enhanced immunity of the breast-fed infant. Another study implied that allergic diseases develop during feeding of cows' milk, but not human milk. It may be that the dearth of uridine nucleotides in cows' milk after 1 month of lactation could be implicated, since these effects were independent of any difference in the gut flora demonstrated in breast-fed, as distinct from formula-fed, infants.

This debate and the problems addressed arose from the use of animals. As indicated elsewhere in this chapter, rodents have both XDH and uridine phosphorylase everywhere. By contrast, in humans, uridine phosphorylase is confined to the liver, and XDH is confined to the liver, intestinal mucosa, and breast milk. This problem highlights the general lack of awareness of the two curious and unexplained differences in the synthesis of nucleotides in humans. First, pyrimidines derived from nucleotide degradation are salvaged at the nucleoside (uridine) level, purines as the base. Second, whereas purine nucleotides are derived exclusively from endogenous sources in humans, pyrimidine nucleotides can be formed from uridine orcytidine ingested in the diet as well. These important differences are underlined by the successful treatment with oral uridine of patients with uridine monophosphate synthase (UMPS) deficiency, which presents generally as macrocytic anemia. A few patients are also immunodeficient. UMPS is the complex catalyzing the last two steps of pyrimidine *de-novo* synthetic (DNS). Long-term follow-up of such patients confirms that humans cannot only survive without pyrimidine biosynthesis, but also reproduce. Thus, uridine certainly can restore immune function in humans as well.

Role of Endogenous Nucleotides, Nucleosides, and Bases in Cellular Metabolism

Purines and pyrimidines are effectively anchored inside the cell as the nucleotide by attachment to a pentose linked to a mono-, di- or triphosphate group (Figure 1a and 1b), principally the triphosphate. It was originally assumed that all reactions of biological significance took place intracellularly at the nucleotide level. Attention has been focused recently on the extracellular regulatory functions of purine nucleosides (base plus pentose), or the bases themselves. The pentose may be either ribose (ribonucleoside) or

2′-deoxyribose (deoxyribonucleoside) bound by the C1 atom through a glycosidic linkage to the N9 atom of the purine group, or to the N2 of the pyrimidine group (Figure 1).

The importance of purine and pyrimidine nucleotides in cellular metabolism is twofold. In addition to their role in the storage, transmission and translation of genetic information (as the polynucleotides DNA and RNA), as mononucleotides, they play an equally vital role in lipid and membrane synthesis (in the form of purine and pyrimidine sugars or lipids), signal transduction, and translation (in the form of guanosine triphosphate, cyclic adenosine monophosphate, and cyclic guanosine monophosphate), as well as providing the energy (adenosine triphosphate (ATP)) that drives many cellular reactions and forms the basis of the coenzymes (nicotinamide adenine dinucleotide, nicotinamide adenine dinucleotide phosphate, flavin adenine dinucleotide, etc.) (*See* **Coenzymes**.)

All cells require a balanced supply of purine and pyrimidine nucleotides for growth and survival, but this may vary from cell to cell, depending on function. Liver, for example, has a very complex nucleotide profile, compared with heart. These nucleotides may be built up by one of two routes: the energetically expensive multistep synthetic route, or the single-step so-called 'salvage' pathway. In normal circumstances, salvage predominates over synthesis. Figure 2 illustrates the different metabolic pathways involved in the *de-novo* synthesis of these nucleotides, as well as the efficient recycling of the nucleotides or bases derived from them during the wear and tear of daily life (muscle work, wound healing, erythrocyte senescence, protein glycosylation, providing essential nourishment for the brain, etc.). Interestingly, whilst this recycling takes place at the base level for purines, it is the pyrimidine nucleosides that are actively recycled in humans, with only a small proportion being degraded further in either case (Figure 2). The importance of these different pathways to the overall control of nucleotide concentrations in the body is the subject of many excellent reviews and for this reason is summarized only briefly here.

Nucleotide Triphosphate Production and Nucleic Acid Synthesis

Purine and pyrimidine ribomononucleotides built up by either the *de-novo* or synthetic routes are phosphorylated via the correponding diphosphate to the triphosphate, which is the active intracellular form of most mononucleotides. In addition to the variety of vital individual cellular functions described above, these ribomononucleotides are essential intermediates in the synthesis of the polynucleotides RNA and DNA respectively (Figure 2), being incorporated into DNA following the formation of the deoxyribonucleotide from the corresponding diphosphate by the enzyme ribonucleotide reductase. The latter is an allosteric enzyme; its activity and specificity are controlled in a complex manner by both purine and pyrimidine ribo and deoxyribonucleotides. This process is particularly active in cells and tissues with a high rate of turnover (e.g., lymphocytes, gut epithelium, skin, bone marrow, etc.). There is approximately five times as much RNA and DNA in the body.

Breakdown of Nucleic Acids

The polynucleotides DNA and RNA, although relatively stable in most tissues, turn over rapidly in dividing cells. Both DNA and RNA first must be degraded to the constituent mononucleotides, which are themselves degraded further. A variety of enzymes capable of hydrolyzing the phosphodiester bonds have been described and include ribonucleases specific for RNA and deoxyribonucleases for DNA as well as nonspecific nucleases, phosphorylases and phosphomonoesterases. The mononucleotides have the highest turnover rate, DNA the lowest. Further catabolism of the resulting monophosphate will differ depending on whether it is a pyrimidine or purine ribonucleotide of deoxyribonucleotide. For example, whereas adenine ribonucleotides are predominantly deaminated at the nucleotide level in humans by AMP-deaminase, deoxy-AMP is not a substrate for this enzyme and must first be degraded to deoxyadenosine and deaminated at the deoxynucleoside level (Figure 2).

Purine and pyrimidine (deoxy) nucleotides are degraded to the corresponding (deoxy) nucleosides by specific 5′ nucleotidases. Different purine endo- or ecto-5′ nucleotidases have been identified with different substrate specificities and may be of particular importance in providing bases for nucleotide resynthesis in tissues where there is rapid cell turnover and massive cell death (e.g., thymus, spleen, bone marrow). As mentioned above, whereas the normal metabolic route for pyrimidines is salvage at the nucleoside level, that for purine nucleosides and deoxynucleosides is degradation to the corresponding base by purine nucleoside phosphorylase, prior to salvage. This degradation is favored by the high intracellular inorganic phosphate and low ribose 1-phosphate levels in most tissues. Interestingly, these phosphorylases are not reactive toward either adenosine or cytidine, or their analogs, in human cells and first must be deaminated at the (deoxy) nucleoside (or nucleotide) level.

Salvage is an active process for both pyrimidines and purines. Consequently, only a small fraction of

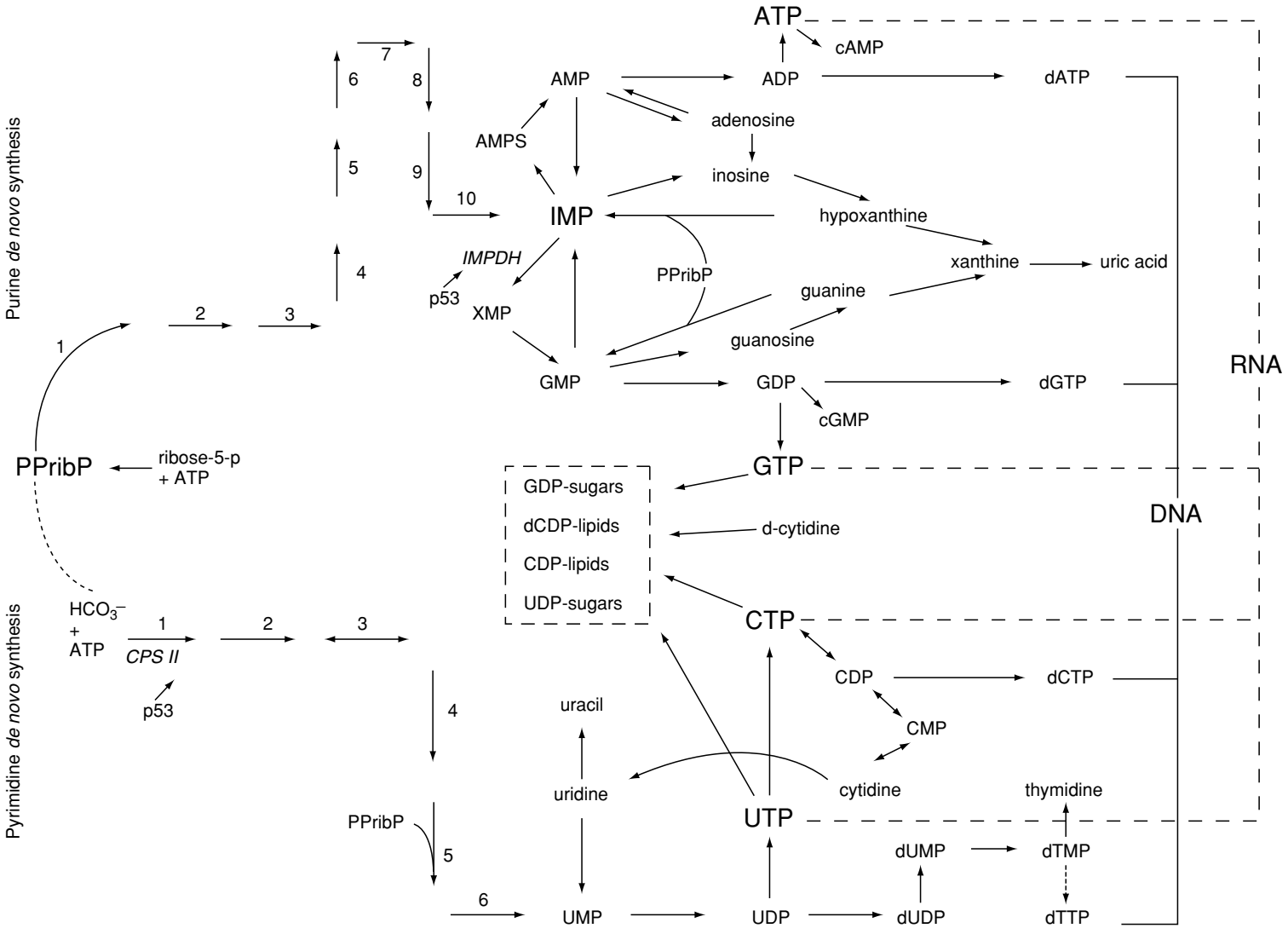

Figure 2 Multistep purine and pyrimidine nucleotide de novo synthetic (DNS) routes and single-step salvage (hypoxanthine, guanine, uridine, cytidine, D-cytidine) pathways, indicating the enzymes inosine monophosphate dehydrogenase (IMPDH) and carbamoyl phosphate synthetase 11 (CPS 11) targeted by analogs which deplete nucleotide pools, reportedly associated with a p53-dependent Go/G1 arrest. Note the involvement of 5-phosphoribosyl pyrophosphate (PPribP) in both purine DNS and salvage as well as in pyrimidine DNS. The role of ATP, GTP, UTP, and CTP in DNA and RNA synthesis, of ATP and GTP in 2nd messenger synthesis, of UTP, CTP, and GTP in the synthesis of the glycoconjugates (dashed box) essential for membrane synthesis, structure, and function and protein glycosylation, is also evident.

the nucleotides turned over daily is actually degraded and lost to the body. The pyrimidine bases, uracil and thymine, derived from nucleosides not recycled are degraded further to the β-amino acids, and there is thus no measurable end product. However, such loss is probably comparable with that for purines, the normal end product of which in humans is uric acid, formed from the precursor purine bases xanthine and hypoxanthine by the action of xanthine dehydrogenase (Figure 2).

Rates of Nucleotide Synthesis and Degradation in Different Cells

It has become apparent that the original concept of endogenous metabolism and its overall control, involving a complex interplay between *de-novo* synthesis and salvage, does not apply to all cells but is governed by a tissue- or cell-specific complement of enzymes and/or controls on them, depending on the function of that cell or tissue. The human erythrocyte, for instance, is anucleate and lacks the ability to use either salvage or *de-novo* synthesis to maintain its ATP levels, being dependent on adenosine scavenged from other tissues for this. In addition, the pyrimidine nucleotides found in nucleated cells are absent from mature erythrocytes, the only pyrimidines normally present being in the form of uridine diphosphate (UDP) sugars. ATP is also the most important purine in both skeletal and heart muscle, adenine nucleotides making up 95 and 90% of the total nucleotide complement, respectively. Although DNA in most tissues is considered relatively stable, it is evident from the two inherited disorders associated with immunodeficiency that cell death and rapid turnover of cells of the hemopoietic system (e.g., extrusion of the nucleus during erythrocyte maturation, or mounting an immune response) normally produces significant amounts of deoxyribonucleosides, as well as ribonucleosides, which must be degraded further. These disorders have highlighted the fact that removal of metabolic waste from DNA catabolism is vital to the normal immune response; failure to do so can result in the accumulation of deoxy-ATP and deoxy-GTP. These deoxynucleotide triphosphates are extremely toxic to T-lineage stem cells, resulting in severe combined immunodeficiency affecting both T- and B-cells in subjects deficient in adenosine deaminase, or a T-cell-specific immunodeficiency, in purine nucleoside phosphorylase deficiency (Figure 2).

Endogenous Nucleic Acid Synthesis in the Gut

It was originally reported from studies in guinea-pigs that the gastrointestinal tract was incapable of *de-novo* synthesis. However, subsequent workers have shown significant synthetic activity in rat intestine, as evaluated by the incorporation of radiolabeled glycine into RNA. The nucleic acid content of intestinal mucosa is high, as is the rate of cell turnover in the luminal villi, and it has been calculated in rat that about 30 mg of endogenous nucleic acid enters the lumen daily. This implies a considerable loss of both purines and pyrimidines, which can only be replaced by *de-novo* synthesis. Studies of nucleotide concentrations in rat intestine have shown both pyrimidine and purine nucleotides at concentrations equivalent to those in liver, supporting active nucleotide metabolism in the intestine.

Intestinal absorption is obviously related to intestinal motility and local blood flow. The presence of adenosine receptors on rat jejunal mucosal cells may be particularly important in the regulation of fluid and electrolyte absorption as well as the active transport of nutrients. Such adenosine may be generated slowly from nucleotides by nucleotidases physically or functionally coupled to the adenosine membrane translocator.

It is evident from the above that the pathways leading to the formation and degradation of nucleic acids are complex, and that many factors determine the origin as well as amount of endogenous pyrimidine or purine nucleotides turned over daily. Moreover, this will differ depending on the cell or tissue.

Role and Fate of Dietary Nucleic Acids

The metabolism of nucleic acids ingested in the diet difers from that of endogenous nucleic acids. The intestinal mucosa plays an important role in this. The above studies confirm that pyrimidine as well as purine nucleotides in intestinal mucosal cells are derived from endogenous sources. However, a growing body of evidence supports a role for dietary nucleotides derived from the gut in intestinal development, turnover, and repair. Suggested effects include enhancement of the host mucosal defense system and an influence on neonatal lipid metabolism and on iron bioavailability, implying a novel role for nutrition in the modulation of gut function.

Most of our knowledge relating to dietary nucleic acid metabolism in the intestine is derived from studies of the absorption of exogenous purine in different species – mouse, dog, rat, pig, etc., as well as humans – which date back as far as the latter half of the nineteenth century. Investigations in animals have shown that whereas the ribose attached to nucleosides derived from RNA was further metabolized, the phosphate was absorbed and excreted in the urine.

Recent studies using ^{13}C-labeled nucleic acid to supplement the diets of rats and chickens have

provided further evidence for the incorporation of dietary pyrimidine nucleosides, but not purine nucleosides, directly into hepatic RNA. The successful lifelong treatment with uridine of patients with hereditary oroticaciduria – a defect in pyrimidine *de-novo* synthesis – confirms that dietary uridine is certainly absorbed and salvaged into mono- as well as polynucleotides in humans. The lack of effect of oral uracil in this disorder also confirms that uracil is a metabolic waste and that pyrimidine salvage occurs at the nucleoside level in humans.

Recent investigations have addressed the supposedly beneficial effects of dietary CDP-choline (citicholine) in patients with stroke or trauma, theoretically by increasing acetylcholine production in cholinergic neurons, as well as the amount of cell membrane. These studies confirmed that, when taken orally, CDP-choline elevates plasma levels of both choline and uridine (not cytidine) in humans, but no benefit was demonstrated in clinical trials.

Metabolism of Dietary Purines by Gut Bacteria

Animal studies have shown that the metabolism of dietary nucleic acid and the corresponding purine nucleotides, nucleosides, and bases in the gut is rapid. Isotope studies demonstrated that up to 50% of the radiolabeled purine is recovered as carbon dioxide within 30 min. The role of gut bacteria in this process was indicated by experiments using the XDH inhibitor, allopurinol, concomitantly, when the radiolabel was recovered *in toto* (in urine and feces only), presumably relating to inhibition of bacterial XDH. Allopurinol not only inhibited purine degradation but also decreased the absorption of dietary purine, a seemingly beneficial effect that may explain the reduction in total purine excretion noted in human subjects on a normal diet taking allopurinol.

Intestinal Mucosa Degrades Dietary Purines to Uric Acid

Interestingly, the above radiolabeling studies in pigs (where, as in humans, XDH activity is significant only in intestinal mucosa and liver), confirmed a lack of any radiolabel incorporation from dietary purine into tissue nucleotides. These studies demonstrated conclusively that dietary purine is degraded to a nonreutilizable form by the intestinal mucosa. Studies in other animal species have confirmed this, purine nucleotides, nucleosides, and bases absorbed from the gut lumen being largely converted to uric acid during passage across the mucosa and released as such in serosal secretions, prior to further degradation by urate oxidase (uricase) in the liver. Thus, in contrast to pyrimidines, humans have no apparent requirement for dietary purines. The intestine thus serves as an effective barrier through the activity of a battery of enzymes capable of rapidly degrading purines already partly processed by gut bacteria to the nonreutilizable metabolic waste, uric acid. Such rapid degradation of purines by the gut presumably reflects an important evolutionary development to protect the integrity of the human genome. The uric acid produced daily in humans thus derives from two sources: catabolism of exogenous as well as endogenous mono and polynucleotides (Figure 3).

Excess Dietary Levels of Nucleic Acids and their Clinical Consequences

As mentioned above, pyrimidines have no measurable end product. The potential toxicity of dietary nucleic acids thus relates predominantly to a single mutational event that has resulted in the fact that the insoluble uric acid, and not the much more soluble allantoin as in other mammalian species, is the only measurable end product of nucleic acid degradation in humans. Although it had long been accepted that the enzyme uricase had been lost in the course of human evolution, more recent studies have established that the absence of uricase activity results from a lack of gene transcription, rather than loss of the gene itself. However, other factors play an equally important role in determining the pathogenesis of nucleic acids ingested in the diet.

Role of Exogenous Purine in Determining Circulating Uric Acid Concentrations

The ingestion of food now known to be rich in purines has been noted for millennia to be high in subjects with what has been designated 'primary' gout, the gout affecting predominantly the middle-aged male. This type of gout is a disorder of affluent societies consuming diets rich in purines; it is rare in women or children. During times of hardship, uric acid concentrations in the population fell considerably, and 'primary' gout almost vanished. The fact that gout was extremely prevalent among wealthy Englishmen for more than three centuries up to World War I is not surprising when the dietary habits of the day are examined. These affluent gentlemen habitually consumed vast meals comprising many courses and frequently including 16 different meats, the majority of which were rich in nucleic acids and other purines.

Type of Nucleic Acid and Toxicity

Detailed studies of dietary nucleic acid absorption in humans were carried out by Zöllner and coworkers in the late twentieth century and confirmed that nucleic acids ingested in the diet exert their major

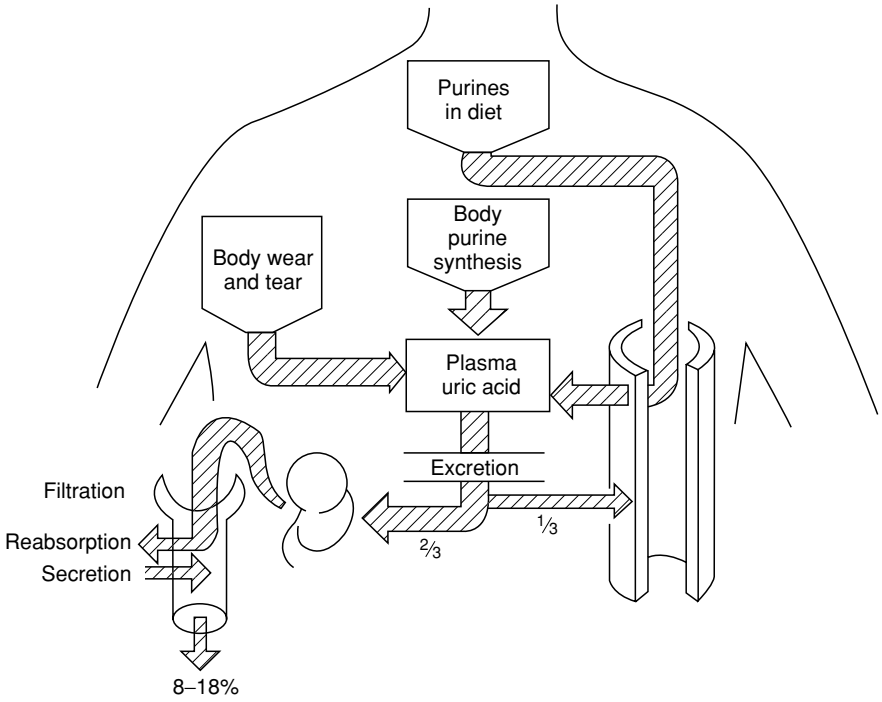

Figure 3 Schematic diagram showing the factors influencing the plasma uric acid, the end product of endogenous (body synthesis plus daily wear and tear) and exogenous (dietary) purine metabolism in humans. The different mechanisms by which this metabolic waste is eliminated (two-thirds by the kidney, one-third by the gut) in normal individuals are also shown. At the bottom left is a simplified version of the complex factors (involving filtration, reabsorption and secretion) interacting in the proximal tubule of the human kidney and resulting in the urinary excretion of only 8–18% of the filtered load depending on age and sex. Reproduced from Nucleic Acids: Physiology, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

effect on uric acid levels. These studies showed that the nature of the nucleic acid ingested is equally important, with the effect of RNA being more than twice that of DNA (Figure 4a). This effect is also evident whether the purine is ingested in the form of nucleic acids, mononucleotides, nucleosides, or bases. Moreover, some forms of purine, e.g., guanosine – the principal RNA degradation product in yeast-rich beverages such as beer (especially real ale) – are absorbed and catabolized more readily than others. A controlled study of diet in primary gout patients compared with control males of comparable age demonstrated that the average daily intake of most nutrients, including purine nitrogen, was similar. However, gouty patients drank significantly more alcohol, predominantly in the form of beer (60 g per day, equivalent to 2.5 l of beer) and had a significantly higher mean plasma uric acid (0.49 mmol l^{-1} compared with 0.39 mmol l^{-1}).

Role of the Kidney or Drugs in the Genesis of Toxicity

The pathological changes that result from uric acid being the exclusive end product of purine metabolism in humans are due entirely to the insolubility of this metabolic waste, coupled with the other peculiarity that primates display, i.e., the renal tubule reabsorbs around 90% of the filtered urate (Figure 3). Net reabsorption is slightly higher in normal males (92%) than in females (88%) and is lower in children of either sex (80–85%). This explains the higher plasma uric acid in adult males and makes them more vulnerable to situations of overproduction, or overingestion. Clearly, sex and age are equally important contributory factors.

It is noteworthy that when normal subjects are fed yeast RNA, the plasma urate rises little with increased intake, the increase in the excretion of urate being dramatic when the rise in plasma urate is modest (Figure 4b). Thus, neither overingestion nor overproduction (unless sudden and massive) is likely to raise the concentration of uric acid in the plasma considerably, if the renal response is normal. It is obviously not so in the majority of gouty males with primary gout, where the striking change is that, for any plasma urate concentration, the excretion of urate is consistently less than in normal subjects. This defect in handling by the kidneys relates to a greatly reduced

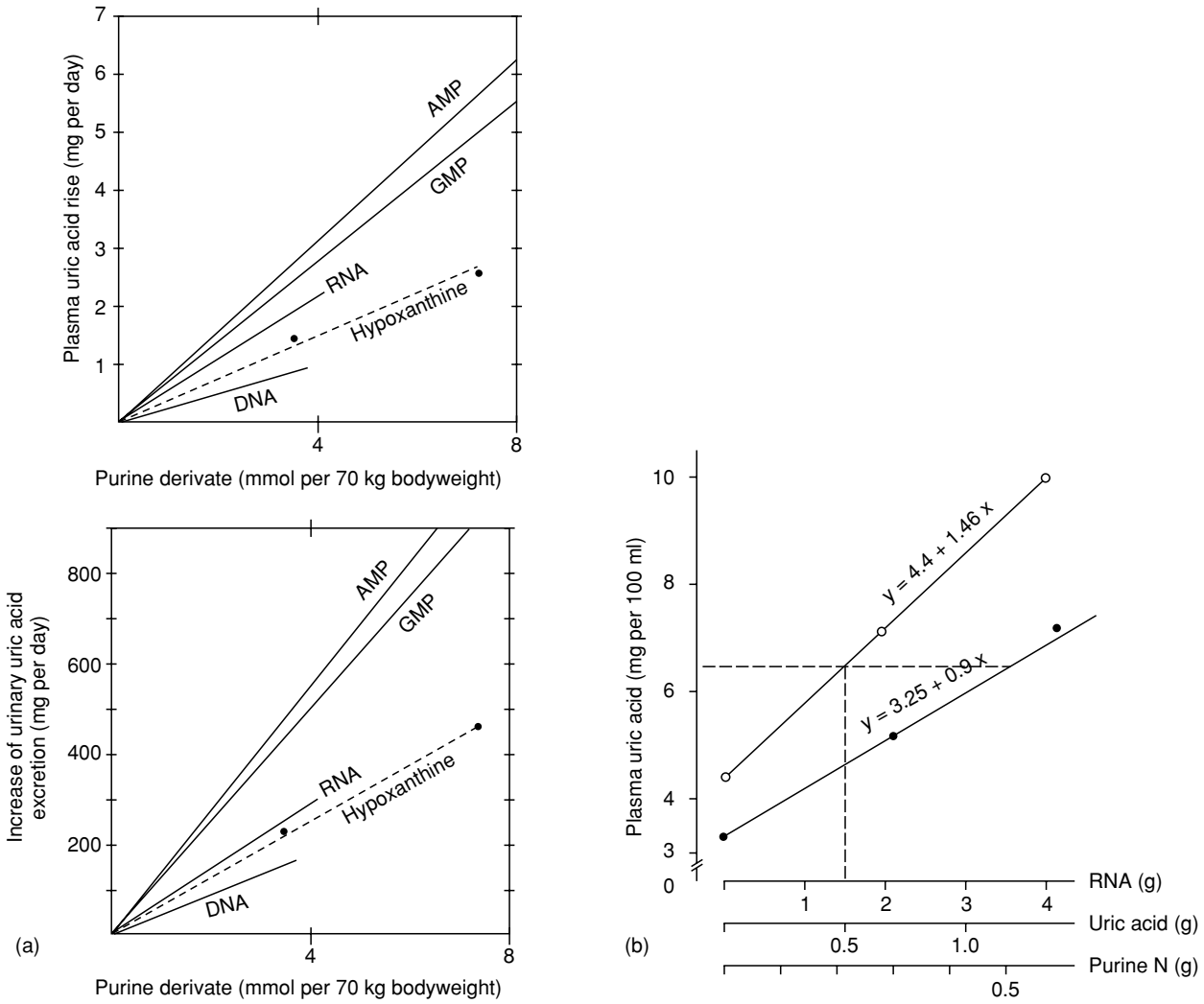

Figure 4 (a) Increase in plasma and urine uric acid (mg per day) in response to a purine load from the different sources indicated on a molar basis. (b) Influence of graded additions of RNA to a purine-free formula diet on plasma uric acid levels of either normouricemic (-•-) or hyperuricemic (-○-) individuals (some of whom had a positive family history of gout). Reproduced from Nucleic Acids: Physiology, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

clearance of uric acid relative to the glomerular filtration rate (FE_{UR}); FE_{UR} is the fractional excretion of uric acid. The mean is 5.4% in gouty males compared with 8.1% for healthy males of comparable mean age (52 years).

The exact nature of this difference is not yet clear, since studies must be reevaluated in the light of recent knowledge of how complex the renal tubular handling of urate is. However, there is no doubt that it involves the transport of uric acid, which occurs predominantly in the proximal tubule, is bidirectional, and has both a secretory and reabsorptive component (Figure 3). Current opinion suggests that the majority of gouty patients underexcrete urate because of a defect in tubular secretion. Confirmation must await identification at the molecular level of the different transporters involved in urate reabsorption and secretion. Clearly, a combination of events is needed to produce hyperuricemia and the clinical syndrome of gout. These are a large intake of readily absorbed purine coupled with a defect in the renal handling of uric acid at the kidney level, which means that the kidney cannot respond to a purine load without an abnormal rise in plasma urate concentration. (*See* **Renal Function and Disorders:** Kidney: Structure and Function.)

However, diet alone may not be the only culprit. Numerous other physiological and pathological agents such as lead are also capable of reducing urate excretion, and hence exacerbate the rise in

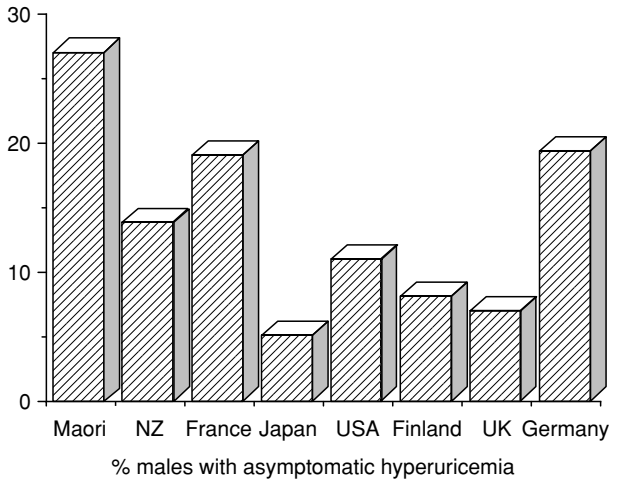

Figure 5 Percentage of Caucasian males, compared with a Polynesian race (the Maori), with asymptomatic hyperuricemia (defined as a plasma uric acid in excess of 0.42 mmol l^{-1}, or 7.0 mg dl^{-1}) in four EU countries, North America, Japan, and New Zealand (NZ).

plasma urate caused by diet. Any of these may lead to an acute attack of gout in a susceptible individual whose plasma urate is already elevated. The best known of the physiological substances are organic acids such as lactate; their overproduction may explain in part the hyperuricemia associated with excessive alcohol consumption coupled with inadequate food intake. Diuretics cause plasma volume contraction, and fractional urate reabsorption increases under these circumstances. The appearance of gout in patients treated with hypotensive agents over long periods currently accounts for over 50% of new presentations in gout clinics. Unusually, the number of elderly females in this group is high, and the mode of clinical presentation is frequently atypical.

Diet may also be important in drawing attention to an unusual subset of patients with juvenile onset of gout; a disorder affecting young males and, unusually, young females and children of either sex equally. Unlike primary gout, where renal function is normal for age, unrecognized, this dominantly inherited disorder is associated with progressive renal disease. Considerable variation in presentation has existed, and there has been no consensus as to whether the gout or the renal disease is the primary factor. However, it is now evident that there are two hallmarks for this disease: the first is hyperuricemia disproportionate to the degree of renal dysfunction; the second is that the degree of renal hypoexcretion of urate is generally extreme, which explains the associated tendency to gout. Moreover, the mean FE_{UR} in this group is only 5.1%, irrespective of age or sex: lower than that of the middle-aged gouty male (5.4%), and even more remarkable considering the high percentage of children and young women in this group. This universally low clearance explains the added susceptibility to dietary purine leading to the isolated attack of gout in children as well as young adults, which has drawn attention to families with 'familial renal disease,' hitherto leading to death in the 30s. Recognition of the correct nature of the familial disorder has underlined the need to measure urate clearance in all kindred members and enabled early diagnosis and treatment of 42% of seemingly healthy children.

This is important because long-term follow-up (>20 years) in these patients treated with plasma uric acid lowering agents has shown that early recognition and treatment ameliorates the progression of the renal disease. These studies implicate uric acid in the etiology of the renal disease, but this has been disputed by others. However, very recent studies suggest that a reappraisal of the pathogenesis and consequences of hyperuricemia in hypertension, cardiovascular disease, and renal disease is warranted, which adds credence to the above hypothesis.

Role of Dietary Nucleic Acid in the Genesis of Urolithiasis

Although overingestion of purines by subjects with a normal renal response does not precipitate gout, it can predispose to uric acid lithiasis. Uric acid stones are relatively common in Australasia and similar countries, where the consumption of purine-rich beverages and food (e.g., beer, seafood, and meat) is high, and in individuals addicted to health foods such as yeast tablets. Vitamin C is also uricosuric, and acute uric acid nephropathy and sometimes urolithiasis have also been reported following therapy with a variety of uricosuric drugs. Perhaps less well recognized is the uricosuric effect of a high-protein diet and the fact that the intake of purine-rich foods also predisposes to calcium stone formation.

Some foods rich in nucleic acid – vegetarian diets consisting of pulses and grains and yeast extracts – have a particularly high adenine content. Although this is not a problem in the normal population, such diets can be potentially lethal to a rare group of subjects with a genetic defect leading to the inability to recycle adenine, overexcretion of adenine, and the even more insoluble uric acid analog 2,8-dihydroxyadenine (2,8-DHA). Patients with this defect generally present in childhood with kidney stones composed of 2,8-dihydroxyadenine. One such case from a commune consuming a macrobiotic diet presented in acute renal failure with severe renal damage, progressing to dialysis and transplantation. Diet can thus be an important precipitating factor in the pathogenesis of inherited disorders associated

with the overexcretion of insoluble purines. (*See* **Macrobiotic Diets; Vegetarian Diets.**)

Reducing Dietary Levels of Nucleic Acids

Diet varies considerably in different countries and must be taken into account when investigating patients for suspected disorders of purine metabolism. Normal ranges for uric acid in plasma and urine differ greatly in the healthy population depending on the country (Figure 5). For example (and in contrast to the nineteenth century), until recently, the majority of subjects in the UK ingested a low purine diet consisting of one meat meal a day, and urinary uric acid excretion above 3.5 mmol per day was considered abnormal. By contrast, the upper limit of normal in Australia has been given as 7.00 mmol per day. However, the increasing affluence of some societies is now leading to an increase in gout, not only in adult Caucasian males, but also in countries such as Japan, where gout was once rare.

Turnover of Exogenous and Endogenous Uric Acid

A useful guide to the dietary purine consumption in different countries can be obtained from the percentage of males with asymptomatic hyperuricaemia. Using this yardstick, the highest consumption in the EU 20 years ago was in France (lovers of pâté and seafoods), the lowest in the UK (Figure 5). The statistics for Germany were from Bavaria where, as in New Zealand and Australia, beer consumption is high. Although similar statistics for diet-related differences are unavailable today, plasma uric acid will have risen along with the increase in obesity and blood pressure, especially in the UK. Race also plays a part, and Polynesians have a genetically low urate clearance compared with Caucasians, as exemplified by the New Zealand Maori (FE_{UR} 4.9% in normouricaemic males, 3.9% in asymptomatic hyperuricemic males).

As we have seen, the body pool of urate, and hence the plasma urate concentration, is the result of a balance between production, ingestion, and excretion. The method for assessing the contribution of diet is to evaluate *de-novo* production of purines by placing the subject on a purine-free diet for 5–7 days and measuring the urinary excretion of urate, which will equal endogenous production. In this way, less than 1–5% of subjects in any country have been found to excrete abnormally large amounts of urate (>3 mmol per day). In such rare cases, an underlying genetic metabolic defect can be generally demonstrated in which the normal feedback controls on *de-novo* nucleotide synthesis and thus endogenous purine production are overridden, resulting in gross uric acid overproduction. Two such defects – both of which are X-linked and generally present in childhood, or adolescence, but sometimes neonatally – have been identified.

Purine Content of Foods

A knowledge of the purine content of specific foods is essential if dietary effects are to be reduced to a minimum. Until recently, such data have been difficult to find but are now available on the web. Most tables give only purine nitrogen, which, as demonstrated by the purine loading studies mentioned earlier, is not always a good guide because of the variation in absorption. Pâté is a particular culprit, as is most offal and organ meat (liver, kidney, heart, brains, sweetbreads), game (venison, pheasant, partridge, grouse), and the nucleic-acid-rich fish and seafoods – herring, kippers, sardines, smelts, sprats, anchovies, salmon, trout, mackerel, crustaceans (crab, lobster, prawns), shellfish (scallops, mussels), and caviar or roe. Purine nitrogen varies and ranges from 50 mg per 100 g in beef steak to 234 mg per 100 g in sardines. Many fresh vegetables, e.g., spinach, peas, beans, lentils, mushrooms, asparagus, and cauliflower, also have a considerable purine content, as have soya and other pulses and grains (porridge and oats, wheat and rye cereals). All meat extracts (Bovril, Oxo) or yeast extracts (Barmene, Tastex) are very rich in purine.

However, many studies have established that humans addicted to diets rich in nucleic acids are generally very reluctant to alter their dietary habits despite the strongest advice to do so. Consequently, therapy to reduce the pathological effects of dietary nucleic acids, namely the elevated uric acid levels, becomes essential.

Therapeutic Approaches to Reducing Uric Acid Levels

Numerous plasma uric acid-lowering drugs are in current use. Some act by increasing the renal elimination of uric acid (uricosuric drugs, e.g., benzbromarone), or restrict its formation (e.g., allopurinol). Allopurinol reduces urine uric acid levels as well. As mentioned earlier, studies in both humans and animals have shown that allopurinol has an additional beneficial effect in reducing dietary purine absorption. Allopurinol is usually a safe drug, but in rare instances, mostly in renal disease, other factors must be considered. The active metabolite of allopurinol, oxypurinol, is handled by the kidney in a fashion akin to uric acid and thus, even normally, has a long half-life. Since excretion of oxipurinol is reduced in renal failure, the allopurinol dose must be reduced to lower the plasma oxipurinol and minimize the risk of bone marrow depression, or other undesirable side-effects, which include epidermal necrolysis and

hepatotoxicity. A recent study pinpointed the poor response to allopurinol in heavy drinkers and related this to the combined effect of ethanol in impairing urate excretion and increasing production. In rare instances, patients have had a severe allergic reaction to allopurinol. In such cases, the potent uricosuric drug, benzbromarone has proved beneficial, even in renal failure patients with kidney function as low as 25% of normal.

Allopurinol and Acute Renal Failure

It is important to note that in patients with genetic uric acid overproduction, allopurinol should be used with care. Its use to avert gout (or uric acid nephropathy) has precipitated acute or chronic renal failure due to xanthine nephropathy and sometimes xanthine stones instead. Xanthine is even more insoluble than uric acid, and as with uric acid, solubility cannot be improved by alkalinization of the urine. Xanthine nephropathy also may occur during massive endogenous nucleic acid breakdown in patients given allopurinol during aggressive therapy with cytotoxic drugs for malignant disorders. In such cases, much lower doses of allopurinol should be coupled with adequate hydration and alkalinization of the urine.

Role in Chemical Carcinogenesis

Substantial evidence from microorganisms and mammalian cells has implicated mutagenic events caused by damage to endogenous DNA as the initiating factor in carcinogenesis. Damage to critical regions of the genome of somatic cells can be produced by a variety of environmental mutagens. Examples of environmental factors affecting humans are cigarette smoke, asbestos, and ultraviolet (UV) irradiation. Recent putative additions to the list are radiation from microwaves and mobile phones.

A high correlation invariably exists between the mutagenic activity of different chemicals and their carcinogenic activity. Epidemiological evidence indicates a relationship between exposure to benzene and nonlymphocytic leukemia in humans, but the significance of DNA adduct formation in this is not clear. Normally, DNA possesses active repair systems to protect against such damage, which can be caused by a variety of chemical and physical agents, including ionizing, radiation, and UV light. Strands may become cross-linked, bases can be altered or lost, and phosphodiester bonds may be broken. UV light induces the formation of pyrimidine dimers, which are recognized and cleaved by specific endonucleases. Failure to repair the damage before DNA replication occurs results in the damaged region becoming the site of somatic mutations, chromosomal rearrangements or amplifications, and aberrant DNA methylation.

In this context, it is of interest that the P53 gene is altered in many human cancers. Recent studies using analogs that target specific enzymes of either purine or pyrimidine nucleotide biosynthesis have identified P53 as a cellular regulator of both nucleotide synthetic pathways (Figure 2). These studies have led to the proposal that P53 is not only a sensor of DNA damage, inducing apoptosis in response to DNA strand breaks, but rather a sensor of nucleotide depletion. Upregulation of P53 expression, activated by nucleotide depletion or related processes, prevents cells entering the S phase when precursor nucleotide pools are low, thus avoiding replication of damaged DNA through a prolonged, but reversible, G0/G1 arrest. The significance of these findings is that elongation of the S phase during nucleotide depletion is known to be associated with increased chromosome breakage. (*See* **Carcinogens:** Carcinogenic Substances in Food: Mechanisms; Carcinogenicity Tests; **Mutagens**.)

See also: **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; Carcinogenicity Tests; **Coenzymes**; **Infant Foods**: Milk Formulas; **Macrobiotic Diets**; **Mutagens**; **Protein**: Synthesis and Turnover; **Renal Function and Disorders**: Kidney: Structure and Function; **Vegetarian Diets**; **Viruses**

Further Reading

Diem K and Lentner C (eds) (1956 and 1970) *Scientific Tables – Chemical Composition of Foodstuffs*, 7th edn, pp. 230–243. Basle: Geigy.

Grahame RG, Simmonds HA and Carrey EA (eds) (2002) *Gont – an at your fingertips guide*. London: Class Publishing.

Grimble GK (1994) Dietary nucleotides and gut mucosal defence. *Gut* 35: S46–S51.

Henderson JF and Paterson ARP (1973) *Nucleotide Metabolism: An Introduction*. New York: Academic Press.

Linke SP, Clarkin KC, Di Leonardo A, Tson A and Wahl GM (1996) A reversible, p53-dependent G0/G1 cell cycle arrest induced by rNTP depletion in the absence of DNA damage. *Genes and Development* 10: 934–947.

Scriver CR, Beaudet AL, Sly WS and Valle D (eds) (2001) *The Metabolic and Molecular Basis of Inherited Disease*, 8th edn. New York: McGraw-Hill.

Secades JJ and Frontera G (1995) CDP-choline: Pharmacological and clinical review. *Methods and Findings in Experimental and Clinical Pharmacology* 17: 1–54.

Stone TW and Simmonds HA (1991) *Purines: Basic and Clinical Aspects*. London: Kluwer.

Zöllner N and Gresser U (eds) (1991) *Urate Deposition in Man and its Clinical Consequences*. Berlin: Springer-Verlag.

Nutraceuticals *See* **Functional Foods**

NUTRITION EDUCATION

S S Deshpande, Central Institute of Agricultural Engineering, Nabibagh, Bhopal, India

Background

Advances in nutrition research have provided means of prevention and control of some of the major crippling nutritional disorder with sever socioeconomic and health repercussions. The next logical step after the discovery of such solutions will be their propagation among the concerned community members. It is here that health and nutrition education has a role to play. The emphasis in these nutrition education programs ought to be on the proper use of natural dietary sources of nutrients and the need for full use of available health services and health and nutrition intervention program launched by government and other agencies.

Nutrition education programs, when properly implemented, have the potential to bring out desired behavior modifications among the communities. But for them to be successful, there are several obvious prerequisites such as the right approach suited to the community appropriately, motivated change agents, suitable educational strategies in terms of media, messages, etc. The ultimate impact of any nutritional education effort hinges on the extent to which the above criteria are satisfied.

Nutrition education has been defined as the process by which beliefs, attitudes, environmental influences, and understanding about food lead to practices that are scientifically sound, practical, and consistent with individual needs and available food resources; nutrition education should be available to all individuals and families. The fundamental philosophy of nutrition education is that efforts should focus on the establishment and protection of nutritional health rather than on crisis intervention. It is needed, regardless of income, location or cultural, social or economic practices, or level of education. Nutrition education must be a continuing process throughout the life cycle as new research brings additional knowledge.

Scope

Education means a change in behavior. It moves the individual from lack of interest and ignorance to increasing appreciation and knowledge and finally to action. Nutrition education offers a great opportunity to individuals to learn about the essentials of nutrition for health and to take steps to improve the quality of their diets and thus their well-being.

Nutrition education must continue throughout the individual's life in order to accommodate for developments in nutrition science and for changing economic circumstances, health requirements, and the new food products appearing in the nation's market. This requires a greatly expanded use of the mass media, and the involvement of governmental and private agencies and universities, as well as the food industries.

Concept

A concept is an idea around which the content of nutrition education curricula can be built.

What is Nutrition Education?

Nutrition education is the process of teaching the science of nutrition to an individual or group. Health professionals have a different role in educating an individual in the clinic, community, or long-term health-care facility. In these settings, the dietician, nutritionist, or nurse serves to assist or enable individuals to incorporate changes in eating patterns and behavior into their lives. The major focus of this type of nutrition is not knowledge and facts, but rather the development of permanent behavioral changes. This is the art of nutrition education – breaking down a large body of knowledge into small, individual components that are represented to a patient or client at a rate and level, at which they are able to absorb and use the information. Effective education is making nutrition information digestible and usable in an everyday setting.

Steps Involved in Nutrition Education

The development of an organized, coordinated nutrition education program requires a series of steps:

- Consumer research is needed to determine consumer use of existing information attitudes, knowledge, levels of awareness and concern, current

nutrition practices, and type of education to which they should be most receptive.
- Nutritional surveillance data are needed to determine the most important public information problems for determining reasonable objectives to be used in formulating the nutrition education.
- Most appropriate target populations for nutrition education are determined on the basis of awareness, nutritional status, current knowledge, and receptivity to nutritional information.
- A nutrition education message should be designed that is appropriate for the needs of the selected population groups.
- The curricula for the nutrition education program must be well thought out and planned. At every step, the learner and the educator should work together to develop goals, objectives, and appropriate activities and resources to meet the goal.
- Evaluation is crucial to any successful educational effort.

Why Nutrition Education?

A number of forces have emerged to stimulate increased recognition of the need for a stronger, more comprehensive and creative effort at nutrition education. These factors include:

- Many serious as well as less serious health problems are the result of lifestyles and personal habits, including habits of food consumption. Individual modifications in lifestyles may improve health status.
- Food Patterns. Differing forms of food vary in nutritive value as well as in taste appeal. The average consumer is often mystified and unaware of how to make the wisest food choices from such a variety. Thus, there is a great need for education to enable consumers to select wisely an appropriate diet from the vast supply of food available.
- Lifestyles are drastically different from those of the turn of the century. The public must be knowledgeable about how to meet nutritional needs without overindulgence in foods.
- Nutrition education is also necessary to aid the consumer to make economical food choices, that is, to save money while getting optimal nutrition. This is especially important for the low-income consumer.
- On a global basis, there is a real possibility of food shortages. Therefore, the consumer should know how to provide nutritional substitutes for unavailable food items.

What to Emphasize in Nutrition Education

It is very difficult to define the basic knowledge that should be conveyed to the public about nutrition. Much will depend upon what the learners already know, how they evaluate their present dietary habits and nutritional status, what their present eating habits actually are, how much they desire to know, and what changes they are willing or can afford to make.

Nutrition Education Strategies

In view of the multifactorial influences on food habits, there is no one way to persuade people to change their food choices. Thus, nutrition education must use a variety of techniques that must be a part of total family and community environments.

In other words, nutrition education involves information exchange as well as techniques to motivate and reinforce improved food habits. Successful nutrition education must include endeavors to make beliefs, attitudes, values, environmental factors, and individual ideas about food conducive to nutritionally sound, practical, and acceptable dietary habits. Nutritional education can be approached in various ways. What might be termed the rational–empirical curriculum design is the approach that is typically used by educators; it is logical and planned, and it involves behavioral objectives, activities that are designed to achieve objectives, and evaluation to determine the discrepancy between planned and achieved outcomes. The idea of self-responsibility for health suggests the basic notion that the individual possesses a dignity, worth, and responsibility to maximize these characteristics. Self-responsibility assumes that the individual has the potential and motivation to make wise judgments about factors that affect their health status. Another strategy has been termed the travel metaphor method. The idea is that there is a body of knowledge, facts, concepts, values and believes that are known and understood by the educator. It is the educator's responsibility to lead the learner through this body of knowledge and beliefs, and point out to the learner what is thought to be of importance. The educator is thus a travel guide whose job is to provide the student with a set of stimuli and the opportunity to take advantage and react positively. It then becomes the responsibility of the student to determine how they will make use of the contacts provided by the educator.

Another approach called the garden metaphor explains how humans grow, develop and flower naturally. The educator has the opportunity to nurture this environment so that the 'child' has the most opportunity to mature and to learn.

Another approach to health and nutrition education takes a strategy of social manipulation. The basic premise is that the matters affecting health are too urgent, too important, and beyond the individual's

power to control on their own. Thus, the role of the health educator is seen as one of attempting to control behavior and to shape the desired responses. This approach is often employed by the mass media.

The manipulative approach is more concerned with the behavior response patterns rather than the manner in which the individual reaches these patterns. The approach is an attempt to reach the desired behavior through a short-cut method and does not take time to prepare the individual to make fully informed, wise choices.

The most appropriate and effective strategy for stimulating improved food habits will depend upon the knowledge, background, personal characteristics, and motivation of the audience. It may also depend upon the size of the target group. Ideally, the approach should be individualized to the target audience's needs, goals, objectives, and other characteristics. A mix of various approaches may be necessary.

Systems Approach to Nutrition Education

The 'systems' approach to nutrition education provides a useful and practical tool to aid the nutrition educator in deciding what should be taught and how it should be taught. Simply defined, the systems approach is a flow of steps needed to achieve the desired objectives. The approach consists of input, process, and output.

A successful nutrition education effort must incorporate managerial elements that are crucial to the success of all program planning. These elements, which could be termed the input, are essential to an effective nutrition education process. They include needs assessment and identification of problems; study of available resources and alternative strategies of intervention; determination of objectives and goals; assignment of resources to implement program; staff development to implement program, and program monitoring and evaluation. Crucial to the entire process is a firm organization base and administrative and budgetary support for nutrition education efforts.

Follow-up and Evaluation

Nutritional counseling follow-up, in order to evaluate learning, is an essential factor, which needs to be included in the plan. After an initial interview, follow-up is usually planned for some time during the next 1–2 weeks or perhaps at a later time in the same week. As clients progress toward a goal, follow-up visits are generally spaced further and further apart. In a follow-up session, the client's progress needs to be measured against the objectives, and the client needs to receive positive feedback about their progress. In the event that no progress has been made or the client has slipped back into old habits, the counsellor needs to pursue the problem assertively. The problem needs to be identified and new goals or methods of achieving the goals established.

Summary

Nutrition education is a broad area; however, the nutritionists or nurses facilitating it are usually concentrated in the acute or long-term care facilities and community clinics. In all aspects of teaching and counseling, instructional objectives need to be established, the learner's knowledge assessed, behavioral objectives written, evaluation tools devised, information needs screened and organized, education materials selected, and instruction critiqued and revised. It is also essential to keep the learner's needs in the forefront and work to maintain a rapport with the client.

Education is not complete until the learning has been evaluated. Formal and informal tools may be used to measure learning against established objectives. Evaluation helps identify if further teaching or clarification of information is needed. To some degree, demonstrating successful learning at evaluation reinforces continuance of new behaviors, which is the goal of nutrition education.

See also: **Community Nutrition**; **Dietary Requirements of Adults**; **Famine, Starvation, and Fasting**; **Food Safety**

Further Reading

American Dietetic Association (1973) Position paper on nutrition education for the public. *Journal of the American Dietetic Association* 62: 429.

Caliendo MA (1986) *Nutrition Education. Definitions and Concepts in Nutrition and Preventive Health Care*, pp. 410–437. New York: Macmillan.

Deshpande SS (1999) Study of nutritional pattern of some selected villages around Bhopal with special reference to soybean consumption. Ph.D. Thesis, Barkatullah University, Bhopal.

Griffin GA (1974) Strategies in Curriculum Development In: *Conference on Education in Nutrition, Looking Forward from the Post*. New York: Columbia University, Teachers College.

Rau P (1982) Experiences in nutrition education. *Nutrition* 22(2): 3.

Robinson CH, Lawler MR, Chenowth WL and Garwick AE (1986) *Nutrition Problems and Programs in the Community. Normal & Therapeutic Nutrition*, 17th edn., pp. 355–357. New York: Macmillan.

SNE Concepts for Food & Nutrition Education (1982) *Journal of Nutrition Education* 14: 1.

NUTRITION POLICIES IN WHO EUROPEAN MEMBER STATES

A Robertson, WHO Regional Office for Europe, Copenhagen, Denmark

Introduction

In December 1992, after more than 2 years of preparatory work by the member states and the joint efforts of the World Health Organization (WHO) and the Food and Agriculture Organization (FAO), the International Conference on Nutrition (ICN) was convened in Rome to:

- increase public awareness of the extent and seriousness of nutrition- and diet-related problems worldwide
- promote effective strategies and actions to address these problems
- encourage the political commitment necessary to do so

Each of the 159 participating countries and the European Community reaffirmed their determination to insure sustained nutritional well-being for all people and committed themselves to achieve this goal by unanimously adopting the World Declaration and Plan of Action for Nutrition. The nine action-oriented strategies of the World Declaration and Plan of Action for Nutrition are summarized in Table 1. These provide a framework for national plans of action.

In 1996 a follow-up to ICN (Table 2) was held in the European region through a FAO/WHO consultation in Warsaw, Poland. Its aim was to assess implementation of the World Declaration. Representatives from European member states and Canada, the USA, the European Commission and the Holy See attended the meeting along with representatives from the United Nations Development Program (UNDP), the United Nations Children's Fund (UNICEF), nongovernmental organizations (NGOs) and the private sector.

Table 1 The nine action-oriented strategies of the plan of action for nutrition

Incorporating nutritional objectives, considerations, and components into development policies and programs
Improving household food security
Protecting consumers through improved food quality and safety
Promoting breast-feeding
Caring for the socioeconomically deprived and nutritionally vulnerable
Preventing and controlling specific micronutrient deficiencies
Promoting appropriate diets and healthy lifestyles
Assessing, analyzing, and monitoring nutrition situations
Preventing and managing infectious diseases (N.B. not included in this analysis, since not mentioned in any country reports)

Methods

Member states were requested to prepare a summary report of their progress for the 1996 follow-up meeting in Poland. Based on the 35 reports received (response rate of 70%) the WHO Nutrition Unit compiled an overview of the situation.

Member States were grouped into eight geographic regions: Balkan, Baltic, Central Asian Republics (CAR), countries of central and eastern European (CCEE), western Europe, southern Europe, Commonwealth of Independent States (CIS), excluding CAR, and Nordic countries (Table 3). The aim of this is to facilitate comparative analysis and interpretation and assist in drawing conclusions and recommendations.

Results

The country data are presented in tables under the headings of the strategies of the Plan of Action for Nutrition adopted at the ICN.

Discussion

Developing Food and Nutrition Policy

Around the time of ICN (1992), some member states in Central Asia, the former Soviet Union and the Balkan region were not fully independent and so were unlikely to have information about the ICN. Thus food and nutrition specialists in some parts of the WHO European region were probably unaware of the ICN and so it is hardly surprising that many countries had not developed national plans of action according to the ICN objectives.

Moreover a few WHO member states are not members of FAO (1998) and others have only recently joined. This probably explains why there were more health than agriculture representatives present – 66% compared with 27% – at the consultation in Poland in 1996.

One of the difficulties when attempting to develop intersectoral policies is the level of commitment that

each stakeholder has to public health versus their own possibly diverging interests (Figure 1). It is improbable that intersectoral policies will be implemented, unless all stakeholders are committed to the process. However, the long-term interests of the agriculture sector, the food industry, wholesale and retail representatives are dependent on policies which have been developed in collaboration with the voluntary sector and consumers. Consumers, the customers of the food industry, should be more involved in the process whereby food policy is developed.

Intrasectoral Collaboration within the Health Sector

Good collaboration between nutrition and food safety is essential because the public and consumers perceive food in a holistic way. Consumers do not compartmentalize food or distinguish between food safety and nutrition – consumers want good wholesome food they can enjoy without fear. There appears to be far more collaboration between the nutritionists and food safety specialists working in central and eastern Europe compared with those in western Europe.

There may be several reasons for the stronger collaboration in central and eastern Europe. Nutrition is a relatively new science and in eastern parts of Europe its evolution seems to be closely linked to 'hygiene' and the sanitary–epidemiology system. In eastern Europe nutrition and food hygiene have evolved from the same postgraduate specialization, usually as a specialization of medicine. In former socialist countries food safety is traditionally under ministries of health, whereas in some western European

Table 2 Member states at the International Conference on Nutrition follow-up consultation, Poland, September 1996

Total no. of FAO/WHO member states invited	*Total no. of countries present*	*Total no. of representatives*	*Percent of health and agriculture representatives from European region*	*Total no. of reports submitted (European region)*
57	40	62	66% health 27% agriculture 7% others	35

FAO, Food and Agriculture Organization; WHO, World Health Organization.

Table 3 Member states of the World Health Organization European region, 1996

Balkan	*Baltic*	*CAR and Turkey*	*CCEE*	*Western Europe*	*Southern Europe*	*CIS*	*Nordic*
Albania	Estonia	Kazakhstan	Bulgaria[a]	Austria	Greece	Armenia	Denmark
Bosnia and Herzegovina	Latvia	Kyrgyzstan[a]	Czech Republic	Belgium[a]	Israel	Azerbaijan	Finland[a]
Croatia	Lithuania	Tadjikistan[a]	Hungary	France	Italy	Belarus	Iceland
Federal Republic of Yugoslavia (Serbia and Montenegro)		Turkey	Poland	Germany	Malta	Georgia[a]	Norway
		Turkmenistan[a]	Romania	Ireland	Monaco[a]	Republic of Moldova	Sweden
Former Yugoslav Republic of Macedonia[a]		Uzbekistan	Slovakia	Luxembourg	Portugal[a]		
Slovenia[a]				Netherlands[a]	San Marino[a]	Russian Federation[a]	
				Switzerland	Spain[a]		
				UK		Ukraine	

[a]No report submitted (Finland was added later).
CAR, Central Asian Republics; CCEE, countries of central and eastern Europe; CIS, Commonwealth of Independent States.

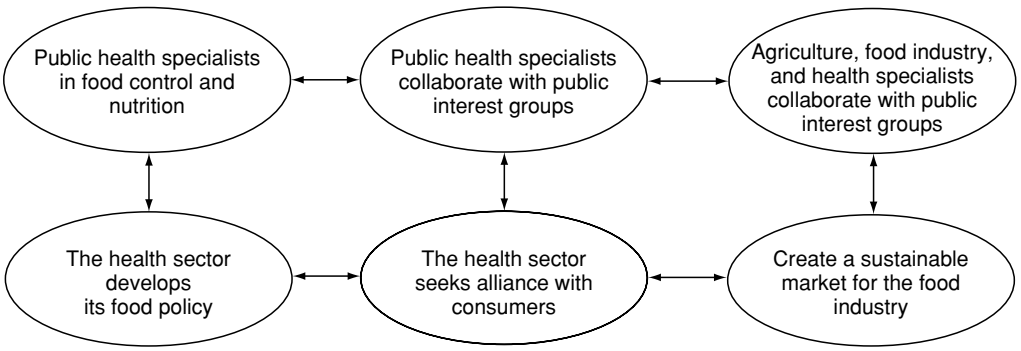

Figure 1 From intra- to intersectoral policy development food priorities for health.

countries only nutrition is under ministries of health and the responsibility for food legislation and enforcement may be with ministries of agriculture or trade in the west. However this is changing, especially in some European Union (EU) member states, where ministries of health are taking a much more proactive role in protecting consumers and their health.

Coordination between public health specialists working in food safety and nutrition is important for many other reasons. One is that from time to time food control authorities may issue warnings about food, e.g., chicken that contains *Salmonella*, or fish and vegetables which are contaminated. Meanwhile, as part of national campaigns, nutritionists promote the consumption of vegetables, fish, and poultry. Public health specialists must deliver consistent and reliable information to avoid public confusion. Closer cooperation between nutritionists and food safety specialists can prevent the promotion of conflicting messages.

Intersectoral Collaboration

Public health specialists dealing with food safety and nutrition should strengthen their collaboration with other sectors dealing with food, e.g., agriculture, food industry, wholesalers and retailers, voluntary sector and public interest groups representing consumer interests. Consumers have a vested interest in supporting the supply of safe food of good quality that is nutritionally healthy and the food industry has a vested interest in supplying healthy consumers. There is an opportunity for the health sector to strengthen health alliances with public interest groups working in the voluntary sector.

Surprisingly, in the country reports, there was very little mention about the importance of collaboration with consumer agencies except by Hungary, Norway, Poland, and Sweden. In these countries there is an understanding that the food industry can profitably increase the availability of vegetables and fruit, low-fat milk, and lean meat, on condition that it is assured a market.

Assessing the Nutrition Situation

Almost all countries reported a need to improve nutrition information systems. According to DAFNE (Network of the Pan-European Food Data Bank Based on Household Budget Surveys: countries participating are Belgium, Denmark, Greece, Germany, Hungary, Ireland, Poland, Portugal, Spain, and the UK), some had no information, while some use secondary data or access nutrition information from household budget surveys. Some had information from one-off surveys that were sponsored through international cooperation. The Nordic countries and the UK seem to have established the most comprehensive surveillance systems that provide information on food consumption and nutritional status on a routine basis. Clearly it will take time for some countries to achieve this level of sophistication but ultimately these data are essential to facilitate food and nutrition policy development.

It seems unlikely that, at least in the near future, nationally representative nutrition surveys will be carried out on a regular basis; therefore other strategies have to be found: members of DAFNE extract nutrition-related data from household budget surveys; in Italy secondary data sources were used to develop a national policy. Mortality data are available from the WHO Health For All database which is accessible through the internet from the WHO website. These data allow comparison of mortality statistics between different countries. FAO food balance sheets provide data such as the percentage energy coming from carbohydrate, fat, protein, and alcohol and the quantity of vegetables and fruits available nationally. These data are directly accessible from the FAO website.

In the future the EU may encourage its member states, and those wishing to join, to develop standard information systems. This will allow comparison of data on food intake, nutritional status, and diet-related behavior in different EU member states and so facilitate the development of cost-effective strategies that enable countries to save money on treating preventable diseases.

Improving Food Security

After the Second World War the main aim in Europe was to increase the supply of food, especially animal products, e.g., meat and milk. However the postwar policy was too successful (Table 4) and in the 1980s until the present time Europe has surplus stocks of butter, meat, and milk.

Since the Second World War, new scientific evidence has led to the level of recommended daily protein intake being reduced. For example the level of protein intake recommended by the FAO, WHO, and the United Nations University (UNU) in 1986 was $0.75\,g\,kg^{-1}\,day^{-1}$. This new level resulted in many countries reducing the level of their national recommended protein intakes for adults from 120 g to around 50 g per person per day. There is now some evidence suggesting that excess intake of some sources of protein, especially from red meat, could be associated with ill health.

Various hypotheses exist to explain how meat consumption may contribute to chronic diseases such as heart disease and certain cancers. Meat intake is associated with an increased intake of saturated fats

Table 4 Food policy in the twentieth century

1950s	*1960s*	*1970s*	*1980s*	*21st century*
Food insecurity and perceived nutrient deficiency, especially protein	Increased food production and consumption, especially of animal protein and fat	Surplus stocks of food, especially animal products	Links identified between mortality from NCDs and diets high in animal fats and low in vegetables and fruit	The opportunity to change food policy to meet health and environmental recommendations

NCDs, noncommunicable diseases.

which are linked to an increased risk of cardiovascular diseases and breast and colorectal cancer. Excessive protein intake may be associated with increased excretion of calcium and so possibly exacerbate the prevalence of osteoporosis. Excess intake of protein is associated with increased risk of renal disease, especially in infants, since the kidneys have to excrete excessive solute loads and so excess protein intake is perhaps linked to the risk of developing high blood pressure later in life. It is advisable that ministries of health review their national recommended intakes to insure that these are in line with current international recommendations.

Lessons learned over the past 50 years suggest future food policies should insure that:

- cereal and potato production should be geared to supply > 50% energy
- vegetables (excluding potatoes) and fruit production should be geared to supply a consumption of at least 400 g per day per person

Given the recommendations, contained in most dietary guidelines, 'to protect, promote and support the consumption of cereals, potatoes, vegetables and fruit,' it is surprising that there was no mention of the need to increase the production of these foods in the country reports. Without a food policy which guarantees food security in the form of cereals, potatoes, vegetables, and fruit, it is difficult to see how national targets for a healthy diet can be achieved.

One of the reasons why families living in countries undergoing economic transition (CEE and CIS) are not suffering from protein/energy deficiency is because many produce a large percentage of their own potatoes, vegetables, and fruit. In Russia, town dwellers produce 88% of their potatoes, 43% of their meat, 39% of their milk, and 28% of their eggs on urban household plots. This important share of production is generated on plots of 0.2–0.5 ha which together constitute only 4% of the total amount of agricultural land in Russia. In addition, imports from abroad have supplemented national food production and currently 50% of chickens and 30% of meat consumed in the Russian Federation is imported. In the short term, imports are essential during times of hardship and economic transition, but in the long term agricultural polices should promote sustainable production and protect local food production to insure national food and nutrition security.

In eastern Europe the agriculture sector forms a much higher percentage of national gross domestic product compared with western Europe. Almost 30% of the population in some CEE countries are employed in agriculture compared with only 1 or 2% in western Europe. Many economists are concerned about the impact when more countries join the EU because the Common Agriculture Policy (CAP) is not sustainable. The CAP receives almost half the EU budget and in 1995 it cost EU taxpayers 39 billion ECU. During the CAP reform process, health recommendations and dietary guidelines should influence the reformed policy in addition to concerns about national employment and the national food security of the EU accession countries.

Protecting Health Through Food Safety and Quality

In some reports there was no mention of the importance of food safety, especially in the countries of southern Europe. As mentioned in the introduction, it is vital to strengthen the collaboration between food control and nutrition agencies. Some western governments are already attempting to bring these two disciplines under one food authority and three-quarters (27 out of 35 countries) did include food safety in their ICN reports. Close collaboration becomes even more essential as the line between food safety and nutrition becomes more and more blurred with the introduction of precooked foods; functional, novel, and special dietary foods; and supplements. Moreover, with the introduction of food into global trade, the public's need for information will be immense. To provide reliable information, both food safety specialists and nutritionists need data on food intake patterns. More effective use could be made of resources if joint surveillance and risk management of food intake are carried out together.

Foodborne diseases are among the most widespread health problems in the contemporary world.

In rich and poor countries alike they impose substantial health burdens, ranging in severity from mild indispositions to fatal illnesses. The emergence of new foodborne diseases is an ominous trend. Epidemics due to newly identified pathogens such as *Campylobacter jejuni, Listeria monocytogenes, Escherichia coli 157:H7* and bovine spongiform encephalopathy (BSE) have hit industrialized countries. *Salmonella* outbreaks are one of the main causes of epidemics in Europe. Global trade will make it more difficult to contain foodborne diseases within national borders and enforce national food laws. In addition, chemical contamination, toxic materials, pesticides, veterinary drugs, and other agrochemicals require constant surveillance to insure their safe use. Similarly, the use of food additives can improve the quality of the food supply but appropriate controls are necessary to insure their proper use.

Most regions included in their country reports the need to harmonize their national food legislation with the EU and World Trade Organization (WTO). Many countries stated they are lining up to join either the EU or WTO (12 countries) and so intend to harmonize their national legislation with international directives set by the EU or Codex Alimentarius. It is essential that health professionals, especially those working in public health, become more informed about international agreements on global food trade and take a full part in the WTO Committee on Sanitary and Phytosanitary Measures (SPS Committee). Only if public health specialists, from both food safety and nutrition, participate can they hope to influence future food policies.

Promoting Breast-Feeding

Clearly there is a lot to be done in the European region to promote breast-feeding. More than one-third (13) of the 35 member states made no reference to breast-feeding in their country report. This does not necessarily mean that there are no active breast-feeding programs in the country (e.g., Romania) but it does mean that there is probably little coordination between those responsible for nutrition policy and those responsible for promoting breast-feeding. Perhaps this is because breast-feeding is regarded as a clinical issue dealt with by midwives and primary health care workers. However, public health nutritionists and hygienists in CCEE and the CIS do have an important role and should be active in the development of national policies to promote breast-feeding.

National breast-feeding initiatives will only be successful if policies are developed and enforced according to guidelines from the Baby-Friendly Hospital Initiative (BFHI) and the International Code for the marketing of breast milk substitutes. It is necessary to national and regional trends (marketing strategies and breast-feeding levels) to interpret the national situation and make recommendations to policy-makers.

The BFHI jointly launched in 1991 by WHO and UNICEF, aims to enable women to choose and practice breast-feeding as the primary source of nutrition up to about 6 months by minimizing obstacles which make it difficult or prevent women from exercising their right to breast-feed and to insure the cessation of free and low-cost infant formula supply to hospitals. The BFHI targets maternity services and hospitals, particularly health workers and those responsible for policies to help mothers succeed in breast-feeding. To become a baby-friendly hospital every facility providing maternity services and care for newborn infants should implement the Ten Steps to Successful Breastfeeding outlined in the joint WHO/UNICEF statement entitled *Protecting, Promoting and Supporting Breastfeeding: The Special Role of Maternity Services* and should end free and/or low cost supplies of breast milk substitutes.

There is clearly a vital need to increase the number of Baby-Friendly Hospitals in Europe and develop and enforce national legislation which is based on the International Code of Marketing of Breast Milk Substitutes.

In some countries there may be conflict between nutritionists, hygienists, and new breast-feeding policies. Some hygiene standards and decrees in the former Soviet countries, such as the recommendation that mothers should be separated from their babies after birth, contradict the current UNICEF and WHO recommendations to promote the 'rooming-in' of babies with their mothers. In addition, National Nutrition Institutes involved with research and development of breast milk substitutes and weaning foods may unintentionally interfere with breast-feeding promotion.

Complementary feeding (weaning) was not discussed in detail in any report and only mentioned very briefly in two country reports, those of Albania and Italy. Perhaps this is because the categories listed in the World Declaration of ICN do not specifically include complementary feeding, only breast-feeding. This was perhaps an oversight by the ICN committees since it is clearly important that all member states address their policies on infant and young child feeding. It is around the age of 6–24 months when mortality and morbidity levels are at their highest. It is likely that the high prevalence of anemia seen in some parts of the European region is due to poor feeding practices, such as the early introduction of foods and drinks, especially cows' milk and tea, before the age of 6 months.

Caring for the Deprived and Vulnerable

In the region around 70% of the population is suffering from physical and mental exhaustion as people cope with increasingly uncertain conditions related to very rapid economic and social change. This is especially true for the countries of the former Soviet Union where policy-makers are not used to developing safety nets for vulnerable groups. The transition to a market economy has resulted in soaring consumer prices due to a reduction in subsidies while salaries have remained low. In addition there is rising unemployment as a result of industrial restructuring and the new monetary polices have hit vulnerable groups in particular. Mechanisms must be developed to identify and protect the vulnerable groups in society.

Caring for vulnerable groups was mentioned in almost three-quarters (22) of the country reports. Many initiatives were related to the physiologically vulnerable, such as children, pregnant women, or the elderly. In addition, welfare safety nets, aimed at insuring a minimum family income by setting minimum wages, pensions, and unemployment benefits, were considered necessary for economically vulnerable groups such as the poor, low-income, unemployed, and refugees or immigrants and those with large or one-parent families.

The gap between the rich and poor is increasing in many countries in Europe, and as a phenomenon of the economic transition in central and eastern Europe, unemployment has appeared in many. If this trend continues, the food intake and nutritional status of vulnerable groups should be monitored closely in order to develop safety nets and prevent food insecurity. Among European industrialized nations the UK was exceptional in the pace and extent of the increase in inequality in the 1980s and by 1990 the level of inequality was almost back to 1930 levels.

Poverty is a growing concern in Europe, especially for those with incomes below societies' poverty lines. In 1994 it was estimated that 13.5% of the Polish population (about 5 million) had average monthly expenditures lower than the relative poverty line. In 1995, 4 million children in the UK lived in families with incomes below 50% of average earnings. Lack of food is associated with poverty, which can be measured by estimating the percentage of disposable income spent on food: in Romania the average figure is around 60%; and in Poland around 40%, compared with 22% in the EU.

Coping within a limited budget often means that healthier, safer foods are not affordable. In Lithuania increased food prices have caused socially deprived people to consume cheaper, less nutrient-dense foods which are more likely to be contaminated. In Poland the prevalence of nutrient deficiencies is highest amongst the unemployed and low-income families with many children. Poor people in the UK consume much lower quantities of vegetables and fruit.

There is growing evidence for the protective effect of vegetables and fruit against chronic diseases such as cancer and coronary heart disease. This suggests that the low intakes of vegetables by poor people may directly increase their risk of these diseases. Inadequate diets also have adverse consequences on a child's health, education, and future employability and there are demonstrable costs to society in each instance.

Micronutrient Deficiencies

There is no doubt that, in the European region of WHO, the most prevalent health problems related to diet are noncommunicable diseases such as cardiovascular diseases and cancers and not deficiency diseases. Nevertheless, micronutrient deficiency does exist in certain European countries under certain circumstances.

Iodine One of the most prevalent nutrient deficiencies is iodine. Iodine deficiency disorders (IDD) are prevalent throughout Europe and only six countries have no IDD (Finland, Iceland, Norway, Sweden, Switzerland, and the UK). Although this problem has been largely eliminated in some countries in western Europe, it is reemerging in some countries in CEE and the CIS. For example, in Albania and Tadjikistan iodine deficiency is severe and in some other countries the prevalence of IDD is increasing.

The main public health strategy recommended to solve this problem is universal salt iodization. This means insuring that all the salt used by the food-processing industry, the mass catering sector, and households is iodized. In some countries there is no universal iodization and only table salt is iodized (Italy, Poland). This could be a problem, especially in western Europe where most of the population get their salt from processed food products and not from home-cooked food to which salt is added during cooking or eating. Therefore nutritional epidemiologists should monitor the source of a population's salt, and so iodine.

In addition to insuring that iodized salt is used in food processing and mass catering, policy-makers should consider the need to include a recommendation to feed iodized salt to animals (cows). In the Nordic countries and the UK iodine is given to cows in the form of 'salt-lick' and so milk and milk products supply around 40–70% of the iodine intake in humans. In addition to insuring elimination of iodine

deficiency, this has the added benefit of limiting the amount of salt consumed by the population.

In CCEE and the CIS, cardiovascular diseases are responsible for 68% of all premature deaths, while in the rest of Europe this figure is only 43%. Moreover, proportionally, cerebrovascular diseases are markedly higher in CCEE and CIS. Because of the strong link between high salt intake and cerebrovascular disease, the WHO recommends a salt intake of no more than 6 g per capita per day. In Italy a campaign on the role of salt in the diet was launched. This campaign aims to eliminate IDD through the promotion of iodized salt, but at the same time preventing hypertension and vascular diseases through limitation of salt intake.

Iron The other main nutrient likely to be deficient in European countries is iron. Iron deficiency was mentioned as a problem by one fourth (9) of 35 member states. Prevalence studies show that anemia is widespread in central Asia and the Caucasus. For example, in Kazakhstan and Azerbaijan up to 70% of children younger than 2 years of age have low hemoglobin levels. In addition, 20–30% of women of childbearing age also have low hemoglobin levels. These problems may not be related solely to iron deficiency in the diet. For example, it has already been mentioned that early introduction of cows' milk is a major cause of iron deficiency in the young. Anemia in adults is associated with the presence of iron-absorption inhibitors in the diet, such as tea and coffee, which is frequently consumed in eastern European countries; or lack of absorption enhancers, such as vitamin C from vegetables and fruit.

Folate Folate deficiency is associated with neural tube defects. Denmark, Poland, and the UK mentioned national strategies related to solving this problem. Folate may also play an important role in the prevention of coronary heart disease by helping to reduce levels of homocysteine. The main food sources of folate are Brussels sprouts, asparagus, spinach, broccoli, cabbage, cauliflower, parsnip, iceberg lettuce, beans, peas, and beef and yeast extracts – yet another reason why national food and nutrition polices and dietary guidelines should promote vegetable production and consumption.

Other micronutrients There may be situations when the only solution to solve micronutrient deficiency is by food fortification of bread or salt, for example (such as for iodine deficiency), or, more rarely, by supplements (during emergencies). However, generally, where possible, WHO promotes primary prevention strategies such as increasing vegetable consumption rather than advocating consumption of nutraceuticals, food supplements, or multivitamin tablets. Population strategies advocating the use of vitamin supplements are not the solution to micronutrient deficiencies in Europe. Vitamin supplements may have many disadvantages: side-effects; nutrient imbalances; toxicity; malabsorption; long-term dependence and lack of confidence in locally produced foods, and, finally, supplements are an unnecessary expense.

Appropriate Diets and Lifestyles

Countries of the CARs and the CIS must pay more attention to the link between diet and noncommunicable diseases (NCD), such as cardiovascular diseases, certain cancers, diabetes, hypertension, and obesity. Some of the country reports focused too much on deficiency of protein and micronutrients. In focusing on deficiency the link between premature mortality from NCD and a diet high in fat, salt, fatty red meat, and fatty/sugary foods and simultaneously low in vegetables and fruit is neglected.

Some countries in the CIS and central Asia mentioned the need to increase the teaching of dietology/dietetics. There is still a tendency in some countries to prescribe many different types of diets for different disorders. Many of these dietary prescriptions have not been scientifically proven. In addition, many of the individuals referred for dietary treatment are suffering from obesity, diabetes, heart disease, high blood pressure, and other conditions related to unhealthy lifestyles. These cases should be treated using the healthy nutrition principles developed for the population. Dental caries is also widespread in many European countries and there can be little doubt that this is related to frequent high intakes of sugar as well as poor oral hygiene.

Some countries, such as Albania, Israel, Kazakhstan, Latvia, Romania, Slovakia, the Republic of Moldova, Ukraine, and Uzbekistan made no mention of the link between diet and NCDs. Moreover, only one-third (12) of 35 member states mentioned obesity (Austria, Azerbaijan, Czech Republic, Denmark, Estonia, Hungary, Lithuania, Luxembourg, Malta, Poland, Turkey, and the UK), despite the fact that the prevalence of obesity is high and appears to be increasing in every country in Europe. For example, the Russian Federation has one of the highest prevalences, where 55% of the female population is overweight. Little mention of obesity probably reflects the lack of data on body weight and height and illustrates the need for countries to collect anthropometric data as part of their health information system.

Physical activity and obesity Obesity is a chronic medical problem caused by a combination of an

energy-dense diet leading to excess energy intake and lack of physical activity. There is abundant evidence that obesity is associated with a high risk of coronary heart disease, hypertension, diabetes mellitus, and gastrointestinal disorders. The risk of cancers of a number of sites (endometrial, renal, colon, gallbladder, and postmenopausal breast cancer) is also linked to obesity. As treatment of obesity is difficult, the need to readjust dietary energy intakes and/or physical activity permanently is essential. Other health advantages of high levels of physical activity include improved mental and psychological health.

Only 17% (six of 35) of member states made reference to the importance of increasing physical activity (Armenia, Estonia, Malta, Poland, Sweden, and the UK). Environments which support improved eating habits and more active living are needed. It is recommended that prevention efforts are focused on population-based public health strategies.

Conclusion

Cooperation and collaboration are needed throughout Europe and there is a need to share information and build alliances. In western Europe there exists the European Academy of Nutritional Sciences (EANS), the Federation of European Nutritional Sciences (FENS), and the Arbeitsgemeinschaft Ernährungsverhalten/Working Association for Nutritional Behavior (AGEV) and many countries have their Nutrition Society. In CIS and CCEE there seems to be a lack of coordination, and nutrition networks or societies should be established to facilitate sharing of information and developments in the area of food and nutrition.

See also: **Anemia (Anaemia)**: Iron-deficiency Anemia; **European Union**: European Food Law Harmonization; **Exercise**: Metabolic Requirements; **Folic Acid**: Properties and Determination; **Food Poisoning**: Statistics; **Food Safety**; **Infant Foods**: Milk Formulas; **Infants**: Breast- and Bottle-feeding; **Iodine**: Iodine-deficiency Disorders; **Obesity**: Epidemiology; **Quality Assurance and Quality Control**; ***Salmonella***: Properties and Occurrence; **World Health Organization**; **Obesity**: Epidemiology **World Trade Organization**

Further Reading

Adams JS and Lee G (1997) Gains in bone mineral density with resolution of vitamin D intoxication. *Annals of Internal Medicine* 127(3): 203–206.

Agra Europe (1997) *East Europe Agriculture and Food*. No. 179. London: Agra Europe.

Bates CJ (1997) Plasma total homocysteine in a representative sample of 972 British men and women aged 65 and over. *Clinical Journal of Nutrition* 51: 691–697.

Braga C (1997) Homocysteine as a risk factor for vascular disease. *Nutrition Research Newsletter* 16: 7–8.

Borgdorff MW and Mortarjemi Y (1997) *Surveillance of Foodborne Diseases: What are the Options?* WHO/FSF/FOS/97.3. Geneva: WHO.

Brenner BM (1982) Dietary protein intake and the progressive nature of kidney disease: the role of hemodynamically mediated glomerular injury in the pathogenesis of progressive glomerular sclerosis in ageing, renal absorption, and intrinsic renal disease. *New England Journal of Medicine* 307: 652–659.

CARAK newsletter (1997). Geneva: WHO.

Consumers in Europe Group (1994) *The Common Agricultural Policy: How to Spend £28 Billion a Year Without Making Anyone Happy*. London: Consumers in Europe Group.

FAO statistical databases (FAOSTAT DATA): http://apps.fao.org/cgi-bin/nph-dp.pl.

Health Education Authority (1996) *Folic Acid – What all Women should Know*. UK: Health Education Authority.

Health Education Authority (1997) *Folic Acid and the Prevention of Neural Tube Defects: A Summary Guide for Health Professionals*.

Hecht K and Steinman E (1996) *Improving Access to Food in Low-income Communities: An Investigation of Three Bay Area Neighborhoods*. San Francisco: California Food Policy Advocates.

Itoh R, Nishiyama N and Suyama Y (1998) Dietary protein intake and urinary excretion of calcium: a cross-sectional study in a healthy Japanese population. *American Journal of Clinical Nutrition* 67: 438–444.

James WPT (1997) *An Interim Proposal on the Creation of a UK Food Standards Agency*. London: FSA.

James WPT, Nelson M, Ralph A and Leather S (1997) The contribution of nutrition to inequalities in health. *British Medical Journal* 314: 1545–1549.

Käferstein FK and Clugston GA (1995) Human health problems related to meat production and consumption. *Fleischwirtschaft* 75(7) 75–77.

Leather S (1996) *The Making of Modern Malnutrition: An Overview of Food Poverty in the UK*. London: Caroline Walker Trust.

Marriott BM (1997) Vitamin D supplementation: a word of caution (editorial). *Annals of Internal Medicine* 127(3): 231–232.

Motarjemi Y and Käferstein F (1996) *International Conference on Nutrition: A Challenge to the Food Safety Community*. 1996 WHO/FNU/FOS/96.4. Geneva: WHO.

National Food Alliance (1995) *Food and Low Income: A Conference Report*. London: National Food Alliance.

National Research Council (1997) *Premature Death in the New Independent States*. London: National Academy Press.

Proceedings of Regional Conference on Elimination of IDD in Central and Eastern Europe, the Commonwealth of Independent States, and the Baltic States (1997) Munich, Germany.

Report of a Working Party to the Chief Medical Officer for Scotland (1993) *Scotland's Health – A Challenge to us*

all: The Scottish Diet. Scottish Office Home and Health Department.
Reports of the Scientific Committee for Food (1993) *Nutrient and Energy Intakes for the European Community*. Luxembourg: Commission of the European Communities, Directorate-General Industry.
Report of the Working Group on Diet and Cancer of the Committee on Medical Aspects of Food and Nutrition Policy (1998) *Nutritional Aspects of the Development of Cancer*. Department of Health report no. 48. London: Stationery Office.
Robertson A (1997) *Priorities for Eliminating IDD in CCEE and CIS*. Regional Conference on Elimination of IDD in Central and East Europe, the Commonwealth of Independent States, and the Baltic States. Munich, Germany.
SAFE Alliance (1997) *The Common Agricultural Policy Yesterday, Today and Tomorrow*. London: SAFE Alliance.
Third European Interdisciplinary Meeting and 20th Annual Scientific Meeting of AGEV (1997) *Public Health and Nutrition*. Abstract book. Copenhagen: AGEV/WZB.
USDA (1997) *Russian Food Consumption: Emerging Demand for Quality*. Newly Independent States and Baltics Update. Agriculture and Trade Report. Economic Research Service WRS-97-S1.
WHO (1990) *Diet, Nutrition, and the Prevention of Chronic Diseases: Report of a WHO Study Group*. Geneva: WHO Technical Report Series no. 797.
WHO (1996) *Guidelines for Strengthening a National Food Safety Programme*. WHO/FNU/FOS/96.2. Geneva: WHO.
WHO (1997) *Food Safety and Globalization of Trade in Food: A Challenge to the Public Health Sector*. WHO/FSF/FOS/97.8. Geneva: WHO.
WHO (1997) *Prevention and Management of the Global Epidemic of Obesity: Report of the WHO Consultation on Obesity*. WHO/NUT/NCD/97.2. Geneva: WHO.
WHO (1998) *Comparative Analysis of the Implementation of the Innocenti Declaration in WHO European Member States*. Copenhagen: WHO Regional Office.
WHO/FAO (1992) *World Declaration and Plan of Action for Nutrition*. International Conference on Nutrition, Rome: FAO/WHO.
WHO website: http://www.who.ch/ and Health for All: http://www.who.dk/www.who.dk/mainframe. htm.
Wilkinson RG (1996) *Unhealthy Societies; The Afflictions of Inequality*. London: Routledge.
Wiseman MJ (1987) Changing in renal function in response to protein restricted diet in type 1 (insulin-dependent)-diabetic patients. *Diabetologia* 30: 154–159.
World Cancer Research Fund in association with American Institute for Cancer Research (1997) *Food, Nutrition and the Prevention of Cancer: A Global Perspective*. Washington: American Institute for Cancer Research.
Wretlind A (1982) Standards for nutritional accuracy of the diet: European and WHO/FAO viewpoint. *American Journal of Clinical Nutrition* 36: 366–375.

NUTRITIONAL ASSESSMENT

Contents

Importance of Measuring Nutritional Status

F Fidanza, Universita degli Studi di Perugia, Perugia, Italy

Aims of Nutritional Status Assessment

Nutrition has been more and more implicated in various health problems. Consequently, improving nutritional status can play an important role in solving or preventing these problems.

Malnutrition – both undernutrition and overnutrition – is widespread both in single patients as well as in population groups. Surgical patients, for example, often suffer from protein-energy malnutrition, and subclinical blood levels of vitamins and minerals are present even in population groups of developed countries. (*See* **Malnutrition**: The Problem of Malnutrition.)

But scientific interest is now moving from deficiencies to a new role for some micronutrients in the maintenance of health and the prevention of degenerative diseases.

Efforts to obtain reliable information about the nutritional status of both patients and population

groups encounter many difficulties. There are several reasons for these problems, the most important of which is the body's ability to adapt to adverse conditions, so that changes in biochemical or functional characteristics become evident only after substantial impairment has occurred. Methods of assessment must therefore be both highly specific and highly sensitive.

The complexity of the interactions between dietary inadequacy, disease, and personal and environmental variables also makes it particularly difficult to determine whether health impairment can be linked specifically to diet and whether or not any nutritional deficiency is secondary to some other defects. These problems can be solved only through continuing basic research to expand knowledge of the patterns of biochemical, physiological, pathological, and behavioral responses to deficiencies or excess of nutrients and to improve methods for applying this knowledge in practical situations.

Nutritional surveys can be cross-sectional (i.e., provide data on prevalence of malnutrition or deficiency) or longitudinal (i.e., provide data on incidence of malnutrition or deficiency and association with other factors). The serial collection of data at population level is a basic part of surveillance or of any intervention program.

A nutritional survey can be very general and comprehensive, in which case it includes the analysis not only of the prevailing nutritional problems of a given population but also of the responsible factors and the consequences. On the other hand, a survey can also be limited to some specific nutrients or selected population or patient groups. The selection is made according the objectives of the study and the funds, personnel, and time available. (*See* **Dietary Surveys: Surveys of Food Intakes in Groups and Individuals.**)

Malnutrition, whether under- or overnutrition, develops in various steps. In the prepathogenic period information on dietary inadequacy can be obtained from food balance sheets at national level and dietary surveys at individual or group level. These can provide an indication of potential nutritional problems of the population or group but do not assess the problems.

Late in the prepathogenic period some early changes in the body nutrient reserves can be present, but methods available are not sensitive enough to assess them.

In the pathogenic period, before the emergence of clear clinical signs, body tissue nutrient levels are modified to such a degree that metabolic and/or functional alterations appear. Biochemical and functional tests are then appropriate for status assessment and/or diagnosis.

Once the clinical horizon has been reached, nonspecific clinical signs and symptoms are present. Anthropometric measurements can be of some use early at this stage, as can biochemical and functional tests.

When symptoms are present clinical nutrition assessment and studies of morbidity data are of great help. The final stage of the nutritional problem can be assessed from analysis of mortality rates and data from postmortem examination, but in this way only information of retrospective conditions is obtained.

Sampling Procedure

In population studies if all the subjects cannot be examined, a very careful sampling procedure has to be chosen, considering first study design and then sample size. Sampling details most appropriate for each study design can be found in epidemiology textbooks. (*See* **Epidemiology.**)

Collecting and Storing Samples for Biochemical and Immunological Tests

Once the population sample is selected it is necessary to consider in great detail the collection and storage of samples for biochemical and immunological tests.

The instructions for sample collection and preparation vary according to the specific assays to be performed. Only general rules are provided here.

Body tissues and excretions can rapidly deteriorate and/or be easily contaminated. In general, to obtain whole blood, serum, or plasma, appropriate syringes or vacuum collection tubes are used. If handled carefully, hemolysis can be avoided.

For some minerals and trace elements, in order to avoid contamination, one must use new plastic (polystyrene or polypropylene) syringes and collection tubes, and wear powder-free latex gloves when collecting or working with specimens.

Urine samples must be acidified and chilled on ice in polypropylene or polyethylene bottles.

Specimens of adipose tissue can be obtained by microbiopsy or needle aspiration by vacuum tube assembly.

The plasma or serum for protein determinations can be stored for 1 month at 2–6 °C and for several months at −20 °C.

Lipid fractions immediately separated from serum or plasma are stable for 1 year or more at −70 °C, if properly handled. Adipose tissue specimens for fatty acid determination can be stored at −20 °C without any other precaution up to 18 months.

For most vitamin determinations, plasma or serum is used and it can be stored at −20 °C for several months or, in a few cases, even for years. Samples should be protected from light for fat-soluble

vitamins and riboflavin analysis. For the enzymatic assays of erythrocyte thiamin, riboflavin, and vitamin B_6 heparinized blood should be stabilized with proper ACD solution (ACD solution is a stabilizer made of citric acid, sodium citrate and D-glucose) for storage at 4–6 °C for 10 days. For the erythrocyte thiamin pyrophosphate (TPP) assay the determination should be carried out within 2 h of drawing blood because of TPP instability. For vitamin C determination, heparinized blood, after stabilization with metaphosphoric acid, can be stored frozen, preferably at −70 °C, for no more than 3 weeks. For vitamin C determination by high-performance liquid chromatography (HPLC), heparinized blood specimens can be stored for a few weeks at −80 °C after addition of reduced glutathione within 20 min of collection. The addition of dithiothreitol increases stability of plasma samples to more than 1 year at −70 °C.

Acidified urine specimens for thiamin, riboflavin, and vitamin B_6 determinations are stable at −20 °C for 3 months. The specimens should be protected from light. Longer storage is possible for biotin and pantothenic acid.

For mineral and trace element determination serum, plasma, whole blood, or blood cells are used with no great problems for storage, but glass containers should be avoided.

For cellular studies of immunocompetence the blood must be rapidly processed; in some cases the testing of frozen cells was successful.

Recommended Methods for Nutritional Status Assessment

Table 1 is a synopsis of some recommended methods. For a detailed description of them as well as for other methods, see the Further Reading section.

These methods can be used for the whole population as well as for single patients. In the clinical setting, nutritional assessment and support are best performed by a nutritional support team, including clinician, nutritionist, and clinical chemist. If needed, advanced methods for the assessment of body composition, nitrogen balance, and some prognostic indices are available to detect patient malnutrition and to predict outcome (sepsis, wound dehiscence, death).

Data Processing and Calculation

The most important stages of data processing are: coding (including a preliminary quality control); data input; quality control data; data bank; data analysis (using parametric and nonparametric techniques); and reporting. Specialized books describe in detail the above stages. For each type of variable, specific calculation methods may be used. Only general rules will be provided here.

For variables not normally distributed (some anthropometric data, ferritin, etc.), a logarithmic transformation can be applied. In some cases the use of nonparametric tests is preferred.

For the use and interpretation of anthropometric measurements an Expert Committee Report of the World Health Organization is highly recommended. Technical framework and detailed guidance on the use and interpretation of anthropometric measurements in pregnant and lactating women, newborn infants, infants and children, adolescents, overweight and thin adults, and adults aged 60 years and over are provided. An extensive series of reference data is included in an annex.

Computer programs are now commonly used to test specific hypotheses. Because of the highly specialized skills required for proper processing and analysis of data, it is recommended that specialists in these fields take part in the planning and execution of evaluation. To improve the quality of this work it is advisable that these specialists are also involved in the data recording – an essential preliminary step of data processing. This means that they must be involved in the design of the study and in the preparation and testing of various forms and/or questionnaires needed for the study.

For data analysis, a data stratification by age groups and sex or any other meaningful stratification like altitude (for hemoglobin), socioeconomic status, and health/disease data, must often be used. Results may be expressed in terms of means and standard deviations or standard errors or as percentages of values (prevalence) above or below an arbitrary cut-off point defining adequacy. Anthropometric indices can be expressed in terms of Z-score (or standard deviation score). If the sample size is large enough, the percentiles or the frequency distributions are very useful. This data presentation facilitates comparison between surveys in different countries; it permits at any time the selection of the percentile or the range of percentiles of reference, i.e., acceptability, and it can represent a combination of the descriptive and interpretative ways for analyzing and presenting collected data.

The final step is the interpretation of data. In general the prevalence of malnutrition or deficiency is obtained using cut-off from a reference population. This reference population and the population under study are often from different areas or the criteria used to specify the reference population are not carefully defined. However, since most reference values are population-specific, it is recommended that data from a clinically healthy reference population of the

Table 1 Nutritional status methodology

Item	*Assessed variable*	*Method or apparatus*
Anthropometry		
Stature	Height	Anthropometer or stadiometer
	Sitting height	Anthropometer
	Length	Measuring table
Body mass	Weight	Beam scale
Circumferences	Mid upper arm circumference (MAC)	Flexible nonelastic tape
	Head	Flexible nonelastic tape
	Chest	Flexible nonelastic tape
	Waist	Flexible nonelastic tape
	Hip	Flexible nonelastic tape
	Thigh	Flexible nonelastic tape
Skinfold	Triceps (TSF)[a]	Skinfold calliper
	Subscapular	Skinfold calliper
	Thigh	Skinfold calliper
Diameters	Biacromial breadth	Anthropometer
	Biiliac breadth	Anthropometer
	Bicondylar breadth	Sliding calliper
	Bistyloid breadth	Sliding calliper
Derived indices	Height for age	Z-score
	Weight for age	Z-score
	Weight for height	Z-score
	Body mass index	Weight/height^2
	Arm muscle circumference	MAC - ($\pi \times$ TSF)
	Arm muscle area	(MAC - ($\pi \times$ TSF))^2/4π
	Waist/hip ratio	
	Waist/thigh ratio	
Body composition		
Atomic level	Carbon	Inelastic neutron scattering
	Nitrogen	Neutron activation analysis
	Others	Neutron activation analysis
Molecular level	Water	Isotope dilution technique or bioelectrical impedance analysis
	Osseous minerals	Dual-photon absorptiometry
	Protein as total body nitrogen	Neutron activation analysis
	Lipid	Hydrodensitometry
Cellular level	Body cell mass	Total body potassium (TBK) or exchangeable TBK
	Extracellular fluid	Dilutometry
	Extracellular solids	Indirect methods (total body Ca by neutron activation analysis)
Tissue system	Subcutaneous and visceral adipose tissue	Computed axial tomography
	Skeletal muscle mass	Indirect methods (24-h urinary creatinine or TBK)
	Nitrogen	Indirect method (neutron activation analysis)
Whole body	Body mass	See anthropometry
	Stature	See anthropometry
	Circumferences	See anthropometry
	Skinfolds	See anthropometry
	Densitometry	Underwater weight
Physical working capacity and physical fitness	Cardiorespiratory efficiency	Treadmill or bicycle ergometer
	Muscle strength	Dynamometers
	Motor performance	Several tests
Energy balance	Energy intake	Individual dietary surveys
	Energy expenditure	Indirect calorimetry
		Doubly labeled water
Lipid		
Essential fatty acids	Serum, erythrocyte membranes, subcutaneous fat tissues	Gas chromatography
Protein		
Transport proteins	Plasma or serum	Immunological
Fibronectin	Plasma	Laser nephelometry
Albumin	Plasma or serum	Dye-binding

Continued

Table 1 Continued

Item	*Assessed variable*	*Method or apparatus*
Creatinine	Urine	Colorimetric (autoanalyser)
		HPLC
3-Methyl histidine	Urine	Colorimetric or fluorometric
		HPLC
Somatomedin C (IGF-1)	Plasma or serum	Radioimmunoassay
Amino acids	Plasma	Automated ion exchange
		HPLC
Vitamins		
Vitamin A (retinol)	Plasma or serum[b,c]	HPLC
Carotenoids	Plasma or serum[c]	HPLC
Vitamin E (α-tocopherol)	Plasma or serum[c]	HPLC
25 (OH)-vitamin D	Serum	Competitive protein-binding
		HPLC
		RIA
Vitamin K_1 (phylloquinone)	Serum or plasma	HPLC
Thiamin	Erythrocytes	Transketolase activity[d]
	Whole blood	HPLC
	Urine[e]	Fluorometric
Thiamin pyrophosphate	Erythrocytes	HPLC
Riboflavin	Erythrocytes	Glutathione reductase activity[d]
	Urine[e]	HPLC
Flavin adenine dinucleotide	Whole blood	HPLC
Niacin	Whole blood	Microbiological
Niacin metabolites	Urine[e]	HPLC
Vitamin B_6	Erythrocytes	Aspartate transaminase activity[d]
Pyridoxal 5′ phosphate	Whole blood	HPLC
Pyridoxal 5′ phosphate	Plasma	Radioenzymatic
4-Pyridoxic acid	Urine[e]	HPLC
Folate (5-Me-THF)	Serum or plasma[e]	Competitive protein-binding
	Erythrocytes	Radioassay
Homocysteine	Plasma	HPLC
Vitamin B_{12} (cobalamins)	Serum or plasma	Competitive protein-binding
Methylmalonic acid	Serum or urine	GC-MS
Biotin	Whole blood	Microbiological
	Plasma/urine	Radioimmunoassay
Pantothenic acid	Whole blood	Microbiological
Vitamin C	Plasma/buffy coat layer	Spectrophotometric
	Whole blood/plasma	HPLC-micromethod
Minerals and trace elements		
Total Ca	Plasma	Colorimetric
Ionized Ca	Plasma	Selective electrode
Bone density	Metacarpal indices	Radiographic or dual-photon absorptiometry
Hydroxyproline	Urine	Colorimetric
Osteocalcin	Serum	Radioimmunoassay
Phosphorus	Plasma or serum	Spectrophotometric
Magnesium	Serum or urine	AAS
	Serum	Colorimetric
Ferritin	Serum or plasma	ELISA
		RIA
		IRMA
Iron	Serum	Spectrophotometric or AAS
Iron binding capacity	Plasma or serum	Colorimetric or radioactive
Protoporphyrin	Erythrocytes	Hemofluorometer
		Fluorometric
Transferrin receptor	Serum	ELISA
Hemoglobin	Blood	Spectrophotometric or electronic counter
Hematocrit	Blood	Special centrifuge or electronic counter
Zinc	Plasma or serum	AAS
	Leukocyte and leukocyte subsets	AAS
	Hair	AAS
	Taste acuity	Threshold of test solutions
Copper	Plasma or serum	AAS or colorimetric
Ceruloplasmin	Serum	Spectrophotometric

Continued

Table 1 Continued

Item	*Assessed variable*	*Method or apparatus*
Superoxide dismutase	Erythrocytes	Spectrophotometric
Selenium	Whole blood or plasma	AAS or fluorometric
Glutathione peroxidase	Erythrocytes, plasma or platelets	Spectrophotometric
Chromium	Serum	Graphite furnace AAS
Manganese	Plasma or serum	Graphite furnace AAS
Immunocompetence		
Lymphocyte count	Blood	Blood cell counter
Delayed cutaneous hypersensitivity	Skin	Recall antigens (multitest)
Complement C_3 and factor B	Serum	Radial immunodiffusion or nephelometric
Secretory IgA	Saliva or tears	Radial immunodiffusion or nephelometric
T-lymphocyte and subset percentage	Blood mononuclear cells	Fluorescence microscopy or cytofluorometric
Lymphocyte proliferation	Blood mononuclear cells	Cell harvester and β-counter
Clinical examination		Various forms
Psychometrics		Various tests

[a]Percent body fat can be computed if population-specific equations are available.
[b]Micromethod available.
[c]Measurable simultaneously.
[d]Micromethod and automation available.
[e]Short-term status.
TSF, triceps skinfold; HPLC, high-performance liquid chromatography; IGF-1, insulin-like growth factor 1; RIA, radioimmunoassay; GC-MS, gas chromatography–mass spectrometry; AAS, atomic absorption spectrometry; ELISA, enzyme-linked immunosorbent assay; IRMA, immunoradiometric assay.

same country or even of the same area, are collected. This is not necessary for nutritional surveillance, for which international references should be used.

In addition, for the interpretation of data, confounding factors must be taken into consideration, because they can alter the status and needs of some nutrients. Good examples are smoking habits, alcohol intake, drug use and abuse, physical activity, and consumption of toxic and antinutritive substances.

Use of Data

The main use of nutritional status data is to assess prevalence of malnutrition or deficiency in cross-sectional studies and/or the study of the association of malnutrition and its determinants in longitudinal studies.

An additional and important use of these data is for surveillance or intervention programs. In this case the data collection is often limited to the most relevant indicators of major risk for the population surveyed or the most vulnerable groups inside the population, such as pregnant and lactating women, children, adolescents, and the elderly. Nutrition-related risk factors of chronic disease have recently been included in the surveillance programs.

The assessment of nutritional status of single patients can be used for predicting outcome, and for establishing the need for, and monitoring the effect of, nutritional support. In this case the measurement of special biochemical markers of stress is highly recommended.

Limitations of Data

Most of the biochemical indices mentioned in Table 1 are static indices (measures of the concentration of a nutrient or its metabolites in a suitable biochemical matrix) that do not give information on the functions of the body. It would be desirable to measure organ functions instead (they depend on the adequate availability of a nutrient or the response to the regulatory process to maintain body stores). However, few sensitive and reliable functional tests are available and some of them are not specific for a single nutrient or not easily applicable in the field.

Other limiting factors are intraindividual variance and the lack of harmonization and standardization of methodology. As yet, there are no official methods for nutritional status assessment and there are too few quality controls and interlaboratory comparisons. There are currently few reference laboratories and official standards from international or national agencies. This makes comparison of data from different studies difficult.

The reliability and validity of the methods used are not always assessed. It is true that in some cases assessment of true validity is almost impossible. But concurrent or other types of validity are available. This is an area that needs much more consideration.

Also important factors which are not always assessed in nutritional epidemiology include sensitivity, specificity, and variability of indicators. The sensitivity can be diagnostic (the capacity to detect the earliest stages of deficiency) and biochemical (reflecting

every stage of the deficiency). The variability can be analytical (instrumental, preinstrumental), and biological (intraindividual, interindividual). The above items are limiting factors in data interpretation.

See also: **Dietary Surveys**: Measurement of Food Intake; Surveys of National Food Intake; Surveys of Food Intakes in Groups and Individuals; **Epidemiology**; **Malnutrition**: The Problem of Malnutrition; Malnutrition in Developed Countries

Further Reading

Fidanza F (ed.) (1991) *Nutritional Status Assessment – A Manual for Population Studies*. London: Chapman & Hall.

Gibson RS (1990) *Principle of Nutritional Assessment*. New York, NY: Oxford University Press.

Jelliffe DB and Jelliffe EFP (1989) *Community Nutritional Assessment*. Oxford: Oxford University Press.

Kinney JM, Jeejeeboy KN, Hill GL and Owens OE (eds) (1988) *Nutrition and Metabolism in Patient Care*. Philadelphia: Saunders.

Sauberlich HE (1999) *Laboratory Tests for the Assessment of Nutritional Status*, 2nd edn. Boca Raton: CRC Press.

van den Berg H, Heseker H, Lamand M *et al*. (1993) Flair concerted action no 10 status paper. *International Journal of Vitamin and Nutrition Research* 63: 247–316.

Wang ZM, Pierson RN and Heymsfield SB (1992) The five level model: a new approach to organizing body composition research. *American Journal of Clinical Nutrition* 56: 19–28.

WHO Expert Committee Report (1995) *Physical Status: The Use and Interpretation of Anthropometry*. Geneva: World Health Organization.

Anthropometry and Clinical Examination

J R Lustig and B J G Strauss, Monash Medical Centre, Clayton, Victoria, Australia

Introduction

A quick glance at a person will provide a gross estimation of nutritional status. Irrespective of training, we are all able to make these assessments: a Sumo wrestler is obese or overnourished, whilst a terminal AIDS patient is wasted and suffers from undernutrition. These assessments, however, are merely qualitative and contribute little to quantifying the nature and extent of these observations.

Table 1 Two-compartment model of body composition

Compartment	*Anthropometric technique used*
Fat mass/distribution	Body mass index
	Skinfold thicknesses
	Abdominal circumference
	Abdominal:gluteal ratio
Fat-free mass	Midarm circumference
	Midarm muscle circumference
	Midarm muscle area

Anthropometry comprises techniques that readily contribute to a more in-depth understanding of body composition and nutritional status, allowing the quantification of observations and the observation of changes with time.

Anthropometric methods are based on a model of body composition that consists of two distinct compartments: fat mass and fat-free mass (see **Table 1**). This two-compartment model of body composition defines the fat-free mass as a compartment consisting of body cells including skeletal muscle, extracellular water, the skeleton, and connective tissue, with the fat compartment consisting of the adipose tissue stores of the body, which has a very small cellular and water component. Anthropometric measurements can indirectly assess these two body compartments and thus provide an index of nutritional status. Alterations in body fat content generally are sensitive to changes in energy balance. Chronic malnutrition is reflected in changes to the protein stores found predominantly in skeletal muscle in the body.

Anthropometric techniques are rapid, portable, noninvasive, and inexpensive. The equipment required includes a tape measure, stadiometer (for measuring height), standardized weight scales, and skin-fold calipers. These techniques are frequently used in nutritional assessment and for monitoring changes in a diverse range of clinical settings. Such settings include tertiary referral centres, isolated rural practices, and population-based epidemiological studies. They are also used in gymnasia, infant welfare centers, schools, insurance assessments, sports clubs, and so on.

Anthropometric Assessment of Fat Mass

Body Mass Index (BMI)

$$\text{BMI} = \frac{\text{weight (kg)}}{\text{height (m}^2\text{)}}$$

Weight should be measured using standardized, good-quality weighing scales. The weight range of the scales should be known and should be appropriate for the subjects being weighed. This is particularly

important for the larger weights. Many instruments are not designed for weighing above 120 kg and are inaccurate when approaching higher weights.

Height is measured in the standing position using a stadiometer. Alternatively, a measuring stick attached to a vertical surface with some form of right-angled headboard can be used. Portable versions have been developed for field work. Shoes and socks should be removed and clothing kept to a minimum in order to allow assessment of posture. The subject should stand upright with the feet, buttocks, and shoulder blades in contact with the vertical surface or stadiometer. The head is positioned so that the Frankfurt plane (the line between the top of the pinna and the outer canthus of the eye) is horizontal. The subject is asked to take a deep breath, and the movable headboard is lowered until it is touching the crown of the head. The height is measured at maximum inspiration with the examiner's eyes level with the headboard to avoid any errors of parallax. The measurement is read to the nearest millimeter, the lower value being used if the level falls between two values.

Because of kyphosis or other postural problems, or, in the case of a patient being confined to bed, knee height has been used as a surrogate for estimating stature, and regression formulae have been developed to estimate height from knee height.

The BMI or Quetelet's index is the most commonly used weight/height ratio in adult populations. The BMI has a reasonable correlation with relative fat mass, as measured by body-density techniques. The correlation between BMI and relative fat mass, as measured by underwater weighing, was found to be 0.82 for a sample of women and 0.70 for a male sample. Another limitation on the usefulness of BMI in predicting body fatness in a given individual relates to the weight nominator. As weight is influenced by muscle, bone, and water as well as by fat, an individual with a well-developed musculo-skeletal system relative to height may have a BMI in the obese range but will not be overfat. BMI should thus be considered in the light of an individual's physical activity level.

Classifications of BMI have been developed, and these define relative levels of adiposity (see Table 2). Such a classification is based on surveys of Caucasian populations and may not apply to other ethnic groups. In addition, there is a good correlation between BMI and morbidity and mortality from obesity, e.g., heart failure, osteoarthritis, and hypertension.

The BMI can also be used as an index of wasting. Population studies have shown an increase in morbidity and mortality, and alterations in immune function as the BMI falls below 18.5, irrespective of the type of population, from undernourished groups, to those with clinical illnesses such as malignancies. The WHO has graded wasting on the basis of the BMI into mild (BMI, 18.5), moderate (BMI < 17.5) and severe (BMI < 16.5).

Table 2 Classifications for body mass index

Body mass index ($kg\,m^{-2}$)	*Health association*
< 16.0	Severe protein-energy malnutrition
16.0–16.9	Moderate protein-energy malnutrition
17.0–18.5	Mild protein-energy malnutrition
20.0–24.9	Apparently healthy, with low health risk
25.0–29.9	Overweight, with increased health risk
30.0–39.9	Obesity, increased macrovascular disease and diabetes mellitus
> 40.0	'Super obesity' with major morbidity

Abdominal:Gluteal Circumference Ratio

The ratio of the abdominal circumference to the maximal gluteal circumference defines the distribution of adipose tissue in the body. The waist is measured with the subject standing erect and undressed. The abdominal circumference is measured at the midpoint of the line between the rib or costal margin and the iliac crest in the midaxillary line. The maximal gluteal (buttock) circumference is also measured with the subject standing erect. It is important to define the sites measured and maintain consistency in these sites in order to compare sets of data.

A waist/hip ratio > 0.95 in males and > 0.85 females is consistent with abdominal obesity.

Abdominal fat distribution is associated with a range of adverse health consequences, including an increased risk of cardiovascular and cerebrovascular disease, impaired glucose tolerance, and hypertriglyceridemia, even when the BMI is in the 'healthy' range.

Skin-fold Thicknesses

Estimation of body fat content from skin-fold thicknesses assumes that measurements of subcutaneous fat represent total body fat. This assumption is not strictly true, as the thickness of subcutaneous fat does not necessarily reflect a constant proportion of total body fat. In addition, there are marked variations in the distribution of subcutaneous fat depending on age, sex, and race. None the less, this method provides useful information and is in routine use. Skin-fold thickness measurements may be taken at single or multiple sites. There is no universal agreement as to which skin-fold best represents total body fat. However, the triceps skin-fold is most frequently selected for a single-site assessment of body fat.

The measurement of the triceps skin-fold is taken at the midpoint of the upper arm. With the elbow bent to 90% the midpoint is located half-way between the acromium process of the shoulder and the olecranon process of the ulna. This point should be marked. The arm is then extended and allowed to hang loosely by the side. Using the thumb and forefinger, the skin-fold is grasped just above the midarm point. The calipers are applied at right angles to the skin at the midarm point and 1 cm in from the surface. The fold should remain held whilst the measurement is taken. Multiple skin-fold sites can be assessed and usually include both limb and trunk sites.

Body density can be estimated from skin-fold measurements. Using regression equations such as Durnin and Womersley's, skin-folds, sex, and age are used to derive body fat and fat-free mass. For example, for males aged 30–39 years:

1. $D\,(\mathrm{gm}^{-3}) = 1.1422 - (0 - 0544 \times \log_{10}$ (Skinfold sum [mm])), where D = density; skinfold sum = biceps + triceps + subscapular + iliac.
2. Percentage body fat or relative fat mass can then be calculated. Several equations have been derived. The Siri equation assumes that the density of fat is $0.900\,\mathrm{g\,cm}^{-3}$ and that the density of the fat-free mass is $1.100\,\mathrm{g\,cm}^{-3}$: % fat = $[(4.95/D) - 4.50] \times 100$.
3. Total body fat is derived as follows: total body fat (kg) = [body weight (kg) × % body fat]/100.
4. Fat-free mass (kg) = body weight (kg) − body fat (kg).

Both the skill of the operator and the character of the subcutaneous fat in an individual can influence the precision of skin-fold measurements. In general, the error of this method is approximately 5%. Depending on the equation used and the population studied, this may range from 3 to 9% body fat mass. Durnin and Womersley derived values for calculation of fat mass for all age groups in both men and women.

Anthropometric Assessment of Fat-Free Mass

Mid-arm Circumference

The upper arm contains both muscle and subcutaneous fat, so that measurement of the circumference of the midupper arm may be used as an index of nutritional status. In underdeveloped countries where the population is often malnourished and with little fat reserves, a change in this measurement can reflect total body protein stores. Serial measurement of the upper-arm circumference may be used to monitor nutritional intervention. Measurement of this circumference is described in the section on skin-fold thicknesses.

Midarm Muscle Circumference

In practice, midarm muscle circumference is used in preference to midarm circumference as a measurement which reflects total body protein stores. The upper-arm circumference comprises central bone surrounded by a layer of skeletal muscle and subcutaneous fat. The midarm muscle circumference, derived from the midarm circumference and the triceps skinfold thickness, takes into account an assessment of both fat and protein stores.

Small changes in muscle mass may not be detected by this technique. In addition, individual variations in the diameter of the humerus are not taken into account in this calculation. Midarm muscle circumference is derived using the following equation:

$$\text{Midarm muscle circumference} = \text{midupper arm circumference} - (\pi \times TSF),$$

where TSF = triceps skinfold thickness (in cm).

Midarm Muscle Area

Midarm muscle area provides a higher correlation with body protein stores than does either of the midarm circumferences. Because it takes into account a two-dimensional assessment, it reflects more accurately small changes in muscle mass. None the less, one appraisal of the equation suggests that it may underestimate significantly the degree of muscle wasting. Midarm muscle area may be calculated using the following equation:

$$\text{arm muscle area} = [C - (\pi \times TSF)]^2/4\pi,$$

where C = midupper arm circumference, and TSF triceps skinfold thickness (in cm).

The coefficient of variation for the measurement of midarm muscle area has been estimated as approximately 7% when made by trained operators. This is therefore not a sensitive enough method by which to detect small changes in muscle mass.

Conclusions

Anthropometry provides a rapid methodology for the indirect quantification of fat and fat-free mass and, thus, the body compositional aspects of nutritional status. The techniques require little in the way of equipment and are readily portable and inexpensive. Although they are insensitive to small changes in nutritional status, they have a useful role to play in assessment of nutritional status and for monitoring

progress over time. In addition, these techniques are useful in population-based field research of nutrition-related disorders.

See also: **Body Composition**; The Problem of Malnutrition; **Nutritional Assessment**: Importance of Measuring Nutritional Status; **Obesity**: Etiology and Diagnosis

Further Reading

Bjorntorp P (1987) Classification of obese patients and complications related to the distribution of surplus fat. *American Journal of Clinical Nutrition* 45: 1120–1125.

Durnin JVGA and Womersley J (1974) Body fat assessed from the total body density and its estimated skinfold thickness: measurements on 481 men and women aged from 16 to 72 years. *British Journal of Nutrition* 32: 77–97.

Forbes GB (1990) The abdomen:hip ratio. Normative data and observations on selected patients. *International Journal of Obesity* 14: 149–157.

Gibson RS (1990) *Principles of Nutritional Assessment*. London: Oxford University Press.

Heymsfield S, McManus CB, Smith J, Stevens V and Dixon DW (1982) Anthropometric measurement of muscle-mass: revised equations for calculating bone-free arm muscle area. *American Journal of Clinical Nutrition* 36: 680–690.

James WPT, Ferro-Luzzi A and Waterlow JC (1988) Definition of chronic energy deficiency in adults. *European Journal of Clinical Nutrition* 42: 909–981.

Lapidus L, Bengtsson Q, Larsson B *et al*. (1984) Distribution of adipose tissue and risk of cardiovascular death: a 12 year follow-up of participants in the population of women in Gothenburg, Sweden. *British Medical Journal* 289: 1257–1261.

Larsson L, Svardsudd K, Welin L *et al*. (1984) Abdominal adipose tissue distribution and risk of cardiovascular disease and death: a 13 year follow-up of participants in the study of men born in 1913. *British Medical Journal* 288: 1401–1404.

Lohman TG (1981) Skinfolds and body density and their relation to body fatness: a review. *Human Biology* 53: 181–225.

Lohman TG (1992) Advances in body composition assessment. *Current Issues in Exercise Science*. Monograph No. 3. Champaign, IL: Human Kinetics.

Siri WE (1961) Body composition from fluid spaces and density: Analysis of methods. *Techniques for Measuring Body Composition*, pp. 223–244. Washington, DC: National Academy of Sciences, National Research Council.

Smalley KJ, Knerr AN, Kendrick ZV, Colliver JA, and Owen OE (1990) Reassessment of body mass indices. *American Journal of Clinical Nutrition* 52: 405–408.

Physical Status. The Use and Interpretation of Anthropometry. (1995) WHO Technical Report 854. Geneva: WHO.

Biochemical Tests for Vitamins and Minerals

A V Lakshmi and M S Bamji[a], National Institute of Nutrition, Hyderabad, AP, India

Introduction

Increasing attention is being given to the definition of marginal malnutrition since the elimination of clinical signs of deficiency *per se* is not sufficient for optimal health. Biochemical measurements on easily available body fluids represent the most objective assessment of nutritional status and often provide subclinical information. It also provides information on micronutrient toxicity. During the development of any deficiency disease, biochemical changes precede clinical symptoms and hence biochemical tests help to identify the disease at the subclinical stage. They also help to confirm the clinical diagnosis.

An ideal biochemical test should be specific, sensitive, and indicative of tissue depletion at an early stage but not immediate dietary intake. However, we do not have an ideal biochemical test for most nutrients and the method selected depends on the situation in which it is applied. Often, more than one test is required. Classification of an individual as severely deficient, marginally deficient, normal, or intoxicated may require different tests. In this chapter biochemical tests used for the nutritional status assessment of fat- and water-soluble vitamins and of a few minerals – sodium, potassium, calcium, iron, iodine, zinc, copper, and selenium are discussed.

Types of Laboratory Tests

Laboratory tests for the assessment of vitamin or mineral nutritional status can be placed in the following categories:

1. Measurement of excretion of nutrient or its metabolite in urine, without or after a bolus load (load return test)
2. Measurement of nutrient level and/or its active metabolites in blood
3. Assay of activity of a vitamin-dependent enzyme and its *in vitro* stimulation with corresponding coenzyme in the blood. (*See* **Coenzymes**.)
4. Measurement of rise in the concentration of a metabolite in the blood or urine (resulting from inadequate intake of the vitamin) after administering a load of the precursor

[a]Current address: Dangoria Charitable Trust, 1–7–1074, Hyderabad 500020, India.

Measurement of excretion of vitamin or its metabolite in urine can be carried out on a 24-h sample, random sample, or in the first voided morning sample of urine. Although a 24-h specimen provides the most accurate estimate, it is difficult to collect it in the free-living population. Measurements in random samples or a first voided urine sample, and expressing the value per gram of creatinine, may be useful for population studies but are not very reliable for individual assessment.

The limitation regarding the interpretation of urinary data is that it is influenced by immediate dietary intake and may not reflect tissue stores. Load return tests may yield better information regarding the state of tissues but they are seldom used now. Certain physiological and pathological conditions, as well as ingestion of drugs, influence urinary excretion of certain vitamins and minerals.

Either whole blood or different components such as plasma, erythrocytes, leukocytes, and platelets has been used for the assessment of vitamin and mineral status. The choice of the blood components is decided by the concentration of the nutrient and its sensitivity to vitamin or mineral depletion. High-performance liquid chromatography (HPLC) technique has revived interest in these measurements due to improved sensitivity. (*See* **Chromatography**: High-performance Liquid Chromatography.)

Details regarding enzymatic tests and metabolite tests will be discussed while dealing with the individual vitamins.

Guidelines for Interpretation of Values

Availability of standard reference material (serum and erythrocyte hemolysates), use of standardized procedures, and strict quality control measures are necessary for arriving at interpretative guidelines. Since reference quality control samples are not available for many vitamins, and different laboratory procedures are adopted by various workers, only tentative guidelines based on the values obtained with apparently healthy, well-nourished controls are available at present for interpretation of data. (*See* **Vitamins**: Determination.)

Fat-soluble Vitamins

Vitamin A

Serum vitamin A level is not a sensitive indicator of vitamin A stores. The levels are however indicative of vitamin A status when the stores are either fully replete or totally deplete and in the absence of infection. Serum levels below 0.35 $\mu mol\,l^{-1}$ are usually associated with clinical signs of deficiency. Serum vitamin A levels above 0.70 $\mu mol\,l^{-1}$ are considered satisfactory and reflect adequate liver reserves (**Table 1**).

Table 1 Tentative guidelines for the interpretation of biochemical parameters of vitamin A status

Biochemical measurement	*Acceptable*	*Medium risk*	*High risk*
Serum ($\mu mol\,l^{-1}$)	> 70	35–70	< 35
Liver ($\mu mol\,g^{-1}$)	> 0.70	0.17–0.70	< 0.17
Relative dose–response (%)	< 20	> 20	
Modified relative dose–response (DR/R)	< 0.03	> 0.03	

DR, 3,4-didehydroretinol; R, retionol.

Marginal vitamin A status can be assessed by the relative dose–response (RDR) test or the modified relative dose–response (MRDR) test. The RDR test consists of determination of the plasma levels of vitamin A at baseline and 5 h after the administration of a small oral or intravenous dose of retinylpalmitate or retinyl acetate. An increase of more than 20% is indicative of inadequate hepatic reserves.

In the MRDR test, 3,4-didehydroretinol (vitamin A_2) is administered instead of retinol. This test requires only one blood sample for analysis. The ratio of dehydroretinol to retinol is used to assess liver vitamin A concentrations (Table 1).

Use of deuterated vitamin A for measuring the total body pool by means of isotope dilution is currently being refined.

Toxicity

High plasma levels of retinyl esters in the presence of high plasma levels of vitamin A is a sign of vitamin A toxicity. In conditions of normal vitamin A status, the plasma levels of retinyl esters is about 5%.

Spectrophotometric, HPLC fluorometric, and calorimetric techniques are used for measuring plasma vitamin A level. (*See* **Retinol**: Properties and Determination.)

Vitamin D

Measurement of serum 25-hydroxycholecalciferol, the major circulating metabolite of vitamin D, is the best indicator of vitamin D status. In healthy adults its concentration in serum varies from 20 to 130 $nmol\,l^{-1}$ and less than 12.5 $nmol\,l^{-1}$ is generally considered inadequate. It can be measured by HPLC or by competitive protein-binding assay. (*See* **Cholecalciferol**: Properties and Determination.)

Vitamin E

Vitamin E status is normally assessed by measuring levels of the vitamin in plasma or serum and other

tissues (red cells, platelets, adipocytes) by spectrophotometric, fluorometric, or HPLC methods. Since the levels of vitamin E and lipids in the serum are directly correlated, their ratio is a better measure of status than serum vitamin E alone (Table 2). The ratio of plasma vitamin E to cholesterol + triglyceride or vitamin E to cholesterol has also been reported to be as good as that of vitamin E to total lipids. (*See* **Tocopherols**: Properties and Determination.)

In vitro hemolysis of red blood cells in the presence of hydrogen peroxide, quantitating the formation of malondialdehyde from hydroperoxides generated by the *in vitro* peroxidation of polyunsaturated fatty acids or erythrocytes exposed to hydrogen peroxide is considered a functional test for assessing vitamin E status. The latter tests, based on quantitation of malondialdehyde formed, are considered to be more sensitive than the hemolysis test.

Vitamin K

In the past, vitamin K status has been assessed by measuring prothrombin time and other clotting times. These measures are however not sensitive to detect subclinical vitamin K deficiency since clotting time does not change until the concentration of prothrombin decreases by about 50%.

Vitamin K is needed for the γ-carboxylation of certain glutamic acid residues in proteins. During vitamin K deficiency, the vitamin K-dependent proteins occur in an undercarboxylated form. Milder forms of vitamin K deficiency can be assessed by measuring urinary γ-carboxyglutamic acid excretion or circulating levels of undercarboxylated prothrombin and osteocalcin. The last parameter is considered to be a very sensitive marker of vitamin K status. Plasma level of phyloquinone has also been used as a measure of vitamin K status. Plasma phyloquinone and urinary γ-carboxyglutamic acid can be determined by HPLC separation followed by fluorometric detection. Undercarboxylated prothrombin and osteocalcin can be determined by radioimmunoassay. (*See* **Immunoassays**: Radioimmunoassay and Enzyme Immunossay.)

Thiamin

The biochemical tests used for the assessment of thiamin nutritional status include measurement of: (1) erythrocyte transketolase activation coefficient (ETK-AC); (2) urinary excretion of thiamin; and (3) measurement of whole blood or erythrocyte thiamin pyrophosphate (TPP) concentration.

TPP is the coenzyme for transketolase and determination of ETK-AC is the most widely accepted procedure for assessing thiamin status. Activity of this enzyme is measured in erythrocyte hemolysate without (basal) and with a saturating amount of TPP added *in vitro* (stimulated). The ratio of stimulated to basal activity is called the activity coefficient (ETK-AC). Basal activity falls and ETK-AC increases in thiamin deficiency (Table 2). Activity can be measured by determining the rate of disappearance of pentose or appearance of hexose. However interpretation of the results is complicated under certain disease conditions (some cancers, uremia, neuropathy, and diabetes) and drug treatment (diuretics, antacids) since ETK activity is affected under these conditions. (*See* **Enzymes**: Uses in Analysis.)

Normal adults receiving adequate dietary thiamin excrete more than 0.3 μmol per 24 h. Urinary thiamin can be measured by fluorometric method. (*See* **Thiamin**: Properties and Determination.)

Measurement of blood or erythrocyte TPP by HPLC appears to be a more sensitive and promising indicator of thiamin status and needs further verification.

Riboflavin

Measurement of urinary excretion of riboflavin, red blood cell riboflavin, and erythrocyte glutathione reductase activation coefficient (EGR-AC) are used for the biochemical assessment of riboflavin nutritional status.

The urinary excretion method is useful for population surveys. On an adequate dietary intake of riboflavin, normal adults excrete more than 0.21 μmol g^{-1} of creatinine. Riboflavin concentration

Table 2 Tentative guidelines for interpretation of blood vitamin levels and enzyme tests

Parameter measured	*Acceptable*	*Medium risk*	*High risk*
Serum vitamin E (μmol l^{-1})	> 16.2	11.6–16.2	< 11.6
Serum vitamin E total lipid ratio	> 0.8		< 0.8
ETK-AC	< 1.15	1.15–1.25	> 1.25
EGR-AC	< 1.2	1.2–1.4	> 1.4
Serum folate (nmol l^{-1})	> 11.0	6.7–11.0	< 6.7
Erythrocyte folate (nmol l^{-1})	> 360	315–359	< 315
Plasma ascorbate (μmol l^{-1})	17–85	11–17	< 11
Leukocyte ascorbate (μmol 10^{-6} cells)	113–300		

ETK-AC, erythrocyte transketolase activation coefficient; EGR-AC, erythrocyte glutathione reductase activation coefficient.

in urine can be measured by fluorometric or microbiological assay or HPLC. (*See* **Riboflavin**: Properties and Determination.)

Red cell riboflavin level may be a better index of tissue status than urinary riboflavin. Fluorometric or HPLC technique can be used to estimate red cell riboflavin and the value can be expressed per gram hemoglobin or packed cell volume. However this test is not commonly used because tissue saturation is difficult to define and the value varies between individuals. Measurement of red cell flavin adenine dinucleotide (FAD) level using HPLC may improve the index.

At present, the EGR-AC test is the most accepted biochemical method for the assessment of riboflavin status. FAD, the coenzyme form of riboflavin, is necessary for the activity of glutathione reductase. The activity of this enzyme is measured in erythrocyte hemolysate or blood hemolysate without (basal) and with added FAD (stimulated). The ratio of stimulated to basal activity is called the activation coefficient (EGR-AC). This ratio increases in riboflavin deficiency (Table 2). EGR activity is assayed by rate reaction techniques. However the EGR-AC test fails to measure riboflavin status when glucose 6-phosphate dehydrogenase deficiency is also present.

During respiratory infections, urinary and blood levels of riboflavin are raised due to mobilization of the vitamin from the liver. There is a transient reduction in EGR-AC values also. Thus assessment of riboflavin status during such infection is difficult.

Niacin

Measurements of the urinary excretion of N^1-methylnicotinamide (NMN) and its oxidation product, N^1-myethyl-2-pyridone-5-carboxamide (2-pyridone) have been used to determine niacin status. Normal adults excrete more than 17 μmol NMN per day, whereas during deficiency its excretion is less than 5.8 μmol per day.

A more recent study suggested that erythrocyte nicotinamide adenine dinucleotide (NAD) concentration may serve as a sensitive indicator of niacin status and that a ratio of erythrocyte NAD to NAD phosphate (NADP) < 1.0 may suggest a risk of developing niacin deficiency. (*See* **Niacin**: Properties and Determination.)

Vitamin B_6

Several direct and indirect measures are used for the assessment of vitamin B_6 status.

Measurement of the plasma pyridoxal phosphate (PLP) level, the coenzyme form of the vitamin, is the most commonly used method and a concentration of 30 nmol l^{-1} is considered to be adequate. However, proper interpretation of plasma PLP data must take into account the factors which affect this value. High protein intake, pregnancy, age, smoking, and an increase in alkaline phosphatase activity are known to reduce PLP concentration. The measurement of both PLP and pyridoxal is suggested to be a better index of vitamin B_6 status during pregnancy. Pyridoxal phosphate levels in plasma can be measured by enzymatic or HPLC techniques.

The second direct measure is urinary excretion of 4-pyridoxic acid. It correlates well with dietary vitamin B_6 intake and this parameter is useful for population surveys. An excretion of > 3.0 μmol per day is considered to be associated with adequate vitamin B_6 status and it can be measured by fluorometric method. (*See* **Vitamin B_6**: Properties and Determination.)

Functional tests for assessing vitamin B_6 status include measurement of PLP-dependent transaminases, such as erythrocyte aspartate aminotransferase (EAST) and alanine aminotransferase (EALT) and their activation coefficient (AC), and tryptophan or methionine load test. They reflect tissue levels of PLP.

EAST or EALT activity is measured without and with added PLP to give an activation coefficient (EAST-AC or EALT-AC) – an enzymatic test analogous to that used for thiamin and riboflavin. Although widely used for assessing vitamin B_6 status, the interpretation of EAST-AC value is not standardized. Values ranging from 1.7 to 2.0 have been used to differentiate between adequacy and deficiency. The EALT-AC value suggested for adequacy of vitamin B_6 status is less than 1.25. These two enzymes can be assayed by colorimetric or rate reaction technique.

The tryptophan load test has been used in population surveys as a functional index of B_6 status. An inadequate status results in an increased excretion of tryptophan metabolites such as kynurenine and xanthurenic acid, since vitamin B_6-dependent enzyme kynureninase is required for their metabolism. This test is usually carried out after administering an oral load of 2–5 g tryptophan and measuring the urinary excretion of xanthurenic acid in a 6–9-h period. However excretion of tryptophan metabolites is influenced by factors like protein intake and pregnancy.

Folic Acid and Vitamin B_{12}

Serum as well as red blood cell folate reflect folate status. While serum folate reflects the dietary intake and readily available tissue reserves, red blood cell folate reflects long-term folate status. Microbiological or competitive protein-binding radioassay is used to measure folate levels in serum and red blood

cells. The guidelines for interpreting these values are given in Table 2. (*See* **Folic Acid**: Properties and Determination.)

The metabolism of histidine to glutamic acid is impaired in folate deficiency, leading to an increase in urinary excretion of formiminoglutamic acid (FIGLU). This occurs particularly after a loading dose of histidine. Normal adults excrete less than 35 mg FIGLU in 24 h after 5 g histidine load.

Deoxyuridine Suppression Test

In folic acid deficiency the activity of the enzyme thymidilate synthalase is impaired. This enzyme catalyzes the conversion of deoxyuridine to thymidine. In this test bone marrow cells or peripheral blood lymphocytes are preincubated with nonradioactive deoxyuridine and then with ^{3}H thymidine. The amount of radioactivity incorporated into the DNA of the cells is increased in folate as well as in vitamin B_{12} deficiency. These two deficiencies can be differentiated by *in vitro* incubation with respective vitamins.

Hypersegmentation of circulating neutrophils is a morphological change that occurs in folate deficiency due to slowed DNA synthesis. There are studies to suggest that this change occurs when plasma folate levels drop lower than 10 nmol l^{-1}.

Measurement of plasma homocysteine level may also indicate the adequacy of folate and, to a lesser extent, vitamin B_{12} status. Vitamin B_6 deficiency also raises plasma homocysteine concentration.

Vitamin B_{12} status is generally assessed by measuring its concentration in serum by microbiological or competitive protein-binding radioassay. Radioassay using pure intrinsic factor measures true vitamin B_{12}, whereas microbiological assay measures a few analogs of vitamin B_{12} as well. The lowest acceptable serum level using microbiological assay is 150 pmol l^{-1}, whereas for the radioassay specific for cobalamine, serum level < 75 pmol l^{-1} is considered deficient. Measurement of plasma vitamin B_{12} bound to transcobalamine II is considered to be more sensitive than total vitamin B_{12} for the assessment of marginal vitamin B_{12} deficiency.

The urinary or plasma methylmalonic acid level, preferably after an oral load of valine (its precursor), is a metabolic test for detecting vitamin B_{12} deficiency. Normal adults excrete less than 100 µmol in 24 h.

Pantothenic Acid

Urinary excretion of pantothenic acid or blood levels of the vitamin can be used to assess pantothenic acid status. Adults taking an adequate amount of dietary pantothenate excrete between 9 and 24 µmol of the vitamin per gram creatinine and the blood levels range between 1.2 and 1.8 µmol l^{-1}. Pantothenic acid levels in blood and urine can be measured by microbiological or radioimmunoassay. (*See* **Pantothenic Acid**: Properties and Determination.)

Biotin

Biotin nutritional status can be assessed by using the following sensitive biochemical parameters: (1) urinary excretion of organic acids such as 3-hydroxyisovaleric, 3-hydroxypropionic, and methylcitric acids, or of biotin and its metabolite bisnorbiotin; (2) assay of the activity of biotin-dependent enzyme propionyl coenzyme A carboxylase with and without *in vitro* addition of biotin in lymphocytes; and (3) measurement of plasma biotin level. Plasma biotin is a less sensitive indicator compared to the other two methods. The normal reported range for plasma biotin is 140–356 pmol l^{-1}. Biotin levels in blood and urine can be determined by microbiological or isotopic dilution assay. (*See* **Biotin**: Properties and Determination.)

Ascorbic Acid

Plasma and leukocyte ascorbate levels indicate ascorbic acid status. The simplicity and reliability of plasma ascorbic acid determination make it preferable for identifying individuals at risk of deficiency due to chronic low intake of the vitamin. However, plasma ascorbic acid levels may be less useful for defining the ascorbic acid status of individuals with adequate or high intake. Leukocyte ascorbic acid concentration reflects more accurately tissue stores of vitamin than plasma level.

Plasma and leukocyte ascorbic acid levels are determined by HPLC or colorimetric techniques. Thresholds for interpretation of ascorbic acid status are given in Table 2. (*See* **Ascorbic Acid**: Properties and Determination.)

Mineral Nutrition

In this section, biochemical methods used for the nutritional assessment of a few macro- and micromineral elements are given. However, the list is not comprehensive. Only those elements for which a laboratory test can give meaningful data from the nutritional point of view are included.

Sodium and Potassium

Dietary deficiency of sodium and potassium cations might occur after increased losses under certain conditions such as vomiting and diarrhea. Their status is generally assessed by measuring serum or plasma levels by atomic emission flame photometry or

atomic absorption spectrophotometry (Table 3). (*See* **Sodium**: Properties and Determination.)

Calcium

In blood, calcium is present almost entirely in the plasma. As ionized calcium in the plasma is under strict homeostatic control ($1.2\,mmol\,l^{-1}$), it cannot serve as a sensitive parameter to assess nutritional deficiency of calcium. (*See* **Calcium**: Properties and Determination.)

Bone density is an indirect measure of calcium status, since 99% of it is present in the bones. It can be measured by radiographic, single- or dual-photon absorptiometry techniques.

Iron

Iron status can be assessed by three types of biochemical tests, reflecting different stages of deficiency. First, bone marrow iron and serum ferritin concentrations indicate iron stores. Serum ferritin levels less than $12\,\mu g\,l^{-1}$ suggest depletion. As an acute-phase reactant, serum ferritin is elevated by acute and chronic infections, inflammatory diseases, malignancies, and liver disorders.

In the second stage of iron deficiency, referred to as iron-deficiency erythropoiesis, the supply of iron to the erythroid marrow is reduced. This can be assessed by the transferrin saturation index. An increase in red cell protoporphyrin level also indicates restricted iron supply to the developing red cell (Table 4).

Recent studies suggest that serum transferrin receptor is a sensitive and quantitative index of tissue iron status (Table 4). The levels are elevated in iron-deficiency anemia. Inflammation or liver disease does not affect serum transferrin receptor levels.

The third stage of iron deficiency is anemia, which can be assessed by hemoglobin level or hematocrit. Hemoglobin levels less than $13\,g\,dl^{-1}$ of blood for adult males and $12\,g\,dl^{-1}$ blood for adult females are considered deficient, whereas during pregnancy hemoglobin concentration less than $11\,g\,dl^{-1}$ is considered to be anemia. (*See* **Anemia (Anaemia)**: Iron-deficiency Anemia.)

Iodine

Urinary excretion of iodine reflects the dietary intake. An excretion of more than 100 μg per day suggests an adequate dietary intake of iodine. It can be measured by the kinetic method, based on iodide-catalyzed reduction of ceric to cerus by arsenic. (*See* **Iodine**: Properties and Determination.)

The only known requirement of iodine in humans is in the synthesis of thyroid hormones. Measurements of serum levels of thyrotropic hormone or triiodothyronine and thyroxine are sensitive functional tests for iodine status.

Zinc

Zinc levels in plasma, urine, hair and cellular compartments of blood, activity of zinc-dependent enzymes such as carbonic anhydrase in red cells and plasma levels of zinc-binding protein metallothionin have been used as indices of zinc nutrition status.

Plasma zinc appears to be the most widely used parameter. It decreases in cases of severe and moderate deficiency of zinc. Plasma zinc concentrations are susceptible to a number of pathophysiological influences. Since erythrocytes and hair have a longer half-life, their zinc content does not indicate acute zinc deficiency. Zinc levels in cells with rapid turnover (neutrophils and monocytes) can be used to detect mild zinc deficiency. Serum thymulin level without and with *in vitro* addition of zinc has been suggested as a sensitive method to assess mild zinc deficiency. Atomic absorption spectrometry is used for the analysis of zinc in urine and blood. Table 5 gives the normal range of tissue zinc. (*See* **Zinc**: Properties and Determination.)

Table 3 Normal ranges of sodium, potassium, copper, and selenium in blood

Parameter	*Range*
Plasma sodium ($mmol\,l^{-1}$)	136–145
Plasma potassium ($mmol\,l^{-1}$)	3.5–5.0
Plasma copper ($mg\,l^{-1}$)	0.8–1.75
Serum ceruloplasmin ($\mu mol\,l^{-1}$)	1.7–2.9
Plasma selenium ($\mu g\,l^{-1}$)	60–120
Erythrocyte selenium ($\mu g\,l^{-1}$)	90–190

Table 4 Assessment of iron status (commonly used cut-off values)

Parameter	*Value*
Serum ferritin ($\mu g\,l^{-1}$)	12
Transferrin saturation (%)	16
Red cell protoporphyrin ($\mu mol\,l^{-1}$)	1.24
Serum transferrin receptor ($mg\,l^{-1}$: normal range)	2.5–8.5
Hemoglobin ($g\,l^{-1}$)	
Men	130
Women	120

Table 5 Normal range of tissue zinc

Parameter	*Range*
Plasma ($mg\,l^{-1}$)	0.7–1.4
Erythrocyte ($mg\,l^{-1}$)	10–14
Urine (μg per day)	400–600
Leukocytes (μg per 10^{10} cells)	80–130
Hair ($\mu g\,g^{-1}$)	124–320

Copper

Copper status is commonly assessed by estimating serum copper concentration or by assaying the activity of the copper-dependent enzyme ceruloplasmin in serum (Table 3). Serum copper levels tend to be higher in women. The values are influenced by certain physiological and pathological conditions. About 95% of copper in serum is in ceruloplasmin, which is an acute-phase reactant protein. Erythrocyte copper-zinc superoxide dismutase has been suggested to be a better index of copper status than serum copper or ceruloplasmin.

Serum copper can be estimated by atomic absorption spectrometry. (*See* **Copper**: Properties and Determination.)

Selenium

Selenium status can be assessed by measuring its levels in urine, plasma, erythrocytes, and blood or by assaying the selenium-dependent enzyme glutathione peroxidase in erythrocytes or platelets.

Plasma selenium is a sensitive index of short-term changes until the plasma level reaches a plateau. Erythrocyte selenium and glutathione peroxidase activity indicates long-term changes because of the long life span of these cells.

The commonly used techniques for measuring selenium in biological fluids include fluorimetry, atomic absorption spectrometry, and mass spectrometry. Table 3 gives the frequently reported normal ranges in plasma and erythrocytes. (*See* **Selenium**: Properties and Determination.)

Selenium is an integral part of glutathione peroxidase and measurement of the activity of this enzyme is a functional test for the assessment of selenium status. The minimal erythrocyte selenium level required to reach a plateau of glutathione peroxidase activity has been estimated to be about 141 $\mu g\,l^{-1}$.

Conclusions

The value of biochemical tests can be improved with a better understanding of their limitation and their relationship to functional consequences. What level of biochemical insult can an individual accommodate without any ill effects? Can generalized interpretative guidelines be applied to all population groups or are there substantial variations? Despite these dilemmas, biochemical tests have an important place in nutritional diagnosis.

See also: **Ascorbic Acid**: Properties and Determination; **Biotin**: Properties and Determination; **Calcium**: Properties and Determination; **Copper**: Properties and Determination; **Folic Acid**: Properties and Determination; **Iodine**: Properties and Determination; **Niacin**: Properties and Determination; **Riboflavin**: Properties and Determination; **Selenium**: Properties and Determination; **Sodium**: Properties and Determination; **Thiamin**: Properties and Determination; **Tocopherols**: Properties and Determination; **Vitamin B_6**: Properties and Determination; **Vitamins**: Determination; **Zinc**: Properties and Determination

Further Reading

Bamji MS (1981) *Vitamins in Human Biology and Medicine*, pp. 1–27, Boca Raton, Florida: CRC Press.

Flair EC Programme (1993) The measurement of micronutrient absorption and status. Flair concerted action no. 10 status papers. *International Journal of Vitamins and Nutrition Research* 63: 247–316.

Sauberlich HE (1984) Implications of nutritional status on human biochemistry, physiology and nutrition. *Clinical Biochemistry* 17: 132–142.

Shils ME and Young VR (eds) (1988) *Modern Nutrition in Health and Disease*, VII edn, pp. 142–446. Philadelphia: Lea & Febiger.

Sokoll LJ, Booth SL, O'Brien ME, Davidson KW, Tsaioin KI and Sadowski JA (1997) Changes in serum osteocalcin, plasma phylloquinone and urinary r-carboxyglutamic acid in response to altered intakes of dietary phylloquinone in human subjects. *American Journal of Clinical Nutrition* 65: 779–784.

Ziegler EE and Filer LJ Jr (eds) (1996) *Present Knowledge in Nutrition*, VII edn, pp. 120–236. Washington, DC: ILSI Press.

Functional Tests

N W Solomons, CeSSIAM, Guatemala City, Guatemala

Background

Reasons abound for assessing human nutritional status, ranging from the population level (epidemiology, public health) to assess risk of nutrient deficiency and excess, and to monitor changes and impacts of interventions to the clinical level, where a practitioner is interested in the patient's nutriture for routine maintenance, as a clue to disease, and in titrating nutritional and disease-specific therapies. (*See* **Epidemiology**.)

One of the options for assessing nutritional status is the approach of *functional assessment*. The purpose of this chapter is to discuss the history, origins, concepts, applications, advantages and disadvantages,

and modern advances in the arena of functional tests of human nutriture.

Historical Perspective

Clinical assessment of nutritional status by examining signs and symptoms is the most venerable approach; in fact, the syndromes of scurvy, beriberi, pellagra, rickets, and anemia had been recognized clinically long before the chemical nature and etiological roles of ascorbic acid, thiamin, nicotinic acid, cholecalciferol, and the hematinic nutrients (iron, folic acid, vitamin B_{12}) were discovered. (*See* **Anemia (Anaemia)**: Iron-deficiency Anemia; **Scurvy**; refer to individual nutrients.)

As *sensitivity* (the ability of a test to find all affected individuals) and *specificity* (the ability of a test to reject all unaffected individuals) of a diagnostic approach must always be considered, clinical examination has its limitations. In terms of sensitivity, clinical examination is a poor index of nutritional deficiency, as clinical manifestations emerge late in the process and only in those most deficient. They represent the tip of the iceberg. The specificity of clinical diagnosis is variable. Whereas scurvy and rickets are unmistakable, anemia has multiple nutritional bases and a host of causes unrelated to nutrition.

The twentieth century saw the chemical identification of the essential nutrients, producing reliable assays to measure nutrients or their metabolites in biological fluids and tissues. The first-line approach for the detection of deficiencies and excess of nutrients shifted from clinical examination to laboratory tests, with the realization that changes in body reserves and tissue content of nutrients decrease before the clinical manifestations of deficiency are expressed. Similarly, in terms of excessive accumulation, an increased burden of a nutrient occurs before overt toxic signs and symptoms develop. Most, but not all, biochemical procedures generate *static* indices of nutritional status. Assessment of individual status is pursued by constructing a normative reference distribution for a healthy, well-nourished population as the standard, and assigning cut-off criteria on the high and low ends that represent excess and deficiency, respectively. For a population assessment, the deviation of its biochemical distribution from that of the reference population is gauged, or the prevalence of individuals with values outside the criteria levels is tabulated.

The concept of *functional* assessment was added to the lexicon of nutritional status evaluation in the 1970s, although its origins go much further back to the Hess tourniquet test for capillary fragility for incipient scurvy of the early twentieth century, dark-adaptation tests for vitamin A status, and prothrombin- and coagulation-time tests for vitamin K adequacy of the middle of the century. However, in 1978 Study Team IX of the Committee on International Nutrition Programs of the National Academy of Sciences of the USA, chaired by the late Professor Doris Howes Calloway, issued a report arguing that the intactness of the physiological and behavioral functions that depended on nutrients was of more interest to policy-makers in governments and development assistant agencies than were the levels of nutrients in the body. In 1983, Noel Solomons and Lindsay Allen produced a *systematic* classification of tests of nutritional status based on measures of physiological performance or behavioral responses.

Definitions

This article compares and contrasts two types of indices that seek to determine the nutriture of individuals: *static* and *functional* indices.

Static Indices

Static indices are tests directed at assessing the *quantitative* content of a nutrient in the organism, either as whole-body reserves or as tissue concentrations. They represent chemically measurements of the nutrient itself, some active or inactive metabolite, or a complex, such as hemoglobin, that contains the nutrient. In the instance of iron status, for example, hemoglobin is a surrogate for iron in target tissues, whereas ferritin reflects its presence in body iron stores. Serum iron, transform saturation, and iron in hair or nails are all examples of static indices of iron status. (*See* **Anemia (Anaemia)**: Iron-deficiency Anemia.)

Functional Indices

Functional indices of nutritional status are those behavioral, physiological, or biochemical functions of the organism dependent on the adequate availability of a nutrient or resulting from the homeostatic regulatory processes that maintain body stores and harmonic internal distribution of some nutrients.

As mentioned, perhaps the first example of a functional test of nutritional status was the capillary fragility test of Hess, followed in the 1930s by dark adaptation tests based on Wald's elucidation of the dependence of retinal cone function on vitamin A. With the advent of isolated radioisotopes and whole-body counting, tests of the absorption, retention, and distribution of several nutrients have been developed. With the molecular biology, genomic, and proteonomic revolutions, new and unexplored

opportunities for functional tests have emerged. (*See* **Ascorbic Acid**: Physiology; **Retinol**: Physiology.)

Classification of Functional Indices

Classification by System

Functional tests can be conveniently classified according to the body system. The categories are as follows: (1) structure and structural integrity; (2) immunity and host defense; (3) transport; (4) hemostasis; (5) reproductive biology; (6) nervous system function; (7) hemodynamics and physical capacity. Some functional tests cannot be classified into any of these, and reside in a 'miscellaneous' catchment category.

Structure and structural integrity Under 'structure and structural integrity' are classified all of those tests that relate to growth and development of children and adolescents. At the level of specific tissues, tests of the integrity of somatic membranes, red cell membranes, capillary walls, and collagen and osseous structures also pertain to this category. The destruction (lipid peroxidation) of membranes is presumably related to the adequacy of vitamin E and of other putative antioxidant nutrients such as selenium, zinc, and copper. (*See* **Growth and Development**; refer to individual nutrients.)

Immunity and host defense In the category of immunity and host defense are included the multitude of tests of phagocytic, humoral, and cellular immunity,as well as the newer tests of messenger (hormonal) function of the cytokine mediators of the inflammatory response and of autoimmunity and autoimmune surveillance. Almost all macronutrients, and most of the micronutrients that relate to cellular proliferation or metabolism, influence the function at some level of host immunity. (*See* **Immunology of Food**.)

Transport Tests of transport function begin at the intestinal level, specifically with its homeostatic regulation of the body reserves. Additionally, at the level of transmembrane movement of nutrients, we have different uptake potentials based on the nutrient status of the organism. In both cases, deficiency should enhance a net inward flux, and sufficiency should balance inward and outward movement. Among the nutrients with regulated uptake at the gut are iron and zinc. Cobalt ion can be used as a surrogate for iron in an isotopic absorption test. The conversion of carotene to vitamin A by the gut is also regulated by the level of body vitamin A stores, and theoretically could be measured by a clinical test that measures the outcome of its uptake and intestinal metabolism. Isotopic turnover studies also fall into the category of tests of nutrient transport, as do the specific tests of mobilization from stores to the periphery, such as the relative dose–response and the modified relative dose–response for vitamin A status assessment. (*See* **Cobalt**.)

Hemostasis Nutrition with respect to vitamin C, vitamin K, and all micronutrients that influence platelet aggregation can be reflected in tests of hemostasis related to capillary structure, bleeding time, or clotting capacity. It should be noted that capillary integrity can be classified under 'structural integrity' or under 'hemostasis.'

Reproductive biology Spermatogenesis and gonadal–pituitary hormone axis function, as well as lactation and reproduction performance in the individual women and, at the population level, fertility and fecundity of the community, birthweights, and peripartum hemorrhage rates constitute functional tests of reproductive biology. (*See* **Menstrual Cycle: Nutritional Aspects**; **Premenstrual Syndrome: Nutritional Aspects**; **Lactation**: Physiology; **Lactic Acid Bacteria**.)

Nervous system function From cognitive performance, social competence, scholastic performance, through sleep pattern to peripheral nerve transmission and, finally, to the specific sensory reception in eyes, ears, nose, and tongue, functional tests in the nervous system cover a ganut of nervous functions. Both macro- and micronutrient deficiencies have been associated with impairment of higher cerebral functions. Micronutrients, such as thiamin with respect to the sixth cranial nerve function, vitamin A, zinc, and vitamin E for retinal function, and zinc and vitamin A for olfactory acuity, are some of the nutrients that can be evaluated in tests of the special senses.

Hemodynamics and physical capacity Functional tests of hemodynamics and physical capacity are based on techniques of noninvasive and work performance physiology. Doubly labeled water ($^2H_2{}^{18}O$) can be used as a long-term marker of energy expenditure. Protein, energy, and iron are the nutrients most regularly thought to be determinant of the performance of the individual in efforts of physical exertion. (*See* **Exercise**: Metabolic Requirements.)

Classification by Type of Test

The alternative form of classification of functional test of nutritional status is by type of test, i.e., by the manner in which the function is approached. These include four levels: *in vitro* tests of *in vivo* functions;

induced responses and load tests *in vivo*; spontaneous *in vivo* activities and responses of organs and tissues; and responses at the level of the individual or population. A selected array of functional tests are listed under the respective classifications described in the following four sections.

In vitro tests of *in vivo* functions

- erythrocyte transketolase activity coefficient;
- erythrocyte glutathione reductase activity coefficient;
- erythrocyte transaminase activity coefficient;
- circulating homocysteine concentration;
- granulocyte chemotactic response;
- lymphocyte chemotactic response;
- bactericidal capacity of granulocytes;
- leucocyte glycolysis;
- granulocyte reduction of nitroblue tetrazolium;
- serum opsonin activity;
- T-cell blastogenesis;
- monocyte interferon production;
- monocyte monokine (interleukin-1; tumor necrosis factor) production;
- erythrocyte fragility;
- ^{75}Se uptake by erythrocytes;
- ^{65}Zn uptake by erythrocytes;
- thymulin activity reconstitution by zinc *in vitro*;
- prothrombin time;
- descarboxy prothrombin levels;
- platelet aggregation;
- D(−)-uridine suppression test;
- ^{14}C-formate conversion in lymphocytes;
- tensile strength of skin strips.

This group of functional tests requires some form of biological sample, either blood to separate the various cellular elements or a skin biopsy. Where specific cellular elements such as platelets, lymphocytes, and granulocytes are concerned, the volumes of blood to be extracted can be considerable. The samples required for erythrocytes are usually small.

In vivo induced responses and load tests

- capillary fragility (Hess) test;
- experimental wound healing;
- collagen accumulation in subcutaneous sponge implant;
- mobilization of inflammatory cells (Rebuck skin window);
- delayed cutaneous hypersensitivity;
- *de novo* antibody formation;
- vasopressor response;
- total-body energy expenditure with $^{2}H_{2}{}^{18}$ metabolism;
- isotopic turnover studies with whole-body counting;
- radioiron absorption;
- radiocobalt absorption;
- zinc absorption tests;
- intravenous magnesium infusion tolerance (urinary excretion) test;
- retinol relative dose–response;
- dihydroxyretinol modified relative dose–response;
- β-carotene bioconversion to retinyl ester challenge
- thyroid radioiodine uptake;
- glucose tolerance test;
- postglucose plasma chromium response;
- postglucose urine chromium response;
- glucose load with exercise for lactate and pyruvate;
- urinary malonaldehyde excretion;
- mixed-function oxidase ($^{14}CO_2$) breath test;
- ^{14}C-histidine ($^{14}CO_2$) breath test;
- ^{14}C-serine ($^{14}CO_2$) breath test;
- histidine load for urinary formimino glutamic acid;
- histidine load test for urinary hydantoin propionic acid;
- purine load test for urinary xanthine;
- sulfur amino acid load test for abnormal sulfur metabolites;
- sodium bisulfite load test for abnormal sulfur metabolites;
- tryptophan load test for urinary xanthurenic acid;
- leucine load test for urinary 3-hydroxyisovaleric acid.

In this class of test, some form of substrate, either natural or artificial, and unlabeled, labeled with a stable isotope, or labeled with a radioisotope, is administered, and its metabolism is monitored. The metabolism is either governed by a nutrient-dependent function of the nutrient in question or related to the body's homeostatic regulation of nutrient reserves.

Spontaneous *in vivo* activities and responses of organs and tissues

- capillary bleeding tendency;
- dark adaptation;
- visual recovery time;
- central scotoma size;
- color discrimination;
- taste acuity;
- olfactory acuity;
- auditory acuity;
- abducens (sixth cranial nerve) function;
- nerve conduction time;
- skin conductivity;
- electroencephalography;
- sleep pattern;

- grip strength;
- weight gain as lean tissue;
- weight gain as fat tissue;
- maximal oxidative capacity;
- maximal work capacity;
- urinary malonaldehyde excretion;
- methyl selenol/methyl trimethyl selenomium ratio in urine;
- pulmonary excretion of volatile hydrocarbons (ethane or pentane);
- peripartum hemorrhage;
- sperm count.

The spontaneous function of organs and tissues, without provoking special handling of substrates or external stimuli, such as taste acuity and dark adaptation, constitute the least invasive and complex, and the most acceptable (to the person examined) of the range of functional tests of nutritional status.

Responses at the level of the whole individual or population

- linear growth velocity;
- total weight gain velocity;
- work productivity and spontaneous play activity;
- birthweight;
- Apgar score at birth;
- infant developmental scale performance;
- infant temperament;
- cognitive performance;
- scholastic performance;
- athletic performance;
- social competence;
- development of secondary sex characteristics;
- lactation performance;
- fertility rate;
- fecundity rate;
- infant mortality rate;
- disease resistance or susceptibility.

Finally, a series of functions of the whole individual or the average statistics for a community population can be used to assess nutritional status, as well as the impact of improvement or deterioration of health and dietary conditions on individual or collective nutriture. However, the more complex and integral the function, the greater is the number of factors – both nutritional and nonnutritional – that influence, mediate, and determine the performance. The quality of educational materials, the skill and motivation of teachers, and the state of civil tranquility are among the covariables that might influence the scholastic performance of children, in addition to the nutrition of the school population. If the first limiting factor in scholastic performance lies among these variables, it is doubtful that any limitation to maximal output from malnutrition could be demonstrated or exposed.

Newer *In vitro* Approaches to Functional Assessment of Nutritional Status Based on Molecular Biological (Genomics, Proteomics) Techniques

The decade since the publication of the first edition of this volume has been marked by dramatic advances in the molecular biology of characterizing and expressing the genetic potential and regulation of the cell at the level of the genetic code information (genomics) and the expression of messenger RNA for specific protein (proteomics). Because nutrients can become limiting in the basic mechanisms, and can induce or repress genetic expression, genomics and proteomics protend a series of approaches to functional assessment:

- nutritional expression of nutrient-dependent or nutrient-containing enzymes or proteins;
- nutritional expression of homeostatic regulatory protein;
- nutritional modification of regulatory elements;
- nutritional pattern of upregulation (or downregulation) of gene expression.

In fact, several laboratories have begun to exploit the second of these methods, in tests for human zinc status based on the expression of mRNA for metallothionein, the intracellular binding protein of divalent metals, in peripheral nucleated blood cells.

Advantages and Disadvantages of Functional Assessment

Advantages

In terms of screening tests in epidemiology and clinical nutrition, functional indices of nutritional status have potential advantages of noninvasiveness, individual accuracy of diagnosis, timeliness and interpretability of nutritional interventions, and policy relevance.

Noninvasiveness Almost by definition, static indices involve the sampling of blood or tissue. Many of the tests of performance are noninvasive in so far as no skin punctures or scrapings are involved. This factor may account for more compliance and cooperation on the part of potential clients or subjects. In an era of hepatitis B and human immunodeficiency virus viral endemicity, nonpenetrating procedures are also welcomed by investigators and laboratory personnel.

Individual accuracy of diagnosis The supreme advantage of functional tests is their potential to provide an accurate diagnosis of status for the *individual* patient, client, or subject. However, this can be accomplished only through *serial* testing before and after a trial of physiological or therapeutic doses of the nutrient in question. Whereas static indices, or functional indices measured in a cross-sectional fashion, must be interpreted as normal or abnormal in a normative fashion, in relation to the distribution in a standard reference population, functional tests repeated in the same individual can be assessed with respect to the degree of incremental change. Thus, an accurate assessment of status in an individual is possible. Those with a nutrient-responsive improvement or normalization in a function are truly deficient. For example, consider two persons with the same dark adaptation threshold, in the highest 10th percentile of the reference population; both receive a course of vitamin A therapy, but only one subject improves. If there is no intrinsic retinal disease in the latter, we can conclude that the threshold is a normal variant. In the former, it is a manifestation of vitamin A deficiency.

Temporal sensitivity in detecting nutritional deficits Another advantage of the functional test is the temporal sensitivity in detecting nutritional deficits. We generally recognize a hierarchy of *sequence* in the stage development of depletion. The earliest changes are those in nutrient-dependent function at the peripheral tissue level, often detectable before the nutrient levels in the tissues themselves have declined. Thus, between a functional deficit or a tissue-concentration decline, conventional wisdom favors the former as the earliest detectable alteration during nutritional depletion. As noted earlier, the signs and symptoms of the clinical deficiency syndrome appear much later in the course of depletion.

Interpretability of nutritional intervention Demonstrating the efficacy and effectiveness, that is whether or not a nutritional intervention, such as the fortification of a food, the distribution of a supplement, or the application of a nutrient-sparing program such as deparasitization or immunizations, is actually having an impact, has been the bane of public health nutrition. Even when normative references are used to determine 'abnormals,' in a baseline survey, a marked shift in the distribution of a population toward a normal distribution after the implementation of an intervention can allow success to be interpreted. The failure to observe any change can be taken as tentative evidence for a null effect. However, in serial testing, one must account for any 'learning effect' or spontaneous improvement that would be expected from prior exposure to, or experience with, the test procedure.

Policy relevance The policy relevance of functional assessment derives, in part, from the consequence of the foregoing discussion. For policy-makers, the demonstration that some aspect of *performance* was improved by a public health intervention generates more conviction and sustainability than simply showing that the level of the nutrient had increased in blood, hair, or urine. It provides a ready answer to the 'so what?' questions that those in charge of legislation or expenditures for health will ask.

Disadvantages

Functional assessment of nutritional status, however, is not a panacea; there are also notable disadvantages in their application. These include poor interpretability in cross-sectional studies, lack of specificity for nutritional status, and specific considerations in young children.

Poor interpretability in cross-sectional studies Status and functional indices share a common limitation when the issue is the nutritional interpretation of a result from a single clinical examination or a cross-sectional study. This limitation is the difficulty in accurately separating those with suboptimal nutritional status from those with adequate status by simple reference to an arbitrary cut-off level. It is a *probabilistic* issue in which only a certain level of *risk* of pertaining to a truly deficient or truly sufficient population can be concluded from a single point-in-time evaluation.

Lack of specificity for nutritional status Almost all of the functions that are measured to assess nutritional status can also be impaired by other (nonnutritional) causes. Thus, in a cross-sectional or single measurement assessment using a functional index of nutritional status, it is impossible to be sure that an impaired or abnormal function is related to nutritional deficiency or to some other pathology or a genetic abnormality. Moreover, even if a longitudinal study with supplementation is employed with a functional index, if the person has *multiple* nutrient deficiencies, and only a single nutrient is supplemented, the remaining deficits may impair a full recovery of the functional abnormality.

Considerations in young children For obvious reasons, tests related to reproductive biology are not applicable to children. Moreover, the use of radioisotopes *in vivo* is relatively or absolutely

contraindicated in young children. Stressful stimuli, such as vasopressors or electrical conductivity electrodes, or stressful performance challenges, such as maximal oxidative capacity, should be proscribed from the assessment of children.

Separating pharmacological from nutritional effects Even when serial or comparative measurements of a functional outcome are coupled to a nutrient-intervention format, the interpretation of the preexisting or prevalent nutritional state must be made with caution and judgment. What is a 'nutrient-responsive' situation, and what is a 'nutrient-deficient' condition? By way of analogy to personal esthetics, the application of make-up to a physically unattractive individual can make them 'presentable.' The appearance of a person already attractive by consensus standards can be enhanced with cosmetics to a state of 'ravishing,' however. Where are the valid boundaries to define an 'esthetic deficit'? Similarly, in household economy, what constitutes a 'deficient' income: when one cannot afford a Mercedes-Benz car? When one cannot afford any vehicle? Or when the funds do not allow for purchase of the basic necessities of food, clothes, shelter, and care? Additional wisdom and a grounded philosophical perspective are needed to differentiate enhancement (nutrient-responsive) from reversal of true deficits (nutrient-deficiency).

Important Applications of Functional Status Assessment

Assessment of Risk from Surgery and Other Major Therapies

Recently, grip strength, i.e., the development of contractive force by the hand musculature, has been shown to have some bearing on the prediction of survival or mortality, or the risks of complications in recovery from surgery or radiation or chemotherapy in association with static indices of anthropometry, immune status, and protein concentrations. Tests of active internal lipid peroxidation, such as the ethane or pentane breath tests, have theoretical promise as complementary indices for the prediction of risk of major therapy of precariously nourished patients.

Evaluation of Public Health Impact of Interventions

Faithful to the spirit of Study Team IX, the 1990s saw an increasing incorporation of functional assessment into the evaluation of public health interventions. Child growth in height and weight is the most commonly used variable. Other indicators of interest in public health nutrition are infection resistance, lactation performance, fertility, work output, and mortality. (*See* **Infection, Fever, and Nutrition**; **Children**: Nutritional Requirements.)

Functional assessment can also be used to gain inferences for improving the balance of nutrients in traditional regional diets. As field intervention studies with supplements of protein and energy failed to show major functional improvements, but the association with animal protein did, the *quality* of the diet in terms of micronutrient (vitamin, mineral) malnutrition was focused on in the 1980s.

Assessment of Dietary Nutrient Intake Requirements

Functional assessment is gaining a place in the continued fine-tuning of our knowledge of human nutritional requirements and dietary recommendations. The dominant concept of where to set the recommendation for the daily intake of a nutrient is at the amount that will prevent the development of signs and symptoms of deficiency in the majority or a population. When possible, the concept of the amount that supports a specific pool size of total-body nutrient content has been introduced. Given that functional alterations are generally the earliest detectable changes with the onset of depletion, the logical approach to defining intake requirements of a nutrient is that level which prevents functional changes in the majority of a population. As experience is gained with interpreting and avoiding pitfalls in the application of functional tests, they will be used increasingly to refine our present understanding of what nutrient intakes to recommend to maintain adequate nutriture. (*See* **Dietary Requirements of Adults**.)

New Frontiers for Functional Assessment

Functional indicators of adverse interactions from excesses or toxicity associated with high doses of intakes of nutrients for use in nutritional toxicology have lagged behind the progress in diagnosis of deficient and marginal states. More thought and research need to go into this domain.

The late James Olson has made a pithy distinction between the functions and actions of constituents in the diet. The former relate to the *essential* functions that define a nutrient. However, at higher doses, these nutrients begin to have actions to favor health or prevent disease that go beyond their capacities to nourish. Bioactive (nutriceuticals, phytochemicals) compounds in foods such as flavonoids, carotenoids, etc. have assumed growing interest. Since the chemical nature of the active substances is often unknown, functional markers seem to be the only variables with which to assess the status of consumers.

Conclusions

The clinician, epidemiologist, investigator, and public health professional – all those involved with nutritional status diagnosis – should be aware of functional assessment, its potential, and its limitations. Any approach to nutritional assessment has its advantages, disadvantages, and limitations. The use of functional indices of nutritional status tends to be less invasive, and more client-friendly, than static assessment tests that seek to measure nutrients directly in body fluid or tissues.

Impairment of a peripheral nutrient-dependent function or the body's response to regulation often occurs before nutrient levels decline in blood and tissues, and well before clinical manifestations of deficiency emerge. When coupled to a trial of nutrient supplementation, the functional approach, unlike the static approach, can provide a definitive, retrospective assessment of the original nutritional status of the host. Functional tests can be classified by the physiological or anatomical system that is measured, or by the level of investigation of the test. As the nature of the metabolism and physiological and biochemical functions of nutrients becomes better understood, newer strategies to exploit the functional assessment of nutritional status will be suggested.

See also: **Anemia (Anaemia)**: Iron-deficiency Anemia; **Ascorbic Acid**: Physiology; **Children**: Nutritional Requirements; **Cobalt**; **Dietary Requirements of Adults**; **Epidemiology**; **Exercise**: Metabolic Requirements; **Growth and Development**; **Immunology of Food**; **Infection, Fever, and Nutrition**; **Lactation**: Physiology; **Lactic Acid Bacteria**; **Premenstrual Syndrome: Nutritional Aspects**; **Retinol**: Physiology; **Scurvy**

Further Reading

Calloway DH (1982) Functional consequences of malnutrition. *Reviews of Infectious Diseases* 4: 736–745.

Chandra RK (1981) Immunocompetence as a functional index of nutritional status. *British Medical Bulletin* 37: 89–94.

Chandra RK (2001) Effect of vitamin and trace-element supplementation on cognitive function in elderly subjects. *Nutrition* 17: 709–712.

Pollitt E (1996) Timing and vulnerability in research on malnutrition and cognition. *Nutrition Reviews* 54: 455–549.

Solomons NW (1985) Assessment of nutritional status: functional indicators of pediatric nutriture. *Pediatric Clinics of North America* 32: 319–334.

Solomons NW and Allen LH (1983) The functional assessment of nutritional status: principles, practices and potential. *Nutrition Reviews* 41: 33–50.

Nutritional Management *See* **Elderly**: Nutritional Management of Geriatric Patients; **Food Intolerance**: Types; Food Allergies; Milk Allergy; Lactose Intolerance; **HIV Disease and Nutrition**; **Hypertension**: Nutrition in the Diabetic Hypertensive; **Infection, Fever, and Nutrition**; **Multiple Sclerosis – Nutritional Management**; **Pregnancy**: Nutrition in Diabetic Pregnancy

Nutritional Policies *See* **Nutrition Policies in WHO European Member States**

NUTRITIONAL SURVEILLANCE

In Industrialized Countries

K F A M Hulshof, M R H Löwik and D C Welten,
TNO Nutrition and Food Research, The Netherlands

Introduction

Especially during the last two decades, evidence has accumulated that prevailing dietary patterns have adverse health effects. Nutritional assessment has become an important topic on the health-policy agenda. This article describes the aim of nutritional surveillance, the availability and usefulness of nutritional surveillance indicators (especially food consumption), and some trends, risk groups, and risk areas. Nutrition-related health problems and surveillance systems differ among countries, in particular between developed and developing countries. This article deals only with industrialized countries in Europe, Canada, Japan, Australia, and the USA. This paper reports the situation in the early 2000s.

Aim of Nutritional Surveillance

As stated by a Joint FAO/UNICEF/WHO Expert Committee, the objectives of nutritional surveillance are as follows: to describe the nutritional status of the population, with particular reference to groups at risk, to contribute to the analysis of causes for changes and differences, to promote decisions by governments on food and nutrition policy issues, to predict future trends, and to evaluate the effects of nutritional programs.

The international conference on Nutrition held in Rome in 1992, adopted a 'World declaration on Nutrition' and a 'Plan of actions for Nutrition.' In this plan, one of the nine action-oriented themes was 'Assessing, analysing and monitoring nutrition situations.' According to these themes, governments should, among other things: identify the priority of nutritional problems in their country, analyze their causes, plan and implement appropriate remedial situations, and monitor and evaluate efforts to improve the situation. Moreover, as a consequence of recent adopted directives of the European Union, the member states have to create systems that will monitor the use and intake of additives.

As illustrated in Figure 1, nutritional surveillance ideally provides information on a wide range of variables, from food availability, distribution and consumption and nutrient utilization (as reflected in nutritional status) to, ultimately, health status and mortality. This results in identification of public health problems that call for specific action and lead to nutrition research priorities (both applied and more fundamental). The data can be obtained from either existing sources, including administrative data, or surveys undertaken specifically for surveillance purposes.

Sources of Dietary Information

Insight into dietary patterns is a core target of nutritional surveillance, since this provides a comprehensive basis for nutritional risk assessment. In principle, three different types of data can be used: food supply data, data from household consumption surveys, and data from dietary surveys among individuals. Each type of data corresponds with a different stage in the food chain and is obtained by different methods.

Food-supply Data

Food-supply data provide information on the type and amount of food available for human consumption, to the country as a whole. The supply is calculated in food balance sheets (FBSs), which are accounts, on a national level, of annual production of food, changes in stocks, imports and exports, and agricultural use and industrial use. Food supply is usually expressed per head of the population in kilograms per year, or grams per day. The per-capita consumption of energy and some additional nutrients is calculated using food composition tables.

Food-supply data refer to food availability, which gives only a crude (overestimated) impression of potential consumption. Food and nutrient losses prior to consumption, owing to processing, spoilage, trimming and waste, may not be adequately accounted for. Furthermore, these data provide no information about the distribution of food among population groups or districts.

International FBSs are prepared and published by the FAO, the Organization for Economic Cooperation and Development (OECD) and the statistical office of the Commission of the European Union (EUROSTAT). The FAO has published FBSs since 1949, also covering the period 1934–1948. Since 1949, FBSs have been compiled on an annual basis from data supplied by about 200 countries. Information is available for all European countries, Australia,

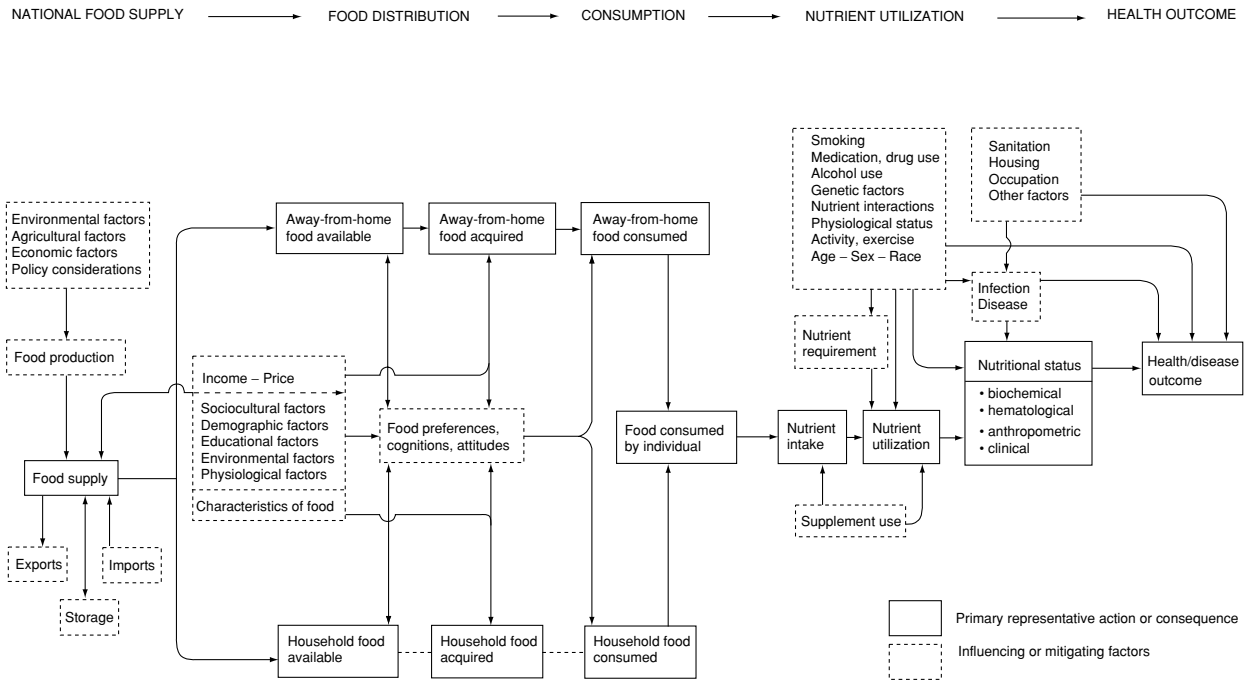

Figure 1 General conceptual model for food choice, food and nutrient intake, and nutritional and health status. Source: Life Sciences Research Office, FASEB (1989) Nutrition and Monitoring in the United States. An Update Report on Nutrition Monitoring. Prepared for the US Department of Agriculture and the US Department of Health and Human Services. DHS Publication No. (PHS) 89–1255. Public Health Service. Washington, US Government Printing Office, September 1989.

Canada, Japan, New Zealand, and the USA. Since 1971, the FAO has included its FBS data in the Interlinked Computer Storage and Processing System of Food and Agricultural Commodity Data (ICS). EUROSTAT publishes FBSs for the 15 member countries of the European Union. The OECD FBSs cover, besides several European countries, Australia, Canada, Japan, New Zealand, and the USA. Although FBSs are compiled in a similar way, they differ in coverage, food grouping and level of processing of commodities (e.g., FAO lists 300 food items classified into 17 food group categories; OECD 70 items in 13 categories) and in nutrient conversion. The FAO and OECD usually publish summaries of FBSs every 3–5 years, with a lag time of 3–4 years between data collection and publication. The ICS supplies figures on magnetic tape, on floppy disk and on Internet (http://apps.fao.org). EUROSTAT publishes supply balance sheets in the Agricultural Statistical Yearbook.

In addition to the international FBSs, many countries publish national FBSs, mostly in statistical yearbooks or special statistical publications. For example, US food-supply statistics are available from 1909 onwards. National FBSs tend to be more up to date. Owing to different methodologies underlying their compilation and presentation, these data can differ from international FBSs.

Despite their limitations, FBSs are useful in that they indicate the (in)adequate aspects of food supply, provide material for planning food supply (production, imports and exports) and give crude indications of (un)desirable changes in terms of potential (adverse) expected health impact. As a result of their long history, FBSs are especially used for assessing trends over time. In contrast to national FBSs, the international FBSs can be used for comparative studies, provided that the FAO and OECD data sets are used separately. Figure 2 illustrates the use of FBS data (FAO) for comparison across countries and for trends over time. This table shows the consumption of meat in five selected countries. Only in the UK has the consumption of meat and meat products remained remarkably steady since 1965. In the other countries, the total consumption of meat has increased considerably, especially in Spain. The same tendency has been observed in other southern European countries and reflects one of the important changes in the Mediterranean diet over the past decades. Such comparisons implicitly assume that the demographic changes across the countries are similar.

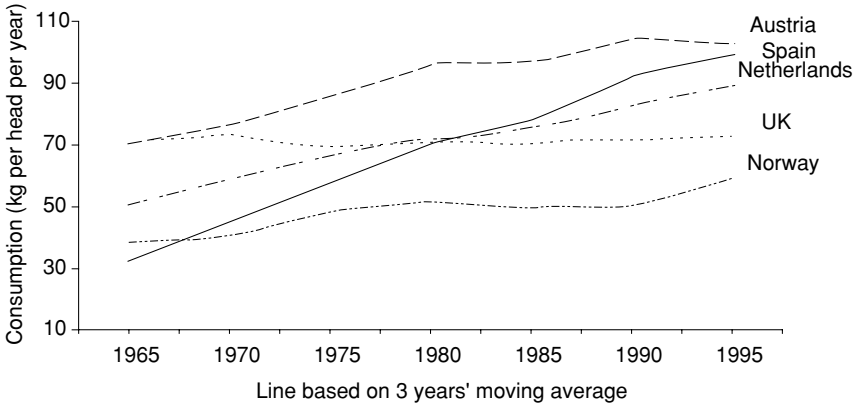

Figure 2 Available consumption of meat and meat products (kg per head per year) in five European countries. Source: FAO Statistical Databases: Food Balance Sheets. http://apps.fao.org. (10.03.2000).

Household Surveys

Food available at the household level may be estimated by budget surveys and by consumption surveys. The first type of survey gives information on the purchases of food in terms of expenditure and is used for economic policy. For example, weights for the construction of consumer price indices can be calculated. In household-consumption surveys, the amounts of foods and drinks brought into the household are also recorded. For the most part, only the expenditures of meals taken outdoors are noted. Some household surveys may even measure changes in food stocks, in addition to acquisition.

In general, household surveys do not provide information on how food is handled within the household, or on actual consumption by its members. Sometimes, the consumption data are converted to individual intake levels. The methods vary from simply dividing the total consumption by the number of people in the households, to assigning factors (consumer units) to persons weighed according to age and sex. In contrast to FBSs, household surveys can supply information on food (and nutrient) patterns in subgroups of households. These groups may be classified by economic, demographic, and other factors, which provide the opportunity for risk group identification.

In most countries, household surveys started in the 1940s or 1950s. Only few countries have a continuous system, some repeat surveys every 3–4 years, and others repeat surveys only every 5–10 years. In The Netherlands, the household budget surveys started in 1951, and since 1978, they have been conducted annually. In Europe, the best-known study is the specialized and ongoing household food consumption survey of the UK. Australia, Canada, and Japan have regularly conducted household consumption surveys. In the USA, the first national household consumption survey was conducted in 1936–1937. Between 1942 and 1965, four nation-wide studies on household consumption were carried out. Since 1965, US household food-consumption surveys have also provided information on food intake at the individual level of household members. At present, a wide range of data on household surveys are available, as shown in the FAO Food and Nutrition Policy Papers and WHO publications. Since the dietary data are based on a variety of methods, the surveys are not very suitable for comparisons among countries. Differences exist in sampling procedures, food grouping, conversion to nutrients and period, frequency, and technique of data collection. Snacks, sweets, soft drinks, and alcoholic beverages are excluded from some studies. Data on the quantity of and/or expenditure on food may be collected by record keeping, by interviews or by both methods. Household accounts for nonfood items can cover a period of 4 weeks, but for foodstuffs, 2 weeks is more usual. The results of household surveys play an important role in nutritional surveillance at the national level, particularly when surveys are carried out annually, which reveals trends in food consumption.

To improve the possibilities for international comparison of these data in Europe, since 1993, an ongoing project of the European Union (DAFNE, DAta Food NEtworking) is harmonizing at the international level dietary exposure data from household budget surveys. DAFNE aims to create a pan-European food databank based on national household budget surveys by the development of the most appropriate way of using food and related data from these surveys. Methods were designed to calculate, for instance, the overall average availability per person per day of comparable food items or groups among DAFNE countries as well as average availability by degree of urbanization and educational level of household head. Distribution and cumulative distri-

Table 1 Average availability of fruit and vegetables (fresh and processed) in DAFNE countries (grams per person per day)

	Greece	*Ireland*	*Luxemburg*	*Norway*	*Spain*	*UK*
Fruit	283	103	234	174	307	133
Vegetables	229	130	180	102	179	157

Source: DAFNE II project, 1998.

bution functions of the availability of food items were estimated. Table 1 shows the differences in availability of fruit and vegetables (fresh and processed) in Greece, Ireland, Luxembourg, Norway, Spain, and the UK.

Individual Dietary Surveys

In contrast to FBSs and household surveys, data from individual surveys provide information on average food and nutrient intake and their distribution over various well-defined groups of individuals. Data more closely reflect actual consumption and can provide additional information on meal patterns, etc.

To collect dietary intake data on an individual level, several methods can be used. Briefly, the methods can be divided into two categories: record and recall methods. Record methods collect information on current intake, keeping a record of all foods and drinks based on menu, household measures and/or weighing, over one or more days. Recall methods reflect past consumption, varying from intake over the previous day (24-h recall) to usual food intake (dietary history or food frequency). Each form has its own strengths and weaknesses, and there is no single ideal method. Details of the available methods for assessing food consumption of individuals are given in numerous reviews and manuals.

To characterize the average intake of food and nutrients and their distribution over various groups of individuals one 24-h recall or 1-day food record is appropriate, provided that the sample is representative of the population under study, and day-of-the-week and seasonal variations are taken into account. To determine the proportion 'at risk' of inadequate intake, the food consumption of each subject must be measured over more than one day, or retrospective information on intake over a longer period may be used (e.g., dietary history method). The appropriate period depends on the purpose of making an estimate, the precision desired, the food component(s) of interest, the intra- and interindividual variation components, and the period over which an intake has to be low or high before health risks are introduced.

Probably all (industrialized) countries, if not all, have carried out small-scale dietary surveys. These surveys provide valuable information, but owing to samples of convenience and different food consumption methods, their usage in national nutrition policy and nutritional surveillance is often limited. Since the mid-1980s, the number of countries that have conducted individual dietary surveys is growing. The USA has one of the most comprehensive nutrition monitoring programs in the world and already has a longer tradition in the field of nutritional surveillance. The National Nutrition Monitoring and Related Research Program (NNMRRP) provides information about the dietary and nutritional status of the US population, conditions existing in the United States that affect the dietary and nutritional status of individuals, and relationships between diet and health. The Continuing Survey of Food Intakes by Individuals (CSFII) conducted by the US Department of Agriculture (USDA), and the National Health and Nutrition Examination Survey (NHANES) conducted by the Health and Human Services (HHS) are the two cornerstones of the NNMRRP surveys. They provide national estimates of food and nutrient intakes in the general US population and in subgroups. The CSFII emphasizes the food and nutrient intake defined by various socio-economic factors. In NHANES, information on dietary intake and health status in the same individuals is available. These surveys also provide the potential for assessing consumption of additives and pesticides in the diet.

Also Australia and several European countries have performed individual surveys on a national basis. Table 2 presents examples from studies conducted since 1985. The surveys differ in coverage of population, methods used to collect dietary data, nutrition-related health indices, etc. In several countries, dietary data were collected using a record method, but the number of record days varied from 1 to 7. A 7-day weighed record is thought to be the most accurate method of dietary assessment. However, this method has a high respondent burden, which can have consequences for the response rate, representativeness of the sample and data quality. Response rates vary widely. Sometimes, weighing factors are used to adjust for sources of nonresponse. Some subgroups (such as ethnic minorities, pregnant or lactating women) do not occur in the population in sufficient numbers to appear in the survey with sufficient representation to allow for separate estimates of their diet and nutritional status that are reliable. Extra sampling can improve the precision of estimates in nutritional assessment in specific groups and is used in several studies, including those in the United States. Special (vulnerable) groups can also be examined in separate studies. For example, in 1998, in The Netherlands, a nation-wide survey was conducted among 8-year-old Turkish and Moroccan children and their

Table 2 Examples of nation-wide food-consumption surveys with nutrient-intake data on an individual level since 1985

Country	*Year*	*Survey*	*Population*		*Sample size*	*Response rate (%)*	*Dietary method*[a]
			Sex	*Age*			
Australia	1985	National Dietary Survey of Schoolchildren	F+M	10–15	5224	65	1d Rcd
Austria	1992	Austrian Study on Nutritional Status (ASNS)	F+M	6–18	2173	na	7d Rcd
	1993–1997	Austrian Study on Nutritional Status (ASNS)	F+M	19–65	2065	69	24h Rcl, DH
Denmark	1985	Dietary Habits in Denmark	F+M	15–80	2242	na	DH
	1995	National Dietary Survey	F+M	1–80	3098	65	7d Rcd
Finland	1992	Dietary Survey of Finnish Adults	F+M	25–64	1861	60	3d Rcd
	1997	Dietary Survey of Finnish Adults	F+M	25–64	2862	72	24h Rcl
			F+M	65–74	290		24h Rcl
France	1993–1994	National Food Consumption Survey (ASPCC Survey)	F+M	18+	1229	na	7d Rcd
	1993–1994	National Food Consumption Survey (ASPCC Survey)	F+M	2–85	1500	na	7d Rcd
	1998–1999	Individual National Food Consumption Survey (INCA)	F+M	3–14	1018	na	7d Rcd
			F+M	15+	1985		7d Rcd
Germany	1985–1989	National Nutrition Survey in Former West Germany	F+M	4–70+	24632	71	7d Rcd
	1991–1992	National Health Survey in East Germany	F+M	18–80	1897	71	DH
	1998	German Nutrition Survey	F+M	18–79	4030	61	DH
Hungary	1985–1988	First Hungarian Representative Nutrition Survey	F+M	14–60+	16641	61	2× 24h Rcl+FFQ
	1992–1994	Hungarian Randomized Nutrition Survey	F+M	18–60+	2559	na	3× 24h Rcl+FFQ
Iceland	1990	Icelandic National Nutrition Survey	F+M	15–80	1240	72	DH
Ireland	1990	Irish National Nutrition Survey (INNS)	F+M	8–18+	1214	na	DH
	1998	North–South Food Consumption Survey	F+M	18–64	1379	66	7d Rec
Italy	1994–1996	INN-CA1995	F+M		2734	47	7d Rcd
Lithuania	1997	Baltic Nutrition and Health Survey	F+M	20–65	2183	73	24h Rcl+FFQ
The Netherlands	1987–1988	Dutch National Food Consumption Survey (DNFCS-1)	F+M	1–79	5898	79	2d Rcd
	1992	Dutch National Food Consumption Survey (DNFCS-2)	F+M	1–92	6218	72	2d Rcd
	1997–1998	Dutch National Food Consumption Survey (DNFCS-3)	F+M	1–97	6250	69	2d Rcd
Norway	1993–1994	National Dietary Survey Among Adults (Norkost)	F+M	16–79	3144	63	FFQ
	1997	National Dietary Survey Among Adults (Norkost)	F+M	16–79	2672	54	FFQ
	1999	National Dietary Survey (Norkost)	F+M	6+12 months	2400	80	FFQ
	1999	National Dietary Survey (Norkost)	F+M	2 years	2010	67	FFQ
Northern Ireland	1986–1987	Diet, Life-style and Health in Northern Ireland	F+M	16–64	616	na	7 d Rcd
Poland	1991–1994	Dietary Habits and Nutritional Status of Selected Populations	F+M	11–14	1126	na	24h Rcl
				18	2193		
				20–65	4945		
Slovak Republic	1991–1994	Asessment of Food Habits and Nutritional Status, etc.	F+M	11–14	3337	na	24h Rcl+FFQ
	1995–1998			15–18	4556		
	1999–2001			19–88	4807		
Sweden	1989	HULK	F+M	1–74	2036	70	7d Rcd
	1997–1998	Riksmaten	F+M	18–74	1215	60	7d Rcd
Switzerland	1992–1993	Swiss Health Survey	F+M	15–74	26000	71	FFQ
UK	1986–1987	The Dietary and Nutritional Survey of British Adults	F+M	16–64	2197	70	7d Rcd
	1992–1993	National Diet and Nutritional Survey: Children Aged 1½–4½ years	F+M	1½–4½	1675	80	4d Rcd
	1994–1995	National Diet and Nutrition Survey: People Aged 65 Years and Over	F+M	65+	1687	80?	4d Rcd
	1997	National Diet and Nutrition Survey: Young People Aged 4–18 Years	F+M	4–18	1701	64	7d Rcd

[a]24h Rcl, 24-h recall; 1d Rcd, 1-day record; FFQ, food-frequency questionaire; DH, dietary history method; na, not available. See also: Löwik MRH and Brussaard JH (2002) EFCOSUM.

mothers, with a Dutch control group. Norway plans to carry out a survey among migrants in 2000–2001.

The presented national dietary surveys provide valuable information for usage in national policy and are central in nutritional surveillance, and, when repeated in a proper way, trends over time can be studied. However, for a detailed evaluation of dietary intake in Europe, there is a need for increasing comparability of sampling designs, dietary methods, and selected population descriptors. The establishment of a Health Monitoring Program in Europe should make it possible to measure health status, trends, and determinants throughout the Community; facilitate the planning, monitoring and evaluation of Community programmes and actions; provide Member States the appropriate health information to make comparisons; and support their national health policies. As part of this program, since the end of 1999, the project European Food Consumption Survey Method (EFCOSUM) has aimed to define a (minimum) set of dietary components that are relevant determinants of health. Moreover, the study aimed to define a method for the monitoring of food consumption in nationally representative samples of all age–sex categories in Europe in order to provide internationally comparable data. This method can be used alone, or as a calibration method for ongoing studies. The project made use of progress in relevant projects carried out until now, such as DAFNE and EPIC, and addressed the possibility for data fusion with other health monitoring studies. Fourteen EU Member States as well as eight other EU countries are participating in EFCOSUM.

Nutritional Status and Health Indices

The assessment of nutritional status includes, in addition to dietary intake, indicators of nutrition-related health status, such as anthropometric measurements, hematological and biochemical tests, clinical signs of deficiencies, and risk factors for diseases associated with diet (e.g., overweight). Furthermore, determinants of food and health-related behavior, such as nutritional knowledge and attitudes, may be studied as well. These indicators can be included in the surveys or studied in separate samples. Several national studies listed in Table 2 studied both dietary intake and nutrition-related health-status indicators. In most surveys, anthropometric data were collected (sometimes self-reported data on body weight and body height); in some countries, also, medical examination and/or biochemical and hematological tests (e.g., Germany, Hungary, UK) were carried out, and information on physical activity was included (e.g., Norway, Sweden).

A major advantage of collecting comprehensive (broad oriented) information at the individual level is that interrelationships can be studied. In studying correlations between diet and nutritional status indicators, one of the characteristics of a cross-sectional study is that mostly low correlations are found. This is attributable to, among other things, intraindividual variation and inaccurate assessment of intake and status indicators. In a cross-sectional design, the observation that a particular dietary factor is positively or inversely associated with relevant variable is meaningful, even when there is a low *P*-value, since this provides suggestive evidence for diet–health relationships which should be studied in more detail. To establish a causal link between diet and health, both intervention and (semi)-longitudinal studies are necessary. Endpoints, such as morbidity and mortality data, provide valuable additional information on the role of nutritional factors in diseases.

Risk Areas and Risk Groups

Nutritional assessment includes a normative evaluation of dietary intake and nutritional status indicators in order to estimate, for instance, the proportion of the population at risk. Nutritional-status indices can be evaluated by comparing them with reference values mostly obtained from healthy adults. Alternatively, predetermined cut-off points (based on consensus reports) can be used. In evaluating dietary intake, the reference values applied in recommended dietary allowances (RDAs) or dietary guidelines are often used. However, the usage of cut-off values is prone to some misclassification owing to (biological) variation within and among individuals. Despite the weaknesses of cut-off points, these criteria are commonly used and often needed to evaluate dietary intake as well as nutritional status parameters.

In most industrialized countries, the principal nutrition-related health problems are related to unbalanced (mostly overconsumption) of some nutrients, particularly energy, fat, and saturated fatty acids. Although the mean intake of energy among adults is mostly lower than the recommendations, the data available from nutritional surveillance indicate a high prevalence of overweight and obesity in several countries. Obesity, defined as a body mass index greater than $30\,kg\,m^{-2}$, is a common condition in Europe and also in the USA. Although Table 3 gives only a rough impression (age groups are not always comparable, the periods in which studies were conducted differ slightly, exclusion criteria might vary, etc.), the data show that the proportion of subjects classified as obese varies among countries. Despite the differences, however, in recent decades, the prevalence of obesity has increased in most

Table 3 Prevalence of obesity (BMI $\geq 30\,kg\,m^{-2}$) in some countries

Country	*Year*	*Age (years)*	*Men (%)*	*Women (%)*
Austria	1993	19–65	6	5
England	1994	16–64	15	16
Finland	1991	20–75	14	11
East Germany	1992	25–65	20.5	26.8
West Germany	1991	25–65	16.0	21.4
The Netherlands	1992	19–65	5.5	9.2
USA	1988–1994	20+	22.3	25.0

European countries as well as in the USA in the past decade. For instance, in The Netherlands, among subjects aged 19 years and over, the proportion of obese subjects has increased during the last 10 years from 6.0% (1987–1988) to 10.4% (1997–1998). At the same time, in The Netherlands, a decrease in energy intake and fat consumption was observed. Data from DNFCS-1 (1987–1988) and DNFCS-3 (1997–1998) suggested that energy intake decreased from 2308 kcal day (9677 kJ) to 2190 kcal day (9191 kJ). This reduction of about 120 kcal day (~485 kJ) was attributable to a decrease in fat consumption (protein intake increased, and carbohydrate intake and alcohol consumption remained constant). This may imply that daily energy expenditure has decreased during the same period and may include an increased sedentary behavior. Findings of the statistical office confirmed a decrease in physical activity in leisure time.

Obesity is associated with several specific health risks, including an increased incidence of hypertension, increased noninsulin-dependent diabetes and high levels of cholesterol and other lipids. Corrected for these factors, obesity in itself has been reported to be an independent factor for cardiovascular disease. In most industrialized countries, a higher prevalence of overweight and obesity is observed among subjects with a lower socio-economic status.

Higher rates of mortality and morbidity have been found among lower socio-economic groups, as compared with higher socio-economic groups. In several studies, food-consumption patterns and nutrient intakes have been more consistent with current dietary guidelines among people with a higher socio-economic status. Recent analyses of 33 studies in 15 European countries, conducted within the framework of the European Union's FAIR program (FAIR-97-3096), showed that, particularly in the north and west of Europe, people with a higher education tend to consume more vegetables and fruits and less fats and oils. Table 4 presents some selected results. Other studies also found that people with a higher socio-economic status have reported eating more wholemeal and brown bread and less whole milk, eggs, and meat. A higher socio-economic status has been associated with a lower intake of fat, saturated fatty acids, and refined sugars, and a higher intake of fiber, although mostly, differences in nutrient intake levels have been quite small. In general, food disparities in relation to consumption levels might possibly explain some of the higher rates of mortality and morbidity among lower socio-economic groups.

Table 4 Average difference (95% confidence intervals) in consumption of selected foods (grams per person per day) between the highest and lowest educational level in nine European countries

	Education (highest minus lowest level)	
	Men	*Women*
Fruit	+24 (+19.0; +29.0)	+26.7 (+21.7; +31.8)
Vegetables	+12.1 (+8.3; +15.8)	+17.5 (+13.7; +21.2)
Fats and oils (added)	−2.9 (−4.0; −1.9)	−3.1 (−3.9; −2.3)
Meat	−32.6 (−36.0; −29.1)	−24.3 (−26.9; −21.8)
Milk total	+46.9 (+38.0; +55.9)	+39.9 (+33.2; +46.6)
Cheese	+9.9 (+8.4; +11.4)	+10.6 (+9.3; +11.8)

Source: Roos and Prättälä (1999): Fair-97-3096 project.

Concerning micronutrients, in most countries, the average intake of most minerals and vitamins appears adequate for the population. In general, iron is an exception in that many subjects have a low iron intake in comparison with recommended values. In most countries, groups with a low intake are young children, adolescents and women of child bearing age. Moreover, the intake of vitamins (e.g., vitamin A and its precursors, vitamin B_6, vitamin C, folic acid), minerals, and/or trace elements (e.g., calcium, magnesium, zinc, iodine) might be not always adequate among certain population groups in several countries. However, it should be noted that the confirmation of nutritional risk obtained by biochemical data is essential for assessment of risk areas and risk groups.

Nutritional Surveillance in the Future

A growing awareness of potential relationships between diet (as part of lifestyle) and health is accompanied by an increasing demand for (comparable) data. Research in the field of molecular genetics and nutrient–gene interactions is promising. Monitoring systems may also provide more information on variations within the human genome. This will enable us increasingly to focus on the specific needs of the individual at potential risk.

Since it might be expected that in the coming years, bioactive components, additives, and contaminants will become increasingly a topic of interest, future food-consumption databases should allow also the

assessment of these nonnutrients. Therefore, there will be a need for more descriptive specificity, such as brand names and more information on the food composition. The use of computer technology and advanced statistical methods may facilitate wider applications of survey data and cost-effective data fusions.

To improve valid comparisons between diet and health at the international level, a better harmonization of data collection, methodology and standardization of data analysis will be needed. Since 1990, this concept has already been implemented in the National Nutrition Monitoring and Related Research Program (NNMRRP) in the USA; in Europe, an increasing number of initiatives as to Pan-European projects are coming up with results.

See also: **Cholesterol**: Properties and Determination; **Coronary Heart Disease**: Etiology and Risk Factor; **Diabetes Mellitus**: Etiology; **Dietary Surveys**: Measurement of Food Intake; Surveys of National Food Intake; Surveys of Food Intakes in Groups and Individuals; **Food Composition Tables**; **Hypertension**: Physiology; **Nutrition Policies in WHO European Member States**; **Nutritional Assessment**: Importance of Measuring Nutritional Status; **Obesity**: Etiology and Diagnosis; **World Health Organization**

Further Reading

Becker W and Helsing E (1991) *Food and Health Data. Their Use in Nutrition-policy Making*. WHO Regional Publications, European Series, No. 34. Copenhagen: World Health Organization.

Briefel RR (1996) Nutrition monitoring in the United States. In: Ziegler EE and Filer LJ (eds) *Present Knowledge in Nutrition*. Washington, DC: ILSI Press.

FAO (1977, 1981, 1988) *Review of Food Consumption Surveys*. Nos 1, 27, 35, 44/Rome: Food and Agriculture Organization.

FAO/WHO (1992a) *International Conference on Nutrition. Major Issues for Nutrition Strategies. Promoting Appropriate Diets and Healthy Life-styles*. Theme Paper No. 5. Rome: Food and Agriculture Organization.

FAO/WHO (1992b) *International Conference on Nutrition. Nutrition and Development: a Global Assessment*. Rome: Food and Agriculture Organization.

Hulshof KFAM and Löwik MRH (1995) Nutritional problems in the industrialized world. In: Velden van der K, Ginneken van JKS, Velema JP, Walle de FB and Wijnen van JH (eds) *Health Matters. Public Health in North–South Perspective*, pp. 162–172. Bohn Stafleu Van Loghum: Houten/Diegem.

Kuczmarski RJ, Carroll MD, Flegal KM and Troiano RP (1997) Varying Body Mass Index cutoff points to describe overweight prevalence among U.A. adults: NHANES III (1988–1994). *Obesity Research* 5: 542–548.

Löwik MRH and Brussaard JH (eds) (2002) EFCOSUM*: European Food Consumption Methods. European Journal of Clinical Nutrition 56(Suppl. 2): 51–596.*

Nelson M and Bingham SA (1997) The assessment of food consumption and nutrient intake. In: Margetts BM and Nelson M (ed.) *Design Concepts in Nutritional Epidemiology*. Oxford: Oxford University Press.

Network for the Pan-European Food Data Bank based on Household Budget Surveys. DAFNE II (1998). Brussels: Commission of the European Union.

Roos G and Prättälä R (1999) *Disparities in Food Habits. Review of Research in 15 European Countries. FAIR-97-3096 Project*. Helsinki: National Public Health Institute.

Seidell JC and Flegal KM (1997) Assessing obesity: classification and epidemiology. *British Medical Bulletin* 53: 238–252.

Trichopolou A and Lagiou P (1998) *Methodology for the Exploitation of HBS Food Data and Results on Food Availability in Six European Countries*. EUR 18357 EN. Brussels: Commission of the European Union, Research DG/B.I-I.

WHO (1976) *Methodology of Nutritional Surveillance. Report of a Joint FAO/UNICEF/WHO Expert Committee*. World Health Organization Technical Report Series 593.

NUTS

R A Stephenson, Maroochy Research Station, Nambour, Queensland, Australia

Introduction

Human consumption of nuts has been recorded from the earliest times and nuts continue as an important component of many diets throughout the world. Despite this, relatively few nut crops have become major commodities on world markets. In fact, most of the lesser known nuts, particularly those from warm climates (**Table 1**), are restricted to traditional diets, usually within relatively narrow regional boundaries. Consequently, information on them, including their composition and nutritional value, tends to be superficial. Only two of the lesser-known nuts of warm climates – macadamia (*Macadamia integrifolia*) and pistachio (*Pistacia vera*) – have developed

Table 1 Some lesser-known edible nuts of warm climates

Name and origin of nut	*Features*
Almondettes (*Buchanania lanzan* Spreng., family Anacardiaceae) Southern Asia	Related to pistachio nut; medium-sized tree; black, single-seeded fruits; pear-shaped kernels, 1 cm, eaten raw or roasted, delicious (combination of almond and pistachio flavor); 51.8% oil, 12.1% protein, 21.6% starch, 5% sugars
Bunya nut (*Araucaria bidwillii*, family Araucariaceae) Australia	Large pine tree; large cones bear starchy nuts *c.* 5 × 3 cm, usually roasted, flavor resembling that of chestnuts
Candle (Tung) nut (*Aleurites* spp., family Euphorbiaceae) China, South-east Asia, Indonesia	Attractive tree; fleshy fruit, containing a single nut, globose, *c.* 3 cm; hard, thick shell; soft, oily white kernel (*c.* 5 g); must be cooked, moderately poisonous raw, often used for oil
Cut nut (*Barringtonia procera, B. edulis, B. novae-hiberniae*, family Lecythidaceae, Brazil nut family) Pacific region	Confusion with taxonomy until recently, variable nut characteristics, trees are easy to propagate from seed or cuttings, common in coastal villages, *B. procera* is a more prolific bearer than the other two species – the fruits are a bit smaller (61 cf. 99 g) but have a higher kernel recovery (9%) and hence higher kernel production per tree (up to 50 kg fruit, 5 kg kernel), irregular bearing up to three times a year, high moisture content when fresh so shelf-life is limited, tasty, nutritious kernels, good nutty taste and long shelf-life when processed, relatively low oil content
Elaeocarpus spp., including silver, blue or bush quandong *E. grandis*, blueberry ash or blue oliveberry, *E. cyaneus*, and *E. dentatus*, family Elaeocarpaceae) Australia, New Zealand, Solomon Islands	Over 200 species, often attractive flowering trees producing blue berries from 12 to 25 mm, usually cooked before eating
Finschia nut (*Finschia chloroxantha*, family Proteaceae) Solomon Islands	Relative of macadamia
Galip nut (*Canarium indicum* L., family Burseraceae, also including related Java almond, *C. commune* and pili nut, *C. ovatum* Engl., adoa/bush ngali found only on Bougainville Island, *C. salmonense*) and *C. harveyi* from the Solomon Islands, Vanuatu, and Fiji Philippines, Moluccas (Indonesia), Papua New Guinea, Solomon Islands, Vanuatu	Tall, buttressed tree; roundish, dehiscent fruit bearing a single 5.5 × 2 cm nut, yield up to 300 kg nut in shell, hard, nonperishable shell, triangular in cross-section, *c.* 3 g, delicious, sweet almond-like flavor; 70–80% oil, 13% protein, 7% starch; remove testa before eating, raw or roasted, oil used for cooking and cosmetics, oleoresin burnt for lighting and incense or used to caulk boats, commercial production in Western Melanesia *c.* 100 000 tonnes nut-in-shell (16 000 t kernel, kernel recovery at 16%) from 2 million trees
Galo nut (*Anacolosa frutescens* (Blume) Blume, family Olacacae) South-east Asia, Philippines	Erect shrub or tree, hard wood used for building, fruit is a drupe (up to 18 g); the green-colored pulp and kernel can be eaten fresh but boiling brings out the delicious taste of the fruit; kernel often roasted; dry pulp and kernel contain respectively 9.5% and 10.7% protein, 4.5 and 7.5% fat, 70.8 and 75.5% carbohydrates, 7.2 and 3.7% fiber, and 7.9 and 2.9% ash; kernel recovery is high (73–85%)
Gnetum nut (*Gnetum gnemon*) Papua New Guinea, Solomon Islands	Some cultivation
Heritiera littoralis Solomon Islands	Common on seashore; some records of being eaten
Kauri pine nut (*Agathis* spp., family Araucariaceae) Pacific region	Attractive rainforest tree; overexploited for timber; limited use as a food source in the Pacific region only
Lotus seeds (*Nelumbo nucifera* Gaertn., family Nymphaceae) Asia, Papua New Guinea	Water lily revered by Buddhists; white, single-seeded carpers embedded in flat-topped, fleshy receptacle which dries when mature so that l-cm seeds rattle in their cavity; bitter green embryos removed before eating; eaten raw before fully ripe; nutty flavor, roasted or boiled when mature; rich in vitamin C, 58% carbohydrate (starchy), 17% protein, 2.5% fat
Macadamia nut (*Macadamia integrifolia* Maiden and Betche, *M. tetraphylla*, family Proteaceae) Australia	Evergreen tree adapted to the fringes of subtropical rainforests; spherical kernel enclosed by a thick, stony shell in a fibrous single-sutured husk; high oil content (> 72%); distinctive, delicate flavor
Moreton Bay chestnut (*Castanospernum australe*, family Fabaceae) Australia Pacific region	Fine spreading tree, large glossy fern-like leaves, deep green but lighter underneath, thick pea-pods with 5-cm 'chestnut' seeds; must be soaked and cooked before eating
Neisosperma oppositifolia	Seashore habitats; small kernel

Continued

Table 1 Continued

Name and origin of nut	*Features*
Okari nut (*Terminalia kaembachii* Warb., family Combretaceae, also *T. catappa* L., the tropical, Indian or sea almond, and *T. impedians*) Papua New Guinea, Vanuatu, South-east Asia, India	Tall tree easy to propagate; leaves clustered at twig tips; yields up to 100 kg large (up to 200 g) flattened, ellipsoid fruit with up to 10% kernel recovery; yields once (March to August) but sea almond bears sporadically throughout the year in low latitudes; stony endocarp encloses white kernel (up to 15 g) consisting of a leaf-like coiled cotyledon, *c.* 8 × 2 cm, 1.5–10 g; delicate almond flavor; contain about 50% sweet, colorless, nondrying edible oil; eaten raw or roasted, dry for long-term storage
Omphalea queenslandiae Solomon Islands	Woody climber; kernel needs processing
Oyster nuts (*Telfairia pedata* (sm. ex Sims) Hook, family Cucubitaceae) Tropical East Africa	Woody stemmed climbing vine, dioecious; large (< 15 kg), deeply ridged, dehiscent gourds containing up to 140 pale yellow seeds (nuts) enclosed in a strong, bitter-tasting, fibrous husk; nuts are flat and circular (3–4 × 15 cm thick); washed, sun-dried and dehusked before eating; raw or roasted; palatable, flavor similar to Brazil nuts; nutritious, *c.* 62% fat, 27% protein, rich edible oil
Pandanus nuts (various species, *Pandanus jiulianettii* Martelli, family Pandanaceae) Papua New Guinea, Pacific region	Wide ecological adaptation from coastal areas up to 1800 m, dioecious 'screw' pine; large, dense multiple fruit (syncarp) up to 30 cm diameter and 16 kg; individual fruits (*c.* 50 or more) separate easily, up to 10 cm long, 1.5 cm in diameter; contain two to four sweet seeds with a coconut taste; oily endosperm; normally roasted, fruitlet flesh of some species can also be eaten; dry seeds for long-term storage; oil content from 10 to 60%; depending on species, hardy stems and leaves also used for handicraft, furniture, and building materials
Pangi, Sis (*Pangium* edule, family Flacourtiaceae) Papua New Guinea, Solomon Islands	Berry an important food source in parts of Papua New Guinea
Paradise (Sapucaia) nut (*Lecythis usitata* Miers., family Lecythidaceae) Brazil, Guyana	Tall Amazon rainforest tree, related to Brazil nut but with superior, sweet delicate flavor (N.B.: some species poisonous); large, dehiscent woody fruits containing 30–40 irregular, oblong nuts resembling Brazil nuts but more rounded with a thinner, softer shell; white, creamy textured kernels; highly nutritious, 62% fat, 20% protein; eaten raw or roasted; edible oil
Pistachio nut (*Pistacia vera* L., family Anacardiaceae) Central Asia, the Middle East and Mediterranean basin	Small, deciduous, dioecious tree; resin ducts through all tree organs; bunches of nuts near shoot tips; harvesting at the correct stage of maturity is critical; fleshy hull surrounds dehiscent, bony shell which encloses the kernel; high in carbohydrates, mainly sucrose (16%), oil, largely unsaturated (55%), and essential amino acids (25%)
PNG oak (*Castanopis acuminatissima*) Papua New Guinea	Of minor importance; sometimes cooked before eating
Quandong (*Santalum acuminatum* Sprague and Summerhayes, family Santalaceae) Australia	Small, semiparasitic trees of desert regions; globular, edible, fleshy fruits; kernel enclosed in pitted, stony shell; oily; harsh aromatic flavor; nutritious, 60% fat, 25% protein; generally roasted
Sago plum (*Cycas revoluta* and other *Cycas* spp., family Cycadaceae) Australia, Papua New Guinea, Pacific region and East Asia	Glossy, fern-leaved plant with sturdy trunk, slow-growing, primitive seed plant. A sago-like meal is made from large seeds; must be soaked for a long time before eating
Souari nut (*Caryocar nuciferum* L., family Caryocaraceae) Brazil, Guianas	Attractive, large tree; fruits round, soft-wooded capsules, *c.* 15 cm diameter, containing two to five large, brown, kidney-shaped nuts up to 5 cm long; edible yellow pulp surrounding nut; kernel enclosed in hard, woody, warty shell up to 1 cm thick, hard to crack; kernel has soft, white, sweet, almond-like flavor; eaten raw or roasted; edible oil
Tahiti (Polynesian) chestnut (*Inocarpus fagiferus* (Parkinson) Fosberg., family Fabiaceae) Pacific Islands	Moderate-sized tree; stout, large (10 × 10 cm) green/brown kidney-shaped, nondehiscent, single-seeded pods borne in terminal clusters; fleshy 'nuts,' boiled or roasted when nearly ripe, taste like chestnuts, palatable, sometimes hard to digest; staple food on some islands; moderately nutritious, 80% carbohydrates (starch), 10% protein, 7% fat

Continued

Table 1 Continued

Name and origin of nut	*Features*
Tallow nuts (*Ximenia americana* L., family Olacaceae) Widespread throughout Tropics	Densely branched shrub, usually deciduous; egg-shaped, juicy, fleshy fruits contain a large, oily seed; white kernels; palatability varies; nutritious, rich in protein and oil; eaten raw or roasted; cooking oil
Water chestnut, Chinese water chestnut (*Trapa* spp., family Trapaceae) Tropical Africa, Central Europe, eastern Asia	Aquatic plant; not a true nut; hard-shelled, woody fruit with four woody, spiny horns contains a single large white starchy kernel; eaten raw, roasted, or boiled; *c.* 16% starch, 3% protein; not particularly nutritious

as commercial crops and will be discussed in more detail. Some commercial processing and marketing of local *Canarium*, *Terminalia*, and *Barringtonia* species has been initiated in the Solomon Islands, Vanautu, and Papua New Guinea in recent years. The pistachio is a temperate crop but has been included because of its requirement for long, hot, and dry summer and autumn seasons for commercial yields of acceptable quality.

The Macadamia Nut

The macadamia is considered to be one of the world's finest gourmet nuts because of its unique, delicate flavor, its fine crunchy texture and rich, creamy color. These features, together with relatively low volumes of production, have led to the successful promotion of macadamia as a luxury dessert nut. Macadamia nuts account for about 2.2% of the world's total tree nut production – approximately 17 800 t of kernel in 1998–99.

Origin

The macadamia nut is the only commercial food crop indigenous to Australia, originating along the fringes of rainforests in south-east Queensland and north-east New South Wales, from 25 °S to 28 °S latitude and within 24 km of the Pacific coast. Of the three species of macadamia, only two are edible – the smooth-shelled *M. integrifolia* and the rough-shelled *M. tetraphylla*. Only the former has been developed commercially. The rough-shelled macadamia, although producing a raw kernel of excellent eating quality, often contains a higher percentage of sugars which may caramelize on roasting, thus detracting from its appearance. The wild *M. ternifolia* produces a small, unpalatable, bitter kernel.

Botanical features

The evergreen macadamia tree is medium to large, attaining a height of up to 18 m and a spread of up to 15 m. It produces a number of vegetative growth flushes per year, with peaks of flushing in late summer and spring in Australia. The oblong to oblanceolate leaves are arranged in whorls of three and often have spiny, dentate margins, and short (5–15 mm) petioles. Three buds are arranged longitudinally in the axil of each leaf. Multiple branches and inflorescences may therefore be produced from each node. The pendulous racemes, 10–15 cm long and bearing up to 200 creamy white flowers, are borne on older wood. Although floral induction occurs in autumn, cool temperatures induce a period of dormancy. Racemes commence growing slowly in late winter and anthesis occurs in early spring. Fewer than 5% of flowers set fruit and many of these abscise 5–6 weeks after anthesis, coinciding with the stage of endosperm development and the commencement of rapid nut growth. Nuts take 6 months to mature.

Nut set is enhanced by cross-pollination. It is therefore recommended that at least two varieties be grown in orchards, often in alternate rows. Activity of pollinating insects, mainly native and domestic bees, is encouraged.

The fruit is a globose follicle in which only one of two ovules develops. In some varieties, however, a small percentage of the nuts produced are twins, resulting from the development of both ovules. Twins are undesirable because of the difficulty of extracting whole kernels. Mature fruits usually, but not always, abscise when the fibrous husk is still green. As the husk dries, it splits along a single suture to release the nut, consisting of a hard, thick, rough, stony, light-tan shell that encloses the kernel.

The rough-shelled macadamia is readily distinguished from the smooth-shelled species. The leaf margins are more serrated, with up to 40 spines on each side, and whereas new leaf growth of *M. integrifolia is* pale green in colour, young *M. tetraphylla* leaves are an attractive pink to red color. Racemes are longer (up to 30 cm) and bear up to 500 reddish-pink flowers.

Ecology

The macadamia occurs naturally in the fringes of subtropical Australian rainforests. Temperature is the major climatic variable determining growth and productivity, the optimum being 25 °C. Although the

mature macadamia is capable of withstanding frosts as low as −6 °C for short periods, longer periods or lower temperatures may severely damage or kill mature trees. Developing inflorescences are particularly susceptible to frost damage. However, there is a low temperature requirement for flowering, the critical minimum above which flowering is suppressed being 20 °C. On the other hand, continuous and prolonged exposure to temperatures greater than 35 °C often produces chlorotic and sometimes distorted growth.

The macadamia tree has several features suggesting adaptation to relatively harsh environments, including sclerophyllous leaves and dense clusters of fine, proteoid roots which develop to enhance nutrient uptake from poor soils, particularly those low in phosphorus. The conditions required for optimum production, however, are quite different from those for survival. Macadamia can be grown in a wide range of soils but not on heavy, impermeable clays and saline or calcareous soils. The trees are most suited to deep, well-drained soils with good organic matter content (3–4% carbon), medium cation exchange capacity with pH of 5.0–6.0.

Production

The commercial development of macadamia occurred in Hawaii where most of the crop is sold locally to tourists as 'suitcase exports' and, to a lesser extent, exported to the mainland USA domestic market where they comprise fewer than 3% of all tree nuts consumed. Australia and Hawaii dominate the world market, with 42% and 33% of the world's total production respectively. Australia produced 7500 t kernel from 12 200 ha in 1998–99, followed by Hawaii (5900 t), South Africa (1700 t), Kenya (950 t), Malawi (700 t), Costa Rica (450 t), Guatemala (450 t), and Brazil (180 t).

Over the last decade, macadamia production has more than doubled and will continue to increase as young plantings throughout the world come into production. The already intense competition from established nut crops has highlighted the need to develop new markets and promote more widespread consumption.

Composition

The quality of the macadamia nut is related to its oil content and composition. Mature nuts contain at least 72% oil (specific gravity, or SG < 1.0) for optimum eating and processing quality. Kernels with SG of 1.0–1.025 are classed as second-grade but can be used for lower-grade products. Third-grade kernels (SG > 1.025) are commercially unacceptable. Oil accumulation does not commence until the nuts are fully grown and the shell hardens. It accumulates rapidly in the kernel during late summer when the reducing sugar content decreases.

The composition of mature, roasted, and salted macadamia nuts is shown in Table 2. As with many oil seeds, the protein is low in methionine. Fresh kernels contain up to 46% sugar, mostly nonreducing sugar. The oil consists of mainly unsaturated fatty acids and is similar in both species, although the proportion of unsaturated to saturated fatty acids appears to be slightly higher in *M. integrifolia* (62:1 compared with 48:1). Detailed fatty acid composition is shown in Table 3. The fatty acid composition and the absence of cholesterol may lead to the promotion of macadamias as a high-energy health food. The major volatile components in roasted macadamia kernels are apparently similar to those found in other roasted nuts, although little detailed information on these is available.

Table 2 Nutritive value per 100-g sample of macadamia nuts (roasted in oil and salted) and pistachio nuts (dried and shelled)

	Macadamia	*Pistachio*
Water (%)	2	4
Food energy (kJ)	3064 (c.732 kcal)	2465 (c.589 kcal)
Protein (g)	7.1	21.4
Fat (g)	78.6	50
Fatty acids (g)		
Saturated	11.4	6.1
Monounsaturated	61.1	33.2
Polyunsaturated	0.14	7.5
Cholesterol (mg)	0	0
Carbohydrate (g)	14.3	25
Calcium (mg)	46.4	135.7
Phosphorus (mg)	203.6	510.7
Iron (mg)	1.8	6.8
Potassium (mg)	332.1	1107.1
Sodium (mg)	264.3	
Sodium (unsalted raw kernel) (mg)	7.1	7.1
Vitamin A		
IU	Trace	250
RE	Trace	25
Thiamin (mg)	0.21	0.82
Riboflavin (mg)	0.11	0.18
Nicotinic acid (mg)	2.14	1.07
Ascorbic acid (mg)	0	Trace
Magnesium[a] (mg)	0.12	
Zinc[a] (mg)	1.4	
Manganese[a] (mg)	0.38	
Copper[a] (mg)	0.33	

[a]From Wenkam NS and Miller CD (1965) Composition of Hawaii Fruits. Hawaiian Agricultural Experiment Station bulletin no. 135. Honolulu.
RE, retinal equivalents. International units, (IU) are used to express vitamin A activity based on a bioassay technique which is calibrated with actual concentration in mg g^{-1}.
Data from Gebhardt SE and Matthews RH (1989).

Table 3 Fatty acid composition (methyl esters) of macadamia nut oil

Fatty acid type	*Content (%)*
Oleate	67.14
Palmitoleate	19.11
Palmitate	6.15
Eicosenate	1.74
Stearate	1.64
Arachidate	1.59
Linoleate	1.34
Myristate	0.75
Laurate	0.62

Data from Cavaletto CG, Dela Cruz A, Ross E, and Yamamoto HY (1966) Factors affecting the stability of macadamia nuts 1. Raw Kernels. *Food Technology*. 20:108–111.

Harvesting and Storage

Macadamia nuts fall from the tree naturally when they are mature. They are either hand-harvested or picked up by machines; efficient machine harvesting requires a smooth, clean orchard floor.

Freshly fallen mature nuts may contain 25% moisture. The husk must therefore be removed as soon as possible to prevent overheating, mold development, and deterioration in quality. The dehusked nuts are initially dried, either artificially or air-dried, to 10% moisture or less before delivery to processors. The nuts are then further dried in silos to 10–15% moisture for longer-term storage, for most efficient cracking of the shell and thus more completed recovery of whole kernels. Drying is performed in stages (at 52 °C down to 45% moisture and then at 77 °C down to 15% moisture) to avoid adverse effects on kernel quality. After cracking and separation of shell from kernel, the product can then be lightly roasted and salted, or packaged raw in bulk in vacuum-filled, foil-laminate bags that help to prevent development of rancidity. Traditionally, kernel has been roasted in coconut oil, although dry-roasting techniques are now more popular. The packaged product is then kept in cold storage to prolong shelf-life. Under these conditions, kernel can be safely stored for at least a year.

The Pistachio Nut

The pistachio has a unique flavor and is popular as a gourmet nut, particularly in the USA. The crop has exacting environmental and cultural requirements for commercial production and these probably limit industry expansion.

Origin

Although the popularity of the pistachio has been recorded from the earliest times, its place of origin is uncertain. The genus *Pistacia* consists of 11 edible species which are widely distributed, including the USA, Africa, and central and eastern Asia. *P. vera* produces the largest nut and is the only species which has a dehiscent shell.

The pistachio appears to have originated on the slopes and plateau country of Iran, Turkestan, and Afghanistan south of the central Asian desert but is found growing wild as far east as India and Pakistan. It is widely cultivated in the Mediterranean basin, and more recently in Asia, the USA and, to a lesser extent, in Australia.

Botanical Features

The pistachio tree is small, dioecious and deciduous. It seldom exceeds 5 m in height and 10 m in diameter. The tree is hardy and can withstand extreme temperatures.

The leathery leaves with raised veins consist of one, three, or five leaflets. Juvenile leaves and those in environments with inadequate chilling have single leaflets, whereas mature leaflets in environments with adequate chilling have five. Resin ducts occur throughout tree tissues, and resin is exuded from any damaged parts.

In spring, inflorescent buds break first and there is often a delay of up to 7 days before vegetative buds break. Flowering may continue through to the latter part of spring, sometimes extending over a 2-week period, depending on variety and season. The transfer of pollen from male to female trees is by wind. Male pollinizer trees must be selected to flower at the same time as female trees in commercial orchards. One male tree per 10 female trees is required. The branched inflorescences are produced in clumps from several nodes just below the shoot tips. Bunches of nuts develop rapidly after pollination, the outer shell often attaining full size within a few weeks. The kernel, however, continues to develop over a 3–4-month period.

The nuts are mature when the outer skin of the husk changes from translucent to opaque. At this stage the husk is readily separated from the shell. The shells of good-quality nuts split to expose the edible kernel which is green or yellow in color with a green or reddish testa. Maturity of different varieties occurs at different times, up to 5 weeks apart, and each must be harvested at optimum maturity. Staining detracts from the appearance and the quality of the nut, resulting in much lower prices. Staining of shells is particularly a problem in humid or showery weather that also predisposes the development of aflatoxin-producing fungi. (*See* **Mycotoxins**: Occurrence and Determination.)

Apart from differences in time of flowering, maturity, and yield, varieties vary in the size and shape of

nuts produced, the percentage of split, empty, and stained shells, and the proportion of shell to kernel.

Ecology

Although essentially a subtropical, semiarid Mediterranean species, the pistachio is grouped with warm climate nuts because of its requirement of long, hot, and dry summers and autumns with an average daily mean of 30 °C for 3 months for good commercial yields. It requires a frost-free period of about 200 days to insure that developing inflorescences are not damaged in spring or by very cold winters. It does, however, require chilling during winter for satisfactory flowering. A total of 1000 h per year below 7.5 °C is required.

Production

The world production of in-shell pistachios in 1998–99 was 232 000 t, of which 157 000 t were exported. The main producing countries are Iran, Turkey, and the USA. The pistachio tree commences bearing at 5 or 6 years and increases from only 1 or 2 kg per year to levels of up to 50 kg of fresh nuts at 15–20 years. There is a tendency for pistachio to be a biennial bearing crop. The pistachio is particularly popular in the USA where it accounts for 9% of total nut consumption. Pistachios are usually marketed in shell and less frequently as kernel, often lightly roasted and salted.

Composition

The pistachio has a similar energy value to many nutmeats such as almonds but is lower than that of macadamia (Table 2). It is high in carbohydrate, mainly sucrose (16%). The fatty acids are essentially monounsaturated, although its polyunsaturated fatty acid content is higher and its saturated fatty acid content lower than in macadamia nuts.

The pistachio is high in minerals, particularly in potassium. A 100-g portion of pistachio kernels provides 92% of the daily thiamin requirement. Refer to individual nutrients.

Harvesting, Handling, and Storage

Mature nuts hang on the tree well under normal conditions and may be left until most are ripe. Under some situations, however, if the harvest is delayed, the husk dries on to the nut. Consequently the nut is stained and quality is lowered. Ripe nuts are readily dislodged by light shaking. Young trees may be harvested by hand but mechanical harvesting is normally used in commercial orchards.

The nuts must be husked and dried as soon as possible after harvesting to maintain high quality and an unblemished appearance and to avoid spoilage. The husked nuts may be sun-dried but drying artificially in silos between 65 and 72 °C is preferred for large commercial orchards. Freshly harvested nuts can contain as much as 45% moisture. This is reduced to about 5% in 10 h.

Some cultivars produce up to 25% empty shells in which kernels have failed to develop. These blanks must be removed from sound nuts prior to drying by flotation separation.

The dried nuts are sorted to remove those with blemishes. They are graded according to size, roasted, salted, and packaged. Most pistachios are marketed in their shells in snack packs.

Future Prospects for Lesser-Known Nuts of Warm Climates

Both pistachio and macadamia nuts are expensive compared with the major nuts. Future expansion of these industries may result in reduced prices and perhaps on promotion of the health benefits of nuts which is now generating a lot of nutritional research activity. As the volume of these commodities entering world trade increases, greater emphasis will be needed on market development and promotion to insure that they can compete with other nut products. It is unlikely that any of the other lesser-known nuts will develop significantly beyond their present regional boundaries.

See also: **Carbohydrates**: Classification and Properties; **Fatty Acids**: Properties

Further Reading

Anonymous (1999) World nut production and exports. *The Cracker* 3: Edition 28, pp. 33–52.

Cavaletto CG, Dela Cruz A, Ross E and Yamamoto HY (1966) Factors affecting the stability of macadamia nuts I. Raw kernels. *Food Technology* 20: 108–111.

Doster MA and Michailides TJ (1999) Relationship between shell discoloration of pistachio nuts and incidence of fungal decay and insect infestation. *Plant Disease* 83: 259–264.

Gebhardt SE and Matthews RH (1991) *Nutritive Value of Foods. Pacific Northwest Extension bulletin no. 357*. Washington State: Pullman.

Henty EE (1985) Some Nut-Bearing Plants in Papua New Guinea. In: Anon (ed.) *West Australian Nut and Tree Crop Association Yearbook 10*, pp. 19–27. Perth: West Australian Nut and Tree Crops Association.

Howes FN (1948) *Nuts. Their Production and Everyday Uses*. London: Faber and Faber.

Maggs DH (1982) *An Introduction to Pistachio Growing in Australia*. Melbourne: Commonwealth Scientific and Industrial Research Organisation.

Maskan M and Karatas S (1999) Storage stability of whole-split pistachio nuts (*Pistachia vera* L.) at various conditions. *Food Chemistry* 66: 227–233.

Menninger EA (1977) *Edible Nuts of the World*. Florida: Stewart.

Rosengarten F (1984) *The Book of Edible Nuts*. New York: Walker.

Stephenson RA (1990) Macadamia nut. In: Bose TK and Mitra SK (eds) *Fruits: Tropical and Subtropical*, pp. 490–521. Calcutta: Naya Prokash.

Stevens ML, Bourke RM and Evans BR (eds) (1996) *South Pacific Indigenous Nuts*. Proceedings of a workshop held from 31 October to 4 November 1994 at Le Lagon Resort, Port Vila, Vanuatu. *ACIAR Proceedings no. 69*, 176pp.

Vallilo MI, Tavares M, Aued-Pimentel S, Campos NC and Neto JMM (1999) *Lecythis pisonis* Camb. Nuts: oil characterisation, fatty acids and minerals. *Food Chemistry* 66: 197–200.

Verheij EWM and Coronel RE (eds) (1991) *Plant Resources of South-East Asia, no. 2. Edible Fruits and Nuts*. Wageningen: Purdoc.

Wenkam NS and Miller CD (1965) *Composition of Hawaii Fruits. Hawaiian Agricultural Experiment Station bulletin no. 135*. Honolulu: Hawaii Agricultural Experiment Station.

O

OATS

G A Hareland and F A Manthey, North Dakota State University, Fargo, ND, USA

Background

Oat is a multipurpose crop that has been grown throughout the world for centuries, generally in cool moist climates. Archeological discoveries have traced oat back to the Greeks, Romans, and Chinese from the first century, but the grain may have originated in areas surrounding the Mediterranean sea in countries of the Middle East. The world oat crop is diverse and includes thousands of commercial cultivars that are grown for multipurpose use. Oat has traditionally provided an inexpensive source of onfarm livestock feed, forage, and bedding. Approximately 70% of the world production of commercially grown oat is used for livestock feed, 20% for human food, and 5% for industrial usage. More than half of the oat crop never leaves the farm where it is produced. Oat is oftentimes grown as an alternate crop to break cycles of soilborne insects and crop diseases. Unlike most other cereal crops, oat has remained at a relatively low market cost and has been confined to growing on marginal soils associated with poor drainage and low fertility. Although oat production has decreased in recent years, high-quality oat is still in demand for human consumption. Studies have shown that oat is nutritious and contains physiologically active fiber components that aid the process of digestion. Oat bran has been reported to have positive effects on lowering serum cholesterol levels in humans. Compared with other cereal grains, however, the functional properties of oat have not been well defined. Breeders can now evaluate gene combinations by using genetic mapping methods to study the world oat collections from which the functional and nutritional merits of oat protein, oil, starch, and bran components will be better understood. This article presents an overview of oat production, breeding and genetics, anatomy, biochemical and nutritional quality, and utilization.

Oat Production

Oat is considered one of the major cereal grains that are grown throughout the world. Yet over the past 10 years there has been a substantial decline in worldwide oat production, primarily because of low price compared with other cereal grains. Between 1994 and 1999, the six-year average worldwide production was 29.5 million metric tonnes (Table 1) compared with an average of 42.8 million metric tonnes between 1986 and 1989. The Russian Federation recorded the highest percent, 40.7%, of the world total area harvested and the UK recorded the highest average yield of 5.81 t ha^{-1} between 1994 and 1999. The 20 countries listed in Table 1 accounted for approximately 92.7% of the area harvested and 90.9% of the world production between 1994 and 1999.

Oat Breeding and Genetics

Oat (*Avena* spp. L.) belongs to the grass family, Gramineae. *Avena sativa* L. (Figure 1), often referred to as common spring or white oat, is a hexaploid containing three distinct diploid genomes, AACCDD, where $2n = 6x = 42$ chromosomes. *A. sativa* is the principal cultivated species grown throughout the world. *A. byzantina* Koch. is a red-oat type adapted to warmer climates where it is grown as a winter oat.

A. fatua L. (Figure 2) is a closely related wild oat hexaploid species that has little or no economic value and is considered one of the worst and ubiquitous cereal weeds in the world. A derivative of a wild/hull-less oat genetic cross can make an attractive ornamental decoration (Figure 3). Other wild oat species of *Avena* include the tetraploids ($2n = 4x = 28$ chromosomes) and diploids ($2n = 14$ chromosomes). The genetic variation among the approximately 21 000 oat genotypes, wild oat species included, has been explored only to limited extents.

Naked (hull-less) oat, *A. nuda* (Figure 4), is an agronomic variant of *A. sativa* and threshes free from its hulls (lemma and palea) during harvest. Naked oat is a species that originated in China

and has been grown for centuries. A single dominant gene controls the naked trait but the degree of nakedness is affected by modifying genes and environmental factors. Naked oat is high in energy, protein, and nutrition. The metabolizable energy content of the groat is comparable to that of corn, which makes naked oat suitable for animal feed rations.

Oat-breeding efforts continue to focus on the production aspects of the crop, usually through hybridization, such as selecting for improved yield, winter hardiness, lodging and disease resistance, plant maturity, hull–groat relationships, seed size, test weight, plant height, and protein content. Breeders are characterizing the germplasm of the thousands of accessions of oat in attempts to enhance specific traits such as protein, oil, beta-glucans, and other nutritional qualities. Genetic maps are being developed to identify molecular markers that are linked to quality and disease resistance traits. The Genetic Resources Information Network (GRIN) database of the US Department of Agriculture disseminates information on oat accessions to breeders for improving the overall quality of oat. Future breeding objectives will likely focus more on oat as a food, as research reports tend to characterize the positive nutritional and physiological properties of oat and oat products in human diets. Future demands by

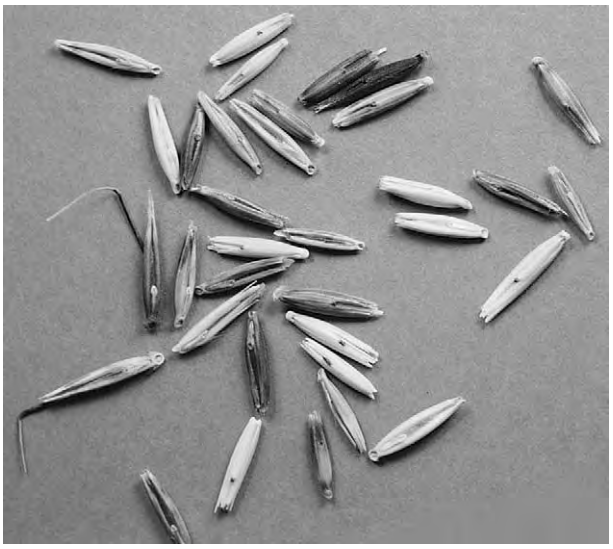

Figure 2 Whole kernels of the wild oat species, *Avena fatua*. Note the characteristic long awns that are attached to two of the kernels.

Figure 1 (see color plate 111) Oat panicle of cultivated *Avena sativa*.

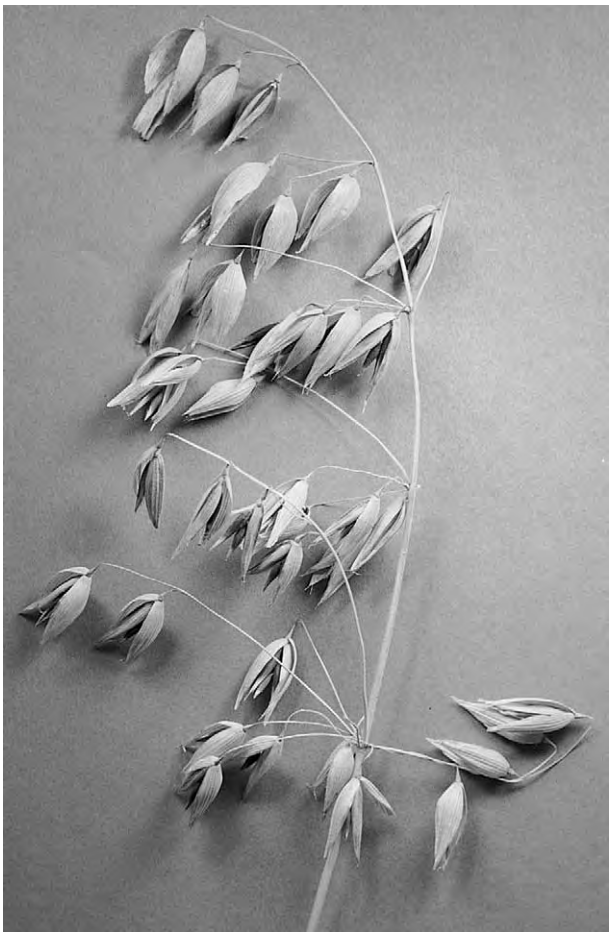

Figure 3 (see color plate 112) Oat panicle of *Avena* spp. From a genetic cross between wild and hull-less oats. The oat derivative can be used for decorative purposes.

Table 1 Six-year average area, yield, and production of oat (1994–99)[a]

Country	*Area harvested ($\times 10^3$ ha)*	*Percent of world total area harvested*	*Average yield (t ha^{-1})*	*Average production ($\times 10^3$ t)*	*Percent of world total production*
Argentina	245	1.5	1.46	368	1.2
Australia	941	5.8	1.50	1416	4.8
Belarus	324	2.0	2.07	672	2.3
Brazil	194	1.2	1.14	216	0.7
Canada	1482	9.1	2.47	3662	12.4
China	416	2.6	2.01	833	2.8
Finland	363	2.2	3.20	1157	3.9
France	141	0.9	4.36	614	2.1
Germany	308	1.9	4.86	1486	5.0
Italy	146	0.9	2.36	345	1.2
Kazakhstan	389	2.4	0.86	335	1.1
Poland	598	3.7	2.49	1485	5.0
Romania	246	1.5	1.55	380	1.3
Russian Federation	6637	40.7	1.17	7838	26.6
Spain	382	2.3	1.36	522	1.8
Sweden	310	1.9	3.66	1132	3.8
Turkey	154	0.9	1.78	276	0.9
Ukraine	550	3.4	1.78	989	3.4
UK	103	0.6	5.81	594	2.0
USA	1190	7.3	2.08	2476	8.4
% of world total		92.7			90.9
World total	16317		1.81	29465	

[a]Values from FAO Statistical Database, Food and Agriculture Organization of the United Nations (http://apps.fao.org/). Included are countries where average area harvested was greater than 100 $\times$ 10^3 ha.

food processors will encourage breeders to develop cultivars with improved milling qualities. Oat cultivars are likely to be adapted to different levels of protein, starch, lipids, and gums, and to changing seed size and shape for easier processing. Oat contains protein of superior quality, highly digestible carbohydrates, and an abundant source of nondigestible fiber. As the functional properties of the biochemical components become more evident, breeders will be encouraged to evaluate other genotypes for their agronomic, nutritional, and quality traits.

Anatomy of Oat Kernel

The compound inflorescence of an oat plant, referred to as a panicle (Figures 1, 3 and 4), is a continuation of the stem, and terminates in a single spikelet. Development of spikelets involves the formation of several florets, of which primary and secondary kernels develop to maturity (Figure 5). Occasionally tertiary kernels develop, but are smaller than the primary and secondary kernels. The hull of an oat kernel is normally attached loosely to the caryopsis. The groat is that portion of the oat kernel after removal of the hull (Figure 6). The major anatomical features of the oat kernel are the hull and groat (Figure 7). The hull contains two tissue layers, a lemma and palea. The outer portion of the groat contains the bran layers, which consist of the pericarp, seed coat, and aleurone cells. The aleurone layer is developmentally and genetically part of the endosperm but adheres to the outer bran layers upon separation of bran from the endosperm during milling. The aleurone is usually a monolayer of cells that represent the thickest part of the bran fraction. Cells in the aleurone produce and secrete enzymes that degrade storage material in the endosperm. The embryo, also part of the groat, is made up of the scutellum and embryonic axis. The scutellum is located between the endosperm and embryo. Scutellum secretes enzymes during germination and is involved in food transfer from the endosperm to the embryo.

The proportional size of the hull and groat affects processing and impacts the economic value of oat. Oat kernels of uniform size and high groat percentages mill more efficiently and yield higher quantities of bran and endosperm than nonuniform and small kernels. Groat percentages range from 65% to 85% of the oat kernel and weigh from 10 to 40 mg in cultivated species. Groat weight of the primary kernel is higher than the groat weight of either secondary or tertiary kernels. Cultivated oat kernels are generally larger and have less hull content than kernels from wild oat species. Environmental growing conditions affect groat percentages. Thin kernels with high hull

Figure 4 (see color plate 113) Oat panicle of naked (hull-less) *Avena nuda*.

content and low test weight generally result from dry conditions after heading.

Oat hulls are cellulosic and fibrous materials that are high in crude fiber and low in caloric value. Hulls decrease bulk density during transportation, but function to keep the groat clean for processing purposes and to protect the groat from mechanical damage during thrashing. Hulls are removed during milling before processing groats into flour or rolled oats. Oat dehulling is accomplished by feeding whole kernels into the center of a high-speed rotor. The kernels are dehulled by the centrifugal force and impact against an abrasive rubber ring on the inside housing of the dehulling machine. The hulls are further separated from the groats by air aspiration based on differences in density. The efficiency of the dehulling process is affected by the quality of oat. Lightweight kernels that are thin or short, and double kernels in which the hull of the primary kernel envelops the secondary kernel, do not dehull as easily as larger and higher test-weight kernels. Whole oat is generally graded for size during the cleaning operation by separating thin and short kernels from larger kernels before dehulling.

The major fractions of the oat groat are the endosperm and bran layers. Their relative proportions are affected by both genotype and environment. Approximate proportions of kernel fractions from cultivated species have been estimated as follows: hull, 24–32%; groat, 65–85%; embryonic axis, 1–2%; bran layers, 27–41% with a total thickness of less than 0.1 mm; and starchy endosperm, 56–68%. The bran contains higher concentrations of vitamins, protein, lipid, minerals, and fiber while the endosperm contains higher concentrations of digestible carbohydrates. The ratio of bran to endosperm has an effect on the total chemical composition of the groat.

Biochemical and Nutritional Quality

Protein

Protein quantity varies within a single cultivar by 3–4% depending on environmental growing conditions and location. The approximate protein concentration in individual fractions of cultivated oat species is: groat, 12–25%; embryonic axis, 25–40%; scutellum, 24–32%, bran, 18–32% and starchy endosperm, 9–17%. Oat hulls contain less than 2% protein in cultivated species, but at higher concentrations in wild oat species. The embryonic axis and scutellum contain higher protein concentrations than other kernel fractions, but together make up only about 3% of the groat size. Thus, most of the protein is actually located in the bran and endosperm. The bran contains about twice the protein concentration as the endosperm, but only about half of the total groat protein because of the difference in relative size of bran and endosperm. Protein content in specific oat fractions varies among genotypes. This is most evident when comparing cultivated species with wild species, which normally have thicker hulls and smaller groats.

Oat groats contain protein of high quality. The amino acids of oat are important for their human nutritional quality; of special interest are the levels of essential amino acids, which the body cannot synthesize. Lysine, the first limiting amino acid in oat, averages about 4.2% of the groat protein, which is higher than other cereals but slightly below the

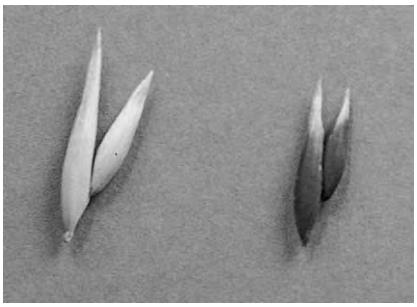

Figure 5 Kernels of cultivated white oat, *Avena sativa* (left), and a derivative of a wild *Avena* sp. (right). The primary (large) and secondary (small) kernels are attached.

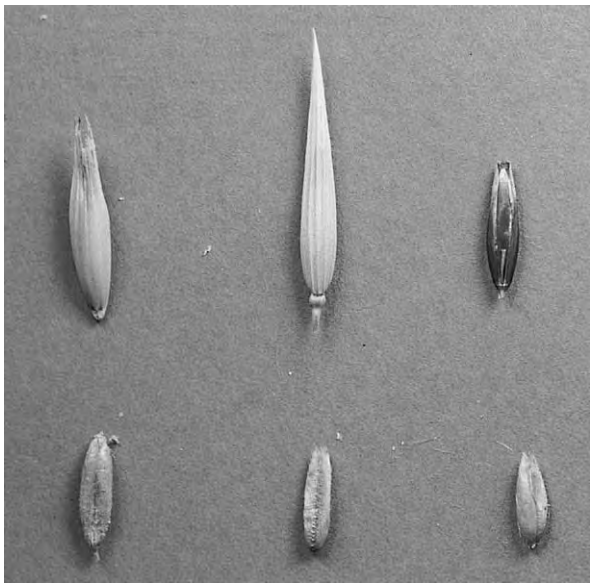

Figure 6 (see color plate 114) Kernels of cultivated white oat, *Avena sativa* (left), hull-less oat, *Avena nuda* (center), and a derivative of a wild *Avena* sp. (right). Whole oat kernels are pictured on the top row and the corresponding groats on the bottom row.

recommended Food and Agriculture Organization/World Health Organization reference standard of 5.5%. The embryo contains the greatest concentration of lysine. Oat ranks above other cereals in lysine content, including rice, because of the high protein concentration in the groat. Other essential amino acid concentrations associated within the groat are: methionine, 2.5%; valine, 6.4%; isoleucine, 3.9%; leucine, 7.4%; phenylalanine, 5.3%; and tryptophan, 1.7%.

Lipids

Oat contains relatively high lipid content compared with other cereal grains. Thus, oat serves as a unique source of energy as a livestock feed. Lipid concentration is genetically controlled and may vary from approximately 2% to 12% in cultivated species. Estimated percentages of free lipids in oat fractions are: hull, 2%; endosperm, 5.2%; aleurone and bran, 6.4%; scutellum, 20.4%; and embryonic axis, 10.6%. The embryonic axis and scutellum together contain the highest concentration of lipids in the oat kernel, but because of its relative size compared with other oat fractions, lipid quantity is low. The endosperm contains low lipid concentration, but contains over 50% of the lipids in the groat. The aleurone layer is very rich in lipid content and constitutes the major source of bran lipids.

Oat lipids are nutritionally important because of the high concentration of polyunsaturated fatty acids, especially linoleic acid. Linoleic acid, an essential fatty acid, is utilized in the synthesis of prostaglandins, which function to regulate smooth muscles like the heart. The approximate percentages of fatty acids in oat lipids are: myristic, 0.4–4.9%; palmitic, 15.6–25.8%; stearic, 0.8–3.9%; oleic, 25.8–47.5%; linoleic, 31.3–46.2%; and linolenic, 0.9–3.7%.

Oat and oat product quality may decrease upon storage because of possible chemical changes in the polyunsaturated fatty acids. Whole oat or undamaged groats that are stored below room temperature and at low moisture levels (less than 10%) show little change in quality. Conversely, damaged oat, groats, or ground oat that are stored for extended periods of time above room temperature and at high moisture have increased levels of enzymatic (lipase) activity. Lipase activity causes the release of free fatty acids from triacylglycerols, which are the major abundant class of oat lipids. Subsequently, unbound polyunsaturated free fatty acids such as linoleic acid undergo oxidation by action of lipoxygenase activity, leading to product rancidity. Lipase activity is highest in the bran because of its association with the aleurone layer, but can be minimized by thermally treating whole oat or groats during processing.

Carbohydrates

Starch is the major carbohydrate in oat and is found primarily in the endosperm. The amylose and amylopectin starch components are present in a ratio of about 1:3, respectively, which is similar to that of wheat and corn. High amylopectin, or waxy oat, and high amylose oat have not been reported in oat genotypes.

Oat contains an abundant source of nondigestible polysaccharides, which are carbohydrates that make up dietary fiber. The nondigestible polysaccharides are classified as either cellulosic or noncellulosic carbohydrates. Cellulose is a linear polymer of beta-1,4-linked glucose units, unlike the alpha-1,4-linked

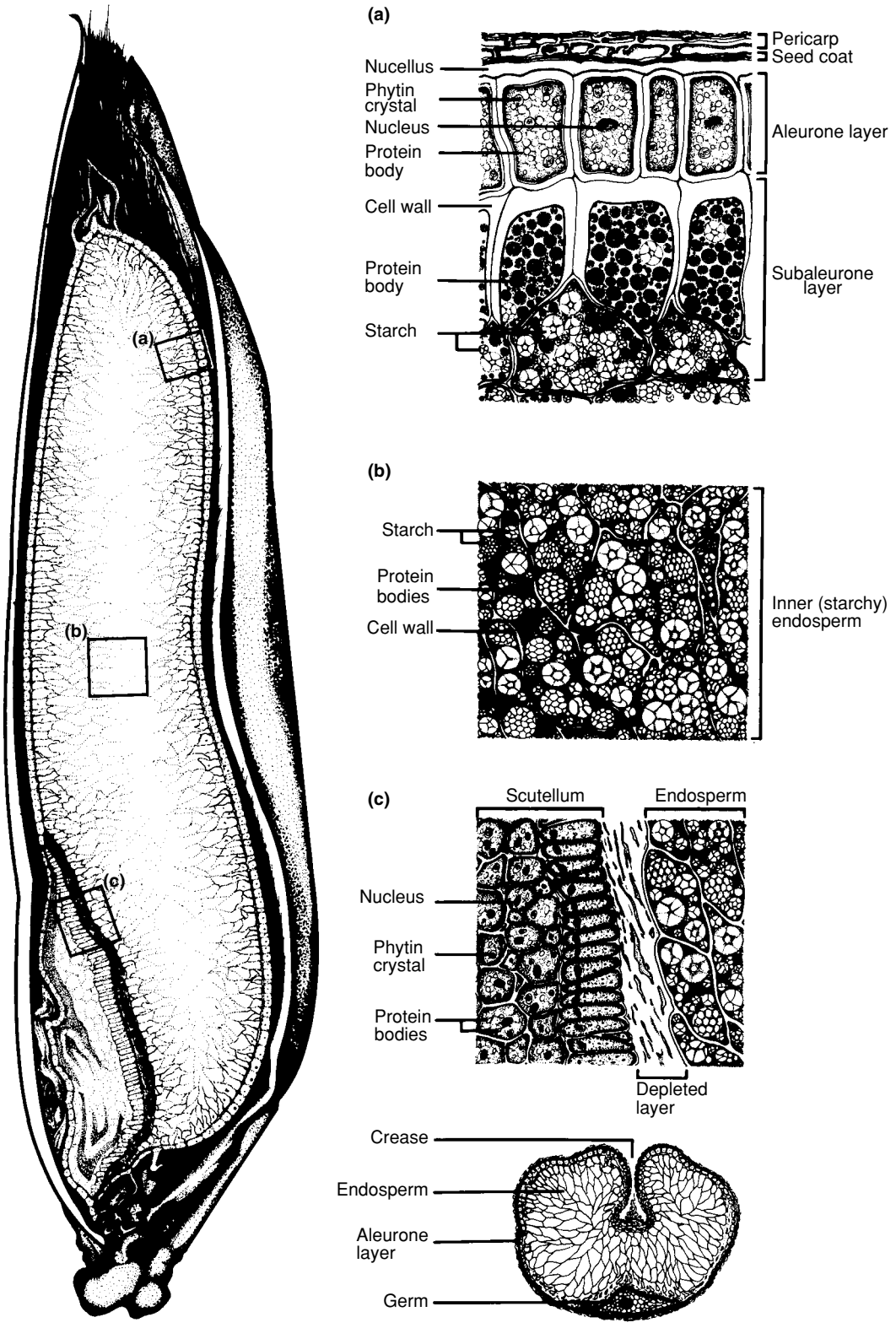

Figure 7 Major structural features of the oat kernel. Oat kernel on the left has been split longitudinally to reveal the approximate size and location of the major tissues. At the lower right is a cross-section view of the groat. (a), (b), and (c) are higher magnifications of portions of the bran, starchy endosperm, and germ. (Reproduced from Webster FH (ed.) (1986) *Oats: Chemistry and Technology*. St Paul, MN: American Association of Cereal Chemists, with permission.

glucose units that make up starch molecules. Noncellulosic polysaccharides include hemicelluloses, pectic substances, and gums and mucilage. Oat gum consists primarily of beta-glucans and pentosans. Beta-glucans are defined chemically as linear molecules of beta-1,3-and beta-1,4-linked D-glucopyranose units that are associated with cell wall structural components in both the bran and endosperm. Beta-glucans may be soluble or insoluble in water and contribute to the high viscosity of cooked rolled oats. Approximately 75% of oat beta-glucans are water-soluble. Pentosans are L-arabinofuranose residues linked as single-unit side chains to a backbone of beta-1,4-D-xylopyranose residues. Pentosans are present mostly in oat hulls and have been used in the commercial production of furfural. Upon acid hydrolysis of oat fiber, the polysaccharides yield sugars and sugar derivatives of D-xylose, L-arabinose, D-galactose, D-glucose, D-glucuronic acid, and 4-*O*-methyl-D-glucuronic acid.

Fiber-containing foods, upon ingestion, have a profound effect on food utilization from mouth to anus. Fiber directly affects gastric emptying time, rate of nutrient absorption from the intestine, fecal bulk, and frequency of bowel movements. Fiber indirectly affects pancreatic hormone secretions, hepatic glucose, and lipid metabolism. Under certain conditions, water-soluble beta-glucans have been reported to function in the small intestine as physiologically active, cholesterol-lowering ingredients. The hypocholesterolemic effects of beta-glucans tend to reduce the potential risk of ischemic heart disease. Oat hulls have little or no beta-glucan and are essentially water-insoluble fiber, which acts as fecal bulk in the large intestine. Total beta-glucan content in groats can vary from less than 2% to greater than 6%. Higher concentrations of beta-glucans are found in oat bran, which is obtained by separating endosperm and bran fractions by sieving. Beta-glucan content in oat bran is often higher than 7%.

Utilization of Oats

Oat is grown primarily for use as a forage crop and feed grain. Pliny, who reported on the human consumption of oat, stated in his *Natural History* that the Germanic tribes of the first century made 'their porridge of nothing else.' Oat later gained wide acceptance in Ireland and Scotland, where a variety of porridges were made. During the nineteenth century in America, oatmeal was sold almost exclusively in pharmacies and was suggested as a food for the infirm. In the late nineteenth century, Ferdinand Schumacher, a German immigrant, helped pioneer oatmeal into the largest selling cereal in the USA. His innovations included large-scale domestic milling, movement of oatmeal from the pharmacy to the grocery, and the development of packaging, brand names, and promotional literature.

Oat and oat products are used extensively in a variety of commercial food products. Thermal processing of oat enhances a unique flavor and aroma that contribute to utilization in cereals and baked-food products. Whole groats and steel-cut (sectioned) groats are processed into flakes. The traditional rolled oats made from whole groats are thick flakes requiring long cooking times, whereas flakes made from steel-cut groats are thinner and require shorter cooking times. Steam applied to sectioned groats produces variations of quick and instant flakes, with cooking times of less than 3–5 min. Oat flour is utilized as an ingredient in a variety of food products, and contains antioxidant activity in products likely to undergo fat oxidation during storage. Oat bran, obtained by sieving coarsely ground groats, is utilized as an ingredient in a variety of hot and cold cereals, and in baked products as breads, cookies, and muffins. The high viscosity of oatmeal makes it useful as a thickener in soups, gravies, and sauces, and a meat extender in meat loaves and patties. Processed oat hulls are used in the production of low-calorie, high-fiber baked products. Studies have shown that oat can be included in diets of celiac patients. The prolamin storage proteins, which cause the chronic condition in the intestine of celiac patients, are denatured by thermally processing the groats. In Sweden, Mill Milk Oat Drink was developed as a supplement for people who lack the ability to produce the lactase enzyme for digesting lactose found in dairy milk. Nu-trim, a thermomechanically sheared product from oat fiber, was developed for its high soluble beta-glucan content and used as a fat substitute in certain low-calorie products.

Oat has numerous industrial applications and potential for other uses. Oat starch has adhesive properties as a glue extender. Oatmeal has been used for cosmetic purposes to absorb dirt and sebaceous secretions from the skin. Oat hulls have cariostatic properties to prevent dental caries. Oat hulls have been used in the production of furfural and other derivatives as components in petroleum extractions, resins, plasticizers, insecticides, and pharmaceuticals; they may be used as a brewery filter aid; a component for making linoleum; and a component for making antiskid tire treads.

See also: **Carbohydrates**: Classification and Properties; **Cereals**: Contribution to the Diet; Dietary Importance; **Fats**: Classification; **Protein**: Sources of Food-grade Protein

Further Reading

Coffman FA (1977) *Oat History, Identification and Classification*. Technical bulletin 1516. Washington, DC: US Department of Agriculture, ARS.

Webster FH (ed.) (1986) *Oats: Chemistry and Technology*. St Paul, Minnesota: American Association of Cereal Chemists.

Welch RW (ed.) (1995) *The Oat Crop: Production and Utilization*. London, England: Chapman and Hall.

Youngs VL and Forsberg RA (eds) (1987) Oat. In: *Nutritional Quality of Cereal Grains: Genetic and Agronomic Improvement*. Agronomy monograph no. 28. Madison, Wisconsin: American Society of Agronomy.

Youngs VL, Peterson DM and Brown CM (1982) Oats. In: Pomeranz Y (ed.) *Advances in Cereal Science and Technology*, vol. 5, pp. 49–105. St Paul, Minnesota: American Association of Cereal Chemists.

OBESITY

Contents

Etiology and Diagnosis

H Pande, Johns Hopkins Bayview Medical Center, Baltimore, MD, USA
L J Cheskin, Johns Hopkins University, Baltimore, MD, USA

Background

The prevalence of obesity world-wide is increasing rapidly. The WHO has designated obesity as a major health problem throughout the world. In 1995, there were an estimated 200 million adults classified as overweight and obese world-wide. By 2000, the number of overweight and obese adults had increased to over 300 million. The obesity epidemic is not restricted to industrialized societies. In developing countries, it is estimated that over 115 million people suffer from obesity-related problems.

In the USA, the incidence of obesity among adults increased by 50% from 1976 to 1994. The most recent data demonstrate a prevalence of overweight (BMI 25+) or obesity (BMI 30+) of 64% and of obesity (BMI 30+) of 30.5%. Nearly 80% of some groups, e.g. middle-aged black women, are overweight or obese.

In most countries of Western Europe, the prevalence of obesity in adults is 15–25%. In the UK, over 20% of women and 17% of men are obese. Obesity is also increasing in prevalence in Latin America and the Caribbean. In South-east Asia, the Middle East, China, Japan, and the Pacific region, a marked rise in obesity is being seen in all populations and is now being recognized as a major public-health problem.

Obesity is clearly associated with increased morbidity and mortality. It is estimated to cause between 280 000 and 325 000 deaths per year in the USA. In 1995, the direct obesity-related health care costs in the USA exceeded $50 billion – nearly 6% of the entire national expenditure for health care.

Etiology

A simple and straightforward definition of obesity is an excess of body fat. However, obesity is more than just an accumulation of excessive amount of fat and its cosmetic consequences. It is a chronic disease with serious health consequences, is a major cause of morbidity and mortality worldwide, and is now considered the most common disorder of nutrition in developed countries.

The etiology of obesity is complex, multifactorial, and incompletely understood, despite significant advances in recent decades. It is thought to result from a

complex interaction between genetic, environmental, metabolic, dietary, and behavioral factors.

The pathogenesis of obesity can be considered the sum of those biological factors, which increase the predisposition towards expansion of adipose tissue mass together with consequences of adaptation to an environment that promotes and possibly rewards increased food intake and decreased physical activity.

When daily energy intake is equal to daily energy expenditure, a state of energy balance exists, and the body weight stays constant. When energy intake exceeds energy expenditure, a state of positive energy balance exists, and body weight increases. Overweight and obesity are usually the result of months and years of positive energy balance, resulting in enlargement of fat cells.

Genetic Factors

It has long been appreciated that obesity runs in families. Familial studies confirm that BMI is highly correlated among first-degree relatives.

Twin studies have shown a similarity in BMI between twins, which is stronger in monozygotic than in dizygotic twins. This correlation was found to persist even when the twins were raised separately.

Adoption studies have shown a significant correlation between the BMI of biological parents and adoptees, but not between the BMI of adoptive parents and adoptees. These studies suggest that the heritability of obesity may be as low as 10% (adoption studies) to as high as 80% (twin studies).

Some of the most compelling evidence for the genetic basis of obesity comes from the discovery of five mutations that cause spontaneous obesity and diabetes in mice. The Agouti (*Ay*), obese (*ob*), fat, diabetes (*db*) and tubby (*tub*) genes have all been cloned. In 1994, the discovery of the *ob* gene and its product leptin in *ob/ob* mice was a significant advancement in our understanding of the genetics of body-weight regulation. Since then, substantial progress has been made in characterizing genes related to human obesity. Genes responsible for previously existing mouse obesity mutations, *Ay*, *db*, *ob*, *fat*, *tub*, have been identified and have led to an understanding of the physiological pathways underlying each mutation.

Leptin, for example, is a 16-kDa hormone produced by adipose tissue, which acts by binding to receptors in the hypothalamus, which in turn alters the expression of several neuropeptides that regulate neuroendocrine function, appetite, and energy expenditure. A decrease in body fat leads to a decreased level of this hormone, which in turn stimulates food intake. In addition, decreased leptin levels activate a hormonal response that is characteristic of the starved state, including reductions in the resting metabolic rate. Increased body fat is associated with increased levels of leptin, which act to reduce food intake. By this mechanism, weight is maintained within a relatively narrow range physiologically.

Mutation of the leptin gene, leading to deficiency of leptin, results in the development of hyperphagia, hypogonadotrophic hypogonadism, and severe obesity with childhood onset (functionally equivalent to the *ob mouse model*). So far, only two kindreds with defects in leptin have been reported. In humans, mutations have also been identified in the leptin receptor in the hypothalamus (equivalent of the *db mouse model*), in the melanocortin 4 receptor (MC4R) and proopiomelanocortin (POMC), which is a precursor for the natural ligand of MC4R (*Agouti model*). Mutation in the prohormone and neuropeptide processing enzyme carboxypeptidase E in the *fat* mouse model and its equivalent in human peptidase PC1 have also been described.

Although monogenic mouse models have played an important role in our understanding of the biology of weight control, these mutations are rare and do not explain the common genetic predisposition to obesity seen in most human populations. They help us in understanding why, within a relatively homogenous environment, some individuals are lean and others are obese. However, only a small fraction of human obesity is due to a deficiency of leptin; instead, human obesity is associated with increased plasma concentrations of leptin. Thus, human obesity is characterized by a central and/or peripheral resistance or decreased responsiveness to leptin. The biological effects of the daily peripheral administration of recombinant human leptin in obese humans are currently being determined in clinical trials but are not expected to be dramatic, considering the nonresponsiveness to the satiety effects of high physiologic levels of leptin in obese individuals.

So far, more than 30 syndromes with a Mendelian pattern of inheritance associated with obesity have been identified. Most of these are pleiotropic syndromes in which obesity is one of the several features.

Prader–Willi syndrome (PWS) is the most common and best-characterized human obesity syndrome with an estimated prevalence of 1:25 000 and autosomal dominant pattern of inheritance. In addition to obesity, it is characterized by hypotonia at birth, small hands and feet, hyperphagia that usually develops between 12 and 18 months of age, mental retardation, short stature, and hypogonadism. Although familial inheritance of PWS is sometimes described, the vast majority of cases are sporadic. It has been established that PWS is most often caused by

a deletion of the paternal 15q11.2–12-chromosome segment.

A second, less common disorder, Bardet–Biedl syndrome, which was previously incorrectly named Laurence–Moon–Bardet–Biedl syndrome, is an autosomal recessive disorder. It shares many characteristics with PWS, including an increased risk of obesity. Studies of affected families have identified several different chromosomal loci (chromosomes 16, 11, 3, and 5) responsible for the syndrome, suggesting that it is a heterogeneous condition, with obesity being one of the common features.

The more common forms of obesity are, however, polygenic, and do not display a Mendelian pattern of inheritance. It is a quantitative phenotype that is probably influenced by numerous susceptibility genes, accounting for variations in energy requirements, energy utilization and storage, and the metabolic characteristics of the muscles, as well as being strongly influenced by environmental factors.

The strong environmental pressure to consume calories in excess of the energy expended exceeds the capacity for homeostatic adaptation in genetically predisposed persons, leading to an energy imbalance favoring fat storage. Genetic determinants of inter-individual variation in obesity and related phenotypes are likely to be multiple and interacting, with most single variants producing only a modest effect. Genome wide linkage studies in obese families and in different ethnic groups have led to the establishment of the 'human obesity gene map,' which contains entries for more than 120 genes and 16 chromosomal regions in which published studies indicate a possible relationship to excess adiposity or related phenotypes.

Landmark discoveries of leptin, uncoupling proteins, and neuropeptides involved in body-weight regulation have highlighted the importance of molecular genetic factors in determining an individual's susceptibility to obesity. These factors alone, however, cannot explain the rapidly rising prevalence of obesity in the past few decades and the current obesity epidemic, as our genes have not changed substantially over this period. It is believed that our genes simply permit us to become obese; the environment, however, may determine whether or not we do, indeed, become obese.

Metabolic Rate

Metabolic rate is the function of genetic, medical and voluntarily modifiable factors. The total daily energy expenditure (TDEE) is the sum of resting metabolic rate (RMR), the thermic effect of food (TEF), and the energy expenditure in physical activity (EEact). EEact can be divided into in activities of daily living and energy expended in additional physical activity.

TDEE has been found to correlate with body weight. The contribution of TEF and RMR to TDEE is on average 10 and 60%, respectively. The available evidence suggests that there is no association, or at best a very small association, between RMR and TEF and a tendency to gain weight. Activity-related energy expenditure contributes to 30% of the TDEE, is the most variable component, and may be the component that predisposes to obesity and can be most readily modified during efforts in weight control. Recent evidence suggests that incidental physical activity, such as large muscle contractile activity, which is, unfortunately largely involuntary, is a significant predictor of TDEE.

Decreases in TDEE can lead to obesity only if energy intake is not reduced to match the new level of energy expenditure. Humans have an effective homeostatic system with the ability to finely match the energy intake to energy expenditure to avoid positive or negative energy balance and changes in body weight. The effectiveness of this system is illustrated by the fact that adults on average consume 1 million calories every year, and a persistent mismatch of even 1% (about 10 000 calories with energy intake more than expenditure) would lead to an accumulation of fat and an increase in body weight of about 30 pounds in a decade. However, usually the body weight of an adult human increases only slightly over the course of a decade, which means that the homeostatic system operates with a remarkable degree of precision.

Endocrine Causes

Endocrine disorders can upset the precise physiologic regulation of metabolic rate and intake normally operating to maintain a stable body weight. These disorders include hypothyroidism, growth-hormone deficiency, Cushing's syndrome, insulinoma, polycystic ovary disease, and primary empty sella syndrome: all can lead to increased body fat. With treatment, patients who are deficient, for example, in either testosterone or growth hormone show a reduction in visceral adiposity when their hormone levels are normalized.

Studies have shown that patients with nonendocrine precipitants of visceral obesity have increased cortisol and insulin secretion combined with low production of sex steroids, such as testosterone in men, and a low rate of secretion of growth hormone. However, these altered endocrine functions seem secondary to the obese state and at least partially

responsible for the development of complications related to the obesity. Weight reduction in general is followed by normalization of endocrine function.

Environmental Factors

Although genetic and metabolic-endocrine factors may explain some of the variability between individuals within a given environment, changes in the environment must be responsible for the increase in the prevalence of overweight and obesity in the population in the past few decades. The rapid increase in obesity that occurred from the 1970s through to the late 1990s suggests that the environment has changed to one that is now strikingly obesity-promoting.

An increase in body weight would occur only when environmental factors overwhelm the ability of the regulatory process to adjust energy intake to energy expenditure. Several changes in the environment have contributed directly to the inability to match energy intake to energy expenditure.

The availability and consumption of energy-dense, high-fat foods is one such factor. Numerous studies in laboratory animals and humans indicate that high-fat diets increase the risk of overeating and development of obesity, and reducing dietary fat results in reduced total caloric intake by reducing the probability of overeating. Moreover, the appetite-suppressant effect of leptin may be overridden by access to high-calorie foods. Body-fat storage also occurs at a greater rate when excess energy comes from fat than when it comes from carbohydrate or protein.

It is becoming clearer that, unlike total energy balance, fat balance is not as closely regulated. A positive energy balance most often results from disruptions in fat balance. It is known that the presence of excess dietary fat does not acutely increase the rate of oxidation of fat for energy. Also, unlike carbohydrates and protein, the capacity for fat storage in humans is virtually unlimited. Excess calories in the form of fat are readily stored as triglyceride in adipose tissue, with a very high efficiency.

However even with a high-fat diet, most, but not all, individuals will become obese. The ultimate amount of weight gained will depend upon the genetic makeup and nongenetic factors like voluntary physical activity.

Another environmental factor is the consumption of more total calories in food as a result of an essentially unlimited supply of convenient, high-energy-density, relatively inexpensive, highly palatable foods, along with an increase in the portion sizes. This increases the likelihood of overeating.

The fast-food and restaurant industry has played a major role in this by using enticing food advertisements, serving large portions of high-calorie foods, and encouraging social eating behaviors. Studies have demonstrated that many of the highly palatable food items that have been introduced into the market by the food companies are optimized to elicit a highly positive reinforcing response, ensuring their continued ingestion and incorporation into our menu of familiar foods.

The relationship between dietary fat consumption and obesity, though, is not universally accepted. The evidence for this association includes the close correlation across nations between the percentage of energy in the diet derived from fat and the prevalence of obesity in that country, as well as laboratory studies in animals and humans, which demonstrate that consumption of a high-fat diet *ad libitum* tends to cause excessive weight gain. However, the USDA nationwide food consumption surveys show that the average fat intake, and the average total calorie intake had decreased between 1977 and 1988. Similar surveys in UK from 1970 and 1990 and studies from France have shown that fat consumption has decreased, yet paradoxically, the prevalence of obesity in USA and other countries has increased.

By contrast, self-reported food intake surveys (National Health and Nutrition Examination Survey) suggests that while dietary fat intake in absolute terms has remained relatively constant over the period, since the total energy intake has increased, the proportional intake from fat has declined. It would therefore be misleading to conclude that dietary fat intake is not likely to play a role in the increase in prevalence of obesity. Obesity prevalence might have risen even faster, if not for the apparent dietary-fat restraint and availability of reduced-calorie food products.

Physical Activity

It is well recognized and appreciated that increased energy intake and decreased physical activity, independent of genetics, would inevitably promote the development of obesity in individuals.

Various studies have shown that there is, indeed, an inverse relationship between physical activity and obesity. Cross-sectional, population, and cohort studies have shown that a low level of physical activity predisposes to obesity, while a high level is protective. In addition, normal-weight, postobese women who report being nonexercisers gained more than twice as much weight over 4 years of follow-up as regular exercisers.

Our increasingly inactive lifestyle may play an important and perhaps dominating role in the increasing prevalence of obesity. There are data to support that

work-related physical activity and energy expended in activities of daily living has declined in the past few decades. In the USA, however, leisure-time physical activity has not changed much since 1985, with 60% of the adults reporting that they are not regularly active, and 25% reporting that they are not active at all.

The amount of energy required for daily living has also decreased, owing to an increase in sedentary activities such as watching television, VCR, remote control, computer interactions, listening to music, etc.

Data from the national personal transportation survey in the USA have shown that the number of annual walking trips has declined significantly, whereas the number of daily car trips has increased by nearly an identical amount.

Decreases in TDEE as a consequence of the above factors would lead to obesity if energy intake were not reduced to match the new level of energy expenditure, which seems to be happening in developed countries. Even additional bouts of exercise, if followed by a high-fat and high-energy diet, can totally negate the energy expenditure of the exercise.

An individual's predisposition to engage in physical activity may be influenced by their genotype. Variations in muscle metabolic characteristics, like oxidative capacity, may play a role in the pathogenesis of obesity. The increased proportion of fast-twitch (type II B) as opposed to slow (type 1) muscle fibers in obese individuals may be responsible for decreased exercise tolerance and increased perception of fatigue by some of these individuals, predisposing them to be less active and gain weight.

Social and Cultural Factors

Social and cultural factors also make important contributions to the pathogenesis of obesity. The range of BMI of a population varies significantly according to the stage of transition to market economy and associated industrialization of a country, urbanization, changing social structures, and socioeconomic status. In the initial stages of the transition, the wealthier sections of society show an increase in the proportion of people with a high BMI, whereas in the later phases of transition, a high BMI among the poor shows increasing prevalence. This is usually accompanied by an increase in childhood and adolescent obesity and obesity-related disorders.

Cultural factors like cuisine, entertaining, and food selection, fat content of diet, attitude toward health, energy intake, body image, and physical activity also play a role in influencing the range of BMI in a population.

Fetal Undernutrition

Maternal malnutrition causing undernutrition of the fetus during intrauterine development may determine the later onset of obesity, independent of genetic factors. In support of the hypothesis is the finding of an inverse relationship between birthweight and adult obesity, blood pressure, and type 2 diabetes, in both men and women, in later life, with the highest BMI, glucose concentrations, and systolic blood pressures being observed in those with the lowest birthweight. The mechanism of this is unclear.

Diagnosis

The diagnosis of obesity is generally not difficult, but is is important to both quantify its degree and determine the relative risk of complications based on body-fat distribution, and other risk modifiers such as personal or family history of complicating conditions.

Body Mass Index

The body mass index (BMI) is accepted as the standard measure of relative body weight, i.e., weight adjusted for height. The BMI was found to correlate significantly with total fat content in the body. It can be used to assess overweight and obesity and to monitor changes in body weight and determine the efficacy of a treatment program.

BMI is calculated as weight in kilograms divided by height in meters squared.

$$\mathrm{BMI} = \mathrm{weight(kg)}/\mathrm{height(m)}^2.$$

The formula using nonmetric measurements is

$$\mathrm{BMI} = 703 \times (\mathrm{weight\ (lb.)}/\mathrm{height\ (in)}^2).$$

These values are independent of age and sex. Calculating BMI is simple, rapid, and inexpensive, and can be applied generally to adults. Although there are a number of accurate methods by which to assess body fat (e.g., total body water, total body potassium, bioelectrical impedance, and dual-energy X-ray absorptiometry), the measurement of BMI is the most practical approach in a clinical setting. This is especially true because no trial data exist to indicate that one measure of fatness is better than any other for monitoring overweight and obese patients during treatment.

Based on BMI, a standard classification of the degree of overweight and obesity has been established (Table 1). A BMI of >25 confers increased risk for a number of health conditions. There is strong evidence that weight loss in overweight and obese individuals reduces risk factors for diabetes, coronary

Table 1 Classification of obesity according to the BMI

Classification	*BMI (kg m^{-2})*
Underweight	< 18.5
Normal	18.5–24.9
Overweight	25–29.9
Obesity	
Class 1	30–34.9
Class 2	35–39.9
Class 3	> 40

artery disease, strokes etc. While the risk of medical complications generally increases with obesity class, an important modifier is the distribution of excess adipose tissue, as estimated most conveniently by the waist circumference.

Waist Circumference

Body fat may be preferentially located in the abdomen (android obesity pattern) or surrounding the hips and thighs (gynoid obesity pattern). The android pattern often reflects an accumulation of fat surrounding the abdominal visceral organs. The presence of excess fat in the abdomen out of proportion to total body fat is an independent predictor of morbidity. Waist circumference is a clinically acceptable measurement for assessing an individual's abdominal fat content, and they positively correlate with each other. The sex-specific cut-off associated with increased relative risk for the development of obesity-associated risk factors, when the BMI is 24.9–34.9, is shown in Table 2. When the BMI is more than 35, the waist circumference is usually more than the cut-off and thus does not have any additional predictive value. The waist circumference is a better marker of abdominal fat content than the waist-to-hip ratio and is the most practical anthropometric measurement for estimating a patient's abdominal fat content before and during weight-loss treatment.

Weight-for-Height Tables

Historically, weight-for-height tables like the Metropolitan Life Insurance Company table were used to define the normal weight range. However, such tables have major limitations, such as a reliance on primarily white reference populations, use of unvalidated estimates of frame size, and derivation of the table from insurance company mortality data.

Assessment of Risk Status

Adults with a BMI of 30 or more have a 1.5–2.0 times relative risk of death from all causes compared with individuals with a BMI of 20–25, with most of the increase being due to cardiovascular causes. It is therefore reasonable to assess an individual's absolute risk status by looking for the presence of the following conditions:

Table 2 Relationship between waist circumference and the risk of obesity-related complications

		Risk		
	Waist circumference	*Low*	*Moderate*	*High*
Men	(in)	< 37	37–40	> 40
	(cm)	< 94	94–102	> 102
Women	(in)	< 32	32–35	> 35
	(cm)	< 80	80–88	> 88

The high-risk cut off for men is a waist circumference of > 102 cm (40 in), and the high-risk cut off for women is a waist circumference of > 88 cm (35 in).

Coronary heart disease Obesity is associated with coronary heart disease (CHD), primarily via its impact on the CHD risk factors of hypertension, dyslipidemia, and type 2 diabetes mellitus. Obesity may also directly confer some increased risk of coronary artery disease (CAD) though. Besides BMI, excessive abdominal fat has also been shown to have a positive correlation with morbidity and mortality from CAD. Body weight, independent of risk factors, is also directly related to the development of left ventricular hypertrophy and congestive cardiac failure, perhaps via increased cardiac workload, and changes in vascular resistance and blood return. Obese individuals should also be evaluated for other cardiovascular risk factors, especially smoking and family history of CAD, as these may magnify the obesity-associated risk.

Type 2 diabetes mellitus Excess body fat is associated with increased insulin resistance and risk of developing type 2 diabetes mellitus. This correlation is strengthened by the fact that a significant majority of people with type 2 diabetes mellitus are obese. Moreover, the dramatic increase in obesity during the past decade has been accompanied by a dramatic increase in the prevalence of type 2 diabetes mellitus. In fact, overweight and obesity, especially in those with a more central body fat distribution, are the major risk factors for the development of type 2 diabetes mellitus.

Obstructive sleep apnea Obstructive sleep apnea syndrome (OSA) is characterized by occlusion of the upper airways and cessation of airflow for at least 10 s during sleep. Obesity is believed to change upper-airway anatomy through increased deposition of periluminal fat. This leads to apnea, arterial hypoxemia, recurrent arousals from sleep, excessive daytime somnolence, and snoring. OSA is more common among the obese, with those having a BMI 30 or more being

at greatest risk. Patients with OSA can develop complications like neuropsychiatric and behavioral disturbances, systemic and pulmonary hypertension, right heart failure, myocardial infarction and stroke. There are data indicating that the symptoms of sleep apnea improve with weight loss.

Hypertension Obesity may influence the blood pressure and predispose to hypertension. There is strong and consistent evidence that weight loss produced by lifestyle modifications reduces blood-pressure levels.

Dyslipidemia Obesity is associated with increased total and low-density lipoprotein cholesterols, increased serum triglycerides, and reduced levels of protective high-density lipoprotein cholesterol. All these lipid abnormalities are associated with an increased risk of developing CAD.

Other disease conditions associated with obesity Other conditions include ischemic stroke, nonalcoholic steatohepatitis, gallstones and cholecystectomy, osteoarthritis, menstrual irregularities, and infertility. Increased body weight is associated with an increased risk for certain forms of cancer, including colon cancer, endometrial cancer, and postmenopausal (but not premenopausal) breast cancer. Maternal obesity is a significant risk factor for the development of gestational diabetes mellitus, and neural tube defects.

Physical examination and laboratory evaluation

The physical exam should include, in addition to anthropometry, a pulse and blood pressure measurement, examination of the thyroid, auscultation of heart and lungs, palpation of the liver and examination of weight-bearing joints. A laboratory evaluation will help further in quantifying the risk of obesity-related complications and in monitoring weight loss. It should include serum chemistry, lipid profile and thyroid-function tests. In certain high-risk individuals, an electrocardiogram, stress test, chest X-ray, sleep study, pulmonary-function tests, ultrasound of gallbladder, liver, and ovaries, etc., may also be appropriate.

Conclusion

The etiology of obesity has been discussed from several perspectives. The concept of a genetic predisposition to weight gain, the biological mechanisms involved in the regulation of body weight, and its individual variation have been discussed. This was followed by a review of our current appreciation of the regulation of energy balance, presented from the point of view that small daily errors in positive energy balance integrated over months and years increasingly lead people to obesity. The complexity of the interaction between genetics, behavior, and the environment was also outlined.

Most available evidence indicates that higher levels of body weight and body fat are associated with an increased risk for the development of numerous adverse health consequences. Medical history should evaluate etiological factors that may have played a role in the development of obesity. Physical examination and laboratory tests should extend this evaluation to risk factors associated with obesity.

See also: **Cholesterol**: Factors Determining Blood Cholesterol Levels; **Coronary Heart Disease**: Etiology and Risk Factor; **Diabetes Mellitus**: Etiology; **Energy Metabolism**; **Exercise**: Metabolic Requirements; **Hormones**: Steroid Hormones; **Hypertension**: Physiology; **Metabolic Rate**; **Obesity**: Fat Distribution; Treatment; Epidemiology

Further Reading

Anonymous (1998) *Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults – The Evidence Report*. National Institutes of Health (published erratum appears in *Obesity Research* 1998 6: 464).

Astrup A (1998) The American paradox: the role of energy-dense fat-reduced food in the increasing prevalence of obesity. *Current Opinion in Clinical Nutrition & Metabolic Care* 1: 573–577.

Bray GA (1999) Etiology and pathogenesis of obesity. *Clinical Cornerstone* 2: 1–15.

Campfield LA and Smith FJ (1999) The pathogenesis of obesity. *Best Practice & Research Clinical Endocrinology & Metabolism* 13: 13–30.

Chagnon YC, Perusse L, Weisnagel SJ, Rankinen T and Bouchard C (2000) The human obesity gene map: the 1999 update. *Obesity Research* 8: 89–117.

Flegal KM, Carroll MD, Kuczmarski RJ and Johnson CL (1988) Overweight and obesity in the United States: prevalence and trends, 1960–1994. *International Journal of Obesity & Related Metabolic Disorders* 22: 39–47.

Hill JO and Peters JC (1998) Environmental contributions to the obesity epidemic. *Science* 280: 1371–1374.

Hill JO, Wyatt HR and Melanson EL (2000) Genetic and environmental contributions to obesity. *Medical Clinics of North America* 84: 333–346.

Kopelman PG (2000) Obesity as a medical problem. *Nature* 404: 635–643.

Lonnqvist F, Nordfors L and Schalling M (1999) Leptin and its potential role in human obesity. *Journal of Internal Medicine* 245: 643–652.

Mokdad AH, Serdula MK, Dietz WH *et al*. (1999) The spread of the obesity epidemic in the United States, 1991–1998. *Journal of the American Medical Association* 282: 1519–1522.

National Task Force on the Prevention and Treatment of Obesity (2000) Overweight, obesity, and health risk. *Archives of Internal Medicine* 160: 898–904.
Sheehan MT and Jensen MD (2000) Metabolic complications of obesity. Pathophysiologic considerations. *Medical Clinics of North America* 84: 363–385.
Tremblay A (1999) Physical activity and obesity. *Best Practice & Research Clinical Endocrinology & Metabolism* 13: 121–129.
WHO (1998) *Obesity: Preventing and Managing the Global Epidemic*. Geneva: World Health Organization.

Fat Distribution

F Deeb, Johns Hopkins Bayview Medical Center, Baltimore, MD, USA
L J Cheskin, Johns Hopkins University, Baltimore, MD, USA

Background

While obesity is easily diagnosed in most cases, and is well known to increase the risk of a variety of disease conditions, less well appreciated is the basic physiology of adipose tissue deposition and the effects of different distributions of adipose tissue on heath risks.

Types of Human Obesity

Obesity is defined as an excess of body weight as a result of an excessive accumulation of fat and results from interaction between the environment and genes. A historically high-fat diet, in concert with decreasing levels of physical activity, has recently resulted in an epidemic of obesity and overweight in much of the world, which now affects at least 64% of Americans. Human adipose tissue can be classified by site into four types: subcutaneous, yellow marrow, interstitial, and visceral. Subcutaneous adipose tissue lies directly under the dermis of the skin and superficial to the fascia overlying the skeletal muscles. More than half of total body fat is generally found in this site. Subcutaneous adipose tissue can be further divided into a deep layer and a superficial layer. One clinical study found that during weight loss, the deep layer decreased to a greater extent than the superficial layer; the significance of this finding is unknown.

Interstitial adipose tissue is found between cells and usually accounts for only a small proportion of total body fat. Yellow marrow is the fat within the bone marrow and is not normally subject to much increase with weight gain.

Variation in the amount visceral adipose tissue (VAT) has the greatest impact on health risks. VAT can be subdivided into retroperitoneal and intraperitoneal adipose tissue. The intraperitoneal adipose tissue can be further divided into mesenteric and omental depots, whose veins drain into the portal vascular system. The liver, too, may accumulate fat, primarily within the hepatocytes. This may occur as a benign result of weight gain (hepatic steatosis or benign fatty liver) or as a manifestation of the often-progressive disease known as nonalcoholic steatohepatitis (NASH), which is often accompanied by obesity and/or type 2 diabetes.

A variety of terms have been used to define differences in adipose tissue accumulation patterns including android/gynoid, apple/pear, central/peripheral, and upper body/lower body. Of note, typical body-fat distribution differs between men and women. Women tend to accumulate more lower-body fat during weight gain, whereas men tend to have upper-body abdominal accumulation of fat, as well as variable lower-body fat accumulation. Purely lower-body weight gain is not generally seen in non-castrated males, though the typical male pattern of predominantly upper-body weight gain is not infrequent in women.

The metabolic consequences of obesity are determined to some extent by the distribution of fat. A number of animal and human studies have found that central obesity poses a greater risk of, among other conditions, cardiovascular disease, type-2 diabetes mellitus, hypertension, gallstones, and gout compared with the risk associated with peripheral obesity. The amount of abdominal visceral fat appears to be the most important corollary of health risk for the individual with obesity. In fact, even in the absence of overall obesity, an excess of VAT appears to be associated with excess health risks.

Regulation of Adipose Tissue Distribution in Humans

In view of the close epidemiologic and metabolic associations between central obesity and disease, the regulation of adipose tissue growth and distribution is important to understand. Age and gender are two important, nonmodifiable determinants of the size of the VAT depot compared to subcutaneous fat. Increasing age and male gender tend to increase VAT magnitude at a given level of total body fat. That is, weight gain in general increases VAT, but to a lesser degree in younger people and in women. However, genetic factors account for a good deal of variability in the amount of fat deposited in VAT or other sites. In addition, some environmental (modifiable) factors appear

important, and include smoking and lack of exercise. A number of endocrine abnormalities have been variably detected in association with central obesity, including low testosterone secretion in men, elevated cortisol and androgens in women, and low levels of growth hormone in men and women. Hypersensitivity of the hypothalamo–pituitary–adrenal (HPA) axis may be the mechanism underlying these abnormalities. Insulin resistance in peripheral tissues, leading to hyperinsulinemia, also plays a key role. All of these hormonal changes have important influences on adipose tissue metabolism and distribution.

At the adipocyte level, insulin and cortisol contribute to lipid accumulation by decreasing lipoprotein lipase (LPL) activity. Growth hormone (GH) and testosterone increase LPL activity, so that the decrease in these hormones often seen in central obesity would also tend to decrease LPL activity and favor further adipose tissue deposition. Because of the higher blood flow, cellularity, and innervation in visceral compared with subcutaneous adipose tissues, the consequences of the hormonal abnormalities described above are believed to be expressed more in visceral adipose tissue. Another factor accounting for the disproportionate health effects of VAT is that the number of androgen and cortisol receptors is higher in VAT than other adipose tissue deposits. Thus, endocrine abnormalities in VAT, e.g., elevated levels of insulin and cortisol, which favor accumulation of fat, and decreased levels of GH and sex hormones, which mobilize fat, are magnified by the higher receptor density, leading to further accumulation of fat in VAT. In various physiological conditions associated with increased visceral fat mass, the balance between the lipid accumulation hormones (cortisol and insulin) and the hormones that prevent lipid accumulation and instead activate lipid mobilization (sex hormones and GH) is shifted to favor the lipid-enhancing hormones. Such conditions include polycystic ovary syndrome, Cushing's syndrome, menopause, aging, GH deficiency, HIV lipodystrophy, and excess alcohol intake. Finally, local synthesis of steroid hormones in adipose tissue may play a role in regulating deposition of fat. Regional measurement of fat turnover reveals higher rates of lipid mobilization in upper-body (visceral) adipose tissue than in lower-body adipose tissue. The order in men is as follows: omental visceral = retroperitoneal > subcutaneous abdominal > subcutaneous femoral adipose tissue.

Effects of Fat Distribution on Health

As noted, body fat distribution is now recognized as an important predictor and modifier of many of the adverse health consequences of obesity. Upper-body obesity and abdominal visceral fat are associated with a variety of metabolic complications, including hypertension, insulin resistance, type-2 diabetes mellitus, dyslipidemia, gout, and premature coronary death.

Evidence suggests that abnormally high adipose tissue turnover in VAT, mediated by high levels of LPL and/or increased sensitivity to catecholamines or decreased sensitivity to insulin's effect on inhibiting lipolysis, results in elevated free fatty acid availability and may explain some aspects of the metabolic consequences of upper-body obesity. Overall, though, the precise mechanism(s) by which excess VAT lead to adverse health consequences is poorly understood.

Heart Disease

Obesity is an independent risk factor for the development of coronary artery disease (CAD). Obesity also increases the risk of CAD indirectly through its adverse effects on insulin resistance, lipid metabolism, and blood pressure. Central obesity is the pattern most highly correlated with an adverse coronary risk profile. Treatment of obesity results in an improved coronary risk profile. Despite the high prevalence of obesity in coronary populations, the effect of weight loss on hard cardiovascular outcomes such as myocardial infarction and death has received relatively little attention.

Endocrine Abnormalities

The endocrine and metabolic complications most frequently seen in obesity are insulin resistance, type 2 diabetes mellitus, dyslipidemia, menstrual abnormalities, and infertility.

Insulin resistance is quite common in severe obesity, and occurs especially with central obesity. Insulin resistance is a central pathophysiologic feature of multiple conditions, the conglomeration of which is referred to as 'Syndrome X.' Syndrome X includes hyperglycemia, hyperinsulinemia, dyslipidemia, and hypertension, which are associated with an increased risk of CAD and stroke. HPA axis hypersensitivity, as evidenced by an increased response to challenges at all levels of the HPA axis, from the adrenals to the central regulatory centers, appears to be associated with obesity, in particular, central obesity. Sex steroid and GH secretions are blunted, despite heightened HPA reactivity. The hyperandrogenicity frequently seen in obese women may be related to increased testosterone production in the ovaries and adrenals.

Obesity in girls is associated with early menarche and puberty. The effect of childhood obesity on boys is more variable; obesity can lead to either early or

delayed puberty. Pubertal gynecomastia is a common problem in obese boys.

Gastrointestinal Conditions

A gastrointestinal disorder for which obesity is a classical risk factor is cholelithiasis. The magnitude of risk is directly related to the severity of obesity but probably not to fat distribution. Since, for each kilogram of fat gained, an additional 20 mg of cholesterol is synthesized and excreted in the bile, this increased cholesterol turnover is the probable mechanism by which obesity increases the likelihood of precipitation of cholesterol gallstones. The risk for gallbladder stones increases during weight loss because of unfavorable changes in bile composition (lithogenicity), increases in the flux of cholesterol through the bile ducts as fat is mobilized, and decreased gallbladder contractility. New gallstone formation especially occurs in subjects who experience rapid weight loss after gastric surgery for obesity or after dieting with a very-low-calorie (< 800 kcal per day) and low-fat diet (as many as 25–35% of patients in some studies develop usually asymptomatic gallstones). The risk for gallstones appears to be lower in patients consuming a low calorie diet (800 kcal per day) containing 15–25 g of fat per day (4% after 10 weeks of dieting).

NASH, a liver disease with frequent sequelae of fibrosis and cirrhosis, occurs most commonly among obese patients and those with type 2 diabetes, and thus is likely to be more common in central obesity. Alanine and aspartate aminotransferase and alkaline phosphatase are the most commonly elevated liver enzyme levels in NASH but are usually no more than twice the upper limit of normal and do not correlate well with histological abnormities. The pathogenesis of liver injury in obese individuals with NASH is in part secondary to the proinflammatory cytokine, tumor necrosis factor α (TNF-α), which has emerged as a key factor in various forms of liver disease. Data from animal and clinical studies suggest that TNF-α mediates not only the early stage of fatty liver but also the transition to more advanced stages of liver damage. A modest weight loss (> 10% of body weight) can normalize liver tests, decrease liver fat content, and decrease liver size.

Another common gastrointestinal condition associated with obesity is gastroesophageal reflux disease (GERD). Abdominal obesity (and pregnancy) are especially associated with GERD, as these conditions result in increased intra-abdominal pressure, which is a central mechanism for reflux of gastric contents into the acid-sensitive esophagus. Abdominal hernias are also associated in particular with central obesity.

Pulmonary Conditions

Pulmonary conditions associated with obesity include obstructive sleep apnea (OSA) and obesity-hypoventilation syndrome (OHS). In OSA, patients suffer from repeated attacks of upper-airway obstruction during sleep. OSA is fat distribution-sensitive because it is believed to be associated with increased fat deposition in the tongue, uvula, pharynx, and hypopharynx. OHS is a condition associated with serious obesity in which an awake patient suffers from hypercarbia and hypoxemia. Abdominal obesity is a mechanical factor in the etiology of OHS, since increased abdominal fat tends to raise the diaphragm, as well as increase fat deposition within the diaphragm and intercostal muscles, and also increase chest-wall weight. Leptin, a protein produced by adipose cells, serves as a pulmonary growth factor and as a modulator of the central respiratory control center. Leptin deficiency or, more likely, leptin resistance is believed to play a role in OHS and OSA.

Assessment of Human Obesity

In order to assess an individual's body fat content and fat topography, a comparison of measures obtained in the individual with age- and gender-specific norms is necessary. This procedure has proved to be challenging in both clinical and research settings. At least three body-fat measures (total body fat content, upper-body fat, abdominal visceral fat) are very important, and must be evaluated not just in a laboratory environment but also in the physician's office and in large-scale population studies. Table 1 lists the current methods available to obtain a direct or predicted measure of total body fat content.

Anthropometry is the cheapest, most widely used method of assessing human-body composition. Anthropometry measurements can be used both clinically and in epidemiological studies to grade the degree of adiposity in either individuals or groups. The measurements are also used to describe the anatomic distribution of adipose tissue and thus classify individuals and groups with regard to the type of obesity (central or peripheral). The anthropometric measures considered most useful in assessing obesity include weight, stature, skinfold thickness, and circumference of the trunk and limbs.

Weight

Body weight is best measured to the nearest $\pm$ 0.1 kg using a beam-balanced scale. Despite the requirement of frequent and careful calibration, spring scales or electronic scales may be used instead. Patients should

Table 1 Methods of estimating body fat and its distribution

Method	*Cost*	*Ease of use*	*Accuracy*	*Measure of regional fat*
Height and weight	$	Easy	High	No
Skinfolds	$	Easy	Low	Yes
Circumferences	$	Easy	Moderate	Yes
Ultrasound	$$	Moderate	Moderate	Yes
Density/immersion	$	Moderate	High	No
Heavy water				
Tritiated	$$	Moderate	High	No
Deuterium oxide or heavy oxygen	$$$	Moderate	High	No
Potassium isotope (40K)	$$$$	Difficult	High	No
Total body electrical conductivity	$$$	Moderate	High	No
Bioelectric impedance	$$	Easy	High	No
Absorptiometry (dual-energy X-ray absorptiometry)	$$$	Easy	High	Yes
Computed tomography	$$$$	Difficult	High	Yes
Magnetic resonance imaging	$$$$	Difficult	High	Yes
Neutron activation	$$$$	Difficult	High	No

Adapted from Bray GA and Gray DS (1988) Obesity. Part I – Pathogenesis. *Western Journal of Medicine* 149(4): 429–441.

be measured either nude or wearing light clothing of known weight.

Stature

The preferred method of measuring stature is by using a wall-mounted stadiometer, measuring to the nearest $\pm$ 1 cm. Plastic stadiometers are available for clinical use that have acceptable accuracy. The position of the patient is very important and should be standardized, the patient standing erect with the head, shoulders, and buttocks against the stadiometer wall. Body weight and stature can be used to calculate the body-mass index (BMI) (or Quetelet's index) or weight corrected for height and to compare this value to percentiles for the distribution of this index by sex, age, and race tabulated from large surveys conducted by the National Center for Health Statistics. The BMI is calculated as the weight (in kilograms) divided by the square of height (in meters) or as the weight (in pounds) multiplied by 704 and divided by the square of the height (in inches). It is important to remember that age and body proportion (leg/trunk length ratio) influence the correlation of BMI with height and that BMI is correlated with lean as well as fat mass. This means that a very muscular individual will have a measured BMI that puts them in the obese category, although they do not necessarily have an abnormal percentage of body weight composed of fat.

Skinfold Thickness

This is a measurement of a double thickness or 'fold' of skin, including the underlying fascia and subcutaneous adipose tissue, which is taken using calipers at standard locations on the body. It is performed by pinching and elevating a skinfold at a specific anatomical site using the thumb and forefingers, and measuring the thickness of the fold with a specially designed caliper. The anatomic sites commonly measured are the triceps (midway between the shoulder and the elbow), the subscapular, suprailiac and paraumbilical area, and the medial thigh and calf. These measurements correlate with, but do not accurately measure, the actual mean thickness of the subcutaneous adipose tissue. This has been illustrated by comparisons of caliper skinfold thickness measurements with radiographic and ultrasound measurements of the true subcutaneous adipose tissue thickness at different anatomic sites. Reproducible skinfold measurement requires training. A major criticism of skinfold thickness methods is that they ignore variations in the amounts of internal adipose tissues, especially the visceral adipose tissue that is associated with so many of the increased health risks of obesity.

Circumferences

Body circumferences can be measured more easily in severely obese patients than skinfold thickness. Circumferences have an additional advantage over skinfold thicknesses in that they can measure internal as well as subcutaneous adipose tissue. However, unlike skinfold measures, circumferences are influenced by variation in the size of underlying muscles and bone. The most useful circumferences for grading or predicting body fat and for describing adipose tissue distribution are the upper arm, chest, waist or abdomen, hip or buttocks, and the thigh (proximal or midthigh). In men, the waist circumference is highly correlated with total body fat. In women, the hip or thigh circumference is more predictive of total body fat. Many circumference indices have been devised in an effort to capture the association between

fatty-tissue distribution and morbidity and mortality. The most popular circumference index is the 'waist/hip ratio' (WHR), followed by the 'waist/thigh ratio' (WTR). The WHR is an independent predictor of metabolic disturbance including insulin resistance, hyperlipidemia, hypertension, and arteriosclerosis. Similar associations have been reported for WTR, as well as for skinfold thickness indices. Recent evidence suggests, however, that the waist circumference alone is as reliable as the WHR in predicting VAT and its associated health risks.

Ultrasound

Ultrasound has the advantages of measuring subcutaneous adipose tissue thickness more accurately than skinfold measures and of being able to measure otherwise inaccessible sites. The two major disadvantages of ultrasound are that an experienced technician is usually required and that the costs are high.

Hydrodensitometry

An obese person floats higher in water than a non-obese person. This is because the density of fat mass is lower than that of fat-free mass (FFM). Underwater weighing is one of the oldest *in vivo* methods of analyzing human-body composition as a two-compartment (fat and fat-free body mass) model. It held the status of being the 'gold standard' for body-fat composition analysis for many years but is both inconvenient and not as precise a measure of body composition as other methods.

Bioimpedance and Conductivity

Total body electrical conductivity and bioelectrical impedance are techniques for predicting body composition based on the measurement of the electrical resistance in the body to a tiny current. This electrical resistance is proportional to body shape and the volume of conductive tissue (body components with high water concentrations such as the FFM and skeletal-muscle mass). As a result, these methods do not estimate body fatness but instead predict FFM; fat must be derived secondarily as the difference between the predicted FFM and the body weight. These tests have the advantage of being simple, quick, and inexpensive, and can be readily used in population studies. Their disadvantages are the variability in the measurement and its dependence on the hydration level of the patient.

Absorptiometry

Dual-energy X-ray absorptiometry (DEXA) was primarily designed to measure the amount of mineral within bone, but can also be used to estimate the amount of fat in soft tissues. DEXA is an outgrowth of dual-photon absorptiometric systems, having been introduced in the early 1990s as faster and more accurate. DEXA is quick and simple to carry out, yields information about regional adipose tissue distribution, because it is a whole body scan, but is costly.

Imaging

Imaging studies, such as computed tomography and magnetic resonance imaging (MRI), are widely considered the most accurate means currently available for assessing body-fat distribution, but the limitations are the high cost, expertise needed, radiation exposure, and limited utility in a research setting. These are the only methods available for precise, accurate quantification of visceral and other internal depots of adipose tissue. MRI is now the method of choice for calibration of field methods of measuring body fat and skeletal muscles *in vivo*.

Isotope Dilution

Dilutional methods include a group of specialized research techniques for measuring body composition of fluid compartments (e.g., total body water (TBW), extracellular fluid, and plasma volume) and the electrolytes within those compartments, such as exchangeable potassium, sodium, and chlorine. They are costly, require specialized training, and are not widely available. Examples are heavy water and neutron activation measures.

Heavy Water

Water labeled with either two isotopes of hydrogen (deuterium, $2H_2O$; tritium, $3H_2O$) or oxygen ($H_2^{18}O$) has been used to quantitate TBW by a dilution method in healthy and diseased individuals.

Neutron Activation

Whole-body counting/*in vivo* neutron activation analysis can be used to estimate adiposity components. This method enables all main anatomic elements to be quantified at the anatomic level *in vivo*. Once the elements have been measured, the proportion of fat, protein, water, and mineral can be calculated by applying simultaneous equations. These techniques are not widely available but are precise.

In summary, an increasingly well-recognized and understood modifier of health risk in obesity is regional fat distribution. While the ideal method of measuring fat distribution has not been defined, efforts should be made to identify those individuals who are at high risk because of an unfavorably high VAT mass. As a marker of health risk, regional

distribution of fat is arguably more important than obesity *per se*, and should be afforded the prominence it warrants in the evaluation and treatment of obesity.

See also: **Body Composition**; **Coronary Heart Disease**: Etiology and Risk Factor; **Diabetes Mellitus**: Etiology; **Hormones**: Pituitary Hormones; **Nutritional Assessment**: Importance of Measuring Nutritional Status; Anthropometry and Clinical Examination; **Obesity**: Etiology and Diagnosis

Further Reading

Bjorntorp P (1996) The regulation of adipose tissue distribution in humans. *International Journal of Obesity and Related Metabolic Disorders* 20(4): 291–302.

Bray GA (1996) Health hazards of obesity. *Endocrinology and Metabolism Clinics of North America* 25(4): 907–919.

Goran MI, Gower BA, Nagy TR and Johnson R (1998) Developmental changes in energy expenditure and physical activity in children: evidence for a decline in physical activity in girls before puberty. *Pediatrics* 101(5): 887–891.

Lo JC, Mulligan K, Tai VW, Algren H and Schambelan M (1998) 'Buffalo hump' in men with HIV-1 infection. *Lancet* 351(9106): 867–870.

Pietrobelli A, Wang Z and Heymsfield SB (1998) Techniques used in measuring human body composition. *Current Opinion in Clinical Nutrition and Metabolic Care* 1(5): 439–448.

Smith SR (1996) The endocrinology of obesity. *Endocrinology and Metabolism Clinics of North America* 25(4): 921–942.

Smith SR, Lovejoy JC, Greenway F *et al*. (2001) Contributions of total body fat, abdominal subcutaneous adipose tissue compartments, and visceral adipose tissue to the metabolic complications of obesity. *Metabolism* 50(4): 425–435.

Tilg H and Diehl AM (2000) Cytokines in alcoholic and nonalcoholic steatohepatitis. *New England Journal of Medicine* 343(20): 1467–1476.

Treatment

L F Donze and L J Cheskin, Johns Hopkins University, Baltimore, MD, USA

Introduction

The purpose of this chapter is to review critically the current state of knowledge about the treatment of obesity. We will discuss various treatment approaches, including the recommended multidisciplinary assessment and treatment; popular, commercial, and self-help approaches; and more aggressive methods such as very-low-calorie-diets (VLCDs), pharmacologic treatment, and surgery. We conclude with recommendations for improving the likelihood of long-term maintenance after weight loss.

Multidisciplinary Assessment and Treatment

Components of a comprehensive or multidisciplinary approach to weight loss typically include medical management, behavior modification, dietary modification, exercise modification, and long-term follow-up. All too often, only one or two elements of a comprehensive approach are considered or implemented in practice. Ideally, all five treatment components will be present in one program, so that overweight individuals will not need to coordinate their own care. True multidisciplinary treatment is most likely to occur in university- or hospital-based programs, with a treatment team consisting of physicians, psychologists, dietitians, and exercise physiologists. A typical treatment process is described below.

Multidisciplinary Assessment

Before treatment can begin, it is important for the treatment team to gain a good understanding of the patient's situation, both in regards to the weight problem and the individual as a whole. Such understanding is typically accomplished via comprehensive evaluations with the patient, including medical, behavioral, nutritional, and exercise/fitness assessments.

Medical assessment Medical evaluation is strongly recommended for obese patients with a body mass index (BMI) of 30 or more, those who have medical comorbidities, and/or individuals who have not seen a medical care provider in a year or longer. This assessment is critical for understanding the etiology of the disorder in each patient and for establishing a reference point for response to therapy. Medical assessment usually consists of a focused medical history, a careful physical examination, and appropriate laboratory tests.

The medical history should include an assessment of patient and family medical history, including obesity, cardiovascular disease (CVD), hypertension, cancer, diabetes, thyroid and other endocrine diseases, and dyslipidemia. Patient history should

include a detailed weight history (lowest and peak adult weight, weight changes and their precipitants, and current weight); history of previous weight-loss attempts; use of tobacco, alcohol, drugs, and medications; and level of motivation. Patients should also be screened for conditions that contraindicate or compromise participation in an exercise program, such as recent myocardial infarction, angina pectoris, disabling osteoarthritis (especially of the knees), severe obesity with restricted mobility, pulmonary disease, or traumatic injury.

The medical care provider should conduct a comprehensive physical examination, with special attention paid to signs of potential comorbidities such as type 2 diabetes mellitus, hypertension, dyslipidemia, and sleep apnea, as well as to possible causes of weight gain such as hypothyroidism, polycystic ovarian syndrome (PCOS), and other endocrine conditions. Blood pressure should be measured with a sufficiently large cuff to give an accurate reading, with the patient in a relaxed state.

Laboratory evaluations can serve as screening tests for certain complications associated with obesity. Blood chemistries should include fasting serum glucose, cholesterol, triglycerides, and liver profile tests. An electrocardiogram (ECG) and complete blood count and urinalysis should be performed to establish a baseline prior to treatment. Thyroid-stimulating hormone (TSH) levels should be checked if there is any suspicion of thyroid dysfunction. Other endocrine and metabolic tests can be performed, if indicated.

Behavioral assessment A clinical psychologist typically conducts the initial behavioral assessment of the patient. This indepth interview covers the patient's psychosocial history, including childhood experience; any trauma, abuse, or unusual events as a child or adult; perceived reasons for weight gain; educational, occupational, and relationship history; and any current stressors that may interfere with the patient's ability to prioritize or adhere to treatment. The patient's diet and weight history are also assessed, including periods of weight gain and losses, the patient's experience on previously attempted diets, patterns of weight maintenance or relapse, and motivation for weight loss. The patient's current eating behavior is assessed by discussing a 'typical day' with regard to eating (such as time of day, foods eaten, and reported hunger or other motivations for eating). Binge-eating behavior and other symptoms of clinical eating disorders are also assessed. Finally, the patient's psychiatric history and current mental status, including history of depression or other disorders, previous and current treatment with psychotherapy or medication, and current depressive (or other) symptoms, are explored.

Dietary assessment A formal dietary assessment is best done by a registered dietitian with training and experience in weight management. An initial nutrition evaluation may consist of a 24-h recall of food intake, a food frequency to determine adequacy and composition of the diet, anthropometric assessment (i.e., waist and hip measurements), and an interview assessing meal patterns, beverage consumption, food preferences, grocery shopping/cooking pattern, restaurant dining, and any religious or cultural customs that may affect the patient's diet. The nutrition assessment is necessary to evaluate the initial nutritional status of the patient and to make appropriate nutritional and treatment recommendations. Although many aspects of diet may be characterized as behaviors, understanding patients' taste preferences and the macronutrient composition of their usual array of food choices is useful in suggesting dietary and behavioral changes that are consistent with the patient's preferences and lifestyle.

The results of the dietary assessment should be interpreted cautiously, because both underreporting and restrained eating (while under observation) are common. Despite these shortcomings, the information gathered can be quite valuable. For example, if an individual reports drinking almost 1000 calories per day in juice and soda, simply switching to noncaloric beverages will likely promote a significant weight loss.

Exercise assessment The exercise/fitness assessment should be performed by a trained exercise physiologist and should explore the patient's usual degree of physical activity, any limiting factors such as joint disease or injuries, preferred types of activity, and measurement of the patient's current level of fitness. A fitness test may consist of a Harvard step test and a test of flexibility, and bioelectrical impedance testing may be used to assess body composition. A formal stress test is not required unless active cardiovascular disease is suspected. The exercise physiologist may also explore lifestyle factors such as usual work hours, social support, and nearby facilities or other resources that could promote increased activity level.

Metabolic testing Metabolic testing of the patient's resting metabolic rate (RMR) via a metabolic cart (indirect calorimetry) is recommended if the necessary equipment and staffing are available. A metabolic test may be performed by an exercise physiologist, dietitian, nurse, or technician. Resting metabolic rate (also called resting energy expenditure,

REE), obtained either by metabolic cart or estimation formulas, is then used to determine daily caloric needs by multiplying RMR by an activity factor ranging from 1.3 (for very sedentary) to 2.0 or more (for very physically active individuals). This estimate of total daily energy expenditure is used by the dietitian to determine the patient's initial caloric requirements and an estimated rate of weight loss, given a specific reduction in energy intake in the diet.

Summary The entire multidisciplinary assessment can be an emotional and draining experience for the patient, who may not have previously talked about (or even thought about) his or her life in such an indepth way, and to a team of clinicians. We have found, however, that only after a comprehensive evaluation by the multidisciplinary treatment team can an optimized and individualized treatment plan be recommended.

Multidisciplinary Treatment

After multidisciplinary evaluations are complete, treatment can begin. The patient may participate in individual and/or group sessions with members of the treatment team. Weekly treatment sessions are recommended initially (for at least the first month of treatment), to provide the support and education necessary for the patient to stabilize his/her eating behavior and implement the suggested dietary and physical activity changes. Once weight loss is under way, less frequent appointments may suffice.

Medical monitoring Ongoing medical management, including vital signs, brief review of systems, discussion of any adverse effects of the diet or weight loss, and periodic repeated blood tests, is recommended for patients on low-calorie diets (< 1200 calories per day) and/or those with serious medical comorbidities. Medication dosage or scheduling (for example, of orally administered glucose-lowering agents or insulin) may need to be adjusted as a result of energy restriction and weight loss.

Behavior modification and psychological treatment Behavior modification in the treatment of obesity is a methodology for systematically modifying a patient's eating, exercise, or other behaviors that may be contributing to the obesity. It is typically the psychologist on the treatment team who suggests or implements behavioral techniques, although these tools may be useful in all components of treatment. The six principles used in behavior modification typically include self-monitoring, stimulus control, contingency management, stress management, cognitive restructuring, and social support.

Self-monitoring is the detailed, daily recording of food intake (or other behaviors) and the circumstances under which the behavior occurs. Food and exercise diaries (**Figure 1**) are used to assess the patient's eating and activity habits, to increase the patient's awareness of his or her own behavior, and to promote positive behavior change. Patients may be encouraged to record time of eating episodes, type and amount of food eaten, feelings of hunger and satiety, place, context, and/or comments (such as, 'Out with friends at happy hour' or 'Not hungry but feeling deprived'). These records are reviewed in treatment sessions, so that specific feedback can be provided to the patient. Research has consistently demonstrated that self-monitoring is associated with improved treatment outcomes, and patients often report that it is one of the most helpful obesity management tools.

Stimulus control involves identifying the environmental cues associated with unhealthy eating and inactivity. Modifying these cues involves strategies such as only eating at the kitchen table and not in front of the television; keeping tempting, high-calorie foods out of the house (or at least, out of sight); and laying out exercise clothes the night before as a reminder to exercise in the morning. Research has suggested that obese individuals are more strongly influenced by external cues to eating than are nonobese people. Therefore, modifying or eliminating these external cues can reduce eating in response to them. For this reason, patients are encouraged to 'turn off all distractions' while eating, in order to reduce environmental and behavioral cues to eating, such as watching television, reading the newspaper, or driving.

Contingency management is the use of rewards for appropriate behavior changes, such as increasing frequency of exercise or reducing intake of fast foods. The rewards can be granted by the patient, the patient's family, and/or the weight management center; however, rewards must be of value and reinforcing to the patient in order to be effective. At some weight-loss programs, patients can earn monetary 'vouchers' towards treatment sessions and products for regular attendance and compliance with treatment recommendations.

Stress management involves the use of problem-solving strategies to reduce or cope with stressful events. Meditation, relaxation procedures, and regular exercise are effective techniques to reduce feelings of stress. However, stress management also involves helping the patient understand the cause(s) of his/her stress in order to make the necessary life or environmental changes to prevent ongoing, chronic stress, which can interfere with weight-loss efforts or exacerbate the weight problem.

SELF-MONITORING RECORD

Name Susan **Date** 1/17/01 **Weight** 205

Water (Please check) Each ○ equals one 240-milliliter glass.

✓ ✓ ✓ ✓ ✓ ✓ ○ ○ ○ ○

Eating Log (Please specify food & beverages consumed, time, and amount)

Time	Food or Beverage	Amount	Calories	Fat (g)	Context or Comments
8am	banana	1	105		
	yogurt	1	90		
	toast	1 slice	80		
10:30	donuts	2	410		at the office
12pm	large salad	1	20		light lunch
	dressing	1 T	50		because of donut binge
	diet coke	1	0		
7pm	crackers	a lot	? 300		starved — while making dinner
8pm	chicken	3oz	140		
	rice	1/2 c	100		
	beans (green)	1 c	35		
9pm	popcorn	4c.	160		watching TV bored

Daily Caloric Intake = 1490

Daily Fat Intake = _____

Physical Activity

Time	Activity	Duration
	no time today!	

Figure 1 Example of a self-monitoring record.

Cognitive restructuring techniques are used to identify and modify a patient's dysfunctional attitudes and beliefs about weight regulation and body image. Examples include the use of affirmations (such as positive self-statements) and visual imagery (such as seeing oneself eating and exercising appropriately or visualizing a realistic body size). Many overweight patients adopt 'all or none' or dichotomous thinking patterns, which need to be combated with more flexible thoughts such as, 'Even though I ate two donuts this morning, the day is not ruined, and I will recommit to sticking to my diet plan, starting now!'

The final behavioral principle is social support. Such support, usually from the patient's family or a support group, is used to maintain motivation and provide reinforcement for appropriate behavior changes. Facilitated support groups are often an integral part of comprehensive weight-loss treatment for these reasons, and they further serve to 'normalize' the ups and downs of the treatment process.

These six behavioral principles are immensely useful in helping patients adhere to a healthy diet and exercise program. However, behavior modification alone may not be enough to address all of the psychological complexities of obesity and associated eating problems. Many patients seeking weight-loss treatment are clinically depressed in addition to being obese. For patients who are introspective and amenable to therapy, psychotherapy may be

recommended, either as part of their weight-loss program or with a therapist outside the program. Depressed patients who are not interested in psychotherapy, cannot afford it, or are severely depressed may be prescribed an antidepressant by the physician on the treatment team or referred to a psychiatrist. Although depressive symptoms often improve with weight loss, adherence to treatment may be easier if the patient is not actively depressed. Individual or group psychotherapy may also be beneficial to address other issues related to obesity, such as low self-esteem, body image, and relationship problems. Binge eating or other eating disorders often coexist with obesity and may require separate treatment by a specialist.

Nutritional management Nutritional management of obesity entails an individualized approach, based on the dietary assessment, within a multidisciplinary setting whenever possible. Diet prescription for weight management involves a caloric deficit to promote weight reduction. For patients with mild to moderate obesity, a caloric deficit of at most 500–750 calories per day is recommended to promote a 0.5–0.7 kg weight loss per week. A low-calorie, individualized food-based diet that is either balance-deficit (reducing the total number of calories while keeping proportions from carbohydrate, fat, and protein basically the same as before) or a fat-deficit diet, with most of the caloric reduction resulting from restriction of fat, can be prescribed. The latter approach is preferable for Americans, whose typical diet is too high in fat. Also, a greater volume of food can be eaten on a diet that emphasizes complex carbohydrates and reduces fat intake to 20–25% of calories consumed. In either case, the focus of a calorie- reduced food-based plan should be on nutritional balance, with calories distributed appropriately among carbohydrates, protein, and fat, based on recommendations outlined in the US Department of Agriculture Dietary Guidelines for Americans or similar World Health Organization guidelines.

The diet must be realistic – that is, based on dietary modification and practical changes in eating habits. Nutritional recommendations should be determined by the patient's current eating habits, lifestyle, ethnicity and culture, other coexisting medical conditions, and potential nutrient–drug interactions. The patient should be advised to drink at least 1.5–2.0 l of water daily, unless contraindicated, e.g., by congestive heart failure, edema, or renal insufficiency. Patients should also be encouraged by the dietitian to self-monitor their food intake, which may include measuring portion sizes and recording and calculating calories, fat grams, and/or carbohydrate grams. If energy intake is prescribed below 1200 calories per day, daily supplementation is usually indicated to ensure adequate vitamin and mineral intake.

Suggesting gradual changes is helpful in altering diet composition. Depending on the initial quality of the patient's diet, the dietitian may focus on revamping one meal at a time, so as not to overwhelm the patient. For example, if a patient's usual breakfast consists of biscuits with gravy, bacon, and sausage from a fast-food restaurant, an alternative of home-prepared oatmeal, yogurt, and fruit may be suggested. Once the patient has incorporated this change, the dietitian may then move on to improving the lunch meal. It is better to recommend dietary changes that are feasible (and achievable) for a patient, rather than prescribing a diet that the patient will reject or not be able to follow. Similarly, reducing the fat content of milk or meats in stages (e.g., 2% to 1% to skim milk) gives the patient a chance to adjust to the new taste before further reducing to a lower-fat version.

Exercise treatment Although regular, moderate physical activity alone results in limited weight loss over the long term, it is an essential and high-priority component of any weight management program. Regular physical activity is thought to be the most important predictor of long-term weight maintenance. Research has shown that patients who diet and exercise regularly are much more likely to maintain weight loss than those treated with diet alone. When performed in conjunction with caloric restriction, regular, moderate physical activity achieves the following effects: increases 24-h energy expenditure; maintains (or minimizes loss of) lean body mass; reduces cardiovascular risk by producing beneficial changes in the lipid profile; has positive psychological effects; improves insulin sensitivity; and may provide other health benefits independent of weight loss.

In determining an appropriate, individualized exercise program, the exercise physiologist and the patient together should consider a plan that: (1) fitsinto the patient's schedule and lifestyle; (2) considers the patient's likes/dislikes; (3) makes use of the patient's resources; and (4) is based on the patient's current level of fitness. The exercise physiologist should ensure that, before beginning a fitness program, all patients can recognize and dealwith abnormal physical responses to physical activity.

Eventually, the exercise goal of any weight management program should be 30–60 min of continuous, moderate-intensity physical activity five to seven times per week. Until a patient can tolerate

30 min of continuous activity, several 10-min periods of physical activity throughout the day can help build the patient's endurance while still burning calories. Additionally, short exercise and rest intervals can be alternated. The intensity, duration, and frequency of activity should be progressively increased, according to the patient's increasing ability. Patients should be encouraged to self-monitor their daily physical activities, including type of activity, time performed, and duration, so that progress can be appreciated and the goal reset a bit higher from time to time. The exercise physiologist may serve as a personal exercise trainer for those who desire additional motivation and accountability.

Popular Weight-Loss Approaches

Many overweight and obese individuals do not seek or have access to the type of multidisciplinary treatment described above, so instead, they turn to popular or fad diets, commercial weight-loss programs, and/or self-help groups. These three types of approaches will be briefly reviewed in the following section.

Popular Diets

Numerous commercially advertised diets are available to consumers today that promise quick results and a 'magic bullet' approach to weight reduction. We believe that it is best to be skeptical, and instill such skepticism in overweight individuals, about diets that are not part of a comprehensive approach to weight management. Many popular diets are based on very limited menus, because monotony may help curtail consumption. Any severely reduced-calorie diet will initially cause diuresis, which makes the diet seem efficacious at first. This diuresis usually results in approximately 2–4% loss of body weight during the first 7–10 days, although most of this weight will be regained when the period of severe caloric restriction ends.

Many popular diets are based on the dieter counting and limiting calories or some other constituent (such as fat, sugar, carbohydrates), with the end result that energy intake is decreased. Another type of popular diet is a fixed-energy-level diet, which limits intake by controlling portion sizes, menu choice, and composition. Popular diets of this type include prepackaged and portion-controlled foods available in supermarkets or from commercial and nonprofit weight management programs. Other popular diets may use an energy-deficit approach (described previously) or prescribe moderate hypocaloric plans, such as 1800 calories per day for men and 1200 calories for women.

A current trend in the USA is that of high-protein, low-carbohydrate plans, such as the Atkins diet, Sugar Busters, and the Carbohydrate-Lover's diet (to name a few). These diets are not nutritionally balanced and have not been scientifically demonstrated to be effective or healthful in the long term. They promote the metabolic state of ketosis, resulting in a large initial diuresis, as described above, which will be reversed once carbohydrates are reintroduced into the diet.

Commercial and Self-Help Approaches

Self-help and commercial groups have the potential to offer the personal involvement and comprehensive approach needed to promote behavior changes in complex problems such as obesity. This potential, combined with the large numbers of people who can be reached, make these groups the most influential approaches to weight management currently available. On the negative side, these groups often generate expectations greater than can be reasonably achieved, sometimes through misleading advertising. In addition, evaluations and outcome statistics of these programs (if existing) have not been made public.

Commercial programs Many commercial (for-profit) programs provide information on dieting and nutrition and may include elements of physical activity, behavioral techniques, the provision of food, and group support. In the USA, we have seen the recent emergence of local, commercial weight-loss centers, run by physicians, in which the primary treatment is medication that is either prescribed or dispensed at the center. Often, preprinted diet plans are given to patients with limited, if any, nutritional or behavioral counseling.

Self-help groups Self-help groups gather people with weight and/or eating problems and operate at little or no cost, generally without professional intervention. These groups offer considerable social support but vary in their philosophy. A group such as TOPS (Take Off Pounds Sensibly) includes more behavioral principles, whereas Overeaters Anonymous is based on the 12-Step model used in addiction programs and focuses on 'compulsive overeating' as the core problem.

Lack of evaluation It is striking how little is known about the effectiveness of commercial and self-help approaches. A few scattered papers have been published on such groups, but the focus has been on attrition rates rather than on the weight

loss and maintenance of the participants. It is possible that the lack of evaluation is due to problems with confidentiality in the self-help groups and with protecting the financial interests of the commercial programs. The closed atmosphere that now surrounds these programs will probably remain so unless regulatory action forces a scrutiny of treatment results.

In light of the lack of evaluation of commercial and self-help programs, and the similar lack of research on the safety of many popular diets, we recommend that all self-initiated weight loss attempts be approved and monitored by the individual's primary physician or an expert in weight management, if the primary physician is not trained or experienced in this area.

Aggressive Approaches to Weight Loss

More aggressive weight management approaches are available under medical supervision and are appropriate for individuals who are severely overweight (generally BMI $\geq$ 35) and/or those with serious medical conditions related to their obesity. Obese patients who have failed multiple weight reduction attempts through traditional programs may also be considered for aggressive strategies. Examples of aggressive approaches include supervised VLCDs, pharmacologic treatment, and surgery.

Very-Low-Calorie-Diets

A medically supervised VLCD (fewer than 800 calories per day) may consist of food, commercially available liquid supplements, or a combination of the two. The first incarnation of these diets was in the form of egg- or milk-based protein combined with water. The use of these severely calorie-restricted diets in medically unsupervised individuals led to at least 58 highly publicized deaths, often resulting from electrolyte imbalances and sudden cardiac death. However, with experience and careful patient selection and monitoring, VLCDs have evolved into comparatively safe weight-loss techniques. It is recommended that patients being considered for VLCDs undergo a comprehensive medical examination and an ECG. Absolute contraindications include the presence of untreated or severe cardiac, hepatic, renal, or thromboembolic disease; type 1 diabetes; cancer; and eating disorders. Weekly medical supervision is recommended while on the VLCD. With full adherence, weight loss averages 1.1–1.8 kg per week, depending on body mass and physical activity level. The initial diuretic phase may be pronounced and accompanied by symptoms such as lightheadedness, headache, or fatigue. Later symptoms may include constipation and cold intolerance. The diet period is followed by progressive refeeding. Throughout the course of dieting, education on nutrition, exercise, and behavior modification will help the patient maintain his/her weight loss during and beyond the refeeding phase. For this reason, multidisciplinary treatment is strongly recommended in combination with the VLCD. Without this education and support, weight gain will likely occur almost as rapidly as weight loss.

Pharmacologic Treatment

Adjuvant anorectic medications may be useful in select obese patients (BMI $\geq$ 30 or BMI = 27–29 with at least one major comorbidity) who have been unsuccessful in previous coordinated weight-loss efforts. Commonly used medications in the treatment of obesity are listed in Table 1. Among the safest and most effective medications currently available are phentermine and sibutramine, both centrally active appetite suppressants.

Phentermine is thought to increase the release and inhibit the reuptake of norepinephrine (noradrenaline) and dopamine from nerve terminals in the brain, thereby suppressing feelings of hunger. The Food and Drug Administration (FDA) has approved phentermine for short-term use (that is, a few weeks to 3 months). Sibutramine, a serotonin and norepinephrine reuptake inhibitor, appears to decrease food intake and may increase energy expenditure as well. Sibutramine has been approved for use for up to 1 year, to aid in weight loss or maintenance of weight loss.

An entirely different mechanism of action is employed by the recently FDA-approved agent, orlistat tetrahydrolipstatin. Orlistat is an inhibitor of pancreatic lipase; thus, it blocks absorption of fat, resulting in about 30% of ingested fat being

Table 1 Medications currently used for treatment of obesity and example trade names

Centrally acting anorexic agents
Phentermine (Fastin, Adipex-P, Ionamin)
Sibutramine (Meridia)
Mazindol (Mazanor, Sanorex)
Diethylpropion HCl (Tenuate)
Phenylpropanolamine (Dexatrim, Acutrim, decongestants)
Ephedrine (decongestants, herbal weight-loss supplements)
Locally acting agents
Orlistat tetrahydrolipstatin (Xenical)

HCl, hydrochloride.

excreted in the stools, with predictable side-effects. Its efficacy in weight loss is probably not only due to the decreased absorption of calories, but also to the Antabuse-like effect when a high-fat eating episode occurs. Orlistat is to be used in obese adults three times per day before meals and has been shown in studies of up to 2 years' duration to improve weight loss and maintenance. Because the malabsorption of fat induced by orlistat also causes malabsorption of fat-soluble vitamins, use of a vitamin supplement is recommended.

Incremental weight loss in controlled trials of anorexic agents tends to be in the range of 0.2 kg per week greater than placebo. This is a definite benefit, but the risks of drug therapy (as illustrated by the fen-phen controversy) must be carefully balanced against the benefits. There is the potential for medications to be prescribed inappropriately, to individuals who do not meet criteria for use or beyond the duration of continued efficacy. Physiologic or behavioral tolerance to these medications, resulting in decreased effects, may occur over time, and some are not recommended, and have not been studied, for longer than 3 months. At best, antiobesity agents may provide certain obese patients with a 'jump-start' to weight loss, or they may be used to improve dietary adherence during a particularly difficult time in the course of treatment.

Surgery

Surgical treatment of obesity may be considered only in carefully selected patients who meet the following criteria: (1) a very high medical risk exists (BMI ≥ 40 or BMI $= 35$–39 with life-threatening or disabling comorbidities); (2) obesity has been present for at least 5 years; (3) no history of or current untreated alcoholism, drug abuse, or major psychiatric disorder is noted; and (4) the patient is between 18 and 65 years of age. In addition, patients should have failed previous attempts at medically supervised weight reduction and should have realistic expectations about the long-term outcome achievable with surgery. A multidisciplinary team approach to screening patients preoperatively, including evaluation with the surgeon, a dietitian, a psychologist, and the patient's primary physician, is critical to the patient's long-term success.

Two proven surgical options are available for the treatment of morbid obesity: purely restrictive operations such as the vertical banded gastroplasty (gastric stapling), and gastric bypass operations such as the Roux-en-Y gastric bypass. These procedures primarily involve either the mechanical restriction of caloric intake by creating a small gastric reservoir, or the induction of malabsorption by bypassing variable lengths of the small intestine. The vertical banded gastroplasty (Figure 2) achieves weight loss by creating a small pouch with a narrowed outlet in the stomach by using a band of synthetic material. With the gastric bypass procedure (Figure 3), a small pouch is created of 20–30 ml, which limits oral intake, and a moderate degree of malabsorption is created by a Roux-en-Y loop of bowel anastomosed to this small gastric remnant.

The Roux-en-Y gastric bypass produces more substantial weight loss than vertical banded gastroplasty and may have longer-lasting effects. The gastric bypass results in weight loss averaging 50–60% of excess body weight, with good weight control documented for up to 10 years. This procedure is not without risks, including post surgical complications, food intolerance, nutritional deficiencies, extreme psychiatric distress, and even death. However, in the majority of patients, obesity-related comorbidities, such as diabetes and hypertension, are either reversed or prevented after obesity surgery, as long as weight loss is maintained.

Maintenance of Weight Loss

Maintaining weight loss seems to be inherently more difficult than losing weight, particularly for

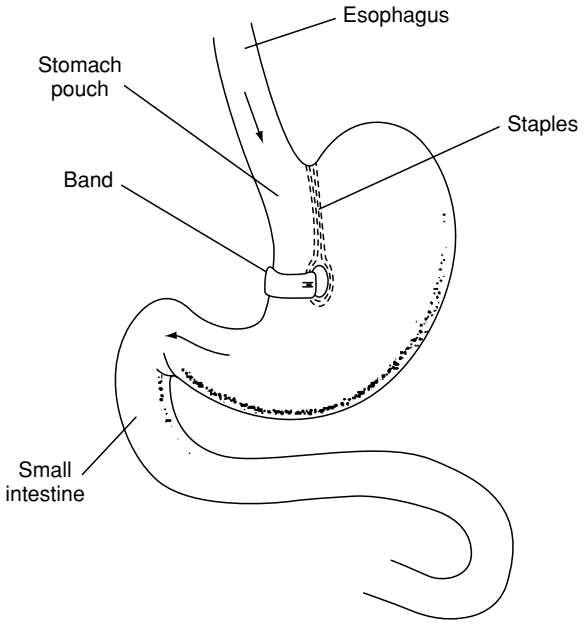

Figure 2 Vertical banded gastroplasty. A vertical staple line reduces the size and accommodation of the stomach. Mesh banding of the distal portion of the newly created gastric pouch slows the exit of food, leading to early satiety. Reproduced with permission from Schaefer DC and Cheskin LC (1998) Update on obesity treatment. *Gastroenterology* 6(2): 136–145.

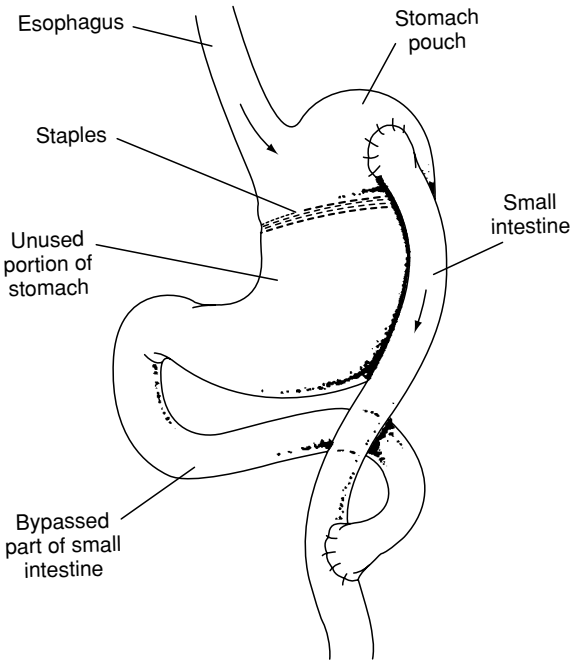

Figure 3 Gastric bypass. A horizontal staple line reduces the size and accommodation of the stomach. A jejunal limb is attached to the new gastric pouch, thus allowing bypass of the duodenum. Reproduced with permission from Schaefer DC and Cheskin LC (1998) Update on obesity treatment. *Gastroenterology* 6(2): 136–145.

individuals who used or were treated with caloric restriction. Unfortunately, the long-term success rate for all weight-loss methods (excluding surgery), defined as losing weight and keeping the majority of it off for at least 5 years, is low – perhaps 5–15%. Even after the most state-of-the-art, medically monitored comprehensive treatment, it is likely that patients will gradually return to their baseline weight within 3–5 years after treatment termination. Given this very high relapse rate, obesity can be considered a chronic condition, much like cigarette smoking or other substance abuse. As such, only the continuous-care model of management, which views obesity as a chronic disease requiring continuous support or contact after the end of formal treatment, has been found to produce significant maintenance of weight loss.

It is probable that some form of continuous or punctuated long-term care will be necessary for many (if not all) obese patients to maintain the dietary and lifestyle changes that produced their weight loss. With ongoing support from a physician or other health care providers, patients will be better able to maintain their changes in eating and activity behavior and the resultant weight loss. Regrettably, optimal strategies for weight maintenance are in a state of evolution and represent the key limitation of our current treatment of this highly prevalent, medically hazardous, and chronic condition. Research has identified continued program contact (e.g. monthly), regular physical activity, nutritional sophistication, and self-monitoring as four predictors of weight maintenance. It not surprising that the strategies helpful in promoting weight loss, such as self-monitoring, physical activity, regular weigh-ins, and attendance at a multidisciplinary program, will be the same strategies necessary for long-term maintenance of weight loss.

See also: **Behavioral (Behavioural) Effects of Diet**; **Body Composition**; **Bulimia Nervosa**; **Exercise**: Muscle; Metabolic Requirements; **Fat Substitutes**: Use of Fat-replaced Foods in Reducing Fat and Energy Intake; **Hunger**; **Metabolic Rate**; **Nutrition Education**; **Nutritional Assessment**: Importance of Measuring Nutritional Status; **Obesity**: Etiology and Diagnosis; Fat Distribution; Epidemiology; **Satiety and Appetite**: Food, Nutrition, and Appetite; **Slimming**: Slimming Diets; Metabolic Conseqences of Slimming Diets and Weight Maintenance

Further Reading

AACE/ACE Obesity Task Force (1998) AACE/ACE position statement on the prevention, diagnosis, and treatment of obesity (1998 revision). *Endocrine Practice* 4(*5*): 297–330.

Bray GA (1998) *Contemporary Diagnosis and Management of Obesity*. Newtown, PA: Handbooks in Health Care.

Brownell KD and Fairburn CG (1995) *Eating Disorders and Obesity: A Comprehensive Handbook*. New York: Guilford Press.

Cheskin LJ (1997) *Losing Weight for Good: Developing Your Personal Plan of Action*. Baltimore, MD: Johns hopkins University Press.

Foreyt JP and Poston WSC (1998) What is the role of cognitive-behavior therapy in patient management? *Obesity Research* 6 (suppl. 1): 18S–22S.

National Institutes of Health Consensus Development Panel (1991) Gastrointestinal surgery for severe obesity. *Annals of Internal Medicine* 115: 956–961.

Perri MG, Nezu AM and Viegener BJ (1992) *Improving the Long-term Management of Obesity: Theory, Research, and Clinical Guidelines*. New York: John Wiley.

Schaefer DC and Cheskin LJ (1998) Update on obesity treatment. *Gastroenterologist* 6(2): 136–145.

Thomas PR (ed.) (1995) *Weighing the Options: Criteria for Evaluating Weight-Management Programs*. Washington, DC: National Academy Press.

Wadden TA and Van Itallie TB (1992) *Treatment of the Seriously Obese Patient*. New York: Guilford Press.

Epidemiology

N E Sherwood and M Story, University of Minnesota, Minneapolis, MN, USA

Introduction

Obesity is a major public health problem in the USA. The dramatic increase in the prevalence of obesity over the last few decades has raised concerns about associated health risks for children, adolescents, and adults. Obesity is considered the second leading cause of preventable death in the USA. Recent data suggest that the estimated number of annual deaths in the USA due to obesity approaches 300 000. A continuation of current trends could lead to further increases in the number of people affected by obesity-related health conditions and premature mortality. Prominent among the health risks associated with obesity are hypertension, type 2 diabetes mellitus, dyslipidemia, stroke, gallbladder disease, osteoarthritis, sleep apnea, respiratory problems, and certain cancers (e.g., endometrial, breast, prostate, and colon). Obesity is also associated with psychosocial problems such as binge-eating disorder and depression. Furthermore, individuals who are obese are adversely impacted by social bias and discrimination.

As the prevalence of obesity has been rising, the economic burden of obesity has been a topic of considerable discussion. The economic costs of obesity have been categorized into three areas, the cost to individuals because of the impact of obesity on health, costs to society due to lost productivity and premature mortality associated with obesity, and costs to individuals and service providers in treating obesity. Recent estimates suggest that obesity-related mortality may account for 7% of US health care costs. Medical care costs alone attributable to obesity in the US were estimated to toal almost $46 billion in 1990. Although no study to date has taken into account the entire range of costs associated with obesity, estimates of the economic burden are as high as $100 billion per year.

Monitoring the prevalence of obesity and associated health behaviors is critical for understanding and documenting the scope of the problem. The goal of this chapter is to review definitions of overweight and obesity in adults and children, present available data regarding the epidemiology of obesity and related risk factors, and discuss implications for future public health research and intervention.

Definitions of Overweight and Obesity

The terms obesity and overweight are often used interchangeably, but they are distinct conditions. Obesity refers to excess adipose tissue, whereas overweight refers to excess weight for height. Because it is difficult to obtain direct measures of obesity and since body weight tends to be highly correlated with adiposity, weight-for-height measures are generally used to classify overweight and obesity in adults and youths.

Adults

Definitions of obesity and overweight have varied over time and in different countries. In the USA, many studies have used the 1959 or 1983 Metropolitan Life Insurance tables of desirable weights for heights and defined overweight as a weight exceeding the midpoint of these standards by a certain percentage. In 1985, a NIH Consensus Conference endorsed the use of body mass index (BMI; $kg\,m^{-2}$) and adopted the 85th percentile of BMI from the second National Health and Nutrition Examination Survey (NHANES II) as the standard for defining overweight. The 85th percentile corresponded well to approximately 20% overweight according the 1983 Metropolitan Life tables. Using this standard, overweight was defined as BMI ≥ 27.8 for men and ≥ 27.3 for women. In 1995, a World Health Organization (WHO) expert committee recommended an alternative classification system for overweight and obesity; overweight was defined as a BMI of 25–29.9 $kg\,m^{-2}$ and obesity as a BMI $\geq 30\,kg\,m^{-2}$. Although the previous definition of overweight in the US used a cut-off of about 27 $kg\,m^{-2}$, epidemiological data show increases in mortality with BMIs above 25 $kg\,m^{-2}$. The category of obesity is further subdivided into three classes: obese class I (BMI 30–34.9), obese class II (BMI 35.0–39.9), and obese class III (BMI ≥ 40).

BMI has been accepted by numerous groups as the most practical and valid screening tool for assessing weight status in individuals and the most useful population-level measure of overweight and obesity. BMI provides meaningful prevalence estimates within and between populations and is useful for identifying individuals and groups at increased health risk and for evaluating the efficacy of obesity interventions. However, it is important to recognize that BMI does not provide information on body fat or body-fat distribution. Body-fat distribution has been identified as an independent predictor of health risk. Individuals with excess abdominal fat are at increased risk of negative health consequences of obesity. In adults,

the waist-to-hip ratio, or waist circumference, is often used as a marker of intraabdominal adipose tissue.

Children and Adolescents

Defining overweight or obesity in children and adolescents is complicated by the normal processes of growth, pubertal development, and body-composition changes. No standard definition exists for child or adolescent obesity. Classifications that have been used include weight-for-height percentiles, relative weight, percent of ideal body weight, skinfold measures, and BMI. BMI is currently the most accepted and widely used measure. For adults, BMI criteria have been based on mortality or mortality outcome research, but no risk-based criteria have been established for youths, as it is difficult to link youth weight status to chronic disease outcomes. Adult BMI criteria also utilize a single cut-off value for both sexes and all ages, which is inappropriate for children and adolescents who are experiencing growth and body-composition changes. Because BMI changes dramatically with age during childhood and adolescence, BMI needs to be assessed using age-specific reference curves. Gender-specific values are also needed for adolescents because of differences in body composition and timing of puberty. During puberty, males increase their lean body mass and decrease the amount of body fat, whereas girls experience rapid fat accretion.

The release of the revised National Center for Health Statistics (NCHS) and the Centers for Disease Control (CDC) growth charts incorporating smoothed gender- and age-specific BMI percentiles based on data from National Health Examination Survey (NHES) and NHANES will provide a useful reference for describing weight status in children and adolescents. It has been suggested by US expert panels that for clinical evaluation of obesity and epidemiologic application, overweight in youths should be defined as a BMI greater than the 95th percentile for age using national reference population data. Youths between the 85th and 95th percentile are considered at risk of becoming overweight.

Prevalence of Obesity: National Trends

Adults

Comprehensive data of trends in the prevalence of obesity provided by national surveys (NHES I (1960–1962); NHANES I (1971–1974); NHANES II (1976–1980); NHANES III (1988–1994)) show that the percentage of obese persons has increased over time. Although there was a small increase in the estimated prevalence of obesity across the first three periods (1960–1980), a much larger increase was observed between the third and the fourth surveys. The percentage of individuals classified as obese increased from 15% in NHANESII to 23% in NHANES III. An additional 33% of US adults in NHANES III were classified as overweight according to the BMI > 25.0 standard. According to these data, an estimated 97.1 million adults (55% of US adults) are classified as overweight.

Data from the Behavioral Risk Factor Surveillance System (BRFSS) also show that the prevalence of obesity has increased in recent years. The BRFSS is a cross-sectional telephone survey of noninstitutionalized adults aged 18 and older conducted by the Centers for Disease Control and Prevention and state health departments. Data from the BRFSS indicate that the prevalence of obesity increased from 12% in 1991 to 18% in 1998. The prevalence of obesity increased across both genders, all socio-demographic groups and in all states. These data are derived from self-reports of individuals; therefore, true rates of overweight and obesity are likely underestimated since individuals tend to underreport their body weight.

Children and Adolescents

Regardless of the method used to classify overweight or obesity, studies have shown high prevalence estimates among children and adolescents, and that rates have increased dramatically since the mid-1960s and continue to rise. NHANES data indicate that the prevalence of overweight (BMI at or above the 95th percentile) among youths doubled from 1976–1980 to 1988–1994, increasing from 8 to 14% for 6–11 year olds and from 6 to 12% for 12–17 year olds. Currently, about 11% of US children and adolescents are overweight, and an additional 14% have a BMI between the 85th and 95th percentiles, indicating that they are at risk for becoming overweight.

Gender Differences in Obesity

Data suggest that the prevalence of obesity is increasing across all populations in the USA, but epidemiological data show that certain subgroups of individuals are at a particularly high risk of obesity. Although the prevalence of obesity has increased over time in both men and women, the prevalence of overweight has remained relatively stable. Data show that men have higher rates of overweight compared with women, but women have higher rates of obesity than men. A major contributor to the high rates of obesity in US adults is weight gain with age. About two-thirds of obese adults became obese in adulthood. Between the ages of 20 and 50 years, weight gain tends to

occur at a rate of 0.45–0.91 kg per year, and major weight gain is most likely to occur between the ages of 25 and 34 among both men and women. Prospective epidemiological data also show that women in the US appear to be at twice the risk of major weight gain as men. The prevalence of overweight and obesity increases with age across both sexes, peaking around the 50–59-year-old age range. After age 70, the prevalence of obesity is lower in adult men and women. This lower prevalence may be due to both obesity-related mortality and decreases in BMI with age.

Among children and adolescents, national data show few gender differences in obesity rates. Males tend to have slightly higher rates of overweight compared with females (e.g., 15% for males compared with 14% for females).

Racial/Ethnic Differences in Obesity

Significant racial/ethnic differences in the prevalence of overweight and obesity have also been observed. The most recent NHANES III data show that among women, African American and Hispanic women are at highest risk, with about two-thirds of women in these ethnic groups meeting BMI criteria for overweight or obesity. Ethnic specific data from studies such as Hispanic HANES (HHANES), which includes Mexican Americans, Cuban Americans, and Puerto Rican Americans, and smaller studies of American Indians and Alaskan Natives, show high rates of overweight and obesity in these populations. In contrast, ethnic-specific data show that Asian American women have the lowest rates of overweight and obesity. Among men, fewer ethnic differences in overweight and obesity have been observed. According to NHANES III data, Hispanic men are at highest risk of overweight and obesity compared with non-Hispanic white men and African American men. According to data from the BRFSS, the prevalence of obesity in Hispanic men almost doubled over the last decade (10% in 1991; 18% in 1997). Data from smaller studies of American Indian and Alaskan Native men also show elevated rates of overweight and obesity compared with all-race data in US men.

As with adults, differences in overweight prevalence by race and ethnicity have been observed among children and adolescents. In NHANES III, the overweight prevalence among youths was considerably higher for male and female Hispanics, and African American girls compared with whites. For example, among 6–11 year olds, 16% of black girls, 14% of Hispanic girls, and 9% of white girls were overweight. Among males aged 6–11, 17% of Hispanic boys were overweight, compared with 10% of white and 12% of black boys. Other studies have shown excessively high prevalence estimates of obesity in American Indian youth. A recent survey of over 12 000 American Indian youths aged 5–17 years of age in the Northern Plains area found that 22% of the males and 18% of the females were overweight (BMI > 95th percentile).

Socio-economic Differences in Obesity

An inverse relationship between socio-economic status and obesity has been observed in the USA and is particularly pronounced for women. Data from NHANES III show that collapsing across all races, women with lower incomes are more likely to be overweight or obese compared with those with higher incomes. Examination of NHANES III all-race data for men shows that the prevalence of overweight does not vary across income groups. The relationship between socio-economic status and obesity, however, appears to vary across ethnic groups and according to gender. There is a strong inverse relationship between income and weight status among nonHispanic White women and a moderate inverse relationship between income and weight status among Hispanic American women. However, among African American women, weight status and income are only weakly associated. A similar trend has been observed in other studies. Although nonHispanic white men did not show any difference in weight status across income groups, examination of similar data in men shows that there is a positive relationship between income and obesity for Hispanic American and African American men.

Compared with adults, studies with children have shown a weak and less consistent relationship between SES and overweight. In NHANES III data, there was an inverse relationship between SES and overweight only for white adolescents. Overweight prevalence was not related to family income in African American or Hispanic youths.

Persistence of Childhood Obesity

Whether childhood-onset obesity leads to an increased likelihood of obesity in later life is an important issue with clinical implications. The likelihood of persistence of obesity from childhood to adulthood is related to the degree and duration of obesity, family adiposity, and age of the child. The likelihood of an overweight infant becoming an overweight adolescent or adult is small. Less than 15% of overweight infants and only about 25% of overweight preschool children will remain overweight into adulthood. Obesity is more likely to persist if it is present during the adolescent years. In general, the later into adolescence overweight persists, the severity of the obesity

and the presence of parental obesity increase the likelihood that obesity will persist in adulthood. A recent Washington state study tracked 850 infants over 21–29 years and found that among obese 6 years olds, about 50% remained obese. By the age of 10–14 years, 80% of obese children with at least one obese parent remained obese.

Global Trends in Obesity

Both childhood and adult obesity are becoming global concerns. Other industrialized countries such as England, Singapore, Japan, Canada, and Australia are also experiencing increasing rates of obesity in both adults and children. Similar to data from the US, women from other countries tend to have higher rates of obesity compared with men, although men may have higher rates of overweight. The WHO reports that the prevalence of obesity is also increasing in countries undergoing economic transition and in developing countries where obesity coexists with undernutrition.

Risk Factors for Obesity in Adults and Children

It is widely accepted that overweight and obesity are multidetermined chronic problems resulting from complex interactions between genes and environments characterized by energy imbalance due to sedentary lifestyles and ready access to a wide variety of foods. Research suggests obesity runs in families and that some individuals are more vulnerable than others to gaining weight and becoming obese. Researchers also agree that there is no single gene that causes obesity, but that multiple genes and gene mutations associated with weight gain are involved. Various mechanisms through which genetic susceptibility to weight gain have been proposed including a low resting metabolic rate, low level of lipid oxidation rate, low fat-free mass, and poor appetite control. Although genetic research is a promising approach for understanding the development of obesity and identifying those at risk for obesity, it should be noted that the rapid increases in rates of obesity and overweight that have been observed have occurred over too brief a time period for there to have been significant genetic changes in the population. Therefore, the following sections briefly review environmental factors associated with the obesity epidemic.

Environmental Influences: Dietary Intake

Increased population levels of energy intake is a likely contributor to the epidemic of obesity. However, research on trends in energy intake over the time period in which the prevalence of obesity has increased has not consistently supported this hypothesis. Data from national surveys show a decrease in energy intake over time, whereas, nutrient intake data from NHANES III show an increase in energy intake over time. A recent review of ecological data regarding the amounts and types of food available in the USA during the past two decades provides interesting data that may shed light on some of the discrepancies found in previous survey data. Specifically, although the availability and purchasing of reduced fat and reduced energy foods has increased over the past few decades, people are eating more meals away from home, eating larger portions, and consuming more convenience foods. These behaviors are more conducive to higher energy intakes and may be contributing to the increased prevalence of obesity.

Environmental Influences: Physical Activity

Low levels of energy expenditure are increasingly recognized as important contributors to obesity and related health conditions. An abundance of cross-sectional research shows that heavier individuals are less active than lighter individuals, and prospective research indicates that changes in physical activity level are associated with changes in body weight in the direction predicted by the energy balance equation. The past century has produced dramatic changes in physical activity patterns in the USA. Machines and labor-saving devices have become so commonplace that the energy expenditure now required for daily life is a fraction of what it was a generation or two ago. Voluntary or leisure-time physical activity has thus assumed central importance in meeting physical-activity requirements. Unfortunately, voluntary physical activity is not very popular. Less than 25% of all adults, adolescents, and children report that they engage in regular physical activity (i.e., a minimum of 30 min of moderate to vigorous activity on most days of the week), the activity level most recently recommended by health experts. The prevalence of regular physical activity also varies according to demographic characteristics. Decreases in physical activity are observed during adolescence, with the decline particularly pronounced among girls. Women are less physically active than men, older adults are less active than younger adults, and African American and Hispanic adults are less active than whites. A lower socio-economic status is also associated with lower levels of physical activity.

Although low levels of leisure-time physical activity likely contribute to the epidemic of obesity, it is

noteworthy that among adults, leisure-time activity has remained stable or increased since the mid-1980s, the time period during which the prevalence of obesity increased. Given this paradoxical finding, it is likely that increases in sedentary activities such as television watching and computer use and decreases in lifestyle, household, and occupational activity that have been less carefully measured have contributed to reductions in overall energy expenditure at the population level. Data from the Americans' Use of Time study show that the amount of free time spent watching television increased from about 4 h per week in 1965 to about 15 h in 1985. Documenting the prevalence of sedentary behavior among children and adults and targeting the reduction of sedentary activity as a health-behavior goal have received considerable attention in recent years.

Future Directions

Understanding and addressing the epidemic of obesity has become a global public-health priority. A recent 1997 WHO report on obesity, entitled 'Obesity: Preventing and Managing the global Epidemic,' noted that overweight and obesity represent a rapidly growing threat to the health of populations world-wide. Both children and adults in developing and developed countries are affected. The report concluded that the spectrum of problems seen across the world is of such magnitude that obesity should be regarded as the principally neglected public-health problem.

The recently released goals of Healthy People 2010 are to reduce the prevalence of obesity among adults from 23 to 15% and to reduce the prevalence of obesity among children and adolescents from 11 to 5%. The etiology of obesity is complex and encompasses a wide variety of social, behavioral, cultural, environmental, physiological, and genetic factors. To achieve these ambitious goals, considerable effort must be focused on helping individuals at the population level modify their diets and increase their physical activity levels, key behaviors involved in the regulation of body weight. A challenge to public-health professionals is to develop educational and environmental interventions that support diet and exercise patterns associated with a healthy body weight. Prevention of obesity should begin early in life and involve the development and maintenance of healthy eating and physical activity patterns. These patterns need to be reinforced at home, in schools, and throughout the community. Communities, government, health organizations, the media, and the food and health industry must form alliances if we are to combat obesity.

A coherent and standard international system for classifying overweight and obesity in children and adults is also needed. The WHO and US agencies recommend using BMI for adults, with BMI ≥ 25 denoting 'overweight' and BMI > 30 denoting 'obesity.' Criteria and methods for assessing and documenting obesity in children and adolescents need to be similarly developed. A common international standard would allow meaningful comparisons of obesity within and across populations and document obesity trends over time.

See also: **Adolescents**; **Children**: Nutritional Problems; **Nutritional Assessment**: Anthropometry and Clinical Examination; **Obesity**: Etiology and Diagnosis

Further Reading

Allison DB, Fontaine KR, Manson JE, Stevens J and VanItallie TB (1999) Annual deaths attributable to obesity in the United States. *Journal of the American Medical Association* 282(16): 1530–1538.

Clinical Guideline on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults. The Evidence Report (1999) National Institutes of Health. National Heart, Lung, and Blood Institute.

Flegal KM, Carroll MD, Kuczmarski RJ and Johnson CL (1998) Overweight and obesity in the United States: prevalence and trends, 1960–1994. *International Journal of Obesity* 22: 39–47.

Harnack L, Jeffery R and Boutelle K (in press) Temporal trends in energy intake in the U.S.: an ecological perspective. *American Journal of Clinical Nutrition*.

Kuczmarski RJ, Flegal KM, Campbell SM and Johnson CL (1994) Increasing prevalence of overweight among US adults. *Journal of the American Medical Association* 272(3): 205–211.

Mokdad AH, Serdula MK, Dietz *et al*. (1999) The spread of the obesity epidemic in the United States, 1991–1998. *Journal of the American Medical Association* 282(16): 1519–1522.

Must A, Spadano J, Coakley EH, Field AE, Colditz G and Foryx WD (1999) The disease burden associated with overweight and obesity. *Journal of the American Medical Association* 282(16): 1523–1529.

Obesity: Preventing and Managing the Global Epidemic. Report of a WHO Consultation on Obesity. 3–5 June 1997. Division of noncommunicable diseases. Geneva: WHO.

Strauss R (1999) Childhood obesity. *Current Problems in Pediatrics* 29: 5–29.

Troiano RP and Flegal KM (1998) Overweight children and adolescents: description, epidemiology, and demographics. *Pediatrics* 101(supplement 3): 497–504.

US Department of Health and Human Services (1996) *Physical Activity and Health: A Report of the Surgeon General*. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and

Prevention, National Center for Chronic Disease Prevention and Health Promotion.
Williamson DF (1991) Epidemiologic analysis of weight gain in U.S. adults. *Nutrition* 7(4): 285–286.
Williamson DF, Kahn HK, Remington PL and Anda RF (1990) The 10-year incidence of overweight and major weight gain in U.S. adults. *Archives of Internal Medicine* 150: 665–672.

Odors *See* **Sensory Evaluation**: Sensory Characteristics of Human Foods; Food Acceptability and Sensory Evaluation; Practical Considerations; Sensory Difference Testing; Sensory Rating and Scoring Methods; Descriptive Analysis; Appearance; Texture; Aroma; Taste

OFFAL

Types of Offal

W F Spooncer, CSIRO, University of Western Sydney, Richmond, NSW, Australia

Introduction

Offals are all the noncarcass parts of slaughtered animals with the exception of hide and skin. Most offals have the potential to be used as human food but this potential can only be realized if the offals are collected hygienically, inspected, and passed fit for human consumption, then cleaned and prepared in an appropriate manner. Offals which are not collected for human use, either because there is no demand for them, or because they are diseased or contaminated, are inedible offals. Inedible offals are either destroyed by burning or burying, rendered to produce tallow and meat meal or, if suitable, used as pet food. (*See* **Meat**: Slaughter.)

Definition of Offals

Because of the variety of different types of offals, both edible and inedible, some attempts have been made to group offals together in categories. Many texts refer to offal as being derived from the 'off fall' from carcass preparation. Some of this 'off fall' may be attractive as edible product while some of it may not be edible at all. To avoid the connotation that offals include unattractive and inedible material, the terms 'variety meats' and 'fancy meats' have been introduced as categories of offals which are edible and prepared and consumed in a recognizable form. Variety meats and fancy meats are items such as heart, tongue, kidney, sweetbread, liver, brain, tail, and tripe. In the USA, where the term variety meat is common, products which are not consumed in a recognizable form but could be used in processed meats are referred to as offal meat. Examples of offal meat are lungs, spleen, udders, and head meat. In this case the word 'offal' is reserved for inedible material.

Other ways in which offals are categorized include the distinction between white and red offals, and between organ offal and muscle offal.

The categories of white and red offal are used to distinguish stomach and intestinal offals from all other offals. This distinction is made because white offals require careful cleaning and preparation to remove intestinal and stomach contents before they can be used as edible material. Intestinal offal should be cleaned in an area separated from the preparation of other offals and carcass meat. Separate facilities for handling white and red offals minimize the chance of red offal being contaminated by intestinal contents.

The distinction between offals which are mainly organ tissue and those which are muscular tissue is made because of the difference in the potential keeping quality of the two types of offals. Organ offals, such as liver, kidney, lungs, spleen, and brain, are not as stable as muscular offals, such as tail, tongue, and diaphragm. In chilled storage, physiological activity in organ tissues can result in deterioration of the flavor and texture. (*See* **Meat**: Preservation.)

Specific Offals

Table 1 lists the items of offal available from slaughtered animals, alternative names, and possible

categorization of the offals. Table 2 lists typical weights of offals from different carcass types. (*See* **Beef**; **Pork**; **Poultry**: Chicken; Ducks and Geese; Turkey.)

According to American surveys, the more popular offals are liver, pigs' feet, oxtail, heart, tongue, pork maw (stomach), brains, sweetbread, tripe, and kidney. These offal items are described below. Other offals listed in Table 1 do not appeal to western tastes but are popular in other markets, particularly in Asian countries.

Liver

Liver is the largest organ in meat-producing species. Its color varies from light red to dark red brown and even black in old animals. The liver is attached by ligament to the anterior abdominal wall and to the stomach by the lesser omentum. During evisceration these attachments are torn or cut, and the liver is removed attached by ligament to the diaphragm and the rest of the pluck (heart, liver, lung, and diaphragm). The liver is separated from the pluck by cutting through the attachment to the diaphragm. The gallbladder and duct are then pulled or cut off the liver.

Liver is composed of specialized liver cells with a network of blood vessels and epithelia-lined ducts between the cells. The cells are held together by a network of connective tissue.

The texture and flavor of liver are affected by species and animal age. Livers from young animals are lighter in color, have more delicate flavor, and are more tender than those from older animals. Lamb and calf liver, and liver from young beef are more suitable for table dishes. Liver from mutton, older bulls, and cows are used in manufactured or processed meats. Pig livers have a strong flavor which is

Table 1 Summary of offal types

Offal type	*Alternative name*	*Category of offal*	
		Variety meat	*Red or white offal*
Organ or glandular offal			
Liver }	Lamb's fry	✓	Red
Heart } Pluck		✓	Red
Lung }	Lights		Red
Brain		✓	Red
Thymus	Sweetbread, heartbread, neckbread	✓	Red
Kidney		✓	Red
Pancreas	Gutbread		Red
Spleen	Melt		Red
Stomach	Pig's maw	✓	White
Rumen	Blanket tripe	✓	White
Reticulum	Honeycomb tripe	✓	White
Omasum	Bible		White
Abomasum	Reed (or vell in calves)		White
Testes	Fries		Red
Udder			Red
Uterus			White
Muscular offals			
Head meat			Red
Cheek			Red
Tail	Oxtail	✓	Red
Diaphragm	Skirt	✓	Red
Tongue (also considered glandular)		✓	Red
Esophagus	Weasand		Red
Others			
Head			Red
Feet			Red
Intestines	Chitterlings (pig only)		White
Ear			Red
Bone marrow			Red

Table 2 Typical weight of offals

	Ovine		*Porcine*		*Bovine*	
	Lamb (g)	*Mutton (g)*	*Porker (g)*	*Baconer (g)*	*Veal (kg)*	*Beef (kg)*
Liver	300–600	650–720	700–1200	1200–2000	0.75–3.0	2.5–8.0
Heart	90–150	180–215	150–250	200–300	0.25–1.0	0.8–2.0
Lung	200–350	400–500	500–850	750–1100	0.65–1.5	1.5–6.0
Brain	75–110	100–130	NA	NA	0.15–0.23	0.23–0.35
Thymus	50–100	NA	NA	NA	0.03–0.125	NA
Spleen	40–65	70–110	70–130	100–150	0.1–0.6	0.6–1.0
Stomach						
Maw	NA	NA	450–600	500–800	NA	NA
Rumen and reticulum	250–500	500–800	NA	NA	0.5–1.5	2.0–10.0
Omasum	30–400	400–500	NA	NA	0.2–1.0	1.0–5.0
Abomasum	80–140	150–200	NA	NA	0.35–0.75	0.7–3.0
Uterus	NA	NA	100–250	200–600	NA	NA
Cheek	NA	NA	1000–1800	1500–2000	0.3–0.7	0.7–1.2
Tail	NA	NA	NA	NA	0.2–0.6	0.55–1.2
Diaphragm	NA	NA	170–250	200–280	0.2–0.6	0.55–1.1
Tongue (short-cut)	50–100	80–150	120–200	150–220	0.15–0.7	0.7–1.6
Head	NA	NA	2000–3000	2800–3800	3.0–5.5	NA
Feet	NA	NA	230–280	300–400	0.2–0.4	NA
Ear	NA	NA	100–300	100–500	NA	NA

NA, not available.

appreciated in Asian countries, but in the west pig livers are used in processed meats rather than in table dishes.

To distinguish between the classes of bovine liver, they may be graded by weight. Livers in the range 1–3 kg are classed as calf livers. Larger livers from beef or ox may be graded into weight ranges of 3–4 kg, 4–5.5 kg, and over 5.5 kg. Ovine livers are not strictly graded but lamb liver should not exceed 700 g.

Heart

Hearts can be collected and trimmed in several styles. Whole hearts are removed from the pericardium and separated from the lungs and remainder of the pluck by cutting through the aorta and pulmonary vein. Whole hearts may be trimmed by cutting off the auricles along with the ossa cordis bones, aorta, and pulmonary vein. Even if the auricles remain in place, whole hearts are trimmed to remove remnants of the aorta, pulmonary vein, and the fibrous rings of the ossa cordis. Trimmed hearts comprise ventricles with associated fat, arteries, and veins, and may or may not include the auricles. They should be well washed to remove blood clots. Large hearts may be split open by cutting through the ventricle wall from the base to the apex.

The thick walls of the ventricles are composed of cardiac muscle, which is tough even in young animals. Cardiac muscle is supported by a network of connective tissue, which also contributes to the toughness of heart meat from older animals.

Tongue

Full tongues consist of the blade of the tongue with the roots attached, including the hyoid bones except the stylohyoid, the larynx, three tracheal rings, lymph nodes, salivary glands, and associated muscle and connective tissue. Short-cut tongues are a more common style of trim and are derived from full tongues by removing the larynx and roots by cutting behind the hyoid bones. Other trims of tongue are prepared with the hyoid bones removed.

The tongue blade is a core of skeletal muscle and connective tissue covered with an epithelial mucous membrane.

Brain

Brains include the cerebrum, cerebellum, and a portion of spinal cord. They are collected after opening the skull by splitting the occipital and maxillary bones. The brain is removed from the skull cavity, leaving the outer skin (the dura mater) in the skull but with a fine membrane (the pia mater and arachnoid meninges) covering the brain tissue. This membrane can be removed when the brain is cooked.

Brain tissue is not supported by a network of connective tissue and, consequently, has a soft, delicate texture.

Sweetbread

Sweetbread is the thymus gland and is only available from young animals. As animals mature, the gland

degenerates into a mass of connective tissue and fat. The sweetbread is collected in two separate portions, although it is a single gland. The main portion, also called neckbread, is found in the neck, either side of the trachea. The other portion is in the chest cavity, close to the heart, and may be called heartbread. After collection, fat and connective tissue is trimmed off the gland.

The thymus gland consists mainly of lymphocytes and is covered by a capsule of connective tissue. The capsule penetrates the gland and divides it into lobules. The texture of sweetbread is delicate but may be tough if sweetbreads are collected from older animals in which the thymus gland has started to degenerate. Sweetbreads are soaked in water to remove blood, blanched to firm the texture, and skinned to remove the capsule before they are prepared for the table.

Kidney

Kidneys are contained in a capsule of fat and remain in the carcass after evisceration. The common way to handle kidneys is to remove them from the capsule for inspection, trim off the ureter and renal blood vessels close to their entry to the kidney and, for beef kidneys, cut out the core of fat and connective tissue from the renal hilus.

Kidneys are made up of renal tubules and small veins and arteries. The structural components are the endothelial and epithelial cells of these vessels, and a network of connective tissue. There is a fibrous capsule of connective tissue around the kidney which is usually removed before kidneys are cooked.

Kidneys from sheep and pigs have the characteristic bean shape. Beef and calf kidneys are divided into lobules.

Pigs' Feet

Forefeet are collected from pig carcasses by cutting the foreleg between the carpal and metacarpal bones. Similarly, hindfeet are collected by cutting between the tarsal and metatarsal bones. Pigs' feet comprise the bones, tendons, skeletal muscle, connective tissue, fat and skin of the trotters, but toenails and all hair should be removed.

The edible part of pigs' feet is fat, muscle, and connective tissue. Prolonged moist cooking of the feet will soften and at least partially gelatinize the collagenous connective tissue and make the flesh tender.

Oxtail

Tails are collected from beef and veal carcasses by cutting off the tail at the sacrococcygeal junction. Tails consist of the coccygeal (tail) vertebrae with associated skeletal muscle, connective tissue, and fat. The last two or three vertebrae may be removed, and some or all of the subcutaneous fat at the anterior end of the tail may be trimmed off.

Tripe

Cattle and sheep stomachs have four compartments, all of which can be used to make tripe. Tripe from the rumen and reticulum is the most common. Rumen tripe is known as blanket tripe and reticulum tripe is known as the honeycomb. The internal surface of the rumen has densely packed papillae, while the reticulum has ridges in the shape of a honeycomb structure. In addition, there are thickened folds in the rumen wall. These folds contain a core of smooth muscle and are not covered by papillae. The fold can be trimmed out of the rumen to produce pillae tripe, known commercially as pillar or mountain chain.

The third compartment of the ruminant stomach is the omasum. The internal wall is in the form of deep, thin folds like the pages of a book. This appearance accounts for the popular name for the omasum, which is 'bible.' Although the omasum has a delicate flavor and texture, it is not commonly used as tripe because of the difficulty of cleaning the stomach contents from between the folds. The fourth part of the ruminant stomach is the abomasum, sometimes called the reed.

The four parts of the stomach are collected in one piece and the omental fat (caul fat) and spleen are removed. The neck of the omasum is cut to separate the rumen and reticulum from the omasum and abomasum. The rumen is cut open and the contents of the rumen and reticulum washed out. The rumen and reticulum are then trimmed into the different tripes, and external fat is trimmed off. Mountain chain tripe does not receive further processing, but blanket and honeycomb tripes may be further cleaned, scalded, and bleached with hydrogen peroxide.

The walls of ruminant stomachs are composed of smooth muscle and connective tissue. The papillae in the rumen are composed of collagen and elastin fibers covered with cornified epithelia, while the ridges and folds in the reticulum and omasum contain smooth muscle as well as connective tissue. There are small, cornified papillae on the folds of the omasum, and the surfaces of the folds are covered by a keratinized mucous membrane. The abomasum has a thick epithelial lining.

Tripes are generally tough because of the high connective tissue content. They contain about 35 g of collagen per 100 g of protein. They require prolonged, moist cooking to tenderize them. Bleached

tripe is treated with caustic soda and has a pH of about 7–9, which increases the water-holding capacity and helps to tenderize the tripe.

Maw

Pig's stomach or maw is a similar organ to the ruminant abomasum. The maw is separated from the rest of the viscera by pulling off the omental fat and spleen from their attachments to the stomach and cutting through the esophagus and duodenum. The maw is then split open and the contents washed out. After cleaning the maw, it is usually scalded in water at about 90 °C to remove the mucous membrane lining.

In common with the abomasum, the maw is composed of a wall of smooth muscle and connective tissue, with an inner lining of thick epithelia and mucous membrane.

Offal for Human Use

Offal consumption varies from about 1.7 kg per person per year in Canada to about 16 kg per person per year in Ireland. Table 3 shows per capita consumption of offal in western countries.

As shown in Table 1, almost all noncarcass parts can be used as food after the carcass and offal have been inspected and found fit for human consumption. (Exceptions are uteri from pregnant animals, and fetuses.) After collection, offals are trimmed and washed to remove contamination by blood, saliva, mucus, and ingesta. They may be distributed frozen, chilled, or vacuum-packed and chilled. Frozen offals held at −18 °C have a useful shelf-life of 6 months; chilled offals have a shelf-life of 2–3 days; and vacuum-packed tail, tongue, skirt, heart, kidney, and liver have a shelf-life of at least 3 weeks at 1 °C.

Table 3 Per capita consumption of offals by countries

	Consumption (kg per head per year)
Australia	4.7
Austria	4.5
Belgium and Luxembourg combined	8.0
Canada	1.7
Denmark	7.0
Republic of Ireland	16.7
Finland	8.1
France	6.6
Greece	6.0
Italy	3.8
Japan	2.5
The Netherlands	2.6
New Zealand	6.0
Norway	2.7
Portugal	5.0
Spain	4.0
Sweden	2.5
Switzerland	5.9
Turkey	2.1
UK	4.3
USA	4.1
Germany	5.7
Yugoslavia	2.9

Offals are consumed both as table dishes and in manufactured meat products such as sausages and pâté. There are many options for preparing and cooking offals, commensurate with the wide ranges of offal types. Some offals require initial treatment such as soaking and precooking. Brains, sweetbreads, and testicles are soaked in cold water to remove blood pigment and improve color. They are then blanched by brief immersion in hot water to firm the texture before they are cooked by other methods. Liver and kidney, particularly from older animals, may also be soaked to leach out the strong flavor associated with these offals. Stomachs and intestines are precooked by prolonged immersion in hot water to remove the mucosal membrane, and soften connective tissues. Tongues are also precooked, after which the thick skin of epithelia on the tongue blade can be peeled off. (*See* **Meat**: Sausages and Comminuted Products.)

Regulations may restrict the way in which offals are used in manufactured and comminuted meat products. In the EC and Australia, offals are considered to be meat but are distinguished from meat of skeletal muscle origin by using terms such as carcass meat and meat flesh, which exclude offal meat. In the USA, heart and tongue are considered to be meat, and other offals have the status of meat by products. The greatest restrictions on the use of offals is in uncooked meat products such as fresh sausages.

In Australia, no offals are permitted in fresh sausage and in the USA only tongue and heart can be used in fresh sausage. In the UK the designated list of permitted offals which can be used in uncooked products includes heart, kidney, liver, and tongue. In general, cooked meat products may contain offals. The Codex Alimentarius Commission has standards which restrict the use of offals in specific products. In summary, these standards are as follows:

1. Canned corned beef. Heart meat is the only offal permitted.
2. Luncheon meat. Luncheon meat with binder may contain offals, with the exception of lungs that have been collected from animals scalded by immersion in water. Only heart and tongue are permitted in luncheon meat without binder.
3. Cooked cured chopped meat. Only heart and tongue are permitted in products which do not contain binder. Offals (except udders, lungs, and

organs and glands from the genital system) may be used in products that contain binder.

Other Uses of Offals

Collection of offals for edible use requires – at least – inspection, washing, and trimming. The cost of these procedures may not be justified by the value of some offals. In this case the offals are disposed of or used in other ways. The main alternatives are to render offals to produce tallow and meat meal, or to use them as pet food. Inspected offal which is not edible because of contamination or disease status is rendered along with other offals which are not required for edible use or pet food, e.g., heads, hooves, and the residues of gut material after collecting sausage casing and tripe.

Offals collected for pet food may be subject to less rigorous inspection, preparation, and trimming than offals for human use and can be handled at a lower cost. The main use of offals is in canned pet foods. These products receive severe heat treatment, and offals which retain form and texture, such as lung and tripe, are most suitable. Liver, heart, and spleen are also used in canned pet food.

There are specialized uses for some offals. Pig and beef pancreas glands are used to produce insulin for the treatment of diabetes. However, insulin from pancreas glands is being replaced by insulin produced from bacterial fermentation.

The mucosa from the lining of lungs and intestines is used to produce the anticoagulant, heparin. A long-standing use of calf vells (abomasum) is in the production of rennet.

See also: **Beef**; **Meat**: Slaughter; Preservation; Sausages and Comminuted Products; **Pork**; **Poultry**: Chicken; Ducks and Geese; Turkey

Further Reading

Codex Alimentarius Commission (1981) *Codex Standards for Processed Meat and Poultry Products and Soups and Broths*. Codex Alimentarius, vol. IV. Rome: Food and Agriculture Organization and World Health Organization.

Gerrard F and Mallion FJ (eds) (1977) *Complete Book of Meat*, pp. 407–413. London: Virtue.

Levie A (1970) *The Handbook of Meat*, pp. 296–301. Westport, Connecticut: AVI.

Patterson JT and Gibbs PA (1979) Vacuum packing of bovine edible offal. *Meat Science* 3: 209.

Spooncer WF (1978) *Yield of Organs and Glands in Relation to Carcass Weight*. Meat Research Report 8/78. Cannon Hill, Queensland: CSIRO Division of Food Research.

Spooncer WF (1988) Organs and glands as human food. In: Pearson AM and Dutson TR (eds) *Edible Meat Byproducts*, pp. 197–217. Essex: Elsevier Science.

Off-flavors (Off-flavours) in Food *See* **Taints**: Types and Causes; Analysis and Identification

Oils *See* **Fats**: Production of Animal Fats; Uses in the Food Industry; Digestion, Absorption, and Transport; Requirements; Fat Replacers; Classification; Occurrence; **Vegetable Oils**: Types and Properties; Oil Production and Processing; Composition and Analysis; Dietary Importance

Okra *See* **Vegetables of Temperate Climates**: Commercial and Dietary Importance; Cabbage and Related Vegetables; Leaf Vegetables; Oriental Brassicas; Carrot, Parsnip, and Beetroot; Swede, Turnip, and Radish; Miscellaneous Root Crops; Stem and Other Vegetables

Olestra *See* **Fat Substitutes**: Use of Fat-replaced Foods in Reducing Fat and Energy Intake

OLIVE OIL

M Tsimidou, G Blekas and D Boskou, Aristotle University of Thessaloniki, Thessaloniki, Greece

Introduction

Poseidon and Athena had a violent disagreement as to who should govern Attica (Greece). To settle this dispute, Zeus declared that he would give the country to whoever offered him the most useful present for mankind. Poseidon then presented a salty spring. Athena presented a twig. This twig would become a staunch tree, capable of living for untold centuries, whose fruit would be edible and from which an extraordinary liquid would be extracted for the preparation of man's food, alleviation of his wounds, strengthening of his system, and lighting his nights. Needless to say, Athena won, and the city, ever since, is named Athens (Figure 1).

The cultivation of the olive tree (Figure 2) dates back before recorded history. The origin of the tree is the area of Mesopotamia. The Minoans (2000–1450 BC) were quite advanced in the cultivation of the olive tree and the production of olive oil, which was also exported. One valuable source of information is the Linear B tablets from the royal archives of Knossos (Crete) and Pylos (Peloponnese) (see Figure 3). Greek mythology, the Old Testament, and Roman literature are full of information regarding the role of the olive tree and its products in religion and everyday life. The cultivation of the olive tree is an essential element of the civilizations developed in the Mediterranean Basin, closely related to their prosperity and cultural achievements.

Olive oil is obtained from the fruits of *Olea europaea* L. The evergreen tree has a life span of several centuries (300–600 years). Its longevity is partially

Figure 2 (see color plate 116) Olive tree in Rethimnon, Crete (Greece), courtesy of V. Zambounis, Athens.

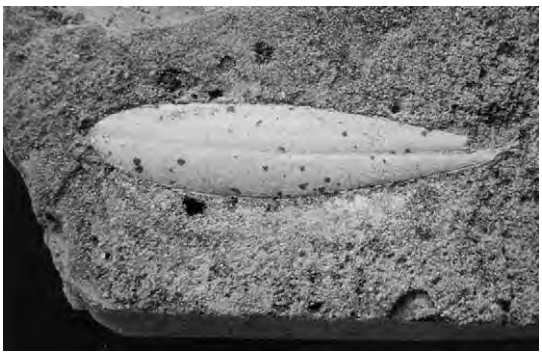

Figure 1 (see color plate 115) Picture of a fossilized olive leaf. Reproduced from Psillakis N and Kastanas E (1999) *The civilization of olive: olive oil*, 2nd edn. Greek Academy of Taste, Karmanor, Iraklion, Crete (Greece), with permission.

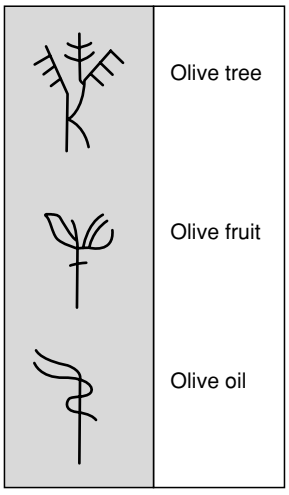

Figure 3 Symbols in the written language Linear B. Reproduced from Psillakis N and Kastanas E (1999) *The civilization of olive: olive oil*, 2nd edn. Greek Academy of Taste, Karmanor, Iraklion, Crete (Greece), with permission.

explained by the characteristic ability to send out shoots and roots from temporary buds at the lower part of the trunk. The height of an old tree may even be as much as 25 m, but trees in systematic plantations have an optimum height of 4–5 m to facilitate manual picking of fruits. The tree is resistant to unfavorable conditions (dryness, infertile soil, etc.) but requires a mild climate (not below −9 °C in winter – a moderate to high temperature in summer). Many species of olive tree are grown all over the world, some of them (>35) suitable for olive oil production. Cultivars used for oil production have a medium-sized fruit containing 17–30% oil when ripe. Among the most important oil-producing cultivars are: *Koroneiki, Megaritiki*, and *Tsounati* (Greece), *Leccino* and *Frantoio* (Italy), *Picual, Hojiblanca*, and *Lechin de Sevilla* (Spain). (*See* **Olives**.)

Olive Oil Trade and Consumption

Olive oil is an important component of the Mediterranean diet, which is now also gaining interest among consumers of northern Europe, USA, Canada, and other countries as a component of healthy eating. Olive oil is almost exclusively produced in Spain, Italy, Greece, Tunisia, Turkey, Morocco, and Portugal (*c.* 31, 31, 18, 4, 5, 2.5, and 2% of the world's production, respectively) (see also Table 1). More than 800 million trees, covering a total area of 25 million acres, yield approximately 2 000 000 tonnes of oil (Figure 4). Though olive oil production accounts for no more than 2% of the volume of edible oils traded world-wide, its market value has a significant share (15%) in international trade because of its high price.

Table 1 World olive oil production

Country	*1998/1999 (tonnes)*	*Percentage share*	*1999/2000 (tonnes, estimated)*	*Percentage share*
European Union	1 698 500	71.5	1 563 000	77
Tunisia	215 000	9	200 000	10
Turkey	170 000	7	70 000	3
Syria	115 000	5	80 000	4
Morocco	65 000	3	40 000	2
Algeria	39 500	1.5	25 000	1
Other countries	70 600	3	55 500	3
Total	2 373 600	100	2 033 500	100

Data from IOOC (1999) Trade Standards Applying to Olive Oil and Olive Pomace Oil. COI/T. 15/NC, Doc. No. 2; rev. 9. Madrid: International Olive Oil Council.

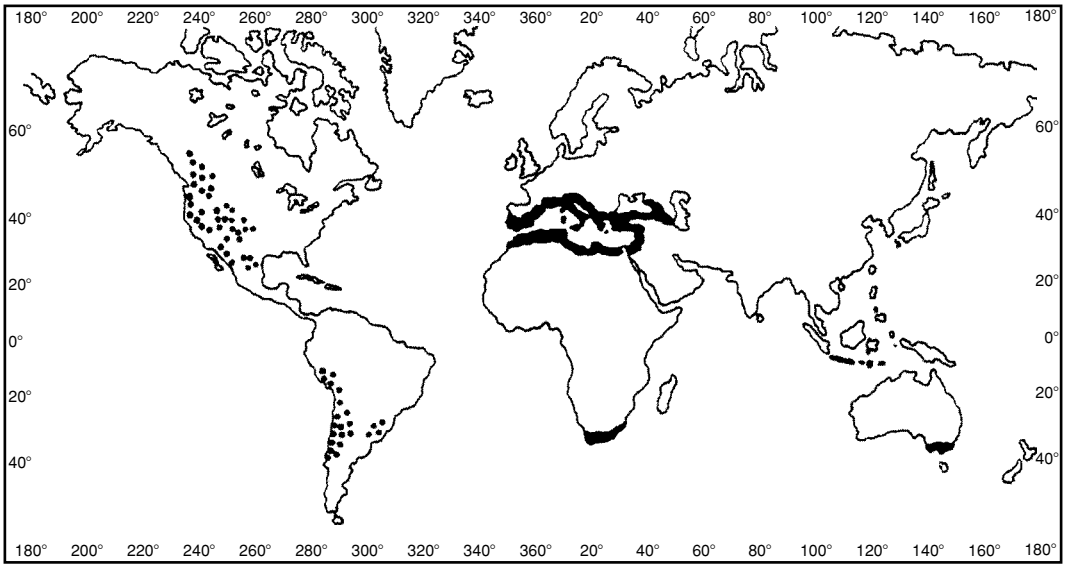

Figure 4 Olive tree growing areas of the world. From Kiritsakis KA (1998) *Olive Oil: From the Tree to the Table*, 2nd edn. Trumbull, CT: Food and Nutrition Press, with permission.

The International Olive Oil Council (IOOC), established in 1959, is the intergovernmental organization responsible for administering the International Agreement on Olive Oil and Table Olives. This agreement has been negotiated at the United Nations and lays down the policy that members should take for standardization of the market for olive oil, olive-pomace oil, and table olives. The Agreement is renewed regularly (for the time being, the 1986 Agreement has been prolonged until 31 December 2000). Principles and provisions set are compulsory in international trade, recommended in domestic trade, and incorporated in the trade standards for olive oil and olive-pomace oil.

According to recent IOOC reports, the increasing production of olive oil in Europe, as well as in the rest of the world, has led to a drop in prices and to a concomitant rise in consumption. Data concerning consumption in 1998/1999 and estimated consumption for 1999/2000 are given in Table 2. Similar trends stand for the year 2001/2002. Olive oil consumption is practically concentrated in the producing countries. The average annual per-capita consumption of olive oil reflects the economic policy and culinary traditions of each producing country. Thus, consumption that is reported for Greek consumers (19.5 kg) is twofold higher than that in Italy or Tunisia. In France and in the countries of central and northern Europe, consumption of olive oil is very low. In the last decade, there has been a significant increase in olive oil consumption. In countries such as Germany, Denmark, UK, and Ireland, the estimated per-capita per-year consumption in the last 2 years has been around 0.5–0.6 kg. In the USA and Canada, similar estimates are approximately 0.8 kg. The results of the 'Seven Countries Study' related the low rate of coronary heart disease among the Cretan population to the regular consumption of virgin olive oil. This was the first scientific evidence for the nutritional role of this 'high-in-monounsaturates' juice. Currently, the interest is focused on the role of olive oil in the prevention of bone and nervous-system functions, digestion, antiageing properties at cell and mitochondrial levels, diabetes, and breast tumors, and also in the preventive role of some of its minor constituents in the oxidation of low-density lipoprotein (LDL). (*See* **Coronary Heart Disease**: Etiology and Risk Factor; Antioxidant Status; Prevention.)

Commercial Types of Olive Oil

Olive oil is marketed in accordance with the designations of Table 3. Virgin olive oil is obtained by only mechanical or other physical means under conditions, particularly thermal, which do not lead to alterations in the oil. This oil has not undergone treatment other than washing, decantation, centrifugation, and filtration. *Extra, fine*, and *ordinary* virgin olive oils are different edible grades of virgin olive oil. *Lampante* is a nonedible grade of virgin olive oil suitable for refining or other technical uses. Refined olive oil is the oil refined by the methods used for other vegetable oils. These include neutralization, decolorization with bleaching earth, and deodorization. Olive oil is an edible blend of virgin olive oil and refined olive oil. Olive-pomace oil is obtained by solvent extraction of the cakes derived from olive milling. The triacylglycerol composition is similar to that of virgin olive oil, but some of the nonsaponifiable compounds differ significantly. Due to the high content of waxes, the oil has to be winterized before refining. *Olive-pomace oil* can be used as a mixture with virgin olive under the designation 'olive-pomace oil.' This oil cannot be called 'olive oil.' Virgin olive oils may also be traded as 'appelation d'origin' products in accordance with the existing legislation (e.g., EC Regulation 2081/92). (*See* **European Union**: European Food Law Harmonization.)

Table 2 World olive oil consumption

Country	*1998/1999 (tonnes)*	*Percentage share*	*1999/2000 (tonnes, estimated)*	*Percentage share*
European Union	1 669 000	70	1 679 000	71
USA	157 000	7	160 500	7
Turkey	97 000	4	60 000	2.5
Syria	88 000	4	85 000	3.5
Tunisia	49 000	2	65 000	3
Morocco	55 000	2	50 000	2
Algeria	35 000	1	29 000	1
Other producing countries	142 000	6	139 000	6
Other nonproducing countries	92 500	4	92 500	4
Total	2 385 200	100	2 359 900	100

Data from IOOC (1999) Trade Standards Applying to Olive Oil and Olive Pomace Oil. COI/T. 15/NC, Doc. No. 2; rev. 9. Madrid: International Olive Oil Council.

Table 3 Quality criteria for olive oils and olive-pomace oils

	Extra virgin olive oil	*Virgin olive oil*	*Ordinary virgin olive oil* [b]	*Lampante virgin olive oil*	*Refined olive oil*	*Olive oil* [d]	*Crude olive-pomace oil*	*Refined olive-pomace oil*	*Olive-pomace oil*
1. *Organoleptic characteristics*:									
Panel test score (scale 1–9)	≥ 6.5	≥ 5.5	≥ 3.5	≤ 3.5					
Odor and taste					Acceptable	Good		Acceptable	Good
Color					Light yellow	Light yellow to green		Light yellow to brownish yellow	Light yellow to green
Aspect at 20 °C for 24 h					Limpid	Limpid		Limpid	Limpid
2. *Free acidity % m/m (expressed as oleic acid)*	≤ 1.0 [a]	≤ 2.0	≤ 3.3	> 3.3	≤ 0.3	≤ 1.5 [e]	No limit	≤ 0.3	≤ 1.5
3. *Peroxide value (mEq peroxide oxygen per kg oil)*	≤ 20	≤ 20	≤ 20	No limit	≤ 5.0	≤ 15	No limit	≤ 5	≤ 15
4. *Absorbance in ultraviolet ($K_{1cm}^{1\%}$)*									
270 nm	≤ 0.25	≤ 0.25	≤ 0.30 [c]	No limit [c]	≤ 1.10	≤ 0.90		≤ 2.00	≤ 1.70
Δ *K*	≤ 0.01	≤ 0.01	≤ 0.01		≤ 0.16	≤ 0.15		≤ 0.20	≤ 0.18

[a] ≤ 0.8; [b] Banning of the category agreed in the EC; [c] After passage of the sample through activated alumina, absorbance at 270 nm shall be equal to or less than 0.11; [d] commercial name is discussed; [e] ≤ 1 proposed.
Data from IOOC (1999) Trade Standards Applying to Olive Oil and Olive Pomace Oil. COI/T. 15/NC, Doc. No. 2; rev. 9. Madrid: International Olive Oil Council.

Olive-oil Extraction

There are three extraction procedures currently used to separate the oil from the other phases in the paste, liquid or solid. These are *pressure*, *centrifugation*, and *percolation*. Specialized manufacturers provide suitable machinery for the proper extraction of olive oil.

Pressure

This is the oldest method but is still in use. In this system, the paste is pressed to release an oily must (olive + vegetation water) (Figure 5). The liquid phase separates from the solid phase with the help of the drainage effect of the mats and the stone fragments. A cake is formed between the mats, while the two liquid phases (oil and vegetation water) are separated by centrifugation.

Figure 5 (see color plate 117) Hydraulic pressure unit (courtesy of ELAIS SA, Piraeus, Greece).

Centrifugation or Three-phase Systems

The crushed olives are mixed with water. A horizontal centrifuge (Figure 6) separates the mass into pomace and must, which is further separated into oil and vegetation water. In the 1970s and 1980s, three-phase centrifuges replaced pressure systems to a great extent to cut processing costs and to reduce olive storage time. A recent innovation to reduce polyphenol losses is the so-called two-phase system, which is able to separate the oily phase without the addition of water. The oily phase is separated from the paste to give oil and water plus husks.

Percolation

A steel plate is plunged into olive oil paste. When it is withdrawn, it will be coated with oil because of the different surface tensions of the liquid phases (the metal phase is coated with a skin of oil). The system

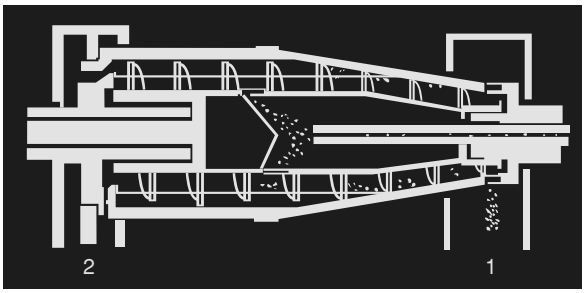

Figure 6 Horizontal centrifuge (decanter): 1, exit of solid phase. 2, exit of liquid phase (oil and water). From Kiritsakis KA (1998) *Olive Oil: From the Tree to the Table*, 2nd edn. Trumbull, CT: Food and Nutrition Press, with permission.

is combined with a continuous horizontal centrifuge to increase the capacity.

Cold Pressing

Cold pressing is the extraction process of olive oil from olive paste at a temperature of less than 25 °C.

Milling Conditions and Quality of Olive Oil

Traditional pressure-extraction systems yield good-quality oil when olives are in good condition and good manufacturing practice is applied. Percolation (selective filtration) also provides good-quality oil. Centrifugation plants have a larger hourly processing capacity and significant practical advantages. In such systems, however, the paste has to be thinned with warm water. This lowers the level of polar natural antioxidants, which have been associated with the storage stability of the oil and its beneficial health effects. Depending on the variety and stage of maturity of olives, different crushing systems are necessary to obtain the optimum polyphenol content and better aroma characteristics. Stone mills yield oils with a good aroma and less intense bitter taste than metal-disk crushers because of a lower polyphenol content. Hammer crushers usually yield oil with a characteristically high content of polyphenols. The malaxation time and temperatures are also critical for the aroma and quality.

Other methods to improve the quality and organoleptic scores during milling are based on the use of pectolytic and cellulolytic enzymes. The International Olive Oil Council, however, does not accept such techniques because olive oil is by definition obtained from the fruits of olive tree only by *'physical means.'*

Olive Oil Composition

The composition of olive oil is primarily triacylglycerols and secondarily free fatty acids, and some 0.5–1.5% of nonglyceridic constituents. The Codex Alimentarius Commission, the IOOC, and the Commission of the European Union provide identity and compositional characteristics for olive oil and olive-pomace oil.

EEC Regulation 2568/91 and subsequent amendments (EC 2632/94, EC 656/95, and EC 2472/97) set the limits for fatty acid composition and content in sterols, waxes, saturated fatty acids in position-2 of triacylglycerols, erythrodiol + uvaol, trilinolein and *trans* unsaturated fatty acids and also for specific absorption at various wavelengths. These limits are very strict and aim at protecting good-quality olive oil, and especially virgin olive oil, from adulteration. The same directives also define analytical methods. (*See* **Fatty Acids**: Properties; Analysis.)

Extended information for the composition of olive oil and other aspects of olive oil chemistry and technology are provided in reviews and books authored mainly by Spanish, Italian, and Greek authors.

Fatty Acids and Triacylglycerols

The fatty acid composition of olive oil ranges from 7.5 to 20.0% palmitic acid, 0.5 to 5.0% stearic acid, 0.3 to 3.5% palmitoleic acid, 55.0 to 83.0% oleic acid, 3.5 to 21.0% linoleic acid, 0.0 to 1.5% linolenic acid, 0.0 to 0.8% arachidic acid, 0.0 to 0.2% behenic acid, and 0.0 to 1.0% lignoceric acid. The olive oil composition may differ from sample to sample, depending on the zone of production, the latitude, the climate, the variety, and the stage of maturity of olives when collected. Olive oil has a fatty acid composition similar to that of hazelnut, almond and high-oleic sunflower oils. This composition differs significantly from any other type of edible oil and fat. (*See* **Nuts**; **Sunflower Oil**; **Vegetable Oils**: Types and Properties.)

The triacylglycerol profile of olive oil, as determined by reversed-phase liquid chromatography, is also different from the profiles of maize, cottonseed, sunflower soybean, and rapeseed oil, and resembles that of hazelnut oil. The main triacylglycerols found in olive oil are: OOO (40–59%), POO (12–20%), OOL (12.5–20%), POL (5.5–7%), SOO (3–7%) (O = oleic acid, P = palmitic acid, S = stearic acid, L = linoleic acid).

Tocopherols

The dietary benefits of olive oil are partly attributed to the presence of α-tocopherol and other natural antioxidants. α-Tocopherol acts not only as a free radical trapping agent but also as a singlet oxygen quencher. This protective effect against photooxidation is enhanced by the presence of β-carotene. The

main homolog of vitamin E forms present in olive oil is α-tocopherol, which makes up approximately 95% of total tocopherols. The other 5% are β- and γ-tocopherols. Good-quality, fresh virgin olive oils have a considerable content of α-tocopherol (200–300 mg kg^{-1}). Refined, bleached, and deodorized olive oils have markedly reduced contents because of the losses of tocopherols during processing. (*See* **Tocopherols**: Properties and Determination.)

Squalene

Squalene ($C_{30}H_{50}$), an intermediate in the biosynthesis of sterols in plant and animal world, is the major olive oil hydrocarbon. Squalene is found in olive oil at a concentration ranging from 0.7 to 12 g per kilogram of oil. The squalene content depends on the olive cultivar and oil-extraction technology, and it is dramatically reduced during the process of refining. There are claims that squalene can enhance the quality of life, if taken continuously, and that its consumption is beneficial for patients with heart disease, diabetes, arthritis, hepatitis, and other diseases. It has been claimed that the high squalene content of olive oil, as compared with that of other human foods, is a major factor in the cancer-risk-reducing effect of this oil.

Carotenoids

The main carotenoids present in olive oil are β-carotene and lutein. Trace amounts of neoxanthin, violaxanthin, cryptoxanthin, and other xanthophyls may be found. The total carotenoids content may range between 1 and 20 mg per kilogram of oil; normal values do not exceed 10 mg kg^{-1}. Oils have been reported that have a higher content of lutein than of β-carotene and *vice versa*. So far, studies have indicated an inverse correlation between β-carotene intake and incidence of cardiovascular disease, but no clear effect on low-density lipoprotein (LDL) oxidation has been demonstrated with β-carotene. (*See* **Carotenoids**: Occurrence, Properties, and Determination.)

Sterols

Phytosterols are functional ingredients. They have an absorption level 20 times lower than that of cholesterol. Phytosterols inhibit the absorption of cholesterol in the body during digestion. Four classes of sterols occur in olive oil, common sterols (4-desmethylsterols), 4-α-methylsterols, 4,4-dimethylsterols, and triterpene dialcohols (erythrodiol and uvaol) (Figure 7). The major class is 4-desmethylsterols. Usual values reported for these classes are 1000–2000 mg per kilogram of oil. Part of the total sterols is present as esters with fatty acids. β-Sitosterol makes up 60–90% of the total sterol fraction. Other sterols found in considerable amounts are δ^5-avenasterol (Figure 8; 5–36% of the total sterol fraction) and campesterol (approx. 3% of the total sterol fraction). The rest of the 4-desmethylsterols present in olive oil are found in trace or very small amounts. Certain sterols are probably responsible for the resistance of olive oil to rapid deterioration at elevated temperatures. (*See* **Functional Foods**.)

Phenolic Compounds

Virgin olive oil contains phenolic substances that affect its stability and flavor. Tyrosol (4-hydroxyphenethyl alcohol) and hydroxytyrosol (3,4-dihydroxyphenethyl alcohol) are usually mentioned as the major constituents. Other phenolic compounds are caffeic acid, *o*-coumaric acid, *p*-coumaric acid, ferulic acid, gallic acid, homovanillic acid, *p*-hydroxybenzoic acid, *p*-hydroxyphenylacetic acid, protocatechuic acid, sinapic acid, syringic acid, tyrosol glucoside, and vanillic acid. Aglycons of oleuropein and ligstroside, the esters of hydroxytyrosol and tyrosol with

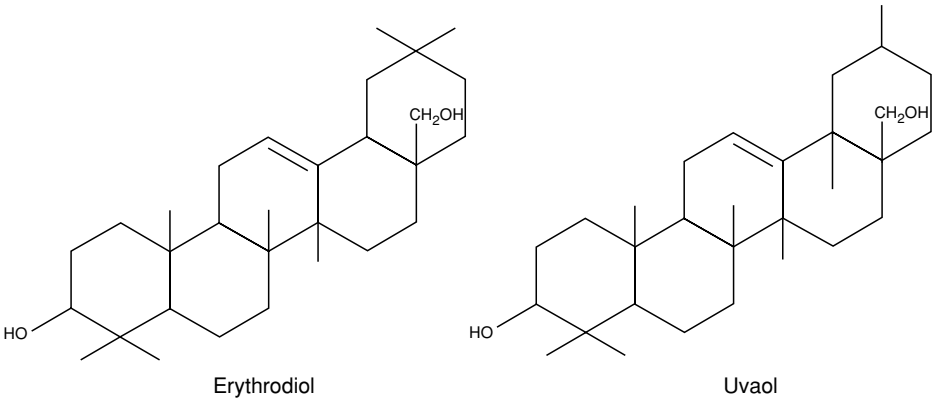

Figure 7 Two triterpene dialcohols used to check purity of virgin olive oil.

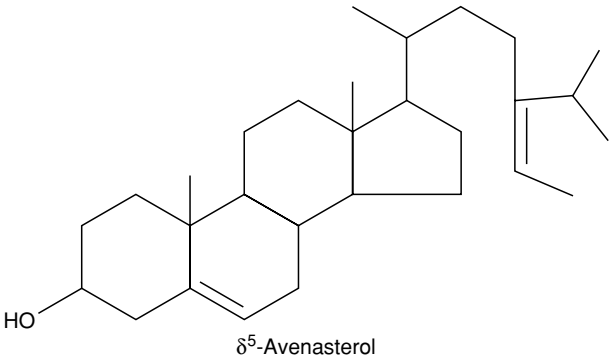

Figure 8 Chemical structure of Delta ^{5}-Avenasterol.

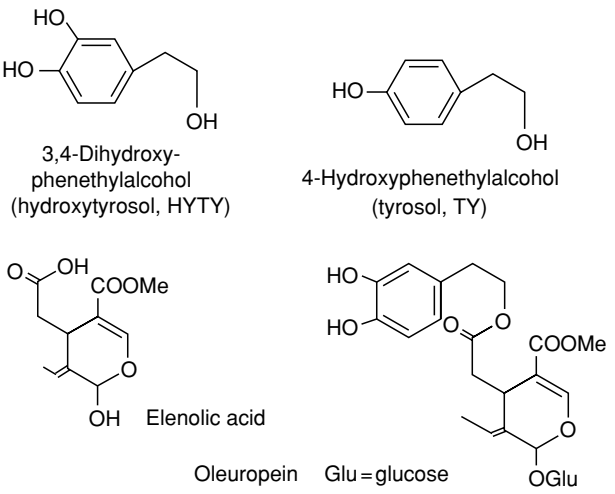

Figure 9 Major phenolic and related compounds in olive fruit and olive oil.

elenolic acid, diacetoxy and dialdehydic forms of these aglycons, elenolic acid, and flavonoids have been also reported to be present in the polar fraction of virgin olive oil (Figure 9).

The content in phenolic compounds differs from oil to oil (a few to more than 400 mg per kilogram of oil, expressed as caffeic acid). When the level exceeds 300 mg kg^{-1}, the oil may have a bitter taste. However, a high polyphenol content appears to be beneficial for the shelf-life of the oil, and there is a good correlation of stability and total phenol or *o*-diphenol content. Among the various phenolic compounds tested for their contribution to the stability, hydroxytyrosol and caffeic acid were found to be the most potent antioxidants.

Dietary antioxidants present in exra virgin olive oil were found to increase the resistance of low-density lipoproteins to *in-vitro* or *in-vivo* oxidation experiments. According to a large number of reports, hydroxytyrosol is the most important biophenol of olive oil similar to the phenolic compounds encountered in green tea and in red wine. Oleuropein also presents valuable functional properties, but it occurs in olive oil only in trace amounts. It is found in abundance in olives and olive leaves. (*See* **Phenolic Compounds**.)

Chlorophyls

The color of olive oil is mainly due to pheophytins α and b. Traces of chlorophyls α and b may be found only in fresh oils. The chlorophyl content in virgin olive oil ranges from trace amounts to more than 30 mg kg^{-1}. The level of chlorophyls (and carotenoids) depends on genetic factors, degree of fruit ripening, and extraction technology. The level decreases as the fruit ripens. In the absence of light, chlorophyls probably act as weak antioxidants. In the presence of light, it is accepted that chlorophyls act as strong oxidation promoters. Chlorophyls are responsible for the yellow/green hues of the oil. (*See* **Chlorophyl**.)

Volatile and Aroma Compounds

More than 100 constituents have been identified in virgin olive oil, mainly hydrocarbons, alcohols, aldehydes, esters, and ketones. However, only a few of these compounds contribute significantly to its aroma. The most important odorants are volatiles with six carbon atoms, arising from the enzymic oxidation of both linoleic and linolenic acids during olive crushing and malaxation. Genetic, agronomic, and technological parameters, such as olive variety, degree of fruit ripening, crushing, and malaxation conditions or extraction system affect the enzymic activity and alter the quality and intensity of the aroma. Certain compounds were found to be responsible for desirable flavor attributes, e.g., (Z)-3-hexenal for the 'green, apple-like' odor, hexanal for the 'grassy' odor, (E)-2-hexenal for the 'green, bitter almond-like' odor, ethyl 2-methylbutanoate for the 'fruity' odor. Other compounds such as octanal, nonenals, decadienals, 1-octen-3-one, acetic acid, 3-methylbutanol, 2-phenylethanol, etc., are responsible for some negative attributes. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Storage and Packing

The oxidative stability of virgin olive oil is mainly related to its characteristic pattern of triacylglycerols (low unsaturation, *iodine value* < 90) and also to the considerable levels of polar phenolic antioxidants and the presence of α-tocopherol. Since the early 1970s, it was found that there was a good correlation between polyphenols and the oxidative stability of the oil, and

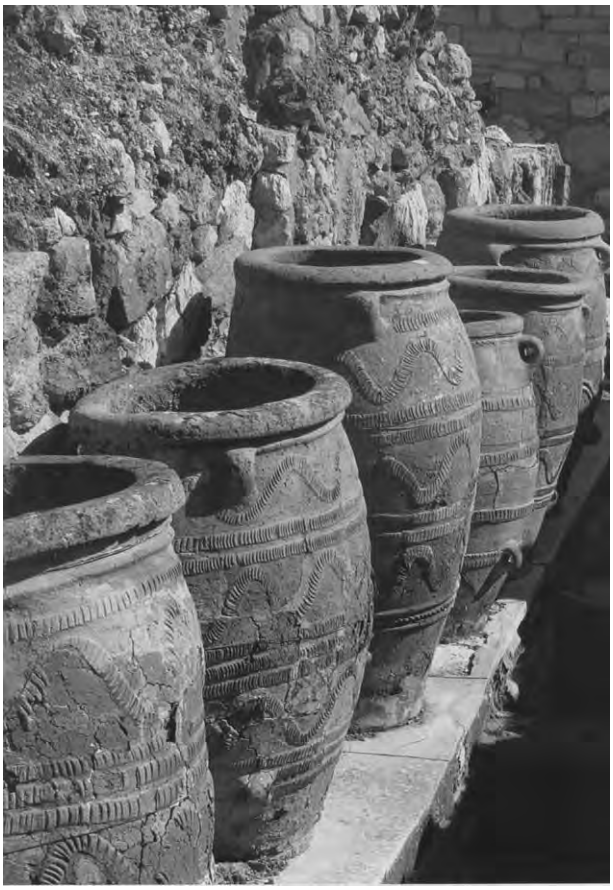

Figure 10 (see color plate 118) Pottery from Knossos Palace. Reproduced from Psillakis N and Kastanas E (1999) *The civilization of olive: olive oil*, 2nd edn. Greek Academy of Taste, Karmanor, Iraklion, Crete (Greece), with permission.

all the researchers agreed that the level of polyphenols or of specific phenols, e.g., tyrosol level, coincided well with the overall oil quality. Thus, oils with a high initial peroxide value may have an unexpectedly long shelf-life due to a high content of natural antioxidants. In the last decade, some researchers focused their interest on the contribution of individual phenols present in virgin olive oil to its oxidative stability. Recent publications indicate that there is an interest in all natural antioxidants of olive oil that are studied in parallel. The importance of polyphenols for olive oil stability explains the efforts of investigators to develop an appropriate technology to recover them from olive oil wastewater.

Bottled virgin olive oil has a shelf-life of around 18 months. This period is very long for a fat product. The resistance of the oil to oxidation may be supported by special storage and packing conditions (see Figure 10).

Storage in dark places is essential for the keepability of the product to prevent the prooxidant activity of chlorophyls in the presence of light. Storage in cool places protected from air access is suggested to prevent autoxidation. Therefore, stainless steel containers are ideal for storage. Glass bottles are resistant to fat and oxygen permeation, but only colored bottles offer light protection.

Traditional and Modern Use

Olive oil is part of a cuisine that is simple, light, and placid, with defined tastes and full of harmony, the so-called Mediterranean cuisine. Olive oil resistance to the development of rancidity is combined with a vast array of flavor and color hues and distinct features due to differences in cultivars of olives from which the oil is extracted. A good-quality olive oil blends perfectly with the greens. The exquisite taste of olive oil is very often complemented by the sharp taste of vinegar, lemon, or tomato. In salads or in cooking, olive oil is usually mixed with herbs and spices, which are also an important element of the Mediterranean diet. Olive oil shows a remarkable resistance during domestic deep-frying of potatoes or in other uses at frying temperatures due to its low unsaturation. It is, therefore, recommended not only as a salad oil but also for cooking and frying.

See also: **Carotenoids**: Occurrence, Properties, and Determination; **Chlorophyl**; **Coronary Heart Disease**: Etiology and Risk Factor; Antioxidant Status; Prevention; **European Union**: European Food Law Harmonization; **Fatty Acids**: Properties; Analysis; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Functional Foods**; **Nuts**; **Olives**; **Phenolic Compounds**; **Sunflower Oil**; **Tocopherols**: Properties and Determination; **Vegetable Oils**: Types and Properties

Further Reading

Boskou D (ed.) (1996) *Olive Oil, Chemistry and Technology*. Champaign, IL: AOCS Press.

Boskou D and Elmadfa I (eds) (1999) *Frying of Food*. Lancaster, PA: Technomic Publishing.

Codex Alimentarius Commission (1993) Revised Norm for Olive Oil, C1 1993/15-FO.

Commission of the European Communities (1991) Regulation No. 2568/91: On the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis. *Official Journal of the European Commission* No. L248.

Fedeli E (1977) Lipids of olives. *Progress in the Chemistry of Fats and other Lipids* 15: 57–74.

Gracian Tous J (1968) The chemistry and analysis of olive oil. In: Boekenoogen HP (ed.) *Analysis and Characterisation of Oils, Fats and Fat Products*, pp. 315–606. London: Interscience Publishers.

Harwood J and Aparicio R (eds) (2000) *Handbook of Olive Oil, Analysis and Properties*. Gaithersburg, MD: Aspen.
IOOC (1996) *International Encyclopedia of Olive Tree*, 1st edn. Madrid: International Olive Oil Council.
IOOC (1999) Trade Standards Applying to Olive Oil and Olive Pomace Oil. COI/T. 15/NC, Doc. No. 2; rev. 9. Madrid: International Olive Oil Council.
Kiritsakis KA (1998) *Olive Oil: From the Tree to the Table*, 2nd edn. Trumbull, CT: Food and Nutrition Press.
Kumpulainen JT and Salonen JT (eds) (1996) *Natural Antioxidants and Food Quality in Atherosclerosis and Cancer Prevention*. Cambridge: The Royal Society of Chemistry.
Newmark HL (1997) Squalene, olive oil and cancer risk: A review and a hypothesis. *Cancer Epidemiology, Biomarkers and Prevention* 6: 1101–1103.
Perrin JL (1992) Les composes mineurs et les antioxygenes naturels de l'olive et de son huile. Mise au point. *Revue Française des Corps Gras* 39: 25–32.
Ryan D and Robards K (1998) Phenolic compounds in olives (a review). *Analyst* 123: 31R–44R.
Tsimidou M (1997) Polyphenols and quality of virgin olive oil in retrospect (a review). *Italian Journal of Food Science* 10: 99–116.

OLIVES

B L Raina, Regional Research Laboratory (CSIR), Jammu, India

Introduction

Olive (*Olea europaea*, family Oleaceae) is an evergreen xerophytic tree grown for its drupes, which yield oil and are also marketed as table or pickled olives. The olive seems to have been cultivated since prehistoric times in the eastern Mediterranean region, and has been closely associated with human religious, sociocultural, medicinal, and nutritional needs. Today, olive cultivation has attained commercial crop status in every country with a Mediterranean climate. Spain, Italy, and Greece are leading producers of olives. Other important olive-producing countries are Turkey, Tunisia, Portugal, France, Morocco, Algeria, Syria, Yugoslavia, Jordan, the USA, Cyprus, Israel, Argentina, and Libya.

Production and Trade

There are about 800 million olive trees in the world, covering about 9 million hectares (ha), with the annual olive fruit production being about 13.7 million metric tonnes (t). Approximately 720 000 t of the production are consumed as table olives, and the remaining 8 680 000 t are used for the extraction of oil (**Table 1**). Nearly 2.3 million t of olive oil and 230 000 t of olive residue oil is produced globally. About 260 000 t of olive oil and 200 000 t of table olives is traded in international markets. Olive oil is exported by 35 countries to more than 100 countries in five continents. Spain is the largest exporter and Italy buys nearly half of the entire quantity traded in the international market. Of the 300 000 t of olive oil produced in Greece, around 65% is consumed on the domestic market, whilst 100 000 t is exported, mostly to the European Union (EU). Seventy-five percent of all the olive oil produced in Greece is edible virgin olive oil, for which the global market is expanding. With its 4.2% share of world olive oil production and 8.7% share of world exports. Turkey ranks fourth and in some years fifth after Tunisia. The annual increase in olive oil production is 1.4% as compared to 4% in respect of other edible oils.

As the world's largest producer, Spain produces between 200 000 and 250 000 t of table olives per year. This is over 60% of EU production. Half is for the domestic market and half is for export. Greece is the second largest producer of table olives and exports of the typical table varieties are greater than

Table 1 Major production of olives and olive oil

Country	*Fruit production (1000 MT)*			*Oil production (1000 MT)*		
	1996	*1997*	*1998*	*1996*	*1997*	*1998*
Spain	4506	5686	3564	1037	1164	777
Italy	2147	3591	2680	420	698	497
Greece	1950	1879	2068	470	457	427
Turkey	1800	520	1550	197	56	56F
Tunisia	1550	500	1000F	329	96	210
Morocco	908	518	650	87	50	66
Syria	648	403	763	140	85	121F
Portugal	284	267	287F	46	43	43
Algeria	313	319	124	50	49	47F
Argentina	92	92F	92F	4	12	12F
World	15 308	14 702	13 757	2848	2770	2316

F, Food and Agriculture Organization estimate.
Source: www.fao.org.

sales on the home market. The production of table olives in Italy is around 86 397 t per year. Table olives account for about 3.4% of the annual crop, and barely 10% of world production. Turkey is one of the world's biggest producers and consumers of table olives, and lies sixth in the export ranking. It is the world's top producer of black olives (29%) and their second biggest exporter (19%).

The crop is very labor-demanding, provides employment to millions, and accounts in many countries for a major share of agricultural income. Almost 25% of the farming income in the Mediterranean basin is from olive products. The olive products, which have been consumed by the Mediterranean people for generations, are increasingly becoming a luxury in countries with a high standard of living.

Botany

Olive is an evergreen tree 3–12 m or even more in height and, in some instances, bears fruit for 1000 years or even more; branches are numerous; leaves are opposite, leathery, lanceolate, dark green above, silvery beneath; flowers are small, white, dimorphic (male and perfect), appearing in early spring in loose axillary clusters. Fruit setting in olive is erratic in certain areas, especially where irrigation is insufficient, the plantation is dense, the pests and diseases are not controlled and, above all, a single cultivar is planted over a wide area, reducing the chances for cross-pollination.

The olive fruit is a globular, oblong, or sometimes crescent-shaped drupe. The pericarp is composed of epicarp or skin, the mesocarp or pulp, and the endocarp or stone (pit) which contains the seed. The fruit reaches its maximum weight 6–8 months after flowering late in spring, and passes through successive shades of straw, pink, and red, before finally turning purplish-black at full maturity. The ripe fruit weighs from 1.5 to 13 g. The pit with its hard shell makes up 13–30% of the fruit weight and the skin 1.5–3.5%. The seed does not exceed 3% of the weight of the fruit. At full maturity, the mesocarp contains 6–10% soluble solids and 15–40% or even more oil, depending on variety. The pericarp contains 96–98% of the total amount of oil, while the remaining 2–4% of the oil is in the kernel. The characteristic bitter glycoside, oleuropein, is more concentrated close to the peel.

Varieties

The principal varieties cultivated for oil have medium-sized fruits and are harvested when ripe. More important among these are Arauco, Bouquetier, Corfolia, Dafnolia, Fratoio, Koroneiki, Lechin, Manzanillo, Morcal, Nevadillo, Picual, Rozzola, Rougette, Smertolia, Taggiasea, Tsounati, and Zorzalena. The varieties preferred for table olive production, with less than 8% oil, include Ascolano, Calamata, Chalkidikis, Conservolea, Gordal, Hojiblanca, Manzanillo, Megariticci, Mission, Sevillano, Throumbolia, Verdale, and Volou.

Physiology and Biochemistry

Fruit Development and Ripening

Along with an increase in fruit weight, various changes in constituents occur during the development of the fruit. In the cultivar Gordal, there is a continuous increase in the oil content, a brief rise in reducing sugar content, followed by a decline till maturity, a sharp initial drop in crude fiber content and a gradual one afterwards, and a low protein and ash content. There are no qualitative changes in chlorophyl and carotenoid pigments in Hojiblanca and Manzanillo olives during growth and ripening. The pigment content in the virgin olive oil varies according to the degree of ripeness of the fruits. Initially the major component is pheophytin, while in oil obtained at the end of the season lutein is the most abundant. The loss of chlorophyl pigments in the fresh fruits and virgin oil is not due to activation of chlorophyllase during oil processing. The β-carotene content and its provitamin A value diminish with ripeness (13.2 μg β-carotene per g oil is reduced to 1.27 μg g^{-1} in Sevillano variety). The considerable range in the lutein/β-carotene ratio (between 1.3 and 5.1 depending on variety) makes this ratio a differentiator of single-variety oils. The amount of coumaric acid, syringic acid, and flavonol varies with maturity. The quantitative differences in the composition of flavonol and flavone glycosides during ripening may be useful in the biochemical characterization of olive cultivars. Cultivar and stage of maturity influence tocopherol and tocotrienol contents. The concentrations of the volatile compounds during three stages of ripeness in four different cultivars are reported to decline with increasing ripeness, although hexyl acetate is highest and hexan-1-ol lowest at the ripe stage. During maturation of the fruit, the size of oil droplets within the olive increases, raising the extractability of the oil from about 20% in the early stage to as much as 90% in fully mature fruit. Furthermore, small fruits are characterized by high oleuropein and low verbascoside contents. The gradual loss of firmness and of anhydrogalacturonic acid content of the pectic chain is associated with increasing enzyme activities. Pectinesterase activity appears

during ripening and polygalacturonase activity mainly during storage. The activity of cellulolytic enzymes increases with ripening of the fruit. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Ripening of Fruit**.)

Constituents of Olive Fruit

The moisture and oil content form 85–90% of the weight of pulp, while the rest comprises organic matter and minerals. The major monosaccharides in fruit pulp are glucose, mannose, xylose, galactose, and arabinose, accompanied by mannitol and rhamnose in certain cultivars. The pulp is rich in potassium. Small amounts of organic acids, such as citric, malic, oxalic, malonic, fumaric, tartaric, lactic, acetic, and tricarballylic (1,2,3 propanetricarboxylic acid) are present. (*See* **Acids**: Natural Acids and Acidulants; **Carbohydrates**: Classification and Properties.)

A variety of phenolic compounds are found in olive pulp; caffeic and ferulic acids are among the simpler of them. The major orthodiphenolic compound typical of olives is oleuropein, which is responsible for the bitter taste of immature olives. Two linear compounds are reported to be isolated from the ethyl acetate extract of residues resulting from Spanish olive oil processing. These compounds are identified as 3-(1-(hydroxymethyl)-(E)1-propenyl) glutaric acid and 3-(1-(formyl)-(E)1-propenyl) glutaric acid. These products are structural components of oleuropein. Oxidation products of oleuropein and other phenolics impart the black color to the fruit as a result of complex interactions between diphenol oxidase and oleuropein. Anthocyanins, specifically the glycosides of cyanidin and peonidin, are responsible for the purple and black color of ripe olives. The highest concentration of anthocyanin pigments has been observed in the epicarp ($0.55\,g\,kg^{-1}$); a complex mixture of flavonoids is found in the mesocarp; derivatives of luteolin and apigenin are found in the endocarp. (*See* **Phenolic Compounds**.)

The skin, pulp, and seed contain different lipid fractions and fatty acids, sterols, triterpene alcohols, dialcohols, and hydrocarbon fractions. Among the differences between the parts of the olive drupe are the presence of erythrodiol and traces of uvaol, oleanoic acid, traces of ursolic acid, and oleanoic aldehydes in the skin only. The distribution of unsaponifiable hydroxy compounds fractions within the drupe and their concentration in oils from various parts of Lingurian olives reveal the presence of sterols, 4-methylsterols, triterpene alcohols (TTA), triterpene dialcohols (TTDA), linear saturated alcohol (LSA), and phytol. Most of the sterols, 4-methylsterols, TTA, and phytol are found in the pulp, while LSA and TTDA are found in the peel. The main hydroxy components are LSA in the peel, TTA in the pulp, and sterols in stone and seed. The kernel oil is characterized by the presence of estrone ester (8 μg per 100 ml oil).

Fifty-six volatile compounds are reported to be present in the leaves, flowers, epicarp, and mesocarp of Lucca and Mission varieties, and 29 compounds, including 3-vinylpyridine, methylvinylpyridine, and *cis*-jasmone, have been identified in these varieties. Higher concentrations of hexanal and *trans*-2-hexanal are found in the Lucca variety, while numerous long-chain hydrocarbons are found in flowers of both varieties. The peroxidase activity and lipid peroxidation intensity in leaves may be of potential use in selecting olive cultivars for yield.

Soil and Climate

Olive does well on moderately saline soils not suitable for most other crops. The soil should be deep and well-drained, so that the roots penetrate deeply to take advantage of underground moisture even in relatively arid areas. The olives can also grow on calcareous slopes provided the fertility of the soil is built up.

Olive can be grown commercially in warm-temperate and subtropical countries located between latitudes 30° and 45°N and 30° and 45°S, at 300–400 m altitude in the north to 1000–1220 m above sea level in favorable sections of the south. The production of olives under dry conditions needs better resource management. This includes use of special tools and traditional water conservation measures. It cannot be grown where the winter temperature falls below −9°C, as this temperature will kill the trees. A certain degree of chilling, differing considerably among cultivars, is necessary for floral induction. The olive blooms relatively late in the spring. A long, relatively hot growing season is required for the fruit to develop to the appropriate size. A consistently clear and dry atmosphere is necessary for maximum yield of fruit. Adequate soil moisture must be present throughout the year to produce heavy yields of the required size of fruit.

Cultural Practices

Propagation

Olives are propagated by hardwood cuttings, or by small softwood cuttings rooted under mist sprays, or by 'ovule' protuberances formed at the base of the trunk. The treatment with IBA+ putrescine HCl and

placing the cuttings in the polyethylene tunnel improve the rooting ability. Although propagation by seeds provides a stronger root system and more fruitful trees, the seedlings produced require grafting and thus a longer time to reach fruiting. The variable nature of the rootstock may bring about variability in the growth and size of the trees. The nursery trees are planted 8–12 m apart in irrigated orchards or 12–22 m apart in unirrigated groves.

IRTA-i.18, a clone selection of the olive Arbequina, is superior to the cultivar for fruit color, fruit yield at a density of 300 trees h^{-1}, crown volume, fruit weight, flesh/stone ratio, and oil content, and identified as suitable for commercialization. The topical pollen application can be used instead of interplanting of pollinizer trees to reduce shotberry incidence in Manzanillo olive groves.

Irrigation and Drainage

For production of a good crop of large fruit size, the trees should have adequate soil moisture throughout the year, especially before flowering time in late winter and early spring and during the summer, when fruit growth takes place and new fruiting wood develops. In the Mediterranean region both oil and pickling olives are largely grown without irrigation. Under rain-fed conditions the olives require 40–50 years to reach full productivity, necessitating the use of many intercrops to provide income during the interim period. The yield of full-bearing olive trees in Spain is about 825 kg ha^{-1}, and water is the main limiting factor for low yield.

In California, the olive groves are grown under irrigated conditions. To insure adequate tree growth and profitable yields, supplementary summer irrigation is necessary. Olive is very sensitive to excessive soil moisture. Poor drainage at any time, if prolonged, will kill the tree.

Fertilizers

The olive tree can obtain the required amount of potassium and phosphorus from the soil. However, nitrogen is utilized by the plants more rapidly, necessitating the application of nitrogenous fertilizers. Under irrigated conditions in California, the growers apply 450 g nitrogen per tree per year just before onset of the winter season. In gravelly, shallow, foothill types of soil, the trees respond well to potassium and boron. In Greece, a fertilizer mixture containing nitrogen, phophorus, and potassium is used.

The organic method of cultivation using *Hedysarum coronarium* cover crop inoculated with a granular formulation of *Rhizobium hedysarii* to supply nitrogen to olive significantly increases the total and organic nitrogen in the soil and has an impact on olive productivity.

Pruning and Thinning

Pruning methods for the bearing trees vary greatly. In the Mediterranean region, heavy pruning is practiced. Heavy pruning under irrigated conditions leads to reduced yields owing to the removal of potential fruit-bearing surfaces, and results in fruit bearing in alternate years. A moderate amount of annual pruning, involving removal of injured, dead, or ill-placed branches, is all that is necessary to spread the fruiting surface uniformly, so that every part of the bearing wood is adequately exposed to light and air. The height of plants in regular plantation is generally maintained at up to 5 m. In some countries, thinning of fruits is achieved by spraying the young fruits with 150 p.p.m. naphthaleneacetic acid. Spray-thinned trees may be expected to bear fruit the following year. The girdling with GA3 plant regulator application (50 or 100 p.p.m.) increases the fruit yield.

Pests and Diseases

Olives have many pest problems, some of them affecting the tree or the fruit itself. Olive fruit fly (*Dacus oleae* Gmel.), olive kernel borer (*Prays oleae* Bern.), olive scale (*Parlatoria oleae* Colvee.), Oleander or ivy scale (*Aspidiotus hederae* (Vallot), greedy scale (*A. camelliae* Signoret), California red scale (*Aonidiella aurantii* (Maskell), and black scale (*Saissetia oleae* Bern.) are the important pests of olive. Two recent reports of beetles injurious to the olive tree involve *Nericonia nigra* and *Noemia semirufa*.

The olive fruitfiy (*Dacus oleae* Gmel.) is a major pest in the Mediterranean region. The insect attacks at early fruit maturity, causing fermentation and premature fruit dropping. Sometimes, owing to adverse climatic conditions, the production loss due to this fly can be as high as 30%. The infestation adversely affects the oil quality by increasing acidity, especially when infestation exceeds 30%. The pest is best controlled by proteinaceous attractant sprays. (E)-2-hexanal emitted by the crushed olives has repellent property to the fruitfly. Chrysopids are the most active and effective predators on olive kernel borer (*Prays oleae* Bern). The pesticides deltamethrin, fenthion, and formothion are also used to control *Dacus oleae*. The use of green immature olives for pickling has become popular in Spain as a means of escape from the olive fly.

The principal disease of the olive tree is 'olive-knot,' caused by *Bacterium savastanoi* Stevens, which leads to knots or tumors on the leaves, twigs, branches, and trunk. There is no proper control for

this disease, except to avoid olive varieties which are susceptible to cold injury and to this organism, or by quarantine. The root-knot, citrus and reniform nematodes can be suppressed by less irrigation frequency and predacious nematode in olive.

Peacock spot (*Cycloconium oleaginum*) and *Verticillium* rot (*Verticillium alboatrum*) are the fungal diseases of olive, occasionally causing heavy losses in production and fruit quality. The problem is serious in humid areas, but can be controlled by various fungicides such as Bordeaux mixture or lime sulfur. Soil solarization of individual diseased trees, when combined with chemical weed control, is effective in controlling *Verticillium* wilt in olive trees. (*See* **Fungicides**.)

Harvesting

Traditionally, the olives are harvested by hand into a receptacle in the autumn or early winter at the appropriate green-yellow stage. Olives for oil extraction are collected by beating the branches of the tree with long poles so that the fruit falls on to cloth sheets laid on the ground. This is done after the fruit has turned black. Beating of the trees tends to promote the spread of olive-knot disease through injury to the fruit and tree. In Greece and elsewhere, the scarcity of farm labor has resulted in the introduction of plastic nets for collection of naturally falling fruit. To obtain better-quality oil, the olives should not be left on the nets for more than 15 days. Chemical sprays to induce fruit abscission are used to aid the mechanical harvesting of olives. Mechanical shakers with stationary cloth, canvas-covered, or mechanized collecting frames are used to a limited extent, mostly for harvesting olives for oil. A 'beating-down' machine for mechanical picking of olives without damaging the trees has been tested in Italy. The machine is particularly suitable for high-yielding trees with well-developed foliage, and for old trees, for which vibrating picking machines are not recommended.

For high-quality table olives, where uniformly ripe fruit is a requisite, the fruit is handpicked using ladders. The finest and most highly prized oil is made from olives picked just before they begin to soften and darken. The time of maturity may vary within the cultivated area, among the trees within the same grove, or within the fruits on the same tree.

The measurement of Brix values allows for rapid, inexpensive estimation of the oil content of olive fruits at the harvest site and is useful in determining optimum harvest time. The respiration rate is another new ripening index for drupes and the harvesting of olives is suggested to be conducted during the climacteric phase of the fruits.

Postharvest Handling and Storage of Fruit

Ideally, the extraction of oil should follow the harvesting of the olives without delay. As this is impractical, the olives must be stored for a period before processing. During storage, several chemical and biochemical changes may occur that lead to the deterioration of the oil. The most serious damage is caused by enzymes indigenous to the olive tissue and those produced by the microflora on it. The pathogenic and saprophytic fungi isolated from stored olives, except *Saccharomyces* spp., increase the acidity of the extracted oil, and no aflatoxins are reported to be detected in samples infected by *Aspergillus flavus*. As the olive tissue respires, heat is produced, which, if not effectively dissipated, accelerates the enzymatic actions. Lipolysis, lipid oxidation, and other undesirable reactions occur, leading to a lowering of the oil quality.

To minimize the effect of these activities, the olives are stored in 25-cm-thick layers in cool buildings. The use of perforated trays for spreading and stacking saves space. The olives are better stored under water in tanks containing mild preservatives, such as 3% salt or 0.03% citric acid and 3% salt, or 2% metabisulfite. Aerobic storage with a low concentration of acetic acid gives stable color and texture even after 3 months; bulk storage can be done in 50% ethanol. Aerobic storage and the maintenance of a CO_2 concentration of 10–15 mg per 100 ml throughout the storage period eliminates shriveling. Fruit maturation is delayed by refrigeration compared with storage at room temperature and the firmness of the fruits stored at room temperature decreases much more rapidly. The sensory and chemical qualities of the oil produced from fruits stored at 5 °C can be maintained for fruit storage durations up to 45 days, but with storage at 8 °C, these qualities are maintained for storage durations up to 15 days. For fruits stored at room temperature, oil quality deteriorates after 7 days of fruit storage. The processing period of oil mills can be extended by freeze-storing of olives at −18 °C for 90 days. This also helps to increase oil extraction. (*See* **Preservation of Food**.)

Processing and Utilization

Oil is extracted traditionally by simple equipment involving crushing, pressing, and separation of oil from the liquor. It is now being replaced by modern mechanical equipment of widely varying designs, either as separate units or combined as a single-stage process. The olive oil recovery processes have been patented in many countries.

The oil obtained from sound fruit by pressing without further treatment is called virgin olive oil. After

the first pressing, the residual pulp is still rich in oil and is usually recrushed and repressed with or without the addition of warm water. Oil obtained from the second pressing tends to have a more intense color and higher acid content, as well as a weaker aroma. This latter oil, together with inferior virgin oil, is further subjected to refining treatments such as neutralization, deodorization, bleaching, and winterization; this helps in removing acid, color, and odor. The oil so obtained is called refined olive oil, which is largely used for mixing with the first extraction to produce edible grades. Refining procedures decontaminate oil containing fenthion residues. The remaining press cake still retains about 4–5% oil, but more than 50% of the total oil in the cake can be obtained in a single step by direct extraction in a continuous centrifugation system, after adding about I % Na_2CO_3 by weight of cake. It is then necessary to dry and further deoil the cake for better stability towards autoxidation. The dried cake is extracted with hexane and the resulting 'sulfur oil' is repeatedly rectified. The extraction with hexane results in an oil with higher acidity and moisture content, but a lower percentage of impurities.

Olive oil has been classified by the United Nations Conference on Olive Oil, held in Geneva in 1961, into four groups according to the method of preparation and acid content:

1. Virgin olive oil. Oil extracted by pressing, free of admixtures, called 'extra' when the oleic acid does not exceed 1 g per 100 g, and 'fine' if the acid content does not exceed 1.5 g per 100 g and the flavor is perfect; 'ordinary' olive oil may contain acid up to 3.0 g per 100 g and have a slight off-flavor; if the oil has a definite off-flavor it is classified as 'lampante.'
2. Refined olive oil. This oil may be called 'pure' when it is refined from virgin oil, and 'second-quality' when it is refined from solvent-extracted oil.
3. Blended olive oil. Blended oil can be called 'pure' when it consists of a blend of virgin and refined oil, and 'blended' when it contains a blend of virgin and second-quality refined oil.
4. Industrial oils. These oils are obtained by extraction of olive residue with solvents.

Storage and Packing of Oil

Olive oil undergoes deterioration during storage. It is important to store the oil extracted from olives of different varieties separately. The storage drums and tanks should be of inert material, impermeable to oil, and lined with epoxy resins, enameled tiles, or glass; and the oil should be stored at a constant temperature of about 15 °C.

Olive oil is packed in several types of retail containers including bottles made of glass, polyvinyl chloride (PVC) or polyethylene, tin-plated cans, and in tetrabriks. The containers should provide protection of the oil from light, and leave a minimum volume of head space. Packing of oil under vacuum and inert gases is also recommended.

Table Olives

Many different methods of converting olives into olive products exist in different olive-growing countries. The traditional method of preparing table olives by removing their bitterness through alkaline treatment, which grew up in the environs of Seville, Spain, is rapidly being replaced by technically advanced methods. Production has been increased and costs have been cut, and the industry has remained highly competitive. Sevillian green olives are the predominant type used (70%), followed by oxidation-darkened olives (25%), and untreated black olives (5%). The chief innovations in the last 25 years have been mass fermentation, development of conservation in an aerobic medium, development of automatic stoning and stuffing machines, developments of pastes, fixing of pack conditions and treatments, and general mechanization of all handling, packing, and storage of the end product.

The major commercial methods are as follows.

Green Fermented Olives (Spanish Method)

The fruits are picked when they are still firm and light green in color. The principal varieties processed in Spain are the Sevillano and Manzanillo. The fruit is covered in large containers with 1.8% lye solution and kept at 28 °C for 4–8 h. The treated olives are frequently washed with water for 24–36 h.

The washed fruits are then covered with brine in fermentation tanks. This changes the color of the olives from deep green to the typical 'olive-green' shade. The product, on attaining equilibrium, contains 6–8% NaCl. Spices may now be added to the fermented olives according to consumer preference. Closed containers help prevent the development of surface yeasts and molds.

The fermented olives, before marketing, are graded according to shape, size, and color. Sometimes the olives are pitted and stuffed with pimentos, onions, almonds, anchovies, or other products. The fermented brine is filtered (or replaced by fresh brine) for covering the olives in bottles or cans prior to pasteurization.

Olives at different stages of ripeness suffer major losses during the pitting process; green unripe olives are excessively firm, and the cell wall of changing-color

olives is too disorganized to give good industrial results. The treatment with high concentration of lye and its 100% penetration into the flesh, irrespective of fruit size, results in significant textural decline and changes in the composition of polysaccharides that are most important to the cell wall structure.

Among the major spoilage organisms, zapatera – a butyric acid producer – is important.

Canned Ripe Olives (American Style)

The principal commercial varieties packed are the Ascolano, Manzanillo, Mission, and Sevillano. Straw-yellow to cherry-red olives are graded according to color and size to insure uniform lye penetration. The fruits are placed in 5.8% brine for 6 weeks or longer in large wooden or concrete tanks, and the brine concentration is gradually raised to 10%. The fruit is then transferred into shallow vats and treated 4–8 times with gradually decreasing strength of lye (3.0–0.5% NaOH). Each treatment is followed by exposure to air for 1–5 days, either with frequent turning in the presence of air or by bubbling air through the immersed fruit. The olives are then leached by frequent changes of water for 5–7 days until all traces of NaOH are removed. (*See* **Canning**: Principles.)

The washed olives are pasteurized, cured in 2–3% NaCl for 2–6 days, graded, and packed into enameled cans covered with 2.5–3.5% NaCl solution, sealed and sterilized at 116 °C for 1 h. Pitting and stuffing of olives are often practiced.

Black (Naturally Ripe) Olives (Greek Method)

The Greek black olive industry uses naturally ripe, fully matured and dark-purple fruits of mainly Calamata, Conservolea, and Magaritici varieties. The fruits are covered with brine, and the concentration is gradually increased to prevent shriveling. The covering brine concentration is kept at 10% NaCl during winter and increased to 15–16% during summer to prevent spoilage. The bitterness is lost within 3–6 months, and the lactic acid formed during fermentation does not exceed 0.5%. The dark-purple anthocyanins of the olives turn light red during fermentation.

The olives, while being prepared for marketing, are exposed to air until the dark color is regained. The fruits are carefully sorted and graded according to international standards, and repacked in fresh brine containing about 8% NaCl and 0.5–0.75% acetic acid.

Antifungal treatment trials of 'Greek-style' black olives indicate that sorbic acid 0.075% is the most efficient preservative, followed by 0.075% benzoic acid and 0.032% calcium proprionate. In the olives inoculated with *Aspergillus parasiticus*, the preservatives totally inhibit aflatoxin formation, while the toxin can be detected at low concentrations in all controls.

The spoilage organisms encountered are zapatera, galazoma, and film-forming pectolytic organisms which disintegrate the 'meat' of the fruit. This last problem can be prevented by means of a thin layer of paraffin oil on the surface of the brine.

Utilization of Byproducts

Olive Cake or Meal

The press residue may be used as a fuel, manure, animal feed, soil conditioner, fiber (cellulose and lignin) for food use, or in single-cell protein production. Sludge obtained from olive-processing plants enriched with nitrogen fertilizer can be applied to olive orchards growing on alkaline alluvial soils and as manure for maize. The spent form of olive oil cake can be incorporated into the diets of rabbits at 30% level. Mixtures of organic acids of low molecular weight and fatty acids can be recovered from the waste gases produced during drying of cake. *Pleurotus eryngii* can be grown on a medium containing olive husks.

Olive Stones

The olive stones find use in molded products, plastics, and furfural manufacture. Activated carbon produced from olive stones has a high adsorption capacity against lead.

Olive Vegetation Water

Tocopherols, flavor compounds, antioxidants, and anthocyanins can be recovered from olive vegetation water. Olive oil deodorization distillate contains squalene in a concentration range of 10–30%. Squalene in high purity with 90% yield can be recovered by supercritical carbon dioxide extraction. The application of 100–200 m^3 waste waters from olive oil-processing plants per hectare, 45 days after sowing, markedly increased barley and sunflower yields and was reported to have no adverse effects on soil microorganisms. (*See* **Antioxidants**: Natural Antioxidants; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Tocopherols**: Properties and Determination.)

Nutritional and Health Aspects

The use of olive oil, whether in medicine or as source of nutrition, has its roots in ancient history. During the last few decades, renewed interest has been generated in the nutritional and health aspects of the olive. The oil is used largely for culinary

purposes, in wool combing, the manufacture of toilet preparations, and in the pharmaceutical industry.

Besides its physiological advantage in being rapidly and completely digested, the oil also has clinical importance e.g., antiulcer activity, effectiveness in combating gallbladder disease, and in lowering plasma cholesterol level. Omega-6-polyunsaturated fatty acid in olive oil may exacerbate adjuvant-induced arthritis. Olive pollen extract can induce an asthmatic response. (*See* **Fatty Acids:** Dietary Importance.)

See also: **Acids**: Natural Acids and Acidulants; **Antioxidants**: Natural Antioxidants; **Canning**: Principles; **Carbohydrates**: Classification and Properties; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Fats**: Classification; **Fatty Acids**: Properties; Dietary Importance; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Fungicides**; **Phenolic Compounds**; **Preservation of Food**; **Ripening of Fruit**; **Tocopherols**: Properties and Determination

Further Reading

Alves Mdo CFGP and Pinheiro Alves Mdo CFG (1991) Rationalization of national olive oil production capacity: a critical review, from harvesting to storage of olive oil. *Revista de Ciencias Agrarias* 14(4): 59–71.

Codex Alimentarius WHO/FAO (1970) Recommended international standard for olive oil, virgin and refined olive residue oil. CAS/RS-33.

Damania AB (1995) Olive, the plant of peace reigns throughout Mediterranean. *Diversity* 11(1–2): 131–132.

di Giovacchino L and Di Giovacchino L (1997) From olive harvesting to origin olive oil production. *OCL Oleagineux, Corps Gras, Lipides* 4(5): 359–361.

FAO (1998) *Production Year Book*, vol. 52. Rome: Food and Agriculture Organization. (www.fao.org)

Fernandez Diez MJ (1971) The Olive. In: Hulme AC (ed.) *Biochemistry of Fruits and Their Products*. New York: Academic Press.

Katsoyannos P (1992) *The Olive Pests and their Control in the Near East*. FAO Plant Production and Protection Paper no. 115, ix + 178 pp. Rome: FAO.

Kiritsakis A and Markakis P (1987) Olive oil: a review. *Advances in Food Research* 31: 453–482.

Mangold HK and Fedeli E (1997) Olives, olive oils and the Mediterranean diet. *Rivista Italiana delle Sostanze Grasse* 74(8): 349–352.

Santos Antunes AF (1980) Technology of olive oil production. *Boletim Instituto de Azeite e Productos Oleaginosis* 8(1): 23107.

ONIONS AND RELATED CROPS

S Khokar, University of Leeds, Leeds, UK
G R Fenwick, Institute of Food Research, Norwich, UK

Introduction

The genus, *Allium* (family Alliaceae), is large and geographically diverse, comprising over 500 species, most of which have little commercial importance. Alliums are distributed across warm temperate and temperate regions and into the subArctic belt. The major centers of genetic diversity are Afghanistan, Iran, and Pakistan with the Mediterranean basin being regarded as a secondary center. In addition to the onion (*A. cepa* L.), cultivated vegetable Alliums include garlic (*A. sativum* L.), leek (*A. ampeloprasum* L.), Japanese bunching onion (*A. fistulosum* L.), rakkyo (*A. chinense* G. Don), Chinese chives (*A. tuberosum* Rottl. ex Spr.), chives (*A. schoenoprasum* L.), and shallot (*A. cepa* L.).

The onion is by far the most important of the cultivated Alliums. Production is mainly in the Northern Hemisphere, with relatively little in unfavorable tropical regions (where shallots are grown). Cultivation experience, cultural attitudes, social beliefs, and culinary practices have resulted in specific Alliums being prized in particular regions; for example, leek and shallots in Western and Northern Europe; kurrat in Egypt and the Eastern Mediterranean; rakkyo, Japanese bunching onion; and garlic in China and Japan. The dramatic increase in global travel, interest in exotic and ethnic foods, and associated culinary practices, allied to improvements in agronomic practices and varietal- and postharvest characteristics have combined to encourage demand for fresh and processed vegetable Alliums. This has been intensified by prevailing enthusiasms for both 'healthy' eating and naturally occurring health remedies and medicines.

Onions and other cultivated Alliums have long been prized for their culinary properties, flavor, and taste and for their use as natural medicines. More

recently, their nutritional and health-enhancing properties have attracted much scientific investigation and consumer interest, thereby encouraging primary production and stimulating postharvest activities and opportunities. This brief article addresses the composition, flavor, postharvest handling, and processing of Alliums, emphasizing their significance for human health and well-being. The focus is mainly, but not exclusively, on onion and garlic.

History

The onion has been cultivated for over 4000 years; the earliest records are found in Ancient Egypt, including carvings dated to the third and fourth dynasties (2700 BC). In addition to being eaten, onions had significance in funeral rituals and embalming. During their Exodus (1500 BC), the Israelites recalled the foods of Egypt, including onion, leek, and garlic; mention was made of onions in Indian manuscripts dating from the sixth century BC. Locally grown onions were among the fermented vegetables given to the builders of the Great Wall of China (third century BC). Greek and Roman authors, notably Theophrastus and Columella, described the botany and cultivation of the onion, whereas others, such as Hippocrates and Dioscorides, focused on its medicinal properties. These culminated in the writings of Pliny, which addressed the cultivation, uses, and history of onions, garlic, and leek, and, together with the work of the military physician Galen, these had a great influence on the practice and writings of herbalists in medieval times and later. More extensive information has been published in the reviews of Fenwick and Hanley.

Production

Onion production in Asia accounts for more than 60% of the global production; together, India and China account for more than a third; in comparison, European production is less than 10%. Asian production of garlic exceeds 80% of the global volume, with China alone producing almost two-thirds. The world production of dry onions and garlic in 1999 is summarized in Table 1.

Table 1 Production of dry bulb onions and garlic ($\times 10^3$ tonnes)

Country	*Onion*	*Garlic*
World	41 527	9 042
Asia	25 511	7 585
Africa	2 780	178
North and Central America	3 370	301
South America	2 796	286
Europe	3 454	266
People's Republic of China	11 290	5 990
India	4 429	451
USA	2 995	224
Turkey	2 300	106
Iran	1 210	
Japan	1 240	
Republic of Korea	936	484
Brazil	921	63
Spain	985	170

From FAO Production Data (1999).

Composition

Onions and garlic are found in kitchens all over the world, being used as cooked vegetables, in salads, and as a flavoring in meat, poultry, and vegetarian dishes. Onions additionally possess excellent processing characteristics and may be canned, pickled, frozen, and fried. Dehydrated onion and garlic powders and flavor extracts also serve to enhance the value of these Alliums in the home and in the food industry. The proximate composition of selected Alliums is listed in Table 2. The cultivar, agronomic and environmental conditions during growth, and postharvest conditions all affect the composition. Protein, fat, and fiber contents are low in comparison with other vegetables, and starch is absent. Carbohydrates include sucrose, glucose, fructose, and fructose polymers. Green onion tops have also been found to be rich in β-carotene and to contain significant amounts of vitamin C. Alliums are rich in amino acids and γ-glutamyl peptides (including precursors of the characteristic pungent flavoring compounds), as well as biologically active secondary metabolites such as anthocyanins, flavonols, and phenolics.

Flavor

Alliums are rich in sulfur-containing compounds that are broken down by endogenous enzymes to yield a number of more volatile, less stable chemical compounds. These are considered to be responsible for much of the biological activity found in Allium extracts and oils, as well as determining the particular flavor characteristics of individual members of the family. (*See* **Flavor (Flavour) Compounds**: Production Methods.)

Flavor Biochemistry

Intact alliums have no pungency, since the volatile products are only released following the interaction of the enzyme, alliinase, with the *S*-alk(en)ylcysteine sulfoxide (alliin, I) which occurs when tissue is damaged or disrupted (Figure 1). The initial products of this enzymic hydrolysis are ammonia, pyruvate, and

Table 2 Composition of cultivated alliums (fresh-weight basis)

	Water (g)	*Energy (kcal)*	*Protein (g)*	*Fat (g)*	*Carbohydrates (g)*	*Sodium (mg)*	*Fiber (g)*	*Carotene (μg)*	*Sodium (mg)*	*Iodine (μg)*	*Folate (μg)*
Onion, fried	65.7	164	2.3	11.2	14.1	10.0	3.2	40	4	6	38
Onion, raw	89.0	36	1.2	0.2	7.9	5.6	1.5	10	3	3	17
Garlic	64.3	98	7.9	0.6	16.3	1.6	4.1[a]	0	4	3	5
Leeks	92.2	21	1.2	1.2	2.6	2.0	2.4	575	6		40
Chives	91.0	23	2.8	0.6	1.7	1.7	1.9[a]	2300	5	1.6	45
Rakkyo	86.0	50	2.2	0.3	12.0						

[a]Nonstarch polysacchrides.

From Bender AE and Bender DA (1999) *Food Tables & Labelling*. Oxford: Oxford University Press.

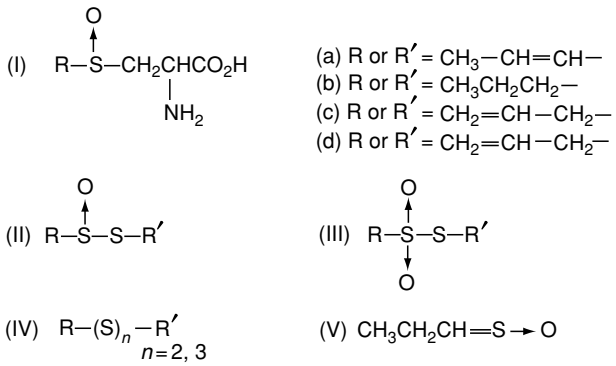

Figure 1 Sulfur-containing flavor volatiles from alliums.

an alk(en)ylthiosulphinate (allicin, II). The latter, which possesses odor characteristics typical of the freshly cut tissue, can undergo further nonenzymic reactions to yield a variety of compounds, including thiosulfinate [III] and di- and trisulfides [IV]. These compounds have slightly differing flavors and odors, and may impart a cooked note to steam-distilled onion or leek oils.

Not all alliums contain the same substituted cysteine sulfoxides; onion contains primarily the *S*-(1-propenyl), *S*-propyl, and, to a lesser extent, *S*-methyl derivatives (Ia–c, respectively). Thus, the typical flavor of onion is due to the presence of propyl- and 1-propenylthiosulphinates and corresponding di- and tri-sulfides. In contrast, garlic contains predominantly *S*-(2-propenyl)cysteine sulfoxide (*S*-allylcysteine sulfoxide, Id) and, and none of the 1-propenyl isomer. Its flavor is consequently due to the presence of 2-propenylthiosulphinate and related di- and trisulfides. Hive and leek contain relatively more *S*-methylcysteine sulfoxide, which yields methylthiolsulfinate (possessing a 'cabbagy' note), dimethyldisulfide, and related compounds.

Since it is possible for both symmetrical (II, $R = R'$) and mixed (II, $R \neq R'$) thiolsulfinates to be formed during the enzymic breakdown of alliins, and since these can yield all of the breakdown products mentioned above and which can subsequently enter into secondary chemical reactions, the composition of the flavor volatiles of freshly chopped onion, leek, and garlic can become very complex. Almost 100 compounds have been identified in freshly cut onion and garlic, and processing – especially that involving thermal treatment – causes additional chemical reactions to occur. These give rise to thiphenes, polysulfides and the products of amino acid–sugar (Maillard) reactions.

Lachrymatory Factor

Alone amongst the common alliums, the onion contains large amounts of *S*-(1-propenyl) cysteine sulfoxide (Ia). This compound is unique because the action of alliinase does not give rise to allicin, but rather to a mixture of unstable isomers of thiopropanal-*S*-oxide (V), which comprise the lachrymatory principles of the onion. The mode of formation of these compounds and their properties suggests that the problems of peeling onions may be reduced if carried out under running water, or if the onion is first chilled.

Estimation of Flavor Strength

The development of simple and reliable methods for the assessment and comparison of strength of flavor in onions and garlic, which may be required to identify cultivars with advantageous processing characteristics, has been the subject of much investigation. Since the enzymic cleavage of flavor precursors by alliinase generates a range of breakdown products, these have been used to measure pungency. The lachrymatory principle has been measured colorimetrically and spectrophotometrically, but these determinations are necessarily confined to onions. A robust method for pyruvate employs the color reaction with 2,4-dinitrophenylhydrazine; however, this cannot be employed for dehydrated products due to the presence of interfering by-products formed during dehydration and storage. Gas-chromatographic methods, with initial flavor concentration, have also been used to measure the headspace volatiles. The

amount and composition of the volatile fraction is largely dependent upon the conditions (degree of tissue disruption, temperature, time), so that all of these variables must be kept constant if reliable comparisons are to be conducted. Alternative approaches involve measurement of alliinase or γ-glutamyltranspeptidase activity. It is possible to monitor the flavor precursors, including γ-glutamyl peptides and alliins; however, such methods provide an indication of *potential* flavor, since the presence of the appropriate enzymes is necessary to release the volatile compounds.

Factors Determining Flavor

Amongst the many factors that have been found to have an effect on flavor are genotype, physiological age and condition, storage, and agronomic and environmental conditions. Thus, the range of flavor intensity in onions may vary 10-fold, with the most pungent being selected for processing. Sprouting of onions during storage leads to an increase in flavor as a result of a rise in transpeptidase levels and the conversion of γ-glutamyl peptides to alliins. Alliums grown under dry or arid conditions are smaller and visually less attractive than those grown in an abundance of water, but generally possess stronger flavors. The application of sulfur to the soil during growth significantly increases flavor, pungency and (in onions) lachrymatory character. The available sulfur is used preferentially for plant growth; only when these growing requirements are met is the secondary metabolic pathway, leading to alliin and allicin, enhanced.

Nutritional and Health Aspects

Garlic bulb and onions, two members of the Allium family, have been widely reported to possess a significant antioxidant activity which has been ascribed to allicin, a known scavenger of peroxyl radicals. Recent studies have indicated that, in addition to this scavenger, other compounds could also be involved. The onion contains a number of quercitin, isorhamnetin, and kaempferol conjugates and, indeed, is one of the major sources of flavonols in European diet. A relatively small range in content of quercitin conjugates for red (110–295 $mg\,kg^{-1}$) and yellow (119–286 $mg\,kg^{-1}$) cultivars have been reported. Flavonoids have been shown to offer protection against degenerative conditions such as cardiovascular disease, cancer, and aging.

Many alliums, including noncultivated wild types, have been extensively used for their therapeutic and medicinal properties. Most traditional medical practices include references to onion, garlic, and other alliums in the form of extracts, decoctions, concoctions, and poultices. Onions and garlic oils have proved effective in the treatment of a number of conditions, including intestinal worms, stomach ulcers, eye disorders, gastrointestinal disturbances, hypertension, and malarial fevers. Beneficial antitumor, hypoglycemic, hypolipemic, antiatherosclerotic, antiplatelet aggregation effects have also been reported. The active principles in many cases have been shown to be the sulfur-containing alliin derivatives.

A number of 'odor-free' garlic products, including capsules, extracts, and tablets, are commercially available. Since the biologically active principles also possess pungency, some care is needed to ensure that the treatment employed to remove odor has not at the same time reduced or removed the medicinally active compounds. (*See* **Garlic**.)

Postharvest Storage

Postharvest losses of onions may range from 10 to 30%, with even higher losses being reported from tropical regions. A number of factors have been shown to contribute to these figures; as a result, cultivars have been selected for improved storage characteristics, methods for sprout control developed, and storage practices and technologies optimized. Postharvest losses may be minimized if damage and bruising are reduced during harvest, storage, and packing; adequate ventilation of storage areas is vital if moisture accumulation on tissue surfaces, and subsequent disease proliferation, is to be avoided.

Dormancy, Sprouting, and Storage

Bulb dormancy is controlled by a complex interaction of growth inhibitors and promoters. Natural dormancy commences with harvest maturity and continues for 4–9 weeks, during which period, the bulbs will neither sprout nor continue to grow due to the influence of inhibitors translocated from the green leaves. However, these inhibitors are gradually destroyed with time. Once dormancy has passed, roots may emerge and leaf shoots appear; sprouting is optimal at 10–15 °C. Onions may be stored without sprouting if maintained at 0–5 °C and 65–70% relative humidity (RH). Sprouting has also been inhibited by controlled-atmosphere storage atmosphere storage at reduced oxygen levels (for example, 5–10% CO_2, 3% O_2, 5 °C), γ-irradiation or preharvest foliar treatment with maleic anhydride (MH). Onions treated with MH and held between −2 and 0 °C and 65–70% RH can be stored for 6–7 months without any deterioration of quality.

Garlic bulbs intended for consumption are held at ambient temperatures and can remain in good condition for several months. However, for extended storage of up to 8 months, bulbs should be stored at −2 to 0 °C and 60% RH. Sprouting most readily occurs between 5 and 10 °C, and inhibition may be achieved, as described for onions. Unlike onions and garlic, leeks are in active growth when harvested; they can have a postharvest life of up to 8 weeks at 0 °C and 95% RH. Controlled atmosphere treatment (for example, 10% CO_2 + 1% O_2) can extend the storage period further. Leeks are stored upright to avoid curvature of the pseudostems.

Storage Diseases

Postharvest diseases of onions, which will reduce the quality and, thereby, the value of the crop, are largely determined by the growing and storage conditions. The major disease of stored onions in temperate regions is neck rot, caused by *Botrytis allii*. This may be controlled by seed and set treatment with benomyl/thiram and effective postharvest drying. Other soilborne diseases (notably, white rot, pink rot, basal rot and smudge) are best controlled by combinations of field treatment and good husbandry.

Processing

Dehydration

Onions were included in the first dehydration experiments carried out in the late eighteenth century. A major program of onion and garlic dehydration was conducted in California in the 1920s, and dehydrated onion was already extensively used in food processing by the end of the next decade. It was the onset of World War II that provided the necessary impetus for the development of commercial food dehydration techniques. Onions for dehydration are generally white or yellow and have a high solids content (18–20% and above) and strong pungency. After preliminary cleaning, peeling, and slicing/chopping, onions are dehydrated at 75–60 °C, the temperature being reduced as the moisture content decreases. The final moisture content of 4% is achieved through warm air circulation. Discoloration may be minimized by effective curing and conditioning whilst the application of fluidized-bed techniques may overcome problems of scorching and agglomeration. In general, 8–10 kg of raw onions produce 1 kg of dehydrated product. Garlic may be dehydrated under similar conditions, with 5 kg of fresh bulbs producing 1 kg of dried product. Experiences of hot-air drying of chives have revealed problems such as an undesirable flavor, texture, and color. Flash-freezing and drying have proved effective in overcoming these problems. Dehydrated leek products are available for use in soups and other products, and dry green onions have been used to replace more expensive shallots and chives. Onion and garlic powders are used in the manufacture of onion and garlic salt. Both dehydrated onion and garlic are prone to discoloration, a problem that can be minimized by optimizing the drying conditions or employing approved additives. (*See* **Drying**: Theory of Air-drying; Physical and Structural Changes.)

Onion and Garlic Oils

Onion oil is obtained by the distillation of minced onions. The oil, obtained in a 0.002–0.03% yield, depending upon the source and processing conditions, has a flavor enhancement 500 times that of the dehydrated product. Garlic oil comprises 0.1–0.25% of the fresh weight of the clove, 1 g being equivalent to 200 g of dried garlic. The intense pungency of these products makes them difficult to use directly, and they are usually diluted in vegetable oil, or encapsulated.

Onion and Garlic Juices

Onion and garlic juices contain both flavor and aroma-active compounds, their precursors and sugars. They may be blended with volatile oils to restore a 'rounder' flavor profile. The juices, viscous and dark brown in color, can be mixed with a support (such as propylene glycol, lecithin or glucose) to provide an oleoresin with a flavor intensity 10 times greater than that of the dehydrated powder.

Others

Canned and bottled onions are used by the catering industry, especially in North America. Onion rings, coated in batter, are widely used in fast food outlets; these are supplied frozen, usually being extruded rather than machine-cut. Pickled onions, popular in the UK, are prepared from both 'brown' and pearl/cocktail onions (28–45 mm and 10–28 mm, respectively). Typically, brown onions are steeped in brine (10%) for 1–4 days, and lactic acid is added to control fermentation. They are then washed, covered in vinegar, and pasteurized; bisulfite may be added to maintain the color. Silverskin onions tend not to be pasteurized since this adversely affects the texture and flavor. Fermentation of Alliums has been described for thousands of years; onions, garlic, and leek are included in a variety of fermented products such as kimchi, ragi, sauces, and pickles. Many varieties of salamis and sausages are also strongly flavored with garlic before drying.

See also: **Drying**: Theory of Air-drying; Physical and Structural Changes; Chemical Changes

Further Reading

Bender AE and Bender DA (1999) *Food Tables & Labelling*. Oxford: Oxford University Press.

Fenwick GR and Hanley AB (1985) The genus *Allium* – parts 1,2 and 3. *CRC Critical Reviews in Food Science and Nutrition* 22: 199–271; 22: 273–376; 23: 1–73.

Rabinowitch HD and Brewster JL (eds) (1990) *Onions and Allied Crops*. Boca Raton, FL: CRC Press.

Rubatzky VE and Yamaguchi M (1997) *World Vegetables: Principles, Production and Nutritive Values*. Florence, KY: ITP Publishing.

Yamaguchi M (1983) *World Vegetables: Principles, Production and Nutritive Values*. Westport, CT: AVI Publishing.

Oranges *See* **Citrus Fruits**: Types on the Market; Composition and Characterization; Oranges; Processed and Derived Products of Oranges; Lemons; Grapefruits; Limes

ORGANICALLY FARMED FOOD

S Stolton, Elm Farm Research Centre, Berkshire, UK

Introduction

As our food system becomes ever more complex, foods become more highly processed, and the variety of foods available at any one time overcomes geographical and seasonal restraints, some consumers have been drawn towards alternative forms of production such as organically farmed food that are seen as more 'natural' and 'healthier' and perhaps of higher quality. The scientific evidence substantiating such claims remains limited, however, and is often the subject of intense debate. In practice, any analysis of the differences between organic and conventional food needs to encompass a wide range of different issues relating to health, environment, society, and the innate preferences of consumers.

What is Organic Farming?

Organic food is produced by farming methods that rely to a large extent on locally available resources and depend on maintaining ecological balances and optimizing the benefits from naturally occurring biological processes, rather than on manipulating the ecosystem through use of agrochemicals and fossil fuels. Organic agriculture aims to encompass agricultural systems that promote environmentally, socially, and economically sound production. By respecting the natural capacity of plants, animals, and the landscape, organic agriculture aims to maximize the quality of food whilst maintaining a sustainable agriculture and environment.

The principal aims of organic agriculture are summarized by the International Federation of Organic Agriculture Movements (IFOAM) in its *Basic Standards for Organic Agriculture and Food Processing*. These aims are:

- to produce food of high nutritional quality in sufficient quantity;
- to interact in a constructive and life-enhancing way with all natural systems and cycles;
- to encourage and enhance biological cycles within the farming system, involving microorganisms, soil flora and fauna, plants, and animals;
- to maintain and increase long-term fertility of soils;
- to use, as far as possible, renewable resources in locally organized agricultural systems;
- to work, as far as possible, within a closed system with regard to organic matter and nutrient elements;
- to work, as far as possible, with materials and substances which can be reused or recycled, either on the farm or elsewhere;
- to give all livestock life conditions which allow them to perform the basic aspects of their innate behavior;
- to minimize all forms of pollution that may result from agricultural practice;

- to maintain the genetic diversity of the agricultural system and its surroundings, including the protection of plant and wildlife habitats;
- to allow agricultural producers a life according to United Nation human rights, to cover their basic needs, and obtain an adequate return and satisfaction from their work, including a safe working environment;
- to consider the wider social and ecological impact of the farming system.

The IFOAM *Basic Standards* have provided a framework for almost all the national regulations and for the international World Health Organization/Food and Agriculture Organization Codex Alimentarius on organic agriculture. They are used by organic farmer organizations worldwide as a common platform. In many countries organic food and farming are regulated by legislation. In the European Union, for example, organic standards and certification procedures are enshrined in the European Community (EC) Regulation 2092/91, which provides a legal definition of organic production in Europe and for organic produce imported into Europe.

What Makes Organically Farmed Food Different?

There have been persistent calls from within the food industry for research that distinguishes clearly what, if anything, differs between organic and conventionally produced foods. Studies have been increasing in both number and scale, although results remain ambiguous. This is a highly political subject and there are regular well-publicised claims to 'prove' the issue one way or another. However, in reality many questions remain to be answered.

The majority of the scientific studies carried out to determine what, if any, special characteristics can be attributed to organically farmed food, as opposed to food produced under other production regimes, have been comparative, using one or two crops to determine general trends.

A comprehensive literature review of over 150 comparative studies of this sort was published in 1997. The review highlighted a number of inherent problems with comparative studies to date, including that there was no clear and consistent definition of 'conventional' agriculture (although primarily this term denotes a system where farming methods attempt to substitute natural processes – for instance, by using inorganic fertilizers or pesticides) and that sample sizes tended to be small – both of which make trend analysis difficult. Most studies reviewed are physicochemical evaluations of concentrations of desirable or undesirable elements in food. Other analytical methods used in comparative studies include electrochemical, plant physiological, and multivariable physicochemical methods, susceptibility to infection, and postharvest behavior. More 'alternative/holistic' methods to determine the quality of organic food have also been developed but these are not generally recognized by conventional scientists. Such methods include the ascending-imaging method, round filter chromatography, copper chloride crystallization, and the measurement of ultraweak photon emissions. The studies carried out using these methods claim to have shown a trend towards organic foods having more 'vital activity' than other food from other production systems.

Other researchers argue that focusing solely on physicochemical evaluations misses many of the wider issues relating to organic food and also misses the reasons why most consumers buy organic food. Just as the organic farming system cannot be fully understood by studying one crop in the rotation, the qualities of organic food cannot be determined by one or two comparative studies of individual food products or the effects of an organic diet on laboratory-reared animals. A more valid approach to understanding the characteristics that may make organic food different is to look at a wider range of attributes that together could be said to describe 'food quality.'

A new framework, or definition, of food quality, combining six distinct criteria was drawn up by a group of international experts involved in the food industry, at a colloquium organized by the UK-based organic research organization, Elm Farm Research Centre, in 1989. Like all such concepts, this can never hope to cover all aspects of quality. However, the definition does incorporate the many issues that have been the focus for research into the quality of organic food and is thus a useful tool for reviewing research findings. In many areas quality characteristics can be measured quantitatively and comparisons between organic and conventional practices can be made; however, our choice of food is often led more by the subjective non-quantifiable values, which can make the judging of quality more difficult, but no less valid.

The Six Aspects of Food Quality

The six criteria of food quality (authentic, functional, biological, nutritional, sensual, and ethical) identify the range of issues that consumers and producers recognize as important with respect to food production and composition. They thus provide a framework within which a more comprehensive analysis of what makes organic food different can be carried out. Each criterion is briefly examined below.

Authentic

> Food that is authentic, traditional or natural and has not been adulterated in production, processing or storage, including the use of genetically modified organisms or their derivatives.

Consistently, research into consumers' attitudes shows that the main perceptions of organic food are linked to phrases such as 'no chemicals/additives/pesticides,' 'natural' and 'healthy.' Although there is an increasing range of processed organic foods being produced, these processes are themselves bound by standards for organic production. According to the IFOAM *Basic Standards*, the general principles of organic food processing and handling are that: 'any handling and processing of organic products should be optimized to maintain the quality and integrity of the product.' Furthermore, the processing methods should ensure that: 'the vital quality of an organic ingredient should be maintained throughout each step of its processing.'

Measuring authenticity is almost impossible because it relates to such a large extent to individual consumer preferences. However, issues like the avoidance of most agrochemicals, the fact that organic food aspires to be locally produced, and the generally reduced processing appeals to a substantial part of the food market. The fact that organic food is also free of genetically modified (GM) food also meshes well with consumer preferences in Europe and beyond. More generally, the less processed nature of most organic food helps fulfill the needs of those consumers who are consciously trying to move away from the highly processed foods that became a dominant part of the industry in the 1980s and 1990s.

Functional

> How appropriate food is to its specific purpose – i.e., food that produces, stores, or cooks well. 'Functionality' is also used by food scientists and technologists to describe ingredients that blend well, or which have a good shelf-life.

Concern about the world's diminishing biological diversity (or biodiversity) has been growing in recent years. Although biodiversity loss tends to be linked in people's minds with threats to rainforests or to rare animals like giant pandas, genetic diversity is also being lost within food production. This loss inevitably affects our choice of the variety of food and the functionality of crops produced.

Loss of variety means loss of potential to adapt to different conditions. A distinction between organic and conventional agriculture is that in the former, encouragement of greater variety is an important part of the overall system, because reduced reliance on agrochemicals to change conditions in the soil means that the plants must themselves be better adapted to local conditions. Organic standards encourage the expansion of varieties grown to gain benefit from those suited to particular growing conditions. For example, conventional wheat grown in an organic system will show a distinct reduction in protein content, whereas selection of a variety suitable for the particular conditions on the farm can allow baking-quality wheat to be grown.

This encouragement of diverse species use is threatened by the general reduction in crop diversity. A survey of some 75 crop species, carried out by the Rural Advancement Fund International, found that about 97% of the varieties given on old US Department of Agriculture lists are now extinct (assuming that varieties that are not stored in US seed banks are probably extinct). This loss is not unique to the USA, for example, indigenous wheat, rice, and sorghum varieties have virtually disappeared from many centers of diversity in areas of Europe, the Middle East, Africa, and Asia. Of the 2500 or so apple varieties that have been grown in the UK, just nine now dominate the retail outlets. Within the EC, this erosion of diversity in crops has become legally obligatory with the introduction of a common catalogue. Varieties not listed are seen as inferior and cannot be legally sold by seed companies.

Organic growing conditions can also contribute to the functionality of food. A review of research on storage degradation by the Soil Association shows that organic crops can have better storage quality and postharvest behavior than nonorganically grown produce. Organic plant production is characterized by a relatively low level of nutrient supply compared to conventional crops. This leads to earlier completion of vegetative growth and an early onset of maturity, which may contribute to a longer shelf-life. Respiration rates and enzyme activity, which can also result in lower storage losses, also tend to be lower in organically farmed foods.

Biological

> How food interacts with the body's functioning. This includes both negative and positive interactions, such as the detrimental effects of some food additives on allergies and beneficial effects of foods in stimulating immunity to disease. The concept also refers to the pharmacological qualities of foods and herbs.

Many consumers feel that organic food can have positive effects on health and this is a major reason for purchase. A comprehensive review of over 400 published papers by the Soil Association of existing research on the differences between organically and

nonorganically grown foods has revealed significant differences to key areas of food quality important to the promotion of good health – food safety, nutritional content, and the observed health effects of organic food. The review notes that, while there has been little research carried out on humans, a number of feeding trials on animals have shown significant improvements in the growth, reproductive health, and recovery from illness of animals fed organically produced food.

Reproductive health is seen as a particularly sensitive indicator of environmental conditions and has been the focus of some research into the effects of organic food on health. It is known that occupational exposure to pesticides can impair reproductive health and it is hypothesized that long-term low-dose exposure to residues through food intake may also have adverse effects, so food with less pesticide residues might be expected to have beneficial results. In the mid-1990s, two small studies in Denmark indicated that semen quality was improved among consumers of organic food and this gained worldwide publicity – however, a larger study could not confirm these results. Several feeding experiments have also been carried out, but over a range of studies of mice, rats, pigeons, rabbits, and hens the reproductive effects of diet have shown no clear trend. In feed experiments with mice and rats the share of stillborn animals and animals that died shortly after birth was significantly higher in litters fed on conventionally produced feed.

The routine use of antibiotics in intensive farming has been linked to the development of 'superbugs' resistant to antibiotics used to combat human illness. As organic standards only allow the use of antibiotics in animals to cure specific problems, it is reasonable to suggest that an organic diet will reduce the build-up of antibiotic residues.

Nutritional

How food contributes to a balanced diet. This recognizes individual food values, such as vitamins, protein, and trace elements, and undesirable elements such as fats and sugar and nitrate and sodium, which in large amounts can be harmful.

Public concern about the presence of pesticide residues, additives, and nitrates in food is the reason many consumers purchase organically produced food. The organic sector has, however, been careful to stress to consumers that, given the state of the environment, and the presence of pesticide residues in soils, air pollution from spray drift, and industrial pollution, no food production method can guarantee a 100% chemical-free crop. There is however evidence that organic foods do have lower levels of pesticide residues than comparable conventional crops, although these results should not be overestimated given the small number of studies.

Concern has also been expressed over increasing nitrate intake from dietary sources, notably from the consumption of vegetables, although the possible health impacts of this remain the subject of debate. There is considerable variation in the levels of nitrate in foods, and in the consumption patterns of different foodstuffs amongst people; for example, the nitrate content of leafy vegetables is generally higher in winter. Comparative studies on farming systems and the related accumulation of nitrate in vegetables have been the subject of several research projects in recent years. In the UK, trials showed significantly lower levels of nitrate accumulation when using composted farmyard manure compared to the use of the soluble compound fertilizer – although fertilization with either type of material did lead to an increase in the level of nitrate in the plants. This trend is confirmed over a range of research studies (the Soil Association reviewed 16 studies, of which 14 showed a trend toward significantly lower nitrate contents in organically grown food), with nitrate content of conventional produce as opposed to organic production being particularly high in nitrophilic leaf, root, and tuber vegetables.

The presence or absence of harmful or potentially harmful substances in food only represents one aspect of the nutrition equation, as our food is made up of a range of nutrients such as vitamins and minerals. The Soil Association reviewed 99 papers that compared the nutritional quality of organically and conventionally grown crops. These studies were assessed for validity in terms of agricultural practice and scientific analysis. Of those studies deemed valid, the Soil Association concluded that vitamin C and dry-matter contents are higher, on average, in organically grown crops. Mineral contents were also higher, on average, in organically grown crops, although the small number and heterogeneous nature of the studies included mean that more research is needed to confirm this finding. Research also indicates a clear long-term decline in the trace mineral content of fruit and vegetables, but to understand the reason for this the influence of farming practices requires further investigation.

As well as having effects on individual elements of crops, the choice of an organic diet may also have an effect on overall dietary habits. A survey of the diet of UK organic consumers concluded that those committed to the consumption of organic food have a different diet from the average consumer – a diet that is more in line with the productive capacity of the organic system. By collecting detailed diet diaries and comparing them with data from the Office of

Population Censuses and Surveys, organic consumers showed a significant swing towards the consumption of fruit and vegetables, with a reduction in the protein food 'centerpieces.' Protein consumption was less dependent on meat, being derived from other sources. Staples showed a greater diversity, with a greater intake of pasta, grains, and cereals, with less concentration on potatoes. The survey data also showed an increased consumption of fibrous foods and a decreased consumption of animal-based as well as sweet and fatty foods.

Sensual

Food which appeals to the senses; the industry uses the term 'organoleptic' to refer to sensual quality of food, i.e. its look (esthetic appeal), taste, feel (as in mouth feel), smell, and sound.

The determination of whether organic food tastes better than food produced from other forms of production has been the subject of considerable research. Although it is widely believed by consumers that organically farmed food tastes better than the alternatives, there is no conclusive scientific evidence to prove this. Partly this is due to individual preferences in taste, which makes the determination of quality through taste alone totally subjective, and is complicated by a whole range of regional factors, such as soils and climate, which make direct comparisons between growing systems difficult (unless trial plots are extremely well matched). The Soil Association also suggests that other influences on taste could be the higher water content in conventionally grown food diluting its flavor, different choices of cultivars, and the fact that organic produce has sometimes been shown to have higher natural sugar content, which may be perceived as a better or worse taste, depending on the product.

Review of feeding experiments with animals has however drawn one unexpected, and so far unexplained, conclusion. When given the choice, animals almost always prefer organic produce to conventionally produced feed.

Ethical

This concept has four related but distinct meanings: environmental (the effects on our environment of food production methods); social (the conditions in which food producers are treated – on the farm, in the factory, on the shop floor); ethical (the morality of the way specific foods are produced – for example, the conditions of animals raised for meat); and political (the effects of the food industry on specific countries).

Research into consumer preferences suggests that there has been a shift in emphasis amongst purchasers of organic food in the last decade, away from people primarily driven by personal health concerns to those with a wider interest in the environmental and social impacts of the food that they buy.

There is a large literature describing the environmental problems of conventional agriculture and the benefits of organic agriculture. For example, long-term research projects have accumulated substantial evidence that organic systems are beneficial to biodiversity. An overview of research findings from 23 of these projects by the Soil Association concluded that in most studies there were important differences between the biodiversity on organic and conventional farms, with generally greater levels of both abundance and diversity of species on organic farms. On average, the research found five times as many wild plants in arable fields and 57% more species, 25% more birds at the field edge, and 44% more species in fields in the autumn and winter, and three times as many nonpest butterflies in crop areas.

In a review of the environmental impacts of organic farming in 18 European countries, an assessment of a series of environmental indicators found that organic farming has more floral and faunal diversity than conventional farms, tends to conserve soil fertility and system stability better and on a per-hectare scale has between 40 and 60% lower CO_2 emissions and is likely to have lower emissions of N_2O, CH_4, and NH_3.

Sociological data comparing organic and conventional farming systems are not so well advanced. The IFOAM began a project on data collection and farm system comparison in 1997 and has developed a methodology which collects data on 13 indicators that provide information on subjects such as production of main and byproducts, recycling, male and female hired and own labor, financial results, energy consumption and production, water use, family expenditure, and education. The methodology is now being tested in Guatemala, India, and Nicaragua. Results to date have shown that self-reliance, which is defined as a percentage of output value that is consumed on the farm, is higher in the organic farms.

Conclusions

Many of the issues described above are not limited to organic production: for example, fresh food is available at the farm gate from many conventional farms and there is a growing fair-trade movement that insures equitable returns for food produced in developing countries. However, organic farming is the only system where all such issues are addressed together and within a legal framework through the organic standards, insuring that consumers can be reasonably

confident that a 'package' of health, environmental, and social issues has been addressed in the production of the organic food that they purchase. When provided with the whole picture, across a range of issues such as those given above, there does seem to be sufficient research to conclude that organically farmed food does have some unique characteristics.

See also: **Antibiotics and Drugs**: Uses in Food Production; **Food and Agriculture Organization of the United Nations**; **Legislation**: Codex; **Quality Assurance and Quality Control**; **World Health Organization**

Further Reading

Alföldi T, Lockeretz W and Niggli U (2000) *Proceedings of the 13th International IFOAM Scientific Conference*. Germany: IFOAM.

Azzeez G (2000) *The Biodiversity Benefits of Organic Farming*. Bristol, UK: Soil Association.

Finesilver T (1989) *Comparison of Food Quality of Organically versus Conventionally Grown Plant Foods*. Ecological Agriculture Projects. Quebec, Canada: McGill University.

Gold M (2000) *Organically Produced Foods: Nutritive Content. Special Reference Briefs Series no. SRB 2000–03*. Beltsville, USA: Alternative Farming Systems Information Center, National Agricultural Library, US Department of Agriculture.

Heaton S (2001) *Organic Farming, Food Quality and Human Health: A Review of the Evidence*. Bristol, UK: Soil Association.

Lampkin N (1990) *Organic Farming*. Ipswich, UK: Farming Press.

Pretty JN, Brett C, Gee D *et al*. (2000) An assessment of the total external costs of UK agriculture. *Agricultural Systems* 65(2): 113–136.

Stolze M, Piorr A, Häring A and Dabbert S (2000) *The Environmental Impacts of Organic Farming in Europe. Organic Farming in Europe: Economics and Policy*, vol. 6. Hohenheim, Germany: University of Hohenheim.

Woese K, Lange D, Boess C and Bögl K-W (1997) A comparison of organically and conventionally grown foods – results of a review of the relevant literature. *Journal of the Science of Food and Agriculture* 74: 281–293.

Woodward L, Stolton S and Dudley N (1992) *Food Quality: Concepts and Methodology*. Hamstead Marshall, Berkshire, UK: Elm Farm Research Centre.

Organoleptic Evaluation *See* **Sensory Evaluation**: Sensory Characteristics of Human Foods; Food Acceptability and Sensory Evaluation; Practical Considerations; Sensory Difference Testing; Sensory Rating and Scoring Methods; Descriptive Analysis; Appearance; Texture; Aroma; Taste

Osmosis *See* **Membrane Techniques**: Principles of Reverse Osmosis; Applications of Reverse Osmosis; Principles of Ultrafiltration; Applications of Ultrafiltration

Osteomalacia *See* **Rickets and Osteomalacia**

OSTEOPOROSIS

J J B Anderson, University of North Carolina, Chapel Hill, NC, USA

Background

Osteoporosis is considered a chronic multifactorial disease because many variables contribute to the development of low bone mass and low bone mineral density accompanied by an increased risk of fracture. This disease usually occurs in late adulthood, following the menopause in women, and a decade or so later in men, but it may appear rarely earlier in life depending on the status of reproductive steroid hormones.

Both hereditary and environmental factors contribute to the multifactorial nature of this disease. Individual differences in constitutional factors as well as lifestyle/environmental determinants influence bone. In terms of lifestyle, both regular physical activities and a healthy diet remain two of the more important contributors to bone health and the maintenance of its functions into late life. The major function, sound ambulation, needs to be preserved as long as possible through exercise and diet. When calcium intake is not adequate, the adaptational role of the calcium regulatory system operates in an attempt to preserve skeletal structure for its functional uses. If the adaptation is not sufficient, bone deterioration, i.e., loss of both mass and density, will lead to fragility fractures. These and related topics are covered in this paper. (Although drugs that help conserve bone are widely used today, this review focuses on dietary risk factors and nutritional therapies.)

Bone Structure

Bone tissue consists of two types within the same specific bone, e.g., a vertebra of the spinal column: trabecular (cancellous) and cortical (compact). Trabecular tissue is the more metabolically active because it has about eight to 10 times more total surface area than cortical tissue, and these surfaces are all largely covered by bone cells that are responsible for new bone formation and old bone resorption (degradation). In the body of a vertebra most of the bone tissue is trabecular, but near the surfaces of the entire vertebra, cortical bone predominates. Therefore, each specific bone or organ contains both types of bone tissue but typically in different locations within the bone. For example, long bones, such as the femur, contain much more trabecular tissue at either end near the hip joint or knee joint and a much greater proportion of cortical bone in the shaft that connects the two ends. This distinction is important because most of the fractures of the bones (organs) occur where more metabolically active trabecular bone tissue exists.

Bone Gain and Bone Loss through the Life Cycle

In early life, during the growth periods of the skeleton, bone acquisition through bone formation dominates. This phase represents the making of the skeletal model, i.e., modeling, and it typically ends by 16–18 years in females and 18–22 years in males. Modeling is characterized by greater formation than resorption and results in a net gain of bone mass or bone mineral content (BMC) by the end of the growth phase. Modeling is completed at the end of linear growth. In remodeling of the skeleton, resorption of bone equals formation; so, the net amount of bone mass remains fairly constant, though some modest gain may still occur until the mid-20s or beyond. During the beginning of the remodeling, the increase in pubertal hormones is responsible for the cessation of growth in length, or height increase. At this time, bone mineral density (BMD) attains a maximum value that becomes the 'healthy norm' for an individual for the rest of their life. (In the definitions of osteopenia and osteoporosis, the healthy means of a population of 20–29-year-old males or females are taken as the standard values for determining osteoporosis late in life (see below.)

In late life, i.e., after the menopause in women and a decade or so later in men, imbalances in the remodeling of the skeleton result in bone losses, so that reductions in both BMC and BMD occur. The loss of estrogens after the menopause and probably the later decline of androgens by men contribute to the increase in resorption, the reduction of formation, or both. The increase in resorption is triggered by increased activity of osteoclasts, and the decline in formation is directly related to decreased activity of osteoblasts. An increase in bone turnover, i.e., increased rates of resorption and formation with resorption dominating formation, greatly accelerates the rate of bone loss. Most individuals are slow or moderate losers of bone, whereas only a small fraction are 'fast losers.' Bone turnover is assessed by measurement of chemical markers of degraded matrix proteins, such as collagen, resulting from

bone resorption and of hormones, especially parathyroid hormone, involved in calcium homeostasis.

The changes in the skeleton over the life cycle reflect early-life gains and late-life losses. When the losses become sufficient to deteriorate to the state of osteoporosis, an individual becomes at risk of fragility fractures of bones such as the vertebrae and proximal femurs (hips). The hip fractures are the most severe and debilitating, and many individuals never recover.

Measurement of BMC and BMD

BMC and BMD are measured today using a technique called dual energy X-radiographic analysis (DXA). Osteopenia, or too little bone, is followed by osteoporosis, or even greater bone loss, which places an individual at great risk of a fracture. Quantitative definitions of osteopenia and osteoporosis have been established by the World Health Organization as follows: osteopenia is between minus 1 and 2.5 standard deviations (SDs) below the 20–29-year-old mean values for women or men; and osteoporosis is greater than −2.5 SDs below the 20–29-year-old means. (These values of individuals at any adult age compared with means of healthy young adults are known as *T*-scores.) According to this definition, a postmenopausal 60-year-old woman with −3.0 SDs below the mean at different skeletal measurement sites does not necessarily have any clinically diagnosed fracture, but she is likely to have one over the next few years without any drug (or drug and diet) therapy.

In addition to the DXA determination of bone status, biochemical markers may also be used to assess the severity of bone turnover and the potential risk of fracture. Therefore, both clinical observations, e.g., loss of height and pain in the lower back, and biochemical marker data may increase the diagnostic significance of DXA measurements that classify an individual in the osteoporotic range or >2.5 SDs below the healthy means.

Types of Osteoporosis

Type I osteoporosis uses the postmenopausal woman as the prototype, although men also rarely may suffer from the abrupt loss of sex steroids that impact greatly on the retention of bone tissue. Type II osteoporosis is age-related and typically occurs in both genders in the later decades of life; the causation of Type II is poorly understood, but it accelerates when the musculoskeletal system functions decline.

Although, for simplicity, only two types of osteoporosis are defined here, other variants clearly exist. One that is of increasing importance is corticosteroid-induced osteoporosis because of the widespread use of prednisone and related drugs in therapy of diseases other than osteoporosis. This type of osteoporosis is called secondary. Other drugs may also induce secondary osteoporosis.

Nondietary Risk Factors for Osteoporosis

Several nondietary risk factors – all potentially adverse or harmful – have been identified that promote the loss of bone and the onset of fractures. Seven environmental/lifestyle factors are listed here:

- thinness with low LBM;
- cigarette smoking;
- excessive alcohol consumption;
- insufficient physical activity;
- drugs – over-the-counter and prescription;
- decline of sensory perceptions;
- falls.

Each of these factors has a risk associated with it, but when two or three exist together at the same time, the risk of an osteoporotic fracture may increase exponentially rather than additively. For example, the small-framed older postmenopausal woman who smokes a pack of cigarettes a day, drinks two or three servings of alcohol a day, and who has little physical activity in her daily life will typically be at great risk of an early hip fracture, i.e., by age 70 years or younger. When the old-old, i.e., greater than 80, suffer declines of acuity of their senses, such as vision and equilibrium, or take medications that result in the same effects, they are much more likely to fall and break their hip.

The most important of these adverse factors, in general, may be the decline in regular physical activity that results in the loss of lean body mass (LBM), characterized by declines in muscle strength and tone, because bone loss follows closely the loss of LBM.

Many elderly take several drugs, i.e., polypharmacy, and some of them become confused or have difficulty maintaining equilibrium. Each of these factors may contribute to falls and consequent hip fractures.

Dietary Risk Factors

In addition to adverse environmental/lifestyle factors, numerous dietary factors may also have adverse effects on skeletal tissue. These deleterious factors are thought to operate throughout the life cycle, not just during late life. The major variables are given here, but this list is not exhaustive:

- low calcium;
- high phosphorus;
- low vitamin D;
- high-animal protein and acid load;
- high-sodium snacks;
- vegetarian diet;
- poor diet in general.

The most common dietary problem associated with the development of osteoporosis has been an inadequate consumption of calcium, at least among Caucasians in Western nations, including pediatric populations. Low calcium intakes in these nations are typically accompanied by high phosphate consumption because of both the ingestion of naturally occurring phosphorus in foods, especially animal proteins, and phosphorus from food fortification. Phosphorus fortification is fairly common in Western nations because of the widespread use of processed foods – foods modified with the many applications that utilize phosphate salts of one type or another. Cola-type soft drink beverages also fit in this category, although, technically, they are not foods. The net result of consuming phosphate-rich foods may be a low calcium (Ca) to high phosphate (P) intake that is worsened when individuals consume little milk or cheese in their usual diets.

If the Ca:P ratio declines to approximately 1 to 4, the serum parathyroid hormone (PTH) concentration becomes elevated (but remains within the normal range). A constantly elevated PTH leads to bone loss and a gradual decline in BMD that eventually becomes osteoporosis. How many years bone takes to get into the osteoporotic range, according to WHO definitions, is not clear, but a long-term dietary pattern of low calcium–high phosphate may even contribute to low bone mass in females before they reach 20 years of age, if fractures among girls and pubertal females are a valid index of low bone mass.

Coupled with limited skin exposure to vitamin D-promoting sunlight, a low intake of vitamin D from foods, especially fortified dairy products in the USA and various deep-sea fish species in much of the world, is now considered a risk factor for low bone mass. The mechanism for this condition is not entirely clear, but a low circulating concentration of 25-hydroxyvitamin D is linked with both a decline in intestinal calcium absorption and an increase in bone turnover with a resultant loss of bone.

A usual high-animal protein intake has been reported by some investigators to contribute to bone loss and an increase in risk for osteoporosis. The mechanism is considered to reside in the increased production of acids, i.e., phosphoric and sulfuric, from the degradation of phosphorus- and sulfur-containing amino acids that are considerably greater in animal than plant proteins, such as those derived from soy. The net effect is an increased loss of calcium ions in urine. This loss has been shown acutely and in short-term experiments, but it has not been established in long-term studies.

High-sodium snack foods have become very popular in Western nations, and they contribute additional sodium to an already high intake that derives from so many foods processed with sodium or salt. Calcium renal losses increase on high-sodium intakes because the kidneys favor sodium reabsorption at the expense of calcium ions. The net loss of calcium comes from bone, and therefore a loss of bone mass may be associated with excessive consumption of sodium-rich snack foods.

Vegetarian diets may also compromise bone health through a number of possible mechanisms, but the low calcium and vitamin D intakes from a vegan dietary pattern may be largely responsible for lower bone mass among vegetarians.

Finally, a poor diet, especially one based on a limited intake of fruits and vegetables, may be deficient in other bone-essential nutrients, such as magnesium, vitamin K, zinc, antioxidant nutrients, and probably a dozen others. Any one or a combination of these deficits may inhibit efficient bone formation. Adverse effects include loss of protection against oxidants, and poor regulation of acid–base balance, thus impacting bone as a major buffering store. Whatever the mechanisms may be, bone health requires a wide variety of nutrients that are best supplied by a varied diet consisting of the recommended numbers of servings of foods each day.

Adaptation to Low Calcium Intakes

Several adaptations to dietary intakes occur after each meal in healthy individuals; a number of these adaptations directly affect calcium homeostasis by impacting on serum PTH and 1,25-dihydroxyvitamin D concentrations. For example, a low calcium intake, especially when coupled with a high phosphate consumption pattern, stimulates PTH secretion via a calcium sensor on the membranes of parathyroid gland cells. Elevated PTH, in turn, stimulates the removal of calcium from bone (see above).

In addition, a low calcium intake stimulates the vitamin D regulatory mechanism by increasing the renal production of the hormonal form of vitamin D, i.e., 1,25-dihydroxyvitamin D, which leads to an increase in intestinal calcium absorption and bone utilization of calcium. The potential problem in older women, and perhaps men, is that the intestinal

adaptation declines, and less calcium can be absorbed with the hormonal stimulus.

Finally, excessive, continuous PTH secretion that may develop late in life with a low Ca:high P ratio diet and low circulating 25-hydroxyvitamin D concentration contributes to bone loss almost across the 24 h of a day. It is now recognized that a persistent and continuous treatment with PTH causes significant bone loss, whereas a discontinuous treatment with PTH, i.e., once daily or once weekly, increases bone mass and density by stimulating osteoblasts to make new bone tissue. Thus, any type of low calcium intake that generates a continuous secretion and elevation of PTH in blood will have serious negative consequences on the maintenance of bone, especially of trabecular bone tissue.

Dietary Prevention and Treatment – Foods and Supplements

The principles of both primary and secondary prevention of osteoporosis are similar: increase calcium in the diet through foods and supplements to 1000 mg or more per day in order to suppress PTH secretion (see above), and assure that individuals obtain sufficient amounts of vitamin D through both foods and supplements: a dosage of 800 IU or more is recommended. Sunshine exposure is also encouraged, depending on geographic latitude and time of the year. Finally, a healthy diet containing virtually all nutrients from foods makes sense from a nutritional perspective. For the elderly, a daily supplement that contains a wide range of nutrients at recommended intake levels is both safe and inexpensive. The current dietary reference intakes should be used to guide consumers in maintaining appropriate amounts of nutrients each day (1997–2001).

Public Health Implications and Conclusions

Since the incidence and prevalence of osteoporosis are increasing as our populations are aging, greater expenditures are anticipated for direct care of hip fracture patients and then their subsequent rehabilitation. Many victims will not survive the first post-fracture year. Preventive strategies, either primary or secondary, that are cheap and effective, must be identified and implemented. Improving the diet may be one of the most easily modifiable approaches for the prevention of osteoporosis, but of course, if individuals cannot ambulate, the dietary changes will have little impact with regard to osteoporosis. Walking and maintaining activities of daily living by the elderly are critical for the prevention of osteoporosis. Any type of minimal exercise program should also yield some benefit to the retention of musculoskeletal function.

See also: **Bone**; **Calcium**: Properties and Determination; Physiology; **Cholecalciferol**: Properties and Determination; Physiology; **Hormones**: Thyroid Hormones; **Smoking, Diet, and Health**; **Sodium**: Properties and Determination; Physiology

Further Reading

Anderson JJB (1999) Plant-based diets and bone health: Nutritional implications. *American Journal of Clinical Nutrition* 70: 539S–542S.

Anderson JJB, Sell ML, Garner SC and Calvo MS (2001) Phosphorus. In: Russell R and Bowman B (eds) *Present Knowledge in Nutrition*, 8th edn. Washington, DC: ILSI Press.

Committee on Nutrition (1999) Calcium requirements of infants, children, and adolescents. *Pediatrics* 104: 1153–1157.

Dietary Reference Intakes (DRIs) (1997–2001) *Several Reports Including: Institute of Medicine (IOM), Food and Nutrition Board. Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride*. Washington, DC: National Academy Press.

Goulding A, Cannan R, Williams SM *et al*. (1998) Bone mineral density in girls with forearm fractures. *Journal of Bone and Mineral Research* 13: 143–148.

New SA, Robins SP, Campbell MK *et al*. (2000) Dietary influences on bone mass and bone metabolism: Further evidence of a positive link between fruit and vegetable consumption and bone health? *American Journal of Clinical Nutrition* 71: 142–151.

Wyshak G (2000) Teenaged girls, carbonated beverage consumption, and bone fractures. *Archives of Pediatrics and Adolescent Medicine* 154: 610–613.

OXALATES

S C Morrison, New Zealand Institute for Crop and Food Research, Lincoln, Canterbury, New Zealand
G P Savage, Lincoln University, Canterbury, New Zealand

Introduction

Oxalic acid and its salts occur as end products of metabolism in a number of plant tissues. When these plants are eaten they may have an adverse effect on mineral bioavailability because oxalates bind calcium and other minerals. While oxalic acid is a normal end product of mammalian metabolism, the consumption of additional oxalic acid may cause stone formation in the urinary tract when it is excreted in the urine. The mean daily intake of oxalate in the diet in the UK has been calculated to be 70–150 mg; tea, rhubarb, spinach, and beet are common high oxalate-containing foods. Soaking and cooking foodstuffs high in oxalate will reduce the oxalate content by leaching.

The consumption of high-oxalate foods is more likely to pose health problems in those who have an unbalanced diet or those with intestinal malfunction. A diet high in oxalate and low in essential minerals, such as calcium and iron, is not recommended. Vegans and lactose-intolerant persons may have a high-oxalate and low-calcium diet unless their diet is supplemented. Vegetarians who consume greater amounts of vegetables will have a higher intake of oxalates, which may reduce calcium availability. This may be an increased risk factor for women, who require greater amounts of calcium in the diet. Persons with an increased absorption rate of oxalate are advised to avoid or eat fewer high-oxalate foods to prevent kidney stone formation. In healthy individuals, the occasional consumption of high-oxalate foods as part of a balanced diet does not pose any particular problem.

Oxalates in Plants

Oxalate can be found in small amounts in many plants. Oxalate-rich foods are usually minor components in human diets but are present in higher quantities in seasonal diets in certain areas of the world (particularly in the tropics) as a component of grains, tubers, nuts, vegetables, and fruits. The highest levels of oxalates are found in the foods listed in Table 1.

In general, oxalate content is highest in the leaves, then in the seeds, and lowest in the stems. Reports show that the stems or stalks of plants, such as amaranth, rhubarb, spinach, and beet, contain significantly lower levels of oxalates than the leaves. In the buckwheat family (e.g., rhubarb, sorrel), there is almost twice as much oxalic acid in the leaves as in the stalk. However, in the goosefoot family (e.g., beet, spinach), oxalic acid is more abundant in the stalk than in the petiole of the leaf. It must be noted that the leaves of rhubarb are rarely eaten and therefore the oxalate content of its leaves is of no concern in human nutrition.

Oxalic acid concentration tends to be higher in plants than in meats, which may be considered oxalate-free when planning low-oxalate diets. Meats, fats, and dairy products contain very low levels of oxalates.

The levels of oxalates in fungi are low when compared to the levels found in spinach and rhubarb. The giant mushroom (*Tricholoma giganteum*), a large, edible fungi, is reported to contain 89 mg $100\,g^{-1}$ dry weight (DW) oxalic acid while tropical species of mushrooms, including termite and ear mushrooms, contain between 80 and 220 mg oxalate $100\,g^{-1}$ DW.

High oxalate levels in tropical plants are of some concern. Taro (*Colocasia esculenta*) and sweet potato (*Ipomoea batatas*) were reported to contain 278–574 mg $100\,g^{-1}$ fresh weight (FW) and 470 mg $100\,g^{-1}$, respectively. Total oxalate levels in yam (*Dioscorea alata*) tubers were reported in the range 486–781 mg $100\,g^{-1}$ DW, but may not be of nutritional concern since 50–75% of the oxalates were present in the water-soluble form and therefore may leach out during cooking. Oca or New Zealand yam (*Oxalis tuberosa* Mol.) contains 80–221 mg $100\,g^{-1}$ FW soluble oxalate. Higher levels of oxalates are usually found in the leaves and highest concentrations of oxalate have been found in the skin of these tropical root crops. Peanut greens, commonly consumed in tropical climates, are reported to contain 407 mg $100\,g^{-1}$. Coriander leaf (*Coriandrum sativum*) contains 1268 mg $100\,g^{-1}$; horsegram (*Macrotyloma uniflorum*) and santhi (*Boernavia diffusa*) contain 508 mg $100\,g^{-1}$ and 3800 mg $100\,g^{-1}$ respectively. Nuts such as peanuts, pecans, and cashews are relatively high in oxalates. Sesame seeds have been reported to contain high quantities of oxalate, ranging from 350 to 1750 mg $100\,g^{-1}$ FW.

Beverages with a high-oxalate concentration include Indian black tea, cocoa drinks, Ovaltine, cola, and certain types of beer.

The oxalic acid content is variable within some species; some cultivars of spinach (Universal, Winter

Table 1 Common foods exhibiting high levels of oxalates

Spinach	Beet	Rhubarb
Swiss chard	Yam, oca, sweet potato	Gooseberries, strawberries
Kale	Peanuts, pecans	Black tea
Purslane	Soybean	Cocoa, chocolate, Ovaltine
Collards	Wheat germ, wheat bran	Cola
Mustard and turnip greens	Sorrel, parsley, amaranth	Beer

Giant) contain 400–600 mg $100\,g^{-1}$, while others range from 700 to 900 mg $100\,g^{-1}$. Oxalic acid accumulates in plants, especially during dry conditions. A study comparing two cultivars of spinach, Magic (cv. summer) and Lead (cv. autumn), revealed that the summer cultivar contained greater amounts of oxalate (740 mg $100\,g^{-1}$ FW) than the autumn cultivar (560 mg $100\,g^{-1}$ FW). Oxalate content has been reported to increase as the plant ages and becomes overripe. The proportion of oxalic acid in the leaves of the goosefoot family can double during ripening. However, in tomatoes, oxalic acid content has been reported to decrease during ripening.

Absorption and Metabolism in Mammals

Calcium can combine with oxalate to form calcium oxalate in the intestinal lumen, making the calcium unavailable for absorption; calcium oxalate is then excreted in the feces. Free or soluble oxalate is absorbed by passive diffusion in the colon in humans; comparative studies between healthy individuals and those with ileostomies indicate that the colon is the principal site for oxalate absorption. However, it is suggested that the small intestine may be the major absorptive site rather than the colon. The absorption of oxalates from individual foods varies depending on their dietary conditions and source; in general, the absorption is quite limited. It has been estimated that 2–5% of administered oxalate is absorbed in humans. Experiments have shown that more oxalate is absorbed when consumed while fasting (12%) compared to only 7% oxalate absorption when consumed with a normal diet. The percentage of oxalate absorption varied markedly from 1% for rhubarb and spinach to 22% from tea, but generally absorption was higher at low doses.

Oxalate is an end product of ascorbate, glyoxylate, and glycine metabolism in mammals. Thirty-three to fifty percent of urinary oxalate is derived from ascorbate, 40% from glycine and 6–33% from minor metabolic pathways and dietary oxalate; dietary oxalate appears to account for only 10–15% of excreted oxalates.

Chemical Properties and Toxic Effects

Oxalic acid forms water-soluble salts with Na^+, K^+, and NH_4^+ ions; it also binds with Ca^{2+}, Fe^{2+}, and Mg^{2+}, rendering these minerals unavailable to animals. However, Zn^{2+} appears to be relatively unaffected. Calcium oxalate ($Ca(COO)_2$) is insoluble at a neutral or alkaline pH, but freely dissolves in acid.

Ingestion of 4–5 g of oxalate is the minimum dose capable of causing death in an adult, but reports have shown that 10–15 g is the usual amount required to cause fatalities. Oxalic acid ingestion results in corrosion of the mouth and gastrointestinal tract, gastric hemorrhage, renal failure, and hematuria. Other associated problems include low plasma calcium, which may cause convulsions and high plasma oxalates. Most fatalities from oxalate poisoning are apparently due to the removal of calcium ions from the serum by precipitation. High levels of oxalate may interfere with carbohydrate metabolism, particularly by succinic dehydrogenase inhibition; this may be a significant factor in death from oxalate toxicity induced in animals grazing pastures containing high levels of *Halogeton glomeratus*. Halogeton (*H. glomeratus*) and wood sorrel (*Oxalis cernua*) are high in oxalates and are known to cause injury to grazing cattle and sheep.

Although garden sorrel is a herb and not normally consumed in large quantities, there has been one report of fatal oxalate poisoning after a man consumed an estimated 6–8 g oxalate in vegetable soup containing 500 g sorrel. Both fatal and nonfatal poisoning by rhubarb leaves is thought to be caused by toxic anthraquinone glycosides rather than oxalates. Experiments involving the consumption of more than 30–35 g day^{-1} of cocoa, a high-oxalate foodstuff, by eight women provoked symptoms of intoxication including loss of appetite, nausea, and headaches. However, cocoa contains theobromine (1500–2500 mg $100\,g^{-1}$) and tannic acid (4000–6000 mg $100\,g^{-1}$), both of which are more toxic than the oxalic acid present (500–700 mg $100\,g^{-1}$). There appears to be a great deal of confusion as to what was responsible for these poisonings and it would be unwise to assume that only one factor was responsible.

Effect on Bioavailability of Minerals

High-oxalate foods have been known to inhibit calcium and iron absorption. Even though vegetables such as spinach, rhubarb, and Swiss chard are high in calcium, the calcium cannot be absorbed due to the presence of oxalates in these vegetables. When calcium absorption from spinach, a high-oxalate and high-calcium food, was compared with calcium absorption from milk, a high-calcium food, the results showed that the calcium from spinach is not readily available (only 5.1% absorbed), probably due to the high content of oxalates. The adverse effect of oxalates is greater if the oxalate-to-calcium ratio exceeds 9:4 (or approximately 2). The oxalate-to-calcium ratio in a food varies widely and can be classified into three groups, as summarized in Table 2.

Oxalate and calcium levels and the oxalate-to-calcium ratio of specific foods are detailed in Table 3.

Foods that have a ratio greater than two, as well as containing no utilizable calcium, have excess oxalate, which can bind calcium in other foods eaten at the same time. Foodstuffs with a ratio of about one do not encroach on the utilization of calcium provided by other products and therefore do not exert any demineralizing effects. However, these foods are not good sources of calcium. Although parsley (*Petroselinum sativum*) contains average levels of oxalate (140–200 mg $100\,g^{-1}$), its high calcium levels (180–290 mg $100\,g^{-1}$) reduce the oxalate-to-calcium ratio to a low level.

Oxalate appears to interfere only slightly with zinc absorption. A counteracting or protective mechanism may prevent the precipitation of zinc by oxalates. Increasing the proportion of magnesium ions in solution was reported to inhibit the binding of calcium and zinc oxalates. This observation explains the minor effect oxalates have on zinc absorption from some leafy vegetables, such as spinach, which has high levels of calcium and zinc, and relatively high levels of magnesium.

Oxalic acid may cause greater decreases in mineral availability if consumed with a high-fiber diet but the decrease may be only temporary. Negative calcium, magnesium, zinc, and copper balances were detected in males consuming a diet containing fiber and oxalates. When spinach was replaced by cauliflower, a low-oxalate vegetable, fiber had no effect on the minerals studied, indicating that the apparent negative balances obtained were due to the presence of oxalic acid.

Adverse Effects of Oxalates

Acute

A number of plants contain calcium oxalate crystals (measured as insoluble oxalate). When ingested, they are not absorbed into the blood stream and remain largely undissolved within the digestive tract, so they have no systemic toxicity, but the sharp raphide crystals can penetrate the tissues of the mouth and the tongue, causing considerable discomfort. Most of the plants that contain calcium oxalate crystals are members of the arum family. It has been suggested that calcium oxalate crystals are responsible for the irritating sensation in kiwi fruit (*Actinidia* sp.) and soluble oxalates are thought to account for the bitter taste present in some oca (*O. tuberosa* Mol.). Conophor seeds (*Tetracarpidium conophorum*) are a popular Nigerian snack, which have a bitter taste when raw but are palatable when cooked. This observation was correlated with a 73% decrease in total oxalate concentration after cooking.

Chronic

Oxalate is poorly absorbed under nonfasting conditions. Once absorbed free oxalate binds to calcium ions to form insoluble calcium oxalate, it remains in the insoluble form.

Free oxalate and calcium can precipitate in the urine and may form kidney stones. These stones consist mainly of calcium oxalate (80%), which is relatively insoluble in urine, and calcium phosphate (5%). Oxalate crystallizes with calcium in the renal vasculature and infiltrates vessel walls causing renal tubular obstruction, vascular necrosis, and hemorrhage, which lead to anuria, uremia, electrolyte disturbances, or even rupture. Kidney stones are becoming more common in men between the ages 30 and 50 years in industrialized countries. The risk factors involved in stone formation are a low volume of urine, increased urinary excretion of oxalate, calcium, or uric acid, a persistently low or high urinary

Table 2 Examples of plants with varying oxalate to calcium ratios

Oxalate-to-calcium ratio	*Examples*
Group 1: Plants with a ratio greater than 2	Spinach, rhubarb, beet, sorrel, cocoa
Group 2: Plants with a ratio of approx. 1	Potatoes, amaranth, gooseberries, currants
Group 3: Plants with a ratio less than 1	Lettuce, cabbage, cauliflower, green beans, peas

Table 3 Oxalate, calcium, and oxalate-to-calcium ratio (Ox:Ca) of some common foods

Foodstuff	*Oxalate (mg 100 g^{-1} FW)*		*Calcium (mg 100 g^{-1} FW)*		*Ox:Ca ratio*
	Range	*Mean*	*Range*	*Mean*	*(mmol l^{-1})*
Group 1					
Rhubarb (*Rheum rhaponticum*)					
cv. Victoria, forced, stewed		260		12.4	9.32
raw	275–1336	805	40–50	45	7.95
Common sorrel (*Rumex acetosa*)	270–730	500	35–45	40	5.56
Red beetroot (*Beta vulgaris*)	121–450	275	121–450	275	5.09
Garden sorrel (*Rumex patientia*)	300–700	500	40–50	45	4.94
Pig spinach (*Chenopodium* spp.)		1100		99	4.94
Purslane (*Portulaca oleracea*)	910–1679	1294	13–236	125	4.60
Spinach (*Spinacia oleracea*)	320–1260	970	80–122	101	4.27
Garden orach (*Atriplex hortensis*)	300–1500	900		100	4.00
New Zealand spinach (*Tetragonia expansa*)		890		100	3.96
Coffee (*Coffea arabica*)	50–150	100	10–15	12	3.70
Cashew (*Anacardium occidentale*)		231		41	2.50
Cocoa (*Theobroma cacao*)	500–900	700	100–150	125	2.49
Beet leaves (*Beta vulgaris* var. *cicla*)	300–920	610	100–120	110	2.46
Rhubarb (*Rheum rhaponticum*)					
cv. Crimson, end of season, stewed		460		91.5	2.23
Group 2					
Potato (*Solanum tuberosum*)	20–141	80	10–34	22	1.62
Amaranth (*Amaranthus polygonoicles*)		1586		595	1.18
Tea (*Thea chinensis*)	300–2000	1150	400–500	450	1.14
Amaranth (*Amaranthus tricolor*)		1087		453	1.07
Rhubarb (*Rheum rhaponticum*)					
cv. Victoria, end of season, stewed		620		266	1.04
Group 3					
Apple (*Malus* spp.)	0–30	15	5–15	10	0.67
Blackcurrant (*Ribes nigrum*)	2–90	50	19–50	35	0.63
Tomato (*Lycopersicum esculentum*)	5–35	20	10–20	15	0.58
Parsley (*Petroselinum sativum*)	140–200	170	180–290	235	0.32
Cabbage (*Brassica oleracea*)	0–125	60	200–300	250	0.11
Lettuce (*Lactuca sativa*)	5–20	12	73–90	81	0.07

FW, fresh weight.

Adapted from Zarembski PM and Hodgkinson A (1962) The oxalic acid content of English diets. *British Journal of Nutrition* 16: 627–634; Gontzea I and Sutzescu P (1968) *Natural Antinutritive Substances in Foodstuffs and Forages*, pp. 84–108. Basel: S Karger; Meena BA, Umapathy KP, Pankaja N and Prakash J (1987) Soluble and insoluble oxalates in selected foods. *Journal of Food Science and Technology* 24: 43–44; Noonan SC and Savage GP (1999) Oxalates and its effects on humans. *Asia Pacific Journal of Clinical Nutrition* 8: 64–74 with permission.

pH, and a low concentration of urinary inhibitors, such as magnesium, citrate, and high-molecular-weight polyanions. Normal urine is usually supersaturated with calcium oxalate. The normal urinary excretion of oxalate is less than 40–50 mg day^{-1} with less than 10% coming from the diet. Intakes of oxalate exceeding 180 mg day^{-1} lead to a marked increase in the amount excreted. Small increases in oxalate excretion have pronounced effects on the production of calcium oxalate in the urine, implying that foods high in oxalate can promote hyperoxaluria (high oxalate excretion) and increase the risk of stone formation. Rhubarb, spinach, beet, nuts, chocolate, tea, coffee, parsley, celery, and wheat bran cause significant increases in urinary oxalate excretion in healthy individuals and have been identified as the main dietary sources in the risk of kidney stone formation. It has been reported that black tea increased oxalate excretion by only 7.9%, compared with increases of 300% and 400% for spinach and rhubarb, respectively. Therefore 2–3 cups a day of black tea would have little effect on the risk of urinary stone formation when compared to spinach and rhubarb. It appears that tea is a significant source of oxalate intake in UK diets.

The main reason for the strong relationship between the risk of calcium stones and urinary oxalate excretion appears to be the effect that the latter has on the supersaturation of urine with calcium oxalate. The amount of oxalate excreted in the urine was higher in individuals with kidney stones than in healthy individuals, suggesting that those with kidney stones absorb more oxalate, consume more oxalate or oxalate-producing substances such as ascorbate, or metabolize more oxalate precursors. Excessive or

increased absorption of oxalate from normal diets is the result of intestinal abnormalities or malfunction. This is termed 'enteric hyperoxaluria' and is the commonest cause of increased renal oxalate excretion. It has been indicated that people with abnormal gastrointestinal absorption are at greater risk for hyperoxaluria and, as a result, kidney stone formation, than healthy individuals and should reduce their intake of oxalate and its precursors, such as ascorbate. A low-oxalate diet has prevented stone formation in some cases involving gastrointestinal disorders associated with hyperoxaluria.

An increase in calcium intake should be accompanied by a lower oxalate consumption, because a low-calcium and high-oxalate diet enhances oxalate absorption and excretion, which carries an even greater risk of stone formation than high calcium excretion. An increase in calcium intake may reduce urinary oxalate excretion by binding to more oxalate in the gut, thus reducing the risk of stone formation. Varying the amounts of calcium does not significantly alter levels of urinary calcium. From experimental work, it has been concluded that hypercalciuria plays, at most, a secondary role in the formation of calcium stones compared with mild hyperoxaluria.

Excessive ascorbic acid (vitamin C) intake may increase urinary oxalate output with an increased risk of forming kidney stones. An excess dose is considered to be 2000 mg of vitamin C per day. However, ascorbic acid doses greater than 500 mg day^{-1} were reported to induce a significant increase in urinary oxalate, and doses of 1000 mg day^{-1} would increase urinary oxalate excretion by 6–13 mg day^{-1}. The recommended daily intake in many countries is in the region of 80 mg.

Effects of Processing

Oxalates may be removed from food by leaching in water but this is not the most effective method as it removes only the soluble oxalate. Although the amount of oxalate in raw soybean (*Glycine max*) is relatively low, soaking and germination of the seed reduced the oxalate concentration. Cooking germinated soybeans reduced oxalate concentration below that in uncooked germinated soybeans. Soaking followed by cooking also proved to be effective, although not as effective as germination. Oxalate content in horsegram seeds (*M. uniflorum*) decreased by 38% when seeds were dehulled (508 and 315 mg 100 g^{-1}, for seed and dehulled seed, respectively). Roasting was found to be the least effective method. Roasting chicory roots was reported to increase oxalate content. Roasting oca (New Zealand yam, *O. tuberosa* Mol.) also increased oxalate levels by 10–26%. This may be caused by the decrease in moisture content, a hypothesis supported by reports of dry tropical leafy vegetables having higher oxalate concentrations than fresh vegetables.

A 40–50% loss of total oxalates by leaching was reported when yam tubers (*D. alata* and *D. esculenta*) were boiled, compared to steaming (20–25%) and baking (12–15%). Cooking proved most effective in reducing total oxalates. However, it must be noted that water-soluble minerals also leach out at the same time. Mineral leaching appears to vary between plant species. Blanching has been reported to decrease the oxalic acid content in spinach. However blanching, by conventional and microwave methods, reduced the oxalic acid content of sweet potato, peanut, and collard leaves only slightly whereas other antinutritional factors such as tannic and phytic acid were reduced significantly. Spinach, orach, and silverbeet are generally eaten after being boiled. However, rhubarb, cocoa, and common and garden sorrel may be consumed in the raw state and therefore should be eaten in smaller quantities.

Fermentation, frequently used in Asian countries, has been reported to decrease the oxalate content of foods. A marked decrease in oxalic acid content was reported in *Icacinia manni* (a starch tuber) after fermentation. Oxalic acid was observed to decrease by 37% (86 to 54 mg 100 g^{-1} FW) during souring of poi (a cooked taro paste) at 20 °C.

Recommendations

Foods high in oxalates should be consumed in moderation to insure optimum intake of minerals from the diet. Although some foods are reported to be high in calcium and other essential minerals, the amount available may be limited due to the presence of oxalates. For instance, spinach is a high-calcium food, yet because of its high oxalate content, the calcium availability is almost negligible. The availability of magnesium, iron, sodium, potassium, and phosphorus may also be restricted.

High-oxalate foods should be cooked to reduce the oxalate content. Soaking raw foods will also reduce the oxalate content but other useful nutrients such as water-soluble vitamins and minerals may also be lost at the same time. Oxalates tend to occur in higher concentrations in the leafy parts of vegetables rather than in roots or stalks.

For the general population, the occasional consumption of high-oxalate foods as part of a balanced diet does not pose any health problems. However, there are some groups of people who may be at risk from oxalate-induced side-effects.

Vegans and vegetarians should be aware that some foods contain high levels of oxalates. The diets of vegans and those persons with lactose intolerance may be low in calcium as a result of the exclusion of dairy products, unless their diet is supplemented by some other high-calcium food products. It is recommended that high-oxalate foods should be accompanied by calcium-rich foods such as dairy products and shellfish. If high-oxalate foods are consumed in conjunction with a low-calcium diet, then the consumer may be at risk of hyperoxaluria, which may lead to kidney stone formation.

Women tend to be more susceptible to calcium and iron deficiencies than men. Osteoporosis is a concern amongst females, especially after menopause. People suffering from fractures should also be aware of the potential effects of oxalates on mineral availability, as high calcium levels are required for bone repair. Once again, consumption of high-oxalate foods with an adequate-to-high calcium intake should pose no health problems. It must also be noted that calcium is only absorbed and used when there are adequate levels of vitamin D in the body, either obtained via the diet or synthesized by the body when exposed to sunlight. Women should eat red meats, which are low in oxalate, to satisfy their iron intake. Adequate levels of vitamin C are required for the absorption of iron, but excess amounts are not advised because ascorbic acid is converted into oxalate.

The risk of stone formation is three times greater in males and they should avoid eating excess amounts of high-oxalate foods. Sufferers of hyperoxaluria and kidney stones are advised to restrict their diet to foods containing low or medium levels of oxalates, as although urinary oxalate arises predominantly from endogenous sources, it can be influenced by dietary intake. Excess vitamin C intake is not recommended in these patients.

Inhabitants of tropical countries should be aware that leafy tropical plants and tropical root crops tend to contain higher levels of oxalates than plants from temperate climates. People living in these areas are at possible risk of stone formation due to hyperoxaluria, and mineral deficiencies if sufficient minerals are not consumed.

See also: **Ascorbic Acid**: Properties and Determination Physiology; **Calcium**: Properties and Determination; Physiology; **Iron**: Properties and Determination; Physiology; **Plant Antinutritional Factors**: Characteristics; **Renal Function and Disorders**: Nutritional Management of Renal Disorders; **Toxins in Food – Naturally Occurring**

Further Reading

Concon JM (1988) *Food Toxicology – Principles and Concepts*, pp. 416–419. New York: Marcel Dekker.

Dobbins JW and Binder HJ (1977) Importance of the colon in enteric hyperoxaluria. *New England Journal of Medicine* 296: 298–301.

Gontzea I and Sutzescu P (1968) *Natural Antinutritive Substances in Foodstuffs and Forages*, pp. 84–108. Basel: S Karger.

Hagler L and Herman RH (1973) Oxalate metabolism I. *American Journal of Clinical Nutrition* 26: 758–765.

Hanson CF, Frankos VH and Thompson WO (1989) Bioavailability of oxalic acid from spinach, sugar beet fibre and a solution of sodium oxalate consumed by female volunteers. *Food and Chemical Toxicology* 27: 181–184.

Linder MC (1991) *Nutritional Biochemistry and Metabolism with Clinical Applications*, 2nd edn. New York: Elsevier.

Massey LK, Roman-Smith H and Sutton RAL (1993) Effect of dietary oxalate and calcium on urinary oxalate and risk of formation of calcium oxalate kidney stones. *Journal of the American Dietetic Association* 93: 901–906.

Meena BA, Umapathy KP, Pankaja N and Prakash J (1987) Soluble and insoluble oxalates in selected foods. *Journal of Food Science and Technology* 24: 43–44.

Noonan SC and Savage GP (1999) Oxalates and its effects on humans. *Asia Pacific Journal of Clinical Nutrition* 8: 64–74.

Ross AB, Savage GP, Martin RJ and Vanhanen L (1999) Oxalates in oca (New Zealand yam) (*Oxalis tuberosa* Mol.). *Journal of Agricultural and Food Chemistry* 47: 5019–5022.

Sangketkit C, Savage GP, Martin RJ, Mason SL and Vanhanen L (1999) Oxalates in oca: a negative feature? In: Jenson J and Savage GP (eds) *Second South West Pacific Nutrition and Dietetic Conference Proceedings*, pp. 44–50. Auckland, New Zealand.

Strenge A, Hesse A, Bach D and Vahlensieck W (1981) Excretion of oxalic acid following the ingestion of various amounts of oxalic acid-rich foods. In Smith LH, Robertson WG and Finlayson B (eds) *Urolithiasis: clinical and basic research*, pp. 789–794. New York: Plenum Press.

Wanasundera JPD and Ravindran G (1994) Nutritional assessment of yam (*Dioscorea alata*) tubers. *Plant Foods for Human Nutrition* 46: 33–39.

Zarembski PM and Hodgkinson A (1962) The oxalic acid content of English diets. *British Journal of Nutrition* 16: 627–634.

OXIDATION OF FOOD COMPONENTS

K L Parkin and S Damodaran, University of Wisconsin-Madison, Madison, WI, USA

Scope

Antoine Lavoisier (1743–94) recognized oxidation as a chemical process, concluding that oxygen was the element responsible for the formation of acidic residues, or oxides, upon combustion of certain substances. A contemporary definition of oxidation is the process by which oxygen is added, or hydrogen or electrons are withdrawn. For a component to be oxidized, another has to be reduced, and reduction can be defined as the withdrawal of oxygen, or the addition of hydrogen or electrons. The component that is oxidized and loses electrons is the reductant and the component that is reduced and gains electrons is the oxidant. Oxidation is distinct from oxygenation; the latter is a noncovalent coordination of oxygen with a component, as in the case where hemoglobin and myoglobin bind oxygen to facilitate oxygen transport in blood and muscle, respectively. In food systems, oxygen is the most common oxidant, although other endogenous and added chemicals can also serve as oxidants. The principal negative effect of oxidation in foods is that flavor quality is lost, giving rise to the defect often referred to as oxidative rancidity. In addition, functional, color, and nutritional qualities of food components can be lost as a consequence of oxidation in foods. However, there are also some oxidative processes in foods that have beneficial effects on quality.

The Basic Process of Oxidation

Oxidation–Reduction Potentials

The potential or thermodynamic favorability for two components to be involved in an oxidation–reduction (redox) reaction can be predicted from the corresponding half-reactions of oxidation and reduction. Table 1 provides a selective list of some standard reduction half-reactions, using the hydrogen half-cell at pH 7.0 as a standard. As the reduction potential (voltage) becomes more positive, the tendency for that half-reaction to take place increases. Thus, the most powerful oxidants in Table 1 are hydrogen peroxide and oxygen. For each component, oxidation half-reactions take place in the reverse direction of what appears in Table 1 and have voltages of the opposite sign of the same magnitude.

The redox potential (E_h) of food systems is dependent on the concentration and redox states of the components of that system. One of the most important components is oxygen and, at limited dissolved oxygen components, E_h is strongly dependent on the oxygen content. 'Reducing' conditions or a very negative E_h (e.g., −400 mV) exist when dissolved oxygen is poised at near-anaerobic levels.

Transition metals such as copper and iron are believed to be involved in oxidation in foods via a redox cycling mechanism. Since there is very little 'free' iron and copper in biological systems, the types and concentrations of chelators present have a marked effect on the redox behavior and thus oxidative activity of these transition metals.

Oxidation in foods is often caused by free radical reactions. There are three stages of free radical oxidation, also referred to as autoxidation when the oxidant is oxygen. The first step, or 'initiation,' involves the formation of a free radical species ($X^{\cdot}$) from a biological component (XH), usually by the abstraction of a hydrogen atom ($H^{\cdot}$) by active oxygen or high-energy irradiation (eqn (1)). 'Propagation' of free radical oxidation processes occurs by chain reactions that consume oxygen and yield new free radical species (peroxy radicals, $XOO^{\cdot}$) or peroxides (XOOH), as in eqns (2) and (3). The products ($X^{\cdot}$ and XOOH, see also eqn (6)) can further propagate free radical reactions. 'Termination' of free radical oxidative reactions occurs when two radical species react with each other to form a nonradical adduct, as in eqn (4).

$$XH \rightarrow X^{\cdot} + H^{\cdot} \quad (1)$$

$$X^{\cdot} + {}^3O_2 \rightarrow XOO^{\cdot} \quad (2)$$

Table 1 Standard electrode potentials of selected reduction half-reactions

Reaction	*Volts*
$H_2O_2 + 2H^+ + 2e^- \rightarrow 2H_2O$	1.77
$O_2 + 4H^+ + 4e^- \rightarrow 2H_2O$	1.23
$Cu^{2+} + e^- \rightarrow Cu^+$	0.15
$Fe^{3+} + e^- \rightarrow Fe_2^+$	0.11
Dehydroascorbate + $H^+ + 2e^- \rightarrow$ Ascorbate	0.054
$2H^+ + 2e^- \rightarrow H_2$	0.00
$RSSR + 2H^+ + 2e^- \rightarrow 2RSH$	−0.39[a]

[a]Estimated value for oxidized disulfide (RSSR) conversion to reduced thiol (RSH) such as for oxidized/reduced glutathione and cystine/cysteine couples.
Standard Conditions: pH 7.0, 1 mol l^{-1} for each component.

$$XOO^{\cdot} + XH \rightarrow XOOH + X^{\cdot} \quad (3)$$

$$X^{\cdot} + X^{\cdot} \rightarrow X - X \quad (4)$$

Activation of Oxygen

In food systems, molecular oxygen (dioxygen; O_2) is generally the source of oxidizing power. Other strong oxidants include the food additives hydrogen peroxide (H_2O_2), calcium and benzoyl peroxides, and bromates ($KBrO_3$). All of these compounds are conjugates of oxygen. However, not all strong oxidants are composed of oxygen, as fluorine and bromine are also strong oxidants.

Although the reaction of ground state oxygen ($^3O_2$, triplet oxygen) with organic compounds is thermodynamically favorable, it is kinetically slow due to the high energy of activation required for oxygen to react. The electron configuration of $^3O_2$ includes two unpaired electrons in the outer shell, yielding a triplet signal in a magnetic field. All known organic compounds are in a singlet state, having all electrons paired with another. Consequently, facile reaction between $^3O_2$ and organic molecules is forbidden due to the incompatibility of their electron 'spin' states.

'Activation' of $^3O_2$ overcomes much of the energy barrier to its reactivity as an oxidant. One means of activation is the excitation of $^3O_2$ to yield singlet molecular oxygen ($^1O_2$); the latter has the two outer electrons paired in a single orbital. Other forms of 'activated' or 'reactive' oxygen result from the first three one-electron reductions to $^3O_2$ in the process of reducing $^3O_2$ to water. These reactive species of oxygen include the superoxide anion radical (O_2^-) and its conjugate acid (HO_2), hydrogen peroxide (H_2O_2), and the hydroxyl radical ($\cdot OH$). The standard reduction potentials of each of these steps is provided in Table 2. The electronic structure of these activated forms of oxygen facilitate their reactivity with biological compounds. The strongest electrophiles (electron seekers), $\cdot OH$ and $^1O_2$ are the most reactive forms of 'active' oxygen, followed in reactivity by O_2^- and then H_2O_2.

Table 2 Standard electrode potentials for univalent reductions of O_2 to H_2O

Reaction	*Volts*
$O_2 + e^- \rightarrow O_2^-$	−0.16 (−0.33)
$O_2^- + e^- + 2H^+ \rightarrow H_2O_2$	0.89
$H_2O_2 + e^- + H^+ \rightarrow \cdot OH + H_2O$	0.38
$\cdot OH + e^- + H^+ \rightarrow H_2O$	2.32

Standard conditions: pH 7.0, 1 mol l^{-1} for each component. For O_2, electrode potential also provided, in parentheses, at 10^5 Pa (0.987 atm).

Some of these active oxygen species can be interconverted, and these processes can be facilitated by the presence of specific catalysts. Activated forms of oxygen can also be formed by γ irradiation and by photosensitization of pigments in foods.

Catalysts of Oxidation Reactions

Catalysts of oxidation reactions can be enzymatic (protein) or nonenzymatic. Transition metals (M^n, reduced form; M^{n+1}, oxidized form) can participate in redox reactions with $^3O_2$ to yield O_2^-/HO_2, as in eqn. (5). The resulting O_2^- can initiate oxidation reactions. Another manner by which transition metals can cause oxidation reactions is by breaking down lipid hydroperoxides (LOOH) (eqn (6)), and the alkoxy radical ($LO^{\cdot}$) so formed can cause further oxidative reactions. Since there are often small quantities of LOOH in food systems this process is probably important. Transition metals can also take part in the interconversion of active oxygen species, as in the Haber–Weiss reaction (eqn (7)). This reaction can be mediated by three partial reactions (eqns (8)–(10)). In the first reaction (eqn (8)), O_2^- acts as a reductant and donates an electron to an oxidized transition metal (e.g., iron). In the second step (eqn (9)), O_2^-, acting as both an oxidant and reductant, undergoes dismutation to form H_2O_2 and $^3O_2$. In the third step (eqn (10)), also called the Fenton reaction, the reduced transition metal donates an electron to H_2O_2 to form the extremely reactive $\cdot OH$, and the transition metal reverts back to its oxidized state to allow another cycle. In food systems, several endogenous components, such as ascorbic acid and thiol compounds, can replace O_2^- as a reductant. This set of reactions also illustrates the participation in oxidative reactions of all activated oxygen species generated by univalent electron reductions of $^3O_2$.

$$M^n + {}^3O_2 \rightarrow M^{n+1} + O_2^- \quad (5)$$

$$M^n + LOOH \rightarrow M^{n+1} + LO^{\cdot} + OH^- \quad (6)$$

$$O_2^- + H_2O_2 \rightarrow {}^3O_2 + OH^- + \cdot OH \quad (7)$$

$$O_2^- + M^{n+1} \rightarrow {}^3O_2 + M^n \quad (8)$$

$$2O_2^- + 2H^+ \rightarrow {}^3O_2 + H_2O_2 \quad (9)$$

$$H_2O_2 + M^n \rightarrow M^{n+1} + \cdot OH + OH^- \quad (10)$$

Other nonenzymatic catalysts include photosensitive pigments in foods. Photosensitive pigments become elevated to an excited triplet state upon the absorption of light energy, and can transfer that energy to $^3O_2$ or other biological components. Some pigments favor transmission of energy to organic

compounds (type I process) which ultimately yield O_2^- and H_2O_2 from $^3O_2$. Other pigments favor transmission of energy directly to $^3O_2$ to yield $^1O_2$ (type II process). Examples of photosensitizers of each of these types in foods are riboflavin and chlorophyll, respectively. (*See* **Chlorophyl**; **Riboflavin**: Properties and Determination.)

Enzymatic catalysts of oxidative reactions usually cause oxidations of specific biological compounds. For example, the enzymes lipoxygenase, polyphenoloxidase, sulfhydryl oxidase and xanthine oxidase are common to foods and cause the specific oxidation of unsaturated fatty acids, mono- and diphenolic acids, protein thiol (cysteine) residues, and xanthine, respectively. Glucose oxidase converts glucose to gluconic acid and also produces H_2O_2. Xanthine oxidase and peroxidase can produce H_2O_2 and O_2^- and $^1O_2$, respectively, and this is dependent on which substrates are being utilized and the level of oxygen present. These active oxygen species may cause the oxidation of other biological compounds, leading to losses in food quality.

Oxidation of Food Components

Lipids

Polyunsaturated fatty acids with 1,4-pentadiene functional units are particularly sensitive to oxidative reactions. Using linoleic acid as an example, oxidation can be initiated by two basic mechanisms, abstraction (autoxidation) and 'ene' addition (Figure 1). Abstraction is when an electron (or hydrogen atom) is removed from the fatty acid by reaction with an electrophilic species such as $^{\cdot}OH$ or $X^{\cdot}$, or by interaction with high-energy radiation. The initial abstraction step yields a fatty acid free radical (initiation step) which can then undergo addition of $^3O_2$ (propagation step) and then abstract an electron from another biological compound. The methylene or 'allylic' hydrogen atoms of the pentadiene structure are most readily abstracted. The resulting free radical ($L^{\cdot}$) can be stabilized by resonance along the original pentadiene structure, and the fatty acid radical tends to undergo addition of $^3O_2$ when the unpaired electron is most 'delocalized' or located at the terminal sites, or C9 and C13, resulting in the formation of first the linoleic acid 9- and 13-hydroperoxyl radicals, and then the 9- and 13-OOH (hydroperoxides) isomers. Further oxidative processes can be initiated by interaction of these hydroperoxides with transition metals as previously described in eqn (6). (*See* **Fatty Acids**: Properties.)

The 'ene' addition reaction that can initiate lipid oxidation is caused by the highly electrophilic $^1O_2$, which will add directly to the double bond since this is where the highest electron density can be found. Thus, a mixture of 9-, 10-, 12-, and 13-OOH isomers are produced by $^1O_2$ reaction with linoleic acid.

Once the fatty acid hydroperoxides and hydroperoxyl radicals are formed, additional initiation reactions can take place for these initial products, being unstable, can be subject to secondary reactions, as shown for linolenic acid oxidation (Figure 2). Oxidation of any remaining double bonds can take place, and in some cases the fatty acid radicals can attack adjacent intramolecular double bonds, forming cyclic structures. Alternatively, hydroperoxyl fatty acids can react with adjacent fatty acids to yield polymerized oxidation products. 'Scission' reactions lead to fracture of the fatty acid chain and result in the emanation of reduced molecular weight ketones and aldehydes. These latter secondary products, being fairly volatile, give rise to the off-flavors and odors that are associated with oxidized foods or oxidative rancidity. One product that can be formed by secondary reactions of oxidizing lipids is malondialdehyde (MDA). MDA is often used by food scientists as an indicator of the degree of oxidation of lipids in foods. In cases where specific secondary products are formed by enzyme reactions, such as by lipid hydroperoxide lyases in freshly cut cucumber and tomato fruits, the resulting volatile compounds are pleasant and contribute desirable aromatic qualities.

Thermally induced oxidation reactions can occur in both saturated and unsaturated lipids at temperatures encountered during processes such as deep-fat frying. Oxidation generally proceeds via the initial formation of hydroperoxides. The high temperatures can cause many isomerization and scission reactions to take place, producing a myriad of secondary or breakdown products such as epoxides, dihydroperoxides, cyclized fatty acids, dimers, and aldehydes and ketones resulting from scission reactions.

Proteins

Proteins, peptides, and amino acids in foods undergo several oxidative changes during food processing. The amino acids that are most susceptible to oxidative degradation are methionine, cysteine (cystine), histidine, and tryptophan. Under severe oxidizing conditions tyrosine, serine, and threonine are also oxidized to some extent. Oxidation of proteins and amino acids is caused by several agents, such as light, γ irradiation, peroxidizing lipids, metal ions, the products of enzymatic and nonenzymatic browning reactions, and food additives such as hydrogen peroxide, benzoyl peroxide, bromates ($KBrO_3$) and azodicarbonamide. (*See* **Amino Acids:** Properties and Occurrence; **Protein**: Chemistry.)

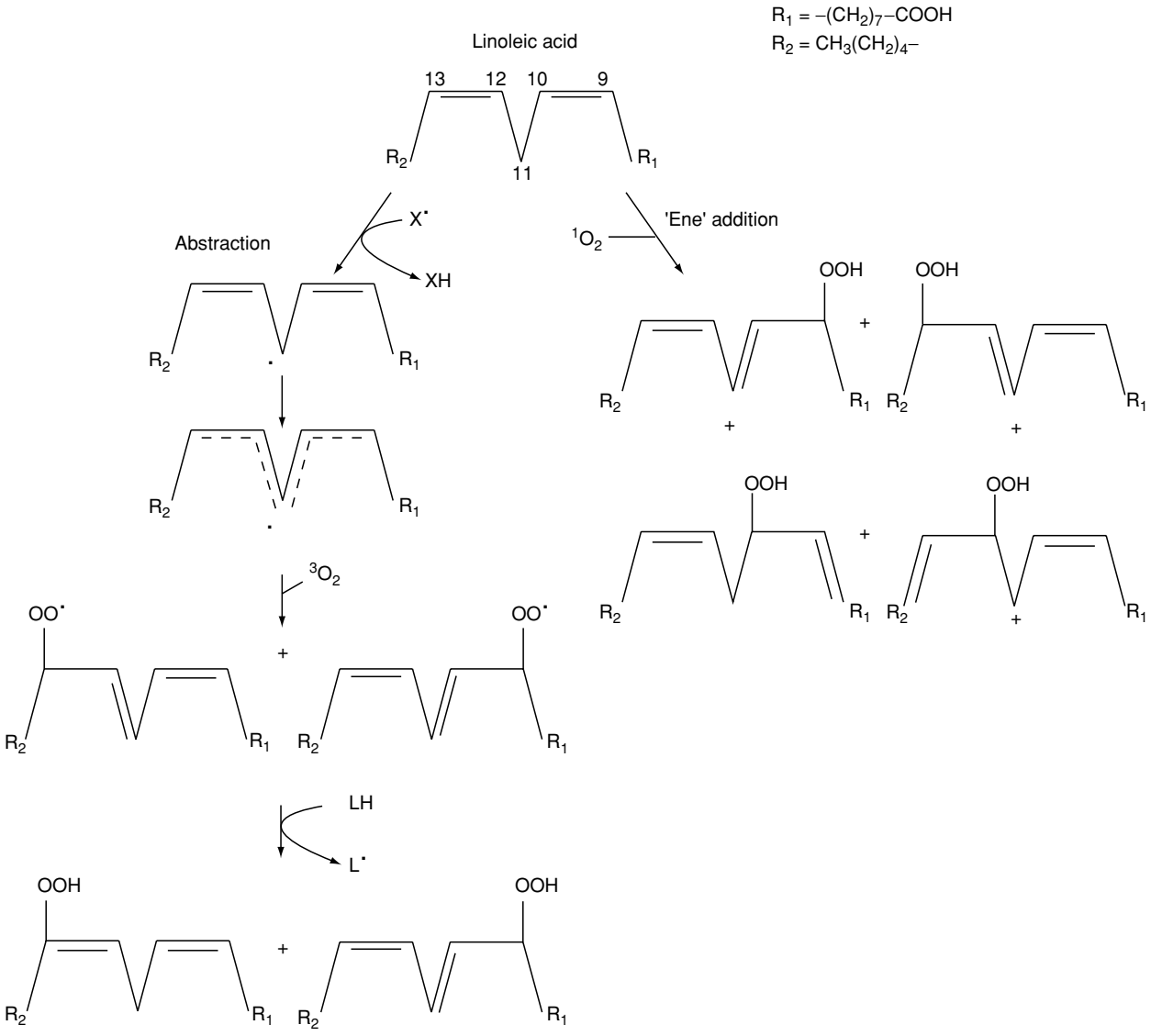

Figure 1 Initiation reactions for the oxidation of linoleic acid. Reproduced from Oxidation of Food Components, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Treatment of proteins with hydrogen peroxide or calcium peroxide causes oxidation of methionine sulfoxide (reversible), which can be further oxidized to methionine sulfones (irreversible) (eqn (11)). Cysteine residues can be oxidized by peroxides or other forms of activated oxygen to yield the sulfenic (Cy-SOH), sulfinic (Cy—SO_2H) and sulfonic (Cy—SO_3H), acid derivatives. Oxidation of cystine residues in proteins results in the formation of mono-, di-, tri-, and tetrasulfoxides.

Free thiol groups in proteins are readily oxidized by atmospheric oxygen to form disulfide cross-links. Free thiol groups also catalyze thiol-disulfide interchange reactions, which often lead to polymerization of proteins. Oxidizing agents such as $KBrO_3$ and azodicarbonamide are often used as additives in wheat flour in order to improve dough formation. These additives are believed to oxidize and block the free thiol groups of protein and nonprotein constituents, and thus prevent the occurrence of thiol–disulfide interchange reactions in the dough. The modulation of oxidation–reduction behavior in dough systems may also be controlled by ascorbic acid, dehydroascorbic acid, and glutathione.

When foods containing photosensitive substances, such as riboflavin and chlorophyl, are exposed to light, the amino acids histidine, cysteine, methionine, tryptophan, and tyrosine are oxidized by the activated oxygen species O_2^-, H_2O_2 and $^1O_2$. γ Irradiation of foods results in the formation of H_2O_2

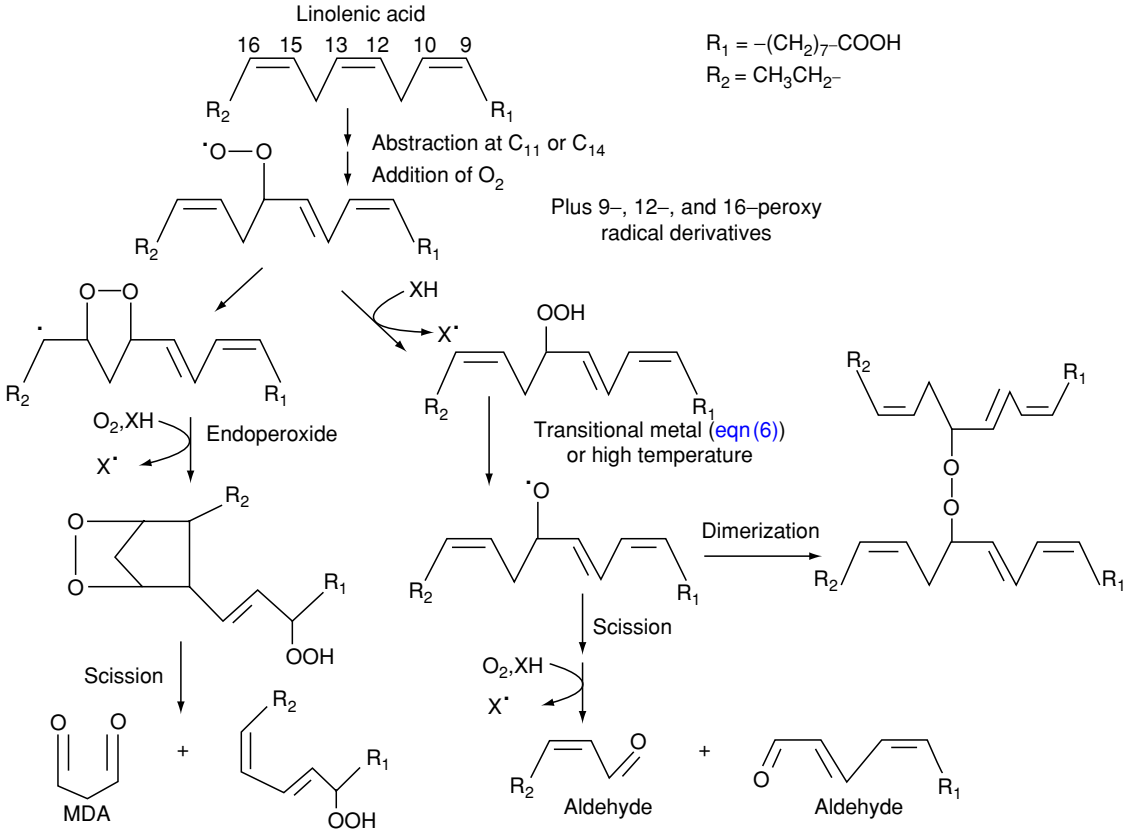

Figure 2 Initiation and secondary reactions for the oxidation of linolenic acid. Reproduced from Oxidation of Food Components, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

$$-CH_2-CH_2-S-CH_3 \rightleftharpoons -CH_2-CH_2-\overset{\overset{O}{\|}}{S}-CH_3 \rightarrow -CH_2-CH_2-\overset{\overset{O}{\|}}{\underset{\underset{O}{\|}}{S}}-CH_3 \qquad (11)$$

through radiolysis of water in the presence of oxygen, which in turn causes oxidative changes in proteins. Tryptophan residues can also be oxidized upon exposure of proteins to acidic conditions.

Substantial oxidation of free amino acids and amino acid residues in proteins occurs in the presence of peroxidizing lipids. Methionine, cysteine, histidine, and lysine are the most susceptible amino acids/residues. Two types of mechanisms, one involving the alkoxy ($LO^{\cdot}$) and peroxy ($LOO^{\cdot}$) free radicals, and the other involving malondialdehyde and other carbonyl compounds, are believed to be involved in the oxidation of proteins by peroxidizing lipids. In the first case, the lipid free radicals react with proteins (P) and induce formation of protein free radicals ($P^{\cdot}$), followed by polymerization of protein molecules (eqns (12)–(16)). In addition to the free radical-induced polymerization of protein molecules, the lipid peroxides formed during the reactions oxidize methionine, cysteine, histidine, and tryptophan residues. The highly reactive malondialdehyde formed in peroxidizing lipids reacts with amino groups of lysyl residues, resulting in intermolecular cross-links.

$$LOO^{\cdot} + P \rightarrow {}^{\cdot}LOOP \underset{O_2}{\rightarrow} {}^{\cdot}OOLOOP \underset{P}{\rightarrow} POOLOOP \qquad (12)$$

or

$$LO^{\cdot} + P \rightarrow LOH + P^{\cdot} \qquad (13)$$

$$LOO^{\cdot} + P \rightarrow LOOH + P^{\cdot} \qquad (14)$$

$$P^{\cdot} + P \rightarrow P - P^{\cdot} \qquad (15)$$

$$P - P^{\cdot} + P \rightarrow P - P - P^{\cdot} \qquad (16)$$

Heat treatment of proteinaceous foods causes several oxidative changes in proteins. While mild heat treatment results in protein denaturation and loss of functionality, severe heat treatment often causes undesirable chemical changes in amino acid residues and complex reactions of proteins with

other food components, such as carbohydrates and lipids. (*See* **Browning**: Nonenzymatic.)

When protein is heated at temperatures above 300 °C, as commonly encountered during broiling and grilling, several amino acid residues undergo thermal decomposition and pyrolysis. Several of these pyrolysis products have been isolated, identified, and shown to be highly mutagenic. The most carcinogenic/mutagenic products are formed from the decomposition of tryptophan, glutamate and lysyl residues. (*See* **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; **Mutagens**.)

Carbohydrates

Carbohydrates are not as sensitive to oxidation reactions as are lipids and proteins. In addition, since many oxidation products are not volatile, the practical consequences of carbohydrate oxidations in foods are limited. Oxidation of food carbohydrates can take place, especially at high temperatures, resulting in caramelization reactions. (*See* **Caramel**: Methods of Manufacture; **Carbohydrates**: Interactions with Other Food Components.)

Some industrial processes employ oxidation reactions to prepare functional derivatives of monosaccharides for the chemical industries. However, the carbohydrate oxidations of most relevance to foods involve enzymatic reactions. Glucose oxidase oxidizes glucose to gluconic acid while simultaneously reducing O_2 to H_2O_2. The enzyme is added to foods to reduce glucose levels (to prevent nonenzymatic browning in eggs to be dried) or oxygen tension (to stabilize beverages and salad dressings from oxidative deterioration).

Carbohydrates can be oxidized by the same free radical mechanisms as described for lipids. Low-molecular-weight carbohydrates such as glucose, mannitol, and deoxyribose are known to react with $^{\cdot}OH$ and produce oxidized derivatives. Again, these derivatives, when present, have little impact on food quality and are thus of little practical significance.

Minor Food Components

Oxidation of minor food components can also influence food quality. Oxidation of ascorbic acid by enzymatic (ascorbic acid oxidase) or nonenzymatic means can compromise nutritional quality. Ascorbic acid, and other organic acids, can be degraded by active oxygen species and by reactions initiated by transition metals. Polyphenol oxidase and tyrosinase are enzymes found in plant and crustacean foods that oxidize phenolic acids and initiate secondary nonenzymatic reactions that are responsible for darkening and often loss of color quality. (*See* **Ascorbic Acid**: Properties and Determination.)

Environmental Factors

Temperature

Generally, as the temperature increases, the rate of oxidation reactions also increases. Rate increases usually follow a $Q_{10} = 2$ relationship, or a doubling in rate for every rise of 10 °C, provided no change in the mechanism of reaction occurs with a corresponding change in temperature and as long as competing reactions have little impact on the reactants. For oxidative reactions caused by enzymes, an optimum temperature exists. This is because enzymes, being proteins, are denatured above a characteristic temperature and will lose biological activity. Another factor is that, as temperature increases, oxygen solubility in water decreases and this could attenuate the temperature activation of oxidative reactions if oxygen was a limiting component in the process.

Moisture

Lipid oxidation reactions generally have rate minima at intermediate water activities (a_w) of about 0.3. As a_w increases above 0.3, rates of oxidation increase, probably due to increased mobility and activity of catalysts. At a_w below 0.3, rates of oxidation increase, perhaps due to solvation and removal of catalysts and reaction intermediates from lipids and into the aqueous phase. (*See* **Water Activity**: Effect on Food Stability.)

Chemical Composition of Food

The relative concentrations of various prooxidants and antioxidants in the food, and their identities and relative reactivities, will greatly influence the rate of oxidation reactions. Antioxidants, such as butylated hydroxytoluene, butylated hydroxyanisole, propyl gallate, ascorbic acid, and sodium bisulfite can be added to impede oxidative reactions. On the other hand, the inadvertent addition of transition metals, such as iron and copper, from processing equipment can hasten oxidative reactions in foods. In addition, the nature of the substrate sensitive to oxidative reactions is important. For example, foods rich in polyunsaturated fatty acids (vegetable and fish oils) are more sensitive to oxidation than those rich in monounsaturated fatty acids (animal depot fats). (*See* **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants.)

Exogenous Factors

Packaging materials and strategies can have an influence on rates of oxidation. Foods containing photosensitive pigments are often packed in opaque or translucent containers to minimize photooxidative

processes by preventing the activation of oxygen by these pigments. The head space of products in containers can also be controlled to inhibit oxidative processes. For example, foods can be sealed in containers after the head space has been flushed with an inert gas such as nitrogen or under vacuum. Both approaches serve to minimize the oxygen available to support oxidative reactions. Products may also be coated, such as with sugar syrups on fruits to be frozen, to minimize the availability of oxygen for oxidative reactions.

Examples in Foods

Many foods are susceptible to oxidative reactions. Plant and fish oils, due to their high levels of unsaturation, can deteriorate in flavor very rapidly if oxidation is allowed to take place. Lipid oxidation gives rise to the 'fishy' flavor and aroma in frozen fish and this limits acceptable storage life. Refining procedures for vegetable and seed oils are partly designed to remove chlorophyl which can cause flavor deterioration by initiating photooxidative reactions. Another light-sensitive food is milk. The endogenous riboflavin can cause photooxidation of lipids and proteins and yield the undesirable 'light-activated flavor.' Early recognition of this problem gave rise to the domestic delivery of milk in opaque containers, in many countries, to prevent its exposure to light. (*See* **Fish Oils**: Composition and Properties; **Milk**: Processing of Liquid Milk; **Vegetable Oils**: Dietary Importance.)

Nuts and high-fat products, such as potato crisps, can be packaged in containers having a modified atmosphere or head space. Nitrogen flushing of containers and packing under vacuum are processes designed to limit the amount of oxygen available for oxidation reactions. (*See* **Chilled Storage**: Use of Modified-atmosphere Packaging.)

Most plant tissues, particularly fresh fruits, brown excessively upon cutting or bruising. This is often due to the presence of polyphenol oxidase, which acts on phenolic acids in these foods. After the initial enzymatic oxidations, a series of subsequent nonenzymatic oxidations convert these phenolic acids into polymers that are responsible for the brown color. (*See* **Phenolic Compounds**.)

Not all oxidative reactions in foods are undesirable. In the extrusion processing of foods at alkaline pH, the oxidative polymerization of proteins results in the desirable texturization in simulated meat products. Protein and lipid oxidation are also recognized for their beneficial effects on dough strengthening in the baking industries. A third example is the emanation of characteristic flavors and aromas upon slicing of fresh cucumbers and tomatoes. The sources of these flavors are unsaturated fatty acids which have been initially oxidized by endogenous lipoxygenase activity and have been transformed further by other enzymes. The chemical 'fermentation' of tea leaves and the 'ripening' of olives are achieved by potentiating polyphenol oxidase enzyme activity on endogenous phenolic acids, causing the development of desirable coloration in these products.

Generally, oxidative reactions that can be controlled can be manipulated to yield beneficial effects, whereas those that cannot often yield detrimental effects on food quality.

See also: **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; **Ascorbic Acid**: Properties and Determination; **Browning**: Nonenzymatic; **Chilled Storage**: Use of Modified-atmosphere Packaging; **Chlorophyl**; **Fatty Acids**: Metabolism; **Fish Oils**: Composition and Properties; **Milk**: Processing of Liquid Milk; **Phenolic Compounds**; **Riboflavin**: Properties and Determination; **Vegetable Oils**: Dietary Importance; **Water Activity**: Effect on Food Stability

Further Reading

Buettner GR (1993) The pecking order of free radicals and antioxidants: lipid peroxidation, α-tocopherol, and ascorbate. *Archives of Biochemistry and Biophysics* 300: 535–543.

Clark WM (1960) *Oxidation–Reduction Potentials of Organic Systems*. Baltimore: Williams and Wilkins.

Frankel EN (1998) *Lipid Oxidation*. Dundee: Theory Press.

Kanner J, German JB and Kinsella JE (1987) Initiation of lipid peroxidation in biological systems. *CRC Critical Reviews in Food Science and Nutrition* 25: 317–364.

Packer L (ed.) (1984) Oxygen radicals in biological systems. *Methods in Enzymology* 105.

Richardson T and Finley JW (eds) (1985) *Chemical Changes in Food During Processing*. Westport, CT: AVI.

OXIDATIVE PHOSPHORYLATION

D A Bender, University College London, London, UK

Background

The total body content of ATP is of the order of 10 g, whereas the daily turnover of ATP is equal to the body weight, some 70 kg. A small number of metabolic reactions involve direct transfer of phosphate from a phosphorylated substrate on to ADP, forming ATP – substrate-level phosphorylation. Under normal conditions, almost all of the phosphorylation of ADP to ATP occurs in the mitochondria, by the process of oxidative phosphorylation – the oxidation of reduced coenzymes linked to the reduction of oxygen to water and (under normal conditions) obligatorily linked to phosphorylation of ADP → ATP. This obligatory linkage of substrate oxidation, reoxidation of reduced coenzymes, and reduction of oxygen to water with the phosphorylation of ADP mean that the availability of ADP controls the rate at which substrates are oxidized. In turn, the availability of ADP to be phosphorylated is dependent on the rate of utilization of ATP in performing physical and chemical work. Thus, energy expenditure in physical and chemical work controls the rate at which metabolic fuels are oxidized, rather than being used to form reserves of (mainly) adipose tissue triacylglycerol.

With the exception of glycolysis and the pentose phosphate pathway, most of the reactions in the oxidation of metabolic fuels occur inside the mitochondria and lead to the reduction of nicotinamide nucleotide and flavin coenzymes. Within the inner membrane of the mitochondrion, there is a series of coenzymes that are able to undergo reduction and oxidation. The first coenzyme in the chain is reduced by reaction with NADH, and is then reoxidized by reducing the next coenzyme. In turn, each coenzyme in the chain is reduced by the preceding coenzyme, and then reoxidized by reducing the next coenzyme. The final step is the oxidation of a reduced coenzyme by oxygen, resulting in the formation of water. Figure 1 shows an overview of this mitochondrial electron transport chain.

Experimentally, the electron transport chain can be dissected into four complexes of coenzymes, which catalyze:

1. Oxidation of NADH leading to the reduction of ubiquinone to ubiquinol. This complex is associated with the phosphorylation of ADP → ATP.
2. Oxidation of reduced flavoproteins and reduction of ubiquinone to ubiquinol. This complex is not associated with phosphorylation of ADP.
3. Oxidation of ubiquinol, leading to reduction of cytochrome c. This complex is associated with the phosphorylation of ADP → ATP.
4. Oxidation of reduced cytochrome c leading to the reduction of oxygen to water. This complex is associated with the phosphorylation of ADP → ATP.

This means that there are three sites in the electron transport chain between NADH and oxygen that are linked to the phosphorylation of ADP → ATP, but only two between reduced flavoproteins and oxygen. Experimentally, this is seen as a ratio of phosphate esterified:oxygen consumed (the P:O ratio) of approximately 3 when substrates that reduce NAD^+ are oxidized, and approximately 2 when substrates that reduce flavoproteins are oxidized.

Experimentally, mitochondrial metabolism is measured using the oxygen electrode, in which the percentage saturation of the buffer with oxygen is measured electrochemically as the mitochondria oxidize substrates and reduce oxygen to water. Figure 2 shows the oxygen electrode traces for oxidation of malate (which is linked to reduction of NAD^+) and succinate (which is linked to reduction of a flavoprotein). The greater consumption of oxygen for oxidation of succinate compared with the same amount of malate reflects the lower P:O ratio for succinate oxidation.

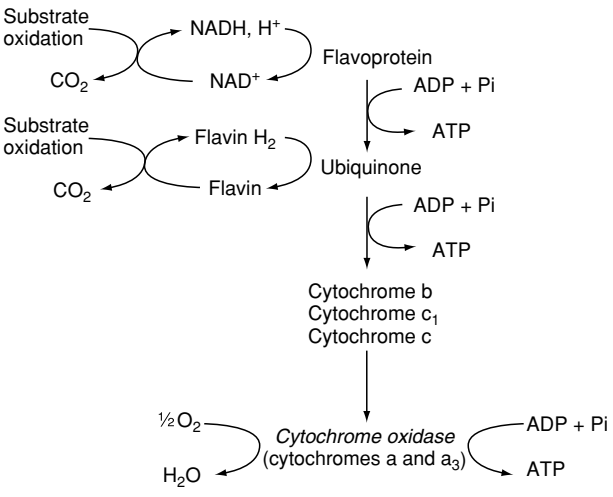

Figure 1 Overview of the mitochondrial electron transport chain.

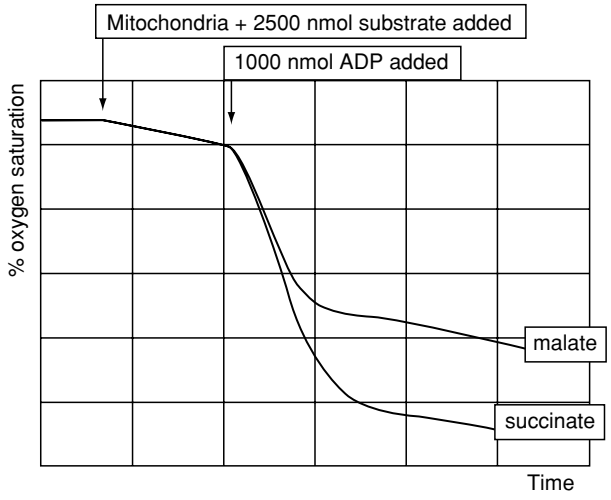

Figure 2 Oxygen electrode traces for mitochondria oxidizing malate and succinate.

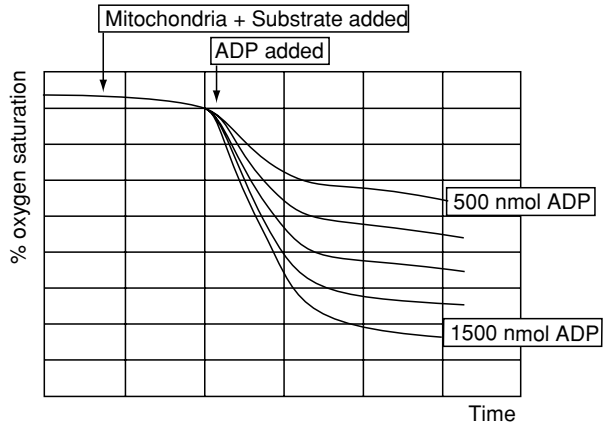

Figure 3 Oxygen electrode traces for mitochondria oxidizing malate with varying amounts of ADP.

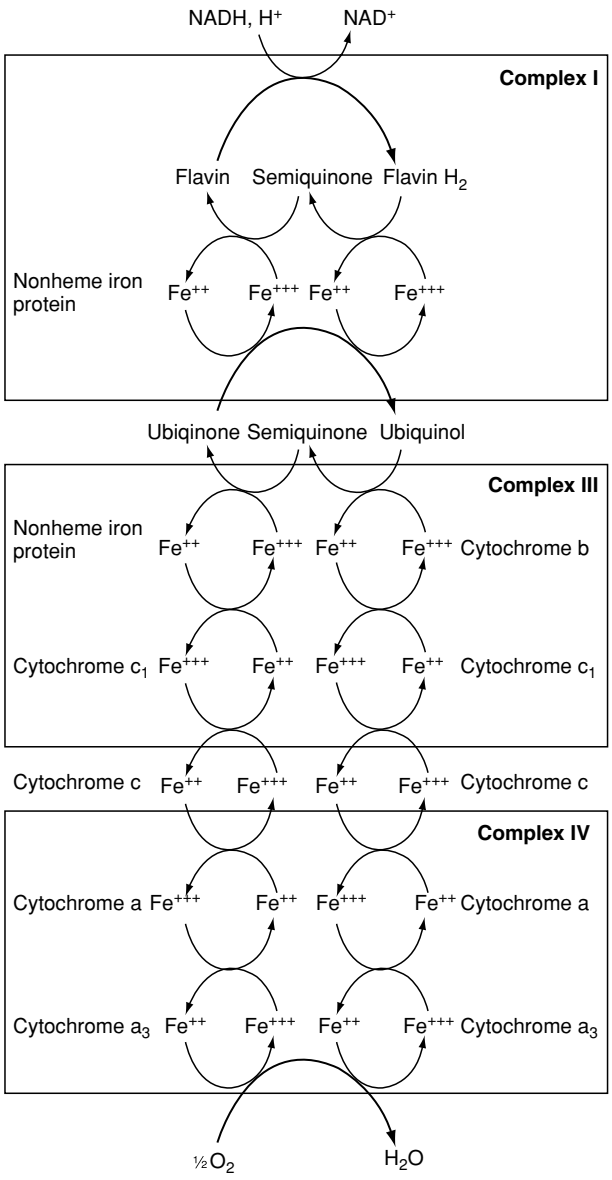

Figure 4 Mitochondrial electron transport chain.

Figure 3 shows the oxygen electrode traces for mitochondria incubated with varying amounts of ADP, and a super-abundant amount of malate. As more ADP is provided, so there is more oxidation of substrate, and hence more consumption of oxygen. This illustrates the tight coupling between the oxidation of metabolic fuels and the availability of ADP to be phosphorylated.

Mitochondrial Electron Transport Chain

The mitochondrial electron transport chain is a series of enzymes and coenzymes in the crista membrane, each of which is reduced by the preceding coenzyme, and in turn reduces the next, until finally the protons and electrons that have entered the chain from either NADH or reduced flavin reduce oxygen to water. The sequence of the electron carriers shown in Figures 1 and 4 has been determined in two ways:

- By consideration of their electrochemical redox potentials, which permits determination of which carrier is likely to reduce another, and which is likely to be reduced.
- By incubation of mitochondria with substrates, in the absence of oxygen, when all of the carriers become reduced, then introducing a limited amount of oxygen, and following the sequence in which the carriers become oxidized. The oxidation state of the carriers is determined by following changes in their absorption spectra.

In order to understand how the transfer of electrons through the electron transport chain can be linked to the phosphorylation of ADP to ATP, it is necessary to consider the chemistry of the various electron carriers. They can be classified into two groups:

Hydrogen carriers, which undergo reduction and oxidation reactions involving both protons and electrons – these are NAD, flavins, and ubiquinone. As shown in Figure 5, NAD undergoes a two-electron oxidation/reduction reaction, while both the flavins and ubiquinone undergo two single electron reactions to form a half-reduced radical, then the fully reduced coenzyme. Flavins can also undergo a two electron reaction in a single step.

Electron carriers, which contain iron (and, in the case of cytochrome oxidase, also copper) undergo oxidation and reduction by electron transfer alone. These are the cytochromes, in which the iron is present in a heme molecule, and nonheme iron proteins, sometimes called iron–sulfur proteins, because the iron is bound to the protein through the sulfur of the amino acid cysteine. Figure 6 shows the arrangement of the iron in nonheme iron proteins, and the three different types of heme that occur in cytochromes:

- heme (protoporphyrin IX), which is tightly but noncovalently bound to proteins, including cytochromes b and b_1, as well as enzymes such as catalase, and the oxygen transport proteins hemoglobin and myoglobin;
- heme C, which is covalently bound to protein in cytochromes c and c_1;
- heme A, which is anchored in the membrane by its hydrophobic side chain, in cytochromes a and a_3 (which together form cytochrome oxidase).

The hydrogen and electron carriers of the electron transport chain are arranged in sequence in the crista membrane, as shown in Figure 4. Some carriers are entirely within the membrane, whereas others are located on the inner or outer face of the membrane. Each of the three complexes in which phosphorylation of ADP → ADP is linked to electron transport forms a membrane-spanning unit.

There are two steps in which a hydrogen carrier reduces an electron carrier: the reaction between the flavin and nonheme iron protein in complex I, and the reaction between ubiquinol and cytochrome b plus a nonheme iron protein in complex III. The reaction between nonheme iron protein and ubiquinone in complex I is the reverse – a hydrogen carrier is reduced by an electron carrier.

When a hydrogen carrier reduces an electron carrier, there is a proton that is not transferred onto the electron carrier, but is extruded from the membrane, into the crista space, as shown in Figure 7. When an electron carrier reduces a hydrogen carrier, there is a need for a proton to accompany the electron that is transferred. This is acquired from the mitochondrial matrix, thus shifting the equilibrium between H_2O and $H^+ + OH^-$, resulting in an accumulation of hydroxyl ions in the matrix.

Similar pumping of protons across the crista membrane occurs in complexes III and IV, although it is less obvious than in complex I. Thus, each complex that is associated with phosphorylation of ADP → ATP pumps protons into the crista space as it transports electrons.

Phosphorylation of ADP Linked to Electron Transport

The result of electron transport through the sequence of carriers shown in Figure 4 is a separation of protons and hydroxyl ions across the crista membrane, with an accumulation of protons in the crista space, and an accumulation of hydroxyl ions in the matrix, i.e., creation of a pH gradient across the crista membrane.

This proton gradient provides the 'driving force' for the phosphorylation of ADP → ATP – a highly endothermic reaction. Protons reenter the mitochondrial matrix, down the proton gradient, through transport pores in the membrane that are an integral part of the mitochondrial primary particles that contain the ATP synthase, and form the transmembrane stalk of the primary particles.

ATP synthase acts as a molecular motor, driven by the flow of protons down the concentration gradient from the crista space into the matrix, through the transmembrane stalk of the primary particle. As protons flow through the stalk, they cause rotation of the core of the multienzyme complex that makes up the primary particle containing ATP synthase.

As shown in Figure 8, there are three ATP synthase catalytic sites in the primary particle, and each one-third turn of the central core causes a conformational change at each active site:

- at one site, the conformational change permits binding of ADP and phosphate;
- at the next site, the conformational change brings ADP and phosphate close enough together to undergo condensation and expel water;
- at the third site, the conformational change causes expulsion of ATP from the site, leaving it free to accept ADP and phosphate at the next part turn.

At any time, one site is binding ADP and phosphate, one is undergoing condensation, and one is expelling

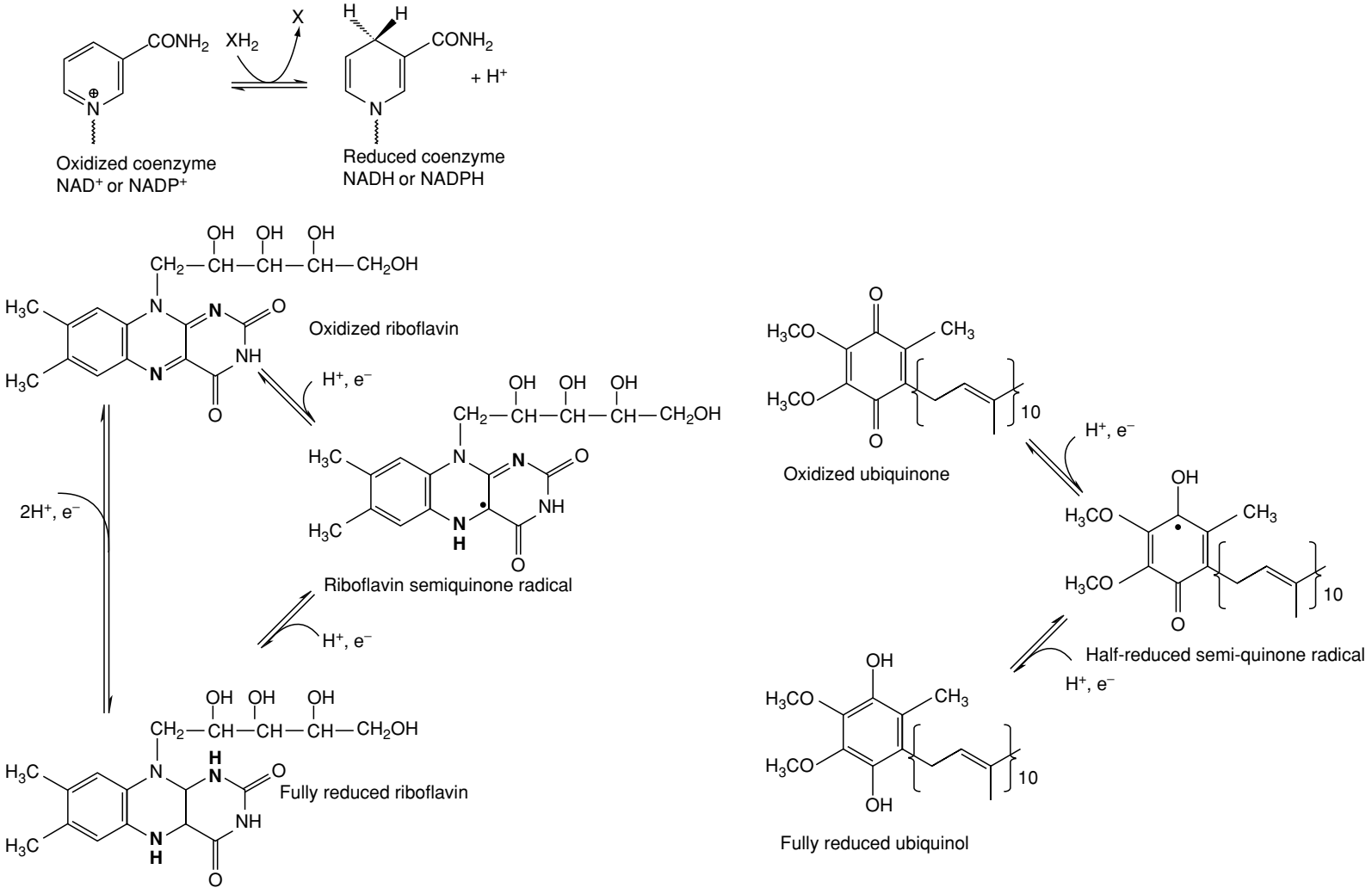

Figure 5 Oxidation and reduction of the hydrogen carriers of the electron transport chain: the nicotinamide nucleotide coenzymes (NAD and NADP), flavins and ubiquinone.

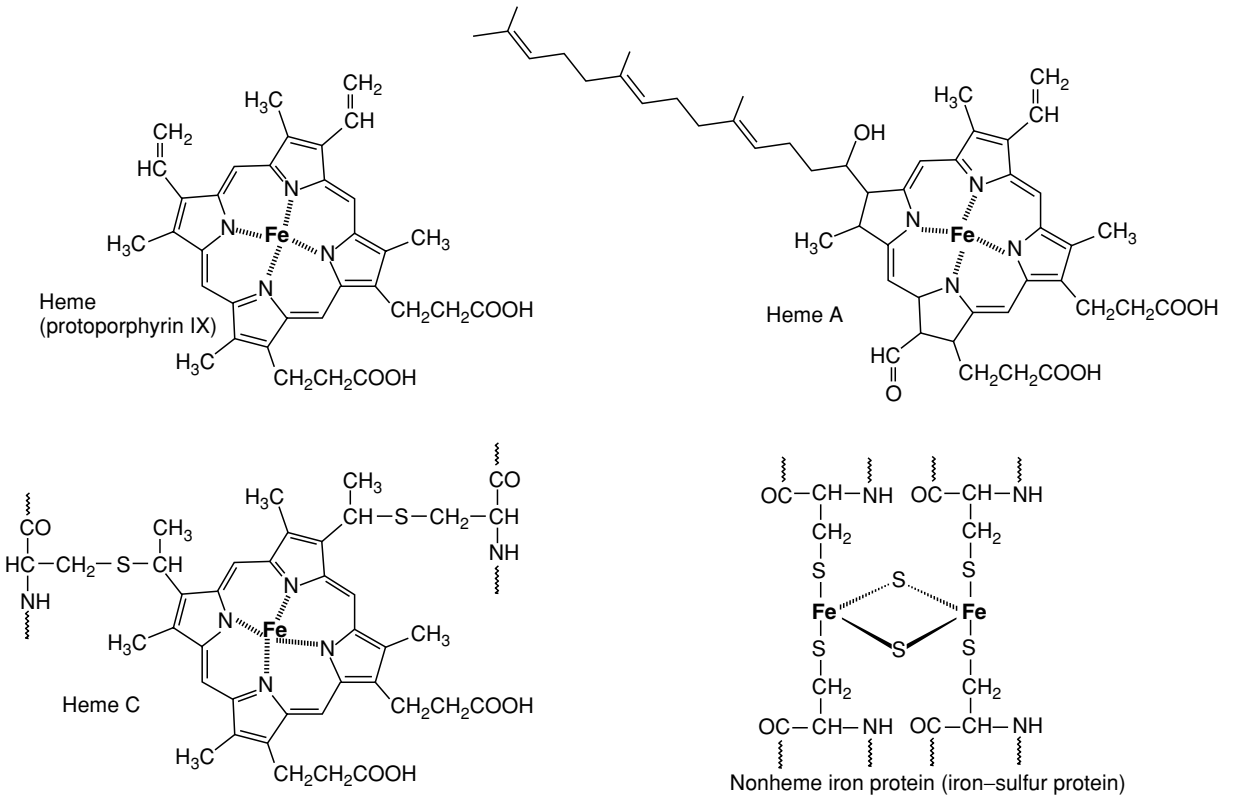

Figure 6 Types of heme in cytochromes and the binding of iron in nonheme iron proteins (iron–sulfur proteins).

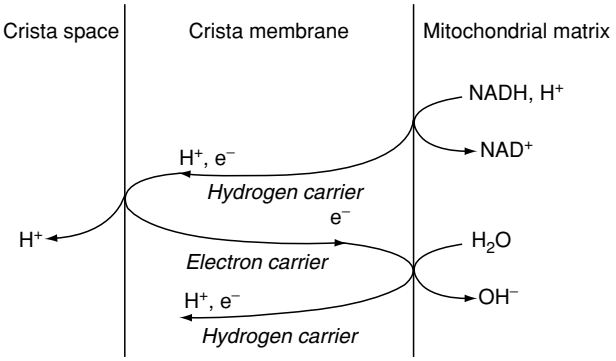

Figure 7 Proton pumping across the crista membrane in complex I.

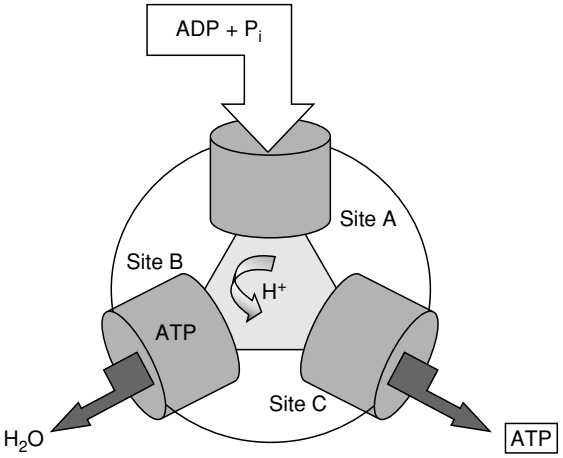

Figure 8 Three active sites of the ATP synthase complex in the mitochondrial primary particle.

ATP. If ADP is not available to bind, rotation cannot occur – and if rotation cannot occur, protons cannot flow through the stalk from the crista space into the matrix.

Coupling of Electron Transport, Oxidative Phosphorylation, and Fuel Oxidation

The processes of oxidation of reduced coenzymes and the phosphorylation of ADP → ATP are normally tightly coupled:

- ADP phosphorylation cannot occur unless there is a proton gradient across the crista membrane resulting from the oxidation of NADH or reduced flavins.
- If there is little or no ADP available, the oxidation of NADH and reduced flavins is inhibited, because the protons cannot cross the stalk of the primary

particle, and so the proton gradient becomes large enough to inhibit further transport of protons into the crista space. Indeed, experimentally, it is possible to force reverse electron transport, and reduction of NAD^+ and flavins by creating a proton gradient across the crista membrane.

Metabolic fuels can only be oxidized when NAD^+ and oxidized flavoproteins are available. Therefore, if there is little or no ADP available in the mitochondria (i.e., it has all been phosphorylated to ATP), there will be an accumulation of reduced coenzymes and hence a slowing of the rate of oxidation of metabolic fuels. This means that substrates are only oxidized when there is a need for the phosphorylation of ADP to ATP, and ADP is available. In turn, the availability of ADP is dependent on the utilization of ATP in performing physical and chemical work.

It is possible to uncouple electron transport and ADP phosphorylation, by adding a weak acid such as dinitrophenol that transports protons across the crista membrane. As shown in Figure 9, in the presence of such an uncoupler, the protons extruded during electron transport do not accumulate in the crista space, but are transported into the mitochondrial matrix, where they react with hydroxyl ions, forming water. Under these conditions, ADP is not phosphorylated to ATP, and the oxidation of NADH and reduced flavins can continue unimpeded until all the available substrate or oxygen has been consumed. Figure 10 shows the oxygen electrode trace in the presence of an uncoupler – there is more or less complete consumption of oxygen regardless of the amount of ADP present.

The result of uncoupling electron transport from the phosphorylation of ADP is that a great deal of substrate is oxidized, with little production of ATP, although heat is produced. This is one of the physiological mechanisms for heat production to maintain body temperature without performing physical work – nonshivering thermogenesis. There are a number of proteins in the mitochondria of various tissues that act as proton transporters across the crista membrane when they are activated.

The first such uncoupling protein to be identified was in brown adipose tissue, and was called thermogenin because of its role in thermogenesis. Brown adipose tissue is anatomically and functionally distinct from the white adipose tissue that is the main site of fat storage in the body. It has a red–brown color because it is especially rich in mitochondria. Brown adipose tissue is especially important in the maintenance of body temperature in infants, but it remains active in adults, although its importance compared with uncoupling proteins in muscle and other tissues is unclear.

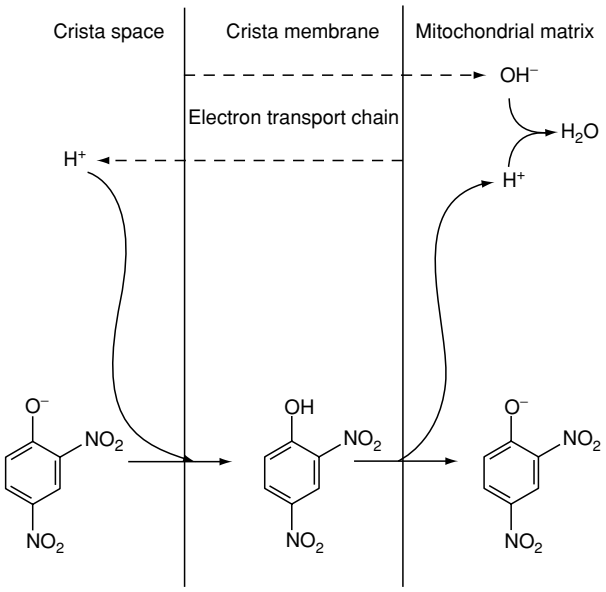

Figure 9 Uncoupling – discharge of the proton gradient by a weak acid.

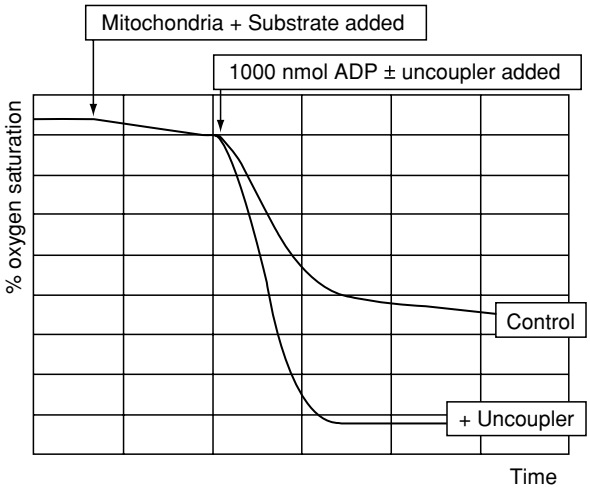

Figure 10 Oxygen electrode traces in the presence and absence of an uncoupler.

In addition to maintenance of body temperature, uncoupling proteins are important in overall energy balance and body weight control. The hormone leptin, secreted by (white) adipose tissue, increases the expression of uncoupling proteins in muscle and adipose tissue, so increasing energy expenditure and the utilization of adipose tissue fat reserves.

Respiratory Poisons and other Inhibitors

Much of our knowledge of the processes involved in electron transport and oxidative phosphorylation has come from studies using inhibitors.

1. Rotenone, the active ingredient of derris powder, an insecticide prepared from the roots of the leguminous plant *Lonchocarpus nicou*. It is an inhibitor of complex I. The same effect is seen in the presence of amytal (amobarbital), a barbiturate sedative drug, which again inhibits complex I. These two compounds inhibit oxidation of malate, which requires complex I, but not succinate, which reduces ubiquinone directly. The addition of an uncoupler has no effect on malate oxidation in the presence of these two inhibitors of electron transport, but results in uncontrolled oxidation of succinate.
2. Antimycin A, an antibiotic produced by *Streptomyces* spp. that is used as a fungicide against fungi that are parasitic on rice. It inhibits complex III, and thus inhibits the oxidation of both malate and succinate, since both require complex III, and the addition of the uncoupler has no effect.
3. Cyanide, azide, and carbon monoxide bind irreversibly to the iron of cytochrome a_3, and thus inhibit complex IV. Again, these compounds inhibit oxidation of both malate and succinate, since both rely on cytochrome oxidase, and again, the addition of the uncoupler has no effect.
4. Oligomycin, a therapeutically useless antibiotic produced by *Streptomyces* spp. Oligomycin inhibits the transport of protons across the stalk of the primary particle. This results in inhibition of oxidation of both malate and succinate, since, if the protons cannot be transported back into the matrix, they will accumulate and inhibit further electron transport. In this case, addition of the uncoupler permits reentry of protons across the crista membrane, and hence uncontrolled oxidation of substrates.
5. Atractyloside (a plant glycoside) and bongkrekic acid (a toxic antibiotic formed by *Pseudomonas cocovenans* growing on coconut – this is named after bongkrek, a mold-fermented coconut product in Indonesia, that becomes highly toxic when *Ps. cocovenans* outgrows the mold). Both compounds inhibit the transport of ADP and ATP across the mitochondrial membrane. Bongkrekic acid fixes nucleotides to the transport protein at the matrix side of the membrane, so that they cannot be released, whereas atractyloside has a higher affinity for the transport protein than does ADP, and so out-competes it for transport into the matrix.

See also: **Adaptation – Nutritional Aspects**; **Adipose Tissue**: Structure and Function of White Adipose Tissue

Further Reading

Bender DA (2002) *Introduction to Nutrition and Metabolism*, 3rd edn. London: Taylor & Francis.

Boyer PD (1997) The ATP synthase – a splendid molecular motor. *Annual Reviews of Biochemistry* 66: 717–749.

Murray RK, Granner DK, Mayes PA and Rodwell VW (2000) *Harper's Biochemistry*, 25th edn. New York: McGraw-Hill.

Oysters *See* **Shellfish**: Characteristics of Crustacea; Commercially Important Crustacea; Characteristics of Molluscs; Commercially Important Molluscs; Contamination and Spoilage of Molluscs and Crustaceans; Aquaculture of Commercially Important Molluscs and Crustaceans

P

PACKAGING

Contents

Packaging of Liquids

S D Deshpande, Central Institute of Agricultural Engineering, Nabi Bagh, Bhopal, India

Background

In modern times, packaging has been identified as an integral part of processing in the food industry. Packaging is a technique of using the most appropriate packaging media for the safe delivery of the contents from the centers of production to the site of consumption. Packaging serves as the vital link in the long line of production, storage, transportation, distribution, and marketing. The package must ensure the same high quality of the product to the consumer, as they are used to getting, in freshly manufactured products. It is important that all products should reach the consumer in a usable condition.

Modern packaging systems for liquid foods are products from a synthesis of demands from producers, distributors, and consumers. The need for hygiene is the primary reason for retail packaging of perishable liquid food products like milk. Although this was realized more than a century ago, packaging techniques for liquid milk were slow in developing. The advent of pasteurization in the 1920s made retail packaging of liquid essential, and the returnable glass bottle was soon to become universal.

The commercial development of plastic materials, starting with polyethylene (PE) in the 1940s, opened up new possibilities for improving hygiene in liquid packaging. PE ultimately became the most frequently used thermoplastic in paper and carton board coating processes, also finding uses in inplant manufacture of packages from reel stock by form–fill–seal techniques. In current efforts to make retailing still more efficient, the focus is on standardized packages and transport wrappings, the aim being to simplify routines and cut costs. One-way cartons are suitable for these requirements. These developments have gradually led to a change in retail patterns in many countries, and the replacement of returnable glass bottles by single-service paper/plastic containers has been seen in many countries.

Today, a product distinction can be made between milk and non-milk on the one hand, and fresh and long-life products on the other. The main products retailed in one-way cartons are still milk and milk products, holding approximately 80% of the carton demand in liquid packaging, but a steady increase in market share for fruit juices, mineral water, sports drinks, vegetable oils and juices, soft drinks, and wine is observed. This trend is likely to continue, ensuring a further potential for the paper bottle.

Functional Packaging

Functional packaging of products contributes to the industrial prosperity of a country through optimal utilization of resources. The packaging has to protect the contents against hazards such as the vagaries of climate and transportation. During the course of the journey, the package would be exposed to varying climatic conditions, often resulting in evaporation and condensation of water of the contents inside the package. Also, atmospheric gases like O_2 and SO_2 contribute to the deterioration of the products by the oxidation of fat-rich products or corrosion of the metal containers. While this is the case in bulk packages, the unit container, which comes directly into contact with the product, must have requisite barriers and protective properties. A scientifically

designed package should, therefore, afford protection against egress or ingress of moisture, flavor loss or odor pickup, light, oxygen, microbial and fungus attack, as well as being compatible with the food packaged. The package must preserve the quality, freshness, and functional performance of the product, afford the requisite shelf-life, and make it possible for the product to reach the consumer in prime condition.

Principles of Production

Filling and sealing machines for paper-based packages for liquids form two groups: those that work from a roll of packaging material, and those that work from premanufactured blanks, the difference reflecting the basic machine philosophy or concepts of companies like Tetra Pak and Elopak. The basic idea of Tetra Pak is that the package should be formed from a roll of packaging material, filled, and sealed in a continuous, closed process. The basic idea of Elopak is that as much of the package production as possible should be included in the converting process. Consequently, the production of blanks, being a highly specialized process, is therefore considered best performed when separated from the food-packaging plant.

Converting

In the converting process, the basic paper is coated on both sides, printed, and provided with scorelines to facilitate creasing when finally forming the package.

Filling and Sealing

When choosing a carton-based packaging system for liquids, there are three differently shaped packages presently predominating the market, namely the gable top, the tetrahedron, and the brick.

The filling and sealing procedure of a gable-top-type package starts with a blank being fed from a magazine. The lay-flat tube is then unfolded and enters a mandrel where the bottom is heated with hot air. The bottom is then folded in accordance with scorelines, and pressure is applied for finishing the bottom sealing. Now an open rectangular box, it is removed from the mandrel on to a conveyer, filled with liquid and the top sealed. The top seal is made using hot air and pressure.

The most striking feature of the tetrahedral package is the shape. The tetrahedral shape requires less packaging material than other designs, as it offers the most favorable ratio of area to volume.

The production of Tetra Brik-type packages from roll-fed machines follows basically the same principles as those for Tetra Standard, but the transverse seams are sealed parallel. The characteristic brick shape is formed after cutting off individual packages from the tube, by folding in the flaps and heat-sealing them.

Materials

The sandwich construction of the two common paper-based laminates used in liquid packaging is shown in Figure 1. If no high-gas barrier is required, the material consists of paper with a polyethylene coating on both sides. The paper layer may consist of unbleached, bleached, or semibleached sulfate pulp or laminates of these. The paper layer, being responsible for much of the machinability and mechanical properties of the package, requires a high and stable quality.

Additional barrier properties are usually provided by aluminum foil, laminated to the board, but the contact surface against the food remains polyethylene.

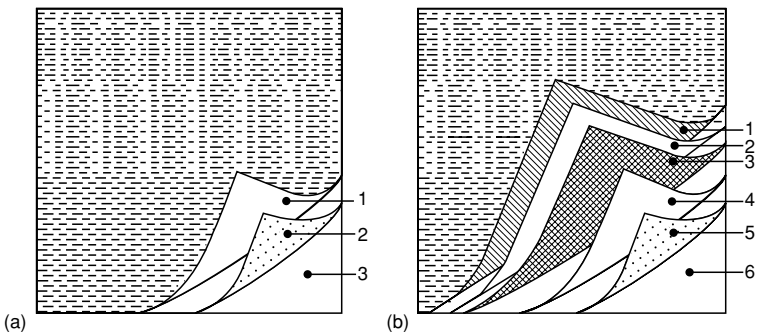

Figure 1 Sandwich construction of two common laminates for carton containers. (a) Typical laminate for short-life products like fresh milk consisting of (1) exterior PE, (2) paper, and (3) interior PE; (b) typical laminate for long-life products consisting of (1) exterior PE, (2) paper, (3) Surlyn, (4) Al-foil, (5) Surlyn, and (6) interior PE.

Packaging materials should provide the following properties:

1. high hygiene standards;
2. sufficient mechanical strength and internal bond;
3. liquid proofness;
4. interness to product;
5. light barrier;
6. low gas permeability;
7. seal ability;
8. machinability in converting and filling processes.

Canning

The process of canning is sometimes called sterilization because the heat treatment of the food eliminates all microorganisms that can spoil the food and those that are harmful to humans, including directly pathogenic bacteria and those that produce lethal toxins. Most commercial canning operations are based on the principle that bacteria destruction increases tenfold for each 10°C (18°F) increase in temperature. Food exposed to high temperatures for only minutes or seconds retains more of its natural flavor. Recent developments include the use of cans made of aluminum, very thin steel, and coated and uncoated plastic. Can openers are unnecessary for cans that have a pullable metal tab or ring attached at the top. Despite the widespread popularity of canned foods, the major limitation of canning is in the quality of the final product. Since food is not a good conductor of heat, excess heat needs to be applied to the container's surface for a period of time to guarantee sufficient heat at the center, or 'cold spot,' in order to destroy all organisms causing spoilage and disease. This method of preserving causes foods to lose juices, texture, flavor, and nutrients. The retort pouch, developed in the 1970s to alleviate this problem, is a 3-layered laminate with flexible plastic films as the outer and inner layers and aluminum foil in the middle. The pouch, which is approximately 19 mm (0.75 in) thick, is filled and sealed under vacuum. Because the pouch has a large surface-to-volume ratio, heat needs to penetrate less than 10 mm (0.38 in) from the surface to the 'cold spot,' thereby yielding greatly improved products.

Selection Criteria for Packaging Materials

Some of the important packaging considerations, which influence the choice of packaging materials, are given below.

Product Protection

The choice of packaging material has to be made depending upon the nature of the product, i.e., susceptibility of the ingredients used to undergo deterioration in contact with water, moisture, gases, light, etc. For products containing high fats, the packaging material must have high oxygen barrier properties. For such applications, material based on nylon, nitrile, or polyester may be considered. However, for products that are likely to deteriorate only by moisture gain or loss, a packaging material with sufficient moisture barrier properties may be adequate. For light-sensitive products, an opaque packaging material may be necessary. The shelf-life is another important criterion that influences packaging material specifications.

Convenience

The convenience for consumers and amenability of the packaging material to provide such features is a predominant factor in the final choice. Features such as reclosability, dispensing capability, stand-up facility, etc. are some examples of packages preferred by the end users.

Sales Appeal and Package Decoration

For some products, it may be necessary for the package to communicate effectively and visually attract the consumer while it is displayed on the shelf. Thus, the choice of packaging material is influenced by such properties as gloss, transparency, and ability to accept multicolor printing.

Product Package Compatibility

This aspect is critical, particularly for edible products. Product package compatibility testing is carried out to assess the possibility of any migration from and to the product. Any loss of product flavor or product tainting with unacceptable odor is also checked.

Packaging Machinery

For products that are mass-produced, the speed of packaging is very important, and the packaging material chosen must run smoothly on the packaging machine used. Here, the mechanical properties of the films and laminates such as tensile strength, rigidity, coefficient of friction, slip properties of the film and laminates, and their susceptibility to develop a static charge need to be considered.

Package Sealing Efficiency

However good a packaging material may be for a given product, its ability to protect the product is determined by its seal strength. Often, the product is spilled around the sealing area while filling. In such cases, sealing of the package with product contamination on the sealing surface is necessary.

Package Strength

Product distribution over longer distances invariably involves multiple handling and long storage periods. These are generally associated with both mechanical and climatic hazards. The choice of material, therefore, has to be made, based on the severity of hazards and the capacity of the material to withstand these in a given distribution situation.

Statutory Requirements

Often, there are certain regulations governing the choice of packaging material, for example, in the packaging of pesticides.

Material Availability

Any user would like to have multiple sources of procuring their packaging material. Therefore, the sources of supply and the availability of material become predominant considerations.

Cost

Sometimes, marketing success depends on how low the product is priced. The ultimate choice of packaging material is often influenced by the cost at which the functions indicated above are achieved.

The criteria discussed above are some of the more important considerations. Depending upon the type of product and the market, many additional considerations may become important. However, by and large, if the above criteria are used while selecting the packaging material, the user is likely to arrive at an optimum choice of packaging material.

Developments in Packaging of Liquid Foods

Liquid food products have been packed, preserved, and transported over the ages by a variety of methods. Wooden containers, earthen pots (glazed pottery or jars), and glass continue to be the major materials, even in the present day. All these had the benefits of being nonreactive to the food material and also efficient to contain the material resisting deformation besides being leak-proof (if properly closed). Over the years, with man's quest for newer technologies, packaging of liquid food products also gained greater importance. The need for developing alternative materials and methods for liquid food packaging has arisen for several reasons, such as to reduce the cost of packaging, improve the shelf-life, provide convenience in using the product, improve the handling during transport or in retail outlets, and also promote consumer acceptance of the product. It was further necessitated due to changed distribution systems, competition, changing consumer needs, and availability of new packaging materials and/or methods.

For transporting in bulk, liquid material can be carried in wooden drums or metal drums. Alternatively, containers can be packed in wooden crated, or fiber mold cases, e.g., glass carboys in crates, polythene, or synthetic plastic carboys can be considered. A new development in this area is the bag-in-box type of packaging for liquids. In examining the shelf-life requirements of the product, microbial spoilage in the case of fresh food products such as milk and fruit juices, due to their high susceptibility to microbial contamination and hence requirement for sterilization/pasteurization, needs to be considered. Packaging material should withstand rigorous treatments during such packaging stages.

As the industry grows, and the demand for convenience increases, the need for quality packaging undoubtedly increases. Today, packaging is cost-effective in the sense that the package is pivotal to the marketing philosophy of the urban markets offering a wide choice at a low cost. Thus, the model food package has to keep down marketing costs and offer the consumer tangible price benefit. The increased speeds on machine lines and use of flexible laminates in place of tin containers may be considered as a development.

As outlined above, the selection of the packaging material and the packaging machines depends upon the food to be packed, and the methods together can be called as a packaging system. Today, fruit juices, wines, water, oil, soup, concentrated vegetable, purées, etc. are packed in a variety of systems such as prepac, Blockpak, Zupack, Purepak, Ceka, Tetrapak, thermoforming systems, and bag-in-box systems. Facilities such as asceptic packaging, hot filling, nitrogen flushing, etc. for the product are also available.

There are the range of cartoning systems and form–fill–seal machines now available and are replacing the glass and tin containers for liquid food packaging. Among these, asceptic packaging is yet another development.

Requirements with respect to hygiene and impermeability are stringent for liquid food products. Sealing must be tight. There are basically three types of machines for the packaging of liquids, viz. vertical form–fill–seal machines horizontal form–fill–seal machines and sachet-forming machines.

All the types can be used for liquid food product packaging. They can have such secondary facilities as gas flushing, vacuum zing, etc. introduced. Standipacks and gusseted sachet packs are quite acceptable for liquid food products for consumer convenience.

Resources and Energy

Presently, one-way carton packages for milk, juices, and other noncarbonated beverages coexist with one-way and returnable alternatives like the glass or plastic bottle. At first sight, the returnable alternatives seem preferable from an energy and ecological point of view. But considering the entire process, from extraction of raw materials, through production and distribution to the handling of waste, this is not the case when compared to one-trip cartons. Investigations show that there are only minor differences, ecologically speaking, between a returnable bottle with a realistic trippage for modern retailing and a plastic-coated paper bottle.

The constituent materials of the dominant packages for the noncarbonated liquid foods, for glass bottles, cans, plastic bottles, and cartons, differ in practically every aspect: raw material consumption, energy consumption, impact on water and air, and amount and quality of waste. The overall energy requirements for the different packages are of the same magnitude (Figure 2), water and air pollutants are comparable, and a large amount of waste is generated in the use of returnable bottles if a bottle does fewer than 20 trips.

Carton-based packages using wood as the basic raw material exploit a renewable resource. Present planting of trees exceeds harvesting, so there will not be any deforestation as a result of the manufacture of paper containers.

Litter and waste may be regarded as minor problems as far as PE-coated cartons are concerned. If a plastic-coated cartonboard is left in nature, complete degradation will usually take several years, the main problem being the degradation of the polymer. Biodegradable PE has been discussed as an alternative outer coating. Although this solution would increase the rate of degradation considerably, it has never found applications. The waste problem is the more important issue when it comes to disposabilities. Nevertheless, the waste from PE-coated cartons is easily handled, does not require separate collection, and may be used as a source of energy when incinerating.

Marketing Aspects and Competition

The main alternatives in liquid food packaging are glass and plastic bottles, cans, and plastic-coated cartons. A packaging system based on cartons is facilitated by low overall costs. In fact, the one-way

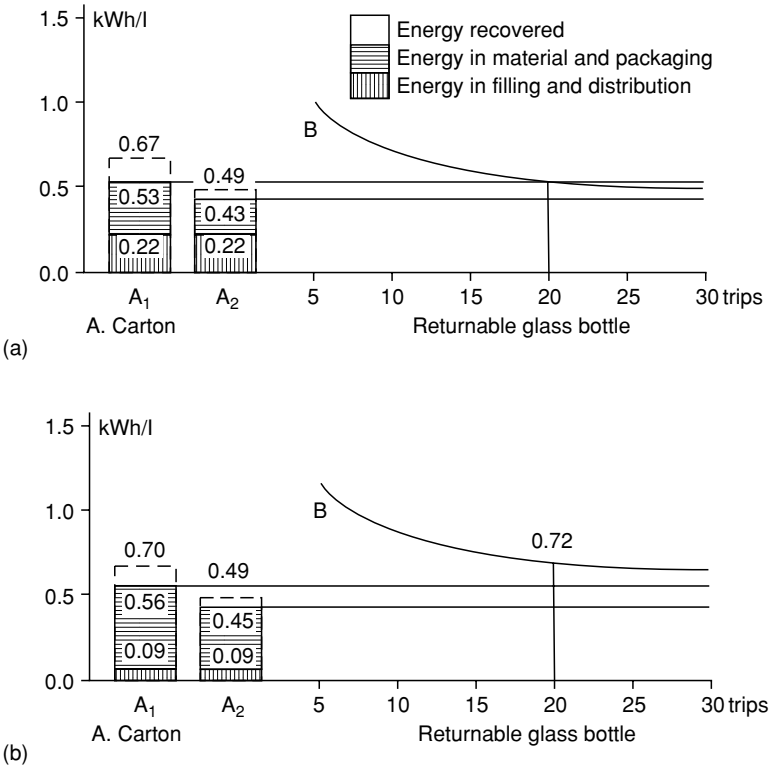

Figure 2 Comparison of energy counts for milk (a) and juice (b) packaged in cartons and returnable glass bottles. A_1 and A_2 represent the energy counts for cartons including and excluding wood energy, respectively. The figures are reproduced from data given in Sundstrom G (1982) *Milk Packages and Energy*. Malmo, Sweden: Sundstrom AB and Sundstrom G and Lundholm MF (1982) *Juice Packages and Energy*. Malmo, Sweden: Sundstrom AB.

carton container is found to be a cheaper alternative than cans and returnable glass bottles. The profitability is largely due to low distribution costs, storage efficiency, maximum use of shelf-space, and low labor costs.

For the working environment, systems based on one-way cartons are regarded as superior. Different products with very different processing qualities can be handled, including mineral water, nectars or fruit juices with a high pulp content, wine, milk, preserved dairy products, and others.

In modern retailing, the package serves the purpose of being an important source of information. Being printable and having four sides available for commercials and content declaration, the carton is its own sale promoter, offering free advertisement displayed on shelves and the consumer's table.

In Western European countries, milk represented somewhat below 80% of liquid carton demand at the beginning of the 1980s, and presently the cartons hold about 60% of the total milk market. There is currently a decline in the total milk market but an increased consumption of cartons for milk, thought to be due to the commercialization of long-life milk and the replacement of returnable bottles in the UK. Cartons have achieved a high penetration of the milk market in most Western countries, so the scope for further progress may be regarded as limited outside the UK. Furthermore, fresh milk (i.e., pasteurized milk) is believed to appeal to popular taste more than sterilized milk, and the market may well turn in favor of fresh milk where proper refrigeration conditions may be achieved.

The second largest product group is for fruit juices, holding about 15% and steadily increasing, and the third largest product group is soft drinks, holding about 5%. Other products for which the cartons are likely to increase in importance are wine, mineral water, vegetable oil, and juices.

Overview

Liquid food packaging has been vital in assuring man of a year-round supply of a variety of foods and has changed us from an agrarian to a cosmopolitan culture.

Significant changes have taken place over the last three decades, and packaging is becoming recognized as a definable and essential technology. It is considered earlier in the food product development cycle and a more organized approach to actual package development is taken.

Currently, package designs are first tested under laboratory conditions (vibration, drop tests, inclined impact) and then confirmed by actual field use, sometimes with little agreement. Improvements in agreement between field experience and laboratory testing will be improved by such developments such as:

1. Shipping hazards recorders, such as that developed by the US Army Natick Development Center (**Figure 3**), to measure and record on tape such events as number and height of drops and environmental conditions such as temperature and relative humidity.
2. In estimating shelf-lives, recognition of diurnal temperature and humidity changes, and the unsteady state nature of oxygen and water vapor permeabilities.
3. In devising more relevant tests to replace archaic techniques, such as the Muller burst test for fiberboard.

Computers can be of great help in the statistical evaluation of the massive environmental and hazards data required, defining any distribution system sufficiently to permit trade-off decisions on product loss versus packaging cost. This will also be essential in handling laboratory test data and will be useful in the graphic mode, for actual package design.

On-line, nondestructive test procedures such as the infrared seal defect technique, the sonic vacuum test technique for cans, and magnetic metal contamination detection procedures will be more plentiful. Energy costs will govern package design to a great degree and will necessitate frequent substitutions for critical materials.

There will be greater emphasis on the nonprotective functions of a packaging system. The flexible

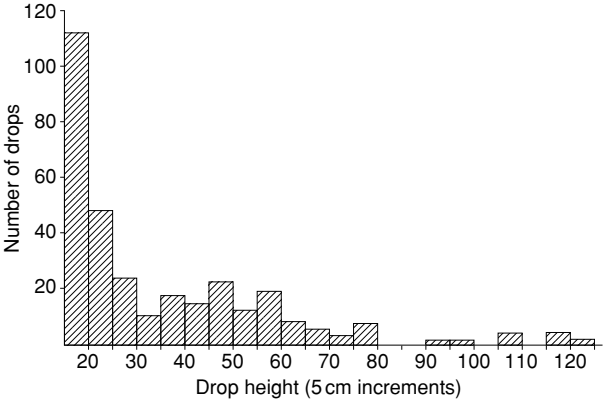

Figure 3 Frequency histogram of recorded drops. Distribution of 312 recorded drops in 5-cm increments. The system activation threshold enables drops of less than 15 cm to be recorded. This represents 1500 days and 80 000 km of shipments. From Barca FD (1975) *Acquisition of Drop Height Data during Package Handling Operations*. Natick, MA: Aeromechanical Engineering Laboratory, US Army Natick Development Center.

packaging system for thermoprocessed foods extends the boil-in-bag heating capability to nonrefrigerated foods. Untended vending systems using microwave energy for in-package product heating appear to be feasible. Liquid foods are being studied with the aim to use the flat configuration to achieve a high product quality and container designs for heating for serving and serving from the container itself.

See also: **Milk**: Liquid Milk for the Consumer; Processing of Liquid Milk; **Pasteurization**: Pasteurization of Liquid Products

Further Reading

Athalye AS (1992) Packaging as a system. In: *Plastics in Packaging*, pp. 15–17. New Delhi: Tata McGraw – Hills.

Barca FD (1975) *Acquisition of Drop Height Data during Package Handling Operations*. Natick, MA: Aeromechanical Engineering Laboratory, US Army Natick Development Center.

Goyal GK (1991) Practice and research related to packaging of indigenous dairy products – A review. *Journal of Indian Dairyman* 43(11): 489.

Gupta TR (2000) Bulk asceptic packaging of fruit pulp and concentrate. 14th Indian Convention of Food Scientists and Technologists Souvenir, pp. 78. Mysore, India: Central Food Technological Research Institute.

Iversen A (1987) Cartons for liquids. In: *Modern Processing, Packaging and Distribution Systems for Food*, pp. 86–90, 105–107. Westport, CT: AVI.

Kumar KR (2000) Global standards for food packaging. 14th Indian Convention of Food Scientists and Technologists Souvenir, pp. 26–32. Mysore, India: Central Food Technological Research Institute.

Lampi RA (1978) Packaging of food – Overview. In: *Encyclopedia of Food Technology and Food Science Series*, vol. 3, pp. 591–597. Westport, CT: AVI.

Menon CPS (1986) Developments in packaging of liquid foods. In: *Packaging of Food Products*, pp. 187–192. Bombay: Indian Institute of Packaging.

Narayanan CK (2000) Trends in health food drinks packaging. 14th Indian Convention of Food Scientists and Technologists Souvenir, pp. 78. Mysore, India: Central Food Technological Research Institute.

Nunn DW (1980) Alternative packaging of milk – a consequence analysis. *CMI Report No*. 790306–1. Fantoft, Norway: Chr. Michelsens Institute.

Raj Baldev (2001) Food and packaging interaction – migration concepts and regulations. *Indian Food Industry* 20(5): 67–74.

Rangarao GCP (2001) Computer simulation of packaging and storage environments and prediction of shelf life of moisture sensitive foods. *Indian Food Industry* 20(6): 65–70.

Sarkar PC and Gupta PC (2002) Shellac-based tin can lacquers for the packaging industry. *Indian Food Packer* 56(2): 69–71.

Sogi DS, Shergill RS and Bawa AS (2000) Effect of packaging materials on storage of lemon juice beverage concentrate. *Journal of Food Science and Technology* 37(3): 296–299.

Sundstrom G (1982) *Milk Packages and Energy*. Malmo, Sweden: Sundstrom AB.

Sundstrom G and Lundholm MP (1982) *Juice Packages and Energy*. Malmo, Sweden: Sundstrom AB.

Sundstrom G and Lundholm MP (1985) *Resource and Environmental Impact of Tetra Brik Aseptic Carton and of Refillable and Nonrefillable Glass Bottles*. Malmo, Sweden: Sundstrom AB.

Packaging of Solids

M Mathlouthi, Université de Reims Champagne-Ardenne, Reims Cedex, France

Background

The aim of packaging solid foods is to protect them against spoilage and to preserve their quality. Selecting a suitable packaging system for this category of products requires a good knowledge not only of the factors affecting their stability but also of the extent to which they are deformed when submitted to mechanical constraints. Both stability and resistance to being deformed are influenced by the water content and water activity (a_w). Most of the physical, biochemical, and microbiological changes of solid foods also depend on a_w. Also, lipid oxidation and other causes of decrease in quality originating from the permeation of gases, water vapor, aroma substances, and other chemicals, or the transfer of energy (light, heat) to an acceptor phase (moisture, fat, etc.) in the product must be minimized through the choice of packaging material, which should be a barrier to matter and or energy transfer. Such a barrier should not be completely tight to take into account a possible release of water vapor from the product in the inner space of the package, which might increase the relative humidity to a limit value of a_w for which the caking of powdery solid foods can occur. In addition, as solid foods are stored and distributed in their packages, one of the objectives of packaging, besides the protection of the product, is to meet the requirements for successful marketing. (*See* **Oxidation of Food Components**; **Water Activity**: Effect on Food Stability.)

Classification

This classification of solid foods is based on the degree to which they are deformed when submitted

to compression or agglomerated and compacted as a result of moisture increase. This depends on whether they are soft, rigid, or powdery products, and may influence the protectivity of the packaging material as well as its acceptance by the consumer. A special class is devoted to food powders for which flowability, density, morphology of particles, and their crystallinity play a major part. Three categories of products are distinguished:

- nondeformable solid foods like toasts, biscuits, crackers, hard cheeses, confectionery, chocolates, etc.;
- deformable products like sandwich loafs, industrial pastry, soft cheeses, meat, and meat products, etc.;
- food powders like refined granulated sugar, wheat flour, spray-dried milk, dairy-based infant formula powders, etc.

The choice of packaging solution depends on the objective sought, namely the nonperception by the consumer of any degradation of the product. As with heat treatment, absolute protection of a packaged foodstuff against alteration is not possible. Thus, protection in this context is only a commercial notion comparable to that of commercial sterilization. It was decided to classify solid foods into deformable and nondeformable products in order to discriminate between those products (deformable) for which the main origin of degradation is the barrier characteristic of the packaging material. In the case of deformation, the appearance of the product becomes the dominant criterion as regards acceptability by the consumer. For nondeformable solid foods, the product may be the origin of a change in structure of the packaging material, leading to a reduction of thickness, pinholes or tears, which modify the transfer of matter or radiation, and provoke biochemical alterations, reduction in shelf-life and, finally, rejection by the consumer. For solid foods, much more than for liquid foods, the concept of an efficient packaging system should be based on the interaction between the product, the packaging material, and the machinery involved.

To generalize for solid foods as a whole, baked goods are taken as a model for this class of food products. Indeed, it is easy to find among baked goods representation examples on nondeformable products like toasts, biscuits, or crackers, and on deformable products like bread or sponge cake. For these products, as for the majority of solid foods, a good criterion of differentiation is the a_w. For rigid baked goods, the a_w is low, about 0.05–0.30, and its value for soft products is around 0.70–0.96. Besides acting as a good reference mark for the aptitude for deforming on which we base our classification, the a_w is associated with most alterations that occur in solid foods.

As a representative product of food powders, white sugar can be taken as an example. Besides the sensitivity to moisture and temperature, the stability of packaged sugar, either granulated or cubes, depends on the stability of crystallization at the surface of the crystals, absence of amorphous or fine particles, and the minimization of the fraction of water content called 'free' water or solvent water on the water vapor sorption curve.

Causes of Alterations

Changes in packaged solid foodstuffs are perceived as being due to a mass or energy transfer across the package providing that there exists a donor (environment, packaging material, etc.) and an acceptor (mobile phase of the solid food). This approach may seem to be complicated, but it has the advantage of relying on classical equations of diffusion. The required characteristics of the packaging material are approached in a dynamic way and deduced from mathematical models of the prediction of shelf-life.

Instead of reviewing all possible matter and energy transfers and their effects on the degradation of solid foods, three major causes of alterations are described: the transfer of water vapor, the transfer of radiation (light), and the transfer of chemicals during storage.

Quality Changes originating from the Transfer of Water Vapor

Nondeformable Products

For this category of foods, an increase in a_w is generally the origin of the alteration and leads to a loss of crispness. However, an increase in moisture may be protective against lipid oxidation in low-moisture foods. In most cases, there remains in the headspace, after sealing the container, enough oxygen to initiate oxidation where an acceptor, e.g., ethylenic bonds in unsaturated fatty acids, exists. It may be noted that, in rigid baked goods with an a_w of 0.30, the risks of biochemical or microbiological alterations are absent.

Deformable Products

It may be recalled that the a_w for these foods is > 0.70, which corresponds to an increase in availability of degradation reaction sites due to an enhanced mobility of water. The risks, in this case, are due to water vapor transfer (gain or loss of weight), chemical and enzymatic reactions, and microbial spoilage. It is also observed that, in this zone of a_w values, a structural rearrangement of one of the constituents (starch) of the product may lead to an alteration known as

staling of the baked goods. (*See* **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage.)

The kinetics of alteration reactions depend on the components of food. This is the case, for example, for the oxidation of lipids. If fats are sequestered by the starchy phase, as in a dry cake (sablé), lipid oxidation does not occur, and the shelf-life of the cake is prolonged. If lipids are also on the surface of the product, as with extruded snacks, lipid oxidation becomes a limiting factor in the quality assurance of packaged solid foods. The availability of water for enzymatic reactions or microbial growth may be limited by the addition of a_w depressors, 'humectants' like sugars, polyols, glycerol, or proteins. To minimize nonenzymatic browning, it is important to reduce the duration of exposure to water vapor pressure at the optimum value ($a_w = 0.65$–0.75) for the Maillard reaction. However, at least for baked goods, the appearance of the product, which is appreciated by the consumer, is partly due to the Maillard reaction. (*See* **Browning**: Nonenzymatic.)

Food Powders

Caking is the major alteration of free-flowing food powders ingredients and results in the transformation of a pulverulent product into lumps, then into agglomerated solid, and finally into a sticky material that has lost its functionality. Such a transformation depends on the temperature and moisture content. Different stages are involved amid which are the bridging of particles, their compaction, and liquefaction. For sugar crystals, the dynamic evolution of a_w and its increase to reach 0.80 can provoke the agglomeration of particles, especially when fine crystals with an average size less than 250 μm are present. When the amount of fine particles increases (above 10%) the a_w value for caking is lowered to around 0.45.

Another origin of instability of food powders is their amorphous structure, which allows their existence below the glass transition temperature (T_g) in an out-of-equilibrium state. It is possible to reach T_g at normal temperatures if the relative humidity is increased. This might be at the origin of a situation where the viscosity is decreased at such a value that allows the product to be transformed into a rubbery material, which sticks to the package and loses its quality.

Alterations Originating from Exposure to Light

Radiation, particularly of ultraviolet light, has a catalytic effect on the quality changes in foods. Deteriorative changes like the destruction of lipid-soluble vitamins, the loss of riboflavin and other vitamins, and changes in proteins and food pigments are accelerated by light. The presence of oxygen is intimately linked to these light catalyzed reactions.

These alterations lead to rancid foods with a changed color and reduced nutritive value. The odor and color changes of the product are obviously perceived by the consumer. In order to limit the rate of light-induced deteriorative reactions, the choice of packaging material for solid nondeformable foods should insure a low oxygen partial pressure and a nontransparent package.

Alterations Originating from Food–Package Interactions

For this group of alterations, the notion of reciprocal mobility of a donor and a receptor acquires all its importance. Indeed, whatever the migration (from the product to the packaging material or from the package to the food), it is essential to have a carrier for the transfer of the migrant chemical. It is also necessary to distinguish between hydrophilic and lipophilic chemicals. For nondeformable products ($a_w \leq 0.30$), the migration of some constituents of the packaging material (low-molecular-weight polyolefins, plasticizers, lubricants, etc.) into the mobile fatty phase of the food may be observed. Likewise, the adsorption of fats by the material (polyethylene, polypropylene, etc.) in contact with solid foods may occur.

For deformable products ($a_w \geq 0.70$), the preponderant mobile phase is water, and only hydrophilic chemicals are implicated in migration, i.e., acrylic derivatives of varnish, glycol derivatives, urea, formaldehyde, etc. In this category of deformable foods, it may be that both the aqueous and fatty phases are mobile, as in some sponge cakes, and both migrations are then observed. The structure and composition of the product help to determine the preponderant mobile phase in contact with the packaging material. Although fat may be fixed to starch and water immobilized by 'humectants,' this category of solid foods is more difficult to control in terms of migration.

Stains of syrup are sometimes observed at the top of paper board package of sugar lumps. This is observed when the lumps are packed at a high temperature (above 40 °C) and not cured in a ventilated room. Such conditions (40 °C and relative humidity above 85%) lead to the diffusion of sucrose in the film of syrup surrounding the crystals. After crystallization of sugar in the film of syrup, and the release of hydration water, the syrup becomes sufficiently mobile to stick to the package. As granulated sugar may act as

an adsorption medium, odors from the environment or other volatiles from the packaging material may be temporarily adsorbed by the sugar.

Preservation

In order to protect a solid food against alteration, it is necessary to place a barrier between it and the environment. This barrier should be adapted to the capacity of adsorption of the food for the factors responsible for the deteriorative reactions. The barrier properties of the packaging material are determined by its permeability to the degradation agents, mainly gas or vapor permeating from outside.

Permeability of Packaging Material

The mass transfer of oxygen, carbon dioxide, water vapor or aroma components, or heat or radiant energy (ultraviolet, infrared, β, γ or microwave radiation) across the packaging material is generally ruled by a relation derived from Fourier or Fick's laws. In the case of unidirectional transfer, under a steady state and at equilibrium, the quantity of permeated material per unit of time (Q) is given by a Fick's first law-type equation:

$$Q = -PA\frac{\Delta p}{l}, \tag{1}$$

where P is the permeability, A is the active surface area of the packaging material, Δp is the difference in pressure (or concentration) on either side of packaging material, and l is its thickness.

The permeability coefficient, P, is a function of the number of variables, which include the structure of the packaging film, the properties of the permeate, the time, the pressure, and the thickness (the temperature is constant). Depending on the sensitivity of the product to water vapor, oxygen, or loss of aroma components, the packaging material composition and barrier characteristics are determined (see Tables 1 and 2).

Prediction of Shelf-life

The classification of solid foods determines the choice of the model transfer equation and affects the prediction of the deteriorative reaction rate. (*See* **Storage Stability**: Shelf-life Testing.)

For nondeformable foods, the most critical transfer is that of water vapor and/or oxygen. The kinetics of oxidative degradation should be adapted to each category of product, while the model equation used for water vapor transfer may be generalized.

For deformable foods with $a_w > 0.70$, water vapor transfer generally occurs from the inside to the outside of the package and is the preponderant factor of alteration. In some cases, more than one transfer has to be taken into account, for instance, water vapor and oxygen. The mathematical model of mass transfer is then obviously more complicated.

In order to calculate the shelf-life of products sensitive to moisture, eqn (2) is applied:

$$\frac{dW}{dt} = -\frac{PA(p_e - p_i)}{l}, \tag{2}$$

where dW/dt is the flow of moisture, and p_e and p_i are, respectively, the water vapor pressure in the exterior and the interior of the package. If storage takes place at a constant temperature, p_e and p_i may be replaced by $a_{w,e}$ and $a_{w,i}$, the values of outer and inner water activities, and eqn (2) can be written in integral form as:

$$t = K \int_{\text{origin}}^{\text{degradation}} f(aw)\, daw, \tag{3}$$

where t is the shelf-life of the product, and K is a constant related to the properties of the product, the permeability of the packaging material, and the conditions of storage. Eqn (3) is only valid when the degradation originates from a change in a_w due to moisture transfer. In the case of oxidative deterioration, two equations are applied, one accounting for oxygen transfer and the other for its adsorption by food. For transfer, the relation is:

$$\frac{d(O_2)}{dt} = P_{O_2} A \frac{\Delta P}{l} m_{o_2}, \tag{4}$$

where $d(O_2)/dt$ is the flow of oxygen into the package, P_{O_2} is the permeability, A is the surface of permeation, Δp is the pressure difference of oxygen between the interior and exterior of the package, m_{o_2} is the absorption capacity of food, and l, as before, is the thickness of the packaging film. It is generally supposed that oxygen is adsorbed by food, and the rate of adsorption of oxygen $d(O_2)/dt$ is given by:

$$\frac{d(O_2)}{dt} = \frac{P_{O2}}{k_1 + k_2 P_{O2}} \tag{5}$$

where P_{O_2} is the partial pressure of oxygen, and k_1 and k_2 are two constants related to the degree to which the food adsorbs oxygen. An increase in the rate of adsorption of oxygen provokes an increase of oxidation. However, many other factors affect this type of degradation (light, heavy metals, water activity, etc.), and it is difficult to find mathematical model accounting for shelf-life as regards oxidation.

Other degradations may be predicted from mathematical models, like nonenzymatic browning on model systems (reducing sugars + amino acid) in the a_w range 0.55–0.85, but a precise prediction of a real degradation in a real product is not possible.

Increase in Shelf-life

When the degradation reactions are known, it is possible to minimize their rate in order to increase the shelf-life of the product. The basis of this approach is the need to reduce the availability of the donor or acceptor of the degradation agent.

If the acceptor is water, its availability is reduced by the addition of humectants (polyols, sugars, proteins) to the formula of the baked goods. Such an addition provokes a decrease in a_w from about 0.85 to 0.75 for a sponge cake, for example, which is then screened from microbial spoilage.

Preservation against oxidation and/or microbial growth may be obtained by changing the inner atmosphere of the package, generally by using a mixture of nitrogen and carbon dioxide. A low partial pressure of oxygen, together with a carbon dioxide/nitrogen atmosphere, contributes to an increase in the shelf-life; for example, industrial pastry with a normal a_w level of about 0.90 can have its shelf-life extended to 3 months. (*See* **Chilled Storage**: Use of Modified-atmosphere Packaging; Packaging Under Vacuum; **Chill Foods**: Effect of Modified-atmosphere Packaging on Food Quality.)

The stabilization of a solid food may also be achieved by using a superficial thermal treatment using radiant energy from infrared or microwave radiation. Such a treatment applied to the packaged product can contribute to an increase in the shelf-life.

Protection of the product against degradation is also obtained by the choice of a good barrier (polymeric film or laminate), which may be transparent or opaque. The barrier properties of the packaging material change if it interacts with degradation agents like water vapor or oxygen. These interactions depend on the hydrophilic or hydrophobic character of the material. For products particularly sensitive to one or other of the degradation agents, the packaging material may include chemicals permitting the scavenging of oxygen, the adsorption of water vapor, the emission of ethanol or carbon dioxide, and so on. The packaging becomes an 'active' barrier of protection and the center of a protective chemical reaction. However, one should remember that adaptation of the packaging material to food protection must be economically feasable and that the objective is not the absolute stability of the product during storage. It is only needed for the solid food to be accepted by the consumer.

Polymeric Packaging Materials used for Solid Foods

The selection of polymeric or laminate film for solid-food packaging involves a certain number of conditions. The material must be compatible with the food in accordance with the legislation and calculated (thickness) to fit with determined barrier properties and shelf-life of the product. Depending on the level of protection and kind of product to be packed, the packaging material will be either a monolayer or a multilayer (coextruded or laminated). It may be coated or metallized, if the aim is to prevent the transfer of water vapor or light, and the structure of a foil reinforced by orientation or plasticizing.

A wide range of films used for baked goods are listed in Table 1. The transfer of gases and water vapor is expressed in commonly used units rather than in SI units. The unit for the transmission rate of gases is given in cubic centimeters of gas in 1 day for 1 square meter of film at a given thickness at constant temperature and pressure. The rate of water vapor transmission is expressed in grams per day per square meter at a specific temperature and a given difference of water vapor pressure (Δp) between the two sides of the packaging material. When the barrier characteristics required are not achievable with monolayer film, laminates of multilayer material are used. The characteristics of such laminates are listed in Table 2. As may be seen in Tables 1 and 2, the basic films are polyolefins, polyethylene, and polypropylene, copolymers like ethylene-vinyl alcohol and other polymers such as polyesters, polyamides (OPA 6 and 66) and polyvinylchlorides. The laminated plastic films are coated with regenerated cellulosics like DM, which is nitrocellulose coated on one side, MXXT, which is polyvinylidene chloride coated, or paperboard or aluminum foils. Different laminating adhesives are used.

Most foods are sensitive to water vapor, and the packaging materials generally used are good barriers to water vapor; polyesters (polyvinyldene chloride, coated or metallized film) and oriented polypropylene are the major materials found. Snacks and crisps contain fats and have a large specific area. Therefore, they require protection against light, which has a catalytic effect on lipid oxidation. The laminates used to protect these products utilize aluminum foils (see Table 2). Metallized polyester is generally coated with vaporized aluminum.

Food powders freely flowing are packaged in polyethylene sealable films, which gives a good water

Table 1 Characteristics of some films used in the packaging of solid foods

Film (thickness 25 μm)	*Gas transmission ($cm^{-3}\ m^{-2}\ day^{-1}$ (dry gas))*			*Water vapor transmission ($g\ m^{-2}\ day^{-1}$)*	
	Oxygen 23 °C	*Carbon dioxide 23 °C*	*Nitrogen 23 °C*	*ΔRH 90% 38 °C*	*ΔRH 75% 25 °C*
LDPE (0.917)	7 400.00	40 000.00	2 800.00	12.50	4.00
HDPE (0.960)	1 600.00	11 400.00	440.00	3.70	1.45
PP-cast	3 040.00	9 760.00	690.00	8.20	3.30
OPP-coextruded	1 550.00	5 280.00	320.00	5.00	1.35
OPP-coated	15.00	88.55	4.50	5.00	2.00
OPP acrylic-coated	1 200.00	4 500.00	250.00	4.60	1.80
OPP-metallized	35.00	108.00	6.50	1.00	
PVC-rigid	120.00	320.00	20.00	32.00	12.00
PVC-oriented	27.00	68.00	20.00	17.50	7.00
PVC-plasticized	190–3 100	430–19 000	53–810	85.00	32.70
PVDC	1.25–14.5	5.0–50.0	0.4–2.5	0.6–3.20	0.25
PS-cast	4 500.00	11 000.00	640.00	170.00	70.00
SAN	900.00	2 800.00	120.00		
Polycarbonate	3 200.00	17 500.00	450.00	178.00	72.50
PET	55.00	240.00	12.40	20.00	7.00
PET–PVDC-coated	8.00	32.00	2.00	8.50	3.40
PET-metallized	0.65	3.4–10	0.20	1.00	0.40
PA6	40.00	200.00		280.00	80–110
OPA6	18.00	120.00	9.00	130.00	28.30
PA 6.6	35.00	140.00	11.00	90.00	15.00–30.00
EVAL (32% ethylene)	0.16	0.45		80.00	32.00
Cellulosic fim 445MXXT A	8.75	80.00	3.65	8.60	3.40

LDPE, low-density polyethylene; HDPE, high-density polyethylene; PP, polypropylene; PVC, polyvinyl chloride; PVDC, polyvinylidene chloride; PS, polystyrene; SAN, styrene acrylonitrile; PET, polyester; PA, polyamide; OPA, oriented polyamide; EVAL, ethyl-vinyl alcohol; MXXT, a PVDC coating.

Table 2 Characteristics of some laminates used in the packaging of solid foods

Laminate	*Gas transmission ($cm^{-3}\ m^{-2}\ day^{-1}$ (dry gas))*			*Water vapor transmission ($g\ m^{-2}\ day^{-1}$)*	
	Oxygen 23 °C	*Carbon dioxide 23 °C*	*Nitrogen 23 °C*	*ΔRH 90% 38 °C*	*ΔRH 75% 25 °C*
Cellulose film 280 XS + PE 40 μm	12.00			4.50	1.10
OPP coextruded 25 μm + OP coextruded 25 μm	650.00			2.60	0.95
PET coated PVDC 12 μm + PE 40 μm	5.00	15.00	1.00	3.70	1.40
M PET 12 μm + PE 80 μm	1.00	4.00	0.20	0.50	0.20
M PET 12 μm + M PET 12 μm + PE 80 μm	< 0.10	< 0.10	0.00	0.15	0.06
OPA6 15 μm end. PVCD + PE 60 μm	10.00	30.00	2.50	5.00	
OPA 6 20 μm + PE 80 μm	40–30				
M OPA 6 15 μm + PE 80 μm	2.00			2.50	
Kraft 45 g m^{-2} + PE 20 g m^{-2} + end. PVDC 20 g m^{-2}	34.00			1.70	0.06
Kraft 60 g m^{-2} + end. PVDC 30 g m^{-2}	15.00			1.90	0.65
PET 12 μm + 119 μm + monomer 20 μm	< 0.20			< 0.10	
Cellulosic film 320 DM + A 19 μm + PE 35 μm	7.15			0.15	0.10
Kraft 70 g m^{-2} + PE 15 g m^{-2} + A 19 μm	4.30			0.10	0.08
A 19 μm + Kraft 70 g m^{-2}	25.40			0.25	0.15
A 19 μm + TPP 20 g m^{-2} wax	28.00			0.40	0.15
30 g m^{-2} TPP 20 g m^{-2}	< 8.00	< 35.00	< 3.00		

M, metallized; Kraft, paper; DM, one-size nitrocellulose coated; A, aluminum foil; XS, cellulosic film coated with PVDC.

vapor barrier and a transparent package allowing the product to be seen. Refined sugar is generally packed in multiwall paper packages, cardboard cartons, or polyethylene coated paper. Finished packs contain 500 g to 5 kg for domestic use. Sacks of up to 50 kg are produced for industrial utilization. An intermediate stage between the bulk transport of granulated sugar and the 50-kg sacks is the big bag called

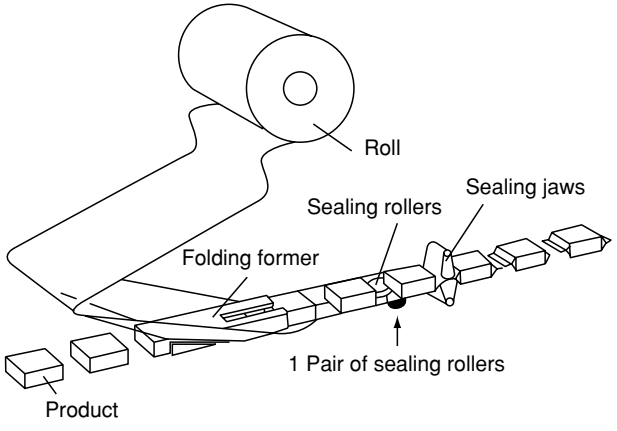

(a)

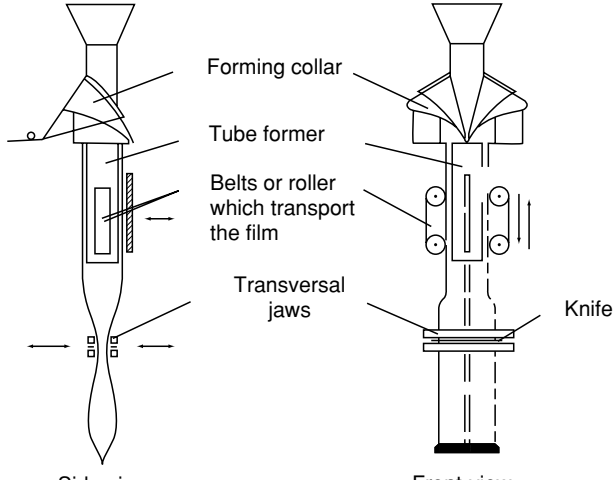

(b)

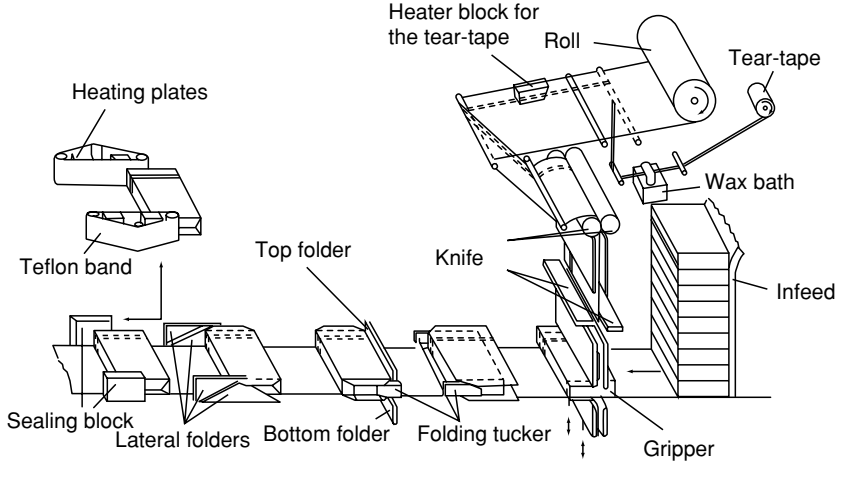

(c)

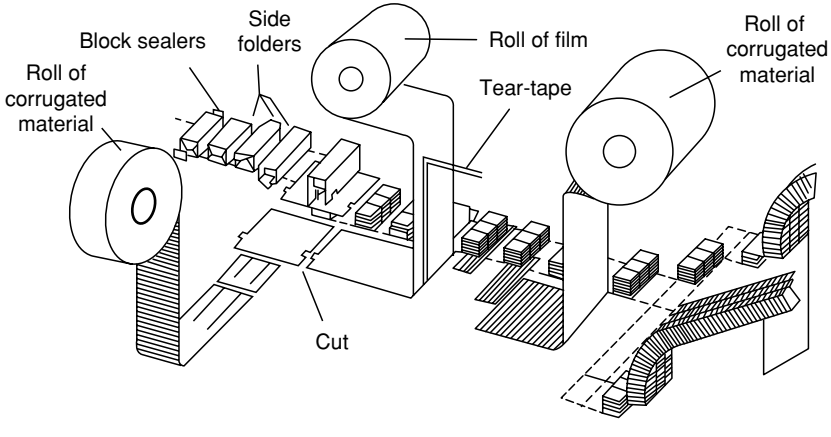

(d)

Figure 1 (a) Horizontal form–fill–seal machine; (b) vertical machine ('pillow bag'); (c) overwrapping machine; (d) overwrapping machine (with adjustment to the circumference of the product).

intermediate bulk container of 1 tonne capacity. This supersack is made of woven polypropylene and it is able to stand unsupported on the pallet.

Packaging Machines

A good packaging solution for solid foods should not ignore packaging machines and the complex interactions between the machine and the packaging material. (Figure 1) shows an example of a horizontal form–fill–seal machine, a vertical pouch form–fill–seal machine, and two overwrapping machines.

The suitability of a packaging machine for a certain packaging material and product can be summarized in the following key points:

- The angle at which the wrapping material comes into contact with the folding former is important, as it may cause a mark or even a cut in the material. The tightness of the packaging film depends on the shape of the former.
- The quality of heat sealing of the packaging material depends on the temperature control, pressure, and dwell time. Optimal results are obtained when the dwell time at the melting temperature of the coating or copolymer is long enough to ensure tightness without damage to the basic material.
- The wrapping material should slip easily after sealing at high temperature. In some cases, it is even necessary to cool the plate fixed above the sealing rollers.
- The quality of cutting depends on the knife, cut angle, position of the knife in relation to the sealing jaws, etc.

Overwrapping also requires compatibility between the machine and the wrapping material, and prevention of static electricity, poor slip under hot conditions, temperature control disruption, and rolling of the film. The packaging material should be sealable on both sides, the sealing area should be sufficient, and the structure (soft or rigid) of the solid food should not affect the shape and sealing surface if a good gas tightness is desired.

Packaging machines for food powders, and especially sugar, cover a large range of sizes of packages, starting with small flat pouches containing between 4 and 10 g of sugar. These are produced using simple machines that assemble sachets by bringing together two webs of polyethylene-coated paper, heat sealing three sides, filling with sugar, and sealing the top. There are also carton packing machines, especially designed for sugar cubes. For ease of handling, 'brick-packs' are preferred. These are produced by forming a polyethylene pack and dropping it into a horizontal solid upstanding pack with similar characteristics to the paper packet.

The quality of packaged food powders should comply with the necessary regulations. Any foreign matter should be eliminated. All packages go through metal detectors to control metal contamination. Likewise, traceability of packets and even full pallets is needed for identification of the product according to EU regulations. A printed code showing the date and line of production, factory of origin, etc. is required. The spread of use of scanning at the shop point of sale imposes the printing of a bar code on each domestic package.

See also: **Browning**: Nonenzymatic; **Chilled Storage**: Use of Modified-atmosphere Packaging; Packaging Under Vacuum; **Chill Foods**: Effect of Modified-atmosphere Packaging on Food Quality; **Oxidation of Food Components**; **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage; **Storage Stability**: Shelf-life Testing; **Water Activity**: Effect on Food Stability

Further Reading

Gary JI, Harte BR and Miltz J (eds) (1987) *Food–Product Package Compatibility*, Lancaster, PA: Technomic.

Hotchkiss JH (ed.) (1988) *Food and Packaging Interactions, American Chemical Society Series*. Washington, DC: American Chemical Society.

Mathlouthi M (ed.) (1986) *Food Packaging and Preservation, Theory and Practice*. London: Elsevier.

Rockland LB and Beuchat LR (eds) (1987) *Water Activity. Theory and Application to Food*. New York: Marcel Dekker.

Rogé B and Mathlouthi M (2000) Caking of sucrose crystals: effect of water content and crystal size. *Zuckerindustrie* 125: 5.

Van der Poel PW, Schiveck H and Schwartz T (eds) (1998) *Sugar Technology, Beet and Cane Sugar Manufacture*. Berlin: Dr Albert Bartens Verlag.

Versanyi I (1985) *Food Packaging Technique [H]*. Budapest: Mezögazdasagi Kiado.

Aseptic Filling

L Mauer, Purdue University, West Lafayette, IN, USA

Background

Aseptic processing is a high-temperature–short-time thermal process to commercially sterilize a product and fill the cooled sterile product into a presterilized package in a sterile environment. Purposes for aseptic

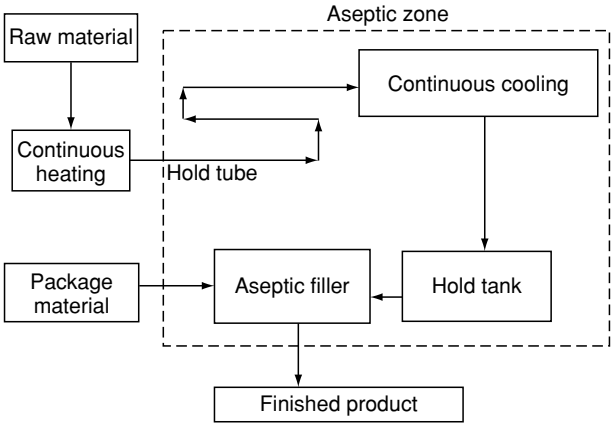

Figure 1 Aseptic process diagram.

processing include extending the storage life of food products, optimizing product quality, and reducing cost. A diagram of an aseptic processing system for consumer products is shown in Figure 1. In the aseptic process, the aseptic filler is designed to sterilize the package material, fill the sterile product into the package in a sterile environment, and then hermetically seal the package. Aseptic filling differs from other traditional methods of food packaging in that the food product and the package are continuously sterilized separately and then meet in the sterile environment provided by the aseptic filler. Important factors in the aseptic filling process are the type of product, type of package, obtaining and maintaining a sterile environment for filling, and the sealing process.

Types of Product

A variety of food products are aseptically packaged. Examples include milk and dairy products, fruit and vegetable juices, fruit juice concentrates, sport and dietetic beverages, tomato products, edible oils, puddings, soups, cheese and soy sauces, and products containing small or large particles of fruits, vegetables, or potatoes. The aseptic process must be designed so that the product is sterilized, the package is compatible with the product composition and storage needs, and the type of filler treats the product gently while maintaining sterility to achieve optimum quality. Products are usually sterilized by high-temperature heat treatments for short periods of time, cooled, and conveyed to the filler via aseptic pumps or nitrogen. Packaging and filling considerations for different product types include the following:

Low-acid Products (pH > 4.5)

The concept of aseptic heat treatments for low vs. high acid foods is similar to the canning principles for these food types. In low-acid foods, such as many milk and vegetable products, pH is not a barrier to pathogenic microbial growth. Therefore, aseptic processing conditions must be designed with stringent controls to achieve a 10-decimal reduction value for *Clostridium botulinum*, the most heat-resistant pathogen able to grow in low-acid foods, both in the food product and in the package materials. Targeting the most resistant pathogen ensures safety because all other pathogens are killed faster than it. The package must attain the same level of sterility as the product; therefore, package materials for low-acid foods may have stricter sterility controls than those for high acid foods. *Bacillus subtilis* is the most resistant organism to hydrogen peroxide sterilization of packages, and a four-decimal requirement for it exists to achieve the required reduction for *C. botulinum* in low acid foods.

Acid and High-acid Products (pH < 4.5)

In acid and high-acid foods, such as fruit juices and many tomato products, pH is a barrier to pathogenic microbial growth, most notably for *C. botulinum*. For these foods, spoilage microorganisms may cause more problems than pathogens. Processing temperatures and holding times may be lower than for low acid foods. Spoilage organisms are much more sensitive to hydrogen peroxide sterilization of packages than *B. subtilis*. *Aspergillus niger* is usually the target organism for dry and moist heat sterilization of packaging materials for use with acid and high-acid products.

Homogeneous vs. Heterogeneous Products

Knowledge of heat-transfer characteristics of a food, quality parameters, flow characteristics, mean residence time, type of heat exchanger or sterilizing system, and heat resistance and required death values for target microorganisms are important factors in designing an effective aseptic process. Homogeneous products contain no particles that disrupt heat transfer. Heterogeneous products contain particles with sufficient size to create a thermal gradient during processing. Therefore, heat treatments for homogeneous liquids may be less damaging than those required to insure commercial sterility of particulates in similar liquids. Pumps used to convey aseptically processed products must be able to maintain both product sterility and integrity; therefore, pumps used for heterogeneous products must limit shear forces so as not to damage particulate structures. Homogeneous liquids are often conveyed to a filler with either a centrifugal pump or sterile nitrogen; however, highly viscous liquids may require use of a positive displacement pump. Heterogeneous products are conveyed to the filler using a positive displacement pump that limits pressure and shear placed on

the particulates. Opening sizes in an aseptic system designed for heterogeneous products must be sufficient to allow passage of the particulates without creating pinch points. In the filler, contamination of the seal area, especially with particulates, must be avoided.

Types of Packages

The type of aseptic package used must be suitable for a product's requirements, package surfaces must withstand sterilization by heat, irradiation, and/or chemicals prior to filling, and the package must contain, protect, and preserve the food product throughout its distribution and shelf-life. Properly aseptically processed products remain microbiologically stable as long as the package remains intact. Often, shelf-life and product quality are limited more by package performance than any other factor. Needs to consider for an aseptic packaging system include:

- Compatibility of product and package – product composition, needs for maintaining quality (barrier, mechanical), and shelf-life requirements.
- Packaging material – type (plastic, metal, glass, laminate, etc.), barrier and mechanical properties, machineability, recyclability, and cost.
- Package form – size, shape, compatible machinery, sealing properties, appeal to consumer, communication of information, and cost.
- Sterilization method (heat, irradiation, chemicals, combination of methods) – efficiency, throughput, residues, compatibility with product/package/environment, worker safety, regulations, and cost.
- Filling equipment – reliability, efficiency, capacity, and cost of installation/operation/maintenance.

Aseptic Package Materials

A significant drive in the conversion from traditional to aseptic processing is the reduction in packaging costs for both materials and transportation. Traditional thermal processes require packages to withstand high temperatures, whether for in-package sterilization or for hot-filled products, as well as vacuum forces created on cooling. Packages that meet these structural requirements include metal, glass, and plastics (rigid, semirigid, some with vacuum panels incorporated into the design, and pouches). The cold aseptic filling allows for a lighter container with a greater design flexibility (squeezable, ergonomic, etc.). By weight, most plastic materials cost significantly less than glass and metals, and less plastic is needed for a cold-filled product than a hot-filled product. Plastics are the most common material used for aseptic packaging; however, plastics are more permeable to oxygen and moisture than either glass or metals. The quality of products may suffer after extended storage in permeable packages. An example of this is oxidation of dairy products in packages permeable to oxygen. To limit permeability, plastics with different barrier characteristics are often combined using lamination or coextrusion methods to optimize barrier properties. This can maximize product quality and package function while maintaining low package costs.

The variety of plastic materials used in aseptic packaging is continuously increasing, along with the use of laminates, coextrusions, and copolymers. Adhesive and thermal lamination methods are used to create multilayer packages, often with a structure based on the diagram in Figure 2. More than one barrier layer may be incorporated into a package design. Commonly used plastics include polyolefins (polyethylene, polypropylene, polystyrene), polyesters (polyethylene terephthalate), vinyl plastics (polyvinyl chloride, polyvinylidene chloride, polyvinyl alcohol), polyamides (nylon), ionomers, acrylics, fluorocarbons, and polycarbonates. Polyethylene (PE) is the most widely used polymer in packages and is available in a variety of densities: low density (LDPE), linear low density (LLDPE), medium density (MDPE), and high density (HDPE). LDPE is used in flexible bags and pouches and is a good heat-sealing material. HDPE used in blow-molded bottles is more rigid and has better barrier properties than LDPE. Polypropylene (PP) provides clarity, stiffness, and heat resistance. Polyethylene terephthalate (PET) has good barrier and clarity properties and a high tensile strength, and is resistant to high temperatures. PET is often used to meet consumer demand for round and recloseable plastic bottles. Ethylene vinyl alcohol

Figure 2 General structure of a laminate aseptic package.

(EVOH) and polyvinylidene chloride (PVDC) are generally the best plastic barriers to moisture and oxygen migration. Metal and glass are used for barrier properties either alone or within a laminate package. A paperboard layer in a laminate package is used for printing and light-barrier purposes.

Package Formation

Packages formed for the aseptic process require minimum levels of contaminating microorganisms prior to the sterilization process, and the packages must pass leak-testing procedures. Packages may be preformed or formed in the filler just prior to filling. Metals are formed into cans, and metal films may be incorporated into laminate structures. Glass is blow-molded into bottles and jars. Plastics and laminated materials are blow-molded or thermoformed into various package shapes. The blow-molding process involves the melting of a glass or plastic, forming a parison (tube), and then blowing the parison into the package shape (designated by the mold) using sterile air or nitrogen. Extrusion, injection, and stretch blow-molding methods are used for different package applications. A sheet of plastic is heat-softened then molded to shape via vacuum, pressure, or matched mold forming in the thermoforming process. Cups and trays are thermoformed packages.

Aseptic Package Systems

Package forms and filler types are interwoven to the extent that discussion of one is not complete without discussion of the other, as the aseptic process hinges on the placing of a sterile product into a sterile package in the sterile filler environment. Common aseptic package forms are discussed in this section, leaving classification of filler types for a later section.

Cartons The brick-style, paperboard packaging used for cartons is recognized as the traditional aseptic package used for juice boxes. The laminate material used for carton formation generally consists of multiple polyethylene, paper, and aluminum foil layers. Outer and inner polyethylene layers provide protection and sealing properties, the paperboard is printed with product information and provides stiffness, and the aluminum foil contributes gas- and light-barrier functions. Cartons are prefabricated, or rolls of the laminate material are sterilized and formed in the filler just prior to filling and sealing. Cartons are sterilized using hydrogen peroxide and hot air.

Bottles The convenience of plastic bottles that are clear, round, resealable, recyclable, and able to be formed in a variety of shapes and sizes is a driving force in the beverage industry. Aseptic milk, creamer, sports, and nutritional beverages in PET, PE, and PP bottles are currently available to consumers. Bottles for use with high-acid products are generally sterilized using steam, a combination of heated air and steam, or heat of formation (extrusion and blow molding). Bottles for low-acid products are sterilized with hydrogen peroxide.

Cups The thermoformed cup is a common aseptic package for puddings, particulates, fruits, purees, and baby food. Package materials for cups include PS, PP, and laminates of PS, PVDC/EVOH, and LDPE materials. Cup lids are often aluminum foil coated with LDPE or other heat-sealant plastic. Cups are generally sterilized using hydrogen peroxide or saturated steam under pressure.

Pouches Pouches for institutional and fast food chain use are easier to open, take up less space for transportation, storage, and waste, and decrease product loss when compared with traditional metal cans. Form-fill-seal systems are used for converting a sheet film of plastic (LDPE, laminate) into a pouch. Plastics formed as lay-flat tubes require fewer seals to form pouches than the sheet films. Aseptic products available in pouches include tomato-based sauces, ice cream mixes, cheese sauces, and puddings. Pouches are often sterilized using hydrogen peroxide.

Bag-in-box The bag-in-box aseptic packaging system is used for bulk packaging (1–10 000 or more liters). The product is filled into a presterilized plastic bag. The bag is placed into a protective container, such as a corrugated box or metal drum, either before or after filling. A laminate of barrier (PVDC, EVOH) and heat sealant (LDPE) plastics is commonly used for bag formation. Tomato, fruit juice, and dairy products are packaged using the bag-in-box system. Preformed bags are sterilized by ionizing gamma irradiation. The exterior of bags for low-acid products is also sterilized with steam or hydrogen peroxide vapor.

Metal cans The metal can traditionally used in thermal processes also has advantages for use with aseptic products despite the added weight over plastic materials. The durability, consumer acceptance, barrier properties, and recyclability of metal cans combined with the ability to sterilize the cans with superheated steam instead of chemicals make the metal can attractive for some applications. Aseptic dietetic beverages, pudding, and cheese sauces are available in metal cans. Superheated steam is used to sterilize metal cans prior to filling.

The Sterile Environment

Obtaining a Sterile Environment

Establishing, maintaining, and validating sterility in an aseptic system are essential. Processes used for obtaining a sterile environment for products, equipment, and packages include thermal, chemical, irradiation, and mechanical treatments along with combinations of these. Thermal processes include saturated steam, superheated steam, hot air, mixtures of hot air and steam, and extrusion/heat of formation (although extrusion may not be acceptable for sterilization purposes). Heat is the most common method for product sterilization and is often used for equipment sterilization as well. The most common chemical used for sterilization is hydrogen peroxide (20–35% concentration), and a combination of hydrogen peroxide treatment followed by hot air is often used to reduce the level of hydrogen peroxide to less than 0.5 p.p.m. on packages and equipment. Peracetic acid and ethylene oxide also have applications; however, regulations for acceptable chemicals, uses, and residue limits must be checked. Chemicals are applied by dipping, spraying, or rinsing, or in vapor form. Irradiation treatments may include ultraviolet radiation, infrared radiation, and ionizing gamma radiation. Mechanical processes are generally designed to reduce initial microbial load to a level the aseptic system is designed to handle. If an initial microbial load is too high, sterility might not be achieved during aseptic processing. Therefore, water rinsing or flushing, air blasting, brushing, and ultrasound methods are used to reduce the initial load on equipment and packaging prior to other sterilization processes.

Equipment and packages must attain the same level of sterility as the products they will come into contact with. Therefore, equipment is sterilized with hydrogen peroxide or a time/temperature steam sterilization at least equivalent to what the product will receive. Thermal processes for equipment used with low-acid foods will be higher, both time and temperature, than for high-acid products. Packages, other than metal cans, may not withstand high temperature treatments; therefore, chemical and irradiation sterilization procedures are commonly used.

Maintaining a Sterile Environment

Once equipment, packages, and products are sterilized, the challenge is to maintain sterility during the filling and sealing operation. This is accomplished using either an overpressure of sterile air or continuous flow of superheated steam, the latter generally being used for the metal can aseptic process. Air is sterilized by filtration, usually using a series of high-efficiency particulate air (HEPA) filters. Since all equipment must attain commercial sterility, the HEPA air filters also must be sterilized. The positive pressure maintained by a continuous sterile air flow prevents contamination during the filling process when the product may be exposed to air. If the integrity of an aseptic zone is compromised because of line stoppages or other occurrence, it is necessary to repeat the initial sterilization process.

Validating Sterility

Validating sterility is a combination of implementing an appropriate hazard analysis critical control point program, documenting product and filler sterilization procedures, filing the process with the appropriate agency, maintaining accurate temperature, time, pressure recording devices, and testing all aseptically processed foods for sterility using appropriate microbiological sampling techniques. Challenge testing with inoculated foods is part of validating a thermal process.

The Sealing Process

At the end of the filling process, packages are often sealed using heated sealing bars, jaws, and plates. The temperature, pressure, and time of sealing must be strictly controlled for each sealing compound to maximize package integrity. Common heat sealant polymers include polyethylene, polypropylene, ethylene vinyl acetate, and polyvinyl chloride. These compounds have demonstrated a desirable mechanical strength of the seal (impact strength, tensile strength, tear strength, seal strength, puncture resistance) and product holding properties (chemical resistance, product compatibility, oxygen and moisture barrier properties). Problems occur when particulates, fibers, or other food contaminants are caught in the seal area and prevent a completely fused seal. To avoid seal contamination, especially with particulate foods, the filler must be designed to avoid or remove product from the seal area prior to fusion.

Types of Fillers

The filling method depends on both the type of product (high or low acid, homogeneous liquid, viscous product, size of particulates, etc.) and type of package (retail, institutional, cup, pouch, bottle, etc.) desired for the end product. Common filler classifications are named for the processes that occur in the filler and are outlined below.

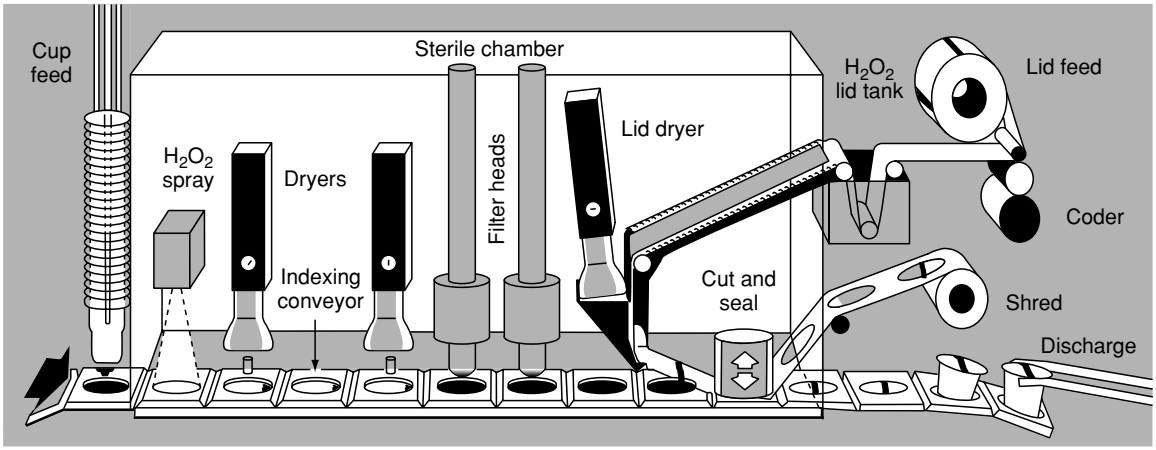

Figure 3 Aseptic filler design.

Fill and Seal

In the fill and seal process, a preformed sterile container (bottles, cups, irradiated pouches, bag-in-box) is aseptically filled and sealed. The fill and seal process for cups is shown in Figure 3. In bulk packaging systems (500-liter bins, large totes, million gallon storage tanks), product is filled directly into a sterilized large storage container. This system is often used for seasonal food products to extend the supply throughout the year.

Form, Fill, and Seal

Brick pack juice boxes and pouches (LDPE, laminate) enter the filler as a roll of packaging material and are sterilized, aseptically formed into the box or pouch shape, filled, and thermally sealed in the form, fill, and seal process.

Thermoform, Fill, and Seal

In the thermoform, fill, and seal process, the packaging material (PS, PP, laminates) enters the filler as roll stock and is sterilized, heated, thermoformed (usually into cups), filled, and thermally sealed with presterilized lid material (aluminum foil coated with LDPE).

Blow Mold, Fill, and Seal

Plastic bottles are formed, filled, and sealed in the blow mold, fill, and seal filling process. An extrudable material (PET, PP, PE) is blow-molded into a container that is filled and sealed in place before the mold opens. For low-acid products, a chemical sterilization after the bottles are formed is used to insure sterility prior to filling.

Advantages and Disadvantages of Aseptic Filling

Advantages of aseptic filling include a high product quality, decreased storage space for packaging materials, decreased weight of packaging materials, decreased package costs, decreased shipping costs for packaging materials, reduced energy consumption, increased flexibility for package options, increased convenience for easy open containers, feasibility of long-term bulk storage, and increased shelf-life and storage time for products. Disadvantages include the high cost of aseptic fillers, potential for product contamination in the aseptic filler owing to the complexity of the system, decreased recyclability of pouch and laminate package materials compared with metal and glass packages, and slower line speeds than traditional process methods. For many, the advantages far outweigh the disadvantages.

See also: **Canning**: Principles; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; **Packaging**: Packaging of Liquids; **Pasteurization**: Principles; **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability; Shelf-life Testing

Further Reading

Chambers JV and Nelson PE (ed.) (1993) *Principles of Aseptic Processing and Packaging*. Washington, DC: The Food Processors Institute.

David JRD, Graves RH and Carlson VR (eds) (1996) *Aseptic Processing and Packaging of Food: A Food Industry Perspective*. New York: CRC Press.

Floros JD (1993) Aseptic packaging technology. Chambers JV and Nelson PE (eds). In: *Principles of Aseptic Processing and Packaging*, pp. 115–148. Washington, DC: The Food Processors Institute.

Holdsworth SD (ed.) (1992) *Aseptic Processing and Packaging of Food Products*. London: Elsevier Applied Science.

Moruzzi G, Garthright WE and Floros JD (2000) Aseptic packaging machine pre-sterilisation and package sterilization: statistical aspects of microbiological validation. *Food Control* 11(1): 57–66.

Reuter H (ed.) (1993) *Aseptic Processing of Foods*. Lancaster, PA: Technomic.

Robertson GL (ed.) (1993) *Food Packaging Principles and Practice*. New York: Marcel Dekker.

Sommers CH, Thayer DW, Pauli G *et al*. (2000) *New Processing and Packaging Technologies*. Washington, DC: National Food Processors Association.

Willhoft EMA (ed.) (1993) *Aseptic Processing and Packaging of Particulate Foods*. London: Blackie Academic & Professional.

Packaging Materials *See* **Chill Foods**: Effect of Modified-atmosphere Packaging on Food Quality; **Chilled Storage**: Use of Modified-atmosphere Packaging; Packaging Under Vacuum; **Packaging**: Packaging of Liquids; Packaging of Solids; Aseptic Filling

PALM KERNEL OIL

K G Berger, Chiswick, London, UK

Introduction

Palm kernel oil is obtained from the oil-rich seed of the oil palm (*Elaeis guineensis Jackqu*). A description of the oil palm and its development into a major crop is given in the entry on **Palm Oil**. The production process of the kernels in the oil mill is also described there.

The quantity of palm kernel oil produced is on average 12% that of palm oil. World production in 1998 was 2.9 million tons.

Production of the Oil

The palm kernels leave the oil mill at about 50% oil content and up to 8% moisture. Further processing typically consists of:

1. Comminution in hammer mills to approximately 2-mm pieces.
2. Cooking to 110–120 °C in a continuous cooker. This breaks the oil cells and reduces moisture to 2.5%.
3. Pressing in a screw press at high pressure to yield crude palm kernel oil and a meal containing 7–9% oil. The meal is used for animal feed.

An alternative process is to submit the broken nuts after step (1) to a cold pressing at low pressure, leaving a meal containing 10–12% oil. This may be further treated by solvent extraction, depending on the market prices of oil and meal.

Refining

Palm kernel oil may be refined either by an alkali or by the physical process, as described in the entry on **Palm Oil**. However, because most of the fatty acids present are of 12 carbon atoms or less, the deodorization temperatures used are lower – typically 220 °C in the alkali process and 230–235 °C in the physical process.

Further Processing

Palm kernel oil is treated by fractionation or hydrogenation and/or interesterification to obtain products with closely defined functionality. These processes are described in the entry on **Palm Oil**.)

Chemical Composition and Physical Properties

The fatty acid composition of palm kernel products is given in **Table 1**, while **Table 2** gives the glyceride composition by carbon number. Typical values for the minor components of palm kernel oil are given in **Table 3**.

Palm kernel oil has a high content of lauric and shorter-chain fatty acids and a low content of unsaturated acids. Coconut oil has a somewhat similar

composition. They contrast strongly with the other common vegetable oils, which contain no shorter-chain fatty acids. In consequence the physical properties are unusual, as will be seen from Tables 4 and 5. Palm kernel oil has a relatively high solid fat content and a rather hard structure at 20 °C, but melts sharply at 28 °C – well below mouth temperature. These properties are accentuated in palm kernel stearin. As is clear from the last two columns of Table 5, hydrogenation raises the solids content further, but the products are still substantially molten at mouth temperature. Hydrogenated palm kernel olein, on the other hand, having a higher content of long-chain saturated acids, has a higher melting point. The products shown in Table 5 are examples of the properties that can be obtained by further processing. A wide range of proprietary products based on palm kernel oil is available, giving variations in physical properties required for specific food applications.

The data shown in Tables 1–4 are average figures taken from an extensive survey of Malaysian palm oil. The figures are not significantly different from those recently obtained for oils from other sources in other producing regions.

Table 1 Fatty acid composition % of palm kernel oil products

Fatty acid	*Palm kernel oil*	*Palm kernel olein*	*Palm kernel stearin*
6:0	0.3	0.4	0.2
8:0	4.2	5.4	1.2–3.5
10:0	3.7	3.9	2.4–3.6
12:0	48.7	41.5	55.6–58.6
14:0	15.6	11.8	18.1–24.7
16:0	7.5	8.4	7.1–7.9
18:0	1.8	2.4	1.5–1.8
18:1	15.0	22.8	2.6–8.8
18:2	2.6	3.3	0.2–1.5

Data from Palm Oil Research Institute of Malaysia, with permission.

Table 2 Triglyceride composition of palm kernel oil products: carbon number by gas–liquid chromatography

Carbon number	*Palm kernel oil (%)*	*Palm kernel olein (%)*	*Palm kernel stearin (%)*
28	0.5	0.3	0.1
30	1.3	1.6	0.5
32	6.4	7.8	3.3
34	8.4	9.3	6.5
36	21.0	18.3	27.5
38	15.6	12.5	24.8
40	9.5	7.6	15.2
42	9.0	9.3	9.2
44	6.8	7.8	5.2
46	5.3	6.5	3.0
48	6.2	8.3	2.4
50	2.5	3.1	0.9
52	2.7	3.4	0.7
54	3.3	4.2	0.8

Data from Palm Oil Research Institute of Malaysia, with permission.

Table 3 Minor components of palm kernel oil typical values

	Refined palm kernel oil (mg kg^{-1})
Carotenoids	Less than 8
Tocopherols and tocotrienols	Less than 33
Sterols[a]	875
Triterpene alcohols	470

[a]Main components β-sitosterol (68%), stigmasterol (14%), and campesterol (9%).

Table 4 Physical properties of refined palm kernel oils

Solid fat content % by NMR at:	*Palm kernel oil*	*Palm kernel olein*	*Palm kernel stearin*
5 °C	72.8	65.6	93.2
10 °C	67.6	56.9	91.6
15 °C	55.7	40.4	90.1
20 °C	40.1	20.9	82.8
25 °C	17.1	1.4	68.2
30 °C			34.6
Slip melting point °C	27.3	23.6	32.2

NMR, nuclear magnetic resonance.
Data from Palm Oil Research Institute of Malaysia, with permission.

Food Uses of Palm Kernel Products

The most important applications of palm kernel oil products are in the confectionery industry, where its high solid fat content and sharp melting characteristics are important. However, palm kernel oil itself melts at too low a temperature for some applications. It is therefore hydrogenated to give a variety of products, one of which is indicated in Table 5. Hydrogenated palm kernel oil is suitable for chocolate-type couvertures for biscuits and sugar confectionery, and for biscuit cream fillings. Hydrogenated palm kernel oil is also used to replace butterfat in filled milk, coffee creamers, and imitation cream. The higher-melting-point grades tend to leave a waxy residue on the palate. They can be improved by fractionating to remove the highest melting point components. A superior and more expensive product is obtained by the direct fractionation of palm kernel oil. The stearin has good contraction when it solidifies and is therefore suitable for molded chocolate. (*See* **Cocoa**: Production, Products, and Use.)

The palm kernel olein produced is the lower-value fraction. It may be hydrogenated (Table 5) to give a range of confectionery fats of somewhat lower quality. Alternatively, palm kernel olein is used in margarine blends. It is a useful component of

Table 5 Solid fat content of further processed palm kernel oil products

	Partly hydrogenated palm kernel oil	*Partly hydrogenated palm kernel olein*	*Fully hydrogenated palm kernel olein*	*Same interesterified*	*Palm kernel stearin*	
					Partly hydrogenated	*Fully hydrogenated*
Melting point (°C)	34	38.6	44.5	35.6	34.5	37.5
Solid fat content % at:						
10 °C	94	85	90	82		
15 °C	92	77	82	74		
20 °C	84	61	72	60	90	90
25 °C	59	39	64	45	83	84
30 °C	27	20	38	25	38	38
35 °C	7	7	22	6	15	17
37 °C		7	18	3	3	4
40 °C	0.5	3	13		2	3

Data from Palm Oil Research Institute of Malaysia, with permission.

Table 6 Interesterified blends using palm kernel oil products

Oil	*%*	*Solid fat content % at:*					*Application*
		20 °C	*25 °C*	*30 °C*	*35 °C*	*40 °C*	
Palm kernel olein	75	53		15	3	0	Biscuit filling cream
Palm stearin	25						
Palm kernel olein	75}	46	24	9			Chocolate soft center
Cotton seed oil stearin	25}[a]						
Same blend	[b]	66	40	23			Couverture
Hardened palm kernel oil	70}	80	63	43	21	3	Whipping cream
Palm stearin	30}						

[a]After interesterification, hydrogenated to iodine value 17.
[b]After interesterification, hydrogenated to iodine value 8.

interesterified mixtures for various applications (Table 6).

Palm kernel oil without modification is used for chocolate coatings for icecream bars, usually in a blend with liquid oil or palm oil to give the right consistency without excessive brittleness.

Palm Kernel Products in Margarine

1. Palm kernel oil forms a eutectic with palm oil in a mixture containing about 30% palm kernel oil. This feature is used to improve the mouth feel (rapid melting) in the following formula:
   Palm oil 63%
   Palm stearin 7%
   Palm kernel oil 30%
2. Palm kernel olein (30 parts) interesterified with palm stearin (70 parts) forms a margarine stock. To make margarine, 60 parts are blended with 40 parts liquid oil.
3. Equal quantities of fully hydrogenated palm kernel oil and fully hydrogenated palm oil are interesterified. A margarine high in polyunsaturated fatty acids is made by blending 12% of this hard stock with 88% of a liquid oil such as sunflower oil.

See also: **Cocoa**: Production, Products, and Use; **Palm Oil**

Further Reading

Applewhite TH (1994) *Proceedings of World Conference on Lauric Oils*. Champaign, Ill: American Oil Chemists Society.

Baldwin AR (1985) Proceedings of World Conference Kuala Lumpur, Malaysia. Processing of palm, palm kernel and coconut oils. *Journal of American Oil Chemists Society* 62.

Siew WL and Berger KG (1981) *Malaysian Palm Kernel Oil Chemical and Physical Characteristics. Porim Technology 6*. Kuala Lumpur: Palm Oil Research Institute of Malaysia.

Tang TS, Chong CL and Yusoff MSA (1995) *Malaysian Palm Kernel Stearin, Palm Kernel Olein and their Hydrogenated Products. PORIM Technology 16*. Kuala Lumpur: Palm Oil Research Institute of Malaysia.

PALM OIL

K G Berger, Chiswick, London, UK

Introduction

Palm oil is obtained from the fruit flesh of the oil palm (*Elaeis guineensis* Jacqu.), a native of the equatorial region of West Africa. Unrefined oil obtained from wild palms has been a traditional food source for the indigenous population for thousands of years. However, its use for edible purposes elsewhere had to await the development of suitable refining processes to produce a bland pale-colored oil. Today, palm oil is the second most abundant edible oil after soya bean oil and is in universal use.

The Oil Palm

The oil palm flourishes best in lowland regions of high rainfall and close to the equator. Optimum conditions are a rainfall of 1700 mm or more per annum, evenly distributed through the year and a position within 10° N and S of the equator, but it also grows in isolated locations outside these limits, i.e., between 15° N and 20° S. The plant can be grown in a variety of tropical soils, regular and sufficient water being apparently more important than soil, provided that the plant nutrient requirements are supplied. Unfavorable soil factors are poor drainage and high laterite or sand content. Special techniques have been developed to enable peat soils to be used.

The oil palm is a single-stemmed plant bearing a number of fronds in a simple head. A mature palm may have up to 50 fully opened fronds. One inflorescence arises from the axil of each leaf; male and female inflorescences occur on the same palm. Both consist of a central stem carrying about 200 flower bearing spikelets. Each spikelet on the male inflorescences carries about 1000 flowers, and on the female inflorescence 15–30 flowers. Pollination is by insects, the most effective being the weevil, *Elaeidobius kamerunicus*, native to West Africa. Ripe fruit develops in about 155 days after fertilization, but the bulk of the oil is synthesized in the final 2–4 weeks. The fruit bunch, containing 1500–2000 individual fruits weighs 20–30 kg.

Considerable variation occurs in the wild palms, and three types have been classified according to fruit type. The most common wild type is the dura, characterized by a relatively thin layer of flesh covering the seed, which consists of a hard thick shell within which lies the kernel. A small proportion of plants, called pisifera, bear fruit with a thick layer of flesh and a small kernel with a very thin or no shell. Another small proportion of plants in the wild, the tenera, has relatively thick flesh with a shell and kernel of intermediate size.

The tenera type was found to be a natural cross (D×P) of dura and pisifera palms. When plantations were first being developed in West Africa (in Belgian Congo, now Zaire), the superior economic performance of the tenera was recognized, and the heritability of shell thickness was discovered. Subsequent development of improved planting material therefore centered on selecting the best dura and pisifera plants for crossing. The continuing efforts of plant breeders have resulted in progressive increases in yield, as shown in Table 1.

A slow development of oil palm plantations occurred in Indonesia after World War I, soon to be followed by Malaya. By 1938, Indonesia had 90 000 ha and Malaya 30 000 ha under oil palm. After 1960, the area under oil palm increased rapidly in Malaysia, reaching 3.2 million ha by 2001. Developments in Indonesia were slower, with 2.4 million ha in 2001. These two countries are by far the largest producers of palm oil and the main suppliers to the world market. Plantation developments in Central and South America have reached about 500 000 ha in all, with Colombia, Ecuador, and Brazil having the largest areas.

A unique feature of the oil palm is that the fruit yields two distinct types of oil. Palm oil from the flesh is the major product, whereas palm kernel oil is

Table 1 Progress through breeding and selection in South-east Asia

Material	*Bunches (tonnes ha^{-1})*	*Mesocarp to fruit (%)*	*Oil to bunch (%)*	*Oil (tonnes ha^{-1})*
Bogor 1878	16.5	58.7	17.6	2.8
Elmina 1933	20.1	58.2	17.0	3.4
OPRS 1969	24.8	64.1	18.3	4.5
Commercial DXP 1978	26.0	80.1	22.0	5.7

Adapted from the Palm Oil Research Institute of Malaysia Bulletin, November 1984 with permission.

obtained from the seed. Its chemical and physical characteristics are quite distinct from palm oil, and it has different applications. (*See* **Palm Kernel Oil**.)

Commercial Importance

The importance of palm oil in the world's oils and fats economy is shown in Table 2, giving the total production of 17 major oils and fats, and individual figures for soya, palm, palm kernel, and rapeseed oils. The figures given are 5-year averages and include production forecasts to the year 2002. Figures in brackets are the percentage of total. Production figures for sunflower oil (not shown in Table 2) are similar to those of rapeseed oil, and that the other oils among the 17 included in the total show little or no growth trend.

Table 2 shows that palm and palm kernel oils together have increased their share of edible oil supplies regularly for more than 20 years, and are forecast to continue to do so. World production of palm oil in 2001 was 23.6 million tonnes of which 17.6 million tonnes were exported to world-wide destinations. Eighty-eight percent of the exports were of Malaysian and Indonesian origin.

In those African countries where the oil palm grows, it continues to hold its position as the major traditional food oil with an annual production of about 1.5 million tonnes in the region. While much of it is still eaten in the crude form, the consumer is also increasingly demanding processed products.

In a number of Latin American countries, palm oil is seen as an efficient means of reaching self-sufficiency, and current production of about 1.4 million tonnes is mainly for local consumption. In contrast, domestic consumption in Malaysia is less than 10% of production. The crop is mainly grown for export, as a successful diversification from rubber. Indonesia, the second largest producer, requires a higher proportion of its production for internal consumption, but is increasing production and exports rapidly.

The importance of palm oil in the world market is based on several factors.

1. It is comparatively cheap, often with an appreciable discount to soya bean oil, the market leader.
2. It has technical attributes, useful in food manufacture, especially its good stability to oxidation, and its natural solid fat content.
3. It is a perennial crop, planted for an economic life of 25–30 years. It mainly grows in equable climates in regions little affected by earthquakes or hurricanes, so that production fluctuates less from year to year than does that of annual crops.

Harvesting and Processing of the Fruit

Various traditional primitive methods of oil recovery are still widely used in West Africa. Typically, the fruit bunch is cut off, and allowed to ferment a few days, so that the fruit detaches easily. The fruit flesh is softened by boiling or by further fermentation, then mashed in a pestle and mortar or under foot. Hot water is added, the oil skimmed, and the water is then boiled off in a separate container. The oil is used without any refining.

In contrast, the modern plantation is usually associated with a mechanical oil mill with a processing capacity of 5–60 tonnes of fresh fruit bunches (FFB) per hour. The harvesting of FFB is still done manually, using an ax, or a sickle-type blade fixed to a pole, depending on the height of the palm. Ripeness is judged by the number of detached fruit that have fallen below the bunch. The frond and then the bunch are cut off. The bunch and any loose fruit are picked up and carried manually, by barrow, in a light animal-drawn vehicle, or with a light mechanical vehicle to the nearest estate road. It is important not to compact the soil close to the palms. Transport is then by lorry to the oil mill. Bunches are emptied into a chute, and filled into sterilizer 'cages' mounted on

Table 2 World production of oils and fats (× 1000 tonnes)

	1968–1972	*1973–1977*	*1978–1982*	*1983–1987*	*1988–1992*	*1993–1997*	*1998–2002*
Total of 17 oils and fats	40 314	45 883	56 848	67 614	80 663	92 314	105 266
Soya bean oil	6 036	8 477	12 639	14 147	16 997	19 508	23 317
	(15)	(18.4)	(22.1)	(20.8)	(21.1)	(21.1)	(21.9)
Palm oil	1 725	2 763	4 482	6 745	10 101	13 878	17 498
	(4.3)	(6.0)	(7.9)	(10.0)	(12.5)	(14.9)	(16.6)
Palm kernel oil	406	437	545	862	1 232	1 655	2 175
	(1.0)	(1.0)	(1.0)	(1.3)	(1.5)	(1.8)	(2.1)
Rapeseed oil	2 039	2 572	3 732	6 009	8 329	9 534	10 829
	(5.0)	(5.6)	(6.6)	(8.9)	(10.3)	(10.3)	(10.3)

Adapted from Oil World Annual. ISTA Mielke, Hanburg, with permission.

wheels, which are then moved into a horizontal cylindrical pressure vessel on rails. Typically, up to 7 × 2.5 tonnes of fruit are cooked in one load. Steam at 304 kPa (3 atm) is applied for about 1 h. The contents of the cage are then fed to a 'bunch stripper' consisting of a horizontal rotating drum with baffles. Bunches are repeatedly lifted and dropped, so that the fruit is shaken out. It is then transported to a 'digester', a vertical steam-jacketed cylinder fitted with rotating beater arms. The fruit is thoroughly mashed and passes directly to a single or a twin-screw press. The press extrudes a liquid at one end, consisting of about 53% oil, 7% finely divided solids, and 40% water, and a press cake containing the fruit fiber and the nuts at the other end. The liquid phase passes through a vibrating screen to a settling tank. After about 2 h, the upper layer is clarified in a sealed centrifuge; the oil is then dried under vacuum and pumped to storage. The lower layer is treated in a sludge centrifuge, or in a three-phase decanter, the oil phase being returned to the settling tank.

The cake formed by the press fiber and nuts is processed in a separate stream. The cake is broken up by rotating arms on a conveyor, and the fiber is removed in a 'polishing' drum. The nuts are partially dried by a warm air stream in a silo, before being cracked in a centrifugal cracker. Nuts drop on to a rotor, which throws them against a peripheral ring of hardened metal. Kernels are separated from shells in a pneumatic column and/or a hydrocyclone, washed, and dried with hot air to below 8% moisture.

The production of palm kernel is usually carried out in a separate factory.

Treatment of Wastes

Three aqueous waste streams arise in the oil mill:

1. condensate from the sterilizing process;
2. aqueous effluent from the centrifuges;
3. waste water from nut processing.

Typically, for every 1 tonne of oil produced, the combined aqueous waste is 2.5 tonnes containing 0.6% oil and 3.9% dissolved and suspended solids. After removal of any supernatant oil, this effluent is treated until the water is of a quality suitable for discharge or reuse in the process. A number of processes have been adopted. Typically, treatment involves:

1. 1–2 days in an open acidification pond;
2. up to 20 days in a tank for anaerobic digestion;
3. about 20 days in a lagoon with vigorous aeration for anaerobic digestion.

The press fiber, after separation of the nuts, is used as fuel in the mill boilers. Consequently, the mill is usually self-sufficient in fuel for steam and electric power. Some shell may be used in the boiler. Alternate uses are as hard-core on estate roads or as raw material for charcoal.

Recently, an alternative treatment process has been developed. The waste stream is partly dewatered in a decanter centrifuge, and then treated in a rotary drier, using heat from the boiler flue, or produced by burning empty fruit bunches. The resulting solid is used as a fertilizer in the plantation.

Refining of Palm Oil

Crude palm oil may be refined by the traditional alkali refining process. In view of its natural strong red color, a highly active bleaching earth may be required. Table 3 gives a flow sheet for a typical process.

Crude palm oil usually contains 3–5% of free fatty acids, and so the process losses in alkali refining tend to be rather high. This has led to the development of physical or steam refining, where the use of a somewhat higher temperature in a modified deodorizer enables the free fatty acids to be removed by distillation instead of neutralization. The process has proved to be more economical and is now being adopted for other oils. A flow sheet for physical refining is given in Table 4.

Further Processing of Palm Oil

Palm oil is subjected to further processes in the refinery in order to make it more widely useful in food applications.

The following processes are carried out on a large scale in order to modify the physical properties:

- fractionation;
- hydrogenation;
- interesterification.

Brief descriptions of the processes will be given here, and the properties of the products will be described in a later section.

The major glyceride components of palm oil range from triolein (melting point 5 °C) to tripalmitin (melting point 66 °C) with a number of mixed glycerides of intermediate melting point.

Crystallization from the melt of this mixture at a controlled temperature, followed by separation of the liquid and solid phases, results in palm olein and palm stearin. The characteristics of these products depend on the temperature chosen for crystallization and the efficiency of the separation process. This may be by centrifugation, filtration on a rotating band filter, or in a plate and frame filter. A

modification of the latter has each frame fitted with an inflatable diaphragm, which enables pressure to be applied to the solids, and results in very efficient separation of the olein. Fractionation from solution in acetone or hexane is more costly and is only used when a sophisticated midfraction is required (see later sections).

Hydrogenation

Hydrogenation is a standard process in the edible oil industry. It involves treating the oil with an activated nickel catalyst with hydrogen under pressure and at an elevated temperature. Vigorous stirring is required, since the liquid and gas phases have to contact at the surface of the catalyst. The oil is neutralized, washed, and bleached before hydrogenation, to avoid deactivation of the catalyst. (*See* **Vegetable Oils**: Oil Production and Processing.)

Interesterification

When a glyceride oil is stirred with sodium methoxide (or other alkali catalyst) at a temperature of about 90 °C, the fatty acid radicals are detached from their original positions, and new glycerides are formed with a random distribution of fatty acids. Examples of interesterified products will be given in later sections. (*See* **Vegetable Oils**: Oil Production and Processing.)

Table 3 Flow sheet for alkali refining

Process step	*Typical conditions*	*Main impurities removed or reduced*
Pretreatment,	0.1% phosphoric acid	Phospholipids
Gum conditioning	80 °C – 20 min	Trace metals
		Pigments
Neutralization	4 N caustic soda	Fatty acids
	20% excess	Phospholipids
		Pigments
Washing		Soap
Drying		Water
Bleaching	1% active earth	Pigments
	80–100 °C, vacuum	Oxidation products
		Trace metals
		Soap residues
Filtration		Spent bleaching earth
Deodorization	240 °C at 133–667 Pa (1–5 torr)	Fatty acids
	90–110 min	Partial glycerides
		Oxidation products
		Pigment
		Decomposition products
Polishing filter		Traces of oil
		Insolubles

Table 4 Flow sheet for physical refining

Process step	*Typical conditions*	*Main impurities removed or reduced*
Gum conditioning	0.1% phosphoric acid	Phospholipids
	80 °C – 20 min	Trace metals
		Pigments
Bleaching	1–2% active earth	Phospholipids
	90 °C – 20 min	Trace metals
	Vacuum	Pigments
		Oxidation products
Filtration		Spent earth
Deacidification and deodorization	260–265 °C at 133–667 Pa (1–5 torr)	Free fatty acids
		Partial glycerides
		Oxidation products
		Pigment decomposition products
		Traces of oil
Polishing filter		Insolubles

Chemical Composition of Palm Oil and Fractions

Average fatty acid composition data from refined palm oil and standard palm olein and stearin, as traded, are shown in Table 5. Tailor-made fractions are also available for specific requirements. The last column shows the composition of a midfraction obtained from palm oil under special conditions, which is suitable for use in confectionery fats. Triacylglycerol compositions for the same products are given in Table 6.

Physical Properties of Palm Oil

Many food products require a consistent or semisolid fat as an ingredient in order to achieve the required structure. An important feature of palm oil is its natural content of solids. The average solid fat content of palm oil products is given in Table 7. Palm stearins of a wide choice of composition and solid fat contents are available, for example, with solid fat contents at 20 °C ranging from 35 to 72%.

Minor Components of Palm Oil

Data for the minor components of palm oil, collected from published sources, are given in Table 8. The main sterols present are β-sitosterol (58%), campesterol (22%) and stigmasterol (11%). The main carotenoids are β-carotene (56%) and α-carotene (35%). The carotenoids are removed in the normal refining process. The tocol content is unusual in having a high proportion of unsaturated tocotrienols. Details for crude and refined products are given in Table 9. The tocols are important as vitamin E and as potent natural antioxidants. (*See* **Antioxidants**: Natural Antioxidants; **Carotenoids**: Occurrence, Properties, and Determination.)

Table 5 Mean fatty acid composition (%)

	Refined palm			*Palm midfraction*[a]
	Oil	*Olein*	*Stearin*	
12:0	0.24	0.27	0.18	
14:0	1.11	1.09	1.27	0.7
16:0	44.14	40.93	56.79	60.9
16:1	0.1	0.1	0.1	
18:0	4.44	4.16	4.93	4.6
18:1	39.04	41.51	29.0	31.0
18:2	10.57	11.64	7.23	2.6
18:3	0.2	0.2	0.2	0.1
20:0	0.2	0.1	0.2	0.3

[a]Data from Britannica Food Ingredients with permission.
Adapted from Siew WL, Tang TS, Oh FCH, Chong CL and Tan YA (1993) Identity characteristics of Malaysian palm oil products: fatty acid and triglyceride composition and solid fat content. *Elaeis* 6(1): 38–46, with permission.

Table 6 Mean triacylglycerol compositon of refined oils (carbon numbers by GLC)[a]

Carbon number	*Palm oil*	*Palm olein*	*Palm stearin*	*Palm midfraction*[b]
C44	0.07	0.09	0.13	
C46	1.18	0.77	3.13	
C48	8.08	3.28	23.72	2.2
C50	39.88	39.52	40.31	78.9
C52	38.77	42.74	25.28	15.1
C54	11.35	12.80	6.86	0.7
C56	0.59	0.67	0.45	
Unidentified				3.1

[a]Carbon numbers are the sum of the carbon atoms in three acyl groups.
[b]Data from Britannica Food Ingredients with permission.
Adapted from Siew WL, Tang TS, Oh FCH, Chong CL and Tan YA (1993) Identity characteristics of Malaysian palm oil products: fatty acid and triglyceride composition and solid fat content. *Elaeis* 6(1): 38–46, with permission.

Table 7 Mean solid fat content of standard refined oils

Temperature (°C)	*Palm oil (mean) %*	*Palm olein (mean) %*	*Palm stearin (mean) %*	*Palm midfraction*[a] *%*
10	53.6	38.27	76.04	
15	39.13	19.89	68.91	
20	26.10	5.67	60.71	89.9
25	16.28	2.05	50.55	82.6
30	10.54		40.39	50.2
35	7.85		34.30	50.2
40	4.64		28.13	0.0
45			22.38	
50			12.45	
55			0.60	

[a]Tempered at 26 °C for 40 h.
Adapted with permission from Siew WL, Tang TS, Oh FCH, Chong CL and Tan YA (1993) Identity characteristics of Malaysian palm oil products: fatty acid and triglyceride composition and solid fat content. *Elaeis* 6(1): 38–46, with permission.
Data from Britannica Food Ingredients with permission.

Table 8 Minor components of crude palm oil typical figures (p.p.m.)

Sterols	490
4-Methyl sterols	360
Triterpenic alcohols	550
Isoprenoid alcohols	80
Other alcohols	130
Tocols	830
Carotenoids	670
Squalene	350
Other hydrocarbons	40

Reproduced with permission from Palm Oil Research Institute of Malaysia.

Table 9 Tocol content of palm oil (p.p.m.)

		α-tocopherol	*α-tocotrienol*	*γ-tocotrienol*	*δ-tocotrienol*	*Total*
Crude palm oil	Mean (n=9)	162	165	324	81	774
	Range	136–241	90–205	273–439	67–94	635–890
Refined palm oil	Mean (n=3)	117	117	158	31	426
	Range	85–180	99–147	67–239	5–62	256–630
Refined palm olein	Mean (n=8)	141	152	218	49	561
	Range	107–163	131–177	113–293	28–68	478–673

Reproduced with permission from Palm Oil Research Institute of Malaysia.

Food Uses of Palm Oil

Frying

Palm oil has a good stability at the high temperatures used in frying (usually 175–185 °C), because of its content of natural antioxidants, the absence of highly unsaturated fatty acids, and the moderate content of linoleic acid. Consequently, palm oil or palm olein is widely used domestically and in industry, especially for deep fat frying whether in a batch process, as in restaurants and fast food outlets such as British 'fish and chip' shops, or in continuous fryers. Palm oil is used for doughnut frying, because the solids content assists the adhesion of the sugar coating. Palm oil is also used for the frozen 'French fry' industry and for instant noodle frying in Japan and China. For potato crisps (American 'chips'), palm olein is preferred, or a blend of palm olein with a more unsaturated oil such as sunflower or soya bean oil.

Bakery Fats

Texturized palm oil is used as such in some types of biscuits. For other biscuits and for cakes, a formulated shortening is required to obtain good aerating properties. Some formulae with a satisfactory performance are given in Table 10.

Vanaspati

Vanaspati may be defined as a cheaper alternative to butterfat, having a melting point of about 37 °C and a granular crystalline structure with little or no free oil. It is generally based on vegetable oils suitably hydrogenated or on a blend of oils. Vanaspati is the customary domestic cooking fat in the Indian subcontinent and the Middle East. When formulated from liquid vegetable oils, the desired structure requires a hydrogenation giving a high content of *trans* fatty acid isomers (30–60%). This is regarded as undesirable according to current nutritional advice. A number of formulae based on palm oil and having low or zero *trans* fatty acids are given in Table 11.

Table 10 Shortening formulae containing palm oil

	1	*2*	*3*	*4*
Palm oil	50			
Hardened palm oil melting point (49–51 °C)	15			
Liquid oil	35			
Palm stearin		35	42	
Hardened rapeseed oil melting point 36 °C		30		
Rapeseed oil		35	40	
Hardened palm oil melting point 42 °C			18	
Palm olein (interesterified)				100

Reproduced with permission from Palm Oil Research Institute of Malaysia.

Table 11 Vanaspati formulae containing palm oil products

	1	*2*[a]	*3*	*4*[b]
Hardened palm olein (melting point 41 °C)	24			
Palm oil	56	80	70	80
Liquid oil	20			
Rice bran oil		20		
Palm stearin			7	
Hardened soya bean oil (melting point 34 °C)			23	20
Trans fatty acids (%)	2.7	Nil	7.5	4

[a]The blend is interesterified.
[b]Current formulae of this type in use in Pakistan.
Reproduced with permission from Palm Oil Research Institute of Malaysia.

Margarines

Margarines can be classified as being for table use in blocks or in tubs, general bakery margarines, and puff-pastry margarines.

The textual properties are the textual properties appropriate for each application, and these can be obtained by blending a variety of ingredients. Table 12 shows some formulae based on palm oil.

Confectionery Fat

Palm oil midfraction (see Tables 5–7) is used as a major component of blends designed to have physical properties like cocoa butter, and compatible with it in mixtures. Palm olein, partly hydrogenated under conditions giving a high content of *trans* fatty acids, is

Table 12 Margarine formulae using palm oil products

	Block (temperate)	*Block (tropical)*	*Tub*	*General bakery*	*Danish pastry*	*Puff pastry*
Palm oil	50	80	50	65	40	50
Hydrogenated palm oil (melting point 44 °C)	20			10	40	50
Liquid oil	30		50	10	20	
Coconut oil				15		
Palm stearin		20				

Reproduced with permission from Palm Oil Research Institute of Malaysia.

useful as a confectionary fat with limited compatability with cocoa butter, or for use in toffee, bakery coatings, and confectionery centers.

Miscellaneous Uses

Palm oil is used instead of butterfat in icecream and in filled milk or coffee whiteners. Partly hydrogenated palm oil is used in dried soups, where its stability against oxidation is important.

Special Products

Using a modified refining process, red palm oil and palm olein, retaining about 80% of the carotenoid content of the crude oil, have become commercially available. Red olein has found uses in the manufacture of attractively golden-colored potato crisps, whereas red palm oil is being used in nutritional intervention studies in South Africa and India. This is potentially a very important use, in view of the high incidence of xerophthalmia in developing countries.

A commercial process is also in operation to obtain a 99% pure concentrate of the tocopherols and tocotrienols from the palm fatty acid distillate, the byproduct of physical refining. The concentrate is used in health supplements.

See also: **Antioxidants**: Natural Antioxidants; **Carotenoids**: Occurrence, Properties, and Determination; **Palm Kernel Oil**; **Tocopherols**: Properties and Determination

Further Reading

Berger KG (1983) Palm Oil. In: Chan HT (ed.) *Handbook of Tropical Foods*, pp. 433–468. New York: Marcel Dekker.

Berger KG (1983) Production of palm oil from fruit. *Journal of American Oil Chemists' Society* 60: 158–162.

Hartley CWS (1988) *The Oil Palm*, 3rd edn. Harlow, UK: Longman Scientific and Technical.

Kheiri SA (1987) Palm Oil. In: Gunstone FD (ed.) *Critical Reports on Applied Chemistry*, vol. 15. London: Society of Chemical Industry.

Siew WL, Tang TS, Oh FCH, Chong CL and Tan YA (1993) Identity characteristics of Malaysian palm oil products: fatty acid and triglyceride composition and solid fat content. *Elaeis* 6(1): 38–46.

Tan BK and Oh FCH (1981) *Malaysian Palm Oil Chemical and Physical Characteristics*, Porim Technology No. 3. Kuala Lumpur: Palm Oil Research Institute of Malaysia.

Palms *See* **Coconut Palm**; **Date Palms**; **Sago Palm**; **Sugar**: Sugarcane; Sugarbeet; Palms and Maples; Refining of Sugarbeet and Sugarcane

Pancreatic Hormones *See* **Hormones**: Adrenal Hormones; Thyroid Hormones; Gut Hormones; Pancreatic Hormones; Pituitary Hormones; Steroid Hormones

PANTOTHENIC ACID

Contents
Properties and Determination
Physiology

Properties and Determination

G F M Ball, Wembley, London, UK

Background

In 1933, a research team led by R. J. Williams isolated from a variety of biological materials an acidic substance that acted as a growth factor for yeast. Williams' team elucidated the chemical structure of the purified substance and named it pantothenic acid because of its apparently widespread occurrence (Greek *pantos*, meaning everywhere). Pantothenic acid was established as a vitamin in 1939, when it was shown to be identical to a 'filtrate factor' required by rats for normal growth, and to a chick antidermatitis factor. Sometimes referred to as vitamin B_5, pantothenic acid is a member of the water-soluble B-group vitamins.

Structure and Physicochemical Properties

The biological activity of pantothenic acid is attributable to its incorporation into the molecular structures of coenzyme A and acyl carrier protein. The molecular structures of pantothenic acid and related compounds are shown in Figure 1. Pantothenic acid ($C_9H_{17}O_5N$; molecular weight = 219.23) is composed of pantoic acid linked by an amide bond to β-alanine. The pantothenic acid molecule, having a chiral carbon atom, exhibits optical isomerism as well as being optically active. Only the D(+) isomer occurs in nature. Synthetic pantothenic acid is a racemic (DL) mixture, and, since only the D isomer is biologically active, this fact must be considered if the DL mixture is to be used therapeutically. Pantothenic acid is a pale yellow oil that is extremely hygroscopic and so is unsuitable for commercial application. For human food supplements, calcium D-pantothenate [$(C_9H_{16}O_5N)_2Ca$; molecular weight = 476.53] is used.

The corresponding alcohol of pantothenic acid, pantothenol (referred to commercially as panthenol), is widely used as a source of pantothenate activity for pharmaceutical vitamin products, because it is more stable than the pantothenate salts, especially in liquid multivitamin products that must be slightly acid to preserve the thiamin content. Pantothenol does not occur naturally and itself has no pantothenate activity, but it is converted quantitatively to pantothenic acid in the body.

The stability of pantothenic acid and its calcium salt in aqueous solution is highly dependent on the pH. In contrast to other B-group vitamins, pantothenic acid becomes more stable as the pH of the solution increases. Solutions of calcium pantothenate are most stable between pH 5 and 7 but, even so, are not stable to autoclaving, and therefore, sterilization by ultrafiltration is necessary for pharmaceutical preparations. Below and above these pH values, solutions of calcium pantothenate are thermolabile. Alkaline hydrolysis yields pantoic acid and β-alanine, whereas acid hydrolysis yields the γ-lactone of pantoic acid. Pantothenic acid is unaffected by atmospheric oxygen and light.

Dietary Sources

Pantothenic acid is widely distributed in foods of both animal and plant origin. In concentration units of mg per 100 g, the vitamin is particularly abundant in liver (8), kidney (3), heart (2.5), egg yolk (4.6) broad beans (4.9), and peanuts (2.7). Lesser amounts are found in beef (0.6), chicken (1.2), potatoes (0.4), broccoli (1.2), oatmeal (1.0), and milk (0.35), but these will be important food sources if consumed in sufficient quantity. Outstandingly high amounts are found in the ovaries of tuna and cod (232) and in royal jelly from the queen bee (50). In contrast, highly refined foods such as sugar, fats and oils, and cornstarch are totally devoid of the vitamin.

Coenzyme A is the major pantothenic acid-containing compound present in foods of both animal and plant origin, accompanied by small amounts of other bound forms (phosphopantothenic acid, pantetheine, and phosphopantetheine). Notable exceptions are human and bovine milk in which free (unbound) pantothenic acid constitutes around 90% of the total pantothenate content.

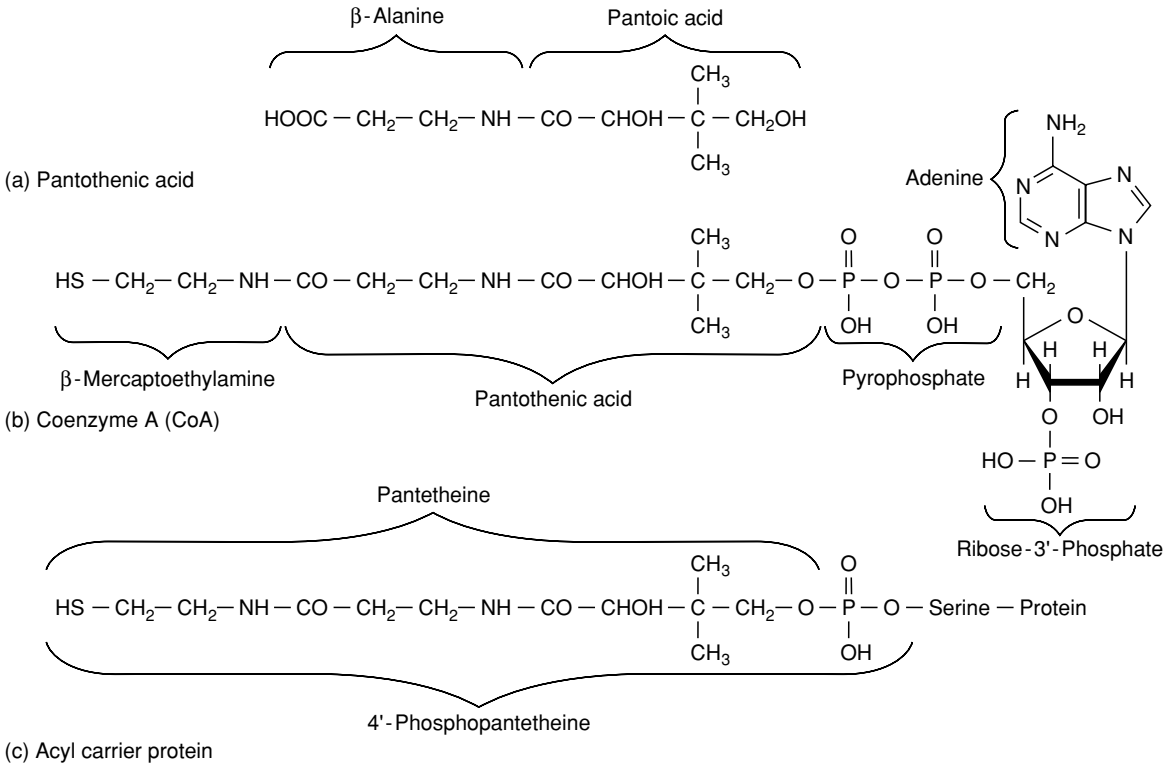

Figure 1 Structures of (a) pantothenic acid, (b) coenzyme A, and (c) acyl carrier protein.

Pantothenic acid has a good stability in most foods during home cooking but is susceptible to leaching. The roasting of meat causes degradation of less than 10%, but the meat drippings contain 20–25% of the initial vitamin content.

Estimates of dietary intakes of pantothenic acid should be based not on the raw food values, but on the cooked food values, otherwise falsely high estimates of intake will be obtained. Table 1 presents data taken from a study in which the pantothenic acid content of 75 processed and/or cooked foods was determined by radioimmunoassay. The foods selected for analysis were those commonly consumed in the USA. Samples were brought to a ready-to-eat stage, following any package directions, and using only the edible portion. Results indicated that the canning of foods incurs large losses of pantothenic acid, as does the conversion of grains to various cereal products, and the processing of meats to produce fat- and cereal-extended products such as frankfurters and sausages.

Little information is at hand regarding the nutritional availability to humans of pantothenic acid in food commodities. In one study, based on the urinary excretion of pantothenic acid, the availability for male human subjects ingesting 'the average American diet' ranged from 40 to 61%, with a mean of 50%.

Analysis

Pantothenic acid in its various bound forms (mainly coenzyme A) is routinely determined by microbiological assay. Other published methods include radioimmunoassay, enzyme-linked immunosorbent assay, gas chromatography, and high-performance liquid chromatography. For all of these techniques, it is necessary to liberate pantothenic acid from its bound forms by enzymatic hydrolysis, because only the free vitamin can be measured. Hydrolysis is not required, however, for calcium pantothenate-supplemented foods or for milk in which the free vitamin predominates.

Extraction from Food

Neither acid nor alkaline hydrolysis is applicable for the liberation of bound pantothenic acid, since the vitamin is degraded by such treatments. The only practicable alternative is enzymatic hydrolysis, and this was successfully accomplished through the simultaneous action of intestinal phosphatase and an avian liver enzyme. This double enzyme combination liberates practically all of the pantothenic acid from coenzyme A, but it does not release the vitamin from acyl carrier protein. The phosphatase splits the coenzyme A molecule between the phosphate-containing

Table 1 Pantothenic acid content of processed and/or cooked foods purchased in Utah, USA

Food[a]	*Pantothenic acid content (mg per 100 g) (mean and standard deviation)*[b]
Breads, cereals, and other grain products	
Bran'ola bread (high-fiber)	0.458 ± 0.044
Rolls, hamburger	0.471 ± 0.075
Ready-to-eat cereals	
Cheerios (oats)	1.341 ± 0.198
Corn Flakes (corn)	0.284 ± 0.032
Wheat Chex (wheat)	0.502 ± 0.042
Rice, white	0.261 ± 0.036
Meat, fish, poultry, and meat products	
Beef, regular ground – pan broiled	0.671 ± 0.048
Pork loin chops – pan broiled	0.650 ± 0.051
Fish filet, frozen, breaded – baked	0.250 ± 0.016
Chicken breast – baked with skin	1.188 ± 0.049
Frankfurters	0.342 ± 0.025
Salami	0.997 ± 0.086
Fruits and vegetables	
Orange juice, frozen, reconstituted	0.197 ± 0.029
Potatoes – baked	0.318 ± 0.045
Potatoes – boiled	0.291 ± 0.018
Potatoes – canned	0.152 ± 0.026

[a]The use of brand names is for identification purposes only and does not imply endorsement of a food product.
[b]Data from Walsh JH, Wyse BW and Hansen RG (1981) Pantothenic acid content of 75 processed and cooked foods. *Journal of the American Dietetic Association* 78: 140–144.

moiety and pantetheine, while the liver enzyme breaks the link in pantetheine between the pantothenic acid and β-mercaptoethylamine moieties. Both enzymes are available commercially as powdered extracts. Liver enzyme preparations contain a relatively high amount of coenzyme A, which is converted to pantothenate during the incubation period, thus creating an unacceptably high blank value. Such preparations can be purified quite simply by treatment with Dowex 1-X4 anion exchange resin. Intestinal phosphatase preparations contain negligible amounts of coenzyme A and do not require purification.

Microbiological Assays

Microbiological methods, as applied to the determination of the B-group vitamins, are based on the absolute requirement of a particular microorganism (the assay organism) for the vitamin in question (in this case, pantothenic acid); that is, the organism can multiply only when the vitamin is present in the surrounding medium. In a typical turbidimetric microbiological assay, aliquots of a standard solution of pantothenic acid, or aliquots of the sample extract containing pantothenic acid, are added to an initially translucent basal nutrient medium, complete in all respects except for pantothenic acid. Following inoculation with the assay organism, the organism multiplies in proportion to the pantothenic acid content of the standard or sample, and the extent of the growth is ascertained by measuring the turbidity produced. Over a defined concentration range, the measured response will be directly proportional to the amount of pantothenic acid present, and, within this range, the sample solution and standard pantothenic acid solution can be compared accurately. The usual assay organism is *Lactobacillus plantarum* (ATTC (American Type Culture Collection) No. 8014), which can also be used for assaying nicotinic acid and biotin. The basal nutrient medium can also be used for assaying nicotinic acid and biotin, with the exclusion of the relevant vitamin from the formulation. Fatty acids are stimulatory in the presence of suboptimal amounts of pantothenic acid, so a preliminary ether extraction step may be necessary.

In the standard turbidimetric procedure, the basal nutrient medium is prepared at twice its final concentration. Multiple aliquots of a standard solution of pantothenic acid and of enzyme-treated extracts of the test food are added to a series of uniform assay tubes in amounts suitable to produce gradations in growth between no growth and maximum growth. The contents of all tubes are diluted with water to the same volume, and an equal volume of the basal medium is added. The tubes are sterilized, cooled to a uniform temperature, and then inoculated with an actively growing culture of *L. plantarum*. The tubes are incubated for 6–24 h at any selected temperature between 30 and 40 °C held constant to ±0.5 °C until growth has reached the maximum permitted by the limiting vitamin present, pantothenic acid. The growth response to standard and test extract is then determined by measuring the turbidity produced. The data obtained from the standards are used to construct a standard curve from which the pantothenic acid concentrations of the various sample aliquots are derived. The use of multiple aliquots allows a validity check to be carried out: the pantothenic acid concentration found should be directly proportional to the volume of aliquot taken. The amount of pantothenic acid present in the original sample is then calculated at the different test levels, and the results are averaged to obtain the final result.

An alternative method, the radiometric microbiological assay, is based upon the measurement of radioactive $^{14}CO_2$ generated from the metabolism of a ^{14}C-labeled nutrient by the test organism. The radioactivity is measured automatically by means of a commercially available gas flow system incorporating an ionization chamber. Sample preparation for this technique is simplified due to the fact that

colored, turbid, or precipitated debris does not interfere with the $^{14}CO_2$ output or detection; furthermore, the scrupulous cleaning of glassware required for turbidimetric assays is unnecessary. An assay for pantothenic acid in human milk and blood is based on the measurement of $^{14}CO_2$ produced from the metabolism of L-[1-^{14}C]methionine or L-[1-^{14}C]-valine by the yeast *Kloeckera brevis* (ATCC 9774). Metabolic CO_2 can also be measured nonradiometrically with the aid of an infrared CO_2 analyzer, which measures automatically the infrared radiation absorbed by the CO_2 band at 4.2 μm.

Radioimmunoassay

The radioimmunoassay is based on the competition for a fixed, but limited, number of antibody binding sites by antigen (a substance capable of binding to a specific antibody) and a trace amount of radiolabeled antigen added to the sample extract. In this case, the antigen is the analyte, pantothenic acid. The presence of larger amounts of unlabeled analyte results in less radioactivity being bound to the antibody.

Wyse and colleagues raised antibodies to pantothenic acid by coupling a bromoacetyl derivative of pantothenic acid with reduced and denatured bovine serum albumin and injecting this immunogen into rabbits. For the assay, each tube contained diluted antiserum, pantothenic acid standard solution or sample extract, and radiolabeled sodium *d*-pantothenate. After incubation, neutral saturated ammonium sulfate was added to facilitate suspension of the antibody-bound pantothenic acid, which was then centrifuged. The washed precipitate, containing antibody-bound pantothenic acid, was dissolved in tissue solubilizer, and the radioactivity was counted in a scintillation counter. The amount of pantothenic acid in each unknown was determined by reference to a standard curve constructed by plotting on logit-semilog paper log concentration of nonradioactive pantothenic acid in the standard vials (ng per 0.50 ml) versus the percentage of the counts bound. Results from the radioimmunoassay and AOAC microbiological assay for pantothenic acid in 75 processed and cooked foods were highly correlated ($r = 0.94$). However, there was a statistically significant difference ($p < 0.05$) between the two assay results for all foods and for the subgroups meats, breads and cereals, and fruits and vegetables. At $p < 0.01$, only meats were significantly different. For breads and cereals, the microbiological assay results averaged 6.6% higher than those of the radioimmunoassay. For fruits and vegetables, the microbiological assay results were 11.6% higher; for meats, the results were 23.2% higher. It was postulated that bacterial enzymes in the assay organism promote further breakdown of bound pantothenic acid, or nonenzymatic breakdown occurs during the long microbiological incubation period.

Enzyme-linked Immunosorbent Assay (ELISA)

An ELISA is an enzyme-linked immunoassay in which one of the reactants is immobilized by physical adsorption on to the surface of a solid phase. In its simplest form, as used in food analysis applications, the solid phase is provided by the plastic surface of a 96-well microtitration plate. The generally preferred format for vitamin assays in food analysis is a two-site noncompetitive assay used in the indirect mode. This format employs two antibodies: a primary antivitamin antibody raised against an immunogen (in this case, a pantothenic acid–protein conjugate), and an enzyme-labeled, species-specific second antibody, which binds specifically to the primary antibody. To perform such an assay, a protein conjugate of pantothenic acid is immobilized to the well surface of the microtitration plate, the attached protein being different to that used for the immunogen. The protein adsorbs passively and strongly to the plastic, and, once coated, plates can usually be stored for several months. To perform the assay, the standard solution or sample extract is added to the well, followed by a limited amount of primary antibody. After incubation, the antibody becomes distributed between immobilized vitamin and free vitamin according to the amount of vitamin initially present. After phase separation, achieved by well emptying and washing, the second antibody is added in excess, and the plate is incubated for a second time. Excess unbound material is removed, and the amount of bound enzyme is determined by addition of substrate and spectrophotometric measurement of the colored product. Unknown samples are quantified by reference to the behavior of vitamin standards.

Finglas and colleagues developed a noncompetitive ELISA, which is highly specific for pantothenic acid and does not recognize coenzyme A, panthenol or pantheneine. The primary antivitamin antibody was raised in rabbits according to the method of Wyse and colleagues (*see* section Radioimmunoassay), and the enzyme-labeled, species-specific second antibody (alkaline phosphatase-labeled antirabbit IgG) was obtained commercially. Microtitration plates were coated with pantothenic acid–keyhole limpet hemocyanin as the immobilized phase of the assay. A high correlation coefficient ($r = 0.999$) was reported when ELISA values obtained for six foods were compared with corresponding values obtained by a microbiological method using *L. plantarum*. Gonthier and

colleagues improved the sensitivity of the ELISA by using an immunogen composed of pantothenic acid coupled to thyroglobulin by a 6-carbon atom linker (adipoyl dichloride). By contrast, the bromoacetyl linker used in Finglas' pantothenic acid–bovine serum albumin immunogen contains two carbon atoms.

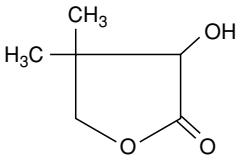

Figure 2 Pantoyl lactone.

Gas Chromatography

Salts of pantothenic acid present in pharmaceutical preparations have been analyzed by gas chromatography after conversion to volatile acetate, trifluoroacetate, or trimethylsilyl derivatives. An alternative approach to derivatization is to chromatograph the pantoyl lactone formed from pantothenic acid by acid hydrolysis (Figure 2). This approach is applicable to foodstuffs, because the hydrolysis reaction liberates the lactone from the free and bound pantothenic acid in the food matrix with a recovery of at least 95%. Davídek and colleagues applied this technique to the determination of pantothenic acid in fresh beef liver, spray-dried egg yolk, soybean flour, whole-grain wheat flour, and dried bakers' yeast. Samples were hydrolyzed by treatment with dilute hydrochloric acid, and the neutralized hydrolysate, after filtration, was extracted with dichloromethane. The combined extracts, to which ethyl laurate was added as an internal standard, were concentrated by rotary evaporation, and then analyzed by gas chromatography using a polar stationary phase of 10% Carbowax 20M and a flame ionization detector. Gas chromatographic results correlated with results obtained by the currently accepted microbiological method ($r = 0.975$), and no significant difference was found between the two sets of results ($p > 0.05$). Davídek and colleagues used a packed column of dimensions 2.4 m × 2 mm i.d. in which the stationary phase was coated on to a porous support material of diatomaceous earth treated with dimethyldichlorosilane. Woollard and colleagues upgraded the column to a more efficient 30 m × 0.25 mm open-tubular capillary column coated with stationary phase (BPX70).

High-performance Liquid Chromatography (HPLC) and Capillary Electrophoresis

The pantothenic acid molecule does not contain a characteristic chromophore, and hence it exhibits only very weak absorbance at 204 nm, owing to the presence of carbonyl groups. Detection at wavelengths below 220 nm is subject to interference from the many organic compounds present in a typical food sample extract prepared for HPLC. The problem of weak and nonspecific absorbance, coupled with the low concentrations of pantothenic acid in foods, has thwarted attempts to apply HPLC to the determination of pantothenic acid in foodstuffs, although the technique has been successfully applied to pharmaceutical products. Chemical treatment of the pantothenic acid molecule to form a derivative that fluoresces or absorbs at a higher wavelength is a possibility, but, so far, reproducible results have not been obtained using such an approach. Refractometry as a means of detection lacks the required sensitivity and specificity.

Recognizing these problems, Woollard and Indyk in New Zealand developed an HPLC method for determining free endogenous D-pantothenic acid in milk and supplemental calcium pantothenate in infant formulas. Sample preparation simply entailed addition of acetic acid to the milk or reconstituted infant formula, followed by centrifugation and membrane filtration. This treatment resulted in a protein- and fat-free extract that could be directly injected (10-µl aliquot). The problems of poor detection specificity in the low-ultraviolet region of the spectrum were overcome by the use of a photodiode array detector that provided multiwavelength detection (selected wavelengths were 200, 205 and 240 nm) and on-line spectral analysis. The HPLC system incorporated an on-line mobile phase degasser – an important feature, as dissolved oxygen constitutes a source of interference at low ultraviolet wavelengths. Among several reversed-phase columns investigated, the column selected for routine use was of dimensions 250 × 4.6 mm i.d. and packed with Luna (Phenomenex) 5 µm C_8 (octyl) of 100 Å pore size, 400 $m^2 g^{-1}$ surface area, and 13.5% carbon load. The Luna material is based on low-acidity silica, exhaustive end-capping, and shielded bonded phase ligand. The mobile phase was phosphate buffer (0.1 M, pH 2.25): acetonitrile (97:3, v/v) delivered initially at 1.4 ml min^{-1} and increased to 1.8 ml min^{-1} at 18 min. Following completion of the sample schedule, the column was purged sequentially with acetonitrile:water (50:50), water (100%), acetonitrile:water (50:50), and finally acetonitrile (100%) for column storage between runs. The retention time of the pantothenic acid was around 15 min, but the next sample was not injected until a major unknown peak with a retention time of about 35 min was removed (Figure 3).

Reversed-phase columns do not retain acidic solutes in the ionic state, but if the mobile phase is buffered to pH 3 or lower, the acidic solute will be nonionized and act as a neutral solute. This technique is known as ion suppression. Under these conditions, residual silanol groups on the silica support will also be nonionized. The net result is retention of undissociated acidic solutes with no peak tailing, owing to the elimination of electrostatic interactions between acidic solutes and silanol groups. A potential problem with certain reversed-phase column packings is that the siloxane (≡Si—O—Si≡) bond linking the alkyl ligand to the silica support is subject to hydrolysis at low pH, resulting in loss of bonded phase. Although longer-chain ligands such as C_{18} are relatively stable at a low pH, short-chain bonded phases, including small endcapping groups, are especially susceptible. The problem of hydrolysis and loss of bonded phase can be minimized by the use of 'shielded' stationary phases, which are sterically protected from attack by hydrolyzing protons. Results obtained by HPLC correlated with those obtained by microbiological assay utilizing *L. plantarum* ($r = 0.971$), and there was no significant difference ($p > 0.05$) between the two sets of results.

Reversed-phase HPLC with ion suppression is unable to separate the D and L enantiomers in synthetic calcium pantothenate. However, separation

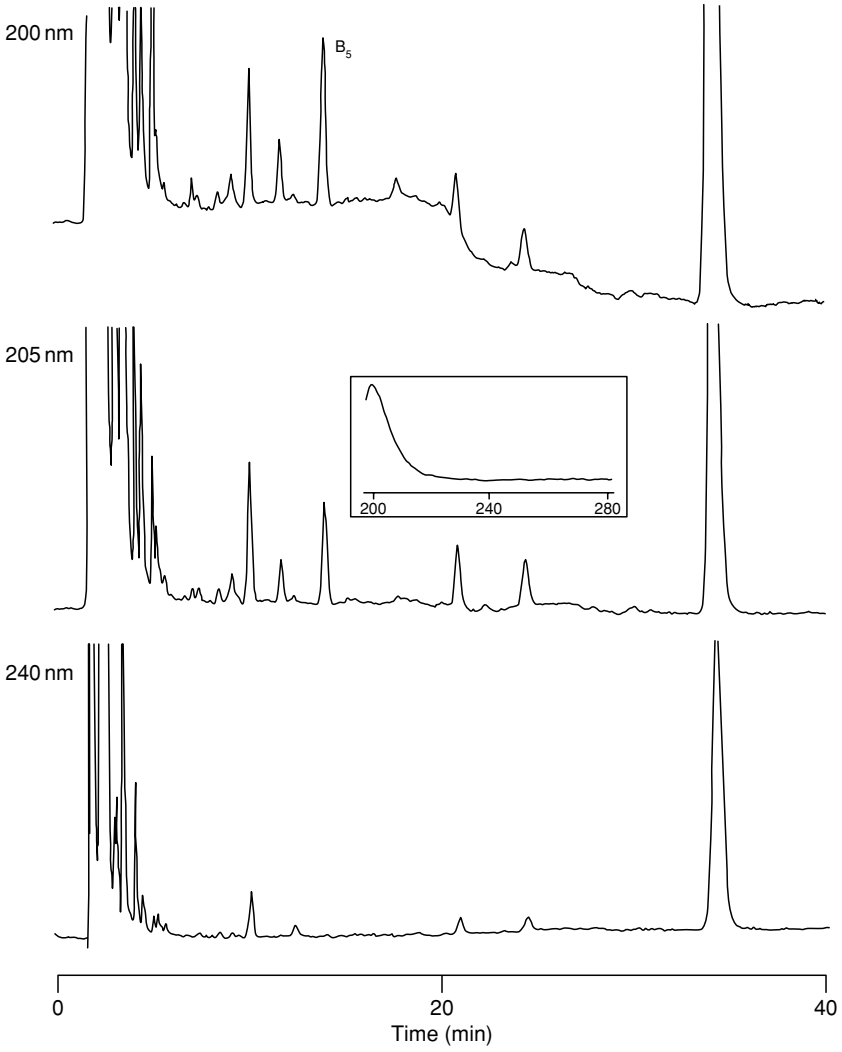

Figure 3 Multiwavelength UV chromatogram (200, 205, 240 nm) of a typical infant formula extract obtained with a Luna C_8 (octyl) column. The insert illustrates a UV spectral scan of pantothenic acid. Chromatographic parameters are given in the text. Reprinted from Woollard DC, Indyk HE and Christiansen SK (2000) The analysis of pantothenic acid in milk and infant formulas by HPLC. *Food Chemistry* 69: 201–208, with permission from Elsevier Science.

of enantiomers in DL pantothenic acid has been reported using a chiral selector in capillary electrophoresis. Optimum separation was obtained using a pH 7.0 phosphate buffer containing the chiral selector (60 mM 2-hydroxypropyl-β-cyclodextrin) and 10% (v/v) methanol. The enantiomers were unresolved when a buffer of pH 3.0 was tried, implying that dissociation of pantothenic acid was necessary for its chiral resolution under the conditions employed.

Overall Appraisal of Analytical Techniques

The 'free' (no enzyme treatment) or 'total' (after enzyme treatment) pantothenic acid content of a food has been traditionally determined by the turbidimetric microbiological assay using *L. plantarum*. This assay has been adopted as an official method by the AOAC on the basis of collaborative study. Once the facilities are in place, the microbiological assay can be used routinely to determine other B-group vitamins with minor changes in protocol. Inherent problems include stimulation or inhibition of bacterial growth by other compounds, and nonlinear response (drift) for various volumes of food extract analyzed. The radioimmunoassay produces results that, although correlated with microbiological assay results, are significantly lower. The use of radioisotopes would not be permitted in the vicinity of commercial food production. The ELISA is a better substitute for the microbiological assay, as results from the two techniques are in good agreement. The high technology of the ELISA is built into the reagents, so the assays are simpler to perform than microbiological assays. Routine use of the ELISA for determining pantothenic acid will depend on the commercial availability of standardized assay kits. Little interest seems to have been taken in gas chromatography, but the technique of chromatographing the lactone hydrolysis product merits further investigation using modern capillary columns. High-performance liquid chromatography is becoming increasingly popular for determining vitamins in foods, although the poor detectability of pantothenic acid limits the sensitivity of this technique. Capillary electrophoresis also suffers from poor sensitivity but, when used with a chiral selector, has the advantage of separating active d and inactive l enantiomers of racemic calcium pantothenate added to foods.

See also: **Bioavailability of Nutrients**; **Chromatography**: Principles; High-performance Liquid Chromatography; Gas Chromatography; **Immunoassays**: Principles; Radioimmunoassay and Enzyme Immunoassay

Further Reading

AOAC (1995) AOAC Official Method 992.07. Pantothenic acid in milk-based infant formula. Microbiological turbidimetric method. In: *Official Methods of Analysis of AOAC International*, Method No 50.1.22, 16th edn. Arlington VA: AOAC International.

Bell JG (1974) Microbiological assay of vitamins of the B group in foodstuffs. *Laboratory Practice* 23: 235–242, 252.

Crawley H (1993) Natural occurrence of vitamins in food. In: Berry Ottaway P (ed.) *The Technology of Vitamins in Food*. Glasgow: Blackie Academic & Professional.

Davídek J, Velíšek J, Černá J and Davídek T (1985) Gas chromatographic determination of pantothenic acid in foodstuffs. *Journal of Micronutrient Analysis* 1: 39–46.

Finglas PM, Faulks RM, Morris HC, Scott KJ and Morgan MRA (1988) The development of an enzyme-linked immunosorbent assay (ELISA) for the analysis of pantothenic acid and analogues. Part II – Determination of pantothenic acid in foods. *Journal of Micronutrient Analysis* 4: 47–59.

Goli DM and Vanderslice JT (1989) Microbiological assays of folacin using a CO_2 analyzer system. *Journal of Micronutrient Analysis* 6: 19–33.

Gonthier A, Boullanger P, Fayol V and Hartmann DJ (1998) Development of an ELISA for pantothenic acid (vitamin B_5) for application in the nutrition and biological fields. *Journal of Immunoassay* 19: 167–194.

Gonthier A, Fayol V, Viollet J and Hartmann DJ (1998) Determination of pantothenic acid in foods: influence of the extraction method. *Food Chemistry* 63: 287–294.

Guilarte TR (1989) A radiometric microbiological assay for pantothenic acid in biological fluids. *Analytical Biochemistry* 178: 63–66.

Kodama S, Yamamoto A and Matsunaga A (1998) Direct chiral resolution of pantothenic acid using 2-hydroxypropyl-β-cyclodextrin in capillary electrophoresis. *Journal of Chromatography A* 811: 269–273.

Morris HC, Finglas PM, Faulks RM and Morgan MRA (1988) The development of an enzyme-linked immunosorbent assay (ELISA) for the analysis of pantothenic acid and analogues. Part I – Production of antibodies and establishment of ELISA systems. *Journal of Micronutrient Analysis* 4: 33–45.

Walsh JH, Wyse BW and Hansen RG (1979) A comparison of microbiological and radioimmunoassay methods for the determination of pantothenic acid in foods. *Journal of Food Biochemistry* 3: 175–189.

Woollard DC, Indyk HE and Christiansen SK (2000) The analysis of pantothenic acid in milk and infant formulas by HPLC. *Food Chemistry* 69: 201–208.

Wyse BW, Wittwer C and Hansen RG (1979) Radioimmunoassay for pantothenic acid in blood and other tissues. *Clinical Chemistry* 25: 108–111.

Physiology

G F M Ball, Wembley, London, UK

Background

The biological activity of pantothenic acid is attributable to its incorporation into the molecular structures of coenzyme A (CoA) and acyl carrier protein. CoA performs multiple roles in cellular metabolism, whereas acyl carrier protein is involved in fatty acid biosynthesis. A wide variety of functional proteins are modified by the addition of acetyl, acyl, and isoprenyl groups through reactions that directly or indirectly involve CoA. These modifications allow the regulation of such important processes as gene transcription, signal transduction, and vision.

Metabolism

Intestinal Absorption

Ingested CoA, the major dietary form of pantothenic acid, is hydrolyzed in the intestinal lumen to pantetheine by the nonspecific action of pyrophosphatases and phosphatase. Pantetheine is then split into pantothenic acid and β-mercaptoethylamine by the action of pantetheinase secreted from the intestinal mucosa into the lumen. Within the alkaline medium of the luminal contents, pantothenic acid exists primarily as the pantothenate anion. Absorption of the liberated pantothenate takes place mainly in the jejunum.

At physiological intakes, pantothenate must move across the brush-border membrane of the intestinal epithelium from a region of lower concentration in the lumen to one of higher concentration in the cytoplasm of the absorptive cell (enterocyte). Such 'uphill' movement requires active transport – a mechanism that depends ultimately upon the expenditure of metabolic energy, i.e., the energy released from the hydrolysis of adenosine triphosphate produced during cellular metabolism. The precise mechanism of pantothenate absorption is secondary active transport in which a transmembrane protein (inappropriately called a carrier) mediates the sodium-coupled transfer of pantothenate across the brush-border membrane. The carrier spans the membrane in a weaving fashion and effects solute transfer through a conformational change in its molecular structure. The immediate energy source for the transport mechanism is the concentration gradient of sodium across the brush-border membrane. The gradient is maintained by the constant extrusion of sodium from the enterocyte by the action of the sodium pump at the basolateral membrane. The sodium pump is driven by metabolic energy and is the primary driving force for pantothenate absorption. As the transport process does not respond to an electrical gradient, it nust be electroneutral, indicating a 1:1 cotransport of Na^+ and pantothenate$^-$ by the same carrier. The mechanism by which pantothenic acid exits the absorptive cell at the basolateral membrane has not been established.

Intestinal microflora have been reported to synthesize pantothenic acid in mice, but the contribution of bacterial synthesis to body pantothenic acid levels or fecal loss in humans has not been quantified.

Unlike other water-soluble vitamins (ascorbic acid, biotin and thiamin) that are absorbed by specific carrier-mediated systems, the absorption of pantothenic acid is not regulated by its level of dietary intake. The absence of clear-cut deficiency symptoms in humans and the lack of toxicity at high doses could explain why a regulated absorption mechanism has not evolved for pantothenic acid.

Tissue Uptake and Metabolism

After absorption, free pantothenic acid is conveyed to the body tissues in the plasma from which it is taken up by most cells. A so-called sodium-dependent multivitamin transporter that mediates placental and intestinal uptake of pantothenate, biotin and the essential metabolite lipoate has been cloned from rat placenta and rabbit intestine. Messenger RNA transcripts of this transporter have been found in many tissues (intestine, liver, kidney, heart, lung, skeletal muscle, brain and placenta) suggesting that this carrier protein may be involved in the uptake of pantothenate, biotin and lipoate by all cell types.

In mammalian tissues (but not in red blood cells), CoA is synthesized from pantothenic acid in five enzymatic steps. Three substrates are needed to synthesize CoA: pantothenic acid, ATP, and cysteine. The rate-controlling step in the synthesis is the conversion of pantothenic acid to 4′-phosphopantothenic acid by pantothenate kinase. Tissue levels of CoA are kept in check by feedback inhibition of pantothenate kinase by CoA, acetyl-CoA, or a related metabolite.

In the event of a drastically reduced intake of pantothenic acid, such as would occur during food deprivation, the liver, and possibly other tissues, is able to maintain nearly constant CoA levels for some considerable time. In fasting rats, pantothenic acid uptake by the liver is stimulated by the natural rise in glucagon, and incorporation of pantothenic acid into CoA is stimulated by glucagon and cortisol.

In contrast to the liver, uptake of pantothenic acid by heart and skeletal muscle of fasting rats is reduced, and yet the rate of pantothenic acid conversion to CoA is increased. Evidently, myocardial and muscle CoA synthesis is not governed by the availability of pantothenic acid to these tissues, but rather is controlled intracellularly by regulation of enzymes involved in the CoA synthetic and/or degradative pathways.

Pantothenic acid derived from the degradation of CoA is excreted intact in urine. The amount excreted varies proportionally with dietary intake over a wide range of intake values. Both fasting and diabetes result in decreased excretion, thus conserving whole-body pantothenic acid under these conditions.

Biochemical Functions of Coenzyme A and Acyl Carrier Protein in Cellular Metabolism

A molecule of pantothenic acid is incorporated into the structures of CoA and acyl carrier protein. Though the functional sulfhydryl group of these coenzymes is not part of the pantothenate moiety, the steric configuration of pantothenic acid is important for enzymatic recognition.

Acetyl-CoA and succinyl-CoA are energy-rich thioesters that play important roles in the tricarboxylic acid cycle. Acetyl-CoA is also required for the acetylation of choline to form the neurotransmitter, acetylcholine. The amino sugars D-glucosamine and D-galactosamine react with acetyl-CoA to form acetylated products, which are structural components of various mucopolysaccharides. The biosynthesis of cholesterol begins with the condensation of two molecules of acetyl-CoA to form acetoacetyl-CoA. The latter reacts with acetyl-CoA to form 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA), which in turn is reduced to the key intermediate, mevalonic acid. CoA is required at two steps in each cycle of the β-oxidation of fatty acids. Acyl carrier protein, as an integral part of fatty acid synthase, is involved in the biosynthesis of fatty acids.

Physiological Roles of Coenzyme A in the Modification of Proteins

Many diverse cellular proteins are modified by acetylation and/or by the covalent attachment of lipids. The modifications fall into three main categories: acetylation, acylation, and isoprenylation. The alterations in protein structure may be relevant to the association of proteins with the plasma membrane or with subcellular membranes, protein–protein binding, or the targeting of proteins to specific intracellular locations. In some cases, the modifications are cotranslational, i.e., they take place on the growing polypeptide chain associated with the ribosome during protein synthesis; in other cases, they are posttranslational.

Most soluble proteins are modified at their amino termini with an acetate group that is donated by CoA. Acetylation alters the protein's binding affinity for receptors or other proteins.

A wide variety of proteins are modified with long-chain fatty acids donated by CoA. The two fatty acids most commonly attached to proteins are myristic acid (14:0) and palmitic acid (16:0). The enzyme linking myristate to amino-terminal glycine residues by an amide bond is *N*-myristoyl transferase. For myristoylation to take place, the protein substrate must have a glycine residue at position 2, immediately following methionine, and preferably a hydroxyamino acid (typically serine) at position 6. Myristoylated proteins include G protein α subunits (signal transduction), ADP-ribosylation factors (vesicular transport), myristoylated alanine-rich C kinase substrate protein (cytoskeletal rearrangements), recoverin (vision), proteins of the immune system, and several enzymes. Palmitoyl transferases link palmitate to the side chains of cysteine residues by a thioester bond. The cysteine residues can reside at any point in the primary structure of the protein; there is little evidence for any specific sequence requirements. Unlike the highly stable amide linkages to myristate, modifications of proteins by palmitate occur in thioester or oxyester linkages that are subject to hydrolysis by esterases. Cycles of palmitoylation and depalmitoylation allow the modified protein to have a regulating function. Palmitoylated proteins include G protein α subunits, many plasma membrane-anchored receptors, cytoskeletal proteins, gap junction proteins, neuronal proteins, and the enzymes acetylcholinesterase and glutamic acid decarboxylase. Palmitate modification is also a prerequisite for the budding of transport vesicles from Golgi cisternae.

Two important isoprenoids, the 15-carbon farnesyl pyrophosphate and the 20-carbon geranylgeranyl pyrophosphate (**Figure 1**), are metabolic products of mevalonic acid. Attachment of either isoprenoid chain is the first step in the modification of proteins bearing a C-A1-A2-X motif, where C is a carboxy-terminal cysteine residue, A1 and A2 are aliphatic amino acids, and X is an undefined amino acid. The attachment is a thioester bond with the terminal cysteine. Isoprenylated proteins include Ras proteins (signal transduction), Rab proteins (vesicular transport), nuclear lamins A and B (assembly and stabilization of the nuclear membrane), G protein γ subunits, and the enzymes phosphorylase kinase and rhodopsin kinase.

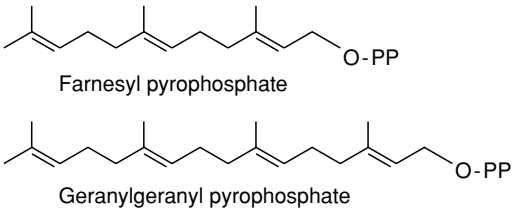

Figure 1 Structures of farnesyl pyrophosphate (C15) and geranylgeranyl pyrophosphate (C20).

The physiological implications of selected acetylation and acylation modifications of proteins are discussed below.

Acetylation of β-Endorphin

Amino-terminal acetylation plays an important role in regulating the biological activity of the brain neurotransmitter β-endorphin. This peptide has morphine-like analgesic activity and also affects sexual behavior and learning. Acetylation deactivates *β*-endorphin by rendering it unable to bind to specific receptors. The modification is posttranslational and occurs before or during the packaging of the peptide into the secretory granules of multineurotransmitter neurons in the pituitary gland.

Histone Acetylation

The DNA in cell nuclei does not exist in the 'naked' state – rather, it is compacted into chromatin by winding around specific DNA-binding proteins called histones. The fundamental repeating unit of chromatin is the nucleosome, which appears in electron micrographs as beads on a string. Each nucleosome consists of core histones (H2A, H2B, H3, and H4), linker histones (H1 or variants thereof) and variable lengths of linker DNA. Two molecules each of the core histones form a barrel-shaped nucleosome core particle, around which 146 base pairs of DNA are wrapped in nearly two complete turns. A model of the octamer of core histones is shown in Figure 2. The linker histone acts as a clamp, preventing the unwinding of DNA from the octameric complex. Each of the four types of core histone comprises a globular, hydrophobic carboxy terminus and an extended hydrophilic amino-terminal tail containing a number of positively charged amino acid residues. The tails lie on the outside of the nucleosome, where they can interact ionically with the negatively charged phosphate groups of the DNA backbone. During the periods between cell division, the 'beads on a string' chromatin filaments form higher-order structures by winding into a solenoid containg six nucleosomes per turn. In these structures, the tails of the core histones still extend outside the nucleosome.

The organization of chromatin into nucleosomes is an essential feature in the regulation of gene transcription – the step in protein synthesis in which messenger RNA is synthesized from DNA. During transcription, the enzyme RNA polymerase II combines with a host of protein transcription factors to form a multiprotein complex at a precise site on the DNA called the promoter. The polymerase moves along the DNA, temporarily unwinding and separating the two strands. As it moves along, RNA is formed by the linking of ribonucleotides under the influence of the enzyme and using one of the DNA strands as a template.

It is necessary to control gene transcription so that only those proteins needed by a particular cell for a specific purpose are synthesized. When a protein is not needed, nucleosomes prevent transcription by impeding the access of factors required to initiate and regulate this process. When protein synthesis is required, changes in cell physiology cause a partial and localized alteration of chromatin structure (chromatin remodeling) in a manner that permits the binding of initiating and regulatory factors.

One important chromatin remodeling system involves the post-translational modification of core histones by acetylation. Nuclear histone acetyltransferases (HATs) catalyze the transfer of an acetyl group from acetyl-CoA on to the ε-amino group of specific lysine residues present exclusively in the amino-terminal tails of each of the core histones. Neutralization of the positively charged lysines reduces the net positive charge of the histone tails and weakens their association with DNA. The displacement of the flexible tails permits subtle changes in nucleosomal structure and a partial unwinding or loosening of the core DNA. The result is an increase in accessibility of transcription factors to their DNA-binding sites. Acetylation does not occur randomly; multiple HATs have specificities for different lysines in the histone tails. Histone deacetylases (HDACs) counter the effects of HATs by restoring the nucleosomes to their transcriptionally repressive configurations.

In the overall scheme (Figure 3), the chromatin structure is transiently and reversibly altered to allow or prevent access of the transcription factors by targeting HATs or HDACs to the core promoter, thereby activating or repressing transcription, respectively. It is now clear that transcriptional activators function by recruiting coactivators, and it is the coactivators that possess HAT activity. Repressors inhibit transcription indirectly by recruiting HDACs via a bridging corepressor.

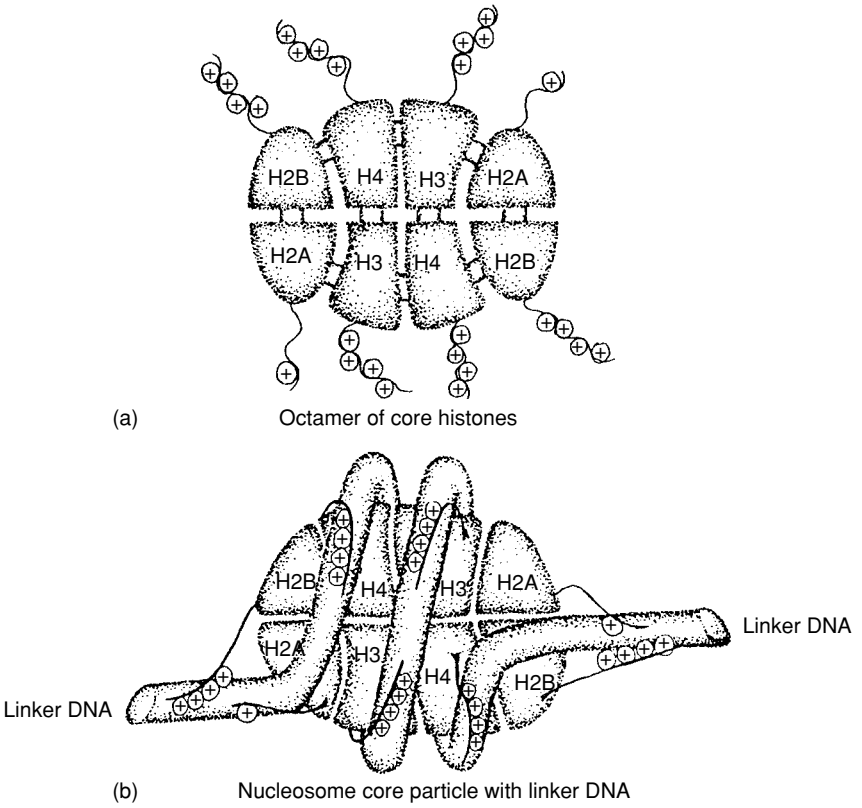

Figure 2 Models for (a) the octamer of core histones, and (b) the nucleosome core particle with linker DNA. From Csordas A (1990) *Biochemical Journal* 265: 23–38 with permission.

α-Tubulin Acetylation

Microtubules are long, stiff, hollow cylinders composed of polymerized α- and β-tubulin dimers. As constituents of the cytoskeleton, microtubules provide structural support for the cell. They also act as lines of transport for the organized movement of mitochondria and other organelles to desired locations within the cell, facilitate delivery of transport vesicles from the Golgi complex to the apical membrane in epithelial cells, and become associated with the centrioles and chromosomes to form the spindle during mitosis and meiosis (cell division). A subset of the α-tubulin is modified, like the histones, by post-translational acetylation of the ε-amino group of specific lysine residues. In contrast to histone acetylation, the acetylation of α-tubulin stabilizes the polymeric structure of the microtubule; deacetylation is coupled to depolymerization.

Acylation of G Proteins

Peptide hormones, being hydrophilic, cannot cross the lipid bilayer of the cell plasma membrane. To overcome this problem, the hormones bind to specific cell surface receptors, and a member of the family of guanine nucleotide-binding regulatory proteins (G proteins) acts as a signal transducer in coupling these receptors to intracellular effector proteins – enzymes that generate the second messenger (e.g., cyclic 3′,5′-adenosine monophosphate, cAMP). The second messenger mediates the biological action of the hormone through the activation of protein kinases.

The posttranslational attachment of palmitate to the G protein α subunit provides the means for reversible translocation of the subunit between the plasma membrane and the cytoplasm. The model shown in (Figure 4) applies to the α_s subunit responsible for stimulation of cAMP synthesis. In the unactivated state, the palmitoylated α subunit–GDP complex, α_{pal}–GDP, is associated with the β/γ subunits and the plasma membrane. Receptor activation stimulates release of GDP and binding of GTP to form active α_{pal}–GTP; the α and β/γ subunits dissociate from each other but remain at the plasma membrane by virtue of their respective palmitate and isoprenyl attachments. Palmitate is rapidly cleaved from α_{pal}–GTP by a palmitoyl esterase, and the depalmitoylated α subunit is released from the membrane into the cytoplasm. Intrinsic GTP hydrolysis converts the active GTP-bound subunits in both membrane and cytoplasm

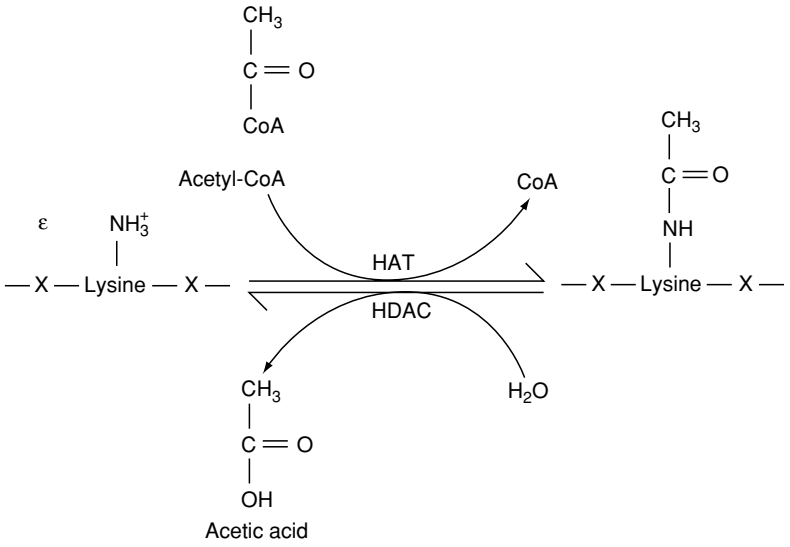

Figure 3 Opposing activities of histone acetyltransferases (HAT) and histone deacetylases (HDAC) in the control of transcription through chromatin remodeling.

into the inactive GDP-bound forms. Reattachment of palmitate to the cytoplasmic subunit by a palmitoyl transferase facilitates the return of the α_{pal}–GDP to the plasma membrane.

The G protein α subunit is further modified by the cotranslational attachment of myristic acid. This increases the affinity of the α subunit with the β/γ subunits, with the plasma membrane, and with the effector protein.

Palmitoylation of Asialoglycoprotein Receptors

The hepatic asialoglycoprotein receptor mediates the endocytosis of desialylated glycoproteins containing terminal galactose or *N*-acetylgalactosamine. (Endocytosis refers to the cellular uptake of macromolecules by entrapment within inward foldings of the plasma membrane, which then pinch off to form intracellular vesicles.) There is evidence that a cycle of palmitoylation and depalmitoylation regulates the ligand-binding activity of the asialoglycoprotein receptor. Inactivation of the receptor by depalmitoylation prevents the rebinding of dissociated ligand molecules and ensures that ligand is shuttled to lysosomes for degradation rather than nonproductively recycled back to the cell surface.

Deficiency in Animals and Humans

Pantothenic acid deficiency has been induced experimentally in many species of animals and birds by feeding diets containing low levels of the vitamin. The wide range of deficiency signs, histopathological abnormalities, and metabolic changes indicate disorders of the nervous system, reproductive system, gastrointestinal tract, and immune system. Rodents are particularly prone to necrosis and hemorrhage of the adrenal glands with consequent impairment of adrenal endocrine function. In young animals, the earliest sign of deficiency is a decline in the rate of growth. Distinctive visible signs are depigmentation of fur in rats and mice, and rough plumage and exudative lesions around the beak and eyelids of chickens. 'Goose-stepping' of the hind legs in pigs and ataxia (falling down to one side) in chicks are associated with demyelination of the motor neurons.

Human pantothenic acid deficiency has been carefully studied in healthy male volunteers given an emulsified artificial diet by stomach tube. In one study, two subjects received the basic diet devoid of pantothenic acid, a second pair received the same diet with added antagonist (ω-methyl pantothenic acid), and a third pair (the controls) received the diet supplemented with pantothenic acid. After about 4 weeks, subjects receiving the antagonist and those in the deficient group began to show similar symptoms of illness. Clinical observations were irritability, restlessness, drowsiness, insomnia, impaired motor coordination, and neurological manifestations such as numbness and 'burning feet' syndrome. The most persistent and troublesome symptoms were fatigue, headache, and the sensation of weakness. Among the laboratory tests, the loss of eosinopenic response to adrenocorticotropic hormone indicated adrenocortical insufficiency.

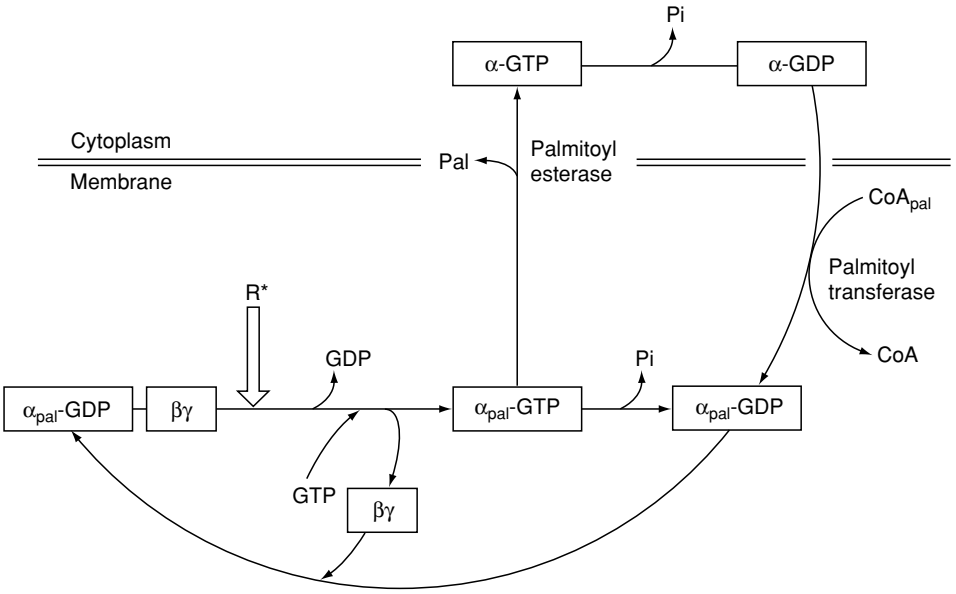

Figure 4 Model of $G\alpha_s$ palmitoylation and depalmitoylation as a means for reversible translocation of the subunit between the plasma membrane and the cytoplasm. R* indicates receptor activation. Modified from Wedegaertner PB and Bourne HR (1994) Activation and depalmitoylation of $G_{s\alpha}$. *Cell* 77: 1063–1070, with permission.

Dietary Intake

A recommended dietary allowance (RDA) for a nutrient is derived from an estimated average requirement (EAR), which is an estimate of the intake at which the risk of inadequacy to an individual is 50%. In the case of pantothenic acid, no data have been found on which to base an EAR, and an adequate intake (AI) is used instead of an RDA by the Food and Nutrition Board of the US Institute of Medicine. The AI for infants up to 12 months old (1.7–1.8 mg day^{-1}) reflects the observed mean intake of breastfed infants. The AI for children aged 1–3 years (2 mg day^{-1}) is extrapolated from adult values. The AIs for children aged 4–13 years (3–4 mg day^{-1}), and adolescents and adults of both sexes (5 mg day^{-1}) are based on pantothenic acid intake sufficient to replace urinary excretion. AIs for women during pregnancy and lactation are 6 and 7 mg day^{-1}, respectively.

There are no known toxic effects of oral pantothenic acid in humans or animals.

See also: **Fatty Acids**: Metabolism; **Oxidative Phosphorylation**; **Tricarboxylic Acid Cycle**

Further Reading

Casey PJ (1994) Lipid modifications of G proteins. *Current Opinion in Cell Biology* 6: 219–225.

Clarke S (1992) Protein isoprenylation and methylation at carboxyl-terminal cysteine residues. *Annual Review of Biochemistry* 61: 355–386.

Food and Nutrition Board of the Institute of Medicine (1998) *Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B_6, Folate, Vitamin B_{12}, Pantothenic Acid, Biotin, and Choline*. Washington, DC: National Academy Press.

Hassig CA and Schreiber SL (1997) Nuclear histone acetylases and deacetylases and transcriptional regulation: HATs off to HDACs. *Current Opinion in Chemical Biology* 1: 300–308.

Hodges RE, Ohlson MA and Bean WB (1958) Pantothenic acid deficiency in man. *Journal of Clinical Investigation* 37: 1642–1657.

Kadonaga JT (1998) Eukaryotic transcription: an interlaced network of transcription factors and chromatin modifying machines. *Cell* 92: 307–313.

Kuo M-H and Allis CD(1998) Roles of histone acetyltransferases and deacetylases in gene regulation. *Bioessays* 20: 615–626.

Plesofsky-Vig N (1999) Pantothenic acid. In: Shils ME, Olson JA, Shike M and Ross AC (eds) *Modern Nutrition in Health and Disease*, 9th edn, pp. 423–432. Philadelphia, PA: Lippincott Williams & Wilkins.

Robishaw JD and Neely JR (1985) Coenzyme A metabolism. *American Journal of Physiology* 248: E1–E9.

Sinensky M and Lutz RJ (1992) The prenylation of proteins. *Bioessays* 14: 25–31.

Smith CM and Song WO (1996) Comparative nutrition of pantothenic acid. *Journal of Nutritional Biochemistry* 7: 312–321.

Tahiliani AG and Beinlich CJ (1991) Pantothenic acid in health and disease. *Vitamins and Hormones* 46: 165–228.

Wedegaertner PB and Bourne HR (1994) Activation and depalmitoylation of $G_{s\alpha}$. *Cell* 77: 1063–1070.

PAPAYAS

J H Moy, University of Hawaii at Manoa, Honolulu, USA

Introduction

Papaya, a mildly sweet, melon-like tropical fruit belonging to the family Caricacae, is a native of tropical America (Central America and the Caribbean, including Mexico, Costa Rica, and the Bahamas). After the Spaniards took the fruit to Luzon Island in the Philippines in the mid sixteenth century, it reached Malacca shortly afterward, and then India. It has been widely grown throughout tropical and subtropical regions such as Australia; Hawaii, Florida, Texas, California, and Puerto Rico in the USA; Peru; Venezuela; various parts of Central and South Africa; and Bangladesh, Pakistan, and India. The Australians call it 'pawpaw' while the Venezuelans call it 'lechosa.'

Fruit size may vary from less than 0.5 kg to 3 kg. A climacteric fruit, it is mainly consumed fresh when ripened after harvest. Ripening is judged by the approximate percentage of yellowness on its skin, and more accurately by measuring its total soluble solids (TSS) contents with a refractometer. Export-grade papayas grown in Hawaii should have a minimum of 11.5% TSS. Firmness, as gaged by touch or by a texture-measuring device, is another way of judging ripening. The firmness is related to the biochemical changes in three fractions of pectins in papayas, and is not a very accurate index of ripeness. Green papayas can also be consumed as a salad or in soups, both of which are quite popular in South-east Asia.

Hawaii, part of the USA, is the world's largest exporter of papayas, with about 16 000 t in 1999, valued at US$14.15 million. Most of the outshipments are to the US mainland, with about 25% going to Japan. Brazil is the second largest exporter, with most papayas going to European countries such as Germany, France, and the UK. Kenya, the Ivory Coast, and Malaysia also ship papayas to European markets. For recent shipment statistics, (*See* **Fruits of Tropical Climates**: Commercial and Dietary Importance.)

Species and Cultivars

Papaya belongs to the family Caricaceae – a small, somewhat anomalous family with four genera and 31 species, of which three are in tropical and subtropical America and one in Africa. The genera and the number of included species are: *Carica*, 22; *Facaratia*, 6; *Farilla*, 1; and *Cylicomorpha*, 2. The edible fruits are found only in *Carica*, of which *C. papaya*, the common papaya, is extensively grown. Other species include: *C. chilensis, C. goudotiana, C. monoica*, and *C. pubesceus*. *C. candamarcensis*, known as mountain papaya, grows to a height of about 2.5 m, tolerates low temperature, and thrives well at an elevation between 1500 and 2000 m. *C. monoica* grows in the Amazon basin.

A large number of papaya cultivars are grown in different parts of the tropical and subtropical regions. However, none are true cultivars because they do not reliably transmit the parental characters to all their progeny. Some of the better known cultivars are listed below, with their average weight in kg in parentheses:

1. Washington (0.9–1.0)
2. Honey dew (1.1–1.5)
3. Coimbatore 1 (1.2–1.5) – a selection from the progenies of cv. *Ranchi*
4. Coimbatore 2 (1.0–1.5) – a high papain-yielding cultivar
5. Coimbatore 3 (0.9–1.0) – a hybrid between Co. 2 and Sunrise Solo
6. Coimbatore 4 (1.3–1.5) – a cross between Co. 1 and Washington (The coimbatore cultivars were released from the Tamil Nadu Agricultural University, Coimbatore.)
7. Pusa Delicious (1.0)
8. Pusa Majesty (0.95)
9. Pusa Giant (1.9)
10. Pusa Dwarf (1.0) (The pusa cultivars were bred at the India Agricultural Research Institute Regional Station, Pusa, Bihar.)
11. Sunrise Solo (0.4–1.0) – red-fleshed
12. Sunset Solo (0.4–1.0) – red-fleshed
13. SunUp (0.4–1.0) – a genetically transformed sunset, red-fleshed
14. Kapoho Solo (0.5–1.4) – yellow-fleshed

15. Waimanalo Solo (0.5–1.4) – yellow-fleshed
16. Rainbow (0.5–1.4) – a transgenic cross between Kapoho and SunUp, yellow-fleshed (The two transgenic cultivars have been grown on the island of Hawaii since 1998)
17. Pink-Fleshed Sweet (0.85)
18. Thailand (1.9–2.2)

Other important cultivars are: Guinea Gold, Sunnybank, and Hybrid 5 (Australia); Hortus Gold and Honey Gold (South Africa); Tai Ning No. 2, Panama red and Solo No. 1 (Taiwan); Red Rock and Cariflora (Florida, USA); Red Fleshed, (Philippines); Madhubindu, Barwani, Ranchi, and Peradeniya (India).

Figure 2 (see color plate 120) Thailand cultivar growing in Thailand, weighing 2.50 kg.

Botany and Horticulture

Papaya is a herbaceous, unbranched plant. Its stem is supported by phloem fibers encircling a hollow stem, which can grow up to 8–10 m high. Its leaves are large, deeply lobe-shaped with long hollow petioles, and spirally crown the stem. Flowers, male and female, can be on the same plant (monoecious) or on separate plants (dioecious), or as male and female parts on the same flower (hermaphrodite).

Papaya fruits are mostly round in the female tree and cylindrical and pear-shaped in the hermaphrodite tree. Underneath the smooth, thin skin (green when immature, orange-yellow when ripe) is a thick layer of deep yellow to orange-red pericarp with an elliptical, central cavity containing many small, round black seeds coated with jelly-like tissues (Figures 1 and 2).

Planting

The three major environmental factors to consider in selecting a site to grow papayas are temperature, moisture (rainfall and soil drainage), and wind. The hermaphrodite papaya plant preferred for commercial orchards is more sensitive to its growing environment than the female papaya plant. Selection of a suitable site is therefore critical. Another condition to consider is the amount of sunlight the site receives to support plant growth and fruit production. Insufficient sunlight results in fruits with inadequate sugar and low yields, and encourages plant diseases that affect papaya production.

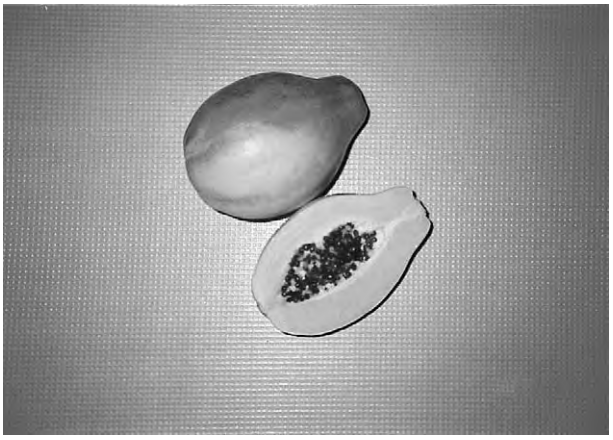

Figure 1 (see color plate 119) Rainbow cultivar growing in Hawaii, weighing 0.66 kg.

The temperature of the site is the most important factor. Optimal production is generally limited to elevation under 160 m within a temperature range of 16–31 °C. Temperature below 16 °C could cause carpeloidy, resulting in 'cat-face' deformity when floral stamens develop abnormally into fleshy, carpel-like structures. High temperature (32–35 °C) may induce female sterility, in which normally hermaphroditic papaya plants produce male flowers, resulting in poor fruit set and production.

A minimum monthly rainfall of 10 cm and an average relative humidity of 66% are considered ideal for papaya growth and production. Drip irrigation can supplement low rainfall. Good soil drainage is also essential. When soil drainage is restricted, papaya is susceptible to fungal root diseases. The plants are severely affected by waterlogging and can be killed when subjected to puddled conditions for even a few hours. The Puna area on the island of Hawaii is well-suited to commercial papaya production because of the very porous lava soils.

Papaya plants must be protected from wind. Plants exposed to constant wind develop deformed, crinkled leaves. When wind stress damage is excessive, the plant growth, fruit set, fruit quality, and productivity will be reduced. Wind-blown dust can cause sap bleeding that harms fruit appearance. In coastal

areas, salt spray carried by wind can desiccate leaves and kill papaya plants. Winds of 64 km h^{-1} can uproot papaya trees growing in mineral soils, especially when accompanied by heavy rain. Windbreaks should be established well in advance of planting a papaya crop. On the other hand, adequate air movement is important in reducing incidence of fungal diseases such as *phytophthora* and *anthracnose*. These diseases can become severe when there is excessive free moisture and high humidity around the plant.

Papaya can either be seeded directly or transplanted into a new field. Direct seeding can be practiced in porous soil. Six-week-old seedlings can be transplanted into less porous fields. Planting in 'virgin' lands or fields in which papaya has not been grown before is preferred because of low disease and insect pressure. However, it is becoming increasingly difficult to find such fields. Other important horticultural aspects of planting papaya are plant sex selection, thinning, fertilizer applications, leaf trimming, weed control, pest management, and disease control.

Transgenic Cultivars

For more than four decades, papaya production in the Hawaiian island chain was severely affected by the papaya ringspot virus (PRV). In the 1960s, papaya planting was forced to move from the island of Oahu to the Puna area of the island of Hawaii. Since 1992, the introduction and subsequent spread of PRV in commercial orchards in Puna has resulted in a significant decrease in papaya production in the state of Hawaii.

The only solution seemed to be to produce a transgenic cultivar that would be resistant to the virus. Faced with this extremely challenging task, molecular biologists from Cornell University, the University of Hawaii, and the Upjoin Company collaborated from 1987 to 1991 to obtain the first transformed plant that appeared to have PRV resistance. It took several more years for plant breeders at the University of Hawaii to inbreed the resistant plant to produce a tree-breeding, redfleshed cultivar that was named UH sunup (also referred to as *SunUp*). Because the papaya industry wanted a yellow-fleshed fruit, plant breeder Richard Manshardt crossed *SunUp* with *Kapoho* to produce a yellow-fleshed F_1 hybrid that was named *UH Rainbow* (also called *Rainbow*).

SunUp is totally resistant to the virus, but *Rainbow*, in greenhouse studies, has shown susceptibility to PRV until about 7 weeks of age. This is referred to as young plant susceptibility. The plants become fully resistant to PRV after 3 months of age. The hybrid *Rainbow* has some differences from the *Kapoho*: (1) a lower sex-segregation ratio; (2) plants grown from seeds of F_1 hybrids do not breed true; (3) a higher sensitivity to calcium deficiency; (4) more sensitive to *phythophthora* and will require timely application of fungicides; and (5) ripens faster than *Kapoho* but slower than *Sunrise* fruits.

These two lines of transgenic cultivars have been rigorously reviewed and approved by three US agencies: the US Department of Agriculture (USDA), the Environmental Protection Agency (EPA), and the Food and Drug Administration (FDA). The USDA determined that these new plants would not contaminate other existing germ plasma, and they would not foster a new strain of virus. The USDA concluded, therefore, that they posed no hazards to agriculture in the USA. The EPA sets tolerance limits for regulated chemicals. The coat protein used in the genetic engineering process was considered a pesticide, to be regulated because its function is to eliminate diseases. The EPA subsequently agreed with University of Hawaii scientists that there was no tolerance required for coat protein. The FDA's job is to protect the public from unsafe products. The FDA was satisfied that the nutrients of the original material (fruit) were retained, and therefore did not require a toxicity test. All three agencies declared the two transgenic cultivars safe for human consumption.

Those wishing to purchase seeds of transgenic papaya cultivars must adhere to a licensing process as well as sign a sublicense agreement with the Papaya Administrative Committee in the state of Hawaii.

Nutrient Composition and Fruit Chemistry

Nutrient Composition

The papaya fruit is a good source of ascorbic acid (vitamin C) and carotenoids (provitamin A), two important nutrients for people in the subtropical and tropical regions. Table 1 shows the nutrient composition of papaya from data published in 1965, 1990, and 1999.

Chemistry and Biochemistry

Sugars are the principal carbohydrates in papayas, with very little starch. Refractometric measurements of total soluble solids (%TSS) show some differences in various cultivars. Those grown in Florida have been reported to range from 5.6 to 7.2%; those in India vary from 6.5 to 13%, while the Hawaiian cultivars measure 11.5–13.5%. The sugars in ripe papayas in which the invertase has been inactivated were reported to be 48% sucrose, 30% glucose, and 22% fructose.

Papaya is a low-acid fruit, with the total titratable acidity of about 0.1% calculated as citric acid. The

Table 1 Nutrient composition of papaya per 100 g of edible pulp

Moisture (%)	86–89
Carbohydrate (g)	9.5–12.2
Protein (N × 6.25: g)	0.36–0.5
Fat (g)	0.06–0.1
Fiber (g)	0.5–0.6
Ash (g)	0.5–0.6
Ascorbic acid (mg)	40–84
Vitamin A (mg)	11–32[a]
Thiamin (mg)	0.027–0.04
Riboflavin (mg)	0.043–0.25
Niacin (mg)	0.20–0.33
Calcium (mg)	10–30
Phosphorus (mg)	10–12
Iron (mg)	0.2–4.0
Energy (cal)	40–48

[a]Vitamin A data assuming 12 μg of *all-trans* β-carotene = 1 μg *all-trans* retinol.

Sources of data: Wenkam NS and Miller: CD (1965) *Composition of Hawaii Fruits*. Bulletin 135. Honolulus, Hawaii: Hawaii Agricultural Experiment Station, University of Hawaii at Manoa; Muthukrishnan CR and Irulappan I (1990) In: Bose TK and Mitra SK (eds) *Fruits*: Tropical and Subtropical, pp. 303–335. Calcutta: Naya Prokash. Moy JH, Paull RE, Bian X, Chung R and Wong L (1999) Quality of tropical fruits irradiated as a quarantine treatment. In: Moy JH and Wong L (eds). *Proceedings of the workshop on the use of Irradiation as a Quarantine Treatment of Food and Agricultural Commodities* pp. 45–53. Honolulu, Hawaii: College of Tropical Agriculture and Human Resources, University of Hawaii and Dept of Hawaii, State of Hawaii.

pH of the pulp ranges from 5.5 to 5.9, far higher than the pHs of other tropical fruits at 3.2–4.5, which explains the low tartness of the papaya fruit. The organic acids in papayas are mainly equal amounts of malic and citric acid, with smaller amounts of ascorbic and α-ketoglutaric acid.

The volatile flavors of papayas were reported to consist of 124 compounds. Linalool is the major component with characteristic fresh papaya aroma and flavor. Benzyl isothiocyanate, another major component, has a pungent off-aroma. Other off-aroma, off-flavor compounds in papaya pulp have been identified as butyric, hexanoic, and octanoic acids and their corresponding methyl esters. These components were analyzed with gas chromatography and mass spectrometry. (*See* **Chromatography**: Gas Chromatography; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Mass Spectrometry**: Principles and Instrumentation.)

Carotenoids are the pigments of the ripe papaya flesh, which is deep yellow. Red-fleshed papaya cultivars additionally contain lycopene, which gives the pulp an orange-red color. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments.)

In addition to papain being an economically important enzyme in green papaya, several other enzymes play a role in the stability and quality of processed papaya products. Pectinesterase (3.1.1.11), the enzyme responsible for gel formation in unheated papaya purée, contributes to some increase in acidity in purée as a result of demethoxylation of the carboxyl groups in the pectins. Thioglucosidase (myrosinase) (3.2.3.1) is responsible for generating benzyl isothiocyanate, a sulfurous, pungent, odiferous compound which contributes to the off-aroma of papaya products. When fruit tissues are macerated, acid phosphatase (3.1.3.2) catalyzes the hydrolysis of the P–O bond of orthophosphoric monoesters, producing ROH and phosphoric acid, which increases the acidity of the purée. Another enzyme, β-fructofuranosidase (invertase) (3.2.1.26), is responsible for the hydrolysis of sucrose to fructose and glucose.

Harvesting, Handling, and Storage

Summer fruits become mature in 22 weeks, while 26 weeks are needed during the winter. Fruits are typically harvested weekly or twice weekly at color-break (a tinge of yellowness on a green fruit) by hand, or with the aid of a cut-off 'plumber's helper' (a rubber suction cup) attached to a pole for hard-to-reach fruits. Fruits can also be harvested with a ladder, or from a large fruit bin mounted on a tractor high-lift. Careful handling with minimal bruising and abrasion helps with postharvest sorting, packing, shipping, and marketing.

To control fungal decay of postharvested fruit, a dip in hot water at 49 °C for 20 min was found to be very effective. Hot-water spray with fungicide or a wax dip containing fungicide is also effective.

Storing papayas at temperatures below 7 °C after harvest can result in chill injury. The recommended storage temperature is 10 °C.

Quarantine Treatment

Fruit flies and other insect pests are prevalent in the tropics and the subtropics. Most soft tropical fruits are prone to infestation by fruit flies. A quarantine treatment approved by the plant quarantine authority of each country must be applied to fruits before they can be exported to noninfested areas. Up until 1984, chemical fumigants such as ethylene dibromide (EDB) were widely used, until tests showed them to be carcinogenic to laboratory animals. Since then, quarantine treatment options are thermal or cold treatment, or irradiation.

Papayas grown in Hawaii are infested with four species of fruit flies: the Oriental fruit fly (*Dacus dorsalis*), the Mediterranean fruit fly (*Ceratitis capitata*), the melon fly (*Dacus cucurbitae*), and the Malaysian fruit fly (*Bactrocera latifrons*). After EDB was banned, a double-dip hot-water treatment (42 °C for

30 min, 49 °C for 20 min) was used but was abandoned after 3 years because a high percentage of treated fruits became injured by the heat. Subsequently, two other thermal treatments were used. One is a modified vapor heat treatment, and the other is called a high-temperature forced-air (HTFA) procedure. Both are very similar. The difference is the relative humidity of the heated air admitted to the treatment chambers. The air in the vapor heat treatment has moisture added to the air throughout, with the relative humidity kept at 90% or above. In the HTFA method, ambient air is heated and admitted into the chamber, and only in the last hour of a 4-h treatment is the relative humidity required to be kept at 90–100%. Both methods require about 4 h to reach the endpoint, which is when the center of the fruit reaches 47.2 °C. Fruits are then cooled with a water spray for about 30 min. Several fruit packers in Hawaii are using these methods, which work quite well. However, both are time-consuming and commodity-specific. Total uniformity in heating is also difficult to achieve because of variation in fruit size, fruit ripeness (affecting thermal conductivity and heat transfer rates), and physical locations within a chamber. After heat treatment, it is common to find lumpy texture and lack of flavor in a few percent of the heat-treated fruits in each batch, caused by enzymes responsible for ripening being inactivated by heat.

Cold treatment requires that fruits be kept at 1 or 2 °C for 12–14 days to immobilize and inactivate fruit fly eggs and larvae before the fruits can be taken to the supermarkets. Not many tropical fruits can tolerate this time–temperature regime, papaya being one of them.

Though somewhat controversial, several decades of research in Hawaii and in other parts of the world have proven that irradiation is the most efficacious quarantine treatment procedure. The treatment is efficient (15–20 min in a commercial irradiator) and effective (all fruits are thoroughly irradiated to cause the fruit fly eggs and larvae to be sexually sterilized, regardless of fruit size and ripeness). Irradiated papayas ripen normally or slightly delayed, and their chemical, physical, nutrient, and sensory qualities are well retained. Radiation sources can be gamma-rays (from cobalt-60), an electron beam (with limited penetration), or X-rays (converted from e-beam). Fruits irradiated will not become radioactive because there is a limit on the energy level used. In April 1995, Hawaii became the first place in the world to use irradiation as a quarantine treatment of its papayas and other tropical fruits. The generic quarantine dose approved by the USDA is 0.25 kGy. Starting in 2001, more countries will be using this technology to treat their fruits for export markets.

Processed Products

Chunks (Refrigerated or Frozen)

After being washed, deseeded, and peeled, papayas can be cut into chunks, then refrigerated or frozen as a convenient food. It is available in some supermarkets in the salad section. Another niche market for this type of product is the airlines. A number of airlines prefer the convenience of serving these prepared fruits to their passengers. The frozen chunks are more suitable for serving with ice cream or for further manufacturing into dessert products.

Canned (Mixed with Other Fruits)

Papaya can be made into a canned product, or as a cocktail of several fruits. Chunks or dices of papayas or mixed fruits can be filled into a can and covered with hot (75 °C), acidified syrup (*c.* 40° Brix with pH adjusted to 3.6–3.8 with citric acid: 40° Brix = 40 g of sugar mixed with 60 g of water at 20 °C). After the air in the head space in the cans (*c.* 6.3–8.0 mm) has been exhausted by steam, the cans are sealed and pasteurized with steam, or in a boiling-water bath. (*See* **Canning**: Principles.)

Purée

Papaya purée is prepared in the form of a free-flowing paste, without seeds, skin, or unwanted fiber. The purée can be an intermediate product used for manufacturing several end products such as juices, nectars, jams, jellies, syrups, and dried fruit rolls or leathers. Papaya purée and other fruit purées in the USA can be manufactured with a high degree of mechanization. Several aspects of purée manufacturing are important to produce a quality product.

1. Ripe papaya fruits should be steamed for 2 min to coagulate the latex in the peel, preventing the latex from entering into the purée. Also, steaming increases purée yield by softening the outer layers of the fruit, and inactivates enzymes in the peel. The fruits are spray-cooled to remove the residual heat.
2. Fruits are sliced, crushed, and dropped into a centrifugal separator. All the skin and most of the seeds are separated. The remaining seeds are separated in a paddle pulper fitted with rubber paddles and a small screen (*c.* 0.80–0.85 mm).
3. The purée is acidified with citric acid (a 50% solution) to a pH of about 3.4–3.6 to inhibit gelatin of the purée. Acidification also helps control microbial activity and enhances the effect of subsequent heat pasteurization.

4. A smooth purée can be obtained by passing the liquid through a paddle finisher with a 0.50 mm screen to remove seed specks and undesirable fiber.
5. The purée is pasteurized in a heat exchanger at 94 °C for 2 min, then cooled to a few degrees above ambient temperature.
6. The purée can be aseptically packed in suitable flexible containers, or frozen in bulk at − 20 °C or below.

Beverages

Juice, drinks, nectars, and cordials can be prepared from papaya purée through formulation by adding appropriate amounts of water, sweeteners, and acidulants to the purée. Commonly used acidulants include citric acid, malic acid, lemon and lime juice. These beverages can be pasteurized in bottles or cans after being acidified to a pH below 4.4.

Dried Products

Several forms of dried products can be prepared from papaya slices. Since papaya is a soft fruit when fully ripe, three-quarter ripe fruits are more suitable for preparing dried slices. Drying can be carried out in a hot-air drier, a vacuum drier or a freeze drier, with increasing product quality but correspondingly higher cost. Solar drying of papaya slices is also a very practical means of preserving papaya with good quality. The final moisture content of dried papaya should be around 5–8% (wet weight basis). Good packaging is important to retain quality and to prevent insects from entering the package. University of Hawaii researchers have designed and built a continuous solar drier on the island of Kauai in Hawaii, capable of drying 450 kg of papaya slices within a 24-h period. This drier uses only renewable energy – solar and biogas.

University of Hawaii researchers have also developed a vacuum-puff freeze drying process to make high-quality juice powder, including papaya. The process involves mixing a papaya purée with a small amount of sucrose, freezing the mixture, and then placing it into a freeze drier. Initially, the purée–sucrose or juice–sucrose mixture puffs into a foam when raised to its freezing point. It then refreezes under a good vacuum. The endpoint is reached when 1% of moisture is left. The freeze-dried product is a crystalline powder, and can be rehydrated into a nectar, similar to freeze-dried coffee. Freeze-dried powders also retain the color and flavor of the original purée or juice with the right degree of sweetness.

Papain

Source

A milky latex in immature papaya fruit contains an enzyme called papain. Papain from the dried latex is in great demand on the international markets, especially in the UK and the USA.

Properties

Papain is a proteolytic enzyme, capable of hydrolyzing or breaking down protein materials.

Applications

In the food, pharmaceutical, textile, and tanning industries, papain is used as a meat tenderizer; for clearing beer; in the manufacture of cosmetics such as face creams and dental creams; in degumming silk and rayon; in the preshrinking of wool; and in tanning leather.

In the medical field, papain can be used to treat necrotic tissues, dyspepsia and other digestive ailments, ringworm and roundworm infections, skin lesions and ulcers, eczema, and other skin diseases and kidney disorders. Papain is used in detecting stomach and intestinal cancers and also in correcting diphtheria.

Production

The latex of a green papaya drips into a container after the skin is slit. Papain is the dried latex in powder form. The collection of latex is very labor-intensive. Papain production is influenced by several factors:

1. Fruit size and shape: oblong fruits 36 cm in length and 28 cm in diameter were found to give the highest papain yield. In general, papain yield increases with increasing fruit size.
2. Fruit maturity: unripe but fully grown fruits yield maximum papain, especially when fruits have been grown to 75–90 days after fruit set.
3. Season: the flow of latex is low if the temperature is below 10 °C.
4. Cultivars: in India and Sri Lanka, cultivars found to have high papain yield were Washington, Philippine, Botanist's Selection, Peradeniya and Coimbatore 1.

Yield

Papain yield varies from 1.23 to 7.45 g per fruit among nine cultivars, with the cultivar Washington recording the highest mean yield of 7.45 g per fruit. In a papaya tree, the total papain yield can vary from 150 g to 227 g.

Role of Papaya in Human Health

The flesh of papaya contains high levels of three carotenoids: β-cryptoxanthin, β-carotene, and lycopene. Different papaya cultivars vary in concentrations and ratios of these carotenoids. It has been reported that many individual carotenoids, including β-carotene and lycopene, have the ability to prevent transformation of cells to cancer phenotypes in a model system *in vitro*. No study has considered carotenoid combinations as they occur in fruits, such as papaya, and the possibility of synergistic effects via multiple inhibitory mechanisms. Research has been proposed to study the effectiveness of various combinations of carotenoids as inhibitors of cancer development in mammalian systems, and the cellular and molecular mechanisms if it proves to be effective. The implication is that fruits rich in carotenoids and lycopene such as papaya and tomato could be beneficial to human health.

See also: **Canning**: Principles; **Chromatography**: Gas Chromatography; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Fruits of Tropical Climates**: Commercial and Dietary Importance; **Mass Spectrometry**: Principles and Instrumentation

Further Reading

Bose TK and Mitra SK (eds) (1990) *Fruits: Tropical and Subtropical*, pp. 303–335. Calcutta: Naya Prokash.

Moy JH, Paull RE, Bian X, Chung R and Wong L (1999) Quality of tropical fruits irradiated as a quarantine treatment. In: Moy JH and Wong L (eds) *Proceedings of the Workshop on the Use of Irradiation as a Quarantine, Treatment of Food and Agricultural Commodities*, pp. 45–53. Honolulu, Hawaii: College of Tropical Agriculture and Human Resources, University of Hawaii, and Dept of Hawaii, State of Hawaii.

Nakasone HY and Paull RE (1998) *Tropical Fruits*, pp. 239–269. Oxford: CAB International.

Nishima M, Zee F, Ebesu R *et al*. (2000) *Papaya Production in Hawaii*, pp. 1–8. Honolulu, Hawaii: College of Tropical Agriculture and Human Resources, University of Hawaii at Manoa.

Shaw PE, Chan HT Jr and Nagy S (eds) (1998) *Tropical and Subtropical Fruits*, pp. 401–445. Auburndale, FL, USA: AgScience.

Wenkam NS and Miller CD (1965) *Composition of Hawaii Fruits*. Bulletin 135. Honolulu, Hawaii: Hawaii Agricultural Experiment Station, University of Hawaii at Manoa.

PARASITES

Contents
Occurrence and Detection
Illness and Treatment

Occurrence and Detection

J Melo-Cristino, Faculdade de Medicina de Lisboa, Lisbon, Portugal
J Botas, formerly of Hospital de Santa Maria, Lisbon, Portugal

Importance of Parasites in Human Infection

Parasites are a heterogeneous group of invertebrate animals, including unicellular microrganisms (protozoa) and multicellular organisms with organ systems (helminths), which can infect a diversity of other animals, including humans.

Human infections are frequently acquired by ingestion of contaminated water or food. They are important all over the world and their prevalence can be very high, especially in tropical and subtropical regions. Many parasites have a worldwide distribution, but others occur in limited endemic areas. However, frequent and rapid travel of people from nonendemic areas (tourists, scholars, business people, military personnel, immigrants, etc.) returning from visits to endemic areas contributes to infection by such parasites.

The spectrum of human disease is extremely wide. Many parasites are not noticed, and the infection is asymptomatic, while others are major human pathogens and are responsible for high morbidity and mortality. Infections may have an acute onset with intense symptoms or may require months or years before becoming clinically evident. Recently, some parasites

have become more important as they have been recognized as major pathogens in immunocompromised hosts, particularly in patients with acquired immune deficiency syndrome (AIDS). (*See* **HIV Disease and Nutrition.**)

Protozoa Involved in Foodborne and Waterborne Human Infection

The most significant foodborne and waterborne protozoa are shown in Table 1. They are classified into phylum Sarcomastigophora (amoebae and flagellates), phylum Ciliophora (ciliates), phylum Apicomplexa (coccidia), and phylum Microspora (microsporidia). Characteristics for differentiation include motility and stages in life cycles and replication.

Motility is accomplished by different mechanisms, namely pseudopodia in amoebae, flagella in flagellates, and movement of rows of cilia in ciliates. Coccidia and microsporidia are essentially nonmotile obligate intracellular parasites.

The life cycles of amoebae, flagellates, and ciliates include two stages, the trophozoite and the cyst. The trophozoite is the vegetative, motile, feeding stage usually found in the intestine. The cyst is the resting, resistant, thick-walled infective stage excreted in feces. Replication is accomplished by binary fission of trophozoites or by development of several trophozoites inside mature cysts.

Coccidial life cycles include stages of asexual development (trophozoite, schizont or meront and merozoite) and stages of sexual differentiation (microgamont and macrogamont) that lead to the production of oocysts, which are shed in feces. The oocyst, containing sporozoites, which may require a period of maturation outside the host, is the resistant and infective stage. Microsporidia multiply by binary fission or multiple fission, producing spores which are excreted in urine or in feces. The spore possessing the infective agent or sporoplasm is the resistant stage.

Helminths Involved in Foodborne and Waterborne Human Infection

The most significant foodborne and waterborne helminths are shown in Table 2. Helminths comprise the phylum Nematoda (nematodes) and the phylum Platyhelminthes (trematodes and cestodes).

Nematodes or roundworms have cylindrical bodies, separate sexes and a complete digestive system. Life cycles can be simple and direct or complex with one or more intermediate hosts. Hosts are defined as intermediate hosts when they harbor an asexual stage of the parasite and definitive hosts when they harbor the sexual stage.

Trematodes or flukes have flattened leaf-shaped bodies, and most are hermaphroditic with an incomplete digestive system. They have complex life cycles, with snails serving as first intermediate hosts and other aquatic animals as second intermediate hosts. Some trematodes may also encyst on aquatic plants.

Cestodes or tapeworms have segmented ribbon-like bodies, are hermaphroditic, and lack a digestive system. Some have direct life cycles, while others need one or more intermediate hosts. Infective stages of helminths include eggs laid by adult worms and asexual stages encysted in animal tissues or on plants.

Occurrence of Parasites in Foods and Water and Mechanisms of Entry into the Food Chain

Parasites have access to food and water by two routes. One is by contamination with the feces of infected humans or animals passing infective resistant stages such as cysts and oocysts of protozoa and eggs of helminths. Another route exists when an infective

Table 1 Foodborne and waterborne protozoa causing human infection

Amoebae	*Flagellates*	*Ciliates*	*Coccidia*	*Microsporidia*
Endolimax nana[a]	*Chilomastix mesnili*[a]	*Balantidium coli*	*Cyclospora cayetanensis*	*Encephalitozoon* sp.
Entamoeba coli[a]	*Dientamoeba fragilis*		*Cryptosporidium* spp.[b]	*Enterocytozoon* sp.
Entamoeba hartmanii[a]	*Enteromonas hominis*[a]		*Isospora belli*[b]	*Nosema* sp.
Entamoeba histolytica/dispar[c]	*Giardia intestinalis*		*Toxoplasma gondii*[b]	*Pleistophora* sp.
Iodamoeba bütschlii[a]	*Retortamonas intestinalis*[a]		*Sarcocystis* spp.	
Blastocystis hominis[d]	*Trichomonas hominis*[a]			
	Trichomonas tenax[a]			

[a]Nonpathogenic for immunocompetent humans.
[b]Infection often serious in immunocompromised patients: most human infections are with *C. parvum*.
[c]*E. histolytica* and *E. dispar* are morphologically identical, but only *E. histolytica* causes disease.
[d]Inclusion in the amoebae group has recently been proposed by some authors.

Table 2 Foodborne and waterborne helminths causing human infection

Nematodes	*Trematodes*	*Cestodes*
Angiostrongylus spp.	*Clonorchis sinensis*	*Diphyllobothrium* spp.
Anisakis spp.	*Fasciola hepatica*	*Dipylidium caninum*
Ascaris lumbricoides	*Fasciolopsis buski*	*Echinococcus granulosus*
Bayliascaris procyonis	*Heterophyes heterophyes*	*Echinococcus multilocularis*
Capillaria hepatica	*Metagonimus yokogawai*	*Hymenolepis diminuta*
Capillaria philippinensis	*Opistorchis* spp.	*Hymenolepis nana*
Dracunculus medinensis	*Paragonimus westermani*	*Multiceps multiceps*
Enterobius vermicularis		*Spirometra* spp.
Toxocara spp.		*Taenia saginata*
Trichinella spiralis		*Taenia solium*
Trichuris trichiura		

viable stage of the life cycle of a parasite is present in animal tissues or on plants.

Environmental contamination with human or animal faeces is correlated with socioeconomic conditions and occurs more frequently in areas with poor sanitation and poor personal hygiene. However, even in developed countries, the use of animal or human feces as a fertilizer may be responsible for the contamination of natural water courses, wells or water supplies. As a consequence, fruits and vegetables may be contaminated during irrigation. Flies and other arthropods may also have an important role in passive transmission.

Prepared food may also be contaminated either by use of polluted water during preparation or by food handlers who are carriers. Asymptomatic carriage is common, and carriers may be reservoirs of the parasites and may excrete infective stages, continuously or intermittently, for long periods of time. Many excreted stages (cysts, oocysts, and eggs) are very resistant and may survive in the environment for weeks or months.

Parasites which may cause human infection transmitted by water or foods contaminated with feces include all protozoa shown in Table 1 and the helminths *Ascaris lumbricoides, Enterobius vermicularis, Trichuris trichiura, Capillaria hepatica, Hymenolepis nana, H. diminuta, Echinococcus granulosus, E. multilocularis*, and *Multiceps multiceps*.

Cryptosporidium parvum and *Cyclospora cayetanensis*, in particular, have been responsible for several large outbreaks resulting from either contaminated drinking water or contamination on food (fruits or vegetables).

Many of them are human intestinal parasites with no known animal reservoirs, including the amoebae, *Dientamoeba fragilis, Isospora belli, A. lumbricoides, En. vermicularis*, and *T. trichiura*. Therefore, their presence implies contamination with human feces. Furthermore, the presence of human commensal or nonpathogenic amoebae also implies contamination with human feces and reflects the possibility of transmission of pathogenic species.

The most significant helminths causing human infection by ingestion of foods containing viable asexual infective stages are shown in Table 3. *Toxoplasma gondii* and *Sarcocystis* spp. are protozoa which may be transmitted to humans by ingestion of infected meat.

Ingestion of raw or improperly cooked meat, as well as sausages, dried, cured, and smoked meat from animals serving as intermediate hosts, is a frequent cause of infection. Examples of parasites transmitted by beef include *To. gondii, Sarcocystis* spp., and *Taenia saginata*. Pork is implicated in the transmission of *To. gondii, Sarcocystis* spp., *Taenia solium*, and *Trichinella spiralis*. Other livestock raw meat may transmit *To. gondii*. Crustaceans involved in transmission of parasites include crabs (*Paragonimus westermani, Angiostrongylus* spp.), prawns and shrimps (*Anisakis* spp., *Angiostrongylus* spp.). (*See* **Shellfish**: Contamination and Spoilage of Molluscs and Crustaceans.)

Uncooked aquatic vegetables, where infective metacercariae encyst, such as water chestnuts and watercress may be responsible for human infection with the trematodes *Fasciolopsis buski* and *Fasciola hepatica*, respectively.

Fate on Processing/Storage

Protozoal cysts or oocysts and helminth eggs are moderately or highly resistant to chlorination at concentrations used to disinfect water, e.g., *Cryptosporidium* spp., *Giardia intestinalis, A. lumbricoides*, and *T. trichiuria*. They are, however, generally susceptible to heat and to freezing at −20 °C.

Helminth-encysted metacercariae or infective larvae in meat, fish, or crustaceans are generally killed by heating and by prolonged freezing at −20 °C. However, they often resist smoking, curing, or pickling, e.g., *Tr. spiralis*.

Table 3 Helminths transmitted by foods containing asexual infective stages

Food	*Nematodes*	*Trematodes*	*Cestodes*
Meat	*Trichinella spiralis*		*Taenia saginata* *Taenia solium*
Fish	*Anisakis* spp. *Capillaria philippinensis*	*Clonorchis sinensis* *Heterophyes heterophyes* *Metagonimus yokogawai* *Opistorchis viverrini* *Paragonimus westermani*	*Diphyllobothrium latum* *Spirometra* spp.
Crustaceans	*Angiostrongylus* spp. *Anisakis* spp.	*Paragonimus westermani*	
Vegetables		*Fasciola hepatica* *Fasciolopsis buski*	

Detection in Foods and Water

Incrimination of foods and water in the transmission of parasites is almost always indirect and based on epidemiological association. Direct isolation of these parasites from foods or water is very difficult. Water-filtration techniques for detection of cysts, oocysts, or eggs may be used, but too often with poor results. Enrichment culture media are not currently available in microbiology laboratories. Thus, it is not possible to recover small numbers of parasites from foods or water by *in-vitro* culture methods. Furthermore, the viability and infectivity of a parasitic stage cannot be assessed.

Direct incrimination of food in human infection is obvious when ingestion of an infected food is the only possible way for the infection to occur. All parasites in Table 3 belong to this group. Encysted larvae of *Tr. spiralis*, *Ta. solium*, and *Ta. saginata* may be detected in raw meat after specific procedures, such as enzymatic digestion. A similar approach to raw fish may have little practical value. Other helminths, e.g., *E. granulosus* and *F. hepatica*, may be detected by the examination of viscera from livestock animals.

Detection in Humans

The diagnosis of a parasitic infection can be made directly by the finding and identification of the parasite, or indirectly by immunological methods detecting specific antibodies. DNA probes, other molecular techniques, and immunological methods using monoclonal antibodies for the detection of specific parasitic antigens in serum and in body tissues have recently been described and are promising powerful diagnostic tools for the future.

Intestinal parasites are currently diagnosed by morphological identification of trophozoites, cysts, oocysts, eggs, or adult worms in feces, whereas tissue parasites are generally diagnosed by immunological methods.

Microscopic examination of fecal specimens by direct wet mount, wet mount after concentration and/or permanent stains, is the most important method for the diagnosis of intestinal or biliary parasites. As excretion is variable, at least three specimens should be examined, the specimens being collected at 2- or 3-day intervals.

Tissue parasites may occasionally be diagnosed by direct detection and identification in biopsy material from lesions. History of traveling into endemic areas combined with appropriate clinical symptoms and radiological, ultrasound, or computed tomography examinations may be very important diagnostic clues. However, detection of rising titers of specific antibodies is the most useful way to establish the diagnosis. Current immunological methods include complement fixation indirect immunofluorescence, indirect hemmaglutination, enzyme-linked immunosorbent assay, immunoelectrophoresis, and double-diffusion tests. In immunocompromised patients, these tests are less satisfactory for the diagnosis of infection. Recently, several easy-to-use antigen detection tests have become available and are of particular value in the diagnosis of *Giardia intestinalis*, *Cryptosporidium parvum*, and *Entamoeba histolytica*. (*See* **Immunoassays**: Principles; Radioimmunoassay and Enzyme Immunoassay.)

Specific Examples

Cryptosporidium spp.

Cryptosporidium spp. have a worldwide distribution and are responsible for enteric infection in humans and many animals, especially cattle and sheep. Infection is acquired by ingestion of water or food containing oocysts, which are excreted in an infective form that does not need maturation outside the gut. Transmission can occur from animals to humans (zoonotic) as well as from humans to humans (nonzoonotic).

Oocysts are highly resistant to chlorination but are sensitive to heat and prolonged freezing. Diagnosis is by microscopic detection of oocysts in fecal specimens stained by the acid-fast technique or immunofluorescence, as well as by antigen detection assays.

Giardia intestinalis

G. intestinalis has a worldwide distribution and is responsible for the most frequently reported human protozoal infection. Giardias is an intestinal infection associated with poor socioeconomic conditions. Infection is acquired by ingestion of food or water contaminated with feces. Cysts are the infective stage excreted in feces of humans and other animals. Asymptomatic carriage and excretion are frequent. Cysts can be found in water, sewage, vegetables, fruits, and other food. They resist chlorination but are killed by heat and by prolonged freezing. Diagnosis is by microscopic identification of cysts in fecal specimens. Trophozoites may sometimes be detected in feces from patients with diarrhea.

Toxoplasma gondii

Toxoplasmosis is a worldwide zoonosis. Prevalence of human infection increases with age and is generally 30–70%. Infection is acquired by ingestion of infective oocysts from cat fecal contamination or by ingestion of improperly cooked meat from animals serving as intermediate hosts, e.g., pork, mutton, and beef. Oocysts need a period of maturation outside the cat bowel. Cysts in meat are killed by heating, by smoking or curing, and by prolonged freezing at −20 °C. Diagnosis is by demonstration of specific immunoglobulin (IgM) antibodies or increasing IgG antibody titers in blood by immunological tests. These tests are not satisfactory in immunocompromised patients with latent or reactivated infections.

Ascaris lumbricoides

A. lumbricoides has a worldwide distribution but is more prevalent in areas of poor sanitation, with 25% of the world's population estimated to be infected. It is the most common helminthic parasite causing human infection. No animal reservoir is known. Infection is acquired by ingestion of water or food contaminated with embryonated eggs from human feces. *A. lumbricoides* eggs need a maturation period outside the gut and can remain infectious for months. Diagnosis is by microscopic detection of eggs in fecal specimens. Occasionally, adult worms are eliminated in feces.

Taenia solium

Ta. solium is prevalent in Africa, South-east Asia, Central and South America, eastern Europe, Spain, and Portugal. The human is the definitive host and the pig the intermediate host. Human intestinal infection is acquired by ingestion of undercooked pork containing a larval stage called the cysticercus.

Humans can also harbor the larval stage of *Ta. solium* in tissues. This potentially severe disease, called cysticercosis, occurs after ingestion of water or foods contaminated with *Ta. solium* eggs (from human feces) or by autoinfection when eggs from adult worms hatch in the intestine, producing a larval form that penetrates the intestinal wall and enters the circulation to reach structures such as muscle, brain, lungs, and eyes. Diagnosis is by macroscopic detection of segments of the adult worm (proglottids) or by microscopic detection of eggs in fecal specimens. Cysticercosis may be diagnosed by radiological or scanning examinations and by detection of rising titers of specific antibodies. Sometimes, identification of cysts after surgery may be the only means of definitive diagnosis.

Trichinella spiralis

Trichinosis, the infection due to *Tr. spiralis*, is a disease of carnivorous animals which accidentally occurs in humans. It has a worldwide distribution but is more prevalent in temperate regions. Human infection is acquired by ingestion of raw or poorly cooked pork or pork products containing encysted larvae. Laboratory diagnosis is confirmed by the detection of spiral larvae in biopsy muscle tissue from the patient, sometimes after trypsin digestion of muscle fibers. Demonstration of rising titers of specific antibodies is also available for diagnosis.

See also: **HIV Disease and Nutrition**; **Immunoassays**: Principles; Radioimmunoassay and Enzyme Immunoassay; **Shellfish**: Contamination and Spoilage of Molluscs and Crustaceans

Further Reading

Casemore DP (1990) Foodborne protozoal infection. *Lancet* 336: 1427–1432.

Gamble HR and Murrell KD (1998) Detection of parasites in food. *Parasitology* 117 (supplement): S97–S111.

Herwaldt BL (2000) *Cyclospora cayetanensis*: a review, focusing on the outbreaks of cyclosporiasis in the 1990s. *Clinical Infectious Diseases* 31(4): 1040–1057.

Smith HV (1998) Detection of parasites in the environment. *Parasitology* 117 (supplement): S113–S141.

Todd EC (1997) Epidemiology of foodborne diseases: a worldwide review. *World Health Statistics Quarterly* 50(1–2): 30–50.

Illness and Treatment

J Botas (formerly of) **and J Melo-Cristino**, Hospital de Santo Maria, Lisboa, Portugal

Protozoal Infections

Illness

Protozoa are a major cause of morbidity and mortality worldwide, either in developing or in western countries. Some of those able to infect humans by the oral route have long been known to cause disease, but others are only now being increasingly recognized as pathogenic in immunocompromised hosts, especially in patients with acquired immune deficiency syndrome (AIDS). They are very diverse in epidemiology, symptomatology, treatment, and preventive measures.

In immunocompetent hosts most infections are asymptomatic, but overt local or systemic symptoms may be present, which may be severe in some cases. However, protozoal infections are particularly troublesome, sometimes life-threatening, in immunocompromised patients.

Acute, self-limited diarrhea is the most common symptom caused by protozoa listed in Table 1, but sometimes more intense manifestations do occur, including abdominal cramps, flatulence, anorexia, vomiting, malabsorption, weight loss, and fever. In AIDS patients, insuperable chronic watery diarrhea with malabsorption and, eventually, wasting syndrome may be due to some of these parasites.

Blood, pus, and mucus in stools are the result of invasion of the intestinal wall and common only in intestinal infection by *Entamoeba histolytica*, which is also able to cause systemic infection, especially liver abscesses, manifested by fever, weight loss, abdominal pain, and hepatomegaly.

Humans may be either intermediate or definitive hosts for *Sarcocystis* spp. and, while intestinal infection is thought to be asymptomatic, parasites in muscle may cause swelling and pain. Microsporidia have only recently been recognized as human pathogens in AIDS and very few cases have been reported in non-AIDS patients. *Pleistophora* spp. has been found in muscle biopsy, *Encephalitozoon* spp. in liver and the central nervous system and *Nosema* spp. in disseminated infection. (*See* **HIV Disease and Nutrition**.)

Treatment

Susceptibility of protozoa to chemotherapy correlates roughly with metabolism and, accordingly, with species. Several drugs are active against different protozoa. For some of them, e.g., diloxanide furoate, the exact mechanism of action is still uncertain. Others, e.g., metronidazole, are broad-spectrum antibiotics, interfering with the DNA of susceptible infectious agents.

Some antibiotics are active directly against protozoa, but others exert their effect on enteric bacteria necessary for proliferation of protozoa, e.g., *Entamoeba histolytica*. Most of these drugs are contraindicated in pregnancy and treatment of pregnant women often must be delayed until after delivery.

Metronidazole is effective against most flagellates and amoebae and is an alternative choice in the treatment of *Balantidium coli*. Other nitroimidazoles (tinidazole and ornidazole) have similar activity and less untoward side-effects, including headache, metallic taste, nausea, vomiting, and diarrhea.

Emetine and dehydroemetine have considerable side-effects, gastrointestinal and systemic, and are

Table 1 Foodborne and waterborne protozoa causing human infection

	Amoebae	*Flagellates*	*Ciliates*	*Coccidia*	*Microsporidia*
Intestinal infection	*Endolimax nana*[a] *Entamoeba coli*[a] *Entamoeba hartmanii*[a] *Entamoeba histolytica* *Iodamoeba bütschlii*[a] *Blastocystis hominis*	*Chilomastix mesnili*[a] *Dientamoeba fragilis* *Enteromonas hominis*[a] *Giardia intestinalis* *Retortamonas intestinalis*[a] *Trichomonas hominis*[a] *Trichomonas tenax*[a]	*Balantidium coli*	*Cryptosporidium* spp.[b] *Isospora belli*[b] *Sarcocystis* spp.	*Enterocytozoon* spp.[c]
Systemic infection	*Entamoeba histolytica*			*Cryptosporidium* spp.[b] *Toxoplasma gondii*[b] *Sarcocystis* spp.	*Encephalitozoon* spp.[c] *Nosema* spp.[c] *Pleistophora* spp.[c]

[a]Nonpathogenic for immunocompetent humans.
[b]Infection often serious in immunocompromised patients. Infection usually with *C. parvum*: avian species of doubtful significance.
[c]Importance as human pathogens, particularly in patients with acquired immune deficiency syndrome (AIDS), has only recently been recognized.

reserved only for extraintestinal amoebiasis. They are not active against *E. histolytica* in the bowel and diloxanide furoate, tetracycline, paramomycin, or iodoquinol must be added in order to treat simultaneous intestinal infection.

Iodoquinol is also effective against *Dientamoeba fragilis* (sensitive also to paramomycin and tetracycline) and is used as an alternative drug against *Balantidium coli*. Infections due to *Giardia intestinalis* may be treated with quinacrine, metronidazole, or furazolidine. Tetracycline is the drug of choice against *B. coli* and *D. fragilis*. Iodoquinol must be used with great caution, as it may cause myelitis and optic atrophy. This is the reason why it is no longer available in most developed countries.

Treatment of protozoal diseases in AIDS patients is far from satisfactory. There is no effective specific therapy for microsporidia or *Cryptosporidium* spp., perhaps the most common agent of infection of the bowel in AIDS. *Isospora belli* and *Sarcocystis* spp. respond poorly, if at all, to antifolates, which include co-trimoxazole, sulfadiazine, and pyrimethamine.

Prevention

Despite high prevalence and easy fecal–oral transmission of protozoa, foodborne and waterborne infections, although surely underestimated, seem rare. Good sanitary conditions and personal hygiene are keystones in preventing these infections. Protozoal cysts and oocysts differ in sensitivity to adverse conditions (heat, desiccation, freezing, chlorination) and some are amazingly resistant.

In countries with poor sanitary conditions, water should be boiled and uncooked vegetables and unpeeled fruits avoided, as they may have been washed with contaminated water. Flies and other arthropods may also be vehicles for fecal contamination of food. But even when a visitor takes every possible precaution to avoid these infections, a local food handler, even in the best hotel, may be a cyst passer not aware of the elementary rules of personal hygiene.

For some species there are several possible animal reservoirs, which lessens the effectiveness of preventive measures. Even for these species with humans as the only hosts, patients in some institutions, especially those who are mentally handicapped, are a difficult group to control. The occurrence of sexual transmission of some protozoa by oral–anal and oral–genital sex is also considered important and is difficult to control.

Helminthic Infections

Illness

Helminths are probably the most common infectious agents of humans. Two of those transmitted orally, *Ascaris lumbricoides* and *Trichuris trichiura*, are each thought to cause 1000 million infections worldwide. Humans may be, for different helminths, definitive or intermediate hosts and clinical manifestations, when present, depend on localization and worm burden. Unlike protozoa, helminths are not generally considered opportunistic pathogens in immunocompromised hosts.

Most intestinal infections caused by helminths listed in Table 2 are asymptomatic. Gastrointestinal symptoms are related to the number of worms in the bowel. Diarrhea and abdominal pain are the most common complaints, but malabsorption and weight loss may occur and, if untreated, infection may even be fatal, e.g., *Capillaria philippinensis*. *Enterobius vermicularis* is frequently responsible for anal pruritus, especially nocturnal, and restless sleep in children.

Some parasites compete for absorption of specific nutrients, with symptoms resulting from their deficiency, e.g., *Diphyllobothrium latum* causes

Table 2 Foodborne and waterborne helminths causing human infection

	Nematodes	*Trematodes*	*Cestodes*
Intestinal infection	*Anisakis* spp.	*Fasciolopsis buski*	*Diphyllobothrium* spp.
	Ascaris lumbricoides	*Heterophyes heterophyes*	*Dipylidium caninum*
	Capillaria philippinensis	*Metagonimus yokogawai*	*Hymenolepis diminuta*
	Enterobius vermicularis		*Hymenolepis nana*
	Trichuris trichiura		*Taenia saginata*
			Taenia solium
Systemic infection	*Angiostrongylus* spp.	*Clonorchis sinensis*	*Echinococcus granulosus*
	Bayliascaris procyonis	*Fasciola hepatica*	*Echinococcus multilocularis*
	Capillaria hepatica	*Opistorchis* spp.	*Hymenolepis nana*
	Dracunculus medinensis	*Paragonimus westermani*	*Multiceps multiceps*
	Toxocara spp.		*Taenia solium*
	Trichinella spiralis		*Spirometra* spp.

deficiency of vitamin B_{12} and consequently megaloblastic anemia and neurological symptoms. (*See* **Cobalamins**: Physiology.)

Systemic infections, although sometimes asymptomatic, are more commonly associated with symptoms which depend on localization of the parasite. Liver and biliary tract are targets for some helminths, e.g., *Capillaria hepatica, Clonorchis sinensis, Fasciola hepatica*, and *Opistorchis* spp., and symptoms of hepatitis and cholangitis may occur. Furthermore, increased incidence of cholangiocarcinoma seems to be associated with these infections.

Spirometra spp. and *Dracunculus medinensis* are present in subcutaneous tissue, causing pain and swelling or chronic ulcers with protrusion of worms. Lung is infected by *Paragonimus westermani*, with resulting bronchopulmonary symptomatology, including cough, hemoptysis, bronchitis, or lung abscesses.

Infections by *Angiostrongylus* spp., *Bayliascaris procyonis*, or *Multiceps multiceps*, when symptomatic, cause central nervous system disease, especially seizures. Visceral larva migrans syndrome (fever, hepatomegaly, and eosinophilia) is caused principally by *Toxocara* spp., but may also be due to *Angiostrongylus* spp., *Anisakis* spp., and *Capillaria* spp.

Larvae of *Trichinella spiralis* may be responsible for fever, edema, myositis and, rarely, encephalitis, pneumonia, or myocarditis. Larvae may encyst in different tissues, resulting in space-occupying lesions when they are big enough. Cysts of *Echinococcus granulosus* are found more frequently in liver and lung and are often multiple, but cysts of *E. multilocularis* have an unlimited germinal membrane and spread either locally or to distant sites like neoplastic metastases.

Treatment

Susceptibility of helminths to chemotherapeutic agents correlates with species even more closely than is the case with protozoa. Some drugs are preferred because they act in the bowel, expelling adult worms, while others must be absorbed and act systemically, in order to kill parasites in tissues. For most of them, the exact biochemical mechanism of action is still unknown. Some cause spastic paralysis of adult worms, e.g., pyrantel pamoate, while others cause flaccid paralysis, e.g., piperazine, with resulting expulsion of worms by peristalsis.

Mebendazole is the drug of choice for intestinal nematodes and replaced thiabendazole, which is much more toxic. In higher doses it is also used in the treatment of infections by *T. spiralis*, *E. granulosus* and *E. multilocularis*. Liver and bone marrow toxicity are not uncommon with high doses.

Other benzimidazole drugs (albendazole and flubendazole) are new alternatives which are perhaps less toxic. Pyrantel pamoate, pyrvinium pamoate and piperazine offer no pharmacological advantages over mebendazole. Mebendazole is also the drug of choice in systemic infections by nematodes, but is much less effective. Removal of worms is the best treatment for *Anisakis* spp. and *D. medinensis*.

Praziquantel is effective against most trematodes and cestodes. It is the drug of choice for either intestinal or systemic infections due to these parasites and failures have only been reported in *F. hepatica* (bithionol is the preferred drug for fascioliasis). Untoward side-effects, including nausea, abdominal pain, and headache, although not uncommon, are mild and transient.

Niclosamide is equally effective against intestinal cestodes, is less expensive, and has fewer untoward side-effects than praziquantel. Although imidazoles might be tried, surgery is still the best treatment for *M. multiceps* and *Spirometra* spp. and should be considered in echinococcosis, in which it is the only treatment when the cyst is calcified.

Although drugs now available to treat helminthic infections are less toxic than before, some authors still prefer, particularly in intestinal infections, treating only heavy infections, or using low doses of drugs to reduce the number of parasites in order to avoid undesirable side-effects. In contrast, reinfection with *Enterobius vermicularis* is so easy that some authors prefer to treat all household contacts. As with protozoal therapy, treatment of pregnant women is better delayed until delivery.

Prevention

Foodborne and waterborne infections are common with both helminths and protozoa, although direct person-to-person, fecal–oral contamination is also responsible for transmitting some infections. Good personal hygiene and sanitary conditions are thus mandatory for prevention and control of these infections. Flies and other arthropods may transmit eggs from feces to food and water. Vegetables may be contaminated either by encysted larvae (aquatic vegetables) or after contact with contaminated water.

Veterinary inspection of meat is very important in parasites which are infectious by macroscopic cysts or larvae. Raw meat or fish may transmit several helminths and should be avoided. Larvae and eggs may be extremely resistant to adverse conditions and remain viable for prolonged periods of time in salted, smoked, or undercooked meat and fish, as well as in chlorinated water. Snails, crayfish, crabs, and prawns may harbor larvae and so should be properly cooked. (*See* **Fish**: Spoilage of Seafood; **Shellfish**:

Contamination and Spoilage of Molluscs and Crustaceans.)

Specific Examples

Cryptosporidium spp.

Cryptosporidium spp. were recently recognized as an important cause of diarrhea in humans. Infection is frequently symptomatic and in immunocompetent hosts symptoms include mild acute self-limited watery diarrhea, sometimes with abdominal cramps, flatulence, nausea, anorexia, and weight loss. In immunocompromised patients, especially AIDS, chronic diarrhea and profound weight loss may occur and are an important cause of morbidity and mortality.

Spiramicin has been used for treatment with doubtful efficacy; no effective therapy is available. Nonspecific symptomatic therapy is important to avoid fluid and electrolyte imbalance, as in acute diarrhea of other etiology. Prevention may be difficult, as oocysts are extremely resistant and fully infectious when passed in stools. Good sanitary conditions and personal hygiene are of utmost importance. Precautions with stools in hospitals and other institutions should be taken in order to avoid outbreaks.

Giardia intestinalis

Giardiasis is now the most commonly reported protozoal infection worldwide. About 20–50% of infected people present symptoms, usually self-limited acute diarrhea, but chronic diarrhea with malabsorption may follow. Treatment with quinacrine has a high degree of efficacy, but the drug is poorly tolerated (vomiting, abdominal pain). Metronidazole and other nitroimidazoles are equally effective and better tolerated. Furazolidine has no advantages over quinacrine, either in efficacy or in untoward gastrointestinal side-effects. Good water control is the most important way of prevention. When necessary, water should be filtered or boiled. Chemical treatment is of doubtful efficacy.

Toxoplasma gondii

Toxoplasmosis in immunocompetent people is generally asymptomatic or may resemble an infectious mononucleosis. Tissue cysts, however, remain viable for life. In immunocompromised patients, such as AIDS patients, transplant recipients and patients submitted to immunosuppressive therapy, either acute or reactivation of latent infection, may be responsible for severe or fatal toxoplasmosis, mostly of the central nervous system. Congenital toxoplasmosis may present several clinical manifestations, some severe (encephalitis, retinitis), depending on the time of pregnancy when infection was acquired.

The combination of pyrimethamine plus sulfadiazine is the most effective therapy for toxoplasmosis and has been the choice in immunocompromised patients. To avoid relapses, it must be maintained for life in AIDS patients and as long as immunodeficiency persists in other cases. Bone marrow toxicity and skin rash are frequent with this combination and may cause replacement of sulfadiazine by clindamycin. Acute toxoplasmic infection in immunocompetent hosts is only to be treated when symptoms are severe or persistent.

Infection in pregnant women, either with or without symptoms, must be treated. Spiramycin can be used throughout pregnancy and after the first trimester, pyrimethamine plus sulfadiazine can be used. None of the available drugs is active on cysts, but only on vegetative forms. Oocysts are present in cat feces in large numbers and so contacts must be avoided by seronegative pregnant women and immunodeficient patients. In meat, cysts do not resist prolonged freezing (−20 °C), heating (66 °C), smoking, or curing.

Ascaris lumbricoides

Clinical manifestations of *A. lumbricoides* infection are diverse, being related to worm burden. Bronchopulmonary complaints may be present, due to migration of larvae through the lungs. If infection is slight, adult worms cause no symptoms, but heavy infections may be responsible for malabsorption or even intestinal obstruction. Other rare manifestations may be due to abnormal worm migration. Mebendazole is the treatment of choice. Pyrantel pamoate or piperazine are effective alternatives. Good sanitary conditions are keystones in prevention. Infection is also avoided by hand washing after contact with soil.

Taenia solium

Human intestinal infection by *Taenia solium* is generally asymptomatic. Symptoms due to cysticercosis depend on organ involvement, the central nervous system being the system most commonly affected (by seizures), followed by orbit, muscle, liver, and lungs. Niclosamide is widely used in the treatment of intestinal taeniasis. It is effective and has few untoward side-effects, but is not active against larvae. Praziquantel is highly active against adults and larvae, and should be preferred. Calcified cysts will not respond to drugs and may justify surgery. Symptomatic treatment of seizures may also be necessary. The most effective prevention is by good sanitary conditions, avoiding contamination of animal and human food and water by human feces. Veterinary

inspection of meat decreases the risk of ingestion of cysticercus, as does thorough cooking of pork.

Trichinella spiralis

Most intestinal and tissue infections by *Trichinella spiralis* are subclinical, symptoms depending mostly on the number of viable larvae and their localization. Diarrhea is the most common symptom due to adult worms in the gut, but abdominal discomfort and vomiting may also occur. Systemic invasion by larvae is more often symptomatic, with fever, myalgia, periorbital edema, and headache. Occasionally, myocarditis, pneumonitis, or encephalitis may occur. Mebendazole and thiabendazole are the drugs of choice for treatment, along with salicylates and bed rest. The most effective method of prevention is thorough cooking of pork.

See also: **Cobalamins**: Physiology; **Fish**: Spoilage of Seafood; **HIV Disease and Nutrition**; **Shellfish**: Contamination and Spoilage of Molluscs and Crustaceans

Further Reading

Casemore DP (1990) Foodborne protozoal infection. *Lancet* 336: 1427–1432.

Mandel GL, Douglas RG Jr and Bennett JE (eds) (1990) *Principles and Practice of Infectious Diseases*, 3rd edn. New York: Churchill Livingstone.

Mehlhorn H (ed.) (1988) *Parasitology in Focus*. Berlin: Springer-Verlag.

Strickland GT (2000) *Hunter's Textbook of Tropical Medicine*. Philadelphia: WB Saunders & Co.

Thompson JH Jr (1990) Parasitology. In: Murray PR, Drew WL, Kobayashi GS and Thompson JH Jr (eds) *Medical Microbiology*, pp. 349–436. London: Wolfe.

PARENTERAL NUTRITION

C C Ashley and L Howard, Albany Medical College, Albany, NY, USA

Definition and History

The development of central venous catheters in the 1960s provided for the safe infusion of hypertonic nutritive solutions. As placement of these catheters became more widespread, so did the use of parenteral nutrition (PN). In the early 1970s, this also became an available therapy for patients with chronic intestinal insufficiency living at home. Today, a variety of vascular access options are available to patients, and portable pumps can be used to infuse outside of the home/hospital setting. PN is sometimes called total parenteral nutrition (TPN), but this is an overstatement because several hard-to-solubilize but important nutrients are left out. Another term, 'hyperalimentation,' has also gone out of vogue, as we learn the downside of overproviding substrate, especially in a stressed patient.

Indications for Therapy

Wide availability does not necessarily imply broad applicability. Only recently has the evidence for and against the use of parenteral nutrition in a variety of clinical settings become better defined. Enteral routes of nutritional support are generally cheaper, safer, and equally efficacious. PN is indicated in those patients where delivery of nourishment via the gut is either unsafe or not possible. Candidates for PN should be carefully selected because the therapy is costly, cumbersome, and risky. A systematic approach to choosing candidates for PN is detailed in **Figure 1**.

The first step in the process requires the clinician to form a clinical judgement about whether the severity and nature of the patient's illness will lead to nutritional impairment. Initial considerations are whether a lengthy course of chemotherapy will provoke nausea and reduce oral intake to negligible levels; whether severe intestinal ileus will make it impossible to use the gut; and whether a planned surgery is likely to significantly reduce the patient's absorptive surface. The second step is to assess whether the patient is at increased risk for malnutrition because of limited nutritional reserves. The physician must determine whether the patient is an alcoholic. Are they already on catabolic drugs such as steroids? Are they vulnerable by virtue of age or chronic illness such as diabetes? The third step is to determine whether the patient is already malnourished. Loss of more than 20% of lean body mass, serum albumin depression signs of vitamin or mineral deficiency independently indicate a high-risk nutritional status. The fourth step is the most critical. The physician must decide whether the quality of life and prognosis can be positively effected by the provision of specialized nutritional support and therefore whether this therapy should be recommended to the patient. The

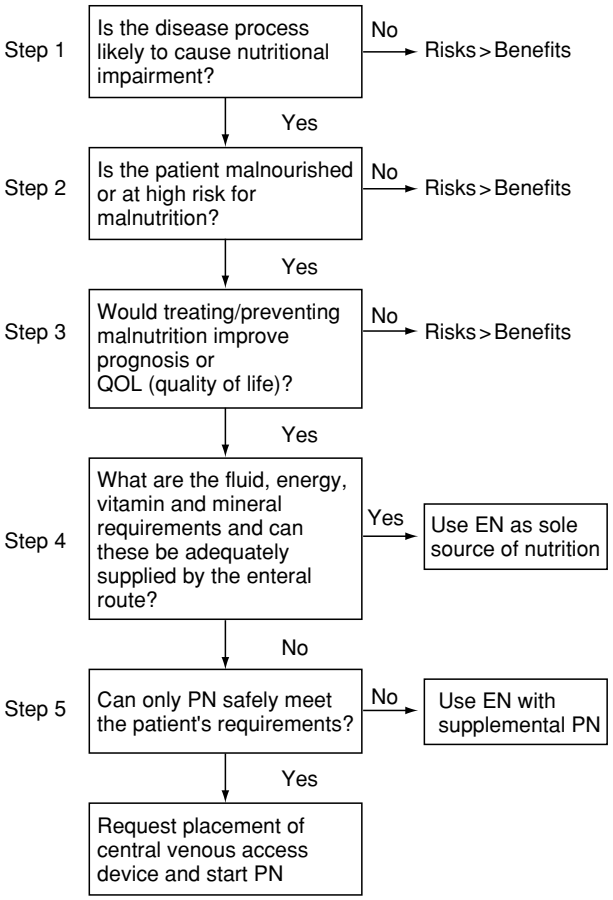

Figure 1 Systematic approach to assessing candidacy for total parenteral nutrition. Adapted from Howard L (1998) Parenteral and enteral nutrition. In: Fauci AS, Braunwald E, Isselbacher KJ and Wilson JD (eds) *Harrison's Principles of Internal Medicine*, 14th edn., chapter 78, figure 78-1, p. 473. New York: McGraw-Hill, with permission.

remaining steps are concerned with choosing the route of delivery for the indicated specialized nutritional support. A summary of common clinical scenarios, where provision of PN is considered, is shown in Table 1.

Evidence Base for PN

Insufficient evidence exists to assess the utility of PN in many specific clinical settings, but there are several broad areas where prospective randomized controlled studies support the use of PN.

Perioperative

Strong evidence exists to support the use of PN in the perioperative setting when the patient has severe protein–calorie malnutrition. The benefit is a modest (several studies found 10%) reduction in postoperative complications.

Table 1 Various indications for parenteral nutrition

Obstruction
Anatomical (congenital malformation, strictures, adhesions)
Functional (prolonged postoperative ileus, intestinal dysmotility)
Hypermetabolic state/wasting syndromes
Critical illness
Multisystem organ failure and bowel ischemia
Severe pancreatitis
Burns
Cancer
Bone-marrow transplantation
HIV
Malabsorption/mucosal disease
Insufficient surface area from short bowel syndrome
Proximal enteric fistula
Severe inflammatory bowel disease
Severe chemotherapy related mucositis
Radiation enteritis

Critical Illness

In critical illness, there is support for the use of PN in patients with severe protein–calorie malnutrition using modest calories. Complication rates are lower when lipids are not infused as a part of this specialized nutritional support. Excessive exogenous carbohydrate enhances hypermetabolism and lipogenesis, leading to increased carbon dioxide production; this makes ventilatory weaning more difficult.

Wasting Syndromes

Cancer Like noncancer patients, cancer patients suffering from severe protein–calorie malnutrition benefit from perioperative PN when undergoing cancer surgery. There are some data to suggest that bone-marrow-transplant patients benefit from PN supplemented with glutamine and that PN during the cytoreductive phase enhances long-term survival.

HIV HIV patients appear to benefit from specialized nutrition support if their wasting reflects poor intake or malabsorption; they do not benefit if their weight loss is principally cytokine-related wasting. The route of nutrition support is more controversial. In one well-controlled study, PN was not superior to enteral nutrition in repleting lean body mass, but it did add more weight as water and fat. One small study showed a longer survival in PN HIV patients compared to those who received diet counseling alone.

Chronic Renal Failure

PN improves survival in acute renal failure. A solution using just essential amino acids is not superior to the usual essential–nonessential combination. In

chronic renal failure requiring dialysis, 15% of patients become significantly malnourished. PN may be indicated if there is severe bowel dysmotility associated with autonomic neuropathy, as may occur with diabetes. Routine use of intradialytic PN has not been shown by randomized studies to clearly alter survival.

Pancreatitis

In pancreatitis, PN can decrease mortality rate and reduce rates of infectious complications when it is used early in the management of the most severe cases.

Short Bowel Syndrome

The prognosis of acute short bowel syndrome has been dramatically improved by PN, so much so that a randomized study of PN vs. enternal nutrition (EN) would be unethical. In the adaptive phase, an open study purported to show that supplementation with growth hormone and glutamine improved bowel adaption, but this was not confirmed in a more rigorous controlled trial.

Inflammatory Bowel Disease

In inflammatory bowel disease, Crohn's disease, or ulcerative colitis, specialized nutritional support (SNS) is often necessary for severe disease associated with protein calorie malnutrition, but there are no clear data that show any advantages of PN over EN. SNS alone is not superior to steroids in reducing remission or preventing relapse; however, it may be the preferred approach in a growing child.

Technical Details

Route

Peripheral PN PN can be delivered via peripheral intravenous access, but this route is appropriate only for formulations of modest osmolarity (e.g., less than 900 mOsm or 10 g of dextrose per 100 ml), since small veins quickly sclerose and become useless for further infusion. Peripheral PN is best used in a supportive role for patients with inadequate oral intake. Patients who are unable to meet their fluid and electrolyte needs can be supported in this way without exposure to the complications of central venous access. However, peripheral PN is not a long-term solution.

Central PN Central access is a prerequisite when infusing formulations of PN that are more complex and more hyperosmolar. The catheter used to deliver PN should be a dedicated line used for nothing else in order to minimize the risk of infection. If additional intravenous therapy is needed, additional access sites should be arranged. Sometimes, multilumen catheters are used, but studies have shown higher infection rates in multilumen catheters versus single lumen catheters.

Conventional central venous line In the setting of an acute requirement for parenteral therapy or as a bridge to the placement of a longer-lasting device, an externalized central venous line (CVL) inserted by the conventional Seldinger technique in the subclavian or internal jugular vein is acceptable for the delivery of PN. This catheter may be placed at the bedside as long as full aseptic precautions are taken. This can be a considerable advantage over other devices that may take several days to arrange.

Peripherally inserted central venous catheter The peripherally inserted central venous catheter (PICC) can be placed at the bedside in the antecubital fossa by specially trained personnel. PICCs placed at the antecubital fossa are subject to mechanical stress from flexion and extension of the elbow. They are a short-term option. For long-term use, especially in patients going home on PN, PICC lines are best placed to the upper arm (brachial vein) under ultrasound and fluoroscopic guidance. Such lines are not subject to the same level of mechanical stress. One disadvantage common to all PICCs is that they are more difficult for home patients to care for without nursing support, because the patient has only one free hand.

Tunneled catheter The chief advantage of tunneled catheters is that an exit site low on the chest wall permits a patient at home to see their catheter site and manipulate the catheter with both hands. This considerably simplifies the act of catheter access and improves patient independence. Tunneled catheters have a dacron cuff just under the skin which aids catheter fixation and prevents bacteria from accessing the blood stream through the tunnel. Patients with an externalized catheter are usually advised to cover their access devices with an adhesive/occlusive dressing before showering or bathing in order to keep it dry. Swimming is not recommended.

Subcutaneous port Subcutaneous ports are often preferred by long-term patients, who have active lifestyles and want to swim. Subcutaneous ports are less obtrusive and therefore cause less disturbance to body image. The disadvantage of these catheters is that they require a needle stick for accessing, but this uncomfortable procedure is made easier by the use of topical anesthetics. Also, it is rarely possible to

clear an infected part and hence removal is necessary. For this reason they are not wise in patients with a history of frequent line sepsis.

Infusion Method

Continuous Once a route is decided upon and available, a method of delivery must be selected. Most patients begin PN as a continuous infusion. Continuous infusion is the least complicated method of delivering PN, since a metabolic steady state is quickly achieved and maintained. This is the method of choice in patients who are critically ill, hemodynamically compromised, have marginal renal function, or are brittle diabetics. However, continuous infusions are not optimal from the perspective of most patients. Long-term patients frequently seek some freedom from their parenteral infusion. If a long-term continuous infusion is necessary, portable pumps are available that allow patients to perform most activities of daily living without needing to disconnect.

Cycled TPN infusions can usually be compressed to a period between 10 and 12 h, depending upon the volume and calories needing to be infused. Most home patients opt to infuse during night time hours. This leaves the patient free for a period of between 12 and 14 h a day to be more active. When the infusion is cycled on and off daily, the risk for dehydration rises especially toward the end of the infusion-free period and in patients with significant gastrointestinal (GI) losses. The risk of both hyper- and hypoglycemia is increased at the start and conclusion of the daily infusion. Swings in serum glucose levels demonstrate the impact of carbohydrate delivered directly to the systemic rather than the portal circulation.

Intermittent In patients requiring less PN support, it may be possible to provide PN every second or third day. The complication of such a regimen is making sure that mineral and vitamin requirements are met.

Establishing a Nutrient Formula

Once the need for parenteral support is established and the route and method of delivery are decided on, the formulation can be determined. The first step is to estimate the total nutritional requirement of the individual patient. The requirement of an individual patient may vary from day to day, depending upon abnormal Gl losses. Over time, bowel adaption occurs, so fluid and nutrient requirements may significantly decrease.

Fluid requirements

Basal The normal daily fluid requirement is 25–35 ml kg^{-1} (for adults) and 120 ml kg^{-1} (for infants). In a 70-kg man, normal losses are 1600 ml of urine daily, 800 ml of insensible losses, 100 ml of stool losses and 350 ml per degree Celsius per day of fever.

Extraordinary fluid losses Fluid losses can vary considerably in patients who require TPN, often because of excessive loss from the GI tract. In the example shown in Table 2, the patient requires twice the usual volume of fluid because of extraordinary losses through his jejunostomy. All fluid losses must be meticulously reviewed in order to ensure adequate replacement. Commonly excessive losses from the GI tract will precipitate a greater oral intake.

Calorie requirement The total amount of calories required is determined by weight and the patient's level of stress. Between 20 and 40 kcal kg^{-1} day^{-1} is usually provided to adults and 80–110 kcal kg^{-1} day^{-1} to infants. In patients who are simply starved but not stressed, 35–40 kcal kg^{-1} day^{-1} (100–110 kcal kg^{-1} day^{-1} for infants) can be effectively utilized. In patients who are moderately stressed and malnourished, slightly lower amounts are provided, e.g., 30–35 kcal kg^{-1} (80–100 kcal kg^{-1} day^{-1} in infants). In patients who are severely stressed, calories are kept modest (< 25 kcal kg^{-1}), since the stress induces a catabolic state that prevents efficient utilization of high levels of calories. Excessive calories precipitate hepatic steatosis and retention of carbon dioxide.

Carbohydrate content Glucose provides 3.4 kcal g^{-1} because it is glucose monohydrate. The carbohydrate content of the formulation varies from 4 to 7 g kg^{-1} day^{-1} and usually accounts for 50–60% of total calories.

Table 2 Daily fluid requirements

Normal man (70 kg)			
35 ml day^{-1} × 70 kg = 2.5 l			
In:	1.2 l of fluid	Out:	1.6 l of urine
	1.0 l of water in food		0.8 l insensible losses
	0.3 l of water from metabolism of food		0.1 l stool
Same patient status post high jejunostomy			
In:	2.2 l of fluid	Out:	1.6 l of urine
	1.0 l of water in food		0.8 l of insensible losses
	0.3 l of water from metabolism of food		4.1 l of jejunostomy losses
	3.0 l TPN		

The patient needs to have enough parenteral fluid to maintain an adequate urine output. Short-bowel patients have an increased risk of kidney stones. Adapted from Howard L (1998) Parenteral and enteral nutrition. In: Fauci AS, Braunwald E, Isselbacher KJ and Wilson JD (eds) *Harrison's Principles of Internal Medicine*, 14th edn., chapter 78, figure 78-4, p. 476. New York: McCraw-Hill.

Fat content Fat solutions provide 10 kcal g^{-1} rather than the expected 9 kcal g^{-1} because the lipid is made iso-osmolar with glycerol. The fat content of the formulation varies from 0.5 to 1.0 g kg^{-1} day^{-1} and usually accounts for 30% of total calories.

Protein content Amino acids provide 4 kcal g^{-1}. The protein content of the formulation varies from 0.8 g kg^{-1} for a nonstressed patient to 1.5 g kg^{-1} for a severely catabolic patient. Amino acids generally account for 10–15% of total calories. A patient whose protein requirements are adequately met will be in nitrogen balance based upon the following formula:

$$\text{Nitrogen balance} = \text{nitrogen intake} - \text{nitrogen loss}$$

(nitrogen intake is 14% of all protein or amino acid intake in g per 24 h; nitrogen loss is urine urea nitrogen in g per 24 h + 4 (a constant for other losses)).

In the long term, protein balance can be assessed by wound healing, resumption of growth, and/or restoration of body composition.

Electrolytes and minerals PN solutions are based on a standard physiologic preparation of electrolytes. The content of specific electrolytes can be varied to replace patient losses or avoid excessive levels in patients with impaired renal or hepatic excretion. Extraordinary losses of alkali and chloride salts are more the rule than the exception in patients with large volumes of GI fluid loss. Serum chemistries must be intensively monitored until a state of equilibrium is established, after which, periodic assessment remains necessary. The necessary electrolyte replacement can be based on the composition of the abnormal fluid losses (see Table 3).

Vitamins and trace elements Parenteral provision of vitamins and trace elements is complicated by systemic infusion and altered enterohepatic cycles. In addition, some micronutrients may adhere to formula bags and delivery tubing or be destroyed by exposure to light (vitamin A). Some nutrients (divalent cations and fat-soluble vitamins) are especially at risk for deficiency in patients with bowel disease and steatorrhea. Others, cleared by the liver, may accumulate in toxic amounts in patients with liver dysfunction and in that setting should be supplied only in reduced amounts (Cu, Mn). Still others are at risk for toxic levels of accumulation in the setting of chronic renal failure (Se, Mb, I). The daily parenteral requirements for essential fatty acids, minerals and vitamins are summarized in Table 4.

Medications and other additives Daily central venous infusions create an opportunity for delivery of a variety of medications by the intravenous route. Any admixing must take into account compatibility with the PN formulation and any cycling of PN delivery that may create an unsafe or inefficacious situation for an added medication (see Table 5).

Complications

Catheter Failure

Infection Catheter-related infection heads the list of serious complications, since it is both common and potentially devastating. More serious than any individual infection is a pattern of infections that increases the risk of endovascular colonization. The incidence of infection is higher in children. Mortality from these infections is low, but the dangers are real,

Table 3 Enteric fluid volumes and their sodium, potassium, chloride and bicarbonate contents[a]

	l day^{-1}[b]	*Na (meq l*$^{-1}$*)*	*K*[c] *(meq l*$^{-1}$*)*	*Cl (meq l*$^{-1}$*)*	*HCO*$_3$[d] *(meq l*$^{-1}$*)*
Oral intake	2–3				
Enteric secretions					
Saliva	1–2	10	30	10	30
Gastric	2	60	9	90	0
Bile	2–3	150	10	90	70
Small bowel	1	100	5	100	20
Colon	Variable	40	100	15	60

[a]Enteric secretions are also rich in divalent cations (Ca, Mg, Zn, Cu), and their loss is increased by steatorrhea, a high bowel fistula, or prolonged gastric suction.

[b]Of the total oral and enteric fluid presented to the upper small bowel, normally 50% is absorbed in the jejunum, 40% in the ileum, and 10% in the colon. In short-bowel patients, the colon may absorb greater amounts, up to 3 l day^{-1}.

[c]Potassium losses are small except in secretions distal to the ileocecal valve. The colon ion exchange is partly controlled by aldosterone, and therefore, Na depletion increases K loss in the stools.

[d]Bicarbonate losses must be replaced in parenteral solutions as acetate or lactate because of potential precipitation of bicarbonate with ingredients such as calcium.

From Howard L (1998) Parenteral and enteral nutrition. In: Fauci AS, Braunwald E, Isselbacher KJ and Wilson JD (eds) *Harrison's Principles of Internal Medicine*, 14th edn., chapter 78, figure 78-5, p. 477. New York: McGraw-Hill.

Table 4 Daily parenteral requirements of essential fatty acids, minerals, and vitamins

Nutrient	*Daily parenteral requirement for an adult*
Essential fatty acids, % kcal	2–4
Calcium, g (meq)	0.2–0.4 (10–20)
Phosphorus, g	0.4–0.8
Potassium, g (meq)	3–4 (75–100)
Sodium, g (meq)	1–3 (45–130)
Chloride, g (meq)	3–4 (85–120)[a]
Magnesium, g (meq)	0.2–0.3 (15–25)
Iron, mg	1–2
Zinc, mg	3–12
Copper, mg	0.3–0.5
Iodine, mg	0.15
Manganese, mg	2–5
Chromium, μg	15–30
Molybdenum, μg	20–120
Selenium, μg	50–100
Ascorbic acid, mg	100
Thiamin, mg	3
Riboflavin, mg	3.6
Niacin, mg	40
Biotin, μg	60
Pantothenic acid, mg	15
Pyridoxine, mg	4
Folic acid, μg	400
Cobalamin, μg	5
Vitamin A, IU	3300
Vitamin D, IU	200
Vitamin E, mg	10
Vitamin K, μg	200

[a]In addition to the requirement for chloride anions, there is a parenteral requirement for bicarbonate equivalents to protect the acid–base balance. These are provided as 90 mmol or more of acetate or lactate per day because of potential precipitation of bicarbonate with ingredients such as calcium.

From Howard L (1998) Parenteral and enteral nutrition. In: Fauci AS, Braunwald E, Isselbacher KJ and Wilson JD (eds) *Harrison's Principles of Internal Medicine*, 14th edn., chapter 78, figure 78-6, p. 477. New York: McGraw-Hill.

Table 5 Compatibilities of selected drugs in parenteral nutrition solutions

Drugs compatible with PN solutions	
Dextrose/amino acids	*Total nutrient admixtures*[a]
Albumin	Albumin
Cimetidine	Cimetidine
Famotidine	Famotidine
Folic acid	Heparin
Heparin	Regular human insulin
Hydrochloric acid	Metoclopramide
Regular human insulin	Phytonadione
Iron dextran	Ranitidine
Metoclopramide	
Phytonadione	
Ranitidine	

Drugs incompatible with PN solutions	
Dextrose/amino acids	*Total nutrient admixtures*
Amphotericin B	Amphotericin B
Ampicillin	Hydrochloric acid
Metronidazole (with $NaHCO_3$)	Iron dextran
Phenytoin	Methyldopa
	Phenytoin

Drugs incompatible with PN solutions with Y-site administration
Acyclovir
Amphotericin B
Cefazolin
Ciprofloxacin
Cisplatin
Cyclosporine
Cytarabine
Doxorubicin
Fluorouracil
Furosemide
Ganciclovir
Methotrexate
Metoclopramide
Midazolam
Minocycline
Mitoxantrone
Potassium phosphate
Promethazine
Bicarbonate salts

[a]Total nutrient admixture or '3 in 1' refers to a solution containing dextrose, amino acids, and fat.

From ME Shils and RO Brown (1999) Parenteral nutrition. In: Shils ME, Olson JA, Shike M and Ross AC (eds) *Modern Nutrition in Health and Disease*, 9th edn, chapter 101, table 101.13, p. 1688. Philadelphia, PA: Lippincott, Williams & Wilkins.

since patients can, through multiple infections, use up their sites for central vascular access. Infection can be limited to the device, spread to the bloodstream, and/or involve local tissue either at the exit site or along the tunnel. Definitive treatment of catheter-related infection usually involves removal of the device. A variety of methods have been tried to reduce the risk of infection in patients dependent upon PN. All patients are instructed to use scrupulous sterile techniques when accessing their device. The device should be accessed only to deliver specialized nutritional support and should not be used for the infusion of other products or blood draws. Some access devices are impregnated with antibiotics or bacteriostatic agents. An antibiotic solution lock after the use of a catheter can reduce catheter sepsis and is used in patients with frequent infection. To remove the biofilm on the internal surface of the catheter which allows invading bacteria to grow and limits antibiotic sterilization – intermittant thrombolytics can be used likewise endoluminal brushes. Endoluminal brushes have been used both to identify infected lines and to clear catheters.

Bleeding Bleeding occurs either locally because of bleeding from a leaking catheter or from a systemic

coagulopathy. Most catheter leaks can be addressed by clamping or crimping the catheter proximal to the leak and then having the catheter repaired. Subcutaneous bleeding around a port usually responds to pressure.

Air embolism Air embolism is a feared complication of central venous access. Patients need to be reassured that dangerous volumes of air are large and rarely happen through an access device. If concern for this complication is real, and the amount of air delivered is unknown, patients crimp off their catheter and lie flat on their left side so that air is trapped in the right auricle or ventricle and does not pass on into the pulmonary artery and the lungs where the 'bubble' will disrupt gas exchange. Over 20 min, even large trapped bubbles will dissolve.

Phlebitis Phlebitis is a relatively common complication with peripherally inserted access devices. Inflammation of the small vessel, which is initially entered, can result in local pain, swelling, erythema, and heat. There may or may not be associated cellulitis. The internal surface of the catheter usually remains sterile. If conservative local therapies are insufficient to provide relief, the line should be removed, since phlebitis can progress to thrombosis and more serious complications.

Migration After initial placement, catheters can migrate in the wrong directions. For example, a line placed via the subclavian vein can end up in the internal jugular. The peripherally inserted central catheters can propagate into the heart from the superior vena cava or right atrium. The resulting endocardial irritation can result in palpitations, skipped beats, and, rarely, more serious arrhythmia. When fluoroscopy is available at the time of the placement, these complications are easily correctable. It is useful to note at the time of insertion just how much of a CVL or PICC remains externalized.

Crimping and breakage If infusion becomes difficult, and the pump signals blockage, the first consideration should be the site of central access. Crimping of the catheter is common with peripherally inserted access introduced through the veins of the antecubital fossa. Repeated flexion and extension at the elbow can introduce a kink in the catheter itself and result in breakage. The diagnosis of catheter breakage is straightforward, since it usually occurs as separation of the catheter from the external hub or as a simple breach in the external catheter wall. Rarely is a catheter sheared off internally. None the less, a central catheter should never be withdrawn with any amount of force, as there have been incidences of catheter embolization requiring extraction under fluoroscopy.

Blockage Catheters can become blocked by thrombus, built-up fibrin and platelet deposits, crystals of infused minerals or drugs, or by lipid precipitate from intravenous fat solutions. Vigorous and frequent flushing with heparinized saline and the use of heparin-impregnated catheters has been useful in reducing the frequency of thrombotic luminal obstruction. Accumulation of fibrinous material on the outer surface of the catheter tip is relatively common and can produce a ball valve effect, allowing infusion but preventing blood withdrawal. Fluoroscopic confirmation of catheter position and clots is useful. Current treatment for thrombus or fibrin is 0.5–1 mg of tissue plasminogen activator. Crystalline material may be dissolved by 0.1 N HCL and lipid by small amounts of alcohol. Checking for blood return from access devices is discouraged because the process of drawing blood into the catheter promotes the process of fibrin and platelet precipitation and contributes to higher rates of intraluminal catheter obstruction and infection, shortening the usable life of the devices. Catheter thrombosis is frequently associated with catheter infection.

Underfeeding

Ideally, the amount of specialized nutritional support delivered to the patient is sufficient to meet the needs of the patient. Because the process of determining those needs is imperfect, it is likely that either too much or too little support will be delivered. Because the complications associated with overfeeding can be dramatic, it is now preferred to err on the side of underfeeding, especially in the patient who is under stress. In order to optimize the patient's response to specialized nutritional support, prospectively determined goals should be used to gauge the success of therapy. The clinician should assess the patient's wound healing, strength to wean from a respirator, cough effectively, and mobility. If such an evaluation is neglected, the patient may inadvertently be denied the full benefit of parenteral support.

Overfeeding

Overfeeding is the provision of one or more of the nutrient components of parenteral nutrition in an amount or at a rate that exceeds the physiologic utilization and excretion of the nutrient.

Hyperglycemia Excessive provision of carbohydrate can occur in both diabetic and nondiabetic patients and is more likely to occur in those who are

critically ill. Hyperglycemia should be treated by reducing the carbohydrate calories. The use of insulin is only appropriate in patients known to be insulinopenic.

Fluid overload Fluid overload is possible in patients who have reduced cardiac reserves and in persons with healthy hearts whose PN volume is added to an already generous regimen of crystalloid or colloid. It is essential to consider the total volume of other fluids delivered when determining a volume for PN.

Fatty liver Delivery of carbohydrate or fat calories in excess of that which can be used by the patient creates hepatic steatosis. In the case of a patient with impaired liver function, this can occur more rapidly. Chronic hepatic steatosis leads to cholestasis and, occasionally, irreversible fibrosis and cirrhosis. This appears to occur more readily in children than adults.

Refeeding syndrome This classic complication of starvation can occur after prolonged starvation from any cause; it happens when the nutrient supply is restored too rapidly. The underfed patient has depleted stores of adenosine triphosphate (ATP) with normal or low circulating levels of potassium, phosphorus, and magnesium. Once metabolism is restored, phosphate is drawn into cells to be coupled to adenosine monophosphate (AMP) and adenosine diphosphate (ADP). Magnesium is also drawn into cells to act as a cofactor for a number of cellular processes, for example, the conjugation of phosphate to ADP and the Na–K ATPase. Finally, potassium is drawn into cells to participate in the vigorous metabolism of newly supplied energetic compounds. This leads to depleted serum magnesium, potassium, and phosphate. This transcellular shift can have dire consequences, as a result of altered membrane conduction, resulting in arrhythmia, edema, and heart failure.

Allergic Reactions

Reactions to any component used in the delivery of PN are possible. Allergic reactions to latex are well known. Autoimmune reactions to heparin are also recognized, and even the small amount present in the lumen of an impregnated catheter is sufficient to trigger the heparin-induced thrombotic thrombocytopenia syndrome. Reactions to the actual components of the nutrient formulation are unusual.

Withdrawal of Support

Once PN is begun in a patient, when does one stop? There are three situations in which withdrawal of therapy is appropriate: first, when the patient can wean from PN to more physiologic EN; second, when desired by the patient; and third, when further treatment is futile, and the risks clearly outweigh the benefits. Assessing PN nutrition repletion and transposing to use of the GI tract is an important clinical step that often has to happen gradually. Every patient must agree to both initiation and stopping PN therapy. Occasionally, after experiencing nutritional rescue by PN, the patient or family desire continuation of therapy and are fearful of the consequences of weaning off. When continued nutritional support is felt to contribute nothing to the patient's quality of life or prognosis, it should be withdrawn like any other futile therapy. This process is ethically straightforward but emotionally charged and medicolegally complicated. The best way to avoid these difficult situations is to avoid PN in every case where its use is not clearly indicated, and the course of therapy is not prospectively defined.

See also: **Bioavailability of Nutrients**; **Body Composition**; **Burns Patients – Nutritional Management**; **Cancer**: Diet in Cancer Treatment; **Children**: Nutritional Requirements; **Dehydration**; **Diarrheal (Diarrhoeal) Diseases**; **Dietary Reference Values**; **Dietary Requirements of Adults**; **Electrolytes**: Analysis; **Enteral Nutrition**; **HIV Disease and Nutrition**; **Infants**: Nutritional Requirements; **Malnutrition**: The Problem of Malnutrition; **Minerals – Dietary Importance**; **Nutritional Assessment**: Importance of Measuring Nutritional Status; **Stress and Nutrition**; **Vitamins**: Determination; **Water**: Structures, Properties, and Determination

Further Reading

Ashley C and Howard L (2001) Evidence base for specialized nutritional support. *Nutrition Reviews* 58: 282–289.

Brooks S and Kearns P (1996) Enteral and parenteral nutrition. In: Ziegler EE and Filer LJ (eds) *Present Knowledge in Nutrition*, 7th edn. pp. 530–539. Washington, DC: ILSI Press.

Dresser RS and Boisaubin EV (1985) Ethics law and nutritional support. *Archives of Internal Medicine* 145: 122–124.

Gonzalez-Huix F, Fernandez-Banares F, Esteve-Comas M *et al.* (1993) Enteral versus parenteral nutrition as adjunct therapy in acute ulcerative colitis. *American Journal of Gastroenterology* 88: 227–232.

Howard L (1998) Parenteral and enteral nutrition therapy. In: Fauci AS, Braunwald E, Isselbacher KJ and Wilson JD (eds) *Harrison's Principles of Internal Medicine*, 14th edn. pp. 472–480. New York: McGraw-Hill.

Howard L, Ament M, Fleming CR *et al.* (1995) Current use and clinical outcome of home parenteral and enteral

nutrition therapies in the United States. *Gastroenterology* 109: 355–365.
Klein S, Kinney J, Jeejeebhoy *et al*. (1997) Nutrition support in clinical practice: Review of published data and recommendations for future research directions. *Journal of Parenteral and Enteral Nutrition* 21: 133–156.
Klein S and Koretz RL (1994) Nutrition support in patients with cancer: What do the data really show? *Nutr Clin Pract* 9: 91–100.
Lipman TO (1998) Grains or veins: Is enteral nutrition really better than parenteral nutrition? A look at the evidence. *JPEN* 22: 167–182.
Semrad CE (2000) Parenteral nutrition. *Clinical Perspectives in Gastroenterology* Nov/Dec: 307–314.
Shils ME and Brown RO (1999) Parenteral Nutrition. In: Shils ME, Olson JA, Shike M and Ross AC (eds) *Modern Nutrition in Health and Disease*, 9th edn. pp. 1657–1688. Philadelphia, PA: Lippincott, Williams & Wilkins.
Viall CD (2000) Home parenteral nutrition: Finances. In: Rombeau JL and Rolandelli RH (eds) *Clinical Nutrition: Parenteral Nutrition*, 3rd edn. pp. 512–528. New York: WB Saunders.

PASSION FRUITS

D B Rodriguez-Amaya, Universidade Estadual de Campinas, Campinas, SP, Brazil

Introduction

Because of the intense, fragrant, and distinctive aroma and flavor, passion fruit has been the object of investigations and is in increasing demand worldwide. Native to Brazil, this fruit owes its name to the delicate and beautiful flowers, some features of which the early Christian missionaries to South America found symbolic of the Passion of Christ.

Species and Cultivars

Passion fruit belongs to the *Passifloraceae* family, which consists of 12 genera and over 500 species, widely distributed in tropical America, Asia, and Africa. *Passiflora*, the principal genus, has approximately 400 known species, about 50–60 of which bear edible fruits, but only a few are of any commercial importance. Many are known only in native markets in South and Central America and the West Indies. Commercial production of passion fruit is based on the purple species *Passiflora edulis* Sims and the yellow form *Passiflora edulis* f. *flavicarpa* Degener. It is still in question whether the yellow passion fruit is a mutant of the purple *P. edulis* or its hybrid with other species. These two forms are commonly called granadilla, parcha, or parchita in Spanish, maracuja in Portuguese, and lilikoi in Hawaiian.

Aside from the skin color, the purple and yellow passion fruits differ in horticultural performance and fruit properties. The purple species is more resistant to cold injury, is less acid, and is considered superior in aroma and flavor. The yellow form is faster growing, has a greater resistance to soil fungi, has more vigorous vines, bears crops over longer periods, and has a greater yield of fruit and pulp, larger fruits, and more acid juice.

P. quadrangularis L. or 'giant granadilla,' widely distributed in the tropics, is the most cultivated species after *P. edulis* in tropical America. It bears the largest fruits, which are more elongated, up to 25 cm long, and fleshy instead of hollow. The seeds are much larger, brownish, and flattened. The rind is not so hard as in the purple and yellow passion fruits. The juice content is much lower and is somewhat inferior in flavor and color. The pulp can be eaten like a melon, with or without the addition of sugar, or cooked with milk. When green, this fruit can be used as a vegetable as green papaya.

P. ligularis Juss, 'sweet granadilla' or 'water lemon,' is cultivated in the mountains of Mexico and Central America. The fruit is mostly eaten out of hand or used in drinks or icecream. Its translucent white pulp is almost a liquid, acid with sweet aroma. The peel is resistant so the fruit can be transported well, without being damaged. Colombia, where this species is cultivated in the Western Cordillera, exports this fruit to Europe.

P. mollissima, 'banana passion fruit' or 'curuba,' grows widely in the Andes and is distributed from Venezuela and Colombia to Peru and Bolivia. The flavor is more astringent and less acid than *P. edulis*. The sieved pulp is mixed with milk and sugar and served as a drink. It is also used in marmalades and desserts and for flavoring icecream

P. maliformis L. is known as 'sweet cup' in the West Indies, 'chulupa' in Colombia, and 'granadilla de hueso' in Ecuador. It is a little-known species but

may have a good future because of its excellent aroma and flavor.

Botany and Horticulture

The purple passion fruit crops best at higher altitudes in the tropics. It is produced commercially in Australia, South Africa, Kenya, Papua-New Guinea, New Zealand, Sri Lanka, and India. The yellow form thrives in humid tropical lowlands. It is the passion fruit most cultivated in Brazil, Peru, Venezuela, Hawaii, and Fiji. Both wild and cultivated passion fruits are also found in Bolivia, Malaysia, Indonesian, Taiwan, and the Philippines. Production of passion fruit is also possible in frost-free areas in the temperate zone, although fruiting is more limited than in the tropical regions.

The plant is a woody, perennial vine that bears a large number of purple or yellow, oval or round fruits 4–7.5 cm in diameter. It climbs with the aid of tendrils that spring from the same axils as the flowers. The alternate evergreen leaves, deeply three-lobed when mature, are finely toothed, 7.5–20 cm long, deep green and glossy above, paler and dull beneath, and, like the young stems and the tendrils, tinged with red or purple, especially in the yellow form. A single, fragrant flower, 5–7.5 cm in diameter, is borne at each node on the new growth. Clasped by three green leaf-like bracts, it consists of five greenish white sepals; five white petals; a fringed corona of straight white-tipped rays, vivid purple at the base; five stamens with large anthers; the ovary; a prominent, centrally located, triple-branched style and three stigmas. The stigmas are at the apex of the androginophore, the anthers situated below them, making self-pollination difficult.

The *P. edulis* Sims bears fruits, weighing about 35 g, with a deep purple leathery skin that wrinkles when the fruits are fully ripe. The *P. edulis* f. *flavicarpa* Degener has a yellow skin softer than that of the purple fruit and is usually larger, weighing about 80 g. It does not shrivel as badly. The rind is about 3 mm thick and has a whitish layer internally, similar to the albedo of citrus fruit. Inside the fruit is a cavity filled with a juicy mass of intensely fragrant, translucent yellow–orange pulp, with a distinctive sour-sweet flavor, in which are embedded numerous small, flattened, oval, edible, black (in purple passion fruit) or dark brown (in yellow passion fruit) seeds.

Although a hermaphrodite, the passion fruit flower is self-sterile. Cross-pollination depends on pollinators such as large bees, the carpenter bee (*Xylocopa varipuncta*) and the honey bee (*Apis melifera*) being the most important. To overcome the lack of natural pollination, hand pollination is practiced in many countries. Artificial pollination has been found to be more efficient, and fruits from hand-pollinated flowers are larger and juicier than fruits from insect-pollinated flowers.

Passion fruit species have different periods for flower opening, which are usually short, rarely passing 8 h. The flowers of *P. edulis* open at dawn and close at midday; those of the *P. edulis* f. *flavicarpa* open at noon and close in the evening. Thus, the time for possible pollination is limited.

The purple and yellow passion fruits rarely hybridize in the field but can be crossed manually. A number of successful hybrids have been produced in Queensland, Florida, and Hawaii. Commercial production in Australia, in fact, now depends almost exclusively on commercial hybrids. Hybrids are intermediate in skin color and character between the two parents and have been selected in terms of agronomic characteristics, including resistance to viral and fungal diseases, high yield, extended cropping season, and flavor quality similar to the purple form.

In countries where the fruit is grown commercially, the usual method of propagation is by seeds. However, vegetative propagation, mainly from rooted cuttings and grafting, is becoming important to obtain disease-resistant plants and high yields.

The plant can grow in different types of soil, preferably sandy-clay soil, provided it is deep, relatively fertile, with good amount of organic matter, has good drainage and a pH between 5.0 and 6.5. A temperature between 21 and 32 °C favors development of this plant, the optimum being 26–27 °C. Long exposure to sunlight (between 10 and 12 h of daylight) is required for flowering. Annual rainfall should be between 800 and 1750 mm and well distributed throughout the year. Heavy rainfall during the peak of flowering will impede pollination. Irrigation is recommended during the dry periods, and pruning is necessary for good production. Frost and strong winds are limiting factors to passion fruit cultivation. The vines are usually supported by trellises in commercial plantations or by wires strung between wooden or concrete posts in small domestic farms.

The plant starts to produce fruit a year after planting, but has a short productive life, varying from 3 to 5 years, although longer periods have been reported. The highest yield is obtained on the second or third year, decreasing thereafter.

On average, 60–70 days are required from pollination to maturation of the fruit. Thus, two or more harvesting seasons per year are possible, depending on the climatic conditions, geographic location, cultural practices, and other factors. In the tropics, production is almost uninterrupted, but in subtropical regions, vegetation ceases in winter, production

being stopped during this period and in the spring months. Thus, in these regions, the productive period is reduced to 6 or 7 months per year.

Upon ripening, the passion fruits fall from the vine and are picked from the ground at least once a week, preferably more often in wet weather. The fruits are transported in lug boxes and, if not processed immediately, placed into cold storage. A temperature of 6.5 °C with a relative humidity of 85–90% is recommended. Under these conditions, the fruits can be stored for 4–5 weeks. Storage temperatures below 6.5 °C cause chill injury; at higher storage temperatures, the fruits suffer more losses due to moldiness.

Diseases and Pests

Production of passion fruit may be adversely affected by various diseases and insect pests. Common diseases are woodiness (thickening and hardening of the pericarp) or the mosaic disease, due to a virus transmissible by aphids, and several diseases of bacterial and fungal origin. Bacterial diseases include grease spot caused by *Pseudomonas passiflorae*, blast due to *Pseudomonas syringae* and bacterial spot caused by *Xanthomonas passiflorae* or *Xanthomonas campestris* pv. *passiflorae*. Diseases brought about by fungi are: brown spot by *Alternaria passiflorae*; leaf spot or septoria blotch by *Septoria passiflorae*; *Fusarium* wilt by *Fusarium* spp., such as *Fusarium oxysporum* f. *passiflorae*, or crown canker by *Fusarium lateritium* and *Fusarium sambucium*, which enter the plant at lesions and cracks; root rot caused by *Thielaviopsis basicola*; antrachnosis by *Colletotrichum gloeosporioides*; cladiosporioses by *Cladosporium herbarium*; and collar rot by *Phytophthora cinnamomi*.

Among the insect pests that can damage passion fruits, the most important are: (1) fruitflies, notably *Dacus dorsalis*, *D. cucurbitae*, or *D. vertebrates*, or larvae of the fruitflies *Anastrepha* spp., *Ceratitis capitata* and *Dacus tryoni*; (2) caterpillars of the butterflies *Dione juno juno* and *Agraulis vanillae*; (3) fruit mites *Tenvipalpus californicus*, *Brevipalpus papayensis*, *B. phoenicis*, *Polyphagotarsonemus latus*, *Tetranychus mexicanus*, *T. desertorum*; (4) pasion fruit or fruit bug *Diactus bilineatus* and *Holymenia clavigera*; and (5) leafhopper *Scolypopa australis*. Purple passion fruit vines can also be damaged by nematodes, such as *Meloidogyne javanica*, *M. arenaria*, *M. incognita*, *Scuttellona truncatum*, *Helicotylenchus* sp., and *Pratelenchus* sp. The yellow passion fruit is nematode-resistant. (*See* **Insect Pests**: Insects and Related Pests; Problems Caused by Insects and Mites.)

Chemical Composition

The chemical composition of the fruit and the juice is known to vary appreciably, being affected by factors such as variety, degree of ripeness, plant status, date of picking, season, climate, locality or region, soil composition, and cultural practices.

Although other tables on the composition of passion fruit are available in the literature, a list based on USDA data is presented in Table 1 because this is constantly updated.

The purple passion fruit juice is a good source of ascorbic acid and carotene, a fair to good source of riboflavin and niacin, and a fair source of mineral matter. On an overall basis, the yellow passion fruit has less ascorbic acid, °Brix, °Brix/acid ratio, and reducing and total sugars, but a higher acid and carotene content.

Carbohydrates constitute the second largest constituents of passion fruit (Table 1), and three sugars, glucose, fructose, and sucrose, make up 86% of the total carbohydrates, the rest being starch. The sugar composition of yellow passion fruit consists of 29%

Table 1 Composition of passion fruit and juice[a]

Constituents	*Fruit, purple*	*Juice, purple*	*Juice, yellow*
Proximate			
Water (g)	72.9	85.6	84.2
Food energy (kcal)	97	51	60
Protein (g)	2.20	0.39	0.67
Total lipid (fat, g)	0.70	0.05	0.18
Carbohydrate, by difference (g)	23.4	13.6	14.4
Total dietary fiber (g)	10.4	0.20	0.20
Ash (g)	0.80	0.34	0.49
Minerals			
Calcium (mg)	12.0	4.0	4.0
Phosphorus (mg)	68.0	13.0	25.0
Iron (mg)	1.60	0.24	0.36
Sodium (mg)	28.0	6.0	6.0
Potassium (mg)	348	278	278
Magnesium (mg)	29.0	17.0	17.0
Zinc (mg)	0.10	0.05	0.06
Copper (mg)	0.09	0.05	0.05
Selenium (mcg)	0.60	0.10	0.10
Vitamins			
Vitamin A (IU)	700	717	2410
Vitamin E (mg α-tocopherol equivalents)	1.12	0.05	0.05
Riboflavin (mg)	0.13	0.13	0.10
Niacin (mg)	1.5	1.5	2.2
Vitamin B_6 (mg)	0.10	0.05	0.06
Folate (μg)	14.0	7.0	8.0
Vitamin C (mg)	30.0	29.8	18.2

[a]Values per 100 g of edible portion.
From USDA Department of Agriculture, Agricultural Research Service (2002) *USDA Nutrient Database for Standard References, Release 15*.

fructose, 38% glucose, and 32% sucrose. The purple passion fruit has 34% fructose, 37% glucose, and 29% sucrose.

The nitrogen content of the juice is reported to vary from 0.10 to 0.19%, and the crude protein content ($N \times 6.25$) from 0.6 to 1.2%. The free amino acids found in purple passion fruit juice are leucine, proline, threonine, valine, tyrosine glycine, aspartic acid, arginine, and lysine.

The total acid content (expressed as citric acid) ranges from 2.4 to 4.8% w/w with an average of 3.4% for the purple passion fruit juice. The yellow passion fruit juice from Hawaii has an acidity ranging from 3.0 to 5.0%, with an average of 4.0%. The predominant acid is citric acid (about 83% of the acids), followed by malic acid (16%) and lesser amounts of lactic (0.87%), malonic (0.20%), and succinic (trace) acids. The same organic acids are found in the purple fruit juice, but the relative amounts are different. Citric acid is also the most abundant (41%), followed by lactic (23%), malonic (15%), malic (12%), and succinic (7.6%) acids.

An early study identified in purple passion fruit the carotenoids phytofluene, α-carotene, β-carotene, and ζ-carotene, of which β-carotene dominated. Later, β-apo-12′-carotenal, β-apo-8′-carotenal, cryptoxanthin, auroxanthin, and mutatoxanthin were also detected. Recently, 13 carotenoids were conclusively identified in yellow passion fruit: phytoene, phytofluene, ζ-carotene, neurosporene, β-carotene, lycopene, prolycopene, monoepoxy-β-carotene, β-cryptoxanthin, β-citraurin, antheraxanthin, violaxanthin, and neoxanthin, with the yellow ζ-carotene predominating. A recent study showed cyanidin 3-glucoside to be responsible for the purple color of the rind of *P. edulis*, accounting for 97% of the total anthocyanin content. Small amounts of cyanidin 3-(6″-malonylglucoside) (2%) and pelargonidin 3-glucoside (1%) were also found.

As in other fruits, the volatile composition of passion fruit is extremely complex, consisting of numerous compounds, and the unique flavor and aroma cannot be attributed to any single constituent. A compendium in 1990 listed 84 esters, 33 aldehydes and ketones, 71 alcohols and acids, and 25 miscellaneous volatiles identified in passion fruit. The volatile constituents present in the highest concentrations are the C-2 to C-8 esters of the C-2 to C-8 fatty acids that occur in many fruits. Some volatiles are probably degradation products of carotenoids, such as linalool, β-ionone, dihydro-β-ionone and the lactone of 2-hydroxy-2,6,6-trimethylcyclohexylideneacetic acid (dihydroactinidiolide), an oxidation product of β-ionone, these compounds being detected in both purple and yellow passion fruit. Dihydro-β-ionone and 1,1,6-trimethyl-1,2-dihydronaphthalene (3,4-dehydroionene), a possible degradation product of β-ionone, have been found only in purple passion fruit. Other compounds, such as edulans, megastigmatrienes, and sulfur-containing compounds, not included in the listing cited above, have also been reported as volatile constituents of passion fruit.

Out of some 300 volatile flavoring constituents of passion fruit drinks, only 22 were found likely to contribute to the passion fruit flavor. Of these, 6-(but-2′-enylidene-1,5,5-trimethylcyclohe-1-ene and (*Z*)-hex-3-enyl butanoate contributed 30 and 11%, respectively, of the passion fruit flavor profile. Several esters present in the juice at relatively high concentrations had negligible flavor impact values.

The flavor and aroma of the purple passion fruit are often described as more pleasing than those of the yellow fruit; this difference is reflected in the volatile composition. The purple passion fruit has higher levels of the major esters ethyl, butyl, and hexyl butanoates, and butyl and hexyl hexanoates, and of the terpene ketones, β-ionone and dihydro-β-ionone. Some important flavor constituents of the purple passion fruit are absent in the yellow fruit: 2-heptyl acetate, butanoate, and hexanoate, the edulans, dihidroionone, 1,1,6-trimethyldihydronaphthalene, and the megastigmatrienes. There are also differences in the terpene hydrocarbons, the yellow fruit having relatively high concentrations of (E)-β-ocimene, myrcene, limonene, and 1,4-*p*-methadiene, whereas only (E)-β-ocimene is present in a significant concentration in the purple variety. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics.)

The immature fruits have lower juice, sugar, ascorbic acid, and carotene contents. These fruits are more acidic and inferior in flavor compared with the partially and fully ripe fruits. Changes in the volatile composition during maturation have also been reported, the compounds of greatest flavor significance reaching maximum concentrations in the fruit that have just fallen from the vine. This observation indicates that the common practice of harvesting passion fruit after it falls is likely to provide fruit of the best flavor, provided that gathering from the ground is not delayed. Another study, however, found the fallen fruit to be lower in soluble solids, sugar, acidity, and ascorbic acid, suggesting that the customary harvesting practice should be avoided. Anthocyanins and carotenoids also increase during ripening.

Production, Utilization, and Processing

Reliable and updated production statistics on passion fruits are not available. Although it is one of the better known so-called exotic fruits, it has not reached the

point of being one of the fruits for which production is monitored by FAO. The major producing countries, listed in decreasing order of area of production in 1987, are: Brazil, Peru, Sri Lanka, Ecuador, Australia, Kenya, South Africa, Venezuela, Papua-New Guinea, Fiji, New Zealand and the USA (Hawaii). These countries account for 80–90% of the world's production of passion fruit. Taiwan and Colombia have been markedly increasing production, a good part of which is destined for export. Commercialization of the fresh fruit is limited, with Kenya being the principal exporter to Europe, especially to the UK. It is estimated that 50% of the world's production of passion fruit juice is exported. Brazil, Peru, and Colombia are the principal exporters of the juice, the major importers being the UK, Germany, France, Switzerland, the USA, and Japan.

After cutting the fruit in half, the pulp and seeds can be scooped with a spoon and eaten as is, or the pulp can be sieved to make a refreshing drink. Because of its intense flavor and high acidity, the passion fruit juice is considered a natural concentrate and is often diluted, sweetened, or blended with other fruit juices. The whole or sieved pulp is also used as a flavoring for yogurt, icecream, sherbet, meringue, or cake topping.

In Australia, there is appreciable consumption of the fresh fruit, although the bulk of passion fruit production is processed into juice, consumed locally as carbonated beverages. The juice is also used as a flavoring for icecream, confectioneries, and tropical fruit salads. Australian consumers are used to eating and drinking passion fruit products with the seeds still present, the seeds being regarded as evidence of passion fruit content. Elsewhere, the seeds must be removed.

Production of passion fruit in Fiji and Papua-New Guinea is largely for export to Australia and New Zealand as frozen, unsweetened pulp or juice. A significant part of production in Fiji is also consumed in homes and restaurants as mixed drinks. Passion fruit products in Sri Lanka include jams and sweetened and unsweetened juices. In India, this fruit is processed into passion fruit squash.

Most of the passion fruit produced in Hawaii is consumed as drinks blended with other fruit juices, such as orange and/or guava, the remainder being used as frozen juice bases. Passion fruit juice is considered too acidic for icecream, but this characteristic is advantageous in the preparation of sherbet. Minor uses are as a flavoring for syrups and as a pie filling.

The principal passion fruit products in Brazilian market are the juice and the frozen pulp, these two products serving as the base for other products such as drinks, yogurt, icecream, confectioneries (cakes, meringues, and chocolate fillings), gelatin, marmalades, and fruit cocktails. Fresh fruits are also marketed. In Venezuela, popular products include passion fruit juice, passion fruit icecream, and a bottled passion fruit and rum cocktail.

Passion fruits are preserved by freezing or thermal processing. Two characteristics of this fruit favor freezing. Firstly, the flavor is extremely sensitive to heat, so it is difficult to heat-process the juice without markedly altering the flavor. Secondly, the high starch content causes the accumulation of gelatinous deposits on the heating surfaces of the heat exchanger, lowering its efficiency, as well as causing deterioration of juice flavor. Enzymatic degradation of starch and centrifugal procedures for removing starch have been recommended.

Production of passion fruit juice can be done manually, as in many small cottage industries, consisting simply of hand-slicing the fruit, scooping out the pulp, and separating out the seeds either through sieving or expression through a cloth.

In Hawaii, the processing of passion fruit is highly mechanized, a centrifugal separator being the favored method for extraction of the pulp. A typical extractor has a capacity of 1725 $kg\,h^{-1}$ of passion fruit with an extraction efficiency of 94%. Its main disadvantages are: (1) some seeds are cut in the slicing operation, which necessitates the use of a fine screen in the finishing operation; and (2) there may be some extraction of skin flavors.

In Australia, the converging cone extractor is the most commonly employed extractor. In another method, a modified apricot-pitting machine and plunger are used. Since Australians are accustomed to consuming passion fruit products with the seeds, further processing to remove the seeds from the pulp is not necessary. Elsewhere, consumers' preference for seedless passion fruit products necessitates further processing.

The extraction unit most commonly utilized in Brazil is a three-stage system, consisting of a cutter, a perforated cylinder with a series of beaters that separates the rind from the pulp and seeds, and a pulper that separates the juice from the seeds and does the finishing of the juice.

Much of the earlier canned passion fruit juice was considerably overcooked. More recent processes make use of slightly higher temperature and shorter heating times. The most successful method for thermal processing of the juice employs a spin cooker. This cooker is an inexpensive, easily constructed unit, utilizing rapid can rotation to transfer heat quickly. Pasteurization to an 88 °C center temperature is achieved in about $1\frac{3}{4}$ min, after which the cans are rapidly cooled with a cold water spray while still

rotating. The color and flavor retention in this method is much better than in any alternative method of heat preservation.

To concentrate passion fruit juice, centrifugal evaporators have been successfully used in Brazil and Australia. The main advantage is the short residence time (0.2–1.0 s), which minimizes heat damage. Passion fruit concentrate can be stored at −18 °C for 6 months, 4 °C for 3 months, and 20 °C for 1 month with good color and flavor retention and no microbial spoilage.

Passion fruit juice may be quick-frozen directly from the finisher. Preferably, a slush-type or scraped-surface freezer should be used to hasten the freezing process, aside from increasing the freezing capacity of the processing plant. The juice may also be frozen directly in containers in an air-blast freezer. This product is sold to manufacturers of juice blends and of foods with passion fruit as an ingredient or as a major flavor component.

According to folk medicine, passion fruit has sedative and muscle-relaxant properties. Medicinal utilization of the leaves has also been cited.

In the extraction of juice from the passion fruit, about two-thirds of the bulk is refuse, of which 90% is rind and about 10% is seeds. Passion fruit rind has been found to be satisfactory as a supplementing feedstuff for dairy cows, and so it is commercially utilized as feed for dairy animals in Hawaii. In Brazil, the passion fruit rind is used as a component of rations for cattle and hogs. Other possible uses of subproducts, such as pectin from the rind and oil from the seeds, have not reached the industrial scale.

See also: **Acids**: Natural Acids and Acidulants; **Amino Acids**: Properties and Occurrence; **Ascorbic Acid**: Properties and Determination; Physiology; **Canning**: Quality Changes During Canning; **Carbohydrates**: Classification and Properties; **Carotenoids**: Occurrence, Properties, and Determination; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Freezing**: Structural and Flavor (Flavour) Changes; **Heat Treatment**: Chemical and Microbiological Changes; **Insect Pests**: Insects and Related Pests; Problems Caused by Insects and Mites; **Minerals – Dietary Importance**; **Vitamins**: Overview

Further Reading

Casimir DJ, Kefford JF and Whitfield FB (1981) Technology and flavor chemistry of passion fruit juices and concentrates. *Advances in Food Research* 27: 243–295.

Chan HT Jr. (1993) Passion fruit, papaya and guava juices. In: Nagy S, Chen CS, and Shaw PE (eds) *Fruit Juice Processing Technology*, pp. 334–377. Auburndale: Agscience.

Choucair K (1962) *Fruticultura Colombiana Tomo II Frutas Tropicales–Subtropicales y de Clima Templado y Frio*, pp. 795–850. Medellin: Editorial Bedout.

Escobar LK (1993) Passion fruits. In: Macrae R, Robinson RK, and Sadler MJ (eds) *Encyclopaedia of Food Science, Food Technology and Nutrition*, vol. 5, pp. 3423–3428. London: Academic Press.

ITAL (1994) *Maracuja. Cultura, matéria-prima, processamento e aspectos econômicos*. Campinas: Instituto de Tecnologia de Alimentos.

Kidoy L, Nygard AM, Andersen OM *et al*. Anthocyanins in fruits of *Passiflora edulis* and *P. suberosa*. *Journal of Food Composition and Analysis* 10: 49–54.

Mercadante AZ, Britton G and Rodriguez-Amaya DB (1998) Carotenoids from yellow passion fruit (*Pasiflora edulis*). *Journal of Agricultural and Food Chemistry* 46: 4102–4106.

Pruthi JS (1963) Physiology, chemistry, and technology of passion fruit. *Advances in Food Research* 12: 203–282.

Ruggiero C (ed.) (1980) *Cultura de Maracujazeiro*. Jaboticabal: UNESP.

Salunkhe DK and Desai BB (1984) *Postharvest Biotechnology of Fruits*, vol. II, pp. 53–58. Boca Raton, FL: CRC Press.

Shibamoto T and Tang CS (1990) 'Minor' tropical fruits – mango, papaya, passion fruit and guava. In: Morton ID, and MacLeod AJ (eds.) *Food Flavours Part C. The Flavour of Fruits*, pp. 221–280. Amsterdam: Elsevier Science.

USDA Department of Agriculture, Agricultural Research Service (2002) *USDA Nutrient Database for Standard Reference, Release 15*.

PASTA AND MACARONI

Contents

Methods of Manufacture

R Cubadda, University of Molise, Campobasso, Italy
M Carcea, National Institute for Food and Nutrition Research, Rome, Italy

Background

The current range of products referred to as pasta is relatively vast, and products vary widely in terms of shape, color, composition, storage requirements, and use. However, all share a basic technology that involves the preparation of a dough made by mixing a flour with a liquid (mainly water), which is then processed (namely by extrusion) to obtain the required shape and dimension of the product itself.

It is a popular belief that pasta originated in China, where it was manufactured with soft wheat flour. From there, it is supposed to have been introduced into Europe (Italy) by Marco Polo in AD 1292 after his travels to the Far East. However, the production of dried pasta, as it is manufactured today, is comparatively recent. The first attempt to produce extruded pasta instead of cut noodles from a sheeted dough was made in Italy in the 1800s using a hand-operated mechanical wooden press. For many years, pasta production was home-based, and only at the turn of this century did it become an industry, owing to the invention of the steam engine and hydraulic presses, and to the development of artificial instead of natural drying. The first continuous press, which replaced the traditional batch method, was introduced in 1933. The final step towards a fully automated system was achieved in the early 1950s, with the introduction of the automatic weighting and packaging equipment in pasta factories.

Today, the whole process, from reception of raw materials to mixing, kneading, extrusion, drying, packaging, and dispatching, can be performed automatically by several pieces of machinery controlled by computers. Moreover, the diffusion of pasta throughout the world and the popularity of the traditional dry product have stimulated the development of production technologies leading to new convenient commodities such as fresh, precooked, and frozen pasta.

Raw Materials and Their Properties

The raw material of choice for pasta production is semolina from durum wheat, although, for various reasons, the industry uses flours from soft and hard wheats, corn, and other cereals. Durum wheat semolina is the only raw material permitted for pasta production by national laws in Italy, France, and Greece. Special pasta is also produced by adding a great variety of other ingredients (i.e., fresh, frozen, or powdered eggs, powdered spinach, tomatoes, carrots and other vegetables, soy protein, wheat gluten, milk protein, vitamins, and minerals) to the two basic components, semolina (or flour) and water. This section deals with the production of pasta using the traditional raw material durum wheat semolina. However, the manufacture of pasta with flour or blends of flour and semolina is also mentioned. (*See* **Wheat**: The Crop; Grain Structure of Wheat and Wheat-based Products.)

Semolina quality requirements vary from country to country. In Germany, Austria, and Switzerland, for example, semolina color is very important, whereas in Italy, the factors affecting cooking quality are synonymous with semolina quality. In today's pasta industry, uniform semolina granulation is also a desirable factor for uniform flow in semolina feeders and for proper dough development in continuous presses. A high ash content is usually indicative of longer-extraction semolina and imparts a dull color. If the wheat is not properly cleaned, or the kernels are smudged and/or damaged by severe mildew, brown and black specks can appear in the semolina. Commercial semolina with fewer than 200 specks per dm^2 is desirable, and gives pasta with a good esthetic appearance. A low level of lipoxidase, which, during pasta processing, may destroy the yellow color given to semolina by natural carotenoid pigments (carotenes and xanthophyls), is also desirable. Nevertheless, the major quality requirement in most traditional pasta consuming areas is the ability of semolina to be processed into pasta with a good cooking quality.

According to the existing literature, gluten quality and quantity are the most important factors affecting cooking quality, but starch and minor constituents such as soluble and insoluble pentosans, lipids,

lipoproteins, various enzymes, and products of enzyme reactions could also be involved. At present, little information is available on the physicochemical changes and on the behavior of starch and minor constituents during processing, cooling, and storage of pasta. It is well known that semolina quality varies widely among durum wheat varieties. Semolinas from certain durum wheat varieties grown in some locations produce pasta with a good cooking quality and/or yellow color, whereas other durum wheat varieties give semolina that produces pasta with insufficient cooking characteristics or pigment. Consequently, attention is paid to check the quality of durum varieties in order to produce the desired semolina.

Common soft wheat flour is less pigmented, translucent, and flinty than durum wheat semolina, and gives a gluten with an elastic strength suitable for making bread, but not for manufacturing a superior pasta product. However, flours milled from common hard wheats with strong gluten, a high protein content, and superior rheological characteristics are acceptable for producing pasta. The color and cooking quality of pasta from common wheat flour can be improved by adding proportional amounts of egg. As far as the water used for pasta making is concerned, it should be pure drinking water with no off-flavors.

Method of Manufacture

In a modern pasta factory, pasta products are formed by extrusion on large automatic machines that perform several processing operations. Production lines for processing short and long pasta are shown in Figure 1. The essential equipment includes:

- A series of metal silos that receive the various types of semolina or flour (coming either directly from a mill adjoining the pasta factory or from outside mills);
- An automatic continuous press;
- A spreader predrier for long pasta or shaker predrier for short pasta;
- A drier; and
- A storage and packing unit.

(*See* **Extrusion Cooking**: Principles and Practice.)

Mixing

The automatic press is one of the most important machines in a pasta factory. It performs three essential processing operations: mixing, kneading, and extruding.

In the mixing stage, water is added to the semolina so that the moisture content of the dough is approximately 30%. The flow of ingredients is regulated by volumetric or gravimetric dosers with constant and proportioned outputs. It is very important that the water is absorbed evenly by the semolina or flour particles to obtain a thoroughly homogeneous dough that will prevent the dried pasta from having faults (white spots, for instance). For this reason, the

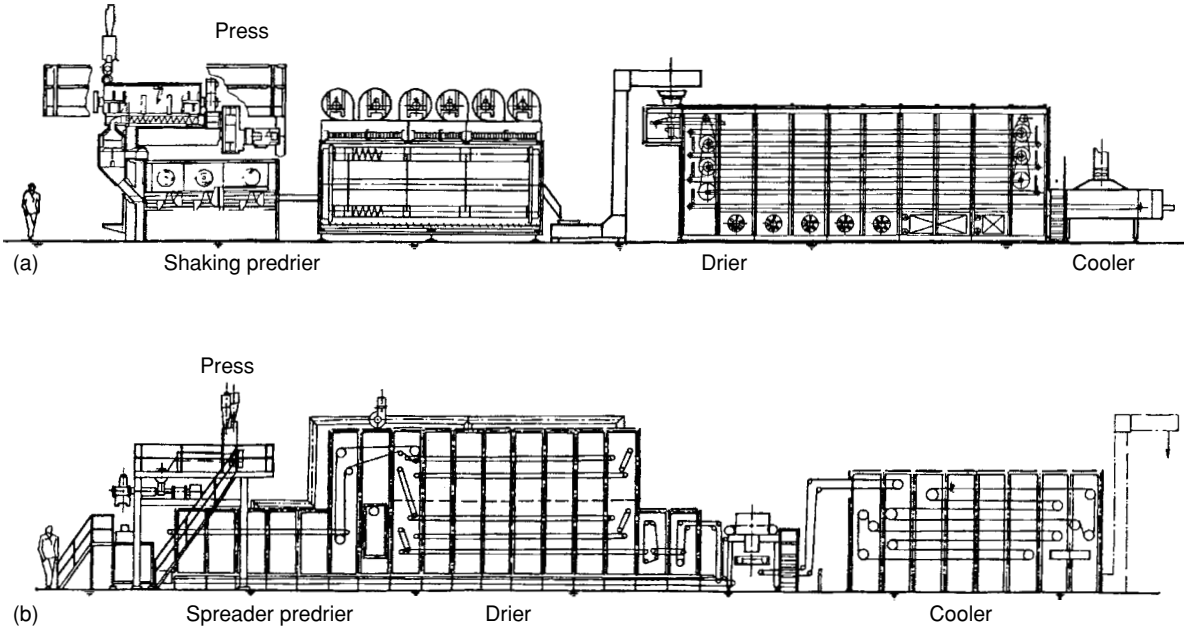

Figure 1 Modern production lines for processing (a) short and (b) long pasta. Reproduced from Pasta and Macaroni, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

semolina particles must have the same size (for a typical semolina, a maximum of 25% of particles should pass through a 0.180-mm mesh sieve), and the time needed for the liquid to be absorbed by the particles has to be evaluated while taking into account their average size and the temperature of both flour and liquid. Steps should be taken to insure that the lower the flour temperature, the higher the temperature of the liquid in the dough.

Depending on the pasta shape and type, warm water at 40–65 °C is generally used. A uniform dough is obtained by mixing the ingredients in a special twin-shaft mixing chamber for 15–17 min for short pasta, and 16–20 and 16–18 min for long pasta and long egg pasta, respectively. The shafts of the mixer turn in opposite directions to limit the amount of dough balling that can occur.

The latest development in press technology is represented by new systems that are able to carry out the dough-making in a very short time (about 2–3 min) instead of the 15–20 min necessary for normal dough preparation. A very short dough-making time increases considerably the line production speed, since it can better match the speed of the following steps (predrying, drying, etc.). For this reason, at present, fine granulometry semolina is preferred, which best suits short dough-making times. Moreover, short dough-making times allow for a reduction in the press size (output capacity being equal) and consequently a better plant compactness and smaller dimensions.

Modern continuous presses are equipped with vacuum-producing equipment to remove any air bubbles from the pasta dough during mixing and just before extrusion. The aim of this operation is to avoid the formation of small bubbles, which can give the end product a white, chalky appearance. In addition, air bubbles can facilitate the oxidation of pigments, resulting in a pale and unattractive pasta, and can weaken the dried product.

Extruding

The extrusion screw is the core of the pasta press, and special attention is paid to its design and construction, just as to the compression cylinder. In fact, it is vital to prevent the dough from being overworked, an occurrence that would adversely affect the quality of the finished product. The dough is forced by the screw into the extrusion head. Even under the best operative conditions, a considerable amount of heat is generated during this process; therefore, the barrel and the extrusion head are equipped with a water-cooling jacket to reduce the heat and to maintain a constant extrusion temperature of no more than 40 °C. Higher temperatures could cause the deterioration of the cooking quality of the finished product.

The process in the press is completed by the extrusion of the dough through a die to make a variety of product shapes that are cut to the required length with a rotary cutter. Until recently, pasta dies were made of bronze. Now, new dies with Teflon on the extruding surface are used. Teflon extends the life of the die and improves the appearance of the product. Bronze dies are still used for small special productions characterized by the rough surface of pasta. Special equipment is required to form particular shapes such as nests, skeins and Bologna-type pasta. The cut pasta is then hung over drying rods or spread evenly, all automatically, ready for the drying stage.

Drying

Drying is the most difficult and critical step in the pasta-making process. The objective of this process is to decrease the moisture level of the extruded pasta from about 30% to less than 12.5% without causing any damage to the finished product. To achieve this, the air temperature and relative humidity must be properly controlled. If the drying is too slow, the pasta products tend to spoil or become moldy. However, if they are dried too rapidly, moisture gradients occur that may cause checking or cracking. Usually, uniform temperature and ventilation are insured in all parts of the dryer. This is achieved through a suitable hot-air circulation system using low-pressure spiral ventilators located in various parts of the dryer. Moreover, automatically controlled periods of hot-air circulation alternate with periods of rest to allow for moisture equilibration between the inner part of the pasta and its surface.

In modern plants, the drying cycle can be subdivided into three distinct stages: (1) predrying, (2) drying, and (3) cooling. Although identical basic components are used, the manufacture of different types of pasta requires machines that are technically different. The most widely used dryer for long pasta is basically made of a single complex consisting of:

- an extruded-pasta supporting structure on one or more levels;
- a separate drive unit for the section on one or more levels;
- a heat ventilation unit;
- an unloading unit; and
- a regulation and control unit.

The initial predrying takes place in the first part of the dryer, where special aluminum or steel rods loaded with pasta are normally at only one level. During this stage, the product moisture is reduced to 17–18%,

as a consequence of quick evaporation of the inside water excess of pasta. Surface drying has to be avoided to prevent closing of the interstices through which the remaining water can reach the surface by capillarity and evaporates. This predrying stage lasts a short time and is followed immediately by a longer drying stage.

For the final drying, driers with more than one level are used. The product hanging on the rods is moved by means of special conveyor chains in the one-level section and with a system of racks, which alternate vertical and horizontal movements, in the section with more than one level. From a technological point of view, the drying differs from the predrying phase for a more gradual moisture reduction and particularly for the alternation of water evaporation and its even distribution inside the product.

The predrying stage of short pasta is generally carried out with vibrating trays, called shaker predryers, whereas there is a great diversification in the machines used for the next drying steps. Two different kinds of equipment are used: the belt dryer and the rotary dryer. The belt dryer is more suitable for large shapes, which must be treated delicately. These dryers receive the product from the shaker predriers by means of conveyor belts. An oscillating band evenly distributes the product on the whole upper belt width. The product, after having reached the far end of the belt, falls on the following ones, repeating this route as many times as the existing number of belt dryer tiers. Drying is achieved by means of aerothermal units in which hot water is kept in forced circulation by a pump. The rotary dryers can be used for both the preliminary and final drying of small- and medium-sized and Bologna-style pasta products. The rotary driers are essentially composed of a rotating drum, supported on rollers and driven by a multiple-speed gearbox. The product advances forward one cell for each revolution of the drum. Drying is achieved by means of a battery of air-circulating fans and heat radiators that are located above the drum.

At present, the drying technology varies greatly from factory to factory and from country to country, the main difference being in the degree of temperature used during the process. However, two systems can be recognized: (1) traditional or conventional drying, which uses 'low' temperatures not higher than 60 °C, and (2) 'high'-temperature drying, which uses temperatures between 70 and 85 °C, even though there is a trend to seek higher temperatures. Of course, the total drying time varies, depending on the system employed. High-temperature drying has been without doubt the most important innovation in pasta technology in recent times. This new technique has the advantage of appreciably reducing the drying time, compared with traditional methods, and of reducing microbial contamination. The former characteristic allows the construction of drying lines that maintain the same production capacity as the traditional lines, but they are smaller, more compact and constructed to facilitate maintenance operations.

With the increasing popularity of high-temperature drying among pasta manufacturers, considerable attention has been given to the effect of this treatment on the cooking quality, color, and chemical composition of the final product. The effect of high-temperature drying on the cooking quality is still controversial. The main reason is that the available information comes from investigations where the method of application of the high temperature during the drying cycle and the intrinsic characteristics of the raw materials are different. However, experimental evidence is accumulating to prove that semolina protein quality and quantity can influence cooking quality jointly or independently as a consequence of the drying technology adopted (i.e., low- or high-temperature drying). Different temperatures modify components and influence cooking quality in different ways. At low temperatures (40–50 °C), protein quantity and quality are equally important in determining the final pasta results, whereas at high temperatures (≥ 80 °C), protein quantity is more important and is strongly correlated with the resultant pasta quality. Some researchers have reported that pasta proteins are in fact polymerized by high drying temperatures, whereas starch granules appear embedded in the protein matrix. As a consequence, there may be an improvement in the cooking quality of pasta. Pasta cooking quality is determined by a physical competition between protein coagulation into a continuous network and starch swelling with spherules scattering during cooking. If the former prevails, starch particles are trapped in the network alveoli, promoting firmness in cooked pasta, whereas if the latter prevails, the protein coagulates in discrete masses lacking a continuous framework, and pasta will show softness and usually stickiness on cooking. High-temperature drying partially overcomes this competition by producing a coagulated protein framework in dry pasta without starch swelling.

Drying conditions also have a significant effect on the retention of lysine in spaghetti. As lysine is the limiting amino acid in wheat, any reduction in its availability has an important bearing on the nutritional quality of pasta. A decrease in the amount of total lysine has been reported as being directly proportional to the increase in temperature. The loss of lysine depends mainly on the blockage of the amino

group as a consequence of the Maillard reaction. (*See* **Browning**: Nonenzymatic.)

Cooling and Packaging

The final drying is followed by a cooling treatment that must be carried out with certain precautions to avoid damage to the pasta through humidity imbalance within the shape. After this stage, the end product is unloaded, sent to storage silos, and then sent to the packaging section.

Pasta products are packaged in various ways, depending on their particular shapes and sizes. The most widely used packages are cellophane bags and packets, cartons, and special trays. All these are automatically weighed, packed, and collected by machines on continuous cycles. Sometimes, semi-automatic machines may be required.

Considerations of Quality

A good cooking quality is the most important requirement for pasta products. Aroma, taste, and, sometimes, uniformity of shape and color and breaking strength of uncooked pasta are also important to the consumer. Although it can be interpreted according to the individual taste and habits of the consumer, the cooking quality of pasta is generally regarded as the capacity of the product to maintain its shape when cooked in boiling water and to retain a good texture after cooking without becoming a thick, sticky mass. The parameters involved in the evaluation of pasta cooking quality can be defined as follows:

- stickiness is the state of surface disintegration of cooked pasta;
- firmness is the resistance of cooked pasta when sheared between the teeth;
- bulkiness is the degree of adhesion of pasta shapes after cooking.

As already mentioned, cooking quality depends essentially on the type of raw materials used and on their intrinsic characteristics, although it can be affected by the manufacturing conditions.

As far as the nutritional quality of pasta is concerned, the nutritive value appears to be completely independent of the use of durum wheat semolina or bread wheat flour. However, the addition of other ingredients such as eggs, vegetables, soy or vegetable protein concentrates and isolates, individual nutrients, etc. produces remarkable changes in the nutritional value of common pasta.

The protein quality of pasta can be influenced by a reduced bioavailability of lysine caused by severe drying conditions. This has to be considered when adding high-quality protein integrators to the raw materials.

See also: **Browning**: Nonenzymatic; **Extrusion Cooking**: Principles and Practice; **Wheat**: The Crop; Grain Structure of Wheat and Wheat-based Products

Further Reading

Cubadda R (1988) Evaluation of durum wheat, semolina, and pasta in Europe. In: *Durum Wheat: Chemistry and Technology*, pp. 217–228. St. Paul, MN: American Association of Cereal Chemists.

Cubadda R (1989) Current research and future needs in durum wheat chemistry and technology. *Cereal Foods World* 34: 206–209.

Cubadda R (1996) Pasta quality: the relationship between raw material properties and production technologies. *Proceedings of 1st World Pasta Congress*, pp. 164–168. Pinerolo, Italy: Chiriotti Editori.

D'Egidio MG, Mariani BM, Nardi S, Novaro P and Cubadda R (1990) Chemical and technological variables and their relationships: a predictive equation for pasta cooking quality. *Cereal Chemistry* 67: 275–291.

Grzybowski RA and Donelly BJ (1979) Cooking properties of spaghetti: factors affecting cooking quality. *Journal of Agricultural and Food Chemistry* 27: 380–384.

Lirici L (1999) From the bar kneader to the modern automatic press: over 100 years of evolution. *Tecnica Molitoria* 50: 101–120.

Pavan G (1996) Pasta: processing trends and possible technological answers. *Proceedings of 1st World Pasta Congress*, pp. 79–94. Pinerolo, Italy: Chiriotti Editori.

Resmini PP and Pagani MA (1983) Ultrastructure studies of pasta. *Food Microstructure* 2: 1–12.

Dietary Importance

L J Malcolmson, Canadian International Grains Institute, Winnipeg, Manitoba, Canada

Background

The term 'pasta' is used to describe products made from a flour–water dough that fit the 'Italian' style of extruded foods. The primary ingredient used in the manufacture of pasta is a coarsely ground flour from durum wheat, called semolina. In the commercial manufacture of pasta, semolina and water are mixed into a dough, which is then extruded through a die to produce the desired size and shape of pasta. The product is then dried or sold fresh. Common pasta products

include spaghetti, vermicelli, linguine, rotini, rigatoni, and macaroni.

In contrast, the term 'noodle' is used to describe an 'Oriental' style of sheeted and cut products made from grains other than durum wheat. Grains used to produce noodles include common or bread wheat, rice, barley, and buckwheat, as well as legumes such as mung bean.

The earliest written record of pasta in the Italian diet was reported to be 1279, although noodle products were most certainly part of earlier cultures of China and the later cultures of many other countries. Today, pasta products have become an important part of the diet in many countries. This has been influenced in part by changes in consumers' perception of pasta. Once thought of as starchy and fattening, pasta is now considered to be nutritious and convenient. Table 1 lists the average per capita consumption of uncooked pasta in selected countries. The low consumption of pasta products in Asian countries such as China and Japan compared with other countries is a reflection of the much higher consumption of noodle products in these countries compared with 'Italian'-style pasta products.

The widespread consumption of pasta throughout the world is likely due to its simple formulation, relative ease of processing, extended shelf-life in the dry form, low cost, and versatility. Pasta's versatility stems from the fact that it is available in numerous shapes and sizes, and can be prepared and served with other foods as an appetizer, main dish, side dish, salad, soup, or dessert, thus allowing menu flexibility for the consumer. Because pasta is relatively bland in flavor, it lends itself to be complemented by sauces and other food ingredients.

Table 1 Average yearly per capita consumption in selected countries

Country	*Per capita consumption (kg)*
Italy	28.0
Venezuela	12.7
Switzerland	9.1
USA	9.0
Russia	7.0
France	7.0
Greece	6.6
Canada	6.3
Germany	5.0
Spain	4.0
UK	2.0
Japan	1.8
China	0.8

Source: Pastavilla Pasta (2001) and Pastificio Pagani (2001).

Nutritional Value of Pasta

Although the nutritional content of pasta can vary widely, depending on the ingredients used in its preparation and added sauces, pasta provides significant quantities of complex carbohydrates, protein, B vitamins, and iron, and is low in fat. A typical 140-g serving of cooked macaroni or spaghetti provides approximately 6 g of protein, 40 g of carbohydrates, and 2 g of dietary fiber, whereas whole wheat pasta provides approximately 8 g of protein, 38 g of carbohydrates, and 3 g of total dietary fiber. The fat content of a 140-g serving of pasta is less than 1 g.

Fortification of pasta products with thiamin, riboflavin, niacin, iron, and, more recently, folic acid is permitted in many countries. The vitamin and mineral content of enriched and whole wheat pasta is provided in Table 2.

In recent years, studies have been undertaken to examine the beneficial role of pasta in the diet. Evidence exists indicating that a diet rich in pasta may lower the incidence of cardiovascular disease, be helpful in managing diabetes, and be useful to individuals with a specific gluten intolerance or individuals suffering from celiac disease.

According to a review published by Costantini, subjects who consumed diets rich in pasta had lower levels of serum low-density lipoproteins and total lipoproteins than those who did not consume pasta-rich diets. This finding suggests that diets rich in pasta may be responsible for lower incidents of cardiovascular disease, provided the pasta is prepared and served using other low-fat ingredients.

The exact role of pasta in the dietary management of diabetes is not fully understood. However, it has been speculated that diets rich in pasta result in reduced rates of both starch uptake and gastric emptying, thereby resulting in a lowered postprandial blood glucose and insulin response.

Table 2 Vitamin and mineral content of cooked pasta (140-g serving)

Product	*Thiamin (mg)*	*Riboflavin (mg)*	*Niacin (mg)*	*Folate (μg)*	*Iron (mg)*	*Calcium (mg)*	*K (mg)*	*P (mg)*	*Mg (mg)*	*Zinc (mg)*
Enriched	0.20	0.20	2.4	10.0	2.0	10.0	44	76	26	0.8
Whole wheat	0.20	0.00	1.0	8.0	1.4	22.0	62	124	42	1.2

K, potassium; P, phosphorus; Mg, magnesium. Source: NutriQuest™ (1999).

Peptides derived from the gliadins of durum wheat have been shown to be less aggressive to celiac disease. Thus, pasta derived from durum wheat may be useful in the diets of individuals suffering from gluten intolerance, depending on the severity of the reaction.

The nutritional value of pasta can be increased by the addition of other food ingredients to increase the protein or fiber content. Research has been undertaken to investigate the effects of various protein sources on pasta quality. Studies have been published showing the successful addition of whey derivatives, nonfat dry milk, soy, corn, legumes (bean, lentil, cowpea), lupin, amaranth, buckwheat, and maize. By blending durum wheat semolina with other plant protein sources, the limiting amino acids lysine and threonine are compensated for by the other protein sources. Protein-enriched pasta products are currently available in some countries. Egg-enriched pasta is also popular in many countries. Whole grain cereals such as barley and legumes as well as the use of whole wheat have also been used to increase the mineral, vitamin, and dietary fiber content of pasta.

The addition of natural colorants such as black squid ink, puréed spinach and carrot, tomato paste, and beet juice has also received considerable attention, and many of these products are currently being sold in the marketplace. The use of herbs and chocolate has also gained popularity in the production of more flavorful product lines. These ingredient additions are made to enhance the flavor and color of the final product, since the nutritional properties are not changed substantially.

Effects of Processing and Cooking on the Nutrient Content of Pasta

The effect of drying and cooking on the nutritional content of pasta has also been studied. Drying conditions have been found to significantly affect the bioavailability of lysine retention in spaghetti. Losses as high as 47% have been found for spaghetti dried at 80 °C and as low as 22% for pasta dried at 45 °C. Other researchers have reported that the extent of lysine losses is a function of both drying temperature and length of exposure to high drying temperatures. With continued improvements in drying technologies, the deleterious effects of high-temperature drying on available lysine will likely be minimized. Furthermore, whether these losses are significant, given that pasta is combined with other foods in the diet, is also questionable. Studies that have examined the loss of vitamins during drying reported higher losses of thiamin and riboflavin with increasing drying temperatures. These losses were smaller for thiamin under industrial drying conditions. The effect of cooking on the retention of B vitamins has suggested that a good portion, namely 50–75% of the vitamins, is retained following cooking. With fortification of B vitamins during processing, these losses are not as significant as they may be in products that are not fortified.

As a wheat-derived staple food, pasta is second only to bread in world consumption. Its worldwide acceptance can be attributed to its low cost, ease of preparation, versatility, long shelf-life, and nutritional and sensory properties. With the greater emphasis towards the consumption of cereal based foods as a result of changes in dietary nutritional recommendations, the importance of pasta in the diet of many countries will continue to remain high.

See also: **Celiac (Coeliac) Disease**; **Coronary Heart Disease**: Etiology and Risk Factor; **Pasta and Macaroni**: Methods of Manufacture; **Wheat**: Grain Structure of Wheat and Wheat-based Products

Further Reading

Costantini AM (1986) Nutritional and health significance of pasta in modern dietary patterns. In: Mercier Ch and Cantarelli C (eds) *Pasta and Extrusion Cooked Foods: Some Technological and Nutritional Aspects, Technoalimenti Food Technology and Nutrition Series, No. 1*. London: Elsevier Applied Science.

Cubadda R (1986) Effect of the drying process on the nutritional and organoleptic characteristics of pasta. A review. In: Mercier Ch and Cantarelli C (eds) *Pasta and Extrusion Cooked Foods: Some Technological and Nutritional Aspects, Technoalimenti Food Technology and Nutrition Series, No. 1*. London: Elsevier Applied Science.

Dick JW and Matsuo RR (1988) Durum wheat and pasta products. In: Pomeranz Y (ed.) *Wheat: Science and Technology*, vol. II. St. Paul, MN: American Association of Cereal Chemists.

Giese J (1992) Pasta: New twists on an old product. *Food Technology* 46(2): 118–126.

Ranhotra GS, Gelroth JA, Novak FA and Mathew RH (1985) Retention of selected B vitamins in cooked pasta products. *Cereal Chemistry* 62(6): 476–477.

Seibel W (1996) Future trends in pasta products. In: Kruger JE, Matsuo RB and Dick JW (eds) *Pasta and Noodle Technology*. St. Paul, MN: American Association of Cereal Chemists.

Pasta Filata Cheese *See* **Cheeses**: Types of Cheese; Starter Cultures Employed in Cheese-making; Chemistry and Microbiology of Maturation; Manufacture of Extra-hard Cheeses; Manufacture of Hard and Semi-hard Varieties of Cheese; Cheeses with 'Eyes'; Soft and Special Varieties; White Brined Varieties; Quarg and Fromage Frais; Processed Cheese; Dietary Importance; Mold-ripened Cheeses: Stilton and Related Varieties; Surface Mold-ripened Cheese Varieties; **Conjugated Linoleic Acid**

PASTEURIZATION

Contents

Principles

R A Wilbey, The University of Reading, Reading, UK

Introduction

Pasteurization has become widely accepted as an effective means of destroying vegetative pathogens in food products, with the least possible damage to the sensory qualities of the product. The heat treatment also reduces the general microbial population, so that an increased shelf-life is normally obtained. Bacterial spores and some heat-resistant enzymes will survive pasteurization processes and limit the shelf-life of the product.

Historical Origins

Cooking is an age-old method of preparing many traditional foodstuffs, and for centuries it was generally appreciated that cooked products would normally take longer to putrefy than if they were raw.

In the latter part of the 18th century there was great interest in understanding the mechanism of putrefaction. Lazzaro Spallanzani demonstrated that putrefaction may not occur in a heated sealed flask of an infusion, but that aerial contamination could result in putrefaction. The presence of microorganisms was demonstrated, and it was recognized that there was a possible division into organisms that could be killed by boiling and those that would survive this heat treatment. Subsequent experiments by Franz Schutz in the early 19th century demonstrated that it was not the air itself that caused spontaneous putrefaction, but a contaminant carried in the air. Similar experiments were carried out by Theodore Schwann at the same time.

Methods for the preservation of foodstuffs were developed in parallel with this pioneering work on the basic understanding of why foods spoil. Carl Wilhelm Scheele used heat for the conservation of vinegar, and Nicholas Appert developed the preservation of foods by heating in cans. In 1824 William Dewees recommended that milk for infants be heated to near to boiling (but not boiled) then cooled as preparation for infant feeding.

The credit for the development of a mild method of processing foods, now particularly associated with milk, has been given to Louis Pasteur, after whom the pasteurization process was named. Pasteur had, amongst his many interests, an interest in fermentations. The poor hygiene conditions associated with the production of food and beverages at that time often led to unwanted fermentations, causing putrefaction and loss of product. His experiments confirmed that fermentations were not spontaneous but the result of microbial metabolism. While some of his earlier work was with lactic fermentations, most of his work in this field was based on alcoholic fermentations, brought about by yeasts. The conversion of ethanol to acetic acid was demonstrated to be brought about by bacteria, subsequently classified as *Acetobacter*. Both yeasts and *Acetobacter* could be destroyed by relatively mild heat treatments, at about 55 °C in an acid medium such as wine, in closed vessels.

While Pasteur's work on beer, wine, and vinegar laid the foundations for hygienic processing, his complementary work on the relation between specific organisms and disease also aided the recognition of the public health implications of hygiene and of heat treatments.

By the late 19th century the economic benefits from improving the shelf-life of milk and other products were appreciated, though the microbiological and public health implications of pasteurization were not fully understood. Pasteurization of wine was adopted in both France and the USA and is still used for some wines, though filtration, higher alcohol levels, and better production hygiene have largely displaced heat treatment.

Introduction to the Dairy Industry

The heat treatment of milk on a commercial scale did not develop until the end of the 19th century, with the production of commercial pasteurizers in Germany and Denmark. The earlier treatment systems were aimed at improving the storage life of milk, often using simple continuous-flow techniques to reduce costs. The realization that milk was a potential carrier for diseases such as tuberculosis led to the development of a low-temperature long-time (LTLT) batch process, the first commercial plant being installed by Charles North in New York in 1907. It was not till 1922 that pasteurization was legally recognized in the UK when the term was defined in the Milk and Dairies (Amendment) Act, using a LTLT process at 62.8–65.5 °C for a minimum of 30 min.

LTLT pasteurization was the first safe method adopted, but the processing of milk was revolutionized by the invention in the UK of the plate heat exchanger, capable of regeneration. The development of a modular heat exchanger that could be relatively easily cleaned, together with a microbiologically effective holding tube system and a flow diversion valve, enabled milk to be heat-treated with safety on a far larger scale than had been possible with the batch-based LTLT system. With better appreciation of the thermal death characteristics of pathogens, this continuous process was able to take advantage of higher process temperatures with a corresponding shorter hold, and became known as the high-temperature short-time (HTST) process. In the UK, as in the European Community, the minimum heat treatment permitted is 71.7 °C for 15 s. It has been suggested that 15 s was originally chosen as the minimum time to allow an adequate margin for the response rate of the temperature-sensing and control system at that time.

Subsequent developments in the design of process equipment have led to the construction of pasteurization plants that may be cleaned in place (CIP), with much greater thermal efficiency and with much more sensitive and responsive instrumentation and control systems. It is now technically possible to pasteurize at higher temperatures with little or no hold – the so-called 'flash' processes.

Development of safe HTST processes for milk insured that in-bottle LTLT pasteurization of milk, pioneered by Charles North in 1911, although capable of producing a high-quality product virtually free of postpasteurization contamination, was too expensive and thus unable to compete in the marketplace. The only in-bottle milk processing that continued was for sterilized milk and this product is also diminishing with the development of continuous UHT processes. (*See* **Heat Treatment**: Ultra-high Temperature (UHT) Treatments.)

The LTLT process has found an application in cream processing. With most UK dairies now producing over half of their milk as semiskimmed plus some skimmed milk, large but variable volumes of cream are produced from the separation and inline standardization processes. Carrying out separation and cream holding at 63 °C avoids problems of varying throughputs with HTST processes.

The drive to minimize energy costs has led to the widespread use of regeneration in plate heat exchangers, where the cold raw material is preheated by the hot pasteurized product. This process can enable heat recoveries in excess of 90%, giving energy savings for both heating and refrigeration. There is a risk associated with this process, however, as in simple systems with one pump, the pressure of the raw liquid will be higher than that of the pasteurized product, so if there were a leak within the heat exchanger there could be contamination of the pasteurized product. Apart from regular checks on the integrity of the heat exchanger, for many products the problem can be avoided by introducing a second pump after the preheater plus some means of maintaining a higher pressure in the downstream section of the heat exchanger, either by flow restriction within the heat exchanger or by a backpressure valve at the exit. Where such pressure drops are undesirable, double heat exchanger plates may be used. These pairs of thinner heat exchanger plates are constructed without gaskets between the pairs so that, in the event of one of the pair failing, the leaking liquid would leak out through the gap between the plates to give a failure indication without the risk of contamination of the other flow path. Maintenance of heat exchanger plates has been eased by the introduction of clip-in gaskets that do not require adhesive.

Basic Aims of Process

The basic aim of pasteurization is summarized by the definition adopted by the International Dairy Federation (IDF): 'pasteurization is a process applied to a product with the aim of avoiding public health hazards arising from pathogenic microorganisms associated with milk by heat treatment which is consistent with minimal chemical, physical and organoleptic changes in the product'.

This definition would be equally applicable to other commodities if one were to substitute the name of that product for 'milk' in the definition.

To achieve the public health objective of pasteurization in a particular product, it is essential to be aware of the pathogens associated, or potentially associated, with that product. The thermal death characteristics of the organism(s) in that product must also be known.

In most of the early work the death characteristics were expressed in terms of a temperature–time combination that would destroy the target pathogen. In the case of milk, tuberculosis was recognized as a major disease associated with milk consumption and *Mycobacterium tuberculosis* was found to be the most heat-resistant pathogen normally associated with milk. Temperature–time combinations needed to destroy *M. tuberculosis* were published by North in 1911, North and Park in 1927, Hammer in 1928 and Dahlberg in 1932; these data are included in Figure 1.

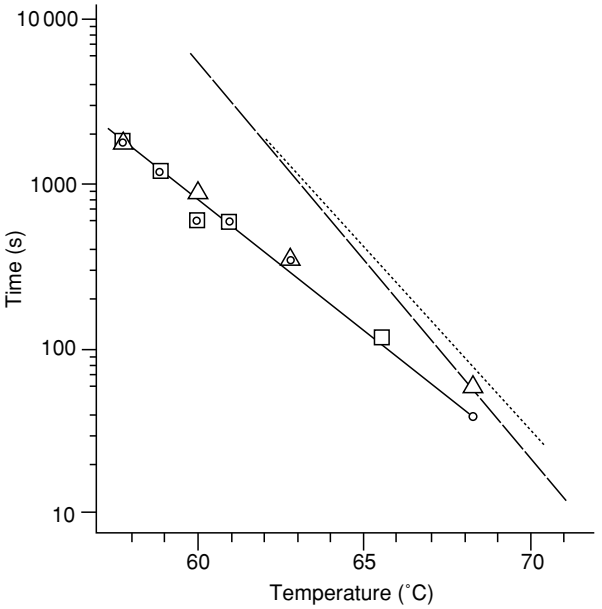

Figure 1 Time–temperature conditions for destruction of *Mycobacterium tuberculosis* in milk. Continuous line, based on data from North and Park (squares), Hammer (triangles), and Dahlberg (circles). Dashed line, virtual loss of alkaline phosphatase. Dotted line, reduction in cream line. Adapted from North and Park (1927) The American Journal of Hygiene 147: VII March, cited by Davis JG (1955) *A Dictionary of Dairying* p. 787. London: Leonard Hill.

This knowledge enabled safer process conditions to be set up for LTLT and subsequently for HTST processes. However, with better understanding of the kinetics of thermal death rates, more quantitative data may now be obtained.

When organisms are subjected to a moist heat above their normal temperature range there is a drop in the number of survivors with time: the logarithm of the number of survivors being inversely proportional to the exposure time. Thus for a given temperature, the time taken for a 10-fold reduction in survivors, the D value, may be obtained. D values are expressed in minutes or seconds and must be accompanied by the temperature, e.g., D_{72} for 72 °C. While D values are a good approximation for the behavior of bacterial populations, there can be a 'tail' of more heat-resistant strains leading to an overestimate of the lethality of the process. The term D may also be used without a subscript to describe the number of orders that the population may be reduced by in the course of a thermal or other process, e.g. a 10^{12} reduction may be described as a 12D process.

The D value will decrease with increasing temperature. The rate of change is usually given as a z value, the z value being the change in temperature required to give a 10-fold change in the D value. Typical z values for mesophiles are 4–6 °C in high water activity (a_w) systems, though some authorities use a value of 8 °C. The data in Figure 1 for *M. tuberculosis* implies a z value of 6.3 °C. Some care is needed in using z values as they can vary with temperature and, as with the D values, will vary with the product being heated. Capillary tubes have been used for the estimation of thermal death data. While this method may be appropriate for low-viscosity foods, increases in viscosity during the heat treatment may give different results to those from pilot and full-scale treatments.

By using the D and z values it is possible to estimate the risks associated with a temperature–time combination, i.e., the probable level of survivors for a given level of contamination in the raw material. This is easy for a batch process such as with LTLT pasteurization, since the hold is easily measured and the contribution of the heating and cooling stages to the overall lethality of the process is small. With HTST processes, however, the temperatures are higher and the heating and cooling stages may have a significant contribution to the overall lethality of the process. Thus it is essential that the process be characterized in terms of temperature and minimum time.

Minimum time is critical as the microbiological risk (particularly the public health risk) is associated with the minimum heat treatment given to the product. Since most HTST processes are continuous, the flow characteristics of the system must be taken into account. From a microbiological viewpoint, turbulent flow in the pasteurizer will give the best results as there will be a narrower spread of flow rates and hence residence times in the equipment. However, in practice, slower flow rates may be needed to conserve desirable product characteristics or to avoid excessive pressures.

Once the plant has been characterized it is possible to analyze the process quantitatively in relation to a given risk. For instance, safe conditions for milk have been established as a minimum hold of 15 s at 72 °C. If the lethality of this process is given an arbitrary value of one unit, for instance P^*, which uses a z value of 8 °C, then we can calculate the contribution of each stage of the process to the overall lethality. An alternative unit, the pasteurization unit (PU) has also been used; in this case, taking 1 min at 60 °C ($z = 10$ °C) as the standard unit so that a safe process for milk at that temperature would need to be ≈60 PU if 63 °C/30 min is taken as the standard for a safe process. A comparison between PU and P^* values is made in Figure 2. A temperature of 10 °C is appropriate for spores but is higher than normal for vegetative pathogens – appropriate z values should be chosen to meet the substrate pathogen conditions applying to that foodstuff. Changes in the pH or a_w of the foodstuff will alter the D and z values.

It will be seen from Figure 2 that the contribution of temperatures below 65 °C to overall lethality in an HTST process is so small that it may be ignored. However, at higher temperatures the effect of heating and cooling becomes more important. This may be illustrated by comparing two model heat treatments in heat exchangers where the rate of change of temperature is 1 °C s^{-1} (Figure 3).

With the HTST process A, the hold at 72 °C provides the main contribution to the lethality of the process; the heating and cooling stages contribute less than 30% to the total lethality of the process. Raising the final temperature to 80 °C as in B extends both the heating and cooling times and introduces exposure to higher temperatures; the result is that, despite there being no hold, the total heat treatment is well in excess of the minimum safe treatment.

While the primary concern in pasteurization is to obtain a safe food, this is irrelevant if the sensory quality of the food is reduced excessively, either by overcooking or due to the persistence of other less temperature-labile factors (microbiological or biochemical). Cooked flavors may be acceptable in some foods, e.g., clotted cream, but not in others such as wine.

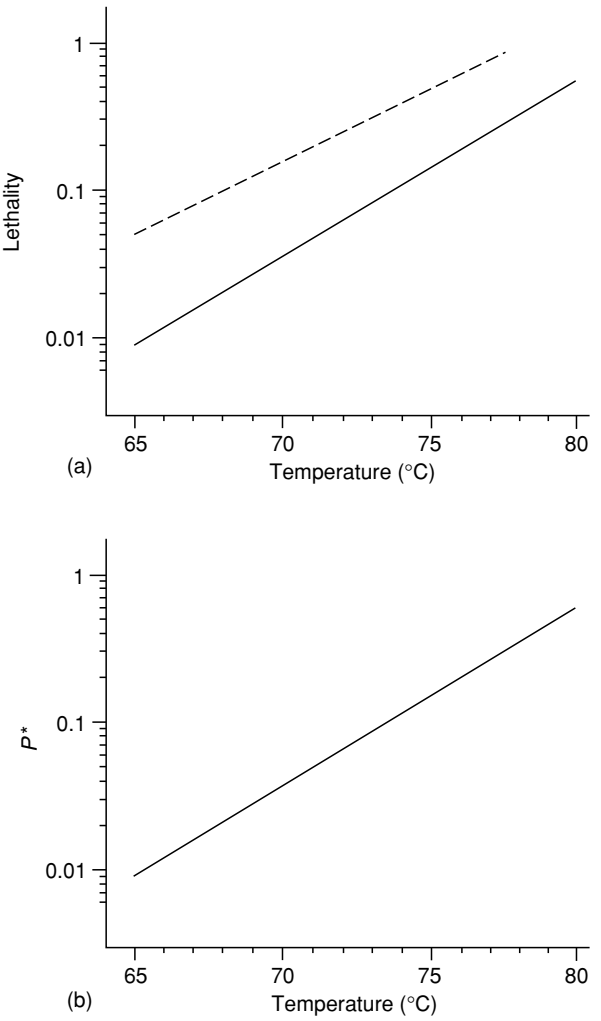

Figure 2 Lethal effect of a 1-s exposure at typical pasteurization temperatures: (a) lethality; (b) P^*. Straight line, P^* ($z = 8$ °C); dashed line, PU ($z = 10$ °C).

Heat treatments may also bring about undesirable changes in the stability or functional properties of food products, often related to protein denaturation. In milk the denaturation of agglutinin (an immunoglobulin fraction) reduces the rate at which cream forms in milk on standing: the denaturation is more pronounced when more severe heat treatments are used so that less cream separates from the milk. This has been an important consideration in processing milk for bottling (Figure 1), where the consumer has associated cream separation with milk quality, but it is not relevant to the production of homogenized milks where cream separation would indicate a processing failure. Similarly, the protein denaturation associated with pasteurization of liquid egg white

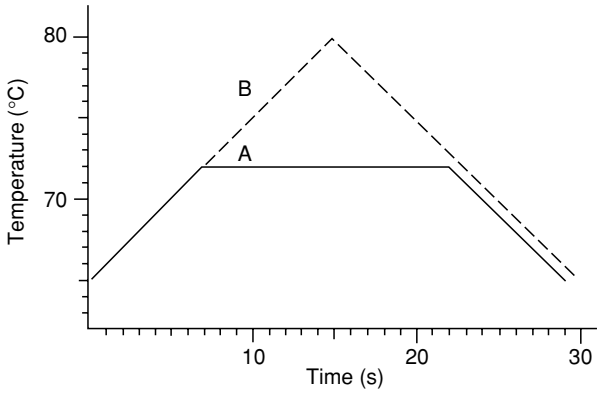

Figure 3 Example of two heat treatment profiles (A and B) with P^* values using the same heating and cooling rates. ($P^* = 1$ for a 15-s exposure at 72 °C with $z = 8$ °C.)

	A	*B*
Heating	0.22	2.4
Hold	1	0
Cooling	0.16	1.8
Total P^*	1.4	4.2

reduces the foaming properties of the product slightly, giving less volume and/or a longer whipping time than with the raw egg white.

Protein denaturation may also be used beneficially to indicate that a satisfactory heat treatment has been given, using an assay of a suitable enzyme occurring naturally in the product. The indicator enzyme should be denatured under conditions slightly more severe than those needed for microbial stability. Ideally the activity of the indicator enzyme should not be subject to wide variation with season or source.

Alkaline phosphatase is denatured under slightly more severe conditions than are required for destruction of *M. tuberculosis*, so the absence of alkaline phosphatase activity is generally used as an indicator for satisfactory pasteurization of milk. The presence of alkaline phosphatase activity may indicate either an insufficient heat treatment or contamination by raw milk. The pasteurization of liquid egg in the UK (minimum 64.4 °C for 2.5 min) is not sufficiently severe to inactivate alkaline phosphatase but will denature α-amylase, whereas the milder treatment required in the USA (minimum 60 °C for 1.75 min) aimed at achieving a 9D reduction in *Salmonella enteritidis*, will leave residual α-amylase activity.

Bacteria are more resistant to heat treatment when the a_w of the medium is lowered. Thus more severe heat treatments are normally used for pasteurization of sweetened products such as icecream and dessert products. Examples of heat treatments are given in Figure 4.

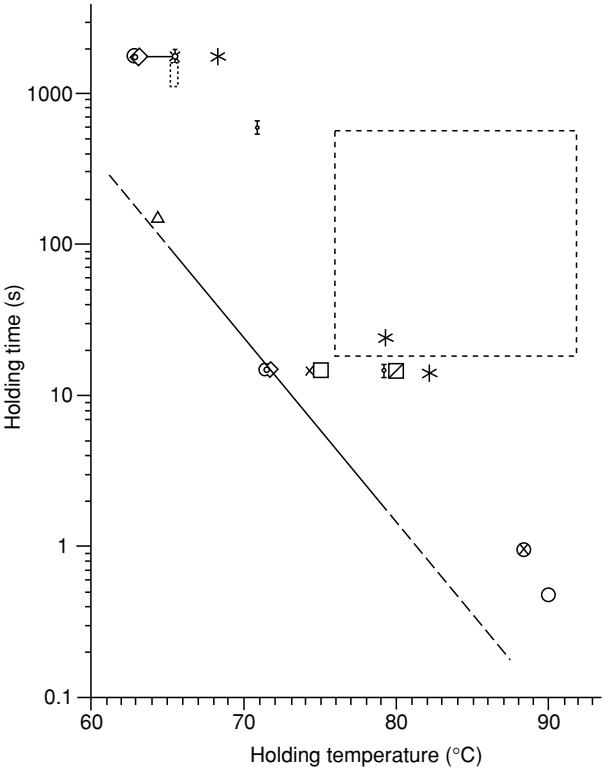

Figure 4 Examples of heat treatments used in pasteurization processes. o Milk (UK); O milk (USA); Δ liquid egg (UK); ◇ cream (UK); □ cream (International Dairy Federation: IDF); ▨ cream (≥35%, IDF); X milk products >10% fat or added sugar (USA); ✳ eggnog and frozen mixes (USA); ⌶ icecream (UK); --- fruit juices; $P^* = 1$ for milk. The dotted boxes indicate the range of values normally associated with low-temperature long-time (LTLT) and high-temperature short-time (HTST) pasteurization processes.

In fruit juices the pH is usually below 4.5 so that growth of pathogenic bacteria will not be supported, though pathogens may survive for some time. Yeasts and some lactobacilli may grow and cause spoilage of the juice, plus molds may grow at the surface. Heat treatments to eliminate yeasts and lactobacilli are more severe than for elimination of vegetative pathogens, e.g., 70 °C/60 s or 85 °C/30 s for citrus juices.

Survival of enzymes can cause problems on storage of fruit juices. In apple juice extraction, polyphenol oxidase will cause rapid browning of cold extracted juices if the juice is not immediately treated with antioxidants such as ascorbic acid or sulfur dioxide. HTST treatment at 89 °C/90 s will denature polyphenol oxidase as well as potential spoilage organisms.

In citrus juices the presence of pectinase will lead to breakdown of the cloud associated with the fresh juices. HTST treatment at 90 °C/10 s or 85 °C/4 min will denature the pectinase. For many fruit juices, including apple and orange, the juice is extracted in the country of origin, heat-treated, and concentrated.

The concentrate is then stored and transported in bulk before reconstitution and final heat treatment.

With any perishable product, the shelf-life is controlled not only by the survivors of the heat treatment but also by the posttreatment contaminants. With mild heat treatments such as the pasteurization of milk the total count may be reduced by two orders so the shelf-life will be a function of the original count, the postpasteurization contamination, and the storage temperature. Fruit juices, though subjected to more severe treatment, will also have their shelf-lives restricted by posttreatment spoilage so that the same juice may be packed as a short shelf-life chilled product or, by aseptic handling and packaging, as a long-life product stable at ambient temperatures. The advantages resulting from avoiding postprocess contamination have led to the retention of in-container pasteurization for many fruit juices and fruit drinks.

The principle of pasteurization has also been applied to the surface treatment of beef and other carcasses following slaughter. Both steam and hot water at 80 °C have been used for reduction of the surface microflora, including *Escherichia coli*. This treatment reduces both the spoilage rate and the potential public health risks associated with raw meat.

See also: **Enzymes**: Functions and Characteristics; Uses in Food Processing; **Fermented Foods**: Origins and Applications; **Heat Transfer Methods**; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; Chemical and Microbiological Changes; Electrical Process Heating; **Lactic Acid Bacteria**; **Milk**: Processing of Liquid Milk; ***Mycobacteria***; **Pasteurization**: Pasteurization of Liquid Products; Pasteurization of Viscous and Particulate Products; Other Pasteurization Processes; **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage

Further Reading

Cunningham FE (1986) Egg-product pasteurisation. In: Stadelman WJ and Cotterill OJ (eds) *Egg Science and Technology*, 3rd edn. Westport: Avi Publishing.

Dahlberg AC (1932) New York Agr. Expt. Sta. Tech. Bull. 203.

Dubos RJ (1960) *Louis Pasteur, Freelance of Science*. New York: Da Capo Press.

Hammer (1928) *Dairy Microbiology*. New York: John Wiley.

IDF (1986) *Bulletin 200: Monograph on Pasteurized Milk*. Brussels: International Dairy Federation.

Kessler HG (1981) *Food Engineering and Dairy Technology*. Freising: Verlag A. Kessler.

Michalski CB, Brackett RE, Hung YC and Ezeike GOI (1999) Use of capillary tubes to validate US Department of Agriculture pasteurization protocols for elimination of *Salmonella enteriditis* from liquid egg products. *Journal of Food Protection* 62: 112–117.

Rees JAG and Bettison J (1991) *Processing and Packaging of Heat Preserved Foods*. Glasgow: Blackie.

Robinson RK (ed.) (2002) *Diary Microbiology Handbook*, 3rd edn, pp. 765 New York: John Wiley.

Wilbey RA (1996) Estimating the degree of heat treatment given to milk. *Journal of the Society of Dairy Technology*. 49(4): 109–112.

Pasteurization of Liquid Products

R A Wilbey, The University of Reading, Reading, UK

Background

Liquid products are relatively easy to pasteurize, since their flow properties permit fast heat transfer by a mixture of convection and conduction. Three basic types of process have evolved to meet the requirements of the food industry: these are batch, in-container, and continuous processes.

Batch Pasteurization

Batch processing represents the simplest approach, where the bulk of the liquid is heated, and possibly cooled, within a vessel with either a jacket or a heating coil. This method was originally used for pasteurization of milk and is still used for small-scale production of icecream and yogurt mixes.

Batch processing suffers from a number of disadvantages. The product flow rate within the tank is relatively slow, and so the heat transfer rates are low, whereas increasing the severity of mixing will increase the power consumption of the plant and may damage the product. The overall heat transfer is also slow because of the relatively low surface:volume ratio, a problem that worsens exponentially with increasing vessel size. Slow heating and cooling rates contribute to a relatively high level of chemical changes in relation to the biocidal effects of the heat treatment, which can limit both the temperature and applicability of the process. Thermal efficiency of batch processing is usually low, as heat is not easily or economically recovered during the cooling process.

Against these disadvantages, batch pasteurization has the advantage of relative simplicity and lower capital cost. Thus, it has been used widely for small-scale processing, particularly where daily outputs are likely to be less than 1000 l per day.

The basic batch pasteurizing setup will usually consist of a jacketed tank with a mixer, temperature sensor and recorder, plus a source of heat and control

system. The mixer should be mounted off-center to avoid excessive cavitation and swirling. A control system is essential to maintain the product at the desired holding temperature for what is usually an extended period, e.g., 30 min at 63 °C for milk or 66 °C for icecream mix. With many small batch pasteurizers, the cooling is carried out within the vessel, and the control system will actuate the changeover from heating to cooling medium.

Apart from the risk of contamination of the product towards the end of the holding period, there are two potential design/operational faults that can occur with batch pasteurizers. Inadequate heating of the head space within the pasteurizer may lead to condensation dripping back into the product with a consequent risk of contamination. The potentially more serious risk is that of cross-contamination from raw materials. The raw material feed must be kept separate from the pasteurized product exit pipe, normally by using a top feed, bottom drain setup, as illustrated in Figure 1. Furthermore, the drain valve must be located close to, or preferably integral with, the vessel, so that all liquid in the drain pipe between the vessel and valve will also receive an adequate heat treatment. The control system and/or the plant operating instructions must insure that the drain valve is closed before any raw materials are filled into the batch pasteurizing vessel.

The different mixing requirements and shear sensitivities have led to a wide range of types of mixers being used in batch pasteurizers. For icecream mix manufacture, an emulsification stage is required, which, on a small scale, can be carried out by using a high-shear mixer. (*See* **Emulsifiers**: Organic Emulsifiers; Uses in Processed Foods), and thus high- and low-shear mixers may be placed into the same vessel.

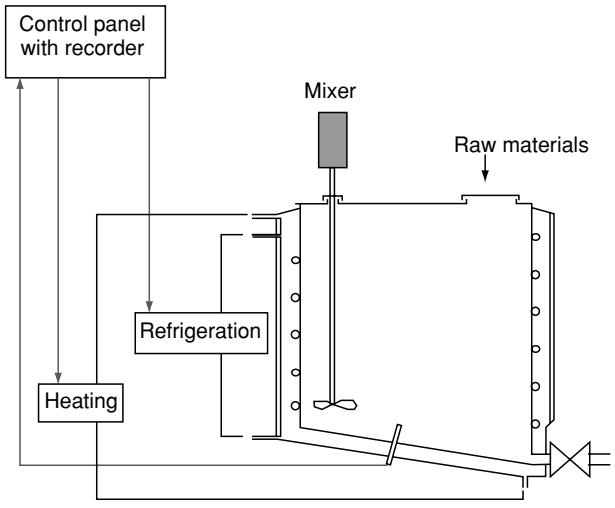

Figure 1 Outline of a simple batch pasteurizer.

Pasteurizer vessels for shear-sensitive products such as cream have used large, slow-moving paddles and rocking coils, though there have been contamination problems resulting from leaks when passing coolant through the paddles.

Where outputs of about 1000 l h^{-1} are required, some processes use two batch pasteurizers in a semicontinuous or flip-flop mode. The vessels are equipped for mixing and heating, together with a temperature sensor, recorder, and control system. In the case of icecream, the emulsification is left to the end of the heat treatment (66 °C for 30 min or 72 °C for 10 min), the heat-treated mix being pumped to a high-pressure homogenizer then cooled by passage through a plate heat exchanger. This plant configuration is outlined in Figure 2. A flip-flop mode of operation is sometimes used for heat treatment of yogurt mix, where holds of 85 °C for 30 min may be applied to the mix (though it is advisable to homogenize the mix before the hold).

Although the semicontinuous mode of operation overcomes some of the objections to batch heat treatment, there remain two serious disadvantages in terms of economics and product quality. It is difficult to recover heat during the cooling phase, thus raising the cost of processing. Product quality can be compromised by the need to empty the tank, after holding, at a rate equivalent to the time taken to prepare, heat, and hold the succeeding batch. This means that the actual hold used will vary from the nominal hold to greater than twice that time when the last part of the batch is removed. Though this excess holding time can be reduced by using three tanks rather than two, there remains a significant average over-processing that can lead to cooked flavors and textural defects in the product. (*See* **Browning**: Nonenzymatic; Enzymatic – Biochemical Aspects; **Protein**: Interactions and Reactions Involved in Food Processing; **Vitamins**: Overview.)

In-container Pasteurization

Batch pasteurization of bulk liquids has a further disadvantage: the pasteurized product must then be dispensed into containers for storage, distribution, and sale. This introduces the likelihood of recontamination of the product and, without the use of aseptic handling systems, limits the shelf-life and requires refrigerated storage to minimize the growth of contaminants. Such postpasteurization contamination (PPC) may be avoided by moving the heat treatment downstream so that pasteurization is carried out in the final container.

The principle of in-container pasteurization is identical to that of batch pasteurization, but in this case, the heat-exchange surface is the container wall,

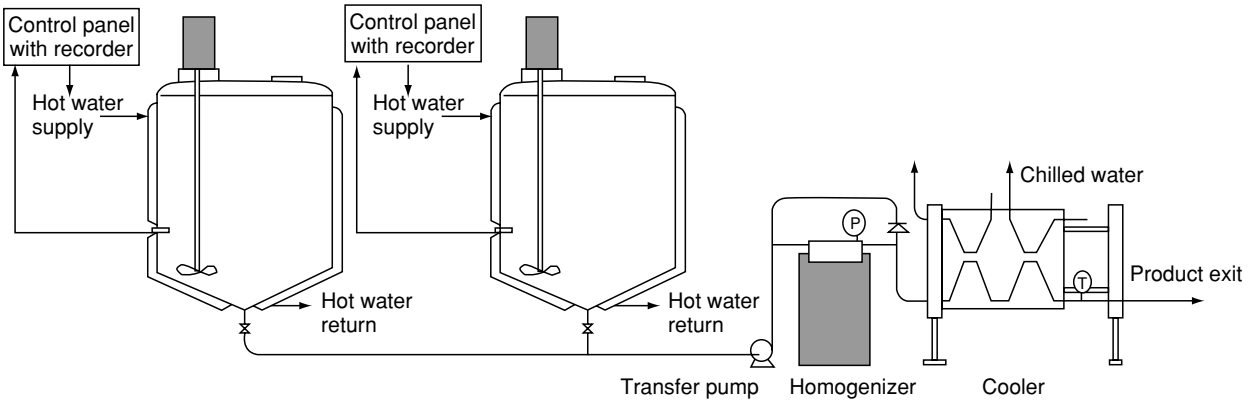

Figure 2 Outline of a semicontinuous pasteurizer.

whether plastic, glass, aluminum, or coated steel. The simplest operations may use a tank of hot water or a steam chamber for heating, followed by immersion in cold water for cooling. Adequate disinfection of the cooling water is essential to avoid contamination, either directly via pinhole leaks or indirectly via contaminated films on the surface of the pack. Unlike bulk batch pasteurization, the product temperature cannot be readily measured during the heat treatment, so test containers with the product must first be set up with temperature sensors and run through the process to establish safe processing conditions. The development of radio telemetry and solid-state logging systems now facilitates continuing checks on product temperatures, but the main control is via the temperature of the heating medium.

The majority of in-container pasteurization operations have moved to continuous processes, based on the conveyance of the containers through a tunnel where the heating and cooling operations may be carried out by spraying water over the containers. The total process time is approximately 1 h, requiring a long tunnel to give the desired holding time and capacity. The width of such tunnels increases with the capacity, as a number of containers will be treated in parallel, so that machines can be up to 25 m long and 7 m wide. Modular construction is essential.

Typical products handled in tunnel pasteurizers are soft and carbonated drinks, fruit juices, beers, sauces, and, occasionally, semisolid products such as jams and puddings. Pasteurization of acid products in this manner can yield products of a very high microbiological quality, where refrigeration is not essential for a long shelf-life.

Tunnel Pasteurizer

The product must first be filled into clean containers and sealed. A scanner or check-weigher is used to guard against under-filling. The stream of containers must then be split into groups corresponding to the number of lanes through the pasteurizer. The conveyer system must be open to permit the heat exchange liquid to drain through and must be made of corrosion resistant material, e.g., stainless steel. A combination of fixed and moving grid may be used to convey the containers on an intermittent rather than continuous basis. The conveyer speed is variable to allow for changes in throughput and holding time.

Heating is carried out by a series of zones in which increasingly hotter water, e.g., 25, 36, and 47 °C, is sprayed over the containers, as illustrated in Figure 3. This incremental heating has two advantages: glass bottles are unlikely to suffer thermal shock, and the incremental system can allow surplus hot water (usually generated by steam injection) to cascade over into the next hottest reservoir, thus reducing the energy loss. The relatively poor heat-transfer characteristics of the containers require an extended heating time, since product flow within the container will be relatively poor, being primarily due to convection currents.

Some energy may be saved by using warm water from the cooling section to preheat the incoming containers, the cooled water then being returned to the cooling part of the tunnel. Process integrity demands that not only water temperature at the various stages be monitored but the flow checked, for instance by monitoring the feed pressure to the spray bars.

Control of In-container Pasteurization

Many tunnel pasteurizers are built for processing a specified product and packaging system and are optimized for that combination. Nevertheless, the performance of the machine must be checked in order for the safety of the process to be proved. This is done by monitoring the center temperature of test containers passing through the pasteurizer under operational conditions. An example of a temperature–time plot

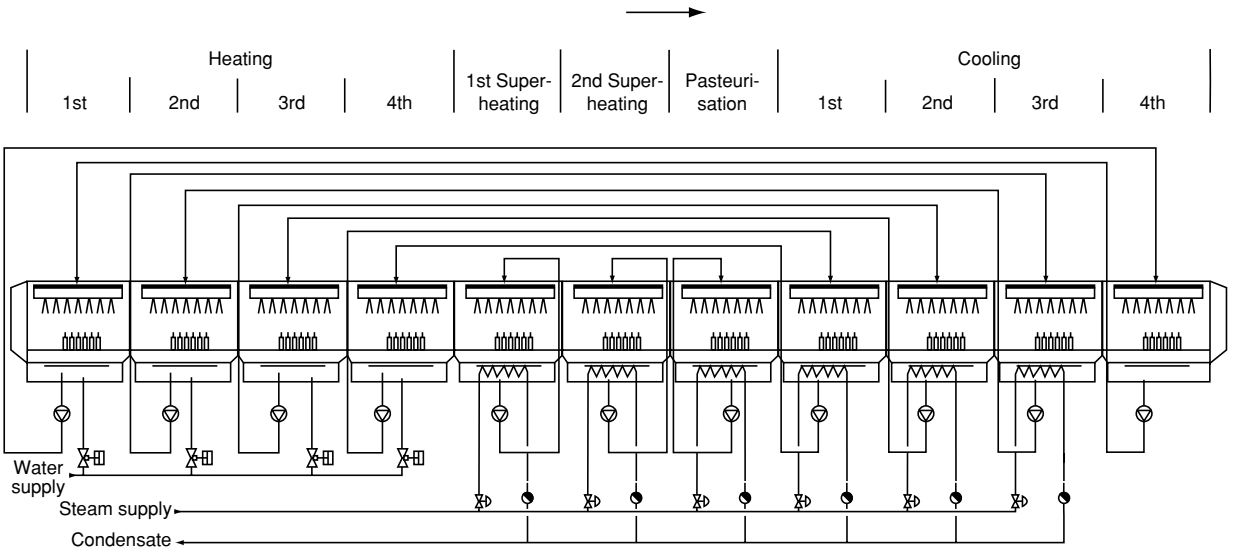

Figure 3 Schematic layout of the water circulation system in the Etna P85 tunnel pasteurizer. (Courtesy of Simonazzi srl.)

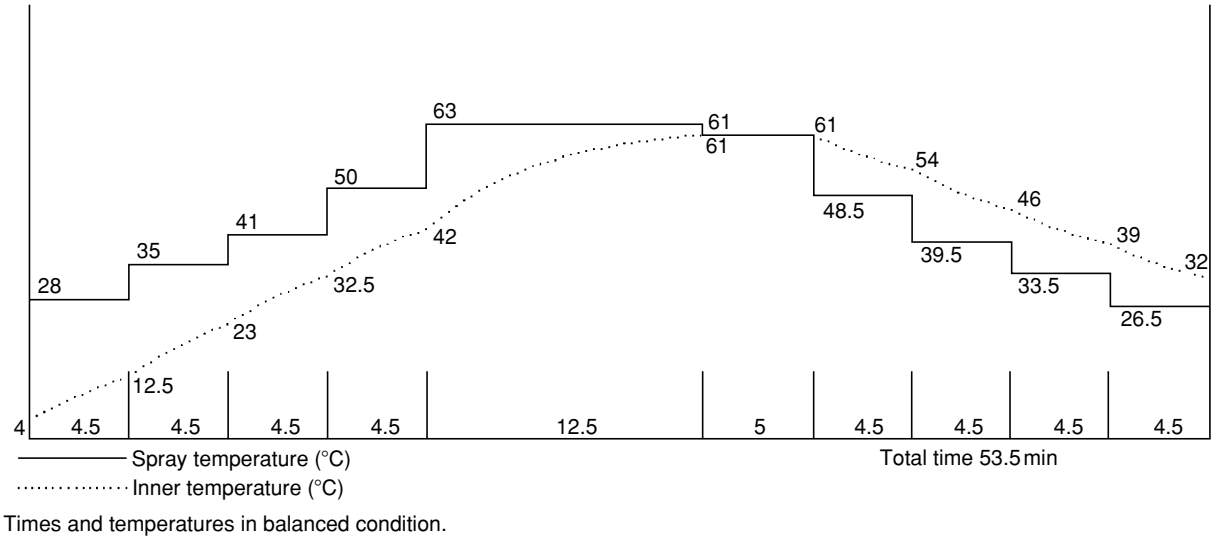

Figure 4 Time–temperature plot for an eight-stage tunnel pasteurizer. Dotted lines indicate water temperatures. (Courtesy of Simonazzi srl.)

for a soft drink is given in Figure 4. The temperature–time data must then be converted into a relevant measure of the lethality of the process, such as the use of pasteurization units (PU). For beers, the PU used is a 1-min exposure at 60 °C, assuming a z value of 7 °C, 20 PU being accepted as a safe process with little effect on flavor, providing the dissolved oxygen has been kept low, preferably below 0.1 p.p.m. (*See* **Pasteurization:** Principles.)

Major problems can occur when the tunnel pasteurizer stops, whether through a fault in the pasteurizer itself or in the production system downstream of it. The risk then is that the containers within the system will be over-processed if no remedial action is taken. Most systems are arranged to shut off the hot water sprays in the event of a stoppage, thus avoiding further heating. Some systems may have extra cooling sprays, the best systems using a computer model of the process to model the actual process during the stoppage, predict when the held product has had sufficient heat treatment, and modify the heat treatment given to the material on start-up, including variation of the conveyer speed.

Continuous Pasteurization Processes

Continuous pasteurization processes take advantage of the shorter exposure times needed at higher

temperatures, and these are often referred to as high-temperature, short-time (HTST) processes. The development of HTST processes has been dependent on the development of hygienic heat exchangers, most notably the plate heat exchanger.

The concept of the plate heat exchanger is brilliantly simple. A pair of plates are separated by a thin elastomer seal, so that the liquid flows as a thin layer, 1.5–5 mm thick, depending upon the seal design, minimizing the distance through which heat must travel. This gives an exceptionally high surface:volume ratio, typically in excess of 500:1 (compared with 5:1 for a 1000-l batch pasteurizer). The plates are rippled, e.g., as shown in Figure 5, so that the flow is broken up, and turbulent flow can be maintained at relatively low velocities over a larger surface area than suggested by the overall dimensions of the plate.

All the plates in the heat exchanger can be pressed with one tool, different heat exchanger configurations being achieved by the presence or absence of the ports cut at the corners of the plate and by the design of the elastomer seal. Plates are now designed so that the seals are held in place without the need for adhesive, speeding up maintenance. The range of flow rates over a plate is limited by the efficiency of heat transfer at lower flow rates, while the pressure drop and possible product damage provides the upper limit. This capacity limitation may be overcome by arranging plates in parallel so that there is an optimal flow across each plate. Similarly, the time permitted

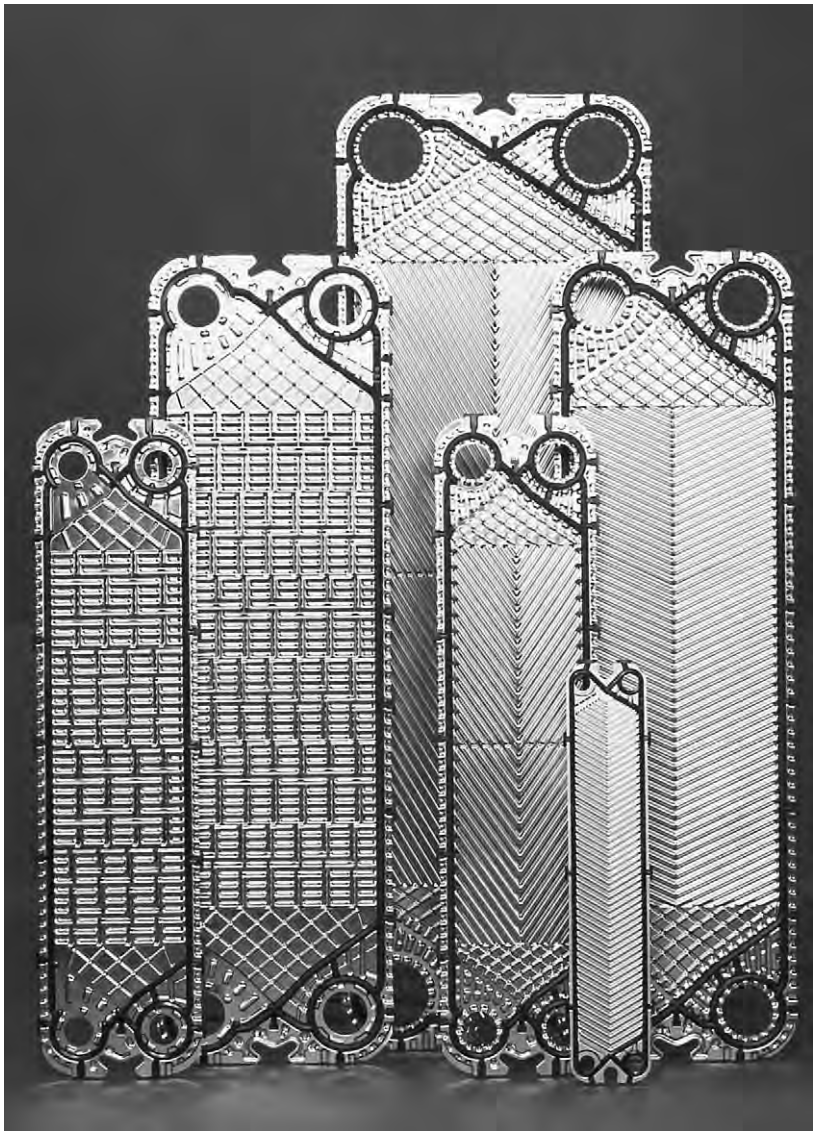

Figure 5 (see color plate 121) Heat exchanger plates. (Photo by courtesy of Tetra Pak Processing UK.)

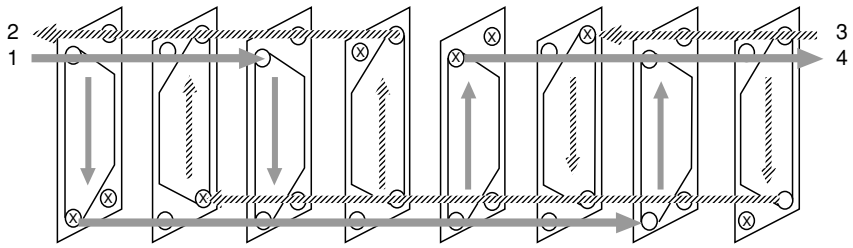

Figure 6 Arrangement of heat exchanger plates in parallel and series. 1, Cold product in; 2, Cool water out; 3, Hot water in; 4, Hot water out; x, blanked-off ports.

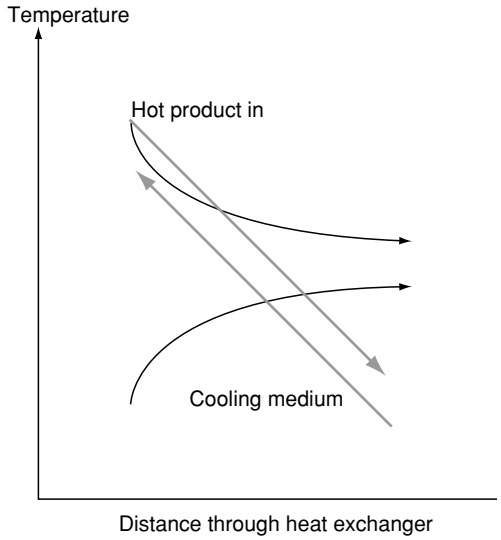

Figure 7 Illustration of co-current (——) and counter-current (——) flow through a heat exchanger using equal flow rates.

for heat exchange can be lengthened incrementally by arranging plates in series, as illustrated in Figure 6.

The heat-exchange medium is normally arranged to flow countercurrent to the product. This reduces the temperature differential between the two media, increasing the overall efficiency of heat transfer, as illustrated in Figure 7. In most cases, the heat-transfer medium will be flowing at a higher rate than the product, but with extremely heat-sensitive products such as fermented milks, similar flow rates must be used to minimize the temperature differential.

With heat-sensitive foods the build-up of a fouling layer will limit the running time of the HTST pasteurizer, owing to both pressure build-up and loss of heat exchange. The nature of the fouling layer will vary with the medium but is most commonly protein, with or without mineral deposition. Fat and carbohydrate components may become associated with fouling layers as they build up. Mineral deposits are common where hard water is used for cleaning and disinfection; additional sequestrant in the cleaning solution or periodic acid cleaning may be required. The running time may be extended by using a wider gap between the plates, though this would require a correspondingly larger heat exchanger to compensate for lower heat-transfer rates. Larger gaps have also been used when liquids have small quantities of suspended solids, e.g., orange juice containing cells that can become lodged within the plate pack and even with back flushing are very difficult to remove. Tubular heat exchangers may be used as an alternative to plates where fouling is a serious problem but thermal efficiency would suffer.

The use of a plate heat exchanger has made it much easier to recover heat by using the hot heat-treated product to heat the incoming untreated liquid. Heat recoveries of up to 95% may be achieved by this 1:1 flow, the recovery being limited by the increasing capital cost of providing further heat exchange capacity as temperature differentials decrease and the hygiene issue of the relatively low flow rates permitting the build-up of a biofilm at temperatures that will permit microbial growth and may eventually cause significant spoilage of the product.

This heat recovery, often referred to as 'regeneration,' may also pose a hygiene risk. This risk arises because, in a simple flow-through system, the pressure of the raw liquid will be greater than that of the heat-treated liquid downstream of it. Thus, any leakage could be from the raw into the pasteurized product and could have serious public health implications. The design of heat exchanger plates seeks to avoid this problem by insuring that the two liquids are separated by either the heat exchanger plate or by two seals, one for each product with a vent to the atmosphere between. Partial failure of one seal is not a major problem as the product would leak on to the floor, but the development of a crack or pinhole crevice could lead to cross-contamination. Most heat exchanger plates are made from a corrosion and stress crack-resistant stainless alloy (e.g., AISI 316), but failure can eventually occur, and checks are needed. The risk can be avoided by using a second pump to drive the downstream part of the process at a higher pressure, raising the pressure in the downstream

regeneration section either through the design of the final cooling section or by incorporating a back pressure valve into the heat exchanger discharge, when this will not damage the product. An alternative system now available is to use double plates; these plate pairs have no internal gaskets other than around the ports, so that any leakage, whether from a port or through a crack in the plate, would leak out of the heat exchanger stack and become obvious. The introduction of the air gap between the pairs reduces the heat transfer rate and would need more plates to achieve the same regeneration efficiency. It must be remembered that the savings from the use of regeneration are twofold; the first saving is from a reduction of heat input, and the second, often greater, saving comes from reduction in the refrigeration requirement.

Figure 8 illustrates a typical arrangement for the pasteurization of liquid milk. Raw milk should be taken from well-mixed bulk storage, *e.g.*, a 100 000–200 000 l silo tank, and fed to a float balance tank, which provides a constant head to the pasteurizer feed pump. Centrifugal pumps are frequently used to feed pasteurizers because of their relatively low cost, but their output is dependent on the pressure that they operate against. In the case of beer, the pump arrangement must allow an over-pressure to maintain the carbon dioxide in solution throughout the process, giving a total working pressure in excess of 8 bar. Shear-sensitive products such as cream require the use of a gentler pump such as a lobe pump, which also has the advantage of a constant output against moderate variations in pressure. Where a centrifugal pump is used, it must be accompanied either by a flow controller or by a second constant-output pump to guarantee a fixed flow rate. This requirement for a constant flow rate is set by the need to guarantee a constant residence time in the holding tube at the completion of heating. The heating is normally divided into two sections, the first preheating section using regeneration and the second heated by hot water or steam.

Sometimes, the preheating section is partially bypassed by a length of pipe with a fine control valve. This may be used to reduce the thermal efficiency during cleaning, but the main use is to control the final temperature of products such as milk for cheesemaking, where the milk will be cooled to incubation temperature (e.g., 30 °C) without the need for final cooling and without upsetting the pasteurization temperature (typically 72–73 °C). A small flow of cold raw milk past the preheater can lower the regeneration efficiency and hence raise the final milk temperature. This implies that the limit for thermal

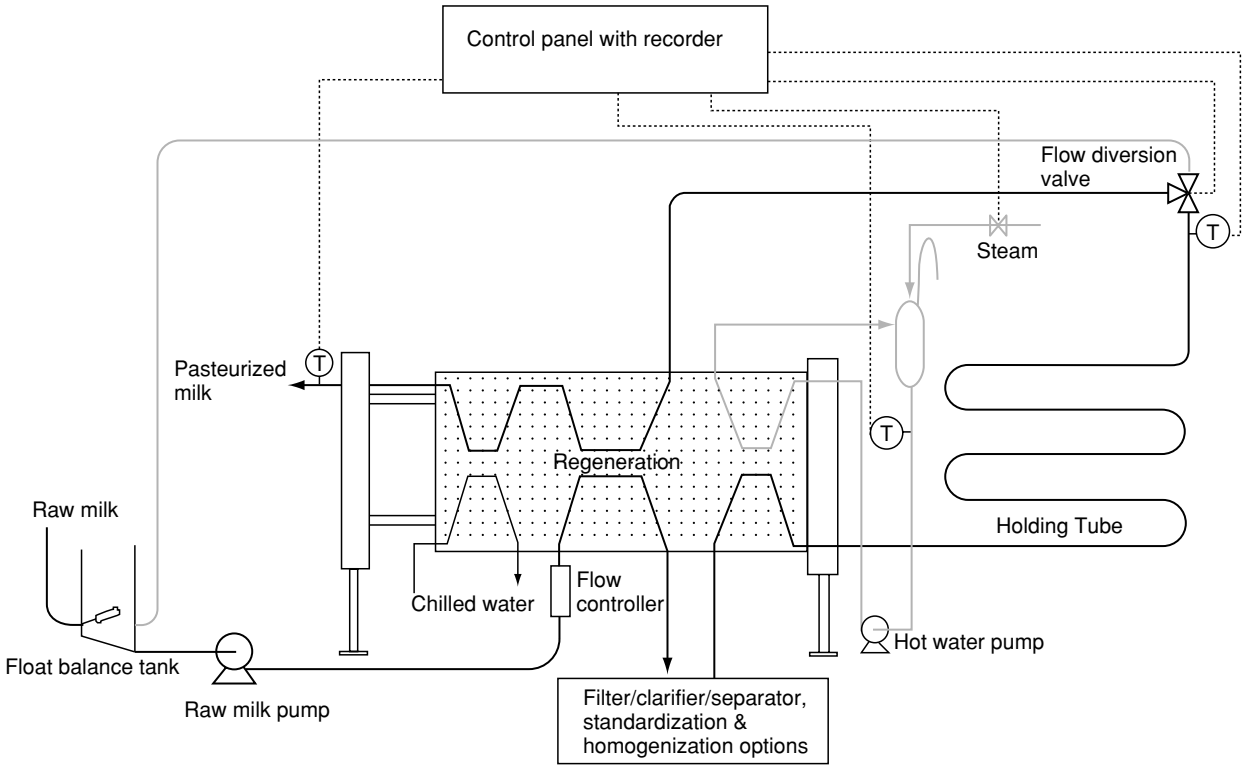

Figure 8 Schematic diagram of a milk pasteurizer.

efficiency is ≈63% when pasteurizing milk for cheesemaking.

In the simpler plants, preheated milk may be filtered before being subjected to final heating, etc. Many modern plants, however, have replaced filtration by centrifugation, sometimes simply for clarification but more often to separate the cream. The cream may be fed back into the skimmed milk to produce standardized milks for subsequent processing. Standardization may be carried out as a continuous process, the milk then passing to a high-pressure homogenizer that will provide a constant output irrespective of back-pressure (thus requiring a pressure-relief valve). The temperature provided by preheating is usually sufficient for homogenization of milk or cream, and incorporation of homogenization at this point avoids contamination from the homogenizer. Higher temperatures (≈65–80 °C) are desirable for icecream emulsification, so the homogenizer may then be placed at the beginning of the holding tube.

Hot water is the most common heating medium, energy being provided by steam injection and the condensate being returned to the boilers. Water temperature is controlled by sensing the water temperature, the controller generating a signal to open the pneumatically operated steam control valve when the temperature drops below the preset level. Some small pasteurizers, e.g., $\leq 200\,l\,h^{-1}$, may use electrical heating only. A few larger pasteurizers may use low-pressure steam as the heating medium, though this is less easy to control and is best avoided for temperature-sensitive products.

Most pasteurizer designs require the product to be held for a fixed period of time in order to achieve the desired lethality. The most reliable method of achieving this is to use a holding tube, the design of which is critical. Any bacteria present in the liquid to be pasteurized can be treated as solid particles in suspension, and the rate at which they progress along the holding tube will depend upon the flow pattern that is induced within it. The flow pattern may be deduced by estimating the Reynolds Number (Re) for that liquid under those process conditions, i.e.,

$$Re = v\,\rho\,d/\mu,$$

where v is the velocity, ρ is the specific gravity at the holding temperature, d is the pipe diameter, and μ is the dynamic viscosity at the holding temperature. When $Re \leq 2000$ for flow in tubes, the flow is said to exhibit streamline or laminar flow, where there is a large variation in velocity across the tube, with the velocity in the center being twice the average velocity, i.e., the minimum time taken for a particle to pass along the tube would be half that of the average, as illustrated in Figure 9. Values of 2000–4000 are said to characterize critical or intermediate flow patterns, where the geometry of the tube, e.g., the incidence of bends, will affect the flow pattern. At $Re \geq 4000$, the flow in a tube will exhibit turbulent flow, where the distribution of velocities is much narrower, and the minimum residence time may be as high as 0.83 that of the mean. Turbulent flow is designed into most holding tubes, but the holding efficiency is assumed to be 0.75. The advantages of turbulent flow in holding tubes are tighter control of the lethality of the process and less chemical damage to the product for that lethality.

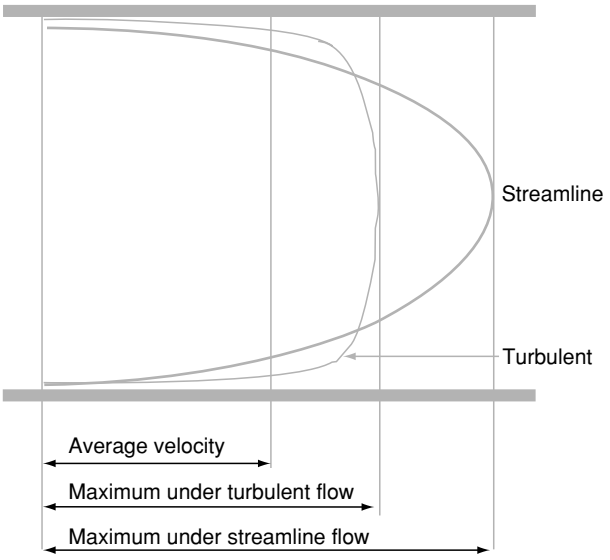

Figure 9 Illustration of the relationships between flow velocities in a tube under turbulent and laminar flow regimes.

The second issue for holding tubes is heat loss. Heat is readily lost even from polished stainless surfaces, particularly with smaller plants where the surface:volume ratio is higher. Holding tubes must never be situated in draughty locations such as close to doors, since sudden draughts can lead to a drop in holding temperature that will cause the process to fail. Where cold and variable ambient conditions are expected, the holding tubes must be insulated, and if a long hold is being used, e.g., holding whole liquid egg at 65 °C for 150 s in the UK or 60 °C for 210 s in the USA, auxiliary heating of a box around the holding tube should be considered.

The temperature of the hot product is normally sensed by a fast responding probe close to the end of the holding tube. The signal generated is used to both provide a permanent record of the heat treatment and to generate a signal to the flow diversion valve. Should the temperature in the holding tube fall outside a preset range (e.g., 72–78 °C for milk) or a failure in any of the services (steam, air, power), the

valve must move into the divert position and the heated product be sent back to the balance tank. Only when the temperature is in the desired range should the valve permit product to flow forward into the downstream, cooling sections of the plant. This action protects the downstream part plant from contamination by under-processed material and hence avoids risk to the consumer. The recorder fitted to the pasteurizer must record the pasteurization temperature, final product temperature, and status of the flow diversion valve throughout the operation.

In most large-throughput plate heat exchangers, the bulk of the cooling will be achieved by regeneration, with final cooling by chilled water. Chilled water may be produced directly by refrigeration or indirectly via an ice bank. Where cooling by glycol or brine solutions is employed, more stringent controls are needed to avoid freezing. In all cases, the coolant must be of good microbiological quality and product contamination avoided. With rheologically simple liquids such as beer and milk, the cooling would use countercurrent flow to achieve the optimal exit temperature. However, with more complex liquids such as cream, where the viscosity is dependent on many factors, including the time–temperature–shear profile during cooling, a mixture of co- and counter-current cooling may be employed. For the thickest creams, there is only partial cooling in the heat exchanger, with the remaining cooling being achieved after filling, in the retail container. This approach demands the highest hygiene standards during filling and effective cooling of the containers.

Tubular Heat Exchangers

Although plate heat exchangers are excellent machines for pasteurization of low-viscosity liquids, there can be problems with those liquids either displaying higher viscosities, a tendency to foul at modest temperatures or containing suspended particles. The introduction of tubular heat exchangers enables a greater path width, e.g., 5–10 mm, so that fouling becomes less critical and particles less likely to become trapped.

In the case of fruit juices containing pulp, simple tubular heat exchangers provide an effective means of pasteurization though with higher capital and running costs. This may be attributed to lower surface:volume ratios than are possible with plate heat exchangers and to problems in heat recovery. The surface:volume ratio can be increased by adopting an annular design, and heat exchange may be improved by adopting a rippled wall profile to promote turbulence. Detail improvements in the design of tubular heat exchangers have reduced the cleaning problems, particularly in permitting product–product heat exchange, and so direct heat regeneration has become possible. Control systems are similar to those for plate heat exchangers.

Although most fruit juices have pH values below 4.5 and are not considered high risk, pear juice, banana purée, and tomato juice may be low acid. Regardless of the public health risk, heat treatment is needed if rapid spoilage is to be avoided. Pasteurization at over 70 °C for 15 s should inactivate vegetative spoilage organisms such as yeasts, mold mycelia, and *Lactobacillus fermentum* in a high acid juice. More severe conditions such as 87 °C for 15 s would be needed for inactivation of spoilage enzymes, and for inactivation of *Byssochlamis* spp. ascospores, temperatures in excess of 100 °C would be needed. Tomato juice has been processed at 115 °C for 15 s.

Concluding Comments

Low-acid products must be kept refrigerated after pasteurization and have a relatively short shelf-life. Postpasteurization contamination poses a major threat to product quality and can only be avoided by using either in-container treatments or aseptic filling techniques. The advantages and disadvantages of the pasteurization processes are summarized in Table 1.

Table 1 Summary of the main advantages and disadvantages of pasteurization systems

Advantages	*Disadvantages*
Batch tank	*Batch tank*
Low capital cost	High energy cost
Low maintenance cost	Poor cooling characteristics
Very flexible	PPC risk depends on downstream hygiene
Suitable for small quantities	
In-container tunnel	*In-container tunnel*
Avoids PPC	High capital cost
High capacity possible	High space requirement
Consistent lethality	Product containers must be heat-resistant
More efficient than batch	Less efficient than HTST
Low cleaning requirement	Higher maintenance than others
Plate heat exchanger	*Plate heat exchanger*
Moderate capital cost	Viscosity limit
Low maintenance cost	Particulates limit
Low space requirement	Shear can damage product
High energy efficiency	PPC risk depends on downstream hygiene
Consistent lethality	Fouling problems with temperature-sensitive products

See also: **Browning**: Nonenzymatic; Enzymatic – Biochemical Aspects; **Emulsifiers**: Organic Emulsifiers; Uses in Processed Foods; **Homogenization**; **Pasteurization**: Principles; **Vitamins**: Overview

Further Reading

Cunningham FE (1995) Egg product pasteurization. In: Stadelman WJ and Cotterill OJ (eds) *Egg Science and Technology*, 4th edn. Binghampton: Haworth Publishing.

Hyde A (1999) Heat transfer processes in breweries – part 3. *Ferment* 12: 30–34.

Lewis MJ and Heppell NJ (2000) *Continuous thermal processing of foods: pasteurization and UHT sterilization*. Gaithersburg: Aspen.

Zufall C and Wackerbauer K (2000) The biological impact of flash pasteurization over a wide temperature interval. *Journal of the Institute of Brewing* 106: 163–167.

Pasteurization of Viscous and Particulate Products

G Tucker, Campden & Chorleywood Food Research Association, Chipping Campden, UK

Background

For food products of high viscosity, the choice to pasteurize in a continuous system has distinct advantages over one where the process takes place in the container (see **Figure 1**). This is because the heat transfer from a heating or cooling medium to a viscous food is far from ideal, but careful design and selection of the heat exchange system can overcome this limitation. Of principal concern are the economics of food production, which become attractive in continuous systems for high volumes and long production runs. Further economic benefits can be realized by minimizing the product's exposure to the adverse effects of high temperatures, long processing times and high shear preparation methods, resulting in improved product quality, reduced processing costs, increased safety, and increased plant throughput. This chapter considers both continuous and in-container processing systems.

Continuous Processing Considerations

A challenge in the design of heat-transfer equipment are the so-called prepared food products, such as tomato products, soups and sauces, and dessert products. These products are normally of high viscosity as well as of complex composition. Also, in most cases, the particle content is significant. With respect to the rheological or flow behavior, the products are typically non-Newtonian, showing, in many cases, quite extraordinary behavior.

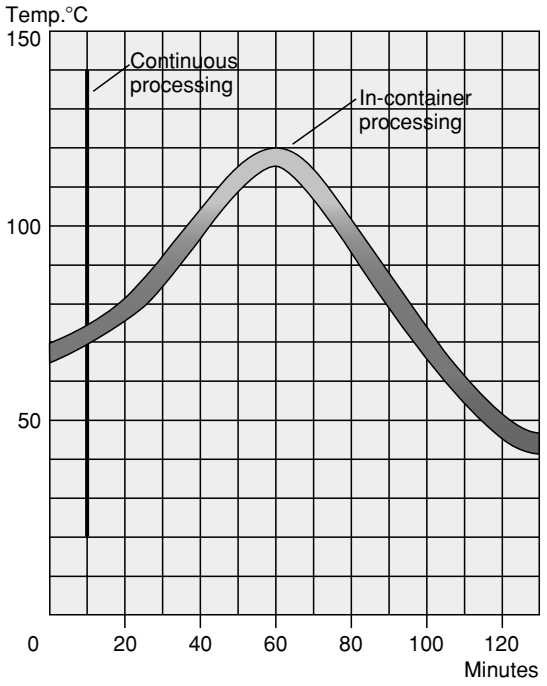

Figure 1 Time/temperature graph for comparison of in-container and continuous processes.

The microbiological demands of a pasteurization process are still to achieve commercial sterility, i.e., the end product must be free from pathogenic and spoilage organisms capable of growth under normal storage and distribution conditions as well as being free from toxins. Depending on the rheological properties of the product and the possible presence of particles, the design and choice of equipment can vary significantly from case to case.

Flow Behavior

In the design and choice of heat exchangers, the flow behavior of the product has to be taken into consideration. The flow behavior will affect, for instance, the residence time distribution and hence the design of heat exchangers and holding cells to obtain the sufficient thermal treatment. The basic difference between laminar (streamline) and turbulent flow is well known (see **Figure 2**), as is the effect on the velocity profile from heating or cooling of the product. For example, the maximum velocity in laminar flow is theoretically twice the mean velocity, and in turbulent

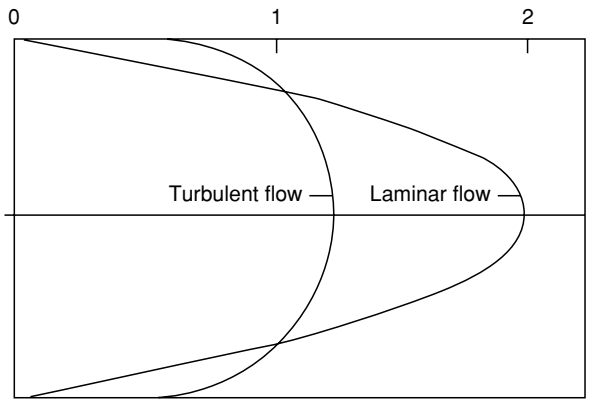

Figure 2 Velocity profiles for turbulent and laminar (streamline) flow showing the ratio of maximum velocity to average velocity.

flow around 10% higher. For viscous products, however, the flow conditions are nearly always laminar.

For liquid foods that have a complex flow behavior the velocity profiles in pipe-flow depend on the flow behaviour index. This index can range from 1.0 in simple materials such as water, milk, or fruit juice, to 0.7 in starch-based sauces, to 0.3 in tomato paste, and in theory can go to 0.0. With decreasing flow behavior index, the velocity profile increases in flatness, which means in practice that the maximum velocity decreases from twice the mean velocity. Maintaining the factor of two for the calculation of the necessary holding cell length thus creates overcooking of the product, but it is a safe assumption and one that most of the industry adopts.

A further complication to the viscous flow behavior arises with additives that give elastic properties, e.g., xanthan or gellan gum. These are sometimes used to enhance the particle-carrying properties for a carrying fluid in continuous processing. The so-called *yield value*, which normally is a measure of the product's willingness to flow by itself, e.g., from a storage tank, is also a measure of the particle carrying abilities. A significant yield value, typical of paste-like products, also adds to the flatness of the velocity profile.

Choice of Heat Exchanger

The choice of optimal heat exchanger depends to a great deal on the flow conditions. Fluids with low viscosities and no particles are preferably treated in a plate heat exchanger (see Figure 3). This is the most economic option. The food product and heating (or cooling) media flow in alternate channels to provide good heat-transfer characteristics. The plates are 0.5–1.25 mm thick, separated by 3–6 mm, and sealed with gaskets.

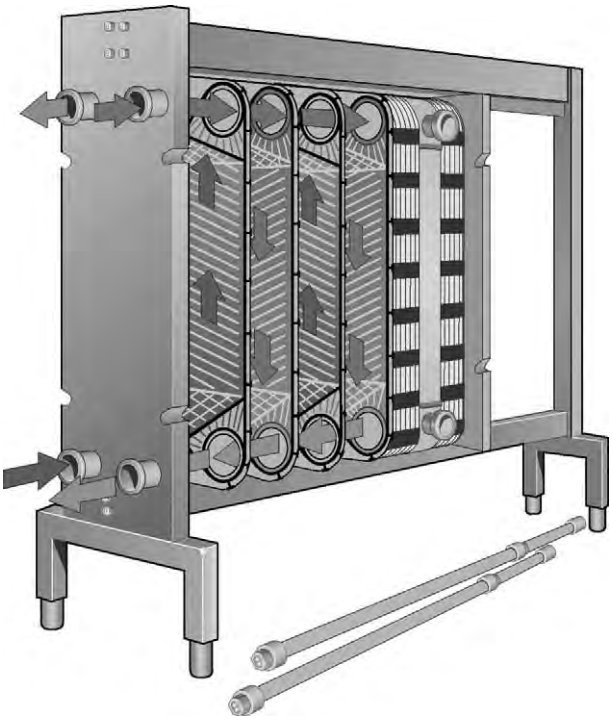

Figure 3 Plate heat exchanger.

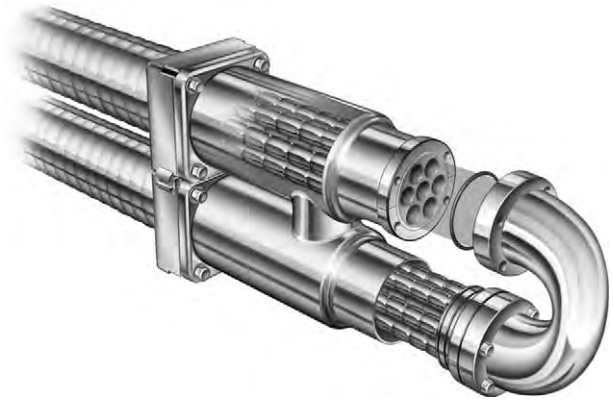

Figure 4 Multitube tubular heat exchanger.

For fruit juices with pulp and fibers up to 5 mm in length, special types of plates are available with wide gaps. Even with foods of higher viscosities, the plate heat exchanger can be utilized as long as the pressures developed are not too high, which can cause failure of the gaskets. Attempts to pasteurize foods with particles can result in blocked flow channels and increased pressure until the gaskets fail.

For fruit juices with fibers up to 15 mm in length, a multitube tubular heat exchanger (see Figure 4) is preferable to the plate heat exchanger. Also, foods of moderate to high viscosity with only small particles (< 5 mm) will flow through a multitube exchanger

without problems. The typical inside diameter of each tube is 14 mm, and these can be grouped in bundles of four, seven, 12, 19, 27, or 37 tubes. Other combinations of tube diameters and tube bundle numbers are available, depending on the equipment supplier. Most commercial tubular heat exchangers are 6 m long.

However, if the fluid is significantly viscoelastic, i.e., exhibits a large yield value, often in combination with a high viscosity, there is a risk of maldistribution across the tubes of a multitube exchanger. In the worst case, the product flow will stop in some of the tubes, causing overcooking of parts of the product and subsequent cleaning problems. Examples of such products are hot break tomato paste or a stiff dessert pudding. In these cases, the concentric tube is the best choice, which has only one product channel and eliminates this risk. At the same time, the narrow gap and the two service medium channels surrounding the product channel provide efficient heat transfer (see Figure 5).

Finally, if large particles are present, the monotube is the optimal choice of tubular heat exchanger (see Figure 6). The drawback with a monotube compared to multi- or concentric tubes is a reduced thermal efficiency due to the thicker product layer. This is illustrated in Figure 7, which shows the cross-sectional flow areas of each tube type. However, to a great extent, the particles present in the product will work as 'internal mixers' and hence will promote heat transfer.

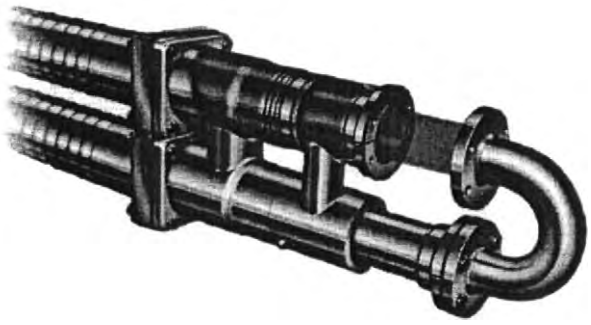

Figure 5 Concentric tubular heat exchanger.

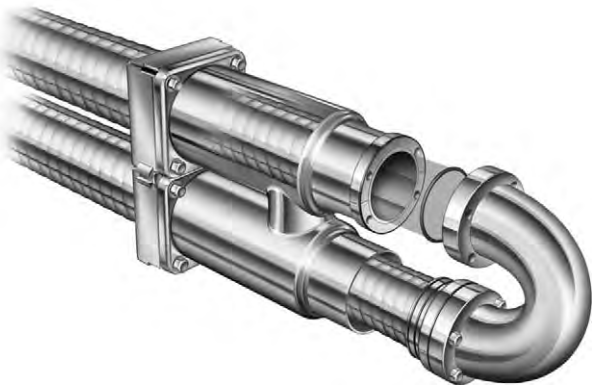

Figure 6 Monotube tubular heat exchanger.

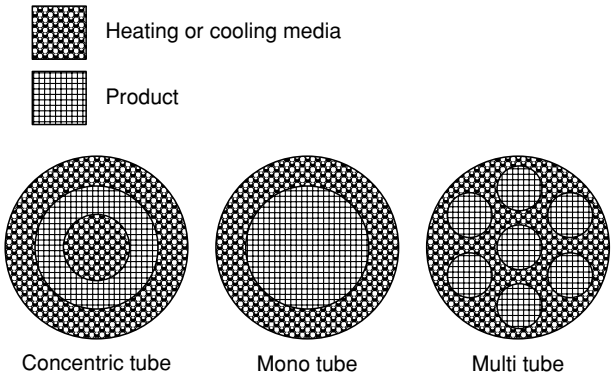

Figure 7 Cross-sectional tube designs for tubular heat exchangers.

If the food viscosity is high, and there is the chance of product fouling the heated or cooled surfaces of a tubular heat exchanger, the scraped surface heat exchanger must be employed (see Figure 8). In principle, a scraped surface heat exchanger is a monotube equipped with a rotating internal scraper. The scraper keeps the heating surface free from any deposits and also promotes turbulence. Hence, this type of heat exchanger is ideal for products of a very high viscosity, possibly also containing large particles.

Heat-transfer Correlations

Design models for heat exchangers are normally based on empirical correlations of the dimensionless Nusselt, Prandtl, and Reynolds numbers. The derived equation is basically of the form $Nu = f(Re, Pr)$. By using dimensionless numbers, only a limited number of experiments have to be performed in which the product and heating/cooling medium flow rates and the product physical properties are varied in order to cover a large range of Reynolds as well as Prandtl numbers. The physical properties are changed preferably by changing the temperatures of the fluids involved. Decreased and increased temperatures can normally vary the viscosity of the product significantly.

The definition of the Nusselt (Nu), Prandtl (Pr) and Reynolds (Re) numbers is as follows:

$$Nu = \frac{\alpha d_h}{\lambda}$$

$$Pr = \frac{c_p \mu}{\lambda}$$

$$Re = \frac{d_h \rho}{\mu}$$

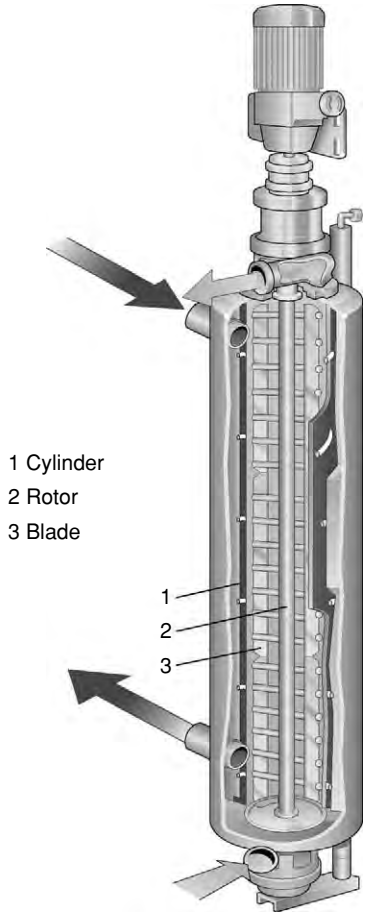

Figure 8 Scraped surface heat exchanger.

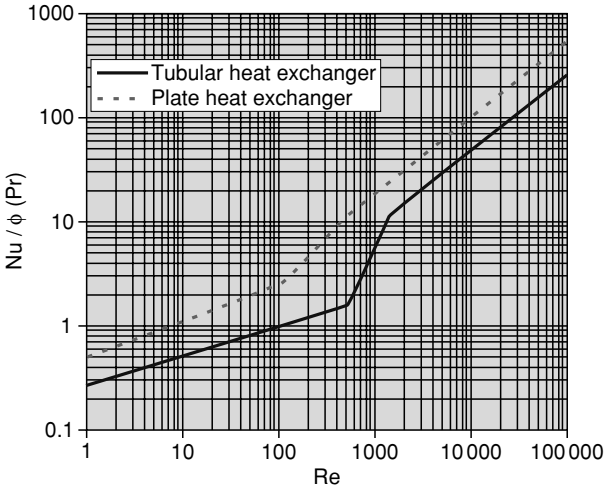

Figure 9 Comparison of *Nu–Pr–Re* correlations for plate and tubular heat exchangers.

where α is the individual heat transfer coefficient (W/m^2K^{-1}); d_h is the hydraulic diameter (m); λ is the thermal conductivity of liquid ($W/m^{-1}.K^{-1}$); c_p is the specific heat of liquid ($J/kg^{-1}.K^{-1}$); μ is the dynamic viscosity of liquid (Pa.s); v is the velocity ($m.s^{-1}$) and ρ is the density ($kg.m^{-3}$). A comparison of tubular heat exchangers with plate heat exchangers shows that the thermal performance, as read off in a typical *Nu–Pr–Re* graph, is better for the plate heat exchanger, mainly due to the complex corrugated pattern including a large number of contact points in the product channel (see Figure 9).

From systematic laboratory tests and from experience of commercial plants, it has been found, however, that to completely model the heat transfer of liquid food products, several more parameters have to be included. Such parameters include, for instance, particle content, particle shape, and size as well as the type of particles. In addition, the softness or hardness of the particles can effect the heat transfer by their influence on the laminar boundary layers at the exchanger surfaces.

Particulate Processes

Flow rates of particulates and carrier liquids should be balanced to give even thermal processing of the product. This relies on controlling the density of both phases to prevent sedimentation or flotation of the particulates. The limiting step will be the heat transfer to the center of the fastest and largest of the particulates, which should be determined by experimentation in the measurement of residence time in the holding tube and of thermal conductivity. Once determined, an estimation of the pasteurization level can be obtained using a mathematical procedure to solve conduction-based heat transfer equations. Before production of the food product commences, a full microbiological validation of the process safety should be conducted.

Batch Processing Considerations

Batch systems fall into two main categories, the most common being where the viscous food is pasteurized in its container, and the alternative is to carry out the processing within a large container (e.g. a closed vessel) and fill into individual containers. The advantage of the former method is that both food and container are pasteurized during the process, which allows the integral package to be commercially sterile and the potential shelf life extended. The disadvantage is that the containers need to be strong enough to withstand high temperatures (80–100°C) and the swings in pressure differential.

Irrespective of the method used, heat transfer from the media to the product thermal center is poor, because of the viscous boundary layers that develop on

the inside surface of the containers. The effect is to form an insulating layer that impairs heat transfer if there is no mixing inside the containers. This results in lengthy processing times to achieve pasteurization and so few companies will process viscous foods without agitation.

In-container Systems

The traditional package is the metal can processed in a steam atmosphere, although glass jars, pouches and flexible plastic or aluminum trays can now be successfully processed in steam or hot water. These containers require an air overpressure to counteract the natural expansion of the gases present in the headspace and those released from the food as it rises in temperature. The desired effect is to push the lid or sides back to their original position and therefore minimize the stress on the seals. Examples of viscous foods pasteurized in their containers are fruit pie fillings, tomato salsas and cook-in-sauces.

The first equipment (introduced by FMC Corporation in the 1930s) to increase heat transfer to this type of food was the reel & spiral cooker-cooler. This imposes rotation of cans around their central axis and allows heating times to be reduced substantially (e.g. from 90 minutes to 15 minutes). Cans travel on a helical reel through the steam heating and water cooling sections, and in doing so rotate at high speeds. The continuous operation of the cooker-cooler allows a high throughput to be achieved and favors high volume production.

More recently, end-over-end container rotation has become popular, where the containers are constrained in baskets that rotate. Being a batch system, this offers greater production flexibility and is not restricted to cylindrical metal cans. However, heat transfer enhancement is not as great as with the cooker-cooler and so the throughput is lower and operating costs higher. Sauce products in plastic pouches are processed in steam and air mixtures, using low rotation speeds (2 to 10 rpm) to induce mixing without damaging the delicate pouch.

In-vessel Systems

The vessel acts in a similar way to a heat exchanger in that it raises the food temperature to that required for pasteurization. A typical vessel size is around 800 to 1 000 kg, and comprises a hemispherical steam-jacketed base with cylindrical sides above. A hinged lid is usually present to reduce heat loss and prevent foreign objects falling into the food. With viscous foods it is essential that the food is well-mixed, otherwise laminar boundary layers develop and the food burns on to the heated surface. Horizontal agitators with scraped surface blades offer the most effective mixing, although recirculating pumps and vertical mixing blades are alternatives. Examples of viscous foods pasteurized this way are fruit preparations, confectionery, cook-in-sauces and tomato products.

Once pasteurized, the food can be filled either hot or cold into the containers. A hot-fill process will only require a short hold time at high temperature to ensure the inside container surfaces are pasteurized. This is usually achieved in a raining water tunnel pasteurizer, although it is possible to omit this step if the food's acidity is high (pH < 3.8), the filling temperature above 95°C and the containers pre-warmed or of low heat capacity. The shelf life of a viscous food of low pH will be many months if hot-filled, and determined by its chemistry. A cold-fill process will not guarantee commercial sterility of the container, and as such requires far greater attention to hygiene in order to minimize the introduction of microbial contamination during filling. Most cold-filled products are sold chilled and have shelf lives up to 14 days.

Heat transfer issues are similar to those for continuous processing in terms of transferring heat from steam or hot water to the core of a viscous food. Boundary layer generation on the inside walls of a food container limit the heating effect, which becomes more detrimental as the food viscosity increases. Agitation of the food container can overcome this resistance to some extent but it is important to match the rotational conditions to the product viscosity, otherwise the benefits will not be realised. Food containers can be rotated either axially in a continuous canning system or end-over-end in a batch retorting system. The continuous system offers greater efficiency for long production runs whereas the batch system offers greater flexibility.

Thermal Process Validation

Although pasteurization of the food is the desired condition, the food is referred to as commercially sterile. A pasteurization process usually operates to 6 log reductions of the target organism (further details on pasteurization treatments can be found in the CCFRA guidelines), and this differs from fully sterilized foods where the intention is to achieve at least 12 log reductions in *Clostridium botulinum* spores. The lower target log reductions for pasteurization are because of the reduced risks associated with the target microbial species when compared with the lethal botulinum toxin. The following equation is used to calculate the process or F-value from heat resistance data on the target pathogenic or spoilage organisms. When the F-value is divided by

the decimal reduction time (D_T) this gives the number of log reductions of surviving spores.

$$F = D_T \ (\log N_0/N)$$

where N_{final} is the final number of organisms after a specific time-temperature history; $N_{initial}$ is the initial number of organisms; D_T is the decimal reduction time at a fixed temperature (T) to reduce the number of organisms by a factor of ten (minutes).

To prove that the pasteurization process has achieved the F-value it will be necessary to conduct validation studies using an approved method. Various methods can be selected, and their choice depends on the costs and on the nature of the food and the process type. Temperature measurements provide the cheapest method but are not appropriate for all foods, and so microbiological methods are required as alternatives.

In a validation using temperature measurements, the heat resistance data for the target organism is used to convert the measurements to log reductions. For a viscous food, the position within a container that heats slowest is usually the geometric center and so the probes are placed at the centers of several containers. To validate a process for a food containing particulates, a large food chunk or piece is usually attached to the end of a temperature probe. While this will work for incontainer processes, this method cannot be used for continuous processes.

Microbiological Methods for Process Validation

These are often referred to as direct methods, but they in fact rely on measuring the achieved log reductions for a process using a non-pathogenic organism and converting this to an F-value using the same equation. If there are no surviving organisms then it is only possible to conclude that the process achieved greater than e.g. 6 log reductions for a 10^6 initial loading. In this situation, there will be uncertainty as to whether the organisms died as a result of the process, during transportation to or from the factory, or if the spores germinated during the come-up time making them more susceptible to destruction at milder temperatures than for the heat resistant spores. Hence, controlling how these tests are performed is critical and the expertise to conduct a test using encapsulated spores or organisms tends to be restricted to a limited number of microbiology laboratories. A microbiological method can be conducted using organisms distributed evenly throughout a food product or concentrated in small beads.

The inoculated pack method is also known as the count reduction method and involves inoculating the entire food with organisms of known heat resistance. For ease of handling, the organisms are usually in the spore form. It is essential that some organisms survive the heat process in order that the containers can be incubated and the surviving organisms counted. The average thermal process received by a container can be calculated using the equation for F-value. If the product is liquid it is relatively easy to introduce the organisms but for solid products it is necessary to first mix the organisms in one of the ingredients to ensure that they are dispersed evenly throughout the container. Typical levels of the inoculum are between 10^3 and 10^5 organisms per container. An alternative is to use a gas-producing organism and estimate the severity of the process by the number of blown cans.

Encapsulating spores or organisms in an alginate bead allows the organisms to be placed at precise locations within a container or within the food particulates. The alginate bead can be made up with a high percentage of the food material so that the heating rate of the bead is similar to that of the food. This method has been used for continuous processes where the food contains particulates that require evaluation at their centers, and conventional temperature sensing methods cannot be used. Large numbers of alginate beads are used to determine the distribution of F-values that can occur in continuous processes as a result of the distribution of particle residence times. Estimating the exact number to use in a test is not straightforward because it depends on the F-value distribution, which is not known until after the test is conducted and the results analyzed. The number of organisms used will be greater than for an inoculated container test and can be of the order of 10^6 per bead. It is also important that not all are destroyed by the heat process otherwise it is not possible to estimate an F-value.

General Conclusions

When pasteurizing a viscous food, the limits to heat transfer are to overcome the development of boundary layers that impair heat transfer rates at the heated surfaces. Some degree of agitation is advantageous irrespective of whether it is a continuous heat exchange process or one processed in a container. If particulates are introduced then this imposes a further limitation to heat transfer rates. The concepts of high temperature short time (HTST) treatments do not apply easily to viscous foods either with or without particulates. Hence, most thermal processes will be considerably longer than those for foods of lower viscosity. Minimizing the thermal impact is a challenge to food processing companies operating in a market where the consumer demands products of

increased viscosity or consistency, but high quality is also essential.

See also: **Heat Transfer Methods**; **Pasteurization**: Principles

Further Reading

Bolmstedt U (2000) *New Food* 3(2): 15–18.
CCFRA (1992) *Pasteurisation Heat Treatments*, CCFRA Technical Manual No. 27. Chipping Camden, UK: CCFRA.
Department of Health (1994) *Guidelines for the Safe Production of Heat Preserved Foods*. London: HMSO.
Holdsworth D (1997) *Thermal Processing of Packaged Foods*. London: Blackie Academic & Professional.
Richardson P (ed.) (2000) *Thermal Technologies in Food Processing*. Cambridge: Woodhead.
Tucker G and Bolmstedt U (1999) *Liquid Foods International* 3(3): 15–16.

Other Pasteurization Processes

G J Swart, Dairy Belle, Olifantsfontein, South Africa
C M Blignaut and P J Jooste, University of the Orange Free State, Bloemfontein, South Africa

Background

Conventional pasteurization processes usually employ continuous heat-transfer mechanisms. The primary aims of these processes are to destroy pathogenic organisms in liquid foods such as milk and to extend the shelf-life of the product for a limited period of time. (*See* **Heat Transfer Methods**.)

In addition to the conventional processes, other methods exist that employ alternative heating methods, no heating at all, or methods that are not generally regarded as true pasteurization processes. These processes may, under certain circumstances, have the same effect on foods as the conventional pasteurization techniques. It is with this in mind that the following processes have been included in this article, namely pasteurization by irradiation, microwave pasteurization, ohmic heating, 'pasteurization' by blanching, as well as lesser-known versions of the pasteurization process, namely 'cold' pasteurization, extrusion pasteurization, thermization, and chemical preservation using hydrogen peroxide.

Pasteurization by Irradiation

Foods are irradiated for different reasons. When the aim of food irradiation is the inactivation of certain spoilage microorganisms that may be present in foods without necessarily leading to sterilization, the process is called 'radurization.' The 'radicidation' of foods, however, is aimed at the inactivation of pathogenic nonspore-forming bacteria and, in some cases, will also inactivate toxigenic fungi, viruses, and parasites. Both radurization and radicidation have an effect comparable to that of heat pasteurization. To understand the process better, the following aspects are dealt with in this discussion, namely the basic theoretical principles, factors influencing the process, uses, product suitability, nutritional implications, and, finally, the acceptability and current status of the process.

Basic Theoretical Principles

The underlying principle of food irradiation (and therefore also irradiation pasteurization) is based on chemical changes caused in foods by a form of energy called 'ionizing radiation.' Different kinds of radiation are included in this concept. Only a few of these are suitable for use in the treatment of foods, namely X-rays, γ-rays (from ^{60}Co or ^{137}Cs), and electron beams (cathode rays and β-rays). Nonionizing radiation, such as ultraviolet radiation, is essentially absorbed at the food surface and does not have effective penetration properties. Although limited work has been done with ultraviolet radiation, research in this regard concentrates more on extending the shelf-life of fruit and vegetables by controlling fungal spoilage, such as fusarium rot (caused by *Fusarium solani*) charcoal rot (caused by *Marcrophomina phaseolina*) and soft rot (caused by *Rhizopus stolonifer*). Ultraviolet lamps are also used to inhibit the formation of bacterial slimes on surfaces. Of the two radiation treatments, processes using ionizing radiation are more important in the context of the present assignment. (*See* **Irradiation of Foods**: Basic Principles.)

In the case of either radurization or radicidation, an irradiation dose of less than 10 kGy will have the desired effect. For radurization the doses required vary within the 1–5 kGy range. Sometimes, even less will suffice, e.g., beer radurization at 500 Gy. An acceptable end product, however, is not necessarily guaranteed. For the radicidation of foods, the radiation doses vary according to the targeted organism. To control *Salmonella* contamination, for example, doses of between 2 and 6.5 kGy have been recommended to reduce the contamination level three- to sevenfold. An estimated sevenfold reduction of other pathogenic nonspore-forming foodborne

bacteria, such as *Shigella*, *Mycobacterium*, *Escherichia*, *Staphylococcus*, *Streptococcus*, and other species, may also be obtained using doses ranging from 5 to 8 kGy. It should always be remembered that the suggested doses may be influenced by the nature of the food to be irradiated. It is, therefore, essential to take note of the factors that generally influence irradiation processes. (*See* ***Escherichia coli***: Occurrence; ***Mycobacteria***; ***Shigella***; ***Staphylococcus***: Properties and Occurrence.)

Factors Influencing the Process

The success of any form of irradiation depends on the influence and interaction of certain factors. These should also be taken into account when deciding on pasteurization by way of irradiation, since it could influence the postirradiation properties of the food. These factors include:

1. the resistance of some organisms to radiation;
2. the rate at which radiation is applied;
3. the water content of the food to be irradiated;
4. the influence of temperature on subsequent effects in the irradiated food;
5. the presence of oxygen and additives during irradiation;
6. the possibility of combining irradiation with other treatments to obtain the desired effect;
7. the penetration ability of the selected type of ionizing radiation within stated radiation limits;
8. the elemental composition of the food (as a whole) to be irradiated;
9. the magnitude and composition of the initial microbial population present in the food to be irradiated;
10. the susceptibility of the product (such as fruit and vegetables) to radiation damage;
11. cost factors with regard to the process.

Since the above-mentioned factors influence the end result, attention should be given to the applications of the process, product suitability, and nutritional implications that may result from irradiation.

Uses, Product Suitability, and Nutritional Implications

In order to understand the use of radurization and radicidation, it is essential to take into account the fact that these treatments should supplement normal food-processing practices. Radurization especially is used solely to prolong the shelf-life of products. As such, it has shown promise with fresh products such as fish, meat, poultry, vegetables, fruit, baked goods, etc. (*See* **Irradiation of Foods**: Applications.)

Radicidation is usually applied to products in which radurization is not normally able to completely eliminate pathogenic nonspore-forming bacteria. Such foods include fruits and fishery products (e.g., shellfish and fishmeal), although it is necessary to mention that radicidation of shellfish has little effect on the presence of viruses. Irradiation of meat products, such as poultry and pork at radiation levels up to 7 kGy, have, however, been successful in controlling pathogens, such as *Salmonella*, *Campylobacter*, and *Listeria* species. These pathogens cannot be controlled by good manufacturing practice alone. (*See* ***Campylobacter***: Properties and Occurrence; ***Listeria***: Properties and Occurrence.)

With regard to nutritive value, it has been shown that normal nutritional values are retained in proteins, lipids, and carbohydrates after irradiation. The nutritional availability of minerals can, however, be altered by the treatment, and a small amount of vitamins may be destroyed. These losses, however, compare favorably with those incurred during conventional processes. These aspects should have a positive influence on the acceptability and current status of radurization and radicidation. (*See* **Irradiation of Foods**: Processing Technology.)

Acceptability and Current Status of the Process

Irradiation of foods is the one process for which safety aspects have been considered very thoroughly. These include aspects such as microbiological implications, effects on nutritional value, and the possible production of toxic substances, carcinogens, and radioactivity in treated foods. It may be concluded that treated foods can be considered safe on all counts (especially at the dose levels required by radurization and radicidation), provided that:

1. approved doses are applied;
2. the foodstuffs to be irradiated do not possess abnormal levels of elements that could be rendered radioactive; and
3. acceptable packaging materials are used.

A limiting factor with regard to both radurization and radicidation processes is the possible development of objectionable changes in the properties of some foods, such as flavor, odor, color, texture, perceived freshness, etc. Negative effects of irradiation can, however, be counteracted by using the treatment in association with other processes. Lower doses of radiation (radurization) accompanied by a refrigeration process are proving to be technically and economically feasible.

It should be mentioned that irradiation can also have a product-enhancing effect, e.g., the enhanced flavor of brandy. Research with regard to this aspect of the process is continuing. Table 1 demonstrates the current interest in low-dosage irradiation and fields

Table 1 Current interest in irradiation processes from differing perspectives

Type of interest	*Author and year of publication*	*Title of paper*	*Irradiation dose applied*	*Source*
Research	Ma CY *et al*. (1990)	Gamma irradiation of shell eggs; internal and sensory quality, physicochemical characteristics and functional properties	0.97, 2.37, 2.98 kGy	*Canadian Institute of Food Science and Technology Journal* 23(4/5): 226–232
Research	Huhtanen CN (1990)	Gamma radiation inactivation of enterococci	Varying	*Journal of Food Protection* 53(4): 302–305
Research	Stevens C *et al*. (1990)	Effect of ultraviolet radiation on mold rots and nutrients of stored sweet potatoes	Varying	*Journal of Food Protection* 53(3): 223–226
Research	Harris T *et al*. (1989)	Poultry meat irradiation – effect of temperatures on chemical changes and inactivation of microorganisms	0.5–10.0 kGy	*Journal of Food Protection* 52(1): 26–29
Research	Dempster JF *et al*. (1985)	Effect of low-dose irradiation (radurization) on the shelf-life of beefburgers stored at 3 °C	1.03, 1.54 kGy	*Journal of Food Technology* 20: 145–154
Research	Modi NK *et al*. (1990)	Effects of irradiation and temperature on the immunological activity of staphylococcal enterotoxin A	9.4, 12.2 kGy	*International Journal of Food Microbiology* 11: 85–92
Research	Paster N *et al*. (1985)	Preservation of a perishable pomegranate product by radiation pasteurization	2.0, 4.0 kGy	*Journal of Food Technology* 20: 367–374
Overview	Various (1989)	Food irradiation: a most versatile twentieth century technology for tomorrow		*Food Technology* July: 75–97
General	Lacey RW and Dealler SF (1990)	Food irradiation: unsatisfactory preservative		*British Food Journal* 92(1): 15–17

Reproduced from Pasteurization: Other Pasteurization Processes, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

of application. It clearly shows the wide interest (both positive and negative) generated by the process.

In conclusion, it should be stated that the current status of the process depends on individual countries and their health standards. Before commercial exploitation of the process by a manufacturer, the government of the country involved should approve or have approved it. Although many countries permit food irradiation, doubts still exist in some quarters with regard to the safety of the process. (*See* **Irradiation of Foods**: Legal and Consumer Aspects.)

Microwave Pasteurization

Microwave processing of foods has been the focus of much research since the 1940s. As a food-processing method, however, it became of practical interest in the 1960s with the development of multikilowatt conveyorized ovens. The microwave principle only came into general use in the 1980s, when microwave heating applications started to increase. One of these applications was pasteurization using microwave energy. To understand the process better, this article concentrates on the basic theoretical principles of the process as well as its applications and acceptability.

Basic Theoretical Principles

Microwaves are electromagnetic radiant energy waves with wavelengths of between 0.025 and 0.75 m and frequencies of approximately 2450 and 915 MHz in the case of food applications. Most countries have standardized on 915 or 896 MHz for industrial use. The energy level for a specific foodstuff is chosen according to the amount of energy lost in the process due to packaging, for example, since this adversely affects the penetration capability of the microwaves. If a high penetration level is required, a microwave frequency with a lower loss factor should, therefore, be selected.

Another important aspect is the fact that microwave energy has an effect on dipolar molecules such as water molecules. Application of microwave energy causes these molecules to oscillate, causing intermolecular friction and, ultimately, producing heat. Heating is, therefore, achieved inside the foodstuff being irradiated, depending on the type of applicator and the foodstuff being treated. All in all, the treatment facilitates control and the prevention of overprocessing in the case of pasteurization.

Applications and Acceptability

Microwave pasteurization as a process has been applied especially to milk and milk products. The Bach process (using a dual-frequency application) claims great success with regard to cultured milk products. In the case of fluid milk, patents exist for a microwave heat exchanger used in the pasteurization of milk. In spite of the fact that different processes, such as

the high-temperature/short-time, multitherm, and in-pack continuous process, have been developed, the feasibility of microwave pasteurization of milk on an industrial scale has not really been established.

Microwave pasteurization has the following advantages and disadvantages:

1. Amino acids may be altered during the process, and this aspect should be looked into in more detail.
2. Microwaves are successful in pasteurization because of their heat-generating action. The effect it has on microorganisms as such, however, needs to be resolved more fully.
3. Microwave-pasteurized products retain the properties of the fresh product.
4. The process shows promise as a high-efficiency, low-energy process. This limits bacterial build-up in usually restricted flow areas of the system.
5. The design of the system is of paramount importance for successful processing.
6. The effectiveness of heating is related to applicator design and can be either more or less even than conventional methods.
7. The operating costs are high.
8. Magnetrons have a relatively short lifespan and are essential in the conversion of electrical energy to microwave energy.
9. Depending on the design, less factory space may be required.
10. Safety is of great importance, since damage can be caused to the eyes and other tissues that absorb microwaves.
11. Both small static microwave ovens (home pasteurization) and complex, moving-belt microwave tunnels (industrial application) may be used. A schematic diagram of a microwave system for processing foods under pressure is given in Figure 1.

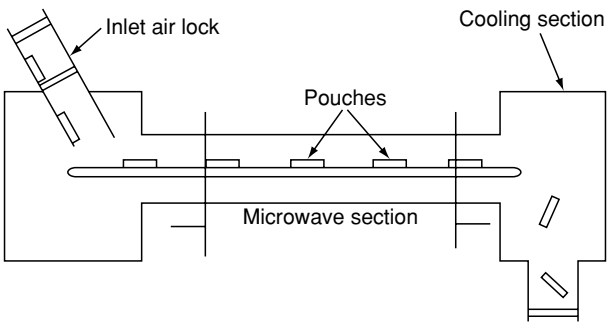

Figure 1 Schematic diagram of a microwave system for processing foods under pressure. From Decareau RV (1985) *Microwaves in the Food Processing Industry*. London: Academic Press, with permission.

12. Microwaves are not necessarily effective in killing bacteria, especially in higher-density products. Manual mixing has been suggested to facilitate more uniform heat distribution. Research in this field is continuing.
13. Heat-sensitive products can be treated to great advantage using microwave technology.

In conclusion, many questions relating to microwave pasteurization and microwave technology in general still remain to be answered. Continuing research should, however, make more general applications a reality.

Blanching

Although blanching is not traditionally seen as a pasteurization process, some authors deem it to be a 'kind' or 'type' of pasteurization. While pasteurization, however, has as its main function the destruction of pathogenic microorganisms, it also destroys certain natural food enzymes in the process. Blanching, however, is applied primarily to inactivate natural enzymes in fruits and vegetables to be processed. Its secondary effect is that it reduces the microbial load. Both pasteurization and blanching are similar in that they employ temperatures below 100 °C.

A process involving acid blanching and the addition of ethylenediaminetetraacetic acid to a canning brine has been a recent topic of research. The results have shown potential with regard to the control of spoilage and botulinal toxigenesis in canned products. This draws a further parallel between blanching (a preparatory method) and pasteurization (a processing method). (*See* **Canning**: Principles.)

With this in mind, the reader is referred to the article on canning for further information.

Ohmic Heating

Ohmic heating is a new food-processing operation which offers a major advantage in the continuous processing and particulate food products.

Basic Theoretical Principle

Ohmic heating occurs when heat is internally generated by the passage of an electrical current through a food product and is in this way similar to microwave heating. In the case of ohmic heating, the penetration depth, however, is virtually unlimited, and heat penetration is more effective. The process depends on the electrical conductivity of the product. Fats, sugars, and syrups are, therefore, not suitable for this type of processing, while foods that contain dissolved ionic salts conduct electricity sufficiently and are,

therefore, suitable for ohmic heating. If both the liquid and solid phases of a product contain sufficient dissolved ionic salts, heat penetration is more rapid than with conventional heating. (*See* **Heat Treatment**: Electrical Process Heating.)

Design of Ohmic Heaters

The ohmic heater is based on a well-known principle, and the apparatus provides an alternative to tubular or scraped-surface heat exchangers for the processing of viscous and particulate food materials. The design of an ohmic heater relies on the principle that an electrical potential is generated across a moving column of product. For safety reasons, the heater column inlet and outlet electrodes are earthed.

Figure 2 illustrates the electrical configuration and construction of the electrical resistance heater.

The heater column is arranged in such a manner that an upward flow occurs in the column while the product is progressively heated to the required temperature. Back pressure is maintained at a constant 1 bar when heating products up to 90 °C, but a pressure of 4 bar is required up to the maximum temperature of 140 °C. The back pressure is controlled by regulation of the pressure on the surface of the product in a presterilized tank using sterile compressed air.

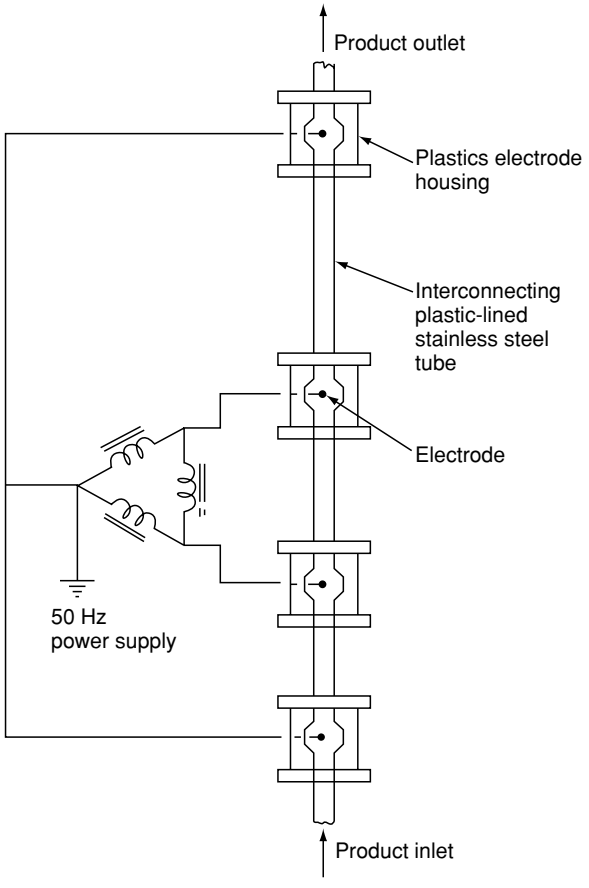

Figure 2 Electrical configuration and construction of the electrical resistance heater. From Skudder PJ, Biss CH and Coombes SA (1988) Ohmic heating: a breakthrough in profitable food manufacture. *Food Review* 15(5): 24, 27.

Acceptability and Applications

The ohmic process offers several advantages over more conventional particulate processing methods. These are:

1. the ability to handle food materials containing particles as large as 25 mm^3;
2. the absence of moving parts (feed pump excluded), which can damage shear-sensitive food products;
3. the ability to heat food products in a continuous flow without the need for any heat-transfer surfaces;
4. a considerably reduced risk of fouling;
5. the generating of heat in the product solids without reliance on thermal conductivity through liquids;
6. the absence of noise during the operation;
7. low maintenance costs;
8. the ease of control and instantaneous start up or shut down;

Examples of food products which have been processed successfully in this manner with minimal structural damage include meat chunks, prawns, diced or sliced carrots, baked beans, and sliced mushrooms. Fruits, such as diced apple and whole strawberries, have also been successfully heat-treated using the ohmic heating principle, and product quality reportedly varies from good to excellent. Ohmic heating has been developed by the APV company as a feasible practical process, and the reader is referred to the article on ohmic heating for further information.

Other Processes

To conclude this article, attention will be given to less well-known, product-specific processes, namely cold pasteurization, pasteurization by extrusion, thermization and chemical preservation by hydrogen peroxide.

Cold Pasteurization

As the name implies, the cold pasteurization technique does not make use of heat to reduce the microbial load. Instead, microporous membrane filters are used to retain the majority of bacteria and yeasts. In this respect, the technique is similar to any other pasteurization process. This process is mainly used in the case of heat-sensitive products, such as draught

beer, wine with less than 17% alcohol (especially sparkling wine) and pulp-free fruit juices. (*See* **Filtration of Liquids**.)

Pasteurization by Extrusion

Although extrusion may not, in general, be comparable with conventional pasteurization, research has shown that the microbial load of thermosensitive food powders, such as egg white powder or whey powder, could be reduced by an extrusion process without any loss of functional properties of proteins and carbohydrates, for example. This advantage, however, depends on the barrel temperatures and contact times, since high temperatures combined with longer contact times lead to a loss of functional properties as a result of mechanisms such as starch gelatinization and protein denaturation. The feasibility of this process as a form of pasteurization is not very clear, and therefore requires further investigation. (*See* **Extrusion Cooking: Principles and Practice**.)

Thermization

The thermization process is a subpasteurization heat treatment of milk at 62–65 °C for 10–20 s, followed by refrigeration. It is used as a prepasteurization treatment of raw milk to safeguard milk quality during prolonged storage in insulated silos. The process is also used as a postpasteurization treatment of dairy products. Research has shown that this process effectively reduces both the total and psychrotrophic bacterial counts, enabling thermized milk to be stored for up to 3 days longer at 8 °C. In this way, detectable sensory changes and the concomitant decrease in shelf-life are limited.

Thermization is only successful if applied to raw milk and products of good quality.

Chemical Preservation by Hydrogen Peroxide

Hydrogen peroxide is undoubtedly a most effective germicidal agent. It has been proven as an effective bacteriostatic and bactericidal agent and has been used in a treatment (30 min; 49–55 °C, 0.08% by weight) that has been suggested as a substitute for pasteurization. The peroxide concentration, temperature used, exposure time, and bacterial species and number all influence the effectiveness of the process. (*See* **Preservation of Food**.)

Treatment by this method destroys coliforms, anaerobic spore-formers and also a number of pathogens. The most resistant pathogen has been found to be *Mycobacterium tuberculosis*. Treatment at levels of 0.115% (v/w) hydrogen peroxide, however, completely inactivates this pathogen. Gram-positive organisms are not inactivated to the same extent as Gram-negative organisms.

Application of hydrogen peroxide as a milk preservative has been attempted in the dairy industry. Certain countries with underdeveloped milk collection industries and high environmental temperatures have expressed interest in this preservation method. The addition of catalase after exposure or the catalase from microorganisms and leucocytes, together with heat treatment, completely destroys or decomposes the hydrogen peroxide. The breakdown products (water and oxygen) are virtually undetectable in milk.

Hydrogen peroxide treatment of milk may be applied to reduce the total bacterial count and to extend the keeping quality. This treatment should, however, be supplemented by approved processing methods to ensure complete destruction of all pathogens. The question of whether hydrogen peroxide treatment of milk may be allowed depends on the specific country and its health regulations.

See also: ***Campylobacter***: Properties and Occurrence; **Canning**: Principles; ***Escherichia coli***: Occurrence; **Extrusion Cooking**: Principles and Practice; **Filtration of Liquids**; **Heat Transfer Methods**; **Heat Treatment**: Electrical Process Heating; **Irradiation of Foods**: Basic Principles; Applications; Processing Technology; Legal and Consumer Aspects; ***Listeria***: Properties and Occurrence; ***Mycobacteria***; **Preservation of Food**; ***Shigella***; ***Staphylococcus***: Properties and Occurrence

Further Reading

Coghill DM, Mutzelburg ID and Birch SJ (1982) Effect of thermization on the bacteriological and chemical quality of milk. *Australian Journal of Dairy Technology* 37(2): 48–50.

Decareau RV (1985) *Microwaves in the Food Processing Industry*. London: Academic Press.

Jay JM (1986) *Modern Food Microbiology*, 3rd edn. New York: Van Nostrand.

Kang KH, Yoon KB and Pack MY (1983) Microbial contamination of raw milk and prevention with hydrogen peroxide. *Korean Journal of Animal Sciences* 25(4): 296–302.

Potter NN (1986) *Food Science*, 4th edn. Westport, CT: AVI.

Skudder PJ, Biss CH and Coombes SA (1988) Ohmic heating: a breakthrough in profitable food manufacture. *Food Review* 15(5): 24, 27.

Urbain WM (1986) *Food Irradiation*. Orlando, FL: Academic Press.

Young GS and Jolly PG (1990) Review: microwaves: the potential for use in dairy processing. *Australian Journal of Dairy Technology* 45(1): 34–37.

PASTRY PRODUCTS

Contents

Types and Production

J B Harte, Michigan State University, East Lansing, MI, USA

Background

There are many types of pastry dough. Shortcrust or standard plain pastry contains four ingredients – flour, fat, water, and salt – that are mixed, rolled, and cut into shapes. Alterations in the proportions of these ingredients, the addition of other ingredients, and the employment of various mixing and manipulation techniques produce popular pastry doughs, such as puff, choux, and strudel pastries. It is the purpose of this article to outline popular pastry varieties and present some common small- and large-scale production methods.

Plain Pastry

Pie Crusts

Shortcrust pastry or piecrusts are prepared by cutting solid fat into the flour and salt, and gradually adding water to obtain the consistency for rolling the dough into the desired shape. A low-protein, pastry flour is used to prevent the development of gluten in the pie dough. Limiting the mixing step after the addition of water and limiting reworking of dough will also help to prevent gluten development that would result in a tough crust.

Medium-flaky crust is obtained by cutting the fat into lumps the size of rice or coarse cracked corn. Flakier crusts are obtained by chilling the water and sometimes the flour to 1.7–4.5°C prior to mixing, cutting fat into larger pieces and/or keeping the mixing step as short as possible. Chilling the dough before rolling also helps to keep the fat firmer. A mealy crust forms when the flour and shortening are mixed thoroughly. The dough should be at about 18 °C after mixing and may be held there for 3–4 h in order to distribute the moisture more evenly. The type of fat used also affects the flakiness or mealiness of the crust. Rolling of the dough should start from the center and continue in the direction that will produce the desired shape and thickness.

Many varieties of pie crust dough, with fat contents ranging from less than 50 to 75%, are used. Pie crusts can also be prepared with a combination of shortcrust dough and short pastry (pâte de sucre or flan pastry) or with short pastry alone. Pie crust dough can be used for two-crust, fruit-filled pies, single-crust, and soft-filled pies. The typical one- and two-crust pies are approximately 23 cm in diameter and serve six to eight people. Crusts are also used for baked meat or fish pies, quiches, tarts, and turnovers. Fruit-filled turnovers can be made from dough scraps.

In two-crust fruit pies, the bottom crust should be slightly thicker than the top in order to support the fruit filling. Placement of cooled fruit filling in the prepared pie crust allows it to bake adequately before the filling begins to boil. Two to four cuts should be made in the upper crust to vent the steam during baking. Crusts are baked at 281–232°C, reducing the moisture content to a low level.

Two types of machines are used to form pies. Pie crusts can be produced commercially on a small-scale basis using a pie press that shapes dough balls into pastry sheets. Alternatively, pies can be formed by automatic sheeting and cutting. The rotary pie machine mixes the dough, reduces its thickness and cuts it into 7.5 × 12.5-cm sections. These pieces are automatically cross-rolled into thin, round pie pastry sheets that are manually placed into the pie plates. The top crust is placed on top after the filling is added to the bottom crust. The rotary pie machine then finishes the pie by crimping, scaling, and trimming.

The straight-lined pie type is another example of a sheeting and cutting machine that is more completely automated than the rotary pie machine. Pie dough is divided in two and separately reduced, cut into 7.5 × 11.5-cm pieces and cross-rolled into the top and bottom round pastry sheets. The pastry is automatically placed into the pie plate, filled, topped with another cut pastry sheet, sealed, crimped, and trimmed.

Short Pastry

Short, pâte de sucre or flan pastry is a rich dough containing a large amount of fat and, unlike standard

pastry, eggs and sugar. This pastry is used for biscuits, tarts, some pies, and the bottom and sides of fruit flans and cheesecakes. If a nonsweet pastry is desired for meat, vegetable, or seafood dishes, other ingredients, such as Parmesan cheese, may be substituted for the sugar. Smooth blends of shortcrust and pâte de sucre doughs are often used for pie shells and cream rolls. Chocolate or cocoa may also be added to these blends.

Puff Pastry

Puff pastry is lighter, flakier, and more delicate than standard shortcrust or flan pastry. It contains a higher fat level and has a unique preparation method that is critical for preparation of a successful product. This pastry can be used for a wide variety of products, including turnovers, shells filled with whipped cream, custards or fruit, vol-au-vents, and sausage rolls, and is considered a staple in commerical bakeries.

The pastry is prepared with a high-protein flour (bread), water, salt and shortening. A firm, plastic shortening, equal to 50% of the dough weight and containing some moisture, is optimum. Other ingredients, such as eggs and acids (cream of tartar, vinegar, or lemon juice) can be used. There are three commonly used methods for mixing and handling the puff pastry dough – the French (roll-in), English (roll-in), and the Scottish (all-in) method.

The French method involves rubbing a portion of pastry fat into the flour to lubricate the gluten. A portion of the water, acid, and eggs is placed into the center of a ring made from the mixed flour and shortening. The dough is mixed well, and the procedure is repeated until all the ingredients, with the exception of the remaining fat, have been added. The dough is rolled into a rectangle, 1.25 cm thick with a thinner outside border, and the fat is rolled to the size of the thicker dough. After the fat is placed on the thicker dough, the thinner dough is folded over on the thicker dough, and the thinner dough is folded over, making a type of pocket. The fat and dough is then rolled into a rectangle. One half-turn of the dough is formed by folding one end of the dough two-thirds of the way down the rectangle and folding the opposite end to cover the first fold (see **Figure 1b**). The new rectangle is three layers of pastry fat separated by dough layers. The dough is relaxed, and the rolling procedure is repeated after turning the dough through 90° for a total of six half-turns. The French method of puff pastry preparation is the most popular but is not used by some, because it is complicated and difficult to produce under high-speed commerical manufacturing processes.

The English method is the most popular in the USA. The dough is sheeted into a rectangular shape, and the roll-in fat is spread uniformly over two-thirds of the area. Three-fold pastry preparation (**Figure 1**) is used to form three dough layers separated by two layers of fat.

The Scottish method is similar to pie-crust preparation, simple, and widely used commercially. All the pastry fat is broken into lumps (5 cm) and mixed loosely in the flour. The remaining ingredients are then added to form a lumpy dough. The rolling procedure is the same as the French method with the addition of one extra half-turn. After the final rolling, the dough is shaped and relaxed before baking. For maximum volume, puffed pastries should be baked at high temperatures (204–212°C). The Blitz method is a modification of the Scottish method that partially develops the gluten prior to adding the firm shortening.

Two mechanical mixers are commonly used in batch commercial pastry production. The low-speed, double-arm mixing machine with a large dough bowl and slow mixing produces a good-quality commerical puffed pastry. The Scotch pastry method must be followed when the high-speed mixer is used in large-scale, high-product-capacity operations. Careful timing is critical with these mixers to prevent overmixing that results in an intimate mixing of fat with the dough and formation of a short rather than a

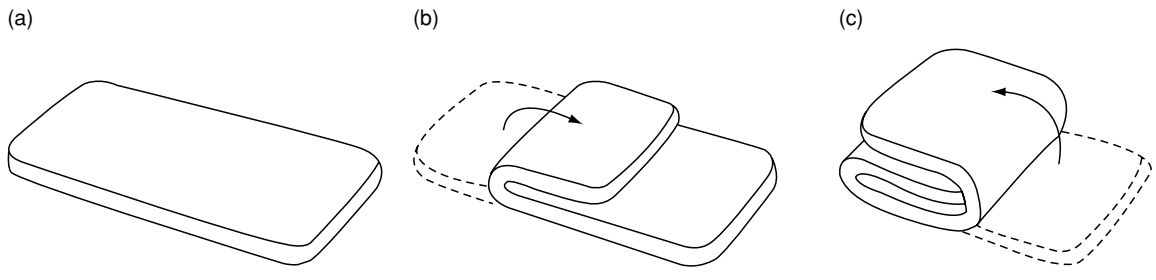

Figure 1 Three-fold pastry preparation: (a) the dough is rolled into a rectangle; (b) one end of the dough is folded two-thirds of the way down the rectangle: (c) the opposite end is folded to cover the first fold.

puffed pastry. After mixing the dough, any of the methods described in the commerical pastry preparation section can be applied. Continuous puff pastry dough-production equipment extrudes the dough and shortening into a double-layer continuous tube with the shortening in the center layer. This is flattened and reduced by a dough stretcher. The sheet is folded, stretched, and thinned. Pieces are cut into their final shape for pastry shells, cheese or savory sticks, napoleons (mile feuilles), turnovers, sugar crisps, etc. Products prepared with this continuous process are usually intended for wholesale distribution. Thus, their size and flakiness are limited to prevent breakage during shipping.

Choux Pastry (pâte é chou)

A wide variety of pastry products are made using choux paste: cream puffs, éclairs, French crullers, soup nuts, and decorative shapes used for pastry garnishing. Choux pastry differs from other pastries in that it is cooked before it is piped or spooned. It is composed of 150–200 parts eggs, 100 parts flour, 100 parts shortening (butter) and approximately 200 parts water. Flour is quickly stirred into a mixture of liquid, seasoning and fat, which is heated to boiling. The fat must be completely dispersed in the hot liquid to form a proper emulsion. Stirring continues until the smooth, gelatinized paste no longer sticks to the pan. The mixture is cooled, and the eggs are beaten into the choux mix one or two at a time, mixing well after each addition. Liquid may be gradually added to make the final paste the desired soft consistency. A small amount of chemical, fast-acting leavening can be added to this liquid to obtain the smooth consistency and maximum pastry volume. The paste is spooned or piped on to parchment-paper-lined pans and baked at 215–226°C until crisp. (*See* **Leavening Agents**.)

Choux pastry can be prepared using a continuous extrusion process. The extruder has groups of processing zones provided by varying the pitch of the screw. Throughout these zones, ingredient mixing, kneading, starch gelatinization, and comminuted/homogenization are accomplished. The final step involves extrusion of the dough through a die, forming the desired choux pastry shape. (*See* **Extrusion Cooking**: Principles and Practice.)

Strudel Pastry

The dough used to prepared strudel can be of the puff pastry type or can be made from water, seasoning, eggs, shortening or oil, and high-protein flour. Optional ingredients include sugar and lemon juice. The fat is mixed or cut into the flour mixture, and the liquid ingredients are added gradually. The dough is made into a ball, kneaded, and shaped into a rectangular form, relaxed, then rolled and gently stretched to obtain very thin, transparent dough. Oil is brushed on the thin layers. Home preparation by hand stretching is difficult. Commercial methods have been developed to prepare thin sheets. Strudel pastry and Fillo (phyllo) dough, which is similar to strudel, are frozen. These are sold in wholesale and retail markets and have wide applications for the home baker. Strudel filling is placed at one end of the pastry, and the dough is rolled over the filling as in the preparation of a Swiss roll. Relaxed strudel is baked at 218–233°C.

Yeast-leavened Pastry

Danish Pastry

Danish pastry is a rich pastry that combines the principles of puff pastry and fermented, sweetened dough. The dough contains high-protein flour (11.5–12%) that is often combined with up to 30% pastry flour for easier handling during lamination. Other ingredients include butter or margarine, yeast, sugar, shortening, milk solids, eggs (optional), salt, and flavoring. Fat containing moisture will help produce steam for leavening and flakiness. The total amount of fat can range up to 50% of the pastry dough. Sheeting and folding fat between layers of dough forms thick films of fat separating dough layers. Two types of Danish pastries are made in the USA. The European style is made from lean sweet dough that has a short shelf-life. The American style is more common and is made from a richer dough with a longer shelf-life.

There are two commonly accepted mixing procedures for the preparation of Danish pastry: conventional (three- or four-fold) or the short mixing time method, and the lamination or the long-mixing-time method, which is used for automated, high-speed sheeting systems. In the conventional three-fold method (Figure 1a–c), all the ingredients are combined with the exception of the rolled-in fat. A rectangle is formed by rolling the dough to a thickness of 1.25 cm. A large portion of roll-in pastry fat is evenly distributed over two-thirds of the dough, and the rectangle is folded in thirds, starting with the portion without fat. The dough is then rolled into a rectangle of 1.25-cm thickness and folded into four equal parts. After retarding, the dough is rerolled, and the procedure is repeated two more times. The four-fold or book-fold technique (Figure 2a–c) involves distributing fat over the center 50% of the dough and folding

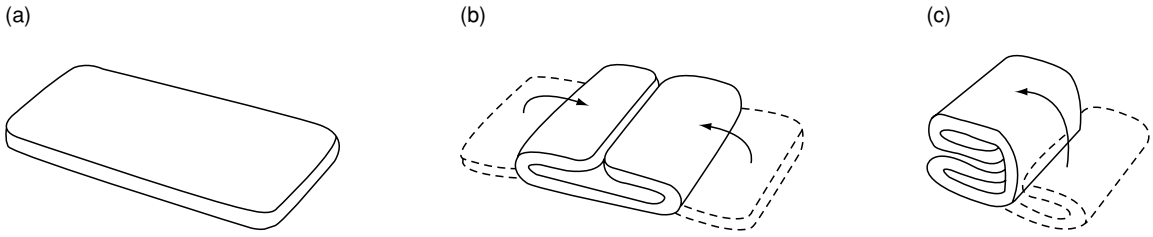

Figure 2 Four-fold pastry preparation: (a) the dough is rolled into a rectangle: (b) both ends of the dough are folded halfway down the rectangle; (c) one folded end is turned to cover the opposite fold.

the two ends of the original rectangle to the dough center. One end is folded over the other fold to line up with the far edge. The retarding and refolding steps are then repeated, similar to the procedure described for the three-fold method.

The gluten structure is developed during the laminating rather than during mixing. The commercial lamination method involves layering the shortening between two pieces of dough. This can be accomplished by extruding dough around the shortening, or high-speed pumping of the shortening between two continuous dough sheets. Automatic or overlapping of 45–60-cm dough pieces is followed by sheeting. This process is repeated two more times before cutting.

The retarded dough may be shaped into a wide variety of pastries, including crescents, sticks, rings, filled pockets, and rolls. Individual pastries are baked at 193–204 °C. Larger Danish products, such as coffee cakes, are baked at lower temperatures, 176–190 °C, to prevent over-browning before the pastry structure is set.

Croissants

The croissant is a popular, sweetened, yeast-raised product that is delicate and flaky. It has a medium-grain texture and the shape of a crescent. Dough composition and preparation techniques are similar to those of Danish pastries, with total fat contents ranging from 25 to 50%. The rolled in fat is a critical component for laminated yeast doughs that should have the same consistency as the dough. It provides the two-dimensional gluten framework that results in the characteristically flaky croissant. Additional ingredients are sugar, salt, whole eggs, yeast, and water. Dough (19°C) is mixed completely, without over-developing the gluten, for 1 min in a vertical high-speed mixer. The dough is extruded, and fat is extruded on top. It is then flattened, rolled and rerolled, and laminated for a total of 36 layers. Most of the gluten development occurs during the lamination step. Retarded dough is rolled and cut into triangles with slightly longer sides than the base. Filling is added, and the dough is rolled with four rotations, leaving the peak touching the baking sheet. The 'rolled' pastries are proofed and then baked at 163–204°C.

Toaster Pastry

Toaster pastries are convenience pastry products that are popular in the USA. These breakfast or snack pastries are made from a filled dough, formed into a shape and size that can fit into a toaster. The pastry dough is prepared from soft wheat flour, sugar, shortening, chemical leavening, salt, flavorings, and colors. It is baked just enough to set the dough and reduce the moisture content, thus extending the product shelf-life. Although many are shelf-stable, many new varieties are frozen. Browning is completed in the toaster.

Commercial Preparation of Pastry

Most pastry products evolved after long periods of small-scale, manual preparation followed by high-speed, automated production in this century. High-capacity production of pastry is made possible by elimination of manual dough manipulation and automation of the pastry production process. Processing time is shortened by elimination of roll-in steps, and reducing resting times between fold-in steps. Variations exist in the continuous production methods for the different pastry types such as pie, Danish, and puff pastries. However, common to all are the sheeting, depositing, lamination, and cutting steps.

An example of an automated pastry production system is presented in Figure 3. The dough is mixed and carried to the sheeter head, or three-roll laminating head, by the dough feeder and conveyer belt. The sheeter head unit consists of three rollers with adjustable speeds for producing perfect sheeted pastry. Two are feed rollers, one with deep grooves and the other with fine triangular grooves. The third roller has a smooth surface and is located underneath the feed rollers. Fat layers can be added to the first dough to

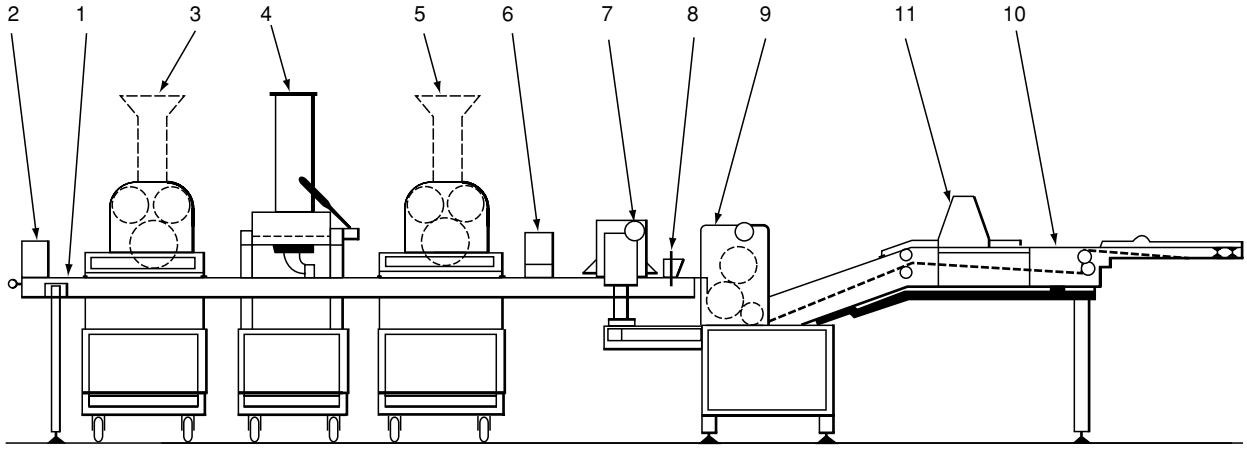

Figure 3 Automated pastry production system. (1) Supply table; (2) flour duster; (3) three-roll laminating head; (4) fat pump; (5) three-roll laminating head; (6) flour duster; (7) cross-roller; (8) guiding rollers; (9) two-roll sheeter; (10) retracting belt; (11) guillotine. From Rijkaart C (1984) Producing the Perfect Pastry. *Food Manufacture* 59: 29 with permission.

produce the flaky texture of pastries. A second sheeter head adds another dough layer over the fat. If the Scottish system is used, all the fat is incorporated into the dough, a fat extruder is not needed, and only one sheeter head unit is required.

Dusting flour boxes distribute small amounts of flour on the conveyor belt, to prevent sticking, and to the top of the dough sheets prior to cross-rolling. Conical cross-rollers travel across the belt width, thinning and spreading the dough sheet to the desired thickness. Further sheeting occurs at the two-roll sheeter, reducing dough thickness. Several systems can be utilized to obtain the desired flakiness by controlling the number of times the pastry is folded.

The book-fold method (Figure 2a–c) involves folding the dough sheet and then turning it through 90 °C and sheeting the dough, repeating the procedure to obtain the optimal flakiness. Another method involves continuous folding and sheeting of the dough. A third method uses a guillotine to cut the dough sheet into pieces (50 × 60 cm) and then stacks then with a retracting conveyor. The dough pieces are placed on another conveyor belt operating at a 90° angle to the first belt. The pieces are stacked five layers high by this process, after which the dough sheets are reduced in thickness. Another method utilizes a curling arm, or roll winder, which coils the dough sheet covered with extruded fat into five layers. A final technique uses a ring extruder that produces a very uniform, hollow dough cylinder, lined with fat. This method is particularly suited for long-shelf-life pastries. For all methods, the dough sheets can be reduced in size by pressing, cross-rolling, and running between a multiroller device. This multiroller allows the dough sheet to be manipulated when in contact with the small rollers, which press the dough on to the larger roller. The small rollers revolve individually as the whole group of small rollers also rotate together in a loop. The dough is allowed to relax when it is between the rollers and is not touching the large roller. A similar stretcher system with a larger number of small rollers, but without the bottom roller, has also been used. The final pastry production steps repeat the sheeting and reducing steps, depending on the necessary dough thickness and desired flakiness. Preparation of individual pastries of varying sizes and shapes depends upon the type of cutting, depositing and shaping equipment utilized.

See also: **Extrusion Cooking**: Principles and Practice; **Leavening Agents**

Further Reading

Doerry W (1998) Formulation and production of puff pastries. *AIB Research Department Technical Bulletin* XX (2): 2–5.

Friberg BO (1990) *The Professional Pastry Chef*, 2nd edn. pp. 14–23, 29–34, 94–97. New York: Van Nostrand Reinhold.

Goodsell GR (1985) Making Danish – basics that determine quality results. *Bakers Digest* 59(3): 16–21.

Loving HJ and Brennesis LJ (1981) Soft wheat uses in the United States. In: Yamazaki WT and Greenwood CT (eds) *Soft Wheat: Production, Breeding, Milling and Uses*, pp. 169–207. St. Paul, MN: American Association of Cereal Chemists.

Matz SA (1987) *Formulas and Processes for Bakers*, pp. 1–26, 39–55, 58. McAllen, TX: Pan-Tech International.

Matz SA (1992) *Bakery Technology and Engineering*, 3rd edition. McAllen, TX: Tan-Tech International, Inc.
Nicolello LG and Dinsdale J (1991) *Basic Pastrywork Techniques*, 2nd edn, pp. 45–46, 53–58. London: Hodder & Stoughton.
Pyler EJ (1988) *Baking Science and Technology*, vol. II, pp. 1107–1183. Meriam, KS: Sosland.
Rijkaart C (1984a) Producing the perfect pastry. *Food Manufacture* 59(7): 29–31.
Rijkaart C (1984b) Producing the perfect pastry. *Food Manufacture* 59(8): 46–47.
Sultan WJ (1990) *Practical Baking*, 5th edn., pp. 217–328, 379–423, 425–485. New York: Van Nostrand Reinhold.

Ingredient Functionality and Dough Characteristics

S L Andrews and J B Harte, Michigan State University, East Lansing, MI, USA

Introduction

Traditional pastry or pie crust is a simple dough made from four primary ingredients – flour, fat, salt, and water. Numerous pastry varieties have been developed by slightly altering ingredients, their amounts, mixing methods, and shaping. Common ingredient modifications include the amount or type of fat, the protein content of flour, the substitution of milk for water, use of chemical leavening agents and yeast, and the addition of other ingredients such as eggs, sugar, acid, and flavorings. It is the purpose of this article to outline pastry ingredient functionality and their proportions and desirable pastry characteristics. US and equivalent UK pastry terms are presented in **Table 1**.

Functionality of Ingredients

Flour

Flour is the main ingredient in pastry, and the type of flour used affects the characteristics of the pastry.

Table 1 US and equivalent UK pastry terms

US	*UK*
Standard plain pastry	Shortcrust pastry
Short pastry	Pâte de sucre or flan pastry
Cookies	Biscuits
French puff pastry	Flaky pastry
Scottish puff pastry	Rough puff pastry

Protein in flour yields gluten wherever the flour is dampened with water and manipulated. Gluten development increases as the flour's protein content increases and as the amount of added water increases. Gluten development can be minimized by incorporating fat and flour before water is added. This method insulates flour proteins from added water and limits gluten development. The extent of gluten distribution throughout pastry and the amount of gluten formed determine whether pastry is crumbly, with a tendency to be compact and tough, or tender and flaky. When dough is rolled and baked, gluten is denatured by heat, which contributes additional toughness. High-protein flours yield more gluten, which results in a more cohesive dough. In contrast, pastry flour and other low-protein flours do not yield as much gluten and tend to make a tender pastry which crumbles and cannot hold its shape. A mixture of high gluten or plain flour blended with a soft cake or pastry flour can be used to achieve the desired protein concentration.

In choux pastry a high-protein (12–13%) flour is used. During baking, egg and flour proteins in choux pastry coagulate, preventing steam and gas from leaving the pastry, and setting the pastry into the desired form. High-protein (13%) flour is also used in yeast-leavened Danish pastries. For croissants, a 12.5% protein flour is recommended. (*See* **Flour**: Analysis of Wheat Flours.)

Fat

Fat contributes tenderness, or shortness, to pastry. Depending on the type of pastry, the fat content can range from 25% to almost 75% of the dough. Fat tenderizes pastry by waterproofing flour particles. The polar groups in water have an affinity for the polar groups in both the protein and starch. Polar carbonyl groups and the double bonds in unsaturated fatty acid moieties make it possible for fat to unite with polar groups on the surface flour particles. The remaining portions of the fat molecule have no affinity for the flour or water and act as a mechanical barrier, preventing contact of the water and protein in the flour.

A fat's ability to interfere with gluten formation is known as its shortening power. Pure fats have more shortening power than do butter or margarine which contain 16% water. Even pure fats, such as lard, hydrogenated shortenings, and oils, exhibit different characteristics in a pastry product. Oil is more dense than lard, which is more dense than shortening. In addition, liquid fats have more spreading power and are able to coat flour more evenly and completely. The higher the ratio of liquid to crystals, the greater the covering power of the fat.

In addition to making pastry tender, fats also contribute desirable flakiness by separating the dough into layers or flakes. Oils, on the other hand, tend to coat each particle of flour. As a result, water contact with the flour is limited, little gluten is developed, and a tender but crumbly, or even greasy, pastry results. A flaky, tender pastry can be made with oil, but, because oil mixes so easily with flour, the mixing must be carefully monitored. Oil and flour should be mixed until the particles are the size of peas. Overmixing will cause particle size to decrease, resulting in a crumbly pastry.

Plastic fats, having the properties of both liquid and crystalline fats, tend to make a more flaky product which is tender. Lard is considered an excellent pastry fat because of its shortening action, desirable plasticity, and dispersibility which enhances flakiness. Selection of fat depends on a number of factors, including ambient temperature, method of processing, and desired crust characteristics (mealy or flaky). (*See* **Fats**: Uses in the Food Industry.)

Puff pastry dough contains up to 100 parts of a good-quality, high-protein flour, 58 parts of water, and two parts of salt. A firm, plastic, shortening, equal to 50% of the dough weight and containing some moisture, is optimum. Fat containing moisture will help to produce steam for leavening and flakiness.

Liquid

The liquid most frequently used in pastry is water, but liquid is often contributed by the addition of milk and eggs, and even fats such as butter and margarine which are 16% water. Liquid is needed to hydrate the flour, develop gluten, and provide cohesion to the dough. Without liquid, the flour particles would not adhere to form a dough. With insufficient liquid the dough crumbles and is difficult to handle. Too much liquid causes excessive gluten formation, producing a tough pastry. While too much liquid is undesirable in pastry doughs, liquid is essential in choux pastry for maintaining the desired soft consistency of the paste. Liquid in the form of water or milk is gradually added to produce the final paste consistency. If the paste is too thick, the baked volume will be low and the shell will be thick.

To promote flakiness, water used in pie dough mixtures should be chilled. This limits homogenization during dough mixing by creating hard fat particles. Ice water may also be used. Liquid is also necessary for leavening. When baked, liquid in pastry dough produces steam, leavening the product and separating the individual flakes. Fat containing moisture will also help to produce steam for leavening and flakiness.

Salt

Salt functions as a seasoning and is not required to produce a successful pastry product. A crust without salt will produce a flat-tasting crust. However, its omission does not affect the mechanical aspects of the pastry.

Other Ingredients

Eggs Eggs are incorporated into several pastries, including choux paste, short (pâte de sucre, or flan pastry), and croissant doughs. Eggs are optional ingredients in yeast-leavened Danish and puff pastry doughs. Egg yolks can function to lubricate the gluten, while the whites aid in baked pastry volume. Eggs contribute richness, structure, and increased keeping qualities. In choux pastry, for which 90–100 parts whole egg is beaten into a mixture of 60 parts high-protein flour (12–13%), 100 parts water, 40 parts shortening, and two parts seasonings to form a very thick paste, the primary functions of eggs are dough emulsification, leavening, and supplying moisture.

Acids Acids such as cream of tartar, vinegar, or lemon juice are used in some pastry formulations. Acids relax the gluten, thus improving the ease of rolling. If too much acid is added, the dough will be sticky and difficult to handle and roll. (*See* **Acids**: Natural Acids and Acidulants.)

Sugar Sugar helps to tenderize the dough, adds extra flavor, and promotes crust browning. Lactose, the simple sugar found in milk, also contributes to product browning. (*See* **Lactose**.)

Milk In some pastries, milk replaces water as the source of liquid. Milk adds crust color by providing lactose, which promotes Maillard browning during baking. Milk solids strengthen the interior cell structure, which helps to maintain pastry flakiness. Milk can also aid in producing a thin, smooth crust. (*See* **Browning**: Nonenzymatic.)

Yeast Yeast is used in croissant and Danish pastry doughs to provide leavening through the production of carbon dioxide. Danish pastry doughs may contain a high concentration of yeast to offset the growth-inhibiting effects of added sugar.

Proportions of Ingredients (by Weight)

For a traditional plain (shortcrust) pastry, 100 parts flour (average 115 g per cup) are used for each 42–59 parts fat. The optimal amount of plastic fats, such as lard, is 44 parts to 100 parts of flour. If less than 44

parts are used it may become difficult to mix in the water without overdeveloping the gluten. When liquid fat such as vegetable oil is used, 42 parts should be used to each 100 parts of flour. If more than 59 parts of fat are used, the pastry becomes crumbly and greasy. Within the recommended range, the smaller the proportion of fat, the greater the likelihood of developing excessive gluten. For each 100 parts flour, 2.5 parts salt are used for flavoring.

A minimal amount of water must be added to develop gluten (flakiness). If too little water is added, the pastry will be crumbly, but too much water will cause excessive gluten development and the pastry will be tough. A mixture of 31 parts water to 100 parts flour is sufficient to provide flakiness without promoting toughness. When a minimal amount of fat is used, careful measurement of liquid becomes even more important. A difference of as little as 2.5 parts water per 100 parts flour can cause a discernible difference in the tenderness of the pastry.

Pastry Characteristics

Tenderness and flakiness are two desirable characteristics of pastry products which can occur together or separately. Up to an optimum, tenderness increases as fat is increased, as a result of the shortening power or ability of the fat to coat the gluten strands, interfering with gluten structure. The fat can also function to coat the flour particles and prevent water from fully hydrating the flour, thus preventing gluten formation. Higher temperatures increase the ability of the fat to spread, coating the gluten and increasing tenderness. Keeping fat and water cold will serve to promote a flaky, tender pastry which is not greasy or too tender. A product which is too tender becomes crumbly and too hard to handle. Overmanipulation, including kneading of pastry dough, has the opposite effect. As the dough is kneaded, rerolled, or mechanically overmixed, the pastry becomes tough as a result of excess gluten formation. Toughness can also be caused by adding too much liquid, or by the presence of excess flour during rolling.

Flakiness results from partially hydrated flour or dough being layered between thin sheets of fat. Pastry layers are separated during baking as steam and, to some extent, air separates the layers. These fat layers must coat the flour sufficiently to produce layers, yet not enough to cause the pastry to crumble. Cutting fat into the flour in large particles promotes flakiness. Medium-flaky pastry is made by cutting fat into small lumps, while a mealy crust will form when the flour and shortening are mixed thoroughly. Use of cold water helps to promote flakiness in the final product. For long flaky pastry the fat is cut into larger lumps and the lumpy, sticky dough is rolled three times, turning three times for each roll. The harder fats do not spread as readily as softer fats and thus are more likely to have flour particles adhering to the fat. Optimum mixing time for desirable tenderness is longer when harder fats are used. Dough rerolling and folding increase flakiness. Increased gluten development can hold steam more efficiently and separate the dough into layers. The resultant increased toughness may be reduced by rolling small pieces of fat between the dough layers.

Dough Retardation and Relaxation

During dough mixing and handling, fat must remain cold enough so that it does not soften during preparation. This is especially true for roll-in fat products, such as puff and Danish pastries. Additional fat folded into the dough creates numerous layers of dough and fat. To prevent changes in fat properties that occur quickly at elevated temperatures, dough is retarded or relaxed at refrigerator temperature between sheetings. To prevent greasiness and decreased volume during baking, cooled pastry should reach ambient temperature slowly.

Leavening

In pastry, leavening occurs when heated liquid creates steam, and individual layers (flakes) are separated. Baking temperature is important to insure sufficient leavening. Too low a temperature will not allow the optimum production of steam for aeration, while an oven temperature that is too high will result in premature setting of the proteins and low pastry volume. Leavening also occurs as air expands during baking.

In yeast-leavened Danish pastry and croissant doughs, yeast is added for additional leavening from carbon dioxide. In choux paste, a small amount of chemical, fast-acting leavening – ammonium bicarbonate and baking powder (1:1) (14 g per 454 g extra liquid) – can be added to this liquid to obtain the smooth consistency and maximum pastry volume. In choux pastry, if eggs are used to soften the paste, the liquid and chemical leavening mixture is not needed. Leavening of the choux paste during baking occurs by expansion of the eggs; carbon dioxide is released from the chemical leavening agents, and steam is formed from heated liquids in the paste. Flour and egg protein coagulate to set the shape of the aerated pastries. (*See* **Leavening Agents.**)

See also: **Acids**: Natural Acids and Acidulants; **Browning**: Nonenzymatic; **Fats**: Uses in the Food Industry; **Flour**: Analysis of Wheat Flours; **Lactose**; **Leavening Agents**

Further Reading

Charley H (1982) *Food Science*, 2nd edn, pp. 243–255. New York: John Wiley.

Daniel AR (1963) *Baking Materials and Methods*, 4th edn, pp. 362–391. Essex: Elsevier Science.

Goodsell GR (1985) Making Danish – basics that determine quality results. *Bakers Digest* 59(3): 16–21.

Loving HJ and Brenneis LJ (1981) Soft wheat uses in the United States. In: Yamazaki WT and Greenwood CT (eds) *Soft Wheat: Production, Breeding, Milling and Uses*, pp. 169–207. St Paul, Minnesota: American Association of Cereal Chemists.

Pyler EJ (1988) *Baking Science and Technology*, vol. II, pp. 1107–1183. Meriam, Kansas: Sosland.

Rijkaart C (1984) Producing the perfect pastry. *Food Manufacture* 59(7): 29–31.

Rijkaart C (1984) Producing the perfect pastry. *Food Manufacture* 59(8): 46–47.

Sultan WJ (1990) *Practical Baking*, 5th edn, pp. 217–328, 379–423, 425–485. New York: Van Nostrand Reinhold.

Pathogens *See* **Emerging Foodborne Enteric Pathogens**; **Food Poisoning**: Classification; Tracing Origins and Testing; Statistics; Economic Implications; **Microbiology**: Classification of Microorganisms

PEACHES AND NECTARINES

C Blattný, Prague, Czech Republic

Introduction

Both peaches and nectarines originated from the same tree – the peach, whose botanical name is *Prunus persica* L. (rose family – Rosaceae). They are a typical stone fruit with the seed in a hard stone; soft pulpy flesh surrounds the stone and the surface is covered with a relatively thin skin. The peach and nectarine are closely related to other species of the genus *Prunus*: apricot, almond, plum, and cherry.

Origin and History

The peach is a native of China, where it was depicted in old paintings and mentioned in Chinese mythology and folklore as far back as the tenth century BC. The peach signifies long life and peach blossom has also been important in China for a very long time. There semiwild forms can be found in all native peach areas from south to north-west and north-east.

It is unclear how peach and nectarines spread to Europe. At least two or three ways are probable. First, when the Romans occupied Syria to the end of the first century BC and found the peach, they named it 'Persian apple.' It must have been important that peaches grow quickly from seed, and this permitted their rapid dispersal from China to the west. The second way was probably across Egypt to Rhodes and the third route was over Greece and the Balkans. Since the beginning of the second millennium peach and nectarines spread throughout Europe, especially in the south and south-east. France played an important role in the development of the peach – at the beginning of the seventeenth century about 30 cultivars of peach were grown there. During the following two centuries this number reached more than 350 cultivars. Also in Italy interest in peach and nectarine increased from century to century. Much archeological evidence of peach seeds has been found there along the central Donau in Hungary and Slovak republic. The Spaniards brought peaches to America (Mexico and Florida) in the sixteenth century. Peaches spread in America very quickly and, especially in California, the peach became the leading fruit grown in orchards.

Requirements of the Peach Tree

Peaches have different requirements according to their old native district. The hardiness, early or late blooming, and requirement of winter cold exposure (chilling) are most important. Most older cultivars need more than 950 h at around 6 °C. However it should also be emphasized that the peach tree is a warmth-seeking botanical species. It needs higher

temperatures in summer with strong intensive sunshine and relatively mild frost in winter – attributes of a mild climate. However, cultivars from south China are suitable for subtropical regions as they need less chilling.

The peach has very similar requirements for environmental conditions as the vine. The soil for peach should be warm, not too heavy, airy, pH 6–7.5, and $CaCO_3$ 3–5%. Selection of the rootstock for given conditions is important. The following seedlings are used in various countries: peach seedling (various types), peach × almond hybrids, almond seedlings, plums seedlings, *Prunus davidiana* (China), and peach × *P. davidiana* hybrids. Peach seedlings prevail: in some countries peach × almond are very often used.

Nectarine cultivars have greater environmental demands. They need warmer summers (without rain), as the fruit's skin tends to crack after rain.

The Fruit

Peach is one of the most variable of all tree fruit species. From the pomological point of view, the cultivars are divided into five groups:

1. True peaches – freestone or semifreestone with fuzzy skin
2. Clings – clingstone with fuzzy skin
3. Nectarines (smooth-skin peaches) – with fuzz-free skin and freestone
4. Brugnons – small clingstone nectarines
5. Peento – a special group of peaches with flat fruit (almost like dried figs) and exposed stone

This division is by no means accurate. It is necessary to emphasize that in some years when there is a cold and rainy summer, a stronger fusion of flesh and stone also occurs with freestone cultivars.

Cultivars of peach and nectarine differ in their height, width, and thickness. Both peaches and nectarines are round or oval-round, oval, oblong, or somewhat flat. The suture leading from the stem pit to the top of the fruit or behind it is a typical feature. The depth of the suture also differs between cultivars. Some have a small tip on the top, some have deeper ones, and a small number of cultivars have a small knoll (like Venus). The skin of true peaches and clings is fuzzy. This fuzz differs according to the cultivars – sometimes it is very fine, sometimes it is dense and very fuzzy. The skin is orange-yellow, white with red cheeks, or red with stripes, or plain red. The flesh is soft, sometimes too firm or too fibrous, yellow, orange-yellow, nearly white, green-white, or red. In old-type cultivars, which have only been selected to a limited degree, the flesh is often farinaceous. The color of flesh around the stone and under the skin is slightly darker and sometimes red.

Nectarines usually have a somewhat smaller fruit and a smooth shiny skin (like plums); the flesh is harder, more like a plum, and not melting like true peaches. The taste differs slightly from true peaches; however it is outstanding too. Greater consumer acceptance is attained of the fruit of nectarines which are more acidic than peaches. The color of the skin is often red to purple-red. In recent years an important development of white-flesh nectarine cultivars has occurred.

The stones of peaches and nectarines are conspicuous by their surface, with different deep branch-like grooves and pits. These characteristic shape, size, grooves, and color of the stone are peculiar to every single cultivar.

Regulations

Commission regulation (EC) no. 2335/1999 lays down marketing standards for peaches and nectarines.

Definition of Produce

This standard applies to peaches and nectarines grown from cultivars of *Prunus persica* L., to be supplied fresh to the consumer; peaches and nectarines for industrial processing are excluded.

Provisions Concerning Quality

The purpose of this standard is to define the quality requirements for peaches and nectarines after preparation and packing.

Minimum requirements Peaches and nectarines must be:

> intact, sound, produce affected by rotting or deterioration such as to make it unfit for consumption is excluded, clean, practically free of any visible foreign matter, practically free from pests, practically free from damage caused by pests, free of all abnormal external moisture, free of foreign smell and/or taste

Peaches and nectarines must have been carefully hand-picked. The development and ripeness of peaches and nectarines must be such as to enable them to withstand transport and handling and to arrive in a satisfactory condition at the place of destination.

Classification

- Extra class
- Class I
- Class II

Table 1 Classification of peaches and nectarines according to size

Diameter	*Size (code)*	*Circumference*
90 mm and over	AAAA	28 cm and over
80 mm – under 90 mm	AAA	25 cm under 28 cm
73 mm – under 80 mm	AA	23 cm under 25 cm
67 mm – under 73 mm	A	21 cm under 23 cm
61 mm – under 67 mm	B	19 cm under 21 cm
56 mm – under 61 mm	C	17.5 cm under 19 cm
51 mm – under 56 mm	D	16 cm under 17.5 cm

Provisions Concerning Sizing

Sizing is determined by circumference or maximum diameter of the equatorial section (Table 1).

Diseases

Peaches and nectarines are subject to attack by a cosiderable number of diseases. Viruses and *Mycoplasma* are the most significant. We will discuss only a few diseases here

- Peach mosaic (vector – eriophyoid mite *Eriophyes insidiosus*)
- Peach yellows and little peach (vector – leafhopper *Macropsis trimaculata*)
- Phony (distributed through the root system)
- X-disease (vectors – more species of the leafhoppers)

These diseases have various symptoms on the leaves, twigs, and fruits. All these diseases are transmitted with buds by propagation, therefore healthy budwood and rootstocks are the most important control measures (together with the fight against vectors). Affected trees have nearly no crop and usually die within a short period.

The fungus diseases leaf curl, brown rot, and mildew are troublesome:

- Leaf curl (*Taphrina deformans*) curls and deforms the leaves early in the season; their color changes to dark red or violet. It must be prevented by applying fungicides just before bud opening.
- Peach mildew (*Sphaeroteca pannosa*) appears in the form of a white powdery substance upon the leaves and fruits. Spraying with fungicides will commonly hold it in check.
- Brown rot (*Monilia fructicola*). Infection begins during flowering but fruit rot often occurs after harvest. Control with fungicide during vegetation and then immediate cooling after harvest are important.
- Gray mold (*Bothrytis cinerea*) occurs especially after a wet spring during storage if the fruit has been infected during gathering. Gentle harvest, undamaged fruit, and temperature about 0 °C during storage and transport are effective control measures.

Harvesting of Fruit

In the 1930s the harvesting season of peaches was about 3 weeks – mid-July to mid-August. Due to intensive breeding work, especially in the USA (and in France and Italy), the harvesting season is now extended to at least 2 months; in the most favorable environmental conditions it may be more than 3 months (mid-May to 3 September). As regards ripening, the cultivar Redhaven grown worldwide is taken as the reference. Most peach cultivars ripen in the midseason 2–3 weeks after Redhaven; more nectarine cultivars ripen before Redhaven; the canning clingstone cultivars have a shorter harvesting period before and 6 weeks after Redhaven.

The harvest differs according to the purpose of using the fruit. For fresh fruit the fruit has to be harvested 2–3 days before it will reach ripeness, with a fully developed flavor. This is called tree-ripe. Therefore it is sometimes recommended to repeat harvesting at least three times according to advancing ripeness. The grade of ripeness can be determined by skin ground color changes from green to yellow in most cultivars (exept for the white-flesh cultivar). This is called minimum maturity, and is used to harvest fruits determined for distant markets and long transport. The ideal time to harvest peaches is when they are beginning their ripening process, but are still firm. At this stage they withstand picking operations and hauling quite well and complete the ripening process in good condition. If picked too soon, potential tonnage is lost because peaches and nectarines grow considerably just before reaching maturity. If the harvest delay is too long, they begin to soften, bruise easily, and shelf-life is reduced. To satisfy consumers that the fruit is ready to eat it is necessary to initiate the ripening process at the right time before the fruit is put on the market. Nowadays detailed ripening protocols for different peach and nectarine cultivars from important international export districts have been developed. For home gardens usually the best time is when the yellow-fleshed cultivars change from yellowish-green to yellow or orange-yellow. White-fleshed cultivars are ready for harvest when the color changes to cream white.

The nearly ripe fruits of the best quality should be gathered and placed in the market boxes that are provided with special niches for each individual fruit to prevent bruising during transport. Rapid transport should take place in refrigerated trucks.

Cultivars destined for processing are usually picked all at once – true peaches by hand and clings with the help of fruit-shaking machines.

Postharvest Operations

Peaches and nectarines are transported from orchard to packinghouse and cooler as soon as possible after harvest. In the packinghouse the fruits of minimum maturity are sorted to eliminate fruits with visual defects, and to select outstanding nice colored and big fruits (for the extra class). Cleaning machinery to remove fuzz from peaches is often used in all major production areas. In the defuzzing operation the peaches are scrubbed with rotating brushes to remove the fuzz and then washed and rinsed with fresh cold water. Sizing segregates fruits by size or weight. It is very important to cool the fruit as quickly as possible to near 0 °C. Also longer chilling at 0 °C is recommend. When the fruits are stored at a higher temperature, between about 2.2 °C and 7.5 °C, chilling injury (internal breakdown) occurs. First the flavor is lost, then the flesh becomes mealy, wooly, or harder, and brown around the pit of the stone. Cultivars of peach and nectarine differ in their susceptibility to chilling injury; nectarines are less susceptible.

Sometimes during harvest or after 1–2 days, black spots or stripes occur on the skin. This is a result of abrasion combined with touch with metal – iron or copper. Therefore it is very important to handle fruit gently during harvest and postharvest hauling operations.

Supplying Fresh Fruit to Consumers

The fruits must be intact, sound, clean, practically free of any visible foreign matter, practically free from pests, free of abnormal external moisture and free of any foreign smell or taste. The fruits must have been carefully picked and sufficiently developed that the ripening process can continue in order to reach the degree of maturity proper to their cultivar characteristic, to withstand transport and handling, and to arrive at their destinations in satisfactory conditions. Usually they are classified to four or seven classes according to the diameter of the widest part of the peach and nectarine.

Use of fresh fruits

Peaches – and, to a certain extent, nectarines – belong to the most sought-after fruits, thanks to their delicious taste and the development of global trade. Most of the peach crop (usually about 60%; in some regions much more) is marketed as fresh fruit; for nectarines, nearly all of the crop is used fresh. Cultivars of true freestone peaches with white and orange-yellow melting flesh, thin skin with not too dense fuzz, and delicious taste, fully ripened are the best for eating fresh. On the market such cultivars are sold in four to five classes. Sliced fresh peaches or nectarines are eaten alone or in combination with other fruit in salads or cocktails. Canned or frozen fruit can be used in the same way. Dried fruits are usually eaten as an energy-giving snack. In homes and confectioneries peaches and nectarines can be used for baked goods, icecream, gelatine desserts, pies, juices, and sherbets.

Some sensitive individuals can experience difficulties caused by a thicker fuzzy skin of certain peach cultivars. This unpleasant characteristic can be solved by a defuzzing operation (with rotating brushes) or by developing large nectarines, as was done in the last decades, especially in the USA.

Processing

The main method of processing is canning, usually about 30% of the whole crop, while a small amount is dried in some regions. The fruits are further used for jams, jellies, juices, brandy and liqueurs, preserves, and other miscellaneous products.

Technological Requirements

Peaches are inspected, graded, halved, pitted, and peeled before canning, freezing, and drying. The shape of the fruit should be round or flattened round and almost uniform; the diameter should be at least 45 mm. The flesh should be soft but firm enough (not farinaceous), of yellow-orange color, and have a taste full of the aroma. The smaller the stones, the better – 8–9% of total fruit weight. Tannin content should not be more than 0.08%. True peaches and clings are used in the processing industry. Cling cultivars are better, as they have firm flesh from the skin to the stone. Therefore after pitting and peeling the flesh is equally firm and does not crumble. The cling cultivars withstand better mechanical harvesting with shaking machines, longer transport, and storage. A further advantage of clings is equal ripening and slower overripening than true peaches. For processing the optimal grade of ripeness has to be measured just before processing starts. Prematurely harvested and processed fruit is too hard and lacks intense taste and flavor. Peaches are usually canned in light to heavy syrup, less so in water or juicy pack. When freezing the syrup, a small amount of ascorbic acid (to prevent browning) is used. Peeled peaches must be sulfured for drying.

Recommended Cultivars

Nowadays grown cultivars have different origins. In earlier times cultivars from the Persian (Iran) region were a base for many old cultivars. But in the middle of the nineteenth century some Chinese clings were imported to America directly from China. Some from North China (with hard winters) required high chilling and some from South China required less chilling. From these Chinese imports accidental hybrids with cultivars grown in that time in the USA have been chosen. One such hybrid was introduced in the 1870s as the cultivar Elberta. It had become one of the most widespread cultivars in a short time and a parent of many new cultivars.

Some examples of worldwide cultivars (2000 AD) are:

- true peaches – freestone and semifreestone – yellow: Redhaven, O'Henry, Spring Lady, Rich Lady, Cresthaven, Maycrest, Springcrest, Elegant Lady, Royal Glory, Fayette; white: Maria Bianca, White Lady
- clingstones: Babygold 5, 7, Vivian, Harbinger, Andross, Loadel, Helford 2, Golden Queen
- nectarines – yellow: Redgold, Fantasia, Flamekist, Supercrimson, Venus, Armking, May Grand, Marycrest; white – Queen Giant, Silverking.

In many districts these has been a shift from middle and late cultivars to early-maturing ones. This resulted from the strong demand for the first peaches and nectarines harvested. This is reflected in the high prices of early peaches and nectarines. In recent years there has been an improvement in the size and color of early peach and nectarine cultivars. Many have firm, white flesh with a delicious flavor and taste. In the typical great peach district a peach cultivar now usually lasts less than 20 years before being replaced by new superior one. Elberta's (1870) longevity is a distinguishing feature. North American cultivars became popular in European districts (Italy, France, Spain, Hungary, and Yugoslavia) and also in South America and North Africa. New cultivars were selected that required less chilling in the USA around the year 1970. This allowed an important increase in peach growing in very warm regions – India, Peru, Thailand, Brasil, Israel, not just in the USA. In the last decade many new cultivars have been developed in all main peach- and nectarine-growing regions – for instance, in the USA about 250, France about 55, Italy about 90, China about 35, Mexico about 20, and in the whole world more than 500 new cultivars.

Table 2 Chemical composition of peaches and nectarines

	$g\,kg^{-1}$
Water	805.0–877.0
Protein	7.0–8.0
Total lipid	0.9–2.0
Sugar	70.0–150.0
Acids	6.0–12.0
Dietary fiber	14.0–20.0
Ash	4.0–5.0

Sugar is composed on average from 13% glucose, 15% fructose, and 72% saccharose. Organic acids fluctuate according to grade of ripeness: 42% citric acid and 58% malic acid.

Nutritional Value

The composition of peaches and nectarines is given in Tables 2, 3 and 4. When ripe, most modern cultivars of peaches and nectarines exhibit a well-balanced proportion of acids and sugars. The content of dietary fiber and mineral salts, principally potassium, is important. Vitamin content is relatively low. It is influenced by the cultivar, how it is grown, the environment, and ripeness grade. There is also a rich content of delicious aromatic substances present. All this contributes to the great popularity of peaches and nectarines.

Production

Worldwide production of peaches and nectarines is increasing. This trend is mainly based on a large

Table 3 Mineral salt content of peaches and nectarines

Mineral salts	$mg\,kg^{-1}$
Calcium	50.0–70.0
Iron	1.1–12.0
Magnesium	70.0–100.0
Phosphorus	120.0–300.0
Potassium	1970.0–2050.0
Sodium	0.0–30.0
Zinc	0.2–1.4
Copper	0.5–0.7
Manganese	0.4–1.1
Selenium	Traces

Table 4 Vitamin content of peaches and nectarines

Vitamins	
Beta-carotene	$0.58–4.40\,mg\,kg^{-1}$
Vitamin E	$7.00\,mg\,kg^{-1}$
Vitamin B_1	$0.17–0.50\,mg\,kg^{-1}$
Vitamin B_2	$0.40–0.53\,mg\,kg^{-1}$
Vitamin B_6	$0.18–0.30\,mg\,kg^{-1}$
Niacin	$6.00–9.90\,mg\,kg^{-1}$
Pantothenic acid	$1.40–1.70\,mg\,kg^{-1}$
Folate	$30.00–34.00\,\mu g\,kg^{-1}$
Vitamin C	$66.00–370.00\,mg\,kg^{-1}$

The content of beta-carotene (provitamin A) depends on the color of the cultivar (white, yellow, orange-yellow, red).

number of new more suitable cultivars and better cultivation methods (fertilization, irrigation, prevention of diseases and pests). These developments have led to the possibility of exploiting moderately favorable environmental conditions. Worldwide production increased over 20 years by more than 50% and in 1996–98 reached a volume of 11 146 million tonnes. The USA produced 1302 million tonnes, and Europe 3418 million tonnes, of which the most important are Italy 1447 million tonnes, Spain 894 million tonnes, Greece 648 million tonnes, and France 468 million tonnes. High production volumes have been registered in China: 2922 million tonnes, Turkey: 372 million tonnes, Argentina: 284 million tonnes, Chile: 278 million tonnes, South Africa 221 million tonnes, and Tunisia 221 million tonnes. The development of international trade was also a great contributory factor: it solved problems of rapid transport without losses, mainly due to standardization and refrigeration all the way from harvest to consumer.

See also: **Canning**: Principles; Cans and their Manufacture; **Cherries**; **Chilled Storage**: Principles; **Controlled-atmosphere Storage**: Effects on Fruit and Vegetables; **Fruits of Temperate Climates**: Commercial and Dietary Importance; Factors Affecting Quality; Improvement and Maintenance of Fruit Germplasm; **Plums and Related Fruits**; **Preservation of Food**; **Ripening of Fruit**

Further Reading

Blattný C (1986) *Koservárenské Suroviny*. Prague: SNTL.

Commission Regulation (EC) no. 2335 (1999) Laying down marketing standards for peaches and nectarines. *Official Journal of the European Communities* L 281/11: 4.11.

Crisosto CH and Kader AA (2002) Peach – postharvest quality maintenance guidelines. http://postharvest.ucdavis.edu/produce/producefacts/.

Crisosto CH and Kader AA (2002) Nectarines – postharvest quality maintenance guidelines. http://postharvest.ucdavis.edu/produce/producefacts/.

Das Grosse Lebensmittel Lexikon, 3rd edn. (1985) Innsbruck: Pinguin Verlag.

Der Kleine Souci Fachmann Kraut, 2nd edn. (1991) *Lebensmitteltabelle fuer die Praxis*, 2nd edn. Stuttgart: Wissenschaftliche Verlagsgeselchaft.

Ensminger AH, Ensminger ME, Konlande JE and Robsen JRK (1995) *The Consise Encyclopedia of Foods and Nutrition*. Boca Raton, FL: CRC Press.

Faust M and Timon B (1995) Origin and dissemination of peach. In: Janiek J (ed.) *Horticultural Reviews*, vol. 17. John Wiley.

Fideghelli C, Della Strada G, Grassi F and Morico G (1998) *The peach industry in the world: present situation and trend. Proceedings of the Fourth International Peach Symposium. Acta Horticulture* 465:

Hubbard EE, Purcell JC and Ott SL (1986) *Variety Changes in the Georgia Peach Industry, 1957–1984*. University of Georgia: Agricultural Experiment Station.

Kutina J *et al.* (1991) *Pomologický Atlas 1,2*. Prague: Zeměděl. Nakl. Brázda.

Kyzlink V (1990) *Principles of Food Preservation*. Amsterdam: Elsevier.

Mareček F *et al.* (1994) *Zahradnický Slovník Naučný*. Prague: ÚZPI.

Stehlík V *et al.* (1969) *Naučný Slovník Zemědělský 1,2*. Prague: UVTI-SZN.

PEANUTS

T H Sanders, North Carolina State University, Raleigh, NC, USA

Origin

Arachis hypogaea L., commonly known as peanut, groundnut, monkey nut, goober, or earth nut because the seed develop underground, is in the division Papiolionaceae of the family Leguminosae. The peanut is only one of a few hundred species of legumes that produces flowers above ground but develops the fruit below ground. Peanuts are native to South America and were cultivated in pre-Columbian native societies of Peru as early as 3000 BC. Peanuts probably originated in the region of eastern South America, where a large number of species are found growing wild. A Bolivian origin is suggested by the wide range in seed and pod morphology documented there. It has been suggested that *A. hypogaea* originated from a hybrid between *A. cardenasii* (nn) and *A. batizocoi* (K. & G.), as both parents occurred in reasonable proximity in Bolivia. Peanuts were widely distributed throughout South and Central America and the Caribbean region during the time of Columbus. Peanuts were probably brought to West Africa from Brazil in the sixteenth century and then to the African east coast and to India. Peanuts from widely

separated regions of the world were brought to Africa and it is regarded as a center of genetic diversity.

Production

Worldwide production figures for peanuts are estimated at slightly less than 30×10^6 t and peanuts are produced on a significant basis in more than 30 different countries. In recent years, India, China, and the USA, the three largest producers, accounted for approximately 65% of the world production. Production figures for the USA are relatively stable at about 1.6×10^6 metric tons. However, production from China and India have been about 10.9×10^6 t and 7.4×10^6 t, although recent estimates from India at 7.4×10^6 t are probably much higher than current production. Although cultivated for many centuries, the economic importance of groundnuts and groundnut oil has increased rapidly only in the past century.

Yield per hectare is highly variable, depending on environment, variety, disease, and other factors. Important increases in peanut yield occurred between 1960 and 1990 and these increases are due to improved technology and management of factors which reduce yield. Peanuts are a high-value crop and an important source of revenue to producers throughout the world, and especially in less developed countries. The exportation of peanuts is mainly from the USA, China, and Argentina, while less developed countries utilize peanuts produced within the country, and the majority of these are processed for oil. The percentage of peanuts crushed for oil is a function of the use of peanuts as edible products and the need for readily available, cheap cooking oil. On a world basis, more than 50% of the peanuts produced are crushed for oil. In India, 75–80% of the peanuts produced are crushed for oil, while in the USA only approximately 12% of recent production is crushed.

World production of peanut oil is estimated at 4.65 million metric tons (Table 1). Extraction methods (hydraulic press, expeller, and/or solvent extractor) may differ widely due to technological advances in some countries; however, residue may contain 1–7% oil and with inefficient equipment the percentage may be much higher.

Peanut Plant

The leaves of peanuts are alternate and pinnate, with three to four leaflets on long petioles (Figure 1). The leaflets, which are obovate and softly hairy, are normally about 3–5 cm long. Peanuts have a distinct main stem and variable numbers of lateral branches. There are two major types of peanuts based on growth habit of the plant – upright with erect central stem and spreading with numerous laterals. The upright type is suitable for mechanical production, and the spreading type is grown more often in less developed farming operations. The root system of peanuts is often expansive and may be found up to a depth of 2 m. Several strains of *Rhizobium* nitrogen-fixing bacteria are found in abundant nodules on main and lateral roots.

Flower, Fruit, and Seed

Peanuts have yellow flowers on axillary branches which are small (about 12 mm long). Depending on genotype, environment, and temperature, flowering starts at about 25 days after emergence and the most

Table 1 Area, yield, and production of peanuts: average for 1995–2000

Country	*Area (million hectares)*	*Yield (tons per hectare)*	*Production (million metric tons)*
India	7.78	0.95	7.39
China	3.89	2.80	10.90
USA	0.58	2.87	1.67
Senegal	0.73	0.90	0.66
Sudan	0.55	0.67	0.37
Brazil	0.09	1.77	0.16
Argentina	0.29	1.49	0.43
South Africa	0.09	1.57	0.15
Other	6.93	0.98	6.77
Total	20.94	1.36	28.51

Source: US Department of Agriculture, Foreign Agricultural Service, Cotton, Oilseeds, Tobacco, and Seeds Division. October 2001.

Figure 1 Peanut plant.

prolific flowering occurs between 5 and 11 weeks after planting (Figure 2).

Under the best of conditions, fewer than 20% of the flowers produce mature fruit and with drought or heat stress the percentage may decrease to fewer than 15%. Genotypes which produce most flowers during the first 2 weeks of the flowering period produce greater numbers of pods. Peanuts are generally self-pollinating, although a small amount of cross-pollination may occur. About 7–10 days after fertilization, the receptacle thickens and elongates into a long carpophore bearing the fertilized ovules to the ground. The carpophore is commonly known as the peg, and the action of burying the immature pod is known as pegging. Growth of the peg is positively geotropic until it enters the soil – approximately 5–7 cm. The tip then turns horizontal and begins to develop into a fruit, commonly referred to as a pod, containing from two to four seeds and ranging in size up to 8.0×2.7 cm. The enlargement of the pod proceeds from base to apex and reaches maximum size about 3 weeks after the penetration of the peg into the soil. As the pod matures, the inner face of the shell becomes increasingly darker in color, associated with an increased tannin content, and becomes very dark brown on maturation. Seeds show large variations in size, shape and color, all of which are fairly stable for any given cultivar. Seed size, together with the seed mass, has been used extensively in agronomic classification of peanuts. Larger seed types are generally preferred for confectionery purposes; mostly small and medium-size seeds are used to produce oil.

Peanut Composition

Peanut composition is greatly affected by variety, environment, and maturity. The interaction of these factors with harvest date and handling determines the specific composition of an individual lot. Average percentages for peanut components are presented in Table 2.

Figure 2 Number of flowers of peanut plant as a function of time.

Oil

Peanut oil content is considered to average about 50% and the triacylglycerol content is generally in the range of 95%. Peanut oil is composed of mixed glycerides and contains a high proportion of unsaturated fatty acids, in particular, oleic (18:1) and linoleic (18:2). Additional fatty acids are palmitic (16:0), stearic (18:0), arachidic (20:0), 11-eicosenoic (20:1), behenic (22:0), and lignoceric (24:0). A comparison of fatty acid content with some other vegetable oils (Table 3) shows that peanuts are considered to have generally undetectable levels of linolenic acid. Peanuts contain higher levels of oleic acid than corn and soybean but lower levels than olive oil. However, peanut lines with high oleic acid concentration have recently become commercially available. The fatty acid profile from these peanuts is similar to that of olive oil.

Several years ago, research at the University of Florida examined approximately 500 peanut lines for fatty acid distribution and identified two lines with oleic-to-linoleic ratios of about 35. Using classical breeding techniques, commercial peanut varieties have been developed incorporating the high oleic acid trait. The developed lines do not have meaningful differences in oil content, flavor, color, or texture. Oxidative stability comparisons of high

Table 2 Typical composition of peanut kernels

Constituent	*Range (%)*	*Average (%)*
Moisture	3.9–13.2	5.0
Protein	21.0–36.4	28.5
Lipids	35.8–54.2	47.5
Crude fiber	1.2–4.3	2.8
Nitrogen-free extract	6.0–24.9	13.3
Ash	1.8–3.1	2.9
Reducing sugars	0.1–0.3	0.2
Disaccharide sugar	1.9–5.2	4.5
Starch	1.0–5.3	4.0
Pentosans	2.2–2.7	2.5

Table 3 Fatty acid profiles of peanut, olive, corn, and soybean oils

Fatty acid	16	18	18:1	18:2	18:3	20	20:1	22	24
Peanut	11.1	2.4	46.7	32.0		1.3	1.6	2.9	1.5
Olive	9.0	2.7	80.3	6.3	0.7	0.4			
Corn	10.9	2.0	25.4	59.6	1.2	0.4			
Soybean	10.6	4.0	23.2	53.7	7.6				

oleic peanut oil versus normal peanut oil were made using extracted, neutralized, and bleached oil from isogenic peanut lines varying only in fatty acid composition. Oil from the high oleic line contained 76.3% oleic and 4.7% linoleic compared to 56.6% oleic and 24.2% linoleic in the normal peanut oil. Oxidative stability was found to be as high as 14.5 times greater for high oleic peanut oil depending on the method of measurement. Details of peanut oil composition and characteristics are provided in the section on vegetable oils.

Extraction of Peanut Oil

Peanut oil is held in cells in an extremely fine emulsion. Damage to cells results in the release of oil. Peanut oil is usually extracted by hydraulic press, expeller, and solvent extractor. These may be used separately or in combination. The oil in the residue from oil recovery by the three methods is about 7, 5, and 1%, respectively. The residue (press cake) from the hydraulic press and expeller is generally used for livestock feed, and that from solvent oil extractors may be finely ground into flour for human consumption, if the nuts were of good quality.

Proteins The total protein content of whole seed peanuts ranges between 22 and 30%. Total protein can be separated into four major fractions: albumins, arachin, nonarachin, or conarachin. The arachin and conarachin fractions represent about 63% and 33% of the total, respectively. The arachin fraction is rich in threonine and proline and poor in lysine and methionine, while the conarachin fraction is poor in phylalanine and tyrosine.

The arachin and conarachin protein fractions are complex in composition and structure. Gel electrophoresis of sodium dodecyl sulfate dissociated proteins showed that these two fractions each contained five different components having molecular weights between 20 and 84 kDa. Two-dimensional polyacrylamide gel electrophoresis of seed polypeptides resulted in the detection of at least 74 major and between 100 and 125 minor peptides among various peanut cultivars and breeding lines. Variation existed among the major polypeptides of the cultivars, as noted by isoelectric points between pH 4.4 and 8.0 and molecular weights between 16 and 75 kDa. There is genetic variability in the mechanism for peanut protein synthesis, indicating that there is a potential for the development of peanut cultivars possessing nutritionally desirable proteins by manipulating protein synthesis.

Nutritional value of peanut protein depends on the amount and kind of essential amino acids available during digestion (Table 4). Amino acid availability in peanuts and peanut products is enhanced by the processing they undergo before consumption. Antinutritive factors, including trypsin inhibitors, hemagglutinins, goitrogens, saponins, and phytic acid are present in raw peanuts, and heat treatment in blanching and roasting of peanuts destroys the trypsin inhibitor and hemagglutinin. Peanuts are generally considered to be deficient in lysine and methionine, thus their relative nutritive value is decreased in comparison with a reference protein. The protein efficiency ratio for peanut protein is generally considered to be 1.5–1.8 and these values are related to the limited availability of lysine and methionine.

Carbohydrates The carbohydrate content of peanut kernels is usually determined by difference from the proximate analysis. It has been found to range from 10 to 20% of the fresh weight. Defatted peanut meal contains about 38% total carbohydrate. The major components are starch, sucrose, and nonstarch polysaccharides. Defatted meal contains about 18% sugars.

The major sugars in peanut are sucrose, stachyose, raffinose, fructose, and glucose. Inositol, verbascose, arabinose, and galactose have also been reported in various studies. The total sugar content has been found to vary with genotype, maturity, growing conditions, storage conditions, and kernel size. Large kernels usually contain less sugar. Cooler growing conditions were found to produce more sugars and sugar content decreases with maturity. Sucrose is the predominant sugar, with the content ranging

Table 4 Amino acid composition of peanuts

Amino acid	*% Total protein*[a]
Aspartic acid	11.89
Threonine[b]	2.43
Serine	5.11
Glutamic acid	19.83
Proline	4.44
Cystine	0.81
Glycine	6.59
Alanine	4.12
Valine[b]	3.14
Methionine[b]	0.51
Isoleucine[b]	3.40
Leucine[b]	5.21
Tyrosine	4.09
Phenylalanine[b]	6.38
Histidine[b]	3.17
Lysine[b]	3.87
Arginine[b]	10.81

[a]Data represent means from several sources.
[b]Essential amino acids.

from 2.7 to 3.5%. Stachyose is the next abundant, at levels of 1.5–2.5%. Other values found are 0.02–0.06% for raffinose and 0.01–0.04% for inositol. Fructose and glucose are usually present at less than 0.02%.

The sugar content is important for peanut flavor. Free sugars are noted to be precursors to roasted peanut flavors. When roasted, there is loss of some sugar content, with the greatest losses being sustained by fructose and glucose.

Vitamins and minerals The vitamin content of peanut seed is shown in Table 5. Peanuts have little or no vitamin A activity. Three forms of tocopherol (α, γ, β) have been found, with γ-tocopherol the highest and β-tocopherol the lowest. Peanuts are considered to be a good source of vitamin E, niacin, and folate.

Peanuts contain several of the minerals essential for human growth and development (Table 6). Recommended daily intake percentages of the quantities present are significant and constitute a good source of minerals.

Table 5 Essential vitamin composition of peanuts

	Amount[a]	*%RDI*
Vitamin E	3.3 IU	11
Folate	41 µg	10
Niacin	3.8 mg	19
Thiamin (B_1)	0.12 mg	8
Riboflavin (B_2)	0.03 mg	2
Vitamin B_6	0.07 mg	4

[a]Based on 1 oz serving of dry roasted peanuts.
RDI, recommended dietary intake.

Table 6 Composition of some minerals in peanuts

	Amount (mg)[a]	*% RDI*
Zinc	0.94	6
Copper	0.2	10
Potassium	186.5	6
Calcium	15.3	2
Iron	0.6	3
Phosphorus	101.5	10
Magnesium	50	13

[a]Based on 1 oz serving of dry roasted peanuts.
RDI, recommended dietary intake.
Source: US Department of Agriculture, Agricultural Research Service, 1998. USDA Nutrient Database for Standard Reference, Release 12.

Production and Handling Practices that Affect Quality

Maturity

Peanut maturation is extremely important in changes of biochemical components related to quality and shelf-life potential. Those facts have been demonstrated collectively in sensory studies to evaluate the relationship of maturity and flavor quality. Mature peanuts tend to have increased intensity of positive flavor descriptors, such as roasted peanutty and sweet aromatic, while immature peanuts are more commonly associated with negative terms, such as painty and fruity fermented. The term 'painty' is related to the common term 'rancid' and indicates the propensity of oil components in more immature peanuts to undergo enzymatic or autooxidation changes that result in decreased shelf-life. The degree of dark roast color is related to maturity level, with immature peanuts roasting darker because of higher free sugars and free amino acids. The relationship of high curing temperature to decreased sensory acceptability in immature peanuts has been demonstrated. Further, decreased sensory acceptability caused by high intensities of the off-flavor term 'fruity fermented' in sized lots was related to darker roast colored immature peanuts in the lot.

Maturity in peanuts is a complex concept because peanuts of all market types are generally marketed in various commercial seed sizes based on thickness, or on count per unit weight. The complexity arises in that all mature peanuts are not large and all immature peanuts are not small. Although size and maturity are positively related, the relationship is not absolute, and peanuts of very different maturity are found consistently in all commercial sizes. This fact is confounded even more in lots that are marketed on a count-per-weight basis wherein large and small seed may be mixed together to give the same count as seed of consistent uniform size. Peanut seed size distribution is affected by the environment in predictable ways. Soil temperature increases tend to shift the mean seed size down while cooler temperatures shift the mean size upward. Irrigation, harvest date, and soil temperature treatments generally produced significant differences among percentages of individual maturity classes in each commercial size. The distributions of maturity classes within commercial sizes are sufficiently different to suggest that flavor, roast color, shelf-life, and other quality factors would be affected in final roast products. Published data indicate that there are sensory differences in medium-grade-size peanuts harvested on a weekly basis over a 6-week period. In these studies, the percentage of immature peanuts in the seed sized lots progressively decreased with time. Roasted peanuts from earlier harvests had lower roast peanut impact and more bitter taste. Although shelf-life studies were not

conducted, the increase in percentage of mature peanuts in progressive harvests predisposes those lots to longer shelf-life.

Environment

The fact that growth, production, and handling practices affect quality, flavor, and shelf-life of peanuts and peanut products is implicit in data from numerous studies, years of manufacturing use, and the wide ranges of flavor and shelf-life observed in those countless situations. However, although studies report specific biochemical constituents and the effect of certain practices and conditions on changes in those compounds, the total effect on final product shelf-life resulting from all these processes must be extrapolated using established quantitative and qualitative relationships of specific constituents to shelf-life potential.

The role of environment on peanut yield has been consistently documented, but the effect on peanut quality factors potentially related to flavor and shelf-life has received little recent research attention. Higher soil temperatures tend to result in smaller seed size distributions, but studies on biochemical comparisons among years do not relate well to specific environmental conditions. Drought and temperature stress have been shown to affect some of the protein factors of peanuts. Soil temperature increases result in decreased concentrations of the free carbohydrates fructose, glucose, sucrose, raffinose, and stachyose in peanuts.

Harvest

When peanut plants are removed from the soil, the peanuts begin to dry, and the final natural processes related to potential flavor and shelf-life are initiated. When peanuts are harvested into inverted windrows, temperatures inside pods near the ground may reach 40–50 °C when air temperatures are no higher than 32 °C. Conditions that slow the drying process with associated high moisture may lead to fungus growth with resulting high free fatty acid values due to microbial lipase activity. Peanuts affected in this way are usually visibly damaged and may be removed electronically as part of the shelling process, or in less developed countries they may be removed by hand picking. Although off-flavors may not normally be thought of as factors that reduce shelf-life, usually the physical, physiological, and biochemical mechanisms that contribute to off-flavor development so disrupt the normal processes that shorter shelf-life should be anticipated. Obviously, processors will not knowingly utilize off-flavor peanuts in formulations, but knowledge or stipulation of growing and handling practices for purchased peanuts may contribute to the production and delivery of more stable products by the processor.

Maturity, as previously discussed, is a primary factor in total quality of peanuts, and the level of total crop maturity at harvest influences many of the handling and storage factors that follow. Moisture content decreases as peanuts mature on the plant, and at harvest, a wide range of moisture percentages are present. The range has variably been ascribed as 20–70% moisture at the time of digging. Given this, and the degree of biochemical and physical development of the various peanuts, it is readily discernible that differences in flavor and shelf-life are still being affected. Peanuts normally air-dry for some period of time before they are harvested. In mechanical harvest situations, peanuts usually dry to an average of 20–25% moisture before being picked and further dried.

The physical process of machine harvest has potential effects on the shelf-life of peanuts due to the force used to pick and transport peanuts into the machine. The physical damage in mechanical harvest, noted as cracked or broken pods, must inherently damage seed integrity, although relationships to quality reduction have not been investigated. Peanuts that are field-dried are subject to various environmental conditions which may affect quality. Harvested peanuts with a high percentage of immature pods will be difficult to dry to an overall safe storage level. High-moisture peanuts going into storage provide an increased opportunity for decrease in shelf-life potential.

Storage

Shelf-life potential may be maintained in storage of in-shell and shelled peanuts. In-shell peanuts are stored in bulk situations in many parts of the world, often at approximately 10% moisture content. In these situations, care must be given to ventilation and moisture control, because as the stored peanuts cool in response to ambient conditions that are usually cool, they dry further. The moisture that is lost must not be allowed to concentrate into the bulk lot because of obvious quality and microbial considerations. Adequate conditions of storage, especially ventilation, are critical to maintaining low levels of free fatty acids and total carbonyl compounds. Because of the removal of the shell, which is a normal protective barrier, shelled peanuts are somewhat more susceptible to quality/shelf-life deterioration. Shelled peanuts may be held in conditions of *c*. 0.5–5 °C and 55–70% relative humidity for periods of a year or more without quality loss.

Moisture content should be in the 7% range and slightly lower levels may impart additional protection. In peanuts stored at *c*. 6 and 9% moisture, the

higher moisture content resulted in reduced flavor quality and changes in amino acids, carbohydrates, and the phospholipid fraction of oil.

Consumption/Uses of Peanuts

Peanut Oil

Peanut oil is used as a cooking oil, especially in deep-fat frying; peanut oil is excellent since it has a smoke point of 229.4 °C. Degradation which may occur during frying results in the increase of free fatty acids and a decrease in smoke point. Crude peanut oil has a bland, slightly beany, nut-like flavor which is removed during refining to produce an oil that is odorless. Peanut oil develops few off-flavors or odors in use as a frying oil. Peanut oil is sometimes used in salad oil and in pourable dressings because of the length of time solids are held in suspension in the oil, although it solidifies from 0 to 3 °C. Refined peanut oil is widely used to prepare pastries, shortening (hydrogenation), oleomargarine, mayonnaise, salad dressings, and other food products. It is superior to oils from corn, soybean, cotton seed, olive, or safflower for making salad dressings to be stored below −12.2 °C.

Roasted Peanuts

Peanut roasting is a common practice in much of the world since roasting brings out the unique flavor. In western countries, particularly the USA, roasted and salted peanuts are one of the popular food items, produced on a commercial scale under a highly mechanized system with excellent quality control and packaging methods. In many European countries, peanuts are roasted and then coated with various flavors. A large portion of the US peanut crop is processed into roasted product for use in peanut butter, salted peanuts, and confections. Although the salted, roasted form predominates, oil-cooked, salted, unsalted, and low-calorie forms such as whole or mixed nuts are also popular.

A typical protocol for dry-roasted peanuts is exposure to 160 °C for 20–30 min. Seed coats are generally removed before applying oil and salt. In oil-roasting, blanched peanuts are immersed for 3–5 min in heated oil. After roasting the peanuts are then subjected to air cooling and salting with about 1.8–2.2% salt. Studies conducted to determine the effect of high-oleic-oil roasting indicated slight increases in shelf-life of normal oleic peanuts. Since peanut seed coats contain phenolics, including tannins, and the germs contain bitter saponin compounds, these are often removed before consumption. The current methods of blanching eliminate the skins and germs so that blanching always precedes oil-roasting, but often follows dry-heat roasting.

Peanuts have a high oil content and defatted peanuts have been developed to meet a specific consumer market. Partially defatted peanuts are obtained by hydraulic pressing to remove about 50% oil, followed by restoration of pressed kernels to their original size and shape by immersion in boiling water containing salt, species, or flavorings. Kernels are then dried and roasted. The pressure applied in the hydraulic press, the holding time, and the initial depth of peanuts in the compression influence the reconstitution of peanuts.

Peanut Butter

Although not a product accepted worldwide, peanut butter and products made from it are the most consumed form of peanuts in the USA, and notable consumption occurs in Canada and The Netherlands. Marketing introductions of peanut butter are still in progress in many European countries. Peanut butter is made by the fine grinding of dry roasted peanuts. Salt, hydrogenated oil, and sugar are usually added. Peanut butter may be made from any variety of peanuts but the blend of two parts Spanish or runner peanuts with one part Virginia peanuts is considered best for the most desirable consistency. Peanut quality, roasting techniques, manipulation of temperature, and grinding result in a wide range of colors, flavors, and consistencies of peanut butters on the market.

After grinding, oil separates from the solid particles unless a crystalline lattice is formed. Since crystals of fat are not present in the natural oil of peanut butter at ordinary temperatures, hydrogenated peanut oil, which is crystalline at these temperatures, is mixed uniformly with the product to provide such crystals. Generally, the incorporation is accomplished above the melting point of the hydrogenated peanut oil to insure more complete dispersion of the hydrogenated oil. In a satisfactorily stabilized peanut butter, crystallization of the added fat (hydrogenated peanut oil) takes place generally throughout the product before the oil–meal mixture can separate. *Trans*-fats infrequently associated with hydrogenated oil have recently been examined in commercial peanut butters and the work demonstrated unmeasurable quantities using current technology.

Confections

Peanuts (split, granulated, whole or ground peanut paste) are used in a wide range of other products,

including confections such as candy bars, dragées, hard-coated peanuts, peanut butter cups, peanut brittle, and a variety of other specialty applications. Shelf-life is important for confectionary products and is greatly affected by the product configuration, moisture content, and storage temperature.

Aflatoxin

Aflatoxin is a potentially carcinogenic compound produced by *Aspergillus flavus* and *A. parasiticus* which can invade peanuts, corn, cottonseed, and other commodities. In technologically advanced countries, control programs and testing programs are in place which significantly reduce the concentrations of aflatoxin over those found in developing countries. Aflatoxin contamination is usually found to be the most serious when drought conditions are present during the last few weeks of growth and pod development. Preharvest contamination is a significant economic problem for the peanut industry. Preharvest aflatoxin contamination in the crop is a problem when drought-type conditions predispose peanuts to insect damage by lesser corn stalk borer and other insects which are known carriers of fungus spores. Invasion of pods and infestation of seed often result in visible damage, and high levels of aflatoxin are characteristically found in damaged kernels. Preharvest invasion can occur with no obvious kernel damage and, although extensive studies have been conducted, the exact mechanism of fungal invasion is not yet known. Aflatoxin contamination is often highest in small, immature seed. Studies have demonstrated a relationship between the water activity-related physiological ability of the seed to produce phytoalexins and the water activity at which the fungus can best invade the seed. As the ability of the seed to produce phytoalexins decreases, there appears to be a point at which the fungus can grow in the seed before decreasing water activity terminates growth of the fungus.

Although generally less frequent, postharvest contamination may be a problem, especially when moisture accumulation and proper temperatures are present in harvest, transport, or storage.

Once the pods are field-dried and stripped from plants, they must be dried to moisture content less than 9% prior to storage. Peanuts stored at moisture content greater than 10% have increased risk for growth of *A. flavus*. Once dried to less than 9% moisture, peanuts can be stored for long periods with little loss of quality if kept under controlled storage conditions. Control of aflatoxin is best addressed during production by the appropriate irrigation of the crop through harvest time. In most contaminated lots, only a relatively small number of kernels is contaminated, and in many cases physical separation is the ideal method. This is based on the assumption that contaminated kernels are either discolored or shriveled. Methods include manual sorting, mechanical sorters, or electronic scanning, which examines each kernel separately and accepts or rejects it on the basis of reflectance. Action levels of aflatoxin in peanuts and other commodities are set and rigorously enforced within producing countries, and import regulations carefully control levels which may be imported.

Peanut Allergy

In recent years, concern for food allergies in general has increased, and concern for peanut allergy is no exception. For unknown reasons, peanut allergy is associated with a higher incidence of fatal food-induced anaphylaxis than any other food allergy. Immediate hypersensitivity to foods occurs in 6–8% of children and about 1% of adults. In the USA, a recent survey suggested that 0.7% of children are allergic to peanuts in varying degrees. Avoidance is the only current method of dealing with food allergy. Research potentials in dealing with peanut allergy will be discussed later.

Five peanut allergens have been identified by different research groups. Two major allergens, Ara h 1 (a vicilin) and Ara h 2 (a conglutin) are recognized by serum immunoglobulin E (IgE) of greater than 90% of peanut allergic individuals. Ara h 3 is recognized by serum IgE of about 40–45% of allergic individuals and is thus considered to be a minor allergen. Ara h 5 (a profilin) is also a minor allergen; IgE is recognized by about 13% of allergic individuals. Refined peanut oil (heat-processed) is not allergenic; however, oils contaminated with peanut protein may indeed produce significant allergic reactions in peanut-sensitive individuals. Cold-pressed oils are more likely to contain peanut proteins than hot-pressed oils.

Current research on peanut allergy is directed at testing an approved asthma drug, anti-IgE monoclonal antibody, against peanut allergy. Development and testing of a vaccine against peanut allergy are progressing towards human trials. A third area of research involves identification of enzymes involved in mediation of the allergic reaction and development of a test kit for individuals (especially children) to determine potential for peanut allergy.

See also: **Aflatoxins**; **Amino Acids**: Properties and Occurrence; **Food Intolerance**: Food Allergies; **Sweets and Candies**: Sugar Confectionery; **Vegetable Oils**: Composition and Analysis

PEARS

C Blattný, Prague, Czech Republic

Introduction

The genus of pear *Pyrus* has about 20 species. Cultivars of the common pear (mainly European and American) belong to *Pyrus communis*. The Asian cultivars belong to the sand pear *Pyrus serotina* = *P. pyrifolia*. The pear is a member of the rose family (Rosaceae).

Origin and History

The common pear came from Central Asia. It was well known and popular in ancient Greece. In the year 287 BC Theofrast, in his *Historia Plantarum*, described four cultivars of pears. The Romans took over the cultivation of pears from the Greeks. In the time of Vergil (50 years BC), Cato described 35 cultivars. The number of cultivars increased very quickly and by the end of the Middle Ages over 200 cultivars were described. But most of the outstanding dessert cultivars that are now widespread were selected out in the eighteenth and nineteenth centuries in France and in Belgium. Most of these old pear cultivars arose out of the selection of seedlings that grew as a consequence of random cross-breeding. Intentional or deliberate breeding led to new cultivars during the twentieth century, but a number of excellent cultivars from the eighteenth and nineteenth century have not been replaced. By the end of the twentieth century it is roughly estimated that 3000 cultivars of the common pear had come into being.

In China the cultivation of sand pear cultivars was known some centuries before Christ. Nowadays, many cultivars of the sand pear are widespread in East Asia. By the end of the twentieth century this sand pear had spread not only to North America, but also in smaller amounts to other continents. This species is also used for hybridization with the common pear, especially in the USA, as hybrids are less susceptible to the dangerous infectious bacterial disease fireblight.

Requirements of the Pear Tree

The pear tree, depending on the species and cultivar, has moderate to high demands of temperature (in winter as well as in summer) and length of vegetation period. Winter cultivars need warmer regions with a longer vegetation period to ripen properly and to attain better quality. Some of the cultivars do not tolerate extensive winter frost, e.g., Bartlett (Williams Christ). Fairly rich, warm, loamy soil, of sufficient depth and with a well-drained subsoil are suitable for the pear. Of all temperate climate fruit species, pear quality is the most influenced by soil properties and weather.

The trees that grow naturally from seeds (or grafts on pear seedlings) reach a height of up to 9 m with trunks of 30 cm or more in diameter. The cultivars of pears can only reproduce by being grafted on a rootstock – usually a pear seedling. In order to create smaller or dwarf trees the quince can be used as the rootstock. Most widespread vegetatively propagated clones of the quince are: quince A, quince C, from East Malling Research Station in England, French Provence Quince BA 29 C, and Le Page series C. In the USA a new hybrid Old Home × Farmingdale 333 (OH × F333) is now also recommended. Certain cultivars like Bartlett (Williams Christ) and Bosc are incompatible with the quince and require an interstem, e.g., cultivar Hardy or Old Home (used in the USA).

Many pear cultivars are self-sterile. They need the pollen of other cultivars to develop fruitfully. Some cultivars have relatively often parthenocarpic fruit (without kernels). Those fruits are usually longer and slimmer.

The cultivars of the Japanese-type Asian pear will pollinate each other. Chinese types need another cultivar for pollination, and when they bloom with the cultivars of the common pear, these cultivars could be very good pollinators for them.

As regards the sand pear's requirements of the environment, these are similar to the common pear, but it seems that suitable districts for peaches will also be best for most sand pear cultivars. *Pyrus betulaefolia* is recommended as standard rootstock for sand pear cultivars. The hybrid OH × F333 is also suitable.

The Fruit

The pear is a typical pome fruit closely related to the apple. The basic shape is pear-shaped and ranges according to the cultivar from bell-shaped, to egg-shaped, conical, globular and teardrop-shaped (Figure 1). Further differences are in size, color, texture, flavor, taste, ripening season, and possibilities for processing and cooling.

The flesh is usually firmer than that of apples; when fully ripe it softens and finally it melts. After

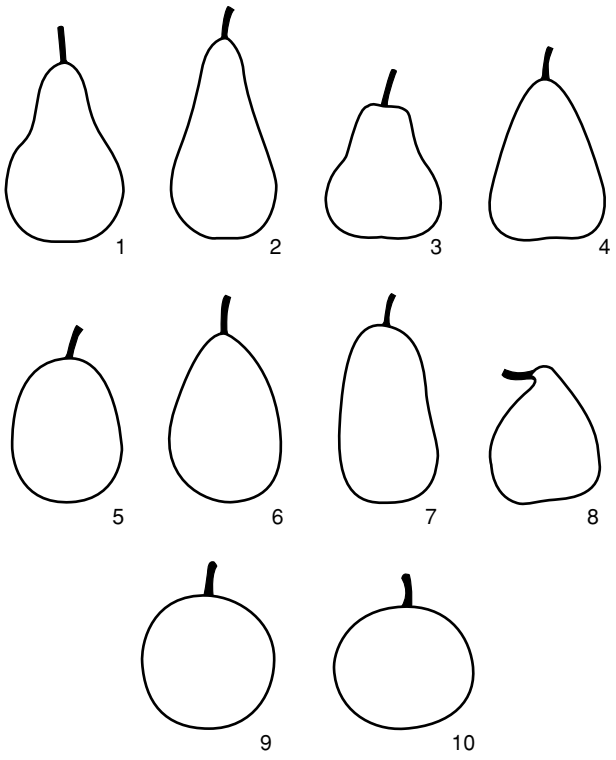

Figure 1 Different shapes of pear. 1, pear-shaped; 2, bottle-shaped; 3, bell-shaped; 4, conical; 5, egg-shaped; 6, teardrop-shaped; 7, cylinder-shaped; 8, fig-shaped; 9, globe-shaped; 10, flat globe-shaped

overripening the flesh quickly melts due to hydrolysis of pectin and total disorganization of enzymes. At optimal grades of ripeness, the fruit of the best dessert cultivars melts like butter. The skin is smooth and its color varies from yellow, yellow green, to brown (maybe russet) and also almost red. Only two cultivars are bright dark red: Max Red Bartlett and Red Anjou. The flavor and taste of cultivars differ greatly – some are spicy like muscat. There is another conspicuous difference between pears and apples. In the center of a pear around the core and along the vessels from the stem to the core, the gritty cells called stone cells are present. The more gentle the cultivar, the fewer small stone cells are present. However, very small stone cells in a very small number are always present in the best dessert cultivars, e.g. Bartlett (Williams Christ). The number of stone cells is not only influenced genetically by also by the quality of the soil, especially for winter cultivars. Most modern dessert cultivars of the common pear have large heavy fruit: medium weight 150–180 g, heavy weight 200–400 g, giant weight up to 750 g.

Asian pears are also called Oriental pears, Chinese pears, and Japanese pears. The Asian pear cultivar differs in its flesh texture – when eaten ripe they remain pleasantly firm, crisp, and juicy like apples but with a different and distinctive texture. They do not change texture after being picked and stored like common pears. Most Asian pears are sweet, flavorful, and low-acid. They are round or flat fruit with yellow or green skin, with bronze-colored and russeting skin or they are pear-shaped with either smooth or russeted skin.

Diseases and Pests

Diseases

Some virus and *Mycoplasma* diseases are relatively widespread. We will only discuss a few here. Ring pattern mosaic gives striking light-green or greenish-yellow rings and a line pattern in the leaves. Yellow veins and red mottle are typical with chlorotic banding of the tertiary and final veins. Both diseases suppress growth in a number of pear cultivars. The disease stony pit occurs wherever pears are grown. The fruits of diseased trees become pitted and deformed with many gritty cells in the tissue near pitty areas. Fruits with many pits are woody and difficult to cut. Either only a few or all of the fruits of diseased trees may develop symptoms, depending on the susceptibility of the cultivar. An affected fruit may have one or a number of pits and is not suitable for eating. The content of dry matter, ash, potassium, calcium, and mallic acid is higher in diseased fruits then in comparable healthy ones. The content of monosaccharides is less. The affected fruits cannot be used to produce sherry and brandy. Very sensitive cultivars are Bosc, Doyenné du Comice, while Bartlett in the USA is relatively resistant.

Fireblight, a bacterial disease, is the greatest problem limiting production of pears in the USA. Fireblight seldom affects Europe. Leaves and young terminal shoots infected by bacteria *Erwinia amylovora* wilt, die, and blacken. Further infection of branches and trunk results in the development of cankers and finally the death of the tree. Recommendations for decreasing the danger of fireblight are:

1. Select tolerant, better resistant cultivars. In the USA recommended hybrids between common and Asian pears include: Moonglow, Kieffer, Seckel, Waite, and Maxim.
2. Control of the vector – green apple aphids (*Aphis pomi*).
3. Avoid any practice that encourages excessive and soft terminal growth.
4. Prune out and burn the infected portion of the tree (better in winter) and avoid overfertilization.

Pear scab is a disease caused by the fungus *Venturia pirina*. The infection starts at blossom time and occurs on the blossoms, more so on the leaves than on the fruits. Infection of young pear fruits may cause the pear to be lop-sided and sometimes cracked. Then the scab spots spread on the whole fruit, especially on that of the susceptible cultivar. After very late infection, small scabs occur which then develop into so-called storage scabs. Spraying during infection time (May–July) is one of the best control measures.

During longer storage some pathological and physiological disorders can occur:

- blue mold rot (*Penicillium expansum*) – after mechanical injury
- gray mold rot (*Botrytis cinerea*) – infection during blossoming or through wounds of mature fruits
- senescent scald – dark brown skin discoloration
- superficial scald – diffuse brown skin discoloration
- freezing injury – translucent, water-soaked appearance of tissue
- bitter pit – brown, corky lesions in the flesh
- watery breakdown – rapid enzymatic flesh breakdown
- carbon dioxide injury – browning of interior walls of carpels and adjacent core tissue
- low-oxygen injury – core browning

Control measures

- effective preharvest disease control procedures
- careful handling during harvesting and postharvest preparation for market to minimize mechanical injuries
- prompt cooling to 0 °C
- maintenance of proper temperature (−1 °C to 0 °C) during storage and transport
- use of postharvest chemical treatments
- use of the controlled atmosphere

Pests

Codling moth (*Laspeyresia pomonella*) larvae feed on the fruits to the core and enlarge an exit hole which they plug with frass. They can have two to three generations and they are the most dangerous pests to apples and sometimes can be pernicious to pears. Spraying when females lay their eggs (June, July) is a good control measure.

Pear midge (*Contarinia pyrivora*) can cause considerable damage to young fruits. The attacked fruitlets become deformed and enlarged, and are later black in the center and full of white maggots, and they crack, decay, and fall down in a short time. They hibernate in the soil. Spraying just before the blossoming can decrease the damage.

Pear leaf midge (*Psylla pyricola*) in some districts is an important pest, resembling a miniature cicade. The psyllas suck on the leaves which margins roll. Maggots produce a honeydew that collects in droplets on the leaves, branches, and the fruit. It is an excellent medium for sooty molds. The crop is usually lower and unsellebled. Another danger to pear psylla is in its transmission of *Mycoplasma*.

Harvesting

The pear has to be gathered from the tree before full ripeness, usually when the basic green color changes to yellowish green. It is very important to determine the correct time for the crop. Prematurely harvested fruits do not ripen but they stay hard permanently and become leathery. Therefore the best time for gathering is when the fruits attain normal size, which is optimal for the cultivar, at the time when the stem parts readily from the branch if the fruit is slightly lifted. Then it is important to get the fruit quickly to cool storage. Overripe fruits soften so quickly that they are nearly useless.

To harvest fruits of Asian pears one must hold them lightly in the palm and pluck them using an upward twisting motion to remove the fruit from the spur. A pulling motion can result in damage as the stalk may be removed from the fruit. The skin of Asian pears is much more susceptible to abrasion and friction marks than common pears. Smooth surface containers such as polystyrene trace should be used to gather fruit. Fruit should be placed into trays with the stem end upward, preferably in single layers.

Chill Storage

Only a very small number of common pear cultivars can endure longer preservation in chill storage (Bosc, Anjou, Doyenné du Comice) if they do make it to the fresh market. These cultivars can be stored in controlled atmosphere (1–2% O_2 + 0–1% CO_2) at 1 °C for up to 4 months (Bosc and Doyenné du Comice) or 6 months (Anjou) while maintaining their capacity to ripen and attain good flavor and texture. Some winter cultivars can be stored in cooler places for some months until spring, but their quality is much lower (Diel, Comtesse de Paris, Nelis). For chilling it is necessary to have optimal temperature −1–0 °C, optimal relative humidity 90–95%, and for ripening, optimal conditions are 15–22 °C and 90–95% relative humidity. This basic schedule for ripening must be introduced after each longer cool storage of all pear cultivars to help the development of flavor and optimal texture of the flesh.

Some of the Asian pear cultivars, e.g., Shinseiki, Hosui, Kosui, Nijisseiki, have an adequate storage life. If fruits are harvested at the recommended maturity, a storage life of 12–20 weeks and subsequent shelf-life of 10–15 days can be expected, depending on the cultivar.

Important World Cultivars

- Common pear
- Summer – Bartlett (= Williams Christ), Dr J. Guyot, Clapp's Favourite, Santa Maria, Max Red Bartlett
- Autumn – Doyenné du Comice, Bosc, Conference, Fondante de Charneuses, Elisabeth, Luise Bonne, Boussock, Hardy, Lucas
- Winter – Diel, Comtesse de Paris, Anjou, Red Anjou, A. Lucas, Passe Crassane, Nelis
- Sand (Asian) pear (recommended in the USA)
- Early – Shinseaki
- August – Hosui, Kosui, Nijisseiki (Twentieth Century), Yo Inashi
- Late – Shinko, Olympic (Korean Giant)
- Sand (Asian) pear (widespread in China)
- Ya Li, Tse Li

Supplying Fresh Fruit to the Consumer

The fruit must be intact, sound, clean, free of any visible foreign matter, free from pests, free of abnormal external moisture, and free from any foreign smell or taste. The fruit must have been carefully picked and sufficiently developed so that the ripening process can continue in order to reach the degree of maturity appropriate to the cultivar, to withstand transport and handling, and to arrive at the destination in a satisfactory condition. Usually fruit is classified into three or four classes according the diameter of the widest part of the pear fruit.

Use of Fresh Fruit

The time to ripeness for consumption varies according to the cultivar: summer cultivars – up to 2 weeks; autumn cultivars – 2–8 weeks; winter cultivars up to 16 weeks. It is greatly influenced by the grade of ripeness during gathering and by storage temperature.

Dessert cultivars of the common pear, when eaten at the right stage of ripeness (the flesh should melt like butter) are delicious fruits. The fruits are used in various cookies, fruit salads, desserts, and gelatine dishes.

A small warning: it should be mentioned that some sensitive people can experience flatulence after eating fruits of the common pear, especially fruits which are not fully ripened.

The Asian pear cultivars are refreshing, very juicy, crispy like apples and, as with common pears, they go well with salads.

Processing

Pears are less often used for processing, compared to apples, due to their relatively low acid content and, above all, to a rapid ripening process which leads to the already-mentioned melting disintegration and liquefaction of the fruit. Selected pear cultivars are outstanding raw material for canning. They must, however, fulfill strict conditions:

1. The shape of the fruits is one of the decisive determinants of industrial preparation, since a suitable shape facilitates basic processing – peeling, halving, and removal of the core. The most suitable is a longish pear form with a wider stem part (for instance, the parthenocarpic fruits of the Bartlett (Williams Christ)). Cultivars that are too round have their core situated too high, which causes difficulties for machine coring. Fruits that are bottle-shaped, that crack, and that are too small are unsuitable.
2. The flesh should be white or faintly yellowish.
3. The skin should be firm, not too thick, clearly demarcated from the flesh, and easily peelable.
4. For factory processing, pears must be of the prescribed size and shape. Fruits of a uniform size and shape reach most uniformly technologically required ripeness. This degree is attained when the basic green color of the skin changes to yellow. Accurate appraisal of the degree of ripeness heralded by the changing color of different cultivars calls for an experienced eye. Many manufacturers use penetrometric methods to measure the firmness of the flesh, i.e., the pressure needed to push a special needle to a certain depth of the fruit. In areas of great production of cultivars destined for canning, for instance the summer Bartlett (Williams Christ) it is possible to employ long-term storage at temperatures -1 to $0\,^{\circ}C$ for 9–12 weeks, or at temperatures $+1\,^{\circ}C$ for 6–8 weeks to extend the processing season. The relative humidity must be 90–95%. The ripening process must follow this chill storage. Similar conditions can be also used for further canning of suitable cultivars like Clapps Favourite, Elisabeth, or Kieffer. The fruits are canned in syrup. To prevent browning of the flesh it is recommended to plunge the peeled halves in a solution of ascorbic or citric acid.

- Sweetening – the same requirements as for canning are applied.
- Drying of pears is relatively widespread in some regions. Pears suitable for drying should be of medium size (3 cm length), typically pear-shaped, slimmer in the stem part, the skin thinner, and the flesh with a small number of stone cells. Dried fruits should have a cinnamon color. Pears should be fully ripe with the highest content of sugar. They are dried whole, halved, or divided into small pieces, with the skin on or peeled.
- Pears are not frequently frozen. The requirements are the same as for canning.
- Brandy – due to the delicious flavor of some cultivars, e.g. Bartlett (Williams Christ), pear brandy is the mildest of the fruit brandies.
- Perry is a famous light alcoholic drink (like light wine) made from special perry cultivars. This drink was produced and beloved in Roman times. It is a drink of choice particularly in Normandy, the UK, Austria, Switzerland, and the USA. The special perry cultivars are highly fertile, with small fruit, and they should have a high content of sugar and acids and a rich content of tannins. Unfortunately, the production of those cultivars declined strikingly during the twentieth century.

Production

The pear is a popular fruit in temperate regions. Their production over the last 10 years in individual countries has developed differently. Worldwide it is increasing – in 1996–98 it reached 14 034 million tonnes, while in 1989–91 it was only 9529 million tonnes. It especially rose in China, by 245%, and amounted to 6401 million tonnes in 1996–98, that is, 45% of world output. The increase can also be observed in South America and Africa – in Chile by *c*. 500%, in Ecuador by 300%, in Argentina by 225%, in Morocco by 430%, in Algeria by 310%, in Tunisia by 210%, and in Egypt by 350%. Even in other warm regions production also increased, for instance, in Lebanon by 390% and in Turkey by 140%.

The changes are the result of the development of international trade which has facilitated refrigerated transport of high-quality dessert cultivars over long distances. Particularly profitable is the supply of the winter market either from southern countries of the northern hemisphere or from countries of the southern hemisphere. Among the largest exporters of pears we find now Argentina and Chile – not only Italy, France, and The Netherlands, as it was some years ago. The decrease in pear output in European countries is chiefly connected to the increase in the cultivation of apples and peaches to replace the pear.

Table 1 Chemical composition of pears

	$g\,kg^{-1}$
Water	775–869
Protein	4–5
Total lipid	3–4
Sugar	124–158
Acids	1–5
Dietary fiber	15–28
Ash	3–4

Sugar is composed on average from 54% fructose, 13% glucose, 15% saccharose, and 18% sorbit. Organic acids fluctuate according to the cultivar. In dessert cultivars malic acid, prevails citric acid is about 10% of

Table 2 Mineral salt content

Mineral salt	$mg\,kg^{-1}$
Calcium	100.0–160.0
Iron	2.0–3.0
Magnesium	60.0–80.0
Phosphorus	110.0–150.0
Potassium	1100.0–1300.0
Sodium	0.0–20.0
Zinc	1.2–2.3
Copper	0.9–1.1
Magnesium	0.4–0.8
Selenium	0.0–10.0

Table 3 Vitamin content

Vitamin	
Beta-carotene	0.15–0.30 mg kg^{-1}
Vitamin E	4.30–5.00 mg kg^{-1}
Vitamin B_1	0.20–0.50 mg kg^{-1}
Vitamin B_2	0.30–0.40 mg kg^{-1}
Vitamin B_6	0.15–0.18 mg kg^{-1}
Pantothenic acid	0.62–0.70 mg kg^{-1}
Folate	70.00–140.00 µg kg^{-1}
Vitamin C	40.00–50.00 mg kg^{-1}

Nutritional Value

The composition of the pear is given in Tables 1–3. Pears contain a good quantity of sugars (rapidly assimilated fructose, glucose, saccharose, and sorbit), as well as pectin, tannin, mineral salts, organic acids, some vitamins, and dietary fiber. There is also a rich content of delicious aromatic substances present. All this contributes to the great popularity of pears.

Further Reading

Bailey LH (1958) *The Standard Cyclopedia of Horticulture*, 17th edn. New York: MacMillan.

Blattný C (1986) *Konservárenské Suroviny*. Prague: SNTL.

Chittenden FJ (1956) *Dictionary of Gardening: the Royal Horticultural Society*, 2nd edn. Oxford: Clarendon Press.

Commission Regulation (EEC) no. 920/89 (1989) Laying down quality standards for carrots, citrus fruit and dessert apples and pears and amending Commission Regulation no. 58. *Official Journal of the European Communities* no. I.97/19.

Crisosto CH (2002) *Asian Pear – Postharvest Quality Maintenance Guidelines*. Dept. of Pomology, University of California, Davis, CA 95616. Available on line at: http: //postharvest.ucdavis.edu/producefacts.

Das Grosse Lebensmittel Lexikon (1985) 3rd edn. Innsbruck: Pinguin Verlag.

Dvořák A, Vondráček J, Kalášek J *et al*. (1978) *Atlas Odrůd Ovoce*. Prague: SZN.

Ensminger AH, Ensminger ME, Konlande JE and Robsen JRK (1995) *The Concise Encyclopedia of Food and Nutrition*. Boca Raton, FL: CRC Press.

Koch V, Ferkl FR, Blattný C *et al*. (1967) *Hrušky*. Prague: Academia.

Kutina J, Kalášek J, Blažek J *et al*. (1991) *Pomologický Atlas 1, 2*. Prague: Zeměděl. Nakl. Brázda.

Kyzlink V (1990) *Principles of Food Preservation*. Amsterdam: Elsevier.

Mareček F, Pekárková E, Blažek J *et al*. (1994) *Zahradnický Slovník Naučný*. Prague: ÚZPI.

Mitcham EJ, Crisosto C and Kader AA (2002) *Pear Anjou, Bosc and Comice. Recommendation for Maintaining Postharvest Quality*. Available on line at: http://postharvest.ucdavis.edu/producefact/fruit/pear2.html.

Senser and Scherz H (1991) *Der Kleine "Souci Fachmann Kraut". Lebensmitteltabelle fuer die Praxis*, 2nd edn. Stuttgart: Wissenchaftliche Verlagsgesellschaft.

Stehlík V, Trantírek J, Hruška L *et al*. (1969) *Naučný Slovník Zemědělský 1, 2*. Prague: ÚVTI-SZN.

PEAS AND LENTILS

G Grant and M Duncan, Rowett Research Institute, Bucksburn, Aberdeen, UK
R Alonso and F Marzo, The Public University of Navarra, Pamplona, Spain

Background

Peas and lentils are potentially rich sources of macro- and micronutrients and bioactive factors. This overview deals with the nutritional properties of the seeds, possible physiological responses (detrimental or beneficial) of animals to the constituent bioactive compounds, and the effects of processing on nutritional quality.

Peas (*Pisum* spp.) and lentils (*Lens* spp.) have been cultivated since around 6000–5000 BC, making them amongst the earliest of the developed legume crop plants. They are highly adaptable and can flourish in a wide variety of habitats and environmental conditions. As a result, they are grown in most regions of the world and provide a staple source of dietary protein and energy for a significant proportion of the population.

Much of the world output of peas and lentils continues to be produced by traditional agricultural methods, whereby they may be grown as part of a crop rotation system or as secondary crops. Large-scale intensive cultivation and industrial processing of these legumes as foodstuffs do, however, occur in some regions. Nonetheless, peas and lentils have been at a commercial disadvantage to seeds such as soyabean in that they do not contain significant amounts of oil. This has limited their commercial value. Protein and starch are the primary products of cultivation of these seeds, whereas with soyabean, they are essentially byproducts of vegetable oil production and therefore comparatively inexpensive. However, food production will need to increase greatly to meet the requirements of a rising world population. This is most likely to be met through use of food sources that grow well in a wide range of environments, in particular those that are adverse to soyabean production. Peas and lentils are good candidates for further development in this regard. Furthermore, since many legume seeds contain bioactive compounds that have been linked to health-promoting effects, peas and lentils may have significant health benefits in addition to being ready sources of protein and energy.

Composition

Although peas and lentils are highly adaptable, their actual composition varies greatly according to region, climate, cultivar, and husbandry. The carbohydrate levels may range between 525 and 740 g kg^{-1}, protein between 156 and 355 g kg^{-1}, and lipid between 6 and 61 g kg^{-1} (**Table 1**). Intensively grown peas and lentils have been developed and selected for yield and thus are less variable in composition. In general, these contain around 240–280 g protein per kilogram, 570–610 g of carbohydrate, about 100 g of lipid,

Table 1 Comparison of general composition and nutritional value of raw legume seeds

	Peas	*Lentils*	*Soyabeans*
N (g kg^{-1})	23–57	29–57	55–82
Protein (g kg^{-1})[a]	145–355	184–355	343–512
Carbohydrate (g kg^{-1})	599–700	500–697	247–370
Lipid (g kg^{-1})	15–62	15–25	181–227
Total fiber (g kg^{-1})	188–347	89–305	93–192
Crude fiber (g kg^{-1})	25–88	25–50	23–39
N digestibility (%)[b]	79–85	79–92	77–82
Biological value (%)	57–72	31–58	48–54
Net protein utilization (%)	45–61	22–46	38–44
Protein efficiency ratio	0.1–2.7	0.4–2.7	0.8–1.6

[a]N × 6.25.
[b]Protein digestibility and utilization by rats.
Data from Savage (1988), Savage and Deo (1989), Huisman and Jansman (1991) *Nutrition Abstracts and Reviews (Series B)* 61: 902–921, Lusas and Riaz (1995) *Journal of Nutrition* 125: 573S–580S, Cuadrado *et al*. (2002) *Journal of Agricultural and Food Chemistry* 50: 4371–4376, Grant *et al*. (1999), and Alonso *et al*. (2000) *Journal of Agricultural and Food Chemistry* 48: 2286–2290, Alonso *et al*. (2001) *Nutrition Research* 21: 1067–1077 .

Table 2 Comparison of the amino acid composition of cultivated raw legume seeds

	Peas	*Lentils*	*Soyabeans*	*Amino acid needs*	
				Human[a]	*Rat*
Alanine[b]	4.4–4.9	2.4–4.2	4.7		
Arginine	8.6–8.9	3.9–7.7	7.8		5.0
Aspartic acid	4.9–11.8	9.9–11.1	12.6		
Cystine	1.1–1.5	1.3–1.5	1.6		
Glutamic acid	12.2–17.1	15.5–16.3	19.4		
Glycine	4.4–4.8	4.1–4.4	4.6		
Histidine	2.2–2.4	1.3–2.8	2.7		2.5
Isoleucine	3.1–4.1	4.0–4.3	4.9	3.5	5.0
Leucine	7.2–8.3	7.0–7.2	8.1	6.5	8.0
Lysine	7.2–8.2	4.3–7.0	6.7	5.0	6.0
Methionine	1.0–1.1	0.8–1.1	1.3	2.5[c]	4.5[c]
Phenylalanine	4.6–5.3	4.9–5.5	5.2	6.5[d]	5.0
Proline	4.1–4.9	2.6–4.2	5.9		
Serine	4.4–4.9	2.9–4.6	5.8		
Threonine	3.6–3.9	2.5–3.6	4.3	2.5	4.0
Tryptophan	0.9–1.1	0.9	1.5	1.0	1.5
Tyrosine	2.9–3.9	1.1–2.5	3.8		4.0
Valine	4.7–5.5	3.0–5.0	5.0	3.5	5.5

[a]Amino acid as g per 100 g of protein.
[b]MIT pattern of human amino acid requirements. *Journal of Nutrition* 130. 1841S–1849S.
[c]Combined sulfur amino acids.
[d]Combined aromatic amino acids.
Data from Lusas and Riaz, (1995) *Journal of Nutrition* 125: 573S–580S, Urbano *et al*. (1995) *Journal of Agricultural and Food Chemistry* 43: 1871–1877, Alonso *et al*. (2000) *Journal of Agricultural and Food Chemistry* 48: 2286–2290, Alonso *et al*. (2001) *Nutrition Research* 21: 1067–1077 and USDA database www.nal.usda.gov.

and 250–300 g of total dietary fiber. The protein and lipid levels are far lower than those in raw soyabean but are similar to the amounts found in other cultivated legumes, such as kidney beans. Carbohydrate levels in peas and lentils are about double that in soyabean, and the dietary fiber content can be two to three times greater.

The levels of individual amino acids in peas and lentils also vary greatly according to region, climate, and cultivar. Again, those in intensively cultivated varieties are more consistent (Table 2). The average amino acid profile of peas is close to the MIT (Massachusetts Institute of Technology) recommended amino acid pattern for humans, but that of lentils may be slightly deficient in the sulfur amino acids and possibly tryptophan (Table 2). However, when compared with the requirements for rats, the most widely used experimental animal, both seeds are deficient in a number of amino acids. The sulfur amino acids are limiting, being present at 47–56% of the requirements of rats. (*See* **Amino Acids**: Properties and Occurrence.)

Pea proteins are readily extractable in neutral or alkaline aqueous conditions. They are mainly of the globulin type, with the 7S globulins (vicilin) tending to predominate over the 11S globulins (legumin). Lentil proteins are also mainly of the globulin type, with legumin being the more prominent form. However, lentil proteins tend to be poorly soluble under neutral or alkaline aqueous conditions if whole seed meal is extracted. In contrast, solubility is high if the lentils have been dehulled. This may be a result of the removal of tannins present in the hulls. The globulin proteins of both seeds generally have low contents of the sulfur amino acids. In contrast, the levels in albumin (water-soluble) fractions from these seeds are quite high, owing to the presence of sulfur amino acid-rich proteins such as the trypsin/chymotrypsin inhibitors.

Protein digestibility values for peas and lentils *in vivo* are around 79–84%. This is similar to the values obtained *in vitro* (80–84%) with activated pancreatic extracts. However, digestibility values *in vitro* are around 65–70% if purified digestive enzymes are used. This discrepancy *in vitro* is likely to be a result of slow or poor digestion of vicilins by the pure enzymes. A similar difference was noted with phaseolin from kidney bean. It was highly resistant to pure proteolytic enzymes but readily degraded when incubated with rat intestinal tissue plus contents. Enzymes or factors, in addition to pepsin, trypsin, and chymotrypsin, were shown to be necessary to facilitate digestion of this protein. The nature of these components is unknown. However, if additional factors are crucial to promote degradation of legume globulins, there is no guarantee that all animals will possess them. There could therefore be

considerable species differences in ability to digest legume globulins. (*See* **Protein**: Quality.)

Purified radiolabeled pea legumin and vicilin are digested by rats almost as effectively as casein (97.4, 97.1, and 98.5%, respectively). Equally, isolated lentil globulins are readily degraded (99%) by rats. The globulins do not therefore appear to be inherently resistant to proteolysis *in vivo*. Digestibility values obtained with raw meals or crude protein extracts are much lower. This is likely to be a result of interference with proteolysis by the seed matrix or seed components, such as protease inhibitors. This would not be a problem with protein isolates prepared from peas or lentils.

Utilization of Raw Legumes

Weight gains and net protein utilization values obtained with young rats fed raw peas or lentils as the sole source of dietary protein (approximately 400 g per kilogram of diet) are low. This is mainly a result of reduced food intake and increased fecal dry matter and N and urinary N excretion. Supplementation of diets with methionine improves food intake, *N* retention, and weight gain, but the overall nutritional quality remains below that of high-quality protein (Table 3).

Inclusion of raw peas at 150–450 g per kilogram of diet severely impairs food intake and growth by weanling pigs. The effects are much less marked in pigs over 5–6 weeks old, and supplementation of diets with tryptophan or methionine greatly improves performance. The maximal inclusion levels recommended are 100 g per kilogram of diet for weaner pigs, 200 g kg^{-1} for grower pigs and 350 g kg^{-1} for finishers. The growth of chicks is severely retarded by inclusion of up to 200 g of raw pea meal in their diet. Supplementation with methionine, cystine, and tryptophan improves performance but only to a limited extent. Older birds appear more tolerant of peas, and with laying hens, up to 330 g kg^{-1} can be incorporated into the diet without any significant adverse effects. Addition of supplementary methionine and lipid to the diet allows this incorporation to be increased to around 440 g per kilogram of diet. Fecal dry matter and N outputs by human subjects are increased by intake of 130 g of lentils per day, but urinary N output is slightly decreased.

Possible Health Benefits

Peas are hypocholesterolemic for rats, pigs, and humans. The mechanisms remain unclear but may be linked to fiber and saponins in the seeds. With rats, the effect may be partly a result of reduced lipid deposition by the animals. Lentils are hypocholesterolemic for rats but not apparently for pigs or humans, at least at the dietary inclusion levels studied.

Bioactive Factors

Peas and lentils contain a number of potentially bioactive components such as protease inhibitors and lectins that may influence body metabolism in a negative or positive manner (Table 4).

Trypsin/chymotrypsin Inhibitors

Peas and lentils contain Bowman–Birk-like trypsin/chymotrypsin inhibitors that inhibit the activity of these digestive enzymes *in vitro* and *in vivo*. These inhibitors are of a relatively low molecular weight (around 8–10 kDa) and are double-headed in that they inhibit two enzyme molecules simultaneously. The original Bowman–Birk inhibitors (BBI) were isolated from soyabean and were found to inhibit

Table 3 Nutritive value of peas or lentils for rats after heat treatment and/or L-methionine supplementation

	Peas			*Lentils*	
	Raw	*Cooked*	*Extruded*	*Raw*	*Cooked*
Unsupplemented					
N digestibility (%)	80–84	84–86	80–87	80–85	82–86
Net protein utilization (%)	50–60	44–54	60–66	27–33	36–40
Biological value (%)	60–75	55–65	75–83	32–40	43–47
Protein efficiency ratio	1.8–2.4	2.0–2.2	1.9–2.3	0.8–1.6	nd
Amino acid supplemented					
Biological value (%)	nd	nd	nd	53–61	69–71
Protein efficiency ratio	2.7–3.3	3.0–3.7	3.1–3.7	2.1–3.0	nd

nd, not determined.

Data from Savage (1988), Savage and Deo (1989, and references therein), and Alonso *et al*. (2000) *Journal of Agricultural and Food Chemistry* 48: 2286–2290, Alonso *et al*. (2001) *Nutrition Research* 21: 1067–1077.

Table 4 Bioactive factors present in raw legume seeds

	Peas	*Lentils*	*Soyabeans*
Trypsin inhibitor (g kg^{-1})[a]	1–8	0.1–4	16.7–27.2
Chymotrypsin inhibitor (g kg^{-1})[a]	0.5–6	0.1–2	0.2–4.9
α-Amylase inhibitor (g kg^{-1})[b]	< 0.1	< 0.1	< 0.1
Lectin (U × 10^{-6} kg^{-1})[c]	5.1–15.1	1.8–7.7	50–350
Lectin (g kg^{-1})[c]	0.3–0.7	~0.8	2–4
Phytic acid (g kg^{-1})	2.2–8.2	1.2–8.1	10.0–14.7
Tannin (g kg^{-1})	0.2–5.0	0.6–2.7	0.2–0.9
Phenolics (mg kg^{-1})	13–27	10–30	8–15
Saponin (g kg^{-1})	1.1–2.5	1.1–4.6	2.2–6.7

[a]Trypsin and chymotrypsin inhibitor is given as the number of grams of enzyme that would be inhibited per kilogram of meal.
[b]α-Amylase inhibitor is expressed as inhibitor equivalents using pure kidney bean α-amylase inhibitor as standard.
[c]Lectin levels are given as hemagglutinating units (U) per kilogram of meal based on the ability to agglutinate rabbit blood cells or as g kg^{-1} based on recoveries obtained during purification of lectin.
Data from Trowbridge (1974) *Journal of Biological Chemistry* 249: 6004–6012, Savage (1988), Savage and Deo (1989), Liener (1991) *Journal of the American Oil Chemists Society* 56: 121–129, Anderson and Wolf (1995) *Journal of Nutrition* 125: 581–588, Huisman and Jansman (1991) *Nutrition Abstracts and Reviews (Series B)* 61: 902–921, Alonso *et al*. (1998) *Food Chemistry* 63: 505–512, Alonso *et al*. (2000) *Journal of Agricultural and Food Chemistry* 48: 2286–2290, Alonso *et al*. (2001) *Nutrition Research* 21: 1067–1077, Cuadrado *et al*. (1999), Grant (1999), and Grant *et al*. (1999).

only trypsin and chymotrypsin in combination. However, in other legume seeds, variants capable of inhibiting two molecules of trypsin or chymotrypsin are also present. This heterogeneity in the reactivity profiles of BBI-like inhibitors makes it difficult to quantify the amounts present in the seeds or seed products. Therefore, for comparison purposes, the amount of each enzyme inhibited by a known amount of product is usually given. Alternatively, enzyme inhibitory units are calculated from the same data.

Trypsin and chymotrypsin inhibiting activity in peas and lentils varies greatly between cultivars (Table 4). Assuming that the mix of inhibitor variants leads to an average profile of one molecule trypsin and one molecule chymotrypsin inhibited per molecule inhibitor, available data suggest that peas contain 0.4–2.8 g of BBI-like inhibitor per kilogram and lentils 0.15–1.4 g kg^{-1}. This is comparable with the levels found in many other legumes. Soyabean has 3.2–3.7 g of BBI per kilogram. However, it also contains 4.5–9.0 g per kilogram of Kunitz trypsin inhibitors. Thus, the total trypsin inhibiting activity in soyabean is far higher than that in peas or lentils.

BBI are highly resistant to proteolysis *in vitro*. Nonetheless, native BBI and BBI-type inhibitors seem to be readily degraded *in vivo*, although ^{125}I-BBI seems to be highly resistant to proteolysis *in vivo*. Native inhibitors have limited effects on digestion and absorption of dietary protein and growth of animals, even when they are included at high levels (≤10 g kg^{-1}) in the diet. However, BBI or BBI-like inhibitors interfere with cholecystokinin-mediated control over pancreas function and trigger hypersecretion of pancreatic trypsin, chymotrypsin and α-amylase in rats, chicks, quails, and humans. In rats, chicks, and quails, this leads to pancreas enlargement, mainly as a result of hyperplasia and hypertrophy of the acinar cells and, in the long term, may lead to tissue dysfunction and disease. The effects of the inhibitors are, however, species-specific. Thus, they induce pancreas growth in young rats, mice, hamsters, guinea-pigs, and chickens but have little or no effect on the pancreas of young pigs, cattle, monkeys, or dogs.

Pancreatic enlargement is evident in rats fed pea diets, although the increase is usually much less than observed in rats given an equivalent intake of raw soyabean. Pancreatic growth was not apparent in rats fed lentil meal, although an inhibitor-enriched fraction of lentil meal did cause enlargement.

Consumption of soyabean BBI by experimental animals appears to significantly reduce the incidence and severity of liver, colon, and mammary cancers that develop as a result of treatment with chemical carcinogens or radiation. The mechanisms by which this occurs, however, remain unclear. BBI acting through localized effects on gut endocrine cells may induce the release of a number of hormones, growth factors, or peptides that interfere with tumor cell metabolism. Alternatively, bioactive fragments of BBI absorbed from the gut may be the main protective agents. The BBI-like inhibitors of field bean have also been shown to have cancer-preventing properties. It is thus possible that pea and lentil inhibitors will have a similar protective effect. (*See* **Trypsin Inhibitors**.)

α-Amylase Inhibitors

α-Amylase inhibitors inhibit the activity of salivary and pancreatic amylase *in vitro* and *in vivo*. They can impair the growth and metabolism of animals when given at high levels in the diet but may have beneficial uses in treatment of obesity or diabetes. They are generally present in very high amounts in *Phaseolus* species (kidney beans, 4.3 g of inhibitor per kilogram). However, the levels in peas and lentils are very low (Table 4) and unlikely to contribute significantly to the effects of these legumes on metabolism.

Lectins

Lectins are defined as carbohydrate-binding proteins/glycoproteins other than enzymes that are present in most plant materials. They are highly resistant to proteolytic degradation *in vivo* and survive passage

through the gastrointestinal tract. If appropriate carbohydrate receptors are present on gut epithelial cells, lectins bind to them and may be taken up systemically. As a result, lectins can potentially interfere with and modify many aspects of gut and systemic metabolism. Individual lectins vary greatly in their effects *in vivo*, but most species appear to be responsive to dietary lectins. (*See* **Hemagglutinins (Haemagglutinins)**.)

Lectins can be separated into eight general categories on the basis of their carbohydrate-binding specificity: complex, fucose, galactose, *N*-acetylglucosamine, mannose, mannose/glucose, mannose/maltose, and sialic acid. Some, such as soyabean agglutinin (galactose-specific), alter gut and pancreas metabolism in rats, in particular causing rapid growth of these tissues, without affecting systemic systems. A few, including kidney bean lectin (complex specificity), have additional effects on systemic hormone balance and lipid and muscle metabolism and can be very deleterious if consumed in high amounts. Others, such as pea and lentil lectins (glucose/mannose-specific) appear to have little or no effect on the body metabolism of rats.

The sensitivity of animals to lectins may, however, vary with species, age, period of exposure, gastrointestinal bacteria, diet composition, and dietary history. Thus, the glucose/mannose-specific jack bean lectin (Con A) has no effect on mature germ-free rats but has limited effects (causes small intestine and pancreas enlargement) in specific pathogen-free rats. It is, however, highly deleterious to rats carrying a salmonella infection, to suckling guinea-pigs, and to quails. The glycoconjugates expressed on the gut surface of very young animals differ greatly from those in mature counterparts; in particular, a high proportion of mannose residues are present. This is also evident in rats with a pathogen infection. In these circumstances, lectins that would not normally affect the gut may be able to bind to it and elicit changes in body metabolism.

The levels of lectins in peas and lentil are low compared with that in soyabean (Table 4) and very low by comparison with kidney bean (15–30 g kg^{-1}). Furthermore, in studies with mature rats, these lectins have no significant effects on metabolism. This would suggest that they are unlikely to cause problems. However, in view of the data with Con A and the number of factors that influence the sensitivity of an animal to lectin, one cannot exclude the possibility of specific circumstances where these dietary lectins have profound effects.

Young chicks do not do well when raw peas are added to their diet in low amounts but will tolerate quite high dietary inclusions if they are a few weeks old. During this period, the gut develops from its very immature form at hatching to its adult form. The gut may be susceptible to the action of pea lectin at the early stages of the maturation period.

Antigenic Proteins

Native 11S globulins (glycinin) and 7S globulins (conglycinin) of soyabean induce very adverse immune reactions in preruminant calves and newly weaned piglets, leading to gut damage, scouring, and poor performance. Pea globulins partially survive gut passage and can trigger some immune responses in preruminant calves. However, the degree to which this occurs is very much lower than that observed in soyabean-fed animals.

Lentils have been linked to allergy problems in a small number of pediatric patients. A number of possible allergens have been identified, including subunits of vicilin. The incidence of intolerance to pea proteins seems to be low.

Phytate

Phytic acid is often the main reserve of phosphorus in legumes. However, it also chelates with minerals and metals, such as calcium, magnesium, zinc, and iron, forming insoluble salts that are not readily absorbed by animals or humans. In particular, it can severely impair availability of zinc and iron. Phytate can also complex with proteins and may thereby reduce digestibility or enzyme activity.

Mineral uptake by pigs and chickens fed with soyabean-based diets is slightly impaired. Addition of phytase, a phytate-degrading enzyme, to the diet appears to counteract this effect, leading to an improvement in mineral uptake and better overall performance by the animals. However, the efficacy of this treatment can be very variable.

Phytic acid levels in peas and lentils are lower than those in kidney beans (11–17 g kg^{-1}) and soyabean (Table 4). None the less, they can still affect mineral metabolism, since iron absorption from a pea protein-based infant formula is significantly enhanced after enzymatic degradation of phytate.

Pea and lentil phytate can clearly have adverse effects on mineral uptake and body metabolism. However, in many cases, their impact is likely to be minimal because the mineral content of the diets is well above the requirements. Pea and lentil phytate, however, may cause significant problems if mineral intake, particularly of zinc and iron, is close to or below requirements.

Dietary phytic acid may have health-promoting properties. It can inhibit α-amylase, limit carbohydrate digestion, and lower blood glucose. There

are also indications that it is hypocholesterolemic and protective against colon cancer. However, the amounts required are quite high. It is unclear whether a normal physiological intake of peas or lentils would provide sufficient phytate to have a significant health-protective effect. (*See* **Phytic Acid**: Nutritional Impact.)

Tannins

Tannins are present in a wide array of plant crops. Legume condensed tannins are oligomers of variously substituted flavan-3-ols, and their antinutritional effects have recently been comprehensively reviewed. These compounds can reduce enzyme activity in the gut, impair gut morphology, lower nutrient (protein, carbohydrate and lipid) digestion and absorption, reduce mineral uptake, and greatly stimulate excretion of endogenous N. Thus, at high dietary intakes, they can adversely affect animal performance.

In contrast to the antinutritional effects, dietary flavan-3-ols have also been suggested to have important roles in disease prevention, particularly of cardiovascular diseases and some forms of cancer. They may act as antioxidant and free radical scavengers, inhibit tumor initiation and promotion, and have antibacterial and angioprotective properties.

The tannin contents of peas and lentils tend to be higher than that in soyabean (Table 4). However, the levels are comparable with those in many other legume seeds. Lentils contain moderate amounts of catechins and proanthocyanidin dimers and trimers, whereas pea samples tend to have low amounts. The compounds may have health-protective effects. However, since questions still remain as to how effectively these compounds or products derived from them are absorbed from the gut and distributed throughout the body, it is difficult to assess whether a normal intake of peas or lentils will provide sufficient flavanols to have a significant beneficial effect. (*See* **Tannins and Polyphenols**.)

Saponins

Triterpenoid saponins, found in leguminous plants, are composed of a triterpene aglycone linked to one, two, or three saccharide chains of varying size and complexity. The levels in peas and lentils are similar to, or slightly lower than, those in soyabean (Table 4) and other legume seeds. They may reduce weight gain if consumed at very high levels. However, at the levels present in peas, lentils, and soyabeans, they are considered to have no significant antinutritional effects. Saponins, however, may have beneficial effects. They have been found to be hypocholesterolemic in a number of species, owing in part to their ability to facilitate adsorption of bile acids to dietary fiber. Evidence of their efficacy in humans is not conclusive. It has also been suggested that saponins may have anticarcinogenic or antioxidant properties. (*See* **Saponins**.)

Fiber

Peas and lentils contain high levels of crude and total dietary fiber (Table 1). The fibers have a high binding capacity for bile acids *in vitro* and also promote fermentation in the hind-gut and production of butyrate *in vivo*. As with other dietary fibers, pea and lentil fibers are generally considered to be beneficial. However, the specific effects of these dietary fibers on metabolism remain unclear. Inclusion of pea or lentil fibers in diets appears to lower postprandial blood triglyceride levels without affecting cholesterol concentrations. However, other studies suggest that pea fibers can influence cholesterol levels in humans. Equally, studies show that pea fibers reduce or delay starch absorption and reduce the blood glucose peak, whereas others show no effect on blood glucose. This variability may be a result of differences in the test meals used or the nutritional/health status of the subjects. Pea fibers are used in the treatment of hypercholesterolemic patients. However, they are not used alone but given as part of a mixture of dietary fibers. (*See* **Dietary Fiber**: Properties and Sources.)

Starches

Legume starches in general tend to be less digestible or more slowly digested than corn starches. This slow release means that their glycemic index is low. They can thus be useful in diets for those with impaired carbohydrate tolerance. Lentils have been found to have beneficial effects on blood glucose profiles of diabetic patients. This is due, at least in part, to the slow release property of the constituent starch combined possibly with the effects of the fiber. (*See* **Starch**: Structure, Properties, and Determination.)

Vitamins and Minerals

Peas and lentils contain significant amounts of important vitamins and minerals. The levels of most are fairly similar in both seeds. However, lentils appear to contain higher levels of ascorbic acid, folic acid and vitamin B_6 than peas but have considerably less vitamin A. (*See* **Vitamins**: Overview.)

Effects of Processing

Various processing methods are routinely used in the preparation of peas and lentils for consumption. These include dehulling, soaking, germination, cooking, autoclaving, extrusion, and fermentation.

In general, dehulling or soaking seems to have limited effects on the levels of bioactive factors or the nutritional quality of the seeds. Dehulling of lentil increases the solubility of lentil proteins, and soaking slightly reduces trypsin inhibitor and phytate content. Germination lowers phytate and slightly reduces trypsin inhibitor, tannin and polyphenol levels in peas, and lowers trypsin inhibitor and phytate in lentil. Nonetheless, it may also increase levels of nonprotein amino acids. Natural fermentation also reduces the levels of bioactive compounds and slightly improves the starch and protein availability from legumes. The efficacy of this treatment is, however, extremely variable, partly because the environmental factors responsible are unknown.

Cooking, autoclaving, or extrusion treatment of peas fully inactivates trypsin inhibitors and lectins, has limited effects on phytate and polyphenols, but greatly reduces tannin concentrations. Starch digestibility is significantly increased, but protein digestibility is improved only to a limited extent (Table 3). The effects of cooking on lentils appear similar, with the exception that the reduction in tannins and polyphenols may be more pronounced than that observed with peas. Prolonged thermal treatment (cooking, autoclaving, or extrusion) can adversely affect nutrient availability from peas and lentils.

Thermal treatment of peas and lentils clearly improves the nutritional value of the constituent proteins (Table 3). With rats, this is evident only after supplementation of the diet with methionine to bring the sulfur amino acid content up to the requirements for rats. The human need for sulfur amino acids is much lower than that of rats and could be met from the levels present in peas and possibly lentils (Table 2). For humans, the biological value of cooked peas or lentils alone may thus be close to that found for the supplemented diets in rats.

Pea-based diets are hypocholesterolemic for rats. This property is retained even after extrusion treatment of the seeds. Some lentil proteins retain their antigenicity, even after cooking.

Digestibility of Heat-treated Globulins

The moderate digestibility of raw legume proteins has mainly been attributed to the presence of antinutritional factors in the seeds and the refractory nature of the proteins. However, although heat-treatment abolishes the activity of the main antinutritional factors, it appears to have remarkably little effect on the digestibility of pea or lentil proteins. In addition, pea and lentil globulins have a very high digestibility *in vivo*, once isolated from the seeds. The proteins are thus not inherently indigestible.

It is possible that interactions between nonprotein components of the seed matrix and the constituent proteins may be preventing digestion of a proportion of the proteins. Heating may not abolish these interactions. Alternatively, globulin-like proteins from some heat-treated legumes have been shown to be less digestible than counterparts from raw seeds. This may be due to the formation of large protein multimers that are not readily accessible to digestive enzymes. Any beneficial effects of heat treatment of seeds on protein availability may be negated, in part, by the development of these poorly digestible multimers of the globulin proteins. This merits further study.

Summary

Peas and lentils are potentially good sources of dietary protein and carbohydrate and micronutrients. They contain a range of bioactive compounds that may be detrimental to consumers. However, some of the bioactive compounds may also be beneficial for health. Thermal treatment greatly improves the nutritional quality of the seeds. Both raw and processed seeds have hypocholesterolemic properties.

See also: **Carbohydrates**: Interactions with Other Food Components; **Cholesterol**: Factors Determining Blood Cholesterol Levels; **Dietary Fiber**: Physiological Effects; Effects of Fiber on Absorption; **Extrusion Cooking**: Principles and Practice; Chemical and Nutritional Changes; **Food Intolerance**: Types; **Hemagglutinins (Haemagglutinins)**; **Legumes**: Legumes in the Diet; Dietary Importance; **Phytic Acid**: Properties and Determination; Nutritional Impact; **Plant Antinutritional Factors**: Detoxification; **Protein**: Interactions and Reactions Involved in Food Processing; **Trypsin Inhibitors**

Further Reading

Bardocz S, Hajos G and Pusztai A (eds) (1999) *Cost 98 – Effects of Antinutrients on the Nutritional Value of Legume Diets*, vol VI. Luxemburg: Office for Official Publications of the European Communities.

Carbonaro M, Grant G, Cappelloni M and Pusztai A (2000) Perspectives into factors limiting *in vivo* digestion of legume proteins: antinutritional compounds or storage proteins? *Journal of Agricultural and Food Chemistry* 48: 742–749.

Castell AG, Guenter W and Igbasan FA (1996) Nutritive value of peas for nonruminant diets. *Animal Feed Science Technology* 60: 209–227.

Caygill JC and Mueller-Harvey I (eds) (1999) *Secondary Plant Products. Antinutritional and Beneficial Actions in Animal Feeding*. Nottingham, UK: Nottingham University Press.

Jansman AJM, Hill GD, Huisman J and van der Poel AFB (eds) (1998) *Recent Advances of Research in Antinutritional Factors in Legume Seeds and Rape Seeds*. Wageningen, The Netherlands: Wageningen Pers.

Martinez San Ireneo M, Ibanez Sandin MD and Fernandez-Caldas E (2000) Hypersensitivity to members of the botanical order Fabales (legumes). *Journal of Investigational Allergology and Clinical Immunology* 10: 187–199.

Rowland I (1999) Optimal nutrition: fibre and phytochemicals. *Proceedings of the Nutrition Society* 58: 415–419.

Sanchez-Monge R, Pascual CY, Diaz-Perales A *et al*. (2000) Isolation and characterisation of relevant allergens from boiled lentils. *Journal of Allergy and Clinical Immunology* 106: 955–961.

Santos-Buelga C and Scalbert A (2000) Proanthocyanidins and tannin-like compounds. Nature, occurrence, dietary intake and effects on nutrition and health (review). *Journal of the Science of Food and Agriculture* 80: 1094–1117.

Savage GP (1988) The composition and nutritive value of lentils (*Lens culinaris*). *Nutrition Abstracts and Reviews (Series A)* 58: 319–343.

Savage GP and Deo S (1989) The nutritional value of peas (*Pisum sativum*). A literature review. *Nutrition Abstracts and Reviews (Series A)* 59: 65–88.

Urbano G, Lopez-Jurado M, Aranda P *et al*. (2000) The role of phytic acid in legumes: antinutrient or beneficial function? *Journal of Physiology and Biochemistry* 56: 283–294.

van der Poel AFB, Huisman J and Saini HS (eds) (1993) *Recent Advances of Research in Antinutritional Factors in Legume Seeds*. Wageningen, The Netherlands: Wageningen Pers.

Pecans *See* **Walnuts and Pecans**

PECTIN

Contents

Properties and Determination

L Flutto, Danisco, New Century, KS, USA

Introduction

Pectin is a high-molecular-weight carbohydrate polymer which is present in virtually all plants where it contributes to the cell structure. The term pectin covers a number of polymers which vary according to their molecular weight, chemical configuration, and content of neutral sugars, and different plant types produce pectin with different functional properties. The word 'pectin' comes from the Greek word *pektos* which means firm and hard, reflecting pectin's ability to form gels.

The gelling properties of pectin have been known for centuries, but the isolation of commercial pectin only started at the beginning of the twentieth century. In this document we highlight the chemistry, origin and production, and the functional properties of pectin.

Chemistry

The Homogalacturonic Acid Backbone

Pectin consists of a chain of galacturonic acid units which are linked by α-1,4 glycosidic bonds. The galacturonic acid chain is partly esterified as methyl esters. Pectin molecules can have a molecular weight of more than 200 000, corresponding to a degree of polymerization up to 1000 units (**Figure 1**).

Though the esters are the most significant components on the galacturonic acid backbone, other chemicals, such as acetyl, can be important in specific pectin types. Commercial pectin can also be partly

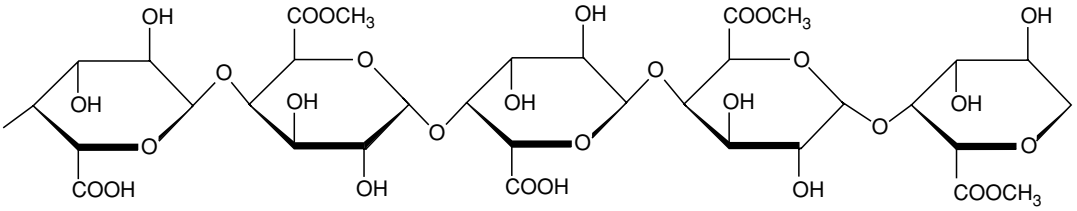

Figure 1 Pectin consists of long sequences of anhydro galacturonic acid completely or partly esterified with methanol.

Table 1 Classification of pectins

DE > 50%	DE < 50% Low-ester pectin (LE)	
High-ester pectin (HE)	No amidation Conventional LE	DA < 25% Amidated LE

DE, degree of esterification; DA, degree of amidation.

amidated with ammonia to form galacturonamide units in the molecular chain.

Some neutral sugars are also included in the homogalacturonic backbone. This is the case for rhamnose but also specifically in apple pectins for xylose.

The percentage of galacturonic acid of the whole molecule is defined as the galacturonic acid content (% GA), which is set at a minimum of 65% in the definition of pectin as a food additive.

The percentage of esterified or amidated galacturonic acid units of the total number of galacturonic acid units in the molecule are respectively defined as the degree of esterification (DE) and the degree of amidation (DA) (Table 1). Regulations limit the DA to a maximum of 25%.

Esterification Pattern

In addition to the number of components on the backbone, their position is of significant importance to the functional properties of pectin. The distribution of esters is of special importance as it affects the local electrostatic charge density of the polymer and so its interaction with other charged molecules, whether ions such as calcium, proteins, or other pectin molecules.

In apple pectin subjected to a mild extraction process, the ester distribution is reported to be almost random, while citrus pectin tends to have a somewhat blockwise distribution.

Neutral Sugars

Pectin always contains varying amounts of neutral sugars such as D-galactose, L-rhamnose, L-arabinose, and D-xylose. Some of these neutral sugars are constituents of side chains to the galacturonan backbone. However, 1,2-linked L-rhamnose is present in the main polygalacturonic chain where it forms 'kinks' in the molecular chain. Xylose is also a very important neutral sugar in apple pectin where it is attached in the homogalacturonic backbone.

It is well documented that neutral sugar side chains are concentrated in relatively short segments of the galacturonic backbone, described as 'hairy regions.' The part of the molecule free of side chains is described as the 'smooth region.' (Figure 2).

Chemical and Physical Properties

Solubility Pectin is soluble in water, but insoluble in most organic solvents. The solubilization rate in water is related to the degree of polymerization and the number and distribution of methyl-ester groups. The pH, temperature, and ionic strength of the solution are of great importance to the rate of pectin dissolution. The calcium content of the water used to dissolve the pectin is of special relevance as it is common that a high-water hardness will translate into an incomplete dissolution of pectin.

Rheology of pectin solutions Pectin solutions are viscous but pectin is not an especially effective viscosifier compared to other gums such as guar gum.

Dilute pectin solutions are almost Newtonian, and they are only slightly affected by the presence of calcium. However, solutions with more than 1% pectin exhibit pseudoplastic behavior and are strongly affected by calcium. There is a continuum in texture starting with water through thixotropic solutions with yield value to stiff gels depending on the pectin type and concentration, level of calcium, and pH.

Stability High-ester pectins are stable at pH levels of 2.5–4.5. Above a pH level of 4.5, β-elimination will occur, depolymerizing the galacturonic acid chain. This mechanism requires an esterified carboxyl group next to the glycosidic bond to be cleaved (Figure 3), so low-ester pectins are more stable at higher pH values.

The pectin molecular structure is quite resistant to heat. At pH around 3.5 pectin is only marginally depolymerized at high temperatures. The heat-stability

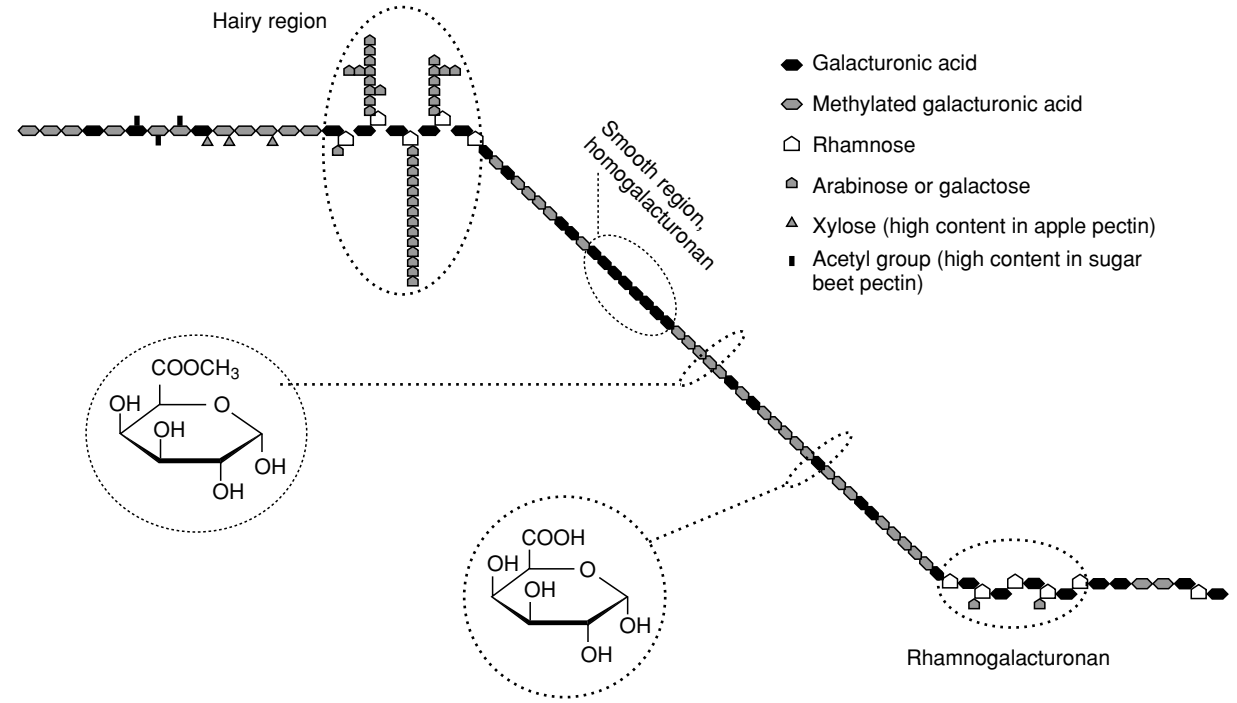

Figure 2 Primary and secondary structure of pectin.

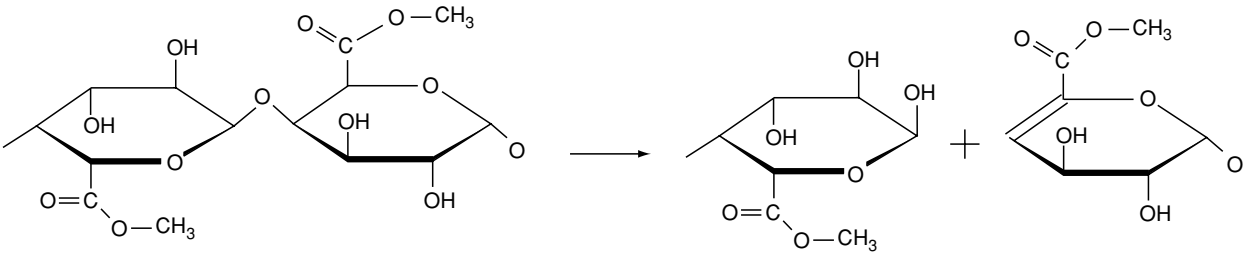

Figure 3 β-elimination mechanism.

of pectin is greatly improved when the water activity of the system is lowered through the addition of sugar.

Origin and Production

Pectin is a natural component of plants, predominantly in the form of pectic substance or protopectin, which is not soluble in water. The pectic substance is an essential part of the plant cell wall structure, acting as a cement for the cellulosic network and a hydrating agent. The exact nature of the pectic substance is not completely understood. It is, however, generally recognized that it is a complex structure in which pectin is attached to other cell wall components such as cellulose, hemicellulose, and proteins by covalent bonds, hydrogen bonds, and/or ionic interactions. In the plant the residual carboxyl groups are partly neutralized with cations of calcium, potassium, and magnesium which are present in the plant tissues.

Pectin is today commercially produced mostly from apple pomace and citrus peels by an extraction process followed by separation, purification, isolation, and then drying, milling, and standardization (Figure 4).

Functional Properties

Pectins are used in a broad variety of food and pharmaceutical products. Their principal properties are gel formation with both high- and low-ester pectins, viscosity build-up, and protein stabilization with high-ester pectins. As methods develop for obtaining a better understanding of the pectin molecular structure, it is likely pectin will be attributed new functional properties in the future.

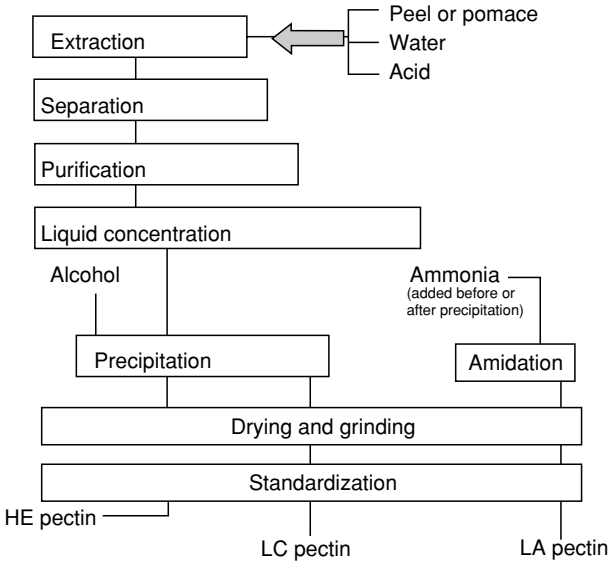

Figure 4 Pectin manufacture flow sheet. HE, high-ester; LC, low ester conventional; LA, low ester amidated.

Gelling of High-Ester Pectins

The ability of pectin to form gels in specific conditions has long been used in the production of jams and preserves. The separation of high- and low-ester pectins in accordance with the 50% esterification rule proves somewhat arbitrary when dealing with gelling mechanisms. It is clear that the dominant factor in the gelling of pectin is highly dependent on its degree of esterification but the complex network structure is most often the result of a combination of several mechanisms. The statement that high-ester pectins will gel with sugars and acid, while low-ester pectins gel with calcium, is certainly valid but certainly calcium can alter the gelling of high-ester pectins, just as pH and soluble solids will affect low-ester pectin gelling.

Gelling Mechanism

It is generally accepted that a high-ester pectin gel is formed by the cross-linking of the polymer in junction zones, in which mainly hydrogen bonds but also hydrophobic attractions between the methyl-ester groups play a part. Calcium bridges may also participate, especially if the esters are distributed in blocks, leaving large parts of the molecule as free acids.

Gelling will occur upon cooling of a media where favorable conditions are met. Cooling is necessary to decrease molecular movement and permit the formation of intermolecular interactions.

As the pectin chains carry negative charges they will tend to repel each other. This repulsion will depend on the charge density of the molecule which can be directly correlated to the pH of the medium and the frequency of free galacturonic acids of the polymer. The higher the pH and the lower the degree of esterification, the higher the charge density and hence the stronger the repulsion.

This repulsion, but more importantly the impossibility of forming hydrogen bonds between ionized pectin chains, are the reasons why a low pH is required for high-ester pectin gelling. At a low pH, typically below 3.6, the repulsion is low enough for the distance between the pectin chains to be reduced sufficiently and hydrogen bonding can occur. In order to achieve sufficient hydrophobic interactions to stabilize the molecular network, the water activity of the system also has to be decreased. Sugars are usually added for this purpose.

So, typically, high-ester pectins will only form gels when the pH is below 3.6 and soluble solids are above 55% (Figure 5).

Upon storage of a cooled gel, it is typical that the texture will still develop into a stronger final gel. This corresponds to a slow reorganization of the network involving the creation of new junction zones or an enlargement of the existing junctions between the pectin molecules.

Parameters Influencing Gelling

The gelling and final gel structure of high-ester pectins is influenced by a great number of parameters, the main ones being the pectin concentration, the degree of esterification, molecular weight, acetylation and branching of the pectin molecule and the pH, ionic strength, water activity, sugar type, and cooling rate of the gelling medium.

Pectin concentration The concentration of high-ester pectin will increase the final gel strength of the system due to the increase in the number of junction zones, increasing the number of chains with elastic activity. An increase in molecular weight would have the same effect. In addition to this expected effect, the gelation rate is also increased with increasing pectin concentration and a power law can be calculated between the two parameters.

Degree of esterification The degree of esterification of the galacturonic acids affects both the charge density of the polymer and the number of sites for hydrophobic interaction. As pectin molecules with a high degree of esterification are less charged, they can form gels at a higher pH and will also start gelling at a higher temperature.

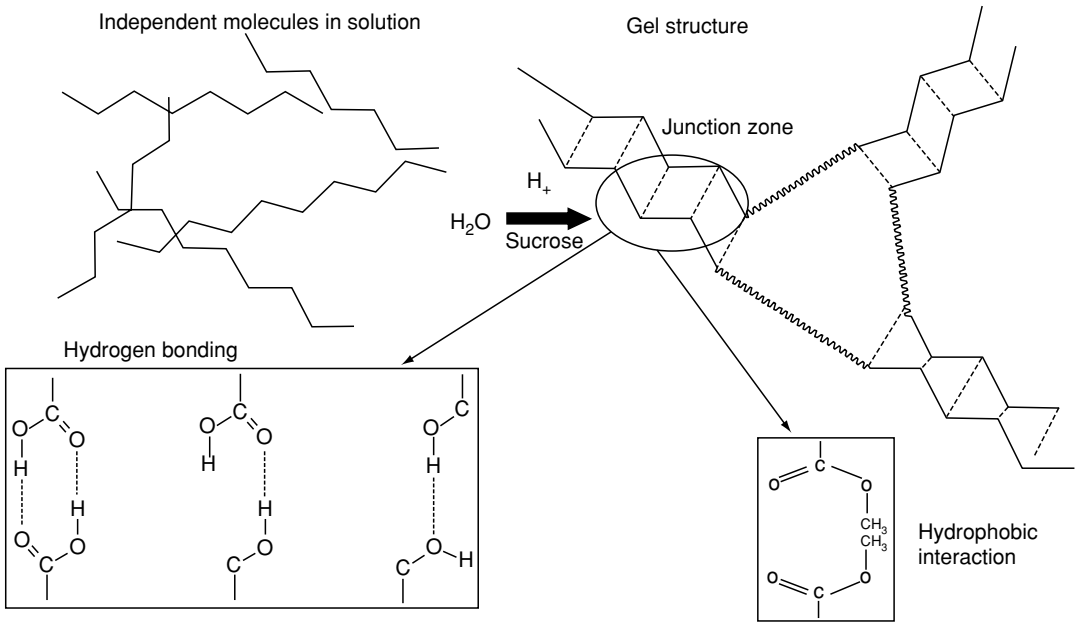

Figure 5 High-ester pectin gels through hydrogen bonding and hydrophobic interactions in an acidic water and sugar matrix.

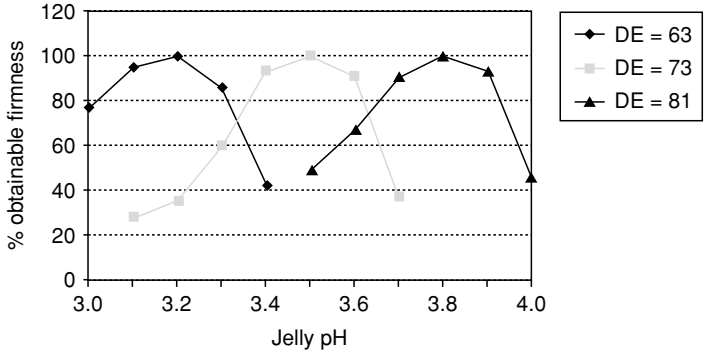

Figure 6 High-ester pectin: the degree of esterification (DE) determines the optimal pH for the gelation.

This last effect forms the basis for the classification of high-ester pectins into rapid-set, medium-rapid-set, slow-set, and extra-slow-set pectins as the degree of esterification is decreased from more than 70% to 50% (Figure 6).

The distribution of the ester groups on the backbone will also affect pectin gelation, as a marked blockwise distribution of esters will result in a significant contribution of calcium gelling. This contribution of calcium gelling will significantly increase the gelling temperature of the pectin (Figure 7).

Acetylation and branching Acetylation sharply decreases the gelling ability of pectin as the size of the acetyl groups does not allow the pectin chains to come close enough for interaction between the molecules.

The effect of neutral sugars on high-ester pectin gelling can be twofold and would need to be studied more extensively. Neutral sugars present on the pectin molecule could result in steric hindrance of intermolecular interaction and thus decrease the ability of the pectin to form gels. However, they could also participate in gelling through hydrophobic interaction and contribute to an increased cohesion of the gel.

pH and ionic strength of the gelling system The lower the pH, the lower the repulsion between the pectin molecules and, thus, the easier it will be for them to interact. This means that a low pH will lead to faster gelling in high-ester pectins. However, below a critical level, the gel strength will be reduced as the gelling is too fast to obtain a well-organized polymer

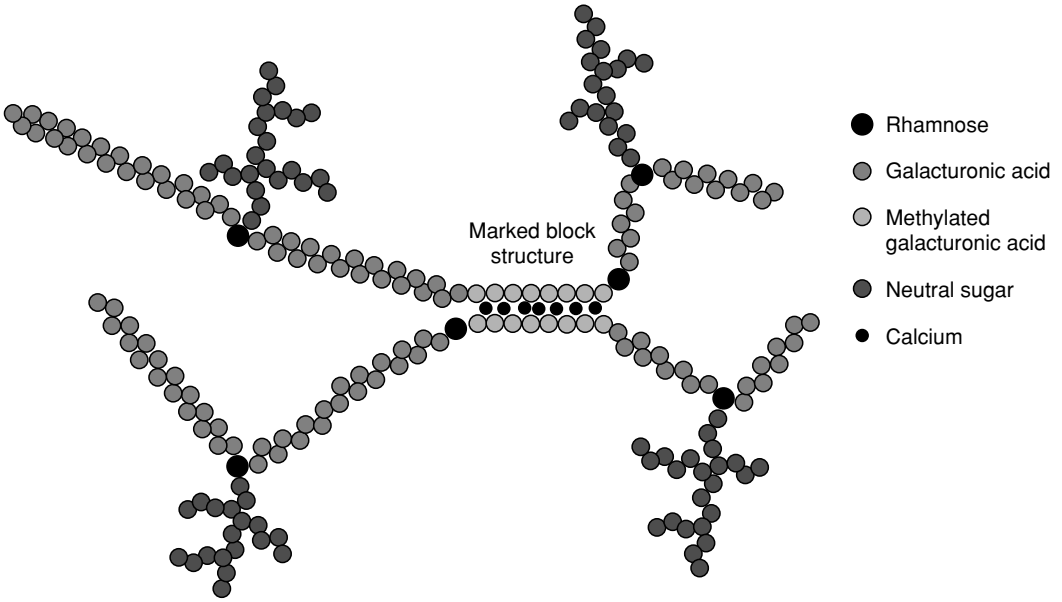

Figure 7 Participation of calcium gelling in high-ester pectin gelation.

network and precipitation can occur. The optimum pH for gelation is controlled by the degree of esterification of the pectin as well as the soluble solids content of the medium (Figure 6).

Through their effect on the neutralization of the pectin molecule, cations present in the system will affect gelling. High ionic strength shifts the optimum pH range towards higher values. This is particularly visible with sodium ions. However, with ions such as calcium or potassium that can bridge high-ester pectin molecules in areas with a low density of ester, there is a possible increase in junction zones. This effect can become significant for high-ester pectin with marked blockwise distribution of ester groups (Figure 7).

Water activity and sugar types Water activity and sugar types will both affect the way hydrophobic interactions can develop between the pectin molecules. As water activity is reduced, the hydrophobic interactions are easier to form, causing faster gelling to occur and the final gel strength to be increased.

The commonest way of reducing water activity in a food system is through the use of sugars. The effect of sugars on hydrophobic interaction and so on gel structure will be specifically linked to their molecular conformation and their interaction with the neighboring water molecules (Figure 8).

Influence of cooling rate and storage temperature Cooling decreases molecular movement and allows polymer molecules to interact at close distances. As the cooling rate is increased, the gelation rate is also increased. However, during rapid cooling and with a low storage temperature, gelling can actually become very slow, reflecting the difficult development of hydrophobic interactions in these conditions. With an intermediate cooling rate and temperature range, hydrogen bonds and hydrophobic interactions together can contribute to the build-up of a network with the highest elasticity.

Properties of high-ester pectin gels Due to the nature of the molecular interactions involved in high-ester pectin gelation, gels made with these pectins will typically not be thermo-reversible or shear-reversible. When submitted to mechanical stress, the broken gel will then show a high level of syneresis.

Gelling of Low-Ester Pectin

Low-ester pectin has traditionally been used for gelling food products when the conditions required to achieve a gel with high ester pectin were not met. Recently however, low ester pectins have also found applications in high sugar and low pH systems because of their specific texture characteristics.

Though the same base mechanism applies, low ester conventional and amidated pectin differ in their gelling properties and offer a broad range of functional properties.

Gelling mechanism The gelation of low-ester pectins is the result of ionic linkage through calcium bridges between carboxylic groups from two pectin

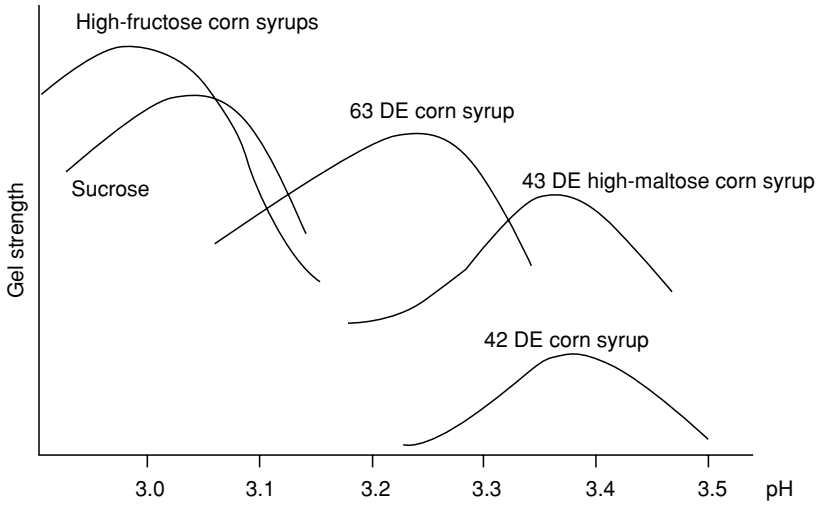

Figure 8 Effect of various sugars on high-ester pectin gel strength. DE, dextrose equivalent.

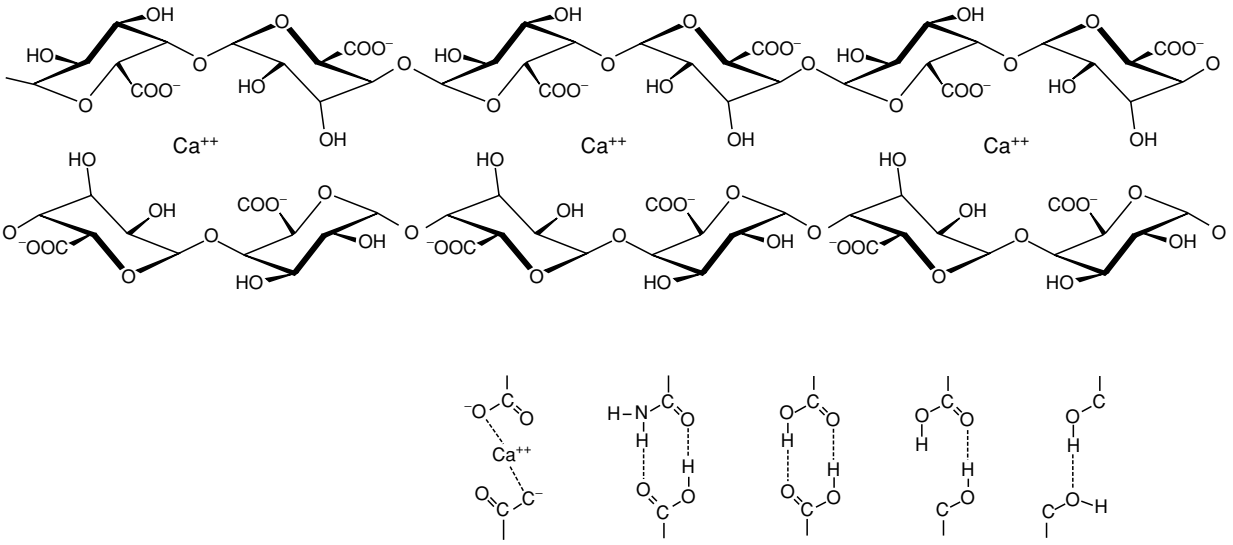

Figure 9 Eggbox model. Schematic overview of established low-ester pectin gel mechanism, showing calcium-induced junction zones. Inset shows detail of the various types of possible hydrogen bonding that participate in the junction zone, together with the calcium chelation (far left).

chains with the participation of hydrogen bonding. The linkage generally takes place upon cooling a pectin and calcium system. The most commonly accepted model of association is the eggbox model (Figure 9). In this model, pectin chains could be bridged by calcium ions, which incorporate in their coordination shells two polyanion oxygen atoms from one pectin molecule and three from another chain. Even though a number of positive ions can bridge pectin molecules, especially magnesium and potassium, calcium is particularly effective in complexing with carbohydrates, largely because its ionic radius (0.1 nm) is big enough to coordinate with many oxygen atoms and because of flexibility with regard to the direction of its coordinate bonds.

The exact calcium requirements to obtain a gel highly depend on the degree of esterification of the pectin, the recipe, and process parameters such as the rate of cooling. An increase in ionic strength, increase in pH, or decrease in the degree of esterification lowers the amount of calcium required to achieve sol–gel transition. An optimum calcium level can be defined for a given pectin in specific conditions. Above this optimum level, pregelation will occur, i.e., gelling will occur at too high temperatures to obtain a coherent gel structure (Figure 10).

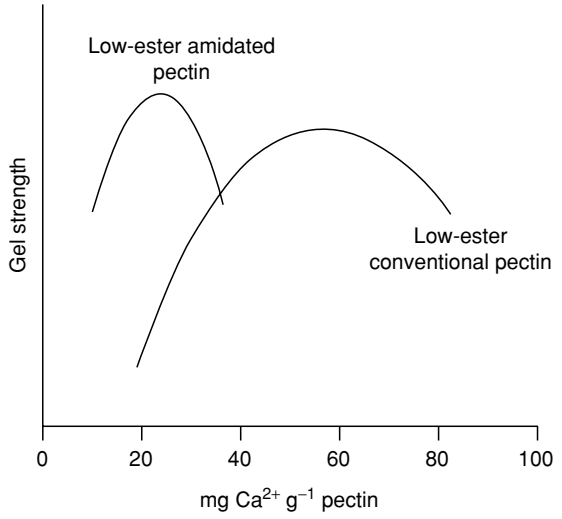

Figure 10 Influence of calcium on low-ester pectin gels.

Gelling is dependent on the length of the junction zones, that is, the number of galacturonic acid units involved in electrostatic bonds with calcium. The bonds are stable when at least seven consecutive carboxyl groups from each chain are involved. If the junction zones formed with calcium become too long, a pectin precipitate may be formed. This may occur with pectin with a very low degree of esterification when large amounts of calcium are available.

The presence of ester or amide groups prevents the formation of junction zones in the interjunction segments of the molecules, making them more flexible. Side chains also prevent aggregation of the pectin molecules through steric hindrance.

The typical high-ester pectin gelling mechanism with hydrogen bonds and hydrophobic interactions can also contribute to the final texture of low-ester pectin gels, especially at low pH and with high soluble solids concentration.

Parameters Influencing Gelling

Although the availability of calcium is a critical factor in the gelation of low-ester pectins, other parameters in relation to the pectin molecule and media have a significant influence on gelling and the final structure obtained. The main parameters are the number and distribution of ester and amide groups as well as the molecular weight of the pectin molecule and the pH, ionic strength, and water activity of the gelling system.

Influence of esterification Because calcium bonds can only occur in esterification-free zones, gel strength increases with a decreasing degree of esterification. For low-ester pectin with an average DE above 30%, the distribution of the esters may be of significant importance as it will control the length of possible junction zones and influence the gelling temperature, the final gel strength, and texture.

Influence of amidation The amidation of pectins was developed in the 1940s as a means of modifying the functional properties of low-ester pectins in order to achieve better gelling control. The exact mechanism by which the amide groups intervene in the gelation remains to be fully explained. However, it is generally accepted that amidation increases the gelling power of low-ester pectins due to the possibility of hydrogen bonding involving amide groups. Gels made with amidated pectins are firmer and require less calcium; they are also more thermo-reversible than ones made with low-ester conventional pectins.

Influence of molecular weight As for any polymer gel, the length of the polymer governs the number of junction zones required to achieve a coherent network. Low-ester pectins with a high molecular weight will exhibit a higher gelation rate, lower calcium requirements for gelling, and an overall more cohesive and elastic gel structure with a reduced tendency towards syneresis.

Influence of pH At low sugar content, as the pH is decreased, the pectin molecules are neutralized with protons, decreasing the probability of junction zones forming with calcium. This translates into higher calcium requirements and a looser gel texture at a low pH.

On the other hand, when the water activity of the system is decreased by the addition of sugars, the high-ester pectin gelling mechanism will start to play a significant part in the gelling and the calcium requirements will then be decreased when the pH is lowered. In usual food systems, availability of natural calcium will also be increased at a lower pH and thus may reduce the need for extra calcium addition (Figure 11).

Influence of ionic strength An increase in gel strength can be observed at a higher ionic strength. This is usually explained by the neutralization of the polymers by the extra ions, which allows the chains to be closer, leading to a more organized and cohesive gel.

Influence of water activity As the solids level increases, calcium requirements decrease. However, for most pectins, a higher solids level accelerates gelling, and increases the setting temperature and the final gel strength. It also reduces the optimum calcium window, thus increasing the risk of pregelation.

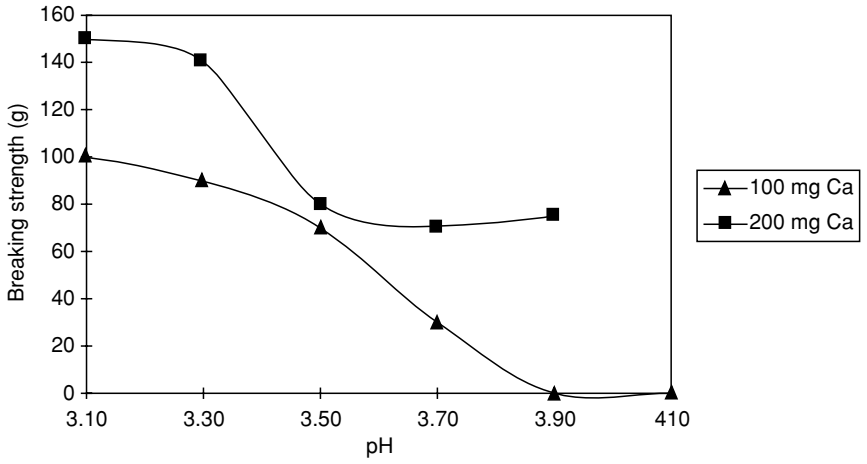

Figure 11 pH influences on Ca-curves in low-sugar system (31% soluble solids), with high reactive amidated pectin. Triangles, 100 mg Ca; squares, 200 mg Ca.

In practice, this leads to a choice of pectin with a higher degree of esterification (less calcium-reactive) at a higher solids level.

Properties of Low-Ester Pectin Gels

The properties of low-ester pectin gels are very dependent on the type of pectin used (conventional or amidated) along with the procedure and formulation employed to make the gel. Typically, they are thermoreversible and show a high degree of thixotropy. In specific conditions it is however possible to obtain heat-resistant or very brittle gels with no shear reversibility.

Protein Stabilization with High-Ester Pectins

In acidified conditions, casein and, more generally, food proteins will tend to agglomerate and sediment if the viscosity of the system is low enough. In these conditions the proteins are also very sensitive to dehydration and can easily become sandy after heat treatment. With the rapid development of acidified dairy beverages worldwide, now expanding with other protein sources such as soy, the need for effective protein stabilizers in a low-pH environment is growing strongly. High-ester pectin has proven to be a very useful stabilizer in these conditions.

Mechanism It is generally accepted that, at sufficient pectin concentrations, the adsorption of the carboxyl blocks of the pectin molecule to the protein surface will stabilize the protein system through steric repulsion (Figure 12). So, the presence of blocks of free carboxyl groups on the galacturonic backbone has an important influence on the protein-stabilizing property of pectin.

Due to the lower proportion of carboxyl groups, high-ester pectin has a weaker electrostatic interaction with protein than low-ester pectin but proves to be more effective. Indeed, it seems to be important that significant parts of the pectin molecule do not interact with the protein surface in order to achieve the steric repulsion effect. It is also the key to minimizing interactions between pectin and cations available in the system.

Both the DE of the pectin and the distribution of the esters on the polymer affect its stabilization properties. Excessively large blocks of carboxyl groups will tend to interact with the ions present in the system, such as calcium, rather than with the protein – an interaction which will lead to an increase in viscosity or even gelation.

Conditions Most food proteins (isoelectric point around 5) can form complex coacervates with anionic polysaccharides such as pectin (isoelectric point around 3.5) in the intermediate-pH region, where the two macromolecules carry opposite net charges: pH above the isoelectric point of the polysaccharide but below that of the protein.

In the case of pectin–protein interaction, the strength of the complex will depend on several factors, such as the distribution of the carboxyl groups on the galacturonic acid backbone, but also on the three-dimensional protein structure and the distribution of ionizable groups on its surface. The whole interaction will also depend on several system parameters such as pH, ionic strength, presence of sugars, or fat.

Through its role in the ionization of both the protein and pectin molecules, pH is the most significant factor affecting electrostatic pectin–protein interactions. It

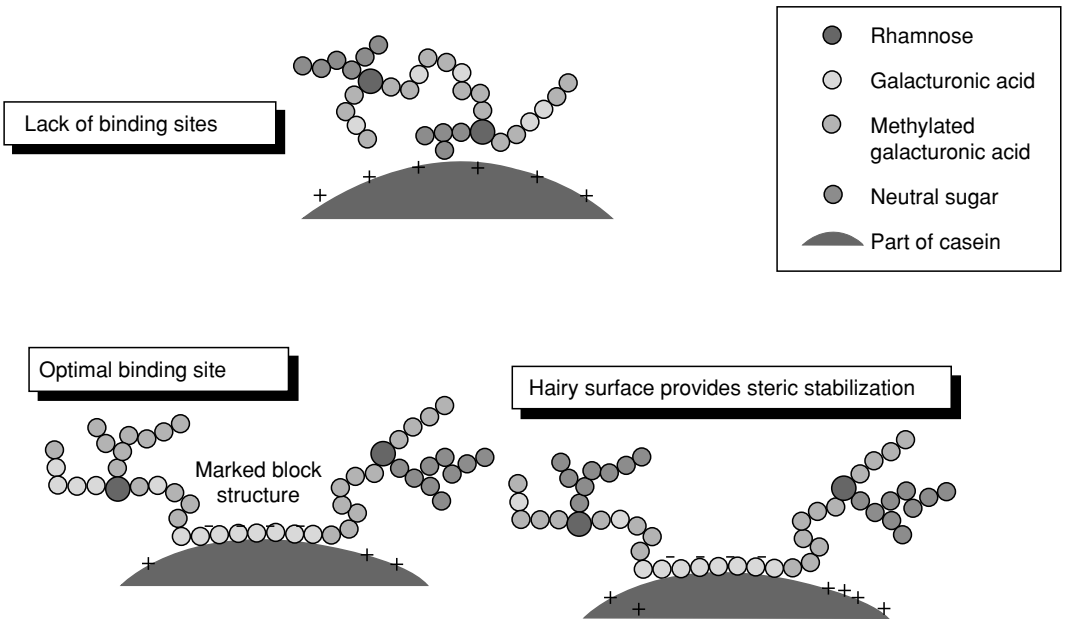

Figure 12 Theoretical picture of interaction between pectin and protein.

also plays a very significant role in the protein structure and how proteins interact in a complex system such as milk.

The optimum pH range for interaction between high-ester pectin and casein is 3.6–4.5. At a lower pH, the block structures of the high-ester pectin will not be sufficiently ionized for proper protein binding as the pH is too far below the pKa of the pectin. Above the isoelectrical point of the protein, the protein–polysaccharide complex is very weak or nonexistent and the electrostatic protein–protein repulsion is dominant. It is, however, clear that this repulsion is not sufficient to stabilize the proteins.

Conclusion

Pectin exhibits a wide range of functional properties and enjoys a very good public image as a natural product derived from fruit. Today it is widely used as a textural ingredient and stabilizer in a variety of food applications, and there is little doubt that its usage will grow as new functionalities are revealed.

See also: **Cholesterol**: Absorption, Function, and Metabolism; **Citrus Fruits**: Composition and Characterization; **Dietary Fiber**: Physiological Effects; **Fermented Milks**: Types of Fermented Milks; **Gums**: Properties of Individual Gums; Food Uses; Dietary Importance; **Jams and Preserves**: Methods of Manufacture; Chemistry of Manufacture; **Pectin**: Food Use; **Protein**: Interactions and Reactions Involved in Food Processing; **Rheological Properties of Food Materials**; **Rheology of Liquids**; **Stabilizers**: Types and Function; Applications

Further Reading

Dickinson E (1998) Stability and rheological implications of electrostatic milk protein–polysaccaride interactions. *Trends in Food Science and Technology* 9: 347–354.

Kravtchenko TP, Voragen AGJ and Pilnik W (1994) Characterization of industrial high methoxyl pectins. *Gums and Stabilizers for the Food Industry* 7: 27–35.

May CD (1990) Industrial pectins: sources, production and application. *Carbohydrate Polymers* 12: 19–99.

Rolin C and De Vries J (1990) Pectin. In: Harris P (ed.) *Food Gels*, pp. 401–434. London: Elsevier.

Thakur BR, Singh RK and Handa AK (1997) Chemistry and uses of pectin – a review. *Critical Reviews in Food Science and Nutrition* 37: 47–73.

Walter RH (ed.) (1991) *The Chemistry and Technology of Pectin*. London: Academic Press.

Food Use

L Flutto, Danisco, New Century, KS, USA

Introduction

Pectin has been used traditionally in food ever since man started cooking fruits and vegetables. As a

Table 1 Function of pectin in food applications

Function	*Food application*	*Typical use level (%)*
Viscosity building	Juice beverage, soft drinks	0.01–0.20
	Sauces	0.10–0.50
	Sorbet (avoids crystallization)	0.10–0.40
Protein stabilization	Yogurt drinks	0.15–0.60
	Milk–Juice drinks	0.15–0.60
	Acidified soy drinks	0.20–0.60
Gelation	Jams, jellies, preserves	0.30–1.20
	Fillings, glazes	0.30–1.20
	Desserts	0.50–1.00
	Confectionery	0.50–2.50

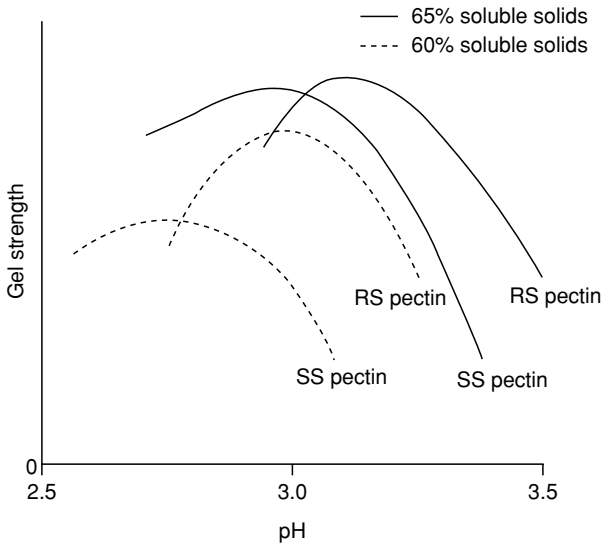

Figure 1 Choice of high-ester pectin and slow-set pectin (SS) content to determine the optimal pH for the gelation. RS, rapid set.

natural component of plants, pectin is a desirable texturizing and stabilizing agent for use in all kinds of processed foods available today. In this chapter, we take a brief look at the most important food uses of pectin and its key properties in each system.

Function in Foods

Pectin is traditionally used for its gelation properties, with more than a third of the pectin commercially available being utilized by the jam and jelly industry. Pectin is used to 'set' or gel jams, jellies, and preserves. The pectin holds fruit pieces suspended within the gel structure, trapping water between cross-linked pectin molecules. The result is a gelled food that is shelf-stable. However, in more complex applications such as fruit preparations for dairy products, pectin is used to impart specific rheological properties during the manufacturing and storage of the fruit preparation, as well as during blending with the dairy product and all during the shelf-life of the finished product. Pectin is also widely used for its unique protein stabilization properties in acidic conditions. The general acceptance of pectin as a natural material and a soluble food fiber has made it desirable for use in a variety of applications (Table 1).

Jams and Jellies

Jams, jellies, and preserves can be divided into two categories. Traditional products have a solids range usually above 60%. Reduced sugar products have a soluble solids content in the range of 25–55%.

High-ester pectins are used in the production of jams, jellies, and preserves with soluble solids above 60% and a pH below 3.6. As high-ester pectin gels are not shear-reversible, it is of utmost importance that the jars be filled at a temperature above the setting temperature of the pectin in order to avoid a broken gel and syneresis. Various high-ester pectins used in recipes with different soluble solid contents will have different optimum pHs (Figure 1).

The choice of pectin will also be dependent on the type of product made. A fast-setting pectin is preferred for preserves where fruit pieces have to be suspended evenly throughout the container. By contrast, a slow setting pectin is used for jellies in order for air bubbles to escape prior to gelling, resulting in a fully transparent product. The size of the packaging will also influence the choice of pectin, with fast-setting pectin recommended for larger jars, as the cooling rate is slower, and a quick gelling is necessary to keep the fruit pieces in place. It is important, however, to note that for very large containers (several kilograms to several hundred kilograms), it is necessary to cool down the product before filling as the cooling rate is too slow in the center of the product, and the pectin is degraded when exposed to high temperatures for an extended time. In this case, a slow-set pectin or even a low-ester pectin is required to allow for cooling prior to filling without breaking the gel network. When filling in small packages and especially in food-service single-serve packages, the use of a slow-set pectin is also required as the filling time can be quite long. The temperature of the product in the holding tank before the packaging line will need to be reduced to avoid pectin degradation in time and insure consistent gelling from the first to the last package of the batch. Low-ester pectins are used when conditions for gelling a high-ester pectin are not met. Typically, this is the case when the sugar content is reduced below 60%. They are, however, also used in high-soluble-solids formulations when a

Table 2 Typical reduced-sugar jam formulations with a low-ester amidated pectin

Ingredients	*Jam final soluble solids content*					
	25%		*35%*		*45%*	
	Dosage (%)	*Soluble solids (%)*	*Dosage (%)*	*Soluble solids (%)*	*Dosage (%)*	*Soluble solids (%)*
High-calcium reactive low-ester amidated pectin	0.8	0.8	0.7	0.7		
Medium-calcium reactive low-ester amidated pectin					0.6	0.6
Locust bean gum	0.1	0.1				
Sugar I	3.6	3.6	2.8	2.8	2.4	2.4
Water I	18.0		14.0		12.0	
Fruit, 10% soluble solids	45.0	4.5	45.0	4.5	45.0	4.5
Sugar II	15.9	15.9	26.9	26.9	37.4	37.4
Water II	21.1		15.1		7.1	
K-sorbate, 20% w/v	0.2	0.04	0.2	0.04	0.2	0.04
Na-benzoate, 20% w/v	0.3	0.06	0.3	0.06	0.3	0.06
Citric acid, H_2O, 50% w/v	As required		As required		As required	
Flavoring	As required		As required		As required	
Total	105.1	25	105.1	35	105.1	45
Evaporation	5.1		5.1		5.1	
Yield	100.0		100.0		100.0	
pH	3.2		3.2		3.2	
Filling temperature (°C)	40–50		45–55		55–65	

Procedure (regardless of targeted soluble solids)

1. Dry-blend the low-ester pectin (the locust bean gum if present) and sugar I, and add the blend to 80 °C hot water I, agitating vigorously
2. Mix fruit, sugar II and water II and heat the blend to approx. 80 °C
3. Add the pectin solution, agitating continuously
4. Evaporate to the desired content of soluble solids
5. Add preservatives and adjust pH with citric acid
6. Add flavoring
7. Cool to filling temperature and fill

spreadable texture is desired. These pectins gel in the presence of calcium ions, which is the main factor controlling the gelling. The soluble solids level and pH are only secondary parameters, but they have an influence on the gelling temperature and the texture.

As the soluble solids are decreased, the greatest challenge for the manufacturer is to keep an even distribution of fruit pieces. Indeed, the main influence of the soluble solids level will be on the viscosity of the jam at high temperatures during the process.

The choice of a low-ester pectin is therefore mostly based on the solids level. The use of low-ester amidated pectins with increasing calcium reactivity is usually recommended when the soluble solids are decreased. This insures a faster set with an even distribution of fruits. In low-brix formulations, the addition of calcium may be required to supplement the calcium brought by the fruits and the water and to insure a fast and homogeneous gelling.

To further avoid the flotation issues, a combination of low-ester amidated pectin and conventional pectin is commonly used. The low-ester conventional pectin forms a network at a higher temperature than the amidated counterpart and helps to keep the fruit well distributed. Other gums such as locust bean gum or guar are also often used in combination with a low-ester pectin to provide viscosity in low-solids formulations (Table 2).

A rapidly growing category in jams and jellies field is products where refined sugars have been replaced by fruit-juice concentrates. These products usually have between 40 and 50% soluble solids and rely on low-ester pectin gels. The only difference with standard reduced sugar recipes is in the ionic balance, as the fruit concentrates add a significant amount of calcium, as well as magnesium and potassium, to the formulation. This influences the choice of pectin as well as the amount of calcium to be added to the product. It is important to position the product on the flat part of the calcium curve and use a pectin with a broad tolerance to calcium in order to obtain a consistent texture, despite the usual variations in calcium content of the natural raw materials (Figure 2).

Fruit Preparations

The fruit preparation industry constantly juggles between retaining the identity of the fruit, perfect fruit

distribution throughout large totes, unique formulations able to accommodate fruits with a varying calcium content, and the texture and flavor of the final dairy product in which the fruit preparations are used. These challenges, combined with the pressure on manufacturers to develop new fruit preparations quickly and the need to create process-friendly formulations, call for the use of highly functional and versatile stabilizers. Pectin has found a natural application in these systems and is widely used.

Use of Pectin in Fruit Preparations

Pectin is used in fruit systems for various purposes. However, fruit preparations should only be thickened and not fully gelled. This requires pectin capable of giving fruit preparations a high degree of thixotropy.

The texture of the fruit preparation can be controlled by various parameters, as outlined in Table 3.

Use of a High-ester Pectin

When more than 55% soluble solids are used, a high-ester pectin with a typical dosage of 0.3% can successfully impart viscosity and control syneresis in a fruit preparation.

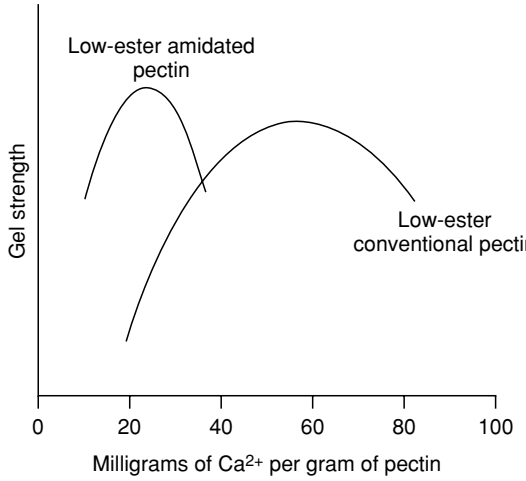

Figure 2 Low-ester-pectin calcium curves.

It is also possible to use a high-ester pectin with less than 55% soluble solids in order to yield a special texture. In these cases, a high-ester pectin typically gives the fruit preparation a high degree of shine and smoothness, although higher dosages are necessary.

Special high-ester pectins developed for their interaction with casein can also be used to obtain a high degree of texture carry-through in the white mass. At a usage level of 1–2%, the pectin gives the fruit preparation the required texture and significantly increases the viscosity of the finished stirred yogurt. It also helps maintain a constant viscosity and reduce syneresis during the shelf-life of the dairy product (Figure 3).

Use of a Low-ester Pectin

Conventional and amidated low ester pectins are widely used in fruit systems. While conventional pectins require a slightly higher dosage and high level of calcium, special ones have extreme processing flexibility and can be used with great success in a broad variety of formulations from 30 to 50 Brix with various pH values. It is common to include low ester conventional pectin in fruit on the bottom formulations. Low-ester amidated pectin can prove more economical with the possibility of lower dosages, although it tends to be more sensitive to calcium variations (see Figure 2). When the calcium level is adjusted, it is possible to achieve an excellent fruit suspension with a high level of shear reversibility. This allows low viscosity during pumping for maximum fruit identity.

Calcium Addition for Texture and Postreaction Control

When using a low-ester pectin, it is usually necessary to add calcium salts to obtain the full texture. The necessary quantity will depend on the calcium source, pH, soluble solids, fruit type, and content of the preparation. Specific calcium salts should be selected according to the speed of release and labeling considerations. Calcium chloride, -lactate, and -citrate is

Table 3 Parameters influencing the texture of fruit preparations

Choice of pectin	High-ester, low-ester conventional, and low-ester amidated pectins perform differently and allow a variety of textures
Pectin dosage	A higher pectin dosage results in a thicker texture
Calcium level	The calcium level needs to be adjusted with consideration for a possible reaction in the white mass; higher calcium levels usually increase viscosity, but pregelation can occur if too much calcium is added
Soluble solids	Higher soluble solids usually result in a shorter texture
Filling temperature	Lower filling temperatures usually result in a more fluid texture
Use of other hydrocolloids	Locust bean gum, guar gum, xanthan gum and starches can be successfully used in combination with pectin to achieve specific textures

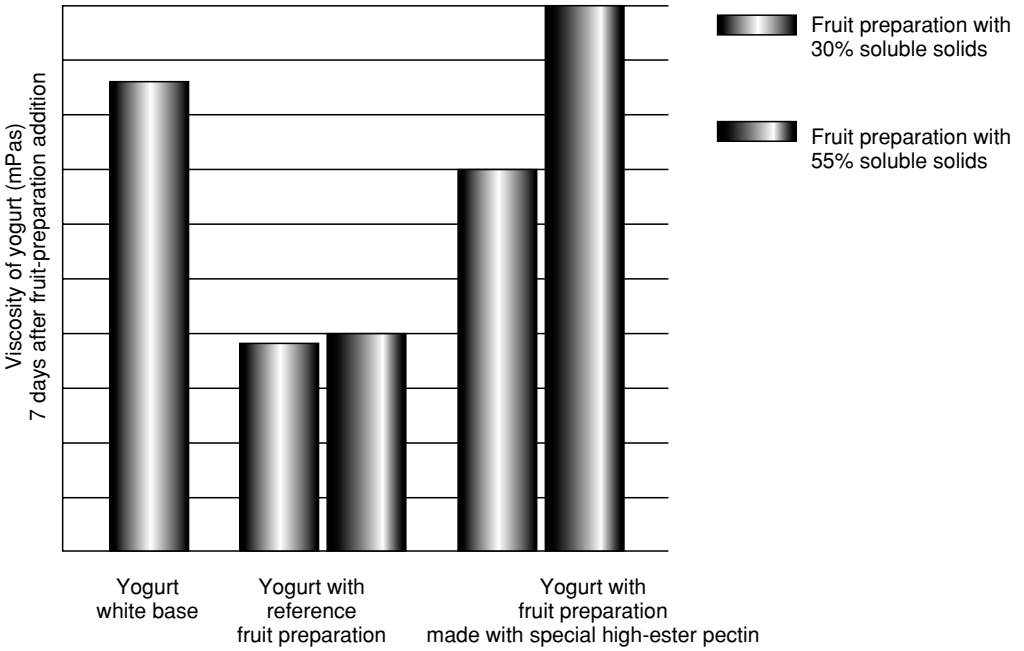

Figure 3 Use of a special high-ester pectin in a fruit preparation to increase the viscosity of yogurt.

most commonly used to supplement the calcium contained in the fruit.

If the calcium content of the fruit preparation is too low, the pectin present in the fruit preparation may react with the calcium of the yogurt, resulting in a grainy texture (stirred yogurt) or a gelled preparation (fruit on the bottom). To avoid this type of postreaction, it is highly recommended that the appropriate amount of calcium is added to the fruit preparation.

Bakery Fillings

In principle, bakery fillings are a form of fruit preparation. However, they deserve a separate paragraph as the demands on these products are fairly specific. Applied on or inside the dough, they usually need to be heat-resistant and have a level of water activity low enough to minimize water transfer from the filling to the dough. This is especially important for long-shelf-life cookies and cereal bars, which tend to be where there is greatest demand. As is the case for fruit preparations for the dairy industry, the trend is for the bakeries to outsource the production of fillings and transport them in larger containers.

Most bakery fillings have a high solids content typically between 60 and 80%. Thus, a high-ester pectin seems a natural choice also because these gels are thermally irreversible. However, high-ester pectin gels are completely nonthixotropic and lose their entire structure after pumping. So, pure high-ester pectin gels, which require cooling prior to filling and mechanical pumping at the bakery, are not adequate for packaging in large containers. High-ester pectin gels would be destroyed by such processes and show syneresis and a very poor bake stability.

High-ester pectins may, however, be used successfully for cold gelation systems, also called delayed setting systems. These are gel systems in which all the gelling conditions for a high-ester pectin are met, with the exception of the pH, which is kept too high. This viscous syrup can then be shipped to the bakery, where the acid is then added upon depositing the filling on the dough.

For other types of bake-stable fillings, specific low-ester pectins can prove ideal. Even if low-ester pectin gels are typically characterized as being thermally reversible (a property used in glazes), they can be made bake-stable by using an optimized formulation. Low-ester conventional pectins are usually preferred, as they show a higher bake stability due to a stronger binding with calcium. The choice of the correct calcium salt as well as its dosage is the key to both the texture of the filling and its heat stability. In very high solids formulations, the use of pectin in combination with alginate can prove useful in improving the filling appearance and its bake stability.

Glazes and Bakery Decor

Glazes are widely used to cover fruit tarts and other types of cakes where they not only add consumer appeal but also extend shelf-life by forming a

barrier to dehydration and microbiological contamination. The high transparency and shine of pectin gels make the use of this hydrocolloid an ideal choice for this application. In some specific applications, it is possible to use combinations of pectin with alginate or carrageenan.

Contrary to most bakery fillings, in glazes and bakery decorations, the heat reversibility of the system is usually of paramount importance. These products are most often sold to bakery shops in concentrated form and reheated and diluted with water before application. The gel must then form quickly and keep for several days, even on fruits with a high water content. In industrial applications, the gel has to be so completely reversible such that the product can be sprayed. Clearly, these requirements are best met with low-ester amidated pectins.

The use of specific calcium sources with controlled release together with calcium sequestrants is necessary to control the gelling of a low-ester pectin and secure an optimum performance after reheating.

Confectionery

Pectin is widely used in traditional confectionery items such as 'pâte de fruit,' or Zefir. However, it is also an important textural agent in a number of other confectionery articles, where its superior flavor release, high transparency, and clean bite are sought after.

Jelly candies typically have a soluble solids content of between 75 and 80% and a pH of around 3.5. Due to the high solids level, it is important to use a very slow setting high ester pectin and include buffer salts in the formulation to avoid any pre-gelling before depositing.

The standard process usually involves making a pectin solution with the buffer salts, heating the sugars and fruits when present, and mixing in the pectin solution before cooking to reduce the water level and obtain the desired solids content. Immediately before depositing, acid, color, and flavorings are added. Pectin jellies can be deposited in starch or rubber molds or in large trays for further cutting. It is also possible to use pectin on jet cooking equipment where all the ingredients are mixed together, and the pectin actually dissolves at a very high temperature (around 140 °C) with all the solids present. This procedure usually requires the use of specific pectins with particular dissolution properties.

In most cases, the pectin chosen for confectionery will be an extra-slow setting type to guarantee a low setting temperature. Even with this type of pectin, the gelling temperature of the system is usually around 80 °C, and the depositing temperature needs to be above this level if the full gelling potential of the pectin is to be used. This is a major difference from the commonly manufactured gelatine candies, which have a much lower setting temperature. Another significant difference is the speed of gelation, as pectin jellies can be demolded quickly after depositing since a firm gel is obtained very quickly. As with all high-ester pectin gels, typical pectin jellies are not thermally reversible.

The buffer salts commonly used with pectin in confectionery include sodium citrate and sodium potassium tartrate. The latter tends to give the jellies a somewhat firmer texture.

High-ester pectin is also often used in combination with other ingredients such as gelatine, starch, or agar to give intermediate textures. In the case of gelatine candy, pectin is also used in small amounts to increase the melting point of the confectionery pieces in warm climates, thus insuring a better appearance throughout their shelf-life without significantly affecting the texture.

In specific candy pieces where a neutral flavor such as vanilla is used, the use of a low-ester pectin is necessary in order to obtain a gel at a higher pH. As with high-ester pectins, it is usually necessary to reduce the setting temperature. In these systems, calcium sequestrants such as sodium hexametaphosphate are used. Turkish delight is one such traditional candy based on a low-ester pectin gel with or without the addition of starch.

High- and low-ester pectins are also widely used for confectionery fillings, which are then either enrobed usually with chocolate or, for example, panned into jelly beans.

Fruit Beverages and Soft Drinks

Pectin is used as a viscosifier in beverages and soft drinks, and high-ester pectins may be used as a mouthfeel improver. This use has been widely developed for juice drinks with a reduced juice content or sugar-free soft drinks.

Low-concentration pectin solutions can be considered Newtonian and show a low viscosity. This is of great relevance for the use of pectin in fruit beverages and soft drinks as the concentration used rarely exceeds 0.5%. Indeed, the clean mouth feel imparted by pectin compared with the tendency towards a slimy mouth feel with some other gums could be related to the low viscosity of pectin solutions at the shear rate applied in the mouth. This property makes pectin an ideal choice when trying to replace the mouth feel lost by the reduction in sugar content.

As most juice beverages and soft drinks contain calcium, pectin with a high degree of esterification is

usually recommended to minimize the calcium sensitivity of the pectin and avoid any risk of gelling. A slight gelling of the product changes the rheology of the solution, resulting in undesirable pseudoplastic behavior. For this reason, the most commonly used pectin is of the rapid setting type. Pectin manufacturers usually offer rapid-set pectins standardized to a viscosity instead of gelling properties, so as to guarantee a consistent performance in a beverage application.

Yogurt Drinks and Other Acidified Protein Beverages

From being largely restricted to certain regions or addressed to specific consumer groups, acidified protein drinks today have become a highly esteemed and popular food world-wide. Aided by the increased focus on health benefits, the market for acidified beverages with proteins is currently enjoying rapid growth.

Perfect stability and a consistent mouth feel are essential to consumers, so the use of stabilizers is necessary in order to prevent any defects such as sedimentation, whey separation, and grainy mouth feel.

At a low pH, the positive net charge of the protein clusters in an acidified drink is insufficient to prevent aggregation of the proteins, and the drink therefore suffers from severe sedimentation of protein aggregates and separation. This lack of physical stability is even more pronounced if a shelf-stable, postpasteurized product is desired as the unprotected proteins are denatured by the heat treatment. Although specific structural details vary, the same general stability issues occur with various protein sources such as casein, whey, soy, or other plant-derived proteins.

The stability problem can be avoided by adding a stabilizer such as a high-ester pectin. With its controlled block structures of nonmethylated galacturonic acid units, high-ester pectins can bind to the protein surface as the blocks are negatively charged at a pH of 3.6–4.5, a typical level in acidified beverages. Bound to the proteins, the long pectin molecules protect them from reaggregation through steric stabilization. Even with a postpasteurized drink, this stabilization can produce a shelf-stable product and prevent unwanted sedimentation and serum separation.

Yogurts and Other Dairy Products

The gelling of low-ester pectins with calcium ions is often used to set dairy products, but low-ester pectins are also used to reinforce the texture of other types of gel network such as the protein network in yogurts and fermented dairy products. Recently, the trend towards low-fat or non-fat dairy products has increased the need to use textural ingredients and stabilizers to restore a pleasant mouth feel and avoid common defects such as syneresis.

Low-ester pectins can be dissolved in neutral milk and are commonly used in the manufacture of gelled neutral dairy desserts. Giving a softer texture with more body and a high level of creaminess than widely used carrageenan, pectin is preferred in premium creamy formulations where the extra cost of pectin can be justified. It is also used when carrageenan is not desired on the label. In addition, pectin is the main component of a variety of regional home-made desserts formulations such as Fruche in Japan. Fruche is a fruit and pectin solution sold in a pouch, which the consumer can mix with milk. After half an hour in the refrigerator, the gelled product can be consumed as a refreshing dessert. In Europe, pectin is often seen in dry mix blends for home-made desserts.

Whether the effect of low-ester pectins on the texture of fermented dairy products comes through gelation with calcium or interaction with proteins is not quite clear. However, low-ester pectin is already widely used on its own or together with gelatine or starch to add texture to yogurts and other fermented products.

Owing to the high level of calcium present, the usage level of the pectin must be low in order to avoid an excessively strong reaction between the pectin and the calcium, resulting in a sandy product. The maximum recommended usage level is around 0.25%. Interestingly, high-calcium-reactive pectins usually show a better performance than their low-calcium-sensitive counterparts. Amidated pectins are usually preferred, even if conventional low-ester pectins do prove useful in specific recipes for syneresis control.

Pectin can also be used in combination with carrageenan for mousse products, where gelatine is not desired. Pectin provides a much better gelation after the mousse mix is whipped, leading to a more sliceable texture compared with carrageenan alone.

Other Applications

Pectin is used in a great number of other applications. For example, it is used in tomato-based products such as barbecue or taco sauce as an alternative to starch. Usually, low-ester conventional pectins are used to thicken the sauce and ensure a nonmelting texture when the sauce is poured over hot food.

See also: **Dairy Products – Nutritional Contribution**; **Gelatin**; **Jams and Preserves**: Methods of Manufacture; Chemistry of Manufacture; **Soft Drinks**: Chemical

Composition; **Sweets and Candies**: Sugar Confectionery; **Yogurt**: Yogurt-based Products

Further Reading

Beli R. Thakur, Singh RK and Handa AK (1997) Chemistry and uses of pectin – a review. *Critical Reviews in Food Science and Nutrition* 37(1): 47–73.

May CD (1990) Industrial pectins: sources, production and application. *Carbohydrate Polymers* 12: 19–99.

Rolin C and De Vries J (1990) Pectin. In: Harris P (ed.) *Food Gels*, pp. 401–434. Amsterdam: Elsevier.

Walter RH (ed.) (1991) *The Chemistry and Technology of Pectin*. London: Academic Press.

PELLAGRA

D A Bender, University College London, London, UK

Introduction

Pellagra is a nutritional disease due to deficiency of the vitamin niacin and the essential amino acid tryptophan. The clinical features of pellagra are dermatitis, diarrhea, and dementia; it is commonly known as the 'disease of the four Ds,' since it is also fatal – the fourth 'D' is death.

During the first half of the twentieth century, pellagra was a major problem of public health in the southern USA, with some 87 000 deaths attributed to the disease between 1900 and 1950. Through the 1950s to 1980s, it continued to be a problem in southern Africa and parts of India, but improvements in food availability and general nutritional status have led to more or less complete eradication of pellagra as an endemic nutritional deficiency disease.

There have been a few reports of outbreaks of pellagra among refugees in southern Africa when rations have been inadequate to meet the demand, and individual cases have been reported among alcoholics and people with Crohn's disease and other gastrointestinal diseases that impair nutrient absorption.

Pellagra may also occur as a result of a variety of (relatively rare) conditions affecting tryptophan metabolism and as a side-effect of a number of drugs that inhibit tryptophan metabolism. In alcoholics, it is not clear whether pellagra is the result of an impairment of tryptophan and niacin metabolism directly attributable to alcohol, or whether it reflects general undernutrition among people who obtain a considerable proportion on their energy needs from alcohol, and hence have a low intake of (nutrient rich) foods.

The first full description of pellagra was given by Casal in 1735, working in Spain; he called it '*mal de la rosa*' – the red disease. The name pellagra was coined by Frapolli in 1771, from the Italian '*pelle*' for skin and '*agra*' for rough, thus describing the most striking feature of the disease, the roughened appearance of the skin, resembling severe sunburn in areas exposed to sunlight. Casal recognized that the underlying cause of pellagra was nutritional, associated with the limited diet of many in central Spain at the time, although he did not associate it with the then recent introduction of maize from Central America.

The spread of pellagra largely followed the introduction of maize as a dietary staple, and by the nineteenth century, it was a major problem around the Mediterranean, in the Balkans, and in the Ukraine, as well as the southern USA and southern Africa. In South Africa, pellagra became a problem following the outbreak of rinderpest in 1897, which killed most of the cattle, leading to a marked deterioration in the diet of people who had previously had ample supplies of milk and meat to supplement their maize-based diet. In the USA, it was the social and economic upheaval of the Civil War, which led to large numbers of subsistence farmers living on a diet based very largely on maize.

The other region where pellagra was a major problem is the Deccan plateau in India, where the dietary staple is jowar (a variety of millet, *Sorghum vulgare*), rather than maize.

Although Casal had described pellagra as a nutritional disease, at the beginning of the twentieth century it was generally assumed to be due to an infectious agent. The nutritional basis of the disease, and the interaction between deficiencies of both niacin and tryptophan, was established mainly by Goldberger and coworkers in the USA between 1913 and 1948. A 'pellagra-preventing factor' isolated from protein-free yeast extract was shown in 1937 to be nicotinic acid, and either nicotinic acid or its amide, nicotinamide, was shown to prevent or cure the disease in both man and experimental animals. The name niacin was coined when it was

decided to enrich foods with the vitamin, since it was considered that nicotinic acid would be unacceptable as a food additive, because of its chemical (but not metabolic) relationship with nicotine. In the USA, 'niacin' is commonly used to mean nicotinic acid, and the amide is niacinamide; elsewhere, 'niacin' is used as a generic descriptor for nicotinic acid and/or nicotinamide.

Other early studies showed that feeding additional protein, and especially the essential amino acid tryptophan, would also prevent or cure pellagra, suggesting that it was a protein-deficiency disease rather than due to a lack of a vitamin. It was not until 1947 that the problem was resolved, when it was shown that tryptophan is a metabolic precursor of the nicotinamide moiety of the coenzymes nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP). The coenzymes can be formed in the body either using preformed dietary niacin or by *de novo* synthesis from tryptophan.

The reason a maize-based diet predisposes to pellagra is that the proteins of maize are particularly poor in tryptophan, so that a diet in which there are few other sources of protein provides insufficient tryptophan for nicotinamide synthesis. Other cereals, such as wheat, barley, rye, rice, and millet, contain enough tryptophan to meet the requirement for niacin synthesis.

Although maize proteins are poor in tryptophan, the other cereal associated with pellagra, jowar, provides a minimally adequate amount of tryptophan to meet requirements. However, the proteins of jowar are considerably richer in leucine than most other proteins, and a diet based largely on jowar provides a considerable excess of leucine. There is a considerable body of evidence that this amino acid imbalance can be a precipitating factor in the development of pellagra when the dietary intake of preformed niacin is extremely low, and the intake of tryptophan is only marginally adequate. Leucine inhibits kynureninase, and hence reduces the rate of oxidative metabolism of tryptophan, resulting in reduced formation of NAD. In addition, leucine competes with tryptophan for tissue uptake, and thus has a further inhibitory effect on the rate of tryptophan oxidative metabolism and NAD synthesis.

All cereals contain preformed niacin. However, this is largely present as a variety of nicotinoyl esters, collectively known as niacytin, which are not hydrolyzed by digestive enzymes to any significant extent, so that most of the niacin present in cereals is nutritionally unavailable. A small proportion of the niacin present as niacytin (up to about 10%) may be biologically available as a result of nonenzymic hydrolysis by gastric acid.

Interestingly, pellagra has not been a problem in Central America, the original home of maize. This is because of the traditional way in which it is prepared. Rather than milling the grain, it is steeped overnight in limewater (calcium hydroxide solution), then squeezed to form the dough from which tortillas are made. This alkaline treatment results in hydrolysis of most of the nicotinoyl esters, so releasing free nicotinic acid, which is nutritionally available. Although maize spread to many countries following its discovery, it was generally milled like other grains rather than being treated in the traditional Mexican manner.

Clinical Features of Pellagra

Dermatitis

Exposure of the skin to modest amounts of sunlight results in a severe sunburn-like dermatitis in sufferers (pellagrins). Mechanical pressure can cause similar lesions, especially around the wrists and ankles. The skin in the affected areas is red and slightly swollen at first, and then becomes rough, thickened, cracked, and dry, with scaling, a shiny surface, and brown pigmentation.

The cause of this photosensitive dermatitis in pellagra is unknown and cannot be attributed to the known metabolic functions of either tryptophan or the nicotinamide nucleotide coenzymes. There is some evidence that there is increased metabolism of the amino acid histidine in the skin in pellagra, resulting in a lower than normal concentration of both histidine and an intermediate in its metabolism, urocanic acid, both of which are believed to have a role in absorbing ultraviolet light, and so minimizing damage to the skin from exposure to sunlight. Treatment of pellagrins with niacin both clears the dermatitis and also increases the concentration of histidine and urocanic acid in the dermis.

The skin lesions of pellagra may be due to secondary zinc deficiency when tryptophan intake is inadequate or its metabolism is disturbed; they resemble the lesions seen in acrodermatitis enteropathica, which is due to a failure to secrete the intestinal zinc-binding ligand, which is believed to be picolinic acid – a tryptophan metabolite. There is some evidence that some pellagrins have a poor zinc status, and the zinc depletion associated with excessive alcohol consumption may be a factor in the development of pellagra among alcoholics.

Diarrhea

Although diarrhea is common in pellagrins, it is not a constant feature of the disease, and indeed in some

cases, there may be chronic constipation. The cause of both the diarrhea and the constipation is almost certainly general nutritional deficiency, resulting in atrophy of the intestinal mucosa and the intestinal musculature.

Dementia

The psychiatric disturbances of pellagra range from mild hallucinations with some psychomotor retardation, through confusion with increasing hallucinations, to severe dementia and anxiety psychosis, with melancholia, intermittent stupor and possibly epileptiform convulsions. In many ways, this resembles schizophrenia, but the dementia of pellagra can be differentiated from schizophrenia and the organic psychoses by the sudden lucid phases that alternate with the most severe mental symptoms.

Although the cause of the psychiatric disturbance in pellagra remains to be firmly established, it is likely that it is largely due to deficiency of tryptophan, as a precursor for the neurotransmitter serotonin (5-hydroxytryptamine), rather than a direct result of inadequate supply of the nicotinamide nucleotide coenzymes in the brain.

The Equivalence of Dietary Tryptophan and Niacin

The nicotinamide ring of the coenzymes NAD and NADP can arise either from preformed dietary niacin or by *de novo* synthesis from quinolinic acid, an intermediate in the oxidative metabolism of tryptophan. For an adult in nitrogen balance, the amount of tryptophan available for oxidative metabolism is about 99% of the dietary intake, since new synthesis of protein is balanced by catabolism of tissue proteins, releasing their tryptophan for metabolism. A variety of studies have shown that under normal conditions, 60 mg of tryptophan are equivalent to 1 mg of preformed dietary niacin, and it is usual to express the total niacin intake in terms of niacin equivalents – the sum of preformed niacin plus 1/60 of the tryptophan. On this basis, average Western diets provide more than enough niacin to meet requirements from tryptophan alone, ignoring preformed niacin. Indeed, reanalysis of dietary records of people who died from pellagra in the USA during the first half of the twentieth century shows that their intake of preformed niacin and tryptophan was (marginally) adequate to meet requirements, suggesting that some other factor may have been involved in precipitating the disease.

Complicating Factors in the Etiology of Pellagra

Deficiency of Vitamins B_2 and B_6

The enzyme kynurenine hydroxylase in the oxidative pathway of tryptophan metabolism is a flavoprotein, and hence its activity is impaired in riboflavin (vitamin B_2) deficiency. The enzyme kynureninase is pyridoxal phosphate-dependent, and its activity is impaired in vitamin B_6 deficiency. This means that deficiency of either vitamin B_2 or B_6 will reduce the rate of oxidative metabolism of tryptophan and hence the rate of formation of quinolinic acid and NAD. If the intake of tryptophan is anyway marginal, this could well precipitate pellagra.

Mycotoxins

A number of mycotoxins cause DNA damage, and activate poly(ADP-ribose) polymerase as part of the DNA repair mechanism. This enzyme uses NAD as the source of ADP-ribose, releasing nicotinamide. While, theoretically, this nicotinamide could be reused for synthesis of NAD, the enzymes involved (nicotinamide phosphoribosyltransferase, nicotinamide deamidase, and nicotinic acid phosphoribosyltransferase) are all more or less saturated with their substrates at normal tissue concentrations. This means that the additional nicotinamide released by poly(ADP-ribose) polymerase cannot be used for NAD synthesis but will largely be methylated to *N*-methylnicotinamide and excreted. Exposure to such mycotoxins may therefore be a factor in the etiology of pellagra, depleting the body of nicotinamide. The one report of an outbreak of pellagra in Mexico was associated with a shipment of maize that had suffered considerable fungal damage.

Estrogens and Progestagens

Through the first half of the twentieth century, when pellagra was a major problem in the southern USA, there was a twofold excess of women over men among those affected. In a number of reports of more recent outbreaks, there is a similar sex ratio between the menarche and menopause. There is no difference in the numbers of males and females affected among either prepubertal children or adults aged over about 40 years. This suggests that estrogens and/or progestagens may have a pellagragenic effect. Estrogen metabolites are competitive inhibitors of kynureninase, and the administration of progesterone results in reduced activity of kynurenine hydroxylase, although *in vitro*, neither progesterone nor its conjugates affects the activity of kynurenine hydroxylase.

When the intake of preformed niacin is low, and that of tryptophan is marginal, the impairment of tryptophan metabolism by estrogens and progesterone may be sufficient to precipitate pellagra more commonly in women than men.

Non-nutritional Pellagra

Pellagra can occur as a result of impairment of tryptophan metabolism due to a variety of diseases that affect the tryptophan oxidative pathway, or as a result of drugs that inhibit one or more enzymes of the pathway, despite an apparently adequate intake of tryptophan. General malabsorption associated with gastrointestinal disease; there have been a number of case reports of pellagra among people with Crohn's disease. In most cases, the condition responds well to supplements of nicotinamide.

Carcinoid Syndrome

Under normal circumstances, about 1% of the daily intake of tryptophan is metabolized by way of 5-hydroxytryptamine in the central nervous system and gut, with the remainder being oxidized by way of kynurenine, and thus available for NAD synthesis. A carcinoid is a tumor of the enterochromaffin cells of the gastrointestinal tract, which forms 5-hydroxytryptamine from tryptophan. The carcinoid syndrome occurs when the tumor has metastasized, usually to the liver, and in extreme cases as much as 60% of the daily intake of tryptophan may be metabolized by way of 5-hydroxytryptamine. The result is a considerable reduction in the rate of oxidative metabolism through kynurenine, and hence a considerable reduction in the synthesis of NAD from tryptophan, resulting in the development of pellagra in a significant proportion of patients.

Hartnup Disease

Hartnup disease is an inborn error of metabolism affecting the membrane proteins that transport the large neutral amino acids (including tryptophan). The same proteins are involved in the absorption of free tryptophan from the gastrointestinal tract into the bloodstream, from the bloodstream into tissues, and in the reabsorption of amino acids from the urine. The result of the defect is a considerable reduction in the amount of dietary tryptophan that is absorbed, as well as a considerable loss in the urine. Thus, despite an apparently adequate intake, there is a deficiency of tryptophan (and other large neutral amino acids), resulting in the development of pellagra.

Inborn Errors of Tryptophan Metabolism

A number of inborn errors of metabolism affecting enzymes of the tryptophan oxidative pathway have been reported, all of which result in the development of pellagra. Such conditions include much reduced activity of tryptophan dioxygenase, the first enzyme of the pathway; low or undetectable activity of kynureninase and kynurenine hydroxylase; and increased activity of aminocarboxymuconic semialdehyde decarboxylase, the enzyme that competes with the nonenzymatic cyclization of aminocarboxymuconic semialdehyde to quinolinic acid, the precursor for NAD synthesis.

Drug-induced Pellagra

A number of drugs can precipitate pellagra, despite an apparently adequate intake of tryptophan. The best-documented such drug is the antituberculosis drug isoniazid (isonicotinic acid hydrazide), although two antiparkinsonian drugs, Benserazide and Carbidopa, are also associated with niacin depletion. These drugs have an indirect effect on tryptophan and niacin metabolism. They are hydrazine derivatives that act as carbonyl-trapping reagents and therefore cause depletion of vitamin B_6 by forming inactive adducts with the metabolically active form of the vitamin, pyridoxal phosphate. Among other effects, this results in impaired activity of kynureninase, and hence a reduced rate of tryptophan oxidative metabolism and NAD synthesis. Although the pellagra responds to supplements of nicotinamide, it is more usual to give supplements of vitamin B_6, at least in combination with isoniazid, in order to minimize the other metabolic effects of vitamin B_6 depletion.

Pellagra has also been reported among cancer patients treated with a variety of chemotherapy agents that cause strand breaks in DNA. This activates poly(ADP-ribose) polymerase as part of the DNA repair mechanism, and so depletes tissue NAD. Indeed, this is the mechanism of action of these drugs; they deplete NAD in the tumor cells to such an extent that there is severe impairment of energy-yielding metabolism, leading to cell death. It has been suggested that patients treated with such chemotherapeutic agents should receive supplements of niacin to protect against the development of pellagra.

See also: **Cereals**: Contribution to the Diet; Dietary Importance; **Drug–Nutrient Interactions**; **Maize**; **Mycotoxins**: Toxicology; **Niacin**: Physiology; **Vitamins**: Overview; Determination

Further Reading

Bender DA (1983) Biochemistry of tryptophan in health and disease. *Molecular Aspects of Medicine* 6: 101–197.

Bender DA (2003) *Nutritional Biochemistry of the Vitamins*, 2nd edition. Cambridge: Cambridge University Press.

Bender DA and Bender AE (1986) Niacin and tryptophan metabolism: the biochemical basis of niacin requirements and recommendations. *Nutrition Abstracts and Reviews (Series A)* 56: 695–719.

Roe DA (1973) *A Plague of Corn: the Social History of Pellagra*. Ithaca, NY: Cornell University Press.

PEPPERS AND CHILLIES

B Pickersgill, The University of Reading, Reading, Berkshire, UK

Introduction

Pepper and chilli have been used to designate many different plants and plant products. This article is concerned principally with *Capsicum* peppers, whose fruits are consumed as a pungent spice or condiment (chilli or cayenne peppers), or as a nonpungent vegetable or source of color and flavor (sweet or bell peppers, paprika, pimentos). Several species of *Capsicum* were domesticated in the American tropics, but *C. annuum* has outstripped all others in global importance. The basic structure of the fruit is similar in all types of *Capsicum*. However, variation in thickness of the fruit wall and thickness of the cuticle affects ways in which the fruits are harvested, handled, and processed. The domesticated peppers vary even more strikingly in the characteristics most valued by humans: color, aroma, and pungency. Their fruits have significant nutritional value, as well as medicinal properties related to their pungency. World production of both spice and vegetable peppers has increased steadily throughout the last two decades.

Pepper, Chilli, Pimento, and Malagueta

The European discoverers of the Americas often named the novel plants they encountered on the basis of real or perceived similarities to those with which they were already familiar. The confusion thus begun in the names associated with *Capsicum* still persists.

'Pepper' is derived from a Greek word originally used for black and white pepper, but subsequently applied to genera belonging to seven different families (**Table 1**), all of which have pungent fruits or seeds. 'Chilli' comes from one of the indigenous languages of Mexico, where *Capsicum* has been consumed for more than 5000 years. The Spanish spelling of this word is 'chile'. In the USA today 'chile' tends to be used for the plant and its fruit, but 'chili' for a cooked dish containing *Capsicum* and various other ingredients. 'Chile powder' thus signifies ground fruits of *Capsicum*, but 'chili powder' also includes garlic, onions, cumin, and other spices.

'Pimento' is derived from the Latin word for pigment, and in medieval times signified spices in general. Pimento is still used as an alternative name for allspice (**Table 1**), but when applied to *Capsicum*, 'pimento' or 'pimiento' in English usage is usually restricted to large red nonpungent fruits, sold canned or used in processed foods.

The name 'malagueta,' used throughout much of Latin America for an extremely pungent small-fruited *Capsicum*, has also been borrowed from an Old World spice and is used for two unrelated genera in the American tropics (**Table 1**).

Botanical Background

Capsicum belongs to the family Solanaceae, which also includes tomato, eggplant (aubergine), and potato. (*See* **Potatoes and Related Crops**: Fruits of the Solanaceae; **Tomatoes**.) *Capsicum* contains about 20 species, all confined to the Americas until Columbus' discovery of the New World. In the Americas, human selection produced much variation in the fruits. Many of these variants were formerly regarded as distinct species, but now only five domesticated species are generally recognized: *C. annuum, C. frutescens, C. chinense, C. baccatum*, and *C. pubescens*. Each was domesticated independently, but contains a similar range of variation in fruit characters.

The bell peppers, paprikas, and pimentos, together with most of the Mexican chillies, are all included in *C. annuum*. This is now the most widespread, and economically the most important, of all the cultivated species of *Capsicum*. There is disagreement over whether *C. frutescens* is or is not a species distinct from *C. annuum*. If it is considered distinct, then the

Table 1 Botanical identities, geographic origins, and uses of plants for which the names pepper, pimento, and malagueta have been used

Scientific name	*Common names*	*Origin*	*Uses*
Anacardiaceae (mango, cashew, pistachio family)			
Schinus molle	Pepper tree, pink peppercorn	Andes	Condiment
Annonaceae (custard apple family)			
Xylopia aethiopica	Guinea pepper, African grains	Tropical Africa	Spice
Xylopia spp.	Malagueto, pimenta de macaco (monkey pepper)	Tropical America	Condiment
Myrtaceae (clove, guava, eucalyptus family)			
Pimenta dioica	Jamaica pepper, pimento, allspice	Caribbean	Spice
Piperaceae (pepper family)			
Piper nigrum	Black, white, green pepper(corns)	India	Condiment
Solanaceae (potato, tomato, tobacco family)			
Capsicum spp.	Red pepper, chilli pepper, capsicum, aji, malagueta	American tropics	Spice and condiment
Capsicum annuum (nonpungent cultivars)	Sweet, bell, green or red pepper, pimento, mango (USA)	Mexico	Vegetable
Rutaceae (citrus family)			
Zanthoxylum spp.	Szechuan pepper, Chinese prickly ash	Far East	Condiment in Chinese cookery
Zingiberaceae (ginger, cardamom, turmeric family)			
Aframomum melegueta	Melegueta pepper, grains of paradise, guinea grains	West Africa	Spice, especially for alcoholic drinks

cultivar 'Tabasco', which is the main ingredient of tabasco sauce, belongs in *C. frutescens*. In Mexico and Central America, wild forms of both species are often lumped together as bird pepper (also known as chile piquin or chiltepin).

The most common cultivated pepper of the West Indies and much of South America is *C. chinense*, which is closely related to *C. annuum* and *C. frutescens* but has a distinctive aroma and flavor. Habanero and Scotch Bonnet types (the latter named from the shape of the fruit) are now finding specialty markets in some developed countries. Nonpungent forms of *C. chinense* are grown commercially in Amazonia, where *C. annuum* does not succeed, and may prove suitable for other parts of the humid tropics.

The remaining two domesticated species are virtually unknown outside the Americas. *C. baccatum* is widespread in Andean South America and southern Brazil. The rocoto, *C. pubescens*, is another Andean species, easily recognized by its dark seeds.

Morphology and Anatomy of the Fruits

The fruit of *Capsicum* is a berry, which is unusual in being hollow, not filled with pulp as in tomato. In wild peppers, the very pungent small red fruits are so palatable to birds that ripe fruits are rarely found. As the fruits ripen, an abscission layer develops between the base of the fruit and the calyx, so birds can readily remove mature fruits from the plant. In domesticated peppers, this abscission layer does not develop, so the fruits remain on the plant until harvested by humans. Domesticated peppers may have much larger fruits than wild peppers, and may be pungent, mild, or nonpungent. Size and pungency are genetically independent characters, so small nonpungent fruits and very pungent large fruits can and do occur. Fruits of domesticated peppers have various shapes, described by local names such as girl's finger, dog's penis, fish eye, and friar's hat. In some European countries, sweet peppers always have pointed fruits, while in others, 'blocky' types that can be stood on the stylar end and stuffed from the stalk end are preferred.

In all *Capsicum* fruits, an outer wall or pericarp surrounds an inner cavity that contains the seeds. Fruits that are used fresh have a thick pericarp, whereas fruits that are used dry or after grinding have a thinner pericarp. Water loss from the pericarp is retarded by an impermeable outer cuticle, which is indigestible and may be tough, but accounts for the relatively long shelf-life of the vegetable peppers. The cuticles of pimentos are removed before canning, but cuticles of bell peppers are not usually removed, so human selection for palatability has produced thinner cuticles in bell peppers than in pimentos. When the fruits are used dried, those that lose moisture quickly without becoming wrinkled or unattractive in appearance are preferred. These fruits thus also often have thin cuticles.

A layer of so-called giant cells immediately under the inner epidermis of the pericarp gives the inner surface of the fruit wall a characteristic blistered appearance which may help to distinguish fragments of *Capsicum* pericarp from other contaminants in the ground spice. The function of the giant cells is not known, but they are not associated with pungency.

In small-fruited chilli peppers, there are two chambers, or carpels, separated by an inner wall, or septum. The seeds are attached to a spongy tissue called the placenta that develops from the septum. Large fruits are often formed from more than two carpels, and therefore have additional septa. The septa are better developed at the stalk end of the fruit, and often fail to reach the tip, so most of the seeds are borne on a hemispherical placenta at the stalk end of the single cavity in the fruit.

Special Characteristics

Quality of pepper fruits, and of the products made from them, depends on their color, aroma, and pungency. Hot peppers also have certain well-known pharmacodynamic effects. Nutritionally, both hot and sweet peppers are valuable sources of vitamins.

Color

The middle layer or mesocarp of the fruit wall is responsible for the color of both ripe and unripe fruits. The cells of the mesocarp contain plastids, which in the unripe fruit usually contain chlorophyl. The quantity of chlorophyl varies in different cultivars. Occasionally chlorophyl is absent and the unripe fruits are creamy white. Immature fruits may also contain sap-soluble, purple anthocyanin pigments. As the fruit ripens, both chlorophyl and anthocyanin usually disappear. Chloroplasts are converted to chromoplasts, and additional chromoplasts form *de novo*. There is a relationship, not yet fully understood, between chlorophyl content of the unripe fruit and carotenoid content of the ripe fruit. Immature fruits that lack chlorophyl may ripen red, but the red is deepest in fruits that are dark green when unripe. (*See* **Chlorophyl; Colorants (Colourants): Properties and Determination of Natural Pigments; Ripening of Fruit.**)

As a red pepper ripens, total carotenoids increase 35-fold and are sequestered within fibrils in the chromoplasts. About half of this increase represents *de novo* synthesis of red pigments. These consist of three ketocarotenoids: capsanthin (30–60% of total carotenoids present), capsorubin (5–15%), and cryptocapsin (about 5%). They are unusual in containing the cyclopentanol ring, unlike red carotenoids of fruits such as tomato (which accumulates lycopene). Lycopene is thought to be an intermediate in the biosynthesis of capsanthin and capsorubin (Figure 1).

Although the carotenoid pigments are formed in the chromoplasts, the quality and quantity of these pigments, which determine the different colors of the mature fruit (Table 2), are controlled by genes located on the chromosomes, not in the plastids. Fruits of the yellow color series cannot synthesize red carotenoids. In *C. annuum*, yellow-fruited plants carry a deletion of the DNA that codes for the enzyme capsanthin-capsorubin synthase, which catalyzes the final step in the production of capsanthin and capsorubin. Recessive alleles of other nuclear gene(s) reduce the quantities of pigments in the fruits, but we do not know where or what these blocks to pigment synthesis may be. Chocolate-brown fruits, such as those used in the famous Mexican sauce mole, are homozygous recessive for an allele of another nuclear gene, which prevents chlorophyl breakdown in the ripe fruit. The combination of red carotenoids and green chlorophyl appears as brown.

Breakdown of the red pigments in harvested fruits, or in powders produced by drying and grinding these fruits, is not prevented by blanching, but is retarded by antioxidants. This suggests that pigment breakdown is not due to enzyme action, but may involve two other processes: an autoxidative degradation which is accelerated by heat, and an autocatalytic destruction in light which involves direct absorption of light energy.

Aroma

Droplets of volatile oil in the mesocarp cells produce the characteristic aroma of the pepper fruit. These droplets increase in quantity as the fruit ripens. The oil is a mixture of methoxypyrazines, aliphatic alcohols, and esters, in which the most important constituent is 2-methoxy-3-isobutylpyrazine. This has one of the lowest odor thresholds of any compound examined, being detectable at levels of 2 parts in 10^{12} parts water. Methoxypyrazines have also been reported from other vegetables such as raw potato and French bean, but at much lower levels than in *Capsicum*. (*See* **Sensory Evaluation**: Aroma.)

The aliphatic alcohols and esters seem to be responsible for the fruity and floral components of the aroma. Although aroma is a valued attribute of fresh chillies, and the different aroma and flavor properties of different chillies are usually well-known to habitual consumers, little is known about differences in composition of the volatile oil either within or between different species of *Capsicum*.

Pungency

Some botanists have used pungency of the fruits to distinguish *Capsicum* from allied genera. The biological significance of pungency is not known, but the sensory effect is produced by a group of vanillyl amides collectively known as capsaicinoids (Table 3), amongst which capsaicin and dihydrocapsaicin

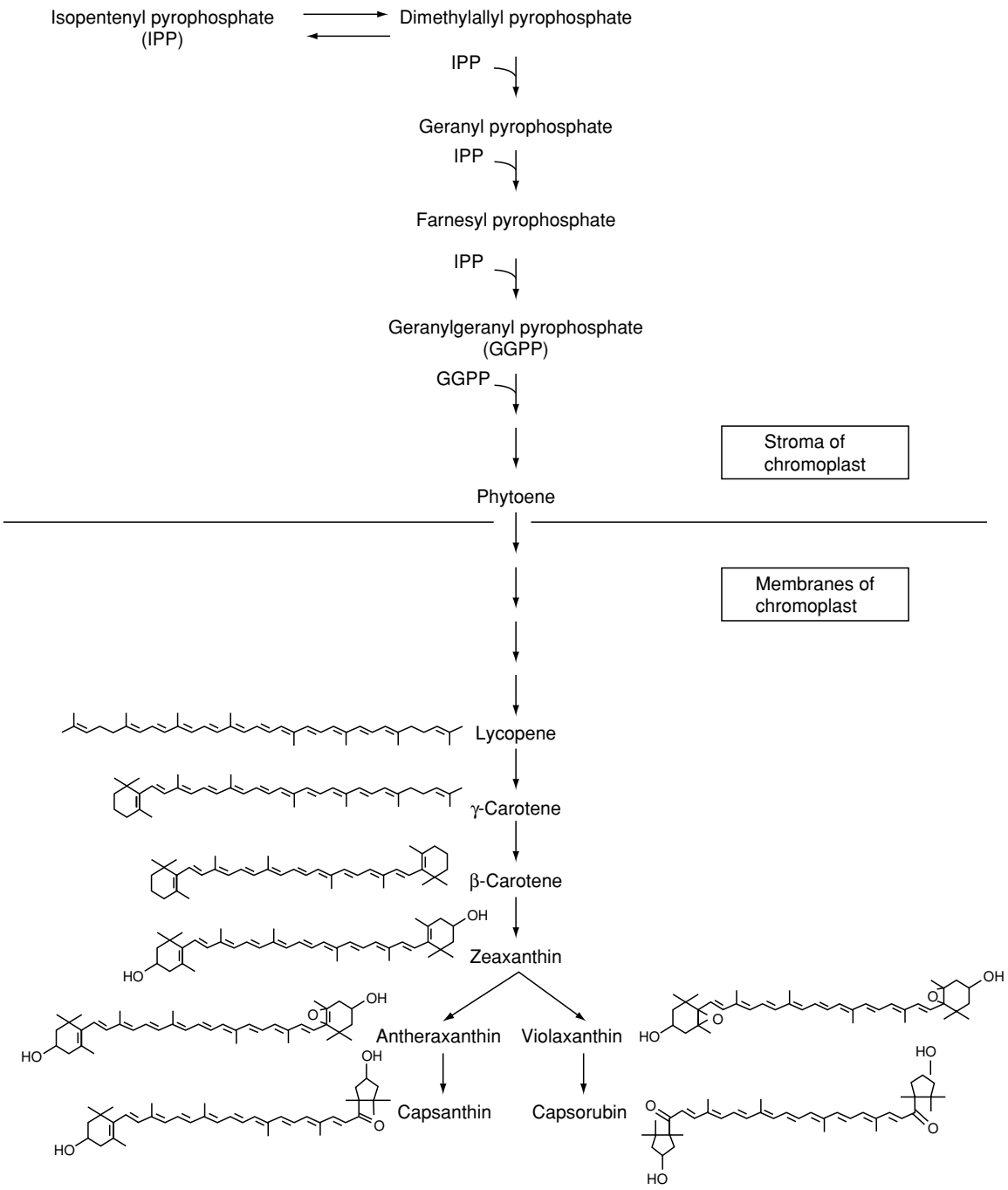

Figure 1 Biosynthesis of carotenoid pigments in fruits of *Capsicum*. Modified from Camara B, Hugueney P, Bouvier F, Kuntz M and Monéger R (1995) Biochemistry and molecular biology of chromoplast development. *International Review of Cytology* 163: 175–247, with permission.

predominate. Capsaicin is insoluble in cold water but soluble in alcohol, acetone, ether, and similar solvents. It is one of the most pungent compounds known, detectable by taste at dilutions of 1 in 15–17 million. Pungent compounds are also present in other plants, for example, ginger and black pepper (Table 3). The capsaicin molecule has three characteristic features: the vanillyl group, the acid–amide linkage, and the alkyl side-chain. Change in any one of these decreases pungency. Piperine from black pepper lacks the vanillyl group and has a short side-chain with a bulky substituent. Its pungency is more than two orders below that of capsaicin. Gingerols and shogaols, from ginger, have a vanillyl group and a long alkyl

chain, but lack the acid–amide bond and thus are also much less pungent than capsaicin. The length of the alkyl side-chain has been varied experimentally in synthetic analogs of capsaicin. Pungency appears first in compounds with a six-carbon side-chain, peaks with the nine-carbon side-chain, then gradually declines as chain length increases. The synthetic nine-carbon analog, *N*-vanillylnonamide, has been used as a substitute for capsaicin.

Biosynthesis of capsaicinoids is thought to involve one pathway from phenylalanine to vanillylamine, and one from valine or leucine to their respective isoacids, followed by condensation of vanillylamine with the activated fatty acids. Two capsaicinoid-like compounds (capsiate and dihydrocapsiate) have been identified from a nonpungent Japanese cultivar of *C. annuum*. These compounds have an ester moiety in place of the amide moiety of capsaicin and

Table 2 Effects of pigment quality and quantity on color of the ripe fruit in *Capsicum*

Pigment quantity	*Pigment quality*	
	Red pigments present	*Red pigments absent*
Normal	Red	Orange-yellow
Slightly reduced	Tangerine	Lemon yellow
Trace amounts only	Pink	Cream

Table 3 Structure and pungency of capsaicinoids compared to pungent principles of ginger and black pepper

Name	*Structural formula*	*Threshold pungency (10^5 Scoville units)*
N-*vanillyl alkyl amides* (*from* Capsicum)		
Capsaicin	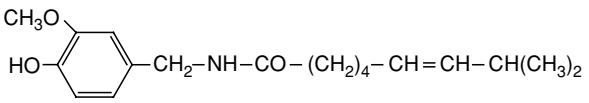	160
Dihydrocapsaicin	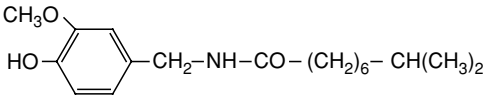	160
Nordihydrocapsaicin	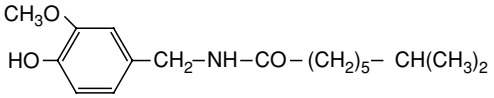	91
Homodihydrocapsaicin	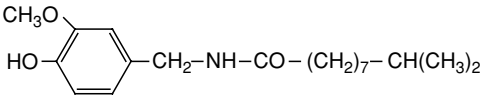	86
Homocapsaicin	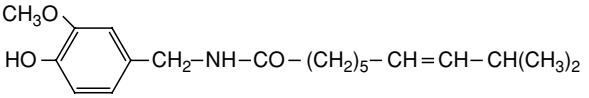	86
N-vanillylnonamide	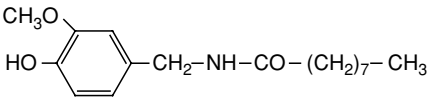	92
N-*vanillyl alkyl ketones (from ginger)*		
Gingerol	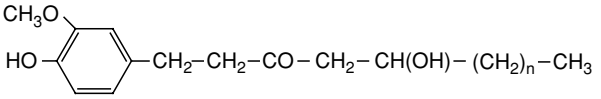	0.8
Shogaol	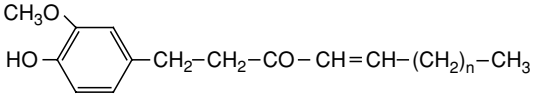	1.5
Substituted piperidine (from black pepper)		
Piperine	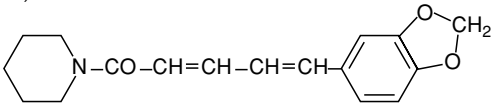	1.0

dihydrocapsaicin, hence do not taste pungent. It is not known whether they are intermediates in the biosynthesis of capsaicin and dihydrocapsaicin.

Continuous light has been reported to induce formation of capsaicinoids during postharvest ripening of nonpungent *C. annuum*, and cells isolated from sweet pepper stem callus will synthesize capsaicin *in vitro* under appropriate conditions. The nonpungent mutant, at least in *C. annuum*, may therefore involve some change in a regulatory or switch mechanism affecting the pathway of capsaicin biosynthesis, rather than a simple block in this pathway.

Although many secondary metabolites of the Solanaceae (such as nicotine) are produced wholly or partly in the roots, this is not true of the capsaicinoids. Experiments in which tomatoes were grafted on to hot pepper rootstocks, and vice versa, showed that capsaicinoids are produced only in the fruits. They usually appear about 2 weeks after flowering and reach a plateau 4 weeks after flowering. Temperature, particularly night temperature, seems to affect both the formation and accumulation of capsaicinoids. Capsaicinoids are synthesized in the epidermal cells of the placenta, in the inner compartment of the endoplasmic reticulum. They are translocated across the cell membrane, through the cell wall, and accumulate in cavities beneath the cuticle covering the placenta. This may cause the cuticle to crack and the capsaicinoids to spread over the seeds and inner wall of the fruit, leading to the widespread, but erroneous, belief that pungent peppers have pungent seeds and pungent fruit walls.

For a long time, pungency of *Capsicum* was assessed organoleptically by the Scoville test, which was simple and convenient but not accurate or reproducible. A specified weight of *Capsicum* was combined with a specified volume of alcohol, diluted with sugar solution to the threshold of taste, and the reciprocal of the dilution was the measure of pungency in Scoville heat units (150 000 Scoville units = 1% capsaicinoids). The major techniques now used for separating and quantifying the various capsaicinoids are gas chromatography, high-performance liquid chromatography, and high-performance thin-layer chromatography. There is still no simple nonorganoleptic test that can be applied in the field. (*See* **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography.)

Pharmacodynamic Properties

Capsaicinoids are responsible for many pharmacodynamic as well as organoleptic properties of *Capsicum*. Hot peppers stimulate saliva flow and may overcome loss of appetite. They also stimulate gastric juices, thereby aggravating stomach ulcers. They increase peristalsis, thus having a laxative action (which has benefited archeologists studying food remains in prehistoric feces). Hot peppers increase perspiration, which may partially explain their popularity in hot climates. Chilli-containing foods reputedly reduce risks of thromboembolism and significantly lower both liver and serum triglycerides. Capsaicinoids induce coughing and sneezing, inflame the skin, and have an irritating effect on mucous membranes of the eyes and nose. This was exploited in the past by some Amerindians, who punished their children by holding them over the smoke of a fire on to which dried chillies had been thrown, and exploited more recently by manufacturers of certain aerosols sold for use against 'muggers.'

The burning and pain associated with hot peppers result from pungent capsaicinoids acting on sensory neurons known as nociceptors. Repeated exposure leads to desensitization. Experimenters who desensitized their tongues to capsaicin found that their taste thresholds for other pungent compounds, such as ginger and mustard, also increased, but their ability to perceive tactile stimuli or basic tastes, such as sweet, salt, sour, or bitter, was not affected. Capsaicin apparently acts via a receptive site in the nociceptor. This site seems also to be involved in the perception of temperatures which are dangerously high (perhaps explaining why pungent foods are perceived as 'hot'). Capsaicin kills the nociceptor, or destroys its peripheral terminals. This has been exploited in the topical use of capsaicin as an analgesic to treat conditions such as shingles and rheumatoid arthritis.

Nutritional Value

Hot peppers are traditionally valued for adding interest to bland, starchy foods and for masking off-flavors in meat and other stored products. Both hot and sweet peppers are also excellent sources of vitamins, particularly vitamins A and C. Carotene (provitamin A) is deficient in many tropical diets. Green peppers have a β-carotene equivalent of 180 μg per 100 g; red peppers contain 4770 μg of carotenoids per 100 g. (*See* **Retinol**: Physiology.)

Fresh peppers are perhaps even more important as a source of vitamin C. The Hungarian chemist Szent-Györgyi won the Nobel Prize for his isolation of vitamin C, first from the adrenal gland, later from paprika and lemon. Fresh paprika fruits contain up to 340 mg per 100 g, i.e. more vitamin C per unit weight than citrus. Much of this is lost when peppers are dried (paprika powder contains only 30–60 mg per 100 g), but about two-thirds is retained in canned *Capsicum* fruit. (*See* **Ascorbic Acid**: Physiology.)

Harvesting, Handling, and Processing

Green bell peppers and fresh green chillies are picked when the fruit has reached its full size but the seeds are not yet fully mature. This is usually about 1 month after flowering, or 70 days after planting. Fruits are picked by hand at 7–14-day intervals over a harvesting period of about 3 months. They are usually picked with calyx and fruit stalk attached, since bacteria and fungi may infect the scar left by removal of the calyx and cause postharvest rots. Harvested fruits can be stored for up to 14 days, preferably under cool humid conditions (7–10 °C and 95% relative humidity). Ventilation may be advisable to remove accumulating ethylene, which would accelerate fruit ripening. Ripe peppers are harvested 2–3 weeks later than green peppers but are handled in the same way. Since the grower has to wait longer before harvest, ripe fruits usually command higher prices than green peppers.

Sweet peppers may also be marketed as dehydrated flakes. Stalks, calyces, placentas, and seeds are removed mechanically and the pericarp is diced, sprayed with sulfite-bisulfite solution and dried by hot air. Pimentos for canning have the tough outer cuticle plus the underlying epidermis removed by roasting or by treatment with lye, then are cored to remove placenta and seeds and canned either whole or diced.

The quality of *Capsicum* powders depends on their color and flavor, which are influenced by how the fruits are harvested and handled. Paprika is always made from red-fruited cultivars. The genotype influences both the quantity of pigment present and how well it is retained after harvest. Traditionally, fruits are harvested when fully ripe and dried in the open air for 3 weeks or more. The red pigments increase for the first 25 days of after-ripening, then remain constant until about 40 days after harvest, when they start to break down. Mechanical harvesting requires that 80–90% of the fruits mature simultaneously. In Hungary, plants are cut by machine at ground level and laid in the field to dry and after-ripen. Artificial drying in hot air is also widely used, but if temperatures exceed 80 °C, fruit color starts to deteriorate. Stalks, calyces, placentas, and seeds are separated from the dried fruits because they would dilute the color, but at least 5% of cleaned seeds are then re-added to facilitate grinding (the precise percentage depends on the intended grade of the final product). The seeds contain fat, and carotenoids are fat-soluble, so adding ground seeds helps to disperse the color evenly, though the fat may become rancid during storage. (*See* **Drying**: Theory of Air-drying.)

Pungent fruits are handled in much the same way as paprika. Grinding is done mechanically, formerly in the importing countries, but costs are high and grinding of chillies is particularly unpopular because of the cough- and sneeze-provoking effects of the dust, so grinding is increasingly left to the producing countries. The ground powder is fumigated to control microorganisms and stored under dry, cool, dark conditions to minimize breakdown of the red pigments. (*See* **Fumigants**.)

Oleoresins are extracted from dried and ground fruits. If the seeds are ground with the fruits, fat from the seeds dilutes both color and pungency, and hence quality, of the oleoresin. However, if the seeds are removed, the cost of production is increased and the yield of oleoresin is decreased. Acetone and ethylene dichloride are frequently used solvents. The extract is distilled to remove the solvent, leaving the concentrated oleoresin.

Global Distribution, Products, and Commercial Importance

Capsicum is adapted to tropical or subtropical climates, but can be grown in temperate climates in glasshouses or under protected cultivation. The Food and Agriculture Organization (FAO) production yearbooks do not distinguish between *C. annuum* and other species, or between pungent and nonpungent peppers. Over the last two decades, total world production of chillies and green peppers (presumably including also ripe fruits of sweet pepper and pimento) increased from about 7 million to nearly 17 million tonnes. Over half of this is produced in Asia. China alone produces over 40% of the world total.

Data in the FAO trade yearbooks combine, under the heading pimento, information for capsicum, cayenne, chillies, paprika, and red pepper. The confusion surrounding the name chilli has already been discussed. Capsicum is a similarly confusing term, which may be used for the nonpungent vegetable peppers, but may also be used for some mildly pungent dried fruits, e.g. Ancho, that are larger than most fruits known as chillies. Applied to the oleoresin, capsicum signifies the most pungent, rather than the least pungent, of the commercial grades. Cayenne is usually made from small-fruited peppers, grown in various parts of the world, and is usually very pungent. However, in the USA cayenne is made from locally grown, large-fruited cultivars and is less pungent. Paprika comes mainly from Hungary, Spain, and the USA and is usually, though not always, nonpungent. Ground red pepper and crushed red pepper have similar names, but ground red pepper is usually milder than crushed red pepper. Red pepper may also refer to ripe fruits of some sweet peppers.

The FAO trade yearbook shows that, in 1997, North America and Europe were net importers of over 100 000 t of 'pimento.' Asia and Africa were net exporters. China contributed about 40% of Asia's exports, thereby earning over US$72 million, while India contributed a further 30% and earned approximately US$40 million.

Oleoresins are valued by some commercial food processors for their greater hygiene (less contamination by microorganisms, insects, or rodent droppings) and more precise standardization of pungency. They are used mainly in the USA and the UK, but elsewhere do not seem to be replacing pepper powders. Three different oleoresins are produced commercially: oleoresin paprika, oleoresin red pepper (or oleoresin chilli), and oleoresin capsicum. Oleoresin capsicum is used mainly in the pharmaceutical industry but also in the production of some foods and beverages, such as ginger ale. One kilogram of oleoresin capsicum replaces approximately 20 kg of cayenne. Oleoresin red pepper is made from less pungent fruits, such as the long, thin peppers much grown in Asia. One kilogram of oleoresin red pepper replaces about 10 kg of red pepper. Oleoresin paprika is used for coloring salad dressings and oleomargarines, partly because it retains its color longer than annatto. One kilogram of oleoresin paprika replaces 12–15 kg of paprika powder.

Probably less than 10% of total global production of hot peppers is for export. The remainder is grown for local use. Mexico leads the world in chilli consumption at 15 g per head per day. Sweet peppers are eaten less regularly than hot peppers but are consumed in larger quantities (up to 20 g per meal). At these levels of intake, *Capsicum* adds considerably more than spice to the diet of some of the world's poorest people.

See also: **Ascorbic Acid**: Physiology; **Chlorophyl**; **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Drying**: Theory of Air-drying; **Fumigants**; **Retinol**: Physiology; **Ripening of Fruit**; **Sensory Evaluation**: Aroma

Further Reading

Andrews J (1984) *Peppers: The Domesticated Capsicums*. Austin, Texas: University of Texas Press.

Camara B, Hugueney P, Bouvier F, Kuntz M and Monéger R (1995) Biochemistry and molecular biology of chromoplast development. *International Review of Cytology* 163: 175–247.

Caterina MJ, Schumacher MA, Tominaga M, Rosen TA, Levine JD and Julius D (1997) The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* 389: 816–824.

Govindarajan VS (1985) *Capsicum* – production, technology, chemistry and quality. Part I: history, botany, cultivation and primary processing. *CRC Critical Reviews in Food Science and Nutrition* 22: 109–176.

Govindarajan VS (1986) *Capsicum* – production, technology, chemistry and quality. Part II: processed products, standards, world production and trade. *CRC Critical Reviews in Food Science and Nutrition* 23: 207–288.

Govindarajan VS (1986) *Capsicum* – production, technology, chemistry and quality. Part III: chemistry of the color, aroma and pungency stimuli. *CRC Critical Reviews in Food Science and Nutrition* 24: 245–355.

Govindarajan VS and Sathyanarayana MN (1991) *Capsicum* – production, technology, chemistry and quality. Part V: impact on physiology, pharmacology, nutrition and metabolism; structure, pungency, pain and desensitization. *CRC Critical Reviews in Food Science and Nutrition* 29: 435–474.

Govindarajan VS, Rajalakshmi D and Chand N (1987) *Capsicum* – production, technology, chemistry and quality. Part IV. Evaluation of quality. *CRC Critical Reviews in Food Science and Nutrition* 25: 185–282.

Kobata K, Todo T, Yazawa S, Iwai K and Watanabe T (1998) Novel capsaicinoid-like substances, capsiate and dihydrocapsiate, from the fruits of a non-pungent cultivar, CH-19 Sweet, of pepper (*Capsicum annuum* L.). *Journal of Agricultural and Food Chemistry* 46: 1695–1697.

Purseglove JW, Brown EG, Green CL and Robbins SRJ (1981) *Spices*, vol. 1. London: Longman.

Somos A (1984) *The Paprika*. Budapest: Akadémiai Kiadó.

Suzuki T and Iwai K (1984) Constituents of red pepper species: chemistry, biochemistry, pharmacology and food science of the pungent principle of *Capsicum* species. In: Brossi A (ed) *The Alkaloids*, vol. 23, pp. 227–299. Orlando, Florida: Academic Press.

Peptic Ulcer *See* **Colon**: Structure and Function; Diseases and Disorders; Cancer of the Colon

PEPTIDES

D González de Llano, Centro Nacional de Biotecnologia, Cantoblanco, Madrid, Spain
C Polo Sánchez, Instituto de Fermentaciones Industriales (CSIC), Madrid, Spain

Introduction

Peptides consist of chains of amino acids linked to each other by amide bonds, also known as peptide bonds. They form a heterogeneous group of compounds, because of the large number of different naturally occurring amino acids and the high potential variability in the order and number (chain length) of residues making up the peptide.

The number of amino acids that a peptide contains is indicated by a prefix, e.g., di-, tri-, tetrapeptide; the term 'oligopeptides' or 'low-molecular-weight peptides' refers to peptides with 10 or fewer residues, while 'polypeptides' is used for peptides with higher molecular weights. Although the transition point from polypeptide to protein is not well defined, proteins are normally considered to have at least 100 residues (mol wt > 10 000). By convention, peptides are referred to by the first three letters of the amino acids making up the chain or by one-letter abbreviations. Also by convention, the amino acid with the free amino group is referred to as the *N*-terminus and is represented on the left, while the amino acid with the free carboxyl group is referred to as the *C*-terminus and is represented on the right. (*See* **Amino Acids**: Properties and Occurrence.)

The *N*-terminal amino group and the *C*-terminal carboxyl group on peptides react chemically in the same way as the α-amino and α-carboxyl groups on free amino acids, and they can be used for detection and quantification purposes.

The acid–base behavior of peptides is dependent upon the free α-amino group on the *N*-terminal residue, the free α-carboxyl group on the *C*-terminal residue, and the ionizable R groups located at intermediate positions.

In the presence of carbonyl compounds, certain peptides take part in the Maillard reaction, leading to the formation of melanoid pigments that contribute to the development of undesirable coloration in foodstuffs. Conversely, there are also lysine-containing peptides that retard browning reactions with glucose, and these may be appropriate for fortifying the lysine content of sugar-containing foods to be cooked. (*See* **Browning**: Nonenzymatic.)

Peptides also play a very important role in determining the rheological properties of foods. Hydrolysis of proteins to peptides during processing (fermentation, ripening, and cooking) of foods may modify food texture.

Peptides in Foods

Peptides are naturally present in foodstuffs. Most simple peptides are the result of partial hydrolysis of protein polypeptide chains. On the other hand, nonprotein peptides have also been recorded in foods. Such peptides usually differ in structure from the peptides derived from proteins, and these structural variations may protect them from the action of peptidases.

The tripeptide glutathione, present in the cells of all higher animals, is an example of a nonprotein peptide. It contains a glutamic acid residue linked by an unusual peptide bond involving its γ-carboxyl group. Glutathione plays a role in the active transport of amino acids, acts as an antioxidant for lipids and as an activator for certain enzymes, and is a coenzyme of glyoxalase. Other nonprotein peptides include the dipeptides carnosine, anserine, and balenine, which are present in vertebrate muscle and contain a β-amino acid (β-alanine) bound to L-histidine or 1-methyl-L-histidine. These histidine dipeptides are present in differing proportions in the muscles of different species and therefore are a mean of determining the source of the meats used in meat or of identifying marine fish species.

Peptides may also be present in foods because they have been used as additives (e.g., sweeteners, flavor enhancers, or bulking agents for light beverages) to improve food quality. Some peptides produced by the enzymatic hydrolysis of proteins possess better functional properties than the parent proteins and consequently are used by the food industry for a variety of purposes.

Peptides have lower molecular weight and less secondary structures, as well as higher number of ionizable groups and exposure of hydrophobic groups than native proteins. These facts imply that solubility, surface activity, foaming, and emulsifying properties may be different from that of the intact protein.

Enzymatic hydrolysis has been reported to enhance the emulsifying capacity of different proteins. Peptides with an inhibitory effect on autoxidation of linolenic acid and hence potentially useful as antioxidants have been found in fermented foods and in protein hydrolysates. Moderately large peptides have

pronounced bulking properties and large peptides act as foam stabilizers; because of these interesting functional properties, enzymatic hydrolysates of soya bean proteins have been used in beverages. Hydrolyzed hydrophobic proteins from gluten (zeins, gliadins) have been prepared with a view to increase solubility and diversify their functional properties.

Analysis of the peptides of enzymatic hydrolysates of trypsin can be used to confirm the presence of nonmeat proteins such as soyabean proteins in meat products. Tryptic peptides of the caseins from milks of different species are distinct and may be used to detect blends of milks. The presence of the casein glycomacropeptide is also useful in detecting adulteration of powdered milks with rennet whey.

Organoleptic Properties

Peptides are tasteless or bitter, with the exception of certain dipeptides of glutamic acid and aspartic acid, which are sweet. Others, like the lower alkane members in the series of methyl esters of L-aspartyl-α-amino-cycloalkane carboxylic acid and a series of L-aspartyl-D-alanyl tripeptides, are sweet but turn bitter or tasteless when the size of the *C*-terminal amino acid ring increases. There is, thus, an important relationship between size and flavor in these peptides. Peptides belonging to L-aspartyl D-alanyl amides are strong sweeteners and may be used as good sugar substitutes. (*See* **Sensory Evaluation**: Practical Considerations.)

Foods may also contain peptides with a flavor similar to that of monosodium glutamate, termed umami taste or relish, which are able to mask the bitter taste produced in foods by bitter peptides or other bitter substances. This group includes hydrophilic peptides, in particular acid peptides that balance flavor in foods.

Warmed-over flavor in cooked–stored–recooked meat has also been related to the higher hydrophobic peptide content in such meats as compared to cooked meats. However, low-molecular-weight peptides play an important role in the flavor formation and intensity of meat, beef broth, and dry cured ham.

Bitter peptides form as a result of proteolytic reactions that take place in foods. These bitter peptides have been identified in soyabean and casein hydrolysates as well as in foods, like cheese, that undergo fermentation of ripening processes with marked proteolytic activity. The bitter taste has been related to the hydrophobic amino acid content and to chain length. An overly high proportion of hydrophobic amino acids gives rise to bitterness; however, above a given molecular weight bitter flavors are no longer perceptible, even though hydrophobic amino acids are present. Proteins do not have a bitter taste even when they contain hydrophobic amino acids, yet hydrolysis of such proteins may indeed yield bitter peptides. Bitter flavors in di- and tripeptides have been observed to increase when the hydrophobic amino acid is located on the *C*-terminal residue and the basic amino acid is on the *N*-terminal residue. Enzymatic activity is extremely important in the development of bitter flavors in foods. Trypsin does not degrade bitter peptides, although peptidases may break down bitter peptides to nonbitter lower-molecular-weight peptides and amino acids.

Bitterness is the most frequent flavor defect limiting cheese acceptability. It has been attributed to the formation and accumulation of bitter peptides derived from casein. During cheese manufacture rennet or rennet substitutes and microbial proteases are the principal agents hydrolyzing the caseins and are thus thought to be responsible for the occasional appearance of bitter taste. The use of rennet substitutes may give rise to the appearance of bitter peptides; for instance, commercial chicken pepsin produces an objectionable level of bitterness in cheese. On the other hand, cheeses made from goats' or ewes' milk are less susceptible to bitter flavor defects than cheeses made from cows' milk.

Aged cheeses, in particular hard cheeses, require long ripening periods in which to acquire the desired organoleptic characteristics. Proteases have therefore been used to accelerate the ripening process and lower costs. However, the proteases give rise to bitter peptides that limit cheese quality. In recent years a considerable amount of work has been carried out on peptidases that break down the bitter peptides derived from the caseins to smaller, nonbitter peptides. The combined use of proteases and peptidases may be an effective means of shortening ripening times, while preventing the development of bitter flavors. (*See* **Cheeses**: Chemistry and Microbiology of Maturation.)

Bitter peptides that contribute to bread flavor form during the fermentation and baking of bread dough as a result of the action proteolytic enzymes on the protein fraction in the flour (gliadins and glutelins).

Analytical Procedures

Spectrophotometric Methods

The *N*-terminal α-amino groups on peptides react quantitatively with reagents such as ninhydrin to form colored derivatives or with *o*-phthalaldehyde and fluorescamine to form fluorescent derivatives. These reactions are useful in making quantitative determinations of peptides in foodstuffs and for detection during chromatographic analysis. Due to the

peptide bond, peptides also take part in reactions which free amino acids do not undergo, like the classic biuret reaction. This reaction consists of the formation, in an alkaline medium, of a colored complex with a transition metal (e.g., copper or nickel) that can be measured quantitatively by spectrophotometry. This reaction is utilized in quantifying peptides. (*See* **Spectroscopy**: Visible Spectroscopy and Colorimetry.)

The peptide content can also he determined by assaying the aromatic amino acids tyrosine and tryptophan, which is readily achieved by measuring absorbance at 280 nm or by measuring the color reactions of these amino acids with certain reagents like Folin-Ciocalteu's reagent. Another analytical method for determining the high-molecular-weight peptide content is by precipitating the peptides with acid dyes (amino black, orange G) and measuring the excess dye remaining in solution. These methods are useful in determining the total peptide contents of foodstuffs, but their drawback is that they provide no information on how many or which peptides are present. For this, other techniques such as chromatography or electrophoresis are needed.

Chromatographic Methods

Paper chromatography and thin-layer chromatography have conventionally been used to separate peptides, particularly for the fractions isolated by gel permeation chromatography (GPC) and ion exchange chromatography (IEC), but they are lengthy, time-consuming procedures with low resolving power and low reproducibility.

Gel filtration chromatography on open Sephadex columns with different pore sizes (G-10, G-25, and G-50) using water or saline or acid buffers as solvents is used extensively in the fractionation of peptides. Detection is performed by measuring absorbance of the column eluate with an ultraviolet detector at 280 nm, the absorbance wavelength for tyrosine and tryptophan, or at 214 nm, the absorbance wavelength for most amino acids and the peptide bond. Peptides elute in fractions according to their molecular weight. Small peptides must normally be separated from the amino acids; gel filtration on columns prepared by hydroxypropylation (Sephadex LH-20), ion exchange on diethylaminoethyl (DEAE)-cellulose or DEAE-Sephadex with Cu^{2+} complexes, or ligand exchange columns with a stationary phase modified with Cu^{2+} are used for this purpose. Ion exchange chromatography on DEAE-cellulose or on Aminex A-5 is normally used for the separation of high-molecular-weight peptides, employing saline buffers as eluent. Detection is carried out at 280 nm or in the visible region after derivatization with ninhydrin.

These methods are now most often used as preparatory techniques for subsequent high-performance liquid chromatographic (HPLC) or electrophoretic analysis of the peptide fractions. (*See* **Chromatography**: Principles.)

High-performance Liquid Chromatography

Because of its versatility, short analysis times, high resolution, and effective separations, and because it is well suited to automation procedures, HPLC is the most widely used method of peptide analysis. Practically all known mechanisms have been employed in the chromatographic separation of peptides, e.g., separation based on molecule size (GPC), on charge (IEC), on hydrophobicity (reversed-phase and interaction chromatography), and even on combinations thereof; of these, however, reversed-phase chromatography is most commonly used to separate mixtures of peptides from foodstuffs (Figure 1). It affords the possibility of changing the stationary phase, the pH, the ionic strength of the aqueous buffer, or the type of organic modifier, of using different gradient shapes, and of working at a variety of temperatures, making this method suitable for analyzing highly different peptides. Moreover, the option of using microcolumns makes it possible to detect quantities on the order of picomols and has been an important step toward HPLC–mass spectrometry coupling. (*See* **Chromatography**: High-performance Liquid Chromatography; Combined Chromatography and Mass Spectrometry.)

Reversed-phase chromatography Reversed-phase columns with pore sizes of 6–10 nm are highly appropriate for separating small peptides. Columns with pore sizes of 30–50 nm give the best results when separating large peptides (mol wt 4000), because the molecules are able to pass freely through the pores, thus permitting greater access to the alkyl chains and increasing column efficiency and load capacity.

The mobile phase is commonly a mixture of water and an organic solvent, normally acetonitrile, methanol, or 2-propanol, although other organic solvents such as methoxyethanol, ethanol, butanol, or tetrahydrofuran are also used. Gradient elution is normally required for reversed-phase HPLC of peptides.

Because of their high polarity, peptides do not interact sufficiently with the hydrophobic chains of the stationary phase in reversed-phase chromatography. On the other hand, they do interact with any free silanol groups in these phases. Salts or strong acids must be added to the stationary phase to block the silanol groups. In order to reduce the polar nature of the peptides, chromatography should be carried

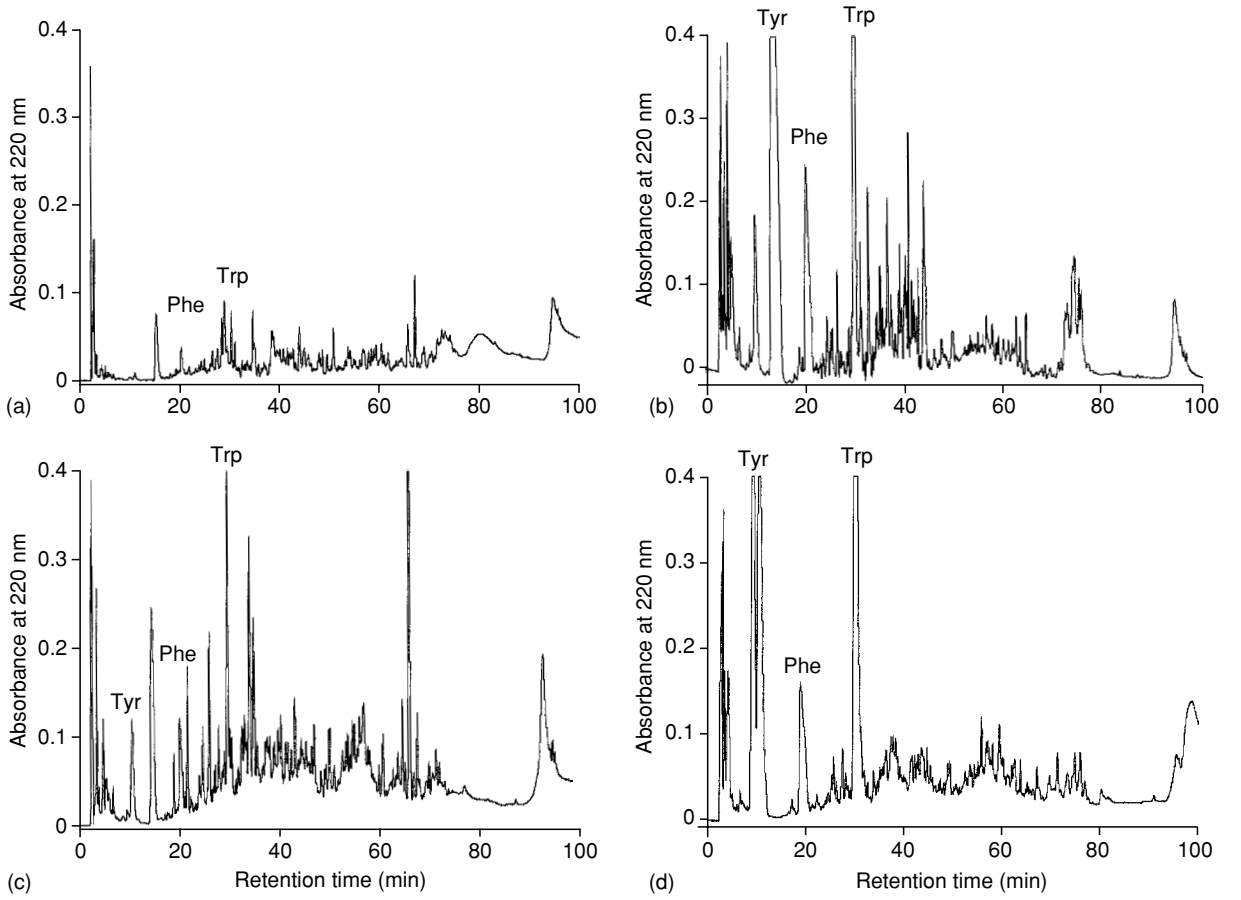

Figure 1 Chromatogram of the peptides of the water-soluble fraction from commercial (a) Vidiago, (b) Cabrales, (c) Beyos, and (d) Peral cheeses. Column: Ultraspher ODS. Solvent: TFA/acetonitrile/water. Gradient elution. Detection at 220 nm. From González de Llano D, Polo C and Ramos M (1995) Study of proteolysis in artisanal cheeses: high performance liquid chromatography of peptides. *Journal of Dairy Science* 78(5): 1018–1024 with permission.

out at a pH of less than 3, so that the carboxyl groups on the acids amino acids (aspartic acid and glutamic acid) are not in their dissociated form. However, at this pH the basic amino acids histidine, lysine, and arginine are charged and must be blocked by forming ion pairs. The selectivity of the chromatographic system can be modified depending on the pH and on the type of ion-pairing reagent used. Trifluoroacetic acid has been widely employed as the ion-pair reagent for peptides.

Ion exchange chromatography IEC has also been used in peptide separations. Two types of ion exchangers are utilized: cationic exchangers to separate neutral and basic peptides, and anionic exchangers to separate neutral and acid peptides. Peptide separations have been carried out on polymeric cationic exchange resins (polystyrene/divinylbenzene copolymers), and indeed automatic peptide analysers, similar to those used in amino acid analysis, have been designed, with postcolumn detection using ninhydrin.

Size exclusion chromatography (SEC) High-performance SEC has most commonly been applied to protein separations, though on occasion it has also been used to separate peptides from foodstuffs. The stationary phase may be inorganic, such as silica or alumina. The disadvantage of inorganic materials is that their active surfaces adsorb charged molecules and hence they must first be deactivated, for instance using organosilanes. Organic materials such as glycerol methacrylate, sulfonated styrene divinylbenzene, and polyesters with surface hydroxyl groups have also been used.

The different separation mechanisms referred to above (reversed-phase HPLC with small or large pore sizes, IEC, and SEC) are complementary, and the best separations are sometimes achieved by combining the various methods.

Detection Peptides are detected in amounts ranging from 100 to 1000 ng by absorbance at between 200 and 230 nm. Detection of peptides containing aromatic

amino acids (phenylalanine, tyrosine, and tryptophan) may be performed at 254 nm; tyrosine- or tryptophan-containing peptides can also be detected at 280 nm. Rapid-scanning ultraviolet/visible detectors based on diode array technology have proved an extremely useful aid in identifying peptides that contain aromatic residues. Wavelengths of the spectrum maxima, convexity interval, and wavelength of the second derivative spectrum maxima, all over in the 190–340 nm range, allow identification of the aromatic amino acids that form the peptides, as well as identification of HPLC coeluted compounds such as the cinnamic derivatives detected in the peptide fractions of wines.

Since numerous solvents and even other sample components may absorb light in the range 200–230 nm, formation of derivatives detectable at higher wavelengths that are more specific for given chromophores is commonly employed. To this end, phenyl isothiocyanate and dansyl chloride are frequently used in peptide analysis as derivatizing agents. Peptides may also be detected using a fluorescence detector to detect either the natural fluorescence of certain amino acids in the peptide sequence or the fluorescence of derivatives formed artificially using such reagents as fluorescamine and *o*-phthalaldehyde. Other reagents used lately for derivatization of peptides are naphthalene-2,3-dicarboxaldehyde, 3-(4-carboxy-benzoyl) 2-quinolinecarboxaldehyde, and 6-aminoquinoyl-*N*-hydroxysuccimidylcarbamate.

Following the significant advances made in liquid chromatography/mass spectroscopy coupling in recent years, mass spectrometry has turned into a detection system that holds great promise for peptide analysis, since it is capable of furnishing structural information and quantitative data that are otherwise difficult to obtain.

The greatest advance in the analysis of peptides has been the coupling with spectrometric techniques, both for identification and for characterization. Their high accuracy, sensitivity, and, in some instances, tolerance to solvents make mass spectrometers ideal detectors for analyses of HPLC-separated peptides. Presently, electrospray ionization and matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry – soft ionization methods – provide the suitable means to determine accurate mass on peptides and proteins with high sensitivity (subpicomol range). Furthermore, both methods are suitable for the creation of peptide ions for analysis using tandem mass spectrometry and collision-induced dissociation methods. Mass spectrometry in conjunction with database searching plays an increasingly important role in the characterization of food peptides.

Electrophoretic Techniques

Despite the important advantages afforded by HPLC in the analysis of hydrophobic peptides, electrophoresis can be useful, because insoluble peptides can be solubilized with detergents. Although a number of different supports have been employed, polyacrylamide gel electrophoresis (PAGE) and sodium dodecyl sulfate–PAGE (SDS-PAGE) have been widely used in the study of peptide formation from proteins during food processing. Another developed method of peptide analysis that may be used either an alternative or a complement to HPLC is capillary electrophoresis. This technique has extremely high resolving power and allows the separation of peptides differing in only a single amino acid; it is simple, fast, and appropriate for quantitative analysis, and only small amounts of sample are required. Peptide separation by CE is now almost a routine technique in food analysis. Moreover the application of two analytical techniques (HPLC-CE, SEC-CE) and coupling mass spectrometry to HPLC have acquired great importance for characterization of complex peptide samples. (*See* **Electrophoresis**: General Principles.)

Determination of Amino Acid Composition and Sequence in Peptides

The amino acid composition of peptides is generally assayed by carrying out acid hydrolysis with 6 mol l^{-1} hydrochloric acid, followed by determination of the individual amino acids by HPLC. Various types of HPLC have been employed to separate the amino acids, but reversed-phase chromatography on C_{18} columns is most commonly used. Detection of the amino acids normally involves derivatization, since maximum absorbance of amino acids is located in a region of the spectrum (214 nm) in which many other compounds also absorb. The most frequently used derivatizing agents are dansyl chloride, phenyl isothiocyanate and *o*-phthalaldehyde.

Peptide sequencing is normally performed by degradation of the *N*-terminal amino acid using phenyl isothiocyanate, Edman's reagent. The terminal amino groups react with the isothiocyanate, forming a phenylthiocarbamylic derivative. When treated with acid in an organic solvent, cyclization takes place and phenylthiohydantoin amino acid is formed; this can be separated from the rest of the chain, which remains intact. The process can be repeated.

The usefulness of Edman degradation in the sequencing of low-molecular-weight peptides is limited, because of the low repetitive yield in the successive sequencing cycles. For such peptides the best results are obtained by coupling sequential Edman degradation with detection by means of

dansylation or by the double coupling method using 4-*N*,*N*-dimethylaminoazobenzene-4′-isothiocyanate and phenyl isothiocyanate.

Mass spectrometry has also become a powerful tool for the sequencing of small amounts of peptides and proteins. Tandem mass spectrometry and ion source collision-induced dissociation produce information specific to the amino acid sequence and the specific covalently modified amino acid. A new method, referred to as protein ladder sequencing, for the *N*-terminal sequencing of peptides utilizes multiple steps of partial Edman degradation chemistry prior to the analysis of the reaction mixture by MALDI-TOF mass spectrometry.

See also: **Amino Acids**: Properties and Occurrence; Determination; **Chromatography**: Principles; Thin-layer Chromatography; High-performance Liquid Chromatography

Further Reading

Gonzalez de Llano D, Herraiz T and Polo MC (1996) Peptides. In: Nollet L (ed.) *Handbook of Food Analysis*, vol. 1, pp. 229–276. New York: Marcel Dekker.

González de Llano D, Polo C and Ramos M (1995) Study of proteolysis in artisanal cheeses: high performance liquid chromatography of peptides. *Journal of Dairy Science* 78(5): 1018–1024.

Matsudaira PT (1989) *A Practical Guide to Protein and Peptide Purification for Microsequencing*. San Diego: Academic Press.

Polo MC, González de Llano D and Ramos M (1992) Derivatization and liquid chromatographic separation of peptides. In: Nollet L (ed.) *Food Analysis by HPLC*, pp. 117–140. New York: Marcel Dekker.

Polo MC, González de Llano D and Ramos M (2000) HPLC of peptides. In: Nollet L (ed.) *Food Analysis by HPLC*, pp. 99–125. New York: Marcel Dekker.

Persimmon *See* **Fruits of Tropical Climates**: Commercial and Dietary Importance; Fruits of the Sapindaceae; Fruits of the Sapotaceae; Lesser-known Fruits of Africa; Fruits of Central and South America; Lesser-known Fruits of Asia

PESTICIDES AND HERBICIDES

Contents

Types of Pesticide

R Pleština, World Health Organization, Geneva, Switzerland

Introduction

Humans have always had to cope with disease, discomfort, and great economic loss caused by pests. Some methods for pest control date back to ancient times. What may be called the 'pesticide revolution' dates back to the early 1940s when dichlorodiphenyltrichloroethane (DDT) was first used as an insecticide. Since that time, in a variety of pest-control activities around the world, both in agriculture and in public health, chemicals continue to play a significant role, and this trend will be sustained for many years to come. Although pesticides are poisons, the majority of people recognize the advantages of their use in view of the benefits they bring to the control of disease vectors and nuisance pests, and particularly to the increase in food and fiber production. Research

on alternatives has produced few materials to take the place of these chemicals. Although the use of pesticides has more or less stabilized, or is even declining in the developed world, it continues to increase in the developing countries owing to development and buildup of pest resistance, because of an increase in the need for food, feed, and fibers, in addition to a limited and constantly decreasing per capita availability of arable land. (*See* **Insect Pests:** Insects and Related Pests; Problems Caused by Insects and Mites.)

There is little doubt that the principles promoted by the concept of integrated pest management will reduce excessive use of pesticides. However, it should be stressed that integrated pest management does not exclude pesticide use, and pesticide treatment should be considered as a supplement to basic sanitation; like drugs, pesticides should always be used with discretion and in conjunction with other measures in order to achieve effective pest control.

Definition and Classifications

The term 'pesticide' is defined in many different ways, each having some specific feature. Most of them, however, define pesticide as a pest-control agent aimed at killing or repelling a pest. The most widely accepted definition is that given in the *International Code of Conduct on the Distribution and Use of Pesticides* (FAO, 1990) which reads as follows:

> Pesticide means any substance or mixture of substances intended for preventing, destroying or controlling any pest, including vectors of human or animal disease, unwanted species of plants or animals causing harm during or otherwise interfering with the production, processing, storage, transport, or marketing of food, agricultural commodities, wood and wood products or animal feedstuffs, or which may be administered to animals for the control of insects, arachnids or other pests in or on their bodies. The term includes substances intended for use as a plant growth regulator, defoliant, desiccant, or agent for thinning fruit or preventing the premature fall of fruit and substances applied to crops either before or after harvest to protect the commodity from deterioration during storage and transport.

Similarly, classifications of pesticides are numerous, depending on the criteria used and their use. Thus, based on the pest they control, pesticides may be grouped as insecticides, fungicides, herbicides, rodenticides, repellents, etc., and based on the chemical structure of the active ingredient, they may be grouped as organophosphates, carbamates, organochlorines, etc. There is also a clear distinction between agricultural and public health pesticides or household pesticides. By formulation, the products may be grouped as liquids, solids, or gases. In addition, there are persistent and degradable pesticides. One classification does not exclude another; all of them are used to make a classification according to the hazard that pesticides pose to humans, and this arises from their intrinsic toxic property. Thus, the World Health Organization elaborated *The WHO Recommended Classification of Pesticides by Hazard* and *Guidelines to Classification 1992–1993* (Table 1). (*See* **Fungicides**.)

This document was approved by the 28th World Health Assembly in 1975 and has since gained wide recognition in a number of Member States and by pesticide registration authorities. Although the classification takes into account acute oral or dermal toxicity, whichever is higher, it also considers any irreversible effect that might be recognized. For practical reasons, a number of pesticides classified as Class III (slightly hazardous) are listed in a separate table as 'Unlikely to present acute hazard in normal use.' *Guidelines to Classification* is prone to frequent revision, based on documented scientific evidence.

Needs and Present-day Use

The costs and benefits of pesticides are currently being evaluated, on the one hand, for their ability to reduce the cost of food and feeds, and to decrease the spreading of vectorborne diseases, and, on the other hand, for the potential impact that pesticides may have on human health and the environment. All of these aspects should not, however, be considered

Table 1 WHO recommended classification of pesticides by hazard

Class	*LD50 for the rat (mg per kilogram of body weight)*			
	Oral		*Dermal*	
	Solids[a]	*Liquids*[a]	*Solids*[a]	*Liquids*[a]
Ia (extremely hazardous)	5 or less	20 or less	10 or less	40 or less
Ib (highly hazardous)	5–50	20–200	10–100	40–400
II (moderately hazardous)	50–500	200–2000	100–1000	400–4000
III (slightly hazardous)	Over 500	Over 2000	Over 1000	Over 4000

[a]The terms 'solids' and 'liquids' refer to the physical state of the product or formulation being classified.
From WHO (1992).

separately from the costs which must be paid to ensure the correct use of pesticides.

It has been estimated that if pesticides were not used, the potential loss of food production would be 45% (30% attributed to pests, including weeds and diseases before harvest, and 15% as a postharvest loss). In addition, in over 100 countries, more than 4×10^8 clinical cases of vectorborne diseases currently exist. For some of these diseases, control of vectors remains the main control measure, and for others, pesticides are considered to be an essential part of disease control. The use of pesticides in the control of household pests should not be neglected, regardless of whether they are used in the control of vectors of diseases or to control nuisance pests. This aspect of pesticide use is of great importance, as pesticides intended for this kind of application are available to the general population.

Although the exact figures for the global use of pesticides, either according to intended use (agriculture, public health, household) or according to regional or country use, are not available, it is certain that the quantities of pesticides used are steadily growing, albeit less in the developed countries than in developing countries, and this trend may continue for some years. A few highly developed countries show a decline in the quantities used as more efficient compounds are applied, and the trend to organic farming increases. (*See* **Organically Farmed Food**.)

In western Europe and North America, herbicides are the most represented groups of pesticides, whereas in the tropics, insecticides are used more than other groups of pesticides. This fact gives rise to ever-growing concern, mainly attributable to the indiscriminate use and frequently to the misuse of pesticides. This concern is supported by the fact that, although only about 20% of pesticides are used in developing countries, they account for the majority of total deaths caused by pesticide poisonings.

Evaluation of the Risk and Testing Procedure

In general pesticides, in common with drugs, are among the most carefully studied chemicals, and this makes the toxicological evaluation of active material possible and meaningful.

According to the widely accepted *International Code of Conduct on the Distribution and Use of Pesticides*, pesticide manufacturers are expected to ensure that each pesticide is adequately and effectively tested in accordance with sound scientific procedures and good laboratory practice. Generated toxicological data and use patterns are then meticulously evaluated by competent international or national institutions, and results made available to the relevant national authorities in order to allow them to introduce the necessary legislation for their regulation, including registration. (*See* **Legislation**: Contaminants and Adulterants.)

In many countries, registration procedures are elaborated upon in great detail, offering possibilities, for preparing precise precautionary measures for their production, transport, storage, and use. Unfortunately, in many areas, the implementation of carefully prescribed safety measures is lacking, and pesticides are greatly misused.

The activities of the WHO in the safety assessment of pesticides were initiated soon after its foundation, and a number of expert committees have been devoted to the toxic hazards arising from pesticides used in public health programs. Subsequently, the experts also became concerned with the safety aspects of pesticides used in agriculture. The evaluation of pesticide residues in food started in 1963 in collaboration with the Food and Agriculture Organization (FAO). Since then, the Joint Meeting of Pesticide Residues in Food (JMPR) has evaluated a large number of pesticides. The main objectives of the WHO Expert Group on Pesticide Residues (the WHO component of JMPR) are consistent with those of the International Programme on Chemical Safety and include the formulation of guiding principles for exposure limits such as acceptable daily intake (ADI) for pesticide residues in food. Tolerances for these substances in air, water, soil, and the working environment are recommended by other WHO expert groups. (*See* **World Health Organization**.)

Although the requirements for toxicological data needed for health-risk assessment vary considerably from country to country, they comprise short- and long-term toxicity studies in several species of experimental animals, data on absorption, distribution, metabolism and excretion of a pesticide, carcinogenicity, reproduction, and genotoxicity, and other special studies for particular classes of compounds.

Evaluation of data is an extremely complicated process, and results are not always equivocal. This is particularly so because extrapolation of animal data to humans is mainly based on the introduction of a safety factor and, when possible, on pharmacokinetic extrapolation. Whenever applicable, observations on humans following occupational and/or accidental exposure are incorporated within the health-risk evaluation process.

Biological Activity of Pesticides

The inherent toxicity of a pesticide to humans is the same regardless of its use, either in agriculture or

vector control activities, or when it is consumed as a residue in food. The difference in untoward effect might occur as a result of the differences in the route and extent of exposure; these influence the dose absorbed, which is most decisive in producing an ill effect.

Of more than 800 chemicals used as pesticides, over 100 have become obsolete or their use has been discontinued because of inefficacy or an unacceptably high risk to human health and to the environment.

Pesticides are, by nature, toxic to some degree to several forms of life, including mammals. However, some of them are such potent toxins that they are considered to be extremely hazardous chemicals. Like other biologically active substances, the active material of a pesticide has to come into contact with its target (such as a cell or an enzyme) in sufficient concentration before it can exert an effect. Therefore, the physical characteristics of the active material, its penetration ability, stability, distribution, persistence, and degradation in the body, etc., considerably influence the likelihood of appearance of untoward effects in humans. These characteristics vary considerably not only from one pesticide to another but also among various preparations of the same active material.

Diversity of chemical structures among pesticides leads to the diversity of the biological effects produced both in pests and in humans; only some of these are sufficiently well understood, and these groups will be described briefly. Several examples of various classes of pesticides are listed in Table 2.

Anticholinesterases

Anticholinesterases are frequently powerful poisons for both insects and mammals. Some of these compounds are too toxic to be used as pesticides (e.g., war gases), while others may have beneficial effects and are used as drugs in clinical practice.

The primary effects of anticholinesterase insecticides, both in insects and in humans, are attributable – entirely or in part – to the inhibition of an enzyme, acetylcholinesterase, in nervous tissue, which leads to the accumulation of acetylcholine, an important chemical transmitter of nerve impulses. If this happens, complex nervous system functions no longer behave normally, causing disturbances in glands, muscles, and parts of the brain.

There are two main groups of anticholinesterase insecticides: organophosphorus compounds (esters, amides or thiol derivatives of phosphoric, phosphonic, phosphorothioic, or phosphonothioic acids) and monomethyl carbamates (methyl N-substituted esters of carbamic acid).

Organophosphorus compounds can be grouped into two categories: those which can act on cholinesterases directly without being modified, and those which first have to be transformed in the body and are therefore called indirect inhibitors. This difference has great importance in the assessment of the hazard. After overexposure to direct inhibitors, signs and symptoms usually appear fairly quickly. Indirect inhibitors produce a more delayed and prolonged action, and when symptoms appear, although exposure might have ceased, they may develop further to cause a critical illness.

Monomethyl carbamate insecticides are all direct inhibitors of cholinesterases. As a rule, they are readily absorbed from the gut, rapidly metabolized, and rapidly excreted. The inhibition of the enzyme by carbamates is very short, rarely exceeding several hours. Therefore, the symptoms of carbamate poisoning, although they may be of rapid onset and severe, are shorter in duration than after poisoning with organophosphorus compounds.

Chlorinated Hydrocarbons

Organochlorine pesticides are more persistent in the environment than most other synthetic organic pesticides. For this reason, and because some insects have developed resistance to this class of compounds, their use is steadily decreasing. They are efficiently absorbed by the gastrointestinal tract, but some of them are also readily absorbed through intact skin.

Although the mechanism of action is not yet fully understood, the major toxic action of this group of chemicals is on the nervous system, both central and peripheral. They act by altering the electrophysiological properties of the cell membranes (particularly nerve axons) disturbing sodium and potassium ion exchange through the membrane. As this is a reversible process, organochlorine insecticides, which vary considerably in toxicity, have little or no obvious effect when dosages are small, even if exposure to them is of a very long duration. It should be borne in mind, however, that many of these insecticides or their degradation products are soluble in fat and can be stored in the body fat, where they remain apparently without effect if not mobilized.

However, their persistence, both in humans and in the environment, has caused public concern, since, in certain situations, they may pose health and environmental problems. In view of the extreme range of toxicities within the group, it is impossible to generalize further. This is particularly so because organochlorine compounds differ considerably in their stability in the environment in different climatic conditions.

Table 2 Examples of some chemical classes of pesticides

Chemical class of compounds	*ISO name*	*Structural formula*	*Primary use*	*Mode of action in mammals*	*Acute oral toxicity for rats (LD50; mg kg^{-1})*	*WHO classification by hazard*
Organo-phosphorus	Parathion	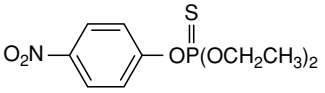	Insecticide	Inhibition of cholinesterases	13	Extremely hazardous
	Dichlorvos	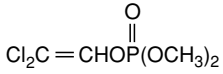	Insecticide	Inhibition of cholinesterases	56	Highly hazardous
	Malathion	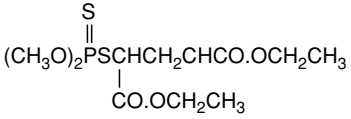	Insecticide	Inhibition of cholinesterases	2100	Slightly hazardous
Monomethyl-carbamates	Aldicarb	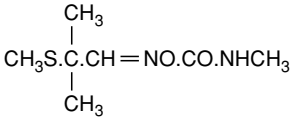	Insecticide	Inhibition of cholinesterases	1	Extremely hazardous
	Propoxur	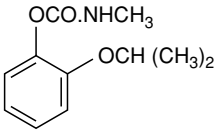	Insecticide	Inhibition of cholinesterases	95	Moderately hazardous
Chlorinated hydrocarbons	Dieldrin	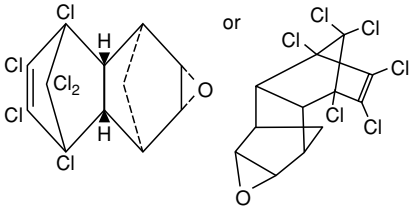	Insecticide	Disturbance of sodium (Na^+) and potassium (K^+) ion transport in nervous membranes	37	Highly hazardous

Continued

	DDT	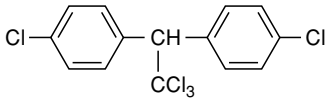	Insecticide	Disturbance of Na^+ and K^+ transport in nervous membranes	113	Moderately hazardous
Pyrethroids	Permethrin	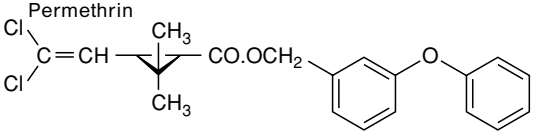	Insecticide	Disturbance of Na^+ and K^+ transport in nervous membranes	500	Moderately hazardous
	Deltamethrin	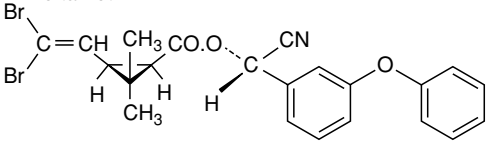	Insecticide	Disturbance of Na^+ and K^+ transport in nervous membranes	135	Moderately hazardous
Substituted phenols	Dinoseb	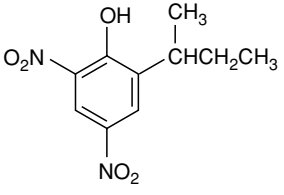	Herbicide	Electron transport inhibition	58	Highly hazardous
	DNOC	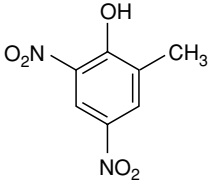	Insecticide	Electron transport inhibition	25	Highly hazardous
Bipyridinium compounds	Paraquat	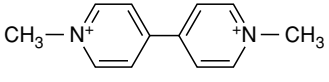	Herbicide	Excessive production of superoxide	150	Moderately hazardous

Continued

Triazines	Atrazine	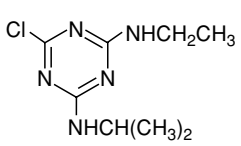	Herbicide	Nonspecific (irritant)	2000	Slightly hazardous
Chlorophenoxy	2, 4-D	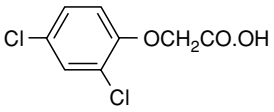	Herbicide	Nonspecific	375	Moderately hazardous
Dithio-carbamates	Maneb	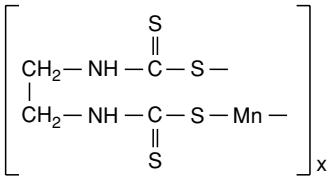	Fungicide	Nonspecific (irritant)	6750	Slightly hazardous
Coumarins	Brodifacoum	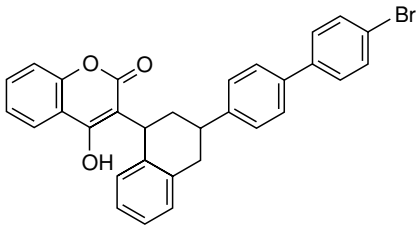	Rodenticide	Reduction of clotting power of the blood	0.3	Extremely hazardous

Pyrethrins and Pyrethroids

Pyrethrum is one of the oldest natural insecticides in use in the world today, and it has one of the best safety records of all insecticides. It is a mixture of several esters, called pyrethrins, which are extracted from flowers belonging to the genus *Chrysanthemum*. Natural pyrethrins are unstable to light and are therefore unsuitable for residual application, particularly in agriculture. This led to the development of several classes of related synthetic compounds which have a higher stability to light and a high insecticidal activity. They are known as pyrethroids. In general, pyrethroids can be toxicologically divided into two classes on the basis of signs of toxicity: those causing mainly tremor and prostration (T syndrome) and those causing choreoathetosis and salivation (CS syndrome).

For both groups, the main biological activity is mediated through the effect on sodium channels along the axon membrane of the nerves, in both insects and mammals. The effect on the sodium-exchange disturbance is fully reversible, the duration of which is different for the two classes of pyrethroids: compounds belonging to the group causing the T syndrome produce a considerably shorter effect on the sodium 'gate' openings than those belonging to the CS group.

Being highly lipophilic, pyrethroids readily pass through cell membranes and are absorbed into the body by all routes following exposure. However, the ratio of the toxic dose by the oral route to that of intravenous injection is very high; this is because of rapid detoxification, mainly by cleaving of the ester bond by esterases.

Present-day evidence indicates that doses of a pyrethroid likely to be encountered both in the workplace and through food consumption would not lead to any serious untoward effects, although reversible transient effects on the skin may be recorded and may be used as a warning response of inadvertent exposure.

Nitro- and Chlorophenols

Among dinitro compounds, the most toxic is dinitroorthocresol (DNOC), applied mostly as a winter wash for fruit trees. This and similar compounds are readily absorbed not only through the lungs and gastrointestinal tract but also through intact skin. Since DNOC is eliminated more quickly from laboratory animals than from humans, animal experiments are of limited value for the assessment of hazard. These compounds are strong metabolic stimulators, exerting a common biochemical action which affects energy at the cellular level, stimulating end metabolic processes independently of their physiological stimulator, which is the thyroid gland.

The major effects observed with excessive overexposure to these chemicals indicate a general effect on the nervous system. As the liver and kidney accumulate high concentrations of nitro- and chlorophenols, they are often also adversely affected. The major mode of action in the acute toxic effects involves uncoupling of the energy conservation process.

In animals and humans, increased basic metabolism leads to hyperpyrexia, tachycardia, hyperventilation, dehydration, and, ultimately, depletion of carbohydrate and fat stores. In the case of a single large exposure, symptoms develop rapidly, and if death occurs, it will probably take place within 24–48 h with characteristically fast rigor mortis. A high environmental temperature might aggravate the hazards from exposure. In nonfatal cases, recovery is complete, but the skin remains stained yellow for a long period of time.

Bipyridinium Compounds

The best-known representative of this class of compounds is paraquat, a widely used herbicide of moderate acute toxicity. However, paraquat deserves special attention as it produces striking lung injury in many species regardless of its route of entrance, most probably because it accumulates in lung tissue in both experimental animals and humans.

Both its herbicidal and toxicological properties are caused by the ability of the parent cation to undergo a single electron addition, forming a free radical which reacts with molecular oxygen, producing a superoxide anion. The oxygen radical damages the cell, presumably through lipid peroxidation.

So far, a large number of paraquat poisonings have been reported, the majority being accidental or suicidal and very few resulting from occupational exposure. Acute fulminant poisoning may lead to death within a few days, and more protracted cases may last for several weeks, even months, resulting in fatal, irreversible pulmonary fibrosis.

The response to treatment of paraquat poisoning, even if started soon after the accident, is not very promising, and the mortality rate remains high.

Anticoagulants

Pesticides with hemorrhagic effects are widely used as rodenticides, exerting their action by disturbing the blood clotting mechanism through inhibition of synthesis of vitamin-K-dependent factors and by decreasing prothrombin production. Some have also been used as drugs in human medicine.

As the desired harmful effect in rodents is achieved only following repeated ingestion of small amounts of these compounds, this reduces the hazard to man, and safety records are remarkable. The existence of a very effective antidote (vitamin K_1), needed in case of poisoning, makes these compounds even safer. Owing to the development of rodent resistance to many existing anticoagulants, there is a present-day trend to develop new rodenticides of a similar or entirely different mode of action.

Fumigants

Fumigants have little in common either in their chemical structure or in their mode of action. The only common characteristic is that all have a relatively high vapor pressure, and, as a rule, they are highly toxic to both pests and humans. Handling requires great skill, and strict safety rules must be followed, including the wearing of specific personal protective clothing (gas masks with specific cartridges). Many countries have therefore posed restrictions on the availability of these fumigants. To this group belong the frequently used hydrogen cyanide, phosphine, methyl bromide, chloropicrin, and some others. (*See* **Fumigants**.)

Miscellaneous Pesticides

Diversity of pesticide structure and biological activity does not allow any further meaningful grouping according to their mode of action. Therefore, several more common classes of pesticides, which do not belong to any of the above-mentioned groups, are outlined below.

Inorganic and organometal pesticides, in particular those containing arsenic, mercury, thallium, and fluorine, were extensively used in the past. Those insecticides containing some types of arsenical active material are considered today to be the only insecticides showing sufficient evidence for carcinogenicity in humans. Therefore, their use today is restricted worldwide or banned. However, numerous pesticides with active ingredients containing zinc, copper, tin or sulfur are still widely used and are considered to be irreplaceable in the control of many plant pests and diseases. As a rule, the toxicity of most of these is rather low, and the health hazard they pose is frequently outweighed by the benefit they bring. (*See* **Heavy Metal Toxicology**.)

Those *pesticides derived from plants* which are still in use are entirely unrelated chemically and toxicologically. They range in toxicity from practically harmless (such as pyrethrins) to highly hazardous (such as nicotine). Several groups of chemicals extracted from some tropical plants (Endod, Neem) are known to be biologically active against various pests, as learned by experience. However, the evidence of safety for humans and beneficial species is still lacking.

Insect growth regulators disturb the metamorphosis of insects by interfering with juvenile hormones needed for proper functioning of this process. Their action is slow and toxicity for humans negligible as the mode of action is specific to insects. Some are even recommended for use in drinking water for control of mosquito breeding sites.

Biopesticides (e.g., *Bacillus thuringiensis* H14) are very useful tools in pests control, both in agriculture and in public health, and have been in operational use for many years. However, research into the genetic manipulation of the bacteria has been intensified, particularly regarding the incorporation of the gene responsible for toxin production into other organisms. The safety aspects of these mutants should be studied carefully.

Health Hazard Arising from Pesticides

A terminological distinction should be made between *toxicity*, which is an innate capacity of a chemical to cause damage, and *hazard*, which is a qualitative term expressing the potential that a pesticide can harm health. Thus, the most toxic pesticides may be handled with little hazard if sensible and disciplined adherence to good practice become routine, and vice versa. Therefore, toxicity and hazard are not synonyms because hazard is a function of two other variables besides toxicity: the extent of contamination and duration of exposure. This can be expressed in the form of an equation:

$$\text{Hazard} = K_1 \text{ toxicity} \times K_2 \text{ contamination} \times K_3 \text{ time}$$

(K_1, K_2, and K_3 are selected measures for a given parameter, e.g., reciprocal value of LD50 (median lethal dose), amount of active ingredient and duration of work.) If any of the three variables on the right-hand side of the equation equals zero, the hazard is also zero.

The toxicity variable can never be zero (by the definition of a pesticide), but it can be acceptably low by selecting a pesticide of favorable mammalian toxicity versus insect toxicity. The toxicity of the active material also depends on the ability of the body to detoxify and/or excrete the active material. If these mechanisms fail, hazard from exposure to an insecticide is increased. Thus, disease of the liver or kidney – the two organs which are mainly responsible for metabolism and excretion of foreign compounds – place an individual in a higher risk group.

The magnitude of the hazard will also be reduced by a reduction in contamination when the pesticide is

applied by an experienced and well-trained operator who is properly equipped and properly protected. To exert a biological action, the active material must reach a sensitive tissue or organ, and it must therefore enter the body by some route. The speed of its action depends largely on the speed with which the active material reaches the blood system. Thus, exposure to volatile insecticides by the respiratory route produces an effect without much delay because the active material, which comes in through the alveolar system, is quickly absorbed into the blood circulation. Insecticide absorption is much influenced by the solvent. Regardless of the kind of compound, the speed of onset of signs of poisoning is faster if the dose is large and oral, and slower if the dose is small and dermal. There are three important routes of entry: oral (gastrointestinal), respiratory (inhalation), and dermal (through intact skin).

The longer the working time, the higher the risk, provided that the two other variables remain unchanged. It is therefore essential to limit the duration of exposure, depending upon the toxicity of the formulation used. It is preferable that work with pesticides be shared if more trained workers are available. In this way, total exposure is reduced and risk diminished.

Assessment of all of the parameters mentioned is essential for those concerned with the safe use of pesticides, as they are crucial in posing specific limitations of availability and use of a particular pesticide in agriculture, public health, or household use.

Diagnosis and Treatment of Pesticide Poisoning

In cases of acute poisoning, placing a correct diagnosis may be very difficult, even for the very experienced clinician, unless the illness is undoubtedly associated with recent exposure (occupational or accidental). This is particularly so because the signs and symptoms vary considerably both in intensity and sequence of appearance. Laboratory tests available are limited, and treatment should never be delayed pending the results of these tests.

Following excessive exposure to anticholinesterases (organophosphorus compounds or monomethyl carbamates), signs and symptoms develop very quickly, and these include nausea, headache, glandular hypersecretion (salivation, lacrimation), muscular weakness, vomiting, and cramplike abdominal pain with diarrhea. Blurred vision caused by an effect on the ciliary body is very indicative of poisoning. In severe cases of poisoning, paralysis of the diaphragm may occur, leading to respiratory depression, coma, and death.

Acute poisoning by chlorinated hydrocarbons is rare nowadays, unless they are swallowed accidentally or with suicidal intent. Signs and symptoms include apprehension and excitement, dizziness, hyperexcitability, disorientation, headache, and convulsions. In some instances, the onset of symptoms may be delayed, and remission of acute symptoms may occur in 1–3 days.

Clinical cases of poisoning with pyrethroids are extremely rare, and limited observations do not permit a detailed description of signs and symptoms in humans. However, from the studies on experimental animals and on the basis of the similarity in mode of action, one may expect similar signs and symptoms to those of organochlorine pesticides.

Acute poisoning with nitro- or chlorophenols resembles the signs and symptoms developed in a thyrotoxic crisis, with increased body temperature being the leading sign.

The peculiarity of paraquat poisoning is usually a delayed effect on the lungs with the development of fatal fibrosis. Without clear evidence of exposure, early diagnosis is virtually impossible.

As anticoagulants disturb clotting mechanisms, laboratory tests may help in clarifying obscure signs of bleeding in an otherwise healthy person.

The management of poisoning includes three essential procedures: (1) alleviation of life-threatening effects, (2) removal of nonabsorbed material, and (3) antidotal or supportive treatment. Any delay in the prompt institution of treatment can prove fatal. Rapid implementation of first-aid measures, removal of the source of contamination, and transport to a hospital, if indicated, may save life.

The sequence of procedures which should be strictly observed is as follows:

1. Check respiration, and make sure that the airway is clear.
2. Give artificial respiration if spontaneous breathing is inadequate.
3. Check the need for decontamination, and proceed if required.
4. Give antidote if available.
5. Collect evidence of exposure.
6. Transport to a medical care facility.

As a rule, emergency treatment starts in the field, continues during transport, and ends in a health center or in a hospital.

Emergency treatment in the field should be directed first towards alleviation of the life-threatening effects. All efforts should be made to maintain normal respiration, and a prerequisite for this is a clear airway.

In cases of serious poisoning with anticholinesterase insecticides, atropine is a drug of choice, to be

given a soon as possible. To save life, it must be given by injection. Oral application may mask the clinical picture and is not recommended.

As soon as the diagnosis of organophosphorus compound poisoning is made, a first injection of 2–4 mg of atropine sulfate should be given intramuscularly, or intravenously in very severe cases. The effects of intravenous atropine application begin within 3–4 min and are maximal about 8 min after injection. Overatropinization is not as dangerous as underatropinization. Persons with manifest peripheral symptoms should also be given 1–2 g of a soluble salt of pralidoxime. Patients poisoned with carbamates must not be given an oxime.

For intoxication with organochlorine compounds, there is no specific antidote. Treatment is aimed at controlling the symptoms, especially hyperactivity and, in some instances, convulsions. Artificial ventilation may be required. Anticonvulsant treatment with barbiturates, diazepam, or paraldehyde should be given in sufficient dosage to calm the patient and prevent convulsions.

The principal management of intoxication by conventional or second-generation anticoagulant rodenticides is administration of phytomenadione (vitamin K_1). In addition to vitamin K_1, a seriously ill patient should initially be given a transfusion of carefully matched whole blood (as little as 50 ml may be effective); transfusions may be repeated daily until the patient's prothrombin time has returned to normal. Prolonged observation of patients affected by second-generation anticoagulants (coumarin derivatives) is required because these compounds are metabolized slowly, and repeated therapy may be indicated.

For intoxication with paraquat, no antidotes currently exist, and management essentially relies upon (1) the use of adsorbents to prevent absorption from the gut and (2) the removal of absorbed paraquat from the body, although these are rarely effective in severe poisoning. One liter of a suspension of Fuller's earth (about $300\,g\,l^{-1}$) or bentonite (about $70\,g\,l^{-1}$) should be administered orally as soon as possible. Activated charcoal may be more effective in adsorbing paraquat and should be used if it is available. In an emergency, use of ordinary soil may be beneficial if these adsorbents are not available. Administration of oxygen is contraindicated in acute poisoning because paraquat is more toxic in the oxygenated lung, and the use of oxygen should be delayed for as long as possible.

If a number of patients are found to be exhibiting symptoms suggestive of poisoning by a pesticide (or other chemical) without a history of exposure, the possibility of the cause being gross contamination of a food item or of drinking water, and being unrelated to any chemical, should be considered.

See also: **Fumigants**; **Fungicides**; **Heavy Metal Toxicology**; **Insect Pests**: Insects and Related Pests; Problems Caused by Insects and Mites; **Legislation**: Contaminants and Adulterants; **Organically Farmed Food**; **World Health Organization**

Further Reading

Aldridge WN (1990) An assessment of the toxicological properties of pyrethroids and their neurotoxicity. *Critical Reviews in Toxicology* 21: 89–104.

FAO (1990) *International Code of Conduct on the Distribution and Use of Pesticides (amended version)*. Rome: Food and Agriculture Organization.

Hayes JW, Jr. and Laws ER, Jr. (eds) (1990) *Handbook of Pesticide Toxicology*, vols. 1, 2 and 3. New York: Academic Press.

WHO (1981) *Environmental Health Criteria 39: Paraquat and Diquat*. Geneva: World Health Organization.

WHO (1986a) *Environmental Health Criteria 63: Organophosphorus Insecticides*. Geneva: World Health Organization.

WHO (1986b) *Environmental Health Criteria 64: Carbamate Pesticides*. Geneva: World Health Organization.

WHO (1990) *Environmental Health Criteria 104: Principles for the Toxicological Assessment of Pesticide Residues in Food*. Geneva: World Health Organization.

WHO (1992a) *IPCS Summary of Toxicological Evaluations Performed by the Joint FAO/WHO Meeting on Pesticide Residues (JMPR) through 1991 (WHO/PCS/92.9)*. Geneva: World Health Organization.

WHO (1992b) *The WHO Recommended Classification of Pesticides by Hazard and Guidelines to Classification 1992–1993 (WHO/PCS/92.14)*. Geneva: World Health Organization.

Worthing CR (ed.) (1990) *The Pesticide Manual, A World Compendium*, 9th edn. Farnham, UK: The British Crop Protection Council.

Types, Uses, and Determination of Herbicides

A M Au, Department of Health Services, State of California – Health and Welfare Agency, Berkeley, CA, USA

Introduction

The major function of herbicides is to prevent or control weed growth in the field. Some herbicides are nonselective, that is, they kill a wide variety of

plants, both desirable and undesirable. Other herbicides are selective – they kill certain kinds of plants while permitting others to survive. Selectivity of herbicides enables a desired crop to grow and produce free of competition from weeds. This article focuses on the classification of herbicides based on mode of action, the methods of application, safety implications, specific examples of uses, stability in the environment, and the analysis of residues in foods.

Classification Based on Mode of Action

Herbicides can be classified in different ways: by chemical name, by chemical characteristics of the compound, by toxicity, or by mode of action. There are two major categories of herbicides classified by mode of action: contact herbicides and translocated herbicides.

Contact herbicides affect only the part of the plant that they touch. Absorption through foliage is minimal. The application, therefore, must be made in sufficient quantity to cover the foliage thoroughly. Examples of contact herbicides are diclofop, dinoseb, diquat, and paraquat. Certain contact herbicides, like diquat and paraquat, are deactivated by soil particles. They must be mixed with clear water and applied directly to the vegetation.

At the molecular level, not all contact herbicides act in the same manner. For example, diquat and paraquat generate phytotoxic free radicals that interfere with the lipid metabolism of the plant and lead to ultimate death, whereas diphenyl ether herbicides cause chlorosis and necrosis resulting from the inhibition of the photosynthetic process. The biochemical mechanism of action of organic arsenicals such as cacodylic acid is not known.

Unlike contact herbicides, systemic herbicides can be translocated to other parts of the plant. They alter the normal biological function of the plant by interfering with certain biochemical reactions. Thus, when applied to foliage or soil, they enter the plant and translocate to their site of action. Examples of translocated herbicides are atrazine, glyphosate 2,4-dichlorophenoxyacetic acid (2,4-D) and simazine. Systemic herbicides, like contact herbicides, also have diverse modes of action at the molecular level. Chlorinated aliphatic acid herbicides, such as trichloroacetic acid (TCA), modify protein structure, causing chlorosis and necrosis in plants. Amide herbicides (e.g., alachlor, metachlor) interfere with both protein and nucleic acid synthesis. Carbamates inhibit protein synthesis only. The phytotoxic activities of thiocarbamates and dithiocarbamates appear to stem from the ability of these compounds to inhibit lipid synthesis. Phenoxy herbicides (e.g., 2,4-D), on the other hand, apparently stimulate protein and RNA synthesis. These stimulations accelerate plant growth and, in turn, contribute to the death of the plants. Unlike the other translocated herbicides listed above, triazine herbicides block photosynthesis as the primary mode of action.

Methods of Application

Uniform and precise application of herbicides to the field is essential. Common herbicide carriers are liquid formulations containing emulsifiers. Since most herbicides are organic compounds that are not readily soluble in water, they are frequently distributed as dry granules.

Liquid herbicides are usually dispensed by sprayers. Large ground sprayers mounted on mobile units such as tractors can apply herbicide over areas as large as a hectare. Recirculating sprayers distribute the herbicide solution horizontally above the crop. Solution not lost in tall weeds is caught in a basin and pumped back to the spray tank for reuse. Most herbicides can be applied to foliage or to the soil by sprayers and granule applicators as broadcast treatments or bands.

There are three basic types of application methods for the distribution of either granular or liquid herbicides. These are ground-operated application, aerial application, and manual application. Most of the application devices can be adapted for use via ground-operated or aerial-operated methods, depending on the vehicle on which the device is mounted. Handheld sprayers, such as backpack sprayers or small plot sprayers, are examples of manual applicators.

Granular herbicide formulations can be applied by drop-type applicator, centrifuge spreaders or pneumatic systems. The drop-type applicator consists of rollers rotating over an orifice in the hopper. The flow of the herbicide granules is controlled via a metering mechanism. The maintenance of this type of applicator is cumbersome. In the centrifuge spreader system, herbicide granules are distributed into the field by two counterrotating disks. In the pneumatic system, the herbicide granules are kept in a hopper, and the flow of granules is controlled by a ground-driven fluted roller coupled to a low-volume, high-pressure airstream. The granules eventually reach the spreader nozzles via two flexible delivery chutes which provide the control necessary for the adjustment and positioning of the treatment.

Recently, because it appears to be cost-effective and environmentally beneficial, the wiper application method (first conducted in 1909 by Mahanay) was reintroduced as an application method for wipe-on herbicides. There are a wide variety of wiper applications, including the Rope-wick, carpet roller,

continuous belt, rotary wiper, and Rogueing glove application. The Rope-wick device is a rig consisting of a series of short, exposed nylon ropes, each end of which is connected to a reservoir of herbicide solution. The solution passes into the ropes by both capillary action and gravitational flow. As the applicator moves through the weed-infested field, the chemical on the soaked wicks is rubbed on to the tops of the tall weeds but not on to lower-growing crop. Spray drift is eliminated. The carpet roller, as the name suggests, is a tractor mounted with a nylon carpeted roller soaked with liquid herbicide through a delivery system from a herbicide reservoir. The design is such that herbicide will only be in contact with the tall weeds without contacting the desired crop. A continuous belt consists of a V-belt, with sponges glued on to it, passing through a herbicide-containing reservoir. An adjustable pressure wheel removes excess herbicide from the sponges prior to its application on to the weeds. This helps prevent herbicide from contacting the crop. A rotary wiper applicator consists of flexible arms that allows wiper rotation around stationary objects to avoid injury to tree trunks. A glove equipped with absorbent pads that has continuous loading of herbicide solution from a reservoir is known as the Rogueing glove application.

A relatively new way of applying herbicide is through a sprinkler system. Liquid herbicide is pumped into the system and sprayed in a manner similar to a water sprinkler system.

Aerial spraying, using small aircraft or helicopters, can cover a large field in a shorter time period. Sprayers similar to those used in the ground operation can be mounted on to the aircraft for application to large areas.

Considerable research has been conducted by chemical companies to devise controlled-release formulations, a technology which has been demonstrated to be safe and effective in the drug industry. The three controlled-release systems are: (1) the chemically bound system, in which the herbicide is bound to a polymer; (2) the microencapsulation system, where the herbicide is coated by a polymer; (3) the matrix encapsulation system, where the herbicide is dispersed within a polymer matrix. This technology renders safer herbicide handling, reduces environmental pollution, and enhances herbicide selectivity.

The treatment of seeds with herbicides is another area of research currently under investigation. Certain herbicides can be applied to certain crop seeds prior to planting. Once planted in the soil, the herbicide moves swiftly away from the seed to the surrounding area to inhibit weed growth. This approach is economical and convenient to use. It also has a lesser tendency to pollute the environment.

Safety Implications

Ideally, herbicides are chemicals designed to cause injury only to undesirable weeds. Due to the biological differences between plants and animals, herbicides have low acute toxicity towards humans. However, most chemicals tend to have more than one metabolic effect on different living organisms. It is such unexpected secondary or side-effects of chemicals that cause major concern. Before the approval for marketing of any herbicide by the government, chemical companies must conduct extensive studies on the effectiveness of new herbicides. In addition, studies on metabolic fate and toxicity to animal species and plants are conducted to insure that the herbicide is safe to be used in the field. The cost of these testing procedures is tremendous. The specific requirements for herbicide registration depend upon the law of the country in which they are used but, in general, herbicides may not be used without registration by some governmental agency that carries out extensive reviews of the studies and tests performed by the chemical companies. It is also the duty of the vendor to provide users with information about all precautions and safe handling procedures of the herbicide. Adherence to those instructions by users is of utmost importance for the avoidance of harmful exposure.

The health effects of herbicides and their environmental impacts have been a major concern of the public. The health and safety issues associated with herbicide use affect not only consumers, but also farmers, formulators, applicators, and field workers, as well as the users of home and garden products. Groundwater contamination by agricultural chemicals has been an issue worldwide and will continue to be important in the years to come. Restrictions on the use and banning of certain herbicides are on the rise. Development of biodegradable herbicides with low residue levels combined with improved delivery systems can help protect the environment.

Depending on the conditions, exposure to herbicides at a high concentration can be fatal. Precautions in handling herbicides should be strictly followed to avoid unnecessary exposure. Safety training of all personnel designated to handle herbicides is imperative, and refresher safety training should also be conducted periodically to keep personnel up-to-date on all safety issues.

Equipment for application should be inspected frequently to insure that it is functioning properly. Malfunctioning equipment should be repaired prior to usage. Emergency procedures covering spillage or accidental poisoning should be established and strictly followed. There should be health surveillance programs for workers to monitor and insure worker

well-being. Personal protective equipment and clothing should be checked for leaks periodically and should be kept clean. Good record-keeping of herbicide inventory and application is necessary to account for usage.

In order to avoid inhalation of herbicides during application, workers should wear respirators. There are many different kinds of respirators, namely the chemical cartridge respirators, powered air-purifying respirators, canister respirators, supplied-air respirators, and self-contained breathing apparatus. There are also certain herbicides that can be absorbed by the skin; users should handle these with due caution and should wear appropriate protective clothing, including face shields.

Care should be given in the storage and transport of herbicides to minimize spillage. In case spillage occurs, proper decontamination should be performed immediately. Maintenance of all equipment for dispensing herbicides should be done routinely. Warnings should be given in advance to alert others of possible herbicide drift during application, and warning signs and restricted-entry signs should be posted to prevent others from entering the treated areas.

Specific Examples of Uses

Weeds are usually defined as undesirable plants. Weeds are often the primary concern of farmers, because they cover many millions of productive acres that could be used to grow beneficial crops. In the past, farmers controlled weeds by manually removing them from the crops. The ancient Romans killed weeds with salt. With farms of small size, manual weed control, such as hand hoeing and pulling, mowing, burning, and machine tillage, is feasible. However, with large farms such labor is extensive and costly. As the size of the farms increased and synthetic herbicides were introduced, farmers began to rely on herbicides to control weeds.

Generally, there are three different types of treatment for the application of herbicides. They are preplanting, preemergence, and postemergence treatments. Preplanting treatment takes place prior to planting; preemergence treatment is done after planting but preceding the crop or weed emergence; and postemergence treatment is performed after the emergence of the crop or weeds.

2-4-Dichlorophenoxyacetic acid (2,4-D) was one of the first synthetic herbicides introduced to control broad-leafed weeds in cereal crops and pastures. 2,4-D is an effective systemic herbicide and is selective for broad-leafed plants but not grasses. It can be used as either a pre- or a postemergence herbicide for corn, but only as postemergence for sorghum. 2,4-D is highly versatile and is used on a variety of crops such as wheat, barley, oats, rice, and sorghum. During the Vietnam war, Agent Orange, a 50: 50 mixture of 2,4-D and 2-4-5-Trichlorophenoxyacetic acid (2,4,5-T), was used extensively over the terrain in Vietnam as a defoliant to clear the way for US troops. 2,4,5-T is usually contaminated with dioxin, a highly toxic chemical compound and known carcinogen. It is due to this notorious contaminant that Agent Orange has been blamed for various illnesses and reproductive problems among those who came in contact with the defoliant in Vietnam.

Paraquat is used as a preemergence treatment for sugar beet. Simazine, on the other hand, is used both as a preplant and as a preemergence treatment for corn. *S*-ethyl dipropylthiocarbamate (EPTC) is incorporated into the soil as a preplant treatment for potatoes.

Herbicides have made it possible to grow more food on less land with less labor and at lower cost. Herbicides are also used to control aquatic weeds which impede water flow in irrigation canals and drainage systems, interfere with fishing, or promote insect-breeding grounds.

Stability in the Environment

For herbicides to exert their effects on weeds, they must be relatively stable in the treated environment. However, the stability of the chemical creates a burden on the environment, especially for those herbicides that find their way into aquifers and contaminate drinking water sources or remain on the crops at the time they reach consumers. In order to insure that the newer generation of herbicides do not linger on after accomplishing their task, research is directed towards synthesizing biodegradable compounds. The ideal herbicide is one that degrades to harmless chemicals after it performs its function and therefore does not persist in nature. Carbamates are one such class of chemicals specifically designed with that goal in mind.

Analysis of Residues in Foods

There is increasing awareness among consumers of the hazard of chemical contamination of food and drinking water. There is particular concern over the implications of food contamination by herbicide residues. Analysis for herbicide residues in food requires methods that identify not only parent structures, but also their metabolites and degradation products in a variety of food matrices. Certain food crops are perishable and therefore cannot wait for lengthy analysis to establish the suitability for consumption. Thus, rapid analytical technology is needed. Multiresidue

methods, which can detect the presence of many herbicides at once, are the methods of choice for determining the presence of a multiple number of herbicides and their degradation products in a food sample.

An analytical process consists of several major steps: the sample preparation, the extraction, the clean-up, the determinations, and the confirmation. These steps are common to the determination of other agrochemical residues, including pesticides, and are discussed in detail in the following article. The basic operation of sample preparation is to separate physically food or plant parts and to chop and blend them. The essence of the extraction process is to remove the target herbicide from the other components in the sample matrix. The main function of the clean-up procedure is to remove interfering constituents, usually by selective partitioning into organic solvents followed by an adsorption or size exclusion chromatographic purification step. The determination step includes separation of the purified samples through thin-layer chromatography, gas chromatography, or liquid chromatography techniques followed by the detection procedure using a variety of specific detectors for the targeted compound. For confirmation purposes, the analyte is further subjected to mass spectrometric analysis. Recently, successful attempts have been made in using gas chromatography–mass spectroscopy (GCMS) as a primary screening method. The GCMS screening technique provides simultaneous results for both the detection and the confirmation of the targeted compound in the sample matrix. This one-step procedure will be the method of choice as it offers both rapid and definitive data. (*See* **Chromatography**: Thin-layer Chromatography; Gas Chromatography; **Mass Spectrometry**: Principles and Instrumentation.)

Improvements to existing analytical technology are well underway to reduce the time- and solvent-consuming extraction and clean-up steps. Supercritical fluid extraction, which is based on the solvent properties of gases such as carbon dioxide at its critical pressure and temperature, can selectively remove the targeted compound from the complex food matrix in a short time. Through such approaches the recovery of the compound can be easily accomplished.

The use of antibodies as analytical tools is a common practice in clinical laboratories. Antibodies have been recently developed for identifying and quantifying herbicides. Antibodies can be isolated from the plasma of an immune-challenged animal or from a hybridoma cell line. An antibody that is specifically generated from a compound will have high selectivity towards that compound even in the midst of other interfering components and can bind to it tightly to form a complex. Therefore, by attaching a tracer to the antibody molecule, one can quantify the amount of antibody complex present, which is also an indication of the amount of the compound in the sample. A variety of tracers are available, for example radioisotopes, fluorescent molecules, and so forth. One of the disadvantages of the immunoassay is that the length of time to generate the specific antibody is relatively long. Typically, it takes approximately a year to develop. However, once it is generated, the immunoassay can be performed in less than half an hour. The triazine immunoassay is now available commercially, and, like most immunoassays, it is specific, sensitive, rapid, and cost-effective. (*See* **Immunoassays**: Principles.)

One of the modes of action of herbicides is by inhibiting photosynthesis. The Hill reaction is one of the processes in the photosynthetic pathway; therefore, a screening technique based on the inhibition of the Hill reaction can be a useful tool in detecting herbicides like triazines and carbamates.

See also: **Chromatography**: Thin-layer Chromatography; Gas Chromatography; **Immunoassays**: Principles; **Mass Spectrometry**: Principles and Instrumentation; **Pesticides and Herbicides**: Residue Determination

Further Reading

Ashton FM and Crafts AS (1981) *Mode of Action of Herbicides*, 2nd edn. New York: Wiley.

FDA (1968) *Pesticide Analytical Manual*, 2nd edn, vol. 1. *Methods Which Detect Multiple Residues*. Washington, DC: US Department of Health Services, Food and Drug Administration.

Klingman GC, Ashton FM and Noordhoff LJ (1982) *Weed Science: Principles and Practices*, 2nd edn. New York: Wiley.

McWhorter CG and Gebhardt MR (eds) (1987) *Methods of Applying Herbicides. Monograph Series of The Weed Science Society of America*, no. 4. Champaign, IL: The Weed Science Society of America.

Newton M and Knight FB (1981) *Handbook of Weed and Insect Control Chemicals for Forest Resource Managers*. Beaverton: Timber Press.

Residue Determination

S Nawaz, Central Science Laboratory, Sand Hutton, York, UK

Background

Continued population growth has led to an increased demand on the world's natural resources. Pesticides are widely used to help increase the yield and improve

the quality of crops. Pesticides are categorized according to their mode of action and include insecticides, herbicides, fungicides, acaricides, nematicides, and rodenticides. Pesticides are also used as plant growth regulators and for public-health purposes. Global sales of pesticides during 1996 were estimated at over US$30 billion. Herbicides, insecticides, and fungicides accounted for 48, 28, and 18%, respectively, of the total sales (see Table 1). There are over 900 chemicals registered for plant protection purposes in the European Union (EU) alone. In addition to the existing pesticides, there are an ever increasing number of new chemicals being granted approval.

As pesticide use can leave undesirable residues, various national and international authorities regulate the use of pesticides and set maximum residue levels (MRLs) in crops. An MRL is the maximum concentration of a pesticide and/or its toxic metabolites legally permitted in food commodities and animal feeds. If pesticides are properly applied at the recommended rates, and crops are only harvested after the appropriate time intervals have elapsed, residue levels are not expected to exceed MRLs. The residue levels in foods derived from commodities that comply with the respective MRLs are also intended to be toxicologically acceptable. In the EU, the regulation of the agrochemical industry, and the setting of MRLs is currently being harmonized across all member states by the EC. In the USA, the Environmental Protection Agency is responsible for such regulation. The Codex Alimentarius Commission (FAO/WHO) has published tables of MRLs which have official status across the world and are used to aid international trade. The monitoring of the residues in foods is often at the microgram per kilogram level or lower and has to be supported by strict analytical quality-control standards, so that the analysis produces unequivocal, precise, and accurate residue data. Before samples are analyzed, the analyst has to demonstrate that the intended analytical method can achieve adequate specificity, sensitivity, linearity, accuracy, and precision at the relevant analyte concentration and in appropriate matrices. The calibration solutions must be prepared using certified reference standards. Residue analyses normally include the metabolites, isomers, and other related compounds included in the MRL definition. Many methods can determine a large number of residues in a single analytical run; these multiresidue methods are in common use and help reduce the total cost of analysis.

Sampling

A representative sample consists of a large number of randomly collected units. It is not always possible to collect large samples because of the cost of transportation and the practicalities of sample handling in the laboratory. Monitoring of pesticide residues for MRL compliance involves analysis of a composite sample, made up of a number of individual units. Recent research has shown that pesticide residues in individual units of fruit and vegetables can exhibit an extremely skewed distribution, and this is likely to add to the difficulty of taking a representative sample. The guidelines for obtaining composite samples for MRL compliance monitoring are published by the Codex Alimentarius Commission and are summarized in Table 2. In addition to checking for MRL compliance, residue analyses are also carried out to investigate other issues such as cases of misuse or the deliberate poisoning of wildlife or domestic animals. In such instances, a more targeted sampling regime may be adopted, and a qualitative analysis may suffice.

Sample Preparation/Subsampling

Samples should be analyzed without any delay, as some pesticide residues may degrade rapidly. If

Table 1 Major classes of chemicals used as pesticides

Type of chemical	*Examples and their primary uses*
Benzimidazole	Benomyl (F), carbendazim (F), thiabendazole (F)
Bipyridylium	Diquat (H), paraquat (H),
Carbamate	Aldicarb (I, N), carbaryl (I), carbofuran (A, I, N)
Dithiocarbamates	Mancozeb (F), ziram (H), thiram (H)
Organochlorine	DDT (I), lindane (I), endosulfan (I)
Organophosphorus	Chlorpyrifos, malathion (I, A), parathion (I)
Pyrethroids	Permethrin (I), cyfluthrin (I)
Substituted phenyl ureas	Diuron, linuron, monolinuron (H)
Triazine	Atrazine (H), simazine (H)

A, acaricide; F, fungicide; H, herbicide; I, insecticide; N, nematicide.

Table 2 Codex guidelines for collection of representative samples

Sample type	*Minimum weight of sample (kg)*
Small or light products (e.g., berries, peas, spinach)	1
Medium sized products (e.g., apples, carrots, potatoes)	1 (at least 10 units)
Large size products (e.g., melons)	2 (at least 10 units)
Dairy products (e.g., cheese, butter)	0.5
Meat, poultry, fish	1
Oils and fats	0.5
Cereals and cereal products	1

immediate analysis is not possible, storage of samples at −20 °C may help minimize the degradation process. Typically, a 20–50-g portion (subsample) is required for analysis. In order to obtain representative subsamples, it may be necessary to grind/mill and thoroughly mix the whole sample, so that any residues present are evenly distributed. This process is especially important because residue levels can exhibit a high degree of variability between individual units. Some pesticides are known to degrade during the processing of fruit and vegetable samples at ambient temperature. Milling frozen food samples in the presence of excess solid CO_2 (dry ice) has been shown to minimize the losses of most pesticide residues during the process.

Domestic mills (e.g., coffee grinder) as well as more specialized mills are used to grind samples of cereals, nuts, and pulses. Manual methods, such as cone and quarter or mechanized devices such as the riffle divider can be used to obtain representative subsamples from such samples. Samples of animal tissues are minced and mixed thoroughly before subsampling. For pesticides (e.g., organochlorines) that accumulate in the fatty tissue of animals, visible layers of fat may be removed for direct analysis. Preparation of homogenous fruit and vegetable samples prior to subsampling may be carried out using domestic food processors and blenders. However, larger specialized mechanical bowl choppers are more suitable for large samples, and heavy-duty choppers may be required to process frozen samples.

Extraction

The extraction step involves the quantitative transfer of pesticide residues from the food matrix into solvent(s). The efficiency of extraction process depends on the physicochemical properties of the solvent(s) and analytes. Important factors include pH, polarity, temperature, sample/solvent ratio, presence of water, and degree of analyte/matrix binding. Most extraction procedures employ organic solvents in the presence of water. The presence of water is critical for extraction of pesticides from cereals and cereal products, as it helps reduce the binding between residues and matrix. Samples of cereals, nuts, pulses, fruit, and vegetable samples are extracted by simple homogenization with an organic solvent. The most commonly used solvents include ethyl acetate, acetone, acetonitrile, hexane, methanol, and dichloromethane. For multiresidue extraction methods, it is not possible to establish the optimum extraction conditions for all residues with differing physical and chemical properties so the choice of solvent polarity is usually a compromise. Although the presence of water may aid the extraction process, most polar pesticides are partitioned into the water phase if the organic solvent used is not water-miscible. Addition of anhydrous sodium sulfate prior to extraction can overcome this problem. The analysis of nonpolar pesticides in fatty products involves extraction of fat using nonpolar solvents such as hexane, *n*-pentane, or light petroleum. After the evaporation of the solvent, the fat is redissolved in an organic solvent prior to the clean-up and determination steps.

Other methods used for residue extraction include soxhlet, where samples are exposed to solvent vapor that is condensed and vaporized repeatedly to exhaustively extract analyte(s). Supercritical fluid extraction (SFE) methods use a gas under high pressure and above the *critical* temperature to extract the residues. This technique is more widely used for samples with low moisture content (cereals) where sample/analyte binding is more common.

For residues that are not suited to multiresidue extraction methods, dedicated extraction methods may be used for single or small groups of closely related pesticides. These methods utilize physical and chemical properties of analyte and solvent to carry out selective extraction of the analyte from matrix. A number of pesticides are normally analyzed using single residue methods and include formetanate, fluazifop, 2,4-D, formetanate, propamocarb, and maleic hydrazide.

Clean-up

Sample extracts not only contain the target analyte(s) but may also contain coextractives, such as plant pigments, proteins and lipids. These coextractives may have to be removed prior to instrumental analysis to avoid possible contamination of instruments and to eliminate compounds that interfere during the determination step. To achieve low detection limits, sample extracts may also require a concentration step, which can be incorporated in the clean-up procedure. Clean-up procedures can lead to losses of residues and increases in the cost of analysis, and can reduce the sample throughput. Therefore, a number of methods utilize minimal clean-up and instead rely on the selectivity of the detector(s).

A number of analytical procedures employ liquid–liquid extraction for clean-up of sample extracts by selective partitioning of analytes between two immiscible solvents. This technique is commonly used during the analysis of fatty samples. Liquid–liquid extraction is not easy to automate and requires the use of large amounts of solvents. (*See* **Analysis of Food**.)

Adsorption chromatography is used in many residue laboratories for the clean-up of sample extracts. This process involves either:

- retention of analyte(s) on a chromatographic column while the coextractives are unretained: (the analyte(s) are then selectively eluted from adsorption medium); or
- retention of coextractives on a chromatographic column while the analyte(s) are unretained.

A number of materials are available for adsorption chromatography, including alumina, florisil, cellulose, diatomaceous earth (celite), carbon (charcoal, graphite), silver nitrate, and silica. Alumina and silica are effective for the clean-up of fatty samples for organochlorine (OC) pesticide residue analysis. Silver nitrate is used for the removal of sulfur-containing intreferences. Carbon has a high affinity for plant pigments and is particularly suitable for the clean-up of green leafy vegetable extracts with a high chlorophyl content.

Chemically modified silica sorbents are widely used for the clean-up of sample extracts. These materials are prepared by the reaction of silanol groups on silica surfaces with silane reagents to form esters containing required functional groups. These sorbents are used in cartridges, disks, membrane filters, and impregnated fiber tips. Some chemically modified silica and other sorbents used for clean-up of sample extract are given in Table 3.

Gel-permeation chromatography (GPC) or size-exclusion chromatography separates molecules on the basis of their molecular size. Large molecules (e.g., lipids, pigments, and polymeric coextractives) elute faster than smaller molecules such as pesticide residues. The reproducibility, suitability to automation, and compatibility with a wide range of pesticide/matrix combinations make GPC a popular clean-up method in many laboratories worldwide. The disadvantages of the technique include the use of large quantities of solvents, limited sample throughput, and incomplete separation of high-molecular-mass pesticides from coextractives. (*See* **Chromatography**: Principles.)

Table 3 Examples of sorbents used for clean-up of sample extracts

Nonpolar	*Polar*	*Ion exchange*
Octadecyl (C18)	Florisil	SCX benzenesulfonylpropyl
Octyl (C8)	Diol (2OH)	PRS sulfonylpropyl
Ethyl (C2)	NH_2 aminopropyl	Water's Oasis™ divinylbenzene: vinylpyrrolidione copolymer
Polypropylene	Silica	SAX trimethylaminopropyl

Chromatography and Determination

The final stage of the pesticide residue analysis procedures involves the chromatographic separation and instrumental determination. Where chromatographic properties of some pesticides are affected by sample matrix, calibration solutions should be prepared in sample matrix. The choice of instrument depends on the physiochemical properties of the pesticide(s) and the sensitivity required. As the majority of pesticides are relatively volatile, gas chromatography (GC) has proved to be an excellent technique for pesticide determination and is by far the most widely used. A typical chromatogram from a multiresidue method is shown in Figure 1. (*See* **Chromatography**: Gas Chromatography.)

Most residue methods employ splitless injection of 1–3 μl of the sample extract. Cold on-column injection is employed when the pesticides are likely to breakdown in a hot injection port. A wide range of GC column types are used for residue analysis, and the choice depends on the physicochemical properties of the analytes. Fused silica capillary columns are most widely used during analysis of pesticides (see Table 4). Typical capillary columns are 25–30 m long with an internal diameter range of 0.1–0.5 mm and a stationary phase of 0.1–10 μm thickness. Nonpolar stationary phases are used for the separation of nonpolar pesticide residues such as OC and pyrethroids. Similarly, more polar pesticide residues (e.g., methadimophos) are separated on relatively polar columns. The conventional semiselective detectors are widely used for residue analysis. Electron-capture detectors (ECDs) are utilized for halogenated pesticides (OCs and pyrethroids). Nitrogen phosphorus detectors are used for organophosphorus (OP) and nitrogen-containing pesticides. Flame-photometric detectors (FPDs) are used for OP and sulfur-containing pesticides, while atomic emission detectors can be used for a wide range of pesticides.

Gas chromatography–mass spectrometry (GC–MS) has been the predominant technique for the confirmation of pesticide residues in the past. Relatively inexpensive bench-top instruments have made the technique more widely available for routine screening in recent years. The resolving power of GC coupled with the specificity of mass spectroscopy provides the most effective means of pesticide residue analysis. A number of ionization techniques are available for GC–MS instruments and include electron impact (EI) and chemical ionization (CI). EI impact ionization

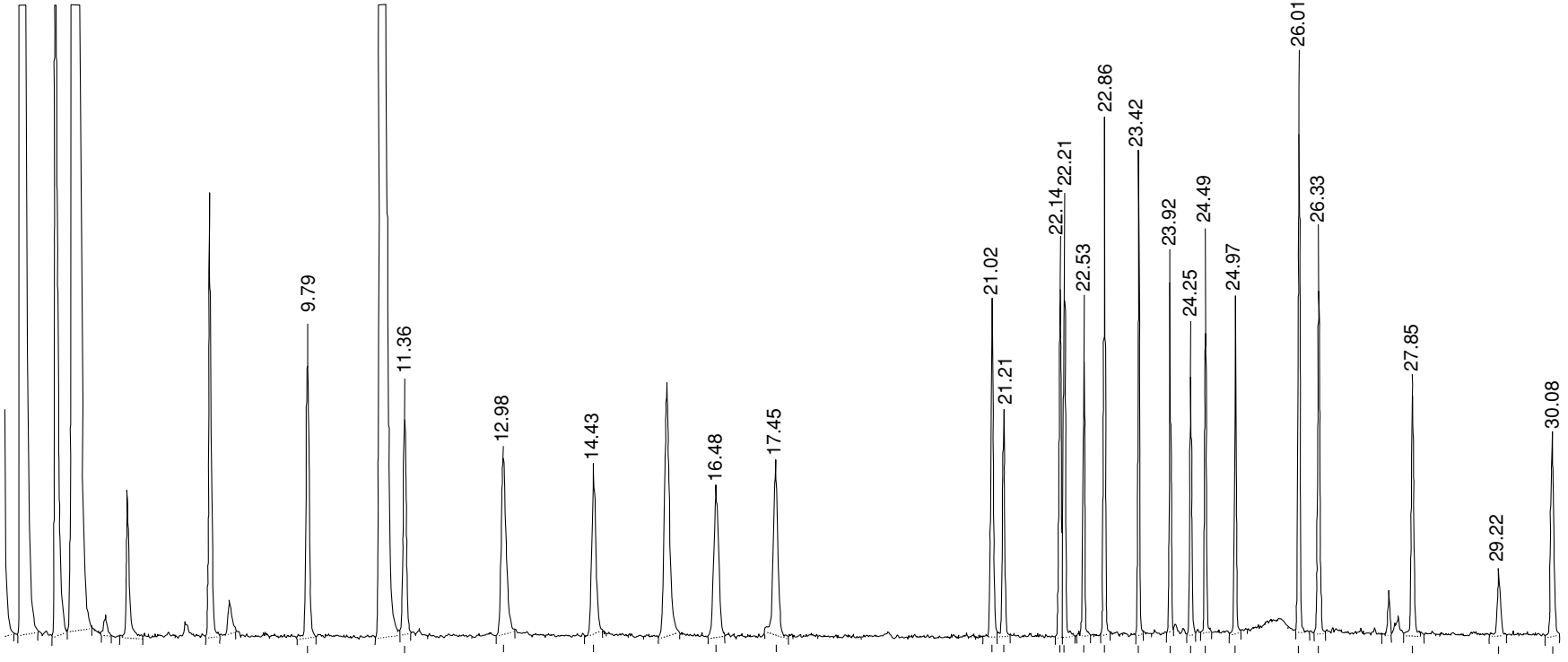

Figure 1 Chromatogram (GC–FPD) obtained from an injection of cauliflower matrix matched calibration solution containing 0.3 ng of each of following pesticides (and their retention times in minutes): heptenophos (9.79), ethoprophos (11.36), monocrotophos (12.98), dimethoate (14.43), fonofos (16.48), diazinon (17.45), parathion-methyl (21.02), malaoxon (21.21), fenitrothion (22.14), pirimiphos-methyl (22.21), malathion (22.53), parathion (22.86), pirimiphos-ethyl (23.42), mecarbam (23.92), methidathion (24.25), tetrachlorvinphos (24.49), profenofos (24.97), ethion (26.01), triazophos (26.33), pyridaphenthion (27.85), azinphos-methyl (29.22), and pyrazophos (30.08).

Table 4 GC stationary phases commonly used for pesticide analysis (listed in increasing order of polarity)

Stationary phase	*Typical uses*
100% methyl silicone (DB1)	Nonpolar pesticides
5% phenyl, 95% methyl silicone (DB-5)	Multiresidue screening
35% phenyl 65% methyl silicone (DB-35)	EPA method 608
50% phenyl 50% methyl silicone (DB-17)	Polar organophosphorus pesticides
14% cyanopropylphenyl 86% methyl silicone (DB1701)	Organochlorine pesticides
50% cyanopropylphenyl 50% methyl silicone (DB-225)	Polar pesticides
Poly(ethylene) glycol (DB-wax)	Polar pesticides

leads to a greater degree of fragmentation of molecules compared with CI. Hence, CI provides a greater sensitivity, but EI provides more spectral information. The detection systems most widely utilized in residues analysis are based on either the quadrupole or ion-trap principle. The quadrupole instruments have limited sensitivity in the scan mode compared with the ion-trap instruments. However, by operating in the selected ion mode, adequate sensitivity can be achieved on quadrupole detectors. A small quadrupole mass selective detector can typically detect over 100 pesticides in food extracts at the relatively high levels (typically 0.2 mg kg^{-1}) using full scan spectra, reducing to lower levels (typically 0.01 mg kg^{-1}) in the selected ion mode. The ion-trap detectors have a higher inherent sensitivity, and this allows screening of clean sample extracts for a wide range of pesticides at low levels (typically 0.05 mg kg^{-1}) in full scan mode. The major advantage of the ion-trap instruments is that the characteristic ions can be selected and then further fragmented to provide added specificity (MS–MS). (*See* **Chromatography**: Combined Chromatography and Mass Spectrometry.)

High-performance liquid chromatography (HPLC) is increasingly being used for the determination of pesticide residues, as it is especially suited to the analysis of nonvolatile, polar, and thermally labile residues that are difficult to analyze using GC. The resolution achieved on HPLC can be comparatively low, and therefore, the use of selective detection systems may be necessary for reliable residue analysis. Ultraviolet (UV) spectroscopy is the most common choice for detection of (OP) residues in environmental samples (e.g., soil, water). Although UV detection is not a very selective technique, it is commonly used for screening purposes due to its low cost, simplicity, and wide application range. Elimination of interferences and optimized chromatography are essential prior to detection in order to enhance the selectivity of UV-based methods. The use of diode array detectors can further enhance the selectivity of UV-detection procedures. Fluorescence detection offers a greater selectivity and sensitivity than UV. Pesticides with inherent fluorescence include dimethoate, ethoxyquin, azinphos methyl, phosalone, thiabendazole, and carbendazim. With the exception of thiabendazole and carbendazim, this technique is not widely used in pesticide residue analysis, as methods based on inherent fluorescence have a poor sensitivity compared with other methods available. Precolumn and postcolumn reaction systems can be employed with HPLC methods, which can help improve the chromatographic separation and detection of analytes. A number of pesticides (e.g., *N*-methyl carbamates, glyphosate, and phenylurea herbicides) are analyzed after derivatization to enable fluorescence detection. The electrochemical detectors are used for a number of pesticide residues (e.g., captan) in relatively clean samples.

The on-line combination of HPLC and mass spectroscopy (HPLC–MS) offers a high sensitivity and specificity, and its use in the field of pesticide residues analysis is growing. There are a number of ionization techniques used to interface HPLC with MS analyzers, of which the most widely used are electrospray and atmospheric pressure chemical ionization. The HPLC–MS methods use soft ionization techniques which typically produce protonated or deprotonated pseudomolecular ions. Therefore, the chromatographic data do not provide structural information except when the ions produced are subjected to successive fragmentation (MS^n). (*See* **Chromatography**: High-performance Liquid Chromatography.)

Derivatization

Some pesticide residues require derivitization to enhance the extractability, clean-up or subsequent chromatographic resolution and determination steps. For example, pesticides with hydroxy groups are not suited to GC analysis, and such an analysis may be possible only after derivatization to esters. Furthermore, esters of certain functional groups can enhance the detection process, e.g., pentafluorobenzyl derivatives produce a high response on the ECD.

Dithiocarbamate pesticides break down to carbon disulfide (CS_2) during analytical procedures. These residues are determined after treatment of samples with acidic tin (II) chloride. Any dithiocarbamate residues in the sample break down to produce CS_2 gas, which can be trapped in the reaction chamber. An aliquot of the gas (headspace) in the reaction chamber is analyzed for CS_2 using GC–FPD. Alternatively, the gas produced can be absorbed into a 2,2,4-trimethyl

pentane, and an aliquot of the liquid layer is then analyzed using GC. This approach is more robust and more amenable to GC–MS analysis compared with the headspace procedure.

Some pesticides containing sulfur may oxidize to form sulfoxide and sulfone derivatives before or during analysis. These products are also toxic and are included in the residue definition for the monitoring purposes. These residues are analyzed after complete conversion of the pesticide and its sulfoxide to the corresponding sulfone and thus enable combined measurement of the pesticide, its sulfoxide, and sulfone residues. The conversion step involves the treatment of sample extracts with potassium permanganate in the presence of 2-methyl propan-2-ol. The sulfone formed is then extracted into an organic solvent and analyzed by GC.

Other Techniques

Enzyme-linked immunosorbent assay (ELISA) methods are used for rapid screening of an individual or a group of closely related pesticides. These methods require little or no sample clean-up, require no expensive instrumentation, and are suitable for field use. ELISA methods are especially suitable for residue analyses that are not possible using multiresidue methods. ELISA kits are available for a number of pesticides, including 2,4-D, aldicarb, carbendazim, thiabendazole, chlopyrifos, diazinon, endosulfan, and metalaxyl.

Confirmation

For regulatory purposes, it is essential that pesticide residues be unequivocally confirmed using MS. However, if an MS method is not available, the sample extract is reanalyzed using a different chromatographic column and/or a different detection system to confirm the initial results.

Emerging Techniques

There are continued improvements in the design of instruments available for residue analysis. The use of GC injectors capable of injecting large volumes can enable determination at low levels. The use of fast GC, which utilizes an improved design for heating the columns, can enable faster chromatographic runs and thus enable quicker analysis. Improvements in GC instrumentation have enabled precise control over temperature and gas flow rate. The use of electronic pressure control devices can enable more reproducible chromatographic runs, thus improving the quality of data. These advances, coupled with more sophisticated software, can enable more reproducible chromatography, with a typical retention time variation of 0.01 s. Improvements in MS instruments will continue to enhance the selectivity of methods.

See also: **Chromatography**: Principles; High-performance Liquid Chromatography; Gas Chromatography; **Mass Spectrometry**: Principles and Instrumentation; Applications; **Pesticides and Herbicides**: Types of Pesticide; Types, Uses, and Determination of Herbicides; Toxicology

Further Reading

Anon (1991) *Concerning the placing of PPPs on the market, Council Directive 91/414/EEC, Brussels: Official Journal No. L 230*. Luxembourg: Office of Official Publications of the European Communities.

Cairns T and Sherma J (eds) (1992) *Emerging Strategies for Pesticide Analysis*. Boca Raton, FL: CRC Press.

Chapman JR (1993) *Practical Organic Mass Spectroscopy, A Guide for Chemical and Biochemical Analysis*. Chichester, UK: John Wiley.

Codex Alimentarius (1993) *Codex Commission Guidelines, Recommended Method of Sampling for Determination of Pesticide Residues*, vol. 2, Section 3. Rome: Food and Agriculture Organization of the United Nations.

Codex Alimentarius (1996) *Pesticide Residues in Food, Maximum Residue Limits*, vol. 2B. Rome: Food and Agriculture Organization of the United Nations.

Fong WG (ed.) (1999) *Pesticide Residues in Foods: Methods, Techniques and Regulations*. New York: John Wiley.

Hill ARC and Reynolds SL (1999) Guidelines for in-house validation of analytical methods for pesticide residues in food and animal feeds. *Analyst* 124: 953–958.

Krause RT (1979) Resolution, sensitivity and selectivity of a high performance liquid chromatographic post-column fluorometric labeling technique for determination of carbamate insecticides. *Journal of Chromatography* 185: 615–624.

Mellon F, Self R and Startin JR (2000) *Mass Spectroscopy of Natural Substances in Food*. Cambridge: The Royal Society of Chemistry.

Sherma J (1981) *Manual of Analytical Quality Control for Pesticide Residues and Related Compounds, USA: EPA 600/2–81–059*. US Environmental Protection Agency, Research Triangle Park.

The Working Party on Pesticide Residues (1997) *Unit to Unit Variation of Pesticide Residues in Fruit and Vegetables*. York, UK: Pesticide Safety Directorate.

Tomlin C (ed.) (1997) *The Pesticide Manual, A World Compendium*, 11th edn. Bracknell, UK: The British Crop Protection Council.

Zooner P Van (1996) *Analytical Methods for Pesticide Residues in Foodstuffs*, 6th edn. The Hague, The Netherlands: Ministry for Public Health.

Toxicology

C K Winter, University of California Davis, CA, USA

Background

The use of agricultural chemicals, collectively known as pesticides, in the past several decades has led to significant reductions in crop losses resulting from insects, weeds, and plant diseases throughout the world. The toxicological properties that pesticides possess also present the potential for impacts upon human health and upon the environment. As an example, agricultural workers involved in the mixing, loading, and/or application of pesticides as well as those working in fields previously treated with pesticides face potential health risks resulting from excess exposure to the pesticides. Consumers are routinely exposed to pesticide residues in their foods, and the potential dietary risks from pesticide exposure have been the subject of considerable government study, regulation, and societal concern.

This review focuses upon the toxicology of the various types of pesticides used, how pesticide residues in foods are regulated, and the magnitude of potential risks faced by consumers from pesticide residues in the food supply.

Pesticides

Classification

The US Federal Insecticide, Fungicide, and Rodenticide Act defines a pesticide as 'any substance or mixture of substances intended for preventing, destroying, repelling, or mitigating any pest, any substance or mixture of substances intended for use as a plant regulator, defoliant, or desiccant, and any nitrogen stabilizer....' Under this broad definition, it is clear that a variety of pesticide types exist to control a wide number of different types of pests. A commonly held perception is that pesticides refer primarily to agricultural chemicals that control insects (insecticides). According to the US definition, however, pesticides also refer to chemicals that control plant diseases (fungicides) and weeds (herbicides) as well as a variety of other 'pests' (Table 1). For the purposes of consistency, all types of pesticides, including herbicides, will be considered under this broad umbrella in this article.

Pesticide Use

According to the US Environmental Protection Agency (EPA), approximately 2 billion kg of chemicals were used as pesticides in the USA in 1997. It should be noted that the majority of pesticide use was not for agricultural purposes. For example, 53% of pesticide use (by volume) involved chlorine or hypochlorites used for disinfection of potable and wastewater pools. 'Conventional' pesticides, defined as those developed or produced exclusively or primarily for use as pesticides, accounted for the remaining 47% of pesticide use by volume; 77% of this total was for agricultural uses, and 12% represented industry/government/commercial use, with the remaining 11% resulting from home and garden use.

Figure 1 shows the relative amounts of a variety of pesticide types used in US agriculture in 1997. Nearly half of the total volume of agricultural pesticide use came from herbicides and plant growth regulators (213 million kg), followed by sulfur/oils (65 million kg) and fumigants/nematicides (63 million kg). Agricultural insecticide use in 1997 was approximately 37 million kg, and 26 million kg of fungicides were used for agricultural purposes.

In terms of trends, total agricultural pesticide use, in terms of kg applied, has decreased slightly since 1979, with the largest drops in use representing insecticides, sulfur, and oils. The use of herbicides has been relatively steady.

Toxicity

Hundreds of different pesticide active ingredients are presently registered by the EPA, and nearly 300 pesticides are considered to have the ability to leave residues on food crops. Some of the more common classes of pesticides and some representative examples are given in Table 2. A comprehensive review of the toxicity of all pesticides is clearly beyond the scope of this article, and those interested in more specific and/or detailed pesticide toxicity information are encouraged to consult articles cited in the Further Reading section.

Table 1 Pesticide types and targets

Pesticide type	*Pest controlled*
Acaricide	Mites
Algacide	Algae
Bacteriocide	Bacteria
Defoliant	Leaves
Fungicide	Fungi
Herbicide	Weeds
Insecticide	Insects
Molluscicide	Snails
Nematicide	Nematodes
Rodenticide	Rodents

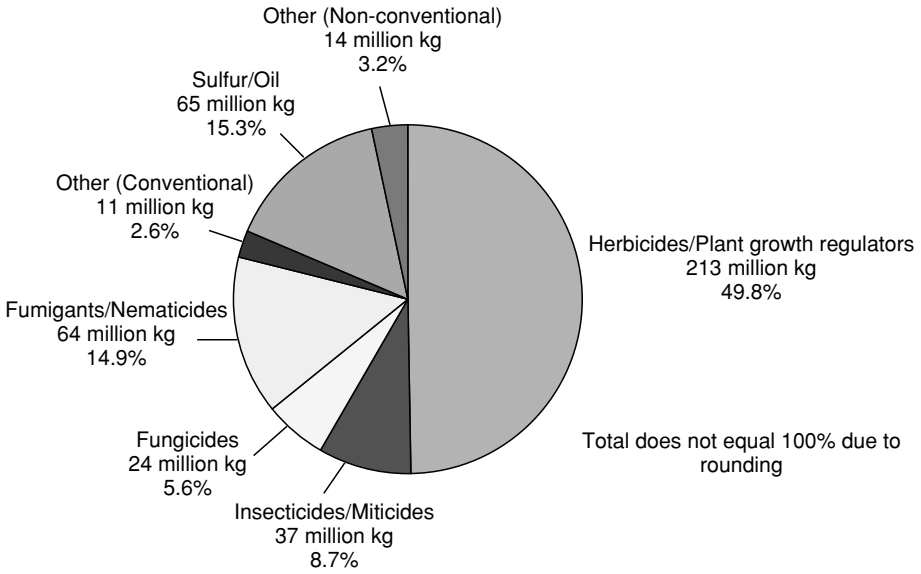

Figure 1 Agricultural use of pesticides in the USA, 1997.

Table 2 Common pesticide classes and representative examples

Pesticide	*Examples*
Insecticides	
Chlorinated hydrocarbons	Dicofol, methoxychlor, DDT, aldrin, chlordane
Organophosphates	Parathion, malathion, chlorpyrifos, azinphos-methyl
Carbamates	Aldicarb, carbaryl, carbofuran
Pyrethroids	Permethrin, cypermethrin
Herbicides	
Triazine	Atrazine, simazine, cyanazine
Phenoxy	2,4-D, 2,4,5-T, MCPA
Quaternary ammonium	Paraquat, diquat
Benzoic acids	Dicamba
Acetanilides	Alachlor, metolachlor
Ureas	Linuron
Fungicides	
Inorganic	Sulfur
Ethylenebisdithiocarbamates	Maneb, mancozeb, zineb
Chlorinated phenols	Pentachlorophenol

Insecticides Insecticides are represented by a variety of different chemical classes involving a variety of mechanisms of toxic action on insects as well as mammals such as humans. Examples of mechanisms of action include metabolic interference of the nervous or muscle systems, desiccation, and sterilization. Common types of insecticides include the chlorinated hydrocarbons, organophosphates, carbamates, and pyrethroids.

The first major synthetic class of insecticides, the chlorinated hydrocarbons, was developed during the 1930s and 1940s. Representative members of this insecticide class include DDT, aldrin, dieldrin, and chlordane. The chlorinated hydrocarbons are very potent nerve toxins to insects, and their initial use led to significant improvements in insect control. Their high insect potency, combined with generally low mammalian toxicity, provided excellent selectivity and insect control that was further enhanced by their high environmental persistence. In subsequent years, however, their resistance to environmental decay, coupled with their widespread continuing use, resulted in environmental build-up and food-chain magnification that resulted in significant environmental and ecological damage. Today, very few chlorinated hydrocarbon insecticides remain registered for use in the USA, because of environmental concerns, although their use may still be significant in many areas of the world. Many recent studies have also associated chlorinated hydrocarbon insecticides with potential adverse effects on human and nontarget organism fertility and reproduction arising from enzyme-inducing or estrogenic properties of the chemicals.

Historically, many of the uses of chlorinated hydrocarbon insecticides were replaced by the organophosphate and carbamate insecticides. In contrast to the chlorinated hydrocarbon insecticides, the organophosphates and carbamates break down quite rapidly in the environment following application. They also have a far greater acute toxicity in mammals. The organophosphates and carbamates are derived as esters of two very different chemical families but share a common mechanism of toxicological action

in both insects and mammals that involves inhibition of cholinesterase enzymes that can disrupt the nervous system.

A newer class of insecticides, the pyrethroids, are synthetic derivatives of pyrethrins (natural extracts from chrysanthemums) that are made to be more light-stable than their natural predecessors and thus more effective as insecticides. The pyrethroids were introduced in the 1970s and have broad-spectrum activity against many insects while possessing a much lower mammalian toxicity than the organophosphates or carbamates. The pyrethroids are frequently used in agricultural pest control, but their use is still limited by their environmental lability, relatively high cost, and their tendency to lose effectiveness through the development of insect resistance.

Herbicides Several different types of herbicides exist (Table 2), and each type has its own mechanism of toxic action on weeds. Since weeds and mammals differ dramatically in terms of metabolic function, herbicides targeting specific metabolic pathways in weeds frequently have little effect, and therefore a low relative toxicity, in mammals including humans.

The timing of herbicide application may determine significantly whether the potential exists for consumers to be exposed to herbicide residues in foods. Some herbicides (preplant herbicides) are applied before a crop is planted, whereas others (preemergent herbicides) are applied after planting but prior to the appearance of weeds, and others (postemergent herbicides) are used after the weeds have germinated. Some herbicides have broad-spectrum activity that makes them toxic to most types of plant life, including the crop being produced. One such example is glyphosate. Recent developments in genetic engineering have led to the development of varieties of soy, corn, and cotton that are resistant to glyphosate, meaning that glyphosate applications made while the crop is growing will control weeds without affecting the crop.

Other herbicides, such as the phenoxy herbicides (2,4-D, MCPA) are more selective in their toxicity; they are toxic to broad-leaf plants but not to narrow-leaf plants such as grasses. Some epidemiological studies have suggested links between agricultural worker exposure to phenoxy herbicides and certain types of cancer.

Fungicides A wide number of different types of fungicides are available for use in agriculture (Table 2). Fungicides control molds and other plant diseases by interfering with the growth and/or metabolic processes of fungal pests. In addition to improving crop yields, fungicides may provide human health benefits by reducing the production of mycotoxins (naturally occurring toxins produced by fungi living on the food crop) such as aflatoxins and fumonisins.

Pesticides suspected as carcinogens The potential carcinogenic (cancer-causing) effects of many pesticides have generated considerable public concern and regulatory scrutiny. Although the consumption of foods containing pesticide residues has not been correlated with the development of human cancers, some epidemiological studies have linked the occupational use of specific pesticides with the development of cancers. Most pesticides considered as carcinogens, however, owe their status as suspected carcinogens to the results of long-term animal toxicology studies in rodents such as rats and mice.

The carcinogenic potential of pesticides and other chemicals is typically determined through long-term animal bioassay studies. In these studies, animals are frequently given a high dose (the maximum tolerated dose, or MTD, that may cause considerable toxicity but does not reduce life expectancy), a lower dose (commonly one-half to one-fourth of the MTD), and a control (zero) dose. Relatively high doses of chemicals are given to the animals to maximize the potential to identify toxicological effects such as tumor production.

The amounts of pesticides that cause cancer in animal studies frequently exceed the amounts that humans are exposed to from the diet by several orders of magnitude. It has been argued that many pesticides considered to be carcinogens receive such classification because of the high doses used in the long-term animal studies and that the results of such studies may not be applicable to human exposure to much lower levels of the pesticides. As an example, it is well established that high-dose toxicity may lead to increased rates of cell proliferation that in itself is linked to the development of cancer in laboratory animals. Human exposure at lower levels of exposure would not be expected to trigger such a response and, as a result, would not be expected to lead to the development of cancer.

Long-term animal studies are also prone to inconclusive and/or contradictory results among different test animals and under different conditions. As a result, the EPA has developed a weight-of-the-evidence evaluation scheme based upon results of any human data and animal testing to determine the likelihood of a pesticide to be carcinogenic. A list of several pesticides considered by the EPA to be 'probable' carcinogens is provided in Table 3. Some pesticides may receive classification a 'possible' carcinogens, whereas many others are considered noncarcinogens.

Table 3 Some pesticides considered to be 'probable' carcinogens by the US Environmental Protection Agency

Pesticide	*Pesticide type*
Acetochlor	Herbicide
Aciflourfen sodium	Herbicide
Alachlor	Herbicide
Amitrol	Herbicide
Cacodylic acid	Herbicide
Chlorothalonil	Fungicide
Creosote	Wood preservative
Cyproconazole	Fungicide
Fenoxycarb	Insect growth regulator
Folpet	Fungicide
Heptachlor	Insecticide
Iprodione	Fungicide
Lactofen	Herbicide
Lindane	Insecticide
Mancozeb	Fungicide
Maneb	Fungicide
Metiram	Fungicide
Oxythioquinox	Insecticide
Pentachlorophenol	Fungicide
Pronamide	Herbicide
Propargite	Insecticide
Propoxur	Insecticide
Terrazole	Fungicide
Thiodicarb	Insecticide
TPTH	Fungicide
Vinclozolin	Fungicide

Regulating and Monitoring Pesticide Residues in Foods

The use of pesticides does not necessarily imply that food residues will occur. Many pesticides are applied to nonfood crops, whereas others such as broad-spectrum herbicides could damage or eliminate a crop if misapplied. The timing of the pesticide application is another important factor; many pesticides are applied to food crops prior to the development of edible portions of the crop, whereas other pesticides used on food crops may not result in residues, because of rapid environmental degradation between the time of application and the time of harvest.

Pesticide residue regulation In cases where the use of a pesticide on a food crop may present the potential to leave a residue in the USA, a maximum permitted allowable residue level, or tolerance, is established. Tolerances are specific to combinations of pesticides and commodities; it is possible for the same pesticide to have different tolerance levels established on different commodities, and several different pesticide tolerances for distinct pesticides may be established on the same commodity.

Although it may seem counterintuitive, pesticide tolerances are not based upon safety but rather represent the maximum expected residue of a pesticide on a particular commodity resulting from the legal use of a pesticide. The maximum levels are determined from the results of controlled field studies performed by the pesticide manufacturer using the 'worst legal case' conditions such as the maximum recommended application rate, maximum number of applications per growing season, and harvesting at the minimal anticipated time following harvest. Pesticide manufacturers typically petition the EPA to establish the tolerances at or slightly above the highest levels determined from the controlled field studies. As such, the values selected for tolerances are determined solely on the basis of agricultural practices but not as a result of human health risk assessments. Pesticide tolerances therefore should be considered to represent enforcement tools to determine whether pesticide applications may have been made in accordance to legal requirements; in cases where residues exceed the established tolerances, it is likely that such residues resulted from misapplication of the pesticides. Such a finding, however, rarely constitutes an 'unsafe' residue according to standard toxicological criteria. Pesticide tolerances, therefore, should be viewed as enforcement tools but not as standards of safety.

Before the EPA grants a tolerance, human health risk assessments are performed to determine the conditions for acceptable pesticide use. Such conditions include the listing of commodities on which a specific pesticide may be used, the target pests controlled by the pesticide, application requirements, and the acceptable interval between application and harvest. In the EPA's assessment of acceptable levels of consumer exposure to pesticides, it considers potential human exposure from all registered (and proposed) uses of the pesticide. If the resulting risk is deemed to be excessive, the EPA will not allow tolerances to be established for specific commodities. If the risks are considered to be acceptable, the tolerances are established, as discussed in the preceding paragraph.

The processes that the EPA uses to determine the acceptability of dietary pesticide risk are quite complicated and subject to ongoing evolution to meet the needs of new regulations, improved toxicology testing, and advances in computational methods.

The first step in evaluating the consumer risks from pesticide residues involves making estimates of the amount of consumer exposure. The maximum legal exposure to the pesticide is frequently calculated by assuming that (1) the pesticide is always used on all food items for which it is registered and/or proposed for registration, (2) all residues on the food items will be present at the established or proposed tolerance levels, and (3) there will be no reduction in residue levels resulting from postharvest effects such as washing, cooking, peeling, processing, and

transportation. This approach leads to the calculation of the theoretical maximum residue contribution (TMRC). Although studies have indicated that the TMRC values may overestimate the actual consumer exposures by factors of 100–100 000, the TMRC provides a starting point from which to estimate consumer risks.

In practice, the TMRC is compared with established toxicological criteria such as the reference dose (RfD) or the acceptable daily intake (ADI) that represent, following analysis of animal toxicology data and extrapolations to human health, the daily exposure levels that do not constitute an appreciable level of risk. In cases where the EPA determines that exposure to a pesticide at the TMRC is below the RfD or ADI, the risks for the pesticide are typically deemed to be negligible, and the EPA allows tolerances to be established for the pesticide on specific commodities. For pesticides that are considered as potential carcinogens, the EPA also requires that the quantitative carcinogenic risk to the pesticide be below one excess cancer per million using models that calculate possible human risks from low levels of exposures to potentially carcinogenic pesticides. Such models are developed from the results from long-term studies on animals given high doses of the pesticides. In cases where the exposures at the TMRC exceed the RfD or ADI, or in cases where the carcinogenic risks at the TMRC exceed one excess cancer per million, the EPA may adopt a refined risk assessment that more accurately expresses exposures. Such refinements may include adjustments of actual pesticide use, the use of more realistic pesticide residue data, and consideration of postharvest effects that may significantly reduce residue levels prior to consumption. In cases where the refined exposure estimates are below the RfD or ADI and where the carcinogenic risks at the refined exposure estimates are below one excess cancer per million, the EPA will typically allow tolerances to be established.

The processes by which the EPA determines human risks from pesticide exposure became more complicated following the passage and adoption of the Food Quality Protection Act (FQPA) of 1996. It is clear that the new requirements of FQPA will require scientists to develop improved methods for assessing risks to pesticides and that once such methods are adopted, the regulatory requirements may be much more stringent and may lead to reductions in the amounts and types of pesticides that may be used on food crops.

Prior to FQPA, the EPA allowed tolerances to be established on a chemical-by-chemical basis and considered only exposures resulting from dietary pathways. FQPA now requires the EPA to establish tolerances only when the risks posed by pesticides represent a 'reasonable certainty of no harm.' In determining what constitutes this 'reasonable certainty of no harm,' the EPA must now consider the *aggregate* exposure to pesticides from dietary, drinking water, and residential sources as well as the *cumulative* exposure from pesticides possessing a common mechanism of toxic action. As such, the EPA may consider the risks from entire families of chemicals rather than the risks from individual chemicals. Another important provision of FQPA is the so-called $10 \times factor$, which requires the EPA to consider applying an additional 10-fold uncertainty factor in cases where infants or children may be more susceptible to the toxicological effects of pesticides than adults. The application of the full $10 \times factor$ would result in a subsequent reduction in the RfD by a factor of 10.

Internationally, many countries adopt the Codex Alimentarius maximum residue limits (MRLs), which, like the US tolerances, exist primarily as enforcement tools to determine if pesticide applications are made following good agricultural practices. Although many US tolerances and Codex Alimentarius MRLs are identical, there are many cases in which the US tolerances are more restrictive and many others in which the Codex Alimentarius MRLs are more restrictive.

Pesticide residue monitoring *Authority* In the US, the Food and Drug Administration (FDA) has the primary responsibility for enforcing tolerances in domestic and imported foods. Domestic food samples are frequently collected near the source of production or at the wholesale level, whereas imported food samples are typically taken at the point of entry into the US. The types and quantities of samples taken by FDA are determined by a variety of factors such as regional intelligence on pesticide use, the dietary importance of specific foods, information on the amount of foods that enter interstate commerce, and pesticide use patterns. Samples are analyzed using multiresidue methods capable of detecting over 200 individual pesticides.

The US Department of Agriculture (USDA) also has a role in pesticide residue monitoring, as it is responsible for the monitoring of meat, poultry, and egg products. It also conducts the Pesticide Data Program (PDP), which collects residue data on fruits, vegetables, and processed foods. Findings from the USDA's PDP are considered to be more representative of the actual food supply than those collected from the FDA's regulatory monitoring programs and are commonly used by EPA to aid in its risk-assessment efforts.

US monitoring of pesticide residues also occurs at the state level. The largest state pesticide regulatory and monitoring program exists in California, and

several other states, including Texas and Florida, have significant pesticide-monitoring programs.

Residue findings The most recent FDA pesticide monitoring residue data are available for 1999. During that year, FDA analyzed 9438 food samples for pesticide residues. More samples were taken from imported foods (6012 samples, or 63.7%) than from domestic foods (3426 samples or 36.3%).

The results of FDA's 1999 monitoring of pesticide residues in imported foods are shown in Figure 2. Overall, 65.0% of the samples showed no detected residues, and violations were identified in 3.1% of the samples. Figure 3 shows the comparable results from domestic foods where 60.2% of the samples showed no detectable residues, and violations were present in 0.8% of the samples.

Pesticide residue violations commonly take one of two forms. The most common form of a violative residue results when residues of a pesticide are detected on a commodity for which no tolerance has been established. This type of violation may result from application of the pesticide to the wrong commodity, uptake from soil contaminated from a prior use of the pesticide on a different commodity, or drift of a pesticide from an adjacent field. The other type of violation results when residue levels are detected in excess of the established tolerance. In 1999, 90% of the import violations occurred where residues were detected on commodities for which tolerances were not established, and only 10% of the violations represented residues in excess of the tolerance. For domestic samples, 69% were detected on commodities for which tolerances were not established, and the remaining 31% of violations occurred when levels exceeding tolerances were detected. Fruits and vegetables were responsible for the highest percentages of residue detections and the greatest number of violations from both imported and domestic foods.

Results obtained from California's 1997 Marketplace Surveillance Program were quite similar to those of the FDA's 1999 monitoring. This program analyzed 5660 samples in 1997, with 62.2% of those originating from California, 6.7% from other US states, and the remaining 31.1% from other countries. The majority of samples (62.1%) showed no detectable residues, whereas 36.7% of the samples contained legal residues. Violations were detected in 1.2% of the samples, and only 12% of the violations represented residues detected in excess of tolerances. The violation rate from imported foods was 2.4%, and the violation rate for domestic foods was 0.7%.

The USDA's PDP analyzed 9125 food samples for pesticide residues in 1999. Foods analyzed included apples, cantaloupe, cucumbers, grape juice, lettuce, oats, pears, spinach, strawberries, sweet bell peppers, tomatoes, winter squash, and corn syrup. Most of the samples (8637) were from fruits and vegetables, with lower numbers of samples collected for oats (332) and corn syrup (156). The majority of samples (79%) was of domestic origin. Overall, 36% of the samples contained no detectable residue, whereas

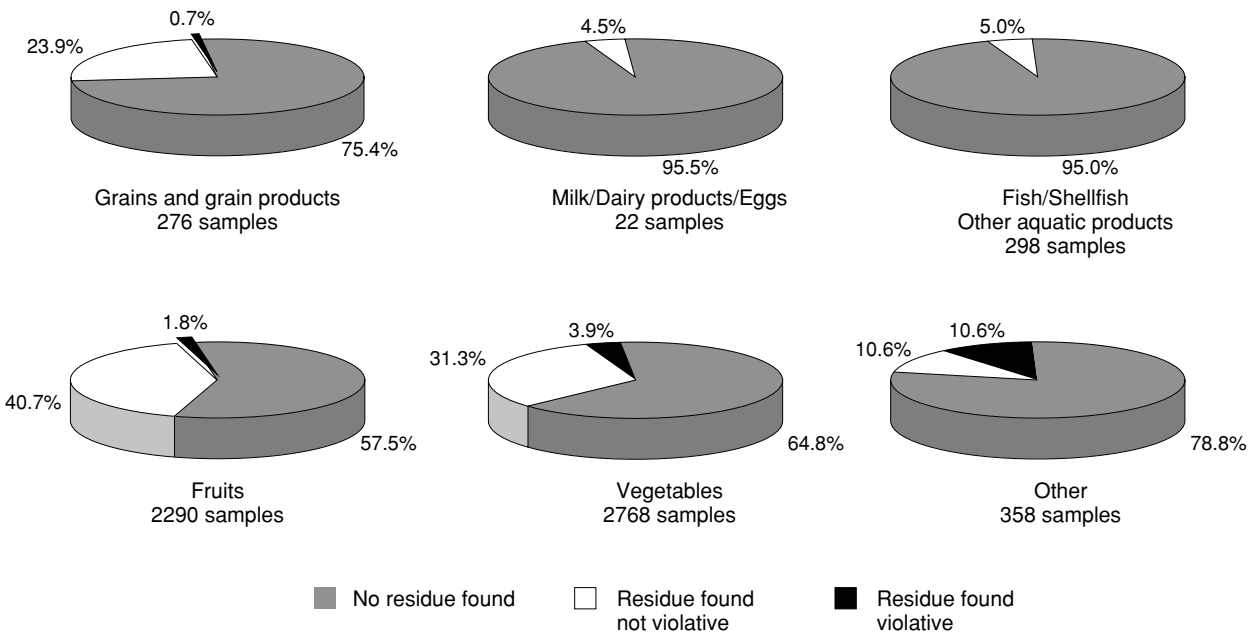

Figure 2 Results of US Food and Drug Administration monitoring of imported foods for pesticide residues, 1999.

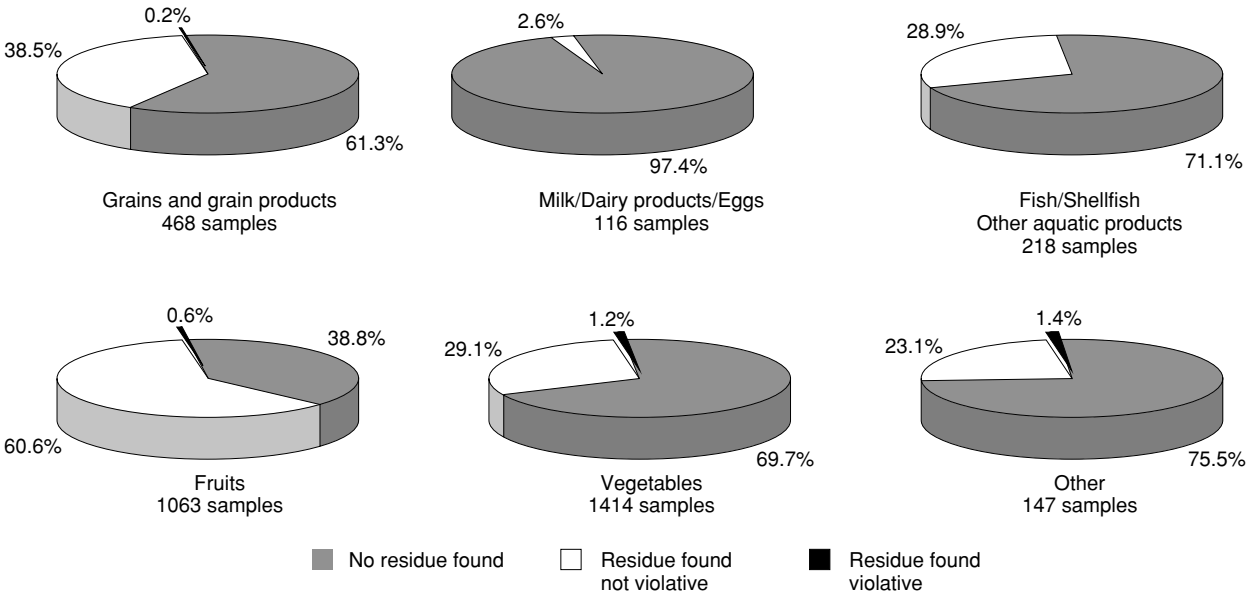

Figure 3 Results of US Food and Drug Administration monitoring of domestic foods for pesticide residues, 1999.

26% contained one residue, and 35% contained more than one residue. Residues exceeding the tolerance level were detected on 0.3% of the samples, and residues of pesticides were detected on commodities for which no tolerances of the pesticides were established on another 3.7% of the samples.

Dietary Risks from Exposure to Pesticides in Food

A common method used to discuss the potential human health risks arising from the consumption of pesticide residues in the diet is to report on the results of regulatory monitoring programs, as has been done in the previous paragraphs. Results indicate that a large percentage of food samples contain no detectable residues and that violation rates are relatively low, particularly for foods grown in the US. Frequently, the recitation of such findings is used as justification for a lack of significant human health risks resulting from pesticide residues in foods.

The problem with adopting such an approach is that it ignores the important, yet confusing, fact that pesticide tolerances are *not* safety standards but rather represent the maximum legal residues expected when pesticides have been used according to directions. Taking this a step further, violative residues should not be considered to be *unsafe* residues in most cases but merely represent cases in which pesticides have been misapplied or have been transported to commodities for which they are not registered. Although cases of pesticide misuse have historically resulted in a small number of incidents of acute poisoning of people consuming tainted foods, it can be reasonably argued that the vast majority of violative pesticide residues do not represent significant health threats to consumers based upon common toxicological and risk assessment criteria.

A more accurate approach for estimating human dietary risks from pesticides is to use exposure data derived from market basket surveys rather than regulatory monitoring data. The FDA, for example, performs its Total Diet Study annually. This study uses a market basket approach, with each market basket containing more than 250 individual food items. Foods are collected by FDA inspectors from three cities in each of the four geographical regions of the USA and are prepared for table-ready consumption prior to analysis for pesticide residues. By combining analytical results from estimates of typical consumption rates of the various food items or their components, it is possible to estimate the typical daily exposure of members of the general population to specific pesticides as well as to break the population down further into subgroups defined by factors such as age, gender, and geographical location.

Although the FDA no longer makes its dietary pesticide exposure estimates from its Total Diet Study available to the public, results reported from the 1991 Total Diet Study indicate that, for most pesticides in a variety of different population subgroups, the typical daily exposure estimates represent only a small fraction (often less than 1%) of the corresponding RfDs or ADIs. To put this into some level of perspective, it should be noted that typical RfDs are derived by first identifying the highest level of exposure to a pesticide that causes no noticeable

signs of toxicity in laboratory animals and then dividing that level by an uncertainty factor (usually 100) that presumably covers potential variability resulting from the animal to human extrapolation and from interhuman variability. Exposure at a level of 1% of the RfD represents an exposure 10 000 times below the level that does not produce noticeable effects in the animals. Such findings provide an illustration of why the majority of health professionals consider the typical human health risks from pesticides in the diet to be much lower than food safety risks posed by such factors as microbiological contamination of foods, nutritional imbalance, environmental contaminants, and naturally occurring toxins. The risks from pesticides in the human diet are clearly not zero, since consumption of tainted foods has caused documented human illnesses throughout the world, and concerns remain regarding potential long-term effects of dietary pesticide exposure.

It is also clear that the potential health benefits resulting from pesticide use should be considered. Pesticide use has resulted in increases in production of a wide variety of food crops, which translates into greater availability and lower consumer costs and thus a greater potential consumption of agricultural products. Epidemiological studies have clearly indicated that diets rich in consumption of fruits, vegetables, and grains may significantly decrease one's risk of heart disease and certain types of cancer. The US National Academy of Sciences, among other scientific bodies, has concluded that the theoretical increased risks from pesticide exposure resulting from increases in consumption of fruits, vegetables, and grains were greatly outweighed by the health benefits of these foods.

See also: **Cancer**: Epidemiology; Carcinogens in the Food Chain; **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; **Food and Drug Administration**; **Food Poisoning**: Classification; **Pesticides and Herbicides**: Types of Pesticide; Types, Uses, and Determination of Herbicides

Further Reading

Fong WG, Moye HA, Seiber JR and Toth JF (1999) *Pesticide Residues in Foods: Methods, Techniques, and Regulation*. New York: John Wiley.

Hayes WJ and Laws ER (1991) *Handbook of Pesticide Toxicology*. San Diego, CA: Academic Press.

National Research Council (1987) *Regulating Pesticides in Foods: The Delaney Paradox*. Washington, DC: National Academy Press.

National Research Council (1993) *Pesticides in the Diets of Infants and Children*. Washington, DC: National Academy Press.

Tweedy BG, Dishburger HJ, Ballantine LG and McCarthy J (1991) *Pesticide Residues and Food Safety: A Harvest of Viewpoints*. Washington, DC: American Chemical Society Symposium Series No. 446.

Winter CK (1992a) Pesticide tolerances and their relevance as safety standards. *Regulatory Toxicology and Pharmacology* 15: 137–150.

Winter CK (1992b) Dietary pesticide risk assessment. *Reviews of Environmental Contamination and Toxicology* 127: 23–67.

Winter CK (2001) Contaminant regulation and management in the United States: The case of pesticides. In: Watson DH (ed.) *Food Chemical Safety, Volume 1: Contaminants*, pp. 295–313. Cambridge, UK: Woodhead.

Winter CK and Francis FJ (1997) Assessing, managing, and communicating chemical food risks. *Food Technology* 47: 85–92.

PH – PRINCIPLES AND MEASUREMENT

D Webster, Formerly of University of Hull, Hull, UK

Background

pH measurements have been, and continue to be, widely used as a rapid, accurate measure of the acidity of fluids of all sorts. There are two methods for measuring pH: colorimetric methods using indicator solutions or papers, and the more accurate electrochemical methods using electrodes and a millivoltmeter (pH meter). The development of the glass electrode, which is convenient to use in a variety of environments, and the development of the pH meter have enabled the widespread application of pH measurement and control to take place. The determination, and hence the control of pH, is of great importance in the food industry.

Basic Theory

In water, molecules (H_2O) are in equilibrium with hydrogen ions (H^+) and hydroxide ions (OH^-) (eqn (1)).

$$H_2O \rightleftharpoons H^+ + OH^-. \quad (1)$$

This ionization is of great importance in the chemistry of aqueous systems, as it enables water to give or take H^+ ions, as required, by other dissolved substances.

Applying the law of mass action to eqn (1):

$$[H^+][OH^-]/[H_2O] = \text{constant}, \quad (2)$$

where [ ] indicates the concentration in units of moles per cubic decimeter (mol dm^{-3}).

In pure water and dilute solutions, the concentration of the undissociated water may be considered constant, hence $[H^+][OH^-] = K_w$, where K_w is a constant called the ionic product of water.

The ionic product varies with temperature, but at about 25 °C, its value is 10^{-14} mol^2 dm^{-6} (K_w (0°C) = $10^{-14.9}$, $K_w(25°C) = 10^{-14.0}$, $K_w(60°C) = 10^{-13.0}$). This means that in pure water, the ionization is very small. As the concentrations of H^+ ions and OH^- ions are equal in pure water, and as $[H^+][OH^-] = 10^{-14}$ mol^2 dm^{-6} at 25 °C, $[H^+]$ (or $[OH^-]$) = 10^{-7} mol dm^{-3}.

To be strictly correct, eqn (1) should be written as eqn (3):

$$2H_2O \rightleftharpoons H_3O^+ + OH^-, \quad (3)$$

where the H^+ ion is attached to a water molecule to form the oxonium ion (H_3O^+). However, the symbol H^+ will be used throughout this article.

When, in a solution, there are an equal number of H^+ and OH^- ions, the solution is said to be neutral, when there is an excess of H^+ ions (> 10^{-7} mol dm^{-3} at 25 °C), it is acidic, and when there is an excess of OH^- ions ([H^+ < 10^{-7} mol dm^{-3}), it is basic or alkaline.

pH Scale

Although concentrations of H^+ and OH^- ions in acidic and alkaline solutions can be expressed in these molar concentrations, a much more convenient method was introduced by Sörensen in 1909. He proposed the use of the H^+ ion exponent pH, defined by the relationship $pH = -\log_{10}[H^+]$. (This may also be expressed as $[H^+] = 10^{-pH}$). To be strictly correct, this equation is $pH = -\log_{10}([H^+]/[1])$, as a logarithm must be dimensionless. From this, it can be seen that pH does not have units – the often-used expression 'pH unit' is wrong.

Using the pH scale, a change of 1 corresponds to a 10-fold change in the H^+ ion concentration, a change of 2 corresponds to a 100-fold change, etc. The pH scale has the advantage that all solutions from 1 mol dm^{-3} acid to 1 mol dm^{-3} alkali can be expressed by positive numbers from 0 to 14. Thus, at 25 °C, a neutral solution, which has $[H^+] = 10^{-7}$ mol dm^{-3}, has a pH of 7, a 1 mol dm^{-3} solution of a strong (completely ionized) acid such as hydrochloric acid, which has $[H^+] = 1 = 10^0$ mol dm^{-3}, has a pH of 0, and a 1 mol dm^{-3} solution of a strong (completely ionized) alkali acid such as sodium hydroxide, which has $[OH^-] = 1$ mol dm^{-3}, hence $[H^+] = K_w/[OH^-] = 10^{-14}$ mol dm^{-3}, has a pH of 14. pH values below 0 and above 14 occur in solutions of strong acids and alkalis of strengths greater than 1 mol dm^{-3}.

As K_w varies with temperature, pH measurements should be made at about 25 °C. The pH of a neutral solution at 0 °C is 7.45 ($K_w = 10^{-14.9}$, $[H^+] = \sqrt{10^{-14.9}} = 10^{-7.45}$ mol dm^{-3}, and at 60 °C, the pH is 6.5 ($K_w = 10^{-13.0}$, $[H^+] = \sqrt{10^{-13.0}} = 10^{-6.5}$ mol dm^{-3}).

To measure pH values, Sörensen used the electrochemical cell:

$$Pt, H_2(g) \mid \text{Solution X} \mid \text{Salt bridge} \mid Hg_2Cl_2 \mid Hg.$$

This representation indicates a hydrogen electrode (hydrogen gas passed over a platinum metal electrode) in a solution X and a calomel electrode (metallic mercury in contact with calomel (Hg_2Cl_2), in contact with a chloride ion solution) separated by a salt bridge (which allows ionic conduction to occur, but prevents solution X and the chloride ion solution from mixing). For such a cell, the difference in electromotive force (emf) ($E_1 - E_2$) between two cells in terms of the H^+ ion concentrations $[H^+]_1$ and $[H^+]_2$ of the two solutions is given by the equation

$$E_1 - E_2 = (RT/F)\log_e([H^+]_2/[H^+]_1), \quad (4)$$

where R is the gas constant, T the temperature, and F the Faraday constant. If $[H^+]_2$ is 1 mol dm^{-3}, E_2 will have a definite value, E^0. Therefore, $E_1 - E^0 = (RT/F)\log_e(1/[H^+]_1)$, or $E_1 = E^0 + (RT\log_e 10/F)p_sH_1 = E^0 + 0.05916\, p_sH_1 V$ (at 25 °C) (p_sH is used here for Sörensen's pH). Sörensen assumed that $[H^+]_1 = \alpha_1 M_1$ where α_1 is the degree of dissociation of the acid in solution 1, and M_1 is its molality. $E^0 = 0.3376$ V at 18 °C.

After this initial work of Sörensen, it was realized that the emfs of cells depend on activity rather than concentration, and other assumptions were incorrect. Sörensen and Linderström-Lang (1924) proposed $p_aH = -\log_{10} a_{H^+} = -\log_{10}[H^+]y_{H^+}$, where a_{H^+} is the activity, $[H^+]$ is the concentration of the H^+ ion in mol dm^{-3}, and y_{H^+} is the activity coefficient of the H^+ ion on the molarity scale. (To be strictly correct, a ratio of concentrations, as earlier, is needed here to give a dimensionless quantity.) These p_aH values can be shown to be related to the original p_sH values by the relationship:

$$p_aH = p_sH + 0.04. \quad (5)$$

Operational Definition of pH

In the previous section, some of the problems of the theory of pH were discussed. Careful study has shown that the numbers obtained depend, in a complex manner, on H^+ ion activity of electrolytes in solution, and it became clear than an operational definition of pH and a standard scale were needed to unify the great variety of pH measurements being made by research scientists and by industry and commerce. Hence, we now have an operational definition that has been endorsed by the International Union of Pure and Applied Chemistry. This definition is

$$pH(X) = pH(S) + (E_S - E_X)F/RT\log_e 10, \quad (6)$$

where E_x is the emf of the galvanic cell.

$$\text{Reference electrode} \,|\, \text{K Cl solution} \,|\, \text{Solution X} \,|\, H_2(g), \text{Pt}$$

and E_s is the emf of the same cell with solution X of unknown pH(X) replaced by solution S of standard pH(S).

In practice, a glass electrode is almost always used in place of the platinum/hydrogen gas electrode. The reference electrode is usually mercury/mercury(I) chloride (calomel), silver/silver chloride, or thallium amalgam/thallium(I) chloride. The standard reference pH is that of an aqueous solution of potassium hydrogen-phthalate of molality $0.05\,\text{mol}\,\text{kg}^{-1}$ at 25 °C. This has a pH of 4.005.

Food and pH

There are many instances when the food scientist needs to measure or control pH. For example, pH control is vital in the clarification and stabilization of fruit and vegetable juices, in the use and control of enzymes and microorganisms, in the preparation of foods by the fermentation of fruit or cereals, in the control of the texture of jams and jellies, and in the stability of the color and flavor of fruits.

The properties (e.g., emulsification and foaming ability) of many colloidal systems are affected by pH – these include proteins, pectins, and gums, all of which occur in foods. Chemical reactions that can occur in foods, particularly hydrolysis, are catalyzed by hydrogen ions, and therefore can be controlled by controlling the pH of the system. (*See* **Colloids and Emulsions**.)

Hydrogen ions affect the rate of growth of molds, yeasts, and, particularly, bacteria. Usually, there is a pH value at which optimum growth occurs, above and below which growth is inhibited. The relationship of pH to sterilization by heat is of particular significance during the canning of foodstuffs. (*See* **Canning**: Principles; **Spoilage**: Bacterial Spoilage; Molds in Spoilage; Yeasts in Spoilage.)

pH is a rough measure of the maturity of fruit. Young fruits initially have a pH close to that of the plant, but this rapidly decreases as the acidity increases. The pH of acidic fruits such as lemons and limes is about 2.0, and that of mildly acidic fruits about 4.0. These values can be compared with vegetables, which have pH values of 5–6. The pH ranges of a number of fruits and other foods is listed in Table 1. (*See* **Ripening of Fruit**.)

Buffer Solutions

Buffer solutions are of prime importance in pH determinations, as they are the standards for both colorimetric and electrochemical methods of determining pH.

Buffer solutions are solutions that resist a change in H^+ ion concentration (pH) when an acid or alkali is added to them. They are made up of a weak acid or weak base and its salt.

The equilibria in a solution of, say, a weak acid and its sodium salt are shown in eqns (7) and (8):

Table 1 pH values of some foods

Food	*pH*
Limes	1.8–2.0
Lemons	2.2–2.4
Gooseberries	2.8–3.0
Plums	2.8–3.0
Pickles	3.0–3.4
Grapefruit	3.0–3.4
Oranges	3.0–4.0
Rhubarb	3.1–3.2
Cherries	3.2–4.0
Pineapples	3.4–3.7
Pears	3.6–4.0
Apricots	3.6–4.0
Tomatoes	4.0–4.4
Bananas	4.5–4.7
Cheese	4.8–6.4
Carrots	4.9–5.3
Spinach	5.1–5.7
Potatoes	5.6–6.0
Peas	5.8–6.4
Tuna	5.9–6.1
Corn	6.0–6.5
Salmon	6.1–6.3
Butter	6.1–6.4
Chicken	6.2–6.4
Drinking water	6.5–8.0

$$HA \rightleftharpoons H^+ + A^- \quad (7)$$

$$NaA \rightleftharpoons Na^+ + A^-. \quad (8)$$

As HA is a weak acid, equilibrium (eqn (7)) at moderate concentration will lie over to the left, i.e., as undissociated acid HA. The equilibrium of the sodium salt (eqn (8)) is a source of A^- ions. If acid (i.e., H^+ ions) is added to this weak acid and salt solution, the H^+ ions will be removed as they will combine with the A^- ions to form undissociated HA. If base (i.e., OH^- ions) is added, the OH^- ions will be removed as they will combine with the H^+ ions to form water and the equilibrium (eqn (7)) will move to the right to supply the H^+ ions required. Hence, the concentration of H^+ ions in the solution (i.e., the pH) will not change significantly. This is the buffer effect.

Common buffer solutions are made from potassium hydrogenphthalate with hydrochloric acid or sodium hydroxide (for pH 3–5), potassium dihydrogenphosphate with sodium hydroxide (for pH 6–8), and boric acid with sodium hydroxide (for pH 8–10).

Biologically, the buffer effect is often more important than the absolute value of pH. For example, buffer chemicals affect the way in which enzymes and microorganisms respond when the pH is changed. Juices from plant tissue have the ability to buffer pH changes to a greater or lesser degree, owing to the presence of organic acids, acid–salt systems, proteins, and acid phosphates.

Measurement of pH

Colorimetric Methods

Determination of pH using the color of acid–base indicators is a very simple technique that can be carried out rapidly and reproducibly. Approximate pH values can be obtained very quickly using pH papers. Under optimum conditions, accurate values of pH can be obtained by colorimetric methods. No doubt, these methods will be used for many years to come, but must become less favored with the advent of portable, cheap, easy-to-use electrochemical pH meters. (*See* **Spectroscopy**: Visible Spectroscopy and Colorimetry.)

Color change of acid–base indicators Indicators are natural or artificial dyestuffs that are weak acids or bases (acids and bases that are only partially dissociated) and have different colors in their acidic and basic forms.

If the indicator is written as HIn and it dissociates (eqn (9))

$$HIn \rightleftharpoons H^+ + In^-, \quad (9)$$

the equilibrium (or ionization) constant is

$$K_{In} = [H^+][In^-]/[HIn], \quad (10)$$

and

$$[H^+] = K_{In}([HIn]/[In^-]). \quad (11)$$

Using $pH = -\log_{10}[H^+]$,

$$pH = pK_{In} + \log_{10}([In^-]/[HIn]); \quad (12)$$

pK_{In} is $-\log_{10} K_{In}$ and is called the indicator constant. The ratio $[In^-]/[HIn]$ determines the color of the indicator in the solution, and eqn (12), therefore, directly relates the color to the pH of the solution.

Eqn (12) shows that when $pH = pK_{In}$, the concentration of the indicator in each form is equal (i.e., $[HIn] = [In^-]$), and that at all pH values, both the acid (HIn) and basic (In^-) forms of the indicator are present in the solution. Our eyes are unable to detect the color of less than about 10% of one form of the indicator in the presence of the other, and the solution will appear to be the 'acid' color when $[In^-]/[HIn] < 1/10$, and the 'alkaline' color when $[In^-]/[HIn] > 10$. Therefore, the solution will be the color of the acid form of the indicator until $pH = pK_{In} - 1$, when the color will appear to change until it is that of the alkaline form of the indicator from $pH = pK_{In} + 1$. That is, the color changes over approximately 2 in terms of pH. In practice, the range for the color change is about 1.6–2 in terms of pH; this corresponds to a 40–100-fold change in the H^+ ion concentration. By choosing indicators with appropriate pK_{In} values, the whole of the pH range may be covered – a selection of some common pH indicators is given in Table 2.

Colorimetric Measurements of pH

Indicator papers For approximate determination of pH values, indicator papers may be used. Of particular value are 'nonbleeding' indicator papers. These contain dyestuffs strongly bound to the cellulose, so that the dye does not bleed into the test solution. These indicator papers are available, covering very narrow pH ranges, enabling moderately accurate pH values to be obtained very cheaply, conveniently, and quickly.

Comparison method This is carried out by adding similar quantities of an indicator to the test solution and a reference solution, and if the two solutions have the same color, they are assumed to have the same pH. The accuracy of the method, therefore, depends on the accuracy of the pH of the reference solution.

The indicator used must have a color at the pH of the test solution intermediate between the acidic and basic forms.

The approximate pH of the solution is first detemined using a 'universal,' or full-range, indicator, when an appropriate indicator for the measurement can be chosen. A quantity of the test solution is measured into a tube and a measured amount of indicator added. Tubes of buffer solutions covering the pH range of the indicator are treated in the same way. The colors of the tubes are viewed through the length of the tube against a white background. The accuracy of the method will depend on the differences of pH of the buffer solutions. These are typically 0.2, and the pH of the unknown solution can then be measured to within ± 0.1.

The Lovibond Comparator The need to use buffer solutions is eliminated by using a Lovibond 2000 comparator, in which the color of the test solution is compared with colored glasses. Standard colored glass rings are fitted into a disk that can be rotated and the colored glasses compared with the test solution. For accurate results, the sample solution and discs must be viewed against north daylight or in white light. Indicator discs for 14 different indicators are available. Special disks are available for the determination of the pH of blood.

With the Nesslerizer attachment, the solution is viewed through the depth of the liquid instead of through its thickness. The proportion of indicator therefore can be considerably reduced, resulting in a significant gain in accuracy, particularly when unbuffered or slightly buffered solutions are being tested.

Sources of error There are a number of possible sources of error when making colorimetric measurements.

1. *Slightly buffered solutions*. As indicators are weak acids or bases they will change the pH of the solution if it is unbuffered or only slightly buffered. Samples of relatively pure water, and solutions of a salt of a strong acid and strong base are the most frequently encountered solutions of this type. It is recommended that an appropriate pH meter is used for such solutions.
2. *Salt effect*. As discussed earlier, equilibrium constants depend on the activities of the components rather than concentrations, and it is only in very dilute solutions that activity equals concentration. Buffer solutions normally used have ionic strengths of up to 0.2 M. If the solution under test has a lower ionic strength than the buffer, the determined pH will be too small. The salt correction that must be applied is also affected by the types of ion present and the indicator used.
3. *Effect of colloids*. Colloid particles in the solution may preferentially absorb either the acid or the alkaline form of the indicator, giving totally erroneous results.
4. *Effect of proteins*. pH values obtained in the presence of proteins are unreliable. The effect of a protein on the determination depends on both the type of protein and the indicator used. Usually, the effect is greater when the protein is on the acid side of its isoelectric point.
5. *Effect of finely divided particles*. Chemical reactions may occur between the indicator and particles in the solution, giving erroneous results.
6. *Colored solutions*. Color comparison is not possible if the test solution is colored.
7. *Nonaqueous solvents*. Literature data on indicators apply to aqueous solutions; the indicator equilibria, hence the color, will be different in other solvents. The color of an indicator in, say, a water–alcohol mixture should be taken as only a rough guide to the true pH.

Table 2 Some pH indicators and their color changes and ranges

Indicator	*Color in acid solution*	*Color in alkaline solution*	*pH range*	pK_{In}
Thymol blue (acid)	Red	Yellow	1.2–2.8	1.7
Bromophenol blue	Yellow	Blue	2.8–4.6	4.0
Methyl orange	Red	Yellow	3.1–4.4	3.7
Methyl red	Red	Yellow	4.2–6.3	5.1
Litmus	Red	Blue	5.0–8.0	
Bromothymol blue	Yellow	Blue	6.0–7.6	7.0
Phenol red	Yellow	Red	6.8–8.4	7.9
Thymol blue (base)	Yellow	Blue	8.0–9.6	8.9
Phenolphthalein	Colorless	Red	8.3–10.0	9.6
Universal[a]	Red, orange, yellow	Green, blue, violet	3.0–11.0	
Full-range[a]	Red, orange, yellow	Green, blue, violet	1.0–14.0	

[a]These are mixtures of indicators; they give a continuous color change over a wide pH range.

Electrochemical Method

In order to measure pH by the electrochemical method, the emf of a reference cell is determined (see earlier). In this cell, the potential of one electrode (the hydrogen electrode) changes with pH, and the potential of the second, reference, electrode does not. Historically, the hydrogen gas electrode was used as the pH electrode, but the discovery by Cremer in 1906 that the potential difference across a glass membrane depends on the H^+ ion concentration on either side of it led to the development of the glass electrode. This is the electrode that is now always used for measuring pH.

Glass electrode In its simplest form, the glass electrode consists of a thin glass bulb containing a solution of constant pH, usually hydrochloric acid or a phosphate buffer with potassium chloride. In this solution is an electrode, usually silver coated with silver chloride.

The exchange of ions in the glass membrane for H^+ ions in the solutions on either side of it to form silanol ($\equiv$Si-OH) layers is the major factor in determining the pH response of the electrode, and special sodium and lithium glasses are used. As this surface layer is crucial to the correct functioning of a glass electrode, the pH-sensitive tip is stored in a pH 7 buffer solution, preferably containing $0.1\,\text{mol}\,\text{dm}^{-3}$ potassium chloride. As it is most important that the tip does not dry out, commercial electrodes are supplied with a storage teat containing the buffer plus potassium chloride solution.

Glass electrodes give a good linear response in the pH range 2–9, and up to pH 14 if the salt concentration in the test solution is not too high.

Combination electrodes Historically, the separate glass and reference electrodes were placed side by side in the test solution. Such electrodes are still available but are now used for special applications. Almost all pH measurements are made with combination electrodes in which the glass and reference electrodes are combined together. Although there are many details that vary from one electrode to another, the basic design is as shown in Figure 1. The glass membrane is at the bottom and is surrounded by the reference electrode. This is usually a silver/silver chloride electrode (silver wire coated with silver chloride in saturated potassium chloride solution), and the reference–test solution junction is at a porous glass frit near the bottom of the electrode. The potassium chloride in the reference cell slowly bleeds through the frit into solutions being tested and must be topped up, when necessary, through the filling hole near the top of the electrode.

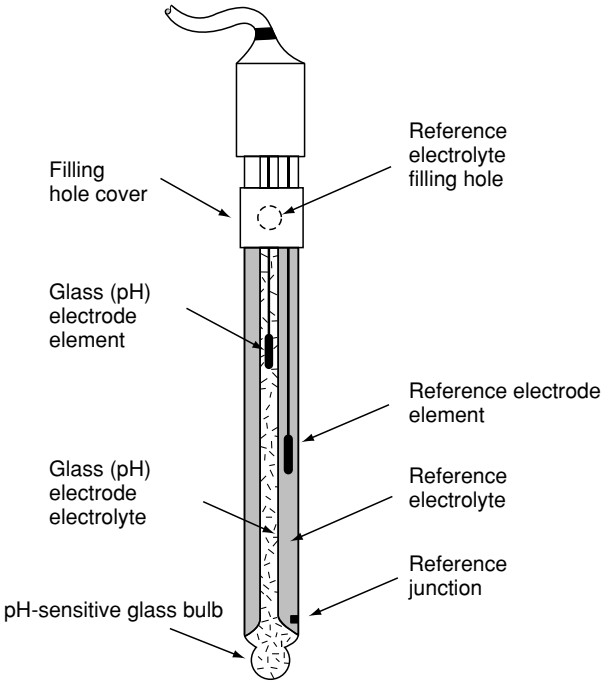

Figure 1 Combination pH electrode. Reproduced from pH: Principles and Measurement, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

For many purposes, this electrode is perfectly satisfactory, but under adverse conditions, the positive flow of reference electrolyte is reversed, and the sample solution contaminates the reference cell. This can be avoided by using a double-junction reference system in which the reference cell is connected through a glass frit with an intermediate electrolyte, which is in turn in contact, through a glass frit, with the test solution. Such an electrode also eliminates clogging owing to precipitation of silver chloride at the glass frit. Also, the intermediate electrolyte can be changed if that being used reacts with the test solution at the glass frit junction.

Special glass electrodes

Amplified electrode The major experimental difficulty in measuring the emf of a pH cell is the high electrical resistance of the glass membrane, commonly 10–500 MΩ. This requires a high-impedance millivoltmeter and short, screened connections to the electrode. One solution to this problem is a glass electrode with a built-in amplifier (Hanna Instruments), where the high-impedance circuitry is in an integrated circuit encapsulated at the top of the electrode together with a mercury battery. This electrode is ideal for industrial pH measuring and monitoring

as long, unshielded cables can be used from the electrode to the pH meter.

Other electrodes pH glass and combination electrodes of many types are available; these include electrodes in plastic bodies (for durability), microelectrodes for small volumes of test solution, electrodes resistant to proteins, and flat electrodes for pH measurement at surfaces.

pH meters A pH meter is a high-impedance millivoltmeter that is designed to convert millivolts to pH. There are many types available at a range of prices. Temperature compensation is important, and many meters have a temperature probe facility for this purpose. Many meters contain microprocessors and are very easy to use.

To make a pH measurement, it is necessary to calibrate the meter first. This is done by using two buffer solutions that, ideally, span the pH of the test solution. After rinsing the electrode, this is placed in the test solution and the pH read from the meter.

Accurate pH measurements can be made without difficulty on test solutions containing dissolved acids, bases or salts. If, however, the sample is virtually pure water (e.g., tap, rain, or boiler-feed water) pH measurements are unreliable. This is because (1) the glass electrode requires a long time to stabilize in a low-ionic-strength solution; (2) the solution has a low buffering capacity and is therefore susceptible to drift as atmospheric carbon dioxide is absorbed; (3) the low conductivity of the solution allows the pick-up of electromagnetic noise; and (4) there is electrical noise created at the reference junction. Special low-conductivity electrode buffer kits have been developed (Russell pH Ltd) for such measurements. In these kits, the glass electrode is of extremely low resistance, and the calibration errors are reduced by using a buffer of low ionic strength.

Very small, portable pH meters are now available. A good example is the Piccolo (Hanna Instruments). This meter weighs only 100 g and uses an interchangeable amplified electrode in a plastic container. It has a built-in temperature sensor that compensates for temperatures from 0 to 70°C, and it measures pH in the range 1–13 with an accuracy and resolution of ± 0.01.

See also: **Canning**: Principles; **Colloids and Emulsions**; **Spectroscopy**: Visible Spectroscopy and Colorimetry; **Spoilage**: Chemical and Enzymatic Spoilage; Molds in Spoilage; Yeasts in Spoilage

Further Reading

Anonymous (1987) *pH Values*, 8th edn. Poole, UK: BDH.

Bates RG (1973) *Determination of pH. Theory and Practice*, 2nd edn. New York: John Wiley.

Covington A (1989) *pH and its Measurement*. London: Royal Society of Chemistry (a chemistry cassette and workbook).

Galster H (1991) *pH Measurement (Fundamentals, Methods, Applications, Instrumentation)* New York: Wiley-VCH.

PHENOLIC COMPOUNDS

M Murkovic, Graz University of Technology, Graz, Austria

Background

Polyphenols play an important role in plants as well as in foods. The main uses of polyphenols in foods are as colorants and antioxidants. A large amount of research on polyphenols has focused on their antioxidant properties, since they are thought to have positive effects on chronic degenerative diseases (cataracts, age-related macular degeneration, central neurodegenerative diseases, and diabetes mellitus), cardiovascular disease, and cancer. Polyphenols are also used as antioxidants in the food industry to increase shelf-life, which can be limited due to deterioration because of oxidation reactions, especially fat oxidation. Not only is oxidized food a problem because of the resultant rancid aroma. It can also have implications on human health, since increased exposure to free radicals can also increase the risk of these degenerative diseases.

The several thousand polyphenols that have been described in plants can be grouped in several classes. Distinction between these classes is drawn first on the basis of the number of constitutive carbon atoms and then in the light of the structure of the basic skeleton. In addition, besides simple soluble forms mainly found in vacuoles, there are also polymerized forms of varying solubility like the tannins or the completely insoluble lignins. The groups of polyphenols dealt with in this chapter comprise anthocyanins,

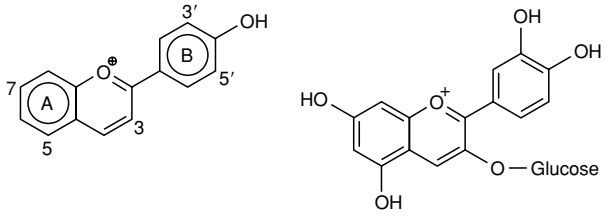

Figure 1 Structure and possible positions of hydroxylation and methoxylation of anthocyanidins (cyanidin-3-*O*-glucoside).

flavan-3-ols, isoflavones, flavones and flavonols, hydroxybenzoic acids, and hydroxycinnamic acids.

Anthocyanins

Anthocyanins are the most important group of water-soluble plant pigments visible to the human eye. With a few exceptions, e.g., betalains, they are universal plant colorants and largely responsible for the cyanic colors of flower petals and fruits. They may also occur in roots, stems, leaves, and bracts, accumulating in the vacuoles of epidermal or subepidermal cells.

The main positions of hydroxylation are 3, 5, and 7 in the A ring and 3′ and 5′ in the B ring (see Figure 1). Methoxyl groups can be attached to 3′ and 5′. In some cases, the 3-hydroxyl group is removed, and a few structures are known that have a hydroxyl group attached to C-6 of the A ring. The anthocyanidins that are commonly found in fruits are pelargonidin, cyanidin, delphinidin, peonidin, petunidin, and malvidin. These are all hydroxylated at the positions 3, 5, and 7. Normally, anthocyanins occur as 3-monosides, 3-biosides, and 3-triosides as well as 3,5-diglycosides and, in some cases, 3,7-diglycosides. The carbohydrates that are attached to the positions 3, 5, and 7 comprise glucose, galactose, rhamnose, arabinose, and xylose.

An example of a glycoside is cyanidin 3-rutinosid (6-*O*-α-L-rhamnosyl-D-glucose), which is very common in fruits, most commonly associated with cyanidin 3-glucoside. Other glycosides are cyanidin 3-sambubioside (2-*O*-β-D-xylosyl-D-glucose), which is present in elderberry, redcurrant, and black raspberry, and cyanidin 3-sophoroside (3-*O*-β-D-glucosyl-D-glucose), the main pigment in raspberries, loganberries, boysenberries, and mangosteen. The latter also occurs in redcurrants and sour cherries, but as a minor component (see Table 1).

The color-active form of the anthocyanins shows an absorption maximum at about 520 nm. This is due to the flavylium cation structure, which occurs only in an acidic environment. An increase in pH above 4 leads to a substantial fading of color. Other color changes occur during oxidation of the anthocyanins, leading to the browning of foods.

Table 1 Anthocyanidin content of selected vegetables and fruits (mg per 100 g fresh weight)

Food	*Total anthocyanidin content*
Fruits	
Blueberry	25–500
Elderberry	300–500
Strawberry	30–50
Cherry, sweet	350–450
Grapes, red	30–750
Vegetables	
Cabbage, red	20–40
Onion, red	10–20

Flavan-3-ols

Flavan-3-ols are components in the structure of proanthocyanidins (condensed tannins) as their monomers. Flavan-3-ols have been obtained from several tree species (*Quercus*, *Castandea*, *Acacia*, *Eucalyptus*, etc.). Four main flavan-3-ols are found in fruits. Two are orthodiphenols hydroxylated in the 3′ and 4′ positions in the B-ring ([+]-catechin and [−]-epicatechin), and two are trihydroxylated in the 3′-, 4′-, and 5′-positions ([+]-gallocatechin and [−]-epigallocatechin). C-2 and C-3 represent two centers of asymmetry in the molecule in such a way that the four mentioned flavan-3-ols in fruits are grouped in two pairs of diastereomers whose absolute configurations are as follows: 2R in all cases and either 3S for (+)-catechin and (+)-gallocatechin or 3R for (−)-epicatechin and (−)-epigallocatechin (see Figure 2).

Unlike in other classes of flavonoids, monomeric flavan-3-ols are generally found in free rather than glycosylated or esterified form in fruits (see Table 2). One exception is the case of the (−)-epicatechin-3-*O*-gallate identified in grapes. This ester has been found in several red grape cultivars.

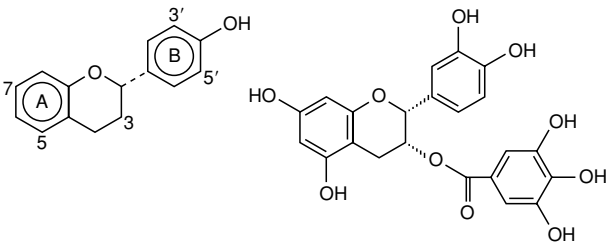

Figure 2 Structure and possible positions of hydroxylation and methoxylation of flavans; ((−)-epicatechin 3-gallate).

Table 2 Content of flavanol monomers in selected vegetables and fruits (mg per 100 g fresh weight)

Food	*Total flavan-3-ol content*
Apricot	25
Apple	8
Strawberry	4
Cherry, sweet	22
Grapes (skin)	19
Barley	3–4
Tea	18–35[a]

[a]Flavanols in g per 100 g of dry leaves.

Isoflavones

The isoflavones are a distinctive but large subgroup of flavonoids. These compounds possess a 3-phenylchroman skeleton that is biogenetically derived by rearrangement of the flavonoid 2-phenylchroman system (see Figure 3).

Despite the restricted distribution of the isoflavonoids in the plant kingdom, the structural variation is surprisingly large. This arises not only from the number and complexity of substituents on the basic 3-phenylchroman system but also from the different oxidation levels in this skeleton and the presence of extra heterocyclic rings. Within the flavonoids, the isoflavones are the largest group of naturally occurring structures.

The enzyme that is responsible for the transformation of the flavones to the isoflavones is named isoflavone synthase. This enzyme catalyzes the isomerization of the two flavones (2S)-naringenin or (2S)-liquiritigenin to the isoflavones genistein and daidzein, respectively.

Soy foods have received considerable attention for their role in disease prevention, especially in relation to heart disease, osteoporosis, and cancer. The research interest in soy focuses on establishing the physiological effects of isoflavones. The occurrence of the isoflavonoids is restricted to only a few plants within the family of the leguminosae and some nonlegume dicotyledons. Soybeans and soyfoods are practically the only nutritionally relevant dietary sources of isoflavones. Isoflavones are weak estrogens in that they bind to estrogen receptors, but they also have important nonhormonal properties as well. Initial speculation that soyfoods, and in particular isoflavones, might promote bone health was based on the estrogenic properties of isoflavones. In ovariectomized rodents, isoflavones retard bone loss almost as effectively as estrogen. Most soyfoods are rich in isoflavones and favorably affect bone turnover and spinal bone mineral density in perimenopausal and postmenopausal women (see Table 3).

Table 3 Content of isoflavones in soy products (mg per 100 g (ml) fresh weight)

Food	*Total isoflavone content*
Soya bean	60–400
Tofu	8–70
Soy flour	80–180
Soya milk	3–180
Soy sauce	1–8

Flavones and Flavonols

Flavones are flavonoids characterized by a nonsaturated 3-C chain and have a double bond between C-2 and C-3, like flavonols, with which they differ by the absence of hydroxyl in the 3-position. It appears that this simple difference in structure between flavones and flavonols has very important consequences in the biogenesis, physiological, and pharmacological roles, and the phylogenetic and chemotaxonomic signification of these compounds. Flavones are widely distributed among the higher plants in the form of aglycones or glycosides (see Figure 4).

The edible portion of some foodstuffs has an unusually high concentration of quercetin. The concentrations of quercetin analyzed in hydrolyzed samples from fruits and vegetables are 284–486 mg kg^{-1} in onions, 110 mg kg^{-1} in kale, 32–45 mg kg^{-1} in French beans, 30 mg kg^{-1} in broccoli, 14 mg kg^{-1} in lettuce, and 8 mg kg^{-1} in tomatoes. The average concentration of quercetin in fruits is 15 mg kg^{-1}, with apples having the highest concentration (21–72 mg kg^{-1}) (see Table 4). Parsley and thyme are also major food sources of flavones.

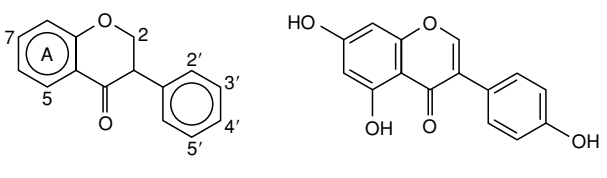

Figure 3 Structure and possible positions of substitutions of isoflavone (genistein).

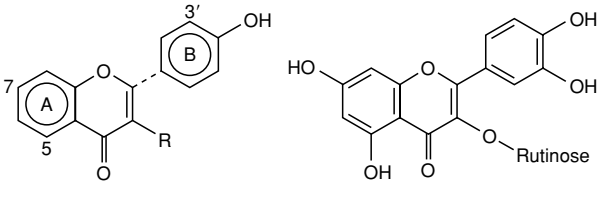

Figure 4 Structure and possible positions of substitutions of flavones (R = OH in flavonols).

Table 4 Content of flavonols and flavones in selected vegetables and fruits (mg per 100 g fresh weight)

Food	*Total flavonol content*
Apple	2–4
Apricot	3
Broccoli	9–11
Kale	32–60
Onion	35
Celery, leaf	95
Sweet pepper, red	1

Hydroxybenzoic Acids

Hydroxybenzoic acids have a general structure of the C6–C1 type derived directly from benzoic acid. Variations in structure lie in the hydroxylations and methoxylations of the aromatic cycle. Four acids dominate the family of these compounds: *p*-hydroxybenzoic acid, vanillic, syringic, and protocatechuic acids. The first three are constituents of lignin, from which they are released by alkaline hydrolysis. Gentisic acid has a particular pattern of hydroxylation, which can be related to that of salicylic acid, but it is much less common. Gallic acid is another well-known plant acid which participates in the formation of hydrolyzable gallotannins. Its dimeric condensation product (hexahydrodiphenic acid) and the related dilactone, ellagic acid, are also common plant constituents.

Hydroxybenzoic acids are commonly present in bound form (see Figure 5). They are the component of a complex structure like lignins and hydrolyzable tannins. The content of hydroxybenzoic acids in foods of plant origin is generally low (see Table 5), but in some berries and vegetables such as onions and horse-radish, the content of hydroxybenzoic acids may be very high.

Figure 5 Structure and possible positions of hydroxylation of hydroxybenzoic acids (gallic acid).

Table 5 Content of hydroxybenzoic acids in selected fruits (mg per 100 g fresh weight)

Food	*Total hydroxybenzoic acid content*
Blackberry	8–27
Blackcurrant	4–12
Raspberry	6–10
Strawberry	2–8

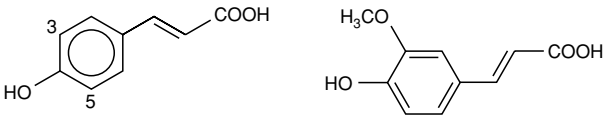

Figure 6 Structure and possible positions of hydroxylation of hydroxybenzoic acids (ferulic acid).

Hydroxycinnamic Acids

Among the fruit phenolics, hydroxycinnamic acid derivatives play an important role because of both their abundance and diversity. They all derive from cinnamic acid and are essentially present as combined forms of the four basic molecules *p*-coumaric, caffeic, ferulic (see Figure 6), and sinapic acids. The free forms of these acids are very rare in fruits. Two main types of soluble derivatives have been identified: first, those involving an ester bond between the carboxylic function of phenolic acid and one of the alcoholic groups of an organic compound, for example chlorogenic acid, which has been identified in numerous fruits; second, those involving a glycosidic bond with one of the phenolic groups of the molecule (e.g., *p*-coumaric acid *O*-glucoside). The diversity of the hydroxycinnamic acids encountered in plants and particularly in fruits thus results from the nature of the bonds and that of the molecules involved.

Hydroxycinnamic acids occur most commonly in foods of plant origin. Of these, caffeic acid is the predominant hydroxycinnamic acid in many fruits, constituting over 75% of total hydroxycinnamic acids found in plums, apples, apricots, blueberries, and tomatoes. However, *p*-coumaric acid is the dominant hydroxycinnamic acid of citrus fruits and pineapple. Hydroxycinnamic acids are widely present in the bound form and are rarely found in free form. Processing of fruits and vegetables (freezing, sterilization, and fermentation) contributes to the formation of free hydroxycinnamic acids in such products (see Table 6). Chlorogenic acid is found in many foods, including apples, apricots, berries, peaches, pears, plums, avocados, and carrots. Chlorogenic acid is the key substrate for enzymatic browning, particularly in apples and pears. Its content is reduced by 70% during browning of the pear skins.

Analysis of polyphenols

Sample Preparation

Prior to extraction, solid food samples have to be homogenized or crushed using a mortar and pestle. Soluble phenolic compounds are generally extracted by means of alcohol–water mixtures. Extraction is performed using a mixture of ethanol (80%) and

Table 6 Content of hydroxycinnamic acid derivatives in selected fruits (mg per 100 g (ml) fresh weight)

Food	*Chlorogenic acid isomers*
Apple juice	6–20
Beer	0.2–2
Coffee beans, green	6–9 g per 100 g
Coffee beans, roasted	0.2–3 g per 100 g
	Hydroxycinnamic acids
Strawberries	1–3
Blackcurrants	5–7
Gooseberries	2–7
Raspberries	2–3
Redcurrants	1–2

water (20%) of the freeze-dried powder of the food to be analyzed. Methanol can be used instead of ethanol, and extraction with methanol gives good results, especially in the extraction of soluble phenolic compounds from fruits. Oxidation is avoided by working at low temperatures and, if necessary, by the addition of antioxidants such as ascorbic acid or butylated hydroxytoluene. To avoid enzymatic oxidation from polyphenol oxidases, samples may need to be heated to a temperature of more than 90 °C for a few minutes. These enzymes catalyze the oxidation of phenols to quinones with subsequent nonenzymatic rapid polymerization. These oxidases can also be inhibited by lowering the pH to below 4.0. Extraction of anthocyanins is normally carried out under cold conditions with methanol containing 1% of hydrochloric acid to obtain the flavylium cation, which is stable under acidic conditions.

For analysis of liquid samples like fruit juices, only dilution, filtration, and, in some cases, clarification are necessary. Carotenoids that can disturb the analysis should be extracted using nonpolar solvents. To remove the sugar moieties from the polyphenols, the glycosides have to be hydrolyzed. The cleaving off be carried out in acidic, basic, or enzymatic conditions. The hydrolysis of anthocyanins to anthocyanidins is often necessary owing to the difficulties of obtaining anthocyanin standards. This can be done by refluxing the dry anthocyanins in methanol acidified with 2 M HCl. Alkaline hydrolysis cleaves the acylated portions of anthocyanins using 2 M NaOH. During hydrolysis, care has to be taken that no polyphenols are destructed or rearrangement takes place so that artifacts are analyzed.

The water–alcohol extract obtained is a raw extract containing numerous nonphenolic substances (carbohydrates, organic acids, proteins, pigments, etc.). Before these crude extracts can be analyzed by high-performance liquid chromatography (HPLC), they should be purified using solid-phase extraction methods. Chlorophylls and carotenoids can be removed by applying conventional depigmentation techniques with petroleum ether.

HPLC Analysis

The columns used for HPLC analysis in the literature are almost exclusively reversed phase. Elution systems are usually binary, with an aqueous acidified polar solvent such as aqueous acetic acid, phosphoric acid, or formic acid and a less polar organic solvent such as methanol or acetonitril. Isocratic elution is mainly used for routine analysis if only a few polyphenols are to be determined.

For detection and quantification after separation, the absorption of ultraviolet or visible light is normally used. For flavonoids, two absorption bands are characteristic: the first in the range of 240–285 nm, which is believed to arise from the A-ring, and the second with a maximum in the range of 300–550 nm, which arises from the absorption of the B-ring (see Table 7). Since the absorption spectrum depends on the chemical substitution and the conjugated system, the different groups of flavonoids (flavanons/isoflavanons, catechins, flavons/flavonols, etc.) have distinct absorption bands. The anthocyanins show absorption above 500 nm, whereas the catechins have only a small maximum at 280 and no absorption at higher wavelengths.

The combination of liquid chromatography with mass spectroscopy for the detection of polyphenols has been increasing in recent years. Mass-selective detection is especially useful in identifying new structures and verifying known substances. With a quadrupole mainly the unfragmented molecule can be detected. Using tandem mass spectrometry, the molecules can be fragmented in a selective way. Thus, information on the chemical structure of the polyphenols can be obtained.

Intake and Bioavailability of Polyphenols

In the Western diet, the phenolic acids, catechins, proanthocyanidins, and anthocyanins dominate uptake. Especially for epidemiologic studies, it is necessary to estimate accurately the amount of

Table 7 Detections wavelengths of flavonoids

Structure	*Wavelength used for analysis (nm)*
Flavons, flavonols, flavonol glycosides	270
Anthocyanins	500–525
Isoflavones	236, 260, 280
Flavanones and glycosides	280, 290
Catechins	210, 280

polyphenols consumed. In general, fruits are richer in polyphenols compared to vegetables. Beverages contribute significantly to the uptake of polyphenols. Red wine, coffee, and black tea contain high amounts of polyphenols. Green tea, which also contains large amounts of these substances, is consumed in East Asia in high amounts. The uptake of polyphenols is estimated to be 50–200 mg per meal. The phenolic acids are constituents of coffee, fruits, vegetables, and cereals. Flavanols, which occur predominantly in onions, apples, and other fruits and vegetables, contribute about 20 mg per day. Catechins are provided by tea and various fruits and proanthocyanidines by various fruits, legume seeds, chocolate, and red wine. Flavanones are found in citrus fruits. The total uptake of flavonoids in the Western diet is estimated to be about 1000 mg per day. Examples of the distribution and quantity of flavonoids consumed daily include: 44 mg from cereals; 79 mg from potatoes, bulbs, and roots; 45 mg from peanuts and nuts; and 162 mg from vegetables and herbs. The largest portion of flavonoid intake comes from cocoa, cola, coffee, tea, beer and wine (420 mg per day), with an additional 290 mg per day from fruits and juices.

The uptake of 100 mg of polyphenols results in a concentration of about 300 μM in the gut. Normally, the plasma levels will not exceed 1 μM. Urinary excretion varies from less than 1% to about 25% depending on their gut absorption, enterohepatic cycling, and metabolism. In particular, the type of polyphenol and the glycosylation influence bioavailability. The uptake of proanthocyanidines is limited to monomers and dimers. A higher degree of polymerization results in a sharp decrease in bioavailability. The metabolism of polyphenols in the small intestine is mainly via glucuronidation, sulfatation, and methylation. The colon microflora contributes to the metabolism of the polyphenols. This has been investigated in detail for several types of polyphenols, and especially for quercetin. Most flavonoids enter the diet as glycosides, with quercetin and rutin being the most common flavonoid glycosides consumed. The hydrophilic nature and their relatively high molecular weight preclude absorption in the small intestine. Furthermore, flavonoid β-glycosides resist intestinal hydrolases, and consequently, flavonoid glycosides can pass unaltered into the large intestine. The resident microflora of the bowel produces glycosidases capable of releasing the aglycone from its sugar. In addition, the resident microflora can cleave the pyrone ring (ring C), producing phenyl acetic and phenyl propionic acids and other derivatives. But since glycosidase activity proceeds at a faster rate than ring cleavage, the intact flavonoid aglycone can persist in the large intestine with a clear potential for absorption.

Polyphenols as Antioxidants

It is well accepted that free-radical-mediated processes can lead to chronic degenerating diseases. The oxidative damage can be found on molecular level as oxidized lipids, proteins, or DNA. These can be detected, especially in patients with atherosclerosis, certain cancers, neurodegenerative diseases, and lung disorders. Many types of reactive oxygen species (ROS) and reactive nitrogen species (RNS) have been shown to induce a certain type of damage that is associated with disease development. These include the superoxide radical, hydrogen peroxide, hydroxyl radical, lipid alkoxyl and peroxyl radicals. Protection against damage induced by ROS/RNS can be achieved in several different ways:

- suppression of free-radical formation by antioxidants;
- scavenging of radicals by antioxidants to inhibit initiation of chain reactions and slow down chain propagation;
- repair mechanisms;
- transition metal chelation.

The polyphenols have structural features that characterize their antioxidative potential. This is the presence of hydrogen-donating substituents and the ability to delocalize the resulting free electron. Owing to the delocalization of the electron, the resulting radical does not have sufficient energy for further radical reactions. The antioxidative properties of flavonoids and other phenolics from natural sources are assessed by determining their activity as scavengers of radicals that arise from lipid oxidation or other biological processes. For evaluation of the antioxidant properties of flavonoids, the chemical structure can be related to the activity. Three structural groups are important determinants for radical scavenging: (1) the catechol structure (3′, 4′-dihydroxy) in the B ring, which is the obvious radical target site; (2) the 2, 3 double bond in conjugation to the 4-keto function, which is responsible for electron delocalization; and (3) the presence of both hydroxyl groups (the 3-OH and the 5-OH) for maximal radical-scavenging potentials. If all of these structural features are present, a maximum radical-scavenging potential would be expected.

Depending on the method of measuring the antioxidative potential, relative reactivities of structurally related polyphenols change. This can be seen especially in low-density-lipoprotein (LDL) oxidation experiments, in which quercetin and luteolin are strong antioxidants, that protect LDL with high efficiency, whereas kaempferol does not. In particular, if chelation of metal ions is the protecting mechanism, quercetin is a very strong antioxidant.

A simple chemical assay broadly applicable in determining the hierarchy of antioxidant activities of polyphenols measures their ability to reduce the ABTS-radical cation (2,2′-azinobis (3-ethylbenzothiazoline-6-sulfonate)). With this method, the relative ability of antioxidant or radical-scavenging substances to scavenge the ABTS cation radical in the aqueous phase is compared to a standard amount of the synthetic antioxidant Trolox, the vitamin E analog. Incubation of $ABTS^{o+}$ with peroxidase (metmyoglobin) and hydrogen peroxide results in the production of the radical cation ($ABTS^{o+}$). This species is blue–green in color and can be detected at 734 nm. Antioxidants or radical scavengers in the added sample suppress this color production to a degree that is proportional to their concentration. In this method, the activity of tested substances is expressed as an equivalent of the millimolar concentration of a Trolox solution. Table 8 shows the antioxidant activities of a range of flavonoids in comparison with the reduction potentials of the flavonoid radicals. It can be seen that the measured reducing properties of the phenolics, in terms of their antioxidant activities in this particular assay system, are consistent with the reduction potentials of their radicals.

The range of reduction potentials of selected flavonoid radicals, relative to those of ascorbic acid and α-tocopherol, is shown in Table 8, from the more oxidizable quercetin to kaempferol. These values reflect the ability of hydrogen-donating antioxidants to scavenge the $ABTS^{o+}$ radical cation, absorbing in the near-infrared region (650–820 nm) compared with that of ascorbic acid (E_7) and Trolox equivalent antioxidant activity (TEAC). Antioxidants reduce the absorbance to an extent dependent on the antioxidant activity. Spectral studies have revealed that the radical site is on the B ring and that the A ring has no influence on the properties of the radicals from the B ring in flavonoids with a saturated C ring, such as catechin. In flavone radicals in which the B ring with a 3′,4′-dihydroxycatechol structure is conjugated through the C ring 2, 3-double bond, evidence suggests that the radical is on the catechol B ring and that spectra resemble those of 3,4-dihydroxycinnamate radicals. The antioxidant activity of flavonoids depends critically on the part of the molecule with more efficient electron-donating properties. In most flavonoids, this is the B ring. Thus, the antioxidant activity of flavonoids as electron or hydrogen donors relates to the reduction potentials and reactivities of the substituent hydroxyl groups. A number of chemical studies on the reactivity of flavonoids with a range of radicals and the stability of the resultant antioxidant radicals have also emphasized specifically the role of the catechol structure in the B ring and unsaturation in the C ring.

Table 8 Reduction potentials (E_7) of flavonoid radicals and the TEAC

	E_7	TEAC
Quercetin	0.33	4.7
Epigallocatechin gallate	0.43	4.8
Epigallocatechin	0.42	3.8
Taxifolin	0.5	1.9
Catechin	0.57	2.4
Rutin	0.6	2.4
Kaempferol	0.75	1.3

Conclusion

Substantial epidemiological evidence links high antioxidant status in human populations with a low risk of degenerative disease. The consumption of large amounts of fruit and vegetables has been shown unequivocally in a considerable number of studies to be associated with a lowered risk of several kinds of cancer in a number of different body sites. It remains to be proven with certainty whether this effect is due to the content in the fruits and vegetables of antioxidant substances, which are the protective agents, but substantial circumstantial evidence points to the likelihood that antioxidant nutrient and nonnutrient substances in fruit and vegetables are the most important anticarcinogenic factors in these foods. Besides the classical antioxidants vitamin C, vitamin E, and β-carotene, many phenolic substances are also present in foods, including the flavonoids, which may also contribute to the total antioxidant potential of the diet and thus may lower the risk of cancer. Similar considerations apply to arteriosclerosis, with its cardiovascular and cerebrovascular risk connotations; it has been clearly shown that antioxidant nutrients can lower the oxidation of low-density lipoproteins in the vascular wall, which is thought to be a major early factor in the development of arteriosclerosis. It is also possible that antioxidants can play an important role in delaying the thrombotic events associated with the onset of heart attack or stroke, and there is thus a possibility that antioxidants have a dual effect in lowering the risk of vascular accidents. If it is proven that antioxidants can lower the risk of these degenerative diseases, it is vital that any enhancement of the diet with antioxidants is safe and free from the risk of any undesirable side-effects.

See also: **Antioxidants**: Natural Antioxidants; Role of Antioxidant Nutrients in Defense Systems; **Chromatography**: High-performance Liquid Chromatography; **Soy (Soya) Beans**: Properties and Analysis; **Soy (Soya) Cheeses**; **Soy (Soya) Milk**; **Soy (Soya) Bean Oil**

Further Reading

Eisenbrand G, Dayan AD, Elias PS, Grunow W and Schlatter J (2000) *Carcinogenic and Anticarcinogenic Factors in Food*. Weinheim: Wiley-VCH.

Harborne JB, Mabry TJ and Mabry H (eds) (1975) *The Flavonoids*. London: Chapman & Hall.

Harborne JB and Mabry TJ (1982) *The Flavonoids: Advances in Research*. London: Chapman & Hall.

Harborne JB (ed.) (1994) *The Flavonoids: Advances in Research Since 1986*. London: Chapman & Hall.

Lindsay DG and Clifford MN (eds) (2000) Nutritional enhancement of plant-based food in European trade (NEODIET). *Journal of the Science of Food and Agriculture* 80.

Macheix J-J, Fleuriet A and Billot J (1990) *Fruit Phenolics*. Boca Raton, FL: CRC Press.

Packer L and Cadenas E (eds) (1996) *Handbook of Antioxidants*. New York: Marcel Dekker.

Packer L, Hiramatsu M and Yoshikawa T (eds) (1999) *Antioxidant Food Supplements in Human Health*. San Diego, CA: Academic Press.

Rice-Evans C and Packer L (eds) (1998) *Flavonoids in Health and Disease*. New York: Marcel Dekker.

Shahidi F and Naczk M (1995) *Food Phenolics: Sources, Chemistry, Effects, Applications*. Lancaster, PA: Technomic.

Shahidi F and Ho C-T (eds) (2000) *Phytochemicals and Phytopharmaceuticals*. Champaign, IL: AOCS Press.

Shahidi F and Wanasundara PKJ (1992) Phenolic antioxidants. *Critical Reviews in Food Science and Nutrition* 32: 67–103.

PHOSPHOLIPIDS

Contents

Properties and Occurrence

B F Szuhaj, Central Soya Company Inc., Fort Wayne, IN, USA

Introduction

Phospholipids have been scientifically studied since the 1700s and became commercially available as lecithin in the 1930s. Their primary commercial source today is the soya bean, but phospholipids can be found in all living cells as part of the cellular membranes. This article covers the properties and occurrence of phospholipids as commercial lecithins, their chemistry, manufacture, composition, specifications, and their potential use in food systems.

Occurrence

The International Lecithin and Phospholipid Society defines lecithin as 'a complex mixture of glycerophospholipids obtained from animal, vegetable or microbial sources, containing varying amounts of substances such as triglycerides, fatty acids, glycolipids, sterols, and sphingophospholipids.' Lecithins are natural surfactants primarily derived from soya beans and eggs. They are found most abundantly in seeds and nuts, eggs, brains, and cell walls, in a concentration range of 0.5–2%. (*See* **Eggs**: Structure and Composition; **Fatty Acids**: Properties; **Soy (Soya) Beans**: Properties and Analysis; **Triglycerides**: Structures and Properties.)

Properties of Phospholipids (Lecithins)

There are three types of properties necessary to define phospholipids and lecithins: (1) chemical, (2) physical, and (3) functional.

Chemical Properties

The chemical composition of deoiled and liquid soya bean lecithin is shown in **Table 1**. There are approximately 17 different compounds in commercial lecithin, including carbohydrates, phytosterols, and minor phytoglycolipids. The three major phospholipids are phosphatidylcholine, phosphatidylethanolamine, and phosphatidylinositol.

Table 1 Chemical composition of soya bean lecithin

	Granular lecithin	*Typical liquid lecithin*
Phosphatides (acetone-insolubles) (%)	95 (minimum)	60 (minimum)
Soya bean oil (%)	2–3	39
Moisture (%)	1	0.7
Fat (g per 100 g product)	90	93
Monounsaturated (oleic acid) (%)	9.2	17.9
Polyunsaturated (linoleic, linolenic acids) (%)	65.9	60.7
Saturated (palmitic, stearic acids) (%)	24.9	20.3
Carbohydrates (g per 100 g of product)	8	5
Approximate composition (100-g sample)		
Fatty acid content (g)	50	66
Fatty acid content (relative composition) (%)		
Linoleic	58.9	54.0
Linolenic	7.0	6.7
Oleic	9.2	17.9
Palmitic	20.3	15.6
Stearic	4.6	4.7
Other fatty acids	0.0	1.1
Total	100.0	100.0
Primary acetone insolubles (g)		
Phosphatidylcholine	23	15
Phosphatidylethanolamine	20	12
Phosphatidylinositol	14	9
Elemental analysis (mg)		
Calcium	65	40
Iron	2	1
Magnesium	90	60
Phosphorus	3000	2000
Potassium	800	440
Sodium	30	10

Reproduced from Central Soya, Chemurgy Division (1989) *The Lecithin Book*. Fort Wayne: Central Soya.

Structure of the Major Phospholipids

The chemical backbone of the major phospholipids is a diacylglycerol molecule with the third carbon attached to a phosphate molecule. Choline, ethanolamine, serine, and inositol can be attached to the phosphate group to change the physical and functional properties, leading to the formation of phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine, and phosphatidylinositol, respectively. The groups attached to positions 1 and 2 (α or β) are C_{14}–C_{18} fatty acids with double bonds associated with the lecithin source. The second carbon of the glycerol molecule, the β-position, usually contains linoleic acid.

Physical Properties

There are two major physical classes of lecithins: (1) fluid, and (2) waxy solids.

The fluid lecithins can have viscosities from 5000 to 100 000 cP, depending on processing conditions and diluents. The low-viscosity products are made through the addition of fatty acids and vegetable oil, depending on the function and stability required. Divalent metal ions like calcium can be added during drying to decrease viscosity. The moisture content can also make a difference. Water levels above 1% will increase the viscosity, eventually to a plastic state.

Deoiled lecithins are waxy solids that can be ground to various particle sizes. They are stable, free-flowing granules or powders.

Functional Properties

Lecithins are multifunctional agents. They can be used for many purposes in a food system, as shown in Table 2. The most popular functionalities are discussed below.

Antidusting agents Lecithins reduce static electricity by wetting dusty particles. They can be used alone or in conjunction with vegetable oils. Oils can be selected for the degree of shelf-life required.

Crystal formation modifier Lecithins retard nucleation in fats and even monoglycerides, reducing graininess in texture.

Emulsifiers Lecithins are most often used as amphoteric emulsifiers. They promote stable formation of oil-in-water and water-in-oil emulsions by reducing

Table 2 Functionality of lecithins

Adhesion aid
Antibleed agent (as in fat bloom)
Anticorrosive
Antidusting agent
Antioxidant
Antispatter agent
Biodegradable additive
Biologically active agent
Catalyst
Color intensifier
Conditioning agent
Coupling agent
Dispersing agent, mixing aid
Emollient, softening agent
Emulsifier or surfactant
Flocculant
Grinding aid
Lubricant
Liposomal encapsulating agent
Machining aid
Modifier
Moisturizer
Nutritional supplement, vitamin source
Penetrating agent
Plasticizer
Promoter
Release agent, antisticking agent
Spreading agent
Stabilizer
Strengthening agent
Suspending agent
Synergist
Viscosity modifier
Water repellent
Wetting agent

Reproduced from Schmidt JC and Orthoefer FT (1985) In Szuhaj BF and List GR (eds) *Lecithins*, p. 187. Champaign: American Oil Chemists' Society, with permission.

the interfacial surface tension between immiscible liquids. (*See* **Emulsifiers:** Organic Emulsifiers.)

Mixing and blending aids Lecithins decrease time and increase efficiency of mixing of unlike ingredients such as sugar and shortening by providing lubricity as well as viscosity reduction at the contact surfaces of the incompatible solids.

Release agents Lecithins provide easy release from metallic surfaces by attaching to the metal surface during hot or cold cooking. They assist in the cleaning of hot surfaces where proteins or batters are applied. They also reduce sticking between frozen food products.

Separating agents Lecithins prevent the adhesion of products that normally stick when in contact, like cheese slices and caramel confectionery.

Viscosity modifiers Lecithins reduce viscosity by coating particles to reduce particle–matrix friction such as in chocolates.

Wetting agents Lecithins provide complete wetting of fatty or hydrophilic powders in aqueous systems. The fatty acids are attracted to the fatty portion and the hydrophilic portions of the molecules actively imbibe water and control the hydration of the powder.

Manufacture

The majority of phospholipids are solvent-extracted from their source. Usually, nonpolar hydrocarbons, like hexane, are used. Soya beans, for example, are cleaned, cracked, and dehulled before flaking and extraction. The hexane is removed from the solvent micelle and the crude soya bean oil is cooled for further refining. Figure 1 shows a flow diagram for the degumming of crude soya bean oil and the production of lecithin. Approximately 2–3% water is added to the crude oil and agitated for at least 30 min. The phosphatides hydrate, swell, and are separated by centrifugation. The wet gums with 50% moisture are dried through a thin-film drier. The important points here are careful drying temperatures and cooling of the product below 20 °C. Lightening the color of the product can be achieved with hydrogen peroxide in the gum stage. The hydrogen peroxide is removed through drying.

Modification

The chemistry and functionality can be altered by simple chemical additions with acids and bases as well as hydrogen peroxide and acetic anhydride. These modifications increase the dispersion and hydration properties of the lecithin. Enzyme modification is also possible with lipases and phospholipases. These changes markedly affect the functionality in emulsification.

Composition of Lecithins

The composition of lecithins and their phospholipids will vary depending on their source: vegetable, animal, or bacterial. There are minor differences within a class but major differences between the sources. Vegetable lecithins are high in phosphatidylcholine, phosphatidylethanolamine, phosphatidylinositol, and phosphatidic acid, but very low in phosphatidylserine and contain no sphingomyelin. Animal lecithins are high in phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine and

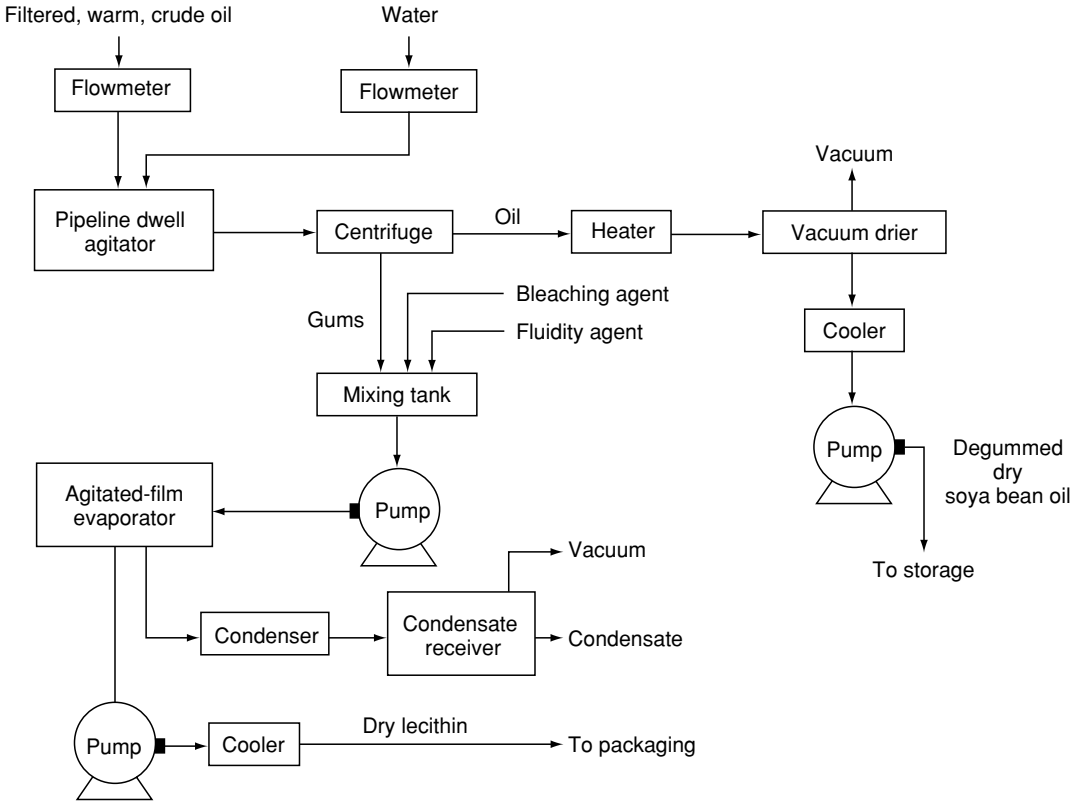

Figure 1 Flowsheet for degumming and crude lecithin production. Reproduced from List GR (1989) In Szuhaj BF (ed.) *Lecithins: Sources, Manufacture and Uses*, p. 149. Champaign: American Oil Chemists' Society, with permission.

sphingomyelin but contain no phosphatidylinositol. Microbes have phospholipids similar to the plant kingdom with high levels of phosphatidylethanolamine and phosphatidylcholine and phosphatidylserine or sphingomyelin.

Specifications of Lecithins

Lecithins may be qualified in several ways, chemically, physically, and functionally, but there are also specifications used to assess quality and purity (Table 3). These include acetone-insolubles (AI), acid value (AV), hexane-insoluble matter (HI), peroxide value (PV), moisture, color, free fatty acids (FFA), divalent metals (DVM), iodine value (IV), and phosphorus. Most of these analytical methods are found in the American Oil Chemists' Society (AOCS) *Official Methods and Recommended Practices*, Section J.

Acetone-Insolubles

Phospholipids are nearly insoluble in cold acetone. This quantitative method should measure the active ingredients in lecithin. Depending on the type of lecithin, the range is 35–98%.

Acid Value

The phosphorus group in lecithins have a titratable acidity that is measured with this volumetric method. FFAs are also measured in this test and should not be confused with phospholipid acidity. The AV range is 20–36 mg of potassium hydroxide per gram.

Hexane-Insolubles

In the processing of soya bean lecithin, particulate matter finds its way through some processes and gives the product a hazy appearance. The HI can be determined by dissolving the product in hexane, centrifuging and gravimetrically measuring the insolubles. The HI range for soya bean lecithins is 0.05–0.3%.

Peroxide Value

The PV is a measure of oxidation in fats and oils. In lecithin, however, the PV usually measures the residual hydrogen peroxide from processing. The range in unbleached products is 0–10 mmol kg^{-1} and in bleached products is 10–75 mmol kg^{-1}.

Table 3 Specifications range for commercial lecithins

Analysis	*Typical range*		
Acetone-insolubles (%)	35–98		
Acid value (mg KOH g^{-1})	20–36		
Hexane-insolubles (%)	0.05–0.3		
Peroxide value (mmol kg^{-1})			
Unbleached	0–10		
Bleached	10–75		
Moisture (%)	0.1–1.0		
Viscosity (Brookfield, 25 °C) (cP)	150–20 000		
Free fatty acids (mg KOH (g^{-1})	1–5		
Iodine value (cg I g^{-1})			
Natural	95–110		
Oil-free	80–90		
	Fatty acid composition (%)		
	Soya bean oil	*Natural*	*Oil-free*
$C_{16:0}$	10.3	15.6	20.3
$C_{18:0}$	4.4	4.7	4.6
Total saturates	14.7	20.3	24.9
$C_{18:1}$	24.5	17.9	9.2
$C_{18:2}$	53.8	54.0	58.9
$C_{18:3}$	7.0	6.7	7.0
Total unsaturates	85.3	78.6	75.1
Unsaturated: saturated ratio	5.8:1	3.9:1	3.0:1

Reproduced from Central Soya, Chemurgy Division (1990), with permission.

Moisture

The water content of lecithins is quite low, at 0.1–1.0%. It can be measured by oven drying but is more accurately determined by the Karl–Fischer method. This water content is so low that there is no measurable water activity for microbial growth. Moisture contents above 1% will change the viscosity from a fluid to a plastic state. (*See* **Water Activity**: Principles and Measurement.)

Viscosity

The Brookfield viscosity at 25 °C will have a range of 150–20 000 cP. Diluents and divalent metals can alter the viscosity to usable levels.

Free Fatty Acids

This method measures the true fatty acid levels in lecithins which range from 1 to 5 mg of potassium hydroxide per gram. Fatty acids are added as additives to adjust the viscosity. (*See* **Fatty Acids**: Analysis.)

Iodine Value

The IV is a traditional method for qualifying lecithin sources. The more unsaturated fatty acids are found in soya bean lecithins, which have a range of 95–110 in natural fluid lecithins, to 80–90 in deoiled lecithin.

Phosphorus

This wet chemical method is an indirect way of measuring the phospholipid content. The typical level of phosphorus is 2.0% in fluid lecithin and 3.0% in deoiled products. The AOCS method Ca 12–55 has an approximation for converting percent phosphorus to the phosphatides in soya bean oil. The equivalent phosphatides content is equal to percent phosphorus $\times$ 30.

Uses of Lecithin

There are many uses of phospholipids in the food industry. As seen from the functional properties, there are multiple functions for lecithins. The following is a listing of the major areas of use:

- margarines – emulsifier, stabilizer, and antispatter;
- confectionery and snack foods – crystallization control, viscosity control, antisticking;
- instant foods – wetting and dispersing agent, emulsifier;
- commercial bakery products – crystallization control, emulsifier, wetting agent, release agent;
- cheese products – emulsifier, release agent;
- meat and poultry processing – browning agent, phosphate dispersant;
- dairy and imitation products – emulsifier, wetting and dispersing agent, antispattering and release agent;
- packaging aid – release agent, sealant;
- processing equipment – internal or external release agent, lubricant.

Refer to individual foods

Applications of lecithins in foods are clearly supported by the Food Chemicals Codex, the European E322 regulations, and they are considered as generally regarded as safe substances by the US Food and Drugs Administration.

Storage and Handling

Lecithins are very stable products. They are shipped in drums or in bulk containers. They can be stored at ambient temperatures for up to 2 years without loss of activity or becoming rancid or spoiling. The water activity is so low that no microbial growth can occur. The products may be heated to 25 °C for easier application. (*See* **Water Activity**: Effect on Food Stability.)

See also: **Eggs**: Structure and Composition; **Emulsifiers**: Organic Emulsifiers; **Fatty Acids**: Properties; Analysis; **Soy (Soya) Beans**: Properties and Analysis; **Spices and Flavoring (Flavouring) Crops**: Tubers and Roots; **Triglycerides**: Structures and Properties; **Water Activity**: Principles and Measurement; Effect on Food Stability

Further Reading

Burner D (ed.) (1991) *Official Methods and Recommended Practices*. Champaign: American Oil Chemists Society.
Central Soya, Chemurgy Division (1990) *The Lecithin Book*. Fort Wayne: Central Soya.
Charalambous G and Doxastakis G (1989) *Food Emulsifiers: Chemistry, Technology, Functional Properties and Applications*. New York: Elsevier.
Hanin I and Pepeu G (1990) *Phospholipids: Biochemical, Pharmaceutical, and Analytical Considerations*. New York: Plenum Press.
Szuhaj BF (ed.) (1989) *Lecithins: Sources, Manufacture and Uses*. Champaign: American Oil Chemists' Society.
Szuhaj BF and List GL (eds) (1985) *Lecithins*. Champaign: American Oil Chemists' Society.

Determination

B F Szuhaj, Central Soya Company Inc., Fort Wayne, IN, USA

Introduction

Phospholipids are a well-known class of lipids that have been thoroughly analyzed over the past three centuries. Their complete analysis was facilitated by the great advances in separation science and qualitative procedures that have occurred in the last 50 years. This article will cover the analysis of phospholipids from their structure and composition, extraction techniques, qualitative and quantitative assays, and industrial methodology. (*See* **Fats**: Classification.)

Structure of Phospholipids

There are at least a dozen compounds that fall into the class of phospholipids. They have a basic structure of a diacylglycerol backbone with a phosphate ester on the α or third carbon of the glycerol molecule. Usually another compound is attached that characterizes the phospholipid. **Figure 1** shows examples of the structures of the major phospholipids. These phospholipids (and their common abbreviations) are:

- phosphatidylcholine (PC, PtdCho)
- phosphatidylethanolamine (PE, PtdEth)
- phosphatidylserine (PS, PtdSer)
- *N*-acylphosphatidylethanolamine (NAPE, *N*-acyl-Ptd-Eth)
- phosphatidylinositol (PI, PtdIns)
- phosphatidic acid (PA, PtdA)
- phosphatidylglycerol (PG, PtdGro)
- plasmologen (PM)
- diphosphatidylglycerol (DPG, diPtdGro)
- lysophosphatidylcholine (LPC, lysoPtdCho)
- lysophosphatidylethanolamine (LPE, lysoPtdEth)

The proper nomenclature for phospholipids has been defined by the 1976 revised recommendations of the International Union of Pure and Applied Chemistry (IUPAC) and the International Union of Biochemistry (IUB) Committee on Biochemical Nomenclature. For example, the term 'lecithin' is permitted for phosphati-dylcholine but the systematic name is 1,2-diacyl-*sn* -glycero-3-phosphorylcholine. The generic name of 3-*sn*-phosphatidylcholine could be used. The abbreviation PtdCho is also allowed. This article will use the common names listed above since the literature has thousands of references with this terminology.

Composition of Phospholipids

The composition of phospholipids depends on the source of the phospholipids. Those from animal, plant, and microbial sources will have different compositions, depending on the nature of the tissue from which the lipids are extracted – for example, brain, liver, or blood. In plants, it will vary on whether they are from soya beans, corn, cotton, rapeseed, or sunflower. In microbial sources it depends upon the organism.

The phospholipid classes are similar within a species, but differ primarily in fatty acid acyl composition around the 1 and 2 positions on the glycerol backbone.

Fatty Acids

The fatty acid chain length is commonly from C_4 to C_{26}, with different degrees of unsaturation from one to six double bonds, which may be at different locations on the acyl group. However, there is a pattern that is relevant to the present discussion. (*See* **Fatty Acids**: Properties.)

In animals the primary fatty acids range from $C_{12:0}$ (lauric acid) to $C_{24:0}$ (tetracosanoic acid). Again, depending on the species and tissue extracted, the fatty acids can be variable.

In the plant kingdom, the primary diacyl groups on the phospholipids will range from $C_{12:0}$ to $C_{18:3}$, i.e., lauric to linolenic acid. There are usually no C_{20} fatty acids and higher, as in the animal kingdom. The degree of unsaturation depends on the origin of the crop, i.e., from temperate or tropical regions. Also, the climate within the zone can make a seasonal difference.

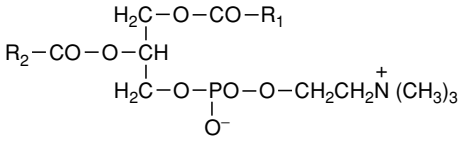

Phosphatidylcholine

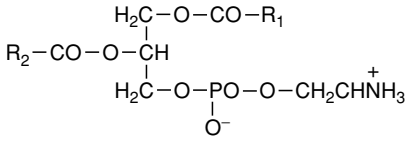

Phosphatidylethanolamine

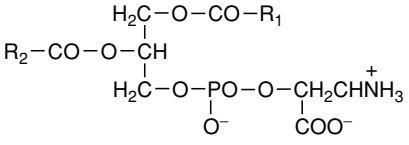

Phosphatidylserine

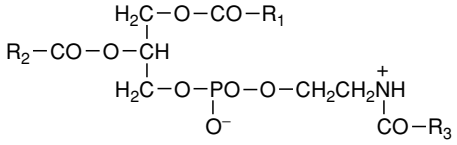

N-Acylphosphatidylethanolamine

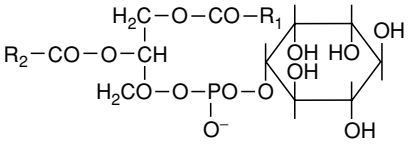

Phosphatidylinositol

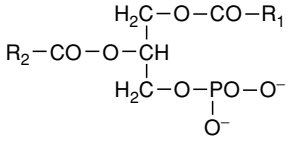

Phosphatidic acid (PA)

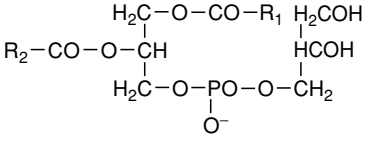

Phosphatidylglycerol

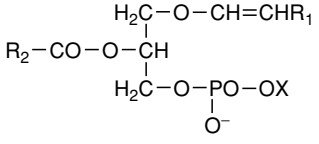

Plasmalogen
X = Choline or ethanolamine

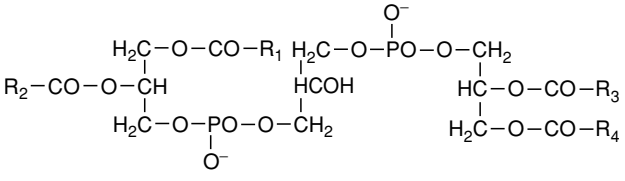

Diphosphatidylglycerol

Figure 1 Structures of the major phospholipids. Reproduced from Scholfield CR (1985) In: Szuhaj BF and List GR (eds) *Lecithins*, p.3. Champaign: American Oil Chemists' Society, with permission.

Microbial phospholipid acyl groups are more similar to those found in the plant kingdom than they are to those from the animal kingdom. The predominant acyl fatty acids range from $C_{16:0}$ to $C_{18:3}$. There are more odd-numbered carbon fatty acids in microorganisms than elsewhere.

A further factor that makes phospholipid composition more complex is the ability to have different fatty acids on the 1 and 2 positions of the molecule. Most research has shown that polyunsaturated fatty acids are usually in the 2 position.

Extraction Techniques

One of the most important factors in phospholipid analysis is the initial extraction procedure. If the analysis is on a finished commercial lecithin there is no problem, but if the analysis is from tissue samples or food samples the extraction technique will be critical in obtaining meaningful results.

Tissue Samples

For many years the Folch extraction of tissue homogenates with chloroform/methanol 2:1 (v/v) has been the method of choice by most researchers. Some found that the use of chloroform/methanol 1:1 (v/v) was preferable and some have used a biphasic system of butanol/methanol with dilute hydrochloric acid. Some have used hexane/2-propanol 3:2 (v/v). Which solvent system used depends largely on the required accuracy, but in most cases chloroform/methanol 2:1 (v/v) is the best solvent to try initially.

Food Samples

Since food samples may have lecithin or phospholipid added rather than being incorporated into the tissues of the food matrix, the dried sample can be ground and extracted with petroleum ether. If the lipids are bound in the product through processing, chloroform/methanol 2:1 (v/v) can be used. Because of environmental considerations di- or trichloroethane should be used in place of chloroform.

Drying of high-moisture products is required, but not by oven drying or air drying, as oxidation of the fatty acids can occur. Freeze drying of the product is preferred.

Qualitative Analysis

There are several ways to detect the presence of phospholipids. Traditional instrumental methods using ultraviolet and infrared are not used as often since thin-layer chromatography (TLC) can give a qualitative and semiquantitative result in one assay. Also, the use of conformational chemical sprays on the TLC plates can further identify the products.

Thin-Layer Chromatography

There are several types of silica gel plates available for TLC. Silica gel G and H are the most useful. Phospholipids may be separated on a 20×20 cm plate in one or two directions. A polar solvent and a nonpolar solvent system are used. The polar system is chloroform/methanol/water (65:25:4, v/v/v) and the nonpolar solvent is petroleum ether/diethyl ether/acetic acid (90:10:1, v/v/v). See the American Oil Chemists' Society (AOCS) recommended practice Ja 7-86 for alternative methodology. (*See* **Chromatography**: Thin-layer Chromatography.)

These TLC plates are air- or oven-dried after separation of 20–50 μg of sample and are sprayed with 10% sulfuric acid and heated to char the phospholipids. Alternatively, they may be sprayed with a phosphorus spray containing molybdenum blue. Phospholipids stain a deep blue on heating the TLC plate. An example is shown in Figure 2.

Nondestructive visualization techniques can be used if the phospholipids are to be determined or the fatty acid composition is to be run. Ultraviolet light and 2′,7′-dichlorofluorescein easily detects lipids on TLC. On prep plates, the bands are scraped off and extracted with chloroform/methanol (1:1, v/v) and the fatty acids converted to methyl esters using boron trifluoride and then determined using gas–liquid chromatography (GLC).

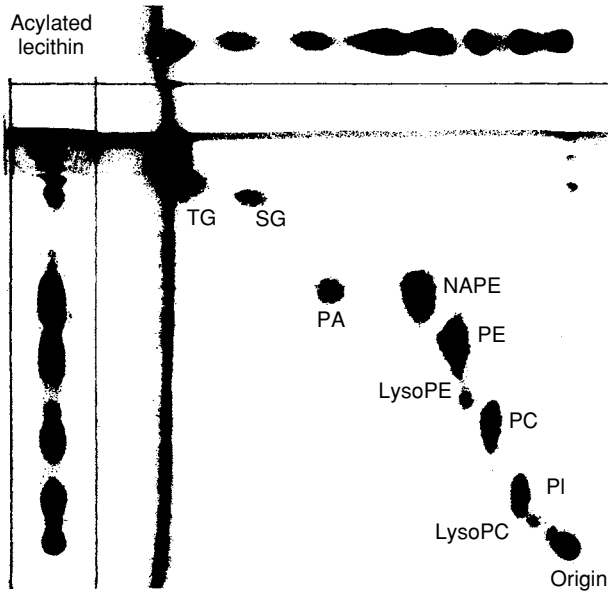

Figure 2 Thin-layer chromatogram of soya bean lecithin in two dimensions: triglycerides (TG), sterol glucosides (SG), phosphatidic acid (PA), *N*-acylphosphatidylethanolamine (NAPE), phosphatidylethanolamine (PE), lysophosphatidylethanolamine (Lyso PE), phosphatidylcholine (PC), phosphatidylinositol (PI), lysophosphatidylcholine (LysoPC). Silica gel plate; first dimension chloroform/methanol/acetic acid/water, 85:15:15:3 (v/v/v/v); second dimension chloroform/acetone/methanol/acetic acid/water (10:4:2:2:1) (v/v/v/v/v). Courtesy of J. Yaste, Central Soya, Food Research, Fort Wayne, Indiana, USA.

Quantitative Analysis of Phospholipids

Column Chromatography

Column chromatography precedes TLC in the separation of phospholipids. The techniques are slow and require good skill with column preparation, flow rates, and solvent removal. Commercial lecithins can be separated by dissolving the crude mixture in petroleum ether and passing it through a deactivated silica gel column with petroleum ether. The phospholipids are adsorbed and do not pass through the column, whilst triglycerides and sterol esters are eluted. The phospholipids are subsequently quantified by TLC and wet phosphorus analysis.

High-Performance Liquid Chromatography (HPLC)

Newer technologies have found that HPLC can separate and quantify phospholipids more quickly and accurately. Separation is carried out on several types of columns, including silica gel and an amino group bonded to the silica surface (μBondapak-NH_2). The columns are eluted with chloroform/methanol gradients, acetonitrile/methanol/85% phosphoric acid, or acetonitrile/methanol/ water. The eluent is measured at 205 nm or detected with flame ionization. Figure 3

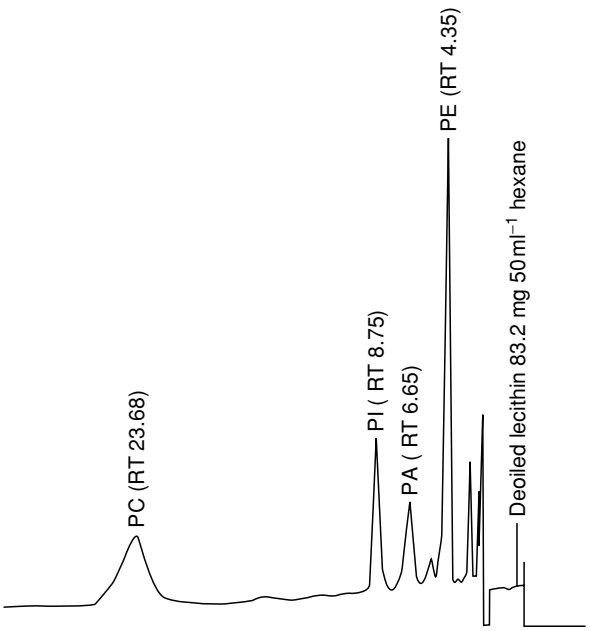

Figure 3 High-performance liquid chromatography of deoiled soya bean lecithin. PC, phosphatidylcholine; PI, phosphatidylinositol; PA, phosphatidic acid; PE, phosphatidylethanolamine. Column, μ Porasil 10 μ 3.98 $\times$ 300 mm; mobile phase, hexane/2-propanol/acetate (8:8:1, v/v/v), buffer pH 4.2; detection, ultraviolet (206 nm); injection, 10 μl; flow rate, 1 ml min^{-1}. Retenton time (RT) in minutes. Courtesy of P. Balazs, Central Soya, Food Research, Fort Wayne, Indiana, USA.

shows an HPLC separation of commercial lecithin, using ultraviolet detection. The mass detector, an evaporative analyzer, has also been successfully used for the HPLC determination of phospholipids. (*See* **Chromatography**: High-performance Liquid Chromatography.)

Densitometry

Densitometric scanning has been used as an indirect method for determining phospholipid content on TLC plates. While the method has some promise, a problem is the quanitative charring of the phospholipid spots. Each phospholipid has a different charring density and this depends on fatty acid composition. Only with proper standards can this method be useful.

Thincography

Thin-rod TLC combines TLC with quantification by flame ionization detection. Rods are used rather than plates but controversy still exists over the suitability of the technique for routine lipid analyses.

Phosphorus Analysis on Phospholipids

Phosphorus analysis is an indirect method for the quantification of phospholipids because the qualitative composition of the sample must be known, if accurate values are to be obtained. With pure phosphatides this will work well, but most separation techniques give mixed phospholipids. The preferred method for phosphorus in lecithins is the AOCS method Ja 5-55. This determines the total phosphorus content of the sample. For commercial lecithins a multiple factor of 30 is used to convert total phosphorus values to acetone-insoluble value.

There are various methods to determine phosphorus through molybdenum blue and molybdovanadophosphate yellow. To improve reproducibility many factors need to be evaluated. This includes the digestion method, chromogenesis, and sensitivity. The AOCS method is the most straightforward and should be used especially in the food area.

Industrial Methods of Analysis

Phospholipids are characterized by a different set of assays than the determination of compounds used for academic or biochemical use. Most commercially available phospholipids come from soya beans and from eggs, and the methods outlined below can be used to qualify or categorize the products.

Acetone-Insolubles (AI)

This method determines the content of phospholipids in commercial lecithins. The method employs AOCS method Ja 4-46. The AI is an approximation of the active ingredients in formulations. Cold lecithin-saturated acetone must be used in this test.

Acid Value (AV)

This method determines the phosphatide and fatty acid content of commercial phospholipids. The method utilized is AOCS method Ja 6-55. Phosphatide acidity is often confused with fatty acid addition to commercial lecithin. It is a combination of organic acids and phosphoric acid.

Peroxide Value (PV)

This is a measurement of oxidative state of commercial phospholipids. It measures the milliequivalents of peroxide per kilogram of sample which oxidize potassium iodide. It also measures the residual peroxide used in process stabilization and bleaching. AOCS method Ja 8-87 is used.

Free Fatty Acids (FFA)

This method utilizes AOCS method Ca 5a-40. When run on the acetone-soluble portion of the AI method, it gives the added fatty acids.

Phosphorus Content

The determination of total phosphorus is an indirect method for quantifying phospholipids. This method (AOCS method Ja 5-55) quantifies phospholipids through a molybdate reaction to a chromophore quantitated by phenolphthalein titration.

Gas Chromatography (GC)

This method is commonly used to measure the fatty acid composition and does not quantify the phospholipids themselves (AOCS method Ce 1-62).

High-performance Liquid Chromatography

This is gradually replacing the older techniques for qualifying and quantifying particular phospholipids in commercial mixtures. A uniform technique is being addressed by the AOCS.

Phospholipid analyses have come a long way since their study by Theuticum, *c*. 1800. Each analyst must choose the best method depending on constraints of accuracy and time.

See also: **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography; **Fatty Acids**: Properties; **Fats**: Classification

Further Reading

Burner D (ed.) (1991) *Official Methods and Recommended Practices*. Champaign: American Oil Chemists' Society.

Central Soya, Chemurgy Division (1990) *The Lecithin Book*. Fort Wayne: Central Soya.

Charalambous G and Doxastakis G (1989) *Food Emulsifiers: Chemistry, Technology, Functional Properties and Applications*. New York: Elsevier.

Hanin I and Pepeu G (1990) *Phospholipids: Biochemical, Pharmaceutical, and Analytical Considerations*. New York: Plenum Press.

Szuhaj BF (1989) *Lecithins: Sources, Manufacture and Uses*, Champaign: American Oil Chemists' Society.

Szuhaj BF and List GL (1985) *Lecithins*. Champaign: American Oil Chemists' Society.

Physiology

T H M Da Costa and M K Ito, Universidade de Brasília, Brasília, DF, Brazil

Introduction

Phospholipids are ubiquitous molecules that, due to their chemical properties, have important structural as well as metabolic roles in the cell. The first phospholipid to be discovered was phosphatidylcholine. It was originally named lecithin after the Greek *lekithos*, which means egg yolk. The concept of phospholipids as mere structural and metabolic inert molecules associated with methological difficulties delayed the interest in these compounds until the 1950s, when elucidation on the biosynthetic pathway of phospholipids began to be published. Today, the metabolic and physiological roles of phospholipids are an active and exciting area of research. In this article we discuss the physical and biochemical properties of phospholipids in cellular and subcellular membranes, the biosynthetic and hydrolytic pathways, and the role of phospholipids in signal transduction.

Functional Role of Membrane Phospholipids

Phospholipids are a major component of cellular membrane and play a pivotal role in the communication between extra- and intracellular space. Phospholipids represent a large class of compounds that have fatty acyl chains esterified to glycerol and a charged or zwitterionic head group (**Figure 1**). The fatty acyl chains usually have an even number of carbon atoms from 12 to 26, with 80% being 16–20 carbon atoms long. They may have up to six double bonds, commonly present as a *cis* isomer in one of the fatty acids that creates a kink in the chain. Differences in the length and saturation of the fatty acid tails are important for their influence in the ability of phospholipid molecules to pack against one another, and for this reason they affect the fluidity of the membrane.

In animal cells phosphatidylcholine (PtdCho) is the major phospholipid, whereas in bacteria phosphatidylethanolamine (PtdEtn) is the predominant species. Sphingomyelin is found in most animal cell membranes and belongs to a different group of phospholipids that fits into the overall pattern of phospholipid structure (**Figure 1**). In many mammalian cells, four major phospholipids predominate in the plasma membrane: PtdCho, phosphatidylserine (PtdSer), PtdEtn, and sphingomyelin. Only PtdSer carries a net negative charge, while the other three are electrically neutral at physiological pH with one positive and one negative charge. Sphingomyelin and PtdCho carry a molecule of choline in their head group (**Figure 1**). The phospholipid bilayer of biological membranes confers the amphipathic feature with hydrophobic and hydrophilic domains. The polar head groups of phospholipids face the aqueous exterior of the membrane while hydrophobic regions of the

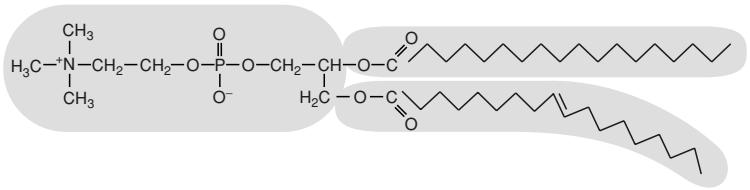

Phosphatidylcholine
(PtdCho or lecithin)

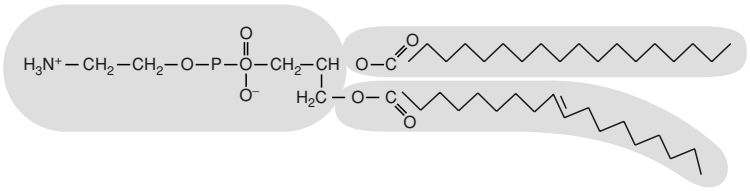

Phosphatidylethanolamine
(PtdEtn)

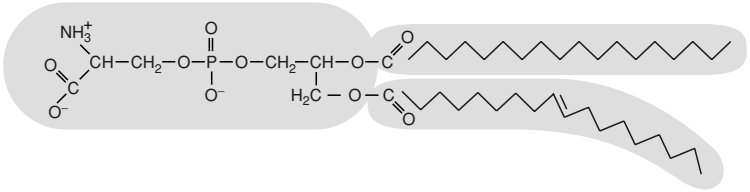

Phosphatidylserine
(PtdSer)

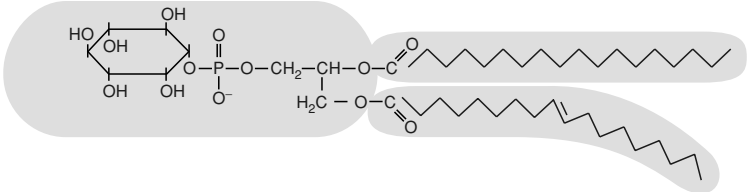

Phosphatidylinositol
(PtdIns)

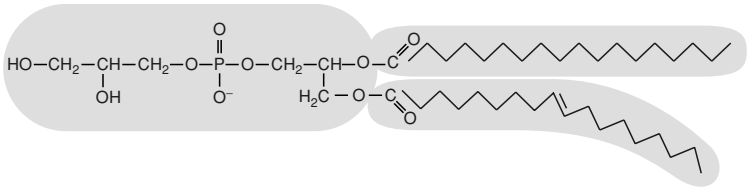

Phosphatidylglycerol
(PtdGro)

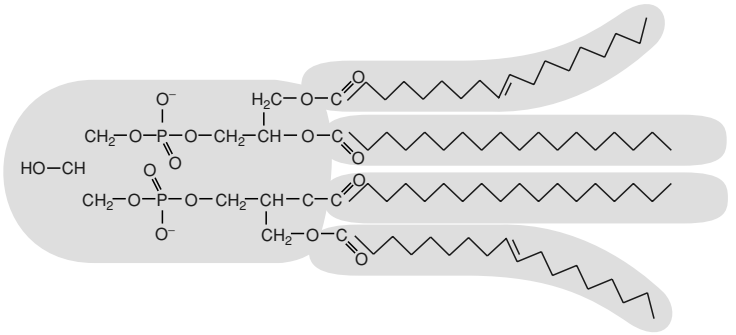

Diphosphatidylglycerol
(cardiolipin)

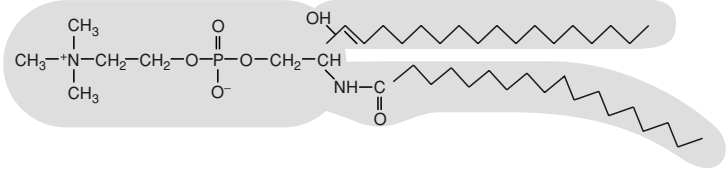

Sphingomyelin

Figure 1 Chemical structure of phospholipids. The head group with electron charge is presented as the oval part of the molecule and the tails are the fatty acyl chains. The presence of a double bond creates a kink in the chain.

lipids are sequestered away from the water to the interior of the bilayer in the most thermodynamically favorable arrangement. Because of such chemical characteristics, membrane phospholipid molecules spontanously form bilayers in aqueous environments and form sealed compartments.

Cellular membranes are organized in mosaic-like structures which are dynamic and mobile. The terms 'fluidity' and 'motility' are used to describe properties of membrane and membrane components, respectively. Fluidy reflects the viscosity within the membrane that depends on both the composition and temperature within the membrane. The change in viscosity or in phase transition (from liquid to gel or crystalline state) is determined by the chain length and saturation of the hydrocarbon chains. A shorter chain length reduces the tendency of the hydrocarbon tails to interact with one another and *cis* double bonds produce kinks in the hydrocarbon chains that make them more difficult to pack together, making membranes more fluid at lower temperatures. Phospholipid mobility is determined by intramolecular and intermolecular motion. Intramolecular motion includes three main types: (1) rotation or vibration about each C–C bond within the fatty acyl chain (segmental motion); (2) rotation of the entire molecule about the long axis perpendicular to the plane of the bilayer; (3) a pendulum-like motion of the fatty acyl chains. Intermolecular motion refers to lateral diffusion of complete phospholipid molecules and occurs by interchange of one lipid molecule for another. These time motions have a different time scale, i.e., segmental motion is very fast, whereas lateral diffusion is slower. Flip-flop or migration of phospholipid molecules from one monolayer to another is a comparatively less common type of intermolecular motion. These movements are possible because of the presence of a special class of endoplasmic reticulum (ER) membrane-bound enzyme, called phospholipid exchange proteins or phospholipid translocators, which catalyze the rapid flip-flop of specific phospholipids from the inner to the outer monolayer of plasma membrane or from one organelle membrane to another, e.g., from microsomal vesicles to mitochondria. The phospholipid movements in biological membranes are presented schematically in Figure 2.

Phospholipid fraction of cellular membranes varies considerably in total amount and in the composition of the fatty acids. This diversity occurs across species, among tissues, and even among organelles in the same cell (Table 1).

The functional role of differences in fatty acid composition of cell membranes and the fatty acid asymmetry between phospholipids from the outer and the inner membrane have an important functional significance. In human red blood cell membranes, most of the PtdSer, which is negatively charged, is located in the inner monolayer. This position is important for the activity of the enzyme protein kinase C which requires the negative charge of phosphatidylserine (Figure 2). Similarly, specific inositol phospholipids are concentrated in the cytoplasmatic half of the plasma membrane where it can be reached by specific enzymes and the product of the phosphatidylinositol molecules serves as intracellular signals.

Biosynthesis of Phospholipids

A more direct approach about the biosynthetic pathways of phospholipids, known as the Kennedy pathway, became possible with the utilization of radioactive and stable isotopes, which permitted advances in the understanding of phospholipid active synthesis and breakdown in the cells and their role as a metabolic pool for cellular signals. These techniques are now combined with genetic manipulation for the study of the pathways involved in phospholipid metabolism.

The ER of eukaryotic cells is the predominant site of phospholipid and sphingolipid assembly, but there are phospholipid biosynthetic enzymes localized in the mitochondria and microsomes. The yeast *Saccharomyces cerevisiae* has become the model of choice for the study of the regulation of phospholipid biosynthesis, owing to the possibility of a biochemical, molecular, and genetic approach. The model system of *S. cerevisiae* has allowed advances in the study of intracellular transport and assembly of lipids into membranes, analysis of the genes, and gene products required for lipid metabolism and definition of the role of lipids in signal transduction. Many of the enzymes were isolated and purified and metabolic pathways were targeted by genetic manipulation of specific genes which were impossible in other eukaryotic organisms. In addition, the pathways of phospholipid metabolism in *S. cerevisiae* are, with few exceptions, similar to those found in higher eukaryotes. They also display the intracellular compartmentalization and subcellular membranes typical of eukaryotes which are important for the study of localization of the phospholipid biosynthesis enzymes.

Figure 3 shows the biosynthetic pathways in *S. cerevisiae*. The important steps are outlined, together with the enzymes responsible for specific reactions. The whole biosynthetic process is highly interconnected to the sorting and delivery of plasma and subcellular membrane. Synthesis of long-chain fatty acids is achieved by the coordinated action of acetyl-coenzyme A (acetyl-CoA) carboxylase, fatty acid synthase and acyl-CoA synthetases (FAAs),

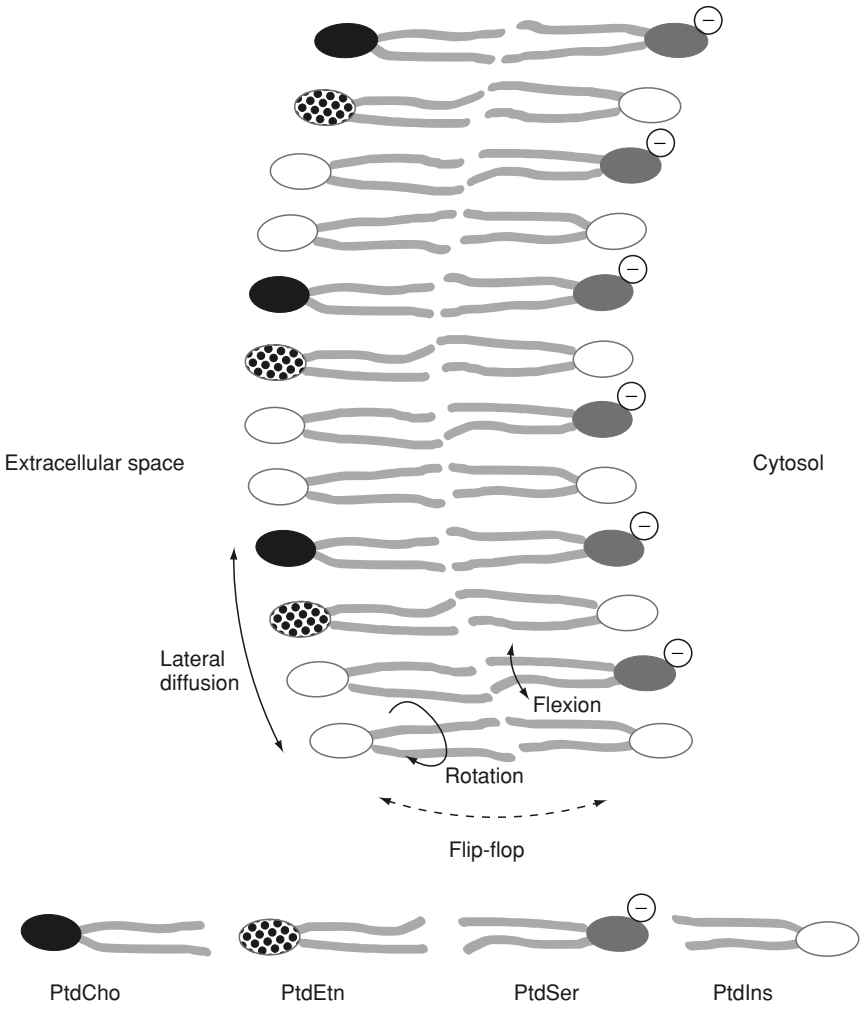

Figure 2 Schematic representation of a biological membrane section, showing the possible phospholipid movements within each phospholipid molecule, between molecules and between the inner and outer phospholipid monolayer. The asymmetry of phospholipid species is also presented, with the predominance of phosphatidylcholine (PtdCho) and phosphatidylethanolamine (PtdEtn) facing the outer half and phosphatidylserine (PtdSer) with a negative charge and phosphatidylinositol (PtdIns) facing the inner half of the membrane.

Table 1 Phospholipid composition of cell membranes

Membrane	*PtdCho*	*PtdEth*	*PtdSer*	*PtdIns*	*PtdGro*	*diPtdGro*	*PtdOH*
Rat liver							
Endoplasmic reticulum	58	17	4	9			
Plasma membrane	56	15		10	2		
Mitochondrial (inner)	< 3	3	25	1	6	2	18
Mitochondrial (outer)	< 5	5	23	2	13	3	3
Nuclear	55	20	3	7			
Rat brain							
Myelin	11	14	7				
Erythrocytes							
Rat	31	15	7	2			
Human	23	20	8	3			
Sheep	1	23	8	1			
Escherichia coli plasma membrane		80			15		5

PtdCho, phosphatidylcholine; PtdEth, phosphatidylethanolamine; PtdSer, phosphatidylserine; PtdIns, phosphatidylinositol; PtdGro, phosphatidylglycerol; diPtdGro, diphosphatidylglycerol (cardiolipin), PtdOH, phosphatidic acid.
Adapted with permission from Jain MK and Wagner RC (1980) *Introduction to Biological Membranes*, p. 36. New York: John Wiley.

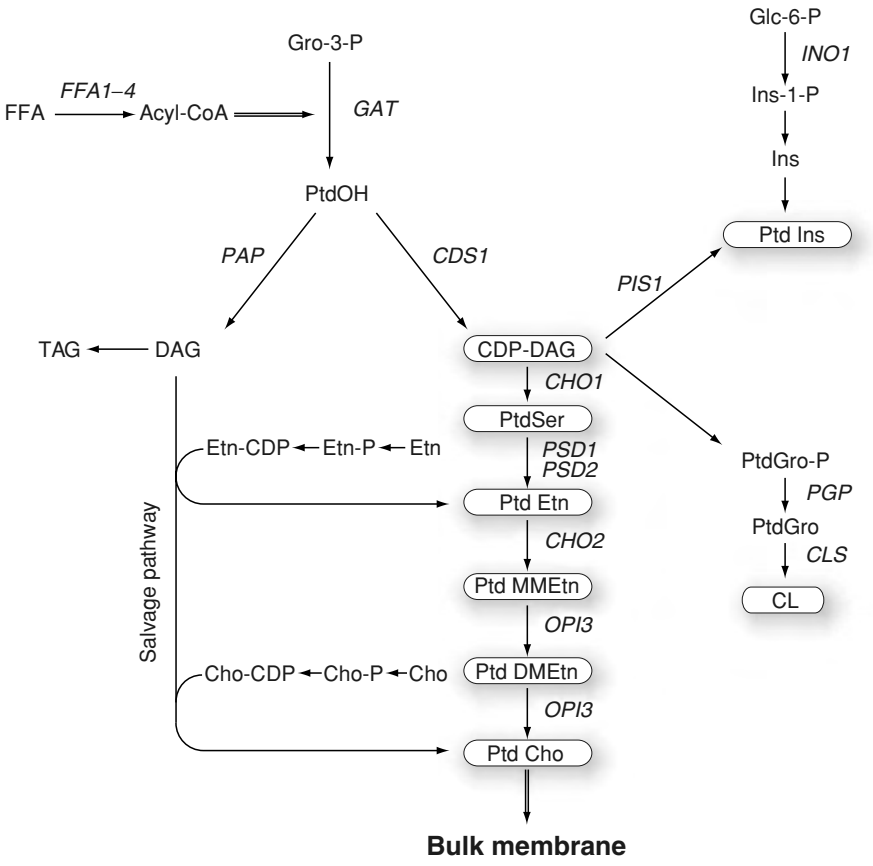

Figure 3 Phospholipid biosynthetic pathways in *Saccharomyces cerevisiae*. FFA, free fatty acid; Acyl-CoA, acyl coenzyme A; Gro-3-P, glycerol-3-phosphate; PtdOH, phosphatidic acid; DAG, diacylglycerol; TAG, triacylglycerol; CDP-DAG, cytidine diphosphate-diacylglycerol; PtdSer, phosphatidylserine; PtdEtn, phosphatidylethanolamine; PtdMMEtn, phosphatidylmonomethylethanolamine; PtdDMEtn, phosphatidyldimethylethanolamine; PtdCho, phosphatidylcholine Cho, choline; Cho-P, choline phosphate; Cho-CDP, cytidine diphosphate-choline. Etn, ethanolamine; Etn-P, ethanolamine phosphate, Etn-CDP, cytidine diphosphate-ethanolamine. Glc-6-P, glucose-6-phosphate; Ins-1-P, inositol 1-phosphate; Ins, inositol; Ptd Ins, phosphatidylinositol. PtdGro-P, phosphatidylglycerol-phosphate; PtdGro, phosphatidylglycerol; CL, cardiolipin. Enzymes abbreviations (italic): *FFA 1–4*, acyl-CoA synthetases 1–4; *GAT*, glycerol-3-phosphate acyltransferase; *PAP*, phosphatidate phosphatase; *CDSI*, CDP-diacylglycerol synthase; *CHO1*, phosphatidylserine synthase; *PSD1,2*, phosphatidylserine decarboxylase 1,2; *CHO2*, phosphatidylethanolamine *N*-methyltransferase; *OPI3*, phosphatidyl-*N*-methyltransferase; *PISI*, phosphatidylinositol synthase; *INO1*, inositol-1-phosphate synthase; *PGP*, phosphatidylglycerophosphate phosphatase; *CLS*, cardiolipin synthase.

-elongase(s) and -desaturase(s). The fatty acids are esterified to sn-1 and sn-2 to glycerol-3-phosphate by the enzyme glycerol-3-phosphate acyltransferase (GAT) (Figure 3). Cytidine diphosphate-diacylglycerol (CDP-DAG) synthase (CDS) converts phosphatidic acid to CDP-diacylglycerol (CDP-DAG), the major intermediate in phospholipid synthesis. The availability of CDP-DAG at specific cellular sites will direct the synthesis of different phospholipids, such as cardiolipin (CL), which occurs exclusively in mitochondria and phosphatidylinositol (PtdIns) and PtdSer to distinct subfractions of the ER or the Golgi apparatus. There is extensive transfer of intermediates and cross-compartment integration for the synthesis of phospholipids. For example, the enzyme phosphatidylserine synthase (CHO1) is localized in the ER, while the steps of conversion of PtdSer to PtdEtn take place in the inner mitochondrial membrane. The PtdEtn synthesized in the mitochondria must again migrate to the ER to insure synthesis of PtdCho, the most abundant phospholipid of *S. cerevisiae*.

In the absence of exogenous choline, PtdCho is synthesized by the *de novo* pathway by a three-step methylation of PtdEtn, which is catalyzed by the ER methytransferases. In the case of mutants with defects in phosphatidylserine decarboxylase (PSD) and methyltransferases, choline and ethanolamine must be provided in the medium to enter phospholipid biosynthesis via the salvage pathway (Figure 3). This route also insures recycling of PtdEtn and

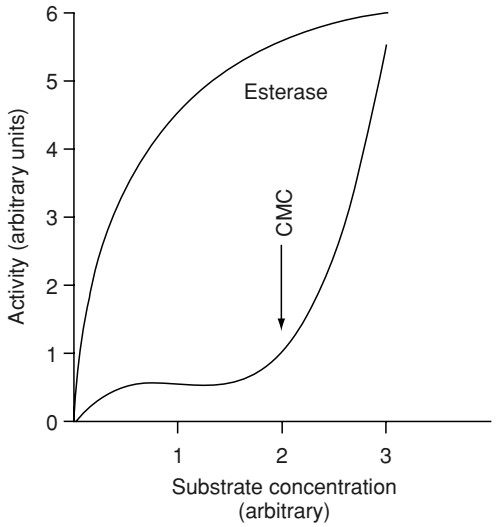

Figure 4 Esterase exhibits Michaelis–Menden kinetics, whereas phospholipase needs critical micellar concentration (CMC) of the substrate for full activity.

PtdCho degradation products and control of local levels of the potent second messenger diacylglycerol.

Degradation of Phospholipids

Undoubtedly, the most important function of phospholipids in the cell membrane relates to their breakdown by the action of phospholipases. The phospholipases are enzymes that hydrolyze phospholipids on water–lipid interfaces and are distinguished from other esterases by their relatively low activity in soluble monomeric phospholipids below their critical micellar concentrations (CMC), but become fully active in aggregated phospholipid structures above their CMC, such as in micelle, bilayers, or hexagonal structures.

As shown in Figure 4, whereas esterases show classical Michaelis–Menten kinetics, the phospholipases may reach more than 1000-fold increase in activity as the substrate phospholipid concentration reaches the CMC. Important factors responsible for this increased rate of hydrolysis include high local substrate concentration, amphipatic substrate orientation at the interface, enhanced diffusion of products from the enzyme to lipid or aqueous moieties, and conformational change of enzyme upon binding to the interface.

Phospholipases play a central role in the activation of various events related to phospholipid degradation. In general: (1) many phospholipases are digestive enzymes found in high concentrations in venoms, bacterial secretions, and digestive fluids of higher animals; (2) many phospholipases have a regulatory function with their products being cellular mediators such as diacylglycerols, inositol trisphosphate, platelet-activating factor, and the eicosanoids. The actions of lipid mediators, known collectively as second messengers, are highly sensitive and under rigid control by a variety of anabolic and catabolic enzymes. Many pathological states, as in inflammation, allergic reactions, and hypertensive states, are related to activation of the cascade of events involving phospholipid degradation and the intracellular second messenger pathway.

Phospholipases are classified as type A_1, A_2, B, C, or D, according to where they act on the substrate phospholipid (Figure 5).

The phospholipases A are acyl hydrolases, which means they remove one acyl group, yielding one fatty acid and lysophosphatide. Phospholipase A_1 and phospholipase A_2 remove fatty acids at positions sn-1 and sn-2, respectively. Phospholipase A_1 is widely distributed in nature and in mammals the major sources are found in the pancreas and the brain. Intracellularly, phospholipase A_1 is dominant in the ER. The best known examples of phospholipase A_2 are snake venom and the pancreatic enzyme that acts mostly on PtdCho and PtdEdn. It is the major phospholipase of mitochondria. Phospholipase B hydrolyzes both acyl groups (sn-1 and sn-2) and also has high lysophospholipase activity. Phospholipase C acts on the glycerophosphate bond while phospholipase D catalyzes the removal of the polar head group. In mammalian cells, phospholipase C is an important initiator of the PtdIns cycle. Phospholipase D is the major phospholipase in many plant tissues and acts specifically on intact phospholipids to give phosphatidic acid and an alcohol. In plants, it seems to be involved in cell turnover and energy utilization. Phospholipase D is also found in bacterial and mammalian sources, where it is involved in replacement of the polar head groups of membrane phospholipids.

Phospholipids as Precursors of Cellular Signal Transduction

In the cell membrane, phospholipids store important precursors of second messengers. Transduction of signals from hormones, across a membrane, involves the coordinated actions of receptor, membrane proteins, and phospholipids that either stimulate or inhibit the synthesis of a second messenger. A number of processes inside the cell are controlled by the level of second messengers.

Among the best described second messenger systems is the G-protein and phosphatidylinositide

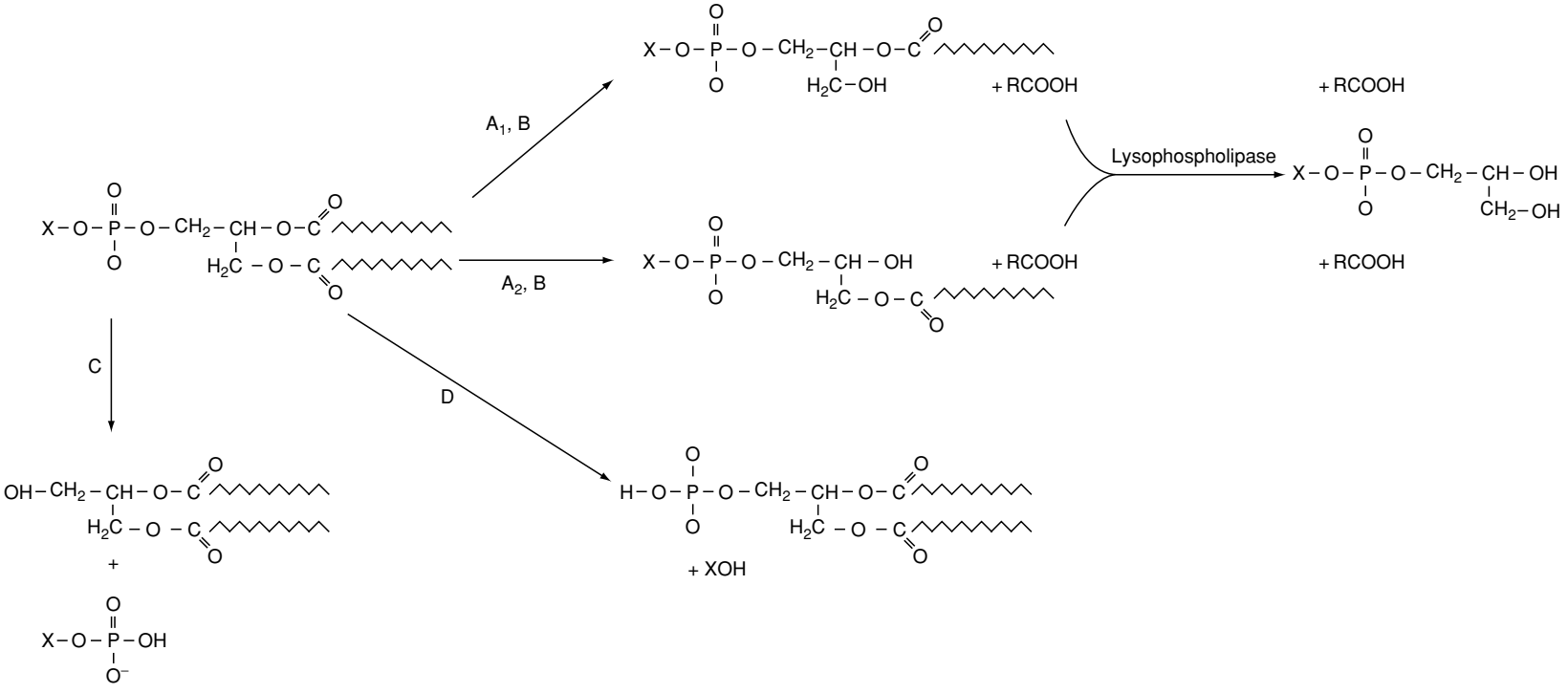

Figure 5 The action of phospholipases A_1, A_2, B, C, and D and lysophospholipase in the hydrolysis of phospholipids. X represents choline, serine, ethanolamine, etc.

system. This system is distinctive in that one stimulus activates membrane reactions that generate two second messengers. The first experimental observations noted that administration of neurotransmitter acetylcholine in pigeon pancreas led to a rapid turnover of PtdIns fraction of membrane phospholipids and release of the digestive enzyme amylase. Similar observations were made in other systems by hormones, neurotransmitters, or growth factors. Despite intensive efforts, the understanding of these mechanisms had not progressed until the early 1980s. It is now recognized that the initial events of inositol phospholipid metabolism occur within 20–30 s of binding of the agonist to the receptor.

Today, the role of specific inositol phospholipids in the phosphoinositide family, namely, phosphatidylinositol 4,5-bisphosphate ($PI(P_2)$), as a membrane-associated storage form of two second messengers, is quite clearly understood.

Figure 6 shows a simplified scheme of the events involved. When an agonist binds to a receptor, membrane G-protein is stimulated to bind guanosine triphosphate. The activated G-protein then acts on the membrane-bound enzyme, phospholipase C, which in turn cleaves $PI(P_2)$ to yield two products, sn-1,2-diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP_3). Both of these products are second messengers. The function of IP_3 is to stimulate the release of calcium ion from intracellular stores, largely from the ER. The increased calcium concentration originates a cascade of events in the intracellular metabolism, including the activation of the membrane-bound protein kinase C (C from calcium). This enzyme requires calcium and PtdSer for its activity. The specific role of the second messenger DAG is to increase the affinity of the protein kinase C for calcium ions. Activated protein kinases will then phosphorylate target proteins inside the cell, which will then be activated and proceed within the cascade. Since many metabolic events are controlled by calcium fluxes and by phosphorylation of specific proteins, the phosphoinositol system shows great ingenuity as a control mechanism. Some of the known target proteins are the insulin receptor, β-adrenergic receptor, and glucose transporter. It is now quite clear that PtdIns is not the only source of DAG which can activate protein kinase C. PtdCho and PtdEtn also seem to serve this role. PtdCho is also a major source of arachidonic acid for the biosynthesis of eicosanoids.

Another well-described example of a regulatory function involving phospholipids in membrane is the generation of another second messenger, known as arachidonate cascade. As described earlier, when an

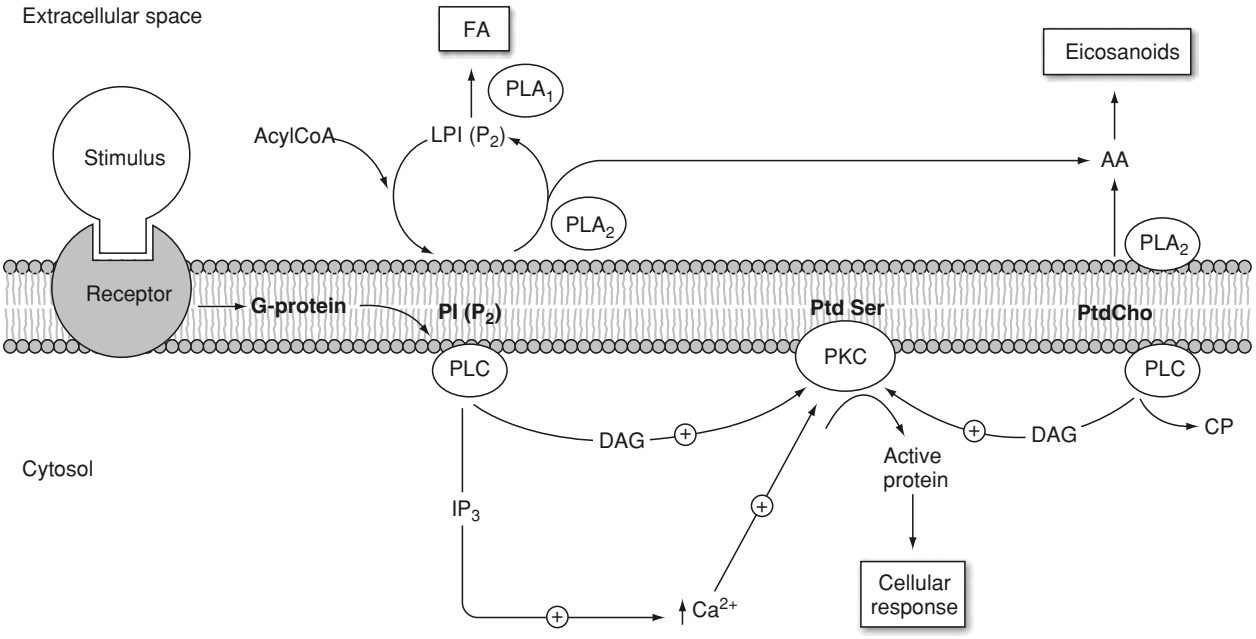

Figure 6 The role of phospholipids in signal transduction. PLA_1, PLA_2, PLC, phospholipases A_1, A_2, C; $PI(P_2)$, phosphatidylinositol 4,5 bisphosphate; IP_3, inositol 1,4,5-trisphosphate; AcylCoA, acyl coenzyme A; DAG, diacylglycerol; FA, fatty acids; AA, archidonic acid; CP, choline phosphate; PKC, protein kinase C; Ptd Ser, phosphatidylserine; PtdCho, phosphatidylcholine; LPI, lysophosphatidylinositol.

agonist stimulates a membrane receptor, a series of events occurs which may lead to the release of 20 carbon polyunsaturated fatty acids from membrane phospholipids, most commonly arachidonic acid (Figure 6). These are tissue-specific stimulants by hormones such as bradykinin or epinephrine, or proteases such as thrombins, to name but a few. Pathological release can occur if membranes are perturbed; for example, the bee sting may stimulate the release of arachidonate from local cell membrane and cause inflammation. This release involves the action of a specific phospholipase A_2 on PtdCho or PtdEdn, yielding arachidonate, and the action of a phospholipase C on PtdIns, yielding a diacylgycerol, which in turn undergoes cleavage to give free arachidonate, which is then converted into a specific eicosanoid in the cell. Eicosanoids, such as prostaglandins, leukotrienes, and thromboxanes are compounds with potent physiological properties which are formed from 20-carbon unsaturated fatty acids into one of a series of eicosanoids according to the enzyme present in the cell and the unsaturated fatty acid released from the membrane. Long-chain polyunsaturated fatty acids of the *n*-3 series, especially from marine origin, when present in the diet will be incorporated into the cell membrane and may replace arachidonic acid in the signal pathway. All eicosanoids are metabolized very rapidly. We still know relatively little about the subsequent effects of eicosanoids at a molecular level, though recent evidence points to their function in the communication pathway with nuclear receptors.

Figure 6 presents events as a coordinated response involving membrane phospholipids and their degradation products. This figure exemplifies the events associated with arachidonic acid release (top) and the mobilization of calcium ions and protein kinase C (bottom). The action of lipid mediators is highly sensitive and under rigid control and when this regulation is not maintained, a number of pathological states, such as those mentioned above, may result.

Current knowledge on the mechanisms involved in intracellular signaling is rapidly advancing and it underscores the importance of the physiological, pathological, as well as the pharmacological properties of this diverse and ubiquitous group of compounds, known as phospholipids. The nutritional significance of phospholipid physiology may be summarized by the fact that foods are the main sources of essential elements of this system.

See also: **Choline**: Properties and Determination; **Essential Oils**: Properties and Uses; **Fatty Acids**: Properties; Metabolism; Gamma-linolenic Acid; **Fish**: Dietary Importance of Fish and Shellfish; **Fish Oils**: Dietary Importance; **Fats**: Classification; Occurrence; **Prostaglandins and Leukotrienes**; **Vegetable Oils**: Dietary Importance

Further Reading

Carman GM and Henry AS (1989) Phospholipid biosynthesis in yeast. *Annual Review of Biochemistry* 58: 635–669.

Eyster KM (1998) Introduction to signal transduction – a primer for untangling the web of intracellular messengers. *Biochemistry and Pharmacology* 55: 1927–1938.

Kohlwein SD, Daum G, Schneiter R and Paltauf F (1996) Phospholipids: synthesis, sorting, subcellular traffic – the yeast approach. *Trends in Cell Biology* 6: 260–266.

Mead JF, Alfin-Slater RB, Howton DR and Popják G (1986) *Lipids: Chemistry, Biochemistry and Nutrition*. New York: Plenum.

Sim E (1982) *Membrane Biochemistry*. Outline Studies in Biology. London: Chapman and Hall.

Vance DE (1991) Phospholipid metabolism and cell signalling in eucaryotes. In: Vance DE and Vance J (eds) *Biochemistry of Lipids, Lipoproteins and Membranes – New Comprehensive Biochemistry*, vol. 20, pp. 205–239. Netherlands: Elsevier.

Waite M (1991) Phospholipases. In: Vance DE and Vance J (eds) *Biochemistry of Lipids, Lipoproteins and Membranes – New Comprehensive Biochemistry*, vol. 20, pp. 269–295. Netherlands: Elsevier.

PHOSPHORUS

Contents
Properties and Determination
Physiology

Properties and Determination

A N Garg, Indian Institute of Technology, Roorkee, U.A., India

Introduction

The extensive and varied chemistry of phosphorus transcends the traditional boundaries of inorganic chemistry because of its vital role in the biochemistry of all living organisms as a constituent of adenosine triphosphate (ATP) and phosphoproteins. It was first isolated from urine by Hennig Brandt in 1669 as a white waxy substance. The spontaneous chemiluminescent reaction of white phosphorus with moist air was the first observed property and was also the origin of its name (Greek: '*phos*,' 'light'; '*pherein*,' 'bearing'). It was also the ancient name for the planet *Venus*, when it appeared before sunrise.

Chemical Properties

Phosphorus is a typical nonmetal placed in group VA of the periodic table. The element phosphorus (P) has an atomic number of 15 and atomic weight of 30.97 with electrons distributed as $1s^2 2s^2 2p^6 3s^2 3p_x^1 3p_y^1 3p_z^1$ and atomic energy levels as shown in Figure 1. Thus, three unpaired electrons, together with the availability of low-lying vacant 3d orbitals, account for the predominant oxidation states III and V in phosphorus chemistry. The most important biological form is the pentavalent oxygen compound phosphate PO_4^{3-}. Phosphorus exists in three main allotropic forms; white, red, and black, each of these being polymorphic. There are at least 11 known modifications, some amorphous, others of some indefinite identity, and all but three of unknown structure. Of these, white phosphorus is soft, waxy, most reactive, and thermodynamically least stable. It has a melting point of 44.1 °C, a boiling point of 80 °C and a specific gravity of 1.82 g cm^{-3}. Red and black forms are heavier with specific gravities of 2.2 and 2.69 g cm^{-3}, respectively. White phosphorus reacts with moist air and gives out light. It ignites spontaneously in air at about 35 °C and is therefore stored in water to prevent combustion. It is soluble in organic solvents such as CS_2 and benzene. If white phosphorus is heated to about 250 °C, or a lower temperature in the presence of sunlight, then red phosphorus is formed. It is a polymeric solid with a melting point of 280.5 °C and sublimes at 430 °C. It is much less reactive than white phosphorus and does not phosphoresce in air. Unlike white phosphorus, red phosphorus need not be stored under water. Its structure is extremely complex, involving a crisscross packing of infinite tubular chains of P atoms. When white phosphorus is heated under high pressure, a highly polymerized form called black phosphorus is obtained. It is also obtained by heating at 220–370 °C for 8 days in the presence of mercury as a catalyst and with a seed of black P. This is thermodynamically the most stable allotrope. It is inert and has a double layer structure with P atoms being bound to three neighbors, as shown in Figure 2. The entire structure consists of a stacking of these double layers with the closest P–P distances within each layer being 2.17–2.20 Å and the shortest P–P distance between layers at 3.87 Å. Therefore, the crystals are flaky like graphite. All forms of phosphorus melt to give the same liquid, which consists of symmetric P_4 molecules with phosphorus atoms occupying corners of a regular tetrahedron. Each atom is directly bonded to the other three atoms P–P = 2.21 ± 0.02 Å. The same molecular form exists in the gas phase at >800 °C and low pressure. The bonding orbitals in P_4 have only 2% of 3s and 3d character. It is most likely that pure

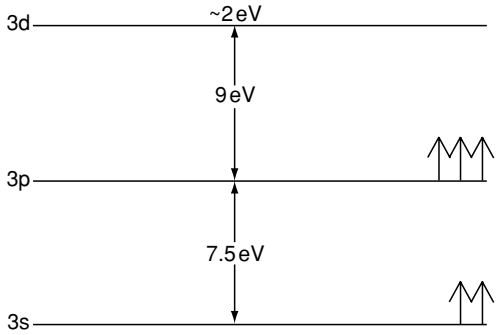

Figure 1 Atomic energy levels in phosphorus.

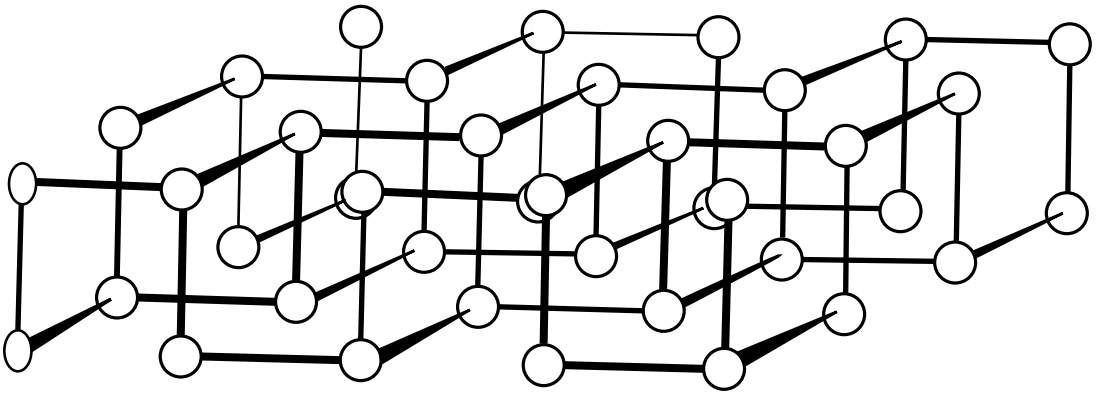

Figure 2 Arrangement of P atoms in the double layers in crystalline black phosphorus.

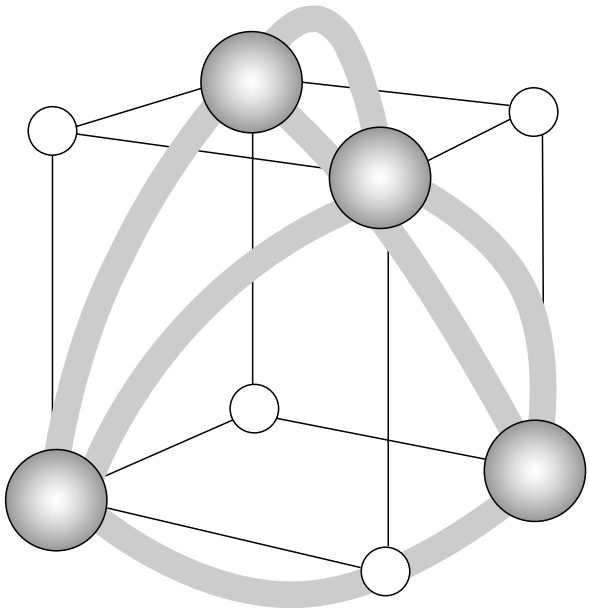

Figure 3 Three-dimensional distribution of valence shell electron density in the P_4 molecule. From Hart RR, Robin MB and Kuebler NA (1965) *Journal of Chemical Physics* 42: 3631–3638, with permission.

3p orbitals are involved, even though the bonds in P_4 are bent with $\angle P\text{–}P\text{–}P = 60^\circ$ and a strain energy of $96\,kJ\,mol^{-1}$, which accounts for its high reactivity. Electronic spectral studies suggest a resonance structure with strong π bonds. The regions of high electron density in the P_4 molecule are shown in Figure 3. The moment of inertia about any of its three major axes is $2.5 \times 10^{-40}\,g\,cm^2$. Thus, the P_4 molecule might be expected to have donor ability.

The only naturally occurring isotope, ^{31}P, has a nuclear spin $I = \frac{1}{2}h/2\pi$ and a large magnetic moment (1.13 NM) but no quadrupole moment. It is suitable for nuclear magnetic resonance (NMR) spectroscopy. In a 10-kG field ^{31}P resonates at 17.24 MHz. Phosphorus forms compounds with all elements except tin, bismuth, and the inert gases. It also reacts readily with heated aqueous solutions to give a variety of products. The bonding and stereochemistry of the phosphorus atom are varied, essentially due to empty d orbitals. It is known in at least 14 coordination geometries with coordination numbers up to 9, though the most frequently met coordination numbers are 3, 4, 5, and 6. Some typical geometries along with orbitals used in the formation of bonds are given in Table 1. Phosphorus has a significant tendency to catenation, forming a series of cyclic compounds $(RP)_n$ where $n = 3$–6, as well as some $R_2\,PPR_2$ type of compounds. (*See* **Spectroscopy**: Nuclear Magnetic Resonance.)

Forms in Foods

Plants need phosphate for healthy growth, especially for the development of roots, flowers, fruits, and seeds, though the requirement is less compared with that for nitrogen and potassium. All foods contain phosphorus in the form of the phosphate anion (PO_4^{3-}), and is consumed by living organisms as such. Very few natural compounds contain phosphorus in any other form. A summary of the types of biologically important phosphorus compounds is presented in Table 2. It is picked up from the soil, where it is present as organic and inorganic phosphates (soluble as well as insoluble) in nucleoproteins, nucleic acids, and the coenzymes nicotinamide adenine dinucleotide (NAD) 2-phosphate, ATP, and other high-energy phosphates. Organic phosphates include sugar phosphates such as glucose 6-phosphate (Figure 4), phospholipids, and pigments. Phosphate in the form of nucleotides serves as a source of a high-energy bond and performs an important function in conserving and providing bursts of metabolic energy.

Table 1 Types of phosphorus compounds and their molecular structure

Numbers of bonds	*Hybridization*	*Geometry*	*Shape*	*Examples*
3	sp^3 (with one lone pair)	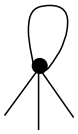	Pyramidal	PH_3
4	sp^3	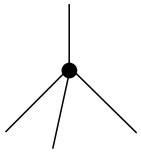	Tetrahedral	PH_4^+, $POCl_3$, $PO(OH)_3$ (d_π–p_π bonding in compounds with PO)
5	sp^3d	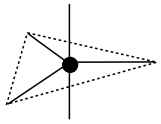	Trigonal bipyramid	PCl_5, PF_5, Ph_5P
6	sp^3d^2	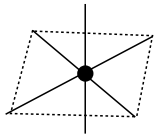	Octahedral	PF_6^-

Table 2 Type of biologically important phosphorus compounds and their functions

Type	*Example*	*Function*
Nucleic acids	DNA	Storage of genetic information
	RNA	Transcription of DNA and protein synthesis
Phospholipid	Phosphatidyl choline	Structural components of membranes
Bone salt	Hydroxyapatite	Bone structure and function
Phosphoproteins	Casein	Nutrient storage
	Glutaminase	NH_3 production; acid–base regulation
	Stathmin	Cell proliferation and differentiation
Sugar phosphate	Glucose-6-P	Glycolysis
Nucleotides	ATP	Energy transformation, molecular activation
	cAMP	Second messenger
	Uridine di-P glucose	Glycogen synthesis

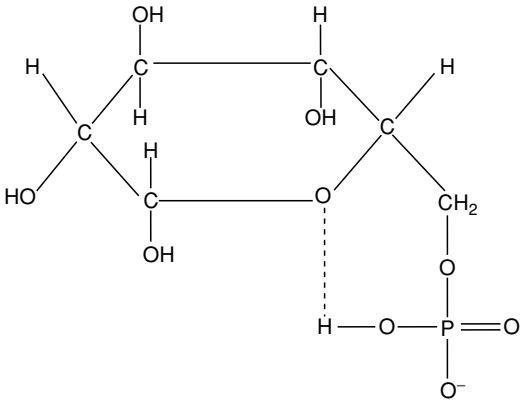

Figure 4

Most of the phosphorus in foods is in the form of organic phosphates that are digested in the intestines to form inorganic phosphates of sodium, calcium, and potassium. A special regulatory mechanism, the parathyroid hormone (PTH)–vitamin D–thyro calcitonin axis is involved in the control of both calcium and phosphorus balance. This hormonal axis controls the absorption rate in the gut, the excretion rate in the kidney, and the storage capacity of bones. Therefore, phosphorus is involved in a multitude of process in the entire life cycle, serving both a structural and catalytic function.

In egg yolk and fish, it occurs in the form of phospholipids, phosphatidylcholine, and phosphatidylethanolamine. In certain foods such as cereal grains and proteins from vegetable sources, 50–80% of the phosphorus occurs in the form of phytin, which is usually the calcium/magnesium salt of phytic acid (the hexaphosphate ester of inositol). Natural starch, particularly of potato, contains phosphoric acid as an ester (Figure 5). (*See* **Phospholipids**: Properties and Occurrence; **Phytic Acid**: Properties and Determination; **Starch**: Structure, Properties, and Determination.)

The phosphate cycle in water is controlled by the biocycle. The inorganic part of the cycle consists of HPO_4^{2-} in solution equilibrium with $H_2PO_4^-$, (PO_4^{3-}), and H_3PO_4. Planktonic algae can absorb inorganic phosphates. Some algae also have a method of

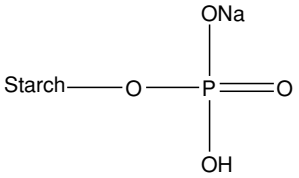

Figure 5

breaking down polyphosphates and even utilizing phospholipids.

Occurrence

Phosphorus is the 11th element in order of abundance in the earth's crust and occurs to the extent of 1120 p.p.m. Elemental phosphorus is not found free in nature but is combined with other elements in the form of inorganic minerals or as components of organic compounds. It occurs in various orthophosphate minerals, notably fluoroapatite [$3Ca_3(PO_4)_2{\cdot}CaF_2$] and also as hydroxyapatite [$Ca_3(PO_4)_2{\cdot}Ca(OH)_2$], chloroapatite [$Ca_3(PO_4)_2{\cdot}CaCl_2$]. About 90% of phosphate rock is used directly to make fertilizers, and the remainder is used to make phosphorus and phosphorus acids. World production of phosphorus is about 1.2 million tonnes per year but is declining. It is obtained by the reduction of calcium phosphate with C in an electric furnace at 1400–1500 °C. Sand is added to remove the slag (calcium silicate) and to drive off the P_4O_{10}, which is then reduced to P by C:

$$2Ca_3(PO_4)_2 + 6SiO_2 \rightarrow 6CaSiO_3 + P_4O_{10}$$
$$P_4O_{10} + 10C \rightarrow P_4 + 10CO.$$

However, in the human body, it is the sixth most abundant element, after O, H, C, N, and Ca. It is an essential constituent of every known tissue and cell in the body and accounts for about 1% of the body weight. Eighty-five percent of it remains in hard tissue and the rest in muscles and blood, as illustrated in Figure 6. Ordinarily, about 70% of phosphorus ingested in foods is absorbed by the body. As phosphate is a major constituent of all plant and animal cells, it is present in all natural foods, mostly in the organic form. Heavy concentrations of phosphorus are found in the meristematic regions of actively growing plants, where it is involved in the synthesis of nucleoproteins. Phospholipids, along with proteins, are significant constituents of cell membranes. Generally, phosphorus is more concentrated in seeds. In plant tissues and juices, it exists as phosphoric acid and the phosphate anion. In some citrus fruits, it occurs up to 2.7% as dihydrogen phosphate anion ($H_2PO_4^-$).

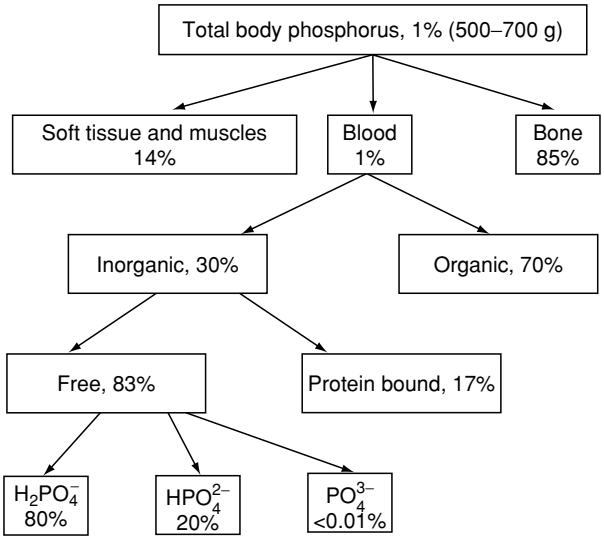

Figure 6 Distribution of phosphorus in the human body.

Table 3 Phosphorus contents in various foods

Food group	*Content (mg per 100 g)*
Cereal grains	100–500
Pulses	300–600
Leafy vegetables	20–500
Fruits	10–500
Nuts and oilseeds	10–1000
Condiments and spices	10–500
Fish and seafood	200–2000
Meat and poultry	150–800
Milk and dairy products	100–1000
Fats and edible oils	10–200
Sugars	10–400
Beverages	10–100

Diets rich in protein and calories also contain P in adequate amounts, regardless of the source of protein, carbohydrates, and fat. Milk and dairy products are the richest source of P in the diet, but P is widely available in other foods. The phosphorus contents in various foods are listed in Table 3. Semirefined brown sugar contains a high level of phosphorus, but refined sugar is completely stripped of the element. In the refinement of cereals, 50–60% of phosphorus is lost. The dietary intake of P can vary substantially with the types of foods consumed. Not all the P in our diet is of natural origin. Polyphosphates are popular additives in many meat products. These enhance the water-binding properties of meat proteins. Total food supplies in the UK and USA provide 0.8–1.5 g of phosphorus per person per day, of which about 10% is added artificially. During pregnancy and the lactation period, however, the requirement for phosphorus is increased. In muscle tissues, phosphorus is present as phosphocreatine (PCr), which synthesizes ATP. Its

deficiency in adults may occur with excessive use of alcohol, prolonged vomiting, liver disease, or hyperparathyroidism. The dietary content of P has been shown to regulate physiologically the serum concentration of PTH and thus indirectly the phosphate homeostasis. Refer to individual food.

Plants lacking phosphorus may develop necrotic areas on the leaves, petioles, or fruits; they may have a general overall stunted appearance, and the leaves may have a characteristic dark to blue–green coloration.

Properties of Phosphates

All phosphates are salts of oxyacids that contain a P = 0 group and at least one P–OH group that ionizes. Some species also have P–H group where the hydrogen atom is not ionizable. Phosphates of metal ions and other cations, mixed metal phosphates, and condensed phosphates are well known because of their commercial and technical importance. Many phosphates, especially long-chain polyphosphates, are known for their toxicity as they adversely affect the osmotic pressure of body fluids and prevent absorption of mineral nutrients. Phosphates are capable of interacting with many of the constituents of food systems, and inactivate metal ions, and are thus important in food processing.

Monosodium phosphate (NaH_2PO_4) is water-soluble and is used as a phosphatizing agent on steel surfaces. Its acidic property is used in effervescent laxative tablets and as a leavening agent in baking powder. Monopotassium phosphate (KH_2PO_4) crystals show a piezoelectric effect and are used in submarine sonar systems. Disodium and dipotassium phosphates are used as buffering agents to maintain pH. This property is used for stabilizing meat. These salts are also used as sequestering agents in the food industry. Sodium orthophosphate (Na_3PO_4) is highly alkaline and finds use in industrial hard surface cleaners. Its aqueous solution is a valuable constituent of scouring products, paint strippers, and grease saponifiers. Its complex with sodium hypochlorite [$(Na_3PO_4{\cdot}11H_2O)_4{\cdot}NaOCl$] releases active chlorine when wetted; this combination of scouring, bleaching and bacterial actions makes the adduct valuable in automatic dish washing powder formulations. Potassium orthophosphate (K_3PO_4) is used to regulate the rate of polymerization of styrene–butadiene rubber. Mono- and diammonium phosphates are used as fertilizers and nutrients. An important property of ammonium phosphate is a flame-retarding agent for cellulose materials. The action depends on its dissociation to ammonia and orthophosphoric acid when heated. Acid so generated catalyzes the decomposition of cellulose to char and smother the flame. Urea phosphate is generally used to flameproof cotton fabrics. A dilute solution of diammonium phosphate, with an initial pH of 7.85, upon boiling evolves ammonia, and the pH drops to 5.78 in 2 h. This property is used for the precipitation of colloidal dyes on wool fabrics. Dicalcium phosphates are used in pharmaceutical tablets as supplements. Natural phosphate minerals are all orthophosphates, the major ones being fluoroapatite; partly carbonated hydroxyapatite makes up the mineral part of teeth. These are important constituents of bones. Calcium orthophosphates are particularly important in fertilizer technology. (*See* **Leavening Agents; Stabilizers**: Types and Function.)

Many phosphate complexes of transition metal ions are known. Ce^{4+}, Th^{4+}, Zr^{4+}, U, and Pu form insoluble phosphates from fairly strong acid solution (3–6 M nitric acid). Condensed phosphates contain more than one phosphorus atom and P–O–P bonds with three main building units – the end unit (Figure 7), middle unit (Figure 8), and branching unit (Figure 9).These units can be readily distinguished by reactivity with water and ^{31}P NMR. These can also be incorporated into linear or cyclic polyphosphates. Linear polyphosphates are salts with the general formula $[P_n O_{3n+1}]^{(n+2)-}$ ($n = 2$–10) such as $M^1_4P_2O_7$

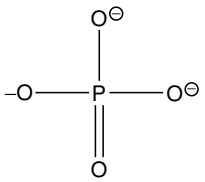

Figure 7

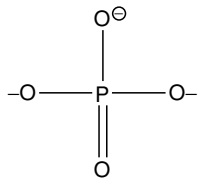

Figure 8

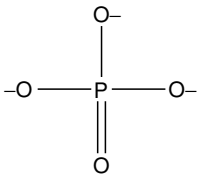

Figure 9

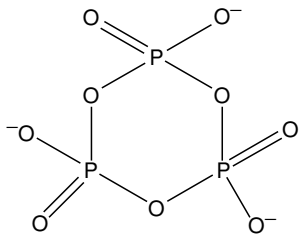

Figure 10

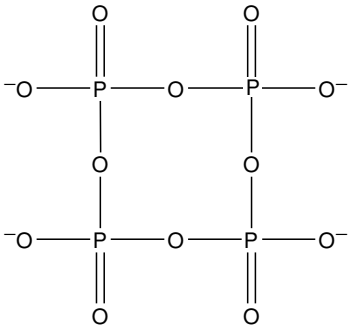

Figure 11

and $M_5^1P_3O_{10}$. Many polyphosphates with different chain lengths have been known. Disodium dihydrogen pyrophosphate, $Na_2^1H_2P_2O_7$ is mixed with $NaHCO_3$ and used in bread making to leaven the bread, that is to make it rise. They react and evolve CO_2 when heated together. Sodium pyrophosphate, $Na_4P_2O_7$, is mixed with starch and a flavoring agent to make instant pudding mixture. At one time, it was also added to soap powders and solutions as a water softener. $Ca_2P_2O_7$ is used as the abrasive/polishing agent in fluoride paste. Cyclic polyphosphates or metaphosphates are the salts with the general formula $[P_nO_{3n}]^{n-}$ with $n = 3$–7 such as $M_3P_3O_9$ (Figure 10) and $M_4P_4O_{12}$. The eight-membered ring of the $P_4O_{12}^{4-}$ (Figure 11) is puckered with equal bond lengths. Condensed phosphates form soluble complexes with many metals. These are usually prepared by dehydration of orthophosphates under various conditions of temperature (300–1200 °C) and also by appropriate hydration of dehydrated species. Chain phosphates are used as water softeners in industry. Polyphosphates aid in controlling the microbiological population on the surface of poultry meat. Some metaphosphates having an infinite chain length are also known, e.g., KPO_3.

Organic phosphates contain phosphate groups linked through OH groups of organic compounds (–C–O–P linkage) such as sugars. Large numbers of phosphate esters, $RO–PO(OH)_2$ are known. These occur in the form of mono-, di-, and triphosphoesters Each form has specific chemical properties leading to different biological functions. These are constituents of numerous highly active intracellular compounds. Release of free energy by hydrolysis of ATP provides the main source of energy for various metabolic processes and for muscle contraction. Intracellular phosphate is a regulator of enzymes in the glycolytic pathways. Some of the alkyl phosphorus compounds are of industrial importance, particularly for solvent extraction of metal ions from aqueous solutions. They extract metal ions by cation exchange and/or by solvation. Among these, di(2-ethylhexyl) phosphoric acid (DEHPA), tri-*n*-butyl phosphate (TBP), and tri-*n*-octylphosphine oxide (TOPO) have been used most extensively. Organic derivatives of fluorophosphoric acid $[FP=O(OH)_2]$ have promising properties as insecticides.

ANALYSIS

Phosphorus is generally detected on the basis of the reaction between orthophosphoric acid and the molybdate ion (MoO_4^{2-}), which gives a yellow-colored precipitate in a strongly acid solution.

Total Phosphorus

Any solution containing phosphorus is fumed with aqua regia almost to dryness, followed by heating with 1 M nitric acid, whence lower oxidation states are oxidized to the orthophosphate (PO_4^{3-}) form. The resultant solution can be used for the estimation of the total phosphorus by gravimetric, titrimetric, or spectrophotometric methods. Some of these have been recommended by the Association of Official Analytical Chemists for the analysis of total phosphorus in vegetables, fruits, cereals, and other foods. These are dried in a silica/platinum crucible and heated over a low Bunsen flame to volatilize organic matter. A 10% sodium bicarbonate solution is added and the contents evaporated to dryness. Oil may be burnt off at a lower temperature without smoking, and finally, ashing is carried out in a muffle furnace at 500 °C. The contents are dissolved in concentrated nitric acid and heated to dryness, and then dilute hydrochloric acid is added.

Gravimetric methods include the formation and weighing of phosphorus as ammonium phosphomolybdate, ammonium magnesium phosphate or pyrophosphate, and quinoline molybdophosphate. On addition of ammonium molybdate solution (12.5 g dissolved in 75 ml of water are slowly added to another solution containing 125 g of ammonium nitrate in 125 ml of water and 175 ml of nitric acid diluted to 500 ml) a yellow-colored precipitate is

obtained, which is filtered, dried at 105 °C, and weighed as $(NH_4)_3PO_4{\cdot}12\ MoO_3{\cdot}H_2O$. On further heating at 450 °C, a dark greenish blue complex, $P_2O_5{\cdot}24MoO_3$, is obtained, which may be weighed. In another method, the ammonium molybdate is replaced by a reagent containing sodium molybdate and quinoline so that quinoline molybdophosphate, $(C_9H_7N)_3[PO_4{\cdot}12MoO_3]$, is precipitated. Alternatively, magnesia reagent (magnesium chloride and ammonium chloride in ammoniacal solution) may be used as the precipitating reagent. The precipitate may be weighed as $Mg(NH_4)PO_4{\cdot}6H_2O$, or after heating in a muffle furnace at 1000 °C as $Mg_2P_2O_7$.

The most common titrimetric method consists of precipitation as $Mg(NH_4)PO_4.6H_2O$, filtration, washing, and dissolution of the precipitate in excess of dilute hydrochloric acid. An excess of standard ethylene diamine tetraacetic acid (EDTA) solution is added, and its pH is adjusted to 10. Excess EDTA is titrated with a standard solution of magnesium chloride or magnesium sulfate using Eriochrome black T as the indicator. Alternatively, the precipitate of ammonium molybdophosphate may be titrated with standard sodium hydroxide solution using phenolphthalein indicator.

Several spectrophotometric methods such as the molybdenum blue method and those using ammonium vanadate and 1-amino-2-naphthol-4-sulfonic acid have been used. A solution containing orthophosphate and molybdate ions condenses in acid solution to give phosphomolybdic acid, which upon reduction with hydrazinium sulfate produces a blue color due to molybdenum blue of uncertain composition. It exhibits a maximum at 620–630 nm. When phosphate, ammonium vanadate, and ammonium molybdate react, a bright yellow-colored complex of phosphovanadomolybdate is formed with a λ_{max} of 460–480 nm. In yet another method, molybdate reagent and 1-amino-2-naphthol-4-sulfonic acid solution (0.5 g with 30 g of sodium bisulfite and 6 g of sodium sulfite made up to 250 ml) are added. After standing for 10 min, the absorbance is measured at 650 nm. Determination of P concentration in plasma and other body fluids is the most common way to evaluate physiological status. The urinary concentration determines mainly how the kidneys handle phosphate. In such cases, the phosphate concentration is measured by the Fiske–Subbarrow colorimetric method. Commonly, the results are reported in units of mass per volume (mg per liter). In all these methods, a blank is always taken, and a standard calibration curve is constructed, from which the concentration of phosphorus is calculated. (*See* **Spectroscopy**: Overview.)

Phosphate is also determined nephelometrically. On adding molybdate strychnine reagent (in two parts: solution A of molybdenum trioxide in 5 M sulfuric acid, solution B of 1.6 g of strychnine sulfate in 500 ml of water; both solutions are mixed before use), a white-colored turbidity is obtained. The phosphorus content is determined from the calibration graph.

In recent years, several instrumental methods, *viz.* inductively coupled plasma (ICP) and/or direct current plasma (DCP) atomic emission spectrometry(AES), spark source mass spectrometry (SSMS), high-performance liquid chromatography (HPLC), flow injection analysis with continuous microwave oven digestion, an enzyme sensor system and X-ray fluorescence (XRF) have been used. Since the advent of body-imaging systems based on nuclear magnetic resonance measurement, it is now possible to detect phosphate in different tissues *in vivo* and determine changes with time under different physiological conditions. Thus, ^{31}P NMR can be used to detect levels of ADP, ATP, PCr, etc., and the metabolic activity of the cells can be determined. This method is used to study mainly the functioning of human liver and heart. Being most sensitive, it opens the most fascinating horizons for understanding the role of phosphorus in different metabolic pathways. With the availability of nuclear reactors, more sensitive methods employing radioactivity measurement have been employed. On thermal neutron irradiation, ^{32}P (β emitter, 1.71 MeV, half-life, 14.3 days) is formed by the (n, γ) reaction. A gas flow proportional counter is used with an aluminum absorber (thickness 27 mg cm^{-2}). Only a small sample size ($\sim$50 mg) is required. A derivative activation analysis method, and those employing Cerenkov and Bremsstrahlung counting have also been employed. Some radiotracer methods employing ^{32}P have been used for investigating the uptake of phosphate ions by plants through roots. (*See* **Chromatography**: High-performance Liquid Chromatography; **Mass Spectrometry**: Principles and Instrumentation.)

Organic Phosphorus

Phosphorus-containing organic compounds are first extracted into benzene or any other suitable solvent. An aliquot is transferred to a beaker, and the solvent is evaporated slowly on a water bath. The residue is heated with concentrated nitric acid and then with a small amount of potassium chlorate to dryness. After cooling, concentrated hydrochloric acid is added repeatedly until a clear solution is obtained. A 2.5% ammonium molybdate and potassium iodide/sodium carbonate solution is added. The flask is immersed in a steam bath for 15–20 min. After cooling, 0.5% sodium sulfite solution is added dropwise until the

iodine color disappears. Finally, the absorbance is measured at 650 nm.

Organophosphorus compounds can also be combusted in a Schoninger oxygen flask to give orthophosphate, which can be absorbed by either sulfuric or nitric acid. Phosphorus can then be determined spectrophotometrically. Recently, malonyl dihydrazide has been proposed for spectrophotometric extraction of organophosphorus compounds in vegetables. Chromatographic separations employing aluminum oxide and thin-layer chromatography (TLC) have also been used. Gas chromatography–mass spectrometry (GC–MS) has been suggested for the confirmation of organophosphorus pesticide residues in foods. A TLC enzymatic method has been suggested for the determination of organophosphorus compounds down to a detection limit of 10^{-10} g. (*See* **Chromatography**: Thin-layer Chromatography; High-Performance Liquid Chromatograpy.)

Other Forms

Phosphorus may also be found as phosphite PO_3^{3-}, which may be determined as mercurous chloride or ammonium magnesium phosphate. In the first case, an acid solution of phosphite reduces mercury(II) chloride solution (eqn 1), and mercury (I) chloride is weighed, the latter being directly proportional to the phosphite concentration.

$$2HgCl_2 + H_3PO_3 + H_2O \rightarrow Hg_2Cl_2 + H_3\ PO_4 + 2HCl \quad (1)$$

In the other case, phosphite is oxidized by nitric acid to phosphate, which is determined as $Mg(NH_4)PO_4{\cdot}6H_2O$ or as $Mg_2P_2O_7$.

Yet another form of phosphorus is hypophosphite (PO_2^{3-}). It is oxidized quantitatively by excess of ceric (IV) sulfate in sulfuric acid solution in 30 min at 60 °C; the excess Ce^{4+} is titrated with ferrous ammonium sulfate to a permanent red end point using ferroin indicator (eqn 2).

$$4Ce^{4+} + H_3PO_2 + 2H_2O \rightarrow 4Ce^{3+} + H_3PO_4 + 4H^+ \quad (2)$$

It is possible to determine both hypophosphite and phosphite by affecting complete oxidation to phosphate. These different forms may also be differentiated by ion chromatography.

See also: **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; **Leavening Agents**; **Mass Spectrometry**: Principles and Instrumentation; **Phospholipids**: Properties and Occurrence; **Phytic Acid**: Properties and Determination; **Spectroscopy**: Overview; Nuclear Magnetic Resonance; **Stabilizers**: Types and Function; **Starch**: Structure, Properties, and Determination

Further Reading

Baudler M (1987) Polyphosphorus compounds, new results and new perspectives. *Argewandte Chemie International Edition in English* 26: 419–441.

Berner YN (1997) Phosphorus, In: O'Dell BL and Sunde RA (eds.) *Handbook of Nutritionally Essential Elements*, p. 63–92. New York: Marcell Dekker.

Berner YN and Shike M (1989) Consequences of phosphate imbalance. *Annual Review of Nutrition* 8: 121–148.

Corbridge DEC (1985) *Phosphorus. An Outline of its Chemistry, Biochemistry and Technology*, 3rd edn. Oxford: Elsevier.

Ellinger RH (1975) Phosphates in food processing. In: Furia TA (ed.) *Handbook of Food Additives*, pp. 617–780. Cleveland, OH: CRC Press.

Goldwhite H (1981) *Introduction to Phosphorus Chemistry*. Cambridge: Cambridge University Press.

Kanazawa T (ed.) (1989) *Inorganic Phosphate Materials*. Tokyo: Kodansha.

Tietz NW (1976) Electrolytes. In: Tietz NW, Carraway WT,Freier EF, Kachmar JF and Rawausley HM (eds) *Fundamentals of Clinical Chemistry*. Philadelphia, PA: Saunders.

Toy ADF (1975) *Chemistry of Phosphorus*. New York: Pergamon Press.

Weginwar RG, Samudralwar DL and Garg AN (1989) Determination of phosphorus in biological samples by thermal neutron activation followed by β-counting. *Journal of Radioanalytical and Nuclear Chemistry Articles* 133: 317–324.

Physiology

J J B Anderson, University of North Carolina, Chapel Hill, NC, USA

Introduction

Phosphorus, an essential nutrient, is required for many different functions in body tissues, both intracellularly and extracellularly. This element, which exists in biological systems as phosphates, is used by cells to make structural molecules and outside of cells to make the crystals of bones and teeth; it serves as a component of intracellular regulatory molecules; it serves as a buffering component in both intra- and extracellular fluids, and it is an important factor in cellular energetics, especially in the mitochondria, where it is used to make most of the high-energy

bonds needed for cellular activities. Lastly, the metabolism of inorganic phosphate is closely linked to that of calcium, and this review will, therefore, also deal to some extent with calcium and its homeostatic control. Adequate phosphorus and calcium intakes are critical not only for skeletal growth, but also for growth and development of soft tissues, especially in neonates. (*See* **Bone**; **Calcium**: Physiology; **Cells**; **Energy**: Measurement of Food Energy.)

Phosphorus, in both the inorganic phosphate (Pi) and organic phosphate forms (Po), is abundant in nearly all foods traditionally consumed. This element is especially rich in animal products, such as meats, fish, poultry, eggs, milk, and other dairy products, but it also exists in cereal grains and most vegetables in good quantities. Because Pi ions are so readily absorbed across the small intestine, with an efficiency of roughly 65–75% in adults and with a somewhat higher efficiency in children, the prompt rise in blood Pi after a meal or snack influences calcium homeostasis, when Pi ions enter the blood from the gut, essentially unaccompanied by the more slowly absorbed calcium ions. An elevation in the plasma concentration of Pi tends to depress the serum calcium concentration, mainly Ca^{2+}, through the formation of a complex of the two ions, and this decline in Ca^{2+} stimulates parathyroid hormone (PTH) release. Although the specific mechanism through which Pi exerts this effect on plasma Ca^{2+} has not been fully established, many experimental data support the existence of this phenomenon.

Dietary phosphorus deficiency, although highly unlikely because of the abundance of this element in foods, contributes to low serum Pi concentration and thereby limits bone mineralization via osteoblasts and the total amount of bone mineral mass deposited in the skeleton. Furthermore, Pi deficiency increases bone turnover, which, during infancy, can lead to rickets.

Dietary Sources of Phosphorus

Although both organic and inorganic forms of phosphorus are widely distributed in foods, 75% or so of the phosphorus in foods is converted to Pi following the various digestive steps in the stomach and upper small intestine. The organic phosphorus usually remains associated with the fat-soluble dietary molecules that are absorbed without digestion of the phosphate groups from these molecules, as is the case with phosphatidylcholine (lecithin).

Phosphate additives in processed foods and in beverages can also make an important contribution to total phosphorus intakes. Concern has been raised about the quantity of phosphate added to cola-type soft drinks, which usually contain approximately 60 mg of phosphate per 12-oz. (336-g) container in the United States, but in reality, much greater amounts of phosphates are added to other processed foods commonly consumed in developed nations. Phosphate additives are used most by the food industry in baked goods, meats, cheeses, and milk products in the United States.

Mean phosphorus intakes of American women, according to recent surveys, range between 900 and 1200 mg per day, depending on age and caloric intake; older women consume less phosphorus than younger women, and active women consume more than sedentary individuals. Men usually consume closer to an average of 1500 mg per day, and their intakes also decline with age and with declining activity and caloric intake. The US Recommended Dietary Allowance (RDA) for each sex over 19 years of age was reduced in 1997 to 700 mg (Institute of Medicine) from 800 mg (NRC), whereas recommended calcium intakes were increased for adults until the end of life. Adequate intakes of phosphorus are readily achieved from foods in the USA, but those of calcium are not. The ratio of Ca:P in typical diets of American adult females is approximately 0.5:1, which raises concern among nutritionists because of the potential adverse effects of nutritional secondary hyperparathyroidism (see below). (*See* **Dietary Requirements of Adults**.)

Table 1 gives representative values of these same foods; the Ca:P ratio of these foods is also given as reference for later discussion of this ratio under Nutritional Secondary Hyperparathyroidism.

Intestinal Absorption of Phosphates

The net absorption of Pi is highly efficient; and the Pi efficiency is more than twice that of calcium absorption, which is usually stated as being between 25 and 30% in adults. The efficiency of Pi absorption by infants has been reported to be as high as 80–90%. Typical meals that contain representative items from all food groups, including dairy, have ratios of Ca:Pi approaching 0.7:1.0. If the actual amount of calcium in the meal is 350 mg, and the amount of phosphorus is 500 mg, then 75% of the phosphorus, or 375 mg, will be absorbed, most of it within the first postprandial hour, but only 30% or 105 mg of calcium will be maximally absorbed, and this process usually takes several hours to be completed. The net effect is that the rise in blood Pi concentration tends to depress the serum Ca^{2+}, perhaps through a reciprocal adjustment in the serum ion concentration product, i.e., $[Ca^{2+}] \times [Pi] = \text{constant}$, a mechanism proposed in the 1940s. Parathyroid hormone (PTH) secretion from the parathyroid glands responds to the depressed serum Ca^{2+} concentration as long as absorbed Pi continues

Table 1 Phosphorus and calcium composition of representative foods in a serving and the Ca:P ratio of these foods

Food item	*Weight (g)*	*Calcium (mg)*	*Phosphorus (mg)*	*Ca:P ratio*
Dairy				
Milk, whole	244	291	228	1.3:1
Milk, 2% fat	244	297	232	1.3
Milk, 1% fat	244	300	235	1.3
Milk, skim	245	302	247	1.2
Cheddar cheese	28	204	145	1.4
Egg, whole, poached	50	28	90	0.3
Meat, fish, and poultry				
Salmon, pink canned	85	167	243	0.7
Shrimp, canned	85	98	224	0.9
Beef, ground, lean with 10% fat	85	10	196	0.05
Liver, beef, fried	85	9	405	0.02
Pork roast, cooked	85	9	218	0.04
Chicken breast, cooked, boneless	85	18	210	0.09
Fruits				
Avocado, raw, skinless	251	23	30	0.8
Orange, whole, peeled	131	54	26	2.1
Pineapple, raw, dried	155	26	12	2.2
Raisins	14	9	14	0.6
Watermelon	926	30	43	0.7
Breads and cereals				
White bread, soft	25	21	24	0.9
Whole-wheat bread, soft	28	14	24	0.6
Bran flakes cereal	35	19	125	0.2
Corn flakes cereal	25	1	9	0.1
Shredded wheat cereal	25	11	97	0.1
Macaroni, cooked	130	14	85	0.2
Noodles (egg), cooked	160	16	94	0.2
Rice, white, instant	165	5	31	0.2
Spaghetti, cooked	130	14	85	0.2
Legumes, nut, seeds				
Almonds, shelled	130	304	655	0.5
Beans, Lima, cooked	190	55	293	0.2
Peas, black-eye, cooked	250	43	238	0.2
Peas, split, cooked	200	22	178	0.1
Vegetables				
Asparagus, cooked	145	30	73	0.4
Beets, cooked	100	14	23	0.6
Beet greens, cooked	145	144	36	4.0
Broccoli, cooked	180	158	112	1.4
Cabbage, cooked	145	64	29	2.2
Carrots, cooked	155	51	48	1.1
Corn, sweet, cooked	140	2	69	0.03
Kale, cooked	110	206	64	3.2
Pepper, sweet, cooked	73	7	12	0.6
Potatoes, baked	156	14	101	0.1
Squash, summer, cooked	210	53	53	1.0
Sweet potatoes, baked	114	46	66	0.7
Tomatoes, raw	135	16	3	5.3
Miscellaneous				
Tomato soup with milk	250	168	155	1.1
Tomato soup with water	245	15	34	0.4
Split-pea soup with water	245	29	149	0.2

Source: Home and Garden Bulletin No. 72, USDA.

to lower the Ca^{2+} concentration. (*See* **Hormones:** Thyroid Hormones.)

Pi absorption occurs by two major routes, transcellular and paracellular, throughout the small intestine and probably also the large intestine. The transcellular route is considered the more significant route, but not much is known about the paracellular pathway in terms of location in the small intestine or the quantitative contribution of this component. The transcellular route involves at least two distinct mechanisms

of entry at the brush-border membrane (mucosa) and probably as many as part of the exit step at the basolateral membrane (serosa). Much of the Pi absorption is considered passive because of cotransport with Na^+ or other cations. Because of the cotransport mechanism with Na^+ following a meal, Pi ions may be much more rapidly absorbed than if a separate and independent mechanism for Pi absorption *per se* were required. Understandings of the mechanisms of Pi absorption remain limited. The hormonal form of vitamin D, 1,25-dihydroxyvitamin D, increases Pi absorption via the intracellular route, but much less is known of this pathway than the vitamin D-mediated Ca^{2+} absorption. (*See* **Cholecalciferol**: Physiology.)

Some secretions of Pi ions into the gastrointestinal tract occur at every level, i.e., salivary glands, stomach, intestine, pancreas, liver, and large intestine. Thus, net Pi absorption represents the difference between total Pi absorption and endogenous Pi secretion. Pi ions not absorbed by the gut pass into the stools.

Blood Concentrations of Phosphates

The distribution in human blood serum is compared with the distribution of calcium in Table 2. Po in the lipid fraction of blood plasma generally represents a fairly stable value in individuals on diets with considerable amounts of phosphorus, i.e., typical Western diets, but it will decline if the diet becomes deficient or much lower in phosphorus. The reason for this change is not clear, but it may result from a homeostatic adaptation to the low-phosphorus intake through which membrane-bound Po groups are cleaved and released to the blood in order to keep the plasma Pi concentration at or near its 'set' level. The set level, which is determined genetically in a species, is regulated by various homeostatic mechanisms; in the case of serum Pi, several hormones, especially PTH, are involved in maintaining serum Pi at or near its set level.

The hormonal form of vitamin D – 1,25-dihydroxyvitamin D – is thought to enhance the intestinal absorption of Pi, as it does for Ca, but relatively little is known about the cellular mechanism of this action. For example, Pi-carrier proteins have been identified at the brush border membrane, but no intracellular Pi-binding proteins have been found after vitamin D treatment in animal models.

Table 2 Serum fractions of phosphates and calcium

Phosphate form	*Percentage of total*	*Calcium*	*Percentage of total*
Free HPO_4^{2-}	44	Free Ca^{2+}	48
Free $H_2PO_4^-$	10	Protein-bound	46
Protein-bound	12	Complexed	3
Cation-bound	34	Other	3

Adapted from Bringhurst FR (1989) Calcium and phosphate distribution, turnover, and metabolic actions. In: DeGroot LJ (ed.) *Endocrinology*, 2nd edn, vol. 2. Philadelphia, PA: WB Saunders.

Pi ions in the blood or extracellular fluids are distributed to all tissues in the body to meet cellular needs and to be taken up by the bone fluid compartment for incorporation in hydroxyapatite crystals. During tooth development, Pi ions are taken up by cells (odontoblasts, ameloblasts) and transferred into the extracellular compartment of the developing hydroxyapatite crystals.

Phosphate Homeostatic Mechanisms

The serum Pi concentration is not as rigorously regulated at any time during the life cycle as that of calcium. Several hormones are involved in Pi homeostasis. PTH, calcitonin, and the hormonal form of vitamin D are considered the major regulators, but many other hormones also affect Pi homeostasis, including insulin, glucagon, growth hormone, estrogens, adrenaline, and adrenal corticosteroids. (*See* **Hormones**: Steroid Hormones.)

An elevation of PTH acts to stimulate renal Ca^{2+} reabsorption while acting to block renal Pi reabsorption during this postprandial period, and at the same time, PTH is thought to be relatively inactive on bone cells because of the dominance of calcitonin. The net effect of PTH actions on different tissues is to try to conserve Ca^{2+} in the face of an elevated plasma Pi. However, if PTH remains elevated after the blocking effect of calcitonin decays, then an increased transfer of Ca^{2+} from the bone fluid compartment to blood and PTH-stimulated osteoclastic resorption together can slowly deplete bone when the dietary phosphorus intake greatly exceeds calcium on a long-term basis. This potential mechanism, then, is considered a highly likely way in which low-calcium diets contribute to osteopenia and subsequent fractures that characterize osteoporosis. The name given to this condition is nutritional 'secondary hyperparathyroidism' (see below). (*See* **Osteoporosis**.)

Regulation of the plasma Pi concentration is not well understood, but PTH and other hormones are involved through their renal, intestinal, and skeletal activities. High Pi intakes are handled readily and promptly by healthy individuals, primarily through increased renal excretion. The problem of a high-Pi intake coupled with low-calcium consumption in a typical dietary pattern, however, is handled at the expense of bone mineral through a chronic elevation of PTH and 1,25-dihydroxyvitamin D. Whether

1,25-dihydroxyvitamin D actually exerts a major enhancing effect on intestinal Pi absorption is not clear, but it does increase cellular transfer of Ca^{2+} from the gut lumen to the blood.

Other hormones that also have an influence on Pi homeostasis are important metabolic regulators, such as insulin following carbohydrate ingestion. Insulin stimulates glucose uptake in many extrahepatic tissues, and Pi moves into the cells along with glucose, probably in a 1:1 molar ratio. However, in situations when glucagon is elevated, glucagon enhances renal excretion of Pi, probably through a mechanism that blocks renal tubular Pi reabsorption. Calcitonin is also phosphaturic. Growth hormone and other anabolic hormones and local tissue factors generally stimulate Pi incorporation into organic structures, much like insulin, to meet the needs of growth, cell division, or structural requirements for tissue maintenance. (*See* **Carbohydrates**: Requirements and Dietary Importance.)

Another aspect of Pi homeostasis involves membrane-bound organic (Po) molecules, such as the phospholipids. Animal models fed low-phosphorus intakes have increased degradation of these Po membrane components, which release Pi to the blood and help to maintain the blood Pi concentration at its set level. These findings strongly suggest that membrane phospholipids have a role in Pi homeostasis.

Estrogen may also enhance Pi transfer into cells because plasma Pi becomes somewhat elevated in postmenopausal women and in women who have had an oophorectomy. This rise may also reflect bone turnover and Pi release to blood from the skeleton. Estrogen-replacement therapy in postmenopausal women causes a slight decline in serum Pi.

Excess Pi in cells can be stored in various organelles, such as mitochondria and endoplasmic reticulum, along with Ca^{2+} as calcium phosphates, which can be solubilized in times of need to be retrieved for cellular requirements.

Functional Roles of Phosphates

Phosphorus as Pi and/or Po exists in all tissues, both intracellularly and extracellularly. In extracellular sites, phosphorus exists in the hydroxyapatite crystals of bones and teeth and in phosphorylated proteins in diverse extracellular matrices. The phosphorylation of Type I collagen in bone may trigger the mineralization process. Phosphates also serve as a buffer system in blood and other extracellular fluids.

Intracellularly, phosphates also serve as important buffers, and Po is a component of numerous classes of molecules, including membranes, high-energy molecules, regulatory proteins, regulatory phospholipids, and nucleoproteins. In the phospholipids of nervous tissue, Po is a critical component of many diverse molecules. Phosphorylation of inositol to phosphatidylinositol and cleavage of inositol triphosphate represent an important regulatory mechanism in cells. In addition, phosphorylation of certain enzymes by a variety of protein kinases and the dephosphorylation of these same enzymes (proteins) by phosphatases is central to activation and inactivation of key regulatory enzymes controlling specific metabolic pathways in cells. Pi uptake by cells for the synthesis of regulatory peptides and regulatory phospholipids is also important in metabolically active cells. Thus, phosphorus is much more widely distributed within cells than is calcium, and it serves many diverse roles.

Pi uptake by cells is enhanced by insulin, but other hormones also increase the uptake of this anion, including estrogens, adrenalin, calcitonin, and many growth factors, such as insulin-like growth factor (IGF-1). Once in the cytosol, Pi ions are used for phosphorylating glucose and related intermediate molecules derived from glucose in a meal. In addition, Pi ions are shuttled across intracellular organelle membranes for use or storage. For example, in mitochondria, Pi ions are essential if oxidative phosphorylation is to be adequately coupled. Mitochondria also store roughly 20% of the cells' Pi as calcium salts. Similarly, the endoplasmic reticulum (ER) uses and stores Pi for phosphorylation of various proteins. Also, the ER contains approximately 30% of total cellular Pi for storage and use in phosphorylation of proteins and other molecules. The nucleus, Golgi complex, and lysosomes contain the remainder of the total Pi.

Phosphate in Health and Disease

The balance of phosphorus is determined by the difference between input and output, and it is kept remarkably constant in healthy individuals until late in the life cycle when lean tissue loss accelerates. The balance is clearly negative prior to death in ill patients because of the death of numerous cells without renewal. Pi absorption declines slightly late in life, i.e., sometime after the fifties, because of a reduced efficiency of absorption and because food energy intake also declines in the elderly. Thus, the overall input of Pi is lowered. Pi excretion also declines slightly in healthy older individuals, but it declines more so if renal function is seriously compromised. Normally, urinary excretion of Pi approximates 67% of Pi consumed in the diet, and this percentage holds throughout life in healthy individuals. Unabsorbed phosphorus makes up nearly all of the remainder of fecal Pi elimination, although sweat and skin losses

do contribute a small percentage to the total excretion of Pi.

Several issues of Pi homeostasis need further explanation because of their potential impact on health and disease in populations of developed nations.

Aging and Renal Function

Serum Pi concentration changes little in women with a healthy renal function, but it does increase when the renal function becomes compromised, though the percent reduction in GFR needed for this increase in serum Pi has not been established. Less is known of changes of the serum Pi concentration in men with increasing age, but the same relationship to renal function probably holds in men as in women. Because intestinal Pi absorption efficiency remains fairly constant in aging women, whereas calcium absorption efficiency declines within a decade or so following the menopause, it is possible that PTH secretion is increased in these women 10 or more years beyond the menopause without any other perturbation. Under these conditions, serum PTH concentrations rise, but they typically remain within the upper limit of normality. Severe reductions in renal function clearly elevate serum Pi concentration. Refer to Renal Secondary Hyperparathyroidism.

Nutritional Secondary Hyperparathyroidism

The major concern of nutritionally induced hyperparathyroidism, even if the PTH level remains within the upper range of normal, is the reduction of both bone mass and bone mineral density. A low Ca:P ratio from a diet that provides too little calcium and plentiful amounts of phosphorus may increase PTH in a fairly persistent manner, which in turn increases bone turnover. If bone formation cannot keep up with PTH-governed bone resorption, bone loss will follow (Figure 1). Although not fully established by research evidence, this scenario of a persistently elevated PTH in response to a low Ca:P ratio ($\ll 0.5{:}1$) is suspected of contributing first to osteopenia and then to osteoporotic bone that is more fragile and at increased risk of fracture (see Table 1 for examples of foods with low Ca:P ratios). Recent reports, for example, have shown that the adverse effects of a low Ca:P diet can be improved and largely overcome by simple calcium supplementation. Because the diets of so many adolescents and adults in the USA and probably adults in other technologically advanced nations with significant use of phosphate additives and inadequate calcium intakes have low Ca:P ratios, it is expected that rates of osteoporotic fractures, especially of hip fractures, will increase in the next several decades.

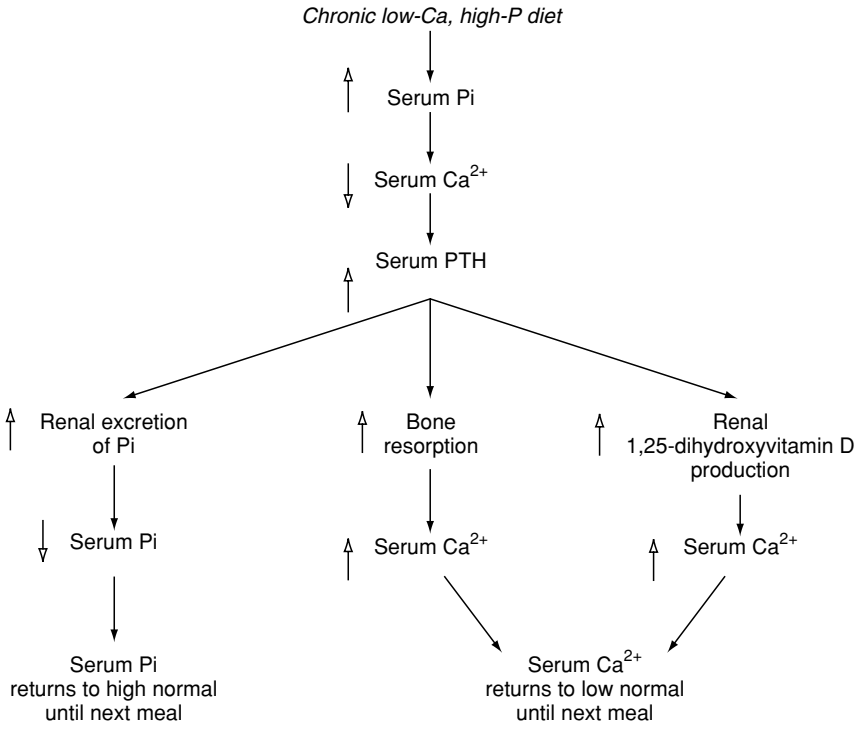

Figure 1 Illustration of steps in development of long-term nutritional secondary hyperparathyroidism: chronic low-calcium, high-phosphorus diet. Reproduced from Phosphorus, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Renal Secondary Hyperparathyroidism

When renal function becomes compromised to such an extent that creatinine, other nitrogenous metabolic products, and Pi are retained abnormally and excessively by the body, then several pathophysiological adaptations occur that have serious effects on health. One of the important adverse effects of the retention of Pi is the rapid and progressive loss of mineral mass. The chronic elevation of serum Pi causes a decline in serum Ca^{2+}, which triggers PTH secretion. The net effect is a constantly elevated PTH concentration that continues to act on bone tissue, i.e., resorption, to try to raise $[Ca^{2+}]$ to its homeostatic set level. Since Pi is also released from bone along with Ca^{2+} during resorption, the serum Pi concentration also increases. Because the kidneys cannot eliminate Pi adequately, $[Ca^{2+}]$ can never be raised to its set level, and bone tissue continues to be degraded as part of an unending vicious cycle.

Various dietary manipulations have been tried to control the loss of bone mass, but no one regimen has been very successful. Reductions of both dietary protein, especially animal protein, and phosphorus have been moderately successful in slowing the progress of chronic renal disease, but the diets are not very palatable or satisfying.

Conclusions

Pi metabolism is much more complex than that of calcium because of the many intracellular pathways that utilize Pi ions at one stage or another. The cytosolic utilization of Pi is closely linked with that of glucose for the formation of glucose 6-phosphate and for triglyceride synthesis through glycerol 3-phosphate formation, as well as with other molecules, during the postprandial period. Pi is utilized by cells for many diverse molecules, including regulatory peptides and phospholipids. Extracellular regulation of Pi is closely associated with that of calcium through PTH and other calcium-regulating hormones. Under typical dietary conditions of excessive phosphorus intake compared with calcium, i.e., low Ca:P ratio, nutritional secondary hyperparathyroidism and the long-term development of osteopenia are likely to result. Food fortification with calcium and calcium supplementation are common ways in which the low Ca:P ratio can be minimized, but individual behaviors aimed at selecting a diet higher in calcium will be needed to overcome the adverse ratio ($\ll 0.5$), despite calcium fortification and/or supplementation. Renal secondary hyperparathyroidism, a serious consequence of renal functional impairment, produces severe bone loss because of altered homeostatic regulation of Pi.

See also: **Aging – Nutritional Aspects**; **Bone**; **Calcium**: Physiology; **Carbohydrates**: Requirements and Dietary Importance; **Cells**; **Cholecalciferol**: Physiology; **Dietary Requirements of Adults**; **Energy**: Measurement of Food Energy; **Hormones**: Thyroid Hormones; Steroid Hormones; **Osteoporosis**

Further Reading

Akesson K, Lau K-H, Johnston P, Iperio E and Baylink DJ (1998) Effects of short-term calcium depletion and repletion on biochemical markers of bone turnover in young adult women. *Journal of Clinical Endocrinology and Metabolism* 83: 1921–1927.

Anderson JJB and Garner SC (eds) (1996) *Calcium and Phosphorus in Health and Disease*. Boca Raton, FL: CRC Press.

Anderson JJB, Sell ML, Garner SC and Calvo MS (2000) Phosphorus. In: Russell R *et al*. *Present Knowledge in Nutrition*, 7th edn. Washington, DC: International Life Sciences Institute.

Barger-Lux J and Heaney RP (1993) Effects of calcium restriction on metabolic characteristics of premenopausal women. *Journal of Clinical Endocrinology and Metabolism* 76: 103–107.

Bringhurst FR (1989) Calcium and phosphate distribution, turnover, and metabolic actions. In: DeGroot LJ (ed.) *Endocrinology*, 2nd edn, vol. 2. Philadelphia, PA: WB Saunders.

Brot C, Jorgensen N, Jensen LB and Sorensen OH (1999) Relationships between bone mineral density, serum vitamin D metabolites and calcium:phosphorus intake in healthy perimenopausal women. *Journal of Internal Medicine* 245: 509–516.

Calvo MS, Kumar R and Heath H III (1990) Persistently elevated parathyroid hormone secretion and action in young women after four weeks of ingesting high phosphorus, low calcium diets. *Journal of Clinical Endocrinology and Metabolism*, 70: 1340–1344.

Calvo MS and Park YM (1996) Changing phosphorus content of the U.S. diet: Potential for adverse effects on bone. *Journal of Nutrition* 126: 1168S–1180S.

Harnack L, Stang J and Story M (1999) Soft drink consumption among US children and adolescents: Nutritional consequences. *Journal of the American Dietetic Association* 99: 436–441.

Institute of Medicine, Food and Nutrition Board (1997) *Dietary Reference Intakes: Calcium, Phosphorus, Magnesium, Vitamin D and Fluoride*. Washington, DC: National Academy Press.

Karkkainen M and Lamberg-Allardt C (1996) An acute intake of phosphate increases parathyroid hormone secretion and inhibits bone formation in young women. *Journal of Bone and Mineral Research* 11: 1905–1912.

McKane WR, Khosla S, Egan KS *et al*. (1996) Role of calcium intake in modulating age-related increases in parathyroid function and bone resorption. *Journal of Clinical Endocrinology and Metabolism* 81: 1699–1703.

National Research Council (1989) *Recommended Dietary Allowances*, 10th edn, pp. 184–187. Washington, DC: National Research Council, National Academy Press.

Slatopolsky E, Dusso A and Brown A (1999) The role of phosphorus in the development of secondary hyperparathyroidism and parathyroid cell proliferation in chronic renal failure. *American Journal of Medical Sciences* 317: 370–376.

USDA (1978) *Nutritive Value of Foods. Home and Garden Bulletin* No. 72. Washington, DC: US Department of Agriculture.

Phylloquinone *See* **Vitamin K**: Properties and Determination; Physiology

Physical Properties of Food *See* **Rheological Properties of Food Materials**

PHYTIC ACID

Contents

Properties and Determination

U Konietzny, Karlsruhe, Dettenheim-Liedolsheim, Germany
R Greiner, Centre for Molecular Biology, Federal Research Centre for Nutrition, Karlsruhe, Germany

Introduction

The proper chemical designation for phytic acid is *myo*-inositol(1,2,3,4,5,6)hexakisphosphoric acid. Salts of this acid, designated as phytates, are found in plants, animals and soil. Phytate has been considered as an antinutrient due to its inhibitory effect on the bioavailability of essential dietary minerals. During food processing and digestion, phytate can be dephosphorylated to produce degradation products, such as *myo*-inositol pentakis-, tetrakis-, tris-, bis-, and monophosphates. Besides the adverse effects of phytate and other highly phosphorylated *myo*-inositol phosphates on mineral bioavailability, some novel metabolic effects of phytate and some of its degradation products have been recognized. Certain *myo*-inositol phosphates have been suggested to have positive effects on heart disease by controlling hypercholesterolemia and atherosclerosis and to prevent renal stone formation. The most extensively studied positive aspect of *myo*-inositol phosphates is their potential for reducing the risk of colon cancer. Furthermore, much attention has been focused on *myo*-inositol with fewer than six phosphate residues, since some of these compounds have been shown to play an important part as intracellular second messengers and some have shown important pharmacological effects, such as the prevention of diabetes complications and antiinflammatory effects. The position of the phosphate groups on the *myo*-inositol ring is therefore of great significance for their physiological function. Thus, it is important to have reliable techniques available to determine qualitatively and quantitatively *myo*-inositol phosphates not only by the number of phosphate groups, but also by the position of the phosphate groups on the *myo*-inositol ring. (*See* **Plant Antinutritional Factors:** Characteristics.)

Structure, Occurrence, and Biological Significance

Phytate is a *meso* compound and consequently possesses a plane of symmetry with either five equatorial and one axial phosphate groups (5-eq/1-ax) or five axial and one equatorial phosphate groups (5-ax/1-eq: Figure 1). The carbon bearing the single axial or equatorial phosphate group is numbered C2 and the other ring carbons can be numbered C1–C6 from a C1 atom either side of C2, proceeding around the ring in a clockwise or counterclockwise fashion. Some less phosphorylated *myo*-inositol derivatives are optically active (Table 1). Their absolute configuration must be clearly defined. According to convention, a counterclockwise numbering gives rise to *myo*-inositol phosphates with a D-prefix and a clockwise numbering to *myo*-inositol phosphates with an L-prefix. The choice of prefix is normally determined by giving preference to that which results in the lowest numbering of substituents (Figure 2). The predominant conformation of the *myo*-inositol phosphates depends on the specific *myo*-inositol phosphate, pH value, type of cations present, and ionic strength. At pH values above 9.5, phytate exists exclusively in the 5-ax/1-eq conformation, whereas with *myo*-inositol pentakisphosphates a small amount of the 5-eq/1-ax conformer is found in equilibrium with the predominant 5-ax/1-eq conformer. In *myo*-inositol phosphates with fewer than five phosphate residues, the *myo*-inositol ring appears to have a conformation in which only the phosphate group at C2 is axially oriented. Below pH 9.5 the 5-eq/1-ax conformer is predominant with all *myo*-inositol phosphates. The binding of some cations such as Cu^{2+} and Ca^{2+} is proposed to occur mainly via phosphate groups at the equatorial position of phytate; other cations such as

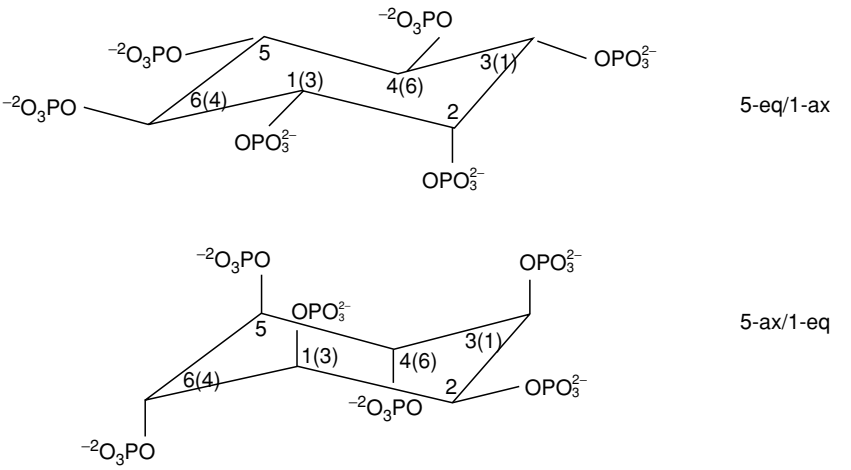

Figure 1 Possible chair conformations of phytate.

Table 1 *myo*-inositol phosphate isomers

No. of phosphate residues	*No. of isomers*	*No. of enantiomeric pairs*	myo-*inositol phosphate isomers*
6	1	0	*I(1,2,3,4,5,6)P$_6$*
5	6	2	*D-I(1,2,3,4,5)P$_5$*/*L-I(1,2,3,4,5)P$_5$* *D-I(1,2,4,5,6)P$_5$*/*L-I(1,2,4,5,6)P$_5$*, I(1,2,3,4,6)P$_5$, *I(1,3,4,5,6)P$_5$*
4	15	6	D-I(1,2,3,4)P$_4$/L-I(1,2,3,4)P$_4$, D-I(1,2,4,5)P$_4$/L-I(1,2,4,5)P$_4$, *D-I(1,2,5,6)P$_4$*/L-I(1,2,5,6)P$_4$, D-I(1,2,4,6)P$_4$/L-I(1,2,5,6)P$_4$, *D-I(1,3,4,5)P$_4$*/L-I(1,3,4,5)P$_4$, *D-I(1,4,5,6)P$_4$*/*L-I(1,4,5,6)P$_4$*, I(1,2,3,5)P$_4$, *I(1,3,4,6)P$_4$*, *I(2,4,5,6)P$_4$*
3	20	8	D-I(1,2,4)P$_3$/L-I(1,2,4)P$_3$, D-I(1,2,5)P$_3$/L-I(1,2,5)P$_3$, *D-I(1,2,6)P$_3$*/L-I(1,2,6)P$_3$, *D-I(1,3,4)P$_3$*/L-I(1,3,4)P$_3$, D-I(1,3,6)P$_3$/L-I(1,3,6)P$_3$, *D-I(1,4,5)P$_3$*/L-I(1,4,5)P$_3$, D-I(1,4,6)P$_3$/*L-I(1,4,6)P$_3$*, *D-I(2,4,5)P$_3$*/L-I(2,4,5)P$_3$, I(1,2,3)P$_3$, I(1,3,5)P$_3$, I(2,4,6)P$_3$, I(4,5,6)P$_3$
2	15	6	D-I(1,2)P$_2$/L-I(1,2)P$_2$, *D-I(1,4)P$_2$*/*L-I(1,4)P$_2$*, D-I(1,5)P$_2$/L-I(1,5)P$_2$, D-I(1,6)P$_2$/*L-I(1,6)P$_2$*, *D-I(2,4)P$_2$*/*L-I(2,4)P$_2$*, *D-I(4,5)P$_2$*/L-I(4,5)P$_2$, *I(1,3)P$_2$*, I(2,5)P$_2$, I(4,6)P$_2$
1	6	2	*D-I(1)PIL-I(1)P*, *D-I(4)P*/L-I(4)P, *I(2)P*, *I(5)P*

Myo-inositol phosphate isomers found in nature are indicated in italics.

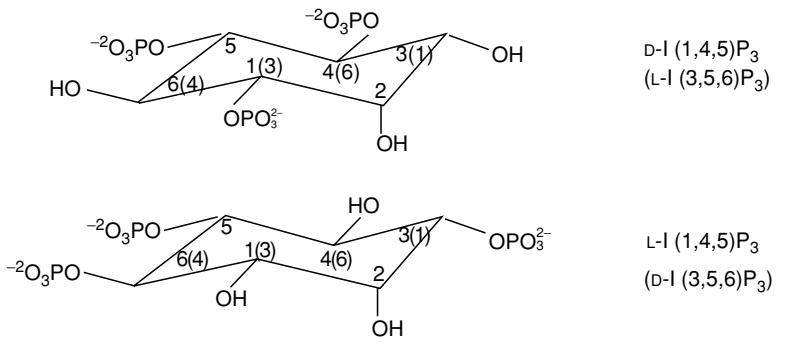

Figure 2 Absolute configuration of D- and L-*myo*-inositol(1,4,5)trisphosphate.

Zn^{2+}, Mn^{2+}, and the alkali metal ions seem to have preference for phosphate groups at the axial position. A single-crystal X-ray analysis of the sodium salt showed the 5-ax/1-eq conformer. In solution, the inositol ring also occurs over a wide pH range in a chair conformation in which five phosphate residues are arranged axially and only the phosphate at C2 is equatorially-oriented. Raman data indicate that alkali metal ions preferentially bind to, and thus stabilize, the 5-ax/1-eq phytate conformer in the order $Li^+ \sim Na^+ > Cs^+$. In contrast, the Raman spectrum of solid Ca_6-phytate is characterized by the 1-ax/5-eq conformer.

Plants

Phytate is ubiquitous among plant seeds and/or grains, comprising 0.5–5% (w/w). It is primarily present as a salt of the mono- and divalent cations K^+, Mg^{2+}, and Ca^{2+} and accumulates in the seeds during the ripening period. In dormant seeds phytate represents 60–90% of the total phosphate. Only a very small part of the *myo*-inositol phosphates exists as *myo*-inositol penta- and tetrakisphosphate of unknown isomeric state. The function of these high phytate concentrations in plant seeds is unclear. It has been suggested that phytate may serve as a store of phosphate, of cations, of the cell wall glucuronate precursor, of high-energy phosphoryl groups and, by chelating free iron, as a potent natural antioxidant. In amoeba, two diphospho-*myo*-inositol pentakisphosphate isomers and one bis(diphospho)-*myo*-inositol tetrakisphosphate isomer are present in concentrations exceeding that of phytate and thus may in fact represent a compact store of high-energy phosphate.

Until now only little is known of the pathway of phytate synthesis in either the plant or animal kingdom. A study in the slime mould *Dictyostelium discoideum* established that phytate synthesis from *myo*-inositol proceeds via Ins(3)P, Ins(3,6)P_2, Ins(3,4,6)P_3, Ins(1,3,4,6)P_4, and Ins(1,3,4,5,6)P_5. Early studies of phytate synthesis in plants led to the proposal that phytate synthesis from Ins(3)P was mediated by phosphoinositol kinase(s) via a series of undefined *myo*-inositol phosphates. Recently the first description of the synthetic sequence to phytate in the plant kingdom was given. From the identities of *myo*-inositol phosphates found in duckweed (*Spirodela polyrhiza* L.), at a development stage associated with massive accumulation of phytate, it was concluded, that synthesis of phytate from *myo*-inositol proceeds according to the sequence D-I(3)P, D-I(3,4)P_2, D-I(3,4,6)P_3, D-I(3,4,5,6)P_4, I(1,3,4,5,6)P_5. An unanswered question that relates to the pathway of phytate synthesis in plants concerns the source of D-I(3)P. Two enzyme activities that are capable of synthesizing D-I(3)P have been identified. These are *myo*-inositol phosphate synthase (EC 5.5.1.4), which converts glucose-6-phosphate, the ultimate source of *myo*-inositol in plants, to D-I(3)P, and *myo*-inositol kinase (EC 2.7.1.64), which converts *myo*-inositol to D-I(3)P. The spatial and temporal distribution and the relative contribution of these two enzymes to phytate synthesis via D-I(3)P are unclear.

During germination, phytate is rapidly hydrolyzed in a stepwise manner by phytate-specific phosphohydrolases (phytases, EC 3.1.3.8, EC 3.1.3.26) or a concerted action of phytases and other phosphatases to supply the nutritional needs of the plant without an accumulation of less phosphorylated *myo*-inositol intermediates. Neither the isomer structure of these intermediates nor the final product of phytate degradation is known to date. From *in vitro* investigations on the stereospecificity of phytate hydrolysis by purified phytases from cereals it was established that these enzymes dephosphorylate phytate in a stereospecific way by sequential removal of phosphate groups via D-I(1,2,3,5,6)P_5, D-I(1,2,5,6)P_4, D-I(1,2,6)P_3, and D-I(2,6)P_2 to finally I(2)P. Moreover, the phytases from bacteria and fungi investigated for phytate degradation release five of the six phosphate groups, and the end product was identified as I(2)P. Thus, the

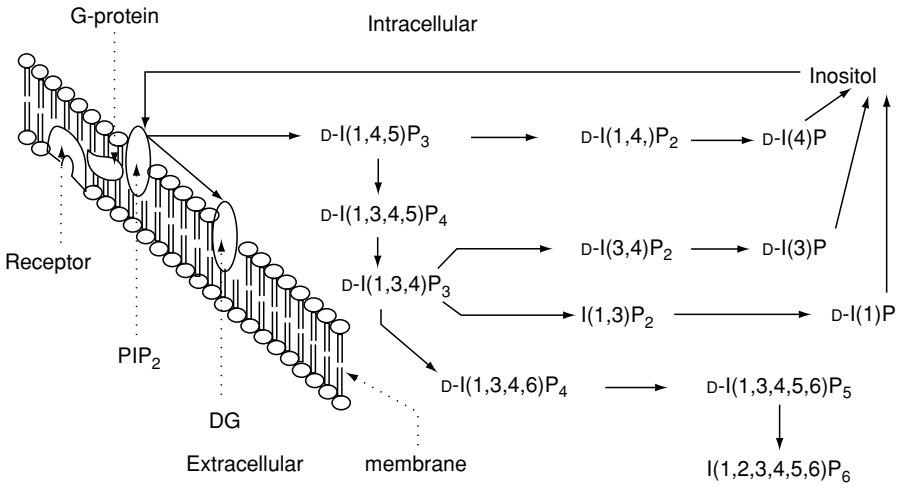

Figure 3 The phosphatidylinositol pathway. PIP_2, phosphatidylinositol 4,5-bisphosphate.

phosphate at C2 seems to be particularly resistant to enzymatic cleavage. (*See* **Enzymes**: Functions and Characteristics.)

Soil

In the soil, phytate as well as less phosphorylated *myo*-inositols are found. Their biological sources are unknown.

Animals

In animal tissue, a considerable number of *myo*-inositol phosphates containing one to six phosphate residues have been found. In most tissues stimulation of the phosphatidylinositol pathway (Figure 3) causes release of D-*myo*-inositol(1,4,5)trisphosphate, which is subsequently metabolized to a wide range of *myo*-inositol phosphate isomers. For D-I(1,3,4)P_3, metabolism is complex, involving both dephosphorylation via D-I(1,4)P_2 to D-I(4)P and phosphorylation to D-I(1,3,4,5)P_4, this latter compound being eventually degraded to D-I(1)P or D-I(3)P. The full significance of this complex metabolism is not clear, but there is now evidence that certain products of the phosphoinositide metabolism play second messenger roles in most cells. D-I(1,4,5)P_3 and D-I(1,3,4,5)P_4 bind to specific receptors and regulate Ca^{2+} release from or movement between intracellular Ca^{2+} stores. D-I(1,3,4,5)P_4 is also the starting point for metabolic pathways generating other *myo*-inositol tetrakisphosphate isomers as well as higher phosphorylated *myo*-inositols. There are no known functions for these higher phosphorylated *myo*-inositols; these metabolites comprise the bulk of *myo*-inositol phosphate content in mammalian cells, but evidence for their association with cell signaling was recently suggested.

D-*myo*-inositol(1,3,4,5,6)pentakisphosphate was also found in the erythrocytes of birds, turtles, and frogs. The functional importance of this compound as a key regulator of oxygen affinity becomes evident with the discovery that erythrocytes of adult birds contain virtually no 2,3-bisphosphoglycerate, the potent allosteric regulator of hemoglobin in mammalian erythrocytes.

Chemical Properties

Numerous studies have been made on the protonation constants of phytate, but the results are often conflicting. This could be due to the fact that the protonation constants of phytate to a large extent are dependent on the ionic strength of the medium. Phytic acid contains six strong acid groups which are completely dissociated in solution (pK_a 1.1–3.2), three weak acid protons (pK_a 5.2–8.0), and three very weak acid protons (pK_a 9.2–12). There unusually high pK_a values for the second protonization step seem to be due to intramolecular hydrogen bonding between the *syn*-axial phosphate residues at C1 and C3 as well as C4 and C6. These pK_a values imply that phytic acid will be strongly negatively charged over a wide pH range and have immense potential for binding positively charged species, such as cations or proteins. Free phytic acid is an unstable compound and decomposes to yield lower *myo*-inositol phosphates and orthophosphate. It is generally isolated as a sodium or calcium salt. In its free form, phytic acid is a light-yellow to light-brown syrupy liquid, soluble in polar solvents (water, methanol, ethanol, 2-propanol, acetone tetrahydrofuran, dimethyl sulfoxide, dimethyl formamide), but insoluble in nonpolar solvents (benzene, toluene, hexane, chloroform). In

contrast to free phytic acid, their salts are very stable compounds. The phosphate groups can be removed hydrolytically by enzymes or acid/heat to yield a large number of homologs and positional isomers ranging from *myo*-inositol mono- to pentakisphosphates (Table 1). In spite of the considerable number of isomers identified *in vivo*, they still represent only a small percentage of the number possible in theory. Above pH 5, there is almost no decomposition of phytate at 100 °C within 10 h. The rate of acid hydrolysis is low – as in other orthophosphoric esters it reaches a maximum at pH 4, e.g., 27% of phytate is cleaved after 6 h at 100 °C, and even the use of strong acids leads to an only moderate increase in the rate of hydrolysis. In 5 mol l^{-1} HCl 47% of phytate is cleaved after 6 h at 100 °C. Final products of acid hydrolysis are *myo*-inositol and *myo*-inositol(2)monophosphate. As with enzymatic cleavage, the axial phosphate at C2 seems to be particularly resistant to hydrolysis. Complete decomposition of phytate was achieved with 3 mol l^{-1} H_2SO_4 at 165 °C for 4 h. Under acidic conditions and higher temperatures in particular, monophosphorylated *cis*-diol groups of the inositols, in competition with hydrolysis, exhibit the phenomenon of phosphate migration. Thus, for example, D-*myo*-inositol(1)phosphate can yield a mixture of D-*myo*-inositol(1)phosphate, *myo*-inositol(2)-phosphate and L-*myo*-inositol(1)phosphate. For this migration, an intermediate formation of cyclic phosphodiesters is essential which is only sterically favored in *cis*-diol groups.

Phytate–Cation Interaction

Phytate forms complexes with numerous divalent and trivalent cations. The stability and solubility of the cation–phytate complexes depend on the specific cation, pH value, phytate-to-cation molar ratio, and the presence of other compounds in the solution. Phytate has six reactive phosphates and meets the criterion of a chelating agent. In fact, a cation can complex not only within one phosphate group or between two or more phosphate groups of one phytate, but also between two or more phytate molecules (Figure 4).

Studying the solubility and relative stability of various phytate–metal complexes by potentiometric titration, the following order of stability at pH 7.4 was found: $Cu^{2+} > Zn^{2+} > Ni^{2+} > Co^{2+} > Mn^{2+} > Fe^{3+} > Ca^{2+}$. Most phytates tend to be more soluble at lower than at higher pH values. The pH value below which the solubilities increase is about 5.5–6.0 for calcium, 7.2–8.0 for magnesium, and 4.3–4.5 for zinc phytate. In contrast, ferric phytate is insoluble at pH values in the 1–3.5 range at equimolar Fe^{3+}-to-phytate molar ratios. Solubility increases above pH 4, reaching 50% at pH 10. When Fe^{3+}-to-phytate molar ratio is increased to 3.5:1, there is increased solubility below pH 2, reaching a maximum of 90% at pH 1.5 and lower solubility at pH values above pH 4. By forming a complex with Fe^{3+} that lacks iron-coordinated water and thus is unable to catalyze the formation of hydroxyl radicals in the Fenton reaction, phytate is a good antioxidant.

Another important fact is the synergistic effect of secondary cations, among which Ca^{2+} has been most prominently mentioned. Two cations may, when present simultaneously, act together to increase the quantity of phytate precipitation. For example, Ca^{2+} enhanced the incorporation or adsorption of Zn^{2+} into phytate by formation of a Ca-Zn phytate. The effect of Ca^{2+} on the amount of Zn^{2+} coprecipitated with phytate is dependent on Zn^{2+}-to-phytate molar ratios. For high Zn^{2+}-to-phytate molar ratios, Ca^{2+} displaces Zn^{2+} from phytate-binding sites and increases its solubility. The amount of free Zn^{2+} is directly proportional to the Ca^{2+} concentration. For low Zn^{2+}-to-phytate molar ratios, Ca^{2+} potentiates the precipitation of Zn^{2+} as phytate. The higher the Ca^{2+} level, the more extensive the precipitation of the

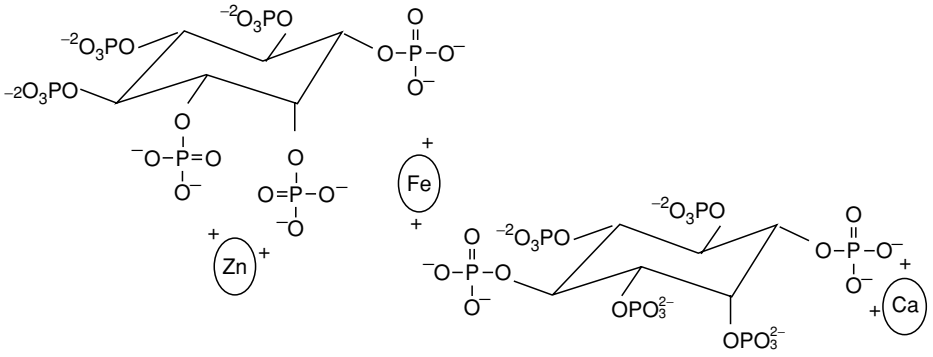

Figure 4 Phytate–cation interaction. DG, diacylglycerol.

ions. Mg^{2+} also has been shown *in vitro* to potentiate the precipitation of Zn^{2+} in the presence of phytate; however, Mg^{2+} has been found to exert a less pronounced effect on Zn^{2+} solubility than Ca^{2+}.

The knowledge about the interaction of the lower *myo*-inositol phosphates with different cations is limited. Recent studies have shown that *myo*-inositol pentakis-, tetrakis-, and trisphosphates have a lower capacity to bind cations (Ca^{2+}, Cu^{2+}, Zn^{2+}, Fe^{2+}, Fe^{3+}) at pH values in the 5–7 range. The capacity to bind cation was found to be a function of the number of phosphate groups on the molecule. The cation-*myo*-inositol phosphate complexes seem to become more soluble as the number of phosphate groups decreases. There is also some evidence for weaker complexes when phosphate groups are removed from phytate. Furthermore, the binding affinity of cations to *myo*-inositol phosphates has been shown to be affected by the orientation of the phosphate groups.

Phytate–Protein Interaction

Phytate interactions with proteins are pH-dependent. Phytate is known to form complexes with proteins at both acidic and alkaline pH (Figure 5). At pH values below the isoelectric point of the protein, the anionic phosphate groups of phytate bind strongly to the cationic groups of the protein to form insoluble complexes that dissolve only below pH 3.5. The α-NH_2 terminal group, the ε-NH_2 of lysine, the imidazole group of histidine, and guanidyl group of arginine have been implicated as protein-binding sites for phytate at low pH values. These low-pH protein–phytate complexes are disrupted by the competitive action of multivalent cations.

Above the isoelectric point of the protein, both protein and phytate have a negative charge, but in the presence of multivalent cations soluble protein–cation–phytate complexes occur. The major protein-binding site for the ternary complex appears to be the unprotonated imidazole group of histidine. The ionized carboxyl group of the protein are also suggested sites. These complexes may be disrupted by high ionic strength, high pH (> 10), and high concentrations of the chelating agents.

Protein–phytate complexation may effect changes in protein structure that can decrease enzymatic activity, solubility, and vulnerability to attack by proteolytic enzymes. Phytate has been shown to reduce the activity of lipase, α-amylase, pepsin, trypsin, and chymotrypsin *in vitro*. The inhibitory effect increases with the number of phosphate groups per *myo*-inositol molecule and the *myo*-inositol phosphate concentration. (*See* **Protein:** Interactions and Reactions Involved in Food Processing.)

Application

Phytate has found industrial application, including uses in the food industry (Table 2). The focus of research on phytates includes occurrence and functions in plant seeds, nutritional significance, preservative applications in food technology, and potential medical and industrial uses. (*See* **Preservatives:** Food Uses.)

Determination

The measurement of *myo*-inositol phosphates in any material requires an initial extraction. The reagents most commonly used to extract *myo*-inositol phosphates from foodstuff and biological samples include 3% trichloroacetic acid and 2.4% hydrochloric acid. Since *myo*-inositol phosphates do not have a

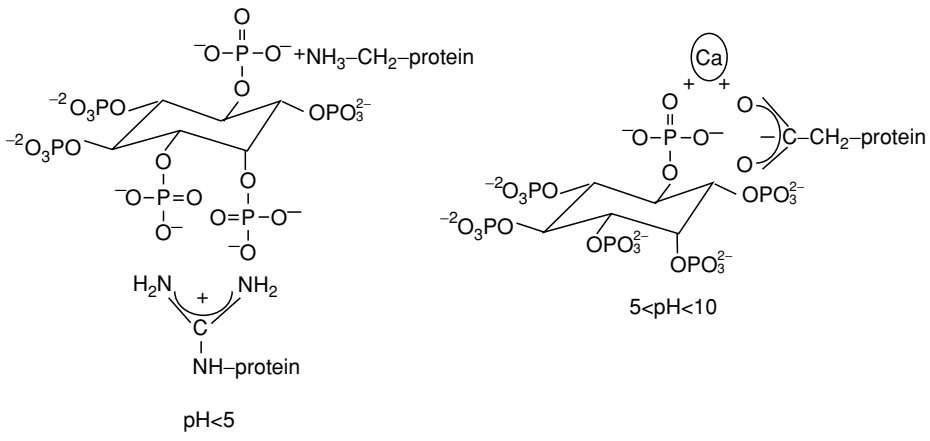

Figure 5 Phytate–protein interaction.

Table 2 Application of phytate

Action	*Application*
Metal chelation	Prevention of color and quality changes in processed agricultural (chestnut, bean sprouts, pickles, asparagus, etc.) and fishery (tuna, clams, shrimps, crabs, etc.) products
	Removing metal ions from wine
	Rust-proofing and dissolving-out prevention inside cans
	Prevention of oxidation in oil/water emulsion-type food such as cream, dressings, butter, chesses, soups
	Additive for etching solution for offset printing
	Anticorrosion agent for paints, antifreezes, and metal surfaces (steel, tin, aluminum, iron)
	Stabilizer for perfumes and cosmetics
	Antioxidant for industrial oils and greases
pH control	Prevention of quality changes by controlling pH value
Fermentation promoter	Improvement of product yield and quality by promoting the growth of microorganisms such as lactic acid bacteria and yeasts (fermented food, antibiotics, methanol, etc.)

characteristic absorption spectrum, nor can they be identified using specific colorimetric reagents, the determination of these compounds has remained a persistent problem.

Qualitative Separation Methods

Qualitative separation and detection of *myo*-inositol phosphates have been developed in the 1950s and 1960s. Paper chromatography has been shown to be useful for separating *myo*-inositol phosphates by the number of phosphate groups. *Myo*-inositol mono- to hexakisphosphate could also be resolved relatively rapidly by electrophoresis. Thin-layer chromatography, even if successfully applicable, has not been widely adopted for the separation of *myo*-inositol phosphates.

Quantitative Separation Methods

Precipitative methods Quantitative methods for determining phytate often employ the addition of a controlled amount of Fe^{3+} to an acidic sample extract to precipitate the phytate. Phytate is subsequently estimated either by determining the phosphate, inositol, or iron content of the precipitate (direct method), or by measuring the excess iron in the supernatant (indirect method). The indirect methods are generally more convenient and reproducible, because the stoichiometric ratio of phosphate to iron in Fe^{3+}-*myo*-inositol phosphate precipitates is affected by several variables, including the way in which the precipitate is washed. These methods are not specific for phytate due to coprecipitation of less phosphorylated *myo*-inositols and should therefore be limited to the analysis of material which contains negligible amounts of these *myo*-inositol phosphates.

Nonprecipitative methods Nonprecipitative methods for *myo*-inositol phosphate determination include ^{31}P-Fourier transform nuclear magnetic resonance (^{31}P-FT NMR) spectroscopy, near-infrared reflectance spectroscopy, low-pressure anion-exchange chromatography, several high-performance liquid chromatographic (HPLC) separation systems, and capillary electrophoresis. The main limitation of ^{31}P-FT NMR and near-infrared spectroscopy is that these methods are specific for phytate only when the sample contains negligible amounts of less phosphorylated *myo*-inositols. Additionally, sophisticated instruments which are not available in most laboratories are required.

Low-pressure anion-exchange chromatography
Low-pressure anion-exchange chromatography is widely used in the determination of *myo*-inositol phosphates. The method currently accepted by the Association of Official Analytical Chemists (AOAC) for measuring phytate in foods and feeds is based on a step gradient (0.7 mol l^{-1} NaCl) anion-exchange method (Dowex AG1-X8). Unfortunately, the anion-exchange resin also retains less phosphorylated *myo*-inositols. The method should therefore be limited to the analysis of material with negligible amounts of these *myo*-inositol phosphates. *Myo*-inositol mono-to hexakisphosphate and even some positional isomers could be resolved using anion-exchange chromatography with a linear eluting gradient of hydrochloric acid or a stepwise elution with either hydrochloric acid or ammonium formate/formic acid solutions of increasing concentrations. Unfortunately, these methods require long elution times (up to 24 h) and a large number of eluate fractions must be hydrolyzed for quantitation as phosphate or inositol, since these systems preclude the use of refractive index and conductivity detection methods. Methods designed by those studying calcium metabolism are dependent on the use of radiolabeled *myo*-inositol phosphates to facilitate detection and quantitation, but it is not feasible to label existing *myo*-inositol phosphates in dietary constituents.

High-performance liquid chromatoraphy More recently, HPLC techniques have been introduced into *myo*-inositol phosphate determination. Purification of crude acid extracts of biological samples is usually required prior to injection on to the analytical HPLC system. The techniques used for detection and quantitation of the *myo*-inositol phosphates is heavily dependent on the system employed for their separation. The *myo*-inositol phosphates may be separated using anion-exchange, reverse-phase, micellar and ion chromatography, and detected/quantified by a variety of techniques, including refractive index, conductivity, indirect photometry, online postcolumn spectrophotometric detection, and offline phosphate or inositol assay. Among these, ion-pair reverse-phase and anion-exchange chromatography are largely used. (*See* **Chromatography**: High-performance Liquid Chromatography.)

Ion-pair reverse-phase chromatography Ion-pair reverse-phase chromatography with refractive index detection has been successfully applied to analysis of *myo*-inositol phosphates. The retention of *myo*-inositol phosphates on reverse-phase packings is markedly increased through the use of ion-pair reagents, allowing the simultaneous separation of *myo*-inositol tris- to hexakisphosphates, but neither *myo*-inositol mono- or bisphosphates nor the individual positional isomers are resolved (Figure 6). However, sample extracts must be passed through anion-exchange resin to remove orthophosphate and concentrate the *myo*-inositol phosphates. Acidic column eluent is then evaporated to dryness to remove hydrochloric acid and reconstituted in water prior to injection on to a silica-based C18 reverse-phase HPLC column. The mobile phase consisted of formic acid/methanol and tetrabutylammonium hydroxide. The affinity of *myo*-inositol phosphates for the stationary phase increases with the increasing number of phosphate groups on the inositol ring and with increasing pH. (*See* **Chromatography**: High-performance Liquid Chromatography.)

Anion-exchange chromatography To date, published procedures involving anion-exchange HPLC fall into two categories: isocratic and gradient ion-chromatographic techniques. The capability of resolving the different *myo*-inositol phosphates depends on the stationary phase used and the chromatography conditions. *Myo*-inositol mono- to hexakisphosphates have been successfully resolved by isocratic elution from low-capacity weak anion-exchange columns. These single eluent systems are compatible with refractive index, indirect photometric, thermospray mass spectrometric, and conductivity detection.

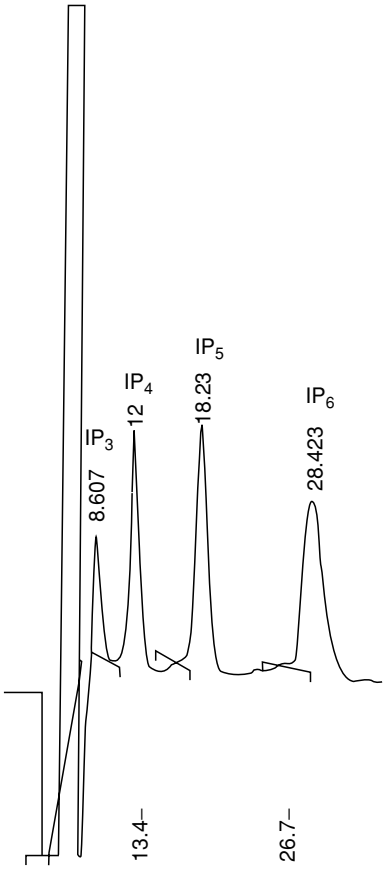

Figure 6 Chromatographic profile of a *myo*-inositol phosphate standard by high-performance liquid chromatography (HPLC) ion-pair chromatography on Ultrasep ES 100 RP18 (2 × 250 mm). The column was run at 45 °C and 0.2 ml min^{-1} of an eluent consisting of formic acid:methanol:water: tetrabutylammonium hydroxide (TBAH: 44:56:5:1.5 v/v), pH 4.25. *Myo*-inositol phosphates were detected by refractive index a standard. Peaks (8.607) IP_3; (12) IP_4; (18.23) IP_5; (28.423) IP_6.

However, in the case of conductivity detection, sensitivity is low unless counterions in the eluent are continously removed using a suppressor column or membrane suppressor system.

In the last few years a number of isomer-specific ion-exchange chromatography methods with gradient elution for separation and quantitation of *myo*-inositol phosphates in the picomolar range have been developed. Eluents with high ionic strength, such as formate, acetate, citrate, phosphate, nitrate, sulfate, sodium chloride, or hydrochloric acid, have been used. The most commonly used detection method with gradient elution is online postcolumn derivatization or complexation reactions followed by spectrophotometric detection. Three approaches have been employed in the postcolumn detection and quantitation of *myo*-inositol phosphates. The first is based on the direct reaction of *myo*-inositol phosphates with a

reagent to form a fluorescent complex or one which has an absorbance in the ultraviolet or visible part of the spectrum. For example, the eluate from the column was mixed, online, with 0.1% $Fe(NO_3)_3$ in 2% $HClO_4$ to form ultraviolet-absorbing phytate–Fe^{3+}–ClO_4 complexes. The use of postcolumn derivatization through ligand-exchange reaction between the iron(III)-sulfosalicylate complex and eluted *myo*-inositol phosphates has been described as an alternative. Furthermore, a complexometric technique based on competition between *myo*-inositol phosphates and the cation-specific reporter dye 4-[2-pyridylazo]resorcinol for the transition metal yttrium has been described. The second approach is based on the online enzymatic hydrolysis of *myo*-inositol phosphates which is then mixed with a molybdate solution in the reaction coil. The colored phosphomolybdate complex may be quantified spectrophotometrically. Finally, *myo*-inositol phosphates may be quantified by online thermospray mass spectrometric techniques.

A remaining problem, however, is to separate isomers from the whole spectrum of *myo*-inositol phosphates in the same run. Separation is generally performed on HPLC columns with gradient elution in two combined systems. *Myo*-inositol mono- to trisphosphates have been acidic gradient-eluted, post-column-derivatized, and ultraviolet-detected and *myo*-inositol bis- to hexakisphosphates have been alkali gradient-eluted and detected using chemically suppressed conductivity detection. The sensitivity of the analysis of *myo*-inositol mono- and bisphosphates was improved 10–100 times by using sodium acetate gradient elution in a sodium hydroxide environment and pulsed amperometric detection.

Capillary electrophoresis Recently capillary electrophoresis has been applied to the determination of *myo*-inositol phosphates. Capillary electrophoresis is attractive, since only a few nanoliters of sample are used in each analysis, there is the potential for concurrent separation of mono- to hexakisphosphate species in the same analysis, and run times are usually short due to the intrinsically high efficiency of the technique. Indirect ultraviolet detection was used to allow the detection of the nonchromophoric *myo*-inositol phosphates. Thus, no derivatization of the compounds is needed. Separations of all six *myo*-inositol phosphate groups in deionized water has been achieved in about 13 min. The position isomers *myo*-inositol(1)phosphate and *myo*-inositol(2)phosphate are easily separated in a phthalate electrolyte system, demonstrating the potential for separating *myo*-inositol phosphate isomers. However, further work is required on the development of capillary electrophoretic methods for the separation and quantiation of the different *myo*-inositol phosphate isomers.

All in all, efficient analytical systems for separation and quantitation of *myo*-inositol phosphate isomers are available. One problem in developing methods to determine *myo*-inositol phosphate isomers is their availability as reference compounds. They may be produced by chemical or enzymatic hydrolysis of phytate. Then identification of these isomers is needed. This requires sophisticated methods. The earliest developed technique is chemical analysis by oxidation with periodate, reduction, dephosphorylation, and subsequent identification of the polyols found. Further information can be obtained from the above-mentioned *cis* phosphate migration. In the past few years, high-resolution nuclear magnetic resonance (NMR) spectroscopy has evolved as a much simpler technique in the identification of the isomeric nature of *myo*-inositol phosphates. Thirty-nine of the 63 theoretically possible *myo*-inositol phosphate isomers can be identified by NMR. Only for the 24 enantiomeric pairs among these isomers (Table 1) absolute configurations are indistinguishable from NMR, but these enantiomers are also not separated on the achiral columns in use. Separation techniques using chiral columns have to be developed to resolve the enantiomers. The absolute configuration of such enantiomers may be determined using high-affinity binding proteins or enzymatic assays.

See also: **Chromatography**: High-performance Liquid Chromatography; **Electrophoresis**: General Principles; **Enzymes**: Functions and Characteristics; **Phosphorus**: Properties and Determination; Physiology; **Plant Antinutritional Factors**: Characteristics; **Preservatives**: Food Uses

Further Reading

Buscher BAP, Irth H, Anderson E, Tjaden UR and van der Greef J (1994) Determination of inositol phosphates in fermentation broth using capillary zone electrophoresis with indirect UV detection. *Journal of Chromatography* A678: 145–150.

Cheryan M (1980) Phytic interactions in food systems. *CRC Critical Reviews in Food Science and Nutrition* 13: 297–335.

Cosgrove DJ (ed.) (1980) *Inositol Phosphates: Their Chemistry, Biochemistry and Physiology*. Amsterdam: Elsevier.

Dean NM and Beaven MA (1989) Methods for the analysis of inositol phosphates. *Analytical Biochemistry* 183: 199–209.

Graf E (ed.) (1986) *Phytic Acid: Chemistry and Application*. Minneapolis: Pilatus Press.

Irvine RF (ed.) (1990) *Methods in Inositide Research*. New York: Raven Press.
Morré DJ, Boss WF and Loewus FA (eds) (1990) *Inositol Metabolism in Plants*. New York: Wiley.
Oberleas D (1971) The determination of phytate and inositol phosphates. In: Glick D (ed.) *Methods of Biochemical Analysis*, vol. 20, pp. 87–101. New York: Wiley.
Potter BVL (1990) Recent advances in the chemistry and biochemistry of inositol phosphates of biological interest. *Natural Products Report* 7: 1–24.
Reddy NR, Pierson MD, Sathe SK and Salunkhe DK (eds) (1989) *Phytates in Cereals and Legumes*. Boca Raton: CRC Press.
Skoglund E, Carlsson N-G and Sandberg A-S (1997) Determination of isomers of inositol mono- to hexaphosphates in selected foods and intestinal contents using High Performance Ion Chromatography. *Journal of Agricultural and Food Chemistry* 45: 431–436.
Turk M, Sandberg A-S, Carlsson N-G and Andlid T (2000) Inositol hexaphosphate hydrolysis by Baker's yeast. Capacity, kinetics, and degradation products. *Journal of Agricultural and Food Chemistry* 48: 100–104.
Xu P, Price J and Aggett PJ (1992) Recent advances in methodology for analysis of phytate and inositol phosphates in foods. *Progress in Food and Nutrition Science* 16: 245–262.

Nutritional Impact

U Konietzny, Karlsruhe, Dettenheim-Liedolsheim, Germany
R Greiner, Centre for Molecular Biology, Federal Research Centre for Nutrition, Karlsruhe, Germany

Introduction

Mixed salts of phytic acid [*myo*-inositol(1,2,3,4,5,6)-hexakisphosphoric acid) are common constituents of foods and feed, since phytate is a naturally occurring compound formed during the maturation of seeds and cereal grains. Depending on the amount of plant foods in the diet and the grade of food processing, the daily intake of phytate can be as high as 4500 mg. On average, the daily intake of phytate was estimated to be 2000–2600 mg for vegetarian diets as well as diets of inhabitants of rural areas of developing countries and 300–1300 mg for mixed diets (**Table 1**).

Phytate behaves in a broad pH region as a highly negatively charged ion and therefore has a tremendous affinity for food components with positive charge(s) (*See* **Phytic Acid**: Properties and Determination). There is a large body of evidence that minerals are less available from foods of plant origin as compared to animal-based foods. Minerals of concern in this regard include zinc, iron, calcium, magnesium, manganese, and copper. The formation of insoluble cation–phytate complexes at physiological pH values is regarded as the major reason for the poor mineral bioavailability, since these complexes are essentially nonabsorbable from the gastrointestinal tract. Furthermore, phytate phosphorus may not be nutritionally available, since phytate is not hydrolyzable quantitatively in the human gut. Consumption of phytate, however, seems not to have only negative aspects on human health. In the last few years, results of epidemiological and animal studies also suggest beneficial effects, such as decreasing the risk of heart disease and colon cancer, but data from human studies are still lacking.

Table 1 Phytate content of foods of plant origin

Food	*Phytate (mg g^{-1} dry matter)*
Cereal-based	
French bread	0.3–0.4
Mixed-flour bread (70% wheat, 30% rye)	0.2–0.7
Mixed-flour bread (70% rye, 30% wheat)	0–0.3
Sourdough rye bread	0–0.3
Wholewheat bread	4.3–6.8
Whole rye bread	2.5–4.8
Unleavened wheat bread	9.2–19.5
Corn bread	5.2–7.1
Unleavened corn bread	11.4–16.3
Oat bran	12.4–29.6
Oat flakes	8.2–10.3
Oat porridge	7.7–10.6
Pasta	2.2–8.6
Maize	11.5–14.2
Cornflakes	0.8–1.3
Rice (polished, cooked)	1.4–2.9
Wild rice (cooked)	16.4–20.1
Sorghum	5.6–9.8
Legume-based	
Chickpea (cooked)	4.9–6.1
Cowpea (cooked)	5.8–10.3
Black beans (cooked)	8.5–11.3
White beans (cooked)	9.1–10.9
Lima beans (cooked)	6.2–9.8
Faba beans (cooked)	10.1–13.7
Kidney beans (cooked)	8.9–11.2
Navy beans (cooked)	7.4–10.6
Soybeans	9.9–14.9
Tempeh	9.1–10.3
Tofu	8.2–9.3
Lentils (cooked)	6.5–9.3
Green peas (cooked)	5.7–7.8
Peanuts	16.5–19.1
Others	
Sesame seeds (toasted)	43.2–55.1
Soy protein isolate	4.3–11.7
Soy protein concentrate	12.4–21.7
Buckwheat	10.3–14.1
Amaranth grain	12.6–14.3

Phytate as an Antinutrient

The main concern about the presence of phytate in the diet is its negative effect on mineral uptake. Most studies have shown an inverse relationship between phytate content and mineral availability, although there are great differences in the behavior of individual minerals (Table 2).

Difficulties in Experimental Approaches

The effect of phytate on mineral absorption is highly controversial, since many investigations have shown a negative effect, but some studies have also shown no effect or even enhancement of mineral uptake. This controversial result gives an idea of the complexity of mineral absorption in the intestine (Figure 1). The differing types of experimental design may explain much of this controversy. *In vitro* studies can only incompletely simulate the physiological factors and physicochemical conditions affecting mineral availability. Also, *in vivo* approaches, widely used in mineral bioavailability studies, are not easily comparable due to the existence of many factors that cannot be reproduced in the different experiments. In addition, part of the variability may arise from differences in the method of phytate analysis and experimental techniques for measuring mineral bioavailability. (*See* **Minerals – Dietary Importance**.)

The solubility of phytate complexes is a critical and perhaps overriding issue, because complexes that are insoluble in the upper small intestine, where maximum mineral absorption normally occurs, are highly unlikely to provide absorbable essential elements. Thus, chemical interactions of phytate in the upper gastrointestinal tract are of particular concern. The form in which many minerals occur in foodstuffs is largely unknown, as is also the form in which they occur in the gut. Thus, predicting the specific interactions of phytate in the gastrointestinal tract and the nutritional implications of these interactions is very difficult. As foods are ingested and the digesta travels through the gastrointestinal tract, phytate may continue to maintain associations developed during ripening or food processing or phytate complexes may dissociate and other chelates form, since binding of phytate with minerals or proteins depends upon the pH value, which changes from low pH in the stomach to about neutral in the upper intestine.

The total composition of experimental diets has a great importance on mineral bioavailability, since the reduced availability of essential minerals depends on several dietary factors, such as the total concentration and composition of minerals or the phytate concentration. Phytate *per se* seems not to have a direct adverse affect on mineral absorption, since the physiological concentration of a single mineral is generally not sufficient for the formation of insoluble phytate complexes in the small intestine. Since calcium concentration in the diet is high enough for complete precipitation of phytate, leading to a coprecipitation of other minerals, the calcium content of the diets is of vital importance to the negative impact of phytate on mineral bioavailability. Calcium clearly augments the adverse effects of phytate on mineral absorption, and numerous other dietary components have lesser effects, both beneficial and adverse. For example, the intake of organic acids such as ascorbic acid and/or meat effectively counteracted the inhibitory effect of phytate, whereas dietary fiber and polyphenols intensified it. Both phytate and fiber have a high potential binding capacity for minerals and, because they are generally presented together in many foods, it is very difficult to separate completely the effects of these two in studies with typical human diets.

Phytate degradation during digestion in the gastrointestinal tract may have a positive effect on mineral absorption, since results of animal and human studies indicated that besides phytate, only the first degradation product of phytate, *myo*-inositol pentakisphosphate, showed negative effects on zinc, iron, and calcium absorption in its isolated form, while *myo*-inositol tetrakis- and trisphosphates had no effect in the concentrations under investigation. Furthermore, phytate phosphorus becomes available for utilization after release from the *myo*-inositol ring. The hydrolysis of phytate in the gastrointestinal tract of monogastric animals, including humans, may be carried out by the action of phytate-degrading enzymes from three sources: dietary phytases, phytases from the bacterial flora in the gut, and intestinal mucosal phytases. The level of dietary calcium is thereby of particular interest. A low calcium concentration may favor increased phytate hydrolysis in the gut, whereas elevated dietary calcium decreases its hydrolysis. It was recently reported that mucosal phytase, if present in the human small intestine, does not seem to play a significant role in phytate digestion, whereas the dietary phytases are an important factor, since these enzymes are active in the stomach for a certain time. Phytate degradation by phytases from the bacterial flora in the colon does not contribute to mineral and phytate phosphorus absorption, since, with the exception of calcium, absorption is negligible there.

Furthermore, the history of food processing appears to be of great interest insofar as it can affect the availability of phytate phosphorus and phytate-associated cations. Depending on the manufacturing process, reduction in phytate content of the foods,

Table 2 Effect of phytate on mineral and phosphorus availability

Zinc	*Iron*	*Calcium*	*Copper, manganese, magnesium*	*Phosphorus*
Zinc was reported to be the essential mineral most adversely affected by phytate. Zinc deficiency in humans was first reported in 1963 in Egyptian boys whose diet consisted in the main of bread and beans. These patients, who were characterized by dwarfism and hypogonadism, showed a response to zinc supplementation of their diet. It became accepted that the presence of phytate in plant products was an important factor in the reduction of zinc absorption from foodstuff. Zinc absorption was shown to be inversely correlated with phytate content of the meal. There is, however, some lack of agreement among studies, particularly with respect to specific foods and their components. In addition, it was shown that phytate not only depresses the bioavailability of dietary zinc, but also substantially reduces the reabsorption of endogenous intestinal zinc	A great deal of controversy exists regarding the effect of phytate on the availability of dietary iron. Much of this controversy may be due to the low absorption of iron in general, the presence of different iron phytates with different solubility, and the existence of two types of food iron, heme and nonheme iron. Heme iron is better absorbed and is little influenced by dietary factors; nonheme iron is less easily absorbed, and its absorption is affected by other dietary factors. Since many human studies indicated that phytate had a very strong inhibitory effect on iron absorption, it is well accepted today that phytate appeared to be the major contributor to the reduction in iron availability in humans, but some other factor(s) also contribute	Human studies indicated that phytate inhibited calcium absorption, but the effect of phytate on calcium availability seems not to be as extreme as on that of iron and particularly zinc. This may be due to the relatively high calcium content of plant-based foods, the ability of phytate degradation by microbial phytases of the gut flora, and the absorption of calcium in the large intestine	Relatively few studies have dealt with the effects of phytate on dietary copper, manganese, and magnesium utilization. Phytate has been shown to decrease their bioavailability in *in vivo* studies. It appears that the effect of phytate on copper, manganese, and magnesium bioavailability is less marked than those for some other essential elements	The fact that phytate phosphorus is poorly available to monogastric animals, including humans, was demonstrated several years ago. Phosphorus is absorbed as the orthophosphate. No conclusive evidence for absorption of phytate or other *myo*-inositol phosphates is available at present. Thus, the utilization of phytate phosphorus by monogastric animals will largely depend on their ability to hydrolyze phytate

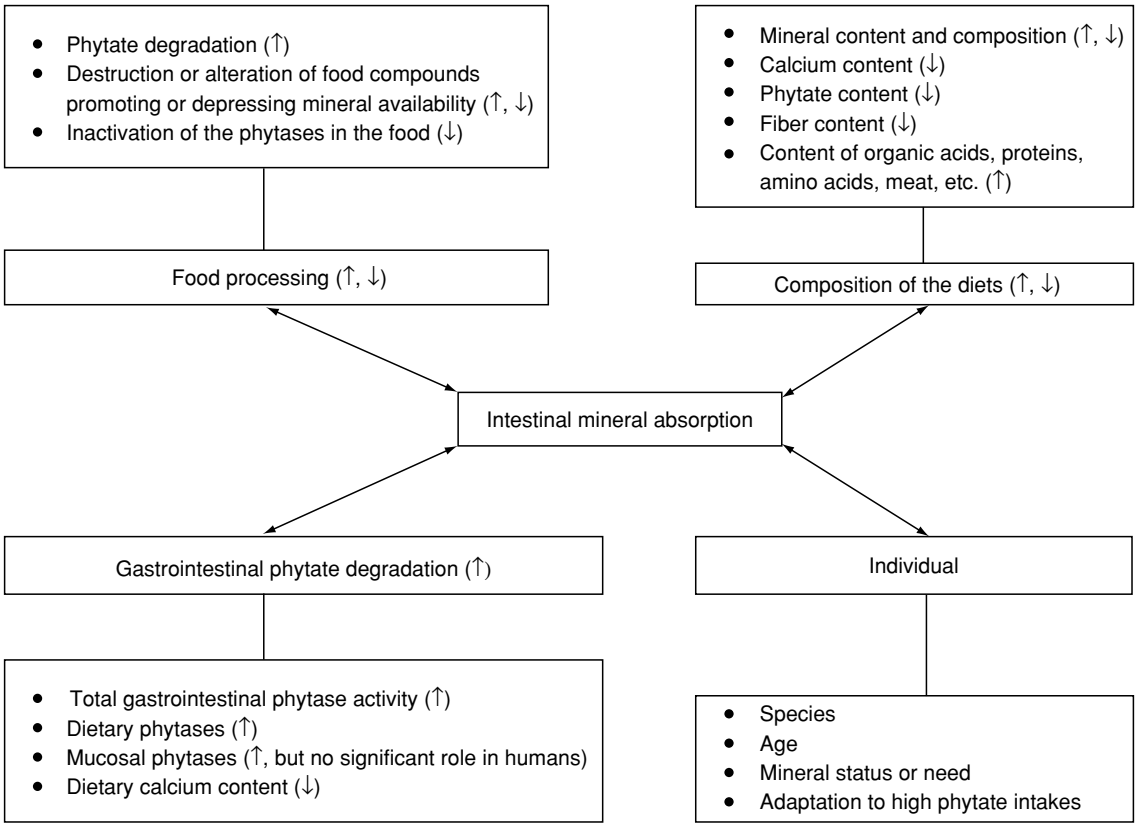

Figure 1 Factors interacting with intestinal mineral absorption. ↑ promoting mineral uptake; ↓ depressing mineral uptake.

destruction or alteration of food compounds promoting or depressing mineral availability, and inactivation of phytate-degrading enzymes may occur. Inactivating the endogenous dietary phytase leads to a limited degradation of phytate in the stomach and therefore to a lower bioavailability of minerals and phytate phosphorus. During food processing and food digestion, phytate can be partially dephosphorylated to yield a large number of positional isomers of *myo*-inositol pentakis-, tetrakis-, tris-, bis-, and monophosphates (*See* **Phytic Acid**: Properties and Determination). Especially with respect to the isomers, their chelating potential has not been carefully studied. This is of great importance, since processed foods and digesta may contain lower *myo*-inositol phosphates in great amounts and human studies have shown that *myo*-inositol tetrakis- and tris-phosphates contribute in high concentrations to the negative effect on mineral absorption. On the other hand, recent evidence has suggested that, besides the adverse effects of phytate and other highly phosphorylated *myo*-inositol phosphates on mineral absorption, lower phosphorylated *myo*-inositol phosphates may improve mineral absorption. Therefore, a reliable and accurate determination of the individual *myo*-inositol phosphates in diets must be utilized before any useful evaluation can be made of their physiological effect on mineral availability (*See* **Phytic Acid**: Properties and Determination).

The source of phytate and minerals also influences the results of experimental studies. Endogenous phytate and added phytate are not necessarily equal. Usually sodium phytate is added to experimental diets; this is capable of chelating the multivalent cations therein, whereas endogenous phytate is already bound to those cations in the seeds and grains. Furthermore, the availability of minerals has mostly been examined after addition of ionic salts to the experimental diets. Results obtained under such conditions may not represent the level of absorption of minerals present in natural sources.

The effect of phytate on mineral absorption also seems to be dependent on the experimental species used. Phytate degradation in the stomach and small intestine varies with the species. For example, it was suggested that rats may not be a good model for assessing mineral absorption from phytate-containing foods due to the existence of rat intestinal phytase activity. In addition, the age of the individuals is of importance, since it seems that phytate digestion decreases with age. The ability of endogenous carriers in the intestinal mucosa to absorb essential minerals

bound to phytate or other dietary substances as well as the mineral status and need of the individual also have to be considered. A possible adaptation to a high-phytate intake is controversially discussed. A long-term study did not show any human adaptation to a high phytate diet with respect to iron absorption, whereas the normal bone and teeth calcification throughout the world in several populations who depend almost exclusively on cereal diets suggests human adaptability towards high phytate consumption with respect to calcium absorption.

Protein

Phytate also interacts with proteins, which may affect protein digestibility negatively. Strong evidence exists that phytate–protein interactions negatively affect protein digestibility *in vitro*. The extent of this effect depends on the protein source. The inhibition of digestive enzymes such as alpha-amylase, lipase, or proteinase by phytate may also be of significance, as shown in *in vitro* studies. This may be due to the nonspecific nature of phytate–protein interactions, the chelation of calcium ions which are essential for the activity of trypsin and alpha-amylase, or the interaction with the substrates of these enzymes. The inhibition of proteases may be partly responsible for the reduced protein digestibility. *In vivo*, phytate has also been considered to inhibit alpha-amylase, as indicated by a negative relationship between phytate intake and blood glucose response. A negative effect of phytate on the nutritive value of protein, however, was not clearly confirmed *in vivo*. While some have suggested that phytate does not affect protein digestibility, others have found improvement in protein utilization with decreasing levels of phytate. This difference may be at least partly due to the use of different protein sources. Thus, the significance of protein–phytate complexes in nutrition is still under scrutiny. (*See* **Protein**: Digestion and Absorption of Protein and Nitrogen Balance.)

Beneficial Effects of Phytate

In view of the above results, the evidence seems overwhelming that high intakes of phytate can have adverse effects on mineral uptake in humans. In the last few years, however, some novel metabolic effects of phytate or some of its degradation products have been recognized. Results of epidemiological and animal studies suggest possible protective effects of certain *myo*-inositol phosphates on heart disease, renal stone formation, and colon cancer in humans. Moreover, the potential beneficial effects of phytate in the prevention of severe poisoning should also be considered. Calcium phytate 1–2% in the diet has been found to protect against dietary lead in experimental animals and in human volunteers. Furthermore, calcium phytate was capable of lowering blood lead levels. Thus, phytate seems to be a helpful means of counteracting acute oral lead toxicity. The effect of calcium phytate on acute cadmium toxicity is still controversially discussed.

Phytate and Diabetes Mellitus

Diabetes mellitus is one of the commonest nutrition-dependent diseases in western society. It may be caused by hypercaloric diets with a high percentage of quickly available carbohydrates. Foods that result in low blood glucose response have been shown to have great nutritional significance in the prevention and management of diabetes mellitus. In this regard phytate-rich foods are of interest, since a negative relationship between phytate intake and blood glucose response was reported. For example, phytate-enriched unleavened bread based on white flour reduced *in vitro* starch digestibility as well as flattening the glycemic response in five healthy volunteers compared to bread without the addition of phytate. The *in vitro* reduction of starch digestion was positively correlated with the *myo*-inositol phosphate concentration and negatively with the number of phosphate groups on the *myo*-inositol ring. It has to be noted that there are also studies which have not found an inhibition of alpha-amylase and starch digestion by phytate. (*See* **Diabetes Mellitus**: Etiology.)

Phytate and Coronary Heart Disease

Heart disease is a leading cause of death in western countries, yet it is low in Japan and developing countries. Elevated plasma cholesterol or, more specifically, elevated low-density lipoprotein (LDL)-cholesterol concentrations have been shown to be one of the risk factors. It has been proposed that dietary fiber or, more specifically, phytate as a component of fiber may influence the etiology of heart disease. Animal studies have demonstrated that dietary phytate supplementation resulted in significantly lowered serum cholesterol and triglyceride levels. This effect was accompanied by a decrease in serum zinc level and in the ratio of zinc to copper. Thus, the hypothesis was put forward that coronary heart disease is predominantly a disease of imbalance in regard to zinc and copper metabolism. The hypothesis is also based on the production of hypercholesterolemia, which is a major factor in the etiology of coronary heart disease, in rats fed a diet with a high ratio of zinc and copper. It was thought that excess zinc in the diet resulted in decreased copper uptake from the small intestine,

since both minerals compete for common mucosal carrier systems. As phytate preferentially binds zinc rather than copper it is presumed that phytate exerts its effect probably by decreasing zinc without affecting copper absorption. It should be pointed out that the support for the preventive role of phytate in heart disease is only based on a few animal and *in vitro* studies. Results from human studies are still lacking. (*See* **Coronary Heart Disease**: Etiology and Risk Factor.)

Phytate and Renal Calculi

The increase of renal stone incidence in northern Europe, North America, and Japan has been reported to be coincident with the industrial development of these countries, making dietary intake suspect. Epidemiological investigations found that there were substantial differences in renal stone incidences between white and black residents of South Africa. Whereas renal stone occurred in the white population with no less frequency than in other Western communities, it was seldom seen in the black South Africans. The major dietary difference is that, compared to the white population, blacks consumed large amounts of foods containing high levels of fiber and phytate. Furthermore, a high-phytate diet has been used effectively to treat hypercalciuria and renal stones in humans. Thus, there is evidence to support the role of phytate in the prevention of renal stone formation.

Experimental evidence indicates that *myo*-inositol bis- and trisphosphates are effective in preventing the formation of hydroxyapatite crystals *in vivo*, which can function as nuclei for stone formation. Thus, the hypothesis was put forward that a high dietary phytate intake results in an increased urinary content of lower *myo*-inosital phosphates, which may act as very effective inhibitors of stone formation. (*See* **Renal Function and Disorders**: Kidney: Structure and Function.)

Phytate and Caries

The higher incidence of caries in industrialized compared to developing countries was suggested to be nutrition-dependent. Phytate lowers the solubility of calcium, fluoride, and phosphate, the major components of enamel. Thus, teeth are more protected against the leading cause of caries, the attack of acids and bacteria. Furthermore, the very high affinity of phytate for hydroxyapatite may prevent the formation of plaque and tartar. (*See* **Dental Disease**: Etiology of Dental Caries.)

Phytate and Cancer

The frequency of colonic cancer varies widely among human populations. It is a major cause of morbidity and mortality in western society. The incidence of cancer, especially large intestinal cancer, has been associated principally with dietary fat intake and is inversely related to the intake of dietary fiber. It was further suggested that the apparent relationship between fiber intake and rate of colonic cancer might arise from the fact that many fiber-rich foods contain large amounts of phytate and that this latter might be the critical protective element, since an inverse correlation between colon cancer and the intake of phytate-rich fiber foods, but not phytate-poor fiber foods, has been shown. A high phytate intake may also be an important factor in reducing the breast and prostatatic cancer mortality in humans. (*See* **Cancer**: Epidemiology.)

Both *in vivo* and *in vitro* experiments have shown striking anticancer potential for phytate. It was demonstrated that a treatment regimen of 1–2% sodium phytate in the drinking water of growing rats significantly decreased the number of colon tumors and tumor volume when treatment was commenced prior to carcinogenic induction or when administered up to 8 months postinitiation. The inhibitory effect of phytate was dose-dependent and a mixture of 1% *myo*-inositol and 1% phytate was more effective in suppressing cell proliferation. Unfortunately, these investigations failed to study and/or report the mineral status of the animals.

A recent study demonstrated that phytate also has antineoplastic effects on another tumor model, the murine fibrosarcoma. Further rat studies indicated that phytate may also reduce the risk of breast cancer.

Not only pure phytate dissolved in the drinking water, but also phytate in experimental diets such as wheat bran has a protective effect against colon cancer, as shown in animal studies. It was concluded that, while having a role, endogenous phytate is not the sole active component in wheat bran. This result clearly emphasizes the fact that dietary components should not be studied individually for their antineoplastic effect. Though phytate at the 2% level significantly reduced the weight gain of the animals, it does not seem to be of major consequence.

How phytate exerts its antineoplastic and antiproliferative action is not understood. It was proposed that the mechanism by which dietary phytate reduced colon cancer was via chelation of iron and suppression of iron-related initiation and promotion of carcinogenesis, since *in vitro* experiments demonstrated that phytate is a powerful inhibitor of the Haber–Weiss reaction, in which iron catalyzes hydroxyl radical formation (**Figure 2**). These radicals are mediators of several tissue damages related to tumor initiation and promotion. Phytate not only suppresses

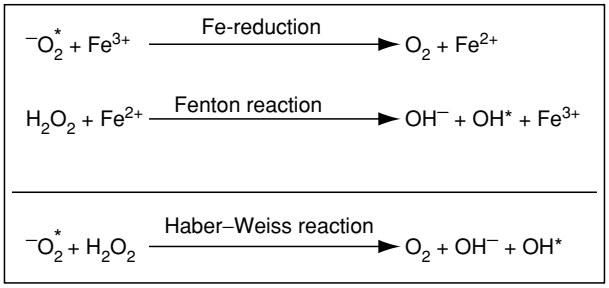

Figure 2 Iron-catalyzed radical formation.

iron-catalyzed hydroxyl radical generation, but also acts to inhibit almost totally iron-catalyzed lipid peroxidation. *Myo*-inositol tris-, tetrakis-, and pentakisphosphate are also shown to be effective. The effects of phytate on cancers distant from the intestine, such as breast cancer, might be simply explained by limitation of iron absorption.

According to a further hypothesis, suppression of tumor incidence in experimental animals may be mediated in part by natural killer cells, since phytate treatment enhanced the activity of natural killer cells involved in the destruction and growth inhibition of tumor cells. It was also suggested that phytate acts by influencing cell regulation. Decreased tumor cell growth may be due to the complexation of zinc and magnesium, which play an important role in cell regulation and gene expression, or due to an influence of extracellular phytate on the intracellular phosphatidylinositol phosphate system, which is important in regulating a variety of physiological and biochemical processes, including cell growth via the second messengers D-Ins(1,4,5)P_3 and D-Ins(1,3,4,5)P_4 (*See* **Phytic Acid**: Properties and Determination). Theoretically, dietary phytate could serve as a precursor for second messengers *in situ* or as a negative-feedback inhibitor for their formation. As a prerequisite, phytate or its hydrolysis products have to be taken up by tumor cells. It was shown that, after exposure of human-transformed cells to [^{3}H]-phytate, a rapid uptake of radioactivity occurred, whereas nontransformed cells showed only a limited uptake. The uptake of radioactivity seems to correspond with tumor cell growth rate. Furthermore, intragastric-administered [^{3}H]-phytate is rapidly absorbed through the upper gastrointestinal tract and distributed to various organs in the rat. Analysis demonstrated that in both studies most of the radioactivity was due to *myo*-inositol and *myo*-inositol monophosphate. The authors suggested that phytate could be absorbed as such or after a variable degree of dephosphorylation, but it cannot be ruled out that phytate was taken up after complete dephosphorylation to *myo*-inositol. From the extracellular area, phytate or its degradation products may be internalized by endocytosis involving fluid-phase pinocytosis, phagocytosis, or receptor-mediated endocytosis, as well as by partitioning into and/or through the plasma membrane. The possibility of receptor-mediated endocytosis was supported by identification of a receptor protein for phytate in rat brain.

Conclusion

Numerous reports suggest that a high phytate intake may suppress absorption of minerals such as zinc, iron, and calcium and that removal of dietary phytate significantly improves the bioavailability of those minerals in humans and animals. Thus, maintaining a diet exceptionally high in phytate may not be entirely without risk. However, there is no uniform agreement on this score, and several studies indicate that phytate may have minimal or no adverse effects on mineral bioavailability in humans. Furthermore, it has been suggested that phytate may exhibit some beneficial health effects, such as reducing the risk of heart disease, renal stone formation, and certain types of cancer.

The most severe effects attributable to phytate in humans have occurred in populations with unrefined cereals and/or pulses as a major dietary component. In particular, zinc and iron deficiency, but not calcium, magnesium, manganese, or copper deficiency, were observed as a consequence of high phytate intake. Therefore, the anticalcifying and rachitogenic effect of phytate has been questioned; it appears that marginal vitamin D status is more important in the etiology of rickets and osteomalacia. Furthermore, marked mineral deficiency syndromes attributed to phytate have not been identified in highly developed countries. Thus, phytate intake does not necessarily result in mineral deficiency. The absorption of minerals depends on the total composition of the meal and in a balanced diet containing animal protein, a high phytate intake does not imply a risk of inadequate mineral supply. Therefore, the recommendation for increasing dietary fiber in western diets would not be expected to have any adverse effect on mineral absorption. The higher phytate intake with whole-grain products will undoubtedly lead to a percentage decrease in mineral absorption, but the absolute amount of absorbed minerals may remain unchanged, because of the large amounts of minerals in these products. In addition, the impact of phytate on phosphorus availability can be considered to be of little consequence in humans, since the phosphorus intakes are usually high, and phytate phosphorus represents only a small portion of the total phosphorus in the diets.

In population groups where phytate-containing foods contribute to poor or marginal mineral status, fortification of some such foods with the corresponding minerals and/or the reduction in phytate content are possibilities to reduce the risk of mineral deficiencies. Since phytate and certain phytate degradation products have been proposed to exhibit health benefits in humans, complete degradation or removal of phytate during food processing should not be followed up, but a controlled degradation to lower *myo*-inositol phosphates. This seems to be a way of avoiding the adverse effects of phytate on intestinal mineral absorption and of utilizing their potential beneficial properties. Controlled degradation is most easily feasible by using phytases. The addition of exogenous or activation of endogenous phytases during food processing has already been shown to result in extensive phytate degradation without affecting other food components.

Phytases are also of interest in areas of intensive animal agriculture. There has been widespread concern in recent years about possible phosphorus pollution from animal manures. The accumulation of phosphorus over time from excessive excretion in manure increases the potential for contaminating water sources from runoff and erosion. The effect of increased content of phosphorus in water is eutrophication, which decreases the available oxygen within the water, thus posing a hazard to fresh-water animal life. Effluent control is therefore a high priority in areas of intensive animal production and, in this context, phytase can become an important waste management tool, since phytase is active in the stomach for a certain time and thus increases phytate hydrolysis during digestion. Phytase not only has the potential to reduce environmental pollution by minimizing the excretion of phosphorus and nitrogen in manure, but also reduces the need for mineral supplementation by increasing the availability of cations bound to phytate.

Because suggestions for the beneficial aspects of *myo*-inositol phosphates are only derived from *in vitro*, animal, or epidemiological investigations, carefully controlled human and animal studies should be carried out to evaluate simultaneously the potential benefits and adverse effects of dietary *myo*-inositol phosphates. If certain phytate degradation products induce physiological effects (the position of the phosphate groups on the *myo*-inositol ring is of great significance for the physiological function), phytase may also find application in food processing to produce foods with improved nutritional value, health benefits, and maintained sensory properties (functional foods). By adding the phytase to the raw material, the antinutritional factor phytate will be degraded to physiologically active *myo*-inositol phosphates during food processing. Thus, foods with a reduced content of phytate and a regulated content and composition of lower *myo*-inositol phosphate with beneficial health effects could be designed.

See also: **Cancer**: Epidemiology; **Coronary Heart Disease**: Etiology and Risk Factor; **Dental Disease**: Etiology of Dental Caries; **Diabetes Mellitus**: Etiology; **Minerals – Dietary Importance**; **Phytic Acid**: Properties and Determination; **Protein**: Digestion and Absorption of Protein and Nitrogen Balance; **Renal Function and Disorders**: Kidney: Structure and Function

Further Reading

Challa A, Ramkishan R and Reddy BS (1997) Interactive suppression of aberrant crypt foci induced by azoxymethane in rat colon by phytic acid and green tea. *Carcinogenesis* 18: 2023–2026.

Erdman JW Jr and Poneros-Schneier A (1989) Phytic acid interactions with divalent cations in foods and the gastrointestinal tract. *Advances in Experimental Biology* 249: 161–171.

Fox MRS and Tao TH (1989) Antinutritive effects of phytate and other phosphorylated derivatives. *Nutritional Toxicology* 3: 59–96.

Harland BF and Morris ER (1995) Phytate: a good or a bad food component? *Nutrition Research* 5: 733–754.

Jariwalla RJ, Sabin R, Lawson S and Herman ZS (1990) Lowering of serum cholesterol and triglycerides and modulation of divalent cations by dietary phytate. *Journal of Applied Nutrition* 42: 18–28.

Jenab M and Thompson LU (1998) The influence of phytic acid in wheat bran on early biomarkers of colon carcinogenesis. *Carcinogenesis* 19: 1087–1092.

Ohkawa T, Ebisuno S, Kitagawa M *et al.* (1984) Rice bran treatment for patients with hypercalciuric stones: experimental and clinical studies. *Journal of Urology* 132: 1140–1145.

Porres JM, Stahl CH, Cheng WH *et al.* (1999) Dietary intrinsic phytate protects colon from lipid peroxidation in pigs with a moderately high dietary iron intake. *Proceedings of the Society for Experimental Biology and Medicine* 221: 80–86.

Sandberg AS, Brune M, Carlsson NG, Hallberg L, Skoglund E and Rossander-Hulthen L (1999) Inositol phosphates with different numbers of phosphate groups influence iron absorption in humans. *American Journal of Clinical Nutrition* 70: 240–246.

Shamsuddin AM, Vuvenik I and Cole KE (1997) IP-6: a novel anti-cancer agent. *Life Science* 61: 343–354.

Shen X, Weaver CM, Kempa-Steczko A, Martin BR, Phillippy BQ and Heany RP (1998) An inositol phosphate as a calcium absorption enhancer in rats. *Nutritional Biochemistry* 9: 298–301.

Thompson LU (1988) Antinutrients and blood glucose. *Food Technology* 42: 123–132.

Vucenik I and Shamsuddin AM (1994) [^{3}H]Inositol hexaphosphate (phytic acid) is rapidly absorbed and metabolized by murine and human malignant cells in vitro. *Journal of Nutrition* 124: 861–868.

Zhou JR and Erdman JW Jr (1995) Phytic acid in health and disease. *Critical Reviews in Food Science and Nutrition* 35: 495–508.

Phytochemicals *See* **Functional Foods**

PICKLING

D M Barrett, University of California Davis, CA, USA

Introduction

Pickling is one of the oldest, and most successful, methods of food preservation known to humans. This article reviews the origins of pickling, the various methods of pickling employed commercially, and the nature of preservative action. The optimization of pickle quality depends on maintenance of proper acidity, salt concentration, temperature, and sanitary conditions.

History and Tradition

It is difficult to suggest a date for production of the first pickled foods, but it is known that both vinegar and spices were being used during biblical times. Fermentation of plant and animal foods was known to the early Egyptians, and fish were preserved by brining in prehistoric times. By the third century BC, Chinese laborers were recorded to be consuming acid-fermented mixed vegetables while working on the Great Wall. In about 2030 BC, northern Indians brought the seed of the cucumber to the Tigris Valley. The Koreans created kimchi from acid-fermented Chinese cabbage, radish, and other ingredients many centuries ago. Corn, cassava, and sorghum were fermented and became staples of the African diet. In the west, acid fermentation of cabbage and cucumbers produced sauerkraut and pickles, products that are still popular today. Early explorers carried kegs of sauerkraut and pickles that prevented scurvy on their voyages.

Peterson in 1977 defined pickling, in a broad sense, as 'the use of brine, vinegar or a spicy solution to preserve and give a unique flavour to a food.' In a 1936 document, the US Department of Agriculture described cucumber pickles as:

> immature cucumbers properly prepared, without taking up any metallic compounds other than salt, and preserved in any kind of vinegar, with or without spices. Pickled onions, pickled beets, pickled beans and other pickled vegetables are vegetables, prepared as described above, and conform in name to the vegetables used.

Literature references concerning the technology of acid fermentation began to appear in the western press in the early 1900s. In 1919, Orla-Jensen isolated strains of *Betacoccus arabinosaceus* from sour potatoes, sour cabbage, and sour dough. Pederson, in a number of classic studies in the 1930s, enumerated and identified the sequence of microorganisms involved in sauerkraut fermentation. *Leuconostoc mesenteroides* was identified as being one of the most important microorganisms for initiation of vegetable fermentation. Numerous investigators have carried out studies on acid-fermented vegetables over the past century. (*See* **Fermented Foods: Beverages from Sorghum and Millet.**)

Advent of New Pickling Methods

Brining vegetables in salt, and the resultant lactic acid fermentation, is an ancient form of preservation. Two new methods of pickling cucumbers, which represent the largest volume of a single vegetable preserved by pickling, have been developed during the 1900s. Both methods use lower salt concentrations and result in a milder product. The first new method, pasteurization and direct acidification, was developed and began commercial production in the 1940s. The second,

refrigeration and direct addition of acid and preservative, was introduced in the 1960s. Relative quantities of cucumbers preserved by the three pickling methods in 1984 were: brine fermentation, 43%; pasteurization, 43%; and refrigeration, 14%.

Pickling remains a major form of preservation in many countries because it: (1) yields desirable organoleptic qualities; (2) provides a means for extending the processing season of fruits and vegetables; and (3) requires relatively little mechanical energy input.

Outline of the Process

Pickled products may be produced by one of three methods, i.e. fermentation, acidification followed by pasteurization, or refrigeration. Because they are relatively straightforward, the methods for producing pasteurized and refrigerated pickled products will be described first.

Pasteurized Pickles

Vegetables which are fresh or only partially fermented may be preserved by the addition of vinegar or acetic acid and subsequent pasteurization. Vinegar alone is not sufficient to insure product safety, and so requires an additional form of preservation such as heat or refrigeration. The steps involved in producing pasteurized or 'fresh-pack' pickles are the following:

- slice, cube, or dice product;
- place in clean container;
- mix water, salt, vinegar, sugar, spices and bring to boil;
- add hot brine to container;
- seal and pasteurize.

Two possible pasteurization methods include: (1) heating such that the center of the jar or can reaches 75 °C, and holding for 15 min, followed by a prompt cooling to 35 °C or below, or (2) heating at 70 °C for 10 min, followed by prompt cooling. (*See* **Vinegar.**)

The pasteurization process essentially destroys spoilage microorganisms and prohibits fermentation from occurring. Both acid-forming bacteria, which are active in brine fermentation, and yeasts, which cause gas production, are destroyed by pasteurization. The pasteurization process inhibits polygalacturonase, the enzyme responsible for pickle softening. Enzyme activity may also be controlled through use of appropriate salt concentrations. Calcium chloride is often added to brine to aid in firming cucumber pickles. Pasteurized pickled products have steadily gained in popularity, and have a very different flavor and texture to that of fermented pickled products. (*See* **Pasteurization**: Principles.)

Refrigerated Pickles

Refrigerated pickle products, which represent the latest development in pickling technology, are produced by direct acidification and addition of sodium benzoate or another preservative. In these nonfermented pickles, the preservative takes the place of pasteurization in preventing spoilage of the product. This process is essentially the same as that used for pasteurized pickles, but instead of pasteurizing, the sealed containers are refrigerated. It is essential that this type of pickle is kept refrigerated throughout the production process and during subsequent consumption.

Fermented Pickles

Cucumbers are by far the most common vegetable fermented; therefore, most of the following processes will be described using cucumbers as an example. There are three general methods which may be used for cucumber fermentation; (1) salt stock; (2) genuine dill; and (3) overnight dill.

Salt stock The first method, salt stock, involves fermentation in 5–8% sodium chloride until all fermentable sugars have been converted to acids and other end products. The cucumbers are then stored in open tanks containing 10–16% salt to maintain stability for up to 1 year. Figure 1 outlines the general process for salt stock pickles. Desalting to an acceptable organoleptic level (2–2.5% salt) is carried out by

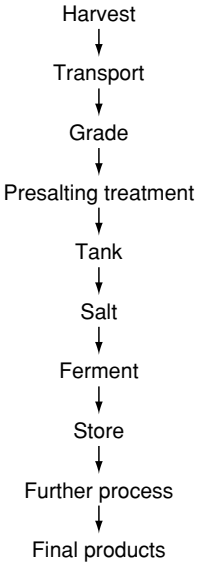

Figure 1 General flowchart for preservation of vegetables by brining. Reproduced from Fleming HP and McFeeters RF (1981) Use of microbial cultures: vegetable products. *Food Technology* 1: 84–88, with permission.

leaching in water. Salt stock pickles make up the largest percentage of fermented pickle products.

The controlled fermentation of pickles has been a goal of the industry for years. In 1973, such a procedure was introduced in the USA. Although the industry has yet to adopt the entire process, addition of some of the modifications outlined has improved pickle quality in general. In particular, most processors now acidify and purge tanks after brine is added. Acidification inhibits the growth of acid-sensitive Gram-positive and Gram-negative bacteria and, therefore, favors the growth of lactic acid bacteria. Purging decreases the incidence of bloating, which results from carbon dioxide production by both the fermenting microorganisms and the cucumber itself.

Genuine dill These pickles are fermented in 4–5% sodium chloride, to which dill weed, garlic, and other spices have been added. It takes approximately 3–6 weeks for fermentation to reach completion, where the lactic acid content is 1.0–1.5% and the salt content is 3–3.5%. This type of pickle does not require desalting, but may be sold as such with the filtered fermentation liquor. Genuine dill pickles are somewhat susceptible to scum yeast development and should be consumed within 12 months of manufacture.

Overnight dill Overnight dill pickles are fermented in 2–4% sodium chloride, containing dill weed and garlic, until the desired level of acidity (0.75–1% titratable acidity as lactic acid) is reached, which should take approximately 1 week. The product must then be refrigerated, and should not be kept longer than 6 months due to its extreme perishability.

Nature of Preservative Action

Two components of the pickling process, acid and salt, are key participants in the preservation of perishable products. Acid, which may be added directly or produced through microbial conversion of indigenous sugars to acids, will lower the pH of the product and inhibit spoilage microorganisms. Only undissociated acid molecules are active in inhibiting microorganisms; therefore, it is important that the acidity of pickles be pH 3.5 or below, when most of the acid present will be in the undissociated form. Salt also acts to inhibit the growth of undesirable bacteria and to delay enzymatic softening. (*See* **Acids**: Natural Acids and Acidulants; **pH – Principles and Measurement**.)

In fermented pickles, microorganisms ferment sugars to lactic acid and also produce enzymes which modify pickle texture. The absence of fermentable carbohydrates is a deterrent to undesirable secondary fermentations which can be initiated by yeasts at pH values below 3.8. Residual sugars can also cause gas production and brine turbidity in finished products if yeast and bacteria growth continues.

Lactic acid bacteria are the primary microorganisms involved in preservation of fermented pickled products. Although these microorganisms represent only a smaller proportion of the total microbial flora present on the surface of plant materials, they predominate under acidic conditions. In cucumbers, *L. mesenteroides* typically grows until the pH begins to drop, then *Pediococcus pentosaceus* predominates, followed by *Lactobacillus brevis* and finally *L. plantarum*. (*See* **Lactic Acid Bacteria**.)

Acidity, salt concentration, temperature, and sanitary conditions are the primary environmental factors which control fermentation. Most vegetables range in pH from 6.5 to 4.6, while fruits range from 4.5 to 3.0. In fruits and fruit juices, yeasts and molds predominate in the more acidic environments. Salt inhibits the growth of undesirable microorganisms and, in addition, it withdraws water and nutrients from plant tissues and allows these to become substrates for lactic acid bacteria.

Low temperatures inhibit the growth of lactic acid bacteria and thus slow fermentation. At 7.5 °C, *Leuconostoc mesenteroides* will grow, but the growth of *Lactobacillus* and *Pediococcus* species is very low. At temperatures in the range 18–23 °C, *L. brevis* and *L. plantarum* exhibit active growth while, at 32 °C, *L. plantarum* and *P. pentosaceus* predominate. Pasteurization is often the final step in pickle processing. The US Department of Agriculture recommends that all pickled products should be pasteurized for safety.

Sensory and Nutritional Attributes

Pickling imparts unique characteristics to fruits and vegetables. Desirable changes in flavor, texture, and color take place in fermented, pasteurized, and refrigerated pickles, and are carefully monitored. However, some of the same bacteria involved in normal fermentation, such as the lactic-acid-formers, may cause spoilage if not destroyed. Selected pickling problems which may affect sensory quality are listed in Table 1.

Pickled cucumbers are composed primarily of water (Table 2). They are not a good source of protein or fat, but are a fairly rich source of vitamin A, vitamin C, and phosphorus. Salt and sugar additions to the brine will affect the sodium and carbohydrate contents. (*See* **Sensory Evaluation**: Sensory Characteristics of Human Foods.) Refer to individual nutrients.

Table 1 Selected pickling problems

Problems	*Cause*
Soft, slippery slimy pickles (discard pickles, spoilage is occurring)	Hard water Acid level too low Cooked too long or at too high a temperature Water bath too short, bacteria not destroyed Jars not airtight Jars in too warm a resting place
Shriveled, tough pickles	Pickles overcooked Syrup too heavy Too strong a brine or vinegar solution Pickles not fresh enough at outset Fruit cooked too harshly in vinegar/sugar mixture
Dark, discolored pickles	Iron utensils used Copper, brass, iron, or zinc cookware used Hard water Metal lid corrosion Too great a quantity of powdered and dried spices used Iodized salt used

Reproduced from McNair JK (ed.) (1975) *All About Pickling*. San Francisco: Ortho Books, with permission.

Specific Examples

Pickled cucumbers are by far the most abundant pickled product available in the western world today. Other common pickled products include sauerkraut, pickled pears, peaches and plums, pickled nuts, relishes, cured meats, fish and poultry, and specialty items such as pickled mushrooms and cherries.

Salt stock is used to prepare sour cucumber pickles, which typically have a final acidity not lower than 2.5%. Sweet pickles are prepared in a similar fashion, except that a sweet, spiced vinegar solution is added to the salt stock.

Sauerkraut is produced through salt-controlled bacterial fermentation. Cabbage selected for sauerkraut should have at least 3.5% sugar to insure an adequate carbohydrate source for bacteria. Shredded cabbage is mixed with salt (2.25% by weight) and the final product contains an average of 1.5–2.0% lactic acid.

Table 2 Nutritive analysis of pickles: the composition of 100 g of edible portion (approx. 1 large dill pickle or 1/2 cup of fresh cucumber pickle slices)

	Fermented dill pickles	*Sweet pickles*	*Sour pickles*	*Fresh pack cucumber pickles*
Water (%)	93	60.7	94.8	78.7
Food energy (J)	46.2	613.2	42	306.6
Protein (g)	0.7	0.7	0.5	0.9
Fat (g)	0.2	0.4	0.2	0.2
Carbohydrate (g)	2.2	36.5	2.0	17.9
Ash (g)	3.6	1.7	2.5	2.3
Calcium (mg)	26.0	12.0	17.0	32.0
Iron (mg)	1.0	1.2	3.2	1.8
Vitamin A (IU)[a]	100	90	100	140
Thiamin	Trace	Trace	Trace	Trace
Riboflavin (mg)	0.02	0.02	0.02	0.03
Vitamin C (mg)	6.0	6.0	7.0	9.0
Phosphorus (mg)	21.0	16.0	15.0	27.0
Potassium (mg)	200.0			
Sodium (mg)	1428.0		1353.0	673.0

Reproduced from McNair JK (ed.) (1975) *All About Pickling*. San Francisco: Ortho Books, with permission. 1 IU = 0.6 μg β-carotene.

See also: **Acids**: Natural Acids and Acidulants; **Fermented Foods**: Beverages from Sorghum and Millet; **Lactic Acid Bacteria**; **Pasteurization**: Principles; **pH – Principles and Measurement**; **Vinegar**; **Sensory Evaluation**: Sensory Characteristics of Human Foods

Further Reading

Binsted R, Dewey JD and Dakin JC (1939) *Pickle and Sauce Making*. London: Food Trade Press.

Fleming HP (1984) Developments in cucumber fermentations. *Journal of Chemical Technology and Biotechnology* 34B: 241–252.

Fleming HP and McFeeters RF (1981) Use of microbial cultures: vegetable products. *Food Technology* 1: 84–88.

McNair JK (ed.) (1975) *All About Pickling*. San Francisco: Ortho Books.

Pederson CS (1930) Floral changes in the fermentation of sauerkraut. *NYSAES Technical Bulletin* 168.

Pederson CS (1979) *Microbiology of Food Fermentations*, 2nd edn. Westport: AVI.

Peterson MS (1977) Pickles. In: Desrosier NW (ed) *Elements of Food Technology*, pp. 690–691. Westport: AVI.

Steinkraus KH (1983) *Handbook of Indigenous Fermented Foods*. New York: Marcel Dekker.

US Department of Agriculture (1936) *Service and Regulatory Announcements*, no. 2, rev. 5. Washington, DC: Food and Drug Association.

Pigments *See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments; Properties and Determinants of Synthetic Pigments

Pilchards *See* **Fish**: Introduction; Catching and Handling; Fish as Food; Demersal Species of Temperate Climates; Pelagic Species of Temperate Climates; Tuna and Tuna-like Fish of Tropical Climates; Demersal Species of Tropical Climates; Pelagic Species of Tropical Climates; Important Elasmobranch Species; Processing; Miscellaneous Fish Products; Spoilage of Seafood; Dietary Importance of Fish and Shellfish; **Fish Farming**; **Fish Meal**

PINEAPPLES

L G Smith, Maroochy Research Station, Nambour, Queensland, Australia

Introduction

The pineapple (*Ananas comosus* (L.) Merr.), botanically a member of the ornamental Bromeliaceae family, originated in tropical South America but is now widely grown in all tropical and subtropical areas of the world. In Spanish-speaking countries the fruit is known as pina; in Portuguese-speaking countries as abacaxi; as ananas in Dutch- and French-speaking former colonies, and as nanas in southern Asia. More than 4.5×10^6 t, both fresh and canned, is marketed worldwide each year from at least nine major countries (Tables 1 and 2), with the major cultivar by far being the large juicy-fruited smooth-leafed cultivar Smooth Cayenne. Wild pineapple varieties still grow in the tropical savanna of South America but most have small, seedy, fibrous fruits. (*See* **Fruits of Tropical Climates:** Commercial and Dietary Importance.)

The Plant, its Appearance, and Physiology

The pineapple is a perennial, monocotyledonous, xerophytic plant, up to 1.5 m high, of herbaceous, lily-like habit, but with tough, spiny-tipped leaves that are waxy on the upper surface and possess a fragile dusty bloom on the underside. The leaves in all but a few cultivars, such as the important Smooth Cayenne, also have numerous formidable barbs along the edges, which make cultivation hazardous. In all varieties the concave leaves channel any precipitation into the plant center for absorption by spongy leaf tissue and roots. Other features which enhance the shallow-rooted plant's adaption to low rainfall include leaves that do not wilt, and its crassulacean acid metabolism (CAM), in which the stomata open at night to take up carbon dioxide rather than in the day, greatly reducing water loss. Malic acid is accumulated during the night and is decarboxylated during the day.

The fruit of the pineapple, botanically a sorosis or a syncarp, comprises spirals of fused fleshy fruitlets radiating from a fibrous but succulent core and topped with a leafy crown or top, an extension of the peduncle or central plant stem of the plant (Figure 1). The crown is planted for reproduction, when available. The commercial fruit, grown as a monoculture of genetically identical plants, is normally seedless due to a genetic self-incompatibility. If cross-pollinated, by wind, natural vectors, or deliberately by humans, as in breeding trials, viable small hard-coated seeds 1–2 mm across will form, capable of producing a new plant and fruit within 2–3 years. The commercial plant produces one fruit 14–24 months after planting, but two or more vegetative shoots (suckers) subsequently produce additional rattoon crops. Fruits become smaller with successive rattoons and usually a maximum of two rattoon crops are commercially viable.

Because pineapple plants cannot tolerate frost or prolonged cold, production is restricted to coastal or near-coastal areas of low or moderate elevation. They tend to thrive on tropical and subtropical islands, where the surrounding water mass maintains a more ideal constant temperature and moderate humidity.

Varieties

The classification of existing pineapple varieties is currently under complete review. The Cayenne variety was previously seen as but one of five recognized groups of *A. comosus*, the other four being Spanish, Queen, Pernambuco, and Mordilona. The five groups are compared in Table 3. Most likely these will soon be reclassified to be just varieties of the one species *A. comosus*, better to reflect the true genetic

Table 1 Pineapple production in countries for years 1980–2000, listed alphabetically

	2000	*1998*	*1996*	*1990*	*1980*
World	13 455 362	12 235 777	12 136 658	11 297 972	10 836 109
American Samoa	20	20	20	20	20
Angola	32 000	38 000	38 000	32 000	36 000
Antigua and Barbuda	150	150	150	150	96
Argentina	3500	3500	3900	3500	3400
Australia	123 000	123 000	127 864	141 603	123 265
Bangladesh	148 580	148 580	148 575	162 445	153 365
Benin	78 440	44 836	89 410	25 000	3000
Bolivia	52 535	46 385	22 050	10 000	7860
Brazil	1 353 480	1 113 219	1 145 981	1 103 897	565 829
Brunei Dar-us-Salam	700	700	650	720	610
Cambodia	16 000	16 000	15 000	12 000	5600
Cameroon	48 000	52 000	48 000	35 000	34 000
Central African Republic	14 000	13 500	12 800	10 700	6600
China	1 318 450	960 982	854 113	697 178	304 004
Colombia	407 753	360 000	329 300	341 790	126 960
Congo, Democratic Republic of	196 000	204 364	201 453	160 000	127 500
Congo, Republic of	13 000	12 500	13 000	11 600	10 200
Cook Islands	50	80	100	200	1950
Costa Rica	400 000	400 000	260 000	160 000	9500
Cuba	19 000	19 000	19 000	22 000	17 832
Côte d'Ivoire	225 675	198 306	250 687	232 543	294 000
Dominican Republic	100 410	107 752	108 620	70 000	20 000
Ecuador	123 597	79 947	57 851	59 021	135 023
El Salvador	5000	4500	6000	15 405	17 870
Ethiopia PDR	2521			200	200
Fiji Islands	1632	2521	3456	3516	3000
French Guyana	3026	3026	3392	1850	80
French Polynesia				5100	1160
Gabon	700	650	650	630	
Ghana	35 000	35 000	35 000	11 000	6600
Guadeloupe	6900	6900	6900	4660	225
Guatemala	102 060	110 450	98 429	67 131	32 326
Guinea	71 858	71 858	65 000	51 072	16 500
Guinea-Bissau	200	200	200	180	50
Guyana	7000	7000	7193	7600	1860
Haiti	3300	3270	2320	2200	2000
Honduras	71 000	72 621	68 500	61 410	30 230
India	1 100 000	1 100 000	1 100 000	881 000	548 980
Indonesia	256 691	326 956	501 111	390 340	180 543
Jamaica	19 275	19 062	19 808	9332	3898
Japan	11 000	12 800	18 800	31 900	56 200
Kenya	280 000	300 000	310 000	225 000	180 000
Korea, Republic of	1500	1280	2419	9115	
Laos	34 000	34 000	32 000	32 000	32 000
Liberia	7000	7000	7000	7000	7000
Madagascar	51 000	51 000	50 000	50 100	53 520
Malaysia	143 000	143 000	163 000	213 000	185 273
Martinique	20 200	20 200	21 150	24 010	18 257
Mauritius	1200	1462	2973	1355	258
Mexico	485 597	480 856	301 406	454 668	622 729
Mozambique	13 000	18 000	16 000	15 500	13 000
Nicaragua	47 000	46 000	45 000	42 000	34 000
Nigeria	881 000	857 000	800 000	763 000	600 000
Panama	28 089	19 000	17 288	12 624	5638
Papua New Guinea	11 000	11 000	14 500	11 800	8700
Paraguay	41 000	38 872	30 676	32 710	32 344
Peru	150 000	127 910	112 835	68 423	57 605
Philippines	1 495 120	1 495 120	1 542 240	1 155 748	1 004 800
Portugal	2000	2000	2000	1500	1605
Puerto Rico	19 204	19 204	18 933	46 812	34 627
Réunion	13 000	13 000	6639	8370	3500

Continued

Table 1 Continued

	2000	1998	1996	1990	1980
Samoa	5700	5700	5700	5500	5100
Seychelles	145	140	130	110	118
South Africa	154 000	133 518	134 121	169 897	221 879
Sri Lanka	62 000	62 000	61 610	41 353	45 651
Sudan	4950	4900	4800	3500	4500
Suriname	400	425	400	255	50
Swaziland	8000	8000	7848	57 000	27 553
Tanzania, United Republic of	75 000	74 000	74 000	70 000	47 000
Thailand	2 280 959	1 786 234	1 986 700	1 865 290	3 688 400
Togo	500	500	500	500	
Trinidad and Tobago	3400	3400	3400	3400	2600
Uganda	500	500	450	400	
United States of America	319 300	301 180	315 000	521 630	596 020
Venezuela, Bolivian Republic of	182 153	252 173	175 745	80 554	77 403
Viet Nam	262 838	195 842	185 000	467 851	336 639
Wallis and Futuna Islands	4	4	4	4	4
Zimbabwe	100	90	100	100	

Data from Food and Agriculture Organization. (www.fao.org)

Table 2 The major pineapple producing countries in decreasing order, for the year 2000, showing fruit production statistics

	Production (Mt)	Area harvest (ha)	Yield ($Mt\,ha^{-1}\,year^{-1}$)
World	13 455 362	718 682	18.7
Thailand	2 280 959	96 768	23.6
Philippines	1 495 120	40 200	37.2
Brazil	1 353 480	56 031	24.2
China	1 318 450	35 300	37.3
India	1 100 000	82 000	13.4
Nigeria	881 000	115 000	7.7
Mexico	485 597	12 143	40.0
Colombia	407 753	6500	62.7
Costa Rica	400 000	13 000	30.8
Hawaii	319 300	8500	37.6
Kenya	280 000	8500	32.9
Viet Nam	262 838	29 080	9.0
Indonesia	256 691	40 000	6.4
Côte d'Ivoire	225 675	5200	43.4
Congo, Democratic Republic of	196 000	8000	24.5
Venezuela, Bolivian Republic of	182 153	9554	19.1
South Africa	154 000	6200	24.8
Peru	150 000	9795	15.3
Bangladesh	148 580	13 792	10.8
Malaysia	143 000	7580	18.9
Ecuador	123 597	12 000	10.3
Australia	123 000	2650	46.4
Guatemala	102 060	4100	24.9
Dominican Republic	100 410	2800	35.9

Data from Food and Agriculture Organization. (www.fao.org)

differences between the five groups, now known to be smaller than originally believed. The Cayenne variety is the most productive and the most widely grown variety around the world. All other varieties have inferior commercial characteristics, from inadequate yield, fruit morphology, or susceptibility to a range of diseases and disorders. Subselections or 'clones' of Smooth Cayenne have been made in different production areas of the world, and more recently many new cross-bred varieties have been produced, particularly for the fresh-fruit market. Breeding and refining new pineapple varieties are very difficult due to the extremely heterozygous nature of the genotype. Plant breeders normally back-cross new varieties to eliminate the more minor unwanted characteristics, but in the pineapple this procedure results in the reintroduction of

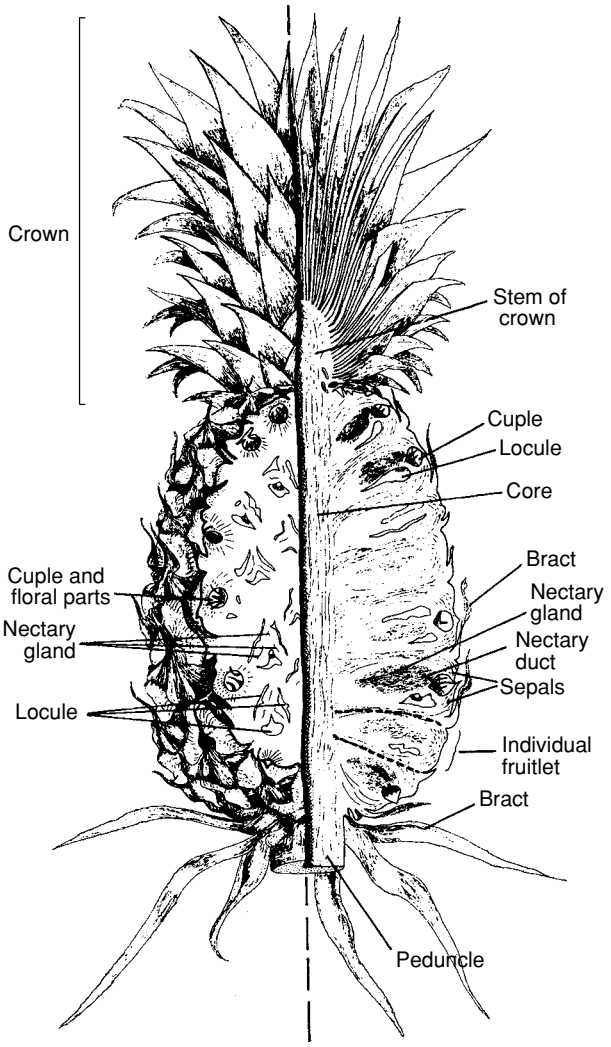

Figure 1 Mature pineapple: on the right, median longitudinal section; on the left, tangential longitudinal section. Adapted from Py C, Lacoeuilhe JJ and Teisson C (1987) *The Pineapple: Cultivation and Uses*. Paris: GP Maisonneuve et Larose, with permission. Reproduced from Pineapples, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

many unwanted genotypes. Genetic engineering is being investigated as a way around the problem of extreme heterozygosity, but to date no new genetically modified variety has been commercially released.

Composition and Nutritional Value

The pineapple fruit is typically a supplementary food rather than a staple. The pineapple's very sweet and sour taste, its mild aromatic flavor, and firm succulent flesh in a large and attractive form make it somewhat unique among foods. While very palatable by itself, it is equally used as a flavoring component, both cold and hot, making many other less attractive but nutritionally sound foods more edible. The dietary composition of Smooth Cayenne, both fresh and canned, is shown in Table 4. The nutrient profile of the pineapple is, in general, similar to many other fruits, containing high levels of carbohydrate and low levels of fat and protein. Dietary fiber constitutes about 14% of the dry matter and, as with most fruits, can be incorporated into a cholesterol-lowering diet. Its vitamin C content is about one-tenth that in citrus, and the level of retinol equivalents (3 per 100 g) is low compared to those of pawpaw (papaya) (153) and mango (356). Refer to individual nutrients.

The total sugar content is the major quality determinant of the fruit for both fresh and canned markets and in commercial practice sugar content is often regularly monitored, using a refractometer calibrated in degrees Brix which, in pineapples, is virtually identical to percentage total sugar. Fresh Smooth Cayenne fruit above 15 ° Brix would be considered of excellent eating quality by most tasters and fruit of 9–12 ° Brix of marginal quality; however, taste preferences towards Smooth Cayenne of a particular Brix level vary considerably between individuals and also between ethnic groups. In canned pineapple the sugar content of the fresh fruit makes less difference to consumer quality.

Variability of Organoleptic Parameters

Wide variations occur in the total sugar content and titratable acidity in fresh pineapples. Not only is the bottom of the fruit always about 3 ° Brix higher than the top but, more importantly, there is wide variation between samples. For example, the 8% total sugars from Queensland fruit in Table 4 is very low compared to fruit from Hawaii or Ivory Coast and is either from immature summer fruit (average mature fruit is 15–18 ° Brix) or is from winter-grown fruit (6–11 ° Brix). Fruit grown under cooler conditions is lower in sugar and higher in acid, although the sour taste of winter-grown pineapples is caused by the low sugar content rather than the high acidity.

Postharvest Changes

The postharvest changes in various parameters in fruit held near 20 °C during 15 days after harvest are illustrated in Figure 2. Total sugar concentration and eating quality remain relatively constant after harvest; acidity and carotene content increase moderately, while the shell color (degree of yellowness of the skin) and ester concentration increase substantially.

Skin Color and Ripeness

Although the fruit normally changes from green to yellow as the fruit ripens on the plant, skin color is a

Table 3 Characteristics of the major pineapple varieties[a]

	Cayenne[b]	*Spanish*	*Queen*	*Pernambuco*[c]	*Perolera*
Main production	Worldwide	Caribbean Malaysia	South Africa Australia	Brazil Venezuela	Colombia Ecuador, Peru
Weight (kg)	1.8–3.5	1–1.8	0.5–1.2	1–1.8	1.8–2.5
Shape	Cylindrical	Spherical	Conical	Conical	Cylindrical
Ripe skin color	Orange-yellow	Reddish yellow	Bright yellow	Green-yellow	Bright red-yellow
Ripe flesh	Near translucent	Near translucent	Opaque	Opaque	Opaque
Flesh fibrosity	Nonfibrous	Fibrous	Crisp	Nonfibrous	Crisp
Flesh color	Pale yellow	Near white	Bright yellow	White-yellow	Bright and pale yellow
Flavor	Sweet and acid	Spicy	Aromatic	Low acidity	Low sugar and acidity
Wilt resistance	*****	*	**	**	**
Nematodes	*****	*	**	**	**
Uses	Fresh, canned Major export	Fresh Minor export	Fresh Minor export	Fresh Minor export	Fresh Minor export

[a]The term 'varieties' is currently under review: see text.
[b]Includes Smooth Cayenne, which is but one variety of the Cayenne group.
[c]Other name abacaxi, but this should be avoided as it means 'pineapple' in Portuguese.
* resistant, ***** very susceptible.
Based on data from Py C, Lacoeuilhe JJ and Teisson C (1987) *The Pineapple: Cultivation and Uses*. Paris: GP Maisonneuve et Larose.

Table 4 Composition of 100 g of pineapple flesh

	Canned[a] *heavy syrup, drained*	*Canned*[a] *pineapple juice, drained*	*Fresh*[a], *peeled*	*Fresh*[b] *peeled*
Proximate				
Water (g)	74.8	84.6	86.0	80–86
Energy (kJ)	350	188	158	[c]
Protein (g)	0.8	0.7	1.0	0.2
Fat (g)	0.0	0.0	0.1	0
Carbohydrate, total (g)	20.4	10.2	8.0	
Sugars, total (g)	20.4	10.2	8.0	10–18
Starch (g)	0.0	0.0	0.0	
Ash (g)	0.2	0.3	0.5	0.3–0.6
Cholesterol (mg)	0	0	0	
Acids, total (g)				0.5–1.6
Total nitrogen (mg)				45–120
Pigments, xanthophylls (mainly carotenoids) (g)				0.2–0.3
Dietary fiber (g)	1	2	2	
Minerals (mg)				
Sodium	1	4	2	14
Potassium	76	140	180	11–330
Calcium	5	6	27	3–16
Magnesium	10	14	11	10–19
Iron	0.3	0.5	0.3	0.05–0.3
Zinc	0.2	0.3	0.2	0
Vitamins				
Retinol equivalents (μg)	3	3	4	
Retinol (μg)	0	0	0	
β-Carotene equivalents (μg)	18	18	25	
Thiamin (μg)	40	30	40	69–125
Riboflavin (μg)	30	30	30	20–88
Nicotinic acid equivalents (mg)	0.3	0.3	0.3	
Nicotinic acid (mg)	0.2	0.2	0.1	0.2–0.3
Vitamin C (mg)	15	14	21	3–25

[a]Queensland fruit; data from *Composition of Foods Australia*. Australian Food Publishing Service, Canberra. 1989.
[b]From Py C, Lacoeuilhe JJ and Teisson C (1987) *The Pineapple: Cultivation and Uses*, Paris: GP Maisonneuve et Larose; data presumably based on fruit from Ivory Coast.
[c]Data not given.

poor indication of palatability in Smooth Cayenne pineapples grown under commercial practices. Green-skinned fruits can taste quite acceptable. Such 'green-ripe' pineapples can either be physiologically 'unripe' (meaning physiologically not fully developed) but have sufficient sugar to taste acceptable, or their flesh may be prematurely ripened before the skin as a result of specific climactic (i.e., climate, nothing to do with the nonclimacteric fruit physiology) conditions during growth. This latter type of 'green-ripe' fruit tend to taste insipid and the condition is now regarded as a physiological disorder. On the other hand, attractive yellow-skinned fruit can taste sour. (*See* **Ripening of Fruit**.)

One way of selecting intact Smooth Cayenne fruit of uniform ripeness is on the basis of fruit specific gravity. Fruits that float in water are less ripe than those that sink. Fruits that sink in a 3% common salt solution are very ripe. This method works well for Queensland and Hawaiian fruit but reportedly not South African. The difference has not been explained.

Fresh Fruit Selection

Probably as a result of the variability of eating quality, there is a degree of folklore applied in the determination of when a harvested pineapple is 'ripe,' with potential purchasers plucking crown leaves, or squeezing, smelling, and tapping the fruit. Leaf plucking is completely useless, but fruits that feel heavy for their size will be riper than lighter fruit of the same size. However, the postharvest nonripening character of pineapple means that once the fruit is picked the eating quality is fixed. There is no advantage for consumers in waiting for the harvested fruit to yellow, as delay only results in the fruit losing volatile flavors and going 'stale.'

Factors Affecting Pineapple Eating Quality

At its best, Smooth Cayenne is a most delightful fresh fruit to eat, but the eating quality of fruits produced from different seasons and growing areas is extremely variable, much more so than other pineapple varieties. The eating quality of Smooth Cayenne is probably the most variable of any commercial fruit. It is difficult, if not impossible, even for an expert, to predict the eating quality, even when the fruit is cut. The eating quality or flavor of pineapple is almost totally dependent on the percentage of sugar in the flesh rather than the degree of visual ripeness. The sugar content is strongly affected by the temperature and light intensity during the last 3–4 weeks of growth, and, as a consequence, the sugar content in the fresh fruit is greatly affected by the season, climate, the degree of maturity at harvest and, to a lesser extent, the farm production methods.

Smooth Cayenne grown in equatorial regions generally have consistent and excellent eating quality. In the subtropical regions, where significant quantities of the Smooth Cayenne of fresh commerce are grown, the eating quality is most often very variable as a result of the daily and seasonal variations of temperature. Many subtropical production areas have developed because they are closer to areas of major urban populations and consumption, but in these areas the seasonal inconsistency of flavor is an important problem, particularly in fresh-market pineapples.

Fruit Physiology

A unique feature of the pineapple is that flowering (with subsequent fruit set) is controllable by growth hormones over an extremely wide range of plant maturity. Modern commercial production practices can regulate date of harvest to fall virtually any day of the year; thus fruit production can be precisely programmed.

Pineapples are nonclimacteric fruits and do not get better to eat (i.e., do not ripen) after harvest, in spite of any subsequent color change (Figure 2). Other members of this same group of fruits include oranges, strawberries, grapes, and cherries. The absence of starch in the fruit is often cited as the cause of the nonripening postharvest behavior, but many climacteric fruits also contain no starch. The physiology involved needs to be better clarified.

Handling, Storage, Refrigeration, and Chilling Injury

Handling

In spite of its tough appearance, Smooth Cayenne is very susceptible to injury and needs to be handled with care. The fruits will readily bruise if they fall more than 20 cm and, once bruised or injured, a ubiquitous fungal disease caused by *Ceratocystis paradoxa* (*Thielaviopsis paradoxa*), which generally spreads quickly throughout the fruit, can readily develop. Commercially, fruit are commonly treated with fungicides such as thiabendazole, benomyl, or triadimefon. In contrast to Smooth Cayenne, Queen fruits develop fewer postharvest rots and in many ways are ideal fresh-market fruits. They keep well without refrigeration and, if free from developing blackheart, have a much longer shelf-life than Smooth Cayenne. (*See* **Fungicides**.)

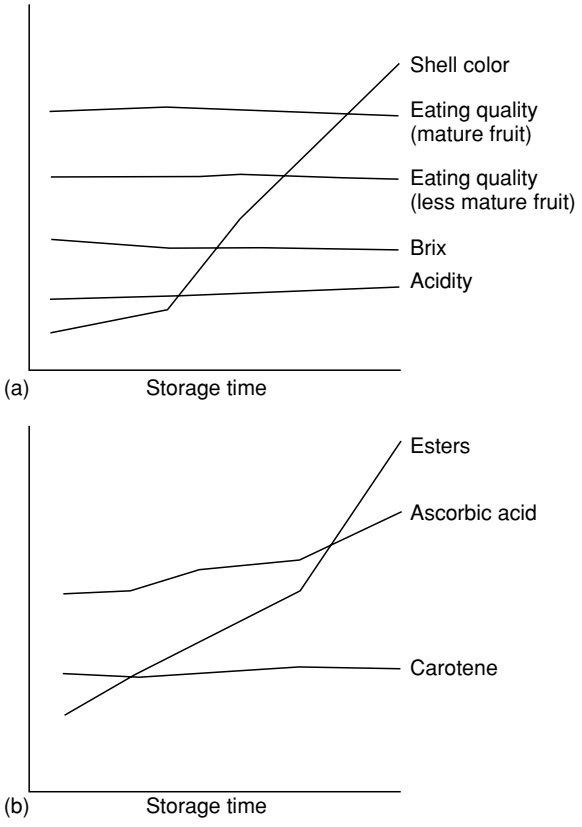

Figure 2 Postharvest changes in chemical and sensory parameters over 15 days. Eating quality data from the author (unpublished). Other data from Singleton (1959), as quoted by Dull GG (1970). The pineapple: general. In: Hulme AC (ed.) The Biochemistry of Fruits and their Products, vol. 2, pp. 303–324. London: Academic Press. Reproduced from Pineapples, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

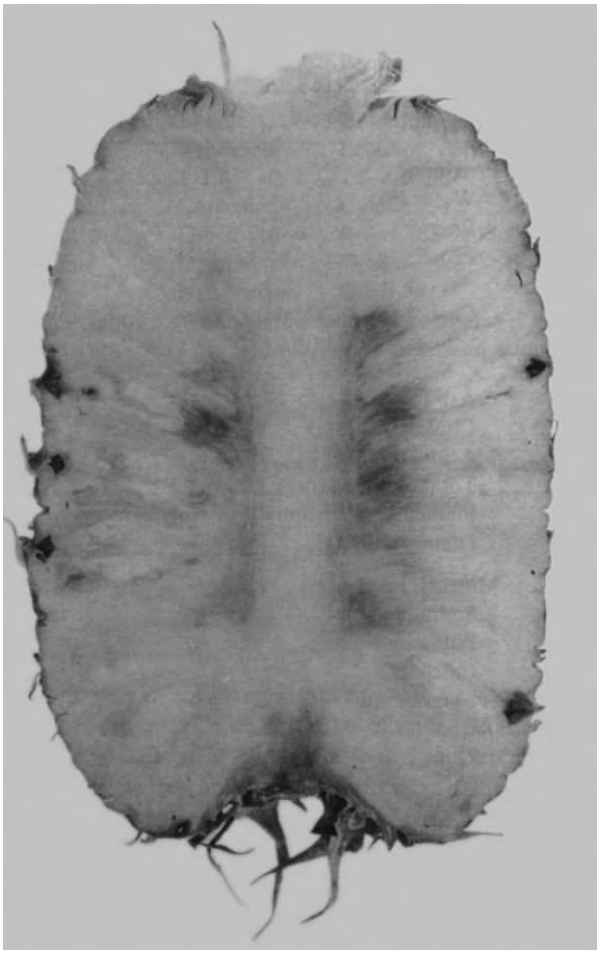

Figure 3 Blackheart or internal browning in pineapple, a serious disorder in fresh-market pineapples which can develop following refrigeration, but which also naturally occurs in subtropical production areas. It is a physiological disorder (i.e., not a pathogenic disease), known as chilling injury and caused by prolonged exposure to temperatures below 21 °C. Reproduced from Pineapples, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Storage and Refrigeration

Although pineapples can keep reasonably well without refrigeration, particularly if the humidity is kept high to prevent desiccation, refrigeration is necessary to maintain a glossy, fresh-looking skin and to reduce the incidence of rots. However, pineapples do not outturn well if kept refrigerated for longer than 2 weeks. They can suffer from 'chilling injury,' a physiological injury caused by cool temperatures, resulting in disruption of the cellular enzymic processes. It is a disorder characteristic of many tropical fruits which will often develop within 2–3 days after removal from refrigeration. In pineapples, chilling injury is manifest most commonly by the fruit developing brown spots adjacent to the core; these are sometimes called endogenous brown spots. With increasing severity, the spots darken further, and in severe cases the complete center of the fruit can become black, giving rise to the other name for the disorder, blackheart (Figure 3).

Chilling injury is a constant threat when pineapples are shipped under refrigeration, and intermittent heavy losses occur. Storage for 3–12 days at temperatures as high as 21 °C have been shown to induce blackheart, but injury incidence is unpredictable, possibly dependent on the intensity of available sunlight during the 3–4 weeks prior to harvest. Temperatures of 15–18 °C induce the greatest injury but, as the fruits are usually shipped at 8 °C for 7–10 days from tropical production areas, they are usually consumed before any injury develops.

Fruits from the Queen group are very susceptible to chilling injury but a Hawaiian variety, known as 53/116, appears to be entirely resistant. This hybrid variety has several quality deficiencies (often

developing a cannonball shape, or lacking a crown), but because of its good blackheart resistance and lo\w acidity, it is slowly developing wider acceptance in Australia. Dipping the fruit in a carnauba-based wax, or applying a lipid-carboxymethyl cellulose coating, can control blackheart. These skin coatings function by reducing the oxygen concentration in the flesh, necessary for the melanin pigment to develop, but the anaerobic flesh can develop an alcoholic flavor. Heating the fruit to 40 °C for 24 h either immediately before or after cool storage also reduces blackheart and improves pineapple eating quality, but is not commercially used because of logistical limitation.

Pineapples that are cool-stored for prolonged periods can also develop other symptoms of chilling injury such as flaccid watery flesh, skin necrosis, crown wilting, and crown necrosis, but fruit with such injuries are often already damaged by internal chilling. Overcoming chilling injury in fresh-market pineapples is currently under investigation.

Controlled-Atmosphere Storage

Controlled-atmosphere storage does not prolong the life of the cool stored pineapples and is not used for international shipping. (*See* **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs.)

Field Blackheart

Pineapples from cooler production areas are particularly susceptible to blackheart, where it can develop in the field following periods of cool overcast weather. Because the disorder shows no external symptoms, affected fruit cannot be identified. In Queensland and Hawaii pineapple canneries close or restrict intake during such periods and thus avoid excessive losses.

Pineapple Products

Canned pineapple is a commodity of major international trade; figures showing exports and imports are shown in Table 1. Pineapples are also processed into a number of other products other than canned fruit. Pineapples are often used as a component of commercially marketed fresh fruit salad, either fresh or frozen. Less ripe fruits freeze best as they have firmer flesh. Soft ripe fruits can develop off-flavors when frozen, as well as turning out very flaccid. Fried pineapple pieces are commonly used in Asian cuisine. Dried pineapple pieces are easily prepared and are being sold in increasing quantities as they retain a more attractive appearance than other fruits that strongly darken when dried. Sliced thin and dried, pineapple 'chips' are also sold as a snack food. Glacé pieces are readily prepared from either the fresh product or, more easily, from canned rings. Pineapple juice is a major item of commerce, as fresh or frozen, full strength, or as a concentrate. Reasonably high in carbohydrates, fermented juice is used to make vinegar. It does not make a good-quality wine but, fermented and distilled, it makes a potent, if somewhat disreputable, alcoholic spirit, available in several Pacific island countries. Cannery waste (crushed skins, etc.) makes a good cattle feed supplement, while old plantation fields are often let out to cattle. Pineapple leaves are used to make cloth or rope in both Taiwan and the Philippines, while use of the whole plant as a source of energy is under current investigation. (*See* **Canning**: Quality Changes During Canning; **Freezing**: Structural and Flavor (Flavour) Changes.)

Bromelain

Pineapple fruits contain a proteolytic enzyme, bromelain, which can constitute nearly half the measured protein in the fruit. Much higher concentrations of the basic isoenzyme (the fruit form is acidic) exist in the stem. Stem bromelin is commercially extracted and sold, as a buff-colored powder, slightly soluble in water and practically insoluble in alcohol, chloroform, and ether. Bromelin is used in the pharmaceutical industry for digestive and antiinflammatory products, in the manufacture of cattle feed and to 'chill-proof' beer (to prevent formation of a haze of proteinaceous material which can form when brewed products are refrigerated). Some recent pharmacological reports indicate that bromelin may have value in modulating tumor growth and blood coagulation, as well as debridement of third-degree burns.

Commercial Farm Production

Modern commercial culture of pineapples is generally capital-intensive, requiring substantial land preparation, machinery for planting, boom (fertilizer) spray and harvesting, plus bulk equipment for cannery fruit and/or harvester-packers for fresh market fruit.

By using a combination of plant hormones and appropriate planting practices, the harvest date and the fruit size and quality can be preset and regulated as required. This makes modern pineapple production function with almost clockwork precision. On modern fresh-fruit farms a year-round cycle is maintained of weekly, if not daily, planting and harvesting. Combined with modern planting machinery, modern herbicides, and large mobile harvesting equipment, production of 1000–2000 t of fruit per staff year is common.

Being parthenocarpic (seedless, but otherwise normal), pineapple propagation is vegetative and can be problematical as only limited planting material is available from any one plant. The best planting material, that producing the most uniform plants and fruits, is the fruit crown, but if the fruit is sold as fresh fruit (as opposed to canning), tops are not available, so that suckers and other vegetative material must be collected and used. Up to 80 000 pieces of planting material (tops, suckers or slips) are required per hectare, planted in double rows. The time of planting and the size and type of planting material affect the date of harvest and fruit size, and are predetermined. Weeds can be readily controlled as the plant is resistant to several effective herbicides (e.g., diuron, fluazifopbutyl, ametryne) which are applied over the rows of growing plants with a boom spray. Ten to fourteen months after planting, a flower-inducing hormone (commonly the ethylene-releasing compound Ethrel, 2-chloroethyl-phosphonic acid) is applied, and the treated plants uniformly produce a single, blue-petalled inflorescence about 8 weeks later. Fruit development occurs without further interruption, but as sunburn can be a serious injury, fruit may be individually covered (bagged) or sprayed with a reflective bentonite clay suspension.

Harvesting pineapples at the correct stage of internal ripeness is important as it is the major factor that determines the fruit quality at consumption. As skin color is an inconsistent guide to eating quality, pickers have to use past experience and market demands.

Sodium naphthaleneacetic acid (NAA) is sometimes applied as a spray on the fruit to increase fruit size by delaying degreening (becoming yellow). Adverse effects on fruit quality from using NAA are reportedly minor but are under current debate. Ethrel is also sometimes sprayed on both fresh and cannery crops to increase flesh yellowness and to 'accelerate ripening.' While very effective and cost-saving on cannery fruit, its use on fresh-market fruit has been shown to be of questionable value as the variability of the eating quality is greatly increased.

See also: **Canning**: Quality Changes During Canning; **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Freezing**: Structural and Flavor (Flavour) Changes; **Fruits of Tropical Climates**: Commercial and Dietary Importance; **Fungicides**; **Ripening of Fruit**

Further Reading

Bartholomew DP and Paull RE (1986) Pineapple. In: Monselise SP (ed.) *Handbook of Fruit Set and Development*, pp. 371–388. Boca Raton, Florida: CRC Press.

Dull GG (1970) The pineapple: general. In: Hulme AC (ed.) *The Biochemistry of Fruits and their Products*, vol. 2, pp. 303–324. London: Academic Press.

Py C, Lacoeuilhe JJ and Teisson C (1987) *The Pineapple: Cultivation and Uses*. Paris: GP Maisonneuve et Larose.

Smith LG (1988) Indices of physiological maturity and eating quality in Smooth Cayenne pineapples. *Queensland Journal of Agricultural and Animal Sciences* 45(2): 213–228.

PINE KERNELS

B T Styles†, University of Oxford, Oxford, UK

Background

Pine nuts, also known as Indian nuts, piñons, or pignolias, have been an important food crop in some areas of the world since prehistoric times. The 'nut' is, in fact, the seed of different species of pine (*Pinus*), nearly all of which belong to a group known as 'soft' or 'white' pines or their relatives. These species are evergreen, coniferous trees whose cones are softly woody and have few scales and two large seeds per scale that lack a wing (see **Figure 1**). The kernel consists of the endosperm tissue of the seed containing the stored food material and the developing embryo (germ). The 'shell' surrounding it is the testa. This must be removed before the kernel can be eaten.

†Deceased.

Sources

Pine nut kernels are almost all obtained from wild forest trees. Piñons from the drier areas of southwest USA and Mexico form the largest wild source, but the Italian pignolia tree has been cultivated and protected in the Mediterranean region for centuries. Almost all the other species grow in upland mountainous areas, but information about them is scanty (**Table 1**). The harvest of pine nuts is only a fraction of the production of cultivated nuts such as pecan, macadamia, walnut, or filbert, and the crop is generally irregular, with bumper harvests occurring approximately every

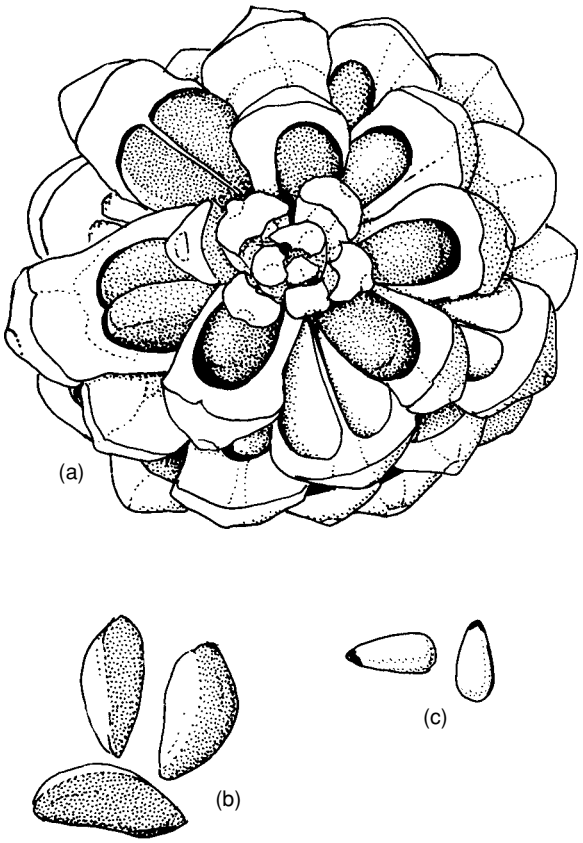

Figure 1 Pignolia (*Pinus pinea*): (a) cone; (b) seeds; (c) polished kernels (all natural size). Reproduced from Pine Kernels, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 1 Sources of pine kernels (*Pinus* species and their distribution)

Group	*Latin name*	*Distribution*
Pion pines	*Pinus edulis*	Southwestern states of USA and Mexico
	P. monophylla	
	P. quadrifolia	
	P. maximartinezii	Mexico
	P. cembroides	
Stone pines		
Italian (pignolia)	*P. pinea*	Mediterranean basin and Turkey
Japanese	*P. pumila*	Northeast Asia, Siberia to Korea and Japan
Korean	*P. koraiensis*	Korea, southeast Siberia and Japan
Siberian	*P. siberica*	Western Russia to Siberia and Mongolia
Swiss	*P. cembra*	Alps and Carpathian Mountains
Chilgoza pine	*P. gerardiana*	Eastern Afghanistan to northern India and Pakistan

5 years. Harvesting is chiefly by an itinerant labor force that moves into the pine forests during the autumn. Green cones are cut from the trees, dried in the sun and opened to release the seeds, or the seeds are collected from beneath the trees. The collection and later preparation can at best be considered as a cottage industry. The kernels of only two or three of the piñon pines and the pignolia pine are important in commercial trading on a large scale. They may be bought in health food shops or delicatessens but are considered expensive, £12.50 per kilogram in the UK.

Kernels of most of the other pines may be found on sale in local markets near the site of collection. There are no reliable statistics on yields of these wild crops, but it has been estimated that about a million kilograms of Colorado pine nuts were collected annually in southwest USA. This quantity has probably not changed a great deal since then.

Storage and Preparation

After harvesting, nuts should be kept unshelled, in a dry, cool, well-ventilated place in cloth or paper bags. Tannins in the shell and in the seed coat of the kernel may function as antioxidants in preserving the fat and oil in the nuts. Fresh kernels stored in closed containers soon go moldy or rancid, but after drying, they can be safely preserved and have excellent keeping qualities for up to 3 years. They freeze well when fresh. Shelling of the nuts can be done domestically by lightly crushing the nuts on a cloth with a rolling pin. Those of the thin-shelled, single-leaf pine can be easily removed with the fingers. The thicker Colorado piñon, stone pine, and pignolia nuts must be cracked mechanically to release the kernels. (*See* **Antioxidants**: Natural Antioxidants; **Oxidation of Food Components**.)

Although the kernels can be eaten raw, roasting is necessary to bring out their full flavor. They may be roasted with or without the shell, the amount of time required depending on shell thickness and moisture content.

Composition and Nutritional Value

The richness and flavor of pine kernels are well known, and they rank highly among all other nuts in food value. There are, however, considerable differences in the properties of the kernels of different species. A comparison of the nutritional value of pine kernels and some commercial nuts is provided in Table 2.

Refuse and Wastage

When compared with other nuts, piñons have low percentages of shell waste or refuse. Shell thickness

varies considerably: 30–35% of seed weight of the very-thin-shelled single-leaf piñon is made up of shell, whereas the thicker-shelled Colorado pine has a waste factor of 42%. Mediterranean pignolia has a particularly thick shell that must always be removed before it is sold. Thus, pine kernels of piñons constitute some 58–70% of the edible portion. The average is lower than that found in all other commercial nut crops except peanuts.

Protein

The average protein content is about 15% for piñon nut kernels and about 34% for pignolias. It is higher than for pecan nuts and about the same as that for English walnut and Brazil nut. It is only greatly exceeded by peanuts. A study in the former Yugoslavia has shown that kernels of pignolias are richer in proteins than pork and goose meat. Kernel protein has a digestibility value almost as good as that of beef and is distinctly better than almost all other nuts. It is near the proportion required for a balanced diet. Colorado and single-leaf kernels are especially rich in tryptophan and cystine. (*See* **Protein**: Requirements.)

Fat (Lipid) and Oil

Piñon pine kernels average about 60% of fatty materials. This is lower than pecan, English walnut, and Brazil nut, which approximate to 70%. Pignolia kernels in the former Yugoslavia are reported to contain 48% of lipid, which is again higher than for fat pork (37%) and goose meat (44%). As Table 2 indicates, the Siberian stone pine is also very rich in fats and oil. This is processed commercially for the production of cooking oil, but no published details of production methods or quantities are available. The fats of piñons, particularly Colorado and single-leaved pine, contain monounsaturated oleic, and polyunsaturated linolenic and linoleic acids. It has been reported that pignolia kernels from the Mediterranean region contain up to 50% of linoleic acid. Commercially produced pine kernels studied contained 46.40 g of fat by weight, of which 6.12 g is saturated (fatty acid not named). (*See* **Fats**: Requirements.)

Table 2 Comparison of the nutritional value of pine kernels and some commercial nuts

Type of nut	*Food content*		
	Protein (%)	*Fat (%)*	*Carbohydrate (%)*
Colorado piñon, Pinus edulis	14	62–71	18
Single leaf piñon, *P. monophylla*	10	23	54
Mexican piñon, *P. cembroides*	19	60	14
Parry piñon, *P.* × *quadrifolia*	11	37	44
Digger pine, *P. sabiniana*	30	60	9
Pignolia pine, *P. pinea*	34	48	7
Siberian stone pine, *P. siberica*	19	51–75	12
Chilgoza pine, *P. gerardiana*	14	51	23
Pecan, *Carya illinoensis*	10	73	11
Peanut, *Arachis hypogaea*	26	39	24
English walnut, *Juglans regia*	15	68	12
Almond, *Prunus dulcis*	21	54	7
Brazil nut, *Bertholletia excelsa*	16	69	8

Typical energy value per 100 g = 556 kcal.
Data from Lanner RM (1981) *The Pion Pine*. Reno, Nevada: University of Nevada Press and Botkin CW and Shires LB (1948) The Composition and Value of Pion Nuts. Bulletin 344, New Mexico Experiment Station.

Carbohydrate

The average oily piñon kernel of Colorado pine contains 19% carbohydrate, but this can increase to as high as 54% in the more starchy kernels of the single-leaf pine. In the Parry pine, it is 44%, but the former is much preferred. (*See* **Carbohydrates**: Requirements and Dietary Importance.)

Other Substances

Pine kernels are extremely rich in phosphorus (6040 mg kg^{-1}) which is about the same as for soya bean, and iron (53 mg kg^{-1}). They also contain significant amounts of vitamin A, thiamin, riboflavin, and niacin. Refer to individual nutrients.

Uses

In the past, pine kernels were a staple component in the diet as well as a subsistence food, but because of their enormous versatility, they now hold a high profile in modern cuisine among people of the developed nations, particularly in the USA, Europe, Asia, and North Africa. As roasted nuts, they are very widely used in the preparation of soups, sauces, and dressings, and may be served as a garnish for fish meals, in mixtures with cooked meat and as part of side-dishes with rice. They form suitable ingredients in cakes, puddings, and biscuits, and as a garnish for icecream. They are also highly appreciated in fruit desserts and vegetable salads. The kernels are so nutritious that a number of recipes have recently appeared under the label of 'Backpackers' Friends.' These products bear such names as Piñon Pemmican and Granola, and may be carried as bars in the rucksack as emergency rations or snacks. Pine kernels will probably become even more popular as the rise in vegetarianism continues.

See also: **Antioxidants**: Natural Antioxidants; **Carbohydrates**: Requirements and Dietary Importance; **Fats**: Requirements; **Oxidation of Food Components**; **Protein**: Requirements

Further Reading

Botkin CW and Shires LB (1948) *The Composition and Value of Piñon Nuts*. Bulletin 344, New Mexico Experiment Station.

Budzynska J (1964) Note sur la composition en acides gras de l'huile des noix de *Pinus pinea*. *Revue Française des Corps Gras* 11: 143–145.

Kaic M (1985) Prehrambenotenoloske osobine sjemenki oraha (*Juglans regia*), crnog oraha (*Juglans nigra*) i pinije (*Pinus pinea*). *Sumarski List* 109: 325–328.

Lanner RM (1981) *The Piñon Pine*. Reno, Nevada: University of Nevada Press.

Little EL (1938) *Food Analyses of Piñon Nuts*. Research notes, Southwestern Forest and Range Experiment Station, Arizona.

Rowe DB, Blazich FA and Weir RJ (1999) Mineral nutrient and carbohydrate status of loblolly pine during mist propagation as influenced by stock plant nitrogen fertility. *Hortscience* 34: 1279–1285.

Siew WL (2001) Crystallisation and melting behaviour of palm kernel oil and related products by differential scanning calorimetry. *European Journal of Lipid Science and Technology* 103: 729–734.

Zakaria S, Hamzah H, Murshidi JA and Deraman M (2001) Chemical modification on lignocellulosic polymeric oil palm empty fruit bunch for advanced material. *Advances in Polymer Technology* 20: 289–295.

Pituitary Hormones *See* **Hormones**: Adrenal Hormones; Thyroid Hormones; Gut Hormones; Pancreatic Hormones; Pituitary Hormones; Steroid Hormones

PLANT ANTINUTRITIONAL FACTORS

Contents

Characteristics

G D Hill, Lincoln University, Canterbury, New Zealand

Background

The growing of plants as a major food source for humans only started about 10 000 years ago with the birth of agriculture. At that stage, human intervention disrupted the chain seed – plant – seed. Because the seeds, tubers, and fruits of a number of plant species are rich in carbohydrates, protein, and fats, plant materials now provide a substantial proportion of the food energy that is consumed by humans and their domestic animals. When that process happened, a number of chemical compounds occurring naturally in plants, and often with a role in protecting the plant from attack by animals – be they insect, ruminant herbivore, or monogastric – became antinutritional factors (ANFs).

Other articles will deal in detail with most of these ANFs, which are sometimes called secondary metabolites. This chapter will consider the plant species involved and their interaction with animals and humans when they are eaten. The chapter will also consider briefly how the effects of ANFs can be overcome. It will not cover their exact chemical structure, their determination, or their biosynthesis.

Plant Antinutritional Factors

Antigens

A number of legume species, but particularly soybean (*Glycine max*) and peanut (*Arachis hypogaea*), produce severe antigenic reactions in both people and animals. Affected persons can go into anaphylactic shock, and unless treatment is quickly applied, death can follow. In a recent case in the USA, a person who was allergic to peanuts died after licking his fingers having touched a chicken leg coated with peanut sauce. Similarly, some airlines now have peanut-free flights. Other plant foodstuffs that are reported to induce allergenic reactions include rice (*Oryza sativa*), hazelnuts (*Corylus avellana*), apples (*Malus sylvestris*), and celery (*Apium graveolens*).

Compounds responsible The proteins that are thought to be responsible for the development of allergenic reactions are generally water-soluble glycoproteins that are resistant to breakdown by heat, acid, and digestive enzymes. In soybean and peanuts, the glycoproteins are thought to be the storage proteins, glycinin and β-conglycincin. These proteins resist digestion and are absorbed into the intestinal mucosa. The body responds by producing immunoglobulins or antibodies in the Peyer's patches of the gut and their associated lymphoid tissue. These antibodies are then released into the gut and react with the antigens to prevent their absorption. Antigens that escape absorption cause inflammation of the intestinal mucosa.

Three mechanisms are involved. An immediate mechanism that involves IgE antibodies and mast cells (type I hypersensitivity), semidelayed, which is brought about by antigen–antibody complexes (type III), and delayed caused by specific activated T-lymphocytes (type IV). The last is so called because there can be a delay of several days between the antigen challenge and the onset of observable symptoms. (*See* **Food Intolerance**: Food Allergies.)

Symptoms Affected animals show poor growth and diarrhea, and may die. In the gut, there are changes in the structure of the mucosa, increased gut motility, and reduced absorption of electrolytes (Na^+, K^+, and Cl^-). Once animals are sensitized, they can remain so for a considerable period. In laboratory animals, chronic stimulation by antigens has given reduced disacchiridase activity, reduced surface area of the brush borders of the microvilli, reduced microvilli height, crypt hyperplasia, cell degranulation, and increased numbers of mast cells.

Minimization of problems In humans, the development of food allergies seems to be related to age. Among young children, up to 8% can suffer from food allergies. In adults, the rate of occurrence is considered to be $< 1\%$. Thus, as people get older, they seem to be less prone to the development of food allergies. Where products such as soybean are used as milk substitutes, it is possible to treat the material by procedures such as steam heating to reduce the level of antigenicity.

α-Galactosides

It is probably not strictly correct to refer to the α-galactosides as ANFs, as they are not toxic. However, their presence in plant foods can cause severe digestive upsets in monagastrics that are embarrassing in humans and can lead to severely reduced growth rates in animals such as pigs.

Compounds responsible The α-galactosides or oligosaccharides comprise the sugars raffinose, stachyose, verbascose, and ajugose. They are simple sugars comprising a sucrose molecule to which are attached α-D-galactopyranosyl units in α-1,6-galactosidic linkages. They are a common carbohydrate present in the seed of grain legumes. It is thought that on seed germination, they provide energy for the emerging plant embryo.

Symptoms When members of the raffinose group of sugars are consumed by monogastrics, the animals do not have the necessary gut enzymes to digest these sugars in the intestinal tract. Consequently, they pass undigested to the large intestine where they are anaerobically fermented by gut bacteria. The products of this fermentation are, among other things, the gases carbon dioxide, methane, and hydrogen. Thus, consumption of legumes rich in these sugars leads to a considerable increase in flatus production. Some legumes can produce up to 137 ml h^{-1} of flatus, causing severe abdominal discomfort, bloating, belching, flatulence, constipation, and diarrhea. In pigs, in particular, high levels of oligosaccharides in their rations can cause severe digestive upsets and reduced growth.

Minimization of problems There is considerable interspecific variation in the level of oligosaccharides in legume seed. Soybean, chickpea (*Cicer arietinum*), and *Phaseolus* spp. tend to contain high levels, whereas lentil (*Lens culinaris*) and mung bean (*Vigna radiata*) contain lower levels. There is also some evidence of intraspecific variation among cultivars within species.

Traditional methods of processing grain legumes prior to their consumption significantly reduced the level of oligosaccharides. In the production of bean sprouts, the oligosaccharide level falls as germination continues. *Tofu*, which is a traditional protein precipitate produced from soybean, has low levels of oligosaccharide. In the production of *tempe*, which is a traditional Indonesian soy-based food, cooked beans are fermented with *Rhizopus oligosporus*. During fermentation, the oligosaccharides are utilized by fungus as an energy source, and the resulting *tempe* contains low levels of oligosaccharides.

Traditional legume-processing methods cannot be used for bulk quantities, however, because of the cost of processing. In the stock food industry, the problem is approached by the plant breeders trying to reduce the level of these sugars in the seed using selection within the species.

Alkaloids

A number of plant species contain significant amounts of alkaloids. Of the alkaloids, the most universally

known is probably nicotine, which is a component of tobacco (*Nicotiana tabacum*) and is ingested on a daily basis by millions of people. Edible members of the plant family Solanaceae, such as potato (*Solanum tuberosum*), tomato (*Lycopersicon esculentum*), and and eggplant (*Solanum melongena*), may contain high levels of glycoalkaloids. (*See* **Alkaloids**: Toxicology.)

Among legumes that are eaten by humans, or used in stock rations, plants of the Genus *Lupinus* contain high levels of alkaloid in both their foliage and their seed. In most lupin species, the alkaloids present are quiniolizidine alkaloids. However, *Lupinus luteus*, which is from the Iberian Peninsula, contains the indol alkaloid gramine.

Compounds responsible The alkaloids present in food plants of the family Solanaceae are glycoalkaloids. They have a stereoalkaloid nucleus, which is attached to a sugar side-chain. Their role is to protect the plant from insect attack. They are also implicated in both fungal and nematode resistance. The alkaloids tend to be concentrated near the epidermis and in the leaves. In green tomatoes, the fruits are protected while ripening by the alkaloid tomatine. As the fruit matures, the tomatine level falls.

In potato, the major alkaloids present are α-solanine and α-chaconine. They are present throughout the plant. Apart from a case in Scandinavia, where plant breeders increased the alkaloid levels in a new cultivar to the level where they produced toxic symptoms in humans, potato alkaloids are not a problem, provided that the tubers have been properly grown and stored. If potato tubers form too close to the soil surface, or are exposed to light after harvest, there is a rapid increase in alkaloid level. Harvested potatoes should be stored in the dark. In modern shopping systems, potatoes are often exposed to light in supermarkets when they are packed in clear plastic bags.

Among lupin species that are grown for their seed, there are a large number of quinolizidine alkaloids (Table 1). However, only eight of these occur in amounts of more than 4% of total alkaloids. The main alkaloids are sparteine, lupanine, 13-hydroxylupanine, albine, multiflorine, angustifioline, tetrahydrorhombifoline, and lupinine. The distribution of the alkaloids tends to vary with plant species. Most of the alkaloid in *Lupinus albus, L. angustifolius*, and *L. mutabilis* is lupanine. In *L. luteus*, the major alkaloid present is lupinine. The gramine in *L. luteus* (an indol alkaloid) can comprise up to 0.12% of total seed dry matter.

Symptoms Both potato and lupin alkaloids affect the nervous system. An excess of potato alkaloids can produce symptoms similar to a bad hangover. There have been deaths reported from the ingestion of excess potato alkaloids. Ingestion of lupin alkaloid leachate by a human has been known to produce blurred vision, an irregular heart beat rate, and urine retention. In animal feeding, the response to foods high in alkaloids is a loss of appetite, because of their bitter taste. This induces a reduction in feed intake and reduces growth rate.

Minimization of problems Fortunately, the alkaloids in both potatoes and lupins are water-soluble. Further, in potatoes, they are concentrated under the skin. Peeling the potato removes most of the alkaloid in the tuber. Soaking and boiling in water further reduce alkaloid levels. In both the altiplano of South America and southern Europe, *L. mutabilis* and *L. albus* are processed to reduce alkaloid levels. The seed is soaked and boiled and then soaked in running water for up to three days, by which time the alkaloid levels are significantly reduced. As noted above, with potatoes, correct storage in the dark, after harvest and prior to use, is also important.

Table 1 Distribution of total alkaloids present in smooth seeded lupin species used in human diets and animal feeding

	Lupin species			
Alkaloid	L. albus	L. angustifolius	L. luteus	L. mutabilis
Sparteine	Trace	Trace	30–50%	5–20%
Lupanine	50–80%	50–80%	Trace	50–70%
17-Oxylupanineo	Trace	Trace	[a]	
4-Hydroxylupanine				7%
13-Hydroxylupanine	5–15%	10–20%		10–20%
Albine	5–15%			
Multiflorine	3–10%			Trace
Angustifoline	Trace	5–20%		Trace
Tetrahydrorhombifoline	Trace	Trace	Trace	4%
Lupinine			40–70%	Trace
Esteralkaloids	1–5%	1–3%	Trace	1–5%

[a]Not detected.

In lupin, since the early 1920s, an alternative approach has been to select for low-alkaloid genotypes. Since then, because of work by plant breeders in Germany, Australia, and South America there are now low-alkaloid genotypes of *Lupinus albus, L. angustifolius, L. luteus*, and *L. mutabilis*. In Australia, low-alkaloid genotypes of rough-seeded Mediterranean lupins have now been bred. Thus, lupin seed is a safe food, and alkaloids are not likely to be a problem, provided the seed is processed before consumption or only low-alkaloid types are used.

Cyanogenic Glycosides

A number of plant species produce hydrogen cyanide (HCN) from cyanogenic glycosides when they are consumed. These cyanogens are glycosides of a sugar, often glucose, which is combined with a cyanide containing aglycone.

Plant species of major importance in human and animal feeding are cassava (*Manihot esculenta*), linseed (*Linum usitatissmium*), various sorghums (*Sorghum* spp), white clover (*Trifolium repens*), and some species of *Lotus*. Lesser quantities are found in the kernels of such plants as almonds (*Amygdalus communis*), apricots (*Prunus armeniaca*), peaches (*Prunus persica*), and apples (*Malus sylvestris*).

When plant material containing the glycoside is consumed, it is broken down by a β-glucosidase to produce a sugar and an aglycone. The aglycone is then acted upon by a hydroxynitrile lyase to produce cyanide and an aldehyde or a ketone.

Compounds responsible The compounds responsible vary with plant species. However, the major cyanogens are amygdalin and prunasin, which are found in fruit kernels. The latter also occurs in bracken fern (*Pteridium aquilinum*), dhurrin, found in members of the genus *Sorghum* and linamarin, found in clovers, linseed, cassava, and lima beans (*Phaseolus lunatus*). Cassava root contains relatively low levels at 53 mg of CN per 100 g of plant tissue. Sorghum forage contains 250 mg of CN per 100 g of plant tissue and lima beans up to 300 mg of CN per 100 g of plant tissue.

Symptoms As cyanide is extremely toxic, one of the most obvious symptoms is death. In the body, cyanide acts by inhibiting cytochrome oxidase, the final step in electron transport, and thus blocks ATP synthesis. Prior to death, symptoms include faster and deeper respiration, a faster irregular and weaker pulse, salivation and frothing at the mouth, muscular spasms, dilation of the pupils, and bright red mucous membranes.

In Africa, where many people consume cassava on a regular basis, many members of the human population are regularly exposed to low levels of cyanide in their diet. This is associated with a condition called tropical ataxic neuropathy. Symptoms include neurological disturbances, which affect vision, hearing, and the peripheral nervous system. There are also raised levels of blood thiocyanate and goiter. Cassava consumption, combined with protein deficiency, which is often common in societies that consume large amounts of starchy tubers, can lead to reduced glucose tolerance and diabetes. There is some evidence that the symptoms can be partially alleviated by the administration of vitamin B_{12} and methionine.

Minimization of problems As with many other problems of plant ANFs, one method of reducing the problem is via plant breeding. Older varieties of white clover used to produce high levels of cyanide. Newer cultivars have much lower concentrations. Similarly, breeding work has been carried out with lima beans to reduce their potential toxicity.

To produce cyanide, plant enzyme systems need to be active, and so heating plant material, and denaturing the enzymes, will generally render it safe. Thus, in the production of linseed meal, which is the residue left over after oil extraction, the material is generally heated sufficiently, while passing through an oil press, to render the enzyme system inactive. The use of entire linseed seeds in bakery products usually leads to their being exposed to high temperatures before they are eaten. In home consumption of cassava, cooking is used to inactivate the enzymes. Roots are also washed, grated, and soaked in water, and further washed to eliminate the cyanide. As with clover and lima beans, plant breeders are also trying to reduce the level of the cyanogen in cassava roots.

Glucosinolates

Glucosinolates are a common component in plants of the family Brassicacea. They have become of more interest in recent years following the considerable increase in the production of rape oil (Canola) from the seed of spring and winter forms of *Brassica campestris* and *B. napus*. Following oil extraction in these, and other *Brassica* spp., a high-protein meal is left, which contains high levels of sulfur amino acids. However, the presence of glucosinolates in the resulting meal has limited the use of the meal in animal feeding because of its tendency to produce goiter and other symptoms associated with low iodine availability. (*See* **Glucosinolates**.)

Compounds responsible Glucosinolates are glycosides of β-D-thioglucose. They contain an aglycone, which, on hydrolysis, can yield an isothiocyanate, a

nitrile, or a thiocyanate. Glucosinolates are responsible for the sharp flavor of such plants as horseradish (*Armoracia rusticana*) various mustards (*Brassica* spp), watercress (*Nasturtium officinale*), and radish (*Raphanus sativus*).

In plant tissues, the glucosinolates are present at the same time as the enzyme myrosinase, which is a thioglucosidase. In intact plants, the enzyme and the substrate occur in different parts of the plant. However, when plant tissues are disrupted, as may occur in chewing, or in pressing and grinding for oil extraction, the two components come together. As a result, the myrosinase breaks the thioglucoside bond on the glucosinolate. The result is the production of glucose and an aglucone intermediate product. The latter breaks down to produce a number of possible toxic components. The most common are isothiocyanates and nitriles. However, this depends on the conditions at the time of breakdown, and other products, such as goitrin, may be produced.

Symptoms The major effect of the breakdown products of the glucosinolates is interference with thyroid function. Goiter can occur in both humans and animals, but it mainly appears in animals when they are fed on rations in which rapeseed meal provides a considerable amount of the protein in the ration. Besides obvious morphological effects on the thyroid gland of animals, differences can be shown in biochemical parameters such as decreased serum thyroxin levels. In pigs, when sows are fed on high-glucosinolate rations, it can lead to abortion of piglets and increased death of piglets after birth.

Minimization of problems As with many other ANFs, moist heating the meal prior to feeding it to the animals or heat extrusion can reduce problems associated with the feeding of glucosinolates in *Brassica* meals. Bacterial fermentation and enzymes have also been used. However, the most significant recent achievement in reducing the problem was wrought by Canadian plant breeders who produced new cultivars of both major oils seed rape species, which contain considerably reduced levels of glucosinolates. These were originally known as '00' rapes but have since been renamed Canola. This major plant breeding advance has meant that considerably greater amounts of *Brassica* meal can now be incorporated into monogastric rations without any deleterious effects on thyroid function and iodine metabolism.

Lectins

Lectins, or phytohemagglutinins, are present in the seed of a number of legume species that are consumed by humans. The most important legume species, which is high in lectin, is *Phaseolus vulgaris*, because of its major role in the diet of central and northern South American populations. *Phaseolus vulgaris* lectin is highly toxic, as is the lectin from scarlet runner bean (*Phaseolus coccineus*) and tepary bean (*P. acutifolius*). At the other extreme, legume species with low to zero lectin activity include chick pea (*Cicer arietinum*), cow peas (*Vigna unguiculata*), faba bean (*Vicia faba*), lentil (*Lens culinaris*), mung bean (*Vigna radiata*), peas (*Pisum sativum*), pigeon pea (*Cajanas cajan*), and soybean (*Glycine max*). (*See* **Hemagglutinins (Haemagglutinins)**.)

In the plant, it appears that lectins are involved in the chemical recognition of the legume root by *Rhizobium* bacteria. Thus, although plant breeding may be used to reduce their absolute concentration in seed, it is unlikely that they could be entirely eliminated from legume seeds because of their importance in biological nitrogen fixation.

Compounds responsible Lectins are either carbohydrate-binding proteins or glycoproteins. They are capable of recognizing and binding carbohydrates in complex glycoconjugates. More than 70 lectins have now been isolated from legumes. They usually contain two or four subunits, each with a single carbohydrate-binding site.

Symptoms The main effect of lectins, when they are included in the diet, is through their interference with digestion in the small intestine. Lectins are resistant to breakdown in the digestive tract and bind to surface receptors. These bound lectins then cause changes in the metabolism of the epithelial cells. These changes include cell hypertrophy and hyperplasia. Indeed, animals fed high levels of toxic lectin can suffer rapid body weight loss, while the weight of the small intestine is significantly increased. Other changes that may occur are changed gut endocrine systems and hormone production, changed gut immune systems and disturbances in gut bacterial ecology, which, combined, can cause a greater sensitivity to bacterial infection. At the gross level, these effects are mediated via reduced growth, diarrhea, reduced nutrient absorption, and increased incidence of bacterial infections. In highly toxic species such as the red kidney bean (*Phaseolus vulgaris*), these symptoms, when combined, cause death.

Minimization of problems As indicated above, it is unlikely that it will be possible to breed the lectins out of legume seed entirely because of their role in the establishment of biological nitrogen fixation. However, it may be possible to select for reduced levels and/or to select for less toxic lectins. Fortunately,

most lectins are heat-labile. Dry *Phaseolus* beans require long soaking and cooking times before they are soft enough to eat, and therefore, for human consumption, most seed is rendered safe before it is consumed. However, there are reports that if beans are heated at 80 °C, lectin levels may be increased, and thus the beans are more, not less, toxic.

For animal feeding, however, lectin-containing seed needs to be processed to reduce lectin levels before the seed can be fed to monogastric animals. Generally, seed can be rendered safe by heat treatment prior to consumption. Other techniques that have been employed include chemical treatment, production of protein isolates, germination, and moist extrusion.

Saponins

Saponins are a diverse group of chemicals, which derive their name from their ability to form soap-like foams in aqueous solutions. They occur in a considerable number of plant species ranging from asparagus (*Asparagus officinalis*) to cucumber (*Cucumis sativus*). They are also present in a number of commonly used herbs and spices such as fenugreek (*Trigonella foenumgraecum*), ginseng (*Panax* spp.), liquorice (*Glycyrrhiza glabra*), and nutmeg (*Myristica fragrans*). (*See* **Saponins**.)

Saponins fall in a gray area between being ANFs or beneficial plant constituents. When monogastric animals such as pigs and poultry are fed rations containing such products as alfalfa (*Medicago sativa*) meal, there is a reduction in growth. However, experiments with a number of animal species have failed to show any significant effect of saponins on a number of growth parameters. As saponins are very bitter, one hypothesis is that their effects are mediated through reduced food intake.

On the positive side, it has been shown that saponins can form complexes with cholesterol in the gastrointestinal tract. This leads to increased excretion of cholesterol and a reduction in blood cholesterol level. With the high level of coronary artery disease in most developed countries, this is seen to have potential medical benefits. The most common form of ingestion by humans is through alfalfa sprouts. Recent publicity would suggest that the sprouts are more dangerous because of bacterial contamination, leading to gastroenteritis, than because of their saponin content.

Compounds responsible All saponins contain an aglycone, which is either sapogenol or sapogenein. The aglycone is linked to one or more sugars or oligosaccharides. However, the saponins fall into two groups, depending on whether the aglycone is triterpenoid or steroidal. Most cultivated plants have a triterpenoid aglycone, whereas in many herbs, it is steroidal. Saponin's structure is complex and depends on the variation in the aglycone structure and the position and nature of attachment of the glycosides to the molecule. Aglycones can be linked to D-glactose, L-arabinose, L-rhamnose, D-glucose, D-xylose, D-mannose, and D-glucuronic acid. The chain lengths are generally linear and comprise two to five saccharide units. This diversity in structure leads to a diversity of biological activity.

Symptoms The major physiological effect of saponins is on cell membrane permeability. It has been proposed that the saponins combine with membrane cholesterol to form permeable micelles in the plane of the membrane. However, as indicated above, the major effect of saponins in reducing intake appears to be the result of their bitter flavor-reducing feed intake, as experiments on a range of animal species have not shown any pathogenic effects of saponin intake.

Minimization of problems Saponins that are present in herbs and spices are likely to be present in such small quantities in the diet that they are unlikely to have any deleterious effect. Alfalfa has been selected by plant breeders to contain low levels of saponins. Similarly, quinoa (*Chenopodium quinoa*) seed, which is a minor Andean food crop, has reported saponin levels ranging from 0.14 to 2.3%. This has allowed for the selection of low saponin genotypes. In the Andes, in households, saponins are removed by soaking, washing, and rubbing to reduce the saponin levels in quinoa grain. Industrially, the seed is milled, or washed and milled.

Tannins

Many plant species contain tannins. However, the seed of a number of legume species and some sorghum lines contain appreciable quantities of tannins. These have dietary impacts on animals that consume them. Tannin is a generic name for polyphenolic compounds derived from plants. They have molecular weights of between 500 and 3000 and are divided into two classes, hydrolyzable and condensed. The former can be broken down by acid, alkali, and some hydrolytic enzymes. The latter are resistant to being broken down.

Faba bean (*Vicia faba*) tannins are involved with the human disease favism. The inclusion of high levels of tannin in rations for a range of monogastric animals reduces weight gain and feed-conversion efficiency. When fed to hens, tannins reduce egg production. However, the consumption of forage with high tannin levels can be advantageous in ruminant animal production. Tannins prevent the formation of the stabilized foams that produce bloat. Further, tannins

form complexes with free protein in the rumen and prevent protein degradation in the rumen. (*See* **Tannins and Polyphenols**.)

Compounds responsible Hydrolyzable tannins have a polyhdric alcohol at their core, the hydroxyl groups of which are partially, or fully, esterified with either gallic or hexahydroxydiphenic acid. They may have long chains of gallic acid coming from the central glucose core. On hydrolysis with acid or enzymes, the hydrolyzable tannins break down into their constituent phenolic acids and carbohydrates.

Condensed tannins are dimers, the simplest of which is procyanidin, or higher oligomers of substituted flavan-3-ols. Monomeric tannins are usually linked by carbon–carbon bonds between carbon-4 and carbon-8 of two flavan-3-ols. The bonds that are formed are highly stable and require heating with strong acids to break them down.

Symptoms In humans, favism produces acute hemolytic anemia. After susceptible subjects eat the beans, symptoms can occur in 5–24 h. The symptoms include headache, vomiting, nausea, yawning, stomach pains, and a raised temperature. The symptoms may subside naturally or, in severe cases, lead to hemolytic anemia, followed by hemoglobinuria. Previously, deaths in children could reach 8%. The disease is due to the lack of an enzyme in red blood cells, glucose-6-phosphate dehydrogenase. It is thought that over 100 million people, of mainly eastern Mediterranean origin, are susceptible to the disease. There has been some suggestion that the high levels of tannins in the diet of people who use sorghum as their main staple may be associated with an increased incidence of esophageal cancer.

The condensed tannins provide an important plant defense against herbivory. In particular, bird damage can be reduced by high tannin levels in crops such as sorghum and insect damage in seed of species like faba beans. The adverse nutritional effects of tannins are due to their reaction with proteins to form indigestible complexes. They can also form complexes with digestive enzymes and reduce the overall digestibility of food. They can irritate the gut lining and stimulate the secretion of mucus. This increases endogenous protein secretion and therefore increases protein demand. They also form complexes with divalent metals and reduce mineral absorption. A common feature of diets high in tannin is weight loss.

Minimization of problems In most legume species, there is a strong relationship between the level of condensed tannin in the seed and their flower color. Plants that have white flowers have low tannin levels in their seed. Thus, in a number of the legume species that have high tannin levels, the development of white-flowered genotypes gives pale-colored seed with low tannin levels.

In both legume seed and sorghum seed, the tannins are located mainly in the seed coat. Thus, dehulling the seed and discarding the hulls reduce the tannin level. Heat treatments do not destroy tannins in sorghum. However, sorghum tannins are unstable under alkaline conditions. Soaking in an alkaline solution followed by washing reduces the level of condensed tannins. The tannins in sorghum can also be neutralized by treatment with anhydrous ammonia. In contrast, in faba beans, the principal thermolabile antinutritional factor is the condensed tannins.

Toxic Amino Acids

In the legume family, there are two toxic amino acids that have a major negative effect on animals or humans that consume them. The tropical tree species *Leucaena leucocephala* contains the amino acid mimosine, which is an amitotic agent. Seed of the hardy annual legume species *Lathyrus sativus* contains a neurotoxic amino acid, generally referred to as ODAP or BOAA. The latter can cause paralysis of the lower limbs when eaten in excess.

Compounds responsible The full chemical name of mimosine is β-[1-(3-hydroxy-4-pyridone)]-α-amino propionic acid. The amino acid ODAP is β-*N*-oxalyl-L-α-β-diamino-propionic acid.

Symptoms Mimosine is an amitotic agent, and one of its major effects is to stop hair growth. It has been tested for a possible role in the chemical shearing of sheep. *Leucaena* seed, which is high in mimosine, is usually only used to a limited extent in human diets as a curry sambal in Indonesia.

The effects of *Lathyrus sativus* seed consumption are somewhat more problematic. Consumption of large amounts of *Lathyrus* seed in the diet often occurs in countries such as India and Bangladesh at times of food shortages as a result of drought. The condition produced is neurolathyrism, which is paralysis of the lower limbs caused by damage to nerve cells in the spinal cord. There is some evidence that the response to *Lathyrus* consumption may be sex-linked, as it seems to have a major effect on young males.

Minimization of problems With regard to mimosine, there have been some attempts to breed genotypes of *Leucaena leucocephala* that are low in mimosine. As it is such a minor component in human diets, it is unlikes to be a major problem. During

World War II in Hawaii, *Leucaena* leaf meal was used in both pig and poultry rations as a substitute for alfalfa meal. Provided that the levels that were fed were less than 10% there were no ill effects. Finally, ruminants can modify the spectrum of their rumen microflora to convert mimosine to 3,4-dihydroxypyrididine, which appears to have no biological activity.

With regard to *Lathyrus sativus* consumption, three approaches have been taken to reduce the problems associated with its consumption. In India, the crop has been banned. However, it apparently has a pleasant flavor and is a common adulterant of red gram (pigeon pea dhal). There is some suggestion that since the ban, the total area sown to the crop has increased. Therefore, the two following methods of dealing with the problem would appear to have a greater potential.

As with many other legume species that contain ANFs, they can be removed by traditional processing methods such as soaking, boiling, and fermentation. The major problem is that often in times of famine, it is not only food that is in short supply. Reduced amounts of wood for cooking often mean that food is cooked for less time, and the ANFs are therefore not destroyed to the same extent. At the same time, because of the drought resistance of the mother plant, it may become the only grain legume that is available, so the total amount consumed, and thus the dose of ODAP, is increased.

In Bangladesh, *Lathyrus sativus* is the most important single pulse crop and is grown on about 82 000 ha. To this end, plant breeders in Bangladesh, in association with workers from Belgium and Canada, have selected lines of *Lathyrus* with low levels of ODAP. The aim is to produce cultivars that are safe to eat with no, or a reduced, requirement for processing.

Trypsin Inhibitors

Trypsin inhibitors that inhibit the activity of the enzymes trypsin and chymotrypsin in the gut, thus preventing protein digestion, are found in many plant species. These species include a range of grain legumes such as common bean (*Phaseolus vulgaris*), cowpea (*Vigna unguiculata*), Lima bean (*Phaseolus lunatus*), peanut (*Arachis hypogaea*), peas (*Pisum sativum*), soybean (*Glycine max*) and winged bean (*Psophocarpus tetragonolobus*). However, they are also found in cereals such as wheat (*Triticum aestivum*) and barley (*Hordeum vulgare*), potatoes (*Solanum tuberosum*), and a number of species in the genus *Cucurbita*. Because of the extensive use of soybean meal in monogastric feeding, the trypsin inhibitors associated with this plant species have been studied most extensively. (*See* **Trypsin Inhibitors**.)

When monogastric animals are fed rations, which contain large amounts of raw soybean meal, there is poor growth, poor hair and feather production, and digestive disturbances. However, there is recent evidence to suggest that consumption of food containing trypsin inhibitors may have a role in combating breast cancer in humans.

Compounds responsible There are many plant-derived trypsin inhibitors. Most of these inhibitors differ in their specificity. Many can inhibit one or two enzymes. Different forms of inhibitor may be present in the same seed, and most of the inhibitors can inhibit trypsin, but they may also inhibit chymotrypsin. However, in legume seed, the two most important inhibitor families are the Kunitz trypsin inhibitor family and the Bowman–Birk trypsin inhibitor family.

The Kunitz inhibitor family was the first family to be isolated. It is a peptide comprising 181 amino acids containing two disulfide bridges with a molecular weight of about 21 000 Da. As this inhibitor inhibits trypsin stoichiometrically to form a stable complex, it is known as a single-headed inhibitor. It primarily inhibits trypsin, but it can weakly inhibit chymotrypsin. It is inactivated by heat and by gastric juices.

The Bowman–Birk inhibitor family is widely distributed in legume seed. It is a smaller peptide molecule and contains 71 amino acids. It contains a high level of cystine and has seven disulfide bridges. The molecular weight is about 8 000 Da. It is a double-headed molecule and inhibits both trypsin and chymotrypsin at two different binding sites. Bowman–Birk inhibitors are resistant to gastric juices and to proteolytic enzymes. There is also a suggestion that they may be resistant to breakdown by heat.

Symptoms Animal response to trypsin inhibitors in the diet varies with animal species. In chicks, rats, and mice, the inhibitors cause pancreatic hypertrophy and increased pancreatic secretion. However, they do not have these effects on pigs, dogs, or preruminant calves. Most monogastrics have reduced growth when fed on rations with high levels of raw soybean meal. Protein digestibility may be reduced, and dietary protein is excreted in the feces. There is also reduced nitrogen and sulfur absorption. One of the effects of the inactivation of digestive enzymes in the intestine is the stimulation of trypsin and chymotrypsin secretion from the pancreas, which can create an increased demand for the sulfur amino acids methionine and cystine. In turn, this leads to increased endogenous loss of both nitrogen and sulfur. Finally, the inhibitors can stimulate the release of

cholecystokinin into the blood stream, which further increases pancreatic secretion.

A number of factors can modify the effect of trypsin inhibitors in the diet. These include the animal species, the age of the animal, other ANFs present in the diet, and the type and level of protein in the diet. It appears that human trypsins are not inactivated by either family of the soybean trypsins.

Minimization of problems As with many other ANFs, the level of trypsin inhibitors can be reduced by moist heat treatments. Loss of activity is associated with temperature, duration of heat treatment, moisture conditions, and particle size. Up to 95% of activity can be destroyed by heating at 100 °C for 15 min. Autoclaving for 15–30 min has also been shown to destroy most trypsin inhibitor activity in a range of grain legume species. However, the results are not universally applicable. In the industrial production of soybean oil and meal, the heating caused during the process substantially reduces the level of trypsin inhibitor activity.

Plant breeders have been able to show that there is considerable variation in the level of inhibitor activity among genotypes of the same legume species. Recently, in peas, breeders have produced isogenic lines that differed only in their level of seed trypsin inhibitor. Similarly, there are also inhibitor-free lines of soybean. Thus, the breeding of plants with reduced levels of trypsin inhibitors in their seed is possible. However, as with a number of other ANFs, it is not clear what role the inhibitors serve in the possible protection of the plant. Until this is resolved, given that they are mainly rendered inactive by the processing of the seed, it may be better to leave the inhibitors in the seed.

Conclusions

It appears that our ancestors, and people in many societies in the developing world today, have been well aware of the ANFs that are present in the plant material eaten. They have long established methods of reducing the risks involved by one or more of the processes of soaking, testa removal, heating, fermenting, germinating, or protein extraction to produce isolates. However, these systems do not always work, and in many developing countries, the problems of already-poor nutrition are exacerbated by the presence of ANFs in the diet.

In the developed countries, in recent years, there has been a swing away from meat-based diets towards the consumption of more legume seed. A result has been an increased number of cases reported where people have become ill after eating legume seed that had not been fully processed. It has been suggested that this is because the need to remove the ANFs is no longer common knowledge.

In the processing of some plant materials to produce products like Canola or soybean oil, the heat that is produced during crushing and milling may be sufficient to render the resulting press cake safe to feed to nonruminants, including humans. Industrial producers prefer to use products that require as little extra processing as possible. To try to meet this demand, plant breeders have been actively trying to reduce the levels of ANFs in plant products. However, the downside is that this often renders plants more susceptible to attack by birds, insects, and fungi while they are growing, and to increased depredations of stored products by insects, particularly weevils.

It is possible that, if the controversy over the use of GM-modified plants for human food can be resolved, plant breeders could remove the ANFs from food plants. The partial protection that the ANFs currently provide could be obtained by the insertion of one or a few specific genes.

See also: **Bioavailability of Nutrients**; **Carbohydrates**: Classification and Properties; **Galactose**; **Glucosinolates**; **Goitrogens and Antithyroid Compounds**; **Heat Treatment**: Chemical and Microbiological Changes; **Legumes**: Legumes in the Diet; **Phytic Acid**: Nutritional Impact; **Hemagglutinins (Haemagglutinins)**; **Plant Antinutritional Factors**: Detoxification; **Sucrose**: Properties and Determination; **Tannins and Polyphenols**; **Trypsin Inhibitors**

Further Reading

Bruneton J (1999) *Toxic Plants Dangerous to Humans and Animals*. Paris: Èditions TEC & DOC.

Cheeke PR (1998) *Natural Toxicants in Feeds, Forages and Poisonous Plants*, 2nd edn. Danville: Interstate Publishers.

D'Mello JPF, Duffus CM and Duffus JH (eds) (1991) *Toxic Substances in Crop Plants*. Cambridge: Royal Society of Chemistry.

Huisman J (1990) *Antinutritional Effects of Legume Seeds in Piglets, Rats and Chickens*. PhD thesis, Wageningen Agricultural University, The Netherlands.

Jansman AJM, Hill GD, Huisman J and Poel AFB van der (eds) (1998) Recent advances of research in antinutritional factors in legume seeds and rapeseed. In: *Proceedings of the 3rd International Workshop on 'Antinutritional Factors in Legume Seed and Rapeseed'. Wageningen, July 1998*. Wageningen: Wageningen Pers.

Liener IE (ed.) (1980) *Toxic Constituents of Plant Foodstuffs*. New York: Academic Press.

Nartey F (1977) Manihot esculenta *(Cassava): Cyanogenesis, Ultrastructure and Seed Germination*. DSc thesis, University of Copenhagen, Denmark.

Raffauf RF (1996) *Plant Alkaloids: A Guide to their Discovery and Distribution*. New York: Food Products Press.
Roberts MF and Wink M (1998) *Alkaloids: Biochemistry, Ecology, and Medicinal Applications*. New York: Plenum Publishing.
Shahidi F (ed.) (1997) Antinutrients and phytochemicals in food. In: *ACS Symposium Series 662*. Washington, DC: American Chemical Society.
Waller GR and Nowacki EK (1978) *Alkaloid Biology and Metabolism in Plants*. Newyork: Plenum Press.
Wink M (ed.) (1999a) Biochemistry of plant secondary metabolism. In: *Annual Plant Reviews*, vol. 2. Sheffield: Sheffield Academic Press.
Wink M (ed.) (1999b) Functions of plant secondary metabolites and their exploitation in biotechnology. In: *Annual Plant Reviews*, vol. 3. Sheffield: Sheffield Academic Press.

Detoxification

I E Liener, University of Minnesota, St. Paul, MN, USA

Introduction

Proteins of plant origin, particularly oil seeds and legumes, provide a valuable source of protein for humans and animals, despite the fact that they contain substances that can adversely affect the nutritional quality of the protein (**Table 1**). Fortunately, in many cases, these substances are heat-labile and can be readily eliminated or inactivated by the heat treatment involved in domestic cooking and commercial processing, or by treatment with chemicals. In other cases, potential toxicants remain innocuous until they are acted upon by enzymes of endogenous origin. Paradoxically, advantage can sometimes be taken of the action of these enzymes to detoxify certain plants employing such traditional modes of food preparation as germination or fermentation.

Protease Inhibitors

Proteins capable of inhibiting mammalian digestive enzymes, such as trypsin and chymotrypsin, are widely distributed in nature and have been shown to retard the growth of animals by virtue of their ability to interfere with the digestion of dietary protein. The animal tends to adapt to this situation by stimulating the secretory activity of the pancreas as well as its size. In extreme cases, such as the long-term ingestion of these inhibitors, the pancreas may exhibit precancerous lesions.

Because of its economic importance, the soya bean, which is very rich in protease inhibitors, has received the most attention with respect to the means whereby these inhibitors can be inactivated. Because they are proteins, they can be denatured by heat treatment, and this is accompanied by a loss in inhibitory activity. In general, destruction by heat treatment is a function of the temperature, duration of heating, particle size, and moisture content, conditions that are carefully controlled in order to minimize thermal damage to the nutritive value of the protein, which may occur if excessive heat is applied. An example of the relationship between the destruction of the trypsin inhibitor and the concomitant improvement in the nutritional value of the protein in shown in **Figure 1**. It is reassuring to note that most commercially available soybean products intended for human consumption such as tofu, soy protein isolates and concentrates, and textured meat analogs have received sufficient heat treatment to reduce the trypsin inhibitor activity to nontoxic levels.

Although the treatment of soya beans and other legumes by boiling in water or exposure to live steam (sometimes referred to as 'toasting') are the most commonly employed methods for inactivating the trypsin inhibitor, other modes of heat treatment or processing have proved to be equally effective. These

Table 1 Examples of naturally occurring toxicants of plant origin, their distribution and physiological effects

Toxicant	*Distribution*	*Physiological effect*
Protease inhibitors	Most legumes	Impaired growth, enlarged pancreas
Lectins	Most legumes	Impaired growth
Goitrogens	Cabbage family	Hyperthyroidism
Cyanogens	Lima beans, cassava	Respiratory failure
Phytate	Most plants	Interference with mineral availability
Tannins	Most legumes	Interference with protein digestion
Oligosaccharides	Most legumes	Flatulence
Mimosine	*Leucaena*	Goiter
β-*N*-oxalyl-α,β-diamino-propionic acid	*Lathyrus sativus*	Lathyrism

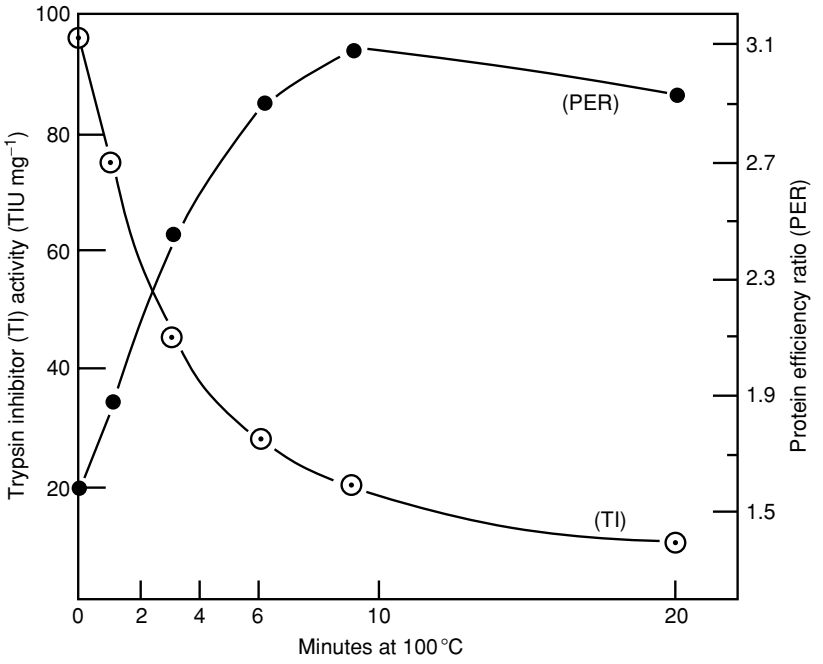

Figure 1 Effect of heat treatment on the trypsin inhibitor (TI) activity and protein efficiency (PER) of soya beans. PER=gain in weight (g)/protein intake (g). Courtesy of Academic Press, New York.

include dry roasting, microwave cooking, γ irradiation and infrared radiation.

Traditional methods of preparing various foods derived from beans generally result in products that are quite low in protease inhibitor activity. For example, tofu comprises mainly protein that has been precipitated from a hot-water extract of soya beans with calcium–magnesium salts, and soya milk is simply a hot-water extract of soya beans that may have been clarified by filtration. In both cases, the boiling or steaming of the soya beans prior to extraction with water serves to inactivate the inhibitors. The same holds true for fermented preparations of soya beans and other legumes such as tempeh, miso, and natto in which the beans have been subjected to boiling prior to the fermentation step.

The trypsin inhibitors are protein molecules whose activity is retained only if the disulfide bonds of the cystine residues remain intact. Thus, cleavage of these disulfide bridges by treatment of soya beans with reducing agents (cysteine, *N*-acetylcysteine, mercaptoethanol, or reduced glutathione) or with sodium metasulfite causes inactivation of the inhibitors.

A search for varieties of soybeans that might be devoid of trypsin inhibitors has led to the indentification of a cultivar that has only about half of the inhibitory activity of most commercial varieties of soya beans. This is due to the absence of the Kunitz trypsin inhibitor, the Bowman–Birk inhibitor, presumably accounting for the remainder of the activity. This particular variety of soya bean has the economic advantage of requiring less heat than conventional soya beans in order to produce a product of comparable nutritional value for animal feeding. (*See* **Plant Antinutritional Factors:** Characteristics.)

Lectins

Paralleling the distribution of protease inhibitors present in oil seeds and legumes is a class of proteins referred to as 'lectins'. These are proteins that exhibit the unique property of being able to bind to specific sugars that are part of the structure of so-called glycoproteins. Their antinutritional effect lies in the fact that, by binding to the glycoproteins located on the surface of the cells lining the small intestines, they interfere with the absorption of nutrients; the result is a failure in growth and eventual death.

The toxic effects of the lectins can be effectively eliminated by heat treatment, essentially under the same conditions as those that inactivate the protease inhibitors. However, there are documented cases where beans that have been eaten raw or only partially cooked have led to serious outbreaks of

gastrointestinal illness. For example, bean-containing stews and casseroles, when cooked in a slow cooker where a relatively low temperature is maintained for a long period of time, retain sufficiently high levels of lectin to cause illness.

With the availability of the soya bean variant devoid of the Kunitz trypsin inhibitor (see above) as well as one that was found to be free of the lectin, it became possible to compare the relative contribution of these two antinutritional factors to the poor nutritive value of raw soya beans. In comparison with conventional raw soya beans, a greater improvement in the growth performance of chicks was noted for the Kunitz inhibitor-free soya beans than that obtained with the lectin-free soya beans. These results may be taken to indicate that the Kunitz trypsin inhibitor is a more important antinutritional factor than the lectins. It is important to note, however, that heat treatment is necessary for the inactivation of both of these factors in order to achieve a maximum nutritive value.

Goitrogens

Goiter-causing agents in the form of glycosides (also referred to as 'glucosinolates') are found in members of the cabbage family, which includes not only cabbage but also broccoli, Brussels sprouts, cauliflower, turnips, rapeseed, and mustard seed. These compounds are biologically inactive as long as they remain bound to glucose ((**1**) Figure 2), but an enzyme, myrosinase, present in the same plant, serves to release the active goitrogenic principle, goitrin ((**2**) Figure 2). The glucosinolates found in the cabbage family appear to pose little risk to human health since the enzyme responsible for the release of goitrin is inactivated by household cooking. The glucosinolates may also be removed to a great extent by leaching out into the cooking water. The common usage of rapeseed meal in the feeding of livestock may prove to be toxic unless it has been treated with moist heat. Alternative methods of detoxification of rapeseed may involve prior extraction of the glucosinolates with water or acetone or by decomposition with iron salts or soda ash. These procedures, however, do not preclude the possibility that some goitrin produced enzymatically prior to processing may still remain in the meal. A more effective means of detoxification is one in which an aqueous slurry of the meal is deliberately allowed to undergo autolysis, which serves to liberate virtually all of the goitrin; the latter is then removed by extraction with water or acetone. Lactic acid fermentation or treatment with a specific fungus (*Geotrichum candidum*) has also been reported to be an effective means for the biological destruction of the glucosinolates in rapeseed meal. The immobilization of myrosinase on a solid matrix offers a promising approach for the hydrolysis of the glucosinolates, provided the goitrogenic end products are subsequently removed by extraction, as described above. (*See* **Plant Antinutritional Factors**: Characteristics.)

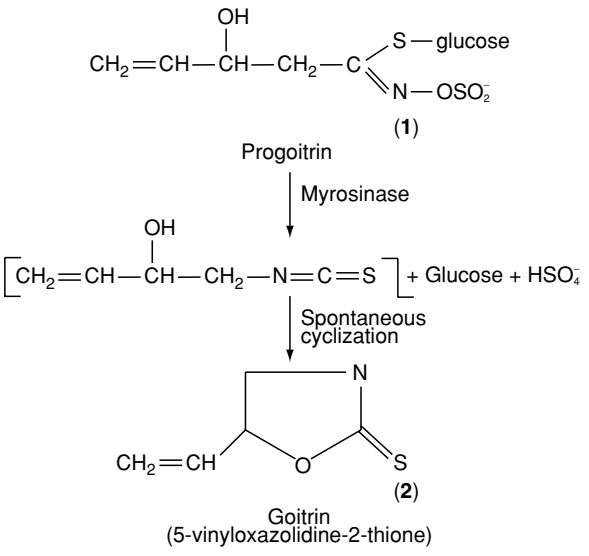

Figure 2 Structure of one of the goitrogenic compounds present in the cabbage family. Progoitrin (1) is biologically inactive, but, upon hydrolysis by the enzyme myrosinase, the active principle goitrin (2) is released.

Cyanogens

Many plants are potentially toxic because they contain glycosides, principally linamarin ((**3**) Figure 3), from which hydrogen cyanide may be released by the action of an endogenous enzyme, linamarinase, when the plant tissue is macerated. Among the cyanogenic

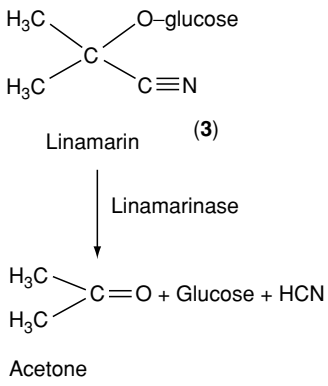

Figure 3 Structure of the major cyanogen present in lima beans. Enzymatic hydrolysis of linamarin (3) releases acetone, glucose and hydrogen cyanide.

plants most likely to be consumed by humans are the lima bean and cassava. Although some tropical varieities of the lima bean may contain toxic levels of cyanide, those varieties consumed in the USA and Europe generally have levels of cyanide below the dosage known to be toxic to humans. Cassava is a staple food item in the tropics and often contains toxic levels of cyanide unless properly processed. The traditional method of preparing cassava involves the removal of the peel, which is particularly rich in linamarin, followed by a thorough washing of the pulp with running water. A further reduction in toxicity is achieved by the application of heat (boiling, roasting, or sun drying), which serves to inactivate linamarinase and to volatilize any hydrogen cyanide that may have been produced. The cyanide content of casava can also be reduced by a lactic acid fermentation to produce a native Nigerian dish called gari. The hydrogen cyanide released by the enzymes elaborated by the fermenting agent is then eliminated by frying. The use of an exogenous source of linamarinase derived from one of several species of fungi and bacteria has also been proposed for enhancing the detoxification of gari. An appreciable degree of detoxification of cassava pulp for use in the feeding of livestock can be achieved by ensilage.

Phytate

Phytic acid ((**4**) Figure 4) or its salt, phytate, is inositol combined with six phosphate groups and is a common constituent of most plants. Its antinutritional effect lies in the fact that it forms a chelate with metal ions such as calcium, magnesium, zinc, and iron to form poorly soluble compounds that are not readily absorbed from the intestine, thus interfering with the bioavailability of these essential minerals. The ability of phytate to bind metal ions is lost when the phosphate groups are removed by hydrolysis through the action of phytase. Heat alone is relatively ineffective in reducing the phytate content of plant materials, but the phytate content can be reduced by taking advantage of the endogenous phytase that accompanies the phytate in separate compartments of the plant tissue. For example, by the simple expedient of allowing ground soya beans to undergo autolysis under conditions that are optimal for phytase activity, one can obtain a significant reduction in phytate content. An exogenous source of phytase is commercially available as a feed additive to livestock and poultry diets. This not only serves to eliminate the mineral binding ability of the phytate that may be present from other plant ingredients of the diet, but also makes more phosphate available to the animal.

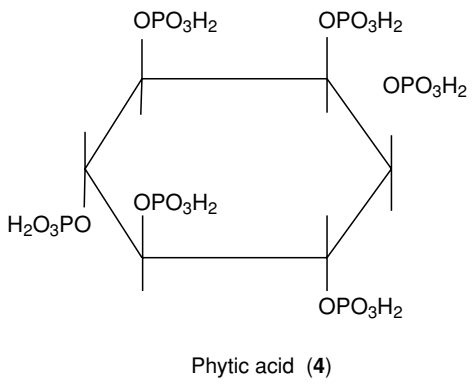

Phytic acid (**4**)

Figure 4 Structure of phytic acid (4). Enzymatic removal of the phosphate groups by phytase serves to eliminate its metal-binding property.

Traditional fermented dishes derived from such plants as soya beans, cassava, rice, maize, cowpeas, or yams have reduced levels of phytate, presumably due to the action of phytase produced by various molds, bacteria or yeasts involved in the fermentation. Even the use of yeast in breadmaking serves to reduce the phytate content of wheat flour. The germination of mature seeds of various legumes results in a great increase in phytase activity with a concomitant reduction in phytate. Other techniques for removing phytate such as ultrafiltration and ion exchange chromatography have been proposed, but these are unlikely to replace the simpler methods already described. (*See* **Phytic Acid**: Properties and Determination; Nutritional Impact.)

Tannins

Oilseeds and legumes contain appreciable levels of polyphenol compounds, broadly referred to as 'tannins' ((**5**) Figure 5). The negative nutritional effects of tannins are diverse and incompletely understood, but the major effect is to interfere with the digestibility of dietary protein. This may be due to the binding of the tannins to the protein to form substrates that are resistant to digestive enzymes or to a direct binding to these enzymes themselves.

Since tannins are located primarily in the seed coat of dry seeds, the physical removal of the seed coat by dehulling or milling markedly reduces the tannin content with a resultant improvement in the nutritional quality of the protein. Soaking in water or salt solution prior to household cooking also causes a significant reduction in tannin, provided the cooking broth is discarded. Treatment with a variety of chemicals such as ammonia, hydrogen peroxide, formaldehyde, polyethylene glycol, or polyvinylpyrrolidone has also proved to be an effective means for reducing the

tannin content of plant sources of protein. Germination leads to more than a 50% loss in tannins in legumes such as the chick pea, green gram, mung bean, and black gram, presumably due to the action of polyphenol oxidase of endogenous origin. (*See* **Tannins and Polyphenols**.)

Flatulence-producing Factors

One of the important factors limiting the use of legumes in the human diet is the production of intestinal gas (flatulence), which is generally attributed to the inability of the gut enzymes to hydrolyze the α-1,6-galactosidic linkages of such oligosaccharides as raffinose, stachyose and verbascose ((**6**), (**7**) and (**8**) respectively Figure 6). In their unhydrolyzed forms, these oligosaccharides become metabolized by the microflora of the large intestine into carbon dioxide, hydrogen, and methane, gases that are responsible for the characteristic features of flatulence, namely nausea, cramps, diarrhea, and the social discomfort associated with the release of rectal gases. Advantage may be taken of the endogenous galactosidases present in the plant tissue by allowing a slurry of the raw ground beans to undergo autolysis under suitable conditions or by treatment with an exogenous source of these enzymes derived from molds or bacteria. Such traditional dishes as tofu and tempeh, which have undergone fermentation by microorganisms, likewise have little flatus activity. Protein isolates from which these oligosaccharides have been eliminated during the course of their isolation are essentially devoid of the offending oligosaccharides. Preparations of galactosidases from microbial sources that have been immobilized as continuous flow reactors or in the form of a hollow-fiber dialyzer have also been employed for removing flatulence-producing factors.

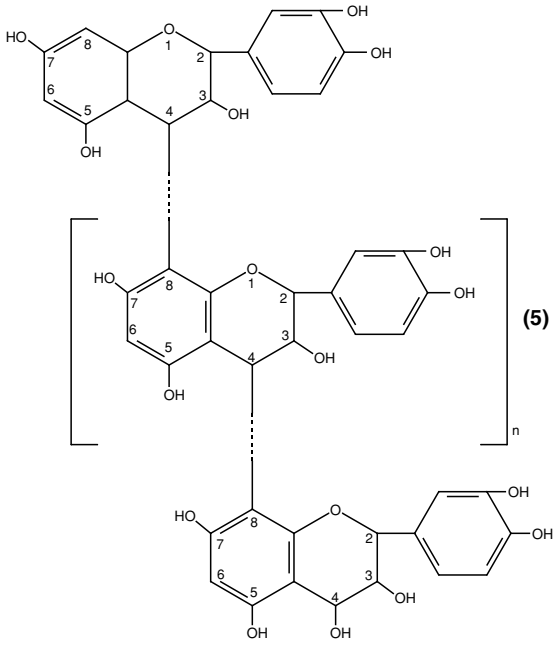

Figure 5 Structure of condensed tannin (5), which, by forming a complex with dietary protein, interferes with the digestion of protein.

Saponins

Saponins comprise a large family of structurally related compounds containing a steroid or triterpenoid aglycone (sapogenin) linked to one or more oligosaccharide moieties. Despite the fact that saponins are widely distributed in the plant kingdom, only a small number of such plants are actually toxic to mammals. It has long been recognized, however, that saponins, such as the glycoside of medicagenic acid ((**9**) Figure 7), which is found in alfalfa, is considered to have an adverse effect on the productive performance of nonruminant animals such as swine and poultry.

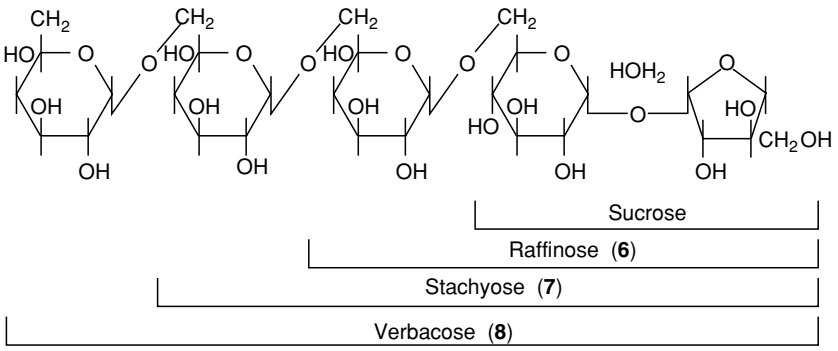

Figure 6 Structure of the oligosaccharides (6, 7, and 8) responsible for flatulence. In the absence of enzymes in the small intestine of humans that are capable of hydrolyzing α-1,6-galactosidic likages, these compounds are metabolized by the microflora of the gut into gaseous end products.

The negative effect of the saponins can be reversed, however, by the inclusion of dietary cholesterol, which interferes with the absorption of saponins by forming an insoluble complex with saponins. By the same token, dietary saponins may exert a positive effect by reducing cholesterol levels in the tissue and serum of experimental animals. (*See* **Saponins.**)

Mimosine

Leucaena leucephela is a tropical legume used primarily as a forage crop for feeding livestock, but its use is limited by the fact that it contains an unusual amino acid, mimosine ((**10**) Figure 8). This compound has an adverse effect on the growth of ruminants because bacteria can convert mimosine to 3,4-dihydroxypyridine ((**11**) Figure 8) which acts as a goitrogenic agent. In Hawaii, ruminants can convert greater amounts of the *Leucaena* before it becomes toxic than ruminants in Australia because of the presence of bacteria in the rumen capable of detoxifying mimosine. In fact, when Australian ruminants were inoculated with these organisms, they became more resistant to the toxic effects of mimosine.

Lathyrism

Human consumption of the legume chickling vetch (*Lathyrus sativus*) causes a neurological condition known as 'lathyrism,' a disease that is common in India and Bangladesh. The causative factor of this disease has been identified as β-*N*-oxalyl-α,β-diaminopropionic acid ((**12**) Figure 9), but over 90% of this toxin can be eliminated by the simple expedient of soaking the seeds overnight in an excess of water followed by steaming, roasting, or sun drying.

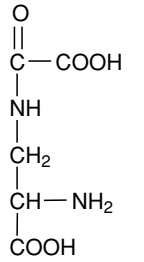

(**12**)

Figure 9 The structure of β-*N*-oxalyl-α,β-diaminopropionic (12), the causative factor of lathyrism in humans.

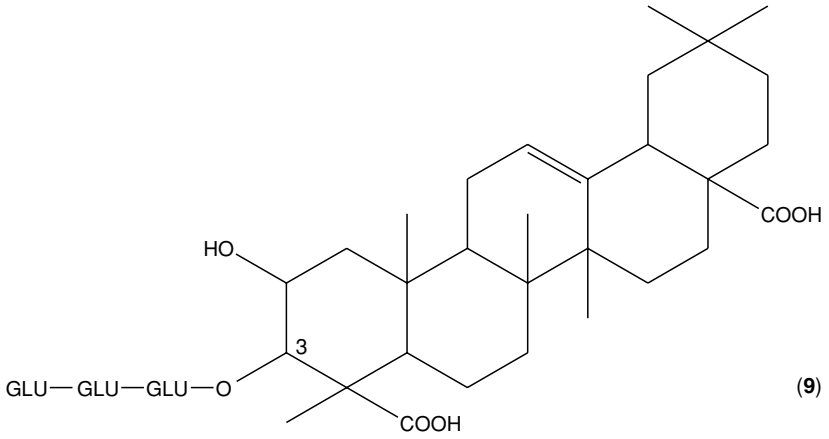

(**9**)

Figure 7 Structure of medicagenic acid (9), a saponin that is present in alfalfa and is responsible for toxic effects in livestock.

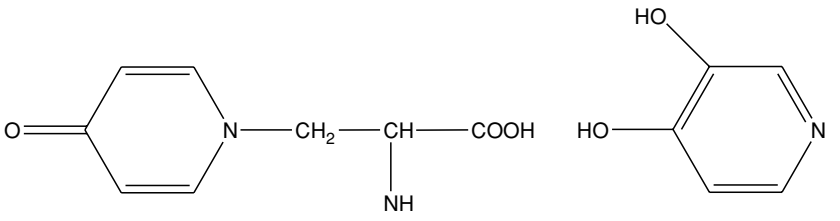

Mimosine
(**10**)

3,4-Dihydroxypyridine
(**11**)

Figure 8 Structure of minosine (10) and its goitrogenic metabolite, 3,4-dihydroxypyridine (11).

See also: **Glucosinolates**; **Phospholipids**: Properties and Occurrence; **Phytic Acid**: Properties and Determination; Nutritional Impact; **Saponins**; **Soy (Soya) Beans**: Properties and Analysis; Dietary Importance; **Tannins and Polyphenols**; **Plant Antinutritional Factors**: Characteristics

Further Reading

Cheeke PR (ed.) (1989) *Toxicants of Plant Origin*, vols. I–IV. Boca Raton, FL: CRC Press.

Douglas MW, Parsons CM and Hymowitz T (1999) Nutritional evaluation of lectin-free soybeans for poultry. *Poultry Science* 78: 91–95.

Jones RJ and Megarrity RG (1986) Successful transfer of DHP degrading bacteria from Hawaiian goats to Australian ruminants to overcome the toxicity of *Leucaena*. *Australian Veterinary Research Journal* 63: 259–261.

Legras JL, Jory M, Arnaud A and Galzy P (1990) Detoxification of cassava pulp using *Brevibacterium* sp. R312. *Applied Microbiology and Biotechnology* 33: 529–533.

Liener IE (ed.) (1980) *Toxic Constituents of Plant Foodstuffs*, 2nd edn. New York: Academic Press.

Liener IE (1987) Detoxifying enzymes. In: King RD, Cheetham PSJ (eds) *Food Biotechnology*, pp. 249–271, vol. 1. London: Elsevier.

Liener IE (1989) Control of antinutritional and toxic factors in oilseeds and legumes. In: Lucas EW, Erikson DR and Nip W-K (eds) *Food Uses of Whole Oil and Protein Seeds*, pp. 249–271. Champaign, IL: American Oil Chemists Society.

Liener IE (1997) Plant lectins: properties, nutritional significance, and function. In: Shahidi F (ed.) *Antinutrients and Phytochemicals in Food*, ACS Symposium Series 662, pp. 31–43. Washington, DC, American Chemical Society.

Majak W (1992) Mammalian metabolism of toxic glycosides from plants. *Journal of Toxicology* 11: 1–40.

Marfo EK, Simpson BK, Idou JS and Oke OL (1990) Effect of local food processing on phytate levels in cassava, cocoyam, maize, sorghum, rice, cowpea and soybean. *Journal of Agricultural and Food Chemistry* 38: 1580–1585.

Moneam NMA (1990) Effect of presoaking on faba bean enzyme inhibitors and polyphenols after cooking. *Journal of Agricultural and Food Chemistry* 36: 1479–1482.

Sessa DJ and Ghoutous PE (1987) Chemical inactivation of soybean trypsin inhibitors. *Journal of American Oil Chemists Society* 64: 1682–1687.

Watson D (ed.) (1998) *Natural Toxicants in Food*. Boca Raton, FL: CRC Press.

PLANT DESIGN

Contents

Basic Principles

H L M Lelieveld and G J Curiel, Unilever Research and Development Vlaardingen, Vlaardingen, The Netherlands

Background

If a plant is not of a proper hygienic design, it will be difficult to clean, and if not clean, the inactivation of microorganisms can be extremely difficult. After inadequate cleaning, harmful microorganisms may grow and affect product quality and safety. This chapter summarizes the basic principles of hygienic plant design to control the microbiological quality and safety of the final product.

Raw materials usually contain microorganisms. When given time and the right environment, they can grow out and increase to very high numbers. In food-processing equipment, the temperature and availability of nutrients are often ideal for the growth of many types of microorganisms. Consequently, when given enough time, such as in areas where the product is stagnant or residing for a long time for any other reason, microorganisms will cause problems. Product that is free from (viable) microorganisms may become recontaminated by microorganisms in the environment (air, people, insects, and other pests), by the processing and packaging equipment and by the packaging materials. Equipment must be hygienically

designed and fabricated, installed in a sufficiently hygienic environment, and be operated by staff who consistently observe the critical hygiene rules that apply.

Requirements for Hygienic Food Plants

Location and Buildings Exterior

Preferably, food plants should not be located near areas where levels of microorganisms, insects, or rodents can be exceptionally high. Therefore, a food plant should not be near a sewage treatment plant or in the middle of an animal farm area. Also, although decorative, there should be no trees, bushes, and plants near the factory. Such flora attracts insects and birds, which are sources of all kinds of pathogenic microorganisms. Also, there should be no breeding places for insects, so there should be no decorative ponds and generally no stagnant water at all in the vicinity. To avoid attracting insects, external lighting should never be mounted to, or near to, the wall of the plant. If positioned at a suitable distance, illumination will be adequate while insects move away from the plant. The outer walls of factories should not provide landing or breeding places for birds or other animals. The lower part of the walls must be rodent-resistant. Also, walls should be deep enough to prevent rodents from digging tunnels underneath to gain access to the plant. Measures must be taken to prevent animals gaining access to any area between the ceiling and roof. Roofs should not retain water (breeding of insects and bacteria) and thus slope such that draining is adequate. Doors must be self-closing; windows and doors that are opened infrequently, including emergency exits, should be such that there is little space between door and frame.

Layout

As the requirements for raw materials differ from those applying to intermediate or final products, the design of food plants should start with defining the hygiene requirements for the product in the subsequent stages of the production process. Based on that inventory, the layout of the plant should be decided, taking into account the various flows: of product, packing material, people, and air. Air, unless freed from microorganisms, should always move away from the exposed final product. Primary packing materials should be clean (complying with the microbiological requirements for the product to be packed). Neither people nor any material should move from the raw materials area to the finished exposed product area unless measures have been taken to prevent the carryover of microbial contamination. To protect microbiologically sensitive products, high-care areas may be needed. The entrance to such an area should have facilities for (compulsory) change of footwear and coats. All areas should have well-designed and well-maintained handwash facilities.

It is essential too, that, if the layout is correct, staff do not nullify the effect of correct measures by not knowing or not following the rules. For example, one is easily tempted to open doors and windows on a hot summer day, thereby completely destroying the airflows needed to prevent product contamination. Also, if possible and unobserved, people are likely to take 'shortcuts' to move from one place to another and thereby entering high-care areas from raw material areas. It is part of correct plant design to ensure that the rules are correct and sensible and that everybody is aware of their importance – and perhaps of the sanctions in case of breach: the health of consumers and the reputation of the factory are at stake.

Construction

Floors Floors are one of the great challenges: on the one hand, for obvious reasons, they should not be slippery. On the other hand, because they must be cleaned, floors must be smooth. In the case of dry floors, the problem is relatively easy to control, but any spilled liquid (water or product) may present an occupational hazard. The hazard selfevidently is related to both floor and footwear, and their interaction must be taken into account. Clearly, a 'dry-floor policy' will help to reduce both hygiene and occupational problems. Smooth concrete floors are acceptable for reception and many raw material storage areas. In areas where product is not protected, ceramic tiles or seamless mortar resin floors are preferred. In high-care areas, often, seamless resin floors are the best choice. Floors must have rounded corners where the floor meets the wall and should slope approximately 2% towards drains or gutters.

Walls Walls should be smooth and cleanable and should be covered by ceramic tiles or a good-quality coating or paint. If sandwich panels are used, all sides should comply with the chemical resistance required. It is important that all seams are reliably sealed. In some cases, seams can be sealed by welding with the parent material. In other cases, and where walls meet the floor, sealants should be used. Also, anything mounted to the wall should be sealed all around to prevent access to insects and dirt. The surface of choice should be strong enough to meet factory conditions, and repairs to any damage should be carried out within a short time. Where walls separate various hygiene areas, care must be taken that any holes for service lines are properly sealed.

Ceilings Panels similar to those used for walls (see above) may be used, and the same requirements apply. Lighting should be preferably built into the ceiling, with the underside flush with the ceiling. All suspended items, including lights at lower levels and exposed beams must be such that the upper side is sloped sufficiently to avoid the accumulation of dust and dirt. Ceilings and anything suspended from them must be easy to clean. Where acoustic ceilings are needed, it should be noted that sound-absorbing panels often cannot be cleaned and that panels that are easily cleanable usually are excellent sound reflectors. Consequently, a compromise will have to be found. Whenever possible, try to avoid noisy equipment in the plant, and when unavoidable, try to contain the noise near its location, where it may be easier to inspect for cleanliness.

Service lines and air ducts Service lines (hot and cold water, drains, electricity, compressed air) may harbor insects and other pests. Preferably, they should be combined in suitable, thus externally, smooth ducts, with all ends effectively sealed against vermin. Where mounted to a flat surface, sealant may be needed to close recesses.

Air ducts as well as cold-water pipes and drains, owing to their lower temperature, often may collect moisture, resulting in drops falling down on to product and food contact surfaces, and these may also support microbial growth. Hence, in areas where this will happen, cold surfaces must be insulated. The insulation should be sealed, sealing requirements being the same as for the ducts mentioned above. If the conditioned air is too humid, water may collect within the air ducts, allowing microbial growth, and so, the air becomes a significant source of contamination. This should be taken into account when designing the air system in the plant.

Equipment

Cleaning Cleanability is very important. Equipment that is not cleanable cannot be decontaminated other than with great difficulty or not at all.

Inactivation of microorganisms It must be possible to treat all product contact surfaces with whatever means has been chosen to inactivate the relevant microorganisms. Where the equipment is intended for pasteurized products, the equipment should preferably be 'pasteurizable.' In other words, it must be ascertained that all surfaces come in contact with water of the correct temperature for the required time. Entrapped air, e.g., in the top of tanks, will hamper pasteurization. Equipment for sterilized products must be made sterile. Steam sterilization is preferable, but again, there are pitfalls: in this case, lower parts of equipment may collect condense and thereby not be treated with steam. The considerably lower degree of heat transfer of water may reduce the temperature too much at such a spot. Where equipment is not sufficiently heat-resistant, chemical means may be used. Also, here, air entrapment must be prevented to ensure treatment of all surfaces.

Migration of microorganisms – aseptic processing and packaging A hygienic plant should protect the product from contamination. The ingress of microorganisms should be limited and, in the case of aseptic processing, fully prevented. For so-called closed plants, microorganisms may migrate to the product through crevices, leaking connections and moving parts of machinery, such as valves and pumps. Prevention is achieved by correct mounting and sealing of parts and by the application of barriers where movable parts are unavoidable. When choosing equipment, it should be realized that reciprocating shafts facilitate the ingress of microorganisms: the shaft transports small amounts of product to the outside, providing nutrients to any microbes present there, and every other stroke, microorganisms are transported to the inside. Rotary shafts display a similar effect, but to a greatly reduced extent. Where ingress must be prevented completely, such as for aseptic processing and packaging, diaphragms should be used to separate the product from the environment. Where this is not possible, such as with rotating shafts, double seals should be used, with an antimicrobial liquid circulating between them.

Growth Where, for functional reasons, such as blending (nonsterile), food products reside in equipment for a long time, measures must be taken to prevent excessive growth. This can be done by keeping the product below or above temperatures suitable for growth of the relevant microorganisms. Where this is done, it must be taken into account that there will be warm spots, e.g., where a shaft enters the equipment (centrifugal pump, rotary stirrer). Where the residence time is too long, because of the presence of avoidable dead areas in equipment, the equipment must be considered unhygienic. Such equipment should not be used for food processing, unless it can be cleaned easily and is used for production runs that are too short to cause microbial problems.

Principle hygienic design requirements for equipment Although hygiene is essential, the measures taken to ensure hygiene should not generate other health problems. Therefore, the principle requirements for the design of hygienic equipment also must take other aspects into account.

The basic principles are as follows:

- To cope with the aforementioned issues, food-processing equipment should meet the following requirements to be acceptable: all product contact surfaces must be noncorrosive, i.e., resistant to the product and chemicals under the operating and cleaning conditions. Furthermore, materials used (basic construction materials, but also auxiliary materials such as coatings, lubricants, sealants, adhesives, and fillers) should not have the potential of contaminating the food product such that the food becomes toxic or allergic.
- All product contact surfaces must be free from crevices and other irregularities that would be difficult to clean. Thus, no metal-to-metal connections (other than by welding or soldering) are allowed, and screw threads must be avoided or otherwise sealed.
- Product contact surfaces must be smooth enough for easy cleaning. Both the Hygienic Engineering and Design Group and 3-A prescribe a surface roughness for product contact surfaces of $R_a = 0.8\,\mu m$ or less.
- All product contact surfaces must be drainable to make it possible to remove all liquids from process equipment. Thus, surfaces and pipes should not be completely horizontal but slope towards drain points, and there should be no ridges, which may hamper draining. Where it is not possible to build equipment in such a way that proper draining is possible, procedures must be designed to ensure that residues of cleaning and disinfection liquids can be removed in another way. Even if no chemicals are used for cleaning, draining is important because many microorganisms can grow easily in residual water, needing only tiny amounts of nutrients.
- Dead legs and product stagnant areas must be avoided or, where unavoidable, should be designed to minimize the residence time. The presence of such areas should be taken into account when deciding on the frequency and procedure of cleaning. Dead legs of tees, if not too long, may be reasonably easy to clean if the direction of the flow of cleaning liquid is towards the dead leg. Special attention should be given to provisions for temporary connections, such as those used for filling transportable containers, road tankers, etc. If not used for some time and not cleaned, they present a large 'dead' area, very much alive with microorganisms.

Fabrication and installation The fabrication must be such that product contact surfaces are not damaged to adversely affect the cleanability. Welds should not be rough and, where necessary, should be polished to obtain a smooth finish. Welding should be done skillfully, as welds otherwise may soon corrode. This is particularly important for welds that cannot be seen anymore after assembly, such as with pipelines. Pipe-bending can be applied to preclude the need for too many welds and too many piping components. In general, the risk of contamination of food product is proportional to the number of microorganisms in the surrounding, and so the nonproduct side of equipment also must be accessible for cleaning. External features should not collect water, product, or dirt in areas that cannot be inspected and cleaned and thus provide places for harboring pests. That means that there must be either no space at all or an insufficiently large space between parts of machinery. The same applies to the space between machinery and the supporting structure, i.e., the floor, wall, ceiling, or any other support.

The installation of components should be such that the assembly is also hygienic. Hygienic components may be built together in such a way that the completed installation is not hygienic because the assembly cannot be drained or cleaned, and creates dead areas. Connections may result in crevices, components may cause obstructions for draining, and in a closed process line, a hygienic diaphragm pressure sensor on an easily cleanable tee may create a large dead area. The food-processing manager should not only decide on hygienic equipment but also keep a close eye on how the contractors intend to build the entire process line. In addition, care should be taken that overhead lines and components are not a source of contamination. Vertical piping is preferred, and horizontal pipework should slope to drains.

Sources of Further Information

There are a number of professional organizations that provide detailed information on the hygienic aspects of plant design. The EHEDG is an organization with about 25 subgroups specializing in the many different aspects of hygienic design and processing. The organization is supported financially by the European Commission to support the development of guidelines and to ensure their availability in many European languages. Currently, over 20 guidelines are available, covering subjects such as hygienic design criteria, the design of hygienic equipment for open and closed plants, designing hygienic pumps and valves, microbiologically safe pasteurization and sterilization of liquid products, hygienic as well as aseptic packaging of food products, and welding to meet hygienic requirements. In addition, the EHEDG has developed methods to determine whether equipment

meets the hygienic design criteria. Equipment complying with the EHEDG criteria may be certified. Much of EHEDG's guidelines can be found in the standards for machinery developed by the European Standardization Organization (CEN, Commission Européenne de Normalisation). Details can be found at www.ehedg.org. In the USA, the 3-A organization develops standards for food-processing equipment. Currently, 3-A provides a self-certification scheme for equipment complying with the 3-A standards, which may result in authorization to use the 3-A symbol on equipment. Information on 3-A can be found at www.3-A.org. Also, the USA-based NSF (National Sanitation Foundation) International produces standards for hygienic equipment, but traditionally focused on the catering area. NSF's website can be found at www.nsf.org. The three organizations have recognized the importance of harmonization between the various regions in the world and increasingly work together. NSF International has an office in Brussels, and one of EHEDG's regional sections is in Japan. EHEDG guidelines are produced with the assistance of 3-A and NSF, and 3-A standards are produced with assistance from EHEDG. Recently, NSF and 3-A have started to produce joint standards. The organizations also all work closely together with national, regional, and global authorities.

Concluding Remarks

It is important to ensure that the plant or process line will deliver safe products. Validation of the cleanability and the disinfectability and, where relevant, imperviousness to microorganisms are essential in the selection and commissioning steps. Not only must the food be safe just after commissioning the plant, but the plant or installations must continue to comply with the hygienic requirements. Therefore, an appropriate scheme for preventive maintenance should be devised.

See also: **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant; Overall Approach; Modern Systems; **Plant Design**: Designing for Hygienic Operation; Process Control and Automation

Further Reading

3-A Sanitary Standards Committees (1999) *Model Document for Preparing 3-A Sanitary Standards and 3-A Accepted Practices*, 2nd edn. McLean, VA: Dairy & Food Industries Supply Association.

Curiel GJ, Hauser G, Peschel P *et al.* (1993) *Hygienic Equipment Design Criteria, EHEDG Document No. 8*. Chipping Campden, UK: Campden & Chorleywood Food Research Association.

Imholte TJ (ed.) (1984) *Engineering for Food Safety and Sanitation*. Crystal, MN: Technical Institute of Food Safety.

Lelieveld HLM, Mostert MA, White B and Holah JT (eds) (2002) *Hygiene in Food Processing* Cambridge, UK: Woodhead.

Shapton DA and Shapton NF (eds) (1991) *Principles and Practices for the Safe Processing of Foods*, Cambridge, UK: Woodhead.

Designing for Hygienic Operation

J P Clark, A. Epstein and Sons International Inc., Chicago, IL, USA

Background

As with facility design, discussed in the previous article, process and equipment design has many elements specific to the food being processed and others that represent common principles. This article discusses a number of those common principles, types of equipment, and common processes.

Overall Considerations

Equipment used in food-processing plants must be appropriate, effective, noncontaminating, easy to clean, and easy to inspect. It must also be safe for the workers using it. In the USA, equipment used in meat, poultry, and fish processing must be specifically approved in advance by the US Department of Agriculture (USDA). Equipment used in dairy processing should be listed in 3A standards (a voluntary industry standard-setting group), and equipment used in baking should satisfy Baking Industry Sanitation Standards Committee requirements. There are other standard-setting groups in other countries and for other industries.

In general, such standards address the aforementioned general requirements in light of the specific conditions. For example, the USDA emphasizes cleaning and inspection, because most meat processing equipment is cleaned daily. Dairy equipment is often cleaned in place, and so surface finish and absence of crevices are important. Baking equipment is less often cleaned with water, but is vulnerable to insect infestation. (*See* **Cleaning Procedures in the Factory**: Overall Approach; **Insect Pests**: Insects and Related Pests.)

As a result of nearly universal sanitation requirements, much food equipment is constructed of stainless steel, usually type 304, but often 316 and occasionally some other alloy. The alloys differ in their weldability and resistance to stress corrosion cracking, which is accelerated by chloride ions found in many foods and cleaning materials. Stainless steels are nonmagnetic, and so fragments that may occur as a result of wear or misuse are not removed by magnets. Metal detectors, which function by measuring changes in inductance, can detect stainless steels and are often used on final packages of foods. (*See* **Corrosion Chemistry**.)

Many polymers are acceptable for food use, especially Teflon, Neoprene, polyethylene, and polypropylene. Carbon steel is acceptable where conditions will not promote corrosion and contamination. Copper is used occasionally, for special purposes such as candy cooking.

Equipment designed for food processing can usually be disassembled with few tools and uses wide screw-threads or flanged connections to avoid creation of places where residues can be caught. Wherever possible, pipelines are welded, taking care to make inside surfaces of welds smooth and flush. Long runs of pipe or tubing must have inspection access and must be installed carefully so that drainage is complete. Dead spots are avoided in pipes so that systems can be cleaned thoroughly.

Clean-in-place

The concept of cleaning in place by flowing cleaning solutions and rinse water through a system in place of food liquids has significantly increased the productivity of food plants. Previous to its wide use, dairies and other food plants were disassembled and washed completely by hand. This consumed so much time and labor that it limited the practical size of plants.

The elements of a typical clean-in-place (CIP) system are chemical storage, solution supply and recovery tanks, supply and return pumps (usually centrifugal), spray devices in tanks, air-controlled automatic valves, manual connection stations (flowverters), additional piping to complete circuits, and automatic control devices such as programmable logic controllers. Empirical design rules have evolved from experience for flow velocities, spray intensities and solution strengths for effective cleaning of typical soils from food plants.

Most experience is with dairy and other liquid-product plants, such as soft drink, fruit juice, and syrup plants. Minimum velocities of $1.5\,m\,s^{-1}$ are generally required in pipelines to achieve sufficient turbulence. Tees and dead legs should not exceed three pipe diameters' length from the flowing stream so as to ensure cleaning. Tanks are sprayed at about 37 liters per minute per meter of circumference. Often, such flow rates surpass the normal process flow rates, and so special pumps may be needed just for CIP. Sometimes, it may be possible to use dual-speed process pumps, with a low speed for process flow and a higher speed for CIP. Such a technique saves investment cost and space.

It is common to design CIP systems so that cleaning solutions and final rinse water are recovered for reuse, to save costs and reduce discharges. Caustic soda solutions (1.5%) and nitric, phosphoric, and citric acids are common cleaning agents. Sometimes, detergents are used, alone or with acids or bases. In addition, sanitizing agents, such as iodophores, quaternary ammonium compounds, and chlorine-releasing agents, may be used, usually as a final rinse that is allowed to sit overnight. In the morning, the sanitizing agent is displaced with fresh water and then with product. (*See* **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant; Modern Systems.)

With a properly designed CIP system, most of the plant waste water should be discharged during cleaning from the CIP unit. This permits monitoring and pretreatment, if necessary.

Some manual labor is involved in cleaning even highly automated systems. Positive-displacement pumps must be disassembled, bypassed, and cleaned by hand because their construction prevents complete cleaning by flushing. It is common practice to require manual connections at flowverters to prevent accidental contamination of food with cleaning solutions. Connections are commonly verified by electronic signals to the controller.

It is difficult to separate CIP design from process system design in many cases; they are best developed together. Physical arrangements, details of nozzles on tanks, valve placement, pump utilization, pipe runs, and other details are significantly affected by CIP.

Food Plant Layout

Overall process flow is dictated in part by location of shipping and receiving and also by constraints on equipment. For example, cookie and cracker baking ovens are as long as 100 m and must run in a straight line, because they use metal bands as baking surfaces, and partially baked pieces cannot be transferred easily. Hydrostatic sterilizers for cans and jars are usually located on outside walls for ease of access and maintenance. Cooling tunnels for chocolate enrobing are usually long and straight.

Dry equipment and processes should not be physically adjacent to wet ones; cold process areas should

be separated from hot ones; 'dirty' areas should be separated from clean ones, and so forth. Such separations are not merely physical, in the sense of walls and doors, but also include separate air-handling units, and perhaps limited access by people from other areas.

Vehicle traffic should be separate from people traffic. Often, it is desirable to keep visitors, who may be customers, off the production floor, in part for their own safety, but also to reduce exposure of food to humans. At the same time, if visitors are likely, it is desirable to offer a decent and controlled view of the process. One solution is an elevated, enclosed walkway with viewing windows. If the process side of the windows is refrigerated, as many food plants are, the windows need to be heated to prevent condensation.

Plant layout can encourage or inhibit communication among workers; with fewer people running modern food processes, good communication is essential. A serpentine layout may be helpful, compared to a long, straight arrangement. Clever use of closed-circuit television can overcome distance and obstructions.

Typical Unit Operations

Some unit operations are so common and important to food processing that they deserve individual discussion.

Mixing

Processes of mixing solids with solids, solids with liquids, and liquids with liquids are all quite common in food processing. For hygienic operation, mixers must be designed to empty completely, to be easily cleaned, and to be easily inspected. Access for addition of minor ingredients and for sampling may also be necessary. Doughs and pastes, such as bread dough, cookie batter, and confection fondants are prepared in specially designed machines with relatively high-powered motors.

Incorporation of gas, especially of air, may be desirable or undesirable for a given case; the mixer design changes accordingly. (High speed, giving high shear, leads to gas incorporation.)

Dispersion of powders into liquids, such as starch or gums into water or syrup, requires high-speed agitation to avoid the formation of partially wet agglomerates, called 'fish eyes.' One technique is to pour solids into a deliberately inefficient centrifugal pump, which is circulating liquid from and to a tank. The inefficiency means that energy is applied to dispersing the solids rather than pumping the liquid.

Dissimilar liquids, such as oil and water, do not form solutions but can be made to form stable suspensions or emulsions by formation of very small droplets of oil and by using surfactants. The small droplets are formed by high-shear agitation and by pumping under high pressure through small orifices or special valves in a homogenizer. Many foods contain natural surfactants, such as proteins and polysaccharides, which reduce surface tension at the interfaces between liquids and so help maintain an emulsion. Milk, icecream and salad dressings are examples of foods containing oils and aqueous solutions in stable suspension. (*See* **Colloids and Emulsions**.)

Heat Transfer

Heating and cooling of foods is critical to many preservation and processing techniques, including pasteurization, freezing, baking, and other types of cooking. One of the more common heat exchangers used in food processing is the plate type. This has dimpled or otherwise embossed plates separated by gaskets and held together by mechanical pressure on a sturdy frame. Hot and cool fluids pass on opposite sides of the plates, providing a large heat transfer surface in a rather compact space. The major advantage of the plate exchanger for food service is the ease with which it can be disassembled for inspection and maintenance. It can also be reconfigured easily by adding or removing plates on the same frame (up to the limit of the frame). (*See* **Heat Transfer Methods**.)

Shell and tube, concentric tube, and spiral tube heat exchangers are also found in food plants. Direct steam injection is also common.

Cooling and freezing are performed in several ways. Wiped-surface heat exchangers are common for freezing flowable fluids, such as icecreams, and for cooling fluids that have been heated for sterilization. Refrigerated plates are used for contact freezing of packages of food, such as vegetables or preplated dinners. Direct contact with very cold vapors or snow from liquid nitrogen or carbon dioxide is used in cryogenic freezing, usually on continuous belts passing through tunnels. Air cooled by refrigeration can also be used in similar tunnels. Cryogenic equipment is less expensive than mechanical refrigeration equipment but is more expensive to operate. The exact balance depends on local costs, but cryogenics are a good choice when minimizing capital is key; mechanical freezing is more common for established, ongoing processes, where operating costs are more critical. (*See* **Freezing**: Principles; Operations; Blast and Plate Freezing; Cryogenic Freezing.)

Steam heating of food in sealed containers, such as cans, jars, and pouches, is commonly performed in pressure vessels known as retorts. Batch retorts may be horizontal or vertical, and may have water in addition to steam or steam alone for heating. For

containers such as jars and pouches, which cannot tolerate internal pressure, air is added during the cooling phase, which uses water, to counter the internal pressure built up during cooking.

There are several approaches to continuous cooking of foods in containers. One of the more efficient is the hydrostatic retort, in which containers are transported in carriers on a chain through a 'U'-shaped tower in which the vertical leg of water serves as a seal to contain a high-pressure chamber. (In both batch and continuous retorts, pressures to reach 120 °C are used so as to shorten the time required to sterilize food without giving it overcooked taste and texture.) (*See* **Sterilization of Foods**.)

Another approach uses a helical track inside a horizontal cylindrical shell. The track rotates, transporting metal cans. Cans enter and leave through a star wheel valve. A similar shell and track is used to cool the cans under pressure.

Conveying

Many foods are solids or are made from solids, and so solids material handling is critical to food processing. Pneumatic conveying, by pressurized air or vacuum, is commonly used to move flour, sugar, and grains in food plants. Pressure conveying permits movement from one source to several delivery points, such as flour to several dough mixers. Vacuum conveying works well for several sources to one delivery point, as in dust collecting. It is also often used to unload bulk material delivery vehicles, such as rail cars or trucks. Dense-phase pneumatic conveying is a variation used for fragile materials, such as sugar, which might be damaged by the high velocities encountered in pressure or vacuum conveying. Typically, for dense-phase conveying, powders are blown from a holding vessel in slugs rather than being entrained in conveying gas.

Screw conveyors are used to move pasty or granular materials, often up inclines. They can be difficult to clean and are very dangerous, but for some purposes, they are uniquely appropriate.

Belt conveyors can be used to move loose food materials, but are best for packages and containers. They can also pose safety hazards, with numerous pinch points and motors, which start and stop under remote control.

Vibratory conveyors are well suited to many food applications, especially for fragile materials such as fried snacks, because they can be quite gentle and are easily cleaned.

No matter what solids conveying system is used, it can rarely be properly designed without specific measurement of the physical properties of the actual materials involved. Solids used in food plants vary widely in the properties that influence conveying and handling, and these properties are difficult to predict from theory or correlations. Examples of key properties include bulk density, particle size distribution, angle of repose, cohesiveness, abrasiveness and moisture content.

Pumping

Many foods are quite viscous and non-Newtonian, so proper design of pumps is challenging. Positive-displacement pumps, especially those using close-fitting lobes and screw-shaped rotors, are common. Such pumps are usually made of stainless steel and are designed for easy disassembly. Centrifugal pumps are more conventional but also likely to be stainless steel and built to be taken apart by hand with few tools.

A special case of pumping, which combines the functions of mixing and heating, is screw extrusion. Extruders may be single or double screw, with heating or cooling in jackets, and with a wide variety of die designs. When starchy materials, such as corn, are extruded under pressure through a die, an expanded foam is formed, which has a desirable texture and density. Pet foods, cereals, and snacks are made in large quantities in this way. (*See* **Extrusion Cooking: Principles and Practice**.)

Typical Food Processes

Many food processes have a number of steps in common, including mixing of ingredients, forming of pieces or shapes, filling of containers or packages, cooking and cooling, packaging for shipment, and unitizing. The specific details obviously depend on the product and circumstances. However, some issues are almost universal.

Formulation

In order to be mixed, a formula must be weighed or measured; this may involved feeding and scaling of particulate solids, metering of liquids and addition of small quantities of key ingredients. Most formulas for foods contain three categories of ingredients: major (percentages in the tens), minor (single-digit percentages), and micro (less than 1%). (Even these distinctions may vary from case to case.) It is common to prepare formulas in batches of several thousand kilograms, even if subsequent processing is continuous. Issues that arise include whether to use one scale, which must then be capable of weighing the entire batch, or multiple scales, sized for accuracy of each category. The cycle time of batching is longer if fewer scales are used, but the costs are lower. Liquid ingredients can be metered using any of several devices.

Sometimes, ingredients are simply added by units, such as entire bags or drums, but this is less precise than scaling, because usually, such units are slightly heavier than their label weight.

Critical issues in any automatic formulating system are reliable feeding and flow of highly variable solids; even such common commodities as flour and sugar can clump, cake, and bridge in hoppers and chutes, disrupting flow and upsetting operations.

Forming

Forming operations are highly specific, but can include such examples as flaking, extrusion, sheeting (rolling into thin layers), cutting, laminating, decorating, dicing, and many others. Usually, these operations are continuous and so need careful metering of a mix, which may have difficult flow properties (doughs or pastes, for instance). Forming operations usually work best when operated at a smooth and constant rate, so the surge capacity ahead of the operation and reliable removal of output are important.

Filling

Primary containers may be filled before or after such processing as cooking, baking, or sterilization. If the material is liquid or a suspension, it can be filled volumetrically; if it is free flowing solid, it may be filled volumetrically, or it may be scaled. Weighing of valuable particulate solids is often done with multicompartment digital scales, in which a computer scans compartments to select several whose summed weight most closely approaches the target. Such scales are very accurate and justify their considerable cost by reducing overweight packages.

It is common for food packages, such as cans of soup, to be filled in several successive steps, with the final 'mixing' occurring in the package. Chicken noodle soup, for instance, has noodles, chicken, and broth each added separately using specialized fillers.

Cooking

The meaning of cooking varies from food to food but almost always involves some heating, directly or indirectly, to cause desirable chemical reactions, remove excess moisture, develop color (usually browning), and reduce microbial populations. Heat may be applied by flames, steam, hot air, oil, microwaves, or hot water. After heating, cooling often occurs by exposure to ambient air but may be accelerated by refrigeration. Cooking almost always involves more than one reaction, some of which may be undesirable, so control is complex. In frying, for example, water is removed and largely replaced by fat, the volume increases, and browning occurs. (*See* **Browning**: Nonenzymatic; **Cooking**: Domestic Techniques.)

Packaging

In addition to a primary package, which may be filled before or after cooking, there are usually additional layers of package to provide additional protection during storage and shipping. These may include cartons and cases. Cartons are usually made of heavy paper and are part of the consumer package, so usually have multicolor printing of brand names, ingredients, use instructions, weight, and nutritional information. Shipping cases are usually removed in a store, are rarely seen by consumers, and so can be made of sturdy corrugated cardboard with minimum printing and graphics.

Plastics are increasingly important as components of food packaging, because they are lighter in weight than metal or glass, can be used in microwave ovens to reheat contents, and do not break as easily as glass. Plastic containers usually cost more than the metal or glass containers they replace but are perceived to add value to the consumer. Recycling of food containers is an increasingly important concern to consumers and complicates the selection of packaging material. Some of the most effective plastic containers are composed of several types of plastic, each providing particular properties, such as barriers to oxygen or strength. Such multicomponent materials are less easily recycled than single-component materials.

Unitizing

Shipping cases of food are usually assembled into larger units, often of about 450 kg on wooden pallets, for storage and shipment. These pallet loads may be wrapped in stretch plastic, wrapped with heat-shrinkable plastic, strapped, or stabilized by glueing boxes or bundles together. Alternatively, large cardboard sheets known as slip sheets may be used to hold a unit load. Special equipment is needed to pick up and move loads on slip sheets, in contrast to the common forklift truck for which wooden pallets are designed, but the slip sheets are less expensive than pallets and take less space in storage and shipping.

In automatic storage systems and racks, used to achieve higher volumetric density in warehouses, special captive or slave pallets may be used, which are not shipped with the product load. This may require an additional transfer operation.

Some short-shelf-life foods are unitized on special carts, baskets, and trays suited for storage on small delivery trucks for direct shipment to stores. Fluid milk, cultured dairy products, and baked goods are often handled this way.

Analysis of the best packaging and unitizing procedure and equipment is a specialized area of engineering and requires balancing investment costs, operating costs and delivery system costs to achieve the best overall arrangement.

Conclusion

Design of food plants for hygienic and effective operations requires consideration of the interaction of the building with the process, selection of the proper equipment, development of an efficient layout, and consideration of the roles of people, materials, and processes. The goals are safety, quality and cost efficiency. Many special pieces of equipment have been developed to aid in reaching these goals.

See also: **Browning**: Nonenzymatic; **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant; Overall Approach; Modern Systems; **Colloids and Emulsions**; **Cooking**: Domestic Techniques; **Corrosion Chemistry**; **Extrusion Cooking**: Principles and Practice; **Freezing**: Principles; Operations; Blast and Plate Freezing; Cryogenic Freezing; **Heat Transfer Methods**; **Insect Pests**: Insects and Related Pests

Further Reading

Matz SA (1976) *Snack Food Technology*. Westport, CT: AVI.

Matz SA (1988) *Equipment for Bakers*. McAllen, TX: Pantech International.

Mercier C, Linko P and Harper JM (eds) (1989) *Extrusion Cooking*. St. Paul, MN: American Association of Cereal Chemists.

Sacharow S and Griffin RC (1980) *Principles of Food Packaging*. Westport, Ct: AVI.

Woodroof JG and Phillips GF (1981) *Beverages: Carbonated and Noncarbonated*. Westport, Ct: AVI.

Process Control and Automation

J P Clark, A. Epstein and Sons International Inc., Chicago, IL, USA

Background

Food processes, like other industrial operations, are measured and controlled in order to obtain consistent, safe, and efficient results. Wherever possible, response to measurements is automated, to reduce human labor and the risk of error. Modern processes rely heavily on computers to monitor operations, record data, and generate reports. Increasingly, computing power is in the form of microprocessors, which are dedicated components of instruments, relieving control computers of functions, simplifying installation, and increasing versatility. This article discusses techniques for measuring typical parameters, especially those found most often in food processes, and the ways in which measurements are used to control processes. The food industry has adopted measurement and control techniques from other industrial processes, including computer-integrated manufacturing (CIM), which is discussed briefly at the end of the article.

Typical Parameters

Weight

Weight, usually of solids, but also of mixes, packages, and storage vessels, is one of the most important and common measurements made in a food plant. Mechanical lever arrangements with counterweights and indicating dials are still widely used, but more common is the use of electronic strain gauges or load cells. These have a quick response and give a signal that is converted to a digital signal that can be read directly by a computer. Load cells need to be calibrated and their limitations understood; they can be sensitive to temperature, moisture, and vibration, and they must be selected for the total weight range they will experience.

The combination of an electronic scale with a variable-speed screw or vibrating cone creates a feeder that can be programmed to deliver a given rate or a given total amount of material. Such a feeder may operate by tracking the loss in weight from a hopper or by weighing the receiving bin (gain in weight). Loss in weight is most common. Periodically, the feed bin must be refilled; during this time, the feeder continues to run at the last speed it had. To minimize the time in which this 'volumetric' feeding occurs, refilling is kept to 5–10% of the cycle time. This means that the refill flow rate is 10–20 times the controlled feed rate; for high flow rates, this can be quite substantial and represents a significant shock to the receiving bin. Refilling can also aerate solids and make them hard to control or force them out of the feeder. Thus, bin sizing, delivery system, and flow control are all integral to successful weighing and delivery of solids.

Checkweighers are special scales placed in line with a package conveyor to weigh individual packages. Usually, a checkweigher controls a kick-out device

to remove over- or underweight packages. Data are also accumulated to calculate averages, standard deviations, and other quality statistics, which are used for statistical process control. This involves identifying process capabilities by measuring performance, then using such data to detect abnormalities and gradually to improve performance. Occasionally, a checkweigher is used to control a filler or other process affecting package weight.

When storage or process tanks and bins are to be weighed, they must be carefully isolated from other supports so that the load cells bear the true weight. This requires flexible connections from delivery pipes and other gear, which may be attached to a typical tank. Sometimes, load cells are installed to permit calibration of other instruments, such as meters, which are regularly used to control formulations.

Flow of Fluids

The traditional method of measuring flow rates of liquids and gases is by measuring pressure drop across an orifice (a small hole in a plate inserted in a pipe). For food fluids, an orifice plate may pose an unacceptable obstruction to suspended solids and may also be hard to clean in place. For homogeneous, clean fluids, this approach is reliable and inexpensive, but it provides a limited turn-down ratio of about 3:1.

Better for measuring homogeneous fluids (liquids or gases) is the vortex meter, which detects small changes in capacitance created as vortices are shed from a bluff body inserted in the fluid path. The vortices, or swirls of fluid, are created and shed in direct proportion to the flow rate. The turn-down ratio can be 10:1 or greater, and the measurement is independent of temperature, density, and viscosity. Installation is simple, and obstruction of the flow path is slight.

An increasingly common technique, which has no flow obstruction and does not require long runs of straight pipe before measurement, is the mass flow meter. A tube is inserted into a pipe as part of the flow path. The tube is vibrated at a resonant frequency. As fluid flows through the vibrating tube, it changes the phase difference of vibration at two points on the tube (or it imparts a twist to a bent tube), which can be detected and converted to a signal proportional to the flow rate. Magnetic or optical detectors measure the location of the tube, and electronic devices convert the signals to flow rates.

Other flow-rate-measuring devices count turns of rotors, vanes, or some other signal of flow rate. For electrically conducting fluids, changes in a magnetic field due to flow can be measured without any obstruction to the flow path. Positive-displacement meters use the flow directly to generate a signal by driving a lobe or piston connected to a shaft; rotations are proportional to flow. If a measurement is used to control flow, it usually signals an air or electrically operated valve, but it could signal a variable-speed drive on a pump.

Temperature

Temperature is a critical measurement in many food processes because it controls sterilization and cooking operations; too high or too low a temperature may be harmful. Thermocouples are inexpensive and reliable. They generate a small but consistent voltage difference at the junction of dissimilar metals, which can be correlated with temperature.

Most common for accurate measurement over larger ranges are resistance temperature detectors, which use the change in resistance of platinum with temperature to measure temperature. A common standard is to set the resistance at 100 Ω for 0 °C.

Bimetallic thermometers are commonly used as temperature indicators – the familiar dial thermometer – but they do not provide an electronic signal. Various solid-state devices, such as thermistors, are used to measure temperature because they give a signal directly compatible with electronic equipment, are compact, and do not require special wiring, as do thermocouples; however, they are not very accurate.

When temperature must be controlled, the signal usually drives a steam, gas, or water valve. Control of temperature in multizone equipment, such as baking ovens, can become very complex, as there are many temperatures to measure and many valves to control, most of which interact with each other. Safety interlocks are also required, to detect temperature excursions, to trigger alarms, to activate fire-suppression equipment, to control exhaust blowers (preventing buildup of gas), and to indicate restoration of operating conditions.

Mercury-in-glass and alcohol-in-glass thermometers are still found in canning operations, as calibration standards for retorts, but are rarely used in other food plants, because of the risk of contamination. (*See* **Canning**: Principles; **Mercury**: Toxicology.)

Temperature can be measured at a distance by infrared instruments. These are useful for checking insulation on overhead lines, roofs, and tall tanks.

A difficult temperature measurement challenge is to follow the temperature history of a container in a continuous sterilizer, such as a hydrostatic retort. The solution is a solid-state measuring and recording device, which fits within or on a container and is plugged to a computer when it is recovered. The data are recorded in semiconductor memory, which is read by computer or special instruments. Similar devices are used to track the temperature history of refrigerated and frozen foods in the distribution chain.

Traditionally, sterilization and pasteurization processes were required to keep ink tracings of temperature history continuously, usually on familiar circular instruments charts. Such records are still seen, but electronic data collection is increasingly accepted. It is critical to ensure the integrity of such data, if it is to displace the ink recorder. (*See* **Pasteurization**: Principles; **Sterilization of Foods**.)

Pressure

Solid-state strain gauges, similar to those at the heart of electronic scales, can also be used to measure pressure, since both parameters involve force. Sanitary pressure gauges involve flush-mounted diaphragms whose deflection generates a signal. The same device can serve as a level indicator in a tank.

Traditional Bourdon tube pressure indicators are found in food plants, especially on utilities such as steam. If used in direct contact with food, they need to be designed to prevent contamination by the food or of the food by any filling material.

Level

The level of fluids is measured by weight or pressure at the bottom, but also can be detected by floats, position detectors, or conductivity switches. The level of solids is more difficult to measure because hydrostatic pressure is not directly proportional to solids height in a bin. A variety of devices, using plum bobs, ultrasonics, and light are used to measure the distance to the top surface of solids, but these suffer from the tendency of solids to remain in a pile with an uneven surface. Bin weight is the best way to control inventory of solids.

Probes can be used to measure capacitance change, as liquids or solids cover more or less of the length, thus giving level indication around the location of the probe. Vibrating reeds or 'tuning forks' generate a change in signal when they are covered, thus indicating the presence of solids or liquid at their location in a bin. Capacitance detectors can also be used as point indicators of level, as can devices that detect change in resistance to rotation of a small paddle.

Composition

The refractive index correlates with the solids content of many liquid food materials, such as syrups, milk, and juices. Thus, instruments to measure refractive index can report results in terms of composition and can be used to control blending operations. Other properties found to correlate with solids content include viscosity and density. On-line instruments exist to measure these properties also.

Direct measurement of composition is quite difficult except by versions of analytical instruments, such as gas chromatographs, mass spectrometers, and pH meters. Ion content can be detected and quantified by ion-specific electrodes, and salt content can be measured by conductivity. Relatively few of these techniques are found directly connected to processes; more often, they are used as off-line quality control devices. (*See* **Chromatography**: Gas Chromatography; **Mass Spectrometry**: Principles and Instrumentation; **pH – Principles and Measurement**.)

Moisture

Moisture is very important in foods and is usually measured by drying a sample. Instruments exist to give moisture determination in minutes, but this is still too slow for direct process control. Water activity can be detected by a humidity sensor over a sample in a closed container, but again it requires minutes. Changes in radio-frequency signals or of microwaves can be used for noncontact, on-line sensing of moisture in some foods.

Moisture, fat, and protein content can be measured quickly for meats and other foods using infrared spectrometery. This has been an off-line technique, but it can be done on the process floor, giving results in time to modify a batch, if necessary. Reflected near infrared is used on-line to measure some composition changes, including moisture and oil content. (*See* **Protein**: Determination and Characterization; **Spectroscopy**: Near-infrared; **Water Activity**: Principles and Measurement.)

Color can be detected and quantified on-line and has been used to control baking ovens, and also to help sort out burnt potato chips (crisps). (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments; Properties and Determinants of Synthetic Pigments.)

Other Parameters

Counting, most often of packages and cases, can easily be done simply by breaking a light beam to a photo detector or by using proximity switches, which detect changes in capacitance or magnetic fields due to the presence of an object. Such switches are also used to detect position and control starting and stopping of machinery, such as accumulation conveyors.

Seal integrity, most often of heat-sealed polymer packages, such as snack foods, is difficult to detect quickly. Visual inspection is the most frequently used technique. Often, such packages have a small amount of inert gas, such as nitrogen, injected to displace oxygen; the 'pillowing' effect that this creates can be used to detect presence of leaks in seals by lightly compressing the package and measuring its resistance to deformation.

The integrity of cans and jars is measured by detecting the deflection of lids due to the internal vacuum. The vacuum is created when hot-filled contents cool or by steam-flushing of the head space. The deflection can be measured physically or by tapping to create a distinctive tone; 'duds' are rejected as likely to have leaked.

Vision systems are complex combinations of optical devices and computers to observe, quantify, and report on shapes. Such systems are becoming less expensive and more useful as the cost of computing power continues to decline. A vision system can check pizzas for the proper count and placement of garnish, such as pepperoni slices; it can look for and discard 'doubles' in a candy bar process; and it can detect the presence and proper placement of labels on containers. Special devices can read bar codes on cases and report data to an inventory management system.

Vision systems will become more complex and useful with time and will assume many duties now performed by humans, such as sorting, grading, trimming, and assembling.

Integration of Processes

Integration of processes means the linking of data from various sources to maintain smooth operation and the use of such data to improve operations. At a primitive level, such techniques as signaling downstream or upstream operations about stoppages can avoid waste and prevent damage to equipment. On a more sophisticated level would be recording data about stoppages for later analysis. A computer can be used to coordinate the signaling and to do the analysis. It can even write the report, recommending process improvements!

CIM encompasses both levels cited plus more. The basic idea is to avoid ever having to reenter data or information once it exists in a computer. For example, direct process data, such as case counts or batch weights, are available in a process control computer. It should not be necessary to obtain a report and reenter that data in another computer in order to calculate daily, weekly, or monthly production results. Nor should it be necessary to correlate data from two or more different sources to identify production problems and detect improvements. In a CIM environment, all data are available on a common data base shared among various computers linked electronically.

Computers included in such a network range from mainframes and minicomputers through desktop or microcomputers to factory-floor devices such as programmable logic controllers. The power found in each of these devices is increasing as the cost decreases, so the capabilities available to those running food plants are constantly improving. Current practice in modern food plants is to control nearly all equipment (pumps, conveyors, ovens, sterilizers, feeders, etc.) through operator-interface terminals, which have color graphic displays and keyboards, keypads, or touch screens.

The displays, of which there may be several dozen, usually show flow diagrams of the process and status of the equipment (on or off, temperature, etc.). Time histories of any measured parameter can be plotted on command, deviations from desired ranges flagged, and correlations tested among measured parameters.

Standard reports are generated, and data can be summarized for use at the next level of management. For example, on the operational level, the instantaneous rate of production is a concern (cans per minute, kilograms per hour, etc.); at the next level, the interest might be cases or pallets per day (calculated by summing all the actual cans, kilograms, cases, or pallets counted); at the plant and corporate levels, totals for weeks, months, quarters, and years are the only concern. If all the results, at whatever level, are calculated from the same basic data, with little human intervention, they will be consistent and available quickly, and people will be relieved of tedious and unnecessary time-consuming labor.

Barriers to this ideal situation have included incompatibilities among computers, languages, programs, and communications media. Additional barriers have been low computer literacy among workers involved and institutional resistance to sharing of information. Evolution and cooperation among vendors have led to the emergence of common protocols for exchanging information electronically, so that many of the previous technical barriers have fallen. Training and the use of computers in early school grades is reducing the literacy problem. The institutional resistance may be the most difficult, but as the significant benefits of computer integration are appreciated, this too will disappear.

Conclusion

The overwhelming importance of computers in process control and business systems, which are increasingly linked together, has a significant impact on food plant design. The need to have data compatible with computers influences the selection of instruments and measurement techniques. Instruments are being made 'smarter' by incorporating computing power in them with microprocessors, leaving only simple connections to control computers. Such smart instruments calibrate themselves, control drift, correct for changing conditions, and are inexpensive to install. The ability of fewer people to manage and control large processes reduces the labor requirements while raising the education and training requirements of the employees.

See also: **Canning**: Principles; **Chromatography**: Gas Chromatography; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; Properties and Determinants of Synthetic Pigments; **Mass Spectrometry**: Principles and Instrumentation; **Mercury**: Toxicology; **Pasteurization**: Principles; **pH – Principles and Measurement**; **Protein**: Determination and Characterization; **Spectroscopy**: Near-infrared; **Sterilization of Foods**; **Water Activity**: Principles and Measurement

Further Reading

Bernard JW (1989) *CIM in the Process Industries*. Research Triangle Park, NC: Instrument Society of America.

Food and Process Engineering Institute (1990) *Food Processing Automation Conference Proceedings*. St. Joseph, MI: American Society of Agricultural Engineers.

Schwartzberg H and Rao MA (1990) *Biotechnology and Food Process Engineering*. New York: Marcel Dekker.

Plantains *See* **Bananas and Plantains**

PLUMS AND RELATED FRUITS

C F Hansmann and J C Combrink, ARC Infruitec-Nietvoorbij, Stellenbosch, South Africa

Background

The European plum (*Prunus domestica*) probably originated in eastern Europe or western Asia around the Caucasus and the Caspian Sea. It has been known in Europe for more than 2000 years. Prunes are the most important subgroup. Others are the greengage, yellow egg, imperatrice, and lombard groups. The Japanese plum (*P. salicina*) originated in China and was domesticated in Japan. An American fruit breeder, Luther Burbank, developed many commercial cultivars from it after its introduction into the USA around 1870. Native plum species occur in some countries, such as North America (e.g., *P. americana, P. hortulana, P. subcordata*) and the UK (*P. spinosa*, blackthorn sloe; *P. institia*, damson or bullace), but these are not extensively grown commercially. Some are used in breeding programs with Japanese plums. Other plum types, used as rootstocks or in plum breeding programs, or occurring in home orchards, are *P. cerasifera* (myrobalan, cherry plum), *P. simoni* (apricot plum), and *P. institia*.

Plums are cultivated over a wide range of climatic conditions. In 2000, world production of plums was estimated at 8 830 000 tonnes (**Table 1**), representing about 8.5% of the world production of deciduous fruit. Fifteen countries produced more than 100 000 tonnes each, or more than 85% of the total production. The ratio of Japanese plum to European plum is approximately 99:1 for Japan, 60:40 for South Africa and Israel, 40:60 for Canada (Ontario) and Italy, and 0:100 for Yugoslavia (Montenegro, Serbia).

Anatomy and Morphology of the Fruit

The plum is a succulent fruit, called a drupe, and is formed by thickening of the ovary wall following fertilization. The floral remnants become senescent and drop off. The skin (epicarp) consists of a layer of elongated living cells (epidermis) covered by a thin film of cutin (cuticle). It protects the underlying tissue and permits exchange of metabolites or gases with the external environment through openings (lenticels). Wax deposited on the cuticle forms a light gray bloom that gives the skin a dull appearance and makes it impermeable to water. The thick, fleshy, edible portion (mesocarp) consists of parenchyma cells that have an active protoplast where all metabolic reactions occur. The protoplast is enclosed by a pectinaceous cell wall. Middle lamellae, which occur between adjacent cell walls, have a high calcium content and play a vital role in maintaining cohesion between cells, thus providing structural rigidity. The single seed is enclosed by a hard stone (endocarp) consisting of isodiametric cells with thick, lignified cell walls (sclerenchyma). The mesocarp usually clings to the endocarp, although some cultivars are freestone (i.e., the mesocarp is free from the endocarp).

Japanese plums are large and attractive, round or heart-shaped, with or without a prominent apex. The surface of some cultivars is covered by a heavy bloom. The flesh and skin color of ripe fruit may vary from

Table 1 World production of plums during 2000

Country	*Production ($\times 10^3$ tonnes)*
Countries producing more than 20 000 tonnes	
Africa	
Algeria	26
Egypt	21
Libya	33
Morocco	39
South Africa	55
Asia	
China	4192
India	74
Iran	130
Iraq	27
Israel	23
Georgia	49
Japan	119
Korea	52
Lebanon	41
Pakistan	81
Russia	120
Syria	26
Turkey	190
Uzbekistan	76
Europe	
Austria	57
Belarus	43
Bosnia and Herzegovina	90
Bulgaria	62
Croatia	40
France	214
Germany	316
Hungary	91
Italy	188
Moldova	27
Poland	107
Romania	345
Spain	156
Ukraine	85
Yugoslavia (Montenegro, Serbia)	370
North and Central America	
Mexico	61
USA	668
Oceania	
Australia	24
South America	
Argentina	69
Chile	158
Countries producing less than 20 000 tonnes	250
Total	8830

Data from Food and Agriculture Organization (2002) *FAOSTAT Agricultural Database (http://apps.fao.org/)*. Rome: United Nations.

yellow to blood red. The skin of the red-colored cultivars changes during ripening from green to red, usually starting at the apex. Fruit texture varies from firm and nonmelting to soft and melting. European plums are usually oval, with bulging of the central side, and compressed bilaterally. Their skin color is blue or purple, and they have a thick, meaty, freestone flesh. Fruit texture is usually firm and semimelting.

Physiology of Plum Fruit

Cultivar, climatic, environmental, and cultural factors significantly affect physiology. Cell division, multiplication, and differentiation of tissue occur during the first few weeks after fertilization. Cell enlargement and maturation follow. Numerous biochemical changes occur in plum fruits during the last few weeks prior to harvest, whilst the fruits are still attached to the tree. Many of these physiological changes are consistent and predictable, and can be used as indices for determining harvesting maturity. Substances synthesized in the leaves through photosynthesis are translocated to the fruit and transformed into products, which eventually determine the quality and nutritional value of the fruit.

The sugar and total soluble solids content increase throughout the period of growth. Organic acids accumulate in the fruit during the early stages of growth and then gradually decrease. The phenolic content of the fruit is high during the early stages, decreases, and then remains contant until harvest. Volatiles, which determine the flavor or aroma, are produced. Wax develops on the skin of the fruit. The fruit attains its full size and optimum maturity, although it is still unripe. (*See* **Phenolic Compounds**; **Sensory Evaluation**: Aroma; Taste.)

During ripening, pectic substances in the cell walls change from an insoluble to a soluble form, resulting in softening of the fruit. The chlorophyl (green pigments) content of the skin decreases, and the carotenoid (yellow pigments) and anthocyanin (red pigments) content increases. Based on its respiration pattern, the plum is a climacteric fruit. The respiration rate is high during and immediately after cell division. As maturity approaches, it decreases to the preclimacteric minimum and then increases irreversibly to a maximum (climacterium) during ripening. At the climacterium, the fruit are soft and sweet, with a characteristic flavor, and are ideal for eating. Subsequently, senescence sets in, whereupon the respiration rate decreases, and the fruits become overripe and decayed. (*See* **Ripening of Fruit**; **Phenolic Compounds**.)

Chemical Composition and Nutritional Value

Food composition tables generally do not specify the cultivar analyzed, nor whether a sample was fresh or cold-stored before analysis. Nevertheless, they are useful guides to the composition and nutritional value of fruit. A typical composition table for plums and related fruits is given in Table 2. The anthocyanins present in plums are 3-glucosides and rutinosides of cyanidin and peonidin. Refer to individual nutrients.

Table 2 Nutrient and mineral content of fresh plums, damsons, and prunes

Nutrient	*Content (per 100 g edible portion)*		
	Plum[a]	*Damson*[b]	*Prune*[a]
Moisture (%)	85.2	69.7	32.4
Energy value (kJ)	230.0	146	1000.0
Carbohydrate (g)	11.0	8.6	55.5
Dietary fiber (g)	2.0	1.6	7.2
Protein (g)	0.8	0.5	2.6
Total fat (g)	0.6	Trace	0.5
Nicotinic acid (mg)	0.5	0.3	2.0
Pantothenic acid (mg)	0.18	0.24	0.46
Riboflavin (mg)	0.10	0.03	0.16
Thiamin (mg)	0.04	0.09	0.08
Folic acid (μg)	2.0	3.0	4.0
Carotene (μg)	295[b]	265	140[b]
Vitamin B_6 (mg)	0.08	0.05	0.26
Vitamin C (mg)	10.0	5	3.0
Vitamin E (mg)	0.65	0.60	[c]
Sodium (mg)	0	2.0	4.0
Potassium (mg)	172	260	745
Calcium (mg)	4.0	22	51.0
Magnesium (mg)	7.0	10	45.0
Phosphorus (mg)	10.0	14	79.0
Iron (mg)	0.1	0.4	2.5
Copper (mg)	0.04	0.07	0.43
Zinc (mg)	0.10	0.10	0.53

[a]Data from Langenhoven ML, Kruger M, Gouws E and Faber M (1991) *MRC Food Composition Tables*, 3rd edn. Parow, South Africa: South African Medical Research Council.
[b]Data from Holland B, Welch AA, Unwin ID, Buss DH, Paul AA and Southgate DAT (1991) *McCance and Widdowson's -The Composition of Foods*, 5th edn. Cambridge: Royal Society of Chemistry and Ministry of Agriculture Fisheries and Food.
[c]No data available.

Taste is largely determined by a balance between the sugar and acid contents. Low acid and high sugar contents result in a bland taste, and high acid and low sugar contents in a sour taste. The contents of individual sugars and acids vary between cultivars, but citric acid as well as malic acid are present in the fruit. For example, in Australia, the cultivar Santa Rosa contains approximately twice as much malic acid as Mariposa, but has a lower soluble solids content. (*See* **Acids**: Natural Acids and Acidulants.)

Harvesting, Handling, and Storage

Plums are harvested in summer, from about December to February in the southern hemisphere and from about July to October in the northern hemisphere. It is important to harvest at the correct stage of maturity. If plums are harvested while still immature, the characteristic flavor will not develop, and fruit quality will be inferior. Plums harvested too ripe are very susceptible to injuries and fungal infection during harvest and postharvest handling, and their storage life will be short.

Dessert Plums

Dessert plums are harvested when they are physiologically mature but unripe. Ripening to a stage at which the fruit has a pleasant taste and aroma occurs during and, mostly, after storage. Plums that are to be sold soon after harvest must be picked at a riper stage than usual, or cold-stored for a few days to promote proper ripening. They require extra care during handling.

Total soluble solids content, estimated by means of a refractometer, and mesocarp firmness, measured by means of a probe (penetrometer) forced into the flesh, are good indicators of maturity. However, these tests are destructive, and pickers use skin color, in combination with these tests, as a maturity index. Fruit size is an additional criterion used to determine optimum maturity stage. Fruits are picked by hand into picking bags and carefully transferred to containers with a capacity of about 500 kg. Plums bruise easily and must be handled carefully during and after harvest. Filled containers are transported to a packhouse. Bruised, cut, or decayed fruit and culls are removed. The plums are sorted according to size and either hand- or mechanically packed into cardboard or wooden containers. These usually have a capacity ranging from about 5 to 20 kg. Trays or padding material immobilize the fruit in the container to prevent transit injuries. Packed containers are unitized (palletized) to facilitate handling and protect the fruit.

To extend the storage life of the fruit, its temperature must be lowered as soon after harvest as possible. Precooling, which removes field heat from the fruit, commences immediately after packing and palletizing. Cold air is distributed through the pallets in such a way that it comes into contact with all fruit. The ideal storage conditions for plums are a temperature of -0.6 to $0\,^{\circ}C$ and a relative humidity of 85–90%. The storage period depends on the cultivar. Most cultivars cannot be stored for longer than 3–4 weeks, but others can be stored for up to 3 months. The fruit of some cultivars are susceptible to chilling injury. To prevent chilling injury, the temperature is raised from $-0.6\,^{\circ}C$ to about $7\,^{\circ}C$ after 7–10 days, and this temperature is maintained for the rest of the storage period. This is called dual-temperature storage. Controlled- or modified-atmosphere storage increases the storage life of some cultivars. (*See* **Controlled-atmosphere Storage**: Effects on Fruit and Vegetables; **Storage Stability**: Mechanisms of Degradation.)

Prunes

Prunes are harvested when fully ripe and in an ideal condition for fresh consumption. Overmature fruit

discolor and break down during drying. The fruit of some cultivars drop from the trees naturally and must be gathered promptly to prevent losses. Soluble solids content and flesh firmness are important maturity indices. As in the case of fresh plums, these indices are related to skin color.

Most prunes are dried rather than sold fresh. Fresh prunes can be cold-stored, but their storage life is shorter than that of dessert plums. Sorted prunes are dipped into a hot caustic solution (generally 0.5% sodium hydroxide at boiling point) for sufficient time to remove the wax layer and cause minute cracks in the skin. This facilitates moisture loss and speeds up drying. The prunes are then packed on to trays and dried in the sun. Prunes can also be dried in mechanical dehydrators, usually tunnel dehydrators. The fruits are washed and placed on trays in the dehydrator without application of a lye treatment. Hot air circulates through the trays, and drying takes place within 24 h. Dehydrators used for drying prunes are frequently operated in the parallel-flow mode (concurrent movement of fruit and heated air through the drying section), with the fruit entering the dehydrator at air temperatures of up to 90 °C. Prunes are gathered when dry but still pliable, with a moisture content of about 20%. Considerable degradation of the red pigments present in the fresh fruit occurs during drying. (*See* **Drying**: Drying Using Natural Radiation; Theory of Air-drying.)

Market Disorders and Diseases of Plum Fruit

Disorders

Disorders are transit- or storage-related. Transit-related disorders include bruising as a result of rough handling, freezing injury owing to temperatures below the freezing point of plums, and shriveling owing to moisture loss caused by low relative humidities in the storage atmosphere. These disorders can be controlled by effective postharvest management.

Storage-related disorders are similar to transit-related disorders but include internal breakdown or internal browning. When a fruit is cut open, internal browning can be seen as a reddish brown discoloration, which is more intense around or near the endocarp. It is responsible for extensive losses during cold storage. It usually occurs at low temperatures and can be prevented by storing fruit at slightly elevated temperatures. Some cultivars are more susceptible than others. Dual-temperature storage is used where plums have to be shipped over long distances.

Diseases

Fungi cause market diseases. The most important postharvest diseases of plums and prunes are brown rot (*Monilinia laxa*, *M. fructicola*), blue mold rot (*Penicillium expansum*), gray rot (*Botrytis cinerea*), mucor rot (*Mucor piriformis*), and rhizopus rot (*Rhizopus* spp.). These fungi cause decay in a wide range of commodities, and their occurrence is not limited to plums. Fungal spores present in the orchard on plant debris or in the air are dispersed to fruit by wind or insects. The spores of *B. cinerea* and *Monilinia* spp. can infect uninjured fruit in the orchard, especially during the 1–2 weeks prior to harvest, when the susceptibility of fruit to infection increases. However, disease symptoms are only expressed during cold storage. The other fungi are wound pathogens that penetrate fruit through skin breaks caused by rough handling or insects. Decay symptoms are related to the fungus involved. Mucor and rhizopus rots develop rapidly, and affected plums become soft and watery, and are covered by black spore masses. Brown rot starts as a small, water-soaked spot, which enlarges rapidly and becomes brown or black. The skin becomes leathery but remains intact. Fruit infected by *B. cinerea* are firm, spongy, and only slightly moist. The skin covering lesions readily slips away when slight pressure is applied. Decay development is slower, and, in advanced stages of decay, gray spores cover the affected areas. Blue mold rot lesions are wet and soft and covered with blue–green mold growth. (*See* **Spoilage**: Molds in Spoilage.)

Control of postharvest decay involves the integration of preharvest factors (soil preparation, spray programs, orchard hygiene, etc.) with sound postharvest crop management. Fruit produced under optimal conditions possess a high natural resistance to infection. Proper sanitation in the orchard and packshed keeps the number of spores to a minimum and is an important precautionary measure. All fungi causing postharvest decay, except *Rhizopus* spp., can grow at 0 °C, and cold storage does not prevent decay development. Chemicals can be used to control decay, but only a few are registered for postharvest use. (*See* **Fungicides**.)

Industrial Uses

Plums have a wide variety of industrial uses. Japanese plums are mostly eaten fresh, although a small percentage are dried. European plums are mostly dried (prunes), cooked (greengages), canned (greengages, yellow egg, damsons, other European plums), or used in jams (plums, greengages, damsons) or jellies (American plums). Prune juice is used as a laxative and plum purée as a baby food. Some plums are used

as rootstocks (e.g., marianna, myrobalan) or in breeding programs (*P. simoni*) and are planted as ornamentals in home gardens. In some parts of Europe, plums are fermented and distilled into a 'brandy.' In Romania, Hungary, and Yugoslavia (Montenegro, Serbia) it is called slivovitz, and in other parts of Europe the name depends on the cultivar used (e.g. Quetsch or Mirabelle).

See also: **Acids**: Natural Acids and Acidulants; **Controlled-atmosphere Storage**: Effects on Fruit and Vegetables; **Drying**: Theory of Air-drying; Drying Using Natural Radiation; **Fungicides**; **Phenolic Compounds**; **Ripening of Fruit**; **Sensory Evaluation**: Aroma; Taste; **Spoilage**: Molds in Spoilage; **Storage Stability**: Mechanisms of Degradation

Further Reading

Chandler WH (1965) *Deciduous Orchards*, 3rd edn. Philadelphia, PA: Lea & Febiger.

Eksteen GJ and Combrink JC (1987) *Manual for the Identification of Post-harvest Disorders of Pome and Stone Fruit*. Stellenbosch, South Africa: Van der Stel Printers.

Harvey JM, Smith WL, Jr. and Kaufman J (1972) *Market Diseases of Stone Fruits. US Department of Agriculture, Agriculture Handbook 414*. Washington, DC: US Government Printing Office.

Holland B, Welch AA, Unwin ID, Buss DH, Paul AA and Southgate DAT (1991) *McCance and Widdowson's The Composition of Foods*, 5th edn. Cambridge: Royal Society of Chemistry and Ministry of Agriculture Fisheries and Food.

Langenhoven ML, Kruger M, Gouws E and Faber M (1991) *MRC Food Composition Tables*, 3rd edn. Parow, South Africa: South African Medical Research Council.

Macheix JJ, Fleuriet A and Billot J (1990) *Fruit Phenolics*. Boca Raton, FL: CRC Press.

Ramming DW and Cociu V (1990) Plums (*Prunus*). In: Moore JN and Ballington JR, Jr. (eds) *Genetic Resources of Temperate Fruit and Nut Crops*, vol. 1, pp. 235–281. Wageningen, The Netherlands: International Society for Horticultural Science.

Salunkhe DK, Bolin HR and Reddy NR (1991) *Storage, Processing, and Nutritional Quality of Fruits and Vegetables*, 2nd edn, vol. 1. Boston, MA: CRC Press.

Teskey BJE (1978) *Tree Fruit Production*, 3rd edn. Westport, CT: AVI.

Wills RBH, Scriven F and Greenfield H (1983) Nutrient composition of stone fruit (*Prunus* spp.) cultivars: apricot, cherry, nectarine, peach and plum. *Journal of the Science of Food and Agriculture* 34: 1383–1389.

POLITICS AND NUTRITION

C Geissler, King's College London, London, UK

Introduction

Politics can be defined as both the art and science of government, as well as the play of competing economic and ideological interests. This article reviews some aspects of government nutrition policies and the effects of disparate interests on nutrition in developing and developed countries.

Nutrition Policies

Governments in both developed and developing countries have over the last two decades increasingly focused attention on nutrition through explicit national nutritional policies. (*See* **Nutrition Policies in WHO European Member States**.)

Hunger and malnutrition were put on the international agenda by the League of Nations in the 1930s and remained an important focus of the United Nations (UN) technical agencies, the Food and Agricultural Organization (FAO), World Health Organization (WHO), and United Nations Children's Emergency Fund (UNICEF), which were created immediately after World War II. The types of nutrition interventions and development policies supported were shaped by the World Food Surveys conducted every decade by the FAO (these estimated the extent and defined the causes of malnutrition) and by post-colonial theories of economic development. During the 1950s and 1960s, the prevailing policy encouraged industrialization and large-scale agriculture in order to increase economic wealth, from which improved nutrition would 'trickle down' rather than be achieved by improved distribution. Nutrition continued to be approached mainly through piecemeal interventions, including child-feeding programs and nutrition education, but also via some integrated village-level 'Applied Nutrition Programs', that addressed the economic, educational, and food resource constraints on good nutrition. (*See* **World Health Organization**.)

By the early 1970s, it had become clear that rapid economic growth had led to increased inequalities, as well as increased absolute poverty in some countries such as Pakistan, Nigeria, and Brazil, resulting in greater levels of undernutrition in the impoverished.

Dissatisfaction with the effectiveness of past approaches in improving malnutrition led to the concept of National Nutrition Planning, based on the coordination of sectoral policies to focus on nutrition. The aim was to remove the barrier to economic growth of malnutrition that results in increased morbidity and mortality, poor educational achievement, low work capacity, increased absenteeism, and low productivity. This integrated, systematic, intersectoral approach was endorsed and promoted by the FAO and the United States Agency for International Development (USAID), that assisted several countries to prepare explicit integrated nutrition policies. (*See* **Malnutrition**: The Problem of Malnutrition.)

In this international context, the need for integrated national food and nutrition policies was also expressed in the developed countries, where the malnutrition of poverty had been largely eliminated but where malnutrition of affluence, reflected in increased rates of coronary heart disease, diabetes, obesity, and cancer, is the current concern. Several countries developed national dietary goals in the 1960s and 1970s to encourage their populations to consume a more healthy diet, but Norway was the first to pass an intersectoral Food and Nutrition Policy in 1975. (*See* **Malnutrition**: Malnutrition in Developed Countries.)

During the 1980s, this extent of integration fell out of favor because of practical difficulties, and the international community concentrated on improving the nutritional impact of agricultural and other sectoral development programs. In the 1990s, the focus moved to micronutrient interventions to eradicate the most prevalent vitamin and mineral deficiencies. However, in 2000, the pendulum of opinion is swinging away from specific 'magic bullet' interventions and back to calls for more integrated approaches to national policy.

Types of Nutrition Interventions

The types of interventions that form part of national nutrition policies tend to be limited to palliative measures such as vitamin supplementation, nutrition education, and child-feeding programs, because the underlying political issues that lead to malnutrition, whether undernutrition or overnutrition, involve fundamental economic and political interests. These are much more difficult and contentious issues and rarely features overtly in nutrition policies. (*See* **Community Nutrition**.)

Famine

One extreme example of the political aspects of nutrition is famine, with mass starvation and death, often attributed to climatic causes such as drought or flooding. Such factors may precipitate famine, but it is only the most vulnerable members of society who starve in this type of famine. To quote Amartya Sen, 'starvation is the characteristic of some people not *having* enough food to eat. It is not the characteristic of there *being* not enough food to eat.' He argues that poverty and famines are related to 'entitlement' or the ability of individuals to obtain food through the legal means available in the society, including production possibilities, trade opportunities, state provisions, and other methods of acquiring food. When individuals can no longer obtain enough food in exchange for their products or services, they suffer from famine, as happens with food price rises owing to local shortages and hoarding. Several examples exist where aggregate food supplies were adequate but certain groups could not afford to buy them, such as the Irish famine of 1846, the Bengal famine of 1943, and more recent famines in the Sahelian countries.

War

War is an extreme example of political crisis and is an increasingly common contributor to the causes of famine through the blockade or diversion of food supplies, the destruction of cropland and the migration of farmers away from their lands in the war zones. War contributed to the famines in Bengal, The Netherlands and Warsaw (World War II) and more recently in Nigeria, Timor, Kampuchea, Laos, Vietnam, Afghanistan, Chad, Uganda, and the Sudan. War and its aftermath also have an strong impact on nutrition, short of famine. One example is the 1990–1991 Gulf War, which resulted in extensive destruction of Iraq's infrastructure and economy so that widespread malnutrition, particularly of children, has occurred.

Refugees

War and political crises also create refugees. The number of people in the world who have either fled their country as refugees, or have been displaced internally, mainly as a result of civil war increased greatly during the 1980s and 1990s and at the beginning of the new millenium stands at around 50 million. Refugees suffer from the same type of diseases as other vulnerable groups in developing countries, the main killers being measles and diarrhea, to which they are more susceptible because of malnutrition. They often receive food inadequate in quantity and quality. If prolonged, this deprivation leads to starvation, and debilitating outbreaks of scurvy, pellagra, beriberi, and other deficiency diseases are now common. This is attributable to the dependency of refugees on the food provided by governments and aid donors, which

has little diversity and is inadequately planned. (*See* **Refugees; Scurvy.**)

Poverty

Famine is only the tip of the iceberg of the larger problem of poverty and access to resources, which affects nutrition on a longer-term basis. Various economic models or political views exist on the causes of poverty and wealth creation, ranging from simplistic models of idleness and ignorance to systematic models such as the Marxist model, which views poverty as an inevitable and even necessary part of the capitalist economic system. These models form the framework of what policymakers believe can be done to break out of poverty and malnutrition, what the implications are for the rest of the economy, and therefore the type of interventions they support.

Poverty reduction

Poverty reduction became a specific target of the international community: an International Development Goal was set in 1990 to halve poverty and was expected to be met within 20 years in most areas of the world. However, the financial crisis later in the decade led to a rise in levels of poverty and inequalities, setting back progress and leading to renewed calls for debt relief.

Debt relief

Some countries are unable to tackle poverty adequately because past loans have led to a level of debt that is now too high for them to repay. For example, in Africa as a whole, governments spend four times as much paying back their debts to developed countries in repayments as they do on health and education. In 1996, the International Monetary Fund (IMF), the World Bank, Regional Development Banks and other external creditors launched the Heavily Indebted Poor Countries Initiative (HIPCI) to reduce the debts of the poorest countries who are deemed eligible by their economic track record over the previous 5–6 years. However, few countries qualified (Uganda and Bolivia), and in 1998, a fundamental review was agreed upon to speed up the process and increase the level of relief.

Equity

Poverty is a major, but not the only, determinant of undernutrition. At the same level of family or national income, some individuals or countries do better than others in terms of nutrition. Internationally, there is a clear relationship between gross national product (GNP) and several indicators of health and nutrition, including food available for consumption, infant mortality, etc. Countries that have done best in improving nutrition in recent decades are those where there is a greater equity or where policies have concentrated on insuring the satisfaction of basic needs, including adequate food. These include a wide spectrum of political ideologies from communist China to capitalist South Korea. China has been a classic example of a country that was still poor but conquered millennial problems of malnutrition and famine – except during the disastrous Great Leap Forward policy of rural industrialization at the end of the 1950s – through effective organization of food production and distribution. Other examples are Kerala state in India, Sri Lanka, Vietnam, Indonesia, Malaysia, and Thailand, which have had better nutrition conditions than other countries with similar GNPs. In contrast, some countries have extensive chronic malnutrition, e.g., Brazil and some other Latin American countries (Bolivia, Ecuador, Gwatemala), most countries in Sub-Saharan Africa, and Bangladesh despite massive aid.

Since the early 1980s, there has been a widespread rise in income inequalities with an acceleration in the 1990s in several countries, including all economies in transition such as China. This has challenged the poverty reduction targets set by the international community.

International Relations

Poverty and malnutrition are controlled not only by national policies and resource distribution but also by the power play of international economic relations. Again, opposing political views shed very different lights on their effects.

On one side are those who believe that the rich industrialized countries and the international agencies do much to assist poor developing countries through technical advice and monetary or food aid, while international corporations provide them with employment opportunities.

On the other side are those who believe that the odds are stacked against poor countries breaking out of poverty, as the countries that became industrialized first now hold economic power internationally, dominating trading and banking systems. In this view, national elites of many countries assist the world powers to dominate their economies through concessions to multinational companies, the acceptance of aid, and requests for finance from the international banking system.

During the 1980s, the latter view received wider acceptance and led to calls for a 'new economic order' by such groups as the Brandt Commission, which warned that the inequalities of opportunity between

the North (the rich industrialized countries) and the South (the poor developing countries) could lead to political and environmental instability, and it was therefore in the interests of both to allow the South greater control of its destiny.

The questions of international food trade, food aid, and the role of international agencies in relation to poverty reduction and nutrition are therefore highly contentious political issues that can be touched on only briefly here.

Food Trade

The economies of the developing countries are affected by the policies of the major producers of food and other essential commodities. For example, the 'world food crisis' in 1973, which was marked by soaring prices of grain and other food commodities on the world market, was the result of several factors. These included climate, the USA policy of deliberately reducing its grain stocks by taking land out of production to maintain producer prices, and Russia's sudden increase in its grain purchases from the world market to support its policy of livestock production. These, and the oil crisis, caused great economic hardship for many developing countries that depended on imports for their food supply, and led to a spate of policies to increase food self-sufficiency.

Similarly, the Common Agricultural Policy (CAP) of the European Union (EU) has had a controversial impact on the Third World through price fluctuations caused by periodic dumping of surplus food on the world market and through tariff barriers that restrict imports. Some trade concessions were made by the EU to the developing countries through the Lome Convention, but this provided little protection in practice. (*See* **European Union**: European Food Law Harmonization.)

These protectionist policies are the subject of international bargaining for free trade that is carried out in periodic rounds of talks in the council on the General Agreement on Trades and Tariffs (GATT). However, these mainly reflect the interests of the seven most economically powerful nations, the USA, Canada, Germany, France, Italy, the UK, and Japan (the G7) and only in the 1986–1992 round included food commodities, prompted by the food price policies that are detrimental to the food trade of the USA and other countries.

GATT was succeeded by the World Trade Organization (WTO) from 1995, which, by the end of the decade, had 134 members, accounting for 90% of world trade and 30 others, including China, negotiating for membership. The costs of participation are high for the poorest countries, which means that the agenda is still dominated by the large economies.

Food Aid

Food and other international aid has often been used as a political weapon, given to politically friendly countries, such as South-east Asia in the late 60s and to Egypt and Syria in 1974, and withdrawn from those that do not conform to the policies of the donor countries, such as Chile in 1971 and Mozambique in 1981.

Food aid has also received much criticism on the grounds that it has been used by the donor countries to dispose of surpluses and to penetrate food markets of developing countries in order to create a long-term trade demand, that it is unreliable, involves excessive opportunity costs, reinforces expensive subsidy programs, provides opportunities for corruption, and is used by governments to keep urban prices down, thereby acting as a disincentive to agriculture and maintaining rural poverty and malnutrition.

Food aid and capital intensive agricultural policies are also seen to contribute to the rapid rural urban migration seen in so many developing countries, where cities are unable to cope with the rate of influx to provide adequate housing, sanitation, water supply, and employment. As a result, the migrants can often only scrape a living through petty trading, begging, and low wage employment. This influx undermines minimum wage legislation or enforcement and keeps the cost of labor, and therefore incomes, low. The consequent inability of the urban poor to purchase adequate diets, and the insanitary living conditions, result in extensive urban malnutrition in many developing countries. Such effects are seen to negate the short-term benefits of food aid used in famine situations, and of child-feeding and food-for-work projects that act as welfare benefits or resource transfers to the poor.

The structure of food aid has altered since it became a permanent international transfer mechanism in the 1954 when the USA enacted the public law PL480, referring to sales to 'friendly nations,' to be paid initially with local money deposited as counterpart funds for use by the USA or with their approval (Title 1), and donations for famine relief and projects (Title II). The latter represents a small proportion contrary to the popular belief that all aid is a gift. Initially, the USA supplied nearly all aid, but now the EU, Japan, Canada, Australia, and other countries contribute about 50%. About 35% of food aid is channeled through the multilateral World Food Program for feeding projects; this reduces political bias. However, 40% is still bilateral, and several countries, such as Bangladesh, remain dependent on food aid and, therefore, politically dependent on the donor. There is widespread recognition of the need for reform of food aid and its international infrastructure

to better handle hunger and meet targets in poverty reduction, but consensus is difficult to reach.

Structural Adjustment

Over the last decades of the millennium, many developing countries experienced severe economic crises owing to rapid changes in oil prices, falling and unstable prices of export commodities, rapidly increasing rates of interest, and increasing dependence on foreign borrowing, resulting in reduced foreign exchange reserves, and inability to service debts.

To deal with the crisis, countries had to implement a variety of 'macroeconomic adjustment policies,' including reductions in government spending. These conditionalities have been rigidly imposed by the International Monetary Fund (IMF) and the World Bank (International Bank for Reconstruction and Development (IBRD)), to obtain new financial loans. These institutions are funded by quotas from members who have voting rights in proportion to their contribution, assessed according to economic status. The USA has the majority of the votes, followed by the European countries, and others very small percentages, so that decisions are effectively in the hands of the major industrialized countries, especially the USA. This banking structure means that the policies of borrowing countries are dictated by the industrialized nations.

The primary aim of macroeconomic adjustment policies is to improve the balance of payments. Therefore, short-term effects on the poor have often been ignored unless they threaten political stability, e.g. through urban riots. Adjustment has frequently included changes that are of particular concern to the poor, such as increased food prices and decreased expenditure on social programs. The effects of these policies on health care, food consumption, incomes, and prices have led to a serious deterioration in indicators of nutrition, health status, and school achievement in several countries. Efforts have been made by UNICEF and other bodies to buffer vulnerable groups from these effects.

Affluence

In most developed countries, the problems of poverty-related malnutrition have been largely overcome by higher incomes and a safety net of welfare benefits, and replaced by the problems associated with affluent diets. The politics of nutrition in this case are very different from those associated with poverty. Only a small proportion of the population is involved in primary food production (2–3% in the US and UK), and the majority of the population depends on the food industry, which is a major industry in economic terms in Europe and the USA: in the UK agriculture, food processing and distribution contribute more than 7% to the GDP. Therefore, the agricultural and food industry lobbies are potentially important in food and nutrition policy formulation. In food policy, nutrition generally has a very low priority. For example, the EU CAP has no explicit nutritional component, its main aims being to support farm incomes and food security for Europe in terms of adequate supplies and stable prices. The emphasis is therefore on quantity, not on nutritional quality. The concept of food quality has lagged behind current values and has generally referred to the purity, hygiene, and richness of nutrients. These are all extremely important, but now that nutritional deficiencies have been largely overcome, the concept of a healthy diet can no longer be related only to purity and abundance. Government policies have also lagged behind current needs such that food legislation and producer incentives continued to favor a high fat diet, including premium producer prices for high-fat milk and inappropriate grading systems for livestock.

Efforts during the 1980s to translate general nutritional advice on healthy eating into explicit policies were initially met with considerable concern and opposition from some sections of the food industry that were threatened with change. The debate included the following arguments: the evidence for the relationship between diet and disease is not adequately conclusive; the level of risk to the population as a whole for certain dietary components such as salt does not warrant blanket nutritional policies, only targeted interventions to those at risk; and in a democratic society, the individual must be free to make his or her own choice of food consumption without coercive pressures from a 'nanny state.'

This initial resistance of the food industry and of government was viewed in some quarters as a conspiracy of vested interests between food producers, government policymakers and advisory committees. However, others viewed it as a natural cautious attitude that is necessary because of the widespread industrial and economic changes entailed. Despite this resistance, the food industry has responded to pressures for change in various ways including research into leaner livestock, the formulation of a variety of low-fat and low-energy products, and voluntary nutritional labeling. Also, since 1993, UK health policy explicitly addresses nutrition and associated 'diseases of affluence,' including obesity and heart disease.

Much stronger political support exists for interventions based only on informed choice, and little for coercive measures to improve the diet such as price manipulation through consumer subsidies and taxation, even though these instruments are used for other purposes. There is therefore much effort put into legislation on food labeling.

The food industry has become progressively more vertically integrated, from production through processing to retailing, by the development of multinational agribusiness and large chain food stores, that increasingly order food directly from the producers. This means that the buyers for these food chains become exceedingly important arbiters of the national diet in deciding the choice available to the consumer.

Emerging Issues

The creation of the WTO and the increasing power of large food corporations have highlighted several issues of control over food, from consumer pressures to national and international governance.

New Technologies

Consumer and environmental lobby groups have put a brake on the use of genetically modified (GM) foods: they disrupted the WTO meeting in Seattle in early 2000 by protests against the trade in GM crops; they caused agribusiness to discontinue the development of terminator genes to prevent plant replication; and consumer pressure in developed countries such as the UK has curtailed the sale of GM products as well as the conduct of field trials.

Resistance to the use of the new technology in agriculture and food production is based on several fears: the safety for consumption; the control over world food supply of the large corporations that develop the technology; the potential detrimental effects on the environment from cross-pollination; and the lack of benefits to developed country consumers as opposed to the seed and pesticide manufacturers. These arguments currently outweigh the arguments for the potential benefits, within appropriate socio-economic structures, to farmers and consumers in other parts of the world by the development of crops that are resistant to adverse climatic conditions and pests and that have higher yields and content of specific nutrients, factors that could expand the areas of crop production, increase total food production, and hence price in line with the expected population growth, and improve nutritional value.

Governance

Such consumer pressures have also led to other issues of governance in relation to national and international political bodies that control economics, including food. For example, consumer and professional pressures following a series of food safety scares in the UK during the 1980s have led to the separation within government of the structures that regulate the food supply from those that regulate agriculture as there are potentially conflicting interests between consumer and producer. A Food Standards Agency was established in 2000 to regulate food 'from farm to fork.' The EU has subsequently proposed a similar structure at European level. This brings to the fore the tensions in governance between national and regional control over food and the extent to which these are dominated by WTO rules in the interests of free trade. Similar issues arise in global governance, underlining the need for new arrangements to make more effective the relationships between the UN, WTO, G7, IMF, and other bodies.

Conclusion

Food and nutrition involve highly political issues that determine both access to sufficient food and the commercial and consumer pressures on a balanced diet.

See also: **Community Nutrition**; **European Union**: European Food Law Harmonization; **Malnutrition**: The Problem of Malnutrition; Malnutrition in Developed Countries; **Nutrition Policies in WHO European Member States**; **Refugees**; **Scurvy**; **World Health Organization**

Further Reading

ACC/SCN (1997) *Nutrition and Poverty. Nutrition Policy Paper 16*. Geneva: ACC/SCN.

ACC/SCN (2000) *Fourth Report on the World Nutrition Situation*. Geneva: ACC/SCN.

Atkinson J (1996) *Farmers, Food and the WTO. Oxfam Policy Briefing Paper.*

Berg A (1973) *The Nutrition Factor. Its Role in National Development*. Washington, DC. The Brookings Institution.

British Nutrition Society (1991) Nutrition and Development: symposium & workshop. *Proceedings of the Nutrition Society* 51(1): 71–107.

Food Industry, Nutrition and Public Health (1997) Symposium. *Proceedings of the Nutrition Society* 56(3): 807–888.

George S (1984) *Ill Fares the Land. Essays on Food, Hunger and Power.* London: Writers and Readers Publishing Cooperative Society.

Jolly R and Cornia GA (eds) (1984) *The Impact of World Recession on Children. UNICEF Report*. New York: Pergamon Press.

Overseas Development Institute (2000) *Reforming Food Aid: Time to Grasp the Nettle? ODI Briefing Paper, January.* London: ODI.

Royal Society (1998) *Genetically Modified Plants for Food Use*. London: Royal Society.

Sen A (1982) *Poverty and Famines: an Essay on Entitlement and Deprivation*. Oxford: Oxford University Press.

Tansey G and Worsley T (1995) *The Food System. A Guide*. London: Earthscan.

UN (1997) *Renewing the United Nations: A Programme for Reform. Report by the Secretary General. New York: UN.*

World Health Organization (1988) *Nutrition Policy Experiences in Northern Europe. Report on a WHO Consultation*. Copenhagen: WHO EUP/ICP/NUT.

POLYCYCLIC AROMATIC HYDROCARBONS

E Menichini and B Bocca, Istituto Superiore di Sanità, Rome, Italy

Introduction

Polycyclic aromatic hydrocarbons (PAHs) are a large class of organic compounds containing two or more fused aromatic rings. In particular, the term PAHs refers to compounds containing only carbon and hydrogen atoms (i.e., unsubstituted parent PAHs and their alkyl-substituted derivatives), whereas the more general term 'polycyclic aromatic compounds' also includes the functional derivatives (e.g., nitro-PAHs) and the heterocyclic analogs (e.g., aza-arenes). The term 'polynuclear' is frequently used for 'polycyclic.' Hundreds of PAHs may be formed. Over 100 PAHs have been identified in atmospheric particulate matter and about 200 in tobacco smoke. PAHs have very low water solubility, and are soluble in many organic solvents and highly lipophilic.

The nomenclature codified by the International Union of Pure and Applied Chemistry (IUPAC) is also used by the Chemical Abstract Service. Some compounds, however, are commonly reported in literature with nonsystematic names. Table 1 lists the PAHs considered for this review. The selection is based on their occurrence data (as being the most frequently determined in environmental and food analysis) and their toxicological concern. Benzo[*a*]-pyrene (BaP) is the most widely studied compound and most information on occurrence and toxicity of PAHs is related to it. In the comparison of different matrices, BaP has frequently been (and still is) used as an indicator for the PAH class, as regards both the levels of contamination and the carcinogenic risk.

Formation and Sources

PAHs are formed during incomplete combustion or pyrolysis of organic material, such as coal, wood, oil products, and garbage. Consequently, their formation is typically associated to the processing of coal and crude oil, power generation plants, incinerating plants, various industrial processes, domestic and residential heating (especially by wood and coal), motor vehicle exhaust, fires (forest, agricultural, cooking), and tobacco smoking. Volcanoes may represent a natural source with a locally relevant impact. Because of these numerous and widespread sources, PAHs are ubiquitous and occur in all environmental

Table 1 Frequently determined polycyclic aromatic hydrocarbons (PAHs) and their carcinogenicity

PAH[a]	*Abbrev.*[b]	*Relative molecular mass*	*IARC classification of carcinogenicity*[c]	
			Animal[d]	*Human*[e]
Phenanthrene	PHE	178.2	I	3
Fluoranthene	FA	202.3	I	3
Pyrene	PY	202.3	I	3
Benz[*a*]anthracene	BaA	228.3	S	2A
Chrysene	CHR	228.3	L	3
Benz[*e*]acephenanthrylene or benzo[*b*]fluoranthene[f]	BbFA	252.3	S	2B
Benzo[*j*]fluoranthene	BjFA	252.3	S	2B
Benzo[*k*]fluoranthene	BkFA	252.3	S	2B
Benzo[*a*]pyrene[g]	BaP	252.3	S	2A
Benzo[*e*]pyrene[h]	BeP	252.3	I	3
Benzo[*ghi*]perylene	BghiP	276.3	I	3
Indeno[1,2,3-*cd*]pyrene	IP	276.3	S	2B
Dibenz[*a,h*]anthracene	DBahA	278.4	S	2A

[a]International Union of Pure and Applied Chemistry (IUPAC) names, ranked according to increasing molecular mass and, for isomers, in alphabetical order.
[b]As used in this text.
[c]Classification according to International Agency for Research on Cancer (IARC, 1987). *Overall Evaluations of Carcinogenicity: An Updating of IARC Monographs*, vols 1–42. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*, Supplement 7. Lyon: International Agency for Research on Cancer.
[d]Degree of evidence for animal carcinogenicity: I, inadequate; L, limited; S, sufficient.
[e]Overall evaluation of carcinogenicity to humans: 2A, probably carcinogenic; 2B, possibly carcinogenic; 3, not classifiable.
[f]Non-IUPAC, commonly used name.
[g]3,4-Benzopyrene in older publications.
[h]1,2-Benzopyrene in older publications.

compartments, which they enter mostly via the atmosphere.

During any formation process, and subsequently in the matrices to which the population is exposed (air, water, soil, and food), PAHs are always present as a class and not as individual compounds, in complex mixtures also containing other chemical classes.

Most PAHs have no use other than for research purposes. Exceptions are limited to an occasional use of some PAHs as intermediates in the production of plasticizers, pigments, dyes, and pesticides; naphthalene is used as a moth-repellant for clothing.

Sources in Foods

In nonprocessed foods, the occurrence is mainly attributed to the environmental contamination: deposition of airborne particles (e.g., grain, fruits, and vegetables), uptake from contaminated soil (e.g., potatoes), absorption from contaminated fluvial and marine waters (e.g., mussels, fish, and shellfish). Common sources in processed foods are thermal treatments (namely grilling, roasting, baking, and frying), and processing procedures. The latter include especially the drying process by combustion fumes, e.g., vegetables oils, and the smoking process by traditional methods.

PAHs may leach into drinking water from coal-tar or asphalt coatings on storage tanks and water distribution pipes.

Toxicological Effects

Carcinogenicity of PAHs

The toxicological concern for PAHs is due to their carcinogenicity; other effects turn out to be by far less important. The International Agency for Research on Cancer (IARC) evaluated the carcinogenic evidence for 43 PAHs. The evidence in experimental animals was evaluated as 'sufficient' for 12 of them, and 'limited' for 12 others. No individual PAH could be evaluated conclusively as a human carcinogen. This is a consequence of the fact that human exposure, via both inhalation and ingestion, always occurs simultaneously with complex mixtures containing a large number of PAHs (with variable composition) and other potentially carcinogenic compounds. Based on an overall evaluation of available data, however, the IARC could classify three PAHs (including BaP) as 'probably' carcinogenic to humans and nine as 'possibly' carcinogenic (Table 1; five possibly carcinogenic PAHs are missing from the table and are not considered in this review, due to the scarcity of their occurrence data and to their very low, or undetected, concentrations reported in any matrix). For a number of PAH mixtures (also containing other potentially carcinogenic chemicals) occurring in occupational environments, there is epidemiological evidence of increased risk for lung cancer and, in some cases, for skin (following dermal exposure) and bladder cancers; it is plausible to attribute these risks partially or predominantly to PAHs.

Carcinogenic Risk of BaP by Ingestion

No adequate epidemiological data are available concerning carcinogenicity by ingestion of PAHs or PAH mixtures. The experimental data are scarce and limited to BaP. The latter has been shown to be carcinogenic in experimental studies when given by diet. Oral administration of BaP to mice induced gastric papillomas and squamous-cell carcinomas and increased the incidence of pulmonary adenomas. Other studies in mice indicated that it can cause carcinomas distal to the point of application. The incidence of gastric tumors was 70% or more in mice fed 50–250 ppm BaP for 4–6 months. In another study with rats, which were fed diets containing BaP, tumors were observed in the forestomach, esophagus, and larynx.

Based on these experiments, the US Environmental Protection Agency made a quantitative assessment of carcinogenic risk for oral exposure to BaP which gave a unit risk of 7.3×10^{-3} ($\mu g\,kg^{-1}$) day^{-1}. This involves an increased risk of 1×10^{-5} for the daily ingestion of about 100 ng BaP per person (1.4 ng BaP kg^{-1} body weight). This ingestion value is roughly in the order of magnitude of the actual BaP intake as estimated in different studies; the actual BaP intake, however, may exceed the value of 100 ng by up to fourfold.

The ingestion of food is by far the major source of PAH exposure for nonsmokers. It has been estimated as contributing up to about 90% of total BaP intake, excluding occupational and tobacco exposure. The remaining is attributed to air inhalation and (about or less than 1%) to drinking water. However, while there is a great deal of evidence which stresses the hazard of inhalation and skin contact as routes of PAH intake, it is less clear what the risk of ingestion is. (*See* **Smoked Foods**: Principles; Applications of Smoking.)

Limit Values of PAH Concentration in Foods

Owing to the potential carcinogenicity of foods contaminated by PAHs, some measures have been taken to limit the PAH content, in particular that due to food processing. They include legal limits associated with smoked foods and water, and recommended limits for refined oils and fats.

Smoked food Germany, Austria, and Poland set a limit of 1 $\mu g\,kg^{-1}$ for BaP in smoked meat and meat

products. The same limit was also set in Germany for smoked cheese and cheese products. In the European Union, the maximum BaP concentration permitted in foodstuffs as a result of using flavorings, including smoke flavorings, is 0.03 μg kg^{-1} (Council directive 88/388/CEE of 22-06-1988). In addition, another Council directive (91/493/CEE of 22-07-1991) concerning fishery products lays down some health conditions under which these products must be smoked, and lists wood materials which cannot be burned in smoking foods. (*See* **Vegetale Oils**: Types and Properties; Composition and Analysis.)

Waters The following limits were set in the European Union to PAH presence in water intended for human consumption (Council directive 98/83/EC of 3-11-1998): 0.010 μg l^{-1} for BaP, and 0.10 μg l^{-1} for the sum of benz[*e*]acephenanthrylene, benzo[*k*]fluoranthene, benzo[*ghi*]perylene, and indeno[1,2,3-*cd*]pyrene. BbFA, BkFA, BghiP, and IP.

Refined oils and fats Some European oil industries set their own guideline values for refined oils and fats: 5 μg kg^{-1} for the sum of seven higher-molecular PAHs (BaP, benzo[*e*]pyrene (BeP), BghiP, dibenz[*a,h*]anthracene (DBahA), perylene, anthanthrene, coronene), and 25 μg kg^{-1} for the sum of 13 PAHs, including the previous seven, plus another six lower-molecular PAHs. (*See* **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; Supercritical Fluid Chromatography; **Spectroscopy**: Visible Spectroscopy and Colorimetry.)

Analysis

The determination of PAHs in food is complex, time-consuming, and requires experienced personnel. Nowadays, it is accepted that it has to be performed under a program of quality control/quality assurance, with the aim of providing fully reliable results.

PAH Selection

There is no standard list of PAHs to be determined and the selection performed by each investigator is commonly arbitrary, based on the available instrumentation and reference standards. Almost invariably, BaP has been determined in every investigation, due to its well-known carcinogenicity. Nowadays, special attention is generally given to the determination of other carcinogenic PAHs, especially in health-related studies. Also when other PAH concentrations are available, BaP is usually taken as a reference compound to compare the contamination and the toxicological potency of different food items. This is supported by the fact that BaP is by far the species of higher toxicological concern, due to the combination of its carcinogenic potency – relatively to the other PAHs – with its concentration levels.

Sample Preparation

Most foods are not homogeneous and must be carefully homogenized prior to analysis. Extraction and clean-up are crucial steps in the determination because of the very low amounts of PAHs and the need to separate them from substances which may interfere during analysis, especially the lipids. The general scheme shown in Figure 1 has been widely used, as such or with some modifications which were later introduced in its application to a wide range of foods and to total diet samples. Modifications to the scheme of Figure 1 most commonly involve the extraction and partitioning solvents (e.g., dimethyl sulfoxide replacing *N*,*N*-dimethylformamide), the chromatographic sorbent (e.g., XAD-2 resin combined with other sorbents or florisil or alumina combined with silica gel), and thin-layer chromatography as the clean-up step. The sonication extraction (e.g., of plants and smoked foods) has also been used efficiently; its advantages are the reduced time of extraction and possibly superior recovery efficiencies and reproducibilities, but the results depend on matrix, solvent, and experimental conditions. Conventional chromatographic columns are often substituted by prepacked commercial cartridges (solid-phase extraction), which give advantages in terms of time and solvents consumed, and of reproducibility performance.

Recently, supercritical fluid extraction (SFE: e.g., of smoked and broiled fish, and toast) and accelerated solvent extraction (ASE: e.g., of smoked meat) have gained attention as rapid alternatives to conventional liquid extraction.

Sample Analysis

Identification and quantification are performed by capillary gas chromatography (GC) or by high-performance liquid chromatography (HPLC). In GC analysis, the most widely used stationary phases are the methylpolysiloxanes, especially SE-54 (5% phenyl, 1% vinyl-substituted) and SE-52 (5% phenyl-substituted) or equivalent phases. A flame ionization detector is almost universally employed: it has an excellent response linearity and, coupled with cold on-column injections, gives an accurate and precise quantification; because of its nonselectivity, however, samples need to be highly purified. Mass-spectrometric detectors are powerful tools in identifying and confirming compounds, and have gained wide acceptance. In HPLC analysis, the most used packed material consists of silica particles chemically bonded to

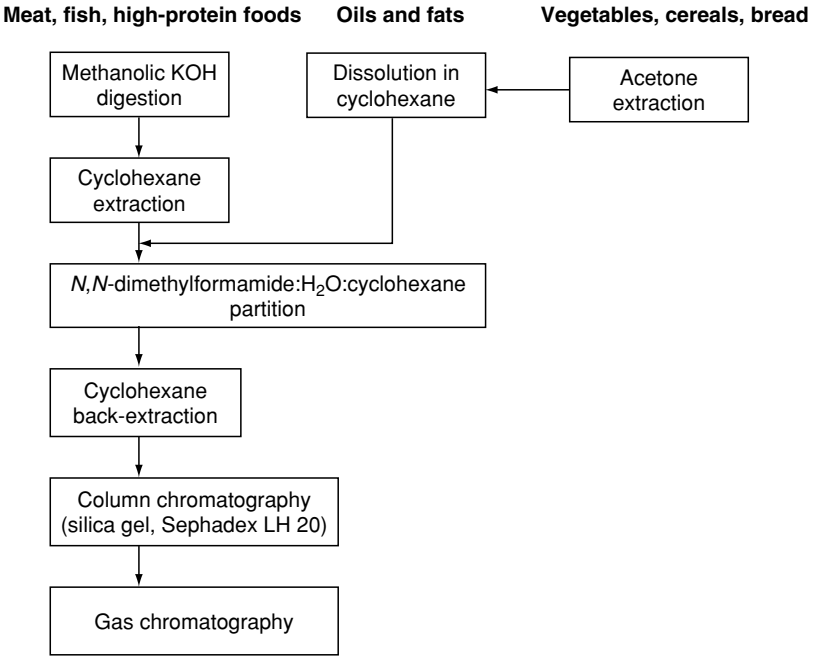

Figure 1 A classical general scheme for the determination of polycyclic aromatic hydrocarbons (PAHs) in food. Adapted from Grimmer G and Böhnke H (1979) Method 4 – Gas chromatographic profile analysis of polycyclic aromatic hydrocarbons in (I) high protein foods; (II) fats and vegetable oils; and (III) plants, soils and sewage sludge. In: Egan H, Castegnaro M, Kunte H, Bogovski P and Walker EA (eds) *Environmental Carcinogens: Selected Methods of Analysis*, vol. 3. *Analysis of Polycyclic Aromatic Hydrocarbons in Environmental Samples*, pp. 163–173. Lyon: International Agency for Research on Cancer.

linear C18 hydrocarbon chains. Ultraviolet (UV) and fluorescence detectors are used, usually arranged in series. The fluorescence detector is more sensitive, and its specificity allows for PAH determination in the presence of nonfluorescing interferences. For confirmation purposes, much information on the isomeric structure may be obtained from UV spectra acquired during the elution of chromatographic peaks by the UV diode-array detector.

Standard Methods

Standard methods for BaP in fats and oils have been prepared by International Standards Organization (ISO), American Oil Chemists' Society (AOCS), and IUPAC. General methods for PAHs in food have been published by IARC and Association of Official Analytical Chemists (AOAC).

Occurrence in Foods

Table 2 summarizes the levels of individual PAHs in various foods, as reported in the literature. The investigations for which the results are included in the table were selected on the basis of their representativeness of typical or (conversely) extreme situations, their recency (since the 1980s), and the wide spectrum of PAHs reported. While not exhaustive, the table aims to give general information on the contamination levels which may be found and to allow them to be compared in different foods. A number of investigations concerning only BaP determination are not included in the table but are discussed in the following sections.

Levels of individual PAHs generally range from below 1 $\mu g\,kg^{-1}$ to a few $\mu g\,kg^{-1}$. Occasionally, they are in the order of ten, and even hundreds of $\mu g\,kg^{-1}$.

The highest PAH levels were found in grilled food, smoked and broiled fish, mussels from polluted waters, and leafy vegetables grown in areas heavily exposed to air pollution. Shellfish such as crustaceans and bivalves do not metabolize PAHs appreciably and may accumulate high PAH amounts.

Food Smoking

There are considerable variations in the PAH contents of smoked foods, even within the same type of food. These are due to the variations in the smoke generation conditions, which include smoke production temperature, type and comminution of wood material, type of generators, time of smoking, and fat content of products. In general, the higher the smoking temperature, the more PAHs are formed.

High levels of individuals PAHs, and BaP in particular, were found in smoked fish (BaP, in the order of 10 $\mu g\,kg^{-1}$); this may be locally of major health concern for populations who consume large

Table 2 Concentrations (μg kg^{-1}) of selected polycyclic aromatic hydrocarbons (PAHs) in various foods

	n	*PHE*	*FA*	*PY*	*BaA*	*CHR*	*BbFA*	*BkFA*	*BaP*	*BeP*	*BghiP*	*IP*	*DBahA*
Meat and meat products													
Meat and meat products	4	3.0	0.5–1.1	0.5	0.02–0.5	0.1–0.6	0.04–1.0	0.01–0.2	0.01–0.6	0.03	0.03–0.6	0.01–0.7	0.01
Poultry and eggs[a]	1		0.9		0.5		0.3	0.2	0.1		0.2	0.2	
Eggs	1		0.1	ND	0.03	ND	0.4	0.01	0.01		ND		
Eggs, after Gulf War[b]	1	18(295)	1.2(3.9)	5.5(27)	4.5(7.8)	5.6(32)	3.5(13)	4.5(6.5)	7.5(19)		1.2(1.5)	8.7(20)	4.7(5.8)
Bacon	1		7.8				0.3	0.05	0.05		3.70	2.5	
Sausages	1				0.04–0.1				0.03–0.3		0.05–0.2	0.05–0.1	
Smoked meat	1				Tr-0.3[k]				0.01–0.1		Tr-0.1	Tr-0.1	
Smoked beef	1				0.02–0.6				0.02–0.4		0.03–0.3	0.04–0.4	
Smoked chicken	1	32	16	20	2.1	6.3	4.6	2.6	3.2		ND	1.7	ND
Smoked sausages	2				0.04–0.4				< 0.01–0.3		0.06–0.3	0.04–1.4	
Grilled sausages	2	3.5–618	1.1–376	1.2–452	0.2–144	0.3–140[k]	ND–92	ND–172[n]	ND–212	ND–81	ND–153	ND–171	ND–8.8
Grilled duck[c]	1	25	20	24	2.4	31	10	5.8	9.2		ND	5.2	ND
Fish and marine products													
Fish	4	3.5	0.8–1.8	0.8	Tr-7.5	0.6–2.9	0.1–2.0	0.04–0.7	Tr-4.5	0.1	Tr-0.9	ND–1.6	0.03
Fish, after an oil spill[d]	1	< 2–101	< 2–123	< 2–145	< 2	< 2			< 2–7.6				
Fish and shrimps, after Gulf War[e]	1	5.8–87	ND–33	ND–68	0.1–5.3	ND–16	0.1–5.8		ND–5.3		0.2–31	0.3–29	0.2–39
Fish, oil-contaminated sea[f]	1	13	1.5	1.7	0.3	1.7	0.4	0.4	0.8		0.3	0.1	0.1
Oysters	1	2.1–4.2	5.1–17	3.1–12	0.8–3	3.2–8.8[k]	4.5–12[m]	see BbFA	0.4–1.0	2.4–6.3	0.4–0.8	0.3–0.6	0.1–0.2[o]
Trout[g]	1	ND–0.07	0.1–0.4	0.8–2.9	0.2–1.5	0.07–0.3	ND		ND	ND		ND	ND
Smoked trout[g]	1	7.6–8.3	27–39	82–175	9.6–16	1.5–2.9	ND–0.08		5.1–8.4	ND		2.0–4.2	3.2–4.0
Smoked fish	3	5–330	1.4–210	1.3–68	ND–86	ND–50	ND–3.9	ND–6.7[n]	ND–40	ND–2.8	ND–25	ND–37	
Smoked fish, traditional kilns[h]	1	65	26	20	2.5	2.5	1.2	0.5	1.2		0.7	1.1	< 0.1
Smoked fish, modern kilns[h]	1	32	9.1	5.3	0.6	0.6	0.1	0.07	0.1		0.03		ND
Smoked or canned oysters and mussels	1	1.9–20	4.5–19	2.6–11	0.8–21	3.9–31[k]	1.2–24[m]	see BbFA	0.2–12	0.7–7.6	0.3–5.7	0.2–6.4	< 0.1–0.5[o]
Broiled fish	1	320	1080	390		390			400				
Vegetables													
Kale	1		117	70	15	62			4.2	7.9	7.7	7.9	1.0
Lettuce	4	0.5–12	0.09–28	0.4–18	0.05–4.6	0.8–4.0	0.1–7.3	ND–17	0.007–6.2	0.07–6.7	0.02–11	0.1–8	
Lettuce, industrial area	1		28				6.1	3.7	5.6		10	2.4	
Potatoes[a]	1		ND		0.4	0.8	0.2	0.1	ND		0.1	ND	
Potato products[a]	1	3.0	0.6		0.1	ND	0.6	0.3	0.3		0.9	0.6	
Tomatoes	2		0.09	0.4	0.006–0.3	0.1–0.5	0.008	0.003	0.003–0.2	0.2	ND		0.04
Soups[a]	1		ND		ND	0.4	0.05	0.08	ND		ND	ND	
Fruits and confectionery													
Fresh fruit[a]	1	7.8	3.6		0.5	0.5	0.1	0.1	ND		ND	ND	
Apples	1		2.4	3.5	0.3	ND	0.3	0.08	0.5		0.02		
Canned fruit and juices[a]	1	ND	1.0		ND	ND	0.1	0.1	0.1		0.1	ND	
Nuts	1	10(17)	1.3(3.0)		0.1(4.2)	4.0(69)	0.2(0.4)	ND(0.1)	ND(0.2)		ND(0.4)	ND(0.4)	
Biscuits, pudding, and cakes	2	2.9	0.5–3.6	0.6–2.4	0.08–2.7	0.09–2.8	0.03–1.3	0.04–1.4	0.04–2.2	0.08–2.9	0.1–2.5	0.1–3.2	
Sugar and sweets	1	ND(3.2)	0.7(2.3)		0.2(4.2)	0.7(36)	0.2(3.5)	0.1(0.5)	0.1(0.4)		ND(0.2)	ND(0.2)	
Cereals													
Breakfast cereal	1		0.2–0.6	0.3–1.2	0.06–0.2		0.02–0.05	0.02–0.07	0.03–0.05	0.06–0.2/	0.06–0.08	0.08–0.15	< 0.01

Bread	3	3.3	0.2–2.8	ND–0.9	0.1–0.8	0.1–1.9	0.04–1.2	0.02–0.06	0.02–0.8	0.06–0.1	0.01–0.5	0.11–0.6	< 0.01–0.01
Pasta, macaroni, and rice	2	2.1	2.5–3.9	ND	0.03–0.4	1.3–1.9	0.04–1.0	0.02–0.5	0.017–0.8		ND–0.6	0.5	
Noodles and pizza	1	4.2	3.7		0.5	2.0	0.5	0.1	0.2		0.5	0.3	
Pizza in a wood-burning oven	1		0.2	0.6	9.1	3.0	0.04	0.02	0.02		0.06		
Oils and fats													
Olive oils	3	0.4–38	Tr-15	ND–14	0.03–0.9	0.4–1.5	0.3–1.4	0.06–0.5	0.1–1.3	< 1	ND–1.9	0.3–2.0	ND–0.1
Extra virgin olive oils	1	1.7–53	Tr-53	0.4–28	Tr-10	Tr-2.3	Tr-1.3	Tr-0.5	Tr-1.2		Tr-0.7	Tr-0.6	Tr-0.3
Various unrefined oils[i]	1	ND–69	0.2–18	0.1–14	ND–6.1	0.1–8.6[k]	ND–8.9[m]	see BbFA	ND–4.1	ND–3.8	ND–4.2	ND–4.3	ND–0.2[o]
Coconut oil, deodorized	1	8.7	23	29	5.7[l]	see BaA	1.0[m]	see BbFA	0.2	0.4	< 0.1		ND
Corn oil	1	0.1	ND	ND	ND	ND	0.09[m]	see BbFA	0.02	0.04	0.1	ND	0.03[o]
Butter	1		0.1	1.8	0.6	ND	0.02	0.03	0.02		ND		
Margarine	3	< 0.2–6.0	0.09–9.0	< 0.1–15	< 0.1–5.2	< 0.2–7.5	< 0.2–9.2	< 0.1–11	0.05–6.0	0.09–6.1	0.02–11	0.03–9.7	0.05–9.2
Beverages													
Wine	2		0.08	ND	0.003	ND	Tr-0.04	Tr-0.01	Tr- 0.009		ND–0.03	ND	ND
Beer	2		0.04	ND	0.1	ND	Tr-0.2	ND–0.1	ND–0.07		ND–0.2	ND–0.1	ND–Tr
Coffee	1		1.2	ND	0.1	ND	0.08	0.02	0.01		0.01		
Tea infusion[j]	1	0.2–0.7	0.007–0.03	0.02–0.2	0.003–0.043				0.002–0.02	0.007–0.056	0.007–0.03		0.005–0.02
Milk	2	ND	0.1–0.2	0.04–0.2	0.01	ND–1.5	0.01–0.06	0.003–0.03	0.01–0.02	ND	0.01–0.03	ND	ND
Milk, after Gulf War[b]	1	3.0(9.8)	3.4(12)	35(145)	2.4(3.7)	8.6(32)	3.1(4.4)	ND	1.5(1.5)		ND	ND	ND

A large part of data was adapted from International Programme on Chemical Safety (1998). *Selected Non-heterocyclic Polycyclic Aromatic Hydrocarbons. Environmental Health Criteria 202*. Geneva: World Health Organization. Unless otherwise specified, an individual value is the mean concentration (if $n = 1$) or the only available mean concentration (if $n > 1$). Concentrations in parentheses are maximum values.
n, number of investigations; ND, not detected; Tr, traces; for PAH abbreviations, see **Table 1**.

[a]Maximum concentrations (mean concentrations were mostly ND).
[b]Kuwait, from locally reared animals.
[c]Grilling time: 0.5 h.
[d]Arabian Gulf.
[e]Kuwait.
[f]Red Sea coast; oil operations and heavy ship traffic.
[g]'Trout' and 'smoked trout' were investigated within the same study.
[h]'Traditional kilns' (27 samples, three kilns) and 'modern kilns' (35 samples, five kilns) were investigated within the same study; mean concentrations are reported.
[i]Olive, safflower, sunflower, maize germ, sesame, linseed, wheat germ oils.
[j]Concentrations in $\mu g\ l^{-1}$.
[k]In sum with triphenylene.
[l]In sum with CHR and triphenylene.
[m]In sum with BjFA and BkFA.
[n]In sum with BjFA.
[o]In sum with dibenz[*a,c*]anthracene.

quantities of these products. Considerable amounts of PAHs were found in home-smoked meat in Iceland, with most of BaP detected in the superficial layers of the meat, and in smoked cheese too. The highest BaP concentrations were found in samples of smoked cereals (up to 160 $\mu g\,kg^{-1}$) and smoked teas (up to 110 $\mu g\,kg^{-1}$).

Traditional smoking techniques, in which smoke from incomplete wood burning comes into direct contact with the product, can lead to extensive contamination with PAHs. Hence, smoke flavorings are used as an alternative. Current commercial smoking practices appear to be effective in controlling the deposition of PAHs. In a recent Spanish survey of commercial smoked products, BaP was found at low levels in almost all samples: up to 0.3 $\mu g\,kg^{-1}$ in meat products and up to 0.9 $\mu g\,kg^{-1}$ in cheese and in fish, with the exception of 2.5 $\mu g\,kg^{-1}$ in a sample of smoked sardine. The oil associated with the fish, if any, may be more contaminated than the fish itself, possibly due to the PAH leaching from the fish.

In a German survey conducted in the 1990s, PAH concentrations were compared between 27 smoked fishery products from traditional smoking kilns and 35 from modern smoking kilns with external smoke generation. The BaP levels from the modern kilns were one order of magnitude lower (on average, 0.1 $\mu g\,kg^{-1}$) than by the traditional systems (1.2 $\mu g\,kg^{-1}$).

Meat and Meat Products

The amount of PAHs formed during cooking depends markedly on the method of cooking and the conditions. In particular, PAH formation during charcoal grilling is dependent upon the fat content of the meat, the time of cooking, and the temperature.

This formation may be due to various causes: the incomplete combustion of charcoal, the transformation of some food components such as triglyceride and cholesterol, or – the most likely source of high PAH levels – the melted fat of the meat. Actually, during charcoal grilling at high temperatures, the fat drippings fall on the hot coals where they are pyrolyzed, producing PAHs which are then volatilized and deposited onto the meat surface.

An investigation of the effect of the method, including broiling on electric and gas heat, charcoal broiling, and broiling over charcoal with a no-drip pan, showed that PAH formation may be minimized by avoiding direct contact of the food with the cooking flame, cooking meat at lower temperatures for longer periods, and using meat with a low fat content.

The total concentration of six carcinogenic PAHs in a lamb sausage increased from 1.9 $\mu g\,kg^{-1}$ when grilled over charcoal under standard barbecue practices (distance between fire and meat 10 cm; temperature at the surface of the meat 200–250 °C) to 13.2 $\mu g\,kg^{-1}$ when heavily grilled for a prolonged time.

An experiment on PAH formation in duck meat during various processing methods showed the highest concentrations of carcinogenic PAHs were due to smoking (1 m distance between smoke generation source and meat), followed by charcoal grilling (1 m distance between charcoal and meat) and roasting (at 200 °C); no carcinogenic PAHs were detected in samples subjected to steaming or to flavoring by liquid smoke. The BaP concentration during smoking increased from 6.9 $\mu g\,kg^{-1}$ after 0.5 h to 13.9 $\mu g\,kg^{-1}$ after 3 h.

Fishes and Marine Products

Processing and cooking may increase PAH levels by the same routes described above for meats. The highest levels of BaP and other PAHs were generally found in smoked fishes, and especially in the smoked skin. In a study of commercial Baltic herrings, BaP was found at about 40 $\mu g\,kg^{-1}$ in a smoked specimen, but at a level one order of magnitude higher in a broiled one (about 400 $\mu g\,kg^{-1}$, which is the highest level of contamination reported). High concentrations are also found in marine products from contaminated sea.

Vegetables

The differences in PAH content may be due to variations in the ratio between the surface area and the weight, to location (rural or industrialized) or to growing season.

It is recognized that broad-leaved vegetables may have particularly high levels of PAHs, due to the deposition of airborne particulate matter. The importance of atmospheric pollution was shown by the high levels detected in lettuce grown close to a highway, with levels decreasing with distance from the road. The PAH profiles in lettuce were found to be similar to those in ambient air, confirming the deposition of airborne particles as the main source of contamination. In an investigation near a coking plant, PAH levels in vegetables with large, rough leaf surfaces (spinach and lettuce) were found to be 10 times higher than in other vegetables (carrots and beans), and this was likely due to deposition from ambient air.

The effect of washing vegetables on reducing PAH contamination due to vehicle exhaust appears controversial: in an investigation on kale, PAH content was not reduced other than at levels lower than about 10% (BaP, 10%). Conversely, in an experiment with lettuce, the higher-molecular PAHs were

considerably reduced (BaP, 67–95%). In no study, however, were lower-molecular PAHs significantly affected by washing.

In an experimental comparison of growth of terrestrial plants in a 'clean-air' chamber and in the open field, the contamination was shown to be due almost exclusively to airborne PAHs and not to synthesis by plants.

Cereals and Dried Foods

Growing crops (wheat, corn, oat, barley) in industrialized areas increases PAH levels in comparison to more remote areas. Grain samples from a heavily industrialized area contained 10 times more PAHs than samples from areas remote from industry. The growth of rye near a high-traffic highway resulted in PAH contamination, which decreased slightly 7–25 m away from the road.

Drying by combustion gases was also found to increase contamination by three- to 10-fold. In an experiment with wheat, drying over a light fuel oil flame increased the BaP deposition from six- to 130-fold depending on the degree of exposure.

Oils

The presence of PAHs in vegetable oils is due to contamination from technological processes (namely, smoke drying of oil seeds or contaminated extraction solvents) or to environmental contamination (e.g., traffic exhaust or industrial emissions).

BaP in olive oils is commonly within the 1 μg kg^{-1} level; higher levels may be present if the oils are from plants exposed to industrial emissions or if they are blended with previously contaminated vegetable oils.

The refining of olive oils (especially the deodorization step) has a marked reducing effect on the light PAHs but not on the heavy ones. The latter may be reduced by treatment with activated charcoal.

High PAH levels may be found in oil seeds. They are usually due to the process of direct drying the seeds, using wood or oil as a fuel. BaP at levels up to 25 μg kg^{-1} were found in corn oils. The PAH content of oil seeds is drastically reduced during refining, particularly by treatment with activated charcoal. This refining method is now widely used.

Dietary PAH Intake

Table 3 shows the dietary intake of PAHs, as estimated from different studies in three European countries. The intakes were calculated on the basis of food consumption surveys, except in one Dutch study where portions of a 24-h intake of food were provided by the participants and analyzed. There are numerous differences between the diets of northern and southern Europe and the criteria adopted in

Table 3 Daily intake of selected polycyclic aromatic hydrocarbons (PAHs) in total diet, as estimated in various countries (mean and, in parentheses, where available, maximum values)

PAH	Intake (μg person^{-1})			
	UK[a]	The Netherlands[b,c]	Italy[c,d]	Range of mean values
PHE	NG	2.69 (5.13)	NG	
FA	0.99	1.32 (2.11)–2.7 (10.4)	4.22	1.0–4.2
PY	1.09	1.6 (5.1)	1.73	1.1–1.7
BaA	0.22	0.16 (0.48)–0.28 (0.65)	0.41–1.29	0.2–1.3
CHR	0.5	1.19 (3.90)–1.2 (5.0)	1.46 (1.70)–2.20	0.5–2.2
BbFA	0.18	0.33 (0.59)	0.65	0.2–0.6
BkFA	0.06	0.12 (0.24)	0.17	0.06–0.2
B[*b*+*j*+*k*]FA	> 0.24	> 0.45	1.10	> 0.2–1.1
BaP	0.25	0.08 (0.35)–0.20 (0.42)	0.17 (0.32)–0.37	0.08–0.4
BeP	0.17	0.14 (1.2)	NG	0.1–0.2
BghiP	0.21	0.16 (0.58)–0.28 (1.03)	0.02	0.02–0.3
IP	ND	0.16 (1.2)–0.27 (0.55)	0.16 (0.20)	ND–0.3
DBahA	0.03	0.04 (0.53)	0.08 (0.17)	0.03–0.1

ND, not detected in any food group; NG, not given; PAH abbreviations: see Table 1. B[*b*+*j*+*k*]FA: sum of BbFA, BjFA and BkFA; BjFA was not determined in any study.

[a]From Dennis MJ, Massey RC, McWeeny DJ, Knowles ME and Watson D (1983) Analysis of polycyclic aromatic hydrocarbons in UK total diets. *Food and Chemical Toxicology* 21: 569–574.

[b]From De Vos RH, Van Dokkum W, Schouten A and De Jong-Berkhout P (1990) Polycyclic aromatic hydrocarbons in Dutch total diet samples (1984–1986) *Food and Chemical Toxicology* 28: 263–268; Vaessen HAMG, Schuller PL, Jekel AA and Wilbers AAMM (1984) Polycyclic aromatic hydrocarbons in selected foods: Analysis and occurrence. *Toxicological and Environmental Chemistry* 7: 297–324.

[c]When a range is reported, it refers to two different studies.

[d]From Turrio-Baldassarri L, Di Domenico A, La Rocca C, Iacovella N and Rodriguez F (1996) Polycyclic aromatic hydrocarbons in Italian national and regional diets. *Polycyclic Aromatic Compounds* 10: 343–349; Lodovici M, Dolara P, Casalini C, Ciappellano S and Testolin G (1995) Polycyclic aromatic hydrocarbon contamination in the Italian diet. *Food Additives and Contaminants* 12: 703–713.

the surveys are not homogeneous. Consequently, the intakes estimated in different studies cannot be directly compared and have to be considered as a rough estimate of the actual intakes. Notwithstanding these limitations, it is worth noticing that they are generally of the same order of magnitude and the differences are relatively small compared with the variation in cancer potency estimates.

The Contribution of the Different Food Groups

The contribution of different food groups to the total dietary intake of PAHs was estimated in the UK (Table 4).

The major contributors were the oil-and-fats group and the cereals group. The former has the highest individual PAH levels, but the latter, although never showing high individual PAH levels, is a major contributor due to its relative weight in the total diet. Although several studies indicated quite large amounts of PAHs in smoked meat and smoked fish, these products made a very small contribution to the pertinent food groups and these in turn were not major components of the diet. Consequently, at least in areas where eating barbecued food is an infrequent activity, this provides a very small part of the dietary intake of PAHs. The third major contributor was vegetables, and this was likely due to the atmospheric fallout of particle-bound PAHs.

These results were substantially confirmed in a subsequent study of the Dutch diet: cereal products were found to contribute most to the daily intake of PAHs (again attributed to the high consumption share), followed by sugar-and-sweets and by oils and fats. It was not possible to explain the surprisingly large share of the sugar-and-sweets group because none of the constituents (sugar, chocolate products, jellies, licorice) was suspicious with regard to high PAH levels. The relatively high contribution of the oils-and-fats group was at least partly attributed to the well-known elevated PAH concentrations possibly present in vegetable oils. Nuts, which were included in this group, contributed 6% to the total PAH intake, and this was most likely due to the roasted peanuts.

Cereals were also found to be the highest contributors in a Swedish study (where they were followed by vegetables, and by oils and fats), and in an Italian one.

A more accurate comparison of the results of all these studies is not possible due to the differences in the reporting criteria adopted.

Table 4 Contribution from individual food groups to the total UK dietary intake of polycyclic aromatic hydrocarbons (PAHs)

	% of total dietary intake	
	BaP	*Sum of six PAHs*[a]
Cereals	30	35
Meat	3	4
Fish	1	2
Oils and fats	50	34
Fruit, sugar and sweets	5	7
Vegetables	8	18
Beverages	0	0
Milk	2	1

Modified from Dennis MJ, Massey RC, McWeeny DJ, Knowles ME and Watson D (1983) Analysis of polycyclic aromatic hydrocarbons in UK total diets. *Food and Chemical Toxicology* 21: 569–574.

[a]CHR, BaA, BeP, BaP, DBahA and IP: for abbreviations, see Table 1.

See also: **Analysis of Food**; **Cancer**: Carcinogens in the Food Chain; **Carcinogens**: Carcinogenicity Tests; **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **Dietary Surveys**: Surveys of National Food Intake; **Flavor (Flavour) Compounds**: Production Methods; **Legislation**: International Standards; **Smoked Foods**: Applications of Smoking; Production; **Smoking, Diet, and Health**

Further Reading

Dennis MJ, Massey RC, McWeeny DJ, Knowles ME and Watson D (1983) Analysis of polycyclic aromatic hydrocarbons in UK total diets. *Food and Chemical Toxicology* 21: 569–574.

De Vos RH, Van Dokkum W, Schouten A and De Jong-Berkhout P (1990) Polycyclic aromatic hydrocarbons in Dutch total diet samples (1984–1986). *Food and Chemical Toxicology* 28: 263–268.

Fazio T and Howard JW (1983) Polycyclic aromatic hydrocarbons in foods. In: Bjørseth A (ed.) *Handbook of Polycyclic Aromatic Hydrocarbons*, pp. 461–505. New York: Marcel Dekker.

Grimmer G and Böhnke H (1979) Method 4 – Gas chromatographic profile analysis of polycyclic aromatic hydrocarbons in (I) high protein foods, (II) fats and vegetable oils and (III) plants, soils and sewage sludge. In: Egan H, Castegnaro M, Kunte H, Bogovski P and Walker EA (eds) *Environmental Carcinogens Selected Methods of Analysis*, vol. 3, pp. 163–173. *Analysis of Polycyclic Aromatic Hydrocarbons in Environmental Samples*. Lyon: International Agency for Research on Cancer.

Guillén MD (1994) Polycyclic aromatic compounds: extraction and determination in food. *Food Additives and Contaminants* 11: 669–684.

IARC (1983) *Polynuclear Aromatic Compounds, Part 1, Chemical, Environmental and Experimental Data. IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans*, vol. 32. Lyon: International Agency for Research on Cancer.

IARC (1987) *Overall Evaluations of Carcinogenicity: An Updating of IARC Monographs* vols. 1–42. *IARC Monographs on the Evaluation of Carcinogenic Risks*

to Humans, Supplement 7. Lyon: International Agency for Research on Cancer.
International Programme on Chemical Safety (1998) *Selected Non-heterocyclic Polycyclic Aromatic Hydrocarbons. Environmental Health Criteria 202*. Geneva: World Health Organization.
Lijinsky W (1991) The formation and occurrence of polynuclear aromatic hydrocarbons associated with food. *Mutation Research* 259: 251–261.
Lodovici M, Dolara P, Casalini C, Ciappellano S and Testolin G (1995) Polycyclic aromatic hydrocarbon contamination in the Italian diet. *Food Additives and Contaminant* 12: 703–713.
Phillips DH (1999) Polycyclic aromatic hydrocarbons in the diet. *Mutation Research* 443: 139–147.
Thomson B and Muller P (1998) Approaches to the estimation of cancer risk from ingested PAH. *Polycyclic Aromatic Compounds* 12: 249–260.
Turrio-Baldassarri L, Di Domenico A, La Rocca C, Iacovella N and Rodriguez F (1996) Polycyclic aromatic hydrocarbons in Italian national and regional diets. *Polycyclic Aromatic Compounds* 10: 343–349.
Vaessen HAMG, Schuller PL, Jekel AA and Wilbers AAMM (1984) Polycyclic aromatic hydrocarbons in selected foods: Analysis and occurrence. *Toxicological and Environmental Chemistry* 7: 297–324.

Polyphenols *See* **Tannins and Polyphenols**

Polysaccharides *See* **Carbohydrates**: Classification and Properties; Interactions with Other Food Components; Digestion, Absorption, and Metabolism; Requirements and Dietary Importance; Metabolism of Sugars; Determination; Sensory Properties

PORK

R W Mandigo, University of Nebraska, Lincoln, NE, USA

Development of the Swine Industry

The earliest known records of swine domestication are from China and date to 4900 BC. Christopher Columbus brought the first pigs to the USA via the Canary Islands in 1493. The early colonists of the USA brought livestock with them from England throughout the 1600s. Gradually, as more grain was grown, larger herds of pigs developed and swine production became a true industry.

By the early 1800s, swine-slaughtering plants had emerged at several large population centers in the USA. The largest of these was at Cincinnati (also known as Porkopolis at that time) because of its prime location on the Ohio river. Early slaughtering plants were located near rivers for four reasons: rivers provided ice to keep the meat cold, transportation for products to be shipped, a source of water, and a place to deposit plant waste materials such as blood.

The advent of mechanical refrigeration led to industry expansion as pork could be processed year-round and kept fresh longer. Development of rail systems and the use of refrigerated rail carriages boosted industry growth as both livestock and meat could be more widely distributed. In the 1800s, pigs were often allowed to roam free in pastures and were fed garbage or what little grain was available. Today, most pigs are raised in large numbers in environmentally controlled buildings with a very specific diet designed to maximize growth.

The first slaughter operations were separate from packing operations. By the 1850s, the two had started to become integrated. Today, many plants have reverted back to specializing in either slaughter and fabrication or pork processing. This specialization has allowed modern-day slaughtering facilities to process around 1000 head per hour.

The pigs in the USA today are thought to have descended from two wild stocks, the European wild boar (*Sus scrofa*) and the East Indian pig (*Sus vittatus*). From these two genetic bases, modern pigs have been refined into several breeds which each have distinguishing characteristics. In addition to hair color and marking differences, breeds have been developed for specific purposes such as prolificacy, bacon production, ham production, and others.

In general, pigs have become leaner as the demand for lard has decreased and the demand for leaner meats has increased. The modern market pig is approximately 71.5% carcass while the remaining 28.5% is mostly byproducts used in the medical and animal feed industries.

Slaughter, Fabrication, and Processing

Procurement

Although pigs of all sizes can be slaughtered, most swine-slaughtering plants have equipment designed to handle pigs marketed at about 105 kg. Swine producers can sell their pigs to slaughter plants in various ways. Pigs can be marketed at a central location such as an auction or a packer buying station or they can be purchased from the farm by a packer buyer. The latter method is the most popular today. A new way to market pigs that has some popularity is contract producing for a specific packer.

After pigs arrive at the slaughtering plant, they are slaughtered within 2–3 h of arrival. If a substantial delay is expected, then they will usually be held in pens for 12–24 h. It is important to allow pigs time to become acquainted with their environment as high-stress conditions can cause undesirable changes in muscle color and juiciness. In pigs with a certain genetic make-up, severe stress can cause the meat to be pale and watery (known as pale, soft, and exudative, PSE). Long-term stress, on the other hand, can lead to a condition where the muscle is dark and very dry-appearing (known as dark, firm, and dry, DFD). The physiological basis for these conditions is an abnormal muscle pH change between the time of death and the completion of rigor mortis. Therefore, it is necessary to minimize stress by proper handling of pigs prior to slaughter. Feed is usually kept from the animals during this period to decrease stomach fill. This improves percentage of carcass dressing, aids evisceration during slaughter, helps minimize contamination, and improves color. (*See* **Meat**: Slaughter.)

Slaughter

All meat sold interstate in the USA must be inspected by the US Department of Agriculture, Food Safety and Inspection Service (USDA-FSIS). Inspection begins *antemortem* and continues throughout the entire slaughter, fabrication, and processing system. This helps ensure the wholesomeness of meat products offered to the consumer. If the meat will be sold intrastate only, some states allow inspection by state-employed inspectors. This inspection, by law, must be equal to federal inspection. Similar inspection regimes apply in Europe and other developed regions of the world.

At the time of slaughter, pigs are herded into a chute where they are stunned to render them unconscious. Stunning can be accomplished by one of several methods, but is mandatory to comply with the US Humane Slaughter Act of 1958. If stunning is not performed correctly, the incidence of poor-meat-quality problems could increase. An example of this is a condition referred to as blood splashing. Blood splashing occurs when small capillaries in muscle tissue burst during stunning, leaving spots of blood in the muscle. The most common method of stunning is accomplished using an electrical current. Other methods include a blow from a captive bolt to the head, moving pigs through a carbon dioxide gas chamber to cause suffocation, or cardiac arrest stunning. The goal of stunning is to render the animals unconscious while allowing the heart to continue to beat and pump blood out of the body.

After stunning, the pig is immediately shackled, hoisted by the hind leg, and the jugular vein is severed with a knife to allow blood to exit the body. Bleeding will remove around 50% of the blood in the body. Much of the remaining blood is removed with the organs during evisceration.

Pigs can either be skinned or they can be scalded and dehaired. To dehair carcasses, the pig is put in a scalding vat of hot water (60 °C) with lye to loosen the hair follicles. The pig is removed from the scalding tank and dehaired mechanically and/or by hand using blades to scrape away the hair. Whole skins from skinned pigs can be saved and made into pigskin for use in garments and for gelatin production.

After dehairing or skinning, the abdominal and thoracic cavities are opened and the internal organs removed. Certain organs are washed, inspected, and saved as edible byproducts. These include the liver, heart, tongue, and brain. Other organs may also be saved as inedible byproducts, and will be discussed later in this article.

The carcass is washed, weighed, and placed in a cooler for chilling. The cooler temperature is maintained at about −2 to 5 °C so that internal areas of the carcass will be adequately chilled in 24 h. While in the cooler, most of the remaining stored energy in the muscle will be depleted and the carcass will undergo

rigor mortis. Some carcasses are processed before they go through rigor mortis. This is called hot boning. Although fabrication is more difficult, the cuts can be chilled faster than whole carcasses, so product that has been hot-boned can be moved through the system faster.

Fabrication

Pig carcasses are fabricated into wholesale cuts, using knives, saws, and other mechanical equipment. The wholesale cuts include the ham, loin, Boston shoulder, picnic shoulder, belly, and spare ribs. The wholesale cuts may be shipped fresh to further processors, food service operations, or to the retail market. Some plants will continue to fabricate the wholesale cuts into retail cuts and trimmings and sell these items. Many plants that slaughter will carry out further processing such as curing and smoking bellies and hams in the same plant. (*See* **Curing**.)

Processing

Further processing of pork can involve many different processes. Processing encompasses many steps, including particle size reduction, the addition of nonmeat ingredients, cooking, smoking, and a variety of packaging procedures. Patties can be made by grinding pork trimmings, with a specified fat level, with salt and other seasonings. Restructured products often have salt and phosphates added to flaked or chunked meat. These ingredients are then mixed and formed into steaks, chops, or roasts. Sausage is another type of further processed meat. Sausages can be cured, smoked, dried, or cooked. Some sausages are coarsely ground while others are finely chopped, forming what is sometimes called an emulsion or batter. Bratwurst, frankfurters, and bologna are sausages that typically contain pork. (*See* **Meat**: Sausages and Comminuted Products.)

The hams and bellies from pork carcasses are usually cured and smoked. The curing may be done by packing in a dry cure or, more commonly, by pumping or injecting a solution containing the cured ingredients into the meat. The curing solution may contain any or all of the following ingredients: water, salt, sugar, phosphate, nitrite, erythorbate, or spices and seasonings. Water is used to add moisture to the product as well as acting as the solvent in a curing solution. Salt solubilizes proteins in meat so the meat particles bind together better. Salt also adds flavor and, at high levels, retards microbial spoilage. Sugar can be used for flavor, to promote browning, and to act as a food source for desirable bacteria in some fermented meat products. Phosphates are added to help retain water in the product and to prevent oxidative rancidity from occurring. Nitrite is the curing agent. Nitrite binds to meat pigments and causes the meat to develop a stable pink color. It also lends flavor, acts as an antimicrobial agent, and helps to prevent oxidative rancidity. Erythorbate is used to speed up the curing process, which is economically important to pork processors. Spices and seasonings are added to develop flavors unique to a particular product. (*See* **Smoked Foods**: Principles.)

Some sausages are cured and may also contain acidulants, starter cultures, or extenders. Acidulants and starter cultures of bacteria can be used to lower the pH of products in order to develop certain flavors and protect the product from spoilage. Nonmeat extenders, such as soya proteins or nonfat dry milk, may be used as inexpensive protein sources.

The US Department of Agriculture (USDA) has regulations regarding the composition of processed meat products and the proper use of some ingredients used in processing. Hams must be labeled according to the amount of protein they contain on a fat-free basis. The amount of phosphates, nitrites, erythorbate, and nonmeat ingredients added are also limited. These regulations, and similar ones in other countries, are designed to provide consumers with high-quality, safe meat products.

Pork is also regulated by the USDA in an attempt to control the disease trichinosis. Trichinosis is caused by the organism *Trichinella spiralis* and can be transmitted from pig to humans by ingestion of infected muscle tissue. Simply put, the regulation states that pork for use in products that are not fully cooked in the home must be certified as 'trichina-free.' Pork may become certified trichina-free by being subjected to specific heat or freezing to kill the organism. (*See* **Zoonoses**.)

Chemical Composition

Fresh pork muscle is 70–75% water, and the protein content ranges from 18 to 22%. There are three main types of protein in the pig: myofibrillar (skeletal), stromal (connective tissue), and sarcoplasmic (pigments). Lipid or fat is another major constituent of fresh pork. It makes up between 5 and 7% of the muscle tissue. Lipids include phospholipids and triacylglycerols (also known as triglycerides). The carbohydrate content of meat is negligible, generally less than 1%. Vitamin and mineral content of fresh pork is usually about 1–2%. (*See* **Fats**: Classification; **Meat**: Structure; Nutritional Value; **Protein**: Chemistry.)

Different cuts in the carcass will have different compositions. This is largely due to the varying fat

Table 1 Composition of various raw pork cuts ($g\,kg^{-1}$ raw meat)

	Moisture	*Fat*	*Protein*	*Ash*
Ham slice	730	55	205	10
Bacon	370	530	95	5
Loin chop	710	70	210	10
Spare ribs	575	245	170	10

level in different areas of the carcass. Table 1 contains some example cuts and their typical raw composition.

Nutrient Value and Dietary Significance of Pork

Pork supplies many nutrients essential for maintenance and growth.

As with other meat, pork is an excellent source of protein. A single 85-g serving of pork contributes 41% of the daily protein requirement for a normal adult male. Not only does pork contain a large amount of protein, this protein is of good quality. (*See* **Protein**: Requirements.)

Pork also contains lipids and fats. About 34% of pork fatty acids are saturated and 66% are unsaturated. Cholesterol is another lipid found in pork. Cholesterol is found in cell membranes in the animal body and is synthesized in the liver of humans and animals. Consumption of animal products, therefore, provides a dietary source of cholesterol which can be used in the body. Cholesterol, like saturated fats, has been associated with increased risk of developing heart disease. The relationship is not well understood, but the American Heart Association recommends keeping dietary intake of cholesterol to less than $300\,mg\,day^{-1}$. One 85-g serving of pork provides about 79 mg of cholesterol or about 26% of the recommended 300 mg. (*See* **Cholesterol**: Role of Cholesterol in Heart Disease; **Fatty Acids**: Properties.)

Pork is an excellent food source for several vitamins and minerals. It supplies large amounts of thiamin, vitamin B_{12}, niacin, riboflavin, and zinc. Pork is also a good source of vitamin B_6, phosphorus, and iron. Dietary iron can be classified into two types, heme and nonheme. Heme iron, which is the major type found in pork, is absorbed more easily and better utilized by the body. Iron is a component of the molecule hemoglobin, which is the major carrier of oxygen in our blood stream. Intake of heme iron is especially important in warding off anemia, which may result from a low level of hemoglobin in the blood. Refer to individual vitamins and minerals.

Pork, when consumed in moderation, is an excellent source of many important dietary nutrients.

Microbiological and Other Hazards

Muscle is essentially sterile prior to death. However, meat destined for human consumption is cross-contaminated with microorganisms by equipment and handling at the time of slaughter and processing. Just as pork is an excellent source of nutrients for our bodies, muscle or meat is also an excellent growth medium for microorganisms. Controlling the growth of microorganisms on pork by acidifying, curing, salting, modified-atmosphere packaging, drying, cooking, or refrigerating is essential. (*See* **Meat**: Preservation.)

Food poisoning can result from consuming pork that has been mishandled, allowing certain microorganisms to grow. Causative organisms of food poisoning may include *Staphylococcus aureus, Bacillus cereus, Salmonella* spp., *Listeria monocytogenes, Yersinia enterocolitica, Clostridium botulinum, C. perfringens*, and *Campylobacter fetus* ssp. *jejuni*. Food poisoning is quite common and, in its mildest forms, is often mistaken for influenza because the symptoms are very similar. Generally, foodborne illnesses are relatively short-lived and more uncomfortable than harmful. However, food poisoning can be a very serious matter. It can be debilitating or even fatal for those with poor immunological defenses such as infants or the elderly. Like other microorganisms, pathogens which cause food poisoning are well controlled by heat, refrigeration, chemicals, or other means mentioned earlier. However, undercooking, improper cooling, or recontamination of cooked food by raw food are common ways that pathogens appear in the food supply. (*See* ***Bacillus***: Food Poisoning; ***Campylobacter***: Properties and Occurrence ***Clostridium***: Food Poisoning by *Clostridium perfringens*; Botulism; ***Listeria***: Properties and Occurrence; ***Staphylococcus***: Food Poisoning.)

Of particular concern in pork is the parasitic nematode *T. spiralis*. This organism forms a cyst in porcine muscle. The organism can be transmitted to humans who consume the contaminated pork, but is readily destroyed by heating the muscle to 62 °C. Processing plants that sell pork which is not likely to be cooked again are required to heat or freeze the meat to certify that it is trichina-free.

Pork-slaughtering and processing plants have rigid sanitation programs that allow production of safe food. Good sanitation at the plant and proper handling throughout the food chain help keep microbial growth under control. Plants producing pork must keep processing temperatures below 10 °C or stop production and sanitize the equipment every 8 h. Most plants keep their working temperature low enough to require cleaning and sanitizing only once every 24 h. Cleaning and sanitizing are not the same

thing, but they are most effective when carried out together. Generally speaking, a plant is cleaned with soap and water to remove residual meat particles. Then the soap is rinsed off and sanitizer is applied to kill microorganisms that survived the cleaning. Common sanitizers used can be chlorine, ammonia, or iodine-based. (*See* **Sanitization**.)

A great deal of effort is spent to clean and sanitize the plant thoroughly, as well as educate employees about the importance of good hygiene in reducing food contamination. Microbiological status of the processing area is monitored daily. In addition, the plant must pass a sanitation inspection by a USDA inspector before the day's production can begin.

Despite all the in-plant efforts to control microorganisms, pork can still be contaminated or growth of microorganisms already present can occur as a result of product abuse in the warehouse, on the delivery truck, in the retail outlet, or in the home. Perishable foods should always be frozen or refrigerated at temperatures below 4 °C. Once cooked, pork should be kept above 60 °C or quickly cooled to under 4 °C. Many microorganisms grow rapidly in the temperature range 4–60 °C. Two very common mishandling problems that occur in the home are failing to refrigerate leftovers promptly and recontaminating cooked product by using the same utensils used with the raw product.

To minimize microbiological hazards and maximize eating quality, the recommendations in Table 2 have been devised as maximum limits for storage of pork. Molds and yeasts are of little concern in fresh pork because the high water activity allows bacteria to dominate. In dried pork items such as pepperoni, molds may grow on the surface. However, mold growth is retarded by a potassium sorbate dip applied by the manufacturer or by vacuum packaging. (*See* **Spoilage**: Bacterial Spoilage; Molds in Spoilage; Yeasts in Spoilage; **Water Activity**: Effect on Food Stability.)

Very few other hazards exist with the consumption of pork. Muscle from pigs is regularly monitored for drug and pesticide residues by the USDA. Incidences of contaminated meat have been isolated and total far less than 1% of the pork supply.

Table 2 Pork storage recommendations[a]

	Refrigerator (2–4 °C)	*Freezer (−18 °C)*
Fresh pork	4 days[b]	3–6 months[c]
Cured pork	7 days	2 months

[a]Packaging and handling prior to the consumer will greatly impact shelf-life of pork.
[b]Ground meat, 2 days.
[c]Ground meat, 1–2 months.

Food Uses and Products

Pork is a very versatile food that can be prepared in many ways. Popular entrée items include pork roast, ham steaks, barbecued spare ribs, and grilled pork chops. Pork is also a common component of many foods such as bacon on pizza or ham dices in a chef's salad.

Pork trimmings can be ground and made into patties or sausages of all types, including frankfurters, pepperoni, and Italian sausage.

Waste or Byproduct Utilization

People in the swine industry say that no part of the pig is wasted. This statement is not far from the truth. The major waste product from the live pig is manure, which is a good nitrogenous fertilizer.

Waste products from swine slaughter and processing are perhaps better referred to as byproducts because they are seldom wasted. Blood, bones, and inedible viscera from pig slaughter are usually dried and ground for use in animal feeds. Gelatin is made from collagenous proteins found in pig skin. Industrial lubricants, plastics, and rubber are made from fat trimmed off pig carcasses. Pig skin fabric can be produced from cured pork skins. Even the pig's hair is sometimes used for brushes or insulation.

Perhaps the most important byproducts of swine slaughter and processing are those used in medicine. Hormones such as pig insulin or heparin can be prepared for human use. Heart valves from pigs have been used to replace damaged or diseased human heart valves and are often used in heart disease research. Skin grafts from pig skin have been successfully used on human burn victims. These are only a few examples of many medicinal uses of byproducts from the swine industry.

Acknowledgments

Appreciation is expressed to Doreen Blackmer for research and editorial assistance.

See also: ***Bacillus***: Food Poisoning; ***Campylobacter***: Properties and Occurrence; **Cholesterol**: Role of Cholesterol in Heart Disease; ***Clostridium***: Food Poisoning by *Clostridium perfringens*; Botulism; **Curing**; **Fats**: Classification; ***Listeria***: Properties and Occurrence; **Meat**: Structure; Slaughter; Preservation; Sausages and Comminuted Products; Nutritional Value; ***Staphylococcus***: Food Poisoning; **Zoonoses**

Further Reading

Mandigo R (1989) *Pork Operations in the Meat Industry*. Washington, DC: American Meat Institute.

National Academy of Sciences, National Research Council, Food and Nutrition Board (1986) *Recommended Daily Dietary Allowances*, 10th edn. Washington, DC: National Academy of Sciences.
National Live Stock and Meat Board (1973) *Hog is Man's Best Friend*. Chicago: National Live Stock and Meat Board.
National Live Stock and Meat Board (1977) *Lessons on Meat*. Chicago: National Live Stock and Meat Board.
National Live Stock and Meat Board (1987) *Exploring Diet and Health*. Chicago: National Live Stock and Meat Board.
Troller J (1983) *Sanitation in Food Processing*. Orlando: Academic Press.
US Department of Agriculture (1983) Composition of Foods; pork products, raw, processed, prepared. *USDA Handbook*, nos. 8–10. Washington, DC: Human Nutrition Information Service.

PORT

Contents

The Product and its Manufacture

E Cristovam, University of Adelaide, Adelaide, Australia
A Paterson, University of Strathclyde, Glasgow, UK

Background

Ports are fortified wines based upon a long tradition with expert blending employed for style, create defined characters and produce the quality factor that makes port highly prized. Production methods differ from table wines primarily in that the fermentation – in some case still from spontaneous endogeneous microfloora of the must – takes place only during the 2–3 days of maceration. During this period, there is extraction of soluble compounds responsible for characteristic color and flavor characters of ports. Addition of brandy terminates the fermentation, leaving a young wine with high amounts of residual sugars. Such fortified wines have been produced in Portugal since the seventeenth century. Originally exported to Britain, since the 1960s France has been the largest importer. Many port houses are still British-owned and blenders follow house styles, believed to yield wines suited to British markets. Port is also used in elegant cocktails and receptions ('Porto de honra'). Three major product styles are recognized – Portuguese, English, and French – based on house cultures. Ports have many myths and traditions associated with Christmas, gout, the British upper classes, and formal dinner parties. Consumer rituals abound, Port should be passed to the left from the host, and ladies should not touch a Port bottle. These wines are established products with an image of luxury, with seasonal peaks in demand dominated by that from the older generation and males. Port wine is a key product for a national economy of Portugal and a symbolic asset representing the country world-wide: unfortunately, demand is decreasing, probably as a result of social and lifestyle changes. However, port is now drunk on formal occasions, and such changes may enhance the demand for higher-quality wines.

History and Origin

The designation 'port' in wines appeared during the second half of the seventeenth century when this fortified product increased in popularity in Britain, replacing Bordeaux wines in English trade. This was enhanced by the Methuen Treaty between England and Portugal in 1703, guaranteeing preferential customs tariffs for port wines to balance similar tariffs for imports of British textiles by Portugal.

At that time such wines, from the Douro in Northern Portugal, were red, dry, and coarse, with an alcohol content of 12–13%. Subsequently, brandy was added to stabilize wines, to reduce spoilage during the long sea journey around Western Europe. However, this fortification practice became central to the production process – removing excess acidity,

rounding the flavor character and creating a characteristic set of attributes that became unique to port wine. In 1756, the Marquis of Pombal set legal boundaries for the Douro Valley vineyards, annual production quotas, and strict rules for cultivation, transportation, and prices. Even a tasting was required before wine was sold. The aim was to safeguard port wine's authenticity, and effectively, a new product was created. In 1921, a further addition was made to the definition of port wines in that these must be *lodged* in the country for a given period before being deemed fit for export.

In the 1860s, phylloxera destroyed most vineyards in the demarcated region. This forced a restructuring of viticulture and winemaking techniques. At the same time, imitations of port started to appear from France and Spain. At the beginning of the twentieth century, in order to counteract such fraud and maintain quality, regulations were imposed for the production, sale, and export of Ports. Denomination of origin was reserved for fortified wines from the Douro region, in Northeast Portugal, with a minimum of 16.5% alcohol. Distillation of Douro wines was prohibited, and fortifying spirit had to be obtained from other regions.

In 1926, a bonded area was created in Gaia – Vila Nova de Gaia – a city opposing Porto on the Douro river mouth (Figure 1). Gaia became part of the demarcated region, and here, port trade companies were forced to build lodges to age wines. Now, shipments can be made from both Gaia and the demarcated Douro region. Lodges maturing in the Douro produce wines with characteristic baked flavor notes through the high temperatures experienced upriver.

By the 1930s, exports were at a peak, but prices soon fell as quality was neglected, and production exceeded market demands. To counter this, legislation was introduced to control the two facets of the port industry – production and trade. In 1932, winegrowers were organized into guilds, now Casa do Douro (CD), responsible for monitoring vineyards, wine quality, and grape brandy production. A year later, a guild of port wine shippers was established, and the Instituto do Vinho do Porto (IVP) was formed to coordinate activities. The IVP regulates the description of ports with respect to dates and indication of age, controls the seal of guarantee, and still coordinates Casa do Douro activities.

Prior to entering or leaving Gaia, all ports are submitted to a chamber of tasters, constituted by experienced specialists. This is overseen by a consultative committee of tasters from individual port firms, chosen on the basis of reputation and expertise, but confirmed and formally appointed by the Ministry of Economy. This gives expert tasting opinions and has the final decision.

In theory, any must produced within the Douro demarcated region can be used for port. However, the IVP sets limits to annual quantities of wine to be 'beneficiado' (improved or benefited). Growers must apply to the CD for the right to produce within this allowance. Applications must list the name of the

Figure 1 (see color plate 122) Vila Nova de Gaia; bonded area in Portugal.

petitioner, the municipality and parish of each property, property names, number of vines over 4 years old, estimated production, and amount of wine to be used for port. Applications are checked against a register that makes reference to a classification system. Class A vineyards are, on average, granted a *benefit* of 600 l per thousand vines and Class E 400 l per thousand vines. Production is quantified in *pipes* – the volume of a traditional 550-l oak cask. Allowances are varied according to vineyard classification, total allowance for the year, and prospects for wine production in that year.

Today, port is a product that requires careful quality management strategies. To protect the industry, there is an upper limit on the number of vintages that can be declared and maxima to the fractions of wine in houses that can be traded in any year.

Production Area

The boundaries of the demarcated Douro region – the oldest in the world – have remained largely unchanged since 1756 (although revised in 1907, with minor alterations in 1921). This region consists of four districts: Vila Real, Bragança, Viseu, and Guarda covering *c*. 250 000 acres, with three subregions: Baixo Corgo, Alto Corgo, and Douro Superior (**Figure 2**), each with individual characters. Vines cover approximately 15% of the region, with the Baixo Corgo dominating. Small vineyards can be found throughout the region, with the larger vineyards mostly located in the Douro Superior. Before planting on the steep mountainous slopes, terraces must be formed. This was done in the past with hammers and dynamite but is now done with bulldozers when possible. Grape production is difficult because of the nature of the soil, schistose with granite in certain areas, and narrow terraces. The climate is harsh with frosty winters, hot (> 37 °C) dry summers, and in certain areas wet springs.

Grape Varieties

Only approximately 10 500 hectares (26 000 acres) are authorized for vineyards that can be used for port production. Most vine varieties are regional, grafted on to rootstock types chosen for grape variety and soil compatibility. The rootstocks used are Rupestris du Lot, Berlandier hybrids, either Riparia strains (420-A, SO4, 161/49) or the Rupestris varieties (R.99 and R.110).

The varied microclimates in the Douro region have resulted in 48 grape varieties being used in port production. However, there has been a rationalization to increase quality and optimize vine planting and grape production. Currently, an official EC recommendation quotes 29 varieties for port: 15 red and 14 white. Desirable red grape varieties produce tannic, intensely colored wines. Until recently, one of the most popular was *Tinta Francisca*, derived from Pinot Noir. Other important varieties are *Touriga Nacional*, *Tinta Roriz* (*Tempranillo* in Spain), *Touriga Francesa*, *Tinta Barroca*, and *Tinto Cão*. The dominant white grape varieties are *Malvasia Fina*, *Viosinho*, *Donzelinho*, and *Gouveio*.

Of the Douro red grapes, *Touriga Nacional* is considered to produce the finest and most complex wine, although the yield is low. It is also a vigorous and adaptable variety, growing in a wide range of conditions. Wines from *Touriga Nacional* have an aroma character suggesting red fruit, often with floral notes, notably of violets. Flavors are complex, elegant, balanced, and rich in nonaggressive tannins, and finishes are long and smooth.

Another red grape, *Tinta Roriz*, grows best on fertile soils, producing large, tight clusters of thick-skinned grapes of a high quality. Wines, less colored than those from *Touriga Nacional*, also have rich aroma characters dominated by red berries, with strong herbal overtones. Flavor characters are strong and astringent from harsh tannins, and lengthy fruity finishes contribute structure in blending.

The most fragrant wines are from *Touriga Francesa*. This variety is adaptable but requires a warm growing season to completely ripen grapes. Strong floral characters, notably of rose, dominate the aroma notes. Flavor, with fruity and floral attributes and earthy overtones, is thought to be well balanced with a long, concentrated finish. These wines contribute aroma and structure to products in blending.

Tinta Barroca is a high-yielding variety with large, loose clusters of thin-skinned, oversized berries. Wines are soft, fragrant, and delicate with a lingering red fruit aftertaste, but are often lacking in complexity and light in color. These wines contribute little to blend structure. *Tinto Cão* produces a smaller grape than other varieties.

A central problem in the Douro is finding a balance between the desirable character attributes in varietal grapes and production yield. The overall must color

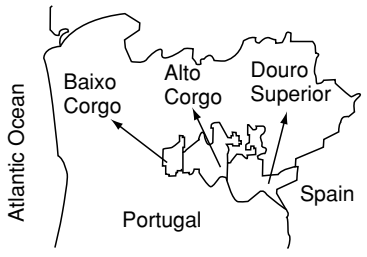

Figure 2 Região Demarcada do Douro.

in young port wines, central to production, is influenced more by cultivar than by season.

Production Method

Harvesting of port grapes is difficult. Grapes are picked entirely by hand, between 15 September and 15 October depending on the weather and region, in vineyards that slope at an average of 45°. After grape destemming, traditionally grape treading was carried out to extract must in a 'lagar' (a large open, shallow stone tank). However, this quaint procedure has generally been replaced by more readily controlled and standardized practices. Currently, most port is produced in closed fermentation tanks often equipped to turn over the wine automatically. Alternatives include open tanks, in which must is pumped around, and concrete or stainless steel tanks, fitted with mechanical stirrers. Grapes are generally broken in centrifugal crushers with pumps. In certain production sites, for total or partial removal of skins, crushed grapes fall into a blender that shreds skins completely. The fermenting must is run off when most sugars have been converted into alcohol after 2–3 days of fermentation. After separation from solids, fermentation is inhibited by *aguardente* (grape brandy) addition. Wine officially becomes port at 19–20% (abv).

Brandy fortification, in addition to quenching the spontaneous fermentation, contributes specific sensory attributes to wine, improves chemical stability, and controls the final sweetness intensity. The sweetness in ports comes only from unfermented sugars. Fermentation must therefore be terminated when the residual wine sugars contribute to the desired sweetness. The quality of brandy added to the fermenting must is also a key factor in determining the chemical composition and final flavor character potential of quality ports.

The nascent port wines are stored in tanks, vats, or casks. It is not clear whether container capacity influences final wine character, but container material certainly does. More important seems to be the storage site, i.e. whether upriver at Douro or by the sea at Gaia – concrete tanks with a tartaric acid coating or stainless steel are often used. Once transported to Gaia, wines must be submitted to the IVP tasters for classification. Group 1 wines are considered 'of immaculate quality, carrying special privileges'. Group 2 wines have slight imperfections but are fit for usage. The final Group, 3, is for wines rejected either by tasters or by analysts.

Traditional Vintage

The traditional process, predominant until 1960, consisted of handpicking of grapes by groups of workers with transport in large baskets either on the shoulders or by cart. At the winery, workers crushed grapes barefoot in the *lagar*, marching back and forth in a line, until the juice floated on top of a pulpy residue. This initiated the fermentation with must darkening as the color was extracted from skins. To maintain the fermentation, paddles stirred the must, and when sufficient sugar was converted to alcohol, the liquid was run off into wooden vats and 110 l of *Aguardente* added for each 440 l of wine.

The nascent port was held in 'toneis' and tasted regularly until it was ready to be transported by expert boatmen down the river Douro in 'barcos rabelos,' in the spring following the harvest, or vintage. Gaia was originally chosen as the bonded area because port pipes could be easily off-loaded from riverboats to lodges and transferred to seagoing vessels.

Modern Vintage

With decreased manpower through migration, major changes have been, and are, taking place in port production. The vinification process, blending policies, maturation practices, and transport to lodges in Vila Nova de Gaia have changed considerably.

The widespread introduction of electricity in the early 1960s led to the replacement of human contributions with the autovinification process. Such mechanization led to establishment of wine centers at larger *quintas*. The result was that farmers sold more grapes, and wine production was centralized to facilitate improved quality control. Closed autovinification plants and open tanks came into use, but certain producers still believe that open-tank fermentations with automatic paddles optimize the color extractions from must.

Since the 1970s, production has moved from individual *quintas* to wineries with a greater plant capacity: currently, both fermentation and storage tanks are considerably larger. By the 1980s, over 80% of port was produced in wineries, with cooperatives producing approximately 40%.

Grapes arrive by lorry in open steel containers and are weighed, tipped into reception tanks, and immediately carried by screw feeders into a crushing and stemming machine. This is typically an upright cylinder, approximately 1 m in diameter, which centrifugally removes stalks to an extent depending on the tannin requirement. The must is pumped into square concrete or triangular stainless steel tanks, autovinification vats, an autovinificator cylinder is clamped on, and a water escape valve is filled. Provided that the temperature is *c.* 16 °C, the must starts fermenting almost immediately. Fermentations last 2–3 days, with sugar contents assessed continuously from

specific gravity. The total sweetness of the port is derived from the original grape sugars.

Young ports stay in wooden vats or cement holding tanks in the Douro until they are ready for transport to Gaia in the spring by the road tankers. The narrow and winding roads of the Douro retain some of the hazards of bygone eras.

White port production uses white grapes, often with limited maceration, and wines are aged under oxidizing conditions. Grape skins acquire a golden-brown hue at harvest, and in order to produce a light gold, rather than a yellow–white, port, it is necessary to minimize skin contact during fermentation. Consequently grape skins are removed after crushing, and the maceration time is also kept to a minimum. Oxidation is not encouraged if complex floral aroma notes and pale yellow color are desired in the final wine.

Maturation and Blending

Young ports are stored in vats or casks made of oak or mahogany in the Douro. During the first winter, wine is racked (pumped off) from the lees into fresh casks. The first tasting in November uses a classification based on sensory attributes. The finest wines, produced in an exceptional grape quality year, are usually set aside with a view to *declaring a vintage*; most are used for blending. Vintages represent highpoints in port wine quality. During the initial 2 years, wines are racked frequently with aeration as often and vigorously as required for the attributes considered desirable after aging, as determined by the winemaker. Winter conditions help precipitate colored matter together with yeast and potassium salts of tartaric acid. Young ports are then observed to 'close up' or show a color intensification. Ports are transported to Gaia for classification in early spring.

Port was traditionally stored in lodges or cellars in Gaia but can now be retained in the Douro. Ruby wines, matured in the Douro, are subjected to its climatic extremes and are considered to lose vigor and freshness. A common view is that older tawny ports should mature in the Douro, and younger ruby ports in Gaia, where temperature and humidity are less variable.

On arrival at Gaia, casks are stored by vineyard and vintage in the cellars of individual port houses. Large cylindrical stainless steel vats and lined concrete are often used for the temporary storage of wines on arrival, but maturation, for at least 2 and usually for 3–40 years or longer, is in wood. Cellar-masters decide between the 550-l 'pipas' (pipes or oak casks), and larger casks (25 000–35 000 l) or vats (550 000 to 750 000 l) (Figure 3).

Figure 3 Wooden casks and vats used for wine maturation.

On arrival in Gaia, a second tasting takes place: wines are classified, and the first *lotas* (blends) are married in large vats fitted with stirrers or pumping. This is on the basis of decisions on the potential of each wine and amalgamating similar wines. There is then separation of the top-quality wines i.e. those to be used when declaring a vintage. Such wines are kept separate for two winters and a summer, and if they do not develop to predefined standards, a vintage will not be declared that year, and the wines will be used for blending.

The purpose of blending is to obtain specific final sets of sensory attributes, port types, and styles. Consequently, young port is always stored in well-seasoned vats or pipes to retain the fresh aroma characters and deep red color with purple hues. As production is based on traditions, wooded ports are created to ensure continuity of style; new blends contain a percentage of previous wines. Previously, individual pipes were blended, but currently, concrete or large wood vats are filled from maturation casks.

Ports are blended from wines varying in age and character. Cellar-masters choose young wines on the basis of fruit flavor notes and color, and older wines for body and 'smoothness.' The objective is a character and quality typical of the style of that port house.

Regulations stipulate aging of ports at least 3 years in cask or bottle prior to release.

Aging

Port wine storage leads to varied changes in the wine character and composition. The type and capacity of containers are selected for the character evolution required, but those from grapes also contribute. Ports are aged under oxidizing conditions except for vintage and late bottled vintage (LBV) products. Young ports are blue–purple in color – gradually turning red, then tawny and progressively paler with age, eventually achieving amber and golden tones (Figure 4). The aging method varies. All red ports are aged for at least 2 years in oak casks and continue maturation either in casks (ruby, tawny, aged tawnies) or in bottles (vintage, LBV). Ruby ports retain an intense dark red color with the sensory characteristics of young wines.

Red ports are aged in wooden vats and pipes, stacked four high to minimize evaporation. With white ports, in order to avoid color gain through prolonged wood maturation, concrete vats are used.

With labor costs increasing, larger vats (5–400 pipes in capacity) are used more often for aging wines. The rate of maturation is inversely related to volume. As a consequence, wines are selected early according to style requirements and then stored in appropriately sized containers. Traditional lodges in the Douro exhibit wine liquid temperature changes of up to 20 °C. The atmosphere is also dry, leading to faster evaporation and maturation rates and baked or maderized flavor notes. In Gaia, the relative humidity is constant (approximately 54%), evaporation lower, and temperature fluctuations under normal storage conditions limited to 14–16 °C. Maturation is slower, producing fresher wines than the Douro.

Wood

Although the character of the grape variety is central to wine flavor development, wood has been considered to play important roles in port maturation. Staves are from local and imported oak (including some used for Scotch whisky casks), chestnut, and Brazilian mahogany. Dried fruits and spicy aroma notes, and smooth mouthfeel (texture) characters, attributed to lengthy wood maturations, characterize premium tawny ports. Traditional guidelines for production include the use of small oak pipes with frequent racking to induce sufficient aeration. Contact between the wood surfaces and the wine, regular racking, and oxygen migration through the staves have been reported to be important variables in port maturation.

Unlike table wines, port is aged in a varied range of container sizes. Wood is often reused many times, depending on the style and cellar practice. New wood is rarely used and, if required, is seasoned thoroughly with a second pressing wine, table wine, and/or young red port before deployment. Consequently, some experts believe that little woody character should be detectable in ports, through the use of extractive-depleted staves, except in the older tawnies matured in the Douro.

LBV ports are stored in large vats, and tawny ports are stored in pipes. Frequent racking with aeration of wines influences maturation. Wood aging of such wines is believed to be primarily an oxidative process with limited parallel extraction of nonflavonoid phenolics from stave lignins and tannin degradations.

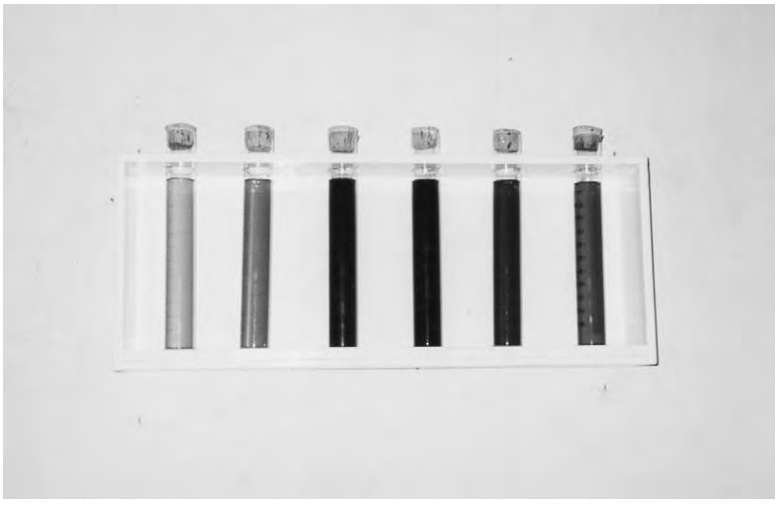

Figure 4 (see color plate 123) Samples of different types of port, showing the range of colors.

Wine maturation is thought to occur mainly from evaporation and oxidation.

Fining and Stabilizing

Fining, traditional for clarifying wines, removes a certain amount of color and tannins to produce softer and 'older' wines. Port is usually filtered: modern methods have replaced the cloth bag method of separating wine from lees. Modern filters consist of a series of woven nylon pads that allow lees to circulate until the pads are evenly coated with solid material. The resulting solid cake then acts as a filter that introduces clarity in the wine. Bentonite may be used with white wines to remove protein.

Refrigeration is now used to stabilize wines: chilling precipitates unstable coloring matter and the potassium salts of tartaric acid. Chilled wine is passed through a vessel that encourages deposition of potassium hydrate tartrate. Chilling is followed by further filtration (continuous refrigeration). There may be wine retention in an insulated tank for up to a week before filtering (Gasquet process). Finally, the clarified product is passed through a plate heat exchanger separated by thin stainless steel plates from incoming wine at ambient temperature.

Refrigeration and sulfur dioxide addition, prior to bottling, allow wines to remain bright in bottle for up to 18 months under normal storage conditions. Brief pasteurization may also be used after cold treatments. In addition, Kieselguhr with pad filtration may be a final processing after fining or refrigeration.

Bottling

In earlier years port was shipped in pipes to England for transfer to squat and bulbous bottles, designed for transporting wine from cask to table. Oxidation was a problem, since stored on their side such bottles did not allow cork contact. In 1775, bottles with longer bodies and shorter necks, were introduced to allow better storage. This change initiated the development of bottle aged vintage ports.

By the 1930s, bottling had become more centralized as suppliers, instead of merchants, could now bottle under customs and excise supervision. Duty was not payable until port was declared from bond. In the 1970s, more shippers began bottling under their own labels instead of shipping in bulk. The result was that it became possible to develop the international brands that currently dominate the consumer's attention.

Port Styles

The IVP governs how producers may use labels to describe port wines. There are four major premium categories of port: vintage, LBV, quinta, and aged tawny (Figure 5).

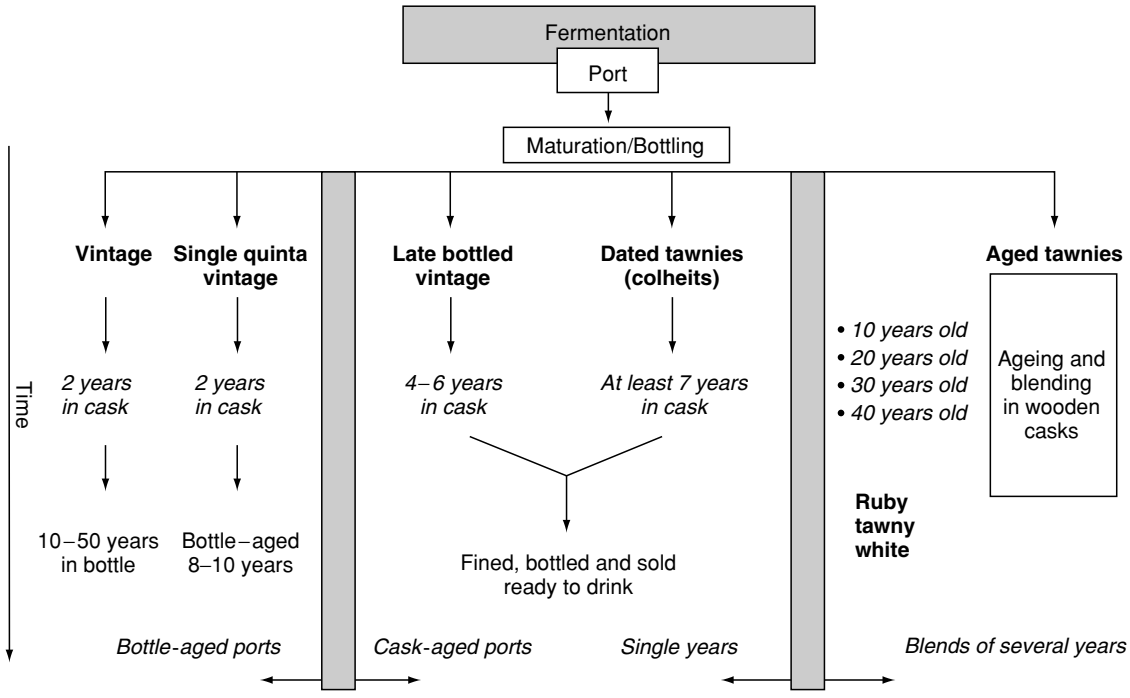

Figure 5 Processes involved in the different types of port production.

Vintage ports are wines of a single harvest, produced in a specifically recognized year, and have sensory characters regarded as exceptional. Such port wines spend 2–3 years in wooden casks or vats prior to bottling. The IVP has rights to descriptions of vintage and corresponding dates. Subsequently, maturation reactions occur in the bottle for 10–20 years and in some cases up to 50 years, with little possibility of oxidation. Such ports are never fined or filtered; the wines generate precipitates, or 'throw a crust,' and require decanting prior to consumption.

LBV ports are wines of a single harvest, aged in wood for 4–6 years. Bottling is between 1 July of the fourth year and 31 December of the sixth year. Bottle maturation reactions may continue for 5 years or more with continuing improvements. Traditional, now rare, LBVs are not fined or filtered, and throw a crust with improvements on bottle aging. However, modern LBVs are treated to remove solids.

Quinta are wines that come from grapes from a single property, named on the label. Such ports are made only in years not deemed good enough to declare a vintage. Single Quinta vintage ports have been available for some time, but currently, the market is expanding to include this designation for both ruby and tawny ports.

Aged tawnies are typically matured for 10, 20, 30, or 40 years, the average of ports blended. Old tawnies have a characteristic gold color. *Dated tawnies* are wines from single harvests, matured in oak casks for at least 7 years with the harvest and bottling date specified on the labels. A special reserve tawny contains wines that have spent 3–4 years in the cask prior to blending.

Wines Matured in Wood

Wood-matured ports, such as tawnies, are treated to encourage color lightening. Blend components are racked at least once a year to remove deposits and ensure oxygen dissolution. During the initial years of aging, container size is thought to be of little importance. Frequent use of wooden containers of considerable size such as 650-l oak casks, or vats, for short periods, stainless steel or tartaric acid-lined tanks.

Racking and the resulting wine aeration lead to oxidation, essential for the maturation of these port wines. During racking, more alcohol may be added, or maturing wine may be 'refreshed' by adding small quantities of younger wine. On creation of the blend, to assist maturation, wines are usually transferred to smaller vats or casks. When the desired degree of maturity has been reached, the final standard blends, with sensory attributes specific to each company, are prepared for bottling and export. As noted previously, no port may be sold for consumption in Portugal or abroad until it is at least 3 years old.

Wines Matured in Bottle

Vintage and LBV ports, bottled after only a few years in casks, have sensory attributes that are more dependent on must characters than on wine treatment. Vintage ports are racked infrequently to prevent excessive aeration, using smaller casks that are considered to provide a more controlled and uniform aeration. Bottles are stored in dark cellars of constant temperature with controlled ventilation and humidity. Color loss occurs slowly compared with wooded wines, and bottled ports retain their characteristic fruitiness, retaining a full flavor and purple–red colors for many years.

See also: **Grapes**; **Port**: Composition and Analysis

Further Reading

Bakker J (1986) HPLC of anthocyanins in port wines: determination of aging rates. *Vitis* 25: 203–214.

Bakker J, Bellworthy SJ, Hogg TA *et al.* (1996) Two methods of port vinification: a comparison of changes during fermentation and of characteristics of the wines. *American Journal of Enology and Viticulture* 47: 37–42.

Bakker J, Bridle P, Timberlake CF and Arnold GM (1986) The colors, pigment and phenol contents of young port wines: Effects of cultivar, season and site. *Vitis* 25: 40–52.

Bakker J and Timberlake CF (1985) An analytical method for defining a Tawny port wine. *American Journal of Enology and Viticulture* 36: 252–253.

Bakker J and Timberlake CF (1986) The mechanism of color changes in aging port wine. *American Journal of Enology and Viticulture* 37: 288–292.

Burnett JK (1985) Port. In: Birch GC and Lindley MG (eds) *Alcoholic Beverages*, pp. 161–169. London: Elsevier Applied Science.

Cristovam E, Paterson A and Piggott JR (in press a) Development of a vocabulary of terms for sensory evaluation of dessert port wines. *Italian Journal of Food Science*.

Cristovam E, Paterson A and Piggott JR (in press b) Differentiation of port wines by appearance using a sensory panel: comparing free choice and conventional profiling. *European Food Research and Technology*.

Fletcher W (1978) *Port: An Introduction to its History and Delights*. London: Philip Wilson.

Fonseca AM, Galhano A, Pimentel ES and Rosas JR-P (1984) *Port Wine. Notes on its History, Production and Technology*. Oporto: Instituto do Vinho do Porto.

Garcia-Viguera C, Bakker J, Bellworthy SJ *et al.* (1997) The effect of some processing variables on non-coloured phenolic compounds in port wines. *Zeitschrift für Lebensmitteluntersuchung und -forschung* 205: 321–324.

Gowell RW (1972) The manufacture of port. *Process Biochemistry* Oct.: 27–29.
Howkins B (1982) *Rich, Red and Rare. The International Wine & Food Society's Guide to Port*. London: William Heinemann.
Reader HP and Dominguez M (1995) Fortified wines: sherry, port and Madeira. In: Lea AGH and Piggott JR (eds) *Fermented Beverage Production*, pp. 159–203. London: Blackie Academic & Professional.
Robertson G (1992) *Port*. London: Faber and Faber.
Williams AA, Langron SP, Timberlake CF and Bakker J (1984) Effect of colour on the assessment of ports. *Journal of Food Technology* 19: 659–671.

Composition and Analysis

E Cristovam, University of Adelaide, Adelaide, Australia
A Paterson, University of Strathclyde, Scotland, UK

Introduction

The demarcated Douro region – established in 1756 – consists of three subregions: Baixo Corgo, Alto Corgo, and Douro Superior (**Figure 1**), each with individual characters. Vines cover approximately 15% of this steep mountainous region, with the Baixo Corgo dominating. Small vineyards can be found throughout the region, with the larger mostly located in the Douro Superior. Vineyards producing grapes for port wines are graded on a point system, biased towards environment – *c*. 75% for vine productivity, altitude, soil, locality, and vine training, 14% from grape cultivars and degree of site slope, and >10% from local features. Grape varieties favored (around 7% points) include Tinta Francisca, Tinta Cão, Touriga Nacional, and Mourisco. Low-scoring musts are utilized in table wines.

Oxidative Aging – Tawny Ports

Tawny ports, matured from 10 to 40 years and particularly important, have maturations dominated by oxidative aging in wooden casks allowing oxygen migration, supplemented by oxygenation during frequent initial rackings. Small 'pipes,' as opposed to large vats, insure high cask surface area-to-volume ratios.

Reactions between wine congeners, together with cask-derived components, are important. Congener oxidations facilitate flavor evolution to characteristic mellowness with dried fruits, caramel, and vanilla notes. Most obvious, and dominating, are relationships between aging and color, with oxidations of red and purple anthocyanins, and losses as insoluble precipitates that include tannins. Tawny ports have yellow, amber, and tawny hues at intensities related to maturation period and cask choice, at rates dependent on original varietal choice and wine composition.

Reactions between acetaldehyde, anthocyanins and catechins yield reddish-blue condensation pigments that darken ports, 'closing up,' over the initial 6 months. Progressive lightening follows, with hue changes and gradual browning from competing aldehyde-induced reactions and direct condensations of anthocyanins with phenolics (**Figure 2**). Less is known about wood contributions to tawny ports spending over 50 years in old and recycled casks than character in red table wines. For tawny ports, color loss is essential and style-defining; in red wines it is detrimental to quality. (*See* **Antioxidants**: Natural Antioxidants.)

Nonoxidative Aging – Vintage Ports

Vintage ports age in bottles for up to 30 years. Initial oxygen, absorbed during bottling, is slowly reduced by reactions with anthocyanins and other wine components. Color is lost more slowly than in cask

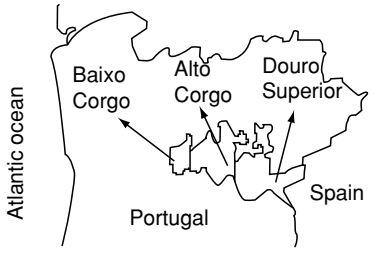

Figure 1 Região Demarcada do Douro – demarcated Douro region.

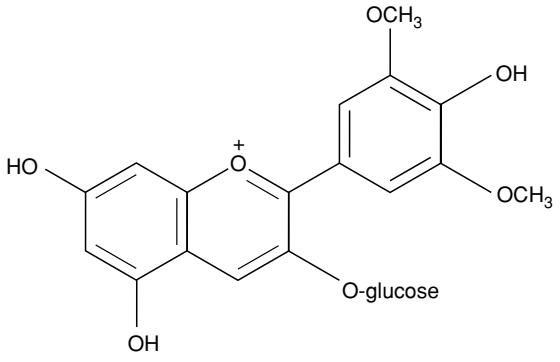

Figure 2 Oxidation/polymerization of phenolic components in ports.

maturations and the initially full-bodied wines retain flavor intensity, fruity characters, purple-red coloration, and freshness of youth for many years. Limited oxidation reduces aldehyde formation, precipitation of insoluble phenolic products, and leucoanthocyanins loss, yielding more condensed products.

Vintage ports show direct relationships between age and volatile acidity, and contents of ethyl acetate, pentyl alcohols, and sodium. This is from either maturation or modern wine-making strategies, including reductions in air contact of grapes; introduction of coating equipment; decanting without aeration; replacement of sodium metabisulfite by SO_2; and quality control in fortifying grape spirit. Clear aging effects include decreases in tartaric acid through deposition of potassium hydrogen tartrate, in citric acid content and total acidity, and increases in glycerol content and changes in alcohols.

Oxidative Reactions

Formation of Esters

In acidic, high-ethanol port wines, esterifications, slow in table wines, proceed at rates influenced by temperature and pH (pH 3.5–4.0), yielding ethyl esters of organic acids – lactic, succinic, malic and tartaric acids. Nonvolatile phenolics influence ester solubilities in the aqueous ethanol, removing such aroma-active congeners from head spaces, influencing perceived character.

Esters of higher (*iso*amyl, hexyl and 2-phenylethyl) alcohols show inverse maturation relationships through hydrolyses with equilibria functions of initial concentrations, those of parent acids and storage temperature. In aging, ethanol and acetaldehyde oxidations yield acetic acid, influencing pH, ester equilibria, and forming ethyl acetate. Such ester formation appears central to desirable flavor changes in fortified wines. (*See* **Oxidation of Food Components**.)

Formation of Carbonyl Compounds: Aldehydes, Acetals

In maturations, important aldehydes are converted to alcohols or combine with sulfur dioxide. Acetaldehyde is most abundant, but other aldehydes are derived from carbohydrate (furfural, 5-methylfurfural, and 5-hydroxymethylfurfural) or lignin degradations (vanillin, syringaldehyde, coniferaldehyde, and sinapaldehyde, cinnamaldehyde). Residual carbohydrates (approximately $100\,g\,l^{-1}$) in port contribute to such reactions.

Furfural, acetaldehyde, and other higher aldehyde contents can be directly related to maturation period, especially at higher temperatures. Woody flavor characters have been linked to aldehyde polymerization.

In acidic wines, acetals are principally formed from reactions between monohydric alcohols, and aldehyde carbonyl groups, notably of acetaldehyde. Oxidations yield flavor-active aldehydes, notably benzaldehydes, also contributing to acetal formation.

Other Reactions

Phenols derived from wood maturations include guaiacol, phenol, *m*-cresol, eugenol, 4-vinylphenol, and 4-vinylguaiacol. Lactones include β-methyl-γ-octactone. Wood-extracted aldehydes include vanillin, syringaldehyde, coniferaldehyde, and sinapaldehyde.

Classic Enological Analyses

Port Acidity

Total titratable acidity values of port wines at pH 7.0 range from 5.86 to $3.45\,g\,l^{-1}$ as tartaric acid. Older wines have increased volatile and total acidities through acetic acid formation. In young ports, total acidity is reduced by tartrate precipitation.

Volatile Acidity

Determined by official Office International de la Vigne et du Vin (OIV) methods, volatile acidity is dominated by acetic acid (around 90%) but propionic, formic, and butyric acids also contribute. Such acids are from yeast and lactic acid bacterial fermentations and wood maturation reactions. Port wines generally have volatile acidity values $< 0.35\,g\,l^{-1}$ acetic acid, although $1.02\,g\,l^{-1}$ was recorded for a 40-year-old tawny and a 1960 Colheita tawny.

Organic acids

Grapes contribute tartaric and malic acids to wines. Tartaric acid concentration varies during grape ripening, and is more related to climatic conditions than varietal character. Tartrate precipitation during maturation decreases wine contents to 0.445–$1.688\,g\,l^{-1}$, independent of age or wine type.

In contrast, malic acid contents, more dependent on grape variety than climate, fall during ripening. Contents are lower in younger port wines, than older, possibly through later picking in recent vintages.

Lactic acid, formed by strain-dependent anaerobic yeast metabolism, is present in port wines in the range $0.3–1.68\,g\,l^{-1}$, independent of wine age.

Type differentiation was achieved on the basis of contents of citric acid and alcoholic strength (Table 1) in 21 red and four white ports, and six Spanish

port-style wines. Australian port-style wines have higher ethanol contents. (*See* **Acids**: Natural Acids and Acidulants.)

Alcoholic Strength

Alcoholic strength (% v/v ethanol) in red ports ranges between 19 and 22%, with white ports as low as 16.5% (light, dry white port). Spirit contributes about 20% initial volume to port wines; subsequent alcoholic strength adjustments dilute wine congeners, and increase precipitation of tartrates and extraction of wood substances. Alcoholic strength changes with water and ethanol evaporation through staves; it is dependent on storage conditions and wine sugar content.

Sulfur Dioxide

Total SO_2 is from summations of iodometric titrations of free and bound components in acid conditions after a standard pretreatment by the Portuguese official method (Portaria 985/82 and pr NP-2220). Values between 9 mg l^{-1} (Colheita tawny, cask wine) and 99 mg l^{-1} (10-year-old tawny, commercial wine) have been reported. In vinifications, SO_2 is added for microbial stability and antioxidant properties. Free SO_2 rapidly binds to ethanol but also hydrogen peroxide, produced during phenol oxidations, reducing further oxidation.

Reducing Sugars and Glycerol

Must origin and quality, shipper requirements, and house style, all influence residual sugar content, typically 80–120 g l^{-1} from the Lane-Eynon method.

Certain sweet red wine musts – geropigas – can be 150 g l^{-1} residual sugar; drier, more fermented musts have 20–50 g l^{-1}. White geropigas are made for blending or specialty products; dry white aperitif ports contain < 50 g l^{-1} sugar. Glucose and fructose concentrations range from 161 g l^{-1} (Colheita tawny, cask wine) to 63 g l^{-1} (30-year-old tawny, commercial wine). Fructose at 60–70 g l^{-1} is more abundant than glucose (40–50 g l^{-1}).

Typical must specific gravities are 1090–1100 for red and 1085–1095 for white ports; fermentations are terminated around 1045. In Australian port-style wines, specific gravities (20/20 °C) of 1036 and 1080 make these even sweeter than the sweetest port wines (Table 2).

The IVP (Instituto do Vinho do Porto) recognizes classes of port wines, varying in sugar content (Table 3).

Glycerol, the second most abundant alcohol, contents are from 3 to 8 g l^{-1}. Glycerol acetals are found as in other aged fortified wines, and *trans* 5 hydroxymethyl-2-methyl-1, 3-dioxolane in port.

Dry Extract, and Folin-Ciocalteu Index

Dry extract values are determined indirectly as density of the residue without alcohol, following replacement of alcohol with an equivalent volume of water in the Portuguese norm (Portaria 985/82 and pr NP-2222) and OIV methods. Values in Colheita tawnies range from 120 to 188 g l^{-1}. Maturation changes include: alcohol and water evaporations; extractive increases; and precipitation of salts,

Table 1 Enological analyses of red and white ports and similar Spanish and Australian wine styles

	Portuguese red ports	*Portuguese white ports*	*Spanish port-styles*	*Australian port-styles*
Alcoholic strength (% vol)	18.6–21.5	19.5–19.9	16.9–19.0	22–23
pH	3.54–3.99	3.54–3.57	3.10–3.72	3.69
Total acidity (g l^{-1})	1.81–3.04	2.06–2.63	2.26–3.09	NA
Volatile acidity (g l^{-1})	0.12–0.49	0.16–0.25	0.26–0.41	NA
Phosphoric acid (g l^{-1})	0.23–0.33	0.23–0.29	0.18–0.28	NA
Tartaric acid (g l^{-1})	0.94–1.45	1.19–1.69	0.97–1.85	NA
Citric acid (g l^{-1})	0.08–0.26	0.11–0.17	0.12–1.03	NA
Malic acid (g l^{-1})	0.06–2.07	0.94–1.39	0.35–0.70	1.66–1.71
Lactic acid (g l^{-1})	0.34–2.42	0.33–0.94	0.34–1.19	0.37–1.20
Succinic acid (g l^{-1})	0.18–0.33		0.17–0.31	NA

NA, not analysed.

Table 2 Sugars in red and white ports, Spanish and Australian port-style wines

	Portuguese red ports	*Portuguese white ports*	*Spanish port-styles*	*Australian port-styles*
Glucose (g l^{-1})	30–48	41–44	24–49	76–118
Fructose (g l^{-1})	48–67	53–59	32–52	93–148
Reduced sugars (g l^{-1})	101–119		69–115	NA

NA, not analysed.

Table 3 Sweetness in port wines

Sweetness level	*Gravity*	*°Baumé*	*Sugars (g l^{-1})*
Very sweet (Lágrima)	> 1034	> 5.0	> 130
Sweet	1018–1033	2.8–5.0	90–130
Semidry	1008–1017	1.4–2.7	65–90
Dry	998–1007	0.0–1.3	40–65
Extra dry	< 998	0.0	< 40

phenolic compounds, proteins, polysaccharides, and other components.

The Folin-Ciocalteu index, from European Economic Community (EEC) recommended method (EEC no. 2676/90), expresses wine total phenols (anthocyanins, tannins, flavones, flavonols, catechins, cinnamic acids, and related compounds).

Tannins and Phenolic Compounds

Tannin levels in ports have been reported to between 400 and 600 mg l^{-1}.

Color in ports, principally from soluble grape-skin anthocyanins (Figure 3), is from malvidin (3-glucoside, 3-*p*-coumarylglucoside, and 3-acetylglucoside, in decreasing order). Varietal differences in such compounds are important. Aldehyde-bridged polymers of anthocyanins and other phenolics dominate over direct condensation of anthocyanins with phenolics, through free aldehyde contents (50–100 mg l^{-1}) of young ports. With aging, anthocyanins rapidly disappear, replaced by polymers of progressively decreasing absorbance, at rates depending on racking and cask choice. Prolonged oxygen diffusion yields wine of tawny coloration influenced by wood extractives and composition.

Wine aeration during filling and racking seems more important for color changes than subsequent oxygen diffusion from studies of varietal port wines aged in miniature recycled-stave casks. The reduced soluble extracts, including ellagic tannins, of recycled wood suggest limited contribution to wine changes, dominated by an evolution in phenolic components. Port maturations are likely primarily driven by oxidation, influenced by grape varietal differences, with inert but permeable casks making little direct contribution. (*See* **Tannins and Polyphenols**.)

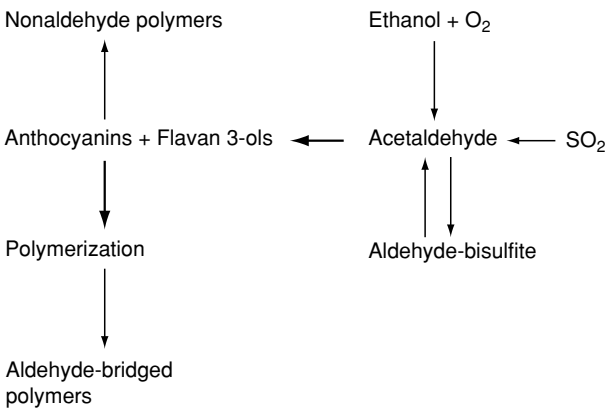

Figure 3 Structure of malvidin 3-glucoside, an anthocyanin occurring in grapes used for port wine production.

Color Measurements

Spectrophotometric measurements at $A_{520\,nm}$ and $A_{420\,nm}$, yield tint ($A_{420\,nm}/A_{520\,nm}$), and color density ($A_{520\,nm} + A_{420\,nm}$) values and at natural pH specify redness and brownness in color, respectively. CIELAB (Commission Internationale de l'Èclairage) 76 L*a*b values (hue, angle, and chroma), are high L* for lightness and low for darkness; a* measures redness and b* brownness. Total pigments and total phenols can also be determined.

Metal Ions

Trace metal contents differentiate between Portuguese ports and similar styles in wine from Spain. Rubidium and manganese contents (Table 4) mirror soil differences, although rubidium varies with maceration intensity, as can lithium, and manganese from vineyard treatments. Iron and aluminum concentrations in Spanish wines may also reflect soil differences.

Volatile Compounds in Port

Thirty-five volatile aroma components were quantified in 20-year and 100-year Australian port-style wines with higher ethyl lactate (110–370 mg l^{-1}) and diethyl succinate (61–130 mg l^{-1}) than in younger wines. Only the former contributes to aroma. *Trans*-β-methyl-γ-octalactone (oak lactone) contents were 0.8–2.2 mg l^{-1}.

Furfural, 2-acetylfuran, 5-methyl furfural, ethyl lavulinate, ethyl furoate, and 5-ethoxymethyl furfural are derived from carbohydrate degradation, heating, and prolonged storage. The older wine had more volatile components from wood and carbohydrate degradation.

Table 4 Metal ions in red and white ports and Spanish port-style wines

	Portuguese red ports	*Portuguese white ports*	*Spanish port-styles*
Rb (mg l^{-1})	0.66–1.43	NA	0.00–0.44
Mn (mg l^{-1})	0.77–1.32	NA	0.20–0.58
Al (mg l^{-1})	0.09–1.81	NA	1.91–4.01
Fe (mg l^{-1})	1.0–3.4	NA	3.6–7.4
Li (μg l^{-1})	17–54	31–37	45–64

NA, not analysed.

In two commercial Portuguese ports, a ruby (young wines retaining an intense dark-red color) and a tawny, 141 volatile flavor components were identified: 81 esters, 14 alcohols, 9 dioxolanes, 6 hydrocarbons, 5 acids, 5 carbonyls, 4 nitrogen components, 3 hydroxy carbonyls, 2 phenols, 2 alkoxyalcohols, 2 alkoxyphenols, 2 halogen compounds, 2 lactones, 2 other oxygen heterocyclics, 1 sulfur component, and 1 diol.

Alcohols were most abundant, followed by esters, dominated by ethyl esters and lower-molecular-weight acetates, the latter increasing with maturation. The tawny port was rich in flavor-active succinates (approximately 100 $\mu g\,g^{-1}$) from slow esterification and *trans*-esterification reactions at wine pH and maturation stave contact. Ethyl lactate was present at 15 and 30 $\mu g\,g^{-1}$ in ruby and tawny ports, respectively, and together with diethyl malate and succinate may indicate age, as can acidic degradation of fructose to 5-hydroxymethylfurfural. Other furan derivatives contribute baked and caramelized flavor notes to older wines. The flavor contribution of β-methyl-γ-octactone, important in oak-aging, is unclear in port wines, as are relationships between maturation and concentration of vitispirane (abundant in the ruby; traces in the tawny).

Reserva ports (single harvest without blending), matured in wood under oxidative aging, show maturation increases in diethyl succinate, ethyl acetate, ethanal, octanal, glyoxal, and methyl glyoxal (Table 5) and also benzaldehyde (oxidation of the alcohol), *trans*-2-octenal, 2-decanone, and 2-tridecanone, but inverse relationships for acetoin and methanol.

Differentiation of red ports (21) from similar Spanish wine (6) on the basis of volatiles concentrations are shown in Table 6.

Characteristic port volatile components are esters and alcohols, the former dominated by ethyl esters, acetates, and succinates. Older tawny ports are rich in, and age-differentiated by, contents of 2-propenoic acid, diethyl succinate, butanedioic acid, oak lactone, ethyl dodecanoate, *iso*amylbutyrate, 1-butanol, decanoic acid, and ethyl lactate, amongst other components. Vintage ports have been differentiated from late-bottled vintage (wines of a single harvest, aged in wood for 4–6 years) from 3-butanoic acid contents.

Table 5 Concentration of volatile compounds in two Portuguese Reserva ports

	Reserva ports	
	1968	*1991*
Glyoxal ($\mu g\,l^{-1}$)	1302.9	669.7
Methyl glyoxal ($\mu g\,l^{-1}$)	1981.2	898.1
Diethyl succinate ($mg\,l^{-1}$)	16.3	2.2
Methanol ($mg\,l^{-1}$)	202	213
Ethanal ($mg\,l^{-1}$)	78.3	94
Ethyl acetate ($mg\,l^{-1}$)	40.4	48
Acetoin ($mg\,l^{-1}$)	27.6	64.3
Octanal ($\mu g\,l^{-1}$)	128	79.7
Benzaldehyde ($\mu g\,l^{-1}$)	561.9	83.4
Trans-2-octenal ($\mu g\,l^{-1}$)	291.8	90.2
2-Decanone ($\mu g\,l^{-1}$)	71.2	13.7
2-Tridecanone ($\mu g\,l^{-1}$)	91.5	65.8

Table 6 Volatile congeners in red ports and Spanish port-style wines

	Portuguese red ports	*Spanish port-styles*
Methanol ($mg\,l^{-1}$)	259–360.3	113.6–275.8
Propanol ($mg\,l^{-1}$)	46.9–66.4	15.9–43.9
Isobutanol ($mg\,l^{-1}$)	93.3–146.7	21.7–30.0
2-methylbutanol ($mg\,l^{-1}$)	62.7–97.1	14.9–23.2
3-methylbutanol ($mg\,l^{-1}$)	237.4–344.1	54.0–92.8
Hexanol ($mg\,l^{-1}$)	3.5–6.6	1.0–2.3
Other higher alcohols ($mg\,l^{-1}$)	477.6–701.3	137.7–220.6
Propanol/isobutanol ($mg\,l^{-1}$)	0.40–0.57	0.72–1.83
Isoamylacetate ($mg\,l^{-1}$)	0.33–0.56	0.13–0.33
Octanoic acid ($mg\,l^{-1}$)	0.56–1.34	0.24–0.66
Ethyl hexanoate ($mg\,l^{-1}$)	0.17–0.49	0.11–0.26
Ethyl octanoate ($mg\,l^{-1}$)	0.33–0.69	0.09–0.27
Other esters ($mg\,l^{-1}$)	0.53–1.30	0.22–0.53

Flavor in Port Wines

Port wine flavors originate in the volatile and nonvolatile components in grapes, from fermentation or maturation. Although ports have traditionally been vinified using endogenous microflora, including a range of yeasts and bacteria, there have been moves towards encouragement of yeast mono- or at least mixed cultures, for example *Saccharomyces cerevisiae* UCD 522. Butanol, isobutanol, and 1-hexanol originate from yeast fermentation of carbohydrate and amino acid sources; hexyl acetate and 2-phenethyl acetate are also fermentation products. Ethyl lactate and diethyl succinate, resulting from esterification reactions, increase in the wines due to storage and long maturations. Oxidation reactions originate mainly from benzaldehyde and ethyl-2-phenylacetate, while 2-phenethanol and oak lactone result from wood extraction.

Few published studies describe flavor components, as opposed to volatiles. Young ports have notably astringent flavor characters through high contents of total and individual phenols with a *harshness* mellowed by wood maturations. Older ports have 'softer' characters with increased flavor complexity.

Flavor characters evolve through oxidation and polymerization of solution nonvolatile phenolic

Table 7 Variables used to model vintage age prediction

Variable	*Code*
Alcoholic grade	A
Volatile acidity	B
Total acidity	C
Glycerol	D
Ethyl acetate	E
Pentyl alcohols	F
Tartaric acid	G
Citric acid	H
Na	I
DO420	J
DO520	K
Sudraud's index (IC)	L
Tone (TON)	M
Garoglio's aging index (IE)	N
Anthocyanins	O
Tone (H*)	P
Chroma (C*)	Q
Clarity (L*)	R
Saturation (S)	S

Tone (H*), chroma (C*), clarity (L*), and saturation (S) are four CieLab parameters.

compounds, changing partitioning of aroma-active components between wine head spaces and liquid. The products of oxidations of congeners and ethanol, depolymerizations and extraction of wood components all influence partitioning. Grape varietal character is central to flavor development, but wood aging provides oxidative processes and nonflavonoid phenolics from lignin and tannin degradations determining final character.

Relating key sensory attributes to concentrations of volatile flavor components would assist understanding of port wine character. Partial least-squares (PLS) regression has facilitated modeling relationships between compositional factors and age in vintage ports. Certain variables (A-I, Table 7) showed a low modeling power (32–49%) in predicting age, while others (J–S) contributed greatly (above 70%). Wine aging is generally understood as polyphenol evolution, using Sudraud, Garoglio, and other indices including color evolution. The importance of carbonyl compounds in vintage port wine aroma has also been recognized.

In further PLS modeling of ports, clear relationships between sensory data and concentrations of head space components, and between sensory and nonvolatile component data, support hypotheses that changes in nonvolatile phenolics influence the perception of flavor character through changing head space congener concentrations. Relationships were also established between such components and total maturation age, through oxidation and polymerization of wine compounds. Prediction of intensity of certain flavor notes has also been possible from solution concentrations of congeners, notably for caramel and smooth characters. (*See* **Flavor (Flavour) Compounds:** Structures and Characteristics; **Sensory Evaluation:** Aroma.)

Port and Replica Wines Production

With the success of the Portuguese product, there have been a number of attempts to produce similar wines, including Starboard, in other countries, notably Australia, California, and South Africa. In 2001 port sales were healthy, suggesting a good future. However, non-Portuguese producers have been required, with changes in international trading, to stop the use of the port descriptor in the labeling of their fortified wines and future exports look problematic.

See also: **Alcohol**: Properties and Determination; **Grapes**; **Port**: The Product and its Manufacture; **Tannins and Polyphenols**

Further Reading

Bakker J and Timberlake CF (1986) The mechanism of color changes in aging port wine. *American Journal of Enology and Viticulture* 37: 288–292.

Bakker J, Bridle P, Timberlake CF and Arnold GM (1986) The colors, pigment and phenol contents of young port wines: Effects of cultivar, season and site. *Vitis* 25: 40–52.

Bakker J, Bellworthy SJ, Hogg TA *et al.* (1996) Two methods of port vinification: a comparison of changes during fermentation and of characteristics of the wines. *American Journal of Enology and Viticulture* 47: 37–42.

Burnett JK (1985) Port. In: Birch GC and Lindley MG (eds) *Alcoholic Beverages*, pp. 161–169. London: Elsevier Applied Science.

Cristovam E, Paterson A and Piggott JR (2000) Development of a vocabulary of terms for sensory evaluation of dessert port wines. *Italian Journal of Food Science* 2: 129–142.

Cristovam E, Paterson A and Piggott JR (2000) Differentiation of port wines by appearance using a sensory panel: comparing free choice and conventional profiling. *European Food Research Technology* 211: 65–71.

Curvelo-Garcia A (1988) *Controlo de Qualidade dos Vinhos*. Portugal: Odivelas.

Fonseca AM, Galhano A, Pimentel ES and Rosas JR-P (1984) *Port Wine. Notes on its History, Production and Technology*. Porto: Instituto do Vinho.

Garcia-Viguera C, Bakker J, Bellworthy SJ *et al.* (1997) The effect of some processing variables on non-coloured phenolic compounds in port wines. *Zeitschrift für Lebersmittel Untersudnung und Forschung A*, 205: 321–324.

Gowell RW (1972) The manufacture of port. *Process Biochemistry* **Oct.**: 27–29.

Howkins B (1982) *Rich, Red and Rare. The International Wine & Food Society's Guide to Port*. London: William Heinemann.

Ortiz MC, Sarabia LA, Symington C, Santamaria F and Iniguez M (1996) Analysis of ageing and typification of vintage ports by partial least squares and independent modeling class analogy. *The Analyst* 121: 1009–1013.

Reader HP and Dominguez M (1995) Fortified wines: sherry, port and Madeira. In: Lea AGH and Piggott JR (eds), *Fermented Beverage Production*, pp. 159–203. London: Blackie Academic and Professional.

Williams AA, Lewis MJ and May HV (1983) The volatile flavour components of commercial port wines. *Journal of Science and Food Agriculture* 34: 311–319.

Williams AA, Langron SP, Timberlake CF and Bakker J (1984) Effect of color on the assessment of ports. *Journal of Food Technology* 19: 659–671.

Postharvest Deterioration *See* **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage; Fungi in Food – An Overview; Molds in Spoilage; Yeasts in Spoilage

POTASSIUM

Contents

Properties and Determination

M A Amaro López and R Moreno Rojas, University of Córdoba, Córdoba, Spain

Background

Potassium is a light, soft metal and a strong reducing agent, discovered by Davy in 1807 in caustic potassium (KOH). It is widely distributed in minerals such as sylvin (KCl), carnallite ($KMgCl_3$ or $MgCL_2{\cdot}KCl$), langbeinite ($K_2Mg_2(SO_4)_3$) or polyhalite ($K_2Ca_2Mg_2(SO_4)_4{\cdot}2H_2O$). Potassium is a very abundant metal and is the seventh most abundant element in the earth's crust (2.59% corresponds to potassium in the combined form). Sea water contains 380 p.p.m. of potassium, making this mineral the sixth most abundant in solution.

Potassium is used extensively, mainly in the form of potassium salts, in various fields such as medicine (iodide as a disinfectant), photographic processing (carbonate), explosives (chlorate and nitrate as powder), defreezing agents, poisons (potassium cyanide), metallurgy and basic chemistry (hydroxide in, for example, strong bases, oil sweetening, CO_2 absorbent, chromate and dichromates as oxidants, carbonate in glass industry), fertilizers (nitrate, carbonate, chloride, bromide, sulfate), detergents and soaps (hydroxide and carbonate), coolants in nuclear reactors (NaK alloy), etc.

Potassium (atomic number = 19) forms part of the group of alkali metals classified between sodium (atomic number = 11) and rubidium (atomic number = 37). It has an atomic weight of 39.098, oxidation state 1 and an electronic orbital structure of [Ar] $4s^1$. This electronic configuration (one electron in the 4s orbital) enables potassium to ionize easily to the cation K^+ owing to the loss of its most external electron. It does not appear as a native metal in nature. Potassium exists in nature in three isotopes: ^{39}K (93.26%), ^{40}K (0.0117%) and ^{41}K (6.73%). ^{40}K is radioactive and responsible for most of the naturally occurring internal radioactivity in the body. This property enables researchers to monitor total body potassium values as a function of age and disease.

The behavior of potassium in its metallic form is very similar to that of sodium, with which it is closely related, both being alkali metals essential for life owing to their involvement in cellular physiological development and the regulation of body fluids, together with

chlorine, thus explaining why it is common for both alkali metals to appear in the form of their corresponding chloride salts. Given this relationship with cellular liquids, potassium, sodium, and chlorine are also known as electrolytes (ions with greater proportions in the composition of organic fluids).

The K^+ ion is the main intracellular cation, with an approximate $[K^+]$ of $5.6\,g l^{-1}$ in cellular fluids, approximately 30 times more concentrated than in plasma or interstitial liquid ($0.15–0.20\,g l^{-1}$). Its high intracellular concentration is regulated by the cell membrane through the sodium–potassium pump. The intracellular K^+ regulates the catalytic action of numerous enzymes, through the attachment of the cation to active locations with negative charges of the enzymatic proteins, modifying the conformation of the molecule and its activity, as well as participating in cellular division processes.

The small percentage of extracellular potassium (2% of body potassium) is of great physiological importance, since it is a critical determinant of neuromuscular excitability (nervous impulse and contraction of bone muscles). At the cellular membrane level, the transport and permeability of energy-dependent potassium, with simultaneous excretion of sodium linked to the Na/K enzyme ATPase, is essential for generating the potentials of membranes required for the proper functioning of nervous and muscular cells. It also helps to maintain the acid–base equilibrium and blood pressure.

Homeostasis of potassium is still the subject of research, although it is known that 90% of dietetic potassium is absorbed in the small intestine and that body potassium (1.6–2 g per kilogram of body weight) is regulated by renal glomerular filtration and tubular secretion, potassium being lost on a daily basis through urine, gastrointestinal secretions (ileum and colon) and, to a lesser degree, sweat. Provided that renal function is normal, it is practically impossible to reach an excessive level of potassium with normal dietary intake, since the kidney is capable of excreting more potassium than it can filter. (*See* **Potassium**: Physiology; **Sodium**: Physiology.)

The effect of potassium on blood pressure has been discussed in recent reports on metaanalysis; these have confirmed that increases in doses of potassium from 60 to 80 mmol per day (2.3–3.1 g per day) may prompt a decrease of 4 mmHg in systolic blood pressure and possibly reduce the number of deaths related to high blood pressure by 25%.

The use of potassium chloride salts as substitutes for sodium chloride, in individuals in whom the intake of sodium is restricted because of problems relating to hypertension, is a questionable alternative, since cardiac arrythmias have been reported in association with the excessive intake of salt substitutes containing potassium chloride. In this connection, hyperkalemia is recommended in healthy individuals for intakes exceeding 17.5 g per day (acute toxicity limit), highly unlikely in normal diets, and certain risks for individuals with renal dysfunctions not detected with intakes of potassium exceeding 5.9 g per day (safest maximum dose). Cardiovascular or neuromuscular complications arising from situations of hypo- and hyperkalemias are resolved favorably by correcting the plasmatic potassium levels. (*See* **Hypertension**: Hypertension and Diet.)

The mean potassium intake in Western populations ranges between 1.6–5.9 g per day and the required dietary intakes of this are met without any problem, thanks to the ubiquitous presence of potassium in both vegetable- and animal-based foods. The Scientific Committee for Food of the European Community recognizes that the Minimum Recommended Intake of potassium is 1600 mg d^{-1}, and the Reference Intake for the population 3100 g d^{-1}.

Analysis of Potassium

Potassium can be analyzed using a number of different methods, although many of these are not commonly used for routine analysis of potassium in food and biological samples. In the case of potassium, there has been a shift from the old gravimetric methods, based on the precipitation of potassium using chloroplatinate or tetraphenylboron, to the spectroscopic methods, mainly flame emission and flame absorbance spectroscopy and which, according to the scientific community, are the most commonly used.

However, there are other techniques for analyzing potassium that are not as 'popular' as those mentioned above, and equally applicable to other mineral elements. These include selective electrodes, nuclear magnetic resonance spectroscopy, X-ray analysis, helium glow photometry, inductively coupled plasma optical emission, inductively coupled plasma atomic fluorescence, ion-scattering spectrometry, and other methods. Another procedure can be used – namely, radioactive dilution of potassium isotopes – when determining both potassium content and its distribution in extra- and intracellular compartments.

Emission and absorbance spectrometry have been the most widely used techniques for analyzing trace elements in biological and food samples. Their widespread use is justified by their analytical specificity, good detection limits, excellent accuracy, and relatively low cost. Both techniques are based on energy modifications of the electronic orbital structure of the atoms of mineral elements in response to certain

stimuli (flame emission spectrometry and incidence of a beam of light of specific wavelength for absorption). (*See* **Sodium**: Properties and Determination.)

The determination of potassium by spectroscopy requires a stage of preparation of samples for analysis that entails the drying, grinding, and destruction of the organic matter of the sample. There are two alternatives for destroying the organic matter of biological and food samples, both with different variants: wet oxidation and dry ashing. Both techniques yield reasonably comparable results, and it is often the analyst who decides on which technique to employ. (*See* **Cadmium**: Properties and Determination; **Spectroscopy**: Atomic Emission and Absorption.)

Flame Emission Spectroscopy

The determination of potassium in biological and food samples, as in the case of sodium, may be performed using flame photometers and atomic absorption spectrophotometers (with the option of working in emission conditions). A specific wavelength of 766.5 nm is used in order to determine potassium, and the detector is adjusted to the response given to the pattern of greatest concentration used. Efforts must be made to avoid possible interference problems owing to the matrix characteristics of the sample in flame emission photometry; this can be achieved by separating the element from the object of study, eliminating the ions responsible for the interferences, or using a compensatory technique. For this purpose, recommended practice for correcting interferences includes the addition of lithium (as an internal standard) both to the samples and to the standards. It is important to obtain a suitable calibration curve, either linear by diluting the samples, or by logarithmic adjustment between emission values and potassium concentration. (*See* **Sodium**: Properties and Determination.)

Flame Absorption Spectroscopy

The determination of potassium using an atomic absorption spectrophotometer requires a light source (wavelength for $K = 766.5$ nm) and an atomization source (flame). For the light source, arc discharge lamps may be used for K, although hollow cathodes are preferred. Standard working conditions for analyzing potassium by atomic absorption spectroscopy are listed in Table 1.

Comparing flame emission photometry and atomic absorption spectroscopy, the detection limits for potassium ($\mu g\,ml^{-1}$) are 0.0005 and 0.005, respectively. Using an atomic absorption spectrophotometer, the accuracy and precision of potassium analysis by flame emission can be obtained (Table 2).

Table 1 Standard conditions for potassium analysis by atomic absorption spectrometry

Spectrometers setting	*Potassium*
Wavelength/slit (nm)	766.5/0.2
Nebulizer	Spoiler
Oxidant	Air
Fuel	C_2H_2
Flame condition	Oxidizing
Optimum concentration range in solution ($\mu g\,ml^{-1}$)	1–10
Detection limit ($\mu g\,ml^{-1}$)	0.005
Sensitivity 1% absorption ($\mu g\,ml^{-1}$)	0.05
Interferences	Ionization

Table 3 shows a series of official methods for determining potassium, based on flame emission and absorption spectrophotometry.

Inductively Coupled Plasma-atomic Emission Spectrometry (ICP-AES)

There are numerous references to the use of this powerful technique in multielement analysis by emission spectroscopy. ICP-AES enables samples with a high degree of variability and minimum interferences to be used. Samples can be prepared by either wet oxidation or dry ashing, and a wide range of concentrations can be used without the need to dilute or concentrate the sample. Also, many of the other mineral elements can be quantified, thus reducing analysis time. These advantages, together with the high speed and excellent instrument stability, make ICP-AES highly attractive.

For potassium determined by ICP-AES in plant tissues, a solution detection limit of 0.06 ($\mu g\,ml^{-1}$) and a sample detection limit of 2.0 ($\mu g\,g^{-1}$, based on a 2-g sample diluted to 50 ml) are recommended. There is also an AOAC Official Method (984.27) for the determination of potassium in infant formula by means of ICP-AES, which recommends the following ICP emission spectrometer operating parameters for potassium; wavelength (nm) = 766.5, no background correction and low standard = 0 and high standard = 200 ($\mu g\,ml^{-1}$). The AOAC also recommends an Official Method (985.01) for the analysis of potassium in plants and pet foods by ICP spectroscopy.

Potassium-selective Electrodes

Potassium-selective electrodes have developed from the first ion-selective glass electrodes, with little ion selectivity, to potassium-selective electrodes based on valinomycin, a highly specific neutral carrier compound for analyzing potassium, to potassium-sensitive electrodes with ion-exchange compounds dissolved in organic solvents that are not very specific

Table 2 Accuracy and precision of the potassium analysis by flame emission

Accuracy	*Certified (g kg^{-1})*	*Found (g kg^{-1})*	*IC (95%)*[a]	*rec (RSD)*[b]
Citrus leaves SRM-1572	18.2 ± 0.6	17.7 ± 0.5	16.8–18.7	97 (3)
Nonfat milk powder NIST-1549	16.9 ± 0.30	16.3 ± 0.41	15.5–17.1	96 (5)
Precision	RSD			
Plant food	1.35			
Dairy food	2.55			

[a]Intervals of confidence (95%).
[b]Recovery percentages and relative standard deviation (RSD).

Table 3 Official methods recognized for potassium determination by flame emission and absorption spectrophotometry

Official methods	
Sodium and Potassium in Seafood	Flame Photometric Method, AOAC Official Method 969.23
Potassium in Fruits and Fruit Products	Rapid Flame Photometric Method, AOAC Official Method 965.30
Potassium and Sodium in Wines	Flame Spectrophotometric Method, AOAC Official Method 963.13
Potassium in Distilled Liquors	Flame Photometric Method, AOAC Official Method 963.08
Potassium in Beer	Atomic Absorption Spectrophotometric Method, AOAC Official Method 987.02
Potassium in Water	Atomic Absorption Spectrophotometric Method, AOAC Official Method 973.53
Potassium in Infant Formula, Enteral Products and Pet Foods	Atomic Absorption Spectrophotometric Method, AOAC Official Method 985.35

because of significant interference induced by the presence of sodium and other cations in organic solutions. Ion-exchange electrodes have faster response times than those of organic solvent solution membranes of valinomycin, although the latter show fewer interferences from other ions. However, neither type of electrode is effective in organic solvents. Furthermore, electrodes measure ion activity and therefore do not detect ions linked to other molecules or potassium that is not in solution.

Ion-selective electrodes are also incorporated into gasometric equipment for measuring electrolytes (Na, K, Cl, and ionized Ca), in oxygenation, ventilation, basic-acidic state, and electrolytic metabolism of patients with respiratory problems. In this context, it is useful to quantify the exchange of gases and electrolytes through cellular membranes; for this purpose, the so-called *anion gap* is calculated, which interrelates with sodium (*anion gap Na*), potassium (*anion gap K*), bicarbonate, and chlorine. This parameter provides information on electrolytic alterations and the presence of toxins in blood, and also helps to control quality in laboratories by interrelating informed values (in normal individuals, not very high values).

Some developments in the field of potassium-selective electrodes have focused on the appearance of novel potassium-selective valinomycins in the development of ion-selective sensors based on new technologies and on the combination of these ion-selective electrodes with other analytical methodologies, as shown in Table 4.

Nuclear Magnetic Resonance (NMR)

NMR is used to study the relationship between metabolism and function in living systems as simple as cells in culture and as complex as human subjects (e.g., the effect of temperature on the sodium/potassium pump in red blood cells). As a noninvasive, nondestructive spectroscopic technique, NMR offers a powerful approach to the study of ion balance in intact biological systems. The aims are to validate new NMR methods, to elucidate tissue-specific mechanisms, and to draw conclusions with respect to basic concepts and design in tissue metabolism and function.

The possibility of determining potassium by NMR is due, on the one hand, to the relatively high concentrations of the cation in cellular tissue and, on the other hand, to the natural abundance of ^{39}K nuclide (NMR sensitivity is about 2000 times less than that of one hydrogen atom) and the short longitudinal relaxation time. The concentrations of intracellular and extracellular potassium in tissues may be quantified by applying impermeable cell-membrane chemical shift reagents.

However, this quantification by NMR measurement depends on the NMR invisibility of the potassium actually present, in accordance with other techniques. Several previous studies have shown that only approximately 20% of cardiac intracellular potassium is visible with current NMR techniques. This NMR sensitivity of potassium (about 60%) represents a serious obstacle to its proper quantification in biological tissues. (*See* **Spectroscopy**: Nuclear Magnetic Resonance.)

Table 4 Analytical methodologies with ion-selective electrodes for potassium analysis

Investigation	*Reference*
A potassium-ion selective electrode with valinomycin-based poly(vinyl chloride) membrane and a poly(vinyl ferrocene) solid contact	Hauser PC, Chiang DWL. and Wright GA (1995) *Analytica Chimica Acta* 302(2–3): 241–248.
Miniaturized ion-selective sensor chip for potassium measurement in a biomedical application	Uhlig A, Dietrich F, Schnakenberg U, Hintsche R and Lindner E (1996) *Sensors and Actuators B: Chemical Abstract Export* 34(1–3): 252–257.
Synthesis of novel potassium selective valinomycins	Dawson JR, Dory YL, Mellor JM and McAleer JF *et al*. (1996) *Tetrahedron* 52(4): 1361–1378.
Optical sensors for sodium, potassium, and ammonium ions based on lipophilic fluorescein anionic dye and neutral carriers	Wang E, Zhu L, Ma L and Patel H (1997) *Analytica Chimica Acta* 357: 85–90.
Recording of neuronal network properties with near-infrared dark-field microscopy and microelectrodes	Holthoff K and Witte OW (1997) *Electrochimica Acta* 42(20–22): 3241–3246.
Flow injection analysis of potassium using an all-solid-state potassium-selective electrode as a detector	Komaba S, Arakawa J, Seyama M, Osaka T, Satoh I and Nakamura S (1998) *Talanta* 46: 1293–1297.
Transferability of results obtained for sodium, potassium and chloride ions with different analyzers	Rodriguez-Garcia J, Sogo T, Otero S and Paz JM (1998) *Clinica Chimica Acta* 275: 151–162.
Plasticizer-free all-solid-state potassium-selective electrode based on poly(3-octylthiophene) and valinomycin	Bobacka J, Ivaska A and Lewenstam A (1999) *Analytica Chimica Acta* 385: 195–202.
Simultaneous detection of monovalent anions and cations using all solid-state contact PVC membrane anion and cation-selective electrodes as detectors in single column ion chromatography	Isildak I and Asan A (1999) *Talanta* 48: 967–978.
Novel sensors for potassium, calcium, and magnesium ions based on a silicon transducer as a light-addressable potentiometric sensor	Seki A, Motoya K, Watanabe S and Kubo I (1999) *Analytica Chimica Acta* 382: 131–136.
Reference ranges of electrolyte and anion gap on the Beckman E4A, Beckman Synchron CX5, Nova CRT, and Nova Stat Profile Ultra	Lolekha PH, Vanavanan S, Teerakarnjana N and Chaichanajarernkul U (2001) *Clinica Chimica Acta* 307: 87–93.

A study of factors affecting ^{39}K NMR detectability in rat thigh muscle showed that the signal may be substantially higher (up to 100% of total tissue potassium) than values previously reported of around 40%, these signals presenting two superimposed components – one broad and one narrow – and involving improvements in spectral parameters (signal-to-noise ratio and baseline roll), together with computer simulations of spectra that enable a spectrum quality with a major effect on the amount of signal detected, largely owing to the loss of detectability of the broad signal component.

Biological and biochemical cellular research in connection with the quantification of ions in intracellular and extracellular compartments, transport and ionic metabolism at the level of different cells and biological tissues, are based on the use of ^{39}K NMR spectroscopy, together with ^{23}Na and ^{31}P NMR studies.

Proper ion equilibrium between intra- and extracellular compartments is required for normal physiological function. Conversely, alterations in membrane ion transport occur in numerous pathological states. By introducing an anionic paramagnetic shift reagent into the medium, NMR signals of intra- and extracellular Na^+ and K^+ can be resolved, enabling ion transport processes to be studied by NMR. Unfortunately, rare NMR active nuclides that are isotopes of the 100% naturally abundant $^{23}Na^+$ and $^{39}K^+$ are not available for tracer kinetic studies of Na^+ and K^+ transport. However, Cs is a biologically active analog of K^+, and the 100% naturally abundant NMR active $^{133}Cs^+$ nuclide can be employed to examine K^+ transport. Other studies have shown the potential of ^{39}K NMR as a useful tool in the study of protein–cation interactions and the binding of K^+ to double-helical DNA.

X-ray Analysis

X-ray analysis is another nondestructive, noninvasive, *in-vivo* technique for determining mineral elements – including potassium – in foods and biological samples. X-ray analysis is based on bombarding a small area of the powdered sample with high-energy X-rays, measuring the Kα line of the element using one of several different detectors according to the wavelength of the emitted radiation. The limitations of this technique are related to scope, in terms of detected elements and their concentration range. X-ray analysis is combined with the use of an electronic microscope for quantifying and locating the mineral element in the sample, even at the subnuclear compartmental level. Another limitation of the X-ray

technique relates to the complexity of the sample preparation process for electronic microscopy.

The bibliography describes different X-ray analysis techniques for measuring potassium in food and biological samples, as indicated in Table 5.

X-ray fluorescence (XRF) can be successfully used for the qualitative and quantitative elemental analysis of various agricultural products. Its simplicity, high throughput and automation possibilities make it useful for screening large numbers of samples. The K content of tea samples has been determined by XRF analysis (an uncommon method for mineral analysis in food), and the findings have been compared with the results obtained using atomic emission techniques, with the conclusion that the XRF system can be used effectively for quantitative analysis of the K content of tea samples.

The methods for preparing the specimens of liquid media of the organism and XRF analysis are simple and fast, entail no disintegration of the sample, and allow measurements of elements at sigma $1 = 0.02$ at concentrations of $12\,\mu g\,ml^{-1}$ to be obtained. The method is preferable and promising for several basic elements (P, S, Cl, K, Ca), which are difficult to measure using other methods.

The use of total reflection X-ray fluorescence analysis (TXRF) in life sciences is considered a powerful analytical tool for simultaneous multielement determination. TXRF is basically an energy-dispersive technique, with the sample being excited in total reflection geometry. This technique only requires minute samples with simple preparation and involves a large dynamic measuring scale. Simultaneous detection of almost all chemical elements and lower limits of detection are achievable in optimized excitation conditions. The preferred sample types are aqueous and acidic solutions, particularly samples digested with HNO_3. Special preparation techniques are required for solids and other samples.

Table 5 X-ray techniques for potassium analysis in food and biological samples

Samples	*X-ray analysis*
Meat and meat products	X-ray fluorescent analysis
Serum and peripheral blood cells	
Fresh green tea, black tea, and tea residue	
Leaves of lettuce	Electron microprobe X-ray analysis
Walls of the mosaic virus-infected wheat leaf cells	Energy-dispersive X-ray analysis
Perisperm tissues of seeds	
Cancerous and normal tissues	Total reflection X-ray fluorescence analysis
In vegetable foodstuffs and their respective cell fractions	
Cancerous breast tissue	
Human brains	Proton-induced X-ray emission analysis
Human atherosclerotic artery wall	

The application of the proton-induced X-ray-emission (PIXE) method has enabled us to determine not only the concentrations of elemental composition, including K, but also their localization in different artery-wall regions. Furthermore, the usefulness of the micro-PIXE method for studying biomedical materials has also been considered.

Radioactive Isotopes

The dilution of radioactive isotopes of potassium is applied to establish the concentration and distribution of the amounts of potassium in the different sample compartments. The dilution of radioactive isotopes enables the volumes of the organic tissue compartments to be determined. The differences in potassium concentrations between the extra- and intracellular spaces may be studied by diluting the radioactive isotopes of potassium.

For this purpose, it is important to take into account that animal and plant tissues present different potassium compartments in which the ion is accumulated, and this conditions the radioactive isotope dilution process. In animal tissue, three cellular compartments (plasma + extracellular fluid, cells, and spaces where ions are closely fixed or form complexes) are described, whereas in plant tissue, only two spaces (cytoplasm and vacuoles) must be considered, since the extracellular space is small in size. This compartmental distribution conditions the process of radioactive isotope dilution.

In animal tissue, the method involves incorporating a known amount of the tracer isotope of potassium into the extracellular compartment and monitoring the changes in plasma isotope concentration until equilibrium is reached with all the compartments. This entails estimating the distribution of potassium (1) in extracellular fluid, using tracer polymers (inulin or polyethylene glycols) that cannot penetrate cell membranes, (2) in cellular volume, by subtracting the extracellular volume from the total distribution space (determined using labeled water, tritiated or deuterated), and (3) in the space where the ion is strongly fixed or forms complexes (this is calculated by subtracting the amount of extracellular potassium + cellular potassium from the amount of potassium isotope injected).

In plant tissue, the dilution of radioactive isotopes of potassium presents limitations owing to the difficulty involved in terms of separating vacuolar volume and cytoplasmic volume. For this reason, more appropriate methods than radioactive dilution methods have been

developed for determining cytoplasmic and vacuolar potassium concentrations, namely ion-selective microelectrodes or X-ray microprobe analysis.

See also: **Cadmium**: Properties and Determination; **Hypertension**: Hypertension and Diet; **Potassium**: Physiology; **Sodium**: Properties and Determination; Physiology; **Spectroscopy**: Atomic Emission and Absorption; Nuclear Magnetic Resonance

Further Reading

Ammon D (1986) *Ion Selective Microelectrodes*. New York: Springer-Verlag.

Benton Jones J, Jr. (1984) Developments in the measurement of trace metal constituents in foods. In: Gilbert J (ed.) *Analysis of Food Contaminants*, pp. 157–202, Barking, UK: Elsevier.

Cappuccio FP and MacGregor GA (1991) Does potassium supplementation lower blood pressure? A meta-analysis of published trials. *Journal of Hypertension* 9: 465–473.

Commission of the European Communities (1993) *Nutrient and Energy Intakes for the European Community. Reports of the Scientific Committee for Food, Thirty-first Series*, pp. 174–178.

Gunther K and von Bohlen A (1990) Simultaneous multielement determination in vegetable foodstuffs and their respective cell fractions by total-reflection X-ray fluorescence (TXRF). *Zeitschrift für Lebensmitteluntersuchung und -forschung* 190(4): 331–335.

Intersalt Cooperative Research Group (1988) Intersalt: a international study of electrolyte excretion and blood pressure. Results for 24 hour urinary sodium and potassium excretion. *British Medical Journal* 297: 319–328.

Luft FC (1990) Sodium, chloride and potassium. In: Brown M (ed.) *Present Knowledge in Nutrition*, 6th edn, pp. 233–240. Washington, DC: International Life Sciences Institute Nutrition Foundation.

Springer C (1987) Measurement of metal cation compartmentalization in tissue by high-resolution metal cation NMR. *Annual Review of Biophysics and Biochemistry* 375–399.

Physiology

M P Navarro, Unidad de Nutrición, Estación Experimental del Zaidin, CSIS Granada, Spain
M P Vaquero, Ciudad Universitaria, Madrid, Spain

Introduction

Potassium is the most abundant intracellular cation in the human body and plays an important role in a variety of cell functions. This review summarizes the main aspects of potassium physiology and nutrition, including absorption, transport, distribution, storage, and excretion, as well as homeostatic mechanisms of K balance.

Role in the Body

Potassium is essential for the muscular, cardiovascular, nervous, endocrine, respiratory, digestive, and renal systems. Cell metabolism depends on the maintenance of a high intracellular K^+ concentration. Many of the body functions of potassium are due to its ionic character: it generates gradients of concentration, potential, and pressure. Potassium is the predominant osmotically active species inside the cell. Together with other ions such as sodium and chloride, which are characteristic of the extracellular fluid, potassium determines osmolarity and plays a major role in the distribution of fluids inside and outside the cell and hence in the maintenance of cellular volume. In addition, potassium participates in the regulation of the acid–base balance and is involved in cellular growth and division, energy transduction, glycogenesis, protein synthesis, hormone secretion, etc. Consequently, its deficiency causes growth retardation, with a pronounced decrease in circulating levels of growth hormone and somatomedin C and inhibition of protein synthesis.

Cell excitability and muscle contraction depend on potassium. The transmembrane electrical potential is determined by the ratio of the intracellular to extracellular potassium and sodium concentrations, in particular that of potassium. Differences in K^+ and Na^+ concentrations across cell membranes are maintained by the specific permeability to each of these ions and by K^+/Na^+-ATPase activity (K^+/Na^+-pump). Thus, K^+ is critical for the excitability of nerve and muscle cells.

During vigorous exercise, potassium is released from muscle cells, leading to an increase in extracellular potassium concentrations which facilitates ongoing muscle contraction and induces vasodilatation, increasing local blood flow. However, liberation of potassium also leads to muscular fatigue. Training reduces the exercise-induced rise in plasma K^+ concentration and also increases the total activity of Na-K pumps in muscle. The potassium internal balance helps to delay the onset of fatigue during exercise and to restore homeostasis during recovery.

Potassium can be defined as a cardioprotector nutrient. It acts directly on the heart, regulating its mechanical and electrical properties. Epidemiological data suggest that potassium intake and blood pressure are inversely correlated. The greatest hypotensive

effect of potassium occurs in hypertensive patients and in subjects with a high sodium intake. The mechanisms involved include: enhancing natriuresis, baroreflex sensitivity, direct vasodilatation, catecholaminergic functions, improvement of glucose tolerance, and effects on the central nervous and the renin–angiotensin–aldosterone systems.

Potassium also protects against stroke, which is the third leading cause of death worldwide (after coronary artery disease and cancer). This electrolyte may decrease ischemic as well as hemorrhagic stroke risk through its effect on blood pressure on one hand, and by inhibiting the formation of free radicals at the endothelial cell level, thus affecting vasomotion, on the other hand. Moreover, potassium inhibits the proliferation of smooth muscle cells, platelet aggregation, and arterial thrombosis. Potassium may also prevent the death of cerebellar neurons.

Interrelationships exist between potassium and other nutrients. Potassium and sodium are strongly metabolically interrelated, principally due to K^+/Na^+-ATPase. This enzyme also provides the driving force for the transport of other solutes, such as amino acids, phosphate, vitamins, and glucose. Potassium depletion, which is more intense if there is a simultaneous excess of sodium, enhances urinary loss of calcium. This interaction may have adverse effects on bone and blood pressure.

Requirements and Daily Intakes

To date, no recommended potassium intake has been unanimously established. The UK report on dietary reference values established a reference nutrient intake of 3500 mg day^{-1} for adults. Lower amounts were recommended for infants (mg day^{-1}): 800 and 850, 0–3 months and 4–6 months, respectively; 700, 7–12 months; 800, 1–3 years; 1100, 4–6 years; 2000, 7–10 years; and 3100, 11–14 years of age. No specific levels were recommended during pregnancy, lactation, or for the elderly. The Nutrition Working Group of the European International Life Sciences Institute (ILSI) suggested a higher recommended daily intake (3900 mg). This recommendation was aimed at promoting a potassium intake similar to that of sodium, in molar terms, i.e., an intake of 100 mmol day^{-1} of each cation (2300 mg and 3900 mg for Na and K, respectively). However, more recently a Na:K ratio below 1 (1600 mg and 3500 mg for Na and K, respectively) has been recommended to prevent hypertension, stroke, and cardiovascular disease.

During exercise potassium needs may be increased owing to higher losses in sweat, especially in hot climates and in unaccustomed individuals. These needs should be satisfied by eating food rich in potassium and exceptionally by potassium supplements.

Published values of nutrient consumption demonstrate that in some populations, sodium intake is higher than recommended, while potassium intake should increase. In the UK, the 1991 report of the National Food Survey Committee showed that the average sodium intake was 170% of the reference intake, while that of potassium was 80% of the reference value. In Spain, data from the same year revealed that the estimated potassium intake was satisfactory but that of sodium was 140% above the reference value.

Potassium is ubiquitous in all kinds of food, but the best sources are vegetables and fruits because they combine a high potassium content with low sodium concentration (Table 1). Potassium is an essential element of all forms of life, whereas sodium and

Table 1 Potassium and sodium content of various foods (mg per 100 g of edible portion)

	Potassium	*Sodium*
Vegetables		
Potatoes	320	11
Cabbage	275–310	7–12
Carrots	235	60–90
Spinach	500–633	65–140
Legumes		
Red kidney beans	1370	18
Soya beans	1730	5
Lentils	940	12
Fruits		
Banana	400	1
Melon	100–310	5–32
Avocado pear	400–600	2–4
Orange	195	2–3
Apple	120	2–3
Dried fruits		
Raisins	1020	60
Figs	970	62
Almonds	780	14
Walnuts	450	7
Cereals		
White bread	100	460–540
Brown bread	175–240	540–636
Rice	110	6
Meat and fish		
Beef, veal, lamb	230–360	52–110
Chicken	320	81
Bacon	350	1860
Herring	320	120
Halibut	410	60
Tuna	400	47
Mussels	320	290
Milk		
Whole cows' milk	140	55
Various		
Chocolate	300	11
Tomato ketchup	590	1120

chlorine are essential for animals but not for many plants. Meat and fish contain important amounts of potassium because they are highly cellular tissues, but they also contain large amounts of sodium chloride, due to the extracellular fluid present in animal food. Raw foods are preferable to processed ones because during food preparation salt is usually added to enhance flavor and retard bacterial growth (e.g., bread, bacon, tomato ketchup, and cookies). Potassium salts, however, are not palatable and are not added to food during processing.

Legumes represent a good source of potassium because they provide at least 1 g potassium per 100 g portion and very little sodium. Dried fruits, for example, apricots, figs, and prunes, should also be mentioned, since they often exceed 0.5 and even 1 g potassium per 100 g portion. Nuts such as pistachios and almonds contain a considerable quantity of potassium, but the addition of salt during roasting upsets the ideal sodium-to-potassium ratio normally present in these foods. Bananas are known to be one of the fruits richest in potassium, together with coconuts, melons, avocados, and kiwi fruit. Milk is the largest dietary source of potassium for infants.

All these data should be considered, taking into account the amount of each food item consumed. For example, a ration of legumes is approximately 80 g while those of meat and milk can easily reach 200 g. Although cereals are low in potassium, they account for almost 75% of potassium ingestion in western diets. There is a wide range of potassium intake by western populations (1560–5850 mg). It is higher in vegetarians and people who consume more food from plant than animal sources. The Mediterranean diet provides a mean daily intake of potassium equivalent to the recommended value (3500 mg).

Although thermal treatments have no effect on dietary potassium, during processes such as boiling, soaking, or canning it can be reduced as a result of leaching, particularly if food is to be cut into pieces and the liquid is not consumed.

In order to protect against hypertension and stroke, except in clinical hypokalemia, potassium from good dietary sources should be encouraged rather than potassium from supplements, because such supplements may have severe adverse effects. Fruits and vegetables also contain magnesium and fiber which seem to have protective effects.

Absorption

Normal subjects absorb about 90% of ingested potassium in the intestinal tract. The net absorption is the difference between fluxes from lumen to blood and from blood to lumen. In the human small intestine, K^+ permeability is high and potassium absorption is carried out across the epithelium of duodenum, jejunum, and ileum by passive mechanisms in response to electrochemical gradients and solvent drag. In the proximal small bowel K^+ is concentrated through the absorption of water, providing a driving force for the movement of this cation across the intestinal mucosa, preferably through the tight junctions between enterocytes. Duodenum and jejunum absorb this ion even more rapidly than water. Indeed, shortly after a meal, the K^+ concentration [K^+] in jejunum rapidly reaches plasma levels. In the ileum, the transepithelial electrical potential difference strongly influences its movement. There is no evidence of active potassium absorption in the small intestine, but the existence of an apical membrane H^+/K^+-ATPase could suggest active K^+ transport.

In the colon K^+ may be secreted or absorbed in response to variations in potassium status: net secretion occurs when the luminal concentration is less than 20–25 mmol l^{-1}, while net K^+ absorption takes place when levels are above 25 mmol l^{-1}. There are two mechanisms for K^+ secretion; the major mechanism involves potential-dependent, passive flux, mostly via the tight union, or by facilitated transcellular diffusion. The second is an active serosa-to-mucosa secretory mechanism in the proximal and distal colon which is the result of uptake across the basolateral membrane mediated by Na^+/K^+-ATPase and Na^+-K^+-Cl^- cotransport and movement across the apical membrane through potassium channels. Active K^+ absorption also occurs in the distal colon, in which the K^+ uptake appears to be an exchange for H^+ across the apical membrane. This process is energized by adenosine triphosphatases (ATPases) located in that membrane. Passive potassium absorption may also take place in the proximal colon.

Bioavailability

The availability of dietary potassium is high because potassium salts are entirely soluble and few dietary factors modify its digestive utilization. Olive oil favors potassium uptake, while some fibers and certain ion exchange resins may decrease K^+ absorption.

Physiological status also affects the nutritive utilization of potassium. Infants absorb a greater proportion of salts than adults. This may be due to the increased permeability of the immature small intestinal epithelium of newborns, and especially to a higher activity of the K^+-absorptive pumps in the colonic apical membrane, while K^+-secreting channels are less relevant. During the second half of pregnancy, fecal potassium decreases. Urinary K^+ output increases in parallel with Na^+ retention; nevertheless,

enough potassium is retained to cover gestational needs. Glomerular filtration rates fall with age, accompanied by limitations of K^+ secretion and Na^+ conservation. Many drugs used in the elderly alter potassium homeostasis.

Transport

Potassium, being an electrolyte, is transported mainly under ionic form in the body fluids. Intracellular potassium concentration is 110–160 mmol l^{-1} of cells while plasma concentration averages 4–5 mmol l^{-1}, of which only 10–20% is bound to proteins.

Figure 1 schematically illustrates the intracellular and paracellular mechanisms of potassium transport. Different potassium transport mechanisms can operate simultaneously. They are usually interrelated and are also linked to other ion transport systems. Each name, except the last one, corresponds to an entire family of transporters.

Na^+/K^+-ATPase activity is responsible for the maintenance of the extra- and intracellular Na^+ and K^+ concentrations against electrochemical gradients. This ATPase is found in the plasma membrane of virtually all animal cells and represents an enormous metabolic energy cost which increases the entropy of the system. It is a carrier protein that pumps $2K^+$ in and $3Na^+$ out of the cell during each cycle of conformational changes driven by adenosine triphosphate (ATP) hydrolysis. This process is electrogenic, meaning that one net positive charge is removed from the cell every pump cycle. Many hormones and various neurotransmitters can modulate K^+/Na^+-ATPase activity.

Another active K^+ transport is controlled by the H^+/K^+-ATPase which ejects H^+ in exchange for K^+. This pump has an important role in some gastrointestinal cells and in the renal tubules.

Other potassium transport systems are driven by the force of ion gradients. The free energy gained during the movement of an inorganic ion down an electrochemical gradient (sum of the concentration gradient and the electrical potential difference) is used as the driving force to pump other solutes against their electrochemical gradient. Thus, the carrier protein acts as a coupled transporter and electroneutrality is maintained. Cotransport depends on ATP only indirectly. Various Na^+-K^+-Cl^- cotransporters, which carry $1Na^+$, $1K^+$ and $2Cl^-$ inside the cell, have been identified in salivary glands, gastrointestinal tract, and renal tubules. The K^+-Cl^- cotransporter, related to the former ones, plays an important role in erythrocytic volume. The first cotransport mediates K^+ influx while the second mediates efflux.

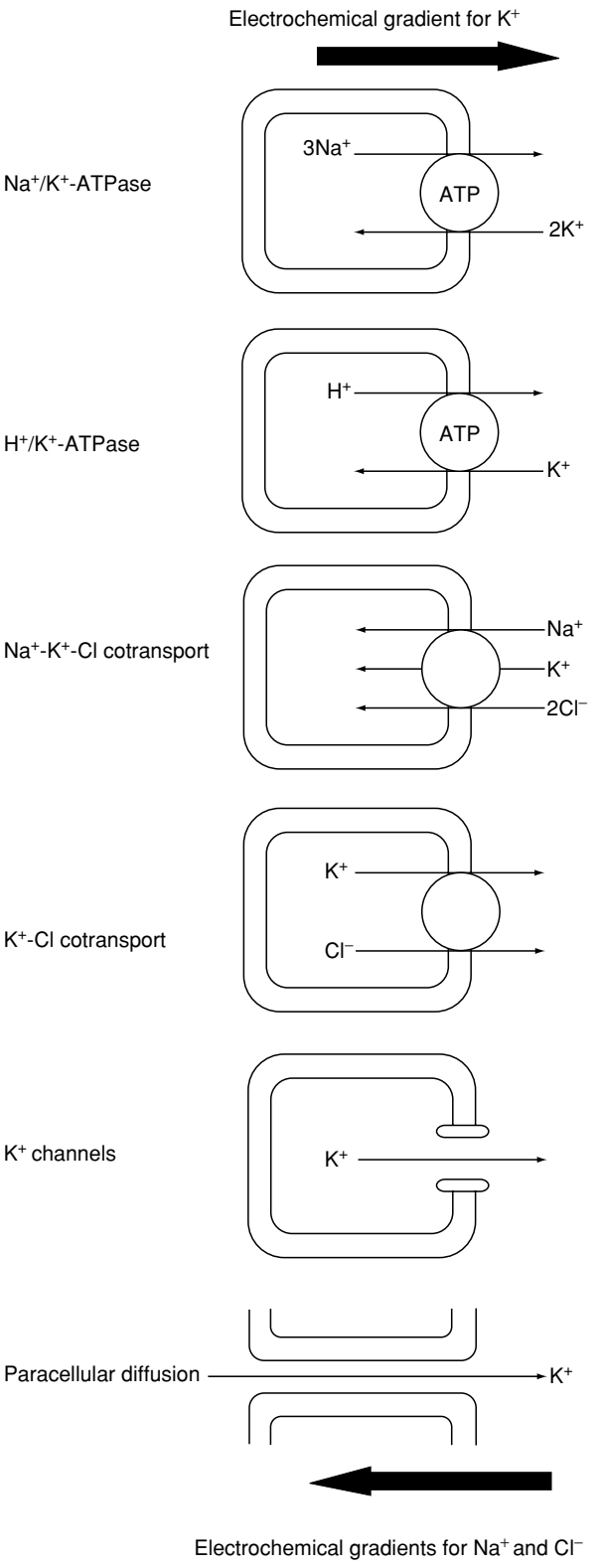

Figure 1 Potassium transport via intracellular and paracellular pathways. ATP, adenosine triphosphate.

Passive transport of K^+ occurs via intracellular and paracellular pathways. The intracellular mechanism involves K^+ channels, which are pores in the plasma membrane made by a specific protein. Channels have 'gates' which open and close in response to specific stimuli: voltage, stretch, ATP, Ca^{2+}, hormones, neurotransmitters, etc. Various stimuli sometimes act together on a channel. K^+ channels exhibit great diversity and participate in many cell functions.

Distribution and Storage

Body potassium concentration is 45–55 mmol kg^{-1} of body weight, thus a 70-kg adult man contains approximately 135 g potassium. Women contain 35–40 mmol kg^{-1} of body weight, and during childhood and in the elderly, body potassium concentration is also lower than that for young men. These differences are due to the lower muscle mass in these groups. Almost 98% of the potassium in the body is inside the cells, making it the major intracellular cation. More than 95% of total potassium is exchangeable.

Total potassium is an index of lean body mass because potassium is only present in the fat-free compartments of the body. It can be measured with K^{40}, a natural isotope present as a small fraction of the total potassium, which emits γ-rays and can be detected by a sensitive whole-body counter, although other isotopes and other techniques that are not based on total potassium are also available.

Studies of body composition in healthy humans have demonstrated that the total body potassium/total body nitrogen ratio is constant. However, in chronic protein-energy malnutrition this ratio can decrease owing to the reduction of muscle proteins and loss of intracellular potassium. The cellular exchange of sodium and potassium is altered, leading to potassium loss and increased intracellular sodium and water. In contrast, exercise and training increase total lean body mass, thereby causing a rise in potassium body content. Increase in muscular mass is only possible if proteins, as well as potassium, phosphorus and magnesium, which are the main intracellular elements, are present in the diet in adequate amounts.

In the extracellular fluid (ECF) an increases of potassium equivalent to 1% of total body potassium may double the concentration of potassium in plasma, resulting in muscle hypopolarization. However, if such an increase is stored in the cells, only minor intracellular changes would result, with little change to the potassium concentration difference across the cell membranes. To accomplish this buffer function, the major reservoir of potassium is the muscle, followed by the liver and erythrocytes, although each cell possesses the capacity for accumulating potassium.

Excretion

About 90% of daily potassium intake is excreted in the urine and 10% in the stool, although this last percentage can be somewhat higher in cases of diarrhea. When dietary potassium is severely restricted, its fecal loss decreases to approximately 3.5 mmol day^{-1}. This presumably represents obligatory potassium losses related to K^+ digestive secretions (salivary, gastric, biliary, and pancreatic), cell desquamation, and mucus secretion. Potassium concentration in sweat is about 10 mmol l^{-1}, so loss through perspiration is small, provided that the climate and exercise conditions are not extreme.

Renal Potassium Handling

The kidney plays the main role in potassium excretion. Renal potassium excretion shows a circadian rhythm, characterized by peak output during the subject's activity.

K^+ is freely filtered by the glomerulus. Usually, urinary potassium excretion is 5–15% of the amount filtered, which indicates the existence of tubular potassium reabsorption. The renal tubules are capable of reabsorbing and secreting potassium in response to various stimuli.

Proximal tubule The proximal tubule is responsible for the reabsorption of approximately 60% of previously filtered potassium. Solvent drag and diffusion have been proposed as the major driving forces of this process, and K^+ is reabsorbed largely via the paracellular pathway. However, some observations suggest participation of a transcellular route of potassium reabsorption by an active mechanism. The fractional rates of K^+ reabsorption are similar to those of sodium and water along the proximal tubule and changes in fluid and potassium transport are closely coupled.

In this region the basolateral K^+ channels have higher conductance than the apical ones, and this is essential to maintain a negative intracellular potential and hence promote reabsorption of positively charged carriers (Na^+-glucose cotransport, etc.)

Loop of Henle The concentration of K^+ increases as the filtrate passes through the loop of Henle, where permeability for potassium is very high. At the end of the descending limb the amount of K^+ present usually exceeds that of the glomerular filtrate. There is a net passive secretory entry of potassium into the proximal straight tubule and the descending thin limb of Henle, which arises from reabsorption in the collecting tubule and partly in the ascending limb. These pathways of K^+ transport constitute potassium recycling in the renal medulla.

In the thick ascending limb potassium is mainly reabsorbed. The mechanism of active transport includes apical uptake by Na^+/K^+-Cl^- cotransport and passive paracellular exit across the basolateral membrane by diffusion. Driven by transepithelial potential difference, reabsorption also occurs through the paracellular pathway.

The role of this limb in potassium secretion has also been recognized. The K^+ secretion through apical K^+ channels is important for net Na^+ and Cl^- reabsorption through Na^+ /K^+-Cl^- cotransport, since luminal Na^+ and Cl^- concentrations are higher than luminal K^+ levels.

Distal tubule and collecting duct These tubules are the major determinants of urinary K^+, since they are able either to reabsorb or secrete potassium at rates which depend on K^+ intake and other factors.

Potassium secretion occurs in 'principal cells' by active uptake across the basolateral membrane by Na^+/K^+-ATPase and passive diffusion into the lumen across the apical membrane by K^+ channels or using a K^+-Cl^- cotransport (Figure 2). K^+ transport follows this route towards the lumen thanks to a favorable electrochemical gradient and the greater K^+ permeability of the apical membrane in comparison with the basolateral one. High Na^+ concentration in the lumen favors K^+ secretion, as the entrance of Na^+ into the cell generates a potential difference that tends to make the interior positive, so K^+ diffuses in the opposite direction. Simultaneously, the enhanced delivery of Na^+ into principal cells facilitates the action of Na^+/K^+-ATPase, and the accelerated potassium uptake favors its secretion towards the lumen.

Potassium reabsorption mechanisms have been detected in the intercalated cells of the distal tubule and collecting duct. It appears that there is a component of potassium reabsorption even during net secretion. Several isoforms of H^+/K^+-ATPases are able to reabsorb K^+ in exchange for H^+. The Na^+-K^+-Cl^- contransport may also introduce K^+ into the cell while the basolateral K^+ channels enable potassium to pass from the cell into the interstitial space. Proton secretion by H^+/K^+-ATPase, together with that by H^+-ATPase, contributes to urine acidification. These enzymes participate in the acid–base balance and potassium status.

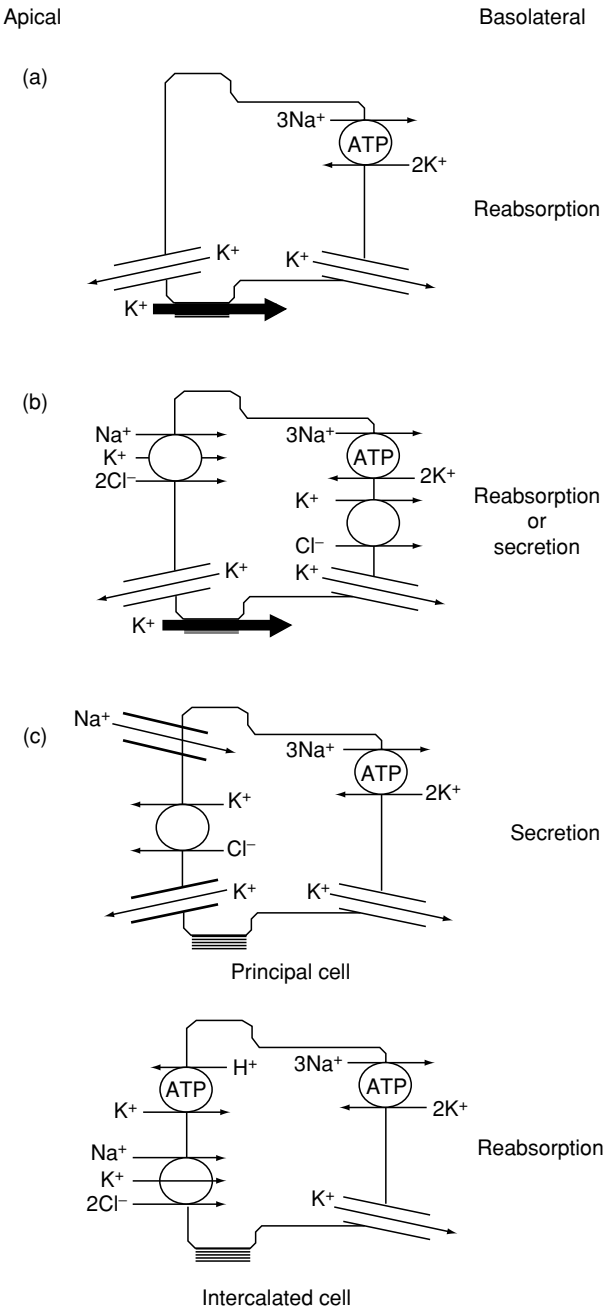

Figure 2 Principal mechanisms of potassium transport in the major tubular segments of the nephron: (a) proximal tubule; (b) thick ascending limb; (c) distal tubule and collecting duct. ATP, adenosine triphosphate.

Potassium Balance and Homeostasis

Intra- and extracellular potassium concentrations must be maintained within narrow limits, despite wide fluctuations in dietary intake. This is achieved, on the one hand, by 'internal balance': distribution of K^+ between the intracellular fluid and ECF carried out mainly by Na^+ K^+-ATPase, and on the other hand, by 'external balance': maintenance of the amount of K^+ in the body through equilibrium between K^+ intake and excretion, which is achieved principally in the kidney but also in the intestine. After a meal, the absorbed K^+ rapidly enters the ECF. The subsequent rise in plasma [K^+] is attenuated by a rapid cellular K^+ uptake (which occurs within minutes). To maintain external K^+ balance, all

of the K^+ absorbed in excess must be excreted slowly by the kidneys.

Potassium Homeostasis

To maintain the equilibrium previously described, several factors which act at different levels contribute to potassium homeostasis (Figure 3).

Internal balance A rise in plasma [K^+], especially with hyperkalemia, and the hormones insulin, epinephrine (adrenaline: by activating β_2-receptors), and aldosterone promote K^+ uptake by muscle, livers, bone, and red blood cells. In contrast, a decrease in plasma [K^+], such as with hypokalemia, and stimulation of α-adrenergic receptors induce K^+ transport from cells to the ECF. Moreover, hyperkalemia stimulates insulin, aldosterone, and epinephrine secretions, while hypokalemia has the opposite effect. These are the major physiological factors involved in internal K^+ balance homeostasis.

Furthermore, other factors which are not normal homeostatic mechanisms influence K^+ movements across the plasma membrane: metabolic acidosis promotes exit of K^+ from cells, whereas metabolic alkalosis favors its uptake. Increased osmolality of the ECF enhances K^+ release by cells. Cell lysis (severe trauma, burns, etc.) induces K^+ release and may produce hyperkalemia. During intense exercise, K^+ is also released from skeletal muscle cells.

External balance There are two hypotheses concerning the homeostatic mechanisms which regulate external K^+ balance:

1. The first involves a peripheral mechanism without central nervous system (CNS) intervention. In this case [K^+] would stimulate urinary K^+ excretion and aldosterone secretion directly.
2. The second attributes K^+ regulation to a reflex mechanism initiated by K^+ receptors in portal vein and liver stimulated by high [K^+]. Afferent fibers would transmit signals to CNS – 'K control center' – and efferent limbs, mediated by several hormones and other factors, would modulate urinary K^+ excretion in response.

In any case, plasma [K^+] and aldosterone are the major physiological regulators of K^+ excretion. High [K^+] increases aldosterone secretion and both act synergistically, stimulating K^+ elimination by the intestine (particularly increasing colonic secretion) and, above all, by the kidney (promoting secretion in the distal tubule and collecting duct). Renal secretion is due to stimulation of the Na^+/K^+-ATPase pump, increased permeability of the apical membrane to K^+, and greater K^+ uptake across the basolateral membrane. Hypokalemia decreases K^+ secretion by antagonist action. Under repletion conditions, potassium transported by Na^+/K^+-ATPase and H^+/K^+-ATPase recycles back into the lumen through K channels; however, when potassium levels are low, little potassium is wasted through luminal channels and passage of K through basolateral channels predominates.

Aldosterone secretion is favored by angiotensin II. On the other hand, aldosterone stimulates Na^+ and

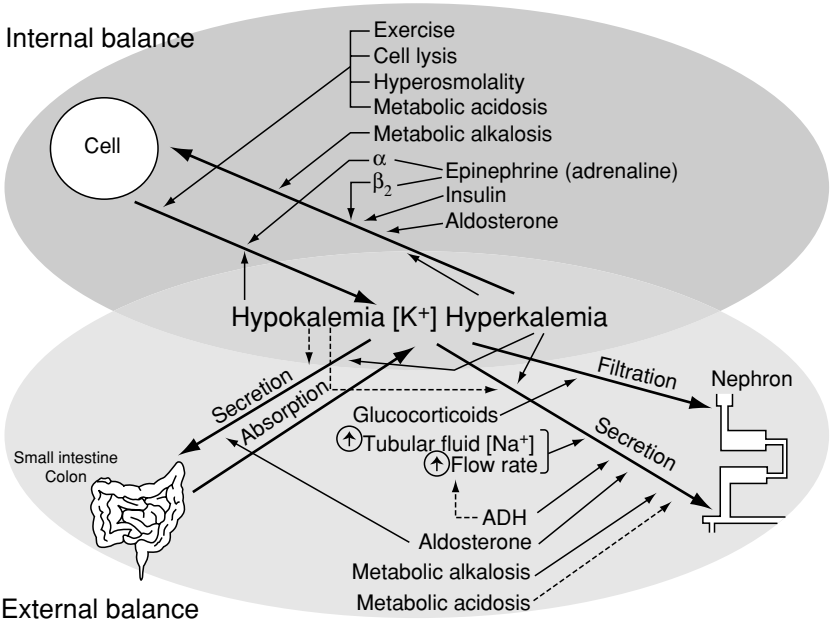

Figure 3 Potassium homeostasis. ADH, antidiuretic hormone. Continuous arrow, increase; dashed arrow, decrease.

water reabsorption and consequently decreases tubular flow at first. However, flux is restored in time, permitting aldosterone to increase K^+ renal excretion.

Antidiuretic hormone (ADH) increases K^+ secretion in the apical membrane of the principal cells in exchange for Na^+ absorption. However, because ADH also decreases tubular flow, changes in ADH levels do not substantially alter urinary K. K^+ secretion rises along the distal tubule in a flow-dependent manner and it is influenced by dietary potassium intake. This means that diuretics may enhance urinary K^+ excretion. However, K^+ excretion and distal flow rate are not completely coupled and, the K^+ balance is maintained through compensatory mechanisms when necessary.

Physiological [Na^+] does not alter K^+ secretion but high [Na^+] in tubular fluid increases K^+ secretion, whereas a fall in concentration has the opposite effect. Acid–base balance is another factor that modulates K^+ secretion: alkalosis increases secretion, whereas acidosis decreases it. Glucocorticoids are kaliuretic, but their effects appear to be a consequence of an increase in the glomerular filtration rate.

K^+ itself exerts a feedback regulation over most of its own control factors. Together, all these mechanisms allow the human body to maintain normal plasma [K^+]. However, the equilibrium can sometimes be altered, causing either hyperkalemia (extracellular [K^+] over $5.5\,mmol\,l^{-1}$) or hypokalemia (extracellular [K^+] below $3.5\,mmol\,l^{-1}$).

Hyperkalemia may be the consequence of a K^+ shift from cells to ECF or the effect of excessive K^+ retention. The first signs are flaccid paralysis, natriuresis, and other minor effects. The most important consequences are ventricular fibrillation and cardiac arrest. Hyperkalemia may be:

- Without potassium excess: caused by K^+ shift from cells (acute acidosis, hyperosmolality, insulin deficiency, β-adrenergic blockers, vigorous exercise, etc.)
- With potassium excess: impaired renal excretion (renal failure, aldosterone deficiency, primary tubule dysfunction). Inappropriately high oral or parenteral K^+ administration.

Hypokalemia may be the result of either intracellular shift or K^+ depletion, or both. Among its clinical sequelae are hyperpolarization of membrane, which affects nerves and muscle activity, and produces renal, cardiac, and metabolic alterations. The hypokalemia may be caused by:

- Intracellular shift: alkalosis, insulin excess, familial periodic paralysis, catecholamine increase, intoxications, etc.
- Extrarenal causes: insufficient K^+ intake, diarrhea, vomiting, skin loss, etc.
- Excessive renal losses: renal tubular acidosis, hypermineralocorticoidism, syndrome of choride depletion, diuretic condition.

See also: **Body Composition**; **Colon**: Structure and Function; **Diarrheal (Diarrhoeal) Diseases**; **Dietary Requirements of Adults**; **Electrolytes**: Water–Electrolyte Balance; Acid–Base Balance; **Hormones**: Adrenal Hormones; Pituitary Hormones; **Hypertension**: Hypertension and Diet; **Legumes**: Dietary Importance; **Potassium**: Properties and Determination; **Renal Function and Disorders**: Kidney: Structure and Function; Nutritional Management of Renal Disorders; **Vegetarian Diets**; **Water**: Physiology

Further Reading

Aizman R, Grahnquist L and Celsi G (1998) Potassium homeostasis: ontogenic aspects. *Acta Pediatric* 87: 609–617.

Berne RM and Levy MN (1998) *Physiology*, 4th edn, pp. 699–754. St Louis: Mosby.

Eaton SB, Eaton III SB, Konner MJ and Shostak M (1996) An evolutionary perspective enhances understanding of human nutritional requirements. *Journal of Nutrition* 126: 1732–1740.

Giebisch G (1998) Renal potassium transport: mechanisms and regulation. *American Journal of Physiology* 274(*Renal Physiology* 43): F817–F833.

Johnson LR, Alpers DH, Christensen J, Jacobson ED and Walsh JH (1994) *Physiology of the Gastrointestinal Tract*, 3rd edn, pp. 2027–2171. New York: Raven Press.

Navarro MP and Vaquero MP (1998) Potassium. Physiology, dietary sources and requirements. In: Sadler M, Caballero B and Strain S (eds) *Encyclopedia of Human Nutrition*, pp. 1573–1580. London: Academic Press.

Nicholes CG and Lopatin AN (1997) Inward rectifier potassium channels. *Annual Reviews of Physiology* 59: 171–191.

Nilius B, Viana F and Droogmans G (1997) Ion channels in vascular endothelium. *Annual Reviews of Physiology* 59: 145–170.

Shils ME, Olson JA, Shilke M and Ross AC (1999) *Modern Nutrition in Health and Disease*, 9th edn, pp. 105–139, 1217–1225. Philadelphia: Lippincott/Williams & Wilkins.

Suter PM (1998) Potassium and hypertension. *Nutrition Reviews* 56: 151–153.

Suter PM (1999) The effects of potassium, magnesium, calcium, and fiber on risk of stroke. *Nutrition Reviews* 57: 84–91.

Wang W, Hebert SC and Giebisch G (1997) Renal K^+ channels: structure and function. *Annual Reviews of Physiology* 59: 413–436.

Yokoshiki H, Sunagawa M, Seki T and Sperelakis N (1998) ATP-sensitive K^+ channels in pancreatic, cardiac, and vascular smooth muscle cells. *American Journal of Physiology* 174 (*Cell Physiology* 43): C25–C37.

POTATOES AND RELATED CROPS

Contents

The Root Crop and its Uses

M V Rama and P Narasimham, Formerly of Central Food Technological Research Institute, Mysore, India

Introduction

The potato (Irish potato, white potato, *Solanum tuberosum* L.) is of ancient origin. It has originated and was first domesticated in South America, even before the appearance of maize. It was initially introduced to Spain and the UK, and gradually to parts of Asia and North America. It has spread around the world over the past 400 years, and gained recognition as an inexpensive and nutritive food in the eighteenth century. It is now among the 10 major food crops of the world and grown in over 140 countries. This article reviews the geographical distribution, varieties, commercial importance, morphology, and anatomy of the tuber, chemical composition, nutritive value, and postharvest handling and storage of potatoes.

Geographical Distribution

The potato (*S. tuberosum* L.) belongs to the Solanaceae family and is largely grown in cool regions where the mean temperature during the growing season does not exceed 18 °C. World production of potato accounts for 47% from 37% acreage of total root and tuber crops. It is the fourth most important food crop in the world, next to wheat, maize, and rice in global tonnage (Table 1). The principal producers are the European and Asian countries, North America, and the Andean countries of Latin America (Table 2). Production shows a yearly increase in both the developing and developed countries, more significantly in the former. Russia, China, USA, Poland, India, Germany, UK, France, Turkey, and Canada are some of the highest potato-producing countries of the world (Table 3). The USA and Canada together account for 8% of the world production from 4% of acreage.

Commercial Varieties and Transgenic Potatoes

There are more than 150 wild species of potato found in Central America, Mexico, and the USA. *S. tuberosum* L. (tetraploid) represents the cultivated species and there are seven other cultivated species, including *S. ajanhuiri, S. goniocalys, S. phureja, S. stenotomum* (diploids), *S.* × *juzepczukii, S.* × *chaucha* (triploids), *S.* × *curtilobum* (pentaploid), which are grown in different parts of Peru, Bolivia, Ecuador, and Venezuela. Many improved local varieties have been developed and introduced in several countries and these may show increased yield, disease resistance, better tuber shape, texture, and quality for processing as well as flavor.

Table 1 World production of potatoes compared to major tuber crops and cereals

Crop	*Area harvested (× 1000 ha)*	*Yield (kg ha^{-1})*	*Production (× 1000 Mt)*
Potato	17 586	164 063	288 522
Cassava	16 601	99 983	165 986
Yam	3802	96 122	36 545
Sweet potato	9002	133 215	199 919
Wheat	213 790	27 313	583 918
Rice	513 177	38 437	588 766
Maize	139 878	43 209	604 400

Data from www.fao.org – Food and Agriculture Organization Statistical database, provisional 1999 production and production indices data.

Table 2 Potato production in parts of the world

Parts of the world	*Area harvested (× 1000 ha)*	*Yield (kg ha^{-1})*	*Production (× 1000 Mt)*	*Percent production*
World	17 586	164 063	288 522	100
Africa	811	111 790	9072	3.1
Asia	9171	148 288	92 058	31.9
Europe	6208	156 305	143 349	49.7
North Central America	306	918 324	28 134	9.7
South America	1030	135 924	14 008	4.8
Oceania	58	326 278	1899	0.6
Developing countries	7585	141 777	107 550	37.3
Developed countries	10 000	180 969	180 972	62.7

Data from www.fao.org – Food and Agriculture Organization Statistical Database, provisional 1999 production and production indices data.

Table 3 Major potato-producing countries of the world (production – × 1000 Mt)

Country	*Year*		
	1995	*1997*	*1999*
Former Soviet Union	39 909	37 039	30 300
China	45 754	45 534	43 477
USA	20 122	20 861	21 840
Poland	24 891	20 776	26 000
India	17 401	19 240	22 100
Germany	10 888	12 438	12 074
UK	6396	7154	7000
France	5882	6500	6500
Turkey	4750	5000	5315
Canada	3374	4050	4260

Data from www.fao.org – Food and Agriculture Organization Statistical Database, provisional 1999 production and production indices data.

In recent years, attempts have also been made worldwide to generate transgenic potato plants and assess the impact of transgenic expressions on diverse parameters such as yield, quality, carbohydrate metabolism, stress physiology, and pest and disease resistance. The outcome of the transgenic approach is promising and already insect- and herbicide-resistant potato plants are in commercial use.

Commercial Importance

The potato is grown as an early or late crop for seed or ware purpose. Even in comparatively harsher climates, it produces more dry matter (DM) and protein per hectare than major cereal crops of the world. It is a major food crop in the temperate zone and is either a staple food or merely a vegetable in tropical and subtropical countries. Tubers are consumed in various forms after cooking or processing. In the UK, USA, and India 65–90% of the production is consumed as food, and a small quantity is used for industrial purposes. In The Netherlands, a large percentage is used for starch production. Starch, alcohol, glucose, and dextrin are the industrial products of potato. World production of potato starch is about 2×10^6 t. Potatoes are generally boiled, cooked in oil, or baked. Dehydrated, frozen, and canned products are also popular. In the USA, 50% of the potato crop is consumed in processed form; the most relished is potato chips. Varieties with high DM content are preferred for potato chips. Frozen products include hash browns (patties), French fries, croquettes (puffs), and mashed potato, while dehydrated products include granules, flakes, diced chunks, and julienne strips. The common canned products are new potatoes, hash, stew, and soups. Papa seca and chuno are the traditional dried potatoes, which form a vital part of the diet in the highland areas of Peru and Bolivia. Dough-like sliceable, storage-stable potato products have also been developed. (*See* **Starch**: Sources and Processing.)

The waste water collected from peeling, cutting, and blanching in the processing industries has been used to recover highly nutritive protein of 80–85% purity. The use of this recovered protein of high biological value is limited as human food, due to the presence of antinutritional factors, but the residual water can be used in preparing a growth medium for single-cell protein production. Potato peel itself is superior to wheat bran, containing high amounts of

Table 4 Advantages of using true potato seeds (TPS) for potato production

Parameters	*Advantages*
Seed rate	25 g of TPS against 2 t seed tubers per hectare
Profit	Planting material cost is very low; saves cost on transport, cold storage, and needs less storage room
Productivity	More productive due to higher yield and more dry matter
Technical feasibility	Can be effectively used in different agroclimatic zones
Less risky/environment-friendly	Built in broad-based resistance to late blight: reduced usage of pesticides

Adapted from Upadhya MD (1994) True potato seed: propagule for potato production for the 21st century. In: Shekhawat GS, Khurana P, Pandey SK and Chandla VK (eds), *Potato: Present and Future*, pp. 15–22. Shimla, India: Indian Potato Association.

minerals and fiber. The water extract of the peel is a nonmutagenic antioxidant with potential antimicrobial activity. Snakin-1 (SN-1), a peptide isolated from the potato tuber, is shown to be active against potato bacterial and fungal pathogens. Further, potato plant having tuber as the sink for carbon and nitrogen is said to serve as a model species for various experiments involving bacteria-mediated gene transfer systems. (*See* **Alkaloids**: Properties and Determination.)

Tuber is also used for propagation. Tuber seed is the costliest input, accounting for nearly 40% of the cost in potato production; the availability of an adequate quantity of virus-free seed stock is still a constraint for increased production, especially in developing countries. Hence, in recent years commercial crops are being raised using true potato seed (TPS), which has several advantages over traditional tuber seeds (Table 4). However, usage of TPS is still limited in developed countries, due to erratic germination and low yield. A host of hybrid TPS families with desirable attributes has been developed by the International Potato Research Center (CIP), Peru, and Central Potato Research Institute (CPRI), India, using *S. andigena* and *S. tuberosum* as parental lines.

Morphology and Physiology of the Tuber

The potato is an underground, modified, fleshy stem, with a shortened axis, developing from the subapical region of a diageotrophic rhizome or lateral shoot. The tuber end that is attached to the rhizome is called the stem end or heel end, and the other end is the bud end or rose end. There are as many as 20 eyes in the axils of leaf scars (eye brows) arranged spirally around a tuber. Each eye has one main bud and several small lateral buds. Physiologically, the youngest bud is the last formed apical bud (Figure 1a). Tubers sprout after completing the inherent dormancy period, which lasts from 6 to 16 weeks after harvest, depending upon the variety. Tubers exhibit apical dominance. Studies have shown that the balance between the two hormones, abscisic acid (ABA) and gibberellic acid (GA_3), in the tubers determines the extent of the dormancy period. Dormancy breaking is also related to protein degradation and mobilization of nitrogen reserves for sprout formation. Morphological characters of the tuber such as shape (round, oval, or elongated), size (small, medium, or big), skin color (light yellow to black or purple), and depth of eye (deep or shallow) determine its acceptability for various end uses.

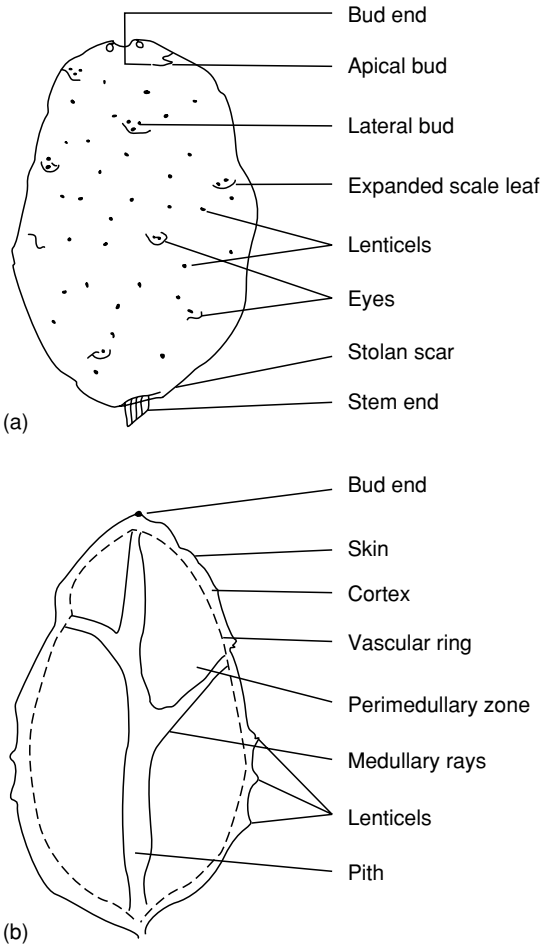

Figure 1 A potato tuber: (a) morphological details; (b) longitudinal section.

Anatomy of the Tuber

The potato tuber is protected by the outermost skin or periderm, consisting of six to 10 layers of suberized cells. Skin color depends on the anthocyanin concentration in the periderm and peripheral cortex. The active periderm of a young tuber can be easily removed, and tuber enlargement is associated with sloughing off the periderm, which is then replaced by a new 'cork' layer, formed from beneath. Similarly, wound healing takes place in damaged tuber tissue by the formation of a wound periderm, which is more impervious than normal skin. A number of lenticels are found in the periderm, and these pores also facilitate the exchange of gases and the entry of pathogens. Inside the periderm is the parenchymatous cortex (0.3–1.0 cm thick) in which food material is stored in the form of starch granules. Cortex (potato flesh) may be white or various shades of yellow depending upon the variety. Yellow-fleshed varieties are sometimes highly prized. The parenchymatous perimedullary zone is seen between the vascular ring and the medullary ray (Figure 1b). Xylem is visible as a ring, while phloem forms many bundles in the cortex and perimedullary zones. Medullary rays run from the stem end to the eyes. The pith present at the center of the tuber is translucent, as it has less starch.

Chemical Composition

The chemical composition of potato varies with variety, soil type, cultural practice, maturity stage, disease, and storage conditions. The distribution of DM and chemical compounds (Figure 2) is extremely heterogeneous within a tuber. The data in Tables 5 and 6 shows potato to be a good source of carbohydrates, vitamins, and minerals. The small fraction of protein present in the tuber consists of many essential amino acids.

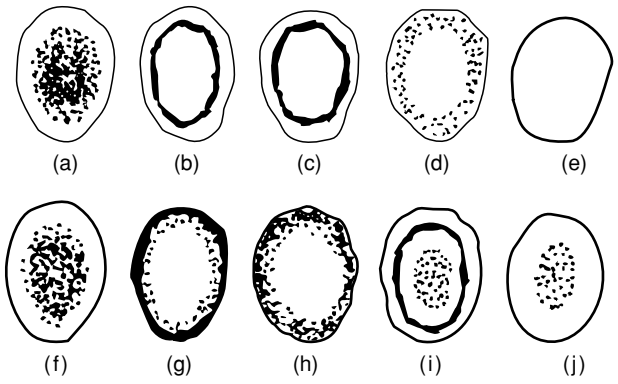

Figure 2 Pattern of distribution of chemical compounds (shaded areas) within a potato tuber. (a) Starch; (b) sugars; (c) proteins and amino acids; (d) fat; (e) crude fiber; (f) vitamins; (g) minerals; (h) organic acids; (i) phenol compounds; (j) alkaloids. Adapted from Rastovski A, Van Es A *et al*. (1987) *Storage of Potatoes. Past Harvest Behaviour. Store Design, Storage Practice, Handling*. Wageningen, The Netherlands: Center for Agricultural Publishing and Documentation.

Carbohydrates

Starch, sugars, and nonstarch polysaccharides constitute the carbohydrate fractions of potato tuber.

Table 5 Chemical composition of the dry matter (22.5%) of the potato tuber

Chemical composition	*Content*[a]
Crude fiber	2.2
Starch	74.2
Total sugar	1.3
Reducing sugar	0.6
Fat	1.0
Total nitrogen (N)	1.2
Protein N	1.0
Protein fractions (% of total protein N)	
Albumin	48.9
Globulin	25.9
Prolamin	4.3
Glutelin	8.3
Minerals	
Calcium	0.02
Magnesium	0.08
Potassium	1.47
Sodium	0.02
Phosphorous	1.87
Iron (p.p.m.)	15.70
Vitamins (mg 100 g^{-1})	
Thiamin	0.73
Ascorbic acid	92.08
Nicotinic acid	10.08
Riboflavin	0.12

[a]Values given as percentages, unless otherwise stated.
Data adapted from Rastovski A, Van Es A *et al*. (1987) *Storage of Potatoes. Post Harvest Behaviour, Store Design, Storage Practice, Handling*. Wageningen, The Netherlands: Center for Agricultural Publishing and Documentation.

Table 6 Essential amino acid composition of potato tuber (dry-weight basis)

Amino acid	*Mean concentration*[a] *(mg g^{-1} tuber)*
Histidine	2.5
Isoleucine	3.4
Leucine	5.0
Lysine	4.9
Methionone + cysteine	1.9
Phenylalanine + tyrosine	7.0
Threonine	3.1
Tryptophan	1.3
Valine	5.7

[a]Average of lowest and highest values reported.
Reproduced from Jadhav SJ and Kadam SS (1998) Potato. In: Salunkhe DK and Kadam SS (eds) *Handbook of Vegetable Science and Technology: Production, Composition, Storage and Processing*, pp. 11–69. New York: Marcel Dekker.

Starch, being the major carbohydrate (16–20% on a fresh-weight basis or fwb), constitutes 60–80% of the dry matter, and it is composed of amylose and amylopectin in a 3:1 ratio. Potato starch gelatinizes above 70 °C. The total sugar content ranges from 0.1% to 0.7% (fwb), and this is chiefly associated with maturity, senescence, and sprouting. The major sugars of potato are glucose, fructose, and sucrose. Trace amounts of melibiose, raffinose, stachyose, glycerol, galactinol, and glucosyl have also been identified. Tubers containing more than 2% (fwb) reducing sugars give dark-colored chips owing to Maillard reactions and are not suitable for processing. In mature stored tubers, starch and sugars exist in a state of dynamic equilibrium. The nonstarchy polysaccharides, such as cellulose, hemicellusose, and pectic substances, constitute about 1.2% (fwb) and are present in the cell walls and middle lamellae. They contribute to the final texture of cooked potato, and act as a source of dietary fiber. (*See* **Browning**: Nonenzymatic; **Carbohydrates**: Requirements and Dietary Importance; Classification and Properties; **Starch**: Structure, Properties, and Determination.)

Proteins and Amino acids

Potato is considered to be low in protein (2% fwb), but is rich in lysine compared to cereal proteins, while the concentration of sulfur amino acids is less than in cereals. The protein is concentrated more in the cortex and pith. The levels of proteins such as albumin, globulin, prolamine, and glutelin are 48.9%, 25.9%, 4.3%, and 8.3% of total protein respectively. Nearly 75% of the nonprotein nitrogen (NPN) occurs as free amino acids and amides. Two-thirds of the NPN fraction is composed of free amino acids and 21 of the amino acids have been identified. The essential amino acids and their concentrations are given in Table 6. Sprouting, storage, diseases, and fertilizer applications influence the concentration of free amino acids in the tuber. Amides like glutamine and asparagine occur in almost equal amounts. (*See* **Protein**: Chemistry; Requirements.)

Vitamins and Minerals

Potato contains substantial quantities of vitamins B and C (Table 5). Vitamin C is present in both oxidized and reduced forms. Freshly harvested tubers may contain 20 mg of ascorbic acid per 100 g; losses of vitamin C occur during long-term storage (40–60%), cooking, and processing (20%). Ascorbic acid content of potato is higher than that of several other vegetables like carrots, pumpkins, onions, and green beans. The vitamin B group comprises thiamin, riboflavin, nicotinic acid, and pyridoxine; folic acid along with pantothenic acid is also present in potato. The ash content is about 1% (fwb), which is equivalent to 4–6% of the DM content. Fat-soluble vitamins occur in traces or are absent. This necessitates supplementation of other food sources rich in vitamin A in a potato diet. The major elements present are phosphorus, potassium, magnesium, sodium, and calcium, with wide ranges in their amounts. Potato is a poor source of calcium and sodium. A small percentage of phosphorus (25%) occurs as insoluble phytic acid. Others, such as boron, copper, zinc, iodine, aluminum, arsenic, nickel, and molybdenum, are found in trace amounts. (Refer to individual minerals and vitamins.)

Lipids and Organic Acids

Approximately 0.1% (fwb) lipid is found in potato; it is concentrated in the periderm. Linoleic, linolenic, and palmitic acids are the major fatty acids. A number of organic acids are present in potato in varying quantities; they contribute to the flavor and buffering of the potato sap. The major organic acids are citric, oxalic, fumaric, and malic acids. Other than phytic acid, nicotinic and chlorogenic acids have also been reported. Chlorogenic acid reacts with ferric iron, forming a complex, which causes darkening after cooking. Enzymatic browning in cut and homogenized potato tissue is caused by the oxidation of tyrosine. Both enzymatic and nonenzymatic browning can be inhibited by treatments with sulfur dioxide and sulfites. Other phenolic compounds found in potato are polyphenols, flavones, anthocyanins, and tannins. Phenols are again associated with after-cooking discoloration of tubers, particularly at the stem end. Tannins being localized in the periderm impart tan coloration to the skin. (*See* **Acids**: Natural Acids and Acidulants; **Fats**: Classification; **Fatty Acids**: Properties; **Phenolic Compounds**; **Tannins and Polyphenols**.)

Enzymes and Pigments

The enzymes reported in potato include amylase, glyoxalase, phosphorylase, tyrosinase, peroxidase, catalase, aldehydrase, phosphatase, sistoamylase, and zymohexase. The phosphorylase and amylase systems form sugars at low temperatures. The probable role of D-enzyme (EC 2.4.1.25; 4-alpha-glucanotransferase) in starch metabolism has been suggested. Transgenic potato plants with reduced D-enzyme activity have been obtained. Polyphenol oxidase, peroxidase, and catalase enzymes involved in the oxidation of phenols bring about browning of the freshly cut surface of the potato. The yellow color of potato flesh is attributable to carotenoids such as α-carotene, auroxanthin,

violaxanthin, lutein, isolutein, and neo-β-carotene. Flavonols, flavones, flavin, and neoxanthin are also found in small quantities. Red-skinned potatoes have anthocyanin in the periderm and outer cortical cells. Chlorophyl is present in tubers exposed to light, and these green potatoes lose their market value. (*See* **Carotenoids**: Occurrence, Properties, and Determination; **Chlorophyl**; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Enzymes**: Functions and Characteristics.)

Antinutritional Factors

The glycoalkaloids, proteinase inhibitors, and the lectins are the major antinutritional factors present in the tubers. The synthesis of toxic glycoalkaloids such as α-solanine and α-chaconine, which are believed to be a part of a disease-resistance mechanism, takes place in damaged and light-exposed tubers; a normal tuber contains an insignificant amount (5–10 mg $100\,g^{-1}$) of these alkaloids. They are recognized as natural products with insect deterrent and antifungal activity, concentrated more in the peel, sprouts, and around the eyes, and absent in the pith. Tubers develop bitterness and off-flavors when the alkaloid content is 25–80 mg $100\,g^{-1}$, and consumption of tubers with more than 20 mg of glycoalkaloid causes fatal illness. It can be eliminated to an extent of 60%, by peeling the tuber, while its accumulation can be controlled by waxing, oil and water dipping, and storing in the dark. The nutritional significance of two other toxic substances, proteinase inhibitors and lectins, has not been much studied, but they are reported to be heat-labile. (*See* **Alkaloids**: Toxicology.)

Nutritive Value

Potato is one of the richest sources of energy and has a greater capacity to supply energy than any other food crop. In some developing countries, potato is considered merely a vegetable rather than a staple food because of the belief that excessive consumption causes flatulence in humans. Experiments have shown that cooked potato starch is easily digestible, and hence forms a valuable food, even for infants. Potato has a slightly lower energy content (335 kJ $100\,g^{-1}$) than other roots, tubers, cereals, and legumes (Table 7), but this is advantageous in overcoming the problem of obesity in the developed world. On the other hand, large quantities of potato have to be consumed in developing countries to meet their populations' daily energy needs. It has been calculated that 100 g of potato can supply 5–7% of the daily energy and 10–12% of the daily protein need of children aged 1–5 years.

Potato protein has a biological value equal to that of soya bean protein and the ratio of total essential amino acids to total amino acids is so balanced that it can meet the needs of infants and small children. It has been calculated that 100 g of potato can supply 3–6% of proteins, the recommended daily allowance (RDA) for adults (depending upon sex and body weight), and one medium-sized potato provides 15 mg of vitamin C, which is about 20% of the recommended allowance (75 mg) per person per day. Eating the tuber with the skin on increases the dietary fiber intake. It is also a most valuable food for those who suffer from excess acidity of the stomach as it has an alkaline reaction. Potato fat is too low to have any nutritional significance, but it does contribute towards palatability. The tuber also provides most of the trace elements needed to maintain good health, and although potato is not a primary source of iron, 100 g of cooked potato can supply between 6% and 12% of the daily iron requirement of children or adult men. About 7% of the US RDA of phosphorus for both children and adults is available from 100 g potato. A low percent of phytic acid present in potato tuber, compared to other staple cereals and vegetables, is an advantage as it allows greater availability of free calcium, iron, zinc, and phosphorus for absorption by the human intestine. (*See* **Amino Acids**: Metabolism; **Ascorbic Acid**: Physiology; **Dietary Fiber**: Physiological Effects; **Energy**: Measurement of Food Energy; **Iron**: Physiology; **Protein**: Quality.)

Table 7 Composition of potato (raw and dried) and other major cereals and root crops (per 100 g edible portion)

Crop	*Energy (kJ)*	*Moisture (%)*	*Crude protein (g)*	*Fat (g)*	*Total carbohydrate (g)*	*Crude fiber (g)*	*Ash (g)*
Potato (raw)	335	78.0	2.1	0.1	18.5	2.1	1.0
Potato (dried)	1343	11.7	8.4	0.4	74.3	8.4	4.0
Rice	1523	12.0	6.8	0.5	80.2	2.4	0.6
Wheat	1389	12.3	13.3	2.0	70.9	12.1	1.7
Sweet potato	485	70.2	1.4	0.4	27.4	2.5	1.1
Yam	444	72.0	2.2	0.2	24.2	4.1	1.0
Cassava	609	62.6	1.1	0.3	35.2	5.2	0.9

Adapted from Woolfe JA (1987) The Potato in the Human Diet. Cambridge: Cambridge University Press.

Handling and Storage

In most countries, potatoes are only harvested at certain times of the year, and harvested tubers must be stored for at least a few months. Harvesting and subsequent processes such as assembling, grading, bagging, transport, and marketing, may cause damage to the tubers; such tubers lose water quickly and are highly susceptible to microbial spoilage. Bruising also causes a physiological disorder called black spot which is as the result of oxidative biochemical reactions involving chlorogenic acid, polypohenol oxidase, amino acids, and tyrosine, resulting in melanin formation (enzymatic browning); bruised black-spotted tubers have less market value and shorter shelf-life. Degree of dormancy, rate of tuber respiration, chemical composition, and ability of wound healing influence the spread of disease-causing organisms through the tuber tissue. Thus curing, a process which promotes wound healing, is necessary before potato storage. Complete curing takes 3–6 days at 20 °C and 85–90% relative humidity (RH), but is faster at higher temperature. The physiological disorders, diseases, and pests of the tubers in the field and in storage are listed in Table 8.

Good storage conditions should prevent excessive sprouting, root development, moisture loss, sugar accumulation, greening, and temperature damage. The postharvest losses in potatoes are estimated to vary from 5 to 40%. For short periods, the tubers can be stored in clamps, in specially designed simple sheds, pits, cellars, or buildings with controlled temperature and humidity. Modern potato storages are equipped with computer-based control systems to monitor and maintain optimum levels of temperature and humidity with air envelop facilities. Long-term storage without sprouting and minimum loss of moisture is possible at 4–5 °C, but temperatures below 6 °C lead to the formation of reducing sugars, and this is not a desirable feature for industrial processing. Tubers can be reconditioned by transfer to a higher temperature, but the sweetness developed by long-term, low-temperature storage (senescent sweetening) cannot be eliminated completely; nearly 30% of the starch is lost as sugars, reducing tuber quality, particularly for processing.

The freezing of tubers takes place at − 1 °C to − 2 °C, and frozen tubers soon become soft and unusable. Potato stores are artificially heated using oil or gas heaters when the outside air temperature is lower than the required temperature. On the other hand, when potatoes are stored above 30 °C, accumulation of carbon dioxide leads to death of cells, causing black heart. In temperate regions, where the ambient temperature falls below 4 °C, potatoes are cooled with outside air. In the tropics and subtropics, refrigeration is used. The recommended temperatures for tubers to be used for different purposes are as follows:

- Seed potato 2–4 °C
- Fresh consumption 4–5 °C
- Chipping 7–10 °C
- French frying 5–8 °C
- Granulation 5–7 °C

(*See* **Storage Stability**: Mechanisms of Degradation.)

Sprouting increases weight loss, softens the tuber, and favors accumulation of glycoalkaloids, thus decreasing marketability and nutritive value. Apart from low-temperature storage, other methods of achieving sprout inhibition include the use of maleic hydrazide (MH), isopropyl-*N*-phenylcarbamate (IPC) and isopropyl-N-(3-chlorophenyl)-carbamate (CIPC), tetrachloronitrobenzene (TCNB), naphthalene acetic acid (NAA), methyl ester of naphthalene acetic acid (MENA), and irradiation. Time, method of application, and concentration of the chemical are important factors, as treatment may have adverse effects such as

Table 8 Potato tuber diseases and pests in the field and during storage

Disease	*Causal agent/organism*
Nonparasitic diseases	
Soft rot, mahogany browning, and net necrosis	Low-temperature storage (chilling injuries)
Black heart and hallow heart	Low oxygen supply
Black spot	Bruising
Pathological diseases	
Bacterial	
Bacterial soft rot	*Erwinia carotovora*
Brown rot	*Pseudomonas solanacearum*
Bacterial ring rot	*Corynebacterium sepedonicum*
Common scab	*Streptomyces scabies*
Fungal	
Powdery scab	*Spongophora subterranea*
Dry rot	*Fusarium coeruleum*
Late blight	*Phytophthora infestans*
Early blight	*Alternaria solani*
Charcoal rot	*Macrophomina phaseoli*
Wart	*Synchytrium endobioticum*
Gangrene	*Phoma exigua*
Pink rot	*Phytophthora erythroseptica*
Insect (pest)	
Potato tuber moth	*Gnoremoschema operculella*
Viral	
Tobacco rattle, mop top, yellow viruses dwarf diseases	

Adapted from Pushkarnath (1976) *Potato in Sub-Tropics*. New Delhi: Orient Longman; and Rich AE (1983) *Potato Diseases*. New York: Academic Press.

Table 9 Sprout yield in naphthalene acetic acid (NAA) and vapor heat (VH)-treated tubers at the end of 3 months of storage

Storage temperature (°C)	*Sprout weight (fresh – g per 100 tubers)*			
	Control	*NAA-treated*	*Control*	*VH-treated*
2 ± 1	17.2	6.2		
10 ± 1	13.9	4.8		
25 ± 5	26.0	2.5	58.7	10.1

Adapted from Rama MV and Narasimham P (1985) *Studies on the effect of various post-harvest treatments for controlling shriveling, sprouting and spoilage of potatoes during storage*. PhD thesis. India: University of Mysore.

Table 10 Physiological loss in weight (PLW), sprouting, and spoilage of potatoes during evaporative cooling storage for 30 days

Parameter	*Ambient storage (25–35 °C)*		*Evaporative cooling storage*	
	Untreated	*NAA-treated*	*Untreated*	*NAA-treated*
PLW (%)	2.6	2.4	1.1	0.9
Fresh sprout weight (g per 100 tubers)	86.1	21.6	70.1	17.1
Spoiled tubers (%)	4.5	0.0	0.0	0.0

Adapted from Rama MV and Narasimham P (1989) *Studies on the effect of various post-harvest treatments for controlling shrveling, sprouting and spoilage of potatoes during storage*. PhD thesis. India: University of Mysore.

inhibition of wound healing, increasing the reducing sugar content, and the problem of toxic residues. TCNB, CIPC, and IPC treatments are to be coupled with cold storage, as they are ineffective at tropical ambient temperatures (> 20 °C). The sodium salt of NAA (1000 p.p.m.) is an effective and economical postharvest sprout-retardant for storing potatoes under tropical ambient conditions (Table 9).

Several naturally occurring volatile and aromatic compounds have proved to be potent sprout inhibitors, but their acceptability for commercial application needs further investigation. Ware potatoes (those potatoes intended for human consumption in contrast to seed potatoes) can be kept sprout-free for a long time by irradiation, but its economic feasibility remains to be ascertained. Irradiated tubers need to be stored below 20 °C, and irradiation often lead to increased *Fusarium* attack and discoloration after cooking. Controlled-atmosphere (CA) storage of potatoes with as low as 5% carbon-dioxide concentration causes black heart and low oxygen concentrations inhibit wound healing; therefore CA storage appears not to be beneficial. Sprout inhibition beyond 3 months at tropical temperatures (22–35 °C) is possible by periodical vapor heat treatment at 60 °C, RH 95% for 60 min (Table 9). This method is superior to manual desprouting, and avoids the use of toxic chemicals. Hot-water dip treatment at 57.5 °C for 20 or 30 min prior to storage at 18 °C enhances storability, by inhibiting sprouting and spoilage by *Fusarium* and *Erwinia*. (*See* **Irradiation of Foods: Applications.**)

Evaporative cooling systems can be used effectively in tropical conditions to store potatoes. The containers needed for this storage can be fabricated locally using brick, sand, bamboo, and metal sheets; it is a cheaper method of storing potatoes, with reduced water loss, at farm level (Table 10). These containers maintain a high RH (85–95%) in the atmosphere and the tuber temperature, which is close to the wet bulb temperature, is the lowest possible for evaporative cooling. Seed potatoes are often exposed to natural or artificial light during storage, as light retards physiological aging, inhibits sprout growth, and increases resistance to fungal infections. On the other hand, exposure of ware potatoes to light should be avoided, as it leads to the formation of the toxic α-solanine and green pigment, chlorophyl; the market value of green potatoes is also low.

Avoiding damage to potatoes and dry, cool storage are necessary to prevent bacterial and fungal attack. Benzimidazole compounds are used to combat fungal diseases. Efficient ventilation in any store is necessary to eliminate excess moisture, which otherwise may enhance rotting. (*See* **Fungicides; Spoilage:** Bacterial Spoilage.)

See also: **Acids**: Natural Acids and Acidulants; **Ascorbic Acid**: Physiology; **Carbohydrates**: Classification and Properties; Requirements and Dietary Importance; **Carotenoids**: Occurrence, Properties, and Determination; **Dietary Fiber**: Physiological Effects; **Energy**: Measurement of Food Energy; **Enzymes**: Functions and Characteristics; **Fungicides**; **Protein**: Chemistry; Requirements; **Spoilage**: Bacterial Spoilage; **Starch**: Structure, Properties, and Determination; Sources and Processing; **Storage Stability**: Mechanisms of Degradation

Further Reading

Anonymous (1984) *A Potato Store Run on Passive Evaporative Cooling*. Technical bulletin no. 11, pp. 1–14. Simla, India: Central Potato Research Institute.

Davis HV (1996) Recent developments in our knowledge of potato transgenic biology. *Potato Research* 39: 411–427.

Harris PM (1978) *The Potato Crop*. London: Chapman and Hall.

Jadhav SJ and Kadam SS (1998) In: Salunkhe DK and Kadam SS (eds) *Handbook of Vegetable Science and Technology: Production, Composition, Storage and Processing*, pp. 11–69. New York: Marcel Dekker.

Pushkarnath (1976) *Potato in Sub-Tropics*. New Delhi: Orient Longman.

Rama MV (1985) *Studies on the effect of various post-harvest treatments for controlling shriveling, sprouting and spoilage of potatoes during storage*. PhD thesis. India: University of Mysore.

Rama MV and Narasimham P (1986) Heat treatments for the control of sprouting of potatoes during storage. *Annals of Applied Biology* 108: 597–603.

Rama MV and Narasimham P (1989) Control of potato (*Solanum tuberosum* L. cv Kufri Jyoti) sprouting by sodium naphthyl acetate during ambient storage. *Journal of Food Science and Technology* 26(2): 83–96.

Rama MV and Narasimham P (1991) Evaporative cooling of potatoes in small naturally ventilated chambers. *Journal of Food Science and Techniology* 28(3): 145–148.

Rastovski A, Van Es A, Van Der Zaag DE *et al*. (1987) *Storage of Potatoes. Post Harvest Behaviour, Store Design, Storage Practice, Handling*. Wageningen, The Netherlands: Center for Agricultural Publishing and Documentation.

Rich AE (1983) *Potato Diseases*. New York: Academic Press.

Rodriguez De Sotillo D, Hadley M and Wolf-Hall C (1998) Potato peel extract: a nonmutagenic antioxidant with potential antimicrobial activity. *Journal of Food Science* 63(5): 907–909.

Shashirekha MN (1988) *Studies on the extension of storage life of seed potato tubers stored under different temperature conditions*. PhD thesis. India: University of Mysore.

Singh G, Kapoor IPS and Kumar Pandey S (1997) Studies on essential oils VII. Natural sprout inhibitors for potatoes. *Pesticide Research Journal* 9(1): 121–124.

Smith O (1977) *Potatoes, Production, Storing, Processing*. Westport, Connecticut: AVI Publishing.

Upadhya MD (1994) True potato seed; propagule for potato production for the 21st century. In: Shekhawat GS, Khurana P, Pandey SK and Chandla VK (eds) *Potato: Present and Future*, pp. 15–22. Shimla, India: Indian Potato Association.

Van Der Zaag DE (1990) Recent trends in development, production and utilization of potato crop in the world. *Asian Potato Journal* 1: 1–11.

Woolfe JA (1987) *The Potato in the Human Diet*. Cambridge: Cambridge University Press.

Yamaguchi M (1983) *World Vegetables: Principles, Production and Nutritive Value*, p. 415. Westport, Connecticut: AVI Publishing.

Fruits of the Solanaceae

M V Rama and P Narasimham, Formerly of Central Food Technological Research Institute, Mysore, India

Background

The family Solanaceae includes about 75 genera and 2000 species of herbs, shrubs, and small trees distributed in the tropical and temperate regions of the world. Important vegetable crops such as tomato, brinjal, pepper, tree tomato, and husk tomato are included in this family. Potato and tobacco are also from this family. This article reviews the geographical distribution, morphology, anatomy, chemical composition, nutritive value, and uses of some solanaceous fruits.

Geographical Distribution, Morphology, and Anatomy

The important fruits of the family solanaceae are shown in **Table 1**. Tomato (*Lycopersicon esculentum*) is the most popular and widely cultivated vegetable and has been consumed by the inhabitants of Central and South America since prehistoric times. It is considered to be a native of the Peruvian and Mexican regions. It is an indispensable fruit today, grown all over the world (outdoors in temperate regions and in green houses in the colder regions), and constitutes about 15% of the world total vegetable production (**Table 2**). World production of tomato for processing is over 20 million tonnes, and in the USA, it ranks second to potato in dollar value, among all the other vegetables grown. Major tomato-producing countries

Table 1 Some important solanaceous fruits

Scientific name	*Common name*
Lycopersicon esculentum	Tomato
L. pimpinellifolium	Redcurrant or grape tomato
Solanum melongena	Eggplant, brinjal, berenjana, aubergine, guinea squash
S. macrocarpon	African eggplant
S. nigrum	Garden huckleberry, wonderberry
S. muricatum	Pepino, melon pear
S. quitoense	Naranjillo, lulo
S. gilo	Jilo
Capsicum annuam	Bell pepper, sweet pepper
C. frutescens	Pimiento, chili, aji, hot pepper, bird chili
Physalis peruviana	Cape gooseberry, uchida
P. ixocarpa	Tomatillo, ground cherry
P. purinosa	Husk tomato
Cyphomandra betacea	Tree tomato, tamarillo

Table 2 World production of major solanaceous fruits (× 1000 tonnes)

Regions	*Tomato*	*Eggplant*	*Capsicum*
World	91 663	21 165	18 024
Africa	10 799	799	2 078
Asia	41 246	19 543	10 715
Europe	19 789	709	2 705
North Central America	13 423	106	2 191
South America	5 990	59	301
Oceania	416		33
Developing countries	56 930	19 852	14 323
Developed countries	34 733	1 313	3 701

Source: Food and Agricultural Organization Statistical Database – Provisional 1999 Production and Production Indices Data www.fao.org
World production of vegetable + melon = 622 428 × 1000 tonnes.

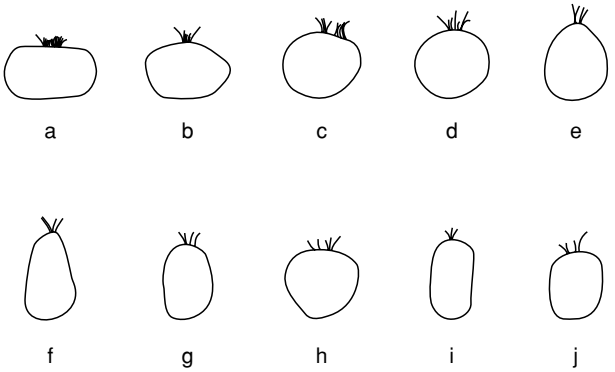

Figure 1 Shapes of tomato cultivars. a, flat; b, oblate; c, square round; d, round; e, pear; f, elongated pear; g, egg; h, oxheart; i, blocky elongated; j, blocky round. Adapted from Gould WA (1992) *Tomato Production, Processing and Technology*, 3rd edn. Baltimore, MD: CTI.

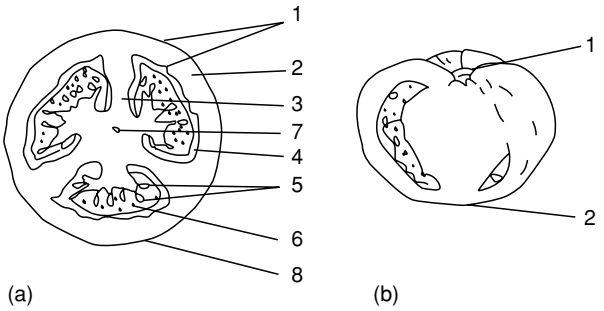

Figure 2 Structure of the tomato fruit: (a) transverse section; (b) longitudinal section. (a) 1, pericarp; 2, outer wall; 3, radial wall; 4, locule; 5, seeds; 6, gel; 7, axile fleshy core; 8, skin. (b) 1, stem scar; 2, blossom end scar.

are China, USA, Spain, Italy, Turkey, and Russia. The fruit is normally round, lobed or pear-shaped, and the diameter ranges from 1 to 12 cm. New cultivars with differently shaped fruits have been developed to suit mechanical harvesting and handling systems (Figure 1). It is a two- to many-loculed berry with a fleshy placenta and many small, kidney-shaped seeds, covered with short, stiff hairs and a jelly-like parenchyma (Figure 2). The tough skin or pericarp thickness varies with the cultivar.

The genus *Lycopersicon* includes red-fruited edible species with carotenoid pigmentation (subgenus *Eulycopersicon*), and green-fruited species with anthocyanin pigmentation (subgenus *Eriopersicon*). *Lycopersicon esculentum* and *L. pimpinellifolium* are red-fruited cultivars, while *L. pissisi, L. peruvianum, L. hirsutum, L. glandulosum*, and *L. cheesmanii* are green-fruited. Varieties evolved on pure-line selection, hybridization, and genetic engineering techniques for market and processing needs are also cultivated in many countries. The Flavr-savr™ is currently a genetically engineered approved tomato variety in the US market. Cultivars suited for home, market, shipping, and processing are listed in Table 3.

The edible *Physalis* species are grown widely in the warmer parts of the world. The fruit of *Physalis purinosa* (husk tomato) is round, yellowish, and acidic-sweet. *Physalis peruviana* (cape gooseberry) is a native of the Andes, and grows from Venezuela to Chile; the fruit is a spherical or ellipsoidal, smooth berry, measuring about 4 cm long and 3 cm wide. The skin color is greenish yellow. *Physalis ixocarpa* (tomatillo or ground cherry) is of Mexican origin; the fruit is a round, green or purplish berry, with a high ascorbic acid content.

Cyphomandra betacea (tree tomato or tamarillo) is a native of Peru and is also grown in India, Sri Lanka, and New Zealand. The fruit is greenish or purple in the early stages and turns reddish at maturity. It has a musky acid taste and tomato-like flavor. The fruit rind is rough with a disagreeable flavor. Several species in the genus *Solanum* are prominent vegetable crops. *Solanum melongena* (eggplant or aubergine) is a popular and staple vegetable crop in Asia, accounting for 92% of the total world production; in Africa

Table 3 Tomato varieties and end uses

Cultivar suitability	
Home use	*Shipping*
Better Boy F_1	Burpees' Big Boy F_1
Red Cherry	Campbell 1327
Market	Empire F_1
Earypark 707	Glamour
Empire F_1	Heinz 1350, 1439
Mountain price F_1	Jet Star F_1
Pik Red F_1	Jubilee
Traveler 76	Morton Hybrid F_1
Processing	Pik Red F_1
Heinz 722, 2653	Rutgers 39
Roma VF	Spring set F_1
VF 134-1-2	Supersonic F_1
UC 82, 204	

Adapted from Madhavi DL and Salunke DK (1998) Tomato In: Salunkhe DK and Kadam SS (eds), pp. *Handbook of Vegetable Science and Technology – Production, Composition, Storage and Processing*, pp. 171–201. New York: Marcel Dekker.

and Europe, its production is 3.7 and 3.3%, respectively (Table 2). The leading country of its cultivation is China, followed by India, Turkey, Japan, Indonesia, and Philippines. It is an extensively grown perennial crop in the tropics and an annual in the temperate zones. The fruit can be round, globose, long, or pear-shaped. Most cultivars have purple to blackish skin, while some have a white–green or mottled green skin, with white flesh; in the white–green types, the mature fruit has yellow skin. There are regional preferences for the color and shape of the fruit; purple, large, more or less round fruits are preferred in North America. Wild eggplants with spiny and bitter fruits are found in India; of the several improved Indian cultivars, black beauty, long black, round purple, Pusa purple long, Pusa purple round, Pusa purple cluster, long purple and hybrids such as Vijay, S-1, S-4, S-5, and S-8 are popular. *Solanum melongena* has three main varieties; see Table 4.

The African eggplant (*S. macrocarpon*) is a perennial crop grown in the Ivory Coast. The fruit resembles the eggplant. Garden huckleberry (*S. nigrum*) is native to North America and has a wide distribution in temperate and tropical regions; the fruit is a berry about 6 mm in diameter, the color being green when unripe and red, yellow, or black when ripe. Jilo (*S. jilo*) is a major crop of Nigeria and a minor crop in Central and Southern Brazil. Immature green fruits with a spherical to oval shape (of 4 cm in diameter and 6 cm long) are harvested, and on ripening, the fruit develops an organish red color.

Naranjillo or Lulo (*S. quitoense*) is native to Ecuador and is cultivated in Ecuador, Peru, Colombia, and some Central American countries. The fruit is spherical, 3–5 cm in diameter, and with a yellow, rough skin when ripe. The pulp containing many white seeds is green and acidic in taste.

Pepino or melon pear (*S. muricatum*) is an ancient cultivated crop of the Andes, chiefly grown along the central coast of Peru and at elevations of 1000–3000 m from Colombia to Bolivia. It is also cultivated to a limited extent in northern Argentina, Chile, New Zealand, and Australia. The fruit is long, ovoid, or ellipsoidal, and the color varies from light green to pale yellow.

Capsicum (hot, chili, bell, or sweet pepper) is the second most important fruit of the Solanaceae, grown as a vegetable and condiment crop in tropical and subtropical regions of the world. The world production of this pepper is shown in Table 2. There are five major recognized cultivated species of *Capsicum* (Table 5) and seven botanical varieties of *C. annuum* (Table 6). The important species are *Capsicum annuam* (annual with flowers borne singly in leaf axils) and *C. frutescens* (perennial with flower clusters in leaf axils). The bell pepper or sweet pepper (*C. annuum*), which is either mildly pungent or nonpungent, has a thick pericarp and is used in flavoring vegetable preparations. The highly pungent fruit of *C. frutescens* has a thin, smooth pericarp and is used as a condiment. Paprika is a nonpungent type of pepper with a thick pericarp, selectively bred for color and flavor.

China, Korea, Indonesia, Sri Lanka, Pakistan, Japan, Mexico, Ethiopia, Spain, Italy, and Hungary are some of the principal chili-producing countries. The world demand for chilies is increasing. The main exporters of chili are India, China, Indonesia, Hungary, Singapore, Malaysia, Mexico, and Japan. The bell pepper, *C. annuum* var. *grossum* (syn. Simla

Table 4 Common cultivated varieties of *Solanum melongena* and their fruit characteristics

Variety	*Fruit characteristics*
Var. *esculentum* (Wees)	Large, pendent, ovoid, oblong berries, 5–30 cm long, shiny
Var. *depressum* (Baily)	Small, pear-shaped fruits, purple in color, 10–13 cm long
Var. *serpentium* (Deeft)	Slender, greatly elongated, 30 cm long, 2–5 cm in diameter, end-curled

Adapted from Lawande KE and Chavan JK (1998) Eggplant (brinjal), In: Salunkhe DK and Kadam SS (eds), *Handbook of Vegetable Science and Technology – Production, Composition, Storage and Processing*, pp. 225–244. New York: Marcel Dekker. *Solanum melongena* var. *incarrum* (Linn) is a nonedible variety.

Table 5 Cultivated *Capsicum* species and their distribution

Species	*Synonyms*	*Distribution*
C. annuum L.	*C. purpureum* *C. grossum* *C. cerasiformae*	Columbia to southern USA, thoughout Asia and America
C. baccatum L.	*C. pendulum* *C. microcarpum* *C. angulosum*	Argentina, Bolivia, Brazil, Columbia, Equador, Peru, Paraguay, etc.
C. frutescens (Tabasco pepper)	*C. minimum*	Columbia, Costa Rica, Guatemala, Mexico, Puerto Rico, Venezuela
C. chinese L.	*C. luteum* *C. umbelicatum* *C. sinense*	Bolivia to Brazil, Costa Rica, Mexico, Nicaragua, West Indies
C. pubescence	*C. eximium* *C. tovari* *C. cardenasii*	Bolivia to Columbia, Costa Rica, Guatemala, Honduras, Mexico

Adapted from Rajput JC and Parulekar YR (1998) Capsicum In: Salunkhe DK and Kadam SS (eds) *Handbook of Vegetable Science and Technology – Production, Composition, Storage and Processing*, pp. 203–224. New York: Marcel Dekker.

Table 6 Botanical varieties of *C. annuum* and their fruit characters

Variety	*Common names*	*Fruit characteristics*
Var. *abbreviatum*, Fingerh	Wrinkled peppers	Ovate wrinkled fruits, about 5 cm long or less
Var. *accuminatum*, Fingerh		Linear, oblong, pungent, pointed fruits about 9 cm long
Var. *cerasiformae* (Miller) Irish	Cherry peppers	Globose, pungent fruits red, yellow, or purple in color, with firm flesh and 1.2–2.5 cm in diameter
Var. *conoides* (Miller)	Cone peppers (Tabasco type)	Erect, conical, pungent fruit, about 3 cm long
Var. *fasciculatum* (Stuart) Irish	Cluster peppers	Clustered, erect, slender, very pungent fruits, about 7.5 cm long
Var. *grossum* (L.) Sendt.	Sweet peppers, paprika	Large inflated fruit with a basal depression, red or yellow thick flesh with mild pungency
Var. *longum* (DC) Sendt.	Long peppers	Mostly dropping fruits with tapering apex

Adapted from Rajput JC and Parulekar YR (1998) Capsicum In: Salunkhe DK and Kadam SS (eds) *Handbook of Vegetable Science and Technology – Production, Composition, Storage and Processing*, pp. 203–224. New York: Marcel Dekker.

mirch, sweet pepper, bullnose capsicum) is cultivated widely in India, Central and South America, Bolivia, Peru, Costa Rica, Mexico, Hong Kong, and almost all European countries. Based on characteristics such as fruit size, color, texture, flavor, pungency, and uses, a 13-group classification has been proposed for cultivars of this species (Table 7).

The *Capsicum* fruit is a pod-like berry with a short, thick peduncle, developing from a bi-carpellary ovary with axile placentation. Many seeds are present in the cavity between the placenta and fruit wall. The unripe fruit is commonly green, but fruits with cream, greenish yellow, orange, purple, and purplish black colors also exist. Ripe fruits are usually red, but sometimes yellow or orange.

Chemical Composition and Nutritive Value

There is considerable variation in the chemical composition of the fruits of this family (Tables 8 and 9). The stage of maturity, cultivar, environmental, and cultural practices influences the chemical composition.

Tomato

Glucose and fructose are the principle sugars in tomato (representing more than 60% of the solids) with small amounts of sucrose. Ripe fruits also contain raffinose; the glucose concentration increases with ripening. Tomato being a climacteric fruit, ripening is associated with increase in respiration and ethylene production: these in turn are integrated with the disappearance of starch, degradation of chlorophyll, synthesis of lycopene, flavor components, and polygalacturonase. The pectic constituents mainly control the texture and firmness of the fruit. In the ripening process, the predominant protopectin in the green fruit decreases, and the pectin increases, making the fruit soft. The protein content is about 1.0% in ripe fruit, and all the essential amino acids except tryptophan are present; other amino acids identified are tyrosine, aspartic acid, glutamic acid, serine, glysine, α-aminobutyric acid, and pipecolic acid; glutamic and aspartic acids occur in greater concentrations.

Citric acid is the principle organic acid of tomato which gives the fruit its typical taste. Along with a

Table 7 Bell pepper groups, fruit character, and uses

Group	*Cultivars*	*Uses*	*Fruit characteristics*
Bell	California wonder Yellow wonder	Fresh market, salads, pitza, meatloaf, and canning	Large, nonpungent with thick flesh (7–12 cm × 5–10 cm)
Pimento	Pimento select	Fresh market, salads, soups, processed meat, and canning	Large, nonpungent, thick-walled (5–10 cm × 5–7 cm)
Squash or cheese	Cheese, antibois and gambo	Processing, canning, freezing, pickling, salad, and culinary purposes	Small to large, thick-walled, nonpungent, green or yellow to red (2.5–5 cm × 5–10 cm)
Ancho	Mild California, big jam, New Mexican chili	Dried, powdered, and culinary purposes	Large, heartshaped, thin walled, less pungent fruits (10–12 cm × 5–7 cm)
Anahein chili	California chili, paprika	Sauces and canning, processed into powder dehydrated pods	Slender, medium-thick flesh, medium to dark green color turning red, sweet to moderate pungency (12–15 cm × 2–3 cm)
Cayenne		For market, pickling, dry powder, sauces culinary purposes	Irregular, wrinkled, highly pungent fruits turning green to red (12–22 × 1.5–6 cm)
Cuben		Fresh market, salads, pickling, and frying	Long, thin walled, irregular fruits of mild pungency, turning yellowish green to red (10–20 cm × 1.5–5 cm)
Jalapeno	Jalapeno	Fresh green pods, dried powder, canning, and sauces	Elongated, round, cylindrical shape, highly pungent, thin-walled, red at maturity (5–7 cm × 2–5 cm)
Small hot	Red chili, sontaka	Dried powder and sauces	Slender, medium- to thick-walled, highly pungent, turning red (4–7 cm × 1–2 cm)
Cherry	Red cherry, large Red cherry, small	Pickling	Small, spherical, pungent fruit turning red (2–5 cm in diameter)
Short wax	Floraljam, caloro, cascabella	Pickle, cooking, sauce, and processing	Medium- to thick-walled, tapering, turning yellow to orange red (5–7 cm × 2–5 cm)
Long wax	Hungarian yellow wax, sweet banana	Fresh market, pickle, sauce, and canning	Pointed or blunt fruits turn yellow to red (8–12 cm × 1–4 cm)
Tabasco	Green leaf Tabasco, Tabasco	Vinegar, sauce, and pickles	Slender, highly pungent fruits, turning yellow to red (2–5 cm × 0.5 cm)

Adapted from Rajput JC and Parulekar YR (1998) Capsicum In: Salunkhe DK and Kadam SS (eds) *Handbook of Vegetable Science and Technology – Production, Composition, Storage and Processing*, pp. 203–224. New York: Marcel Dekker.

small amount of malic acid, traces of acetic, formic, lactic, and succinic acids have also been detected. The lipid fraction of the tomatoes is composed of triglycerides, sterols, sterol esters, free fatty acids, and hydrocarbons. The phenolics reported are caffeic, ferulic, chlorogenic, and *p*-coumaric acids. The chief coloring materials of tomato at the mature green stage are chlorophyll a and b, while at the ripe stage, carotene and lycopene contents dominate, contributing 7 and 87%, respectively. The final color of tomato depends upon the ratio of the carotene and lycopene. The various volatile compounds in ripe fruit are alcohols, aldehydes, carbonyls, and sulfur compounds, contributing to the flavor quality.

Tomato is regarded as an essential protective food. It is a rich source of ascorbic acid: On a fresh-weight basis, the vitamin C content averages about 25 mg per 100 g, varying with the variety, maturity, and season.

Table 8 Macronutrient content (per 100-g edible portion) of some solanaceous fruits

Crop	Energy (J)	Water (g)	Protein (g)	Fat (g)	Carbohydrates (g)
Tomato (green)	97	93.0	1.9	0.1	3.6
Tomato (ripe)	84	93.8	1.2	0.3	4.2
Eggplant	109	92.0	1.6	0.3	4.0
Chili pepper	487	65.4	6.3	1.4	24.8
Bell pepper	109	92.0	1.3	0.2	6.0
Tree tomato	202	85.9	1.5	0.3	11.3
Naranjillo	118	92.0	0.7	0.1	6.8
Tomatillo	134	92.0	0.4	1.0	6.3
Cape gooseberry	223	83.0	1.8	0.2	11.1
Husk tomato	105	92.0	0.7	0.6	5.8
Pepino	134	92.0	0.4	1.0	6.3

From FAO (1972) *Food Composition Table for Use in East Asia*. Rome: Food and Agriculture Organization; Gopalan C, Ramasastri BV and Balasubramanian SC (1993) *Nutritive Value of Indian Foods*. Hyderabad, India: Indian Council of Medical Research; Yamaguchi M (1983) *World Vegetables, Principles, Production and Nutritive Value*, pp. 291–310. Westport, CT: AVI.

Table 9 Vitamin and mineral content (per 100-g edible portion) of some solanaceous fruits

Crop	Vitamin (mg)[a]					Mineral (mg)			
	A (IU)	B_1	B_2	Nicotinic acid	C	Calcium	Iron	Magnesium	Phosphorus
Tomato (green)	320	0.07	0.01	0.4	31	20	1.8	15	36
Tomato (ripe)	385	0.06	0.04	0.6	23	7	0.6	12	30
Eggplant	124	0.08	0.07	0.7	6	22	0.9	16	37
Chili pepper	576	0.37	0.51	2.5	96	86	3.6	24	120
Bell pepper	530	0.07	0.08	0.8	103	12	0.9	13	34
Tree tomato		0.04	0.04	1.4	17	13	0.8	34	24
Naranjillo	170	0.06	0.04	1.5	65	8	0.4		14
Tomatillo	80	0.05	0.05	2.1	2	7	0.4	23	40
Cape gooseberry	2380	0.05	0.05	0.1	180	10	2.0	31	67
Husk tomato	380	0.05	0.02	2.1	0.2	7	0.4	23	40
Pepino	200	0.08	0.04	0.5	32	18	0.8		14

[a]Vitamin A content is given in IU.

From FAO (1972) *Food Composition Table for Use in East Asia* Rome: Food and Agriculture Organization; Gopalan C, Ramasastri BV and Balasubramanian SC (1993) *Nutritive Value of Indian Foods*. Hyderabad, India: Indian Council of Medical Research; Yamaguchi M (1983) *World Vegetables, Principles, Production and Nutritive Value*, pp. 291–310. Westport, CT: AVI.

Research has shown that the inclusion of tomato in the diet can prevent deficiencies of vitamins and minerals. One medium-size, raw tomato provides 47% of the B vitamin, 33% of vitamin A, and 1% of the energy of the daily dietary requirement of an average person. The mineral content varies between 0.3 and 0.6%; apart from calcium, potassium, magnesium and iron, the tomato contains zinc, boron, iodine, cobalt, and aluminum in trace amounts.

The tomato contains a glycoalkaloid, tomatine, and traces of solanine. Narcotine is present in unripe fruit. The tomatine content is lowest in the pink stage of ripeness and increases slightly in the fully ripe tomato.

Brinjal or Eggplant

This contains about 4% carbohydrate, composed of sucrose, glucose, and fructose. Brinjal protein has a high biological value and contains amino acids such as arginine, histidine, lysine, tryptophan, leucine, isoleucine, and valine, although the protein is relatively poor in lysine and tryptophan, isoleucine, and methionine. Brinjal contains a higher percentage of vitamins than many other vegetables. However, it is not a rich source of vitamin B_2. The physical characteristics such as the shape, color, and presence of spines on the calyx have also been found to influence the composition; hence the green, white, and purple varieties are often compared for chemical composition. Dark-purple-skinned varieties contain more vitamins than those with white skins. The white cultivar contains twice as much crude fiber as the purple and green cultivars. The amino acid content is higher in purple cultivars and lowest in the white. Purple cultivars are poor in potassium and chloride content, while green cultivars are rich in these minerals. The polyphenol oxidase activity is highest in the purple type, thus making the fruit turn dark faster when cut surfaces are exposed to air. Variations in the activities of other enzymes have been observed in differently

colored fruits. The main pigment of the fruit is an anthocyanine, nasunin; lycopene and lycoxanthin are also present. The seed oil is reported to be rich in linoleic acid.

The fruit peel contains a bitter principle, solasonine. The phenolic compounds present in the fruit are chlorogenic acid, neochlorogenic acid, scopoletin, and caffeic acid; a trace amount of hydrocyanic acid is also present. Higher concentrations of glycoalkaloids (20 mg per 100 g fresh weight) result in a bitter taste and off-flavor.

Pepper

Pepper is a good source of vitamins A and C, and is superior to tomato and eggplant in this respect. The vitamin C content varies with the variety and maturity stage (160–210 mg in green chilies and 113–160 mg in red ripe fruits). The pungent principle of chili is capsaicin (a crystalline, colorless compound), which is concentrated more in the placenta (90%) that connects the seeds with pericarp. A highly irritating vapor is liberated on heating capsaicin. Yellow chilies are more suitable for capsaicin extraction. An African chili variety, 'Mombusa,' is known for its high capsaicin content. The green fruits contain chlorophyll a and b. The coloring matter of ripe fruit includes capsanthin, capsorubin, zeacanthin, cryptoxanthin, and α- and β-carotene.

Other Solanaceous Fruits

Husk tomato, tomatillo, and cape gooseberry are also important sources of vitamins and minerals (Table 9). Tree tomato contains substantial amounts of carbohydrates and minerals. It is rich in vitamin C and is also a good source of provitamin A.

Handling and Storage

Tomato

Tomato fruits are harvested at different stages of maturity (Table 10), depending upon the use, and are often harvested at the mature green stage for the market and ripened either in transit or during storage. For canning or juice extraction, the fruit is harvested at the ripe stage. The postharvest losses in tomato may range from 5 to 50%; the highest loss is attributable to mechanical damage, and other causes may be physiological disorder, diseases, and heat and chilling injury.

Packing-house operations include cleaning, grading, waxing, wrapping, and packing. Fruits are transported to the market in bamboo baskets, corrugated board cartons, or wooden boxes. In general, the optimum storage temperature and relative humidity (RH) for tomato storage is 12 °C and 86–95%. However, depending upon the stage of fruit maturity at harvest, storage conditions vary. Mature green tomatoes can be stored at 13–18 °C and 85–90% RH for 3–4 weeks as the fruits ripen at this temperature without chilling injury. Holding them below 10 °C for more than 24 h should be avoided to overcome the symptoms of chilling injury. However, storing green fruits above 18 °C results in early ripening, which leads to rapid deterioration in the quality of the fruits, while at about 30 °C, the red pigment formation is hindered, and the ripe fruit develops an orange to yellow color. Fruits at the breaker stage are less sensitive to chilling injury than immature green fruits. Ripe tomatoes can be preserved for 7–10 days at 3–5 °C and 85–90% RH. Fully ripened fruits are held between 1.7 and 4.4 °C until they are ready for use. Prolonged storage of fully ripe tomato causes softening and loss of color of the fruits.

Controlled-atmosphere and hypobaric storage can also prolong the shelf-life of the tomato. The packaging of tomatoes in polyethylene bags provides a modified atmosphere, which extends the shelf-life. Postharvest treatment with gibberellic acid markedly retards ripening, while 2,4-dichlorophenoxyacetic acid, naphthalene acetic acid, and ethylene hasten ripening, and this facilitates the marketing of mature green tomatoes. Ethephon, an ethylene-releasing compound, has been used commercially to accelerate

Table 10 Maturity stages of tomato

Stage	*Characteristics*
Immature green	Seeds and jelly-like substance surrounding the seeds not yet formed
Mature green	Fully grown fruit shows a brownish ring on the stem scar, and blossom end turns yellowish green; seeds are surrounded by a jelly-like substance
Turning (breaker stage)	One-quarter of the surface at blossom end appears pink
Pink	Three-quarters of the surface appears pink
Hard ripe	Nearly all red or pink, but flesh is firm
Over-ripe	Fully red and soft

Adapted from Gould WA (1992) *Tomato Production, Processing and Technology*, 3rd edn. Baltimore, MD: CTI.

ripening. Chemicals such as captan, dithane, and thiram have the greatest potential as fungicidal treatments on tomatoes.

Brinjal

Brinjal is a heavy vegetable and requires care in handling, even though it is not as perishable like the tomato. The fruit becomes edible from the time it attains one-third or the full growth size. Over-mature fruits are dull, seedy, and fibrous, sometimes tasting bitter. After harvest, the graded fruits are packed in special crates or bushel baskets. In tropical countries, chilling injury and certain pests and diseases cause high losses of eggplants. Temperatures below 10 °C cause chilling injury, the symptoms of which are pitting, surface bronzing, and browning of the seeds and pulp. Fruit rot and anthracnose caused by *Phomopsis parasitica* and *Gleosporium melongense*, respectively, are the most serious postharvest diseases. Hot water (50 °C) and a low concentration of dithiocarbamate provide good disease control. The fruit can be stored successfully for 2–3 weeks at 10.0–12.8 °C and 92% RH. Controlled-atmosphere storage has not been found to be beneficial in extending the storage life of the eggplant. Small fruits are not well suited for long-term storage, and chilling injury is more common in tender fruits. Prepackaging of brinjal in adequately ventilated polyethylene bags (100-gauge) increases the shelf-life, and treating brinjal with a fungicidal wax emulsion extends the shelf-life by 30–40%.

Pepper

Chili, used as a raw vegetable, is harvested green, but for processing, only red ripe chilies are harvested. The important postharvest losses in chili are weight loss, chilling injury, and microbial spoilage. At the recommended storage conditions of 7.2 °C and 85–90% RH, green peppers can be stored for 2–3 weeks, and ripe peppers at 5.6–7.2 °C and 90–95% RH can be stored for about 2 weeks. Storage in paper bags rather than in polyethylene bags can prolong the shelf-life of hot peppers. Bell peppers are harvested when they attain their full size and are still green. Storage below 7 °C results in chilling injury and, subsequently, *Alternaria* rot. Controlled-atmosphere storage at various temperatures has been recommended to extend the shelf-life of bell peppers, while hypobaric storage is not beneficial as it greatly increases the weight loss. Fruits are commonly wax-coated to reduce the moisture loss before shipment. Ripe fruit can be preserved at room temperature (20 °C) for only a week, but for 12–14 weeks at 3–4 °C after dipping in hot water at 50 °C for 10 min. A common storage disease is bacterial decay by *Erwinia carotovora*.

Uses of Solanaceous Fruits

Fresh, ripe tomato fruits are refreshing and appetizing, and are consumed raw in salads or after cooking. Unripe fruits are usually cooked and eaten. Large quantities are processed in southern European countries and in California. Tomatoes can be processed into juice, purée, paste, ketchup, sauce, soup, and powder. Special varieties have been evolved to meet these processing requirements. The tomato cultivars for juice should have a bright color, rich flavor, and high total acidity. They should also be juicy and not meaty. Tomatine, a glycoalkaloid present in the fruit, yields, on hydrolysis, tomatidine, which can be transformed into hormones such as progesterone and testosterone. The oil of tomato seeds is used in the soap and paint industry. The pressed cake can be used as a feed for livestock or a fertilizer. Tomato also has several medicinal applications and is used in the preparation of traditional medicine in Japan, Greece, and Peru. The hot water extract of dried fruits has been used in the treatment of ulcer wounds and burns, and tomato has been reported to reduce the risk of prostrate and digestive tract cancers.

Brinjal is valued more as a vegetable during the autumn when other vegetables are scarce. It is a fairly good source of calcium, iron, phosphorus, and vitamin B. Sun-dried or hot-air-dried eggplant powder with good organoleptic and nutritive qualities has been prepared. Treatment with brinjal is recommended in liver complaints. It has been reported to stimulate the intrahepatic metabolism of cholesterol and produce a marked drop in blood cholesterol level.

The garden huckleberry (*S. nigrum*) leaf has great medicinal value. The tender leaves are boiled like spinach and eaten in many parts of India. Ripe fruits are used as a substitute for raisins in plum puddings. The berry is considered to possess tonic, diuretic, and cathartic properties, and is useful in the treatment of heart disease. The berries also find use as a domestic remedy for fever, diarrhea, ulcers, and eye problems.

Peppers are used green, or after ripening, in a great variety of ways. Chilies, when taken with food, stimulate the taste buds and increase the flow of saliva. The commercial red pepper(s) consists of the fruits of a small, pungent variety, which is dried and ground to a fine powder. Dry chili powder and green chilies are commonly used in various culinary preparations and in the preparation of sauces, soup, ketchup, and salads. Pepper sauce, prepared in a variety of ways, consists of the fruit of pungent varieties preserved in brine or strong vinegar. Both red and

green bell peppers are sold frozen and are also canned, along with other vegetables. Tabasco sauce is prepared from the juice of a pungent variety, expressed by applying pressure. Paprika, a Hungarian condiment, is made from fruit ground after removing the seeds. Pepper is used in pickles of various kinds. The sweet varieties are stuffed and baked. *Capsicum* preparations are used as a counterirritant in neuralgia and rheumatic disorders. Green chilies are a good source of L-asparaginase, which has an antitumor effect and is used in the treatment of acute lymphocytic leukemia.

The oleoresins have varying color properties and pungency, are prepared from the dried fruits of *Capsicum*, and are used in the pharmaceutical, food, and cosmetic industries. The three types of oleoresins are oleoresin capsicum, oleoresin red pepper, and oleoresin paprika; oleoresin capsicum is used in making pain-balms and vaporubs. The nonpungent type of oleoresin is chiefly used as a coloring agent, while the pungent type is used for flavoring and coloring snack foods. Chili seed cake, a byproduct of seed oil, has a protein content of 27–29% and can be used as a fertilizer or as an animal feed. Chili color is a new product, which can replace artificial food colors.

The melon pear fruit is cooked and eaten when immature, while the mature fruit has a fleshy pulp with an aroma and flavor similar to those the cucumber. Lulo is cultivated mainly for the juice, especially in Ecuador, Peru, Columbia, and some Central American countries. The young shoots of jilo are finely chopped and used in soups in Nigeria. Tomatillo can be used to make chili sauce and is used for dressing meats in Mexico. The ground cherry, which is acidic-sweet, is used for preserves and sometimes for sauces, and the Cape gooseberry is eaten raw or preserved as a pickle.

See also: **Peppers and Chillies**; **Potatoes and Related Crops**: The Root Crop and its Uses; Processing Potato Tubers; **Tomatoes**

Further Reading

Bajaj KL, Bansal KD, Chadha ML and Kaur PP (1990) Chemical composition of some important varieties of eggplant (*Solanum melongena* L.). *Tropical Science* 30: 255.

Chadha YR (1972) *The Wealth of India, A Dictionary of Indian Raw Materials and Industrial* products, vol. IX, pp. 383–393. New Delhi: CSIR.

Dighe AH (1995) *Studies on Biochemical and Nutritional Composition of Promising Brinjal* (Solanum melongena L.) *Cultivars*. M.Sc. thesis, Mahathma Phule Agricultural University, Rahuri, India.

Gopalan C, Ramasastri BV and Balasubramanian SC (1993) *Nutritive Value of Indian Foods*. Hyderabad, India: Indian Council of Medical Research.

Gould WA (1992) *Tomato Production, Processing and Technology*, 3rd edn. Baltimore, MD: CTI.

Habson G and Greirson D (1993) Tomato. In: Seymour GB, Taylor JE and Tucker GA (eds) *Biochemistry of Fruit Ripening*, p. 405. London: Chapman & Hall.

Lawande KE and Chavan JK (1998) Eggplant (brinjal). In: Salunkhe DK and Kadam SS (eds) *Handbook of Vegetable Science and Technology – Production, Composition, Storage and Processing*, pp. 225–244. New York: Marcel Dekker.

Madhavi DL and Salunke DK (1998) Tomato. In: Salunkhe DK and Kadam SS (eds) *Handbook of Vegetable Science and Technology – Production, Composition, Storage and Processing*, pp. 171–201. New York: Marcel Dekker.

Pruti JS (1998) *Major Spices of India – Crop Management and Post-harvest Technology*, pp. 180–238. New Delhi, India: Indian Council of Agricultural Research.

Rajput JC and Parulekar YR (1998) Capsicum. In: Salunkhe DK and Kadam SS (eds) *Handbook of Vegetable Science and Technology – Production, Composition, Storage and Processing*, pp. 203–224. New York: Marcel Dekker.

Salunkhe DK and Desai BB (1984) *Postharvest Biotechnology of Vegetables*, vol. 1, pp. 39–82. Boca Raton, FL: CRC Press.

Wall MM, Waddell CA and Bosland PW (2001) Variation in β-carotene and total carotenoid content in fruits of capsicum. *Hortscience* 36(4): 746–749.

Yamaguchi M (1983) *World Vegetables, Principles, Production and Nutritive Value*, pp. 291–310. Westport, CT: AVI.

Processing Potato Tubers

M G Lindhauer, N U Haase and B Putz, Federal Centre for Cereal, Potato and Lipid Research, Detmold, Germany

Background

The convenience aspect of processed potatoes (*Solanum tuberosum* ssp. *tuberosum* L.) has become of significant importance in the total use of potatoes by humans. Especially in some industrialized countries, the percentage of potato products in relation to total use of potatoes as foodstuff has reached a value of up to 65% (e.g., in the USA). To get an overview, the main directions of potato processing are discussed in this article.

Raw Material for Processing

Potato processing requires special quality profiles of the raw material. Therefore, the number of varieties suitable for processing is limited. Plant breeders make a great effort to select new varieties with better agronomical behavior and with higher internal quality. Today, only a few varieties with a known processing quality are grown worldwide.

Several specific aspects must be observed:

- The oxidative potential of the harvested tubers should be low to prevent discoloration during the first processing steps.
- Key constituents like protein, minerals, and vitamins should be high because of inevitable losses during processing. Also the total content of solids should be high to make processing more worthwhile, by achieving an higher input-to-output ratio.

Typical nutrient contents of selected potato products are listed in Table 1.

Storage Aspects

Storage has to provide high-quality potato tubers to the processing plant for several months. Good storage management should prevent excessive loss of moisture, development of rot, and excessive sprout growth. Also the loss of potato constituents (e.g., ascorbic acid) should be low. Finally, the accumulation of reducing sugars must be controlled to prevent dark-colored products.

After harvest, potatoes pass through a period of several months without any sprouting. The following sprout growth is little or absent at temperatures below 5 °C. However, reducing sugars accumulate in excessive quantities, resulting in undesirable dark-brown colored products. Therefore, potatoes for processing are stored at 8–10 °C. At that temperature sprouting must be controlled. Experiments have been conducted with gamma-irradiation of tubers or treatment of vines with maleic hydrazide. Also several chemical compounds are known to inhibit sprouting (e.g., ethylene, volatile monoterpenes, ethanol, abscisic acid, and naphthalene), but in practice isopropyl-*N*-(3-chlorophenyl) carbamate (CIPC) is preferred. To prevent the use of antisprouting agents, potato breeders have created cold storable varieties without cold sweetening to avoid the problem described.

Immature tubers with considerable amounts of free sugars will not reduce that level during storage. Also storehouses with little gas exchange capacity lead to high contents of reducing sugars inside tubers. This is because of the increasing carbon dioxide concentration in the storage atmosphere. An increase of carbon dioxide towards 0.5% in the atmosphere already shows a significant deterioration of baking color. Depending on the geographical region, storehouse equipment must include a mechanical ventilation and cooler. The relative humidity should be 90% minimum.

Potato Processing for Food

A schematic overview of several processing techniques is given in Figure 1.

Common Techniques

Washing Washing systems remove dirt and attached soil from the tubers. Stones can be removed through stone catchers or through exploitation of the different specific density in comparison to tubers. Often the washing water is taken from the overflow of waste water decanters within processing.

Peeling With few exceptions, all tubers are peeled before processing. Abrasive and steam peelers are common, while in the past lye (caustic) peelers were also used. Because of concern about environmental effects, the importance of that type of peeler has decreased. In abrasive peelers, designed with carborund layers or knives or combinations thereof, the amount of peeling waste can be up to 60% of tuber weight, if the tuber shape is irregular or deep potato eyes exist. Most single cells are disrupted and oxidative enzymatic discoloration must be avoided (often by additives, such as sulfur). Steam peelers remove much less of the tuber cortex. Most of the tissue with high nutritive value achieves further processing.

Table 1 Typical nutrient concentrations of processed potatoes

	Per 100 g edible material					
	Energy (kJ) (kcal)	*Main constituents*				
		Protein (g)	*Fat (g)*	*Carbohydrates (g)*	*Potassium (mg)*	*Ascorbic acid (mg)*
Cooked potatoes	298 (71)	2.0	traces	14.8	443	14
Potato chips (crisps)	2241 (536)	5.5	39.4	40.6	1000	8
French fries	1215 (290)	4.2	14.5	35.7	926	28
Mashed potatoes	314 (75)	2.0	1.9	12.2	259	9

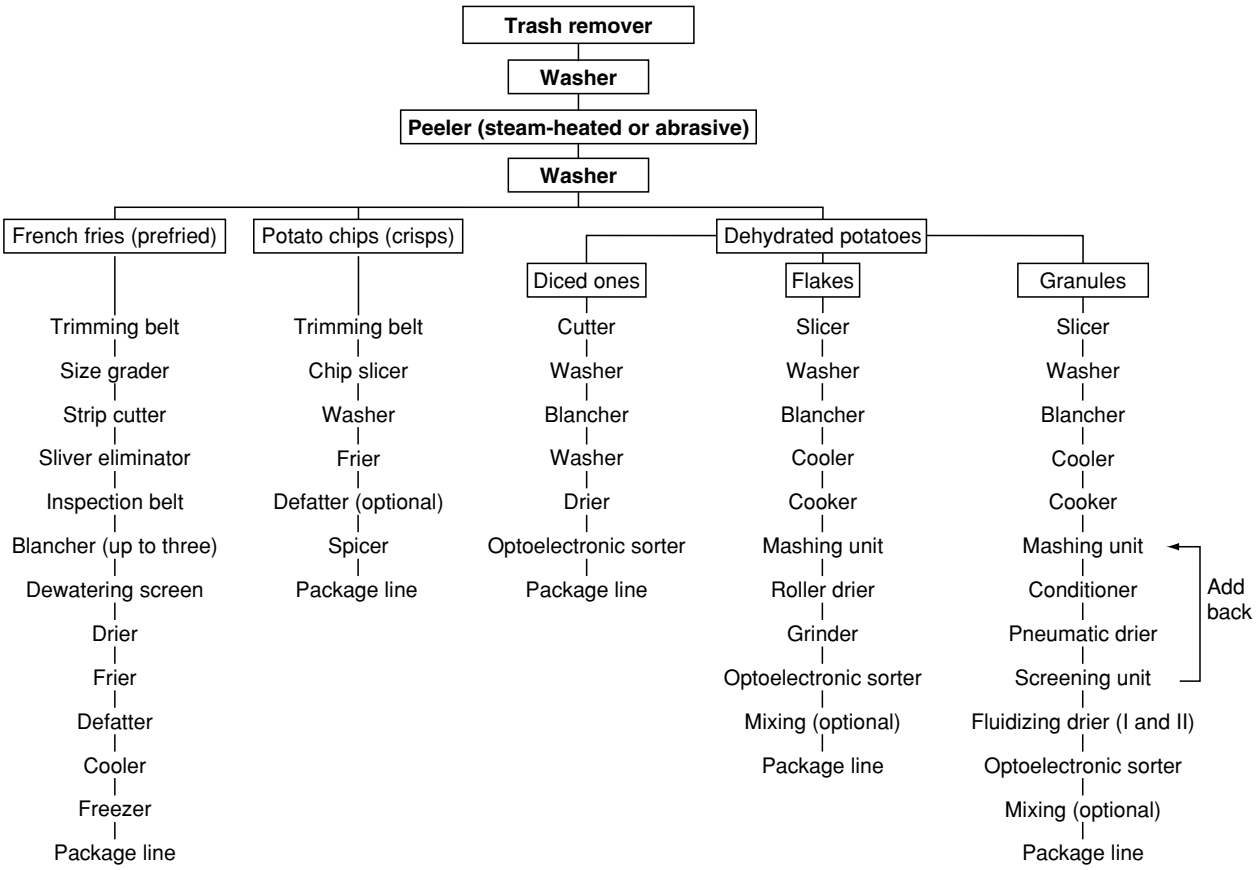

Figure 1 Schematic description of main processing lines.

Tubers come into contact with high-pressure steam (up to 15 bar) for a few seconds. The outer cell layers are softened, and the cohesion of the tissue is loosened. Through fast relaxation of the vessel the peel is separated from the tubers and can be removed by brushes or by a hard water jet. The main disadvantage is the layer of cooked tissue. Therefore, the preferred application is when making French fries and dehydrated products, whereas potato chips (crisps) are predominantly produced using abrasive peelers.

Cutting Several types of cutting exist depending on final use. The stream of tubers is directed to specially designed knives with the help of centrifugal power or water pressure. High-pressure water knives need precise positioning of tubers. Except for when slicing potato chips (crisps), using smooth and crinkled knives, cutting is carried out in several steps. Disks are cut first. Then, stripes are cut with rotating knives. If cubes are to be manufactured, a third cutting is done using chopping knives. Stripes for French fries production can be cut using water knives. High-pressure water leads tubers through a pipe with a cutting fence at its end.

Sorting Irregular cuts must be removed, if the shape of single pieces is a quality criterion, e.g., when manufacturing French fries, and thus is done before and after the main processing. Separation takes place by screening or by roller sorters. Optoelectronic sorters allow an additional sorting according to discoloration. Single pieces are visually inspected by charge-coupled device (CCD) cameras. Next to light and dark differentiation, state-of-the-art systems also offer the possibility of color differentiation. Discolored pieces are removed from the main stream by an ejection system using compressed air.

Blanching Blanching means a short-term heat treatment (70–100 °C) of raw potatoes or potato pieces. Several reactions occur. Enzymes are inactivated (e.g., polyphenoloxidase) to prevent discoloration. Often the peroxidase system is regarded as a key enzyme mirroring the efficiency of enzyme inhibition. Reducing sugars are leached off, to minimize the Maillard reaction, as well as starch pastes, to reduce fat uptake during frying. Blanching for about 20 min at 70–75 °C results in a firm structure. Native pectin methyl esterase is activated to reduce cross-linking

of pectin, and free carboxyl groups may react with calcium or magnesium liberated from starch granules after gelatinization to form a thermostable pectin network. The addition of calcium to the blanching water intensifies that reaction.

Blanching takes place predominantly in a continuous screw motion through a water bath or through a steam chamber. Depending on the raw material and the shape of single pieces, blanching parameters vary widely.

Dehydrated Potato Products

Dehydrated diced potatoes Dehydrated diced potatoes have been produced since the beginning of the twentieth century. Their first use was as a staple food. Today, snack industry and other food manufacturers (e.g., for dumplings, hash brown) are the main users of dehydrated diced potatoes. Because of different outlets, no uniform quality level can be addressed, but discoloration is to be prevented, flavor should be fairly typical of the varietal characteristics, and texture of the reconstituted product should be similar to that of fresh prepared potatoes.

Processing Preliminary inspection by hand or by optoelectronic sorters immediately after washing removes potatoes that are unfit for processing.

Standard peeling equipment uses steam or abrasive peeling. The expense of the inspection and trimming of peeled tubers can be reduced by optoelectronic inspection belts.

Mechanical dicers which can be adjusted to various cutting sizes (e.g., cubes, disks) cut the tubers. Irregular cuts depend on the sharpness of the cutter blades and on the mass flow of potatoes. Released solids (up to 15%) cannot be avoided but should be reduced as far as possible. Often, free starch is separated from the wash water and is used for special purposes.

Blanching for 2–10 min at 90–100 °C reduces adverse enzymatic reactions. Under- and overblanching negatively affects the texture and appearance of the finished product. Therefore it is to be controlled periodically. After blanching, gelatinized starch must be removed from the surface by a further water bath or a rinsing spray.

Sulfiting during or just after blanching protects against nonenzymatic browning reaction (darkening after cooking) during drying and reduces adverse oxidation during storage. Chemicals to be used are sodium sulfite, sodium bisulfite, sodium metabisulfite or combinations thereof. The European Community tolerates up to 400 mg sulfite as sulfur dioxide per kg ground dehydrated potatoes.

Potato drying is a two-stage process by migration of water to the surface and by removal of surface water by evaporation. Slow dehydration at a low temperature (e.g., 30 °C) gives a hard dense product, while a higher temperature results in a more porous material with better rehydration capacity. Hot air driers, constructed as hordes drier with a batch charging, or as a continuous-belt drier (conveyor drier) are used in most cases. During the first stage of drying moisture is readily removed and the drying rate is high. Therefore, the temperature can be high (70–90 °C). Within the second stage of drying, progress is much less and temperature is reduced to 50–70 °C. Potato size, shape, and raw material characteristics influence the efficiency of water removal. Microwave-assisted drying reduces the drying time drastically. The rapid diffusion of moisture through the surface will tend to eliminate surface hardening. Rehydration potential rises. It is most suitable at moisture levels between 6 and 30%, but often energetic and economic disadvantages prevent utilization.

Screening and inspection by pneumatic separators follows drying, to remove small pieces, discolored pieces or peeling left on potato dices, and black spots.

Dehydrated mashed potatoes

Dehydrated flakes Potato flakes are dehydrated mashed potatoes made by drum drying a thin layer of cooked and mashed potatoes, and breaking the sheet of dehydrated solids into a suitable size for packaging. Next to retail consumption in a pure state or prepared with milk solids and other constituents, flakes are taken as a food ingredient for potato snacks (e.g., potato chips made from a dough with uniform size and curvature, and for dehydrated French fries). A finely ground potato flour for soups and baked goods is also available.

Processing Raw potatoes should be washed to remove adhering soil, and sorted (by hand or by optoelectronic sorters) to remove defective tubers.

Peeling may be done by any process described above. In most cases steam peeling is used. Trimming may ignore some peel and defects, because of a final product control by optoelectronic sorters.

The tubers are sliced into disks of 1 cm thickness to obtain a more uniform heat treatment during precooking and cooling. Thinner slicing is possible, but flavor deteriorates and loss of solids clearly increases.

Precooking leads to gelatinization of starch and enzyme inactivation. Cooling below 20 °C for about 20 min reduces the level of free starch in the final flakes. For a long time that effect was described as retrogradation of gelatinized starch, but that reaction needs several hours to become visible. More evidence is given in toughing cell wall components (e.g.,

reduced pectin esterification) for an increased resistance against shear forces during drying.

For cooking, steam of atmospheric pressure is injected into a screw conveyor. To determine the proper degree of cooking, time must be evaluated periodically. Overcooking results in a poorer texture of the finished product, whereas undercooking causes excessive loss of mash from the applicator rolls and a high level of defects.

Mashing immediately follows cooking to avoid cell rupture of cooled and firmed potatoes. Also mixing with additives is easier in the hot stage. Next to emulsifiers (0.5% monoglyceride, to complex free gelatinized starch), chelating agents like sodium acid pyrophosphate (which reacts with iron, preventing darkening after cooking) and antioxidants like butylated hydroxyanisole (which extends storage life and provides reduced rancidity) are used widely. Also sulfur can be added to prevent nonenzymatic browning. The proper dosage of additives is difficult because of high losses during drying (up to 75% of antioxidants is lost and 50% of sulfur dioxide).

Drying takes place on a single-drum drier fitted out with four to six applicator rolls. The drum (e.g., 1.5 m diameter and 5 m length) is steam-heated. The temperature on the surface is between 140 and 160 °C. The mash is applied above the top roll while the direction of drum rotation is downward. Therefore, the distance between the top roll and the drum surface determines the rate of flow of mash to the rolls below. Each applicator roll applies a layer of potato cells to the drum. The distance of the applicator rolls to the surface of the drum must be controlled very carefully, because cell rupture is influenced by distance. A wider distance may result in a nonuniform deposition of potato cells and lumps from larger agglomerates can be in the product. Finally, a knife removes the dry sheet. It drops down into a screw conveyor which transfers the roughly broken sheet to a collecting tank. Then it is ground with hammer mills, creating a product with acceptable texture. High amounts of fines is necessary to prevent a sticky reconstituted product. The shelf-life depends on the atmosphere in the final products packages. Autoxidation of free fatty acids remains a problem.

Dehydrated granules Potato granules are dehydrated single cells or aggregates of cells dried to about 6–7% moisture after prior cooking. The principal process used is the 'add-back' process. Potatoes are partially dried by adding back previously dried granules to give a moist mix, which is air-dried after conditioning. The result is potato flakes, but the product is much less sticky, because most cells remain intact.

Processing The first steps of processing are as for flake processing. Raw potatoes are washed and sorted (by hand or by optoelectronic sorters) before steam peeling. Trimming removes the rest of peel and the defects, and slicing results in disks with 1–2 cm thickness to obtain a uniform heat treatment.

During precooking sulfur dioxide is added to prevent nonenzymatic browning.

Water cooling below room temperature for about 20 min reduces the level of broken cells in the final product.

Steam cooking at atmospheric pressure in a screw conveyor for about 30–50 min gives fully cooked potatoes ready for mashing and mixing with dry add-back granules. Repeated mild shearing and pressing of the added granules against the cooked mash causes separation of the latter into single cells with a low content of broken cells.

The moist mix is cooled to approximately 15–25 °C and held for about 1 h at that temperature (conditioning) by moving the moist mix on rubber belts with a depth of product of 10–20 cm, to avoid excessive compression. Since some agglomeration does occur during conditioning, gentle mixing is advised after that. Very large agglomerates and bruised tissue can be separated by a scalping reed. Poor granulation is a self-perpetuating process. Therefore, the amount of large granules will increase progressively with continued recycling. Because the larger particles do not absorb moisture rapidly enough, the process is to be controlled periodically.

Flash drying with a heated air stream, an inlet temperature of 170 °C, and an outlet temperature of 80 °C reduces the moisture content to about 12%. Material coarser than 60–80 mesh is screened off and returns to the process as add-back. Further drying takes place in a two-chamber fluidizing drier with alternate hot- and cold-air streams towards a final moisture content of about 6%.

For retail purposes the granules can be agglomerated with fat, milk solids, and flavors to give a ready-to-eat product. In that case only rehydration with hot water is necessary.

French Fries

French fries are deep-fat-fried potato strips. Commercially distributed French fries (UK: chips) are semifinished, and must be fried or heated before consumption (par-fried = partially fried). Shelf-life depends on the kind of distribution, either fresh (cooled) or deep-frozen. With respect to the desirable quality profile, large and long oval-shaped tubers (50 mm up) are preferred.

Processing Sorted tubers are cleaned. Subsequent peeling is with steam peelers. After removal of the

skin, strips are cut either by mechanical knife systems or by water knives (e.g., 6×6, 9×9, 11×11, 13×13 mm). In an inspection unit small pieces and defects are removed. The total amount of waste can be up to 50%, but most rejected material is processed to other products, e.g., pommes croquettes. In a washing unit free starch and other solids are washed off. Blanching reduces enzyme activity. A layer of pasted starch reduces fat uptake during frying. To prevent nonenzymatic discoloration, sulfur is often added to the blanching water. After washing, sometimes a second blancher with a low concentration of glucose is implemented to adjust the content of reducing sugars for a more stable frying color. Also dipping in a water bath containing a coating substance occurs, to enhance the texture stability for about 5–10 min after final frying. Predriers remove free water from the surface and the outer layer. Fat uptake will be reduced by the higher dry-matter content. Frying units can have a capacity of $4\,t\,h^{-1}$ of product, or more. The fat content of such units can be more than 5000 l of fat. Temperature is between 140 and 180 °C. The potato strips are moved through the oil bath by conveyor belts. To enhance texture behavior and capacity, two frying units can work in a line. Normally, the total frying time is between 3 and 5 min. After frying a defatting unit removes free fat from the surface by vibration or by hot-air streams. Mechanical cooling reduces the temperature of prefried strips to 4 °C. That product can be packaged and distributed as par-fried fresh French fries. Alternatively, the French fries are frozen inside tunnel freezers, packaged, and distributed as par-fried deep-frozen French fries.

Potato Chips

Potato chips (UK: crisps) are thin slices (1.1–1.6 mm) of fresh potatoes which have been normally peeled and fried in hot oil to reduce the water content to less than 2%. For processing, special varieties with a very low content of reducing sugars (< 0.15% fresh weight) are required.

Processing After cleaning and visual inspection, the tubers are peeled abrasively with losses up to 25%. Thin slices are cut with smooth or crinkled knives. In a washing unit free starch and other solids of disrupted cells are washed off. Optionally, blanching may reduce the content of reducing sugars. Predriers remove free water from the surface and the outer layer. Frying units can have a capacity of $4\,t\,h^{-1}$ of product or greater. The oil content of such units can be more than 5000 l of oil. Temperature is up to 190 °C. The potato slices are of lower specific density than the oil. Therefore specially designed rotating paddles dip and move the slices forward. Total frying time 2–4 min. The slices absorb up to 50% oil. That high amount of oil must be continuously replaced. After frying, a defatting unit removes free fat from the surface by vibration or by hot-air streams. The hot slices are seasoned and packaged.

Others

Peeled potatoes Peeled table potatoes are produced with abrasive peelers, either carborund or knives or combinations thereof. Shelf-life depends on the hygienic status. Preservation with chemicals (e.g., sulfur) leads to a shelf-life of 8 days, whereas a substitution of air by nitrogen or carbon dioxide or combinations thereof leads to a shelf-life of 10 days.

Potato dough Several potato products prepared of dough are available, either fresh or deep-frozen (e.g. pommes croquettes, pommes duchesses). Dough preparation of cooked potatoes and several other ingredients can be from fresh potatoes (e.g., small pieces from French fries production) or from dehydrated potatoes. A final frying or heating is necessary.

Sterilized potato products Peeled potatoes are available as canned potatoes. After peeling, diced or whole tubers are blanched with water at 85 °C, filled in cans and covered with brine, sealed, and sterilized at 121 °C. In addition to table potatoes, other potato products are also sealed in plastic bags and sterilized.

Potato Processing for Industrial Purposes

Ethanol

Ethanol originates from sugar fermentation of yeast (*Saccharomyces cerevisiae*). Processing of ethanol production integrates these steps: first starch is broken down by steaming the potatoes. After mashing the potatoes, yeast fermentation runs in a batch process for up to 3 days. Finally the crude spirit is distilled and rectified.

Starch

Starch is the main component of solids in potato tubers. In comparision with other types of starch, isolated potato starch offers several technological advantages for applications in the food and nonfood industry. Also plastic substitutes (e.g., shopping bags) have become of strong interest. The main disadvantage of potato starch separation is the campaign separation outside the winter season: destroying the cells by rasping liberates the starch granules. After

fruit water separation, starch is washed out in jet extractors of various types. In addition to welded sieves, plate sieves improve the efficiency of this step. After desanding, crude starch milk is purified by countercurrent washing with demineralized fresh water, dewatered, and dried. Optimum engineering results in starch recovery rates of at least 97–98%.

Fruit water is separated and concentrated by decanting for subsequent protein recovery by isoelectric precipitation and heat coagulation. The remaining deproteinized fruit water can be concentrated by ultrafiltration or reverse osmosis. Further evaporation results in a potato protein liquid.

See also: **Browning**: Nonenzymatic; Enzymatic – Biochemical Aspects; Toxicology of Nonenzymatic Browning; Enzymatic – Technical Aspects and Assays; **Chilled Storage**: Use of Modified-atmosphere Packaging; Packaging Under Vacuum; **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Cooking**: Domestic Techniques; **Drying**: Theory of Air-drying; **Potatoes and Related Crops**: The Root Crop and its Uses

Further Reading

Agblor A and Scanlon MG (1998) Effects of blanching conditions on the mechanical properties of French fry strips. *American Journal of Potato Research* 75: 245–255.

Bergthaller W, Witt W and Goldau H-P (1999) Potato starch technology. *Starch/Stärke* 51: 235–242.

Boskou D and Elmadfa I (1999) *Frying of Food*. Lancaster, PN: Technomic.

Lisinska G and Leszczynski W (1989) *Potato Science and Technology*. London: Elsevier Applied Science.

McMinn WAM and Magee TRA (1997) Physical characteristics of dehydrated potatoes – part I and part II. *Journal of Food Engineering* 33: 37–48, 49–55.

Mensah-Wilson M and Gierschner K (2000) An improved method for the production of canned potatoes. *Fruit, Vegetable and Potato Processing* 85: 132–139.

Moreira RG, Castell-Perez ME and Barrufet MA (1999) *Deep-Fat Frying*. Gaithersberg, MD: Aspen, USA.

Talburt WF and Smith O (eds) (1987) *Potato Processing*, 4th edn. New York: AVI Book.

Willard M (1993) Potato processing: past, present and future. *American Potato Journal* 70: 405–418.

POULTRY

Contents
Chicken
Ducks and Geese
Turkey

Chicken

A J Maurer[†], Formerly of University of Wisconsin-Madison, Madison, MI, USA

Introduction

Chicken is one of the most widely accepted muscle foods in the world. Its high-quality protein, relatively low fat content, new products, and generally low selling price because of favorable feed conversion make chicken a high-demand food in the marketplace. Furthermore, the absence of cultural or religious taboos allows increased chicken production and consumption worldwide.

This article reviews specific characteristics of chicken, its chemical composition and nutritive importance, production advances, slaughter and further processing, food uses and products, and microbiological problems.

The USA leads the world in poultry meat production, followed by the former Soviet Union, China, and Brazil. Consumption of chicken in the USA has increased dramatically over the last 20 years. **Table 1** shows the changes in chicken consumption and

Table 1 Meat consumption per capita (kg) in the USA

	1994	*1995*	*1996*	*1997*	*1998*
Beef	30.3	30.5	30.6	30.3	30.5
Pork	24.0	23.8	22.2	22.1	23.6
Poultry	40.4	40.3	40.8	41.1	41.9

Source: http://meat.tamu.edu/consum.html.

[†]Deceased.

compares those changes with other meats. Broiler consumption in the USA continues to increase.

Chemical Composition, Nutrition, and Dietary Significance

Chicken is a very digestible source of high-quality proteins. Its muscles are differentiated into light meat (primarily the breast) and dark meat (the legs). These muscles vary in myoglobin content as well as in fat content. Table 2 lists the chemical composition of selected chicken parts, and shows some of the proximate changes that occur on cooking. (*See* **Protein**: Food Sources.)

Chicken meat is low in total fat and saturated fat, and light meat is lower in fat than dark meat (Table 2). Most of the fat in poultry is deposited under the skin and is easily removed by pulling the skin from carcasses or parts. Because chicken fat has a high ration of unsaturated to saturated fatty acid (approximately 30% saturated), the fat melting point is relatively low and liquefies easily. Chicken fat is also prone to oxidative rancidity. Although most consumers believe otherwise, sodium and cholesterol contents in chicken are similar to those in most meats. Light meat contains more protein, and dark meat contains higher levels of fat and cholesterol. Older chickens generally contain more fat and less moisture. Cooking tends to concentrate the level of cholesterol and protein because water is lost in the heating process. (*See* **Cholesterol**: Properties and Determination; **Fatty Acids**: Properties; **Sodium**: Properties and Determination.)

Chicken contains all the essential amino acids, and is a good source of B vitamins and minerals such as iron and phosphorus. Poultry meat also provides potassium, calcium, magnesium, and copper, but carbohydrates and fiber are negligible. Light meat contains more nicotinic acid than dark meat, although both are good sources. Conversely, dark meat contains more riboflavin, iron, and zinc than light meat. (*See* **Amino Acids**: Properties and Occurrence; Refer to individual vitamins and minerals.)

Economics of Fast-Growing Broiler Chickens

Chicken broilers have been bred to a uniform size specifically for meat yield. A common broiler strain combines Cornish genes for conformation and fleshing with the White Plymouth Rock for white feathers and faster growth. White-feathered birds are the norm for commercial broiler production because of their cleaner-looking defeathered carcasses.

Many factors have contributed to the efficiency and economics of broiler production. Advances in breeding, nutrition, disease control, and management practices have enabled the broiler industry to produce a chicken weighing 1.8 kg in 6–7 weeks. Vaccines, antibiotics, confinement rearing, and computer-balanced rations assist in producing broilers with a feed conversion ratio of less than 2 kg of feed per kg gain. Vertical integration in the broiler industry has increased efficiency even more by combining most of the activities in producing, processing, and marketing broilers under the same ownership and management.

Chicken Processing

Chicken slaughter and processing have evolved into a sophisticated and automated procedure whereby a processing plant can handle as many as 20 000 broilers per h. Transforming live chickens into a ready-to-cook form involves live bird catching, crating and hauling, unloading, hanging on a conveyor line, stunning, slaughtering and bleeding, scalding,

Table 2 Proximate analysis of chicken broilers (per 100 g edible portion, wet-weight basis)

Nutrient	*Breast meat with skin (raw)*	*Breast meat without skin (raw)*	*Breast meat with skin (roasted)*	*Leg meat with skin (raw)*	*Leg meat with skin (roasted)*
Water (g)	69.46	74.76	62.44	69.91	60.92
Protein (g)	20.85	23.09	29.80	18.15	25.96
Energy (kJ)	772	462	827	785	974
Lipid					
Total (g)	9.25	1.24	7.78	12.12	13.46
Saturated (g)	2.66	0.33	2.19	3.41	3.72
Monounsaturated (g)	3.82	0.30	3.03	4.89	5.24
Polyunsaturated (g)	1.96	0.28	1.66	2.65	3.00
Cholesterol (mg)	64	58	84	83	92
Ash (g)	1.01	1.02	0.99	0.85	0.92
Sodium (mg)	63	65	71	79	87

Reproduced from Posati LP (1979) *Composition of Foods. Poultry Products, Raw, Processed, Prepared.* Agriculture Handbook 8–5. Washington, DC: Science and Education Administration, US Department of Agriculture.

defeathering, eviscerating, inspecting, chilling, grading, packing, and shipping. (*See* **Meat**: Slaughter.)

Assembly of Live Birds

Broilers should be taken off feed 8–12 h before the time of slaughter to reduce fecal contamination during processing. Bruising in chickens can be minimized by removing feeders and waterers prior to the arrival of the catching crew, and by careful handling of the birds as they are caught and placed into coops. Novel mechanical methods (e.g., herding, sweeping, and vacuum systems) for catching chickens are also available and commercially used by a few companies in Europe.

Hauling and Unloading

After the birds are loaded into wooden or plastic crates or specially built compartments, they are transported via open-sided trucks to the slaughter plant. Large-volume, slow-speed fans or evaporative cooling provides good ventilation and comfort for the birds in the holding shed.

Trailer loads move into the plant unloading area as needed. Various methods (e.g., manual bird removal, or dumping the birds through side-doors) unload the crates. Birds are hung by their shanks on shackles attached to an overhead monorail conveyor for transfer through the slaughtering operations. The unloading area should be dimly lit, and a breast rub bar can be used to calm the birds. According to UK regulations, birds should be slaughtered within 3 min after they are suspended from shackles.

Stunning, Killing, and Bleeding

Stunning is used to enhance bleeding and feather release. The heads of the birds are usually dragged through an electrically charged water bath. For consistent stunning, the water should contain from 0.1 to 1.0% sodium chloride. A fine mist of water spray should be directed at the bird's feet to provide a positive electrical contact prior to entering the stunner. The shackles complete the circuit and cause an electric current to run through the body of the bird as a result of the applied voltage (approximately 50 V). Overstunning can rupture blood vessels and break carcass bones due to excess muscle contraction. It is essential that the electric shock does not kill the bird; bleeding must be the cause of death.

Killing can be either manual or mechanical. In the manual operation, a skilled worker with a sharp knife severs the jugular vein and carotid artery by cutting across the side of the neck at the base of the bird's head. In mechanical neck cutting, a guide bar with grooved rollers holds the head rigid and extends the neck in preparation for the cuts. The bird's head is guided across a single, revolving, circular blade or between a pair of revolving blades.

Bleed time should be 1.5–2 min, and the blood lost accounts for approximately 4% of the live weight. The shackles convey the birds through a blood tunnel, so that the blood can be collected and disposed of properly.

Scalding

After bleeding, the birds are scalded to loosen the feathers by immersion in agitated hot water. Wing flapping and struggling should have ceased by the time birds enter the scalder. In general, a soft- or semiscald temperature of 50–54 °C is used, with an immersion time of 1.5–2.5 min. For the fresh chilled broiler market, low-temperature scalding permits retention of the yellow cuticle on the chicken skin and retards skin drying.

Chickens are subscalded at 59–60 °C for cuticle removal if there is a preference for lighter-skinned birds, if the carcasses are to be frozen, or if the parts are to be battered and breaded. Higher temperatures greatly facilitate feather removal, and necks and wings of semiscalded birds are often scalded at the subscald conditions to achieve cleaner carcasses. In the USA, the US Department of Agriculture (USDA) recommends a minimum overflow from the scalder of 11 per bird to reduce build-up of contamination. Various chemicals added to scald tanks assist feather removal by reducing surface tension and enhancing wetting of the feathers. An alternative to immersion scalding with separate defeathering machines is a combined spray-scalding and defeathering system. This type of scalding may reduce contamination spread from bird to bird.

Defeathering

After the carcasses leave the scalder, they enter a series of online defeathering machines. The picking machines consist of banks of counter rotating stainless-steel disks or drums, with rubber 'fingers' mounted on them. As the birds are conveyed through the rotating picker fingers, the feathers are rubbed or plucked from the carcass. Continuous water sprays flush away the feathers.

Remaining pinfeathers are removed by hand. In the USA, the birds pass through a gas flame to singe the fine hairs. A final step in the defeathering area is an outside spray wash.

The head removal operation may take place in the defeathering area between the picker and the outside bird washer, or in the eviscerating room to facilitate crop removal. The bird's neck passes through a device

which restrains the head as the overhead conveyor pulls the body forward, separating the neck vertebrae at the base of the skull. Some head removal devices can include a set of rollers to separate the neck bones and a rotating knife to sever the neck skin instead of tearing it; yield is improved with this system.

The birds pass through an automatic hock cutter which severs the shanks at the hock joint and causes the carcass to drop onto a conveyor for transfer to the evisceration area. In the UK, the feet of the broilers are cut off just above the spur by means of a rotating knife. The severed feet remain on the shackles and are removed mechanically on the return line.

The scalding, defeathering, and other operations occur in a separate portion of the plant, apart from the evisceration and final processed bird area. The transfer conveyor takes the defeathered carcasses into the evisceration area, where the carcasses are rehung on the clean evisceration line.

Evisceration

The first cutting operation in the evisceration area involves removing the preen or oil gland, located on the top side of the tail next to its base. A sharp, short-bladed knife is used to remove the gland manually, or a machine with a cutting blade can be used online.

An opening cut into the body cavity is then made using a knife, a special drill-like vent cutter with a rotary blade, or an online automated machine. Making an incision in the abdomen and removing the vent can be performed manually or by machine to complete the opening process for evisceration.

Removing or drawing the viscera from inside the bird can be performed by hand or by machine. To retain the identity of the viscera within the carcass and be in a position for easy inspection, the viscera are draped uniformly to one side of the tail but remain attached to the carcass until the carcasses have passed the inspection for wholesomeness.

Mechanized removal of the viscera is very common in large processing plants (Figure 1). An eviscerating spoon is mechanically inserted into the cavity and withdraws the viscera. Back-up workers may be necessary to remove viscera missed by the machine.

Mandatory USDA inspection of every bird for wholesomeness is performed under the supervision of an inspector employed and trained by the government. If evidence of unwholesomeness is detected, the carcass or affected parts are trimmed or condemned.

Workers remove the giblets (heart, liver, and gizzard) from the viscera of inspected birds and clean and chill them. Then the lungs are removed with a vacuum lung gun, or they can be withdrawn with the viscera on a mechanical eviscerator. The neck is generally cracked with shears, the crop and windpipe are removed, and the neck is then pulled off by stationary guide bars.

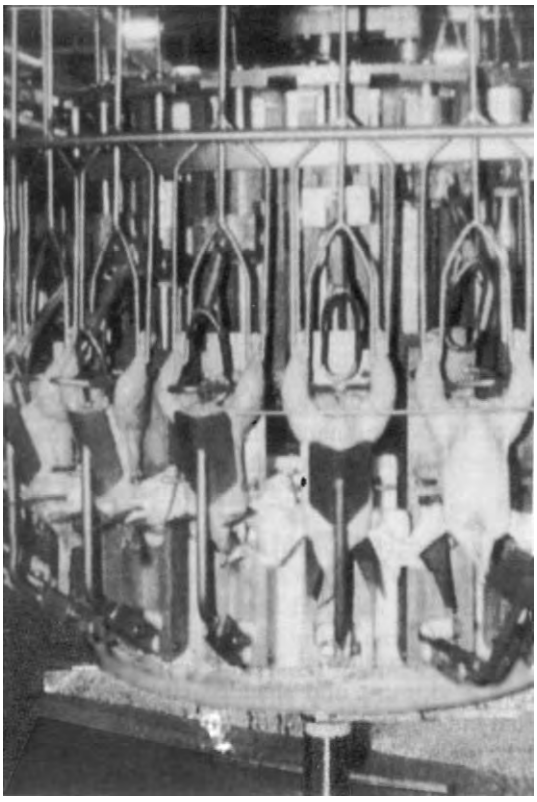

Figure 1 Automatic chicken eviscerator. Reproduced from Poultry: Chicken, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Finally, the shackles convey the carcasses through a bird washer where spray nozzles rinse the inside and outside of the carcasses.

Chilling

The most common chilling operations in the USA immerse carcasses in long flowthrough tanks containing agitated chlorinated water or slush ice. The USDA regulations require that a chiller overflow rate be maintained at 2 l of water per chilled bird, in order to minimize microbial build-up in the chill water.

The chickens are first placed in a prechiller, containing water at 10–18 °C, and then into a slush ice chiller at 0–1 °C. Chickens must be chilled to 4 °C or lower within 4 h to meet USDA requirements; the chilling time generally needed to obtain 4 °C carcasses is about 40–60 min.

After chilling, carcasses are hung by one leg on a drip line for 2.5–4 min for draining, and are then conveyed to the packing area. Although the meat absorbs some moisture during washing and chilling, the drained moisture level is strictly regulated. The

drip line shackles are usually weighing devices which drop the sized birds into appropriate bins.

Air chilling of eviscerated carcasses is used extensively in Europe. Spraying the carcasses with water at intervals avoids weight loss during evaporative chilling. (*See* **Meat**: Preservation.)

Grading

Carcass grading in the USA is a voluntary practice, but is often required by purchasers. Government graders sort the birds into A, B, and C categories according to body conformation, fleshing, fat cover, deformities, bruises, and defects such as pinfeathers, disjointed or broken bones, and missing parts. Europe uses similar grading standards.

Packing

Giblet packs containing a heart, liver, gizzard, and neck are stuffed into the body cavity of chilled, sized carcasses. The carcasses then go into overwrapped trays which are heat-sealed, or into polyethylene bags which are clipped or taped shut ready for market. Another common method in the USA for distributing broiler carcasses from the processing plant is to put them into corrugated, wax-impregnated boxes. After boxing, ice covers the birds to keep the skin moist during shipment. Nearly half of the broilers in the USA leave the processing plant as ice-packed birds. An increasingly popular packing method is chill-packing, whereby a $-6\,^{\circ}C$ air blast lowers carcass temperatures to between $-2\,^{\circ}C$ and $-1\,^{\circ}C$. Refrigeration then maintains that temperature during marketing of the birds.

Waste Products

Poultry processing results in large amounts of waste waters, semisolids, and solids, which require separation and treatment before being discharged into the environment. Waste material that can be reclaimed and, for example, used for animal feeds or fertilizer helps to reduce the overall load for disposal.

There are several ways to handle poultry offal (heads, feet, viscera, inedible parts, condemned whole birds, feathers, and blood). Usually, all offal except the blood and condemned birds is floated in water from the processing areas to an accumulation area for removal by trucks. Screens and rotating drums remove water and separate offal components. Because of its high oxygen demand for decomposition, blood is usually handled separately.

Most of the offal solids goes to a rendering plant, while liquid generally goes through a primary and secondary lagoon treatment system. Some plants can discharge the liquid effluent into a municipal water system after removing the fat and as many solids as possible.

Another method of waste treatment is spray irrigation or spreading the processing effluent on the land. This type of waste disposal requires large fields to prevent overloading. Runoff into streams is a major concern. (*See* **Effluents from Food Processing**: Composition and Analysis; Disposal of Waste Water.)

Food Uses and Products

Most broilers in the USA are sold as fresh (unfrozen) carcasses, but frozen deboned meat is used in further processing. A decline in the sale and consumption of whole chickens in the USA has been counterbalanced by a steady increase in the sale of cut-up carcasses and parts. More than half (53%) of the processed carcasses in 1990 were marketed as cut-up broilers and parts.

Cut-Up

The cut-up operation is usually mechanized with motor-driven equipment and shielded circular blades. Individual cut-up stations or online automated machines can be used. A popular method of packaging is to place the cut-up parts in a tray containing an absorbent pad to collect seepage and then overwrap the tray. The bagged whole carcasses and tray packs may be prepriced before delivery to the retail market.

Deboning

The spectacular growth in the sale of value-added, further-processed poultry products in recent years has placed a heavy demand on the production of deboned meat. Poultry meat has traditionally been removed from the carcass by hand with a sharp knife, while the carcasses hang from special shackles on a slow-moving line or are positioned on static or moving cones.

Several automatic deboning machines are also available for removing breast fillets and thigh and drumstick meat. One system operates by holding a particular chicken part in position above a contoured recess in a base plate. The machine forces the meat from the bone into the depression. Another meat-deboning system pushes the bone lengthwise out of the carcass part (e.g., drumstick) and strips off the meat in the process.

Currently, deboning is primarily done on aged chilled carcasses. Hot-stripping the muscles and skin from defeathered but uneviscerated carcasses is being studied. Some researchers are also testing hot deboning, the removal of meat after evisceration but prior to chilling. Generally, hot deboned meat is tough,

which greatly limits its potential uses. Some poultry meat, especially that of mature hens, is cooked prior to hand deboning.

After the major muscles are deboned from the carcass, the remaining frames (also necks, backs, and low-value parts) can be more completely deboned by special machines. In the first stage of mechanical deboning, a grinder reduces bone and meat particle size. In the second step, a pressure system squeezes the ground meat and bone against a perforated screen or microgrooved cylinder. Advances in mechanical deboning have eliminated pregrinding and produced a more texturally attractive product. An alternative system is a batch operation. Pressure from a ram forces meat to flow from the bones through a sieve screen; the bone cake is ejected, and a new batch is introduced.

Further-Processed Products

Much of the hand- and mechanically deboned chicken meat is directed into value-added products. Deboned meat can be marketed 'as is' fresh, as whole breast or split breast fillets, strips, and chunks. Alternatively, deboned meat can be tumbled, marinated, chopped, formed, ground, emulsified, or prepared in a number of ways for sale.

Many nonmeat ingredients are used in further processing. Salt helps to extract proteins for improved binding and texture, acts as a preservative, and enhances flavor. Up to 0.5% phosphate increases water-holding capacity and final product yield. Other ingredients, such as sweeteners, spices, binders, and curing salts, are used in the wide array of poultry products on the market. (*See* **Curing**.)

The driving force behind poultry product development is the consumer. Fat content in meats is a current concern; many people want healthier diets, and the low fat in poultry is a major reason for its popularity. Fat pads are removed from many whole carcasses in response to consumer desires. As new poultry products are designed, diet and health, microwave ovens, and an aging population cannot be ignored.

The fast-food restaurants have had a tremendous influence on new poultry product popularity for items such as nuggets, tenders, marinated breast fillets, and frozen–fried parts. Ethnic foods are popular in the USA, especially Italian, Mexican, and Chinese. Different flavors in poultry products are in demand, including cajun, barbecue, and honey. Spicy chicken wings are an example of a popular seasoned finger-food appetizer. In addition to less fat, smaller portions, and different flavors, consumers want less salt and cholesterol in their chicken products. The food must be of high quality, convenient, safe, and nutritious.

In the retail supermarket, low-fat chicken products such as patties, rolls, sausages, and sliced luncheon meats are selling well. Poultry frankfurters and bologna are generally lower in fat than similar red-meat products and are a healthy alternative. (*See* **Meat: Sausages and Comminuted Products.**)

Most recent entries among the new poultry products are boneless chicken breasts stuffed with ingredients such as cheese and broccoli or wild rice and mushrooms. These products are sold in a variety of packages and combinations, but are almost always ready for cooking in the microwave. Other gourmet items include chicken Kiev (breaded chicken breasts filled with garlic butter and parsley), chicken cordon bleu (filled with ham and cheese), and herb-roasted chicken dinner (low-fat, low-cholesterol, low-sodium meal). Even the luncheon market has new chicken 'lunchable' products, which are individual packages containing items such as cooked sliced poultry, cheese slices, and crackers.

Microbiological Concerns

Modern poultry husbandry and processing techniques have greatly improved the quality of poultry meat. However, chicken can carry many kinds of organisms. The two major concerns are control of spoilage organisms which cause consumers to reject the product due to odor or flavor, and the minimization of pathogenic organisms which may (under faulty handling) lead to a health hazard.

Spoilage Microorganisms

Although poultry is refrigerated, and even frozen, for shelf-life extension, spoilage will invariably occur owing to the growth and metabolic activities of specific types of bacteria. The psychrophiles, such as *Pseudomonas*, can generally grow at refrigerator temperatures and are responsible for most of the spoilage. Spoilage organisms are present in high numbers when an off-odor becomes apparent (10^7 cells per cm^2) and even greater when slime formation occurs from coalescence of colonies (*c.* 10^8 cells per cm^2). (*See* **Spoilage: Bacterial Spoilage.**)

Pathogenic Microorganisms

Poultry is considered to be a major source of *Salmonella*, which can spread during processing, and even after cooking if raw meat preparation surfaces were not cleaned before placing a cooked product on the same surface. However, normal cooking procedures destroy *Salmonella* and it will not grow well under refrigeration. The most common problems are undercooked poultry or poultry contaminated after cooking. According to data on the vehicle of

foodborne salmonellosis outbreaks in the USA between 1973 and 1987, only 3.8% (30 of 790) of the cases were caused by chicken.

In addition to *Salmonella*, other pathogens sometimes found in poultry include staphylococci, *Campylobacter, Listeria* spp., clostridia, and coliforms. Careful product handling, proper refrigeration, and adequate cooking will almost always insure product safety. (*See* ***Campylobacter***: Campylobacteriosis; ***Clostridium***: Food Poisoning by *Clostridium perfringens*; Botulism; ***Listeria***: Listeriosis; ***Staphylococcus***: Food Poisoning.)

Sources of Bacteria

Microbes in poultry-processing plants are ultimately found on the product and come from three main sources: the birds (feet, feathers, intestinal contents), the environment (water, air, supplies), and the workers. Careful management at the production site and adherence to good manufacturing practices at the processing plant (e.g., filtered air, cool temperatures, thorough cleaning and sanitation, and good worker hygiene) will minimize final product contamination with bacteria.

See also: **Amino Acids**: Properties and Occurrence; ***Campylobacter***: Campylobacteriosis; **Cholesterol**: Properties and Determination; ***Clostridium***: Food Poisoning by *Clostridium perfringens*; Botulism; **Curing**; **Effluents from Food Processing**: Disposal of Waste Water; Composition and Analysis; **Fatty Acids**: Properties; ***Listeria***: Listeriosis; **Meat**: Slaughter; Preservation; Sausages and Comminuted Products; **Sodium**: Properties and Determination; **Spoilage**: Bacterial Spoilage; ***Staphylococcus***: Food Poisoning

Further Reading

Anonymous (1990) *Broiler Industry Marketing Practices, Calendar Year 1989*. Washington, DC: National Broiler Council.

Bean NH and Griffin PM (1990) Foodborne disease outbreaks in the United States, 1973–1987: pathogens, vehicles, and trends. *Journal of Food Protection* 53(9): 804–817.

Brant AW, Goble JW, Hamann JA, Waback CA and Walters RE (1982) *Guidelines for Establishing and Operating Broiler Processing Plants*. Agriculture Handbook 581. Washington, DC: Agricultural Research Service, US Department of Agriculture.

Cunningham FE and Cox NA (1987) *The Microbiology of Poultry Meat Products*. Orlando: Academic Press.

Mead GC (1989) *Processing of Poultry*. Essex: Elsevier Science.

Mountney GW (1976) *Poultry Products Technology*, 2nd edn. Westport, CT: AVI.

North MO (1984) *Commercial Chicken Production Manual*, 3rd edn. Westport, CT: AVI.

Posati LP (1979) *Composition of Foods. Poultry Products, Raw, Processed, Prepared*. Agriculture Handbook 8–5. Washington, DC: Science and Education Administration, US Department of Agriculture.

Putnam JJ (1990) *Food Consumption, Prices, and Expenditures, 1967–88*. Statistical bulletin no. 804. Washington, DC: Economic Research Service, US Department of Agriculture.

Stadelman WJ, Olson VM, Shemwell GA and Pasch S (1988) *Egg and Poultry-Meat Processing*. Cambridge: VCH.

Ducks and Geese

A J Maurer[†], Formerly of University of Wisconsin-Madison, Madison, WI, USA

Introduction

Ducks and geese (waterfowl) are a delicacy to many people, while others object to the higher amounts of fat in the carcasses compared to broilers and turkeys. However, breeding and mass selection programs are improving the meat-to-bone and meat-to-fat ratios in waterfowl. Annual duck consumption on a ready-to-cook basis in the USA is only about 0.2 kg per person. Goose consumption is somewhat less, at approximately 0.01 kg per person per year. Although chickens and turkeys dominate the world poultry industry, in parts of Asia ducks are commercially more important than broilers (chickens), and there are more geese than turkeys in areas of Europe. This article reviews duck and goose processing and preparation, meat composition and nutrition, food uses and products, waste products, and microbiological problems.

Processing and Preparation

With a few exceptions, duck and goose processing is similar to that of broilers and turkeys. For a more complete slaughter and evisceration procedure, (*See* **Poultry**: Chicken.)

In the USA, White Pekin-type ducks are generally slaughtered at 7 weeks of age and weigh about 3.2 kg live. White Muscovy ducks are also raised for meat, but they are a slower-growing breed and commercially

[†]Deceased.

less popular than the White Pekin. Geese are 3–5 months old at slaughter, and their average live weight is 4.5–6.8 kg. The average ready-to-cook weight of geese varies, but is generally about 4.5–5 kg. Feather maturity (lack of pinfeathers) is an important factor in determining the best time to slaughter waterfowl. (*See* **Meat**: Slaughter.)

Processors transport ducks and geese to slaughter plants on open trucks or trailers. The loading and unloading process is unique as the birds are not usually caught by their legs; they are herded on to the conveyance, and off into holding pens at the processing plant. Processors use turkey-sized crates when there are relatively few geese to transport.

Ducks and geese are driven from holding pens through chutes on to scales for weighing and then into the shackle hanging area. Care in herding the waterfowl is essential to avoid pile-ups, skin scratches, damaged legs, and smothered birds.

It is important to minimize struggling by the birds when hanging them on motorized conveyor line shackles as excessive flapping can result in bruises. In smaller processing plants, killing funnels can be used to restrain and position the birds for slaughter.

The birds are usually stunned with an electric current and then bled. Sometimes a cut inside the mouth is used, but a common method of killing the birds is to cut the outside of the throat on the left side at the base of the jaw, severing the left jugular vein and carotid artery.

Following bleeding, both ducks and geese can be scalded or dry-picked. However, the latter method is slow and laborious. Generally, ducklings proceed through immersion scalders containing agitated water at 58–63 °C. Geese can be scalded in a similar commercial scalder, or they can be hand-scalded for a smallscale operation. Scald water temperature for geese should be 63–66 °C. The duration of the scald varies from 1.5 to 3 min, usually longer for geese than for ducks. The time and temperature will also change depending on the age of the bird, time of year (season), and density of feathering. The lower the temperature, the longer the scald. A little detergent or an alkaline defeathering agent can be added to facilitate thorough wetting of the feathers. Processors can hand-scald waterfowl by pulling the bird repeatedly through the water against the lay of the feathers.

After scalding, the birds may be rough-picked by hand, picked by a conventional rubber-fingered picking machine (online), or placed in a spinner-type picker. Because pins and down remaining on the carcass are difficult to remove, it is common practice to finish rough-picked birds by dipping each bird into melted wax specially formulated for this purpose.

The defeathered birds should be surface-dried (a jet of compressed air can be used) just long enough to allow the wax to adhere. The waxing operation is usually mechanized and often includes conveying the ducks in and out of two wax tanks. A resin-based microcrystalline wax is used for greater resilience and improved stripping.

Good wax penetration and adhesion are achieved in the first tank with immersion in 90 °C wax for 15 s. The second wax dip at 71 °C puts a heavier coating on the birds to thicken the wax for good pulling power and cleaner stripping. After waxing (and sometimes between wax dips), a cold-water spray over the birds or a dip into a tank of cold water will cool and harden the wax to a tacky state. The wax is then removed by hand or by a rubber-fingered wax stripper. Some processing plants dip the wax-picked birds into water at 82 °C to tighten the skin for easier manual removal of any remaining pinfeathers. Grasping pinfeathers between the thumb and a dull knife will assist in this operation.

The wax is reclaimed by remelting and straining out the pins, down, and feathers. Occasionally, the used wax should be 'cooked down' to remove all water that may have been mixed with it by emulsion.

After completely removing feathers, pins, and down, processors eviscerate the carcasses by making an opening cut in the abdomen and pulling out the viscera. Waterfowl viscera are somewhat difficult to withdraw, so that evisceration is most often done manually. However, processing plants in Europe are using automatic duck evisceration and automatic head and trachea pulling with processing rates of 4000 ducklings per h.

Each carcass on the eviscerating line is inspected for wholesomeness before the withdrawn viscera are detached. After the inedible viscera and lungs are discarded, the heads, tracheas, and feet are removed; the giblets are cleaned and wrapped for later placement in the body cavity. In many processing plants, only the two lobes are cut from gizzards because the inner lining is difficult to peel.

Following evisceration, birds receive an inside and outside spray wash before proceeding through a chilling treatment of cold tap water or ice and water. Airagitated cooling vats, a continuous immersion chiller, or air-spray chilling can be used.

After carcass cooling and draining in commercial waterfowl plants, each carcass is graded for conformation, fleshing, fat covering, and defects such as pinfeathers and exposed flesh. Processors then vacuum-package carcasses in barrier bags, heat-shrink the bags, and quick-freeze the packaged carcasses. Young ducks and geese may also be sold fresh, unfrozen. (*See* **Meat**: Preservation.)

Composition and Nutrition

Dressing percentages and meat yield of ducks and geese will vary with breed, age, sex, weight, and grade. The processing loss from live to ready-to-cook (carcass, neck, and giblets) for ducks is approximately 30%, 25–32% for geese, and 24% for broilers.

As the data in Table 1 show, raw ducks and geese contain more fat, less water, and less protein than are found in broiler chickens. Roasting causes fat and some water loss, thereby concentrating (increasing) the protein content. Because of the relatively high fat content of waterfowl, the carcasses become rancid more easily, and frozen shelf-life is shorter than that of broilers or turkeys. (*See* **Fats**: Digestion, Absorption, and Transport; **Protein**: Food Sources.)

Skinning the carcass can remove much of the fat in poultry. For example, the fat content of raw duckling meat with skin is 39.3%; without skin, it is 6.0%. Values for goose are 33.6% with skin and 7.1% without skin. The fat in ducks and geese is highly unsaturated, as is the case in all poultry. Duck and goose skin (including separable fat) makes up 34–38% of the ready-to-cook carcass, but broiler skin is about 15% of the carcass. Conversely, the meat yield from ready-to-cook carcasses for ducks and geese is 34–47%, compared to 52% for broilers.

Duck is a good source of thiamin, and goose is an excellent source of phosphorus. The other vitamins and minerals are also present in ample amounts. Iron content is especially high in waterfowl, contributing to the darker color of the breast meat. (*See* **Phosphorus**: Properties and Determination; **Thiamin**: Properties and Determination.)

Food Uses and Products

Most duck and goose meat is consumed in whole carcass form. The carcasses are roasted in the same way as other poultry, but they yield less cooked edible meat. Because waterfowl contain more fat than broilers or turkeys, they do not require basting during roasting. If the carcasses are excessively fat, it helps to puncture or scratch the skin to permit some fat to cook out during roasting. Thorough cooking of waterfowl (to an internal temperature of at least 85 °C) is important to attain a crispy skin and for complete customer satisfaction.

Food manufacturers sell some further-processed waterfowl products. Duckling breast portions, semi-boneless halves, marinated breasts, fully cooked roast half duckling, and other gourmet items are available to the institutional trade. A few shops market specialty duck and goose products, such as smoked duck, boneless breast of duckling, smoked goose leg, or goose liver sausage.

Goose livers are popular in delicacies such as pâté de foie gras; the product must contain at least 30% goose liver. For a period of time in Europe and in parts of the USA, many people force-fed, or noodled, their geese. The starchy noodling ingredients (corn, wheat, barley, and rye) used in this frequent forced feeding produced an extra large liver (weighing nearly 1 kg each), which was sold at premium prices.

Goose fat is sometimes used in place of butter. It can also be used in frying, baking, cooking, and preparation of gravies, broths, and soups.

Waste Products

Duck and goose slaughtering produces the usual waste products of blood, feathers, feet, heads, inedible viscera, grease, debris, and cleaning water. The control and handling of these waste materials depend on the standards set by individual countries and local authorities. The methods of handling can include disposal to a public sewer after the fat and as many solids as possible have been removed, or treatment

Table 1 Proximate analysis of ducks, geese, and broilers (100 g edible portion, wet-weight basis)

Nutrient	*Duck meat with skin (raw)*	*Duck meat with skin (roasted)*	*Goose meat with skin (raw)*	*Goose meat with skin (roasted)*	*Broiler meat with skin (raw)*
Water (g)	48.50	51.84	49.66	51.95	65.99
Protein (g)	11.49	18.99	15.86	25.16	18.60
Energy (kJ)	1697	1415	1558	1281	903
Lipid					
Total (g)	39.34	28.35	33.62	21.92	15.06
Saturated (g)	13.22	9.67	9.78	6.87	4.31
Unsaturated (g)	23.77	16.55	21.53	12.77	9.47
Cholesterol (mg)	76	84	80	91	75
Ash (g)	0.68	0.82	0.87	0.97	0.79
Sodium (mg)	63	59	73	70	70

Reproduced from Posati LP (1979) *Composition of Foods. Poultry Products, Raw, Processed, Prepared*. Agriculture Handbook 8–5. Washington, DC: Science and Education Administration, US Department of Agriculture.

on-site and disposal of the filtered effluent to a lagoon or spray irrigation system.

In general, the first stage of any effluent system is the removal of coarse solids by screening. A fat trap or dissolved air flotation can then remove fine solids, fat, and grease. The effluent is further cleaned by either anaerobic digestion or aeration. (*See* **Effluents from Food Processing**: Composition and Analysis; Disposal of Waste Water.)

In contrast to other poultry-processing waste treatment problems, ducks and geese provide several valuable byproducts. The Far East is a good market for frozen duck feet, where they are stuffed with pork and considered a delicacy. Duck tongues are also valued by orientals as hors d'oeuvres. Some parts (heads and offal) go into mink or other animal food.

Other important byproducts are duck and goose down and feathers, used chiefly by the bedding and clothing industries. The poultry slaughter plant rinses and centrifuges the wet feathers to decrease the water content prior to shipping them to a feather-processing plant. A machine separates the down (15–25% of the feather mixture) and feathers, and washes and dries them. About five ducklings or three goslings are needed to produce 0.45 kg of dry feathers.

On a smaller scale, home processors can wash feathers in soft, lukewarm water which includes either a mild detergent or a little borax and washing soda. After rinsing, feathers are spread out to dry.

Microbiological Problems

In general, microbiological problems in ducks and geese are similar to those in other poultry. Pseudomonads are the main spoilage organisms. However, because of the very low duck and goose meat consumption, few illnesses have been attributed to waterfowl. The combination of scalding at 60 °C, followed by immersion in molten wax at *c.* 90 °C to aid final removal of feathers, appears to have a beneficial effect on the microbial quality of the finished product. (*See* **Spoilage**: Bacterial Spoilage.)

See also: **Effluents from Food Processing**: Disposal of Waste Water; Composition and Analysis; **Fats**: Digestion, Absorption, and Transport; **Meat**: Slaughter; Preservation; **Phosphorus**: Properties and Determination; **Poultry**: Chicken; **Protein**: Food Sources; **Spoilage**: Bacterial Spoilage; **Thiamin**: Properties and Determination

Further Reading

Baeza E, Dessay C, Wacrenier N, Marche G and Listrat A (2002) Effect of selection for improved body weight and composition on muscle and meat characteristics in Muscovy duck. *British Poultry Science* 43(4): 560–568.

Cunningham FE and Cox NA (1987) *The Microbiology of Poultry Meat Products*. Orlando: Academic Press.

Franson JC, Hoffman DJ and Schmutz JA (2002) Blood selenium concentrations and enzyme activities related to glutathione metabolism in wild emperor geese. *Environmental Toxicology and Chemistry* 21(10): 2179–2184.

Mead GC (1989) *Processing of Poultry*. Essex: Elsevier Science.

Mountney GW (1976) *Poultry Products Technology*, 2nd edn. Westport, CT: AVI.

Orr HL (1978) *Duck and Goose Raising*. Publication 532. Ontario: Ministry of Agriculture and Food.

Posati LP (1979) *Composition of Foods. Poultry Products, Raw, Processed, Prepared*. Agriculture Handbook 8–5. Washington, DC: Science and Education Administration, US Department of Agriculture.

Stadelman WJ, Olson VM, Shemwell GA and Pasch S (1988) *Egg and Poultry-Meat Processing*. Cambridge: VCH.

Turkey

A J Maurer[†], formerly of University of Wisconsin-Madison, Madison, WI, USA

Introduction

Chickens and turkeys dominate the world poultry industry. The increasing popularity of poultry meat, including turkey, comes from low cost (value for money), a healthy nutritious image, and availability in a variety of convenient forms. Efficiencies in integrated turkey production, processing, and marketing have helped maintain favorable retail prices. This article reviews specific characteristics of turkey, processing of turkeys, food uses and products, microbiological concerns, waste products, and the nutritional profile of turkey meats and products.

Specific Characteristics

Turkey consumption has increased in many countries throughout the world, with Israel as the leader (over 9 kg per person in 1986). Less than one-third of this per capital consumption (2.7 kg) is of whole-carcass birds. Clearly, the growth in turkey popularity has come in cut-up parts and more fully prepared products.

Turkey hens in the USA, sold at about 16 weeks of age, provide almost 6 kg of ready-to-cook carcass. Toms (male turkeys) are slightly older, at 20–24

[†]Deceased.

weeks, and yield 10–12 kg of marketable carcasses. Desirable characteristics of turkeys include heavy bird weights without excess fat, a high dressing yield, and an ample proportion of valuable parts. The breast (white meat) is 35–40% of the ready-to-cook carcass. Leg meat, 25–30% of the carcass, contains more myoglobin and is darker than breast meat.

Other desirable turkey meat attributes include tenderness, bland flavor (allows further processing and seasoning), and good functional properties, such as protein extraction, protein gelation, water holding, meat binding, and emulsification. Turkey meat is similar to broiler meat in its composition and qualities, easily digestible, high in protein, and low in fat.

Processing of Turkeys

Turkey processing refers to slaughtering, feather removal, evisceration, and chilling. Other operations, such as inspection and packaging, are also important steps in processing turkeys for market. Most of the processing procedures are similar to those used for broilers.

Procurement

Special trailers or trucks with built-in cages haul turkeys to the slaughter plant. If crates are used to transport turkeys, they are considerably higher than those used for chickens. Handlers must catch, load, and unload turkeys very carefully to minimize bruises and broken bones. Escalator-type loaders are sometimes used to elevate turkeys to the built-in truck cages. The turkeys should be off feed 8–12 h prior to slaughter to minimize fecal contamination in the processing plant. Large-volume, slow-moving fans or evaporative cooling provides good ventilation and comfort for the birds in holding sheds at the processing plant.

Because turkeys are large and heavy, shackle line height in the unloading area and truck cages must be adjusted to the same level to minimize lifting required by the hangers. Workers hang the turkeys by both feet to the overhead moving shackles.

A dimly lit hanging area discourages the birds from struggling and flapping. Some processors install a smooth plastic bar parallel to and slightly below the overhead conveyor line so that the turkey's breast rubs against it to provide a soothing effect. The turkeys should move from the hanging area to the stunner within 6 min.

Stunning

The shackle line drags the heads of the birds through 0.1–1.0% saline water which contains a submerged electrode. Careful control of the voltage and current is vital. Too little current will not immobilize the birds, and too much current can cause a violent muscle contraction, often resulting in broken clavicle fragments in the muscle tissue.

Slaughter

Killing is usually performed manually by cutting across the side of the neck at the base of the bird's head, severing a jugular vein and carotid artery. Bleed time should be at least 2 min in a blood tunnel to collect the blood for proper disposal. Workers in some turkey slaughter plants cut both sides of the neck for more complete bleeding. (*See* **Meat**: Slaughter.)

Scalding

After bleeding, processors scald turkeys by immersion in agitated hot water to facilitate feather release. For some fresh chilled markets, they semiscald turkeys at 50–52 °C to retain the epidermal cuticle, which reduces skin dehydration. Most processors, however, scald turkeys at approximately 60 °C for 2–2.5 min. The USA has a requirement for an overflow from the scald tank of approximately 11 of water per bird to help float debris from the water.

Defeathering

After scalding, the shackle line carries the birds through a series of picking machines containing rotating rubber fingers on disks or drums which rub or pluck the feathers from the carcasses. Workers remove the remaining pinfeathers by hand.

The birds pass through a gas flame to singe off the filoplumes or hairs protruding from the skin surface. The final process on the kill line is a thorough washing of the external surface of the carcass with pressurized water jets.

The carcasses then pass through a shank-cutting station. In some plants, an automatic tendon puller removes the shanks and pulls up to nine of the main sinews from the drumstick. At this point, a conveyor transfers the carcasses from the killing, scalding, and defeathering area into another room for evisceration.

Evisceration

For evisceration, turkeys hang from both legs with their heads also placed in a center slot of the shackle, creating a three-point suspension. This presents the birds horizontally, breast-up, for easier cutting and eviscerating. Plants equipped with mechanical eviscerating equipment use the two-point leg suspension.

Evisceration begins with an incision made through the abdominal wall. The cut continues around the vent and enlarges in the abdomen for easier removal

of the viscera. In some plants, a bar (transverse) cut in the abdomen allows evisceration and later trussing of the legs. A mechanical vent cutter with a revolving cylindrical blade is another method to make an opening cut and remove the vent.

Processors withdraw the viscera through the abdominal opening, taking special care to avoid damaging the intestines or spilling the contents. The viscera are left attached to the carcass and draped over the outside of the bird for inspection. An inspector then examines each bird (inside, outside, and viscera) for wholesomeness. Carcasses can be trimmed to remove damaged parts or dressing defects, or condemned if unfit for food.

Workers harvest, clean, and save the giblets (heart, liver, and gizzard). They clean the gizzard by splitting it, washing away the contents, peeling off the hard lining, and washing. They discard the inedible viscera, and remove the lungs with a special tool or vacuum lung gun.

As one of the last operations, the head is cut or pulled off, and the esophagus and crop are removed from the front of the carcass. Finally, the skinless neck is cut off, washed, and retained for packing. Following evisceration, the birds must be washed inside and out before chilling.

It is possible to perform some of the slaughter and evisceration operations mechanically, but the variability in turkey carcass sizes presents unique problems. Fully automated equipment for turkey processing is still being developed.

Chilling

Most processors chill poultry meat in cold water or water and ice, but they use air chilling extensively in Europe, where carcasses are more often soft-scalded and sold fresh. The most-used chillers in the USA drop the birds from the evisceration line into a prechiller which also serves as a very effective washer. The agitated water has a temperature of less than 18 °C. The birds travel to a second chiller with a water temperature of less than 2 °C. Inspectors carefully monitor chiller water overflow, and carcass exit temperature and water uptake.

After chilling, graders evaluate the carcasses according to conformation, fleshing, fat covering, and defects such as pinfeathers and exposed flesh. (*See* **Meat**: Preservation.)

Packaging

In some plants, turkey carcasses move directly to a cut-up or deboning line. If turkeys are sold as whole carcasses, workers place a giblet packet in the crop cavity and a neck in the body cavity. Processors truss the legs into the bar cut, or use a metal or plastic hock lock. These turkeys will probably be frozen, in which case they are placed in an oxygen-impermeable, shrink-film bag. After the air is evacuated, the bags are clipped shut and passed through a hot-water shrink tunnel prior to freezing.

The freezing process is very rapid, with the first step being a brine or blast freezer to set the carcass surface color. Final freezing and storage occur in a holding freezer, where the birds can be stored for relatively long periods of time until they are moved into market channels.

Food Uses and Products

About 75% of the turkeys produced in the USA are cut up or further processed. In some countries, such as Israel, France, and Italy, the proportion of cut-up or further-processed turkey is 90% or more. However, whole birds are still popular at holiday times. Some of the whole carcasses are injected with various flavored solutions to create a self-basting, juicier carcass when roasted.

Deboning

Deboning of chilled turkey carcasses usually occurs as they hang from special shackles on a slow-moving line, or sit on deboning cones which may be static or moving. Precise cuts are made to remove parts, breast meat, thigh meat, and trimmings from the skeleton. The drumsticks (and also the thighs) are often deboned on automated machines, whereby the drumstick bone may be pushed lengthwise out of the meat; thigh pieces may be compressed in a specially designed mold, squeezing the meat away from the bone. More deboning automation will occur in the near future.

After removal of intact muscles using knives or semiautomated procedures, the remaining frames, necks, and backs are usually mechanically deboned. One system uses an auger principle, pressing the meat and bone against a perforated cylinder screen or microgrooved cylinder. The alternative system is a hydraulic-press-type batch design which squeezes the meat through a series of stationary filter rings. Depending on the incoming meat materials (meat-to-bone ratio), screen perforation size, and machine adjustments, the mechanically deboned meat can be a very fine purée or have a particle size of 5 mm, which is ideal for some sausages.

Further Processing

Hand- and mechanically deboned meats are useful in many further-processed products. The term 'further processing' encompasses procedures such as deboning, size reduction, injection, tumbling, massaging, reforming, and emulsifying. The principles of

water-holding, protein extraction, protein gelation, and meat binding are very important in making further-processed products. Other processes, such as battering, breading, cooking, and freezing, may take place. Further processing also refers to whole carcasses that are basted, marinated, or smoked.

The types of products now available are increasing, ranging from cut-up portions to reformed roasts,-breasts, rolls, steaks, hams, burgers, frankfurters, bolognas, coarse-ground sausages, salamis and bacons. More recently, ready-prepared meals have utilized an increasing amount of turkey meat. (*See* **Meat**: Sausages and Comminuted Products.)

Consumers with changing careers and households are the driving force behind trends in the consumption of poultry products. Smaller families often lack time but can afford the convenience of further-processed poultry. Diet and health, microwave ovens, and an aging population are important factors in new product development. Ethnic foods, especially Italian, Mexican, and Chinese, are popular in the USA. Different flavors and marinades, such as honey and barbecue, are well liked in turkey products. Smaller portions, less fat, less salt, and low cholesterol are in vogue. Poultry appetizers, finger foods, sliced luncheon meats, and center-of-the-plate, cooked, vacuum-packaged, chilled boneless breasts are very popular. New turkey products could soon include sous-vide (vacuum-sealed, cooked, refrigerated, ready-to-heat) entrées, or possibly surimi (minced washed gelled protein, usually from fish) seafood-style favorites.

Microbiological Concerns

Turkey meat quality is highest immediately after processing. The maintenance of acceptable quality depends on initial microbial levels and measures taken to minimize organism growth and prevent further contamination. The two major concerns are spoilage organisms which cause odors or off-flavors, and pathogenic organisms which may, under faulty handling such as undercooking or temperature abuse, lead to a health hazard. The cutting, deboning, handling, mixing, and packaging of turkey meat also increase possible microbial contamination and growth. (*See* **Spoilage**: Bacterial Spoilage.)

Although turkey and turkey products are refrigerated or frozen for shelf-life extension, spoilage can occasionally occur as a result of the growth and metabolic activities of specific types of bacteria. Psychrophiles, such as *Pseudomonas*, can grow at refrigerator temperatures and cause problems. Turkey has reached spoilage conditions when an off-odor becomes apparent (10^7 cells per cm^2) or when slime formation occurs (10^8 organisms per cm^2). It is important to maintain refrigeration temperatures at 0–4 °C to minimize microbial growth in turkey and turkey products.

Food pathogens are more serious than spoilage organisms because the food product may not look or smell spoiled. Some of the pathogens of concern in turkey are *Salmonella*, *Staphylococcus*, *Campylobacter*, *Listeria*, and coliforms. Recently, *Salmonella* has been a major worry for consumers, and turkey carcasses do harbor the organism. Surveys in the USA indicate that in food-borne salmonellosis outbreaks only 4.5% are caused by turkey. To prevent food-borne illness from poultry, it must be kept refrigerated and cooked properly, and cross-contamination or postcooking contamination from unclean utensils or equipment must be avoided. (*See* ***Campylobacter***: Campylobacteriosis; ***Listeria***: Listeriosis; ***Staphylococcus***: Food Poisoning.)

Bacteria can come from many sources. At the processing plant, bacteria arrive on the feet and feathers of the birds; they are present in the intestinal contents, and can also come from the workers and the environment (air and water supplies). Bacterial problems can be minimized by following good production and manufacturing practices such as feeding clean feed and keeping the litter dry at the production site, using clean hauling equipment, filtering incoming air at the processing plant, monitoring the water supply, eviscerating carefully, chlorinating chiller water, insisting on good worker hygiene, and using an approved plant clean-up and sanitation program.

Utilization of Waste Products

Poultry processing results in large amounts of highly polluting waste waters, semisolids, and solids, which must be separated and treated before being discharged into the environment. Where practical, the use of waste products for livestock food or fertilizer reduces the overall load for disposal. Rendered poultry byproducts as an animal feed ingredient can provide 50–60% protein. Feather meal and dried blood also have value as a feedstuff. Such products must be carefully processed before being recycled as animal feed, to avoid microbial contamination.

The types of waste and byproducts differ at varying stages of processing. Manure, feathers, blood, viscera, flesh debris, grease, and cleaning water are examples of the pollutants to be treated and either used or discarded.

The methods of disposal are to a public sewer, or treatment on-site followed by disposal to a water course or to fields. Preliminary treatment of turkey-processing wastes using a coarse-solids screen separator as well as a fatty-matter trap, or chemical

flocculation combined with a dissolved air flotation system, will significantly reduce the pollution potential of the effluent before discharge or further biological treatment. Based on the strength of the effluent, a secondary biological treatment uses a mixed culture of microorganisms for anaerobic digestion or aerobic treatment. The cheapest and most cost-effective options for the disposal of stabilized processing sludges arising from the biological treatment of poultry processing effluents are to spread them on the land or to discharge them at land-fill sites. (*See* **Effluents from Food Processing**: Composition and Analysis; Disposal of Waste Water.)

Nutritional Significance

Table 1 shows the nutrient composition of selected raw and cooked turkey meats, and Table 2 lists the composition of several further-processed turkey products.

Turkey and its products have a favorable reputation as nourishing and healthy foods. The composition of raw turkey meat depends on factors such as diet, age, sex, and growth environment. Processed product nutrition (Table 2) is a result of incoming meat and nonmeat ingredients as well as cooking processes and formula variations by different manufacturers. Salt and fat can vary markedly, depending on whether salt was needed for protein extraction and binding, or whether skin (with adhered fat) was added for juiciness and flavor.

Consumers recognize turkey as a good protein source. Turkey meat is easily digestible, contains all the essential amino acids, and is a good source of the B vitamins and iron. (*See* **Amino Acids**: Metabolism; **Vitamins**: Overview.)

Turkey has a relatively low fat content, and the fat is only about 30% saturated. Because the fat is highly unsaturated (more than beef and pork), it is a softer type of fat and prone to oxidation. As turkeys grow, they deposit more fat under the skin. The fat content is higher in dark meat, and protein content is greater in light meat. Moisture and some fat are lost during heating, and protein is concentrated in cooked turkey.

Turkey products are somewhat similar to other processed meat products in protein content but are higher in moisture and lower in fat and energy. This favorable nutritional profile for turkey is one of the major reasons for its increasing popularity. (*See* **Protein**: Food Sources.)

Table 1 Proximate analysis of turkey (per 100 g edible portion, wet-weight basis)

Nutrient	*Breast with skin (raw)*	*Breast with skin (roasted)*	*Leg with skin (raw)*	*Leg with skin (roasted)*	*Light meat only (raw)*
Water (g)	70.05	63.22	72.69	61.19	73.82
Protein (g)	21.89	28.71	19.54	27.87	23.56
Energy (kJ)	659	794	605	874	483
Lipids					
Total (g)	7.02	7.41	6.72	9.82	1.56
Saturated (g)	1.91	2.10	2.06	3.06	0.50
Unsaturated (g)	4.32	4.25	3.89	5.59	0.69
Cholesterol (mg)	65	74	71	85	60
Ash (g)	0.91	1.03	0.89	0.99	1.00
Sodium (mg)	59	63	74	77	63

Reproduced from Posati LP (1979) *Composition of Foods. Poultry Products, Raw, Processed, Prepared*. Agriculture Handbook 8-5. Washington, DC: Science and Education Administration, US Department of Agriculture.

Table 2 Proximate analysis of selected turkey products (per 100 g edible portion, wet-weight basis)

Nutrient	*Ham*	*Roll (light)*	*Salami*	*Frankfurter*	*Loaf (breast)*
Water (g)	71.38	71.55	65.86	62.99	71.85
Protein (g)	18.93	18.70	16.37	14.28	22.50
Energy (kJ)	538	617	823	949	462
Lipids					
Total (g)	5.08	7.22	13.80	17.70	1.58
Saturated (g)	1.70	2.02			0.48
Unsaturated (g)	2.67	4.24			0.73
Cholesterol (mg)		43	82	107	41
Ash (g)	4.23	2.00	3.42	3.53	4.18
Sodium (mg)	996	489	1004	1426	1431

Reproduced from Posati LP (1979) *Composition of Foods. Poultry Products, Raw, Processed, Prepared*. Agriculture Handbook 8-5. Washington, DC: Science and Education Administration, US Department of Agriculture.

See also: **Amino Acids**: Metabolism; ***Campylobacter***: Campylobacteriosis; **Effluents from Food Processing**: Disposal of Waste Water; Composition and Analysis; ***Listeria***: Listeriosis; **Meat**: Slaughter; Preservation; Sausages and Comminuted Products; **Protein**: Food Sources; **Spoilage**: Bacterial Spoilage; ***Staphylococcus***: Food Poisoning; **Vitamins**: Overview

Further Reading

Bean NH and Griffin PM (1990) Foodborne disease outbreaks in the United States, 1973–1987: pathogens, vehicles, and trends. *Journal of Food Protection* 53(9): 804–817.

Cunningham FE and Cox NA (1987) *The Microbiology of Poultry Meat Products*. Orlando: Academic Press.

Mead GC (1989) *Processing of Poultry*. Essex: Elsevier Science.

Mountney GW (1976) *Poultry Products Technology*, 2nd edn. Westport, CT: AVI.

Nixey C and Grey TC (1989) *Recent Advances in Turkey Science*. Oxford: Butterworth-Heinemann.

Posati LP (1979) *Composition of Foods. Poultry Products, Raw, Processed, Prepared*. Agriculture Handbook 8–5. Washington, DC: Science and Education Administration, US Department of Agriculture.

Stadelman WJ, Olson VM, Shemwell GA and Pasch S (1988) *Egg and Poultry-Meat Processing*. Cambridge: VCH.

POWDERED MILK

Contents

Milk Powders in the Marketplace

M A Augustin and C L Margetts, Food Science Australia, Werribee, Victoria, Australia

Background

The conversion of milk into milk powder enables milk to be preserved in a convenient form for transport and for later use. The drying of milk was used initially as a means of dealing with surplus milk as it preserves the nutrients in milk. This allowed consumers to obtain a liquid milk product similar to fresh milk by reconstituting the milk powder with water. The transformation of liquid dairy products into powder has also paved the way for the effective utilization of byproducts such as skim milk and buttermilk from cream and butter processing and whey from cheese manufacture. Although skim and buttermilk powders were originally considered as byproducts of the dairy industry, these are now regarded as valuable food ingredients.

Commercial manufacture of milk powders began in the second half of the nineteenth century. Through the latter half of the nineteenth century and the early twentieth century, advances in drying technology made possible by improvements to drum and spray-drying equipment led to the growth of the milk powder industry. Drum drying was the major method used for the production of milk powders in the first half of the twentieth century. Today, spray drying is the method most widely used in the commercial manufacture of milk powders. However, drum drying is still used for the production of some specialized milk powders.

The export of milk powders from dairying countries such as Australia, New Zealand, and Europe has made possible the establishment of the recombined dairy products industry in countries that are not self-sufficient in milk. Recombining operations are increasing in South-East Asia, North Asia, Middle East, Africa, and South America, and this has led to a greater demand for milk powders.

A range of powders can be derived from milk (**Figure 1**). Today, dried milk and dairy-based powders represent a significant portion of the trade of dairy products. Milk powders are used directly by consumers and as ingredients in a range of manufactured dairy and food products. This article discusses the production and applications of traditional dried milk products and new tailor-made formulated milk powders on the market. Brief descriptions of casein and whey-based powders, although not strictly milk powders, have been included for completeness.

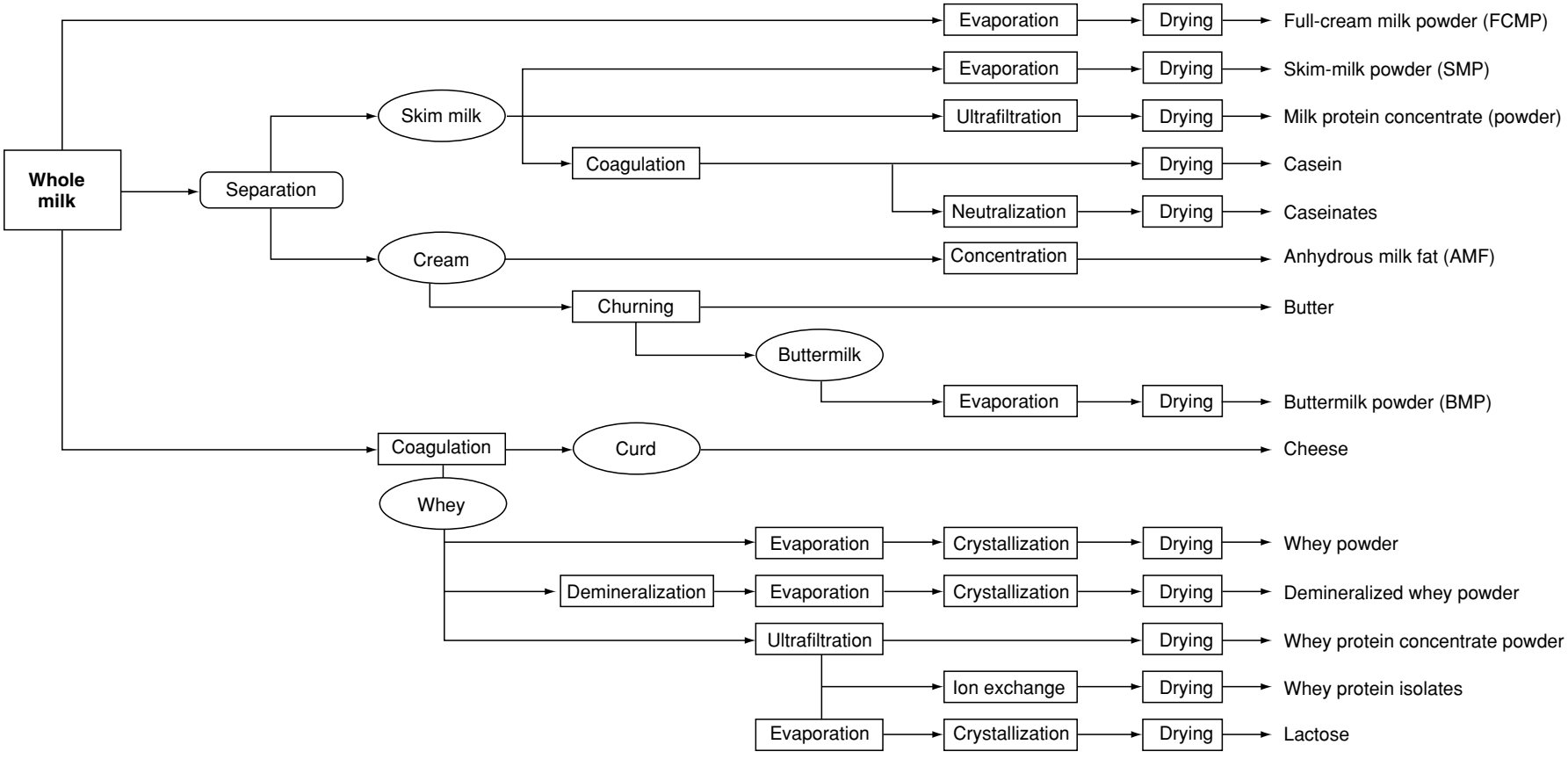

Figure 1 Milk ingredients. Reproduced with permission of the Australian Dairy Corporation.

Principal Spray-Dried Milk Powder Products

Skim, full-cream, and buttermilk powders are the traditional milk powders and are the major milk powders in the marketplace. The typical compositions of these powders are given in Table 1. The powders are dried to achieve specified standards for various grades of milk powders (Table 2).

Production of Skim, Full-Cream, and Buttermilk Powders

The main steps in the commercial manufacture of milk powders are: the preparation of the milk, heat treatment of the milk, concentration by evaporation, and spray drying. During the production of full-cream milk powder, there is an additional step, involving the homogenization of the concentrate prior to drying. Figure 2 shows the main steps in the manufacture of milk powders. Where a powder with instant properties is required, an agglomeration step is included during powder manufacture.

Preparation of milk This step defines the final composition of the powder. For the production of skim-milk powder, the milk is first separated into skim milk and cream by centrifugal separation. The skim milk should have a fat content of < 0.1%.

Table 1 Typical composition of milk powders

Constituent	*Skim-milk powder*	*Full-cream milk powder*	*Buttermilk powder*
Moisture (%)	3.0	2.25	3.0
Fat (%)	0.7	26.75	5.0
Protein (N × 6.38: %)	36.0	26.0	34.0
Lactose (%)	51.0	38.0	48.0
Ash (%)	8.2	6.0	7.9
Calcium (%)	1.31	0.97	1.3
Phosphorus (%)	1.02	0.75	1.0

Adapted from *Standards for Grades of Dairy Milks Including Methods of Analysis* (1990). Bulletin 916. American Dairy Products Institute.

For full-cream milk powders, the milk is normally standardized, generally by blending cream or skim to obtain a fat content of 25–28% in the final powder. Supplements such as vitamins and minerals may be added to the skim or full-cream milk to enhance the nutritional value of these powders. Buttermilk powders are made from the aqueous fraction remaining after churning of cream during buttermaking or during manufacture of anhydrous milkfat.

Heat treatment of milk The milk has to receive a heat treatment prior to concentration. A range of heat treatments may be used. These range from low-heat treatment (72 °C for 15 s) for pasteurization to high-heat treatments (e.g., 85 °C for 30 min or ultrahigh-temperature (UHT) at 140 °C for a few seconds). A primary purpose of the heat treatment is to achieve a required microbiological specification. The heat treatment also reduces the enzyme activity in the milk and increases the shelf-life of the powder products. Another consequence of the heat treatment is the denaturation of the whey proteins in milk. Appropriate heat treatments can be used to develop desired physical, chemical, and functional properties in milk powder products.

Skim-milk powders are generally classified on the basis of the heat treatment received during powder manufacture. Although the heat treatment classification is not used in the grading of milk powders, they are used as a general guide to the selection of powders for specific applications. The most commonly used heat classification is that of the American Dried Product Institute which is based on the amount of undenatured whey protein in the powder (Table 3). (*See* **Powdered Milk**: Characteristics of Milk Powders.)

Concentration of milk This is the first stage for removal of water from milk. The milk is concentrated at a low temperature by vacuum evaporation. The temperature of the milk reaches a maximum of 70 °C in modern evaporators. Generally, about 90% of the water is removed by the evaporator, as removal of

Table 2 Specifications for milk powders

Constituent	*Skim-milk powder*[a]	*Full-cream milk powder*[a]	*Buttermilk powder*[a]
Moisture (%)	Max. 4.0	Max. 4.5	Max. 4.0
Fat (%)	Max. 1.25	Min. 26.0, max. 40.0	Min. 4.5
Protein (%)			Min. 30.0[b]
Titratable acidity (%)	Max. 0.15		0.10–0.18
Insolubility index	Max. 1.25 ml	Max. 1.0 ml	Max. 1.25 ml
Scorched particles	Not more than disc B (15.0 mg)	Not more than disc B (15.0 mg)	Not more than disc B (15.0 mg)
Bacterial estimate	Not more than 50 000 per g	Not more than 50 000 per g	Not more than 50 000 per g

[a]Extra-grade spray-dried powder.
[b]Label should specify the minimum protein content.
Adapted from *Standards for Grades of Dry Milks Including Methods of Analysis* (1990). Bulletin 916. American Dairy Products Institute.

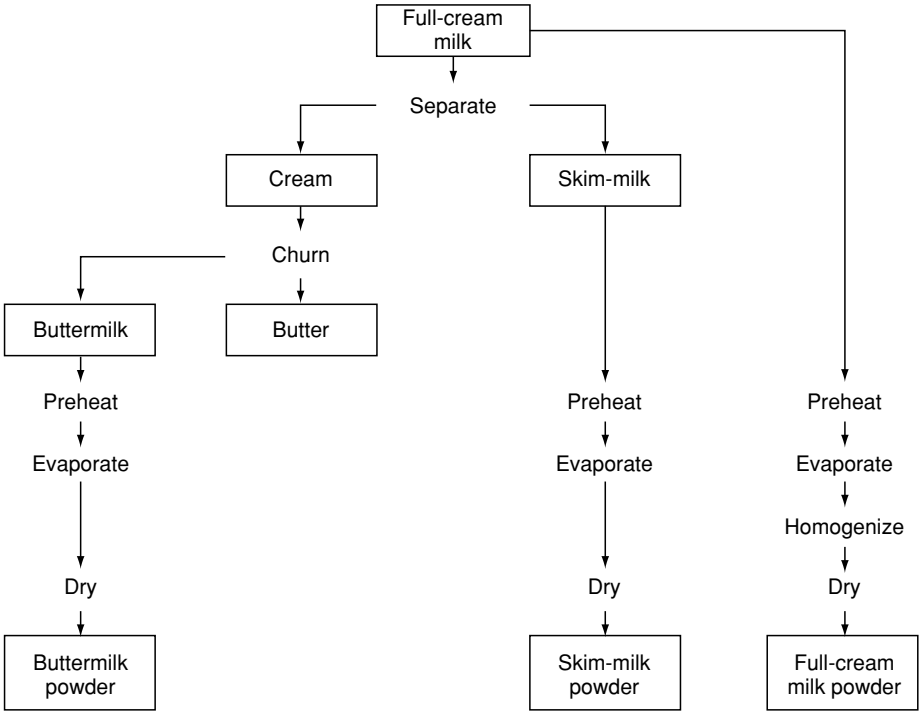

Figure 2 Schematic diagram of manufacture of milk powders.

Table 3 Heat classification of skim-milk powder

Heat class	*Whey protein nitrogen index (mg undenatured whey protein N g^{-1} powder)*
Low heat	Not less than 6
Medium heat	1.51–5.99
High heat	Not more than 1.5

Adapted from *Standards for Grades of Dry Milks Including Methods of Analysis* (1990). Bulletin 916. American Dairy Products Institute.

water by evaporation requires 16–20 times less energy per kg water than spray drying. The milk concentrate of about 48–50% total solids is then fed into the spray drier.

Homogenization of milk concentrate Homogenization of full-cream milk concentrates is carried out to reduce the 'free-fat' content (i.e., solvent-extractable fat) of the powder. During homogenization, the fat globule size is reduced, proteins are transferred to the surface of the globule, and the globule is stabilized. Homogenization of the milk concentrate results in powders with improved flow properties and resistance to caking and clumping.

Spray drying Spray drying transforms the fluid milk concentrate that is fed into the hot air of the drier to a powder. The moisture content of the spray-dried powder is usually 3–4% (w/w). The main stages of spray drying are atomization of the concentrate into a spray through pressure nozzle atomizers or rotary atomizers, the mixing of the atomized particles with the hot air, the evaporation of moisture from the surface of the particles, and separation of the powder from the air. Single- and double-stage spray driers are used in milk powder manufacture. Commercial milk powders are typically dried to 4% moisture or less.

Agglomeration of milk powders For the production of milk powders with instant properties, the milk powders are agglomerated. This process yields large powder agglomerates. The wettability and dispersibility of powders are improved by agglomeration. Surfactants (e.g., lecithin) may be used during agglomeration of full-cream milk powders to improve their instant properties in cold water. (*See* **Agglomeration**; **Drying**: Spray Drying; **Evaporation**: Basic Principles; **Homogenization**.)

Applications of Skim, Full-cream Milk, and Buttermilk Powders

Milk powders are used in many applications. Instant milk powders, which dissolve readily in water, are used by consumers as a substitute for fresh milk and in beverage mixes. Also available in the market are a range of nutritionally enriched milk powder products that have been tailored to meet the needs of consumers at various stages of life. These

Table 4 World use of skim-milk powder in products

Product	*% Used*
Condensed milk	30%
Ultrahigh-temperature (UHT) fluid	26%
Icecream	18%
Cultured products and yogurts	9%
Bakery	5%
Cheese	4%
Other products	3%

Based on figures from *Dairy Foods* Jan 1999, 100(1): 15.

Table 5 Typical compositions of acid casein, caseinates, and coprecipitate

Constituent	*Acid casein*	*Sodium caseinate*	*Calcium caseinate*	*Coprecipitate*
Moisture (%)	9.0	3.5	3.5	4.0
Fat (%)	1.0	1.0	1.0	1.0
Protein (N × 6.38: %)	88.0	91.4	90.9	89.0
Lactose (%)	0.1	0.1	0.1	1.5
Ash (%)	2.2	4.0	4.5	4.5

Adapted from Chandan R (1997) *Dairy-Based Ingredients*. St Paul, Minnesota: Eagan Press.

include powders fortified with various nutrients. Most common in the market are milk powders enriched with calcium, iron, and folate.

Milk powders have major applications as ingredients in manufactured dairy and processed food products. A significant amount of milk powder is used in the manufacture of traditional recombined dairy products such as evaporated milk, sweetened condensed milk, and UHT milk in countries which do not have an adequate supply of fresh milk. Milk powders are also used as ingredients in a range of food products, including icecream, cultured milks and yogurts, chocolate, confectionery, bakery products, soups, and sauces. Buttermilk powders are used as replacers for skim-milk powder in applications where enhanced dairy flavors are desired. Table 4 gives the proportion of skim-milk powders used in various applications. Their ability to bind water, thicken and gel, and their emulsifying and foaming properties make milk powders valuable food ingredients. These properties of milk powders can be modulated by the amount of heat treatment received by the powder during manufacture. The characteristics of milk powders that make them useful in food applications are discussed in more detail elsewhere. (*See* **Powdered Milk**: Characteristics of Milk Powders.)

Dried Casein Products

Caseins are the principal milk proteins, accounting for ~80% of the proteins in milk. They can be separated or precipitated from milk. Among the major products in this class of dairy products are acid and rennet casein, caseinates, and coprecipitates.

Production of Casein Products

Acid casein is made by direct acidification of skim milk to pH 4.6 by mineral or organic acids. Rennet casein is made by adding rennet to skim milk. The casein is separated, washed, and dried. Acid casein is insoluble in water. This limits the applications of acid casein in some applications. To obtain a casein product that is soluble, the acid casein is neutralized with an alkali and spray-dried to obtain caseinates. Various types of alkali have been used. Sodium, potassium, and calcium caseinates are the most common types of caseinates on the market. Coprecipitates are total milk protein powders; they contain both casein and whey proteins. The proteins are precipitated from skim milk under controlled conditions of heat, acid, and calcium concentration. The composition of casein products is given in Table 5.

Applications of Casein Products

Casein products have good emulsifying, whipping, and water-binding properties which make them suitable in applications such as desserts, confectionery, bakery products, salad dressings, processed meats, soups, and sauces. They have applications in dairy-based products such as coffee whiteners, yogurt, icecream, and processed cheese and imitation cheese products. As they are high-protein products, they are also used as supplements in dietetic foods. (*See* **Casein and Caseinates**: Uses in the Food Industry.)

Dried Whey and Whey Protein Concentrates

Whey, the serum remaining after the manufacture of cheese or casein, can be converted into powder. Sweet whey (pH ~6.3) is obtained during rennet-coagulated cheese manufacture whereas acid whey (pH < 5.1) is produced during cottage cheese or casein manufacture. Acid whey has a higher mineral content than sweet whey.

Production of Whey-Based Powders

Whey is preheated gently, evaporated, and dried for production of whey powders. To facilitate drying, the lactose in the whey concentrate is precrystallized for the production of nonhygroscopic whey powders. For the production of reduced-lactose or reduced-mineral whey powders, a portion of these components can be selectively removed from the whey. Whey can also be concentrated by membrane separation (ultrafiltration, with or without diafiltration). These

Table 6 Typical composition of whey-based powders

Constituent	*Sweet whey powder*	*Acid whey powder*	*Dried whey protein concentrates*
Moisture (%)	3.5–5.0	3.5–5.0	3.0–4.0
Fat (%)	1.0–1.5	0.5–1.5	1–10
Protein (N × 6.38: %)	11.0–14.5	11.0–13.5	34–80
Lactose (%)	63.0–75.0	61.0–70.0	10–55
Ash (%)	9.8–12.3	9.8–12.3	4–8

Adapted from Chandan R (1997) *Dairy-Based Ingredients*. St Paul, Minnesota: Eagan Press.

processes separate the proteins from the water, lactose, and minerals. The concentrated protein solutions are then spray-dried to obtain dried whey protein concentrates with a range of protein contents, typically between 34 and 80% protein. Table 6 gives the typical compositions of dried whey and whey protein concentrates. Whey protein concentrates with a lower protein content (25% protein) and whey protein isolates (>90% protein) can also be produced and are on the market.

Applications of Whey-Based Powders

Dried whey, reduced-lactose and reduced-minerals whey and whey protein concentrates are used as ingredients in many food applications, including confectionery, bakery products, snack foods, yogurts, dips, desserts, meat products, pasta products, ice-cream, soups, sauces, beverages, and processed cheese products. Whey protein concentrates have been also used as economic egg-white replacers in food formulations. With increasing protein content, the whey protein concentrates provide greater nutritional value as well as improved functional properties such as emulsification, foaming, water binding, viscosity building, and gelling to foods. Whey protein isolates are widely used as nutritional supplements in sports drinks and health foods such as nutritional bars and protein supplements. (*See* **Whey and Whey Powders**: Production and Uses; Protein Concentrates and Fractions.)

Other Milk Powder Products

Nowadays, a range of other types of milk powders may be formulated to achieve a desired fat, protein, lactose, or mineral content, as well as to obtain target functional attributes in the powder. These include a range of cream powders, high 'free-fat' milk powders for the chocolate industry, high-heat 'heat-stable' milk powders for evaporated milk, milk protein concentrate powders, skim milk/whey powder blends, lactose-hydrolyzed powders, yogurt powders, and a variety of customized formulations for use in target food applications. These tailor-made ingredients are finding their way into the marketplace, as food manufacturers increasingly demand ingredients with enhanced performance in their applications.

The major steps in the manufacture of skim milk/whey powder blends and specialized milk powders generally include the same essential unit processes of heating, concentration, and drying that are used in conventional milk powder manufacture, except that sometimes additional processing steps are required during the preparation of the milk or milk concentrate prior to drying.

Cream and High-Fat Powders

Cream and high-fat powders, containing 40–75% fat, may be produced by spray drying of cream or milk with an increased fat content. A typical cream powder obtained from drying of cream contains ~70% fat. Cream powders were developed more than 40 years ago. The early powders were difficult to dry and handle because of their high fat content. Improvements in formulation science and drying technology have enabled cream powders with superior properties to be made. These powders are becoming more prevalent in the marketplace and can be used as an alternative to fresh cream or in formulae where milk fat is required. Some of their applications include chocolate, confectionery, icecream, desserts, soups, and sauces.

High Free-Fat Milk Powders

High 'free-fat' powders for use in chocolate manufacture have traditionally been produced by roller-drying. In roller-dried powder, greater than 90% of the fat in the powder is readily extractable. In addition, roller-dried powders are flaky and do not contain much air. These properties of roller-dried powders make them more suitable for chocolate manufacture than conventional spray-dried powders that have a high level of entrapped air and a low free-fat content. The use of high free-fat powders in chocolate reduces the requirement for cocoa butter in chocolate making. As roller-drying is more expensive than spray drying, there have been attempts to increase the level of free-fat in milk powders intended as ingredients for chocolate. Nowadays, some high free-fat powders are produced by spray drying and have improved performance compared to conventional milk powders in chocolate manufacture. (*See* **Fats**: Uses in the Food Industry.)

High-Heat Heat-Stable Milk Powders

High-heat heat-stable milk powders have been processed to insure that they are suitable for use in the manufacture of recombined evaporated milk. For

the manufacture of this product, reconstituted milk powder is recombined with fat to produce a concentrate containing typically 26% total solids (18% skim-milk solids and 8% fat) and the concentrate is sterilized. A heat-stable milk powder is required in these applications, as the milk concentrate that is prepared from the powder has to withstand in-can sterilization (120 °C for 12 min) without coagulation or excessive thickening.

Milk Protein Concentrate Powders

In the production of milk concentrate powders, skim-milk concentrates obtained using a membrane process that concentrates the casein and whey proteins without precipitation are spray-dried. A range of milk protein concentrates with varying protein, lactose, and mineral contents may be made. The degree of concentration and process conditions used during membrane processing dictates the composition of the final powder. These powders are used in various applications where traditional milk powders and milk protein products have been used previously. As the protein content of these powders is higher than that of skim-milk powder, they have better functional properties.

Skim Milk/Whey Powder Blends

Skim milk/whey powder blends are made by replacing a portion of the skim-milk solids by whey-based solids. These replacers can be dry blends of skim-milk powders and whey-based solids or, alternatively, skim milk and whey may be blended prior to heat treatment, concentration, and drying. There are currently no specifications for the composition of these powders which contain a higher ratio of whey protein to casein than conventional milk powders. Blends are formulated to obtain the desired composition and functionality in the target applications, such as confectionery, bakery, icecream, and yogurts. Skim milk/whey powder blends are used as economic alternatives to skim-milk powder in many recombined dairy products. (*See* **Whey and Whey Powders: Production and Uses; Protein Concentrates and Fractions.**)

Lactose-Hydrolyzed Milk Powder

Lactose-hydrolyzed milk powders are niche products that have been developed for people who are lactose-intolerant. For the production of lactose-hydrolyzed milk powders, an enzyme, β-galactosidase (EC 3.2.1.23), that hydrolyzes lactose to glucose and galactose is added to the milk and allowed to act until the desired degree of hydrolysis is obtained; the enzyme is then inactivated by heat treatment prior to further processing. These powders may be used as alternatives to milk powders.

Yogurt Powder

A small amount of yogurt powder is available in the market. Milk is fermented prior to spraying to yield powder with a yogurt flavor. The powder has applications in confectionary as fillings and pastes, bakery products, and in a range of soups, sauces, and dips.

Milk Protein Hydrolysates

A number of milk protein, casein, and whey protein hydrolysates have been developed. These products are aimed at the nutritional supplement and sports nutrition markets. The hydrolyzed products are soluble and are quickly absorbed by the body.

Milk Powder Production Worldwide

Milk powder is the third largest category of dairy products, after cheese and fresh dairy and liquid milk products. World production of milk has continued to grow at the rate of about 1% a year through both increased cow numbers and better yields from herds. Stagnant sales of drinking milk in high-milk-production countries have meant that there has been a steady increase in manufacturing milk available for drying and milk powder production, particularly in Argentina, New Zealand, and Australia (Tables 7 and 8), with the last two countries exporting the majority of their production (Tables 9 and 10).

China and India have also shown steady increases in their production of full-cream milk powder, which has been used mostly for home consumption to shift production of milk products internally to cover areas and times of low production (Table 8). The

Table 7 Skim-milk powder production ('000 tonnes)

Country	*1995*	*1996*	*1997*	*1998*	*1999*
Argentina	36.6	36.6	40.0	38.0	81.0
Australia	229.6	237.9	230.8	272.8	275.4
Belarus	31.4	30.7			
Brazil	60.0	50.0	55.0	58.0	
Canada	71.1	64.6	66.2	69.8	77.4
Czech Republic	65.6	51.6	33.6	31.9	34.6
Estonia	19.2	21.6	16.1	14.1	
European Union (15 countries)	1187.7	1186.0	1130.4	1074.0	1113.9
Japan	189.4	200.3	199.9	201.8	191.1
Lithuania			30.1	26.4	17.4
New Zealand	127.1	172.0	177.0	160.0	227.0
Poland	118.0	120.6	119.9	131.3	110.9
Russia		107.0		79.0	83.0
South Africa	18.7	9.0	10.5	22.1	9.3
Switzerland	28.9	26.9	27.2	29.1	29.6
Ukraine	34.6	28.0	19.8	22.4	19.2
USA	564.0	477.6	548.0	515.0	625.1

Adapted from *The World Dairy Situation 2000* (2000) Bulletin 355, Table 14. International Dairy Federation.

Table 8 Full-cream and semi-skimmed milk powder production ('000 tonnes)

Country	*1995*	*1996*	*1997*	*1998*	*1999*
Argentina	146.5	162.3	166.0	207.0	269.0
Australia	113.0	133.7	126.5	144.8	172.0
China	352.0	358.0	391.0	422.0	552.0
Czech Republic	29.6	29.2	22.5	25.9	21.9
European Union (15 countries)	942.6	877.5	898.8	927.4	896.0
India	82.0	103.8	115.2	120.0	130.0
Japan	30.6	23.7	18.9	18.7	17.8
Lithuania	15.4	13.4	13.2	16.4	16.0
New Zealand	342.0	337.2	396.0	375.0	373.0
Poland	40.0	35.2	39.8	39.3	32.2
Russia	124.0	107.0	89.0	79.0	83.0
South Africa	11.6	9.7	13.0	10.8	11.3
Switzerland	10.3	9.9	9.3	8.0	8.0
Ukraine	11.5	6.3	4.8	3.2	3.1
USA	74.8	58.8	55.4	64.6	53.5

Adapted from *The World Dairy Situation 2000* (2000) Bulletin 355, Table 13. International Dairy Federation.

Table 9 World trade in skim-milk powder – exports ('000 tonnes)

Country	*1995*	*1996*	*1997*	*1998*	*1999*
World	958	1074	975	1130	1200
Argentina	22	17	15	29	
Australia	168	205	199	238	240
Canada	45	30	31	41	
European Union	227	282	175	272	370
New Zealand	127	183	166	174	157
Poland	77	112	101	83	
Other countries	270	152	177	114	

Adapted from *The World Dairy Situation 2000* (2000) Bulletin 355, Table 15. International Dairy Federation.

Table 10 World trade in full-cream milk powder – exports ('000 tonnes)

Country	*1995*	*1996*	*1997*	*1998*	*1999*
World	1154	1302	1392	1420	1430
Argentina	55	62	97	149	
Australia	93	109	110	139	
European Union	540	571	588	571	580
New Zealand	278	341	359	362	370
USA	16	27	20	17	20
Other countries	227	254	315	331	

Adapted from *The World Dairy Situation 2000* (2000) Bulletin 355, Table 15. International Dairy Federation.

USA and the European Union continue to be major producers and exporters of all forms of milk powder (Tables 7–10).

During the 1990s, skim-milk production increased an average of only 1.7% compared with the average annual increase in full-cream milk powder of 7%. Over this period, conventional skim-milk powder lost market share as a livestock feed component. It also faced increased competition in the food ingredient market from full-cream milk powders and specialist dairy powders, such as whey powders and other milk components, for the production of both dairy and general food products. The demand for dairy products made from preserved milks and manufactured food items with dairy ingredients such as bakery goods, confectionery, processed meats, and beverages continued to grow in all markets.

Factors which have influenced the increased consumption and demand for a greater variety of dairy products and manufactured foods containing dairy products throughout the period include:

- an increased middle class with more discretionary income
- changes in consumption patterns as diet becomes increasingly international in style
- increased identification of dairy products as a healthy food
- newly identified applications for dairy ingredients and components
- growth in food service establishments worldwide
- growth in food-manufacturing capacity in key markets
- extended availability of dairy technology as an increasingly concentrated dairy industry moves into new markets and expands its processing capability in developing countries
- better distribution systems to support and extend the availability of dairy products in developing countries

Whilst milk production has outstripped demand for dairy products in high-production countries, the demand for dairy products and components, including powders, has continued to grow in developing areas. Key markets for imported milk powders in the 1990s included South-East Asia, Africa, and Latin America. Algeria, Japan, the Philippines, and Mexico were the world's largest importers of skim-milk powders throughout the latter parts of the 1990s, with Algeria, Brazil, Venezuela, and Saudi Arabia taking large amounts of the world's production of full-cream milk powders (Tables 11 and 12).

Economic factors, such as the financial crises suffered by Brazil, some South-East Asian and Eastern bloc countries and changing agreements on world trade policies, affected trade in milk powders through the 1990s, with downturns occurring in affected economies. Import trends (Tables 10 and 11) show the influences of these factors on changes in the sales of milk powders throughout this period. Other factors which affect the demand for milk powders and prices obtained on international markets include:

Table 11 World trade in skim-milk powder – imports ('000 tonnes)

Country	*1995*	*1996*	*1997*	*1998*	*1999*
World	1190	958	1074	975	1130
Algeria	108	53	40	43	
Brazil	54	34	29	24	
European Union	43	61	74	65	75
Japan	87	75	73	57	57
Mexico	107	127	130	149	
Philippines	104	79	98	90	
Russia				31	109

Adapted from *The World Dairy Situation 2000* (2000) Bulletin 355, Table 16. International Dairy Federation.

Table 12 World trade in full-cream milk powder – imports ('000 tonnes)

Country	*1995*	*1996*	*1997*	*1998*	*1999*
World	1165	1154	1300	1390	1420
Algeria	75	78	91	120	
Brazil	217	116	100	100	
Malaysia	60	62	65	50	
Mexico	30	30	30	20	
Philippines	36	42	52	47	
Russia				35	35
Saudi Arabia	63	69	63	64	
Singapore	30	26	26	20	
Venezuela	66	66	56	80	

Adapted from *The World Dairy Situation 2000* (2000) Bulletin 355, Table 16. International Dairy Federation.

- protectionist policies, such as subsidies and market price supports
- quotas and stockpiles in storage in producing countries
- domestic consumption by producing countries
- policies on self-sufficiency and import reduction by importing countries
- competitiveness of major exporting countries
- currency fluctuations

Future of Milk Powders in the Marketplace

Consumers and food product manufacturers recognize that milk powders have many desirable attributes. The nutritional, physical, and physiological functionality of milk powders, coupled with the ease with which they can be handled and stored, make them sought-after ingredients in the market place. This is evidenced by the growing demand for milk and dairy-based powders. As users of ingredients are placing more stringent requirements for consistent performance and enhanced functional performance in food, the trend towards tailored milk powder ingredients that are matched to specific food application sectors is expected to grow.

See also: **Agglomeration**; **Casein and Caseinates**: Uses in the Food Industry; **Drying**: Spray Drying; **Evaporation**: Basic Principles; **Fats**: Uses in the Food Industry; **Homogenization**; **Powdered Milk**: Characteristics of Milk Powders; **Whey and Whey Powders**: Production and Uses; Protein Concentrates and Fractions

Further Reading

American Dairy Products Institute (1990) *Standards for Grades of Dry Milks Including Methods of Analysis*. Bulletin 916. Chicago: ADPI.

Anonymous (1996) Milk materials multiply. *Ingredients and Analysis International* 18: 30–35.

Ashton D, Brittle S and Shaw I (2000) Dairy outlook to 2004–05.*Australian Commodities* 7: 66–72.

Chandan R (1997) *Dairy-Based Ingredients*. St Paul: Eagan Press.

Early R (1998) Milk concentrates and milk powders. In: Early R (ed.) *The Technology of Dairy Products*, 2nd edn, pp. 228–300. London: Blackie Academic and Professional.

Hall CW and Hendrick TI (1971) *Drying of Milk and Milk Products*, pp. 1–16. Westport: AVI.

Hardcastle S, Gleeson T and Topp V (1999) Dairy outlook to 2003–04. *Australian Commodities* 6: 69–78.

Hunziker OT (1935) *Condensed Milk and Milk Powder: Prepared for Factory, School and Laboratory*, 5th edn, pp. 457–527. La Grange, IL: Hunziker.

International Dairy Federation (1990) *Recombination of Milk and Milk Products*, special issue no. 9001. Brussels: International Dairy Federation.

International Dairy Federation (1999) *Third International Symposium on Recombined Milk and Milk Products*. Special issue no. 9902. Brussels: International Dairy Federation.

International Dairy Federation (2000) *The World Dairy Situation 2000*. Bulletin no. 355. Brussels: International Dairy Federation.

Knipschildt ME and Andersen GG (1994) Drying of milk and milk products. In: Robinson KR (ed.) *Modern Dairy Technology*, vol. 1, 2nd edn, pp. 159–254. London: Chapman & Hall.

Masters K (1985) *Spray Drying Handbook*, 4th edn.London: Godwin.

Pisecky J (1986) Standards, specifications, and test methods for dry milk products. In: MacCarthy D (ed.) *Concentration and Drying of Foods: Proceedings of the Kellogg Foundation 2nd International Food Research Symposium (Cork: 1985)*, pp. 203–220. London: Elsevier Applied Science.

Rasmussen KW (2000) World market for milk powder in for a boom. *Scandinavian Dairy Industry* 2/00: 46–47.

Singh H and Newstead DF (1992) Aspects of proteins in milk powder manufacture. In: Fox PF (ed.) *Advanced Dairy Chemistry – 1: Proteins*, pp. 735–765. London: Elsevier Applied Science.

Wong NP, Jenness R, Keeney M and Marth EH (1988) *Fundamentals of Dairy Chemistry*, 3rd edn. New York: Van Nostrand Reinhold.

Characteristics of Milk Powders

M A Augustin, P T Clarke and H Craven, Food Science Australia, Weribee, Victoria, Australia

Background

Milk powders are used by consumers as a substitute for fresh milk and as ingredients for the manufacture of a range of processed food products. In order to be acceptable to consumers and users of ingredients, it is essential that milk powders are of a good quality. Milk powders are manufactured to meet certain specifications and standards for composition. These have been developed for milk powders by authorities such as the American Dairy Products Institute, the International Dairy Federation, the Food and Agricultural Organization of the United Nations and national food authorities in individual countries. In addition, a range of other technical specifications have been developed for the characterization of milk powders to ensure that they have the required functional performance in specific target applications. Milk powders may be similar in composition but have different functional properties.

There are many types of milk powders in the market place. This article focuses on the characteristics of skim and full-cream milk powders, which are the major types of milk powders produced. The microbiological quality, physical and chemical attributes of these milk powders, and their functional properties are discussed. Aspects of deteriorative changes that may occur in milk powders during transport and distribution that have an impact on the sensory properties of powders and their performance as food ingredients are included. The production, composition, and applications of various types of milk powders have been discussed elsewhere. (*See* **Powdered Milk**: Milk Powders in the Marketplace.)

Microbiological Aspects

Standards for Quality and Safety

Milk powder is a microbiologically stable product. It has a water activity of 0.3–0.4, which is too low to support the growth of microorganisms. However, after milk powder has been reconstituted, it is susceptible to microbial growth and spoilage in a similar manner to pasteurized milk. Provided milk powder is protected from moisture contamination before use, the numbers of microorganisms present generally decrease during storage, although the numbers of spores may remain constant.

Although milk powder does not support the growth of microorganisms, the microbiological content is an important consideration in the subsequent use of the powder. For this reason, government bodies and customer groups have developed microbiological limits or specifications that apply to certain groups of microorganisms that may be present in milk powder. These specifications may relate to expectations of raw milk quality, hygiene during manufacture, microbial safety, or compatibility with the intended use of the milk powder.

Common end-product standards relate to the total number of bacteria (mesophilic aerobes), coliforms, *Salmonella*, and *Staphylococcus aureus*. Criteria may also be applied for *Bacillus cereus*, *Listeria*, thermophiles, Enterobacteriaceae and spore-forming bacteria. The standards developed by the International Dairy Federation, for example, are shown in Table 1. Many countries have either adopted these standards or developed their own local specifications based on the principles of the International Commission on Microbiological Specifications for Foods (ICMSF).

The microbiological count of milk powder is influenced by both the numbers and types of microorganisms in the raw milk and the processing conditions under which the milk powder is produced. In powders subjected to a high heat treatment, the microorganisms present will be predominantly spore-formers, belonging to the genus *Bacillus*. When heat treatment is less severe, vegetative cells

Table 1 Microbiological specifications for milk powder, as recommended by the International Dairy Federation

Criteria[a]	*Total count (per gram)*	Salmonella *(per 25 g)*	*Coliforms (per gram)*	Staphylococcus aureus *(per gram)*
m	50 000	0	10	10
M	200 000	na	100	100
n	5	15	5	5
c	2	0	1	1

[a]For a production batch, n = number of samples that must be tested, c = number of samples that may exceed the microbiological limit specified as m, and M is the maximum allowable microbiological limit specified for any of the samples examined.
na = not applicable.

of thermoduric bacteria will be present, with their proportion to spore-formers decreasing with the intensity of the heat treatment applied. Vegetative cells of pathogenic bacteria and Gram-negative milk spoilage bacteria are destroyed during the heat treatment.

Coliforms, *Salmonella*, and other Enterobacteriaceae are killed when the milk is heated prior to evaporation; however, they may contaminate milk powder if conditions are not sufficiently hygienic during drying. These bacteria can enter the dryer through the intake air from the factory environment, or the equipment used to dry or transport the milk powder. Cracks in dryers have been shown to be a particularly significant source of *Salmonella*. Here, *Salmonella* are harbored in the insulation material. Although dryers operate at a high temperature, the concentrated milk offers protection to the bacteria, and they will survive heating at dryer air inlet and outlet temperatures. *Salmonella* spp. are significant pathogens, and several notable outbreaks of illness have been attributed to the presence of this organism in milk powder.

Staphylococcus aureus is significant, as certain strains can produce a heat-stable toxin that is not destroyed during powder manufacture. Although *Staphylococcus aureus* is common in raw milk, it does not normally grow to produce toxin unless the milk is stored at a high temperature prior to processing. The risk of toxin production increases with temperature and storage time. Although the bacteria will be killed during the process, the toxin remains and can be detected only through specific tests. Large outbreaks of illness have been attributed to the presence of *Staphylococcus aureus* toxin in milk powder.

Another bacterium of potential significance in milk powder is *Bacillus cereus*. This is commonly found in milk, and its spores may survive heat processing. Specialty powders such as infant formula often have specifications for this bacterium, owing to the potential risk of the growth of this organism in warmed milk and sensitivity of the target group of consumers.

Sometimes, yeasts and molds or their toxins, and *Listeria* are included in powder specifications. Yeasts and molds may be significant spoilage organisms if powder is contaminated with moisture, and *Listeria* may contaminate powder from the factory environment, especially if the environment is not kept dry.

In the milk powder process, milk is subjected to heat whilst concentrated under vacuum. Such conditions are conductive to the growth of thermophilic *Bacillus* species that may form biofilms in the process lines. When this occurs, the product may be contaminated with thermophiles that can reach more than 10^6 per gram in long production runs. Thermophiles may sporulate in the process, leading to the presence of large numbers of thermophilic spores in the powder. The spores can be extremely heat-resistant and may not be completely destroyed when the reconstituted powder is used in ultrahigh-temperature (UHT) processes. They are significant because they may cause sterility failures or spoilage in other heated products. If not properly cleaned from the plant between production runs, residues of thermophiles will seed subsequent batches of milk powder.

Although milk powder is a microbiologically stable product, the microbial quality of the raw milk may influence the shelf stability of the powder. Some bacteria present in raw milk, particularly *Pseudomonas* species, produce heat-stable spoilage enzymes, including proteases and lipases, that remain active in milk powder over many months. Experience has shown that lipase can act in full-cream milk powder to degrade milk fat to cause rancidity and other objectionable flavors. Proteases retain activity in milk powder and degrade milk proteins to cause objectionable flavors after the milk powder has been reconstituted. Proteases and lipases may be particularly detrimental in recombined milk products, or if milk powder is used to prepare UHT milk. Here, very low levels of protease and lipase may cause spoilage during long storage periods. (*See* ***Bacillus***: Occurrence; Detection; Food Poisoning; **Biofilms**; ***Listeria***: Properties and Occurrence; **Pasteurization**: Principles; ***Salmonella***: Properties and Occurrence; **Spoilage**: Bacterial Spoilage; Fungi in Food – An Overview; Molds in Spoilage; Yeasts in Spoilage; ***Staphylococcus***: Properties and Occurrence.)

Control of Microorganisms

The manufacture of microbiologically sound milk powder is dependent upon processing good-quality raw milk under hygienic conditions. To ensure the supply of good-quality milk, farm milk should be tested regularly for microbial quality. Many countries now use the total count test to monitor levels of bacteria in farm supplies. Thermoduric counts are sometimes used also. Raw milk ideally should be stored at less than 5 °C and used within 72 h of collection to minimize bacterial growth.

The pasteurization of milk is important and is normally identified as a critical control point. An example of process criteria for pasteurization would be heating of the milk for at least 15 s at 72 °C or 5 s at 80 °C.

Within the factory, application of good manufacturing practice is essential to minimize the risk of milk powder contamination with undesirable types or levels of microorganisms. To achieve this, consideration must be given to the design of the premises and control of staff or vehicular movement to separate raw materials from drying areas. Manufacturing

equipment and the processing environment must be maintained, cleaned, and sanitized to ensure that microbial build-up and spread are prevented. Staff must be trained in practices to maintain high standards of hygiene. A supply of good-quality water and air for the process is also essential. Many factories now have ongoing monitoring systems for *Salmonella* and *Listeria* in place. If these bacteria are detected in the processing environment, special clean-up regimes and extra product testing are implemented.

The modern approach to ensuring that milk powder is microbiologically safe involves preventative management to ensure manufacture under appropriate conditions of hygiene. Many factories now either have in place, or are moving towards, the hazard analysis critical control point (HACCP) system. Although end-product testing is still used to verify compliance and to detect gross process failures, it cannot be relied upon to ensure the safety of a batch of product. Testing can be labor-intensive and time-consuming, taking up to 7 days to obtain final results. To overcome these problems, samples may be composited and rapid techniques for detection of pathogens based on ELISA or DNA methods applied. These methods have advanced efficiencies in testing, and product can now be cleared in 24–48 h. (*See* **Hazard Analysis Critical Control Point; Quality Assurance and Quality Control.**)

Physical Properties

The physical properties of milk powders are governed by process variables, the type of dryer, and the composition of the milk. The physical properties of milk powders play an important role in their use as food ingredients. Their ability to be readily incorporated into products and to perform specific functions in a food formulation can be influenced by many physical properties. There are many physical attributes that must be taken into account when either evaluating a current product, setting specifications for new products or designing or modifying a drying system. Tighter and more demanding specifications have meant that powders are now often manufactured on specialist dryers designed specifically to produce the best possible product of defined specification.

Moisture

The final moisture content is critical for several reasons and is therefore defined in all powder specifications. It can affect functionality and microbiological quality, and is an economic consideration in the manufacture of powders. There are several factors during manufacture that can influence the moisture content of powders. These include the characteristics of the concentrate fed into the dryer, the type of atomization used, and the operating conditions during drying.

Insolubility Index

The insolubility index of a powder is a measure of the degree to which it can be readily solubilized in water prior to use. It is related to the amount of sediment obtained under defined conditions of mixing milk powders. The main reason for loss of solubility is the temperature of the particles during the primary stage of the drying process where the majority of the moisture is removed. During this stage, an impermeable crust can form on the particle surface that severely restricts water removal, leading to the production of case-hardened particles and subsequent loss of solubility.

Bulk Density and Particle Density

Bulk density is the amount of powder by weight that is present in a defined volume. It is usually expressed as g ml^{-1} and is obtained by measuring the volume of a fixed weight of powder after it has been tapped for a defined number of times. A high bulk density is very important in packaging and transportation, and is desirable as it can significantly reduce costs. The bulk density is influenced by a range of factors. These include the amount of air entrapped in the powder particles (occluded air), the overall density of the particle (determined by the composition), the air between the individual powder particles (interstitial air), the particle size distribution and the particle shape. The bulk density of powders is influenced by dryer design and configuration (Tables 2 and 3). Particle density is the density of the solids (determined by the composition), which determines the particle density, together with the amount of occluded air.

Particle Size Distribution

The individual particles produced during drying can vary greatly in size. The distribution of particle size then can be further altered by the degree of agglomeration or after grinding. An indication of the range of particle sizes obtained from different dryer configurations is given in Table 4.

Interstitial Air and Occluded Air

Interstitial air is the amount of air that exists between particles or agglomerates as well as the air inside porous agglomerates. The sphericity of the particles, the particle size distribution, and the degree of agglomeration determine the amount of interstitial air. To obtain minimum interstitial air, the particles need to be smooth, have a range of particle sizes, and be in compact agglomerates.

Table 2 Ability of various spray dryers to manufacture nonagglomerated and agglomerated skim milk powder with low or high bulk densities

Type of drying process	*Chamber configuration*	*Postprimary treatment*	*Atomizer type*	*Nonagglomerated skim milk powder*		*Agglomerated skim milk powder*	
				Low BD[a]	*High BD*[a]	*Low BD*[a]	*High BD*[a]
Single stage	Conventional	None	Rotary/nozzle	Yes	Yes	No	No
	Tall form	None	Nozzle	Yes	Yes	No	No
	Conventional	Cooling bed	Rotary/nozzle	Yes	Yes	No	Yes
	Tall form	Cooling bed	Nozzle	Yes	Ideal	No	Yes
Two stage	Conventional	External bed	Rotary/nozzle	Yes	Yes	No	Yes
	Tall form	External bed	Nozzle	Yes	Ideal	No	Yes
	Compact	Integrated bed	Rotary/nozzle	Ideal	Ideal	No	No
Three stage	Compact	Integrated bed + external bed	Rotary/nozzle	Yes	Yes	No	Yes
	Multistage	Integrated bed + external bed	Rotary/nozzle	Yes	Yes	Ideal	Ideal
	Integrated belt	Integrated belt	Nozzle	Ideal	No	Yes	Yes

[a]BD = bulk density; nonagglomerated skim milk powder: low bulk density $\leq 0.72\ g\ ml^{-1}$; high bulk density $\geq 0.72\ g\ ml^{-1}$; agglomerated skim milk powder: low bulk density $0.30–0.50\ g\ ml^{-1}$; high bulk density $0.45–0.55\ g\ ml^{-1}$.
Adapted from Pisecky J (1997) *Handbook of Milk Powder Manufacture*, p. 79. Copenhagen: Niro A/S.

Table 3 Ability of various spray dryers to manufacture nonagglomerated and agglomerated full-cream milk powder with low or high bulk densities

Type of drying process	*Chamber configuration*	*Postprimary treatment*	*Atomizer type*	*Nonagglomerated full-cream milk powder*		*Agglomerated full-cream milk powder*	
				Low BD[a]	*High BD*[a]	*Low BD*[a]	*High BD*[a]
Single stage	Conventional	None	Rotary/nozzle	Yes	Yes	No	No
	Tall form	None	Nozzle	Yes	Yes	No	No
	Conventional	Cooling Bed	Rotary/nozzle	Yes	Yes	No	Yes
	Tall form	Cooling Bed	Nozzle	Yes	Yes	No	Yes
Two stage	Conventional	External bed	Rotary/nozzle	Yes	Yes	Yes	No
	Tall form	External bed	Nozzle	Yes	No	Ideal	No
	Compact	Integrated bed	Rotary/nozzle	No	Ideal	Ideal	No
Three stage	Compact	Integrated bed + external bed	Rotary/nozzle	Yes	Yes	Yes	No
	Multistage	Integrated bed + external bed	Rotary/nozzle	Ideal	Yes	No	Ideal
	Integrated belt	Integrated belt	Nozzle	Yes	No	No	Yes

[a]BD = bulk density; nonagglomerated full-cream milk powder: low bulk density $\leq 0.63\ g\ ml^{-1}$; high bulk density $\geq 0.63\ g\ ml^{-1}$; agglomerated full-cream milk powder: low bulk density $0.30–0.50\ g\ ml^{-1}$; high bulk density $0.45–0.55\ g\ ml^{-1}$.
Adapted from Pisecky J (1997) *Handbook of Milk Powder Manufacture*, p. 79. Copenhagen: Niro A/S.

Table 4 Mean particle size obtained from dryers of different configuration

Powder characteristics	*Dryer configuration*	*Particle size (μm)*
Individual particles	Concurrent with pneumatic conveying	20–200
	Tall form – tower	30–250
Flakes	Roller dryer	200–5000
Loose agglomerate	Mixed flow with integrated fluid bed	100–400
– open structure	Concurrent with integrated fluid bed	100–200
Compact agglomerate	Concurrent spray dryer with integrated belt	300–2000
– porous structure	Mixed flow with integrated fluid bed	100–400

From personal communication (E. Refstrup), Niro A/S, Denmark.

Occluded air is the amount of air entrapped within the powder particles. It is affected by the preheat treatment of the original milk, with a higher pretreatment of milk resulting in less occluded air, and the amount of air incorporated in the concentrate. Higher total solids generally result in lower occluded air. Powders atomized by a nozzle contain less air than rotary atomized powders, despite improvements to the modern rotary atomizers. Gentle drying also reduces the level of occluded air, and therefore, the use of multistage dryers is recommended for the production of powders with low occluded air.

Flowability

With the ever-increasing diversity of use of milk powders today, the need for properties such as

flowability is increasing. Powders are used in applications ranging from dispensing machines through to the large-scale recombining operations that utilize mechanical handling and dosing. For both agglomerated and nonagglomerated powders, a better flowability can be obtained by producing larger powder particles with smooth and rounded particle surfaces within a narrow particle size distribution. Flowability is also influenced by other factors such as total fat in the powder and the amount of 'free fat.'

'Free Fat'

'Free fat' in powder is defined as the fat fraction that is extractable by organic solvents under specific conditions of solvent type, time, and temperature of extraction. In most instances, 'free fat' is considered a defect. The exception is where 'free fat' is required for a specific application, e.g., chocolate manufacture. One of the most critical influences of 'free fat' is the moisture content of the powder. If the moisture is too low (< 2.5%), 'free fat' increases and then decreases as the moisture content is raised from 2.5 to 4–5% but increases again if the moisture content is > 6–7%.

Instant Properties

Very fine powder particles are difficult to handle and have poor reconstitution properties. Agglomeration of powders allows water to permeate the powder particles more readily, breaking up the agglomerate and allowing the individual powder particles to dissolve. Instant milk powder is highly soluble and designed to reconstitute completely in water at both hot and cold temperatures. The other properties required in instant powders are wettability and dispersibility. The wettability of a powder is measured by determining the time taken for a given amount of powder to pass through the surface of water. Wettability may be enhanced by lecithination. The dispersibility of a powder is a measure of how completely a powder dissolves under controlled conditions. Other tests carried out on milk powders related to their instant properties include slowly dispersible particles, coffee test, white flecks number (minute particles that are seen on the surface of reconstituted milk), and the sludge test.

Color

The color of a powder is determined by composition, preheat treatment, drying conditions, and particle size distribution. Scorched particles can be a visual defect that will often show up as deposits on the bottom of mixing vats and in strainers.

Other Properties

Apart from the properties described above, there are others that influence a powder's acceptability. These include the mechanical stability of the powder, which influences the degree of agglomeration breakdown during transport and storage, hygroscopicity, which is related to the degree of water attraction a powder exhibits, and cakiness, an attribute that is a measure of the extent to which a powder adheres to itself, especially under compression. (*See* **Agglomeration**; **Drying**: Spray Drying; **Rheological Properties of Food Materials**.)

Chemical Characteristics

The chemical properties of milk powders are determined by the composition of the milk and the heat treatment applied during powder manufacture.

Chemical Composition

Skim and full-cream milk powders are obtained by dehydration of skim milk and full-cream milk to ~ 4% moisture. Full-cream milk is usually standardized to a fat:solids-nonfat ratio of 1:2.67 to meet the 26% legal minimum fat content for this powder. The protein content of skim milk powders may be standardized also. Variations in milk composition owing to factors such as cow breed, feed, stage of lactation, and season are reflected in the composition of milk powders. The American Dairy Products Institute standards for skim and full-cream milk powder compositions are as follows: skim milk powder should have a maximum fat content of 1.25% and a maximum water content of 4.0%, whereas full-cream milk powder should have a minimum fat content of 26% and a maximum water content of 4.0%. Control of the moisture content of milk powders to a maximum of 4% is essential for good shelf-life stability. Table 5 shows the range of values observed in milk powder.

Table 5 Composition of milk powders[a]

Constituent	*Skim milk powder*	*Full-cream milk powder*
Moisture (g per 100 g)	3–5	2–4
Fat (g per 100 g)	0.7–1.3	25–28
Crude protein (g per 100 g)	35–37	25–27
Lactose (g per 100 g)	49–52	36–38
Citric acid (g per 100 g)	1.8–2.1	1.3–1.4
Ash (g per 100 g)	7.5–8.0	6.0–7.0
Sodium (mg per 100 g)	400–550	370–420
Potassium (mg per 100 g)	1550–1750	1150–1350
Calcium (mg per 100 g)	1200–1300	900–1000
Magnesium (mg per 100 g)	110–140	85–100
Phosphorus (mg per 100 g)	950–1050	700–770
Chloride (mg per 100 g)	~1100	750–800

[a]Adapted from Walstra P and Jenness R (1984) *Dairy Chemistry and Physics*, pp. 418–419. New York: John Wiley.

Another important indicator of milk powder quality is the titratable acidity of the reconstituted powder. This is an indicator of the microbiological quality of the milk. The American Dairy Products Institute sets a maximum of 0.15% for titratable acidity of skim milk powder.

Heat-treatment Classification

The characteristics of milk powder can be influenced by the heat treatment received by the milk powder during manufacture. The time and temperature of the preheat treatment affects the level of whey-protein denaturation. The whey-protein nitrogen index, which is a measure of the undenatured whey-protein nitrogen in the powder and was developed by the American Dairy Products Institute, is commonly used to classify powders into low-heat, medium-heat and high-heat milk powders. Typical preheat treatments used for the manufacture of these powders are listed in Table 6. As the composition of milk, including the initial level of whey proteins in milk, can vary with season, the same heat treatment can result in a different whey-protein nitrogen index. Other methods for heat classification of milk powders, such as the heat number, cystine number, and thiol number also may be used as a measure of the heat treatment given to the milk during powder manufacture. (*See* **Heat Treatment**: Ultra-high Temperature (UHT) Treatments.)

Functional Properties of Milk Powders

When milk powders are used as ingredients in food applications, they contribute to the physical attributes of the food. The ability of milk powders to impart desirable properties to food is related primarily to functional properties of milk-protein components in the powders. These functional properties include solubility, hydration, heat stability, viscosity, gelling, foaming, and emulsifying. In milk powders, the functional properties of the milk proteins may be modulated by heat, ions, and other components. Heat treatment of milk prior to concentration and drying is the most common method used to alter the functional properties of milk powders. Milk powders with the same composition given different preheat treatments prior to concentration and drying have different functional attributes when used as ingredients.

Table 6 Heat classification of skim milk powder

Heat class	*Whey protein nitrogen index*[a] *(milligrams of undenatured whey protein N per gram of powder)*	*Preheat treatment of milk*[b]
Low heat	Not less than 6	72 °C for 15 s
Medium heat	1.51–5.99	75 °C for 3 min
High heat	Not more than 1.5	90 °C for 10 min 120 °C for 2 min

[a]From American Dairy Products Institute (1990) *Standards for Grades of Dry Milks including Methods of Analysis, Bulletin 916*.
[b]A range of other preheating conditions may be used to achieve a desired whey protein nitrogen index.

Solubility

Solubility is a fundamental functional property that is a prerequisite for most other desired functionalities. The solubility of milk powders is dependent on pH. Proteins have a minimum solubility at the isoelectric pH, and solubility is increased on the acid and alkaline side of this pH. Caseins, the major proteins in milk, are least soluble at pH 4.6.

Hydration

Hydration is related to the ability of the milk proteins to bind or entrap water. Caseins hold about 3.3 g of water per gram, whereas undernatured whey proteins hold ~0.4 g of water per gram. Heat denaturation of whey proteins increases the water holding to 2.5 g of water per gram. Milk powder contains other components, such as lactose, that bind water in addition to the protein. Skim milk powders have a water sorption of 0.96–1.28 g water per gram, depending on the conditions used during powder manufacture.

Heat Stability

Heat stability is an important property in certain applications such as the manufacture of recombined evaporated milk. Single-strength milks made from low-, medium- or high-heat milk powders have a similar heat stability to fresh milk. They are heat-stable at the pH of milk (pH 6.7), being able to withstand coagulation for up to ~20 min at 140 °C. However, for adequate heat stability of evaporated milks under sterilization conditions (e.g., 120 °C for 12 min), high-heat milk powders are necessary. Heat stability is affected by the pH of the milk, mineral content, and other components in the milk (e.g., lecithin, urea).

Viscosity

The viscosity of milks reconstituted from milk powders is dependent on their state of dispersion, concentration of solids, and temperature. Increasing the concentration of milk solids increases the viscosity. Decreasing the temperature increases the viscosity, but heating milk to a temperature that results in denaturation of whey proteins also increases the viscosity.

Gelation

Milks reconstituted from milk powders have the ability to form gels under similar conditions to those required for the formation of gels from fresh milk, i.e., by rennet action for formation of rennet gels and by acidification of milk under quiescent conditions.

Foaming and Emulsifying

Milk powders can be used in applications where foaming and emulsifying properties are required. The ability of milk proteins in the milk powders to stabilize foams and emulsions may be exploited when these properties are required. (*See* **Aerated Foods**; **Emulsifiers**: Uses in Processed Foods; **Mixing of Powders**.)

Functional Requirements of Milk Powders in Major Food Applications

For milk powders to have the desired performance in food applications, the functional characteristics of the powders have to be matched to the application. This requires an understanding of the required functional properties of the milk powder ingredients in the target application.

Milk Powders for Recombined Dairy Products

A significant amount of milk powder is used in the manufacture of reconstituted and recombined dairy products. In these applications, the milk powders are combined with water and milkfat to reestablish the fat:solids-nonfat:water ratio of milk or other dairy products. Some of the major applications of milk powders in the recombination industry are for the preparation of pasteurized fluid milk, UHT milk, cream, evaporated milk, sweetened condensed milk, yogurt and cultured dairy products, recombined cheese, and icecream. Different functionalities of the milk powder ingredients are needed in these various recombined dairy products. Table 7 lists the major functional requirements of milk powders for recombined dairy products.

Table 7 Functional requirements of milk powders in recombined dairy products and selected processed foods

Product	*Functional properties required in milk powder*	*Heat treatment of milk powder*
Pasteurized milk	Good flavor Emulsifying	Low–medium heat
UHT milk	Good flavor Heat stability Emulsifying	Low–medium–high heat
Cream	Good flavor Emulsifying	Low–medium heat
Evaporated milk	Heat stability Viscosity	High heat
Sweetened condensed milk	Viscosity	Low–medium heat
Yogurt	Water-binding Viscosity Gelling	Low heat[a]
Cheese	Rennetability	Low heat
Icecream	Foaming/whipping Emulsifying	Low–medium–high heat
Confectionery	Water-binding Foaming/whipping Emulsifying Heat stability	High heat
Bakery	Water-binding Foaming/whipping Emulsifying Gelling	High heat

[a]If a low-heat milk powder is used, the yogurt milk has to be given a high-heat treatment during yogurt manufacture. Alternatively, a high-heat milk powder may be used, in which case, the yogurt milk requires only a low-heat treatment to pasteurize the milk during yogurt manufacture.

Pasteurized milks and UHT milks These products have a similar composition to fresh milk. For pasteurized milks, low-heat or medium-heat powders are used to obtain a flavor similar to milk and to minimize heat-induced flavors. In the case of UHT milks, any type of powder can be used, as single-strength milks made from low-, medium-, or high-heat powders are stable to UHT conditions.

Evaporated milks It is essential to use high-heat powders for this application to obtain evaporated milk with the desired viscosity. A high-preheat treatment improves the heat stability of a recombined milk concentrate (typically 26% total solids; 18% solids-nonfat: 8% fat) to in-can sterilization conditions used in its manufacture. Additionally, high-heat powders are screened using heat-stability tests to ensure that they withstand sterilization without excessive thickening or coagulation.

Sweetened condensed milk This is a traditional dairy product containing 74% total solids (20% milk solids nonfat: 8% fat: 46% sucrose). The most important physical attribute of this product is its viscosity. Low- and medium-heat powders are used in this application. Milk powders given a high-heat treatment (e.g., 85 °C for 30 min) cannot be easily processed, because the high viscosity of concentrates made from these powders also results in rapid age thickening during storage of the product. There are a number of viscosity tests that may be used as indicators of suitability of powders for sweetened condensed milk manufacture.

Yogurt Milk powders may be used as a partial or total replacement for fresh milk in this application. Viscosity development, gelling, and good water-binding properties are necessary for the production of high-quality yogurts. These properties are obtained in yogurt by preheating the yogurt milk at a temperature that causes significant denaturation of whey proteins (e.g., 90 °C for 10 min). Low-heat milk powder may be used if a high-heat treatment is given during yogurt manufacture. If a high-heat milk powder is used, the yogurt milk requires only a low-heat pasteurization treatment during yogurt manufacture.

Cheese Only low-heat milk powders are suitable for recombined cheese manufacture. This ensures good rennetability of the reconstituted milk. With a high-heat treatment of milk, there is association of the denatured whey proteins with the casein, which hinders the reaction of the rennet.

Ice cream Milk powders contribute to the flavor and texture of ice cream. The milk powder aids in the emulsification of the ice cream mix and has a role in the development of an aerated matrix. (*See* **Condensed Milk**; **Evaporation**: Basic Principles; Uses in the Food Industry; **Recombined and Filled Milks**.)

Milk Powders for Selected Food Applications

Milk powders are used as functional ingredients in a number of processed foods.

Chocolate and confectionery products Milk powders contribute to the flavor, color, and texture development in chocolate and confectionery applications. The emulsifying properties of the milk proteins influence the miscibility of the ingredients used in chocolate and confectionery, hence influencing flow properties and texture. In confectionery products such as toffee, good water-binding properties of milk proteins contribute to the texture of these products. The Maillard reaction, which is the reaction of the amino groups of the proteins with reducing sugars in the formulation, is responsible for color development and for the production of caramelized flavors; the lactose in milk powders participates in the Maillard browning reactions.

Bakery products High-heat milk powders are useful in bakery applications. In addition to enhancing the nutritive value of cereal-based baked goods, milk powders contribute to the texture and flavor of these products. Their emulsification and foam-stabilization properties and their ability to participate in the Maillard browning reaction are important requirements in bakery applications.

Other applications The functional properties of milk powders also make them useful in a number of other applications, such as processed meat products, soups, gravies, and dips.

Characteristics of Stored Milk Powders

The characteristics of milk powder are dependent on the quality and composition of the raw milk and the manufacturing process used during its manufacture. However, even if milk powders are manufactured to meet the desired standards and specifications, changes in the properties of milk powders may occur during storage and distribution. The composition of the powder, the type of packaging material used, and the conditions of handling and storage influence the shelf-life of the powder.

Deterioration of milk powders resulting from Maillard browning, lactose crystallization, and oxidation of fat may lead to flavor and physical defects in the powder. It may also affect the functionality of the milk powder when it is used in a food product. Some of the changes that may occur during storage include the development of a brown color, a reduction in pH, reduced solubility, development of off-flavors, and reduced heat stability of powders.

See also: **Biofilms**; **Condensed Milk**; **Emulsifiers**: Uses in Processed Foods; **Evaporation**: Basic Principles; Uses in the Food Industry; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; **Mixing of Powders**; **Pasteurization**: Principles; **Rheological Properties of Food Materials**; **Powdered Milk**: Milk Powders in the Marketplace; **Quality Assurance and Quality Control**; **Recombined and Filled Milks**

Further Reading

American Dairy Products Institute (1990) *Standards for Grades of Dry Milks Including Methods of Analysis, Bulletin 916*, rev ed. Chicago, IL: ADPI.

Early R (1998) Milk concentrates and milk powders. In: Early R (ed.) *The Technology of Dairy Products*, 2nd edn, pp. 228–300. London: Blackie Academic & Professional.

International Commission on Microbiological Specifications for Foods (1998) *Microorganisms in Foods 6, Microbial Ecology of Food Commodities*. London: Blackie Academic & Professional.

International Dairy Federation (1984) *General Code of Hygienic Practice for the Dairy Industry and Advisory Microbiological Criteria for Dried Milk, Edible Rennet*

Casein and Food Grade Whey Powders, Bulletin No. 178. Brussels: International Dairy Federation.

International Dairy Federation (1990) *Recombination of Milk and Milk Products, Special Issue No. 9001*. Brussels: International Dairy Federation.

International Dairy Federation (1991) *IDF Recommendations for the Hygienic Manufacture of Spray Dried Milk Powders, Bulletin No. 267*. Brussels: International Dairy Federation.

International Dairy Federation (1999) *3rd International Symposium on Recombined Milk and Milk Products, Special Issue No. 9902*. Brussels: International Dairy Federation.

Knipschildt ME and Andersen GG (1994) Drying of milk and milk products. In: Robinson KR (ed.) *Modern Dairy Technology*, 2nd edn., vol. 1, pp. 159–254. London: Chapman & Hall.

Masters K (1997) Spray dryers. In: Baker CGJ (ed.) *Industrial Drying of Foods*, pp. 90–112. London: Blackie Academic & Professional.

Mettler AE (1994) Present day requirements for effective pathogen control in spray dried milk powder production. *Journal of the Society of Dairy Technology* 47: 95–107.

Pisecky J (1986) Standards, specifications, and test methods for dry milk products. In: MacCarthy D (ed.) *Concentration and Drying of Foods: Proceedings of the Kellogg Foundation 2nd International Food Research Symposium (Cork: 1985)*, pp. 203–220. London: Elsevier Applied Science.

Pisecky J (1997) *Handbook of Milk Powder Manufacture*. Copenhagen: Niro A/S.

Walstra P and Jenness R (1984) *Dairy Chemistry and Physics*, pp. 418–419. New York: John Wiley.

POWER SUPPLIES

Use of Electricity in Food Technology

D Graham, R and D Enterprises, Walnut Creek, CA, USA

Introduction

The food industry in the USA consumed 17.4 billion kcal in 1999. Fifty-eight percent of total food industry energy costs are from electricity. The industry purchases 64.0 billion kWh of electricity annually at a cost of $3.36 billion. Purchased energy is 1.3% of the value of product shipments for the food industry compared with 1.7% for the total US industry.

Major uses of electricity include freezer and refrigeration compressors, conveyors, air handling, pumping, lighting, process controllers and monitors, and packaging forming and sealing. Natural gas is the primary source of thermal energy for ovens, fryers, dryers, evaporators, and boilers, with relative costs being the main reason. Closely related uses of electricity in agricultural food production include crop irrigation, pest control, produce disinfection, grain harvesting and storage, weed reduction, livestock waste management, and fish farming.

Irradiation, ohmic heating, microwave processing, ozonation, freeze concentration, nonthermal pasteurization, and the use of heat pumps are examples of relatively new electrotechnologies that the food industry may apply increasingly in the future. Application of these technologies provides opportunities for food processors to improve operating efficiencies and helps insure the quality and safety of processed food products.

Electricity Use by Food Industry Sectors

Food energy costs rose steadily over the past two decades. In 1996, purchased electricity comprised over half of the total energy used in food processing. Electrical consumption by grain milling, meat processing, preserved fruits and vegetables accounts for 23, 19, and 13%, respectively, of the total food industry use of electricity. Some food processors operate cogeneration facilities using surplus heat to produce electricity, but approximately 92% of total electricity used by the food industry in the USA is purchased from utility companies. Approximately 87% of the electricity is used by motor-driven equipment such as compressors, pumps, mixers, grinders, fans, etc. **Table 1** compares the cost of electricity with total energy costs in food processing. Electricity consumption has increased approximately 3% annually from 48.9 billion kWh in 1986 to 69.1 billion kWh in 1999. **Table 2** shows the trends in electric energy cost from 1980 to 1996.

New Electrotechnologies for Food Processing and Preservation

Electron beam irradiation, X-ray, microwave processing, membrane separation technology, ozonation, ohmic heating, high pressure pasteurization, infrared

Table 1 Electricity vs total energy used in food processing

	Electricity ($ million)	*Total energy ($ million)*	*Electricity as a percentage of the total*
Total food industry	3364	5799	58.0
Meat products	614	884	69.5
Dairy products	401	632	62.4
Preserved fruit/vegetables	459	804	57.1
Grain mill products	683	1142	59.8
Bakery products	241	426	56.6
Sugar/confectionery	177	424	41.7
Fats and oils	180	456	30.8
Beverages	345	584	59.1
Miscellaneous	464	447	59.1

Source: 1996 Annual Survey of Manufactures, US Dept. of Commerce, with permission.

Table 2 Growth of electricity use in food processing as a percentage of the total energy cost

	1980	*1986*	*1994*	*1996*
Total food	41.0	51.8	57.4	58.0
Meat products	51.4	59.6	69.2	68.5
Dairy products	47.3	58.8	66.1	64.3
Preserved fruit/vegetables	37.3	52.1	55.8	57.1
Grain mill products	44.0	53.0	57.5	59.8
Bakery products	43.3	47.9	55.7	56.6
Sugar/confectionery	n/a	35.9	34.9	41.7
Fats and oils	26.8	40.1	42.5	39.4
Beverages	43.1	52.1	61.1	59.1
Miscellaneous	48.0	54.8	53.8	59.1

Source: Annual Survey of Manufactures. US Dept. of Commerce, with permission.

radiation, laser, ultraviolet light, freeze concentration, and pulsed electric fields are developing technologies that offer future potential for improvements in food safety, quality, and more efficient processing.

Irradiation of Food Products

Treatment of a food product with a controlled source of irradiation can be very effective for improving the safety of the food product, since the irradiation can destroy any viable food pathogens that may be present. Food products are not made radioactive when irradiated, and no toxic byproducts are formed. Irradiation has been used successfully on a variety of products, including cereals, fruits, vegetables, prepared foods, dairy products, meat and poultry, beverages, spices, and seasonings. Products can be irradiated after being sealed in packaging material, which reduces the possibility of contamination by handling after treatment.

Safety of irradiated foods Controversy has surrounded the use of food irradiation since its discovery in the late 1940s. Much concern has resulted from fear that the irradiation process might make the food radioactive. Treatment of food with gamma irradiation does not inject radioactivity into the food. High-energy electrons pass through the food product and, if sufficient energy is applied, kill microorganisms that they impact. In approving food irradiation, the Food and Drug Administration (FDA) reviewed over 400 toxicity studies, including animal feeding studies before issuing any approvals. The FDA concluded that food irradiated up to 1 kGy is "wholesome and safe for human consumption, even where the food that is irradiated may constitute a substantial portion of the diet . . . "

Electron beam irradiation may provide an effective way to control harmful *E. coli* 0157:H7 and other emerging pathogens in ground beef and other processed meat products such as pork sausage and refrigerated poultry. Electron beams (gamma rays) from machine sources up to 10 million electron volts (10 MeV) and X-rays generated from machine sources up to 5 MeV are permitted commercially. Similar energetic rays also can be produced by decay from sealed units of radioactive sources such as cobalt-60 or cesium-137. Gamma rays are used effectively for sterilizing a variety of products including food packaging materials, syringes, bandages, and other heat-sensitive pharmaceutical products. X-rays also are effective in some applications. One kilogray is 1 J of energy absorbed per kilogram of the substance being irradiated. The FDA has approved electron beam radiation providing absorbed doses of radiation from 0.2 to 7.0 kGy, depending of the particular food product. Table 3 lists the radiation levels approved by the FDA for food products.

Microwave Heating

Microwave ovens found extensive use in home food preparation and in commercial food service during the past two decades. Industrially, microwaves are used primarily for tempering, thawing large blocks of frozen meat, and preheating. Studies of microwave application for sterilization of food products have shown limited success.

Microwaves are electromagnetic waves similar to TV, radio, light, radar, or infrared waves, with the main difference being the frequency of the wave motion. Food products contain molecules such as water, salts, and proteins that have dipolar properties. Microwave energy passing through a food product causes these dipolar molecules to vibrate. The resulting internal friction produces heat. Thus, the food product is heated from within, and heating time is greatly reduced compared with heating by external application of heat. This rapid heating provides improved food quality and energy savings in many applications.

Table 3 Food irradiation processes approved by the FDA

Year approved	*Food product*	*Dose (kGy)*	*Purpose*
1963	Wheat and wheat flour	0.2–0.5	Disinfestation of Insects
1964/5	White potatoes	0.05–0.15	Inhibit sprouting
1983	Spices and seasonings	30 max.	Disinfestation of insects
1985	Pork, fresh unheat-processed	0.3–1.0	Control of *Trichinella spiralis*
1985/6	Dry of dehydrated enzymes	10 max.	Control of insects and pathogens
1986	Fruit	1 max.	Disinfestation and delay ripening
1986	Fresh vegetables	1 max.	Disinfestation of insects
1986	Herbs and spices	30 max.	Control of microorganisms
1986	Vegetable seasonings	30 max.	Control of microorganisms
1990	Poultry, fresh or frozen	3 max.	Control of microorganisms
1995	Meat, frozen and packaged (limited to use by NASA)	44 minimum	Sterlization
1995	Animal feed and pet food	2–25	Control of *Salmonella*
1997	Red meat, uncooked, chilled	4.5 max.	Control of microorganisms
1997	Red meat, frozen	7.0 max.	Control of microorganisms

Table 4 Microwave energy penetration

	Material temperature (°C)	*Half power depth (cm)*	
		915 MHz	*2450 MHz*
Ice	−12	1600	780
Water	2	4	0.6
Water	45	14	2
Water	75	21	3
Water	95	30	5
Beef, lean	−51	70	46
Beef, lean	−18	10	8
Beef, lean	4	2	2
Beef, lean (freeze-dried)	−18	550	190

Microwaves are in the radio frequency portion of the electromagnetic spectrum between 300 and 300 000 MHz. Microwaves are generated by a magnetron that converts electrical energy into an electromagnetic field with centers of positive and negative charges that change direction billions of times a second. In the USA, the Federal Communication Commission limits commercial microwave sources to 915 and 2450 MHz.

Microwave heating is used for cooking or partial cooking of meat, chicken, other prepared foods, rapid thawing of frozen meat, poultry, and seafood, drying of pasta, fruits, and cereal products, freeze drying of meat and juices, and, pasteurization of milk, yogurt, beer, and wine. Table 4 shows the penetration of microwave energy into a variety of food substances at different temperatures.

Ohmic Heating

Ohmic heating describes an innovative process in which an electric current is passed directly through a food product to generate heat by internal resistance without any need for intermediate heat exchangers. The electric energy is applied from electrodes inserted into the product and is converted directly to thermal energy within the food product with an overall conversion efficiency of 90%.

Food products sterilized by ohmic heating undergo less damage to flavor and color as compared with conventional heating methods. Electrical conductivity of the food product influences the flow of electric current through the food and can cause uneven heat distribution. This has been troublesome with some prepared foods that contain ingredients of varied composition, size, and shape. Ohmic heating can process a wide variety of pumpable products including stews, dairy sauces, eggs, custards, soups, diced fruits, pie fillings, etc. Microorganisms present in the food are killed by heat, just as in conventional heat processing. The temperatures required to assure commercial sterility are identical to the temperatures required in conventional heat processing. The ohmic advantage stems from the rapid temperature increase due to internally generated heat, making conductance of heat through the food to the center of the mass unnecessary, and the 90% conversion of electrical energy to heat. Conventional heat exchangers achieve only about 50% transfer of heat to the food product. Figure 1 shows a typical electrode assembly for ohmic heating.

A typical 75-kW commercial unit capable of heating $750\,kg\,h^{-1}$ uses a 5-cm-diameter electrode housing. A 300-kW, $3000\,kg\,h^{-1}$ unit requires a 10-mm-diameter housing. A basic ohmic heating system typically consists of four electrodes in series connected by short spacer tubes. The top and bottom electrodes are connected to one phase of a three-phase transformer and are grounded. Each of the two middle electrodes is connected to one of the other phases of the transformer. The voltage required for heating the product is varied by adjusting the primary side of the transformer.

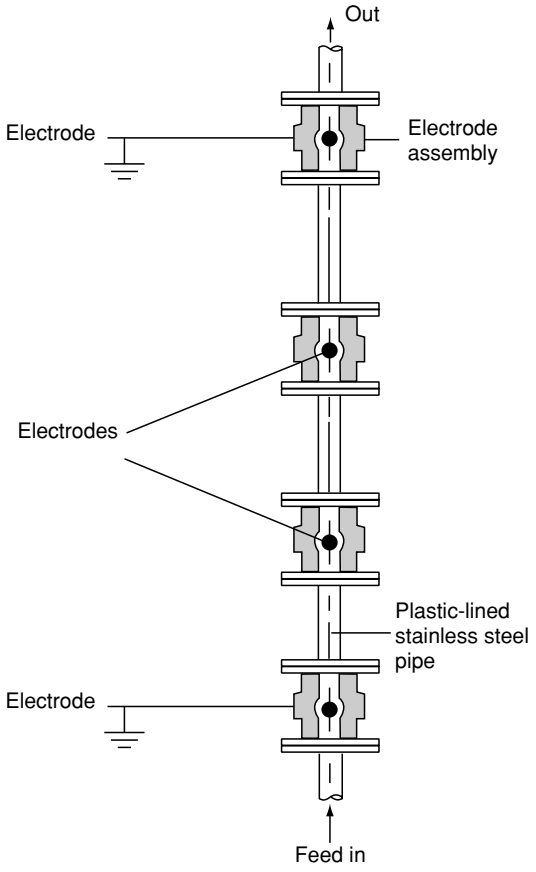

Figure 1 Typical electrode assembly. Drawing courtesy of APV.

Membrane Technology

Membrane technologies separate fluid mixtures in many different applications. They can separate materials based on size, electrical nature, physical properties, and are used widely in the food processing and chemical industries, as well as in water treatment. All membrane processes are electrically driven, using high-velocity/high-pressure pumps.

This technology originated in the early 1930s with cellulose acetate membranes, and has developed rapidly as new and improved media have been created. The arsenal of membranes now includes a wide range of polymeric membranes, ceramic, and sintered stainless steel. Among many applications are: concentration of fluid products; sterilization by removal of bacteria or viruses; removal of constituents such as proteins, acids, and fats from process streams; desalination; purification; demineralization; refining of oils; and clarification.

Membrane separation could be considered as a modern, highly efficient, versatile form of the traditional filtering process. The objective is still the same, i.e., to remove one component such as small solid particles from a second component such as a liquid or to separate a given molecular component from a solution.

The efficiency with which separations can be made and the wide range of components that can be processed with membranes are noteworthy. Separations can be based on particulate size differences of a less than 1 μm down to a molecular level. Separation depends on the passage of specific molecules through a semipermeable membrane while the membrane retains other molecules.

In traditional simple filtration, the osmotic pressure due to differences in concentration on the two sides of a membrane, a differential pressure, and/or gravity were the only driving forces. In membrane processes, pumps apply high pressure and high cross-flow velocity to the solution being filtered to speed up the separation process. High-velocity turbulent flow at the membrane/liquid interface continuously cleans the membrane surfaces and extends their useful life. In some instances, when strongly highly ionic systems are being filtered, an electrical potential will be applied to accelerate the process and increase selectivity of the separation process.

Membrane separations typically are classified into subtechnologies based on the pore size of the filter media. Common classifications include microfiltration (coarsest), ultrafiltration, nanofiltration, and reverse osmosis (finest). Figure 2 illustrates the tremendous range and flexibility of the filtration spectrum. Particle sizes are measured in microns and alternatively stated as Daltons. (MWCO, molecular weight cutoff).

The ability to concentrate delicately flavored solutions such as fruit juices or dairy products without the application of heat provides the food processor with the potential to significantly improve product quality. The ability to separate components selectively from a product stream or effluent stream can lead to the creation of new products, allow recycling or reclaiming of valuable ingredients, or simplify and reduce the cost of effluent disposal.

Membrane filtration has been commercially used in the USA since about 1930, primarily in the water and beverage industries. Widespread applications and use were initially limited due to the lack of practical membranes. Early membranes tended to clog easily, were hard to clean, and had short process lives. Development of new polymeric membranes, or ceramic membranes in the 1980s provided the food processor with filtering systems with greater durability, greater efficiency, far more versatility and lower costs. These membranes have a higher capacity, are less prone to clog, are easier to clean, and withstand multiple

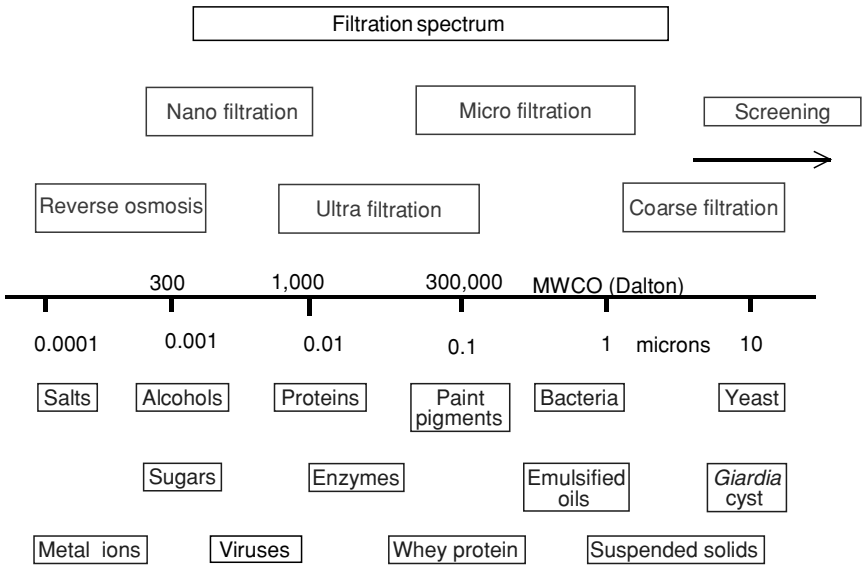

Figure 2 Membrane filtration spectrum. Drawing courtesy of J. Strasser, Process & Equipment Technology.

filtration/cleaning cycles. Many of today's routine commercial applications of membrane separation technology were only dreams two decades ago.

The cost for many types of membranes has dropped significantly in the past 10 years, e.g., from $800 to $350 for similarly sized spiral wound modules. Energy efficiencies have improved partially due to lower pressure differential required across the membranes, higher throughput and more sophisticated computer controls. These efficiencies, coupled with lower cost membranes, have made membrane separation processes not only more cost-effective but also more dependable and easier to maintain and operate.

Membrane separation technology is finding multiple industrial applications not only in food processing but also in process water and plant effluent treatment. Most food processors are very aware that water management has become a major consideration to an efficient operation. Effluent discharge requirements have become more stringent, dependable supplies of process water can no longer be taken for granted, and both effluent disposal and water supply costs have increased.

Ozone: Antimicrobial Agent

Ozone (CAS No. 10028-15-6) is a gas at ambient and refrigerated temperatures and is partially soluble in water. In the stratosphere, UV and lightening generate ozone naturally from atmospheric oxygen; ozone is generated industrially by passing oxygen through a high-voltage electric corona discharge field.

Ozone (O_3) is a molecule composed of three oxygen atoms, in contrast to ordinary oxygen (O_2), which has only two oxygen atoms. Because of its atomic structure, ozone is an unstable gas that quickly decomposes into ordinary oxygen, especially in water. This tendency to decompose naturally makes ozone an important chemical. As ozone changes into oxygen, the extra oxygen atom splits off from each ozone molecule. These free oxygen atoms have two important characteristics: they are toxic to microorganisms, and they oxidize many chemical compounds, usually changing them into nontoxic substances.

Ozonation is used extensively for sanitizing bottled water, and may become widely used as an antimicrobial agent in processing foods for human consumption. Ozone is a powerful oxidant, can be dissolved in water, and can be used effectively to kill harmful microorganisms such as *Listeria*, *E. coli*, *Salmonella*, and other pathogenic bacteria, viruses, fungi, and cyst organisms such as *Giardia*. Ozone degrades to oxygen in a few minutes and thus leaves no chemical residue like chlorine, iodine, or other common disinfectants. Current food uses in the USA include aqueous ozone for cleaning sausage curing racks, washing apples, garlic cloves, strawberries, and fresh-cut produce including celery and lettuce; gaseous ozone is used in dry storage of onions, garlic, potatoes, and citrus products.

A schematic of a typical corona discharge system used to generate ozonated water from dry air or oxygen is shown in Figure 3. Figure 4 shows an ozone system installed in a food processing plant. In the foreground is a pressure swing absorption (PSA) air dryer. The sealed cabinet in the corner houses the ozone generator; the cabinet to the right contains the electronic operating controls for the system.

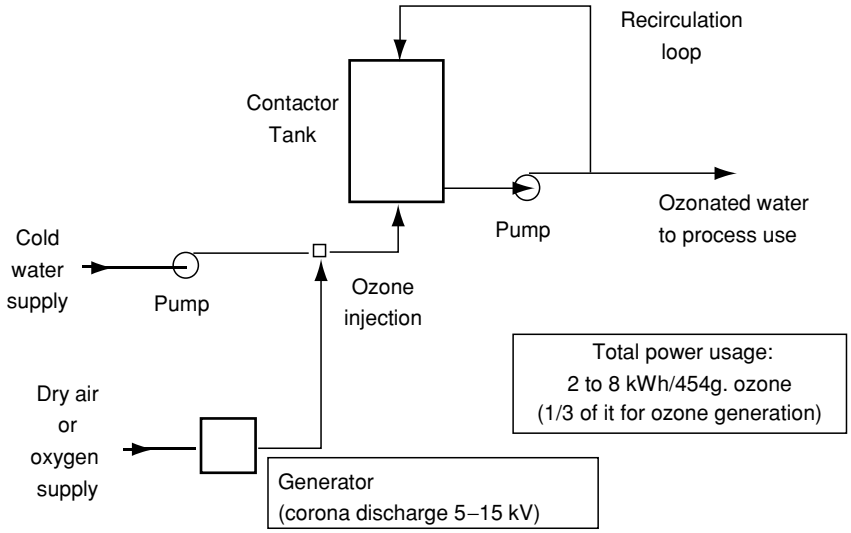

Figure 3 Typical system for generating ozonated water.

Figure 4 Ozone system installed in a grain-processing plant. Photo courtesy of RGF Environmental Systems.

Freeze Concentration

Separation is an important step in every food processing operation. Evaporators heated with steam produced from coal or gas fired boilers is the dominant separation process used in the food industry. Freeze concentration (FC) technology is an electrically driven alternative for conventional evaporation technologies. The dairy industry is the food industry's largest user of energy for evaporation. FC offers both energy and economic savings to dairy processors using this innovative freeze concentrate technology. In addition, the food processing industry and the environment benefit by shifting from fossil fuels to more environmentally clean electrical energy.

FC is a process for removing water from food by crystallizing the water and mechanically separating ice crystals from the unfrozen liquid medium. The FC concentration process has been installed in more than 100 locations throughout the world. It has helped improve the quality and reduce the operating costs for concentrating orange juice, grapefruit juice, mandarin juice, coffee extract, vinegar, and ice beer, and to remove toxic organic residues from problematic process waste streams in the chemical and refining industries.

FC stemmed from fundamental research at the University of Eindhoven in 1960, and the first prototypes were developed by Grasso/Grenco in The Netherlands from 1970 to 1975. Early commercial applications were mainly for the production of soluble coffee in 1975 and orange juice in 1980. Pioneering work by the Electric Power Research Institute (EPRI), the Dairy Research Foundation (DRF), and Niro Process Technology successfully applied the freeze concentration process to milk and milk products. Freeze concentration yields milk products with a remarkably improved product quality, no heat degradation, lower operating costs, reduced waste burden, and a sustained continuous operation.

Freeze concentration provides several unique advantages over traditional thermal concentration technologies. FC products have a better product quality with no heat degradation and are clearly superior in taste, color, and functionality compared with other commercial concentrates. In addition, the improved operating efficiencies reduce the processors' bottom line cost. Freezing consumes 80 kcal (335 kJ) per kilogram of water made into ice versus 526 kcal (2200 kJ)

per kilogram of water converted to steam. The bottom-line additional cost of FC concentrate is no more than 6 cents per kilogram of milk solids higher than the cost of the best thermal evaporators of \$2.05 to 2.20 kg^{-1}.

How freeze concentration works The process works by pumping the milk through a scraped surface heat exchanger to form crystals. It is then fed to the recrystallizer, where the small crystals are mixed with a population of larger ice crystals. The larger crystals are grown through a ripening process. A continuous supply of small crystals from the scraped surface heat exchangers provides the fuel for this ripening growth. The ice crystals need to be removed to complete the concentration step. The wash column provides the most efficient method to remove these ice crystals and minimize the loss of product. Figure 5 provides a simplified schematic of the process.

The ripening process takes advantage of the equilibrium temperature at which crystals form. The equilibrium temperature of the solution is when crystals neither grow nor melt. When placed in the same concentration solution, small crystals have a slightly lower equilibrium temperature than larger crystals. The bulk solution temperature of a mixture of small and large crystals will be somewhere between the equilibrium temperature of the small and large crystals. The smaller crystals will be in a warmer environment and will melt, and the larger crystals will be in a colder environment and will grow. The heat of crystallization is exactly balanced by melting and growing crystals. The driving forces are therefore very small, producing spherical crystals, as shown in Figure 6.

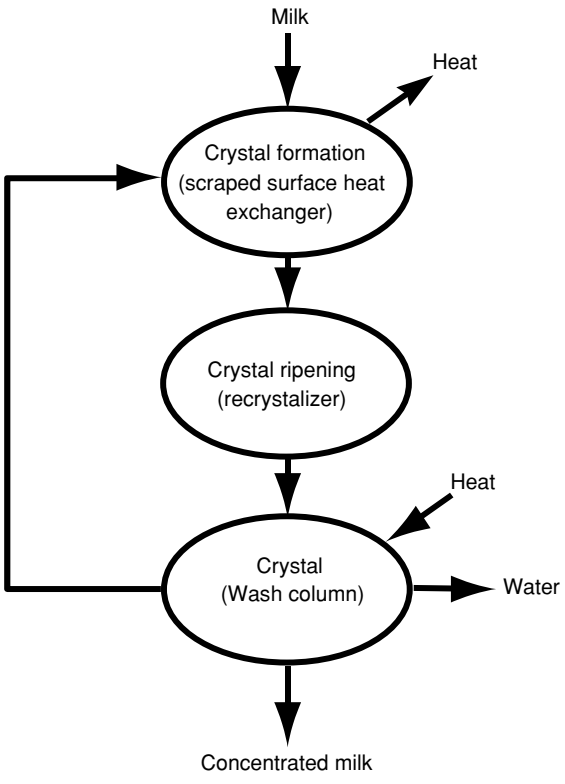

Figure 5 Overview of the freeze concentration process.

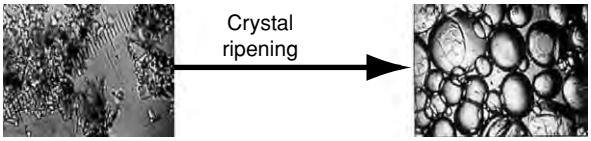

Figure 6 Ripening of crystals.

Once crystals are formed in the solution, the remaining liquid is concentrated in the solute(s) contained in the original feed because a portion of the original water is now in the form of ice crystals. The ice crystals need to be removed to complete the concentration step. The wash column provides the most efficient method to remove these ice crystals and minimize the loss of product.

Other Developing Electrotechnologies

Ultra-high-pressure processing Pressure can kill microorganisms in food products at room temperature with little or no damage to the food product. The most promising results to date have been with acidic liquid foods such as orange juice. Ultra-high-pressure processing, also called high-pressure processing (HPP) can extend the shelf-life of food products without the application of heat or chemical treatment.

In HPP, food products are exposed to pressures in the range of 103 422–620 000 kPa for a few minutes. Spores and common food enzymes appear to tolerate high pressures and are not destroyed.

Pulsed electric fields Pulsed Electric Field (PEF) is a unique nonthermal method of inactivating microorganisms, including many of the common food pathogens, without heating the product to the usual pasteurization temperatures. The destruction or inactivation of the microorganism is achieved by the breakdown of the microorganism's cell membranes during exposure to electric fields.

PEF applies multiple short pulses (1 μs each) of high-intensity electric fields (20–50 $kV\,cm^{-1}$) between two electrodes. The food product being treated flows between the two electrodes and is exposed to the electric field. The degree of treatment is adjusted to the characteristics of the food product being processed by:

- exposure time (related to flow rate and fluid volume of the electrode chamber);

- frequency of the pulses;
- intensity (kV cm^{-1}) of the electric field; and
- shape of the pulse wave (rate of build and decay of the electrical intensity).

Even though this technology has been applied only recently in the USA, it was described in German Patent 1,237,541 issued in 1960. PEF can be an effective pasteurization process, but the FDA has expressed concerns about the consistency of some of the microbial destruction data. These apparent inconsistencies may be due to the effect of the growth phase, type of organism, media, and pH. PEF presently is *not* recommended as a food sterilization method. Promising test results have been obtained with liquid eggs, milk, emulsions, juices, rice wine, some sauces, fermented products, and water.

See also: **Convenience Foods**; **Crystallization**: Basic Principles; **Effluents from Food Processing**: On-Site Processing of Waste; **Filtration of Liquids**; **Freezing**: Operations; **Irradiation of Foods**: Applications; **Membrane Techniques**: Applications of Reverse Osmosis; **Pasteurization**: Pasteurization of Liquid Products; **Radioactivity in Food**

Further Reading

A Century of Food Science (2000) Chicago, IL: Institute of Food Technologists.

Castellari M, Arfelli G, Riponi C and Amati A (2000) High hydrostatic pressure treatments for beer stabilization. *Journal of Food Science* 65(6): 975–977.

Gerber J (1998) The electric farm. *EPRI Journal* 23(3): 8–17.

Graham DM (1997) Use of ozone for food processing. *Food Technology* 51(6): 72–75.

EPRI (1993) *Electroconductive (Ohmic) Heating for Continuous Sterilization of Solid–Liquid Food Mixtures*. Palo Alto, CA: EPRI.

EPRI (1997) *Pulsed Electric Field Technology*. Palo Alto, CA: EPRI.

EPRI (1999) *Resource Guide for Food Irradiation*. Palo Alto, CA: EPRI.

EPRI (2000) *Ozone and UV for Grain Milling Systems*. Palo Alto, CA: EPRI.

EPRI (2000) *Food Industry 2000: Food Processing Opportunities, Challenges, New Technology Applications*. Palo Alto, CA: EPRI.

EPRI (2001) *Freeze Concentration of Dairy Products: Phase 3 and Phase 4*. Palo Alto, CA: EPRI.

Kim JG and Yousef AE (2000) Inactivation kinetics of foodborne spoilage and pathogenic bacteria by ozone. *Journal of Food Science* 65(3): 521–528.

McDonald CJ, Lloyd SW, Vitale MA, Petersson K and Innings F (2000) Effect of pulsed electric fields on microorganisms in orange juice using electric field strengths of 30 and 50 kV/cm. *Journal of Food Science* 65(6): 984–989.

Marcotte M, Trigui M, Tatiboquet J and Ramaswamy HS (2000) An ultrasonic method for assessing the residence time distribution of particlate foods during ohmic heating. *Journal of Food Science* 65(7): 1181–1186.

Marquez VO, Mittal GS and Griffiths MW (1997) *Journal of Food Science* 62(2): 399–401.

Moore T (1994) A separable feast. *EPRI Journal* 19(6): 17–23.

Saldo J, Sendra E and Guamis B (2000) High hydrostatic pressure for accelerating ripening of goat's milk cheese: proteolysis and texture. *Journal of Food Science* 65(4): 636–640.

Vega-Mercado JR, Powers GV, Barbosa-Canovas G and Swanson BG (1995) Plasmin inactivation with pulsed electric fields. *Journal of Food Science* 60(5): 1143–1146.

Prader–Willi Syndrome *See* **Developmental Disabilities and Nutritional Aspects**: Down Syndrome; Prader–Willi Syndrome

Prawns *See* **Shellfish**: Characteristics of Crustacea; Commercially Important Crustacea; Characteristics of Molluscs; Commercially Important Molluscs; Contamination and Spoilage of Molluscs and Crustaceans; Aquaculture of Commercially Important Molluscs and Crustaceans

PREBIOTICS

M B Roberfroid, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

Background

The microbial flora that live symbiotically within the large intestine represent an essential element of the physiology of most mammals. The flora constitute a complex ecosystem that needs to be correctly fed in order to maintain a well-balanced composition in which the health-promoting bacterial species quantitatively predominate over the potentially harmful species. Foods that resist hydrolysis by the digestive enzymes and are not absorbed in the upper part of the gastrointestinal tract, including the small intestine, become substrates for the colonic microflora. These dietary components pass into the large bowel, where most of the indigenous intestinal microflora are located in a healthy human individual. A wide variety of dietary carbohydrates, especially resistant starch, dietary fiber, some polyols and nondigestible oligosaccharides, have such characteristics, and they provide quantitatively the majority of colonic food, i.e., dietary components available for bacterial fermentation in the colon. The colonic fermentation of such 'malabsorbed,' 'non-digestible,' or 'resistant' carbohydrates (oligosaccharides and polysaccharides) plays a role in salvaging part of the energy of these dietary components, in controlling transit time, stool bulking, and stool frequency, in influencing nutrient, especially mineral, bioavailability, in producing short-chain fatty acids that are known to play physiological roles such as control of mucosal motility and epithelial cell proliferation, or in modulating immune activity and endocrine functions. Amongst colonic foods, it has been shown recently that some components are especially beneficial to health. These compounds have been called 'prebiotics,' defined as 'nondigestible food ingredients that beneficially affect the host by selectively stimulating the growth and/or activity of one or a limited number of bacteria in the colon.' Prebiotics are thus colonic foods, but they are more than simply malabsorbed, nondigestible, or resistant carbohydrates; when they reach the large bowel, they have specific effects that promote the growth of advantageous species of bacteria to the detriment of adverse species. The inclusion of prebiotics in the diet can lead ultimately to a marked change in the composition of the colonic microflora, e.g., by selectively stimulating the growth of bacteria that are generally recognized as being beneficial for health and, at the same time, reducing the number of potentially harmful bacteria. The most efficient prebiotics identified today stimulate the growth of bifidobacteria, and sometimes lactobacilli, whilst reducing the numbers and activities of potentially pathogenic organisms. (*See* ***Bifidobacteria*** **in Foods**; **Lactic Acid Bacteria**.)

The contents of the human gut are not readily accessible for microbiological analysis; therefore, demonstration of changes in the composition of the fecal microflora is often used as a surrogate marker for the prebiotic effect. However, for such a demonstration to be convincing, it is critical that as many components of the fecal microbiota as possible are measured. These should include at least bacteroides, bifidobacteria, clostridia, eubacteria, Gram-positive cocci, coliforms, lactobacilli, total aerobes, and total anaerobes. Simple stimulation of growth of bifidobacteria and/or lactobacilli is insufficient to substantiate a prebiotic property without determining the effects on other fecal microorganisms, since it is the selectivity of effect that determines classification as a prebiotic. Clearly, studies using pure bacterial cultures are of very limited, if any, value in this respect, unless they are supported by mixed culture work in a well-validated set-up. But the ultimate proof must come from human studies in which correctly collected and stored fecal samples are analyzed for their composition in terms of bacterial species that are further well characterized using classical techniques combined with a conventional microbiological approach towards identification and/or modern molecular genotyping methods. Indeed, it is the effect of the prebiotic in a competitive ecological environment that is important.

The Prebiotics

At present, the food components for which a prebiotic effect has been reported are nondigestible carbohydrates that consist of mixtures of oligomers of different chain lengths and are characterized by the average number of osyl moieties, referred to as the 'degree of polymerization' (DP). They have been classified as 'nondigestible oligosaccharides,' or NDOs, the osidic bond of which is in a spatial configuration that confers resistance to the hydrolytic activities of the digestive enzymes of the upper gut. But once they reach the large bowel, the NDOs are fermented by at least some of the colonic bacteria. This fermentation produces gases which are excreted, and short-chain fatty acids, which may be absorbed by the host and

thus represent a source of energy. The NDOs affect the growth and, ultimately, the selective proliferation of these bacteria. Currently, there are few data on the relative efficiencies of different prebiotic oligosaccharides, or on their selectivity at a species, or even genus, level; however, the newly developed methodologies based on hybridization with specific rRNA probes will help us to make progress in that direction. It is often the case that a prebiotic effect is claimed for certain NDOs, or other dietary carbohydrates, without a full and careful investigation of their fermentation profile. Such data can never be accepted as a proof for a prebiotic attribute. (*See* **Carbohydrates**: Classification and Properties.)

Among the nondigestible oligosaccharides, those composed primarily of fructose units occupy a leading position in food science. Fructooligosaccharide is used as a generic name for all nondigestible oligosaccharides composed mainly of fructose. Strictly, however, the linear β-(2–1)-fructans are inulin-type fructans, which are different from levans, which are β-(2–6), often branched, fructans. The inulin-type fructans are by far the most extensively studied compounds and are clear market-leader prebiotics. They are composed of β-D-fructofuranose moities attached by β-2–1 linkages. The first monomer of the chain is either an α-D-glucopyranosyl or a β-D-fructopyranosyl residue. They constitute a series of homologous oligosaccharides derived from sucrose and may be represented by the formula GF_n or FF_n (where G = glucose and F = fructose). The naturally occurring NDOs, which are extracted from the roots of the chicory plant (*Cichorium intybus*), are a mixture of either GF_n (α-D-glucopyranosyl-[β-D-furanosyl]$_{n-1}$-D-fructofuranoside) or GF_n + FF_n (β-D-fructopyranosyl-[β-D-fructofuranosyl]$_{n-1}$-D-fructofuranoside) molecules, with the number of fructose units varying from 2 to ~60–65 units. As food ingredients, they are available as native inulin (inulin ST, average DP 10) and high-molecular-mass inulin (inulin HP, average DP 20), enzymatically hydrolyzed inulin (oligofructose, average DP 4), and a mixture of inulin HP and oligofructose (synergy 1), all of which occur naturally in a variety of food plants such as garlic, onion, asparagus, artichoke, banana, and wheat. Oligofructose can be produced by enzymatic conversion of sucrose to give a mixture composed of only GF_2, GF_3, GF_4 (average DP 3.8), sucrose, glucose, and fructose. The galactooligosaccharides, or transgalactooligosaccharides, are a potentially important class of prebiotics; they are produced industrially from lactose by transglycosylation reactions and consist of galactosyl derivatives of lactose with β1–3 and β1–6 linkages. The purported bifidogenic (i.e., stimulation of growth of bifidobacteria) nature of fructooligosaccharides and galactooligosaccharides is explained, at least in part, by the linkage specificity of the bifidobacterium β-fructosidase and β-galactosidase enzymes, respectively. These enzymes are cell-bound.

The glucose-based maltooligosaccharides and xylooligosaccharides are candidate prebiotics; however, specific bacterial enzymes for the degradation of these molecules have not been identified.

Malabsorption of the Nondigestible Oligosaccharides

The β configuration of the anomeric C-2 in the fructose moieties of the inulin-type fructans as well as in the galactose monomer subunits of galactooligosaccharides explains why they are resistant to hydrolysis by human digestive enzymes (α-glucosidase, maltase-isomaltase, and sucrase), which are mostly specific for α-osidic linkages. In normal physiological conditions, the inulin-type fructans and the galactooligosaccharides are resistant to acid hydrolysis in the stomach. The most convincing data have been obtained from human intervention studies with ileostomy volunteers. These studies show that 86–88% of the ingested dose (10–30 g) of inulin or oligofructose is recovered in the ileostomy effluent, supporting the claim that these carbohydrates are practically nondigestible in the small intestine of man. Using an intubation technique in human volunteers, it has been concluded that fructooligosaccharides are malabsorbed in the human small intestine (89% recovery). The small, but still significant, loss of fructooligosaccharides in the upper part of the gastrointestinal tract could be due to fermentation by the microbial population colonizing the ileum, especially in ileostomy patients, and/or to hydrolysis of the lowest-molecular-mass oligomers such as the enzymatically synthesized oligomers. In a recent review of the malabsorption characteristics of inulin-type fructans, the authors concluded that 'inulin and oligofructose pass through the small bowel without degradation and (furthermore) without influencing the absorption of nutrients and minerals especially calcium, magnesium and iron.'

There are fewer data on the resistance of other oligosaccharides to digestion in the human upper gastrointestinal tract than for the inulin-type fructans. The available evidence comes predominantly from *in vitro* experiments or is based on hydrogen production and exhalation as an indirect marker of colonic fermentation, or on stimulation of growth of specific fecal microorganisms in animal models. No *in vivo* human data are available. Thus, the nondigestibility of isomaltooligosaccharides, soybean

oligosaccharides, galactooligosaccharides, palatinose condensates, or xylooligosacchairdes *in vivo* remains to be demonstrated.

Fermentation in the Large Bowel: the Prebiotic Effect

The large bowel is by far the most heavily colonized segment of the human gastrointestinal tract, with up to 10^{12} (mostly anaerobic) bacteria per gram of gut content. These bacteria belong to a wide variety of genera, species, and strains. Through the process of fermentation, these colonic bacteria produce a wide variety of metabolites, among which the short-chain fatty acids represent salvage of part of the energy of malabsorbed food components, especially malabsorbed carbohydrates, and they play important systemic physiological roles.

Evidence for the fermentation of inulin-type fructans by bacteria colonizing the large bowel has come from *in vitro* and *in vivo* studies. At nutritional doses (up to 20–40 g per day), these malabsorbed carbohydrates are fermented quantitatively and are not excreted in the feces; the products of fermentation are gases and short-chain fatty acids, mainly acetate, butyrate, and propionate. Compared with most other malabsorbed carbohydrates (e.g., resistant starch and dietary fiber), the colonic fermentation of inulin-type fructans is accompanied by a significant change in the composition of the colonic microbiota due to selective proliferation of bifidobacteria and a concomitant reduction in the number of other bacteria, like bacteroides, fusobacteria, and pathogenic clostridia. On the basis of the results of well-designed human studies that have shown significant changes in the composition of human fecal flora, it can be concluded that inulin-type fructans (5–15 g per day for a few weeks) are prebiotic. But even though some studies showed a significant reduction in the number of pathogenic clostridia, the health benefits (e.g., reducing the risk of intestinal infections) of such a change in the composition of the colonic microbiota have yet to be established. A recent report has shown that oligofructose (daily dose of 6 g (3×2 g)) had no therapeutic value in patients with irritable bowel syndrome. But in an experimental model of necrotizing enterocolitis in quails, data have been reported that support the hypothesis that oligofructose might prevent overgrowth of the bacteria known to play a role in this pathology in preterm neonates.

For the other NDOs, studies *in vivo* have been performed with doses ranging from 3 to 15 g per day, given for periods of 1, 2, or 3 weeks. For soybean oligosaccharides, a dose of 10 g given twice daily for 3 weeks significantly increased the number of bifidobacteria, whilst slightly decreasing clostridia counts. A dose of 3 g per day increased bifidobacteria, bacteroides, and eubacteria. For the galactooligosaccharides, an increase in bifidobacteria and lactobacilli in response to doses ranging from 3 to 10 g per day has been reported. Similarly, a daily dose (5 or 10 g) of galactosylsucrose stimulated the growth of bifidobacteria after 1 and 2 weeks of ingestion. A dose of isomaltooligosaccharide of 13.5 g per day for 2 weeks significantly increased the number of bifidobacteria in adult and elderly volunteers. Early results suggest that platinose condensate may stimulate the growth of bifidobacteria. For all these NDOs, except the galactooligosaccharides, only a single human intervention study has been performed, and this will need to be repeated before any prebiotic effect can be substantiated. Moreover, as discussed above, great care should be taken to quantify the component species of the fecal microflora and to identify changes in its composition.

Physiological Effects in the Gastrointestinal Tract

The fermentation of prebiotics in the colon has a series of consequences that affect large-bowel physiology. Firstly, it contributes to the production of short-chain fatty acids, thus creating a more acidic environment, which is favorable for the development of bacteria like bifidobacteria or lactobacilli but unfavorable for the growth of potentially pathogenic species like *Clostridium* spp. or *Escherichia coli*. Furthermore, in this acidic environment, ammonia and amines become protonated and are thus much less absorbable and hence more readily excreted. Secondly, it leads to a proliferation of colonic bacteria, which increases fecal mass and thus contributes to a beneficial bulking effect and a regularization of stool production. As a consequence of these two large-bowel processes (i.e., production of acids and proliferation of bacteria), only part of the metabolizable energy of the NDOs is salvaged. As compared with absorbed carbohydrates (e.g., nonresistant starch or sucrose), NDOs represent, weight for weight, a lower energy value for the host. Caloric values of 1–2.1 kcal g^{-1} (4.1–8.8 kJ g^{-1}), i.e., 25–50% of the caloric value of sucrose, have been reported, especially for inulin-type fructans. But, as stated recently by a group of European experts, 'all carbohydrates that are more or less completely fermented in the human colon should be given a caloric value of 1.5 kcal g^{-1} (6.3 kJ g^{-1}).' In fact, the daily intake of these carbohydrates is likely to remain small, probably often not representing more than $\sim 5\%$ of total daily energy intake. Therefore, it is not justifiable to

spend a great deal of effort in trying to give, for such carbohydrates, a precise calorie value, the determination of which will depend on the protocol used and probably also on the diet in which they are included.

Another physiological consequence of the consumption of NDOs that has been reported recently for the inulin-type fructans is an increased bioavailability of Ca^{2+}. Such an effect has been studied extensively in rat and hamster. These studies have led to the conclusion that, most probably because of their malabsorption and colonic fermentation, these food ingredients facilitate Ca^{2+} absorption from the large bowel compartment, thus complementing the process that takes place in the small bowel. A change in colonic pH, production of short-chain fatty acids, and increase in mucosal concentration of the protein calbindin in the colon have been proposed as hypotheses to explain that effect. Besides increased Ca^{2+} bioavailability, it has been shown both in rat and in hamster that feeding fructooligosaccharides increases Ca^{2+} concentration and improves the structure and density of the bones. More recently, three human trials (two in adolescents and one in adults) have shown that supplementing the diet every day with 16.8 g of oligofructose, 8 g of a mixture of oligofructose and high-molecular-mass inulin, or 40 g of inulin significantly increases the apparent absorption of Ca^{2+} by 10–12%. One of these studies used measurement of Ca^{2+} balance, and two used a double stable isotope technique.

NDOs, because they are malabsorbed and fermented in the colon, may be considered to be part of the dietary fiber complex. In particular, it has been shown that inulin-type fructans have a fecal bulking effect that is comparable on a weight-for-weight basis with that of soluble fiber such as pectin. Moreover, an internationally validated method derived from the AOAC method for dietary fiber analysis has been developed to quantify inulin and oligofructose in plants and food products. For the purpose of food labeling, the NDOs are thus classified as dietary fiber in most countries. (*See* **Dietary Fiber**: Properties and Sources; Determination; Physiological Effects.)

Prebiotics and the Risk of Colon Cancer

Over the last two years, experimental reports have been published that repeatedly demonstrate that feeding inulin-type fructans to rats previously treated with a colon carcinogen (i.e., dimethylhydrazine or azoxymethane) reduces the incidence of the so-called aberrant crypt foci in the colon. In one of these studies, the synbiotic approach that combines oligofructose (prebiotic) and bifidobacteria (probiotic) was reported to be more active than either the prebiotic or the probiotic alone. Even though still experimental, these data suggest that inulin-type fructans might play a role in reducing the risk of developing preneoplastic lesions and possibly cancer in the colon. Moreover, such an effect might not be limited to the inulin-type fructans. Indeed nondigestible oligosaccharides may exert anti-carcinogenic effects first because they have been shown to beneficially affect certain biomarkers known to be associated with cancer risk (e.g. increase in apoptotis in colonic mucosa, reduction of bacterial β-glucuronidase and nitrate reductase, pH, and conversion of a dietary carcinogen to its genotoxic metabolite in caecal contents or feces of NDOs fed rats and human volunteers respectively), and second because they stimulate the growth of lactic acid bacteria for which evidence of anti-genotoxic and anti-carcinogenic effects have been reported. (*See* **Colon**: Cancer of the Colon.)

Conclusion: Prebiotics, what Benefit(s) for Human Health?

Prebiotics have nutritional properties that, according to the present state of knowledge, derive mainly from resistance to the hydrolytic activities in the upper part of the gastrointestinal tract of monogastric organisms followed by extensive fermentation in the large bowel, which leads to significant and possibly health-beneficial changes in the composition of the colonic microbiota. The gastrointestinal target functions that are associated with a balanced microflora together with an optimal gut-associated lymphoid tissue (GALT) are relevant to the state of well-being and health and to the reduction of the risk of diseases. The colonic microflora is a complex ecosystem, the functions of which are a consequence of the combined action of the microbes that, besides interacting with the GALT, contribute to salvage of nutrient energy and yield metabolic end products like the short-chain fatty acids (SCFAs) that play a role in cell differentiation, cell proliferation, and metabolic regulatory processes. It is generally assumed that the group of potentially health-promoting bacteria includes principally bifidobacteria, lactobacilli, eubacteria, and bacteroides, which are, and possibly should remain, the most important genera in humans. Changes in the composition of the fecal flora, a recognized surrogate marker for the residual colonic microbiota, can be considered as a marker, both indicator and factor, of large-bowel functions. They might play a role in gastrointestinal infections and diarrhea, constipation, irritable bowel syndrome, inflammatory bowel diseases, and colorectal cancer. Probiotics (e.g., lactobacilli or bifidobacteria), prebiotics (like chicory inulin and its hydrolysate oligofructose), and synbiotics (a

combination of probiotics and prebiotics) are recent concepts in nutrition that are being used to support the development of functional foods targeted towards gut functions. Their effects may include:

- stimulation of the activity of the GALT (e.g., increased IgA response, production of cytokines, etc.);
- reduction of the duration of episodes of rotavirus infection;
- change in the composition of the fecal flora to reach/maintain a composition in which bifidobacteria and/or lactobacilli become predominant in number, a situation that is considered optimal;
- increase in fecal mass (stool bulking) and stool frequency;
- increase in calcium bioavailability via colonic absorption (e.g., inulin).

By reference to the recently published European consensus on scientific concepts of functional foods, prebiotics, especially inulin-type fructans (but also synbiotics) are good candidates to be recognized as functional food ingredients for which claims shall become authorized. Such claims should relate to enhanced gastrointestinal functions (composition of colonic flora, bulking effect and bowel habit, Ca^{2+} bioavailability) or risk of developing a disease like colon cancer.

See also: **Carbohydrates**: Digestion, Absorption, and Metabolism; **Colon**: Cancer of the Colon

Further Reading

Gibson GR and Roberfroid MB (1995) Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. *Journal of Nutrition* 125: 1401–1412.

Gibson GR and Roberfroid MB (eds) (1999) *Colonic Microbiota, Nutrition and Health*. Dordrecht: Kluwer Academic.

Roberfroid MB (1999) Concepts in functional foods: the case of inulin and oligofructose. *Journal of Nutrition* 129 (supplement): 1398S–1401S.

Roberfroid MB and Delzenne N (1999) Dietary fructans. *Annual Review of Nutrition* 18: 117–143.

Van Loo J, Cummings J, Delzenne N *et al*. (1998) Functional food properties of non-digestible oligosaccharides. A consensus report from the ENDO project. *British Journal of Nutrition* 81: 121–132.

PREGNANCY

Contents

Metabolic Adaptations and Nutritional Requirements

D J Naismith, University of London, London, UK

Predicted Energy and Nutrient Costs

Women during pregnancy are generally believed to be nutritionally 'at risk.' There are obvious energy and nutrient costs arising not only from the growth of the fetus and placenta, but also from enlargement of the maternal reproductive tissues and the deposition of a substantial energy reserve in the form of fat. These have been determined by direct chemical analyses and by indirect measurements (**Table 1**) and have been used as the basis for predicting additional nutritional needs for a successful outcome in pregnancy.

Importance of Body Fat

The relationship between nutrition and reproduction, however, begins at an earlier stage. In late childhood, body fat accounts for about 12% of body weight in both boys and girls. With the onset of puberty, the proportion of body fat rises in girls to reach around

Table 1 Calculated energy and nutrient costs of pregnancy

Energy/nutrient	*Cost*
Energy (fat reserve, new tissue synthesis increase in basal metabolic rate)	335 MJ (80 000 kcal)
Protein (maternal reproductive tissues, conceptus)	910 g
Calcium (fetal skeleton)	28 g
Iron	300–400 mg

17% at menarche, and 24% in the physiologically mature woman some 12 months later. The attainment of this additional fat deposit, amounting on average to 226 MJ (54 000 kcal) of stored energy, is believed to be the major determinant of fertility in normal healthy women. Young women who, for professional reasons, must maintain a low proportion of body fat, such as track athletes and ballet dancers, are frequently infertile, and the cessation of menstruation is an early symptom of the wasting disease anorexia nervosa. Should conception occur, total body fat is calculated to be sufficient to meet the energy costs of pregnancy (excluding further deposition of fat during gestation) and to sustain a 3-month period of lactation. (*See* **Anorexia Nervosa**; **Fats**: Requirements; **Lactation**: Human Milk: Composition and Nutritional Value.)

Effects of Undernutrition

The validity of this argument was revealed in a study of the effects of acute severe malnutrition that occurred in the winter of 1944 and in the following spring in cities of the western regions of The Netherlands. Immediately before the famine, the nutritional status of the population had been generally satisfactory. Thereafter, however, the situation deteriorated rapidly, and food from all sources provided no more than 2.1–2.5 MJ (500–600 kcal) per day. The incidence of amenorrhea rose dramatically in the young female population. At the height of the famine, the number of infants conceived fell to 50% of the previous conception rate. Nevertheless, in the most vulnerable women, exposed to famine in the third trimester, the average birthweight declined to a maximum of about 300 g below the prefamine level. Since the body of a healthy newborn infant contains more than 500 g of fat, most if not all of the deficit in birthweight may have been attributable to an inability to accumulate fat, priority being given to lean tissue growth.

Acute severe malnutrition, however, occurs rarely among normally well-nourished populations, and might give a misleading picture of the interaction of undernutrition and reproductive performance. In most developing countries, chronic moderate malnutrition is commonplace, and nutritional surveys invariably report mean birthweights significantly below the average for Europe and North America. Although this has commonly been thought to reflect the poor nutritional status of the mothers, it has been impossible to isolate the putative influence of maternal diet on fetal growth from that of other environmental factors associated with endemic malnutrition, perhaps the most important being the smaller stature of the mothers. Even within an affluent society, an association between birthweight and maternal height and weight-for-height is found. (*See* **Malnutrition**: Malnutrition in Developed Countries.)

Nutrition Intervention and Pregnancy Outcome

A number of well-controlled nutrition intervention studies were carried out in developing countries in the early 1970s, no doubt with the aim of demonstrating the need for, and benefit to be derived from, diet supplementation during pregnancy. The subjects lived in countries in which protein–energy malnutrition was endemic. Energy intakes ranged from 5.86 to 7.53 MJ (1400–1800 kcal) per day. Given the small stature of the mothers, such a marginal plane of nutrition would just satisfy the needs of a nonpregnant woman. The dietary supplements provided energy alone or energy with protein, and were administered on a scale that often exceeded the estimated additional costs for energy and protein during pregnancy. The findings were as surprising as they were informative.

Only in one study, conducted in rural Guatemala, was a significant increase in mean birthweight achieved, amounting to 117 g. Although most women showed no improvement in reproductive performance, the response was very variable, those with the lowest customary energy intakes experiencing the greatest benefit, and the proportion of low-birthweight infants (< 2500 g) was consequently reduced.

It was evident from these studies that there exists an 'energy threshold,' around 6.7–7.1 MJ (1600–1700 kcal) per day, above which diet supplementation has no effect, but below which fetal growth may be stimulated, the improvement being directly related to the degree of energy deprivation. It was also revealed that the major determinant of fetal growth was the maternal energy supply; the inclusion of protein in the supplements had no effect independent of the energy it provided.

Dietary Habits in Well-nourished Populations

The results of the various intervention studies clearly indicate that if a woman is adequately nourished,

then pregnancy entails no appreciable additional energy (food) cost. It could be argued that the fetus is remarkably protected from the influence of maternal malnutrition, and that women exposed to generations of deprivation have simply adapted to a limited food supply. Small stature is often cited (incorrectly) as an example of adaptation. Studies of the dietary habits of women living in affluent circumstances, however, have shown that this is not so.

As long ago as 1953, analyses of the diets of 2300 well-nourished women in Nashville, USA failed to show any association between maternal food intake and birthweight. Furthermore, on average, no increase in food consumption during pregnancy was apparent (Figure 1). During the second trimester, an increase of around 251 kJ (60 kcal) per day was noted, but during the last trimester, when a marked acceleration in fetal growth occurs, energy intake fell by 753 kJ (180 kcal) per day. This surprising discovery, ignored for 30 years, has been amply confirmed by studies in the 1980s in Paris, Cambridge, and Glasgow.

How, then, is the obvious requirement for increased nutrition during pregnancy (Table 1) to be reconciled with the mother's failure, even under the most favourable environmental circumstances, to increase food intake? Only a small fraction of the fat stored in the first two trimesters of pregnancy is mobilized before term, and there exists no *ad-hoc* store of minerals and water-soluble vitamins. The answer to this question is to be found in the complex mechanisms whereby nutrients are transported from the maternal circulation to the fetus, and in the adjustments that occur in the maternal physiology in response to the gravid state.

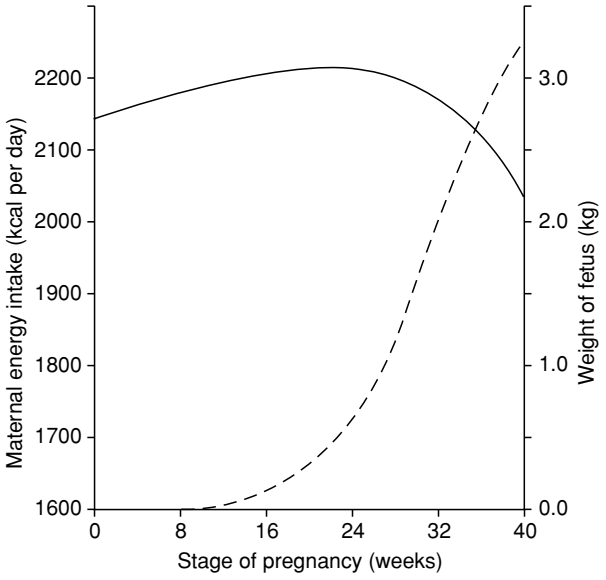

Figure 1 Diagram showing changes in maternal energy intake (—) and weight of the fetus (– – –) at different stages of pregnancy. Reproduced from Pregnancy: Metabolic Adaptations and Nutritional Requirements, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Role of the Placenta

The role of the placenta is discussed in **Pregnancy**: Role of Placenta in Nutrient Transfer. It is primarily an organ for nutrition of the fetus, but functions also as an endocrine gland throughout the greater part of gestation, regulating nutrient utilization by the mother.

Glucose provides the energy for fetal growth, the fetus being unable to oxidize fatty acids. Transfer of glucose across the placenta is by 'facilitated diffusion,' and is therefore influenced by the maternal blood glucose concentration. A prolonged low blood glucose concentration, as may occur in mothers in the developing world engaged in hard physical work, or an abnormally high concentration associated with gestational diabetes can therefore lead to growth retardation, or to excessive growth and fat deposition (macrosomia). From early pregnancy, the fasting blood glucose concentration is reduced, a change that may favor transfer to the fetus since the placenta is particularly effective at extracting nutrients from low concentrations in the blood. Glucose tolerance following a glucose load is also progressively relaxed so that the glucose concentration remains elevated for a longer period, again facilitating placental uptake. Whether or not the metabolic response to carbohydrate consumed as part of a meal is similarly affected is unclear. (*See* **Glucose**: Maintenance of Blood Glucose Level; Glucose Tolerance and the Glycemic (Glycaemic) Index.)

Fat-soluble vitamins also cross the placenta by diffusion, maternal blood levels being maintained by dietary input and mobilization of tissue stores. In contrast, amino acids, minerals, and the water-soluble vitamins are actively transported, the concentrations in the fetal circulation being much higher than in the maternal circulation. Furthermore, there is evidence that some vitamins are chemically altered by the placenta to a form that is unable to diffuse back into the maternal bloodstream, so reinforcing the one-way transport mechanism. Thus, the fetus has prior claim on the nutrient supply, much of which is maintained, by homeostatic regulation, within narrow concentration limits in the maternal plasma. In adverse nutritional circumstances, it is the mother, not the fetus, who is 'at risk.' (*See* **Amino**

Acids: Metabolism; **Minerals – Dietary Importance**; **Vitamins**: Overview.)

Metabolic Adjustments in Pregnancy

Under normal circumstances, the integrity of the tissues is assured by homeostatic regulation of the absorption, excretion and catabolism of the nutrients. During reproduction, these control mechanisms are adjusted, under the influence of hormones secreted by the placenta, in order to make available an additional nutrient supply for growth of the conceptus.

Absorption of the Nutrients

The macronutrients, fat, carbohydrate, and protein, are almost completely digested and absorbed in the small intestine, so that no appreciable change occurs in pregnancy. The absorption of minerals, however, is carefully regulated. Most Western diets provide a considerable excess of calcium, and no more than around 30% is normally absorbed. From early in pregnancy, the blood concentration of the physiologically active form of vitamin D (1,25-dihydroxycholecalciferol) is elevated, leading to an increased uptake of calcium from the gut, thus allowing the mother to satisfy the needs of her developing fetus without recourse to her own skeletal calcium or the need to increase her dietary intake. Only when her customary calcium intake or the biosynthesis of vitamin D is severely restricted is calcification of the fetal skeleton compromised. (*See* **Calcium**: Physiology; **Cholecalciferol**: Physiology.)

As with calcium, the proportional absorption of iron from a diet providing 12–14 mg per day is very small (about 10%), but as gestation advances, iron uptake increases progressively and may exceed 40%. This adjustment in absorption, in association with a reduction in iron loss resulting from the cessation of menstruation, is sufficient to satisfy the iron cost of pregnancy. Judged by normal standards, the characteristic fall in hemoglobin concentration that is noted in pregnancy would indicate a state of anemia. Blood volume, however, is considerably increased so that the total amount of hemoglobin in circulation, and consequently the oxygen-carrying capacity of the blood, is actually raised. Supplementation of the diet of women who begin pregnancy in an iron-sufficient state may therefore be injurious as it has been shown to increase red cell size and could adversely affect blood flow in the small capillaries. (*See* **Iron**: Physiology.)

Other minerals, such as copper and zinc, may also show improved absorption. (*See* **Copper**: Physiology; **Zinc**: Physiology.)

Urinary Excretion of Nutrients

The glomerular filtration rate is increased during pregnancy. It is suggested that the presence of traces of glucose in the urine at some stages of pregnancy in most women merely indicates that more glucose has been filtered from the plasma than can be reabsorbed by the kidney tubules at that time. Some women, however, may lose large amounts of glucose in their urine, particularly those who develop gestational diabetes.

The urinary excretion of most nutrients is altered, but whether or not this reflects specific modifications in kidney function is unclear since the picture is obscured by withdrawal of nutrients from circulation by the placenta, which changes with time. The pattern of excretion of amino acids is most consistent. In nonpregnant women, about 1% of total nitrogen excretion is accounted for by free amino acids, but urinary loss rises in early pregnancy, reaching a plateau in midpregnancy at a level approximately double that found in the nongravid state. Most of the increase is accounted for by the nonessential amino acids, possibly reflecting selective uptake by the placenta. Since the plasma amino acid concentrations show little change, the large increase in urinary excretion cannot be attributed to kidney overload and is most likely the result of altered hormonal status affecting kidney function. The return to normal levels of excretion during lactation supports this conclusion.

One example of altered kidney function that favors conservation has clearly been identified. The amino acid taurine, a derivative of the nonessential amino acid cystine, shows a dramatic and sustained reduction in excretion from early pregnancy. It has been suggested that taurine may act as a membrane stabilizer, as an inhibitory neurotransmitter or neuromodulator, and as a growth modulator in fetal tissues. Tissue concentrations are particularly high in the fetus, and the suppression of urinary excretion of taurine during pregnancy is seen as a means of satisfying the needs of the developing fetus, which lacks the ability to synthesize taurine *de novo*.

Metabolism of Protein

The rate of accretion of nitrogen by the fetus and maternal reproductive tissues is known to rise 10-fold between the first and last quarters of gestation. Since, on average, no appreciable change in food intake, and thus protein intake, occurs in pregnancy, one would expect to find a positive nitrogen balance, rising continuously throughout pregnancy. In the few nitrogen-balance studies carried out on pregnant women, however, no differences have been found between the values for nitrogen retention measured

at different stages of gestation. This apparent paradox was resolved by the study of reproduction in the rat. (*See* **Protein**: Digestion and Absorption of Protein and Nitrogen Balance.)

During the first 2 weeks of pregnancy, when competition between the dam and her fetuses for nutrients is minimal, a substantial reserve of protein is built up in the muscles (the 'anabolic phase'). During the final week, the period of rapid fetal growth, the protein reserve is broken down, the amino acids released being taken up by the placenta for growth of the conceptus. This 'catabolic phase' is not influenced by the protein content of the maternal diet, and is controlled by the hormone progesterone. (*See* **Hormones**: Steroid Hormones.)

Evidence for a similar biphasic system of protein metabolism in women was provided by measurement of the excretion of the amino acid 3-methylhistidine. This amino acid is found predominantly in muscle. Histidine undergoes methylation after its incorporation into the peptide chains of the contractile proteins actin and myosin, and, in the course of muscle protein turnover, the 3-methylhistidine released is not reutilized, but is quantitatively excreted in the urine. In pregnancy, the excretion of 3-methylhistidine rises sharply during the last trimester, indicating the hydrolysis of an amount of protein approximating the estimated protein cost of pregnancy.

Such a redistribution of amino acids from maternal to fetal tissues would not be detected in measurements of nitrogen balance.

This cyclic course of protein metabolism has important nutritional implications. The protein cost of pregnancy is distributed over the whole gestational period, and the influence of acute or chronic maternal undernutrition on fetal growth is minimized.

Nitrogen retention occurs when the intake of protein-nitrogen exceeds nitrogen losses from the body, largely the products of amino acid catabolism in the urine (mostly urea) and undigested dietary protein in the feces along with cells shed from the gut epithelium. If, during pregnancy, the catabolism of amino acids were in part suppressed, the fraction spared could be used for fetal protein synthesis. In the case of a woman existing on 40 g of protein per day, as in many developing world populations, a reduction of less than 9% in amino acid oxidation would spare enough protein to cover the entire estimated requirement for pregnancy.

The hypothesis was tested in the rat. The activities of two rate-limiting enzymes (alanine aminotransferase and argininosuccinatesynthetase), which regulate the oxidation of amino acids and the conversion of amino-nitrogen to urea, were measured in the livers at different stages of gestation. The activities of both enzymes were markedly depressed by the end of the first week, and declined even further (by around 50%) as pregnancy advanced. A parallel change in plasma urea was also noted.

Evidence for a similar adjustment in amino acid catabolism in pregnant women was later obtained in a metabolic study using urea labeled with a stable isotope of nitrogen. Measurements made in the last trimester and in the postpartum postlactational period showed a reduction of 30% in the rate of urea synthesis in late pregnancy and a similar fall in the plasma urea concentration.

The mechanism for protein sparing, which operates progressively throughout pregnancy, combined with the biphasic system of early storage and later breakdown of stored protein, ensures a supply of amino acids commensurate with the demands of the growing fetus. The suppression of hepatic amino acid oxidation was also shown to be induced by the placental hormone progesterone. The fetoplacental unit thus indirectly controls its supply of amino acids as well as its energy needs.

Metabolism of Energy

As stated earlier, the dietary energy supply is the major determinant of fetal growth. In healthy pregnant women, energy balance becomes positive during the first trimester, probably in response to the rising secretion of progesterone. The purpose of the augmented fat reserve, amounting, on average, to some 4 kg of fat, is primarily to subsidize the high energy cost of lactation, but a small amount is mobilized in late pregnancy to provide an alternative fuel for use by the maternal tissues and enhance the availability of glucose for use by the fetoplacental unit. The human fetus derives its energy almost exclusively from the oxidation of glucose. (*See* **Energy**: Energy Expenditure and Energy Balance.)

The discrepancy between the measured energy intakes of healthy pregnant women and the values predicted for energy expenditure has yet to be explained. There is no doubt that some saving is made in energy expenditure from a reduced level of physical activity, but this is difficult to quantify.

The alterations in carbohydrate metabolism that are characteristic of pregnancy could also lead to the sparing of energy. In pregnancy, there is an increased output of insulin in response to a glucose stimulus, and reduced glucose uptake by the peripheral tissues (muscle and adipose tissue). Insulin antagonism has been attributed to the action of placental lactogen. Consequently, more of the ingested carbohydrate is directed to the liver, the organ that maintains the blood glucose concentration. The conversion of

carbohydrate to fat is a very costly process. Approximately 20% of the energy that is available from the direct oxidation of glucose is lost as heat if the glucose is first converted to fat, and the fat is later oxidized to produce energy. This adjustment in carbohydrate utilization, therefore, not only conserves energy for anabolic purposes, but also safeguards the fetal energy supply. (*See* **Carbohydrates**: Digestion, Absorption, and Metabolism.)

Dietary Recommendations for Pregnancy

Growth of the fetus is little affected by transient or prolonged moderate undernutrition of the mother. Relatively small deficits in birthweight may be accounted for by a lower proportion of body fat resulting from maternal dietary energy restriction in late pregnancy, from the diversion of glucose (the fetal fuel) to the muscles should the mother be engaged in hard physical work throughout pregnancy, or from malfunction of the placenta.

The security of the fetus is provided by the placenta, an organ designed not only to ensure a priority claim on all nutrients present in the maternal circulation, but also, by the secretion of hormones, to modulate the homeostatic regulation of nutrient utilization at all levels – absorption, excretion, and catabolism, in order to augment the nutrient supplies. No government committee responsible for devising dietary guidelines would be so incautious as to suggest that women during pregnancy require no more food than when in the nongravid state, although all evidence points very clearly to that conclusion. One obvious *a priori* condition would be that nutritional status should be satisfactory before conception and throughout pregnancy.

Attention has been focused on the latter half of pregnancy, when fetal growth is most rapid, the need for nutritional input is at its greatest, and the effects of maternal food deprivation most apparent. These considerations led a Joint FAO/WHO (Food and Agriculture Organization, World Health Organization) Ad Hoc Expert Committee in 1973 to propose an additional 628 kJ (150 kcal) per day for the first trimester, rising to 1464 kJ (350 kcal) per day for the second and third trimesters, the stage of pregnancy when women, unencumbered by professional advice, would spontaneously reduce their food consumption. In the light of continuing research, however, estimates of energy and nutrient requirements are being revised in a downward direction. In 1974, the US report on Recommended Daily Allowances proposed an additional daily supplement of 30 g of protein, 400 mg of calcium and 30–60 mg of iron for the pregnant woman, acknowledging that such a high intake of iron could not be met by the iron content of habitual US diets. The implication was that pregnancy was a clinical condition that required therapeutic intervention. One decade later, the proposed increments in energy and calcium were little changed, but the supplement of iron was reduced to 15 mg, and the protein allowance was changed to anticipate the pattern of accretion by the fetus, rising from 1.2 g per day in the first trimester to 10.7 g in the final trimester of pregnancy. (*See* **Dietary Requirements of Adults**.)

The UK dietary recommendations for pregnancy have consistently been on a less generous scale. Over a similar period, the daily allowance for energy has fallen from 1004 kJ (240 kcal) in the second and third trimesters to 837 kJ (200 kcal) during the third trimester only. Protein is unchanged at 6 g per day throughout pregnancy, but the recommendation for calcium has fallen from 700 mg per day in the third trimester to zero. Likewise, no recommendation was made for an increase in iron intake in the 1991 report on Dietary Reference Values, compared with the small increase of 1 mg per day in the earlier 1979 edition.

There is no doubt that as scientific opinion changes, other values will also be reduced, and nutritional guidelines for pregnancy ultimately will correspond to the dietary practices of healthy women satisfying their natural appetites on a well-balanced diet.

See also: **Anorexia Nervosa**; **Calcium**: Physiology; **Carbohydrates**: Digestion, Absorption, and Metabolism; **Dietary Requirements of Adults**; **Energy**: Energy Expenditure and Energy Balance; **Fats**: Requirements; **Hormones**: Steroid Hormones; **Iron**: Physiology; **Lactation**: Human Milk: Composition and Nutritional Value; **Malnutrition**: Malnutrition in Developed Countries; **Minerals – Dietary Importance**; **Premenstrual Syndrome: Nutritional Aspects**; **Protein**: Digestion and Absorption of Protein and Nitrogen Balance; **Vitamins**: Overview

Further Reading

Campbell DM and Gilmer DG (eds) (1983) *Nutrition in Pregnancy*. London: Royal College of Obstetricians and Gynaecologists.

Department of Health (1991) *Dietary Reference Values for Food Energy and Nutrients for the United Kingdom*. London: The Stationery Office.

Hytten F and Chamberlain GVP (eds) (1980) *Clinical Physiology in Obstetrics*. Oxford: Blackwell Scientific.

Naismith DJ (1983) Maternal nutrition and fetal health. In: Chiswick ML (ed.) *Recent Advances in Perinatal Medicine*, vol. I, pp. 21–39. Edinburgh: Churchill Livingstone.

National Research Council (1989) *Recommended Daily Allowances* 10th edn. Washington, DC: National Academy Press.

Role of Placenta in Nutrient Transfer

Y Kudo[†] **and C A R Boyd**, University of Oxford, Oxford, UK

Introduction

During intrauterine life, the developing fetus effectively receives all of its nutrition across the placenta, which, in the human, has a chorionic villus structure in which maternal blood directly superfuses the external surface, the trophoblast of the placenta. This trophoblast forms an unusual epithelium separating the maternal plasma from the fetal extracellular space of the villus core through which the fetal capillary circulation flows. For the fetus, the trophoblast is thus equivalent to the epithelium lining the small intestine of the newborn, since, from a nutritional perspective, it must absorb those molecules required for both growth and maintenance of the organism. The trophoblast also has an important role in acting as the lung for the fetus (since all gas exchange between mother and baby must occur across this surface), as the fetus's kidney (since excretion from the conceptus to the mother occurs across this structure), as well as being an important endocrine tissue secreting peptide and steroid hormones into the mother. The placenta itself also plays a very substantial role in intermediary metabolism, so that it cannot be assumed that there is no metabolism of absorbed nutrients during solute movement from mother to fetus.

As an epithelium, the trophoblast has two surfaces, the one facing the mother and the other facing the fetus. These surfaces differ structurally; for example, microvilli are found projecting into the maternal bloodstream at the apical surface, where they form a brush border, whereas the basal surface facing the fetus does not have this surface specialization. From the functional point of view, it is the differences in the distribution of transport membrane proteins (channels, carriers, and pumps) that determine the overall transport of nutrients from mother to baby, at least as far as water-soluble molecules are concerned. In the placenta, all transport appears to be across the trophoblast since, uniquely, the cells that compose this structure are fused to form a syncytium; this is in contrast to other epithelia where transport between adjacent epithelial cells allows a functional paracellular route to lie in parallel to the route through the epithelial cells themselves (the transcellular route). The question in the human as to whether there is a special route available for the transport of large molecules (MW 5000) is not yet resolved: certainly, at term, the human placenta does appear to be able to permit the transport of larger molecules at a slow rate from baby to mother. This route is unlikely to be of nutritional significance, since the rate of transport is low, but it may be important with regard, for example, to immunological sensitization.

The placenta, together with the growing fetus, has a very substantial metabolic energy requirement, and in the human, ATP synthesis appears to be met by the very substantial rate of glucose delivery from the mother. The glucose transporter systems that are found in both the apical and basal surfaces of the human trophoblast appear to be the type named GLUT1; in other words, they are sodium-independent facilitated transporters that are not regulated by insulin. The Michaelis constant (K_m) for glucose transport at both surfaces is relatively high (approximately 30 mM), and the maximal transport rate (V_{max}) is very substantial. The result of this is that glucose delivery across the brush-border membrane of the placenta will be in a direction and rate that are dependent solely on the chemical driving force from mother to baby (maternal–fetal plasma glucose concentration). It seems likely that this fundamental property is the basis for the macrosomia ('large-for-dates') found in the babies of mothers with elevated plasma glucose concentration, as typically found in diabetes mellitus. Transport of glucose, which is stereospecific, may be inhibited by glucose analogs that share the chemical structure of D-glucose at carbon 1; for example, both 3-0-methylglucose and 2-deoxyglucose are transported, whereas α-methylglucoside (with a methyl group on carbon 1) is not. The transport of other monosaccharides has been studied rather little; in the human, fructose is transported much more slowly than glucose itself (in contrast to other nonprimate species). The question of regulation of carbohydrate delivery across the placenta is not fully resolved since some glucose transporters may be regulated by phosphorylation, and the gradient for transplacental glucose delivery will also depend upon factors regulating glucose utilization and production by the placenta itself; little is known of the physiological regulation of either of these processes. (*See* **Carbohydrates**: Digestion, Absorption, and Metabolism; **Glucose**: Maintenance of Blood Glucose Level.)

In contrast to glucose transport, amino acid transport can be powered. The overall gradient of amino acid between maternal plasma and fetal plasma varies

[†]Deceased.

for individual amino acids. Typically, individual amino acid concentrations are twice as high in fetal as in maternal plasma. Current understanding of the mechanisms responsible for this relates to the distribution of the membrane transport proteins between the two faces of the trophoblast and in particular to the distribution of sodium-coupled transporters that are found predominantly (although not exclusively) in the brush border. Recent work using isolated membranes that reseal to form artificial structures (vesicles) has been useful in establishing the numbers and properties of such transporters. In addition to the direct effect of sodium ions in moving amino acids into the trophoblast across the brush-border membrane against a concentration gradient, these transporters are often electrogenic and are thus also driven physiologically by the membrane potential. One example of such a process is the transport system called 'A,' which uses alanine, serine, and proline as transported substrates and accumulates these amino acids in the trophoblast against a concentration gradient. These amino acids then leave the trophoblast across the basal membrane by a different transport system. Other amino acids may be transported via tertiary active transport; for example, leucine is found in higher concentrations in fetal plasma than in maternal plasma, but it is not itself a substrate for sodium-coupled transport; rather, it appears to exchange with amino acids, such as alanine, that have been accumulated in the trophoblast as just described.

The cationic and anionic amino acids are unusual in that, having their own charge, they will be accelerated or retarded by the membrane potential in crossing each of the plasma membrane surfaces of the trophoblast. For cationic amino acids (lysine, arginine, histidine), entry into the placenta appears to be largely by system y^+ (Na^+-independent), whereas exit into the fetus involves an electroneutral system (y^+L), which exchanges the cationic amino acid for a neutral amino acid (e.g., leucine) and a sodium ion, thus effectively solving the problem of permitting positively charged amino acids to exit against an inside-negative membrane potential. For anionic amino acids (glutamate and aspartate), very high intratrophoblast concentrations are achieved by a transport system that is coupled to K^+ efflux as well as Na^+ entry. Essential amino acid requirements for the fetus are different from those of adults. However, it is not clear whether transport of specific amino acids across the placenta ever becomes rate-limiting for fetal growth. In the human, intrauterine growth retardation not associated with other disease has been shown to be associated with reduced placental delivery of amino acids through specific systems, e.g., associated with decreased function of system A. IGF (insulin-like growth factor) and IGFBP (insulin-like growth factor binding protein) are now recognized as having an important role in either normal or abnormal fetal growth via controlling placental amino acids and glucose transport (e.g., IGF-I (insulin-like growth factor-1) selectively enhances system A activity). In certain unusual metabolic disorders (e.g., maternal phenylketonuria (PKU)) maternal levels of one particular amino acid may be elevated; this results in competition between this amino acid and others that share the same transporters. Some of the abnormalities found in the developing babies of such mothers may be a consequence of nutritional deprivation of tyrosine, for example, owing to competition by raised maternal phenylalanine levels for the delivery of this amino acid across the placenta. The fact that amino acid transport across the placenta involves a family of transport proteins with overlapping substrate (amino acid) specificities means that the nutritional consequences for the fetus of changing the level of one amino acid in the mother will be complex. This follows because, in contrast to placental glucose transporters, the K_m and V_{max} of the amino acid transporters are relatively low.

Lipid transport across the placenta in relation to human nutrition has been studied less rigorously, in part because it is likely to be perfusion- rather than membrane-limited, since the lipid-soluble nature of such a substrate allows ready transmembrane transport. Nutritionally, the nervous system of the developing fetus requires substrate delivery of precursors for myelin synthesis. Studies suggest that placental binding proteins may provide a pool of essential fatty acids for fetal utilization. (*See* **Amino Acids**: Metabolism; **Fatty Acids**: Metabolism.)

Transport of the inorganic cations of sodium and potassium involves both channels and transporters. Sodium transport into the trophoblast is coupled to the entry of those solutes (which include both organic and inorganic molecules) powered by secondary active transport. The extrusion of sodium across the basal surface of the trophoblast is likely to be a result of sodium pumping by Na^+/K^+ ATPase activity. In contrast, potassium, accumulated in the trophoblast, as in other epithelia by the sodium pump, requires channel-mediated release to account for its movements between mother and baby. Potassium channels have recently been shown to be sensitive to modulation in this tissue (e.g., by G proteins, by arachidonic acid, and by pH). These regulatory factors may themselves be controlled by circulating factors in both mother and fetus. It is clear that fetal plasma potassium is carefully regulated by control of placental transport of this cation.

The divalent cation of calcium is found in higher concentrations in fetal than in maternal plasma. Active transport processes are therefore involved in placental calcium metabolism, and these are unusual in that regulation of transport is clearly precisely controlled. It seems likely that calcium extrusion across the basal surface of the trophoblast is ATP-driven and regulated by calmodulin. Entry of calcium across the brush border is now known to be via ECaC (epithelial calcium channels) that are regulated: membrane potential hyperpolarization activates, and intracellular calcium inhibits, these channels. Fetal parathyroid hormone and vitamin D are likely to regulate all of these events.

Iron is also transported from mother to fetus in a regulated fashion, and again, there is a greater concentration of iron in the fetal circulation. In some species, transport of iron has been shown to be active in that it is inhibited by anoxia. The role of the transferrin receptor found in the human placental microvillus membrane is now known to be related to the mechanisms of iron entry into the cytosol. From this compartment, iron has to leave, but the mechanisms responsible for this are not fully understood. (*See* **Calcium**: Physiology; **Hormones**: Thyroid Hormones; **Iron**: Physiology; **Potassium**: Physiology; **Sodium**: Physiology.)

Anion entry into the placenta has been studied using isolated membrane preparations. These studies show that for monovalent ions (chloride and other halides), two routes are available, an exchange and a conductive route, the latter likely to be via channels. The anion exchange system appears to be functionally linked to the transport of organic anions (bicarbonate, lactate) from placenta into the maternal circulation. Phosphate transport is also regulated and appears to involve a sodium-dependent cotransporter at the maternal-facing surface and an efflux mechanism (possibly driven by the membrane potential) at the basal surface of the tissue. As for calcium transport, phosphate delivery is regulated by parathyroid hormone concentration in the fetus.

Trace-element delivery across the placenta also involves specific placental binding proteins analogous to those found in adult liver; however, membrane transport is also required to allow such ions to gain access to and from the trophoblast. The nature of such transporters varies greatly, although the role of the divalent cation transporter (DCT-1) in the placenta may be more generally important. For zinc, there is evidence that the histidine amino acid transporter is responsible for delivery of a histidine–zinc complex, whereas for transition-metal oxides, it appears that anion exchange is important. The transport of iodide across the human placenta is also likely to be by anion exchange since SCN can inhibit it, but the mechanisms responsible for the concentration of this element in the fetal compartment are not clear. Selenate appears to share a pathway with sulfate for entry across the brush border, and the transport of both of these ions is inhibited by blockers of anion exchange. This pathway appears to be shared with those available for transport of the trace elements chromium and molybdenum as chromate and molybdate. (*See* **Trace Elements**.)

Vitamin transport also is highly specific for individual substrates; thus, for ascorbic acid, a sodium-independent transporter for the reduced from of this nutrient has been described in the brush-border membrane; this transporter appears to be functionally coupled to a placental system that maintains ascorbate in this chemical form. The Na^+-dependent multivitamin transporter that transports pantothenate, biotin, and lipoate is expressed in human placenta.

See also: **Amino Acids**: Metabolism; **Calcium**: Physiology; **Carbohydrates**: Digestion, Absorption, and Metabolism; **Fatty Acids**: Metabolism; **Glucose**: Maintenance of Blood Glucose Level; **Hormones**: Thyroid Hormones; **Iron**: Physiology; **Potassium**: Physiology; **Sodium**: Physiology; **Trace Elements**

Further Reading

Bain MD, Copas DK, Taylor A *et al.* (1990) Permeability of human placenta *in vivo* to four non-electrolytes. *Journal of Physiology* 431: 505–513.

Boyd CAR and Kudo Y (1990) Placental amino acid transport. *Biochimica et Biophysica Acta* 121: 169–174.

Hoenderop JGJ, van der Kemp AW, Hartog A *et al.* (1999) Molecular identification of the apical Ca^{2+} channel in 1,25-dihydroxyvitamin D_3-responsive epithelia. *Journal of Biological Chemistry* 274: 8375–8378.

Kudo Y and Boyd CAR (1996) Placental tyrosine transport and maternal phenylketonuria. *Acta Paediatrica* 85: 109–110.

Parkkila S, Waheed A, Britton RS *et al.* (1997) Association of the transferrin receptor in human placenta with HFE, the protein defective in hereditary hemochromatosis. *Proceedings of the National Academy of Sciences, USA* 94: 13198–13202.

Shennan DB and Boyd CAR (1988) Trace element transport in placenta. *Placenta* 9: 333–343.

Sibley CP, Birdsey TJ, Brownbill P *et al.* (1998) Mechanisms of maternofetal exchange across the human placenta. *Biochemical Society Transactions* 26: 86–91.

Wang H, Huang W, Fei YJ *et al.* (1999) Human placental Na^+-dependent multivitamin transporter. *Journal of Biological Chemistry* 274: 14875–14883.

Safe Diet

R B Fraser and F A Ford, University of Sheffield, Sheffield, UK

Introduction

Would-be pregnant and pregnant women have heightened concerns about their diet, which can be summarized by the common question, 'Will it harm my baby?' Because the question is so common, many sources of advice are available in the lay press, not all as scrupulously researched and scientifically based as they might be. Another source of potentially misleading advice is the food industry and written media, the agents of which possess a lack of objectivity that may not be obvious to the casual reader. Articles proliferate with titles such as 'How to have a beautiful baby' and 'How to have a perfect baby'. These tend to contain a mixture of dietary and lifestyle advice with the implication that, if followed, the undesirable outcomes of pregnancy such as miscarriage, congenital malformation, and fetal death are avoidable. Professionals in the field of nutrition must be aware of this for the following reason. If a woman learns of such advice only after her pregnancy has failed in some way there is a potential for a lifelong burden of guilt. For this reason purveyors of advice should restrict themselves to that which has been demonstrated in scientifically valid experiments. In particular, they should be guarded in their extrapolations from observational studies and animal experiments.

This article highlights the principal areas in which women have dietary concerns. Some of these are dealt with in more detail elsewhere and are cross-referenced.

Periconceptional Nutrition

In an ideal world, pregnancies would be planned but in the real world a high proportion are not. It is important to remember that many women who have not planned their pregnancies and have not subsequently followed the advice outlined below may need reassuring that they have probably not caused harm to their baby.

The work of Frisch in the USA identified the link between body composition and ovulation in the human female. She suggested that fat must comprise at least 22% of body weight for the maintenance of ovulatory cycles and also observed that in normal postpuberty women fat is about 28% of body weight. It is well recognized that the relationship is to the fat content of the body rather than absolute body weight, as trained athletes of average or above-average body weight may have very low body fat content and may be oligo- or amenorrheic. Frisch observed that amongst trained athletes who became fit after a normal menarche, 60% continued regular cycles, but 40% had irregular cycles and presumably associated subfertility. Other causes of secondary amenorrhea leading to infertility include psychological stress, thyrotoxicosis, and the various malabsorption syndromes. Eating disorders such as anorexia nervosa and bulimia nervosa which may have been underdiagnosed in the past are now been seen as contributing to the problem. Particularly for bulimics, a remission in the disorder may allow the body fat to build up to a stage where ovulation and conception can take place, but the disorder may relapse during the pregnancy, leading to serious nutritional deficiencies for the mother and possibly for the fetus.

The Nurses Health Study in the USA has also produced some interesting information about the relative risk of menstrual cycle irregularity, not only in underweight but also in overweight women. For women with a body mass index (BMI) below 20, at the age of 18 years, ovulatory infertility was found with a relative risk of about 1.2 compared to women with a BMI between 20 and 25. Interestingly, however, the relative risk of ovulatory infertility was 1.5 in those with a BMI of 28 and more than 2 in the obese group with a BMI above 30. About half of the risk is associated with polycystic ovarian syndrome (PCOS) in which ovulatory infertility and obesity coexist but there is still a doubling in relative risk of ovulatory infertility in women with a BMI above 30 who do not have ultrasonically detectable polycystic ovaries.

Vitamin Intakes and Congenital Malformations

Folic Acid

In 1991, the Medical Research Council (MRC) vitamin trial reported that a significant reduction in the recurrence of neural tube defects (NTD) had been obtained by a daily dietary supplement of 4 mg of folic acid in women who were at high risk. Women at risk of a recurrence of NTD should be advised to take a folic acid supplement of 5 mg (5000 μg) per day when planning a pregnancy and continue with it until 12 weeks' gestation. A later study reported that first occurrence of NTD could also be prevented by daily supplements containing 800 μg of folic acid. In the UK, in line with many other nations, women are advised to take a prepregnancy supplement of 0.4 mg (400 μg) of folic acid as a daily medicinal supplement from when they begin trying to conceive

until 12 weeks' gestation, in addition to eating a folate-rich diet and breads and breakfast cereals fortified with folic acid.

The mechanism for the effect of folic acid in reducing the risk of NTD is not clear. The most likely explanation is that it overcomes genetically determined defects in folate metabolism that interfere with normal neural tube development. One other hypothesis is that it might decrease the likelihood of an affected pregnancy surviving, and there is some concern that folic acid supplementation could be associated with a higher risk of fetal death. However, the generally accepted explanation for this is that folic acid might initially permit the survival of affected or nonviable pregnancies to a point where they are recognized as spontaneous abortions.

Educational campaigns conducted since 1995 have had a substantial effect in increasing the number of women in the UK who are aware of the link between folic acid and NTD. However, such education can only have a limited effect because it is estimated that only 50% of all pregnancies are planned. NTD arise very early in gestation and by the time a woman realizes that she is pregnant it is usually too late to prevent an NTD by taking supplements. For this reason it has been suggested that the most effective way of reducing NTD in the UK would be to fortify the food supply with folic acid so that even women with unplanned pregnancies would be less likely than at present to have offspring with NTD. The Department of Health Committee on Medical Aspects of Food and Nutrition Policy (COMA) has recently recommended universal fortification of flour at 240 μg $100\,g^{-1}$, which would reduce the risk of NTD by 41% without resulting in unacceptably high intakes in any group of the population.

Retinol

It has been known for a long time from animal studies that high levels of vitamin A in the form of retinol are teratogenic in the periconceptional period. This led to the Department of Health in 1990 advising women to avoid food sources containing high levels of the retinal form of vitamin A in the periconceptional period and indeed during pregnancy. The Department of Health advises that very high intakes of retinol, i.e., more than 10 times the recommended dietary allowance either in liver/liver products or vitamin/fish liver oil supplements, can damage the developing embryo. This advice remains current and all women of child-bearing age should avoid liver and its products and should not take supplements containing more than four times the recommended daily amount of retinol.

The same risks do not seem to be present for vitamin A derived from β-carotene. Manufacturers of multivitamin supplements recommended for pregnancy have generally recognized this and switched the source of vitamin A to β-carotene.

Nutritional Management of Common Symptoms in Pregnancy

Heartburn

Heartburn is thought to be caused by gastroesophageal reflux. Although occasionally experienced in the first trimester, it is generally more common in the last trimester and occurs in 30–50% of women.

Small, frequent meals or snacks are usually tolerated better than large, well-spaced meals. Common foods cited as causing heartburn as spicy and fatty foods, fizzy drinks, citrus fruits, fruit juices, and cucumber. Milk and milk products can help to relieve symptoms but antacids are frequently used.

Nausea and Vomiting

Psychological factors, changing hormone levels, hunger, altered carbohydrate metabolism, and vitamin deficiencies have all been proposed as possible causes for nausea and vomiting, but none has been confirmed.

Symptoms may start before the woman knows she is pregnant or in later pregnancy but commonly they are worst between weeks 6 and 10, and subside by about 13 weeks. Nausea is experienced at any time of day or night and can be either slight or severe. It often becomes worse when the stomach is empty, and eating small, frequent meals based on starchy carbohydrates may relieve it. Morning sickness is common and consuming dry biscuits or toast before getting up can relieve this.

Nausea can also be triggered by traveling, fried and spicy foods, and smells such as coffee, perfume, and cigarette smoke.

Some women just feel nauseated while others actually vomit as well; this may cause minor weight loss but rarely causes nutrient deficiency. Women need reassurance that not eating proper cooked meals or losing some weight and their taste alterations will cause no problems for their developing fetus. The more severe cases of pregnancy vomiting (hyperemesis gravidarum) require hospital admission, intravenous fluids and, sometimes, parenteral nutrition.

Constipation

Constipation is common at all stages of pregnancy. It may be related to a general reduction in motility in the gastrointestinal tract, with prolonged transit

times and increased water resorption from the stool. General advice about constipation is also suitable for pregnancy, i.e., increased intake of fiber, particularly cereal fiber, and increased fluid intake.

Constipation may be aggravated by the consumption of iron tablets; if it is not appropriate to reduce or stop them, bulking agents may be prescribed.

Qualitative Aspects of Diet

The following section refers to common questions of dietary safety; some arise because of suspicion of harm when items are included in the diet, others for the paradoxical reason that their omission from the diet might be dangerous.

Alcohol

As far as alcohol is concerned, the picture is confusing. There is no doubt that a heavy intake of alcohol can damage the unborn baby and cause miscarriage. Many women choose to give up alcohol and this seems to be a sensible practice but there is no evidence of harm from occasional drinking or the consumption of less than 2 units per day (Table 1). Despite this, many women do give up alcohol when trying to conceive or whilst pregnant, and this seems a sensible but not mandatory practice. (One unit is 15 g of absolute alcohol, e.g., 0.28 l of beer, one glass of wine, or one measure of spirits.)

A well-defined group of anomalies referred to as the fetal alcohol syndrome is now recognized. The major defects of affected infants are weight and length below the 10th centile for gestational age, and microcephaly. Microcephaly is a condition of small head size associated with an underdeveloped brain. Such children are likely to be mentally retarded and their physical growth is stunted. The syndrome has been reported in up to 40% of the infants of women drinking more than 6 units of alcohol per day.

Definite harm has been recorded to the offspring of women drinking between two and six units of alcohol per day and long-term growth retardation and mental retardation to offspring of those drinking more than 6 units a day (fetal alcohol syndrome).

Table 1 Congenital malformation rates and maternal alcohol consumption

Drinks per day	*Malformation rates (per 1000)*[a]
None	78
Less than one	77
One to two	83

[a]No statistically significant difference.

From Mills JL and Graubard BI (1987) Is moderate drinking during pregnancy associated with an increased risk for malformations? *Pediatrics* 80: 309–314.

A recent study showed that a woman's alcohol intake is associated with decreased fertility, even among women with a weekly alcohol intake corresponding to five or fewer drinks. This finding needs further corroboration, but it seems reasonable for the moment to encourage women to avoid alcohol if they are having trouble conceiving.

Caffeine

Concern about the detrimental effects of caffeine during pregnancy is not new but to some extent the literature is conflicting. Caffeine is present in tea, coffee, cola drinks, and many over-the-counter remedies for colds and allergic symptoms. Animal experiments with high doses have shown that some congenital malformations may be induced but there is no evidence of harm in humans. Current recommendations are that pregnant and breast-feeding women should limit their caffeine intake to 300 mg day^{-1}, which is equivalent to 4 cups ordinary-strength coffee per day (or 6 cups of tea or 7 cans of cola). A recent review from the UK Committee on Toxicity of Chemicals in Food, Consumer Products and the Environment (DoH 2001) issued a statement on the reproductive effects of caffeine in October 2001. A meta-analysis of studies of maternal caffeine intake during pregnancy and the risk of spontaneous abortion or low birth weight, compared maternal caffeine intakes during pregnancy of more than 150 mg/day with less than 150 mg/day. Calculated odds ratios were significantly increased for spontaneous abortion (miscarriage) (odds ratio = 1.36; 95% confidence interval, 1.29–1.45) and low birth weight (odds ratio = 1.51; 95% confidence interval, 1.39–1.63), for low birth weight (< 2.5 kg).

Two studies, which examined the links between caffeine intake and pregnancy in SIDS, received a huge amount of publicity and caused a lot of worry for women. However, the results were conflicting.

Peanuts

Peanut allergy is increasing in children, although the cause is unclear. The use of peanuts and peanut oil in the British diet has increased rapidly over the last few years and it is thought that being exposed to peanuts at a young age may cause the allergy. However, it is not known if this can happen as a result of a mother eating peanuts when she is pregnant or breast-feeding. Current advice is that women, their partners, and existing children who have a family history of atopic disease should avoid peanuts.

Over-the-Counter Remedies

As a general rule, self-prescribing in pregnancy should be kept to a minimum. Those drugs that

have been used most widely and have been shown to be safe should be preferred to anything new or untested. Small doses of simple analgesic drugs such as aspirin and acetaminophen (paracetamol) appear to be safe, as do common cold remedies, which often have these drugs as their active contents.

No problems have been reported with the use of homeopathic remedies in pregnancy. There has been much interest in possible therapeutic benefit from the consumption of oil of evening primrose in pregnancy. It has been hypothesized that the linoleic acid and γ-linoleic acid in the compound may stimulate production of vasodilatory prostaglandins and be of benefit in lowering blood pressure in abnormal states of pregnancy such as preeclampsia. The limited studies so far reported do not support this hypothesis.

Recent reports have compared women who are on diets rich in natural fish oil, the Faroese, with women on the Danish mainland whose diets contain much less fish oil. The Faroese had fewer premature babies, fewer problems with blood pressure, and heavier babies than the mainlanders. Unfortunately, these studies cannot be interpreted as showing uniform benefit from fish oil supplements as the Faroese also lost more babies as stillbirths. A further report, comparing women in Aberdeen with women in the North Sea Orkney Islands, showed the latter to have average birth weights 250 g higher. After correction, a small but significant proportion of the difference was found to be related to genetic or environmental factors. The latter included a diet in Orkney containing 30% more fish than in Aberdeen.

Ginger root may act as an antiemetic by a local effect in the stomach. Evidence of benefit in pregnancy nausea exists.

Garlic is widespread in the human diet. There has always been a recognition of possible benefits of garlic supplements on a wide range of disorders, including hypertension, hyperlipidemia, and thrombosis. No studies of supplementation in pregnancy have been reported.

Vitamin and Mineral Supplementation

Some pregnant women benefit from iron and/or folic acid supplements when dietary deficiency provokes anemia. Although self-medication with vitamin and mineral supplements is widespread, there is a good case during pregnancy to restrict such supplements to those prescribed by doctors or preparations containing less than 100% of the recommended dietary allowance. Vitamin A in the form of retinol and its analogs can cause congenital malformations in megadosage, and vitamin D supplementation is thought to have been the cause of idiopathic hypercalcemia. This was a congenital syndrome consisting of abnormal facial features, mental retardation, and abnormal calcium metabolism. It was common in the UK in the late 1950s and early 1960s when vitamin D supplementation of a variety of foods was practiced. (The relationship of periconceptional multivitamins to NTDs is discussed elsewhere.)

It seems clear that some members of dark-skinned races are at risk of vitamin D deficiency when resident in temperate climates. Their newborns may be at risk of hypocalcemia, a cause of convulsions in the newborn period. These children may also have delayed growth in the first year of life. Both problems may be corrected by vitamin D supplementation in pregnancy.

At various times, maternal zinc deficiency has been proposed as a cause of both congenital malformation and retarded fetal growth. No valid evidence has yet emerged of any benefit derived from maternal zinc supplementation with regard to either outcome in humans.

Food Safety in Pregnancy

Pregnant women share the concerns of the general public about food poisoning but there is a great deal of confusion in the mind of the public and the clinicians treating them about which foods are safe to eat during pregnancy.

Listeria monocytogenes

This organism is widely distributed in the environment and the low incidence of infection suggests that infection requires the ingestion of a large dose by a susceptible individual. Listeriosis is a significant problem in pregnancy because infection of the mother can cause fetal death following transmission of the organism across the placenta. It is not certain at the present time whether the apparent increase in fetal loss from this infection represents improved ascertainment or a genuine increase in disease frequency.

Most women do not realize that the symptoms of *Listeria* infection are not the same as those for some other foodborne pathogenic organisms: *Listeria* symptoms are often described as mild 'flu-like symptoms' rather than diarrhea and vomiting. The perceived risk of *Listeria* infection in pregnancy for many women appears to be much greater than the actual risk.

During pregnancy the following foods should be avoided to reduce the risk of *Listeria* infection:

- Mold-ripened cheese and some types of sheep and goats' milk cheese
- Pâté – meat, fish, and vegetable (unless canned or ultra heat-treated)

- Cook–chill foods that have not been thoroughly reheated

However, many health professionals advise women to avoid 'unpasteurized soft cheeses' which, to most pregnant women, means avoiding all types of cream and cottage cheese together with yogurt and fromage frais. In fact, it is only mold-ripened cheeses with either a skin such as that found on Brie and Camembert or a mold inside such as Danish blue that should be avoided, together with sheep and goats' milk cheese that has not been made with pasteurized milk. Unnecessary avoidance of these dairy foods will substantially decrease calcium intake.

Toxoplasmosis

Toxoplasma is a protozoan organism which can cross the placenta when primary infection occurs in pregnancy. It can lead to fetal death, mental handicap, and/or blindness in an estimated 30% of offspring of infected mothers. To reduce the risk of infection during pregnancy women are advised to:

- Avoid eating raw and undercooked meat.
- Always wash vegetables and salads well to remove soil.
- Wear gloves when gardening.
- Avoid unpasteurized sheep and goats' milk.
- Wash hands after contact with cats and kittens and avoid contact with stray cats.
- Wear rubber gloves when cleaning out cat litter trays or get somebody else to do this job if possible.

(*See* **Vegetarian Diets**.)

Salmonellosis

Salmonella bacteria are one of the most common causes of food poisoning in the UK, giving rise to sickness and diarrhea. Steering clear of raw eggs and products containing uncooked egg and exercising appropriate food hygiene can reduce this risk.

Campylobacteriosis

The main source of *Campylobacter* infection is raw milk and poultry but a large number of foods have been implicated in outbreaks. Again, campylobacter may not have any direct effect on the fetus, but the Department of Health advises women in the UK to avoid this distressing illness.

Shellfish

Pregnant women are advised not to eat oysters and other shelled seafood, such as prawns, mussels, and crab, unless they are part of a hot meal and have been thoroughly cooked. When raw these foods may be contaminated with harmful bacteria and viruses.

Bovine spongiform encephalopathy

In 1996 evidence of a new variant of Creutzfeldt–Jacob disease (CJD) began to be observed, which is most probably the result of eating beef infected with bovine spongiform encephalopathy (BSE). A particular source of concern at the moment is vertical transmission which, although common in cattle, has not been shown to occur in humans. However, there has recently been a possible case reported in the media of mother to child infection.

Other concerns of pregnant women in the area of food safety include the hazards of fungicide and pesticide residues. and possible harmful effects of natural food toxicants. At present it seems unlikely that there are toxic residues on foods which can be associated with harm in pregnancy, but this observation should not be taken as grounds for complacency. One of the main difficulties for scientists wishing to explore this matter is the limited availability of reproducible assays for the minute quantities to be studied. Similarly, the study of natural food toxicants for possible harmful effects on the fetus is in its infancy.

Conclusions

It is important to monitor continuously any effects our dietary habits may be having on child-bearing and child development, but we should be able to achieve this without causing unnecessary scares. Most women who consume a good general diet, and are not addicted to tobacco, alcohol, or other drugs, can contemplate pregnancy without concern that their dietary habits can 'harm the baby'.

See also: **Caffeine**; ***Campylobacter***: Campylobacteriosis; **Cholecalciferol**: Properties and Determination; Physiology; **Coffee**: Physiological Effects; **Folic Acid**: Properties and Determination; Physiology; **Food Intolerance**: Food Allergies; **Food Poisoning**: Classification; **Food Safety**; **Garlic**; ***Listeria***: Properties and Occurrence; Detection; Listeriosis; **Retinol**: Properties and Determination; Physiology

Further Reading

Department of Health (2001) Committee on Toxicity of Chemicals in Food, Consumer Products and the Environment. Statement on the Reproductive Effects of Caffeine 2001/06. London: HMSO.

Department of Health and Social Security (1990) *Vitamin A and Pregnancy*. PL/CMO (90) 11, PL/CNO (90). London: Stationery Office.

Department of Health and Social Security (2000) *Folic Acid and the Prevention of Disease*. London: Stationery Office.

Frisch RE and McArthur JW (1974) Menstrual cycles: fatness as a determinant of minimum weight for height

necessary for their maintenance and onset. *Science* 185: 949–951.
Jensen KT, Hjollund NH and Henriksen TB (1998) Does moderate alcohol consumption affect fertility? Follow up study among couples planning first pregnancy. *British Medical Journal* 317: 505–510.
Karen G (ed.) (1990) *Maternal–Fetal Toxicology. A Clinician's Guide*. New York: Marcel Dekker.

Maternal Diet, Vitamins, and Neural Tube Defects

M L Watkins, National Center on Birth Defects and Developmental Disabilities, Atlanta, GA, USA
S L Carmichael, March of Dimes/California Birth Defects Monitoring Program, Oakland, CA, USA

Background

The most interesting work relative to neural tube defect (NTD) etiology and prevention in the last few years has focused on nutritional factors. This chapter summarizes what is known about the relation between NTDs and various nutritional factors, with particular emphasis on folate, and includes a discussion of the methodologic challenges inherent in identifying such associations.

Neural Tube Defects

NTDs result from failure of normal closure of the embryonic neural tube at one or more of five separate closure sites. Complex forces are involved, with varying closure mechanisms at different sites within the neural tube. Of the several NTD types (**Figure 1**), spina bifida is the most common. The extent of physical disability (e.g., paralysis, bowel and bladder dysfunction) associated with spina bifida depends on the lesion level, size, and extent of spinal nerve involvement. Developmental, learning, and social problems are common. Most children born with spina bifida in developed countries now survive because of extensive medical and surgical treatment. Many of these children, however, face lifelong physical and developmental disabilities. The death rate for affected infants in the USA is about 10%, but the rate is much higher in countries with less accessible medical care. Clinical deterioration and secondary conditions occur in many affected children in a variety of body systems, resulting in complex medical and surgical care. Because of a lack of information about the natural history of untreated children, a lack of randomized trials, and a lack of diagnostic criteria for neurologic complications, there are controversies about optimal care. The economic costs of spina bifida are huge. The role and effectiveness of intrauterine myelomeningocele repair is controversial; clinical trials have been recommended.

NTDs affect at least 300 000 newborns worldwide each year. NTD rates vary considerably around the world, with northern China reporting some of the

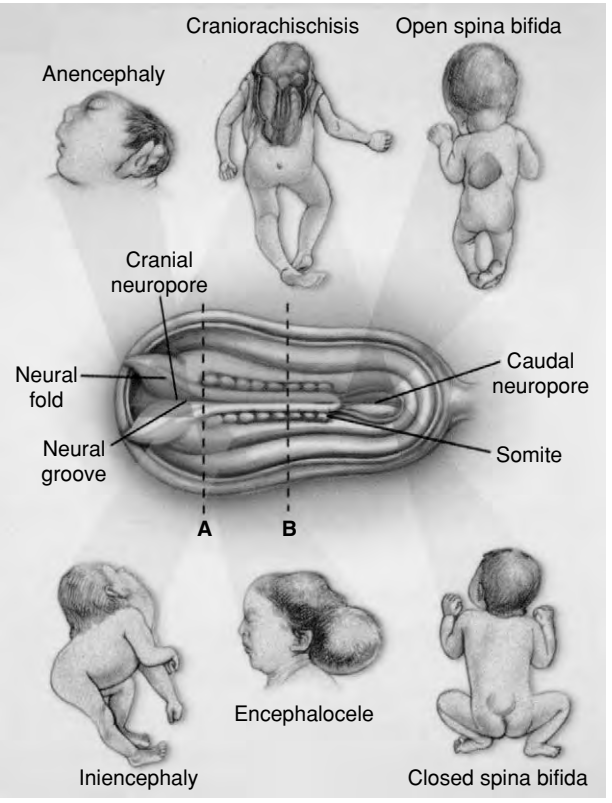

Figure 1 Neural tube defects. Schematic view (from above) of a neural tube. Main clinical types of neural tube defects. Anencephaly (lateral view): the absence of the brain and skull can be total or partial, always fatal. Craniorachischisis (posterior view): anencephaly occurs with contiguous bony defect of the spine and exposed neural tissue. Open spina bifida (posterior view): bony defect of the vertebrae (in this case, of the lower thoracic vertebrae), accompanied by exposure of neural tissue and meninges that are not covered by skin. Closed spina bifida (posterior view): bony defect of the vertebrae and, if present, the herniated meninges and neural tissue, are covered by skin. Extent of the disability associated with spina bifida related to the level and size of the lesion and the extent of spinal nerve involvement. Encephalocele (lateral view): the brain and meninges herniate through a skull defect. Iniencephaly (lateral view): the dysraphic process in the occipital region is accompanied by severe retroflexion of the neck and trunk. Spina bifida and anencephaly are more common than encephalocele, craniorachischisis, or iniencephaly. Adapted from Botto LD, Moore CA, Khoury MJ and Erickson JD (1999) Neural-tube defects. *New England Journal of Medicine* 341: 1509–1519, with permission.

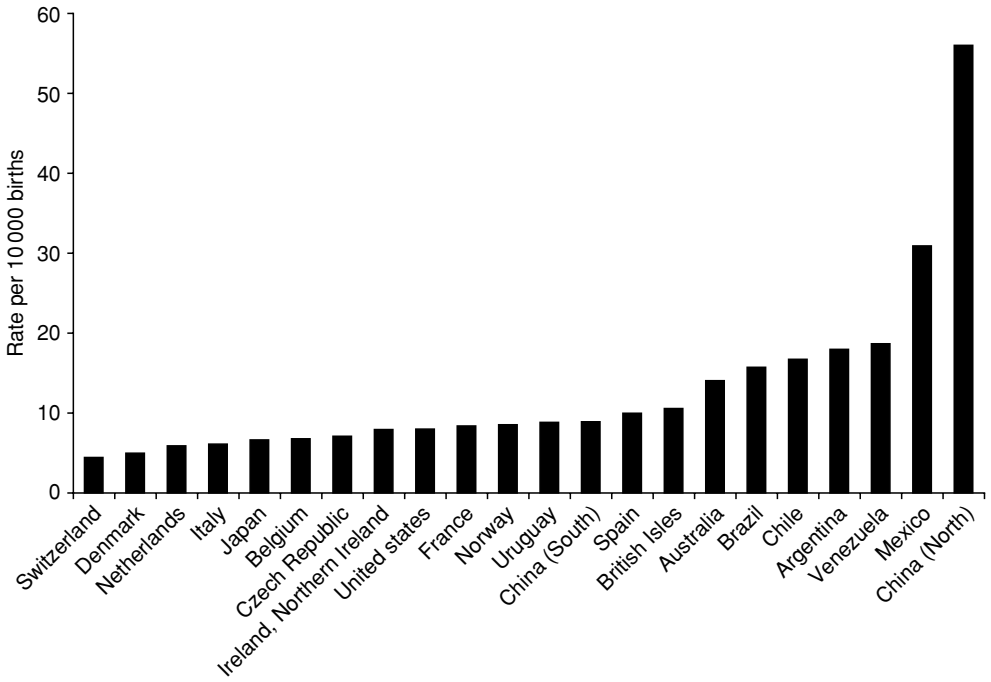

Figure 2 Rates of spina bifida and anencephaly by country. Cases reported in pregnancy terminations are included, where available. Rates for Latin-American countries and Japan are estimated from hospital-based registries. Rates for other locations are based on data from birth defects registries that usually monitor only part of a country's birth population. The following countries and registries are included: Switzerland (selected counties), Denmark (Odense), The Netherlands (North), Italy (Campania, Emilia-Romagna, Toscana), Belgium (Hainaut-Namur), Czech Republic, Ireland (Dublin, Belfast), United States (Atlanta, Hawaii), France (Bouches-du-Rhone, Paris, Strasbourg, Central East France), Norway, Spain (Basque country), UK (Glasgow, North Thames), Australia (Victoria, South Australia), and China (selected counties). Data from the International Center for Birth Defects and the European Registration of Congenital Anomalies (1998) *World Atlas of Birth Defects*, pp. 20–31. Geneva: World Health organization.

highest rates (Figure 2). Both environmental and genetic factors likely explain the variation in rates. Of NTDs, 80% are thought to be 'multifactorial,' meaning that they are influenced by both genetic and environmental factors. A minority of NTDs are caused by chromosomal abnormalities, single gene mutations and teratogens. Risk factors include maternal diabetes, use of folic acid antagonists, fever or hyperthermia in early pregnancy, low socioeconomic status, and obesity. In the USA, NTD rates are higher among Hispanic whites than non-Hispanic whites; blacks have the lowest rates. NTD-rate monitoring is complicated by the increasing use of prenatal diagnosis and selective termination, which results in lower birth prevalence. Because the neural tube develops very early in pregnancy (closure by 28 days postconception) before most women are aware they are pregnant, prevention efforts are challenging, especially in populations with high unplanned pregnancy rates. Approximately 95% of all NTD-affected pregnancies occur in women with no previous history, which is referred to as 'occurrence.' However, women who have already had an NTD-affected pregnancy have a 10-fold greater risk for having an NTD-affected subsequent pregnancy than women who have not. Having more than one NTD-affected pregnancy is referred to as 'recurrence.'

The increased prevalence of NTDs in lower socioeconomic groups, decreasing NTD rates, and seasonal patterns of occurrence are consistent with a dietary etiology for NTDs. Certain nutrients (specifically folic acid) were thought to be involved because of their role in human growth and because birth defects occurred with maternal use of folic acid antagonists (e.g., aminopterin, when used as an abortifacient, induced fetal NTDs). Also, animal data have demonstrated NTD inducement resulting from various nutritional deficiencies and excesses, suggesting the plausibility of nutritional causes of NTDs.

Methodologic Issues in the Study of Nutrition and NTDs

However plausible the link between nutrition and NTDs, studying these associations presents inherent challenges. Approaches to measuring nutritional status can be broadly classified as dietary, biochemical, anthropometric, or genetic. Many of the

difficulties in studying nutrition and birth defects stem from the fact that most studies are retrospective in nature, resulting from the relatively low prevalence of NTDs. This limitation has varied implications, depending on the particular approach used to measure nutritional status.

The food-frequency questionnaire is a tool commonly used to assess dietary intake retrospectively. This method produces reasonably valid and reliable, comprehensive, and semiquantitative data on usual or average intake of foods and nutrients. The validity of the data depends, however, on several factors, including the appropriateness of the food list for the study population; the accuracy of the nutrient database; and variability in individual nutrient requirements and in the bioavailability and absorption of nutrients from foods (which are affected by, for example, food processing, fiber intake, alcohol consumption, cigarette smoking, and infection). Challenges related to the analysis of dietary data include whether to analyze intake from foods and supplements separately or together; foods versus nutrients; or single versus multiple nutrients simultaneously. The high correlation that exists between nutrients makes it particularly difficult to isolate the independent effects of individual nutrients.

Depending on the particular nutrient, biochemical measures of nutritional status may vary by time of day, by the tissue from which the measurement is made (serum and toenail zinc, for instance, reflect more short-term versus long-term zinc status, respectively), and by various host factors (e.g., genes, behaviors such as smoking and alcohol consumption, health status, physical activity, stress, and medications). Furthermore, pregnancy itself is a time of changing nutritional requirements and absorption. These factors must be taken into consideration when interpreting results from individual studies and when comparing results across studies.

Anthropometric measurements (e.g., height and prepregnancy weight) can be recalled reasonably well. Whether retrospectively collected data on more complex measures (e.g., skinfold thickness and waist-to-hip ratio) can serve effectively as proxies for these parameters around the time of conception is questionable.

Genetic material is most commonly obtained from blood samples or buccal smears. The challenge is in discovering functionally relevant polymorphisms and in having sample sizes large enough to detect meaningful differences in risk. Recent advances in technology are enabling more rapid identification and analysis of polymorphisms, and innovative approaches to epidemiologic study design are being developed that enable smaller numbers of subjects for examination of certain types of genetic hypotheses.

Folic Acid and NTDs

In 1930, Dr. Lucy Wills discovered a factor that cured the nutritional deficiency anemia of pregnancy among women in India. This factor was later isolated from spinach and named 'folic acid' (*folium* is Latin for 'leaf'). Because humans are unable to synthesize folate, they must depend solely upon dietary sources. Folate-rich food sources include green leafy vegetables, grains, legumes, certain fruits, and liver. Because heat, ultraviolet light, and air inactivate food folate, food processing, preparation, and cooking can reduce the amount of food folate ingested by an estimated 50–95%. *Bioavailability*, the extent to which folates are available for use at the cellular level, varies widely across foods. The bioavailability of synthetic folic acid (monoglutamate form), which is used in cereal grain fortification and multivitamin/mineral supplements, is estimated to be about twice that of food folate (polyglutamate form).

Studies in the mid-twentieth century linked nutrition and NTDs. Lower vitamin C and folate levels in one study and poorer diet quality (in terms of macronutrient and fresh fruit and vegetable intake) in another were found among women with NTD-affected children. Intervention and observational studies followed (Figure 3). The most convincing and solid evidence for a preventive effect of folate is provided by two intervention trials. The international Medical Research Council (MRC) trial, a large UK-sponsored *recurrence* prevention trial, was conducted at 33 sites in seven countries. Women who had a previous NTD-affected pregnancy were randomized to one of four vitamin-use groups. Among 1195 pregnancies with known outcomes, folic acid (alone or with other vitamins) was associated with a 72% reduction in risk.

In 1992, a Hungarian randomized trial provided strong evidence for the efficacy of folic acid-containing multivitamin supplementation to prevent NTD *occurrence*. Results from this trial, combined with results from several observational studies (Figure 3) provided convincing evidence that occurrence is preventable. As a result, several countries recommended in 1992 that women of childbearing potential consume periconceptional folic acid daily (most commonly 400 μg) to prevent NTDs. The recommendation was made for all women capable of becoming pregnant (not just those planning a pregnancy), because these birth defects occur before many women are aware that they are pregnant, and because many pregnancies are unplanned. Since then, a large, nonrandomized community intervention in China demonstrated significant NTD reductions associated with the use of a 400-μg supplement containing only folic acid. Risk reductions were greater (85%) in the

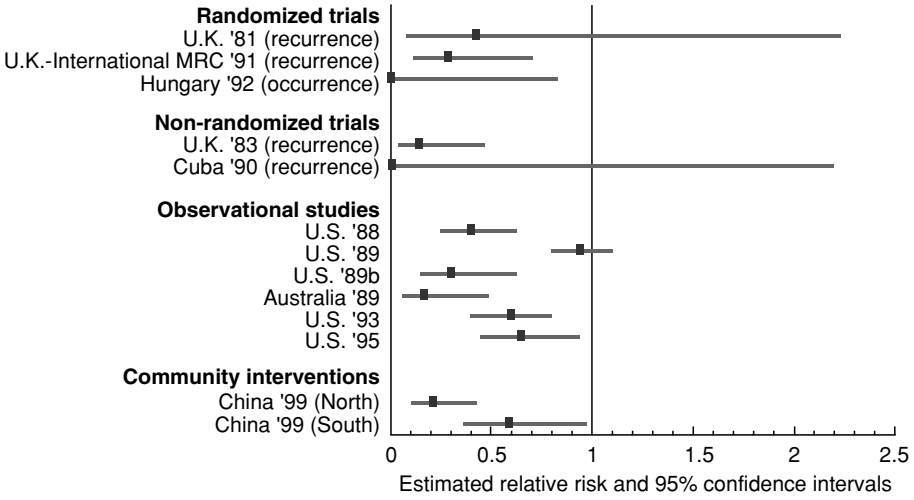

Figure 3 Risk of neural tube defects and use of folic acid or multivitamin supplements: summary of studies 1980–95, by country and year of publication. The figure shows risk estimates (boxes) and 95% confidence intervals. Risk estimates < 1 indicate a reduction in risk among the 'exposed' mothers (periconceptional use of folic acid or multivitamins). If the upper limit of confidence intervals is > 1, the risk estimate is not statistically significant. Some studies used folic acid (varying doses); some used multivitamins. Two randomized controlled trials (UK'81 and UK-international-MRC'91) and two nonrandomized trials (UK'83 and Cuba'90) examined the efficacy of folic acid with or without other vitamins among women with a previously affected pregnancy (recurrence). One randomized trial (Hungary'92) assessed the efficacy of a multivitamin containing folic acid among women without a previously affected pregnancy (occurrence). Of the observational studies, five (US'88, US'89, Australia'89, US'93, US'95,) were case-control studies, whereas one (US'89b,) was an observational cohort. The community intervention study (China'99) was conducted in two areas, one in Northern China and one in Southern China; it assessed the effectiveness of 0.4-mg folic acid supplements (without other vitamins) among women without a previously affected pregnancy in areas of high (North) and low (South) occurrence rates of neural tube defects. From Botto LD, Moore CA, Khoury MJ, and Erickson JD (1999) Neural-tube defects. *New England Journal of Medicine* 341: 1509–1519, with permission.

high-prevalence northern countries than in the lower-prevalence southern counties (40%). Additional evidence for the efficacy of folic acid comes from another study that found that periconceptional exposure to folic acid antagonist medications more than doubles the risk for NTDs and that folic acid supplementation reduced the risk elevation associated with certain folic acid antagonists.

Efficacy Issues

Based on the risk reduction observed in several studies, estimates are that at least 50% of NTDs could be prevented by the use of folic acid. The proportion of NTDs that are preventable may be less in low-prevalence areas than in high-prevalence areas. Some studies also suggest a smaller reduction in risk for NTDs associated with folic acid use for Hispanics than for non-Hispanic whites or blacks, and for obese women compared with average weight women. These observations require more study. The minimum daily effective dose of folic acid for NTD prevention is unknown; it may be < 400 μg for occurrence prevention, but determining the effectiveness of lower dosages is difficult because of the size and cost of studies necessary to assess the relative efficacy of various dosages.

The reasons for variability in folate efficacy are unknown, largely because the biologic mechanism by which periconceptional folic acid use protects against NTDs is unknown. Folic acid plays a role in transfer of methyl groups in the amino acid methylation cycles, a process essential for recycling homocysteine back to methionine (Figure 4). It is also a cofactor for enzymes involved in DNA and RNA synthesis, providing single carbon units for the synthesis of nucleotide bases, a process that is essential to the rapid cell division that occurs in early fetal development. Extrinsic folic acid may work by increasing tissue folate levels enough to override a failure in folate metabolism, or it may compensate for deficiency or metabolic defects related to other nutrients closely related to folate metabolism. Recent research has focused on the role of mutations in the genes that code for enzymes involved in folic acid metabolism.

Approaches to Increasing Folic Acid Consumption

Although the mechanism is unknown, a significant proportion of NTDs can be prevented with folic acid. The challenge is in how to translate science into public health policy and practice. Table 1 summarizes three approaches for increasing folic acid consumption.

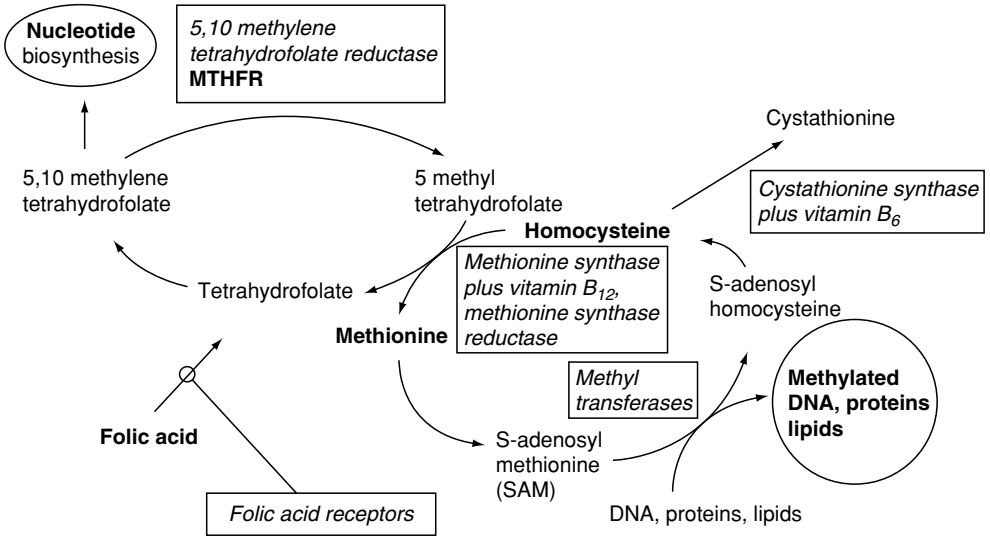

Figure 4 Metabolic roles of folic acid and related dietary factors. Simplified schematic of some of the metabolic processes that involve folic acid and other related factors. Folic acid (bottom left), after entering the cell (aided by folate receptors), participates in the transfer of carbon units used for the synthesis of nucleotides (top left) or, through the conversion of homocysteine to methionine (center), for methylation purposes. These processes are regulated by many factors including enzymes (in rectangular boxes) and vitamins other than folic acid (e.g., vitamins B_6 and B_{12}). The activity of some enzymes (e.g., methionine synthase) may be influenced by other enzymes (e.g., methionine synthase reductase). Adapted from Botto LD, Moore CA, Khoury MJ, and Erickson JD (1999) Neural-tube defects. *New England Journal of Medicine* 341: 1509–1519, with permission.

Table 1 Ways to increase folic acid/folate consumption to prevent neural tube defects

Approach	*Advantages*	*Disadvantages*	*Status*
Fortify staple foods	Proven to increase blood folate level Reaches almost all women Behavior change not necessary unless higher intakes needed Inexpensive	All consumers (not just target population) increase intake; concern that some will have excess intake Variable staple consumption results in variable folic acid intake Sustained behavior change required for consumption of higher amounts	Mandatory fortification in the USA, Canada, and several Central and South American and eastern European countries; proposed in the UK
Increase use of supplements	Efficacy proven by clinical trials Proven to increase blood folate level Relatively inexpensive	Requires sustained behavior change Reservations of some medical and nutritional professionals	Varies by country and survey question (e.g., in the USA, 30% of reproductive-age women report regular use; in The Netherlands, 54% of women report 'any periconceptional use' and 21% 'appropriate use')
Increase consumption of foods with high levels of natural folates (e.g., fruits and vegetables)	Other benefits of healthy diet that includes fruits and vegetables	Efficacy not proven Much smaller increases in blood folate levels than supplements or fortified foods Requires significant and sustained behavior change May be expensive	Various healthy diet campaigns under way

Fortification of a food staple provides wide population coverage at low cost without requiring behavior change. The optimal fortification level is controversial. Some experts had concerns that daily folic acid intakes in excess of 1 mg could adversely affect some persons, particularly persons with untreated cobalamin deficiency. Folic acid can ameliorate the anemia associated with cobalamin deficiency, possibly leading to a failure to detect and treat the cobalamin deficiency, which could result in the initiation or

progression of neuropathy. Although some experts in the study of birth defects advocated a higher fortification level, the US Food and Drug Administration ruled that, effective January 1, 1998, all flour, corn meal, pasta, and rice labeled as 'enriched' be fortified with folic acid at 140 mg per 100 g of cereal grain product. Canada and several Central and South American and eastern European countries also have mandated fortification of wheat flour. The fortification level in some countries (e.g. Chile) was established at a higher level, with the goal of achieving a higher average daily intake of folic acid. In 2000, the UK Committee on Medical Aspects of Food and Nutrition Policy recommended folic acid fortification at 240 μg per 100 g of flour. At the time of writing, this proposal is under review but not yet approved, although voluntary optional fortification is allowed in the UK. Serum and red blood cell folate levels among reproductive aged women in the USA have more than doubled since fortification of cereal grain products. In addition, preliminary analyses suggest an approximate 20% decline in NTD rates postfortification.

Although multivitamin supplements containing folic acid are of proven efficacy in preventing NTDs, their use requires a sustained behavior change by most women. Educational campaigns in the USA have increased awareness and knowledge about folic acid more than they have increased actual supplement use. In 2001, although 80% of childbearing aged women in the USA had heard of folic acid, only 29% took a daily vitamin containing folic acid, up only slightly from 25% in 1995. In the UK and in The Netherlands, public-awareness campaigns significantly increased both knowledge about, and use of, folic acid. For example, after a 1995 campaign in The Netherlands, use of folic acid for 'any period around conception' increased from 25% before the campaign to 54% after, and 'appropriate' use increased from 5 to 21%. Whether increases in supplement use can be sustained and whether levels of use plateau at a certain level is unknown. Supplement use is less common among younger women, less educated women, and those with lower incomes – populations that are often harder to reach through health education efforts.

Attempts to increase population consumption of folate-rich foods (e.g., fruits and vegetables) through programs like the 'Five-a-Day' campaign in the USA have generally resulted in only modest increases. One small study found that neither dietary advice nor folate-rich foods significantly increased women's red cell folate levels, whereas supplements and fortified foods did. This could result from the lower bioavailability of food folate and/or the challenges of making dietary behavior changes.

Folate-related Nutrients

As noted previously, the preventive effects of folate supplementation are more likely to result from its compensation for metabolic errors closely related to folate metabolism, rather than overt folate deficiency. Folate is involved in an intricate network of reactions that include many enzymes, nutrients, and other substrates that could affect NTD risk through a variety of mechanisms (**Figure 4**). The conversion of 5-methyl tetrahydrofolate to tetrahydrofolate produces the substrate necessary for the transmethylation of homocysteine to methionine; the enzyme methionine synthase (MS) catalyzes the reaction, and vitamin B_{12} is a cofactor for this enzyme (**Figure 4**). Homocysteine may also be metabolized to cystathionine via cystathionine β synthase (CBS), of which vitamin B_6 is a cofactor.

Hyperhomocysteinemia, which has been shown to cause NTDs in animal models, may reflect impaired metabolism of folate, B_{12}, B_6, or homocysteine, may reflect a defect in the enzymes 5,10-methylenetetrahydrofolate reductase (MTHFR), MS, or CBS, or may be harmful on its own. Several studies have reported an increased NTD risk associated with high homocysteine levels in humans, in the amniotic fluid of mothers carrying affected fetuses, in maternal serum during pregnancy and postpartum, and among children with NTDs, although results have not been entirely consistent.

Animal models indicate that methionine is also important to neural tube closure. Although the mechanism is unknown, methionine is necessary for protein synthesis and transmethylation reactions. Two population-based, case-control studies have reported a 30–50% reduction in risk for NTDs with increasing quartiles of methionine intake. Other studies have been inconsistent: some have demonstrated higher levels of methionine in the amniotic fluid of mothers carrying NTD-affected fetuses, whereas another study reported no difference in maternal serum methionine post-partum.

Several studies have indicated lower levels of vitamin B_{12} in the amniotic fluid and serum of women with affected pregnancies. Studies of methyl malonic acid (an indicator of B_{12} deficiency or a defect in B_{12} metabolism) and B_{12} carrier proteins also suggest a defect in B_{12} metabolism among case mothers. Other studies have not supported these differences or have been inconsistent in their findings depending on the timing of and tissues used for measurement. Although the results are mixed, they do not seem to provide strong evidence for a relation between B_{12} deficiency and NTD risk. Few studies have examined the association between vitamin B_6 and risk for NTDs;

significant differences have not been observed, but studies have involved only small sample sizes.

There is some evidence that polymorphisms in the genes for MTHFR, MS, and CBS contribute to altered levels of these nutrients and to NTD risk, but again, results have been mixed.

The individual contributions of each of these indicators of metabolic disturbances to NTD risk, and their interdependence, are unclear at this point. Folic acid supplementation might play a role in overcoming the negative effects of many of these disturbances. However, until the exact causes of NTDs are determined, the understanding of the implications of these findings and the way in which they should be used for prevention purposes remains limited.

Other Nutrients

At least 30–50% of NTDs are folate-resistant; therefore, other nutrients have been implicated as possible causes of these cases. The evidence for the role of vitamin A, zinc, and inositol in NTD etiology is reviewed here. Other nutrients – including vitamin C, riboflavin, iron, calcium, magnesium, selenium, lead, and copper – also may be associated with increased NTD risk, but evidence is very limited.

The finding that isotretinoin, a synthetic derivative of vitamin A, is teratogenic led to investigations of whether naturally derived vitamin A is associated with increased risk for NTD-affected pregnancy. Most studies have found no evidence of an association of high vitamin A levels with NTD risk, based on intake from foods and supplements and on serum samples taken during the first trimester of pregnancy or postpartum. The safe level of vitamin A intake is still debatable, in part because of its potential association with other types of malformations and its association with other negative health outcomes. The effects of deficient vitamin A intake on NTD have not been explored.

Zinc is important to cellular growth and differentiation and to folate absorption. Various experimental studies in animals and case reports of women with specific disturbances of zinc metabolism have suggested an association between zinc and increased risk for NTD-affected pregnancy. Results for analytic studies of zinc and NTDs in humans are inconsistent. Postpartum studies have reported NTD case mothers to have levels of zinc in the hair that were higher, lower, or the same as those in control mothers, higher levels in toenails, and the same as control mothers in serum levels. Studies of pregnant women have reported that mothers with NTD-affected pregnancies have higher levels of zinc in serum, lower levels in white blood cells, and lower dietary intakes, whereas other studies have reported no difference in serum or amniotic fluid. Despite the number of studies, inference is limited by the wide variability in zinc measurement and by the limited understanding of the relationships between these measures.

Inositol is a 6-carbon polycyclic alcohol that is present in all cells, with especially high levels in nerve cells. It protects against NTDs in the curly-tailed mouse, a folate-resistant NTD model, leading some researchers to recommend supplementation with a combination of folate and inositol. However, the applicability of these findings to humans has not been studied.

Limitations

In general, the proposed associations reviewed here are biologically plausible and supported by animal models. Unfortunately, systematically testing these associations in humans is difficult, and most existing evidence is clinical or experimental and speculative. Sample sizes for many of the studies have been small (often involving < 20 case-patients). Study population characteristics and selection methods have been highly variable; very few have been population-based. Few studies have adjusted for potential confounding or effect modification by other maternal characteristics, including status with respect to other nutritional factors or individual nutrients.

The ability to measure maternal nutritional status is also limited. Because pregnancy is a time of tremendous metabolic change, the ability of measurements taken later in pregnancy to represent the status during the time of neural tube closure is questionable. For example, the fetal contribution to levels of nutrients in amniotic fluid is uncertain and may vary by timing of measurement and by nutrient. Also, many women begin taking multivitamin/mineral supplements soon after the time of neural tube closure.

The cause of observed variability in nutrient levels is usually unknown, limiting the understanding of how to use such information for preventive measures. Nutrient metabolism is highly complex, and many reactions are interdependent. For example, depending on their cause, high homocysteine levels could be remedied with supplementation of B_{12}, methionine, folate, or B_6; low calcium intake may result in increased bioavailability of lead, reduced bioavailability of vitamin B_{12}, and, accordingly, reduced methionine synthase activity; and calcium, fiber, protein, and methionine are all important to zinc bioavailability. Therefore, it is critical that future studies examine interrelationships, including nutrient–nutrient interactions as well as gene–nutrient interactions, whenever possible.

Other Nutritional Factors

Evidence for a link between tea and NTDs is weak. A study in England and Wales showed an excess of anencephaly among women drinking three or more cups of tea per day, but no dose–response effect was observed, and the relation was observed only in regions with higher NTD prevalence rates. Furthermore, a study in Boston, Philadelphia, and Toronto found no elevated risk for NTDs associated with tea drinking. Nor is the link between coffee and NTDs strong. In the Boston study, no significant differences in coffee consumption were observed between the mothers of NTD-affected infants and control mothers. Also, no association with decaffeinated coffee or cola drinks has been found. A Finnish study found no association between central nervous system defects and coffee intake. Although there are case reports of NTDs occurring among infants of mothers who consumed significant amounts of alcohol during pregnancy, epidemiologic studies have provided little evidence for a link between NTDs and alcohol, especially for mothers consuming minimal or moderate amounts.

Some food contaminants have been proposed to be NTD teratogens. Potato blight, caused by a fungus (*Phytophthora infestans*) that rots potatoes, was hypothesized to cause NTDs in the early 1970s, although most evidence does not support this hypothesis. Fumonisin, a potent mycotoxin prevalent in corn and a carcinogen and cause of various animal diseases has been suggested as a possible cause of high NTD rates among persons living along the Texas–Mexico border who commonly consume corn products. Recent research suggests that fumonism may be involved in NTD formation by inhibiting folate uptake by cells. Also, fumonisins have been shown to produce NTDs in mouse whole embryo cultures; folic acid has ameliorated this teratogenic effect. More research is needed to establish the link between fumonisin consumption and NTD risk in humans.

About one-fourth of childbearing age women in the USA and almost as many such women in Canada, the UK, and Continental Europe, are obese, and obesity prevalence is increasing rapidly. Studies consistently find a twofold elevated risk for NTDs among children of obese women compared with mothers of average weight, although risk estimates vary with the body mass index cutoffs and referent groups used (Figure 5). Most studies found a higher risk estimate for spina bifida than anencephaly and found no risk elevation for encephalocele. Reasons for the increased risk are unknown. Potential explanations include teratogenic metabolic abnormalities associated with obesity including elevated levels of glucose, insulin, uric acid, and endogenous estrogens. Obese women may be more likely to be nutritionally deficient because of a self-imposed diet restriction, poor choice of foods, or increased requirement for folate or other nutrients. More studies are needed to help elucidate the mechanism. The population burden of NTDs associated with obesity is substantial and increasing.

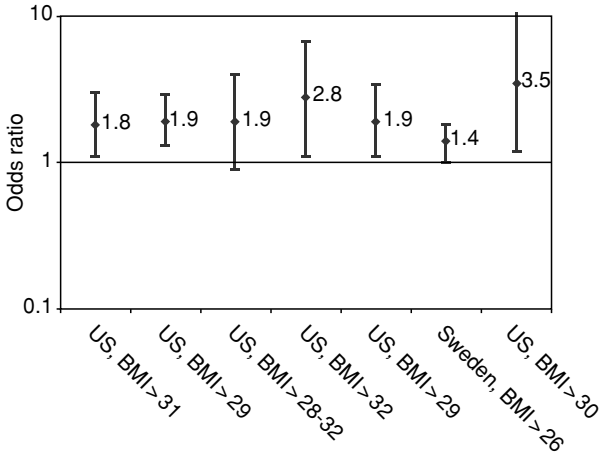

Figure 5 Risk of neural tube defects in offspring of obese women, odds ratios, and 95% confidence intervals. All studies except one (Sweden) were performed in the USA. The body mass index (BMI) used to define obesity varied (see *x*-axis). The referent group was average-weight or nonobese women.

Conclusion

There is still much to be learned about the complex link between nutrition and birth defects. Evidence is emerging regarding the role of nutritional factors in the etiology of birth defects other than NTDs (e.g., heart defects and orofacial clefts). Despite the challenges in studying nutritional causes of birth defects, the prevention potential is substantial and worthy of pursuit.

See also: **Cobalamins**: Properties and Determination; **Folic Acid**: Properties and Determination; Physiology; **Food Fortification**; **Nutritional Assessment**: Importance of Measuring Nutritional Status; Anthropometry and Clinical Examination; Biochemical Tests for Vitamins and Minerals; Functional Tests; **Obesity**: Epidemiology; **Retinol**: Properties and Determination; Physiology; **Vitamins**: Determination; **Zinc**: Properties and Determination; Physiology; Deficiency

Further Reading

Anonymous (2000) Folic acid and the prevention of disease. Report of the Committee on Medical Aspects of Food

and Nutrition Policy. *Reports on Health & Social Subjects* 50: i–xv, 1–101.
Berry RJ, Li Z, Erickson JD *et al*. (1999) Prevention of neural-tube defects with folic acid in China. *New England Journal of Medicine* 341: 1485–1490.
Botto LD, Moore CA, Khoury MJ and Erickson JD (1999) Neural-tube defects. *New England Journal of Medicine* 341: 1509–1519.
Centers for Disease Control and Prevention (1992) *Recommendations for the Use of Folic Acid to Reduce the Number of Cases of Spina Bifida and Other Neural Tube Defects. MMWR 41 (No. RR-14)*, pp. 1–7.
Cogswell ME, Perry GS, Schieve LA and Dietz WH (2001) Obesity in women of childbearing age: risks, prevention, and treatment. *Primary Care Update Ob/Gyns* 8: 89–105.
Elwood M, Elwood H and Little J (1992) Diet. In: Elwood JM, Little J and Elwood JH (eds) *Epidemiology and Control of Neural Tube Defects*, pp. 521–603. Oxford: Oxford University Press.
Institute of Medicine (1998) *Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B_6, Folate, Vitamin B_{12}, Pantothenic Acid, Biotin, and Choline: A Report of the Standing Committee on the Scientific Evaluation of Dietary Reference Intakes and its Panel on Folate, Other B Vitamins, and Choline and Subcommittee on Upper Reference Levels of Nutrients*. Washington, DC: National Academy Press (see http://www.nap.edu/openbook/ 0309065542/html).
International Centre for Birth Defects and European Registration of Congenital Anomalies (1998) *World Atlas of Birth Defects*, pp. 20–31. Geneva: World Health Organization.
International Clearinghouse for Birth Defects Monitoring Systems (1998) *Annual Report 1998 with Data for 1996*, pp. 100–101. Rome: International Centre for Birth Defects.
Kirke PN, Molloy AM, Daly LE *et al*. (1993) Maternal plasma folate and vitamin B_{12} are independent risk factors for neural tube defects. *Quarterly Journal of Medicine* 86: 703–708.
Mills JL, Scott JM, Kirke PN *et al*. (1996) Homocysteine and neural tube defects. *Journal of Nutrition* 126: 756S–760S.
Refsum H (2001) Folate, vitamin B_{12} and homocysteine in relation to birth defects and pregnancy outcome. *British Journal of Nutrition* 85 (supplement 2): S109–S113.
Rosenquist TH and Finnell RH (2001) Genes, folate and homocysteine in embryonic development. *Proceedings of the Nutrition Society* 60: 53–61.
Scott JM, Weir DG and Kirke PN (1995) Folate and neural tube defects. In: Bailey LB (ed.) *Folate in Health and Disease*, pp. 329–361. New York: Marcel Dekker.
Sells CJ and Hall JG (eds) (1998) *Neural Tube Defects.-Mental Retardation and Developmental Disabilities Research Reviews* 4(4). New York: Wiley-Liss.
Wald NJ (1994) Folic acid and neural tube defects: The current evidence and implications for prevention. CIBA. *Foundation Symposium* 181: 192–208.

Preeclampsia and Diet

E Abalos, G Carroli and L Campodónico, Centro Rosarino de Estudios Perinatales, Rosario, Argentina
J Villar, World Health Organization, Geneva, Switzerland

Background

Hypertensive disorders during pregnancy are one of the main causes of maternal death worldwide, most of these deaths being attributed to eclampsia. Hypertensive disorders occur in 6–8% of all pregnancies, contributing significantly to stillbirths and neonatal morbidity and mortality. Babies are also at increased risk of intrauterine growth restriction, low birth weight and preterm delivery. Pregnant women with hypertension, either newly diagnosed or pre-existing, are prone to the development of potentially lethal complications, notably abruptio placentae, disseminated intravascular coagulation, cerebral hemorrhage, pulmonary edema, hepatic failure, and acute renal failure. The etiology of hypertensive disorders related to pregnancy, particularly preeclampsia, remains unknown. The most important consideration in the classification of the disease is differentiating hypertensive disorders that antedate pregnancy from those that are pregnancy-specific, from which the more ominous are preeclampsia and eclampsia. Preeclampsia is a pregnancy-specific syndrome of reduced organ perfusion secondary to vasospasm and activation of the coagulation cascade. Although our understanding of this syndrome has increased, the criteria used to identify the disorder remain a subject of confusion and controversy. In chronic hypertension, elevated blood pressure is the primary pathophysiologic feature, whereas in preeclampsia, increased blood pressure is important primarily as a sign of the underlying disorder. As might be expected, the impact of the two conditions on mother and fetus is different, as is their management **Table 1**.

Classification

There are controversies about the definition of hypertensive disorders during pregnancy, and several classifications have been suggested. Recently, the National High Blood Pressure Education Program Working Group on High Blood Pressure in Pregnancy updated the 1990 report, and classified the hypertensive disorders during pregnancy

Table 1 Effectiveness of nutritional interventions in hypertension during pregnancy and preeclampsia

Intervention	*Hypertension during pregnancy*		*Preeclampsia*	
	Practice	*Research*	*Practice*	*Research*
Nutritional advice	No evidence		No effect RR = 0.89 (0.42–1.88)	
Balanced protein (<25%) / energy	No evidence		No effect RR = 1.20 (0.77–1.89)	
Isocaloric balanced protein (<25% of total energy)	No evidence		No effect RR = 1.00 (0.57–1.75)	
Energy/protein restriction for high PI or high weight gain	No effect RR = 0.97 (0.75–1.26)		No effect RR = 1.13 (0.59–2.18)	
Salt restriction	No effect RR = 0.97 (0.49–1.94)		No effect RR = 1.11 (0.46–2.66)	
Calcium	Possibly beneficial for women at high risk RR = 0.45 (0.31–0.66) and with low baseline intake RR = 0.49 (0.38–0.62)	RCT in progress	Possibly beneficial for women at high risk RR = 0.21 (0.11–0.39) and with low baseline intake RR = 0.32 (0.21–0.49)	RCT in progress
Iron and folate	No effect RR = 1.15 (0.41–3.18)		No evidence	
Folate	No effect RR = 1.26 (0.90–1.76)		No evidence	
Magnesium	No evidence	Needed	No effect RR = 0.87 (0.57–1.32)	Needed
Fish oil	No effect RR = 0.96 (0.86–1.07)		Possibly beneficial (data from low quality studies) RR = 0.70 (0.55–0.90)	
Zinc	No effect RR = 0.87 (0.65–1.15)	Needed	No evidence	Needed
Antioxidants	No evidence	Needed	Possibly beneficial for vitamins C and E (data from one RCT) RR = 0.46 (0.24–0.91)	Needed

as (1) chronic hypertension defined as hypertension observable before pregnancy, or diagnosed before the 20th week of gestation; (2) preeclampsia, which is a pregnancy-specific syndrome usually occurring after 20 weeks' gestation, determined by hypertension with proteinuria; (3) preeclampsia superimposed on chronic hypertension; and (4) pregnancy-induced hypertension or gestational hypertension, which is transient hypertension detected for the first time after midpregnancy if preeclampsia is not present at the time of delivery and blood pressure returns to normal by 12 weeks post partum (a retrospective diagnosis). The system suggested by the International Society for the Study of Hypertension in Pregnancy (ISSHP) defines hypertension as a diastolic blood pressure of 90 mmHg or above on two consecutive occasions at least 4 h apart, or a single diastolic blood pressure of 110 mmHg or more. The definition of preeclampsia has the same criteria for high blood pressure, but with the addition of significant proteinuria, usually at least 300 mg per 24 h or 1+ on dipsticks.

Pathophysiology of Preeclampsia

Preeclampsia is a syndrome with both fetal and maternal manifestations. The maternal disease is characterized by vasospasm, activation of the coagulation system, and perturbations in many humoral and autacoid systems related to volume and blood pressure control. The pathologic changes in this disorder are primarily ischemic in nature and affect the placenta, kidney, liver, and brain. Of importance, and distinguishing preeclampsia from chronic or gestational hypertension, is that preeclampsia is more than hypertension; it is a systemic syndrome, and several of its 'nonhypertensive' complications can be life-threatening, even when blood pressure elevations are quite mild.

The cause of preeclampsia is not known. Many consider the placenta as the pathogenic focus for all manifestations of preeclampsia because the delivery of both the baby and the placenta is the only definitive cure of this disease. There is no disease without the placenta. Thus, research has focused on the changes in the maternal blood vessels that supply

blood to the placenta. Failure of the spiral arteries to remodel is postulated as the morphologic basis for decreased placental perfusion in preeclampsia, which ultimately may lead to early placental hypoxia. Oxidative stress and inflammatory-like responses may also be important in the pathophysiology of preeclampsia.

Research on how alterations in the immune response at the maternal interface might lead to preeclampsia addresses the link between placenta and maternal disease. A nonclassical human leukocyte antigen (HLA), HLA G, is expressed in normal placental tissue and may play a role in modulating the maternal immune response to the immunologically foreign placenta. Placental tissue from preeclamptic pregnancies may express fewer or different HLA G proteins, resulting in a breakdown of maternal tolerance to the placenta. Additional evidence for alterations in immunity in pathogenesis includes the disease prominence in nulliparous gestations with subsequent normal pregnancies, a decreased prevalence after heterologous blood transfusions, long cohabitation before successful conception, and observed pathologic changes in the placental vasculature in preeclampsia that resemble allograft rejection. Finally, there are increased levels of inflammatory cytokines in the placenta and maternal circulation, as well as evidence of increased 'natural killer' cells and neutrophil activation in pre-eclampsia.

The mechanisms underlying vasoconstriction and altered vascular reactivity in preeclampsia remain obscure. Research has focused on changes in the ratio of vasodilative and vasoconstrictive prostanoids, since there is evidence suggesting decrements and increments in the production of prostacyclin and thromboxane, respectively. More recently, investigators have postulated that the vasoconstrictive potential of pressor substances (e.g., Angiotensin II and endothelin) is magnified in preeclampsia as a consequence of a decreased activity of nitric oxide (NO) syntheses and decreased production of NO-dependent or -independent endothelium relaxing factor (EDRF). Also under investigation is the role of endothelial cells (the site of prostanoid, endothelin, and EDRF production), which, in preeclampsia, may be dysfunctional, owing perhaps to inflammatory cytokines (e.g., tumor necrosis factor α) and increased oxidative stress. Other systems postulated to play a role in pre-eclamptic hypertension are the sympathetic nervous system, calciotrophic hormones, insulin, and magnesium metabolism.

Finally, some nutritional factor deficiencies have been postulated as playing a role in the pathogenesis of preeclampsia. This chapter discusses their possible role on the hypertensive disorders of pregnancy.

Possible Role of Nutrition in the Pathophysiology of Preeclampsia

Nutrient supplements in addition to food sources of nutrients are provided to populations either to increase the intake among those with a deficiency (in order to prevent or treat functional outcomes related to such a deficit) or to obtain a pharmacological, perhaps nonnutritional effect, among individuals with an adequate intake of the nutrient in question.

Epidemiological observations have long suggested a role for nutritional deficiencies in preeclampsia (i.e., calcium, proteins, vitamins, etc.). However, intervention evaluations have failed to confirm such promising observations. We will describe here the evidence from randomized-controlled trials that support the relationship between different nutrients and preeclampsia.

Calcium

There is considerable evidence linking calcium intake and hypertension during pregnancy from observational and experimental studies. However, there is still no satisfactory explanation for the mechanisms involved in the calcium-mediated effect on blood pressure prevention. It was postulated that parathyroid hormone could be involved in this relationship. Demonstrated alterations in extracellular calcium homeostasis in preeclampsia include hypocalciuria and decrease serum levels of calcitriol. Increased parathyroid hormone (PTH) and decreased plasma ionized calcium concentration have not been consistently observed. Also, consistent abnormalities of intracellular calcium metabolism have been described in pre-eclamptic women, such as increased intracellular free calcium concentration in platelets and lymphocytes. Increases in intracellular free calcium concentration in circulating cells are hypothesized to result from fluctuation in hormones or vasoactive substances that cause similar alteration in vascular smooth muscle. Pregnancy is a state of high calcium requirements as a result of fetal demands while maternal adaptive mechanisms are partially inhibited. These phenomena lead to the hyperparathyroid state of pregnancy. An increase in parathyroid hormone serum levels would involve an increase in free intracellular calcium. Thus, the concentration of intracellular free calcium in vascular smooth muscle cells determines the degree of tension and is the trigger for muscular contraction. So, the vasoconstrictive effect, with a rise in blood pressure, results from an increase in vascular smooth muscle tension.

Antioxidant Agents

An additional role for nutrition in the genesis of preeclampsia could be nutritional factors that strengthen oxidative stress, leading to preeclampsia. A nutritional factor could be the deficiency of antioxidant intake, specifically vitamins C and E. Vitamin C is central for the neutralization of both water-soluble and lipid-soluble free radicals. As a water-soluble molecule, its ability to neutralize free radicals in the aqueous compartment is clear. Also, ascorbate is not made in humans and must come from the diet. Vitamin E, a potent antioxidant, has been suggested to play a role in preventing preeclampsia. Vitamin E is not usually reduced by dietary deficiencies, and reductions are more likely due to consumption.

Other Nutrients

Nutritional factors other than antioxidants can also contribute to oxidative stress. Hyperhomocysteinemia can be a result of dietary deficiencies. Hyperhomocysteinemia is a risk factor for preeclampsia and atherosclerosis, and endothelial function is said to be altered, at least in part, by the genesis of oxidative stress. Vitamins B_6 and B_{12} and folic acid are involved at different steps in the metabolic pathway for removing or recycling homocysteine to methionine. Dietary deficiencies of any of these micronutrients can increase circulating homocysteine. Preeclampsia is characterized by increased triglycerides that favor the formation of small, dense low-density lipoprotein (LDL). This lipoprotein variant has increased access to the subendothelial space, where it is sequestered from bloodborne antioxidants. The relevant role of triglycerides in the genesis of preeclampsia is indicated by the fact that they are increased long before clinically evident disease. Similarly, free fatty acids are increased in preeclampsia, and this increment was observed months before the diagnosis. Recent studies indicate that this effect may be secondary to altered copper binding by albumin to which large amounts of free fatty acids are bound. Unbound copper is a potent stimulator of free radical formation. Normally, this effect of copper is prevented by protein binding (quantitatively, primarily to albumin). However, with fatty acid binding, albumin binds copper differently. In this configuration, copper bound to albumin maintains its ability to participate in redox reactions. Thus, it appears that increased free fatty acids can also contribute to oxidative stress.

All of these nutritional alterations may be agreeable to dietary modification raising the possibility of nutritional prophylaxis.

Nutritional Interventions and Hypertensive Disorders of Pregnancy

Prevention

The ability to prevent hypertensive disorders of pregnancy is limited by a lack of knowledge of its underlying etiology. Prevention is focused on identifying women at higher risk of developing pregnancy-induced hypertension or preeclampsia during pregnancy, followed by close clinical and laboratory monitoring to recognize the clinical symptoms of the disease in its early stages. These women and their pregnancies can then be selected for more intensive monitoring or delivery. Although these measures do not prevent the disease, they may be helpful for preventing some adverse maternal and fetal sequelae.

As part of many other nonpharmacological interventions, some dietary interventions were proposed to prevent the development of pregnancy-induced hypertension and preeclampsia.

Nutritional advice in pregnancy Literature was reviewed in order to assess the effects of advising pregnant women to increase their energy and protein intakes on the outcome of pregnancy, and maternal and fetal/infant morbidity and mortality. Nutritional advice appears to be effective in increasing pregnant women's energy and protein intake, but the implications for fetal, infant, or maternal health cannot be judged from the available evidence. Preeclampsia prevention was assessed only in one small trial involving 136 women with no beneficial effects [relative risk (RR): 0.89, 95% confidence intervals (CI) 0.42–1.88].

Protein/energy supplementation The effect of balanced protein/energy supplements for pregnant women on gestational weight gain and pregnancy outcomes was assessed on a Cochrane systematic review. Preeclampsia prevention was evaluated in three trials involving 516 women, with no significant beneficial effects (RR: 1.20; 95% CI: 0.77–1.89). However, these trials had methodological flaws, alternate allocation, and large lost-to-follow-up for the main outcomes, so results should be taken cautiously. In another Cochrane systematic review, only one trial involving 782 women evaluated preeclampsia prevention when isocaloric balanced protein/energy supplements were given to underweight pregnant women, showing no effect (RR: 1.00; 95% CI: 0.57–1.75).

Energy/protein restriction for obese pregnant women Excessive weight gain during pregnancy

has long been recognized as a clinical sign of edema and impending preeclampsia. Epidemiological studies suggest that a high maternal weight is positively associated with the risk of preeclampsia.

Energy/protein restriction for high weight-for-height or weight gain during pregnancy was assessed in a Cochrane systematic review. Preeclampsia was evaluated in two trials (284 women) showing no reduction in the risk of occurrence (RR: 1.13; 95% CI: 0.59–2.18), and the same as for pregnancy-induced hypertension (three trials, 384 women; RR: 0.97; 95% CI: 0.75–1.26). The limited evidence available suggests that protein/energy restriction of pregnant women who are overweight or show a high weight gain is unlikely to be beneficial and may be harmful to the developing fetus. Although weight reduction may be helpful in reducing or preventing high blood pressure in nonpregnant women, it is not recommended during pregnancy, because there is no effect on preventing preeclampsia, even in obese women. Clinicians frequently ask pregnant women to restrict their food intake in an attempt to prevent preeclampsia, despite the absence of evidence that such advice is beneficial.

Salt restriction Even in an early phase of pregnancy, marked hemodynamic changes occur, including a fall in vascular resistance and blood pressure and a rise in cardiac output. To compensate for the increased intravascular capacity, the kidney retains more sodium and water. Apparently, the set point of sodium homeostasis shifts to a higher level at the expense of an expansion of extracellular volume. In nonpregnant women, a strong positive association of sodium intake with blood pressure has been established, but the relationship between sodium intake and blood pressure in human pregnancy remains obscure up to date. For decades, a low salt diet has been often recommended as treatment for edema, with the idea that restricting salt intake would treat, and also prevent, preeclampsia. Recently, this practice was questioned, and even a high sodium intake was proposed for preeclampsia treatment and prevention.

Concerns about the effect of a low sodium diet during pregnancy on maternal nutritional status led researchers to investigate whether such changes could alter other nutrient intakes. It was shown that the reduction in sodium intake also caused a significant reduction in the intake of energy, protein, carbohydrates, fat, calcium, zinc, magnesium, iron, and cholesterol. Even though women are no longer advised by many clinicians to alter their salt intake during pregnancy, this is still current practice in many other settings around the world.

A recently published Cochrane systematic review evaluates the effect of the advice about low dietary salt intake during pregnancy. The review includes two trials with data reported for 603 women. Both compared nutritional advice to restrict dietary salt with advice to continue a normal diet. Women with established preeclampsia were not enrolled, so this review provides no information about the effects of advice to restrict salt intake for the treatment of preeclampsia. No effect was found in preventing pre--eclampsia (RR: 1.11; 95% CI: 0.46–2.66) or pregnancy-induced hypertension (RR: 0.97; 95% CI: 0.49–1.94). Women's preferences were not reported, but authors presumed that a low-salt diet was not very palatable and was therefore difficult to follow.

Calcium supplementation A role for altered calcium metabolism in the pathogenesis of preeclampsia is suggested by epidemiological evidence linking low dietary levels of calcium with increased incidence of the disease. In agreement with these observations, several modifications in calcium metabolism have been observed in pre-eclamptic women and in calcium-supplemented mothers.

A Cochrane systematic review of calcium supplementation during pregnancy has been published. Authors pre-specified comparison groups taking into account women's risk of hypertensive disorders of pregnancy (low versus increased) and women's baseline dietary calcium intake (low, < 900 mg per day, versus adequate, ≥ 900 mg per day).

High blood pressure with or without proteinuria was evaluated in nine trials, involving 6604 women. Overall, there is less high blood pressure with calcium supplementation (RR 0.81; 95% CI: 0.74–0.89), but there is a variation in the magnitude of the effect across the subgroups. The effect was considerably greater in women at high risk of developing hypertension (four trials, 327 women: RR: 0.45; 95% CI: 0.31–0.66) than in those at low risk (11 trials, 6894 women: RR: 0.68; 95% CI: 0.57–0.81). Taking into account women's calcium intake, the effect was also greater in those with low baseline dietary calcium (five trials, 1582 women: RR: 0.49; 95% CI: 0.38–0.62) than in those with adequate calcium intake (four trials, 5022 women: RR: 0.90; 95% CI: 0.81–0.99).

There is a reduction in the risk of preeclampsia, evaluated on 10 trials involving 6864 women (RR: 0.70; 95% CI: 0.58–0.83). When predefined subgroups are considered, there is a significant reduction in women with low baseline dietary calcium intake (six trials, 1842 women: RR: 0.32; 95% CI: 0.21–0.49), but not in those with adequate calcium intake (four trials, 5022 women: RR: 0.86; 95% CI: 0.71–1.05).

Preeclampsia was considerably reduced in women at high risk of hypertension (four trials, 557 women: RR: 0.22; 95% CI: 0.11–0.43), and less consistently in those at low risk of hypertension (six trials, 6307 women: RR: 0.79; 95% CI: 0.65–0.94).

The results from the largest trial conducted by NIH, which studied low-risk women with adequate baseline calcium diet, and in whom all women in both groups received low-dose calcium supplementation as part of their routine antenatal care, showed no significant effect on hypertension and preeclampsia. Based on this, authorities from developed countries where adequate dietary calcium intake is common discourage the use of routine calcium supplementation during pregnancy. Evidence from this review supports the idea that calcium supplementation might benefit women at high risk of gestational hypertension and women with a low dietary calcium intake at risk of developing preeclampsia. However, further methodologically sound randomized controlled trials with adequate sample size are needed to confirm or reject this hypothesis.

Iron and folate supplementation Numerous trials involving various populations of pregnant women with normal hemoglobin levels have evaluated the effects of iron and/or folate supplementation on several outcomes, some of them including hypertensive disorders of pregnancy. A Cochrane systematic review of two trials involving 87 women with normal hemoglobin levels, in which iron and folic acid were compared with no treatment, showed no effect on the occurrence of gestational hypertension (RR: 1.15; 95% CI: 0.41–3.18). Preeclampsia was not evaluated. Another Cochrane review that included two trials involving 696 pregnant women already receiving iron, in which women were allocated to receive folic acid or no treatment/placebo, showed no effect on prevention of gestational hypertension either (RR: 1.26; 95% CI: 0.90–1.76).

Although evidence shows that iron and folate supplementation is not effective in preventing hypertensive disorders during pregnancy, they should be prescribed for other established beneficial effects on pregnancy such as prevention of anemia.

Magnesium supplementation Magnesium is one of the essential minerals needed by humans in relatively large amounts. Magnesium works with many enzymes in regulating body temperature and synthesizing proteins as well as maintaining electrical potentials in nerves and muscle membranes. Magnesium occurs widely in many foods; dairy products, breads and cereals, vegetables, and meats are all good sources. It is therefore not surprising that genuine clinical magnesium deficiency has never been reported to occur in healthy individuals who eat standard diets. However, dietary intake studies during pregnancy consistently demonstrate that many women, especially those from disadvantaged backgrounds, have intakes of magnesium below recommended levels. Observational studies based on medical records have reported that magnesium supplementation during pregnancy was associated with a reduced risk of fetal growth retardation and preeclampsia, and that magnesium intake was associated with increased birth weight. Stimulated by these encouraging epidemiological studies, randomized clinical trials have been undertaken to evaluate the potential benefits of magnesium supplementation during pregnancy on pregnancy and neonatal outcomes.

A Cochrane systematic review of these randomized-controlled trials was carried out in order to assess the effects of magnesium supplementation during normal or high-risk pregnancies on maternal, neonatal, and pediatric outcomes. Results from two trials (474 women) showed no apparent effect of magnesium treatment on prevention of preeclampsia (RR: 0.87; 95% CI: 0.57–1.32). However, these results may have been confounded by the fact that in the largest trial all women (both magnesium-supplemented and placebo groups) received a multivitamin and mineral preparation containing low doses of magnesium. Trials also have a poor methodological quality, especially related to concealment of allocation, which could give biased results. Authors conclude that dietary magnesium supplementation of pregnant women cannot be recommended for routine clinical practice because of the poor methodological quality of the current evidence.

Fish oil supplementation Studies of nonpregnant subjects suggest that fish oil, rich in long-chain *n*-3 fatty acids, has a moderate effect on blood pressure in normotensive as well as hypertensive individuals. A meta-analysis of controlled clinical trials of the effect of fish oil on blood pressure has demonstrated a significant reduction in systolic and diastolic blood pressure in untreated hypertensive nonpregnant individuals, but found no other significant effect on normotensives. Fish oil has been shown to interfere with prostaglandin metabolism, and its effect on blood pressure has often been assumed to be due to such interference. Epidemiological studies suggested that marine diets could have a preventive effect on early delivery and hypertensive disorders of pregnancy.

Fish oil supplementation during pregnancy was evaluated in a systematic review of two trials (5135 women), showing no effect on pregnancy-induced hypertension (two trials, 5135 women; RR: 0.96;

95% CI: 0.86–1.07). There was a statistically significant, but modest, reduction in the rate of preeclampsia (RR: 0.70; 95% CI: 0.55–0.90). However, this reduction is strongly influenced by a large nonrandomized trial conducted in 1942, in which vitamins and minerals were given to women in addition to fish oil. Another seven trials of fish oil supplementation involving more than 2000 women have been published recently, none of which demonstrates any differences in the incidence of hypertension and preeclampsia between groups. Based on current evidence, fish oil supplementation is not recommended during pregnancy.

Zinc supplementation Zinc is thought to play an important role in many biological functions, including protein synthesis and nucleic acid metabolism. There are controversies in the literature in demonstrating the relationship between low serum zinc levels with abnormalities on pregnancy outcomes such as pregnancy-induced hypertension, prolonged labor, post partum hemorrhage, preterm or postterm pregnancies, small-for-gestational-age babies, or poor perinatal outcomes.

The role of routine zinc supplementation during pregnancy on outcomes for both mother and newborn was assessed in a Cochrane systematic review. Routine zinc supplementation in pregnancy had no detectable effect on gestational hypertension (four trials, 1962 women; RR: 0.87; 95% CI: 0.65–1.15). However, there appears to be inconsistency between trials regarding the effects from other pregnancy outcomes. This may be related to varied population characteristics of women recruited in the various trials, as some included normal pregnant women with no systemic illness, other studies specifically selected women at high risk of low zinc status, and in one study, participants were selected on the basis of proven low plasma zinc levels. There is at present no evidence of overall benefit from routine as opposed to selective zinc supplementation in pregnancy on pregnancy-induced hypertension or preeclampsia.

Vitamin supplementation In terms of vitamins A, E, and C, an oxidant/antioxidant imbalance has been suggested among the pathogenic factors involved in preeclampsia. As vitamin E is one of the most important antioxidants in body components, its levels and their relation with circulating levels of lipids peroxides on pre-eclamptic women have been intensively studied in recent years. As with other antioxidants, several studies found decreased vitamin E levels in serum from women with gestational hypertension and preeclampsia compared with controls. However, these findings could not be demonstrated in other studies. Increased ascorbate radical formation and ascorbate depletion were also found in plasma from women with preeclampsia. Recently, a randomized-controlled trial involving 283 women at high risk of developing preeclampsia was conducted. Women were randomly assigned to vitamin E and vitamin C supplements, or placebo at 16–22 weeks of gestation. Authors found a significant reduction in the risk of developing preeclampsia in the vitamin-supplemented group compared with controls (RR: 0.46; 95% CI: 0.24–0.91). The authors concluded that supplementation with vitamins C and E may be beneficial for preventing preeclampsia in women at increased risk of the disease. However, these findings need to be evaluated further in different settings and populations, as well as in low-risk women.

The role of vitamin A in pregnancy-induced hypertension and preeclampsia is a subject of controversy. It was proposed as a chain-breaking antioxidant on free radical cascade. Some studies found significantly reduced serum vitamin A levels in pre-eclamptic and eclamptic women when compared with those found in healthy women in the third trimester. To date, no trials have been published that assess the effect of vitamin A supplementation on pregnancy-induced hypertension or preeclampsia. A double-blind, cluster randomized trial of low dose supplementation with vitamin A or β-carotene carried out in Nepal in 44 646 married women showed a 40% reduction in maternal mortality related to pregnancy in vitamin A-supplemented women. However, differences in cause of deaths, including preeclampsia/eclampsia, could not be reliably distinguished between supplemented and placebo groups. The future prospect of giving vitamin A supplements for the prophylaxis/management of pregnancy-induced hypertension and preeclampsia needs to be evaluated further before it is recommended.

Treatment

The objectives of treatment for established preeclampsia, or pregnancy-induced hypertension are to prevent eclampsia as well as other severe maternal complications. Close maternal evaluation is aimed at observing the progression of the condition, both to prevent maternal complications and to determine whether fetal well-being can be assessed. As this disorder is often completely reversible and usually begins to abate with delivery, an imbalance between the mother's condition and the chance of the fetus surviving without any significant neonatal complications *in utero* or in the nursery must be continuously evaluated. Even though the only definitive treatment of preeclampsia is delivery, several nonpharmacological approaches were proposed as part of an overall

strategy of management of the disease to achieve these goals.

Unfortunately, there is no information from randomized controlled trials related to dietary approaches on the management of the disease in its mild to moderate stage, when conservative management is decided on.

Pre-existing (Chronic) Hypertension

Mild and uncomplicated chronic hypertension during pregnancy has a better prognosis than preeclampsia. However, there is an increased risk of superimposed preeclampsia and possible complications if pre-existing renal disease or systemic illness is present. The primary aim of therapy, if necessary, is to prevent cerebrovascular complications and to avoid the progression to superimposed preeclampsia with worse prognosis. Nonpharmacological management of this condition during pregnancy remains controversial.

In a published review of management of mild-chronic hypertension during pregnancy, no trials were found that compared non-pharmacological interventions with either pharmacological agents or no intervention in pregnant women. This comprehensive search identified 50 randomized controlled trials, but they involved either normotensive women or women with a history of preeclampsia. For the management of established chronic hypertension during pregnancy, no relevant evidence could be located to assess the effects of non-pharmacological interventions, such as limiting activity, diet modifications or stress reduction.

Weight reduction during pregnancy, even in obese women, is generally not recommended to improve pregnancy outcomes. As weight reduction may be helpful in reducing blood pressure in nonpregnant women, for obese hypertensive women planning a pregnancy, weight reduction before conception is advisable. Even though obesity may be a risk factor for superimposed preeclampsia, there is no evidence that limiting weight gain during pregnancy reduces its occurrence.

Pregnant women with hypertension have a lower plasma volume than do normotensive women, and some studies suggest that the severity of hypertension correlates with the degree of plasma volume concentration. For this reason, sodium restriction is generally not recommended during pregnancy for the reduction of blood pressure. In addition, an increase in plasma volume concentration is a risk factor for intrauterine growth restriction. If, however, a pregnant woman with chronic hypertension is known to have salt-sensitive hypertension and has been treated successfully with low salt diet before pregnancy, it is reasonable to continue some sodium restriction for blood pressure control during pregnancy, but not for preventing superimposed preeclampsia.

High alcohol intake is related to hypertension in nonpregnant women but is not associated with an increased risk for gestational hypertension, preeclampsia, or eclampsia. There is no conclusive evidence of adverse effects on pregnancy outcomes, including fetal growth, at levels of consumption below 120 g of alcohol per week. However, there are suggestions that excessive consumption of alcohol can cause or aggravate maternal hypertension.

There is no reliable information from well-designed, randomized controlled trials assessing the best dietary approach for the management of pre-existing hypertension during pregnancy. Recommendations come from experts' consensus and authorities' statements. It seems that mild-to-moderate pre-existing (chronic, essential) hypertension without any risk factor should be managed in the same way as in the nonpregnant state. However, additional concerns include effects on fetal well-being (mainly intrauterine growth restriction) and worsening of hypertension, particularly as a result of superimposed pre-eclampsia.

Conclusions

In short, based on the available data from systematic reviews (presented in the Appendix), we can conclude that there is some evidence that calcium supplementation in populations with a low calcium intake and/or at risk of developing pregnancy-induced hypertension could be beneficial. Antioxidants (particularly vitamins E and C) are promising, but there is a need for adequately sized and designed randomized controlled trials to confirm these findings before any widespread recommendation. Although pregnant women living in developing countries could be exposed to several other nutrient deficiencies, no evidence precludes recommending other nutrient supplementations as part of their routine antenatal care in order to prevent the occurrence of pregnancy-induced hypertension or preeclampsia.

See also: **Antioxidants**: Natural Antioxidants; **Calcium**: Properties and Determination; **Copper**: Properties and Determination; **Hypertension**: Physiology; **Obesity**: Etiology and Diagnosis; **Pregnancy**: Metabolic Adaptations and Nutritional Requirements; **Vitamins**: Overview

Further Reading

Atallah AN, Hofmeyr GJ and Duley L (2002) Calcium supplementation during pregnancy for preventing hypertensive disorders and related problems (Cochrane

Review). In: *The Cochrane Library, Issue 3*. Oxford: Update Software.

Belizán JM, Villar J and Repke J (1988) The relationship between calcium intake and pregnancy-induced hypertension: up-to-date evidence. *American Journal of Obstetrics and Gynecology* 158: 898–902.

Duley L and Henderson-Smart D (2002) Reduced salt intake compared to normal dietary salt, or high intake, in pregnancy (Cochrane Review). In: *The Cochrane Library, Issue 3*. Oxford: Update Software.

Ferrer RL, Sibai BM, Murlow CD *et al*. (2002) Management of mild chronic hypertension during pregnancy: a review. *Obstetrics and Gynecology* 96: 849–860.

Kramer MS (2002a) Balanced protein/energy supplementation in pregnancy (Cochrane Review). In: *The Cochrane Library, Issue 3*. Oxford: Update Software.

Kramer MS (2002b) Energy/protein restriction for high weight-for-height or weight gain during pregnancy (Cochrane Review). In: *The Cochrane Library, Issue 3*. Oxford: Update Software.

Kramer MS (2002c) Isocaloric balanced protein supplementation in pregnancy (Cochrane Review). In: *The Cochrane Library, Issue 3*. Oxford: Update Software.

Kramer MS (2002d) Nutritional advice in pregnancy (Cochrane Review). In: *The Cochrane Library, Issue 3*. Oxford: Update Software.

Kulier R, de Onis M, Gülmezoglu AM and Villar J (1998) Nutritional interventions for the prevention of maternal morbidity. *International Journal of Gynaecology and Obstetrics* 63: 231–246.

Mahomed K (2002a) Folate supplementation in pregnancy (Cochrane Review). In: *The Cochrane Library, Issue 3*. Oxford: Update Software.

Mahomed K (2002b) Iron and folate supplementation in pregnancy (Cochrane Review). In: *The Cochrane Library, Issue 3*. Oxford: Update Software.

Mahomed K (2002c) Zinc supplementation in pregnancy (Cochrane Review). In: *The Cochrane Library, Issue 3*. Oxford: Update Software.

Makrides M and Crowther CA (2002) Magnesium supplementation in pregnancy (Cochrane Review). In: *The Cochrane Library, Issue 3*. Oxford: Update Software.

Moutquin JM, Garner PR, Burrows RF *et al*. (1997) Report of the Canadian Hypertension Society Consensus Conference: 2. Nonpharmacologic management and prevention of hypertensive disorders in pregnancy. *Canadian Medical Association Journal* 157: 907–919.

National High Blood Pressure Education Program (1990) Working group report on high blood pressure in pregnancy. *American Journal of Obstetrics and Gynecology* 163: 1689–1712.

Olsen SF, Secher NJ, Tabor A *et al*. (2000) Randomised clinical trials of fish oil supplementation in high risk pregnancies. *British Journal of Obstetrics and Gynaecology* 107: 382–395.

Report of the National High Blood Pressure Education Program Working Group on High Blood Pressure in Pregnancy (2000) *American Journal of Obstetrics and Gynecology* 183: S1–S22.

Roberts JM and Cooper DW (2001) Pathogenesis and genetics of preeclampsia. *Lancet* 357: 53–56.

Villar J and Belizan JM (2000) Same nutrient, different hypotheses: disparities in trials of calcium supplementation during pregnancy. *American Journal of Clinical Nutrition* 71 (supplement): 1375S–1379S.

Appendix

The Cochrane Collaboration

The Cochrane Collaboration is an international non-for-profit organization that aims to help people make well-informed decisions about healthcare by preparing, maintaining, and promoting the accessibility of systematic reviews of the effects of healthcare interventions. The Cochrane Collaboration's work is based on principles of collaborative efforts in order to avoid duplication by good management and coordination to maximize economy of effort. A variety of approaches such as scientific rigour, insuring broad participation, and avoiding conflicts of interest are promoted to minimize bias. Published reviews are keeping up to date by a commitment to insure that Cochrane Reviews are maintained through identification and incorporation of new evidence.

Randomized Controlled Trial

This is an epidemiologic experiment in which the investigators randomly allocate eligible subjects into intervention groups, usually called 'study' and 'control' groups, to receive or not to receive one or more interventions that are being compared. The results are assessed by comparing outcomes in the treatment and control groups.

Systematic Review

A systematic review is a systematic search and critical evaluation of all primary studies answering the same question. Statistical methods may or may not be used to analyze and summarize the results of the studies included in the review. The metaanalysis involves the use of statistical techniques within a systematic review to integrate the results of the included studies. Systematic reviews are important tools that help clinicians, health providers, researchers, and policymakers to summarize the existing information in order to make evidence-based decisions.

Relative Risk (RR)

The ratio of risk in the intervention (exposed) group to the risk in the control (unexposed) group. A RR = 1 indicates no difference between comparison groups. For harmful or undesirable outcomes, a RR that is less

than 1 indicates that the intervention was effective in reducing the risk of that outcome.

Confidence Interval (CI)

This is the range within which the 'true' value is expected to lie with a given degree of certainty (e.g., 95 or 99%). Confidence intervals represent the probability of random errors.

Nutrition in Diabetic Pregnancy

A Dornhorst and G Frost, Nutrition and Dietetic Research Group, Imperial College of Science and Medicine, London, UK

Introduction

Gestational diabetes mellitus (GDM) is glucose intolerance first recognized in pregnancy. For the majority of women this is a transient deterioration in glucose tolerance which is due to the hormonal changes of pregnancy. In a minority of women the glucose intolerance continues after pregnancy and these women have either had previously unrecognized type 2 diabetes or are in the early stages of type 1 diabetes. Older, obese women of high parity are most at risk of GDM, especially when from an ethnic background with a high prevalence of type 2 diabetes.

Gestational diabetes predisposes to accelerated fetal growth which is the main contributor to the obstetric complications of GDM. Both cesarean section rates and birth trauma are increased in GDM pregnancies. The stillbirth rate is increased when GDM remains either unrecognized or untreated. Many women with GDM are obese, with a prepregnancy body mass index (BMI) $>29\,kg\,m^{-2}$; this itself is associated with an increased risk of a large-for-gestational-age (LGA) infant, operative delivery, pregnancy-induced hypertension, and thromboembolic disease. Neonatal complications of GDM include transient hypoglycemia and hypocalcemia. The long-term sequelae of GDM are future type 2 diabetes in the mother and future obesity, insulin resistance and type 2 diabetes in the child.

Dietary management of GDM can alone, or in combination with insulin, reduce maternal hyperglycemia sufficiently to improve pregnancy outcome. Providing dietary education and advice that extends beyond the pregnancy is also important and this has the potential to lessen the future risk of GDM in subsequent pregnancies as well as type 2 diabetes in the mother and her child.

Despite the recognition of the importance of dietary management for GDM, there is little consensus regarding the optimal diet to prescribe. There are too few controlled dietary studies during a GDM pregnancy with sufficient nutritional detail to provide the necessary guidelines. There are no long-term prospective studies on whether the dietary advice given during pregnancy reduces future recurrence of GDM or diabetes in either the mother or her child in later life. Different and sometimes conflicting dietary advice is advocated for the management of GDM. Debate surrounds the energy content of the diet and optimal proportions and type of carbohydrate and fat that should be prescribed. The minimum weight gain that is safe in the obese GDM woman still needs to be defined. Below we attempt to provide a rational framework on which we believe dietetic advice should be given for the management of GDM. We discuss the potential short- and long-term benefits to the mother and her child which may be gained from dietary intervention while emphasizing the need for future dietary studies.

Background: General Dietetic Principles Applicable to Pregnancy

Energy Requirements in Pregnancy

Pregnancy is an anabolic state. During pregnancy, sufficient energy sources need to be acquired to cover the products of conception, the fetal–placental unit, and the changes in maternal tissues. This last includes mammary, uterine, adipose deposits, and the expansion in blood volume. The basal metabolic rate increases by 15–26% during pregnancy to meet the metabolic demands of these newly synthesized tissues. The total energy cost for pregnancy was theoretically calculated by Hytten and Leitch in the 1960s to be 355 640 kJ (85 000 kcal), equivalent to an extra 1191.3 kJ (285 kcal) a day. Subsequent longitudinal studies during pregnancy, using the latest physiological techniques, in well-nourished nonobese women confirm these original calculations. These early calculations have formed the basis of the dietary recommendations for pregnancy. Cross-sectional and longitudinal nutritional studies during pregnancy have repeatedly shown that these energy costs are seldom met by an equivalent increase in food intake. The increase in dietary energy appears to provide only 20% of the total energy costs of pregnancy.

How any one women meets the energy needs of pregnancy is highly variable and it is this variability

and adaptability to environmental pressures that allow pregnancies to occur under adverse circumstances, such as food deprivation or the need to perform physical labor during pregnancy. These highly adaptive processes include limiting adipose deposition, reducing diet-induced thermogenesis, and limiting physical energy expenditure. Any reduction in diet-induced thermogenesis is either small or of marginal significance. It has been estimated that a 20% reduction in physical activity during pregnancy alone would save the necessary energy to complete pregnancy. Recent work, however, suggests that significant energy savings from reducing physical activity only occur in women with high physical expenditures before pregnancy and those involved in high physical work loads during pregnancy. Under conditions of extreme calorie restriction, a fall in basal metabolic rate has been observed in the first half of pregnancy.

Rigid guidelines for increasing dietary energy consumption in pregnancy in well-nourished and obese pregnant women are probably misguided. Overall, there is a poor correlation between dietary intake and gestational weight gains. In nonobese and underweight women, pregnancy outcome improves when the minimal weight gain targets are achieved. In obese mothers minimal or no weight gain in pregnancy does not appear to jeopardize the pregnancy. As most women with GDM are obese, these women should not be expected to meet their energy costs of pregnancy through increased dietary intake alone. Despite firm evidence, it is generally proposed that all obese women should have weight gains in pregnancy to account for the products of conception and significant calorie restriction should be avoided.

Postabsorptive and Postprandial Metabolism in Normal Pregnancy

The metabolic changes in pregnancy insure optimal maternal fetal fuel transfer. Glucose is the primary fetal oxidative substrate and by late pregnancy the fetus utilizes an estimated 17–26 g glucose per day. This high obligatory fetal oxidative metabolism results in an increase in maternal carbohydrate oxidation and an accompanying rise in the 24-h respiratory quotient. Glucose is preferentially diverted from the mother to the fetus by the fall in maternal insulin sensitivity. This reduction in insulin sensitivity is due to the antagonistic actions of placental hormones, many of which are also lipolytic. The rise in circulating free fatty acids in late pregnancy also contributes directly to the increase in peripheral insulin resistance. Increased maternal insulin resistance helps maximize the transfer of glucose postprandially to the fetus as glucose uptake into maternal peripheral tissues is impaired due to the insulin resistance. Postprandial glucose and insulin concentrations rise during pregnancy in well-nourished women consuming western-style diets and insulin resistance falls. By contrast, little or no rise in postprandial glucose concentrations or insulin occurs in rural populations consuming more traditional low glycemic index diets and in these women insulin resistance does not fall. This observation strongly suggests that habitual diets and lifestyles can influence maternal glucose tolerance and insulin sensitivity.

The high postprandial insulin levels associated with decreased maternal insulin sensitivity facilitate maternal fat deposition. The metabolic milieu of pregnancy is very permissive for fat accumulation and even undernourished women have a 2-kg increase in their adipose stores. In well-nourished women body fat increases by approximately 9% with 2–5 kg of tissue being accumulated by the end of pregnancy. Fatty acid oxidation falls in late pregnancy and this too will further favor adipose accumulation.

Metabolic changes in maternal fat metabolism insure a constant supply of fetal glucose in the postabsorptive state. This is achieved by a faster switchover from lipogenesis to lipolysis. Human placental lactogen and other placental hormones increase maternal lipolysis which generates ketone bodies and glycerol – two important gluconeogenic substrates that help spare maternal glucose and amino acids for the fetus. Ketone bodies, but not nonesterified acids, can cross the placenta and be used as fetal fuels. Despite lower fasting glucose values, hepatic glucose output increases in pregnancy due to the glucogneogenic nature of many of the pregnancy-related hormones and the increase in hepatic insulin resistance. This increase in hepatic glucose output under fasting conditions also helps maintain a steady supply of glucose to the fetus.

Metabolic Changes Associated with GDM

A degree of deterioration in glucose tolerance in pregnancy is characteristic of well-nourished women consuming western-style diets. However, only in a minority of women is the deterioration sufficient to warrant the diagnosis of GDM. Most women remain glucose-tolerant through their ability to treble their insulin secretion in the face of increasing insulin resistance. Women who do develop GDM have subtle abnormalities of β-cell function and, despite being able to maintain glucose tolerance outside pregnancy, are unable to when faced with the high insulin resistance levels encountered in pregnancy. A degree of β-cell dysfunction is universal in women who

develop GDM, both during and following pregnancy. By the time GDM has developed insulin resistance, both peripheral and hepatic insulin resistance are higher than for glucose-tolerant women. There appears to be a relatively greater β-cell defect in the nonobese women who develop GDM than the obese women in whom insulin resistance is a greater contributing factor. The combination of inadequate β-cell secretion and increased insulin resistance results in inadequate nonesterified free fatty acid (NEFFA) suppression postprandially in GDM. Postprandial increases in NEFFA levels would be anticipated to reduce insulin sensitivity further and compromise β-cell function.

The Principle of Dietary Management GDMs

The dietary advice given in GDM needs to lessen the metabolic abnormalities while insuring adequate nourishment for the mother and fetus. Dietary advice should prevent unnecessary weight gain in all pregnancies, especially in obese women. Dietetic advice should in addition provide general guidelines for healthy eating beyond the pregnancy.

General Dietary Recommendations for GDM

The diet should include the recommended vitamins and minerals for pregnancy and potentially harmful foods such as uncooked meats and soft cheese should be avoided. Alcohol is potentially harmful to pregnancy and should be actively discouraged in the first trimester and severely limited for the rest of pregnancy.

When diabetes predates the pregnancy, 5 mg of folate acid should be started before conception and continued throughout the period of organogenesis (12th gestational week/4th week from last menstrual period). Higher doses of folic acid than 400 μg are recommended for nondiabetic pregnancies for the prevention of neural tube defects despite any direct evidence that folate protects against glucose-mediated neural tube defects. All women with GDM diagnosed before the 12th gestational week should start 5 mg of folic acid, as should all previous GDM women in future pregnancies. A recent large nutritional analysis of over 1000 British pregnant women showed relatively low folate intakes – fewer than 10% of women were taking folate supplements in early pregnancy.

Dietary antioxidants protect diabetic rodents against the teratogenic effects of hyperglycemia. Although similar evidence is not available for human pregnancies, insuring adequate antioxidants in the diets of diabetic mothers during organogenesis seems reasonable.

Calcium and vitamin D supplements during both pregnancy and lactation should be considered for Asian Indian women and others with poor sunlight exposure and/or low calcium intakes. Recently a relationship between insulin resistance and certain vitamin D receptor polymorphism has been reported, providing a potential association between the risk of vitamin D deficiency and diabetes.

Recommended Maternal Weight Gains for Normal Pregnancies

The published guidelines on recommended maternal weight gains are based on large obstetric surveys in the USA (Table 1). The maternal weight gain required to minimize the frequency of small-for-gestational-age (SGA) infants is higher for underweight (BMI $< 19.8\,kg\,m^{-2}$) than overweight or obese women. When the prepregnancy BMI is $> 35\,kg\,m^{-2}$ with little or no maternal weight gain, there is no added risk of an SGA infant. Overweight (BMI 26.1–29 $kg\,m^{-2}$) and obese (BMI $> 29\,kg\,m^{-2}$) women are more susceptible to giving birth to an LGA infant and this risk increases further as maternal weight gain increases. The US obstetric recommendation for a 7-kg minimum weight gain for all obese women may not be universally appropriate, especially for morbidly obese women (BMI $> 35\,kg\,m^{-2}$).

The dietary management of GDM women includes the management of obese pregnant women. Unnecessary weight gain in pregnancy contributes to postpartum obesity, which will increase the likelihood of future GDM recurrence, diabetes, and other obesity-related comorbidities. The American Diabetic Association (ADA) has endorsed dietary guidelines for management of diabetes in pregnancy recommending a daily calorie allowance based on prepregnancy ideal body weight (IBW) and current pregnancy weight. These guidelines recommend 36–40 $kcal\,kg^{-1}$ when IBW < 90%, 30 $kcal\,kg^{-1}$ for IBW of 90–120% and 24 $kcal\,day^{-1}$ for IBW 121–150% and as little as 12–18 $kcal\,kg^{-1}$ during pregnancy

Table 1 The 1990 guidelines of the US Institute of Medicine on maternal weight gain targets according to prepregnancy body mass index

	Underweight ($< 19.8\,kg\,m^{-2}$)	*Normal weight* ($19.8–26\,kg\,m^{-2}$)	*Overweight* ($> 26\,kg\,m^{-2}$)
Weight gain term target	12.5–18 kg	11.5–16 kg	7.0–11.5 kg

when the prepregnancy IBW > 150% IBW. The ADA currently endorses the minimum recommended weight gain of 7.0 kg for obese (BMI > 29 kg m^{-2}) women, as published by the American College of Obstetricians and Gynecologists. No equivalent weight or daily calorie guidelines exist for the UK.

Calorie Restriction in the Obese Woman with GDM

To date there are no long-term studies on the psychological or physical development of infants whose obese mothers are mildly calorie-restricted in late pregnancies for GDM. Long-term follow-up studies are available from infants born to well-nourished Dutch women with severe calorie-restricted daily allowance of 800 kcal, in late pregnancy during the 5½-month famine of 1944–45. These infants were thinner at birth but had normal subsequent childhood development and were reported to be less obese than other recruits on entry into the army at 18 years old. However, subsequent follow-up has shown that among this cohort of children born to calorie-restricted mothers in late pregnancy, the incidence of glucose intolerance and diabetes was increased in middle age. It is these long-term studies combined with genuine concerns relating to the harmful effect of maternal ketosis, caused by calorie restriction, on fetal neurophysiological and cognitive development that have resulted in a general reluctance to advocate calorie restriction in pregnancy.

While no one recommends severe calorie restriction in pregnancy, there is probably a place for modest calorie constraint for obese women with GDM. Women with GDM are relatively more ketosis-resistant to mild calorie restriction than glucose-tolerant women; this may be attributable to their higher hepatic glucose outputs. Theoretically, modest calorie-restricted diets that are given as small frequent meals containing slowly absorbed carbohydrates will help minimize maternal ketosis as these diets are associated with attenuated insulin responses which delay lipolysis and therefore ketogenesis.

We have previously reported that in obese GDM women modest calorie restriction of 20–25 kcal kg^{-1} day^{-1} from the 24th gestational week reduces the frequency of LGA infants compared with obese glucose-tolerant women not dieted. In this study the GDM women gained only half the weight of the controls from 28 weeks to term (1.7 ± 1.6 versus 4.1 ± 3.1 kg) and this was associated with fewer LGA infants than for obese non-GDM controls. Other studies using hypocaloric diets have also shown improvements in glycemic control with mild calorie restriction in pregnancy.

On today's evidence it would seem appropriate to limit weight gain in GDM pregnancies to the bottom rather than the top for those recommended for average, overweight, and obese women, while setting no minimum weight gain for the grossly obese (> 34 kg m^{-2}). This would insure that following pregnancy there would be no overall weight gain in the overweight women and potentially allow weight loss in the morbidly obese woman.

The Optimal Amount and Type of Dietary Carbohydrate and Fat for GDM

Glucose crosses the placenta in both a concentration-dependent manner and by specific glucose transporters. The reduction of postprandial peak glucose concentrations has a greater impact on limiting accelerated fetal growth than lowering postabsorptive levels. Therefore diets that prevent excessive postprandial placental uptake of glucose should be the basis of our dietary advice to all women with GDM.

Controversy surrounds the optimal dietary proportion that should be taken as carbohydrate. Recent advice from the USA has advocated limiting the overall percentage of dietary carbohydrate to < 45%, as in short-term studies this has been shown to improve glycemic control. The current guidelines approved by the ADA suggest limiting carbohydrate to 40% of the total energy content while increasing dietary fat to 40%. Like others, we would argue that it is the type rather than the absolute amount of carbohydrate that dictates postprandial glycemia. The glycemic index of different ingested carbohydrates quantifies their different glycemic responses. Starches, fiber, and refined sugars are the main dietary carbohydrates and their rates of intestinal absorption are different. The highest postprandial glucose and insulin values occur with refined sugars, the most rapidly absorbed, while the lowest values are with the soluble fibers that have the slowest absorption rates. In pregnancy, if the carbohydrate is predominantly refined, any increase above 45% of the total dietary energy content is associated with a deterioration in glycemic control, especially if given first thing in the morning. However, 60% of the total energy content of the diet can be consumed as low glycemic index carbohydrates with no change in glucose tolerance or an actual improvement. Both during and outside pregnancy the introduction of low-glycemic-index diets has been shown to reduce insulin sensitivity. There is a degree of gastric stasis in pregnancy and this itself can further lower the glycemic index of many carbohydrates.

The basis of the dietary management of GDM is to reduce postprandial glycemia. This theoretically can

be achieved either by limiting the total amount of carbohydrate ingested or insuring that the dietary carbohydrates consumed have a low glycemic index. Reducing carbohydrates for breakfast has also been advocated by some as the postprandial glucose values tend to be highest mid-morning. However, limiting carbohydrates will result in a higher proportion of the diet consumed as fat, which is more energy-dense and is likely therefore to result in weight gain. In addition, diets that emphasize limiting the proportion of carbohydrate over fat are, we believe, sending out the wrong educational message. In practice it is often difficult to achieve compliance when greater than 50% of the energy content of the diet is given as carbohydrate. This is especially so with the low-glycemic-index carbohydrates, which produce greater satiety than refined sugars or fats.

Dietary Fat

We believe the current American recommendations, endorsed by the ADA, that 40% of the GDM diet should be derived from dietary fat, are misguided. The short-term dietary studies which demonstrated a benefit of high-fat diets on postprandial blood glucose values may, as discussed above, have been accounted for by the high glycemic index of the carbohydrates used in these studies. In addition, promoting high-fat diets in women destined to develop diabetes and therefore cardiovascular disease is highly questionable. High-fat diets have also been linked with both β-cell toxicity and increased insulin resistance. There are currently insufficient data to know if chain length and degree of saturation of fatty acids are important dietary factors in determining glucose tolerance during pregnancy.

Diet and Insulin Therapy for GDM

The aim of medical management of women with GDM is to insure maternal glycemia can be kept at a safe level. The aim of management is not to delay insulin treatment if required. The fasting and postprandial glucose levels at which insulin therapy should be introduced in addition to diet have been specified as a fasting value $> 5.5\,\mathrm{mmol\,l^{-1}}$ and 1-h postprandial value $> 7\,\mathrm{mmol\,l^{-1}}$. However, the time that dietary management should be tried before starting insulin is less well-specified. Even when very tight glycemic targets are being achieved, most women can be managed with diet alone. The women who require insulin are the most metabolically compromised and have the highest perinatal complications and the fastest deterioration postpartum for the development of diabetes. It is important to recognize these women early in pregnancy and not to lay any sense of failure or noncompliance on them or to delay the introduction of early insulin treatment. Insulin is also occasionally introduced later in pregnancy for obstetric rather than pure glycemic reasons; these would include evidence of accelerated fetal growth and unexplained polyhydramnios.

The principles of dietary management for GDM change little with the introduction of insulin. One should continue to limit weight gain in obese women while providing sufficient carbohydrate snacks throughout the day to prevent hypoglycemia.

Whenever insulin is prescribed with the intention of achieving euglycemia, the frequency of hypoglycemia increases. Although short periods of hypoglycemia are not detrimental to the fetus, it is both frightening and unpleasant for the woman. In addition, after an episode of hypoglycemia the blood glucose rises acutely, both through the action of the counterregulatory hormones and the usual subsequent consumption of a sugary drink. Frequent episodes of hypoglycemia often result in women chasing these high-rebound glucose levels by increasing their insulin dosage and this can result in further hypoglycemic attacks.

Women on insulin should have a regular intake of carbohydrate throughout the day. This needs be taken with each meal, as well as in a snack between meals and at bedtime. To avoid unnecessary calorie intake, snacks that are low in fat should be advised: fruit is ideal for this purpose. Meanwhile, low-glycemic-index carbohydrate should be advised for meals. This strategy will increase the absorption period postprandially, thereby minimizing the risk of hypoglycemia. The use of the soluble fiber supplement guar gum can be introduced if blood glucose levels cannot be controlled adequately without increasing insulin to levels that are causing hypoglycemia.

Long-Term Dietary Advice for the Mother and her Child

Women with GDM are at greatly increased risk of future diabetes. Many of these women are obese and all have at least one child at increased risk of adolescent obesity and diabetes. Given this, it is obvious that dietary advice, education, and reinforcement should continue beyond the pregnancy. Women should be made fully aware of their added risk of both future GDM pregnancies and type 2 diabetes and the importance of not gaining weight postpartum. Low-fat diets have been shown to minimize the recurrence of subsequent GDM pregnancies. Suitable healthy living advice should be given to all women following a GDM pregnancy, emphasizing the benefits of such a

diet for the whole family. Ideally, all these women would receive annual long-term dietetic input within the community. The case for continual follow-up of these women will become more apparent with the result of the ongoing diabetic prevention studies currently being carried out among high-risk groups, of which previous GDM women are included.

The Need and Feasibility of Future Dietary Studies in Pregnancy

There remains a lack of good randomized studies on the dietary management of GDM. Such studies are required for both short-term pregnancy outcomes and long-term outcomes for the mother and her child. One of the main difficulties in conducting such studies is the control arm: even when no dietary advice is given, once diagnosed with GDM women make lifestyle changes based on family beliefs of information gathered from a variety of sources. Also if the health care providers are aware of the diagnosis they too unintentionally are likely to influence lifestyle factors. The need to blind both the women and the health care staff of the diagnosis is difficult and often considered unethical, as, if ignored, GDM does carry a risk to that pregnancy.

Summary

Gestational diabetes is a common complication of pregnancy. Controlling maternal hyperglycemia with diet alone and diet and insulin can reduce the risk of inappropriate accelerated fetal growth. The prescribed diet for GDM needs to include regular meals and snacks. These should contain a large component of slowly absorbed carbohydrate. As many women with GDM are obese, the diet must not allow excessive maternal weight gain as this further compromises pregnancy outcome and adds to the mother's risk of future diabetes. Dietary advice in pregnancy should extend beyond pregnancy as a woman with GDM has a lifetime risk of future diabetes, as does her child. There remains a need for future dietary studies in the management of GDM.

See also: **Diabetes Mellitus**: Etiology; **Obesity**: Etiology and Diagnosis; **Pregnancy**: Metabolic Adaptations and Nutritional Requirements

Further Reading

Åberg A, Rydhström H, Kallen B *et al.* (1997) Impaired glucose tolerance during pregnancy is associated with increased fetal mortality in preceding sibs. *Acta Obstetrica Gynecologica Scandinavica* 76: 212–217.

Adams KM, Li H, Nelson RL *et al.* (1998) Sequelae of unrecognised gestational diabetes. *American Journal of Obstetrics and Gynecology* 178: 1321–1332.

American College of Obstetricians and Gynecologists (1993) Nutrition during pregnancy. *ACOG Technical Bulletin* 1–7.

American Diabetes Association (1995) Nutritional management. In: *Medical Management of Pregnancy Complicated by Diabetes*, 2nd edn. Virginia: American Diabetes Association. pp. 47–56.

Berkowitz GS, Lapinski RH, Wein R *et al.* (1991) Race/ethnicity and other risk factors for gestational diabetes. *American Journal of Epidemiology* 135: 965–973.

Bianco AT, Smilen SW, Davis Y *et al.* (1998) Pregnancy outcome and weight gain recommendations for morbidly obese women. *Obstetrics and Gynecology* 91: 97–102.

Buchanan TA, Kjos SL, Montoro MN *et al.* (1994) Use of fetal ultrasound to select metabolic therapy for pregnancies complicated by mild gestational diabetes. *Diabetes Care* 17: 275–283.

Butte NF, Hopkinson JM, Mehta N *et al.* (1999) Adjustments in energy expenditure and substrate utilisation during late pregnancy and lactation. *American Journal of Clinical Nutrition* 69: 299–307.

Casey BM, Lucas MJ, McIntire DD *et al.* (1997) Pregnancy outcomes in women with gestational diabetes compared with the general obstetric population. *Obstetrics and Gynecology* 90: 869–873.

Catalano PM, Huston L, Amini SB *et al.* (1999) Longitudinal changes in glucose metabolism during pregnancy in obese women with normal glucose tolerance and gestational diabetes mellitus. *American Journal of Obstetrics and Gynecology* 180: 903–916.

Dabelea D, Pettitt DJ, Hanson RL *et al.* (1999) Birth weight, type 2 diabetes, and insulin resistance in Pima Indian children and young adults. *Diabetes Care* 22: 944–950.

Dornhorst A and Frost G (1997) The potential for dietary intervention postpartum in women with gestational diabetes. *Diabetes Care* 20: 1635–1637.

Dornhorst A and Rossi M (1998) Risk and prevention of type 2 diabetes in women with gestational diabetes. *Diabetes Care* 21 (suppl. 1): B43–B49.

Durnin JVGA (1987) Energy requirements of pregnancy: an integrated study in five countries. Background and methods. *Lancet* ii: 895–897.

Felig DS and Naylor CD (1998) Eating for two: are guidelines for weight gain during pregnancy too liberal? *Lancet* 351: 1054–1055.

Freinkel N (1980) Banting lecture 1980: pregnancy and progeny. *Diabetes* 29: 1023–1035.

Freinkel N (1985) Effects of the conceptus on maternal metabolism during pregnancy. In: Leibel BS and Wrenshall GA (eds) *On the Nature and Treatment of Diabetes*, pp. 679–691. Amsterdam: Exerpta Medica.

Harris MI (1988) Gestational diabetes may represent discovery of pre-existing glucose intolerance. *Diabetes Care* 11: 402–411.

Hawdon JM and Aynsley-Green A (1996) Neonatal complications, including hypoglycaemia. In: Dornhorst A and Hadden D (eds) *Diabetes and Pregnancy: An Inter-*

national Approach to Management, pp. 303–318. Chichester, UK: John Wiley.
Hay WWJ (1991) The role of placental–fetal interaction in fetal nutrition. *Seminars in Perinatology* 15: 424–433.
Hay WWJ (1994) Placental supply of energy and protein substrate to the fetus. *Acta Paediatrica Supplement* 405: 13–19.
Hytten FE (1980) Nutrition. In: Hytten FE and Chamberlain G (eds) *Clinical Physiology in Obstetrics*, pp. 163–192. Oxford, UK: Blackwell Scientific.
Hytten FE and Leitch I (1964) *The Physiology of Human Pregnancy*. Oxford, UK: Blackwell Scientific.
Institute of Medicine (1990) *Nutrition During Pregnancy: Weight Gain and Nutritional Supplements*. Washington, DC: National Academy Press.
Jacobsen JD and Cousin L (1989) A population-based study of maternal and perinatal outcome in patients with gestational diabetes. *American Journal of Obstetrics and Gynecology* 161: 981–986.
Janerich DT (1990) Alcohol and pregnancy. An epidemiologic perspective. *Annals of Epidemiology* 1: 179–185.
Jovanovic L (1998) American Diabetes Association's fourth international workshop-conference on gestational diabetes mellitus: summary and discussion: therapeutic interventions. *Diabetes Care* 21 (suppl. 2): 131–137.
Kalhan S, Rossi K, Gruca L *et al*. (1997) Glucose turnover and gluconeogenesis in human pregnancy. *Journal of Clinical Investigation* 100: 1775–1781.
Kühl C (1991) Insulin secretion and insulin resistance in pregnancy and GDM. *Diabetes* 40(suppl. 2): 18–24.
Major CA, Henry MJ, Veciana M *et al*. (1998) The effects of carbohydrate restriction in patients with diet-controlled gestational diabetes. *Obstetrics and Gynecology* 91: 600–604.
McFarland MB, Langer O, Conway DL *et al*. (1999) Dietary therapy for gestational diabetes: how long is long enough? *Obstetrics Gynecology* 93: 978–982.
Metzger BE, Coustan DR and the Organizing Committee (1998) Summary and recommendations of the fourth international workshop-conference on gestational diabetes mellitus. *Diabetes Care* 21 (suppl. 2): 161–167.
Moses RG (1996) The recurrence rate of gestational diabetes in subsequent pregnancies. *Diabetes Care* 19: 1349–1356.
Moses RG, Shand JL, Tapsell LC *et al*. (1997) The recurrence of gestational diabetes: could dietary differences in fat intake be an explanation? *Diabetes Care* 20: 1647–1650.
National Research Council (1989) *Recommended Dietary Allowance*, pp. 33–34. Washington, DC: National Academic Press.
O'Sullivan JB (1982) Body weight and subsequent diabetes mellitus. *Journal of the American Medical Association* 248: 949–952.
Peterson CM and Jovanovic-Peterson L (1991) Percentage of carbohydrate and glycaemic response to breakfast, lunch, and dinner in women with gestational diabetes. *Diabetes* 40 (suppl. 2): 172–174.
Pettitt D, Aleck K, Baird HR *et al*. (1988) Congenital susceptibility to NIDDM: role of intrauterine environment. *Diabetes* 37: 622–628.
Silverman B, Metzger B, Cho NH *et al*. (1995) Impaired glucose tolerance in adolescent offspring of diabetic mothers. Relationship to fetal hyperinsulinism. *Diabetes Care* 18: 611–617.
Spellacy WN and Goetz FC (1963) Plasma insulin in normal late pregnancy. *New England Journal of Medicine* 268: 988–991.
The MRC Vitamin Research Study Group (1991) Prevention of neural tube defects: the results of the Medical Research Council Vitamin Study.
Wolever TMS, Bentum-Williams A, Jenkins DJ *et al*. (1995) Physiological modulation of plasma free fatty acid concentrations by diet. *Diabetes Care* 18: 962–970.

PREMENSTRUAL SYNDROME: NUTRITIONAL ASPECTS

F Ford, University of Sheffield, Sheffield, UK

Background

Many women seek medical help for symptoms related to premenstrual syndrome (PMS). Physicians, dietitians, and nutritionists often recommend some form of nutrition advice, including incorporating extra vitamin B_6, vitamin B_1, vitamin E, calcium, magnesium, essential and ω-3 fatty acids, herbal remedies, carbohydrates, low-fat and vegetarian diets, and chocolate. Although much (but not all) of this advice is part of a healthful diet, there is relatively little sound scientific evidence that they relieve or prevent symptoms.

PMS is a condition characterized by a number of physical, emotional, and psychological symptoms that appear in the second half of the menstrual cycle, after ovulation. It can affect women in the reproductive age

from adolescence until the menopause. There is a list of menstrual symptoms identifying over 150 different symptoms linked to the menstrual cycle. The main symptoms reported are irritability, depression, anxiety, weight gain, edema, breast pain, fatigue, and headache. Greene and Dalton first used the term premenstrual syndrome in 1953. The first author to describe the syndrome known as premenstrual tension was Frank, in 1931. The intensity of symptoms may vary considerably between women and for each cycle. Severe cyclical symptoms cause suffering to many women in their daily activity and personal relationships.

PMS is perceived to be a common complaint, but data on prevalence vary according to different measurement scales and cultural variations. It is estimated that up to 95% of women suffer mild symptoms, and 5–10% of women have symptoms severe enough to disrupt their lives in the 2 weeks before the onset of menstruation. PMS can be defined as a regular pattern of symptoms occurring just before the start of menstruation and which gradually abate soon after the start of bleeding. Severe PMS can be defined as causing functional impairment in work, relationship, or usual activities. The social consequences of premenstrual syndrome are important and have been the focus of much research. It has been reported that PMS is responsible for a high incidence of crimes, alcoholism, school absence, and admittance to hospital for accidental injury.

The etiology of PMS is unknown, although many theories have been suggested. Factors such as estrogen levels, deficiency of progesterone and progesterone metabolites, increase of adrenal activity, subclinical hipoglucosemy, increase of prolactin and monoaminoxidases, imbalance of renin–angiotensin–aldosterone, deficiency of pyridoxine and other vitamins, and neuroendocrine dysfunction have all been cited.

The diagnosis of the PMS is essentially based on clinical symptoms. A prospective evaluation of three consecutive months that show these clinical symptoms occurring during the luteal phase of the menstrual cycles characterize PMS. However, diagnosis of the condition may be difficult, as there is no standardized diagnostic tool. Women's experiences of PMS are so varied that it is difficult to fit them into categories or even subgroups of symptoms. In the past, some doctors have regarded PMS as a psychological syndrome, so women report finding it difficult to receive satisfactory treatment and attention. This might be one of the reasons why it is a condition that women often diagnose and treat themselves.

Treatment of PMS has been largely empirical. The following dietary components have been manipulated for the treatment of PMS.

Vitamin B_6 (Pyridoxine)

Vitamin B_6 is involved in the production of prostaglandin E2 (which contributes to myometrial relaxation) and in the utilization of magnesium, so higher levels of vitamin B_6 could also influence dysmenorrhea cramps. Vitamin B_6 has been used in the doses of 50–500 mg daily in the second half of the menstrual cycle. It has been suggested that vitamin B_6 may act by correcting a deficiency at the hypothalamic end of the complex psycho-endocrine reproductive pathways. Pyridoxal 5′ phosphate, the active form of vitamin B_6, serves as the coenzyme of a wide variety of enzymes of amino acid metabolism. For example, it serves as a cofactor in the metabolism of tryptophan (the precursor of serotonin) and also in the metabolism of tyrosine (leading to dopamine and noradrenaline) and glutamate (leading to γ-aminobutyric acid). Low levels of dopamine and serotonin lead to high levels of prolactin and aldosterone, thus explaining the fluid retention, and the effect on the neurotransmitters could explain the psychological symptoms in PMS.

In recent years, a systematic review on the efficacy of vitamin B_6 in the treatment of PMS concluded that:

- Randomized placebo controlled studies of vitamin B_6 treatment for PMS were of insufficient quality to draw definitive conclusions.
- Limited evidence exists to suggest that 100 mg of vitamin B_6 daily (and possibly 50 mg) is likely to be beneficial in the management of PMS.
- Vitamin B_6 was significantly better than placebo in relieving overall premenstrual symptoms and in relieving depression associated with PMS, but the response was not dose-dependent.
- No conclusive evidence was found of neurological side-effects with these doses.
- A randomized controlled trial of sufficient power and quality is needed to compare vitamin B_6 with placebo to establish definitive recommendations for treatment.

A systematic review published in the Cochrane Database concluded that no definite results could be reported regarding the efficacy of vitamin B_6 in improving dysmenorrhea.

Calcium

Previous reports have suggested that disturbances in calcium regulation may underlie the pathophysiologic characteristics of PMS and that calcium supplementation may be an effective therapeutic approach. In one study, each participant received 6 months of treatment, involving 3 months of daily calcium supplementation

(1000 mg of calcium carbonate) and 3 months of placebo. Three premenstrual factors (negative affect ($p = 0.05$); water retention ($p = 0.005$); pain ($p = 0.05$)) and one menstrual factor (pain ($p = 0.05$)) were significantly alleviated by calcium.

Another study evaluated the effect of 1200 mg of elemental calcium on the luteal and menstrual phases of the menstrual cycle in 497 women with PMS, in a prospective, randomized, double-blind, placebo-controlled, multicenter clinical trial. During the luteal phase of the treatment cycle, a significantly lower mean symptom complex score was observed in the calcium-treated group for both the second ($p = 0.005$) and third ($p < 0.001$) treatment cycles. By the third treatment cycle, calcium effectively resulted in an overall 48% reduction in total symptom scores from baseline compared with a 30% reduction in placebo. All four-symptom factors were significantly reduced by the third treatment cycle.

Magnesium

Results from a recent metaanalysis have suggested that magnesium supplements of 500 mg per day may be a promising treatment for one of the major symptoms of PMS, i.e., dysmenorrhea. Magnesium treatment has been shown to be more effective than placebo, with those taking magnesium requiring significantly less absence from work and less use of additional medication. Minimal adverse effects have been experienced in both magnesium and placebo groups with no significant differences shown. However, overall, there has been little conclusive evidence for the effectiveness of magnesium for symptom relief. Although all the trials have been randomized and double blind with adequate methodological quality, they were all small trials, and there was poor measurement and reporting of pain outcomes. Two of the included trials had withdrawal rates of over 30%. Many of these participants may have failed to complete the trial due to a lack of efficacy or adverse effects from the treatments, and so the high withdrawal rates have an impact on the strength of the evidence. It is also unclear which treatment regimen was more effective, as the trials both administered treatment in quite different ways, one daily and the other during menses only, and at different doses. Therefore, no strong recommendation can be made about the efficacy of magnesium until a further evaluation is carried out. A randomized controlled trial of magnesium with a larger number of participants and adequate outcome measurement is needed.

The largest of the included trials also reported data on the levels of prostaglandin F2 α in menstrual blood. Overproduction of this prostaglandin has been shown to be a substantial contributing factor to the painful cramps associated with dysmenorrhea. Women taking the magnesium therapy had substantially lower levels of PGF2 α in their menstrual blood than those on placebo ($p < 0.05$), which mirrored the therapeutic decrease in pain experienced by the participants. This highlights the possible biological rationale behind magnesium therapy for dysmenorrhea; it inhibits the biosynthesis of PGF2 α as well as having a role in muscle relaxation and vasodilation.

Vitamin B_6 and Magnesium

Magnesium was shown to be no different in pain outcomes from both vitamin B_6 and a combination of vitamin B_6 and magnesium by one small trial. The same trial also showed that a combination of magnesium and vitamin B_6 was no different from placebo in reducing pain. However, vitamin B_6 alone was more effective at reducing pain than both placebo and a combination of magnesium and vitamin B_6. In a recent systematic review, no definite conclusions could be made about the efficacy of the combination of vitamin B_6 and magnesium supplements.

Vitamin B_1 (Thiamin)

Vitamin B_1 plays an important role in metabolism, and vitamin B_1 deficiency can be characterized by fatigue, muscle cramps, various pains, and a reduced tolerance to pain, all factors that could be associated with dysmenorrhea and PMS. Vitamin B_1 was shown to be an effective treatment for dysmenorrhea, taken at 100 mg daily, although this conclusion is tempered slightly because it is based on only one large randomized controlled trial (RCT). This RCT trial demonstrated a significant effect of vitamin B_1 taken daily compared with placebo in reducing pain. Results from phase one of the trial only, before participants were crossed over to the other treatment group, were used when calculating odds ratios. After crossover, those swapped from vitamin B_1 treatment to placebo maintained pain relief. The authors of the trial interpreted this as a curative effect of vitamin B_1 treatment. The length of the trial (2 months of either treatment) makes this interpretation difficult to confirm, as no longer-term follow-up was made, and the strong placebo effect that is typical in dysmenorrhea trials could account for some of the maintenance of effect. Vitamin B_1 needs to be assessed further by a randomized trial, preferably a large multicentered trial with different study populations that would generate generalizable

results that could confirm the positive results of the above trial.

Vitamin E

It has been suggested that vitamin E has analgesic and antiinflammatory properties. A randomized trial of vitamin E for rheumatoid arthritis has shown a significant reduction in pain parameters, which lends further support to this theory. One trial comparing a combination of 100 mg per day of vitamin E and ibuprofen with ibuprofen alone on dysmenorrhea showed no significant difference between the two groups for the outcome of pain relief after 1 month of treatment. Both treatment groups experienced excellent treatment responsiveness, although this conclusion is based on a comparison of baseline scores with post-treatment scores, which is more subject to bias than if a placebo control group had been used. Overall, the addition of vitamin E supplements to the diet has no further effect on pain over that of standard treatment with ibuprofen. The treatment time of 1 month of each treatment was a limitation of the trial, as the efficacy of a dietary therapy cannot really be assessed adequately over such a short period of time.

Essential Fatty Acids

Findings from uncontrolled studies have indicated that essential fatty acid (EFA) metabolism may be abnormal in women with PMS. EFAs are the precursors of prostaglandins. Low levels of prostaglandins are thought to lead to increased sensitivity to prolactin. Women with PMS may be abnormally sensitive to normal amounts of prolactin. This hypothesis – that there is a deficiency in EFAs, leading to low levels of prostaglandin E1 that may attenuate the effects of prolactin – is one rationale for using EFAs as a treatment for women with PMS.

Evening primrose oil (EPO) contains two EFAs – linoleic acid and γ-linolenic acid (GLA). Linolenic acid is needed for the synthesis of prostaglandin E, and GLA is needed for the synthesis of prostaglandin E1. Whilst EPO is generally accepted as a safe product, reported side-effects include occasional nausea, indigestion, and headache. A small number of less common side-effects have also been documented. A potential risk of inflammation, thrombosis, and immunosuppression with prolonged use of GLA has been described.

EPO is available in most countries without a prescription and is heavily promoted for women as being an effective treatment for a range of conditions including PMS. It is, however, an expensive treatment that many women have to pay for themselves. In a systematic literature search of clinical trials of EPO for the treatment of PMS with a view to performing a meta-analysis, only seven placebo-controlled trials have been found, and randomization was clearly indicated in only five trials. Inconsistent scoring and response criteria made statistical pooling and hence a rigorous metaanalysis inappropriate. The two most well-controlled studies failed to show any beneficial effects for EPO, although, because the trials were relatively small, modest effects cannot be excluded. From the current evidence, EPO is of little value in the management of PMS.

ω-3 Fatty Acids

One trial has compared ω-3 fatty acids (fish oil) and placebo, and data have shown that the treatment is significantly more effective than placebo after 2 months' administration. A meta-analysis has examined the outcome of the use of additional medication and showed that the fish oil treatment group consumed significantly fewer tablets than the placebo group. Minimal adverse effects were significantly more likely in the fish oil group compared with the placebo group. None of the adverse effects were particularly serious (nausea, acne exacerbation, and difficulty swallowing capsules), but they were acute enough to cause some women to discontinue treatment. One methodological limitation of this trial is the poor reporting of data. Pain relief was reported, as the average of the two groups after the treatments allocated was crossed-over, and so it is not an adequate assessment of efficacy. The trial was of short duration, only 2 months in each treatment arm, which may not be enough time to properly assess the effect of a dietary intervention.

Levels of polyunsaturated fatty acids (PUFAs) have been correlated with menstrual pain, with higher levels of the ω-3 fatty acids associated with milder menstrual symptoms. The PUFA ω-6 is metabolized into the specific prostaglandins associated with dysmenorrhea, and it appears that the ratio of ω-3 to ω-6 is associated with menstrual symptoms, therefore a diet higher in ω-3 fatty acids is possibly associated with less dysmenorrhea.

Herbal Remedies

It is advisable that women consult a qualified medical practitioner who specializes in complementary medicine before taking unproven herbal remedies. There is a Practitioner Directory, which lists a selection of the best doctors and practitioners who specialize in treatment using complementary medicine.

Japanese Herbal Combination (Toki-Shakuyaku-San)

The combination herbal remedy Toki-shakuyaku-san (TSS) has been compared with placebo in one trial. Pain was assessed using a visual analog scale (VAS). TSS was significantly more effective in reducing VAS scores, after 2 months of treatment. This difference was maintained after a 2-month untreated followup period. The use of additional medication in the treatment group was significantly less than in the placebo group. One major limitation of this trial was that participants were included in the trial using a particular traditional Chinese medicine diagnosis of a complex set of symptoms as well as the typical diagnosis of PMS. It is not clear how this would translate into Western therapeutics, or if this remedy would help women without this particular pattern of diagnosis.

Agnus Castus

Agnus Castus has been reported to be a hormone regulator and tonic for the nervous system, being used for PMS to treat mood swings, depression, water retention, and breast pain. The efficacy and tolerability of agnus castus fruit (*Vitex agnus castus* L extract Ze 440) with placebo has been investigated in women with PMS in a prospective, randomized, placebo controlled study. The main efficacy variable was a change from baseline to end point (end of the third cycle) in women's self-assessment of irritability, mood alteration, anger, headache, breast fullness, and other menstrual symptoms including bloating. Results have shown a significant improvement in the active group compared with the placebo group ($p < 0.001$). Some women have reported mild adverse events, none of which caused discontinuation of treatment. It has been concluded that a dry extract of agnus castus fruit is an effective and well-tolerated treatment for the relief of symptoms of PMS.

Hypericum perforatum

A pilot observational study has been carried out to investigate whether *Hypericum perforatum* could relieve the symptoms of PMS in a small group of women. Participants took hypericum tablets for two complete menstrual cycles (1×300 mg of hypericum extract per day standardized to 900 μg of hypericin). The degree of improvement in overall PMS scores between baseline and the end of the trial was 51%, with over two-thirds of the sample demonstrating at least a 50% decrease in symptom severity. The results of this pilot study suggest that there is scope for conducting a RCT to investigate the value of hypericum as a treatment for PMS.

St John's Wort

St John's Wort has been reported to raise serotonin levels, which helps to alleviate the mild to moderate depression associated with PMS. It should not be taken with prescription antidepressants, progesterone, or other hormone treatments.

Dong Quai – (*Angelica sinensis*) or Chinese Angelica

Dong Quai (*Angelica sinensis*), or Chinese Angelica, is sometimes known as 'women's ginseng' and is used for PMS.

Natural Progesterone

Natural progesterone is sometimes prescribed where low progesterone levels are indicated in PMS. Overall, there is insufficient evidence to recommend the use of any herbal remedies apart from *Agnus castus* for the treatment of PMS. The Complete German Commission E Monographs list a number of herbal products as being used for dysmenorrhea or menstrual disorders, yet no evidence from clinical trials to support the use of any of these herbal products has been found to date.

Chocolate

Chocolate's ingredients satisfy many aspects of PMS. Its high sugar content induces the brain to make new serotonin and make the chocolate eater feel calm and relaxed. Moreover, chocolate is usually eaten by itself rather than as part of a meal. Thus, there is little chance that protein foods will interfere with the body's ability to make new serotonin. The high butterfat content of chocolate may also change premenstrual mood. High-fat foods can have a numbing effect on mood and on mental and physical energy, and the combination of the fat and sugar content of the chocolate is reported to soothe and tranquilize women. However, chocolate has other ingredients – like phenylethylamine – that are also claimed to have positive mood effects. There are no reliable studies proving this, but anecdotal evidence supports the view that chocolate may have some ingredients that women with PMS crave.

Carbohydrates

The effect of a carbohydrate-rich beverage on mood, appetite, and cognitive function was investigated in a small group of women with PMS. Twenty-four women with confirmed PMS took a beverage or placebo for three menstrual cycles after a 1-month

placebo run-in period, in a double-blind placebo-controlled study. Patients were tested for mood, cognition, and food cravings, using an interactive computer-telephone system, during the luteal phase of the menstrual cycle. The beverage, but not the placebos, significantly decreased self-reported depression, anger, confusion, and carbohydrate craving, and improved memory word recognition.

Isoflavones (Plant Estrogens)

These are often derived from soya and are suggested as a natural hormone replacement for postmenopausal women. However, many are also advocating their use for PMS. There is little evidence for a beneficial effect on PMS, and in fact supplementing with soya protein and other sources of plant estrogens is now controversial, many experts believing that they can also have a harmful effect. They should certainly not be taken if a woman has had breast cancer or a family history of the disease.

L-Tryptophan

Serotonin reuptake inhibitors have been shown to be beneficial in the treatment of PMS. The efficacy of L-tryptophan, which acts specifically on serotonergic neurons, was investigated in a RCT. Results suggested a significant ($p = 0.005$) therapeutic effect of L-tryptophan relative to placebo for the cluster of mood symptoms comprising the items of dysphoria, mood swings, tension, and irritability. These results suggest that increasing serotonin synthesis during the late luteal phase of the menstrual cycle has a beneficial effect in patients with PMS.

Low-fat Vegetarian Diets

A low-fat vegetarian diet has been associated with increased serum sex-hormone binding globulin concentration and reductions in body weight, dysmenorrhea duration and intensity, and PMS duration. The symptom effects might be mediated by dietary influences on estrogen activity.

The UK PMS society states, 'a healthy diet, particularly one which is low in fat and high in fibre can relieve PMS,' and they recommend the following:

- Eating three main meals per day with three smaller snacks in between, all of which should contain starchy foods. This is particularly important in the luteal phase of the cycle.
- Cutting down on sugar and sugary food, as these provide sudden bursts of energy rather than a steady release, which will help combat fatigue and mood swings.
- Eating fiber helps avoid premenstrual constipation, but it is important to drink plenty of water or sugar-free drinks (six to eight glasses a day) to prevent bloating associated with PMS.
- Having one pint of milk a day of either skimmed or semi-skimmed, or an alternative source of calcium such as low-fat cheese or yogurt,
- Reducing salt intake.
- Eating five portions of fruit and vegetables per day.
- Reducing alcohol intake to no more than 2 units per day (1 unit = half a pint (284 ml) of lager or beer, or one small glass of wine, or one measure of spirits).
- Limiting tea and/or coffee to no more thtrun an 5 cups per day, as excess caffeine can make premenstrual symptoms worse.

See also: **Calcium**: Properties and Determination; **Carbohydrates**: Metabolism of Sugars; **Cocoa**: Chemistry of Processing; Production, Products, and Use; **Essential Fatty Acids**; **Herbs**: Herbs and Their Uses; **Magnesium**; **Thiamin**: Properties and Determination; Physiology; **Tocopherols**: Properties and Determination; Physiology; **Vegetarian Diets**; **Vitamin B_6**: Properties and Determination; Physiology

Further Reading

Barr SI, Janelle KC and Prior JC (1995) Energy intakes are higher during the luteal phase of ovulatory menstrual cycles. *American Journal of Clinical Nutrition* 61: 39–43.

Budeiri D, Li Wan Po A and Dornan JC (1996) Is Evening Primrose Oil of value in the treatment of PMS? *Controlled Clinical Trials* 17: 60–68.

Dalton K (1984) *The Premenstrual Syndrome and Progesterone Therapy*. London: William Heineman.

Freeman EW (1997) Premenstrual syndrome: current perspectives on treatment and aetiology. *Current Opinions in Obstetrics and Gynecology* 9: 147–153.

Horrobin DF and Manku MS (1989) Premenstrual syndrome and premenstrual breast pain: disorders of essential fatty acid (EFA) metabolism. *Prostaglandins, Leukotrienes and Essential Fatty Acids Review* 37: 255–261.

O'Brien PMS (1993) Helping women with premenstrual syndrome. *British Medical Journal* 307: 1471–1475.

Proctor ML and Murphy PA (2002) Herbal and dietary therapies for primary and secondary dysmenorrhoea. Cochrane Menstrual Disorders and Subfertility Group. *Cochrane Database of Systematic Reviews. Issue 1.*

Schellenberg P (2001) Treatment for the premenstrual syndrome with agnus castus fruit extract: prospective, randomised, placebo controlled study. *British Medical Journal* 322: 134–137.

Speroff L, Glass RH and Kase NG (1994) *Clinical Gynaecologic Endocrinology and Infertility*. Baltimore, MD: Williams & Wilkins.

Wyatt KM, Dimmock PW, Jones PW and O'Brien PMS (1999) Efficacy of vitamin B-6 in the treatment of premenstrual syndrome: systematic review. *British Medical Journal* 318: 1375–1381.

PRESERVATION OF FOOD

M F Sancho-Madriz, California State Polytechnic University, Pomona, CA, USA

Introduction

Food preservation consists of the application of science-based knowledge through a variety of available technologies and procedures, to prevent deterioration and spoilage of food products and extend their shelf-life, while assuring consumers a product free of pathogenic microorganisms. Shelf-life may be defined as the time it takes a product to decline to an unacceptable level. Deterioration of foods will result in loss of quality attributes, including flavor, texture, color, and other sensory properties. Nutritional quality is also affected during food deterioration. Physical, biological, microbiological, chemical, and biochemical factors may cause food deterioration. Preservation methods should be applied as early as possible in the food production pipeline and therefore include appropriate postharvest handling before processing of both plant and animal foods (Figure 1). Processing techniques usually rely on appropriate packaging methods and materials to assure continuity of preservation. Handling of processed foods during storage, transportation, retail, and by the consumer also influences the preservation of processed foods.

Selection of technology and procedures for food preservation depends on factors inherent to the product, common pathogenic and spoilage microorganisms, and cost. Product-inherent factors include customary ways of consuming the particular food, sensitivity to heat or other principle used to inactivate microorganisms, and other physical and chemical characteristics of the food.

Rationale and Goals of Food Preservation

Many food products are seasonal, and from year to year the yields of agricultural production vary depending on weather, pests, and other factors. During the production season, especially when the yields are high, the offer of fresh products often exceeds the demand. The products that are not consumed in the fresh market can be used in processed form for long periods of time and by people in regions located far from the production areas. Processed foods not only reduce the waste of food but also offer the consumer many convenient features that make preparation at home easier and faster. Food preservation plays an important role in special circumstances like wars or space missions far from earth. In the nineteenth and twentieth centuries many new developments in food preservation occurred during war times, in both industry and government-owned laboratories. More recently, the development of foods suitable for space missions that spend a significant amount of time away from earth has posed new challenges to food scientists.

If the food production system is pictured as a long pipeline that carries food from producers to consumers, one will discover a number of leaks in the pipeline that prevent large amounts of food reaching their final destination. Those leaks or losses represent different forms of food deterioration and spoilage, and are likely to happen from the beginning of food production (Figure 2).

Food preservation aims to prevent and reduce the loss of food in the production system and extend its shelf-life. While developed nations have been very successful in reducing food losses and increasing efficiency of the production system, food deterioration remains an important issue in developing nations and very poor countries where undernutrition of the population is common. In many instances the food problems of those countries could be significantly reduced by the implementation of adequate postharvest handling and food preservation methods without increasing food production. (*See* **Food Security.**)

Appropriate postharvest handling of food products is essential in food preservation, especially for products sold in the fresh market or as minimally processed items. Exposure to sunlight and heat, high

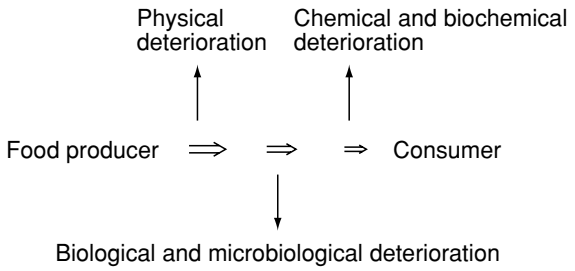

Figure 1 Losses in the food production pipeline.

Food preservation = Adequate postharvest handling
+ food-processing technology + packaging technology
+ adequate transport and storage

Figure 2 Main components of food preservation.

relative humidity, rain, dirt, insects, and rodents increases food deterioration by accelerating chemical reactions in the food and favoring microbial growth. Fast cooling of produce after harvest, storage under temperature- and humidity-controlled conditions, protection of the products by physical barriers, and minimizing the time between harvest and processing are all forms of reducing postharvest losses.

Factors that Cause Deterioration and Spoilage of Food

The factors that cause food deterioration may be grouped under physical, biological and microbiological, and chemical and biochemical factors. Most of the time more than one factor is responsible for the spoilage of a food product and some factors may favor or enhance the effect of others. Light and other forms of radiation, heat, cold, moisture loss, or gain, and the application of force that may alter the structure of the food are examples of physical factors that cause food deterioration. Chemical and biochemical factors include reactions of food components with oxygen or with each other, and reactions catalyzed by enzymatic activity. Maillard browning is caused by a chemical reaction between proteins and carbohydrates and it results in changes in color, flavor, and odor of the food product. This type of browning is desirable in baked products and roasted coffee amongst others, but it is detrimental in foods where browning is not desired. Oxidative rancidity is a chemical change in unsaturated fats and oils and it produces off-flavors and odors. These and other chemical reactions affect sensory qualities of food and can significantly alter the nutritional value of products. (*See* **Browning**: Enzymatic – Biochemical Aspects; **Oxidation of Food Components**; **Spoilage**: Chemical and Enzymatic Spoilage.)

Biological factors include birds, rodents, insects, and parasites. (*See* **Insect Pests**: Insects and Related Pests; Problems Caused by Insects and Mites.) These may consume or destroy the food and contaminate it with pathogenic or spoilage microorganisms. Bacteria, yeasts, and molds are different types of microorganisms that may cause food deterioration and foodborne illness. (*See* **Aflatoxins**; **Microbiology**: Classification of Microorganisms; **Mycotoxins**: Classifications; Occurrence and Determination; Toxicology; **Spoilage**: Bacterial Spoilage; Fungi in Food – An Overview; Molds in Spoilage; Yeasts in Spoilage.) Some viruses may also be transmitted through food but they are not associated with the deterioration of products. Not all microorganisms are detrimental to food; some are indeed very useful and have been utilized since ancient times for the production of fermented products like wine, yogurt, and cheese. Fermented cassava and corn are typical in some African and Latin American countries. (*See* **Fermented Foods**: Origins and Applications; Dietary Importance.)

Because consumers assume that processed foods are safe, the goals of food preservation must include not only measures for the prevention of quality deterioration, but also destruction of pathogenic microorganisms that may cause foodborne illness. A food product that does not show evidence of deterioration may not necessarily be free of pathogenic microorganisms. (*See* **Food Poisoning**: Classification; **Food Safety**.) A preventive strategy for assuring safe food products is the Hazard Analysis Critical Control Point (HACCP) system. HACCP considers biological, physical, and chemical hazards. (*See* **Hazard Analysis Critical Control Point**.)

Factors that Affect Microbial Growth

In addition to the common nutrients found in most foods, microorganisms require certain conditions to survive and grow. The control and optimization of those factors is a fundamental principle of food preservation. The main factors are discussed below.

pH The degree of alkalinity or acidity of a water solution is based on the concentration of hydrogen ($H+$) or hydroxyl ($OH-$) ions. The term pH is by definition the negative logarithm of the hydrogen ion concentration. The pH scale ranges from 0 to 14, and pure water has a pH of 7. Each microorganism will grow only within an optimum pH range. (*See* **pH – Principles and Measurement**.)

Water activity Water activity is a measure of the availability of water in a food product. Solutes dissolved in water bind the water molecules and make them not available for microorganisms to use. Water activity may be defined as the ratio of the vapor pressure of water in food to the vapor pressure of pure water at the same temperature. Microorganisms require a minimum level of water activity in order to grow and thus water activity reduction can be used to control microbial growth. (*See* **Water Activity**: Principles and Measurement; Effect on Food Stability.)

Oxygen Microorganisms that require the presence of free oxygen for their growth are known as aerobes, whereas oxygen inhibits the growth of anaerobes. When they can tolerate the presence of oxygen to a certain level they are known as facultative anaerobes.

Temperature An optimum temperature range is required for microbial growth. Bacteria have been classified into groups depending on their required

temperature range. The temperature range for psychrotrophs is 14–20 °C, for mesophiles is 30–37 °C, and for thermophiles is 50–66 °C. Some thermophiles can grow at temperatures up to 77 °C.

Traditional Methods of Food Preservation

The following section provides a brief description of traditional food preservation methods currently used by the food-processing industry. For more details on each method see the indicated articles within this encyclopedia or the suggested publications for further reading. (*See* **Traditional Food Technology.**)

Acidification

The use of acid to preserve foods, whether by fermentation of the product or by addition of acid foods or acids, is a very common method to preserve foods that does not require a high level of technology or special processing equipment. The acid acts as a preservative by controlling microbial growth and pH meters are used to determine if the food has reached the target pH level. (*See* **Acids**: Properties and Determination; Natural Acids and Acidulants; **pH – Principles and Measurement.**) Acidified foods, sometimes called pickled foods (*see* **Pickling**), have been defined by the US Food and Drug Administration (FDA) as low-acid foods to which acid(s) or acid food(s) is (are) added to produce a product that has a finished equilibrium pH of 4.6 or less and a water activity greater than 0.85. The pH value under 4.6 is necessary to prevent the growth of *Clostridium botulinum*; however, an additional preservation method such as pasteurization, refrigeration, or chemical preservatives is used to destroy or control the growth of deterioration microorganisms. (*See* ***Clostridium***: Botulism.) Foods have different degrees of buffering capacity against acids and therefore different foods require different amounts of acid to reach a desired pH level; the higher the buffering capacity, the higher the amount of acid needed to acidify the product. Other changes in fermented foods such as metabolites produced by some microorganisms also have an effect on microbial inactivation. (*See* **Beers**: History and Types; **Cheeses**: Types of Cheese; **Fermented Foods**: Origins and Applications; **Fermented Milks**: Dietary Importance; **Wines**: Types of Table Wine; **Yogurt**: The Product and its Manufacture.)

Thermal Processing

The development of canning as a preservation method for foods can be traced back to Nicolas Appert (1749–1841), a French chef who published his procedures in 1810. Appert was awarded 12 000 francs for his proposal for preserving foods by heating them in hermetically sealed containers. Not only did Appert not have scientific training but also microorganisms were not known at the time he developed his method, long before Louis Pasteur laid the foundations for the science of bacteriology.

The degree of microorganism destruction achieved by thermal processing varies, depending on the specific temperature and time of the thermal treatment. Thermal process design is based on a target microorganism and desired number of logarithmic reductions in microorganism concentration. Blanching, pasteurization, commercial sterilization, and sterilization are different kinds of thermal processing. Blanching is a mild heat treatment utilized mostly for enzyme inactivation but it also serves other functions. Although it reduces the number of containing microorganisms on the surface of foods, it is not intended as a sole preservation method. Pasteurization involves heating the product under atmospheric pressure without exceeding the boiling point of water (100 °C). Milk and other low-acid foods (pH > 4.6) are pasteurized to eliminate pathogenic microorganisms and extend their shelf-life in combination with refrigeration or other preservation methods. In acid or acidified foods pasteurization is used to destroy spoilage microorganisms and sometimes to inactivate enzymes.

The US FDA defined commercial sterility of thermally processed food as the condition that renders the food free of microorganisms capable of reproducing under normal nonrefrigerated conditions of storage and distribution, and viable microorganisms (including spores) of public health significance. According to the FDA definition, commercial sterility may be achieved by the application of heat alone or combined with water activity control. (*See* **Canning**: Principles; **Heat Treatment**: Ultrahigh Temperature (UHT) Treatments; **Pasteurization**: Principles; Pasteurization of Liquid Products; Pasteurization of Viscous and Particulate Products.)

Concentration by Evaporation

Concentration by evaporation consists of partially removing water from liquid foods by the application of heat. Native Americans used natural evaporation in wooden or bark vessels to manufacture maple syrup using sap. Water removal causes a reduction in water activity. The rate of heat transfer into the food and the rate of mass transfer of vapor from the food are the factors that determine the rate of evaporation. This method of food preservation has a high level of energy consumption and is therefore more expensive. It offers the convenience of a concentrated product that the consumer can dilute at home and it

reduces the cost of transportation and packaging. For example, a can of concentrated orange juice weighs less than the single-strength juice and it requires a smaller package. Destruction of heat-sensitive vitamins and loss of aroma are important issues in evaporation; however, the addition of vitamins and the use of aroma-recovery systems reduce the effect of those factors. (*See* **Evaporation**: Basic Principles; Uses in the Food Industry.)

Dehydration

In drying or dehydration, water is removed from the food by hot air or heated surface driers. Examples of the former include cabinet, tunnel, conveyor, and fluidized bed driers, and the latter include drum and vacuum shelf driers. As water in the surface evaporates it is replaced by water from inside the food by several mass transfer mechanisms, resulting in a reduction of the water content and water activity of the product. In addition to preserving the food, dehydration reduces the weight and bulk of the food, lowering transportation and packaging costs. Despite added convenience, dehydration also has a significant effect on the sensory properties of food. Some dehydrated products such as prunes or raisins are consumed in dehydrated form or used as ingredients in recipes; however, other dehydrated products such as dry milk or vegetables in a soup mix are reconstituted with water before consumption. (*See* **Drying**: Drying Using Natural Radiation; Fluidized-bed Drying; Spray Drying; Dielectric and Osmotic Drying; Hygiene.)

Freeze-drying or lyophilization is a method that accomplishes dehydration of the food by sublimating the water. When water sublimates it goes directly from a solid to a gas without passing through the liquid phase. Freeze-dried foods exhibit superior sensory and nutritive qualities when compared with products dehydrated by other methods. This method can be used to dehydrate high-value solid and liquid foods such as shrimp, strawberries, coffee, and juices, and it is also used by the pharmaceutical industry. (*See* **Freeze-drying**: The Basic Process.)

Refrigeration or Chilled Storage

Chilled storage in refrigerated chambers at temperatures above freezing is a widely used food preservation method. Refrigeration temperatures usually range from 0 to 7 °C in commercial and household refrigerators. The low temperatures lower the rate of metabolic reactions in unprocessed fruits and vegetables and other chemical reactions in foods. Microbial growth is usually slowed at refrigeration temperatures because metabolic reactions of microorganisms are enzyme-catalyzed and their rate depends on temperature. Refrigeration will preserve perishable foods for days or weeks, depending on the food. This method of preservation has very mild effects on sensory and nutritive attributes of products; however, it does not prevent food deterioration in the same degree and for as long as freezing or most other preservation methods. Refrigerated foods require low storage temperatures during transportation, retail, and home storage and their use is limited in rural areas of developing nations. (*See* **Chilled Storage**: Principles; Attainment of Chilled Conditions; Quality and Economic Considerations; Microbiological Considerations.)

Freezing

Foods were frozen in ancient times using ice and snow. Frozen storage cabinets were developed in the nineteenth century. Nowadays, food is frozen using several types of industrial mechanical refrigerators through cooled surfaces, and cooled liquid or air. Cryogenic freezing uses liquid nitrogen or carbon dioxide in solid or liquid form in direct contact with the food. Frozen food storage requires temperatures that maintain the food in frozen condition, usually −18 °C or less, and will preserve foods for months or years if properly packaged. A proportion of the water in the food is frozen and the concentration of solutes in unfrozen water increases, lowering the water activity. Freezing usually stops microbial growth but it does not destroy bacteria and molds. The parasite *Trichinella spiralis* and fish parasites are killed during frozen storage. Freezing temperatures significantly reduce the rate of chemical reactions in foods. Freezing has a low effect on nutritive quality of the food but sensory qualities, especially texture, may be affected by the formation of ice crystals. Fish and seafood, meats, fruits, and vegetables have been sold in frozen form for a long time. Baked goods and other prepared foods have become popular for their convenience, especially because of the use of microwave ovens in households. (*See* **Freezing**: Principles.)

Salting, Sugaring, and Curing

The addition of large amounts of salt or sugar to food is an old method of food preservation. Fish preservation by salting was used in Mesopotamia, Egypt, China, and the Mediterranean basin since ancient times. Meat was also preserved in old Mesopotamia by the addition of salt. Salt has also been used to preserve butter, cheese, and milk curds. Jams and preserves are the most common products preserved by adding sugar. When salt or sugar is added water moves from inside the cells to the outside solutes by osmosis, causing a partial dehydration of the cell, known as plasmolysis, that interferes with microorganism multiplication.

Both salting and sugaring have a significant effect in lowering water activity in food.

Curing is a method to preserve meats that also changes the flavor, color, and tenderness of the product. With more effective preservation methods available, the main purpose of curing is to produce characteristic products with unique flavor and to preserve the red color of meat after cooking. The main ingredients for curing or picking meat are sodium chloride, sodium nitrate and/or sodium nitrite, sugar, and spices. Sodium nitrate and nitrite have been linked with the formation of nitrosamines, compounds shown to be carcinogenic in animal studies. Nitrates and nitrites are naturally present in some vegetables and other foods and therefore cured meats may not be the main source of these compounds in human diets. (*See* **Curing**; **Jams and Preserves**: Methods of Manufacture; Chemistry of Manufacture.)

Smoking

Smoking is an old method of food preservation and it continues to be used today for fish and meats. The smoke is obtained by burning hickory or a similar wood and it contains formaldehyde and phenolic compounds that have antimicrobial properties. The heat also dries the food, increasing preservation. Smoking of foods is currently used more for its unique flavor properties than for its preservative action.

Chemical Preservatives

A number of chemical substances are used to inhibit the growth of microorganisms in foods. Sodium benzoate, sodium and calcium propionate, sorbic acid, ethyl formate, and sulfur dioxide are examples of commercially used food preservatives. Preservatives must undergo a review process that includes data on toxicological aspects prior to their approval. The allowed amount of a preservative in a product varies depending on the substance and the food. The pH level is also an important consideration as preservatives have an optimum pH range for their antimicrobial activity. Antioxidants prevent oxidative rancidity of fats and oils. Examples of commonly used antioxidants are butylated hydroxyanisole (BHA), butylated hydroxytoluene (BHT), tertiary butylated hydroquinone (TBHQ), and propyl gallate. Ascorbic acid (vitamin C) and tocopherols (vitamin E) are also used as antioxidants. (*See* **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; Synthetic Antioxidants, Characterization and Analysis; Role of Antioxidant Nutrients in Defense Systems; **Preservatives**: Classifications and Properties; Food Uses.)

Increased interest in natural antimicrobials has been noted in recent years due to toxicological concerns about synthetic preservatives and market trends that show increased popularity of foods with no artificial substances. Several naturally occurring substances have been identified and studied for their antimicrobial properties.

Role of Packaging in the Preservation of Foods

Adequate food packaging is essential and it must be used in combination with all food preservation methods. Packaging must protect the processed food against chemical attack, physical damage and contamination with microorganisms, insects, and rodents. Hermetic packaging refers to containers that are completely sealed against the ingress of vapors and gases and therefore also resistant to bacteria, yeasts, molds, and dust. Glass and metal cans are the most common hermetic containers. Nonhermetic containers may allow gas exchange but should provide protection against microbial contamination. Other packaging materials are plastics, paper, paperboard, and combinations of more than one material by lamination. Combinations are also achieved by simultaneous extrusion of two or more layers of different polymers to form a single film. Packaging must also be nontoxic, act as a barrier to moisture loss or gain, protect against oxygen and odor absorption, protect the food against ultraviolet light, be tamper-resistant or tamper-evident, and be compatible with the food. Food packages have other important functions in marketing the product and educating consumers on product use and handling. Food packages often carry the date of manufacturing or an expiration date, and some include a code that allows companies to trace the product in case of a recall. Expiration dates inform the consumer of the approximate shelf-life of the product. Shelf-life depends on the type of food, the processing method, packaging, and storage conditions. (*See* **Packaging**: Packaging of Liquids; Packaging of Solids.)

Active packaging is an innovative concept that has received a lot of attention in recent years. Active packaging may be defined as a type of packaging that changes the condition of the packaging to extend the shelf-life or improve safety or sensory properties while maintaining the quality of the food. Major active packaging techniques involve the absorption of oxygen, ethylene, moisture, odors, or carbon dioxide; they may also involve the release of antimicrobial agents, carbon dioxide, antioxidants, and flavors.

Controlled- or Modified-Atmosphere Storage and Packaging

Modification of the gas composition in storage rooms or inside the food packaging reduces the rate of respiration of fresh fruits and vegetables and also inhibits microbial and insect growth. Lowering respiration rates reduces biochemical and enzymatic activities that promote ripening and senescence. Gas modifications include oxygen reduction and increase in carbon monoxide or nitrogen. This method may also be combined with chilled storage, resulting in extended shelf-life and high quality of the products. (*See* **Chilled Storage**: Use of Modified-atmosphere Packaging; Packaging Under Vacuum.)

Aseptic Packaging

Aseptic processing and packaging is a sophisticated food preservation method where the food is sterilized or commercially sterilized outside the container and then placed in previously sterilized containers and sealed in an aseptic environment. Aseptic systems use ultrahigh-temperature (UHT) sterilization, a fast heating treatment at temperatures higher than pasteurization temperatures. Paper and plastic packaging materials are sterilized, formed, filled, and sealed in a continuous operation at the end of the processing line. Aseptic packaging is also used with metal cans, large plastic or metal drums, or large flexible pouches. Packages for aseptic processing may be sterilized using heat, chemicals, irradiation, or a combination of these methods.

Nontraditional or Alternative Methods of Food Preservation

Several alternative methods for food preservation are currently used or considered for future use. Some methods have been approved by regulatory angencies only for certain products and after a review of available scientific information on kinetics of microbial inactivation, toxicological aspects, and the effects of the treatment on the food products. The US FDA has identified several research needs applicable to alternative technologies. The following are brief descriptions of alternative food-processing methods evaluated by the FDA.

Irradiation

Irradiation is a preservation method where food is exposed to radiation. Irradiated food does not become radioactive. Machine or radionuclide radiation sources are used. Machine sources include electron acelerators and X-ray generators, and radionuclide sources include radioactive materials that give off ionizing gamma-rays. Bacteria, molds, yeasts, and insects are inactivated by irradiation. In 1963 the US FDA first approved irradiation for use on wheat and wheat flour. FDA requires that foods that have been irradiated bear both a logo and a statement that the food has been irradiated. The safety of irradiated foods has been studied extensively. (*See* **Irradiation of Foods**: Basic Principles.)

Microwave and Radiofrequency Processing

Electromagnetic waves of certain frequencies generate heat in foods by dielectric and ionic mechanisms. Microwave and radiofrequency heating have the advantage that they require less time than conventional heating, particularly for solid and semisolid foods. Industrial microwave pasteurization has been used for at least 30 years, but not radiofrequency processing systems. Food does not heat uniformly during microwave processing, and this remains an important issue when considering this technology. Several methods have been used to improve the uniformity of heating but equipment design can significantly influence processing parameters and establishing general conclusions has proved difficult.

Ohmic and Inductive Heating

In ohmic heating electric currents are passed through the food in order to heat it. This technology is also known as Joule heating, electrical resistance heating, electroheating, and electroconductive heating. Inductive heating is a process that induces electric currents within the food by the use of oscillating electromagnetic fields generated by electric coils. Microbial death kinetics data have been published for ohmic heating but not for inductive heating. The main advantage of ohmic heating is its ability to heat materials in a fast and uniform fashion, including products with particulates. Potential future uses of ohmic heating are in dehydration, evaporation, blanching, and extraction.

High-Pressure Processing (HPP)

HPP is also known as high hydrostatic pressure (HHP) or ultrahigh pressure (UHP) processing. During HPP, liquid or solid foods are subject to pressures between 100 and 800 MPa, at temperatures below 0 °C to above 100 °C and for times ranging from a millisecond pulse to more than 20 min. Temperatures ranging from 45 to 50 °C during treatment appear to increase the inactivation of pathogens and spoilage microorganisms. Temperatures in the range of 90–100 °C, together with pressures from 500 to 700 MPa, have been used to inactivate spore-forming bacteria. The effect of HPP works instantaneously

and uniformly throughout the mass of food, independent of size, shape, and food composition. The temperature of the food increases by approximately 3 °C per 100 MPa of applied pressure. Water activity and pH are critical product factors in microbial inactivation by HPP. Process factors that affect HPP include pressure, time at pressure, time to achieve treatment pressure, decompression time, treatment temperature, product initial temperature, vessel temperature distribution at pressure, packaging material integrity, and concurrent processing aids.

Pulsed Electric Fields (PEF)

This method involves the application of pulses of high voltage (20–80 $kV\,cm^{-1}$) to foods placed between two electrodes. Most PEF systems are at the laboratory phase with not many commercial systems in use. Application of PEF is restricted to foods that can withstand high electric fields, have low electrical conductivity, and do not contain or form bubbles. Although various theories have been proposed to explain the microbial inactivation by PEF, the most studied are electrical breakdown and electroporation. Process factors that affect microbial destruction by PEF include electric field intensity, pulse width, treatment time, and temperature.

Ultraviolet (UV) Light

This type of processing involves the application of radiation from the UV region of the electromagnetic spectrum. Microbial inactivation occurs by DNA mutations upon absorption of the UV light and exposure must be at least 400 $J\,m^{-2}$ in all parts of the product. Critical factors during UV light processing include the transmissivity of the product, the radiation path length, the geometric configuration of the reactor and the power, the wavelength, and physical arrangement of the UV source(s). UV light is used for bottled-water processing and for sanitizing food contact surfaces. There has been an increased interest in using UV light as a preservation method for fruit juices to replace pasteurization and other thermal treatments that have more impact on sensory attributes. (*See* **Ultraviolet Light**.)

Other Alternative Technologies

High-voltage arc discharge, pulsed light technology, oscillating magnetic fields (OMF), ultrasound, and pulsed X-rays are other types of technology that have been explored for their potential to inactivate microorganisms and may show promise for use in food preservation; however, more extensive research is needed at this time.

See also: **Antioxidants**: Natural Antioxidants; **Browning**: Enzymatic – Biochemical Aspects; **Canning**: Principles; **Chilled Storage**: Principles; Microbiological Considerations; Use of Modified-atmosphere Packaging; Packaging Under Vacuum; ***Clostridium***: Botulism; **Curing**; **Drying**: Drying Using Natural Radiation; Fluidized-bed Drying; Spray Drying; Dielectric and Osmotic Drying; Hygiene; **Evaporation**: Basic Principles; Uses in the Food Industry; **Fermented Foods**: Origins and Applications; **Food Poisoning**: Classification; **Food Safety**; **Food Security**; **Freeze-drying**: The Basic Process; **Freezing**: Principles; **Hazard Analysis Critical Control Point**; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; **Irradiation of Foods**: Basic Principles; **Mycotoxins**: Classifications; **Oxidation of Food Components**; **Pasteurization**: Principles; **Pickling**; **Preservatives**: Classifications and Properties; **Spoilage**: Bacterial Spoilage; Molds in Spoilage; Yeasts in Spoilage; **Storage Stability**: Mechanisms of Degradation

Further Reading

Borgstrom G (1968) *Principles of Food Science*, vol. 1. *Food Technology*. London: Collier-MacMillan.

Buffler CR (1993) *Microwave Cooking and Processing*. New York: Van Nostrand Reinhold.

Connor JM and Schiek WA (1997) *Food Processing, An Industrial Powerhouse in Transition*, 2nd edn. New York: John Wiley.

deMan JM (1999) *Principles of Food Chemistry*, 3rd edn. Gaithersburg, MD: Aspen.

Fellows PJ (2000) *Food Processing Technology*. Cambridge, UK: Woodhead Publishing.

Food and Drug Administration (2000) *Code of Federal Regulations, Title 21*. Washington, DC: US government.

Food and Drug Administration (2000) *Kinetics of Microbial Inactivation for Alternative Food Processing Technologies: Executive Summary*. Washington, DC: US government.

Heldman DR and Hartel RW (1997) *Principles of Food Processing*. New York: Chapman and Hall.

Jay JM (1998) *Modern Food Microbiology*, 5th edn. Gaithersburg, MD: Aspen Publishers.

Potter NN and Hotchkiss JA (1995) *Food Science*. New York: Chapman and Hall.

Rhaman MS (ed.) (1999) *Handbook of Food Preservation*. New York: Marcel Dekker.

VanGarde SJ and Woodburn M (1994) *Food Preservation and Safety: Principles and Practice*. New York: Chapman and Hall.

PRESERVATIVES

Contents

Classifications and Properties

B L Wedzicha, University of Leeds, Leeds, UK

Background

Preservatives are substances added to food to inhibit microbial spoilage. Chemical food spoilage by enzymatic and nonenzymatic mechanisms may be controlled by specific additives, e.g., antioxidants and antibrowning agents, which are described in the relevant articles. Substances such as common salt, sugar, vinegar, or spices, which are effective antimicrobial agents under appropriate conditions, are not regarded as food preservatives. Reactions of food preservatives, which are not part of the chemistry of food preservation but nevertheless affect food quality, will also be considered. The chemical reactions that lead to cured meat color and flavor are reviewed elsewhere.

Need for Preservatives

The use of preservatives to extend the life of foods and improve their safety has been recognized for thousands of years. Traditional methods of preserving foods have relied on fumigation by sulfur dioxide, curing with brines containing nitrites and nitrates, and smoking, in which phenolic antimicrobial agents are added to the food. These processes not only ensure that stored food is of acceptable microbiological quality but also frequently provide it with unique organoleptic characteristics.

Food preservatives ensure the safety and organoleptic quality of food when sold in the type of marketing environment now found in the developed countries. A particular problem associated with multiuse packs of food, e.g., beverages and spreads, is that an initially sterile, pasteurized or hygienically assembled, food becomes infected with bacteria, fungi, and yeasts, and begins to spoil. The instability might arise because the food has been formulated in such a way that it has little or no natural protection against microorganisms. For example, a low-fat spread contains significantly larger water droplets than butter and can, therefore, support microbial growth more readily. A preservative is particularly useful in such a product. A relatively short increase in storage life can be of significant economic and practical advantage. This is well illustrated with advantages gained by extending the shelf-life of bread by a few days.

Food preservatives are often used in conjunction with other methods of food preservation, e.g., some preservatives allow a milder thermal treatment to be applied for the preservation of food with consequent improvement in textural and nutritional properties.

Structures of Food Preservatives

Substances commonly used to preserve foods are listed in Table 1. These represent a wide range of molecular structures and tendency to form ions. The carboxylic acids sorbic, benzoic, formic, acetic, lactic,

Table 1 Substances recognized as suitable for use as food preservatives

Sorbic acid and Na^+, K^+, Ca^{2+} salts	E200–E203
Benzoic acid and Na^+, K^+, Ca^{2+} salts	E210–E213
Ethyl, methyl, and propyl *p*-hydroxybenzoate and Na^+ salts	E214–E219
Sulfur dioxide, Na^+ and Ca^{2+} sulfites and hydrogen sulfites (bisulfites), Na^+ and K^+ disulfites (metabisulfites or pyrosulfites)	E220–E227
Biphenyl	E230
Orthophenylphenol	E231
Thiabendazole	E233
Formic acid and Na^+, Ca^{2+} salts	E236–E238
Na^+ and K^+ nitrite and nitrate	E249–E252
Acetic acid and Na^+ and K^+, Ca^{2+} salts	E260–E263
Lactic acid	E270
Propionic acid and Na^+, K^+, Ca^{2+} salts	E280–E283
Hydrogen peroxide	
Nisin	

Those that are approved for use within the EC have their E-number shown alongside. The presence of any substance in this list does not imply that its use is permitted throughout the world. Biphenyl, orthophenylphenol, and thiabendazole are used for surface treatment.

and propionic are monobasic. Aqueous solutions of sulfur dioxide are commonly referred to as sulfurous acid, despite the fact that H_2SO_3 does not exist in measurable amounts. It is established that SO_2 does not interact significantly with water molecules, and the term 'molecular SO_2' is more appropriate. Aqueous solutions of SO_2 behave as solutions of a dibasic acid. The hydrogen sulfite ion is also in equilibrium with the disulfite ion. Significant amounts of disulphite ion are formed in concentrated systems. When sodium or potassium disulfite is used as a food preservative, the salt is hydrolyzed to the hydrogen sulfite ion.

The nitrites are salts of nitrous acid, which is also the source of the nitrosating species, N_2O_3, and oxides of nitrogen, particularly NO.

Regardless of the chemical form of a preservative when it is added to food, its ionic state is determined largely by the pH of the food and the pK_a of the acid. The pK_a values of those preservatives that have an acid group capable of ionizing within, or close to, the pH range of foods are given in Table 2. The fraction, x, of an acid that remains undissociated at a given pH may be found using

$$x = (10^{(\mathrm{pH}-\mathrm{p}K_a)} + 1)^{-1}.$$

Since significant changes in x occur in the range (pK_a − 2) to (pK_a + 2), the proportions of ionized and unionized forms of the acids listed in Table 2 change substantially over the pH range of foods. However, the hydroxybenzoate esters do not ionize significantly in the pH range of foods and are, therefore, uncharged.

An essential feature of food preservatives is that their molecules should have an appropriate balance of lipophilic and hydrophilic behavior, so as to be capable of traversing relatively nonpolar membranes and yet be sufficiently soluble in the aqueous environment of the microorganisms. This applies to the undissociated carboxylic acids, which will consequently tend to *partition* between the aqueous and oil phase of food. The ionized acids are relatively insoluble in nonaquous environments. A partition coefficient, P, is defined by

$$P = C_{\mathrm{oil}}/C_{\mathrm{aq}},$$

where c refers to the concentration of undissociated acid in each phase. Values of P vary from 0.17 and 2.5 for propionic and sorbic acids, respectively, to 26 and 88 for the ethyl and propyl esters of *p*-hydroxybenzoic acid. The latter two preservatives therefore show a marked bias for the oil phase. Numerous surfactants exist in solution as micellar structures, which are also found in the lipid bilayers of cell membranes. The interior of such structures is markedly hydrophobic, and carboxylic acids are able to partition into micelles. When expressed in the same way as the oil–water partition coefficient, but using mole fractions instead of concentrations, values of the micellar partition coefficients for benzoic and sorbic acids are in the range 1000–2000. The presence of an oil phase and/or surfactants in solution can significantly reduce the availability of food preservatives.

Table 2 pK_a values of commonly used food preservatives

Sorbic acid	4.76
Benzoic acid	4.18
Sulfur dioxide	1.86
Formic acid	3.75
Nitrous acid	3.40
Acetic acid	4.76
Lactic acid	3.08
Propionic acid	4.88

Mechanism of Antimicrobial Action

Carboxylic Acids and Esters

The antimicrobial action of a food preservative depends on the concentrations of both the ionized and unionized forms, and the specific efficacy of each. In general, the undissociated acid is the better antimicrobial agent, but the presence of the ionized form cannot be neglected, particularly when, at high pH, its concentration may be much larger than that of the undissociated acid. The minimum concentration of an undissociated acid required to inhibit microorganisms is generally one to two orders of magnitude smaller than that of the corresponding anion.

There is no general theory to explain the mechanisms whereby these carboxylic acid and ester preservatives exert their effect. Transport of these preservatives across cell membranes is passive, but accumulation in the cell causes a reduction in the protonmotive force through the membrane and a reduction in pH within the cell. These contribute to the cessation of cell growth or cell death. Certain preservatives could also exert specific effects on metabolic enzymes. Sorbic acid is alleged to react with the sulfhydryl groups of fumarase in catalase-positive bacteria, molds and yeasts, and aspartate, succinic, and yeast alcohol dehydrogenase. It is suggested that this might apply to sulfhydryl-containing enzymes in general. Benzoate inhibits enzymes in oxidative phosphorylation and at various stages in the tricarboxylic acid cycle.

Microorganisms are able to metabolize some food preservatives when present at sublethal concentrations. A significant example is the conversion of

sorbic acid to hexadienol by certain strains of lactic acid bacteria. This product can react with ethanol to form 1-ethoxy-2,4-hexadiene and 2-ethoxy-3,5-hexadiene, which give rise to a geranium-type odor, occasionally detected in wines treated with the preservative.

Sulfur Dioxide and Sulfites

Molecular sulfur dioxide passes into yeast cells by passive transport and possibly into some other microbial cells by active transport. As with carboxylic acids, it interferes with membrane transport processes. However, sulfites are much more chemically reactive than any of the carboxylic acids referred to above. Thus, the sulfite ion acts as a powerful nucleophile causing the cleavage of disulfide bonds of proteins; it reacts with coenzymes (NAD^+), cofactors, and prosthetic groups (flavin, thiamin, heme, folic acid, and pyridoxal). Consequently, a broad range of metabolic enzymes are inactivated, and structural proteins that contain disulfide bonds may be denatured. Prior to the death of yeast cells treated with sulfite, there is a rapid decrease in ATP content. This has been attributed to the inactivation of glyceraldehyde-3-phosphate dehydrogenase by sulfite species. The additive also reacts with carbonyl constituents of the metabolic pool, to form hydroxysulfonates. When treated with sublethal concentrations of sulfite, yeasts tend to excrete increased amounts of acetaldehyde. This is due to the trapping of this metabolic intermediate as the particularly stable hydroxysulfonate, thereby preventing its conversion to ethanol. Glycerol is formed instead of ethanol by reduction of glyceraldehyde-3-phosphate to glycerol-3-phosphate, which is subsequently dephosphorylated.

Nitrite Ion

Nitrite ion is believed to exert its action on the phosphoroclastic system of enzymes, which causes the conversion of pyruvate to acetate and is a source of ATP. Evidence for this involvement comes from the fact that in the presence of nitrite, the ATP concentration in the cell rapidly decreases, and pyruvate is excreted. Nitrite ion is in equilibrium with nonionic forms, including NO. This is a good ligand for iron, and the mechanism of inhibition is likely to be the reaction of NO with the nonheme iron-containing pyruvate:ferredoxin. In aerobic bacteria, the heme iron of cytochrome oxidase is also a likely target for NO. Inhibition of enzymes with sulfhydryl groups as a result of the formation of *S*-nitroso products is also possible at relatively high nitrite concentrations.

When nitrite-containing bacteriological media are heated, they become more inhibitory towards *Clostridium botulinum* than media that have been heated prior to the addition of the preservative. This so-called Perigo effect can be modeled using mixtures of cysteine, nitrite, and iron(II) salts. When heated, such mixtures give rise to iron–sulfur-bridged complexes, of which Roussin's black salt is a well-known example. These compounds have been shown to be effective inhibitors of clostridial spores.

Nisin

Nisin is a polypeptide (molecular weight 3500) that usually exists as a dimer. It is produced by *Lactococcus lactis* and may be formed naturally in cheese. The polypeptide chain contains L-amino acids and the unusual sulfur-amino acids lanthionine and β-methyl-lanthionine. A specific enzyme, nisinase, has been widely reported, and its formation is possibly one reason why many lactic acid bacteria inactivate nisin. The peptide is also susceptible to α-chymotrypsin but not to pepsin, trypsin, and carboxypeptidase A, among other proteolytic enzymes.

Chemical Reactivity of Preservatives towards Food Components

Chemical reactions between food preservatives and components of microbial cells, or with food components where there are implications with regard to antimicrobial action, have been described above. However, some food preservatives, and particularly sorbic acid, sulfur dioxide, sulfites, and nitrite ions are capable of more extensive reactivity with food components. This may lead to the formation of reaction products of toxicological importance and a reduction in the concentration of available preservative.

Sulfur Dioxide and Sulfites

It is well known that the concentration of sulfite species in a food decreases with time, and the shelf-life may be limited by this reactivity. The conversion of sulfite to sulfate through autoxidation is catalyzed by transition metal ions and involves the formation of oxidizing free radicals. There is evidence to suggest that the facile autoxidation of ascorbic acid, which is less easily controlled by antioxidants, may promote the oxidation of sulfite in foods containing the vitamin. The free-radical intermediates ($^{\bullet}OH$, $^{\bullet}O_2^-$) in sulfite autoxidation are known for their ability to cause the oxidation of unsaturated organic compounds in the absence of suitable antioxidants. Free-radical sulfonation of unsaturated organic compounds by $^{\bullet}SO_3^-$ is also possible. An important nucleophilic reaction of the sulfite ion is its addition

to the α,β-unsaturated carbonyl moiety of 3,4-dideoxyosulos-3-enes formed as reactive intermediates in Maillard and ascorbic acid browning. This reaction can cause a considerable depletion of the additive in foods susceptible to these forms of nonenzymatic browning.

Nitrite

The reaction of secondary amines with nitrosating species derived from nitrite to give *N*-nitroso compounds is very well known and has been given much publicity on account of the possible carcinogenicity of the products. The reaction proceeds at a maximum rate in the pH range 2.25–3.4 and is catalyzed by weakly acid anions, notably thiocyanate. In meat products, the reaction is most probably associated with the adipose tissue, and it is suggested that nonionic amine and the nitrosating species N_2O_3 partition into the nonaqueous phase where a facile reaction takes place. A similar explanation has been suggested for the catalytic effect of surfactants and cell membranes on nitrosation, but in this instance, the reaction takes place within the micellar environment. *C*-Nitrosation of phenolic components of food, particularly in smoked products, leads to the formation of nitrosophenols, which can readily oxidize to the corresponding nitro compounds. The *C*-nitrosation of activated methylene groups is also established; suitable reactants in food include the 3-deoxyosuloses formed as intermediates in Maillard and ascorbic acid browning reactions, when the product is an oxime. *S*-Nitroso compounds are readily formed by the nitrosation of thiols and represent a reversibly bound form of the preservative.

Chemical Interactions between Preservatives

Sulfite species, nitrite, and sorbic acid are chemically incompatible. Sulfite ion reacts readily with nitrite ion to form the imidodisulfonate ion. The mechanism of the reaction between sorbic acid and sulfite species depends on whether or not oxygen is present. Under aerobic conditions, a sulfite-mediated oxidation of sorbic acid takes place. Under anaerobic conditions, a much slower nucleophilic addition of sulfite ion gives 5-sulfo-3-hexenoic acid. Mixtures of sorbic acid and nitrite ion give rise to ethylnitrolic acid and 1,4-dinitro-2-methylpyrrole, both of which give a positive result in the Ames mutagenicity test.

See also: **Acids**: Properties and Determination; Natural Acids and Acidulants; **Nisin**; **Nitrates and Nitrites**; **Preservatives**: Analysis

Further Reading

Branen LA, Davidson PM, Salminen S and Thorngatwe III JH (eds) (2002) *Food Additives*. New York: Marcel Dekker.

Davidson PM and Branen LA (eds) (1993) *Antimicrobials in Foods*. New York: Marcel Dekker.

Gould GW (ed.) (1989) *Mechanisms of Action of Food Preservation Procedures*. London: Elsevier Applied Science.

Lueck E (1980) *Antimicrobial Food Additives. Characteristics, Uses, Effects*. Berlin: Springer-Verlag.

Wedzicha BL (1984) *Chemistry of Sulphur Dioxide in Foods*. London: Elsevier Applied Science.

Wedzicha BL (1988) Distribution of low-molecular weight food additives in dispersed systems. In: Dickinson E and Stainsby G (eds) *Advances in Food Emulsions and Foams*, pp. 329–371. London: Elsevier Applied Science.

Food Uses

B L Wedzicha, University of Leeds, Leeds, UK

Factors Affecting Choice

General Considerations

The type of preservative applicable to a particular need is determined by the composition of the food, the type of microbial spoilage that takes place, and the desired shelf-life. Important compositional variables include pH, since the efficacy of a given preservative generally decreases with increase in pH. The maximum pH at which a preservative is useful is often quoted. Some components of food can also potentiate the action of preservatives, e.g., chelating substances such as ethylenediaminetetraacetic acid or critic acid can render preservatives more effective and at the same time protect the food from other forms of spoilage. Salt is a common food component that acts synergistically with several food preservatives. Ideally, the use of preservatives should be regarded within a 'preservative systems' approach to include interactions, rather than in terms of individual substances. This concept is extended to include other methods of food preservation, e.g., the use of thermal processes in conjunction with chemical preservatives, to optimize the use of each. Such effects of food components and processing on antimicrobial activity also make theoretical predictions of the efficacy of a given preservative unreliable, and the levels of application are usually decided upon empirically.

Preservatives are also chosen for their specific physical properties such as solubility in particular foods and ease of handling. Calcium sorbate, for instance, is sparingly soluble in water (1.2 g per 100 g water) and is suitable for the surface treatment of foods such as cheeses as it does not quickly dissolve in surface moisture or migrate into the cheese. Chemical stability is important with regard to the handling of sulfites; the disulfites (metabisulfites) tend to be the most stable sulfite species towards autoxidation and dissolve in water to give solutions with pH values in the middle of the pH range of foods.

Cost may be an important consideration; sulfites are the cheapest to use, whilst sorbic acid, sorbates and the *p*-hydroxybenzoate esters are the most expensive.

Spectrum of Antimicrobial Action

In general, benzoic acid, its salts and esters of *p*-hydroxybenzoic acid have the broadest spectrum of antimicrobial activity and are useful against many spoilage bacteria, fungi, and yeasts. Benzoic acid cannot, however, be relied upon to preserve foods against bacteria effectively because, whilst food poisoning and spore-forming bacteria are inhibited at normal levels of use in low pH foods, many spoilage bacteria are more resistant. Benzoic acid is most suitable for foods at pH 2.5–4, whilst the esters of *p*-hydroxybenzoic acid may be effective to pH 7 or higher. However, the latter tend to be less effective against bacteria and particularly Gram-negative organisms. Thus, in relatively acidic products where yeasts and molds are the greater causes of spoilage, benzoic acid and its salts offer an effective means of preservation.

Sorbic acid and its salts are useful additives against yeasts and molds but are less effective against bacteria. They provide antimicrobial activity up to pH 6.5 and are good preservatives for foods with a high fat content and pH, e.g., low-fat spreads, processed cheeses; the *p*-hydroxybenzoate esters tend to be too soluble in the nonaqueous component of such foods, whilst benzoate is relatively ineffective at the higher pH values.

Sulfur dioxide and sulfites also represent broad-spectrum antimicrobial agents in acidic foods, particularly beverages. In general, this preservative is more effective against bacteria than molds and yeasts, but Gram-positive bacteria are less susceptible than Gram-negative bacteria. The effect against yeasts is greatest around pH 3.5 and falls off markedly as the pH is raised above this value. There is a tendency for desirable fermentation yeasts to show a greater resistance to this preservative than undesirable 'wild yeasts' and lactic acid bacteria. This is exploited in the use of sulfites during wine-making. Despite the relatively high pH associated with meat products such as the British sausage (often pH 6–6.5), sulfites tend to inhibit enterobacteriaceae, and particularly salmonellae, in this environment. In some situations, sulfites are added also to inhibit chemical spoilage, notably enzymic browning caused by oxidation of *o*-diphenols, other forms of oxidative spoilage, and nonenzymatic browning of reducing sugars and ascorbic acid.

Propionic acid and its salts are effective against molds despite the fact that some species of *Penicillium* can grow on media containing as much as 5% propionic acid. Yeasts and most bacteria are generally less affected. However, a specific target for propionic acid is the rope-forming *Bacillus subtilis* in bread, which is effectively controlled at pH 6. The preservation of a wide range of foods with vinegar (acetic acid) has been known since ancient times; acetic acid is more effective against yeast and bacteria than molds. Its effect on bacteria is, however, predominantly due to the associated reduction in pH. *Acetobacter* and some lactic acid bacteria are unaffected by this preservative. Lactic acid is similarly a preservative associated with traditionally prepared and particularly fermented foods. As with acetic acid, its antimicrobial action is mainly due to a reduction of pH to below that at which bacteria can grow. It does, however, inhibit spore-forming bacteria at pH 5.0 but is ineffective against yeasts and molds.

The most important effect of nitrite ion in food is its inhibition of *Clostridium botulinum* in cured meat products. Whilst it also has an inhibitory effect towards a wide range of different species of bacteria, some, particularly *Salmonella*, lactobacilli, and *Clostridium* perfringens, tend to be more resistant. Nitrite and nitrate are essential for the development of cured meat color and flavor.

Thiabendazole, orthophenylphenol, and biphenyl inhibit fungi and are suitable only for surface application.

Specific Applications

Beverages

Nonalcoholic beverages, including fruit juices, are usually preserved with benzoic acid and its salts, esters of *p*-hydroxybenzoic acid, sorbic acid and its salts, or the sulfites. Typical concentrations of benzoic acid required are in the range of 500–1000 mg kg^{-1}. In the event that these levels lead to a noticeable taste, benzoic acid is replaced wholly or partly by mixtures of methyl and propyl esters of *p*-hydroxybenzoic acid (usually 2:1, methyl:propyl) at a combined

concentration of 300–500 mg kg^{-1}, or sorbic acid and its salts at 200–1000 mg kg^{-1}. Whilst sulfites are effective antimicrobial agents in beverages at concentrations of 50–500 mg of SO_2 per kilogram, they tend to undergo significant oxidation when beverage containers are repeatedly opened and air admitted. A combination of sulfite with another preservative, e.g., sorbic or benzoic acid, is frequently used for fruit juices such that the sulfite acts to control forms of chemical spoilage, and lactic and acetic acid fermentation, whilst the second preservative acts as a longer-lasting agent against yeasts and molds. A significant disadvantage of adding sulfites to beverages colored with anthocyanins is that these are decolorized, even at low concentrations of preservative, and additional sulfite-stable food-color agents are required. Fruit juice concentrates used for the manufacture of normal-strength juice may contain any of the preservatives used in the final beverage. In practice, it is not feasible to add sufficiently high concentrations of the carboxylic acid or ester preservatives to the concentrate to provide an acceptable level of additive in the finished product. In general, such concentrates need to be protected against browning, which demands high levels (e.g., 2.5–10 g of SO_2 per kilogram) of sulfite; this also acts as a preservative. In any case, the concentration of preservative in the finished product is a result of further addition after dilution and blending.

Sulfites are the major preservatives for alcoholic beverages. Their most important function is the control of the so-called wild yeasts in grape juice, and after fermentation, the preservative is required at levels of 50–150 mg of SO_2 to inhibit further microbial action. A specific problem associated with the use of sulfites to preserve wine, cider, and perry is that fruit that has been affected by molds may contain significant amounts of acetaldehyde, pyruvic acid, α-ketoglutaric acid, and other aldehydes and ketones, which are able to bind the preservative reversibly as hydroxysulfonate adducts. When present in this form, sulfite is unable to act as a preservative; however, the concentration of hydroxysulfonates is included in the analytical concentration of the preservative in the beverage. When grapes affected by *Botrytis* are used for wine-making, as much as 80% of the total preservative can be found as hydroxysulfonate adducts. During fermentation in the presence of sulfites, yeasts form sulfite-binding compounds and particularly acetaldehyde. Sorbic acid is also a good preservative for wine, and despite its chemical reactivity towards sulfite, mixtures of the two are advocated; in such a situation, sorbic acid serves to prevent further fermentation, whilst the sulfite protects against chemical and bacterial spoilage. However, depletion of sulfite through chemical reactions could result in malolactic bacteria metabolizing sorbic acid to intermediates, which ultimately lead to off-flavors. In general, potassium sorbate is the preferred salt on account of its higher solubility than that of the sodium salt. In some wines, however, potassium tartrate may precipitate, in which case, the sodium salt is used.

Fruit Products

Sulfur dioxide is an effective fumigant in the control of postharvest decay of grapes caused by the fungus *Botrytis cinerea*. Raspberries may be similarly treated for *Botrytis* and *Cladosporium*. Another unique application of sulfite is for the storage, as pulp, of soft fruit to be used for jam manufacture. A typical 'sulfite liquor' would contain up to 3000 mg of SO_2 per kilogram of solution and is capable of preserving fruit for up to 2 years with a high retention of ascorbic acid. Whilst the anthocyanins are bleached in this process, the color of the fruit reappears as the preservative is 'lost' during jam-making. An endopolygalacturonase enzyme from *Rhizopus sexualis* is not inactivated by sulfite and can cause the breakdown of strawberries in pulp. Jam and related products, e.g., fruit purees and pie fillings, may be successfully preserved with benzoic acid (1000 mg kg^{-1}), *p*-hydroxybenzoate ester (methyl:propyl ester, 3:1; 700 mg kg^{-1}), sorbic acid (800–1500 mg kg^{-1}), and sulfite (100 mg of SO_2 per kilogram in jam).

Dried fruits are frequently prepared with the aid of sulfite as an antibrowning agent. In such products the concentration of additive may be as high as 2500 mg of SO_2 per kilogram, and this will also effectively preserve the product. When a relatively moist unsulfited product is desired, a preservative is necessary, and a potassium sorbate dip or spray to give 200–500 mg of sorbic acid per kilogram, or similar application of sodium benzoate at 1000 mg of benzoic acid per kilogram is effective.

Vegetable Products

One of the most common traditional methods of preserving vegetables is by pickling with vinegar, acetic acid contributing to the reduction of pH, a specific antimicrobial effect and a characteristic taste. Typically, raw vegetables are immersed in 0.5–3% acetic acid solution, but some yeasts and molds are still capable of causing long-term spoilage. The combined use of acetic acid with benzoic acid, *p*-hydroxybenzoate ester mixtures (methyl:propyl, 2:1) or sorbic acid (all up to 1000 mg kg^{-1}) allows lower acetic acid concentrations and safeguards against spoilage. In sweet relishes, sorbic acid is more effective than benzoic acid on account of the relatively

high pH. The selective control of yeasts, molds, and putrefactive bacteria is the reason why sorbic acid is useful at levels of 500–2000 mg kg^{-1} in vegetable fermentations, where it allows the growth of the desired lactic acid-producing organisms.

Fresh vegetables are generally dipped in sulfite solutions to prevent enzymatic browning, though this also protects against microbial spoilage. It is feasible to preserve vegetarian burger-type products with sulfite. However, dehydrated vegetables are microbiologically stable but require sulfite as an anti-browning agent for the dehydration process and subsequent storage, at a final level of some 2500 mg of SO_2 per kilogram of dried product.

Baked Products

Propionic acid is widely used in the making of bread and cakes to inhibit surface mold and *Bacillus subtilis*, which causes rope. The sodium salt is used chiefly in confectionery products, whilst the calcium salt is used for bread. Levels of propionic acid up to 3000 mg kg^{-1} may be found in such applications. Sorbic acid is also an effective preservative for cereal products, at concentrations similar to those of propionic acid, and inhibits the mold *Trichosporon* found in rye bread and cakes. Unfortunately, sorbic acid also inhibits yeast. A potential solution could be to use sorboyl palmitate, which has no antimicrobial activity but degrades on heating, liberating sorbic acid. Mixtures of esters of *p*-hydroxybenzoic acid (methyl:propyl, 3:1) at 300–600 mg kg^{-1} are useful preservatives for cakes (particularly fruit cakes) but inhibit yeasts and cannot be used in bread. Some components of flour confectionery, e.g., icing, fillings, and marzipan, can be preserved with benzoic acid, esters of *p*-hydroxybenzoic acid, and sorbic acid.

Dairy Products

Apart from the small-scale use of hydrogen peroxide to preserve milk by reducing the severity of heat treatment required for sterilization, the major use of preservatives in dairy products is for cheese. Propionic acid is formed naturally during the ripening of certain cheeses but may be used as an added preservative for surface treatment to prevent mold growth. Nitrite ion, which is formed from added nitrate in cheese, inhibits fermentations caused by clostridia and coliforms such as *Enterobacter*, which are undesirable because they may give rise to the production of carbon dioxide with resultant blowing of the cheese. Nitrite ion is to be avoided in cheeses where the production of propionic acid is important because nitrite also inhibits the bacteria required for its formation. Nisin is a good antimicrobial agent for cheese because it is particularly effective against clostridia such as *Cl. tyrobutyricum* in cheeses with a higher pH. The egg white antimicrobial enzyme, lysozyme, is employed for a similar purpose. Cheesemaking represents a major use of sorbic acid, where it serves two purposes: to inhibit surface mold growth, achieved by dipping, dusting (preferably with calcium sorbate), or incorporation of the preservative into wax coatings or packaging films, or to protect packs of processed cheese after opening, by incorporation of the preservative into the cheese at levels of 500–700 mg of sorbic acid per kilogram.

Meat and Fish Products

The curing of meat is probably the best known use of nitrites and nitrates. They are used in conjunction with salt, which has the effect of reducing the water activity sufficiently to reduce spoilage by *Pseudomonas* and related organisms, whilst nitrite prevents the growth of germinating spores. Curing is regarded as protection against food poisoning caused by *Clostridium botulium*. Eventual spoilage in cured meat is caused by lactic acid bacteria if the salt content is low, or by micrococci and vibrios if it is high together with a variety of molds and yeasts. Off-flavors are associated with hydrolysis of fat. The storage life of cured meat is extended by smoking as a result of additional antimicrobial agents from the smoke; again, when the meat ultimately spoils, it is dominated by lactic acid bacteria. Typical concentrations of sodium nitrite and nitrate in cured meat are 50–200 and 500 mg kg^{-1}, respectively.

Sulfites are added to a small number of comminuted meat products (sausages, burgers) at levels of 600 mg of SO_2 per kilogram, to achieve a maximum of 450 mg of SO_2 in the finished product. Whilst the total sulfite content of such a product falls slightly during storage, a large proportion becomes reversibly bound to carbonyl compounds of microbial origin. The success of sulfite in this application is the result of its inhibition of salmonellae, whilst spoilage of sulfited meat is confined to Gram-positive microflora consisting of lactobacilli and *Brochothrix thermosphacta*. The result is that the spoilage is associated with a sour odor, unlike the spoilage of untreated meat, which gives rise to a 'putrid' smell. The use of sulfites in meat products leads to the destruction of thiamin and is not permitted in important meat sources of the vitamin. Mold growth on the surface of dry sausage, during the drying period, may be controlled by dipping the casings in a solution of potassium sorbate prior to stuffing.

Fish may be preserved in acetic acid (10–30 g kg^{-1}) by the traditional process of marinating. Whilst this

provides basic protection against pathogenic microorganisms, lactobacilli may grow, and the use of acetic acid should be combined with a mild thermal process or other antimicrobial agents. Benzoic acid is slightly more effective than sorbic acid in this application (1000 mg kg^{-1}), and esters of *p*-hydroxybenzoic acid (300–600 mg kg^{-1}) may also be used.

Fat Products

Although benzoic acid had been used for some time for the preservation of margarine at concentrations of 800–1500 mg kg^{-1}, it is far from ideal for this application owing to the relatively high pH of the food. However, it is a good preservative for mayonnaise, though improved performance against acid-producing bacteria can be obtained by using mixtures of benzoic and sorbic acids. Margarine and low-fat spreads can be preserved most effectively by means of sorbic acid.

Miscellaneous Uses

Food ingredients that are normally manufactured in the form of solutions, or have a water activity sufficient to sustain microbial growth, may be treated with preservatives. Examples include flavoring preparations, gelatin, malt extracts, and antifoam agents.

Carbon Dioxide

Carbon dioxide is not a preservative in the usual meaning of the word. However, in most of the modified-atmosphere-packaged foods, it does help to preserve the food. It specifically inhibits the growth of the typical fast-growing oxidative Gram-negative bacteria that quickly spoil food, e.g., meat and fish. Slower-growing, less obnoxious microorganisms, such as lactic acid bacteria, often then take over, but the shelf-life can be substantially extended. The extent of such inhibition by CO_2 is greatly enhanced as the temperature is reduced. It has been speculated that this results mainly from the increased solubility of the gas at low temperatures. The shelf-life increases for meat and fish can be by two to three times if chill control is effective.

See also: **Alcohol**: Properties and Determination; **Bread**: Chemistry of Baking; **Cakes**: Chemistry of Baking; **Fish**: Miscellaneous Fish Products; **Jams and Preserves**: Chemistry of Manufacture; **Meat**: Preservation; Extracts; **Preservatives**: Analysis

Further Reading

Branen LA, Davidson PM, Salminen S and Thorngatwe III JH (eds) (2002) *Food Additives*. New York: Marcel Dekker.

Davidson PM and Branen LA (eds) (1993) *Antimicrobials in Foods*. New York: Marcel Dekker.

Hayes PR (1985) *Food Microbiology and Hygiene*. London: Elsevier Applied Science.

Lueck E (1980) *Antimicrobial Food Additives. Characteristics, Uses, Effects*. Berlin: Springer-Verlag.

Wedzicha BL (1984) *Chemistry of Sulphur Dioxide in Foods*. London: Elsevier Applied Science.

Analysis

B L Wedzicha, University of Leeds, Leeds, UK

Background

Food preservatives constitute a group of compounds of widely different molecular structures; they are organic and inorganic substances with different functional groups and tendencies to form ions. There are no procedures that are generally applicable to the analysis of preservatives as a class of food additive; the procedures are specific to the preservative being analyzed. The lowest concentrations of commonly used preservatives are of the order of a few milligrams per kilogram of food, and, with few exceptions, recommended or statutory methods of analysis are designed to give a good accuracy at levels of 10 to > 1000 mg of preservative per kilogram of food. The question of the lower limit of detection is rarely an issue, unless it is desired to use small sample sizes, e.g., < 1 g, or to determine whether or not a food or its ingredients had been treated with a preservative. For solid foods, small sample sizes often lead to nonrepresentative sampling and should be avoided. Not all the procedures described constitute official methods of analysis. Frequently, for routine analysis, a food manufacturer would use a rapid or cheap analytical technique standardized against an official method. The official status of given procedures varies from country to country.

Organic and inorganic acid preservatives may be added in the form of the undissociated acid or a variety of salts. In food, the ionic composition is determined largely by concentration and pH, but it is generally impossible to predict this accurately for any given situation. In order to avoid complications with the specification of the amount of preservative in a food, this is usually referred to as the weight-for-weight concentration of the undissociated acid, e.g., benzoic acid, sorbic acid, or sulfur dioxide. Nitrite and nitrate levels are expressed in terms of the weight of the sodium salt.

There are, of course, a very large number of possible analytical procedures available for each preservative. Those given here represent a selection to illustrate the variety of methods recommended for use on food samples.

Carboxylic Acids and Esters

Extraction

All the carboxylic acid and ester preservatives are steam-volatile from acidified food samples, and steam distillation offers an effective means of separating them from the sample. The nonionic nature of these preservatives in acid solution also allows extraction with a variety of organic solvents, e.g., diethyl ether, and dichloromethane. Such organic solvents may also be used to isolate, or concentrate, preservatives obtained in large volumes of steam distillate. The subsequent detection and determination of preservatives depend on the substance to be analyzed. (*See* **Acids**: Properties and Determination.)

Analysis of Extracts and Distillates

Rapid screening of extracts and distillates for sorbic and benzoic acid and the esters of *p*-hydroxybenzoic acid is possible by thin-layer chromatography (TLC) on silica gel. If separated on plates containing an F_{254} (fluorescent indicator), these acids may be readily visualized under UV light. Acetic and propionic acids can be identified by paper chromatography using an acid–base indicator for visualization of spots.

The total acidity in steam distillates is an acceptable measure of the concentration of an acid preservative provided that only one such preservative is used and that a blank for the food sample is known. Propionic and acetic acids can be separated using a silica gel column with butanol–chloroform solvent and the eluate titrated for quantitation. Solvent extracts of food samples can only be subjected to direct spectrophotometric analysis with caution. For example, benzoic acid, extracted into diethyl ether from tomato products, jams, jellies, and soft drinks, may be determined by measuring its absorbance at 267.5, 272, and 276.5 nm, and using the average of the highest and lowest wavelength readings as the background value for the absorbance at 272 nm. In general, solvent extracts and steam distillates can be analyzed directly by reverse-phase high-performance liquid chromatography (HPLC) or by gas–liquid chromatography (GLC) after forming trimethylsilyl or methyl esters. Internal standards for chromatography are chosen to have extraction and chromatographic behavior similar to that of the preservative being analyzed. Standards are generally added to the food before extraction, and any derivatization is carried out on both the analyte and the standard. Examples of internal standards include phenylacetic and caproic acids for gas chromatographic analysis of benzoic and sorbic acids, respectively. (*See* **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography.)

Esters of *p*-hydroxybenzoic acid are determined in steam distillates or extracts either by reversed-phase HPLC or by direct spectrophotometry as *p*-hydroxybenzoic acid after saponification.

Oxidation of sorbic acid to malonaldehyde by means of dichromate and subsequent reaction with thiobarbituric acid (TBA) offers a specific nonchromatographic assay of this preservative in steam distillates from all types of food. Frequently, crude solvent extracts can be analyzed in the same way without any undue interference from other TBA-reactive substances either present in food or formed during dichromate oxidation.

Enzymatic Methods

Specific enzymatic procedures for the determination of acetic and formic acids are available. Acetic acid is determined by conversion to pyruvate (using ATP and coenzyme A in the presence of acetylcoenzyme A synthetase), which reacts in turn with oxaloacetate (citrate synthetase) obtained by the reduction of malate by NAD^+ (malate dehydrogenase). The rate of utilization of NAD^+ is a measure of acetic acid concentration. Formic acid is determined more simply by measuring the rate of utilization of NAD^+ during the oxidation of the acid to hydrogen carbonate (formate dehydrogenase). In general, sample preparation for enzymatic methods involves the preparation of aqueous extracts of solid foods, separating fat, precipitating protein (perchloric acid), decolorizing strongly colored samples (charcoal), and adjusting to pH 7–8.

Sulfur Dioxide and Sulfites

Classification

Sulfur dioxide and sulfite are considered to exist in food in two forms: free and bound. The former includes all species of sulfur in oxidation state +4, i.e., SO_2, HSO_3^-, SO_3^{2-}, and $S_2O_5^{2-}$. The latter term is used to describe that preservative which is in the form of hydroxysulfonate (carbonyl-HSO_3^-) adducts. Legislation requires that total (free + bound) preservative be determined. In food processing, the concentration of free additive determines its

antimicrobial effect and is frequently measured. A specific problem that exists with this preservative is that itsmeasured concentration falls with time as a result of reactions with food components, and as a result of autoxidation after packaged foods are opened and exposed to air. Samples must be analyzed without delay and exposure to air minimized.

A broad classification of analytical methods for sulfur dioxide and sulfites is based on whether or not a separation procedure is involved. Methods are usually referred to as direct, when a specific analytical reaction is applied to the whole food sample, and indirect, when the additive is recovered before determination.

Direct Methods of Analysis

The simplest analysis of sulfites in foods involves iodimetric titration after allowing for a blank representing the reactivity of the food sample itself towards iodine. This is normally found by adding an aldehyde or ketone (formaldehyde, acetaldehyde, acetone) to combine with free sulfite before the blank titration. The free sulfite content of a food is determined by acidifying the sample and titrating with iodine. The total (free + bound) sulfite may be determined by raising the pH of the sample to pH 10 to decompose any hydroxysulfonates before acidifying and titration. Iodimetric titration is applicable to beverages and aqueous extracts/homogenates of solid foods. Dark-colored beverages may be titrated using electrometric detection of the end point.

Sulfites in solution may be determined spectrophotometrically by reaction with tetrachloromercurate + pararosaniline + HCHO reagent, 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB), or malachite green. These reactions depend on the nucleophilic behavior of the sulfite ion and suffer interference from nucleophiles in foods, particularly thiols. Despite this, the pararosaniline method has been used widely for direct determination of the preservative in beverages, solutions of sugars, and aqueous extracts or homogenates of solid foods including fruits and vegetables, and has been adapted successfully for an 'autoanalyzer' system.

Few chromatographic methods of analysis of sulfites in foods have gained acceptance, mainly because of the problems associated with detection of the species. Ion-exclusion chromatography of the sulfite ion with amperometric detection is the best-established chromatographic method available and can be applied directly to beverages and aqueous extracts of foods.

Other direct analyses include the use of polarography and sulfur dioxide-sensitive electrodes but are not in widespread use. Enzymatic methods of analysis make use of the oxidation of sulfite to sulfate + H_2O_2 (sulfite oxidase) and assay of H_2O_2 with NADH (NADH-peroxidase). The rate of formation of NAD^+ is used to determine the sulfite concentration. Sulfite oxidase may also be immobilized at the tip of an oxygen-measuring electrode for a direct indication of the rate of utilization of oxygen by the enzyme in sulfite oxidation. The pH requirement for these enzymatic analyses is slightly alkaline (pH 7.5–8), and the method can give only the total additive concentration.

Indirect Methods of Analysis

All indirect methods of analysis of sulfur dioxide and sulfites in foods involve the conversion of the ionic forms to gaseous SO_2 as the method of separating the additive from the food. The Monier–Williams distillation technique, described in 1927, is still the standard by which other methods are evaluated. This method, and variants of it, involve the distillation, under reflux, of the food sample acidified with HCl, H_2SO_4, or H_3PO_4, in a gentle stream of an inert gas. SO_2 is trapped in H_2O_2 or iodine. Some variations of this technique include the codistillation of SO_2 with water with a downward sloping condenser, to speed up the procedure. The most serious interference arises from the presence of volatile compounds, which are oxidized by H_2O_2 to acid, or react with iodine. Such interfering components are found in onions, leeks, and cabbage. Small samples may be analyzed by replacing the conventional trapping agents by pararosaniline reagent or DTNB, which allow spectrophotometric determination of SO_2 at a much greater sensitivity than is possible by titration. Polarographic analysis of the distillate offers a good alternative. Desorption of SO_2 from cold acidified samples enables free preservative to be determined; prolonged boiling (e.g., 1.75 h) or pretreatment with alkali causes the decomposition of hydroxysulfonates and gives the total preservative present. In general, distillation methods do not require any sample preparation and are equally applicable to solid and liquid foods.

An automatic distillation unit ahead of an 'autoanalyzer' with spectrophotometric detection provides an effective automatic indirect method of analysis for liquid samples and aqueous extracts. An ingenious approach to separating dissolved SO_2 from an acidified food sample is by diffusion through polytetrafluoroethylene membrane. This may be used in a flow injection arrangement where the diffusion cell carries streams of acidified liquid food or extract solution on one side of the membrane and a suitable reagent for spectrophotometric analysis on the other.

Analysis of gaseous SO_2 in the headspace above acidified liquid foods or homogenates of solid foods, or transfer of SO_2 through the headspace from the sample to a reagent for spectrophotometric analysis in a microdiffusion cell, offers an alternative indirect method of analysis.

Nitrites and Nitrates

The most important uses of nitrite and nitrate as food preservatives are in meat products, and the majority of analytical procedures for food have been devised especially for this application. There are, however, numerous published procedures for the determination of the ions in water, and soil, and those occurring naturally in plant material. (*See* **Nitrates and Nitrites**.)

Since nitrites and nitrates are readily soluble in water, aqueous extracts from homogenates of the food at neutral or weakly alkaline pH are sufficient. A low extraction pH leads to the loss of nitrite. Apart from pH, important variables in extraction are time and the addition of heavy-metal salts (e.g., $BaCl_2$, $CdCl_2$, $HgCl_2$). A proportion of nitrite present in food may exist in bound form as nitrosothiols, nitrosylmyoglobin, etc.; methods of extraction that involve the heavy-metal ions and long extraction times (e.g., 1–4 h) cause the release of such bound additives. Techniques based on short extraction periods (e.g., 10 min) at neutral pH measure mainly the free nitrite. Nitrosyl hemoproteins may be decomposed to species convertible to nitrite if the extracting medium contains 80% acetone.

Extracts generally contain dissolved and suspended protein. This is removed, and solutions are clarified by means of zinc acetate + potassium ferrocyanide, alumina cream, potash alum or zinc sulfate + borax. In general, the reagents allow recoveries of nitrite of >98% with standard deviations no greater than 0.75%, from meat samples, with the exception of the procedure involving the use of alumina cream, for which the recovery is slightly lower at 95%.

Nitrate may be determined directly using a reaction in which it is used as a nitrating agent for variety of organic substrates and particularly the xylenols (dimethylphenols); the nitro-compounds so formed may be determined spectrophotometrically. This reaction can also be used to determine the concentration of nitrite if it is first oxidized to nitrate with permanganate. Alternatively, nitrate may be reduced to nitrite by spongy cadmium or enzymatically with NADPH (nitrate reductase). The rate of formation of NADP in the enzymatic procedure may be used to measure nitrate concentration directly. The most widely used reaction for the determination of nitrite is that with an aromatic primary amine (usually sulfanilic acid) in acid solution to form a diazonium salt, which is coupled to an aromatic amine or hydroxy compound to give an azo dye. This can be determined spectrophotometrically. Typical coupling agents have included 1-naphthylamine and 1-naphthol, which are carcinogenic, and 1-naphtholsulfonic acid and *N*-1-naphthylethylenediamine dihydrochloride are now frequently recommended. Both oxidizing and reducing agents interfere; ascorbic acid is a particular problem.

The sample blank can represent a significant error in analysis. There is no completely satisfactory way of preparing this; suggestions have included the passing of the deproteinized sample through an ion-exchange resin to remove nitrate and nitrite before determining the blank.

Other Preservatives

Biphenyl

Steam distillation of citrus peel homogenates and extraction of distillate with heptane provides a sample that may be analyzed for biphenyl by TLC on silica gel with heptane as the developing solvent. Quantitative analysis is by extracting biphenyl from the TLC medium with ethanol and measuring the absorbance of resulting solutions at 248 and 300 nm. The absorbance at 300 nm is used to correct for background absorbance. GLC methods of analysis are also available.

Thiabendazole

Thiabendazole may be extracted from food with acid (HCl), in which it is subsequently reduced with zinc in 30% glycerol+phenylenediamine. Subsequent oxidation with ferric ions gives rise to a blue complex, which is extracted into butanol for spectrophotometric measurement.

Lactic Acid

Standard methods for the determination of lactic acid involve its conversion to the Fe(III) complex for spectrophotometric measurement, or by GLC after complete methylation. In an enzymatic procedure, lactic acid is converted to pyruvate by NAD^+ (lactate dehydrogenase); the product is trapped by reacting with glutamate (glutamate–pyruvate transaminase) in order to displace the lactic acid–pyruvate equilibrium to the right. The rate of utilization of NAD^+ is used to measure lactic acid concentration. Sample preparation is generally as described for carboxylic acids above.

Hydrogen Peroxide

Spot tests for the presence of hydrogen peroxide in milk use either a solution of vanadium pentoxide in H_2SO_4 (pink or red color) or KI/starch (blue color).

Quantitative analysis is with peroxidase and a suitable substrate. The frequently recommended *o*-dianisidine is carcinogenic, and substrates such as guaiacol or 2,2′-azinobis (3-ethylbenzthiazolidine-6-sulfonic acid) are preferred. Assays on milk samples are carried out after precipitation of the protein by adjusting to pH 4.5 with HCl.

Nisin

The complexity of procedures for isolating and chemically quantifying nisin has led to the development of microbiological assays as normal analysis procedures. One international unit of nisin activity is the extent of inhibition of a test microorganism caused by approximately 25 ng of pure nisin. A suitable test microorganism is *Streptococcus agalactiae*. A particular assay procedure is an adaptation of the methylene blue reduction test; the delay in decolorization of the dye is proportional to nisin concentration over a 10-fold concentration range. A new method of analysis involves the enzyme-linked immunosorbent assay, which is much more straightforward than the microbiological assay and should become popular.

See also: **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **Immunoassays**: Radioimmunoassay and Enzyme Immunoassay; **Nitrates and Nitrites**; **Preservatives**: Classifications and Properties

Further Reading

Furia TE (ed.) (1972) *Handbook of Food Additives*. Boca Raton, FL: CRC Press.

Helrich K (ed.) (1990) *Official Methods of Analysis of the Association of Official Analytical Chemists*, 15th edn and First and Second Supplements (1990 and 1991). Arlington, VA: Association of Official Analytical Chemists.

Usher CD and Telling BM (1975) Analysis of nitrate and nitrite in foodstuffs: a critical review. *Journal of the Science of Food and Agriculture* 2: 1973–1805.

Wedzicha BL (1984) *Chemistry of Sulphur Dioxide in Foods*. London: Elsevier Applied Science.

Preserves *See* **Jams and Preserves**: Methods of Manufacture; Chemistry of Manufacture

PRETERM INFANTS – NUTRITIONAL REQUIREMENTS AND MANAGEMENT

T F Fok, The Chinese University of Hong Kong, Shatin, NT, Hong Kong

Physiological Considerations

Recent advances in neonatal care have resulted in a significant improvement in the survival of infants who are born premature, notably the very low birthweight infants (birth weight < 1500 g). In the developed world, most infants with a birth weight over 1000 g are able to survive, and the survival rate of those weighing between 700 and 999 g is also showing progressive improvement. The nutritional needs of these infants pose a special challenge to scientists and clinicians: they have a high demand for calories and nutrients to cope with their rapid growth rate, while at the same time, the physiological and biochemical immaturity of their digestive systems and metabolic pathways render them inefficient in metabolizing and assimilating the nutrients.

Carbohydrate, protein, and fat comprise the major nutrients and sources of calories. Carbohydrate is available in human and formula milk in form of lactose. Lactose is hydrolyzed in the small intestine by the enzyme lactase present in the brush border at the villus tip. Preterm infants have a relatively low concentration of intestinal lactase, which is only about 70% of that of term infants. However, most preterm infants seem to be able to tolerate lactose in human milk or milk formulas, and only a small amount of lactose is excreted in the stools. In

formulas specially designed for preterm infants, the carbohydrate content is enriched by the addition of glucose polymers. Preterm infants are able to digest glucose polymers, apparently through the action of salivary amylase and mucosal α-glycosidase.

Protein is required for the formation of new tissues and is therefore essential for growth and repair. Because of the immaturity of the amino acid metabolic pathways, preterm infants are not able to completely metabolize several amino acids. Hence, excessive protein intake might lead to incomplete amino acid catabolism resulting in elevation of plasma concentration of amino acids, hydrogen ion, and ammonia. Also, because of the immature amino acid metabolism, preterm infants are not able to synthesize some amino acids that are nonessential in older subjects. These amino acids, notably cysteine, taurine, and glycine, must be provided in the diet.

The fat component of milk consists mainly of triglycerides. The digestion of fat begins in the intestinal lumen when the triglycerides are solubilized and hydrolyzed into fatty acids by the combined action of bile sat and lipase. In the preterm infants, pancreatic lipase and intraluminal bile salts are relatively deficient, and triglyceride hydrolysis is carried out mainly by the action of endogenous lingual lipase and gastric lipase, and also the lipase present in human milk. Infants with gestation less than 26 weeks might have impaired fat digestion owing to a deficiency of gastric lipase. The second phase of fat digestion takes place in the intestinal mucosa, where fatty acids are reesterified to form chylomicron, which is secreted into the lymphatics. Some of the fatty acids remain unesterified and are directly absorbed into the portal venous system.

Medium chain triglycerides (MCT) are incorporated into the intestinal mucosal cells without the need for intraluminal lipase or bile salt. In the intestinal mucosa, MCT are hydrolyzed by mucosal lipase, and absorbed directly into the portal venous system. Thus the addition of MCT to preterm infant formulas should theoretically enhance fat absorption. A number of studies showed that MCT increase energy absorption and better weight gain in preterm infants, but others failed to demonstrate any benefit of MCT in energy or nitrogen balance.

Besides being a concentrated source of calories, fat is also an important component of phospholipids, which are essential for cellular functions and the formation of a large variety of bioactive metabolites including surfactant and the prostaglandins. Some fatty acids, including linoeic acid and linolenic acid, cannot be synthesized by the body, and are known as essential fatty acids. Deficiency in these fatty acids in animal fetus has been associated with long-term impairment in learning and visual function. Linoleic acid and linolenic acid are metabolized to arachidonic acid and docosahexaenoic acid, respectively, both of which are essential for growth and development, especially of the central nervous system. In the preterms, the formation of arachidonic acid and docosahexaenoic acid is relatively insufficient, especially when the supply of calories is suboptimal.

Arachidonic acid (20:4*n*-6) and docosahexaenoic acid (22:6*n*-3) belong to the *n*-3 and *n*-6 long-chain polyunsaturated fatty acids (LCPUFA) with double bonds at the six and third position, respectively. LCPUFA are present in human milk but absent from formula milk. In a randomized control trial in 447 full-term infants, supplementation with *n*-3 and *n*-6 LCPUFA did not seem to have any beneficial or adverse effect. In another randomized control trial on 56 healthy full-term infants, supplementation of infant formulas with 0.35% DHA or with 0.36% DHA and 0.72% arachidonic acid results in an increase of 7 points on mental development index at 18 months of age.

Nutritional Requirements of Preterm Infants

Energy

Energy intake from food is required to maintain the resting metabolism, normal body temperature, and growth. A small amount of energy is excreted in the feces and urine. Normally, a very low birth-weight infant can retain up to 85–95% of the energy intake starting at 2–3 weeks of age. The resting metabolic rate of a preterm infant is around 40 kcal kg^{-1} day^{-1} in the first week, increasing to 50 kcal kg^{-1} day^{-1} by 2–3 weeks of age, and is higher in small for gestational age infants. The energy expenditure due to activity is highly variable, and the reported values range from 3.6 to 19 kcal kg^{-1} day^{-1}. Energy spent on thermoregulation should be negligible in infants being nursed in a thermal neutral environment. Published data on the energy of growth vary greatly as the amount of energy varies with the composition of the tissues synthesized. A reasonable estimate is approximately 3–4.5 kcal per gram weight gain, or an average of 10 kcal kg^{-1} day^{-1}. The total energy requirement, including that for the basal metabolic need, growth, stored energy, energy excreted, and energy stored, activity, is approximately 90–120 kcal kg^{-1} day^{-1}. Infants who have suffered from intrauterine growth retardation require a higher energy intake. Infants with medical or surgical complications also have significantly greater demands for energy.

Carbohydrate

Carbohydrate is an important and readily usable source of energy. An adequate supply of glucose also serves to prevent catabolism of body tissues. During the early days of life when the infant is sick from the complications of prematurity, glucose is often given via the intravenous route to maintain a normoglycemic state. Any shortage of exogenous glucose supply at this stage will result in hypoglycemia, which may cause injury to the brain and potentially other organs since glucose is the major, if not the only, source of energy for the metabolism of the nervous system, red blood cells, the renal medulla, and the retina. Excessive glucose supply, however, results in hyperglycemia, which causes glycosuria and osmotic diuresis, and stimulation of pancreatic islet cells resulting in rebound hypoglycemia when the glucose supply is interrupted. In most preterm infants, the glucose infusion rate should be 4–6 mg kg^{-1} min^{-1}. The extremely preterm infants may develop hyperglycemia, even at this rate of glucose infusion, because of inadequate insulin secretion, or peripheral insulin resistance. These infants may require a reduction of glucose intake or insulin therapy. When the infant becomes stable and is taking enteral milk feeding, carbohydrate should contribute to about 40–50% of the total caloric intake (10–14 g per kilogram of body weight), which is the proportion of carbohydrate content in human milk.

Protein

The estimation of protein requirement and the requirements of most other nutrients were based on the need of the growing fetus of the same gestation. Daily protein requirements appeared to be 2–4 g kg^{-1}. In extremely low birth-weight infants, the amount could be increased in stepwise fashion over a span of 2 weeks. This should be accompanied by a concomitant increase in calories intake, since protein synthesis requires energy. For each gram of protein deposited, the body spends about 10 kcal of energy. A deficient energy intake will result in the breakdown of endogenous protein and a negative nitrogen balance.

Fat

Fat provides a concentrated source of energy and is also required to provide the essential fatty acids. As recommended by the European Society of Paediatric Gastroenterology and Nutrition Committee on Nutrition, the fat intake of enterally fed preterm infants should be 40–55% of the total calories intake, or 4.4–6.0 g per 100 kcal. The Committee also recommended a linoleic acid content of 4.5–10.8% of the total energy content of infant formulas, and a linoleic acid: α-linolenic acid ratio of 5:1 to 15:1.

Carnitine facilitates the transport of long-chain fatty acids through the mitochondrial membrane and is therefore essential for the oxidation of fatty acids for energy production in the heart and skeletal muscle. The European Society of Paediatric Gastroenterology and Nutrition Committee on Nutrition recommended that infant formulas contain at least 7.5 μmol of carnitine per 100 kcal. This amount should be available in human milk. Carnitine is now added to some of the commercial parenteral nutrition fluid. There is, however, no clinical evidence that supplementation of carnitine to the parenteral lipid infusion has any physiological benefit.

Vitamins and Minerals

Vitamin A Vitamin A is important for the growth and differentiation of epithelial cells, and its deficiency results in pathological changes of the epithelium, including that of the airways and lungs. The vitamin is transferred to the fetus from the mother mainly during late gestation, and therefore the vitamin content is significantly lower in the preterm infants than in the term infants. Several studies have demonstrated a protective effect of vitamin A supplementation against bronchopulmonary dysplasia in preterm infants. Excessive vitamin A intake results in intoxication manifesting with increased intracranial pressure, bone and joint damage, mucocutaneous lesions, liver damage, and hematological changes.

The dietary intake of vitamin A, as recommended by the World Health Organization, is 333 IU kg^{-1} day^{-1}. This amount may not be sufficient for preterm infants, and a daily intake of 1500 IU kg^{-1}, 700–1500 IU kg^{-1}, and 700 IU kg^{-1} has been recommended for infants weighing less than 1000 g, 1000–1750 g, and 1751–2500 g, respectively. Infants at risk of developing BPD, namely the extremely preterm infants with respiratory distress requiring supplemental oxygen and mechanical ventilation, might be protected from BPD by additional vitamin A intake. It has been demonstrated that, in these infants, biochemical vitamin A deficiency could be corrected with intramuscular administration of 5000 IU of vitamin A given three times per week for 4 weeks, which was associated with a slight reduction in the risk of chronic lung disease.

Vitamin E Vitamin E is a biologic antioxidant important for the integrity of the lipid layers of cell membranes. Its deficiency in preterm infants has been associated with hemolytic anemia, and impaired phagocytosis. Clinical trials have been carried out to evaluate its protective effects

against bronchopulmonary dysplasia, retinopathy of prematurity, and intraventricular hemorrhage. However, there is no convincing evidence that vitamin E protects preterm infants from any of these conditions.

Because of the limited tissue store and poor dietary absorption of vitamin E, and the rapid growth of the preterm infants, The American Academy of Pediatrics Committee on Nutrition recommended that preterm infants should receive 5–25 IU supplemental oral vitamin E per day.

Vitamin K Vitamin K is a cofactor in the metabolic conversion of intracellular precursors of vitamin K-dependent clotting factors 2, 7, 9, and 10. Vitamin K deficiency results in hemorrhagic disease of newborn. Purely breast fed infants are particularly prone to develop vitamin K deficiency, since the concentration of the vitamin is very low in human milk. The current recommended daily intake of vitamin K for newborn infants is the same as that for children and adults, namely 1 mg kg^{-1} day^{-1}.

Vitamin D Vitamin D is obtained from the diet or produced endogenously in the skin after exposure to ultraviolet light. The vitamin is hydroxylated to 25-hydroxyvitamin in the liver, which is further hydroxylated to its most active ingredient, 1,25-dihydroxy vitamin D in the kidney. The process is stimulated by parathyroid hormone and deficiency in calcium and phosphate. Vitamin D increases intestinal calcium and phosphorus absorption; 1,25-dihdroxy vitamin D also maintains normal circulating calcium concentration by its direct action on mobilization of bone calcium and phosphorus to the extracellular fluid. The role of vitamin D in the cause of osteopenia and ricket of the newborn is not clear. It appears that mineral (calcium and phosphorus) intake is the most important factor affecting bone mineralization, since infants with low mineral intake developed ricket despite supplementation of vitamin D of up to 2000 IU day^{-1}, whereas ricket could be prevented by high mineral intake and moderate vitamin D (400 IU day^{-1}) supplementation. Thus, a daily enteral intake of vitamin D of 400 IU should be sufficient for the preterm infants. For infants on parenteral nutrition, a daily intake of 40–160 IU kg^{-1} has been recommended, with a maximum intake not exceeding 400 IU day^{-1}.

Water-soluble vitamins The water-soluble vitamins, vitamin C, vitamin B_1 (thiamin), vitamin B_2 (riboflavin), vitamin B_3 (niacin), vitamin B_6 (pyridoxine), pantothenic acid, and biotin, are cofactors for several important enzymatic reactions, and deficiencies can result in abnormal substrate utilization. The exact requirements for these vitamins by preterm infants are not clear. The vitamins are water-soluble, and therefore, excessive intake will be excreted in the urine. The current recommended dietary allowance (RDA) for the term neonate is: vitamin C 30 mg day^{-1}, vitamin B_1 300 μg day^{-1}, vitamin B_2 400 μg day^{-1}, vitamin B_6 300 μg day^{-1}, niacin 5 mg day^{-1}. The RDA of biotin and pantothenic acid has not been established, but doses of 10 μg day^{-1} and 2 μg day^{-1}μg have been suggested for biotin and pantothenic acid, respectively. For preterm infants, it has been suggested that those weighing > 1500 g should receive the RDA, whereas smaller infants can be supplemented with half the oral doses. Very low birth-weight infants on total parenteral nutrition should be supplemented with 25–30 mg of vitamin C per kilogram per day, 35 μg of vitamin B_1 per kilogram per day, 150–200 μg of vitamin B_2 per kilogram per day, 150–200 μg of vitamin B_6 per kilogram per day, 4–6 mg of niacin per kilogram per day, 5–8 μg of biotin per kilogram per day, and 1–2 mg of pantothenic acid per kilogram per day.

Folic acid is a water-soluble vitamin that is important for DNA synthesis and cell division, especially in tissues with high rates of cell multiplication, such as the bone marrow and intestine. Preterm infants are predisposed to folic acid deficiency because of limited intrauterine hepatic stores and because of their rapid postnatal growth. The American Academy of Paediatrics Committee on Nutrition for Preterm Infants recommended that preterm infants < 2000 g should receive 50 μg of folic acid daily, with a minimum intake of no less than 4 μg per 100 kcal, which is the same as that for term infants. Preterm infants on total parenteral nutrition should be given folic acid supplement at a dose of 56 μg kg^{-1} day^{-1} (140 μg kg^{-1} day^{-1} for full term) usually in the form of a multivitamin preparation.

Vitamin B_{12} is a water-soluble vitamin. Its deficiency may result in megaloblastic anemia and neurological changes during late infancy. The minimal dietary intake for preterm infants, as recommended by the American Academy of Pediatrics Committee on Nutrition, is 0.15 μg per 100 kcal, which is the same as that for term infants. Preterm infants on total parenteral nutrition should be given vitamin B_{12} supplement at a dose of 0.3 μg kg^{-1} day^{-1} (1 μg kg^{-1} day^{-1} for full term).

Minerals

Calcium, phosphorus, and magnesium Calcium and phosphorus are the major constituents of bone. Magnesium is also an important constituent of

bone, but a large amount is also found in the muscle and intracellular fluid. About 80% of these minerals accrue in the fetus from 25 weeks of gestation to term. The peak accretion rate occurs between 36 and 38 weeks of gestation. The amount of minerals present in human milk and ordinary infant formulas, especially that of calcium and phosphorus, is not sufficient to cope with this rate of mineral accumulation. Osteopenia and rickets are common among the extremely preterm infants, and the condition can be prevented by high calcium and phosphorus intake. The recommended daily intakes of calcium, phosphorus, and magnesium for preterm infants are $210\,mg\,kg^{-1}$ ($5.25\,mmol\,kg^{-1}$), $140\,mg\,kg^{-1}$ ($4.52\,mmol\,kg^{-1}$) and $10\,mg\,kg^{-1}$ ($0.42\,mmol\,kg^{-1}$), respectively.

Sodium and potassium Preterm infants, particularly those with a birth weight less than 1500 g, have significant urinary sodium loss during the first 10–14 days of life because of high fractional excretion rates of sodium. During this time when the infants are sick and unstable, serum sodium and potassium values are affected by a number of factors including the amount of fluid intake, the rate of evaporative fluid loss, which is affected by environmental temperature and humidity, and therapeutic interventions such as mechanical ventilation. These infants are normally given only parenteral fluid. Restriction of parenteral sodium intake during the first 2 days of life is a common practice in neonatal intensive care units and may prevent the development of hypernatremia. Additional potassium intake is also not necessary during this period. Subsequently, the daily requirement is 2–3 mmol kg^{-1} for each of sodium and potassium. The amount of supplementation has to be adjusted according to the serum sodium and potassium levels, which should be monitored frequently for as long as the infant is on intravenous feed. Preterm infants on enteral feeding also require 2–3 mmol of sodium per kilogram per day, which is greater than that required by term infants. The sodium content in term human milk and commercial infant formulas designed for the feeding of term infants is too low to meet the requirement of very low birth-weight preterm infants and may lead to hyponatremia when given to these infants as the sole source of sodium. Special formulas for preterm infants provide 2.5–3.5 mmol of sodium per kilogram per day at full feeding. During the stable and growing period, sodium requirements are usually met with a daily intake of $2–3\,mmol\,kg^{-1}\,day^{-1}$. The potassium requirement of preterm infants seem to be the same as that of term infants (2–3 mmol $kg^{-1}\,day^{-1}$).

Trace minerals Eight trace minerals are nutritionally essential for the human: iron, zinc, copper, selenium, chromium, molybdenum, manganese, and iodine.

Preterm infants have low iron stores because iron accumulation occurs mainly at the third trimester at a rate of $1.6–2.0\,mg\,kg^{-1}\,day^{-1}$. This, together with the rapid growth rate of preterm infants, and the frequent blood sampling during the early postnatal days, results in a high incidence of iron deficiency anemia after age 4 months. The incidence has been decreased with iron supplementation. At the same time, these infants are also often given multiple transfusions, which may restore the iron load and impair intestinal absorption of iron.

The American Academy of Pediatrics Committee on Nutrition recommended that iron should be started no later than 2 months of age in very low birth-weight preterm infants and should be continued for at least the remainder of the first year. The recommended dose is $2\,mg\,kg^{-1}\,day^{-1}$ up to a maximum of $15\,mg\,day^{-1}$. Others have suggested that iron should be started at 2 weeks of age or when the body weight doubles the birth weight. Iron supplementation is particularly important when the infant is being treated with erythropoietin because of increased erythropoiesis.

Zinc is important for protein metabolism and cell growth. Fetal accretion of zinc takes place mainly during the third trimester of pregnancy, and preterm infants have a lower total store at birth than term infants. The daily enteral intake of zinc, as recommended by the American Academy of Paediatrics and the European Society of Paediatric Gastroenterology and Nutrition, is $600\,\mu g\,kg^{-1}\,day^{-1}$ and $720–1400\,\mu g\,kg^{-1}\,day^{-1}$ respectively.

Copper is an essential component of many oxidative enzymes and is important for the integrity of cell membranes. Its deficiency presents clinically with anemia, neutropenia, osteoporosis, decreased pigmentation of the skin and hair, dermatitis, failure to thrive, diarrhea, hypotonia, and psychomotor retardation. The recommended daily enteral intake of copper for preterm infants is $100–150\,\mu g\,kg^{-1}$. For infants on long-term parenteral nutrition, the recommended daily intake is $20\,\mu g\,kg^{-1}$. Copper supplementation is not necessary for those on only short-term TPN. As copper is primarily metabolized in the liver, and reduced hepatic excretion may cause liver cirrhosis, it should be withheld from infants with cholestasis.

Iodine is essential for the manufacturing of thyroid hormones. Transient hypothyroidism has been reported in preterm infants taking $10–30\,\mu g\,kg^{-1}\,day^{-1}$ of iodine. Excessive intake of iodine, however,

also suppresses the thyroid gland. The recommended daily intake of iodine for stable growing preterm infants is 30–60 $\mu g\,kg^{-1}\,day^{-1}$. For infants on total parenteral nutrition (TPN), a daily intake of 1 $\mu g\,kg^{-1}$ is recommended.

Selenium is an essential constituent of the enzyme glutathione peroxidase. Its deficiency results in cardiomyopathy affecting children and adults. Clinical selenium deficiency has, however, not been reported in preterm infants. The recommended daily enteral intake of selenium for stable and growing preterm infants is 1.3–3 $\mu g\,kg^{-1}$. Infants on parenteral nutrition should be supplemented with selenium at a dose of 2 $\mu g\,kg^{-1}\,day^{-1}$.

Chromium is a cofactor involved in glucose homeostasis. Chromium deficiency has not been described in newborns. The daily recommended intake for stable and growing preterm infants is 0.1–0.5 $\mu g\,kg^{-1}$. The recommended daily parenteral intake for infants on TPN is 0.2 $\mu g\,kg^{-1}$.

Molybdenum is required for the function of certain enzymes, including xanthine oxidase. Its deficiency has not been described in the newborns. The recommended daily enteral intake for stable preterm infants is 0.3 $\mu g\ kg^{-1}$. The recommended intake for infants on total parenteral nutrition is 0.25 $\mu g\ kg^{-1}\,day^{-1}$.

Manganese is required for the function of pyruvate carboxylase, superoxide dismutase, and glycosyl transferase. Manganese deficiency has not been reported in the newborns. The recommended intake for stable growing preterms is 0.75–7.5 $\mu g\,kg^{-1}\,day^{-1}$. The recommended parenteral intake for infants on total parenteral nutrition is 1 $\mu g\,kg^{-1}\,day^{-1}$. Since manganese is metabolized in the liver, and its accumulation has been implicated as a cause of cholestasis and damage to the central nervous system, the element should be withheld from infants with cholestasis or impaired liver function.

Dietary Sources of Nutrients

Carbohydrate

The lactose content of mature human milk is 7.4–9.9 $g\,dl^{-1}$. The lactose content of preterm milk is 6.99 $g\,dl^{-1}$ after 26–31 weeks of gestation and 6.24 $g\,dl^{-1}$ after 32–36 weeks of gestation. In preterm formulas, the carbohydrate content is about 8.5 $g\,dl^{-1}$ and consists of lactose and glucose polymers in equal amounts. Compared with lactose, the addition of glucose polymers to formulas adds fewer osmotic particles per unit weight. Its use has allowed a substantial increase in the carbohydrate content of formulas without any significant increase in osmolality.

In infants receiving parenteral nutrition, carbohydrate is usually given in the form of intravenous glucose. As described above, the glucose infusion rate should be sufficient to maintain a normoglycemic state, and most preterm infants will require a glucose infusion rate of 4–6 mg $kg^{-1}\ min^{-1}$. The extremely preterm infants may not be able to tolerate glucose load and may require a reduction in the infusion rate or insulin therapy.

Protein

Mature human milk cannot provide sufficient protein for preterm infants since its protein content is only 1.2 $g\,dl^{-1}$. To obtain a daily amount of protein of 3 $g\ kg^{-1}$, the infant has to consume 250 ml of milk per kilogram per day. The milk produced by the preterm baby's own mother contains significantly more protein, especially during the first few weeks of life, and is more suitable for the infant. However, as the protein content of the preterm milk declines rapidly after delivery, and becomes quite similar to that of term milk after 2 weeks, it is now common practice to enrich the protein content of expressed breast milk with human milk fortifier. Such supplementation has resulted in a greater weight gain and nitrogen retention.

Formulas specially designed for preterm infants have a higher protein content than breast milk – ranging from 1.8 to 2.4 $g\,dl^{-1}$ and providing about 2.2–3.2 g of protein per 100 kcal.

Infants on parenteral nutrition are given protein in the form of amino acid mixture. Currently available amino acid mixtures provide up to 3 $g\,kg^{-1}\,day^{-1}$ of protein intake, and are able to meet the requirements for growth and nitrogen retention.

Fat

The fat content of mature expressed human milk is around 3.5–4.5 $g\,dl^{-1}$. The fat content of premature milk is lower, and its content and composition vary, depending on the mother's dietary fat intake. The fat content in the milk obtained by manual expression (expressed breast milk) is higher than drip breast milk. During tube feeding, substantial reduction in fat intake may result from the deposition of fat on the wall of the gastric tube.

Preterm formulas often contain 20–50% MCT. MCT are absorbed directly into the portal venous system without the need for intraluminal bile salts, which are relatively deficient in the preterms. The benefit of adding MCT oils to preterm formulas is controversial.

Several commercially available premature formulas are now supplemented with LCPUFAs.

Preterm infants on parenteral nutrition are given intravenous lipid, which contains about 45–55% linoleic acid and 6–9% linolenic acid. Normally, infusion of 0.5–1 g of lipid per kilogram per day of lipid will prevent essential fatty acid deficiency. More lipid can be given to provide more energy but should not exceed the rate of lipid clearance from the plasma. Most infants over 28 weeks of gestation can clear from the plasma up to 2 g of lipid per kilogram per day, and most infants more than 32 weeks of gestation can clear up to $3\,g\,kg^{-1}\,day^{-1}$. Infants less than 28 weeks of gestation are less able to clear lipid and have a high chance of hyperlipidemia when receiving intravenous lipid.

Intravenous lipid is hydrolyzed by endogenous lipoprotein lipase. The use of MCT in the lipid preparation results in more rapid clearance of lipid from the blood. There is, however, no clinical evidence that MCT preparation offers any physiological or metabolic advantage to the preterm infants.

Vitamins and minerals

Vitamin A The concentration of vitamin A in preterm milk is approximately $300\,IU\,dl^{-1}$, which is lower than that in term milk. The concentration is, however, variable. Preterm formulas are fortified with a vitamin A concentration of 250–1000 IU dl^{-1}. Multivitamin preparations usually contain about 1500 IU of vitamin A per milliliter.

Vitamin E The vitamin E content of preterm human milk is sufficient to satisfy the requirement of the preterm infant. The American Academy of Pediatrics Committee on Nutrition recommended that formulas designed for preterm infants should meet the minimum requirement of 0.7 IU per 100 kcal and 1.0 IU of linoleic acid per gram. The currently available preterm formulas are fortified with vitamin E well in excess of this amount. The requirement of vitamin E by preterm infants receiving parenteral nutrition is not clear. It has been recommended that infants weighting < 1000 g at birth should receive supplemental vitamin E of $3.5\,IU\,day^{-1}$, and those weighting 1000–1500 g should receive $4.7\,IU\,day^{-1}$.

Vitamin K The American Academy of Pediatrics recommended that in order to prevent hemorrhagic disease of newborns, vitamin K should be given to term infants at birth in the form of vitamin K_1 administered intramuscularly at a dose of 1 mg. Oral vitamin K can be used, but at a dose of 2 mg at birth, and should be repeated at 1–2 weeks and 4 weeks after birth. For preterm infants, an intramuscular dose of 0.3 mg is sufficient. It has been reported in one study that intramuscular vitamin K is associated with a higher incidence of childhood leukemia and suggested that it should be replaced by oral vitamin K. This recommendation was not supported, however, by other studies, and intramuscular injection remains the most certain way of vitamin K administration. Formulas designed for preterms are fortified with vitamin K well in excess of the daily requirement of the infants.

Preterm infants on TPN should be supplemented with vitamin K in the form of a multivitamin preparation. The amount provided usually exceeds the daily requirement of $10\,\mu g\,kg^{-1}$.

Vitamin D Preterm infants are often deficient in their mineral (calcium and phosphate) intake but the quantity of vitamin D in both human milk and formulas is sufficient. Infants on parenteral nutrition should be supplemented with vitamin D in form of a multivitamin preparation. Care is required to ensure that the dose is not excessive.

Water-soluble vitamins Human milk produced by healthy mothers on a balanced diet is a good source of the water-soluble vitamins vitamin C, vitamin B_1 (thiamin), vitamin B_2 (riboflavin), vitamin B_3 (niacin), vitamin B_6 (pyridoxine), pantothenic acid, biotin, folic acid, and vitamin B_{12}. Infant formulas are now all fortified with a sufficient quantity of these vitamins to meet the infant's requirements.

Minerals

Sodium and potassium Mature human milk contains 0.7–0.9 mmol per 100 ml, which is inadequate to meet the requirement of the preterm infants. Preterm milk has a higher sodium content during the first 3 weeks postpartum (1.3–1.7 mmol per 100 ml), which declines to 0.8 mmol per 100 ml by the 4th postpartum week. Formulas designed for preterm infants have a higher sodium content of 1.4 mmol per 100 ml and are able to provide 2.5–3.5 mmol of sodium per kilogram per day to infants receiving a full feed of $200\,ml\,kg^{-1}\,day^{-1}$. The potassium content in preterm formulas, and also in term and preterm human milk, contain is sufficient to meet the requirement of 2–3 $mmol\,kg^{-1}\,day^{-1}$ in preterm infants.

Calcium, phosphorus, and magnesium Preterm human milk contains only 40 mg of calcium per 100 kcal and 20 mg of phosphorus per 100 kcal, which are insufficient to satisfy the need of the preterm infants. The mineral content of the milk can be enriched by the addition of a human milk fortifier, the use of which has been associated with improved bone mineral content of the infants. Ordinary infant

formulas also cannot provide the necessary amount of calcium and phosphorus. Formulas designed for preterm infants contain 94–183 mg of calcium per 100 kcal and 50–70 mg of phosphorus per 100 kcal. Consumption of these formulas has been associated with improved bone mineral content of the preterm infants. The magnesium content of both human and formula milk is adequate for preterm infants.

Preterm infants, especially very low birth-weight infants, often have a transient period of hypocalcemia during the first few days of life, and require intravenous calcium supplement. In these infants, hypocalcemia can be prevented or treated with calcium infusion in the form of 10% calcium gluconate at a rate of 1.25–1.88 mmol (50–75 mg) of elemental calcium per kilogram per day in the first 2–3 days of life.

When the minerals are being delivered to infants on TPN, it is important that the concentrations of calcium and phosphorus do not exceed the solubility product, or else the minerals will precipitate. A calcium:phosphorus ratio of 1:1 to 1.3:1 by molar ratio, or 1.3:1 to 1.7:1 by weight, results in the highest calcium and phosphorus retention. It has been recommended that the TPN solution should contain 12.5–15 mmol (500–600 mg) of calcium per liter, 12.5–15 mmol (390–470 mg) of phosphorus per liter, and 1.5–2 mmol (36–48 mg) of magnesium per liter. This would provide the infant with 1.5–2.5 mmol (60–90 mg) of calcium per kilogram per day, 1.5–2.25 mmol (47–70 mg) of phosphorus per kilogram per day and 0.18–0.3 mmol (4.3–7.2 mg) of magnesium per kilogram per day.

Iron Human milk contains very little iron. Infants fed on breast milk should be supplemented with ferrous sulfate, which provides 2–4 mg of elemental iron per kilogram per day. Preterm formulas are fortified with 1.33 mg of iron per 100 ml or 1.67 mg of iron per 100 kcal, which should be sufficient for a fully fed infant.

For babies on prolonged TPN, 0.1–0.2 mg of parenteral iron per kilogram per day should be administered together with the TPN solution.

Trace minerals The zinc content in human milk declines rapidly during the first few postpartum months, and might not be sufficient to meet the need of the growing preterm infant. Preterm formulas contain sufficient zinc to provide the recommended daily intake in fully enterally fed infants. It has been suggested that preterm infants on human milk should be supplemented with $500\,\mu g\,kg^{-1}\,day^{-1}$ until 6 months of age. For infants on parenteral nutrition, it has been suggested that the TPN solution should be supplemented with 150 μg of zinc per kilogram per day during the transitional period after birth.

The copper content in human milk and preterm formulas is sufficient to meet the need of preterm infants. The iodine content of human milk varies with the dietary intake and geographic location of the mother. In most developed countries, human milk contains sufficient iodine to meet the need of the preterm infants. Preterm formulas are also fortified with an adequate amount of iodine. The amounts of other trace minerals (selenium, chromium, molybdenum, and manganese) in both human and formula milk are also adequate for the preterm infants.

See also: **Energy**: Intake and Energy Requirements; **Fats**: Requirements; **Folic Acid**: Physiology; **Magnesium**; **Minerals – Dietary Importance**; **Phosphorus**: Physiology; **Potassium**: Physiology; **Protein**: Food Sources; Requirements; **Selenium**: Physiology; **Sodium**: Physiology; **Thiamin**: Physiology; **Tocopherols**: Physiology; **Vitamin K**: Physiology; **Zinc**: Physiology

Further Reading

Birch EE, Garfield S, Hoffman DR, Uauy R and Birch DG (2000) A randomized controlled trial of early dietary supply of long-chain polyunsaturated fatty acids and mental development in term infants. *Developmental Medicine and Child Neurology* 42(3): 174–181.

Catzeflis C, Schutz Y, Micheli JL *et al.* (1985) Whole body protein synthesis and energy expenditure in very low birth weight infants. *Pediatric Research* 19(7): 679–687.

Evans JR, Allen AC, Stinson DA *et al.* (1989) Effect of high-dose vitamin D supplementation on radiographically detectable bone disease of very low birth weight infants. *Journal of Pediatrics* 115(5 Part 1): 779–786.

Freymond D, Schutz Y, Decombaz J, Micheli JL and Jequier E (1986) Energy balance, physical activity, and thermogenic effect of feeding in premature infants. *Pediatric Research* 20(7): 638–645.

Greene HL, Hambidge KM, Schanler R and Tsang RC (1988) Guidelines for the use of vitamins, trace elements, calcium, magnesium, and phosphorus in infants and children receiving total parenteral nutrition: report of the Subcommittee on Pediatric Parenteral Nutrient Requirements from the Committee on Clinical Practice Issues of the American Society for Clinical Nutrition [published errata appear in *American Journal of Clinical Nutrition* 1989; 49(6): 1332 and 1989; 50(3): 560]. *American Journal of Clinical Nutrition* 48(5): 1324–1342.

Gross SJ (1983) Growth and biochemical response of preterm infants fed human milk or modified infant formula. *New England Journal of Medicine* 308(5): 237–241.

Hanning RM and Zlotkin SH (1989) Amino acid and protein needs of the neonate: effects of excess and deficiency. *Seminars in Perinatology* 13(2): 131–141.

Kashyap S, Schulze KF, Forsyth M *et al.* (1990) Growth, nutrient retention, and metabolic response of low-birth-weight infants fed supplemented and unsupplemented

preterm human milk. *American Journal of Clinical Nutrition* 52(2): 254–262.

Lucas A, Stafford M, Morley R *et al*. (1999) Efficacy and safety of long-chain polyunsaturated fatty acid supplementation of infant-formula milk: a randomised trial [see comments]. *Lancet* 354(9194): 1948–1954.

Lundstrom U, Siimes MA and Dallman PR (1977) At what age does iron supplementation become necessary in low-birth-weight infants? *Journal of Pediatrics* 91(6): 878–883.

Putet G (1993) Energy. In: Tsang RC, Lucas A, Uauy R and Zlotkin S (eds) *Nutritional Needs of the Preterm Infant. Scientific Basis and Practical Guidelines*, pp. 15–28.Cincinnati, OH: Digital Educational Publishing.

Reifen RM and Zlotkin S (1993) Microminerals. In: Tsang RC, Lucas A, Uauy R and Zlotkin S (eds) *Nutritional Needs of the Preterm Infant. Scientific Basis and Practical Guidelines*, pp. 195–207. Cincinnati, OH: Digital Educational Publishing.

Schulze KF, Stefanski M, Masterson J *et al*. (1987) Energy expenditure, energy balance, and composition of weight gain in low birth weight infants fed diets of different protein and energy content. *Journal of Pediatrics* 110(5): 753–759.

Tyson JE, Wright LL, Oh W *et al*. (1999) Vitamin A supplementation for extremely-low-birth-weight infants. National Institute of Child Health and Human Development Neonatal Research Network [see comments]. *New England Journal of Medicine* 340(25): 1962–1968.

Vileisis RA (1987) Effect of phosphorus intake in total parenteral nutrition infusates in premature neonates. *Journal of Pediatrics* 110(4): 586–590.

PROBIOTICS

S E Gilliland, Oklahoma State University, Stillwater, OK, USA

Background

Probiotics may be defined as selected viable microorganisms that, following consumption in a food or feed, have potential for improving health or nutrition of man or animal. Bacteria in this group may be used to ferment food or are added to food as dietary supplements. Foods for human consumption containing these organisms are sometimes referred to as functional foods. A number of different bacterial species have been suggested as probiotics (**Table 1**). The major species that have been considered over the years, however, are *Lactobacillus acidophilus*, *Lactobacillus casei*, and *Bifidobacterium* species, and *Bifidobacterium longum*.

Table 1 Bacteria used as probiotics

Major bacteria
Lactobacillus acidophilus
Lactobacillus casei
Bifidobacterium longum
Bifidobacterium bifidum
Others
Lactobacillus ruterii
Lactobacillus johnsonii
Bifidobacterium lactis
Lactobacillus plantarum

Dairy products provide an excellent carrier for these probiotic bacteria especially *Lactobacillus acidophilus*, *Lactobacillus casei*, and *Bifidobacterium* species. Most of these can readily utilize lactose as an energy source for growth; thus, an important requirement for their growth in the intestinal tract is provided by the milk. Additionally, many of them have been used for centuries in the manufacture of cultured dairy products and thus are generally regarded as safe. This means that there should be no objection to their being used as dietary adjuncts in milk products. In the early 1900s, Eli Metchnikoff advocated that man should consume, on a daily basis, milk fermented by *Lactobacillus* species to displace microorganisms in the intestinal tract that produce toxic substances that could shorten man's life. Thus, he felt that consumption of such fermented milk would prolong life. This theory was the beginning of what we refer to today as foods containing probiotic bacteria.

Several dairy products containing probiotic bacteria are marketed today, and the most widely encountered product is yogurt. It is not uncommon to read on the label of a yogurt product that it was made with a culture including *Lactobacillus acidophilus* and/or *Bifidobacterium* species. These are not the bacteria that have traditionally been used to manufacture yogurt. The traditional yogurt starter cultures are *Streptococcus salivarius* ssp. *thermophilus* and *Lactobacillus delbrueckii* ssp. *bulgaricus*. Neither of these two organisms is traditionally listed as a probiotic, because they are not expected to survive and grow in the intestinal tract, whereas

probiotics do. In the USA, nonfermented milk products are available that contain added cells of *Lactobacillus acidophilus* and/or bifidobacteria. Following supplementation of the pasteurized milk with these two organisms, it is stored under refrigeration so that the probiotic bacteria do not grow in the milk. In this case, milk serves as a carrier for the organisms, and the product, of course, does not have the sour or tart taste normally associated with fermented dairy products. This provides a product containing the probiotics for those people who do not desire the taste and flavor associated with fermented products.

Potential Benefits

When Eli Metchnikoff proposed his theory to prolong life, the primary benefit he proposed from the consumption of the fermented milk product was to control the intestinal microflora. Since the interest in probiotics was renewed some 20–30 years ago, several other potential benefits have come to light (Table 2). While these potential health and nutritional benefits appear to be very important to mankind, most of them have not been thoroughly proven. In order to firmly establish whether or not the benefits can actually be produced, relatively large clinical trials will be required, especially in the case of humans. A large number of factors must be considered when undertaking such a monumental task. A very important aspect is the selection of the probiotic bacteria to be utilized. This will be discussed in a later section of this chapter. Most of the possible benefits listed in Table 2 have to do with human health and nutrition. However, with the pending ban on the use of subtherapeutic levels of antibiotics in livestock feed, there is tremendous interest in the role of probiotics as supplements for animal feed. The primary interest here focuses on controlling intestinal pathogens as well as improving nutrient utilization. The subtherapeutic antibiotics, which have been used for years as livestock feed supplements, have been shown in many cases to improve the growth and performance of livestock. While this may be due in part to control of undesirable microorganisms in the digestive tract, other mechanisms are also likely. Probiotic bacteria offer very good opportunities for use as livestock feed supplements to achieve these goals.

Table 2 Potential health and/or nutritional benefits from probiotics in the diet

- Control of intestinal pathogens
- Modulation of the immune system
- Control lactose maldigestion
- Hypocholesterolemic action
- Control of colon cancer
- Improved utilization of nutrients

Control of Intestinal Infections

In the literature, there are many studies on the use of *Lactobacillus* species to control intestinal infections. Unfortunately, many of these earlier studies were not well designed. In most cases, the probiotics, primarily *Lactobacillus* species, were used as a therapeutic rather than a preventative agent. In many cases, proper controls were not included in the studies, so valid observations as to the effectiveness of the probiotic could not be made. Additionally, in many of these studies, there was very little information concerning the particular culture of probiotic involved, especially with regard to origin or characteristics giving it the potential for producing the desired effect.

While it is possible that they may function in a therapeutic manner, it is more reasonable to consider probiotics as a prophylactic treatment for intestinal infections. This has been shown in a well-designed study involving germ-free chickens. In this study, germ-free chickens were fed a culture of *Lactobacillus acidophilus* and 2 days later were challenged with either *Salmonella typhimurium* or *Staphylococcus aureus*. In the same study, another series of germ-free chickens were fed the pathogenic organisms initially and then fed *Lactobacillus acidophilus* (as a therapeutic treatment). In the therapeutic experiments, the *Lactobacillus acidophilus* had minimal effect, but those animals fed the *Lactobacillus acidophilus* initially (prior to challenge with the pathogen) survived much better than did those animals that were challenged with the pathogen first. These results suggest that it is important to have the *Lactobacillus acidophilus* initially (i.e., as a prophylactic treatment) to ward off intestinal pathogens. The scientists conducting this study also indicated that it was desirable to provide the probiotic organism on a continuing basis in a diet for best control of the pathogens. Other studies involving animal models have confirmed these types of results. It is difficult to perform such studies in humans, since it is very difficult, if not impossible, in a university setting to obtain approval for doing challenge studies with intestinal pathogens using human test subjects. However, there have been some well-designed studies in which feeding products containing selected probiotic bacteria have been shown to be effective in helping control naturally occurring intestinal disorders brought on by intestinal pathogens, especially in young children. Based on what is known about the variation among strains and species of this group of bacteria, it is

Table 3 Suggested mechanisms through which probiotics might control intestinal infections

- Production of antimicrobial substances
- Competition for nutrients (not likely)
- Competitive exclusion
- Stimulation or modulation of the body's immune system

very important to properly select a strain or culture for use as a probiotic to control intestinal infections.

Several mechanisms have been suggested to be responsible for the control of intestinal pathogens by probiotic bacteria (Table 3). Probiotic bacteria can produce several types of antimicrobial agents, including acid, bacteriocins, and hydrogen peroxide. However, because of the proteolytic enzymes in the intestine, the role of bacteriocins (most of which are easily inactivated by proteolysis) may be very limited. Because of the low levels of oxygen, peroxide formation should be minimal. Acid (especially acetic acid) is likely the most involved of these antimicrobial agents. Some have suggested competition for nutrients as a mechanism, although this is unlikely. Competitive exclusion, where probiotic bacteria occupy the binding sights on the intestinal wall, thus preventing attachment of pathogenic organisms, may be a main factor in some cases. Stimulation or modulation of the body's immune system, which is discussed in the next section, appears to be a very likely mechanism.

Modulation of the Immune System

Stimulation or modulation of the body's immune system by probiotic bacteria can have health benefits. Probiotic bacteria can cause the body to secrete antimicrobial substances into the intestines. These antimicrobial substances then in turn inhibit the growth of undesirable microorganisms. This is a very interesting and challenging area of research and is currently receiving considerable attention throughout the world. While most of the research on modulation of the immune system has involved animal studies, several studies have focused on the influence of feeding cells of *Lactobacillus acidophilus* on the immune system of humans. In these studies, an increase in the secretion of substances, which are considered inhibitory for certain intestinal pathogens, into the intestine was observed. The ability to modulate the body's immune system through the consumption of probiotic bacteria provides a great opportunity to control intestinal pathogens as well as potentially to have other benefits for the consumer.

Lactose Maldigestion

A number of terms, including 'lactose malabsorption' and 'lactose intolerance,' have been used to describe the inability of a person to digest lactose adequately. 'Lactose maldigestion' seems to be a more accurate term to describe this condition in humans, since it refers to the inability of a person to digest lactose adequately. This is primarily due to the absence of sufficient amounts of the enzyme β-galactosidase in the small intestine.

Microorganisms in the intestinal tract can influence the ability to digest lactose. For instance, dietary lactose is highly toxic to germ-free chicks, whereas chicks with a normal intestinal flora, are able to digest the lactose. This difference has been attributed to the lactose-hydrolyzing enzymes provided by the microorganisms in the intestinal tract of conventional chicks.

Some cultured or culture-containing milk products have been shown to be beneficial. For instance, the consumption of milk containing cells of *Lactobacillus acidophilus* has been shown to improve the utilization of lactose by lactose maldigestors. A similar phenomenon can be observed by the consumption of cultured yogurt containing viable starter bacteria. The primary requirement for these probiotic and yogurt bacteria to provide this benefit is that they contain adequate levels of the enzyme β-galactosidase (Figure 1). The traditional bacteria used to manufacture yogurt, *Lactobacillus delbrueckii* ssp. *bulgaricus* and *Streptococcus salivarius* ssp. *thermophilus*, are not expected to survive and grow in the intestinal tract, because they lack bile tolerance. However, when they interact with bile in the intestinal tract, their permeability is altered, permitting lactose to enter the cells to be hydrolyzed. This results in the organisms being able to hydrolyze lactose at a more rapid rate than they would in milk in the absence of bile. Thus, without growing in the intestinal tract, these two species of bacteria can provide a benefit for those people unable to digest lactose adequately. Research has shown that, while the permeability of these cells was altered in the presence of bile, it was not sufficient to completely rupture the cells. The β-galactosidase remains in the cell having the increased permeability.

Bile has a similar effect on *Lactobacillus acidophilus* in that the permeability of the cells increases

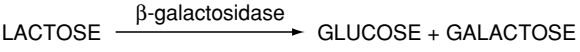

Figure 1 Action required by probiotics to provide benefit for lactose maldigestion.

in the presence of bile enabling lactose to enter more freely and be hydrolyzed. Thus, the mechanism whereby cells of *Lactobacillus acidophilus* can alleviate the problems of lactose maldigestion is similar to that of yogurt. However, *Lactobacillus acidophilus* has the added advantage that it can grow in the presence of the bile, and so new cells can be produced during its growth in the intestine, providing even more of the enzyme necessary for hydrolyzing lactose. When preparing a product to which cells *Lactobacillus acidophilus* are added, it is important that the cells have been grown in a medium containing lactose as a sugar source to insure that cells contain adequate levels of the enzyme β-galactosidase. If not, the amount of enzyme probably will be too low to produce the desired benefit with regard to lactose maldigestion as rapidly as desired.

Hypocholesterolemic Actions

The risk of coronary heart disease in hypercholesterolemic persons can be significantly reduced by lowering serum cholesterol levels. Germ-free animals on an elevated cholesterol diet accumulate approximately twice as much cholesterol in the blood as conventional animals on a similar diet. This suggests the possibility that microorganisms in the intestinal tract can influence serum cholesterol levels. Consumption of organisms such as *Lactobacillus acidophilus* can aid in the control of serum cholesterol levels in humans. Two studies published in the 1970s showed this to occur. In one of the studies, infants fed a formula supplemented with *Lactobacillus acidophilus* exhibited reduced serum cholesterol levels associated with increased numbers of lactobacilli in the feces, compared with infants receiving the control formula. This suggests that the lactobacilli may be associated with the reduction of serum cholesterol levels. In another study, a group of men fed milk fermented by the natural flora in raw milk exhibited unexpected reduced serum cholesterol levels. The fermentation apparently included what was described as a culture of *Lactobacillus*, although the organism was not identified.

Because of these studies, considerable research has focused on the potential for *Lactobacillus acidophilus* or related bacteria to exert beneficial influences on serum cholesterol levels. *Lactobacillus acidophilus* growing in a laboratory medium supplemented with cholesterol and bile salts under anaerobic conditions will assimilate cholesterol. This ability to assimilate cholesterol varies tremendously among strains of this species of bacteria. Results from a feeding trial using pigs on a high-cholesterol diet as an animal model revealed that supplementation of the diet with a strain of *Lactobacillus acidophilus* that assimilated cholesterol during growth in a laboratory medium had a significant beneficial effect on serum cholesterol, whereas a strain that did not assimilate cholesterol during growth in a laboratory medium did not. This suggests the importance of screening cultures for use as a probiotic to produce a beneficial effect on serum cholesterol levels. In another feeding trial involving pigs with diet-induced hypercholesterolemia, animals receiving cells of *Lactobacillus acidophilus* exhibited a significant positive correlation between the reduction of serum cholesterol levels and the concentration of cholic acid in the serum. The decline in serum cholesterol was initiated more rapidly and reached a significantly lower level in the animals fed the lactobacilli than in those not fed the lactobacilli. Such data suggest the possibility that the lactobacilli interfered with the enterohepatic circulation of bile acids, which can be an important factor in reducing serum cholesterol levels. This also suggests another activity of the lactobacilli, which may be important in controlling serum cholesterol levels. *Lactobacillus acidophilus* has the ability to deconjugate bile acids through the action of bile salt hydrolase. Free bile acids being less well absorbed from the small intestine than conjugated acids are more likely to be excreted from the body. Replacement of these excreted bile acids involves the synthesis of new bile acids from cholesterol in the body. This has a tendency to lower the cholesterol pool in the body. This represents the major pathway for removing cholesterol from the human body. An additional advantage for the deconjugation is that free bile acids do not support the absorption of cholesterol for the intestinal tract as well as conjugated acids.

The mechanism whereby consumption of fermented milk products containing cells of *Lactobacillus acidophilus* or other probiotics may reduce levels of serum cholesterol appears to include their ability to assimilate cholesterol and/or deconjugate bile acids (Table 4). Some of the cholesterol assimilated by *Lactobacillus acidophilus* is incorporated in the cellular membranes of the organisms. *Bifidobacterium longum* also can incorporate some cholesterol into its membrane, but *Lactobacillus casei* does not. All three species have the ability to deconjugate bile salts.

Table 4 Activities of probiotic cultures important for imparting a hypocholesterolemic effect

- Assimilation of cholesterol
- Deconjugation of bile salts

Control of Colon Cancer

Some lactic acid bacteria can produce anticarcinogenic or antimutagenic compounds, which can result in the inhibition of some types of tumor cells in animals. For example, rats that had been implanted with ascites tumor cells and fed milk fermented with *Lactobacillus acidophilus* exhibited a lower proliferation of tumor cells than those in the control group. The conclusion was that the *Lactobacillus* species produced substances during growth in the intestine that was antagonistic toward the proliferation of these tumor cells.

Other studies have shown that certain fermented milks have a beneficial influence on the proliferation of colon tumor cells in rats. Feeding rats cells of *Lactobacillus* reduced the activity of three enzymes capable of converting procarcinogens into carcinogens in the intestinal tract. Similar results have been shown in human studies. These enzymes were β-glucuronidase, azoreductase, and nitroreductase. Reduction of the level of activity of these enzymes in the intestinal tract can reduce the chances of carcinogens being formed. The enzymes apparently are of microbial origin. Thus, the potential anticarcinogenic activity of feeding *Lactobacillus acidophilus* may, in this case, be related to the control of undesirable microorganisms in the intestinal tract. In other studies, the consumption of milk containing cells of *Lactobacillus casei* activated macrophages in mice. This observation was based on increases in enzyme activity associated with macrophages. This suggests another potential means, whereby the consumption of a selected probiotic culture may influence proliferation of tumor cells in the body.

Other Benefits

One of the main interests in the livestock industry in using probiotic bacteria as feed supplements is to improve nutrient utilization resulting in improved growth and/or feed efficiency of the livestock. While there is great interest in this area, there is very little evidence reported in the scientific literature to indicate how such benefits might occur. It is possible that such benefits can result from the action of certain enzymes associated with the probiotic bacteria. This could be similar, for instance, to the beneficial effect of yogurt bacteria and *Lactobacillus acidophilus* on lactose maldigestion in humans that results from β-galactosidase associated with the bacterial cells. Selection of a culture of *Lactobacillus acidophilus* with a high level of amylase activity for inclusion as a probiotic in the diet of young pigs can improve starch utilization in the young animal. In a feeding trial involving the use of such a culture, improved growth and feed efficiency of weaning aged pigs were observed as a result of including the selected culture in the diet. The benefit was attributed to the amylase activity of the culture. It is possible that other enzymes that could improve the digestion of other nutrients may be present in some bacteria used as probiotics that could result in improved growth and feed efficiency of the livestock. Of course, the improved growth may relate to the control of undesirable microorganisms in the animal's intestinal tract. There is also interest in the livestock and pet-food industries in using probiotics to help control intestinal infections in animals.

Selection of Probiotic Culture

While there are a number of potential benefits from the inclusion of a probiotic culture in the diet, it is essential that a culture be properly selected for such use. One culture, that is one strain of one species, should not be expected to provide all of the potential benefits that might be derived from probiotics. To help insure that the culture is successful and has a positive effect, certain requirements are to be met (Table 5). Since, in most cases, the sight of action of the probiotic organisms is the intestinal tract, it is generally accepted that the organism should be a normal inhabitant of the intestinal tract. There is strong evidence in the literature indicating that many of these probiotic bacteria exhibit host specificity; for example, a strain of *Lactobacillus acidophilus* originating in the intestinal tract of a calf should not be expected to function equally well in the intestinal tract of a pig or of a human. Thus, it is desirable to use a selected strain of bacteria that originated in the host species for which the product is to be used. If it is to be used as a probiotic for humans, it is highly desirable that the probiotic organism originated in the human intestines. In order to avoid problems with regulatory agencies, it is desirable that the organism of choice has a history of having been used for producing fermented food products.

Table 5 Important factors to consider in selection of a probiotic culture

- Normal inhabitant of intestine
- Host specificity
- Acid tolerance
- Bile tolerance
- Stability during production and delivery
- Production of desired health/nutritional benefit
- One strain would be unlikely to provide all possible benefits

The organism should be resistant to gastric acidity to survive passage through the stomach. The main organisms considered for use as probiotics (i.e., lactobacilli) are acid-tolerant. However, resistance to acidity varies among strains and species, so it must be considered. If the organism is expected to grow and function in the intestinal tract, it must have a certain degree of bile tolerance. The bile resistance based on the ability of the organism to grow in the presence of bile acids varies tremendously among strains of each species of probiotic bacteria. The minimum level of bile tolerance required is not known, but it is suggested that the most bile-resistant strain that will provide the desired benefit be selected. The probiotic culture must be stable during production and storage of the food or feed containing it. This is often the most challenging factor in providing a probiotic organism that has adequate levels of viability at the time of consumption. The most important characteristic of course is that the probiotic culture produce the desired effect. Therefore, it is desirable that some sort of laboratory screening test for the desired effect be implemented. This is not always easy. However, if the organism for instance is to be used for improving lactose utilization, it would be desirable to select one having a high level of β-galactosidase activity. If it is to be used for controlling serum cholesterol levels, it is desirable that the organism not only assimilate cholesterol but also have a high degree of ability to deconjugate bile acids. In order to develop useful screening tests, it is of course essential that we understand how the probiotic organism functions to provide the desirable benefit. In many cases, we simply do not know the answer to this mechanism at this point.

A major factor in selection of a specific organism for use as a probiotic is to remember that one strain should not be expected to provide all potential benefits. Far more is required in the selection process than to select a culture just because it is *Lactobacillus acidophilus*, for example. There is a great variation among strains of this bacterial species. Such variation also occurs among strains of other species of probiotic bacteria.

Dietary Delivery of the Probiotic

If the probiotic organism is expected to provide its benefit during growth and action in the intestinal tract, it is reasonable to assume that the organism must be viable at the time of consumption. It is also desirable and recommended that the probiotic be consumed on a continuing basis. The numbers of viable probiotic bacteria required in a food or feed product to provide the benefit are not known. The actual numbers required probably will vary with the organism and the expected function. Some have suggested, particularly with regard to milk products, that the probiotic should be present at a level of one to two million per milliliter; however, these numbers are not based on scientific data but are based more on an economic situation reflecting the cost of the organism. Research is needed to answer the question with regard to the optimum numbers required. It is essential that the organism be stable and that the desired characteristics be maintained during the production of the probiotic culture, during its storage, and delivery in the food or feed. This is often the greatest challenge in the use of probiotic bacteria. The stability of the organism during storage and delivery can be greatly influenced by the conditions under which the bacteria are grown. For instance, the growth of *Lactobacillus acidophilus* at pH 5 has resulted in cultures that are more stable during storage when added to refrigerated milk than the same cultures grown at a higher pH. The storage stability, especially during frozen and refrigerated storage, varies among strains of a given bacterial species used as a probiotic. Thus, this becomes a factor to consider in selecting the organism for use as a probiotic.

The use of probiotic bacteria such as *Lactobacillus acidophilus* in conjunction with yogurt can be accomplished by either including the probiotic as part of the starter culture or adding cells of the probiotic bacteria after the product has been manufactured. Of course, it would be desirable to use the probiotic as part of the inoculum (i.e., starter culture) used in the manufacture of yogurt in order to yield higher numbers of the probiotic bacteria during the incubation period for the fermentation. However, most of the probiotic bacteria do not grow well under such conditions, and very little increase in numbers is achieved during the incubation process involved in the manufacture of yogurt. Research is needed to provide means of increasing the numbers of probiotic bacteria that grow during the manufacture of a product such as yogurt when the organism is used as a part of the inoculum for manufacturing the fermented milk. In summary, once a desired probiotic culture has been selected, it is important to develop procedures for producing the culture, storing the culture, and delivering the culture in the food or feed without damaging the desired characteristics.

See also: ***Bifidobacteria* in Foods**; **Cholesterol**: Properties and Determination; Role of Cholesterol in Heart Disease; **Colon**: Cancer of the Colon; **Immunology of Food**; **Lactic Acid Bacteria**; **Lactose**; ***Salmonella***: Salmonellosis; ***Staphylococcus***: Food Poisoning

Further Reading

Bezkorovainy A and Catchpole RM (1989) *Biochemistry and Physiology of Bifidobacteria*. Boca Raton, FL: CRC Press.

Fuller R (ed.) (1992) *Probiotics the Scientific Basis*. New York: Chapman & Hall.

Gilliland SE (1990) Health and nutritional benefits from lactic acid bacteria. *FEMS Microbiology Reviews* 87: 175–188.

Klein G, Pack A, Bonaparte C and Renter G (1998) Taxonomy and physiology of probiotic lactic acid bacteria. *International Journal of Food Microbiology* 41: 103–125.

Ouwehand AC, Kirjavainen PV, Shortt C and Salminen S (1999) Probiotics: mechanisms and established effects. *International Dairy Journal* 9: 43–52.

Robinson RK (ed.) (1991) *Therapeutic Properties of Fermented Milks*. Barking, UK: Elsevier Science.

Sissons JW (1989) Potential of probiotic organisms to prevent diarrhoea and promote digestion in farm animals – a review. *Journal of the Science of Food and Agriculture* 49: 1–13.

Tamine AY and Robinson RK (1999) *Yoghurt Science and Technology*, 2nd edn. Boca Raton, FL: CRC Press.

Process Control *See* **Plant Design**: Basic Principles; Designing for Hygienic Operation; Process Control and Automation; **Instrumentation and Process Control**

Processed Cheese *See* **Cheeses**: Types of Cheese; Starter Cultures Employed in Cheese-making; Chemistry and Microbiology of Maturation; Manufacture of Extra-hard Cheeses; Manufacture of Hard and Semi-hard Varieties of Cheese; Cheeses with 'Eyes'; Soft and Special Varieties; White Brined Varieties; Quarg and Fromage Frais; Processed Cheese; Dietary Importance; Mold-ripened Cheeses: Stilton and Related Varieties; Surface Mold-ripened Cheese Varieties; **Risk Assessment**

PROSTAGLANDINS AND LEUKOTRIENES

S Katayama, Saitama Medical School, Iruma-gun, Saitama, Japan
J B Lee, State University of New York, New York, USA

Background

Prostaglandins (PGs) were discovered in the 1930s and were first chemically isolated and identified in sheep seminal vesicles and renal medulla in the 1960s. Although together with thromboxanes (TXs) and leukotrienes (LTs), the PGs possess the most potent and most divergent biological activities of any naturally occurring compounds, their true physiological role in many instances remains unknown. From a pathophysiological viewpont, an absolute or relative deficiency of PGs relative to TXs has been implicated in the etiology of hypertension, thrombosis, and atherogenesis for many years. Since PGs and TXs are derived from essential fatty acids, the role of dietary intake of these fatty acids in PG and TX production, particularly in relation to cardiovascular system, will be examined. The possible role of dietary essential fatty acids on LT and PG production in relation to inflammation, the immune system, and gastric function will also be discussed.

Chemistry

In the 1930s, it was reported that fresh human semen causes rhythmic contractions of human endometrium. The active substance was called 'prostaglandin.' In the 1960s, PG was determined to be a mixture of biologically active compounds, which were isolated from sheep seminal vesicles and identified as PGE_{1-3} and PGF_{1-2}. These compounds were independently isolated from extracts of rabbit kidney

medulla and identified as PGA_2, PGE_2, and PGF_{2a}. Later, in 1975, the compound TXA_2, which is a potent vasoconstrictor and platelet aggregant, was isolated, and in 1976, prostacyclin (PGI_2) was discovered. Prostacyclin is synthesized mainly by endothelial cells, prevents platelet aggregation, and is vasodilatory. Studies of the slow-reacting substance of anaphylaxis (SRS-A) showed that it was composed of certain LTs, which are products of leucocyte membrane arachidonate.

The PGs are composed of a basic 20-carbon fatty acid chain containing a cyclopentane ring, the so-called hypothetical prostanoic acid. The carbons are numbered 1–20 from the carboxyl to the terminal group. The designations of PGE_1, PGE_2, and PEG_3 refer only to the number of double bonds in the aliphatic side-chains. The PG_2s are the most abundant naturally occurring class. For PG_1s, the precursor is 8,11,14-eicosatrienoic acid (dihomo-γ-linolenic acid), and for PG_2s, the precursor is 5,8,11,14-eicosatetraenoic acid (arachidonic acid). The PG_3s can be formed from 5,8,11,14,17-eicosapentaenoic acid (EPA).

The LTs were named because they were discovered in leucocytes and because the common structural feature is a conjugated triene. Various members of the group have been designated alphabetically, and the subscript denotes the number of double bonds.

Analysis of Prostaglandins and Leukotrienes

Prostaglandins and leukotrienes are present in many different biological samples, including plasma. However, eicosanoid quantification in the peripheral circulation does not accurately reflect *in vitro* biosynthesis and disposition because of rapid turnover, low concentration, metabolism by the lung and liver, and sample processing problems. For example, plasma concentrations of PGE_2 and TXB_2 determined by gas chromatography have been reported to be 12 and several picograms per milliliter, respectively. Urinary samples can be utilized to evaluate *in vivo* PGs or LTs for the kidneys, whereas pleural fluids or bronchial lavages can be used to evaluate pulmonary production. The most accurate reflection of local synthesis is obtained by incubations of slices or homogenates of specific cells and/or tissues with measurement of PG, TX, or LT production *in vitro*.

The most precise method to determine PGs and LTs is undoubtedly gas chromatography linked to mass spectrometry. However, this type of analysis is not widely available. Radioimmunoassay using a specific antibody against various PGs or LTs and ^{3}H- or ^{125}I-labeled tracer are available, highly specific, sensitive, and relatively inexpensive, and have been widely utilized for eicosanoid analysis.

Prior to radioimmunoassay, extraction of arachidonic acid metabolites from plasma, urine or tissue homogenates may be required. The most common method is to extract acidified aqueous solution with an oraganic solvent such as diethylether or ethyl acetate, followed by purification by column chromatography with a octadecylsilyl (ODS) silica (SEP-PAK C-18) cartridge, silicic acid or sometimes by high-performance liquid chromatography.

Biosynthesis of Prostaglandins and Leukotrienes

Arachidonic acid and its metabolite byproducuts (the arachidonic acid cascade) are important mediators of a number of physiological phenomena. As shown in Figure 1, the rate-limiting step in the formation of the metabolites of arachidonic acid seems to be the initial step, i.e., the release of free arachidonic acid from the cell membrane phospholipid pool mediated through activation of phospholipase A_2. Phospholipase C may play another role in arachidonic acid release through liberating a diglyceride, which is then hydrolyzed by another lipase to yield arachidonic acid. Following the release of arachidonic acid, enzymes are involved in its subsequent metabolism. Prostaglandin synthase, i.e., cyclooxygenase (COX), is a key enzyme in controlling the extent of prostanoid biosynthessis. It possesses two enzymatic activities: COX activity, the conversion of arachidonic acid (AA) to PGG_2 and peroxidase activity, the conversion of PGG_2 to PGH_2 catalyzing a two-electron reduction of PGG_2 to PGH_2. PGH_2 is the common precursor for a number of biologically important prostanoids, i.e., PGI_2, TXA_2, PGE_2, PGD_2, and $PGF_{2\alpha}$, the formation of which occurs through the action of the respective synthetase called PGI synthase, TXA synthase, PGE synthase, PGD synthase, and PGF synthase. COX is distributed in most mammalian tissues. Two isoforms of COX have been identified and cloned. COX-1, originally cloned from ram seminal vesicles and subsequently from mouse tissues and human endothelial cells and platelets, is considered to be constitutive, whereas COX-2, cloned from chicken embryonic fibroblasts and mouse 3T3 fibroblasts, is highly inducible. COX-1 activity is inactivated by aspirin through acetylation of the serine residue and inhibited by other nonsteroidal antiinflammatory drugs (NSAIDs), whereas COX-2 is inhibited by glucocorticoids. Various biological activities are due to binding of prostanoids to the receptors on the cell surface. With the advent of the molecular biology technique, eight prostanoid receptors have been

cloned so far; TP, IP, FP, DP and four classes of EPs (EP_1, EP_2, EP_3, and EP_4). Prostanoid receptors have seven transmembrene domains and are members of the superfamily of the rhodopsin receptor. These receptors are classified into three groups; IP, DP, EP_2, and EP_4 are the stimulatory receptors, which generate cAMP, resulting in vasodilatoion; TP, FP, and EP_1 are the G protein-linked receptors that increase intracellular Ca^{2+} concentration, resulting in vasoconstriction; and EP_3 is an inhibitory receptor, which inhibits the elevation of cAMP, resulting in vasodilation.

In the second pathway of arachidonic acid metabolism, 5-lipoxygenase forms a series of products named LTs, as illustrated in Figure 1. 5-Hydroperoxyeicosatetraenoicacid (5-HPETE) may be enzymatically dehydrated to LTA_4 (5(S)-*trans*-oxido-7, 9-*trans*-11, 14-*cis*-eicosa-tetraenoic acid). Subsequent enzymatic hydrolysis of LTA_4 results in LTB_4 (5(S)-12(*R*)-dihydroxy-6-*cis*-8, 10-*trans*-14-*cis*-eicosatetraenoic acid). LTA_4 also reacts with sulfydryl compounds by addition of gluthathione (Glu-Gys-Gly) by glutathione-(*S*)-transferase, yielding LTC_4 (5(*S*)-glutathionyl-7, 9-*trans*-11, 14-*cis*-eicosatetraenoic acid). LTD_4 can be then produced from LTC_4 through elimination of glutamyl residue by a γ-glutamyl transpeptidase. The remaining peptide bond in LTD_4 is hydrolyzed to give LTE_4 by a dipeptidase.

Physiological Action of PGs and LTs

The relative amounts of compound formed depend on the tissue or cell being studied. When platelets are activated by collagen or thrombin, unstable TXA_2 is predominant, resulting in a rapid and irreversible aggregation of platelets and contraction of vascular smooth muscle cells. However, vascular endothelial

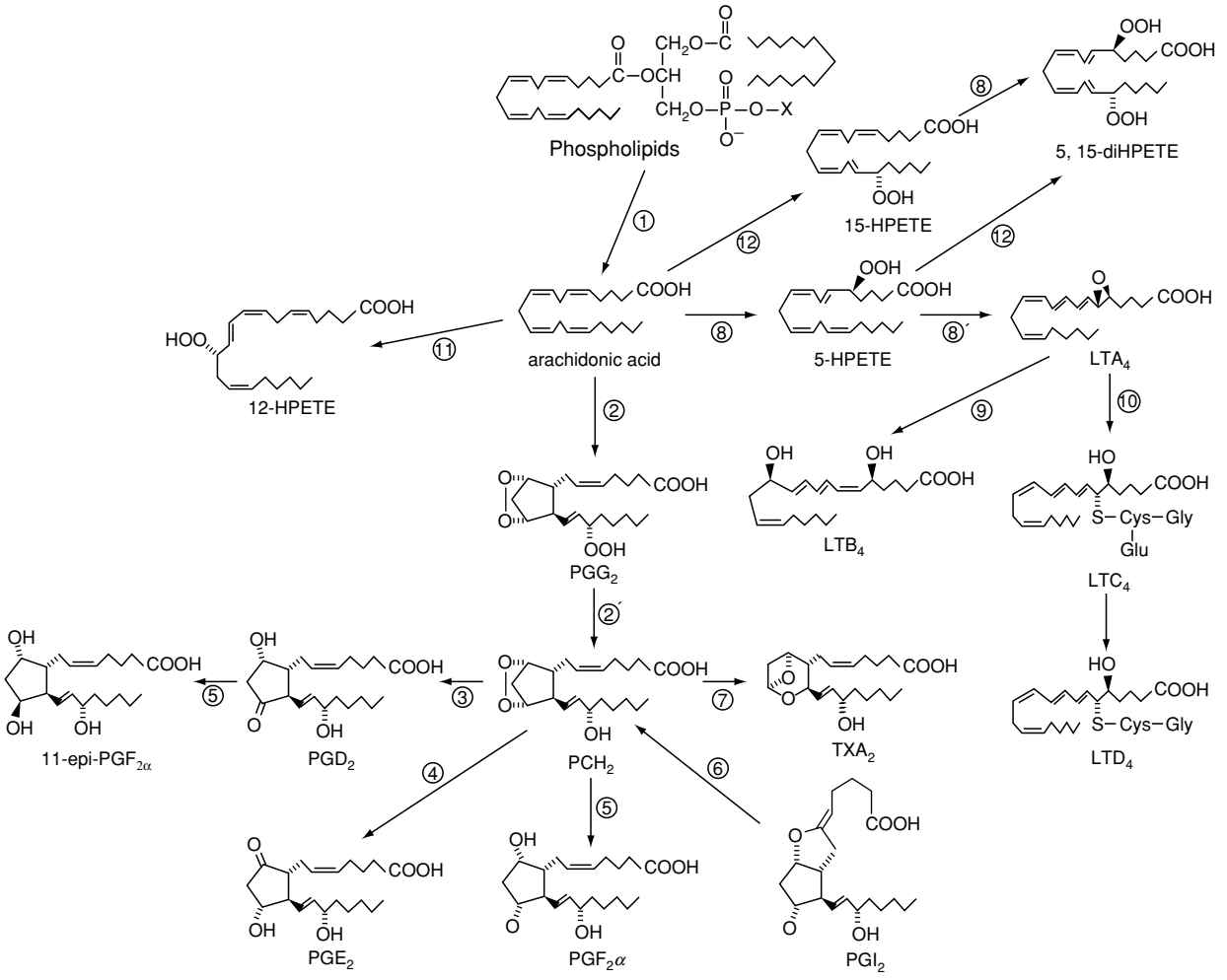

Figure 1 Metabolic pathways of prostaglandin (PG) and leukotriene (LT) from arachidonic acid. The name of enzyme that catalyzes each step is as follows: 1, phospholipase A_2; 2, cyclooxygenase; 3, PGD synthase; 4, PGE synthase; 5, PGF synthase; 6, PGI synthase; 7, TXA synthase; 8, 5-lipoxygenase; 8′, dehydrase; 9, hydolase; 10, glutathione-*S*-transferase; 11, 12-lipoxygenase; 12, 15-lipoxygenase.

cells form unstable PGI_2, which has a potent vasodilating and antiaggregatory action similar to stable PGE_2, Neutrophils as well as monocytes or macrophages produce PGE_2, whereas PGD_2 is formed in mast cells and basophils. In kidneys, renal interstitial cells and collecting duct cells synthesize PGE_2, collecting dust cells synthesize PGE_2, the renovasculature endothelia synthesize PGI_2, and glomerular mesangial endothelia from PGF_{2a}. In addition to PGE_2, other PGs are also produced in the gastrointestinal tract. PGE_1, PGE_2, and their analogs, administered exogenously, suppress gastric acid secretion and possess the property of protecting the gastric mucosa against necrotizing agents such as absolute ethanol or hydrochloric acid (cytoprotection). This is the reason why NSAIDs cause a variety of problems in the gastrointestinal tract including irritation and ulceration of the stomach. Since COX-1 appears to play a major role, treatment with COX-2 selective inhibitors appears to be associated with less gastrointestinal damage than conventional NSAIDs.

Prostaglandins play many physiological roles in human reproduction, e.g., menstrual regulation, pregnancy, and induction of labor. PGE_1, PGE_2, or their analogs are now widely used clinically for induction of labor or abortion. PGE_2 and PGI_2 produce or augment vasodilation and enhance bradykinin or serotonin-induced pain, resulting in redness, heat, swelling, and pain in the inflammatory processes. PGs such as PGE_2 and PGI_2 have also been shown to increase during pyrogen-induced fever. In fact, PGE_2 administered into the hypothalamus increases body temperature, whereas PGD_2 deceases body temperature, indicating a possible role of PGs in human thermoregulation. PGs, TXs, and some monohydroxy derivatives of arachidonic acid (hydroxyeicosatetraenoic acids) also have chemotactic effects on polymophonuclear leucocytes. LTB_4 is chemotactic not only for neutrophils and eosinophils but also for monocytic macrophages. These phenomena are considered to be the initial events of the inflammatory response. Polymorphonuclear leucocytes only survive for hours in the inflammatory area, whereas monocytes may stay for weeks and finally may be transformed to fibroblasts, initiating the repair process of the wounds. Furthermore, monocytes can present antigens to cells capable of producing antibodies and can synthesize all the members of the arachidonic acid cascade. Monocytes are also capable of forming interleukin, interferon, complement, and protease, all of which participate in tissue disruption. The lipoxygenase system is more important in subacute and chronic inflammation, whereas the COX system produces or modulates acute inflammation. The antiinflammatory action of aspirin, indomethacin, and other nonsteroidal antiinflammatory agents is produced almost entirely by PG synthesis inhibition, as shown by studies revealing inhibition of PG-cyclooxygenase with aspirin.

LTC_4, LTD_4, LTE_4, and their trans isomers do not have any effects on chemotaxis, enzyme release, or leucocyte aggregation. However, they possess potent biological activities, which were formerly attributed to SRS-A release from sensitized lungs treated with a specific antigen. These compounds may be important mediators in asthma and other acute hypersensitivity reactions. Contraction of guinea-pig ileum and other smooth muscle by LTs exhibits a slow onset and relaxation, which is the basis of the original designation as SRS-A. This distinguishes these substances from histamine, bradykinin, and PGF_{2a}. LTC_4 and LTD_4 are respectively about 200-fold and 20 000-fold more potent than histamine in promoting small airway contraction. In addition, these LTs cause rapid arteriolar contraction and promote plasma leakage in postcapillary venules. They also slow the rate of mucus clearance from the airways of patients with asthma after the inhalation of antigen and increase the amount of mucus glycoprotein synthesis in the airways.

With the advent of cloning of various prostanoid synthetases and their receptors, vast numbers of prostanoids analogs synthesized so far will be retested. Diverse physiological actions of various prostanoids will be shortly reevaluated and reclassified, based on their specific receptors using a specific agonist and antagonist. For example, inhibition of vasopressin-induced water reabsorption in kidney or histamine-induced gastric acid secretion can be explained by EP_3-mediated inhibition of adenylate cyclase. Furthermore, genetic engineering such as prostanoid receptor knockout mice may be able to clarify a specific physiological role of PGs and LTs.

Effects of Diet on PG Formation

Polyunsaturated Fatty Acids

It was first shown in 1974 that dietary deficiency of essential fatty acids in animals results in hypertension, which depends on a high sodium intake, thus being very similar to renoprival hypertension or experimental 'salt sensitive' hypertension. This was later confirmed in 1983, and the blood-pressure response was attributed to a dietary deficiency of linoleic acid. However, it would appear unlikely that this is a cause of human hypertension, since selective deficiency of essential fatty acids in the human diet is rare.

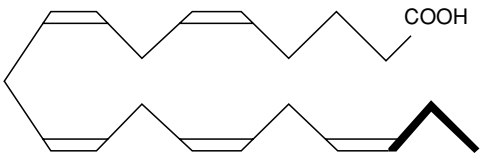

Figure 2 Eicosapentaenoic acid (EPA).

A low prevalence of atherosclerosis and a low mortality from myocardial infarction among Greenland Inuit, despite a diet high in fat and cholesterol, suggested that dietary fish may have some properties that could prevent coronary artery disease. The Inuit consume 5–10 g daily of the long-chain *n*-3 polyunsaturated fatty acids EPA (C20; 5*n*-3) (Figure 2) and docosahexaenoic acid (C22; 6*n*-3), which are present in fish oils. When *n*-3 fatty acids are included in the diet, EPA inhibits conversion of linoleic acid and competes with arachidonic acid for the 2-acyl position in membrane phospholipids, reducing plasma and cellular levels of arachidonic acid. In addition, EPA competes with arachidonic acid as the substrate for cyclooxygenase. The platelets produce biologically inactive TXA_3 instead of TXA_2. However, PGI_2 synthesis in endothelial cells is not markedly inhibited, and any newly synthesized PGI_3 has the same potency as PGI_2. Thus, a diet abundant in EPA may inhibit platelet aggregation and cause vascular dilatation, resulting in blood-pressure reduction. In fact, a metaanalysis including 1356 subjects from 31 placebo-controlled trials confirmed the blood-pressure-lowering effects of dietary fish oils. The hypotensive effect was dose-dependent; there was little change in blood pressure at doses lower than 3 g per day (corresponding to 250 g of salmon), but above this, a 0.66 mmHg fall in systolic blood pressure was demonstrated per 1.0-g increase in *n*-3 polyunsaturated fatty acid. The hypotensive effect appeared greater in patients with hypertension, hypercholesterolemia, or coronary heart disease. Mechanisms by which *n*-3 polyunsaturated fatty acids may act might be attributed to an attenuated constrictor responses to vasoactive agents, such as noradrenaline and angiotensin II, an attenuated sympatho-adrenal stimulation and/or modulated production or release of nitric oxide from vascular endothelial cells. However, the safety and long-term benefits of this treatment have been questioned because PGE_2 synthesis is inhibited and replaced by PEG_3 from EPA in the fish oils. Since renal blood flow is dependent on PGE_2 in renal disease and during volume contraction, any decrease in PGE_2 might result in more severely compromised renal function.

The other merit of dietary fish or fish oil supplements may be its lowering effect on plasma lipid. The principal effect has been found to be a reduced production of triacyglycerols and very-low-density lipoprotein (VLDL) in the liver. Because VLDL is a precursor of low-density lipoprotein (LDL), a secondary reduction in LDL cholesterol is also seen. Such conditions are considered to reduce the atherogenicity of lipoprotein particles. Futhermore, *n*-3 fatty acids have been reported to improve insulin sensitivity in skeletal muscle, possibly through a decrease in triacylglycerol accumulation. However, fish oil supplementation to diabetic patients resulted in no change in glycemic control in one study and an actual deterioration in glycemic control in another study.

Increased intake of polyunsaturated fatty acids has also been reported to result in reduced gastric acid secretion and an increase in gastric mucosal protection. Diets enriched in fish oil or dihomo-γ-linolenic acid also reduce LTB_4 formation in leucocytes or monocytes, suggesting a possible role for fish oil diets in suppressing inflammation in such disorders as rheumatoid arthritis or systemic lupus erythematosus.

Carbohydrate

The increase in urinary excretion of TXB2 associated with a decreased 6-keto-$PGF_{1\alpha}$ has been demonstrated in diabetic animals and humans. An overproduction of PGE_2 has also been considered to be one of the reasons for glomerular hyperfiltration. In fact, enhanced growth of mesangial cells under high-glucose conditions can be explained, at least in part, by a decreased cyclic adenosine monophosphate production via EP_4, which augments EP_1 function in conjunction with the overproduction of PGE_2. Diabetic pregnancy is associated with growth disturbances and malformations in the offspring. Embryonic dysmorphogenesis induced by high concentration of glucose is protected by addition of arachidonic acid or PGE_2. COX inhibitors such as indomethacin or aspirin yield embryonic dysmorphogensesis similar to that caused by glucose, indicating that this disturbance may be partly mediated via altered metabolism of arachidonic acid.

Proteins

Protein intake has profound effect on renal hemodynamics and excretory function. A high protein intake has been shown to have deleterious effects on the kidneys, especially in preexisting renal disease, indicating a rationale for dietary protein restriction to lessen renal parenchymal damage. However, not all proteins are equipotent in relation to their renal

effects. A lower glomerular filtration rate and renal plasma flow, as well as reduced excretion of albumin and some proteins, have been reported in vegan and lactovegetarian individuals compared with omnivorous subjects. In fact, acute (more than 80 g) or chronic loading (more than 1 g per kg of body weight per day) with animal proteins has resulted in an increase in renal plasma flow and glomerular filtration associated with a higher clearance of albumin than that on vegetable protein diets. Vasodilatory PGs such as PGI_2, which may be partly dependent on a rise in glucagon secretion from the pancreas, have been proposed as a possible mediator of such renal effects of animal proteins, since the administration of indomethacin was found to abolish the renal response to a meat meal.

Minerals

The isolation and identification of renal PGE_2 and PGA_2 between 1965 and 1967 provided biochemical support for the renal anitihypertensive-endocrine function. Numerous studies in humans and animals thereafter have shown that PGE_2 has a potent vasodilatory and natriuretic action and hence a pivotal role in sodium homeostasis. In fact, PGE_2 administration augments renal blood flow and urinary sodium excretion. In addition, PGE_2 stimulates renal renin secretion and inhibits vasopressin-mediated water reabsorption. During a low sodium intake, the renin–angiotensin axis is activated by volume depletion, resulting in antinatiuresis and vasoconstriction. Renal PGE_2 synthesis is stimulated by elevated circulating angiotensin II, which then counteracts the action of angiotensin II, leading to normalization of renal blood flow, blood pressure, and sodium excretion (Figure 3). This is the rationale for the beneficial effects of a low-sodium diet on these cardiovascular parameters. The same can be said for diuretic therapy, since the renomedullary synthesis of PGE_2 from arachidonic acid is directly stimulated by furosemide. The natriurietic and antihypertensiv effects of furosemide are mediated by this newly synthesized PGE_2, since both effects are inhibited by indomethacin. While COX-1 has long been recognized to be involved in normal kidney function, COX-2 is now seen to have a distinct role. COX-2 has been localized in both the macula densa and the interstitial cells of the medulla. Chronic sodium deprivation was found to increase COX-2 levels in the macula densa, and COX-2-generated prostanoids may be important mediators of renin production and tubuloglomerular feedback.

Dietary potassium deficiency leads to hypokalemia and impairment of maximal urinary concentrating ability. Since PGE_2 antagonizes vasopressin-induced water reabsorption, an increase in renomedullary PGE_2 production was postulated to underlie this concentrating defect. However, during chronic dietary potassium depletion secondary to dietary potassium deprivation in animals, PGE_2 synthesis by the renal medulla is markedly reduced, and thus current evidence does not support a role for PGs in hypokalemic polyuria.

Oral calcium supplementation has been reported to decrease blood pressure in some patients with essential hypertension and hypertensive rats. However, in other studies, calcium supplementation produced no effect on blood pressure in normotensive or hypertensive subjects, Although calcium supplementation has been shown to increase urinary PGE_2 excretion in normotensive women, there was no associated fall in blood pressure.

Spices

Spices have been consumed by many cultures for thousands of years. In fact, in classical Indian (Ayurvedism), Chinese, or Graeco-Arabic systems of medicine, some spice-containing plant materials were used to prevent or treat some diseases. Some spices have

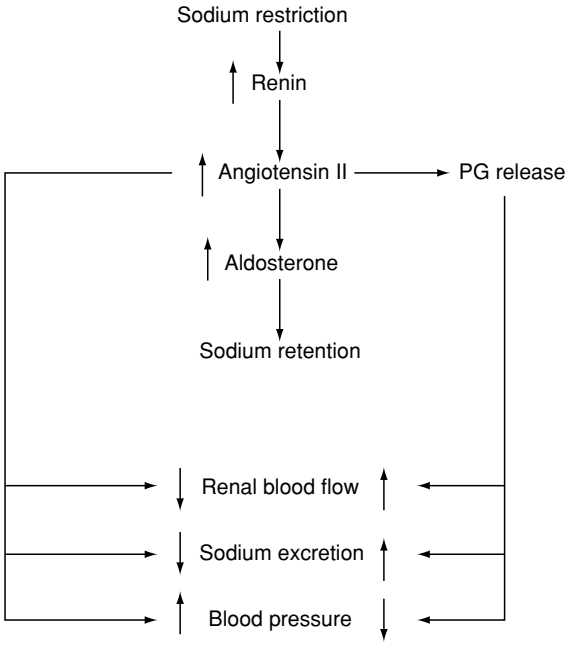

Figure 3 Scheme whereby volume depletion secondary to dietary sodium restriction leads to an increase in PG release from the renal medulla and resultant physiological actions of angiotensin II. From Lee JB (1980) Prostaglandins and the renin–angiotensin axis. *Clinical Nephrology* 14: 159–163, with permission.

been shown to inhibit platelet aggregation. These include onion, garlic, ginger, cloves, omum, cumin, and turmeric. Extracts from these spices were found to inhibit platelet COX, allowing less TXA_2 to be produced. As COX is inhibited, more substrate (arachidonic acid) becomes available to the lipoxygenase pathway, as shown by the observation that 12-HPETE increases in platelets treated with extracts from ginger, cloves, cumin, and turmeric. However, onion extracts have been demonstrated to inhibit the 5- or 12-lipoxygenase in platelets. This might at least partially explain the observation that crude ethanolic extracts of onion prevented allergen-induced bronchoconstriction in animals and humans. Ginger, which is demonstrated to have both antihistaminic and antioxidant action, also inhibits both the COX and 5-lipoxygenase pathway. Garlic extracts have also been demonstrated to inhibit COX activity and LTC4 formation.

See also: **Calcium**: Physiology; **Fatty Acids**: Metabolism; Dietary Importance; *Trans*-fatty Acids: Health Effects; **Fish**: Dietary Importance of Fish and Shellfish; **Hypertension**: Physiology; Nutrition in the Diabetic Hypertensive; **Potassium**: Properties and Determination; Physiology; **Renal Function and Disorders**: Nutritional Management of Renal Disorders; **Sodium**: Physiology

Further Reading

Bergstrom S, Ryhage R, Samuelsson B and Sjövall J (1962) The structure of prostaglandin E_1, F_1 and F_2. *Acta Chemica Scandinavica* 16: 501–507.

Chin J and Dart AM (1995) How do fish oils affect vascular function? *Clinical and Experimental Pharmacology and Physiology* 22: 71–81.

Fiorelto P, Trevisan R, Valeho A *et al*. (1990) Impaired renal response to a meat meal in insulin-dependent diabetes: role of glucagon and prostaglandins. *American Journal of Physiology* 258: F675–683.

Katayama S, Maruno Y, Inaba M *et al*. (1991) Effect of dietary calcium on renal prostaglandins. *Postaglandins, Leukotrienes and Essential Fatty Acids* 42: 197–200.

Knapp HR and Fitzgerald GA (1989) The antihypertensive effects of fish oil. *New England Journal of Medicine* 320: 1037–1043.

Kontessis P, Jones S, Dodds R *et al*. (1990) Metabolic and hormonal responses to ingestion of animal and vegetable proteins. *Kidney International* 38: 136–144.

Lee JB and Katayama S (1985) Prostaglandins, thromboxanes and leukotrienes. In: Wilson JD and Foster DW (eds) *Williams Textbook of Endocrinology*. Philadelphia, PA: WB Saunders.

Moncada S, Gryglewski R, Bunting S and Vane JR (1976) An enzyme isolated from arteries transforms prostaglandin an unstable substance that inhibits platelet aggregation. *Nature* 263: 633–665.

Morris MC, Sacks F and Rosner B (1993) Does fish oil lower bood pressure? A meta-analysis. *Circulation* 88: 523–533.

Narumiya S, Sugimoto Y and Ushikubi F (1999) Prostanoids receptors: Structures, properties, and functions. *Physiological* Reviews 79: 1193–1226.

Oates JA, Fitzgerald GA, Branoh RA *et al*. (1988) Clnical implications of prostaglandin and thromboxane A_2 formation. *New England Journal of Medicine* 319: 689–698, 761–767.

Samuelsson B (1983) Leukotrienes: Mediators of immediate hyper-sensitivity reaction and inflammation. *Science* 220: 568–575.

Vane JR, Bakhle S and Botting RM (1998) Cyclooxygenase 1 and 2. *Annual Review of Pharmacology and Toxicology* 38: 97–120.

Wentzel P and Ericksson UJ (1998) Antioxidants diminish developmental damage induced by high glucose and cyclooxygenase inhibitors in rat embryos *in vitro*. *Diabetes* 47: 677–684.

PROTEIN

Contents

Chemistry

J K P Weder and H-D Belitz[†], Technische Universität München, Garching, Germany

Background

Like peptides, proteins are formed from amino acids through amide linkages. Covalently bound hetero constituents can also be incorporated into proteins. For example, phosphoproteins such as milk casein or phosvitin of egg yolk contain phosphoric acid esters of serine and threonine residues. Glycoproteins, such as κ-casein, various components of egg white and egg yolk, collagen from connective tissue, and serum proteins of some species of fish contain one or more monosaccharide or oligosaccharide units bound *O*-glycosidically to serine, threonine, or δ-hydroxylysine or *N*-glycosidically to asparagine.

The structure of a protein is dependent on the amino acid sequence (the primary structure) which determines the molecular conformation (secondary and tertiary structures). Proteins sometimes occur as molecular aggregates which are arranged in an orderly geometric fashion (quaternary structure).

Amino Acid Sequence

Amino Acid Composition and Subunits

Sequence analysis can only be conducted on a pure protein. First, the amino acid composition is determined after acid hydrolysis. The procedure (separation on a single cation exchange resin column and color development with ninhydrin reagent) has been standardized and automated (amino acid analyzers). More recently, derivatization of amino acids to form fluorescent derivatives, e.g., with *o*-phthaldialdehyde (OPA) and 2-mercaptoethanol, followed by reversed-phase high-performance liquid chromatography (RP-HPLC) has been introduced (precolumn derivatization). The use of chiral thiols in this procedure allows the separation of amino acid enantiomers.

Selenocysteine has been detected as a relatively new protein amino acid in glutathione peroxidase (EC 1.11.1.9) in 1973, and later also in other oxidoreductases and other proteins (e.g., in selenoprotein P, 10 residues per molecule). The incorporation of selenocysteine occurs during ribosomal protein synthesis cotranslationally by a complex mechanism in which a specific $tRNA^{(ser)sec}$ is first loaded with serine, then transformed into selenocysteyl-$tRNA^{(ser)sec}$ with selenophosphate as the selenium donor and bound with its anticodon region to a UGA codon of the mRNA. Some selenocysteine, however, is also formed by posttranslational selenium–sulfur exchange, e.g., in selenoprotein P.

It is also necessary to know the molecular weight of the protein. This is determined by gel column chromatography, ultracentrifugation, sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) or mass spectrometry (matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, MALDI-TOF-MS, or electrospray ionization mass spectrometry, ESI-MS). Furthermore, it is necessary to determine whether the protein is a single molecule or consists of a number of identical or different polypeptide chains (subunits) associated through disulfide bonds or noncovalent bonding.

[†] Deceased.

Dissociation into subunits can be accomplished by a change in pH, by chemical modification of the protein, such as by succinylation, or with denaturing agents (urea, guanidine hydrochloride, sodium dodecyl sulfate). Disulfide bonds, which are also found in proteins consisting of only one peptide chain, can be cleaved by oxidation of cystine to cysteic acid or by reduction to cysteine with subsequent alkylation of the thiol group to prevent reoxidation. Separation of subunits is achieved by chromatographic or electrophoretic methods.

Terminal Groups

N-Terminal amino acids can be determined by treating a protein with 2,4-dinitrofluorobenzene (DNFB, Sanger's reagent) or 5-(dimethylamino)naphthalene-1-sulfonyl chloride (dansyl chloride). Another possibility is the reaction with cyanate, followed by elimination of the *N*-terminal amino acid in the form of hydantoin, and separation and recovery of the amino acid by cleavage of hydantoin. The *N*-terminal amino acid (and the amino acid sequence close to the *N*-terminus) is accessible by hydrolysis with aminopeptidase M (EC 3.4.11.2), in which case it should be remembered that the hydrolysis rate is dependent on amino acid side chains and that proline residues are not cleaved. A special procedure is required when the *N*-terminal residue is acylated (*N*-formyl- or *N*-acetyl-amino acids, or pyroglutamic acid).

Determination of *C*-terminal amino acids is possible via the hydrazinolysis procedure. The *C*-terminal amino acid is separated from the amino acid hydrazides, e.g., by a cation exchange resin, and identified. The *C*-terminal amino acids can be removed enzymatically by carboxypeptidase A (EC 3.4.17.1), which preferentially cleaves amino acids with aromatic and large aliphatic side chains; carboxypeptidase B (EC 3.4.17.2), which preferentially cleaves lysine, arginine, and amino acids with neutral side chains; carboxypeptidase C (EC 3.4.16.1), which cleaves with less specificity and cleaves proline; carboxypeptidase P (EC 3.4. 16.1), hydrolyzing almost all amino acids (including proline, aspartic acid, and glutamic acid; release of serine and glycine is retarded); or carboxypeptidase Y (EC 3.4. 16.1), which shows high catalysis rate if the penultimate and/or terminal residue is an aromatic or aliphatic amino acid including proline (release of glycine and aspartic acid is considerably retarded).

Partial Hydrolysis

Longer peptide chains are usually fragmented. The fragments are then separated and analyzed individually for amino acid sequences. Selective enzymatic cleavage of peptide bonds is accomplished with trypsin (EC 3.4.21.4), which only cleaves Lys-X and Arg-X bonds; chymotrypsin (EC 3.4.21.1), which cleaves peptide bonds with less specificity (Tyr-X, Phe-X, Trp-X and Leu-X); and, more recently, with endoproteinase Arg-C (EC 3.4.22.8), hydrolyzing only Arg-X bonds; endoproteinase Asp-N (EC 3.4.24.33), hydrolyzing X-Asp and X-CySO_3H bonds; endoproteinase Glu-C (EC 3.4.21.19), hydrolyzing Glu-X or Glu-X and Asp-X bonds depending on pH and buffer; or endoproteinase Lys-C (EC 3.4.21.50), hydrolyzing only Lys-X bonds. The enzymatic attack can be influenced by modification of the protein.

The most important chemical method for selective cleavage uses cyanogen bromide (BrCN) to attack Met-X linkages. Separation of peptide fragments is achieved by gel and ion exchange column chromatography using a volatile buffer as the eluent (e.g., pyridine, morpholine acetate) which can be removed by freeze-drying of the fractions collected. Recently the separation of peptides and proteins by RP-HPLC has gained great importance, using volatile buffers mixed with organic, water-soluble solvents (e.g., acetonitrile) as the mobile phase.

The fragmentation of the protein is performed by different enzymatic and/or chemical techniques, at least by two enzymes of different specificity. The arrangement of the peptides obtained, in the same order as they occur in the intact protein, is accomplished with the aid of overlapping sequences.

Sequence Analysis

The Edman degradation is by far the most important method in sequence analysis. It involves stepwise degradation of peptides with phenyl isothiocyanate, starting at the *N*-terminus of the polypeptide. The resultant phenylthiohydantoin is either identified directly or the amino acid is recovered. The stepwise reactions are performed in solution or on peptide bound to a carrier, i.e., to a solid phase. Both approaches have been automated (sequencer). Carriers used include resins containing amino groups (e.g., aminopolystyrene) or glass beads treated with aminoalkylsiloxane. The peptides are then attached to the carrier by carboxyl groups (activation with carbodiimide or carbonyl diimidazole, as in peptide synthesis) or by amino groups. Nowadays, Edman degradation has been miniaturized and is performed on gas-phase or pulsed-liquid-phase sequencers, in which the peptide is adsorbed on a polybrene-treated glass-fiber disk and the reagents are applied either as vapors in a carrier gas stream or in small liquid doses separated by carrier gas bubbles.

Methods for C-terminal sequencing have also been reported. After activating the polypeptide at the α-carboxyl group with acetic anhydride, the activated polypeptide is reacted with thiocyanic acid (formed from ammonium thiocyanate with an acid) and the resulting thiohydantoin is cleaved with a base to give the thiohydantoin of the C-terminal amino acid and the polypeptide reduced by one amino acid residue. Another procedure uses diphenyl phosphoroisothiocyanatidate (DPP-ITC), pyridine for reaction and cyclization, and sodium trimethylsilanoate for cleavage. A modification of the first procedure has been automated including *S*-alkylation of the thiohydantoin with 2-(bromomethyl)-naphthalene and cleavage of the C-terminal amino acid derivative simultaneously with the formation of the next thiohydantoin by addition of ammonium thiocyanate plus trifluoroacetic acid. It should be mentioned that C-terminal sequencing procedures are by far less sensitive than Edman degradation and that only a few steps can be interpreted without doubt.

Methods other than chemical degradations can provide additional information. These include determination of terminal residues with amino- and carboxypeptidases, as already discussed, or mass spectrometry. Mass spectrometric sequencing can be done by MALDI-TOF-MS (peptide ladder sequencing and postsource decay (PSD)-MALDI) or ESI-MS (atmospheric pressure ionization collision-induced decay, API-CID; single or multiple fragmentation of selected ions, MS-MS or MS^n). Peptide ladder sequencing combines a modified Edman degradation (incomplete by the addition of 5% phenyl isocyanate) with the determination of the molecular masses of the resulting mixture consisting of phenylcarbamoyl derivatives of peptides stepwise reduced by one amino acid. Because of the small difference between the molecular mass of aspartic and glutamic acid and their amides, results can only be interpreted up to 5 kDa. The other two methods include fragmentation of the peptide chain, and unknown sequences can only be interpreted up to 2 kDa without doubts. The main area of mass spectrometric sequencing is the confirmation of already known sequences. None of the methods can differentiate between leucine and isoleucine.

Recently, amino acid sequences of proteins have increasingly been deduced from the nucleotide sequences of the corresponding genes. Examples of food proteins are gliadins and glutenins from wheat. If the gene responsible for the expression of a protein is known, the sequence analysis of the nucleic acid can be performed much easier than that of the protein itself.

Examples for amino acid sequences of food proteins are summarized in Tables 1–3 (legume 11S globulins, legume 7S globulins, wheat proteins). Further examples (milk proteins, collagen, monellin, and thaumatin) are presented in the first edition of this Encyclopedia.

Conformation

Information about conformation is available through X-ray crystallographic analysis of protein crystals and through the determination of the H–H distances by proton nuclear magnetic resonance (NMR) spectroscopy in solution. In principle, the conformation of the protein in crystalline form can be assumed to be similar to that of the protein in solution. In 1960 the structure of myoglobin (17.8 kDa) was elucidated with a resolution of 0.2 nm. Individual atoms are well revealed at 0.11 nm. Such a resolution has only been achieved with few proteins. Reliable localization of the C^{α} atoms of the peptide chain requires a resolution of less than 0.3 nm. Most protein conformations have been studied with a resolution between 0.2 and 0.3 nm and, more recently, between 0.15 and 0.2 nm.

Extended Peptide Chains

X-ray structural analysis and other physical measurements of a fully extended peptide chain reveal the lengths and angles of bonds. The peptide bonds have partial (40%) double-bond character with π-electrons shared between the C′–O and C′–N bonds. The resonance energy is about 83.6 kJ mol^{-1}. Normally the bond has a *trans* configuration, i.e., the oxygen of the carbonyl group and the hydrogen of the NH group are in the *trans* position; a *cis* configuration only occurs in exceptional cases (e.g., in small cyclic peptides or in proteins before proline residues). Because of the partial double-bond character, six atoms of the peptide bonds, C_i^{α}, C_i', O_i, N_{i+1}, H_{i+1} and C_{i+1}^{α}, lie in one plane. For a *trans* peptide bond, ω_i is 180°. The position of two neighboring planes is determined by the numerical value of the angles ψ_i (rotational bond between a carbonyl carbon and an α-carbon) and ϕ_i (rotational bond between an amide nitrogen and an α-carbon). For an extended peptide chain $\psi_i = 180°$ and $\phi_i = 180°$. The position of side chains can also be described by a series of angles χ_i^{1-n}.

Secondary Structure (Regular Structural Elements)

The primary structure gives the sequence of amino acids in a protein chain whereas the secondary and tertiary structure reveal the arrangement of the chain in space. The peptide chains are not in an extended or

Table 1 Amino acid sequences of 11S globulins from legumes

Glycinin A_2B_{1a} from soya bean, α polypeptide

```
          5         0         5         0         5         0         5         0         5         0
  1 L R E Q A Q Q N E C Q I Q K L N A L K P D N R I E S E G G F I E T W N P N N K P F Q C A G V A L S R
 51 C T L N R N A L R R P S Y T N G P Q E I Y I Q Q G N G I F G M I F P G C P S T Y Q E P Q E S Q Q R G
101 R S Q R P Q D R H Q K V H R F R E G D L I A V P T G V A W W M Y N N E D T P V V A V S I I D T N S L
151 E N Q L D Q M P R R F Y L A G N Q E Q E F L K Y Q Q Q Q Q G G S Q S Q K G K Q Q E E E N E G S N I L
201 S G F A P E F L K E A F G V N M Q I V R N L Q G E N E E E D S G A I V T V K G G L R V T A P A M R K
251 P Q Q E E D D D D E E E Q P Q C V E T D K G C Q R Q S K R S R N
```

Glycinin A_2B_{1a} from soya bean, α polypeptide

```
          5         0         5         0         5         0         5         0         5         0
  1 G I D E T I C T M R L R Q N I G Q N S S P D I Y N P Q A G S I T T A T S L D F P A L W L L K L S A Q
 51 Y G S L R K N A M F V P H Y T L N A N S I I Y A L N G R A L V Q V V N C N G E R V F D G E L Q E G G
101 V L I V P Q N F A V A A K S Q S D N F E Y V S F K T N D R P S I G N L A G A N S L L N A L P E E V I
151 Q H T F N L K S Q Q A R Q V K N N N P F S F L V P P Q E S Q R R A V A
```

Legumin J from pea, β polypeptide

```
          5         0         5         0         5         0         5         0         5         0
  1 L A T S S E F D R L N Q C Q L D S I N A L E P D H R V V S E A G L T E T W N P N H P E L K C A G V S
 51 L I R R T I D P N G L H L P S F S P S P Q L I F I I Q G K G V L G L S F P G C P E T Y E E P R S S Q
101 S R Q E S R Q Q Q G D S H Q K V R R F R K G D I I A I P S G I P Y W T Y N H G D E P L V A I S L L D
151 T S N I A N Q L D S T P R V F Y L G G N P E T E F P E T Q E E Q Q G R H R Q K H S Y P V G R R S G H
201 H Q Q E E E S E Q Q N E G N S V L S G F S S E F L A Q T F N T E E D T A K R L R S P R D E R S Q I V
251 R V E G G L R I I K G R T E E E K E Q S H S H S H R E E K E E E E E E E E E D E E E K Q R S E E R K N
```

Legumin J from pea, β polypeptide

```
          5         0         5         0         5         0         5         0         5         0
  1 G L E E T I C S A K I R E N I A D A A R A D L Y N P R A G R I S T A N S L T L P V L R Y L R L S A E
 51 Y V R L Y R N G I Y A P H W N I N A N S L L Y V I R G E G R V R I R S C E L P T N T M F D N K L R K
101 G H L V V V P Q N F V V A E Q A G E E E G L E Y V V F K T N D R A A V S H V Q Q V F R A T P S E V L
151 A N A F G L R Q R Q V T E L K L S G N R G P M V H P R S Q S Q S H
```

Amino acids are given in the International Union of Pure and Applied Chemistry one-letter notation. The two polypeptides of glycinin and legumin are connected by a disulfide bond Cys(α84)-Cys(β7) and Cys(α89)-Cys(β7), respectively. The two remaining cysteine residues of the α subunits (10/43 and 13/46, respectively) probably form an intrachain disulfide bridge.
Data from Wright DJ (1987) The seed globulins. In: Hudson BJF (ed.) *Developments in Food Proteins*, vol. 5, pp. 81–157. London: Elsevier Applied Science.

unfolded form (ψ_i, $\phi_i \neq 180°$). It can be shown with models that ψ_i and ϕ_i, at a permissible minimum distance between atoms can assume only particular angles. It has been shown for many proteins empirically that they have values of ψ_i, ϕ_i pairs within the permissible range. When a multitude of equal ψ_i, ϕ_i pairs occurs consecutively in a peptide chain, the chain acquires regular repeating structural elements (secondary structure). Examples of some important secondary structures are shown in Figures 1–3.

β-Sheet Three regular structural elements (pleated-sheet structures) have values in the range of $\phi = -120°$ and $\psi = +120°$. The peptide chain is always lightly folded on the C_α atom, thus the R side chains extend perpendicularly to the extension axis of the chain, i.e., the side chains change their projections alternately from $+z$ to $-z$. Such a pleated structure is stabilized when more chains interact along the x axis by hydrogen bonding, thus providing the cross-linking required for stability. When adjacent chains run in the same direction, the peptide chains are parallel. This provides a stabilized, planar, parallel sheet structure (Figure 1). When the chains run in opposite directions, a planar, antiparallel sheet structure is stabilized (Figure 2). The lower free-energy, twisted sheet structures, in which the main axes of the neighboring chains are arranged at an angle of 25°, are more common than planar sheet structures.

The β-structures can also be regarded as special helical structures with a continuation of two residues per turn. With proline, the formation of a β-structure is not possible.

Helical structures There are three regular structural elements in the range $\phi = -60°$ and $\psi = -60°$, in which the peptide chains are coiled like a threaded screw. These structures are stabilized by intrachain hydrogen bridges which extend almost parallel to the helix axis, cross-linking the CO and NH groups, specifically the CO group of amino acid residue i with the NH group of residue $i+3$ (3_{10}-helix), $i+4$ (α-helix) or $i+5$ (π-helix). The α-helix, and for polypeptides from L-amino acids exclusively the right-handed α-helix, is the most common (Figure 3). The 3_{10}-helix is only observed at the ends of α-helices, whereas the existence of a π-helix is hypothetical. A helix is characterized by the angles ϕ and ψ, or by the

Table 2 Amino acid sequences of 7S globulins from legumes

β-Conglycinin from soya bean, α subunit

```
          5         0         5         0         5         0         5         0         5         0
  1   V E E E E E C E E G Q I  P R P R P Q H P E R E R Q Q H G E K E E D E G E Q P R P F P F P R P R Q P H
 51   Q E E E H E Q K E E H E W H R K E E K H G G K G S E E E Q D E R E H P R P H Q P H Q K E E E K H E W
101   Q H K Q E K H Q G K E S E E E E E D Q D E D E E Q D K E S Q E S E G S E S Q R E P R R H K N K N P F
151   H F N S K R F Q T L F K N Q Y G H V R V L Q R F N K R S Q Q L Q N L R D Y R I L E F N S K P N T L L
201   L P H H A D A D Y L I V I L N G T A I L T L V N N D D R D S Y N L Q S G D A L R V P A G T T F Y V V
251   N P D N D E N L R M I A G T T F Y V V N P D N D E N L R M I T L A I P V N K P E R F E S F F L S S T
301   Q A Q Q S Y L Q G F S K N I L E A S Y D T K F E E I N K V L F G R E E G Q Q Q G E E R L Q E S V I V
351   E I S K K Q I R E L S K H A K S S S R K T I S S E D K P F N L G S R D P I Y S N K L G K L F E I T Q
401   R N P Q L R D L D V F L S V V D M N E G A L F L P H F N S K A I V V L V I N E G E A N I E L V G I K
451   E Q Q Q R Q Q Q E E Q P L E V R K Y R A E L S E Q D I F V I P A G Y P V M V N A T S D L N F F A F G
501   I N A E N N Q R N F L A G S K D N V I S Q I P S Q V Q E L A F P R S A K D I E N L I K S Q S E S Y F
551   V D A Q P Q Q K E E G N K G R K G P L S S I L R A F Y
```

Vicilin from pea

```
          5         0         5         0         5         0         5         0         5         0
  1   R S D P Q N P F I F K S N K F Q T L F E N E N G H I R L L Q K F D Q R S K I F E N L Q N Y R L L E Y
 51   K S K P H T I F L P Q H T D A D Y I L V V L S G K A I L T V L K P D D R N S F N E R G D T I K L P A
101   G T I A Y L V N R D D N E E L R V L D L A I P V N R P G Q L Q S F L L S G N Q N Q Q N Y L S G F S K
151   N I L E A S F N T D Y E E I E K V L L E E H E K E T Q H R R S L K D K R Q Q S Q E E N V I V K L S R
201   G Q I E E L S K N A K S T S K K S V S S E S E P F N L R S R G P I Y S N E P G K F F E I T P E K N P
251   Q L Q D L D I F V N S V E I K E G S L L L P H Y N S R A I V I V T V N E G K G D F E L V G Q R N E N
301   Q Q E Q R K E D D E E E E Q G E E E I N K Q V Q N Y K A K L S S G D V F V I P A G H P V A L K A S S
351   N L D L L G F G I N A E N N Q R N F L A G D E D N V I S Q V Q R P V K E L A F P G S A Q E V D R I L
401   E N Q K Q S H F A D A Q P Q Q R E R G S R E T R D R L S S V
```

Phaseolin from garden bean, β subunit

```
          5         0         5         0         5         0         5         0         5         0
  1   A T S L R E E E E S Q D N P F Y F N S D N S W N T L F K N Q Y G H I R V L Q R F D Q Q S K R L Q N L
 51   E D Y R L V E F R S K P E T L L L P Q Q A D A E L L L V V R S G S A I L V L V K P D D R R E Y F F L
101   T Q G D N P I F S D N Q K I P A G T I F Y L V N P D P K E D L R I I Q L A M P V N N P Q I H E F F L
151   S S T E A Q Q S Y L Q E F S K H I L E A S F N S K F E E I N R V L F E E E G Q Q E E G Q Q E G V I V
201   N I D S E Q I E E L S K H A K S S S R K S H S K Q D N T I G N E F G N L T E R T D N S L N V L I S S
251   I E M K E G A L F V P H Y Y S K A I V I L V V N E G E A H V E L V G P K G N K E T L E F E S Y R A E
301   L S K D D V F V I P A A Y P V A I K A T S N V N F T G F G I N A N N N N R N L L A G K T D N V I S S
351   I G R A L D G K D V L G L T F S G S G E E V M K L I N K Q S G S Y F V D G H H H Q Q E Q Q K G S H Q
401   Q E Q Q K G R K G A F V Y
```

Data from Wright DJ (1988) The seed globulins – part II. In: Hudson BJF (ed.) *Developments in Food Proteins*, vol. 6, pp. 119–178. London: Elsevier Applied Science, and references therein.

parameters derived from these angles: n, the number of amino acid residues per turn; d, the rise along the main axis per amino acid residue; and r, the radius of the helix. Thus, the equation for the pitch, p, is $p = n \cdot d$.

Reverse turns An important conformational feature of globular proteins is the reverse turns, β-turns, or β-bends. They occur at 'hairpin' corners, where the peptide chain changes direction abruptly. Such corners involve four amino acids residues, among them frequently proline. Glycine is favored in the third position of β-bends on the basis of energy considerations. Different types of β-turns are known, for which different amino acids are allowed.

Super secondary structures Analysis of known structures has demonstrated that regular elements can exist in combined forms. Examples are the coiled-coil α-helix, chain segments with antiparallel β-structures (β-meander structure) and combinations of α-helix and β-structure (e.g., βαβα).

Tertiary and Quaternary Structure

Proteins can be divided into two large groups on the basis of conformation: (1) fibrillar (fibrous) or scleroproteins, and (2) folded or globular proteins.

Fibrous proteins The entire peptide chain is packed or arranged within a single regular structure for a variety of fibrous proteins. Examples are wool keratin (α-helix), silk fibroin (β-sheet), and collagen (a triple helix). Stabilization of these structures is achieved by intermolecular binding (electrostatic interaction and disulfide linkages, but primarily hydrogen bonds and hydrophobic interactions).

Table 3 Amino acid sequences of gluten proteins from wheat

HMW subunit of glutenin (x-type), 1Bx7 (Cheyenne)

```
          5         0         5         0         5         0         5         0         5         0
  1 E G E A S G Q L Q C E H E L E A C Q Q V V D Q Q L R D V S P G C R P I T V S P G T R Q Y E Q Q P V V
 51 P S K A G S F Y P S E T T P S Q Q L Q Q M I F W G I P A L L R R Y Y P S V T S S Q Q G S Y Y P G Q A
101 S P Q Q S G Q G Q Q P G Q E Q Q P G Q G Q Q H Q Q P G Q R Q Q G Y Y P T S P Q Q P G Q G Q Q L G Q G
151 Q P G Y Y P T S Q Q P G Q K Q Q A G Q G Q Q S G Q G Q Q G Y Y P T S P Q Q S G Q G Q Q P G Q G Q P G
201 Y Y P T S P Q Q S G Q W Q Q P G Q G Q Q P G Q G Q Q S G Q G Q Q G Q Q P G Q G Q R P G Q G Q Q G Y Y
251 P I S P Q Q P G Q G Q Q S G Q G Q P G Y Y P T S L R Q P G Q W Q Q P G Q G Q Q P G Q G Q Q G Q Q P G
301 Q G Q Q S G Q G Q Q G Y Y P T S L Q Q P G Q G Q Q L G Q G Q P G Y Y P T S Q Q S E Q G Q Q P G Q G K
351 Q P G Q G Q Q G Y Y P T S P Q Q S G Q G Q Q L G Q G Q P G Y Y P T S P Q Q S G Q G Q Q S G Q G Q Q G
401 Y Y P T S P Q Q S G Q G Q Q P G Q G Q S G Y F P T S R Q Q S G Q G Q Q P G Q G Q Q S G Q G Q Q G Q Q
451 P G Q G Q Q A Y Y P T S S Q Q S R Q R Q Q A G Q W Q R P G Q G Q P G Y Y P T S P Q Q P G Q E Q Q S G
501 Q A Q Q S G Q W Q L V Y Y P T S P Q Q P G Q L Q Q P A Q G Q Q P A Q G Q Q S A Q E Q Q P G Q A Q Q S
551 G Q W Q W Q L V Y Y P T S P Q Q P G Q L Q Q P A Q G Q Q G Y Y P T S P Q Q S G Q G Q Q G Y Y P T S P
601 Q Q S G Q G Q Q G Y Y P T S P Q Q S G Q G Q Q P G Q G Q Q P R Q G Q Q G Y Y P I S P Q Q S G Q G Q Q
651 P G Q G Q Q G Y Y P T S P Q Q S G Q G Q Q P G H E Q Q P G Q W L Q P G Q G Q Q G Y Y P T S S Q Q S G
701 Q G H Q S G Q G Q Q G Y Y P T S L W Q P G Q G Q Q G Y A S P Y H V S A E Y Q A A R L K V A K A Q Q L
751 A A Q L P A M C R L E G S D A L S T R Q
```

HMW subunit of glutenin (y-type), 1Dy10 (Cheyenne)

```
          5         0         5         0         5         0         5         0         5         0
  1 E G E A S R Q L Q C E R E L Q E S S L E A C R Q V V D Q Q L A G R L P W S T G L Q M R C C Q Q L R D
 51 V S A K C R S V A V S Q V A R Q Y E Q T V V P P K G G S F Y P G E T T P L Q Q L Q Q G I F W G T S S
101 Q T V Q G Y Y P G V T S P R Q G S Y Y P G Q A S P Q Q P G Q G Q Q P G K W Q E P G Q G Q Q W Y Y P T
151 S L Q Q P G Q G Q Q I G K G Q Q G Y Y P T S L Q Q P G Q G Q Q G Y Y P T S L Q H T G Q R Q Q P V Q G
201 Q Q P E Q G Q Q P G Q W Q Q G Y Y P T S P Q Q L G Q G Q Q P R Q W Q Q S G Q G Q Q G H Y P T S L Q Q
251 P G Q G Q Q G H Y L A S Q Q Q P G Q G Q Q G H Y P A S Q Q Q P G Q G Q Q G H Y P A S Q Q Q P G Q G Q
301 Q G H Y P A S Q Q E P G Q G Q Q G Q I P A S Q Q Q P G Q G Q Q G H Y P A S L Q Q P G Q G Q Q G H Y P
351 T S L Q Q L G Q G Q Q T G Q P G Q K Q Q P G Q G Q Q T G Q G Q Q P E Q E Q Q P G Q G Q Q G Y Y P T S
401 L Q Q P G Q G Q Q Q Q G Q G Q Q G Y Y P T S L Q Q P G Q G Q Q G H Y P A S L Q Q P G Q G Q P G Q R Q Q
451 P G Q G Q H P E Q G K Q P G Q G Q Q G Y Y P T S P Q Q P G Q G Q Q L G Q G Q Q G Y Y P T S P Q Q P G
501 Q G Q Q P G Q G Q Q Q G H C P T S P Q Q S G Q A Q Q P G Q G Q Q I G Q V Q Q P G Q G Q Q G Y Y P T S V
551 Q Q P G Q G Q Q S G Q G Q Q S G Q G H Q P G Q G Q Q S G Q E Q Q G Y D S P Y H V S A E Q Q A A S P M
601 V A K A Q Q P A T Q L P T V C R M E G G D A L S A S Q
```

LMW subunit of glutenin, clone LMWG-1D1 (Chinese spring)

```
          5         0         5         0         5         0         5         0         5         0
  1 R C I P G L E R P W Q Q Q P L P P Q Q T F P Q Q P L F S Q Q Q Q Q L F P Q Q P S F S Q Q Q P P F W
 51 Q Q Q P P F S Q Q Q P I L P Q Q P P F S Q Q Q Q L V L P Q Q P P F S Q Q Q Q P V L P P Q Q S P F P Q
101 Q Q Q Q H Q Q L V Q Q Q I P V V Q P S I L Q Q L N P C K V F L Q Q Q C S P V A M P Q R L A R S Q M L
151 Q Q S S C H V M Q Q Q C C Q Q L P Q I P Q Q S R Y E A I R A I I Y S I I L Q E Q Q Q V Q G S I Q S Q
201 Q Q Q P Q Q L G Q C V S Q P Q Q Q S Q Q Q L G Q Q P Q Q Q Q L A Q G T F L Q P H Q I A Q L E V M T S
251 I A L R I L P T M C S V N V P L Y R T T T S V P F G V G T G V G A Y
```

α-Gliadin clone A 1235 (Cheyenne)

```
          5         0         5         0         5         0         5         0         5         0
  1 V R V P V P Q L Q P Q N P S Q Q Q P Q E Q V P L M Q Q Q Q Q F P G Q Q E Q F P P Q Q P Y P H Q Q P F
 51 P S Q Q P Y P Q P Q P F P P Q L P Y P Q T Q P F P P Q Q P Y P Q P Q P Q Y P Q P Q Q P I S Q Q Q A Q
101 Q Q Q Q Q Q T L Q Q I L Q Q Q L I P C R D V V L Q Q H N I A H A S S Q V L Q Q S S Y Q Q L Q Q L C
151 C Q Q L F Q I P E Q S R C Q A I H N V V H A I I L H H H Q Q Q Q Q P S S Q V S Y Q Q P Q E Q Y P S
201 G Q V S F Q S S Q Q N P Q A Q G S V Q P Q Q L P Q F Q E I R N L A L Q T L P A M C N V Y I P P Y C S
251 T T I A P F G I F G T N
```

γ-Gliadin clone gene A (Yamhill)

```
          5         0         5         0         5         0         5         0         5         0
  1 N M Q V D P S G Q V Q W P Q Q Q P V L L P Q Q P F S Q Q P Q Q T F P Q P Q Q T F P H Q P Q Q Q F P Q
 51 P Q Q P Q Q Q F L Q P Q Q P F P Q Q P Q Q P Y P Q Q P Q Q P F P Q T Q Q P Q Q L F P Q S Q Q P Q Q P
101 Y P Q Q P Q Q P F P Q T Q Q P Q Q Q F P Q S Q Q P Q P F P Q P Q Q P Q Q S F P Q Q Q P S F I Q P S L
151 Q Q Q L N P C K N L L L Q Q C R P V S L V S S L W S M I W P Q S A C Q V M R Q Q C C Q Q L A Q I P Q
201 Q L Q C A A I H S V V H S I S M Q E Q Q Q Q Q Q Q Q Q Q Q Q Q Q G M R I L L P L Y Q Q Q Q V G Q Q G
251 T L V Q G Q Q G I I Q P Q Q P A Q L E A I R S L V L Q T L P T M C N V Y V P P E C S I I K A P F A S I
301 V T G I G G Q
```

The high-molecular-weight (HMW) subunit (*x*-type) contains one intramolecular disulfide bond (Cys (10)-Cys (17)), whereas the remaining two cysteines (Cys (32) and Cys (758)) form intermolecular disulfide bonds probably with other HMW subunits. In HMW subunits (*y*-type), the two adjacent cysteines (Cys(43) and Cys(44)) form two intermolecular disulfide bonds with two adjacent cysteines of another *y*-type HMW subunit, Cys(513) forms an intermolecular disulfide bond with low-molecular-weight (LMW) subunits, whereas the linkage of the remaining four cysteines is still unknown (probably one intermolecular disulfide bond, Cys(10)-Cys(22)). The LMW subunit contains three intramolecular disulfide bonds (Cys(127)-Cys(162), Cys(135)-Cys(155) and Cys(163)-Cys(260)), whereas the remaining two cysteines (Cys(2) and Cys(210)) form intermolecular disulfide bonds with other LMW

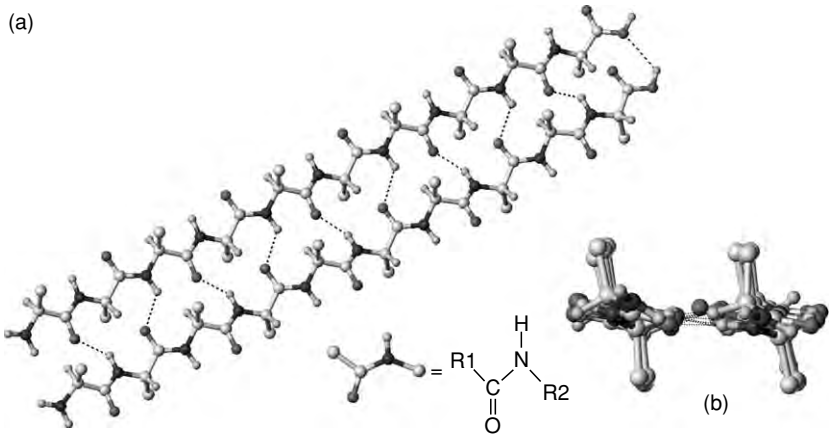

Figure 1 Parallel β-sheet ($\phi = -119°$, $\psi = +113°$, $n = 2.0$, $d = 0.32$ nm, $r = 0.11$ nm; see text). (a) Perspective view of two decapeptide chains consisting of 10 alanine residues each. The *N*-terminus of both chains is on the left; hydrogen atoms of the methyl groups have been omitted for clearness; hydrogen bridges are presented as broken lines. (b) View in the direction of the chain axis.

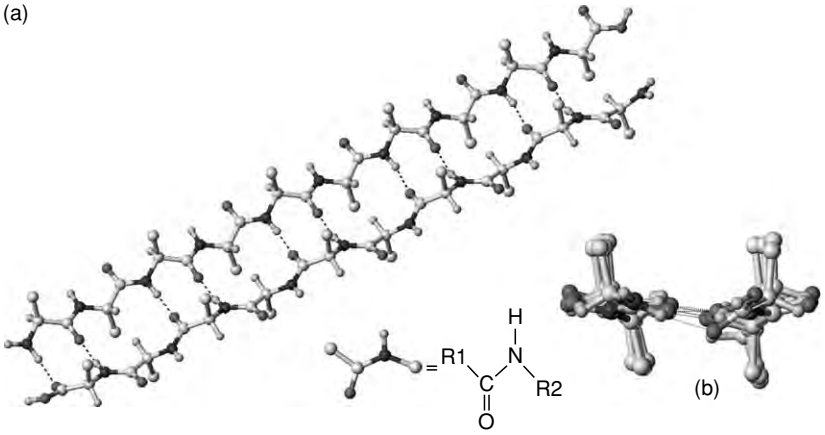

Figure 2 Antiparallel β-sheet ($\phi = -139°$, $\psi = +135°$, $n = 2.0$, $d = 0.34$ nm, $r = 0.09$ nm). (a), (b) Cf. **Figure 1**, but the *N*-terminus of the upper chain is on the left and that of the lower chain is on the right.

Globular proteins Regular structural elements are mixed with randomly extended chain segments (random-coiled structures) in globular proteins. The proportion of regular structural elements is highly variable: 20–30% in casein, 45% in lysozyme, and 75% in myoglobin. Five structural subgroups are known in this group of proteins: (1) α-helices occur only; (2) β-structures occur only; (3) α-helical and β-structural portions occur in separate segments on the peptide chain; (4) α-helices and β-structures alternate along the peptide chain; and (5) α-helices and β-structures do not exist. The process of peptide chain folding is not yet fully understood. It occurs spontaneously, probably arising from one center or from several centers of high stability in larger proteins. Folding of the peptide chain packs it densely, by formation of a large number of intramolecular noncovalent bonds. Data about the nature of the bonds involved are provided in Table 4.

subunits (head to head or head to tail), *y*-type HMW subunits or glutenin-bound γ-gliadins. In α-gliadins, three intramolecular disulfide bonds occur (Cys(120)-Cys(150), Cys(151)-Cys(241) and Cys(163)-Cys(249)). In γ-gliadins, four intramolecular disulfide bonds occur (Cys(157)-Cys(191), Cys(165)-Cys(184), Cys(192)-Cys(282) and Cys(204)-Cys(290)).

Data from Müller S and Wieser H (1997) The location of disulphide bonds in monomeric α-type gliadins. *Journal of Cereal Science* 26: 169–176, Shewry PR and Tatham AS (1997) Disulphide bonds in wheat gluten proteins. *Journal of Cereal Science* 25: 207–223 and references therein.

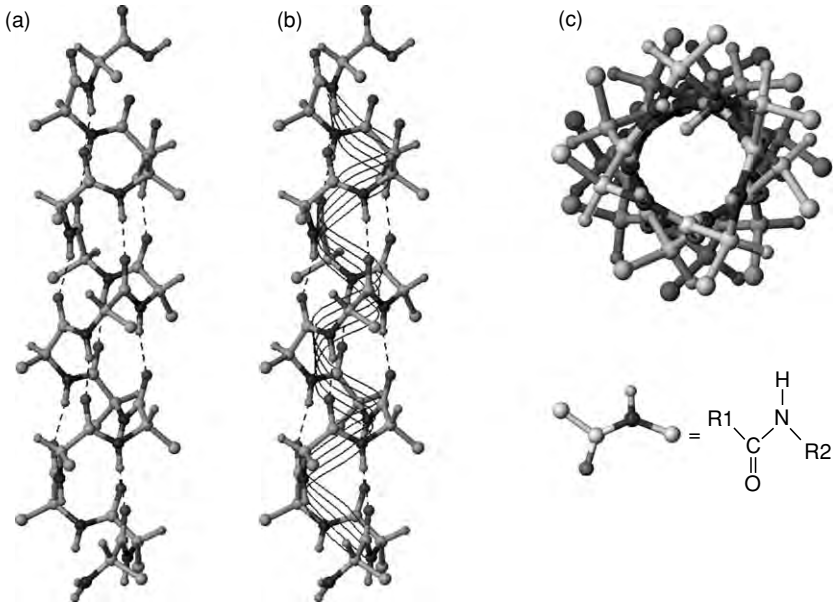

Figure 3 Right-handed α-helix ($\phi = -57°$, $\psi = -47°$, $n = 3.3$, $d = 0.15$ nm, $r = 0.23$ nm). (a) Perspective view of a helical peptide chain consisting of 15 alanine residues; the *N*-terminus is at the bottom; in other respects it is the same as **Figure 1**. (b) Same representation as (a) with a lined band elucidating the right-handed helix running from bottom to top. (c) View in the direction of the helix axis.

Table 4 Bond-types in proteins

Type	*Examples*	*Bond strength (kJ mol^{-1})*
Covalent bonds	-S-S-	−230
Electrostatic bonds	-COO^- . . . H_3N^+-	−21[a]
	> C=O . . . O=C <	+1[a,b]
Hydrogen bonds	-O-H . . . O <	−17
	> N-H . . . O=C <	−13
Hydrophobic bonds	-Ala . . . Ala-	−3
	-Val . . . Val-	−8
	-Leu . . . Leu-	−9
	-Phe . . . Phe-	−13
	-Trp . . . Trp-	−19

[a]For a dielectric constant of the surrounding medium $\varepsilon = 4$; [b]weak repulsion due to the negative partial charge of the oxygens in a polarized C–O double bond.
Data from Belitz H-D and Grosch W (1999) *Food Chemistry*, 2nd edn. Berlin: Springer–Verlag.

The hydrogen bonds formed between main chains, main and side chains, and side-side chains are of particular importance for folding. The portion of polar groups involved in hydrogen bond build-up in proteins of > 8.9 kDa appears to be fairly constant at about 50%. The hydrophobic interaction of the nonpolar regions of the peptide chains also plays an important role in protein folding. These interactions are responsible for the fact that nonpolar groups are folded to a great extent towards the interior of the protein globule. The surface areas accessible to water molecules have been calculated for both unfolded and native folded forms for a number of monomeric proteins with known conformations. The proportion of the accessible surface of a stretched state which tends to be fixed in the interior of the globule as a result of folding is a simple linear function of the molecular weight.

Proteins with disulfide bonds fold at a significantly slower rate than those without disulfide bonds. Folding is not limited by the reaction rate of disulfide formation. Therefore the folding process of disulfide-containing proteins seems to proceed in a different way. The reverse process, protein unfolding, is very much slowed down by the presence of disulfide bridges which generally impart great stability to globular proteins. This stability is particularly effective against denaturation. An example is the Bowman–Birk inhibitor from soya beans, which inhibits the activity of trypsin and chymotrypsin. Its tertiary structure is stabilized by seven disulfide bridges. The reactive sites of inhibition are Lys(16)-Ser(17) and Leu(43)-Ser(44), i.e., both sites are located in relatively small rings, each of which consists of nine amino acid residues held in ring form by a disulfide bridge. The thermal stability of this inhibitor is high (**Figure 4**).

Quaternary structures In addition to the free energy gain by folding of a single peptide chain, association of more than one peptide chain (subunit) can provide

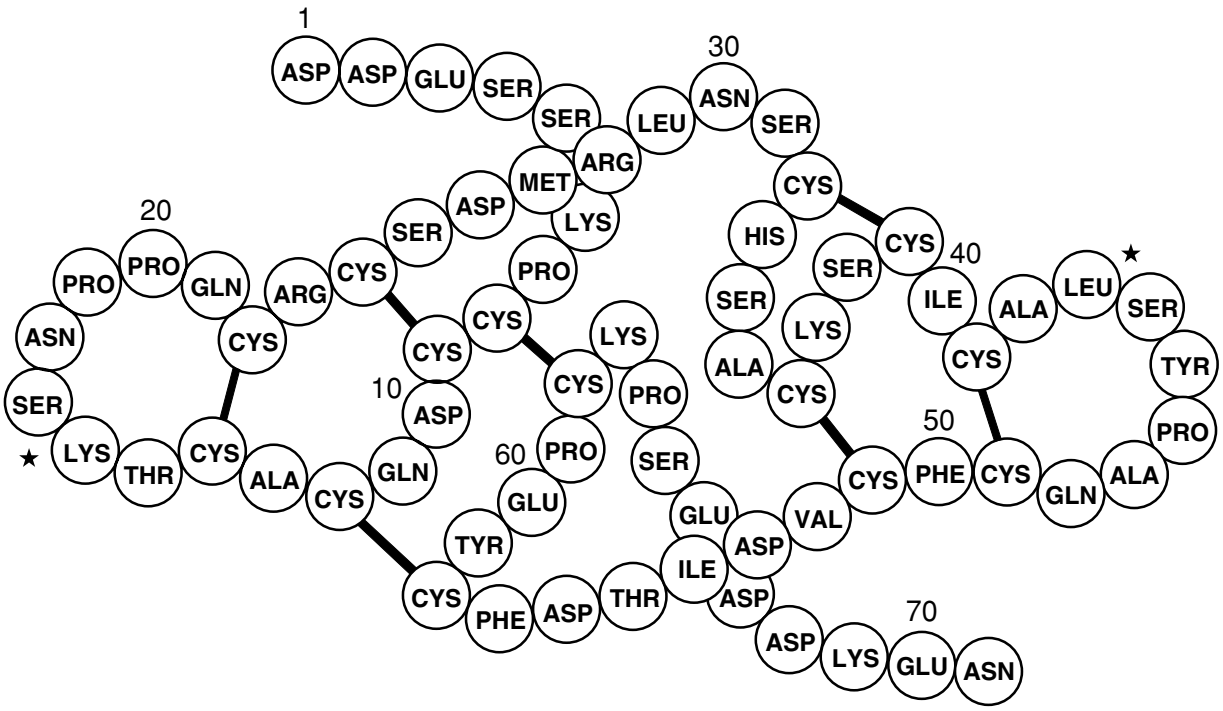

Figure 4 Bowman–Birk proteinase inhibitor from soya beans. Solid black lines, disulfide bridges between half-cystine residues; asterisks, trypsin (Lys-Ser) and chymotrypsin (Leu-Ser) reactive sites. Adapted from Ikenaka T, Odani S, and Koide T (1974) Chemical structure and inhibitory activities of soybean proteinase inhibitors. In: Fritz H, Tschesche H, Greene LJ and Truscheit E (eds) *Bayer-Symposium V Proteinase Inhibitors*, pp. 325–343. Berlin: Springer-Verlag, with permission.

further gains in free energy. For example, hemoglobin (four associated peptide chains) $\Delta G^\circ = -46\,\mathrm{kJ\,mol^{-1}}$ and the trypsin–trypsin inhibitor complex (association of two peptide chains) $\Delta G^\circ = -75.2\,\mathrm{kJ\,mol^{-1}}$. In principle, such associations correspond to the folding of a larger peptide chain around several structural domains without covalently binding the subunits.

Examples for protein conformations are given in Figure 5, which shows the sweet proteins monellin and thaumatin.

Denaturation

The term 'denaturation' denotes a reversible or irreversible change of native conformation (tertiary structure) without cleavage of covalent bonds (except for disulfide bridges). Denaturation is possible with any treatment that cleaves hydrogen bridges, or ionic or hydrophobic bonds. This can be accomplished by changing the temperature, adjusting the pH, increasing the interfacial area, or adding organic solvents, salts, urea, guanidine hydrochloride, or detergents such as sodium dodecyl sulfate. Denaturation is generally reversible when the peptide chain is stabilized in its unfolded state by the denaturing agent and the native conformation can be reestablished after removal of the agent. Irreversible denaturation occurs when the unfolded peptide chain is stabilized by interaction with other chains (as occurs for instance with egg proteins during boiling). During unfolding, reactive groups, such as thiol groups, that were buried or blocked may be exposed. Their participation in the formation of disulfide bonds may also cause an irreversible denaturation. Denaturation of biologically active proteins is usually associated with loss of activity. The fact that denatured proteins are more readily digested by proteolytic enzymes is also of interest.

Physical Properties

Dissociation

Proteins, like amino acids, are amphoteric. Depending on pH, they can exist as polyvalent cations, anions, or zwitterions. Since α-carboxyl and α-amino groups are linked together by peptide bonds, the uptake or release of protons is limited to free terminal groups, and mostly to side chains. In contrast to free amino acids, the pK_a values fluctuate greatly for proteins since the dissociation is influenced by neighboring groups in the macromolecule. For example, in lysozyme the γ-carboxyl group of Glu(35) has a pK_a of 6–6.5,

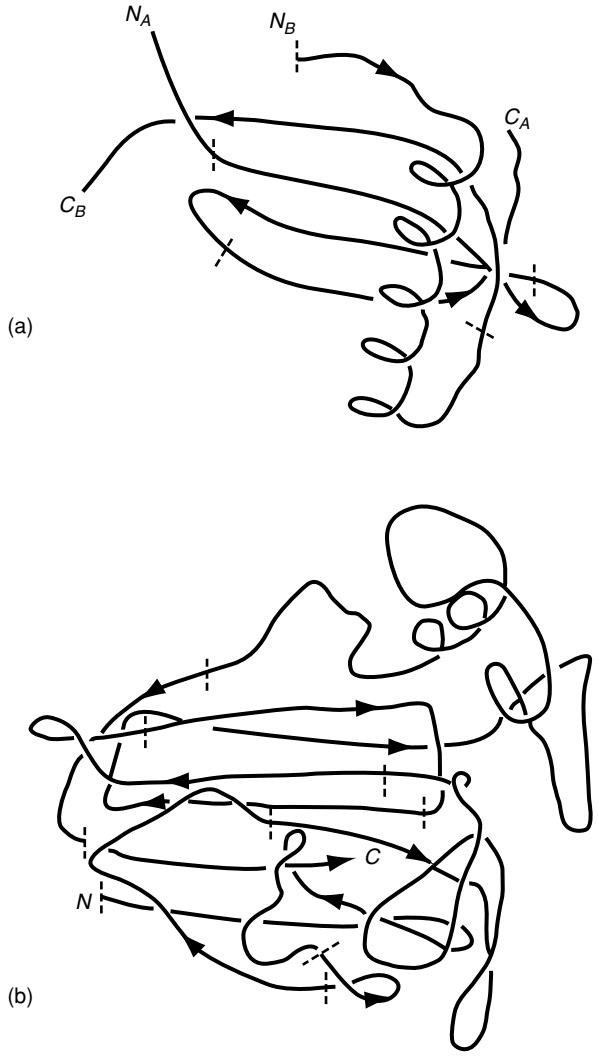

Figure 5 Conformation of the peptide chains of the sweet proteins (a) monellin and (b) thaumatin. β-sheet, |→; α-helix, ෆ; β-turn, ⊃; N_A, N_B, C_A, C_B denote the *N*- and *C*-termini of the *A* and *B* chains. Adapted from Ogata C, Hatada M, Tomlinson G, Shin WC and Kim S-H (1987) Crystal structure of the intensely sweet protein monellin. *Nature* 328: 739–742, with permission.

while the pK_a of the β-carboxyl group of Asp(66) is 1.5–2, of Asp(52) is 3–4.6, and of Asp(101) is 4.2–4.7. The total charge of a protein, which is the absolute sum of all positive and negative charges, is differentiated from the so-called net charge which, depending on the pH, may be positive, zero, or negative. By definition the net charge is zero and the total charge in general is maximal at the isoelectric point. Lowering or raising the pH tends to increase the net charge toward its maximum, whereas the total charge becomes less than at the isolectric point.

Since proteins interact not only with protons but also with other ions, there is a further differentiation between an isoionic and an isoelectric point. The isoionic point is defined as the pH of a protein solution at infinite dilution, with no other ions present except for H^+ and OH^-. Such a protein solution can be acquired by extensive dialysis (or, better, electrodialysis) against distilled water. The isoionic point is constant for a given substance whereas the isoelectric point is variable, depending on the ions present and their concentration. In the presence of salts, e.g., when binding of anions is stronger than that of cations, the isoelectric point is lower than the isoionic point. The reverse is true when cation binding is dominant. In most cases the shift in pH is consistently positive, i.e., the protein binds more anions than cations. At its isoelectric point a protein is the least soluble and the most likely to precipitate (isoelectric precipitation) and it is at its maximal crystallization capacity. The viscosity of solubilized proteins and the swelling power of insoluble proteins are at a minimum at the isoelectric point.

Optical Activity

The optical activity of proteins is due not only to asymmetry of amino acids but also to the chirality resulting from the arrangement of the peptide chain. Information on the conformation of proteins can be obtained from recording the optical rotatory dispersion (ORD) or the circular dichroism (CD), especially in the range of peptide bond absorption wavelengths (190–200 nm). The Cotton effect occurs in this range and reveals quantitative information on secondary structure. An α-helix or a β-structure gives a negative Cotton effect, with absorption maxima at 199 and 205 nm, respectively, whereas a random-coiled conformation shifts the maximum to shorter wavelengths, i.e., results in a positive Cotton effect.

Solubility, Hydration, and Swelling Power

Protein solubility is variable and is influenced by the number of polar and apolar groups and their arrangement along the molecule. Generally, proteins are soluble only in strongly polar solvents such as water, glycerol, formamide, dimethylformamide, or formic acid. In a less polar solvent such as ethanol, few proteins have appreciable solubility (e.g., prolamins). The solubility in water is dependent on pH and on salt concentration. At low ionic strength, the solubility rises with increase in ionic strength and the solubility minimum (isoelectric point) is shifted to a somewhat lower pH. This shift is due to preferential binding of anions to the protein. As a rule, neutral salts have a twofold effect on protein solubility. At low concentrations they increase the solubility ('salting-in' effect) by suppressing the electrostatic protein–protein interaction (binding forces).

The logarithm of the solubility (S) is proportional to the ionic strength (μ) at low concentrations: $\log_{10}S = k{\cdot}\mu$.

Protein solubility is decreased ('salting-out' effect) at higher salt concentrations due to the ion hydration tendency of the salts. The following relationship applies (S_0, solubility at $\mu = 0$; K, salting-out constant): $\log_{10}S = \log_{10}S_0 - K{\cdot}\mu$. Cations and anions in the presence of the same counter ion can be arranged in the following orders (Hofmeister series) based on their salting-out effects: $K^+ > Rb^+ > Na^+ > Cs^+ > Li^+ > NH_4^+$; SO_4^{2-} > citrate^{2-} > tartrate^{2-} > acetate$^-$ > $Cl^- > NO_3^- > Br^- > I^- > SCN^-$. Multivalent anions are more effective than monovalent anions, whereas divalent cations are less effective than monovalent cations.

Since proteins are polar substances, they are hydrated in water. The degree of hydration (grams of water of hydration per gram protein) is variable. It is 0.22 for ovalbumin (in ammonium sulfate), 0.06 for edestin (in ammonium sulfate), 0.8 for β-lactoglobulin, and 0.3 for hemoglobin.

The swelling of insoluble proteins corresponds to the hydration of soluble proteins in that insertion of water between the peptide chains results in an increase in volume and other changes in the physical properties of the protein. The amount of water taken up during swelling can exceed the dry weight of the protein by several times. For example, muscle tissue contains 3.5–3.6 g of water per gram of protein (dry matter).

Chemical Reactions

The chemical modification of proteins is of importance for a number of reasons. It provides derivatives suitable for sequence analysis, identifies the reactive groups in catalytically active sites of an enzyme, enables the binding of protein to a carrier (protein immobilization), and provides changes in protein properties which are important in food processing. In contrast to free amino acids, except for the relatively small number of functional groups of the terminal amino acids, only the functional groups in protein side chains are available for chemical reactions.

Lysine Residue

Reactions involving the lysine residue can be divided into several groups: (1) reactions leading to a positively charged derivative; (2) reactions eliminating the positive charge; (3) derivatizations introducing a negative charge; and (4) reversible reactions. The last are of particular importance.

Reactions which retain the positive charge Alkylation of the free amino group of lysines with aldehydes and ketones is possible, with a simultaneous reduction step. A dimethyl derivative (Prot-N$(CH_3)_2$) can be obtained with formaldehyde.

Guanidination can be accomplished by using *O*-methylisourea as a reactant. α-Amino groups react at a much slower rate than ε-amino groups. This reaction is used analytically to assess the amount of biologically available ε-amino groups.

Derivatization with imido esters to amidines is also possible. The reactant is readily accessible from the corresponding nitrile. Proteins can be cross-linked with the use of a bifunctional imido ester.

Treatment of the amino acid residue with amino acid carboxyanhydrides yields a polycondensation reaction product, made up of the protein as main chain and peptide side chains. The length of the side chains depends on reaction conditions. The carboxyanhydrides are readily accessible through interaction of the amino acid with phosgene.

Reactions resulting in a loss of positive charge Acetic anhydride reacts with lysine, cysteine, histidine, serine, threonine, and tyrosine residues. Subsequent treatment of the protein with hydroxylamine leaves only the acetylated amino groups intact. Acylation is possible with *N*-hydroxysuccinimidyl acetate, succinimidyl 3-(*p*-hydroxphenyl)propionate (after radioiodination for radioactive labeling, Bolton–Hunter reagent) or *S*-ethylthiotrifluoroacetate. A photocleavable amine-protecting group can be introduced by reaction with 6-nitroveratryl chloroformate.

Sulfonylation occurs with dansyl chloride (5-(dimethylamino)-naphthalene-1-sulfonyl chloride), which also reacts with cysteine, histidine, and tyrosine. Carbamoylation with cyanate attacks α- and ε-amino groups as well as cysteine and tyrosine residues. However, their derivatization is reversible under alkaline conditions. Thiocarbamoylation with fluorescein isothiocyanate (FITC) or tetramethylrhodamine isothiocyanate (TRITC) is used for fluorescent labeling of proteins.

Arylation is possible with DNFB (2,4-dinitrofluorobenzene) and 2,4,6-trinitrobenzenesufonic acid (TNBS). DNFB also reacts with cysteine, histidine, and tyrosine. 7-Chloro-4-nitrobenz-2-oxa-1,3-diazole, which also reacts with cysteine, and 4-fluoro-3-nitrobenzenesulfonic acid, a reactant which has good solubility in water, are also of interest for derivatization of proteins.

Deamination can be accomplished with nitrous acid. This reaction involves α-and ε-amino groups as well as tryptophan, tyrosine, cysteine, and methionine residues.

Reactions resulting in a negative charge Acylation with dicarboxylic acid anhydrides, e.g., succinic anhydride, introduces additional carboxyl groups into the protein. Introduction of a fluorescent acid group is possible by interaction of the protein with pyridoxal 5-phosphate, followed by reduction of an intermediate Schiff base, or with fluorescamine.

Reversible reactions *N*-Maleyl derivatives of proteins are obtained at alkaline pH by reaction with maleic anhydride. The acylated product is cleaved at pH < 5, regenerating the protein. The half-life (τ) of ε-*N*-maleyl lysine is 11 h at pH 3.5 and 37 °C. More rapid cleavage is observed with the 2-methylmaleyl derivative (τ < 3 min at pH 3.5 and 20 °C) and the 2,2,3,3-tetrafluorosuccinyl derivative (τ very low at pH 9.5 and 0 °C). Cysteine binds maleic anhydride through an addition reaction. The *S*-succinyl derivative is quite stable. This side reaction is, however, avoided when protein derivatization is carried out with *exo-cis*-3,6-endoxo-1,2,3,6-tetrahydrophthalic acid anhydride.

Acetoacetyl derivatives are obtained with diketene. This type of reaction also occurs with cysteine and tyrosine residues. The acyl group is readily split from tyrosine at pH 9.5. Complete release of protein from its derivatized form is possible by treatment with phenylhydrazine or hydroxylamine at pH 7.

Arginine Residue

The arginine residue of proteins reacts with α- or β-dicarbonyl compounds, e.g., with glyoxal, phenylglyoxal hydrate, 2,3-butanedione (diacetyl) or nitromalondialdehyde, to form cyclic derivatives. The nitropyrimidine resulting with nitromalondialdehyde absorbs at 335 nm. The arginyl bond of this derivative is not cleaved by trypsin but it is cleaved in its tetrahydro form, obtained by reduction with sodium borohydride. In the reaction with benzil, an iminooxo-imidazolidine derivative is obtained after a benzilic acid rearrangement. Reaction of the arginine residue with 1,2-cyclohexanedione is highly selective and proceeds under mild conditions. Regeneration of the arginine residue is again possible with hydroxylamine.

Glutamic and Aspartic Acid Residues

These amino acid residues are usually esterified with methanolic hydrochloric acid. There can be side reactions, such as methanolysis of amide derivatives or *N,O*-acyl migration in serine or threonine residues. Diazoacetamide reacts with a carboxyl group and also with the cysteine residue to carboxamidomethyl derivatives.

Amino acid esters or other similar nucleophilic compounds can be attached to a carboxyl group of a protein with the help of a carbodiimide. Amidation is also possible by activating the carboxyl group with an isoxazolium salt (Woodward reagent) to an enolester and its conversion with an amine.

Cystine Residue

Reductive cleavage of cystine occurs with sodium borohydride and with thiols. Thiols used to reduce cystine to cysteine include 2-mercaptoethanol and, more effectively, dithiothreitol (DTT, Cleland's reagent) or dithioerythritol (DTE). More recently, bis(2-mercaptoethyl)sulfone and tris(2-carboxyethyl)phosphine have been introduced, which are both water-soluble and very effective, and also tri-*n*-butylphosphine, which is less water-soluble. The phosphines have the advantage that the excess used for reduction does not react with thiol reagents used for subsequent thiol modification.

Cleavage of cystine is also possible by a nucleophilic attack. The nucleophilic reactivity of the reagents decreases in the following series: arsenite and phosphite > alkanethiol > aminoalkanethiol > thiophenol and cyanide > sulfite > OH^- > *p*-nitrophenol > thiosulfate > thiocyanate. Complete cleavage with sulfite requires that oxidative agents (e.g., Cu^{2+}) are present and that the pH is higher than 7. The resultant *S*-sulfo derivative is quite stable in neutral and acidic media and is fairly soluble in water. The *S*-sulfo group can be eliminated with an excess of thiol reagent. Cleavage of cystine residues with cyanides (nitriles) is of interest since the thiocyanate formed in the reaction is cyclized into a 2-iminothiazolidine under cleavage of the *N*-acyl bond. This reaction can be utilized for the selective cleavage of peptide chains. Initially, all the disulfide bridges are reduced with DTT, and are then converted to mixed disulfides through reaction with 5,5′-dithio-bis(2-nitrobenzoic acid). These mixed disulfides are then cleaved by cyanide at pH 7.

Electrophilic cleavage occurs with Ag^+ and Hg^+ or Hg^{2+}. Electrophilic cleavage with H^+ is possible only in strong acids (e.g., 10 mol l^{-1} hydrochloric acid). The sulfenium cation which is formed can catalyze a disulfide exchange reaction. In neutral and alkaline solutions a disulfide exchange reaction is catalyzed by the thiolate anion.

Cysteine Residue

A number of alkylating agents, e.g., iodoacetic acid, iodoacetamide, ethylenimine, and 4-vinylpyridine, yield derivatives which are stable under the conditions for acid hydrolysis of protein. The reaction with ethylenimine gives an *S*-(2-aminoethyl) derivative

and, hence, an additional linkage position in the protein for hydrolysis by trypsin. Iodoacetic acid, depending on the pH, can react with cysteine, methionine, lysine, and histidine residues. Maleic anhydride, 2-nitrobenzyl bromide, 2-nitrohydroxybenzyl bromide, and methyl *p*-nitrobenzenesulfonate are also alkylating agents. A number of reagents make it possible to measure thiol group content spectrophotometrically (azobenzene-2-sulfenyl bromide, 5,5′-dithio-bis(2-nitrobenzoic acid) (DTNB, Ellman's reagent), *p*-hydroxymercuribenzoate, *N*-ethylmaleimide, *N*-phenylmaleimide). Especially suitable for the specific and very sensitive labeling of cysteine-containing peptides prior to isolation are *N*-[4-(4′-dimethylaminoazobenzene)]maleimide (DABM), N-(7-dimethylamino-4-methyl-3-coumarinyl)maleimide (DACM), *N*-(1-pyrenyl)maleimide, and 4-aminosulfonyl-7-fluoro-2,1,3-benzoxdiazol (ABDF).

Cysteine is readily converted to the corresponding disulfide, cystine, even under mild oxidative conditions, such as treatment with iodine or potassium hexacyanoferrate(III). Stronger oxidation of cysteine, and also of cystine, e.g., with performic acid, yields the corresponding sulfonic acid, cysteic acid.

Methionine Residue

Methionine residues are oxidized to methionine sulfoxide with hydrogen peroxide. The sulfoxide can be reduced, regenerating methionine, using an excess of thiol reagent. With performic acid, methionine sulfone is formed.

α-Halogen carboxylic acids, β-propiolactone, and alkyl halides convert methionine into sulfonium derivatives, from which methionine can be regenerated in an alkaline medium with an excess of thiol reagent. Reaction with cyanogen bromide, which splits the peptide bond on the carboxyl side of the methionine molecule, is used for selective cleavage of proteins.

Histidine Residue

Diethyl pyrocarbonate reacts with histidine to form *N*-(ethoxycarbonyl)histidine. With iodoacetamide, *N*-1-(carboxamidomethyl)-, *N*-3-(carboxamidomethyl)-, or *N*-1,*N*-3-di(carboxamidomethyl)histidine are formed.

Selective modification of histidine residues present on active sites of serine proteinases is possible. Substrate analogs such as halogenated methyl ketones inactivate such enzymes (e.g., 1-chloro-3-tosylamido-7-amino-2-heptanone inactivates trypsin and 1-chloro-3-tosylamido-4-phenyl-2-butanone inactivates chymotrypsin) by *N*-alkylation of the histidine residue.

Tryptophan Residue

N-Bromosuccinimide oxidizes the tryptophan side chain and also tyrosine, histidine, and cysteine. The reaction is used for the selective cleavage of peptide chains at the carboxyl group of tryptophan residues and the spectrophotometric determination of tryptophan. Other oxidative cleaving reagents are *o*-iodosobenzoic acid and 3-bromo-3-methyl-2-(2-nitrophenylmercapto)-3H-indole. Selective modification of histidine is possible with 2-hydroxy-5-nitrobenzyl bromide (Koshland reagent I) and 2-nitrophenylsulfenyl chloride.

Tyrosine Residue

Selective acylation of tyrosine can occur with 1-acetylimidazole as a reagent. Diazotized *p*-arsanilic acid reacts with tyrosine (*ortho* substitution) and with histidine, lysine, tryptophan, and arginine. Tetranitromethane introduces a nitro group into the *ortho* position. Radioactive labeling of tyrosine is possible with $Na^{125}I$ and chloramine T (*ortho* mono- and di-iodo derivatives).

Bifunctional Reagents

Bifunctional reagents enable intra-and intermolecular cross-linking of proteins. Examples are bifunctional imidoester, maleimides, fluoronitrobenzene, and isocyanate derivatives.

See also: **Amino Acids**: Properties and Occurrence; Determination; **Chromatography**: Principles; High-performance Liquid Chromatography; **Enzymes**: Functions and Characteristics; **Mass Spectrometry**: Principles and Instrumentation; Applications; **Peptides**; **Protein**: Food Sources; Functional Properties; Interactions and Reactions Involved in Food Processing; **Spectroscopy**: Nuclear Magnetic Resonance

Further Reading

Belitz H-D and Grosch W (1999) *Food Chemistry*, 2nd edn. Berlin: Springer-Verlag.

Davies JS (ed.) (1993–99) *Amino-acids, Peptides, and Proteins*, vols 24–30. Cambridge: Royal Society of Chemistry.

Deutscher MP (ed.) (1990) *Methods in Enzymology*, vol. 182. *Guide to Protein Purification*. San Diego: Academic Press.

Gray WR (1993) Dislfide structures of highly bridged peptides: a new strategy for analysis. *Protein Science* 2: 1732–1748.

Ikenaka T, Odani S and Koide T (1974) Chemical structure and inhibitory activities of soybean proteinase inhibitors. In: Fritz H, Tschesche H, Greene LJ and Truscheit E (eds) *Bayer-Symposium V 'Proteinase Inhibitors'*, pp. 325–343. Berlin: Springer-Verlag.

Jones JH (ed.) (1983–92) *Amino-acids, Peptides, and Proteins*, vols 14–23. London: Royal Society of Chemistry.
Kellner R, Lottspeich F and Meyer HE (1999) *Microcharacterization of Proteins*, 2nd edn. Weinheim: Wiley-VCH.
Lottspeich F and Zorbas H (eds) (1998) *Bioanalytik*. Heidelberg: Spektrum Akademischer Verlag.
Lundblad RL and Noyes CM (1984) *Chemical Reagents for Protein Modification*, vols I and II. Boca Raton: CRC Press.
Matsudaira PT (ed.) (1989) *A Practical Guide to Protein and Peptide Purification for Microsequencing*. San Diego: Academic Press.
Müller S and Wieser H (1997) The location of disulphide bonds in monomeric γ-type gliadins. *Journal of Cereal Science* 26: 169–176.
Ogata C, Hatada M, Tomlinson G, Shin WC and Kim S-H (1987) Crystal structure of the intensely sweet protein monellin. *Nature* 328: 739–742.
Shewry PR and Tatham AS (1997) Disulphide bonds in wheat gluten proteins. *Journal of Cereal Science* 25: 207–227.
Shimada K and Mitamura K (1994) Derivatization of thiol-containing compounds. *Journal of Chromatography B* 659: 227–241.
Singh R and Whitesides GM (1994) Reagents for rapid reduction of native disulfide bonds in proteins. *Bioorganic Chemistry* 22: 109–115.
Wingler K, Böcher M, Flohé L, Kollmus H and Brigelius-Flohé R (1999) mRNA stability and selenocysteine insertion sequence efficiency rank gastrointestinal glutathione peroxidase high in the hierarchy of selenoproteins. *European Journal of Biochemistry* 259: 149–157.
Wittmann-Liebold B, Salnikow J and Erdmann VA (eds) (1986) *Advanced Methods in Protein Microsequence Analysis*. Berlin: Springer-Verlag.
Wright DJ (1987) The seed globulins. In: Hudson BJF (ed.) *Developments in Food Proteins*, vol. 5, pp. 81–157. London: Elsevier Applied Science.
Wright DJ (1988) The seed globulins – part II. In: Hudson BJF (ed.) *Developments in Food Proteins*, vol. 6, pp. 119–178. London: Elsevier Applied Science.

Food Sources

J K P Weder and H-D Belitz[†], Technische Universität München, Garching, Germany

Protein

Amino acids, peptides, and proteins are important constituents of food. They supply the required building blocks for protein biosynthesis. In addition, they directly contribute to the flavor of food and are precursors for aroma compounds and colors formed during thermal or enzymatic reactions in production, processing, and storage of food. Other food constituents, e.g., carbohydrates, take part in such reactions. Proteins also contribute significantly to the physical properties of food through their ability to stabilize gels, foams, doughs, emulsions, and fibrillar structures.

Table 1 World-wide annual protein production ($\times 10^6$ tonnes)

Source	*1993*	*1994*	*1995*	*1996*	*1997*	*1998*
Cereals	177	183	177	193	196	191
Oil seeds	53.1	60.7	60.2	62.1	65.6	68.0
Legumes	11.8	12.1	11.8	11.6	11.9	12.2
Root crops	8.5	8.1	8.4	8.9	8.4	8.4
Vegetables	5.1	5.3	5.6	5.9	6.0	6.1
Meat	37.9	39.1	40.2	40.6	42.0	42.9
Milk	18.0	18.2	18.4	18.5	18.7	18.7
Fish and aquatic animals	11.3	12.2	12.6	13.0	13.3	
Eggs	4.8	5.1	5.3	5.6	5.9	6.0
Total	328	344	340	359	368	353[a]

[a]Excluding fish and aquatic animals.
Data from FAO (1999) *FAO Year book Production*, vol. 52, 1998. FAO Statistics Series No. 148. Rome: Food and Agriculture Organization and by courtesy of Dr. Rehbein (Federal Research Centre for Fisheries, Hamburg, Germany, for fish and aquatic animals), protein calculated from production with data from Scherz H and Senser F (eds) (2000) *Food Composition and Nutrition Tables*, 6th edn. Boca Raton, FL: CRC Press.

Table 1 shows the most important protein sources and their contribution to world-wide production of protein. Cereals contribute to protein production by more than half, followed by oil seeds and meat. Besides plants and animals, algae (*Chlorella*, *Scenedesmus*, and *Spirulina* spp.), yeasts and bacteria may be used for protein production (single-cell protein (SCP)). Common carbon sources are glucose, molasses, starch, waste water, higher *n*-alkanes, and methanol. Yeasts of the genus *Candida* can be grown, for example, on paraffins, and they deliver about 0.75 t of protein per tonne of alkane. Bacteria of the genus *Pseudomonas* deliver about 0.3 t of protein per tonne of alcohol in aqueous methanol. Because of the high nucleic acid content of yeasts and bacteria (6–17% of dry mass), isolation of the protein from biomass is necessary. The importance of SCP in the future will depend on the total protein market, especially on prices and functional properties of individual proteins.

Table 2 provides data on the average protein content and the nutrient density – the ratio between the amount of protein (g) and the total energy value (MJ) of the digestible constituents – of selected foodstuffs. The protein content varies as follows: >20%

[†]Deceased.

Table 2 Protein content (%) of selected foodstuffs

No.[a]	Foodstuff	Average protein content		Nutrient density ($g\,MJ^{-1}$)
		Raw product	Edible portion	
Milk and dairy products				
1	Human milk	1.13	1.13	3.92
2	Cow's milk	3.33	3.33	12.05
3	Buffalo milk	4.01	4.01	8.94
4	Ewe's milk	5.27	5.27	13.17
5	Goat's milk	3.69	3.69	13.14
6	Condensed milk (7.5% fat)	6.49	6.49	11.78
7	Dried milk whole	25.20	25.20	12.50
8	Dried skimmed milk	35.00	35.00	23.05
9	Yoghurt (1.5–1.8% fat)	3.55	3.55	16.88
10	Blue cheese (50% fat dm[b])	21.10	21.10	14.35
11	Camembert cheese (40 % fat dm)	22.50	22.50	19.72
12	Cheddar cheese (50% fat dm)	25.40	25.40	15.43
13	Edam cheese (40% fat dm)	24.80	26.10	19.93
14	Emmental cheese (45% fat dm)	26.98	28.70	18.02
15	Feta cheese (45% fat dm)	17.00	17.00	17.24
Eggs				
16	Chicken egg, whole	11.35	12.90	19.98
17	Chicken egg, yolk	16.10	16.10	11.03
18	Chicken egg, white	11.10	11.10	53.37
Meat				
19	Beef (muscles only)	22.00	22.00	48.41
20	Veal (muscles only)	21.30	21.30	54.33
21	Pork (muscles only)	22.00	22.00	49.68
22	Mutton (muscles only)	20.40	20.40	43.13
23	Lamb (muscles only)	20.80	20.80	42.41
24	Corned beef, American	25.30	25.30	28.94
25	Luncheon meat	14.70	14.70	12.36
26	Meat extract	56.60	56.60	54.13
27	Sausage 'Cervelat'	19.89	20.30	12.43
28	Chicken, for roasting	14.73	19.90	28.70
29	Duck	14.48	18.10	19.17
30	Goose	9.89	15.70	11.10
31	Turkey	14.75	20.20	22.48
Fish				
32	Carp	9.36	18.00	37.22
33	Cod	13.28	17.70	54.55
34	Flounder	7.42	16.50	53.85
35	Halibut	16.08	20.10	49.67
36	Herring	12.74	18.20	18.80
37	Mackerel	12.16	18.70	24.66
38	Mullet	10.61	20.40	40.32
39	Redfish	8.74	18.20	41.09
40	Salmon	12.74	19.90	23.65
41	Sardine	11.45	19.40	38.98
42	Trout	10.14	19.50	45.09
43	Tuna	13.12	21.50	22.90
Cereals				
44	Barley (whole grain)	9.84	9.84	7.35
45	Maize (whole grain)	8.54	8.54	6.17
46	Corn flakes	7.15	7.15	4.77
47	Oats (whole grain)	11.69	11.69	7.91
48	Rolled oats	12.53	12.53	8.09
49	Rice, polished	6.83	6.83	4.68
50	Rye (whole grain)	8.82	8.82	7.08
51	Rye flour, type 997	6.86	6.86	5.22
52	Rye bread	6.22	6.22	6.76
53	Sorghum	10.30	10.30	6.97
54	Wheat (whole grain)	11.73	11.73	8.95
55	Wheat flour, type 405	9.84	9.84	6.98
56	Wheat flour, type 1700	11.23	11.23	8.76
57	Wheat bread	7.61	7.61	7.54
Roots and tubers				
58	Beetroot	1.19	1.53	8.73
59	Cassava	0.74	1.00	1.75
60	Celeriac	1.13	1.55	20.18
61	Potato	1.63	2.04	6.84
62	Potato flakes	8.60	8.60	6.26
63	Sweet potato	1.32	1.63	3.55
64	Taro	1.68	2.00	4.40
65	Topinambur	1.68	2.44	18.72
66	Yam	1.68	2.00	4.77
Leaves, stems, and flowers				
67	Artichoke	1.15	2.40	25.68
68	Broccoli	2.01	3.30	29.88
69	Brussel sprouts	3.47	4.45	29.41
70	Cauliflower	1.53	2.46	25.94
71	Chinese leaves	0.94	1.19	22.85
72	Red cabbage	1.17	1.50	16.24
73	Soya bean sprouts	4.59	5.53	26.15
Vegetable fruits				
74	Aubergine	1.03	1.24	17.12
75	Cucumber	0.44	0.60	11.59
76	Green peppers	0.90	1.17	13.65
77	String beans	2.25	2.39	17.24
78	Tomato	0.91	0.95	12.98
Legumes and oil seeds				
79	Bean, white dry	21.09	21.30	19.32
80	Chick pea, dry	19.80	19.80	17.02
81	Lentil, dry	23.50	23.50	17.61
82	Pea, dry	22.67	22.90	19.89
83	Peanut	25.25	25.25	10.75
84	Soya bean, dry	28.00	33.73	24.98
Fruits				
85	Apple	0.31	0.34	1.49
86	Apricot	0.82	0.90	4.90
87	Apricot, dried	5.00	5.00	4.91
88	Banana	0.77	1.15	3.07
89	Cherry (morello)	0.80	0.90	4.00
90	Date, dried	1.61	1.85	1.58
91	Fig	1.30	1.30	5.00
92	Fig, dried	3.50	3.54	3.34
93	Grape	0.65	0.68	2.38
94	Grape, dried (raisin)	2.46	2.46	2.09
95	Orange	0.72	1.00	5.58
96	Peach	0.70	0.76	4.33
97	Pear	0.44	0.47	2.02
98	Plum	0.56	0.60	2.92
99	Strawberry	0.80	0.82	6.03
Mushrooms				
100	Champignon	2.69	2.74	41.10
101	Chanterelle	0.96	1.57	32.89
102	Edible boletus	2.88	3.60	42.32
103	Edible boletus, dried	19.70	19.70	37.66
104	Morel	1.38	1.66	41.44
105	Oyster mushroom	2.35	2.35	52.07
106	Truffles	5.53	5.53	48.99

[a]These numbers are used in **Table 3**.

[b]dm, dry matter.

Data from Scherz H and Senser F (eds) 2000 *Food Composition and Nutrition Tables*, 6th edn. Boca Raton, FL: CRC Press.

Table 3 Amino acid compositions of selected foodstuffs

	Foodstuffs[a]												
Amino acids	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
Alanine	0.056	0.130	0.130	0.220	0.140	0.270	0.910	1.25	0.160	0.730	0.890	0.760	0.890
Arginine	0.051	0.130	0.110	0.180	0.130	0.230	0.920	1.28	0.130	1.65	0.820	0.900	1.02
Aspartic acid	0.120	0.290	0.310	0.460	0.300	0.550	1.99	2.74	0.280	0.830	1.67	1.81	2.00
Cystine	0.024	0.028	0.048	0.060	0.083	0.066	0.230	0.310	0.027	0.120	0.130	0.210	0.250
Glutamic acid	0.220	0.790	0.480	1.07	0.780	1.50	5.51	7.57	0.700	4.97	4.95	6.62	6.15
Glycine	0.036	0.076		0.110	0.074	0.180	0.560	0.770	0.086	0.510	0.460	0.470	0.550
Histidine	0.031	0.095	0.078	0.130	0.079	0.170	0.660	0.920	0.095	0.990	0.730	0.800	0.930
Isoleucine	0.077	0.220	0.200	0.280	0.230	0.450	1.61	2.24	0.220	1.19	1.52	1.81	1.48
Leucine	0.130	0.360	0.370	0.530	0.390	0.720	2.47	3.43	0.380	2.14	2.19	2.52	2.65
Lysine	0.086	0.280		0.460	0.340	0.520	1.96	2.72	0.280	2.38	1.65	2.07	2.37
Methionine	0.024	0.090	0.097	0.140	0.094	0.170	0.620	0.860	0.092	0.520	0.590	0.770	0.780
Phenylalanine	0.054	0.180	0.162	0.260	0.180	0.340	1.22	1.70	0.190	1.22	1.20	1.45	1.44
Proline	0.120	0.340		0.550	0.470	0.720	2.55	3.50	0.420	2.35	2.38	3.05	2.91
Serine	0.059	0.210	0.230	0.320	0.210	0.410	1.43	1.97	0.220	1.33	1.29	1.58	1.57
Threonine	0.063	0.160	0.180	0.240	0.230	0.330	1.16	1.61	0.160	0.920	0.840	0.980	1.13
Tryptophan	0.022	0.049	0.053	0.070	0.050	0.088	0.350	0.490	0.042	0.210	0.310	0.290	0.400
Tyrosine	0.056	0.180	0.183	0.260	0.240	0.320	1.28	1.78	0.180	1.02	1.08	1.30	1.33
Valine	0.081	0.240	0.220	0.320	0.280	0.480	1.73	2.40	0.270	1.46	1.61	1.81	1.88
	Foodstuffs[a]												
Amino acids	*14*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*28*	*29*
Alanine	0.920	0.890	1.03	0.830	1.69	1.64	1.53	1.44	1.29			1.44	
Arginine	1.00	0.890	1.28	0.680	1.54	1.54	1.53	1.44	1.39	1.63	0.790	1.39	1.10
Aspartic acid	1.58	1.46	1.76	1.23	2.34	2.40	2.43	2.15	1.89			2.27	
Cystine	0.170	0.310	0.310	0.290	0.280	0.280	0.310	0.290	0.170	0.320	0.120	0.300	
Glutamic acid	5.76	1.81	2.20	1.64	4.13	3.97	3.91	4.30	3.05			3.69	
Glycine	0.510	0.530	0.620	0.500	1.56	1.34	1.42	1.43	1.09			1.40	
Histidine	1.02	0.330	0.440	0.280	0.850	0.800	0.990	0.630	0.600	0.890	0.360	0.610	0.410
Isoleucine	1.73	0.930	1.09	0.740	1.25	1.29	1.27	1.21	1.02	1.24	0.560	1.29	0.940
Leucine	2.99	1.26	1.63	1.08	1.95	1.89	1.92	1.80	1.69	1.92	1.03	1.78	1.40
Lysine	2.39	0.890	1.30	0.740	2.31	2.05	2.20	2.00	1.89	2.13	1.01	2.04	1.56
Methionine	0.790	0.450	0.470	0.470	0.650	0.600	0.720	0.560	0.530	0.690	0.330	0.640	0.450
Phenylalanine	1.61	0.800	0.790	0.760	1.06	1.02	0.980	0.920	0.820	0.910	0.540	0.910	0.710
Proline	3.73	0.590	0.780	0.500	1.28	1.15	1.21	1.02	0.970			1.05	
Serine	1.66	1.15	1.62	0.920	1.14	1.15	1.12	1.04	0.820			0.920	
Threonine	1.14	0.710	1.01	0.580	1.15	1.13	1.25	1.09	0.890	1.11	0.550	1.01	0.790
Tryptophan	0.430	0.230	0.290	0.200	0.290	0.300	0.310	0.290	0.200	0.230	0.120	0.280	
Tyrosine	1.61	0.590	0.780	0.460	0.890	0.880	0.910	0.770	0.820	0.850	0.460	0.760	
Valine	2.12	1.12	1.24	0.980	1.32	1.31	1.42	1.18	1.12	1.27	0.710	1.18	0.870
	Foodstuffs[a]												
Amino acids	*30*	*31*	*32*	*33*	*34*	*35*	*36*	*37*	*38*	*39*	*40*	*41*	*42*
Alanine	0.970		1.45	1.42	1.24	1.77	1.52	1.88	1.55	1.68	1.67	1.60	1.55
Arginine	0.980	1.21	1.27	1.21	1.16	1.37	1.18	1.16	1.36	1.19	1.33	1.31	1.40
Aspartic acid	1.41		2.31	2.01	2.07	2.15	2.31	2.12	2.90	2.28	2.22	2.32	2.36
Cystine		0.270	0.180	0.250	0.150	0.240	0.240	0.230	0.310	0.220	0.290	0.220	0.170
Glutamic acid	2.34		3.19	3.13	3.18	3.01	3.23	3.17	3.47	3.30	3.23	3.04	3.33
Glycine	0.990		1.00	0.940	0.930	1.15	1.13	1.41	1.38	1.10	1.63	1.24	1.47
Histidine	0.440	0.520	0.420	0.520	0.450	0.480	0.520	0.840	0.610	0.420	0.660	0.460	0.570
Isoleucine	0.740	1.01	1.00	0.990	0.920	1.27	1.04	1.09	1.09	1.14	1.16	1.19	1.07
Leucine	1.32	1.47	1.68	1.69	1.60	1.94	1.75	1.80	1.95	1.78	1.77	1.87	1.78
Lysine	1.24	1.74	2.11	2.05	1.82	1.56	1.75	1.73	2.04	1.90	2.02	2.28	2.02
Methionine	0.380	0.530	0.590	0.600	0.580	0.800	0.660	0.640	0.660	0.640	0.700	0.640	0.660
Phenylalanine	0.660	0.770	0.890	0.840	0.700	0.680	0.750	0.840	0.860	0.840	0.910	0.910	0.920
Proline	0.760		0.640	0.820	0.600	0.810	0.840	0.790	0.790	0.750	1.00	0.850	0.850
Serine	0.620		0.990	0.990	0.860	1.27	1.05	0.900	0.950	1.04	1.01	1.10	0.970
Threonine	0.700	0.810	1.04	0.970	0.920	0.990	1.04	0.970	1.02	1.01	1.11	1.12	1.08
Tryptophan	0.200	0.160	0.210	0.240	0.210	0.260	0.210	0.270	0.320	0.200	0.260	0.240	0.240
Tyrosine	0.510	0.280	0.740	0.710	0.640	0.680	0.670	0.640	0.740	0.560	0.720	0.810	0.680
Valine	0.970	0.950	1.05	1.09	1.06	1.30	1.21	1.21	1.24	1.04	1.39	1.45	1.25

Continued

Table 3 Continued

	Foodstuffs[a]												
Amino acids	*43*	*44*	*45*	*46*	*47*	*48*	*49*	*50*	*51*	*52*	*53*	*54*	*55*
Alanine	1.61	0.560	0.790	0.800	0.720	0.790	0.500	0.520	0.410	0.300	0.880	0.510	0.370
Arginine	1.25	0.560	0.420	0.240	0.850	0.870	0.570	0.490	0.330	0.420	0.380	0.620	0.430
Aspartic acid	2.88	0.680	0.620	0.540	1.11	1.29	0.780	0.680	0.610	0.480	0.710	0.700	0.480
Cystine	0.290	0.220	0.140	0.160	0.320	0.390	0.110	0.190	0.140	0.130	0.100	0.290	0.240
Glutamic acid	3.52	2.81	1.78	1.86	2.90	3.08	1.58	2.57	2.05	1.92	2.29	4.08	3.66
Glycine	1.17	0.540	0.430	0.340	0.780	0.850	0.410	0.500	0.430	0.320	0.430	0.720	0.420
Histidine	1.09	0.210	0.260	0.240	0.270	0.300	0.170	0.190	0.180	0.240	0.220	0.280	0.220
Isoleucine	1.21	0.460	0.430	0.330	0.560	0.610	0.340	0.390	0.320	0.260	0.580	0.540	0.460
Leucine	2.17	0.800	1.22	1.24	1.02	1.13	0.660	0.670	0.540	0.470	1.36	0.920	0.820
Lysine	2.21	0.380	0.290	0.180	0.550	0.500	0.290	0.400	0.280	0.300	0.260	0.380	0.240
Methionine	0.610	0.180	0.190	0.170	0.230	0.240	0.170	0.140	0.120	0.060	0.200	0.220	0.170
Phenylalanine	1.05	0.590	0.460	0.430	0.700	0.780	0.390	0.470	0.360	0.350	0.440	0.640	0.550
Proline	0.880	1.26	1.02	0.970	0.870	0.840	0.420	1.25	0.840	0.720	1.55	1.56	1.45
Serine	1.05	0.540	0.520	0.470	0.740	0.710	0.410	0.450	0.340	0.350	0.420	0.710	0.660
Threonine	1.18	0.430	0.390	0.320	0.490	0.530	0.280	0.360	0.310	0.250	0.440	0.430	0.320
Tryptophan	0.300	0.150	0.070	0.050	0.190	0.190	0.090	0.110	0.070	0.050	0.110	0.150	0.120
Tyrosine	0.970	0.390	0.380	0.270	0.450	0.570	0.260	0.230	0.220	0.170	0.250	0.410	0.320
Valine	1.42	0.580	0.510	0.440	0.790	0.810	0.490	0.530	0.410	0.330	0.580	0.620	0.490

	Foodstuffs[a]												
Amino acids	*56*	*57*	*58*	*60*	*61*	*62*	*63*	*68*	*69*	*70*	*71*	*72*	*74*
Alanine	0.490	0.240		0.086	0.110	0.350							
Arginine	0.600	0.310	0.027	0.044	0.120	0.610	0.065	0.190	0.280	0.110	0.080	0.110	0.042
Aspartic acid	0.660	0.390		0.164	0.430	1.83							
Cystine	0.250	0.180		0.004	0.020	0.140	0.025				0.010	0.030	
Glutamic acid	3.75	3.15		0.283	0.460	1.89							
Glycine	0.630	0.290		0.047	0.120	0.340							
Histidine	0.250	0.180	0.021	0.024	0.040	0.290	0.029	0.063	0.110	0.049	0.026	0.027	0.021
Isoleucine	0.520	0.380	0.049	0.048	0.100	0.330	0.068	0.130	0.210	0.110		0.043	0.063
Leucine	0.860	0.590	0.053	0.075	0.140	0.590	0.084	0.160	0.230	0.170		0.061	0.077
Lysine	0.350	0.200	0.082	0.074	0.130	0.690	0.066	0.150	0.250	0.140	0.058	0.071	0.034
Methionine	0.210	0.130	0.005	0.018	0.030	0.080	0.028	0.050	0.040	0.048	0.032	0.014	0.007
Phenylalanine	0.590	0.420	0.026	0.047	0.100	0.420	0.079	0.120	0.150	0.077	0.047	0.032	0.054
Proline	1.57	0.960		0.040	0.110	0.470							
Serine	0.740	0.390		0.049	0.100	0.460							
Threonine	0.390	0.250	0.033	0.044	0.090	0.360	0.068	0.120	0.160	0.110	0.052	0.042	0.043
Tryptophan	0.150	0.080	0.013	0.012	0.030	0.070	0.028	0.037	0.050	0.034	0.020	0.012	0.011
Tyrosine	0.370	0.210		0.025	0.080	0.500	0.071			0.035	0.039		
Valine	0.600	0.390	0.047	0.073	0.130	0.490	0.110	0.170	0.240	0.150	0.070	0.046	0.073

	Foodstuffs[a]												
Amino acids	*75*	*76*	*77*	*78*	*79*	*80*	*81*	*82*	*83*	*84*	*85*	*88*	*90*
Alanine				0.026	0.740		1.29	0.480	0.810	1.53	0.015	0.046	
Arginine	0.045	0.023	0.100	0.018	1.49	1.48	2.24	3.71	3.46	2.36	0.008	0.054	0.040
Aspartic acid				0.121	2.45		3.16	1.92	3.31	3.99	0.101	0.115	
Cystine			0.024	0.001	0.230	0.280	0.250	0.450	0.430	0.590	0.001	0.002	
Glutamic acid				0.337	4.33		4.49	3.46	5.63	6.49	0.025	0.105	
Glycine				0.018	0.950		1.30	0.590	1.64	1.42	0.009	0.042	
Histidine	0.008	0.014	0.049	0.013	0.700	0.530	0.710	0.770	0.710	0.830	0.006	0.077	0.040
Isoleucine	0.019	0.045	0.110	0.023	1.49	1.14	1.19	1.88	1.23	1.78	0.010	0.038	0.060
Leucine	0.025	0.045	0.140	0.030	2.26	1.46	2.11	2.34	2.03	2.84	0.016	0.085	0.062
Lysine	0.026	0.050	0.140	0.029	1.87	1.37	1.89	2.13	1.10	1.90	0.015	0.057	0.044
Methionine	0.007	0.260	0.260	0.220	0.350	0.310	0.580	0.003	0.009	0.022			
Phenylalanine	0.014	0.054	0.073	0.024	1.40	0.960	1.40	1.39	1.54	1.97	0.009	0.034	0.051
Proline				0.016	0.980		1.22	0.490	1.43	1.82	0.010	0.040	
Serine				0.028	1.38		1.51	0.980	1.83	1.69	0.012	0.049	
Threonine	0.016	0.049	0.093	0.023	1.15	0.700	1.12	1.57	0.850	1.49	0.008	0.038	0.049
Tryptophan	0.004	0.009	0.027	0.006	0.230	0.160	0.250	0.350	0.320	0.450	0.002	0.018	0.049
Tyrosine			0.050	0.012	0.970	0.660	0.840	1.22	1.19	1.25	0.005	0.021	
Valine	0.021	0.032	0.130	0.023	1.63	0.980	1.39	1.82	1.45	1.76	0.012	0.057	0.076

Continued

Table 3 Continued

	Foodstuffs[a]										
Amino acids	*92*	*94*	*95*	*96*	*99*	*100*	*101*	*102*	*103*	*105*	*106*
Alanine	0.160	0.091	0.029	0.039	0.044		0.044				
Arginine	0.090	0.305	0.073	0.017	0.037	0.200	0.076	0.260			0.650
Aspartic acid	1.14	0.087	0.122	0.090	0.191						
Cystine	0.140	0.006	0.003	0.009	0.007	0.014	0.120	0.290	2.08		0.150
Glutamic acid	0.380	0.118	0.066	0.139	0.126						
Glycine	0.260	0.063	0.023	0.015	0.034		0.035				
Histidine	0.090	0.051	0.012	0.017	0.016	0.057	0.028	0.220	1.59	0.050	0.100
Isoleucine	0.140	0.047	0.020	0.013	0.019	0.110	0.040	0.030	0.210	0.110	0.160
Leucine	0.140	0.075	0.032	0.028	0.044	0.120	0.086	0.120	0.840	0.170	0.400
Lysine	0.140	0.071	0.039	0.029	0.034	0.170	0.039	0.190	1.35	0.150	0.490
Methionine	0.040	0.013	0.008	0.030	0.001	0.023	0.009	0.058	0.420	0.050	
Phenylalanine	0.120	0.047	0.020	0.018	0.025	0.074	0.065	0.100	0.730	0.100	0.190
Proline	0.130	0.157	0.189	0.027	0.027		0.040				
Serine	0.260	0.051	0.043	0.033	0.033		0.049				
Threonine	0.120	0.055	0.020	0.027	0.026	0.087	0.088	0.110	0.750	0.120	0.380
Tryptophan	0.040	0.005	0.007	0.005	0.015	0.024	0.048	0.210	1.46	0.030	0.020
Tyrosine	0.220	0.010	0.013	0.020	0.029	0.066	0.058	0.120	0.860		0.180
Valine	0.290	0.071	0.033	0.039	0.025	0.090	0.052	0.078	0.560	0.140	0.250

[a]Numbering as in **Table 2**; units are grams of amino acid per 100 g of edible portion.
Data from Scherz H and Senser F (eds) 2000 *Food Composition and Nutrition Tables*, 6th edn. Boca Raton, FL: CRC Press.

(cheeses, meat, legumes, oil seeds); 10–20% (fish, eggs); 5–10% (cereals); and < 5% (milk, roots, tubers, vegetables, fruits, mushrooms).

A considerable number of proteins have been isolated from various foods and characterized by physical and chemical properties. Cereals and cereal products are amongst the most important staple foods of mankind. Proteins provided by bread consumption in industrial countries meet about one-third of the daily requirement. The major cereals are wheat, maize, rice, barley, sorghum, oats, millet, and rye. Wheat and rye have a special role since only they are suitable for bread-making. With the example of wheat, the cereal proteins have been separated by Osborne, on the basis of their solubility, into four fractions: the water-soluble albumins, the salt-soluble globulins, the 70% aqueous ethanol-soluble prolamins, and the remaining glutelins. In the literature, Osborne fractions derived from different cereals are often designated by special names, e.g., gliadin and glutenin for wheat prolamins and glutelins, respectively. The levels of Osborne fractions differ amongst cereals with albumins amounting to 4–44%, globulins 3–12%, prolamins 2–48%, and glutelins 24–77% of the whole protein fraction. Each of the Osborne fractions consists of a larger number of proteins. Albumins and globulins contain the enzymes, whereas prolamins and glutelins are storage proteins. The glutelins can be separated into two subfractions after reduction of the disulfide bonds: the low-molecular-weight (LMW) subunits and the high-molecular-weight (HMW) subunits. Wheat prolamins and glutelins, both fractions together also designated wheat gluten, are responsible for the characteristic rheological properties of wheat dough. Wheat prolamins consist of ω-, α-, and γ-gliadins, designated according to their electrophoretic mobility. In addition, the LMW subunit fraction of wheat glutelins also contains some gliadins (ω5-, ω1,2-, and γ-gliadins). The amino acid sequences of some gluten proteins from wheat are shown in the article '**Protein: Chemistry**' (HMW subunit of glutenin, x-type and y-type; LMW subunit of glutenin; α-gliadin; γ-gliadin).

Meat and meat products are other important staple foods, in particular in industrial countries. The main meat-producing warm-blooded animals are pig, cattle, poultry, sheep, goats, and buffalo. Meat proteins, i.e., the proteins of the muscle, are divided into three groups: proteins of the contractile apparatus (myofibrillar proteins), soluble proteins (sarcoplasma proteins), and insoluble proteins (connective tissue and membrane proteins). The myofibrillar proteins of a typical mammalian muscle amount to about 60% of total muscle protein, with myosin (29%) and actin (13%) as their predominating components and about 20 minor components including connectin, tropomyosins, troponins, and actinins. The sarcoplasma proteins form about 30% of total protein. They contain most of the enzymes, in particular those of glycolysis and the pentosephosphate cycle, but also considerable amounts of creatine kinase (2.7% of total protein), myoglobin, and some hemoglobin. The insoluble proteins contain collagen as the main component, besides elastin, insoluble enzymes, and cytochrome *c*. In connective tissue, collagen forms a triple-stranded helix composed of α-helices.

Covalent cross-links are formed between the fibers of collagen during maturation and aging, thus improving its mechanical strength. When heated, collagen fibers shrink or are converted into gelatine, depending on the temperature. The structure of the gelatine obtained after cooling depends on the gelatine concentration and temperature gradient. Collagen contains two unusual amino acids, 4-hydroxyproline and 5-hydroxylysine. Since the occurrence of the former is confined to connective tissue, its determination provides data on the extent of connective tissue content of a meat product. The primary structure of bovine skin collagen is given in the first edition of this encyclopedia (Belitz H-D (1993) Protein-Chemistry. In: Macrae R, Robinson RK and Sadler MJ (eds) *Encyclopaedia of Food Science, Food Technology, and Nutrition,* pp. 3781–3791. London: Academic Press) showing the characteristic three-amino-acid repeats with glycine in the first position, often followed by proline and hydroxyproline.

Milk and dairy products form a further important group of staple foods. Milk generally means cow's milk, but milk from buffalo, goats, and sheep is of importance in some regions. Milk proteins, in particular the caseins, play an important role in processing to yield dairy products such as cheese and sour milk products. The caseins, first isolated by Hammarsten in 1877, make up about 80% of total milk proteins. They have been separated later into different fractions: α_{s1}-, α_{s2}-, κ-, β-, and γ-caseins, constituting 34, 8, 9, 25, and 4% of total protein, respectively. Each of these fractions occurs in the form of different genetic variants, designated A, B, C, etc., depending on the breed from which they have been isolated. In cheese-making, the specific cleavage of κ-casein by chymosin (EC 3.4.23.4) into para-κ-casein and a glycopeptide (so-called, though not always containing a sugar moiety) reduces the solubility of the casein complexes and casein micelles, thus leading to their aggregation followed by gel formation (curd formation). The whey proteins (about 20% of total protein) consist of β-lactoglobulins, α-lactalbumins (both in different genetic variants), serum albumin, immunoglobulins, and proteose-peptone. Also, more than 40 enzymes occur in the whey protein fraction, but in much lower quantities than the other components. Whey proteins can be incorporated into the curd using several new processing methods of cheese-making in order to increase the yield and also to reduce waste water or whey treatment costs. The primary structures of some proteins from bovine milk are shown in the first edition of this encyclopedia (Belitz H-D (1993) see above) (α_{s1}-, α_{s2}-, β-, and κ-casein; α-lactalbumin; β-lactoglobulin).

Legumes (pulses) are very important staple foods in some parts of the world, e.g., soya beans in South-east Asia and common beans in Latin America. Other legumes, some of greater regional importance, include peas, peanuts, chick peas, broad beans, and lentils. Legume proteins, when fractionated, according to Osborne, in a similar way to cereal proteins, yield three fractions: albumins, globulins, and glutelins. The portion of the fractions varies, depending on the legume species, but globulins always predominate. The globulins are subdivided, initially according to sedimentation during ultracentrifugation, into 11S and 7S globulins (legumins and vicilins, respectively). Again, the subfractions derived from different legumes are sometimes designated by special names, e.g., glycinin and arachin for soya bean and peanut legumin, as well as β-conglycinin and phaseolin for soya bean and common bean vicilin, respectively. Soya protein isolates, produced by diluted alkali extraction of defatted soya bean flakes followed by acid precipitation, are texturized and flavored for use as meat substitutes or are added to foods to raise their protein level and improve their processing qualities such as the water-binding capacity or emulsion stability. The isolates contain about 95% protein and consist of 11S and 7S globulins. The similarity between the caseins from bovine milk and the globulins from soya beans is reflected by the production of some typical Asian foods such as soy milk, soy curd (tofu), and soy cheese (sufu). The primary structures of some legume proteins are shown in the article '**Protein**: Chemistry' (glycinin, pea legumin; β-conglycinin, pea vicilin, phaseolin.)

Eggs are used as a food not only because of their excellent nutritional quality but also because of their functional properties. Eggs generally means chicken eggs; those of other birds (geese, ducks, plovers, seagulls, quail) are less important. Egg proteins are divided into those of egg white and those of egg yolk. Egg white proteins (about 10% of total egg white) are ovalbumin, conalbumin (ovotransferrin), ovomucoid, ovomucin, lysozyme, ovoglobulin G_2, ovoglobulin G_3, and some minor components (54, 12, 11, 3.5, 3.4, 4, 4, and 2.5% of total egg white protein, respectively). Ovalbumin, conalbumin, ovomucin, and the ovoglobulins G contribute to foam formation and foam stability. Yolk proteins (about 17% of total yolk) are phosvitin, the livetins, and the protein moieties of high-density lipoproteins (HDL) and low-density (LDL) lipoproteins (13, 31, 36, and 20% of total yolk protein, respectively). Apart from phospholipids, LDL and proteins are responsible for the emulsifying effect of whole eggs or egg yolk alone. Owing to the ability of all egg proteins, except ovomucoid and phosvitin, to coagulate

Table 4 Daily requirement of essential amino acids (milligrams per kilogram of bodyweight)

Amino acid	*Infants (3–4 months)*	*Children (2 years)*	*Schoolboys (10–12 years)*	*Adults*
Histidine[a]	28	?	?	[8–12]
Isoleucine	70	31	28–30	10
Leucine	161	73	44–45	14
Lysine	103	64	44–60	12
Threonine	87	37	28–35	7
Total SAA	58	27	22–27	13
Total AAA	125	69	22–27	14
Tryptophan	17	12.5	3.3–4	3.5
Valine	93	38	25–33	10
Total	714	352	216–261	84

[a]Evidence is accumulating that histidine is essential even for adults.
SAA, sulfur amino acids (methionine+cystine).
AAA, aromatic amino acids (phenylalanine+tyrosine).
Data from WHO (1985) *Energy and Protein Requirements: Report of a Joint FAO/WHO/UNU Expert Consultation*. WHO Technical Report Series 724. Geneva: World Health Organization.

when heated, egg products are important food-binding agents.

The amino acid compositions of selected foodstuffs are shown in Table 3. The nutritional quality of a food protein depends on the absolute content of essential amino acids, the relative proportions of essential amino acids, and their ratios to nonessential amino acids. In addition, the digestibility of the food protein, the influence by other food components such as dietary fibers, polyphenols, or proteinase inhibitors, and also the total food energy intake are of importance. (*See* '**Protein**: Requirements'; **Protein**: Quality.) The daily requirements of humans at different ages for essential amino acids are compiled in Table 4. During pregnancy and lactation, the first 6 months, and after 6 months, the daily requirement increases by 13, 24, and 18%, respectively. The biological value of a protein is generally limited by the following amino acids:

- lysine: deficient in proteins of cereals and other plants;
- methionine: deficient in proteins of bovine milk and meat;
- threonine: deficient in wheat and rye;
- tryptophan: deficient in casein, corn and rice.

The biological values of some important food proteins are given in the article '**Protein**: Quality'. The highest value observed so far is for a blend of 35% egg and 65% potato proteins (one chicken egg and 500 g of potatoes).

See also: **Amino Acids**: Metabolism; **Rheological Properties of Food Materials**; **Protein**: Requirements; Functional Properties; Quality; **Single-cell Protein**: Algae; Yeasts and Bacteria

Further Reading

Altschul AM (ed.) (1974–1985) *New Protein Foods*, vols. 1–5. New York: Academic Press.
FAO (1999) *FAO Yearbook Production*, vol. 52, 1998. FAO Statistics Series No. 148. Rome: Food and Agriculture Organization.
Fox PF (1992) *Advanced Dairy Chemistry – 1: Proteins*. London: Elsevier Applied Science.
Hudson BJF (ed.) (1982–1991) *Developments in Food Proteins*, vols. 1–7. London: Elsevier Applied Science.
Hudson BJF (ed.) (1994) *New and Developing Sources of Food Proteins*. London: Chapman & Hall.
KeShun L (1997) *Soybeans: Chemistry, Technology, and Utilization*. London: Chapman & Hall.
Klostermeyer H, Schmandke H, Soeder CJ *et al.* (1993) Proteins. In: Elvers B, Hawkins S, Russey W and Schulz G (eds) *Ullmann's Encyclopedia of Industrial Chemistry*, 5th edn, vol. A22, pp. 289–339. Weinheim: VCH Verlagsgesellschaft.
Kulp K and Ponte JG, Jr. (2000) *Handbook of Cereal Science and Technology*, 2nd edn. New York: Marcel Dekker.
Lasztity R (ed.) (1995) *The Chemistry of Cereal Proteins*, 2nd edn. Boca Raton, FL: CRC Press.
Nakai S and Modler HW (eds) (1996) *Food Proteins: Properties and Characterization*. Weinheim: Verlag Chemie.
Nakai S and Modler HW (eds) (2000) *Food Proteins: Processing Applications*. New York: Wiley-VCH.
Sadler MJ, Strain JJ and Caballero B (eds) (1999) *Encyclopedia of Human Nutrition*, vols. 1–3. London: Academic Press.
Scherz H and Senser F (eds) (2000) *Food Composition and Nutrition Tables*, 6th edn. Boca Raton, FL: CRC Press.
Sikorski ZE, Pan BS and Shahidi F (eds) (1994) *Seafood Proteins*. New York: Chapman & Hall.
Welch RAS, Burns DJW, Davis SR, Popay AI and Prosser CG (eds) (1997) *Milk Composition, Production and Biotechnology*. Wallingford, UK: CAB.
WHO (1985) *Energy and Protein Requirements: Report of a Joint FAO/WHO/UNU Expert Consultation*. WHO Technical Report Series 724. Geneva: World Health Organization.

Determination and Characterization

D Otter, New Zealand Dairy Research Institute, Palmerston North, New Zealand

Background

The nutritional and functional value of proteins in food have resulted in considerable research efforts devoted to their analysis. This work has been

undertaken on a number of different levels, i.e., total protein content, defining and quantifying the individual protein components, and characterization of protein biological activity (direct or indirect) and functional properties.

Food Matrix

Food manifests itself in a wide variety of guises from the basic, raw material such as fruit through to highly refined processed foods. Food contains a large number of other components, as well as proteins, such as carbohydrates, fats, and minerals/salts. These components all interact to form the food matrix. It is usually necessary to separate the protein fraction from the other nonprotein material before any meaningful analysis of the individual proteins can be undertaken. The effect of the other components in the food matrix must be considered, however, when characterizing the protein biological activity and functional properties.

Total Protein Content

The gross analysis of protein in food is often required prior to the characterization of the individual protein components. The amount of protein is also necessary during the purification of the individual proteins to determine their relative purity. The three principal protein analysis methods are total nitrogen, UV adsorption, and chemical or dye-binding reactions.

The Kjeldahl method, in which the sample is digested with acid, a boiling aid, and a catalyst to produce an ammonium salt that is then reacted with alkali, and the resulting free ammonia is determined either titrimetrically or photometrically, results in an approximate protein level. Interference by nonprotein nitrogen-containing material and inaccuracies caused by differences in the nitrogen content of various proteins must be considered.

UV adsorption at either 280 nm (aromatic band of tyrosine and tryptophan) or 205–220 nm (peptide band) provides a rapid, nondestructive, and sensitive estimate of protein content. This is subject to interference from nucleic acids and other compounds and can be used to quantify purified proteins if their extinction coefficients are known.

A number of chromophore or dye-binding assays are available, including the Biuret, Lowry, and bicinchoninic acid reagents and the Coomassie Blue dye. The methods differ in their sensitivity and require a prepurified calibrating protein for quantitative results.

Individual Protein Characterization

Electrophoresis

Electrophoresis encompasses a wide range of techniques used for the separation and analysis of complex mixtures of proteins and peptides, and for the purification of small amounts of these different compounds. Separation is based on the differential movement of charged particles under the influence of an electric field where the overall motion is the product of the charge on the particles and the field strength, and where this movement is balanced by the frictional resistance of the medium. Four parameters, viz. differences in molecular size, charge or hydrophobicity, and specific interactions with other molecules, can be used to optimize the separation of molecules from each other, thus making electrophoresis a very powerful analytical technique. In the separation of proteins and peptides, the charged amino acids and carboxyl and amino terminal groups contribute to the overall charge on the protein and result in migration to either the anode or the cathode when an electric field is applied.

There are presently two methods commonly used for the electrophoretic separation of proteins and peptides. Gel electrophoresis utilizes a 'solid,' immobilized matrix to retard the movement of the sample through the aqueous phase. Initially, cellulose acetate strips and starch gels provided this matrix, although they have generally been superseded by polyacrylamide. Capillary electrophoresis (CE) occurs in free solution in very narrow bore capillary tubes and with high separation voltages, and is the electrophoretic analog of high-performance liquid chromatography (HPLC). Narrow capillary tubes allow rapid separation through their excellent heat dissipation, thus minimizing convective mixing and diffusion spreading of the separated zones.

Polyacrylamide gel electrophoresis Polyacrylamide gels are three-dimensional networks of acrylamide reacted with the bifunctional reagent *N,N′*-methylene-bis-acrylamide (abbreviated as Bis) via a free-radical initiated vinyl polymerization mechanism. The pore size of the gel is very reproducible and is directly related to the ratio of acrylamide to Bis. The resulting gels are described in terms of %T, the concentration (w/v) of acrylamide and Bis, and %C, the weight percentage of the cross-linker in T. For proteins, %T values of 5–10% result in gels with relative molecular mass (M_r) ranges of 20 000–200 000 Da. Separation of proteins in complex samples based on size, net charge, and hydrophobicity is possible using different polyacrylamide gel electrophoresis (PAGE) formats.

Sodium dodecyl sulfate–polyacrylamide gel electrophoresis This technique separates proteins according to their subunit size. The natural charge of the protein is masked by saturation with the anionic detergent sodium dodecyl sulfate (SDS), which also denatures the protein degrading the secondary, tertiary, and quaternary protein structure and aiding solubilization. The inclusion of a reducing agent such as 2-mercaptoethanol breaks any disulfide bonds, resulting in rod-like protein monomers containing one SDS molecule for every two amino acid residues. In an electric field, the protein–SDS complexes therefore migrate towards the anode at a rate dependent on the number of SDS molecules (or the M_r of the protein) and the pore size of the gel.

The M_r of the sample proteins can be determined by direct correlation of their migration distance with proteins of known M_r. Resolution of the protein bands in SDS–polyacrylamide gel electrophoresis (SDS–PAGE) can be improved by using a discontinuous gel system, where a stacking gel is used to concentrate the protein sample into sharp bands before the main separation gel. Differences in pH and ionic strength between the stacking gel and the electrophoresis buffer (isotachophoresis) are used to achieve this. When a sample contains proteins with a wide range of M_r values, a gradient polyacrylamide gel (e.g., 5–20% acrylamide) can be used.

The determination of M_r of glycoproteins and lipoproteins by SDS–PAGE can often result in erroneous M_r estimates due to the different binding characteristics of SDS to carbohydrate and lipid moieties.

Native gel electrophoresis This electrophoretic technique is generally used when the biological activity of the sample to be detected is required. It is generally considered to be a 'gentle' method with a lower resolving power than SDS–PAGE. Separation is based on the native charge, size, and shape of the sample molecules and on the properties of the separating gel. Homogeneous or gradient gels can be used, and the sample must be soluble, as any aggregated or precipitated material will not enter the gel matrix. Identification of individual proteins within these gels can be difficult due to the complex separation mechanisms.

Urea gel electrophoresis Urea gel electrophoresis is similar to SDS–PAGE in that it uses a detergent, 6–8 M urea, and a reducing agent, usually 2-mercaptoethanol, to ensure sample solubilization and disulfide bond disruption. In this method, however, the urea does not impart any charge on to the individual proteins, and separation is due to both charge and subunit size. In food analysis, the method is particularly useful for the separation of the individual casein proteins from the casein micelles in milk.

Isoelectric focusing Isoelectric focusing (IEF) is a high-resolution technique where proteins are separated according to their isoelectric points within a continuous pH gradient. The high resolving power allows the separation of compounds differing by only 0.01 pH units in p*I*. Separation is achieved by initially establishing a pH gradient along the gel using a mixture of low-M_r carrier ampholytes such as amino acids or small peptides. The sample components then travel through the pH gradient under an electric field until their individual charges reach zero, i.e., when they attain the pH corresponding to their isoelectric point. For this technique, a low polyacrylamide or agarose gel is used as it is only a stabilizing medium and does not participate in the sample separation. As urea carries no charge, samples can be run under either native or denaturing conditions (8 M urea).

Two-dimensional electrophoresis With the advent of proteomics, two-dimensional gel electrophoresis has become an indispensable tool with which to tease apart the individual proteins in cell systems. Proteins in a sample are initially separated according to their p*I* by isoelectric focusing (often in 8 M urea), and then a sample strip is applied at right angles along the edge of a SDS gel. Electrophoresis in the second direction is therefore according to subunit size, and the resulting pattern provides a unique 'fingerprint' of the sample.

Protein/peptide fixing and detection Proteins are immobilized in the gels after separation by fixing, usually with trichloroacetic, methanolic acetic or sulfosalicylic acid. Dyes such as Coomassie brilliant blue R-250, silver nitrate or Sypro Orange or Red are then used to visualize the protein bands.

Capillary Electrophoresis

Capillary electrophoresis (CE) includes a range of methods where samples are analyzed electrophoretically in capillary tubes. The capillaries may contain buffers, sieving gels, or more classic HPLC porous beads. As in HPLC, the numerous separation mechanisms available contribute to the versatility of CE with the different modes generally accessible simply by altering the buffer composition. Separation is based on differences in component migration in an electric field where the migration is a sum of the electric force on the component and the frictional drag through the medium. This is often highly dependent on the pH and composition of the running

buffer and is also affected by the electroosmotic flow (or bulk flow) of liquid in the capillary.

Capillary zone electrophoresis Fundamentally, capillary zone electrophoresis (CZE or free zone electrophoresis (FZE)) is the simplest CE format, principally because the capillary is only filled with buffer. Separation of both anionic and cationic components in a sample occurs when they migrate in discrete zones at different velocities due to electroosmotic flow. The selectivity of the separation can be altered using additives or coatings that alter the charge and hydrophobicity of the capillary wall.

CZE has a diverse range of applications for protein analysis such as purity validation, screening protein variants, and contaminants, and in conformational studies. It is also well suited for analysis of posttranslational modifications such as N- or C-terminal modifications, phosphorylations, carboxylations or N-glycosylations, and qualitative peptide mapping to detect subtle differences in proteins.

Micellar electrokinetic chromatography Micellar electrokinetic chromatography (MEKC) is a hybrid of electrophoresis and chromatography which uses surfactants in the running buffer, thus making the separation of neutral species possible. The surfactant molecules form micelles above their critical micelle concentration that can then interact with the neutral components. MEKC can be used for charged, uncharged, hydrophilic, and hydrophobic compounds such as peptides and amino acids.

Capillary gel electrophoresis The size-based separation of macromolecules such as proteins and nucleic acids is affected in capillary gel electrophoresis (CGE) by the electrophoresis of the compounds through a polymer network that acts as 'molecular sieve.' It is directly comparable to PAGE and uses polymer matrices such as acrylamide, dextran, and agarose.

For protein chemistry applications, the separation of native and SDS-complexed proteins in either the reduced (by 2-mercaptoethanol) or nonreduced state has been achieved using either cross-linked polyacrylamide or linear dextran polymers.

Capillary isoelectric focusing Capillary isoelectric focusing (CIEF) is similar to IEF–PAGE and separates proteins and peptides according to their p*I* values. It is a 'high-resolution' technique with a resolution of 0.005 p*I* units and less. Ampholytes are used to form a pH gradient within the capillary, and the proteins to be separated migrate (or are focused) through the ampholyte medium until they become uncharged at their p*I* values.

The p*I* values of proteins can be determined using protein standards with known p*I* values. Protein isoforms, proteins whose separation by other methods can be problematic, such as immunoglobulins and hemoglobins, and dilute biological solutions have all been successfully analyzed by CIEF.

Capillary isotachophoresis This is a 'moving boundary' electrophoretic technique where the separated zones move at the same velocity sandwiched between two buffer systems (called the leading and terminating electrolytes). The method creates very sharp boundaries between the zones and is mainly used to separate anions or cations.

Whilst it has not been used extensively for proteins, it has great potential for separating components of very similar mobility such as removing small amounts of impurity in recombinant proteins.

Capillary electrochromatography Capillary electrochromatography (CEC) is a true hybrid of electrophoresis and chromatography. It uses capillaries packed with chromatographic material, and the separation is based on electrophoresis and the chromatographic distribution of the various components between the stationary solid phase and the liquid mobile phase. It can also be compounded by interactions between the sample components and the capillary walls.

In theory, CEC has similar applications to CZE and MEKC. In practice, however, macromolecules such as proteins have an increased surface adsorption to both the capillary wall and the chromatographic matrix that restricts its usefulness. The problems should not be as severe with peptides.

Liquid Chromatography

Liquid chromatography (LC) consists of a family of methods in which two phases are used to separate structurally different analytes in a mixture. For the separation and purification of proteins, a variant of LC, HPLC, is now routinely used. This method uses small-diameter, wide-pore, solid-phase packings, and several techniques are available based on differences in the protein properties of size, charge, and hydrophobicity. The methods all use a stationary solid phase immobilized in a column and a liquid mobile phase. The protein mixture is introduced on to the column via the mobile phase, separation occurs via interaction with the stationary phase, and the individual components are detected usually at 214 or 280 nm using the UV-absorptive properties of proteins.

Reversed-phase chromatography Separation of proteins by reversed-phase chromatography (RPC)

depends on the reversible adsorption/desorption of the proteins, which have varying degrees of hydrophobicity, to a hydrophobic stationary matrix. The nature of the hydrophobic binding interaction is thought to be the result of a favorable entropy effect in areas adjacent to the hydrophobic regions where there is a higher degree of organized water structure, although the actual mechanism has yet to be fully elucidated. The matrices are generally silica-based or synthetic organic polymers and contain covalently bound alkyl chains of different lengths such as *n*-octadecyl (C18), *n*-butyl (C4), or phenyl groups. The strength of the hydrophobic interaction increases as the alkyl chain size increases.

Proteins are initially introduced on to the chromatographic matrix in conditions that promote binding, e.g., 5% organic modifier (acetonitrile) containing a weak hydrophobic ion-pairing reagent such as trifluoroacetic acid (0.1%, pH $\sim$ 2), and then eluted with an aqueous gradient of the organic modifier (5–60% v/v). The resulting separations are generally quick, and the proteins are usually highly resolved.

RPC has only recently become the method of choice for the separation and characterization of proteins/peptides. This is because of concerns that the organic solvents, e.g., acetonitrile, routinely used for the mobile phase can lead to protein denaturation and loss of biological activity. Strategies such as using short columns of short alkyl chain length, e.g., C4, organic modifiers such as ethanol or isopropanol, and omitting trifluoroacetic acid have been developed to allay these apprehensions.

Hydrophobic interaction chromatography Separation of proteins by hydrophobic interaction chromatography (HIC) is similar to RPC and utilizes hydrophobic interactions between the proteins in a mixture and the hydrophobic ligands (usually butyl, octyl, or phenyl groups) attached to an immobile matrix. HIC is different to RPC, however, in its use of aqueous mobile phases of high ionic strength and neutral pH that do not denature proteins. The ligands in HIC are also less hydrophobic and less densely bound to the solid matrix than in RPC matrices.

Proteins are bound to the column in a mobile phase that contains a high salt concentration, e.g., 1–3 M ammonium sulfate, and then eluted with a descending salt gradient that resolvates the proteins and makes binding thermodynamically unfavorable. This differs from RPC, where an organic solvent gradient is used to neutralize hydrophobic binding interactions. The protein-binding selectivity and strength can be manipulated by the mobile-phase pH and temperature, and elution of strongly adsorbed hydrophobic proteins can be achieved using ethylene glycol to lower the polarity of the mobile phase.

Affinity chromatography Affinity chromatography uses unique aspects of the biological or individual chemical structure of a protein to affect its purification. By selecting an interacting ligand which has a high natural specificity to the target protein, highly selective separations can be achieved. The protein to be purified or quantified is specifically and reversibly adsorbed by the complementary ligand immobilized on an insoluble matrix. Purification factors of up to several thousand can be achieved, together with further concentration during the elution step and high recoveries of the initial active protein.

Whilst the method is now routinely used for the purification of monoclonal antibodies and their corresponding antigens, receptor proteins, DNA-binding proteins, and 'tagged' recombinant proteins, its use in protein analysis is less prevalent.

Immobilized metal affinity chromatography Immobilized metal affinity chromatography (IMAC) is a specialized variant of affinity chromatography where the proteins or peptides are separated according to their affinity for metal ions that have been immobilized by chelation to an insoluble matrix. At pH values around neutral, the amino acids histidine, tryptophan, and cysteine form complexes with the chelated metal ions (e.g., Zn^{2+}, Cu^{2+}, Cd^{2+}, Hg^{2+}, Co^{2+}, Ni^{2+}, and Fe^{2+}). They can then be eluted by reducing the pH, increasing the mobile phase ionic strength, or adding ethylenediaminetetraacetic acid (0.05 M) to the mobile phase. This technique is especially suited for purifying recombinant proteins as poly-histidine fusions and for membrane proteins and protein aggregates where detergents or high-ionic-strength buffers are required.

Size-exclusion chromatography Size exclusion (or gel filtration) chromatography (SEC or GFC) separates proteins in solution according to differences in their size and shape as they travel through a solid phase (gel) matrix. As most proteins are spherical, their shape, or radius of gyration, will approximate to their M_r. The gel matrices are usually cross-linked three-dimensional polymers of dextran, agarose, or silica that contain pores of controlled size ranges to allow free transfer of the liquid phase and limited diffusion of the proteins to be separated. By using matrices with pores of carefully controlled sizes, different M_r ranges of proteins can be separated.

Separation of a mixture of proteins occurs when the passage of the smaller proteins, peptides, and

other molecules through the column is retarded by their diffusion into the pores of the gel beads. In comparison, the large proteins do not diffuse into the pores, effectively only 'see' a smaller volume of the mobile phase, and are said to elute with the 'solvent front.' Hence, the proteins are eluted from the column according to their size, with the large proteins eluting earlier. The M_r of the eluting proteins can be calculated by using suitable reference proteins.

The gel matrix and buffer systems are designed to minimize any adsorption of the protein to the gel matrix (e.g., 20 mM sodium phosphate, pH 7 or 50–100 mM NaCl for silica matrices). Denaturants such as 0.1% SDS or 6 M guanidine hydrochloride can also be added to the mobile phase to disrupt aggregates, promote the formation of uniform, rod-like conformations, and reduce interactions between the sample and the column matrix.

Ion-exchange chromatography Separation of proteins by ion-exchange chromatography (IEXC) is based on the interaction between proteins containing different charged groups and the oppositely charged moieties covalently linked to a chromatographic matrix. Thus, negatively charged proteins (i.e., in a mobile phase with a pH greater than the isoelectric point of the proteins) will separate on a positively charged solid phase (such as diethylaminoethyl groups) using anion exchange. The converse situation occurs for positively charged proteins usually using a negatively charged carboxymethyl ion exchanger and a pH 5–6 buffer. The technique is very powerful and can separate two proteins differing by only one charged amino acid.

Proteins are initially introduced on to the column under conditions that encourage binding. They are then sequentially eluted from the matrix usually by increasing the ionic strength of the eluting buffer by including a salt gradient or by changing the pH of the eluting buffer. Separation is due to different proteins having different degrees of interaction with the ion exchanger because of differences in their charges, charge densities, and distribution of charges on their surfaces.

Immunoassays

Immunochemical techniques require an antibody (or immunoglobulin) to recognize and bind to a three-dimensional site on an antigen giving rise to a measurable secondary effect. This reaction is of necessity very specific for the given antigen and can be monitored by a number of detection systems, including precipitation of the antibody–antigen complex, agglutination, or the use of labels, e.g., radioactive, enzymatic, or fluorescent.

Marker-free immunoassays These techniques rely on the ability to measure the antigen–antibody complex 'physically' and include turbidity, nephelometry, and immunodiffusion. The first two methods rely on differences in the light-scattering properties of the antigen–antibody complex and are not generally considered to be as sensitive or accurate as other immunoassays. This may change with the introduction of surface plasmon resonance that detects changes in the refractive index of the reaction solution close to a sensor surface and is sensitive down to the picomole level.

Immunodiffusion covers a number of formats where antigen and an excess of antibody are allowed to react as they diffuse through a permeable gel. The end product of the reaction is an immunoprecipitate that is 'trapped' in the gel.

In single radial immunodiffusion (RID), the antigen is added to a well in the gel and diffuses radially through the gel that contains the antibody. Initially, the antigen is in excess of the antibody, resulting in soluble complexes. The antigen diffuses further until the antibody concentration is greater than the antigen, and a precipitation ring forms.

In the double diffusion (Ouchterlony) immunoassay, the antigen and antibody are applied to separate wells and diffuse towards each other until a precipitation line is formed. This method can be used to compare the specificity of an antibody for different antigens.

Immunoelectrophoresis combines electrophoresis, initially to separate the antigens (proteins), and then gel immunodiffusion, to visualize them by precipitation with specific antibodies. A variation of this method, rocket immunoelectrophoresis, permits quantification of the antigen by performing the electrophoesis in a gel that contains antibody. A precipitation line in the shape of a 'rocket' is formed whose peak height depends on the concentration of the antigen.

Labeled-reagent immunoassays Immunoassay sensitivity can be increased by several orders of magnitude by using labeled reagents instead of marker-free immunoassays. The label can be an enzyme, a radioactive antigen or antibody, or a fluorescent chromogen. There are a large number of methods available based on either competitive or noncompetitive formats. In competitive assays, free antigen competes with labeled or solid-phase bound antigen for a limited amount of antibody. In noncompetitive methods, the amount of antigen is limited and is

bound by specific labeled antibodies. Various formats such as microtiter plates, dipsticks, immunodots, and immunofiltration devices can be used.

Other Technologies

Protein bioassays When proteins or peptides have specific functions, e.g., enzymes, folate-binding protein, or biologically active peptides, individual bioassays are often used to determine their bioactivity. These assays are often more informative to the researcher than the more physical data. For proteins (and enzymes), this activity is usually also indicative of the conformational state of the protein with protein denaturation correlating to loss of activity.

Mass spectrometry Recent developments in mass spectrometry, such as electrospray ionization and matrix-assisted laser desorption ionization, have made it possible to determine the M_r of proteins and peptides that weigh in excess of 100 000 Da and to analyze minute amounts of these compounds down to the femtomole level. This has resulted in mass spectrometry becoming a common adjunct to the UV adsorption detector for the detection and characterization of proteins and peptides by HPLC and CE.

See also: **Analysis of Food**; **Biosensors**; **Chromatography**: Combined Chromatography and Mass Spectrometry; Principles; High-performance Liquid Chromatography; **Enzymes**: Uses in Analysis; **Immunoassays**: Principles; Radioimmunoassay and Enzyme Immunoassay; **Mass Spectrometry**: Principles and Instrumentation; Applications; **Protein**: Chemistry; Food Sources; Requirements; Functional Properties

Further Reading

Andrews AT (1986) *Electrophoresis: Theory, Techniques, and Biochemical and Clinical Applications*, 2nd edn. Oxford: Oxford University Press.

Ashurst PR and Dennis MJ (eds) (1998) *Analytical Methods of Food Authentication*. London: Blackie Academic & Professional.

Hames BD and Rickwood D (eds) (1990) *Gel Electrophoresis of Proteins: A Practical Approach*, 2nd edn. Oxford: IRL Press.

Oliver RWA (ed.) (1998) *HPLC of Macromolecules: A Practical Approach*, 2nd edn. Oxford: IRL Press.

Pomeranz Y and Meloan CE (1994) *Food Analysis: Theory and Practice*. New York: Chapman & Hall.

Scopes RG (1993) *Protein Purification: Principles and Practice*, 3rd edn. New York: Springer-Verlag.

Sorensen H, Sorensen S, Bjergegaard C and Michaelsen S (1999) *Chromatography and Capillary Electrophoresis in Food Analysis*. Cambridge: The Royal Society of Chemistry.

Requirements

A E Bender[†], Fetcham, Leatherhead, Surrey, UK

Introduction

The protein requirement is defined as the lowest level of dietary protein that will balance the losses of nitrogen (N) in persons maintaining energy balance. The body proteins are constantly undergoing breakdown and resynthesis at rates that vary from tissue to tissue. The amino acids released by tissue breakdown are mostly reused for synthesis but there is some loss – the obligatory nitrogen loss – by oxidative catabolism.

It is necessary to specify that the subject is in energy balance because utilization of dietary protein is influenced by energy intake. It has been shown that at any given level of dietary protein, the addition of energy improves nitrogen balance until it reaches a plateau, which represents the limitation imposed by protein. Any change above or below the energy needs of the subject is likely to influence nitrogen balance. This is why the current international recommendations consider energy and protein requirements together. (*See* **Energy:** Measurement of Food Energy; Energy Expenditure and Energy Balance.)

If more protein is consumed than is physiologically required, the excess is metabolized as a source of energy. However, there has been some concern that excessive protein intakes may contribute to demineralization of bone and deterioration of renal function in patients with renal disease; it is therefore suggested that the upper limit should be 1.5 g of protein per kg of body weight per day.

Earlier Recommendations

According to the perceived wisdom of the day, figures for requirements for protein have varied enormously (**Table 1**). In the early days such figures were effectively pronouncements by leading physiologists and it was not until the discussions of the Food and Agriculture Organization (FAO) of the United Nations that figures were based on scientific evidence. At intervals of about 10 years, namely 1957, 1965, 1973 and 1985, revised FAO recommendations have been made in the light of increasing knowledge.

The first of these reports, in 1957, emphasized the proportions of the various amino acids required and

[†] Deceased.

Table 1 History of recommended protein intakes (adult man)

Author	*Recommended intake*
Playfair (1865)	57–184 g (observed)
Voit (1881)	118 g (recommended)
Chittenden (1905)	50–55 g (observed and recommended)
Sherman (1920)	33–40 g at 70 kg body weight (0.59 g per kg)
League of Nations (1936)	1 g per kg body weight
FAO (1957)	0.35 g kg^{-1}
FAO (1965)	0.71 g kg^{-1}
FAO (1973)	0.57 g kg^{-1}
FAO (1985)	0.86 g kg^{-1}

FAO, Food and Agriculture Organization.

they quantified dietary protein requirements in terms of a theoretical reference protein of perfect amino acid composition, i.e., satisfying the experimentally determined needs for each amino acid. The reason for terming this a theoretical protein is that dietary protein is most effectively assimilated only at low levels of intake (as used in the laboratory assessment of protein quality) and efficiency falls off as the level rises. The efficiency of utilization of the theoretical protein is constant at all levels of intake.

National Recommended Dietary Allowances (RDAs)

While the FAO reports represent the conclusions of expert committees, each national authority has compiled its own figures based on the opinions of the committee involved and taking national habits into account. This has led to different figures in different countries. Thus, the RDA for protein for adult males is 81 g day^{-1} in France, 64 g day^{-1} in Germany, 70 g day^{-1} in the Netherlands, and 54 g day^{-1} in Spain. In practical food terms, the amount of protein one would recommend in the diet must depend on the total food (energy) intake; otherwise, heavy work would apparently call for provision of the extra energy from pure carbohydrate and fat. Two examples of the practical food approach that obviate this difficulty are the figure of 10–15% of the total energy intake as protein recommended by the Scandinavian countries, and 10% (69 g of protein on average) in the 1979 UK tables.

The 1991 UK revision (Table 2) adopted the FAO/World Health Organization (WHO)/United Nations University (UNU) figures. For adult males aged 19–50 years this is listed as 42.6 g of protein for males weighing 74 kg as the estimated average requirement, with 55.5 g as the reference nutrient intake – the term introduced in the 1991 revision and which is equivalent to RDA. For women weighing 60 kg, the corresponding figures are 36 and 45 g day^{-1}. The figures are based on milk or egg protein which are taken as completely digestible. For diets based on unrefined cereal grains and vegetables, a correction factor for digestibility of 85% is applied; for diets based on refined cereals, the correction factor is 95%. (*See* **Dietary Reference Values**.)

Table 2 UK dietary reference values for protein

Age	*Weight (kg)*	*Estimated average requirement (g day^{-1})*	*Reference intake (g day^{-1})*
0–3 months	5.9		12.5
4–6 months	7.7	10.6	12.7
7–9 months	8.8	11.0	13.7
10–12 months	9.7	11.2	14.9
1–3 years	12.5	11.7	14.5
4–6 years	17.8	14.8	19.7
7–10 years	28.3	22.8	28.3
Males			
11–14 years	43.0	33.8	42.1
15–18 years	64.5	46.1	55.2
19–50 years	74.0	44.4	55.5
50+ years	71.0	42.6	53.3
Females			
11–14 years	43.8	33.1	41.2
15–18 years	55.5	37.1	45.4
19–50 years	60.0	36.0	45.0
50+ years	62.0	37.2	46.5
Pregnancy			+6
Lactation			
0–6 months			+11
6+ months			+8

Reproduced from Department of Health (1991) *Dietary Reference Values for Food Energy and Nutrients for the United Kingdom*. Report on Health and Social Subjects 41. London: Her Majesty's Stationery Office.

Methods of Assessment of Needs

There are two methods of determining nitrogen requirements. The one used until the 1985 FAO report was the factorial method in which the losses of nitrogen from the body are measured to establish the amount that is required. The second method, adopted in the 1985 report, measures the intake of nitrogen needed to maintain nitrogen equilibrium.

Factorial Method

On a diet free from protein the body loses nitrogen in the urine as urea, creatinine, uric acid, and ammonium salts, together with small amounts of other compounds; this is termed the endogenous loss. There are also losses in the feces from bacteria, mucosal cells shed from the lining of the intestine, and residues of digestive juices that have not been reabsorbed; this is termed metabolic loss.

In addition, there are small losses from skin cells, hair, finger nails, and body fluids.

These obligatory losses are measured on subjects fed a diet free from protein. On such a diet the urinary output of nitrogen does not attain a constant level but falls sharply for 4–6 days and then very slowly. The levels reached are 2–3 g day^{-1} or 30–45 mg per kg of body weight. Forty-five milligrams of nitrogen approximates to 2 mg of nitrogen per basal kcal. The two phases of this decline are taken to indicate two processes. The first rapid fall is loss from organs with a rapid protein turnover and is called labile body protein. The second is largely from muscles and skin. The obligatory nitrogen loss is taken as the figure at the end of the rapid fall. This relative inexactitude is one of the disadvantages of the factorial method.

The 1965 report took the figures of 46 mg nitrogen per kg (2 mg per basal kcal) for urine loss, 20 mg kg^{-1} for fecal loss, and 20 mg for other losses. In the light of better evidence, the 1973 report changed these findings to 37, 12, and 5 mg respectively. This accounts for the marked reduction in protein recommendations between 1965 (0.71 g kg^{-1}) and 1973 (0.57 g kg^{-1}). At first consideration, obligatory losses should be balanced by an intake of the same amount. This was assumed in the 1965 report but it was later realized that when an individual is supplied with the same amount of nitrogen as his or her obligatory loss, he or she remains in negative nitrogen balance. It was found necessary to increase that figure by 30% to achieve nitrogen equilibrium. An additional 30% was added to allow for individual variation. All measurements are made as nitrogen and then converted into protein by multiplying by the factor 6.25.

While the basic data show a small difference for nitrogen losses between males and females, figures are not sufficiently firm to allow any distinction. The difference between final protein figures is due to the different standard body weights used for males and females; in the 1973 report this was 0.71 g of protein per kg of body weight per day, which is 39 g for females and 46 g for males. This was called the practical allowance and based on full utilization of the protein in the diet, i.e., high-quality protein, termed reference protein.

An extra allowance is made for protein of lower quality, i.e., protein with net protein utilization (NPU) values of 70–80 in developed countries and 60–70 in less developed countries (with values as low as 50–60 in special conditions); this is stated in the 1973 report.

Apart from the difficulty of deciding the level of the endogenous loss (since there is no clear sharp break in the curve), most workers did not measure the cutaneous losses because of the difficulty of doing so, and they accepted published values which were derived from only a limited number of experiments.

Nitrogen Balance Method

Nitrogen balance is the difference between nitrogen excreted from the body and nitrogen ingested in the diet (of which the greater part by far is protein). During growth, pregnancy, lactation, and recovery from convalescence, the body is in positive nitrogen balance since it is retaining nitrogen for the purpose of synthesizing new protein tissues. During dietary deprivation, most illnesses, and certain types of stress, the body loses nitrogen and is in negative balance. The healthy adult is in nitrogen equilibrium. The basis of this method of determining nitrogen requirements is to feed the subjects a series of diets with different levels of protein while measuring nitrogen excretion, then to interpolate to nitrogen equilibrium (zero nitrogen balance).

In early studies the diets included very low levels of protein and a zero level but since the nitrogen balance response is not linear throughout the entire submaintenance range, recent studies use levels around the expected range of requirements. An allowance has to be made for the variable losses of nitrogen in sweat, which are considerable in heavy work in a hot climate.

It takes some time at a given level of dietary protein to achieve a steady state, i.e., adjustments in urine output do not immediately follow changes in nitrogen intake, so that diets are fed for periods of 1–3 weeks at each intake level. These short-term nitrogen balance determinations do not take account of adaptation to low levels of dietary protein as experienced in developing countries. Long-term studies require several months, but these are usually limited to a single level of protein intake which would provide evidence that a particular diet is adequate.

It must be borne in mind that in these detailed consultations about protein (and energy) requirements and recommendations, the FAO is largely concerned with the adequacy of diets in poorly fed populations of developing countries. In the well-fed western world there is never a problem of protein shortage in healthy people as long as enough food is consumed to satisfy hunger. The fundamental physiological considerations, of course, still apply.

Direct nitrogen balance studies are the currently accepted method of determining protein requirements but they suffer from the lack of long-term balance studies, the absence of independent validation of an optimal state of protein nutrition, and lack of knowledge of the functional significance of

the size of the total nitrogen pool and rate of turnover of tissue proteins. There are no functional indicators of protein inadequacy at a stage before clinically detectable changes occur.

The limitations of the data are indicated by the small numbers of studies at the time of the 1985 report – nine short-term studies of single protein sources on a total of 93 subjects, eight short-term studies on the typical mixed diets of eight countries on a total of 73 subjects, and six long-term studies, 24–89 days, five on egg and one on milk, on a total of 34 subjects.

The short-term studies provide an estimate of a mean daily requirement of 0.63 g of highly digestible good-quality protein – a figure slightly higher than the 1973 'safe level.' The long-term studies suggest that 0.58 g kg^{-1} is a reasonable estimate.

From all the available results it is suggested that 0.6 g per kg of body weight per day is the average requirement for good-quality proteins such as meat, milk, egg, and fish.

The coefficient of variation was 12.5%, i.e., 2 SD equals 25%. This provides the currently accepted figure of 0.75 g of protein per kg of body weight per day as the safe level of intake for an adult.

Table 4 Safe level of protein intake (good-quality protein)

Age	*g per kg body weight*	*g per day*
3–6 months	1.85	13
9–12 months	1.50	14
1–2 years	1.2	13.5
2–3 years	1.15	15.5
3–5 years	1.10	17.5
5–7 years	1.0	21
7–10 years	1.0	27
10–12 years		
Male	1.0	34
Female	1.0	36
12–14 years		
Male	1.0	43
Female	0.95	44
14–16 years		
Male	0.95	52
Female	0.9	46
16–18 years		
Male	0.9	56
Female	0.8	42
Adult		
Male	0.75	49 (65 kg body weight)
Female	0.75	41 (55 kg body weight)
Pregnancy		+6
Lactation		+17.5

Reproduced from WHO (1985) *Energy and Protein Requirements. Report of the Joint FAO/WHO/UNU Expert Consultation*. WHO Technical Report Series 724. Geneva: World Health Organization.

Table 3 Values for the digestibility of protein in humans

Protein source	*True digestibility (mean ± SD)*	*Digestibility relative to reference proteins*
Egg	97 ± 3	
Milk, cheese	95 ± 3	100
Meat, fish	94 ± 3	
Maize	85 ± 6	89
Rice, polished	88 ± 4	93
Wheat, whole	86 ± 5	90
Wheat, refined	96 ± 4	101
Oatmeal	86 ± 7	90
Millet	79	83
Peas, mature	88	93
Peanut butter	95	100
Soya flour	86 ± 7	90
Beans	78	82
Maize plus beans	78	82
Maize plus beans plus milk	84	88
Indian rice diet	77	81
Indian rice diet plus milk	87	92
Chinese mixed diet	96	98
Brazilian mixed diet	78	82
Filipino mixed diet	88	93
American mixed diet	96	101
Indian rice plus beans diet	78	82

Reproduced from WHO (1985) *Energy and Protein Requirements. Report of the Joint FAO/WHO/UNU Expert Consultation*. WHO Technical Report Series 724. Geneva: World Health Organization.

Table 5 Protein losses related to surgery

Time of loss	*Protein loss (g)*
Before surgery	
Bleeding peptic ulcer	90 (over 5 days)
Long-bone fracture	140–190 (over 10 weeks)
Limb burn	Up to 1000
During surgery (blood loss)	
Gastric surgery	9–18
Pneumonectomy	20–60
Thyroidectomy	3–12
Catabolic response after surgery	
Herniorrhaphy	18 (over 10 days)
Gastric resection	50–175 (5–10 days)
Appendectomy	50 (10 days)
Bed rest (inactivity), 3–4 weeks	300

Quality and Digestibility

Earlier recommendations for protein requirements were based on 'good-quality' protein or reference proteins such as eggs or milk, i.e., proteins with NPU close to 1.0 (100 in the older nonmenclature). Thus, the recommendation of 0.75 g protein per kg of body weight is for protein of this high quality and the amount has to be increased proportionately to compensate for lower quality. For example, 0.75 g of NPU

Table 6 Approximate protein loss (g) during first 10 days following injury or operation

	Loss of tissue	*Hemorrhage or exudate*	*Catabolic phase*	*Total*
Simple fracture of femur		200	700	900–1100
Muscle wound	500–750	150–400	750	1350–1900
35% Burn	500	150–400	750	1400–1650
Gastrectomy	(20–180)	20–100	625–750	645–850
Untreated typhoid fever			675	675

1.0 is equivalent to 1.07 g of quality 0.7, or 1.25 g of NPU 0.6.

In practice, the quality of the mixture of dietary proteins in underdeveloped countries (which rely very largely on a single staple food, usually a cereal), is about 0.7, and the addition of protein-rich foods such as meat and milk – cereals, meat, and milk are the major sources of dietary protein in the developed countries – increases this to only 0.8. Consequently, the 1985 report lays less emphasis on the quality of the protein because the figure recommended is higher than the 1973 figure and digestibility correction alone is adequate. Digestibility was not mentioned at that time.

Only four amino acids are likely to limit the protein quality of mixed diets: they are lysine, the sulfur amino acids, threonine, and tryptophan, and diets in developing countries generally meet the adult requirement for these amino acids. Quality can be taken care of by calculating the proportion of the most limiting amino acid compared with that in the requirement pattern and then increasing the protein figure in that ratio. (*See* **Amino Acids**: Metabolism.)

Differences in digestibility arise from the nature of the cell walls of the food, the presence of dietary fiber and polyphenols, and from chemical reactions that alter the release of amino acids during digestion. Adjustments are therefore made to the amount of food protein to be ingested by comparing the digestibility of the protein food with that of reference proteins (Table 3).

The digestibility of a complete diet can be calculated by using the values for individual sources of protein and multiplying by the proportion of those foods in the diet.

Diets based on coarse, whole-grain cereals and vegetables are generally about 85% digested, while those based on refined cereals are 95% digested; meat, milk, and eggs are 100% digested.

Extra Allowances

Figures for protein requirements are based on measurements made on young adults with corrections for age, sex, and body weight. They apply to healthy adults in nitrogen equilibrium. There are obviously greater needs for growing children, pregnant women, and lactating mothers, and these are shown in Table 4 (the UK figure for lactation, in Table 2, has been modified). The extra needs for growth are small compared with the nitrogen required to maintain the tissues, even in young children, except for babies under 1 year of age. For example, the average requirement for growth of boys 4–5 years of age is 0.36 g of protein per day, while for maintenance it is 12.6 g. It is accepted that different amounts of protein may be laid down from day to day as part of the normal process of growth. The body has very limited capacity for storing amino acids; it is therefore necessary to provide enough protein every day for possible extra demands of higher rates of growth. For this reason the extra demands for growth were estimated at 50% greater than the theoretical growth rate. (*See* **Children**: Nutritional Requirements; **Lactation**: Human Milk: Composition and Nutritional Value; Physiology; **Pregnancy**: Safe Diet.)

Trauma

Protein requirements for healthy adults take into account the 'normal stresses of everyday life' but there are increased losses of nitrogen from the body in pain, anxiety, and psychological stresses, with very large losses in trauma, chronic disorders, and parasitic infections. Clearly, these will vary enormously with the severity and duration of the condition. Tables 5 and 6 show some examples of these losses which must be made good during convalescence.

See also: **Amino Acids**: Metabolism; **Children**: Nutritional Requirements; **Dietary Reference Values**; **Energy**: Energy Expenditure and Energy Balance; Measurement of Food Energy; **Lactation**: Human Milk: Composition and Nutritional Value; Physiology; **Pregnancy**: Safe Diet

Further Reading

Department of Health (1991) *Dietary Reference Values for Food Energy and Nutrients for the United Kingdom*. Report on Health and Social Subjects 41. London: Her Majesty's Stationery Office.

FAO (1957) Protein requirements. *FAO Nutritional Studies*. 16. Rome: Food and Agriculture Organization.
FAO (1965) Protein requirements. Report of the FAO/WHO Expert Committee. *FAO Nutrition Report Series* 37. Rome: Food and Agriculture Organization.
FAO (1973) Energy and protein requirements. Report of the Joint FAO/WHO Expert Committee. *FAO Nutrition Report Series* 52. Rome: Food and Agriculture Organization.
WHO (1985) Energy and protein requirements. Report of the Joint FAO/WHO/UNU Expert Consultation. *WHO Technical Report Series* 724. Geneva: World Health Organization.

Functional Properties

J K P Weder and H-D Belitz†, Technische Universität München, Garching, Germany

Introduction

Standardization of food properties, in order to meet nutritional, physiological and toxicological demands, and requirements of food-processing operations, is a perennial endeavor. Food production is similar to a standard industrial fabrication process: on the one hand is the food commodity with all its required properties; on the other hand are the components of the product, each of which supplies a distinct part of the required properties. Such considerations have prompted investigations into the relationship in food between macroscopic physical and chemical properties and the structure and reactions at the molecular level. Reliable understanding of such relationships is a fundamental prerequisite for the design and operation of a process, either to optimize the process or to modify the food components to meet the desired properties of the product.

Functional Properties

In common with other food constituents, proteins contribute significantly to the physical properties of foodstuffs, especially through their ability to build or stabilize gels, foams, doughs, emulsions and fibrillar structures. **Table 1** shows some typical examples of functional properties of proteins in relation to important food systems. Foaming, gelling, and emulsifying properties will be discussed in more detail.

† Deceased.

Foaming Properties

Proteins act as foam-forming and foam-stabilizing agents in various foodstuffs, e.g., in baked products, sweets, desserts, and beer. The properties of various proteins are different: serum albumin foams very well, whereas ovalbumin does not. Mixtures of proteins, e.g., egg white, can be particularly effective. In that case, the globulins start the formation of foam; ovomucin is important for its stabilization, and ovalbumin and conalbumin are responsible for the heat-setting properties.

Foams are dispersions of gases in liquids. Proteins stabilize such systems by forming flexible, cohesive films around the surface of the gas bubbles. During whipping, the protein is adsorbed at the interface via hydrophobic areas, followed by partial unfolding (surface denaturation). The decrease of the surface tension, caused by protein adsorption, facilitates the formation of new interfaces and further gas bubbles. The partially unfolded proteins associate under film formation.

The foamability of a protein molecule depends on its diffusion rate and the ease with which it is denatured. These parameters in turn depend on the molecular mass, the surface hydrophobicity, and the stability of the conformation.

Foams collapse because large gas bubbles grow at the expense of smaller ones ('*disproportionation*'). Therefore, the stability of a foam depends on the stability of the protein film and its permeability for gases. The stability of the film in turn depends on the amount of adsorbed protein and on the ability of the adsorbed protein to associate. Generally, surface denaturation exposes additional amino acid side-chains that can participate in intermolecular interactions. The stronger the cross-linking, the more stable is the film. The pH of the system should be near the isoelectric points of the involved proteins because the association is promoted by a small net charge.

In summary, the ideal foam-forming and foam-stabilizing protein is characterized by a low molecular mass, high surface hydrophobicity, good solubility, a small net charge at the pH of the food, and easy denaturability.

Foams are destroyed by lipids and organic solvents such as higher alcohols, which, because of their hydrophobicity, displace proteins from the surface of the gas bubbles without being able to form stable films. Egg yolk, for example, prevents the whipping of egg white, even at low concentrations. The disturbance of protein association by the lecithins is responsible for that effect.

The foaming properties of proteins can be improved by chemical and physical modification. Partial

Table 1 Typical functional properties caused by proteins in food systems

Functional property	*Mode of action*	*Food system*
Solubility	Protein solvation, pH-dependent	Beverages
Water absorption and binding	Hydrogen bonding of water, entrapment of water (no drip)	Meat, sausages, bread, cakes
Viscosity	Thickening, water binding	Soups, gravies
Gelation	Protein matrix formation and setting	Meat, curds, cheese
Cohesion–adhesion	Protein acts as adhesive material	Meat, sausages, baked goods, pasta products
Elasticity	Hydrophobic bonding in gluten, disulfide bridges in gels (deformable)	Meat, bakery
Emulsification	Formation and stabilization of fat emulsions	Sausages (e.g., Bologna), soup, cakes
Fat adsorption	Binding of free fat	Meat, sausages, doughnuts
Flavor binding	Adsorption, entrapment, release	Simulated meat, bakery, etc.
Foaming	Formation of stable films to entrap gas	Whipped toppings, chiffon desserts, angel cakes

Modified from Kinsella and Srinivasan (alias Damodaran S) (1981).

enzymatic hydrolysis produces smaller molecules of a higher diffusion rate, better solubility, and higher surface hydrophobicity. The disadvantages are the lower film stability and the loss of heat coagulability. The introduction of charged or neutral groups or a partial thermal denaturation (e.g., of whey proteins) can also improve the desired properties. Recently, the addition of strongly basic proteins (e.g., clupeines) has been tested, which obviously increases the protein association within the films and allows the foaming of lipid-containing systems.

Emulsifying Properties

Emulsions are disperse systems of two or more immiscible liquids. They are stabilized by emulsifiers, compounds that form interfacial films and prevent the dispersed phase from coalescing. Proteins are able to stabilize emulsions (e.g., milk) due to their amphipathic nature. The emulsifying properties of a protein depend on the rate at which it diffuses into the interface, on its adsorbability there, and on the deformability of its conformation through interfacial tension (surface denaturation). The diffusion rate depends on the temperature and the molecular mass of the protein, which in turn can be influenced by the pH and the ionic strength. The adsorbability depends on the exposure of hydrophilic and hydrophobic groups, and thus on the amino acid composition, as well as on the pH, ionic strength, and temperature. The stability of the conformation depends on the molecular mass, amino acid profile and number of intramolecular disulfide bonds of the protein.

Thus, a protein with ideal qualities as an emulsifier for oil-in-water emulsions would have a relatively low molecular mass, a balanced amino acid composition in terms of charged, polar and apolar side-chains, good solubility in water, marked surface hydrophobicity, and a relatively stable conformation.

Gel Formation

Gels are disperse systems of at least two components in which a solid phase (dispersed phase) forms a cohesive network in a liquid phase (continuous phase). Gels are characterized by their lack of fluidity and their elastic deformability. They are placed between solutions with repulsive forces between molecules of the dispersed phase predominating, and precipitates with strong intermolecular forces predominating. A differentiation is made between two types of gels, the '*polymeric networks*' and the '*aggregated dispersions*.' Intermediate forms are also known.

Examples of *polymeric networks* are the gels formed by gelatin or by polysaccharides such as agarose or carrageenan. The formation of a three-dimensional network takes place by aggregation of unordered fibrous molecules via limited ordered structures, e.g., double helices. Gels of this type are characterized by a low concentration of polymer (*c*. 1%), transparency and fine texture. Gel formation is started by adjusting to a suitable pH, by adding suitable ions, or by heating followed by cooling. Since the aggregation takes place mainly via hydrogen bonds, which are easily solved when heated, polymeric networks are '*thermo-reversible*,' i.e., they are formed when a solution is cooled, and they melt again when they are heated.

Examples of *aggregated dispersions* are the gels formed by globular proteins after denaturation by heat. The thermal unfolding of the protein exposes amino acid side-chains, which can take part in intermolecular interactions. The subsequent association leads to small spherical aggregates, which combine into linear strands whose interaction results in the three-dimensional gel network. Because of the unordered type of aggregation, gel formation requires a relatively high protein concentration (5–10%). To

avoid the formation of coarse and fairly unstructured gels, in particular in the area of the isoelectric point, the aggregation rate should be slower than the unfolding rate. The degree of unfolding necessary to start aggregation seems to depend on the particular protein. Since partial denaturation exposes mainly hydrophobic groups, the aggregation takes place predominantly via intermolecular hydrophobic bonds, and '*thermoplastic*' *(thermo-irreversible)* gels are formed. This type of gel does not liquefy when heated, but can soften or shrink. Besides hydrophobic bonds, disulfide bonds formed from exposed thiol groups can also contribute to cross-linking, as can electrostatic bonds between proteins with different isoelectric points in heterogeneous systems (e.g., egg white).

Addition of salts improves the gel formation. The moderate increase in ionic strength promotes interaction between charged macromolecules or molecule aggregates through charge shielding without precipitation. An example is the improvement of thermal gelling of soya globulins in soya milk by calcium ions (soya curd, tofu).

Modification of Functional Properties

Modification of proteins is still far from being a common method in food processing, but it is increasingly being recognized as essential, for two main reasons. First, proteins fulfill many functions in food, and some of these can be served better by modified than by native proteins. Second, persistent nutritional problems the world over necessitate the utilization of new raw materials. Modifying reactions can insure that such new raw materials (e.g., proteins of plant or microbial origin) meet stringent standards of food safety, palatability, and acceptable biological value. This review includes several protein modification reactions that are being used or are being considered for use. They involve chemical or enzymatic reactions or a combination of both. Examples have been selected to emphasize existing trends. Modification of the properties of proteins is possible by changing the amino acid composition or the size of the molecule, or by removing or inserting hetero constituents. Such changes can be accomplished by chemical and/or enzymatic reactions.

From a food processing point of view, the aims of protein modification are:

- blocking the reactions involved in deterioration of food (e.g., the *Maillard* reaction);
- improving some physical properties of proteins (e.g., texture, foam stability, whippability, solubility);
- improving the nutritional value (increasing the extent of digestibility, inactivation of toxic or other undesirable constituents, introducing essential ingredients such as some amino acids).

Chemical Modification

A selection of chemical reactions of proteins that are pertinent to, and of current importance in, food processing is shown in Table 2.

Acylation Treatment with succinic anhydride generally improves the solubility of a protein. For example, succinylated wheat gluten is quite soluble at pH 5. This effect is related to disaggregation of high-molecular-weight gluten fractions. In the case of succinylated casein, the modification shifts the isoelectric point of the protein (and thereby the solubility minimum) to a lower pH. Succinylation of leaf proteins improves the solubility as well as the flavor and the emulsifying properties. Succinylated yeast protein not only has an increased solubility in the pH range of 4–6 but is also more heat-stable above pH 5. It has better emulsifying properties, surpassing many other proteins, and has an increased whippability.

Table 2 Chemical reactions of proteins significant in food

Reactive group	*Reaction*	*Product*
—NH_2	Acylation	—NH—CO—R
—NH_2	Reductive alkylation with HCHO	— $N(CH_3)_2$
—NH_2	Reductive glycosylation	—NH—CH_2—$(CHOH)_n$—CH_2OH
—$CONH_2$	Hydrolysis	—COOH
—COOH	Esterification	—COOR
—COOH	Amidation ('glycosylation')	—CO—NH—CH(CHO)—$(CHOH)_n$H
—OH	Esterification	—O—CO—R
—SH	Oxidation	—S—S—
—SH	Alkylation	—S—R
—S—S—	Reduction	—SH
—CO—NH—	Hydrolysis	—COOH + H_2N—

Introduction of aminoacyl groups into protein can be achieved by reactions involving amino acid carboxy anhydrides, amino acids and carbodiimides, or by *t*-butyloxycarbonyl (BOC)-amino acid *N*-hydroxysuccinimide esters with subsequent removal of the amino-protecting group (BOC). Feeding tests with casein with attached methionine, as produced by this method, have demonstrated a satisfactory availability of methionine. Such covalent attachment of essential amino acids to a protein may avoid the problems associated with food supplementation with free amino acids: losses during processing, development of undesired aroma due to methional, etc. With β-casein as an example, it was shown that the association of a protein is significantly affected by its acylation with fatty acids of various chain lengths.

Alkylation Modification of proteins by reductive methylation of amino groups with formaldehyde and sodium borohydride ($NaBH_4$) retards *Maillard* reactions. The resultant methyl derivative, depending on the degree of methylation, is less accessible to proteolysis. Hence, its value from a nutritional and physiological point of view is still an open question.

Sugars such as glucose, fructose, maltose, or lactose can be bound to the amino groups of proteins (casein, legumin, β-lactoglobulin) by reductive glycosylation with sodium cyanoborohydride. Whereas neutral sugars favor foaming capacity and foam stability, charged carbohydrates improve emulsifying properties. Amino sugars can also be bound to the carboxyl group of proteins through an amide bond by the carbodiimide method.

Reactions involving cysteine and cystine Disulfide bonds exert a strong influence on the properties of proteins. Wheat gluten can be modified by reduction of its disulfide bonds to sulfhydryl groups and subsequent reoxidation of these groups under various conditions. Reoxidation of a diluted suspension in the presence of urea results in a weak, soluble, adhesive product (gluten A), whereas reoxidation of a concentrated suspension in the presence of a higher concentration of urea yields an insoluble, stiff, cohesive product (gluten C). Additional viscosity data have shown that the disulfide bridges in gluten A are mostly intramolecular, whereas those in gluten C are predominantly intermolecular.

Disulfide cleavage in soya protein by 0.1 M sodium sulfite produces an adhesive with an acceptable viscosity and improved adhesive strength and hydrophobicity compared with unmodified soya protein when spray-dried, but a very low viscosity when freeze-dried.

Enzymatic Modification

Of the great number of possible enzymatic reactions with protein as a substrate, only a small number have been found so far to be suitable for use in food processing.

Dephosphorylation With β-casein as an example, it was shown that the solubility of a phosphoprotein in the presence of calcium ions is greatly improved by enzymatic dephosphorylation.

Plastein reaction The enzyme-catalyzed formation of larger polypeptides (about 3 kDa) from peptide fragments of a hydrolysate, i.e., the formation of peptide bonds, is called the '*plastein* reaction.' Amino acid esters that are added to the hydrolysate are incorporated into the resulting products, which are called '*plasteins*.' Incorporation of amino acid esters into the polypeptide is affected by the nature of the amino acid side-chain and by the length of the alkyl chain of the ester. Amino acids with hydrophobic side-chains and long-chain alkyl esters are preferred. Thus, an amino acid with a short side-chain should be applied as an ester of a long-chain alcohol. The plastein reaction can help to improve the biological value of a protein. An example is the plastein enrichment of zein with tryptophan, threonine, and lysine.

Enrichment of a protein with selected amino acids can be achieved with the corresponding amino acid esters or, equally well, by using suitable partial hydrolysates of another protein. For example, the enrichment of soya protein with sulfur-containing amino acids is possible through '*adulteration*' with the partial hydrolysate of wool keratin. The protein efficiency ratio (PER) value of such a plastein is significantly improved. In this way, the production of plasteins with an amino acid profile very close to that recommended by the Food and Agriculture Organization and the World Health Organization (FAO/WHO) can be achieved from very diverse proteins, e.g., from leaf, bacterial and algal proteins.

The plastein reaction also makes it possible to improve the solubility of a protein, e.g., by increasing the content of glutamic acid. Soya protein (24% glutamic acid) yields a plastein with 25% glutamic acid and a Glu-plastein with 42% glutamic acid. Soya protein has a pronounced solubility minimum in the pH range of 3–6. The minimum is much less pronounced in the case of the plastein, and the glutamic acid-enriched Glu-plastein has a satisfactory solubility over the whole pH range and is also resistant to thermal coagulation. Proteins with an increased content of glutamic acid show an interesting sensory

effect: partial hydrolysates of such plasteins do not taste bitter, but exhibit a pronounced 'meat broth' flavor.

Elimination of the bitter taste from a protein hydrolysate is also possible without incorporation of hydrophilic amino acids. Bitter-tasting peptides, such as leucyl-phenylalanine, which are released by partial hydrolysis of proteins, react preferentially in the subsequent plastein reaction and are incorporated into higher-molecular-weight peptides with a neutral taste.

The versatility of the plastein reaction is also demonstrated by examples wherein undesired amino acids are removed from a protein. A phenylalanine-free diet, which can be prepared by mixing amino acids, is recommended for certain metabolic defects. However, the use of higher-molecular-weight phenylalanine-free peptides is more advantageous from a sensory and osmotic point of view. Such peptides can be prepared from proteins by the plastein reaction. First, the protein is partially hydrolysed with pepsin (pepsin A, EC 3.4.23.1). Then, treatment with pronase (mycolysin, EC 3.4.24.31) under suitable conditions preferentially releases amino acids with long hydrophobic side-chains such as phenylalanine. The remaining peptides are separated by gel chromatography and subjected to the plastein reaction in the presence of added tyrosine and tryptophan ethyl esters. This yields a plastein that is practically phenylalanine-free and has a predetermined ratio of other amino acids, including tyrosine. The plastein reaction can also be carried out as a one-step process, thus putting these reactions to economic, industrial-scale use.

Associations involving cross-linking Cross-linking between protein molecules is achieved with peroxidase (EC 1.11.1.7). The cross-linking occurs between tyrosine residues when a protein is incubated with a mixture of peroxidase and hydrogen peroxide (H_2O_2). Incubation of a protein with a mixture of peroxidase, H_2O_2, and catechol also results in cross-linking. In this case, oxidative deamination of lysine residues takes place, followed by aldol and aldimine condensations, i.e., reactions analogous to those catalyzed by lysyl oxidase in connective tissue.

Texturized Proteins

The world protein production for nutrition is currently about 22% from animal sources and 78% from plant sources. The plant proteins are primarily from cereals (53%) and oilseed meal (18%). Some nonconventional sources of protein (single cell proteins, leaves) have also acquired some importance. Proteins are responsible for the distinct physical structure of a number of foods, e.g., the fibrous structure of muscle tissue (meat, fish), the porous structure of bread, and the gel structure of some dairy and soya products.

Many plant proteins have a globular structure and, although available in large amounts, are used to only a limited extent in food processing. In an attempt to broaden the use of such proteins, a number of processes were developed in the mid-1950s that confer a fiber-like structure to globular proteins. Suitable processes yield products with cooking strength and a meat-like structure. They are marketed as meat extenders and meat analogs and can be used whenever a lumpy structure is desired.

Starting material The following protein sources are suitable for the production of texturized products: soya; casein; wheat gluten; oilseed meals such as from cottonseed, groundnut, sesame, sunflower, safflower or rapeseed; zein (corn protein); yeast; whey; blood plasma; or packing plant offal such as lungs or stomach tissue. The required protein content of the starting material varies and depends on the process used for texturization. The starting material is often a mixture such as soya protein with lactalbumin, or protein plus acidic polysaccharide (alginate, carrageenan, or pectin).

The suitability of proteins for texturization varies, but the molecular weight should be in the range of 10–50 kDa. Proteins of less than 10 kDa are weak fiber builders, whereas those higher than 50 kDa are disadvantageous due to their high viscosity and their tendency to gel in the alkaline pH range. The proportion of amino acid residues with polar side-chains should be high in order to enhance intermolecular chain binding. Bulky side-chains hinder such interactions, so that the proportions of amino acids with these structures should be low.

Texturization The globular proteins are unfolded during texturization by solving the intramolecular bonds. The resultant extended protein chains are stabilized through interactions with neighboring chains. In practice, texturization is achieved in one of two ways:

- The starting protein is dissolved, and the resultant viscous solution is pressed through a spinning nozzle into a coagulating bath ('*spin process*').
- The starting protein is moistened slightly, and then, at high temperature and pressure, the molten protein is extruded with shear force through the orifice of a die ('*extrusion process*').

In the *spin process*, the starting material (protein content >90%, e.g., soya protein isolate) is suspended in water and dissolved by the addition of alkali. The 20% solution is then aged at pH 11 through constant stirring. The viscosity increases during this step as the protein unfolds. The solution is then pressed through the orifices of a nozzle (5000–15 000 orifices, each with a diameter of 0.01–0.08 mm) into a coagulating bath of pH 2–3. The bath contains an acid (citric, acetic, phosphoric, lactic, or hydrochloric) and, usually, 10% sodium chloride. Spinning solutions of protein–acidic polysaccharide mixtures also contain earth alkali salts. The resulting protein filaments are extended further (to about two to four times the original length) in a 'winding up' step, and are bundled into thicker fibers with diameters of 10–20 mm. The molecular interactions are enhanced during stretching of the fibers, thus increasing the mechanical strength of the fiber bundles.

The adherent coagulation solvent is then removed by pressing the fibers between rollers, followed by passing a neutralizing bath (sodium bicarbonate plus sodium chloride) of pH 5.5–6 and, occasionally, also a hardening bath (concentrated sodium chloride). The fiber bundles may be combined into thicker cords with diameters of 7–10 cm. Additional treatment involves passage of the bundles through a bath containing a binder (a protein that coagulates when heated, such as egg protein) and other additives such as modified starch or other polysaccharides, aroma compounds or lipids. This treatment produces bundles with improved thermal stability and aroma. A typical bath for fibers that are processed to produce a meat analog might consist of 51% water, 15% ovalbumin, 10% wheat gluten, 8% soya flour, 7% onion powder, 2% protein hydrolysate, 1% sodium chloride, 0.15% monosodium glutamate, and 0.5% pigments. Finally, the soaked fiber bundles are heated and chopped.

In the *extrusion process*, the moisture content of the starting material (protein content about 50%, e.g., soya concentrate) is adjusted to 30–40%, and additives (sodium chloride, buffers, aroma compounds, pigments) are incorporated. Aroma compounds are added in fat as a carrier, when necessary, after the extrusion step to compensate for aroma losses. The protein mixture is fed into the extruder (a thermostatically controlled cylinder or conical body that contains a polished, rotating screw with a gradually decreasing pitch) that is heated to 120–180 °C and develops a pressure of $3–4 \times 10^6$ Pa. Under these conditions, the mixture is transformed into a plastic, viscous state in which solids are dispersed in the molten protein. Hydration of the protein takes place after partial unfolding of the globular molecules and stretching and rearrangement of the protein strands along the direction of mass transfer. The process is affected by the rotation rate and shape of the screw, and by the heat transfer and viscosity of the extruded material and its residence time in the extruder. As the molten material exits from the extruder through a die of 4 mm diameter, the water vaporizes, leaving behind vacuoles in the ramified protein strands.

The extrusion process is more economical than the spin process. However, it yields fiber-like particles rather than well-defined fibers. A great number and variety of extruders are now in operation. As with other food processes, there is a trend toward developing and utilizing high-temperature/short-time (HTST) extrusion cooking.

See also: **Aerated Foods**; **Casein and Caseinates**: Uses in the Food Industry; **Colloids and Emulsions**; **Emulsifiers**: Organic Emulsifiers; **Enzymes**: Uses in Food Processing; **Functional Foods**; **Protein**: Chemistry; Quality; Sources of Food-grade Protein; **Rheological Properties of Food Materials**; **Single-cell Protein**: Yeasts and Bacteria; **Soy (Soya) Beans**: Processing for the Food Industry; **Soy (Soya) Cheeses**; **Stabilizers**: Types and Function; **Wheat**: Grain Structure of Wheat and Wheat-based Products

Further Reading

Damodaran S and Paraf A (eds) (1997) *Food Proteins and their Applications*. New York: Marcel Dekker.

Froning GW (1988) Nutritional and functional properties of egg proteins. In: Hudson BJF (ed.) *Developments in Food Proteins*, vol. 6, pp. 1–34. London: Elsevier Applied Science.

Hettiarachchy NS and Ziegler GR (eds) (1994) *Protein Functionality in Food Systems*. New York: Marcel Dekker.

Kinsella JE (1978) Texturized proteins: fabrication, flavoring, and nutrition. *Critical Reviews in Food Science and Nutrition* 10: 147–207.

Kinsella JE and Shetty KJ (1978) Yeast proteins: recovery, nutritional and functional properties. *Advances in Experimental Medicine and Biology* 105: 797–825.

Kinsella JE and Srinivasan D (1981) Nutritional, chemical and physical criteria affecting the use and acceptability of proteins in foods. In: Solms J and Hall RL (eds) *Criteria of Food Acceptance*, pp. 296–332. Zurich: Forster Verlag.

Mitchell JR (1986) Foaming and emulsifying properties of proteins. In: Hudson BJF (ed.) *Developments in Food Proteins*, vol. 4, pp. 291–338. London: Elsevier Applied Science.

Morrissey PA, Mulvihill DM and O'Neil EM (1987) Functional properties of muscle proteins. In: Hudson BJF

(ed.) *Developments in Food Proteins*, vol. 5, pp. 195–256. London: Elsevier Applied Science.
Morrissey PA, Mulvihill DM and O'Riordan D (1991) Functional properties of blood proteins. In: Hudson BJF (ed.) *Developments in Food Proteins*, vol. 7, pp. 125–166. London: Elsevier Applied Science.
Phillips LG, Whitehead DM and Kinsella J (1994) *Structure–Function Properties of Food Proteins*. San Diego, CA: Academic Press.
Pomeranz Y (1991) *Functional Properties of Food Components*, 2nd edn. San Diego, CA: Academic Press.
Poole S and Fry JC (1987) High-performance protein foaming and gelation systems. In: Hudson BJF (ed.) *Developments in Food Proteins*, vol. 5, pp. 257–298. London: Elsevier Applied Science.
Puigserver AJ, Sen LC, Clifford AJ, Feeney RE and Whitaker JR (1978) A method for improving the nutritional value of food proteins: covalent attachment of amino acids. *Advances in Experimental Medicine and Biology* 105: 587–612.
Watanabe M and Arai S (1988) The plastein reaction and its applications. In: Hudson BJF (ed.) *Developments in Food Proteins*, vol. 6, pp. 179–217. London: Elsevier Applied Science.
Whitaker JR, Shahidi F, Munguia AL, Yada RY and Fuller G (eds) (1998) *Functional Properties of Proteins and Lipids, ACS Symposium Series 708*. Washington, DC: American Chemical Society.

Interactions and Reactions Involved in Food Processing

J K P Weder and H-D Belitz†, Technische Universität München, Garching, Germany

Introduction

The nature and extent of the chemical changes induced in proteins by food processing depend on a number of parameters, for example, food composition and processing conditions, such as temperature, pH, or the presence of oxygen. The consequences of such reactions may be desirable or undesirable. For example, the biological value of proteins may be decreased by destruction of essential amino acids, conversion of essential amino acids into derivatives that are not metabolized, or reduced digestibility of protein as a result of intra- or interchain cross-linking. The formation of toxic degradation products is also possible. The nutritional/physiological and toxicological assessment of changes induced by processing of foods is a subject of some controversy and opposing opinions.

Reactions with Carbohydrates

Many foodstuffs contain reducing sugars and amino compounds such as proteins, peptides, amino acids, and amines. Reactions between these components are usually classed under the term '*Maillard reaction*' or 'nonenzymatic browning.' They occur especially at a higher temperature, low water activity and during longer storage. Reactive sugars are glucose, fructose, maltose, lactose, and, to a smaller extent, reducing pentoses. On the side of the amino components, primary amines are more important than secondary amines because their concentration in foods is usually higher. Exceptions are, for example, malt and corn products, which have a high proline content. In the case of proteins, the ε-amino groups of their lysine residues react predominantly, but guanidino groups of arginine residues can also react. These reactions result in:

- Brown pigments, known as '*melanoidins*,' which contain nitrogen in variable amounts, and have varying molecular masses and solubilities in water. Little is known about their structure. Browning is desired in baking and roasting, but not in foods which are only slightly colored or have a distinct color (condensed milk, faintly colored dried soups, tomato soup).
- Volatile compounds, which often contribute to the aroma. Aroma formation through the Maillard reaction is desired in cooking, baking, roasting, and frying. However, the generation of off-flavors in food during storage, especially in the dehydrated state, or during pasteurization, sterilization, or roasting is not desired.
- Bitter substances, which are partially desired (coffee), but can also cause an off-flavor, e.g., in grilled meat or fish.
- Reductones, which have highly reductive properties and contribute to the stabilization of foods against oxidative deterioration.
- Losses of essential amino acids (e.g., lysine, methionine).
- Mutagenic compounds.
- Cross-linking of proteins.

The Maillard reaction of the ε-amino group of lysine prevails in the presence of reducing sugars, for example, lactose or glucose, which yield the protein-bound Amadori compounds, ε-*N*-deoxylactulosyl-1-lysine or ε-*N*-deoxyfructosyl-1-lysine, respectively. Lysine is not biologically available in these forms.

† Deceased.

Acid hydrolysis of such primary reaction products yields lysine as well as the degradation products furosine and pyridosine in a constant ratio. A nonreducing sugar (e.g., sucrose) can also cause a loss of lysine when conditions for sugar hydrolysis are favorable, leading to the formation of the reactive sugars glucose and fructose.

Very recently, two intensely red-colored compounds were isolated from brown–orange melanoidins with mass > 10 000 Da obtained by thermal treatment of an aqueous solution of casein and furan-2-carboxaldehyde (furfural), a well-known product of the Maillard reaction. These compounds were identified as the so-far unknown chromophoric amino acids (*S*)-2-amino-6-{4-[(*E*)-1-formyl-2-(2-furyl)ethenyl]-5-(2-furyl)-2-[(*E*)-(2-furyl)methylidene]-2,3-dihydro-3-oxo-1*H*-pyrrol-1-yl}hexanoic acid and its 2-[(*Z*)-(2-furyl)methylidene] isomer.

In the late 1970s, it was shown that charred surface portions of grilled fish and meat, as well as the smoke condensates, isolated during grilling, exhibit highly mutagenic effects in microbial tests (Ames test with *Salmonella typhimurium* mutant strain TA 98). Model experiments showed that pyrolysates of amino acids and proteins are responsible for those effects. Table 1 lists several mutagenic compounds (pyridoindoles, pyridoimidazoles, pyridocarbazoles, and tetraazafluoranthenes) that were isolated from such pyrolysates.

Mutagenic compounds can also be formed at lower temperatures from amino acids and proteins. The compounds listed in Table 2 were isolated from meat extract, deep-fried meat, grilled fish, and heated model mixtures of creatinine, an amino acid (glycine, alanine, or threonine), and glucose. The compounds, mainly imidazoquinolines and imidazoquinoxalines, are formed from creatinine, subsequent products of the Maillard reaction (pyridines, pyrazines), and amino acids. Their toxicity is based on the heteroaromatic amino function.

Pentosidine (N^5-[*N*-(5-amino-5-carboxy-1-pentyl)-pyrido[4,5-*b*]imidazol-2-yl]-2,5-diaminopentanoic acid), a cross-linked amino acid in which one arginine and one lysine residue are linked together by a pentose, was recently detected in foods. The low levels found in commercial thermally processed food samples (up to 4 mg kg^{-1} protein in milk products, up to 40 mg kg^{-1} protein in roasted coffee beans and bakery products) indicate that pentosidine does not contribute significantly to food protein cross-linking.

CROSSPY (1,4-bis-(5-amino-5-carboxy-1-pentyl)-pyrazinium radical cation), a cross-linked amino acid in which two lysine residues are linked together through reaction with two molecules of glycolaldehyde, a Maillard reaction product, was more recently isolated from heated aqueous solutions of bovine serum albumin and glycolaldehyde and detected also in toasted wheat bread crust and dark brown roasted cacao as well as coffee beans by ESR (electron spin resonance) spectroscopy.

Reactions with Lipid Oxidation Products

Products Formed from Hydroperoxides

Fatty acid hydroperoxides formed thermally or enzymatically in food are usually degraded further. This degradation can also be of nonenzymatic nature. In nonspecific reactions involving heavy metal ions, heme(in) compounds or proteins, hydroperoxides are transformed into oxo, epoxy, mono-, di-and trihydroxy carboxylic acids. Unlike hydroperoxides, i.e., the primary products of autoxidation, some of these derivatives have a bitter taste. Such compounds are detected in legumes and cereals. They may play a role in other foods rich in unsaturated fatty acids and proteins, such as fish and fish products.

Lipid–Protein Complexes

Studies related to the interaction of hydroperoxides with proteins have shown that, in the absence of oxygen, linoleic acid 13-hydroperoxide reacts with *N*-acetylcysteine, yielding an adduct that consists of

Table 1 Mutagenic compounds from pyrolysates of amino acids and proteins

Mutagenic compound	*Short form*	*Pyrolyzed compound*
3-Amino-1,4-dimethyl-5*H*-pyrido[4,3-*b*]indole	Trp-P-1	Tryptophan
3-Amino-1-methyl-5*H*-pyrido[4,3-*b*]indole	Trp-P-2	Tryptophan
2-Amino-6-methyldipyrido[1,2-*a*:3′,2′-*d*]imidazole	Glu-P-1	Glutamic acid
2-Aminodipyrido[1,2-*a*:3′,2′-*d*]imidazole	Glu-P-2	Glutamic acid
3,4-Cyclopentenopyrido[3,2-*a*]carbazole	Lys-P-1	Lysine
4-Amino-6-methyl-1*H*-2,5,10,10*b*-tetraazafluoranthene	Orn-P-1	Ornithine
2-Amino-5-phenylpyridine	Phe-P-1	Phenylalanine
2-Amino-9*H*-pyrido[2,3-*b*]indole	AαC	Soya globulin
2-Amino-3-methyl-9*H*-pyrido[2,3-*b*]indole	MeAαC	Soya globulin

Modified from Chen C, Pearson AM and Gray JI (1990) Meat mutagens. *Advances in Food and Nutrition Research* 34: 387–449.

Table 2 Mutagenic compounds from various heated foods and model systems

Mutagenic compound	*Short form*	*Food/model system*[a]
2-Amino-3-methylimidazo[4,5-*f*]quinoline	IQ	1, 2, 3
2-Amino-3,4-dimethylimidazo[4,5-*f*]quinoline	MeIQ	3
2-Amino-3,8-dimethylimidazo[4,5-*f*]quinoxaline	MeIQx	2, 3
2-Amino-3,4,8-trimethylimidazo[4,5-*f*]quinoxaline	4,8-DiMeIQx	2, 3, 5, 6
2-Amino-3,7,8-trimethylimidazo[4,5-*f*]quinoxaline	7,8-DiMeIQx	4
2-Amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine	PhIP	2

[a]1, meat extract; 2, deep-fried meat; 3, grilled fish; 4, model mixture of creatinine, glycine and glucose; 5, as 4, but with alanine; 6, as 4, but with threonine.
Modified from Chen C, Pearson AM and Gray JI (1990) Meat mutagens. *Advances in Food and Nutrition Research* 34: 387–449.

Table 3 Amino acid losses occurring in proteins exposed to peroxidized lipids

Reaction system		*Reaction conditions*		*Amino acids lost*[a] (% loss)
Protein	*Lipid*	*Time*	*Temperature (°C)*	
Casein	Linoleic acid ethyl ester	24 h	55[b]	Lys(10), Thr(10), Val(10), Ala(9), Tyr(8), Phe(8), Ser(8), Arg(8), Asp(8)[c,d]
Casein	Linoleic acid ethyl ester	4 days	60[e]	Lys(50), Met(47), Ile(30), Phe(30), Arg(29), Asp(29), Gly(29), His(28), Thr(27), Ala(27), Tyr(27)[c,d]
Ovalbumin	Linoleic acid ethyl ester	24 h	55[b]	Met(17), Ser(10), Lys(9), Ala(8), Leu(8)[c,d]
Ovalbumin	Linoleic acid ethyl ester	4 days	60[e]	Lys(50), Met(42), Leu(22), His(21), Val(21)[c,d]
Ribonuclease	Linoleic acid	40 min	37[b]	Met(99), Tyr(62), His(54), Lys(51), Cys(40)[c]
Trypsin	Linoleic acid	40 min	37[b]	Met(83), His(12)[c]
Lysozyme	Linoleic acid	8 days	37[b]	Trp(56), His(42), Lys(17), Met(14), Arg(9)

[a]Only the most labile amino acids are listed.
[b]Aqueous system.
[c]Trp not analyzed.
[d]Cys not analyzed.
[e]80% relative humidity.
Data from Gardner HW (1979) Lipid hydroperoxide reactivity with proteins and amino acids: a review. *Journal of Agricultural and Food Chemistry* 27: 220–229 and references therein.

several isomers. However, in the presence of oxygen, covalently bound amino acid–fatty acid adduct formation is significantly suppressed; instead, oxidized fatty acids are formed.

The difference in reaction products is explained by different reaction pathways. The thiyl radical, derived from cysteine by abstraction of a hydrogen atom, is added to the epoxyallyl radical only in the absence of oxygen. In the presence of oxygen, oxidation of cysteine and fatty acids to their oxidized forms occurs with a higher reaction rate than in the previous reaction. As a consequence, a large portion of the oxidized lipid from a protein-containing food stored in air does not have lipid–protein covalent bonds and, hence, is readily extracted with a lipid solvent such as chloroform/methanol (2:1, v/v).

Protein Changes

Some properties of proteins change when they react with hydroperoxides or their degradation products. This is reflected by changes in food texture, decreases in protein solubility (formation of cross-linked proteins), changes in color (browning), and changes in nutritive value (loss of essential amino acids).

The radicals generated from hydroperoxides can abstract hydrogen atoms from proteins, preferentially from the amino acids tryptophan, histidine, lysine, arginine, cysteine and tyrosine, wherein the nitrogen-, sulfur- or phenolic hydroxyl-containing group reacts. Protein radicals combine with each other, resulting in the formation of a protein network.

Malonaldehyde is generated under certain conditions during lipid peroxidation. As a bifunctional reagent, malonaldehyde can cross-link proteins through a Schiff base reaction with the ε-amino groups of two lysine residues. The Schiff base product is a conjugated fluorochrome that has distinct spectral properties (λ_{max} excitation $\sim$350 nm; λ_{max} emission $\sim$450 nm). Hence, it can be used to detect lipid peroxidation and the reactions derived from it with the proteins present.

Table 4 Products formed from amino acids exposed to peroxidizing lipids

Reaction system		*Compounds formed from amino acids*
Amino acid	*Lipid*	
Histidine	Methyl linoleate	3-(Imidazol-4-yl)lactic acid, 3-(imidazol-4-yl)acetic acid; histamine, valine, aspartic acid, ethylamine
Cysteine	Ethyl arachidonate	Cystine, hydrogen sulfide, alanine
	Linoleic acid	Cystine, cysteic acid, cystine disulfoxide
Methionine	Methyl linoleate	Methionine sulfoxide
Lysine	Methyl linoleate	Diaminopentane, aspartic acid, glycine, alanine, α-aminoadipic acid, pipecolinic acid, 1,10-diamino-1,10-dicarboxydecane

Modified from Gardner HW (1979) Lipid hydroperoxide reactivity with proteins and amino acids: a review. *Journal of Agricultural and Food Chemistry* 27: 220–229.

Table 5 Formation of unusual amino acids by alkali treatment of proteins

Name	*Formula*
3-N^6-Lysinoalanine (R═H) 3-N^6-Lysino-3-methylalanine (R═CH_3)	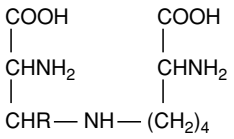
3-N^5-Ornithinoalanine (R═H) 3-N^5-Ornithino-3-methylalanine (R═CH_3)	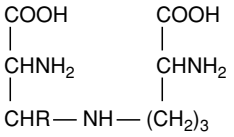
Lanthionine (R═H) 3-Methyllanthionine (R═CH_3)	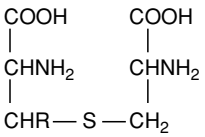
3-Aminoalanine (R═H) 2,3-Diaminobutyric acid (R═CH_3)	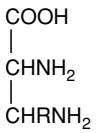
3-N^{π}-Histidinoalanine	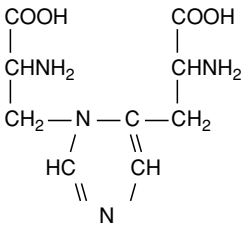
3-N^{τ}-Histidinoalanine	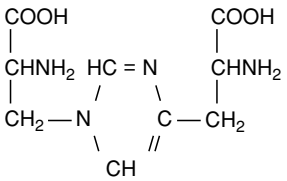

Reactions resulting in the formation of a protein network like that outlined above have practical implications; for example, they are responsible for the decrease in solubility of fish protein during frozen storage. Also, the monocarbonyl compounds derived from autoxidation of unsaturated fatty acids readily condense with free amino groups of the proteins, thus forming Schiff bases that can provide brown polymers by repeated aldol condensations. The brown polymers are often nitrogen-free since the amino compound can be readily eliminated by hydrolysis. When hydrolysis occurs in the early stages of aldol condensations (after the first or second condensation) and the released aldehyde, which has a powerful odor, does not re-enter the reaction, the condensation process results not only in discoloration (browning) but also in a change in aroma.

Decomposition of Amino Acids

Studies of model systems have revealed that protein cleavage and degradation of side-chains, rather than the formation of protein networks, are the preferred reactions when the water content of protein–lipid mixtures decreases. Examples for the extent of losses in amino acids in a protein in the presence of an oxidized lipid are presented in Table 3. The strong dependence of this loss on the nature of the protein and on reaction conditions is obvious. Degradation products obtained in model systems of pure amino acids and oxidizing lipids are listed in Table 4.

Reactions Under Alkaline Conditions

Losses of available lysine, cystine, serine, threonine, arginine, and some other amino acids occur at high pH values. Hydrolysates of alkali-treated proteins often contain some unusual compounds, such as ornithine, β-aminoalanine, lysinoalanine, ornithinoalanine, lanthionine, methyllanthionine and D-*allo*-isoleucine, as well as other D-amino acids. The formation of these compounds is based on the following reactions: 1,2-eliminations in the case of hydroxy amino acids and thio amino acids

Table 6 Lysinoalanine content of various foods

Food	*Origin/treatment*	*Lysinoalanine (mg kg^{-1} protein)*
Frankfurter	CP[a], as purchased	0
	Boiled	50
	Oven-baked	170
Chicken thigh	CP, raw	0
	Oven-baked	110
	Cooked in microwave oven	200
Egg white	Fresh	0
	Boiled (3 min)	140
	(10 min)	270
	(30 min)	370
	Pan-fried (10 min at 150 °C)	350
	(30 min at 150 °C)	1100
Dried egg white	CP	160–1 820[b]
Condensed milk	CP	360–540
Milk, evaporated	CP	590–860
Skim milk, evaporated	CP	520
Milk product for infants	CP	150–640
Infant food	CP	< 55–150
Simulated cheese	CP	1070
Corn chips, pretzels, tortillas, etc.	CP	170–500
Cocoa powder	CP	130–190
Whipping agent	CP	860–50 000
Soya protein isolate	CP	0–370
Hydrolyzed vegetable protein	CP	40–500
Sodium caseinate	CP	45–6 900
Calcium caseinate	CP	250–4 320
Yeast extract	CP	120

[a]Commercial product.
[b]Variation range for different brand name products.
Data from Sternberg M and Kim CY (1977) Lysinoalanine formation in protein food ingredients. *Advances in Experimental Medicine and Biology* 86B: 73–84 and Haagsma N and Slump P (1978) Evaluation of lysinoalanine determinations in food proteins. *Zeitschrift für Lebensmittel-Untersuchung und -Forschung* 167: 238–240.

result in 2-aminoacrylic acid (dehydroalanine) or 2-aminocrotonic acid (dehydroaminobutyric acid). In the case of cystine, the eliminated thiolcysteine can form a second dehydroalanine residue. Alternatively, cleavage of the cystine disulfide bond can occur by nucleophilic attack on sulfur, yielding dehydroalanine residues through thiol and sulfinate intermediates. Intra- and interchain cross-linking of proteins can occur through addition of amino and thiol groups containing amino acid side-chains to dehydroalanine residues. Ammonia may also react via an addition reaction. Acid hydrolysis of such a cross-linked protein yields the unusual amino acids listed in Table 5. Ornithine is formed by cleavage of arginine.

Reaction of the imidazole ring of protein-bound histidine with dehydroalanine residues results in two histidinoalanine isomers, depending on the nitrogen that reacts. The two isomers, 3-N^{τ}- and 3-N^{π}-histidinoalanine, are formed in a constant ratio of about 8:1. The histidinoalanine content of commercial foods containing milk protein (50–1800 mg kg^{-1} protein) is comparable with their lysinoalanine content.

Histidinomethylalanine and lysinomethylalanine do not contribute significantly to protein cross-linking. Their demonstration failed in numerous acid hydrolysates of various milk products. Model studies showed that β-elimination of threonine takes place much more slowly than that of serine and that the reactivity of the resulting dehydroaminobutyric acid towards nucleophiles is also much lower.

The formation of D-amino acids occurs through abstraction and subsequent back addition of a proton via the C^{α} carbanion. For example, the extent of D-amino acid formation is 30.1, 24.9, 18.8, 15.8, 8.0, 6.6, and 5.7% for aspartic acid, phenylalanine, glutamic acid, alanine, leucine, valine, and proline, respectively, when soya protein is exposed to 0.1 N sodium hydroxide (pH $\sim$ 12.5) at 65 °C for 3 h. The reaction with L-isoleucine is particularly interesting. L-Isoleucine is isomerized to D-*allo*-isoleucine, which, unlike other D-amino acids, is a diastereomer and thus has a retention time different from L-isoleucine,

making its direct determination by amino acid analysis possible.

Heating proteins in a dry state at neutral pH results in the formation of isopeptide bonds between the ε-amino groups of lysine residues and the β- or γ-carboxamide groups of asparagine or glutamine residues. These isopeptide bonds are cleaved during acid hydrolysis of protein and, therefore, do not contribute to the occurrence of unusual amino acids. But their formation reduces protein digestibility. A more drastic heat treatment of proteins in the presence of water leads to more pronounced degradation. Alkaline treatment of a protein isolate of sunflower seeds shows that serine, threonine, arginine, and isoleucine concentrations are markedly decreased with increasing concentrations of sodium hydroxide. New amino acids (ornithine and *allo*-isoleucine) are formed. Initially, the lysine concentration decreases but increases at higher concentrations of alkali. Lysinoalanine behaves in the opposite manner. The formation of lysinoalanine is influenced not only by pH but also by the protein source. A pronounced reaction occurs in casein, even at pH 5.0, due to the presence of phosphorylated serine residues, whereas noticeable reactions occur in gluten from wheat or in zein from corn only in the pH range of 8–11. Table 6 lists the content of lysinoalanine in food products processed industrially or prepared under 'usual household conditions.' The lysinoalanine content is obviously affected by the food type and by the processing conditions.

Reactions Under Oxidative Conditions

Oxidative changes in proteins primarily involve methionine, which forms methionine sulfoxide relatively readily. After *in vivo* reduction to methionine, protein-bound methionine sulfoxide is apparently biologically available.

See also: **Amines**; **Amino Acids**: Properties and Occurrence; Determination; Metabolism; **Browning**: Nonenzymatic; Toxicology of Nonenzymatic Browning; **Carbohydrates**: Classification and Properties; Interactions with Other Food Components; **Fatty Acids**: Properties; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Mutagens**; **Oxidation of Food Components**; **Peptides**

Further Reading

Chen C, Pearson AM and Gray JI (1990) Meat mutagens. *Advances in Food and Nutrition Research* 34: 387–449.

Eriksson C (ed.) (1981) *Maillard Reactions in Food. Chemical, Physiological and Technological Aspects*. Oxford: Pergamon Press.

Friedman M (ed.) (1991) *Nutritional and Toxicological Consequences of Food Processing. Advances in Experimental Medicine and Biology 289*. New York: Plenum Press.

Fujimaki M, Namiki M and Kato H (eds) (1986) *Aminocarbonyl Reactions in Food and Biological Systems*. Amsterdam: Elsevier.

Gardner HW (1979) Lipid hydroperoxide reactivity with proteins and amino acids: a review. *Journal of Agricultural and Food Chemistry* 27: 220–229.

Haagsma N and Slump P (1978) Evaluation of lysinoalanine determinations in food proteins. *Zeitschrift für Lebensmittel-Untersuchung und -Forschung* 167: 238–240.

Henle T, Walter AW and Klostermeyer H (1993) Detection and identification of the cross-linking amino acids N^{τ}- and N^{π}-(2′-amino-2′-carboxy-ethyl)-L-histidine ('histidinoalanine,' HAL) in heated milk products. *Zeitschrift für Lebensmittel-Untersuchung und -Forschung* 197: 114–117.

Henle T, Schwarzenbolz U and Klostermeyer H (1997) Detection and quantification of pentosidine in foods. *Zeitschrift für Lebensmittel-Untersuchung und -Forschung A* 204: 95–98.

Hofmann T (1998) Studies on melanoidin-type colorants generated from the Maillard reaction of protein-bound lysine and furan-2-carboxaldehyde – chemical characterisation of a red coloured domaine. *Zeitschrift für Lebensmittel-Untersuchung und -Forschung A* 206: 251–258.

Hofmann T, Bors W and Stettmaier K (1999) Radical-assisted melanoidin formation during thermal processing of foods as well as under physiological conditions. *Journal of Agricultural and Food Chemistry* 47: 391–396.

Labuza TP, Reineccius GA, Monnier VM, O'Brien J and Baynes JW (eds) (1994) *Maillard Reactions in Chemistry, Food, and Health*. Cambridge: Royal Society of Chemistry.

Ledl F (1990) Chemical pathways of the Maillard reaction. In: Finot PA, Aeschbacher HU, Hurrell RF and Liardon R (eds) *The Maillard Reaction in Food Processing, Human Nutrition and Physiology*, pp. 19–47. Basel: Birkhäuser.

O'Brien J, Nursten HE, Crabbe MJC and Ames JM (eds) (1998) *The Maillard Reaction in Foods and Medicine*. Cambridge: Royal Society of Chemistry.

Phillips RD and Finley JW (eds) (1989) *Protein Quality and the Effects of Processing*. New York: Marcel Dekker.

Sternberg M and Kim CY (1977) Lysinoalanine formation in protein food ingredients. *Advances in Experimental Medicine and Biology* 86B: 73–84.

Waller GR and Feather MS (eds) (1983) *The Maillard Reaction in Foods and Nutrition. ACS Symposium Series 215*. Washington, DC: American Chemical Society.

Walter AW, Henle T, Haeßner R and Klostermeyer H (1994) Studies on the formation of lysinomethylalanine and histidinomethylalanine in milk products. *Zeitschrift für Lebensmittel-Unterschung und -Forschung* 199: 243–247.

Whitaker JR and Fujimaki M (eds) (1980) *Chemical Deterioration of Proteins. ACS Symposium Series 123*. Washington, DC: American Chemical Society.

Quality

G S Gilani and N Lee, Bureau of Nutritional Sciences, Health Canada, Ottawa, Ontario, Canada

Background

The quality of a dietary protein is determined by the pattern and concentration of indispensable or essential amino acids, the digestibility of the protein, and the bioavailability of its amino acids. To satisfy the protein requirements of the body, the diet must supply both enough of the indispensable amino acids (IAA) in required proportions and enough of the total amino nitrogen for synthesis of dispensable or nonessential amino acids. These are needed to support the synthesis of body protein and the production of other nitrogen-containing compounds such as hormones and neurotransmitters involved in a range of physiological functions.

An IAA is defined as an amino acid that cannot be synthesized in the body, or at least not in adequate amounts. Nine amino acids – histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine – are not synthesized in adequate amounts by mammals and are, therefore, IAA for humans. Dispensable amino acids (alanine, arginine, aspartic acid, cyst(e)ine, glutamic acid, glycine, hydroxyproline, proline, serine, and tyrosine) are not essential in the diet, as they can be synthesized from IAA and amino nitrogen. Since cystine can replace part of the requirement for methionine, and tyrosine a part of the requirement for phenylalanine, cystine and tyrosine are also included when considering IAA contents of diets, and these pairs are then expressed as total sulfur amino acids (methionine + cystine) and total aromatic amino acids (phenylalanine + tyrosine). Most dietary proteins contain a mixture of all 20 common amino acids in varying proportions.

The reported quality of a protein may vary depending on the function measured, for instance growth versus maintenance, because these differ in the amount and pattern of IAA they require. In all cases, however, the quality of a protein source is limited by the IAA available in least amounts (the first limiting amino acid) relative to requirements. Occasionally, a protein source may be colimiting in more than one IAA. There are several different ways that have been developed to assess how well a protein could be expected to meet human protein and amino acid requirements.

Methods of Assessing Protein Quality of Foods

The only true measurement of protein quality for human use is an evaluation, in humans, of growth, nitrogen balance, or metabolic balance or some other appropriate test, preferably performed with suitable subjects from the target population of interest. Such evaluations directly reflect how well the food or food product meets human needs and this is a function of the proportion and content of IAA, the digestibility of the protein, and the bioavailability of amino acids in the food or food product.

Conventional or 'long-term' nitrogen-balance experiments which measure dietary nitrogen retained in the body (nitrogen intake − fecal nitrogen output − urinary nitrogen output) in humans are often considered to be the standard or reference method for assessing protein quality of foods. Nitrogen balance is used as a surrogate for protein balance and is assumed to reflect how well protein intake is maintaining body protein homeostasis and organ function. Nitrogen balance can be expressed in two ways: one relates nitrogen intake to nitrogen retained, as measured by the net protein utilization (NPU) method, and the second relates nitrogen absorbed to nitrogen retained, as measured by the biological value (BV) (see **Table 1**).

Nitrogen-balance studies are used to determine the minimum intake of a test or food protein that will maintain nitrogen balance. The amount of a protein required to maintain nitrogen balance is a measure of its quality. Therefore, the accepted nitrogen balance technique to measure protein quality in humans is to feed graded levels of a test protein for 10–21 days. Shorter-term human nitrogen balance assays or those based on single test levels can provide useful information, but at this time, it is not clear without further validation and critical evaluation whether these methods have the capacity to predict, in quantitative terms, the nutritional value of a source of protein.

More recently, some research has focused on protein and nitrogen metabolism during different periods

Table 1 Comparison of features of rat assays for protein quality evaluation

	Credits protein used in both growth and maintenance	*Description/equation*	*Comments*
Growth methods			
PER (standardized) (protein efficiency ratio)	No, growth only; no protein-free diet group included	Weight gain of rats on test or casein control diets per gram of protein consumed, standardized to casein PER of 2.5	Simple but relatively costly and lacking other attributes including proportionality to protein quality, accuracy, precision, and reproducibility
NPR (net protein ratio)	Yes, protein-free diet group included	Weight gain on test or control diet plus weight loss on nonprotein diet per gram of protein consumed	Overcomes the major weakness of PER, but the assumption is that the quality of protein required to prevent weight loss is equivalent to that needed for maintenance; usually conducted in 2 weeks, half the time required for standardized PER; comparable to NPU (below) except estimated from body weight change, not nitrogen balance
RNPR (relative net protein ratio)	Yes, protein-free diet group included	NPR for test protein divided by NPR for reference protein expressed as a fraction of 100	Values are proportional with respect to protein quality within reasonable limits; more accurate and reproducible and cheaper than PER
NU (nitrogen utilization)	Yes, no protein-free diet group included; factor of maintenance calculated	Body weight gain + 0.1 (initial + final body weight) divided by nitrogen consumed	Factor of maintenance is calculated instead of measured
RNU (relative nitrogen utilization)	Yes, no protein-free diet group included; factor of maintenance calculated	100 × NU (test protein)/NU(reference protein)	RNU gives almost the same results as RNPR
Nitrogen balance methods			
NPU (net protein utilization)	Yes, protein-free diet group included	The proportion of ingested protein (N×6.25) that is retained for maintenance and/or growth Equation: 100 × (N retained on protein diet − N retained on protein-free diet/g N ingested on protein diet) Note: N retained = N intake − fecal N output − urinary N output	Relates N intake to N retained; this index includes the factor of digestibility; nitrogen retained can be measured either by the balance method or by carcass analysis; nitrogen balance is used as a surrogate for protein balance and is assumed to reflect how well protein intake is maintaining body protein homeostasis and organ function; NPU is the product of BV and TPD
BV (biological value)	Yes, protein-free diet group included	The proportion of nitrogen absorbed that is retained for maintenance and/or growth purposes Equation: 100 × (N retained on protein diet − N retained on protein-free diet)/N ingested − (fecal N on protein diet − fecal N on protein-free diet)	Relates N absorbed to N retained; thus, as an index, it excludes the factor of digestibility
TPD (true protein digestibility coefficient)	Yes, protein-free diet group included	The proportion of ingested protein (N×6.25) that is absorbed Equation: 100 × (N ingested on protein diet − (fecal N on protein diet − fecal N on nitrogen-free diet))/N ingested on protein diet	This is not a measure of protein quality, only digestibility; it is needed for calculating protein digestibility-corrected amino acid score; this test requires an adaptation period of 4 days and a fecal collection period of 5 days

of the day in relation to eating and fasting. The metabolic fate and retention of dietary nitrogen during the postprandial phase may be of particular importance in determining the quality of a given protein. An experimental approach, termed 'net postprandial protein utilization,' measures the efficiency of postprandial nitrogen retention in humans after the ingestion of different protein sources and appears to be able to differentiate between proteins of similar high quality. Postprandial nitrogen retention may offer a sensitive *in-vivo* method to discriminate between protein sources of differing quality. It is too early, however, to say whether this will become a routine method of protein-quality evaluation.

Although they are the gold standard, human studies cannot be carried out on a routine basis for reasons of cost and ethics. Therefore, *in-vitro* and animal assay techniques have been developed that correlate well with data from human experimentation for each food product.

Routine Protein-quality Evaluation Methods

Criteria for a valid and useful assay for routine protein-quality evaluation include accuracy, precision, reproducibility, proportionality to protein quality, and low cost. The ability of assays to accurately measure protein quality depends on a number of factors including whether they account for protein requirements for growth alone or for growth plus maintenance.

Chemical and Microbiological Assays

The chemical assays are the amino acid scoring methods which involve chemical analysis of the amino acid composition of the protein of interest and comparison of its IAA profile to that of a reference profile. As a routine assay, this method has been standardized as the protein digestibility-corrected amino acid score (PDCAAS), which will be discussed later. Microbiological, or *in-vitro*, assays use microorganisms that require the same amino acids as humans. The most successful of these uses the protozoans, *Tetrahymena pyriformis* or *T. thermophila*, which require the same 10 IAA as the growing rat and human. The use of this method is limited mainly by the fact that the growth of *T. pyriformis* can be inhibited by factors in the food unrelated to protein quality, including several food additives and spices.

Rat Assays

The rat is the most common species used in a protein-quality bioassay. Rat assays are based on the measurement of either growth or nitrogen balance in young animals in response to feeding defined amounts of one or more test proteins or a control protein (see also Table 1).

The protein efficiency ratio (PER) was the first method adopted for routine assessment of protein quality of foods. The PER is a standardized method in which a test diet and a casein control diet, both containing 10% protein ($N \times 6.25$) are fed to weanling rats for a period of 4 weeks. The PER values are calculated by dividing the weight gain of the rats by the amount of protein consumed over this period. To compare PER values of diets determined by different laboratories, the PER values are corrected to an assumed value of 2.5 for casein control. It has been recognized that the PER method, although simple, does not meet any other criteria of a valid bioassay, as noted above. The failure of the PER assay to properly credit protein used for maintenance purposes is the chief criticism of the method. The poorer the quality of dietary protein, the larger is the error that is introduced because of this failure to make allowance for maintenance. For this reason, PER values are not proportional to protein quality, i.e., a PER of 2.0 is not twice as good as a PER of 1.0.

The net protein ratio (NPR) method overcomes the major weakness in the PER assay by adding the weight loss of a group of rats fed a zero-protein diet to the weight gain of the rats fed the test protein. However, it has to be assumed that the protein required to prevent weight loss of rats fed the nonprotein diet is equivalent to the protein needed for maintenance. As normally carried out, NPR values are uncorrected to any standard. However, relative NPR (RNPR), where the NPR of a test protein is expressed relative to a value of 100 for the NPR of a reference protein, is expressed on a scale of 1–100. NPU and BV, two nitrogen-balance methods mentioned earlier, are also expressed on a scale of 1–100.

The relative net protein ratio (RNPR) method at 2 weeks is shorter than the standard PER method (4 weeks) and therefore is less expensive. Also, RNPR values (unlike PER values) are proportional to each other in protein quality within reasonable limits. RNPR is more accurate and reproducible than PER.

The nitrogen utilization (NU) method is a modification of NPR, the factor of maintenance being calculated instead of measured:

$$\begin{aligned} \text{NU} = {} & \text{body weight gain of test group} \\ & + 0.1\ (\text{initial} + \text{final body weight}) \\ & / \text{nitrogen consumed by test group.} \end{aligned}$$

The factor, 0.1, derives from experimental observations of the growth of rats on protein-free diets. The relative NU (NU of test protein/NU of reference protein × 100) gives essentially the same values as RNPR. For example: lactalbumin, RNPR of 87 vs. RNU of 87; for egg white, RNPR of 97 vs. RNU of 97; for wheat gluten, RNPR of 32 vs. RNU of 31; for wheat gluten + soy protein, RNPR of 64 vs. RNU of 64. The RNU method eliminates the feeding of a 'zero' protein diet to a group of rats; thus, less labor is required for the RNU procedure than for the RNPR procedure.

Limitations of Rat Assays

It is well known that the requirements of rats for sulfur amino acids are much larger than those of humans. Thus, a limitation of any rat growth assay is that it will underestimate the protein quality of any protein limiting in sulfur amino acids such as soybean protein, peanuts, and other legumes or pulses. (*See* **Protein**: Sources of Food-grade Protein.)

Protein Digestibility-corrected Amino Acid Score (PDCAAS) Method

A method based on a comparison of the amino acid content of food with human requirements (amino acid scoring system) is considered internationally to be the most suitable approach for routine assessment of the protein quality of foods. The amino acid score should be corrected for incomplete digestibility of protein, and for the unavailability of individual amino acids, especially those that are susceptible to damage by processing treatments.

The FAO/WHO has recommended the PDCAAS as the preferred method for routine prediction of protein quality of properly processed and highly digestible food products for human nutrition. In other words, the foods should contain minimal amounts of residual antinutritional factors, and the digestibility of the protein should be a good approximation of the bioavailability of individual amino acids.

Analyses Required for the Determination of PDCAAS

Proximate composition Levels of moisture, total nitrogen, fat, and crude protein are determined according to standardized AOAC methods. Total protein is then calculated by using a nitrogen-to-protein conversion factor of 6.25. When the true nitrogen-to-protein conversion factor is known for a given protein, this should be used.

Amino acid profile Amino acids in foods have commonly been analyzed by ion-exchange chromatography of protein hydrolysates. Hydrolysis with 6 M HCl at 110°C for 24 h is routinely used for all amino acids except tryptophan and sulfur amino acids. Since sulfur amino acids and tryptophan are destroyed to varying degrees by 6 M HCl hydrolysis, a performic acid followed by the HCl hydrolysis and an alkaline (4.2 M NaOH) hydrolysis are widely used for the accurate determination of sulfur amino acids (methionine as methionine sulfone and cysteine/cystine as cysteic acid) and tryptophan, respectively. Amino acids may also be determined accurately and more rapidly by high-performance liquid chromatography methods using precolumn phenylisothiocyanate-derivatized amino acids. Amino acids should be expressed on the basis of mg per gram of protein.

Amino acid score Amino acid ratios (milligrams of an IAA in 1.0 g of test protein per milligram of the same IAA in 1.0 g of reference protein × 100) for nine IAA are calculated by using the FAO/WHO's suggested pattern of amino acid requirements as the reference protein for all ages except for infants. This pattern contains, in milligrams per gram of protein: histidine, 19; isoleucine, 28; leucine, 66; lysine, 58; methionine + cystine, 25; phenylalanine + tyrosine, 63; threonine, 34; tryptophan, 11; and valine, 35. For infants under 1 year of age, the amino acid composition of human milk is used as the basis for scoring: (milligrams per gram of protein): histidine, 26; isoleucine, 46; leucine, 93; lysine, 66; methionine + cystine, 42; phenylalanine + tyrosine, 72; threonine, 43; tryptophan, 17; and valine, 55. The lowest amino acid ratio is termed the amino acid score.

Protein digestibility The classic method for determining the digestibility of proteins and amino acids has been the balance method, in which the protein or amino acid excreted in the feces is subtracted from the amount ingested and the value expressed as a percentage of protein intake. This calculation gives an apparent digestibility value. To determine true digestibility, it is necessary to correct for the amount of fecal protein or amino acid excreted when the subject or animal is consuming either a protein-free diet or a diet with just enough of a highly digestible protein to prevent excessive loss of body protein. Therefore, the true digestibility of protein and amino acids can be calculated by the following equations:

$$\text{True protein digestibility} = [\text{PI} - (\text{FP} - \text{MFP})/\text{PI}] \times 100,$$

where PI is the protein intake, FP is fecal protein, and MFP is metabolic fecal protein.

$$\text{True amino acid digestibility} = [\text{AAI} - (\text{FAA} - \text{MFAA})/\text{AAI}] \times 100,$$

where AAI is the amino acid intake, FAA is fecal amino acid, and MFAA is metabolic fecal amino acid.

The amounts of protein and amino acids in the feces of rats fed the protein-free diet are used as the estimates for MFP and MFAA, respectively.

Since true digestibility measurements take into account the fecal protein or amino acids which are not of dietary origin, the true digestibility of a protein or amino is always higher than the apparent digestibility. Unlike apparent digestibility, the true protein digestibility of a food is not affected by the dietary conditions under which the food is fed to the animal such as the protein content of the test diet. It is a more accurate measure of amino acids that are absorbed from the gut and therefore gives a better representation of protein quality than apparent digestibility.

Protein/amino acid digestibility is most frequently determined using rats. The rat balance method is well established, and the procedure for the determination of true protein digestibility has been standardized. It has been recommended by the FAO/WHO that when human balance studies cannot be used, the standardized rat fecal-balance method should be used. The true digestibility of protein is considered to be a reasonable approximation of the true digestibility of most amino acids (as determined by the rat balance method) in mixed diets. Therefore, the FAO/WHO has recommended that amino acid scores be corrected only for true digestibility of protein.

Protein Digestibility-corrected Amino Acid Score (PDCAAS)

The PDCAAS of a test food is calculated as follows:

$$\text{True protein digestibility coefficient} \times \text{amino acid score}.$$

PDCAAS above 100 are considered as 100.

Protein Digestibility of Some Diets and Common Foods

Values for the true digestibility of protein in diets from India (54–75%), Guatemala (77%), and Brazil (78%) are markedly lower than the values in North American diets, including vegetarian diets, 88–94%, suggesting that protein digestibility is of greater concern in diets of some developing countries. The poor digestibility of protein in the diets of some developing countries is due to the use of less-refined cereals and pulses (such as beans and lentils). Low true digestibility values (63–65%) have also been reported in studies with children fed millet and ragi-based diets in India.

A comparison of the true digestibility of protein in some common foods for human adults has indicated that animal protein sources (meat, fish, poultry, eggs, properly processed milk protein products) and flours and breads made from low-fiber wheats, wheat gluten, farina, peanuts, and soy-protein isolates have high true protein digestibilities of 94–99%. Whole corn (except high amylose containing opaque-2) and flour or bread of high-fiber wheat, polished rice, oatmeal, triticale, cottonseed, soy flour, and sunflower have intermediate true protein digestibilities of over 85%. Ready-to-eat cereals (corn, wheat, rice, or oat) have lower true protein digestibilities of 70–77%, probably due to the high heat involved in their processing. Various types of dry beans (*Phaseolus vulgaris*) and millet also have low true protein digestibilities of 75–79%.

PDCAAS for Some Common Foods or Food Products

The availability of amino acid composition and protein digestibility data for a number of foods has made it possible to calculate their PDCAAS. Animal protein products such as egg white powder, casein, ground beef, beef salami, skim milk powder, tuna and chicken frankfurters have high PDCAAS of about 100 (97–100). Soybean protein products also have high PDCAAS (89–99) and are marginally deficient in sulfur amino acids in some cases. The scores for various types of beans, lentils, and peas range from 47 to 71, and these products are first limiting in sulfur amino acids for human nutrition. Wheat gluten and sunflower-protein isolate are severely limiting in lysine and have PDCAAS of 25 and 37, respectively.

Effects of Supplementation and Complementation on Protein Quality

The low protein quality of a vegetable protein source can be improved by the addition of supplementary protein or a limiting amino acid, or by protein complementation. Protein supplementation means to increase the protein quality of a food having a low-quality protein by the addition of a moderate amount of another food having a high content of the indispensable amino acid which is limiting in the low-quality protein. For example, egg, milk, meat, and fish proteins are rich in lysine and sulfur amino acids, and so they would be supplementary for lysine deficient proteins such as cereals, and for proteins deficient in sulfur amino acids such as beans, peas, and lentils. These animal proteins can also fully satisfy all indispensable amino acid requirements for adults when ingested as the sole protein source at a level of 0.7 g per kilogram of body weight per day.

Protein complementation means that the quality of two proteins is higher than either of the two components. Proteins which complement each other have different limiting amino acids. When blended in the right proportions, cereals and legumes (beans and peas) complement each other because cereals are limiting in lysine, and legumes are deficient in sulfur amino acids. As an example, the BV of pea flour is 43% and is limited by methionine, with a surplus of lysine. Maize meal protein with a BV of 35% is limited by lysine but with a surplus of sulfur amino acids. The average of the BV of these two protein sources is 39%. In practice, however, the BV of an equal mixture of these two proteins sources is found to be 70% due to the complementary effects. In the same way, traditional combinations of vegetable proteins consumed in some countries, such as rice–chick peas in Asia, wheat–bean in the Near East, and maize–bean in the Americas, have a good protein quality.

Applications of Protein-quality Measurement

There are two distinct purposes for measuring protein quality: to determine whether a food or diet will meet human amino acid needs, and for regulatory purposes to ensure that foods are of the appropriate protein quality for the intended use and that protein claims are valid.

Protein-quality Measurement of Mixed Diets

It is recognized that the consumption of mixed diets in the developed world poses no problem of inadequacy in protein content and quality. However, the food supply is undergoing many changes, which may significantly alter the types and quality of protein available. Many new foods are being produced for various purposes through genetic modification. Increasing numbers of foods consumed in the West are processed, and new methods of processing are constantly being developed. Products of the plant and animal world that have not been previously used as human food are being developed for that purpose. These situations point to a continuing need for surveillance of potential effects on protein quality of foods in the diet, if not the diet as a whole.

Some subgroups in developed countries consume protein from a single source, rely on a small variety of foods, or eat foods with proteins of inferior quality. To the extent that these people do not have control over the source of the proteins in their foods because they may be relying on commercially formulated products, they must be considered at risk, and the quality of their dietary proteins must be known. Examples of such groups are infants and young children relying on commercial formulas; people using formulated products to lower their energy/food intake, generally for weight reduction; people changing their eating patterns to become vegetarian, for example, and possibly relying on prepackaged substitute foods of lower protein quantity and quality; and individuals with various illnesses requiring therapeutic formulated diets. There are also subpopulations who, for economic or other sociocultural reasons, might become at risk because of their decreased ability to procure high-quality proteins or because of making uninformed dietary choices. These might include the poor, the elderly, teenagers who are undergoing rapid growth and often eating unbalanced diets, and pregnant and lactating women, especially the growing number of teenage mothers. Depending on the issue of concern, either protein-quality assessment or protein-status assessment will have an application.

Regulatory Use of Protein-quality Assessment of Foods for Human Nutrition

Methods currently used by regulatory bodies Since the adoption of the PDCAAS method by the FAO/WHO, this method has been widely accepted. For example, in the USA, PDCAAS has replaced PER as the official method of protein-quality assessment, including for protein declaration in nutritional labeling of all foods except infant formulas. However, PER remains the official regulatory biological method for protein-quality assessment of infant formulas sold in USA. PER also remains the official method for evaluating protein claims for foods sold in Canada. The chemical index (amino acid score using human breast milk as reference protein) is used for assessing the protein quality of infant formulas in the UK. The Codex Alimentarius Commission requires that the quality of protein in an infant formula not be less than 85% of that of casein as determined by the PER method. The EC requires that, for an equal energy value, an infant formula must contain an available quantity of each essential and semiessential amino acid at least equal to that contained in the reference protein (breast milk), but for calculation purposes, the concentrations of methionine and cystine may be added together. For infant formulas based on partial protein hydrolysates, the EC has an additional requirement, specifying that the PER and the NPU must be at least equal to those of casein.

Limitations of the PDCAAS As adopted by the FAO/WHO, the PDCAAS method is truncated in that it requires the adjustment of all values that

exceed 100 down to 100. This means that, unlike biological assays such as PER, the PDCAAS does not recognize the additional value of high-quality animal proteins (*see* **Protein**: Sources of Food-grade Protein) and their potential for acting as supplements to low-quality proteins.

Since the PDCAAS method was adopted, it has been found that there are differences in the metabolic fate and retention of dietary nitrogen from different protein sources when measured acutely during the postprandial period in humans. One method, called 'net postprandial protein utilization,' is designed to measure the efficiency of postprandial nitrogen retention following consumption of various dietary proteins. Results obtained so far have placed several food protein sources in the order: milk > soy > pea > wheat, with values of 120 (sulfur amino acids), 99 (sulfur amino acids), 73 (sulfur amino acids), and 36 (lysine), respectively. This order is the same as that of the nontruncated PDCAAS values for these sources, suggesting that there may be valid reasons to use the nontruncated PDCAAS values. The official PDCAAS method puts both milk and soy at 100.

PDCAAS includes a rat-feeding study, but this only measures protein digestibility. Thus, it may not adequately reflect the adverse effect on the bioavailability and utilization of amino acids caused by naturally occurring growth-depressing factors such as glucosinolates and isothiocyanate in mustard flour, trypsin inhibitors, and hemagglutinins in soy protein and beans (*Phaseolus* spp.), or antinutritional factors formed during processing such as lysinoalanine in alkaline-/heat-processed lactalbumin and soy protein. As a result, the PDCAAS method would overestimate the protein quality of such sources. In the absence of upper safe limits for these antinutritional factors, the use of the PDCAAS method may be inappropriate for predicting the protein quality of such protein products and any foods where these products may be the sole source of protein, such as infant formulas and enteral nutritionals.

The PDCAAS method assumes complete biological efficiency of supplemental amino acids in improving protein quality which may not be true, especially in the case of low-quality proteins. Markedly lower relative protein efficiency ratio or RNPR values have been seen with amino acid-supplemented zein (3–44%) compared to the PDCAAS value (71%), which suggests incomplete biological efficiency of the supplemental amino acids and overestimation of protein quality by the PDCAAS method. The crystalline amino acid added to a low-quality protein that also has poor digestibility may be absorbed more quickly than the amino acids from the protein, so protein synthesis does not take place.

The PDCAAS method does not accurately predict the quality of proteins colimiting in more than one IAA such as soy when used in infant formulas. The amino acid requirements for infants are not the same as the FAO/WHO's suggested pattern of amino acid requirements used in PDCAAS. Since the identification of the true first limiting amino acid would be difficult in cases where the protein is colimiting in more than one IAA, an improvement in protein quality with supplementary amino acid(s) should not be assumed without biological testing.

Recommended method for regulatory purposes The PDCAAS is the most appropriate routine method for properly processed food where protein digestibility is a good approximation of bioavailability of the amino acids and where there is known to be no or only low levels of antinutrient factors. Where these conditions do not apply, a biological method should be used, whether this is PER, RNPR, or relative nitrogen utilization (RNU). Probably the least expensive and yet still valid and reliable method, and one that gives results that are proportional across a range of protein qualities, is the RNU method. The RNU method gives essentially the same values as the better known RNPR rat assay for routine protein-quality assessment of foods except that, as described in the section Rat Assays, it replaces measurement of protein needed for maintenance with a factor for maintenance. The elimination of the need for feeding a 'zero' protein diet to a group of rats makes the RNU method less costly than the RNPR method.

See also: **Protein**: Sources of Food-grade Protein

Further Reading

Bodwell CE, Adkins JS and Hopkins DT (1981) *Protein Quality in Humans: Assessment and* in vitro *Estimation*. Westport, CT, AVI.

FAO/WHO (1991) *Report of the Joint FAO/WHO Expert Consultation on Protein Quality Evaluation. FAO Food and Nutrition Paper 51*. Rome: Food and Agriculture Organization of the United Nations.

Friedman M (1996) Nutritional value of proteins from different food sources. A review. *Journal of Agricultural and Food Chemistry* 44: 6–29.

Pellett PL and Young VR (1980) *Nutritional Evaluation of Protein Foods. UNU Food and Nutrition Bulletin Supplement*. Tokyo: United Nations University.

Reeds P, Schaafsma G, Tomé D and Young V (2000) Summary of the workshop and recommendations. *Journal of Nutrition* 130: 1874S–1876S.

Sarwar G (1987) Digestibility of protein and bioavailability of amino acids in foods. *World Review of Nutrition and Dietetics* 54: 26–70.

Sarwar G (1997) The protein digestibility-corrected amino acid score method overestimates quality of proteins containing antinutritional factors and of poorly digestible proteins supplemented with limiting amino acids in rats. *Journal of Nutrition* 127: 758–764.

Sarwar G and McDonough FE (1990) Evaluation of protein digestibility-corrected amino acid score method for assessing protein quality of foods. *Journal of the Association of Official Analytical Chemists* 73: 347–356.

Tomé D and Bos C (2000) Dietary protein and nitrogen utilization. *Journal of Nutrition* 130: 1868S–1873S.

Digestion and Absorption of Protein and Nitrogen Balance

P W Emery, King's College London, London, UK

Background

The assimilation of protein into the body has classically been viewed as involving digestion of protein to amino acids in the gut, absorption of these amino acids through the gut mucosa, and transport of amino acids in the plasma for subsequent metabolism, including protein synthesis, in other tissues. However, it is now clear that substantial quantities of peptides containing up to five amino acid residues can also be absorbed into the intestinal mucosa. Here they are largely hydrolyzed to free amino acids before being released into the circulation, so that the circulating concentration of peptides is normally very low, although many other tissues are capable of taking up and utilizing peptides. There is also growing evidence that very small but detectable amounts of whole proteins can be absorbed. It should also be noted that the bacterial flora of the colon can play a significant part in the supply of nitrogenous molecules to the body.

Digestion

The digestion of protein is a series of hydrolytic reactions which split the α-peptide bonds between adjacent amino acid residues in the primary sequence of the polypeptide chain. Peptide bonds involving the side chains of amino acids (the γ- and β-carboxyl groups of glutamate and aspartate and the ε-amino group of lysine) are not hydrolyzed by human enzymes and are thus not biologically available.

Protein digestion begins in the stomach with the action of pepsins, which are hydrolytic enzymes which preferentially cleave peptide bonds adjacent to phenylalanine, tyrosine or leucine residues. The human stomach secretes at least three distinct pepsins with different pH optima ranging from almost neutral, corresponding to the condition of the stomach contents immediately after a meal, to pH 1.2, as would be found several hours later.

Further digestion then takes place in the small intestine, under the influence of enzymes which originate in the acinar cells of the exocrine pancreas. These are mainly endopeptidases which hydrolyze particular classes of peptide bonds within the polypeptide chain (see **Table 1** for their specificity) to produce a large number of small-peptide fragments.

Table 1 Digestive proteolytic enzymes

Enzyme	*Proenzyme*	*Source*	*Substrate*	*Specificity*
Endopeptidases				
Pepsin A	Pepsinogen A	Stomach	Polypeptides	Hydrophobic/aromatic amino acids
Pepsin B (parapepsin)	Pepsinogen B	Stomach	Gelatin	
Pepsin C (gastricin)	Pepsinogen C	Stomach	Polypeptides	Tyrosine residues
Trypsin	Trypsinogen	Pancreas	Polypeptides	Arginine, lysine residues
Chymotrypsin	Chymotrypsinogen	Pancreas	Polypeptides	Tyrosine, phenylalanine, tryptophan, leucine residues
Elastase	Proelastase	Pancreas	Elastin and similar proteins	Alanine residues
Elastase II	Proelastase II	Pancreas	Elastin and similar proteins	Leucine, phenylalanine, methionine residues
Enteropeptidase		Small intestine	Trypsinogen	$Lysine^{6}$-isoleucine
Exopeptidases				
Carboxypeptidase A	Procarboxypeptidase A	Pancreas	Peptides	C-terminal except aspartate, glutamate, arginine, lysine, proline
Carboxypeptidase B	Procarboxypeptidase B	Pancreas	Peptides	C-terminal arginine, lysine
Aminopeptidase		Small intestine	Peptides	N-terminal except arginine, lysine
Tripeptidases		Small intestine	Tripeptides	Mainly N-terminal
Dipeptidases		Small intestine	Dipeptides	Various

The final stage of protein digestion is accomplished by a wide array of aminopeptidases and di- and tripeptidases manufactured in the absorptive cells of the intestinal epithelium. Some of these act in the lumen of the gut, but most of the activity is located within the brush-border membrane or within the cytoplasm of the cells. These are mainly exopeptidases, which remove single amino acids from the N-terminal (aminopeptidases) or the C-terminal (carboxypeptidases).

Both pepsin and the pancreatic proteolytic enzymes are synthesized in the form of inactive proenzymes (also termed zymogens) containing one or more terminal peptide regions which prevent the formation of the three-dimensional structure that confers catalytic activity. The terminal peptide of pepsinogen (containing 42 amino acids) is hydrolyzed by hydrochloric acid in the stomach. Trypsinogen is activated by the removal of a hexapeptide from the amino-terminal by enteropeptidase (formerly known as enterokinase), an enzyme secreted by the small intestine. Trypsin in turn activates all the pancreatic proenzymes, including itself. Hence the presence of trypsin inhibitors in some legumes plays a crucial role in reducing the digestibility of these otherwise valuable sources of protein.

Absorption

Absorption of the products of digestion involves active transport into the enterocyte, metabolism within the cell, and diffusion down a concentration gradient into the portal circulation. Absorption occurs throughout the length of the small intestine, but is confined to the cells in the top third of the villus.

Amino acids and peptides are transported into cells by a number of active transport systems, most of which are linked to the simultaneous entry of sodium ions. Sodium enters the cell down a chemical concentration gradient; this is maintained by the membrane 'pump' which exchanges sodium for potassium ions and is fueled by the hydrolysis of adenosine triphosphate (ATP). The specificity of the amino acid transport systems in the gut is broadly similar to that of the systems which are found in other tissues, particularly the brush border of the kidney tubule, and this is summarized in Table 2. Evidence from *in vitro* studies suggests that small peptides may be transported more rapidly than free amino acids, but these studies often have not been carried out under the conditions which occur *in vivo* after a normal meal.

After transport across the intestinal wall, amino acids can be degraded, metabolized to other compounds, including other amino acids, incorporated into proteins, or released unaltered into the circulation. The rate of protein synthesis in the small intestinal mucosa is amongst the highest of any tissue in the body. This is no doubt related to the high rate of cell turnover and the requirement for secretion of digestive enzymes. These cells also utilize energy at a high rate, and the amino acid glutamine, supplied both from the lumen of the gut and from the circulation, is a major fuel. Some of the ammonia which is formed during the utilization of glutamine is released into the gut, from where it may subsequently be reabsorbed and transported to other tissues to be used for the synthesis of nonessential amino acids.

Most of the peptides which are absorbed into the mucosal cells are hydrolyzed to single amino acids before being released into the portal circulation. However, some intact peptides have also been observed to enter the portal circulation, and this has raised some interest because of the possibility of peptides with hormonal activity being absorbed intact.

Table 2 Intestinal amino acid transport systems

System	*Sodium dependence*	*Preferred amino acids*
A	Yes	Alanine, serine, glycine, methionine, proline
L	No	Leucine, isoleucine, valine, methionine, phenylalanine, tyrosine, tryptophan, histidine
ASCP	Yes	Alanine, serine, cysteine, proline
Ly	Yes	Lysine, histidine, arginine, ornithine
x^-_A	Yes	Aspartate
x^-_G	Yes	Glutamate
x^-_C	No	Aspartate, glutamate, cystine
y^+	Yes	Lysine, arginine, histidine
β	Yes	β-alanine, taurine
$b^{0,+}$	No	Lysine, leucine
Gly	Yes	Glycine, sarcosine
N	Yes	Histidine, glutamine, asparagine
imino	Yes	Proline

The circulating concentration of peptides is normally very low. This may suggest that only small quantities of peptides are released into the portal circulation. However, it has also been observed that when peptides are infused directly into the circulation they are rapidly metabolized, probably extracellularly, since the plasma concentration of the free amino acid constituents of the peptides rises almost immediately. This means that glutamine, an amino acid which is notoriously unstable during storage in solution, and sparingly soluble amino acids such as tyrosine can be supplied in parenteral nutrition mixtures in the form of stable and soluble peptides which are readily available to the tissues.

Very small quantities of whole proteins can also be absorbed directly from the intestinal lumen, possibly by diffusion through tight junctions between cells. There may also be some endocytosis of food proteins by immunologically active cells within the intestinal epithelium, particularly in the areas known as Peyer's patches, and these proteins may subsequently be transported by lymphocytes to distant sites. The absorption of intact proteins from maternal milk makes an important contribution to the development of passive immunity by neonates. Where it persists into later life it is likely to be a major factor in the development of food allergies.

It is now known that some small molecules can be absorbed from the colon as well as the small intestine, so that the products of bacterial metabolism can be an important source of nutrients. In the case of nitrogenous compounds the main process appears to be secretion of urea into the colon, bacterial hydrolysis of this urea to ammonia, and subsequent absorption of the ammonia into the body. The ammonia may then be used for the synthesis of amino acids by transamination. This process is referred to as urea salvage, since it results in rather less urea being excreted in the urine than has been produced by the liver. This may make an important contribution to the maintenance of nitrogen balance when protein intake is low. There is also some evidence that essential amino acids synthesized by colonic bacteria may also be absorbed into the body, but the quantitative significance of this observation is not clear.

Amino Acid Pools

Amino acids are transported in the plasma and are taken up into tissues by mechanisms similar to the systems referred to in Table 2. Amino acids are also present in blood cells, but the interaction of these amino acids with other tissues may be quite different from that of plasma amino acids. For example, the plasma amino acid concentration increases as blood traverses the gastrointestinal tract after a meal, whereas the amino acid content of blood cells decreases.

The intracellular free amino acid content varies between different tissues, and is affected by hormonal and nutritional influences. However, the intracellular concentration is greater than the plasma concentration for most amino acids in most tissues. The distribution ratio (intracellular concentration divided by plasma concentration) is close to unity for many of the essential amino acids, particularly in muscle, but is severalfold higher for some of the nonessential acids, with a value as high as 70 having been reported for aspartic acid in liver. The interpretation of changes in amino acid concentrations in particular situations is complex. Changes in concentration may result from changes in outflow to or inflow from other tissues, or pools of other metabolites in the same or different tissues, including utilization for protein synthesis and supply from protein degradation.

Only about 2% of the amino acids in the body are present as free amino acids, the rest being present as protein. Again there is considerable variation between different amino acids and different tissues in the ratio of the free to protein-bound amino acids, with much lower values generally being found for essential than nonessential amino acids. Given the currently accepted values for rates of protein synthesis, it can be shown that in some tissues the entire free pool of some essential amino acids is incorporated into protein within seconds, emphasizing the rapidity with which free amino acids must be resupplied.

Nitrogen Balance

The state of protein metabolism in the body as a whole may be assessed by measuring nitrogen balance. This is because almost all the nitrogen in the body is in the form of protein, and the nitrogen content of a wide range of proteins is relatively constant, at around 16%. Thus, if the body is in positive nitrogen balance, it must be laying down new protein, and this is normally associated with growth in children or recovery from illness. Conversely, if the body is in negative nitrogen balance, it must be losing tissue protein, either because the diet is inadequate or because of a pathological response to injury or illness. Nitrogen balance in an individual will fluctuate from day to day by a few grams either side of zero, but over a period of weeks a nongrowing adult will normally be in zero nitrogen balance.

Every protein in the body has a specific functional role. These roles include the catalytic function of enzymes, the contractile function of muscle proteins, the structural role of connective tissue proteins, roles

in transport, immunological recognition, hormone receptors, and many others. There is no inert store of protein to be drawn on in times of need, as exists for fat (adipose tissue triacylglycerol) and carbohydrate (glycogen); any sustained period of negative nitrogen balance is likely to have functional consequences. Nevertheless, different tissues do lose proteins at different rates when the body goes into negative nitrogen balance, with muscle protein (including the smooth muscle of the gut) being lost in the early stages while the liver and the heart are relatively protected.

Measurement of Nitrogen Balance

Nitrogen balance is defined as the difference between intake and output, and may be formally represented by the following equation:

$$\text{Balance} = I - (U + F + M)$$

where I is nitrogen intake, U is urinary nitrogen excretion, F is fecal nitrogen excretion, and M is the sum of all the other routes by which nitrogen is lost from the body.

Nitrogen intake is mainly in the form of protein and can conveniently be measured by the Kjeldahl method, which measures virtually all nitrogen except that bonded to oxygen (e.g., nitrates and nitrites). The same method is thus also appropriate for measurement of the various components of nitrogen excretion. Accurate determination of nitrogen balance requires direct analysis of duplicate portions of diet, although a reasonable approximation can be made using a weighed-food record and tables of food composition. An alternative approach is to feed the subjects a constant amount of a diet which has been specially formulated using known quantities of ingredients.

The urine is the major route for excretion of nitrogen, so that urinary nitrogen is the major determinant of nitrogen balance and reliable collection of 24-h urine specimens is required for measuring nitrogen balance. Most of the nitrogen in urine is in the form of urea, which is the end product of amino acid oxidation. Other nitrogenous compounds in the urine include ammonia, creatinine, and uric acid, as well as small quantities of peptides, amino acids, and other small molecules. The amount of ammonia depends on the body's acid–base status and the amount of uric acid varies with nucleic acid intake. The amount of creatinine is fairly constant from day to day within a healthy individual, and shows a good correlation with muscle mass. When protein intake is low, the proportion of urea in the urine will decrease, so that urinary urea content is not a reliable indicator of total urinary nitrogen.

Fecal nitrogen is mainly composed of bacterial cells from the large intestine, together with mucosal cells which have been sloughed off from the intestinal wall, and some remnants of undigested food proteins. On a normal mixed diet fecal nitrogen amounts to a fairly constant 8% of nitrogen intake, although on very-high-fiber diets and particularly those containing large quantities of legumes this proportion may increase. In the measurement of nitrogen balance it is conventional to use a nonabsorbable marker to mark the beginning and end of the fecal collection, or to use a continuous fecal marker.

The other routes by which nitrogen is lost from the body include the shedding of hair, nails and dead skin cells, sweat, saliva, semen, blood lost during menstruation or removed for clinical testing, and the ammonia exhaled in the breath. Dermal losses vary with the rate of sweating and with protein intake. Many of these losses are small and hard to measure, so that a single figure of 0.5 g of nitrogen per day may be taken to approximate the sum of these miscellaneous routes.

The most important use of nitrogen balance is in evaluating the adequacy of dietary intake. It is the major criterion which is used to assess quantitative protein requirements. Such studies have demonstrated that energy intake is also a major determinant of nitrogen balance. Measurement of nitrogen balance is also the basis of many biological methods for measuring protein quality.

Amino Acid Imbalance and Antagonism

It is normally the case that once the minimum requirements for total nitrogen and for each essential amino acid have been met, any surplus protein intake has no deleterious effect on nitrogen balance. However, there are reports of the addition of a very large quantity of a single essential amino acid causing a reduction in the biological value of an otherwise adequate protein. This has been called amino acid imbalance and appears to occur only when protein intake is marginally adequate. It can thus be overcome by increasing the intake of either total protein or of the limiting amino acid.

The related phenomenon of amino acid antagonism refers to a situation in which the addition of one amino acid affects the efficiency of utilization of another related amino acid. For example, the addition of one of the branched-chain amino acids (leucine, isoleucine, or valine) will impair the utilization of dietary protein, and this can be overcome by adding the other two branched-chain amino acids. It is tempting to speculate that this is because of competition for membrane transport, or because of increased activity of the common pathway by which these

amino acids are oxidized, but no such mechanism has yet been confirmed.

See also: **Amino Acids**: Properties and Occurrence; Determination; Metabolism; **Enzymes**: Functions and Characteristics

Further Reading

Alpers DH (1986) Uptake and fate of absorbed amino acids and peptides in the mammalian intestine. *Federation Proceedings* 45: 2261–2267.

Bender DA (1985) *Amino Acid Metabolism*, 2nd edn. Chichester: John Wiley.

Bender DA and Bender AE (1997) *Nutrition: A Reference Handbook*. Oxford: Oxford University Press.

Calloway DH, Odell ACF and Margen S (1971) Sweat and miscellaneous nitrogen losses in human balance studies. *Journal of Nutrition* 101: 775–786.

Gardner MLG (1984) Intestinal assimilation of intact peptides and proteins from the diet – a neglected field? *Biological Reviews* 59: 289–331.

Jackson AA (1995) Salvage of urea-nitrogen and protein requirements. *Proceedings of the Nutrition Society* 54: 535–547.

Waterlow JC, Garlick PJ and Millward DJ (1978) *Protein Turnover in Mammalian Tissues and the Whole Body*. Amsterdam: North Holland.

Webb KE (1986) Amino acid and peptide absorption from the gastrointestinal tract. *Federation Proceedings* 45: 2268–2271.

Yudilevitch DL and Boyd CAR (eds) *Amino Acid Transport in Animal Cells*. Manchester: Manchester University Press.

Synthesis and Turnover

P W Emery, King's College London, London, UK

Background

It is self-evident that growth must involve the synthesis of new protein. Equally, there are situations, usually associated with illness or undernutrition, when the organism loses protein, indicating that a mechanism must also exist for degrading tissue proteins. However, not until the 1940s was there direct evidence that the processes of protein synthesis and degradation continue to occur in healthy nongrowing adult animals which are neither gaining nor losing weight. The amino acid tyrosine, labeled with the stable isotope ^{15}N, was fed to an adult rat, and after 10 days only half the isotope had been excreted in the urine while most of the rest had been incorporated into tissue proteins. Protein turnover includes both the synthesis of protein from amino acids and the breakdown of protein to amino acids. Much experimental work over the last 40 years has been devoted to measuring the rates at which these two processes occur and identifying the factors which modulate them. It is now clear that both processes occur continuously and simultaneously, and it is estimated that an adult man synthesizes and degrades about 300 g of protein each day, which is three or four times as much as is consumed in the diet.

Mechanism of Protein Synthesis

The major biochemical pathways by which protein is synthesized in cells are now well established. Information specifying the primary structure is carried in the sequence of base pairs in the DNA molecules in the nucleus. This information is transcribed by the formation of short-lived molecules of messenger ribonucleic acid (mRNA) with base sequences complementary to those of the deoxyribonucleic acid (DNA). The mRNA moves out of the nucleus into the cytosol, where it associates with ribosomes to allow translation of the message as a sequence of amino acids. Every three bases on the mRNA specify a single amino acid which then binds to the ribosome in association with a specific molecule of transfer RNA (tRNA). A peptide bond is formed to join the new amino acid to the growing polypeptide chain, and the deacylated tRNA is released from the ribosome.

Even when the completed polypeptide has been released from the ribosome it may undergo considerable posttranslational modification. Some amino acid residues may be hydroxylated or methylated, while others may form covalent links with other parts of the molecule as the protein adopts a stable three-dimensional structure. Nonprotein molecules, particularly carbohydrates, may be added. Finally, parts of the protein may be removed altogether, possibly after being transported to a different compartment of the cell, to allow activation of an inactive precursor or to facilitate secretion of export proteins.

Regulation of the rate of protein synthesis within a given cell occurs at two levels. There is specific control of the types of protein synthesized, which occurs mainly by controlling the transcription of particular genes. There is also general control of the total amount of protein synthesized, which occurs mainly at the level of translation. This includes regulation of the number of ribosomes present in the cell, and hence its capacity for protein synthesis, as well as regulation of the efficiency of translation, which

appears to be achieved largely by changes in the rate of initiation of peptide synthesis, although elongation and termination may also be modulated under some circumstances.

Protein Degradation

The mechanisms by which proteins are broken down are much less well understood. A large number of proteolytic enzymes have been identified, with differing substrate specificities and conditions for optimal function. This may reflect the complexity and variety of the substrates on which they operate, and may also indicate a degree of cooperative activity, since once a protein is committed to degradation it appears to be broken down to its constituent amino acids very rapidly.

Many of the proteolytic enzymes appear to operate within a specific organelle, the lysosome. This provides the acidic environment in which enzymes such as the cathepsins are optimally active. Lysosomes can engulf and degrade large structures such as whole mitochondria, but individual soluble proteins can also enter the lysosome from the cytosol.

Lysosomes are believed to be responsible for the majority of protein degradation, particularly in tissues such as the liver where the overall rate of protein turnover is high. However, there are also a number of proteinases which operate in the cytosol, and are active at neutral or even alkaline pH. These include the following: the calpains, which are activated by calcium ions and may be particularly active in muscle; a large multicatalytic proteinase sometimes called a proteosome; and a series of enzymes which commit proteins to degradation by binding them to a 76-amino-acid polypeptide called ubiquitin.

There also exist extracellular proteinases which degrade specific extracellular proteins such as collagen.

The control of protein degradation may be exercised at both a general and a specific level, as with protein synthesis. Some of the factors that affect the rate of protein degradation are intrinsic to the protein substrate itself, for example, increasing size and acidic nature favor more rapid degradation. At least some protein degradation appears to require energy. There also appears to be a requirement for continuing protein synthesis, although this may simply reflect the fact that the enzymes which degrade proteins are, of course, proteins themselves. However, this association may at least partly explain the observation that rates of protein synthesis and degradation often change in the same direction in response to external influences.

Measurement of Protein Synthesis and Breakdown

Having identified the pathways by which proteins are synthesized and degraded it is desirable to measure the rates at which these processes occur. Such measurements can be made *in vivo*, in humans or experimental animals, or *in vitro*, in perfused or incubated tissues or organs, cell cultures, or cell-free systems. Unfortunately there are quantitative differences between the results obtained from different systems, which limits the usefulness of *in vitro* measurements. For example, rates of protein synthesis measured in incubated muscle preparations tend to be lower, and rates of protein breakdown are always considerably higher than those observed *in vivo*. This occurs despite careful attention being paid to factors which are known to affect the viability of the model such as tissue oxygenation, muscle integrity, stretch and electrical stimulation, as well as nutrient supply and hormonal environment. Hence such preparations are always in net negative protein balance, whereas *in vivo* muscle switches between negative balance, in the postabsorptive state, and positive balance after a meal.

The ultimate aim is to be able to measure the turnover rates of individual proteins. This can be done *in vitro*, and it may soon be possible to make such measurements *in vivo* as well. However, for technical reasons the methods that are currently available for use *in vivo* are only able to give reliable data for turnover rates in whole tissues or even the whole body. Nevertheless, such data do provide a useful insight into the mechanisms by which nutritionally significant quantities of protein are gained or lost, and such measurements are being made increasingly frequently in both clinical and agricultural research.

Measurement of Whole-Body Protein Turnover

The measurement of protein turnover necessitates a model which is simple enough to analyze fully but can be reasonably interpreted in terms of what happens in a living animal or person. The model used most commonly is the two-pool model, shown in Figure 1. All the protein in the body is represented by one pool, while the other pool represents the total amount of one particular free amino acid (both intracellular and extracellular). The flux of that amino acid (Q) is defined as the sum of all the processes by which the amino acid leaves the free pool:

$$Q = O + S + M$$

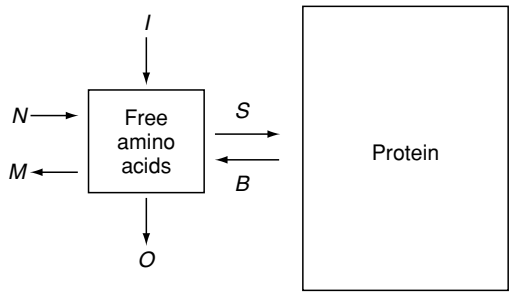

Figure 1 Two-pool model for investigating whole-body protein turnover. *I*, intake of amino acid from dietary protein; *S*, protein synthesis; *B*, protein degradation; *O*, amino acid oxidation; *N*, synthesis of amino acid; *M*, metabolism to other compounds.

In this equation O is the rate at which the amino acid is oxidized, S is the rate at which it is incorporated into protein, and M is the rate at which it is metabolized to other compounds. Measurements are normally made in a steady state, when the amino acid pool is neither expanding nor contracting, so that flux is also equal to the sum of all the processes by which the amino acid enters the pool:

$$Q = I + B + N$$

where I is the intake of the amino acid from dietary protein, B is the rate at which it is released by protein breakdown, and N is the rate at which it is synthesized. Judicious choice of amino acid allows these equations to be simplified: for essential amino acids, $N = 0$, and several of the amino acids have no metabolic fates other than oxidation, so that $M = 0$. Dietary intake can be controlled during the period of measurement, and is often set to zero. Thus, if the rates of flux and oxidation can be measured, the rates of incorporation of amino acid into protein (i.e., protein synthesis) and release from protein (i.e., protein degradation) can be calculated. Knowledge of the proportion of this amino acid in tissue proteins then allows the overall rates of protein synthesis and degradation to be calculated.

There are two commonly used methods for measuring amino acid flux. One involves the administration of a single dose (d) of an amino acid (often glycine) labeled with ^{15}N and measurement of the cumulative urinary excretion of label in an end product of amino acid metabolism, usually either urea or ammonia, over a period of up to 60 h. The flux is calculated from the following equation:

$$Q = d(E/et)$$

In this equation e is the cumulative excretion of the labeled end product and E is the total excretion rate of that end product. Oxidation rate can be estimated from the rate of excretion of urea nitrogen expressed as a proportion of the nitrogen flux. This method is noninvasive and therefore particularly suitable for certain clinical situations. However, from a theoretical point of view it has major disadvantages in that the long period necessary to obtain a representative collection of labeled end product allows some isotope to reenter the free amino acid pool after incorporation into rapidly turning over proteins, leading to an underestimate of flux. In addition, the fact that very different values for flux are obtained using different end products even in the same experiment suggests that the end products may not actually be derived from the same pool as the precursors for protein synthesis. This therefore casts doubt on the validity of the method, particularly when comparing different clinical and metabolic conditions.

The other method for measuring protein turnover is to give a constant infusion of labeled amino acid (at a rate i) and to measure the isotopic enrichment of the amino acid in the plasma when this reaches a steady-state value (L_{max}), at which time isotope is entering and leaving the plasma pool at the same rate. The flux is then calculated from

$$Q = i/L_{max}$$

L_{max} is expressed as the ratio of plasma enrichment to infusate enrichment. The most commonly used amino acid is leucine labeled with ^{13}C in the carboxyl group so that the oxidation rate can be obtained by measuring the rate at which ^{13}C-labeled carbon dioxide ($^{13}CO_2$) is excreted in the breath. By giving priming doses of labeled amino acid and bicarbonate, steady-state labeling of both plasma leucine and breath CO_2 can be reached very quickly and the whole procedure can be completed in 4 h, although the need to take repeated blood samples is considered a practical disadvantage. The major theoretical limitation is that the enrichment of free amino acids in the plasma may not be the same as that at the site of protein synthesis, which is likely to be intracellular and will thus vary between different tissues anyway. The enrichment of the intracellular amino acid pool will be lower than that of the plasma pool, so that this method tends to underestimate the true flux rate. When ^{13}C-leucine is being used as the tracer the enrichment of its transmination product, α-ketoisocaproic acid, in plasma is often taken as an estimate of the enrichment of intracellular free leucine.

Measurement of Protein Synthesis in Individual Tissues

The rate of protein synthesis in an individual tissue can be measured directly by administering a labeled amino acid and measuring the amount of the

labeled amino acid that has been incorporated into the proteins within the tissue after a given time. The enrichment of the precursor pool for protein synthesis during the incorporation time must also be measured in order to calculate the rate of protein synthesis. This can be problematical since the enrichment of the free amino acid pool following administration of a tracer dose of labeled amino acid follows rather complex kinetics, and erroneous conclusions have been drawn from experiments in which the enrichment of the precursor pool was not measured.

One way of overcoming this problem is again to give a constant infusion of the labeled amino acid, so that the free amino acid enrichment remains constant for a period of several hours. Since the practical approach is similar this can conveniently be combined with measuring whole-body protein turnover. The major limitation of this technique is again that the enrichment of free amino acids reaches different steady-state values in the plasma and within the cells. It is not clear which value best reflects the precursor pool for protein synthesis, since there is some evidence that extracellular amino acids may be incorporated directly into proteins. Furthermore, in some tissues which contain proteins that turn over very rapidly, there is the possibility of some reentry of label from protein after several hours of infusion.

A method which largely overcomes these problems involves giving a very large single dose of labeled amino acid. This effectively floods all the free amino acid pools, so that they reach virtually the same level of isotopic enrichment, and this value remains constant, or declines slowly and linearly for an adequate period, after which the tissue sample is taken. This period is usually 10 min in rats or mice and 90 min in humans.

The practical drawback to both these methods for measuring protein synthesis in tissues is the need to take a sample of the tissue. This can be achieved in animals which can be killed at the end of the experiment, but studies in humans are restricted to those tissues which can safely be sampled by percutaneous biopsy (such as muscle) or those which can be sampled during scheduled surgical operations.

An alternative approach which obviates the need for tissue sampling involves sampling arterial and venous blood simultaneously during a constant infusion of a labeled amino acid. The decrease in amino acid enrichment between arterial and venous blood samples reflects the rate at which free amino acid is being produced in the tissue, and at steady state, for an amino acid that is not synthesized in that tissue, this is assumed to equate with the rate of release by protein degradation. This value can then be subtracted from the net difference in amino acid concentration across the tissue to calculate the rate of amino acid incorporation into tissue protein, so long as the amino acid has no other metabolic fate within that tissue. This method has been used to measure protein synthesis and degradation in muscle using deuterated or tritiated phenylalanine as the tracer. However there are practical limitations, including the difficulty of measuring small differences in isotopic enrichment, measuring blood flow accurately, and reliably sampling venous blood from a single tissue.

Measurement of Protein Degradation

The method outlined above produces values for the rate of protein degradation as well as protein synthesis. This is important since there are no other methods for measuring protein degradation in single tissues directly. Other methods based on isotopic labeling are limited by the problem of reutilization of amino acids after release from protein. Thus many workers have concluded that the most reliable estimate of protein breakdown is obtained from the difference between protein synthesis and the net rate of change of the protein mass of the tissue. This involves sequentially killing closely matched groups of animals over a period of days, and thus cannot give any information about short-term fluctuations in protein degradation.

One approach to estimating protein degradation specifically in skeletal muscle involves 3-methylhistidine. This amino acid is formed by post translational modification of certain histidine residues in the myofibrillar proteins actin and myosin, and once it is released from degraded protein it cannot be reutilized for protein synthesis. 3-Methylhistidine is not metabolized, except for *N*-acetylation in certain species, including the rat; assessment of its production rate, from measurement of urinary excretion, should therefore indicate the rate of myofibrillar protein breakdown. Unfortunately, it is now known that many other tissues, including smooth muscle and skin, contain significant quantities of actin, and can make appreciable contributions to 3-methylhistidine excretion, and this has led to erroneous conclusions about muscle protein degradation in certain circumstances such as fasting. Some useful data have been derived from measurements of arteriovenous differences in 3-methylhistidine across muscles, but results based solely on measurement of urinary excretion of 3-methylhistidine should be interpreted with extreme caution.

Rates of Protein Synthesis

Whole-body protein turnover has been measured in a number of species. Values (per kg of body weight per

day) reported for mature animals range from 40 g in the mouse to 3 g in the cow and 4–5 g in humans (Table 1). Both synthesis and degradation are considerably increased in young growing animals, with values of 12 g kg^{-1} day^{-1} being reported for premature infants, falling to 7 g kg^{-1} day^{-1} in 1-year-old children.

Whole-body protein turnover represents the sum of the contributions from many different tissues. The contribution of each tissue depends on the fractional rate of protein synthesis in the tissue and on the proportion of the body's protein mass present in that tissue. Table 2 shows representative data for a number of tissues in the rat; isolated values are available for several other species, including humans but these do not yet form a comprehensive picture.

It should be noted that protein turnover in different tissues is controlled independently. Thus a simple measurement of whole-body protein turnover may fail to detect changes in opposite directions in two different tissues, such as muscle and liver.

Regulation of Protein Synthesis and Degradation

The rates of both protein synthesis and degradation change in response to nutritional and other environmental influences. For example, muscle protein synthesis increases rapidly after a meal, as the tissue goes temporarily into positive balance, then subsequently declines as amino acids are mobilized during the postabsorptive phase. There is a similar cycle of positive and negative balance with feeding and fasting at the level of the whole body, although the changes in protein synthesis and degradation are more equivocal. Whole-body protein degradation is suppressed by feeding, but the rate of whole-body protein synthesis does not always increase. The magnitude of the changes in protein balance is affected by the habitual diet, so that people who are habituated to a high-protein diet lose more protein during the postabsorptive period. This has important implications for the estimation of protein requirements, which will appear to be higher in people who are accustomed to a high protein intake.

It is possible that the concentrations of free amino acids within the cell have an influence on the rate of protein synthesis, but this view is hard to reconcile with the observation that the Michaelis constants (K_m) of the amino acyl tRNA ligases are generally well below the intracellular amino acid concentrations, so that the tRNAs are normally fully charged. Moreover, intracellular amino acid concentrations have been observed to increase in starvation, when the rate of protein synthesis is known to go down. Thus it is likely that other factors such as hormones play a more important role in regulating protein synthesis.

However, it is still possible that the concentration of one or another amino acid may affect protein turnover. Tryptophan is the amino acid which is generally present in the lowest concentrations in intracellular free amino acid pools, at least in mammalian tissues, and there is evidence from *in vitro* studies using rat liver cells that changes in tryptophan supply can affect the rate of protein synthesis by affecting the aggregation of ribosomes into polysomes (the form in which they are active). However, the circumstances in which such large changes in tryptophan concentration would occur *in vivo* are not clear.

In vitro studies have also suggested that the branched-chain amino acids, or leucine alone, may regulate protein synthesis in skeletal muscle, and that the deaminated metabolite of leucine, α-ketoisocaproic acid, may regulate protein degradation. However, injection of a large quantity of leucine has been

Table 1 Representative values for whole-body protein turnover (i.e., synthesis or degradation) rates in adult animals

Species	*Protein turnover (g kg^{-1} body weight day^{-1})*
Mouse	40
Rat	20
Rabbit	10
Sheep	6
Human	4–5
Cow	3

Table 2 Contributions of tissues to whole-body protein synthesis (average values for 100 g rats)

Tissue	*Protein content (% of whole body)*	*Fractional synthetic rate (% per day)*	*Absolute synthesis (% of whole body)*
Whole body	100	34	100
Muscle	45	17	23
Skin	10	64	19
Liver	4	90	12
Gut	4	100	11
Bone	3	90	8
Remainder	34	26	27

shown to have no effect on muscle protein synthesis *in vivo*.

Recent evidence has indicated that the rate of protein synthesis in muscle correlates with the intracellular concentration of glutamine, particularly in circumstances where both are reduced, including injury and infection. However, increasing muscle glutamine concentration has not been shown to alter the balance between protein synthesis and degradation.

Of the many hormones that are known to affect protein turnover, insulin has perhaps received the most attention. Studies in rats show that insulin increases protein synthesis, and this may be the major mechanism by which food intake affects protein metabolism. In contrast, insulin's main action in humans appears to be to decrease protein degradation. Glucocorticoids depress muscle protein synthesis, and thyroid hormones increase protein breakdown. Testosterone and its synthetic derivatives appear to exert their anabolic effects either by increasing protein synthesis or by decreasing protein degradation, depending on concentration.

The rate of protein synthesis in muscle is affected by physical activity. Exercise causes muscles to hypertrophy by increasing the rate of protein synthesis. During normal growth an increase in the length of long bones tends to stretch the muscles which are attached to them, and this stimulates muscle protein synthesis, resulting in the growth of the muscle.

Factors that Depress Protein Synthesis

One of the major reasons for studying protein turnover in relation to nutrition is to investigate the mechanisms by which growth rate is reduced or by which lean body mass is lost in situations of undernutrition or disease. It is perhaps not surprising that retardation or even cessation of growth is associated with a reduction in the rate of protein synthesis. There is now considerable evidence that the net loss of lean body mass in adults suffering from chronic diseases such as cancer as well as primary muscle-wasting diseases is also caused by a decrease in protein synthesis. Protein breakdown appears not to be increased, and may even be decreased in many tissues as a secondary response to decreased protein synthesis. This has major implications for therapy, which should be directed towards augmenting protein synthesis rather than suppressing protein breakdown.

Several factors appear to be involved in the depression of protein synthesis. Patients with chronic diseases often have inadequate intakes of protein and energy, and this is likely to depress protein turnover. Lack of physical activity, particularly in bedridden patients, is another important factor. Changes in conventional hormones are likely to be involved, as are cytokines such as the interleukins and tumor necrosis factor. These are peptides produced by the cells of the immune system in response to infection and injury which, apart from their immediate local action, also appear to have systemic effects, including suppressing muscle protein synthesis. This presumably evolved as an adaptive response to acute injury or infection, when it would be more important to insure the supply of substrates for such processes as gluconeogenesis and acute-phase protein synthesis than to synthesize muscle proteins. However, this now appears as a maladaptive response to chronic disease, since prolonged or severe muscle wasting is associated with considerable morbidity and mortality. At the cellular level there is some evidence that these cytokines, and indeed some conventional hormones, may act by modulating the production of eicosanoids. In this context prostaglandin $F_{2\alpha}$ appears to increase the rate of protein synthesis while prostaglandin E_2 promotes protein breakdown.

See also: **Amino Acids**: Properties and Occurrence; **Exercise**: Muscle; **Nucleic Acids**: Properties and Determination; Physiology; **Protein**: Functional Properties

Further Reading

Emery PW (1999) The protein synthetic response to feeding: a fundamental question revisited. *Nutrition* 15: 241–242.

Garrow JS and Halliday D (eds) (1984) *Substrate and Energy Metabolism in Man*. London: John Libbey.

Millward DJ (1990) The hormonal control of protein turnover. *Clinical Nutrition* 9: 115–126.

Pacy PJ, Price GM, Halliday D *et al*. (1994) Nitrogen homeostasis in man: the diurnal responses of protein synthesis and degradation and amino acid oxidation to diets with increasing protein intake. *Clinical Science* 86: 103–118.

Rennie MJ, Edwards RHT, Emery PW *et al*. (1983) Hypothesis: depressed protein synthesis is the dominant characteristic of muscle wasting and cachexia. *Clinical Physiology* 3: 387–398.

Rennie MJ, Hundal HS, Babij P *et al*. (1986) Characteristics of a glutamine transporter in skeletal muscle may have important consequences for nitrogen loss in injury, infection and chronic disease. *Lancet* ii: 1008–1012.

Waterlow JC (1984) Protein turnover with special reference to man. *Quarterly Journal of Experimental Physiology* 69: 409–438.

Waterlow JC and Stephen JML (eds) (1981) *Nitrogen Metabolism in Man*. Essex: Elsevier Science.

Waterlow JC, Garlick PJ and Millward DJ (1978) *Protein Turnover in Mammalian Tissues and in the Whole Body*. Amsterdam: Elsevier Science.

Deficiency

C Leitzmann, Justus-Liebig University, Giessen, Germany

Background

Protein deficiency indicates a lack of body protein or a relative deficiency of one or several essential amino acids. Thus, protein deficiency is synonymous with a negative nitrogen balance. The deficiency can result from a protein-deficient diet or other events, such as diseases, and it must be distinguished from the multifactorial syndrome of kwashiorkor. However, the clinical features and physiological effects are generally similar to those of kwashiorkor.

In this article the causes and groups at risk of protein deficiency are described. Its clinical features, physiological effects and treatment are explained. The adaptation to low-protein intake and protein reserves of the body are discussed.

Causes and Groups at Risk

Causes

Although the main cause of protein deficiency is a protein-deficient diet, seen in children in developing countries, it can also occur in patients suffering from various diseases. In this sense, protein deficiency (as well as other nutritional deficiencies) is a secondary consequence of the particular disease. (*See* **Malnutrition**: The Problem of Malnutrition.)

Secondary protein deficiencies can be ascribed to six causes:

1. Irregular food habits, e.g., in the case of chronic alcoholics. The diet of alcoholics can be severely deficient in protein.
2. Inability to digest and absorb the protein that is consumed; this occurs in patients with chronic gastrointestinal disorders, such as celiac disease or enteritis. (*See* **Celiac (Coeliac) Disease**.)
3. A disturbed protein metabolism, which may exist in patients with cirrhosis of the liver but also in patients with hormonal disorders or in some cases of diabetes.
4. A continuous loss of protein; this predominates in patients with diseases such as chronic renal disease, bleeding or exudative gastroenteropathy. High losses of albumin into the urine are indicators of the nephrotic syndrome. (*See* **Renal Function and Disorders**: Kidney: Structure and Function.)
5. An increased protein turnover, which is characteristic in cases of infection, fever, or gastroenteritis.
6. Enhanced catabolism of protein, with increased nitrogen losses, seen in patients with severe injuries, especially burns, or in postoperative stress. (*See* **Burns Patients – Nutritional Management**.)

Protein losses after operations depend on the kind of surgery (Table 1). If patients are unable to eat normally, insufficient food intake worsens the protein deficiency.

Groups at Risk

The groups at risk can be classified according to the causes of protein deficiency (see above) or better according to socioeconomic parameters. According to socioeconomic parameters, by far the largest group is the poor population in developing countries, primarily children in the second year of life. This group is at risk from primary protein deficiency because of a protein-deficient diet, caused by economic, ecological, and political factors. In developed societies, groups at risk are those who adhere to extreme diets or suffer from anorexia nervosa or bulimia nervosa. (*See* **Anorexia Nervosa**; **Bulimia Nervosa**.)

Clinical Features

Fatty Liver

Protein deficiency contributes to fatty infiltration of the liver and results in hepatomegaly. This can be seen in healthy subjects receiving a low-protein diet as well as in protein-deficient animals fed adequate quantities of other nutrients. Because of a decreased synthesis of β-lipoproteins needed for the transport of triglycerides, the fat accumulates in small droplets within the cells, first accumulating in the periphery of the lobules and then spreading to the center of the

Table 1 Postoperative protein losses

Operation	*Loss (grams per day)*
Stomach removal	110–112
Strumectomy	72
Cholecystectomy	71
Herniotomy	11
Amputation of the breast	9–18
Fractures of long bones	86–312
Skull injuries	67
Hip joint dislocation	17

From Welsch A (1986) *Krankenernährung*, 5th edn, p. 410. Stuttgart: Thieme Verlag.

lobules. In spite of this, the liver functions are well maintained. (*See* **Fats**: Digestion, Absorption, and Transport.)

Muscle Wasting

Because of increased protein catabolism, skeletal muscle is often wasted, and subjects limit their physical activity. In many cases, the parallel prevailing edema conceals the wasting.

Edema

Edema is a frequent sign in protein deficiency. It generally causes swelling of the tissues but usually appears in the feet and lower legs. The etiology of edema has not been completely clarified. Protein deficiency or, rather, hypoalbuminemia is only one of the factors.

Other common features of protein deficiency are changes in the hair; it becomes easily plucked, and there are changes in texture and color. Also common are skin changes, with areas of desquamation and hypo- and hyperpigmentation. Protein deficiency also leads to mental changes, lethargy, fatigue, anorexia, and some degree of anemia, and is frequently associated with infections. In children, severe protein deficiency leads to growth retardation.

Physiological Effects

Organs and systems can be changed in protein deficiency, and this may alter their physiological functions.

Cardiovascular System

Muscle wasting in the heart leads to reduced cardiac output and poor circulation. As a result, the pulse rate of many patients is low and even impalpable, and the extremities are cold and pale.

Renal Function

Structural abnormalities of the kidney have not been observed, but the glomerular filtration rate may be diminished. The concentrating power can also be impaired, which may be attributable to a low blood urea concentration or an accompanying potassium deficiency.

Pancreas

Atrophic changes and fibrosis of the pancreas have been observed both in children dying of protein malnutrition and in animals fed low-protein diets. The atrophy affects the acinar cells, responsible for exocrine secretion, but not the islets, responsible for hormone (glucagon and insulin) production. A reduced number of secretory granules have also been observed. These changes result in impaired pancreatic function with a diminished secretion of pancreatic enzymes into the intestine. Glucagon and insulin secretion are also lowered in severe cases of protein deficiency.

In another abnormality of the pancreas in protein-deficient animals, the protein content and phospholipid levels are reduced, and triglyceride and cholesteryl ester levels are increased. This altered pancreatic lipid composition in protein-deficient animals may lead to changes in the integrity of pancreatic membranes and can predispose to injury.

Gastrointestinal Tract

The mucosal epithelial cells of the small intestine make up the tissue with the highest turnover rate and are therefore very vulnerable to protein deficiency. The intestine wall becomes thin, the mucous membrane is smooth and atrophic, and the villi are frequently flattened. The enzyme-secreting cells and the pancreatic acinar cells may be affected in a similar way. As a result, the production of digestive enzymes, primarily the disaccharidases, is reduced, and the absorptive and digestive capacity of nutrients is impaired.

A feature seen in biopsis after protein deficiency is the intense cellular infiltration in the mucosa and submucosa, especially with lymphocytes and plasma cells. The colon can be less affected, but the muscle coats may also be atrophic, and the surface epithelium may show infiltration of plasma cells.

Endocrine System

The pituitary gland responds effectively to the stimulus of protein deficiency; thus, the secretion of human growth hormone is normal or, possibly, supranormal, and the thyroid-stimulating hormone is elevated. (*See* **Hormones**: Thyroid Hormones; Pituitary Hormones.)

The thyroid may be atrophic, with reduced colloid tissue, and the vascular cells are flattened and inactive. The thyroid function appears to be normal, but there are data suggesting a reduced function of the thyroid gland. The concentrations of thyroxine-binding proteins and of thyroxine are reduced.

The adrenal gland appears atrophic, but nevertheless, plasma cortisol levels are elevated and may lead to metabolic disturbances. A reduced insulin secretion, as seen after glucose tolerance tests, can also be predominant in protein deficiency. (*See* **Glucose**: Glucose Tolerance and the Glycemic (Glycaemic) Index; **Hormones**: Adrenal Hormones.)

Immunological System

In protein deficiency, the lymphoid tissue – primarily the thymus gland, but also lymph nodes and spleen – is

atrophic. First, the production of thymus-dependent lymphocytes (T cells) is reduced and is responsible for the diminished cell-mediated immunity to infectious agents, as demonstrated in specific tuberculin skin tests. Second, the phagocytic activity of neutrophils and antibody formation are reduced. The body therefore becomes more susceptible to infections which can be fatal. (*See* **Immunology of Food**.)

Treatment

The dietary treatment of protein deficiency depends on the cause of the deficiency. In patients with nephrotic syndrome, characterized by massive losses of albumin into urine, the intake of protein should be increased to 90–120 g per day for adults when hepatic synthesis of albumin can compensate in part the urinary losses. (*See* **Kwashiorkor; Marasmus**.)

In cases of acute renal failure, the protein intake should be only 20 g per day. This reduces protein metabolism and the production of urea. In patients with liver cirrhosis, the protein intake should also be reduced because a high-protein diet can precipitate hepatic encephalopathy.

Special Feeding Methods

Protein deficiency can be treated by tube feeding, intravenous feeding or supplementary feeding. Tube feeding is indicated in patients with severe malnutrition who are unable to eat. Parenteral nutrition is essential when the small intestine is unable to digest and absorb nutrients, or after severe injuries such as burns.

In mild protein deficiencies, an increased oral intake of protein is sufficient to meet dietary needs. Several special preparations containing high levels of protein are available on the market.

Complications

Since protein deficiency is frequently accompanied by infections, dehydration, and deprivation of vitamins and electrolytes, these conditions need special attention. In most cases of severe protein deficiency, therefore, treatment has to start with fluid replacement and, if indicated, with antibiotic therapy.

Adaptation to Low-protein Intakes

Waterlow (1986) described two possibilities of nutritional adaptation, which he called the first and second line of defense. The first line of defense is the ability to achieve nitrogen balance at various levels of protein intake. The second line of defense is necessary after the capacity to economize nitrogen is exhausted, resulting in a reduced lean body mass or reduced growth rate in children. In this section, only the first line of defense is discussed.

Nitrogen Balance

Nitrogen equilibrium is a state in which, for a given intake of nitrogen, an equivalent amount of nitrogen is lost from the body via urine, feces, skin, sweat, etc. When protein intake is low, dietary protein is used more efficiently; this is exemplified by the excretion of nitrogen, particularly in the form of urea, which is reduced and so contributes to the restoration of nitrogen balance. In this situation, the pathways of amino acid synthesis predominate.

The liver plays an important role in the adaptive process, since it is the only organ which can transform the nitrogen from amino acids into urea. The alterations in the nitrogen pathway are mainly brought about by changes in the activity of liver enzymes. Experiments with animals have shown that in protein deficiency, urea cycle enzymes, such as argininosuccinase, are low, and the amino-acid-activating enzymes are high. Table 2 shows this finding in children with severe protein malnutrition, soon after admission to hospital and after recovery.

The metabolic activity of the gastrointestinal tract is important in the adaptive process. Normally, one-third of the urea produced is passed into the bowel and can be hydrolyzed by the gut microflora. Thus, urea nitrogen becomes available for metabolic interaction, e.g., for the synthesis of nonessential and essential amino acids. However, the necessary carbon skeletons can be the limiting factor in adaptation. In protein deficiency, a greater proportion of urea is retained by the body. (*See* **Microflora of the Intestine: Role and Effects**.)

Recent investigations have shown that the colon is permeable to urea and amino acids, thus supporting the view that hydrolysis of urea plays an important nutritional role.

Table 2 Enzymatic activity of the livers of children with malnutrition

	Amino-acid-activating enzymes	*Argininosuccinase*
Soon after admission	1.44	1.06
1–2 months later	0.91	1.46

The figures for the amino-acid-activating enzymes are the means of 18 measurements and are expressed in micromoles of phosphorus exchanged per milligram of protein; for argininosuccinase, the figures are the mean of 11 measurements and expressed in micromoles of urea per milligram of protein. The changes on recovery are statistically significant. From Passmore R and Eastwood MA (1986) *Human Nutrition and Dietetics*, 8th edn., p. 52. Edinburgh: Churchill Livingstone.

Table 3 Relative losses of protein in different organs and tissues from rats over 7 days

Organ or tissue	*Loss (percentage of primary content)*
Liver	40
Prostate gland	29
Seminal vesicle	29
Gastrointestinal tract	28
Kidney	20
Blood plasma	20
Heart	18
Muscle, skin, skeleton	8
Brain	5
Eyes	0
Testicle	0
Adrenal gland	0

From Kraut H (ed.) (1981) *Der Nahrungsbedarf des Menschen: 1 – Stoffwechsel, Ernährung und Nahrungsbedarf, Energiebedarf, Proteinbedarf*, p. 148. Darmstadt: Steinkopff Verlag.

Plasma Albumin

Plasma albumin reacts very sensitively to a reduced protein intake. Although the albumin turnover is reduced by about 36% in studies with adults, the plasma albumin concentration is reduced by only about 7%. This maintenance of intravascular circulating albumin mass is attributable to a reduced breakdown and a shift of albumin from the extravascular to the intravascular compartment.

Factors Affecting Adaptation

Among the factors that can affect the adaptation to low protein intake are infections, diarrheal diseases, and injuries. In infections, protein from muscle and skin is needed for the immune response, and this leads to a negative nitrogen balance. In diarrheal diseases, the malabsorption has a negative influence on the adaptive process; in patients with injury, the severe losses of nitrogen have the same effect.

Energy balance is also essential for nitrogen balance because of its nitrogen-sparing effect. Thus, if protein deficiency is accompanied by energy deficiency, the adaptation to a low protein intake cannot be achieved completely.

Protein Reserves

The body of a human adult (65 kg) contains 12 kg of protein, about 50% of which is found in muscles. The amount of body protein depends on, among other things, the dietary protein and carbohydrate intake; if carbohydrate is lacking, the amino acids are utilized for gluconeogenesis. (*See* **Glucose**: Function and Metabolism.)

Protein reserves are not comparable to special fat depots, and not all body proteins can serve as protein reserve. Reserves are primarily organs which contain labile body protein such as liver, plasma (with proteins such as albumin and enzymes) and the gastrointestinal tract, which has an assessed protein content of about 700–800 g. Although the protein turnover rate in muscle is very slow, this tissue is a very important protein reserve owing to its large mass.

In protein deficiency, the labile body proteins are metabolized first; when deficiency is long-term, all organs are affected to various extents (Table 3). The well-fed human adult can lose about 3 kg of protein without any disturbances to his or her health.

See also: **Anorexia Nervosa**; **Bulimia Nervosa**; **Burns Patients – Nutritional Management**; **Celiac (Coeliac) Disease**; **Fats**: Digestion, Absorption, and Transport; **Glucose**: Function and Metabolism; Glucose Tolerance and the Glycemic (Glycaemic) Index; **Hormones**: Adrenal Hormones; Thyroid Hormones; Pituitary Hormones; **Immunology of Food**; **Kwashiorkor**; **Malnutrition**: The Problem of Malnutrition; **Marasmus**

Further Reading

Elmadfa I and Leitzmann C (1998) *Ernährung des Menschen*, 3rd edn. Stuttgart: Verlag Ulmer.

Jackson AA (1990) The aetiology of kwashiorkor. In: Harrison GA and Waterlow JC (eds) *Diet and Disease in Traditional and Developing Societies*, pp. 76–113. Cambridge: Cambridge University Press.

Khaled MA, Kabir I and Mahalanabis D (1995) Effect of protein energy supplementation on oxidative stress in malnourished children. *Nutrition Research* 15: 1099–1104.

Kraut H (ed.) (1981) *Der Nahrungsbedarf des Menschen: 1 – Stoffwechsel, Ernährung und Nahrungsbedarf, Energiebedarf, Proteinbedarf*, pp. 140–153. Darmstadt: Steinkopff Verlag.

Latham MC (1990) Protein–energy malnutrition. In: Brown ML (ed.) *Present Knowledge in Nutrition*, 6th edn, pp. 39–46. Washington, DC: International Life Sciences Institute – Nutrition Foundation.

Latham MC (1997) *Human Nutrition in the Developing World, FAO Food and Nutrition Series No. 29*, pp. 127–146. Rome: FAO.

Passmore R and Eastwood MA (1986) *Human Nutrition and Dietetics*, 8th edn pp. 40–53, 279–291, 403–407, 452, 482–501. Edinburgh: Churchill Livingstone.

Siegenthaler W, Kaufmann W, Hornbostel H and Waller HD (eds) (1984) *Lehrbuch der inneren Medizin*, pp. 14.17–14.19. Stuttgart: Thieme Verlag.

Trowell HC, Davies JNP and Dean RFA (1982) *A Nutrition Foundation's Reprint of Kwashiorkor*, pp. 122–160. New York: Academic Press.

Waterlow JC (1986) Metabolic adaptation to low intakes of energy and protein. *Annual Reviews of Nutrition* 6: 495–526.

Heat Treatment for Food Proteins

L Mauer, Purdue University, West Lafayette, IN, USA

Background

Amino acid composition and sequence determine the native structure, functionality, and nutritional quality of a protein in a set environment. During food processing, heat is often added to the protein's environment, and this addition of energy can change any or all of the structural, functional, and nutritional characteristics of the native protein. Foods are complex systems, and it is important to recognize that pH, water activity, food composition, and interactions of these with temperature also affect protein properties to varying extents. Common heat treatments used in processing food proteins, effects of heat on protein structure and function, effects of heat on protein interactions with carbohydrates and lipids, and nutritional effects of heating proteins are discussed.

Common Heat Treatments on Food Proteins

At home or on an industrial scale, the purposes of common heat treatments of foods containing proteins are similar: to change texture or function, improve safety and quality, and control enzymes by altering physical, chemical, and biological protein properties. Foods are baked to change texture and improve safety, vegetables are blanched to inactivate enzymes, canning temperatures used for low acid foods are designed to prevent toxin production by pathogenic microorganisms, pasteurization and sterilization are designed to kill pathogens and inactivate enzymes, and extrusion modifies the texture of protein-containing foods. Mild heat treatments, such as incubation, generally do not cause the same extent of change as higher heat treatments and may be used to promote enzyme activity instead of destroying it. It is worthwhile to note that proteins usually are most stable to heat at their isoelectric points.

Dry-heat Food Preparation

Dry-heat food preparation methods at temperatures ranging from 160 to 230 °C include baking, roasting, grilling, and frying. In addition to microbial destruction, dry-heat methods alter the texture of protein foods by heat gelation of proteins, denature enzymes and pigments, and with extensive heat may form thermally induced mutagens. When meats are heated, myofibrillar proteins denature, then form a gel matrix, enzymes such as myosin and actomyosin are inactivated, and oxidation of denatured myoglobin pigments turns cooked meats brown. On baking, the elastic wheat gluten network in bread dough expands to contain leavening gases until the temperature is high enough to gelatinize starch, around 65 °C. Gluten protein denaturation and gelation occur at higher temperatures than starch gelatinization, beginning around 74 °C, and continuing for the remaining baking time. The Maillard reaction between proteins and carbohydrates produces the browning of bread crusts during baking.

Moist-heat Food Preparation

Moist-heat food preparation methods at temperatures ranging from 65 to 100 °C include blanching, boiling, steaming, scalding, and poaching. Pressure canning foods can reach temperatures in excess of 116 °C. Blanching is a process to inactivate enzymes by dipping foods, usually vegetables, into boiling water for a short time period. This enzyme inactivation prevents quality loss by color, texture, or flavor changes during frozen storage. Blanching also decreases initial microbial load on foods as well as wilts products such as spinach for tighter packing. Boiling occurs at or near 100 °C, depending on elevation, and functions to inactivate enzymes, denature proteins, change texture, and potentially destroy toxins.

Higher temperatures achieved under pressure, such as for pressure canning, inactivate enzymes, destroy pathogenic and many spoilage microorganisms, and prevent toxin formation during storage. Poaching eggs leads to denaturation and coagulation of egg white proteins. Steaming fish causes texture changes as a result of denaturation and gelation of proteins. Scalding milk may initiate unfolding of whey proteins and improve select functional properties desirable in preparing bakery foods.

Microwaving

Microwaving combines properties of dry- and moist-heat food preparations with the advantage of reducing food cooking times. Heat produced by friction of rotating food molecules, most notably water, in response to magnetron-generated microwaves denatures and coagulates food proteins as do other heating methods. However, foods cooked in a microwave do not brown as they would in dry-heating methods. The crust of microwaved bread dough does not brown, because the air in the microwave does not

reach high enough temperatures, and the steam generated does not allow the surface to dry sufficiently for nonenzymatic browning to occur.

Pasteurization

The pasteurization process is designed to destroy any pathogenic microorganisms that might be present in the food product. Temperatures used for pasteurization also reduce the total microbial load, thereby increasing shelf-life, and may inactivate select enzymes that lead to quality loss during storage. Pasteurization of milk, designed to destroy the pathogen *Coxiella burnetti*, is accomplished by a time/temperature integrated process. Lower temperatures take longer times (63 °C for 30 min, 71.5 °C for 15 s), whereas higher temperatures require much shorter process times to accomplish the same result (88.5 °C for 1 s). At temperatures above 78 °C, esterase, a lipase that may cause hydrolytic rancidity of milk products, is inactivated. Pasteurization of milk also may denature whey proteins, and higher temperatures (ultrahigh temperature processing) or longer process times (such as vat pasteurization) may cause cooked and heated flavors primarily from the volatile sulfur compounds released by β-elimination of disulfide bonds in denatured whey proteins. The high temperatures used for sterilization of aseptic products lead to microbial safety but do not inactivate all enzymes. During extended storage, these enzymes may cause age gelation of aseptically processed milk products. Pasteurization of liquid eggs is a low-temperature, long-time process, either 60 °C for 3.5 min or 64 °C for 2.5 min, designed to destroy pathogenic microorganisms, specifically *Salmonella*, without coagulating the egg proteins.

Extrusion

Extrusion is a high-temperature/high-pressure/high-shear process used to convert soy proteins to textured vegetable protein meat analogs. The texture changes during extrusion cooking result from orientation and denaturation of proteins followed by cross-linking into a network of fiber-like proteins. The extruded vegetable protein network mimics the fibrous texture of meat products. Pressure and shear individually can denature proteins, and lower temperatures may be used in combination with these forces to achieve the same level of denaturation as higher temperatures alone.

Incubation

Temperature is the most important factor affecting the rate of enzyme-catalyzed reactions and also influences the stability of the enzymes. As temperature is increased, reaction rates increase until a temperature at which the enzyme loses activity is passed, often around 50 °C. In the range of 10–40 °C, every 10 °C rise in temperature is accompanied by a 1.8 increase in reaction rate for the enzyme chymosin. The enzyme calf rennet (chymosin) is added to milk in the cheese-making process to cleave κ-casein from the casein micelle and induce micelle aggregation. The maximum rate of aggregation is achieved at 40 °C, and no aggregation occurs below 18 °C or above 60 °C. At temperatures above 50 °C, denaturation of the enzyme causes it to lose activity. Therefore, incubation at 40 °C will maximize chymosin activity, and increasing temperatures above 50–60 °C can be used to stop the enzyme. Temperatures used for cheese ripening also are controlled to maximize desirable enzyme activity for flavor development.

Heat-induced Changes in Protein Structure

Common changes in protein structure as a result of thermal processing include denaturation, aggregation, and thermal degradation. Elevated temperatures also increase rates of deleterious chemical reactions in proteins that lead to oxidation, isomerization, Maillard browning, deamidation, desulphuration, and other β-elimination reactions; however, many of these reactions may occur at temperatures as low as 0 °C as well.

Denaturation

Heat denaturation of proteins involves configurational changes in the thermodynamically stable native structure of the protein via unfolding or alteration of the quaternary, tertiary, or secondary structure as a response to heat exposure. Denaturation may disrupt hydrogen and disulfide bonds, hydrophobic interactions, and salt bridges, but peptide bonds remain intact. The primary structure, or amino acid sequence, of the protein remains unchanged, as does the molecular weight. A loss of ordered structure generally occurs in the entropy-driven transition from a native to a denatured protein. Temperatures at which denaturation occurs vary greatly with protein source and type. Some proteins unfold a few degrees above temperatures at which they function, whereas others, such as wheat gluten and milk β-casein, require much higher temperatures for denaturation. Globular dairy whey proteins denature at much lower temperatures than casein proteins that have more random coil native structures.

Functional properties affected by heat denaturation include solubility, emulsifying capacity, gelation

capacity, foaming properties, and enzyme or biological activity. After mild (or insufficient) heat treatments, protein denaturation may be reversible. For enzymes, this reversible unfolding always occurs prior to irreversible inactivation. If a thermal process does not irreversibly inactivate target enzymes, the enzymes may be able to refold and potentially cause quality issues in food products during storage. Residual enzyme activity in aseptically processed foods can lead to problems such as thinning of puddings due to amylase activity, age gelation in ultra-high-temperature processed milks, and separations in orange juice caused by pectin methylesterase activity.

Aggregation

Denaturation, or at least a partial denaturation, of a protein is usually required prior to its ability to aggregate and form a gel or a precipitate. Unfolding of a protein exposes reactive groups that may then form intermolecular cross-links via covalent, hydrogen, ionic, or other bonding. Gels are formed when the cross-linked protein network is extensive enough to form a continuous phase and trap water. Precipitates are formed when the aggregated proteins become insoluble and settle out of a solution.

Gelation The cross-linking of proteins and entrapment of water during gelation may be either reversible or irreversible. The polymeric networks formed by hydrogen bonding of gelatin molecules are thermoreversible. On cooling, the gelatin bonds to form a gel, and on reheating, the hydrogen bonds break, and gelatin returns to the dispersed phase in the solution. Conversely, denatured globular proteins form thermoirreversible gels. Whey, soy, and egg white proteins first denature and then interact via disulfide, hydrophobic, and ionic bonds to form gels when exposed to heat. These gels are heat-set and may stiffen, instead of liquefy, when exposed to additional heat. The term 'coagulation' is often applied to irreversible heat-setting of proteins.

Precipitation Precipitation of food proteins often is controlled by pH, enzymes, or salt concentration adjustments for food ingredient isolation and applications; however, heat also may destabilize proteins, causing them to precipitate. Solubility and hydrophobicity, which are affected by temperature, also influence protein precipitation. As temperatures increase, the likelihood of protein precipitation increases as hydrophobicity increases and solubility decreases. Denaturation of whey proteins in milk causes them to precipitate, adhere to the cooking vessel, and possibly scorch with continued heating.

Thermal Degradation

Thermal degradation of proteins occurs at temperatures higher than those for denaturation, and as with denaturation reactions, temperatures for thermal degradation vary greatly with protein type. In addition to the quaternary, tertiary, and secondary structure disruption in denaturation, thermal degradation also disrupts the primary structure and peptide bonds of proteins. Effects of thermal degradation on functionality of proteins are more severe than the effects of denaturation. Types of protein thermal degradation include hydrolysis, racemization, and pyrolysis.

Hydrolysis Hydrolysis of proteins for use as functional or flavor ingredients often is accomplished using acid and enzyme techniques; however, high temperatures also may be used to hydrolyze proteins into peptides, especially at high or low pH. Hydrolysis of caseins may occur at 140 °C, and heat-induced hydrolysis of meat connective tissue proteins may increase their solubility. Partially hydrolyzed proteins may be more digestible than native proteins due to the unfolding of the protein structure.

Racemization Racemization, or isomerization, of amino acid residues from L-isomers to D-isomers occurs when heating proteins either above 200 °C or at alkaline pH. This may reduce protein digestibility, because D-amino acids are less absorbed and hydrolyzed than L-amino acids in the digestive tract. Isomerase or racemase enzymes also contribute to these conversions.

Pyrolysis Pyrolysis is a high temperature degradation of organic materials in nonoxidative conditions. Amino acid pyrolysis products are formed at temperatures above 250–300 °C. Free radicals formed during pyrolysis may attach to other amino acids, thereby forming new heterocyclic compounds. Some of these compounds are thermally induced mutagens.

Functional Changes in Heat-treated Proteins

Research indicates that protein structure is optimized for function as opposed to stability. As a result, when addition of heat destabilizes protein structure and stability, functionality also is affected. Effects of heat on protein functionality can be positive or negative, and the degree of heat treatment on a specific protein often differentiates between an increase or decrease in the desired functionality. Addition of some heat may slightly unfold a protein, thereby

increasing emulsifying, gelation, and foaming properties. In designing functional and nutritional foods, the extent of protein denaturation is controlled to attain desired functional characteristics, such as whey proteins designed for use as fat replacers. Higher heat treatments, especially those leading to thermal degradation, may in turn decrease these functional properties as the protein unfolds more or loses primary structure. Water–protein interactions dominate protein functionality in food structure; therefore, effects of heat on hydrophobicity, solubility, emulsifying and gelation capacities, foaming properties, and enzyme activity are important.

Hydrophobicity

The surface hydrophobicity of proteins may either increase or decrease as a result of heat treatments. A thermally induced unfolding of protein molecules exposes hydrophobic sites, thereby increasing hydrophobicity. Conversely, protein aggregation in response to heat results in decreased exposure of hydrophobic sites, thereby decreasing the surface hydrophobicity of aggregated proteins.

Solubility

The solubility of proteins depends on the nature of the protein surfaces in contact with the environment (usually water). To generalize, a protein with a hydrophilic surface will be more soluble in water than a protein with a more hydrophobic surface. As temperatures rise from 0 to 40 °C, most proteins exhibit increasing solubility; however, hydrophobic proteins such as β-casein show the opposite solubility trend in this temperature range and may be most soluble around 4 °C. As temperatures rise above 40 °C and proteins unfold, more hydrophobic sites are exposed, and the solubility of the proteins will decrease. These temperature-dependent changes in solubility will influence emulsifying, gelation, and foaming properties, since solubility influences the amount of protein available for reactions.

Emulsifying Capacity

The emulsifying capacity of proteins results from amphipathic structure and is measured as the volume of oil that can be emulsified per gram of protein in an oil-in-water system. The protein must be somewhat soluble in water in order to act as an emulsifier, and the increase in surface hydrophobicity that occurs with partial denaturation of most proteins will increase their emulsifying capacities. When higher temperatures have caused extensive denaturation and decreased protein solubility, the emulsifying capacity also will decrease.

Gelation Capacity

The gelation capacity of proteins is measured as the amount of water that can be bound or trapped per gram of protein. The effects of heat on gelation capacity are similar to those on emulsifying capacity. Partial denaturation may increase gelation capacity, whereas extensive denaturation will decrease it.

Foaming Properties

The foaming capacity of a protein is measured as the amount of interfacial area that can be created by whipping the protein. Foam stability is measured as the time required to lose either 50% of the liquid or 50% of the volume from the foam. Generally, heating a globular protein to achieve partial denaturation will increase foaming properties. As the structure unfolds and exposes hydrophobic sites, it may be able to adsorb more quickly to air–water interfaces and lower interfacial tension, thereby trapping more air and increasing the foaming capacity. Extensive heat denaturation of proteins will decrease their ability to form foams.

Enzyme Activity

Temperature controls the rate of enzyme-catalyzed reactions and influences enzyme stability. As discussed in the incubation section, moderate temperatures can optimize the rate of enzyme-catalyzed reactions, and higher temperatures may denature (unfold) the enzymes, thereby causing them to lose activity. The temperatures at which enzymes denature vary with the type of enzyme. Chymosin, added to milk for cheese-making, loses activity above 50 °C, whereas pectinmethylesterase present in orange juice is stable to much higher heat treatments.

Heat-induced Protein–Carbohydrate Interactions

Perhaps the most notable protein interaction with other food ingredients is the browning produced in the Maillard reaction between proteins and carbohydrates. At high temperatures, low water activity, and/or extended storage times, proteins may react with reducing sugars to form brown pigments in the Maillard reaction (nonenzymatic browning). Reactive groups in proteins for the Maillard reaction are primary amines, usually the ε-amino groups of lysine residues. Examples of reducing sugars are glucose, fructose, lactose, and maltose. When the free ε-amino group of lysine reacts with a reducing sugar, the lysine is no longer nutritionally available. Lysine is often the limiting amino acid in protein quality of grain products, and a decrease in the available lysine due to the Maillard reaction decreases the overall

protein quality of the food. Nonenzymatic browning is desirable in bakery products such as breads, cooked meats, and caramels for which browning contributes to color and flavor development. However, too much browning produces burnt or off-flavors in these products. Nonenzymatic browning also is undesirable in dried milk powders, infant formula, dehydrated potatoes, dried fruits, and white wine.

Heat Effects on Protein–Lipid Interactions

Heat treatments may affect protein–lipid interactions in terms of free-radical formation, changes in emulsifying capacity, and alteration of conjugated lipoprotein structure. Lipid–protein free radicals may be formed when free radicals produced by oxidation of unsaturated lipids react with proteins. High temperatures greatly increase the rate of oxidation of sulfur-containing amino acids via reactions with oxidized lipids. Cysteine and histidine free-radicals may then cross-link and induce aggregation of proteins. As discussed in the emulsifying capacity section, partial denaturation of globular proteins may expose hydrophobic sites and increase emulsifying capacity, thereby increasing the ability of proteins to interact with lipids; however, higher heat treatments will decrease this ability. Heat also will denature proteins in conjugated lipoprotein structures and affect the functionality of these, especially in membrane systems.

Nutritional Changes in Heat-treated Proteins

Heat-induced changes in protein structure can exhibit both advantageous and negative effects from a nutritional standpoint. Heat denaturation of globular proteins, such as dairy whey proteins, often leads to increased digestibility and sometimes increased nutritional value as the structure is unfolded and therefore more susceptible to proteolytic attack by digestive enzymes. Proteinaceous antinutritional factors, commonly found in legume and oilseed proteins as well as milk and egg proteins, can be inactivated with sufficient heat, usually moderate heat treatments, thereby increasing the biological availability and digestibility of select proteins. Heat inactivation of trypsin and chymotrypsin inhibitors present in legumes and oilseeds also may protect the pancreas. Once heat denatured, the lectins (phytohemagglutinins) present in legumes and oilseeds no longer bind to intestinal cells, causing malabsorption of nutrients. Heat inactivation of ovomucoid and ovoinhibitor in eggs and plasmin and plasminogen activator inhibitors in milk prevent their protease inhibitory activities. Once denatured by heat, egg avidin no longer inhibits biotin absorption. Toxic proteins, including *Clostridium botulinum* neurotoxin and *Staphylococcus aureus* enterotoxin, also are inactivated by heat, although toxin destruction usually requires higher heat treatments than inactivation of antinutritional factors.

Although heat improves safety and some nutritional aspects of food, heating foods also may detract from nutritional quality, especially during high heat or long time processes such as charcoal grilling. Protein cross-linking can decrease proteolysis, thereby decreasing digestibility. The D-amino acid residues formed in racemization are less digestible than the original L-amino acids. As a result of Maillard browning, there is a loss of available lysine and consequent decrease in protein quality. Melanoidins and heterocyclic amines produced in this reaction and during pyrolysis have been shown to be mutagenic or carcinogenic and classified as thermally induced mutagens or carcinogens. Consumption of small amounts of these compounds over extended periods of time may lead to serious health problems.

See also: **Amino Acids**: Properties and Occurrence; Determination; Metabolism; **Browning**: Nonenzymatic; Toxicology of Nonenzymatic Browning; **Heat Treatment**: Chemical and Microbiological Changes; **Pasteurization**: Principles; **Plant Antinutritional Factors**: Detoxification; **Protein**: Functional Properties; Interactions and Reactions Involved in Food Processing

Further Reading

Brown A (ed.) (2000) *Understanding Food Principles and Preparation*. Belmont, CA: Wadsworth.

Creighton TE (1993) *Proteins. Structure and Molecular Properties*. New York: W.H. Freeman.

Fágáin CÓ (ed.) (1997) *Stabilizing Protein Function*. New York: Springer.

Fennema OR (ed.) (1996) *Food Chemistry*, 3rd edn. New York: Marcel Dekker.

Friedman M (ed.) (1991) *Nutritional and Toxicological Consequences of Food Processing*. New York: Plenum Press.

Hoyem T and Kvale O (eds) (1977) *Physical, Chemical and Biological Changes in Food Caused by Thermal Processing*. London: Applied Science.

Kinsella JE and Soucie WG (1989) *Food Proteins*. Champaign, IL: American Oil Chemists Society.

Phillips RD and Finley JW (eds) (1989) *Protein Quality and the Effects of Processing*. New York: Marcel Dekker.

Priestly RJ (ed.) (1979) *Effects of Heating on Foodstuffs*. London: Applied Science.

Stadelman WJ and Cotterill OJ (eds) (1995) *Egg Science and Technology*, 4th edn. New York: Food Products Press.

Zayas JF (1997) *Functionality of Proteins in Foods*. New York: Springer Verlag.

Sources of Food-grade Protein

G S Gilani and N Lee, Bureau of Nutritional Sciences, Health Canada, Ottawa, Ontario, Canada

Background

Comparison of data from different countries has indicated that countries with low gross national products also tend to have a low food energy availability. In addition, there are major differences in the pattern of foods in the diet in such countries compared with wealthier countries. In particular, the availability of animal-protein foods is significantly lower in poor countries. Compared to the developed world, a greater proportion of the protein consumed by people in developing countries is derived from low-quality protein sources such as cereals. Moreover, the patterns of indispensable amino acid (IAA) composition in the proteins are such that there is a significant difference between the diets of rich and poor people with respect to lysine. Standard tables of amino acid composition show that the values for cereals range from 26 to 38 mg of lysine per gram of protein, while the values for animal foods range from 70 to 100 mg of lysine per gram of protein. A recent analysis of the world's balance of dietary IAA demonstrated that average lysine intakes were low in poor, developing countries. This would indicate a considerable risk of inadequate levels of intake, especially in vulnerable subgroups such as children and pregnant or lactating women. Strategies to alleviate the possible global problems of a low lysine supply could include increasing dietary diversity as well as either supplementing cereal-based foods with food-grade proteins high in lysine or fortifying cereals with lysine.

Food-grade Proteins

In this chapter, sources of food-grade protein are defined as products made by processing certain food ingredients in a manner that significantly increases their protein content. They are often referred to as protein flours, protein concentrates, and protein isolates. Some examples are milk-protein isolate, whey-protein concentrate, casein, lactalbumin, egg-white solids, gelatin, soybean-protein flours, concentrates, and isolates, pea-protein concentrate, canola-protein concentrate, mycoprotein, cottonseed protein, and wheat gluten.

These protein sources could be used to improve the protein nutriture of populations by increasing the protein content of the diet, increasing the quality of the proteins in the diet, or a combination of both. It should be noted that increasing the protein quantity and/or quality of a diet is ineffective if energy requirements are not met.

Animal-protein Products

Most animal sources of food-grade protein (except gelatin) are excellent sources of all the IAA (Table 1) and have a high protein digestibility (*See* **Protein: Quality**). High-protein-quality animal products include egg proteins, casein, milk protein isolate, lactalbumin, whey-protein concentrate, beef, tuna, and chicken meat products. These products are especially high in key IAA such as lysine, methionine+cystine, threonine, and tryptophan, which are limiting in poor-quality vegetable-protein products. Among the animal-protein products, egg white contains the highest amount of methionine+cystine, and lactalbumin and whey-protein concentrate contain the highest amounts of lysine and tryptophan, while egg white, tuna, beef, and chicken meat products contain the highest amounts of threonine. Therefore, these animal-protein products could serve as supplements to low-quality proteins limiting in one or more of these IAA.

The protein efficiency ratio (PER) values of most high-quality animal-protein products are higher than that of the standard for the method, which is casein (Table 2). In contrast, proteins with protein digestibility-corrected amino acid scores (PDCAAS) exceeding 100% are truncated to 100%, implying that values exceeding 100% do not contribute any additional benefit in humans. This is not true, especially when these proteins are used as supplements to poor-quality proteins. Therefore, there is an urgent need to revise the PDCAAS method to permit the use of untruncated PDCAAS values for high-quality animal-protein products so that their supplementary value, when combined with lower-quality protein sources, can be identified by their PDCAAS values.

Gelatin is a unique animal protein product which is devoid of tryptophan and is also limiting in all other IAA. This complete lack of nutritional value has not stopped some individuals from using it as the basis for weight-loss drinks. After several illnesses and deaths were associated with the use of these products in very-low-calorie diets, Canada introduced a requirement that any product containing gelatin or collagen or hydrolyzed gelatin or collagen must carry a cautionary statement advising people not to use it as a sole source of nutrition.

Table 1 Protein (on a percentage dry weight basis) and IAA compositions and protein quality expressed as PDCAAS of animal-protein products

Nutritional parameter	*Casein*	*Milk-protein isolate*	*Lactalbumin*	*Whey-protein concentrate*	*Egg white*	*Tuna*	*Minced beef*	*Chicken breast*	*Gelatin*
Total protein % (N × 6.25)	90–95	93	80	79	87.7	89.1	95.2	90.2	96.8
IAA : expressed as a percentage of the respective IAA in the FAO/WHO (1991) scoring pattern[a]									
Histidine	158	153	105	110	118	289	168	179	42
Isoleucine	193	157	200	207	188	223	149	161	50
Leucine	154	141	215	191	133	123	117	115	42
Lysine	145	131	209	172	120	153	137	141	66
Methionine + cystine	140	156	224	196	266	143	131	148	38
Phenylalanine + tyrosine	182	157	108	124	168	108	111	113	32
Threonine	135	123	165	168	138	145	124	123	51
Tryptophan	118	127	209	191	133	110	112	113	0
Valine	194	157	163	154	194	133	130	137	57
Protein digestibility (%)	99	98	99	98	97	95	94	97	98
PDCAAS (%)	100	100	100	100	100	100	100	100	0

[a]FAO/WHO (1991) Recommended Amino Acid Scoring Pattern (containing, in milligrams per gram of protein included: histidine, 19; isoleucine, 28; leucine, 66; lysine, 58; methionine + cystine, 25; phenylalanine + tyrosine, 63; threonine, 34; tryptophan, 11; and valine, 35.).

Table 2 PER and PDCAAS values of some high-quality animal-protein products

Product	*PER (casein = 2.5)*	*PDCAAS (%)*
Casein + methionine	3.1	100
Whey-protein concentrate	3.0	100
Egg-white solids	3.0	100
Lactalbumin	2.8	100
Skim milk powder	2.8	100
Milk-protein isolate	2.8	100
Chicken breast	2.8	100
Turkey breast	2.8	100
Minced beef	2.7	100
Beef salami	2.6	100
Tuna	2.6	100

Vegetable-protein Products

The Codex Committee on Vegetable Proteins (CCVP) has developed an international general standard for vegetable protein products (VPP). This standard applies to VPP intended for use in foods and prepared by various separation and extraction processes from proteins of vegetable sources except single-cell proteins. The VPP are intended for use in foods requiring further preparation and for use by the food-processing industry. The VPP standard applies to food products produced by the reduction or removal from vegetable materials of the major constituents (water, oil, starch, and other carbohydrates) in a manner to achieve a protein (N × 6.25) content of 40% or more. CCVP has also developed international standards for soy-protein products (flour, concentrate, and isolate) and for wheat gluten. These VPP are extensively used in international trade.

Soybean-protein Products

Soybeans are dehulled and solvent-extracted to remove the oil. The residual solvent is then removed and the flakes dried, resulting in defatted soy flakes. Soybean-protein products are mostly made from the defatted flakes. They are not consumed directly but instead find wide application as versatile ingredients in various food systems, including bakery, breakfast cereals, beverages, infant formulas, enteral nutritionals, dairy, meat, and dairy and meat analogs. They not only boost protein content in these food systems but also provide many functional properties, including gelling, emulsifying, water-holding, and fat-absorbing properties. There are four major types of soybean protein products, including flour, concentrates, isolates and textured soybean protein. The Codex Alimentarius Commission has established international compositional standards for soybean protein flours, concentrates, and isolates.

Soybean flour comes in many types including full-fat, low-fat and defatted, as well as toasted and textured. Defatted soybean flour, the most common type, is produced by grinding the defatted flakes. It has a protein content of about 50% (N × 6.25). The soybean-protein flour is mainly used as an ingredient in bakery products. However, full-fat soybean flour has been gaining popularity in recent years as newer technology has enabled it to be used as a starting material for the production of soy milk and tofu and as an ingredient in many other food systems.

Soybean protein concentrate is traditionally made by aqueous alcohol extraction of defatted soy flakes. The resulting product has a protein content of about 70% (N × 6.25), with the remaining portion being mainly insoluble carbohydrates. The protein

concentrate can be further processed by thermal processing and homogenization for a better functionality. Alternatively, soybean protein concentrate can be made by an acid-leach method to retain isoflavones and other beneficial phytochemicals, and to prevent protein denaturation. Soybean protein concentrate is widely used in the meat industry to bind water and emulsify fat, and as a key ingredient for many meat alternatives. It is also used for protein fortification of various types of food.

Soybean-protein isolate is produced by alkaline extraction of the defatted flour, followed by precipitation at acid pH. As a result, both soluble and insoluble carbohydrates are removed. The resulting product has a protein content of about 90% ($N \times 6.25$), and is light in color and bland in flavor. Soybean protein isolate is the most refined soybean protein product and possesses many functional properties, including gelation and emulsification. Consequently, it may be used in a wide variety of food systems, including processed meat, meat analogs, soup and sauce bases, nutritional beverages, sole-source foods such as infant formulas and enteral nutritionals, and dairy analogs.

Textured soybean proteins are produced by thermoplastic extrusion of soybean protein flour or concentrate under moist heat and high pressure to impart a fibrous texture. The textured products, which come in many sizes, shapes, colors, and flavors, are mainly used in processed meats and snack foods, and as meat alternatives.

Soybean protein products contain a number of antinutritional and bioactive components such as trypsin inhibitors (causing potential adverse effects on pancreatic function and growth in animal models), phytates (causing potential adverse effects on mineral status), phytohemagglutinins (with the unique property of being able to agglutinate red blood cells), and isoflavones (causing potential hormonal and goitrogenic effects). However, residual levels of these naturally occurring bioactive components, and of those formed during alkaline/heat processing such as lysinoalanine (an unnatural amino acid derivative) would be expected to be minimal in properly processed soybean protein products.

Adverse nutritional effects following consumption of uncooked and lightly heat-treated soybean protein have been attributed to the presence of endogenous inhibitors of digestive enzymes (about 56–64 mg per gram of protein of trypsin inhibitor activity) and saponins (0–10 μg per gram of protein) and to poor digestibility. To improve the nutritional quality of soy foods, inhibitors and saponins are generally inactivated by heat treatment or eliminated by fractionation during food processing. Soybean protein products also contain about 1–2% of phytic acid, which has long been recognized to interfere with the absorption of minerals, especially zinc. Phytic acid is stable to cooking and probably is not degraded during texturization by extrusion. Although long considered to be antinutritional factors, because of their estrogenic properties, soybean isoflavones are now of great interest because of their possible role in preventing many chronic diseases. Soy-protein flour has an isoflavone profile approximating that of soybeans (4028–4808 μg per gram of protein). Protein isolates contain reduced levels of isoflavones (690–1096 μg per gram of protein) compared to flours as a result of aqueous processing during manufacture. A concentrate made by aqueous alcohol extraction is very low in isoflavones (73 μg per gram of protein) because the isoflavones are especially soluble in aqueous alcohol and are thus largely removed during processing.

Soybean protein flours, concentrates, and isolates are good sources of all IAA, but some products could be marginally deficient in sulfur amino acids (Table 3). Although the addition of methionine to properly processed soybean protein products consumed by adults has no significant effect on the nutritional value when nitrogen intake is adequate, the addition of methionine to soy milk preparations increases the nitrogen retention by malnourished children and improves the protein quality of soy-based formulas for infants. The PER value of the methionine-supplemented soybean protein products is equivalent to that of casein. Because soybean-protein products have high levels of lysine, they can be used to fortify cereal proteins, which are limiting in lysine.

Wheat Gluten

According to the CCVP specifications, wheat gluten is a food product prepared by the wet extraction from wheat or wheat flours of certain nonprotein constituents (starch and other carbohydrates), in a manner to achieve a protein content of 80% or more ($N \times 6.25$) on a dry-weight basis. Although high in protein digestibility, wheat gluten is drastically limiting in lysine, with a PDCAAS of 36% (Table 4). Therefore, this protein source should not be used as the sole source of protein nutrition. However, the low protein quality of gluten can be improved by the addition of supplementary protein high in lysine, by supplementary lysine, or by protein complementation. (*See* **Protein**: Quality.)

Other Vegetable-protein Products

Besides soy protein products and wheat gluten, processing technology has also been developed for the

Table 3 Protein (on a percentage dry weight basis), IAA compositions, protein quality expressed as PDCAAS, and contents of minor bioactive components in soybean protein products

Nutritional parameter	*Flours*	*Concentrates*	*Isolates*
Total protein % (N × 6.25)	56–59	70–72	90–92
IAA: expressed as a percentage of the respective IAA in the FAO/WHO (1991) scoring pattern[a]			
Histidine	137	131	131
Isoleucine	164	171	175
Leucine	118	120	126
Lysine	110	110	115
Methionine + cystine	104	108	100
Phenylalanine + tyrosine	140	141	146
Threonine	115	132	112
Tryptophan	118	127	118
Valine	131	143	148
Protein digestibility (%)	86	94	96
PDCAAS (%)	89	100	96
Trypsin inhibitor activity (mg per gram of protein)	56–64 (raw flour), 16–19 (toasted flour)	8–11	2–29
Phytic acid (μg per gram of protein)	26–37	18–31	11–22
Isoflavones (μg per gram of protein)	4028–4808	104	690–1096

[a]FAO/WHO (1991) Recommended Amino Acid Scoring Pattern (containing, in milligrams per gram of protein included: histidine, 19; isoleucine, 28; leucine, 66; lysine, 58; methionine + cystine, 25; phenylalanine + tyrosine, 63; threonine, 34; tryptophan, 11; and valine, 35.).

Table 4 Protein (on a percentage dry weight basis), IAA compositions and protein quality expressed as PDCAAS of some vegetable-protein products

Nutritional parameter	*Wheat gluten*	*Sunflower-protein isolate*	*Pea-protein concentrate*	*Peanut flour*	*Canola-protein concentrate*
Total protein (%) (N × 6.25)	83	93	55	60	69
IAA: expressed as a percentage of the respective IAA in FAO/WHO (1991) scoring pattern[a]					
Histidine	131	140	126	100	126
Isoleucine	114	153	153	114	150
Leucine	95	95	117	91	118
Lysine	**38**	**38**	133	**55**	**98**
Methionine +cystine	132	100	104	84	177
Phenylalanine +tyrosine	119	135	135	128	112
Threonine	68	94	123	72	123
Tryptophan	87	125	**79**	71	136
Valine	103	146	142	101	150
Protein digestibility (%)	94	97	92	91	92
PDCAAS (%)	36	37	73	50	90

[a]FAO/WHO (1991) Recommended Amino Acid Scoring Pattern (containing, in milligrams per gram of protein included: histidine, 19; isoleucine, 28; leucine, 66; lysine, 58; methionine + cystine, 25; phenylalanine + tyrosine, 63; threonine, 34; tryptophan, 11; and valine, 35.).

preparation of a number of other VPP such as pea-protein concentrate, cottonseed-protein products, sunflower-protein isolate, peanut flour and canola (rapeseed)-protein concentrate (Table 4). However, among these protein products, only the pea-protein concentrate is distributed and used internationally. Pea-protein concentrate is a highly functional and nutritious protein concentrate derived from field peas. Careful processing to prevent protein denaturation has resulted in a wide range of useful functional properties. Moreover, due to its high lysine content, it can be used to greatly improve the nutritional quality of cereal proteins. Pea-protein concentrate has a protein digestibility of about 92% and a PDCAAS of 73%, with tryptophan and sulfur amino acids being the limiting IAA.

Like cereal proteins, sunflower-protein isolate and peanut flour are first limiting in lysine (Table 4). Sunflower-protein isolate has a high protein digestibility of 94%, and peanut flour has a somewhat lower protein digestibility of 91%. The PDCAAS values for sunflower-protein isolate and peanut flour are about 37 and 50%, respectively, because of the low lysine concentration.

In contrast to peanut and sunflower, canola (rapeseed) protein concentrate is one of the more balanced sources of IAA among VPP (Table 4). It is marginally limiting in lysine, and, along with its moderately low

protein digestibility of 92%, this gives it a PDCAAS of about 90%.

The properly processed defatted cottonseed flours (about 60% protein, N × 6.25) (glandless or deglanded (liquid cyclone process)) have a PER equivalent to that of casein. Heat and moisture introduced after removal of lipids significantly decrease the PER. Fractionation of the various types of cellular proteins during isolate preparation also produces both favorable and unfavorable effects. Storage protein isolates (99–100% protein) are low in lysine and the sulfur amino acids, and low in protein quality, but the nonstorage protein isolates (76–85% protein) are high in all IAA and in protein quality (PER > 2.5). The degree of fractionation of the seed proteins and, consequently, the nutritive value of the isolates are dependent upon the extraction characteristics of the flour and on the isolation process. The quality of defatted cottonseed products also depends upon the level of physiologically active or 'free' gossypol. Gossypol is a binaphthyl, dialdehyde, polyhydroxyl pigment which is highly reactive and deleterious to monogastric animals, including humans. All of the gossypol in the seed is in discrete extracellular glands. During processing, the glands usually are ruptured, and consequently, gossypol is either removed with the oil or converted to a physiologically inactive form through interaction with other cellular constituents. Glandless cultivars of cotton seed have also been developed. Edible cottonseed protein products, approved by the US FDA, have a maximum of 0.045% 'free' gossypol. IAA data and PER values of cottonseed-protein isolates are listed in Table 5. The protein quality of the minor, nonstorage-protein isolate is higher than that of the major storage-protein isolate, while the protein quality of the classical protein isolate is intermediate between the nonstorage and storage-protein isolates.

Table 5 Protein (on a percentage dry weight basis), IAA compositions and protein quality expressed as PER of cottonseed-protein isolates

Nutritional parameter	*Nonstorage protein*	*Storage protein*	*Classical*
Protein (%, N × 6.25)	76–86	99–100	92
IAA: expressed as a percentage of the respective IAA in the FAO/WHO (1991) scoring pattern[a]			
Isoleucine	146	121	135
Leucine	120	95	110
Lysine	109	55	82
Methionine + cystine	145	76	100
Phenylalanine + tyrosine	140	152	150
Threonine	114	85	99
Tryptophan	134	109	120
Valine	149	143	145
PER	2.7–3.2	1.3–1.7	2.0–2.2

[a]FAO/WHO (1991) Recommended Amino Acid Scoring Pattern (containing, in milligrams per gram of protein included: histidine, 19; isoleucine, 28; leucine, 66; lysine, 58; methionine + cystine, 25; phenylalanine + tyrosine, 63; threonine, 34; tryptophan, 11; and valine, 35.).

Table 6 PER and PDCAAS values for some properly processed vegetable-protein sources

Product	*PER (casein = 2.5)*	*PER as a percentage of casein*	*PDCAAS (%)*
Rapeseed-protein concentrate	2.6	104	96
Rapeseed-protein isolate	2.4	96	86
Soy-protein isolate	1.7	68	99
Pea-protein concentrate	1.2	48	68
Kidney beans	1.1	44	68
Pinto beans	0.5	20	63
Lentils	0.3	12	52
Peanut butter	1.7	68	52
Sunflower-protein isolate	1.1	44	37
Rolled oats	1.8	72	63
Wheat cereal	1.4	56	40
Wheat gluten	0.9	36	25

As can be seen from Table 6, the PER method of protein-quality determination gives a lower score than PDCAAS for proteins that are low in sulfur amino acids such as soy protein products, beans, and peas. This is because of the higher requirements for methionine and cysteine by rats compared to humans. In contrast, those first limiting in lysine, such as sunflower-protein isolate, have a PER that corresponds closely to PDCAAS.

Addition of beef to vegetable protein sources improves significantly the protein quality of soy, pea, peanut, sunflower and wheat-gluten-protein sources but only marginally improves that of rapeseed (Table 7).

Mycoprotein Products

Mycoprotein is the ingredient name for a food-grade protein source that has been available for food use only since 1985. As the name suggests, this protein is made from a type of fungus, *Fusarium venenatum* (PTA 2684), that was found in the wild and developed specifically for the production of food-grade protein. Mycoprotein production follows a series of steps. The organism is first grown in an aerobic, axenic, continuous fermentation system using food-grade carbohydrate substrates and other components needed for growth. The mycelium of the fungus is then heat-treated to reduce the ribonucleic acid content to safe levels. Single-cell protein sources tend to be high in nucleic acids, which can lead to excessive blood uric acid levels, a problem for people prone to gout. Once

Table 7 Protein quality expressed as RNPR values of beef, some potential meat extenders, and their mixtures with beef

Protein sources	*RNPR (%)*
Beef	100
Rapeseed-protein concentrate	95
Rapeseed-protein isolate	90
Soybean-protein concentrate	71
Soybean-protein isolate	67
Pea-protein concentrate	55
Peanut flour	53
Sunflower-protein isolate	40
Wheat gluten	34
Beef + nonmeat products (50:50 protein mixtures)	
Rapeseed-protein concentrate	96
Rapeseed-protein isolate	94
Soybean-protein concentrate	83
Soybean-protein isolate	80
Pea-protein concentrate	79
Peanut flour	80
Sunflower-protein isolate	85
Wheat gluten	85

Table 8 Protein (on a percentage dry weight basis) and IAA compositions and protein quality of mycoprotein

Nutritional parameter	*Mycoprotein*
Total protein (%) (N × 6.25)	56
IAA: expressed as a percentage of the respective IAA in the FAO/WHO (1991) scoring pattern[a]	
Histidine	184
Isoleucine	186
Leucine	130
Lysine	143
Methionine + cystine	116
Phenylalanine + tyrosine	141
Threonine	162
Tryptophan	145
Valine	177
Protein digestibility (%)	78
PDCAAS (%)	91

[a]FAO/WHO (1991) Recommended Amino Acid Scoring Pattern (containing, in milligrams per gram of protein included: histidine, 19; isoleucine, 28; leucine, 66; lysine, 58; methionine + cystine, 25; phenylalanine + tyrosine, 63; threonine, 34; tryptophan, 11; and valine, 35.).

the RNA level is reduced, the suspended solids are centrifuged and recovered in the form of a paste with 75% water content. This is the mycoprotein.

The vegetative growth of fungi, the mycelium, is in the form of hyphae – thread-like cells that permeate the growing medium. In the case of *F. venenatum* (PTA 2684), these hyphae, after the heat treatment, become insoluble and are of a length and diameter that make them suitable for creating products that mimic meats. To do this, binders, flavorings, and other ingredients are added and the mixture further processed to achieve the desired organoleptic and physical properties.

The introduction of this product onto the market required considerable toxicity and nutritional testing because it was a novel food not previously used for human consumption. As part of these tests, the protein quality was evaluated. Based on both PDCAAS and animal and human studies, it was determined that the protein quality of this mycoprotein is similar to that of soybeans. It has now been on the market under the brand name, Quorn™, in the UK and Europe for some 10–15 years and has secured considerable consumer acceptance. It has recently been introduced to the USA.

Mycoprotein is typically 12% water, 3% fat, 3% available carbohydrates, 6% fiber, and 2% ash. B complex vitamins and mineral nutrients are also present in small amounts. The amino acid profile and protein-quality indices are shown in Table 8.

A number of other features make mycoprotein of interest beyond being simply a protein source. It can be modified so as to function as either a fat replacer or a cereal replacer. Clinical studies have found that it can help control blood lipids and blood glucose as well as appetite, effects that may relate to its dietary fiber content.

Leaf-protein Concentrate

Green leaves are protein factories producing good-quality protein. Because of the low concentration, high fiber content, and palatability problems, however, potential sources are rarely exploited for human food use. When leaves are used as food, they are not generally considered important protein sources but rather as vegetables, such as lettuce, spinach, and cabbage. However, some green forage crops produce several times more protein per unit area than grain crops, and there may be circumstances where extracting the protein for direct human use has economic and ecological advantages over feeding it to domestic animals.

Consideration of the possible suitability of this protein for use as human food goes back at least to the 1940s and research to develop a potential product was conducted in the 1960s and 1970s in many parts of the world. In brief, it has been found that leaf-protein concentrate (LPC) can be obtained from harvested leaves by first pressing out the juice. The soluble protein in the juice is then heat-coagulated, the serum drained away, and the solid material dried and powdered. This produces 'whole' leaf protein concentrate (WLPC), which is about 45–60% crude protein, or 34–45% true protein. There are different types of protein in leaves that can be separated by coagulating at different temperatures. The green

Table 9 Protein (on a percentage dry weight basis) and IAA compositions and protein quality expressed as PDCAAS of alfalfa whole-leaf-protein concentrate

Nutritional parameter	*Whole-leaf-protein concentrate*	*Green-leaf-protein fraction*	*White-leaf-protein concentrate*
Total protein (%) (N × 6.25)	61.9	45.3	88.7
IAA: expressed as a percentage of the respective IAA in the FAO/WHO (1991) scoring pattern[a]			
Histidine	127	127	154
Isoleucine	194	221	195
Leucine	141	155	142
Lysine	107	101	113
Methionine + cystine	139	126	148
Phenylalanine + tyrosine	182	177	185
Threonine	146	139	169
Tryptophan	125	87	218
Valine	177	200	205
Protein digestibility (%)	88	87	96.4
PDCAAS (%)	94	76	100

[a]FAO/WHO (1991) Recommended Amino Acid Scoring Pattern (containing, in milligrams per gram of protein included: histidine, 19; isoleucine, 28; leucine, 66; lysine, 58; methionine + cystine, 25; phenylalanine + tyrosine, 63; threonine, 34; tryptophan, 11; and valine, 35.).

fraction is dominated by proteins from the chloroplasts, while the white-protein fraction is largely from the cell cytoplasm. The nature of these fractions is such that the white protein appears to be more suitable for incorporation into foods for humans since it is bland-tasting, white or off-white in color, and of a higher protein quality.

Any plant whose leaves can be eaten safely by humans is a potential source of LPC, which can also be used for animal feed. Some of these sources may currently be largely waste material or may be used as animal feed such as cassava leaves, or may be weeds such as dandelions and water hyacinth. Candidates should not have any toxic or antinutritional factors or unpalatable flavors, and preferably should provide high protein yields per unit area, be readily available, and be easy to harvest and process.

Investigations have shown that the amino acid profile of whole-leaf protein is very similar from one plant species to another. Taking alfalfa as an example, the protein digestibility of the WLPC is about 88%, and the amino acid score is 107, giving a PDCAAS of 95. Lysine is the limiting amino acid. The white-protein fraction has a PDCAAS of 100 (109) (see **Table 9**). The WLPC also contains significant amounts of vitamin A precursors, iron, and other nutrients which may be reduced in the white-protein fraction.

While an industrial process for producing WLPC has been developed, there appears to be little, if any, commercialization of this material for human use at present. Efforts have been made to set up facilities in developing countries as a way of converting this underutilized protein and nutrient source to useful, nutritious food and feed.

Microbial-biomass Protein

The original term for protein derived from yeast and bacteria was 'single cell protein.' The term 'microbial-biomass protein' was coined to permit the inclusion of protein from multicellular fungi, such as the mycoprotein described above. The interest in microbial-biomass protein was prompted by the same concerns of global protein shortages that brought about the development of LPC. It was also considered to be a possible way to produce useful food while reducing waste products and surplus raw materials resulting from some types of industrial and agricultural activities such as pulp mill waste liquor. The former Soviet bloc countries had, in the 1980s at least, about 100 plants producing microbial biomass protein, but very few existed elsewhere. The further development of these potential protein sources would seem to depend on the economical availability of a suitable substrate, the assurance that safe food materials are being produced, the lack of other more familiar or accessible alternatives, and the technology and facilities to produce attractive food products.

See also: **Celiac (Coeliac) Disease**; **Legislation**: Codex; **Mycoprotein**; **Protein**: Food Sources; Quality; **Soy (Soya) Beans**: Dietary Importance; **Wheat**: Grain Structure of Wheat and Wheat-based Products

Further Reading

Animal Feed Resources Information System, Feed Resources Group, Animal Production and Health Division, Food and Agriculture Organization of the United Nations, http://www.fao.org/ag/aga/agap/frg/afris/data/465.htm

Bickoff EM, Booth AN, de Fremery D *et al.* (1975) Nutrition evaluation of alfalfa leaf protein concentrate. In: Friedman M (ed.) *Protein Nutritional Quality of Foods and Feeds, Part 2*. New York: Marcel Dekker.

Friedman M (1996) Nutritional value of proteins from different food sources. A review. *Journal of Agricultural and Food Chemistry* 44: 6–29.

Graham R (2001) Mycoprotein – A meat alternative new to the U.S. *Food Technology* 55(7): 36–50.

Leaf for Life, http://www.leafforlife.com/pages/leaf_concentrate.htm

Pellett PL (1996) World essential amino acid supply with special attention to South-east Asia. *Food Nutrition Bulletin* 17(3): 204–224.

Pirie NW (ed.) (1975) *Food Protein Sources, International Biological Programme 4*. Cambridge: Cambridge University Press.

Protein Concentrates *See* **Whey and Whey Powders**: Production and Uses; Protein Concentrates and Fractions; Fermentation of Whey; Principles of Dialysis; Applications of Dialysis

PULSES

D D Dalgetty, B-K Baik and B G Swanson,
Washington State University, Pullman, WA, USA

Introduction

Of all the plants utilized by man, only grasses are more important than legumes. The term legume comes from the Latin word 'legumen,' which means seeds harvested in pods. *Leguminosae* is the Latin name of the third largest family of flowering plants. Of the 18 000 different species within the family *Leguminosae*, humans consume only a mere 24 species as food. A common name used to identify the food legumes is pulses. Pulse is a word most commonly used in the UK to denote edible legumes that store their energy in the form of starch. Other terms for pulses include beans or starchy legumes. Notable members of the pulses include peas (*Pisum sativum*), chickpeas (*Cicer arietinum*), and the common bean (*Phaseolus vulgaris*). Table 1 gives a list of these and other common pulses with their Latin designations.

Another very important category of food legumes are the oilseeds. As the name implies, these legumes store their energy in the form of lipid rather than starch. The two most common oilseeds are soybeans (*Glycine max*) and peanuts (*Arachis hypogaea*).

Food legumes can be further partitioned into separate classifications based on their growing climates as cool- and warm-season legumes. The four main species of cool-season food legumes are peas,

Table 1 Food legumes Latin and common names

Latin name	*Common name*
Arachis hypogaea L.	Groundnut, peanut
Cajanus cajan (L.) Millsp.	Pigeonpea, red gram, Congo pea, Arhar, Tur, Gongo pea
Cicer arietinum L.	Chickpea, Bengal gram, garbanzo gram
Glycine max (L.) Merr.	Soybean, soya
Lablab purpureus (L.) Sweet	Hyacinth bean, Egyptian bean, Val
Lathyrus sativus L.	Khesari, chickling vetch, grasspea
Lens culinaris Medik.	Lentil, Masur
Lupinus albus L.	White lupine
Lupinus angustifolius L.	Blue lupine, New Zealand blue lupine
Lupinus luteus L.	European yellow lupine
Macrotyloma uniflorum (Lam.) Verdc.	Horse gram, Madras gram, Kulthi
Phaseolus lunatus L.	Lima bean, butter bean
Phaseolus vulgaris L.	Bean, common bean, Franch bean, field bean, haricot bean, pinto bean, navy bean, dry bean
Pisum sativum L.	Common or garden pea, dry pea
Psophocarpus tetragonolobus (L.) DC.	Winged bean, Goa bean, four-angled bean, Manila bean, princess pea
Vicia faba L.	Broad bean, faba bean, horse bean
Vigna aconitifolia (Jacq.) Marechal	Moth bean, mat bean
Vigna mungo (L.) Hopper	Urd, black gram, mung bean
Vigna radiata (L.) Wilczek	Green gram, golden gram, mung bean
Vigna umbellata (Thumb.) Owhi & Ohashi	Rice bean, mambi bean
Vigna unguiculta (L.) Walp. ssp. unguiculata	Cowpea, black-eyed pea, crowder pea
Voandzeia subterranea (L.) Thouars	Bambarra groundnut

chickpeas, lentils, and faba beans. Cool-season legumes are very widespread geographically, with production concentration in temperate and subtropical climates. Some production occurs in the tropics at high altitudes. Cool-season legumes make up nearly 60% of the world pulse production and occupy around 40% of the world area. However, the distribution of each species over this area is very uneven. In developing countries, chickpeas and lentils are dominant in production and are a very common staple of the diet. The other two crops, peas and faba beans, are typically produced in more developed countries. Yields of these crops are much higher than for chickpeas and lentils.

The warm-season food legumes include the common beans and other tropical pulses such as cowpeas, lupines, and pigeonpeas. Aside from the common bean, most warm-season food legumes are produced in small amounts in specific regions. Warm-season legumes are well adapted to harsh conditions such as those in Africa where beans are an essential crop. These species are also very common in the Asian and Mediterranean regions.

Table 2 World production of major food crops (production in thousand metric tons)

Legume	*Total production*	*Developed nations*	*Developing nations*
Soybeans	92,982	64,401	28,581
Groundnuts	18,580	1,801	16,740
Dry beans	14,195	2,116	12,079
Peas	9,807	6,552	3,225
Chickpeas	6,158	346	5,812
Broad beans	4,224	464	5,759
Lentils	1,292	231	1,061

Distribution and Use of Edible Food Legumes

Among the cultivated legumes, soybeans, peanuts, dry beans, peas, broad beans, chickpeas, and lentils are the major crops produced worldwide. The other legumes are produced in more local settings in some countries depending on the climate needed to support the growth and food habits of the population. Table 2 shows the production (in tonnes) of the major food legumes in developed and developing nations. The breakdown of total production of food legumes into regions and countries of the world is illustrated in Tables 3 and 4. In many developed countries, the lack of pulse production is caused by low yields, uncertain harvests, slow maturation, and sensitivity of legumes to growing conditions at all periods of development and loss due to pest damage. Another reason for low production is cost in energy to prepare legumes with ensured digestibility.

Soybeans

Soybeans (*Glycine max* (L.) Merr.) are grown primarily throughout the world as a food crop. Nearly 46% of the world soybean production is in the USA. Other soybean-producing countries include Brazil, Argentina, China, Japan, Korea, India, Indonesia, Thailand, and the Philippines. The highest yields per hectare of soybeans have been achieved in the USA. (*See* **Soy (Soya) Beans**: The Crop.)

Groundnuts or Peanuts

The peanut (*Arachis hypogaea* L.) makes up a large part of the economy in many developing nations where most of its production occurs. The major peanut-producing nations are India, China, and the USA, of which India contributes 30% of the total world production. Peanuts are also produced in

Table 3 Regional food legume production (production in thousand metric tons)

Legume	*Africa*	*North and Central America*	*South America*	*Asia*	*Europe*	*Oceania*	*Russia*
Soybeans	241	63,088	14,456	15,047	672	103	600
Groundnuts	5,522	1,985	1,121	10,407	27	63	1
Dry beans	1,289	1,945	2,780	7,188	678	5	76
Broad beans	761	61	117	5,588	486	4	–
Peas	342	229	94	5,194	529	107	3,517
Chickpeas	227	260	26	6,839	113	–	–
Pigeonpeas	151	30	4	1,938	–	–	–
Lentils	56	71	53	816	75	–	3
Vetches	53	–	–	165	144	1	589
Lupines	11	–	7	1	62	55	114
Unspecified pulses	850	15	4	2,631	336	25	32

Table 4 Food legume production by nation (production in thousand metric tons)

Nation	*Soybeans*	*Peanuts*	*Other legumes*
World	92,982	18,580	44,701
China	7,477	3,873	6,858
India	650	5,700	11,107
Russia	460	–	6,508
US	61,970	1,561	–
Brazil	12,810	–	3,011
Mexico	672	–	1,430
Ethiopia	–	–	1,002
Trinidad and Tobago	–	–	1,356

most African countries. The highest peanut yields are in Europe, where only small amounts of peanuts are actually produced.

French Beans

French Beans (*Phaseolus vulgaris* L.) are widely cultivated in semitropical regions of the world. The leading producer of beans is Brazil, followed by India, China, the USA, and Mexico. Nearly 50% of the world production occurs in Asia with over 14 000 cultivars.

Peas

Peas (*Pisum sativum* L.) are produced mainly in cool areas. China and Russia contribute about 80% of the world production. *P. sativum* has been divided into the subspecies: *hortense, syracium, absyssinicum, jomardi, elatius*, and *arvense*. Garden peas have been separated into two types: smooth and wrinkled seeded. The wrinkled seeded peas tend to be sweeter than the smooth peas. (*See* **Peas and Lentils**.)

Chickpeas

One of the most important legumes in the developing world is the chickpea (*Cicer arietinum* L.). Among developing nations, India produces nearly 80% of the world production, with other major producers being Pakistan, Turkey, and Myanmar. World chickpea production has seen a decrease in recent years due to reductions in yield. Chickpeas are subgrouped into *Kabuli* and *Desi*. *Kabuli* chickpeas are very large creamy seeds and grow well under irrigated conditions. *Desi* chickpeas are much smaller and produce the best yields.

Pigeonpeas

Pigeonpeas (*Cajanus cajan* (L.) Millsp.) are a very important crop in Asia, Africa, and South America, ranking sixth in total world production of the food legumes, with more than 90% produced in India.

Lentils

The lentil (*Lens culinaris* Medik) is an important crop in developing nations, accounting for only 2% of total world pulse production with 75% of their production in Asia. Several lentil cultivars exist, with differences based on flower color and shape. Lentils are grouped into small and large seeded subgroups. The large seeded lentils are grouped into the subspecies *macrosperma* Baroulina, and the small seeds are grouped into *microsperma* Baroulina. Lentils are grown as a cold-season crop throughout the subtropics of India, North Africa, Central America, and South America. (*See* **Peas and Lentils**.)

Cowpeas

Cowpeas (*Vigna unguiculata* L.) are cultivated for their pods and seeds. The seeds are used as a vegetable and sometimes roasted like peanuts, while the roots are consumed in South Africa. Cowpeas are grouped into 'crowder peas,' 'brown crowders,' and 'black-eyed,' depending on the color of the seed and its arrangement in the seedpod. Cowpeas are produced in Asia, Africa, and the Mediterranean.

Consumption of Food Legumes

Food legumes are distributed throughout the whole world, with different species being popular in different regions. Legume preference for different areas of the world is due to the availability of the food in those areas. Availability results directly from environmental conditions favoring their growth. For people in developing countries of the world, where standards of living are limited due to a lack of buying power, legumes are a crucial part of the diet. Since the prices of other food commodities are unmanageable for the average family, pulses receive an enhanced level of consumption. Although legumes constitute a major portion of the diet in many developing nations, their most prominent use is in the form of feed for animals. In developed nations, where yields are significantly higher, the low market value of legumes provides an inexpensive source of animal feed. The popularity of processed cereal crops and knowledge of the antinutritional effects of legumes in these countries also inhibit the consumption of food legumes.

Africa

Cowpeas and groundnuts are the primary legume crops of West Africa, where consumption of 10–17 g per day per person is not unusual. In the humid, high-altitude regions of Central and East Africa, legume consumption is unusually high. Broad beans, chickpeas, and lentils are staples, with hyacinth beans,

lupines, lathyrus beans, and peas having local importance.

Asia

The leading consumer of legumes in Asia is India, whose average pulse consumption is around 34 g per person per day, with some variation between individual states. The most popular legume of the Indian nation is the chickpea, but pigeonpeas, black gram, mung beans, horse gram, and pea are also very important. In western Asia, popular legumes include beans, chickpeas, lentils, and peas. The haricot bean from America has also been successfully introduced into western Asia. South-east Asians do not consume large amounts of legumes, but a variety are grown for their use as vegetables. In the eastern Asian nations of China, Japan, Korea, and Indonesia, the dominant legume crop is the soybean. Soybeans are the basis of many traditional food products of eastern Asia. Groundnuts and pulses, along with the soybean, are very important foods in Indonesia.

In Asian countries, the consumption of pulses tends to increase with income level. This is primarily due to the predominance of soybeans in eastern Asia, which are of a higher value than most pulses. Typically, legume consumption tends to increase with decreasing income level, but because of the importance of the soybean in so many Asian dishes, the opposite is true. Higher income levels also show an increase in the consumption of animal products, but this does not change the consumption level of soybeans.

Latin America

By far the most predominant food crop of Central America and Mexico is maize, followed closely by the haricot bean. Although many pulses have been introduced into this region, their consumption and production are low. An important commercial crop of this region is the groundnut, which is also consumed in small amounts as a snack.

Pulse consumption in the islands of the Caribbean is much higher than in Central America and Mexico. Pulses such as cowpeas, kidney beans, and pigeonpeas are popular when mixed with cereals in main dishes. Most of the pulse consumption in the Mediterranean is in lower-income families, where as much as 50 g per person is consumed daily.

In most South American countries, legumes are a traditional part of the diet, except in the beef-producing countries of Argentina, Bolivia, and Uruguay. The main pulse crop of South America is the haricot bean, which is indigenous to this region of the world. Other important crops include lima beans and lupines. Chickpeas, lentils, peas, broad beans, and pigeonpeas have been introduced to South America and are quite successful. Another indigenous legume crop of South America is the groundnut, which is grown primarily for oil extraction, but some are consumed whole as food.

Europe

Legume consumption in Europe is rather high in England and in the poor Mediterranean countries (about 5 kg per year or 13–14 g per day) of southern and eastern Europe. The summers in southern and eastern Europe are hot and dry, enabling legume seeds to grow well and at a relatively low cost. Consumption of legumes is substantially lower in central and eastern European countries (about 2 kg per year or 5–6 g per day). However, in France, the average daily consumption of pulses and nuts is around 10 g.

North America

Large quantities of groundnuts are consumed in both the USA and Canada. Most pulses are consumed in the warmer, poorer areas of the southern USA. The highest USA consumption of 24 g per day is in rural nonfarm families of very low income, with the lowest consumption rates in high-income families.

Nature and Composition of Food Legumes

Seed Structure

The seeds of all legumes are generally very similar in structure, with major differences being in the method of energy storage. Pulses store their energy in the form of carbohydrate, and oilseeds store energy in the form of lipid. The major components of mature legume seeds are the seed coat (8%), the cotyledons (90%), and the embryo (2%). Like cereals, legumes do have an endosperm at the early stages of development, but as the seed matures, the energy is stored in cotyledons. The seed coat is the only source of protection for mature legume seeds. Therefore, legume seeds are unusually vulnerable to breakage, which can lead to damage to the cotyledons.

The seed coat and the cotyledons of legume seeds are very tightly attached to one another, which can make hull removal very difficult. The cotyledons of legume seeds, much like the endosperms of cereals, are made up of starch and protein. Within the cotyledon, the most abundant component is starch, which is surrounded near the outside of the cotyledons by protein bodies. The ability to remove the hull from the cotyledons is very important when trying to recover these fractions.

The outsides of legume seeds are typically characterized by three structures, which include the hilum,

the micropyle, and the rhape. The hilum is the point at which the seed breaks away from the stalk and looks like a large oval scar. The micropyle, which is next to the hilum, is a small opening in the seed coat where the pollen tube enters the valve. The raphe is a ridge opposite the micropyle and represents the base of the stalk. The hilum and the micropyle are very important to permeability and water absorption of the seed coat of legume seeds. Both vary in appearance, depending on the size and shape of the seed.

Legume seeds are very similar, but they do differ from each other in color, shape, size, and seed coat thickness. In general, the thickness of the seed coat is directly proportional to the lipid content of the seed. Oilseeds have much thicker seed coats than the pulses, which are starch-rich.

Chemical Composition

With the exception of peanuts, chickpeas, and soybeans, legumes can be described as containing approximately 10% moisture, 21–25% protein, 0.8–1.5% lipid, 60–65% total carbohydrates, and 2.2–4.0% ash. Table 5 illustrates the proximate composition of some commonly consumed food legumes. The energy provided by legumes falls in the range of 345–350 kcal per 100 g. The exceptions which all contain higher amounts of lipid are consequently of a higher caloric value. Along with a high fat content, soybeans contain the highest amount of crude protein of all food legumes. Due to the high protein and fat content of soybeans and peanuts, the total carbohydrate content of these legumes is much lower than in other legume species.

Carbohydrates Carbohydrates of legumes, much like the carbohydrates of other grains, can be divided into water-soluble components, such as sugars and pectin, and insoluble components, such as starch and cellulose. Within both the soluble and insoluble fractions of legume carbohydrates are components that can be used as sources of energy and those that cannot be utilized because they are resistant to human digestive enzymes. (*See* **Carbohydrates**: Classification and Properties.)

The most dominant carbohydrate in most legumes is starch, but some seeds such as winged beans, soybeans, and those seeds containing gum tend to lack starch. Legume starch is located in the cotyledon as granules embedded in a very thick protein matrix. The average size of legume starch granules ranges from 25 to 28 μm, varying between species. The two primary constituents of starch are amylose and amylopectin, which are glucose homopolymers. Starch is the primary storage carbohydrate in legumes, but other carbohydrates such as mono-, di-, and oligosaccharides as well as cyclitols and sugar alcohols have received attention because of their association with flatulence.

The mono- and disaccharide content of legumes is generally less than 4%. Soybeans have been reported to contain as much as 7% sucrose. Among the oligosaccharides found in legumes are raffinose, stachyose, and verbascose. These compounds are galactosides of sucrose containing one, two, and three molecules of galactose, respectively. The human digestive system does not contain the enzyme α-galactosidase, which is required for the digestion of such oligosaccharides. Without the ability to digest these compounds, they are subject to occasional anaerobic fermentation. This results in gas production or flatulence, which can be very discomforting.

The other carbohydrate fraction of legumes is dietary fiber, which is composed primarily of cellulose, hemicelluloses, and lignin. Legumes contain a wide range of unusual storage dietary fibers with unique properties. The water-soluble gums making up the dietary fiber of legumes have the ability to form gels in water, which provides unique cooking characteristics. Both the soluble and insoluble fractions of dietary fiber are very beneficial nutritionally as they play key roles in laxation, blood-glucose attenuation, and blood-cholesterol attenuation.

Protein The protein of legume seeds is located primarily within the cotyledon, with only a small portion present in the seed coat. Quantitatively, legumes are an excellent source of protein, but qualitatively, they have many problems. Probably the most notable problem with legume protein is the lack of sulfur-containing amino acids. Another major problem is the presence of proteinase inhibitors, which lower the digestibility of legume proteins considerably. (*See* **Protein**: Food Sources.)

The amino acid composition of legumes has been widely studied and has shown that lysine levels are very high, while cysteine and other sulfur amino acid

Table 5 Proximate composition of common food legumes

Legume	*Protein (%)*	*Lipid (%)*	*Crude fiber (%)*	*Ash (%)*	*Starch (%)*
Chickpeas	24.7	6.0	1.2–13.5	3.08	37.2–50.0
Lentil	27.7	0.9	3.8–4.6	2.56	34.7–52.8
Lupine	47.4	1.2	3.0	3.3	0.3–0.5
Mung beans	30.5	0.9	1.2–12.8	3.5	37.0–53.6
Navy beans	28.0	–	3.4–6.6	–	27.0–52.7
Pinto beans	28.0	1.9	4.3–7.2	–	51.0–56.5
Smooth peas	27.8	1.1	4.6–7.0	3.1	36.9–48.6
Soybean	49.9	23.3	2.4–5.5	5.3	0.2–0.9
Wrinkled peas	26.1	1.8	7.6	3.5	24.0–36.6

contents are quite low. This property makes legumes a natural complement to cereal products, which contain high levels of sulfur amino acids and relatively low levels of the essential amino acid lysine. In general, legumes exhibit a large variation in amino acid content, depending mainly on the environmental conditions.

Legume protein has often been categorized as poor and of low biological value. This has mainly been attributed to the low concentration of sulfur amino acids and poor digestibility. Besides proteinase inhibitors, the poor digestibility of legumes is caused by trypsin inhibitors, lectin, tannins, and phytates. Trypsin inhibitors not only reduce the digestibility of legume protein but also remove a significant fraction of cysteine, which is already low. Lectin is a phytohemagglutinin that combines with cells lining the intestinal wall and causes nonspecific interference with absorption of nutrients. Tannins are known to form complexes with proteins, which decrease the digestibility and amino acid availability. Phytic acid interacts with protein to form phytate–protein complexes that are insoluble. The most efficient way to improve the digestibility of legume protein is through heat processing. By heating in the presence of water, proteins can easily denature and become subject to proteolytic enzymes to aid digestibility. (*See* **Protein**: Quality.)

Lipids In general, food legumes tend to have higher lipid contents than cereals. The lipid content of pulses ranges from 1 to 7%, while soybeans and peanuts contain 20–45% lipids. Due to their high lipid content, soybeans and peanuts are very good sources of edible oil. The fatty acid composition of legumes varies, with the predominant fraction being the unsaturated oleic and linoleic acids. Oils from temperate-zone legumes tend to have more unsaturated fatty acid components than those of the tropic regions. Some legumes also contain considerable amounts of linolenic acid. (*See* **Fats**: Occurrence.)

Vitamins and minerals The vitamin content and availability in legumes vary as a result of both genetic and environmental factors. Legumes are very good sources of B vitamins. The thiamin contents of legumes are usually equivalent to, or higher than, those of cereal grains. Legumes contain small amounts of riboflavin but tend to be a good source of nicotinic acid. The pantothenic acid content of legumes is less than that of cereal grains, but in comparison with most common foods, legumes are good sources of folic acid. In general, legumes contain very small amounts of vitamins.

The mineral content of legumes is in the range of 2.5–4.2%. The most predominant mineral found in legumes is phosphorus, averaging around 300 mg per 100 g. The calcium and iron contents of legumes are variable, but in general, legumes are good sources. Minerals are more concentrated in the seed coat of legumes than in the cotyledons, but because of the size of the cotyledon, most of the total mineral content is found in this fraction.

Antinutritional factors The most prevalent antinutritional factors in legumes are those dealing with protein, but there are also a number of other factors that decrease the nutritional value of food legumes. One such factor is the presence of goitrogens in oilseeds. Goitrogens are thioglycosides or glucosinolates that react with iodine and lead to the formation of goiter in humans. This effect can be counteracted in soybeans by heating, but not in peanuts. Another problem with legumes is the presence of mineral chelating agents such as phytic and oxalic acid. This can lead to a decrease in the availability of certain trace minerals, thus causing anemia. (*See* **Plant Antinutritional Factors**: Characteristics.)

Cyanogenic glycosides that are naturally present in legumes can be cleaved by the action of the β-glucosidase enzyme producing cyanide, which is a toxicant. Other antinutritional compounds in legumes include α-amylase inhibitors, saponins, allergens, and toxic amino acids that have been known to exhibit antivitamin and antihormonal activity.

Processing and Food Use of Legumes

Storage of Food Legumes

Food legumes are stored at all levels from farmer to trader in various types of storage units until the next season's harvest. Farmers generally retain a portion of their harvest for family consumption, for seed, and for animal feed. In developing nations, legume storage is done at the trader and government levels as insurance against future crops of low yield or poor quality and against fluctuation in price and market demand. Losses of legume seeds due to storage are quite high in comparison to other crops. These losses are the result of physical, biological, and engineering factors. Physical factors causing loss are temperature, moisture, and oxygen concentration. Biological causes of loss are birds, rodents, insects, mites, bacteria, and yeast. Engineering factors resulting in loss are storage structures and process handling equipment.

Food Uses of Legumes

Typically, the leaves, stalks, and seed coats of legumes are used as animal feed in many underdeveloped

countries of the world. These parts of the plant are generally not used as human food at a household level but may find use in some bulk products. Humans usually consume legumes in the form of whole seeds in a variety of ways. One of the most common ways in which legume seeds are consumed by humans is in the form of a dhal. A dhal is a legume seed that has been physically or mechanically separated at the cotyledon, resulting in dehusked, split cotyledons that are then soaked in water and cooked. Cooked dhals are seasoned appropriately and consumed with cereal-based products such as bread or cooked rice. Seasoning for legume seeds varies considerably by region, with common seasonings being onion, garlic, cilantro, green chilies, ginger root, black mustard seeds, cumin seeds, anise seeds, coriander, and turmeric. To prepare these dishes, the seeds and seasonings are heated separately and then mixed and boiled, resulting in a soup-like product. In addition to this common method of consumption, a variety of different products are prepared from food legumes.

Soaking legumes in water is often the first step to prepare the seeds for consumption or processing. Seeds are usually soaked for up to 24 h to loosen the hulls and help remove antinutrients such as tannins, oligosaccharides, and β-glysosides. In many cases, food legumes are germinated to form sprouts. Legume sprouts are a very important part of the diet in the Far East. By germinating legumes, the flavor and nutritional properties of the seeds are improved through the breakdown of antinutrients and phytic acid, which leads to flatulence.

Another legume process traditionally used in Africa and Asia is milling or removal of the outer portion of the seed. Milling occurs at all levels from the home kitchen to large-scale commercial operations. Prior to milling, legume seeds are often pretreated using wet and dry methods to aid in hull removal. The wet method involves mixing legume seeds with small amounts of water, followed by draining and sun drying. The dry method, commonly used in India, involves sun drying legume seeds that have been treated with oil and a very small amount of water. Both methods cause shrinkage of the endosperm, thus loosening the seed coat for easy removal by milling. The resulting legume flour is used to make a variety of breads and pastries.

One of the oldest methods of food processing that is often used with food legumes is fermentation. In many parts of the world, fermented legume products serve a very important function in the diet. Most fermented legume products are widely used in Asia and India but are rapidly gaining popularity in the western hemisphere. Legume-based fermented products are popular because of their texture and organoleptic characteristics. By fermenting legumes, it is possible to eliminate undesirable flavor, improve digestibility, prolong shelf-life, and increase the safety of the food. Popular legume-based fermented products include soy sauce, shoyu, miso, sufu, natto, and tempeh. Food legumes are also very popular in the canning industry. Most of this is the increased popularity of refried beans, soups, and other ready-to-eat dishes.

Byproduct Processing

Whole legumes, although produced and used widely throughout the world, have a low market value in comparison to other food commodities. To increase the value of legumes, some processors isolate functional byproducts from whole seeds. By far the most popular legume byproduct, due to its high availability, is protein. Legume protein concentrates and isolates are prepared by air classification and solvent solubilization. Air classification is a physical method in which legume flour is separated by density into a protein concentrate and a starch-rich fraction. In solvent solubilization techniques, soluble protein is removed from insoluble material by centrifugation, followed by an acid precipitation. Other byproducts of food legumes include starch and dietary fiber, which can be used to fortify and increase the functionality of various food products.

See also: **Beans**; **Goitrogens and Antithyroid Compounds**; **Legumes**: Legumes in the Diet; Dietary Importance; **Peanuts**; **Peas and Lentils**; **Plant Antinutritional Factors**: Characteristics; **Soy (Soya) Beans**: The Crop

Further Reading

Arora SK (1983) *Chemistry and Biochemistry of Legumes*. India: Oxonian Press.

Hedley CL (2001) *Carbohydrates in Grain Legume Seeds: Improving Nutritional Quality and Agronomic Characteristics*. New York: CABI.

Huisman J, van der Poel TFB and Liener IE (1988) Recent advances of research in antinutritional factors in legume seeds. In: *Proceedings of the First International Workshop on 'Antinutritional Factors (ANF) in Legume Seeds,'* November 23–25, 1988, Wageningen, The Netherlands.

Matthews RH (1989) *Legumes – Chemistry, Technology and Human Nutrition*. New York: Marcel Dekker.

Nwokolo E and Smartt J (1996) *Food and Feed from Legumes and Oilseeds*. New York: Chapman & Hall.

Salunkhe DK and Kadam SS (1989) *Handbook of World Food Legumes: Nutritional Chemistry, Processing Technology, and Utilization*, vols. 1–3. Boca Raton, FL: CRC Press.

Salunkhe DK, Kadam SS and Chavan JK (1985) *Postharvest Biotechnology of Food Legumes*. Boca Raton, FL: CRC Press.

Smartt J (1990) *Grain Legumes: Evolution and Genetic Resources*. Cambridge: Cambridge University Press.

Sosulski F and Garratt MD (1976) Functional properties of ten legume flours. *Journal of the Canadian Institute of Food Science and Technology* 9: 66–69.

Summerfield RJ and Roberts EH (1985) *Grain Legume Crops*. London: Collins Professional and Technical Books.

Pyridoxine *See* **Vitamin B_6**: Properties and Determination; Physiology

Q

Q Fever *See* **Zoonoses**

QUALITY ASSURANCE AND QUALITY CONTROL

P J Molnar, Central Food Research Institute, Budapest, Hungary

Background

A principal factor in the performance of a food enterprise is the quality of its products. There is a worldwide trend towards more stringent customer expectations with regard to food quality and safety. Accompanying this trend has been a growing realization that continual improvements in quality are often necessary to achieve and sustain good economic performance.

Today, every customer demands quality from their supplier. In particular, the last customer in the chain of suppliers and customers – the consumer – has increasingly placed vociferous demands on quality in the sense that (1) the future of the world will not be endangered by harm to the environment resulting from food production and (2) the consumer will suffer no economic damage or injury to health as a result of the consumption of food components.

Definition and Description of Food Quality and Safety

The quality of food can be defined as a composite of characteristics which affect the ability of foods to satisfy definite requirements and which determine its fitness for consumption.

Quality in this sense can be discussed under three headings:

1. Sensory quality, the basic characteristics of which are the product's color, appearance, texture, tenderness, juiciness, taste, and smell; these characteristics, being easily perceptible, are the basis of evaluation and preference on the part of consumers. (*See* **Sensory Evaluation**: Sensory Characteristics of Human Foods.)
2. Nutritive value, which determines the content and level of essential nutrients (biological value), some microbial components (e.g., desirable lactic bacteria), energy content, dietary value, digestibility, and availability. (*See* **Bioavailability of Nutrients**.)
3. Convenience and technological quality, characterizing the easy handling and functionality of a food product, to manufacturers and tradesmen and then to consumers; these qualities are based on characteristics such as shelf-life, processing level, good price–quality relation, durability, ease of storage and transportation, dimensions, functionality, and usefulness in the production of definite food products. (*See* **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability; Shelf-life Testing.)

Wholesomeness, also called food safety or food hygiene, may be defined as the result of all conditions and measures that are necessary during the production, processing, storage, distribution, and preparation of food to ensure that it is safe, sound, wholesome, and fit for human consumption. The aim of all these conditions and measures is to protect the health of people as consumers and to ensure fair practice in the food trade.

The wholesomeness of food determines its fitness for consumption, the classification of which should be based on specified sanitary criteria. The bases of these criteria are the assumptions that a food product must not (1) be harmful to human health, (2) have undergone deterioration (spoilage), (3) have been adulterated, and (4) have been produced, distributed, or stored under improper sanitary or physical conditions.

Factors harmful to food consumers include the following:

1. Pathogenic microorganisms and their toxins.
2. Pathogenic parasites and products of their metabolism.
3. Foreign substances, or their degradation products, with toxicity to humans. (*See* **Parasites**: Occurrence and Detection; **Spoilage**: Bacterial Spoilage; Molds in Spoilage; Yeasts in Spoilage.)

In relation to pathogenic microorganisms and parasites, there is an obligatory sanitory rule, which states that they must not occur in foods owing to their decisively negative influence on human health. Foreign substances may be differentiated as follows:

1. Food additives introduced intentionally to foods to improve their utilization quality, and
2. Technical and accidental contaminations present in food as residues resulting from technological processes or environmental pollution. These include toxic metals, pesticides, medicines, fodder supplements (premix), and growth stimulants. (*See* **Contamination of Food**; **Food Additives**: Safety.)

The main components of food quality and safety are shown in Figure 1.

Role of Quality Assurance and Quality Control

Quality assurance can be characterized by the quality loop for processed food (Figure 2). Quality assurance is consistent only if the origins of lack of quality are eliminated, or the emergence of faults is excluded from the beginning: problems result in, rather than originate from, exceeded limits of contamination, adulteration, etc. The quality-assurance system is based on the following:

1. Product planning and development, including the results of marketing and market research.
2. Education and training for management and quality staff as well as for each counterpart.

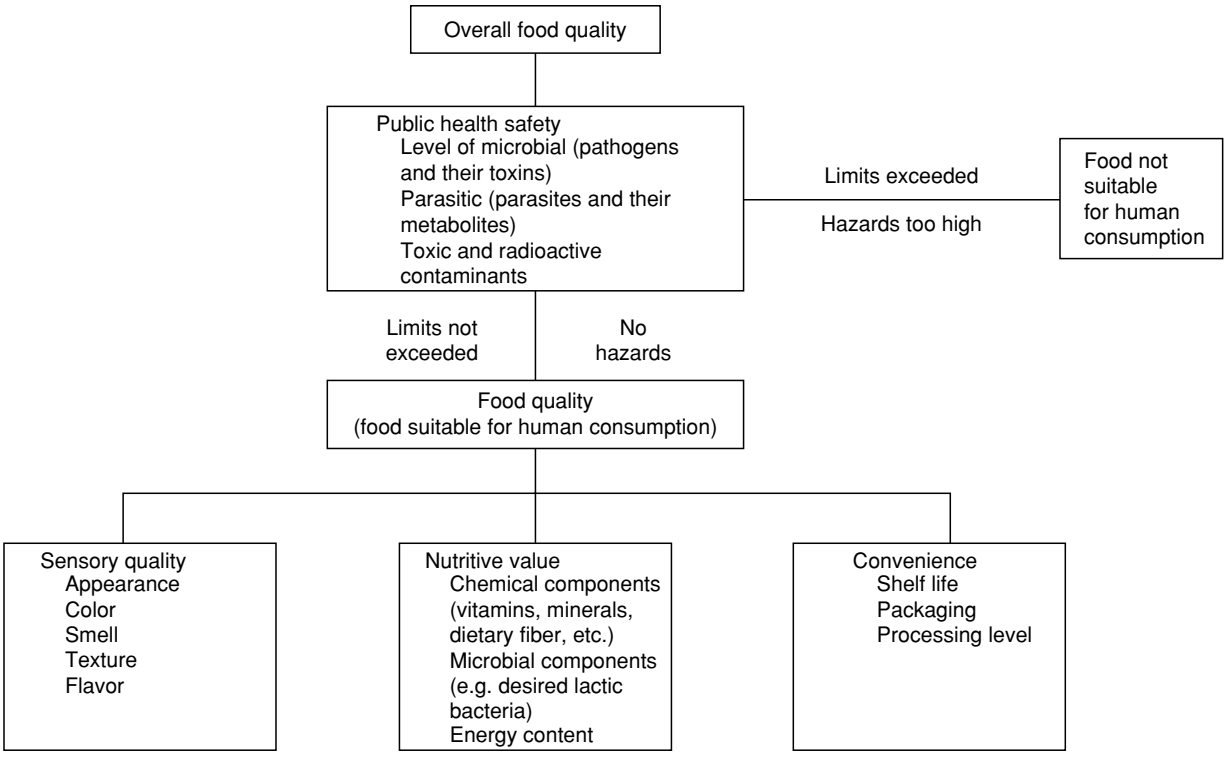

Figure 1 Main components of food quality and safety.

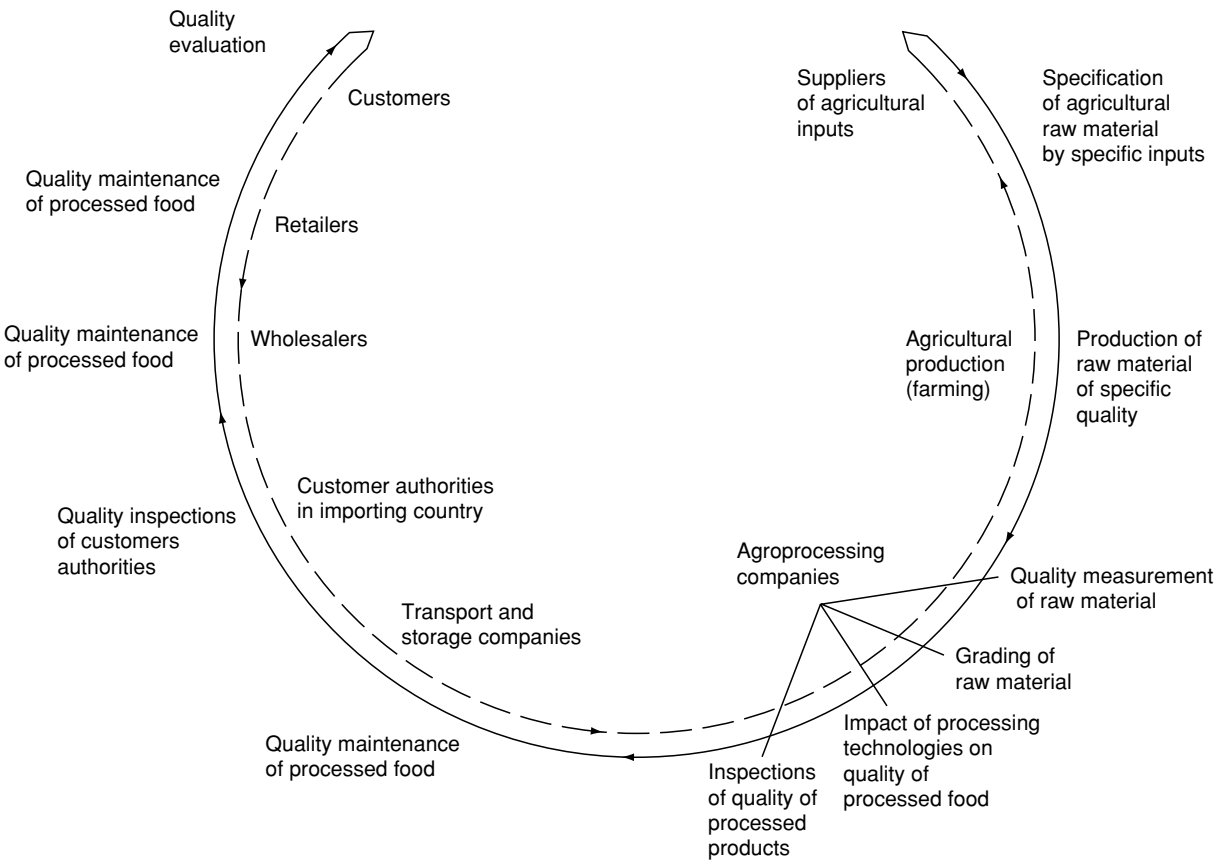

Figure 2 Quality loop for processed food: flow of products and flow of information on quality requirements. ———, main food chain; — — background information flow.

3. Application of a statistical control method and sampling inspection plan in process control and for final inspection.
4. Responsibility for the quality in each step of the quality loop.
5. Use of an information and motivation system on product quality.

Assisted by the internal quality system of a company and with additional commitment from the external quality partners (suppliers and customers), the product-oriented quality-assurance system is structured in the following steps:

1. Definition of requirements and responsibilities with regard to the quality of individual activities, processes, and materials.
2. Definition of the desired properties and required parameters of the product.
3. Definition of materials, technologies, packagings, storage conditions, number and qualification of personnel, etc.
4. Analysis of possible sources of faults and factors affecting quality.
5. Definition of the quality-assurance actions: information, education, training, motivation, parameters of investigations, methods of analysis and sampling documentation, etc.
6. Risk assessment and hazard analysis critical control point (HACCP).
7. Definition of responsibilities for certain quality-assurance actions.

Appropriate quality-control procedures should be part of the process to ensure product safety and quality and hence part of the company quality assurance system. Figure 3 shows the quality control of food manufacturing, including inspection and laboratory investigation, in a general flow chart.

All plant operations should be controlled by written procedures and records of performance data. The range and complexity of procedures vary, depending upon the size and scope of the operation. However, management personnel of even the smallest operation should be aware of the need for quality control consistent with the type of food products being manufactured, stored, or distributed. In larger, more complex operations, the inspector should determine whether

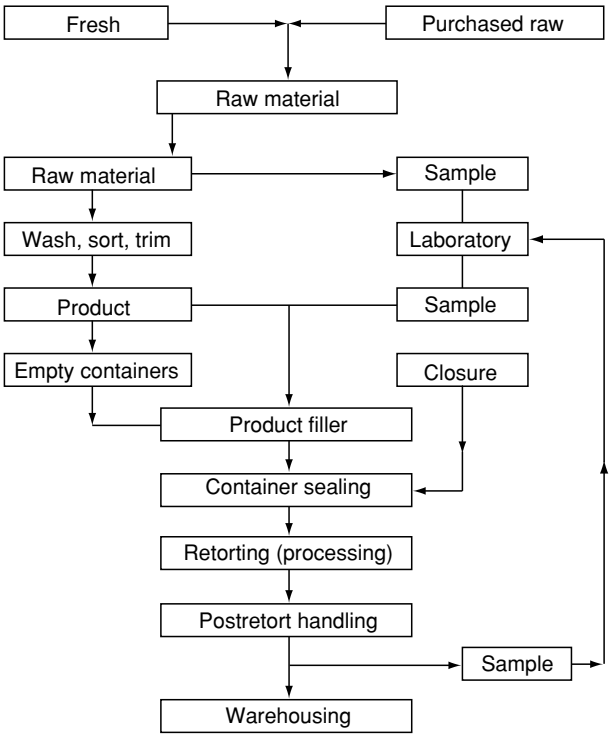

Figure 3 General flow chart including quality control.

written control procedures exist and should review records detailing results of quality-control testing relating to product safety or other factors where regulatory requirements exist.

The quality-control department analyses any planned production changes. Final approval for production should not be given until all quality-control records have been reviewed and properly endorsed by the quality-control management team.

Laboratory personnel and equipment must be evaluated to determine whether or not they are able to carry out all their responsibilities under the firm'sspecifications and existing regulations. Laboratory inspection should include spot checks of in-process controls made by personnel on the production lines.

Implementation in Food Production (Good Manufacturing Practice)

The guidelines on good manufacturing practice (GMP), elaborated and published initially for drugs, were later modified for human foods and oriented mainly to food safety. The formulated criteria apply in determining whether or not the facilities, methods, practices, and controls used in the manufacture, processing, packing, or holding of food conform with, or are operated or administered in line with, GMP to ensure that food for human consumption is safe and has been prepared, packed, and held under sanitary conditions.

More recently, systematic, rule-based approaches to quality assurance have been widely applied, most notably that based on the HACCP program. Such techniques provide ways of applying our current knowledge of microbial ecology of food in a systematic manner to ensure that nothing is overlooked and that the risk of microbiological food contamination is minimized. (*See* **Hazard Analysis Critical Control Point**.)

The HACCP program includes an assessment of potential microbiological, chemical, and physical hazards, prescribes the elimination of avoidable hazards, and sets tolerances for the hazards that cannot be eliminated in the processing of food. It defines the appropriate control measures, the tests to be carried out, and the criteria for product acceptance. It gives a rational systematic documented procedure which can be used for organizing and implementing the entire quality-assurance program.

However, GMP includes not only food safety, but also, as a part of the quality-assurance system, the application of product specifications. The specification for a product is usually the description of the requirements with which the product has to comply. These can have a regulatory, commercial, or in-house basis. Specifications are an integral part of product development, manufacture, marketing, and sales. From primary producer to the retail trade, each industry has its own specific requirements in the description of the following characteristics of their products:

- name;
- product description;
- specific properties;
- sensory attributes;
- chemical parameters;
- physical data;
- microbiological parameters;
- shelf-life;
- storage conditions;
- packaging and labeling;
- handling;
- proper sampling and test methods.

From a company viewpoint, specifications can have an internal and external function. The internal function comprises purchasing criteria, incoming goods control, and recording of knowledge and information in a systematic and accessible way. The external function includes sales criteria, sampling procedures, and quality control of finished products. It simplifies

functional contacts between companies and can serve as a base for product comparisons. This has led to the development of quality system standards and guidelines that complement relevant product requirements given in the technical specifications.

The International Standards Organization (ISO) 9000 series of standards set by the ISO embodies a rationalization of the many and various national approaches in this sphere which must be applied in adequate form under the circumstances of the food industry. According to this standard, an increasing number of food companies have a quality-assurance handbook or quality manual covering the following main topics:

1. General quality policy.
2. Description of the quality system and the responsibilities and interrelationships of all operating and staff functions.
3. General description of quality planning requirements, with specific details for each product category where appropriate.
4. Policy requirements for specific elements of the system (relating to GMP) and procedures which implement these policies.

The quality manual is of direct benefit not only to the producer (e.g., cost reduction, help in product liability cases, or improved competitively), but also to the consumer, who will perceive its effect as a high level of quality that will be produced and maintained by the application of an integrated and documented quality-assurance system.

Food-quality and safety programs in many countries are currently in a period of reevaluation and change. Risks to consumers resulting from hazards in food largely depend on the means available to prevent, eliminate, or reduce the risks to acceptable levels through the food chain. Management of foodborne risks in these terms is a complex independent activity, and risk-management strategies must include consideration of all options, including general approaches such as adherence to GMP, good agricultural practice, good veterinary practice, consumer education, and labeling.

Total quality management (TQM) is a modern term to describe how firms are becoming more successful today. The results of TQM are significant in productivity gains, unit cost reduction, elimination of most consumer complaints, and winning consumers by always giving them what they expect. TQM based on GMP and HACCP includes quality assurance and quality control as well as problem-solving by team work, and by the quality tools, risk management, cost of quality and rewards.

Implementation in the Food Laboratory (Good Laboratory Practice)

Whereas quality assurance of the work performed by laboratories was formerly based largely on confidence in the individual laboratory and in individual members of staff, a more formalized requirement has recently been developed for factual documentation of quality in the results of the investigations presented. In this respect, the work has proceeded largely along international lines where, for example, the ISO has been extremely active.

Incorrect results from a laboratory may have serious health or economic consequences as a result of an unsuitable food incorrectly being assessed as suitable, or by unnecessary rejection of an acceptable food. To ensure reliable analytical results and avoid mistakes, the laboratories must continuously check and improve the quality of analytical work.

The quality manual, which consists of general advice for 'quality assurance,' should be supplemented with more detailed regulations for individual laboratories. Consequently, it is recommended that each laboratory prepare its own quality-assurance directives in a quality manual as a supplement to the general principles. A quality manual is individual and therefore differs from laboratory to laboratory, and so each laboratory needs to prepare its own manual. However, the following information is necessary in almost every case:

1. Requirements placed on the laboratory: staff (management, experts, chemists, physicists, microbiologists, etc.); premises (type and extent of the activities); apparatus (equipment, instruments, etc.); glassware and plastic equipment; chemicals, gases, and solvents.
2. Sampling (sampling and sampling staff do not necessarily belong to the laboratory).
3. Sample reception at the laboratory: marking and keeping records; storage of samples before and after analysis.
4. Sample preparation.
5. Selection of methodology.
   - *Defining method* (type I) is one which determines a value that can only be arrived at in terms of the method *per se* and serves for calibration purposes.
   - *Reference method* (type II) is designated where type I methods do not apply. It should be selected from type III methods and should be recommended for use in cases of dispute and for calibration purposes.
   - *Alternative approved method* (type III) is one which meets the criteria required by the Codex

Committee on Methods of Analysis and Sampling for methods that may be used for control, inspection, or regulatory purposes.

- *Tentative method* (type IV) is one which has been used traditionally or else has been recently introduced but for which the criteria required for acceptance by the Codex Committee on Methods of Analysis and Sampling have not yet been determined.

6. Documentation and reporting: registration; laboratory register; analytical documentation; use of computers and electronic data processing; method description.
7. Quality assurance of analytical results: acceptance criteria for an evaluation of analytical results; standard substances, recovery tests, and reference materials; repeated determinations; intercalibration and collaborative tests; control chart.

The intention with a quality manual is to provide a general survey of the factors that influence analytical reliability at the food laboratories. The implementation of good laboratory practice and the guidelines should be of value to all the staff working within the laboratory and should inspire the creation of efficiently operating quality-assurance systems for laboratory work within the food companies. (*See* **Analysis of Food**.)

See also: **Analysis of Food**; **Bioavailability of Nutrients**; **Contamination of Food**; **Food Additives**: Safety; **Hazard Analysis Critical Control Point**; **Parasites**: Occurrence and Detection; **Sensory Evaluation**: Sensory Characteristics of Human Foods; **Spoilage**: Bacterial Spoilage; Molds in Spoilage; Yeasts in Spoilage; **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability; Shelf-life Testing

Further Reading

ALINORM 97/9-Rev. 1 (1997) *The Application of Risk Analysis Principles in Codex*. Rome: Codex Alimentarius, Food and Agriculture Organization.

ASQC (1986) *Food Processing Industry Quality, Control System Guidelines*. Milwaukee, WI: American Society for Quality Control.

FAO (1990) *Manuals of Food Quality Control: 5. Food Inspection*. Rome: Food and Agriculture Organization.

FAO (1995) *The Use of Hazard Analysis Critical Control Point (HACCP) Principles in Food Control*. Rome: Food and Agriculture Organization.

FAO (1997) *Risk Management and Food Safety*. Rome: Food and Agriculture Organization.

Gould W A (1992) *Total Quality Management for the Food Industries*. Baltimore, MD: CTI.

ISO (1983) *ISO Guide 25, EN 45001*. Geneva: International Standards Organization.

ISO (1994a) *ISO 9001 Quality Systems. Model for Quality Assurance in Design, Development, Production, Installation and Servicing*. Geneva: International Standards Organization.

ISO (1994b) *ISO 9002 Quality Systems. Model for Quality Assurance in Production, Installation and Servicing*. Geneva: International Standards Organization.

ISO (1994c) *ISO 9003 Model for Quality Assurance Final Inspection and Test*. Geneva: International Standards Organization.

ISO (1994d) *ISO 9004 Quality Management and Quality System Elements. Part 1: Guidelines*. Geneva: International Standards Organization.

ISO (1994e) *ISO 9004 Quality Management and Quality System Elements. Part 2: Guidelines for Services*. Geneva: International Standards Organization.

Nordic Council of Ministers (1990) *Quality Assurance Principles for Chemical Food Laboratories*. Copenhagen: Nordic Council of Ministers.

Quarg *See* **Cheeses**: Types of Cheese; Starter Cultures Employed in Cheese-making; Chemistry and Microbiology of Maturation; Manufacture of Extra-hard Cheeses; Manufacture of Hard and Semi-hard Varieties of Cheese; Cheeses with 'Eyes'; Soft and Special Varieties; White Brined Varieties; Quarg and Fromage Frais; Processed Cheese; Dietary Importance; Mold-ripened Cheeses: Stilton and Related Varieties; Surface Mold-ripened Cheese Varieties

QUINOA

S A Valencia-Chamorro, Escuela Politécnica Nacional, Quito, Ecuador

Introduction

Quinoa is a typical crop of the Andean region. It has been cultivated since ancient times, and was a staple food of the Inca empire. However, production of quinoa was almost completely reduced after the Spanish conquest; foreign grains were widely produced instead.

During the 1970s, initial interest in the crop was increased. At the same time, higher plantation areas for cultivation of quinoa in South America and new markets for the grain in the USA and Europe were developed. Thus, quinoa is now of great interest to the scientific world, and it may become a food for the future.

The origin of quinoa and its distribution in the Andean region have been reviewed, as well as its actual distribution in North and South America and Europe. It is important to mention the strong and dedicated interest in cultivating this grain outside the borders of the former Inca empire. Classification of the grain, cultivation practices, and morphology of the plant have been described.

Quinoa has been recognized as an extremely nutritious grain all over the world thanks to both the high quantity and good quality of its protein content as regards its essential amino acid content. Thus, an extensive review of the chemical composition and nutritional value of quinoa has been made, including its main nutritionally disadvantageous factors.

Finally, several traditional and new forms of using quinoa have been described. Some future perspectives with regard to the germplasm collection, agronomic and agricultural practices, postharvest and industrial uses of this promising grain, which has been called the 'mother grain,' have been included.

Origin and Distribution

Quinoa (*Chenopodium quinoa*, Willd) is an indigenous crop of the Andean region of South America. It is one of the oldest crops of the American continent. Archeological findings in northern Chile showed that quinoa was used prior to 3000 BC. In Ayacucho, Peru, evidence has been shown that quinoa was cultivated before 5000 BC.

The quinoa plant was widely cultivated in the entire Andean region, in Colombia, Ecuador, Peru, Bolivia, and Chile, before the Spanish conquest. However, habits and traditional foods of natives were substituted by foreign crops such as wheat and barley. Therefore, quinoa was cultivated either in small plantations in rural areas for domestic consumption or as borders for other crops such as potatoes or maize. For that reason it was classed as food for poor people.

Quinoa is grown in the highlands from 5°N in southern Colombia to 30°S in north-eastern Argentina. At sea level, it is cultivated between 36 and 40°S, in central Chile.

In the late 1970s the main production areas of quinoa were widely described in Colombia, Chile, the Andean valleys in Peru, Altiplano in Bolivia and Peru, and the highlands of Ecuador. In Peru and Bolivia, this crop has been of great importance. It is cultivated not only for domestic consumption but also for export. Common efforts made by governments and research institutes have contributed to increase the production of quinoa in Andean countries. The cultivation of quinoa has now spread from the Andean region to several countries in the world. In the 1980s, in the USA, quinoa was cultivated in the Colorado Rockies. Nowadays, it has became a commercial crop. In the UK, in 1989 quinoa was grown commercially. In Manitoba, Canada, quinoa was grown for domestic consumption; farmers expected to have a viable crop for Canada. Quinoa was introduced to Denmark in 1984. Further improvements in the crop could make quinoa a promising new crop for European agriculture.

In order to promote the regional interchange of the excellent genetic material of quinoa among research institutes and universities, an American and European trial of quinoa was undertaken in the late 1990s. In the project, sponsored by the Food and Agriculture Organization (FAO: regional office for Latin America and the Caribbean), 25 cultivars selected from 10 different countries were tested: eight from Peru, four from Bolivia, two from Ecuador, the UK, Denmark, and Chile respectively, and one each from Argentina, Brazil, Colombia, and The Netherlands. The best cultivar from each country taking part in the experiment was used as a local control. Therefore, it is expected to be able to identify, evaluate, and select promissory genotypes of quinoa with high yield, and to provide the technological knowledge of cultivation practices and production of

quinoa suitable for the agroecological and food requirements of the producer and national consumer.

Classification

Quinoa belongs to the Chenopodiceae family, genus *Chenopodium*. Its botanical name is *Chenopodium quinoa*, Willd. Common names used in the Andean region are: quinua, kiuna, parca (Ecuador, Peru, Bolivia); supha, jopa, jupha, jiura, aara, ccallapi, vocali (Bolivia); quinhua (Chile); and suba, pasca (Colombia).

The classification of quinoa was first made from the color of the plant and fruits. Subsequently it was based on the morphological types of the plant. Despite the wide variation observed, quinoa is considered to be one single species. For practical reasons, quinoa, like maize, was classified as a race.

The most extensive collection of different races of quinoa belongs to Peru and Bolivia; each has over 2000 ecotype samples. However, other collections do exist in Argentina, Colombia, Chile, Ecuador, England, the USA, and the Former Soviet Union.

Quinoa collected in Ecuador, Peru and Bolivia has been classified into 17 races; however, more races may exist. Two types of inflorescence are described:

1. Glomerulates – small groups of flowers (glomeruli) originate from tertiary axes
2. Amaranthiformes have glomeruli originating mainly from secondary axes

According to this, the races of quinoa are classified as follows: first, glomerulate inflorescence: Cajamarca, Copacabana, Cuzco, Challapata, Cochabamba, Sicuani, Junín, Ancash, Glorieta, and Dulce; second, amaranthiforme inflorescence: Achacachi, Puno, Real, Potosi, Puca, Sucre, and Pichincha.

Quinoa grows from sea level to the Andean highlands. Thus, one of the most useful classifications is that describing five ecotypes: sea-level, valley, subtropical, salar, and altiplanic (Table 1).

Cultivation Practices

The cultivation of quinoa is related to the crop rotation seen in potatoes. This is the usual practice that improves quinoa yield and preserves soil fertility. Moreover, the biological cycle of several pathogenic microorganisms is broken down. Together with residues of fertilizer previously applied to the crop, nitrogen is sometimes applied. Cultivation of the quinoa plant requires loose soil that can retain an adequate amount of moisture.

Quinoa tolerates a wide range of acidic conditions of the soil, from pH 6.0 to 8.5. The plant is not affected from around $-1\,^{\circ}C$. However, it tolerates high temperatures not above $35\,^{\circ}C$. Quinoa is frost-resistant when the frost occurs before flowering; after that significant damage may occur. Quinoa flowers are sensitive to frost. Quinoa is drought-resistant. It is able to develop even in regions where annual rainfall is in the range 200–400 mm.

The planting season varies from August, in the Andean highlands, extending through December, and in some areas from January to March. Seeds may be spread, but weed control and mechanized practices become difficult. Quinoa is planted in rows (row spacing range 40–80 cm) when mechanized agriculture practices are used. In dry areas, seeds can be deposited at the bottom of the furrows; once

Table 1 General categories of quinoa

Ecotypes	*Location*	*Growth altitude (m)*	*Varieties*	*Characteristics*
Sea-level	South of Chile	< 500	Chilean varieties	Unbranched, long day plants, yellow, bitter seeds
Valley	Andean valley	2000–4000	Blanca de Junín, Rosada de Juní, Amarilla de Matangani, Dulce de Quitopamba, Dulce de Lazo	Big plants, branched, short growth period
Subtropical	Subtropical area of Bolivia (Yungas)	2500–3000		Plants with intense green color that turns orange as they mature; small seeds, white or orange
Salar	Bolivian Salares	3700–3800	Real	Plants adapted to salty and alkalinic soils; bitter seeds; high saponin content
Altiplanic	Area around Lake Titicaca	3500–4000	Cheweca, Kancolla, Blanca de Julí	Short plants with straight stems; short growth period; resistant to frost

planted, the seeds are covered with soil. When rain is more abundant, seeds are deposited on the top of the ridge.

Sowing density may vary according to the region. It has been reported from 0.4 to 0.6 g m^{-2} in Bolivian Altiplano, from 0.5 to 2.3 g m^{-2} in Puno, and from 0.8 to 1.4 g/m^{-2} in Ecuador. A density of 1.2 g m^{-2} has been recommended in Puno for mechanical drilling. However, sowing density could be related to the climatic conditions of the region where it is cultivated.

At physiological maturity quinoa is harvested. The grains become dry and hard, making it difficult to break them with a finger nail. Physiological maturity may be reached within 70–90 days after flowering. Depending on the variety, it takes between 5 and 8 months for a plant to mature.

Traditional harvesting is done manually. The plants are either pulled or cut with a sickle, then placed in windrows to dry completely. Threshing is performed by rubbing the panicles by hand against a stone or using threshing on the floor with sticks, animals, or vehicles followed by winnowing. Mechanical threshers have been applied using stationary threshers, some of which are adapted from those used for cereals. The yield of quinoa can be in the range of 45–500 g m^{-2} depending on the variety and growing conditions.

The most important fungus disease is downy mildew (*Peronospora farinosa*), which requires high humidity and temperature as ideal conditions to grow. However, it succeeds in low humidity and low temperature (6–10 °C) found in the north Altiplano. The main symptom is chlorotic lesions on the upper surfaces of the leaves, with a white or purple mycelium on the lower surfaces.

The disease brown stalk rot is produced by *Phoma exigua* var. *fovaeta*. Low temperature, high humidity, and wounds in the plant, such as those produced by hail, favor the growth of pathogen. Dark brown lesions with a vitreous edge (5–15 cm) on the stem and inflorescence are the main symptoms. The stem is often shrunken, the plant may become chlorotic, and progressive defoliation toward the apex may occur.

Kcona kcona (*Scrobipalpula* sp.) is probably the most serious pest of quinoa. When drought periods and high temperatures are present, insects attack intensely. Larvae destroy first leaves and inflorescence. Later on, when the plant is mature, larvae destroy the panicle and grains. Sometimes, a white powder around the base of the plant is seen as a result of grain destruction. Treatment is performed just before harvest to prevent contamination of seeds and consequently postharvest losses.

Morphology of the Plant

Quinoa is not a true cereal grain: it is a pseudocereal, which is dicotyledonous. In contrast, cereals are monocotyledonous. In spite of that, the composition of cereals and quinoa is similar as regards the main components.

Quinoa, as a plant, grows 1–3 m high (Figure 1). The seeds can germinate very fast, in a few hours after being exposed to moisture. Roots can reach depths of up to 30 cm. The stem is cylindrical (3–5 cm diameter); it can be either straight or with branches, and its color is variable. Depending on the variety it changes from white, yellow, or light brown to red. Leaves are shaped like a goose's foot. They are formed by petioles and lamina; petioles are long-channeled on their upper side. Lamina is polymorphous in the same plant; rhomboidal or triangular in the lower lamina of leaves, triangular or lanceolate in upper leaves.

The flowers are incomplete; they do not have petals. Quinoa has both hermaphrodite flowers, located at the distal end of a group, and female flowers located at the proximal end. Quinoa inflorescence is full of bunches (racemose), which emerge on the upper part and do not have branches. The arrangement of flowers in raceme is considered to be the panicle; the length of the panicle varies from 15 to 70 cm high. Flowers can be clustered in different forms – either amaranthiforme or glomerulate.

Figure 1 (see color plate 124) Quinoa plant. Courtesy of Ing. Carlos Nieto C. Pronaleg-Iniap.

Quinoa is a fruit of the Chenopodium family. The fruit of quinoa is an achene. It produces small, circular-shaped seeds, about 2 mm diameter (250–500 seeds per gram: Figure 2). It is covered by perigonium, which is the same color as the plant: white, yellow, gray, light brown, pink, black, or red. It is easily removed when it is dried. Another two layers enclose the seed. Pericarp adheres to the seed; it contains saponins which confer the bitter taste characteristic of quinoa. Episperm encloses the cylindrical seed as a thin layer. The embryo can be up to 60% of the seed weight. It forms a ring around the perisperm. The high protein content in quinoa, unlike cereals, is explained by the high proportion of embryo.

Figure 2 (see color plate 125) Quinoa seeds. Courtesy of Silvia Valencia Ch.

Chemical Composition and Nutritional Value

The diet of ancient inhabitants of the Inca empire has generated interest due to its extremely nutritious quality. Quinoa seeds contain carbohydrates, protein, fat, minerals, and vitamins. The chemical composition of quinoa depends on the variety and the environment of its cultivation.

Quinoa seeds

Protein The protein content of quinoa seeds varies from 8 to 22%, which on average is higher than common cereals such as rice, wheat, and barley (Table 2). However, it presents less than 50% of the protein found in most legumes. In quinoa, most of the protein is located in the embryo.

In pseudocereals, such as quinoa, albumins and globulins are the major protein fraction (44–77% of total protein), which is greater than prolamins (0.5–7.0%). Using a modified Osborne's method, a protein fraction of quinoa was reported – 75.1% of albumins + globulins and 19.4% of glutelins (insoluble); no prolamins were found. Thus, quinoa is considered to be a gluten-free grain because it contains very little or no prolamin. Quinoa provides a nutritional, economical, easy-to-prepare, flavorful food source which is of particular relevance for people with gluten intolerance, such as those with celiac disease.

Quinoa has a good balance of the amino acids that make up the protein. It is exceptionally high in lysine (Table 3), an amino acid which is not overly abundant in the vegetable kingdom. It is also a good complement for legumes, which are often low in methionine and cystine.

The nutritional evaluation of quinoa protein has been reported in several studies. The protein

Table 2 Chemical composition of quinoa and some cereals and legumes (g 100 g^{-1} dry wt)

	Quinoa	*Barley*	*Maize*	*Rice*	*Wheat*	*Oat*[b]	*Rye*[b]	*Bean*	*Lupine*	*Soya*
Protein	16.5	10.8	10.22	7.6	14.3	11.6	13.4	28.0	39.1	36.1
Fat	6.3	1.9	4.7	2.2	2.3	5.2	1.8	1.1	7.0	18.9
Fiber	3.8	4.4	2.3	6.4	2.8	10.4	2.6	5.0	14.6	5.6
Ash	3.8	2.2	11.7	3.4	2.2	2.9	2.1	4.7	4.0	5.3
Carbohydrates	69.0	80.7	81.1	80.4	78.4	69.8	80.1	61.2	35.3	34.1
kcal 100 g^{-1}[a]	399	383	408	372	392	372	390	367	361	451

[a]kcal 100 g^{-1}: 4 × (%protein + carbohydrates) + 9 × (%fat)

Source: [b]Kent N (1983) Chemical composition of cereals. In: *Technology of Cereals*, 3rd edn. pp. 27–48. Oxford: Pergamon Press; Koziol MJ (1992) Chemical composition and nutritional evaluation of quinoa (*Chenoodium quinoa* Willd). *Journal of Food Composition Analysis* 5: 35–68.

Table 3 Essential amino acids in quinoa and other foods (g 100 g^{-1} protein)

	Quinoa	*Maize*	*Rice*	*Wheat*	*Bean*	*Milk*	*FAO*[a]
Hiistidine	3.2	2.6	2.1	2.0	3.1	2.7	2.6
Isoleucine	4.9	4.0	4.1	4.2	4.5	10.0	4.6
Leucine	6.6	12.5	8.2	6.8	8.1	6.5	9.3
Lysine	6.0	2.9	3.8	2.6	7.0	7.9	6.6
Methionine[b]	5.3	4.0	3.6	3.7	1.2	2.5	4.2
Phenylalanine[c]	6.9	8.6	10.5	8.2	5.4	1.4	7.2
Threonine	3.7	3.8	3.8	2.8	3.9	4.7	4.3
Tryptophan	0.9	0.7	1.1	1.2	1.1	1.4	1.7
Valline	4.5	5.0	6.1	4.4	5.0	7.0	5.5

[a]As reported by Koziol (1992): see below.
[b]Methionine + cystine.
[c]Phenylalanine + tyrosine.
FAO, Food and Agriculture Organization.
Source: Koziol MJ (1992) Chemical composition and nutritional evaluation of quinoa (*Chenoodium quinoa* Willd). *Journal of Food Composition Analysis* 5: 35–68.

Table 4 Mineral composition (mg kg^{-1} dry wt) and vitamin concentration (mg 100 g^{-1} dry wt) of quinoa and some cereals

Minerals (mg kg^{-1})	*Quinoa*	*Wheat*	*Rice*	*Barley*
Ca	1487	503	69	430
Mg	2496	1694	735	1291
K	9267	5783	1183	5028
P	3837	4677	1378	3873
Fe	132	38	7	32
Cu	51	7	2	3
Zn	44	47	6	35
Vitamins (mg 100 g^{-1})				
Thiamin (B_1)	0.38	0.55	0.47	0.49
Riboflavin (B_2)	0.39	0.16	0.10	0.20
Niacin (B_3)	1.06	5.88	5.98	5.44
Ascorbic acid (C)	4.00	0	0	0
α-Tocopherol	5.37	1.15	0.18	0.35
β-Carotene	0.39	0.02	NR	0.01

NR, not reported.
Source: Koziol MJ (1992) Chemical composition and nutritional evaluation of quinoa (*Chenoodium quinoa* Willd). *Journal of Food Composition Analysis* 5: 35–68.

efficiency ratio (PER) in raw debittered quinoa was 78–93% that of casein. These figures increased when quinoa was cooked, and became 102–105% those of casein. Similar results were found when quinoa from the San Luis Valley of Colorado was used. Thus, the quality of protein in quinoa matched that of the milk protein casein. (*See* **Amino Acids**: Metabolism; **Protein**: Quality.)

Carbohydrates The major component in quinoa is carbohydrates, which varies from 67 to 74% of the dry matter. Starch is about 52–60%. The starch compound is located in the perisperm of the seed; starch can be present as simple units or as spherical aggregates. The amylose content is about 11%, which is lower than in cereals, for example, rice (17%), wheat (22%), and barley (26%).

The diameter of quinoa starch granules is in the range of 0.4–2.0 μm. Starch granules in quinoa are smaller than those reported for maize (range 1–23 μm) and for wheat (2–40 μm). Small-granule starches exhibit a higher gelatinization temperature; for quinoa the gelatinization temperature range is 57–64 °C. Other carbohydrates are found in small amounts, such as monosaccharides (2%) and disaccharides (2.3%), crude fiber 2.5–3.9%, and pentosans 2.9–3.6%. (*See* **Carbohydrates**: Classification and Properties.)

Fat, vitamins, and minerals Quinoa contains from 2 to 10% fat. Quinoa and soya oils exhibit similar fatty acid composition; thus, quinoa is a rich source of essential fatty acids such as linolenic (18:2*n*-6: 52%) and linolenic (18:3*n*-6: 4%). Quinoa is a good source of minerals. It contains more calcium, magnesium, iron, and zinc than common cereals, and the iron content is particularly high (Table 4). Polishing and washing quinoa seeds reduce the mineral content to some extent 12–15% in the concentration of iron, zinc, and potassium, 27% loss of copper and 3% loss of magnesium. Quinoa contains more riboflavin (B_2) and α-tocopherol than rice, barley, or wheat (Table 4). Quinoa seeds can be a source of vitamin E.

Nutritional disadvantages Saponins and phytic acid are the main disadvantageous factors in quinoa. Other inhibitors, trypsin inhibitor, and tannins are present in low levels.

Trypsin inhibitor in eight varieties of quinoa (range 1.36–5.04 TIU mg^{-1}) was lower than for soybean (24.5 TIU mg^{-1}). Trypsin inhibitor is a thermolabile compound which is inactivated by heat treatments.

Polyphenols (tannins) are present in small amounts (0.53 g 100 g^{-1} in whole quinoa seeds), which are reduced after scrubbing and washing with water (0.23 g 100 g^{-1}).

Saponins A bitter taste compound called saponin is located in the outer layers of quinoa seeds. This protects them from birds and insects.

Saponins are glycoside compounds which occur in two groups. According to the nature of the sapogenin moiety they are conjugated with hexoses, pentoses, or uronic acids. The sapogenins are steroids (C27) or triterpenoids (C30). Using a gas chromatography method, the sapogenins oleanolic acid, hederagenin, 30-*o*-methylspergulagenate, and phytolaccagenic acid are identifiable in sweet and bitter genotypes of quinoa.

Quinoa saponins are soluble in methanol or water. They have strong detergent properties which form very stable foam in water solutions, and reduce the superficial tension of aqueous solutions. They also show hemolytic activity and are in general toxic to cold-blooded animals which obtain oxygen directly from water. Saponins are also present in common foodstuffs such as peanuts, asparagus, garlic, onion, and spinach.

The amount of saponins present depends on the variety of quinoa. It is higher in bitter-flavor varieties than in sweet, or low-saponin, varieties. Quinoa comprises saponins from 0.1 to 5%.

The saponins of quinoa seeds are reduced to low levels after dry polishing and washing with water. These levels are apparently nontoxic to humans.

The reduction of plasma cholesterol and bile salt concentration has been attributed to the presence of certain saponins in the diet. However, some saponins can form insoluble complexes with minerals, such as zinc and iron, which make the minerals unavailable for absorption in the gut.

Phytic acid Phytic acid (myoinositol 1,2,3,4,5,6-hexakis dihydrogen phosphate) is found in most cereals and legumes at concentrations of 1–3% dry matter. It is also found in some fruits and vegetables.

In cereals, phytic acid is located in the germ. In quinoa seeds, phytic acid is located in the external layers as well as in the endosperm. It has been reported that mean (value) phytic acid concentration, in five varieties of quinoa, was 1.18 g $100\,g^{-1}$.

Studies *in vivo* and *in vitro* have shown that phytic acid interferes with mineral absorption in the gut of humans, because of its ability to form insoluble complexes with divalent minerals such as iron, zinc, and calcium. Even small amounts (0.5 μmol g^{-1}) of inositol hexaphosphate or pentaphosphate may reduce the solubility of iron.

Inositol hexaphosphate (IP_6) was mainly found in varieties from Ecuador: sweet INIAP-Tunkahuan (11.3 μmol g^{-1}) and bitter INIAP-Ingapirca (8.6 μmol g^{-1}). These figures were almost completely reduced to 0.3 μmol g^{-1} after fermentation of the germinated quinoa flour. At the same time a five- to eightfold increase in the amount of soluble iron was found.

Quinoa Leaves

The leaves of quinoa are compared to spinach as regards flavor. Quinoa leaves are cooked as a green vegetable or used raw in salad. Leaves contain (on a dry basis) carbohydrates 4.8 g 100 g^{-1}; protein 3.3 g $100\,g^{-1}$; fat 1.8 g $100\,g^{-1}$; ash 3.3 g $100\,g^{-1}$, fiber 1.9 g $100\,g^{-1}$. The protein concentration of quinoa leaves is similar to spinach; however, it contains slightly more isoleucine (5.8 g $100\,g^{-1}$ protein) and valine (7.5 g $100\,g^{-1}$ protein). The amount of fatty acids such as palmitic (16:0; 16.7%) and stearic (18:0; 1.3%) is higher than in the grains. Quinoa leaves are a rich source of vitamin A: they contain 2085 μg RE (retinol equivalents) $100\,g^{-1}$ (fresh wt), and vitamin E 2.9 mg α-TE (alpha-tocopherol) $100\,g^{-1}$. Fresh quinoa leaves contain more magnesium (83 mg $100\,g^{-1}$ fresh wt) and sodium (289 mg $100\,g^{-1}$ fresh wt) than spinach leaves. Using a gas chromatography method, sapogenins were detected in the leaves of four sweet and bitter genotypes of quinoa. Sapogenins increased as the plant matured. After 120 days of sowing, the sapogenin content on the leaves of sweet genotypes varied between 0.013 and 0.017 g $100\,g^{-1}$ (dry matter) and in bitter varieties varied between 0.02 and 0.16 g $100\,g^{-1}$ (dry matter). Hederagenin was the major sapogenin present in the leaves. (*See* **Phytic Acid**: Properties and Determination; Nutritional Impact.)

Uses

Quinoa has a natural seed coating containing saponins, which encases the seed and confers the bitter taste which is characteristic of quinoa. Saponins must be removed before consuming. External coatings are removed using either a wet or dry process. The traditional wet process used in rural areas is hand scrubbing in alkaline water. This process is used on a commercial scale; it involves abrasive dehulling to remove the external coverings, followed by a thorough washing. However, this method has economical and ecological restrictions: the water demand is high and waste water is contaminated with saponins, which are toxic to cold-blooded animals. Moreover, wet seeds must be dried immediately, or they may germinate after a few hours in wet conditions.

A dry method is also used. The seeds are scrubbed and polished in order to remove, as fine powder, external coatings. The equipment used to polish other grains has been adapted for use with quinoa seeds, with excellent results. This method presents several advantages; no water is needed, no heat treatment to dry the seeds is required, and no environmental contamination is produced. This method is best suited to sweet varieties (low saponin content) of quinoa.

A combination of dry and wet processes is applied to bitter varieties (high saponin content) of quinoa. Saponin is first removed by polishing, when most of the saponin is removed. Then, saponins that remained in the seeds are washed with water, followed by a dry

process. Any of the processes described above makes the quinoa ready for use by the consumer or further processing such as grinding.

After removing the saponins, quinoa seeds can be boiled in water (15–20 min) and served as a grain. Cooked seeds swell to about two or three times their original size. Seeds become transparent, with tiny white bands circling across the midsection.

In Chile, Ecuador, Peru, and Bolivia, the whole seeds are used in soups, salads, casseroles, chilli and stew, as well as roasted and ground in several kinds of desserts.

Quinoa can be eaten as a rice replacement, as a hot breakfast cereal, or boiled in water to make an infant cereal. The seeds can even be popped like popcorn. Seeds can be ground and used as a flour, or sprouted. The sprouts need to turn green before they can be added to salads.

Quinoa flour can be mixed with maize or wheat flour. Several levels of substitution of quinoa flour have been reported, for instance, in bread (10–13% quinoa flour), noodles, and pasta (30–40% quinoa flour) and sweet biscuits (60% quinoa flour). All yield products of excellent quality. Quinoa flour can also be drum-dried and extruded, providing products with good physical, sensorial, and nutritional qualities. Solid-state fermentation of quinoa with *Rhizopus oligosporus* Saito was performed, giving a good-quality tempeh.

In Bolivia, in 1975, the government adopted a resolution mandating that 5% of quinoa flour must be added to all pastas, crackers, and breads.

Leaves, like the seed, can also be cooked, made into a spinach-like dish, or may be served raw in a salad. Tonics, puddings, and syrups can also be prepared from the leaves. The foaming qualities of saponin are sometimes used to produce a frothier chicha.

In industry, saponins from quinoa have multiple purposes. They are used as soap for washing hair or clothes, in a compound for a fire extinguisher, or in photo processing. Dried stalks of the plant are used as fuel, or may be used in preparing bleach or dyes.

Future Perspectives

The nutritional excellence of quinoa has been known since ancient times in the Inca empire. Nowadays, quinoa has been recognized for its nutritional benefits all over the world, and for its protein, mineral, and vitamin content.

The importance that quinoa could play in nutritional behavior is being emphasized, not only in developing countries but also in the developed world. In the Andean countries, quinoa crops could play an important role in the future of their economies, giving a new export market, as well as in national subsistence. Moreover, quinoa could be a strategic crop used to complement the diet in rural/marginal regions where energy-protein malnutrition affects most of the population of the developing countries. Quinoa, as the 'mother grain,' represents an exotic and healthy rediscovery in the developed world.

Germplasm collection should continue in countries of the Andean region. Agronomic research, including plant density, potential cultivation, phenology, morphology, physiological maturity, yield, and weed control, should be performed. Further research is needed in order to study the adaptability of different cultivars to 'new homes of quinoa' in the USA and Europe. Using mechanized agriculture may facilitate mechanical harvesting of the grain, reducing post-harvest losses.

Improving methods for removing saponins without significant modification of nutritive value are encouraged. The selection of sweet genotypes with a very low saponin content in the seeds, large grain, and high yield are the main breeding goal. Sweet genotypes could be selected early in plant development in order to speed up the selection process. Further research is needed to find markers for indirect selection for sweet genotypes.

The need for intensive cultivation of quinoa should be emphasized; this could meet quality and quantity needs by the food industry. Besides, aggressive promoting campaigns should be carried out to encourage greater consumption of the grain. Finally, quinoa is been promoted as an extremely healthy food – a supergrain – of the future (gluten free). It is a food of the twenty-first century.

See also: **Amino Acids**: Properties and Occurrence; Metabolism; **Phytic Acid**: Properties and Determination; Nutritional Impact; **Protein**: Food Sources; Requirements; Quality; **Saponins**; **Starch**: Sources and Processing; Functional Properties; **Vitamins**: Overview

Further Reading

Berti DM, Serri GH, Wilckens ER, Urbina PM and Figueroa CI (1997) Determination of physiological maturity and the optimum maturity for harvesting quinoa (*Chenopodium quinoa*, Willd) in Chillan. *Agro-ciencia (Chile)* 13: 135–141.

Jacobsen S-E (1993) Quinoa (*Chenopodium quinoal*, Willd): *A Novel Crop for European Agriculture*. PhD. thesis. Denmark: Royal Veterinary and Agricultural University.

Koziol MJ (1992) Chemical composition and nutritional evaluation of quinoa (*Chenoodium quinoa* Willd). *Journal of Food Composition and Analysis* 5: 35–68.

Kuhn M (1998) Pseudocereals: a challenge for further research and product development. In: *Increasing the*

Utilisation of Sorghum, Buckwheat, Grain Amaranth and Quinoa for Improved Nutrition. International Association for Cereals Science and Technology, Cereals Conference Symposium Challenges in Speciality Crops. 16th ICC Cereals Conference. Vienna, Austria. London, UK: Institute of Food Research.

Mastebroek H, Limburg H, Gilles T and Maevin H (2000) Occurrence of sapogenins in leaves and seeds of quinoa (*Chenopodium quinoa* Willd). *Journal of Science Agriculture* 80: 152–156.

Mujica A, Jacobsen SE and Izquierdo J (1998) *Libro de Campo: Prueba Americana y Europea de Quinua (Chenopodium quinoa* Willd.). Red de Cooperación Técnica en Produción de Cultivos Alimenticios de la Oficina Regional de la FAO para América Latina y el Caribe. Puno, Peru: Universidad Nacional del Altiplano.

Nieto C and Vimos C (1992) *La Quinoa, Cosecha y Poscosecha. Algunas Experiencias en Ecuador.* Boletín Divulgativo no. 224. Quito, Ecuador: Departmento de Comunicación Social del INIAP.

Nieto C and Fisher V (1993) *La Quinua un Alimento Nuestro*. Proyecto Desarrollo Comunitario, Bélgica Ecuador. Proyecto 3 P-90-160 INIAP-CIID. Santa Catalina, Ecuador: Programa Cultivos Andinos.

Risi J and Galwey NW (1984) The chenopodium grains of the Andes. Inca crops for modern agriculture. *Advances in Applied Biology* 10: 145–216.

Ruales J (1992) *Development of an Infant Food from Quinoa* (Chenopodium quinoa, *Willd). Technological Aspects and Nutritional Consequences*. PhD. Thesis. University of Lund, Sweden.

Tapia M, Gandarillas H, Alandia S *et al*. (1979) *La Quinua y la Kañ iwa*. Cultivos Andinos. Bogotá, Colombia: Centro Internacional de Investigaciones para el Desarrollo (CIID), Instituto Interamericano de Ciencias Agrícolas (IICA).

Valencia S, Svanberg U, Sandberg A-S and Ruales J (1999) Processing of quinoa (*Chenopodium quinoa* Willd): effect on the *in vitro* iron availability and phytate hydrolysis. *International Journal of Food Science Nutrition* 5: 203–221.

Whali C (ed.) (1990) *Quinoa: Hacia su Cultivo Comercial*. casilla 17-110-6053, Quito, Ecuador: Latinreco.

R

Radiation *See* **Drying**: Theory of Air-drying; Drying Using Natural Radiation; Fluidized-bed Drying; Spray Drying; Dielectric and Osmotic Drying; Physical and Structural Changes; Chemical Changes; Hygiene; Equipment Used in Drying Foods

RADIOACTIVITY IN FOOD

J F Diehl, Karlsruhe, Germany

Principles

Radioactivity is the process of spontaneous disintegration of unstable atomic nuclei, a process associated with emission of ionizing radiation. Radioactive substances, of natural origin or man-made, are present in all foodstuffs. This article describes the radionuclides of interest and considers the internal radiation exposure of man resulting from incorporated radionuclides. The relative contribution of natural and man-made radionuclides to the resulting effective dose is discussed.

Naturally Occurring Radionuclides

Radioactive species of certain elements are continuously formed in the upper atmosphere by interaction of cosmic radiation with elements present in the atmosphere ('cosmogenic radionuclides'). Thus, carbon-14 (^{14}C) and tritium (^{3}H) are produced by interaction of cosmic-ray neutrons with nitrogen. They disintegrate, ^{14}C with a half-life of 5730 years, ^{3}H with a half-life of 12.3 years, emitting electrons or β-*radiation*. (The *half-life*, also called the physical half-life or radioactive half-life, is the time required for the disintegration of half of the atoms of a given radioactive substance.) The unit of radioactivity is the becquerel (Bq); 1 Bq corresponds to one nuclear disintegration per second. Formerly, the Curie unit (Ci) was used; 1 Ci equals 3.7×10^{10} Bq.

Other radioactive species were formed in the nuclear processes associated with the earth's origin ('primordial radionuclides'). A representative of this group is potassium-40 (^{40}K), which disintegrates with a half-life of over 10^{9} years, emitting electrons and γ *radiation*, an electromagnetic radiation. The adult human body (70 kg) contains about 13 kg of carbon with 3000 Bq of ^{14}C, 50 kg of water with 20 Bq of ^{3}H, and 140 g of potassium with 4300 of Bq ^{40}K. The major cosmogenic and primordial radionuclides and some of their properties are listed in Table 1. The α *radiation* emitted by some of the radionuclides listed in the table consists of particles containing two protons and two neutrons (i.e., helium nuclei).

With regard to the cosmogenic radionuclides and ^{40}K, data concerning activity concentrations in foods and in the human body are rather uniform in the whole world, whereas the radionuclides of the uranium and thorium series may be present at considerably higher levels (compared with Table 1), depending on local geological conditions. The population living along the Kerala coast of India, for instance, has an annual intake of 40 Bq of radium-226 (^{226}Ra) and 2000 Bq of radium-228 (^{228}Ra), and similar intakes are found in the population of the volcanic Minas Gerais area of Brazil. Another case of elevated intake relates to aboriginals of uranium-rich areas of Western Australia, where the annual intake of lead-210 (^{210}Pb) from the carcasses and offal of sheep and kangaroos can be as high as 3000 Bq. Arctic lichen accumulate polonium-210 (^{210}Po), and reindeer graze mostly on lichens in the winter, so that a further accumulation occurs in reindeer meat. Laplanders consuming reindeer meat may have a 10-fold higher intake of ^{210}Po than populations of the temperate latitudes. Refer to individual foods.

Table 1 Radionuclides naturally occurring in foods

Radionuclide	*Radiation emitted*	*Half-life*	*Concentration in foods*	*Annual intake (Bq per person)*	*Human body weighing 70 kg contains (Bq)*
Cosmogenic					
Carbon-14	β	5730 years	230 Bq per kilogram of C	20 000	3 000
Tritium	β	12.3 years	0.4 Bq per kilogram of water	500	20
Beryllium-7	γ	53 days	2 Bq per kilogram of leafy vegetables, 3 Bq per kilogram of cereals	1 000	
Sodium-22	β/γ	2.6 years		50	
Primordial					
Potassium-40	β/γ	1.3×10^9 years	30.9 Bq per gram of K	33 000	4 300
Rubidium-87	β	4.7×10^{10} years	3 Bq per kilogram of food		300
Uranium series					
Uranium-238	α/γ	4.5×10^9 years	< 2 Bq per kilogram of vegetables, mushrooms < 0.05 Bq per kilogram of other foods	5	1
Radium-226	α/γ	1600 years	50 Bq per kilogram of brazil nuts < 5 Bq per kilogram of mushrooms 2 Bq per kilogram of cockles 0.5 Bq per kilogram of eggs 0.7 Bq per kilogram of peanuts < 0.1 Bq per kilogram of other foods	19	2
Lead-210	β/γ	22.3 years	2 Bq per kilogram of shellfish < 0.1 Bq per kilogram of other foods	32	14
Polonium-210	α/γ	138.4 days	2 Bq per kilogram of mussels 10 Bq per kilogram of reindeer meat 0.7 Bq per kilogram of fish < 0.1 Bq per kilogram of other foods	55	12
Thorium series					
Radium-228	β	5.8 years	25 Bq per kilogram of brazil nuts < 0.1 Bq per kilogram of other foods	13	1
Thorium-228	α/γ	1.9 years	40 Bq per kilogram of brazil nuts < 0.1 Bq per kilogram of other foods	1.3	0.3

Source: UNSCEAR, 1982 and 1993 Reports, updated from other sources.

The occurrence of certain natural radionuclides in food plays a role in analytical methods. Fossil fuels do not contain ^{14}C or tritium, because their age is much greater than those radionuclides' half-lives. This is of interest in food control because adulterations of alcoholic beverages with synthetic ethanol, natural vinegar with acetic acid, or plant extract flavoring with synthetic flavors can be recognized by a lower ^{14}C and ^{3}H content of the synthetic product. (*See* **Adulteration of Foods**: Detection.)

In contrast to α and β radiation, which cannot penetrate matter deeply, γ radiation has a high penetrating power. The ^{40}K in a body can thus be determined by measuring γ radiation outside this body. Whole-body counters for humans and animals are available that permit measurement of ^{40}K and other γ-emitting radionuclides. Since adipose tissue is more or less free of potassium, lean body mass can be determined in this way. The principle can also be used for nondestructive estimation of the fat content of meat.

Much of the knowledge of metabolism in animals and plants is based on tracer studies with radioactively labeled substances. In radioimmunoassays (RIA), which permit extremely sensitive analyses of veterinary drugs, anabolics, pesticides, bacterial toxins, and many other substances in food, and in the RAST inhibition (radioallergosorbent test) for detection of food allergens, radioactive labeling is utilized, usually with tritium. (*See* **Analysis of Food; Immunoassays**: Radioimmunoassay and Enzyme Immunoassay; **Protein**: Synthesis and Turnover.)

Man-made Radionuclides

Consequences of Atmospheric Explosions of Nuclear Weapons

Since the first nuclear test explosion in the desert of New Mexico on 16 July 1945, large amounts of radioactive fission products have been released into the stratosphere and have eventually settled on the earth's surface as 'nuclear fallout.' Short-lived radionuclides such as ^{131}I, a β and γ emitter with a half-life of 8 days, disappeared quickly, whereas long-lived species remained in the biosphere for many years, the most important species being ^{90}Sr, a β emitter with a half-life of 29 years, and ^{137}Cs, a β and γ

emitter with a half-life of 30 years. The total release of ^{90}Sr and ^{137}Cs by atmospheric tests was about 600 and 960 PBq, respectively (P, peta $= 10^{15}$). Fallout radioactivity in food reached a peak in 1964, when an agreement between the former Soviet Union and the USA to stop atmospheric nuclear weapons tests became effective. This is demonstrated in Figure 1 for ^{90}Sr in total diet samples collected in New York City. The majority of nuclear weapons tests were carried out in the northern hemisphere. The extent of radioactive contamination was therefore considerably lower in the southern hemisphere, as shown for Argentina.

Yearly averages of specific activity of radiocesium and radiostrontium in total diet samples from Germany are shown in Figures 2 and 3, respectively. Results obtained during the 1960s and 1970s in other countries of the north 30–70° latitudes, which received the highest fallout deposits, were similar. After the peak of 1964, cesium activity (Figure 2) decreased faster than strontium activity (Figure 3). This was a consequence of binding of cesium ions to clay minerals in the soil, which makes this element rather unavailable to the roots of plants, whereas the uptake of cesium ions deposited on the leaves of plants is quite rapid. In most soils, the availability of strontium is much higher than that of cesium. Uptake of ^{90}Sr ions through plant roots continued long after deposition of fallout had decreased to below a measurable level; indeed, it continues even today, so that the present food supply still contains traces of the radiostrontium deposited in the 1950s and 1960s. The distribution of strontium and cesium in the human and animal body resembles that of calcium and potassium, respectively. In vertebrates, radiostrontium accumulates in the bones, whereas radiocesium is rather evenly distributed in muscle tissue.

Consequences of Nuclear Accidents

A severe accident occurred on 10 October 1957 at the Windscale Works, later known as Sellafield, in northwest England, when a graphite fire broke out in a uranium reactor primarily designed for production of plutonium. The accident released 740 TBq of ^{131}I, 22 TBq of ^{137}Cs, 0.07 TBq of ^{90}Sr and other radionuclides (T, tera $= 10^{12}$). Authorities restricted the consumption of milk in an area of 500 km^2 for various periods of time. Environmental monitoring of radioactivity was not well developed at that time, and the extent of contamination remained largely unknown.

The accident at the Three Mile Island reactor in Pennsylvania, USA, on 28 March 1979 released less than 1 TBq of ^{131}I and no ^{90}Sr or ^{137}Cs. Radioiodine at barely detectable levels was found in some foods produced in the area.

According to information provided by the former Soviet Government, the catastrophe of the Chernobyl reactor on 26 April 1986 released 260 PBq of ^{131}I, 38 PBq of ^{137}Cs, 18 PBq of ^{134}Cs, 8 PBq of ^{90}Sr, 5 PBq of ^{241}Pu, and large quantities of many other radionuclides during a period of 10 days. Shifting winds carried the radioactive clouds over various parts of the former Soviet Union, Scandinavia, Central Europe, the Balkan countries, and Turkey. The dominant radionuclide during the first 2 weeks was ^{131}I.

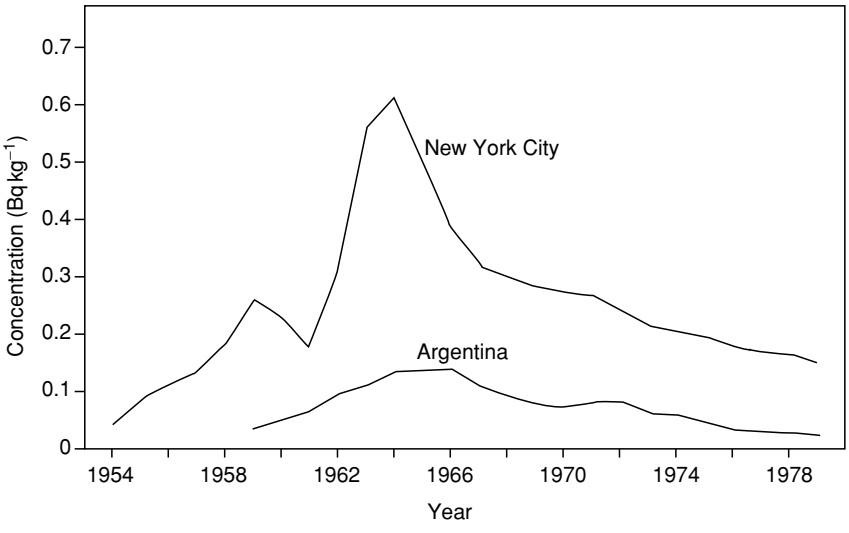

Figure 1 Strontium-90 in the total diet of Argentina and New York City, 1954–1979 (Bq kg^{-1}). Source: Ionizing Radiation: Sources and Biological Effects, United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR), 1982 Report to the General Assembly, New York, United Nations, p. 217.

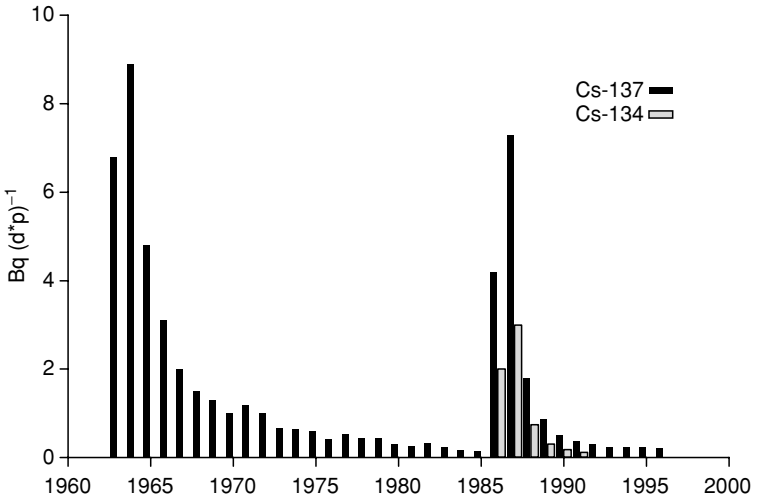

Figure 2 Annual average intake of radiocesium (Bq per day per person); daily total diet samples collected in the Federal Republic of Germany, 1963–1996. Source: Federal Research Center for Nutrition, Karlsruhe, data bank 'Leitstelle Umweltradioaktivitöt.'

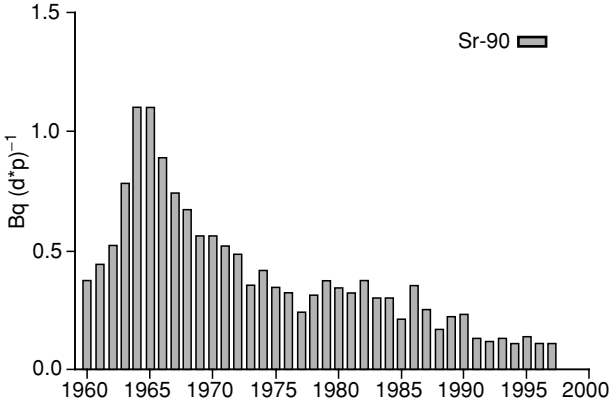

Figure 3 Annual average intake of radiostrontium (Bq per day per person); daily total diet samples collected in the Federal Republic of Germany, 1960–1997. Source: Federal Research Center for Nutrition, Karlsruhe, data bank 'Leitstelle Umweltradioaktivitöt.'

Cows' milk in some parts of Austria, Germany, and Switzerland reached activity concentrations of several hundred becquerels per liter during the first week of May. Owing to its short half-life, radioiodine disappeared quickly from the environment, and the two cesium nuclides became dominant. As shown in Figure 3, the levels of ^{90}Sr were little affected in Central Europe. Released radionuclides of low volatility, such as plutonium, were deposited mostly in the near vicinity of the accident site.

The results presented in Figure 2 allow a comparison of radiocesium in total diet after the Chernobyl accident with that observed at the time of atmospheric testing of nuclear weapons. The data of the preChernobyl years represent ^{137}Cs alone (measurable levels of ^{134}Cs were not present in the fall-out from weapons tests), whereas those of the first postChernobyl years also show ^{134}Cs, which – owing to its short half-life – had fallen to below measurable levels by 1990. It may seem surprising that Figure 2 indicates a higher radiocesium ingestion in 1987 than in 1986. This is mostly due to the fact that in 1986, radioactivity from Chernobyl played a role only from May to December. With the exception of milk and fresh fruit and vegetables, most of the food and feed consumed in 1986 came from the harvest of 1985; only when the preChernobyl supplies of grain, potatoes, and frozen and canned foods had been largely exhausted, was the peak daily intake of 11 Bq of ^{137}Cs and 5 Bq of ^{134}Cs reached in March 1987.

Activity concentrations of ^{137}Cs in various foodstuffs in the years 1986 and 1996 are presented in Table 2, with the natural ^{40}K-levels for comparison. As in Figure 2, the data are based on the situation in

Table 2 Typical activity concentrations of ^{137}Cs in various foodstuffs of the postChernobyl harvest (1986) in Germany, compared with annual (arithmetic) mean concentrations in the year 1996, and with concentrations of ^{40}K

Food	*Bq per kilogram fresh weight*		
	^{137}Cs (1986)	*^{137}Cs (1996)*	*^{40}K*
Wheat	7	0.11	130
Barley	12	0.15	140
Rye	16	0.21	140
Potatoes	3	0.17	130
Peas	7	0.04	98
Carrots	5	0.15	80
Apples	20	0.13	40
Wild blueberries	140	54	22
Wild mushrooms	(up to 2×10^4)	170	95–170
Beef	50	1.1	115
Pork	19	0.33	115
Veal	41	1.3	115
Venison (roe deer)	(up to 10^4)	50	100
Hens' eggs	5	0.1	23
Cows' milk	8	0.2	46
Fish from the Baltic Sea	4	9	90–140
Fish from the North Sea	3	0.5	90–140

Source: Federal Research Center for Nutrition, Karlsruhe, data bank 'Leitstelle Umweltradioaktivitöt.'

the Federal Republic of Germany. In all areas, the local deposition of Chernobyl fallout varied considerably, mostly depending on the amount of rainfall during the critical days ('washout effect'); the radioactivity in crops and animal products differed accordingly, samples taken in southern Germany generally showing two- to fivefold higher contamination than those taken further north. In Switzerland, milk produced in the Kanton of Ticino had 10 times higher activity concentrations of ^{131}I and radiocesium than milk from the area around Geneva. Data given in the 1986 column of Table 2 can therefore only serve as a rough approximation. In field crops and products obtained from farm animals, the ^{137}Cs concentration had returned to the preChernobyl level by 1996; venison, wild growing mushrooms and blueberries, and fish from the Baltic Sea still showed elevated levels of radioactivity, although these do not seem to be high when compared with the natural levels of ^{40}K. Accumulation of cesium in mushrooms is species-dependent; in the first years after the Chernobyl accident some species, such as *Xerocomus badius*, reached radiocesium levels of many thousands of becquerels per kilogram in some samples, whereas other species growing in the same habitat accumulated very little radioactivity. The sometimes very high concentrations of radiocesium in venison are at least partly, perhaps primarily, due to consumption of mushrooms by the animals. High radiocesium activities in wild growing blueberries and other species of forest berries are probably due to the generally more acidic pH of forest soils, which makes cesium more available to absorption by plant roots.

Radiation Exposure from Incorporated Radionuclides

Natural Exposure

Radioactivity alone, measured in becquerels, does not provide enough information to evaluate possible risks associated with the intake of radionuclides, because one becquerel of one radionuclide does not usually have the same biological effect as one becquerel of another radionuclide. It is the amount of energy deposited in living tissues that primarily determines the possible biological effects of radiation. The radiation energy imparted per gram of tissue, the type of radiation (α, β, γ), the distribution of the particular radionuclide in the body and its rate of excretion, and the different vulnerability to radiation of different tissues and organs must all be taken into account. Following procedures developed by ICRP (the International Commission for Radiological Protection), the *effective dose* can be calculated, taking all these factors into consideration. It was formerly called the 'effective equivalent dose;' it is an approximate measure of the actual ('effective') risk of genetic and somatic radiation damage to man during the remaining life time. For brevity it is simply referred to as 'dose' in the following discussion – unless the full term is required for clarification. The dose unit is called the sievert (Sv), and 1 Sv is defined as 1 J kg^{-1}.

Following the ICRP procedure, the annual dose caused by ingestion of various naturally occurring radionuclides has been estimated to be 380 μSv for adults (world-wide average), with ^{40}K making the largest single contribution (Figure 4). A comparison of Figure 4 with Table 1 shows that the radionuclides of the uranium and thorium series contribute significantly to the ingestion dose in spite of their low activity concentration in the diet. This is partly due to the higher vulnerability of biological systems to α radiation as compared to β and γ radiation, but the long biological half-life of some of these α emitters also plays a role. (*Biological half-life* is the time required for the human or animal organism to eliminate half of a given substance).

It would be difficult – and not advisable – to try to achieve a total ingestion dose of much less than 380 μSv by resorting to special diets. Potassium (and with it ^{40}K) is an essential mineral under close homeostatic control in the body. Carbon (and with it ^{14}C) is contained in all foods. The elements of the uranium and thorium series are present in most foods of plant

or animal origin, although at very low concentrations. Populations living in areas where the soils are rich in uranium/thorium minerals or individuals regularly consuming reindeer meat, brazil nuts, or shellfish can reach annual exposure levels of 1000 μSv.

Ingested radionuclides are not the only source of radiation to which man is inevitably exposed. Inhalation of the natural radionuclides ^{222}Rn and ^{220}Rn causes an annual dose of 1000 μSv (Figure 4). This must be seen together with an annual external dose of f650 μSv, of which 250 μSv come from cosmic radiation and 400 μSv from terrestrial radiation, for instance the gamma radiation emitted from ^{40}K present in the soil and in building materials. The dose from ingested natural radionuclides thus contributes 19% of the total dose of 2030 μSv from all natural sources.

Actual dose levels can vary widely above and below the average values presented in Figure 4. The contribution from cosmic radiation increases with increasing altitude; that from terrestrial radiation and from inhalation fluctuates considerably with local geological conditions. In Switzerland, for instance, the average dose from cosmic radiation is 350 μSv, from terrestrial radiation 450 μSv and from inhalation 1600 μSv; the internal dose of 380 μSv contributes 14% to the total dose of 2780 μSv from all natural sources. Recent studies from several countries tend to assign higher dose contributions to inhalation of radon, sometimes also to cosmic and terrestrial radiation, with the effect of lowering the contribution from ingestion to 12% and even less.

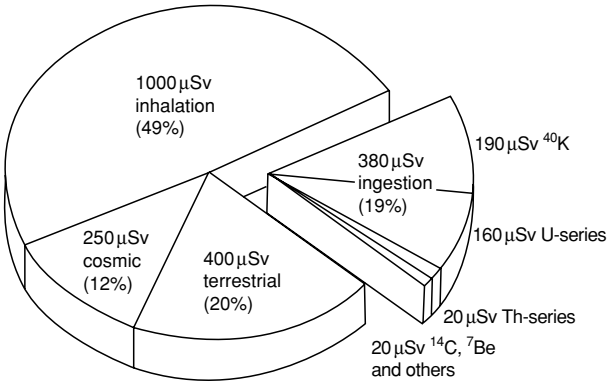

Figure 4 Effective dose from natural sources of radiation. The dose from incorporated radionuclides contributes 19% of the total of 2030 μSv per person.

Man-made Exposure

The average intake of ^{137}Cs from fallout in 1964 reached 9 Bq per day in Central Europe (Figure 2), or 3280 Bq in the whole year. For adult individuals, this meant an effective dose of 46 μSv. The daily intake of 1.1 Bq of ^{90}Sr (Figure 3), or 400 Bq per year, added 14 μSv. In all, the intake of fallout radioactivity up to 1985 in adults living in Central Europe resulted in a lifetime dose of 240 μSv from ^{137}Cs and 160 μSv from ^{90}Sr.

The effect of the Chernobyl disaster on the internal dose received by adults in the Federal Republic of Germany during the years 1986–1993 is demonstrated in Table 3. The data were calculated from the intake (Bq per year) with total diet (Figures 2 and 3), using the ICRP's conversion factors. For comparison, the external dose from radiocesium deposited on the ground is also shown. Inhalation of radionuclides occurred only during the few days when the radioactive cloud was present, and the dose resulting from inhalation was negligible, compared with that from ingestion and from external radiation. An ingestion dose of 3 μSv caused by ^{131}I during 1986 was not included in the table, because radioiodine had disappeared after a few weeks. It should be noted that a small part of the exposure caused by ^{137}Cs and a large

Table 3 Annual effective dose from ingestion and from external exposure to man-made radionuclides (adults, Germany)

Calendar year	*Ingestion dose (μSv per person)*				*External dose resulting from radiocesium (μSv per person)*
	^{137}Cs	*^{134}Cs*	*^{90}Sr*	*Sum of columns 2–4*	
1986	21.5	14.6	4.4	40.5	70
1987	37.8	22.0	3.3	63.1	30
1988	9.2	5.4	2.2	16.8	25
1989	4.5	2.2	2.8	9.5	21
1990	2.5	1.7	2.9	7.1	20
1991	1.8		1.6	3.4	19
1992	1.6		1.5	3.1	18
1993	1.2		1.6	2.8	17
Total	80.1	45.9	20.3	146.3	220

Source: Federal Research Center for Nutrition, Karlsruhe, data bank 'Leitstelle Umweltradioaktivitöt.'

part (around 90%) of that caused by ^{90}Sr originated in the nuclear test fallout of the 1960s rather than the Chernobyl accident. Except for the year 1987, the dose from external radiation was always higher than that from ingestion, and the relative importance of external exposure will grow as time progresses. Disregarding the contribution from ^{90}Sr and adding that from ^{131}I, the ingestion dose attributable to the Chernobyl accident reached about 129 μSv by 1993 and is unlikely to exceed a total of 140 to 150 μSv in the future. In the UK, the average effective dose resulting from ingestion of radionuclides from Chernobyl has been estimated to be less than 20 μSv. These calculated lifetime doses are small compared with an *annual* dose of 0.38 mSv from ingestion of natural radionuclides. They are also small compared with the ICRP *annual* dose limit of 1 mSv for the general public. The large regional variations in deposition and in activity concentrations of groups of foodstuffs observed in Germany, Switzerland, and other countries were not reflected in similarly large variations of the internal radiation dose. Even farm families nowadays buy much of their food in supermarkets, where products from many parts of Europe and from overseas are offered. As a result, ingestion dose levels in different areas of one country rarely differ by a factor of more than two.

A dissimilar situation prevails in the area surrounding the accident site. There, the highest contribution to the effective dose received during the first year came from external gamma radiation (50%), and about 25% each from inhalation and ingestion. The radionuclide contributing the highest fraction of the total dose was ^{131}I. Retrospective calculations indicate that some 270 000 living in an area of 10 000 km^2 have received a total dose of > 50 mSv, most of it during the first year. The former Soviet Government set a 'temporary dose limit' of 100 mSv (combined internal and external) for the population during the first year after the accident. In an effort to minimize the number of persons exceeding this limit, some 135 000 people were evacuated from a 30-km zone around the reactor. For various periods of time, the sale of various locally produced foodstuffs was restricted. Among the inhabitants who stayed or returned in the meantime, those living in towns could purchase uncontaminated food brought in from other regions. However, many of the people in the area live on small farms, and they and their domestic animals eat crops from their farms. Radiocesium in the soil of the family farm is now the main source of human contamination. As a consequence of high ^{131}I-exposure during the first few weeks after the accident, both from inhalation and food intake (especially milk), the incidence of thyroid cancer in children and adolescents has sharply increased in the more severely affected areas of Belarus and the Ukraine in the years since 1990, with a less pronounced increase in the bordering area of the Russian Federation. Other consequences reported in the media, such as an increased incidence of leukemia or other forms of cancer, have not been confirmed by scientific investigations. However, since the peak in the incidence of radiation-induced cancer may occur more than 10 years after the exposure, final conclusions concerning the long-term health effects attributable to radioactive contamination must await the results of ongoing epidemiological studies.

Regulation and Control

Most countries have established systems of governmental control of environmental radioactivity in the 1960s, including surveillance of food. In Europe, the EURATOM Treaty, which came into effect in 1958, obliged member countries to organize the surveillance of radioactivity in air, water, and soil, and to report the results annually to the European Commission. Milk was later included in the system, followed by certain other foodstuffs. After the Chernobyl accident, many national and regional authorities had taken steps to limit the intake of man-made radionuclides by setting maximum permitted levels of activity concentration in foods. These limits varied widely, and such variations did not inspire confidence in a wary public. The Commission of the European Communities announced a European Community Regulation on 31 May 1986, setting the following limits for radiocesium (i.e., ^{137}Cs plus ^{134}Cs): milk and milk products 370 Bq l^{-1}, baby food 370 Bq kg^{-1}, all other foods 600 Bq kg^{-1}. These limits initially applied only to foods imported into the EC, but were subsequently applied also to the internal market. Some countries outside the EC adopted more stringent standards, some going so far as to require the complete absence of ^{134}Cs and ^{137}Cs or even – forgetting the presence of natural radionuclides – 'zero radioactivity.' This caused considerable difficulties in world food trade. In an effort to overcome the confusion, the Codex Alimentarius Commission has proposed the Guideline Levels cited in Table 4. A higher dose per unit intake factor (Sv/Bq ratio), as it applies to the α-emitting nuclides of americium and plutonium, demands a lower Bq kg^{-1} limit than the lower intake factor of the γ emitters, with the bone-seeking β emitter ^{90}Sr taking a position in between. Because of the high sensitivity of the young thyroid to radioiodine, the limit for ^{131}I in infant food and milk is lower than in food destined for general consumption.

Table 4 Codex Alimentarius Guideline Levels for radionuclides in foods

Dose per unit intake factor ($Sv\,Bq^{-1}$)	*Representative radionuclides*	*Guideline level ($Bq\,kg^{-1}$)*
Foods destined for general consumption		
10^{-6}	^{241}Am, ^{239}Pu	10
10^{-7}	^{90}Sr	100
10^{-8}	^{131}I, ^{134}Cs, ^{137}Cs	1000
Milk and infant foods		
10^{-5}	^{241}Am, ^{239}Pu	1
10^{-7}	^{131}I, ^{90}Sr	100
10^{-8}	^{134}Cs, ^{137}Cs	1000

Guideline Levels are intended for use in regulating foods moving in international trade. When the Guideline Levels are exceeded, governments should decide whether, and under what circumstances, the food should be distributed within their territory or jurisdiction. The levels indicated above are designed to be applied only to radionuclides contaminating food moving in international trade following an accident and not to naturally occurring radionuclides that have always been present in the diet. The Guideline Levels remain applicable for one year following a nuclear accident.

Countermeasures

In the 1960s and again after Chernobyl, many proposals for the removal of radionuclides from the food chain or for the reduction of activity concentrations in foods have been published. Radioactive contaminants adhering to the surface of vegetables and fruit can be largely removed by washing, removal of outer leaves, or peeling. In the process of milling, much of the radioactivity in grain remains in the bran, low-extraction flour being more or less free of radioactive contamination. Dairy processing of milk produces butter that is essentially free of nuclides of cesium or strontium. If ^{131}I is present, the butter can be stored until the radioiodine has disintegrated sufficiently. Acid precipitation of curd leaves almost all radiocesium, -strontium, and -iodine in the whey, thus producing decontaminated cheese, whereas in enzymatic (rennet) precipitation, most of the radiostrontium originally present in milk is found in the curd. Many studies, some of them on a semi-industrial scale, have been carried out on the removal of cesium and strontium ions from liquid food (milk, fruit juices, tomato juice) by ion exchange or electrodialysis. Efficient decontamination can be achieved, but the treatments are relatively expensive. Some authors have described undesirable effects on sensory quality or on nutritional value. Many of these studies were probably carried out with a war situation in mind, when food would be at a premium. Should another nuclear disaster occur during peacetime, it will be more economical under most circumstances to bring in uncontaminated products from other areas than to try to decontaminate food by procedures more costly than simple washing or peeling.

Another approach is the reduction of contamination during agricultural production of food. Contamination of farm animals can be reduced by keeping them indoors during the fallout period. Root uptake of radiocesium in relatively acid soils can be minimized by liming. The fecal excretion of radionuclides can be greatly increased and their uptake into meat and milk minimized by adding adsorbents, such as bentonite or complexing compounds, such as ammonium ferric hexacyanoferrate (Prussian Blue), to the feed. The efficiency of such methods is high, and the cost is moderate.

The assessment of the radiological impact of releases of radionuclides to the terrestrial and freshwater aquatic environments requires knowledge of the transfer of various radionuclides from soil to plants, plants to animals, and food to man, and of the many factors that influence the rates of transfer. Numerous studies have been carried out with this aim since the 1960s; tables of transfer coefficients are available. Complex mathematical models, such as ECOSYS, have been developed to permit prediction of the biological consequences of a major radioactive release. Experience after the Chernobyl catastrophe has shown that these systems, when fed with solid data, allow reasonably accurate predictions to be made. In contrast, many of the population dose estimates announced soon after the accident under the pressure of circumstances later turned out to be several-fold too high.

See also: **Adulteration of Foods**: Detection; **Analysis of Food**; **Immunoassays**: Radioimmunoassay and Enzyme Immunoassay; **Protein**: Synthesis and Turnover

Further Reading

Carter MW (ed.) (1988) *Radionuclides in the Food Chain*. Berlin: Springer.

Codex Alimentarius (1995) *Levels of Radionuclides, Volume 1 – 1995, Section 6.2, CAC/GL 5-1989*. Rome: Codex Alimentarius Commission.

Diehl JF, Frindik O and Müller H (1989) Radioactivity in total diet before and after the Chernobyl reactor accident. *Zeitschrift für Lebensmittel-Untersuchung und -Forschung* 189: 36–38.

European Commission (1996) *The Radiological Consequences of the Chernobyl Accident. Proceedings of the First International Conference, 18–22 March 1996, Minsk, Belarus. Report EUR 16544. Brussels*.

IAEA (1989) *Measurement of Radionuclides in Food and the Environment. A Guidebook*. Vienna: International Atomic Energy Agency.

IAEA (1994) *Handbook of Parameter Values for the Prediction of Radionuclide Transfer in Temperate Environments*. Vienna: International Atomic Energy Agency.

McDonald P, Jackson D, Leonard DRP and McKay K (1999) An assessment of ^{210}Pb and ^{210}Po in terrestrial foodstuffs from regions of England and Wales. *Journal of Environmental Radioactivity* 43: 15–29.

Ministry of Agriculture, Fisheries and Food (1994) *Radionuclides in Foods. Food Surveillance Paper No. 43*. London: HMSO.

Müller H and Pröhl G (1992) ECOSYS-87. A dynamic model for assessing radiological consequences of nuclear accidents. *Health Physics* 64: 232–252.

Rühm W, Steiner M, Kammerer L, Hiersche L and Wirth E (1998) Estimating future radiocaesium contamination of fungi on the basis of behaviour patterns derived from past instances of contamination. *Journal of Environmental Radioactivity* 39: 129–147.

UNSCEAR (2000) *Sources and Effects of Ionizing Radiation, UN Committee on the Effects of Atomic Radiation (UNSCEAR), 2000 Report to the General Assembly*. New York: United Nations. (Also earlier reports of the UNSCEAR series.)

Woodman RFM and Nisbet AF (1999) Derivation of working levels for radionuclides in animal feedstuffs for use following a nuclear accident. *Health Physics* 77: 383–391.

Radioimmunoassay *See* **Immunoassays**: Principles; Radioimmunoassay and Enzyme Immunoassay

Raising Agents *See* **Leavening Agents**; **Yeasts**

Raisins *See* **Grapes**

Raman Spectroscopy *See* **Spectroscopy**: Overview; Infrared and Raman; Near-infrared; Fluorescence; Atomic Emission and Absorption; Nuclear Magnetic Resonance; Visible Spectroscopy and Colorimetry

RAPE SEED OIL/CANOLA

N A M Eskin and R Przybylski, University of Manitoba, Winnipeg, Manitoba, Canada

Introduction

Oilseed rape species that produce canola oil and meal are from the *Brassica* genus in the Cruciferae family. It was first cultivated in India almost 4000 years ago. Large-scale plantings of oilseed rape were first reported in Europe in the thirteenth century. The Brassica species probably evolved from the same common ancestor as wild mustard (*Sinapis*), radish (*Raphanus*), and *Eruca*.

Development of Canola

Rapeseed was first introduced into Canada from Poland in 1936. Because of oil shortages during

World War II, *Brassica rapa* seeds were tested for agronomic adaptation. At the same time, rapeseed (*B. napus*) seeds of Argentinean origin were obtained from the USA and grown under contract in the Province of Saskatchewan. During the war years, acreage increased with the first rapeseed crushing plant built in 1945. The first Canadian *B. napus* cultivar, Golden, was registered in 1954, followed by Echo, the first Canadian cultivar of *B. napus* in 1964. Oil shortages during the war suggested that Canada must develop its own domestic oil supply. Because of its agronomic adaptation, rapeseed was selected.

Early rapeseed cultivars, however, had very high levels of eicosanoic and erucic acids in the oil as well as glucosinolates in the meal. These factors were of concern if rapeseed were to be marketed for human consumption and approved by the Food and Drug Directorate of the Department of National Health. High levels of erucic acid were shown to produce fatty deposits in the heart, skeletal muscles, and adrenals of rodents as well as impair the animal's growth. Plant-breeding programs were initiated to reduce the level of erucic acid in rapeseed oil. In 1959, Liho, a rapeseed line containing low levels of erucic acid was identified. A program of backcrossing and selection was conducted to transfer the low-erucic trait into agronomically adapted cultivars. This led to the release of the first low-erucic-acid cultivar of *B. napus*, Oro, in 1968 and the first low-erucic-acid *B. rapa* cultivar, Span, in 1971. Because of health concerns related to high levels of erucic acid, over 95% of the rapeseed grown in Canada in 1974 were low-erucic-acid varieties.

Glucosinolates were also considered detrimental in rapeseed meal fed to poultry and ruminants. Its hydrolyzed products, isothiocyanates and other sulfur-containing compounds, were shown to interfere with the uptake of iodine by the thyroid gland, contribute to liver disease, and reduce growth and weight gain in animals. Consequently, breeders realized that if rapeseed meal were to be used in animal feed, the glucosinolate content had to be reduced. A Polish line with a low-glucosinolate trait, Bronowski, was identified by Dr. Keith Downey in the late 1960s. Breeding efforts to introduce this trait into low-erucic lines by Dr. Baldur Stefannson resulted in the release of the world's first low-erucic, a low-glucosinolate cultivar of *B. napus*. This was followed in 1977, by the release of the first low-erucic, low-glucosinolate cultivar of *B. rapa*, Candle, by Dr. Keith Downey. Approximately 80% of the total Canadian rapeseed acreage in 1980 consisted of the double low cultivars. (*See* **Glucosinolates**.)

The name 'canola' was registered by the Western Canadian Oilseed Crushers in 1978. The name included those cultivars containing less than 5% erucic acid in the oil and 3 mg per gram of aliphatic glucosinolates in the meal. The canola trademark was subsequently transferred to the Canola Council of Canada in 1980. In 1986, the definition of canola was amended to *B. napus* and *B. rapa* lines with less than 2% erucic acid in the oil and less than 30 μmol per gram of glucosinolates in the air-dried oil-free meal and was added to the 'generally recognized as safe' (GRAS) list of food products in the USA.

It was much more difficult to introduce the low-erucic-acid trait into the European rapeseed lines because they were primarily of the winter type. This extended the time required to produce each generation and crosses between spring low-erucic-acid rapeseed (LEAR) cultivars and winter lines resulting in undesirable segregates. Nevertheless, the development of European LEAR varieties was accomplished within 15 years. European acreage of rapeseed declined during the 1970s as a result of health concerns. In 1977, the low-erucic-acid trait was mandatory in Europe in 1977. Initially, the new LEAR cultivars produced lower yields and oil content compared with the traditional cultivars. Subsequent plant breeding overcame this problem with European production of LEAR increasing substantially by 1984.

Distribution and Adaptation

Canola is the only oilseed crop adapted to temperate regions. Its ability to germinate and grow at low temperatures permits production in cooler regions and at higher elevations compared with other oilseed crops. As a result, it can be grown as a winter crop in Europe where the winters are not too severe. The major producers of canola and rapeseed are China, European nations, India, and Canada, with Canada, China, and India accounting for over 65% of the world's total production.

Different species of *Brassica* crops are grown throughout the world, although *B. napus* and *B. rapa* are both grown in Canada. *B. napus* is higher-yielding as well as slightly higher in oil and protein content than *B. rapa*. However, *B. rapa* matures 2–3 weeks earlier and can be seeded in areas with a shorter growing season (less than 95–110 days). In addition, *B. rapa* is shorter, more shatter-resistant and frost-resistant, containing less chlorophyl and fiber than *B. napus*. The main canola species grown in western Canada are spring types, except for a small amount of high-erucic-acid rapeseed (HEAR) grown for industrial purposes. Winter cultivars of both *B. napus* and *B. rapa* are grown throughout Europe, except Sweden. In northern Sweden, spring types predominate. In the Indian

subcontinent, high-erucic-acid, high-glucosinolate spring *B. juncea* and some *B. rapa* types are grown. Winter types of *B. napus* and *B. rapa* are grown in central and southern China, whereas spring types predominate in western China. Spring *B. juncea* is also found in China. Asia still grows mainly the HEAR varieties.

Development of New Cultivars

During the last 20 years, breeding goals in Canada included increasing yields, oil and protein contents, earlier maturity, development of yellow seeded varieties, reduced green seed, and improved disease, insect, and herbicide resistance. The first triazine tolerant cultivar was released in Canada in 1984, followed by agronomically improved Tribute.

Genetic breeding is now used to modify oil quality. Using mutagenesis, a low-linolenic (C18:3)-acid cultivar, Candle, was developed. Similar techniques by researchers at Pioneer Hybrid produced a mutation that blocked the desaturation of C18:1 to C18:2, producing a high-oleic (C18:1) oil. The resulting mutation was then crossed with low-linolenic (C18:3) cultivars producing a high-oleic/low-linolenic line. A successful transformation by Calgene produced canola strains high in stearate (C18:0) and high in laurate (C12:0), with future plans for high-palmitate (C16:0) lines.

Oil Production

Canola/rapeseed oil accounts for 16% of the world supply of edible oils compared with 32 and 18% for soybean and palm and palm kernel oil, respectively. In 1997, canola oil accounted for 70.4% of all deodorized fats and oil, 72.8% of the vegetable oils, and more than 80% of the salad oils produced in Canada. The growing demand for canola products resulted in 424 000 tonnes of oil to the USA in 1996–1997, accounting for over 80% of the total oil exports. (*See* **Vegetable Oils**: Types and Properties.)

Canola Oil Extraction

On an 8% moisture basis, canola seeds contain between 38 and 44% oil. The most efficient method for extraction involves mechanical pressing or expelling in which 60% of the total oil is released. The remaining 40% is solvent-extracted, the common practice for most other oilseeds. The various stages involved are summarized below. (*See* **Vegetable Oils**: Oil Production and Processing.)

Seed Cleaning

The first step is designed to remove weed seeds, grain seeds, other foreign material as well as any canola seed fragments. This involves a combination of aspiration and screening. The foreign material removed accounts for less than 2.5% of the remaining seed and is used for animal feed.

Preconditioning

This involves preheating the seeds prior to flaking to temperature of 30–40 °C. This is achieved directly with a fluid bed type using hot air or steam or indirectly in rotary kilns equipped with steam-heated coils. This renders the seed more pliable, preventing it from shattering.

Flaking

The preheated seed is then flaked on smooth-surface rolling mills. This can be performed using a single-stage operation or as two separate stages; flakes of 0.2–0.3 mm thickness are obtained.

Cooking

Flakes are indirectly heated on steam-heated surfaces in stack cookers or in rotary cookers with steam coils. The cooking temperature is held between 75 and 100 °C. In order to adjust the moisture content of the material to 5–7% for the expeller operation, ambient air may be admitted to the cooking equipment. Besides coelescing minute oil drops into large drops, this step alters the protein and makes the oil more extractable. In addition, it also inactivates a number of enzymes, particularly myrosinase, responsible for the hydrolysis of glucosinolates to sulfur-containing compounds, lipases, and phospholipases responsible for the formation of nonhydratable phosphatides.

Oil Extraction by Prepressing

The heat-conditioned flakes are then subjected to continuous screw-presses or expellers to reduce the oil content from around 42 to 16–20%. The expellers are designed with a rotating screw shaft in a cylindrical barrel to allow oil to flow between the flat steel bars of the barrel while retaining the solid material or presscake.

Solvent Extraction

Before extracting with solvent, the presscake is cooled with ambient air to minimize hexane vaporization when it enters the extractor. The solids, now at around 80 °C, are washed with hexane in several stages. In the first stage, hexane is already high in oil content (miscella), followed by progressively leaner

miscella and ultimately pure hexane. A variety of different extractors are available, all using some type of countercurrent extraction system. The oil content in the meal is reduced to around 1%, depending on the type of equipment and throughput rate, as well as on the type of cooking, flaking, and prepress operations conducted. The resulting meal and miscella are stripped to insure the production of solvent-free meal and crude oil. The prepressed and solvent-extracted canola oils are usually blended before subjecting them to degumming.

Canola Oil Refining and Processing

Crude canola oil is converted to an edible oil using the same processing principles applied to other vegetable oils. The following steps are typical for the overall production of refined edible oils for food application.

Degumming

This process removes phosphatides from the crude oil, present at around 1.0–1.5% and measured as phosphorus, 400–600 mg kg^{-1}.

Two methods are used:

1. Degumming with water in which oil at 80 °C is intensely mixed with 2% water for 5–30 min in an agitation tank or reactor. The majority of the phosphatides are precipitated in suspension and then removed by centrifugation leaving around 0.25–0.5% phosphatides or 100–200 mg per kilogram of phosphorus in the oil. Nonhydratable phospholipids are not removed by this process.
2. Degumming with acid and water, also known as 'superdegumming,' also removes nonhydratable phosphatides. This involves intensively mixing the crude oil with 0.1–0.5% of a 50% solution of citric or malic acids for up to 15 min, depending on the mixing rate. Then, 2% of water is added, before or after cooling the mixture to 25–45 °C and is mixed in a continuous stirred reactor for 1–3 h. The precipitated phosphatide particles are removed by centrifugation. The phosphatides remaining range from 5 to 50 mg per kilogram of phosphorus, depending on the amount of nonhydratable phosphatides in the crude oil.

Refining

Degummed oils are further purified by bleaching followed by either alkali refining or physical refining process, winterization, and deodorization.

1. Bleaching. Adsorptive bleaching is performed with 1–3% of acid-activated clays to remove colored components, metals, and phospholipids. Oil is mixed with clay at 100 °C under vacuum for up to 30 min at 10–15% moisture, and then clay is removed by filtration. Chlorophyls will be reduced to below 0.05 mg kg^{-1}. Other adsorbents are rarely used for canola oil.
2. Alkali refining. This is the most commonly used process to remove phosphatides and free fatty acids (FFA). Firstly, oil is mixed with 0.05–0.1% concentrated phosphoric acid to precipitate phospholipids. Secondly, the oil is treated with 12% aqueous sodium hydroxide to neutralize FFA and phosphoric acid, and precipitate phosphatides. In both cases, intensive mixing and a short contact time at 40–90 °C are applied, followed by centrifugation, and then the oil is washed with water to remove soaps. Alkali refining reduces FFA and phosphorus to below 0.05% and 3 mg kg^{-1}, respectively, chlorophylides are reduced by up to 70%, and iron and copper are removed.
3. Physical refining. The acid-degummed oil first undergoes phosphoric acid pretreatment followed by bleaching with acid-activated clay. The bleached oil is then subjected to physical (steam) refining which removes the free fatty acids by steam distillation in the deodorizer.
4. Winterization. Salad oils are winterized to remove wax esters, which can cause sediment formation in oil. Oil is chilled to 5 °C and filtered to reduce wax content to below 50 mg kg^{-1}.
5. Deodorization. This process is performed for up to an hour at 225–260 °C under vacuum, 2–4 mm of mercury with 1–3% steam blown through the oil to remove flavor components. During this process, some tocopherols and sterols are removed from oil together with FFA. When oil is cooled to 60–90 °C, additives are added, and spurging with nitrogen gas is performed.
6. Hydrogenation. Hydrogenation modifies the melting and crystallization behavior of the fat or oil as well as rendering it more resistant to oxidative and thermal damage. This process is performed at 160–200 °C and 100–300 kPa pressure in the presence of catalyst, usually nickel. During this process, saturated and isomerized fatty acids are formed that change the physical properties of fat. Hydrogenated fats containing *trans* fatty acids have a tendency to crystalize to the β′ form producing products with a smooth and pleasing mouth feel and good spreadability. Health concerns have resulted in the reduction of *trans* isomers by searching for new catalysts or by blending. An alternative method to hydrogenation used in Europe and only to a limited degree in North America is interesterification.
7. Interesterification. This process is used to obtain fats similar to hydrogenation by applying chemical

or enzymatic rearrangement of FA in glycerides within oil and between oils. Chemical process is performed with an alkaline catalyst, sodium methoxide or sodium hydroxide, at 100–150 °C for up to an hour. The enzymatic process conditions are dictated by the enzymes applied. The main advantage of this method is that no *trans* fatty acid isomers are formed.

Properties of Canola Oil

Physical

The physical properties of canola, including density and viscosity, the main engineering parameters, are presented in Table 1.

The density and viscosity of canola oil decrease with temperature increase in a linear and exponential fashion, respectively. Most of the canola oil physical parameters are within the range found for other vegetable oils and are governed by composition of fatty acids.

Chemical

The average composition of canola, rapeseed, and soybean oils is presented in Table 2. (*See* **Fats**: Classification.)

Triacylglycerols or triglycerides are the main components of vegetable oils composed of glycerol and three fatty acids attached to it. The composition of triglycerides defines the physical properties of oils and their oxidative stability. (*See* **Fats**: Classification.)

Tocopherols are the natural and most efficient antioxidants produced by plants. There are four isomers of tocopherols produced, and canola oil contains about twice as much γ isomer than α, both isomers being present in important amounts. The tocopherol content in canola oil is at the higher end among vegetable oils. (*See* **Tocopherols**: Properties and Determination.)

Phytosterols or plant sterols have been neglected for many years, although the nutritional importance of some of these components is receiving considerable attention. Some plant sterols appear to lower plasma cholesterol by inhibiting the absorption of dietary cholesterol and reabsorption of biliary cholesterol. In fact, canola oil contains high amounts of sterols second only to corn oil.

The sensitivity of canola oil to deterioration is often related to the presence of pigments such as chlorophyls. They are capable of transferring light energy into chemical energy and can initiate formation of radicals responsible for oxidative degradation of oils. The content of chlorophyls in canola oil is dependent on the maturity of seeds used for processing, as even fully matured seeds contain a few p.p.m. of these components.

The presence of components containing sulfur in the structure is normal for oils produced from oilseeds from the Cruciferae family because of the presence of glucosinolates. In canola oil, minor fatty acids with sulfur in the structure have been detected.

The composition of fatty acids in canola oil is unique when compared with other oils. Canola oil contains the lowest amount of saturated fatty acids and is second after olive oil in the content of oleic acid. The presence of essential fatty acid, such as linoleic and linolenic, is intermediate, both acids being important nutrients in the control of cardiovascular diseases. The main fatty acid that differentiates canola oil from rapeseed oil is erucic acid, which was reduced to less than 2% in canola.

Table 1 Physical properties of canola oil

Parameter	*Value*
Relative density ($g\,cm^{-3}$; 20 °C/water at 20 °C)	0.914–0.917
Refractive index (n_D 40 °C)	1.465–1.467
Viscosity (kinematic at 20 °C; $mm^2\,s^{-1}$)	78.2
Smoke point (°C)	220–230
Flash point, open cup (°C)	275–290
Specific heat ($J\,g^{-1}$ at 20 °C)	1.910–1.916
Thermal conductivity ($W\,m\,^0K^{-1}$)	0.179–0.188

Table 2 Composition of canola, rapeseed, and soybean oils

Component	*Canola*	*Rapeseed*	*Soybean*
Triglycerides (%)	94–99	92–99	93–99
Phospholipids (%)	Up to 0.1	Up to 0.1	Up to 0.2
Free fatty acids (%)	0.2–1.2	0.5–1.8	0.2–1.0
Tocopherols (p.p.m.)	700–1200	700–1000	1700–2200
Plant sterols (p.p.m.)	8810	6900	4600
Unsaponifiables (%)	0.5–1.4	0.5–1.4	0.5–1.6
Chlorophyls (p.p.m.)	5–35	5–35	Trace
Sulfur (p.p.m.)	3–15	5–25	Nil
Iodine value	110–126	97–108	120–135
Saturated fatty acids (%)[a]	6–7	7	15
Oleic (18:1)[b] (%)	61	15	24
Linoleic (18:2) (%)	21	14	54
Linolenic (18:3) (%)	11	9	8
Erucic (22:1) (%)	0.2	45	Nil

[a]Percentage in oil.
[b]Abbreviation of fatty acids.

Nutritional Properties

Dietary fats serves as an important source of energy and essential fatty acids, and serves as a carrier for fat-soluble vitamins. In addition, it makes an important contribution to the feeling of satiety, palatability of food products, and textural properties of foods.

Recently, interest in dietary fat originated from its involvement in the etiology of cardiovascular diseases, cancer, obesity, and hypertension. It is not disputed importance of lipids but type and the amount of fats to be responsible for the development of these chronic diseases, cancer, obesity, and hypertension. Lipids, particularly the type and amount of fats, are now considered responsible for the development of these chronic diseases. Nutritional recommendations in Europe and North America stipulate a reduction of total fat intake to 30% and saturated fat intake to less than 10% of total energy. Over the past 30 years, there has been a major shift from the consumption of animal origin fats to vegetable oils. In Canada, canola oil has become the major oil consumed and processed. The shift to vegetable oils did not reduce saturated fat consumption because a significant number of these oils were used as hydrogenated fats in margarine and shortening. Where unsaturated fatty acids were hydrogenated, both saturated fatty acids and *trans* isomers were produced to obtain fat with proper technological properties such as a higher melting point and crystallization structure and texture. *Trans* isomers have a similar effect on plasma cholesterol level to saturated fatty acids. The composition of fatty acids in canola oil is compatible with nutrition recommendations directed to reduce the consumption of saturated fat and to have a ratio of omega-6 to omega-3 within 3–5. Canola oil has a ratio of 2 and has been found to be equally effective as soybean, sunflower and corn oils in reducing plasma levels of total and LDL cholesterol in normal and hyperlipidemic subjects. Diets containing canola oil also increased the content of long-chain polyunsaturated fatty acid in plasma and platelet phospholipids indicating the transformation of linoleic and linolenic acids. Consumption of canola oil alters clotting activity and thrombogenesis, and protects against cardiac arrhythmias. All these findings and the positive ratio of linoleic to linolenic acids in canola oil make this oil suitable as a source of fat for infant formulas and as part of any nutritious diet.

See also: **Brassicas**; **Functional Foods**; **Soy (Soya) Bean Oil**; **Triglycerides**: Structures and Properties; **Vegetable Oils**: Types and Properties; Oil Production and Processing; Composition and Analysis

Further Reading

Downey RK, Kramer JKG, Sauer FD and Pygden WJ (eds) (1983) *High and Low Erucic Acid Rapeseed Oil: Production, Usage, Chemistry, and Toxicological Evaluation*. New York: Academic Press.

Eskin NAM, McDonald BE, Przybylski R *et al*. (1996) Canola oil. In: Hui Y (ed.), *Edible Oil and Fat Products: Oils and Oilseeds, Bailey's Industrial Oil & Fats Products*, 5th edn., vol. 2, pp. 1–95. New York: John Wiley.

McDonald BE (1999) *Canola Oil: Nutritional Properties. Bulletin*. Winnipeg, Canada: Canola Council of Canada.

Przybylski R (1999) *Canola Oil: Physical and Properties. Bulletin*. Winnipeg, Canada: Canola Council of Canada.

Shahidi F (1993) *Rapeseed Production, Chemistry, Nutrition and Processing Technology*. New York: Van Nostrand Reinhold.

Vaisey-Genser M and Eskin NAM (1987) *Canola Oil: Properties and Performance. Publication No. 60*. Winnipeg, Canada: Canola Council of Canada.

RASPBERRIES AND RELATED FRUITS

M P Pritts, Cornell University, Ithaca, NY, USA

Background

Raspberries are a high-value crop owing to their unique flavor, exacting climatic requirements, high costs of production, and perishability. This article describes the relationships among various raspberry species and relatives, discusses the commercial raspberry industry, describes the morphology, anatomy, and chemical composition of fruit, reviews harvesting and handling techniques for fresh fruit, and presents various food uses for the raspberry.

Taxonomy and Commercial Importance

Raspberries are a diverse group of flowering plants that are closely related to blackberries. Both raspberries and blackberries belong to the genus *Rubus*. Taxonomists recognize 12 subgenera within *Rubus*, but only the raspberries (*Idaeobatus*) and blackberries (*Eubatus*) have obtained commercial significance. Species of a third group, the arctic raspberries (*Cylactis*), are harvested from the wild, and are popular in Scandanavia for making liqueurs, but these are not cultivated. The fruit of raspberries detaches from the receptacle when picked, leaving a white torus attached to the plant and a hollow fruit. This characteristic distinguishes them from

blackberries in which fruit abscission occurs behind the receptacle.

The genus *Rubus* is a member of the rose family (Rosaceae), which also includes important fruit crops such as apples, pears, cherries, peaches, plums, and strawberries. Wild raspberries occur on five continents, but are most abundant in the northern hemisphere. The temperature and subtropical region of eastern Asia is recognized as the center of origin where the most diversity exists. More than 200 species have been identified, but only a few are important commercially. These include the European red raspberry (*R. idaeus* ssp. *vulgatus* Arrhen.), the North American red raspberry (*R. idaeus* ssp. *strigosus* Michx.), and the black raspberry (*R. occidentalis L.*) of the eastern USA. Hybrids between the red and black raspberry are commonly called purple raspberries because of the fruit and cane color, and these were once given the specific rank of *R. neglectus* Peck. However, most taxonomists do not recognize hybrids as distinct species. Interspecific hybrids with blackberries have also been made; some of these are commercially important, such as Tayberry, Loganberry, Boysenberry, and Youngberry.

Several other species within the genus *Rubus* have edible fruit, or have been used by plant breeders to improve cold hardiness and resistance to diseases and insects in raspberries. Examples are *R. glaucus* Benth., a South American tetraploid black raspberry that is probably a raspberry × blackberry hybrid, *R. leucodermis* Torr. and Gr. (the western North American black raspberry), and *R. spectabilis* Pursh. (salmonberry); the Asiatic species *R. coreanus* Miq., *R. phoenicolasius* Maxim. (Japanese wineberry), *R. parvifolius* Nutt. (trailing raspberry), *R. ellipticus* Sm. (golden evergreen), *R. illecebrosus* Focke (strawberry raspberry), *R. kuntzeanus* Hemsl. (Chinese raspberry), and *R. nivens* Thumb.; the Hawaiian species *R. macraei* Gray and *R. hawaiiensis* Gray (Akala berries); and the arctic raspberries of Europe (*R. arcticus L.*) and North America (*R. stellatus* Sm.).

Commercial Industry

Raspberries were first introduced into cultivation in Europe nearly 450 years ago. By the early nineteenth century, more than 20 cultivars of red raspberry were grown in both England and the USA. English cultivars were then exported to the USA, where crosses between them and North American seedlings gave improved cultivars. Red raspberries are the most widely grown, whereas black raspberries are popular only in certain regions of the eastern USA. The progeny of black and red raspberries have purple fruits and canes; these types are popular in eastern North America. Yellow-fruited *R. idaeus*, caused by a recessive mutation, is also grown on a limited scale for specialty markets. *R. occidentalis* genotypes with yellow fruit are not grown commercially.

The three major raspberry production regions are (1) Russia, (2) Europe (mostly in Poland, Hungary, Serbia, Germany, and the UK), and (3) the Pacific Coast of North America and Canada (British Columbia, Washington, and Oregon). Much of the fruit produced in these regions is harvested mechanically and processed. In other production regions, such as eastern North America, nearly all the production is for the fresh market. Many other countries, such as Chile, New Zealand, and Australia, have a significant production as they supply the fresh market during winter in the northern hemisphere. World production is estimated at more than 400 000 tonnes.

Patterns of production in North America shifted dramatically in the early 1900s. In 1920, New York State growers (East Coast) harvested more than 4000 ha. The systemic 'mosaic virus disease' infected most of the planting stock, and the processing raspberry industry collapsed in this region. With the development of tissue culture propagation techniques, virus-indexing of nursery stock, and breeding for resistance to the virus vector, the raspberry processing industry redeveloped on the West Coast. In Britain during World War II, little attention was paid to maintaining the health of raspberry stock. It was not until the 1970s that the raspberry industry reorganized with certification programs and heat-treatment therapy for the elimination of viruses. Currently, there is interest in both Europe and North America in greenhouse raspberry production to supply local markets during winter and spring.

Varieties

Two types of bearing habits are found in commercial red raspberries. The first type is called a 'summer-bearing' habit. Canes originate from either crown buds or adventitious root buds in early spring. Canes elongate during the growing season, forming fruit buds in the axils of leaves in the autumn when temperatures decrease and day lengths shorten. The plants become dormant for winter, and then the buds on the cane grow the following spring once the chilling requirement has been fulfilled. The chilling requirement varies considerably among summer-bearing varieties, ranging from a few hundred hours to more than 1800 h. The lateral axillary buds on dormant canes contain both leaf and flower primordia. At the onset of warm weather, buds break, and flowering occurs about 6–10 weeks later. Fruiting occurs in early to late summer, depending on variety,

and then the entire cane senesces. While these second year canes (floricanes) are flowering, first-year canes (primocanes) are growing from the crown or roots. These primocanes will fruit the following year. More than 40 summer-bearing red raspberries are grown commercially, and these change with the release of new, improved varieties. Among the major varieties originating in North America are 'Boyne,' 'Canby,' 'Killarney,' 'Meeker,' 'Reveille,' 'Taylor,' 'Titan,' 'Tulameen,' and 'Willamette.' The Glen and Malling series are important varieties from the UK. The Scottish Crops Research Institute is the leading institution in the world for raspberry variety development.

The second type of growth habit is called 'fallbearing.' In some varieties, fruiting laterals will develop from the top of first year primocanes after they reach a certain height, without any chilling. If the growing season is sufficiently long, fruit can be harvested from the upper portion of these canes through autumn. The lower portion of the cane will fruit the following summer, if it is allowed to remain in the field. The major variety worldwide is 'Heritage;' other important varieties are 'Amity,' 'Autumn Bliss,' 'Polana,' and 'Autumn Britten.'

At least 13 varieties of black raspberries are grown in North America. The most successful are 'Allen,' 'Bristol,' 'Haut,' and 'Jewel.' 'Royalty' is the most popular purple raspberry. Several yellow or golden raspberry varieties are grown on a small scale for specialty markets.

Morphology and Anatomy of Fruit

Each raspberry flower contains 60–160 ovaries. Each ovary contains two ovules, but one usually aborts after differentiation. About 1 month after pollination, the ovaries ripen simultaneously to form the fruit.

Fruits ripen in three phases. Pollination is followed by a period of rapid cell division. In the second phase, cell division slows while the embryo develops, and the seed coat hardens. Finally, very rapid growth occurs as a result of cell enlargement. Each phase lasts 10–12 days. Considerably variation in fruit size exists, with a range of 1–5 g. Under certain conditions, fruit can exceed 10 g.

As with most climacteric fruit, ethylene production in the raspberry begins in the receptacle when the fruit starts to color and peaks when fully ripe. Respiration, however, decreases as ripening proceeds. Growth stresses help to fracture the middle lamella and walls of cortical cells where the fruit attaches to the receptacle. Cell-wall breakdown and disintegration are complete when the fruit is fully ripe. The detachment force is small, usually 25 g, when fruits are ripe. A 12-fold difference in detachment force can exist between underripe and overripe fruit. This differential is exploited by mechanical harvesters.

The raspberry fruit is not a true berry, but rather an aggregate of many individual drupelets. Each drupelet is anatomically analogous to a cherry with a hard endocarpic seed (pyrene) surrounded by a fleshy mesocarp and an outer exocarp. The fleshy mesocarp is composed of thin-walled, turgid parenchymatous cells. Just below the exocarp is a thin layer of oval, collenchymatous cells. Large seeds are undesirable, but there is a relationship between seed size and drupelet size. Small seeds tend to be associated with small drupelets. An average seed has a mass of about 1 mg, and comprises 4–5% of the total mass of a berry. A 100-g sample of raspberries may contain more than 4000 seeds.

The cohesion of individual drupelets results from the entanglement of unicellular, epidermal hairs that are most abundant on the sides and base of the drupelet. In some black raspberry varieties, fusion of the cuticle or wax also contributes to drupelet cohesion. Drupelets cannot normally be separated without tearing the exocarp. Considerable variation in fruit firmness and drupelet cohesion exists among commercial varieties. If the percentage of developed drupelets is low, cohesion will be poor. Firmness is related to cell diameter and tissue compactness.

Chemical and Nutritional Composition

Fruit Composition

The main constituent of the raspberry fruit is water (approximately 87%). Of the remaining solids, 9% are soluble and the rest insoluble. Pectins comprise 0.1–1.0% of the soluble fraction, but this amount decreases with ripening owing to hydrolysis.

The main sugars are glucose, fructose, and a smaller amount of sucrose. These comprise the major soluble component of the juice. A typical ripe raspberry fruit contains 5–6% sugar. Citric acid is the second largest component of the soluble fraction; raspberries contain very little malic acid, but at least 10 other acids in trace amounts. The amount of acid in the fruit increases early in development, then decreases as the fruit begins to ripen. The balance between the sugars and acids is important for consumer acceptance. A fruit with a low sugar:acid ratio will taste tart; one with a high ratio will taste bland. A typical pH of a ripe raspberry fruit is 3.0–3.5; the ratio between sugars and acids (w/w) is approximately 1.0. Fruits grown under warm, dry summers (daytime temperatures near 25 °C) are sweeter, less acid, more aromatic, and more highly colored. Hot weather (temperatures greater than 30 °C) reduce the

aroma of the fruit, and wet weather reduces the sugar content.

A large number of volatile compounds are found in raspberry fruit, but most are present at less than 10 p.p.m. and are below the threshold of detection by the human nose. Compounds include alcohols, acids, esters, carbonyls, ketones, and other hydrocarbons (naphthalene and related compounds). A particular ketone, 1-(*p*-hydroxyphenyl)-3-butanone has an odor very characteristic of raspberry.

Considerable changes occur in phenolic compounds as fruits mature. Raspberry juice from ripe fruit contains 0.10–0.14% polyphenols, mainly as catechin and chlorogenic, ferulic, and neochlorogenic acids.

Raspberry fruits contain small amounts of vitamins; only vitamin C is present at a significant level. Amino acids include alanine, serine, asparagine, glutamic acid, glutamine, (gamma)aminobutyric acid, valine, leucine, and aspartic acid. The high levels of antioxidants and ellagic acid in raspberries have been highlighted in the popular press as public interest in cancer prevention increases. Table 1 presents data on the vitamin and nutrient content of red raspberries.

Table 1 Reported values or ranges of nutrient content in 100 g of fresh raspberries

Nutrient	*Amount*
Water (g)	84–87
Food energy (kcal)	31–49
Protein (g)	0.42–1.40
Fat (g)	0.20–0.55
Carbohydrate (g)	5.8–11.6
Fiber (g)	3.0–7.4
Ash (g)	0.40–0.51
Minerals (mg)	
Calcium	22–50
Iron	0.57–1.20
Magnesium	18–30
Phosphorus	12–50
Potassium	130–221
Sodium	0–2.5
Zinc	0.46
Copper	0.07–0.21
Manganese	1.01
Sulfur	17.3
Chlorine	22.3–22.8
Boron	71–125
Vitamins (mg)	
Carotene	0.05–0.08
Thiamin	0.01–0.03
Riboflavin	0.03–0.10
Pantothenic acid	0.24–0.30
Nicotinamide	0.20–1.00
Vitamin B_6	0.06–0.90
Vitamin C	13–38
Tocopherols	0.3–4.5

Data from: Jennings DL (1988) *Raspberries and Blackberries – Their Breeding, Diseases and Growth*. New York: Academic Press, Green A (1971) Soft fruits. In: Hulme AC (ed.) *The Biochemistry of Fruits and their Products*, vol. 2, pp. 375–410. New York: Academic Press and Souci *et al.* (1986/7) Wissenschaftliche Verlagsgedellschaft, Stuttgart, Germany.

Fruit Color

The color of raspberry fruit is imparted by anthocyanins. The anthocyanin molecule in raspberry consists of cyanidin and pelargonidin with glucose attached at the 3 position. Additional glucose, rhamnose, or xylose sugars may be present in various combinations to give diglycosides or triglycosides. Fruits with a preponderance of pelargonidin glycosides have an orange–red color, as opposed to a deep red color with cyanidin glycosides. The type and concentration of anthocyanins in the fruit are controlled by a few major genes. Fruits with a yellow color are produced when one or more of these genes suppress anthocyanin production. Molecules that are complexed with the anthocyanin also affect color development, and the pH of the fruit has a small effect. During storage, raspberry fruits increase in anthocyanin, darken, and become more blue, owing in part to an increase in cytoplasmic pH.

Handling and Storage of Fresh Produce

Harvest Considerations

Raspberries have one of the highest respiration rates of any fruit. This, coupled with their thin skin and sugary interior, makes them among the most perishable of all fruits. With any given variety, fruit on a bush will ripen over a period of several weeks. Harvesting the same planting frequently (once every 2 days) is critical. Fruit harvested before it is fully ripe will have a much longer shelf-life than fully ripe or overripe fruit but will be lower in sugar and anthocyanins. The optimum stage of maturity for the raspberry occurs when the berry first becomes completely red but before any darker hues develop.

Fruit quality for fresh market raspberries usually declines as the season progresses. Marketing channels must be open before the first berries ripen, as these are likely to be of the highest quality and largest size for the season. Berries should not be touched before harvest, and only undamaged berries with a good appearance should be placed in the pack. The magnitude of injury caused by human pickers can be so great as to mask any other causes of deterioration. Overripe or damaged berries should be harvested and discarded, because they are susceptible to molds.

Botrytis is the most common pathogen of raspberry fruit. Once the mold growing on overripe berries sporulates, large amounts of inoculum will be present to infect other ripening fruit. Overripe berries also attract ants, wasps, and other pests.

Containers holding approximately 150 g of raspberries are typically used. Wide, shallow containers are preferable to deep containers; each should have no more than four layers of raspberries to prevent crushing. Many different types of containers are available, but among the most popular today are plastic clamshells.

Postharvest Considerations

The objective of postharvest handling of raspberries is to slow down the respiration and transpiration rate of fruit. Respiration and transpiration result in shrinkage and reduced sweetness. Conditions that slow down the respiration process are low temperatures, high carbon dioxide levels, and low oxygen in the storage chamber. Transpiration is slowed down by high humidity.

Temperature is the easiest environmental variable to modify for extended storage of raspberries. A 5 °C reduction in temperature reduces the respiration rate by approximately 50%: at 0 °C, the respiration rate is 24 mg of carbon dioxide per kilogram of raspberry per hour; at 5 °C, the rate is 55; at 10 °C, the rate is 92; at 15 °C, the rate is 135; and at 20 °C, the rate is 200. Rapid movement of cold, humid air through the berries is essential during the first few hours after harvest to slow down respiration and prevent deterioration. Large growers may have a precooling facility, specifically designed for removing field heat, and an additional storage cooler. For every hour's delay in cooling, the shelf-life can be reduced by 1 day. Growers take advantage of natural night cooling by harvesting fruit as early in the morning as possible. Mechanical harvesting is often done at night.

Once the berries are cool, containers are wrapped in plastic to prevent water loss from the fruit and condensation on the berries when they are removed from the cooler. The plastic is not removed until the temperature of the berries approaches ambient temperature. The storage room can be maintained as low as − 1 °C. Berries will not freeze at or above this temperature, because the sugars in the fruit depress the freezing point.

When the temperature is lowered, the amount of moisture in the air is reduced. For raspberries, it is critical to maintain a humid atmosphere (90–95%) simultaneously with a low temperature to prevent water loss from the fruit. Special cooling units designed to maintain a high humidity are required for raspberries. At 25 °C and 30% relative humidity, fruits lose water 35 times faster than at 0 °C and 90% relative humidity.

A high carbon dioxide (15–20%) and/or low oxygen (5–10%) atmosphere will reduce respiration and mold growth. *Botrytis*, *Rhizopus*, *Alternaria*, *Penicillium*, and *Cladosporium* can cause postharvest fruit rots, depending on the storage temperature and carbon dioxide level. Modified atmospheres are used frequently when raspberries are transported long distances. Special semipermeable wraps are sometimes used by shippers to create a modified atmosphere (low oxygen, high carbon dioxide) within individual containers. Low-oxygen atmospheres will extend the shelf-life of raspberries, but bad-tasting aldehydes and alcohols can accumulate in the raspberry fruit when oxygen is limited for an extended period of time. Off-flavors and browning also develop when raspberries are held under elevated carbon dioxide levels for an extended period. Raspberries are considered to be at moderate risk of physiological injury from high-carbon dioxide or low-oxygen atmospheres.

The loss of raspberries from harvest to the consumer's table has been estimated at more than 40%. A 14% loss occurs from farmer to wholesaler, a 6% loss occurs from wholesaler to retailer, and 22% is lost between the retailer and consumer. Most of these losses result from poor handling of berries after harvest.

Fruit Uses

Raspberries have been eaten fresh for thousands of years. Medicinal uses of raspberries are frequently found in the literature, with references to raspberry leaf tea dating to the sixteenth century. In the early 1900s, black raspberry juice was extracted, concentrated, and used as an edible dye for foodstuffs, such as meat. Raspberries were also dehydrated for long-distance transport. Today, raspberry fruits are either harvested by hand and eaten fresh, or machine-harvested and processed. Approximately 27 000 tonnes are processed each year. The major products of processing are IQF, block frozen, purée frozen (8–15° Brix), juice (about 9° Brix), concentrate, canned, aseptic packs, and preserves. These products are then packaged and sold directly to the consumer, or reprocessed into jam, jelly, dessert topping, pie filling, icecream, yogurt, and other desserts. Raspberry juice is usually blended with apple, pear, or grape juice, because the flavor is too intense for direct consumption. Recently, the demand for fruit wines has increased, and raspberries make one of the better wines. Some wineries add raspberry juice to grape wine to obtain a less expensive raspberry-flavored wine. Raspberry beer is also made at breweries and meaderies. The popularity of raspberries continues to grow as many raspberry-containing products are now on supermarket shelves.

See also: **Fruits of Temperate Climates**: Factors Affecting Quality; **Spoilage**: Molds in Spoilage

Further Reading

Crandall PC and Daubeny HA (1990) Raspberry management. In: Galletta GJ and Himeirick DG (eds) *Small Fruit Crop Management*, pp. 157–213. Englewood Cliffs, NJ: Prentice Hall.

Green A (1971) Soft fruits. In: Hulme AC (ed.) *The Biochemistry of Fruits and their Products*, vol. 2, pp. 375–410. New York: Academic Press.

Jennings DL (1988) *Raspberries and Blackberries – Their Breeding, Diseases and Growth*. New York: Academic Press.

Kader AA (1997) A summary of CA requirements and recommendations for fruits other than apples and pears. In: Kader AA (ed.) *Proceedings of the Controlled Atmosphere Research Conference* 3: 1–34. Davis, CA.

Ourecky DK (1975) Brambles. In: Janick J and Moore J (eds) *Advances in Fruit Breeding*, pp. 98–129. West Layfayette, IN: Purdue University Press.

Pritts M and Handley D (1989) *Bramble Production Guide. Northeast Regional Agricultural Engineering Service Bulletin 35*. New York: Cornell University.

RECOMBINED AND FILLED MILKS

W Kneifel, University of Agricultural Sciences, Vienna, Austria

General Considerations

The manufacture of recombined milk and milk products is a technology that has emerged over the past three decades. During this time, there has been a worldwide development in the establishment of dairy recombination plants, mainly in countries that have a limited indigenous dairy industry, or in regions suffering from seasonal deficiencies in milk supply. Recombined dairy products are therefore ensuring a year-round supply for many households in developing countries, and also enabling the extension of the local dairy industries. However, for the industrialized countries, milk recombination offers the opportunity to transfer raw materials (milk powders, anhydrous milk fat, etc.) from surplus production areas to deficiency areas, in order to compensate for the abovementioned problems and to open up new markets. Hence, it should be taken into consideration that world stock levels and prices of butter and milk powder are strongly affecting the development of recombination industry as well as the trend of prices of recombined dairy products.

Definitions

In principle, most types of dairy products can also be produced in a reconstituted or recombined form. In the *Code of Principles Concerning Milk and Milk Products*, the Codex Alimentarius Commission of the Food and Agriculture Organization and World Health Organization has considered the relation between conventional and recombined milk products: it was determined that all products must be based on the use of fresh milk, reconstituted milk, or recombined milk. Recombined products can be subcategorized into seven main groups: milk, cream, evaporated milk, sweetened condensed milk, cultured milk products, butter, and cheese. According to definition, a *reconstituted milk product* is the product resulting from the addition of water to the dried or condensed form of the product in the amount necessary to reestablish the specified water: solids ratio. A *recombined milk product* is the product resulting from the combining of milk fat and milk solids-nonfat (SNF) in one or more of their various forms, with or without water. This recombination must be made in order to reestablish the product's specified fat: SNF ratio and solids: water ratio. Among these two fundamental categories of special milk products, the following designations are also commonly distinguished: *toned milk* is a product containing locally produced milk enriched with reconstituted skim-milk solids in order to obtain the average milk composition (e.g., high-fat milk, such as buffalo milk, is adapted to cows' milk composition, or local cows' milk showing adverse variations in composition is adjusted to normal composition). *Blended milk* is a product in which recombined milk is used in mixture with local fresh milk in order to restore the fluctuations in local milk production. *Filled milk* is a recombined product in which milk fat is partly or completely replaced by locally available vegetable oils.

Ingredients Used for Recombination

The quality of recombined dairy products is directly influenced by the quality of the ingredients used. Nutritional, technological, and sensory specifications

of raw materials must be considered in order that the recombined products resemble the original products. Typical compositions of some selected products are shown in Table 1.

Dried Milk and Milk Protein Powders

Skim-milk powder is generally the most usual component for providing milk proteins in the manufacture of recombined dairy products. *Whole-milk powder* (prevailingly 'instantized' types) can be used as a source for both SNF and milk fat, but oxidative changes in the fat phase, leading to sensory deterioration, are limiting its application. Certain proportions of *buttermilk powder* can be used more or less advantageously to enhance typical flavor characteristics. In addition, different types of *whey protein powders* and *caseinates*, and, recently, *retentate powders* (for the manufacture of cheese) made from ultrafiltered milk are coprocessed because of attributing certain functional properties to the final products.

Table 1 Typical composition of recombined milk products

Ingredients	*Product formulations (kg t^{-1})*			
	Recombined milk		*Recombined cream*	*Recombined sweetened condensed milk*
Skim-milk powder	80.5	85	30	229
Buttermilk powder	9.7		45	
Anhydrous milk fat	33.3	35	190	88.5
Water	876.4	880	733.2	250
Carrageenan			0.3	
Emulsifiers			1.5	
Sucrose				432
Seed lactose				0.5

(*See* **Casein and Caseinates**: Uses in the Food Industry; **Whey and Whey Powders**: Production and Uses.)

Reconstitutability, heat stability, and viscosity properties are the most important parameters, limiting as well as enabling the utilization of milk powders in recombined products. These properties are determined mainly by the preheat treatment applied during milk powder production as well as by seasonal variations in raw milk composition. Because a number of milk proteins are heat-labile, the extent to which they are denatured reflects the heat treatment applied during milk powder manufacture. This effect is also used as an indicator for its suitability to be applied in a diverse array of recombined products. The amount of undenatured whey protein (as characterized by the 'whey protein index' or the 'heat number' analysis) is usually taken as a measure for milk powder classification, but for certain purposes, dried milk is also classified, based on other analytical indicators (see Table 2). Besides these parameters, moisture and fat content, solubility index, bulk density, flowability, wettability, scorched particles, rennetability, emulsification properties, titratable acidity, sensory aspects, and bacteriological requirements are also included in various milk powder specifications. Most of these properties can be influenced to a certain extent by applying a defined powder technology. However, some particular factors are additionally improved by using certain additives to the product. (*See* **Whey and Whey Powders**: Protein Concentrates and Fractions.)

Milk powder is usually received in 25-kg multiwall bags, with an inner layer consisting of polyethylene. Alternatively, large bulk bins with a capacity of 200–1000 kg are used.

Table 2 Heat classification and applicability (*) of skim-milk powder to recombined dairy products

	Categories of skim-milk powder				
	Extra-low heat	*Low heat*	*Medium heat*	*Medium-high heat*	*High-heat*
Classification parameters					
Whey protein index (ADMI[a], IDF)	nd	≥ 6.0	5.9–4.5	4.4–1.5	≤ 1.4
Heat number (IDF)	nd	≤ 80	80.1–83.0	83.1–88.0	≥ 88.1
Cysteine number	26–33	34–41	42–49	50–60	≥ 60
Recombined products					
Pasteurized milk		*	*	*	
UHT milk			*	*	
Sterilized milk			*	*	
Evaporated milk					*[b]
Sweetened condensed milk		*	*		
Yogurt		*	*		
Cheese	*	*			
Butter					*
Icecream		*	*	*	*

[a]American Dairy Products Institute (formerly American Dry Milk Institute), Chicago, Illinois, USA.
[b]Specially manufactured, 'heat-stable,' high-heat powder is used.
nd, no data available.

Fats and Oils

Among the traditional dairy-based fat sources – *anhydrous milk fat* (AMF), *unsalted butter*, *anhydrous butter oil*, *butter oil* – AMF has been used most because of its better keeping quality. In order to maintain a good storage stability (6–12 months) of AMF, suitable packaging in steel drums under an inert atmosphere must be applied. Recently, fresh frozen milk fat and soft-fat fraction (produced by fractional crystallization from milk fat) have also been successfully employed in recombined products. Because of their pronounced organoleptic properties, the use of vegetable oils for the manufacture of filled dairy products is restricted to a few oil sources. Only *palm oil*, *coconut oil*, *soya bean oil*, and, to a certain extent, *maize oil* have proved to be suitable. For filled milks, highly refined, bleached vegetable oils are suggested. Some of the oils contain considerable amounts of natural antioxidants. However, all fat components have to be regarded as very sensitive to oxidative and lipolytic rancidity, leading to the formation of objectionable off-flavors. Specifications regarding fats or oils usually contain parameters such as fat and moisture content, fatty acid composition, maximum concentration of free fatty acids, rancidity, peroxide value, trace elements levels, and organoleptic properties. (*See* **Butter**: The Product and its Manufacture.)

Water

Water is the major constituent in many of the recombined dairy products. It is therefore essential that the water used fulfils several standards for *drinking water*, as laid down by the World Health Organization. Besides sensory properties, physical, chemical, (micro)biological, and radiological factors are important. It has been shown that some physicochemical parameters strongly coinfluence deposit formation during thermal processing of recombined products. Today, many countries impose established limits and specifications for drinking water. According to the recommendations of the International Dairy Federation (IDF), water to be used for the recombination of dairy products should not exceed the following maximum salt concentrations: total hardness, 100 μg of calcium carbonate ($CaCo_3$) per gram of water; chloride, 100 $\mu g\,g^{-1}$; sulfate, 100 $\mu g\,g^{-1}$; nitrate, 45 $\mu g\,g^{-1}$.

Additives

Additives can fulfil several targets in recombined products. Owing to certain losses during production and storage, the raw materials used for recombination may contain a lower vitamin content than conventionally produced milk products. In this case, the final vitamin content of the recombined product would be reduced. Fortification with *fat-* and *water-soluble vitamins*, either as single components or in readily prepared mixtures, is therefore applied frequently or is even obligatory. In general, it was noted at the last IDF seminar on milk recombination, in 1988, that recombined dairy products are of a high nutritive value and comparable to conventional dairy products. (*See* **Vitamins**: Overview.)

Antioxidants exhibit a very complex function in preventing the generation of oxidized flavor of the fat components in the product, Since some particular vitamins (tocopherols and ascorbic acid) are capable of inhibiting free radical chain reactions in the lipid phase, they are commonly added as antioxidants. (*See* **Antioxidants**: Natural Antioxidants.)

Stabilizers and *emulsifiers* (carrageenan, alginates, gelatin, lecithin, and glycerolmonostearate) are used to stabilize the fat phase and also to improve the texture and the mouth-feel of the products. (*See* **Emulsifiers**: Uses in Processed Foods; **Stabilizers**: Types and Function.)

Different *salts* (sodium citrates, phosphates, calcium salts, and sodium chloride) are used to assist the reconstitution properties, heat stability, and coagulation, and to act as stabilizing reagents.

Besides *rennet*, the addition of *glucono-δ-lactone* helps to improve the coagulation properties of milk in the manufacture of recombined cheese. Pregastric *lipases* are applied to enhance the development of the desired cheese flavor.

Sugar, as needed for recombined sweetened condensed milk, must fulfil high-quality requirements: sucrose should be of a refined 'A1 granular quality' and of a 'water white' color; the lactose seed material must be finely ground with a maximum particle size of 10 μm.

Depending on the wishes of the consumers, various natural or synthetic *flavors* and *colors* of different origin are also added. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments; Properties and Determinants of Synthetic Pigments; **Flavor (Flavour) Compounds**: Production Methods.)

Recombination Technology

The processes for recombination of dairy products range from rather simple (e.g., recombined pasteurized milk) to more advanced technologies, such as those applied in the manufacture of defined cheese varieties or sweetened condensed milk. Milk recombination plants are normally built with capacities of up to 15 000 l h^{-1}. Parallel lines are installed to achieve higher performances. In general, the

recombination technique has been derived from the conventional processes of dairying, which were further adapted and supplemented by procedures such as weighing, dissolving, mixing, and filtration. 'Tailor-made' recombination systems are frequently built in order to meet the individual requirements. The raw materials are processed using discontinuous weighing, and semicontinuous or continuous recombination systems. Typical equipment for batchtype milk recombination basically consists of the following parts: a jacketed *mixing vat*, which is equipped with a calibrated sight glass or mounted on load cells, so that its contents can be exactly measured, and is thermostatically controlled by circulation through a heat exchanger; a *powder–liquid blender* or a *dumping funnel* with a circulating pump for adding and dispersing powder; equipment for *melting milk fat*; a *clarifier* or *duplex filters*; a two-stage *homogenizer*; equipment for *pasteurization*; a *packaging* line; equipment for *refrigerated storage* of the products.

Modern, continuously working equipment for milk recombination with automatic fat dosage, as designed by Alfa Laval (Lund, Sweden), is shown in Figure 1. It consists of *mixing tanks* (6), in which warm water at 40–50 °C is metered. Skim-milk powder is supplied to the *funnel* (4) and transported automatically into the mixing tank by water flowing through a bypass line. The agitator in the mixing tank is started at the same time as the *circulation pump* (5). After completely dissolving the milk powder, AMF is added via a *measuring tank* (3) from the *fat storage* and *melting tanks* (1). Then, the whole mixture is agitated until complete fat dispersion is obtained. The process is repeated in the next tank when all ingredients have been mixed in and added to one tank. The mixture is continuously drawn from the full mixing tank by a *pump* (7), which propels the mixture through *duplex filters* (8), which remove foreign particles. After preheating in the *heat exchanger* (11), the product is pumped through a *homogenizer* (10), where the fat dispersion is completed. A *vacuum deaerator vessel*

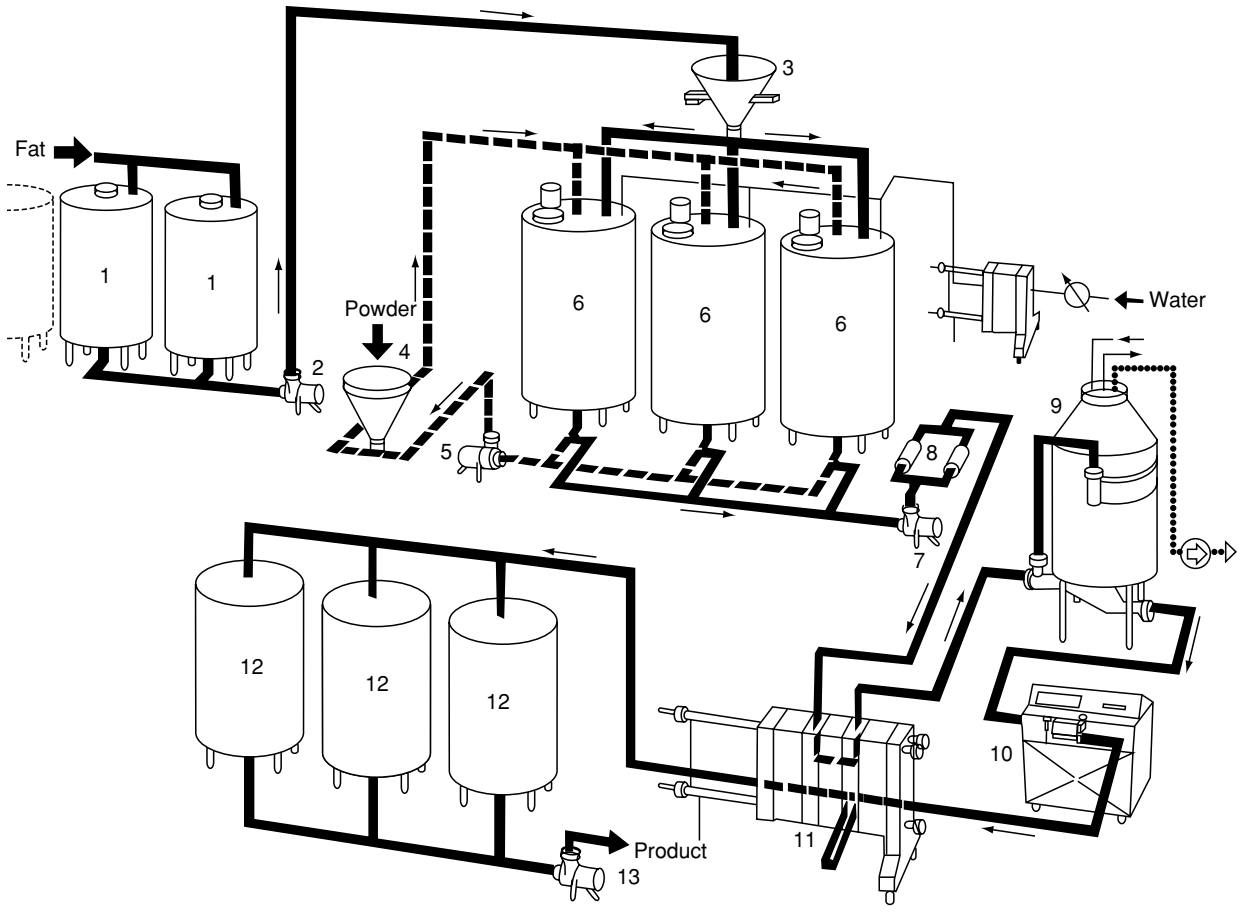

Figure 1 Plant for the manufacture of recombined milk with fat supply to mixing tanks. 1, Tanks for fat; 2, pumps; 3, weighing tank; 4, powder-blending funnel; 5, circulation pump; 6, mixing tanks; 7, pump; 8, filters; 9, vacuum deaerator; 10, homogenizer; 11, heat exchanger; 12, storage tanks; 13, pump. For an explanation, see text. (By courtesy of Alfa Laval Food Engineering AB, Lund, Sweden.)

(9) installed in the line before the homogenizer can be used to eliminate the air occluded in the powder particles as well as that picked up during powder mixing. The homogenized product is then cooled in the *plate exchanger* (11) before storage in buffer tanks or packaging.

Besides high-capacity recombination systems such as those shown in Figure 1, other automatic processors can be used advantageously. The Primodan (Holbaek, Denmark) recombiner, for example, is equipped with an adjustable steam-heated butter melter and is capable of milk production capacity of up to 25 000 $l\,h^{-1}$. Another fully automated and energy-saving milk powder dissolver is available from Jongia (Leeuwarden, The Netherlands).

Mixing and Fat Melting

In smaller plants, the powder bags are mostly handled manually; they are emptied directly into the mixer. In larger plants, automatic emptying systems are installed, and the powder is transported pneumatically into storage silos, from which it is mechanically conveyed to the mixing system via a weighing hopper or a screw-feeder.

Milk powder is usually dissolved at high stirring or mixing rates. Simple mixing devices consist of a centrifugal pump attached to a tank filled with the defined amount of warm water, with a powder funnel mounted between. The pump is connected to the jacketed tank with a pipe enabling sufficient circulation. The milk powder is transported into the tank by the vacuum thereby obtained.

In general, the choice of mixer used for dissolving the powder and dispersing the fat depends on the viscosity and/or the dry matter content of the product mix. For recombined products of low viscosity (pasteurized, sterilized and ultra-heat-treated (UHT) fluid milk, stirred yogurts), the following systems are used: turbine or propeller mixers (e.g., Alfa Laval; APV Rosista, Horsens, Denmark; Scanima, Aalborg, Denmark) driven by an electric motor; Venturi jet mixers (e.g., Danish Turnkey Dairies, Aarhus, Denmark); static mixers (Primodan, Jongia); high-speed blenders (TMP mixer, APV Rosista; Tri-Blender, Tri-Clover, Kenosha, WI, USA). The best dispersion is achieved when applying a continuous recirculating motion in the mixing tanks. High-speed blenders are also suited for the preparation of high-viscosity recombined products (e.g., evaporated and sweetened condensed milk, icecream). Another category of mixers are the disc mill mixers, which are also mainly used in the production of recombined sweetened condensed milk. The so-called Disko-Eulgiermix (Figure 2a), capable of simultaneously homogenizing whilst recirculating, and of working under vacuum conditions,

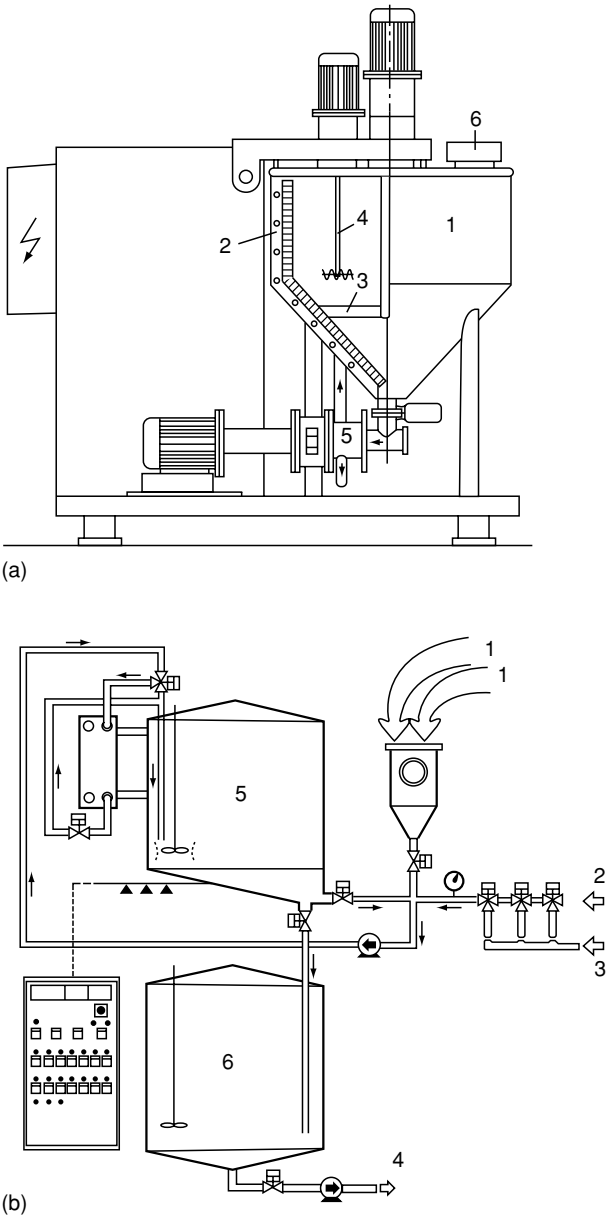

Figure 2 Examples of special mixing units. (a) Disko-Emulgiermix (courtesy of Balik GmbH, Vienna, Austria): 1, vacuum mixing tank; 2, thermostatically controlled jacket; 3, anchor-type stirrer; 4, disk mill dissolver; 5, flow-through homogenizer; 6, milk powder intake. (b) Weighmatic (courtesy of Alfa Laval Food Engineering AB, Lund, Sweden): 1, powder intake; 2, liquid inlet; 3, cleaning-in-place inlet; 4, product outlet; 5, weighing tank; 6, buffer tank. Reproduced from Recombined and Filled Milks, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

was introduced recently (Balik, Vienna). It is suitable for preparing products of different viscosity ranges. By applying a defined vacuum, foaming and picking up of air can be avoided during mixing. Another example for a sophisticated mixing unit is the

Weighmatic (Figure 2b) two-tank system (Alfa Laval), which has been specially designed for the manufacture of recombined icecream and confectionery. The whole mixing process is performed with high accuracy under computer-controlled conditions.

Different systems are used for melting the milk fat. If the fat is packed in cans, the melting process is carried out by placing the cans in water at 80 °C for 2–3 h. Drums containing AMF are either stored in rooms at 45–50 °C for 24 h or rapidly treated in steam tunnels or in a hot-water-heated Primodan butter melter before use. The fat or oil is added into the mixing tank after complete dissolution of the milk powder has been achieved. Depending on the equipment, the melted fat is either pumped or discontinuously added into the mixing vessel.

Filtration, Homogenization, and Pasteurization

From the mixing tank, the product is transferred via a balance tank to a separation unit, consisting of either duplex filters, made from stainless steel with nylon nets, or a clarifier, in order to remove extraneous matter and undissolved particles. Most recombined milk products are homogenized. Depending on the product type, different conditions of homogenization are employed. For most of the products, two-stage homogenization should be preferred. Typical pressures are 14 MPa plus 3.5 MPa for whole milk, 17 MPa plus 3.5 MPa for evaporated milk. (*See* **Filtration of Liquids**.)

Pasteurization is carried out mainly following the conventional dairy technology, using continuous heat exchangers rather than batch heating. The latter system is used only in small-scale production units. Typical conditions are 73 °C for 15 s for whole milk, 75–80 °C for 20–30 s for evaporated milk, 86–92 °C for 30 s for sweetened condensed milk. (*See* **Pasteurization**: Principles.)

Normally, the recombined milk flows from the production line to the filling station, thereby passing buffer tanks. They should be of the aseptic type in the case of sterilized or UHT milk. (*See* **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; **Sterilization of Foods**.)

Future Perspectives

Scientific progress has contributed considerably to the development of improved technologies for recombination of dairy products as well as to the successful use of special dairy-based ingredients. Examples such as the manufacture of recombined feta-type and other cheeses, using retentate or high-protein powders from ultrafiltration technology, confirm this trend. In this context, the functional properties of various components should be well recognized, since they can be utilized specifically. Furthermore, in many of the countries with a recombination industry, long-life milk products would allow a more convenient distribution and storage. Following this aim, modern processes such as UHT technology, cleaning-in-place cleaning facilities and aseptic packaging lines must be considered as factors of growing importance. (*See* **Cheeses**: White Brined Varieties.)

See also: **Antioxidants**: Natural Antioxidants; **Butter**: The Product and its Manufacture; **Casein and Caseinates**: Uses in the Food Industry; **Cheeses**: White Brined Varieties; **Emulsifiers**: Uses in Processed Foods; **Filtration of Liquids**; **Flavor (Flavour) Compounds**: Production Methods; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; **Pasteurization**: Principles; **Stabilizers**: Types and Function; **Sterilization of Foods**; **Vegetable Oils**: Types and Properties; **Vitamins**: Overview; **Whey and Whey Powders**: Production and Uses; Protein Concentrates and Fractions

Further Reading

Al-Tahiri R (1987) Recombined and reconstituted milk products. *New Zealand Journal of Dairy Science and Technology* 22: 1–23.

International Dairy Federation (1982) *Milk for the Millions. Proceedings of the IDF Seminar on Recombination of Milk and Milk Products, Singapore 1980. Bulletin No. 142.*

International Dairy Federation (1990) *Recombination of Milk and Milk Products. Proceedings of the International Seminar, Alexandria, Egypt, 1988. Special Issue No. 9001.*

Kjaergaard Jensen G and Nielsen P (1982) Reviews of the progress of dairy science: milk powder and recombination of milk and milk products. *Journal of Dairy Research* 49: 515–544.

Mann EJ (1988) Recombined and reconstituted milk. *Dairy Industries International* 53: 15–16.

Pedersen PJ (1985) Plants for recombination. In: Hansen R (ed.) *Evaporation, Membrane Filtration and Spray Drying in Milk Powder and Cheese Production*, pp. 373–383. Vanlose, Denmark: North European Dairy Journal.

Recommended Dietary Allowances *See* **Dietary Reference Values**

Refining *See* **Sugar**: Sugarcane; Sugarbeet; Palms and Maples; Refining of Sugarbeet and Sugarcane; **Spices and Flavoring (Flavouring) Crops**: Tubers and Roots; Properties and Analysis; **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage; Fungi in Food – An Overview; **Soy (Soya) Beans**: The Crop; Processing for the Food Industry; Properties and Analysis; Dietary Importance

Refrigeration *See* **Chilled Storage**: Principles; Attainment of Chilled Conditions; Quality and Economic Considerations; Microbiological Considerations; Use of Modified-atmosphere Packaging; Packaging Under Vacuum

REFUGEES

A M Madden, London Metropolitan University, London, UK

Introduction

A refugee is a person who, 'owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group, or political opinion, is outside the country of his nationality, and is unable ... to avail himself of the protection of that country,' as defined by the United Nations High Commission for Refugees (UNHCR). It is estimated that there are approximately 12.1 million people in the world today who have legal status as refugees and that a further 1.2 million people, described as asylum seekers, are in the process of applying for refugee status. It is also estimated that another 25–30 million people have fled their homes for the same reasons as refugees, but as they have not left their country of origin, they are classed instead as displaced people. These groups are distinguished from economic migrants who retain, at least theoretically, the protection of their country of origin. The problems relating to food and nutrition for refugees, asylum seekers, and displaced people are similar and, in many situations, indistinguishable. For simplicity in this article, the term 'refugee' will be used to include all asylum seekers and displaced people except where there is a significant legal difference which pertains to their ability to purchase food or obtain employment.

Since data were first collected in 1951, the total number of individuals categorized as refugees has increased substantially overall and in most regions of the world (Table 1). War and its consequences have led to the movement of the greatest numbers of refugees during this period, as typified in Europe after the Second World War, in South-east Asia following the prolonged and geographically widespread effects of the Vietnam war, and in India where 10 million people fled to India in 1971 following the Bangladeshi war of independence. In many situations, the effects of war have been exacerbated by famine,

Table 1 Estimated number of refugees by region (in thousands)

	1951	*1970*	*1980*	*1990*	*2000*
Africa	5	998	4154	5891	3611
Asia	42	159	2728	7944	5375
Europe	1221	646	574	1468	2427
Latin America and Caribbean	120	110	179	1197	38
North America	519	519	942	618	629
Oceania	180	44	315	110	68
Total[a]	2116	2480	8894	17 229	12 148

[a]Including some uncharacterized by region.

Source: governmental data compiled by United Nations High Commission for Refugees (2000 and 2001). United Nations High Commissioner for Refugees, UNHCR (2000) *The State of the World's Refugees*. Oxford: Oxford University Press.

either directly or in combination with drought, as in the Horn of Africa during the 1980s.

In addition to large movements of populations whose motivation is primarily survival, some people become refugees as individuals and for personal reasons. These may include political persecution, conscientious objection to war or military service, or fear of attack for refusing to comply with cultural expectations, for example, wearing restrictive clothing or undergoing genital mutilation. Inevitably, the nutritional implications of refugee-hood for such individuals are very different from those affecting refugees displaced in large numbers as a consequence of armed conflict.

The food and nutritional issues affecting refugees vary considerably depending on their geographical homeland and area of relocation, the health and nutritional status of individuals before displacement, the duration of displacement, external factors such as conflict and loss of infrastructure, and the availability of humanitarian relief. In addition, psychosocial and cultural influences may also have a major effect on the nutritional well-being and food needs of these people which in total represent almost 1% of the world's population. Against this background, the food and nutrition of refugees can be considered in two separate contexts: first, the emergency provision of food in refugee camps and, second, the nutritional well-being of refugees after moving to countries of asylum.

The Emergency Provision of Food in Refugee Camps

Article 25 of the Universal Declaration of Human Rights (1948) states that 'Everyone has the right to a standard of living adequate for health ... including food ....' While recognizing the indispensable need for food, it is impossible to consider the nutritional requirements of refugees in isolation without acknowledging the essentiality of a place of physical safety and access to clean water, sanitation, and health care facilities. These factors, along with nutrition and food aid, represent the five pillars which support the Humanitarian Charter, an interagency collaboration aiming to increase the effectiveness of humanitarian assistance to people affected by disasters. Each pillar of the Charter is in turn supported by a set of universal minimum standards. The standards for nutrition focus on general nutrition support to the population and nutrition support to those suffering from malnutrition; the standards for food aid consider the food rations required, targeting of food, resource management, logistical operations including transport and storage, and food distribution to the people in need.

Food Rations for Refugees

Food rations for refugees generally comprise a staple cereal, vegetable oil, and sometimes pulses and salt. Some rations are deficient in overall energy content and provide inadequate micronutrients. As a result, refugees may suffer from diseases of nutrient deficiency, including protein energy malnutrition, beriberi, pellagra, and scurvy. In 1997, the average nutritional requirements for refugee food rations were increased by the World Food Programme and UNHCR to 2100 kcal (8782 kJ) per person per day, in line with World Health Organization recommendations (Table 2). These figures are based on a defined demographic profile which, if different from the population under consideration, will necessitate adjustment to the average requirements.

In order to prevent or correct micronutrient deficiencies, the World Food Programme and UNHCR recommend that at least some of the food rations should be provided in the form of blended foods, particularly for refugees who are totally dependent on food aid. Blended foods are a precooked mixture of cereal and oil seed and sometimes a pulse such as chickpeas to which a vitamin and mineral mixture has been added. Generally, they are expensive as most are produced in the USA, although the World Food Programme has also procured some produced locally to requirements.

Powdered or modified milk should not be included in general food distribution to refugees because of the potential hazards associated with inappropriate dilution, contamination and, in some populations, lactose intolerance. Support for breast-feeding is paramount as facilities for safely preparing formulae are rarely available in refugee camps; in situations where sanitation is inadequate, the death from diarrheal disease

Table 2 Average nutritional requirements used for initial planning purposes in refugee camps, based on a defined demographic pattern

Nutrient	*Mean population requirement*
Energy	2100 kcal (8782 kJ)
Protein	10–12% total energy (52–63 g) but < 15%
Fat	17% total energy (40 g)
Vitamin A	0.5 mg retinol equivalents
Thiamin	0.9 mg (or 0.4 mg per 1000 kcal intake)
Riboflavin	1.4 mg (or 0.6 mg per 1000 kcal intake)
Niacin	12.0 mg (or 6.6 mg per 1000 kcal intake)
Vitamin C	28.0 mg
Vitamin D	3.2–3.8 μg calciferol
Iron	22 mg (low bioavailability, i.e., 5–9%)
Iodine	150 μg

Source: The Sphere Project (2000) *Humanitarian Charter and Minimum Standards in Disaster Response*. Oxford: Oxfam Publishing, with permission.

is 14 times higher in artificially fed babies than in those who are breast-fed.

Suitability of Food Aid

Although in extreme starvation people might consider eating any potential food, it is well recognized that the acceptability of foodstuffs provided for refugees in humanitarian relief operations must be considered in order to optimize any nutritional benefit. Unpopular, unpalatable, or difficult-to-prepare foods may be exchanged for preferred items as observed, for example, in camps in Kenya, Guinea, and Zaire, where refugees sold maize at a financial loss in order to buy rice and nonfood requisites. Although the resulting diets had an improved micronutrient content, the total energy content fell, further jeopardizing nutritional well-being. Generally, blended cereal foods are well accepted, particularly by people who are already familiar with porridge-type foods. In Ethiopian refugee camps, where only a limited variety of food items were available for approximately 6 months, sensory-specific satiety or ration fatigue has been observed. It has been suggested that the provision of sugar, chilli, and other spices may alleviate poor appetites induced by repeatedly eating the same foods; although flavorings and condiments contribute little nutritional value of their own, they may help refugees eat sufficient to maintain an adequate intake.

Nutritional Well-Being After Moving to Countries of Asylum

While considerable attention has been given to the nutritional needs of individuals living temporarily in refugee camps, relatively less regard has been paid to the nutritional well-being of refugees once they have settled in a host country which is often geographically and culturally distant from their homeland. Governmental data indicate that approximately 418 000 people applied for refugee status in European countries during a 12-month period to March 2001, with approximately 18% of applications being received in the UK (Table 3).

The provision of housing, legal assistance, and basic health care is usually amongst the first priorities dealt with by organizations assisting newly arrived refugees and asylum seekers, with education and psychological support, primarily for the young, also receiving attention. Few systematic studies have been undertaken to evaluate the nutritional status of refugees living in western countries, although there are a number of factors which suggest that they may be at nutritional risk (Table 4).

Table 3 Asylum applications lodged in Europe (2000–2001)

Country of asylum	*Number of applications*	*Share (%)*	*Ranking*
Germany	80 879	19.4	1
UK[a]	75 290	18.0	2
Netherlands	41 944	10.0	3
Belgium	41 642	10.0	4
France	39 043	9.3	5
Austria	21 995	5.3	6
Total	417 743	100	

[a]UK data are number of cases (average of 1.3 individuals per case).
Source: Governmental data (1 April 2000–31 March 2001) compiled by United Nations High Commission for Refugees.

Table 4 Factors influencing the nutritional status of refugees after resettlement

Nutritional status on arrival
Poverty
Cultural factors
Communication
Psychological issues

Nutritional Status on Arrival

The nutritional status of some individuals may be suboptimum when they arrive in the host country and this will be influenced by their general health, the nature of their departure from their homeland, the journey undertaken in flight, and any intervening periods spent in camps en route. Refugees from some countries, for example, Ethiopia, Somalia, and Sudan, may have been exposed to communicable and parasitic diseases, including tuberculosis, malaria, and schistosomiasis, which may further impair an already diminished nutritional status. Individuals arriving from European countries are less likely to suffer from obvious nutritional depletion, generally because of their better initial nutritional status, shorter periods of displacement, and prior vaccination to common communicable diseases, although inevitably there are exceptions. Some refugees who have endured long overland journeys, often covertly crossing several international borders, may arrive depleted and at risk of weight loss and subclinical micronutrient deficiency after consuming an inadequate diet for several weeks or longer.

At present, many templates for the health screening of newly arrived refugees in the UK do not make reference to nutritional variables. This is a cause for concern, particularly for infants and children who are most likely to be at nutritional risk and for whom the consequences of nutritional depletion are greatest. Evaluation of their nutritional status on arrival is, therefore, imperative.

Not all newly arrived refugees have nutritional problems, however, and many are generally in good health and physically fit. There is evidence that the health status of some refugees deteriorates and is worse 2–3 years after arrival in the host country. In some cases the cause for this deterioration may relate specifically to their situation as refugees, for example with tuberculosis and mental health problems, while in others, it may be a consequence of adopting an unhealthy urban lifestyle and diet. Obesity has been observed in Chilean and Bosnian refugees in Sweden, even after adjustment for socioeconomic status, and there is a high incidence of dental caries in young Vietnamese refugees which increases with the length of stay in the host country.

Poverty

After arriving in a host country, many refugees live in poor housing, receive limited financial support, and have difficulty obtaining paid work. As a consequence, their ability to purchase food and eat a healthy diet is constrained. Policies of dispersal in the UK have resulted in many refugees being housed in areas away from their initial point of entry, but typically, refugees settle in the poorest areas of large cities. Many of these deprived urban areas, where few supermarkets are located, are considered 'food deserts' and, in the absence of adequate cheap transport, the local residents have little choice but to buy their food from small corner shops. Generally, these sell a limited range of food products and few fresh fruit and vegetables, for which they charge relatively high prices. Asylum seekers are initially housed in temporary accommodation until their refugee status is legally confirmed, which may take several months or longer. Such accommodation is usually found in bed-and-breakfast establishments or small hotels where there are few cooking facilities and limited storage for food. Inevitably, it is difficult for people living in these situations with limited money to eat a healthy diet.

The financial support provided for refugees varies between host countries, the legal status of the applicant and, in some cases, where the application for refugee status is made. In most countries, no benefit is paid to people applying for asylum until they are granted legal refugee status unless they can demonstrate financial hardship and then they may receive reduced levels of benefit (Table 5). For example, asylum seekers in Germany receive their accommodation, a food parcel, some clothing, and a small sum of money for 'personal requirements' each month. Financial benefit is only paid in exceptional circumstances and then, for the first 12 months, at a rate 20% lower than that given to other members of the community. Similarly, in the UK, asylum seekers are supported through a virtually cashless system, receiving vouchers to a value of 30% less than income support paid to other recipients. Vouchers can be exchanged for food and basic necessities only at designated shops but cannot be traded for money. This has caused some difficulty for refugees wishing to purchase halal meat from specific outlets. The introduction of food vouchers is unpopular and has been widely criticized by a number of refugee and nutrition organizations as inflexible and potentially compromising nutritional intake. Applications for asylum may take several months or longer to be processed, during which a single adult aged over 25 years would receive £36.54 per week in vouchers. The quantity and quality of food that can be purchased for this figure is limited, particularly if access to cheaper shops and cooking facilities are not available. If individuals are granted legal refugee status they are entitled to cash benefits at the usual value of income support to other UK claimants and may also be entitled to a back-dated cash sum. On being granted legal status, refugees are entitled to move from temporary

Table 5 Eligibility of asylum seekers for benefits and employment

Country of asylum	*Financial and other benefits*	*Employment eligibility*
Australia	Financial hardship only: 89% of standard benefits	Dependent on visa (if any)
Belgium	Theoretically eligible but local authority-dependent	Provisional work permit if claim deemed admissible
Canada	Welfare assistance available if meet financial criteria	Eligible to seek work after submitting appropriate form
France	Financial help as a cash sum. Eligible for social security	No access to labor market unless authorized
Netherlands	Free board and lodging plus small monthly cash sum	Not eligible
Sweden	Eligible for social benefit, paid daily if necessary	No work permit required if decision expected < 4 months
Switzerland	Eligible for welfare benefits at reduced rates paid 'in kind'	Not eligible for first 3 months
USA	Information not available	Not eligible

Adapted from Bloch A and Levy C (1999) *Refugees, Citizenship and Social Policy in Europe*. Basingstoke: Macmillan Press.

accommodation to a more permanent residence where the cash sum could be used to purchase kitchen and other essential household items which might help facilitate an improved food intake.

The term 'refugee' embraces people from a multitude of backgrounds and includes people with a wide variety of employment skills. Although varying greatly with area of origin and gender, a high proportion of refugees are well-educated and professionally trained. However, obtaining work is forbidden in most host countries until legal refugee status is established, and even then, is often very difficult, even for professionals whose qualifications may not be recognized (Table 5). As a result, there is a high level of unemployment in most refugee communities and, of those who are working, many are employed in low-paid, unskilled, and sporadic jobs. Whilst not influencing food intake directly, the difficulty in obtaining work and consequent relative poverty exacerbates the problems already outlined and makes it difficult for individuals to work their way out of financial hardship.

Cultural Factors

Refugees are not a homogenous group which can be considered together, coming as they do from many different countries from all geographical regions of the world and bringing with them a wealth of culture and tradition which may differ considerably from those of their host country (Table 6).

In addition to providing dietary energy and essential nutrients, food plays important social and cultural roles for most people. The foods eaten, the methods of preparation and serving, and the context in which they are consumed often have significant social meaning and help define group identity. For refugees displaced from their homeland and often separated from their families and friends, the importance of food symbolism may grow and in some cases may represent one aspect of their lives over which they have some control. However, difficulties arise when familiar food items are unavailable or can only be purchased at great expense or from specialist shops in areas where there is a refugee community of significant numbers. Similarly, familiar cooking implements or facilities may not be available and preclude the preparation of specific dishes. Whilst most refugees quickly become skilled at adapting available substitutes and learning new skills, food intake will by necessity change and may lead to a diet which is nutritionally inadequate compared to their traditional fare.

Food habits may also need to be adapted in terms of who prepares the meals and with whom food is shared. Refugees who traditionally live in extended families where women frequently undertake most of the food preparation and do not have outside employment may find themselves living in smaller and more isolated nuclear family groups where a large shared meal is no longer a major focus of the day. Many refugees are young men who may have had little practical experience of preparing their own food. A study of single male Ethiopian refugees living in Ottawa, Canada, found that their total energy intake was related to their domestic skills and that a lack of such skills was the most commonly perceived dietary problem.

Children who may have been expected to contribute to food preparation and cooking in their homeland may be influenced by their host country peers, who generally have less responsibilities, and as a result wish to avoid participation in household tasks which might be viewed by others as exploitation. In Minneapolis, USA, 30–45% of adolescent refugees from South-east Asia report primary responsibility for preparing the evening meal. A Kurdish mother in London described how 'at home, an eight-year-old could cook a meal. Now they can't do anything. They just watch television all day'. Some children may select a more western-type diet in order to try to fit in with their contemporaries or to distance themselves from painful memories of their homeland while others may cling to traditional habits and avoid the stress of trying unfamiliar foods.

Communication

Some, but not all, refugees arrive in the host country speaking the local language. For those who do not, an inability to communicate easily may increase difficulties experienced in many aspects of life, including shopping for food. This becomes an increasing problem when the available foods are unfamiliar and ingredients or cooking instructions cannot be read.

Table 6 Origin of asylum applicants in Europe (2000–2001)

Nationality	*Number of applications*	*Share (%)*	*Ranking*
Iraq	37 902	9.1	1
Yugoslavia, Former Republic of	37 174	8.9	2
Afghanistan	32 992	7.9	3
Iran, Islamic Republic of	27 853	6.7	4
Turkey	24 647	5.9	5
Russian Federation	15 489	3.7	6
Bosnia and Herzegovina	13 468	3.2	7
Sri Lanka	11 986	2.9	8
India	9516	2.3	9
Sierra Leone	8321	2.0	10
Total	417 743	100	

Source: Governmental data (1 April 2000–31 March 2001) compiled by United Nations High Commission for Refugees.

For refugees who may be familiar with growing much of their own food or shopping in a market, food shopping can present some difficulty. Generally, school-aged children and men are more likely than women to speak the host country language and yet women are often charged with shopping for food. Whilst interpreters may sometimes be available, they generally assist with legal issues or essential health matters and are rarely used for what may be considered more mundane food-related communication. However, health authority interpreters in the UK provide a free service for hospitals and general practice surgeries and should be available for nutritional consultations if required.

Psychological Issues

Refugees may have faced many diverse experiences before arriving in their host country. Some have suffered violence, loss, separation, family disruption, and terrifying journeys. Concerns may continue even when a place of perceived safety is reached, including uncertainty over asylum, anxiety for relatives left behind, difficult living conditions, hostility from local people, and the challenges of settling into a new life. As a consequence, many refugees report psychological problems, with two-thirds of adult refugees in the UK experiencing anxiety or depression and 90% of Bosnian children in the USA showing signs of posttraumatic stress disorder. Other symptoms, which may be exacerbated by poverty and isolation, include poor sleep patterns, loss of memory, panic attacks, and agoraphobia. Loss of appetite and lack of interest in food may also be a problem.

Approaches to psychological well-being vary between cultures. Refugees from some backgrounds may be uncomfortable with western-style counseling or 'talking treatments' and prefer instead to attend to practical issues which they consider will make their lives better. Interestingly, food and nutrition are perceived by many as 'safe' subjects and have been used as a focus for discussion groups in both Canada and the UK as a way of providing practical and emotional support for refugee women in a nonthreatening environment. The cooking of traditional recipes and sharing of food has also been used as therapy in a group of London-based refugees recovering from torture.

See also: **Eating Habits**; **Ethnic Foods**; **Famine, Starvation, and Fasting**; **Infants**: Breast- and Bottle-feeding; **Nutritional Assessment**: Importance of Measuring Nutritional Status; Anthropometry and Clinical Examination; **Politics and Nutrition**; **World Health Organization**

Further Reading

Aldous J, Bardsley M, Daniell R *et al*. (1999) *Refugee Health in London*. London: The Health of Londoners Project.

Bloch A and Levy C (1999) *Refugees, Citizenship and Social Policy in Europe*. Basingstoke: Macmillan Press.

Burnett A and Peel M (2001) Health needs of asylum seekers and refugees. *British Medical Journal* 322: 544–547.

Hilderbrand K, Boelaert M, Van Damme W and Van der Stuyft P (1998) Food rations for refugees. *Lancet* 351: 1213–1214.

Mears C and Young H (1998) *Acceptability and Use of Cereal-Based Foods in Refugee Camps*. Oxford: Oxfam GB.

Richman N (1998) *In the Midst of the Whirlwind. A Manual for Helping Refugee Children*. Stoke on Trent: Trentham Books.

Sellen DW and Tedstone A (2000) Nutritional needs of refugee children in the UK. *Journal of the Royal Society of Medicine* 93: 360–364.

The Sphere Project (2000) *Humanitarian Charter and Minimum Standards in Disaster Response*. Oxford: Oxfam Publishing.

Trafford P and Winkler F (2000) *Refugees and Primary Care*. London: Royal College of General Practitioners.

United Nations High Commissioner for Refugees, UNHCR (2000) *The State of the World's Refugees*. Oxford: Oxford University Press.

Wallace R and Brown A (1999) *Asylum Practice and Procedure*. London: Trenton Publishing.

Regulations *See* **Legislation**: History; International Standards; Additives; Contaminants and Adulterants; Codex

RELIGIOUS CUSTOMS AND NUTRITION

P Christian, Johns Hopkins University, Baltimore, MD, USA

Religious Dietary Laws and Nutritional Status

Food in all human societies is more than just a source of nutrition required for survival. It is deeply entrenched in the religious, social, and economic aspects of life and carries with it various symbolic meanings, many of which express the relationship between humans and their deities. Among numerous factors that shape dietary practices, religious laws are most prevailing. Religious dietary regulations which determine food-related behaviors of individuals are codified in the written religious texts and holy scriptures, although different interpretations of these may also be found. The extent to which individuals follow food-related prescriptions or proscriptions may depend on their religiosity – the degree to which they adhere to a set of religious beliefs. Some dietary habits may not have their origins in religious doctrines, but are adhered to due to a strongly held cultural or traditional belief common to the members of a religious group and may be tied to a religious or historical event.

The origins of religious dietary laws have been extensively debated by both ancient and modern-day scholars, although no single theory appears to be well-accepted. Among theories that exist, the ones that are related to dietary laws being divine commandment, or having allegorical parallels or esthetic origins are considered subjective and emotional in nature. Other theories, based on reasons of health and sanitation, ethnic identity, and ecology, are generally more well-accepted. For example, reasons of health and sanitation are considered the best scientific explanations for the prohibition of carrion, pork, and blood in Judaism and Islam. Supporters of this theory contend that the cause–effect relationship between ingestion of specific foods and the manifestation of disease are reflected in the dietary laws prescribed in the scriptures. On the other hand, the ecologic theory propounds that because the environment is more favorable towards the availability of certain types of flesh foods but not others, proscriptions regarding eating the latter type of animal foods makes economic sense. The hypothesis pertaining to ethnic identity suggests that dietary laws and codes set aside one religious/ethnic group from another and tend to foster religious and ethnic cohesiveness.

Noteworthy are the shared attributes of the dietary regulations across different religions. All major religions have prohibitions, either permanent or temporary, regarding consumption of animal foods. In Hinduism and Buddhism, killing living creatures is abhorred and meat consumption is forbidden. Judaism and Islam forbid the consumption of pork, and meat intake is restricted on fasting days among orthodox Christians, and on Fridays among Catholics. Most religions also advocate religious fasting, usually associated with religious observances. Fasting is considered to be an act of penitence and is practiced in order to gain spiritual merit. Most religious fasts last over a brief period of time with either total abstinence from food or from specific food items. The Ramadan fast, observed every year by Muslims, worldwide, is the most studied of all religious fasting customs. The Eastern Orthodox Church observes two major fasts: the Lent fast and the Advent fast in the 40 days preceding Easter and Christmas, respectively.

Religious dietary prescriptions or proscriptions as outlined by dietary laws and guidelines can be broadly classified into ceremonial, periodic, or habitual (Figure 1). Ceremonial prescriptions, such as

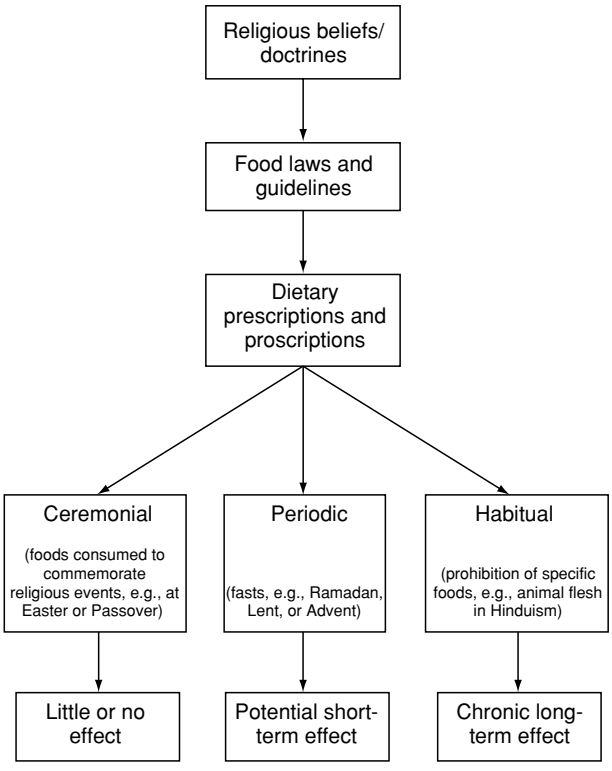

Figure 1 Religious dietary laws and types of nutritional effects.

eating lentil soup on Good Friday, symbolizing the tears of the Virgin Mary, or lamb on Easter, or 'seder' plate with foods symbolic of events in Jewish history at Passover, may have little or no nutritional effects and are not considered here. Periodic customs such as religious fasting may result in a potential short-term alteration in a group's nutritional status. Dietary guidelines that prohibit all animal foods, such as in Hinduism and Buddhism, are habitual and likely to have chronic long-term effects on nutritional status and health. Prohibition of pork in Judaism or Islam falls in this category of dietary proscriptions but its nutritional effects are unknown. Data are scant on the relationship between religious customs and nutritional status. Also, certain food practices may be latent in modern times, with little known about their nutritional effects. None the less, particular groups of individuals, following clearly defined and unique dietary customs of their religion, have been studied, to some extent, in order to understand better the nature of such customs in relation to their nutritional and health outcomes. These cover religious dietary customs of each of the five major religions of the world – Christianity, Islam, Judaism, Hinduism, and Buddhism – and are described below in more detail.

The Fast of Ramadan

One of the most common religious practices uniformly followed by Muslims throughout the world is that of fasting in the month of Ramadan. Fasting in Ramadan is one of the Five Pillars of Islam, the others being the statement of faith, prayers five times a day, giving your dues to the poor, and pilgrimage to Mecca. The holy scriptures of Qur'an reveal the reasons for fasting: the main one is the Revelation of Qur'an, a guidance for mankind and the criterion of right and wrong, which occurred in this month (2.185). Fasting is also considered a training in self-control, of restraining one's needs and desires, and in deepening spiritual life. According to Prophet Mohammed, fasting can also improve health, and Muslims fast for better health.

Fasting during Ramadan is partial or controlled, as Muslims abstain from food and drinks during the hours from sunrise to sundown. Fasting is obligatory for every adult Muslim, although sick individuals or those traveling, and pregnant and lactating women are exempt but have to make up for the missed days. Elderly people and children under the age of puberty are also exempt, although the former group is expected to give food, cash, or kind to a needy person for each day of missed fast. Fasting is forbidden for women during menstruation but they are also required to make up for the missed days.

On the night before fasting commences and each night thereafter, Muslims consume a predawn meal called 'sahoor', after which they abstain from eating or drinking until sundown. Immediately after sunset the fast is broken with a meal called 'iftar'. There are primarily three types of changes that occur in the food-related behaviors of Muslims during this month. One is the change in the frequency and time of meal consumption, with long periods of fasting and short periods of food intake which might result in gorging. Frequency of eating decreases from an average of 4.4 times to 2.8 times per day during Ramadan. The second change that occurs is in the types of foods consumed during Ramadan. It is customary to prepare and consume special foods which are rich in protein and carbohydrates as well as eat fresh or dried fruits and vegetables. The third notable change is in the sleeping pattern, since the number of hours of sleep is reduced due to the predawn partaking of breakfast.

Fasting During Ramadan and Nutritional Status

Ramadan fasting has drawn considerable interest among nutritional scientists as it provides an excellent opportunity to study the effects of restricted energy intake and altered periodicity of meals on changes in body weight, composition, fluid, and electrolyte balance, and metabolic and biochemical changes, especially those of serum lipid concentrations in free-living populations. Most data are derived from studies using small numbers of subjects who are examined before and after the Ramadan fast. Rarely are nonfasting control groups used to account for seasonality or other changes that occur during the month, nor are the long-term effects of these annual fasts described.

Energy intake and weight changes during Ramadan

Fasting during Ramadan results in moderate body weight losses in both men and women, primarily due to declines in energy intake. Body weight decreases range between 0.3 kg and almost 2.5 kg, with energy intake reductions, in some cases, as high as 20–25%. Significant reductions in skinfold thickness and elevation in plasma glycerol concentrations suggest that decreased body fat may be the main component of body weight loss. Decreased water intake by more than half a liter per day can cause dehydration, also contributing to a reduction in body weight. A pattern of slower decline in weight observed in a few settings may be attributed to metabolic adaptation and reduction in physical activity over time.

Although in general there is a reduction in body weight during Ramadan fasting, some individuals or

groups may experience no change, or in some cases, even weight gain. The weight increase may be related to increased consumption of foods high in sugars and protein which is common during Ramadan. It is also customary to eat meals with friends and relatives, which is conducive to higher food intake. Further, a decline in respiratory exchange ratio reflecting increased mobilization of body fat is suggestive of a metabolic adaptation to fasting.

Dehydration due to abstinence from drinking fluids over long periods of time during Ramadan has been a cause for some concern, especially in hot tropical climates. Negative fluid balance among fasting individuals is evident by increased hematocrit, serum sodium, calcium, protein, creatinine, urea, and modified electrolyte balance. Increased concentration of urine and decreased urine volume and salt retention are some of the compensatory mechanisms to achieve fluid balance. The season in which the month of Ramadan falls may have a big effect on the evaporative water loss that can occur. Studies have shown conflicting results: some suggest no negative fluid balance while others show this may be due to environmental factors such as temperature and humidity of the place and season in which the studies were carried out.

The effects of Ramadan fasting on changes in carbohydrate metabolism are not well-known. Although fasting levels of glucose and insulin tend to remain unchanged, or decline initially but then return to normal later, circadian profiles of gastric pH, plasma gastrin, insulin, glucose, and calcium are modified during Ramadan. Increased concentrations of urea and uric acid suggesting increased catabolism have not always been confirmed.

The effect of Ramadan fasting on plasma lipids and lipoproteins has received some interest. Almost consistent reports of no change in total cholesterol and low-density lipoprotein-cholesterol (LDL-C) and beneficial increases in high-density lipoprotein cholesterol (HDL-C), apo A-1, and apo A-1 to apo B ratio are found. While these effects may be sustained for about a month after Ramadan ends, the impact of such transient annual changes on cardiovascular disease remains unknown and open for research.

Pregnancy and Lactation and Ramadan Fasting

The rules of exemption from fasting in certain conditions such as during pregnancy or while breastfeeding an infant appear to have varying degrees of acceptance and interpretation. In some cultures, pregnant women may either be unaware of this exemption or may choose to fast with the rest of the family rather than alone, at a later time, due to the inconveniences of a different meal pattern and cooking. Cases of unsuccessful lactation due to fasting have led to unhygienic alternatives being offered to the young infant in lieu of breast milk, resulting in adverse health effects. The effects of dehydration due to abstention from drinking water during Ramadan on lactational performance suggest that, although lactating women may lose some of their total body water during the daylight hours, plasma indices of dehydration over a 24-h period remain in a normal range. This may occur because women adapt by superhydrating themselves overnight and by restricting urinary output. Despite these mild changes in their state of hydration, there may be marked changes in breast milk osmolality, lactose, sodium, and potassium concentrations, reflecting disturbed milk synthesis. Accounts of drastic effects of fasting on young breast-feeding infants is found in the 1930s when severe forms of xerophthalmia (clinical signs of vitamin A deficiency) occurred in infants whose mothers participated in long-term religious fasts.

Scientific knowledge on the health and other impacts of fasting during pregnancy is scant, although metabolic consequences of fasting during pregnancy, such as declines in concentration of serum glucose, insulin, lactate, creatinine, and a rise in triglycerides and hydroxybutyrate suggests a state of a phenomenon termed as 'accelerated starvation.' Although pregnancy outcome, including birth weight of infants whose mothers observed Ramadan fast during pregnancy, does not appear to be affected in affluent settings, increased rates of low birth weight, stillbirths, and perinatal mortality are observed among populations living in poor, more stressful environments. It is hard to determine, however, whether the adverse outcomes are a direct result of the fasting. Interestingly, increased full-term births have been recorded during the 25-h-long fast of Yom Kippur.

The Hindu Conception of Vegetarianism: The Sacred Cow

In the pre-Aryan Indus valley civilization, it was the bull, not the cow, that was venerated. Cattle were prized assets and considered objects of worship, as recounted in the hymns of Rig-Veda, the holy Hindu scriptures written in 1500–100 BCE. The Atharva-Veda, which came later, is where the first accounts of the prohibition of the slaughter of cows are found. While meat eating was practiced, and cows were occasionally slaughtered for guests and at times of festivity, the sacredness of the cow, manifest as the Vahan (carrier) of god Shiva became complete in the great epic Mahabharata. Here the sanctity of a cow is recognized and attributed to being a source of food (milk), promoting fertility, removing evil influences,

and assisting the soul in crossing the river of death at the end of life's journey. Even the Dravidian agricultural civilization recognized the practical value of cattle to provide manure as a fertilizer or to work pulling ploughs, apart from providing milk and its products which had great nutritional value to families and communities.

The taboo of killing living creatures at large is tied to the concept of ahimsa or nonviolence that prevails widely in Hinduism, Buddhism, and Jainism. Vegetarianism is practiced by devout Hindus (especially Brahmin priests of the highest caste), Buddhists, and Jains. It is difficult to assess the nutritional effects of this habitual religious prescription of vegetarianism in populations which are primarily vegetarians due to a lack of adequate control omnivorous groups from the same population. Nonvegetarian groups living in the same environment may come from differing religious or socioeconomic backgrounds. Further, nutritional deficiencies common in developing countries where vegetarianism is practiced (for example, India) may occur due to multiple environmental and other factors, which makes it hard to study the solitary effects of a vegetarian diet. However, there is a significant amount of literature gathered over the past few decades of the effects of vegetarianism as practiced among Hindus, Buddhists, and other Asian groups who have emigrated to Europe or North America. The diet and nutritional status of these populations has been compared with that of omnivores of indigenous origin living in the same environment and with similar socioeconomic backgrounds.

Hindu Vegetarianism and Nutritional Status

The nutritional adequacy of vegetarian diets (predominantly lactovegetarian) adhered to by certain religious groups has focused on minerals such as iron, zinc, copper, and vitamins B_{12} and D. The observation that infants born to predominantly vegetarian Hindu women have consistently low birth weight, although birth length is comparable to those of the indigenous Caucasian populations, has resulted in studies to examine whether nutritional deficiency, either of total calories or specific micronutrients such as zinc or iron might be a factor, although genetic differences cannot be completely ruled out. Zinc and copper intakes of vegetarian, pregnant Hindu women are lower compared to those of European pregnant women. However, no association is found between measures of zinc or copper status and birth weight. Moreover, plasma zinc and copper concentrations are surprisingly similar in both vegetarian and nonvegetarian Asians, whether living in India or Europe, suggesting that vegetarianism *per se* may not be associated with poor intrauterine growth and that ethnic differences are likely to explain the detrimental birth weight among Asians.

Influence of maternal vegetarianism on essential fatty acid status of newborn and pregnancy outcome has also been examined in Asian vegetarians compared with white omnivores. Although intake of eicosapantaenoic acid and docosahexanoic (DHA) acid are absent in the diets of vegetarians and infants of these women have significantly lower DHA in their plasma or cord blood, these factors are, again, not related to the outcome of pregnancy such as birth weight, length, and head circumference of infants. The prevalence of iron-deficiency anemia is higher among emigrant females and male vegetarian populations than that observed in omnivorous Caucasian populations in North America and western Europe, which may be related to the high fiber and phytate content of the former's diets. The major source of cereal grains in these populations is unleavened bread (chapatti). Although total iron intake is similar in both groups, the nonheme form is the major source of iron among vegetarians.

Diets of vegetarian Gujaratis (from the state of Gujarat in India) living in the UK have also been found to be lower in vitamins D and B_{12}, although intakes of other vitamins, such as A, B_1, B_2, and folic acid, are similar to those of the European nonvegetarian controls. It is interesting that subclinical vitamin D deficiency among vegetarian Asians (mostly Gujarati) in the UK resolves spontaneously during the summer months. No such seasonal variation is found among the Caucasian population, consistent with their adequate vitamin D status. Increased exposure to sunlight during the summer months followed by cold dark winter months in a population with dietary insufficiency of vitamin D may be an explanation for this seasonal phenomenon. Osteomalacia and rickets are more likely to occur in Asians living in the UK, especially the vegetarian Hindus, and young Asian children also have lower serum 25-hydroxycholcalciferol concentrations compared to white children.

Risk of tuberculosis (TB) in immigrant Asians (especially Gujaratis) is also found to be high and consistently higher than that recorded in their country of origin. TB rates are three to five times higher among vegetarian Asians than either nonvegetarian Muslims or Asians consuming a mixed diet. A range of socioeconomic, migration, and lifestyle variables fail to explain the risk of TB, suggesting a contributory role of dietary factors in determining susceptibility to disease. Vitamin D deficiency has previously been linked to impaired host defense against TB by increased mycobacterial reactivation. An early summer peak in notified cases of TB in the UK among Asians

is indicative of low postwinter trough concentration of 25-hydroxycholecalciferol contributing to the reactivation of latent infection.

The other deficiency frequently found among vegetarian Hindus residing in Europe is that of vitamin B_{12}. Milk is a source of vitamin B_{12}, but boiling milk to extract the fat, which is a common practice among Indians, can destroy much of the vitamin. Apparently healthy lactovegetarian Indians tend to have lower concentrations of vitamin B_{12} in serum and urine compared to nonvegetarians. This deficiency is not, in most instances, due to a metabolic defective absorption caused by the lack of intrinsic factor which results in pernicious anemia, but due to nutritional cobalamin deficiency. Although vitamin B_{12} deficiency is common on the Indian subcontinent, the evidence of a large-scale problem of megaloblastic anemia became apparent through studies of vegetarians who emigrated on a large scale from East Africa to UK. In India, however, many cases described as nutritional megaloblastic anemia may be due to both cobalamin and folate deficiencies.

Seventh Day Adventists and Vegetarianism

Not all the effects of vegetarian diets prescribed by certain religions have negative nutritional effects. In fact, when availability and diversity of food are not found to be confounding factors, the beneficial effects of a vegetarian diet have been well demonstrated. A good example is that of the most widely studied of modern religious groups, the Seventh Day Adventist (SDA) Christians who espouse primarily lactoovovegetarian diets and refrain from smoking or drinking. While these proscriptions are not found in the Bible, the SDAs maintain that the dietary practices are based on the general theological consideration that the human body is the temple of the Holy Spirit and should be protected. Numerous studies have documented the improved benefit of the SDA dietary and lifestyle factors on their health and survival. Rates of cancer and cardiovascular disease mortality are much lower among Adventists from the USA, Australia, and the Netherlands compared to non-Adventist groups. Because most SDAs do not smoke or drink alcohol, and have varying degrees of meat consumption, it is hard to distinguish between the beneficial effects of vegetarian diets versus lifestyle factors. However, there is evidence to suggest that among SDAs, vegetarians are healthier than nonvegetarians but this may not be ascribed only to absence of meat intake.

All-cause mortality among SDAs is shown to be negatively associated with green salad consumption and positively associated with consumption of eggs and meat. Low meat and saturated fat intake and high intakes of carbohydrates and dietary fiber may be associated with these marked protective effects. Low levels of cholesterol and blood pressure have been observed among some SDA groups. The protective effect against major sites of cancer may be due to higher intakes of foods providing vitamins A and C or in some SDA groups the lower frequency of digestive tract cancers may be due to consumption of broiled - meat or fish, which produces little or no benzo(α)pyrene and nitrosamines, which are carcinogenic products of heated or burnt meat. Higher consumption of all or dried fruits has been associated with reduced risks of lung, prostate, and pancreatic cancer among the Adventists adjusted for sex, age, and cigarette smoking history.

Although there might be some concern regarding inadequate intakes of certain trace minerals obtained through a vegetarian diet, dietary intakes and status of various nutrients such as iron and zinc have been found to be adequate among SDAs, despite the high content of total dietary fiber and phytate of a vegetarian diet and low intakes of iron and zinc from flesh foods. Body mass index, vitamin A, and hemoglobin concentrations among vegetarian and nonvegetarian SDAs are generally similar, although instances of low vitamin B_{12} have been reported in countries where food fortification with vitamins is not done and supplementary vitamins are not consumed.

Buddhist Vegetarianism

Similar beneficial effects of a vegetarian diet have been observed among young modern Chinese Buddhists. Their diet is high in carbohydrates with a moderate fat content since rice and soybean provide most of their calorie and protein requirements. The blood concentrations of cholesterol, glucose, and uric acid concentration of a vegetarian Buddhist are lower, and that of apo A-1 to apo B ratio higher, compared to Chinese omnivores, suggesting a protective effect of a vegetarian diet for chronic degenerative diseases. Similarly, the blood pressure of Buddhist vegetarians is significantly lower than those of nonvegetarians of similar age, sex, and body mass. This might explain in part the lower all-cause mortality observed among Zen Buddhist priests who consume a traditional Japanese diet comprising vegetables and seafood with low amounts of meat.

Jewish Dietary Laws and Nutrition

Although many accounts of Jewish dietary laws, their origins, and scientific rationalizations are found

in the literature, the effects of Jewish dietary prescriptions and proscriptions on the nutritional status of Jewish people are not well-examined. The Jewish dietary laws, called the Kashruth, are codified in the Torah in the books of Deuteronomy 14 and Leviticus 11. The Kashruth pertains to the type of foods allowed for consumption and the methods of their preparation. All fruits and vegetables and herbivorous, ruminant animals (domesticated and wild) whose hooves are wholly cloven are categorized as 'clean' and permitted. Pork is forbidden and only aquatic creatures that have at least one fin and scale are permissible. Animals are to be slaughtered by bleeding them to death after cutting through the neck below the larynx (the ritual slaughter is called shehitah). Also, animals with certain defects or injuries are not permitted. Only animals treated thus are considered kosher. Consumption of blood is disallowed, as is consumption of certain portions of 'clean' animals such as the sciatic nerve and certain fat portions. Carrion meat is also forbidden and it is prohibited to cook or eat meat and milk together. Thus, the dietary laws are complex but whether their practice poses any adverse or beneficial effects on the nutritional status of the Jewish people who follow them is unclear. There is no evidence to show that lack of pork meat in the diet results in any adverse nutritional outcome. Among the ultraorthodox Hasidic Jews living in Montreal who are strict followers of the Kashruth, all-cause mortality is low, primarily due to a strikingly lower mortality from diseases of the circulatory system, although community cohesiveness and affiliation may, to some degree, mediate these beneficial effects. Dietary intakes of Hasidic Jews are lower in energy, fat, and saturated fat compared to those of other Montrealers, Canadians, and to the NHANES II sample, although intake of cholesterol is higher in the Hasidics. Orthodox Jewish men and women living in Jerusalem also have a significantly lower consumption of total and saturated fats and higher carbohydrates compared to secular Jews. These dietary differences are reflected in the lower concentrations of plasma cholesterol, triglycerides, and LDL-C concentrations observed in the former group. A substantially higher risk of myocardial infarction among secular versus orthodox Jews is also observed.

Conclusion

Religious customs pertaining to diet can have potentially significant effects on the health and nutritional well-being of individuals depending on the degree to which they are prescribed. There remains a paucity of scientific data on the effects of dietary habits of religious origin on the nutritional status of populations, unconfounded by other cultural, social, and economic influences. There appear to be changes in aspects of nutritional status associated with chronic long-term or periodic shorter-term dietary practices, although the long-term implications of the latter changes on health are unclear.

See also: **Bioavailability of Nutrients**; **Coronary Heart Disease**: Etiology and Risk Factor; **Iron**: Properties and Determination; **Minerals – Dietary Importance**; **Vegetarian Diets**; **Yogurt**: Yogurt-based Products

Further Reading

Dwyer JT (1988) Health aspects of vegetarian diets. *American Journal of Clinical Nutrition* 48: 712–738.

Glanz K and Kolonel LN (1998) Culture, religion, diet and health: challenges and opportunities. *Nutrition* 14: 238–239.

Grivetti LE and Pangborn RM (1974) Origin of selected Old Testament dietary prohibitions. *Journal of the American Dietary Association* 65: 634–638.

James EO (1979) Cultural and religious taboos related to food. *Progress of Food Nutrition Science* 3: 67–77.

Jiggets JI (1949) *Religion, Diet, and Health of Jews*. New York: Bloch.

Johnston PK and Sabate J (1999) Third International Congress on Vegetarian Nutrition. Proceedings of a symposium held in Loma Linda, CA March 24–26, 1997. *American Journal of Clinical Nutrition* 70: 3(S).

Rashed A (1992) The fast of Ramadan. *British Medical Journal* 304: 521–522.

Register UD and Sonnenberg D (1973) The vegetarian diet. *Journal of the American Dietary Association* 62: 253–261.

Robson JRK (1980) *Food, Ecology and Culture. Readings in the Anthropology of Dietary Practices*. New York: Gordon and Breach.

Sakr AH (1975) Fasting in Islam. *Journal of the American Dietary Association* 67: 17–21.

Shatenstein B and Ghadirian P (1998) Influences on diet, health behaviours and their outcome in select ethnocultural and religious groups. *Nutrition* 14: 223–230.

RENAL FUNCTION AND DISORDERS

Contents

Kidney: Structure and Function

T T Rosa, University of Brasilia, Brasilia, Brazil

Background

The human kidneys are located in the lumbar area, as a pair in the retroperitoneal space, and are composed of about a million nephrons, which are the functional units. Each nephron contains a glomerulus which carries out blood filtration and there is tubular system, divided into 14 segments, which are different morphologically and functionally. The combined action of these structures produces urine. The glomerulus is located in the outer part of the kidneys, called the cortex, and the tubules are located in the inner part, called the medulla. These structures are connected by interstitial tissue and contained by a fibrous capsule, named Bowman's capsule, in a similar shape to a bean (Figure 1).

Within the kidney are located the renal hilum, as well as the renal pelvis, formed by the exit of the renal vein, which it drains from the vena cava; the renal artery, a branch of the abdominal aorta artery; the lymphatics; and a nerve plexus. In addition to these components are those of the urinary branches. These are the renal pelvis and ureter, which transports urine to the bladder, from where it exits via the urethra, completing the system of renal excretion (Figure 2).

Glomerular Ultrafiltration

The first stage of the process of urine formation is glomerular ultrafiltration. About 170 l of plasma is ultrafiltered every 24 h, representing more than 30 times volume. The resulting product, in the form of urine, corresponds to less than 1% of that volume. The difference is the result of the balance between the reabsorption process and the secretion process accomplished by the tubular system.

Marcello Malpighi first described the structure which is now called the glomerulus in 1662. As a result, the corpuscle was just known as corpuscles of Malpighi for a long time. Carl Ludwig, in the

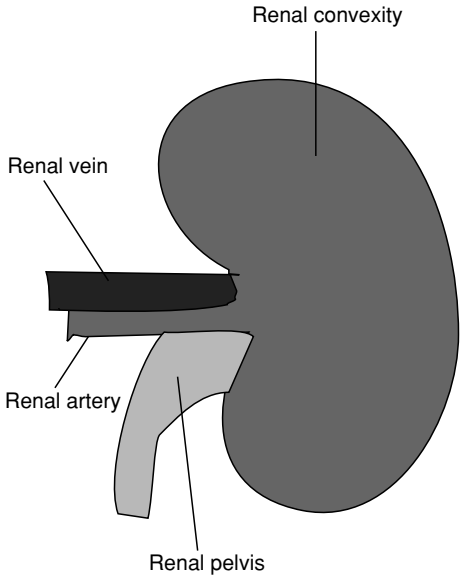

Figure 1 Macroscopic renal structure.

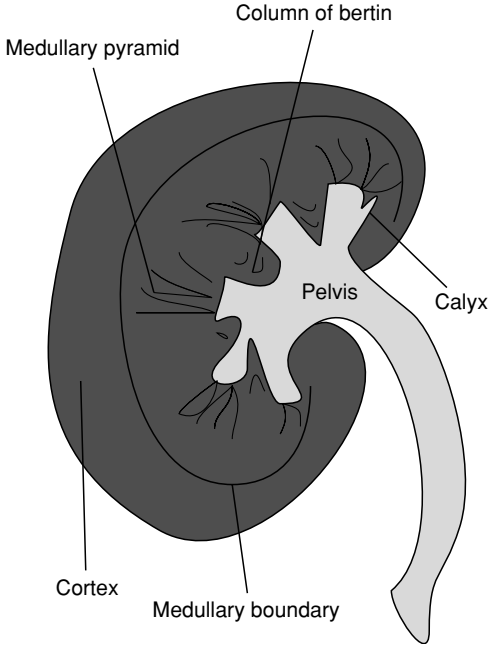

Figure 2 Renal hilum.

nineteenth century, described the ultrafiltration process as the passage of water, electrolytes, and small molecules, from the glomerular capillaries to Bowman's space. In 1924, this theory was proven: glucose, chloride, but no proteins were detected in the Bowman's space. All the energy for that process is created by the heart, acting at the level of the renal microcirculation, and this characterizes ultrafiltration as a physical phenomenon.

Glomerular filtration rate (GFR) in humans is approximately 120 ml min^{-1}. It is the result of ultrafiltration achieved by about 2 million nephrons.

As a physical phenomenon ultrafiltration results from a balance of pressure between the glomerular capillary and the internal space of Bowman's capsule. The glomerular tuft forms a conductive system in parallel, located between the afferent and the efferent arterioles (Figure 3). The pressures involved in this process are as follows:

$$\Delta P_H = P_{IG}.P_{CB} \quad (1)$$

where P_H represents the variation in hydrostatic pressure resulting from the difference between P_{IG}, the hydrostatic intraglomerular pressure, and P_{CB}, pressure inside the Bowman's capsule. It is similar to the physical principle behind a domestic water filter.

The concentration of proteins in the plasma also interferes with the process, as it adds to the filtrate a pressure that is opposed to the hydrostatic pressure, called oncotic pressure (P_O). That pressure represents a vector that addresses the filtered flow in an inverse direction, in other words, from Bowman's capsule to the glomerular capillary once the protein concentration is minimum in the Bowman's capsule, close to zero, because the wall of the glomerular capillary does not allow proteins to pass. In that way, the effective filtration pressure (P_F) is expressed in agreement with eqns (2) and (3):

$$P_F = (P_{IG}.P_{CB}) - P_O \quad (2)$$

or

$$P_F = \Delta P_H - P_O. \quad (3)$$

The renal plasma flow should be adapted to allow a balance of those pressure forces and the permeability of the glomerular wall should be unaffected if the process is to happen in a physiologic way. This last variable is represented by a variable called the coefficient of hydraulic conductance and represented by the symbol K_f. In comparison to the domestic filter, this coefficient measures the porosity of the filter. In the glomerulus it is defined by the product of the intrinsic permeability of the glomerular wall (K) for the total surface available for the filtration process (S), and is represented as follows:

$$K_f = K \times S. \quad (4)$$

The decrease in K_f determines a decrease in GFR, that it is not as pronounced as variations due to alterations in hydraulic and oncotic pressure.

To tune so many variables involved in glomerular ultrafiltration it is necessary for there to be self-regulation, so that variations in systemic blood pressure (SBP) do not interfere markedly in the process of urine formation that begins with ultrafiltration. If such a mechanism did not exist, dynamic variations in SBP would affect GFR.

Self-regulation is effected by an increase or decrease in resistance of the afferent and efferent arterioles defined by vasoconstriction or vasodilatation of these arterioles, if SBP falls or rises, respectively. This vascular movement is caused by three factors: vasomotor reflex, when the vein contracts its smooth muscle and reduces its diameter, when there is an increase in hydrostatic pressure within; metabolite production in response to the increase in blood flow in those arterioles, causing vasoconstriction; and the mechanism of glomerular tubule feedback.

In this last factor, the tubular structure is involved; this is called the macula densa and is located in the distal convoluted tubule. Alterations in the load of filtered sodium give rise to renin secretion; renin is produced in the cytoplasm of differentiated cells, and is present in the macula densa and in cells contiguous to the glomerulus, which comprise the justaglomerular device. Production of increased renin results in an increase in circulation of angiotensin II level, by a cascade of reactions shown in Figure 4. Angiotensin II stimulates local receptors in the afferent

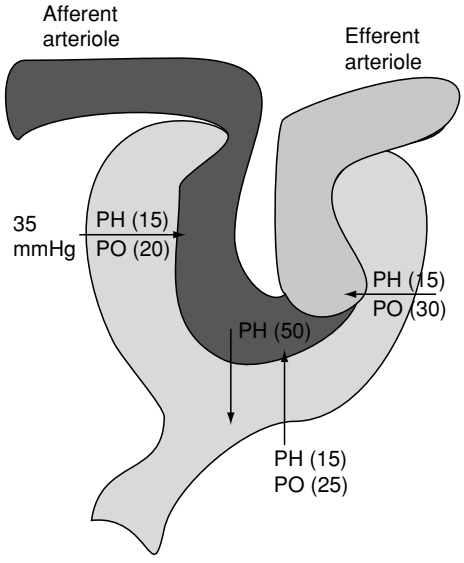

Figure 3 Renal microcirculation.

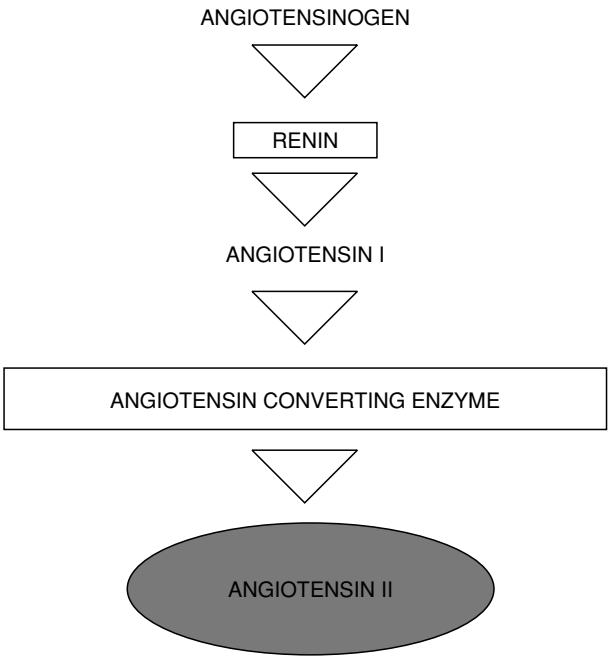

Figure 4 The renin–angiotensin system.

arterioles promoting vasoconstriction and a decrease in flow in the nephrons and a consequent decrease in glomerular filtration rate, balancing out the process of glomerular filtration associated with the secretion and tubular reabsorption, and resulting in the glomerular tubular balance.

To estimate GFR it is necessary to carry out clearance of a substance that is only filtered, and is not reabsorbed or secreted by tubules. Clinically, it is preferred for a substance to be produced endogenously, rather than having to administer an exogenous drug. Calculation of the endogenous creatinine clearance is commonly used. The clearance of any substance is always given by the product of its urinary concentration (U) and urinary flow (V), divided by its plasma concentration (P). Urinary flow is always calculated in ml min^{-1}.

In fact creatinine is not ideal for calculation of GFR, as it is partly secreted by the renal tubules, and appears in the urine as a result of filtration and a small amount of secretion, but results are very close to those obtained with inulin, an exogenous substance, which is totally filtered, and is used in experimental assays for the exact determination of GFR. Besides, the method of creatinine dosage is easy and cheap, which is a distinct advantage in clinical use.

In contrast, the simple creatinine level in plasma gives us an idea of how renal filtration works, because it is always in balance in the organism, as are almost all endogenous products. That means that its excretion should equal its production. Mathematically we have: creatinine production ($Prod_{Cr}$) = urinary excretion of creatinine (U_{Cr}) × urinary flow (V). As with the creatinine, the excreted load is the same as the filtered load, and can be written: $Prod_{Cr}$ = plasma concentration of creatinine (P_{Cr}) × GFR, that is: $P_{Cr} = Prod_{Cr} \times \text{GFR}$. As creatinine production is constant, its plasma concentration is an inverse hyperbolic function of GFR.

Most components in plasma are filtered by the kidney.

Tubular Reabsorption

Unlike glomerular filtration and tubular secretion, tubular reabsorption is not part of the process of renal excretion. In fact it is a renal mechanism that saves essential element, to the homeostasis of the internal environment, such as water, glucose, and amino acids. The transport of fluid from the inside of the tubule to the internal environment through the tubular wall is complex and involves several mechanisms. Some of these spend energy of the organism (adenosine triphosphate or ATP), while others are made passively, by gradients created among the compartments involved. Those gradients may be physical (differences in the electric charge or in the pH), chemical (differences in the concentrations of several substances) or osmotic (difference of concentration of solutes). There is also so-called facilitated transport, such as when glucose transport is facilitated by sodium transport. Those substances that need carriers are located in the membrane of the tubular cells which are linked to the substance to be transported in the border with the inner tubule. The carriers make transport through the cytoplasm to the border with the renal interstice, where the peritubular capillaries are located, liberating them to the internal environment of the organism.

Of the 170 l plasma that is filtered, and one more small rate of secretion, tubular reabsorption is about 1.5–2 l day^{-1} to the external environment, in the form of urine. The difference is returned to the internal environment by tubular reabsorption.

The main site responsible for reabsorption is the convoluted proximal tubule, which is the tubular portion that is located immediately after the glomerulus. These cells have quite a heterogeneous form and the borders contain microvilli that provide a larger contact surface with the fluid to be reabsorbed, optimizing the process. Despite this, reabsorption, as well as secretion, takes place along the whole renal tubular system.

The main substances reabsorbed in the renal tubules are amino acids, glucose, electrolytes, bicarbonate, and urea.

Tubular Secretion

Tubular secretion is the term given to the transport of substances from the medullary peritubular capillary to the inside of the tubule. It happens predominantly in the final part, the straight section, of the proximal convoluted tubule. It is also a method of renal excretion and can be calculated by the clearance of substances that are not filtered, or reabsorbed, but which appear in the urine, such as paraaminohippurate. This is a salt of paraaminohippuric acid and it is used to calculate renal plasma flow, after excretion. The transport mechanisms involved in secretion are very similar to reabsorption, especially as regards energy (ATP) expenditure.

The main substances secreted by the kidneys are urea, hydrogen, organic acids, and exogenous substance such as drugs and antibiotics.

Renal Metabolism of Water

Water makes up the largest part of the human organism, comprising about 80% of the total. The sensation of thirst originates in the hypothalamic center of thirst, which is sensitive to a variation in plasmatic osmolality of at least 1% and it is located in paraventricular and supraoptical nuclei. An increase in plasmatic osmolality stimulates the secretion of antidiuretic hormone (ADH) from the hypophysis as a result of those hypothalamic impulses. ADH acts, among other structures, at the level of the collector ducts, in the distal nephron, promoting the reabsorption of water and the consequent normalization of plasma osmolality.

In this way, the final urine under the action of ADH is more concentrated, it has larger osmolality. When there is an excess of water in the organism, an inverse mechanism results; there is inhibition of ADH and the urine produced has less osmolality. The first situation is more frequent, when maximum capacity of urinary concentration is reached, whose resultant osmolality can be up to 1200 mosm ol l^{-1}.

The mechanism by which nephrons concentrate urine is called countercurrent mechanism (similar to that described by thermodynamics) and it involves the proximal segments, the loop of Henle, and the distal segments of the renal tubules. It acts by an established osmolar gradient in the renal medulla and in inside the tuble, as a result of active transport of NaCl and the diffusion of H_2O through membranes with special permeabilities along the long loop nephrons, which are especially adapted for that function, located in the boundary between cortex and medulla. The result of that complex mechanism is a concentration of urine in such a way that the plasma osmolality is maintained constant and normal.

Endocrine Functions of the Kidneys

One of the most important hormones produced by the kidney is 1α-25-dihydroxycholicalciferol, the active form of vitamin D_3, and it acts directly in the renal metabolism of calcium and of phosphorus. A fall in plasma calcium stimulates the secretion of parathormone (PTH) in the parathyroid glands and so, they act in the renal medulla, promoting the synthesis of activated vitamin D_3. This acts in the small intestine, increasing the reabsorption of calcium in the bone tissue, increasing demineralization, and in the renal tubules, increasing reabsorption, at the same time as it promotes an increased excretion of phosphate, as phosphaturia.

A reduction in phosphate in the plasma stimulates vitamin production, directly activated by vitamin D_3, by the renal medulla and, in the absence of PTH renal tubular reabsorption of phosphate and greater excretion of calcium (calciuria) are promoted.

Thus the product of plasma concentration between calcium and phosphorus is maintained constant.

Action of Kidneys in the Acid–Base Balance

In addition to the important mechanism of acid–base control accomplished by the lungs, promoting the elimination of the carbon gas in its volatile form, the kidneys also participate in controlling the metabolism of bicarbonate (HCO_3) and of the ion H^+, which are respectively reabsorbed and secreted by the renal tubules.

This balance is fundamental to maintain organic biochemical reactions that keep constant blood pH, at around 7.4.

The transport of those two ions takes place in the proximal convoluted tubule and in the convoluted distal tubule. It happens in a connected way: one ion H^+ secreted for each reabsorbed HCO_3 molecule. In the proximal convoluted tubule, bicarbonate is regenerated from the inner tubule by two principal mechanisms: excretion of carbonic acid dissolved in the tubular fluid and secretion of fixed acids, such as sulfuric, hydrochloric, and phosphoric, among others. At the level of the distal convoluted tubule, there is reabsorption of HCO_3 starting from the cellular synthesis of ammonia with the secretion of the ammonium ion.

Renal Control of Blood Pressure

In the distal convoluted tubule is located the macula densa: it has longer cells and cytoplasmic granules. Close to it there are segments of afferent and efferent arterioles that have a greater amount of cytoplasmic

granules. These granules correspond to renin, the proteolytic enzyme that is synthesized there. This group of structures is termed the juxtaglomerular apparatus.

The main incentives for renin production are a reduction in blood pressure with a consequent fall in hydrostatic pressure of the afferent arteriole, and a reduction in renal plasma flow and variations in the concentration of sodium in the distal convoluted tubule.

Renin acts on a substratum called angiotensinogen, a duodecapeptide, of hepatic synthesis, transforming it to the decapeptide called angiotensin I. This undergoes action by another proteolytic enzyme, called angiotensin-converting enzyme (ACE), which is mainly synthesized in the capillary endothelium, creating the octapeptide angiotensin II, which is a substance which has a larger effect on the vasoconstrictor produced by the organism.

Receivers exist for angiotensin II in the afferent and efferent arterioles, altering GFR, when there is an increase in that peptide in the circulation. In the systemic circulation there is an increase in peripheral resistance and therefore also in blood pressure. There are also receivers in the adrenal cortex which when stimulated synthesize aldosterone, a hormone whose main action is to promote reabsorption of sodium in the distal convoluted tubule.

Thus it is easy to understand that arterial blood pressure is controlled by this renal mechanism and dysfunction of is system, called the renin–angiotensin–aldosterone system, will result in arterial hypertension.

In short, the renal structure is adapted to achieve vital functions for the human organism. The most difficult stage in the evolution of animal life on earth was considered to be the adaptation from an aqueous environment, the kidneys are essential to make life on earth viable, saving water and mineral salts and maintaining the necessary metabolic homeostasis for the hemodynamic adaptation and of the pH of blood.

See also: **Electrolytes**: Analysis; Water–Electrolyte Balance; Acid–Base Balance; **Renal Function and Disorders**: Nutritional Management of Renal Disorders; **Water**: Physiology

Further Reading

Goldman L (2000) *Cecil Textbook of Medicine* 21st edn. Philadelphia: WB Saunders.

Levine SA (ed.) (2000) *Brenner & Rector's The Kidney*, 6th edn. Philadelphia: WB Saunders & Co.

Miller RD (2000) *Anesthesia*, 5th edn. Philadelphia: Churchill Livingstone.

Nutritional Management of Renal Disorders

B E Engel, Royal Infirmary, Stoke-on-Trent, UK

Introduction

The nutritional state of the body is dependent on the intake of adequate amounts of nutrients in food and fluids and the ability of the body to utilize these substances and dispose of the excess and the waste products created from metabolic processes. One of the main functions of the kidney is the disposal of some of these waste products, in particular, the nitrogenous waste from the breakdown of protein as well as excretion of fluid, sodium, potassium, phosphate, hydrogen ions, and organic acids, in order to achieve homeostasis. The kidney also has a synthetic function converting 25-hydrocholecalciferol to 1,25-dihydroxycholecalciferol ($1,25(OH)_2D$ – the active form of vitamin D), and in producing the hormone erythropoietin (EPO), which is important in promoting the formation of red blood cells. In many types of kidney disease, these functions are impaired, resulting in the accumulation of toxic waste products in the blood, and the development of hypertension, bone disease, anemia, and hyperlipidemia.

The main aims of dietary management in kidney disease are therefore to:

1. Reduce the accumulation of waste products and thus prevent or ameliorate uremic toxicity and the adverse effects of alterations in various metabolic pathways.
2. Maintain or improve nutritional status (identifying and treating malnutrition at an early stage).
3. Maintain or improve the kidney patient's quality of life.

Finally, dietary interventions may help contribute to slowing down the progression of the disease.

The various kidney diseases that lead to end-stage renal failure (ESRF) show many similarities in their clinical and pathological features, and thus dietary management is not necessarily modified by the original cause of the disease. In this chapter, the main subjects discussed are: (1) chronic renal failure, (2) nephrotic syndrome, (3) ESRF requiring dialysis and transplantation, (4) acute renal failure, and (5) renal stone disease.

Chronic renal failure (CRF)

Giovanetti described CRF as a chronic reduction of glomerular filtration rate (GFR) below the normal

Table 1 Main causes of chronic renal failure

Glomerulonephritis	Hypertension
Chronic pyelonephritis	Polycystic kidney disease
Obstructive, reflux, analgesic nephropathies	Systemic disease: scleroderma, systemic lupus erythematosus
Diabetic nephropathy	Vasculitis
Renal transplant failure	Myeloma, amyloidosis

Table 2 Stage of renal failure related to creatinine level

Stage	*Plasma creatinine concentration (μmol l^{-1})*	*Glomerular filtration rate (ml min^{-1})*
Early	120–150	40–60
Moderate	150–300	20–40
Advanced	> 400	10–20
End-stage renal failure	> 600 – requires RRT	< 10

range (*c.* 120 ml min^{-1}) irrespective of the underlying cause (Table 1). The increase in plasma constituents such as creatinine and urea is used as a general marker to indicate the severity of the disease (Table 2). Reciprocal plots of creatinine against time are commonly used to chart the progression of CRF. The excretion of creatinine (formed from creatine and phosphocreatine in muscle tissue) occurs predominantly by renal filtration. A decrease in GFR below 50% of normal leads to a rise in plasma creatinine. However, as creatinine production is proportional to lean body mass and is also influenced by animal protein intake, a loss of lean body mass or decreasing appetite during the course of CRF can give the false impression that the rate of decline of renal function is slowing down. In addition, a plasma creatinine of 1000 µmol l^{-1} may indicate a GFR of < 10 ml min^{-1} in a 70-kg man, whereas a creatinine of 600 µmol l^{-1} would indicate the equivalent level of renal function in a person of lower body weight.

Protein Metabolism and Requirements

A well-planned and monitored reduction in protein intake in CRF patients will decrease the amount of nitrogenous waste, phosphorus, potassium, and acids that must be dealt with by the kidney. Urea is the main metabolite of protein, and in CRF, plasma urea increases as GFR decreases. A decrease in the urea:creatinine ratio after starting a reduced protein diet may be used as a measure of compliance with the diet. Other factors, however, may raise plasma urea further, e.g., a high dietary protein intake, drugs such as frusemide and steroids, gastrointestinal bleeding, and a hypercatabolic state leading to muscle catabolism. Urea itself is nontoxic, but its elevation correlates with the rise in plasma levels of more toxic compounds, e.g., guanidines, amines, phenols, and organic acids. Plasma urea levels are therefore used as a surrogate marker for these toxins.

Most of the research in this area has focused on the effect of a low-protein diet on slowing down the progression of renal disease. The diets have aimed to provide 0.6 g of protein per kilogram of ideal body weight (IBW), which is the minimum requirement advised by the World Health Organization. It is still recommended that at least 50% of the protein should be of high biological value (HBV) – essentially from meat, fish, and diary products in order to provide all essential amino acids, although vegan and vegetarian diets can maintain the nitrogen balance as long as proteins are obtained from a combination of low biological value (LBV) sources, such as grains, pulses, etc. The protein intake from food can be reduced further to 0.3 g kg^{-1} IBW if the diet is virtually vegetarian and supplemented with essential amino acids or their keto analogs. In practice, even with considerable support from dietitians, it seems difficult – certainly in Western society – to maintain this level of protein restriction. In the largest trial ever completed on this subject, the Modification of Diet in Renal Disease, patients' actual intake was 0.7–0.8 g protein per kilogram per day. At this level, overall nutritional status was maintained throughout the 2 years of the trial, although the energy intake was less than the recommended 35 calories per kilogram per day. A reduction of protein intake by 0.2 g kg^{-1} resulted in a 30% reduction in the rate of decline of GFR, i.e., a prolongation of the time to ESRF by 41% in patients with advanced renal failure. Some clinicians argue that this level of improvement is not significant enough to justify a diet that is difficult to achieve, requires a lot of support from dietitians as well as family, and may lead to nutrient deficiencies in the long term. There are none the less other benefits from reducing protein intake: the improved biochemical indices that result from this intervention will help reduce some of the classic symptoms of uremia that are related to the accumulation of protein metabolites and imbalances particularly in phosphorus, potassium and lipid metabolism (Table 3).

Energy Requirements

There is no evidence to suggest that requirements are affected by renal disease itself. Calculation of energy requirements should be made depending upon the need for weight loss, maintenance, or gain in the normal way, e.g., Schofield or Harris–Benedict equations or 30–35 kcal kg^{-1} IBW. As hyperlipidemia (increased triglycerides and/or cholesterol) and some degree of glucose intolerance are common, the source of calories should be from complex carbohydrates

and mono-polyunsaturated fats (with overall fat intake providing 30% of the energy) in line with advice for healthy eating.

Vitamins and Minerals

Some of the clinical features of uremia may be linked to subclinical vitamin and mineral deficiencies. Dietary protein, phosphate, and potassium restrictions will limit the intake of various micronutrients, and even when intake is adequate, the metabolic pathways may be affected, reducing the availability or efficacy of the nutrients. Specifically, a low-protein diet was found to contain low levels of B vitamins (B_1, B_2, B_6, B_{12}), iron, calcium, and zinc. A low-potassium diet will limit folic acid and vitamin C intake. However, there may be reduced renal losses of some nutrients, and cases of hypervitaminosis A have been recorded. It is recommended that vitamin and mineral status should be part of a renal patients regular nutritional assessment.

Phosphorus and Calcium

Hyperphosphatemia and hypocalcemia are characteristic of CRF. A decreasing GFR will lead to a reduced renal clearance of phosphate, even at the moderate stage of CRF, and eventually will lead to an increased plasma phosphate. Simultaneously low levels of active vitamin D cause a reduction of calcium absorption from the gut, hence lowering the plasma calcium. These two factors stimulate parathyroid hormone (PTH) release, which in turn causes bone resorption. The release of calcium from the bone is an attempt to normalize plasma calcium levels, unfortunately at the expense of the integrity of the skeletal system (and a further rise in plasma phosphate), resulting in osteodystrophy. A raised plasma calcium × phosphate product ($> 5\ \text{mmol}^2$) also results in calcification of soft tissues and has been linked with an increased risk of cerebrovascular disease (stroke) in these patients.

A low-phosphorus diet (*c.* 0.3–0.5 mmol per kilogram IBW per day) (Table 4), in combination with phosphate binders such as calcium carbonate, has been prescribed to maintain plasma phosphate within the normal range. A phosphorus intake at the lower recommended level can only be achieved with some degree of protein restriction, particularly the HBV proteins. This diet will also limit calcium intake, hence there is a dual role for calcium tablets: as a binder of phosphorus and to provide a calcium supplement. Prescription of 1,25-dihydroxy or 1-hydroxy cholecalciferol can help to suppress PTH production. Calcium levels must be monitored regularly, as, due to the medication, there is a risk of hypercalcemia from increased intestinal absorption. Plasma levels for phosphate and calcium should be maintained within the normal ranges. As yet, there is no consensus regarding the optimum level of renal failure at which to advise a reduction in phosphorus intake, i.e., it may be of benefit to reduce intake before plasma levels exceed the normal range, as PTH may already be increasing in order to maintain homeostasis.

Table 3 Uremic manifestations

System affected	*Uremic symptoms*
Cardiovascular	High blood pressure, fluid overload, pericarditis, cardiomyopathy
Gastrointestinal	Nausea, vomiting, gastritis
Peripheral nervous system	Neuropathy, myopathy
Central nervous system	Confusion, drowsiness, impaired higher mental functions, insomnia, coma
Blood	Anemia
Skeletal	Renal osteodystrophy, bone pain, fractures
Skin	Pruritis, pigmentation, calcification

Potassium

Hyperkalemia does not tend to develop until the GFR $< 10\ \text{ml min}^{-1}$ (i.e., ESRF). It can occur in moderate CRF under certain circumstances such as increased catabolism: infection, trauma, acidosis, or medication: angiotensin-converting enzyme (ACE) inhibitors, cyclosporin A, and EPO. A raised plasma potassium $> 6.5\ \text{mmol l}^{-1}$ can increase the risk of asystolic cardiac arrest. Hence, it is advised that levels should be maintained within the normal range using a combination of dietary restrictions of high-potassium foods and appropriate cooking methods such as

Table 4 Phosphate (mg):protein (g) ratio of some common foods

Phosphate:protein ratio < 10	*Beef, lamb, turkey, chicken, pork, cod, haddock, tuna, corned beef*
Phosphate:protein ratio 10–15	*Bacon, ham, meat paste, plaice, plaice, herring, mackerel, salmon, cottage cheese, white bread, white pasta*
Phosphate:protein ratio 15–20	*Liver, pilchards, mussels, prawns, soya milk, peanuts, pulses, white rice*
Phosphate:protein ratio > 20	*Milk, icecream, cheese, yogurt, pate, roe, sardines, whitebait, nuts, chocolate, wholemeal bread, brown rice, brown pasta, bran flakes, sponge, scones*

NB: Although wholemeal products are higher in phosphate, patients are still encouraged to have them as part of a healthy diet and to help bowel function.

boiling vegetables and discarding the water. It is not possible to state an absolute amount for the advised intake; 0.6–1 mmol kg^{-1} IBW will help to keep the plasma levels below 5.5 mmol l^{-1}. The use of potassium-losing diuretics and calcium resonium – a potassium binder – may be appropriate.

Sodium and Water

Kidney diseases such as diabetic nephropathy and some forms of glomerular nephritis are characterized by retention of sodium and water with resulting hypertension and edema. However, interstitial nephritis, e.g., chronic pyelonephritis, may initially present with a failure to concentrate urine, leading to sodium and water loss and potential dehydration. The nonpharmacological approaches that have been used to treat hypertension in the nonrenal population may be of some benefit in the renal population: dietary sodium restriction, weight control, and control of other risk factors for coronary heart disease. Sodium restriction also potentiates the action of ACE inhibitors, diuretics, and β blockers. As the kidney loses its ability to respond rapidly to changes in sodium intake, any restrictions should be implemented gradually. The dietary changes would include not adding salt to cooking or at the table and reduction of processed, tinned, or convenience foods. Compromises have to be made for the elderly patient who relies on these foods as dietary staples. To improve the palatability of the diet, a wide range of herbs and spices can be used. The recommended level of restriction is usually 80–100 mmol (1.8–2.3 g) of sodium per day.

Table 5 gives a summary of dietary advice for conservative management (prior to starting RRT).

Other Nutritional Concerns

Malnutrition A spontaneous decline in nutritional intake has been recorded in patients approaching ESRF. Unfortunately, loss of body weight is often masked by increasing levels of fluid, giving the impression that the weight is stable. A poor nutritional state when the patient enters the dialysis program will have a negative affect on morbidity and mortality, so it is important to monitor these patients as they approach ESRF irrespective of whether any dietary restrictions have been advised.

Anemia The anemia of CRF has many causes, including (1) decreased production of EPO, (2) poor iron, folate or B_{12} intake due to a low protein diet or a poor appetite, (3) poor iron absorption due to drug interactions, and (4) chronic blood loss due to disturbed clotting mechanisms. Anemia contributes to extreme fatigue, loss of appetite, and taste changes. Plasma levels of ferritin, hemoglobin, vitamin B_{12}, and red blood cell folate help determine the need for iron, folate and B_{12} supplements, and EPO injections.

Hyperlipidemia Hyperlipidemia (raised cholesterol and triglycerides) is a common complication of renal failure. There is evidence in animal models that it may increase the rate of progression of the disease, but this has not been shown in humans except in the progression of diabetic nephropathy. There is, however, a greatly increased risk of cerebro- and cardiovascular disease, which accounts for 60% of deaths in renal patients. HMG CoA (3 hydroxy, 3 methyl giutaryl coenzyme A) reductase inhibitors are being used to reduce cholesterol levels, and trials of fish oil supplements are promising. Although diet and drug interventions have been shown to reduce risk factors, including improving the lipid profile, as yet, there have been no studies to show that this actually results in decreased mortality in renal patients. Extrapolating from the nonrenal population, it seems wise to encourage the principles of healthy eating, weight loss if the body mass index is higher than 30, and regular exercise.

Table 5 CRF diet (conservative management prior to ESRF)

Nutrient (recommended daily intake)	*Rationale*
Protein 0.6–1.0 g kg^{-1} IBW	As described in the text. With the lower-protein diet, the patient is given a meal plan comprising HBV (60%) and LBV (40%) protein foods: to insure all the amino acid requirements are met
Energy 35 kcal kg^{-1} IBW or use Schofield equations, for example	Normal adult requirements. Calorie sources should not exacerbate hyperlipidemia, i.e., total fat 30% of energy, low saturated fat. Encourage good glycemic control in diabetics. NB: Monitor for signs of malnutrition
Phosphate 0.3–0.5 mmol kg^{-1} IBW	Plasma levels should be maintained between 0.8 and 1.5 mmol l^{-1}.
Sodium 80–100 mmol per day. Fluid – no restriction usually required	Salt restriction will help potentiate action of antihypertensives
Potassium 0.6–1 mmol kg^{-1}	Levels should be maintained between 3.5 and 5.5 mmol l^{-1}. Restriction may not be necessary until patient has reached ESRF
Vitamins and minerals	May be required with lower protein diets and if potassium restriction is advised

Nephrotic Syndrome (NS)

NS is characterized by proteinuria >3 g per day, hypoalbuminemia, edema, and hyperlipidemia. It is usually also associated with hypertension. NS appears as an additional complication in a number of types of kidney disease (Table 6). Treatment involves antihypertensive agents, immunosuppressants, diuretics, salt-poor albumin, and dietary modifications.

Protein

Historically, a high protein diet was advised on the grounds that protein losses in the urine needed to be replaced. However, studies comparing low (0.8 g kg^{-1}) and high protein (1.5 g kg^{-1}) diets showed that proteinuria was reduced with the lower protein diets, although not all studies showed a corresponding increase in plasma albumin levels. It has been shown that a high protein intake causes increased permeability and hyperfiltration in the basement membrance of the kidney glomeruli, and this may be the mechanism that explains the poor results of high protein diets in NS. In addition, phosphate, lipid and renin levels tended to improve on the lower protein diets. An intake of 0.8–1 g of protein per kilogram per day has been recommended.

Salt

The kidneys tend to retain sodium and water in NS, and patients benefit from a moderate dietary sodium restriction: 80–100 mmol per day, and if edema is present, a fluid restriction is also necessary. Sodium restriction can potentiate the antihypertensive and antiproteinuric effects of ACE inhibitors.

Fat

Both triglycerides and cholesterol are raised in NS. The incidence of myocardial infarction has been reported as 5.5 times greater in NS compared with normal. Lipid lowering agents are used in conjunction with standard lipid-lowering advice.

Micronutrients

Iron, copper, zinc, and vitamin D are also lost as a result of proteinuria. Vitamin D losses can result in derangements of calcium metabolism, and supplementation may be necessary.

Table 6 Renal diseases which may develop nephrotic syndrome

Glomerulonephritis
Diabetic nephropathy
Amyloidosis
Systemic lupus erythematosus
Vasculitis
Toxic glomerulopathy (gold, penicillamine)

Renal Replacement Therapy

Hemodialysis (HD) and continuous ambulatory peritoneal dialysis (CAPD)

The nutritional problems of renal patients do not stop once renal replacement therapy is started. Even with dialysis, it is not possible to maintain body fluid, potassium, and phosphate levels within the normal ranges without some degree of dietary restriction. On the positive side, some of the symptoms of uraemia such as nausea, taste changes and gastrointestinal disturbances improve. Usually, there is an increase in appetite within 1 month of starting dialysis.

Dialysis itself has a number of negative effects on the nutritional state (Tables 7 and 8). In addition, some global factors affect the patient's well-being:

- The procedure itself is time-consuming, and traveling to and from hospital for unit HD severely disrupts the patient's day, including meal times.
- It is important that the risk of infections from the procedures be minimized as the catabolic effects of line infections and peritonitis can have long-term consequences for the patient's nutritional status.
- Underdialysis – as measured by urea and creatinine clearance – leads to impaired nutritional intake. There are now set standards that describe the amount of dialysis required in order to minimize morbidity and mortality.
- Depression can affect the patient for a number of reasons: inability to keep full-time employment, financial problems, body-image problems – as the treatment can be disfiguring, restrictions on traveling, dietary restrictions, and physical weakness due to the long-term metabolic effects on bone and muscle metabolism.
- It is important to maintain normal hemoglobin levels as anemia causes decreased appetite and poor physical functioning. Injections of EPO and iron are normally required to achieve this.

A team of professionals is necessary to treat the renal patient and help them achieve the best quality of life. In addition to medical and nursing staff, this should include a dietitian, social worker, physiotherapist, and counselor.

It is estimated that 30 – 70% of dialysis patients fall into the moderate- to high-risk categories for malnutrition. Protocols for routinely assessing the nutritional status of each patient must be in place in order to identify problems early. Subjective global assessment has been validated for use in dialysis

Table 7 HD diet

Nutrient (recommended daily intake)	*Rationale*
Protein 1.1–1.3 g kg^{-1} IBW	From nitrogen balance studies (although numbers were small and may not be applicable to all ethnic groups). 8–12 g of amino acids lost per dialysis. Possible poor intake on dialysis days
Energy 35 kcal kg^{-1} IBW or use Schofield equations.	Normal adult requirements. 25 g of glucose lost if glucose free dialysate is used
Phosphate 0.5 mmol kg^{-1} IBW	Plasma levels should be maintained between 1.2 and 1.7 mmol l^{-1}. Phosphate is not well dialyzed, so restriction is necessary, aided by phosphate binders
Sodium 80–100 mmol per day. Fluid 500 ml + previous day's urine output	Urine output decreases once dialysis starts. Salt restriction will help prevent thirst. Maximum of 2 kg interdialytic weight gain recommended. Fluid overload leads to left ventricular heart disease and hypertension
Potassium 1 mmol kg^{-1}	Plasma levels should be maintained between 3.5 and 6.5 mmol l^{-1}

Table 8 CAPD diet

Nutrient (recommended daily intake)	*Rationale*
Protein 1.2–1.5[a] g kg^{-1} IBW	From nitrogen balance studies (although numbers were small, and nitrogen balance was achieved at lower levels in some individuals). 9 g of protein/amino acids lost per day.
Energy 25–30 kcal kg^{-1} IBW + energy from dialysate or use, e.g., Schofield equations	Normal adult requirements. 70% of glucose is absorbed from the 'bag' of dialysate fluid. This can provide 300–1000 calories per day, depending on the glucose concentration in the bag. This leads to hypertriglyceridemia, obesity, and poor glycemic control in diabetics
Phosphate 0.5 mmol kg^{-1} IBW	Plasma levels should be maintained between 1.1 and 1.6 mmol l^{-1}. Phosphate is not well dialyzed, so restriction is necessary, aided by phosphate binders
Sodium 80–100 mmol per day. Fluid 500 ml + previous day's urine output. Additional fluid may be taken if the patient ultrafiltrates (removes fluid during dialysis) well	Urine output decreases once dialysis starts. Salt restriction will help prevent thirst. Overdrinking leads to overuse of the higher sugar containing bags of dialysis fluid and the problems described above
Potassium ~1 mmol kg^{-1} IBW	Levels should be maintained between 3.5 and 5.5 mmol l^{-1}. As dialysis is continuous, hyperkalemia is less of a risk than on HD. The diet can be more varied with respect to high-potassium foods
Vitamins (HD and CAPD)	There is no consensus on the routine use of vitamins. Folate, pyridoxine, and vitamin C levels are reported to be low without supplementation. Folic acid supplements may also reduce homocysteine (risk factor for CVD). Fat-soluble vitamins are not supplemented, except vitamin D, as discussed earlier.

[a]Larger protein requirements applicable during peritonitis.

patients; this combines medical history, including dietary assessment, with a physical examination of fat and muscle stores. There is a wide range of nutritional supplements available for the first-line treatment of malnutrition, and for the more severely malnourished patient, gastrostomy feeding (HD and CAPD) and intradialytic parenteral nutrition (IDPN) in HD patients have been used successfully. An amino acid containing CAPD solution is available and may be of use in certain cases of hypoalbuminemia caused by a persistent low protein intake.

The dialysis diet should be tailored to the individual patient taking into consideration their medical treatment and psychosocial factors. It is also guided by regular monitoring of biochemical profiles that usually occur monthly for HD and every 2 months for CAPD patients. For the well-nourished patient, healthy eating principles should continue to be encouraged within the constraints of the potassium restriction.

Transplantation

The positive effects of transplantation are to relieve the patient from dietary and other lifestyle restrictions imposed by dialysis, and the patient experiences a general improvement in well-being. This in itself can lead to large gains in body weight as the patient eats more in general. In addition, the immunosuppressive therapy can exacerbate the long-term metabolic effects of CRF such as bone disease, muscle

wasting, and cardiovascular disease. The main aims of dietary therapy are to reduce the risk factors for cardiovascular disease, which is the cause of 60% of deaths in transplant patients. All the aspects of 'Healthy Eating' should be encouraged. Patients may be reluctant at first to increase their intake of fruits and vegetables and even oily fish, as these would have been restricted on the dialysis diet. Intensive dietary counseling prior to discharge with regular follow up in outpatient clinics has been shown to reduce the average weight gain from 12 to 6 kg in the first year. Lipid levels can also be improved with advice on appropriate fat and sugar intake. An increasing amount of literature is showing that encouraging exercise at all stages of renal disease helps to improve physical mobility, and extrapolating from the nonrenal population may further reduce risk factors for CVD.

Acute Renal Failure (ARF)

ARF can be defined as an abrupt decline in renal function. Initially, there is oliguria (< 400 ml per day) or anuria, retention of the end products of protein metabolism, acidosis, and electrolyte imbalance. Biochemical profiles show a rapid increase in urea and creatinine, and also potassium and phosphate levels can increase. ARF is usually reversible, and as renal function returns, there may be polyuria to an extent where potassium and phosphate levels can drop below normal, and salt and water depletion can occur. As the causes of ARF are so varied, the effect on nutritional requirements may be negligible or, in the case of severe trauma or sepsis, cause a hypercatabolic state. Protein and energy requirements are calculated using standard formulae and stress factors and are related to the underlying causes. ARF in itself has specific effects on protein, fat, and carbohydrate metabolism, but the literature is less clear about how this impacts on the final nutrient prescription. ARF can be divided into three categories:

1. Noncatabolic: nontraumatic causes such as obstruction and interstitial nephritis from drugs.
2. Catabolic: postsurgery, rhabdomyolysis, and hemolytic uremic syndrome.
3. Hypercatabolic: major trauma (road traffic accident), burns, and sepsis.

Noncatabolic ARF

As in CRF, the aims of treatment are to slow down the build-up of uremic toxins and fluid overload in order to prevent or alleviate symptoms, to maintain good nutritional status, and to prevent weight loss. Dialysis may or may not be necessary. Nausea and lack of appetite are common features, and the main concern is insuring the patient is meeting their nutritional requirements sometimes with the help of dietary supplements. Restriction of dietary protein < 0.8 g protein per kilogram is not recommended. It is usually acceptable to aim for a 'normal' protein intake of 1 g per kilogram IBW per day. Advice on salt intake – aiming for 80–100 mmol per day, fluid restriction – 500 ml plus PDUO (previous days urine output) and potassium restriction < 1 mmol kg^{-1} may be necessary. In practice, all that is usually required is a simple diet sheet advising the patient and their relatives on suitable snacks and drinks. The usual 'treats' that people receive in hospital, i.e., fruit juice, fruit, and chocolate, may not be acceptable. High plasma phosphate levels can be controlled by phosphate binders.

Catabolic and Hypercatabolic ARF

The patient will almost certainly require dialysis treatment as the rate of production of waste products will otherwise cause severe and potentially fatal uremia. The type of dialysis will also affect nutrient requirements and the ability to provide them through nutritional support. Intermittent HD may be sufficient in noncatabolic or moderately catabolic patients, but as the severity of catabolism increases, continuous dialysis may be required, e.g., continuous arteriovenous HD (CAVHD). Losses of 9–13 g of protein per day can occur with this therapy. With 1.5% glucose in the dialysate solution, there is a net gain of glucose of 6 g h^{-1}, which represents an energy intake from dialysis of 550 kcal per day. The effects of dialysis on nutrient losses and gains should be taken into consideration when prescribing the feeding regimen.

Protein requirements Catabolic patients require in the region of 9–14 g of nitrogen (56–87 g of protein) per day. Hypercatabolic patients may break down up to 40 g of nitrogen (235 g of protein) per day, but the liver cannot deaminate more than 20 g of nitrogen (117 g of protein) per day, so this is the maximum that can be replaced by nutritional support.

Carbohydrate Glucose utilization is limited by ARF, and excessive amounts can cause lipogenesis, fatty liver, and increased carbon dioxide production. Provision of carbohydrate should not exceed the maximum oxidation rate of 4 mg per kilogram IBW per minute.

Fat As the metabolism of fat is impaired in ARF, an upper limit of 1 g of fat per kilogram per day is recommended.

Fluid and electrolytes With intermittent HD, a feeding regimen containing reduced fluid, sodium, and potassium may be necessary and may limit the macronutrient provision from an enteral or parenteral feed. Continuous dialysis and ultrafiltration obviate the need for a fluid (and usually also electrolyte) restriction.

Phosphate Hyperphosphatemia is a common feature of ARF, and initially, TPN (total parenteral nutrition) regimens can be phosphate-free. Anabolism and dialysis will lower serum phosphate, and the requirements would be reassessed at this stage.

Vitamins and minerals Requirements for micronutrients are not well documented. Water-soluble vitamins tend to be given daily in the standard amounts and trace elements given only once or twice a week. Fat-soluble vitamins are not required in the short term.

Renal Stone Disease (Nephrolithiasis)

Renal stone formation occurs as a result of increased urinary concentration of promotor substances and decreased concentration of inhibitor substances (Table 9). This is partly influenced by dietary factors but can also be linked to congenital abnormalities, as seen in polycystic kidneys, horseshoe kidneys, and medullary sponge kidney or to short bowel syndromes such as Crohn's disease. The stones vary in composition, but the main types include combinations of calcium, oxalate, and phosphate. The size of the stones can vary from microscopic to the large 'staghorn' calculi, and can lead to kidney failure if the kidney or urinary tract becomes obstructed.

A number of dietary factors can contribute to the management of stone disease.

Fluid

An adequate fluid intake of 2–3 l a day is encouraged to ensure a minimum urine output of 2 l per day. Specifically, patients are told to drink 250 ml every four waking hours, with 250 ml at meals. Compliance with this single measure has been shown to reduce new stone formation by diluting the concentration of stone-forming salts, and may prevent new stone formation in up to 60% of patients with ideopathic calcium urolithiasis.

Table 9 Risk factors for renal stones

Low urine volume particularly < 1000 ml per day
High urinary concentrations of:
Calcium
Oxalate
Low urinary concentrations of:
Citrate
Urine acidity

Calcium

Forty to 50% of people who develop stones have a higher level of calcium in the urine (daily excretion $>4\,mg\,kg^{-1}$, although absolute amounts are also relevant).

There are a number of dietary factors that can cause hypercalcuria:

- excessive intake of calcium or vitamin D particularly as supplements;
- high salt intake; and
- high protein intake, particularly from animal protein.

The ideal level of calcium intake has not been ascertained – what is known is that a low dietary intake of calcium will exacerbate intestinal absorption of oxalate and promote stone formation. In a number of patients, a low calcium diet will also escalate an underlying tendency to osteoporosis. Intake of calcium supplements, however, is associated with an increased risk of stone formation, and therefore, on balance it is advisable to aim for the normal reference nutrient intake for calcium (700–800 mg per day) from diet, not supplements. There is also evidence showing that calcium excretion increases with sodium intake and protein intake. A moderate intake of no more than 100 mmol of sodium per day is recommended, and 1 g of protein per kilogram of body weight is thought to be appropriate.

Oxalate

Calcium oxalate stones account for up to 75% of stones formed. Oxalate is mostly formed endogenously; only 10% comes from the diet, although the proportional change in excretion is quite significant if calcium intake is decreased or oxalate intake increases. Therefore, it still helps to avoid or reduce intake of high oxalate foods such as tea, rhubarb, beetroot, chocolate, cocoa, nuts, spinach, and strawberries. American sources also name gooseberries, sweet potato, leek, parsley, okra, and wheat bran. Unfortunately, there are no comprehensive lists of oxalate-containing foods and their bioavailability. High vitamin C intake increases oxalate production, although the doses stated are in excess of 1 g per day.

Other dietary factors that may have a role are dietary fiber and phytates, which reduce the amount of calcium absorbed by the gut, but this may then increase the amount of oxalate absorbed.

Uric Acid

Uric acid excretion is involved in calcium and also uric acid stone formation. Uric acid is an end product of purine metabolism, and dietary restriction of purines and animal protein will reduce its urinary excretion. Purine-rich foods include liver, kidney, brain, anchovies, sardines, herring, mackerel, mussels, scallops, goose, partridge, and yeast.

In summary, stone formation is multifactorial, so in order to give appropriate dietary advice, it is necessary to have an analysis of the stone composition and urinary concentrations of the risk factors (Table 9). It is then necessary to take a detailed diet history in order to identify excesses or deficiencies or general trends in dietary intake.

See also: **Calcium**: Physiology; **Energy**: Intake and Energy Requirements; **Hyperlipidemia (Hyperlipidaemia)**; **Malnutrition**: The Problem of Malnutrition; **Phosphorus**: Physiology; **Potassium**: Physiology; **Protein**: Requirements; **Sodium**: Properties and Determination; **Vitamins**: Overview; **Water**: Physiology

Further Reading

Brown MA and Whitworth JA (1992) Hypertension in human renal disease. *Journal of Hypertension* 10: 701–712.

Engel B *et al*. (1998) *Setting Standards and Achieving Optimal Nutritional Status*. Birmingham, UK: British Dietetic Association.

Giovannetti S (1989) *Nutritional Treatment of Chronic Renal Failure*. Dordrecht: Kluwer Academic.

Greenberg A (1994) *Primer on Kidney Diseases*. London: Academic Press.

K/DOQI (2000) Clinical practice guidelines for nutrition in chronic kidney failure. *American Journal of Kidney Diseases* 35(6) (supplement 2): s1–s137.

Lazarus JM (1993) Nutrition in haemodialysis patients. *American Journal of Kidney Diseases* 21(1): 99–105.

Levey AS (1998) *Controlling the Epidemic of CVD in CRD*. USA: National Kidney Foundation.

Levey AS, Adler S, Gaggiula AW *et al*. (1996) Effects of dietary protein restriction on the progression of advanced renal disease in the modification of diet in renal disease study. *American Journal of Kidney Diseases* 27(5): 652–663.

McCann L (1996) Subjective global assessment as it pertains to the nutritional status of dialysis patients. *Dialysis and Transplantation* 25(4): 190–225.

Massy ZA, Jennie Z *et al*. (1995) Lipid lowering therapy in patients with renal disease. *Kidney International* 48: 188–198.

Mitch W and Klahr S (1998) *Handbook of Nutrition and the Kidney*, 3rd edn. Philadelphia, Pennsylvania, USA: Lippincott-Raven.

RCP and Renal Association (1997) *Treatment of Adult Patients with Renal Failure: Recommended Standards and Audit Measures* 2nd edn. UK: Royal College of Physicians and Renal Association, Lavenham Press.

Smith CL, Davis M and Berkseth RO (1992) Dietary factors in calcium nephrolithiasis. *Journal of Renal Nutrition* 2(4): 146–153.

Toigo G, Aparicio M, Altman PO *et al*. (2000) Expert Working Group Report on nutrition in adult patients with renal insufficiency. *Clinical Nutrition* 19(3): 197–207.

Rennin *See* **Cheeses**: Types of Cheese; Starter Cultures Employed in Cheese-making; Chemistry and Microbiology of Maturation; Manufacture of Extra-hard Cheeses; Manufacture of Hard and Semi-hard Varieties of Cheese; Cheeses with 'Eyes'; Soft and Special Varieties; White Brined Varieties; Quarg and Fromage Frais; Processed Cheese; Dietary Importance; Mold-ripened Cheeses: Stilton and Related Varieties; Surface Mold-ripened Cheese Varieties

Residue Determination *See* **Antibiotics and Drugs**: Uses in Food Production; Residue Determination; **Contamination of Food**; **Fumigants**; **Pesticides and Herbicides**: Types of Pesticide; Types, Uses, and Determination of Herbicides; Residue Determination; Toxicology

Resistant Starch *See* **Starch**: Structure, Properties, and Determination; Sources and Processing; Functional Properties; Modified Starches; Resistant Starch

RETINOL

Contents
Properties and Determination
Physiology

Properties and Determination

D C Woollard, AgriQuality NZ, Auckland, New Zealand
H E Indyk, Anchor Products, NZMP, Waitoa, New Zealand

Physical Properties

Retinol (9,13-dimethyl-7-[1,1,5-trimethyl-6-cyclohexene-5-yl]-7,9,11,13-nonatetraene-15-ol) is a pale yellow crystalline powder or oily mass, depending on purity. It is insoluble in water but readily miscible with most organic solvents. Its formula, formula weight, and certain other characteristic physical properties are summarized in Table 1, along with those of the two commercially significant ester derivatives.

Chemical Properties

All-*trans* retinol, considered the parent compound of the vitamin A group, is a complex unsaturated alcohol. Its trivial name originated with the historical recognition of its role in vision. This generic structure, illustrated in Figure 1, shares a common β-ionone ring, with attached conjugated isoprenoid side chain.

Chemical and structural modifications to the ring, side chain, or polar functional group generate many retinoids possessing a wide spectrum of properties. In nature only relatively few of these compounds exhibit significant vitamin A activity. These include the predominant all-*trans* forms of retinol (originally designated vitamin A_1) and its esters, retinal, retinoic acid, and their associated 9-, 11-, and 13-*cis* isomers. The *cis*-isomers are selectively interconvertible with the *trans*-forms in the body and thereby express fractional biological activities of approximately 0.25, 0.50, and 0.75, respectively, relative to the all-*trans* vitamer. Some marine and fresh-water fish contain significant quantities of the cyclic diene 3-dehydroretinol (originally designated vitamin A_2), which possesses approximately 40% of the bioactivity of all-*trans* retinol.

Vitamin A activity is commonly expressed in international units (1 IU is equivalent to 0.300 μg all-*trans* retinol, 0.344 μg retinyl acetate, or 0.549 μg retinyl palmitate), or, more recently, retinol equivalents (1 RE is equivalent to 1.0 μg all-*trans* retinol). Such units have been useful in nutritional studies where contributions from numerous vitamin A active congeners may be combined to yield a single value. For the food scientist, it is probably more appropriate to express vitamin A content as the summation of each retinoid, in absolute mass units (usually μg).

The dominant structural feature of retinol is the extensive conjugated double-bond system, to which many of the physicochemical and biological properties may be attributed. It is also the principal factor in the lability of vitamin A. Thus, retinol and its derivatives are particularly sensitive to oxidizing conditions and are rapidly destroyed by heat, light,

Table 1 Physical properties of retinol (all-*trans*) and its esters

Property	*Retinol*	*Retinyl acetate*	*Retinyl palmitate*
Formula	$C_{20}H_{30}O$	$C_{22}H_{32}O_2$	$C_{36}H_{60}O_2$
Formula weight	286.46	328.50	524.88
Melting point (°C)	63–64	57–59	28–29
λ_{max} (nm)[a]	325	326	326
Extinction coefficient, $E^{1\%}_{1cm}$	1820	1530	960
Molar absorptivity, ε	52 140	50 260	50 390
Fluorescence			
Excitation max. (nm)	325	325	325
Emission max. (nm)	470	470	470

[a]In 2-propanol. Spectral properties vary slightly between protic and aprotic solvents.

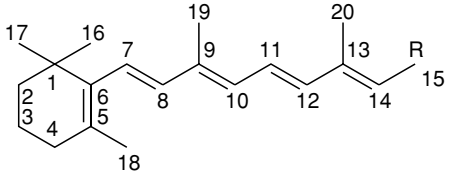

Figure 1 Structures of the basic retinoids. The *cis*-isomer positional variants are indicated by the side-chain numbering system.
R=CH_2OH all-trans retinol
CH_2OCOCH_3 all-trans retinyl acetate
$CH_2OCO(CH_2)_{14}CH_3$ all-trans retinyl palmitate
CHO all-trans retinal
COOH all-trans retinoic acid

and acids when in solution. Isomerization, oxidation and, ultimately, molecular cleavage may all occur concurrently, depending on the extent of environmental stress. Cleavage may result in the formation of volatile β-ionone fragmentation products which have importance in the development of off-flavor in some foods. Retinol is, however, relatively resistant towards alkali, and most degradation processes are minimized if it is maintained under inert gas at low temperature and in the absence of short-wavelength light.

In solution, several fat-soluble antioxidants offer protection, while in foods and food extracts, accompanying lipids and endogenous antioxidants such as phospholipids, tocopherols, and carotenes act to stabilize retinol until they are sacrificially depleted. Both thermal processing and storage expose vitamin A in foods to the risks of considerable loss (5–40%), although published data are often highly variable. (*See* **Antioxidants**: Natural Antioxidants.)

Retinyl esters are considerably more stable than the parent alcohol, a feature which is exploited in the principal commercial formulations available to the food industry. Although minor differences in absorption efficiency have been reported, dietary retinyl esters are considered to exhibit identical biological activities as a result of their conversion to retinol in the intestinal wall.

Occurrence and Forms in Foods

All forms of vitamin A found in foods ultimately derive from the provitamin A carotenoids, which are ubiquitous in the higher plants and lower animal organisms. Humans obtain preformed vitamin exclusively from animal sources, while carotenoids are gained from foods of both plant and animal origin.

Retinol and its derivatives are not widely distributed in foods. Fish liver oils are by far the most concentrated natural source, while animal liver, milk (and dairy products), and eggs contain significant quantities. The average western diet is generally assessed as satisfying the recommended daily intake level of 1000 RE total vitamin A, with about 25% supplied by β-carotene. (*See* **Carotenoids**: Occurrence, Properties, and Determination; Physiology.)

Retinoids in foods occur mainly as mixtures of retinyl esters, with lesser contributions from retinol itself. The predominant forms are long-chain esters, notably palmitate, with contributions from stearate and oleate. However, eggs are an exception, where unesterified retinol is the major form. Certain foods also contain contributions from retinal (eggs and fish roe) and *cis*-isomers (mainly 13-*cis*), the latter particularly in foods subjected to processing.

Table 2 Preformed vitamin A content of selected foods, expressed as retinol

Food	*Retinol content (range) of edible portion ($\mu g\,100\,g^{-1}$)*[a]
Milk	32–45
Butter	800–1000
Eggs	140–250
Beef	2–5
Liver (lamb)	7000–10 000
Mackerel	25–50
Cod liver oil	15 000–30 000

[a]To convert to IU per 100 g, multiply by 3.33.

Vitamin A deficiency is one of the prevalent diet-related issues and has received global attention. Consequently, nutritional tables are replete with information regarding retinol distribution in foods. Table 2 lists the vitamin A content of a few representative foods. Carotenoid-rich vegetable sources are not included. Refer to individual foods.

Use in Food Fortification

Increasing the vitamin A intake amongst populations at risk of retinol deficiency is a simple expediency and has been employed for several decades, both prophylactically and therapeutically. Many of the specific symptoms of severe deficiency (e.g., xerophthalmia, keratomalacia) are usually reversible, provided other nutritional criteria are similarly satisfied. In the developed nations, potential risks of subclinical deficiency are also avoided through food enrichment, thereby insuring satisfactory intake. (*See* **Food Fortification**.)

Fat-based foods such as margarine, milk, and infant formula provide an ideal carrier in which to supply supplemental vitamin A. Dried milk, by reason of cost and convenience, is globally the preferred foodstuff and additionally offers a high-quality protein and mineral enrichment medium appropriate to nutritional rehabilitation programs. An increasing range of alternative food items are being used as carriers, ranging from specialized dietary beverages to breakfast cereals.

In most cases, supplementation is performed with synthetic all-*trans* retinyl acetate or palmitate. The esters in the case of margarine are usually added directly in an oil carrier, often in the presence of appropriate antioxidants. Where additional stability is required, or when dried food products are involved, it is common practice to use a powder preparation in which vitamin A (and stabilizer) is deposited in a suitable carrier, such as gelatin carbohydrate, although alternatives (e.g., acacia gum) have recently been developed. However, uneven vitamin

distribution is a common problem when dry-blending food products.

Water-miscible vitamin A powders have been developed containing emulsifiers that facilitate reconstitution into aqueous solution. These solutions can then be processed directly into fluids such as milk prior to drying. Using the wet-blending technique, the protective environment of the additive is hopefully substituted by intimate interaction with the lipid phase of the finished food. The principal advantage of this supplementation route is an enhanced vitamin homogeneity in the finished product. However, the added ingredients are generally found to be less stable than their endogenous counterparts. (*See* **Emulsifiers: Uses in Processed Foods.**)

During food storage there are significant advantages in using packaging which provides effective oxygen and light protection. The use of cans allows nitrogen purging of the head space, thereby greatly enhancing the oxidative stability of the food and extending its shelf-life, particularly in developing countries with extreme climates.

Extraction and Clean-up

During analytical procedures, precautions are mandatory to exclude exposure of vitamin A to ultraviolet radiation, thermal, and oxidative stresses. Protective antioxidants and exclusion of air are essential to the success of the analysis. Purity of vulnerable retinoid standards is important during quantitative investigations, and spectrophotometric procedures are generally necessary to ascertain accurately their purity concentration.

Alkaline digestion (saponification) is most commonly employed as the first stage during analysis of retinol in foods. A representative sample is homogenized and digested in ethanolic potassium hydroxide or similar media. Saponification has the threefold effect of eliminating the bulk of lipid materials, releasing the vitamin from within the sample, and converting the various esters into free retinol. The procedure can be performed under reflux temperatures, or at ambient temperature (for longer periods), the details of which are determined by the sample matrix. In some food samples where the vitamin A concentration may be low, a preliminary fat extraction step will assist in achieving the required assay sensitivity.

Retinol is partitioned from the digest into a binary organic solvent, commonly consisting of hexane and a polar modifier such as an ether. The solution is washed with water, dried, and usually concentrated by evaporation, to provide a crude extract. Direct spectrophotometric or fluorometric assay of retinol may be impractical in poorly characterized foods, owing to interfering coextractives, and further clean-up is often necessary. Traditionally this has been achieved by open-column or thin-layer chromatography, using silica or alumina stationary phases, while gel-permeation chromatography has also been advocated. The laborious nature of these clean-up techniques has largely been overcome through the online automated use of prepacked, solid-phase extraction columns.

As the majority of native and supplementary retinol in foods exists in esterified form, there are analytical advantages to be gained through avoiding preliminary saponification, providing the nature of the food allows such simplified methodology. The total lipid fraction of such foods may therefore be directly extracted into organic solvent, which is then ready for analysis of the intact esters. Advantages include a reduction in the number of manipulative steps and the decreased risk of analyte loss through retinol degradation.

Chromatographic and Other Methods of Determination

The detection of retinol is based on its extensive conjugated double-bond functionality, which provides an intense and relatively specific chromophore in the long-wavelength ultraviolet region. The sensitivity and selectivity of spectral analysis are consequently enhanced, irrespective of the separation techniques used. In addition, retinol exhibits a strong fluorescence which offers selectivity advantages and often reduces the need for preliminary clean-up. In situations involving well-characterized foods, and where discrimination between retinol and related compounds is unnecessary, spectrophotometric or fluorometric measurement of the crude extracts can often supply a reliable estimate, providing the limitations of such strategies are recognized. (*See* **Spectroscopy: Fluorescence.**)

The traditional Carr–Price method for determination of retinol in foods has exploited the reaction with antimony trichloride to produce a blue complex in direct proportion to vitamin A concentration. Although the color is transient and difficult to control, the reaction is still employed in laboratories with limited access to modern chromatographic instrumentation. Other Lewis acids, notably trifluoroacetic acid, will react in a similar way and offer some manipulative advantages over the Carr–Price approach.

It is now accepted that a reliable estimate of vitamin A content in foods is best achieved through exploiting chromatography to separate coextracted substances and, in particular, to differentiate the active retinoid species. Most variants of chromatography have been

used during attempts to separate and quantify the retinoids. Thin-layer chromatography (TLC) and low-pressure liquid chromatography, while reasonably successful, have traditionally lacked the resolving power for reliable quantitative measurements, although high-performance TLC shows potential. Gas-liquid chromatography has not been widely utilized, as retinol (and its esters) degrades at elevated temperatures. The use of high-performance liquid chromatography (HPLC) has dominated vitamin A measurement over the last 20 years and has been largely responsible for the rapid proliferation of knowledge regarding this vitamin. This nondestructive technique facilitates a rapid analysis with minimum sample preparation. Recently, variants of micellar capillary electrophoresis (CE) have been applied to fat-soluble vitamin separations. Although representing an alternative to HPLC techniques, the application of CE to the hydrophobic vitamins is currently in its infancy. (*See* **Chromatography**: High-performance Liquid Chromatography; Thin-layer Chromatography; Gas Chromatography.)

Since isomerization can be a common phenomenon during food production, *cis* and *trans* differentiation may be required, with the application of appropriate factors to account for their selective bioactivities. In addition, inactive oxidation products such as oxyretinoids, epoxyretinoids, and their fragmentation products, as well as various retroretinoids, can exist in many food samples. Published literature is available for the determination of retinol isomers and esters, 3-dehydroretinol, retinal, and retinoic acid in a wide variety of foods. These analyses are accomplished using normal- or reversed-phase HPLC, often with simple isocratic solvent systems. Normal-phase is generally the more efficient mechanism for isomer separations, while esters are best resolved under reversed-phase conditions. Figure 2 illustrates the isomeric distribution in a saponified sample of cod liver oil utilizing normal-phase conditions.

The estimation of retinal and retinoic acid in foods is not frequently required. They exist in some foods as metabolites of the live animal but their contribution to the vitamin A pool is negligible compared to that of retinol. Despite the occurrence of 3-dehydroretinol in fish and particularly fish liver oils, retinol remains the primary source of vitamin A owing to its higher biopotency. Thus retinol, and its esters, are generally the retinoids of major interest during routine food analysis.

The specific details of HPLC procedures used for retinol analysis are too numerous to cover comprehensively. However, there are several features common to most schemes, notably the almost universal use of ultraviolet (either dual-wavelength

Figure 2 A 1-g sample of commercially refined cod liver oil was saponified with alcoholic potassium hydroxide and extracted into hydrocarbon. A portion was analyzed by normal-phase high-performance liquid chromatography using a Waters Radial-PAK silica column with a mobile phase of hexane:2-propanol (96:4 v/v) at 1.5 ml min^{-1}. Measurement was by fluorescence detection at 325 nm (excitation) and 470 nm (emission). Peak identification: 1, 11,13-di-*cis*-retinol; 2, 11-*cis*-retinol; 3, 13-*cis*-retinol; 4, 9,13-di-*cis*-retinol; 5, 9-*cis*-retinol; 6, all-*trans*-retinol.

or photodiode array) or fluorescence detection. Both electrochemical and mass spectrometric detection techniques have been reported but are not currently in widespread use. Normal-phase separations are generally performed on 5–10-μm particulate silica columns, replacing the earlier use of alternative adsorbents (alumina, keiselguhr, etc.). Amino-, cyano- and diol-bonded silica is sometimes advocated to circumvent the problems associated with moisture on underivatized silica. The primary mobile-phase component is usually a hydrocarbon in binary combination with a polar organic modifier, although ternary systems are sometimes recommended. Reversed-phase separations are usually confined to C_{18} bonded-phase columns, although other hydrophobic derivatives are sometimes used (C_2, C_8, C_{30}, and phenyl). Mobile phases are often a simple binary

combination of methanol (or acetonitrile) and water, and in some cases may be entirely nonaqueous.

Many foods, including poorly characterized or unknown samples, will usually be assayed subsequent to saponification, which offers the option of employing either chromatographic mode, and permits analytical simplification through conversion of multiple esters to a single retinol form. Milks, dairy products, and fortified cereals, may, alternatively, be tested by direct injection of a total lipid extract. Figure 3 illustrates a typical chromatogram obtained for a milk-based infant formula by this procedure. While retinyl acetate is the additive, the long-chain endogenous esters are visible as a composite peak, which are resolvable by independent nonaqueous, reversed-phase chromatography, as illustrated in the insert.

In fats and oil products, retinyl esters can be assayed by direct dissolution in hydrocarbon, followed by clarification and injection into the chromatographic system. Analytical success with these strategies relies on the selective nature of spectral detection modes, the narrow elution window of the various esters, and their extinction equivalence.

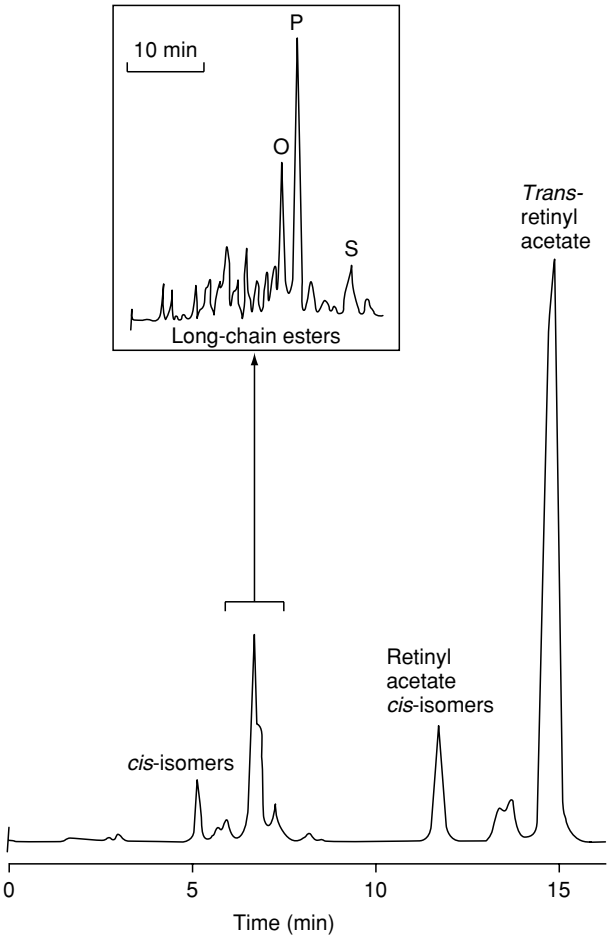

Figure 3 The lipids from 2 g of milk-based infant formula (22% fat) were extracted without saponification into 5 ml of organic solvent: 100 μl was injected directly into high-performance liquid chromatography (HPLC) equipped with a silica column and a fluorescence detector (325 nm excitation, 470 nm emission). Mobile-phase composition was hexane:2-propanol (99.92:0.08 v/v). Insert: The endogenous esters were collected and separated by nonaqueous reversed-phase HPLC on a C_{18} column with an isocratic mobile phase of acetonitrile:dichloromethane (80:20). The major identified long-chain esters are retinyl oleate (O), retinyl palmitate (P), and retinyl stearate (S).

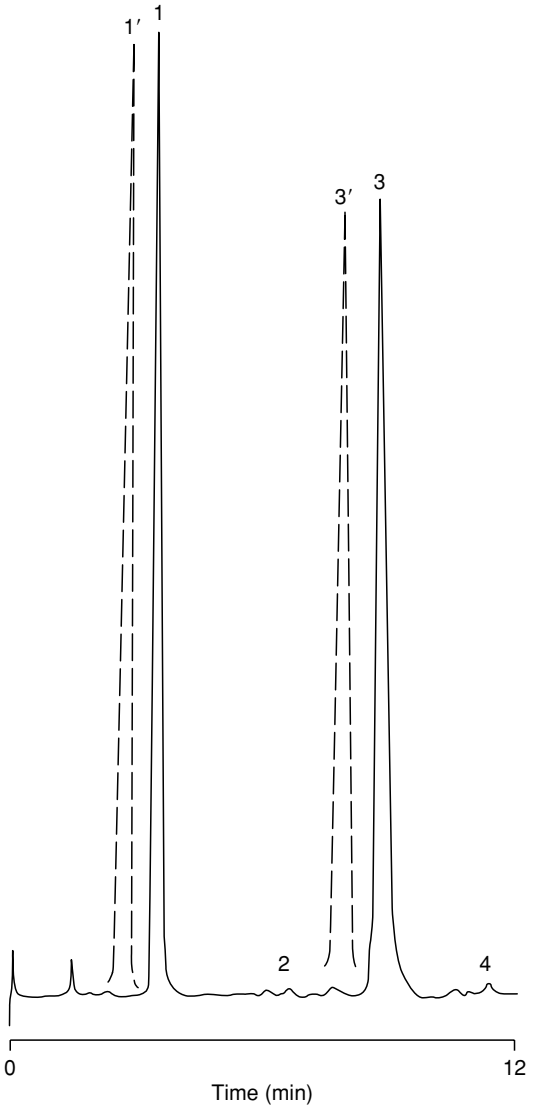

Figure 4 Retinyl acetate was extracted from a multivitamin formulation without saponification, using dimethyl sulfoxide (DMSO). High-performance liquid chromatographic (HPLC) analysis was achieved on a Waters C_{18} Radial-PAK column with a methanol (100%) mobile phase flowing at 2 ml min^{-1}. Detection was achieved by ultraviolet at 280 nm, 0.3 aufs. Vitamins D, E, and K can be assayed concurrently. Peak identification: 1, retinyl acetate (vitamin A); 2, cholecalciferol (vitamin D_3); 3, tocopheryl acetate (vitamin E); 4, phylloquinone (vitamin K_1). The elution positions of retinol (1′) and tocopherols (3′) are also indicated.

Limitations of these approaches include the requirement for normal-phase conditions in order to avoid triglyceride column fouling, and anhydrous or isohydric eluents to obtain reproducible retention. In addition, any nonesterified retinol present in the sample will not be concurrently viewed owing to its tenacious retention on silica relative to the esters.

High-potency pharmaceutical preparations and vitamin premix formulations can also be successfully assayed without the need for saponification. Thus, providing the carrier materials are free from significant fat, the retinyl esters can be extracted into a suitable solvent (e.g., dimethylsulfoxide), insuring release from encapsulation, diluted into ethanol, clarified, and injected on to a reversed-phase column. Such products usually contain vitamins D, E, and K, all of which can be concurrently assayed using both wavelength and sensitivity attenuation. However, under these conditions, it is difficult to analyze for retinyl acetate and palmitate concurrently, as the latter is highly retained by nonpolar columns. This problem can be circumvented by performing the analyses subsequent to saponification and thereby recovering total vitamin A as retinol. Figure 4 illustrates a typical chromatogram of retinyl acetate in a pharmaceutical product. The elution position of retinol is also indicated.

See also: **Antioxidants**: Natural Antioxidants; **Carotenoids**: Occurrence, Properties, and Determination; Physiology; **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; **Emulsifiers**: Uses in Processed Foods; **Food Fortification**; **Spectroscopy**: Fluorescence

Further Reading

Ball GFM (1988) *Fat-Soluble Vitamin Assays in Food Analysis. A Comprehensive Review*. London: Elsevier Applied Science.

Ball GFM (1998) *Bioavailability and Analysis of Vitamins in Food*, pp. 115–161. London: Chapman and Hall.

Bates CJ (2000) Vitamins: fat and water soluble: analysis. In: Meyers RA (ed.) *Encyclopedia of Analytical Chemistry*, pp. 1–35. Chichester: John Wiley.

De Leenheer AP, Lambert WE and Boxdaer R (2000) Vitamin A: retinol, carotenoids and related compounds. *Modern Chromatographic Analysis of the Vitamins*, Chromatographic Series, vol. 84, pp. 1–74. New York: Marcel Dekker.

Eitenmillar RR and Larder WO (1999) *Vitamin Analysis for the Helath and Food Sciences*, pp. 3–75. Boca Raton, USA: CRC Press.

Friedrich W (1988) *Vitamins*, pp. 63–140. Berlin: de Gruyter.

Klaui HM, Hausheer W and Huske G (1970) Technological aspects of the use of fat-soluble vitamins and carotenoids and of the development of stabilised marketable forms. In: Morton RA (ed.) *Fat Soluble Vitamins; International Encyclopaedia of Food and Nutrition*, vol. 9, pp. 113–159. Oxford: Pergamon Press.

McLaren DS and Frigg M (2001) *Sight and Life. Guidebook on Vitamin A in Health and Disease*. Basel, Switzerland: Roche.

Noll GN and Becker C (2000) HPLC of non-polar retinoid isomers. *Journal of Chromatography A* 881: 183–188.

Olson JA (1991) Vitamin A. In: Machlin LJ (ed.) *Handbook of Vitamins*, pp. 1–57. New York: Marcel Dekker.

Pitt GAJ (1985) Vitamin A. In: Diplock AT (ed.) *Fat-Soluble Vitamins. Their Biochemistry and Applications*, pp. 1–75. London: Heinemann.

Thiex N, Smallidge R and Beine R (1996) Sources of error in vitamin A analysis. *Journal of the Association of Official Analytical Chemists* 79: 1269–1275.

Thompson JN (1986) Review: official methods for measurement of vitamin A. *Journal of the Association of Official Analytical Chemists* 69(5): 727–738.

United States Department of Agriculture (2002) USDA National Nutrient Database for Standard Reference, Release 15. Beltsville, MD, USA: Agricultural Research Services.

Wagner AF and Folkers K (1975) *Vitamins and Coenzymes*, pp. 280–307. New York: Robert E. Krieger.

Wyss R (1990) Chromatography of retinoids. *Journal of Chromatography* 531: 481–505.

Physiology

H C Furr, Craft Technologies, Wilson, NC, USA
M M McGrane, University of Connecticut, Storrs, CT, USA

Absorption, Bioavailability, Transport, and Distribution

'Vitamin A' is the collective term for compounds that show the biological properties of retinol, including maintenance of epithelial tissue and visual function. This classification includes retinol, retinyl esters, and retinal. Retinoic acid is included, even though it does not sustain visual function, because it has been found to be the active metabolite that carries out many of the functions of vitamin A. These are isoprenoid compounds, having in common an 11-carbon polyene chain attached to a trimethyl substituted cyclohexenyl ring (**Figure 1**). The term 'retinoids' refers to all compounds, natural or synthetic, that show some biological activity typical of vitamin A, such as promoting differentation of cells in culture; not all retinoids can support all the functions of vitamin A, such

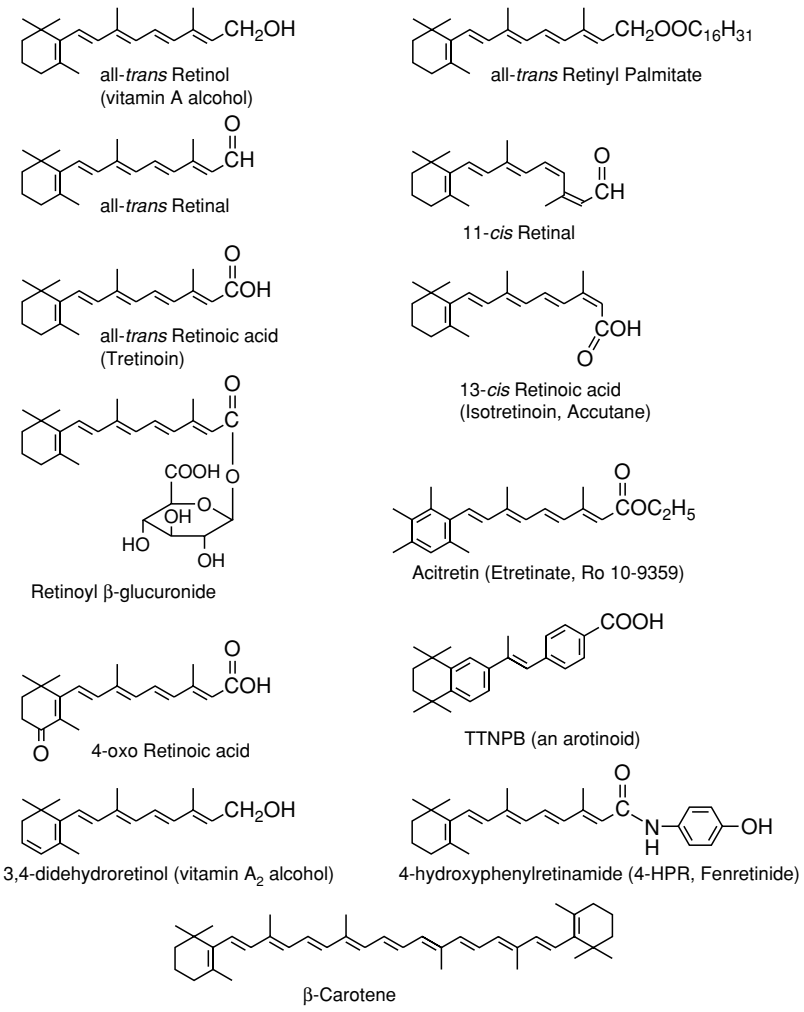

Figure 1 Chemical structures of some representative retinoids.

as vision. Vitamin A compounds are not found as such in plant tissues, but rather are characteristic of the animal kingdom; the notable exception is 13-*cis* retinal, which serves as a chromophore in the purple membrane of certain halobacteria.

Dietary vitamin A is provided by two sources: preformed vitamin A and provitamin A carotenoids. Preformed vitamin A (mostly as esters of retinol with long-chain fatty acids) comes from animal products or from dietary supplements; retinyl esters, e.g., retinyl acetate and retinyl palmitate, are more stable chemically than is free retinol. Provitamin A carotenoids arise mostly from plant products: β-carotene is the most active and is widely distributed in plants, but other carotenoids (such as α-carotene and β-cryptoxanthin and β-apo-carotenals) can be important sources of vitamin A from particular foods (*See* **Carotenoids**: Physiology). The relative importance of these sources of vitamin A is very dependent on diet. Other carotenoids, such as lycopene and xanthophyls (including lutein and zeaxanthin), are major carotenoids in some foods and may have other important physiological functions, but they have no provitamin A activity in higher animals.

Typical estimates of dietary vitamin A absorption efficiency are 70–85%. Estimates of carotenoid absorption are usually much lower, but are confounded by slow intestinal absorption and rapid metabolism; there is considerable species variability in absorption efficiency and in metabolism of carotenoids. Animal-feeding studies show that the biological matrix of food carotenoids has profound effects on their bioavailability.

Since both vitamin A and carotenoids are lipids, intestinal micelle formation with bile acids is essential for their absorption. Human subjects with impaired bile acid formation or flow (e.g., with biliary atresia) may require intramuscular supplementation with vitamin A and the other fat-soluble vitamins. Within the intestinal lumen, vitamin A esters (retinyl esters)

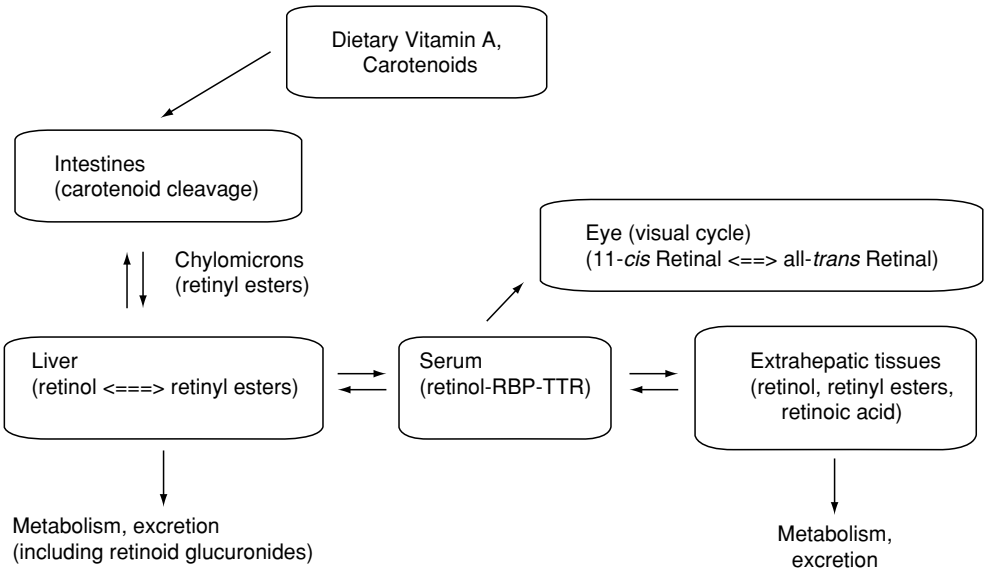

Figure 2 Overview of vitamin A metabolism.

are hydrolyzed to free retinol and are absorbed as such; this free retinol is promptly reesterified within the intestinal cells. (Figure 2). Provitamin A carotenoids, such as β-carotene, are often cleaved within intestinal cells; this metabolism is predominantly by central cleavage (mediated by the enzyme carotene 15,15′-dioxygenase), although asymmetric cleavage followed by chain shortening may play a role. Retinal (also called retinaldehyde or vitamin A aldehyde), the final product of carotenoid cleavage, is enzymatically reduced to retinol and is then esterified. The physiological ligand for this esterification process seems to be retinol bound to an intracellular retinol-binding protein (CRBP II, molecular weight (MW) approximately 14 600 Da, one of several small intracellular retinoid-binding proteins); the primary retinyl ester synthesizing activity transfers fatty acid from phosphatidyl choline (lecithin:retinol acyltransferase, LRAT), although an acyl-coenzyme A-dependent esterifying activity (acylCoA:retinol acyltransferase, ARAT) is also present.

Regardless of their dietary source, the retinyl esters are incorporated in the core of chylomicra and transported in the lymph. After removal of triacylglycerols by lipoprotein lipase as the lipoprotein particle circulates through peripheral tissues, the chylomicron remnants are rapidly taken up by the liver, and the vitamin A esters are hydrolyzed by retinyl ester hydrolase there. The resulting retinol is then either reesterified (primarily by LRAT, although ARAT activity may be important at high retinol concentrations) and stored in the liver, or released into the plasma as a complex with plasma retinol-binding protein (RBP). The two cell types involved in liver vitamin A metabolism are the hepatocytes (parenchymal cells) and the lipocytes (also called stellate cells, Ito cells, or fat-storing cells). The hepatocytes are the major site of RBP synthesis and retinol-RBP release; some retinyl esters are also found here. The lipocytes store retinyl esters in cytoplasmic lipid droplets, which also contain triacylglycerols and some cholesteryl esters within a phospholipid coat. It has been suggested that RBP transfers vitamin A, as retinol, between lipocytes and hepatocytes.

Because most forms of vitamin A are hydrophobic, the transport, metabolism, and function of vitamin A are dependent on a series of binding proteins, each specific for its ligand and tissue. Plasma RBP typically has molecular weight of approximately 21 000 Da in mammalian species; the complete amino acid composition has been determined for several species, and the gene from some species has been cloned. RBP binds retinol with 1:1 stoichiometry. The hydrophobic retinol molecule fits into a 'barrel' within the protein, shielded from interactions with the aqueous environment. Holo-RBP, in turn, binds transthyretin in plasma (TTR; formerly called prealbumin); a TTR tetramer can bind up to four molecules of RBP. Usual concentrations of human plasma RBP are 1.9–2.4 μM (40–50 $\mu g\,ml^{-1}$); typically, total circulating RBP is 80–90% saturated with retinol ligand. Although retinol, retinal, and retinoic acid bind to RBP with similar affinity, retinol is present in highest concentrations and is the predominant ligand. Retinyl esters have much less affinity for RBP and are transported by serum lipoproteins instead.

The release of holo-RBP from liver hepatocytes is carefully controlled to maintain levels of circulating

retinol within narrow limits, but the mechanism of this control is not yet clear. In the absence of adequate vitamin A, apo-RBP accumulates in the liver, ready to be released as soon as vitamin A is available (this is the basis of the Relative Dose Response assay for vitamin A deficiency, as discussed below). Clearly, this mechanism has developed because of the dichotomy of vitamin A action: vitamin A is essential, in small amounts, for proper differentiation and maintenance of epithelial cells, but excesses of vitamin A are toxic and must be avoided by the organism.

Metabolism, Storage, and Excretion

Vitamin A in excess of immediate requirements is stored in the liver as esters of long-chain fatty acids (Figure 2); the ratio of liver retinol to retinyl esters decreases as total liver vitamin A increases, but typically, 95% of total liver vitamin A is present as the esters. Retinyl palmitate is the primary ester in the liver of the human, rat, cow, sheep, rabbit, cat, frog, trout and polar bear, although significant amounts of other esters are found (especially oleate and stearate), and the liver retinyl ester composition can be affected by dietary fatty acid composition. (Remarkably, the predominant vitamin A ester in rat adrenal cortex is retinyl stearate.) Liver frequently contains as much as 90% of total body vitamin A, although other organs, such as kidney, testes, and adrenal glands, also contain detectable retinyl esters. The efficiency of liver storage of vitamin A is particularly noteworthy in the polar bear, where concentrations as much as 36 μmol (10 380 μg) per gram of liver have been reported. Typical values for human liver vitamin A (autopsy specimens) are 0.44–0.74 $\mu mol\ g^{-1}$ (126–211 $\mu g\ g^{-1}$) in the USA.

Kinetic studies carried out in rats have shown that there is extensive cycling of liver vitamin A stores, and extensive recycling of vitamin A between liver and other tissues; this assures an ample supply of retinoid to tissues in response to immediate needs. It is certain that retinol-RBP is the major form of transport from liver to peripheral tissues; although the transport of vitamin A from other tissues back to the liver may be via retinyl esters transported by lipoproteins (normally present at about 5–10% the serum concentration of retinol-RBP), transport via retinol on RBP synthesized in the outlying tissues may be an important pathway (mRNA for plasma RBP has been detected in a number of other tissues in addition to liver). In vitamin A-deficient rats, the recycling is even more extensive, and catabolism of vitamin A is markedly diminished.

The rat has served as a valuable model for vitamin A metabolism in the human, particularly for aspects that can not be studied readily in humans. In the rat, catabolism of vitamin A occurs by oxidation of the cyclohexenyl ring (particularly at the 4-position), epoxide formation at the 5,6-positions, hydroxylation of the ring methyl groups, and chain shortening. The resulting metabolites are generally inactive, although some may have a little biological activity. These more polar retinoids are excreted in the urine and in the bile.

Retinol can be reversibly oxidized to retinal (vitamin A aldehyde); retinal can be oxidized to retinoic acid, but retinoic acid cannot be converted back to the other forms. Thus, retinol, retinyl esters, and retinal have equal biological activity because they are freely interconverted; retinoic acid fills some (but not all) of the functions of retinol, but it and its derivatives are not stored. Typical human serum concentrations of retinoic acid are 5–10 nM (1.5–3 $ng\ ml^{-1}$), compared with typical retinol concentrations of 1–2 μM. Retinoyl β-glucuronide and retinyl β-glucuronide (formed in the liver from retinoic acid and retinol, respectively) are secreted into the bile but can be hydrolyzed and reabsorbed in the intestine (enterohepatic circulation); both these compounds have vitamin A activity in a variety of tests, are found in human blood, and may have regulatory roles in vitamin A function.

Roles in the Body

Vision

The major roles of vitamin A are in vision and in control of gene expression (including differentiation of epithelial tissues, and in the immune system, as well as a host of other systems). Metabolism of vitamin A in the retina of the eye is unique, in keeping with the unusual role of vitamin A in that tissue. Vitamin A is stored in the retinal pigment epithelium as retinyl esters. All-*trans* retinyl esters are simultaneously hydrolyzed and isomerized to 11-*cis* retinol by an isomerohydrolase enzyme that uses the free energy of hydrolysis of the retinyl esters to drive the isomerization reaction. 11-*cis* Retinol, a compound unique to the eye, is then oxidized to 11-*cis* retinal by an alcohol dehydrogenase. 11-*cis* Retinal is transferred from the pigment epithelium cells to the rod cells by interstitial retinoid binding protein (IRBP), a distinct binding protein (MW 140 000). In the rod cells, 11-*cis* retinal binds (as a Schiff base with the ε-amino group of a specific lysine residue) to the protein opsin to form the visual pigment, rhodopsin. When a photon of light is absorbed by a rhodopsin complex, the 11-*cis* retinal is isomerized to all-*trans* retinal and released from the protein complex; the resulting conformation change of the protein initiates a cascade of reactions, resulting in a neural signal to the brain. The protein opsin is then available to bind another molecule of 11-*cis*

retinal for another round of the visual cycle. The all-*trans* retinal that was released from the protein complex is transferred back to the pigment epithelium via IRBP, enzymatically reduced to all-*trans* retinol, and esterified again. In contrast to the high turnover rates of vitamin A in other tissues, vitamin A in the eye is highly conserved, with little leakage back to the liver. Prolonged vitamin A deficiency, however, leads to reduced sensitivity to light, usually first noted as impaired vision at night (night blindness). These effects of vitamin A deficiency are generally reversible by subsequent vitamin A supplementation.

In a very different role of vitamin A in the eye, the cornea depends on vitamin A for proper cell differentiation and for secretion of protective glycoproteins. In vitamin A deficiency, these tissues are susceptible to attack by opportunistic bacteria; such attack may result in permanent scarring and permanent vision loss. These effects of vitamin A deficiency, unlike those in the retina, may not be reversible by subsequent vitamin A supplementation. Such vitamin A-dependent corneal degeneration (given the general name 'xerophthalmia') accounts for an estimated 500 000 new cases of blindness in preschool children in the world each year.

Control of Gene Expression

Overview of control of gene expression Vitamin A and its metabolites are required for embryonic development, cellular differentiation, growth, reproduction, adult tissue homeostasis, vision, and maintenance of the immune system. Over the past several years, it has become clear that retinoids mediate many of their pleiotropic effects by regulating the activation or inhibition of gene transcription. It is well documented that retinoids regulate the expression of many genes involved in their own transport and metabolism. The genes for different retinoid receptor subtypes, the cellular retinol binding proteins (CRBP) I and II, and the cellular retinoic acid binding protein II (CRABP II) are regulated by either all-*trans* retinoic acid (RA) or 9-*cis* RA (Figure 3). These vitamin A metabolites exhibit hormonal action by binding to their nuclear receptors that interact with DNA response elements within target genes to affect gene transcription. The retinoid receptors are members of the steroid/thyroid/retinoid superfamily of nuclear receptors. There are two distinct families of nuclear retinoid receptors: the retinoic acid receptors (RARs) and the retinoid X receptors (RXRs). It has been suggested that the two classes of receptors have a distinct functional activity, the RARs being important in cell development and differentiation, and the RXRs being important in metabolic regulation. Nuclear receptors of the steroid/thyroid/retinoid superfamily can act as either inhibitors or activators of transcription. The mechanism whereby these regulatory properties are propagated is via the recruitment of accessory proteins to the nuclear receptor. In the absence of ligand, such as all-*trans* RA, corepressor proteins are bound to retinoid receptors and cause inhibition of transcription of the target gene. Conversely, binding of ligand causes a conformational change in the receptor such that corepressor proteins are released, and coactivator proteins are recruited to the retinoid receptor. The association of coactivators with the nuclear retinoid receptor causes induction of transcription of the target gene, initiated by retinoic acid binding to its specific nuclear receptor.

Nuclear retinoid receptors As indicated previously, there are two distinct classes of nuclear retinoid receptors, the RARs and RXRs. Both RARs and RXRs bind the naturally occurring metabolites of vitamin A with a similar order of affinity (all-*trans* RA > retinal > retinyl acetate > retinol), but RXRs require from 10- to 40-fold more ligand for comparable transcriptional activation. A downstream metabolite of

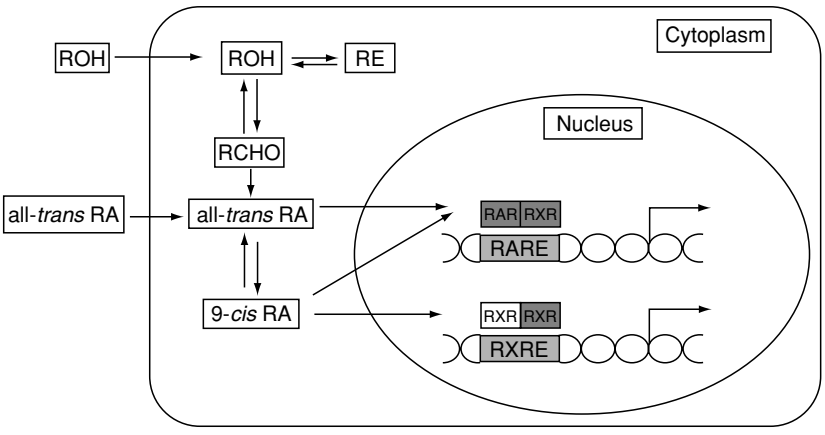

Figure 3 Retinoid action in the nucleus.

all-*trans* RA, 9-*cis* RA, has been shown to be the naturally occurring, high-affinity, ligand for the RXR. The binding of ligand to nuclear retinoid receptors induces dimer formation and initiates transcriptional activation of a target gene via the DNA response element(s) in the promoter/regulatory region of the gene. Retinoid receptors act in a dimeric conformation and bind to specific RA response elements (RARE) within the 5′promoter/regulatory regions of target genes. The DNA response elements are usually a direct repeat (DR) of six nucleotides in a specific sequence separated by a defined number of nucleotides that vary depending on the type of DNA response element. RAR/RXR heterodimers can activate or repress gene transcription, depending on the type of RARE to which they bind in the promoter region of a target gene. If the RAR/RXR heterodimer binds to a DR-5 RARE, the RAR/RXR complex activates transcription in response to the RAR ligand all-*trans* RA. In contrast, RAR/RXR heterodimers bound to a DR-1 RARE exhibit no response to RAR-specific ligands and repress the response to RXR-specific ligands such as 9-*cis* RA. Therefore, the RAR/RXR heterodimer can be stimulatory, neutral, or inhibitory, depending on ligand, all-*trans* RA or 9-*cis* RA, and the DNA RARE to which it is bound. This is partially explained by the fact that there is a specific polarity to the binding of RAR and RXR heterodimeric partners, determined by the DNA itself. RAR/RXR heterodimers bind with the RXR occupying the upstream half-site of a DR-5 element, but this orientation is reversed on a DR-1 element. Further investigation of this polarity of RAR/RXR heterodimer binding has shown that transactivation by RAR bound to a DR-5 RARE involves the ligand-dependent dissociation of a bound corepressor (discussed below), but RAR bound to a DR-1 RARE does not respond to ligand by dissociation of a bound corepressor. Therefore, transcription is not stimulated in the later response. In this regard, the DNA itself is considered an allosteric regulator of receptor–accessory protein interaction.

All of the RAR and RXR subtypes, α, β, and γ, are encoded by separate genes at distinct chromosomal loci. However, the protein products encoded by these genes have certain common modular structures, particularly in the DNA-binding and ligand-binding domains, which define, in part, their membership within the steroid/thyroid/retinoid superfamily. The amino acid sequence of the RARs can be divided into six regions (A–F). The amino acid sequence of the RXR genes has a similar domain structure to the RARs. Among the retinoid receptors, the DNA-binding domain (C) is a highly conserved region of approximately 70 amino acids that contains two zinc fingers; the zinc-finger configuration is responsible for specific DNA response element recognition. According to Evans and colleagues, a sequence of three amino acids within the first zinc finger (the P box) determines recognition of the half-site sequence of a DNA response element, and five amino acids in the second zinc finger (the D box) are critical for recognition of the spacing between half-sites. Nuclear magnetic resonance spectroscopy of the DNA-binding domain indicates a secondary structure comprising three α-helices; the first helix is postulated to fit across the major groove of the DNA response element. In this orientation, the third helix interacts with the minor groove of the DNA, such that Lys and Arg residues make contact with the phosphate backbone and stabilize protein–DNA interactions. This third helix in the DNA-binding domain is also thought to mediate protein–protein interaction required for dimer formation. The ligand-binding domain (E) is also highly conserved among the RAR and RXR genes and mediates the ligand-dependent transcriptional activation function. The amino-terminal region (A) is less well conserved and is subtype-specific. Both regions A and B are responsible for ligand-independent transcriptional activation. The amino-terminal part of domain D may function as a nuclear localization site.

In addition to subtype differences in RARs and RXRs, there are also different mRNA isoforms for each gene. These isoforms are generated by different transcription start sites and alternative splicing, which affects the 5′ untranslated region (UTR) and the A domain such that either a deleted or altered A region is combined with a common B–F region.

The biological activity of vitamin A metabolites is further extended by the interaction of the retinoid receptors with other members of the receptor superfamily. The RXRs serve as heterodimeric partners with RARs, thyroid hormone receptors (TR), and the vitamin D receptors (VDR) and have been shown to increase the affinity of RAR, VDR, and TR, for their respective response elements (favored DR motifs DR-2/DR-5, DR-3, and DR-4, respectively), and to activate transcription in the presence of the appropriate ligand. This indicates a central role for the RXR as a heterodimeric partner with the other nuclear receptors of this superfamily. Interestingly, only one ligand is required to stimulate the heterodimeric complex and affect transcriptional activation. However, the RXR is unique in that it can function as a homodimer. In this capacity, it can stimulate gene transcription favoring the DR-1 hormone response element and is dependent upon the 9-*cis* RA ligand. The RXR also forms a heterodimeric complex with the peroxisome proliferator-activated

receptor (PPAR) and, for example, mediated by a DR-1 motif in the promoter of the acyl-CoA oxidase gene, stimulates transcription in response to both 9-*cis* RA or clofibric acid (a peroxisomal proliferator). Recently, it has been shown that dietary polyunsaturated fatty acids (PUFA) activate PPAR, and via PPAR–RXR heterodimer formation, dietary PUFA regulate both lipid and glucose metabolism. Clarke refers to dietary PUFA as 'fuel partitioners' that act, via PPAR activation, to increase glycogen storage and fatty acid oxidation while decreasing triglyceride synthesis.

The multiplicity of nuclear retinoid receptors reflects developmental and tissue-specific differences in their expression. In the mouse embryo, RARα is ubiquitously expressed, whereas RARβ and RARγ are more restricted, and nonoverlapping, in their expression. RXRβ genes are widely expressed in the mouse embryo, whereas RXRα is found in liver, kidney, intestine, and skin, and RXRγ is more restricted still. In the adult, both RARα and RXRβ are found in most tissues, whereas RARβ, RARγ, RXRα, and RXRγ are more limited. RXRα is highly expressed in liver, kidney, spleen, and visceral tissues, indicating a possible role for this RXR in retinoid, lipid, and carbohydrate metabolism. The demonstrated RXR responsiveness of numerous genes including the CRBP II, apolipoprotein (apoAI), acyl-CoA oxidase and the phosphoenolpyruvate carboxykinase (PEPCK) genes further supports this hypothesis.

Nuclear coactivators and corepressors associated with nuclear retinoid receptors Recently, several coactivators and corepressors that associate with nuclear retinoid receptors and affect their ability to regulate target genes have been cloned and characterized. It has been established that retinoid binding induces conformational changes in nuclear retinoid receptors and promotes their association with a diverse group of nuclear proteins that function as coactivators of transcription. Alternatively, unbound receptors recruit nuclear proteins that repress transcription of target genes. It has been demonstrated that all-*trans* RA binding to RAR, with subsequent activation of transcription of a target gene, requires dissociation of corepressor proteins and recruitment of coactivator proteins. It is thought that corepressors contribute to the formation of condensed chromatin structure, prohibiting activation of target genes. Conversely, most coactivators cause a loosening of chromatin conformation, allowing nuclear receptors to activate transcription of target genes.

Regulation of the PEPCK gene by retinoids The multifaceted action of vitamin A is indicated by its role in regulation of the gene encoding PEPCK, the rate-determining enzyme of gluconeogenesis in liver (Figure 4). All-*trans* RA and 9-*cis* RA, in equal concentrations, induce the PEPCK promoter by approximately sevenfold in hepatoma cells in culture. Studies in mice *in vivo* have shown that vitamin A deficiency causes a decrease in PEPCK gene expression. Specifically, with food deprivation, hepatic PEPCK mRNA levels are not induced in vitamin A-deficient mice. Induction of the PEPCK gene can be reestablished with retinoid treatment of vitamin A deficiency. Retinoid regulation of the PEPCK gene has been tested *in vivo* in two lines of transgenic mice. Results from these experiments indicate that expression driven by both the −460/+73 and −355/+73 PEPCK DNA sequences is decreased with vitamin A deprivation. Subsequent treatment of vitamin A deficiency with all-*trans* RA increases expression driven by the −460/+73 promoter or the minimal −355/+73 promoter. The second retinoid isomer, 9-*cis* RA, however,

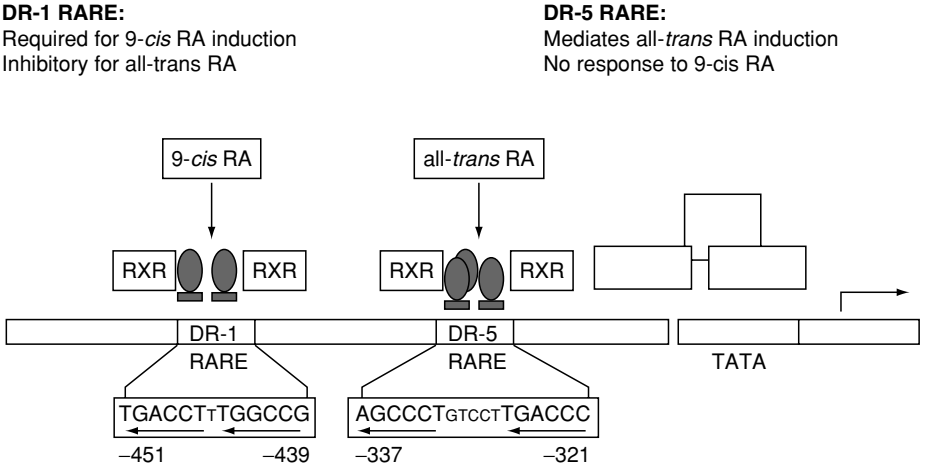

Figure 4 Retinoic acid response elements involved in regulation of the PEPCK gene.

increases expression driven by the −460/+73 promoter, but not the −355/+73 promoter. Therefore, the two active retinoid isomers in liver have different molecular signaling mechanisms and activate the PEPCK gene at different DNA response elements.

Additionally, it has been shown that appropriate developmental regulation of the PEPCK gene during fetal development and the perinatal period is dependent on adequate vitamin A nutriture. Vitamin A deficiency causes a decrease in hepatic PEPCK mRNA during late gestation, a stage in development when the PEPCK gene is first expressed in liver. At birth, with adequate vitamin A nutriture, the PEPCK gene is rapidly induced. This is consistent with the need of the newborn to conduct gluconeogenesis because of removal from the maternal glucose supply. The rapid induction of PEPCK gene expression at birth is significantly decreased with vitamin A depletion in the newborn. Overall, adequate vitamin A is required for the appropriate developmental expression of the PEPCK gene in late gestation and at birth. The fact that vitamin A is an active regulator of expression of the PEPCK gene indicates a significant role for this nutrient in the regulation of carbohydrate metabolism overall. This is a new area of investigation of vitamin A regulation of gene expression as it expands the known role of vitamin A beyond the regulation of genes involved in vitamin A metabolism itself.

Human Health

Infection and Disease

Early in the study of vitamin A requirements, it was appreciated that vitamin A-deficient animals are more susceptible to disease. Attempts to extend these observations to human health long gave inconsistent results, however. Recently, it has been shown that supplementation of children with vitamin A reduces the severity of measles and diarrhea (although vitamin A supplementation may not reduce the incidence and duration of diarrhea). Although vitamin A prophylaxis may not reduce measles severity, supplementation of measles patients with vitamin A (200 000 international units (IU), i.e., 60 000 μg of RE, 209 μmol) during infection does reduce hospital stay and mortality rate and is now recommended by WHO in communities where vitamin A deficiency is prevalent. Currently, it is not believed that vitamin A supplementation can provide any benefit for other respiratory diseases (nonmeasles pneumonia). The possible role of vitamin A status in tuberculosis, malarial infection, and HIV infection is not clear. However, vitamin A public health programs can have a tremendous impact on child morbidity and mortality in developing countries.

The mechanism(s) by which retinoids affect the immune system have not yet been delineated, but it seems safe to assume that control of gene expression is involved. In at least some experiments, the function of vitamin A can be filled by all-*trans* or 13-*cis* retinoic acid, consistent with current knowledge of retinoids and gene expression. Both antibody-mediated and cell-mediated immune responses are reduced in vitamin A deficiency. Interpretation of human studies of the relationship between vitamin A status and immune response is complicated by the fact that plasma retinol concentrations (often used as a convenient measure of vitamin A status) fall during infection even in vitamin A-replete subjects, and chronic infection reduces vitamin A absorption and mobilization.

Dermatology One of the first pharmaceutical applications of retinoids was the use of 13-*cis* retinoic acid (isotretinoin) to combat cystic acne, ichthyosis, and follicular keratosis; 13-*cis* RA inhibits sebaceous gland size and activity by promoting CRABP II mRNA expression. The process of keratinization (differentiation and maturation) of skin cells is affected both by deficiency and by excess of retinoids. Dehydroretinol and its esters (vitamin A_2) are found in human skin (more so in epidermis than in dermis) and may have specific roles there. Retinoids are now used clinically to treat psoriasis, ichthyosis, Darier's disease, and other disorders of keratinization, acne, and photoaged skin. Synthetic retinoids, such as etretinate and the ethyl ester of the arotinoid TTNPB (Figure 1), as well as the naturally occurring retinoyl β-glucuronide, have been developed in attempts to reduce the toxicity of the naturally occurring retinoic acids without losing efficacy.

Cancer prevention and treatment Retinoids can inhibit cell transformation and tumor development in a variety of cell and animal models, but the effects are not uniform with all malignancies. In general, it seems that retinoids may not prevent initiation of tumors, but more likely interfere with the promotion process. They exert these effects by modulating growth, differentiation, and apoptosis of cells; most, if not all, of these effects may be explained by the role of retinoids in control of gene expression.

Retinoids have been tested against a variety of cancers in human and animal studies. *N*-(4-Hydroxyphenyl)-retinamide (4-HPR, Fenretinide) has been tested clinically for prevention of recurrence of human breast cancer; although results are promising, some toxicity (owing to interference with normal retinol metabolism) has been observed. Incidence of skin tumors associated with xeroderma pigmentosum is reduced by treatment with 13-*cis* retinoic acid.

Perhaps the most dramatic application is the use of retinoic acid (tretinoin) against promyelocytic leukemia; this disease results from a chromosomal translocation producing the fusion of the RARα gene with the PML gene to produce an oncogene and responds well to high doses of all-*trans* retinoic acid.

Requirements and Recommended Intakes

Because of confusion among the various units for presenting vitamin A values, the concept retinol equivalent (RE) has been proposed. One retinol equivalent equals 1 μg of all-*trans* retinol, either free or as the retinyl component of a retinyl ester; 1 RE = 3.33 IU of vitamin A, or 3.5 nmol of retinol or retinyl ester. Although the exact vitamin A value of carotenoids depends on several factors (*See* **Carotenoids**: Occurrence, Properties, and Determination), the retinol equivalent was defined as 6 μg of all-*trans* β-carotene or 12 μg of other provitamin A carotenoids (10 IU of provitamin A carotenoid). To account for updated findings on provitamin A value of carotenoids, the retinol activity equivalent (RAE) has been defined more recently as 1 μg of all-*trans* retinol, 12 μg of all-*trans* β-carotene, or 24 μg of other provitamin A carotenoids.

Liver concentrations provide the best appraisal of vitamin A status. Human liver specimens can be difficult to obtain, but analysis of autopsy specimens can be useful in evaluating the vitamin A status of a population. As indicated above, serum retinol levels are maintained nearly constant and are not generally useful in assessing vitamin A status, except when liver vitamin A reserves fall well below $0.07\,\mu mol\,g^{-1}$ (20 μg per gram of liver). Serum (or plasma) retinol levels are normally 1–2 μM (290–570 $ng\,ml^{-1}$) across a wide range of mammalian species.

Because liver concentrations of vitamin A are not readily measured, and serum retinol values are not an adequate indicator of an individual's vitamin A status, several indirect methods have been developed. Conjunctival impression cytology (CIC) evaluates the development of squamous metaplasia (enlarged epithelial cells) and loss of goblet cells from the conjunctiva of the eye by histological examination of cells transferred from the cornea to filter paper. The relative dose response (RDR) depends on the release of retinol-RBP from liver into serum after a large oral dose of vitamin A (typically, 450 μg of retinyl acetate in human studies); as described above, apo-RBP accumulates in the liver in vitamin A depletion. The RDR is calculated as:

$$\mathrm{RDR} = \frac{A_5 - A_0}{A_5} \times 100\%,$$

where A_5 represents the serum retinol concentration at 5 h after the oral dose, and A_0 represents the fasting serum retinol concentration at the time of the dose. Studies in both humans and rats indicate that an RDR value greater than 20% indicates liver vitamin A reserves less than $0.07\,\mu mol\,g^{-1}$ (< 20 μg per gram of liver), i.e., inadequate vitamin A status. The RDR assay requires, of course, that the oral dose be normally absorbed (an intramuscular injection has been used in human subjects with biliary atresia), and that protein metabolism is normal: protein deficiency or zinc deficiency or liver cirrhosis impairs the dose response. An alternate approach, the modified relative dose response (MRDR), uses the vitamin A analog 3,4-didehydroretinol (vitamin A_2, a form found in some freshwater fish and found in very low levels in human skin). The ratio of vitamin A_2 to vitamin A_1 in serum retinol at 5 h after an oral dose is used as a measure of vitamin A status, with high values (ratio > 0.03 after an oral dose of 100 μg per kilogram of body weight) indicating poor vitamin A status. Chlorinated vitamin A analogs have been used in a similar fashion. The RDR and MRDR, as well as CIC, have been used successfully in human population studies. Isotope dilution of tracer-labeled vitamin A (radioactive or stable-isotope labeling) has been used to estimate human vitamin A status, but technical difficulties have so far prevented general use of the technique.

Nutritional requirements for vitamin A have not been well defined because of the diversity of vitamin A functions. In animal studies, intakes of 3–8 RE per day per kilogram of body weight (10–28 nmol per day per kilogram of body weight) cure deficiency symptoms and intakes of 30–60 RE per day per kilogram of body weight (100–210 nmol per day per kilogram of body weight) produce optimal growth. Kinetic studies of vitamin A metabolism in rats show that irreversible loss of vitamin A is decreased on low vitamin A intakes. Studies on vitamin A requirements have not yet addressed functional criteria such as immune function and possible anticancer effects.

To provide an *adequate and safe human intake*, current WHO/FAO dietary recommendations (1988) and suggested recommended dietary intakes (RDIs) for adults are based on intakes of 9.3 RE per day per kilogram of body weight (33 nmol of vitamin per day per kilogram of body weight). The European Union Scientific Committee for Food has recommended population reference intakes of 700 RE per day (men) and 600 RE per day (women), with estimated average requirements of 500 and 400 RE, respectively, and lowest threshold intakes of 300 and 250 RE, respectively. In contrast, the recommended dietary allowances (RDAs) of the US National

Research Council (1989) attempted to set an *optimal intake level*, and so recommended a higher RDA of 1000 RE per day for men and 800 RE per day for women (the difference being due to a lower body weight, on average). However, the most recent recommendations from the US Institute of Medicine are calculated to *maintain an adequate body pool size* and set an RDA of 900 μg of RAE for adult males, 700 μg of RAE for adult females.

Toxicity

Over 600 individual cases of human vitamin A toxicity have been reported, owing to either acute (single or a few large doses ingested over a brief period of time) or chronic intake (moderately high doses taken frequently for periods of months or years). Acute toxicity in human adults is reported from doses of 300 000–10 000 000 RE (1000–35 000 μmol); chronic doses of 15 000–300 000 RE (50–1000 μmol) have produced hypervitaminosis A. Symptoms include headache, vomiting, diplopia, alopecia, dryness of mucous membranes, desquamation, bone abnormalities, and liver damage. However, single oral doses of 60 000 RE in oil have been successfully used in vitamin A intervention programs for preschool children, with transient toxic symptoms observed in no more than 3% of subjects. Massive doses of β-carotene are not toxic, but may be less efficiently absorbed and used than vitamin A itself. Rodent animal models have been extremely valuable in elucidating vitamin A requirements and metabolism, but the rat seems to be much less susceptible to hypervitaminosis A than is the human.

The toxic effects of high vitamin A intakes are mediated by serum retinyl esters; retinol-RBP concentrations are maintained at normal levels in hypervitaminosis A, but serum retinyl esters are markedly elevated, bypassing the normal homeostatic controls on vitamin A transport. Some carnivores, including the dog, are unusual in having high fasting levels of serum retinyl esters, presumably reflecting differences in lipoprotein metabolism in these species; the implications of this for vitamin A metabolism and for resistance to hypervitaminosis A are not clear.

The most tragic consequences of excessive vitamin A intake are teratogenicity (specifically, malformations of the cranium, face, heart, thymus, and central nervous system) and embryotoxicity. Acidic retinoids such as those that have been used in dermatology are particularly potent, as they can attain high serum levels and can readily pass the placental barrier. Although large intakes of vitamin A itself (> 7500 RE per day, 26 μmol per day) early in human pregnancy can cause birth defects (perhaps owing to metabolism of retinol to retinoic acid?), serum concentrations of retinol and retinyl esters are normally maintained at moderate levels during pregnancy. It is assumed that these teratogenic effects are related to the important role of retinoids in differentiation of cells and that these effects are mediated via the nuclear receptor proteins. (Interestingly, the retinoid β-glucuronides are much less teratogenic than retinoic acid.) In view of these teratogenic effects occurring at less than 10 times the recommended daily intakes, a consensus of several professional organizations is that women should avoid vitamin A supplements during the first trimester of pregnancy, and that subsequent supplements, if taken at all, should be prudently limited to 8000 IU (2400 μg RE, 8.5 μmol) per day, although 10 000 IU (3000 μg RE, 10 μmol) appears to be safe.

The previous two decades have been exciting in retinoids research, with the disclosure of the roles of retinoids in control of gene expression. Continuing challenges in vitamin A research include: (1) the development and confirmation of indirect indices of vitamin A status; (2) elucidation of the mechanism of control of serum retinol-RBP concentrations; (3) more exact determination of vitamin A requirements for specific functions (not only growth and prevention of blindness, but also immune function and cell differentiation in individual tissues); and (4) definition of the role of vitamin A in differentiation in specific tissues, perhaps leading to chemical design of distinctive retinoids to combat specific cancers and dermatologic diseases.

See also: **Carotenoids**: Occurrence, Properties, and Determination

Further Reading

Blomhoff R (1994) *Vitamin A in Health and Disease*. New York: Marcel Dekker.

Chambon P (1996) A decade of molecular biology of retinoic acid receptors. *FASEB Journal* 10: 940–954.

Chambon P, Zelent A, Petkovich M *et al.* (1991) The family of retinoic acid nuclear receptors. In: Saurat JH (ed.) *Retinoids: Ten Years On*, pp. 10–27. Basel: Karger.

Clarke SD, Thuillier P, Baillie RA and Sha X (1999) Peroxisome proliferator-activated receptors: a family of lipid-activated transcription factors. *American Journal of Clinical Nutrition* 70: 566–571.

Collingwood TN, Urnov FD and Wolffe AP (1999) Nuclear receptors: coactivators, corepressors, and chromatin remodeling in the control of transcription. *Journal of Molecular Endocrinology* 23: 255–275.

Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc (2001), pp. 65–125. Washington, DC: National Academy Press. (available online at http://books.nap.edu/books/0309072794/html/65.html).

McKenna NJ, Xu J, Nawaz Z *et al*. (1999) Nuclear receptor coactivators: multiple enzymes, multiple complexes, multiple functions. *Journal of Steroid Biochemistry and Molecular Biology* 69: 3–12.

Mangelsdorf DJ, Umesono OK and Evans RM (1994) Retinoid receptors. In: Sporn MB, Roberts AB and Goodman DS (eds) *The Retinoids*, 2nd edn, pp. 319–349. New York: Raven Press.

Nalubola R and Nestel P (1999) *The Effect of Vitamin A Nutriture on Health: A Review*, pp. 1–83. Washington, DC: ILSI Press (available at www.ilsi.org/hnipubs.htm).

Napoli JL (1996) Retinoic acid biosynthesis and metabolism. *FASEB Journal* 10: 993–1001.

Napoli JL (1999) Retinoic acid: its biosynthesis and metabolism. *Progress in Nucleic Acid Research and Molecular Biology* 63: 139–188.

Ross AC (1999) Vitamin A and retinoids. In: Shils ME, Olson JA, Shike M and Ross AC (eds) *Modern Nutrition in Health and Disease*, pp. 305–327. Baltimore, MD: Williams & Wilkins.

Sporn MB, Roberts AB and Goodman DS (eds) (1994) *The Retinoids*, 2nd edn. New York: Raven Press.

Reverse Osmosis *See* **Membrane Techniques**: Principles of Reverse Osmosis; Applications of Reverse Osmosis; Principles of Ultrafiltration; Applications of Ultrafiltration

RHEOLOGICAL PROPERTIES OF FOOD MATERIALS

J Lefebvre, Institut National de la Recherche Agronomique, Nantes, France

Introduction

Rheological properties of ingredients and products are involved in all stages of food, from formulation to consumption. Their knowledge is therefore important to many aspects of food science and technology, such as modeling of processing operations (pumping, mixing, heat exchange, etc.), design of the processing and packaging machines, process control, characterization of ingredients, intermediary and end products, and instrumental evaluation of texture. They even concern nutrition, since foodstuffs have to be chewed and swallowed before being digested. The interest of food scientists in rheology is as old as rheology itself, and some of the pioneering work in this comparatively young discipline has been carried out on food materials.

During the last decade, the most active research areas in food rheology dealt with the characterization of the viscoelastic behavior of liquid, semiliquid, or gel food systems. Whereas flow properties are of prime importance in all aspects of processing of liquid materials, viscoelasticity measurements often constitute in practice the only feasible approach (along with microscopy) to obtain information on the structure of complex concentrated systems (solids and liquids) and on how it is affected by composition, or by physical, chemical, or biochemical treatments. They provide a unique tool to adapt formulations or to develop new engineered foods to meet the demands of changing technologies and evolving consumers' requirements.

However, rheometry on food systems has long been limited to the study of flow behavior (viscometry). Because of theoretical and practical difficulties, and especially because of the lack of affordable rheometers with high enough sensitivity, the viscoelastic properties of food systems have only been evaluated empirically, using an impressive variety of instruments and methods of testing, adapted to different specific types of materials or problems, but which did not allow the expression of measurements in terms of relevant physical variables because of the complexity of the deformations and of the mechanical sollicitation patterns they implemented. This situation has completely changed since the mid-1980s with the development of comparatively low-priced performing commercial rheometers. Viscoelasticity studies in shear are now within the reach of the average food science laboratory, opening up new, exciting perspectives. But this does not mean they are easy to carry out and to interpret, and a working knowledge of the essential concepts has not spread as fast and wide as the equipment. The first purpose of this article, focused on viscoelasticity, is to give an outline of these concepts.

Food materials display an amazing variety of rheological behavior. They can nevertheless be reduced to a relatively small number of basic types. The main one along the views outlined above corresponds to transient networks. It will be illustrated here with a few examples which will also be used to point out some

fundamental or practical problems which are especially important or critical in the case of food systems.

Only behaviors in shear (more precisely, in rotational geometries) will be considered, as shear is the most important type of deformation in rheological studies with regard to the orientations of this short article.

Solids and Liquids: Betwixt and Between

Rheology emerged as a science to describe, analyse, and rationalize the complex mechanical behavior displayed by most real materials, for which the classical Hookean elastic solid and the Newtonian viscous fluid represent only limiting behavior, which is seldom encountered in practice. Two types of deviation from the ideal behavior can occur, either separately or together: time-dependence and nonlinearity.

Let us revisit briefly these classical notions with the help of a rheological experiment which is the retardation test (creep and recovery): a stress σ is applied suddenly at time $t = 0$ and maintained till $t = \sigma$. Then, the sample is suddenly unloaded; the strain γ is recorded during the loading period (creep) and after the removal of stress (recovery); the results are expressed in terms of the compliance function $J(t) = \gamma(t)/\sigma$ in both parts of the test.

Viscoelasticity

In time-dependent behavior, the response depends on the duration of excitation. The general case is that of viscoelastic behavior, illustrated schematically in Figure 1c and d. The responses of the ideal elastic solid and of the ideal viscous liquid are shown for comparison (Figure 1a and b); they are instantaneous and defined by a constant value of the compliance $J_e = \gamma_e/\sigma$ (for the solid) or the compliance rate $dJ(t)/dt = 1/\eta$ (for liquid of viscosity η). To reach steady state (i.e. the equilibrium compliance J_e in the case of a viscoelastic solid, or a constant rate of

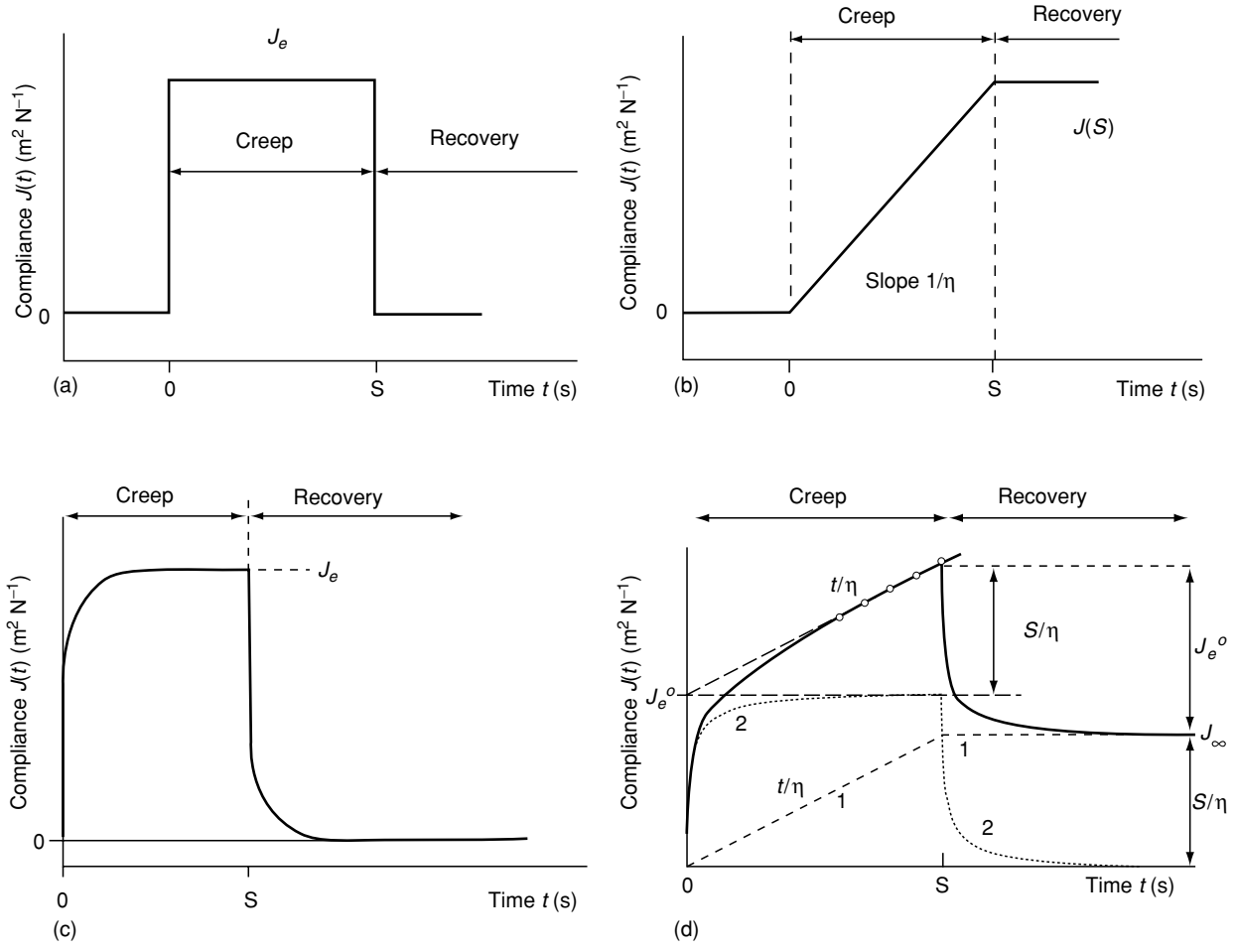

Figure 1 Schematic responses to a retardation test: (a) elastic solid; (b) viscous liquid; (c) viscoelastic solid; (d) viscoelastic liquid within the linearity range. Solid curve: measured total compliance. Curve 1 (dashed line): contribution of flow. Curve 2 (dotted line): recoverable compliance $J_r(t)$.

compliance $dJ(t)/dt$ in that of a viscoelastic liquid) requires a certain length of time after loading. And it is the same after unloading for compliance to reach a constant value. This value is zero in the case of a viscoelastic solid: deformation is entirely reversible, although not instantaneous. In contrast, for a viscoelastic liquid, the strain gathered during creep is progressively recovered after unloading, but only incompletely; the residual compliance J_∞ corresponds to the irreversible part of the deformation, i.e., to the contribution of flow to the total strain at the end of the creep period: $J_\infty = S/\eta$. Flow contribution to strain becomes dominant ($J(t)$ becomes a linear function of time) only after a creep time $t > t_m$; for $t < t_m$, elasticity dominates the response, although flow begins as soon as stress is applied to the material. By subtracting flow contribution t/η from the compliance in creep, the reversible contribution to the compliance is obtained. This contribution similar to the response of the viscoelastic solid is called the recoverable compliance $J_r(t)$; it is identical, in the case of linear behavior (see below), to the part of the final compliance $J(S)$ which is recovered during the recovery part of the test. The plateau value J_e° of $J_r(t)$ can be taken as a gauge of the elastic character of the liquid. The product $t_m = \eta J_e^\circ$ is the longest retardation time of the material and corresponds to t_m.

Such behavior simultaneously displays elastic and viscous characteristics. Elasticity dominates at short times of observation; the behavior appears to be increasingly viscous as the time scale of the experiment broadens. Thus, a given material can display either solid- or liquid-like behavior, depending on the time scale of the deformation process. This scaling is represented by the Deborah number $D = \tau/\Theta$, where τ is a characteristic time of material response (infinite for an elastic solid, zero for a viscous liquid) and Θ is a characteristic time of the experiment. A material will display a solid-like behavior (high D values) if it responds slowly (large τ) or if its response is observed at short time scales (small Θ). Although the distinction between solid and liquid subsists in theory for viscoelastic materials, it can be difficult to draw it in practice. Many materials which are felt as solids in common experience are indeed viscoelastic liquids.

Linearity and Nonlinearity

The notion of linearity, which simply amounts to the proportionality of the response to the solicitation for the ideal solid or liquid (J_e° or η independent of σ, respectively), has to be generalized to the time dependence for viscoelastic materials. The viscoelastic behavior is linear if the response to a sequence of arbitrary solicitations is identical to the sum of the responses to these stimuli if they were acting independently. This is the Boltzmann superposition principle which can be expressed as:

$$\gamma(t) = \int_{-\infty}^{t} J(t-t')\frac{d\sigma(t')}{dt'}dt' \qquad (1)$$

in terms of the stress history. As to the retardation test, linearity means that:

- $J(t)$ is independent of σ;
- The values of the steady-state limiting compliance J_e° and viscosity η are the same when obtained from the creep or from the recovery parts of the test;
- The curve of the recoverable compliance against the recovery time ($t-S$), obtained from recovery, superimposes on that of the recoverable compliance against creep time, obtained by subtracting the contribution of flow from the creep compliance.

Most real materials behave linearly only up to some limit of strain or strain rate limit, which is a material property. Whilst handling and processing the material generally involves solicitations which exceed this limit (which may be fairly low), characterization of its viscoelastic behavior within the linearity range is a prerequisite for studies in the nonlinear regime. In the linear regime, the results of any of the classical rheological tests are independent of the amplitude of the solicitation and they can also be used to predict the behavior of the material in the other tests. Beyond the linearity range, this is no longer the case.

Classical Rheological Tests

The characterization of the rheological behavior of a material requires the study of its viscoelasticity and, in the case of a liquid, that of its steady flow behavior. The rheologist uses very simple 'system of loads' patterns and deformations. Transient regime and dynamic tests are used to study viscoelasticity. Flow behavior is investigated using viscometry.

Very often, the study of viscoelasticity is restricted to the linear range, for the reasons hinted at above. Although within the linearity domain all viscoelasticity tests are equivalent in theory and their results can be interconverted using theoretical relations or semi-empirical approximation methods, their respective applications differ in practice according to the materials and the instruments and the experimentalist should view them as complementary.

Transient Regime Tests

We have already given the principle of the retardation test (creep and creep recovery). This test corresponds to the 'natural' operating mode of stress-controlled rheometers. The transient test with strain-controlled

instruments is the (stress) relaxation test: at time zero, the sample is suddenly deformed to a given value of strain which is then kept constant while the decay of stress with time is recorded. The experiment is analyzed using the relaxation modulus function $G(t) = \sigma(t)/\gamma$, which does not depend on the strain in the linear domain of behavior. The modulus and the compliance functions are linked in the linear regime by the integral equation:

$$t = \int_0^t J(t-u)G(u)du = \int_0^t J(u)G(t-u)du \quad (2)$$

so, $G(t) \neq 1/J(t)$, except in the plateau regions of these functions. Ideal elastic solids store integrally and indefinitely the mechanical energy imput: stress does not relax. For viscoelastic liquids, stress relaxes gradually and vanishes after a long enough time; at intermediate times, a region, called the viscoelastic plateau G_N°, where the decrease of the modulus slackens momentarily, is frequently observed on bilogarithmic scales (Figure 2a). Viscoelastic solids do not relax completely: at long time values, the modulus approaches a constant value G_e, the equilibrium modulus (Figure 2b).

Stress relaxation tests are not easy to perform on liquids with low elasticity or short characteristic times. In such cases, it is preferable to submit the material to a sudden constant shear rate $\dot{\gamma} = d\gamma/dt$ until an equilibrium value σ_e of stress is reached, after which the shear rate is suddenly stopped while stress continues to be recorded (Figure 3). Both parts of this test can be analyzed in term of relaxation modulus:

$$G(t) = (1/\dot{\gamma})|d\sigma(t)/dt| \quad (3)$$

The retardation test is particularly interesting for viscoelastic liquids since the viscoelastic response can be very simply separated from the steady flow contribution (Figure 1d), in contrast to the relaxation test; the latter is better suited for solid materials.

Transient tests allow materials with very long response times to be studied; the upper time limit for the experiments only depends on the stability of the material and the patience of the operator. But since in practice the establishment or the release of a constant value of the stress, strain, or strain rate takes some time, transient tests do not measure correctly the response at short times. Besides, the response often shows at starts and stops transient damped oscillations due to mechanical coupling between the sample and the measuring device. These methods are applicable to materials with characteristic times longer than, say, 1 min.

Transient tests usually require relatively large 'system of loads' to give an accurately analyzable response; this may be a problem for materials with a low linearity limit.

Dynamic Measurements

Dynamic measurement consists of submitting the sample to a sinusoidal 'system of loads' at an angular frequency ω rad s^{-1}. When stress is imposed, strain is measured, and conversely: this type of test can be performed either with stress- or with strain-controlled

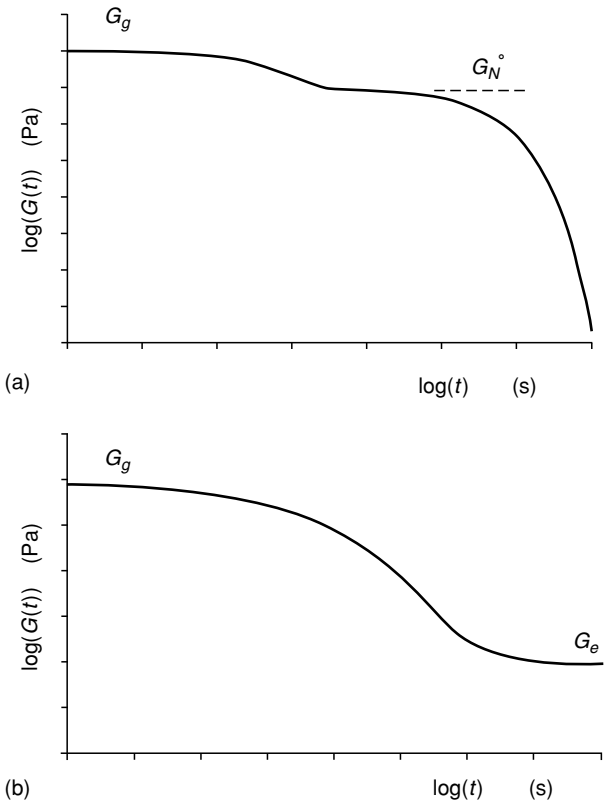

Figure 2 Schematic responses to a stress relaxation test: (a) viscoelastic liquid; (b) viscoelastic solid.

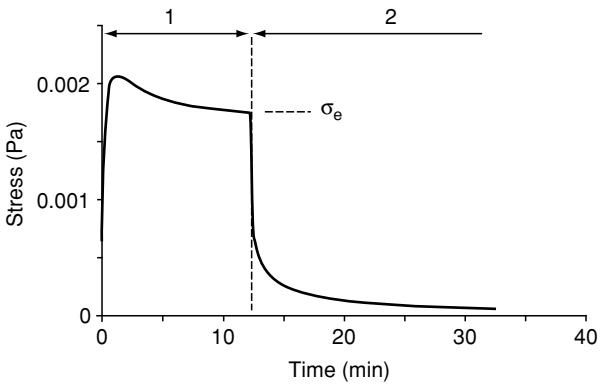

Figure 3 Response of a 3% bovine serum albumin solution (isoionic conditions) to the sudden application of a shear rate of 0.05 s^{-1} over 12 min (step 1) followed by its sudden stop (stress relaxation: step 2).

instruments. Within the range of *linear* viscoelastic behavior, the response of the material is also sinusoidal with the same frequency as the excitation, but out of phase. If a strain $\gamma = \gamma_o \cos(\omega t)$ is imposed, the stress will be $\sigma = \sigma_o \cos(\omega t + \delta)$; the phase angle δ is zero for an ideal solid and $\pi/2$ for an ideal liquid. The results are often expressed in terms of the storage and loss moduli $G' = (\sigma_o/\gamma_o)\cos(\delta)$ and $G'' = (\sigma_o/\gamma_o)\sin(\delta)$, which are the real and 'imaginary' parts of the complex modulus $G^*(\omega) = G'(\omega) + iG''(\omega) = \sigma(t)/\gamma(t)$, respectively. The viscoelastic behavior will be described by the variations of G' and G'' with ω (the mechanical spectrum of the material); however, the two moduli are not independent over the frequency domain. Alternatively, the results can be expressed in terms of the real and 'imaginary' components of the complex compliance, or (for liquids only) of the complex viscosity: $J' = G'/(G'^2 + G''^2)$, $J'' = G''/(G'^2 + G''^2)$, and $\eta' = G''/\omega$, $\eta'' = G'/\omega$. Although these representations contain strictly the same information as G', G'', they can prove more interesting for viscoelastic liquids because they weigh dissipative processes differently on the frequency scale. Qualitatively, a measure at frequency ω in a dynamic test is equivalent to a measure at time $t = 1/\omega$ in a transient test.

Figure 4a and b show the schematic theoretical mechanical spectra of a viscoelastic liquid and a viscoelastic solid, respectively. Whereas $G'(\omega)$ is necessarily a monotonously increasing function, $G''(\omega)$ shows maxima which correspond to concentrations of dissipative processes in the material over certain frequency regions. Two loss peaks are visible on the liquid-type spectrum. They encompass an intermediate frequency range where $G' > G''$ and keeps a nearly constant value G_N° – the viscoelastic plateau; at this time scale, when energy dissipation is showing a relative minimum, the material is transiently quasielastic. On both sides of the plateau, $G'' > G'$ and both moduli are strongly frequency-dependent. At high frequencies, G' and G'' first both increase approximately as ω^a with $a \sim 0.5$ (the transition zone); then G' tends towards its very high 'glassy' limit G_g, whereas G'' drops: the 'glassy behavior' corresponds to solicitations which are too fast for the structural elements in the material to respond. Below the viscoelastic plateau, $G' \propto \omega^2$ and $G'' \propto \omega$; this is the terminal region of the mechanical spectrum, corresponding to flow. For viscoelastic solids, the terminal region is not observed; at low enough frequencies, G' reaches a constant value – the equilibrium modulus G_e. Frequently, the loss peaks overlap strongly; then the viscoelastic plateau reduces to an inflection in G'.

Dynamic measurements are very popular. They are easy to perform, accurate even with fairly fluid materials, and cause minimal perturbation to the sample since the amplitude of strain is controlled and can be kept to quite low values (minimum value $\sim 0.1\%$ on strain-controlled instruments, but $\sim 1\%$ on stress-controlled ones). They permit the study of materials with a low linearity limit.

However, some limitations of dynamic tests should not be overlooked:

- Beyond the linearity limit, the response signal is no longer sinusoidal. The calculation of the real and 'imaginary' components of the rheological functions, as outlined above, performed automatically by the rheometer softwares, is then invalid.
- The frequency range is restricted for practical reasons. Because of mechanical problems, rotational rheometers cannot operate above 200–500 rad s^{-1}, but the upper limit may be as low as ~ 10 rad s^{-1} when rather fluid materials are tested. Instrument softwares generally permit operation down to 10^{-4} rad s^{-1}, but below 10^{-3} rad s^{-1} measurements take an exceedingly long time. The practical range of operation is thus 10^{-3}–100 rad s^{-1}, corresponding to the 10^{-2}–10^{-3} s^{-1} time scale, extending well below the lower limit of transient

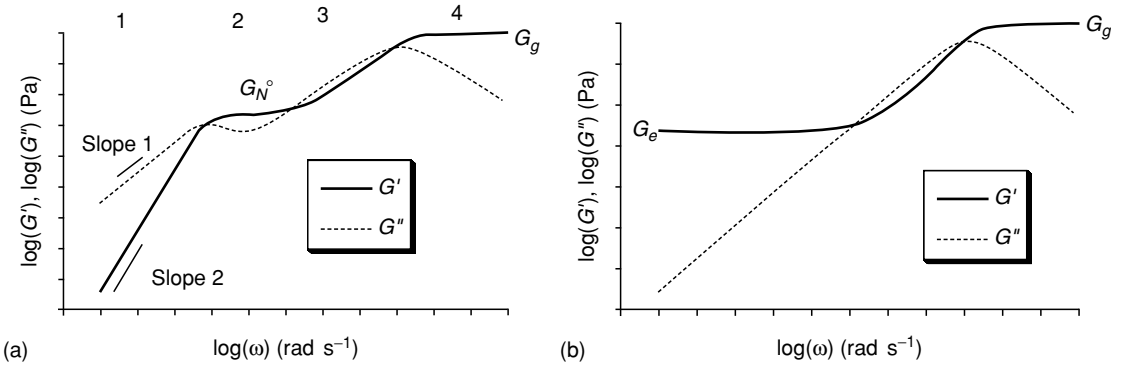

Figure 4 Dynamic measurements. Schematic mechanical spectrum of: (a) a viscoelastic liquid: 1, terminal zone; 2, viscoelastic plateau; 3, transition zone; 4, glassy zone; (b) a viscoelastic solid.

tests but stopping short for many materials, as will be shown later.

Viscometry

Flow behavior is usually investigated through viscometric experiments: a shear rate (or stress) is imposed on the liquid and the resulting response (shear stress or rate) is measured; viscosity is defined as $\eta = \sigma/\dot{\gamma}$. Viscometry is primarily concerned with the steady-state response.

Simple liquids and solutions of low-molecular-weight molecules are Newtonian (constant viscosity) over the wide shear rate range of practical interest, which extends roughly from $\sim 10^{-4}$ s^{-1} (sedimentation) to 10^{4} s^{-1} (spraying); chewing and swallowing or pumping and mixing involve intermediate typical shear rate values (10^{1}–10^{3} s^{-1}). But for most materials, viscosity depends on the shear rate (or stress), although generally it tends to be Newtonian at low enough shear rates, and in many applications non-Newtonian flow properties are important.

The usual behavior is shear-thinning: as $\dot{\gamma}$ increases, η decreases from an initial Newtonian plateau η_o to a final one η_{oo}; but one or both of them may often be out of the experimental window (Figure 5). Shear-thickening behavior is exceptional in food systems.

Non-Newtonian flow curves are usually recorded by applying shear rate (or stress) ramps. However, artifacts can result from the inertia of the measuring device or from the viscoelasticity of the sample. It is better to apply a series of shear rate (or stress) steps, waiting at each step for the steady-state response. Performing retardation tests under different stress values is almost the only method applicable to highly viscoelastic liquids (Figure 6) or to very viscous inelastic ones; it allows viscosity measurements at very low shear rates.

Many relatively inelastic shear-thinning liquids display viscosities which also depend on the duration of shear. When viscosity decreases during shear, the behaviour is said to be thixotropic; the reverse case is called antithixotropy; the changes may be completely or partially reversible. Thixotropy is a time-dependent behavior which in principle is to be distinguished from viscoelasticity. But in practice, these two properties may be difficult to differentiate; besides, thixotropic materials often show appreciable

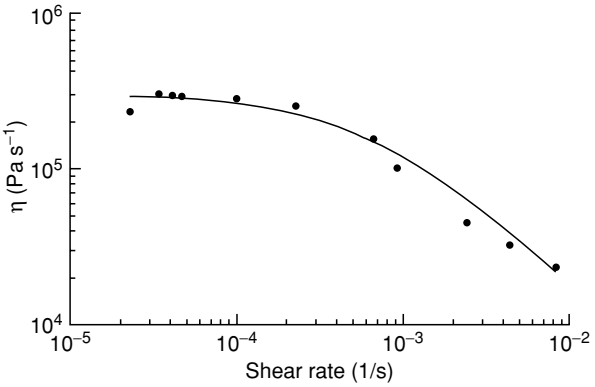

Figure 6 Flow curve of a toothpaste determined from the analysis of creep tests carried out at different values of the applied stress.

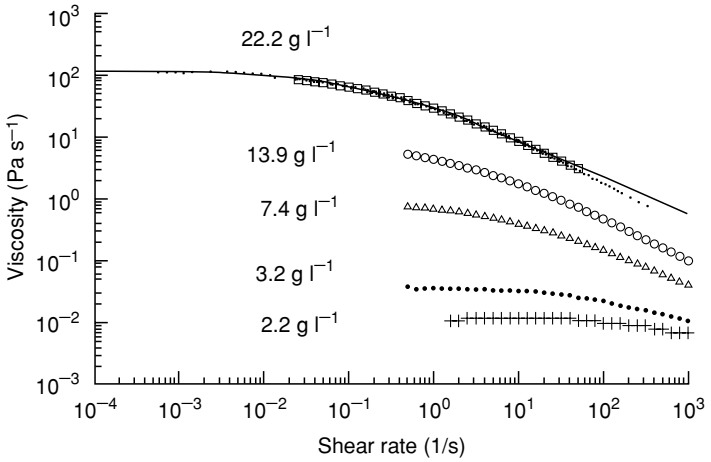

Figure 5 Shear-thinning behavior. Flow curves of solutions of a high-viscosity hydroxyethylcellulose sample at different concentrations. Above a concentration of ~ 2 g l^{-1}, the solutions exhibit typical shear-thinning behavior. The master curve (dots) is built by collapsing the different curves on the 22.2 g l^{-1} one by shifting them along the two axes. The continuous line shows the fit of Cross equation: $(\eta - \eta_\infty) = (\eta_o - \eta_\infty)/[1 + (\dot{\gamma}/\dot{\gamma}_c)^m]$ to the master curve data, where η_o = initial Newtonian plateau viscosity; η_∞ = high shear rate Newtonian viscosity (taken as zero here since it is negligible compared to η); $\dot{\gamma}_c$ = critical shear rate; and m = dimensionless exponent. The fit gives $\eta_o = 118$ Pa s^{-1}, $\dot{\gamma}_c = 0.15$ s^{-1} and $m = 0.60$.

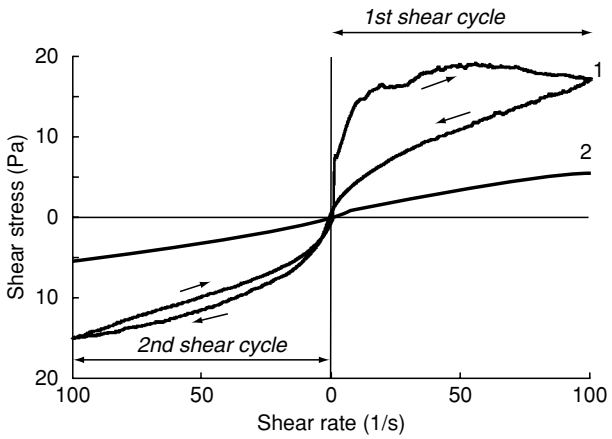

Figure 7 Thixotropic loops. Application of two successive shear cycles (0–100–0 s^{-1}) to a guar gum galactomannan/xanthan mixture (ratio 90/10 w/w) in water at a total polysaccharide concentration of 0.5% w/w (curve 1). The first cycle causes drastic irreversible damage to the structure of the mixed system. A residual slight thixotropic behavior is observed during the second cycle and remains unchanged upon subsequent cycles.

No thixotropy is displayed by guar gum solutions or xanthan solutions; the results for a 0.5% guar gum galactomannan solution are given for comparison (curve 2).

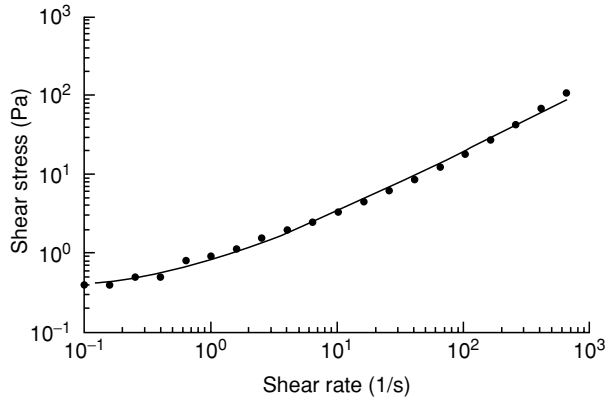

Figure 8 Apparent yield stress behavior. Flow curve recorded at 60 °C of a wheat starch paste prepared in the presence of sodium caseinate (concentrations 6% starch, 7.5% caseinate). The continuous line is the fit of the Herschel–Bulkley equation: $\sigma = \sigma_o + k\dot{\gamma}^n$; it gives an apparent yield stress value $\sigma_o = 0.34$ Pa.

viscoelasticity at the same time. When the material is submitted to a shear rate cycle from zero to a maximum and back to zero, a thixotropic loop is observed (Figure 7). This is a popular way to spot the existence or the apparition of thixotropy, but it only allows comparative work; moreover, there are risks of confusion with inertia effects or with the manifestation of viscoelasticity, which also shows as loops. The responses to sequences of step changes of shear rate are used for thorough studies, but thixotropy is a difficult topic.

Yield stress behavior or plasticity, another type of complex flow properties, has been the subject of much discussion. Plastic material would not flow until the stress has exceeded a yield value σ_o, below which viscosity is infinite (solid behavior). Many materials show apparent yield stress behavior over limited ranges of stress or shear rate (Figure 8). However, whenever it has been possible to go down to low enough stresses or shear rates (using creep test, for instance), such materials were found to flow, but with the peculiarity that viscosity drops from very high values to relatively low ones over quite a narrow range of stress in the region of low shear rates (Figure 9).

Problems in Rheological Characterization of Food Systems

Structural Basis of Viscoelasticity in Food Systems

Viscoelasticity is the manifestation of the internal structure of the material, i.e., the distribution in space of the structural elements and their interactions. Thixotropy and plasticity derive from flow-induced large-scale structural changes.

Structural elements can be molecules, macromolecules, or particles (solid particles, liquid droplets, gas bubbles), or associations of such elements (aggregates, micelles, etc.) and their combinations. The two simplest cases to consider are the solutions of polymer-like chain macromolecules (polysaccharides, gelatin) and the dispersions of (quasi-)spherical particles (including globular protein solutions). When dilute, such systems show no appreciable viscoelasticity in practical conditions – the longest relaxation time of a macromolecular chain or of an emulsion droplet is shorter than $\sim 10^{-2}$ s. Beyond a certain polymer concentration or disperse phase volume fraction in the system, an elasticity of entropic nature develops due to topological constraints between macromolecular chains (entanglements), or from crowding effects between the particles which result in a local order. Semidilute polysaccharide solutions should normally display this type of viscoelasticity. This is indeed the case for guar gum galactomannan solutions; their mechanical spectrum (Figure 10) shows in the upper region of the experimental frequency range the beginning of the viscoelastic plateau due to entanglements. But locust bean gum galactomannan (LBG) solutions progressively depart from this behavior upon aging, with the development of a viscoelastic plateau at low frequencies (Figure 11): the system slowly changes into a gel. This reflects the establishment of relatively long-lived linkages between the chains, probably due to hydrophobic interactions between the 'smooth' regions (devoid of

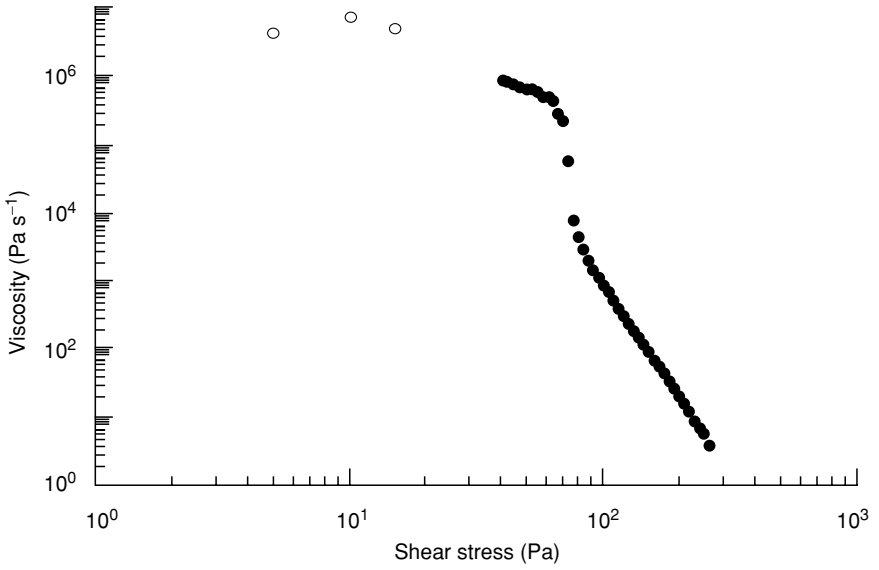

Figure 9 Apparent yield stress behavior. Flow curve of a commercial mayonnaise. The filled circles correspond to the results recorded during the imposition of a shear rate ramp, the empty circles to data obtained by application of retardation tests. The latter data demonstrate unequivocally that the behavior is in fact shear-thinning.

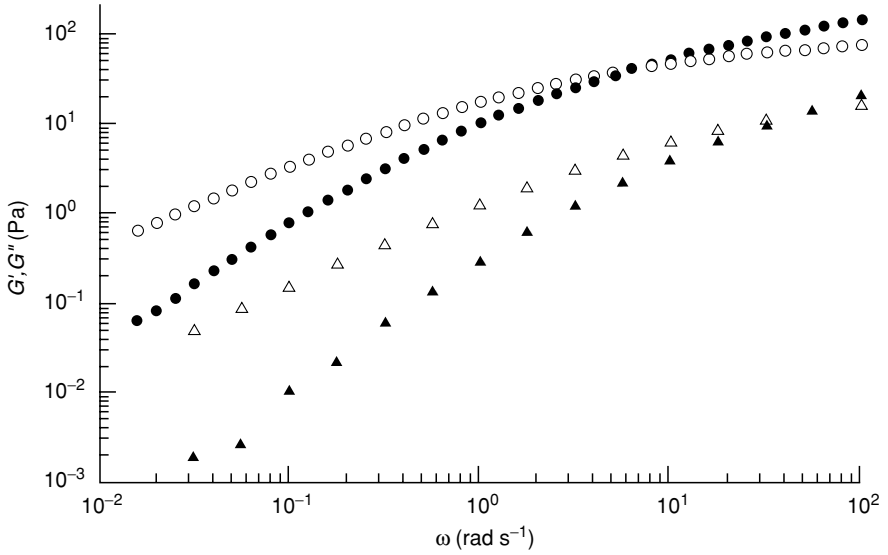

Figure 10 Typical mechanical spectra of biopolymer solutions: guar gum galactomannan solutions in water (20 °C) at two concentrations: 1.87% w/w (circles) and 1% w/w (triangles). Filled symbols, storage modulus G'; open symbols, loss modulus G''.

galactose groups) of the mannose backbone, which do not exist in the case of guar. Similarly, a 3% bovine serum albumin (BSA) isoionic solution departs from the purely viscous and Newtonian behavior expected for dilute hard sphere dispersion (Figure 3), indicating that protein molecules are strongly interacting.

Because of their chemical structure and of the specific properties of water, food macromolecules generally interact through hydrogen and hydrophobic interactions. As a result, viscoelasticity of the transient network type (noncovalent gels) can be observed at relatively low concentrations. Transient networks also result from partial destabilization (flocculation) of suspensions or emulsions due to colloidal interactions.

Monitoring Structural Transitions

The example of LBG solution (Figure 11) showed that viscoelastic measurements can be used to follow structural changes in a system. In this case, the change

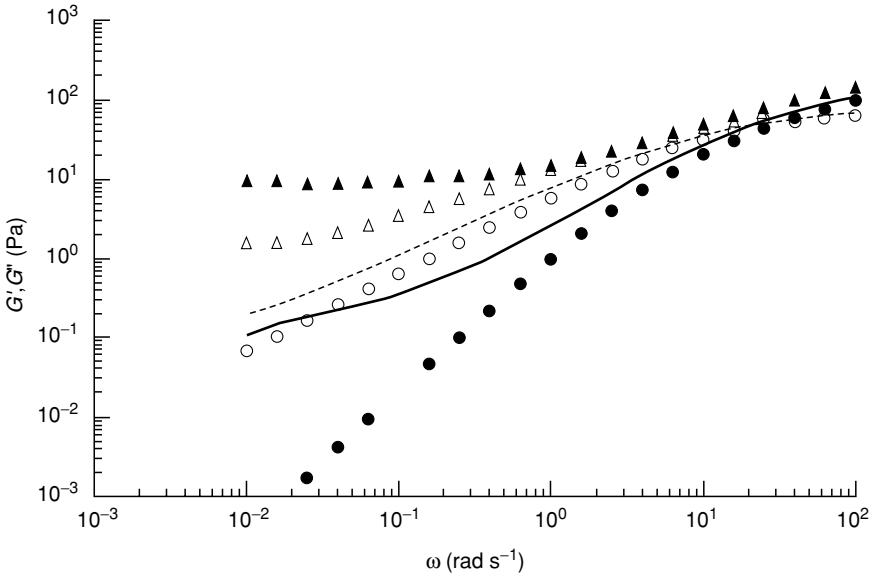

Figure 11 Progressive departure from the solution state during aging of a locust bean gum galactomannan solution (concentration 10 g l^{-1}; 20 °C). Mechanical spectra of the freshly prepared solution (circles), of the system after 6 days' aging (lines), and of the system after 15 days' aging (triangles). Filled symbols and solid line, G'; open symbols and interrupted line, G''.

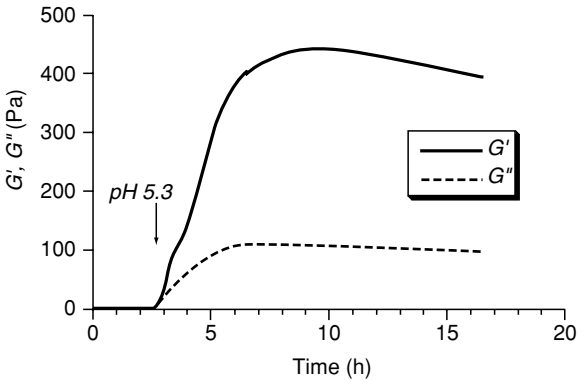

Figure 12 Monitoring lactic acid bacteria milk clotting using dynamic measurements at fixed frequency (40 °C, $\omega = 6$ rad s^{-1}, strain amplitude $\gamma_o = 0.02$). The gelation starts after about 150 min fermentation, when the pH has decreased to 5.3. The decline in the G' curve after ~11 h is an artifact due to syneresis. The experiment was performed using coaxial cylinders to minimize the incidence of syneresis; with cone–plate or plate–plate geometries, syneresis causes slippage and its effect is seen much earlier, even when measuring devices with striated or sanded surfaces are used.

was slow enough to allow viscoelasticity to be characterized over a certain frequency range at different steps of its evolution. However, structural changes are generally rather fast, and once started, can rarely be stopped at will. Single-frequency dynamic measurements, since the time required for each individual measure is short provided that $\omega > 0.01$ rad s^{-1}, are very precious indeed to monitor the kinetics of structure changes, as exemplified in Figure 12 for milk clotting, an instance of colloidal gel formation. But they do not characterize rheologically the evolution of the system. Mechanical spectra can be obtained in the classical way only when the system remains practically unchanged during the time necessary to record them. However, advanced techniques which are outside our scope require much shorter times than with the classical frequency sweep, and allow study of systems undergoing relatively fast structural changes.

In classical follow-up experiments at fixed frequency, the choice of frequency may be critical. No change would be detected in LBG solutions if $\omega = 10$ rad s^{-1} were selected to monitor aging (Figure 11)!

The Need to Enlarge the Experimental Window

Figure 13 shows the mechanical spectra of food systems we know to differ largely in nature and in texture or consistency. Nevertheless, these spectra are qualitatively similar; those of the commercial mayonnaise and the 61 mg ml^{-1} BSA heat-set gel even superimpose. They show the viscoelastic plateau region of transient networks of the colloidal gel type, with gently ascending moduli curves and with G' higher than G'' over the frequency window but relatively large G''/G' ratios (0.1–0.5); in the case of the gluten sample, however, the upper end of the plateau is visibly approached. Beyond this assignation to a general type of structure, such flat spectra allow only a rough estimation of the plateau modulus. But quantitative comparison can be misleading, since the viscoelastic plateau often extends beyond both limits

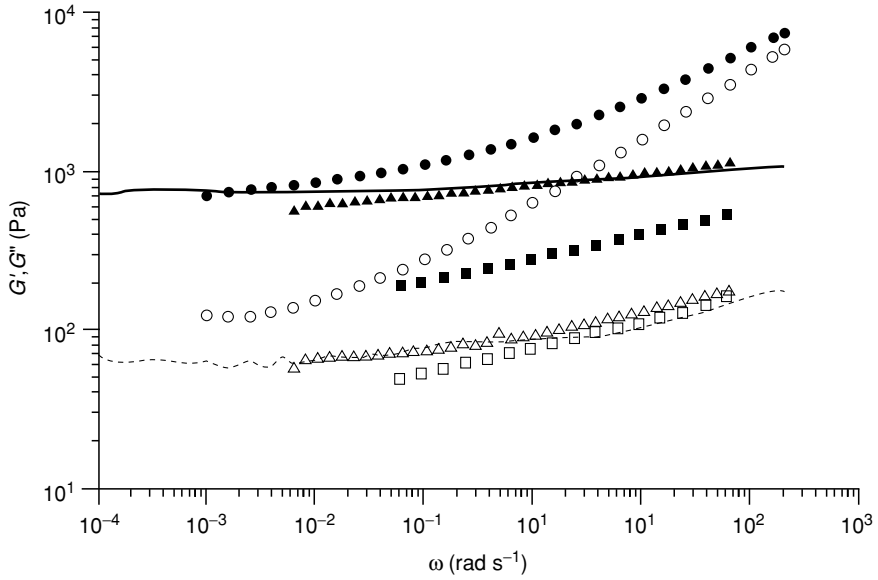

Figure 13 Mechanical spectra obtained at 20 °C by dynamic measurements on various systems of the colloidal gel type. Circles: fully hydrated gluten from a commercial wheat cultivar. Triangles: heat-set gel of bovine serum albumin (61 mg ml^{-1}, pH 7, 0.05 mol l^{-1} NaCl). Squares: experimental yogurt. Lines: commercial mayonnaise. Filled symbols and solid line: G'; empty symbols and interrupted line: G''.

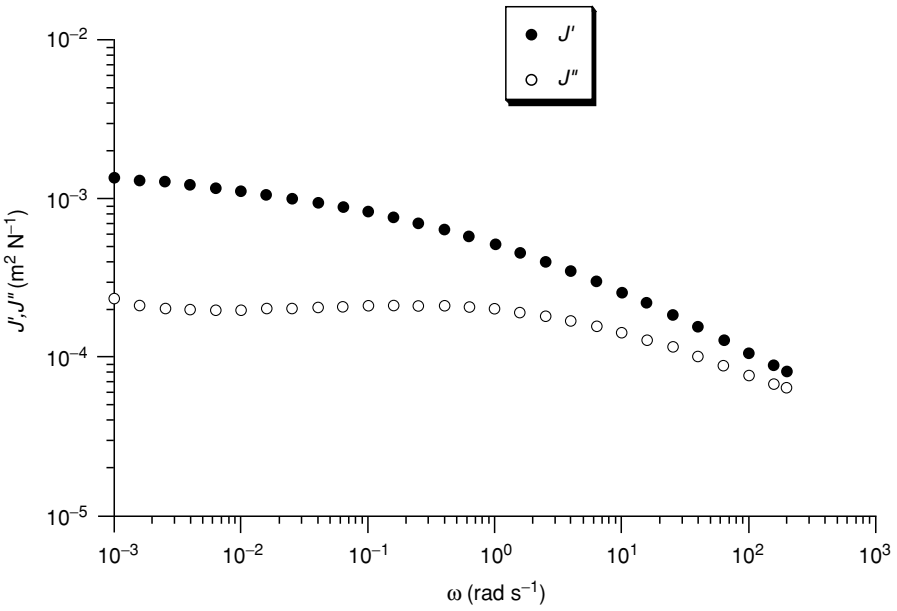

Figure 14 Mechanical spectrum of the same gluten as shown in Figure 13, but plotted as J' and J'' versus angular frequency. The curve of J'' shows a peak with a maximum close to $\omega = 0.16$ rad s^{-1}; it bounds the viscoelastic plateau on the high-frequency side.

of the frequency window of dynamic measurements. Plotting the data as $J'(w)$, $J''(w)$ is sometimes more telling, as shown in Figure 14 for the same gluten as in Figure 13; $J''(w)$ displays a maximum at the right of the frequency range, which at least allows location of the upper end of the viscoelastic plateau on the time scale and possibly opens the way to modeling this region of the behavior.

Enlarging the time scale of observation is required anyway for a full-scale rheological characterization. In the case of classical polymer systems, the time–temperature superposition principle, which means that a change in the temperature of the system is equivalent to a shift of the time scale of the behavior, allows one to describe the viscoelastic behavior over a very large frequency range from data obtained over

a restricted one at different temperatures. This principle does not in general apply to food systems since their elasticity is rarely entropy-driven; anyway, the range of temperature which could be covered is quite narrow because of the presence of liquid water. One has then to combine dynamic measurements with transient tests which permit one to extend the observation window to longer times. Let us give a few illustrations.

A gluten and a reconstituted dough are compared in Figure 15. The mechanical spectra of a gluten and of a 'synthetic' dough prepared by mixing it with starch do not differ much on the 10^{-1}–10^{2} rad s^{-1} range (Figure 15a). The moduli are somewhat larger for dough, because starch granules act as a filler; in fact, their presence also makes dough behavior extremely nonlinear and the dynamic data are only approximate. But when the materials are submitted

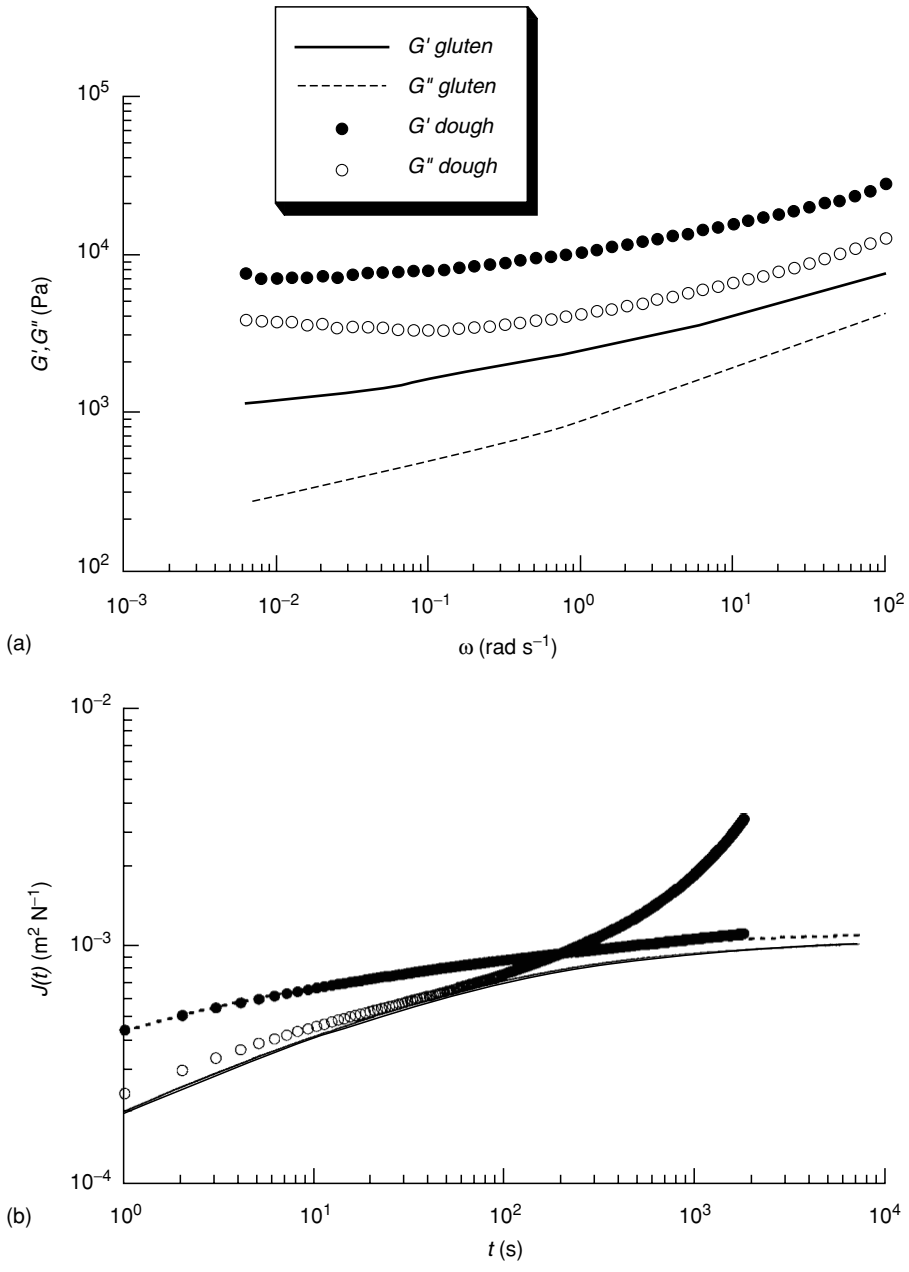

Figure 15 Viscoelastic behavior at 20 °C of a reconstituted dough compared to that of the gluten used to prepare it. (a) Mechanical spectra obtained by dynamic measurements under 0.05 strain amplitude over the 10^{-2}–10^{2} rad s^{-1} frequency range. (b) Creep curves under 350 Pa stress. Symbols: creep compliance $J(t)$; lines: recoverable compliance $J_r(t)$. Filled symbols and dotted line, gluten; empty symbols and solid line, dough.

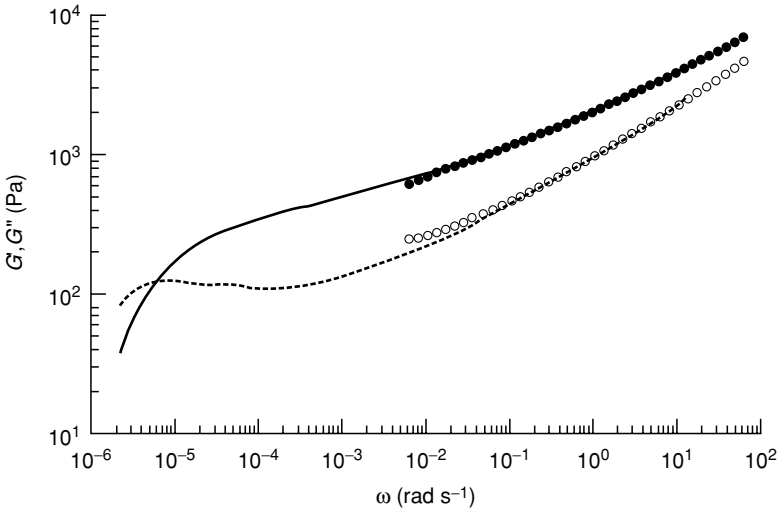

Figure 16 An example of a combination of the retardation test with dynamic measurements in order to describe the viscoelastic behavior of a material over a large time scale. The material is a fully hydrated gluten from an experimental wheat line. Mesurements were performed within the linearity range at 20 °C. The data from the recovery part of the retardation test were converted from the time to the frequency domain using an approximation method. Symbols: dynamic measurements: filled symbols, G'; empty symbols, G''. Lines: recovery data: solid line, G', dotted line, G''.

to a retardation test in the same conditions, the difference becomes striking after 100-s creep (Figure 15b); analysis of the results shows that whereas J_e° is about the same ($\sim 10^{-3}$ m^2 N^{-1}), viscosity is much smaller for dough ($\sim 5\ 10^5$ Pa s^{-1}) than for gluten (10^7 Pa s^{-1}) at comparable shear rates (7×10^{-4} versus 2×10^{-4} s^{-1}).

When the tests are performed within the linearity range, it is possible to convert the data of relaxation or retardation tests from the time to the frequency domain. Combining the data of the retardation test with those of dynamic measurements allowed us to obtain the mechanical spectrum of a weak gluten sample over about eight logarithmic frequency decades (Figure 16). The viscoelastic plateau extends over more than seven frequency logarithmic decades, the terminal region of the spectrum being entered only below $\sim 10^{-5}$ rad s^{-1}. Extended viscoelastic plateaus result from strong interactions and multilevel structures within the material. Such cases are not exceptional in food systems and highlight the fact that rheology is a matter of time.

See also: **Colloids and Emulsions**; **Emulsifiers**: Organic Emulsifiers; Phosphates as Meat Emulsion Stabilizers; Uses in Processed Foods; **Gums**: Properties of Individual Gums; Nutritional Role of Guar Gum; **Pectin**: Properties and Determination; Food Use; **Protein**: Sources of Food-grade Protein; Quality; Heat Treatment for Food Proteins; **Starch**: Structure, Properties, and Determination; Sources and Processing; Functional Properties.

Further Reading

Axelos MAV, Renard D, Chevillard C and Lefebvre J (1999) Viscoelastic properties of some biopolymer systems in relation with their nano-structure. In: Stokke BT and Elgsaeter A (eds) *Wiley Polymer Group Review Series*, vol. 2, pp. 330–342. New York: Wiley.

Barnes HA (2000) *A Handbook of Elementary Rheology*. Aberystwyth: Institute of Non-Newtonian Fluid Mechanics.

Barnes HA, Hutton JF and Walters K (1989) *An Introduction to Rheology*. Amsterdam: Elsevier.

Ferry JD (1980) *Viscoelastic Properties of Polymers*, 3rd edn. New York: Wiley.

Hill SE, Ledward DA and Mitchell JR (1998) *Functional Properties of Food Macromolecules*, 2nd edn. Gaithersburg: Aspen.

Macosko CW (ed) (1994) *Rheology. Principles, Measurements, and Applications*. New York: VCH.

Marin G (1988) Oscillatory rheometry. In: Collyer AA and Clegg DW (eds) *Rheological Measurement*, pp. 297–343. London: Elsevier.

Powell RL (1988) Rotational viscometry. In: Collyer AA, Clegg DW (eds) *Rheological Measurement*, pp. 247–296. London: Elsevier.

Rao MA and Steffe JF (eds) (1992) *Viscoelastic Properties of Foods*. London: Elsevier.

Ross-Murphy SB (1994) Rheological methods. In: Ross--Murphy SB (ed) *Physical Techniques for the Study of Food Biopolymers*, pp. 343–392. London: Blackie.

Whorlow RW (1992) *Rheological Techniques*, 2nd edn. Chichester: Ellis Horwood.

RHEOLOGY OF LIQUIDS

A E Bell, University of Reading, Whiteknights, Reading, UK

Background

Food materials are multiphase, multicomponent systems whose rheological characteristics frequently present process engineers with difficult handling and pumping problems.

The food industry has, almost since its inception, been concerned with the physical properties of its products, in terms of both their processing and their eating characteristics. Consequently, considerable care must be taken in designing plant and process if the texture, consistency, and quality of the final product are not to be compromised. The selection of appropriate pumps and conveying and mixing systems is especially germane to food materials that may be shear-dependent and/or abrasive in nature. (*See* **Rheological Properties of Food Materials**.)

Although it is a relatively easy matter to describe mathematically the flow behavior of 'simple' liquids, food solid/liquid mixtures present unique problems in describing their flow behavior.

Liquid Behavior

When assessing the behavior of a liquid system, it is usual to investigate the stresses generated when the sample is sheared, that is to say, the force per unit area produced when the 'layers' of liquid move over each other in the direction of flow. This gives rise to the classical definition of a shear viscosity (Figure 1):

Shear rate, $\gamma = V/h$ (s^{-1})
Shear stress, $\sigma = F/A$ (Pa)
Shear viscosity, $\eta = \sigma/\gamma$ (Pa s),

where V is the velocity, F is the force, and A is the area over which it acts.

The range of shear rates observed in the food industries is vast, from values as low as $10^{-2}\,s^{-1}$, for the self-leveling of a liquid due to surface tension, to $10^{6}\,s^{-1}$ seen in the spray drying of slurries. (The chewing and swallowing of foods generate shear values of about 10^{1}–$10^{2}\,s^{-1}$.) This presents us with considerable difficulties with regard to most food systems, as very few behave in an ideal or Newtonian manner, with the shear stress proportional to the shear rate over all of the possible shear rate values.

Since viscosity is essentially related to shear rate and stress, it is possible to operate a viscometer in one or two modes: (1) by applying a rotation and measuring the torque generated (controlled stain), or (2) by applying a fixed torque and measuring the speed generated (controlled stress). These two methods have their advantages and disadvantages; for example, the controlled-strain instruments are relatively cheap, robust, and well understood but usually have a limited shear range (1–1000 s^{-1}). Controlled-stress instruments, however, have large shear rate ranges and can carry out other forms of testing (creep, oscillatory, etc.) but do tend to be somewhat expensive.

Working Liquids

When we work a liquid system, we apply energy to it. This 'force' has a number of characteristics that we can define: (1) the area over which it is applied, (2) the distance over which it acts, and (3) the rate at which it is applied. We may then define the stresses involved (force/area) and the strains produced (relative deformation). If a more complex analysis is required, these forces may be resolved further into three components, the tensile or normal forces, perpendicular to the plane of the applied force, and two shear or tangential forces parallel to it.

The strain (distance moved) when processing or measuring a sample has a considerable effect on its mechanical behavior/flow properties. If a sample is moved such that its molecular domains or structural elements no longer interact/overlap, the network may become 'overloaded' and ultimately break down (Figure 2). For example, a sample of an emulsion

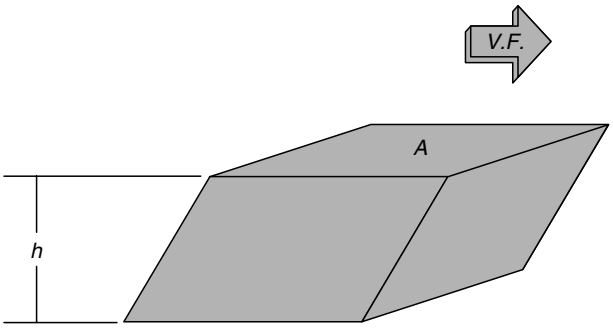

Figure 1 Classical shear viscosity. Reproduced from Rheology of Liquids, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

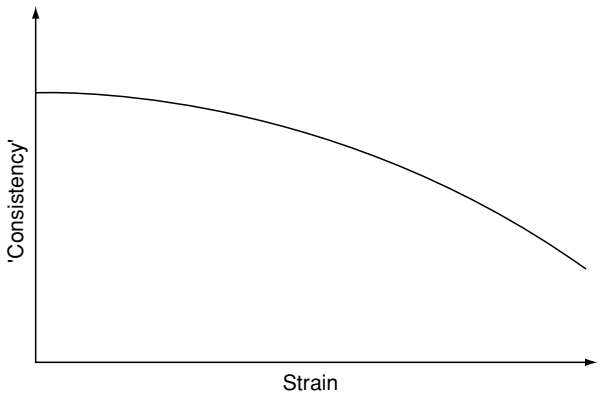

Figure 2 Effect of increasing strain. Reproduced from Rheology of Liquids, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

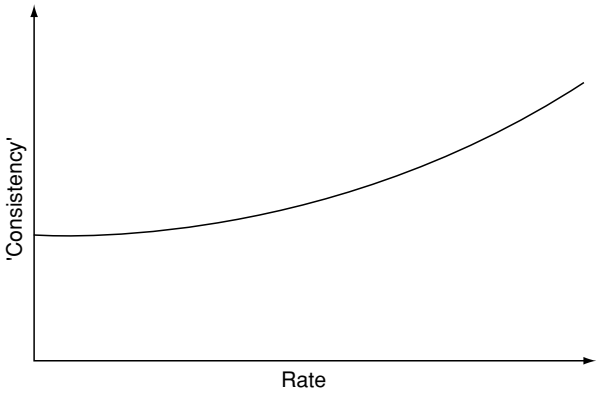

Figure 3 Increasing interaction with shear rate. Reproduced from Rheology of Liquids, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

such as mayonnaise, which is stabilized by electrostatic or steric means, may have a very low value of 'critical strain,' after which structural damage to the emulsion will occur, and it is very important that testing and processing be carried out in such a way so as not to exceed this point. (*See* **Colloids and Emulsions.**)

While, in general, increasing the strain tends to reduce interactions, increasing the shear rate may give rise to either increases or decreases in the apparent degrees of interaction. If we consider the effect of increasing the shear rate on entangled long-chain polymer systems (e.g., a long-chain polysaccharide), at zero shear rate, these chains will be entangled by Brownian motion and will have a degree of interaction based on physical constraints. On moving one of the chains slowly with respect to the others, very little resistance would be encountered, as the molecules would have sufficient time to allow any rearrangements needed to take place. However, if the same process were carried out at a higher rate, the entanglement would have insufficient time for rearrangement and would resist such movement. This would produce a material that appeared to increase its degree of interaction at higher shear rates. Such a system may become unworkable if, for example, one attempted to pump it too quickly (Figure 3). (*See* **Carbohydrates:** Classification and Properties.)

However, if a system were stabilized by, for example, some form of hydrogen bonding, an increase in shear rate might reduce the effectiveness of the interactions. This would lead to a fall in the apparent consistency (Figure 4).

The two parameters, strain and rate, when combined give an overall picture of the behavior of the

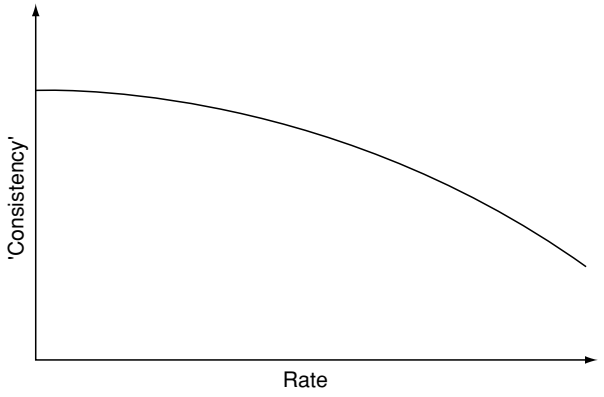

Figure 4 Decreasing interaction with shear rate. Reproduced from Rheology of Liquids, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

liquid. This rheological 'plane' may be very flat (basic Newtonian behavior) or, more commonly, steeply curving (nonideal behavior). The consequences of not having assessed the physical characteristics over a suitable testing regime may prove disastrous in terms of predicting processing behavior.

Behavior as a Function of Time

In the description of the different types of non-Newtonian behavior, it was implied that, although the viscosity of a fluid might vary with shear rate, it was independent of the length of time that the shear rate was applied, and also that replicate determinations at the same shear rate would always produce the same viscosity. This must be considered as the ideal situation, since most non-Newtonian food materials are colloidal in nature, and as such, the flowing

elements, whether particles or macromolecules, may not adapt immediately to the new conditions. Therefore, when such a material is subjected to a particular shear rate, the shear stress and, consequently, the viscosity will decrease with time. Furthermore, once the shear stress has been removed, even if the structure that has been broken down is reversible, it may not return to its original structure (rheological ground state) instantly. The common feature of all these materials is that, if they are subjected to a gradually increasing shear rate followed immediately by a shear rate decreasing to zero, the downcurve will be displaced with regard to the upcurve, and the rheogram will exhibit a hysteresis loop.

In the case of plastic and pseudoplastic materials, the downcurve will be displaced to below the upcurve (Figure 5), whereas for dilatant substances, the reverse will be true (Figure 6).

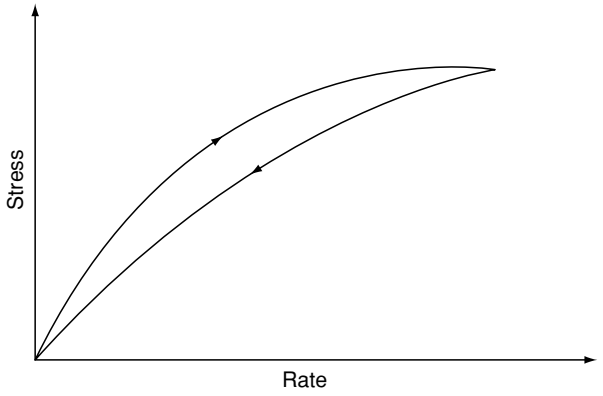

Figure 5 Plastic/pseudoplastic behavior. Reproduced from Rheology of Liquids, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Figure 6 Dilatant behavior. Reproduced from Rheology of Liquids, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

The presence of the hysteresis loop indicates that a 'breakdown' in structure has occurred, and the area with the loop may be used as an index of the degree of 'breakdown.'

The term that is used to describe one such type of behavior is *thixotropy*, which means 'to change by touch.' Strictly, this term should be applied only to an isothermal sol–gel transformation. It has, however, become common to describe as thixotropic any material that exhibits a reversible time-dependent decrease in apparent viscosity. Thixotropic systems are usually composed of asymmetric particles or macromolecules that are capable of interacting by numerous secondary bonds to produce a loose three-dimensional structure, so that the material is gel-like when unsheared. The energy that is imparted during shearing disrupts these bonds so that the flowing elements become aligned, and the viscosity falls, since a gel–sol transformation has occurred. When the shear stress is eventually removed, the structure will tend to reform, although the process is not immediate and will increase with time as the molecules return to the original state under the influence of Brownian motion. Furthermore, the time taken for recovery, which can vary from minutes to days depending upon the system, will be directly related to the length of time that the material was subjected to the shear stress, since this will affect the degree of breakdown.

In some cases, the structure that has been destroyed is never recovered, no matter how long the system is left. Repeat determinations of the flow curve will then only produce the downcurve that was obtained in the experiment that resulted in the material's destruction. Such behavior should be referred to as 'shear destruction' rather than thixotropy, which, as will be appreciated from above, is a misnomer.

An example of such behavior are the gels produced by high-molecular-weight polysaccharides, which are stabilized by large numbers of secondary bonds. Such systems undergo extensive reorganization during shearing such that the three-dimensional structure is reduced: the gel-like nature of the original is then never recovered.

The occurrence of such complex behavior creates problems in quantitative classification because not only will the apparent viscosity change with shear rate, but there will be two 'viscosities' that can be calculated for any given shear rate (i.e., from the upcurve and the downcurve). It is usual to attempt to calculate one viscosity for the upcurve and another for the downcurve. This must assume, of course, that each of the curves achieves linearity over some of its length; otherwise, a defined shear rate must be used: only the former situation is truly satisfactory. Each of

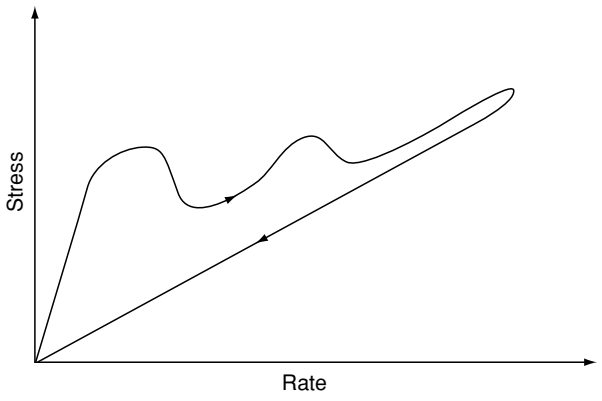

Figure 7 Complex flow behavior. Reproduced from Rheology of Liquids, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

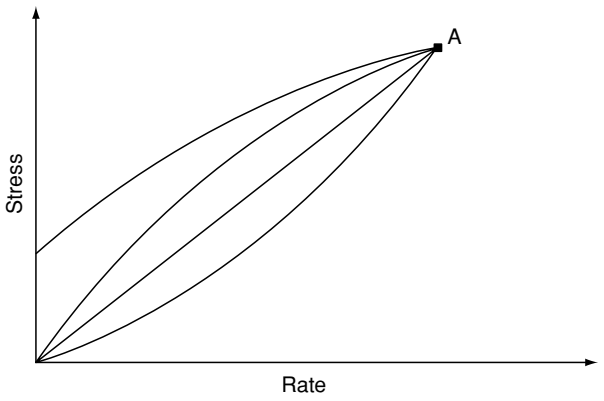

Figure 8 Convergence of flow behavior at a single rate. Reproduced from Rheology of Liquids, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

the lines used to derive the viscosity may be extrapolated to the shear stress axis to give an associated yield value. However, only that derived from the upcurve has any significance, since that derived from the downcurve will relate to the broken-down system. Consequently, the most useful index of thixotropy can be obtained by integration of the area contained within the loop. Of course, this does not take into account the shape of the up- and downcurves, and consequently, two materials may produce loops of similar area but have completely different shapes representing totally different flow behaviors. In order to prevent confusion, it is best to adopt a method whereby an estimate of area is accompanied by yield value(s). This is of particular importance with flow curves that exhibit complex upcurves.

This situation is typical of the type of loop obtained with, for example, some samples of white soft paraffin, where the upcurve exhibits a number of bulges (Figure 7). Those at a lower shear rate are thought to be associated with the initial loss of the three-dimensional structure, while the smoother deviations occurring at the higher shear rates are associated with molecular reorientation. Such behavior is common in food and pharmaceutical systems and is one of the major causes of difficulties in their evaluation.

With such a wide range of rheological behavior, it is extremely important to carry out measurements that will produce meaningful results. It is crucial, therefore, not to use a determination of viscosity at one shear rate (such as would be acceptable for a Newtonian fluid), since it could lead to completely erroneous comparative results.

Figure 8 shows rheograms that are an example of four different types of flow behavior, all of which intersect at point A, which is equivalent to a shear rate of $100\,s^{-1}$. Therefore, if a measurement were made at this one shear rate, all four materials would be shown to have the same viscosity, although they all possess different properties and behavior. Single-point determinations are probably an extreme example but are used to emphasize the importance of properly designed experiments.

It should be noted that while more complex rheological measurements may give rise to parameters other than viscosity (for example, the storage and loss moduli), the arguments given here still apply, and time and shear dependence remain major problems when measuring changing food systems.

See also: **Carbohydrates**: Classification and Properties; **Colloids and Emulsions**; **Rheological Properties of Food Materials**

Further Reading

Ferry JD (1980) *Viscoelastic Properties of Polymers*, 3rd edn. New York: Wiley.

Larson RG (1998) *The Structure and Rheology of Complex Fluids*, 688 pp. Oxford, UK: Oxford Press.

Lewis MJ (1987) *The Physical Properties of Foods and Food Processing Systems*. Chichester, UK: Ellis Horwood.

RIBOFLAVIN

Contents
Properties and Determination
Physiology

Properties and Determination

R Bitsch, Friedrich-Schiller-University, Jena, Germany

Physicochemical Characteristics

Riboflavin, the name prescribed for vitamin B_2 by the International Union of Pure and Applied Chemistry – International Union of Biochemistry, is the biologically active component of the prosthetic group of flavoproteins. Older designations are lactoflavin, ovo-, or uroflavin, which originated from the first isolation. It comprises a tricyclic, nitrogen-containing ring-system, chemically defined as a substituted isoalloxazine with a ribitol side-chain at N_{10}. With the systematical nomenclature, the molecule is named 7,8-dimethyl-10-(1-D-ribityl)-benzo[g]pteridin-2,4-dion. The empiric formula is given as $C_{17}H_{20}N_4O_6$ with a molecular weight of 376.36 (Figure 1). The substance is a yellow to orange–yellow powder and crystallizes in sharp, optically active needles, melting under decomposition at 275–282 °C. The optical rotation is $[\alpha]_D^{25} = -112$ to -122°C at a concentration of 50 mg in 10 ml of 0.02 N alcoholic NaOH). Riboflavin is sparingly soluble in water with a yellow–green fluorescence, very sparingly soluble in ethanol, and insoluble in other organic solvents. It is readily soluble in dilute alkali, but rapidly deteriorated, being accelerated by light. In neutral or acid solutions, it is more stable, even against heat and atmospheric oxygen. Nevertheless, all operations with riboflavin solutions should be handled in the dark or in red light. The light sensitivity of the flavins also has consequences for the stability of vitamin B_2 containing foods. In milk stored in day light, the riboflavin content decreases by more than 80% within hours. Processing losses in foods are given with approximately 20%.

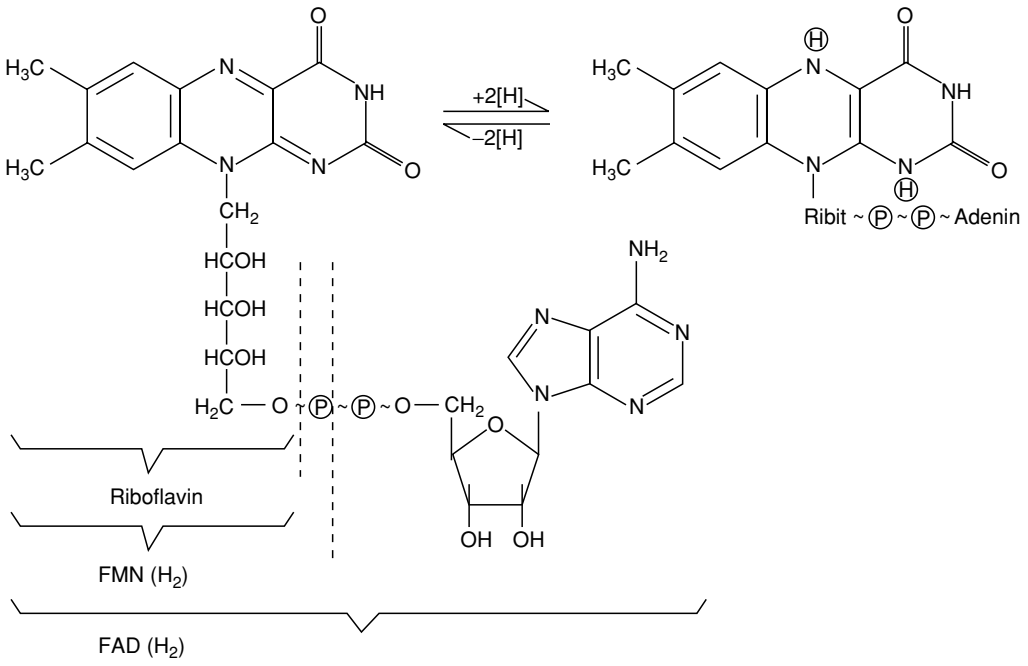

Figure 1 Riboflavin and its coenzyme forms.

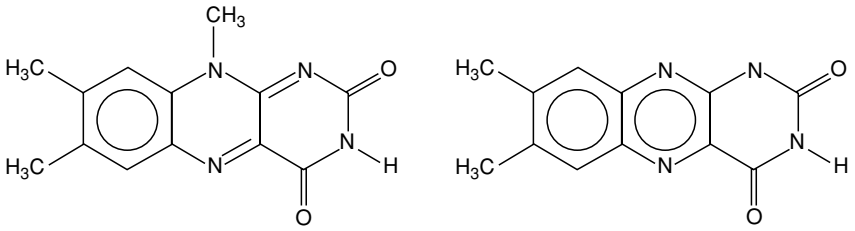

Figure 2 Photolysis products of riboflavin (lumiflavin and lumichrome).

Table 1 Physicochemical data of flavines[a]

Property	*Riboflavin*	*Riboflavin-5-phosphate-Na (FMN-Na)*	*Flavin adenine dinuclectide-Na$_2$ (FAD-Na$_2$)*
Molecular weight	376.36	478.34	829.6
Melting point	278–282 °C (decomp.)	280–290 °C (decomp.)	
Solubility	Readily soluble in dilute alkalis (decomp.)	Soluble in water at pH 7 (30 g l^{-1}, 25 °C)	Soluble in water
	Sparingly soluble in water (70–100 mg l^{-1}, 27 °C)		
	Very sparingly soluble in ethanol (45 mg l^{-1}, 27 °C)	Very sparingly soluble in ethanol	
	Insoluble in acetone, chloroform, ether, benzene		
Absorbance (UV/vis)	223, 267, 374, 444 nm (0.1 N–HCl)		
Fluorescence	Yellow–green fluorescence of aqueous solutions at pH 6–7 (ex. 444 nm, em. 530–565 nm)		ex. 400–500 nm, em. 530 nm
Stability	In neutral and acidic medium relatively stable against heat and oxygen. Rapid photolysis by UV- and visible light to blue fluorescent lumichrome (acidic and neutral solutions) or yellow–green fluorescent lumiflavin (alkaline solutions). After reduction by dithionite, Zn–HCl, sodium hydrosulfite, etc. reversible conversion to leucoflavin (dihydroflavin)		

[a]According to *The Merck Index*, 11th edn. Rahway NJ: Merck & Co. and modified from Eitenmiller RR and Landen WO, Jr. (1999) *Vitamin Analysis for the Health and Food Sciences*, pp. 229–337. Boca Raton, FL: CRC Press.

Light exposure of alkaline solutions decomposes riboflavin to lumiflavin (7.8.10-trimethylisoalloxazine), and in neutral or acidic solutions, the blue fluorescent lumichrome (7.8-dimelthylalloxazine) is formed (Figure 2 and Table 1). Riboflavin is found in free form only in the retina of the eye, whey, and urine. In other biological systems and tissues, the predominant forms are flavin mononucleotide (FMN) or riboflavinphosphate and flavin adenine dinucleotide (FAD). Free riboflavin and FMN-monosodium salt are the most important commercial products. FMN and FAD are effective as coenzymes or prosthetic groups in more than 100 enzymes in the plant and animal kingdoms. The accepted designation for both compounds is incorrect, because FMN is not truly a nucleotide and FAD is not a dinucleotide. The designations arose from the discovery of riboflavin, and have been adopted ever since. Pure riboflavin was first isolated from yeast, egg white, and whey in 1933 (by Kuhn and Wagner-Jauregg). Simultaneously, the so-called 'yellow coenzymes' were discovered (by Warburg and Christian) in yeast and could be identified as FMN. FMN is considerably more soluble in water than riboflavin. Its solubility in organic solvents and its stability against heat and light exposure are comparable with that of riboflavin. The flavoproteins FMN and FAD belong to the enzymatic class of oxidoreductases. The conjugated structure of the pteridine moiety of the basic molecule determines its coloring and light absorbance in the UV and visible spectrum. Absorbance maxima for the flavins are listed in Table 1. Likewise, catalytic hydrogen transfer in the flavin enzymes takes place at both conjugated of bound N_1 and N_5 atoms of the isoalloxazin molecule. The oxidized form of the flavoproteins, the flavoquinone, is thereby reversibly reduced to the colorless flavohydroquinone or leucobase (see Figure 1). The absorbance is lost, because double bonds are dissolved. As an intermediate product during this reaction, a radical semiquinone is formed. During the redox process, the flavin molecule changes its conformation, which is planar in the oxidized state

and folded in a butterfly configuration in the reduced state.

The characteristic yellow–green fluorescence of riboflavin and FMN in aqueous solutions exhibits maximal intensity at pH 6–7. The fluorescence of FAD is about 10–20% of that of riboflavin with the same spectral distribution but exhibiting a maximal intensity at pH 3. The fluorescence of flavins is quenched when bound to proteins. The irradiation products of the flavins also fluoresce, and their spectrum is shifted to the blue (lumichrome) or green (lumiflavin) spectral range. Riboflavin can be quantified by conversion to lumiflavin and by measuring its fluorescence intensity in chloroformic solutions with exitation (ex.) at 270 nm, and emission (em.) at 418 nm.

Methods of Determination

Initially, direct fluorimetry of the intact flavins or after hydrolysis to FMN or free riboflavin have been most frequently used for quantitation of vitamin B_2 active components. Alternatively, the lumiflavin method based upon the conversion of the chloroform-insoluble riboflavin into chloroform-soluble lumiflavin after photodegradation in alkaline solution can be used, despite a transformation rate of only 60–70%. Currently, microbiological assays and HPLC procedures with an enhanced sensitivity are widely used (see also Table 2).

In principle, riboflavin is extracted from dry materials by autoclaving in 0.1 N HCl for 30 min or in a water bath for 1–2 h. By using acid extraction, flavins are released from their protein binding, and FAD and FMN are converted to free riboflavin. In protein-rich materials, acidic extraction may be followed by enzymatic hydrolysis for complete liberation of flavins. It is important to carry out all operations relating to the extraction and quantitation of flavins and riboflavin under dim light and in brown or dark vessels.

Direct Fluorimetry

Direct measurement of the native fluorescence of riboflavin may be applicable for pharmaceutical preparations or food samples with a relatively high vitamin content. After acid extraction, any remaining proteins are removed by precipitation at pH 4.5. Interfering substances are subsequently oxidized with potassium permanganate and excess oxidant destroyed with hydrogen peroxide or sodium bisulfite. The fluorescence of the clear sample solution, commonly spiked with riboflavin as an internal standard, is registered by difference measurement before and after reduction of riboflavin with sodium bisulfite to the leuco form for correction by unspecific fluorescence (ex. = 440 nm; em. = 565 nm).

The detection limit of this method is about 0.1 μg of riboflavin per gram of sample. The direct fluorimetry of flavins had been adopted by the US Pharmacopoeia for the riboflavin analysis in tablets and injections, and by the AOAC for the riboflavin determination in foods, as in milk-based formulas, for example.

Lumiflavin Method

In this method, riboflavin is photodegraded in alkaline solution at pH 10–12 to the stronger fluorescent lumiflavin (ex. = 270 nm; em. = 418 nm). This reaction is specific for riboflavin, since other flavins interfering with riboflavin fluorescence are not converted into lumiflavin. Lumiflavin is separated from riboflavin and other fluorescent substances by extraction with chloroform. The transformation into lumiflavin is not quantitative and ranges between 60 and 70%. Reliable results can be obtained only by comparison

Table 2 Methods of riboflavin determination

Method	*Principles*	*Specials*	*Detection limits*
Direct fluorimetry	Fluorescence measurement of riboflavin at 440 (ex.)/565 (em.) nm	Conversion of FAD and FMN to riboflavin by acidic hydrolysis	0.1 μg B_2 per gram of sample
Lumiflavin method	Fluorescence measurement after irradiation at pH 10–12 (ex. 270, em. 418 nm)	Lumiflavin is separated by chloroform extraction	0.02 ng of B_2
Microbiological assay	Turbidimetric measurement of bacterial growth response dependent on the B_2 content	Interference of growth response with nutrients/food components	0.1–0.5 ng of B_2
HPLC analysis	Separation of flavins on RP-C_{18} materials with aqueous/organic solvents and fluorimetric/UV detection	Differentiated analysis of flavins, combined analysis with other B-vitamins	0.21 ng of riboflavin 0.89 ng of FMN 11.15 ng of FAD 0.02 ng of B_2 (precolumn irradiation to lumiflavin)

with a riboflavin standard under identical irradiation conditions for the sample and standard solution.

Although the original lumiflavin method is rarely used, the principal methodology has been applied to high-performance liquid chromatography (HPLC) analyses of riboflavin content in foods with detection limits of 0.02 ng of riboflavin per injection.

Microbiological Assay

Besides bioassays, as in chicken and curative rat growth tests, for example, microbial assays are the first widely used test methods, measuring the biological activity of preparations containing total vitamin B_2. The most frequently used test organism is *Lactobacillus casei* ssp. *rhamnosus* (ATCC No. 7469). The detection limit is indicated to 0.5 ng of riboflavin absolutely. Occasionally, other test organismus are proposed such as *Enterococcus faecalis* (ATCC No. 10100) or *Leuconostoc mesenteroides* (ATCC No. 9135) with an enhanced sensitivity of 0.1 ng per milliliter of riboflavin. Microbiological assays with *L. casei* have been approved by the Association of Official Analytical Chemists International Methods for riboflavin determination in vitamin preparations. The growth response of bacteria is proportional to the riboflavin content of the medium and can be measured turbidimetrically. Earlier techniques determined metabolically formed lactic acid by titration. However, comparisons of turbidimetrically after 16 h of incubation yielded results that agreed with those obtained by titration of lactic acid after 72 h. Thus, turbidimetric measurement is currently preferred in view of its shorter incubation time and easier handling.

The growth response of *Lactobacillus* species differs significantly between free riboflavin and FMN and FAD. Acidic or enzymatic treatment is therefore needed for releasing and converting the flavins into riboflavin, as described in Methods of Determination, followed by adjustment of the extract to pH 4.5 before incubation. In case of enzymatic release using takadiastase or clarase, it should be noted that some enzyme preparations contain variable traces of riboflavin.

Starch, glycogen, free fatty acids and other lipids, and protein degradation products can interfere with the test by either stimulating or inhibiting bacterial growth. Lipids can be removed either by filtration or by ether or petroleum ether extraction before hydrolysis. Proteins are precipitated at pH 4.5 and starch split by the acidic or enzymatic hydrolysis step. As with the other analytical procedures, the microbiological test should be carried out in dim light. Furthermore, it is important that the glass vessels used do not release any traces of alkali and should thus be cleaned with acid before used, because the flavins are alkali-sensitive. Hence, it is advisable to dissolve riboflavin standards in 0.02 M acetic acid. An optimal extraction procedure applicable for microbiological or HPLC determination of thiamin and riboflavin has been worked out by the European Measurement and Testing Program. The steps of the extraction are detailed in autoclaving food samples in 0.1 N HCI at 121 °C for 30 min and adjusting thereafter to pH 4.0 with 4 M sodium acetate buffer (pH 6.1), followed by the addition of takadiastase (0.1 g per gram of sample), incubation at 37–45 °C (18 h), and then filtration or centrifugation after cooling.

HPLC Procedures

HPLC methods enable separate quantitation of individual flavins or simultaneous analysis of riboflavin, including FMN and FAD. A broad variety of HPLC techniques are described, with variations in the column material, mobile phase, separation mode, and detection system as well. Of the numerous analytical and technical variations communicated during the last 20 years, the separation techniques can be categorized as follows.

In most cases, reversed-phase (RP) C_{18}-materials as stationary phases, and aqueous/organic mobile phases on the basis of water, methanol, acetonitril with or without phosphate or acetate buffer have been used in an isocratic mode. Ion-pair chromatography with sodium salts of hexane or heptane sulfonic acid in the mobile phase (ion-interaction chromatography) was rarely used. In some cases, the separation of riboflavin and flavins was accomplished using the gradient elution technique. Ion-pair HPLC seems more favorable only when riboflavin is to be determined simultaneously with other B-vitamins (e.g., B_1, B_6, niacin, folic acid, B_{12}), because this technique results in better peak shapes.

The retention times of vitamin B_2 active compounds are between 5 (FAD) and 20 (riboflavin) min. Fluorescence monitoring is preferred for detection, and UV or visible absorbance detection at 270 or 446 nm is restricted to early HPLC analyses or measurements of pharmaceuticals and enriched foods. The limits of fluorescence detection at 450/520 nm (ex./em.) for the native flavins were found to be 0.55 pmol (0.21 ng) of riboflavin, 1.96 pmol (0.89 ng) of FMN, and 14.19 pmol (11.15 ng) of FAD. In an earlier HPLC method, the detection limit could be increased to 0.02 ng of riboflavin/injection by irradiation of the sample extract and conversion of the extracted riboflavin to lumiflavin prior to HPLC separation.

In principle, the extraction techniques and sample clean-up suitable for HPLC analysis resemble those

Table 3 HPLC analysis of flavins[a]

Food sample/matrix (compounds determined)	*Extraction procedure*	*Column*	*Mobile phase*	*Detection (nm)*	*Detection limit*
Dairy products (total riboflavin)	Disperse in water, clean up on C_{18} cartridge	Biosil ODS-5S C_{18}	H_2O/CH_3OH/acetic acid = 65/35/1	UV 270	10 ng
Fruits, vegetables (total riboflavin)	0.1 N HCl (30 min) at 100 °C, incubation with mylase (38 °C)	Altex Ultrasphere ODS 5 μm	H_2O/CH_3OH = 60/40 + 5 mM heptanesulfonic acid, pH 4.5	Fluorescence 450/530 (ex./em.)	0.2 ng per injection
Potatoes, vegetables (total riboflavin)	0.1 N HCl (30 min) at 121 °C, incubation with takadiastase (45–50 °C)	μ-Bondapack C_{18}	CH_3OH/H_2O = 30/70	Fluorescence 450/510 (ex./em.)	0.1 ng per injection
Dairy products (FAD, FMN, riboflavin) sorboflavin as internal standard	Homogenize with 6% , containing 2 M urea pass through C_{18} solid phase, elute with 10% / CH_3OH = 4/1	LC-18, 3 μm (Supelco)	14% acetonitril in 01. M KH_2PO_4, pH 2.9	Fluorescence 450/530 (ex./em.)	FMN 2.5 nM FAD 3 nM Riboflavin 2.5 nM
Dairy products, raw and cooked meats, cereals (FAD, FMN, riboflavin) 7-ethyl-8-methyl-riboflavin as internal standard	Homogenize with CH_3OH/CH_2Cl_2 = 9/10 mix with citrate-phosphate buffer, pH 5.5 (containing 0.1% sodium azide), centrifugation	2PLRP-S, 5 μm column temperature 40 °C	Acetonitril/0.1% sodium azide in 0.01 M citrate-phosphate buffer, pH 5.5, gradient elution	Fluorescence 450/522 (ex./em.)	96–113% recovery
Cereals, various foods (thiamin and total riboflavin)	0.1 N HCl, autoclave (15 min) at 125 °C, adjust to pH 4.0–4.5 (2 N NaOAc), incubation with claradiastase at 50 °C (3 h), 50% TCA and heating at 90 °C (15 min), adjust to pH 3.5 (2 N NaOAc), filtration	μ Bondapak C_{18}	CH_3OH/0.005 M phosphate buffer pH 7.0 = 35/65	Fluorescence 440/520 (ex./em.)	80–96% recovery
Dairy + meat products, fruits, vegetables, flour, baked products, beer, coffee (total riboflavin)	0.05 M (0.1 N) H_2SO_4, autoclave (20 min) at 121 °C, adjust to pH 4.5 (2.5 M acetate buffer), incubation with claradiastase at 45 °C (overnight); filtrated sample passed through C_{18} Sep-Pak, elution with 40–70% CH_3 OH	Spherisorb ODS 2.5 μm	H_2O/CH_3OH = 65/35	Fluorescence 445/525 (ex./em.)	20 pg
Blood (FAD, FMN, riboflavin)	10% TCA at 4 °C (30 min), adding NaOAc-buffer, centrifugation	Hypersil ODS 5 μm	0.3 M KH_2PO_4/CH_3OH = 83.3/16.7 (pH 2.9)	Fluorescence 470/525 (ex./em.)	FMN 15 nM FAD 20 nM Riboflavin 10 nM
Serum, urine (riboflavin) isoriboflavin as internal standard	Homogenize/mix with TCA (100 gl^{-1}) centrifugation, pass through Sep-Pak C_{18} (serum)	ROSIL C_{18} HL 5 μm	$CH_3OH/H_2O/CH_3COOH$ = 36.7/63.7/0.1	Fluorescence 450/530 (ex./em.)	10 μgl^{-1}

[a]Modified according to Ball GFM (1998) *Bioavailability and Analysis of Vitamins in Foods*, pp. 294–305. New York: Chapman & Hall; Eitenmiller RR and Landen WO, Jr. (1999) *Vitamin Analysis for the Health and Food Sciences*, pp. 229–337. Boca Raton, FL: CRC Press; De Leenheer AP, Lambert WE and Nelis HJ (eds), (2000), *Modern Chromatographic Analysis of Vitamins*, 3rd edn. New York: Marcel Dekker.

Table 4 Riboflavin content of foods[a]

Content (mg per 100 g) in ingested food	*Milk and milk products*	*Meat and meat products*	*Cereals*	*Vegetables and legumes*	*Fruits and nuts*	*Fish*
> 3.00		Pig's liver				
2.50–3.00		Beef liver, calf liver				
2.00–2.50		Beef kidney, calf kidney				
1.50–2.00		Pig's kidney				
1.00–1.50		Pig's heart				
0.70–1.00			Wheatgerms			
0.50–0.70	Camembert type, cheddar type, Danish blue type, Parmesan		Wheat bran			
0.30–0.50	Cream Brie (50%), Edam type, fresh cheese (skim milk),	Eggs, lamb				Eel, mackerel
0.20–0.30	Fresh cheese (50%), Gouda cheese (45%), yogurt (low-fat)	Goose, pork (lean) veal		Green cabbage, soybean meal	Cashew nuts, hazel nuts, sallow thorn	Flounder, herring, pilchard, plaice
0.15–0.20	Cottage cheese, cows' milk (fresh)	Beef (lean), chicken, duck, turkey	Oat flakes, wholemeal	Broccoli (boiled), mangold, soybean sprouts, spinach (boiled)	Avocado	Haddock, salmon
0.10–0.15				Asparagus (boiled)	Peanuts (roasted)	

[a]According to Souci SW, Fachmann W and Kraut H (2000) *Food Composition and Nutrition Tables*. Stuttgart: Medpham Scientific.

used for quantification by the other methods discussed. Most often, vitamin B_2 active substances are released from the food matrix by autoclaving with 0.1 N mineral acid (HCI or H_2SO_4) followed by enzymatic digestion with papain, takadiastase, or claradiastase. Solid-phase clean-up prodecures on C_{18}-materials or Florisil are often used for support prior to injection. The extraction procedure depends on the type of analysis, whether the total riboflavin content, i.e. the sum of FAD, FMN, and free riboflavin is to be determined, or whether the separated quantitation of FAD and FMN in addition to free riboflavin is required. Treatment with potassium permanganate can be omitted, because possibly remaining extraneous fluorescent substances are chromatographically separated.

Combined acid and enzymatic hydrolysis is advantageous, particularly for foods with a high starch or protein content, to liberate the bound flavins. Occasionally, autoclaving with dilute mineral acid is incomplete, particularly for the conversion of FMN to riboflavin, as indicated by the appearance of an FMN peak in the following chromatogram. In addition, FMN may be partially converted during acid hydrolysis to biologically active isomeric riboflavin phosphates, which can be separated from FMN by HPLC analysis, and are ignored when calculating total riboflavin. The combination of acidic and enzymatic digestion is thus advisable prior to HPLC analysis for the determination of total riboflavin. When using this extraction procedure and following HPLC analysis, the riboflavin content of various foods such as breakfast cereals, porridges, milk, and milk products correlates well with the microbiological assay if *Lactobacillus casei* is used. The simultaneous determination of FAD and FMN besides free riboflavin requires non-degradative techniques, such as extraction of flavins by methanol/dichloromethane followed by partitioning with citrate buffer (pH 5.5) or extraction with 6% formic acid containing 2 M urea. The HPLC separation and quantitation of the flavins were carried out using internal standards (7-ethyl-8-methyl-riboflavin, sorboflavins, isoriboflavin, nicotinamide, and others). This analytical device can be used in the flavin analysis (FAD, FMN, and riboflavin) of milk and dairy products, fruits and vegetables, meats, and cereal products. In many cases, HPLC methods have been developed for the simultaneous or sequential determination of riboflavin and thiamin in foods. Both vitamins are

extracted by a common procedure using acidic and enzymatic digestion.

For biological samples, such as blood or plasma/serum, the preferred extraction medium is 5–10% trichloroacetic acid, which is suitable for denaturating the relatively weak protein binding, while keeping the phosphorylated forms intact. In this way, flavins in whole blood can be analyzed by isocratic RP-HPLC against an external standard. A list of usual HPLC procedures is given in Table 3, and an overview of the methods of riboflavin determination is given in Table 2.

Food Content

Vitamin B_2 is widely distributed in animal and vegetable foods. Protein-rich foods of animal origin are, as a rule, considerable sources of this vitamin with good bioavailability. Particularly rich in riboflavin are offal, such as heart, liver, and kidney. Because of the relative heat stability, only minor vitamin losses occur during preparation. Likewise, pasteurization of milk causes losses of less than 10%. However, exposure to day light may result in a remarkable decrease in the riboflavin content in foods, depending on the surface area exposed. Vitamin losses occur during the milling of cereals. White flour with a low extraction rate contains about one-third of the vitamin content of whole grain flour. Of special importance for the dietary habits in Western countries are milk and dairy products despite only a medium riboflavin content. If these are omitted from the diet, it may be difficult to achieve an adequate vitamin intake (Table 4).

See also: **Analysis of Food**; **Chromatography**: High-performance Liquid Chromatography; **Enzymes**: Functions and Characteristics; **Milk**: Dietary Importance; **Spectroscopy**: Fluorescence

Further Reading

Ball GFM (1998) *Bioavailability and Analysis of Vitamins in Foods*. New York: Chapman & Hall. pp. 294–305.

De Leenheer AP, Lambert WE and van Bocxlaer JF (eds) (2000) *Modern Chromatographic Analysis of Vitamins*, 3rd edn. New York: Marcel Dekker.

Eitenmiller RR and Landen WO Jr. (1999) *Vitamin Analysis for the Health and Food Sciences*, pp. 229–337. Boca Raton, FL: CRC Press.

Friedrich W (1988) Vitamin B_2: Riboflavin and its bioactive variants. In: *Vitamins*. New York: Walter de Gruyter.

Official Methods of Analysis of AOAC International (1999) 16th edn., vol. II. Gaitherburg, MD: AOAC International.

Souci SW, Fachmann W and Kraut H (2000) *Food Composition and Nutrition Tables*. Stuttgart: Medpharm Scientific.

Strohecker R and Henning HM (1965) *Vitamin Assay Tested Methods*. Darmstadt: Verlag Chemie.

The Merck Index (1989) 11th edn. Rahway NJ: Merck & Co.

Physiology

D B McCormick, Emory University School of Medicine, Atlanta, GA, USA

Introduction

Following the earlier identification of riboflavin and the two most common flavocoenzymes, the more recent recognition of the greater diversity of natural flavins has led to a broader appreciation of the multiple functions and metabolic processing of these important compounds. Much of the progress in this area and detail of function of the flavins and flavoproteins have been given in the periodic symposia held on this subject. This article reviews the principal features of the physiological handling of riboflavin and its natural derivatives in the mammalian body and, where known, in the human.

Digestion and Bioavailability

Riboflavin and lesser amounts of natural derivatives are released by digestion of complexes, mostly flavoproteins, contained within foods. Coenzyme forms of the vitamin, mainly flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN), are released from noncovalent attachment to proteins as a consequence of gastric acidification. Nonspecific hydrolyses of the coenzyme forms by pyrophosphatase and phosphatase occur in the upper small intestine. By such actions, FAD is converted to FMN, which is further converted to riboflavin. Several per cent of 8α-(amino acid)riboflavins originally in covalent attachment as 8α-FAD linked to certain enzymes, notably of mitochondrial origin, are also released by such hydrolases that function together with proteolysis of the attached protein chains, which begins in the stomach with pepsin and continues in the small intestine with trypsin, chymotrypsin and exopeptidases. Traces of other ring and side-chain substituted flavins are similarly released by combinations of the above actions on non-covalently and covalently bound flavins. Riboflavin 5′-glycosides, for example, are cleaved by glycosidases present in the succus entericus. The digestive processes and locale for release of

flavins from ingested material is shown in Figure 1. (*See* **Coenzymes**.)

While the efficiency of release of riboflavin from noncovalently bound forms is essentially complete in the normal gastrointestinal tract, the vitamin is not recovered intact from flavin covalently linked to protein. Since the latter is less than 10% of total flavin within the diverse foods ingested by people and most mammals, the average bioavailability of riboflavin is fairly high. (*See* **Bioavailability of Nutrients**.)

Absorption and Transport

Riboflavin and a fraction of flavin metabolites,including ring-altered forms, e.g., 8α-(amino acid)riboflavins, and side-chain derivatives, e.g., 7,8-dimethyl-10-formyl-methylisoalloxazine, are absorbed primarily in the proximal small intestine by a saturable transport system that is rapid and approximately proportional to dose before levelling off. This saturation level is achieved with about 25 mg of the vitamin given in a single bolus to adult humans. Bile salts appear to facilitate the uptake, and a modest amount of flavin circulates via the enterohepatic system. The initial uptake of riboflavin by enterocytes is Na^+-dependent and reflects an adenosine 5′-triphosphatase (ATPase)-involved active cotransport system. Metabolic trapping by conversion to FMN and FAD occurs before release of the vitamin to circulation by nonspecific pyrophosphatase and phosphatase. (*See* **Bile**.)

Circulatory transport of flavin involves loose association with albumin and tight associations with some globulins. A subfraction of immunoglobulin G (IgG) has been found to bind avidly a small portion of the total free flavin in blood, and several immunoglobulins contribute significantly to plasma transport of the vitamin. Some riboflavin-binding proteins in plasma are pregnancy-specific, including the classic case of the estrogen-induced egg-white protein. These proteins have at least some portion of the binding domain in common and are essential for fetal development. Placental transfer of riboflavin in the human and other mammals involves binding proteins that help vector the vitamin and enhance supply to the fetus.

Uptake processes for flavins by mammalian cells have some characteristics in common, but there are both qualitative and quantitative differences among different cell types. Entry of riboflavin appears to be carrier-mediated (facilitated) at physiological concentrations of the vitamin, since there is relative specificity to a saturable component that is responsible for initial rapid uptake. A riboflavin-binding protein has even been isolated from the plasma membrane of rat liver cells. The nonepithelial hepatocyte does not depend on Na^+ for riboflavin import, as do bipolar epithelial types such as the enterocyte or renal proximal tubular cell. Slower passive diffusion becomes more evident when the facilitating transporter is exceeded by pharmacological levels of the vitamin. In all cases, metabolic trapping of riboflavin by

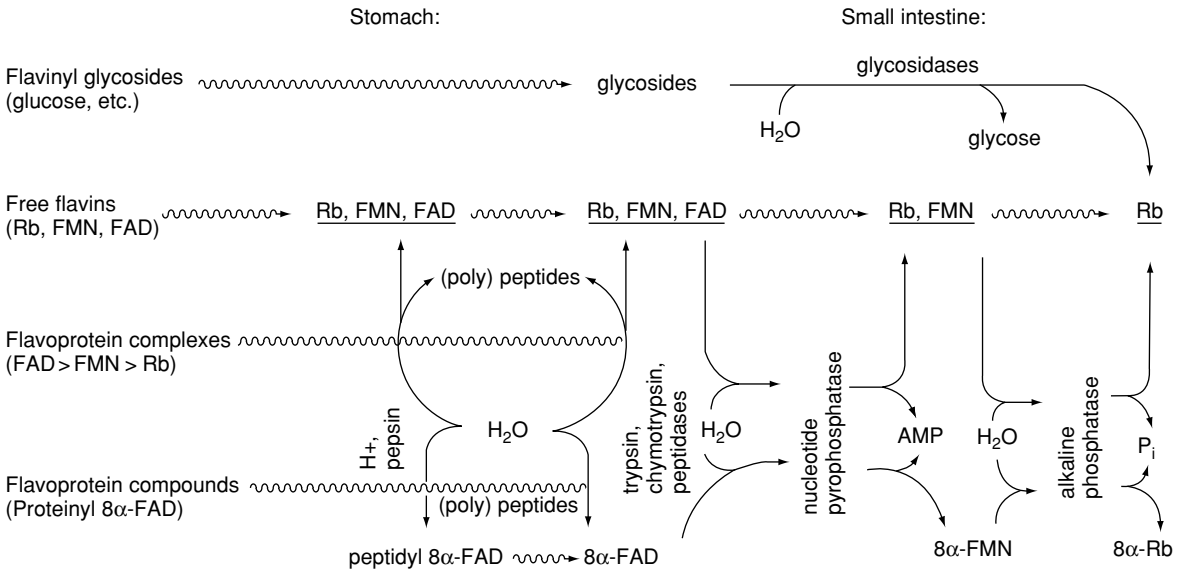

Figure 1 Digestion of flavins in monogastric mammals. Rb, riboflavin; P_i, inorganic phosphate. Reproduced with modifications from Riboflavin: Physiology, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

phosphorylation dependent upon cytosolic flavokinase follows passage of the vitamin through the plasma membrane. Release of riboflavin from cells requires hydrolysis of FMN by nonspecific phosphatases.

Cellular Interconversions

The metabolic interconversions of riboflavin and flavocoenzymes are summarized in Figure 2.

Conversion of riboflavin to coenzymes occurs within the cellular cytoplasm of most tissues, but particularly in the small intestine, liver, heart, kidney, and brain. The obligatory first step is the adenosine 5′-triphosphate (ATP)-dependent phosphorylation of the vitamin catalyzed by flavokinase, which utilizes Zn^{2+}. The FMN product can be complexed with specific apoenzymes to form several functional flavoproteins, but the major portion is further converted to FAD in a second ATP-dependent reaction catalyzed by FAD synthetase, which utilizes Mg^{2+}. It is clear that the biosynthesis of flavocoenzymes is regulated by supply of riboflavin (flavin status), competition for ATP (energy status), and hormonal balances. Thyroxine and triiodothyronine stimulate FMN and FAD synthesis in mammalian systems. This seems to involve a hormone-mediated increase in an active form of flavokinase. As a product of the synthetase, FAD is also an effective inhibitor at this second step and may help regulate its own formation. FAD is the predominant flavocoenzyme present in tissues where it is mainly complexed with numerous flavoprotein dehydrogenases and oxidases. Several percent of the FAD also becomes covalently attached to specific amino acid residues of a few important apoenzymes. Examples for the human include the 8α-N^3-histidyl FAD within the mitochondrial dehydrogenases for succinate, dimethylglycine and sarcosine, and the 8α-*S*-cysteinyl FAD within monoamine oxidase, also of mitochondrial localization. (*See* **Hormones**: Thyroid Hormones.)

Turnover of covalently attached flavocoenzymes requires intracellular proteolysis, and further degradation of the coenzymes involves nonspecific pyrophosphatase and specific 5′-nucleotidase cleavage of FAD to AMP (adenosine monophosphate) and FMN and action by nonspecific phosphatases on the latter.

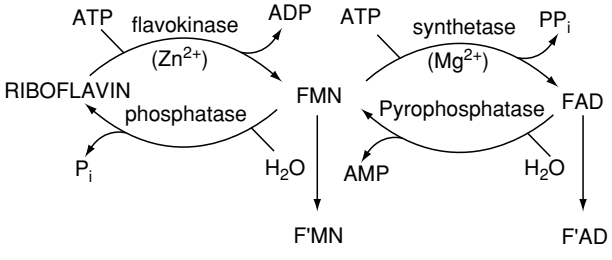

Figure 2 Metabolic interconversions of riboflavin and flavocoenzymes. Reproduced from Riboflavin: Physiology, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Storage and Catabolism

There is little storage of riboflavin as such, since most exists within flavocoenzymes that are in relatively tight associations in holoenzymatic systems. During such severe deficiency of the vitamin as leads to death of experimental animals, there is a reduction in the level of extractable flavin that can approach about half of that found in an optimally supplemented control. Hence, there is moderately effective retention of riboflavin by its metabolic commitment to bound forms; however, even a modest deficit of the vitamin is reflected in the decrease in function of certain flavoproteins well before full-blown symptoms of deficiency.

Though certain bacteria, especially of the *Pseudomonas* genus, can extensively degrade both the ring system and side-chains of flavins, mammals are more limited in their abilities to catabolize the vitamin. The considerable diversity of flavin metabolites in and from mammals reflects the composite of reactions of photochemical processes on the skin, microfloral activities in the gastrointestinal tract, as well as somatic actions both directly on flavin and on derivatives presented to cells by circulatory recovery from dermal tissue and by enterohepatic retrieval from the gut. The diverse flavin-related products identified from humans and other mammals are summarized in Figure 3.

Cleavage of the D-ribityl side-chain at position 10 is mainly, if not entirely, attributable to light and intestinal microflora. Both can lead to partial fragmentation to form the 10-formylmethylflavin. This can be oxidized by alimentary bacteria of the ruminant and human to form the 10-carboxymethylflavin, and a further fraction of the formylmethyl compound is interconverted with the 10-hydroxymethylflavin as a result of pyridine-nucleotide-dependent dehydrogenase in tissue. Lumichrome-level compounds not only can result from complete removal of the side-chain by microflora, which can be decreased by antibiotic administration, but can accompany lumiflavin as a photoproduct from action of light on flavin within the dermal tissue. Catabolites of riboflavin that primarily derive from oxidations within tissues

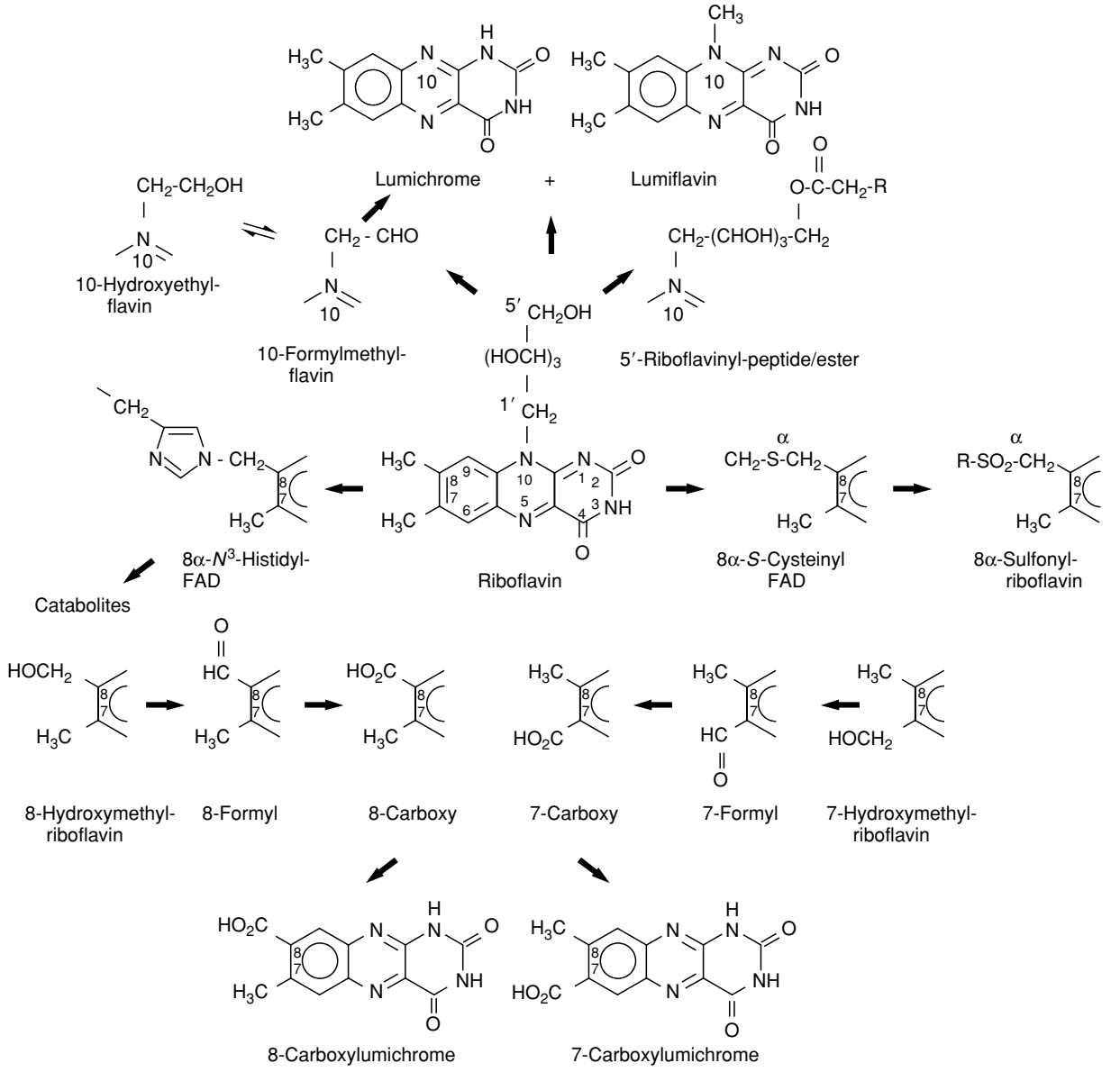

Figure 3 Photochemical, microfloral, and cellular catabolism of riboflavin within mammals. Reproduced from Riboflavin: Physiology, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

are the 7- and 8-hydroxymethylriboflavins (7α- and 8α-hydroxyriboflavins). These, and products from further oxidation of the hydroxymethyl functions to formyl and carboxyl groups, reflect microsomal mixed-function oxidase activity. Other flavin catabolites include those from 8α-(amino acid)riboflavins released from covalently bonded FAD. An 8α-sulfonylriboflavin may derive from the 8α-cysteinyl-FAD of monoamine oxidase. A peptide ester and a glucoside, both linked to the 5′-hydroxymethyl terminus of the vitamin, have also been found.

Excretion and Secretion

Since no isoalloxazine (flavin) can be biosynthesized within the cells of mammals that lack riboflavin synthetase, excretion and secretion reflect dietary intake and catabolic and photodegradative events. Essentially all known catabolites of riboflavin have been detected in urine; many of the lumichrome-level compounds are also in feces. For normal adults eating varied diets, riboflavin comprises 60–90% of urinary flavin, 7-hydroxymethylriboflavin 3–7%,

8α-sulfonylriboflavin 2–15%, 8-hydroxymethylriboflavin 1–8%, 10-hydroxyethylflavin 1–7%, riboflavinyl peptide ester up to 5%, with traces of lumiflavin and, sometimes, the 10-formylmethyl- and carboxymethylflavins.

The presence in milk of 'lactoflavin,' an early name for riboflavin, led to the recognition of this food as a good source of the vitamin. For milk from both cows and humans, the flavin in highest concentration other than the free vitamin is FAD, which can comprise over a third of total flavin. Much of this is hydrolyzed to FMN by pasteurization. Fairly significant quantities of the 10-(2′-hydroxyethyl)flavin are notable, since this catabolite has antivitamin activities, as reflected in competitive inhibition of both cellular uptake and subsequent flavokinase-catalyzed phosphorylation of riboflavin. Hence, this catabolite, which may reach 10–12% of flavin in cow's milk, modestly subtracts from the biological activity of this food. Several percent of both 7- and 8-hydroxymethylriboflavins are also present, with more of the former. Smaller amounts of other catabolites, including the 10-formylmethylflavin and lumichrome, comprise most of the rest. (*See* **Milk**: Dietary Importance.)

Biochemical Functions

In bound coenzymic forms, riboflavin participates in oxidation–reduction reactions in numerous metabolic pathways and in energy production via the respiratory chain. A variety of chemical reactions are catalyzed by flavoproteins. The redox functions of a flavocoenzyme (Figure 4) include one-electron transfers, during which the neutral, oxidized quinone level of flavin is half reduced to the radical semiquinone, which can exist within natural pH ranges as neutral or anionic species. A further electron transfer can lead to a fully reduced hydroquinone. In addition, a single-step, two-electron transfer from substrate to flavin can occur with hydride ion transfer, e.g., from reduced pyridine nucleotide, or by base abstraction of a substrate proton together with carbanion addition.

There are flavoprotein-catalyzed dehydrogenations that are both pyridine-nucleotide dependent and independent, reactions with sulfur-containing compounds, hydroxylations, oxidative decarboxylations, dioxygenations, and reduction of oxygen to hydrogen peroxide. The intrinsic abilities of flavins – to be varyingly potentiated as redox carriers upon differential binding to proteins, to participate in both one- and two-electron transfers, and to react in reduced (1,5-dihydro) form with oxygen – permit wide scope in their operation.

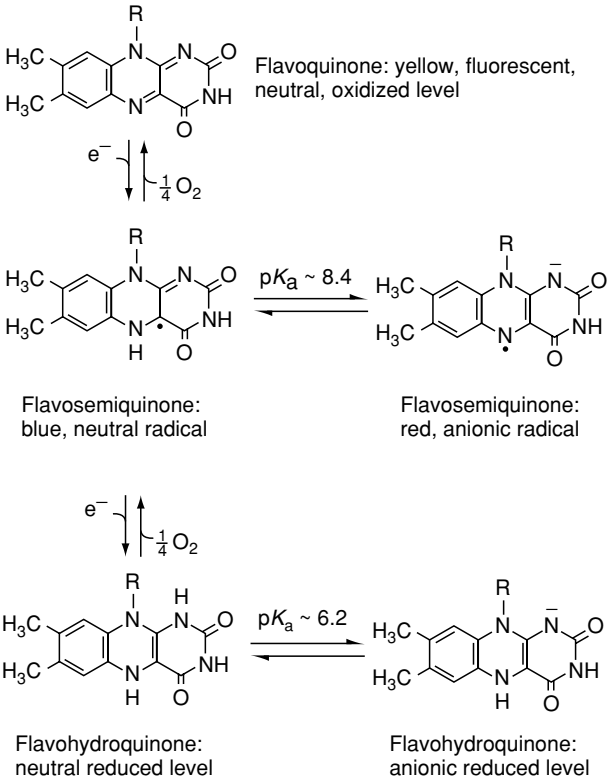

Figure 4 Oxidation–reduction states of flavocoenzymes functioning physiologically. Reproduced from Riboflavin: Physiology, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Requirements and Intakes

The requirement levels for riboflavin, in contrast to those for thiamin, are not raised when energy utilization is increased. Because of the interdependence of protein, energy intake, and metabolic body size, however, allowances calculated on these three bases do not differ significantly. Clinical signs of deficiency in adults can be prevented with intakes of riboflavin above 0.4 mg per 1000 kcal, but over 0.5 mg per 1000 kcal may be required to maintain tissue reserves in adults and children as reflected in urinary excretion, erythrocyte riboflavin, and erythrocyte glutathione reductase. From these considerations, the riboflavin allowances are now computed as 0.6 mg per 1000 kcal for people of all ages. This leads to US Recommended Dietary Allowances (RDAs) ranging from 0.4 mg per day for early infants to 1.7 mg per day for young adult males. However, for elderly people and others whose daily calorie intake may be less than 2000 kcal, a minimum of 1.2 mg per day is recommended in the USA. Since pregnancy imposes extra demands, reflected by decreased excretion

and an elevated FAD stimulation of erythrocyte glutathione activity, an additional 0.3 mg per day is recommended. The lactating woman secretes approximately 35 μg per 100 ml of milk for an output of about 0.26 mg per day (750 ml) during the first 6 months and 0.21 mg per day (600 ml) during the second 6 months. Since the utilization of the additional riboflavin for milk production is assumed to be 70%, an additional intake of 0.5 mg is recommended for the first 6 months and 0.4 mg for the second. (*See* **Lactation**: Physiology.)

Small amounts of riboflavin, largely as digestible coenzymes, are present in most plant and animal tissue. Good sources are eggs, lean meats, milk, broccoli, and enriched breads and cereals. Such losses as occur during cooking are largely attributable to leaching of the heat-stable but light-sensitive flavins into water.

When supplementation or therapy with riboflavin is warranted, oral administration of five to 10 times the RDA is usually satisfactory.

Deficiency Causes and Symptoms

Pure, uncomplicated riboflavin deficiency is probably never encountered in patients, but is accompanied by multiple nutrient deficiencies. Ariboflavinosis can result from such primary and secondary factors as commonly affect supply or utilization of other nutrients as well. Inadequate dietary intake most commonly related to limited availability of food, but sometimes exacerbated by poor storage or processing, remains the major cause. In addition, anorexic persons rarely ingest adequate amounts of riboflavin and other nutrients.

Decreased assimilation results from abnormal digestion, absorption, or both. Lactose intolerance as a result of lactase insufficiency, mostly encountered among Blacks and Asians, argues against such people consuming nonlactase-treated milk, which is a good source of the vitamin. Malabsorption can occur as a result of tropical spure, celiac disease, malignancy and resection of the small bowel, and gastrointestinal and biliary obstruction. Poor absorption also results from disorders that increase motility and decrease gastrointestinal passage time, such as diarrhea, infectious enteritis, and irritable bowel syndrome. (*See* **Food Intolerance**: Lactose Intolerance.)

Rather rarely encountered, but usually significantly improved by therapeutic treatment with riboflavin, are certain inborn errors where the genetic defect is in formation of a normal flavoprotein. Cases in this category include fatty acid desaturases in which specific defects have been found for the mitochondrial FAD-dependent dehydrogenases for short-chain, long-chain, and multi-chain acyl-CoAs (acyl coenzyme As). The young patients have a lipid storage myopathy, often accompanied by carnitine insufficiency, and exhibit glutaric aciduria. A low, FMN-dependent pyridoxine 5′-phosphate oxidase activity due to an erythrocyte deficiency of FMN, confirmed by response to oral riboflavin, was reported in the majority of subjects with D-glucose 6-phosphate dehydrogenase deficiency. Such cases seem to have an accelerated conversion of FMN to FAD so that glutathione reductase is saturated. This contrasts with heterozygous β-thalassemia, in which there is an inherited slow erythrocyte conversion of riboflavin to FMN, a decrease in subsequent FAD, and a high stimulation of the erythrocyte glutathione reductase by extraneous FAD.

Defective utilization can result from disturbances in hormonal production, certainly relating to thyroid hormone, but less likely as a result of taking oral contraceptives. Phenothiazine derivatives appear to impair use of riboflavin.

Increased destruction of riboflavin occurs during treatment of neonatal jaundice with phototherapy. In this case, the side-chain of the vitamin is photochemically destroyed, as it is involved in the photosensitized oxidation of bilirubin to more polar, excretable compounds.

The finding that phenobarbital induces microsomal oxidation of the 7-methyl function of the vitamin lends credence to the belief that long-time use of barbiturates may jeopardize flavin status.

Enhanced excretion of riboflavin occurs in catabolic patients undergoing nitrogen loss. The relationship of the vitamin to protein status has long been recognized. Also, certain antibiotics and phenothiazine drugs increase excretion of riboflavin. (*See* **Drug–Nutrient Interactions**.)

Increased requirements can, of course, be the consequence of one or more of the above-mentioned factors. For example, protein–calorie malnutrition commonly accompanies a diminution in both absorption and utilization of riboflavin. Systemic infections, even without gastrointestinal involvement, sometimes lead to increased requirements that can result from decreased intake, defective absorption, poor utilization, and increased excretion. (*See* **Protein**: Deficiency.)

Clinical deficiency of riboflavin has been reduced by feeding a riboflavin-deficient diet and/or by the administration of an antagonist such as galactoflavin. The deficiency syndrome is characterized by sore throat, hyperemia and edema of the pharyngeal and oral mucous membranes, cheilosis, angular stomatitis, glossitis (magenta tongue), seborrheic dermatitis, and normochromic, normocytic anemia associated

with pure red cell cytoplasia of the bone marrow. As noted above, some of these symptoms, e.g., glossitis and dermatitis, when encountered in the field, may have resulted from other complicating deficiencies. Severe riboflavin deficiency can also affect the conversion of vitamin B_6 to its coenzyme and even curtail conversion of tryptophan to niacin. (*See* **Niacin**: Physiology; **Vitamin B_6**: Properties and Determination.)

Toxicity

Toxicity from ingestion of excess riboflavin by experimental animals or humans is doubtful. The capacity of the human gastrointestinal tract to absorb orally administered riboflavin may be less than 30 mg in a single dose. The limited solubility and absorptivity of this vitamin as encountered in multivitamin preparations and natural foodstuffs, and its ready excretion as typical of water-soluble vitamins, normally precludes a health risk. There is one report of EEG (electroencephalogram) abnormalities in two patients during long-term treatment with riboflavin and niacin.

See also: **Bile**; **Bioavailability of Nutrients**; **Coenzymes**; **Drug–Nutrient Interactions**; **Food Intolerance**: Lactose Intolerance; **Hormones**: Thyroid Hormones; **Lactation**: Human Milk: Composition and Nutritional Value; Physiology; **Milk**: Dietary Importance; **Niacin**: Properties and Determination; **Protein**: Deficiency; **Vitamins**: Overview; Determination; **Vitamin B_6**: Properties and Determination

Further Reading

Edmondson DE and McCormick DB (eds) (1987) *Flavins and Flavoproteins*. Berlin: Walter de Gruyter.

McCormick DB (1999) Riboflavin. In: Shils ME Olson JA, Shike M and Ross AC (eds) *Modern Nutrition in Health and Disease*, pp. 391–399. Baltimore, MD: Williams and Wilkins.

McCormick DB (1989) Two interconnected B vitamins: riboflavin and pyridoxine. *Physiological Reviews* 69: 1170–1198.

McCormick DB (1990) Riboflavin. In: Brown ML (ed.), *Present Knowledge in Nutrition*, pp. 146–154. Washington, DC: International Life Sciences Institute–Nutrition Foundation.

McCormick DB (1997) Coenzymes, Biochemistry. In: Dulbecco R (ed.) *Encyclopedia of Human Biology*, vol. 2, pp. 847–864. San Diego: Academic Press.

Merrill AH, Jr, Lambeth JD, Edmondson DE and McCormick DB (1981) Formation and mode of action of flavoproteins. *Annual Review of Nutrition* 1: 281–317.

Müller F (ed.) (1991) *Chemistry and Biochemistry of Flavoenzymes*, vols. I–III. Boca Raton, FL: CRC Press.

RICE

B O Juliano, Philippine Rice Research Institute Los Baños, Laguna, Philippines

Global Distribution, Varieties, and Commercial Importance

Rice is the most important crop in the world in terms of total developing world production (480×10^6 tonnes of rough rice) and the number of consumers (2.5 billion) dependent on it as their staple food. It is widely grown in over 100 countries in every continent (except Antartica), from 53 °N to 40 °S, and from sea level to an altitude of 3 km. The total production in 1998 was 568×10^6 tonnes of rough rice. Asia accounts for 134 of the 149×10^6 ha world area that is used for rice cultivation. The mean rough rice yield was 3.74 tonnes ha^{-1} in 1998. About 90% of the world's rice is grown and consumed in Asia. The major rice producers in 1998 were China, India, Indonesia, Bangladesh, Vietnam, and Thailand. In terms of water regime, 54% of the total rice area was irrigated in 1991, 25% rainfed lowland, 13% upland, and 9% as flood-prone wetland, corresponding to total rough rice production of 76% from irrigated, 16% from rainfed lowland, 4% from upland, and 4% from flood-prone wetland. Since the rice-growing area is shrinking, rice production must keep up with the 1.7% per year increase in population in tropical Asia through increased yield.

There are estimated to be about 100 000 rice varieties; only a small proportion is actually widely cultivated. They vary in grain weight, size and shape, degree of dormancy, longevity and seedling vigor, and some have red to purple–black pigments. About 75% of Asia's rice area is planted with varieties of improved semidwarf plant type with erect leaves. The newer, improved varieties have a similar yield

potential to the first variety (IR 8) but have a better resistance or tolerance to biotic and abiotic stresses. 'Super' rice being bred has fewer tillers but with all of them productive (bearing panicle).

Only about 4–5% of the world's rice production enters the international trade. The major exporters in 1997 were Thailand, Vietnam, India, the USA, and Pakistan. The major importers in 1997 were Iran, Brazil, Nigeria, the Philippines, Iraq, Saudi Arabia, Malaysia, South Africa, and Côte d' Ivoire.

Morphology and Anatomy of the Spikelet or Grain

The rice grain (rough rice or paddy) consists of an outer protective covering, the hull (husk), and the edible rice caryopsis or fruit (brown, cargo, dehulled, or dehusked rice) (Figure 1). Brown rice consists of the outer layers of pericarp, seed coat, nucellus, and germ or embryo, which are the maternal tissues, and the endosperm. The endosperm consists of the aleurone layer and the starchy or inner endosperm. Pigment is confined to the pericarp, but there is a varietal difference in the extent of retention of pigment with the degree of milling. The aleurone layer encloses the embryo.

The inedible hull constitutes 16–28% (mean 20%) of rough rice weight. Brown rice consists of 1–2% pericarp, 4–6% aleurone plus nucellus and seed coat, 1% embryo, 2% scutellum and 90–91% endosperm. The aleurone and embryo cells are rich in lipid bodies (spherosomes, 0.2–1.5 μm) and in protein bodies (aleurone grains) containing inclusions of phytic acd bodies or globoids (1–3 μm).

The endosperm cells are thin-walled and packed with amyloplasts containing polyhedral compound starch granules about 3–9 μm in size. Protein occurs mainly in the form of large (1–2 μm) and small (0.5–0.8 μm) spherical protein bodies and crystalline protein bodies (2–4 μm). Spherical protein bodies (PB I) are rich in prolamin (alcohol-soluble protein), and crystalline protein bodies (PB II) are rich in glutelin (alkali-soluble protein). Spherosomes are present in the subaleurone or the two outermost cell layers of the endosperm.

Chemical and Nutritional Composition

Rice has one of the lowest protein contents (7%) among the cereals. The bran layers and embryo are richer in nonstarch constituents than the milled rice (Table 1). The major nutritional advantage of brown rice over milled rice is its higher content of B vitamins. Although higher in minerals, bran phytic acid and probably dietary fiber in the aleurone form complexes with minerals and proteins, reducing their bioavailability. The energy content of brown rice and bran is higher than that in milled rice, owing to the higher fat content. Rice has no vitamin A, C, or D.

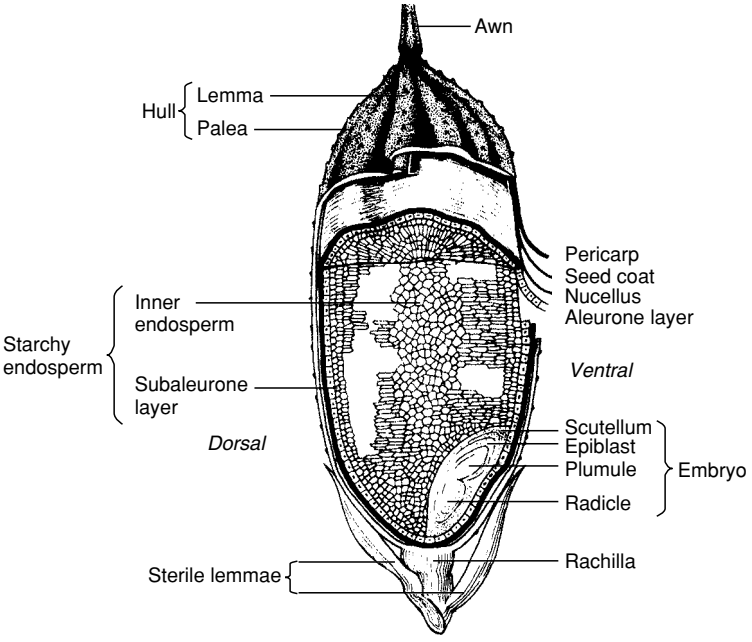

Figure 1 Longitudinal section of the rice grain. From Juliano BO (1984) In: Whistler RL, BeMiller JN and Paschall EF (eds) *Starch Chemistry and Technology*, 2nd edn, p. 509. Orlando, FL: Academic Press, with permission.

Table 1 Comparison of nutrient composition of brown rice, milled rice and rice bran

Property	*Amounts per 100 g*		
	Brown rice	*Milled rice*	*Rice bran*
Moisture (g)	14.0	14.0	14.0
Energy content (KJ)	1520–1610	1460–1560	1670–1990
Energy content (kcal)	363–385	349–373	399–476
Crude protein (g)	7.1–8.3	6.3–7.1	11.3–14.9
Crude fat (g)	1.6–2.8	0.3–0.5	15.0–19.7
Crude fiber (g)	0.6–1.0	0.2–0.5	7.0–11.4
Crude ash (g)	1.0–1.5	0.3–0.8	6.6–9.9
Available carbohydrates (g)	73–87	77–89	34–62
Total dietary fiber (g)	2.9–4.0	0.7–2.3	17–29
Water-insoluble fiber (g)	2.0	0.5	15–27
Sugars (g)	1.9	0.2–0.5	6.4
Thiamin (mg)	0.3–0.6	0.02–0.11	1.2–2.5
Riboflavin (mg)	0.04–0.14	0.02–0.06	0.18–0.43
Niacin (mg)	3.5–5.3	1.3–2.4	26.7–49.9
Pantothenic acid (mg)	1.4	1.0	6.8
Vitamin B_6 (mg)	0.5	0.2	3.7
Folate (μg)	19	8	58
Vitamin E, α-tocopherol (mg)	0.8–2.5	< 0.01–0.30	3–15
Calcium (mg)	10–50	10–30	30–120
Phosphorus (g)	0.17–0.43	0.08–0.15	1.1–2.5
Phytic acid P (g)	0.13–0.27	0.02–0.07	0.9–2.2
Iron (mg)	0.2–5.2	0.2–2.8	8.6–43.0
Zinc (mg)	0.6–2.8	0.6–2.3	4.3–25.8

From Juliano BO (ed.) (1985) *Rice: Chemistry and Technology*. St. Paul, MN: American Association of Cereal Chemists; United States Department of Agriculture (USDA) (1998) *Nutrient Database for Standard Reference, Release 12*. Riverdale, MD: Nutrient Data Laboratory, USDA Beltsville Human Nutrition Research Center (available at http://www.nal.usda.gov/fnic/food comp/).

Although cereal proteins are deficient in lysine, rice protein has one of the highest lysine contents among them, corresponding to an amino acid score of 59% in milled rice based on the Food and Agriculture Organization/World Health Organization/United Nations University (FAO/WHO/UNU) amino acid pattern of 5.8% lysine as 100% (Table 2). The solubility fractions of protein are about 15% of albumin–globulin (water- and salt-soluble), 20% prolamin (PB I) and 65% glutelin (PB II) in milled rice. Bran proteins are 66–98% albumins. Prolamin is poor in lysine but rich in sulfur amino acids. The high lysine content of rice protein is due to the low prolamin content.

The energy digestibility is higher in milled rice than in brown rice owing to the lower dietary fiber and phytic acid levels, as verified by the poor energy digestibility of rice bran (Table 2). The true digestibility (TD) of milled rice protein is also higher than that of brown rice, but the biological value (BV) is lower, resulting in a similar net protein utilization (NPU). Bran protein has a lower TD but a higher BV than brown and milled rice proteins. The amino acid score corrected for TD in rats proposed by the FAO as a protein quality index shows similar values to NPU for the rice proteins. Black or purple rice has a lower brown rice NPU (72%) and higher tannin level (0.6%) than red (NPU 83% and 0.2% tannin) and nonpigmented rices (NPU 97% and $\leq$ 0.02% tannin), but their milled rices have identical NPUs. Rice

Table 2 Essential amino acid profile and energy and nitrogen balance in growing rats of raw brown rice, milled rice and rice bran

Property	*Brown rice*	*Milled rice*	*Rice bran*
Arginine (g per 16 g N)	7.1	7.9	7.6
Histidine (g per 16 g N)	2.4	2.2	2.5
Isoleucine (g per 16 g N)	4.0	4.1	4.0
Leucine (g per 16 g N)	7.9	7.9	7.3
Lysine (g per 16 g N)	3.6	3.4	4.6
Methionine (g per 16 g N)	2.2	2.2	2.2
Methionine + cystine (g per 16 g N)	3.3	4.2	4.4
Phenylalanine (g per 16 g N)	4.9	5.1	4.5
Phenylalanine + tyrosine (g per 16 g N)	8.5	8.3	7.5
Threonine (g per 16 g N)	3.5	3.4	4.0
Tryptophan (g per 16 g N)	1.2	1.1	0.8
Valine (g per 16 g N)	5.6	5.8	6.3
Amino acid score[a] (%)	63	59	80
Digestible energy[b] (percentage of total)	94.3b	96.6a	67.4c
True digestibility (TD)[b] (percentage of diet N)	96.9b	98.4a	78.8c
Biological value[b] (percentage of digested N)	68.9b	67.5b	86.6a
Net protein utilization[b] (percentage of diet N)	66.7b	66.4b	68.3a
Amino acid score × TD (%)	61	58	63

[a]Based on 5.8 g of lysine per 16 g of N as 100%.

[b]In each line, mean values followed by the same letter are not significantly different at the 5% level (by Duncan's multiple range test).

From Eggum BO, Juliano BO, and Maniñgat CC (1982) Protein and energy utilization of rice milling fractions by rats. *Qualitas Plantarum Plant Foods for Human Nutrition* 31: 371–376.

United States Department of Agriculture (USDA) (1998) *Nutrient Database for Standard Reference, Release 12*. Riverdale, MD: Nutrient Data Laboratory, USDA. Beltsville Human Nutrition Research Center (available at http://www.nal.usda.gov/fnic/good comp/).

complements legumes in amino acid composition for human diets.

Cooking and parboiling reduces the TD in growing rats by 5–15%, with a corresponding increase in BV but little change in NPU. However, lysine digestibility remains close to 100%, and cystine digestibility drops to about 82%. The fraction that remains in the feces as fecal protein particles represents the lipid-rich core of large PB I, with less than 1% lysine in its protein but is rich in cystine. PB II is readily digested.

Starch varies in apparent amylose content (as determined by iodine colorimetry): waxy, 1–2%; very low amylose, 2–10%; low, 10–20%; intermediate, 20–25%; and high, 25–33%, all on a milled rice dry-weight basis. The glycemic index of cooked brown rice tends to be lower than that of cooked milled rice owing to the higher amounts of phytic acid and fiber in brown rice. Among cooked milled rices, the glycemic index decreases with increasing amylose content regardless of the cooking method (Table 3). Processing, including parboiling, tends to reduce the glycemic index. Resistant starch is $\leq 5\%$ and may not be as discriminating as the glycemic index in humans.

Breeding efforts to improve the nutritional value of rice grain include: a higher micronutrient density, (higher iron, zinc, and vitamin A), low phytic acid (high phosphate) content, low prolamin content absence of lipoxygenase-3 activity and allergenic globulin, and transgenic rice endosperm with vitamin A, with the soybean glycinin gene, and with the soybean ferritin gene.

Grading, Handling, and Storage

There is no international standard for milled rice size and shape. The International Rice Research Institute (IRRI) uses the following scale for size: extra long, > 7.50 mm; long, 6.61–7.50 mm; medium, 5.51–6.60 mm; and short, < 5.50 mm. For grain shape based on the length:width ratio, the following scale is used: slender, > 3.0; medium, 2.1–3.0; bold, 1.1–2.0; and round, ≤ 1.0. Grades are based on grain size and shape, degree of milling, percentage head or

Table 3 Comparison of glycemic index and *in vivo* resistant starch of various rice and rice products

Rice food[a]	*Amylose content*[a] *(%)*	*Glycemic index*[b]			In-vivo *resistant starch (%)*
		Percentage of glucose	*Percentage of bread*	*(no.)*	
Brown rice					
Waxy purple	0–2	78 ± 8	112 ± 11	(1)	
Low amylose	15–20	81 ± 7	116 ± 10	(3)	
Intermediate amylose	20–25	55 ± 5	79 ± 6	(3)	
High amylose	25–33	66 ± 7	94 ± 10	(1)	
Milled rice					
Waxy/low amylose	0–2/15–20	88 ± 3	126 ± 4	(3)	
Long grain	20–33	56 ± 2	81 ± 3	(13)	3.9
Intermediate/high amylose	20–33	59 ± 3	83 ± 5	(3)	
Instant rice	15–25	91 ± 4	128 ± 4	(2)	5.1
Milled parboiled rice					
Low/intermediate/high amylose	15–33	47 ± 3	68 ± 4	(13)	3.6 ± 0.9
Specialty rices	20–25	54 ± 1	78 ± 1	(4)	
Rice pasta, brown	15–20	92 ± 8	131 ± 11	(1)	
Rice noodles (Chinese)	25–33	58 ± 4	83 ± 5	(1)	
Rice noodles (Thai)	25–33	54 ± 5	77 ± 7	(2)	
Brown-rice cakes, molded	15–20	86 ± 9	123 ± 6	(2)	
Brown-rice cakes, molded	25–33	61 ± 5	87 ± 7	(1)	
Puffed rice	15–20	86 ± 7	123 ± 11	(3)	
Rice Bubbles	15–20	88 ± 7	126 ± 10	(2)	
Rice Krispies	15–20	82 ± 4	117 ± 5	(1)	
Rice Chex cereal	15–20	89 ± 4	127 ± 5	(1)	2.9
Rice bran	15–20	19 ± 3	27 ± 4	(1)	
White bread with 20% rice bran	25–33	55 ± 8	79 ± 12	(1)	

[a]Apparent amylose content: waxy 1–2%, low 15–20%, intermediate 20–25%, and high 25–33%.

[b]Mean ± standard deviation. Based on blood plasma glucose value for the glucose diet as 100%, and that for the white bread diet as 100%, respectively (GI of glucose is 0.70 of the GI of white bread). Number of reports in parentheses.

From Food and Agriculture Organization (1998) *Carbohydrates in Human Nutrition*. Rome: FAO; Foster-Powell K and Brand Miller J (1995) International Tables of Glycemic Index. *American Journal of Clinical Nutrition* 62: 869S–893S; Panlasigui LN (1989) *Glycemic Response to Rice, Ph.D. dissertation*, University of Toronto; Jenkins DJA, Cuff D, Wolever TMS *et al*. (1987) Digestibility of carbohydrate foods in an ileostomate: relationship to dietary fiber, *in vitro* digestibility, and glycemic index. *American Journal of Gastroenterology* 82: 709–717.

whole-grain milled rice, immature grains, damaged (discolored) and heat-damaged grains (chalky grains, red grains and red-streaked grains), aroma, and organic and inorganic extraneous matter.

An important property of harvested grain is its moisture content: 14% on a wet-weight basis is considered a safe storage value; grains become susceptible to fissuring from moisture adsorption stress below critical moisture contents of 12–16%, depending on the variety. Aging of sun-/oven-dried rough rice for up to 2 months after harvest at ambient temperature $< 15\,^{\circ}C$ improves milling yields and makes milled rice expand more during cooking and become more flaky. Heating during grain drying accelerates aging.

Processing and Food Uses

The per-capita consumption of milled rice is 86 kg per year in 1999 in Asia as compared with 4–31 kg in other continents. It is consumed mainly as milled rice. About 20% of rice is consumed as parboiled rice. Parboiling consists of boiling or steaming steeped rough rice until the hull starts to open, and then cooling and drying the gelatinized grain. Diffusion of bran B vitamins into the endosperm occurs during parboiling. Mycotoxins are a potential problem in parboiled rice. Milling involves dehulling, followed by removal of the outer 7–10% of the brown rice, either by friction or by abrasion, removing most of the pericarp, seed coat, nucellus, aleurone layer, and the germ. In the Engelberg mills, hull and bran are removed together in one step with high grain breakage. Milling is done in several steps in modern cone mills with tempering in between to minimize breakage. Many modern mills have shifted to milling at >14% moisture to minimize grain breakage and to moisture mist treatment (through hollow shaft) during milling to soften the bran and improve surface gloss.

There are various ways of cooking milled rices. In tropical Asia, they are prewashed to remove dirt, but this results in losses of B vitamins and fat. Presoaking for 30 min reduces cooking time (particularly of parboiled rice). Grain cracking during soaking is minimized by adjusting milled rice moisture content to 15% by high-pressure hydration to minimize cracking. Rice may be cooked in the amount of water that it will absorb, or boiled in excess water and the cooking liquor discarded. A steaming step is used in Indonesia and also for waxy rices. Oil or ghee may be added in the Middle East, and this reduces surface stickiness. Enriched rice premixes containing iron and B vitamins resistant to washing have been developed for rice, but have not been popular because of additional expense; the enriched rice can be readily distinguished from ordinary grain.

Apparent amylose content (AC) correlates positively with water absorption and volume expansion during cooking, and with the hardness of boiled rice. South Asian consumers prefer high-AC intermediate-gelatinization temperature (GT $> 70\,^{\circ}C$) rice, Southeast Asians prefer intermediate-AC, intermediate-GT rice. *Japonica* rices preferred in Japan, Korea, Taiwan, and northern China are mainly low-AC, low-GT ($\leq 70\,^{\circ}C$), with some having an intermediate AC in Europe. Only in Laos and north and north-east Thailand is steamed waxy rice consumed as a staple. The amylopectin of low-GT, high-AC rice has a longer mean chain length than the amylopectin of intermediate-GT, high-AC rice. Among waxy and low-AC rices, low-GT rice is a softer cooked rice than high-GT rice, but among intermediate- and high-AC rices, intermediate GT rice is a softer cooked rice than low-GT rice. Longer-chain amylopectin contributes to the flaky cooked rice of low-GT, high-AC rices. Amylopectin staling is lower in low-GT cooked rice than in higher-GT rice within each amylose type and is reversed by heating.

Specialty rices include Italian Arborio rice for making risotto, waxy rice (sweet rice with an opaque grain), Thai Jasmine rice (aromatic low-AC long grain) and Punjab Basmati (aromatic intermediate AC long grain that elongates when presoaked rice is cooked). 2-Acetyl-1-pyrroline was the first major aroma principle identified in raw and cooked aromatic rices, and others are being characterized.

Various rice products are prepared for which specific AC types are preferred (Table 4). Freshly and well-milled rice is preferred for rice products to prolong shelf-life by minimizing fat rancidity in the stored products. Parboiled rices are preferably high and intermediate AC, whereas extruded and flat noodles use mainly aged, high-AC, low-GT rices. Rice with a low-starch GT is preferred in rice puddings, breads and cakes, and beer adjuncts. Waxy and low-AC rices are preferred for rice wines (for a higher ethanol yield) and in frozen sauces, desserts, snacks, and sweets, because of their slow staling rate. Rice crackers are prepared from waxy and nonwaxy rices. Parboiled rice is preferred over raw rice for *idli* (pudding) and *dosai* (cake) with rice:black gram usually fermented at a 3:1 weight ratio. The thermophysical properties of starch such as the glass transition temperature, gelatinization temperature, enthalpy of raw starch, staled amylopectin melting (45–60 °C), amylose–lipid complex I (< 100 °C) and II (>100 °C) melting, and staled amylose melting (> 130 °C) of gelatinized starch affect the properties of rice

Table 4 Apparent amylose content (AC) type and other properties reported as preferred for various processed rice products[a]

Rice-based product	*AC type*				*Other properties*
	Wx	*Low*	*Int.*	*High*	
Parboiled rice	+	+	++	++	
Precooked/quick cooking	+	+	+	+	Based on table rice AC
Expanded rice	+	+	+	+	AC not a major factor
Expanded rice, molded		+	++	+	Waxy does not expand
Rice cereals/snacks		+	+	+	Low fat, texture affected by AC
Extruded rice food		+	+	+	Low fat
Rice-based infant food		+	+		Low fat
Rice flour and starch	+	+	+	+	Wet milling process, freshly milled rice
Rice crackers/biscuits	++	++	+	+	Japanese arare, senbei
Rice puddings	+	+			Japanese *uiro*
Rice breads		+	+	+	Low GT
Unleavened rice bread				++	Pakistani *roti*
Rice cakes, steamed	+	+	+		Fermented/nonfermented
Rice cakes, baked	+	+	+	+	Celiac disease, low GT
Flat noodles/rice paper		+	+	++	Low shear process
Extruded rice noodles			+	++	Hard gel consistency
Rice pasta		+	+	+	Raw and parboiled rices
Rice frozen sauces	+	+			Gel stability
Rice desserts/sweets	+	+			Gel stability
Fermented rice foods		+	+	++	Parboiled *idli, dosai*
Rice wines	+	++			Low protein, low fat
Beer adjunct		+			Low GT, low fat
Rice in batters and fried foods	+	+	++	+	
Rice in thickeners	+	+			Gel stability

[a]Wx, waxy; Int., intermediate; GT, gelatinization temperature.
Reported preferred AC type as raw material for rice products.
++ preferred more often than +.
Juliano BO (1998) Varietal impact on rice quality. *Cereal Foods World* 43: 207–211, 214–216, 218–222.

products in addition to the amylose–amylopectin ratio and protein content.

The use of rice bran in cereal products has increased in recent years owing to the hypocholesterolemic effect, in humans, of the factor(s) in the high (up to 4.4%) unsaponifiable fraction of its oil, such as τ-oryzanol (cycloartenol, 24-methylene-cycloartanol and campestrol esters of ferulic acid), tocopherols and tocotrienols, an analog of tocopherol (vitamin E). Defatted bran has no hypocholesterolemic activity, unlike in oats, where the active principle is soluble β-glucan. Inactivation of antinutrition factors – trypsin inhibitor, oryzacystatin, and hemaglutinin-lectin – and lipase and lipoxygenase that are concentrated in the rice bran by heat treatment and extrusion cooking improves the shelf-life of the bran and its nutritional value to poultry. By contrast, the antinutrition factor phytic acid is heat-stable. Phytic acid in rice bran and brown rice has been recently reported to have some medicinal activity in preventing some types of cancer. Phytic acid content of rice bran is highest among cereal brans (3–8% phytic acid, Table 1).

The total rice bran oil production in 1986–1988 was 600 000 tonnes per year, mainly in India, China, Japan, and Vietnam, according to FAO. This represented only 13% of potential bran oil production, requiring immediate oil extraction or stabilization because of bran lipase. The high hull contamination (70%) of rice bran produced by Engleberg mills makes oil extraction from their 'bran' uneconomical. Essential fatty acid content of rice oil is 34.2% 18:2 and 1.5% 18:3. Levels of oryzanol, tocopherols, and tocotrienols and unsaponifiable matter differ among crude oils, and oryzanol, tocopherols, and tocotienols may be reduced by up to 90% by conventional refining and deodorizing.

See also: **Amino Acids**: Properties and Occurrence; **Starch**: Resistant Starch

Further Reading

Chikubu S, Ishitani T and Ohtsubo K (eds) (1996) *Science of Rice* (in Japanese). Tokyo: Asakura Publishing Company.

De Datta SK (1981) *Principles and Practices of Rice Production*. New York: John Wiley.

Food and Agriculture Organization (FAO) (1985) *Rice Processing Industries. Proceedings of the FAO/UNDP Regional Workshop, Jakarta, 15–20 July 1985*. Bangkok: FAO Regional Office for Asia and the Pacific.

Huke RE and Huke EH (1990) *Rice Then and Now*. Manila: International Rice Research Institute.
International Rice Research Institute (IRRI) – Centro International de Agricultura Tropical (CIAT) – West African Rice Development Association (WARDA) (1997) *Rice Almanac*, 2nd edn. Manila: IRRI; Cali, Colombia: CIAT; Bouaké, Côte d'Ivoire: WARDA.
Khush GS and Toenniessen GH (eds) (1991) *Rice Biotechnology*. Wallingford, UK: CAB International; Manila: International Rice Research Institute.
Juliano BO (ed.) (1985) *Rice: Chemistry and Technology*, 2nd edn. St. Paul, MN: American Association of Cereal Chemists.
Juliano BO (1993) *Rice in Human Nutrition*. Rome: Food and Agriculture Organization; Manila: International Rice Research Institute.
Juliano BO (1997) *Rice Products in Asia*. Bangkok: Food and Agriculture Organization Regional Office for Asia and the Pacific.
Juliano BO (1998) Varietal impact on rice quality. *Cereal Foods World* 43: 207–211, 214–216, 218–222.
Juliano BO and Villareal CP (1993) *Grain Quality Evaluation of World Rices*. Manila: International Rice Research Institute.
Marshall WE and Wadsworth JI (eds) (1994) *Rice Science and Technology*. New York: Marcel Dekker.
Satake T (1990) *Modern Rice Milling Technology*. Tokyo: University of Tokyo Press.
Tsuno Foods and Rice (1998) *First International Symposium on Disease Prevention by IP6 and Other Rice Components, Kyoto, June 8–9, 1998, Abstracts*. Wakayama, Japan: Ito-gun.
United States Department of Agriculture (USDA) (1998) *Nutrient Database for Standard Reference, Release 12*. Riverdale, MD: Nutrient Data Laboratory, USDA Beltsville Human Nutrition Research Center (available at http://www.nal.usda.gov/fnic/food comp/).

RICKETS AND OSTEOMALACIA

J J B Anderson and S U Toverud, University of North Carolina, Chapel Hill, NC, USA
S C Garner, Duke University Medical Center, Durham, NC, USA

Introduction

The terms 'rickets' and 'osteomalacia' refer to conditions in which mineralization is impaired, usually as a result of nutritional deficiencies of vitamin D and/or calcium. Although both conditions share common physical and microscopic abnormalities, rickets, which is only seen in infancy and childhood, has more severe manifestations in deformities of bone because of the greater plasticity of the young skeleton. Even though deficiency of vitamin D is a frequent cause of rickets and osteomalacia, vitamin D nutriture of individuals and populations has not been well investigated, in part because of the significant contribution of sunlight exposure to the nutritional status of vitamin D and in part because of the difficulty of accurately assessing vitamin D intakes from foods and beverages. Although rickets still remains a problem in much of the developing world, preventive strategies have led to its virtual disappearance in the developed world – with a few notable exceptions – because of food fortification and supplementation of this essential fat-soluble vitamin. Osteomalacia and the less severe hypovitaminosis D, on the other hand, appear to be increasingly more serious public health problems in the developed nations. As longevity continues to increase, the older segments of populations in affluent nations are at risk of insufficient vitamin D.

Dietary Vitamin D Intakes in the USA

No national survey findings exist on vitamin D intakes by populations. A few studies involving small numbers of subjects have shown that intakes by elderly women and shut-in females and males have been consistently below recommended intakes; low intakes have also been verified indirectly from measurements of serum 25-hydroxyvitamin D concentrations of subjects living at northern latitudes during winter and spring months.

Vitamin D Recommendations

Recommended intakes, also known by the specific term dietary reference intakes (DRIs) in the USA, of vitamin D have been designated in 1997 as adequate intakes (AIs), rather than recommended dietary allowances (RDAs). In its recent reevaluation the Food and Nutrition Board of the Institute of Medicine did not think that sufficient data were available to establish with scientific certainty new RDAs for vitamin D. **Table 1** lists the recently established AIs for vitamin D. Intakes that meet these AIs should

Table 1 Recommended adequate intakes (AI) of vitamin D in the USA

Life-stage group: age and gender	*AI ($\mu g\ day^{-1}$)*
Infants 0–1 years	5
Children	
1–3 years	5
4–8 years	5
Males	
9–13 years	5
14–18 years	5
19–30 years	5
31–50 years	5
51–70 years	10
> 70 years	15
Females	
9–13 years	5
14–18 years	5
19–30 years	5
31–50 years	5
51–70 years	10
> 70 years	15
Pregnancy	
< 18 years	5
19–30 years	5
31–50 years	5
Lactation	
< 18 years	5
19–30 years	5
31–50 years	5

1 μg = 40 IU of vitamin D.
From Institute of Medicine, Food and Nutrition Board (1997) *Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride*. Washington, DC: National Academy Press, with permission.

prevent rickets or osteomalacia. Lower intakes of vitamin D by individuals who get adequate sunlight on a daily basis should also be able to prevent these diseases.

Primary Causes and Abnormalities

Human vitamin D deficiency, which may result from inadequate dietary intake and/or skin biosynthesis of vitamin D (Figure 1), is a primary cause of rickets. The active hormonal form of vitamin D, 1,25-dihydroxy-vitamin D ($1,25(OH)_2D$, also called calcitriol, has an important role in intestinal absorption of calcium and phosphate, bone metabolism and mineralization, and calcium homeostasis. Other roles are not yet well understood and include conservation of calcium and phosphate by the kidney, interaction of the hormone with the immune system, skin and skeletal muscle, and enhanced differentiation of cells. Rickets may also be caused by inadequate calcium intake alone, as in the case of children living in Nigeria. It is characterized by several clinical features and laboratory measurements of blood (Table 2).

Rickets describes the disordered growth and mineralization of the growth plate of the long bones. The resulting abnormalities, widening of the growth plate, delayed conversion of the growth plate cartilage to bone, and irregularly arranged bone trabeculae,

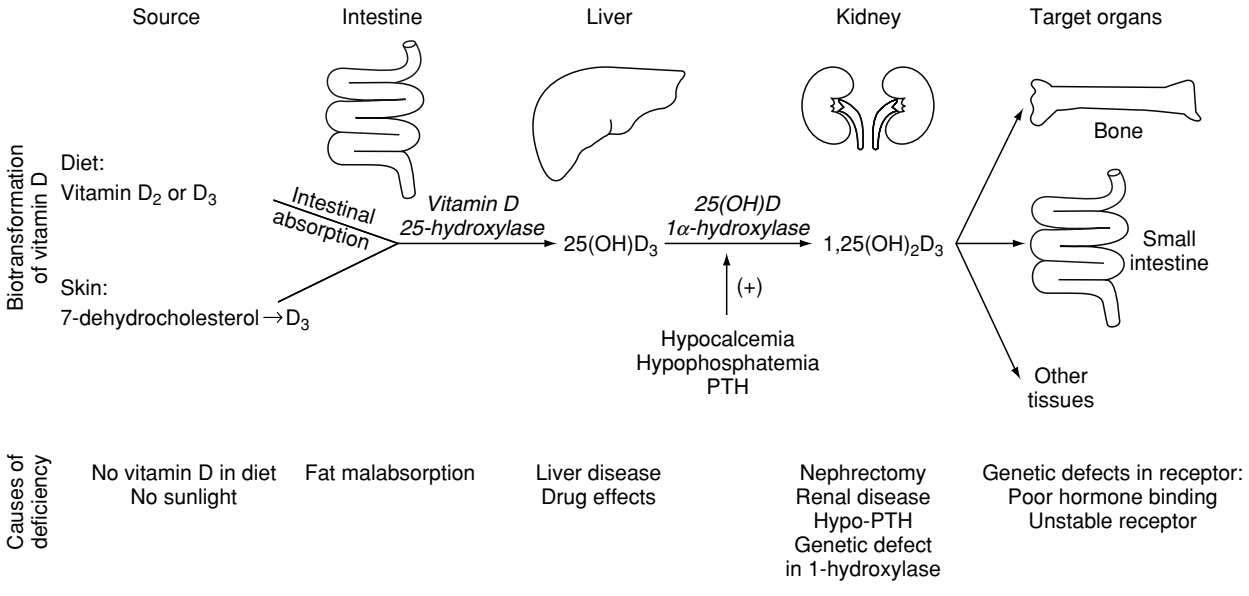

Figure 1 Causes of vitamin D deficiency. PTH, parathyroid hormone. Reproduced from Cholecalciferol and Ergocalciferol/Rickets and Osteomalacia, *Encyclopaedia of Human Nutrition*, Sadler MJ, Strain JJ and Caballero B (eds), 1999, Academic Press.

Table 2 Characteristic clinical features and blood serum measurements in rickets and osteomalacia

Age	*Clinical features*	*Blood serum measurements*
Children	Skeletal deformations (rickets) Impaired growth Undermineralized bone	Hypocalcemia Hypophosphatemia Secondary hyperparathyroidism Low 25-hydroxyvitamin D Elevated alkaline phosphatase
Adults	Undermineralized bone (osteomalacia) Fractures	Hypocalcemia Hypophosphatemia Low 25-hydroxyvitamin D Elevated alkaline phosphatase Elevated osteocalcin[a]

[a]This finding has not been consistently reported: see Vieth R (1990) The mechanisms of Vitamin D toxicity. Bone and Mineral 11: 267–272. Reproduced from Cholecalciferol and Ergocalciferol/Rickets and Osteomalacia, *Encyclopaedia of Human Nutrition*, Sadler MJ, Strain JJ and Caballero B (eds), 1999, Academic Press.

are only seen when the vitamin D deficiency occurs during childhood or adolescence, i.e., before the growth plates close.

Osteomalacia describes abnormalities resulting from delayed and reduced mineralization of mature bone, whether trabecular or cortical. This mineralization defect leads to a widened area of unmineralized bone matrix on the bone surface, called osteoid. Osteomalacia is observed at all ages.

Infants and young children suffer from growth retardation primarily because of the stunted growth of the long bones, and may become bow-legged because of the reduced mechanical strength of these bones. Skeletal muscle weakness and increased urinary excretion of amino acids may also occur.

Radiographic evidence of rickets includes widening of the growth plate of the long bones of both arms and legs, and reduced bone density. In osteomalacia of the adult, undermineralization of the bones may manifest as pseudofractures, which are radiolucent lines on the radiographs that mimic stress fractures. However, the pseudofractures are usually bilateral and often involve nonweight-bearing bones such as the scapula.

Biochemical Changes

Biochemical changes in most cases of rickets and osteomalacia include decreased serum calcium and phosphate levels because of the insufficient intestinal absorption of these ions, increased serum parathyroid hormone levels as a result of the reduced serum calcium concentration, and elevated serum alkaline phosphatase concentration. (Increased alkaline phosphatase results from increased osteoblastic activity related to bone formation.) In cases of inadequate supply of vitamin D or in liver disease, levels of serum 25-hydroxyvitamin D (25(OH)D, a metabolite) are invariably low, but the levels of $1,25(OH)_2D$ (the active hormone) may still be normal, depending on the severity of the condition. Kidney disease will lead to low serum $1,25(OH)_2D$ levels, while serum 25(OH)D may be normal.

Vitamin D Insufficiency and Seasonal Variation in Biosynthesis

A role for vitamin D insufficiency, rather than a frank deficiency, in the elderly is now being recognized as a contributing factor in low bone mass and associated fractures in the aging population. The age-related reduction in synthesis of vitamin D in elderly women and men living in institutions or confined to the home is often made worse by limited exposure to direct sunlight and a decreased intake of vitamin D. The dietary intake of vitamin D may be particularly low in Europe, in comparison with the USA where most dairy products are fortified with vitamin D. Supplements of vitamin D and calcium have proved effective at increasing bone mineral density while decreasing the rate of hip and other nonvertebral fractures in nursing-home residents in Europe.

Large numbers of elderly residents in the northern parts of North America and of Europe have been shown to have depressed circulating levels of 25-(OH)D and, in some reports, also low serum concentrations of $1,25(OH)_2D$, particularly during winter. Such seasonal differences have become well established in these geographic regions, but $1,25(OH)_2D$ insufficiency without symptoms of a full-scale deficiency have only recently been documented. Because of the increasing life spans of so many elderly people, and because so many of them reside in nursing homes or similar facilities which permit little exposure of residents to sunlight, frank deficiency of vitamin D is likely to become more common and may even reach epidemic proportions. The public health consequences of osteomalacic bone tissue and increased fracture incidence could be enormous. Accordingly, some investigators have raised

questions about the adequacy of the current allowances of vitamin D for the elderly.

Secondary Causes

Causes of rickets and osteomalacia include not only insufficient dietary intake of vitamin D or inadequate exposure to sunlight, but also diseases that may interfere with absorption and biotransformation of vitamin D_3 or D_2, as well as conditions leading to decreased tissue responsiveness to the hormone and to renal loss of phosphate (Figure 1). In fact, relatively few cases of rickets in the western world today are due simply to inadequacy of supply of vitamin D, and most of these are due to strict dietary or religious practices that involve vegetarian diets devoid of sources of vitamin D or dress codes that prevent exposure of the skin to sunlight.

Patients suffering from malabsorption of fats may have inadequate absorption of vitamin D from the small intestine. Liver disease may interfere with the conversion of vitamin D to 25(OH)D. Patients whose kidneys have been removed or are severely diseased will have decreased vitamin D hormone production and may develop osteomalacia.

Finally, inborn errors of metabolism may also lead to rickets. Vitamin D-dependent rickets type I is caused by a defect in the kidney enzyme 25(OH)D-1-hydroxylase, which is responsible for the synthesis of $1,25(OH)_2D$. In vitamin D-dependent rickets type II the receptor for $1,25(OH)_2D$ is defective and cannot mediate the genomic action of the hormone. In a third condition, called X-linked hypophosphatemia, defects occur in the renal reabsorption of phosphate, leading to low circulating phosphate levels. Rickets and osteomalacia are present in this condition despite the presence of normal adult plasma $1,25\text{-}(OH)_2D$ concentrations.

Public Health Consequences of Vitamin D Deficiency

Vitamin D, derived either from the diet or skin production, has become an increasingly important molecule in human function because of the extended longevity of human populations in many nations.

Much greater efforts by health organizations are needed to reduce the incidence and prevalence of rickets and osteomalacia in both the developing and the developed nations of the world. Assurance of adequate calcium intakes must be the first priority, and then emphasis needs to be placed on daily sunlight exposure for infants and children.

In recent years, supplements of vitamin D (400 IU or greater) have been increasingly recommended for elderly people, especially those who have become institutionalized or shut-in because of limited mobility or confining conditions and diseases. Since the elderly have declining capabilities of skin biosynthesis and vitamin D conversion, larger doses of vitamin D supplements are being recommended for this section of the population. This recommendation is supported by findings of low circulating blood concentrations of 25(OH)D or $1,25(OH)_2D$. Such low blood levels of vitamin D metabolites are now being defined as characteristic of vitamin D insufficiency rather than frank deficiency.

Oral therapy with 1α-hydroxycholecalciferol has become fairly widespread in the USA among elderly subjects with low bone mass or fractures. The 1α form must be converted in the liver to $1,25(OH)_2D$ by 25-hydroxylation. Dosages are generally low because of concerns about toxic effects. The efficacy of this therapy has not yet been fully evaluated.

Life-Cycle Changes in Vitamin D Production and Metabolism

The skin production of vitamin D does not remain constant throughout the life cycle, and it depends on the changes in the skin tissue itself as well as the environmental conditions in which people live. The skin of older individuals has considerably less capacity, for example, to produce vitamin D from its precursor. In the elderly, the concentration of 7-dehydrocholesterol is diminished, and the rate of conversion to vitamin D_3 is also reduced. When one considers that the elderly usually have reduced sun exposure as well, it becomes clear that they are at great risk of developing vitamin D insufficiency unless they have an adequate dietary supply of vitamin D.

Alterations in the vitamin D endocrine system in old age, such as a decrease in circulating $1,25(OH)_2D$ levels, have been implicated in the pathogenesis of postmenopausal osteoporosis, because reduced serum $1,25(OH)_2D$ levels have been found in some women afflicted with this condition. However, there is no evidence indicating a general age-related decrease in serum $1,25(OH)_2D$ concentration. A recent study of a large group of healthy white adults ranging in age from 20 to 94 years revealed no decline in serum 25(OH)D or $1,25(OH)_2D$ concentrations in either sex. In another study, involving women aged 26–88 years, serum $1,25(OH)_2D$ levels were actually higher after age 65 years than in the youngest subjects, but the rate of intestinal calcium absorption was unaltered, suggesting decreased responsiveness to the hormone with increased age followed by a compensatory increase in hormone synthesis.

Special Problems of Low Vitamin D Intakes

Four categories of individuals who may have significant deficiencies of vitamin D, whether from inadequate diet or insufficient skin exposure, are the elderly, strict vegetarians (vegans), infants and young children who have been maintained on breast milk for periods of a year or longer, and migrant populations. Each of these groups is characterized below.

Elderly

Elderly individuals typically stay indoors most of the time, and some may even be shut-ins who need routine assistance with activities of daily living. It is this latter group of elderly who are most likely to be vitamin D-deficient and who need daily supplementation of vitamin D, independent of season and latitude of their homes, as well as calcium supplementation. Low circulating serum 25(OH)D concentrations provide a strong indication of secondary hyperparathyroidism and bone loss. Long-term vitamin D treatment may improve bone health both by enhancing intestinal calcium absorption and reducing secondary hyperparathyroidism. An elderly group of women living in nursing homes in Lyons, France, had significant hypovitaminosis D that was readily corrected by administration of a combination of vitamin D (800 IU day^{-1}) and calcium (1200 mg day^{-1}) for 18 months, at which time nonvertebral fractures were significantly reduced.

Vegetarians

Vegans are at risk of vitamin D deficiency if they have little or no exposure to sunlight, because they typically consume unfortified plant foods. In addition to the risk of osteomalacia, vegans have other nutritional risk factors that put them at greater risk for osteoporosis.

Long-Term Breast-Fed Infants and Young Children

Rickets is occasionally reported in the USA in infants who are breast-fed for long periods of time. Breast milk is low in vitamin D and these nursing infants are given no other source of vitamin D in their diets. The risk of vitamin D deficiency is even greater in infants with dark skin, as increased skin pigmentation has also been established as reducing the skin's biosynthesis of vitamin D.

Migrant Populations at Risk

Other groups at risk for vitamin D deficiency are dark-skinned populations, especially Muslims, from the Middle East or other Asian nations who have migrated to the UK and other northern European countries. These immigrants are vulnerable to vitamin D deficiency because of inadequate exposure of their skin to sunlight and little or no consumption of vitamin D-fortified dairy products. Cultural practices result in adults, as well as children, wearing clothing that allows very little skin exposure to sunlight. Supplementation with vitamin D and calcium is effective in improving bone health in these populations.

Summary

While no accurate estimates of the prevalence of either rickets or osteomalacia exist, hypovitaminosis D is probably underdiagnosed in otherwise reasonably healthy adults, certainly in hospitalized adult patients. The strong assumption is that vitamin D consumption, and/or skin exposure for vitamin D biosynthesis, are insufficient to permit adequate formation of 25(OH)D, the circulating store of this fat-soluble nutrient. The result of a low 25(OH)D level is a fall in serum calcium and a continuous elevation of serum parathyroid hormone concentration. (Parathyroid hormone is the culprit responsible for increased bone turnover and decreases in bone mass and bone density.) Correction of this problem can be simply made by treatment with vitamin D supplements, typically accompanied by calcium supplements, and the consumption of vitamin D-rich foods, fortified or not. Prevention of this condition among older populations, however, appears currently to be woefully inadequate.

See also: **Aging – Nutritional Aspects**; **Bone**; **Calcium**: Physiology; **Cholecalciferol**: Properties and Determination; Physiology; **Inborn Errors of Metabolism**: Overview; **Osteoporosis**; **Vegetarian Diets**

Further Reading

Anderson JJB (1999) Plant-based diets and bone health: nutritional implications. *American Journal of Clinical Nutrition* 70 (suppl): 539S–542S.

Anderson JJB and Garner SC (eds) (1995) *Calcium and Phosphorus in Health and Disease*. Boca Raton, FL: CRC Press.

Anderson JJB and Miller CP (1998) Lower lifetime estrogen exposure among vegetarians as a possible risk factor for osteoporosis: a hypothesis. *Vegetarian Nutrition* 2: 4–12.

Brazier M, Kamel S, Maamer M *et al.* (1995) Markers of bone remodeling in the elderly subjects: effects of vitamin D insufficency and its correction. *Journal of Bone and Mineral Research* 10: 1753–1761.

Chapuy MC, Arlot ME, Duboeuf F *et al.* (1992) Vitamin D_3 and calcium to prevent hip fractures in elderly women. *New England Journal of Medicine* 327: 1637–1642.

Clemens T, Henderson S, Adams J and Holick M (1982) Increased skin pigment reduces the capacity of skin to synthesize vitamin D_3. *Lancet* 1: 74–76.

Dawson-Hughes B, Harris SS, Krall EA and Dallal GE (1997) Effect of calcium and vitamin D supplementation on bone density in men and women 65 years of age or older. *New England Journal of Medicine* 337: 670–676.

Egsmose C, Lund B, McNair P, Lund B, Storm T and Sørensen O (1987) Low serum levels of 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D in institutionalized old people: influence of solar exposure and vitamin D supplementation. *Age and Ageing* 16: 35–40.

Feldman D, Glorieux FH and Pike JW (eds) (1997) *Vitamin D*, chapters 40–43. London: Academic Press.

Fomon S and Strauss R (1978) Nutrient deficiencies in breast-fed infants. *New England Journal of Medicine* 299: 355–357.

Glorieux FH (ed.) (1991) *Rickets*. Nestlé Nutrition Workshop Series, vol. 21. New York: Raven Press.

Gloth FM III, Gundberg CM, Hollis BW, Haddad JG Jr and Tobin JD (1995) Vitamin D deficiency in homebound elderly persons. *Journal of the American Medical Association* 274: 1683–1686.

Holick MF (1996) Vitamin D and bone health. *Journal of Nutrition* 126: 1159S–1164S.

Holick MF (ed.) (1999) *Vitamin D: Molecular Biology, Physiology, and Clinical Applications*. Totowa, NJ: Humana Press.

Institute of Medicine, Food and Nutrition Board (1997) *Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride*. Washington, DC: National Academy Press.

Jacques PF, Felson DT, Tucker KL *et al*. (1997) Plasma 25-hydroxyvitamin D and its determinants in an elderly population. *American Journal of Clinical Nutrition* 66: 929–936.

Key LL Jr (1991) Nutritional rickets. *Trends in Endocrinology and Metabolism* 2: 81–85.

Lamberg-Allardt C, Karkkainen M, Seppanen R and Bistrom H (1993) Low serum 25-hydroxyvitamin D concentrations and secondary hyperparathyroidism in middle-aged white strict vegetarians. *American Journal of Clinical Nutrition* 58: 684–689.

National Research Council, Food and Nutrition Board (1989) *Recommended Dietary Allowances*, 10th edn. Washington, DC: National Academy Press.

Oginni LM, Sharp CA, Worsfold M *et al*. (1999) Healing of rickets after calcium supplementation. *Lancet* 353: 296–297.

Okonofua F, Gill DS, Alabi ZO *et al*. (1991) Rickets in Nigerian children: a consequence of calcium malnutrition. *Metabolism* 40: 209–213.

Ooms ME, Roos JC, Bezemer PD, van der Vijgh WJF, Bouter LM and Lips P (1995) Prevention of bone loss by vitamin D supplementation in elderly women: a randomized double-blind trial. *Journal of Clinical Endocrinology and Metabolism* 80: 1052–1058.

Sharp CA, Oginni LM, Worsfold M *et al*. (1997) Elevated collagen turnover in Nigerian children with calcium-deficiency rickets. *Calcified Tissue International* 61: 67–94.

Thomas MK, Lloyd-Jones DM, Thadani RI *et al*. (1998) Hypovitaminosis D in medical inpatients. *New England Journal of Medicine* 338: 777–783.

Vieth R (1990) The mechanisms of vitamin D toxicity. *Bone and Mineral* 11: 267–272.

Webb AR, Kline L and Holick MF (1988) Influence of season and latitude on the cutaneous synthesis of vitamin D_3: exposure to winter sunlight in Boston and Edmonton will not promote vitamin D_3 synthesis in human skin. *Journal of Clinical Endocrinology and Metabolism* 67: 373–378.

Zeghoud F, Vervel C, Guillozo H *et al*. (1997) Subclinical vitamin D deficiency in neonates: definition and response to vitamin D supplements. *American Journal of Clinical Nutrition* 65: 771–778.

RIPENING OF FRUIT

A R N Teixeira and R M B Ferreira, Universidade Técnica de Lisboa, Tapada da Ajuda, Lisboa Codex, Portugal

Introduction

Most fleshy fruits attain physiological maturity at an unripe state unsuitable for immediate consumption. The ripening process that follows, either on the tree or after picking, is the result of a number of physiological and biochemical processes which are revealed in a sequence of changes in color, texture, aroma, and taste, leading eventually to a physiological state at which the fruit is commercially considered as edible. However, in the context of sensorial changes the meaning of ripening is rather ambiguous as it is based on a subjective appreciation of a very intricate process that involves a number of metabolic changes. **Table 1** illustrates some of the most important organoleptic and physiological changes that occur during ripening and the biochemical processes most likely to be responsible for them.

The ripening process is mainly concerned with changes in components formed during fruit growth and development. So, many fruits are able to ripen after being detached from the mother plant, provided physiological maturity has been attained before picking.

Respiratory Patterns – Climacteric and Nonclimacteric Fruits

Even after detachment from the plant, fruits continue to display for some time the metabolic activities characteristic of living organisms. As changes in respiratory rates are good indicators of alterations in general metabolism, fruit respiration is useful as a reference to physiological state, which determines the potential storage life of a fruit. The first systematic studies of fruit respiration were carried out by Kidd and West in the 1920s with Bramley Seedling apples. Fruits, picked at different stages of development and stored at various temperatures, exhibited a typical respiratory pattern characterized by a decline in the respiration rate after picking which, sooner or later, depending on the storage temperature, was followed by a temporary rise in respiration (Figure 1). The visible changes normally associated with ripening (color, texture, and flavor) began at or shortly after the maximum rate of respiration. The sudden change in the respiration rate was recognized by Kidd and West as a critical phase in the life of the fruit, denoting the transition from maturity to senescence, and was named 'climacteric.' The same authors showed later that a similar increase in respiration also occurred in apples attached to the tree. In this situation the rise in respiration takes place at a slower rate but, eventually, reaches higher peaks than those attained by the detached fruits. Subsequent work indicated that many other ripening fruits exhibit a similar pattern of respiration, most of them, mainly tropical and subtropical fruits, displaying higher climacteric peaks than apples (Figure 2a). On the other hand, some fruits show a continuous decline in respiration throughout maturation and ripening (Figure 2b) and have been termed 'nonclimacteric.' A list of commercially important climacteric and nonclimacteric fruits is given in Table 2.

Table 1 Organoleptic and metabolic changes that occur during ripening

Organoleptic and physiological changes	*Biochemical changes*
Texture	Rise in water-soluble pectins and decreasing degree of esterification through pectolytic activity. Partial cellulose breakdown
Color	Degradation of chlorophyll and unmasking of underlying pigments. Synthesis of new pigments (carotenoids or anthocyanins)
Aroma and taste	Qualitative and quantitative changes in carbohydrates and organic acids. Synthesis of alcohols, esters, and other volatile compounds
Rise in respiration	Autocatalytic synthesis of ethylene. Increased protein synthesis and altered metabolic control of respiratory pathways

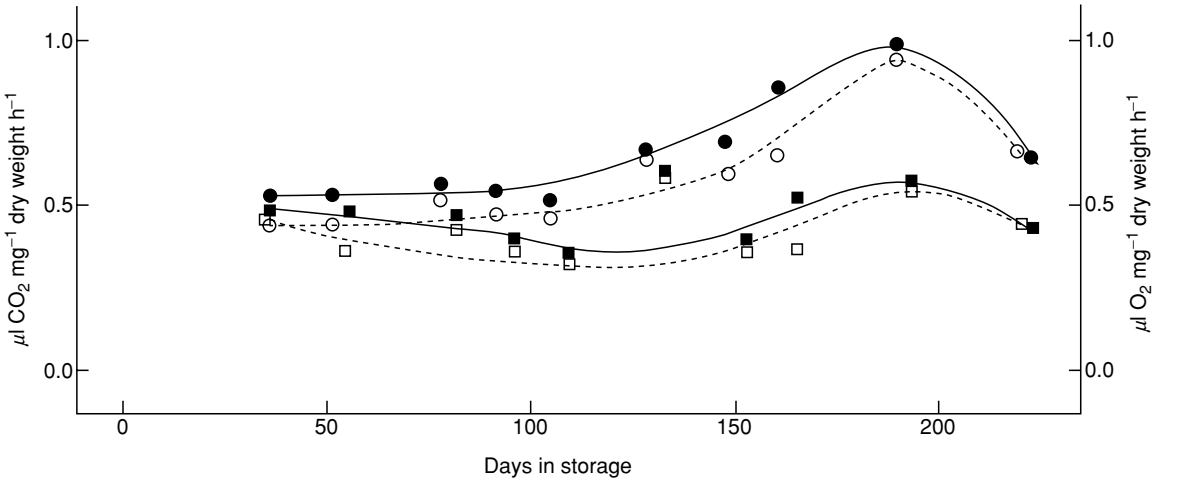

Figure 1 Climacteric rise in respiration of Rocha pears. Carbon dioxide evolution in fruits stored at 0 °C (filled circles) and −1 °C (open circles) and oxygen absorption at 0 °C (filled squares) and −1 °C (open squares). From Teixeira AR, Carmona MA, Barreiro MJ, Silva MJ and Cabral ML (1978) A counservaçáo frigorifica de pera Rocha. *Agronomia Lusitana* 39: 57–84, with permission.

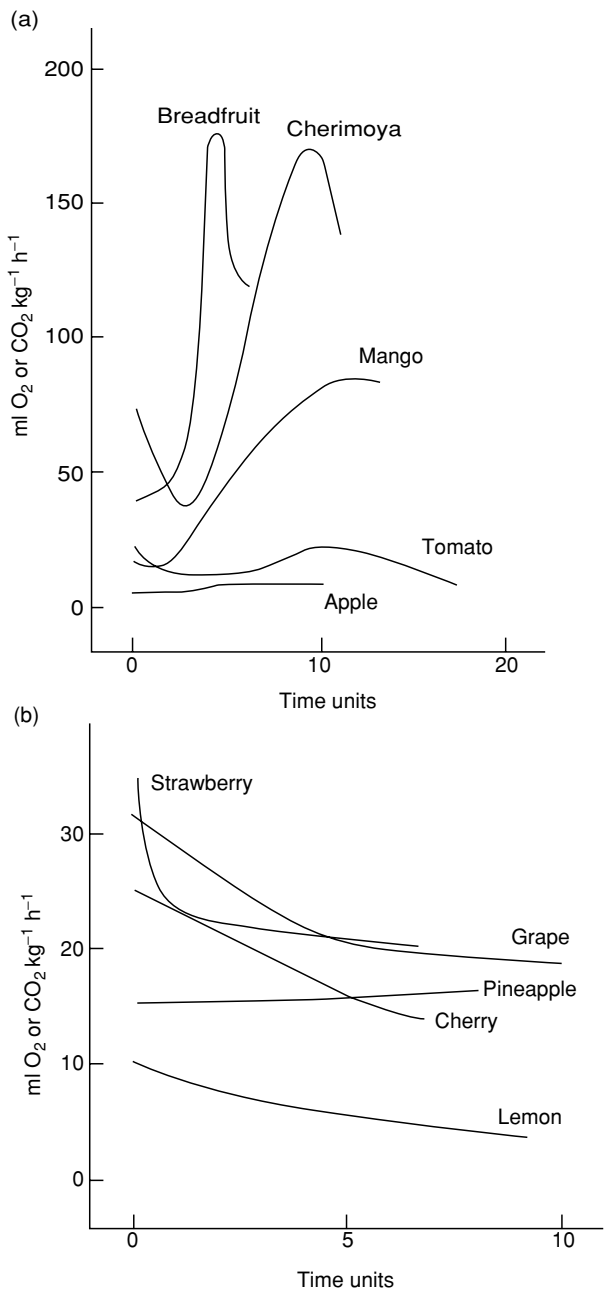

Figure 2 Respiration patterns of some (a) climacteric and (b) nonclimacteric fruits. From Biale JB and Young RE (1982) Respiration and ripening in fruits – retrospect and prospect. In: Friend J and Rhodes MJC (eds) *Recent Advances in the Biochemistry of Fruits and Vegetables*, pp. 1–39. London: Academic Press, with permission.

Table 2 Fruits with climacteric and nonclimacteric patterns of respiration

Climacteric	*Nonclimacteric*
Apple	Blueberry
Apricot	Cacao
Avocado	Cherry
Banana	Cucumber
Mango	Grape
Papaya	Grapefruit
Peach	Lemon
Pear	Lychee
Plum	Olive
Tomato	Orange
Watermelon	Pineapple
	Strawberry

Climacteric Behavior and Ethylene Production

It was recognized early in this century that ethylene can cause physiological responses in various plant tissues. The suggestion that fruit tissues themselves release gaseous substances that may affect the ripening behavior of nearby fruits was made by Coussins, who reported, in 1910, that gases emanating from oranges caused premature ripening of bananas. It was later showed that the active substance released by ripening fruit was ethylene and that in most climacteric fruits the rise in ethylene production occurred at or prior to the beginning of the respiratory upsurge. So the gas was considered responsible for the initiation of the climacteric and the related ripening processes and was named the 'ripening hormone.'

The climacteric or nonclimacteric behavior of fruits has been associated with the capacity of their tissues to synthesize ethylene. In climacteric fruits the respiratory increment is always accompanied, and often preceded, by an increase in ethylene levels within their tissues. This increase in endogenous ethylene induces an autocatalytic process of ethylene synthesis that, apparently, triggers off the respiratory upsurge of climacteric fruits. On the other hand, in nonclimacteric fruits the ethylene levels are always low and usually tend to decrease slowly during ripening at a rate approaching the rate of diminishing respiration. Climacteric and nonclimacteric fruits also differ in their responses to ethylene applied exogenously. The autocatalytic response of climacteric fruits to applied ethylene leads to the onset of climacteric respiration which, once initiated, cannot be stopped. Even if the exogenous gas is removed, the endogenous levels of ethylene will continue to increase once its catalytic synthesis starts. Thus, the respiratory rate at the climacteric peak does not depend upon the concentration of ethylene initially present. The respiratory rate of nonclimacteric fruits also responds to the application of exogenous ethylene, but does not exhibit an autocatalytic effect. Thus, in nonclimacteric fruits, the magnitude of the response varies proportionally to the concentration of the gas applied and the process can be reversed; that is, the respiration rate decreases to basal levels upon

removal of ethylene, and can be reproduced by renewed applications of the gas. Nevertheless, the application of exogenous ethylene accelerates the ripening in nonclimacteric fruits as it does in climacteric fruits when furnished in the preclimacteric phase.

The belief that ethylene is the only substance directly responsible for the climacteric rise in respiration and the ensuing ripening has recently been under close reexamination. The existence of a strong link between ethylene and respiration is well documented in the literature, being observed in many other plant tissues besides fruits. However, other volatile substances that are produced during ripening when applied exogenously may mimic the respiratory changes caused by ethylene. On the other hand, the onset of increased ethylene synthesis does not always precede the respiratory upsurge in climacteric fruits. The literature reports a variety of fruits in which the increase in ethylene production either coincides with (e.g., Cox's Orange Pippin apples and Anjou pears) or follows (e.g., avocado var. Fuerte and plum var. Wickson) the rise in respiration. It has been suggested that, in these fruits, ethylene is not the only factor that determines the initiation of the climacteric and of ripening. The minimum ethylene production rate that triggers off the rise in respiration, although varying between species and cultivars, is always very low (frequently less than 1 μl kg^{-1} h^{-1}, which is approximately equivalent to an internal concentration of 2 p.p.m.). So, the accurate determination of the relative timing of initiation of the rises in the rates of ethylene synthesis and respiration is rather difficult, particularly if, at this stage, other substances are being formed whose effects add up to those of ethylene. Notwithstanding the evidence suggesting that the stimulation of respiratory activity in fruits and other plant tissues may be induced by a number of factors, in the majority of climacteric fruits the autocatalytic ethylene synthesis begins before the respiratory upsurge and in most nonclimacteric plant tissues respiratory activity is tightly coupled to ethylene concentrations. Apparently, ethylene stimulates metabolism in general and may alter the control mechanism of existing metabolic pathways and induce changes in gene expression.

Climacteric Behavior and Ripening

If the link between ethylene and climacteric respiration is partially resolved, the role of the climacteric respiratory rise in ripening remains obscure. It is now evident that many of the physiological changes associated with ripening (e.g., softening, color changes, and ethylene synthesis) can be, with some manipulation, separated from the respiratory climacteric. Thus, some fruits can initiate softening and pigment synthesis without an increase in respiration and the application of certain plant hormones and of protein synthesis inhibitors may prevent ripening changes without obviating the climacteric rise in respiration. Such observations do not support the opinion that the increased rate of respiration supplies the extra adenosine triphosphate (ATP) required as the energy source for physiological transformations that occur during ripening, such as synthesis of proteins, nucleic acids, and pigments. In fact, the estimated energy demands of ripening are much lower than those provided by climacteric respiration; apparently, the energy generated by the basal respiratory rate is sufficient to drive the ripening events.

The evidence does not uphold the existence of an integrated relationship between the respiratory climacteric and ripening, but rather suggests that the rise in respiration during ripening may simply be a facet of ethylene action. The suggestion that ethylene may act by lowering the organizational resistance of the cell, mainly by increasing membrane permeability, has some experimental support. On the assumption that fruit tissues, like other living tissues, react to any injurious situation by changing their metabolism in order to maintain a homeostatic condition and that such changes are driven by respiration, it has been suggested that the respiratory climacteric is largely a homeostatic response, in an attempt to counterbalance the injury imposed on mitochondria by the initiation of senescence. That is, the degradative effects of incipient cellular senescence, probably accelerated by ethylene, impose a stress on mitochondria which leads to an increased respiratory rate. Such a suggestion is supported by the fact that mitochondrial structure and function, including coupling, are kept intact throughout the ripening process.

Effects on Texture

The texture of a fruit depends on the turgor of the cells as well as on the presence of supporting structures like cell walls and the cohesiveness of the cells. The latter property is conferred by the middle lamella, the intercellular structure that cements together the primary walls of contiguous cells and that is composed mainly of pectic material. Structurally, fruit cell walls are similar to primary cell walls of other plant tissues whose main components are polysaccharides, namely cellulose, hemicelluloses, and pectins. Essential for the structural integrity of cell walls is the presence of calcium. Calcium ions are particularly important in cell-to-cell adhesion, the cementing effect being primarily attributed to the formation of a calcium–pectate complex in the

middle lamella. The role of calcium in the maintenance of cell wall structure is particularly important in fruits in which the middle lamella gives a larger contribution as a cell wall component than in other plant tissues. The extensive softening that takes place during the ripening of many fruits is mainly attributed to changes occurring in cell walls, including the middle lamellae. (*See* **Cellulose; Hemicelluloses**.)

In ripe fruit, cell walls and middle lamellae are partly dissolved as a result of the action of a number of enzymes that break down pectins and cellulose, of which polygalacturonase, pectinesterase, and cellulase are the most extensively studied and whose activities are often correlated with the rate at which softening occurs during ripening. It is apparent from such studies that the major changes occurring in cell walls during ripening are associated with depolymerization and solubilization of the pectic substances due to the action of polygalacturonase and pectinesterase, respectively. The solubilization of pectin during the ripening of apples has been attributed to diminished levels of calcium ions in the cell walls and there is evidence that the softening of the apple fruit is associated with the transfer of divalent cations, particularly calcium, from the cell wall into storage compartments inside the cell. Postharvest calcium treatments that are carried out to prevent some physiological disorders, such as bitter-pit in apples, confer greater firmness to the fruits during storage.

The role of cellulase in tissue softening is not clear. Though it seems likely that hydrolysis of cellulose would weaken the cell wall structure, the experimental evidence is rather inconclusive as to its importance as a primary cause of fruit softening.

Effects on Sweetness

In fresh fruits the main compounds responsible for sweetness are soluble carbohydrates, i.e., sugars. It is apparent from the degree of sweetness of different ripe fruits that their sugar contents vary widely. Even within a particular species the sugar concentration in fruits may depend on the variety and on the environmental conditions and agricultural practices to which the mother plant has been subjected, prior to harvesting. As most climacteric fruits are harvested before ripening, considerable changes in sugar content may occur during storage. The average total sugar contents of most ripe fruits lie in the range 5–10% of fresh weight. Outside this range are, typically, lime and lemon with lower sugar contents (0.5–3.0%), and grape, cider apple and other fruits used for alcoholic fermentation purposes with higher contents (13–20%). (*See* **Carbohydrates**: Classification and Properties.)

The main individual sugars responsible for sweetness in fruits are the monosaccharides glucose and fructose and the disaccharide sucrose. Further monosaccharides may also be present in fruits, but their contribution to sweetness is negligible. The relative proportions of glucose and fructose vary with species but in many fruits glucose levels exceed those of fructose. This does not necessarily mean a higher contribution of glucose for sweetness since the sweetening power of glucose is 2–3 times lower than that of fructose. Two important exceptions are apples and pears in which fructose may be present in concentrations up to three times those of glucose. In grapes, berries, and oranges, both sugars are often present in similar amounts. (*See* **Fructose**; **Sucrose**: Properties and Determination.)

Sucrose is an important product of photosynthesis and the main form in which carbon is translocated from the leaves to other parts of the plant, including fruits. So, it is not surprising that in most fruits sucrose is the main respiratory substrate utilized for the provision of energy and intermediary metabolites for biosynthesis. The first step in the metabolism of sucrose, catalyzed by β-fructofuranosidase (invertase), is the hydrolysis of the disaccharide to glucose and fructose. These facts explain the predominance of glucose, fructose, and sucrose in most fruit tissues. The sucrose concentration varies from fruit to fruit but in most fruits the total hexose content exceeds that of sucrose, and in some fruits (e.g., cherry, grape, and tomato) sucrose is almost absent.

A significant fraction of the sugar translocated to the growing fruit is metabolized and used in various biosynthetic processes, another fraction being stored after conversion into starch, the most common reserve carbohydrate in plants. Starch breakdown is one of the most important biochemical changes that occur during the ripening of many fruits. Some fruits may contain, when mature but unripe, large amounts of starch, the levels of which decrease dramatically during ripening. For instance, bananas may contain up to 20% starch at the mature green stage and less than 1% when fully ripened. Large reductions in starch content also occur during the ripening of other fruits, notably mango, apple, and pear. The increased levels of sucrose, glucose, and fructose present in such fruits when ripened result primarily from the enzymatic hydrolysis of starch. However, some fruits do not accumulate starch during their development, attaining their mature unripe state with little or no starch at all (e.g., melon, pineapple, plum, and grape). In these fruits, carbohydrates are stored in the form of soluble sugars, their sweetness depending, largely, on the transport of sucrose from other parts of the plant, leaves being usually the main source.

Consequently, the optimum sugar content of these fruits cannot be reached if they are harvested before ripening. (*See* **Starch**: Structure, Properties, and Determination.)

Effects on Aroma and Taste

Aroma and taste are considered to be the two major components of flavor, a rather subjective attribute to which all the other senses contribute in an integrated manner. Most of the changes in aroma and taste, together with changes in color and texture, that are responsible for the commercial acceptability of fruits take place during the ripening process. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics.)

In spite of the important advances in flavor chemistry made in the last decades, our understanding concerning the metabolism of the compounds involved and the way they interact in the perception of aroma and taste is very incomplete. Aroma, being the sensation associated with smelling, requires the presence of volatile, relatively hydrophobic compounds. On the other hand, substances responsible for taste and other nonspecific saporous sensations are water-soluble and most nonvolatile, usually being present at higher concentrations than those responsible for aromas.

Despite the relatively low lipid contents of most fruits, lipid metabolism plays an important role in a number of ontogenic events, including ripening and senescence. Many of the most important aliphatic fruit volatiles are formed through an oxidative degradation of unsaturated fatty acids such as linoleic and linolenic acids. For instance, the action of lipoxygenase and a lyase on linolenic acid leads to the formation of 2-*trans*-hexenal or 2-*trans*-6-*cis*-nonadienal (depending on the site-specificity of the enzymatic attack), these being important aroma constituents of tomatoes and cucumbers, respectively. The aldehydes and ketones generated by lipoxygenase action can be converted further, by the action of dehydrogenases, to the corresponding alcohols, which usually have stronger aromas than the parent carbonyl compounds. For example, the dehydrogenation of 2-*trans*-6-*cis*-nonadienol gives 2-*trans*-6-*cis*-nonadienal, which is important in the aroma of cucumbers and melons.

The class of compounds that most frequently makes the major contribution to the pleasant fruity aromas developed during ripening is the esters of carboxylic acids. These are generated by esterification of the alcohols formed by reduction of the aldehydes derived either from oxidation of long-chain fatty acids or from amino acids. The mechanism of ester formation has not been satisfactorily explained but it is thought that esterification of the alcohols occurs with acyl-coenzyme A derivatives of fatty acids acting as acyl donors. While all the esters present may contribute to the characteristic aroma, certain individual esters have been associated with the aroma of specific fruits (Table 3). Some exceptions are known. For example, the aroma of peaches is attributed to the presence of lactones and that of raspberries to the ketone 1-(*p*-hydroxyphenyl)-3-butanone.

Volatile terpenoids are largely responsible for the characteristic aroma of citrus fruits. The monoterpenes citral and limonene exhibit distinct aromas of lemons and limes, respectively. Sesquiterpenes are also characteristic aroma compounds, such as β-sinensal in oranges and nootkatone in grapefruits.

Of the four basic taste sensations (sweet, sour, bitter, and salt), the dominant changes that occur during fruit ripening involve an increase in sweetness and a decrease in sourness (mainly due to acidity). However, changes in bitterness (due to terpenoids) are important in citrus fruits, as are changes in astringency (a taste sensation due to the association of phenolic substances, including tannins, with proteins in the saliva) that occur during the ripening of bananas, grapes, and other fruits. The changes in the concentration of acids during fruit development and ripening differ with the type of fruit. The optimal acidity is rather high for citrus fruits, somewhat less for pome fruits, and even less for tomatoes. Most fruit tissues accumulate excess acid in the preripening stage of development but, when ripe, each fruit has a range of acid-to-sugar ratio corresponding to optimum taste. In most fruits, citric and malic acids are the major contributors to the desired degree of acidity. (*See* **Acids**: Properties and Determination; **Phenolic Compounds**.)

In postharvest fruits, as in other plant living tissues, organic acids are in a constant state of flux. They may arise or be used up through the operation of

Table 3 Volatile esters associated with the aroma of individual fruits

Ester	*Fruit*
Benzyl benzoate	Cranberry
Ethyl butanoate	Orange
Ethyl 2-methylbutanoate	Melon
Ethyl 3-methylbutanoate	Blueberry
Isopentyl acetate	Banana
Methyl anthranilate	Concord grapes
Ethyl 2-methylbutanoate	Apple
Methyl (3-methylthio)propanoate	Pineapple
Methyl and ethyl derivatives of *trans*-2-*cis*-4-decadienoate	Bartlett pear

glycolysis, the citric acid cycle and, possibly, the glyoxylate cycle. Much of the loss of organic acids observed during ripening is attributed to their oxidation in respiratory metabolism as suggested by the increased respiratory quotient (RQ, the ratio of carbon dioxide evolution to oxygen absorption) during the climacteric. The RQ is approximately 1.0 when sugars are the respiratory substrates, increasing to about 1.3 when malate or citrate is the substrate, and further increasing to 1.6 when tartrate is respired. The rise in RQ during the climacteric is often accompanied by an increased activity of certain decarboxylation enzymes, such as malic enzyme and pyruvate decarboxylase. There is evidence that malic enzyme is synthesized *de novo* during the climacteric of pome fruits.

Changes in Color

During the ripening of almost all fruits there occurs a disruption in the organization of chloroplasts and their reorganization into chromoplasts. As the photosynthetic apparatus is dismantled and the chlorophyll degraded, the color of existing carotenoids is unmasked. This phenomenon may be responsible for the yellow-to-red colors of the ripe fruits, as in bananas where yellowing is due to the unmasking of carotenoids as a consequence of chlorophyll degradation. Although carotenoids are normally synthesized in green tissues, in most fruits that contain carotenoids ripening is accompanied by increased biosynthesis. During postharvest storage such synthesis may be affected by the surrounding atmosphere (decreased oxygen and increased carbon dioxide or nitrogen concentrations inhibit pigment development in the endocarp of some cultivars of oranges and in tomatoes), by light (it appears that light may stimulate its synthesis but it is not required for its induction; in detached green tomatoes it has been shown that light increases carotenoid synthesis, and red light is more effective than white or green light; far red light has an inhibitory effect which suggests a phytochrome-mediated response) and temperature (its effects vary from fruit to fruit but in most cases the optimal temperature for carotenoid synthesis is relatively low; it is around 24 °C for lycopene formation in tomatoes, the pigment not being formed above 30 °C but in other fruits in which lycopene is the main carotenoid, such as watermelons and red grapefruits, its synthesis is not affected by temperature). (*See* **Colorants (Colourants)**, Properties and Determination of Natural Pigments.)

The color of many ripe fruits, including pomes, several types of berries, grapes, oranges, cherries, peaches, plums, bananas, and figs, is due to the production of anthocyanins. These are conspicuous water-soluble plant pigments of phenolic character that accumulate in the vacuoles and are responsible for the pink, red, violet, and blue colors of fruits, flowers, and vegetables. The colours of anthocyanins are strongly affected by pH and can be intensified and stabilized by intermolecular associations with a number of metal ions such as those of aluminum, iron, manganese, and copper. The anthocyanins are typically located in the epidermal layers of the fruits but, like carotenoids, may also be present in the flesh. Anthocyanin synthesis during ripening is strongly stimulated by light. Apparently, this stimulatory effect has at least two mechanisms. By increasing the photosynthetic rate, light increases the availability of the carbohydrate needed to supply the energy and building blocks for anthocyanin biosyntheis. The second photoreaction involves the activation of phytochrome, a photoreceptor pigment that mediates several physiological processes in plant cells, including the synthesis of anthocyanins. Preharvest applications of some growth regulators, such as alar (*N*-dimethylaminosuccinamic acid), can induce earlier formation of anthocyanins in fruits.

Postharvest Storage of Fruits

After picking, the ripening of fruits can be controlled by temperature adjustment or by changing the concentration of gases (usually carbon dioxide and oxygen) in the surrounding atmosphere. These two major methods of fruit storage are usually referred to as 'air refrigeration' and 'controlled-atmosphere storage,' respectively, the latter being normally combined with some degree of refrigeration. Although effective in retarding ripening, both methods of environmental control impose stresses on fruits that may lead to different types of physiological disorders.

Low-temperature Storage

Lowering the storage temperature is one of the commonest methods for extending the shelf-life of fruits. With the exception of tropical and subtropical fruits, the lowest temperature that, in theory, still permits normal metabolism is near the freezing point of a fruit, which in most cases lies between 0 and −2 °C. The reduction of storage temperature decreases the rate of respiration and of metabolism in general by its direct effect on the rate of the enzyme-catalyzed reactions. In climacteric fruits, low-temperature storage also delays the onset of ripening by retarding the autocatalytic production of ethylene and so extending the preclimacteric phase of ripening. Lowering the temperature also reduces the degree of response to exogenous ethylene.

Controlled-atmosphere Storage

A decreased rate of respiration and of other metabolic reactions, including ethylene synthesis, can also be achieved by altering the composition of the atmosphere in which the fruits are stored, usually by increasing the carbon dioxide and decreasing the oxygen concentrations. In controlled-atmosphere storage the proportion of the gases is carefully controlled, usually within $\pm$ 1% of the desired value. When the control of the gas concentrations is less accurate, the practice is named modified-atmosphere storage. For example, in a confined atmosphere the carbon dioxide-to-oxygen concentration ratio may be increased simply as a consequence of the normal respiratory process or by sublimation of solid carbon dioxide (dry ice). Another type of controlled-atmosphere storage is hypobaric storage in which the fruits are kept in partial vacuum (0.1–0.3 atm). The decreased partial pressure of oxygen and the reduction of internal ethylene levels due to the increased diffusibility of the gases formed within the tissues retard the ripening of the fruits. (*See* **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs.)

The efficacy of controlled-atmosphere techniques in extending the storage life has been well demonstrated for a number of fruits. Nevertheless, their commercial use has been somewhat limited. Controlled-atmosphere techniques are expensive and some of the more profitable fruits either do not respond favorably to atmosphere regulation or have such a rapid market turnover that the need for an improved storage procedure is curtailed. Nevertheless, conventional controlled-atmosphere storage is currently used for apples and pears and the practice of modified-atmosphere consumer-sealed packs is becoming more popular due to the discovery of new packaging materials having a selective gas permeability that allows the retention of adequate gas concentrations when equilibrium is reached. Hypobaric storage is successfully employed in the storage of cut flowers but, to our knowledge, it has not been utilized for the commercial storage of fruits. (*See* **Chilled Storage**: Use of Modified-atmosphere Packaging.)

Physiological Disorders

Different fruit species or even different cultivars of the same species may exhibit different tolerances to low temperatures and to modified atmospheres. Exposure of a particular fruit either to a temperature or to an atmosphere with a carbon dioxide-to-oxygen concentration ratio outside the recommended range usually causes injuries that lead to a decrease in both quality and storage life. The extent of such injuries depends strongly on the duration of exposure.

The mechanism by which physiological disorders are induced is not well understood. Chilling injuries are generally attributed to a deregulation in metabolism leading to the underproduction of some essential metabolites and the overproduction of substances whose accumulation has toxic effects. For example, accumulation of ethanol, acetaldehyde, and oxaloacetic acid has been associated with low-temperature breakdown of certain fruits, such as apples. Visible effects of cold injury include tissue necrosis, surface pitting, internal browning, and failure to ripen. Some fruits (e.g., pears) normally exhibit a low susceptibility to cold injury, but for most fruits there is a critical temperature below which the fruits may be damaged, particularly if submitted to prolonged exposure. As recovery may often take place after short exposures to potentially damaging temperatures, it is believed that a minimum time of continuous exposure is required for injury to occur. In fact, interrupted exposures are often effective in diminishing damage. The lower temperature limit tolerated by fruits varies with their origin. Many fruits from temperate regions (e.g., apples) may tolerate temperatures in the range 0–4 °C, whilst the banana becomes susceptible at temperatures below 12 °C. Subtropical fruits, such as pineapples, avocados, and citrus fruits, show intermediary limits, often below 8 °C.

Identical metabolic imbalances are induced by extremes in atmospheric composition. Toxic levels of ethanol, acetaldehyde, and succinic acid have been shown to accumulate in certain fruits prior to injury symptoms. Some fruits, such as citrus and pome fruits, are rather susceptible to high carbon dioxide levels and the use of low carbon dioxide concentrations (1–5%) is recommended. For more tolerant fruits, such as cherries, grapes, plums, and strawberries, higher concentrations (15–30%) are usually advised. During storage, a minimum concentration of oxygen is required to support respiration. Below this level, anaerobic metabolism takes over and ethanol is produced. Although many fruits appear to tolerate oxygen concentrations below 5%, the exact limit for normal ripening is variable, depending on carbon dioxide concentrations, temperature, and duration of exposure.

Ionizing Radiation

Although this remains an isolated rather than an extensive method of food preservation, the application of ionizing radiation is effective in delaying the ripening of fruits. Depending on the fruit, doses of 0.3–3.0 kGy can prevent, temporarily or permanently, the onset of fruit ripening. However, the minimum dose required to achieve this effect often

exceeds the maximum tolerable dose. Even when successful in prolonging storage life, fruit irradiation is frequently hampered by a rapid softening of the tissues, possibly due to a direct or indirect activation of pectic enzymes. Actually, the mechanisms responsible for the observed effects of ionizing radiations on fruits are far from clear, but this is not surprising if one considers how numerous are the doubts over the biochemical and physiological control of fruit ripening even under normal conditions. (*See* **Irradiation of Foods**: Applications.)

See also: **Acids**: Properties and Determination; **Carbohydrates**: Classification and Properties;, **Cellulose**; **Chilled Storage**: Use of Modified-atmosphere Packaging; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Fructose**; **Irradiation of Foods**: Applications; **Phenolic Compounds**; **Starch**: Structure, Properties, and Determination; **Sucrose**: Properties and Determination

Further Reading

Abeles FB, Morgan PW and Saltveit ME (1992) *Ethylene in Plant Biology*. San Diego: Academic Press.

Bleeker A and Schaller G (1996) The mechanism of ethylene perception. *Plant Physiology* 111: 653–660.

Brady CJ (1987) Fruit ripening. *Annual Review of Plant Physiology* 38: 155–178.

Ecker J (1995) The ethylene signal transduction pathway in plants. *Science* 268: 667–675.

Frenkel C (1984) Factors regulating the respiratory upsurge in developing storage tissues. In: Palmer JM (ed.) *The Physiology and Biochemistry of Plant Respiration*, pp. 33–46. Cambridge: Cambridge University Press.

Friend J and Rhodes MJC (eds) (1981) *Recent Advances in the Biochemistry of Fruits and Vegetables*. London: Academic Press.

Gray J, Pictor S, Shabbeer J, Schuch A and Grierson D (1992) Molecular biology of fruit ripening and its manipulation with antisense genes. *Plant Molecular Biology* 19: 69–87.

Hulme AC (ed.) (1970) *The Biochemistry of Fruits and their Products*, vol. 1. London: Academic Press.

Lieberman M (ed.) (1983) *Post-harvest Physiology and Crop Preservation*. New York: Plenum Press.

Mattoo A and Suttle J (eds) (1992) *Plant Hormone Ethylene*. Boca Raton: CRC Press.

Salunkhe DK and Desai BB (1984) *Postharvest Biotechnology of Fruits*, vol. 1. Boca Raton: CRC Press.

Solomos T (1988) Respiration in senescing plant organs: its nature, regulation and physiological significance. In: Noodin LD and Leopold AC (eds) *Senescence and Aging in Plants*, pp. 111–145. San Diego: Academic Press.

Tucker GA and Grierson D (1987) Fruit ripening. In: Davies DD (ed.) *The Biochemistry of Plants*, vol. 12, pp. 265–318. San Diego: Academic Press.

Yang SF (1987) The role of ethylene and ethylene synthesis in fruit ripening. In: Thomson WW, Nothnagel EA and Huffaker RC (eds) *Plant Senescence: Its Biochemistry and Physiology*, pp. 156–166. Rockville: American Society of Plant Physiology.

RISK ASSESSMENT

C L K Chan, Safe Food Production NSW, Sydney, Australia

Risk Assessment Applied to Food Safety

As its name implies, risk assessment involves the evaluation of risks. The techniques for achieving this have been developed, improved, and utilized over many decades in a variety of fields, including the insurance industry, financial market analysis, project management, and environmental impact on public health. In recent years, there has been rapidly growing interest in applying risk-assessment techniques to food safety. In this context, the risk referred to is the risk of food-safety hazards causing human diseases.

Food may contain a myriad of hazards that can cause a variety of adverse health effects, ranging from mild symptoms through to serious illnesses and injuries and, in extreme cases, death. These hazards are usually classified as physical, chemical, or biological. Physical hazards may include fragments of glass, metal, plastic, or bone that could sustain injury or choking when the food is ingested. Chemical hazards include contaminants such as environmental pollutants, chemical residues derived from packaging material or product contact surfaces, as well as residues of agricultural or veterinary chemicals used in association with food production. Natural toxins are also chemical hazards in some food. Biological hazards are due predominantly to bacterial or viral pathogens but may also include other microorganisms such as nematodes and algae. The number of

food safety hazards that can potentially be present in food is large, and the complete elimination of them to achieve zero risk is not practically possible.

Risk is defined as a function of both the probability of a hazard causing an adverse health effect in consumers and the magnitude of that effect. Hence, it may be possible for a food to contain a certain pathogenic bacterium, by virtue of its exposure along the production and supply chain to different environments that may be sources of this pathogen. However, the risk of the microbiological hazard may be considered low because of a low probability and/or usually low level of contamination, low consumption of the food, and/or the minor nature of the ill effect that would be caused. Food producers, manufacturers, and the public generally have a notion of the magnitude and likelihood (hence risk) of various food-safety problems. This may be derived partly from experience and partly from impression. However, this is frequently unreliable. It has been demonstrated in surveys that the general public regard the issues of food irradiation and genetically modified food to be high-risk compared to problems of microbiological contamination, when reality is evidently quite the reverse.

Risk assessment provides a scientific framework for the estimation of food-safety risks, thus providing a reliable foundation for policy-making, strategy development, and practical management of food safety.

Risk Assessment: Principles and Practices

Risk assessment is defined as a process to evaluate scientifically the probability of occurrence and severity of known or potential adverse health effects resulting from human exposure to foodborne hazards.

Although the framework according to different organizations may vary somewhat, risk assessment generally consists of the following steps or components. These are:

1. Hazard identification: the identification of biological, chemical, and physical agents capable of causing adverse health effects and which may be present in a particular food or group of foods.
2. Exposure assessment: the quantitative and/or qualitative evaluation of the likely intake of biological, chemical, and physical agents via food as well as exposures from other sources, if relevant.
3. Hazard characterization: the qualitative and/or quantitative evaluation of the nature of the adverse health effects associated with the biological, chemical or physical hazard. Where possible, a dose–response assessment is performed. This is a determination of the relationship between the magnitude of exposure (dose) to a chemical, biological, or physical agent and the severity and/or frequency of associated adverse health effects (response).
4. Risk characterization: the process of determining the qualitative and/or quantitative estimation, including attendant uncertainties, of the probability of occurrence and severity of known or potential adverse health effects in a given population based on hazard identification, hazard characterization, and exposure assessment. The output of risk characterization is known as the risk estimate.

This framework represents a logical flow of thoughts in arriving at an estimate of food safety risks, as illustrated in Figure 1.

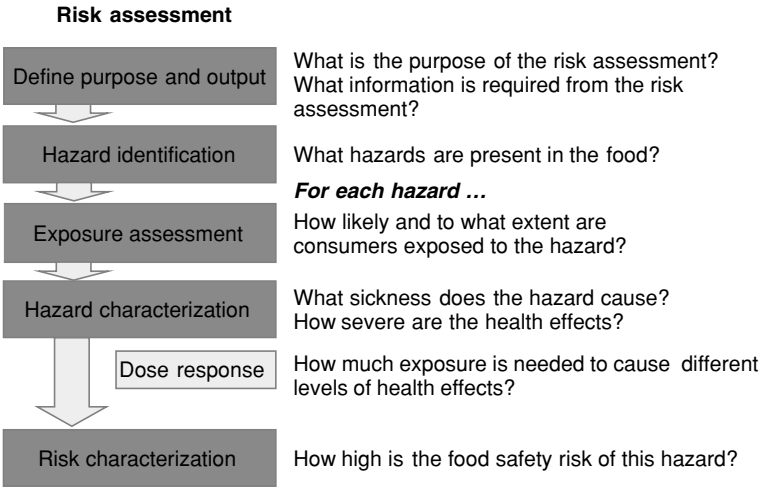

Figure 1 Components of risk assessment and the conceptual questions addressed.

A risk assessment analyzes the state of affairs at one point in time, utilizing the best available information at that time. As situations change, and significant new information becomes available, a reassessment or update should be carried out. A risk assessment can be thought of as a live and continuing process.

Risk Assessment of Chemical Hazards

In hazard identification, information as to what constitutes a hazard is frequently limited. The 'weight-of-evidence' approach is usually adopted, relying on a combination of relevant sources of information, including epidemiological studies, animal toxicological studies, *in-vitro* assays, and quantitative structure–activity relationships. In the majority of cases, clinical, and epidemiological data are unavailable or inadequate, and much reliance is placed on animal studies. These would include long-term chronic studies as well as short-term acute toxicity studies. They are designed to identify the range of toxicological effects, such as cancer, reproductive/developmental effects, neurotoxic effects, and immunotoxic effects. They can also provide information on the relevance of these effects for human risks through characterizing the mechanism of action. *In-vitro* studies may provide information on genotoxicity, pharmacokinetics, and pharmacodynamics.

One particular issue with animal toxicological studies is the need to use a relatively high dosage in order to identify the toxicological effects. However, realistic contamination levels in food are typically orders of magnitudes lower, and the significance of the identified health effects is uncertain. This can pose major issues for hazard characterization.

Animal experiments provide information on dosage at which no ill effects occur in the test animals; this is referred to as the 'no observed effect level.' A dose–response assessment would then need to extrapolate this animal data. There is much uncertainty in this, because the nature of the hazard and mode of human metabolism may change with dose, and the response in humans may be different to that of the experimental animals used. These uncertainties are compensated for with the use of safety factors. Frequently, a multiplication factor of 10 is used to accommodate the differences between how humans and animals might react to the hazard. Another factor of 10 is typically used to take account of the fact that some humans are more sensitive than others to the hazard in question. This yields an 'acceptable daily intake' (ADI) figure. If dietary exposure is below the ADI, it is assumed that there will be no ill effect. This is known as the 'threshold approach.' For genotoxic carcinogens, this approach is not considered appropriate, as it is assumed that there is a finite risk, even with the lowest possible dose. In such cases, a level of negligible/acceptable risk is set, frequently at one in a million.

Exposure assessment is carried out using diet studies. Direct monitoring of residues of the chemical hazard (e.g., organochlorine pesticides) in human tissues and body fluids is also gaining importance as a means of exposure assessment.

Risk characterization then draws on the outputs of hazard identification, hazard characterization, dose–response assessment, and exposure assessment to derive an estimate of the likelihood of adverse health effects in human populations as a consequence of the exposure. For threshold-acting hazards, risk is characterized by comparing exposure to ADI. Risk is notionally zero if exposure is lower than ADI. For nonthreshold acting hazards, risk is calculated as a function of exposure and potency.

As an example of how such a risk assessment can be used, a study may be carried out on the existing or proposed use of an agricultural or veterinary chemical in association with food-production animals. If the estimated risks are deemed unacceptable, consideration can be given to various possible risk-mitigation measures, such as means of reducing exposure. If the risks can be reduced to an acceptable level through such means, the chemical may be considered for registration. The setting of 'safe/tolerance levels' for various chemical residues in food is another application of risk assessment of chemical hazards.

Risk Assessment of Biological Hazards

Biological hazards include pathogenic strains of bacteria, viruses, helminthes, protozoa, and algae. Sometimes, toxins that some of these organisms may produce, e.g., enterotoxins produced by *Staphylococcus aureus* when growing to high numbers in food, are also grouped with biological hazards. Of these, foodborne pathogenic bacteria and viruses present, are by far, the greatest concerns in regard to food safety.

Microbiological risk assessment follows basically the same framework as chemical risk assessment. However, it has its unique set of difficulties and challenges. The list of foodborne pathogens is extensive, and pathogens can be found in abundance in the natural environment, including that for the production, harvesting, processing, storage, transportation, display, retail, and preparation of food. Some of the microbiological hazards, e.g., bacteria, can multiply rapidly in the food as it goes through the production and supply chain, whereas others can also produce toxins or form heat-resistant spores under certain conditions. The survival and proliferation of these

organisms are affected by the complexity of the food matrix, parameters of the manufacturing processes, as well as the temperature profile to which the food is exposed along the production supply chain. So, it is not unusual to find that a specific pathogen in a certain food matrix has caused many cases of foodborne illness, whereas the same pathogen in another food may rarely have been implicated. Moreover, the virulence and infectivity of microorganisms can change, depending on their interaction with the host and the environment. Genetic material can be transferred between microorganisms, leading to the transfer of characteristics such as antibiotic resistance and virulence factors. There is also evidence of very large differences in virulence between different strains of a pathogen, e.g., *Listeria monocytogenes*, that are almost impossible to distinguish by traditional taxonomic methods, and there are usually no easily identifiable virulence markers.

Human susceptibility to specific pathogens can also vary widely. For example, young, old, pregnant, and immunocompromised consumers are highly susceptible to the foodborne pathogen *Listeria monocytogenes*, but the healthy adult population is rarely affected. All these factors can make hazard identification even less clear-cut in comparison with identification of chemical hazards. Like the latter, epidemiological studies remain the most desired information base with which to advise biological hazard identification. Structure–activity relationship studies are used in the absence of, or in addition to, reliable epidemiological data. They compare the disease agent in question with a similar agent whose health effect is better understood.

Exposure assessment aims to define the pathway by which people will become exposed to the pathogen in the food, how much of the pathogen they are likely to ingest, and how frequently. This requires information on the occurrence rate and level of the pathogen in the food in question. These data are usually obtained by conducting microbiological surveys of the food. If surveys are not carried out at the point of consumption, predictive microbiology can be used to 'extrapolate' from the survey point (e.g., the end of production) to the point of consumption, based on typical temperature–time conditions to which the food is exposed, in order to estimate the dose exposure.

A dose–response assessment for microbiological hazards presents an even greater difficulty compared to chemical hazards. Human studies are very limited, and animal studies are generally not used. The relevance of the former is also restricted because of the variability in susceptibility between persons. There have been attempts to derive a dose–response relationship from available epidemiological data, but this approach has yet to be fully developed.

Risk characterization in microbiological risk assessment would include one or more of the following:

1. Risk estimate: depending on the purpose of the risk assessment, this may be expressed as a risk ranking of the hazards considered. This may be done in terms of high-, medium-, and low-risk categories, or in accordance with some arbitrary scale that would also show more quantitatively the relative risk levels. The risk estimate may also be expressed as the probability of human illness for defined exposure scenarios, e.g., in terms of x in a million chance of illness or x cases of illness per year in a specific population.
2. Characterization of uncertainty, variability, and confidence in the risk estimate: the uncertainties associated with each step or component of the risk assessment are identified and, if possible, quantified. Similarly, variability that reflects the heterogeneity of the population under consideration and the diversity in exposure is also identified and quantified. Confidence in the risk estimate is then discussed with reference to the adequacy of data and the weight of evidence.
3. Sensitivity analysis. This is used to show the influence of different variables on the outcome of the risk assessment and the range of response that may be encountered. It can also be used to highlight the impact on the risk estimate if various assumptions are changed.
4. Identification and evaluation of risk mitigation options: various options for mitigating or controlling the major risks are identified and evaluated in terms of their effectiveness.

The risk estimate may be expressed qualitatively, semiquantitatively, or quantitatively, depending on the purpose of the risk assessment, and the availability of relevant data and resources to allow quantification of the risk estimate. Accordingly, the risk assessment is referred to as a qualitative, semiquantitative, or quantitative risk assessment. A qualitative study is much less resource-intensive and may be adequate for some applications, such as prioritization of hazards for further investigation. A quantitative risk assessment may be necessary for purposes such as the setting of food standards and critical limits.

Risk Assessment of Physical Hazards

The possible health effects of physical hazards, such as glass, metal, and bone fragments, are generally obvious. There has been little demand to quantify

physical hazards in a systematic way in the context of food safety. However, attempts have been made to quantify risks according to the size and nature of extraneous objects.

Risk Assessment and Risk Analysis

Risk assessment is not an end in itself and is usually conducted for a specific purpose or objective. For example, it may be conducted to identify the higher-level risks of a food or manufacturing process, with the view of controlling those risks at acceptable levels though better policy setting, strategy formulation, and/or implementation/improvement of process controls. To achieve such an objective, risk assessment has to feed into a process of *risk management*, and both processes have to be carried out in conjunction with *risk communication*. The three elements, risk assessment, risk management, and risk communication, together are referred to as *risk analysis*.

Risk management is the process of weighing policy alternatives in the light of the results of risk assessment and, if required, selecting and implementing appropriate control options, including regulatory measures. It is responding to the findings of risk assessment and determining the most practical and effective way of achieving the purpose or objectives identified. Risk communication is the interactive exchange of information and opinions concerning risk and risk management among risk assessors, risk managers, consumers, and other interested parties. This ensures that the purpose, objectives, considerations, processes, findings, uncertainties, and implications of the risk assessment and risk management are clearly understood by all stakeholders throughout the conduct of the risk analysis. This tends to increase the quality of the process owing to a broadening of the input base for relevant information and expertise. It increases the transparency of the risk analysis and fosters ownership of the process and output. The interrelationship between the elements of risk analysis is illustrated in Figure 2.

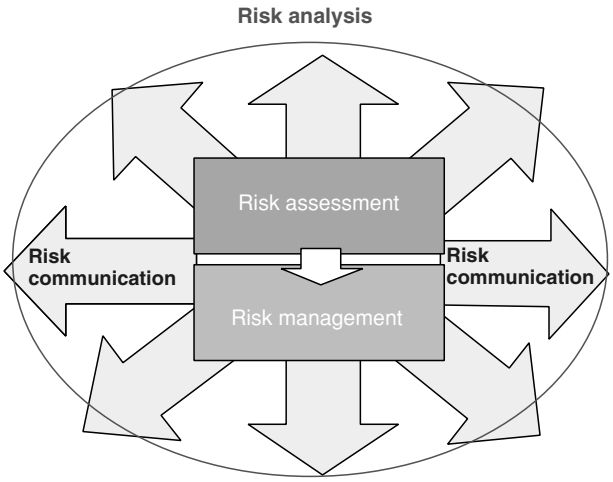

Figure 2 Components of risk analysis.

Application of Risk Analysis

Government Policy Setting

Developed countries have come to recognize food safety as a major social, economical, and political issue. The monetary and social costs of foodborne illnesses to society are very significant. There is also an increasing public awareness of food-safety issues, particularly after a spate of high-profile food-safety incidents, such as the BSE crisis in the UK, the dioxin contamination episode in Belgium, and the Snow Brand milk-poisoning case in Japan. As a result, there is an increasing expectation that the government fulfills a role of protecting the public from undue exposure to food-safety hazards, and in significantly reducing foodborne disease.

In 1997, President Clinton of the USA announced his Food Safety Initiative that recognized the importance of risk assessment as a scientific tool in achieving the goal of improving food safety. Major agencies such as the US FDA, FAO/WHO, and the European Commission have indicated that risk analysis underpins the formulation of their food-safety policy.

Whether at company, industry sector, or national level, food-safety hazards are numerous, whereas the resources for improving food safety are finite. In practice, resources must become more focused on reducing those risks that have the greatest consequences for human health. Risk assessment assists in such decision-making by providing a scientific estimation of the risks of various food hazards, and providing risk-management options that are scientifically sound. This allows the risk managers to prioritize and weigh up the options available. This may lead to government agencies formulating legislative requirements on food production, setting food standards, encouraging industry-based codes of practice, and/or undertaking educational programs.

Standard Setting

Food standards, such as those that specify the maximum allowable level of contaminants in food, or the amount of food additives permitted, are an important instrument for the assurance of food safety. This is particularly the case in international trade, where the food commodity being traded is manufactured in another country, and the history and conditions of

manufacture may not be fully known. International trade on food is based on agreements achieved at the World Trade Organization (WTO), including the Agreement on Sanitary and Phytosanitary Measures and the Agreement on Technical Barriers to Trade. Under these agreements, the Codex Alimentarius Commission (CAC) has the role of implementing the Joint FAO/WHO Food Standards Program, establishing international food quality and safety standards. These standards are adopted or used as models by many countries in their domestic food legislation and regulations.

The CAC takes a risk-based approach to setting of standards and has enunciated the principle of using sound scientific analysis and evidence as the basis of all its food standards. In practice, this means the use of risk assessment and risk analysis for the justification and derivation of food standards. This has set the scene for many countries also to adopt risk analysis as a tool for developing domestic standards.

Determination of Equivalence

It is the sovereign right of a county to determine a desired level of public health protection to be achieved in food production. The exporting country need not apply the same regulations for processing and production of food as the importing country, provided that the controls used will achieve the same level of safety. Risk assessment is a scientific tool that can be used to assess whether the imported food is of an equivalent level of safety as food produced locally or deemed acceptable from other sources. This provides a means of preventing artificial barriers to trade. In recognizing the importance of this science-based approach to fair trade, the WTO has required each member country's food safety measure to be based on risk assessment.

Equivalent systems, once established between trading partners, can result in acceptable export certification of shipments with minimum additional control. This simplifies trade procedures and reduces or eliminates the need for stringent and prescriptive trading conditions.

Underpinning HACCP

Hazard Analysis Critical Control Point (HACCP) has been recognized worldwide as the desired approach to achieve safe food production. There is a close relationship between hazard analysis within the context of HACCP and risk assessment/risk analysis. Much of the information required for hazard analysis is also required for the hazard identification and hazard characterization steps of risk assessment, such as identifying the nature, source, prevalence, and anticipated level of hazards. However, HACCP and risk assessment are different in terms of their scope, application, and focus.

HACCP is generally carried out at company level for a specific food product manufactured by a specific process and may be specific to a single factory, or even one particular production line. The objective is to manage and control the identified significant hazards at acceptable levels, thus ensuring production of safe food from the production line in question. Hazard analysis in HACCP generally does not systematically quantify risks, although a prioritization of hazards is involved, which can be considered an intuitive estimation of relative risk levels.

Risk assessment and risk analysis are normally carried out by government agencies, industry peak bodies, and the like. They have a broader focus than one specific manufacturing process and find application at a policy or strategic level. Usually, the study is carried out for a specific hazard in the context of a food commodity group, e.g., *Salmonella enteritidis* in egg and egg products. Sometimes, it is also carried out for a specific industry sector or food commodity grouping, e.g., significant hazards in seafood produced in a specific region. Systematic quantification (to the extent possible) of the risks of significant hazards is a fundamental element of such a study. The risk estimates provided by a risk assessment will give guidance in prioritization and strategic planning for the management of the risks. The proposed risk mitigation options then feed into the risk-management process to formulate practical work plans and risk control measures.

Owing to the commonality between HACCP and risk assessment/risk analysis, there is also an important interrelationship between them. Hazard identification conducted as an element of risk assessment is generally done more systematically and thoroughly, by virtue of the greater resources and expertise usually made available for such studies. In risk assessment, the hazards are also scientifically evaluated in terms of risks. This distinguishes those hazards that pose a significant threat from others that are no more than perceptions or are of a low impact. Therefore, risk assessment, where available, should provide the foundation for HACCP. The latter should address all the hazards of higher risks identified in the risk assessment and adopt one or more of the risk-management options provided, customized appropriately for the specific circumstances wherein the HACCP plan operates.

Benefits and Limitations

Risk assessment, with its clearly defined framework, provides the necessary discipline to ensure rigorous

problem definition, through examination of relevant data, assimilation of existing knowledge, and identification of variability and uncertainties. This provides for a structured, scientific, and objective inquiry into issues of food safety. Apart from providing a quantification of food-safety risk, a risk-assessment study also produces a report that serves as a database of relevant information, a documentation of considerations justifying the risk estimate and risk-management options, as well as a listing of critical information gaps. The transparency, objectivity, and scientific basis of the study and its findings engender critical thinking and understanding, and provide a firm foundation for sound decision-making, strategic planning, and effective action.

The major problem in conducting a risk assessment, in particular quantitative risk assessment, is the frequent lack of relevant data, whether epidemiological, prevalence, and/or dose–response data. It is also frequently necessary to make broad assumptions on which the robustness of the study rests heavily. As a consequence, this means that the conclusions of a risk assessment are sometimes accompanied by large uncertainties. In most cases, the generation of necessary additional data is not feasible or is very labor- and time-intensive.

The lack of relevant data may also mean that the study can only be qualitative, or at best semiquantitative. As such, the full benefits of a quantitative risk assessment cannot be realized, and the findings may be little more than what is already known based on an intuitive estimation of risk or impressions. In such cases, the main value of the risk assessment is restricted to having documented all relevant considerations in a systematic and transparent manner and identifying the knowledge gaps that need to be filled for future analyses.

Quantitative risk assessment requires the use of simulation modeling and/or animal studies. These models or studies cannot always be validated, and the reliability of the assessment findings would depend heavily on the validity of the modeling or the relevance of the animal studies. Related to this is the large variation in the susceptibility of people to foodborne hazards. Relevant data are generally much more available for the general population than for minority groups that may be more susceptible. Consequently, the risk levels determined may be an underestimation for certain consumer sectors.

A substantial quantitative study is very costly. For example, some quantitative microbiological risk assessments have taken up to 30 person-years to complete. This means that only resource-rich agencies, organizations, or collaborative teams are able to undertake such studies. However, in recognition of this, many major risk-assessment reports have been made freely available through the scientific literature and the Internet. This has allowed others to make use of the collated information and findings.

While the risk assessment process is both objective and systematic, subjectivity cannot be avoided in using the findings in risk management. This is because an acceptable level of food-safety risk is neither well defined nor universally agreed upon. Ultimately, the risk-management decision of what is an acceptable risk for a particular situation has to take account of cost–benefit analysis and public perception.

Risk assessment, being a scientifically based process, is designed to assess true risks to human health. However, the public's reaction to food safety is largely based on perceptions. Frequently, what the public perceives as unacceptable hazards in food and reacts emotively to is not necessarily proven as a major cause of foodborne disease. Nevertheless, issues surrounding these perceived hazards could bring about severe backlash on businesses, industries, and government by consumers and export markets. From a business viewpoint, the perceived risks are as real as the true food-safety risks. However, perceived risks are outside the domain of scientific food-safety risk assessment but may be dealt with as a commercial risk by other techniques.

See also: **Food and Agriculture Organization of the United Nations**; **Food Safety**; **Hazard Analysis Critical Control Point**; **Legislation**: International Standards; Additives; Contaminants and Adulterants; Codex; **World Health Organization**

Further Reading

Buchanan RL and Whiting RC (1998) Risk Assessment: A means for linking HACCP plans and public health. *Journal of Food Protection* 61(11): 1531–1534.

Codex Committee on Food Hygiene (1999) *Principles and Guidelines for the Conduct of Microbiological Risk Assessment. CAC/GL-30*. Codex Alimentarius Commission.

Codex Committee on Food Hygiene (2001) *Draft Response to the Codex Executive Committee Regarding Clarification of the Terms 'Hazard Analysis' and 'Risk Analysis'*. Thirty-fourth Session, Bangkok, Thailand.

FAO (2000) Report of the Joint FAO/WHO Expert Consultation on Risk Assessment of Microbiological Hazards in Foods. Rome: FAO.

Frewer LJ, Howard C, Hedderley D and Shepherd R (1997) The elaboration likelihood model and communication about food risks. *Risk Analysis* 17(4): 759–770.

Hathaway SC (1997) Development of food safety risk assessment guidelines for foods of animal origin in international trade. *Journal of Food Protection* 60(11): 1432–1438.

Herber RFM, Duffus JH, Christensen JM, Olsen E and Park MV (2001) Risk assessment for occupational exposure to chemicals. A review of current methodology. *Pure and Applied Chemistry* 73(6): 993–1031.

ILSI Risk Science Institute (2000) *Revised Framework for Microbial Risk Assessment*. Washington DC: ILSI Press, International Life Sciences Institute.

International Commission on Microbiological Specifications for Food (ICMSF) Working Group on Microbial Risk Assessment (1998) Potential application of risk assessment techniques to microbiological issues related to international trade in food and food products. *Journal of Food Protection* 61(8): 1075–1086.

Lammerding AM and Paoli GM (1997) Quantitative risk assessment: an emerging tool for emerging foodborne pathogens. *Emerging Infectious Diseases* 3(4): 1–7.

McNab WB (1997) A literature review linking microbial risk assessment, predictive microbiology, and dose–response modelling. *Dairy, Food and Environmental Sanitation* 17(7): 405–416.

Potter ME (1996) Risk assessment terms and definitions. *Journal of Food Protection* 59 (supplement): 6–9.

Van Schothorst M (1997) Practical approaches to risk assessment. *Journal of Food Protection* 60(11): 1439–1443.

WHO (1995) Report of a Joint FAO/WHO Expert Consultation *Application of Risk Analysis to Food Standards Issues*. Geneva: WHO.

WHO (1999) Report of a Joint FAO/WHO Expert Consultation *Risk Assessment of Microbiological Hazards in Foods*. Geneva: WHO.

Roller Milling Operations *See* **Flour**: Roller Milling Operations; Analysis of Wheat Flours; Dietary Importance

Root Vegetables *See* **Cassava**: The Nature of the Tuber; Uses as a Raw Material; **Potatoes and Related Crops**: The Root Crop and its Uses; Fruits of the Solanaceae; Processing Potato Tubers; **Vegetables of Temperate Climates**: Commercial and Dietary Importance; Cabbage and Related Vegetables; Broccoli-type Brassicas; Leaf Vegetables; Oriental Brassicas; Carrot, Parsnip, and Beetroot; Swede, Turnip, and Radish; Miscellaneous Root Crops; Stem and Other Vegetables; **Vegetables of Tropical Climates**: Commercial and Dietary Importance; Root Crops of Uplands; Root Crops of Lowlands; Edible Aroids

Roughage *See* **Dietary Fiber**: Properties and Sources; Determination; Physiological Effects; Effects of Fiber on Absorption; Bran; Energy Value

RUM

L Fahrasmane and A Parfait, Institut National de la Recherche Agronomique, Petit-Bourg, French West Indies, Guadeloupe

Introduction

The term 'rum' is used to describe the distillates originating from the fermented products of sugarcane (*Saccharum* spp.), that is to say fresh juice, molasses (the thick brown liquid separated from raw sugar at the end of the sugar manufacturing process), and syrup obtained from sugarcane juice concentration (**Table 1**). The fermented liquid must be distilled under 96% ethanol by volume (Directive 89/1576/EEC).

Rum production has always been a way to use and value molasses. Thus, in the French West Indies islands, *de sucrerie* rum, obtained from molasses, is produced from distilleries attached to sugar plants. However, in the Caribbean French colonies, Guadeloupe and Martinique and also in Haiti, some small sugarcane planters who were disadvantaged, during

Table 1 Average percentage of components in sugarcane, and raw materials used in the distillery

Component	*Cane and distilleries raw material*				
	Cane	*Juice*	*Syrup*	*HTM*	*Molasses*
Water	69.0	78.0	29.0	16.0	20.0
Sugar	16.0	20.0	66.0	77.0	62.0
Nonsugar	3.0	2.0	5.0	7.0	18.0
Fiber	12.0				

HTM, high-test molasses.

Table 2 Wash components, as a percentage, at the beginning of the nineteenth century

Component	*North Caribbean*	*Jamaica*
Molasses	10	6
Scum	30	36
Slops	20	50
Water	40	8

the first half of the eighteenth century, by the development of central sugar plants and a general reduction in sugar sales began to distil rum directly from the fresh juice of their harvest or from the syrup resulting from the juice concentration. This was how *agricole* rum was born. It is produced by independent distilleries, outside the sugar industry, using the raw material from farming.

Nowadays the world rum market is dominated by the light rums which are produced from molasses, in modern productive distilleries. Light rum is marketed white or aged, and drunk as an ingredient in long drinks. However, traditional rums are still produced and sold. These are characterized by a mixed fermentation involving alcohol-producing yeast strains and cofermenting bacteria. They are distilled at a lower ethanol content than the light rum (about 80% by volume) and are richer in aromatic compounds. Jamaican type (also known as heavy rum or *grand arôme*, is a typical example of these traditional products.

History and Uses

The British term 'rum' referred to the alcoholic liquors distilled from fermented sugarcane products, coming from the British colonies of the Caribbean. They were generally of a better quality than those produced in the French colonies, where they were referred to as tafia and guildive. Indeed, rum has had and still does have various names throughout the world, including some archaisms, such as kill-devil, the obsolete word rumbullion, guildive which is probably a French adaptation of kill-devil, tafia, grappe, arrack (Mauritius), and gavine (South Africa). Nowadays, it is rhum in French, aguardiente and ron (in Spanish-speaking Latin America), cachaça, pinga, aguardente (Brazil), clairin (Haitian traditional rum, with 30–35% alcohol content), toaka (Madagascar), and kana (Guinea-Bissau).

Although the cultivation of sugarcane and its transformation into sugar date back to time immemorial, rum production is more recent; it developed with sugar production in the Caribbean islands in colonial plantations in the seventeenth century. Every plantation had its distillery. Rum appeared when the world sugar production, which was only from sugarcane, was taken from the Mediterranean area to America. The increase in sugar production generated important molasses volumes. The European colonists discovered that they could take advantage by distilling fermented washes from the molasses and sugary wastes such as scum.

Among the first writers to mention rum was the father Du Tertre, who stayed in the French West Indies from 1640 to 1657. At much the same time, Robert Ligon in his *History of the Island of Barbados* (1657) described the preparation of a fermented beverage called punch. Father Labat, who arrived in the West Indies in 1694, described the elaboration of guildive in his *Nouveau voyage aux isles de l'Amérique* (*New Journey to the American Islands*; 1722).

In those days, rum was a byproduct of the sugar industry. The manufacturing process was simple: the fermentable sugar was seldom derived from cane juice but came from the scum that comprised the sugary impurities raised during juice boiling, and molasses rich in sucrose were drained from the sugar production process. These sugary resources, diluted – around 100 g l^{-1} of sugar – in water and/or distillation waste waters, were spontaneously fermented (Table 2). Then the fermented wash was single- or double-distilled in a pot still. The manufacturing process in use in the seventeenth and eighteenth centuries was diverse. The quality of the product was often poor, or frankly bad, because of the low microbial quality of the raw materials (scum, molasses, waste waters), the rudimentary condition of the fermentation process, and above all the poor resolving power of the distillation apparatus. Spirits were often derived from washes inside which acetic, butyric, and even putrid fermentation had occurred, and the pot stills were unable to provide the rectification to remove undesirable compounds. Moreover, the alcohol content was weak, often under 42% (by volume), especially in the products from the French colonies.

The phylloxera crisis in the nineteenth century had involved a high demand in alcoholic products including those from the rum distillery. In the meantime there was development of industrial processes with rational management and mass production. Pasteur's works on fermentation allowed microbial control in the washes. This evolution and new knowledge were applied in rum production where there was a chemical comprehension of distillery products and pursuit of productivity. During the second part of the nineteenth century appeared an untraditional process, which featured microbial control and bacteria elimination in the washes, a plant with voluminous fermentation vats, and a multistage column for distillation which involved a high level of rectification of the distillates.

Rum was first a drink for slaves and seamen. It was considered as a coarse drink, not appropriate for the palates of the upper classes. Admiral Penn, after he took Jamaica in 1655, instituted the daily rum ration distribution to his sailors. Admiral Vernon, who was nicknamed Old Grog – grog is a mixture of one portion of rum, two portions of water and some lemon juice – replaced the Admiral Penn ration, with grog in 1731. Rum was distributed daily to sailors of the British Royal Navy until recent times. In the West Indies the British blended rum with various foodstuffs, including tea, sugar, lemon, and cinnamon and named the cocktails punch. In the eighteenth century there were many rum liquors known by the name island liquor.

The British encouraged rum production from the beginning, because it was favorable for traffic with the North American Indians, and the slave trade on the African coast. Whisky production was low at that time. Barbados, Jamaica, and New England were colonies where rum production was developed. In 1765, two distilleries were situated in Liverpool, where they produced rum to supply ships bound for Africa.

From the eighteenth century, rum-based drinks became numerous. An important feature of rum utilization is its ready association with fruit aroma and vegetable extracts in the mixing of punch, cocktails, planteurs and, more recently, rum and cola; some mixes are sold ready-to-use. Rum has the distinction of being an excellent basic ingredient for composite drinks. The earliest English terms, punch and grog, are still largely used, although the beverages are rather different from those in the eighteenth century. Nowadays one can find, for example:

- curaçao from the Dutch West Indies, produced after soaking orange peel in rum: shrub is very similar to curaçao
- cocktails which are mixtures of rum with other liquids, mainly tropical fruit juices and spices: punch and planteur from the French West Indies are similar
- Alexandra is a sophisticated mixture of aged rum and cream
- in Brazil there is a variety of Batida, which is a mixture of fruit and cachaça
- daïquiri or daiquiri was created in 1898 by an American in Cuba. Originally it was a light rum like Bacardi mixed with lemon juice
- Cuba libre, one of the most popular drinks in North America, is a mixture of Coca-Cola and rum

Another very interesting feature of rum which is not often emphasized is its use in eaten products, confectioneries, icecreams, chocolate, and foods. For example, in France, more than 40% of marketed rum is used as an aromatic resource, in confectionery. Baba (sponge-cake) is one of the best-known confectioneries containing rum. Traditional rum and heavy rum are used to flavor various foods because both are rich in aromatic compound and can endure physical hard treatment, such as the high temperature encountered in food processing. Light rum, on the other hand, is poor in aromatic compounds and is not suitable as a flavoring agent in foodstuffs.

Manufacturing Process

The first stage of rum elaboration is preparation of the mash. The goal is to favor the alcoholic fermentation of sugars by yeasts, either occurring naturally or introduced in the fermentation media by the distiller. During fermentation there is also a production of heat which in small vats is quickly dissipated, but in larger vats may raise the temperature such that it kills the yeast. When fermentation is finished, the mash is distilled either in a pot still in continuous distillation columns, or in a multistage distillation column (Figure 1). This operation makes it possible to separate from the fermented media the volatile components and concentrate them in the distillate. The resulting raw rum is at a higher percentage in ethanol (around 80%) than the marketed products, so it is diluted before sale. It contains aromatic volatile compounds which distil together with the ethanol. These congeners are generally in small quantities (0.0001–0.001% of the ethanol, in weight). Depending on their proportion, they add significantly to the taste and aroma of the raw rum.

More than 95% of the rum produced in the world is derived from molasses and high-test molasses (HTM). Molasses is rich in sugar, and has a density comparable to honey: 1.4. It also contains nonsugar

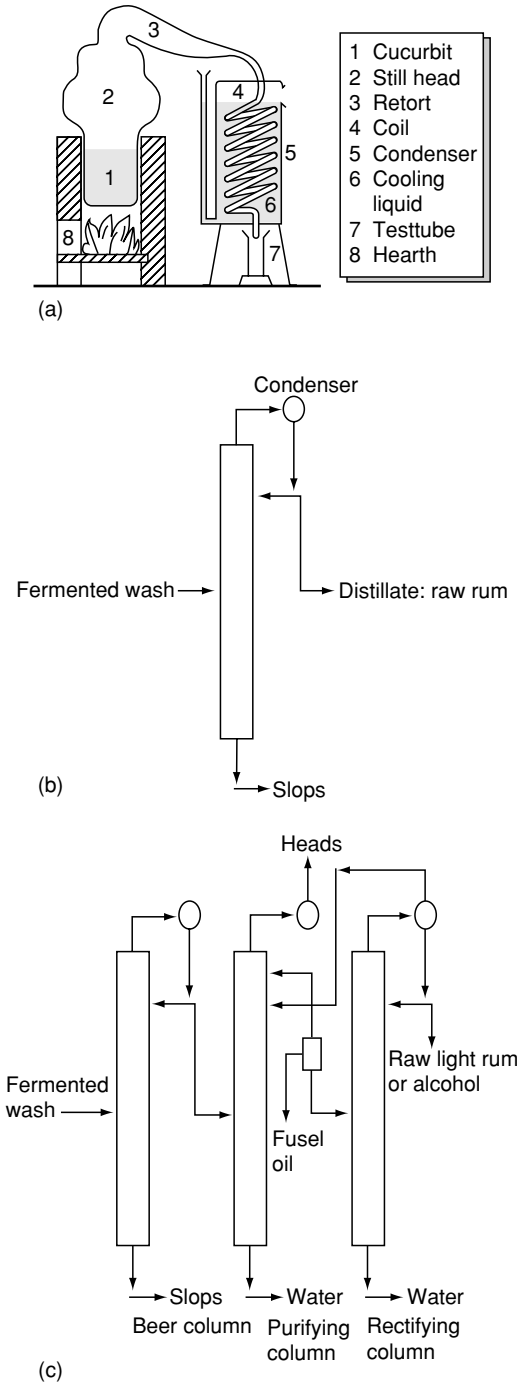

Figure 1 Rum distillation devices. (a) Intermittent pot still; (b) continuous single distillation column; (c) multistage column.

compounds such as minerals, organic acids, phenols, and others chemicals generated by the thermal treatment during sugar manufacturing.

Rum from sugarcane juice is a special feature of traditional production in the French West Indies.

Before the eighteenth century, the essential elements of rum production were:

- a mash comprising molasses and scum as the source of sugar, which was diluted with waste water (slops) and/or water in varying proportions (Table 2)
- in this mixture sugars were spontaneously fermented over a 7–15-day period by the abundant flora of *Schizosaccharomyces* yeast and various bacteria. Often, the molasses were introduced repeteadly, three times, to avoid fermentation failure
- the fermented washes were distilled in a pot still (Figure 1a) with double distillation. The kettle was manufactured of red copper, and a tap at the bottom was used to drain the spent liquor. The still head was also of red copper, sometimes tin-plated. The condenser was made from copper or pewter.

The rums obtained were between 41 and 45% ethanol content by volume. Developments in distillation engineering and microbiology were destined to improve these early processes dramatically.

During the eighteenth century, distillation equipment evolved, with the use of the retort, making it possible to obtain marketable spirit from a pot still without double distillation. In some devices the distillates were around 80% ethanol content. The continuous single distillation column (Figure 1b) was developed in the nineteenth century, in order to increase production. The distillates were also more consistent and the raw rums contained up to 90% ethanol – sometimes even 95%. The Creole, the Savalles, and the Barbet columns were typical of this period; the Barbet columns produced very aromatic distillates.

Multicolumn devices appeared during the nineteenth century (Figure 1c) and are used extensively today. The Coffey still was introduced for the distillation of grain whisky in Scotland and used largely in the British colonies. It produces distillates with 84–85% ethanol content. It initially comprised two columns: the analyzing column and the rectifying column, as used for grain whisky distillation. Today, multicolumn stills with three stages (beer, purifying, and rectifying columns) or even four columns are used for rum distillation.

The other great advance in rum production in the twentieth century concerned microbiology. In the early days, the spontaneous flora was used, and was often initiated with a natural starter by using a piece of wool soaked in washes and then carefully dried and kept to be introduced into the new fermentations. After Pasteur's studies of wine and beer fermentations, some researchers considered replacing the spontaneous or natural starter by pure fermentation using selected yeast strains. The aim was to improve

productivity and avoid bacterial contamination, since it was thought by some that the bacterial flora was detrimental to the flavor of the products. On the other hand some microbiologists, such as Allan and Ashby, believed that bacteria played an important part in producing the rum bouquet.

The application of pure fermentation involved modifications in fermentation management, in particular growth of starter cultures with selected yeast strains, and the use of antiseptics to limit bacterial presence and activity. The major change was fermentation by *Saccharomyces* yeast strains rather than *Schizosaccaromyces*. The result was an increase in yield but the quality of the products fell, through an increase in chemical neutrality. Roques, in 1927, was commissioned by the French Ministry of Agriculture to carry out a study which concluded that 'rums produced from pure and rapid fermentations are characterized by low levels of acid and ester, as well as relatively high contents in higher alcohols'. Most traditional rum producers consequently gave up the use of selected yeasts and decided that 'wild' fermentations gave the best results, producing rums with richer flavor.

At that time, which was the Prohibition period, light rum production was expanding in the greater West Indies and in the English-speaking lesser West Indies. Here rum was produced using modern plants and pure culture yeast. Under these conditions fermentations are rapid, the distillation is carried out with multistage devices, and the resulting distillates are low in aromatic compounds. These light rums formed the basis of the modern rum industry. However, traditional aromatic rums are still produced in the French West Indies and in Jamaica.

The distillery products are sold white (without aging), or after aging, generally in oak vats for many months.

Typology and Raw Rum Treatments

Rums, like many others alcoholic beverages, are often classified and labeled according to their geographical area or country of origin. However, this geographical classification does not necessarily correspond to the organoleptic properties of the spirit and, indeed, the US Federal Administration does not correlate the geographical designation with rum type.

From the analytical and sensory viewpoints, rums are most often described according to their aromatic intensity. This property largely relates to the microbial status of the fermentation and distillation equipment (Table 3).

Rum can be divided into four categories (Table 4): light, heavy, *de sucreride,* and *agricole*.

Table 3 Basic rum types and processing features

Raw material			
	Molasses		*Cane juice*
PF, MC Light	MF, PD, SC Heavy	MF, SC *de sucrerie*	MF, SC *Agricole*

PF, pure fermentation (yeast only);
MC, multistage column distillation;
MF, mixed fermentation (bacteria and yeast);
PD, pot still distillation;
SC, single-column distillation.

Table 4 Basic rum types: appellation, aspect, and post-distillation treatments

Not traditional	*Traditional*		
W, A, SC Light	W Heavy	W, A, SC *de sucrerie*	W, A, SC *Agricole*

W, white, a, aged; SC, straw-colored.

Light rum, of which Bacardi is the market leader, is the result of improvements in distillation engineering and fermentation flora control. Fermentation is rapid – 12–20 h duration – and uses cultured yeast with minimization of the bacterial flora by various means. In Cuba, rum with lightened features appeared around 1870. A gold medal was awarded to Don Facundo Bacardi, in 1876, at the international exhibition of Philadelphia for his light products. Similarly, a light rum or common was manufactured in Jamaica in 1882. In the French West Indies, traditional producers were not satisfied by the aromatic profile of the light rums and by the 1930s, they returned to traditional production processes. In modern light rum, the content of compounds other than ethanol (not alcohol: NA) is often under 60 g impurities per hectolitre (hl) of pure alcohol.

Traditional rums are a specialty of the French overseas territories: Guadeloupe, Martinique, and La Réunion. While part of this production is exported, the traditional products from Haiti and Brazil are solely consumed locally. The raw materials for traditional rums may be cane juice or molasses, giving respectively *agricole* rum and *de sucrerie* rum. The starting gravity is around 1035 for media from sugarcane juice, and 1060 for those from molasses. The production is on a cottage industry basis with small plants, because productivity is not a major goal. Length of fermentation is 20–48 h and features the presence and activity of a spontaneous bacterial flora that contributes to the aromatic properties of the product. The ethanol content of the final fermentation media is about between 3.5 and 8% by volume and the

marketed products have an aromatic content over 225 g Na hl^{-1} of pure alcohol.

Heavy rum or *grand arôme* (Wedderburn and Plummer in Jamaica) is molasses-based. The wash is diluted with prefermented distillery waste water (slops) instead of water. The wash containers are often wood-made. The fermentation is a true ecosystem, where only *Schizosaccharomyces* yeast strains are able to perform alcoholic fermentation; meanwhile there is important spontaneous bacterial activity. Length of fermentation is long: 7–15 days. Heavy rums, which are a peculiar form of traditional rums, are not used as a beverage, but as an ingredient of blends, and as an aromatic agent in confectionery and cooking. The content of aromatic compounds is very high compared to the other rums – generally between 800 and 1500 g NA hl^{-1} of pure alcohol. Rumverschnitt, or German rum, is a blend of potato alcohol with heavy rum, providing aromatic flavors.

Light rum and the traditional rums, heavy rum, *agricole*, and rum *de sucrerie*, are the four basic rum types. Depending on postdistillation treatments they will have various color and organoleptic properties (Table 4). There are three color categories that generally relate to the length of maturation; the longer the maturation period, the darker the color. However, in some raw rums coloration can be derived from added caramel and some aged rums may be treated to remove the color.

White rums are potable spirits sold without coloring. They are made in two ways:

- They can be obtained from raw rum which, after a short maturing period to eliminate the pungent highly volatile compounds, is diluted to the commercial strength.
- Alternatively, white rum is sometimes made from matured spirits that have been treated with activated charcoal, and mixed with sugarcane alcohol. The Cuban Havana Club Silver dry is one such product.

Straw-colored rums are matured for several months in huge oak casks, during which time they adopt a light coloration, resulting from exchanges with the cask walls. The organoleptic properties of these products are halfway between white rum and aged rum.

Old rums are obtained after maturation in previously used oak barrels, of 170–650 liters capacity, generally for a period of 3 years. During the aging process, many changes occur in the raw spirit, which becomes darker, by leaching compounds from the barrel to the spirit. There are also changes to the organoleptic properties of the spirit, for example, oxidation and hydrolysis reactions involving lignin from the barrels, producing aromatic aldehydes. The very old rums are appreciated in much the same way as other matured spirits like malt whisky and brandy.

The ethanol content of rums varies from one product to another. The rum produced in the English-speaking Caribbean islands is often 43% by volume, but in the French-speaking islands it is 50% by volume. The aged products are generally lower in alcohol content than the white ones – between 37 and 45% by volume. However, there are products with very high ethanol content, such as Jack Iron (79% by volume) in Carriacou, Strong Rum (79% by volume) in Grenada, and Puncheon (75% by volume) in Trinidad. In western markets there are many distributor brands. Those distributors often use raw rum from the West Indies to elaborate their products.

Rum Market

Rum was the world's leading spirit in the 1980s with an annual volume of 9 500 000 hl of pure alcohol, while the annual volume of whisky and vodka respectively was of 9 000 000 hl of pure alcohol. Nowadays, after vodka and whisky, rum is the third spirit consumed through the world, with more than 80 million 9-liter cases sold in 1998. The US market leads, with a sales volume of 12 million cases, amounting to $1.5 billion, of which half is sold by Bacardi, followed by Captain Morgan, Castillo, and Havana Club. Bacardi, with $1840 million, was second in the list of top 100 spirit brands world retail sales value in 1995 after Stolichnaya vodka, which was first with $2090 million sales. The second rum brand was Captain Morgan which was 67th in this Top 100, with $290 million. The world's six leading rum brands ranked by retail value are shown in Table 5.

With regard to consumption, USA has the highest rate, followed by the Philippines, Mexico, India, Germany, Colombia, the Czech and Slovak republics, Venezuela, Canada, and France.

Table 5 The world's top rums in 1995

Rank	*Brand*	*Brand owner*	*Sales*[a]
1	Bacardi	Bacardi Martini	18.76
2	Tanduey	Tanduay Distillery	14.00
3	Captain Morgan	The Seagram Co.	1.90
4	Gran Reserva	Santa Teresa	1.40
5	Castillo	Bacardi Martini	1.34
6	Cacique	The Seagram Co.	1.25

[a]In millions of 9-liter cases.

Conclusions

Rum is a product steeped in history, pictures, and flavor, associated with seasonings and condiments. It is exceptional among spirits because of its range of products, uses, and variety of applications, as a beverage and as a flavor ingredient in foods.

Nowadays knowing about dietetics reveals that food value is not only nutritional. On an individual basis, food is also a source of pleasure and symbols. The composition of rum and its appreciation are plenty of elements to satisfy consumers' expectations.

The modern consumer search for local products opens a perspective for traditional foodstuffs which are vehicles of aroma.

On all alcoholic beverage packaging, in France, it is a legal obligation to display: *A consommer avec modération* (i.e., to be taken in moderation). Now, can I propose a toast?

See also: **Alcohol**: Properties and Determination; **Sugar**: Sugarcane; **Whisky, Whiskey, and Bourbon**: Products and Manufacture; **Yeasts**

Further Reading

Arroyo R (1945) *Studies on Rum*. Research bulletin 5. Puerto Rico: Agricultural Experimentation Station, University of Puerto-Rico.

Fahrasmane L and Ganou-Parfait B (1997) *De la Canne au Rhum*. Paris: INRA Editions.

Guillaume J (1946) *Le Rhum, sa Fabrication et sa Chimie*. Annecy, France: Depollier.

Hamilton E (1995) *Rums of the Eastern Caribbean*. PO Box 771, Culebra, Puerto-Rico 00775: Tafia Publishing.

Huetz de Lemps (1997) *Histoire du Rhum*. Paris: Éditions Desjonqueres.

INRA (1975) *Symposium International sur le Rhum et Alcools Dérivés de la Canne à Sucre. Annales de Technologie Agricole* 24: 3–4, 3237–495.

Kervegant D (1946) *Rhums et Eaux-de-vie de Canne*. Vannes, France: Editions du Golf.

RYE

M G Lindhauer and J Dreisoerner, Federal Centre for Cereal, Potato and Lipid Research, Detmold, Germany

Background

Botanically, the rye (*Secale cereale* L.) belongs to the family of grasses (Gramineae). The rye plant is characterized by a culm, ranging in length from 65 to 200 cm, depending on the cultivar. The 5–20-cm-long four-cornered ear consists of single spikelets covered by lemmas with long awns. The blue-green grains are 5–9 mm in length and can be easily removed from the spike when ripe. For cultivation, winter rye varieties are preferred, as they are normally higher yielding than summer rye cultivars.

Rye is thought to have originated around 8000 BC in south-west Asia, probably in the area of the present states of Turkey, Iraq, and Iran. From that area, rye spread out for centuries, although the exact route is not known. About the eighth century BC, farmers in Central Europe considered rye as a weed in wheat and barley crops, as it often exhibited a better growth performance and yields, obviously due to its drought resistance and ability to grow on poor soils, than the primary cereals. Because it is often a weed in older cereals, rye (together with oats) is also called a 'secondary cereal'.

Nowadays, rye is a specialty of the northern latitudes between 50 and 65° degrees latitude. Its main cultivation area is in the north-west part of the eastern hemisphere, roughly between the Ural mountains in the east and the North Sea in the west. During the sixteenth and seventeenth centuries, rye was eventually brought by European settlers from Central Europe to North America and western South America, and during the nineteenth century, it was introduced into Brazil, Argentina, Uruguay, Australia, and South Africa.

Agronomic Significance

Rye is a cereal with modest requirements in terms of soil, fertilization, and climate. Particularly on poor sandy soils, as in northern and north-western parts of Poland and Germany, which have a limited nutrient and water availability, and because of its relatively good winter hardness, rye has an overall better performance than other cereals and should not be replaced in the crop-rotation system. But as a result of hybrid breeding rye can now compete in yield formation even on more fertile soils.

The greatest rye producer used to be the former Soviet Union, producing 11–13 million tonnes per year. Now, this has been split into two main parts: Russian Federation 6–8 million tonnes and Belarus 3 million tonnes. The second highest producing country is Poland, with 5–6 million tonnes, followed by Germany, with 4–5 million tonnes. The Nordic and Baltic countries also have a specific tradition in rye utilization and are relatively small producers; yearly production fluctuates with the weather conditions at the time of sowing and the overwintering conditions.

Yield performance shows a broad range depending on the agricultural system, especially on the availability of resources like fertilizer, pesticides and high-yielding varieties, and on climatic conditions. Statistical data for 1999 show that Germany had the highest yield, with 57.9 dt ha^{-1} (decitons per hectare), followed by Switzerland (53.7 dt ha^{-1}) and Scandinavia (Sweden and Norway: 47.9 dt ha^{-1}; Denmark: 46.5 dt ha^{-1}). In major rye-producing countries like Russia, the yields were 13.5 dt ha^{-1}, and in Poland 23.1 dt ha^{-1}. For comparison, the average yields in North America were 20.6 dt ha^{-1} in 1999.

The most important diseases in rye are snow mold (*Microdochium nivale*), leaf rust (*Puccinia recondita*), powdery mildew (*Erysiphe graminis*), foot rot (*Pseudocercosporella herpotrichoides, Fusarium* ssp.), head blight (*Fusarium graminearum, F. culmorum, Microdochium nivale*) and ergot (*Claviceps purpurea*). This latter fungus is typical for cultivated grasses prefers to grow on rye. Instead of a rye kernel it is up to 2.5 cm long outside black violet and inside grayish white sclerotium develops in a spikelet and as such it is clearly visible in an infected ear in the field. During harvest it gets into the bulk grain. The sclerotia of Claviceps purpurea are known as ergot; they contain specific poisonous alkaloids, so-called ergot alkaloids, which cause different symptoms, from dizziness to cramps, resulting in neuronal disorders in the case of chronic poisoning. Since the ninth century, ergot poisoning has been implicated in several epidemics in central Europe during the Middle Ages. Today, ergot in rye is no longer a problem due to the use of efficient cleaning and sieving systems in flour mills preventing ergot from contaminating rye flour. The EU has specified a threshold value for ergot in flour of 0.05%.

Rye Breeding

For farmers, the most important objective is a stable grain yield. In this respect, much progress has been made during the past 20 years in hybrid rye breeding. Hybrid rye breeding started in 1970 at the University of Hohenheim, Germany. At present, 17 hybrid cultivars are registered on the official descriptive list of plant varieties in Germany occupying almost 60% of the total rye acreage in this country. Hybrids are characterized not only by higher yields but also by a higher resistance against preharvest sprouting. This is a phenomenon specifically more pronounced in rye than in other cereals, meaning that under wet weather conditions, the grains already start to germinate on the culm. Preharvest sprouting causes high harvest losses, as the flour prepared from such preharvest sprouted grain loses its baking performance and is no longer suitable for breadmaking.

As well as breadmaking quality, progress has been made in improving other important traits such as kernel weight, lodging resistance, and resistance against diseases (e.g., leaf rust). Hybrid varieties are also licensed and released in other member countries of the EU as well as in Poland. Hybrid rye breeding programs are also in progress in Russia and Australia. By crossing rye and wheat, a new cereal species, *Triticale* (from Triti*cum aestivum* and *Se*cale *cereale*), has been generated, which has replaced rye in some production areas.

Rye and Nutrition

In the food culture of several European countries such as Russia, the Baltic and Scandinavian countries, Polish and German rye besides wheat has a long tradition as a raw material for breadmaking and for the production of a variety of other different products. Although the total production of rye has diminished in recent years, its use as a human food increased slightly during the 1990s (Table 1). In 1990, the food consumption of rye in the world was about 8 million tonnes, which was about 35% of the total production.

Milling Products

There are many similarities between the milling of rye and wheat. In both cases, the same type of machinery with only slight modifications in the procedure is used. By using different equipment and milling techniques, a large variety of rye products can be obtained (Table 2).

During the milling process, the first machine flours contain after sifting mainly starch flour granules only (i.e., they have a low flour extraction grade), while the following flours with a higher extraction grade contain increasing amounts of bran and other fibrous material. The extraction rate is strongly correlated with the ash (mineral) content. The higher the amount of bran and fiber in the final flour, the higher the ash content. Wholegrain rye flour has an ash

Table 1 Rye product consumption in the world in some selected countries

Area/country	*Rye consumption (kg per capita)*		
	1985	*1990*	*1995*
World	1.77	1.08	1.38
USA	0.2	0.4	0.3
Canada	0.5	0.5	1.6
China	0.7	0.5	0.2
USSR	8.4	3.8	
Former USSR			21.3
Poland	57.0	32.5	36.4
Germany	21.4	13.8	11.5
Finland	20.4	17.7	15.8
Denmark	20.3	18.4	15.5
Sweden	14.8	11.7	12.4
Norway	8.9	9.2	7.2

From *FAO Statistical Yearbooks*.

Table 2 Milling products

Whole grain rye
Steel cut rye
Crushed rye grains
Malted and crushed rye grains
Whole-grain flour
Sifted flour (variations in ash content)
Rye bran
Rye flakes
Breakfast cereals (muesli and others)
Sourdough rye bread mix

content of about 2 g per 100 g, whereas the most frequently used flours for breadmaking have ash contents of about 1 g per 100 g (with an extraction rate of about 80%).

Breadmaking

Unlike wheat, the gluten protein of rye is able to form a three-dimensional network as basic structure of the dough rye proteins are less able to swell. This is also valid for rye starch. However, rye contains a large amount of nonstarch polysaccharides of high water binding capacity, i.e., swelling capacity. They are collectively called pentosans. Conditioned by sour dough – today, dough acidification steps are often undertaken instead – certain rye proteins are also able to swell. The soluble proteins, starch, and the pentosans cause the gas-holding capacity of rye flour during dough preparation. The swelling ability can be measured ('swelling index') and provides information on the dough-rising and gas-holding capacity. Other specific methods such as techniques using falling numbers and amylograms are available to predict the dough-forming suitability of rye flours. Pentosans, especially, are responsible for the higher shelf-life and freshness of rye breads than with wheat breads.

The most typical rye bread made from whole grains or wholemeal is that prepared with sour dough. In this method the main ingredients, flour, water, and starter culture, are mixed and fermented for about 8–18 h. During the fermentation period, the lactic acid bacteria and the sourdough yeasts grow, and as a result of the microbial activity and enzymatic reactions of the microflora, flavor precursors are synthesized. The main components formed are lactic and acetic acid. After fermentation, more flour, water, and other ingredients are mixed into the sourdough to make the dough. The dough is then left to rise for a short period, after which the breads are shaped, left to rise again, and baked.

Bread Types

The use of rye is mainly based on local traditional nutritional practices. Consequently, in those countries with significant rye production and consumption, a wide variety of different types of rye bread have been established. The flavor and structure of rye bread are quite different from those of wheat bread. They vary depending on the flour type, ingredients, dough preparation (type and amount of sourdough), baking conditions, and time, as well as the size and type of bread.

Traditional rye bread is the dark sour bread, popular in Finland, the Baltic states, Russia, Belarus, and Poland. This tradition has changed somewhat on its way to the West, for instance to Sweden, Denmark, or Germany. In Germany, the total consumption of rye bread is steadily decreasing, especially that of wholemeal rye bread. Mixtures of different amounts of wheat and rye flour are preferred.

In addition to the typical rye bread crispbreads, thin crispbreads, rye rolls and buns, rolls, buns, and breads containing wheat/rye mixtures and parbaked rye products enrich the variety of rye products, but in different amounts in individual countries.

The most typical rye bread in Finland, Denmark, Russia, and the Baltic countries is wholegrain bread made by the aid of sourdough. Crispbread originated from Sweden, and nowadays is eaten throughout the world but with a greater preference in the Nordic countries. There are three different types of rye crispbread: normal yeast-fermented, sourdough-fermented, and cold crispbread, the third being baked without yeast. The dough has the right texture from a foaming process in which air is incorporated into the cooled dough, which also leads to the almost white color of the finished bread. Crispbread has a long shelf-life owing to its very low water content (5–7%).

Other Rye Products

Different types of rye groats and flakes are available for baking, porridges, and breakfast cereals. These products are generally produced from whole-grain rye. In the Nordic countries, several new types of rye products have been developed, among which breakfast cereals have gained a certain popularity. Breakfast cereals with up to 55% rye are available; the rye of these products is flaked and precooked and sometimes even extruded to increase the crispiness and improve the taste. Table 3 lists the different types of rye products.

In the light of modern knowledge on the nutritional value of rye, much effort has been made in different rye-producing countries to invent attractive rye-based dishes to reverse decreasing consumption.

Chemical Composition

Table 4 compares the most frequent constituents in whole grains of rye, wheat, and oats. The outer layer of the endosperm, the aleurone layer, is rich in proteins, minerals, and vitamins, especially B-vitamins (cf. Table 5). Fat is mostly located in the germ, and starch in the central endosperm. Rye is essentially a good source of minerals, such as potassium, magnesium, zinc, manganese, and fluoride.

From Tables 4 and 5, it is clear that rye wholegrain or wholemeal is also an important source of dietary fiber. The dietary fiber content of typical rye bread is about three times higher than that of white wheat bread. Even at the same flour extraction rate, rye flour contains higher amounts of dietary fiber than the respective wheat flour.

Furthermore, rye grains contain significant amounts of lignans: 47 mg of secoisolari ciresinol (SECO) per 100 g and 65 mg of matairesinol (MAT) per 100 g, so-called phytoestrogens. Also, other phenolic compounds are present in rye, e.g., tannin and phytic acid. All constituents are considered bioactive compounds.

Table 3 Other rye products

Rye porridge
Breakfast products
Pastries
Rye pasta
Rice–rye mixture
Snack products
Crispbread sandwich
Rye hamburger
'Maemmi' – a Finnish Easter pudding
'Kalakukko' – a rye-dough-covered baked dish containing ingredients like fish, meat, and/or vegetables

Health Benefits

Nutritionists worldwide recommended increased consumption of wholegrain products as a source of dietary fiber. Rye is an excellent raw material for healthy and tasty food with a high fiber content (Table 4). In cereals, fiber is mainly located in the outer cell layers of the kernels, especially in the bran. Wheat and rye have similar bran contents, but rye contains more cell wall within the endosperm and thus has a higher fiber content.

There are two main types of dietary fiber: soluble and insoluble. Water-insoluble dietary fiber includes,

Table 4 Main constituents of whole grain meal of rye, wheat, and oats

Component	*Percentage dry matter*		
	Rye	*Wheat*	*Oat groats*
Protein	10–15	12–14	13–16
Fat	2–3	3	6–7
Starch	55–65	67–70	54–64
Ash	2	2	2
Total dietary fiber	15–17	10–13	11–13
Soluble fraction of dietary fiber	3–4	1–2	3–5

From Andersson R, Westerlund E, Tilly Z-C and Aman P (1993) Natural variations in the chemical compositions of white flour. *Journal of Cereal Science* 17: 183–189; Clydesdale FM (1994) Optimizing the diet with whole grains. *Critical Reviews in Food Science and Nutrition* 34: 453–471; Hörkönen H, Pessa E, Suortti T and Poutanen KL (1997) Distribution and some properties of cell wall polysaccharides in rye milling fractions. *Journal of Cereal Science* 26: 95–104; Lasztity R (1998) Oat grain. A wonderful reservoir of natural nutrients and biologically active substances. *Food Review International* 14: 99–119; Vollendorff NW and Marlett JA (1991) Dietary fibre methodology and composition of oat groats, bran and hulls. *Cereal Foods World* 36: 565–570; Welch R (ed.) (1995) *The Oat Crop*. London: Chapman & Hall.

Table 5 Mineral and vitamin content of rye bread and the percentage of recommended daily intake (RDI) that 100 g (three slices) of wholegrain rye bread provides

Nutrient	*Amount in 100 g of rye bread*	*Percentage of RDI (for women)*	*Percentage of RDI (for men)*
Vitamin E	1.0 mg of tocopherol	14	10
Thiamin	0.18 mg	16	13
Riboflavin	0.2 mg	15	13
Niacin	1.2 mg	8	6
Folate	43 μg	14	14
Iron	2.7 mg	18	27
Zinc	2.3 mg	33	26
Selenium	3.2 μg	8	6
Calcium	31 mg	4	4
Potassium	400 mg	13	11
Magnesium	75 mg	27	21
Fiber	9.9 g	40	40

From Kujala T (2000) *Rye: Nutrition, Health and Functionality*. Finland: Rakennuspaino Oy.

among others, cellulose and lignin, whereas arabinoxylan and β-glucan are partly insoluble, partly soluble. Arabinoxylans, often subsumed under the term pentosans, are the main fiber component of rye.

In the classical understanding of dietary fiber, it is common knowledge that a recommended intake of 25–35 g of fiber per day reduces the risk of overweight and constipation. Dietary fiber increases the fecal volume and reduces the intestinal transit time.

Research results in recent years have provided good evidence that dietary fiber, especially soluble fiber, contributes to a number of health benefits, such as a reduced risk of heart disease, decreased serum cholesterol level, reduced risk of elevated blood glucose levels and diabetes, strengthening of the immune system, and protection against different kinds of cancer (e.g., colorectal cancer, breast cancer, etc.).

Besides dietary fiber, rye grain contains relatively high concentrations of lignans, which are converted to weakly estrogenic mammalian lignans in the colon. Such phytoestrogens affect sex hormone metabolism and function by influencing certain sex hormone metabolizing enzymes at the cellular level. In addition, they have a great number of other interesting biological activities, possibly making them anticarcinogenic compounds.

Animal Feed and Alcohol Production

Depending on the total rye production and consumption of rye as food in an individual country, varying amounts of rye between 30 and 80% are used for animal feed. There are, however, restrictions on the use of rye in animal feed.

Those components with beneficial effects on human health, the pentosans, create some problems in the digestion of monogastric animals such as pigs and poultry–young animals are much more affected than older animals, with symptoms of maldigestion, reduced growth, and weight increase. Therefore, much effort has been made to improve rye-based feed by adding specific enzymes and to find uses for feed to alluviate the apparent overproduction of rye in the EU and the rest of the world. One such use is the use of rye and rye products for feeding fish and crustaceans in specialized farms in Asia and elsewhere.

There is also a tradition of producing alcohol from rye. Famous brand names of such spirits exist in nearly all of the rye-growing countries, but the total amount of rye used for these is comparably small.

See also: **Alcohol**: Properties and Determination; **Bread**: Breadmaking Processes; **Cereals**: Contribution to the Diet; Dietary Importance; **Milling**: Characteristics of Milled Products

Further Reading

Andersson R, Westerlund E, Tilly A-C and Aman P (1993) Natural variations in the chemical compositions of white flour. *Journal of Cereal Science* 17: 183–189.

Bushuk W (2001) *Rye: Production, Chemistry and Technology*, 2nd edn. St. Paul, MN: AACC.

Clydesdale FM (1994) Optimizing the diet with whole grains. *Critical Reviews in Food Science and Nutrition* 34: 453–471.

FAO (2001) *Production Yearbook*, vol. 53, 1999. Rome: Food and Agriculture Organization.

Hörkönen H, Pessa E, Suortti T and Poutanen KL (1997) Distribution and some properties of cell wall polysaccharides in rye milling fractions. *Journal of Cereal Science* 26: 95–104.

Kujala T (2000) *Rye: Nutrition, Health and Functionality.* Finland: Rakennuspaino Oy.

Lasztity R (1998) Oat grain. A wonderful reservoir of natural nutrients and biologically active substances. *Food Review International* 14: 99–119.

Miedaner T (1997) *Roggen: Vom Unkraut zur Volksnahrung*. Frankfurt: DLG-Verlag.

Nyman M, Siljeström M, Pedersen B *et al.* (1984) Dietary fibre content and composition in six cereals at different extraction rates. *Cereal Chemistry* 61: 14–19.

Pyler EJ (1998) *Baking Science and Technology*, 3rd edn. Merriam, IN: Sosland.

Stear CA (1990) *Handbook of Breadmaking Technology.* Barking, UK: Elsevier Science.

Vollendorff NW and Marlett JA (1991) Dietary fibre methodology and composition of oat groats, bran and hulls. *Cereal Foods World* 36: 565–570.

Weipert D (1997) Processing performance of rye as compared to wheat. *Cereal Foods World* 42: 706–712.

Welch R (ed.) (1995) *The Oat Crop*. London: Chapman & Hall.

S

SACCHARIN

J F Lawrence, Health and Welfare Canada, Ontario, Canada

Introduction

Saccharin has been in use as an artificial sweetener for about 100 years. It was discovered accidentally by researchers at the Johns Hopkins University in 1879 during studies on the oxidation of *o*-toluene sulphonamide. Its structure is shown in Figure 1. It is not metabolized but is excreted unchanged by the human body and thus contributes no calories to the diet. The safety of the substance has been the subject of debate for many years, and the controversy continues with arguments both for and against its use as an artificial sweetener.

Sweetness

Saccharin is about 300–500 times as sweet as sucrose, depending upon its concentration and the type of food medium in which it is used. Compared with sucrose, it has a slow onset of sweetness that increases to a maximum and then persists. The major drawback of saccharin is its bitter metallic aftertaste, which is particularly evident at higher concentrations. Because of this, efforts have been made to mask the aftertaste by adding substances such as cream of tartar, lemon flavor, pectin, ribonucleotides, glycine, gentian root, and artificial sweeteners such as aspartame or cyclamate. Combinations of saccharin with aspartame or cyclamate have proved to be very successful for many applications. In these cases, it has been found that saccharin exerts a substantial synergistic effect on the sweetness of either aspartame or cyclamate. For example, a mixture of 20 mg of aspartame and 4 mg of saccharin is equal in sweetness to either 45 mg of aspartame or 35 mg of saccharin if used alone in a cup of coffee. Not only does saccharin exhibit a synergistic effect with aspartame, but it apparently improves the stability of the latter in acidic soft drinks, enabling extended storage of such products. Saccharin/cyclamate mixtures were widely accepted in the USA until the ban on cyclamate in 1969. At present, saccharin/aspartame concentrations are commonly used in the USA in diet soft drinks. In some European countries, saccharin/cyclamate combinations are still permitted.

Production, Physical, and Chemical Properties

Commercial production of saccharin has been carried out by two different chemical processes. The Remsen–Fahlberg method involves reaction of toluene with chlorosulfonic acid to produce *o*- and *p*-toluene sulfonyl chlorides. The *o*-isomer is separated and then treated with ammonia to form *o*-toluene sulfonamide. The last compound is oxidized to *o*-sulfamoyl benzoic acid, which, upon heating, cyclizes to saccharin.

The Maumee process involves diazotization of methyl anthranylate with sodium nitrite in the presence of hydrochloric acid to form 2-carbomethoxybenzenesulfonyl chloride, which is converted to saccharin after amidation and acidification. The calcium or sodium salts of saccharin are produced by reaction of the respective hydroxides with saccharin.

Saccharin and its calcium or sodium salts are white crystalline powders with intensely sweet tastes. A dilute aqueous solution of saccharin is about 500 times sweeter than a solution with an equal concentration of sucrose, whereas the salts are about 300 times sweeter than a solution containing an equivalent weight of sucrose. The melting point of saccharin is 228.8–229.7 °C, whereas the salts decompose above 300 °C. The excellent stability of saccharin and its salts under normal food-processing makes it ideally suited technically as a noncaloric sweetener in many different products. It is unaffected, for example, by the heat processing required for jams, jellies, and canned fruits. The salts of saccharin are much more water soluble than saccharin (1 g of salts per 1.5 ml of water versus 1 g of saccharin

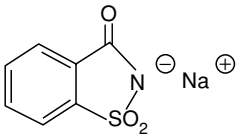

Figure 1 Structure of saccharin (sodium salt). Reproduced from Saccharin, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

per 290 ml of water). Aqueous solutions of sac charin are slightly acidic, whereas the calcium and sodium salts are neutral or slightly basic. They are very stable in aqueous solution over a wide pH range, although below pH 2.5 at elevated temperatures saccharin has a tendency to hydrolyze slowly to *o*-sulfamoyl benzoic acid and *o*-sulfobenzoic acid.

Food Uses

Because of its stability, saccharin (usually as its calcium or sodium salt) has been used in a wide variety of food products. These include carbonated and noncarbonated soft drink beverages, table-top sweeteners, dry beverage bases, canned fruits, gelatin desserts, cooked and instant puddings, salad dressings, jams, jellies, preserves, chewing gum, and baked goods. Typical amounts of sodium saccharin used in food products are, for example, 2.7 mg per ml of table-top sweetner, 0.32 mg per ml of carbonated soft drinks, 0.19 mg per ml of noncarbonated soft drinks, 0.9 mg per ml of jams and jellies and 2.2 mg per stick of chewing gum.

Nonfood uses of saccharin include applications in cosmetics such as toothpastes, mouthwashes, and lipstick, pharmaceuticals, such as coatings on pills, cattle feed, electroplating, and, in Japan, as an intermediate in the production of the rice blast fungicide, probenazole.

World commercial production of saccharin in 1900 was estimated to be about 190 tonnes. This increased more or less steadily up to the late 1970s, when world annual production reached approximately 5000 tonnes.

Metabolism and Safety

Many feeding studies with a variety of animal species have shown that saccharin is excreted unchanged. In rats, saccharin is excreted rapidly. However, there is some accumulation in the bladder, but the sweetener is completly cleared in 3 days after removal of saccharin from the diet. No metabolites have ever been detected. Studies have shown that the compound is not metabolized by liver microsomal enzymes, it is not genotoxic, and it does not bind to DNA after oral administration. The metabolic profiles of saccharin fed to dogs, rabbits, guinea-pigs, and hamsters are similar.

The greatest concern relating to metabolic effects of saccharin has been the uncertainty about its carcinogenicity to humans. Several early carcinogenicity studies in rats that included *in utero* exposure had indicated that feeding of saccharin may lead to bladder tumors. However, subsequent studies seem to indicate that it is not saccharin itself that causes the effects, but some other still unclear mechanism that operates in a dose-dependent fashion. Also, recent (1998) studies in monkeys have shown that saccharin is without carcinogenic effect on the primate urinary tract. A number of extensive epidemiological studies in humans have shown no association between saccharin consumption and urinary bladder cancer.

Regulatory Status

In 1972, because of the uncertainty of the safety of saccharin, the US Food and Drug Administration delisted saccharin and in 1977 proposed to ban its use in foods and beverages. However, public protest led congress to impose a moratorium on the ban, which has been extended up to the present. In Canada, saccharin was banned in 1978 for use in foods. However, it is permitted to be sold in pharmacies as a table-top sweetener. It is currently under review for other applications. In most other countries throughout the world, saccharin is permitted, although its use is restricted to varying degrees.

Analysis

A number of methods for the determination of saccharin in foods and pharmaceutical preparations have been reported. These include gas chromatography with flame ionization or electron capture detection, thin-layer chromatography, ultraviolet absorption spectrophotometry, and high-performance liquid chromatography (HPLC) with ultraviolet absorbance detection. The last technique is the most commonly used method at present, especially for the determination of saccharin in foods and beverages. For liquid samples, minimal sample preparation is required. The samples are filtered and perhaps diluted before being analyzed directly by HPLC. Solid foods are usually mixed with water, which extracts the saccharin. An aliquot of the aqueous extract is then directly analyzed by

HPLC. The sensitivity of the technique is more than adequate for the concentrations of saccharin normally encountered in foods. Saccharin has been incorporated into a multisweetener HPLC method.

See also: **Aspartame**; **Chromatography**: High-performance Liquid Chromatography; **Sucrose**: Properties and Determination; Dietary Importance

Further Reading

Franta R and Beck B (1987) Alternatives to cane and beet sugar. *Food Technology* 40: 116–128.

International Agency for Research on Cancer (1990) *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans. Some Non-nutritive Sweetening Agents*, vol. 22. Lyon: IARC.

Kretchmer N and Hollenbeck CB (1991) *Sugars and Sweeteners*. Boca Raton, FL: CRC Press.

Lawrence JF and Charbonneau CF (1998) Determination of seven artificial sweeteners in diet food for preparations by reverse-phase liquid chromatography with absorbance detection. *Journal of the Association of Official Analytical Chemists* 71: 934–937.

O'Brien-Nabors L and Gelardi RC (1991) *Alternative Sweeteners*. New York: Marcer Dekker.

Saccharose *See* **Sucrose**: Properties and Determination; Dietary Importance

Safety of Food *See* **Cleaning Procedures in the Factory**: Types of Detergent; Types of Disinfectant; Overall Approach; Modern Systems; **Food Poisoning**: Classification; Tracing Origins and Testing; Statistics; Economic Implications; **Hazard Analysis Critical Control Point**; **Laboratory Management**: Chemical Safety; Microbiological Safety; **Emerging Foodborne Enteric Pathogens**

SAGO PALM

J J Lal, Centre for Development of Science and Technology, Thrissur, Kerala, India

Introduction

The term sago means an edible starch extracted from the pith-like center of several Asian palms (including *Metroxylon sagu*) or sometimes of cycads. Palm is the common name for members of the plant family Palmae. Palmae is a large family and includes tropical trees, shrubs, and vines. Most members of this family are tree-like, characterized by a crown of compound leaves, and terminating in a tall, woody, unbranched stem. Natives adopt different techniques to separate sago from the stem.

Cycas revoluta

Cycas revoluta, popularly known as sago palm, is not a real palm but a cycad. It is a slow-growing, evergreen, long-lived, medium-sized palm-like plant. It is the most popular species in the genus *Cycas*. This plant is popularly known as King Sago. It is one of the cheapest and most readily available sources of food starch. It is a native of China and Japan and is widely grown as an ornamental plant in India. In combination with protein from sources such as fish and nuts, and vitamins from wild fruits, it forms a diet that has nourished communities in underdeveloped regions for many centuries. Even today it is a major energy food for many poor populations and is also a valuable famine food in these regions. The Japanese show special affinity for this plant and it plays an

important role in Japanese culture. The genus also occurs in the Malaysian region and South-east Asia, extending to Micronesia, Polynesia, Madagascar, and East Africa, and constitutes one of the oldest living plants on the earth. It plays an important role in the life of the locals as a source of cheap food in the form of sago and also plays a negative role by producing some toxins. These toxins are very strong and found in most parts of the plant. Consumption of the parts of the plant or even the processed flour from the plant has been implicated in a number of disease conditions.

Morphological and Anatomical Features

This attractive, ornamental tree is about 2 m in height and has branched stems with multiple heads. The plants grow in xerophytic areas like exposed slopes and other sunny places where water is scarce. It is hardy to $-9\,^{\circ}C$ and can also withstand high temperatures. It has a tuberous stem at a young age and rough, thick columnar stem in the adult stage and it is covered with an armour of thick, persistent leaf bases and at the apex with a crown of large compound leaves in apparent whorls. Once or twice a year, they produce a complete circle of new leaves which emerge all at once. The main rachis of the young leaves is in-curled and the leaflets are in-rolled, as in ferns. The general appearance of the plant is similar to that of a small palm and it is popularly known as cocopalm by the natives.

Internally the stem consists of an outer protective layer: the epidermis, large parenchymatous cortex, and large pith-containing thin-walled cells densely filled with starch or sago grains, numerous mucilage canals, and vascular bundles. Sago is the edible starch that can be extracted from the stem of cycads.

There are two types of leaves: scale leaves and foliage leaves. They are arranged in close spiral succession, alternating with each other. The scale leaves are brown, persistent, more numerous than foliage leaves, and play a protective role. The foliage leaves are large, 60 cm long, compound, and pinnately divided into more than 100 leaflets with a revolving margin, thus giving the plant its species name, *Cycas revoluta*. The midrib does not have veins and projects beyond the apex, ending in a spine. Petioles are thick and quadrangular in shape.

There is a taproot system which produces branches and some of the lateral roots become apogeotropic, growing vertically upward on the surface of the ground, where they branch repeatedly and form dichotomously branched coral-like masses, coralloid roots or corallorhiza that are inhabited by blue-green algae, *Anabaena cycadacearum*. This cyanophycean algae helps the plant to fix atmospheric nitrogen and increases the health and fertility of the soil. The surface of the coralloid roots is beset with lenticel-like apertures, suggesting their respiratory function.

Propagation

Sago palm is extensively propagated by vegetative methods by means of adventitious buds or bulbils, which are formed in the basal part of the stem as well as in the main trunk in mature plants. The plant is asexual (sporophyte) and dioecious in nature, i.e., the male and female structures, the cones, develop on separate plants. The male cones are borne singly and terminally on the main stem. Further apical growth of the stem is continued by the development of an axillary bud at the base of the cone and it becomes the new stem apex and therefore the stem of the male plant is sympodium. The male cone consists of a central woody axis bearing microsporophylls in a close and compact spiral. Each microsporophyl bears a large number of microsporangia, bearing microspores. There is no true, compact, and properly organized female cone.

The male gametophyte phase of the life cycle begins in the microsporangium and the pollen grain is considered to be an immature male gametophyte. The male gamete (sperm) develops inside the pollen grains. The large, top-shaped sperm ranges in size between 180 and 210 μm and it is visible to the naked eye. The megasporophylls arise spirally in acropetal succession and are loosely arranged on the stem like the ordinary crown of foliage leaves. The apical growing point of the female plant thus grows without any interruption and the stem in the female plant is monopodial. The megasporophylls are large, ranging from 15 to 20 cm in length. Each megasporophyl is divided into an upper broad leafy portion and a lower stalk-like portion and the ovules are borne laterally on the lower stalk-like portion: their number varies from two to 10. The female gametophyte develops inside the ovule.

The plant is generally believed to be wind-pollinated. After fertilization, the seeds develop over the summer and are ready to be removed in January or February. Mature seed is differentiated into an outer orange-red-colored fleshy layer, stony middle layer, and fleshy inner layer. Within these seed coats, there is a fleshy female prothalus, the endosperm, which functions as the food storage region of the seed. Seed will usually germinate in 3–6 months, but may require more than 3 years of growth to reach a small bulb-like structure.

Toxicity

Sago palm is known to be poisonous and sago separation includes careful processes to remove these toxins, before they are edible. Intake of sago before proper processing to remove toxins can cause vomiting, liver damage, and even death. Recent evidence indicates that the toxins have neurotoxic effects and now it is considered to be a slow poison.

It is remarkable that many indigenous people in different parts of the world have quite independently adopted different processing techniques. Fleshy seeds, leaves, unprocessed flour from stem pith, and pollen grains are considered to be toxic to livestock and to the population who live in proximity to these plants, and this has created interest in the pharmaceutical industry. The general symptoms of *Cycas* toxicity include diarrhea, headache, dizziness, vomiting, and mortality. The major toxic principle in sago palm poisoning is a glycoside named cycasin and other toxic glycosides include macrozamin, neocycasin, and an amino acid β-methylamino-L-alanine (BMAA). The active component of the glycosides is methylazoxyglycoside (MAM), which is formed by the cleavage of cycasin (glycosides) with the help of an enzyme, found in microbes in the gut. MAM is known to be hepatotoxic and carcinogenic. It causes apoptosis (programmed cell death) in the brain, thus reflecting a link between cycad ingestion and development of neurodegenerative disorders. BMAA is a nonprotein amino acid and its chemical structure shows similarities with beta-*N*-oxalylamino-L-alanine (BOAA), the neurotoxin from *Lathyrus sativus*. There are reports of incidence of amyotrophic lateral sclerosis (ALS) and parkinsonism dementia (PD) after *Cycas* ingestion. Cycasin and its aglycone MAM impair both rodent and human β-cell function, which may lead to the death of pancreatic islet cells, a very high prevalence of glucose intolerance, and diabetes mellitus.

Utilization of Sago Palm

Stem

Stem starch has been widely used as a food source. Pith and cortical cells contain a large amount of starch and yield sago grains and hence the plant is known as the sago palm. It has been estimated that a mature stem (i.e., beginning to flower) of about 1 m yields about 2 kg of sago. *Cycas* starch is a constituent of poor people's food and is used in times of food scarcity.

To separate the sago, the trunk of the plant is split and its pith-like center is chopped and ground into powder and the starch is removed. Before sago separation there is repeated washing, settling, and discarding of the water to eliminate the toxins. Ingestion of sago before proper washing leads to vomiting and dizziness. In Japan, the seeds and stems are often used to prepare wine. In Africa, to prepare bread from cycads, the central pith of the stem is taken in animal skin and buried in the ground for about 40–42 days until it is partially fermented. This fermented stem is ground up with water and is shaped into bread and baked. Crushed bark and seeds or megasporophyls, when mixed with coconut oil, are used in South. India as a poultice for sores and wounds. Since it contains a toxic glycoside, its constant use is said to cause intestinal disorders and it is considered to be genotoxic. Fiber obtained from the stems is used to make cloth and rope.

Leaves

The strong and leathery leaves keep their fresh green color long after being cut off from the plant and hence they are used as a nursery stock for funeral wreaths and church festivals under the name palm branches. The leaves are also used to decorate marriages and festivals. Leaves are used to make brooms and baskets, and to thatch huts. Leaf is also a source of fiber in the manufacture of cloth, ropes, and twines. Young fleshy leaves are used as vegetables. Consumption of leaves without proper detoxification is harmful and prolonged ingestion of leaves is toxic.

Roots

The roots are a source of sago and in Africa it is believed that burying the roots by a house could protect from lightning.

Seeds

Mature palms produce seeds in the female cones. A large plant may yield over 1000 seeds. Mature seeds are large and orange-red in color. The seeds attract animals, birds, rodents, and fruit-eating bats, which help seed dispersal. Intake of these seeds by domestic animals is toxic and leads to vomiting, dizziness, salivation, paralysis, and even death, but in wild animals, it is found to be nontoxic. To separate the sago, mature seeds are split open and the kernel is collected and sun-dried for 1 or 2 days. After thorough drying, the toxins are eliminated from the seeds by immersing it in water for 3–6 h: the water is stirred well and the foamy water is drained off. This washing process is repeated at least seven times for the total removal of toxins. Thereafter it is dried in sunlight and ground to prepare flour. Sago flour is pure white

in color and is similar to that of wheat or rice flour. Natives prepare bread, steam cakes, several indigenous dishes/sweets, and even wines. Preparations from the sago palm flour are considered to be an effective medicine to cure hemorrhoids (piles). It is also observed that pregnant women demand sago palm seed flour preparations as part of their pregnancy cravings, particularly if they were in the habit of consuming sago during their childhood days. Young seeds on boiling yield a decoction which is often used as a purgative.

Pollen Grains

Pollen grains formed inside the microsporophyls are considered to have medicinal properties. In India 'quacks' claim it as a folk medicine which has several health benefits, obtained from 'kailasam' (the place of Lord Shiva). It has been reported that pollen is heavily laden with cycasin and BMAA and is toxic.

Metroxylon Sagu

Metroxylon sagu is a true palm belonging to the order Arecales, family Palmae, and subfamily Calamoideae. It is commonly grown in wild swampy areas of Malaysia, Indonesia, and New Guinea. This palm grows between 10 °N and 10 °S latitudes and up to an elevation ranging from 700 to 1000 m. It is also found growing in dry lands. In certain regions of the world, this palm is a provider of staple food. Its tremendous utility favored its distribution from its place of origin to far-off distant places, including South-east Asia. It attains a natural height of 15 m. The stem is straight, stout, and without branches and has a unique crown of large pinnate leaves. The roots are unbranched and go straight down to the soil. Sago palm grows quickly and it dominates in tropical peat swamp rainforests, where other crop plants do not thrive. Sago palm is a kind of semiwild species but not a well-established crop plant. In the wild and in cultivation, sago palm is mostly propagated vegetatively through its suckers.

Sago palm is tolerant of highly acidic conditions (low pH) coupled with high concentration of metals in the soil like aluminum, iron, and manganese, which inhibit the growth of other plant species. It has the ability to grow abundantly in heavy clay soils, which inhibits the growth of other plant species. In wild conditions, a cluster of sago palms can be seen. Each sago cluster consists of several mature plants and suckers. People used to harvest the mature trunk, leaving the immature ones and suckers. Thus, there is a possibility for the continuous harvesting of mature palms at intervals, thereby sustaining sago production.

Individual suckers at first grow into a rosette of leaves. Then it produces a stout trunk. Sago palm reaches sexual maturity at about 12–15 years. It is a monocarpic plant – it flowers only once in its life and then dies. Insects help to pollinate the flowers. The fruits take about 1½–2 years to mature. Just before the onset of flowering, the palm is harvested to obtain maximum yield. During this stage, the palm converts its stored nutrients into starch, which accumulates in the trunk.

Edible starch is extracted from the pith-like center of sago palm when the tree is about 12 years old. During harvesting, the entire trunk is cut into pieces. The marrow of the stem is then chopped out as finely as possible and its starch is then washed out and separated from the cellulose. Starch separated from the stem is dried to yield sago meal.

Productivity of sago palm is four times that of paddy (*Oryza sativa*). A single sago palm yields about 150–300 kg sago. This starch is an important item in the diet in some parts of Eastern Asia. It is used in various food items and also to stiffen cloth material in the textile industry. Desiccated products made from sago starch can be stored for exceptionally long periods. Sago starch has a multitude of uses. Sago is widely used to produce sago pearls. Sago pearls can be boiled, either alone or mixed with other foods, and consumed directly as a carbohydrate source. Sago is also widely used, together with rice, corn, and potatoes, in the manufacture of noodles in Malaysia. It is also used for the production of monosodium glutamate, soft drinks, and various syrups. The other uses of sago include the manufacture of biodegradable plastics, ethanol, and citric acid. Sago meal is very digestible and hence it can be used to feed livestock, pigs, and even poultry.

In addition to the production of starch from the stem, the large and strong leaves are useful. Leaves are best for roof thatching, and for making bags, baskets, and cages. Leaves contain long fibers (1.69–3.59 mm) which have a high cellulose content. Hence, it is also useful for papermaking.

The sago beetle is one pest which attacks the sago palm. It bores into the trunks of the palms and lays its eggs inside. But the natives consume the larvae of these sago beetles as a highly esteemed delicacy. They consume it either raw or in cooked form. Thus, the pest attack has proved a boon to the local population. Unfortunately, however, apes and wild boars tend to damage palms at the young stage by digging up and eating young palms during the sucker stage.

See also: **Alcohol**: Properties and Determination; **Amino Acids**: Properties and Occurrence; **Bread**: Breadmaking

Processes; **Flour**: Analysis of Wheat Flours; **Liver**: Structure and Function; **Protein**: Food Sources; **Starch**: Sources and Processing; **Toxins in Food – Naturally Occurring**; **Wines**: Production of Table Wines

Further Reading

Brandis D (1984) *Indian Trees*. London: Archibals Constable.

Chamberlain CJ (1919) *The Living Cycads*. New York: Hafner (Reprinted 1969).

Durlach J, Bac P, Durlach A, Bara M and Guiet-Bara A (1997) Are age related neurodegenerative diseases linked with various types of magnesium depletion? *Magnes Research* 10(4): 339–353.

Eizirik DL and Kisby GE (1995) Cycad toxin-induced damage of rodent and human pancreatic beta-cells. *Biochemistry and Pharmacology* 31: 355–365.

Giddy C (1984) *Cycads of South Africa*, 2nd edn. Cape Town: C. Struik.

Gobe GC (1994) Apoptosis in brain and gut tissue of mice fed a seed preparation of the cycad *Lepidozamia peroffskyana*. *Biochemistry and Biophysical Research Community* 30: 327–333.

Macmillan HF (1994) *Tropical Planting and Gardening*. Jodhpur, India: Scientific Publishers.

Pan M, Mabry TJ, Beagle JM and Mamiya BM (1997) Nonprotein amino acids from *Cycas revoluta*. *Phytochemistry* 45(3): 517–519.

Parker SP (1992) *Encyclopedia of Science and Technology*, 7th edn. New York: McGraw-Hill.

Watt JM and Breyer-Brandwijk MG (1962) *The Medical and Poisonous Plants of Southern and Eastern Africa*, 2nd edn. Edinburgh and London: E & S Livingstone.

Whitten Tony and Whitten J (1996) *Indonesian Heritage: Plants*. Singapore: Didier Miller.

Yoshimura K and Sagawa N (1918) The chemical constituents in *Cycas revoluta*, Thunb. 1. *Nippon Kagaku Zasshi* [Chemical Society of Japan] 39: 1116–1121. Abstracted in: *Chemical Abstracts* 13: 758–759.

SALAD CROPS

Contents

Dietary Importance

K Gupta and M K Rana, CCS Haryana Agricultural University, Haryana, India

Introduction

Vegetables that are normally consumed raw or uncooked are known as salad crops. They are seasonal, attractive in appearance, and are valued for their appetizing succulence, bulk, and nutritional importance in human diet. Advances in education and communication now mean that people have become health-conscious and aware of quality of food for their better health. Salad is becoming a popular addition in the routine food habits of masses of people all over the world. Salad vegetables consumed raw supply more nutrients compared to cooked ones. All salad vegetables are highly perishable in nature, and their natural fresh form, which is preferred by quality-conscious consumers, is rapidly lost. However, with improved packaging technology, refrigerated transportation systems, and advanced storage facilities, the demand for salad vegetables, especially celery, lettuce, and many others, is increasing day by day.

Types of Plant in the Salad Category

Since there is a vast number of salad vegetables grown and consumed in various parts of the world, it is essential that these must be so grouped that information on them can be accessed readily. They may be classified according to botanical traits, season of growth, edible part of the plant used, and suitability of the soil reaction for their successful cultivation (**Tables 1–3**.) However, as salad vegetables are consumed fresh, it may be pertinent to classify them particularly on the basis of edible portions consumed (**Table 2**).

Table 1 Names of salad crops in nine languages

English	*Danish*	*Dutch*	*French*	*German*	*Italian*	*Portuguese*	*Spanish*	*Swedish*
Artichoke	Artiskok	Artisjok	Artichauft	Artischocke	Carciofo	Alcachofra	Alcachofa	Kronartskocks
Asparagus	Asparges	Asperge	Asperge	Spargel	Asparago	Aspargo	Espárrago	Sparris
Broccoli	Broccoli	Broccoli	Chou-brocoli	Brokkoli	Cavolo broccolo	Brocolos	Brécol	Broccoli
Brussels sprouts	Rosenkal	Spruitbool	Chou de Bruxelles	Rosen kohl	Cavolo di Bruxelles	Couve de Bruxelas	Col de Bruselas	Brysselkal
Cabbage	Kal	Kool	Chou	Kohl	Cavolo	Couve	Col repollo	Kal
Celery	Selleri	Selderij	Céleri	Sellerie	Sedano da erbucci	Aipo	Apio	Selleri
Celeriac	Knoldselleri	Knolselderiji	celeri-rave	Knollensellerie	Sedano rapa	aipo de cabega	Apio-nabo	Rotselleri
Chinese cabbage	Kinesisk kal	Chinese kool	Chou de Chine	China kohl	Cavolo cinese	Couve da China	Col de China	Sallad kal
Endive	Endivie	Andijuie	Chicorée frisée	Endivie	Indivia	Chicoria	Escarola	Endiviesallet
Kale	Gronkal	Groene kool	Chou vert	Grunkohl	Cavolo a foglia riccia	Couve galega frisada	Sin cabeza	Gronbal
Kohlrabi	Knude kal	Koolrabi	Chou-rave	Kohlrabi	Cavolo-rapa	Couve-rabano	Colirabano	Kalrabbi
Lettuce	Salat	Sla	Laitue	Salat	Pattuga	Alface	Lechuga	Sallad
Parsley	Persille	Peterselie	Persil	Petersilie	Prezzemola comune	Salsa frisada	Perejil	Persitja
Parsnip	Pastinak	Pastinaak	Panais	Pastsinake	Pastinaca	Pastinaca	Pastinaca	Pasternacka
Rhubarb	Rabarber	Rabarber	Rhubarbe	Rhabarber	Rabarbaro	Ruibarbo	Ruibarbo	Rabarber
Rutabaga	Kalrabi	Kool raap	Chou-navet	Kohlrube	Navone	Rutabaga	Colinabo	Kal rot
Spinach	Spinat	Spinazie	Epinard	Spinat	Spinaci	Espinafre	Espinaca	Spenat

Adapted from Stadhouders PJ (ed.) (1990) *Elsevier's Dictionary of Horticultural and Agricultural Plant Production*. New York: Elsevier Science.

Nutritional and Dietary Significance

Salad vegetables, categorized as protective foods, are rich in the macro- and micronutrients and fiber that play a significant role in the maintenance of human health. In contrast, salad vegetables are relatively inexpensive. They are quite popular, easy to grow, ship well, and have a long period of harvest but short storage life. They are important for neutralizing the acids produced in the stomach during the course of digestion of meat, cheese, and other fatty foods. They are of value as roughage that helps to prevent constipation. A certain amount of bulky food is necessary for good health. Most salad vegetables, particularly the leaf types, such as celery, cabbage, Chinese cabbage, lettuce, parsley, and spinach, are characterized by a high water content and a relatively high percentage of cellulose or fiber because of their succulence and large bulk. The leafy and root salad crops ordinarily help in the digestion of highly concentrated foods (Tables 4–7.).

The carbohydrates of vegetables comprise starch and sugars, mainly sucrose, glucose, and fructose, which are easily available to the body and constitute the main source of energy. Vegetables also provide unsaturated lipids containing essential fatty acids such as linoleic, linolenic, and arachidonic acid, which are required for normal growth and maintenance of the body. Salad vegetables provide a variety of proteins to the human body.

Salad crops ranking highest in vitamin A are carrot, turnip greens, spinach, mustard, and dandelion (Tables 5–7.). Measured in levels of vitamins and minerals, artichokes rank seventh among the top vegetables. Spinach, kale, and collard contain high levels of vitamins and minerals, particularly vitamin A, calcium, phosphorus, and iron. A single serving meets the daily requirement of vitamin A and C and 13% of the total calcium requirement of an adult.

Beet crops and carrots are considered primarily as a source of carbohydrates, especially sugars. Carrot roots contribute a significant proportion of the daily requirement for vitamin A in US diets for adults, and in addition to vitamins and minerals they provide important dietary fiber.

Likewise, celery is also a good source of calcium, phosphorus, and iron, and is rich in carotene, the precursor of vitamin A, and in riboflavin, niacin, and thiamin. One hundred grams of celery contain protein 6.3 g, minerals 2.1 g, carbohydrates 1.6 g, and fiber 1.4 g.

The nutritive value of cactus is comparable to those of other popular fruits. They are rich in sugars (sucrose 0.2%, glucose 7.0%, fructose 4.8% of fresh fruit mass); vitamin C 22 mg; minerals (calcium 27.6 mg, magnesium 27.2 mg, sodium 0.8 mg potassium

Table 2 Botanical names, common names, and edible plant parts of vegetables/salad crops

Common name	*Botanical name*	*Family*	*Edible plant part*
Monocotyledony			
Onion family			
Chive	*Allium schoenoprasum* L.	Alliaceae	Leaf
Chinese chive	*Allium tuberosum* Rottler ex. Sprengel	Alliaceae	Leaf and immature flower
Lily family			
Asparagus	*Asparagus officinalis* L.	Liliaceae	Shoot
Carpetweed family			
New Zealand spinach	*Tetragonia tetragoniodes* (Pall.) O Kuntze	Aizoaceae	Tender shoot and leaf
Dicotyledony			
Carrot family			
Chervil	*Anthriscus cerefolium* (L.) Hoffm	Apiaceae	Leaf
Celery	*Apium graveolens* L. var. dulce (Mill.) Pers	Apiaceae	Petiole leaf
Celeriac	*Apium graveolens* L. var. rapaceum (Mill.) Gaud.	Apiaceae	Root, leaf
Fennel	*Foeniculum vulgare* var. azoricum (Miller) Thell.	Apiaceae	Leaf
Florence fennel	*Foeniculum vulgare* var. dulce Fiori	Apiaceae	Leaf base
Parsnip	*Pastinaca sativa* L.	Apiaceae	Root, leaf
Parsley	*Petroselinum crispum* (Mill.) nym. var. Crispum	Apiaceae	Leaf
Turnip-rooted parsley	*Petroselinum crispum* (Mill.) nym. Tuberosum	Apiaceae	Root, leaf
Italian parsley	*Petroselinum crispum* (Mill.) Nym. var. neapolitanum	Apiaceae	Leaf
Sunflower family			
Endive	*Cichorium endiva* L.	Asteraceae	Leaf
Chicory	*Cichorium intybus* L.	Asteraceae	Leaf
Indian lettuce	*Lactuca indica* L.	Asteraceae	Leaf
Asparagus lettuce	*Lactuca sativa* L. var. asparagina Bailey	Asteraceae	Stem
Head lettuce	*Lactuca sativa* L. var. capitata L.	Asteraceae	Leaf
Romaine lettuce	*Lactuca sativa* L. var. longifolia Lam.	Asteraceae	Leaf
Dandelion	*Taraxacum officinale* Wiggers	Asteraceae	Leaf, root
Mustard family			
Curled mustard	*Brassica juncea* (L.) Czernj. and Coss. var. crispi folia	Brassicaceae	Leaf
Small leaf mustard	*Brassica juncea* (L.) Czernj. and Coss. var. foliosa	Brassicaceae	Leaf
Gemmiferous mustard	*Brassica juncea* (L.) Czernj. and Coss. var. gemmifera	Brassicaceae	Stem and axillary bud
Involute mustard	*Brassica juncea* (L.) Czernj. and Coss. var. insoluta	Brassicaceae	Leaf
Wide petiole mustard	*Brassica juncea* (L.) Czernj. and Coss. var. patipa Li.	Brassicaceae	Leaf
Rutabaga	*Brassica napus* L. var. napobrassica (L.) Reichb.	Brassicaceae	Root and leaf
Chinese kale	*Brassica oleracea* L. var. alboglabra Bailey	Brassicaceae	Young flower stalk and leaf
White cabbage	*Brassica oleracea* L. var. capitata L.	Brassicaceae	Leaf
Portuguese cabbage	*Brassica oleracea* L. var. costala DC	Brassicaceae	Leaf and Inflorescence
Brussels sprouts	*Brassica oleracea* L. var. gemmifera Zenk.	Brassicaceae	Axillary bud
Konlrabi	*Brassica oleracea* L. var. gongylodes L.	Brassicaceae	Enlarged stem
Broccoli	*Brassica oleracea* L. var. italica Plenck	Brassicaceae	Immature flower stalk
Spinach	*Brassica perviridis* Bailey	Brassicaceae	Leaf
Chinese cabbage	*Brassica rapa* L. var. pekinensis (Lour.) Olsson	Brassicaceae	Leaf
Garden cress	*Lepidium sativum* L.	Brassicaceae	Leaf
Water cress	*Nasturtium officinale* R. Br.	Brassicaceae	Leaf
Buckwheat family			
Rhubarb	*Rheum rhabarbarum* L.	Polygonaceae	Petiole
Valerian family			
Italian corn salad	*Valerianella eriocarpa* Desv.	Polygonaceae	Leaf
European corn salad	*Valerianella locusta* (L.) Laterrade em. Betcke	Polygonaceae	Leaf

161 mg, phosphorus 15.4 mg, iron 1.5 mg $100\,g^{-1}$ edible part) and amino acids.

Chicory, chives, endive, kale, mustard, parsley leaves, spinach, spring onions, and turnip greens are excellent sources of vitamins A and C. Vitamin A ranges from 2500 to 14 000 IU and vitamin C from 30 to 180 mg per 100 g of the edible part (Table 7). The plain-leaf endive (escarole) and the dandelion (leaves) are highly rich in vitamin A (14 000 IU) and contain vitamin C 100 and 35 mg $100\,g^{-1}$ edible plant part, respectively. Sprouting broccoli is most nutritious amongst the cole crops, especially for calcium, iron, vitamins, and protein. As regards quality, kohlrabi is similar to red cabbage but superior to turnip and rutabaga as it has less bitter principle.

Parsley and watercress are good sources of iron, calcium, vitamins A and C, and riboflavin. Spinach,

amongst the highly nutritious salad vegetables, is the best source of vitamin A, providing 8100 IU $100\,g^{-1}$ edible part. This amount also provides vitamin C 51 mg and calcium 93 mg. Spinach is also a good source of iron, phosphorus, and potassium, with 3.1, 51, and 470 mg $100\,g^{-1}$ edible plant part, respectively. The quality of spinach is related to its bright green color, tenderness, and flavor.

Table 3 Classification based on life cycle and optimum temperature (season)

Cycle/vegetable	*Season*
Perennial	*Cool-season crops*
Artichoke	*Hardy*
Garlic	Asparagus
Tomato	Leek
Asparagus	Parsley
Dandelion	Rhubarb
Watercress	Turnip
Chicory	Cabbage
Onion	Dandelion
Rhubarb	Watercress
Chive	Garlic
	Onion
Biennial	Kohlrabi
Beet	Radish
Cabbage	
Celery	*Half-hardy*
Leek	Carrot
Parsley	Celeriac
Carrot	Fennel
Celeriac	Lettuce
Parsnip	Artichoke
Cauliflower	Beet
Florence fennel	Celery
Kohlrabi	Chicory
Turnip	Endive
Annual	*Warm-season crops*
Cucumber	Tender–tomato
Lettuce	Very tender – cucumber
Radish	
Endive	
Mustard	

Turnip greens are considered valuable in the diet primarily because of vitamin A and C contents and the minerals calcium and iron. Leafy green and yellow salad vegetables are important sources of vitamin A and vitamin C, and contain appreciable quantities of thiamin, niacin, and folic acid.

Industrial Use

The consistent availability of fresh and processed vegetables in the world market indicates the immense economic significance of the vegetable industry as a major segment of the agricultural economy. The dehydration of onions is done at industrial level in the USA, especially in California. Asparagus, tomatoes, lettuce, cucumber, and carrot are produced commercially and processed into different food products on an industrial scale. Some cauliflower, cabbage, rat-tailed radish, turnip, and carrot are stored, canned, and shipped. The commercial production of kale and spinach is confined along the East Coast and in California, respectively. Sauerkraut is the main processed product prepared from cabbage.

Chinese cabbage is fermented to produce the Korean dish kimchi, which includes either fermented Chinese cabbage or fermented radish along with hot peppers, garlic, and other ingredients. Commercial quantities of chives are sold to processors, and chopped leaves are mixed with cottage cheese or cream cheese for the retail market. Lettuce leaves are sometimes used as a substitute for tobacco in cigarettes. Certain species of cactus are used for making jam and syrups, alcoholic drinks, honey and cheese, and fruit juices, and soap is made from the leaves.

Celery has been used as a garnish and condiment in Europe, Egypt, and China since old times. Probably the plants used for these purposes were of the 'smallage' type, with hollow, narrow petioles. Presently

Table 4 Nutritional constituents of some salad crops (per 100 g edible part)

	Macronutrients					*Vitamins*					*Minerals*		
	Water (g)	*Energy (KJ(kcal))*	*Protein (g)*	*Fat (g)*	*Carbohydrate (g)*	*A (IU)*	*C (mg)*	*B_1 (mg)*	*B_2 (mg)*	*Niacin (mg)*	*Ca (mg)*	*P (mg)*	*Fe (mg)*
Asparagus	91	100 (26)	2.5	0.2	5.0	900	33	0.18	0.20	1.5	22	62	10.0
Artichoke	85	84 (20)	2.7	0.2	2.3	160	11	0.06	0.08	0.8	53	78	15.0
Beet greens	91	100 (24)	2.2	0.3	4.6	6100	30	0.10	0.07	0.4	119	40	3.3
Beetroot	88	180 (43)	1.6	0.1	9.9	20	10	0.03	0.02	0.4	16	33	0.7
Kale	83	220 (53)	6.0	0.8	9.0	10 000	186	0.16	0.26	2.1	249	93	2.7
Mustard	90	130 (31)	3.0	0.5	5.6	7000	97	0.11	0.22	0.8	183	50	0.3
Rhubarb	95	67 (16)	0.6	0.1	3.7	100	9	0.03	0.39	0.3	96	18	0.8
Spinach	51	100 (26)	3.2	0.3	4.3	8100	51	0.10	0.22	0.6	93	51	3.1
Turnip greens	90	117 (28)	3.0	0.3	5.0	7600	139	0.21	–	0.8	246	58	1.8

Table 5 Nutritional constituents of some salad crops (data per 100 g basis)

Crop	Macronutrients					Vitamins					Minerals		
	Water (g)	Energy (KJ(kcal))	Protein (g)	Fat (g)	Carbo-hydrate (g)	A (IU)	C (mg)	B_1 (mg)	B_2 (mg)	Niacin (mg)	Ca (mg)	P (mg)	Fe
Chicory	95	63 (15)	1.0	0.1	3.2	Trace				18	21	0.5	
Chinese cabbage	95	58 (14)	1.2	0.1	3.0	150	25	0.05	0.04	0.6	43	40	0.6
Endive (curly)	93	84 (20)	1.7	0.1	4.1	3300	10	0.07	0.14	0.5	81	54	1.7
Garden cress	82	280 (67)	5.8	1.0	8.7			0.15			360	110	26.6
Lettuce	96	54 (13)	0.9	0.1	2.9	330	66	0.60	0.06	0.3	20	22	0.5
Mustard	90	130 (31)	3.0	0.5	5.6	7000	97	0.11	0.22	0.8	183	50	3.0
Parsley	85	184 (44)	3.6	0.6	8.5	8500	172	0.12	0.26	1.2	203	63	6.2
Spinach	91	109 (26)	3.2	0.3	4.3	8100	51	0.10	0.20	0.6	93	51	3.1
Watercress	89	138 (33)	2.9	0.2	4.9	2803	13	0.12	0.38	0.8	290	140	4.6

Table 6 Nutritional constituents of some salad crops (per 100 g edible portion)

	Macronutrients					Vitamins					Minerals			
	Energy (KJ(kcal))	Water (g)	Protein (g)	Fat (g)	Carbo-hydrate (g)	A (IU)	B_1 (mg)	B_2 (mg)	Niacin (mg)	C (mg)	Ca (mg)	Fe (mg)	Mg (mg)	P (mg)
Carrot (roots)	130 (31)		89	0.8	0.2–6.6	1300	0.05	0.04	0.3	6	30	1.4	17	36
Celeriac roots	84 (20)	88	1.5	0.3	3.5	0	0.05	0.06	0.7	8	43	0.7	20	115
Celery self-blanching petioles	29 (7)	96	0.7	0.1	1.2	90	0.03	0.02	0.3	7	25	0.3	10	27
Green (petiole)	34 (8)	95	0.9	0.1	0.2	120	0.03	0.04	0.3	10	70	0.5	14	34
Cucumber fruit (immature)	50 (12)	96	0.6	0.1	2.2	45	0.03	0.02	0.3	12	12	0.3	15	24
Florence fennel	63 (15)	93	1.1	0.1	2.6	100	0.04	0.02	0.4	9	44	0.8	23	38
Mustard greens	63 (15)	91	2.7	0.2	0.1	5300	0.08	0.11	0.8	70	140	2.0	48	45
Parsley leaves	67 (16)	90	2.2	0.3	1.3	5200	0.08	0.11	0.7	90	125	2.0	79	40
Radish (Chinese winter icicle, scarlet globe)	54–58 (13–14)	94	0.6–0.1	0.1	0.1–0.2	0	0.03	0.02	0.3	21	27	0.8	22	27
Tomato fruit	80 (19)	94	0.9	0.1	3.7	1700	0.10	0.02	0.6	21	6	0.3	10	16
Turnip (tops and roots)	75 (18)	92	1.8	0.2	0.2	2700	0.05	0.07	0.5	70	125	1.5	45	45
Turnip greens	58 (14)	91	1.5	0.3	0.1	3400	0.07	0.10	0.6	60	90	1.1	31	42

the celery seeds are used as a pleasant flavoring agent in tomato juice, sauces, soups, pickles, pastries, salads, and other savoury dishes.

Fennel (*Foeniculum vulgare*), a tall, hardy, aromatic perennial of the parsley family, is native to southern Europe, and the Mediterranean area, especially in the vicinity of the sea. Fennel is derived from the Roman *Foeniculum*, a variety of fragrant hay, and has been used as a condiment by the Chinese, Indians, and Egyptians since long ago. *F. vulgare* var. dulce Alef is cultivated in France for its young stems, which are used in salads, and for its seeds; young shoots are used as a vegetable and herb in Italy. The leaves are feathery green, finely divided, and there are golden yellow flowers. The fresh needle-shaped leaves are used to flavor fish sauces and to garnish mackerel.

Biological Properties

The medicinal significance of the above-mentioned salad vegetables in human life is briefly described below.

Asparagus shoots used as salad are useful to cleanse the bowels and to cure jaundice and other stomach disorders. The cactus may reduce sugar and cholesterol levels in human blood and its seeds are used in pharmaceuticals. Globe artichoke is known for its medicinal significance in reducing blood clotting and capillary resistance, and in neutralization of some acids. The juice extracted from globe artichokes is used as a remedial treatment of gastrointestinal disorders.

Chicory is used for the remedy of warts. The boiled leaves of dandelion are used as a diuretic and for liver, kidney, and urinary disorders. The extract taken from

Table 7 Nutritional constituents of some salad crops (per 100 g edible portion)

Crop	Macronutrients					Vitamins					Minerals			
	Energy (KJ(kcal))	Water (g)	Protein (g)	Fat (g)	Carbo-hydrate (g)	A (IU)	B_1 (mg)	B_2 (mg)	Niacin (mg)	C (mg)	Ca (mg)	Fe (mg)	Mg (mg)	P (mg)
Cabbage (white, red, Savoy),	20	92	1.2–2.0	0.1	0.4	200–1000	0.06	0.03	0.3	31–60	35–51	0.4–0.7	17–28	34–42
Cauliflower (snowball)	22	90	2.2	0.3	0.9	40	0.08	0.06	0.6	72	30	0.5	12	60
Celtuce (leaves)	12	94	1.1	0.4	1.3	3500	0.09	0.12	0.5	33	59	0.8	38	34
Celtuce (stalk)	12	94	0.6	0.2	0.4	70	0.02	0.02	0.6	6	18	0.3	17	43
Chicory (leaves)	13	92	1.7	0.3	1.1	4000	0.06	0.10	0.5	24	100	0.9	30	47
Chicory (roots)	23	80	1.4	0.2	4.6	0	0.04	0.03	0.4	5	41	0.8	22	61
Chives (leaves)	20	92	2.8	0.6	1.1	6400	0.10	0.18	0.7	79	81	1.6	55	51
Endive (leaf type)	11	95	1.3	0.2	1.2	2500	0.07	0.08	0.4	8	42	2.0	20	30
Endive (curled type)	20	93	1.7	0.1	4.1	3300	0.07	0.14	0.5	10	81	1.7		54
Globe artichoke	20	83	2.7	2.3	2.3	160	0.08	0.06	0.8	11	53	1.5	48	78
Lettuce (butter head)	11	96	1.2	0.2	1.2	1200	0.07	0.07	0.4	9	40	1.1	16	31
Lettuce (Cos/ Romaine)	16	94	1.6	0.2	2.1	2600	0.10	0.10	0.5	24	36	1.1	6	45
Lettuce (crisp head) (Great Lakes)	11	95	0.8	0.1	2.3	300	0.07	0.03	0.3	5	13	1.5	7	25
Onion (green bulb)	21	90	1.3	0.2	4.0	330	0.06	0.05	0.3	32	62	0.5	25	43
Onion (green leaves)	19	92	2.0	0.2	3.4	5000	0.07	0.14	0.2	45	80	1.0	24	30

In some cases, the analytical data relating to a given crop will appear to vary between articles. These differences may arise because of contrasts in experimental techniques and/or raw materials employed by the original authors, and hence no attempt has been made to standardize the available information.

spinach leaves is used to clean wounds and cure warts in the UK. According to documentary evidence in the literature, watercress has been used as a medicinal plant from the first to the nineteenth centuries; its principal use is as an antiscorbutic.

The plants of the Brassicaceae family have been claimed to possess cancer-preventive properties. Increased consumption of broccoli, particularly among women, appeared to improve survival against lung cancer. Broccoli may also play an important role in lowering the serum cholesterol levels. Plants of the Crucifereae family contain *S*-methylcysteine sulfoxide, which is principally responsible for reducing blood cholesterol levels in the body.

Sulforaphane, a substance isolated from broccoli and some other crucifers, helps to prevent the growth of tumors developed by cancer-causing toxin in mice. This may also have future applications in human cancer treatment. The discovery of antithyroid properties of cabbage causing goiter in rabbits and other studies have related the occurrence of endemic goiter in humans in Belgium, western China, and Canada and in animals to the consumption of rutabaga, turnip, cabbage, kale, rape, and other crucifers.

Researchers and breeders increasingly are concerned about the concentration of glucosinolates in cruciferous crops and their influence on humans and animals. Glucosinolates as such are not injurious, but their hydrolytic forms cause serious damage. When plant cells are crushed, the myrosinase (thioglucoside-glucohydrolase) enzyme becomes active and starts hydrolyzing the glucosinolates present in the cells into various breakdown products, such as isothiocyanates (mustard oils), thiocyanates, and nitrites. However, there is increasing evidence associating regular consumption of cabbage and other *Brassica* vegetables with the reduced incidence of some forms of cancer, probably due to the presence of glucosinolate-derived products.

Several claims have been made that celery seeds improve longevity and memory, eradicate skin wounds and cure leprosy. It has also been used in the UK for the treatment of rheumatism since the nineteenth century and is still used today as a nerve stimulant in India. In the Orient, celery is used as the medicine tranians; the boiled seed oil is inhaled as vapor to cure headaches.

In past centuries fennel has been involved in medicine as a carminative, stimulant, and stomachic, and

as a cure for asthma, earache, toothache, rheumatism, and cough. Even today it is being used as one of the principal ingredients of 'gripe water' given to small babies with digestive troubles and recommended by herbalists to obese persons to reduce their weight.

Leek and other vegetables of the genus *Allium* are known to have aphrodisiac, blood purification, and antithrombotic properties. The antibiotic effects of *Allium* are due to the occurrence of sulfur compounds, the bactericide being allicin. Onion has been prized for its medicinal significance since Gerarde. In the *Great Herbal* (1596) it was stated that 'the juice of onions snuffed up into the nose purges the head ... Forged with salt, rue and honey ... it is good against the biting of mad dogs ... Anointed upon a bald head in the sun brings up the hair again very speedily.' Onion and garlic are used as folk medicine to cure headaches, remove parasites, and for heart and circulatory disorders. In China, onion tea is taken to treat headache, cholera, fever, and dysentery.

Ancient Greek and Roman physicians used to cure jaundice, headaches, and fevers with extract of wild lettuce leaves. The extract for the treatment of cancers of the uterus, liver, and other tumors was also used. Unlike its close wild relatives, the cultivated form *Lactuca sativa* does not contain the bitter principles sesquiterpene lactones, lactucin, and lactupicrin, having sedative properties. The other form of lettuce, *L. virosa* L., is used in the pharmaceutical industry for manufacturing the drug lactucarium.

Rhubarb was prescribed in the literature for liver and stomach disorders and to eradicate ringworm. In Europe towards the end of the Middle Ages a huge demand for the purgative drug derived from its dried roots established a steady trade from central Asia.

Use in the Home as Raw or Industrially Processed Food

Freshness contributes prominently to crispiness and taste of the salads that are customarily consumed uncooked and in a raw state. A small fraction of salad vegetables is also processed to convert them into more delicious and durable food products. Ordinarily carrot, radish, turnip, cauliflower, cucumber, and chillies are used to make pickles in addition to their use in fresh salads. Celery, lettuce, onions, and tomatoes, the most popular and important salad vegetables, are available throughout the year. However, due to their medicinal importance in human diet, carrots, cabbage, lettuce, mustard, onion, rhubarb, spinach, and watercress are used as salad.

Artichoke is generally boiled or steamed before it is consumed. However, some people prefer to take it raw, when it is harvested at a young stage. The bottoms and hearts are canned, frozen, marinated, or kept in oil, and also used to produce pulp. Inulin, the carbohydrate stored in the Jerusalem artichoke tubers, on hydrolysis yields fructose, and is therefore recommended for diabetes mellitus.

Numerous cultivated forms of cabbage (*Brassica oleracia* var. capitata), with a wide range of features, are available in nature. One or another variety, e.g., Spring Green, January King, Savoy, Round-Headed, White Cabbage, and Red Cabbage, can be made available throughout the year. These are particularly used in the frozen-food industry and in prepared salad. Many types of Chinese cabbage are extremely important in the diets of Asians, used raw in salads, cooked as greens, or fermented and preserved for later use. Fermented Chinese cabbage or radish, along with hot pepper, garlic, and other ingredients, forms the popular Korean dish kimchi.

In Mexico, the USA, Chile, Brazil, Spain, Italy, Greece, and Northern Africa, certain genera of cactus form an important part of people's dietary requirements. The tender young pads of *Opuntia* and *Nopalea* species of cactus, known as Nopalitos, are extensively used as fresh green or salad vegetable and fruit in Mexico and Texas. With the excellent quality and flavor of fresh fruits, its young leaves serve as a nutritious vegetable and salad dish and immature fruits for making mock-gherkins.

Celery, chervil, leeks, and tomatoes are used to make delicious soups and stews. Rhubarb and tomatoes are also used in the preparation of sauces. Chervil, chicory, leeks, fringed-leaf endive, fennel leaves, and parsley are also used to flavor and garnish the food. Lettuce, cabbage, radish, turnip, and cucumber are industrially fermented to lettuce kraut, sauerkraut, sauerriipen, pawtsay, dill pickles, sour pickles, and salt stocks. These fermented products are relatively more appetizing, palatable, and flavorful than the unfermented ones.

Finely ground powder prepared from dried, roasted chicory roots is used as a substitute of coffee or in blends with coffee. Chicory and endive are grown as potherbs and their leaves are used to some extent only. Chives are used in salads or in omelets and other dishes. A close relative of the globe artichoke, a leafy plant known as cardoon is grown for its edible roots, or more often for its petioles, which may be used either as a cooked vegetable or eaten raw in salads.

Parsley, being appealing in flavor and taste, may be used both on and in most foodstuffs except sweets, as a seasoning and garnish in soups, salads, fish, meats, and vegetables. Its foliage is in greater demand than its seeds. In California, USA, West Germany, Belgium, Hungary, Canada, Spain, and France, industrial

dehydrated parsley flakes are produced on a commercial scale.

Toxic Substances

Asparagus contains saponins and cholinesterase inhibitors. Carrot roots hold carotatoxin, which causes neurotoxic symptoms. The leaves of rhubarb and sorrel have a high content of oxalic acid, which on ingestion reacts with soluble calcium and synthesizes insoluble calcium oxalate, and as a result inhibits the absorption of calcium. Hence, because of their toxicity, they should not be consumed in huge quantities. The glucosides present in rhubarb appear to cause strong irritation. White-colored oca tubers possess a slightly higher oxalic acid content.

Brassica plants possess cholinesterase enzyme inhibitors and a number of different goitrogens that cause goiter in humans. Various species of mustard contain glucosinolates, which are responsible for several digestive disorders. Parsley and celery contain psoralens and terpenoids, which cause dermatitis, whereas some species of parsley and celery also contain alkaloids and a cholinesterase enzyme inhibitor. Lettuce is known to be an accumulator of a nitrate form of nitrogen. Spinach contains oxalic acid and saponins, and also tends to accumulate high levels of nitrates, which may be reduced to nitrites. Nitrites may convert hemoglobin to methemoglobin, interfering with the normal transportation of oxygen in the blood.

Some species of cucumber have a high content of glycoalkaloid compounds, usually the glycosides (cucurbitacins), which contribute a bitter taste to the fruit, and their higher concentrations appear to be very toxic. Hot pepper contains a high amount of amide or resinoid capsaicin, which may sometimes cause gastric disorders and skin irritation.

The cultivated forms of vegetables are generally considered as safe if consumed appropriately, and have many alimentary benefits that contribute to the maintenance of health. However, excess ingestion of beneficial substances can produce harmful effects, whereas only low concentrations of some toxic compounds can be endured by the body. The important consideration is the dosage that makes the toxic compound effective to produce injurious effects, and where and how it acts in the body (Table 8).

Table 8 Toxins present in salad crops

Name of crop	*Toxins*
Artichoke	Grosshemin
Brown mustard	Progoitrin
Cabbage	Glucocapparin (progoitrin)
Cabbage leaves	Spirobrassinin
Celery	Xanthotoxin
Chicory	Cichoralexin
Cucumber	Cucurbitacin
Fennel	Anethole
Garden beet	Malonic acid
Hot pepper	Capsaicin
Lettuce	Lactucin, lactucopicrin
Mustard	Gluconasturtiin
Parsley	Malonic acid
Parsnip	Xanthotoxol
Rhubarb	Oxalic acid
Spinach	Coumestrol, histamine/oxalic acid

See also: **Onions and Related Crops**; **Tomatoes**; **Vegetables of Temperate Climates**: Cabbage and Related Vegetables; Oriental Brassicas; Carrot, Parsnip, and Beetroot

Further Reading

Howard FD, MacGillivray JH and Yamaguchi M (1962) *Bulletin No. 788*. Berkeley, CA: Agricultural Experimental Station, California.

Lincoln CP (1987) *Vegetables: Characteristics, Production and Marketing*. New York: John Wiley.

Maynard DN and Hochmuth GJ (1997) *Knott's Handbook for Vegetable Growers*. New York: John Wiley.

Rubatzky VE and Yamaguchi M (1997) *World Vegetables: Principles, Production and Nutritive Values*. New York: International Thompson.

Ryder EJ (1979) *Leafy-Salad Vegetables*. Westport: AVI.

Salunkhe DK and Desai BB (1984) *Post Harvest Biotechnology of Vegetables*, vol. II. Boca Raton: CRC Press.

Wood R (1979) *Salad Crops*. London: Marshall Cavendish.

Wynuewood, Pennsylvania (1969) *The Book of Spices*. Philadelphia, USA: Livingston.

Yamaguchi M (1983) *World Vegetables: Principles, Production and Nutritive Values*. Westport: AVI.

Leaf-types

K Gupta and M K Rana, CCS Haryana Agricultural University, Haryana, India

Introduction

Leafy salad vegetables are amongst the most important and universally used vegetable crops grown throughout the world, and are also available throughout the year. They are attractive in appearance and have value for their appetizing succulence, bulk, and nutrients. These vegetables mostly have crisp, tasty,

and tender edible leaves. Leafy salad vegetables are nutritionally important sources of protein, fiber, minerals, and vitamins. Their prominence as an important crop group has been heightened over several decades through awareness of consumers regarding nutritional and other obvious values they offer to the diet as fresh greens. As these crops have become a daily table staple, their acreage has expanded to a remarkable degree over the past few decades, making them an important segment of the agricultural and marketing industries, especially in the USA.

Nearly all vegetables can be consumed raw, but few of them are commonly referred to as salad crops, of which only a few are considered as leafy salad vegetables based on the criterion that their leaves alone are consumed. Leafy salad crops are defined as those crops from which leaves and associated parts are harvested for use as raw vegetables. Their succulent leaves and petioles are susceptible to damage by pests, poor production and handling systems, and environmental extremes over which growers have little or no control.

There are four major leafy salad crops – celery (*Apium graveolens*), chicory (*Cichorium intybus*), endive (*Cichorium endivia* L.) and lettuce (*Lactuca sativa* L.) – and other lesser vegetables used primarily for salads, flavoring, or garnishing the foods. The latter include Brussels sprouts (*Brassica oleracea* var. gemmifera), cabbage (*B. oleracea* var. capitata), Chinese cabbage (*B. pekinensis*), chervil (*Anthriscus cerefolium*), dill (*Anethum graveolens*), garden cress (*Lepidium sativum*), mustard (*B. juncea*), parsley (*Petroselinum crispum*), spinach (*Spinacia oleracea*), and watercress (*Rorippa nasturtium-aquaticum*). Most salad crops thrive best during cooler and moist seasons, and do not tolerate hot weather conditions.

Global Distribution, Commercial Importance, and Varieties

Brussels sprouts (*B. oleracea* var. gemmifera) are similar to cabbage when prepared as salad for table purposes. This crop has been grown around Brussels and Belgium for hundreds of years and is an important vegetable in Europe, particularly in England. Out of Europe it is not important, although it is widely grown in Australia and California.

Improved Long Island, Catskill, Early Morn Dwarf Improved, and Friither Zwerg Kvik are important varieties belonging to the dwarf type. Bedfordshire, Cambridge no. 1, 3, and 5, Irish Elegance, and Sherradian are some of the important cultivars grown in England and Ireland. Early cultivars such as Breda (Netherlands), Weibulls Rapid (Sweden), and Wilhelmsburg (Germany) have light green sprouts. Late hardy cultivars include Red Vein, Hild's Idea, DeRoshy Amager, and Polarstjernen.

Cabbage is one of the most important and oldest recorded vegetables, cultivated throughout the world. The word cabbage (*B. oleracea* var. capitata L.) is probably derived from the French word cobbache meaning head, and the name *Brassica* from the Celtic name bresic for cabbage. Cabbage was mentioned in the early literature and was probably in general use as early as 2000 or 2500 BC. In 1987, the total area under cabbage crop was 1 682 000 ha with a production of 38 132 000 tonnes. Asia has the largest production area (811 000 ha) followed by the former Soviet Union (420 000 ha) and Europe (296 000 ha).

Cabbage and Chinese cabbage (*B. pekinensis* and *B. chinensis*) are annual potherbs grown as salad vegetables. In China they have been cultivated since the fifth century and were introduced to the USA in the late nineteenth century. In the USA, Chinese cabbage is grown throughout the year in California, Florida, and Hawaii. The importance of Chinese cabbage is also increasing in European countries, notably the UK, France, and Germany. The most popular varieties are Chihili and Wong Bok, grown in the USA. Two more or less distinct species are grown: petsai (*B. pekinensis*) and pak choi (*B. chinensis*). Shantung Cabbage, Chefoo Cabbage, Pekin Cabbage, Chou De Chine and Celery Cabbage are the known varieties of Chinese cabbage.

Celery (*Apium graveolens*) is an herb, widely cultivated for its edible swollen leaf petioles. Green and white varieties are available. Celeriac is a variety of celery (*A. graveolens* var. rapaceum) with an edible tuberous base to the stems. Celery was first cultivated as a medicinal crop as early as 400 BC in Egypt and in the Roman Empire. The Greeks used it to make a celery wine called selinites. Crowns made of celery stems and leaves were awarded to winners of athletic games at the Isthmus of Corinth.

Chicory (*Cichorium intybus* L.), a member of the family Compositae (Asteraceae), is an important perennial salad vegetable. It is native to Europe and Asia, and may have originated from the Mediterranean region. Chicory is also known as French endive, witloof, witloof chicory, and succory. The first mention of chicory cultivation was in 1616 in Germany. The blanched cluster of leaves (head) called witloof chicory, which has a delicate flavor, is widely used in France, Belgium, and The Netherlands. The dried, roasted, and ground roots are also used as a coffee substitute or in blends with coffee.

Radichetta and Rosso de Verona are the important chicory cultivars in the USA and Italy, respectively. Penninck, Christaens, Mueninck, and Van Espen are Belgian cultivars, while Normato, Malina, Bubbal

Blank, and Slusia Meilof are Dutch cultivars. Flambor, Bergere, and Zoom are chicory hybrids. Tardivo and Mitado are the best cultivars for forcing, giving highest yield and best quality produce. Best forcing varieties under hydroponic system in terms of yield and quality are Tardivo, Mitado, Nunhems 825-76 and Nunhems 820-76, Zoom, Alvaro, and Barola. For very early forcing, Daliva, Arnova, Zoom, Mitiva, and Wivro are the outstanding cultivars.

Chervil is native to Europe and has finely divided pinnate leaves. It was cultivated in England in 1597 and in America by 1806. The origin of salad chervil (*Anthriscus cerefolium* L.) lies in the region of south-eastern Europe and western Asia. Cress is also known as pepper grass or garden cress, and is grown as a salad crop all over the world. A native of Iran and Europe, it spread from Persia to India, Egypt, and Greece. It was mentioned by Xenophobe in 400 BC, and has been cultivated in England since 1548. In the USA it is cultivated and also grows wild. Four types of cress are cultivated: common, curled, broad-leaved, and golden cultivars. Chervil may also be used as a garnish, in soups or as a cooked vegetable.

Endive (*Chichorium endivia* L.), belonging to the family Compositae (Asteraceae), is an annual or biennial popular salad vegetable in Europe. Its popularity has also increased in the USA. It is believed to be a native of Egypt or India. It is indigenous to eastern Mediterranean areas and may have originated from a cross between *C. intybus* and *C. pumitum*. The Egyptians used it at a very early period. During the thirteenth century, it reached northern Europe. It was noted in the USA in 1806. Endive is an important market garden crop in Florida, USA. Its cultivation began in England in 1686 and in France in 1826. Endive is classified into two general groups: (1) the curled or fringe-leafed cultivars, and (2) the broad-leaved cultivars. The Green Curled Ruffic, Deep Heart Fringed, Green Curled Pancalier, and White Curled are curled or fringe-leafed types. Varieties representing the Broad-Leaved class are broad-leaved Batavian, Full-Heart, Batavian or Escarole and Florida Deep Heart.

Kale (*B. oleracea* var. acephala) is a nonheading type of cabbage and belongs to the family Cruciferae (Brassicaceae). It is a hardy food and a fodder crop. It can tolerate temperatures as low as $-15\,^{\circ}C$ as well as high temperatures. There are various types of kale, but curly-leafed types are grown for human consumption. Cultivation of kale probably began seven millennia ago; the Greeks grew them as early as 600 BC. Kale varieties have been classified as Marrow-Stem Kale, Leafy Kale, Thousand-Head Kale, and Curly Kale. The most popular varieties of kale in the USA are Dwarf Blue Scotch and Dwarf Green Scotch.

Lettuce, an annual and a most popular cool-season leafy salad vegetable of great commercial importance, belongs to the tribe Cicoreae and family Compositae. The cultivated lettuce (*Lactuca sativa* L.) is related to wild lettuce (*L. scariola*), a common weed in the USA. Heading lettuce is a relatively recent cultivated crop, being first described as cabbage lettuce in 1543. It is grown throughout temperate, tropical, and subtropical climates. Both cultivated and wild lettuce species are native of Europe and Asia, and have been in cultivation for the last 2500 years. Evidence from tomb paintings indicates that it was cultivated before 4500 BC. In India, the Portuguese and the British introduced it during the sixteenth century. The crop spread throughout the Mediterranean region during the Greek and Roman era and from there to the rest of Europe. It was recorded as being cultivated in the New World by 1494. Lettuce is known as green gold in California, the most important vegetable-producing state in the USA.

Lettuce production and consumption have increased several-fold over the last two decades throughout temperate countries. However, in tropical and subtropical countries, its popularity is still confined to urban populations. Lettuce ranks sixth in acreage and third in value among 22 principal vegetables. World lettuce production is estimated at about 3 million tonnes and it is grown on more than 300 000 ha. Lettuce is grown in all continents, but the greatest consumers and producers are the USA (91 000 ha, of which 60 000 ha are in California) and Europe (total European Community area 80 000 ha). Lettuce is also grown in large areas of south-eastern Australia, Japan, China, Israel, northern Mexico, Chile, Argentina, Brazil, and Peru. The value of the US crop is nearly $800 million and in the UK lettuce is the single most valuable vegetable crop, with a value of £78 million. It is widely used in all salad preparations, and has a somewhat bizarre use as a substitute for tobacco.

Four types of lettuce – head-type (*Lactuca sativa* L. var. capitata), leaf-type (*L. sativa* L. var. crispa), cos-type (*L. sativa* L. var. longifolia), and asparagus lettuce or stem-type (*L. sativa* L. var. asparagina) are grown in the USA. The crisp head-type is the most popular and is extensively grown in the USA. About 70% of US crisp head lettuce is produced commercially in California, 16% in Arizona, and the rest in several other states. The American type has recently become increasingly popular in the UK and Spain. Crisp head lettuce is often called Iceberg lettuce, when marketed as a trimmed overwrapped head. However, this is somewhat misleading since there is

a crisp head cultivar called iceberg, which is unsuitable for iceberg production. Butter head and cos (Romaine) are popular in northern and southern Europe. Leaf lettuce is grown mainly in Mediterranean areas and also in South America. Stem lettuce is cultivated mainly in the orient.

One hundred and fifty distinct varieties have been reported, of which 20–25 are commercially important. They have been classified into three general classes:

1. Butter head varieties:
   (a) Bunching varieties
   (b) Cabbage heading varieties
2. Crisp head varieties:
   (a) Bunching varieties
   (b) Cabbage heading varieties
3. Cos-type varieties:
   (a) Lanceolate-leaved varieties
   (b) Lobed-leaved varieties
   (c) Spatulate-leaved varieties

The crisp head class includes most of the important commercial varieties, of which New York 515 and Imperial 44, 152, 456, 615, and 847 are the most popular. Great Lakes and Penn Lakes are other important crisp head varieties: the former has several strains. Big Boston, White Boston, May King, Salamander, and Wayahead are the common varieties of butter head type, but they are of relatively little commercial importance. Black-Seeded Simpson, Early-Curled Simpson, Grand Rapids, and Slow Bolt are the main varieties of leaf or bunching lettuce. Paris White (White Cos) and Dark Green are the cos-type varieties.

Cultivars include Sea Green, Empress Calsweet, Green Ice, Cabrillo, Cal K-60, Calmaria, Chaparral, Mesaverde, Morangold, Morgangourd, and Royal Oak Leaf. In India Great Lakes (head-type) and Chinese yellow (loose-leaf-type) are commonly grown cultivars. Imperial 859 and Slow Bolt are also improved cultivars of lettuce, and the Indian Agricultural Research Institute, New Delhi, has released all four cultivars.

Mustard (*B. juncea* or *B. alba*) is a member of the cabbage family. It is used both as a salad ingredient and as a green vegetable. There are several species of mustard used for spice, flavoring, oil seeds, green fodder, manure, and greens or salad. In addition to being grown for its seeds, white mustard may be grown as a salad plant, cover crop, and green manure. It is principally an oilseed crop, but is commonly grown as a leafy vegetable in northern parts of India, China, and tropical and subtropical regions. All the *Brassicas* are native to central Asia and the Himalayas, with secondary centers in India and eastern China. The commercially important white mustard (*Sinapsis alba*), brown mustard (*B. nigra*), and Ethiopian mustard (*B. corinata*) are grown all over the world for various uses. They are particularly common in the USA, Canada, the UK, and most countries of Europe, India, China, and the South-east Asian countries. In the USA, mustard for salads and greens is grown in Texas, California, Florida, Georgia, Louisiana, and Mississippi. There are two principal salad or greens cultivars, namely Florida (broad leaf) and Southern Giant (curled leaf), which are grown in the USA. Varieties Ford Hook Fancy, Ostrich Plume, and Southern Giant are the popular varieties of *B. juncea* grown in the USA.

Parsley (*Petroselinum crispum* Mill.), a biennial or short-lived perennial cool-season crop, is native to the Mediterranean region and Europe. Parsley is popular as a garnish for salads, sandwiches, and cooked dishes; as flavoring for soups and pasta, and as a salad ingredient – only in small amount owing to its pungent flavor. The Hamburg type is used mainly in soups and stews. The swollen root of turnip-rooted parsley is eaten as a cooked vegetable. It was popular in early times among the Greeks and Romans. It is valued as an aid to digestion and in suppressing odors of onion and wine. It has been in cultivation for over 2000 years. Parsley is produced commercially in Texas, California, New Jersey, Florida, and New York. Parsley was introduced into England from Sardinia in 1548 and into America by 1806.

Five types of parsley varieties – plain-leafed, celery-leafed or Neapolitan, curled, fern-leafed, and turnip-root types – are available. The curled type is most common but plain and turnip-root types are also grown in the USA. The best-known varieties are Moss, Extra Double Curled, Fern-Leafed, Curled Dwarf, Ever Green, Extra Triple Curled, Plain, and Dark Green Italian. Morgo was developed in Sweden, and Hamburg is a turnip-rooted variety.

Spinach (*Spinacia oleracea* L.) belongs to the family Chenopodiaceae. The name spinach is derived from the Spanish word espinaca. Spinach is annual for leaf production and biennial for seed production. It is the most important potherb or green vegetable grown throughout the world and is used raw in salads or mixed with other salad vegetables. Its leaves and tender shoots are consumed fresh or processed and canned. Spinach is native to central Asia, most probably Iran (Persia). Spinach cultivation developed during the Greek and Roman civilizations about 2000 years ago. It was introduced and known in China as early as the seventh century (647 AD). Arabs introduced spinach into North Africa and therefrom the Moors introduced it into Spain in the eleventh

century and into America in 1806. The crop was known in Germany in the thirteenth century and elsewhere in Europe in 1552. Spinach is of two types: prickly-seeded and smooth-seeded.

The major commercial areas for growing spinach are California, mostly in Ventura country; the Winter Garden region of the lower Rio Grande Valley of Texas; Maryland, New Jersey, and Colorado. California contributes one-third of the total spinach produced for industrial processing. One hundred and twenty-one spinach varieties obtained from American and European sources were tested and it was concluded that only five distinct varieties exist. Varieties were classified into Savoy leaf and flat leaf types, instead of the original four groups – Norfolk or Bloomsdale, round-leafed, thick-leafed, and prickly-leafed. Juliana, Virginia Savoy, Old Dominion, Hollandia, Viroflag, Amsterdam Giant, King of Denmark, and Nobel are important varieties of these two classes. The varieties Viro flag, Fruremona, and Wiremona are good for processing. Beltsville Arlington (Virginia), Beltsville (Maryland), Winter Haven (Texas), and Davis (California) are long-standing varieties in different countries; many hybrids are also available.

Watercress, a unique perennial salad vegetable having worldwide distribution, commonly grows in flowing water as a wild or cultivated crop. It is believed to be a native of the eastern Mediterranean and Europe. It grows wild in streams in the cool regions of the world, but is also found growing wild above 1500 m in the Himalayas of Nepal and hilly areas of the USA. Most production is in Virginia, Maryland, Pennsylvania, and California. It was used as a medicinal plant in AD 77 as it has antiscorbutic properties. In Europe, watercress formed one of the tributes to the high Kings of Erin (Ireland) during the Dark Ages. The first large-scale cultivation of watercress is suggested to have been in southern Germany, near Erfurt, around 1750. It was grown in England for the first time in 1808 and in France in 1811 where large-scale cultivation was started to supply the markets of London and Paris.

Watercress is cropped for its side stems and leaves. It has a hot mustard/radish-like flavor and pungent taste, and is generally used in mixed salads, as a garnish, in sandwiches and soups. Seedlings with the hypocotyl and green cotyledons are cut and used in the UK for salads and sandwiches. Its elongated shoot with vigorous basal branching produces a sward crop. Repeated cutting leads to increased branching and as many as 10 crops a year may be harvested. For commercial outdoor cultivation it is planted in beds covered with slowly flowing clean water. There are two main cultivars: Green Cress, which is diploid, and fertile, and Brown Cress or Winter Cress, which is triploid and sterile. Cress (*Lepidium sativum* L. f. Cruciferae) is a much-branched annual with linear leaves. The Moss Curled (curled) variety is branching, dark green, and with a mild, but distinct flavor.

Features of the Edible Parts

In Brussels sprouts, instead of a single head at the top of the stem, miniature heads are born in the axils of leaves distributed all along a tall stem, which is crowned with a cluster of loose leaves. The dwarf varieties of Brussels sprouts have short stems, mostly less than 50 cm long. In dwarf cultivars sprouts are of medium size and the distance between sprouts is small. Tall late hardy cultivars are preferred for a long growing season. Early cultivars have light green sprouts. The small solid heads 5–7 cm in diameter are tender and delicious.

Michihili, the most common improved variety of Chinese cabbage, is uniform in size and heading. The heads are usually 46 cm long and 10 cm thick, and cylindrical in shape with a tapered tip. The outer foliage is dark green, the inner leaves are blanched, and the midrib is white and broad. The old Chinese cultivar, Won Bok, has light green leaves. The petsai, which resembles the cos lettuce, produces a much larger, elongated and compact head; the leaves are slightly wrinkled, green, thin, and much veined. The midrib is broad and light in color. Another variety, pakchoi, resembles Swiss chard in its growth habit. The leaves are long, dark green, and oblong or oval, and do not form a solid head.

In the case of Radichetta chicory, the leaves are dark green, narrow, and notched, and the Italian cultivar is a red, nonforcing type.

Endive is a plant with numerous radical leaves, which are smooth, lobed, more or less deeply cut and spreading into a rosette; the stem is hollow, from 50 cm to over 1 m high, ribbed and branching. Endive grows as a loose head of leaves, which are usually strongly ruffled and serrated. The outermost leaves are green and bitter in taste, but the inner leaves are light green to whitish in color.

Kale is closely related to the wild cabbage, and most of its forms bear a rosette of leaves at the top of the stem. Its tender young leaves are used for salad. It also produces edible inflorescence like cauliflower and broccoli.

Lettuce forms a rosette of large, longish leaves, which are somewhat spoon-shaped and more or less undulated/toothed at the edges. From the center of the rosette springs a nearly cylindrical stem, which narrows very rapidly and branches at about one-third of its height. It is furnished with clasping

leaves, which encircle the stem and become narrower as they approach the top. There are six morphological types of lettuce: crisp head, butter head, cos, leaf, stem, and Latin, depending on changes in leaf shape and development of the heading character.

Crisp head-type lettuces have unbranched stems, which generally remain less than 30 cm long, owing to the growth being arrested at an early stage. As the growing point continues to form leaf primordia, a rosette of sessile leaves arises. The first leaf unfolds normally, and can reach a length of over 30 cm. After some time, other leaves are produced which unfold only partially and form a layer, embracing the laterally formed leaves that do not expand. This continuous segregation and growth of young leaves develops a head. Crisp head cultivars are usually large, heavy, brittle-textured and tightly folded with green outer and white or yellowish inner leaves.

Butter head comprises heads of very soft, buttery, textured crumpled leaves, which have less prominent veins and midribs than in crisp head types. The variety Imperial has relatively soft textured light to medium green leaves with serrated or wavy edges. The leaves of Great Lakes are brittle with ruffled margins, and are bright green to yellow with prominent ribs. Leaves of Empire are light green, deeply serrated and crisp, while the heads are often slightly elongated or conical. Butter head varieties are sometimes referred to as cabbage lettuce. Their broad oily, soft-textured, crumpled leaves form relatively small, loose heads. There are a number of other cultivars, which vary in size, color (including red), and appearance.

Leaf lettuce primarily produces a rosette of leaves and has no heading tendency. Like butter head lettuce, it forms a loose head, but the leaves are elongated, resembling those of cos lettuce.

Cos (Romaine) lettuce has elongated heads of long leaves with heavy midribs. The outer leaves are coarse and tough in appearance and dark green, but the inner leaves are fine-textured and light green. The eating quality of cos lettuce is good and it is very popular in southern Europe and the Mediterranean region.

Stem lettuce is sold as Celtuce in the USA; the stems are peeled and used as salad or cooked vegetable.

Latin lettuce leaves are somewhat elongated, but more leathery than Romaine. It is very different in appearance from other types of lettuce, having a thickened, elongated stem, which is peeled and used as salad or cooked vegetable.

Mustard leaves are very strongly flavored and pungent. The inner younger leaves are milder and are best suited for salad. The basal rosette leaves of mustard vary in form, while the stem leaves are mostly entire. Leaves are large, broadening towards the apex, and have smooth or curled margins. Florida Broad Leaf is a vigorous, large, and erect plant with large, thick, broad, oval leaves with toothed margins. Southern Giant (curled leaf) is large and upright with large, wide leaves and curly leaf margins.

Parsley is a rosette of divided leaves on a short stem. The curled-leaf-type is most common and attractive. Its three subtypes – double-curled (moss curled), evergreen, and triple-curled – are distinguished by the degree of leaf curling, coarseness of the leaf, and plant growth habit. Moss curled (double-curled) has a stem about 30 cm tall with vigorous, compact and very dark green leaves, which are finely cut, deeply curled, and frost-resistant. Leaves of evergreen type are coarsely cut, while extra triple-curled leaves are finely cut and very closely curled. The triple-curled variety has slightly shorter leaves, which are closely curled. The leaves of the variety Paramount are tall, very uniform, triple-curled, and a very dark green color. The plain leaf parsley (Dark Green Italian) has deeply cut leaves but no curling or fringing. Plain-leaved types include plain (singles) and have flat leaves. The leaves of the Dark Green Italian cultivar are heavy and glossy green.

Spinach, in general, produces rosettes of fleshy leaves, which may be crinkled (Savoy) or smooth in the vegetative phase. The leaves are ovate, rounded or triangular, succulent, and borne on a short stem. In the second stage of growth, the stem elongates, producing a seed stalk with narrow, pointed leaves. The shape, size, and color may differ and be characteristic of a varietal group. Varieties used for processing have large, smooth, dark, or medium green leaves with a semierect or spreading habit.

The watercress plant has three parts: first is an aerial stem portion with an apex, and leaves with no roots projecting above the water; second is a stem portion beneath the flowing water, which has foliage and a system of adventitious roots that arise from the leaf axils and remain free in water; and third is the normal roots. Watercress leaves are pinnately compound with three to 11 leaflets, rounded or oblong, and slightly fleshy with smooth margins.

Chemical and Nutritional Composition

Brassica vegetables have isothiocyanates, thiocyanates, and nitrites as major sources of their characteristic flavor, i.e., pungent aroma and bitter taste. Brussels sprouts are rich in vitamin A and ascorbic acid and contain appreciable amounts of riboflavin, niacin, calcium, and iron. Kale, being a rich source of

vitamin C, carotenoids, and calcium, ranks high among the greens in nutritive value.

Cabbage is rich in vitamin C and high in minerals and proteins. Chinese cabbage provides 25 mg vitamin C, 240 IU equivalents of vitamin A, and moderate amounts of calcium, iron, phosphorus, and potassium per 100 g of raw product. The cold-hardy varieties possess higher soluble sugar content. The sugar to dietary fiber ratio and soluble protein content are related to flavor quality.

The fringed-leaf endive is quite an ornamental potherb and is more popular as a salad vegetable than the broad-leaf type which is used mainly unblanched in stews and soups. Curly endive provides 81 mg calcium, 3300 IU equivalents of vitamin A and 10 mg vitamin C as well as moderate amounts of phosphorus, potassium, and iron per 100 g edible part. Escarole has 14 000 IU equivalents of vitamin A and vitamin C 100 mg $100\,g^{-1}$ edible part. All endives contain moderate amounts of minerals. Chicory is rich in vitamins containing 1400 IU equivalents of vitamin A and vitamin C 100 mg $100\,g^{-1}$ edible part.

Lettuce, rated as fourth, behind tomato, citrus and potato on overall consumption basis, ranks 26th on a nutritional scale. Butter head lettuce, producing a large number of green leaves, is more nutritious than the crisp head type. Cos and leaf lettuces are still more nutritious, mainly because of their high carotenoids (1900 IU vitamin A equivalents) and vitamin C (18 mg) content per 100 g fresh weight. Crisp head lettuce supplies 6 mg ascorbic acid, 330 IU equivalents of vitamin A and calcium 20 mg $100\,g^{-1}$ fresh weight. However, butter head lettuce contributes 8 mg ascorbic acid, 970 IU equivalents of vitamin A and calcium 35 mg $100\,g^{-1}$ fresh weight. Other nutrients like phosphorus, iron, sodium, and potassium, are present in varying amounts in the different lettuce types. Although lettuce has a very high water content (94–95%), it can still prove a useful source of minerals. Leaf texture, color, crispness, taste, flavor, and resistance to early bolting are considered the quality characters.

The young inner leaves of mustard that are tender and mild in taste are consumed as salad. Mustard greens are highly nutritious. A 100-g serving of cooked mustard contains 48 mg vitamin C, 5800 IU equivalents of vitamin A, 138 mg calcium and moderate amounts of other elements.

Healthy parsley leaves with their familiar agreeable flavor are an excellent source of carotenoids, vitamin C, and iodine, a good source of potassium, and a moderate source of iron, sodium, and phosphorus. However, its contribution to the diet is negligible. Parsley per 100 g edible part contains 8500 IU equivalents of vitamin A, 172 mg vitamin C, and 203 mg calcium. (*See* **Fatty Acids**: *Trans*-fatty Acids: Health Effects.)

Spinach is the best vegetable source of provitamin A, providing 8100 IU equivalents per 100 g edible part, and it also contains 51 mg vitamin C and 93 mg calcium. Spinach is a good source of iron (3.1 mg), phosphorus (51 mg), and potassium (470 mg). However, the calcium is said to be unavailable due to the presence of oxalic acid, forming calcium oxalate. It may also contain appreciable amounts of nitrate and nitrite, which are injurious to human health. (*See* **Minerals – Dietary Importance; Nitrates and Nitrites.**)

Raw watercress leaves and stews supply 79 mg ascorbic acid, 4900 IU equivalents of vitamin A, calcium 151 mg per 100 g edible part and moderate to large amounts of other elements.

Postharvest Handling and Storage

Baskets are used to pick and carry Brussels sprouts to the packing houses. The individual boxes are overwrapped with a transparent film. Film packaging is useful in preventing moisture loss because the transpiration rate of Brussels sprouts is relatively high. The 0 °C temperature coupled with 95–98% relative humidity are optimum storage conditions for Brussels sprouts to keep them in marketable condition for about 3–5 weeks. The sprouts can be stored for a longer period at -2 °C temperature. A controlled atmosphere, having 2.5–5.0% O_2 and 5–7.5% CO_2, maintains the quality of Brussels sprouts held at 5 or 10 °C but not at 0 °C. Oxygen levels below 1% can cause internal discoloration.

Cabbage heads are usually harvested with a knife or sickle, retaining few leaves below the head intact to provide a kind of protection. For prolonged storage sufficient care is taken to prevent injuries due to bruising. The healthy heads are packed in perforated polyethylene bags and stored at low temperature immediately after harvest. Refrigeration at 1.6 °C or lower and low relative humidity is required to check the storage disorders. The lower the temperature, the longer will be the storage life, but severe freezing damages the quality. Solid heads are comparatively more prone to freezing injury than merely firm heads. A temperature of 0 °C with 98% relative humidity is advocated for storage up to 6 months.

All cole crops should be stored free from ethylene to avoid premature senescence and tissue injury. Prevention of excessive moisture loss is essential during the course of storage. The major advantages of controlled-atmosphere storage include slight reductions in weight loss and retardation of general senescence,

such as yellowing, toughening, and loss of the flavor of cabbage. White cabbage stores well in 5% O_2 plus 2.5–5% CO_2. However, if O_2 is maintained near zero or CO_2 at 15% or higher for 1 month or longer, there is discoloration of internal leaves, even though the external leaves of cabbage may remain normal. Some varieties which are stored for off-season marketing can be maintained in an acceptable condition for 2–3 months, and up to 6 months with the use of modified-atmosphere storage (1% O_2 and 5% CO_2 at 0–1 °C and 98–100% relative humidity). (*See* **Packaging**: Packaging of Solids; **Storage Stability**: Parameters Affecting Storage Stability.)

Chinese cabbage is harvested by hand at leaf base; the outer diseased or damaged leaves are trimmed off and packed in flat baskets, boxes, and crates of various sizes. The head should be packed loosely and preferably upright in crates. The produce can be stored for several weeks at 0 °C temperature and 95% relative humidity. However, at this temperature with 98% relative humidity, the Chinese cabbage can be kept in marketable condition for about 4–6 months. Proper spacing should be maintained in between containers for air circulation in storage. A low concentration of oxygen has been reported to be beneficial in prolonging the storage life of Chinese cabbage.

The compact, brightly colored and marketable size heads of chicory are harvested by cutting the plants from the base, and their damaged or diseased leaves are trimmed. Such heads, after washing and vacuum cooling or hydrocooling to 0 °C temperature, are packed in layers in wire-bound crates or cartons or baskets for transit. The smaller package is much desirable. The heads are stored for 2–3 weeks at 0 °C temperature and 90–95% relative humidity. Chicory can be stored for a longer period with crushed ice in and around the package.

Endive and escarole are not adapted to long storage and even at 0 °C temperature they cannot be kept satisfactorily for more than 2–3 weeks. Vacuum cooling or hydrocooling can be helpful in retaining their fresh appearance. The relative humidity in rooms, where endive or escarole is kept, should be maintained above 95% to avoid wilting. They can be kept somewhat longer if stored with crushed ice in or around the packages.

Lettuce is harvested as soon as its plants reach an acceptable size and firmness; harvesting should be completed before the leaves become tough in texture and bitter in taste, and prior to the emergence of seed stalk. The lettuce grown for market is allowed to grow to its full size and develop a solid head while, for home consumption it is often harvested before the head is well developed. The heads are hand-cut just below the lowest leaf; loose and damaged leaves are removed, and the produce is washed prior to packaging.

Generally three types of market preparation – naked pack, source-wrapped, and bulk pack – are adopted:

1. Naked pack: the heads are cut, trimmed, and packed in two tiers (12 heads each) in cardboard boxes.
2. Source-wrapped: after being cut and trimmed, the heads are wrapped, heat-sealed, and placed 24 per carton.
3. Bulk pack: cut heads, collected in bulk containers, are transported to a central packing house where they are cored, shredded, washed, and cooled immediately after harvest. Cooling is predominantly done by a vacuum method, particularly for head lettuce shipped to distant markets. To aid vacuum cooling, clean water is sprinkled on lettuce heads prior to carton closure if they are dry and warmer than 25 °C. Refrigerated trucks are used for the shipment of lettuce.

For long storage lettuce should be packed in perforated plastic films, used either as individual head wraps, or in carton or crates with liners to maintain a high humidity around the lettuce. To save on space and weight, trim two 'wrapper' leaves before packing, rather than keeping all five to six leaves, as extra wrapper leaves are not required to maintain quality. When lettuce reaches the market, wrapper leaves are removed and discarded to improve the appearance. Removal of wrapper leaves at harvest and wrapping the heads with plastic film reduces the amount of injury sustained by lettuce during packing.

Polyethylene-lined (0.04 mm thick) packaging checks lettuce leaf rot by 30–40% compared with open polyethylene bags. Mechanical injury is the major cause of quality loss and waste in unwrapped lettuce. Packaging of Romaine lettuce in unsealed polyethylene-lined packaging, in spite of high relative humidity, markedly decreases the rate of decay during storage. This beneficial effect of packaging is due to the high CO_2 and low O_2 levels in the package as a result of respiration. Lettuce should be precooled to − 1 °C just after harvest and stored at 0 °C temperature and 98–100% relative humidity. The rate of general deterioration at 15 °C is observed to be about 2.5 times higher than that at 0 °C.

Lettuce, being highly perishable, deteriorates quickly. The respiration rate increases rapidly and storage life decreases concomitantly as storage temperature increases from 0 to 25 °C. Leaf lettuce respires at about twice the rate of head lettuce. Lettuce can be stored for 3–4 weeks at about 0 °C, which is

twice a long as at 3–4 °C, if it is brought in good condition at the storage facility. Lettuce is easily damaged by freezing, therefore, storage room temperature must be kept above the highest freezing point (− 0.2 °C) of lettuce.

A limited use of controlled-atmosphere storage with about 1.5% CO_2 and 3–5% O_2 has been recommended for storing lettuce in good condition. This atmosphere prevents russet spotting and butt discoloration. High CO_2 (above 2.5%) and low O_2 (below 1%) levels in storage atmosphere are injurious to lettuce. Thus the problems of russet spotting, bacterial soft rot, and carbon dioxide injury causing serious losses can be checked. Kinetin application delays chlorophyl loss and slows down the increase in abscissic acid content. Two days prior to harvest a combined application of GA_3 10 ppm and the cytokinin isopentenyl-adenine 0.1 ppm delays lettuce senescence most effectively.

Small bundles of mustard greens are made after discarding the old and diseased leaves. It is handled like other greens at 0 °C temperature and commonly shipped with inpackage and top ice to maintain freshness. Vitamin content and quality are retained so long as wilting is prevented. Due to high rates of respiration, kale is prone to rapid desiccation, therefore provision of rapid cooling, low temperature, and high-humidity storage conditions is important until the commodity is consumed.

Harvesting of parsley, cutting outer and larger leaves only, extends over a long period of time. The loosely packed or bunched leaves are washed thoroughly before packing in crates or bushel baskets. Crushed ice is usually placed in the package, and shipment to distant markets is made under refrigeration. Much of the crop goes into the institutional trade (restaurants and hotels) and for dehydration. At 0 °C temperature and high humidity parsley can be stored up to 2 months.

Harvesting of spinach usually begins when the plants have attained five to six leaves and continues until before the seed stalks emerge. The plants are cut just close to the soil surface for market, while for processing purposes about 5 cm above the soil surface allows regrowth. Harvesting should be done only during the cooler parts of the day to minimize wilting. The crop should not be cut immediately after rain or heavy dew since the leaves, being at their full turgor, are injured easily. Yellow and diseased leaves are removed, and the produce is handled gently to prevent bruising. As washing in tanks usually results in injury to the leaves, spraying with water on a moving belt is desirable. Round bushel baskets, hampers, and crates are used to pack spinach, and crushed ice is generally placed in the containers when refrigerated trucks are used for long shipments.

Wilting of spinach on account of rapid air movement and low humidity makes it less attractive and causes loss of ascorbic acid content; however, this is of much less significance than unfavorable temperatures. From packaging till its final retail sale continuous refrigeration near 0 °C is generally recommended. Spinach can be hydrocooled or vacuum-cooled equally well. In case of hydrocooling, to remove excess water, it should be centrifuged for about 3 min. Chlorinated water (100 p.p.m.), if used for washing or cooling, avoids bacterial build-up on the produce. Spinach remains usable up to 24 days at 1.6 °C, 7 days at 10 °C, or 2–2.5 days at 18 °C. In spinach stored at 20 °C for 6 days the chlorophyl and ascorbic acid contents decrease greatly and yellowing increases. CO_2-enriched atmosphere (10% CO_2) retards yellowing of spinach and keeps the product in good quality for about 3 weeks at 5 °C. Thus, controlled-atmosphere storage extends the shelf-life of spinach by about 1 week at 5 °C.

The top 15 cm watercress shoots are harvested. Watercress is a highly perishable crop and needs prompt handling after harvesting. Improper handling of watercress will result in loss of its color and the leaves become slimy. It should be precooled by either hydrocooling or vacuum cooling. It is bunched and usually packed in alternate layers with flake ice and kept cool during shipment by spraying with cold water and/or ice. Watercress packed in waxed cartons with top ice holds up well for 2–3 weeks at 0 °C with high (95–100%) relative humidity. A similar storage life is possible using permeable polyethylene crate liners and package icing to minimize wilting. Naked bunches of watercress are highly perishable, even at 0 °C, and may be kept for 3–4 days. The use of polyethylene film liners may be justified to prevent moisture loss under some circumstances.

See also: **Brassicas**; **Vegetables of Temperate Climates**: Cabbage and Related Vegetables; Oriental Brassicas

Further Reading

Jones HA and Rosa JT (1928) *Truck Crop Plants*. New York: McGraw-Hill.

Kalloo G and Bergh BO (1993) *Genetic Improvement of Vegetable Crops*. Oxford: Pergamon Press.

Maynard DN and Hochmuth GJ (1997) *Knott's Handbook for Vegetable Growers*. New York: John Wiley.

Peirce LC (1987) *Vegetables: Characteristics, Production and Marketing*, pp. 163–268. New York: John Wiley.

Ryder EJ (1979) *Leafy Salad Vegetables*. Westport: AVI.

Ryder EJ (1981) 'Sea-Green' lettuce. *Horticultural Science* 16(4): 571.
Ryder EJ and Whitaker TW (1980) The lettuce industry in California. A quarter century of change, 1954–79. *Horticultural Reviews* 2: 164–207.
Salunkhe DK and Desai BB (1984) *Post Harvest Biotechnology of Vegetables*, vol. II. Boca Raton: CRC Press.
Salunkhe DK and Kadam SS (1998) *Handbook of Vegetable Science and Technology: Production, Composition, Storage and Processing*. New York: Marcel Dekker.
Thompson RC (1951) Lettuce varieties and culture. *USDA Farmers Bulletin no. 1953*, p. 42.
Thompson HC and Kelly WC (1957) *Vegetable Crops*, 5th edn. New York: McGraw-Hill.
Tigchelaar EC (ed.) (1980) New vegetable varieties list XXI. The Garden Research Committee, American Seed Trade Association. *Horticultural Science* 15(5): 565.
Watt BK and Merrill AL (1963) *Composition of Food*. USDA Agricultural Handbook No. 8. Washington, DC: US Department of Agriculture.
Wood R (1979) *Salad Crops*. London: Marshall Cavendish.
Yamaguchi M (1983) *World Vegetables: Principles, Production and Nutritive Values*. Westport: AVI.

Other Types of Salad Crops

K Gupta and M K Rana, CCS Haryana Agricultural University, Hisar, Haryana, India

Introduction

Of all the leaf-stem types of salad crops, celery (*Apium graveolens*) is the most important. The other crops of this group are: American wintergreen (*Gaultheria procumbens* L.), asparagus (*Asparagus officinalis*), baby corn (*Zea mays* var. crassa), broccoli (*Brassica oleracea* var. *italica*), dandelion (*Taraxacium officinale* Wig.), florence fennel (*Foeniculum officinale* D.C.), globe artichoke (*Cynara scolymus* L.), kohlrabi (*Brassica oleracea* var. gongylodes), oca (*Oxalis tuberosa* Molina), and parsnip (*Pastinaca sativa* L.). These crops are consumed in different forms, especially in temperate countries.

Global Distribution, Commercial Importance, and Varieties

American wintergreen (*Gaultheria procumbens* L.) belongs to the family, Ericaceae, and *Gaultheria* was named after Dr. Gaultheria, a physician from Quebec in about 1750. It is commonly consumed as salad crop for its tender leaves and stem in America and Western countries, as well as Asia to Australia.

Asparagus (*Asparagus officinalis*) is native to Europe and Asia, and it is the young shoots that are the edible parts. There are green and purple varieties of asparagus. Forcing the plants with the exclusion of light produces white spears, which are very popular in Europe. The green variety of asparagus is the one commonly canned and is most popular in the UK.

Baby corn (*Zea mays* var. rugosa or sacchorata), a form of maize, is an ancient crop, as revealed from cob remnants, dating back to about 5000 BC, found at the Tehuacan cave excavations in Mexico. Domestication is estimated to have begun about that time. Baby corn is quite nutritious and is the most popular salad vegetable in United States and Western countries, where its tender ears, canned in 2% citric acid, are imported from developing countries, mainly from Thailand. It is an ingredient of vegetable biryani, mixed vegetables, Chinese food preparations, etc. Soup, crispy vegetable salad, and pickles made from baby corn are other value-added delicacies in big hotels and airlines. In addition to significant nutrition, baby corn imparts diversity to meals because of its texture and flavor. The recognized cultivars of baby corn in India are Prakash, Golden Baby, and Composite Kesari.

Broccoli (*Brassica oleracea* var. italica), a rapidly developing floral vegetable, is usually harvested when the curd is compact and immature. The name broccoli refers to the young shoots that develop in some species of genus *Brassica* ('brocco' is an Italian word for 'shoot'). In the United States, it is also known as Italian broccoli, suggesting its Italian origin. Its peduncle, basically a stem, is consumed raw as salad. The broccoli, believed to be the ancestor of cauliflower, is a vegetable of the highest quality and of commercial importance in USA as it is the favorite vegetable of Americans. Broccoli was one of the top vegetables exported from the USA in 1995. Broccoli and asparagus accounted for about 90% of the $102.6 million of exported vegetables from the USA to Japan in 1994. A large part of the US production goes to the freezing industry. The top broccoli-producing states in 1992 were California, Arizona, and Texas.

Broccoli is gradually becoming a popular vegetable in other parts of the world. Italian Green Sprouting and Calabrese are the most popular varieties of broccoli. The early and medium-maturing strains of Italian Green Sprouting, Midway, Green Mountain, and Grand Central, are dark bluish green in color and produce large compact heads and many lateral shoots after the central head is cut. The cultivar De Cicco, the main winter broccoli grown in Texas, is light green in color and produces abundant lateral shoots.

Celery (*Apium graveolens* L.) belongs to the family Apiaceae (Umbelliferae), and comprises three distinct cultivated varieties that are grown in different parts of the world for different economic purposes. These varieties are classified below, according to the botanical parts of the plant consumed.

1. *Apium graveolens* var. dulce, common name 'celery:' the plants of celery develop succulent, solid petioles that are used mostly raw as salad.
2. *Apium graveolens* var. rapaceum, common name 'celeriac' or 'root celery.' It is characterized by the development of enlarged hypocotyl or root tissue. It is consumed mostly cooked in stews and soups or grated on salad.
3. *Apium graveolens* var. secalinum, common name 'smallage' or 'leaf celery:' the plants have leafy, slender, and often hollow petioles. Its leaves are used as a condiment garnish and for medicinal purposes.

The genus *Apium* is spread world-wide, particularly in the Mediterranean basin, Australia, New Zealand, South Africa, and South America. Among 14 species, *Apium graveolens* is the only cultivated type. The Mediterranean basin is considered to be the center of its origin. Argentina and Chile are the countries richest in *Apium* species. Celery is important in Europe and America. Its habitat extends from Sweden southwards to Algeria, Egypt, and Ethiopia and in Asia from the Caucasus (Bluchistan) to the mountains of India. Wild celery can be found in damp or marshy areas of different European countries, Egypt, and parts of Asia. It was first used as a medicinal plant and as a food in AD 1500. Homer's '*Odyssey*' mentions salinon (celery) grown in 850 BC, but this early form, later called 'smallage,' was a leafy plant, quite pungent and bitter, and used exclusively for medicines.

Celery was demonstrated as a true vegetable in Italy in the sixteenth century. At that time, the first reports of the three horticultural varieties, 'celery,' 'celeriac,' and 'smallage,' appeared. Selection for solid and succulent petioles resulted in 'celery' types, whereas, selection for enlarged hypocotyls gave rise to 'celeriac' types.

Fresh use of celery began in the eighteenth century, and its use as a vegetable in the USA was recorded in 1806. It is used in mixed salads or as an appetizer, in which case, it is often stuffed with cream cheese or peanut butter. In the USA, a large part of the celery crop is consumed in the raw state, but considerable quantities are used in vegetable juices, soups, and stews and as cooked vegetables.

In the eighteenth century, two varieties of solid-stem Italian celery available in France were 'Solid Celery' and 'Striped Rose.' In 1875, Chemin selected a line known as 'Solid Golden White Celery,' which became the variety 'Paris Golden Yellow Self-Blanching.' In 1884, Pascal selected a green celery line from this variety, which was known in the seed trade as 'Pascal Celery.' In 1887, both the varieties were introduced to North America under the names of 'White Plume' for the self-blanching type and 'Giant Pascal' for the green type.

The word 'dandelion' comes from the French 'Dent de Lion' meaning 'lion's tooth' and reflects the jagged shape of the leaf. Previously, it was considered a noxious weed, but now, cultivated dandelion has been developed from the wild species (*Taraxacum officinale*), a member of the Compositae family. It is a great favorite in Europe, and to a lesser extent, in the USA. Its leaves are used raw in salads or cooked as greens. The flowers can be fermented to produce wine. The roots, similar to those of chicory, may be used as a coffee substitute.

Fennel is produced commercially in the USA, but most of its seed is imported from India, Argentina, and Bulgaria. Volatile oil (anethole constituent) of fennel seeds is an adaptable flavoring agent used in the manufacture of pickles, perfumes, soaps, liquors, cough drops, and chicory candy.

Florence fennel (*Foeniculum officinale*) is also a leaf-stem type salad crop. It is a native of southern Europe and resembles wild fennel. The type 'Carosella' (*Foeniculum piperitum*, D.C.), extensively used in Naples and scarcely known anywhere else, is a species very closely related to *F. officinale*. The plant is used while running to bloom. The fresh and tender stems, available only from the end of March until June, are served raw and still enclosed in the expanded leaf stalks; Italian fennel (*F. dulce*), a short-lived perennial also called finnochio, is cultivated as an annual. The base of its leaves becomes enlarged and bulb-like and, when blanched, has the texture of celery being primarily used for flavoring.

Globe artichoke (*Cynara scolymus* L.), belonging to family Asteraceae (Compositae) in tribe *Cynarae*, is named so as to distinguish it from the Jerusalem artichoke (*Helianthus tuberosus*). It is an important vegetable in Mediterranean countries, where 90% of the world artichoke crop is produced. Italy, Spain, France, and Argentina are the main producers. *Cyanara cardunculus* L. var. sylvestris Lam (the wild cardoon) is widely spread in the western Mediterranean basin (Spain, Morocco, Algeria, Tunisia, and Sicily). *C. sibthorpiana* Boiss and Heldr is located in the central part (the Peloponnese, Aegean Islands, and Cyrenaica), and *C. syiaca* Boiss is found in the eastern part (Palestine).

Kohlrabi (*Brassica oleracca* var. gongylodes), belonging to the family Cruciferae (Brassicaceae), is a biennial herbaceous swollen bulb-like stem vegetable grown as an annual. Writers in ancient Rome had described kohlrabi, which was developed in northern Europe in the fifteenth century. It is a minor crop grown in temperate countries and in some subtropical areas. The tubers or swollen stems, which develop above the ground, may be due to secondary growth in thickness. The tuber is harvested before it becomes tough or woody.

Oca (*Oxalis tuberosa* Molina) belongs to the family Oxaliolaceae. It is an ancient crop cultivated by preColumbian people and in the Andean highlands. Commercial production occurs in Mexico, Central America, and the South Island of New Zealand.

Parsnip (*Pastinaca sativa* L.) belongs to Umbelliferae (Apiaceae) family. It is a native biennial of Europe and Western Asia and is cultivated for its large, fleshy taproots and also for its leaves. It was introduced to the West Indies in 1564 but grows well in the tropics only at high altitudes. The Romans and Greeks used it for medicinal purposes and food in ancient times.

Rhubarb (*Rheum* spp.) is a very early and high-yielding vegetable crop, well suited for cultivation in climatically cool regions. The UK, The Netherlands, Germany, the Nordic countries, Canada, and the USA are the main cultivation regions for rhubarb. It is a minor commercial crop plant but very common in gardens. Rhubarb (*Rheum rhaponticum*) is a long-lived perennial plant, with edible stalks, and was originally native to Siberia.

Morphology of Edible Parts

American wintergreen plants are erect, or prostrate, with creeping stems, giving rise to erect branches (15 cm high and leafy at the top). The lower leaves are alternate and serrated with short petioles, but the top leaves are oval, 2.5–5 cm long, apiculate, glaucous, and shining above. The flowers are solitary or in racemes or panicles, and scarlet-colored.

The edible portion of asparagus is the spear that develops from the crown. Green asparagus, which is the greater portion of the crop, is cut 2.5–5 cm below the soil surface. Asparagus spears are usually harvested before excessive elongation occurs to keep the tenderness. The spears are harvested when the tip is about to emerge or just protrudes through the soil surface.

Baby corn is a herbaceous monocotyledon and warm season annual. Plants are monoecious with male flowers borne as the terminal inflorescence (tassels) on the main stem, and female flowers borne separately as lateral inflorescence (ears) developing at the leaf axils. Plants produce one or more ears. Baby corn is the young ear of maize plant harvested before fertilization when the silks have just emerged. The stem-like, dehusked, light yellow-colored, and tender ears of baby corn are used in salads.

The broccoli plant forms a kind of head consisting of green buds and thick fleshy flower stalks. The terminal head is rather loose and green in color, and the flower stalks are longer than those of cauliflower. The internodes are long, and the plant produces axillary flowering shoots in addition to the terminal inflorescence. The color of the flower buds forming the edible portion varies from white to green, depending on the cultivar. The inflorescence comprises fully differentiated flower buds, although it is considered to be a good-quality vegetable if these reach the open stage. The sprouts in the leaf axils develop strongly, especially after removal of the terminal head. The terminal head and the sprouts with bud clusters may be used in salads but also as a cooked vegetable.

The celery plant is characterized by enlarged, tender, edible petioles or leaf stalks. The petioles are broad, with sheathing bases. Celery is a biennial; however, a few strains and related wild species are cultivated as annuals. The plants start bolting (producing seed stalks) when exposed to cold temperatures for a longer time, which makes them unmarketable. The most common form bears five to 12 thick petioles in a tight bunch. A less popular form is turnip-rooted celery, known as celeriac (var. rapaceum). This is a biennial and is grown in the cool season, but it is also sensitive to prolonged cold temperatures. In the first year, the plant normally produces a tight cluster of petioles and leaves attached to a very compressed stem. Celery leaves are divided and pinnate with almost triangular toothed leaflets, which are smooth and dark green in color. Celery has a broad, furrowed, and concave leaf stalk. In the second year, the stem becomes furrowed and branched. Early vegetative growth is characterized by a fair amount of spreading, but leaves arising from the apex of the short stem form compact and elongated heads. The root system of celery is fibrous and spreads with many feeder roots close to the soil surface. The plant bears compound clusters (umbels) of small, white, perfect flowers with five petals and five stamens. The seed, which is actually a fruit formed from two compressed carpels enclosing the actual seed, is very small.

Dandelion produces deeply toothed leaves in a basal tuft and bears a single, composite, yellow flower on a hollow stalk. The cultivated forms of dandelion produce larger leaves than the weedy forms. Improved cultivars, mostly of French origin,

are tender, less bitter, slightly greener and larger than the wild types.

Florence fennel is an annual, very distinct, low-growing and thickened plant with a very stout stem that has joints very close to the base. The leaves are very finely cut and are of a glaucous green color. The leaf stalks are very broad, of a whitish green hue, and overlap one another at the base of the stem, forming a kind of head that varies in size; the inside is white and sweet. The height of the plant, even when grown for seed does not exceed 60–75 cm. The broad umbel with stout, stiff rays bears greenish flowers. The seed, twice as long as it is wide, is flat on one side and convex on the other and marked by five thick yellowish ribs that occupy almost the entire surface of the skin.

In globe artichoke, the edible part is the young capitulum, which is harvested when it is still in rapid growth. The enlarged receptacle (bottom) and the tender, thickened bases of the bracts are eaten. If the capitulum is picked at an early stage, all of the 'heart' can be eaten after removal of the tough outer bracts. The part of the stem just below the young head is edible. In the case of the Jerusalem artichoke, the underground tubers, which are oblong or extremely elongated and weigh 80–120 g, are eaten.

Kohlrabi (*Brassica oleracea* var. gongylodes) is often called turnip-rooted cabbage and is closely related to cabbage. The stem, which is the edible part, is greatly enlarged immediately above ground. It is an excellent vegetable if used before it becomes woody, tough, and stringy. The most popular varieties of kohlrabi in India are White Vienna, Green Vienna, Purple Vienna, and Earliest Ear Furt.

The compact, bushy growth of oca (perennial) is between 20 and 30 cm in height. Tubers can be eaten immediately after harvest, but their flavor is better when exposed to the sun in order to reduce bitterness. Oca is consumed after boiling, baking, frying, or pickling and can also be eaten uncooked. Leaves and young shoots are eaten in salads or as cooked vegetables.

Chemical and Nutritional Composition

Artichoke has interesting nutritional characteristics related to its high content of phenolic compounds, flavonoids, inulin, fiber, and mineral salts. The phenolic compounds are mainly derivatives of caffeic acid. Extracts containing cynarin (1,5-dicatteoyl quinic acid) have some effects on hepatobiliary disease, hyperlipidemia, and cholesterol metabolism. The pharmaceutical industry mainly uses the leaves.

Baby corn is a good source of pro-vitamin A. Celery is an important source of minerals and vitamins. Celery leaves also contain higher levels of vitamin C, carotene, and essential oils. The shape, color, crispness, pithiness, stringiness, ribbiness, petiole cracking, and flavor contribute to the quality of stalk and petioles. Celery is quite popular in the Western world due to its low fat content (0.6 g per 100 g of edible parts.) The composition varies at different stages of maturity; reducing and total sugars show a marked decline in the leaf blades but an increased concentration in the petiole from harvest until the end of storage. Nitrogen increases in leaf blades and petioles towards the end of storage.

Little information is available regarding the nutritional/chemical composition of Florence fennel and American wintergreen vegetables. However, it is reported that dandelion leaves are an excellent source of provitamin A (14 000 IU), vitamin C (35 mg) and calcium (187 mg) per 100 g of edible parts.

The nutritional qualities of oca tubers, with 85% carbohydrate (on a dry-weight basis) are comparable to potato, also presenting starch of good quality. (*See* **Ascorbic Acid**: Properties and Determination; **Calcium**: Properties and Determination; **Carotenoids**: Occurrence, Properties and Determination; **Retinol**: Properties and Determination; **Starch**: Structure, Properties, and Determination.)

Postharvest Handling and Storage

Asparagus loses its freshness and quality quickly after harvest; it is desirable to market or process it as soon as possible. Cooling of the spears soon after harvest helps in maintaining quality and improves its marketability considerably. Hydrocooling is the best way to reduce the temperature of asparagus spears rapidly. Chlorine is usually added in water used for cooling to minimize the bacterial population. The spears should be kept at a low temperature, as high temperatures result in the synthesis of more ethylene and rapid senescence. The losses in tenderness, flavor, vitamin C, and reducing and total sugars are also pronounced at higher temperatures. In addition to general deterioration, spear elongation and development of decay take place at moderately high temperatures. An increased rate of fiber development is directly correlated with spear length and age, and also with high postharvest temperatures.

Asparagus is packed either in plastic film bags or in boxes or crates without bunching. When the loose-pack method is adopted, the butts are cut off, and spears are laid in boxes lined with oily paper with a layer of wet moss placed in the bottom. Packing only spears with tightly closed tips can extend the postharvest life of asparagus. The packaging of asparagus in film bags maintains the marketability, edible, and

nutritional qualities by reducing moisture loss, preventing fiber formation, and decreasing the loss of ascorbic acid, respectively. However, impermeable film bags can produce off-flavors due to anaerobic decomposition. Such packaging would lead to uncontrolled accumulation of CO_2 and depletion of O_2 with possible damage to the product. The film must be adequately permeable to exchange these gases and to dissipate the respiratory heat.

Asparagus spears must not be placed horizontally, as their tips are negatively geotropic and turn upward. Containers with wide bases and narrow tops should always be used to accommodate the long tapered spears. To prevent desiccation, particularly at butt ends, a high humidity is essential; this can be achieved by placing the butts of asparagus on wet pads. Another way to maintain a high relative humidity inside the package is to pack spears in perforated films. Since spears continue to elongate after harvest, therefore, headspace should be provided in packaging to avoid tip damage.

Asparagus can be stored successfully for about 3 weeks at 2 °C, but high-quality asparagus can be maintained in good condition only for 10–14 days under such conditions. Fresh asparagus is highly perishable and deteriorates rapidly at temperature above 5 °C. Storage at 0 °C, coupled with 95% relative humidity, is generally recommended to prolong the marketable life of asparagus for about 3–4 weeks, but asparagus suffers from cold damage if it is held longer at this temperature. The recommended handling temperature is 2 °C along with a relative humidity greater than 95%. Chilling damage occurs with prolonged holding at temperatures lower than 2 °C. Rapid sorting, packing, and storage at 0–6 °C and 90% relative humidity help maintain a high quality.

Controlled-atmosphere storage of asparagus reduces postharvest deterioration and prevents the development of off-flavors, water loss, chlorophyll loss, and toughening. Exposure of spears to $7 \pm 2\%$ CO_2 for 24 h retards bacterial rots. No cold damage is observed at 1.7 °C and 15% CO_2 concentration. Controlled-atmosphere storage of asparagus helps to reduce decay, retardation or even reversal of toughening, retention of a more desirable color, sugar, organic acids and soluble proteins than spears stored in air. The shelf-life of several leafy vegetables including asparagus has been reported to be doubled by immersing the cut ends or dipping the vegetables in solutions of cycocel (CCC).

Broccoli is harvested shortly before the buds begin to open while their heads remain compact and cut to a length of about 20–25 cm. The axillary shoots, as they develop, are cut at a similar stage to that of the central head. Harvested broccoli heads are further trimmed to a length of approximately 15 cm, and bunches weighing about 700 g are tied together for packaging. Bunches are further packed in crates holding 28–30 bunches. The respiration rate of freshly harvested broccoli is very high in comparison with that of asparagus, spinach, or sweetcorn. Hence, to reduce the respiration rate rapidly, broccoli must be cooled soon after harvest and stored at a low temperature and O_2-free environment in order to prolong the shelf-life. Broccoli should be hydrocooled to 4.4 °C or packed in ice immediately after harvest and kept at 0 °C to maintain a good marketable condition. Film wraps can be useful in maintaining a high relative humidity and extending the storage life.

There is no definite stage of maturity at which celery should be harvested for a satisfactory quality. Celery plants are cut off just below the soil surface. Since the plants wilt easily, they should be moved quickly from the field to the packing house for preparatory operations like stripping to remove suckers and diseased or damaged leaves, precooling, sorting, washing, grading, and packaging prior to marketing. Celery is generally hydrocooled or vacuum-cooled. If ice is used in crates or waxed cartons, it is advisable to store celery at around 0 °C to avoid decay. To avoid wilting, packaging in ventilated wraps is advantageous. Box or crate liners should have about one hole per square foot. Celery can be stored at a temperature of 0–0.6 °C and 90–95% relative humidity for about 8 weeks. Trenching in the field or storing celery in the pits is practiced in the fall when there is a danger of freezing before the crop is harvested, but usually for only a short period of time. An O_2 level less than 5% slightly retards yellowing of celery, and a CO_2 level above 2.5% may be damaging in controlled-atmosphere storage. Therefore, all wraps must be perforated to avoid potentially dangerous accumulation of CO_2 or reductions in O_2 below 1%. Celery in storage absorbs foreign flavors, and so it should be kept away from the odors of other products.

Dandelion plants are trimmed, washed, and stored in containers at 0 °C and 90–95% relative humidity to preserve their quality.

Florence fennel plants are harvested just above the taproot. The upper petiole and leaf blades are trimmed relatively close to the bulb. Fennel has a good postharvest life if stored at 0 °C and high relative humidity.

Globe artichoke flower buds need to be handled carefully to avoid bruising and desiccation, as physical injury can accelerate decay as much as a 10 °C rise in temperature. The harvested buds are placed into suitable containers for transport to the packing house, where they are graded according to their size

and quality. Artichoke buds must be precooled as quickly as possible to retain their quality after harvest. Wilting of artichoke can be prevented by hydrocooling or by washing the bracts in cool water. Globe artichoke is then packed in different layers in paper-lined boxes of different sizes followed by immediate cold storage. Frosts can injure artichoke buds in the field, but in the USA, an industry promotional program effectively markets frost-affected buds, unless damaged severely, claiming that they are 'winter kissed' and, therefore, taste better.

The storage life of globe artichoke is related to its initial quality, since injured and decaying buds at harvest deteriorate 1.5 times more rapidly than sound buds, especially when they are held at 4.4 °C or above. Holding artichoke below 4.4 °C can effectively prevent decay. Storage at a temperature of 0–2 °C and a relative humidity of 95% can help maintain a good quality for about 10 days. A temperature of around 0 °C for prolonged storage and 1.7–2.8 °C for normal marketing periods is satisfactory.

Kohlrabi tubers can be stored in a good condition for 2–3 months at a low temperature (0 °C) and high relative humidity (98–100%) since a high humidity is essential to prevent shrivelling and toughening. Packaging in perforated film can be used to reduce moisture loss. Kohlrabi with attached leaves at 0 °C has a storage life of only 2 weeks.

Parsnip is harvested either by ploughing or hand-digging. The roots are usually lifted after removing the foliage. Harvests are delayed until low temperatures or frosts occur because the root quality is improved under such conditions. With the advancement of root maturity, starch starts converting into sugars, and this is accelerated by cold exposure. Parsnip can be stored in the field unless freezing is severe. Roots also store well at a temperature of 0 °C and relative humidity greater than 90%. A high humidity is required to extend the postharvest life because roots desiccate easily. Like carrots, if parsnip roots are exposed to ethylene during storage, they become bitter. Oca's storage characteristics are relatively good if properly cured, and moisture loss is minimized.

See also: **Storage Stability**: Mechanisms of Degradation; **Vegetables of Temperate Climates**: Cabbage and Related Vegetables; Carrot, Parsnip, and Beetroot; Miscellaneous Root Crops

Further Reading

Kalloo G and Bergh BO (1993) *Genetic Improvement of Vegetable Crops*. Oxford: Pergamon Press.

Lincoln CP (1987) *Vegetables: Characteristics, Production and Marketing*, pp. 173–268. New York: Wiley.

Ryder EJ (1979) *Leafy Salad Vegetables*. Westport: AVI.

Salunkhe DK and Desia BB (1984) *Post Harvest Biotechnology of Vegetables*, vol. II. Boca Raton, FL: CRC Press.

Salunkhe DK and Kadam SS (1998) *Handbook of Vegetable Science and Technology: Production, Composition, Storage and Processing*. New York: Marcel Dekker.

Wood R (1979) *Salad Crops*. London: Marshall Cavendish.

Yamaguchi M (1983) *World Vegetables: Principles, Production and Nutritive Values*. Westport: AVI.

Root, Bulb, and Tuber Crops

K Gupta, G Talwar, V Jain, K Dhawan and S Jain, CCS Haryana Agricultural University, Haryana, India

Introduction

Root crops and bulbs (radish, carrot, beets, onion, garlic, etc.) are being consumed (raw or cooked along with other vegetables) increasingly for salad, have prominent fleshy, marketable, underground edible structures, generally have a long storage life, are economically important, and used and grown worldwide for market as well as domestic consumption. Most root crops are hardy to frost. Radish and turnip are susceptible to heat injury.

Edible, underground salad crops are classified as:

1. Enlarged taproot: beet, carrot, radish, salsify, turnip, parsnip, celeriac, rutabaga.
2. Tubers: ahipa, potato beans, Chinese artichoke, coleus potato, Jerusalem artichoke, jicama.
3. Bulbs: onion, leek, garlic, shallot.

Root salad is eaten raw as an appetizer, served as a side dish. Root salad has natural inherent energy and affects our bodies. Root crops tend to give coldness, warmth, and strength to the body, make the body alkaline-acidic (**Table 1**), and are used for different therapeutic applications. Their names are given in different languages in **Table 2**. Carrot and radish can be dipped in sauce and eaten as an appetizer. Some salads are used in cooking to add flavor. Garlic has a distinct flavor that adds to the taste of many meat and vegetable dishes. Leek and onion are also used to add flavor to soups and stews.

Table 1 Commonly used vegetables/salad, their energies, flavor, organs, and therapeutic applications, according to the traditional Chinese concept of food and nutrition

Name	*Energy/taste*	*Organs*	*Therapeutic conditions*
Beet	Neutral: sweet	Spleen	Nourishes fluids; rheumatic pains, congested chest, poor energy circulation
Carrot	Neutral: sweet	Lung, spleen	Diuretic and digestive, improves eyes, removes swellings and tumors, indigestion, cough, dysentery, difficult urination, skin eruptions, chronic diarrhea
Garlic	Warm: pungent	Lungs, stomach, spleen	Warms, expels cold, promotes energy, circulation, destroys worms, kills parasites, arthritis, cold, abdominal pain, edema, diarrhea, dysentery, whooping cough
Onion	Warm: pungent	Lungs, stomach	Diuretic, expectorant, external application for *Trichomonas vaginalis* and ulcers, common cold, headache, constipation, cold, abdominal pain, dysentery, dysuria, mastitis, nasal congestion, facial edema
Radish (seed)	Cool: sweet, pungent	Lungs, stomach	Detoxifies, promotes digestion, eliminates hot mucus discharge, expels cold; abdominal swelling due to indigestion, laryngitis due to continual cough with mucus discharge, vomiting of blood, nose bleeding, dysentery, headache, bloating, hoarseness, diabetes

Table 2 Names of root salads in different languages

	English	*Chinese*	*French*	*German*	*Italian*	*Japanese*	*Russian*	*Spanish*
Root crops	Carrot	hu luo bo	carotte	mohre	carota	ninjin	mos kov kul	zanahoria
	Red beet	gen tian cai	betterave	rote riibe	bietola da orta,	biito, shokuyo	burnaja svekla	remolacha hortelana,
			Rough, Potagere	rote Beete	barba bic. tola, d insalata	bito, kaensa	stolovaja	r. roja, r. de mesa
	Radish	ying tao luo bo	radis, ravonet	garten-Rettich, Radies, Radieschen,	rafano, ravanello, radice	hatsuka daikon		rabanito, rabanete, rabanillo,
	Turnip	Wujing	navet, n.potagar	mairube, speiseriibe	rapa, r. coltivata	kabu	turneps	nalso
	Rutabaga, Swede, Swede, Turnip	ruidian wu jing	rutabaga, navet de suede	steckrube	rutabaga, navone,	rutabaya	brjukva	rutabaga, nabo succo n. de suecta.
Bulb crops	Garlic	da suan	ail, a. ordi-Naire, a. blane	knoblauch	aglio, a. domestico	ninniku	cesnok	ajo ajo comun
	Leek	jiu cong	poireau, Porreau	porree	porro porretta	liiki	luk porkej	ajo porro a. puerro
	Onion	yang cong	oignon	speisezwiebel, Kiichenzwiebel,	cipolla	tam an egi	luk repcatyj	cebolla

Nutritive Composition

Raw root salad crops and bulbs provide varying amounts of water, carbohydrates, minerals, and vitamins (Tables 3, 4 and 5). Carrot, rutabaga, and coleus are high in vitamin A and are within the reach of most consumers.

Root and Tuber

Ahipa: *Pachyrhizus ahipa* (Wedd.) Parodi

Potato beans: *P. tuberosus* (Lam) Sprengel

Other names: ajipa, ashipa, frijol chuncho, Central American tambeans.

Table 3 Chemical composition of root salad crops (per 100 gm fresh weight basis)

Crop	*Energy (Cal.)*	*Water (%)*	*Proteins (g)*	*Fats (g)*	*Carbohydrates (g)*	*Dietary fiber (g)*	*Total ash (g)*
Arracacha	104	73.0	0.8	0.20	24.9	0.60	1.10
Beet root	43	87	1.6	0.1	9.9	0.9	0.97
Carrot	125	86.0–88.8	0.7–0.9	0.2–0.5	6.0–10.0	1.2–2.4	1.1
Coleus	19	77.0	1.3	0.30	22.0	1.10	1.0
Garlic	131	64.0	6.2	0.35	27.9	1.0	1.3
Jerusalem Artichoke	72	79.0	2.2	1.30	16.6	0.85	1.6
Leek	56	83.0	1.9	0.30	12.6	1.4	1.1
Onion	36	90.8	1.3	0.18	8.0	0.52	0.40
Parsnip	76	79	1.7	0.5	17.5	2.0	1.10
Red radish	49	95.4	0.7	0.2	1.9	1.9	–
Rutabaga	46	87	1.1	0.1	11.0	1.20	0.79
Salsify	81	78	2.9	0.6	18.0	1.4	0.96
Shallot	72	79.8	2.4	0.10	16.8	0.7	0.8
Squaw Root	154	60.0	4.6	–	30.8	1.80	–
Table beet Green	23	92	2.0	0.35	4.5	1.7	1.90
Turnip	30	92	1.0	0.2	6.6	0.90	0.74
White radish	64	93.0	0.8	0.1	2.9	6.6	2.72
Yacon	–	76.0	1.3	3.50	–	3.80	–
Yam Bean	40	87.4	0.2	1.20	8.8	0.60	0.30

Table 4 Vitamin contents of root salad crops (mg/100 gm fresh weight basis)

Crop	*Vit. A (IU)*	*Thiamine*	*Riboflavin*	*Niacin*	*Vitamin C*
Arracacha	–	0.06	0.04	3.4	28.0
Beet root	20	0.03	0.05	0.4	10
Carrot	1100	0.04	0.02	0.2	4
Coleus	540	0.05	0.02	1.00	1.0
Garlic	–	0.20	0.08	0.50	11
Jerusalem Artichoke	20	0.19	0.08	1.20	7
Leek	75	0.08	0.05	0.45	15
Onion	40	0.05	0.03	0.15	9
Parsnip	30	0.08	0.09	0.2	16
Red radish	Trace	0.03	Trace	0.4	17
Rutabaga	580	0.07	0.07	1.1	43
Salsify	10	0.04	0.04	0.4	11
Shallot		0.06	0.20	0.20	8
Squaw Root	Trace	0.11	0.12	3.0	13.0
Table beet Green	6250	0.1	0.21	0.4	31
Turnip	Trace	0.04	0.07	0.6	36
White radish	–	0.03	0.02	0.5	24
Yam Beans	Trace	0.05	0.03	0.3	20.0

IU = 0.3 μg Vitamin A alcohol.

Ahipa and potato beans belong to family Fabaceae (Leguminoseae) and have long trailing vines. Ahipa is native to the south-western Amazon, near Bolivia or northern Argentina. Production and consumption are restricted to Bolivia and Peru. Potato beans are widely adaptated to tropical, subtropical, and temperate regions and resemble yam beans which are native to South America. *P. ahipa* vines are shorter than those of *P. tuberosus*. Leaves are entire and rhomboidal ovate in shape. To enhance root enlargement, developing flowers are removed. The roots of *P. ahipa* are smaller and mature earlier than those of *P. tuberosus*. Potato bean roots have a succulent, crispy, white interior. Swollen tubers are like roots, and are mainly eaten raw for salad: all other plant parts are toxic. Only tuberous roots are harvested. The thin, pale yellow skin does little to prevent desiccation; the root is commonly left in the ground until needed.

Arracacha: *Arracacia xanthorrhiza*

Other names: apio, Peruvian carrot, Peruvian parsnip, mandioquina-salsa, zanahoria, racacha.

Arracacha belongs to the family Apiaceae (Umbellifereae), and is a dicotyledonous perennial which is a

Table 5 Mineral contents of root salad crops (mg/100 gm fresh weight basis)

Crop	*Sodium*	*Potassium*	*Calcium*	*Magnesium*	*Phosphorus*	*Iron*
Arracacha	–	–	26	–	52	0.9
Beet Root	105	335	16	23	33	0.7
Carrot	40	240	34–80	9.0	25–53	0.4–2.2
Coleus	–	334	17	–	–	6.0
Garlic	–	50	52	–	–	1.1
Jerusalem Artichoke	–	478	18	15	80	3.0
Leek	16	230	55	25	43	1.9
Onion	6	156	26	11	33	0.4
Parsnip	12	541	50	–	77	0.7
Radish (red)	11	240	19	5	20	0.6
Radish (white)	27	220	30	15	25	0.4
Rutabaga	8	257	66	16	39	0.4
Salsify	13	380	47	23	66	1.5
Squaw Root	12	340	440	32	165	6.5
Table beet green	148	565	121	89	40	3.3
Turnip (medium)	50	207	30.0	150	32	0.3
Yacon	–	high	–	–	–	–
Yam Bean	4	144	17	16	18	0.5

native of the Andean highlands from Venezuela to Bolivia. It is considered to be important as the white potato in some areas of the Andes. It is the preferred staple food of much of South America and is very popular in southern and central Brazil, where it is grown commercially. Wild types have been found in Peru and considerable diversity is found in Ecuador.

Arracacha is similar to carrot and celery, and has a mild carrot or celery-like flavor. Enlarged, fleshy, lateral roots resemble white carrots; the flesh is yellowish and a ring of cambium tissue is purple. The cultivars can vary in color, from white (blanca), to cream yellow (amarilla) and orange-purple (morada). Productive plants develop several (as many as 10) lateral carrot-sized roots clustered around the main rootstock. Root length typically ranges from 5 to 25 cm with diameter from 2 to 6 cm. The yield ranges between 5 and 15 t ha^{-1}. Dark green, purple-tinged petioles resemble celery and are eaten raw or cooked. Roots must be cooked and are commonly fried, boiled, or baked. They have an unusual taste, resembling a mixture of celery, carrot, and parsnips, roast chestnut combined with celery and carrot flavor. Starch content varies from 10 to 25%. This vegetable is easily digestible and is often prepared for infants and elderly.

Beet: *Beta vulgaris* subsp. Rubra

Beet is a fairly hardy biennial, which belongs to the family Chenopodiaceae. It is grown primarily for roots and has been cultivated as a potherb or green throughout recorded history. It is an important home garden and market vegetable crop and is also grown for shipping to distant marketplaces and for processing. Beets are believed to be native to the Mediterranean area of Europe and North Africa and North America.

The major producers of sugar beets are Russia, France, the USA, Germany, Poland, and Italy. Most commercial production of beet is for processing and this was an $8 million industry in 1997.

Four general classes of beetroot varieties are reported which are oblate or top-shaded, oval, half-long and long conical. Only varieties belonging to the root oblate class are of commercial importance. Based on red or yellow color, each class can be subdivided into varieties that also differ in form, earliness, and vigor, as shown in Table 6.

Beet is a good source of carbohydrates (sugar). It contain two pigments; the deep red color is beta-cyanin and beta-xanthin is a yellow pigment.

Freezing, drying, or canning can preserve beets. Uniform dark red color without zones or rings and tender flesh free of fiber are desirable processing qualities. Beets to be dehydrated are harvested at slightly later stage of maturity than those used for canning. Detroit Dark Red and Cylindra are the best varieties for canning and pickling. Beets are also added to soups and beet juice is used for coloring in many different products, as a primary ingredient of borscht (a Russian stock soup).

The best storage temperature for beets is nearly freezing point, 0 °C (32 °F) with 95° humidity. Topped beets store well for 3–5 months.

Carrot: *Daucus carota* sp. sativus (Hoffm.), (2*n* = 18)

Carrot is a biennial, belonging to the family Apiaceae, and is an important vegetable for its fleshy edible,

Table 6 Classification of garden beets

Description	*Cultivars*
Table beets	
Flat shape	
Red	Flat Egyptian
Yellow	Burpee Golden
Top shape	
Light red	Light Red crosby
Red	Crosby Egyptian – popular for market
Dark red	Early Wonder – popular for market
Yellow	Yellow turnip
Globe shape	
Red	Detroit Dark Red, Pacemaker – good for processing and all-purpose
Yellow	Yellow Tankard
Elongate	
Red	Long Dark Blood, Cylindra
Yellow	Long Yellow
Black	Long Black
Benching beets	
Flat shape	Green Top Bunching
Red	
Canning beets	
Globe shape	Ruby Queen
Red	

colorful roots. It varies in colour from white, yellow, orange, light purple, deep red to deep violet. It is widely grown in spring, summer, and fall in temperate countries and during the winter in tropical and subtropical regions.

Afghanistan is considered a center of origin, because the greatest diversity of wild types is found there. South-west Asia and eastern Mediterranean regions are considered to be secondary centers of diversity of wild type and domestication. Purple and yellow carrot variants were introduced into Europe, then to India, China, and Japan.

Carrot production for the fresh market increased from $355 million to $441 million in the year 1997. The market for processing carrots is considerably smaller than the fresh market, with a production value of $37.5 million in 1997. In the USA 99 000 acres of carrot is grown for fresh-market sales, of which 70 900 acres of carrot is grown in California. The total area for processing carrots in the USA is 23 800 acres.

Carrots may be grouped into two types: (1) temperate or European cultivars, which are biennial, orange in color and uniform in thickness with a small core; and (2) tropical or Asiatic cultivars, which are annual, red in color, more juicy, and have a bigger core and a heavier top. The most important bunching varieties of carrot in the USA are Imperator, Gold Spike, and Gold Pak, which have long slender roots with a good smooth exterior whereas the varieties Red Cored Chantenay, Royal Chantenay, and Autumn King are good colored and popular for processing. Nantes is a good home garden variety.

Some important and commonly grown varieties of carrot in India are Amsterdam, Barlikuner, Chantenay, Royal Chantenay, Delattya, Honey Sweet, Karotina, Pusa Kesser, Pusa Meghalli, Rubica, and Pusa Yamadagni.

Carrot is one of the most nutritious vegetables consumed both raw and processed. Gazraila is a sweet dish prepared by cooking shredded carrot in milk and sugar and is much liked in North India.

Carrots are low in calories and are an excellent source (ranging from 41 to 475 $\mu g\,g^{-1}$ fresh weight) of β-carotene, a precursor of vitamin A, which is good for eyesight. A serving size of 78 g carrot has only 35 cal and 2 g dietary fiber, along with 270% of the recommended dietary allowance (RDA) of vitamin A, 10% of the RDA of vitamin C, and 2% of the RDA for calcium. Carrot roots are a rich source of carotenoids, which contains different types of carotene, i.e., β-carotene (45–80%), α-carotene (15–40%), γ-carotene (2–10%) and others (3–6%). Thiamin, riboflavin, niacin, folic acid, and vitamin C are also present in appreciable amounts in carrot roots.

Some volatile constituents are also present, such as α-pinene, camphene, β-pinene, myrcene, α-terpinene, *p*-cymene, limonene, γ-terpinene, terpinolene, caryophyllene, β-bisabolene, γ-bisalolene, heptenol, dedocenal, myristicin, and falacarinol in stem-distilled oil of carrots.

The flavor of carrots is influenced by free sugars glucose, fructose, and sucrose, which contribute to sweetness; volatile mono and sesquiterpenoids, which contribute to harshness, 2-nonenal, which imparts a cooked flavor; isocoumarin and other phenolic compounds, which impart bitterness; 2-methyoxy-3 butypyrazine, which contributes to the carrot aroma; free amino acids, which contribute to the delicate flavor; and ionones, which impart a floral off-flavor. Carrot roots contain several times more sucrose than glucose or fructose.

Carrots are a versatile vegetable that can be served alone, raw or cooked, or, in combination with meats, baked, fried, sautéed, and pickled. A small portion of total production is used for making the fermented product kanji, an appetizing drink.

Carrots are processed into products such as canned carrots, dehydrated carrots, juice, beverages, candy, preserves, and halva. Carrot leaves are used to extract protein. Carrot tops and roots are fed to animals to increase milk yield and used in the preparation of poultry feed.

Carrots possess many medicinal properties and are used in ayurvedic medicine. Carrot is considered a

popular remedy for jaundice. Its juice has diuretic and nitrogen-balancing properties, and thus has been found to be effective for the elimination of uric acid. β-carotene acts as a free radical-trapping agent and a singlet oxygen quencher in biological systems. β-carotene has been reported to reduce the risk of cancer of the lungs, cervix, easophagus, and stomach, and cataract formation. Carrot extract has antioxidant properties. The combination of carrot and orange juice was found to reduce the oxidation of low-density lipoproteins in habitual cigarette smokers. Carrot seeds have been reported to possess antifertility activity in female albino rats.

Carrots are harvested when the roots are 2–2.5 cm in diameter. Fall carrots, if mulched in certain areas, can be left in the ground all winter and harvested as needed. After harvest, prompt cooling to 5 °C or below extends the storage life of carrot roots. Mature roots can be stored for 7–9 months at 0–1 °C temperature coupled with 90–100% relative humidity.

Chinese artichoke: *Stachys affinis* Bunge

Other names: Chryorogi, Japanese artichoke, knot root, kon loh (China).

The Chinese artichoke is a perennial herb, belonging to the Lamiceae (Labiateae) family, and is grown for its tubers (cylindrical) while constricted internodal segments, appearing bead-like on a square-shaped stem, and petiolated ovate leaves form an opposite pair on the stem. Flowers are white or pink. The tubers are usually pickled, and are a popular food in China and Japan and are also eaten fried or fresh in salads. Much of the carbohydrate (about 60%) is the tetrasaccharide, i.e., stachyose. Tubers are dug, washed, and eaten fresh or after cooking. They are commonly stored in damp sand.

Coleus potatoes: *Coleus parviflorus* (Chinese potato)

Other names: Sudan or Hausa potato (*Solenostemon rotundifolius* (Poir), J.K. Morton: *Coleus rotundifolius*, Livingston; kaffir or hausa potato: *Pleetranthus esculantus* (N.E. Br.).

The coleus potato is native to southern India. It is commonly called kourka and is also grown in tropical and subtropical regions of Asia, Sri Lanka, and Africa. The plant is an annual, belong to the Lamiaceae (Labiatae), family. It has simple leaves with serrate margins, oppositely attached to a succulent, square-shaped stem. Clusters of white-fleshed, small, round, brown or black starchy tubers are present at the base of the primary stem. Leaves can be used for flavoring. Tubers are eaten raw or cooked, pickled, and eaten as condiments in India. Leaves and roots have an aromatic flavor, and contain about 11–20% carbohydrate and 1–5% proteins. Harvest should not be delayed, as quality declines with overmaturity. It is stored well in cool, well-ventilated conditions.

Jerusalem artichoke: *Helianthus tuberosus* L. (2*n* = 102)

Other names: girasole, topinambour, and sunchoke. Jerusalem artichoke.

The Jerusalem artichoke is a herbaceous perennial plant, cropped as an annual, belonging to the Asteraceae (Compositae) family. Plants can be found growing wild from Kansas north to Minnesota and as far east as Nova Scotia in Canada and central USA. The Jerusalem artichoke is endemic to North America. Native American Indians were utilizing the tubers long before European colonization. It is not an artichoke: derived from the Italian, *girasole articocco*, it tastes like a globe artichoke, *Cynara scolymus*.

The fleshy, thin-skinned tubers are variable in size, shape, and highly branched, and the inside is white with a crispy and watery texture. Tuber colors very from white to yellow and sometimes red or purple. Tuber carbohydrate content ranges from 14.5 to 20%; of this, 65–82% is fructose and the inulin content is 22.4% in some varieties. Only two varieties, Mammoth white French and Sutton's white, are widely grown. V71, D1, D2, Patate Vilmorin, Comber Challenger, and Moder are also important cultivars.

A unique feature of Jerusalem artichoke is the production of inulin, a polymer of fructose (which stores carbohydrate and is a suitable food for diabetics). Tubers are used raw in salad or cooked like potatoes and are also used for animal feeding. Tubers are utilized for sugar extraction, fuel alcohol production, as a source of fructose, and for the preparation of 5-hydroxy methyl furfural.

Tubers are harvested in the late fall when stems are completely senescent or killed by cold temperature. Stems are cut away to facilitate harvesting. Yields vary by 15–40 t h^{-1}. Careful handling during harvest and postharvest is necessary because the thin-skinned tubers bruise easily and are subject to rapid desiccation. Harvested tubers can be stored for several months at 0–2 °C and at 95% relative humidity.

Parsnip: *Pastinaca saliva* l./P. *sylvestris* (2n = 22)

Parsnip belongs to the family Umbellifereae and is native to the Mediterranean region. It was introduced to the West Indies in 1564 but grows well in the tropics, but only at high altitudes. The Romans utilized the wild form for medicine and food.

Parsnips have a sweet taste and a slight mucilaginous texture. The edible portion is a long, tapered,

fleshy axis formed by hypocotyls. Large roots have an excellent, sweet flavor, and can be appreciated better when they are cooked and served cold. Small miniparsnips can be grated raw and served in salad. Smooth-skinned 'Lancer' is a recommended Miniparsnip. Roots have been reported to contain three phytotoxic, mutagenic, and photocarcinogenic furocoumarins that persist through normal cooking. They will melanize skin and have been used to treat depigmented skin and also psoriasis. Workers handling parsnips have frequently reported dermatitis.

Radish: *Raphanus sativus* L. ssp. Salivus (2n = 18)

The radish belongs to the Crucifereae (Brassicaceae), mustard family. It is a popular salad crop for home gardening and fresh market (Table 7). Raphanus is a latinized form of the Greek rephanos, which means 'easily reared.' Radish was a common food crop in Egypt. The eastern Mediterranean region, China, and middle Asia are considered to be the origin of radish. China is the center of origin because wild radish still exists in China. The ancient Greeks prized radishes so much that they made small replicas of them in gold. In 1544, a German botanist reported seeing radishes that weighed 100 lb (45 kg). Radishes were common in England in 1586 and were among the first European crops introduced to America by the Spaniards. By 1629, they were being cultivated in Massachusetts.

Table 7 Classification of radish cultivars depending on the season

Summer radish (approximately 25 days to maturity)	*Cultivar*
Round	
Red	Cavelier, Cherry-Beauty, Comet, Sparkler, Early Scarlet
Red/white	Sparkler
White/yellow	Golden of Globe, Snowboll
Half-long	French Breakfast
Red/white	Icicle
White	
Long-rooted	Cinnicinnati Market, white Icicle, long Scarlet, Short Top, Strasburg, Stuttgart, white Vienna
All-season radish (approximately 45 days to maturity)	Summer cross hybrid
Winter radish (approximately 60 days to maturity)	
Round	
Black	Round Black Spanish
Elongated	
Red	China Rose
White	White Chinese

The varieties evolved in India can broadly be categorized into two groups: Asiatic (tropical) or annual types and European (temperate) or biennial types. Several Asiatic or tropical varieties such as Japanese White, Pusa Desi, Pusa Chetki, and Pusa Reshmi are grown extensively in India. The Indian Agricultural Research Institute (IARI), New Delhi has released Pusa Himani as an improved variety of radish, having white skin, crisp flesh, and sweet flavor. Cultivar CO-1 has recently been released by Tamilnadu Agricultural University, Coimbatore.

The pink color of the radish roots is due to the presence of anthocyanin pigment. The characteristic pungent flavor and taste of radish roots are due to the high content of volatile alkaloid isothiocyanate (*trans*-4-methyl thiobutenyl isothiocyanate: MTB-ITC). For salad purposes, selection is towards low pungency.

Although radish is consumed worldwide, it contributes little to nutrition. Radish is low in calories and is a good source of vitamin C (15–40 mg $100\,g^{-1}$ fresh wt). One serving size of 85 g has only 15 cal and this will provide 30% of the RDA of vitamin C and 20% of the RDA for calcium. Radish contains glucose as the major sugar and smaller quantities of sucrose and fructose. Pectin and pentosans are also reported to be present. Pink-skinned radishes are generally richer in ascorbic acid than white-skinned ones. Radish leaves are a good source of extraction of protein on a commercial scale. The seeds are a potential source of nondrying fatty oil suitable for soap maturing.

Radish roots are eaten raw as salad and as a side item and cooked as vegetables. The very characteristic flavor is popular in Japan, the Philippines, and Hawaii, and roots are used to prepare food products such as takuwan and cabaizuku. It increases appetite, produces cooling effects, and prevents constipation. Radish roots are considered to be good for patients suffering from liver trouble, gallbladder ailments, hemorrhoids (piles), jaundice, and enlarged spleen. The seeds are said to be carminative, diuretic, expectorant, and peptic. The seed oil is used to make soap.

The edible roots become ready for harvesting in about 25–60 days depending on the variety. Radishes grown for the home garden are pulled by hand and laid in bunches of 6–12 in the field and washed to remove soil. They are usually packed in baskets or crates and iced for transport to market. The tops of winter radish are removed before they are stored, Mechanical harvesters are used for commercially grown crops in the USA.

Salsify: *Tragopogon porrifolius*

Salsify is a biennial plant, belonging to the family Asteraceae and native to Europe, North Africa, and Asia. From the appearance of its foliage, the plant acquired the name of 'goat's beard' and from the taste of the root 'vegetable oyster.' It is grown for its long, tapering roots, frequently branched, which are used for flavoring. The long, grass-like narrow leaves form a rosette and the plant can reach 1.2 m (4 ft) in height, with purple daisy-like flowers. Salsify is very hardy and the root may be harvested well into the winter with no damage. Little salsify is grown commercially in the USA. Roots are ready for use in mid-fall to late spring, and are lifted and stored for winter in boxes of sand.

Salsify contains the highest amount of protein (2.9%), carbohydrates (18.0%), phosphorus (66 mg $100\,g^{-1}$), and potassium (380 mg $100\,g^{-1}$) among root salad crops (beet, carrot, radish, turnip, rutabaga, and parsnips).

Roots, leafy shoots (known as chards), buds, and open flowers can be used as raw salad. Salsify is cooked with its peel on and after cooking the peel is removed and the vegetable is served hot or cold with dressing. Flowers as used raw or cooked.

Squawroot: *Perideridia gairdneri* Mathias

Other names: *yampah, epos* root.

The squawroot is perennial, and belongs to the Apiaceae family. It is grown wild in the western USA (southern Oregon, northern California) and British Columbia. Enlarged storage roots are collected in the summer and fall by local people and are eaten raw or cooked. They are also dried for use during summer. Epos was served as food for natives and early settlers in these regions. The starchy dark-skinned roots resemble nuts with a white flesh.

Turnip and Rutabaga: *Brassica rapa* L. ssp. Rapa ($2n = 20$); *B. napus* L. ssp. Napobrassica ($2n = 38$)

These are biennial herbs, belonging to the family Cruciferae, and are cultivated as annual root crops for both animal and human consumption. Rutabaga is evolved as a cross of cabbage (18 chromosomes) with turnip (20 chromosomes), followed by spontaneous chromosome doubling. Turnip and rutabaga have been grown for nearly 4000 years and were known by the Greeks and Romans at the beginning of the Christian era. Their original home is considered to be temperate Europe. They reached Mexico in 1586, Virginia in 1610, and New England in 1628. The early European settlers introduced them to North America in the seventeenth century.

Although both are similar in flavor, most common cultivars can be distinguished by their foliage appearance and root shape, size, and color. The enlarged fleshy storage organ in turnip and rutabaga varies in form, and these may be round, cylindrical, or flattened globes. The exterior may be white, green bronze, or purple color and the interior is white or light yellow. Their nutritional values also differ. Turnip generally has hairy leaves and a small white flesh root, whereas rutabaga has smooth, large leaves with a large, yellow-fleshed root. Turnip is a short-season crop; rutabaga requires a full season.

Purple-Top White Glove and Seven Top are the important cultivars of turnip whereas American Purple-Top and Laurentian are the important cultivars of rutabaga.

A cup of turnip greens provides 165% of the RDA of vitamin C, 182% of vitamin A, one-sixth of iron and thiamin RDA and one-fifth of riboflavin.

Turnips are used in roasts, stews, soups, and casseroles and as a boiled vegetable or sliced in salads. They may be eaten wholes or halved or mashed and served with a variety of sauces. When young and tender, turnip greens are cooked and eaten.

The flavor of rutabaga makes it suitable not only as a cooked vegetable but also in soups, salads, and baked goods. Rutabaga may be boiled or roasted with meat and eaten in pieces or boiled and mashed. Fresh uncooked rutabaga can be shredded into salads.

Rutabagas have much more vitamin A than turnips, considerably more ascorbic acid and more of the other vitamins, but less iron. This comparison refers only to the roots.

Turnip and rutabaga are of best-quality roots (mild and tender) when they are medium-sized, are commonly packed in transparent film bags, and require storage temperatures of 0–2 °C (32–35 °F) and relative humidity of 90–95%.

Yacon: *Polymnia sonchifolia* Poepp. and Endl.

Other names: jiquama, jacon, apple of the earth.

The yacon is perennial, belonging to the Asteraceae family, and is cold-tolerant; overground portions are damaged by frost and reestablish from the underground portion. The crop is grown as an annual for its foliage and as a perennial for the underground tuberous roots. Wild forms of yacon are found in the Andean regions of Colombia, Ecuador, and Peru. Plant height is approximately 1.5 m. Stems are hairy and green with purple markings. Lower auxiliary stems bear yellow or orange flowers. After flowering, the tops become senescent and die. Usually four or more developing tuberous roots form clusters that can be separated. Shapes vary from spindle to round. The thin skins are tan or purplish brown color.

The flesh can be white, pale yellow, or orange and is of varying sweetness.

Moisture and protein ranges are 70–85% and 6–7%, respectively, in roots. Stems and leaves are often fed to animals and can be used as cooked vegetables or eaten raw. Roots are high in potassium. Carbohydrate is stored as inulin, a polymer of fructose, and 65% dry weight is sugar.

Before use, skin must be removed because it is resinous. The crisp-textured tuberous roots are juicy and frequently eaten raw in snacks and salads, and also boiled or baked. The tubers are a rich source of fructose.

Yacon production is more than $20\,t\,ha^{-1}$ and is stored well at 5 °C and in the dark for a few weeks.

Yam beans/jicama: *Pachyrhizus erosus* L. Urban

Other names: sincamas or sinkamas (Philippines), fanko and sargott (China), dolique bulbeus (France), jimaca (Spanish), xicamatl (Nahuatl Indian).

Yam bean is a perennial, belonging to the Fabaceae (leguminoseae) family. It is a native of tropical America, from Mexico to northern South America. It is widely grown for its tuberous roots in these regions, and in the Philippines, and South China and is a domestic crop in Bolivia, northern Argentina, and the Amazon river headwaters.

Jicama (yam bean) is a herbaceous climbing vine. Leaves are entire or lobed rhomboidal shape. The tuberous root shape can be a flattened sphere, much like a turnip, elongated, spindle-like, and lobed. The skin is colored tan to light brown and flesh is white, has a watery crisp texture and sweet taste. In Mexico, two major types are grown:

1. Jicama de leche: this type has a dark skin, slender spindle-shaped root, and a milky taste.
2. Jicama de aqua: this type has a light skin, turnip-shaped root, is very succulent, and has a sweet watery taste.

The young brown-skinned and white-fleshed roots are sweet and watery and often eaten raw as a snack food and appreciated for their mild flavor and succulent crispy texture. Roots are harvested by hand-digging, yielding $15-45\,t\,ha^{-1}$ and can be stored at 13–15 °C for more than a month.

Bulbs

It is monocotyledonous, assigned to Amaryllideceae (Liliaceae), a distinct family of Alliaceae and has bulb or corms, belonging to the *Allium* genus. Scaly leaves are the underground edible portion and are valued for their pungency.

The genus *Allium* contains about 500 widely distributed species. The most important cultivated species are onion (*A. cepa* L.), leek (*A. ampeloprasum* L.), Japanese bunching of onion (*A. fistulosum* L.), rakkyo (*A. chinesensis* G. Don), Chinese chive (*A. tuberosum* Rott. Ex. Spr), chives (*A. schoenoprasum* L.), and shallots (*A. cepa* L.).

These crops are native to the temperate regions of South Africa and the Mediterranean. The alliums have been cultivated for food, medicine, or religious purposes and spices. The best-known feature of the alliums is their characteristic smell and taste.

The remarkable sulfur-containing compounds they contain characterize alliums, giving them their distinctive smell and pungency. In alliums most of the sulfur is in the form of various nonprotein amino acids, including the precursors (*S*-methyl, *S*-propyl, *S*-(1-propenyl) and *S*-allyl) of the volatile flavor compounds. These precursors are odorless, nonvolatile amino acids of the general name 5-alk (en) yl cysteine sulfoxides. When fresh tissue is damaged, flavor precursors react under the control of the enzyme allinase (*S*-alk (en) yl-L-cysteine sulfoxide lyase) to release the highly reactive sulfenic acids plus ammonia and pyruvate. Allium crops are almost totally lacking in starch but store carbohydrates as fractans (polymers of fructose), including sucrose, glucose, and fructose.

Garlic: *Allium sativum* L. ($2n = 16$)

Garlic is a cool-season, bulbous perennial plant of the Alliaceae family. The name comes from the Welsh word garleg, transformed into the English word 'garlic.' Carl Linnaeus gave the name *Allium sativum* to garlic in 1754. The Romans disliked the strong flavor of garlic but they fed it to their laborers and soldiers for strength and courage. In the world of the occult, garlic is believed to be a protection against known and unknown evil. Garlic wreaths hung on an outside door are said to ward off witches.

Garlic is native to middle Asia and the Mediterranean region. Central Asia, in particular the high mountain plateaus of Pamir and Tian Shan, is the original home of garlic. Garlic culture appeared considerably later in Europe and the rest of Asia, and was introduced to North and South America by the Spanish and Portuguese at the beginning of the fifteenth century. Anglo-Saxon settlers cultivated garlic in Australia, neighboring islands, and in Central and Southern Africa at the beginning of the nineteenth century. Garlic was used in Egypt before 2000 BC and in China and India for more than 1000 years. With the increased consumption of garlic in almost all countries, both home and international trade have developed rapidly. China is the highest producer of garlic,

followed by South Korea, Spain, and India. World production of garlic in 1990 was estimated to be 3 012 000 metric tons. Asia leads the world in garlic production, with more than 82% of total world supply, with China producing 63% of world supply.

The biggest exporters of garlic are Spain, Italy, France, and the USA. Fresh garlic and tinned foods are the target of much interest. Garlic for the fresh market was $261.5 million in 1997. Per capita consumption of garlic was 94 g (2.1 lb) per person in 1996. Most of the 37 000 acre of garlic grown in the USA is in California. The Republic of Korea consumes the most garlic.

The garlic bulb is composed of a disk-like stem, thin dry scales, which are the bases of foliage leaves, and smaller bulblets or cloves (usually 10 or more). The edible part of garlic is termed the clove; this is a small bulb that, together with other cloves, forms an aggregate or cluster covered with the outermost leaf, a white or purplish papery sheath. Cloves are formed from auxiliary buds of younger foliage leaves and enclosed by dry outer scales. Each clove consists of a protective cylindrical sheet, a single thickened storage leaf sheath and a small control bud. Garlic leaves have solid, thin blades.

Cultivars There are two different types of garlic: those that send up a seed stalk (hard-neck cultivars) and those that do not (soft-neck cultivars).

1. Rocabole type (hard neck):
   a. Spanish roja produces 6–13 cloves per bulb and cloves are brown to reddish purple and are easy to peel.
   b. Carpathian produces 6–10 cloves per bulb. The bulbs are large and uniform with bulb wrappers that have purple blotches. Carpathian has a hot and spicy flavor.
   c. German Red produces 10–15 cloves per bulb. It is very vigorous and has a deep green large bulb. The cloves are light brown with some purple at the base.
2. Artichoke type (soft neck):
   a. Inchellium Red produces 4–5 clove layers with 8–22 cloves. The bulbs can be over 7 cm in diameter and have a mild, lingering flavor.
   b. California Early produces four clove layers with 10–22 cloves. The cloves can have a tan to off-white color with pinkish blush. The flavor is mild and slightly sweet.
3. Silver Skin types (soft neck):
   a. Mild French produces four clove layers with 13–16 cloves. The color of the cloves is reddish pink blush on a yellow-white background and it is better adapted to hot climate. It has a sharp taste when raw but has a smooth, nutty taste when cooked.
   b. Silver Skin produces 15–20 cloves, usually in five layers. The clove color is off-white to tan with pink blush. It is a good producer of large bulbs that taste mild and sweet.

Two distinct cultivars, Fawari and Royalle Gaddi, are reported from India. Odeskii-13 and Imeruli-23 are grown in CIS. The principal varieties grown in the USA for drying are California Late (Pink and Italian) and California Early (White or Mexican and Creole).

Garlic is largely used fresh and stored as a condiment and a flavoring agent in soups, stews, tomato combinations, spaghetti, sausages, pickles, and salads. Spray-dried garlic products are also available. Garlic is used to enhance the smell and flavor of salted meat and fish. Dried garlic in powdered or granulated form is sometimes used in place of fresh bulbs. A number of garlic products, including capsules, extracts, garlic oil, and tablets, are now marketed and are odor-free. Garlic salt has much wider culinary potential than garlic powder. One tablespoon is equivalent to a clove of fresh garlic. Garlic butter has many uses in home and restaurant cooking and many restaurants offer garlic bread with meals.

Garlic has been used medicinally and there is some belief in its mythical properties. Garlic is a gastric stimulant, and helps in the digestion and absorption of food. Allicin has a hypocholesterolemic action and blood sugar-lowering effect. Garlic possesses antibacterial, insecticidal, and insect repellent properties, as well as nematicidal and larvicidal properties. Ancient folk medicine says that a cold can be cured if a person rubs the soles of the feet with cut garlic cloves or if a tonic of honey and garlic is drunk. Garlic preparations are given in whooping cough, lung disease, stomach complaints (such as to heal intenstinal ulcers) and disorders resulting from childbirth and specifically to treat for sore eyes and earache.

Garlic is harvested when the foliage collapses and starts to senesce. In a warm, dry climate, the bulbs may be left to dry and cure indoors or in windows in the field for about 1–3 weeks, using the foliage to protect the bulbs from sun injury. After drying, plants are tied in small bundles and then stored hanging on bamboo sticks or ropes. Trimmed bulbs can be stored in 2-m-deep bins or stacks provided with forced ventilation, as well as in shallow containers or sacks. Cold air (down to 0 °C) of low relative humidity is best for ventilating the stacks. Garlic cloves are stored best at 1 °C. Garlic cloves sprout most rapidly at 4–18 °C.

Leek: *Allium ampeloprasum* (Porrum group), (2*n* = 4x = 32)

The leek is herbaceous, biennial, and belongs to the family Alliaceae. Leak probably originated in the eastern Mediterranean area and spread to Europe by the Middle Ages; it was introduced to North American by European settlers. Leek are popular year-round as a salad vegetable in northern and western Europe and China. France is the largest producer of leeks in the world. Leeks are especially important in northern European countries such as Belgium, Denmark, and The Netherlands. Wild forms extend into western and southern Russia.

Leek is a tall, hardy robust herbaceous biennial with white, narrowly ovoid bulb and broad flat leaves. It resembles a green onion but is larger and, unlike onion and garlic, it does not form bulbs or cloves. Leeks are made up of sheaths of basal leaves that can be 15–25 cm long and 5 cm in diameter. The leaf base of leek stores some reserve, and has a milder and more delicate flavor than onion. When tender, it is eaten raw; is also cooked with other vegetables or used as a flavoring in soups and stews.

Cultivars Leek varieties have been classified into a number of major types that differ in pseudostem length and slenderness, leaf color, winter hardiness, and the tendency for bulbiness. Types are:

1. Bulgaarse Reuzen cv. Longa: very long, high-yielding leeks with pale green leaves and little frost resistance.
2. Danish cv. King Richard: long, high-yielding types with pale to mid-green leaves, suitable for late summer production and frost-susceptible.
3. Franse Zomer Selections cv. Pancho: long leek with mid-green leaves, limited frost resistance, and a tendency to form bulb.
4. Swiss Giant Selections cv. Tilina: medium to long leaves, with pale to mid-green, rather spreading leaves, frost-resistant.
5. Blauwgroene Herfst cv. Verina: medium shank length, dark leaf varieties, tendency to form bulb.
6. Autumn Mammoth selection cv. Goliath: medium shank length, mid-green, spreading leaves, little bulbing, very frost-hardy, high-yielding.
7. Giant winter cv. Winterreuzen: variable in appearance between varieties; all have good frost resistance.
8. Blauwgroene Winter cv. Bleu Star: short leeks with dark leaves and a high degree of bulbiness, very hardy, low-yielding.

The cultivars listed by American seedmen include London Flag, American Flag, Elephant or Monstrous, Carentan, Giant Nussenburgh, and Lyon. Acadia, Derrick, and Electra are suitable for fall cultivation in Belgium, while Alberts, Brizzard, and Carina are better for winter. Longa, Odin, Kilima, Goliath, Cyberia, and Artico are promising cultivars in Canada. Highest yield of leeks was obtained from King Richard, Verina, Tivi, Kazan, Kilima, and Albana.

Commercially important leek cultivars include the following:

- Carina has heavy thick, long white shanks and is good for overwintering.
- Titan is an extra long, early type that has vigorous growth. The leaves are dark green with white stems.
- Alaska has dark blue-green foliage and is tolerant to subfreezing temperatures.
- Pancho is considered an early leek, has thick white shanks, and resists bulbing.
- American Flag is very hardy and cold-resistant, and has large, thick shanks; it blanches to clear white with blue-green leaves.

Slender minileeks, which are very tender, may be eaten raw or cooked. The sharp flavor of leeks disappears upon boiling. They are used extensively in soups and stews because of their mild flavor and mucilaginous character.

Compared to onion, leek contains more protein and minerals. The energy value of 100 g of edible portion of leek is also higher than that of onion.

Leeks are harvested when their stalks reach a size of 2.5 cm or more in diameter and are tenderer. Hydrocooling, icing, or vacuum cooling to preserve the freshness of harvested leeks has been adopted. Leeks can be stored at near 0 °C (32 °F) and about 90% relative humidity for 2–3 months.

Onion: *Allium cepa* (2*n* = 2x = 16)

The onion is a biennial herbaceous, belonging to the Alliaceae family, and is typically grown for bulbs as annuals. Onions are native to southern Asia and the Ural mountains and have long been valued in China and India for their pungent flavor. The onion is an essential ingredient of the cuisine of many regions; it was the staple food of workers at the pyramids in Egypt; it has been worshipped since before the Christian era; and it has also been used in ancient Druid rites. The onion is the oldest salad crop known to humans.

Onions have been grown in England for centuries. The onion was introduced into the West Indies by the Spanish and spread to all parts of America, including California, Oregon, Colorado, Texas, and Georgia. China, India, CIS, Turkey, Japan, Pakistan, Mexico,

Egypt and Chile and Spain are the major producers and exporters of onion, whereas Malaysia, United Arab Republic, Canada, Japan, Lebanon and Kuwait are the major importers.

India, Indonesia, the former Soviet Union and the USA are the leading producers of onion. Onion is a major commodity of international trade. Imports in 1987 by Germany, France, the UK, and the USA were valued at US$265 m. Annually, the US onion crop is the third most valuable commercial vegetable ($426 m in 1988). Onion production for fresh market and processing increased from $581.5 m to $648.4 m in 1997.

The size, shape, and color of skin and flesh, single centers, skin retention, firmness, dormancy, pungency, and concentration of soluble solids are important quality traits for bulb production. Onions are becoming the mainstay of our diets. The bulbs may be white, red, brown, or yellow; round, piriform, or flat, soft, or firm. Dormancy is important because onions are often stored and marketed when prices are higher. Higher soluble solids are important to the dehydrating industry, producing onion chips or powder, which is used as seasoning.

Consumers in the USA prefer a less pungent onion, whereas many tropical onions are prized for their strongly pungent flesh.

The edible part of the onion is the bulbs, which are composed of concentric, fleshy, enlarged leaf bases or scales. The outer leaf bases lose moisture and became scaly. The inner leaves generally thicken as the bulb develops.

Table 8 lists important onion cultivars, and Table 9 gives a classification.

Starlite, Sunlite, BRI, Buffalo, Utalian Red Torpedo, California, Early Red, and Sweet Sandwich are important onion cultivars. Kyoto market

Table 8 Important onion cultivars

Country	*Cultivars*
India	Pusa Red, Pusa Ratnar, Nasik Red, Patna Red, N404, N207-1, white Patna, N-53, N2-4-1, Bellary Red, Bellary Big Onion, B-780, Udaipur 101, Udaipur 103, Hisar-2, Hisar-II, Kalyanpur Red Round, Arka Pragati, Arka Niketan, Arka Kalyan, Bangalore Rose and Punjab 48, Udaipur 102, Pusa White Round, Pusa White Flat, Phule Safed, Phule Suwama, Basawant 780
Pakistan	Phulkara, Faisalabad, Early and Desi Red, Local White (Kasmir)
Bangladesh	Faridpuri Vati, Taherpuri
Egypt	Giza6, Beherl
Israel	Haemek, Moab
Spain	Babosa, Valenciana, Temprana
South Africa	De Wildt, Pyramid, Bon Accord, Hojein
Mexico	Cojumatlah
Brazil	Baia Performe, Grano 502, Texas Early Grand
USA	Brigham Yellow Globe, Yellow Globe Danvers, Early Yellow Globe, Mountains Danvers, Ebenzer, Red Wethersfield, South Port Red Globe, California, Early Red, South Port White Globe, White Cresole, White Portugal or Silverstein, Yellow Bermuda, Crystal Wax, Early Grano, Yellow Sweet Spanish, White Balbosa, White Sweet Spanish, Barletta, Beltsville, Bunching and White Lisbon

Table 9 Classification of onion cultivars

Day length	*Scale color*	*Pungency*	*Representative cultivars*	*Characteristics*
Short (European type)	Brown	Sweet	Awahia	Very mild, soft-fleshed, and unsuitable for storage
	Red	Sweet	Red Granex	
	Red	Pungent	Red Creole	
	White	Sweet	White Granex, Crystal Wax	
	Yellow	Sweet	Grans, Granex	
	Yellow	Pungent	Yellow Croele	
Intermediate	Brown	Moderate	Cochise Brown	Relatively soft-fleshed and primarily used for the fresh trade
	Red	Moderate	Stockton Early Red	
	White	Moderate	Fresno White	
	Yellow	Moderate	Riaeto	
Long (American type)	Brown	Pungent	Australian Brown	Very pungent, hard, and store well
	Red	Pungent	Carmen, South Port, Red Globe	
	White	Pungent	White Lisbon, Ivory	
	Yellow	Sweet	Fiesta, Sweet Spanish	
	Yellow	Pungent	Autumn Spice, Downing Yellow Globe	

(multistem), Ishikuro, and White Evergreen are the best known and Paris Silver Skin, Barletta, Brunswick (Reddish Brown), and Purplette (Purple) are important traditional cultivars.

Bulb onions may be sliced and used in sandwiches and tried as onion rings or blooming onions. They are used at both immature (spring onion: bulbs are rudimentary) and mature bulb stages as a vegetable and spice as well as food for cattle and poultry. Onion can be eaten raw or cooked. Mild-flavored, colorful bulbs are often chosen for salads. Onions have many uses as folk remedies and recent reports suggest that onions play a part in preventing heart disease and other ailments. In many countries pickled *Allium* bulbs are eaten in large amounts. Onion is liked by the rich and the poor alike as a vegetable and for flavoring and pickling. Onions for pickling are either common yellow-brown-skinned varieties or special white Silver Skin types.

1-propenyl-sulfenic acid is produced on cutting; it spontaneously rearranges its chemical structure to form the tear-inducing thiopropanal *S*-oxide. Onion bulbs boiled whole, without previously damaging the tissue, lack flavor, since the enzyme will be destroyed before it has access to the flavor precursors.

Onions are generally dried and pickled. In some parts of West Africa, fermented preparations are made from crushed onion leaves and tops. These products are used to flavor food at times when onion is not available. Onion scales may be sun-dried for this purpose. Commercially prepared onion products include dried flakes and powder, usually made from white cultivars, with high dry-matter contents and high pungency. Onion oil has a flavor strength 500 times and onion juice 10 times that of dried products.

Canned and bottled onion is used in industries in North America. Onion rings are common products in fast-food industries. Pickled onions are popular in the UK: onions covered in vinegar and pasteurized are used. Onions are said to be stimulant, diuretic, and expectorant. Onion oil is used to treat stomach ulcers, eye disorders, gastrointestinal disturbances (flatulence), high blood pressure, malarial fevers, scorpion bites, and intestinal worms. The onion is known to possess insecticidal, antibacterial, antifungal, anti-tumor, hypoglycemic, hypolipidemic, and antiatherosclerotic properties. This has been attributed to its sulfur-containing compounds.

Onion bulbs are harvested when the tops fall over and begin to turn yellow. The ideal diameter of an onion bulb is 5–10 cm. Onions are dug and allowed to dry out in the open sun for a few days to toughen the skin. The dried soil on the bulb is removed by gentle brushing. The stems are cut, leaving 5–7 cm attached to the bulb. Onion bulbs are thoroughly covered before being stored in crates; these crates, filled with topped bulbs, are usually stacked in the field or in open curing sheds, and covered with boards, and roofing paper to protect the onions from injury by sun and rain. For fresh usage, a temperature of 0 °C and relative humidity of about 65–75% for storing are recommended. Onions can also be stored at high temperatures of 29.2–34.7 °C: 3–5% O_2 plus 10% CO_2 is superior to storage in air at 4.4 °C.

Shallot: *Allium cepa* (Aggregatum group)

A multiplier onion is an onion-like plant, which originated in western Asia, and is popular in southern USA and some European countries – France in particular. In the tropics, shallots are often grown in areas where onion culture is difficult because the climate is humid. The shallot has a very short growing season of only 2 or 3 months. The term shallot refers to the vegetatively propagated forms of *Allium cepa* var. ascalonicum. Bulbs usually propagate the crop. A single plant produces a cluster of from two to more than 15 distinct small bulbs at the base. Shallots ripen earlier than onion grown from seeds and generally give a lower yield. Shallots are reasonably hardy; they have a distinctively flavored onion-like bulb, and are a good substitute for onion as they are easily grown and fast-maturing, and most cultivars store for an exceptionally long time, much longer than onions. The shallot grows well where temperatures are too high for bulb onions over 30 °C (86 °F). They are normally roundish in shape, but there are some long and banana-shaped forms. The flesh is reddish or creamy (yellow). Red is considered to be better-flavored but it is smaller, slower-growing, and may not store as well.

Varieties/cultivars There are number of different forms of shallots:

1. The Old Shallot: bulbs are small, about the size of a walnut, 2.5 cm in diameter, 2.75 cm high and with an irregular pyriform shape. When ripe, the outer skin is silvery green or brownish, and the inner scales are tinged with purple or brown. The bulbs are produced in clusters of 5–8 or 10. Leaves are about 30 cm tall and bright green in color. This is the earliest variety, and it keeps well and is rarely cultivated.
2. Large Brown: bulbs are about 5 cm in diameter, 6 cm high, the outer skin is reddish brown, and the inner skin is tinged deep purple and fleshy. Leaves are 45 cm long, erect, and deep green.
3. Jersey Shallot: bulbs are fairly large, uneven, and irregular, often divided into side bulb and divided crowns.

4. Yellow or Dutch shallots: bulbs are round or slightly flattened, scales are greenish before drying, and straw-colored when ripe.

Ampenan, Cloja, Bima, Bima Kuning, Bauji, Balizo, Suminep, Bawang Lambung, Betawi Cipenos, and Hejakuning cultivars are reported from Indonesia.

Traditional cultivars Traditional cultivars include Long Keeping Yellow; Long Keeping Red, Hative de Niort, New Yellows. Sarite Atlantic Creation (grown for seed), New Reds: Pikant and Deli Cato.

Shallots are used in salads and cooking. Used as bulb onions, they are superb pickled. Shallots may contain more fat (Suminep) and soluble solids. Shallots are also used in certain sauces. In Ethiopia small local red shallots are highly valued in a traditional sauce to accompany injera bread made from wheat flour.

Shallots are pulled by hand, when 70–80% of the leaves have turned yellow. The outer skin is peeled off and the roots are trimmed after washing, and they are tied into 1.5-kg bunches. Shallot bulbs are dried for 5–14 days in the field and stored in warm conditions (30 °C or 86 °F).

See also: **Onions and Related Crops**; **Vegetables of Temperate Climates**: Oriental Brassicas; Carrot, Parsnip, and Beetroot; Swede, Turnip, and Radish

Further Reading

Bremster JL (1994) *Onion and other Vegetable Alliums*. Wallingford, UK: CAB International.

Decoteau DR (2000) *Vegetable Crops*. New Jersey: Prentice-Hall.

FAO Production Yearbook (1990) Rome: Food and Agriculture Organization.

FAO Production Yearbook (statistic series) (1993) Rome: Food and Agriculture Organization.

Kalloo G and Bergh BO (eds) (1993) *Genetic Improvement of Vegetable Crops*. Oxford: Pergamon Press.

Lancaster JE and Boland MJ (1990) Flavour biochemistry. In: Rabinowitch HD and Brewster JL (eds) *Onions and Allied Crops*, vol. 3, pp. 32–72. Boca Raton, Florida: CRC Press.

Larkcom J (1995) *Salads for Small Gardens*. London: Hamlyn.

Larkcom J (1995) *Vegetables for Small Gardens*. London: Hamlyn.

Leshey Tiersa LAC (1992) *The Herbs of Life, Health and Healing Using Western and Chinese Techniques*. Indian Medical Series no. 65. Delhi, India: Sri Satguru Publications.

Pierce C (1987) *Vegetables: Characteristics. Production and Marketing*. New York: John Wiley.

Rubatzky VE and Yamaguchi M (1997) *World Vegetables. Principles, Production and Nutritive Values*. New York: Chapman & Hall.

Salunkhe DK and Desai BB (eds) (1984) Onion and garlic. In: *Post Harvest Biotechnology and Vegetables*, Part II, pp. 23–38. Boca Raton, Florida: CRC Press.

Salunkhe DK and Kadam SS (eds) (1998) *Handbook of Vegetable Science and Technology*. New York: Marcel Dekker.

Wien HC (ed.) (1997) *The Physiology of Vegetable Crops*. Wallingford, UK: CAB International.

Yamaguchi M (1984) *World Vegetables: Principles, Production and Nutritive Values*. Westport, Connecticut: AVI.

Salad Dressings and Oils *See* **Dressings and Mayonnaise**: The Products and Their Manufacture; Chemistry of the Products

Salami *See* **Meat**: Sources; Structure; Slaughter; Preservation; Eating Quality; Sausages and Comminuted Products; Analysis; Nutritional Value; Hygiene; Extracts

Salmon *See* **Fish**: Introduction; Catching and Handling; Fish as Food; Demersal Species of Temperate Climates; Pelagic Species of Temperate Climates; Tuna and Tuna-like Fish of Tropical Climates; Demersal Species of Tropical Climates; Pelagic Species of Tropical Climates; Important Elasmobranch Species; Processing; Miscellaneous Fish Products; Spoilage of Seafood; Dietary Importance of Fish and Shellfish; **Fish Farming**; **Fish Meal**

SALMONELLA

Contents

Properties and Occurrence

C Wray, Greendale Veterinary Laboratories Ltd, Knaphill, Woking, UK

Background

The clinical and postmortem picture of typhoid fever was first accurately described by French workers in the early nineteenth century. The most striking contribution on the natural history of typhoid before the opening of the bacteriological era was the work of William Budd, who elucidated many of the epidemiological aspects of the disease. Ebert had observed in 1880 the typhoid bacillus in autopsy specimens from a patient who had died from typhoid and the organism was isolated in 1884 by Gaffky. Calf paratyphoid was first recorded as an endemic disease in northern Europe in the middle of the nineteenth century and in 1885 the organism now known as *Salmonella choleraesuis* was first isolated from pigs by Salmon and Smith, who considered the organism to be the cause of swine fever (hog cholera). *Salmonella* was first shown to be the cause of food poisoning by Gartner in Germany, who isolated an organism which he called *S. enteritidis* from the meat of a cow that had diarrhea and was slaughtered, and from the spleen of a man who had eaten about 800 g of meat from the cow and died a day later. Subsequently *Salmonella* were isolated from other species of animals and the new strains were initially named after the disease condition (e.g., *S. abortusovis, S. typhimurium*) and each was considered a different species. With the refinement of serological techniques for identification, the serological subdivisions have increased and the number of new *Salmonella* serovars continues to increase annually. Each newly recognized serovar is now customarily named after its place of isolation.

Taxonomy and Serological Classification

The genus *Salmonella* belongs to the bacterial family Enterobacteriaceae, whose definition is: straight rods, generally motile with peritrichous flagella, grow on nutrient agar, facultatively anaerobic, ferment glucose often with the production of gas, reduce nitrates to nitrites, and the oxidase test is negative. Some serovars have peculiarities, e.g., the avian serovar *S. gallinarum* is nonmotile and nonmotile variants of other serovars are occasionally observed. The classification of *Salmonella* has been controversial for many years, and according to the latest nomenclature, which reflects advances in *Salmonella* taxonomy, the genus consists of only two species: *S. enterica* and *S. bongori*. *S. enterica* is divided into six subspecies which are distinguishable by certain biochemical characteristics. These subspecies are:

Subspecies I = subspecies *enterica*
Subspecies II = subspecies *salamae*
Subspecies III = subspecies *arizonae*
Subspecies IV = subspecies *diarizonae*
Subspecies V = subspecies *houtenae*
Subspecies VI = subspecies *indica*

Thus the full correct nomenclature of an isolate would be, e.g., *Salmonella enterica* subspecies *enterica* serovar Enteritidis, rather than, as in the past, *Salmonella enteritidis*. This, however, is cumbersome and the concept that is currently in use at many laboratories is to write *Salmonella enteritidis*.

Most isolates of *Salmonella* from domestic animals belong to subspecies *enterica*, whose behavior is, with few exceptions, regular and may be classified as typical for the species. The principal biochemical tests by which *Salmonella* can be identified are given in Table 1. It should be pointed out however that variable test results may occur because, with a few exceptions, no test is 100% positive or negative.

Strains of *Salmonella* are classified into serovars by the characterization of their surface antigens, as described in the Kauffmann–White–Le Minor scheme. The identity, known as the *Salmonella* serovar, is designated by the identification of two sets of antigens. The somatic antigens, known as the O antigens, are present on the main body of the organism. They are polysaccharide antigens, identified by agglutination with specific antisera, usually prepared in

Table 1 Biochemical reactions of typical *Salmonella enterica* subsp. *enterica* strains

Test	*Reaction of typical strains*	*Test*	*Reaction of typical strains*
Motility	+	Fermentation of:	
Nitrate reduction	+	Glucose	+
Oxidase	−	Mannitol	+
Oxidation/fermentation	Fermentation	Maltose	+
Urea hydrolysis	−	Lactose	−
Indole production	−	Sucrose	−
H_2S production	+	Salicin	−
Citrate utilization	+	Adonitol	−
Sodium malonate	−	Dulicitol	+
Growth on potassium cyanide	−	Lysine decarboxylase	+
Methyl red	+	Arginine dihydrolase	+
Voges–Proskauer	−	Ornithine decarboxylase	+
Gelatin liquefaction	−	Deamination of phenylalanine	−
O-nitrophenyl-β-D-galactopyranoside	−		

rabbits, to give the serogroup, e.g., *S. typhimurium* is a group B *Salmonella* due to the presence of the O4 somatic antigen. Most *Salmonella* possess two phases of flagellar antigens (termed H), which are present on the protein flagellar structures that extend from the surface of the bacterium and such antigens are described as diphasic. Some serovars possess only one flagellar antigen and are described as monophasic. Examples of antigenic formulas are:

Group B (O4)	*S. agama*	O 4,12 : H i : 1,6	Diphasic
	S. typhimurium	O 1, 4, (5), 12 : H i : 1,2	Diphasic
Group C1 (O7)	*S. bonn*	O 6,7 : H 1,v : e,n,x	Diphasic
Group D1 (O9)	*S. enteritidis*	O 9,12 : H gm : –	Monophasic

Currently *c*. 2500 different *Salmonella* serovars have been identified, although only a limited number are encountered in practice. Some *Salmonella* serovars possess a capsular polysaccharide which is termed the Vi antigen. It is only of practical importance in *S. typhi* and *S. paratyphi*, when its presence may mask agglutination with somatic antiserum.

Salmonella also possess fimbriae, which are a family of polymeric proteinaceous surface organelles, but until recently little was understood about their variety and functions. Several different fimbrial types may be expressed during the organism's lifetime and it is believed that some may assist intestinal colonization and others assist their survival.

Further subdivision for epidemiological studies of the more frequent serovars may be obtained by phage typing, which evaluates the degree of susceptibility (or resistance) of isolates to a set of selected bacteriophages (viruses active against *Salmonella*). Such schemes are in routine use in many laboratories for the subdivision of *S. typhi, S. typhimurium*, and *S. enteritidis*, and other serovars of more local importance. Other methods used to show the relatedness of isolates during disease outbreaks include biotyping, antibiograms, and molecular genetic techniques to examine plasmid and chromosomal DNA. In recent years, many different molecular genetic techniques have been used to fingerprint *Salmonella* during disease outbreaks, but it should be pointed out that while one method may work with a particular serovar, it may not work with a different serovar.

Antibacterial Resistance

In general the prevalence of antibacterial resistance in *Salmonella* is low and to some extent relates to the serovar. However over the years different phage-types of *S. typhimurium* have become resistant to a range of antibacterials and caused epidemics in humans and other animals. Often the resistance genes are carried on mobile genetic elements, e.g., plasmids, and thus may be transmissible to other serovars and other bacterial species. Recently there has been an epidemic in many different countries of a multiple-resistant clone of *S. typhimurium* DT104 in which the resistance genes are incorporated into the chromosome.

Detection of *Salmonella*

Conventional isolation of *Salmonella* is accomplished by using cultural methods, which are often based on its phenotypic properties. *Salmonella* and other members of the Enterobacteriaceae are more resistant to novobiocin, selenite, tergitol, and bile salts, especially desoxycholate, than other bacteria. *Salmonella* are also more resistant than other Enterobacteriaceae to some dyes, e.g., brilliant green and malachite green; however, these properties are not sufficient for a true selective isolation and no medium is presently available that would allow the isolation of only *Salmonella* and no other bacteria. Because of this drawback, most differential media often include

lactose, sucrose, and salicin with a pH indicator, since most *Salmonella* fail to produce acid from these substrates. (*See* ***Salmonella***: Detection.)

Rapid Methods for the Detection of *Salmonella*

The time span that is necessary to obtain a negative result can thus take up to 96 h when the whole cultural method has to be applied and presumptive positive results may take up to 48–96 h. Many alternative methods have consequently been introduced in recent years to reduce analytical time and also save staff time and media requirements. These may include modification of conventional methods by including specific substances to media used either for isolation or confirmation or the development of tests to detect specific characteristics of *Salmonella*. Other rapid methods do not involve the isolation of the organism and only provide a presumptive result. This allows negative samples to be screened out, but positive samples still have to be investigated further for confirmation. Of these methods, techniques based on electrical conductance and impedance are in use in many laboratories. When microbial growth and metabolism take place in a culture medium, changes in conductance of the medium occur but the reliability of the method largely depends on the performance of the selective medium used for the assay. Immunological techniques comprise the largest group of alternative methods and many test kits are available. These techniques include immunodiffusion, enzyme-linked immunosorbent assays (ELISA), and the use of specific antibodies to 'capture and concentrate' the organism. Nucleic acid-based assays are also available, including the use of DNA probes to detect the organism and the polymerase chain reaction (PCR) to amplify a targeted region of the *Salmonella* genome. Rapid methods, however, still require up to 24 h and for many of the methods no worldwide standardization for the validation and approval of the technique is available.

Epidemiology

Although primarily intestinal parasites of human beings and other animals, *Salmonella* are widespread in the environment and commonly found in human sewage, farm effluents, and any material subject to fecal contamination. Salmonellosis has been recognized in all countries of the world, but appears to be most prevalent in areas of intensive animal production, especially in poultry and pigs. Although the disease may affect all species of domestic animal, young animals and pregnant animals are the most susceptible. Many animals may be infected but show no clinical signs. People are also highly susceptible to infection, either by direct contact with infected animals or through their products, as occurs in foodborne salmonellosis.

Host Range

While all *Salmonella* are considered to be potentially pathogenic, different serovars differ widely in their host range and pathogenic syndromes that they produce. Some serovars show a degree of host adaptation and primarily infect only one species of animal. Thus *S. typhi* and *S. paratyphi* A, B, and C usually infect only human beings and typhoid and paratyphoid are considered as human disease entities. *S. choleraesuis, S. abortusovis*, and *S. gallinarum/pullorum* occur in pigs, sheep, and poultry respectively, and *S. dublin* appears to have a predilection for cattle. In contrast, *S. typhimurium* affects all species of animals and is one of the most common causes of human food poisoning. Companion animals are a source of *Salmonella* infection; the serovars relate to their feed and the extent to which they scavenge. *Salmonella* have been isolated from wild animals and birds, reptiles, fishes, and insects.

Animal Infections

Most *Salmonella* infections in farm animals are likely to be acquired from the same species of animal, especially the host-adapted serovars. Purchased animals may acquire infection on their home premises, in transit, in markets, in collecting centers, or in lorries. Poultry may be infected vertically by egg-borne transmission of *Salmonella* when the contents or the shell may be contaminated or in the hatchery by cross-contamination. *Salmonella typhimurium* and other serovars may be introduced into herds and flocks by contaminated feedstuff. Many batches of animal protein, e.g., meat and bone meal, fishmeal, pelleted feeds, and vegetable proteins have been found to be contaminated with *Salmonella* and many new serovars have been introduced into countries in imported feed ingredients. Recontamination of feedstuffs after processing is a major problem and heat treatment, e.g., pelletization, and treatment with organic acids, e.g., propionic or formic acid, is used to produce a *Salmonella*-free product.

Infection may also be introduced by other domestic animals as well as by feral animals and birds. Rats and mice may acquire *Salmonella* infection; available evidence suggests that they do not play a major role in the spread of disease but that they may prolong the persistence of infection on a farm.

Environmental Considerations

Salmonella may survive for long periods in infected feces and slurry, where their survival depends on a

number of factors, especially the climatic conditions. Intensive livestock production has created problems of excreta disposal. The number of animals per unit area of land has vastly increased in many countries and it has been necessary to increase application rates of animal waste on to the land. One survey found that 10% of cattle slurry samples contained *Salmonella*, with survival periods ranging from 11 to 12 weeks.

Biological treatment of human sewage removes most of the suspended and dissolved organic matter, but it does not necessarily remove pathogenic bacteria. Where sewage sludge is used as a fertilizer, many samples have been found to be contaminated with *Salmonella*. There are numerous reports of the isolation of *Salmonella* from rivers and streams and of their pollution by sewage effluent and slurry runoff. Further spread from contaminated environments may also take place by wild animals and birds.

Transmission to Humans

Typhoid fever usually follows the ingestion of drinking water contaminated by human carriers of *S. typhi* or food prepared or handled by such carriers who do not take adequate hygienic precautions.

In contrast, acute gastroenteritis (foodborne disease/poisoning) is usually acquired from consumption of foods which may be directly or indirectly contaminated with *Salmonella*. A wide variety of animal species have been shown to be capable of harboring the organisms, and in the developed world turkeys, chickens, swine, and cattle have often been found to be infected in abattoir studies. Cross-contamination may readily occur during transport to the abattoir and during slaughter. Modern poultry-processing plants are designed with emphasis on speed, and during the slaughtering process there are many points where cross-contamination may occur. Recent surveys have found that up to 30% of chicken carcasses may be contaminated with *Salmonella*. Table eggs have been found to be an important vehicle for the transmission of *S. enteritidis* to humans and undercooked or raw eggs or dishes made from them, such as mayonnaise or mousse, have been responsible for many outbreaks of food poisoning.

The progressive trend toward mass processing and distribution of food products has been an important factor in the increased incidence of *Salmonella* foodborne disease. Milk has also been shown to be an important source of infection and has been associated with massive outbreaks of human salmonellosis. Most outbreaks have been associated with the consumption of raw milk, although pasteurization plant failure has also resulted in large outbreaks. Salads and other vegetables which have been contaminated in washing or 'freshening' with polluted water have also been shown to be associated with human salmonellosis. Many other products, e.g., beansprouts, chocolate, have been found to be the source of *Salmonella* infection and dried milk became contaminated with *Salmonella* when the production machinery was contaminated. Whatever the source, many outbreaks occur as a consequence of the delay between preparation/cooking and consumption of the food, which allows the numbers of *Salmonella* to increase rapidly. The important food-associated risk factors are inadequate cooking or reheating, storage at inappropriate temperatures, and cross-contamination from raw to cooked food.

Person-to-person spread has been demonstrated on many occasions and may be of importance in young children and groups living under poor socioeconomic conditions where effective sanitation is lacking. In these situations, flies may serve as vectors for *Salmonella* transmission. Dog and wild animal feces on footwear have been implicated as a route by which infection can be carried into households. Person-to-person spread may also occur in hospitals, nursing homes, and mental institutions, where a number of large outbreaks have occurred. Amplification of infection in these institutions may occur from contaminated food or from asymptomatic carriers; babies are at special risk. Likewise, contaminated equipment has been shown to be vehicles of infections.

Causation of Disease

Salmonella cause disease by invading the intestinal mucosa and multiplying in the gut-associated lymphoid tissue. From the infected tissue the organisms are drained to the regional lymph nodes, where further spread is usually prevented if the host defense mechanisms are successful. In humans nontyphoidal *Salmonella* serovars usually cause a localized disease which is manifest as an acute gastroenteritis. If, on the other hand, the host's defense mechanisms are unable to prevent the spread of the organism, systemic disease results, as commonly occurs with typhoid. The etiology of diarrheal disease is poorly understood, and although several putative toxins, e.g., enterotoxins, cytotoxins, have been described, their significance is far from clear. (*See* ***Salmonella***: Salmonellosis.)

Whether infection follows exposure to any pathogenic microorganisms depends in part on the number of organisms that enter the host's tissues. From volunteer studies in humans, it has been suggested that large inocula of *Salmonella* are necessary to cause illness. However, investigations of outbreaks of *Salmonella* foodborne disease have suggested that the

infective dose is often low; in six of the 11 outbreaks the actual dose ingested was calculated as $< 10^3$ organisms. However, many factors may affect the infective dose, not least of which are strain variations, e.g., S. *choleraesuis*, although a rare cause of human infection, may cause serious disease. Children and older people are more susceptible; other concurrent illnesses (e.g., acquired immune deficiency syndrome (AIDS)) may increase susceptibility. The use of antibiotics for other indications may predispose to foodborne salmonellosis, even when the *Salmonella* are sensitive to that antibiotic. There is a fivefold increase in risk when the *Salmonella* are resistant.

Prevention of Transmission

The overall quality of meat and edible byproducts depends heavily on their microbiological quality. This is very dependent on the contamination level of the production equipment, as well as on the skin and intestinal microflora of animals and on the technology and techniques used in slaughtering and processing. The fecal microflora of animals is the predominant source of pathogens on the final dressed carcass, and the modern slaughter line offers many risks of cross-contamination. In general, such cross-contamination is much less pronounced with carcasses of beef and sheep than with pork and poultry carcasses. A system to insure quality and safety must be designed for the entire chain of operations, from the farm to the consumer. This system must include the identification of potential microbiological problems and control measures at all stages of the entire meat production chain, as well as the development of Hazard Analysis Critical Control Point procedures. Although the hazards may differ from country to country and vary with animal health, slaughter procedures, type of meat products, and storage and distribution, the main areas of focus are the slaughter process itself and disinfection of tools and equipment.

Pasteurization will prevent milk-borne salmonellosis, although outbreaks have occurred on a number of occasions when faults have developed in plants. One of the most effective end-product treatments is gamma-irradiation, which can destroy *Salmonella* on poultry carcasses and prolong shelf-life, although possible changes in flavor and texture may occur. There may be some consumer resistance toward the use of nuclear energy.

Equally important is the education of caterers and domestic consumers in all aspects of hygienic preparation of food. Cross-contamination may occur readily in kitchens, especially when raw and cooked foods are prepared on the same surfaces.

Public Health Aspects

In many countries the results of tests which identify the presence of *Salmonella* in animals have to be reported to the appropriate government department and, in addition, separate orders may relate to animal feeds.

Likewise, on the human side, it may be necessary to report if a patient is suffering from or is suspected of having foodborne disease. Following the reporting of *Salmonella* incidents in both humans and animals, further investigations may be carried out. Regulations may also apply to the production and sale of food.

Economic and Social Impact

The financial costs of *Salmonella* infections are not only associated with investigations, treatment, and prevention of human illness; they may also affect the whole food production chain.

In the public sector, resources are necessary for surveillance, treatment of patients, and investigations of outbreaks. Financial burdens may be imposed on industry, especially the food industry. Likewise, the cost to affected individuals may be extremely high, especially should death result from *Salmonella* infections. The costs associated with Salmonellosis may be subdivided into tangible costs, which can be measured in monetary terms (e.g., medical costs and loss of work), and intangible costs, such as loss of leisure, pain, discomfort, etc., which cannot be readily costed. A study of infectious intestinal disease in England between 1993 and 1996 suggested that the true rate of salmonellosis in the community was 220/100 000 population and the estimated cost per human case was £660.

An analysis of five *Salmonella* outbreaks in North America gave direct costs which ranged from $36 400 to $62 million. In the UK, an outbreak of *S. ealing* infection associated with powdered milk was estimated to have cost £14.6 million in the late 1980s.

There have been few studies into the costs and benefits of preventing *Salmonella* infection, but it has been suggested that for every £1 spent on investigation and curtailment of an outbreak, there is a saving of £5.

See also: **Food Poisoning**: Classification; ***Salmonella***: Detection; Salmonellosis

Further Reading

Doyle MP (ed.) (1989) *Foodborne Bacterial Pathogens*. New York: Marcel Dekker.

Ewing WH (ed.) (1986) *Edwards and Ewing's Identification of Enterobacteriaceae*, 4th edn. New York: Elsevier.

Guinee PAM and van Leeuwen WY (1987) Phage typing of *Salmonella*. In: *Methods in Microbiology*, vol. 11, pp. 158–190. New York: Academic Press.

Le Minor L and Popoff MY (1987) Request for an opinion. Designation of *Salmonella enterica* sp. nov. as the type and only species of the genus *Salmonella*. *International Journal of Systemic Bacteria* 37: 465–468.

Reeves MW, Ewins GM, Heiba AA *et al*. (1989) Clonal nature of *Salmonella typhi* and its genetic relatedness to other salmonellae as shown by multilocus enzyme electrophoresis and proposal of *Salmonella bongori* comb. nov. *Journal of Clinical Microbiology*. 27: 313–320.

Saeed AM (1999) Salmonella enterica *Serovar* enteritidis *in Humans and Animals: Epidemiology, Pathogenesis and Control*. Ames, Iowa: Iowa State University Press.

Varnam AH and Evans MG (1991) *Foodborne Pathogens: An Illustrated Text*. London: Wolfe.

WHO Collaborating Centre for Reference and Research on *Salmonella* (1997) *Antigenic formulas of the* Salmonella *Serovars*. Paris: Institut Pasteur.

Wray C and Wray A (eds) (2000) Salmonella *in Domestic Animals*. Wallingford, Oxon: CABI Publishing.

Detection

T Humphrey, PHLS Food Microbiology Research Unit, Heavitree Exeter, Devon, UK
P Stephens, Oxoid Limited, Basingstoke, Hampshire, UK

Detection

In the developed world, the majority of cases of human salmonellosis is zoonotic and results from the colonization/infection of food animals and resultant contamination of products derived from them. *Salmonella* are also ubiquitous, and a diverse range of foodstuffs – including dried milk, fermented sausage, beansprouts, chocolate, and yeast-based flavoring – have been implicated as vehicles for human salmonellosis. In the UK, and elsewhere in the last decade, most cases or outbreaks of *Salmonella* infection have resulted from the consumption of contaminated eggs or poultry meat. Natural water systems, particularly those receiving human and animal waste, are also frequently contaminated.

The ubiquity of *Salmonella* and the wide range of materials with which they have been associated have meant that numerous media and techniques have been developed for their isolation. It is not the intention in this chapter to go into detail on the possible virtues or disadvantages of individual methods. Attention is focused instead on the fundamental aspects of the isolation and detection of *Salmonella*, particularly from food and environmental samples. Special emphasis is given to sublethal injury and its impact on recovery of the bacteria and the possible inhibition by other organisms. The application of rapid methods is also discussed. (*See* **Food Poisoning**: Tracing Origins and Testing.)

Sublethal Injury and its Impact on Detection

Salmonella are primarily enteric organisms and, whether in the gut or on contaminated foodstuffs, they are usually present as part of a mixed microbial population. Exceptions to this are seen with invasive, host-specific strains such as *S. typhi* in humans and *S. enteritidis* in poultry. The latter organism, for example, can be isolated in pure culture from both the tissues of infected chickens and the contents of intact shell eggs. It is generally necessary, however, to apply techniques which selectively encourage the growth of *Salmonella* and to use agars which permit the organism to produce characteristic colonies. The similarity in antibiotic resistance patterns between *Salmonella* and other enterobacteria has largely precluded the use of antibiotics and selection has relied upon the use of chemicals such as selenite, tetrathionate, and deoxycholate.

When microorganisms, particularly Gram-negative bacteria, are exposed to unfavorable conditions such as high or low temperature, they may become sublethally injured. This can be defined as 'a sensitivity to either selective agents or conditions to which normal cells are resistant.' In *Salmonella*, injury can be manifested by, for example, lowered resistance to the chemicals mentioned above. The organisms may also be less able to grow, particularly in selective media, at elevated incubation temperatures, such as 43 °C, which are used to improve the selectivity of isolation media when samples like sewage are examined.

Techniques used routinely for clinical specimens when large numbers of the target pathogen may be present and where it is relatively unaffected by its environment are not suitable for use with either food or environmental samples. These might have been exposed to conditions damaging to bacteria and they may also contain only low numbers of *Salmonella*. The direct inoculation of food or environmental samples into selenite broth or another selective medium may lead to false-negative results.

The principal site of injury in the bacterial cell is the outer membrane, and alterations in permeability allow the ingress of selective agents that are normally excluded. If cells are exposed to more extreme conditions, damage can also be caused to the chromosome and ribosomes. Much work remains to be done on the

response of *Salmonella* to adverse conditions and on the full impact of the exposure of these bacteria to different environments and their subsequent ability to grow in culture media. *Salmonella*, like other bacterial cells, are able to repair the lesions causing sublethal injury, either partially or completely, if they are incubated in a nonselective medium at 37 °C. Two media, buffered peptone water (BPW) and lactose broth, are the most commonly used, with the buffering capacity of BPW giving it an advantage over the other medium – particularly when foods containing fermentable substrates are examined.

Studies have demonstrated that the isolation rate from selective broth cultures is greatly improved if they are inoculated with an actively growing population of *Salmonella*. It is thus essential that initial incubation in BPW allows the organism both to overcome the effects of sublethal injury and to multiply.

The period of recovery or repair is often referred to as preenrichment. Incubation is usually for 18–24 h, although some workers have reported improved isolation rates when preenrichment is extended to 72 h. This is dependent on the material under investigation. *Salmonella* do not appear to be able to compete well with some other bacteria. Recovery media, by necessity, are nonselective and permit the growth of the majority of aerobic and facultative organisms present in the sample. Where this encourages the proliferation of either *Proteus* or *Pseudomonas* spp. – as with liquid raw egg and chicken meat – or lactobacilli – as with dairy products – *Salmonella* may be inhibited. This can be particularly pronounced when incubation is prolonged (Table 1).

The addition to BPW of novobiocin and cefsulodin, which inhibit *Proteus* or *Pseudomonas* spp., respectively, but have little effect on the growth of *Salmonella*, was found to improve the isolation rate of *Salmonella* from naturally contaminated samples of raw liquid egg (Table 1). The recovery of *Salmonella* from dried milk can also be improved if, after 2–4 h initial incubation, brilliant green at a final concentration of 0.0002% is added to the preenrichment culture.

Table 1 The influence of period of preenrichment and the addition of novobiocin and cefsulodin on the isolation of *Salmonella* from samples of liquid raw egg

Culture medium	*No. of* Salmonella*-positive samples*[a] *after incubation for*	
	24 h	*48 h*
Egg/BPW only	32	14
Egg/BPW plus novobiocin (5 μg ml^{-1}), cefsulodin (10 μg ml^{-1})	74	70

[a]110 samples were examined. Rappaport–Vassiliadis broth was used as the selective medium.
BPW, buffered peptone water.
Data from Humphrey TJ and Whitehead A (1992).

The presence of any number of *Salmonella* in a processed or cooked food is regarded as being potentially hazardous and most microbiological guidelines suggest an absence of the bacteria in 25 g of product. There is also evidence that, with some foodstuffs, particularly those rich in fat, and with certain vulnerable individuals such as the very young or the elderly, the infective doses of *Salmonella* may be low. Recent studies have also demonstrated that sublethally injured cells of *Campylobacter jejuni* are capable of repair in the intestinal tract. *Salmonella* may also be able to do this and it is important that sensitive techniques are used for their detection in the laboratory. In this respect, the importance of effective preenrichment cannot be overemphasized. As a general rule, it is prudent, when examining food and environmental samples from any source, to carry out primary incubation in BPW at 37 °C for a full 24 h. The ratio of sample to medium should be at least 1:10 and, where appropriate, the growth of competing flora may be suppressed by the addition of antibiotics.

Selective Culture

The necessity for selective culture techniques in microbiology has been recognized for over 100 years and much of the early work was centered upon *Salmonella*. Selective isolation is usually in two stages: growth in an enrichment broth followed by plating on a selective agar. Selective media make use of either dyes or chemicals. A formulation much in use in clinical laboratories is based on selenite. It is believed that this compound inhibits bacteria either by reaction with sulfhydryl groups or by the formation of analogs of sulfur amino acids. Selenite broth has been used successfully for many years and modifications to improve the isolation rate have included the addition of cystine and the incorporation of fermentable sugars, such as either mannitol or lactose, which prevent increases in the pH. Selenite is less selective at alkaline pH values.

Selenite can be used for direct inoculation with specimens such as feces or used following preenrichment. It is probably less suitable, however, for the latter when compared to other media. The medium can inhibit many bacterial types, but two important exceptions – *Proteus* and *Pseudomonas* spp. – can cause problems, particularly as they are capable of preventing the growth of *Salmonella* and can also resemble them on selective agars.

Salmonella have been shown to possess the enzyme tetrathionate reductase, which is absent in many

other bacteria. This provides a selective advantage and has been used in the development of another commonly used enrichment medium, tetrathionate broth. Tetrathionate is usually employed in combination with brilliant green and bile salts, the medium being known as Müller–Kauffman tetrathionate broth. It is probably more selective than selenite, but more care may be required in the control of incubation temperatures.

Malachite green, in combination with high concentrations (4%) of magnesium chloride and a medium pH of 5.0–5.2, has been shown to be effective for the isolation of *Salmonella*. The medium was originally developed for use with feces but modifications in the last 10–15 years, which have seen a reduction in the malachite green concentration from 0.012% to 0.004% and use of soya peptone rather than tryptone, have meant that the formulation, usually known as Rappaport–Vassiliadis (RV) broth, is more suitable for selective culture following preenrichment. This medium is probably the one of choice in the examination of nonclinical samples and numerous studies, on many different materials, have confirmed the superiority of RV broth over both selenite and tetrathionate.

The selective nature of the above broths is often finely balanced and can be disturbed by either a large inoculum, relative to broth volume, or overlong incubation. A change in incubation temperature can also profoundly affect the toxicity of the medium.

An inoculum-to-broth ratio of 1:10 is recommended for both selenite and tetrathionate broths, although following preenrichment culture there would appear to be no reduction in sensitivity if smaller volumes of BPW are used. The size of the inoculum is critical with RV broth and its ratio to the medium should be at least 1:100. With this medium, studies have indicated that more samples are *Salmonella*-positive after 48 h rather than 24 h incubation, although the differences are often not significant. In contrast, extended incubation significantly reduces the effectiveness of tetrathionate and selenite broths and, with the latter, overgrowth of competing flora occurs frequently. These media can also become more toxic to the target organism and *Salmonella* may die if incubation is continued much beyond 24 h. Care also has to be taken in controlling incubation temperature. Selenite and tetrathionate can be used at either 37 °C or 43 °C and are more selective at the higher temperature. Some studies have shown, however, that *Salmonella* will not always survive in these media at 43 °C. Initial studies on RV broth recommended 43 °C but it is now accepted that incubation at 41 ± 0.5 °C increases sensitivity.

Studies with particular samples or in particular situations have demonstrated that one selective medium may be superior to another. No one medium, however, can be relied upon always to maximize the isolation rate with all samples. There may thus be advantages in using two selective broths (**Figure 1**).

An important disadvantage with traditional culture techniques is the time taken to confirm whether a food sample is *Salmonella*-positive or -negative. There have been many attempts to make *Salmonella* testing more rapid and some of the newer methods are discussed below.

At first sight, it would seem possible to reduce the length of either preenrichment or selective broth culture, but in the past this approach has been largely unsuccessful. A number of publications have stated that recovery from injury is complete within 8 h at 37 °C. This is dependent on the culture medium used and the degree of cellular damage. In some systems, lag times of up to 72 h have been reported, and other studies have indicated that repair of damaged cells is not complete until some hours after growth has begun. Thus, subculture of preenrichment broths after 6 h rather than 24 h incubation was found to reduce the isolation rate. Newer work, however, has suggested that it may be possible to shorten traditional culture methods without any loss of sensitivity by using a combination of delayed release of selective agents into an optimized recovery medium. The new method makes use of a peptone-based medium that contains oxygen-consuming sterile outer membrane fragments from *Escherichia coli*, which protect bacteria by reducing oxidative stress. This allows damaged *Salmonella* cells to recover more rapidly than with conventional media. The method also makes use of delayed release capsules that release the selective components of RV broth after about 5 h at 42 °C. Trials to date indicate that this method is at least equivalent in performance to standard methods.

Selective agars should be so formulated that the competing flora is suppressed while the target organism is able to form discrete, characteristic colonies. A degree of selectivity is achieved by the addition of either dyes like brilliant green or chemicals such as sodium deoxycholate. The resistance of *Salmonella* to selective agents is similar to that of many other bacteria and this can create difficulties in media formulation. Thus, the detection of *Salmonella* on selective agars relies heavily on diagnostic aspects. Principal among these is the inability of most strains to ferment lactose and the production of hydrogen sulfide from thiosulfate at a neutral pH (leading to black colonies). More recently, the advent of affordable chromogenic substrates has led to the development of a number of new selective differential media for *Salmonella*. These

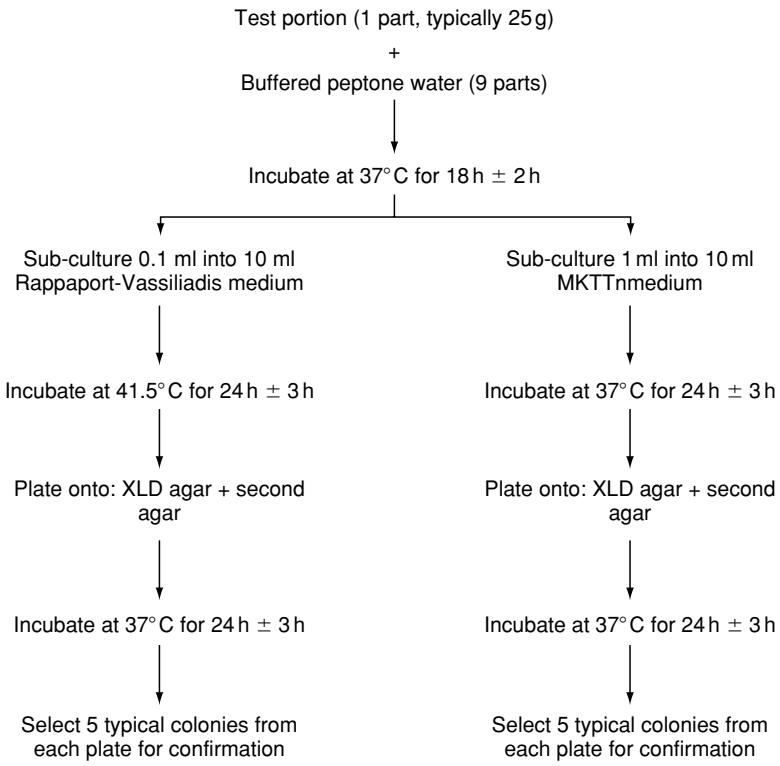

Figure 1 International Standards Organization (ISO) standard method for *Salmonella* isolation (ISO 6579:1997). BGA, brilliant green agar; XLD, xylose lysine deoxycholate.

new media often utilize a combination of traditional diagnostic markers with new substrates for the detection of galactosidase, glucoronidase, and esterase activities. Additionally, old diagnostic tools such as acid production from carbohydrate utilization have been applied to new substrates, e.g., propylene glycol utilization by *Salmonella*.

As with enrichment broths, many selective agars are available, but none has been shown definitively to be suitable for all applications and the best course of action would be to use two agars in parallel. The most commonly used media appear to be bismuth sulfite, xylose lysine deoxycholate (XLD), and brilliant green agar (BGA).

Many workers have attempted to improve selectivity by the incorporation of antibiotics and/or other selective agents. Of these, novobiocin, which prevents the growth of *Proteus* spp., seems to be the most useful.

Confirmatory Tests

Other organisms, particularly members of the family Enterobacteriaceae, can resemble *Salmonella* on selective agars and confirmation of '*Salmonella*-like' colonies is necessary. Tests take the form of biochemical confirmation, which usually involves assessment of urease and lysine decarboxylase activity, fermentation of dulcitol, indole production, growth in the presence of potassium cyanide, utilization of sodium malonate and, more recently, pyrrolindonylarylamidase (PYR) activity. These reactions, in combination with serological tests, are usually sufficient for identification, but additional tests may sometimes be necessary. If these are to form part of a laboratory routine, the use of commercial kits may be cost-effective. Antibodies against somatic (O) and flagella (H) antigens are used to confirm/identify *Salmonella*-like isolates. Somatic antigens are composed of polysaccharide, while those from the flagella are proteinaceous. Testing will usually comprise slide agglutinations with polyvalent O and H antisera, followed by the use of sera raised against specific antigens.

Rapid Methods

The time taken to confirm or deny the presence of *Salmonella*, especially in food and environmental samples, can create difficulties and, with perishable foods, can mean that products may be distributed and consumed before test results are known. It is not surprising that considerable attention has

been paid to the development of rapid methods of detection.

The new methodologies can be divided into two broad categories. There are those designed to detect or isolate viable cells and those that detect cellular components such as DNA. There are concerns that the latter methods may detect either dead cells or DNA released from them and that cultural techniques may lack the required sensitivity.

Conventional culture techniques, as discussed earlier, can be long-winded, cumbersome, and expensive. How much better it would be if it were possible to detect pathogenic bacteria by the direct application of a test to a food sample. *Salmonella* spp., and presumably other foodborne pathogens, appear to have unique DNA sequences that could allow detection. *Salmonella*-specific DNA probes and polymerase chain reaction (PCR) methods have been developed.

These techniques compare well to conventional culture when used on enrichment broths. The above studies were designed to obviate the need for selective culture and serological/biochemical identification. There are other examples, and various automated PCR-based systems are now available. Most of them require an enrichment period which delays results but sensitivity can be of the order of 3 CFU per 25 g of food.

Other new approaches include flow cytometry. These methods have faced the criticism that they are still too slow for the rapid pace of modern processed-food production methods. Tests would be much easier to perform and much more rapid if the modern technologies could be applied directly to samples. Such an approach, which continues to be investigated, is not without problems. One is due to the fact that DNA is heat-stable. Thus, intact DNA will be present in processed foods. DNA in both a free form and within dead cells of *Salmonella* can also survive for considerable periods in sea water. One possible way to overcome this problem is to include a culture step as a means of detecting viable cells. Not only will this increase the length of the test but it may not be sufficient to allow the growth of more severely damaged cells, as discussed earlier.

An alternative nucleic acid target to determine viability would be an RNA molecule that exists as many copies in a cell and which exhibits rapid turnover. Messenger RNA (mRNA) may prove to be a suitable target as it has a reported half-life of 2 min or less, and PCRs have been published for the detection of viability in *Escherichia coli* and the thermophilic *Campylobacter* spp. The rate of degradation of the mRNA is influenced by the method of inactivation, however, and this has to be considered when assessing food-processing conditions.

'Rapid' methods clearly have a place for the detection of *Salmonella* spp. in food samples. With many of the current tests, time is saved because selective plating and biochemical and/or serological confirmatory tests are not necessary. 'Rapid' tests would appear to offer advantages over traditional culture systems. It is likely that improvements will be made in both speed and sensitivity. This may encourage a wider use of the new technologies. If this is the case, it is to be hoped that people continue to bear in mind the vital roles of proper sample handling and the importance of using initial culture systems, which facilitate the recovery and growth of injured organisms.

See also: ***Escherichia coli***: Occurrence; Detection; Food Poisoning; Occurrence and Epidemiology of Species other than *Escherichia coli*; Food Poisoning by Species other than *Escherichia coli*; **Food Poisoning**: Tracing Origins and Testing

Further Reading

Bailey JS, Cox NA and Blankenship LC (1991) Comparison of an enzyme immunoassay, DNA hybridisation, antibody immobilisation and conventional methods for recovery of naturally occurring *Salmonella* from processed broiler carcases. *Journal of Food Protection* 54: 354–356.

Barrow GI and Feltham RKA (1993) *Cowan and Steel's Manual for the Identification of Medical Bacteria*, 3rd edn. Cambridge: Cambridge University Press.

Candish AAG (1991) Immunological methods in food microbiology. *Food Microbiology* 8: 1–14.

Cudjoe KS and Krona R (1997) Detection of *Salmonella* from raw food samples using Dynabeads anti-*Salmonella* and a conventional reference method. *International Journal of Food Microbiology* 17: 55–62.

de Boer E (1998) Update on media for isolation of enterobacteriaceae from foods. *International Journal of Food Microbiology* 45: 43–53.

Fierens H and Huyghebaert A (1996) Screening of *Salmonella* in naturally contaminated feeds with rapid methods. *International Journal of Food Microbiology* 31: 301–309.

Howie A (1999) A new development in *Salmonella* testing. *European Food and Drink Review*, winter issue, 53–56.

Mansfield L and Forsythe S (1996) Collaborative ring-trial of Dynabeads anti-*Salmonella* for immunomagnetic separation of stressed *Salmonella* cells from herbs and spices. *International Journal of Food Microbiology* 29: 41–47.

Reissbrodt R, Vielitz E, Kormann E, Rabsch W and Kuhn H (1996) Ferrioxamine E-supplemented pre-enrichment and enrichment media improve various isolation methods for *Salmonella*. *International Journal of Food Microbiology* 29: 81–91.

Saeed AM (ed.) (1999) *Serovar Enteritidis in Humans and Animals: Epidemiology, Pathogenesis and Control*. Ames, Iowa: Iowa State University Press.

Speck ML (1984) *Compendium of Methods for the Microbiological Examination of Foods*, 2nd edn. Washington, DC: American Public Health Association.

Salmonellosis

C Wray, Greendale Veterinary Laboratories Ltd, Knaphill, Woking, UK

C A Hart, University of Liverpool, Liverpool, UK

Introduction

Salmonellosis is the general term for human illness caused by members of the genus *Salmonella*; however from the epidemiological point of view it is necessary to differentiate the enteric fevers caused by *S. typhi* and *paratyphi*, A, B, and C from foodborne gastroenteritis caused by any of the other *Salmonella* serovars. The infectious nature of typhoid fever has been known for more than a century and the causative organism was one of the earliest to be isolated and characterized. The transmission of *S. typhi* occurs most often by water which has been contaminated by feces or urine from infected human carriers or food prepared and handled by such carriers who do not take adequate hygienic precautions. The organism appears to be only a parasite of humans and the higher apes and the link between poor sanitation and other epidemiological aspects of the disease was elucidated by William Budd before the opening of the bacteriological era.

In contrast, gastroenteritis (foodborne disease) is usually acquired from consumption of foods which may be directly or indirectly contaminated with *Salmonella*. Whether infection follows exposure to any pathogenic microorganism depends in part on the number of organisms (infective dose) that enter the host's tissues. Factors that may predispose to human salmonellosis are given in **Table 1** and the following section. There are four main clinical patterns of *Salmonella* infection in humans: (1) typhoid (enteric fever); (2) acute gastroenteritis; (3) septicemia, with or without localized infection; and (4) asymptomatic infection.

Table 1 Factors affecting the infectious dose of *Salmonella*

Bacteria-associated factors
Serovar, genotype
State and nature of contaminated material ingested (dry, moist, liquid, and does it allow bacterial multiplication?)
Resistance to antibacterials
Stress status of the bacteria
Host-associated factors
Age of host (very young and aged are most susceptible)
Effect of therapy (prior treatment with antibacterials, antacids, etc.)
Presence of other disease (malaria, human immunodeficiency virus (HIV), etc.)
Immune status of the host
Genetic status of the host

Enteric Fevers

Typhoid and paratyphoid fevers usually follow ingestion of drinking water contaminated with feces containing *S. typhi/paratyphi* A, B, or C, or food prepared or handled by a person who is a carrier of the organism. Direct transmission from one individual to another may occur via the fecal–oral route, especially in children, because of their poor attention to personal hygiene. While large-scale epidemics of typhoid fever have been traced to human fecal contamination of food, poor sanitary practice resulting in contamination of drinking water by fecal material containing *S. typhi* appears to be a major factor in many countries. Of foodborne outbreaks, 504 cases of typhoid occurred in Aberdeen; this was caused by the ingestion of tinned corned beef from Argentina. The meat was contaminated in Argentina, where river water containing unprocessed sewage was used to cool the containers after sterilization. The water seeped through the seams of the cans and reached the contents. In Aberdeen, infection developed in individuals who ate the corned beef, but also in those who ate meat that became contaminated by the machine that had been used to slice the corned beef.

In Mexico, in the early 1970s a large outbreak of typhoid was caused by a chloramphenicol-resistant strain of the organism, which had contaminated water. The lack of susceptibility of the organism to chloramphenicol was responsible for the death of many patients who were infected with it. The resistance determinant was a useful epidemiological marker when the organism caused disease in the USA, and a number of European countries.

Clinical Findings: Typhoid Fever

Typhoid fever is the best-studied form of salmonellosis and is well-described in the literature. It is potentially fatal for young children, particularly during the first year of life, and for the elderly, but in endemic areas a degree of immunity may develop in the population. Clinically, paratyphoid tends to be less severe than typhoid.

Illness usually occurs 7–10 days after consumption of infected water or food, and the initial symptoms tend to be both vague and variable. Headaches tend to be common, as does a general feeling of lethargy, along with abdominal pain/discomfort, and constipation is more usual than diarrhea. The fever increases progressively during the first week, and then decreases, although a biphasic response may be seen. Rose spots representing cutaneous vasculitis may be seen on the abdomen and enlargement of the spleen may be felt on abdominal palpation. The initial symptoms resemble those of acute gastroenteritis and *Salmonella* may be isolated from the feces. Then for a week or more the organism may not be isolated from the feces, although blood cultures are positive. This is then followed by further fecal shedding. In about 25% of patients *Salmonella* may be isolated from the urine. Fatalities usually arise from perforation of the intestine after the bacteria have been localized in the Peyer's patches, and/or pneumonia, but otherwise the disease runs a slow course over 5–6 weeks and then leaves the patient in a longer period of gradual convalescence.

The pathogenesis of the disease is of interest. From the small intestine the organisms pass via the lymphatics to the mesenteric lymph nodes, where, after a period of multiplication, they invade the blood stream via the thoracic duct and the liver, gall bladder, spleen, kidney, and bone marrow become infected during this bacteremic phase in the first 7–10 days of the disease. From the gall bladder a further invasion of the intestine occurs and lymphoid tissue, particularly Peyer's patches, are involved in an acute inflammatory reaction, followed by necrosis, sloughing, and the formation of the characteristic typhoid ulcers. Hemorrhage of varying degree may occur and, less frequently, perforation through a necrotic Peyer's patch may complicate the illness and result in death. *S. typhi* is present in large numbers in the inflamed tissues in the ulcers and it may localize in the kidneys and be found in other lesions that occur as a complication or sequela of typhoid fever, e.g., acute suppurative periosteitis, renal and hepatic abscesses, bronchopneumonia, and ulcerative endocarditis.

In 2–5% of convalescents the organism persists in the body, sometimes for an indefinite period, and in such chronic carriers the organism is most commonly found in the gall bladder or, more rarely, in the urinary tract.

Diagnosis

Typhoid fever should be suspected in any patient who has visited an endemic area; the diagnosis is confirmed by recovering the organism from the feces, urine, blood, or other tissues. Blood cultures are positive in about 80% of patients and in the early stages of illness, i.e., 7–10 days, it is the most reliable diagnostic test. Bone marrow cultures and skin biopsies of rose spots are, however, more likely to be positive for *S. typhi* than blood cultures. *S. typhi* and *paratyphi* can be isolated from the feces throughout the illness but are most frequently found during the second and third weeks. Repeated examination of the feces may be required before isolation is successful. The tube agglutination test (Widal reaction) for detecting antibodies to the somatic, flagellar H and the envelope Vi antigen has been used for the serological diagnosis and positive titers occur about day 7–10 after infection. Occasionally positive titers develop earlier, but they may be delayed and a negative result at an early stage of the illness therefore may be inconclusive (cross-reactions with other bacteria and previous typhoid may give false-positive results). Other serological tests, such as enzyme-linked immunosorbent assays (ELISAs) have also been developed to detect different antigens of the bacillus. Vaccination may give rise to false-positive serological reactions and, if the patient has a history of vaccination, it is advisable to take a second sample a week or so later to determine whether there is a rising titer.

Chemotherapy

The treatment of enteric fevers necessitates the use of antibacterial drugs. Chloramphenicol, ampicillin/amoxicillin, trimethoprim/sulfamethoxazole are the drugs of choice against sensitive *Salmonella*. Multiple drug-resistant (MDR) isolates of *S. typhi*, resistant to the first three antimicrobials (chloramphenicol, ampicillin, and co-trimoxazole) have emerged in Southeast Asia and in 1998 in Kenya. Therapeutic options include fluoroquinolones or azithromycin. Thus, because MDRs of *S. typhi* have been reported, all isolates should be tested for their antimicrobial susceptibility. Proper management of the fluid and electrolyte balance is important in all patients, especially the young and the old.

Immunoprophylaxis

Vaccines have been available for the prevention of enteric fevers for many years. The classic inactivated vaccine TAB protects against both typhoid and paratyphoid A and B and extensive field trials in endemic areas have shown a good degree of protection, although a high percentage of vaccinates develop local and systemic reactions.

In recent years, two new typhoid vaccines have become available. A vaccine prepared from the Vi capsular polysaccharide antigen has given good results in field trials. One dose protects for 3 years and it has a low incidence of side-effects. It is not licenced for

children less than 18 months of age because they may have a suboptimal response.

A live oral preparation produced from an attenuated strain (Ty 21a) of *S. typhi* is available in enteric-coated capsules; it is administered in three doses on alternate days. Significant side-effects have not been reported, but transient mild nausea, vomiting, abdominal cramps, diarrhea, and urticaria may occur in less than 1% of vaccinates. In mass trials it has not proved as effective as the previous vaccines.

Acute Gastroenteritis

Acute gastroenteritis is usually acquired from consumption of foods which may have been directly or indirectly contaminated with *Salmonella*. A wide variety of animal species have been shown to be capable of harboring the organisms without showing clinical signs, and in the developed world, turkeys, swine, chickens, and cattle have often been found to be infected in abattoir studies. Animal products such as milk and eggs have also been shown to be important vehicles of infection. Many other foodstuffs, including vegetables and salads, have also been found to be associated with *Salmonella* outbreaks. It should also be remembered that person-to-person spread may also be of importance.

Clinical Findings

The incubation period ranges from 8 to 48 h following the ingestion of contaminated food. The most common symptoms are diarrhea, abdominal pain, fever, and headache. Extremely fluid stools persist for 3–4 days, and these may contain mucus and occasionally blood. The temperature may be slightly elevated in most patients but usually returns to normal within 1–2 days. In the young and the elderly the symptoms may be more severe and last for a week or more.

Salmonella gastroenteritis is usually a self-limiting illness and fatalities are uncommon. Sigmoidoscope and biopsy examinations have demonstrated that the colon is the major site of infection. The changes in the colon range from edema of the lamina propria with a focal or diffuse inflammatory infiltrate to a more intense inflammation with disruption of the surface epithelium and multifocal microabscesses. In more severe cases, edema, vascular congestion, infiltration of the lamina propria with polymorphonuclear leukocytes, and abscess formation have been recorded.

Chemotherapy

Although *Salmonella* are usually sensitive *in vitro* to many antibiotics, their use for the treatment of uncomplicated gastroenteritis until recently has generally been contraindicated by the lack of favorable effect on the course of disease and by a prolongation of *Salmonella* shedding. Following the introduction of the fluoroquinolones, a number of clinicians have advocated their use for the treatment of *Salmonella* gastroenteritis because of their efficacy in reducing the severity of symptoms, the duration of illness, and *Salmonella* shedding. However, this has been followed by increasing development of resistance.

Salmonella Bacteremia (Septicemia)

A transient bacteremia may occur in patients with gastroenteritis caused by *Salmonella*, although periods of longer persistence occur in the blood of individuals with one of the forms of enteric fevers or those suffering from underlying disorders. Chronic bacteremia is commonly associated with *S. choleraesuis* infection, and the fatality rate is usually two to three times that caused by typhoid fever. Metastatic infectious foci, of which the most important are intravascular lesions, osteomyelitis, and meningitis, are common complications of disease caused by *S. choleraesuis*. This organism may lie dormant in some tissues, usually bone marrow, and then later become activated.

The serovar most commonly associated with bacteremia is *S. typhimurium*. This is perhaps not surprising because it is the most frequent cause of salmonellosis in humans and other animals. The high-grade bacteremia suggests the presence of focal intravascular infections in either the large arteries or the endocardium, but whether this is a primary or secondary effect is not certain. In sub-Saharan Africa, non-typhi *Salmonella* (NTS) are the most common isolates from blood culture in children (< 50% of isolates). This is not related to human immunodeficiency virus (HIV) infection but occurs in the malaria season and especially in children with severe malarial anemia. In HIV-infected adults NTS bacteremia is an early acquired immunodeficiency syndrome (AIDS)-defining illness: relapse after therapy commonly occurs.

Focal infections, such as osteomyelitis, arthritis, and meningitis, are not uncommon sequelae to *Salmonella* infections. Any area of the skeleton, especially the long bones, may be invaded. Two types of arthritis have been described: bacteremia may give rise to suppurative arthritis with the organism in the joint fluid, or a sterile inflammatory reaction involving the synovia may occur as a sequel in about 2.5% of cases of gastroenteritis.

Chemotherapy

Chemotherapy is usually similar to that for typhoid fever, although prolonged antibiotic therapy for

localized infection, e.g., osteomyelitis, may give rise to resistant organisms.

The Salmonella Carrier State

Although *Salmonella* carriage occurs more commonly after illness, it may also occur in individuals who have not been ill. All patients with *Salmonella* gastroenteritis usually shed up to 10^6–10^9 organisms per gram feces early in the course of disease and excretion generally persists for 2–4 weeks, with 0.2–0.6% of cases continuing to shed the organism for 1 year or more. The persistence of *Salmonella* does not relate to the severity of the disease.

A number of factors are involved in the pathogenesis of chronic carriage. Age is important: young children tend to be prolonged convalescent carriers but do not usually develop the chronic state. Adults in their sixth decade or older are likely to become carriers, with the biliary tree believed to be the site of multiplication.

Pathogenesis of Infection

There is little information concerning the number of organisms involved in many outbreaks of naturally acquired disease. Volunteer studies indicated that *c.* 10^6–10^9 organisms were necessary to cause disease in 50% of the healthy individuals participating. In the case of *S. typhi*, none of the individuals became ill when given 10^3 organisms: of those given 10^5, 10^7, and 10^9 organisms, 25, 50, and 95% respectively developed disease. The clinical manifestations did not vary with the size of the inoculum, although the incubation periods were longer in those receiving the smaller doses. However, volunteer studies in healthy adults involving only a few serovars may not be directly transferable to naturally acquired infection. Where it has been possible to obtain the vehicle, e.g., chocolate, cheese, and desserts, the number of *Salmonella* have often been $< 10^2\,g^{-1}$, especially in foods with a high fat content or a good buffering capacity which may protect the *Salmonella* from the gastric acidity. It has also been demonstrated that people on antacids, where gastric acidity has been lowered, or on antibiotics, which may alter the gut flora, are more susceptible to infection. The minimal infective dose can also vary with age and the state of health. Factors which help to determine the infective dose of *Salmonella* are given in Table 1.

Following ingestion of *Salmonella*, the organisms pass through the stomach where the bactericidal activity of the stomach produced by the low pH of gastric juice regulates the number of organisms entering the small intestine. The course of events then depends on multiple factors, of which the most important is probably the serovar of the *Salmonella*. Thus, *S. typhi* after an initial period of multiplication invades the intestinal mucosa, primarily it is thought in the M cells of the Peyer's patches of the distal ileum, where a cellular reaction may take place in the lamina propria. Although the infiltrate is composed of mononuclear cells, in the case of *S. typhi*, it contains predominantly polymorphonuclear leukocytes when other noninvasive *Salmonella* produce infection. It has been suggested that differences in phagocytic effectiveness of the two types of phagocytes may explain the high incidence of bacteremia with *S. typhi* and the relatively low occurrence with other non-invasive *Salmonella*. The major factor for intestinal penetration is encoded by genes that are clustered in a large (40 kb) area of the chromosome, designated *Salmonella* pathogenicity island 1 (PI 1). In *Salmonella* there are at least five PIs, of which two are associated with type III secretion systems, which export proteins in response to bacterial contact with cells. Although the onset of systemic and localized infection is fairly similar, the symptoms produced are not identical. Localized infection is usually manifested as acute gastroenteritis, whereas only one-third of typhoid patients develop diarrhea, usually several days after the onset of fever. The contribution of toxins in the etiology of diarrhea has never been conclusively demonstrated, although several putative toxins have been described, and it is now suggested that the type III export system on PI 1 secretes one or more proteins with enterotoxic activity.

To cause bacteremia, *Salmonella* must have strategies to overcome each of the host's deep-tissue defenses, including resistance to phagocytic cells, complement, specific antibodies, and cellular immunity. It must be able to reach its preferred site of replication, i.e., the liver and spleen, where intracellular multiplication occurs, which protects it from the host's defense mechanisms. If the bacteria are not contained, they may then invade the blood and replicate rapidly.

See also: **Food Poisoning**: Classification; ***Salmonella***: Properties and Occurrence; Detection

Further Reading

Baumler AJ, Tsolis RM and Heffron F (2000) In: Wray C and Wray A (eds) *Salmonella in Domestic Animals*, pp. 57–72. Wallingford, Oxon: CABI.

Rubin RH and Weinstein L (1977) *Salmonellosis: Microbiologic, Pathologic and Clinical Features*. New York: Stratton International Medical Book.

Winstanley C and Hart CA (2001) Type III secretion systems and pathogenicity islands. *Journal of Medical Microbiology* 50: 116–126.

Salting *See* **Curing**; **Smoked Foods**: Principles; Applications of Smoking; Production; **Preservation of Food**

SAMNA

S A Abou-Donia and E I El-Agamy, Alexandria University, Alexandria, Egypt

Introduction

Although similar to ghee and other concentrated milk fats, samna is of especial importance to the dairy industry in Egypt. This article deals specifically, therefore, with the production and properties of samna as traditionally manufactured in the rural areas of Egypt.

Samna is pure, clarified milk fat, produced in Egypt. It is generally prepared from buffuloes' milk or cows' milk, which constitute 63.5% and 35% respectively of total milk output in Egypt, while small amounts of samna are prepared from sheeps' milk and goats' milk, which constitute 1% and 0.5%, respectively, of total milk output.

The main object of the primitive dairy industry in the rural districts of Egypt is to separate milk fats for making butter, and to make the remainder into products which are consumed as such or, after storage, throughout the year. In lower Egypt, farmers put fresh milk in shallow or deep earthenware pots, martrad or shalia, and leave it undisturbed in a warm, dark place till the cream rises and the milk coagulates. The cream layer is removed and beaten into butter, which is converted into samna. The presence of earthenware pots, such as used in cream-making, in the tomb of king Horaha of the first dynasty (3200 BC), indicates that the art of samna-making was known to the Ancient Egyptians.

Samna making by the Traditional Method

Samna is usually made from butter produced from ripened cream; however, nowadays, in some cases, it can be made using cream prepared from salted or unsalted whey. The resulting samna is called whey-cream samna.

The traditional process of samnamaking depends primarily upon the properties of the butter with respect to its acid degree value, taste, odor, and weight. Weight is important to determine both the yield of the samna and the amount of salt to be added to the butter prior to the making process.

Liquefaction of Butter and Addition of Salt

Butter is placed in a stainless-steel or aluminum vat; iron or copper pans should not be used, for if the samna is contaminated with heavy metals, oxidation of the fat is accelerated. The butter is then heated with continuous stirring until liquefied (50–60 °C), and the salt is added at a rate of about 2–4% of butter weight.

The addition of salt leads to the following: (1) a decrease in the water content of the resultant samna because the salt increases the boiling point of the water in the butter; (2) it helps in the fat separation that arises from the difference in density between fat and nonfat phases; (3) it increases the quantity and shelf-life of the byproduct (morta); (4) it plays an important role in the precipitation of the proteins in butter during boiling. In addition, some makers of samna believe that the addition of high levels of salt during the making of samna is necessary in order to store samna for a long time. This view is misguided for the following reasons: (1) salt is not a fat-soluble material and hence does not have a preservative effect; (2) contamination of the salt with heavy metals, e.g., iron or copper, accelerates fat deterioration; (3) salt is a good absorber of water, and thus the presence of salt in samna can increase the moisture level and hence risk of rancidity. In general, after the butter has been heated and salted, it is filtered through a cloth, such as cheesecloth, to remove any foreign matter mixed with the butter. In the case of a good-quality butter, this step is omitted.

Butter Boiling and Ripening

Butter boiling and ripening are the most important parts of the process of samnamaking with respect to the quality of the resultant product. They are performed as follows. The liquefied butter is heated further to 90–96 °C and, in the course of heating, a foam is produced which is known as the boiling foam. At this point, the level of heating should be reduced.

After the disappearance of the foam, regular boiling or simmering of the fat starts. After a certain time, floating particles appear, which are principally composed of proteins and phospholipids.

Some makers prefer to remove the floating particles, but others do not. In general, at 103–107 °C most particles precipitate to the bottom of the vat. At 107–110 °C the amount of nonfat particles increases. At 110–115 °C all the nonfat suspended particles are precipitated to form morta, which has the same color as samna, i.e., bright yellow. At 115–125 °C, the color of morta becomes dark and the characteristic cooked odor of samna starts to appear. Moreover, a foam suddenly appears, at which point the heating should be stopped (this foam is called the 'ripening foam'). Excessive heating after this stage leads to the following: (1) production of a darker product; (2) the precipitated nonfat particles start to disperse and suspend again and will be difficult to separate; (3) the samna acquires an undesirable flavor; (4) the keeping quality of the resultant samna is very poor, owing to the weak body derived from the crystallization of the fat.

The quality of samna depends on the control of the boiling and ripening processes and the expertise of the maker. For example, continuous stirring during boiling is very important to avoid any overcooking defects, as is the time at which the heating and stirring should be stopped. All these aspects must be controlled by the samnamaker.

Decantation and Filtration

Decantation and filtration are carried out when the samna is slightly hot. The clear fat layer is decanted by pouring it into another vat until just above the level of precipitated particles (morta). The latter portion of samna is passed through cheesecloth at least twice, to retain any precipitated particles, and the clear fat is then added to the rest.

Packaging of Samna

The packing of samna is considered a vital process. Samna should be filled into storage cans with sufficient temperature to expel air bubbles inside the cans.

Many different containers can be used for packaging samna, e.g., jars of dark-colored glass to avoid light-induced oxidation, tin-cans free of rust, or there are special pots (which are known as barany and mostly used in Egypt) made from clay with a glazed inner surface.

Too hot ambient temperatures and direct light should be avoided for samna storage.

Factors Affecting the Keeping Quality of Samna

1. An increase in the acidity of the butter or cream decreases the keeping quality of samna.
2. The heat treatment of the fat during the manufacturing process should be not less 110 °C and not more than 125 °C, since heat treatment (115–125 °C for 20 min) causes noticeable changes in the fatty acid contents and induces changes in unsaponifiable matter components.
3. The presence of traces of heavy metals, e.g., iron and copper.
4. The presence of air or oxygen inside pots or cans.
5. The storage temperature.
6. The exposure of samna to light during storage.
7. A high level of moisture and/or nonfat matter in the samna.
8. The presence or absence of natural antioxidants.
9. The source of cream prepared for buttermaking, since samna results from whey cream has a lower keeping quality due to higher rancid flavor and increase in peroxide, thiobarbituric acid, and acid values.
10. Samna produced in winter has a better keeping quality than that made in summer, spring, or autumn. This is presumably due to the higher phospholipids and unsaponifiable matter content and lower levels of unsaturated fatty acids.
11. The keeping quality of cows' milk samna is better than that of buffaloes' milk due to the higher contents of phospholipids and vitamin E in cows' milk fat. The chemical composition of samna is shown in Table 1.

Characteristics of Good-Quality Samna

1. Samna produced from cows' milk fat has a golden-yellow color (owing to the high content of β-carotene), but in the case of buffalo milk fat, the product has a white, slightly greenish color.

Table 1 Chemical composition of samna

Constituent	*Cows' milk samna*	*Buffaloes' milk samna*
Water	0.3–1.0%	0.3–1.0%
Fat	98–99%	98–99%
Solid not-fat	0.2–0.3%	0.2–0.3%
Phospholipids (mg 100 g^{-1})	93.2	76.6
Sterols (mg 100 g^{-1})	241	316
Cholestrol (mg 100 g^{-1})	33.05	42.66
Carotenoids (μ g^{-1})	4–15	0.2–1.0
Vitamin A (μg g^{-1})	3–8	3–8
Vitamin D (μg g^{-1})	7–10	7–10
Vitamin E (μg g^{-1})	30–45	20–30
Vitamin K	Traces	Traces

Table 2 The fat content of samna made from cows' milk and buffaloes' milk

Fat constant	*Cows' milk samna*	*Buffaloes' milk samna*
Reichert–Meissl value	≥22	≥25
Polenske value	≤2.7	≤2.7
Kirschner value	≥19	≥22
Saponification value	≥220	≥222
Iodine value	29.65	37.90
Butyrorefractometer reading	40–44	40–43

Sometimes hydrogenated oils or other fats are added to samna and to detect this kind of adulteration, milk-fat constants should be measured, e.g., Reichert–Meissl, Polenske, Kirschner, iodine, and saponification values.

2. The samna should be free of cooked, slightly sweet flavor and it should be free of rancidity.
3. It should have a gritty texture at 10–20 °C.
4. The fat content should be not less than 99.5%, the moisture not more than 0.3%; the expected chemical properties of fat(s) are shown in Table 2.
5. The main flavor compounds in samna are methyl ketones, 2-enals, and 2,4 dienals.

Antioxidants used in Samnamaking

The most common cause of deterioration in samna is oxidative rancidity resulting from exposure to light. To prevent this phenomenon, antioxidants are added. These are categorized as follows:

1. Natural antioxidants, which are present in samna, derived from milk, e.g., phospholipids and vitamin E (tocopherol).
2. Antioxidants derived from nonfat matter (proteins) during the heat treatment process of samnamaking, e.g., sulfhydryl groups (–SH groups).
3. Some natural materials that may be added to certain brands of samna during packaging, e.g., soya bean and wheat flours, dried date, fenugreek grains, and safflower at a rate of 0.5–1%.
4. Chemical substances, e.g., propylgallate at a rate of 0.01–0.003%, but these materials are rarely used.

Nutritive Value of Samna

Samna is mainly composed of pure fat and thus it is considered a good source of energy. Moreover, samna provides significant quantities of fat-soluble vitamins, e.g., A, D, and E. In addition, the use of samna in cooking improves the taste of meal, and hence overall intake may be increased.

The Byproduct (Morta) of Samnamaking

Morta is a byproduct of samnamaking and, if butter rather than cream is used, the amount of morta is usually equivalent to 6% of the weight of butter. During boiling of butter, more than 85% and 75% of cows' and buffaloes' butter cholesterols were separated with morta respectively. The composition of morta depends primarily upon the steps in and conditions of samnamaking but it is generally as follows:

- water: 10–18%
- fat: 40–65%
- nonfat solids: 10–25%
- ash and salt: 10–15%
- cholesterol: buffaloes' milk samna 131.6 mg $100\,g^{-1}$ versus 193.6 mg $100\,g^{-1}$ in cows' milk samna

The nutritive value of morta is mainly attributable to the high content of phospholipids. Morta is consumed as is, or used in making mish cheese.

See also: **Buffalo**: Milk; **Ghee**; **Goat**: Milk; **Sheep**: Milk

Further Reading

Abd El-Hady SM, Dawood AH, Montasser EA and Shinana ME (1992) Evaluation of whey samna. In: *Proceedings of the 5th Egyptian Conference for Dairy Science and Technology*, pp. 104–113.

Abou-Donia SA (1984) Egyptian fresh fermented milk products. *New Zealand Journal of Dairy Science and Technology* 19: 7–18.

Fahmy AH (1961) *Samna Manufacture*, 1st edn, pp. 19–25. Cairo: Modern Commercial Press.

Ghita El, Abd-Rabo FH, Ismail AA and Dabiza NM (1992) Studies on cholesterol and cholesterol esterase contents in milk and some dairy products. *Egyptian Journal of Food Science* 20(2): 221–231.

Hassan MNA, Rady AH and Ibrahim MK (1989) Stability and keeping quality of butter oil and samna. *Egyptian Journal of Dairy Science* 17: 93–104.

Kankare V and Antila V (1988) Production and properties of Indian ghee and Egyptian samna. *Elintarvikey lioppilas* 23: 28–30.

Lambert LM (1975) *Modern Dairy Products*, 3rd edn, p. 307. London: Food Trade Press.

Mahran GA, Ahmed NS, Hofi AA and Asker AA (1975) Phospholipid contents in buffaloes butter oil 'samna' as affected by processing. *Egyptian Journal of Dairy Science* 3: 51–54.

Mahran GA, Hamzawi LF, Ahmed NS and Sayed AF (1992) Seasonal composition variation of buffalo milk fat in relation to keeping quality of the resulting fatty products. *Egyptian Journal of Food Science* 20(2): 263–272.

Warner JN (1976) *Principles of Dairy Processing*, p. 253. New Delhi: Wiley Eastern.

SANITIZATION

P N Hoffman, Public Health Laboratory Service, London, UK

Principles

Sanitization is usually defined as those disinfection processes appropriate to use in the context of food. Disinfection is 'the elimination of microorganisms such that they may remain only at levels that are not harmful to health', but there are, in practice, many minor variations on this general definition. This elimination may be by removal through cleaning, by killing in processes such as by chemicals or heat, or a combination of these methods. Such a definition is, in practice, reliant on the context in which it occurs and outlined in the following paragraph.

The level of microorganisms designated as a 'safe' level will vary according to their location: more bacteria could be present on a 'safe' floor than on a 'safe' chopping board. It will also depend on the type of microorganism present: for a pathogen such as *Shigella*, where very small numbers can give rise to infection, only a total absence would be safe; for nonpathogenic (i.e. which do not cause disease) spoilage organisms, low numbers present neither a threat to health nor any immediate threat to quality. The process of disinfection does not normally include the killing of bacterial spores. These are highly resistant dormant forms in the life cycle of members of the genera bacillus and clostridia, such as *Bacillus cereus* and *Clostridium botulinum*. These are normally only eliminated by the more stringent process of sterilization, that is the total elimination of all microbial life, preferably in a process that has a high quality assurance. Some chemical disinfectants are capable of killing bacterial spores and are termed 'sporicides.'

Mechanisms and Kinetics

Chemical disinfectants kill their target microorganisms by reaction with one or more components vital to the internal integrity and function of that cell. This reaction can be chemical, physico-chemical, or both. A chemical reaction is one where covalent bonds within molecules are broken or formed or where the charge on an ion is altered. A physicochemical reaction is one where non-covalent bonds, i.e., areas of hydrophobic affinity or of noncovalent polarized attraction are disrupted. These reactions follow the kinetics of chemical reactions in general; thus, the killing of microorganisms by disinfectants can be similarly categorized. As the end effect, the death of microbes, may be the result of a single reaction or, more likely, of a combination of many reactions, the kinetics of microbial death can become somewhat more complex than the more simple chemical reactions. The death of microorganisms in these circumstances approximates to the kinetics of a first-order chemical reaction; the rates of microbial death occur in a logarithmic manner. To illustrate this, if 1 000 000 bacteria are acted on by a lethal chemical agent, they will lose a proportion, rather than a set number, of their population with each unit of time. If a lethal agent kills nine out of 10 of a particular bacterium every minute (i.e., one base ten logarithmic reduction per minute), 1 000 000 bacteria at time zero (i.e., the start of the reaction) will be reduced to 100 000 after 1 min; to 10 000 in 2 min, and so on. After 6 min, there will be one bacterium left. After 7 min there will be (in theory) 0.1 of a bacterium remaining. A tenth of a bacterium does not exist, but in statistical terms, this means one-tenth the chance of a surviving bacterium, or one bacterium surviving in every 10 samples processed. This continues, so that at 12 min from the start, there will be one-millionth of a bacterium left; in effect, one bacterium in every million samples. Thus, the concept of disinfection, as with the concept of safety, takes on a statistical element.

Microbial Resistance to Chemical Disinfectants

Microbial resistance to chemical disinfectants is a less well-defined phenomenon than the better-known analogy of microbial resistance to antibiotics. With antibiotics, the concentration of an agent achievable in the body will determine the so-called 'break point.' This provides a naturally set level of antibiotic concentration against which to judge resistance. If an organism is killed or inhibited below this break point, it is considered sensitive; if it can grow at or above the break point, it is resistant. Attainable concentrations of chemical disinfectants are not similarly fixed and usually can be varied within any given use situation. Thus, 'resistance' to chemical disinfectants has to be considered within these constraints but can still be a useful concept.

There are several possible mechanisms by which resistance to chemical disinfection exists.

- Innate resistance: This implies that there is either no susceptible target within the microbial cell or that the microbial cell is in some way impermeable to the agent. An example of the former is the resistance of viruses that comprise only nucleic acid and protein (such as the picornaviruses, a group of viruses that include a number of viruses that infect by ingestion, an example of which is poliovirus) to disinfectants such as phenolics and quaternary ammonium compounds. An example of the latter is the high resistance of bacterial spores to most disinfectants. Bacterial spores are a 'survival' form of certain bacterial genera that show extreme resistance to chemical and physical inactivation. Only a few chemical disinfectants (sporicides) are capable of inactivating bacterial spores.
- Acquired resistance: In this form of resistance, certain bacteria can adapt to survive or grow in the presence of a particular disinfectant. This can be through a process of 'training', whereby a bacterium is exposed to a sublethal concentration of a disinfectant and can then bring in to play a variety of phenotypic adaptations, such as capsule production, that can enable it to survive increasing concentrations of the disinfectant. Another mechanism is by the acquisition of transmissible genetic elements that confer resistance to lethal agents. Although this is better known with bacterial resistance to antibiotics, it does also exist with disinfectants.

Reasons for Disinfection Failure

It is not only genuine resistances, as above, that can result in failures of a disinfection process. Many other factors can affect the performance of a disinfectant and lead to a suboptimal disinfection result:

- Neutralization of a chemical disinfectant can occur by chemical or physical reaction. Disinfectants, by their very nature, will react with molecules in a microbial cell and thereby disrupt its vital functions. They will similarly react with nonliving organic matter, removing disinfectant molecules from effective solution. If there is an excess of organic matter and the disinfectant has a high affinity for it, this can lead to a disinfection failure due to neutralization of the disinfectant. The disinfectant will have reacted with nonliving organic matter, leaving insufficient for microbicidal purposes. Although excess organic matter is the most commonly encountered disinfectant neutralizer, others do exist. There can be reactions between disinfectant and neutralizing compounds, converting the disinfectant compound to a nonmicrobicidal reaction product. These are specific chemical reactions and tend to occur in unusual circumstances. The other main source of disinfectant neutralization is from physical reaction with surfactants, charged, or polarized large molecules, again removing disinfectant molecules from active solution. Disinfectants that are especially susceptible to this are those whose activity depends on charge or polarization within their molecule. This makes them both prone to these charge-based interactions and also means that, once their charge or polarization has been disrupted, their microbicidal capacity is neutralized.
- Failure to make contact with the target organism is a common source of ineffective disinfection. Organic matter, as well as neutralizing disinfectants (see above), can form a barrier to disinfectants, shielding microorganisms within and facilitating their survival. In this way, microorganisms can withstand a disinfection procedure that, to the operator, should have been effective, i.e. the correct concentration of disinfectant applied for the right time to act against a microorganism innately susceptible to that disinfectant. Some disinfectants, mainly those with innate or added surface active capabilities, are better able to overcome shielding barriers whilst others, especially those that coagulate proteins, are particularly prone to this shortcoming. However, even those disinfectants with surfactant properties cannot be expected to penetrate thick layers of viscous or solid matter. A clean surface is an essential prerequisite to routinely efficient disinfection. Organisms within crevices in surfaces are usually protected from contact with disinfectants. This protection is enhanced by the presence of organic matter. In situations when it may be difficult for a disinfectant to penetrate a crevice, it will probably be impossible for it to penetrate that crevice when it is also filled and protected by organic matter. So, not only an absence of soiling, but also a crevice-free surface, is needed. Crevices can be innate, as in the grain of a wooden chopping board, a result of poor design, as in machinery or pipework or a byproduct of use, as in knife scours in a plastic chopping board.

Another situation where disinfectants do not work due to a failure to make contact with their target is that of poor-quality application such that the agent does not make contact with all the surfaces intended for disinfection. On an accessible surface, this could be a result of human error, perhaps due to inept or badly instructed staff, to equipment that does not apply disinfectant to areas that are difficult to reach, or to failure to remove extraneous objects that

may shield a surface from a directly applied or sprayed disinfectant. Inaccessible surfaces, such as those within pipework or other enclosed vessels, present a more substantial problem. Ideally, the application of disinfectants should be considered at the design stage, in the form of either a modification allowing free access to all surfaces or 'cleaning in place'-style dispensing nozzles as permanent internal features.

The concentration of a disinfectant will also determine the microbicidal efficacy. Disinfectants are formulated to work at a specific concentration. Sometimes, different concentrations for different situations are stated, for example, in the presence or absence of organic soiling. If the concentration is too low, inefficient microbicidal action will result; if the concentration is too high, the consequences could be wasted resources, corrosion problems, or taint and toxicity problems. Disinfectants should be made up for use accurately. A 'splash in a bucket' approach cannot insure accuracy. Similarly, if a disinfectant is going to become substantially diluted during its use, it should be formulated to be near its designated use concentration after, rather than before, this dilution.

All disinfectants need adequate time in which to work. The exposure time of a microorganism to a disinfectant can be determined by the method of application; for example, it is very convenient to use a volatile surface disinfectant, such as alcohol, in the form of a wipe. The surface to be disinfected will be dry and useable within seconds. The very convenience of this form of application is also what can make it a less effective method.

Another factor that will determine the efficacy of surface disinfection is the wetting ability of a disinfectant. When a disinfectant without a detergent or other surface active agent (i.e., which has no wetting ability) is spread on a surface, the film that it forms will turn rapidly into discrete droplets with dry areas in between these droplets. No substantial disinfection will occur in these dry areas, which will comprise the majority of the area to which liquid was initially applied. A wetting agent, usually a detergent, will allow the disinfectant to remain as a continuous film on the surface and exert its full action until it is taken out of effective solution by drying and so becoming unavailable to microbial cells. The wetting ability also enhances a disinfectant's ability to penetrate or remove layers of organic matter.

As with chemical reactions in general, the temperature at which a disinfectant acts will affect the speed of its activity. Unless otherwise specified, disinfectants are formulated to work around normal ambient temperatures. There are use situations where they might be expected to work at low temperatures, for instance, in a refrigerated food production unit or out of doors in a cold climate. For a disinfectant to work effectively in such situations, it will need a higher concentration, a longer exposure time, or both. From the same principles, if the temperature is higher than normal ambient, a more rapid and efficient disinfection will occur. Temperature itself starts to become lethal to vegetative bacteria (not bacterial spores) and other microbes around 60–65 °C, at which temperatures, the time needed for heat disinfection is in the range of several minutes to hours. At 80 °C, disinfection will occur far more rapidly, needing only seconds for susceptible microbes to be killed at this temperature.

Other Factors to be Considered in Disinfectant Choice

Chemical disinfectants have a high reactivity with biological systems, this being an integral part of their microbicidal mechanism. As living systems share many biochemical similarities, it is not unusual for disinfectants to have a toxic effect on humans. Toxicity may occur through one or more routes of contact, such as through the skin, ingestion, inhalation, or absorption into mucous membranes of the eyes, nose, etc. There are two separate considerations to the use of chemical disinfectants in catering and the food industry: risk to the user and risk to the consumer.

The major toxic risk is to the users of chemical disinfectants. It is they who will handle concentrated disinfectant solutions or solids, where any toxicity will be many times that of the use dilution; who will make up the use dilution, involving splash and skin contact risks during this process; who will apply the disinfectant with attendant risks of skin contact and inhalation; and who will dispose of the disinfectant, with more contact and inhalation risks. Chemical disinfectants, especially in their undiluted form, must always be assessed for hazard before use. Any hazard must be eliminated by handling procedures insuring minimal contact (especially uncontrolled contact such as splashing), and use of effective and appropriate personal protective equipment (gloves, eye protection, etc.), where contact cannot be ruled out. As a general rule, prevention of operator contact with toxic agents by containment of the agent is preferable to the use of personal protective equipment.

Consumer toxicity, whilst much rarer, is a serious consideration in terms of the number of people affected as well as commercial considerations. The dilution factors of disinfectants that find their way from food-processing areas to consumers via a food product will be immense. Nevertheless, the

consequences of such an occurrence are equally immense in terms of the potential numbers of people affected, as well as commercial considerations. Control is achieved by segregation of significantly toxic chemicals from food-handling areas or routes into those areas.

Taint is a problem allied in many ways to toxicity in terms of origin and consequences. The problem here is that certain disinfectants can impart characteristic and undesirable taints to foods. Taint can be caused by extremely low concentrations of these chemicals, usually in terms of a few parts per billion of a contaminant in food. The disinfectants most implicated in taint problems are those with phenol-derived molecules in them. These should be regarded with suspicion unless an assurance of lack of taint can be given. As with toxicity, such disinfectants are best excluded from use in food-handling areas completely.

Some disinfectants, particularly those of an oxidative nature, can cause or accelerate corrosion of a variety of metals. Whilst this is not directly connected to microbicidal issues, it is a factor that must be considered in the wider context of the practicalities of disinfectant use. Particularly implicated in this are hypochlorite disinfectants, which can start rusting carbon steels in minutes. Corrosion is dependent on a combination of factors, mainly: the disinfectant, its constituents and concentration; the materials that are in contact with the disinfectant; and the contact time and temperature.

Characteristics of Commonly Used Disinfectants

Hypochlorites

These have a wide applicability in food-associated use. Their advantages are a low toxicity, low taint, wide microbicidal spectrum, and low cost. Their disadvantages are corrosivity to some metals (good-quality stainless steels are usually fairly resistant to corrosion) and ready inactivation of low-hypochlorite concentrations by organic matter. Hypochlorites are available as sodium hypochlorite, a liquid ('bleach') or sodium dichloroisocyanurate ('NaDCC') and chloramines, both soluble solids. Hypochlorites will decay in liquid solution, dependent on exposure to light, elevated temperature, and the presence of impurities. The solid hypochlorites are stable on dry storage.

Iodine

The advantages of iodine are a low toxicity, low taint, and a wide microbicidal spectrum. The disadvantages include some corrosion to metals (though less than hypochlorites), ready inactivation of low-iodine concentrations by organic matter, and their comparatively high costs. Relatively insoluble in water, iodine preparations are usually supplied as a complex with polyvinylpyrrolidone ('PVP') or solubilized with detergents. Iodine dissolved in potassium iodide ('Lugol's iodine') or alcoholic solutions ('tincture of iodine') are available for medical use but have no role in food hygiene.

Quaternary Ammonium Compounds (QACs)

The advantages of these are a low toxicity, low taint, cheapness, and noncorrosivity. They all possess some cleaning ability and are convenient to use. Their disadvantages are a ready inactivation by a wide range of materials and an often incomplete microbicidal spectrum. They are useful as general hygiene agents but must be used with care if disinfection is a critical step. Benzalkonium chloride is the most commonly used QAC.

See also: ***Bacillus***: Occurrence; **Cleaning Procedures in the Factory**: Types of Detergent; Overall Approach; ***Clostridium***: Occurrence of *Clostridium botulinum*; **Microbiology**: Detection of Foodborne Pathogens and their Toxins

Further Reading

Ayliffe GAJ, Coates D and Hoffman PN (1993) *Chemical Disinfection in Hospitals*, 2nd edn. London: Public Health Laboratory Service.

Bock SS (ed.) (1991) *Disinfection, Sterilization and Preservation*, 4th edn. Philadelphia, PA: Lea & Febiger.

Hobbs BC and Roberts D (eds) (1993) *Food Hygiene and Food Poisoning*, 6th edn. London: Edward Arnold.

Russell AD, Hugo WB and Ayliffe GAJ (1999) *Principles and Practice of Disinfection, Preservation and Sterilisation*, 3rd edn. Oxford: Blackwell Scientific Publications.

SAPONINS

G P Savage, Lincoln University, Canterbury, New Zealand

Background

Saponins are a heterogeneous group of glycosides that are widely distributed in plants of agricultural importance, particularly legumes. Many of these legumes are staple items of the human diet. Foods particularly rich in saponins are soya beans (*Glycine max*), chickpeas (*Cicer arietinum*), and beans derived from *Phaseolus vulgaris*. (*See* **Beans; Cereals**: Dietary Importance; **Legumes**: Legumes in the Diet; **Pulses; Soy (Soya) Beans**: Properties and Analysis.)

When saponins are agitated in water, they form a soapy lather. Other properties generally ascribed to this wide group of compounds are hemolytic effects on red blood cells, cholesterol-binding properties, and a bitter taste. These properties characterize particular types of saponins and are not necessarily shared by all members of the group. From a biological point of view, some of these properties are beneficial, whereas others are considered to be adverse.

Of particular interest is the observation that dietary saponins reduce plasma cholesterol levels in primates, thus having the potential to lower the risk of coronary heart disease in humans.

Chemical and Physical Properties of Saponins

Saponins consist of an aglycone unit linked to one or more carbohydrate chains (Figure 1). The aglycone or sapogenin unit consists of either a sterol or the more common triterpene unit. In both the steroid and triterpenoid saponins, the carbohydrate side-chain is usually attached to the 3 carbon of the sapogenin.

Saponins possess surface-active or detergent properties because the carbohydrate portion of the molecule is water-soluble, whereas the sapogenin is fat-soluble. The stability and strength of forage saponin foams are affected by pH, and this may have an effect on the development of bloat in ruminants. Saponins are remarkably stable to heat processing, and their biological activity is not reduced by normal cooking.

Isolation of saponins from plant material involves extraction with a polar solvent after removal of lipids, with petroleum ether or chloroform, followed by various purification techniques. A number of chromatographic procedures have been used to separate individual saponins. (*See* **Chromatography**: Principles.)

The analysis of saponins is complex and potentially subject to considerable errors during their isolation, separation and quantification stages. Thus, many early reports of the saponin content of food plants and processed food should be treated with caution. The data presented in Table 1 were obtained using rigorous methodology and are the most reliable data available at present. More recent studies have shown that many legume seeds contain several saponins, e.g., five different saponins have been separated from soya beans.

Biological Effects of Saponins

Ingested saponins can influence animal performance and metabolism in a number of ways.

Sensory Properties of Saponins

Recent studies have shown that a bitter or astringent taste is related to amounts of soya saponin I isolated from pea and soya flour. It is possible that saponins are a contributing factor to the undesirable organoleptic properties that humans associate with some legumes and legume products. The bitter taste of saponins may be responsible for the low palatability of lucerne to ruminants and may explain the reduced feed intake observed in many experiments when using this material as a forage. Many stock will discriminate against feeds containing high levels of lucerne meal.

Erythrocyte Hemolysis

Saponins have pronounced hemolytic properties when given intravenously, the degree of effect on the red blood cells varying among different mammalian species. The release of hemoglobin from red blood cells is the direct result of the interaction of saponins with membrane-bound sterols, which causes an increase in the permeability of the plasma membrane, bringing about the destruction of the cell. (*See* **Sensory Evaluation**: Taste.)

The hemolytic activity of saponins has been widely used as a means of detecting and assaying saponins in plant materials. The potentially toxic effects of

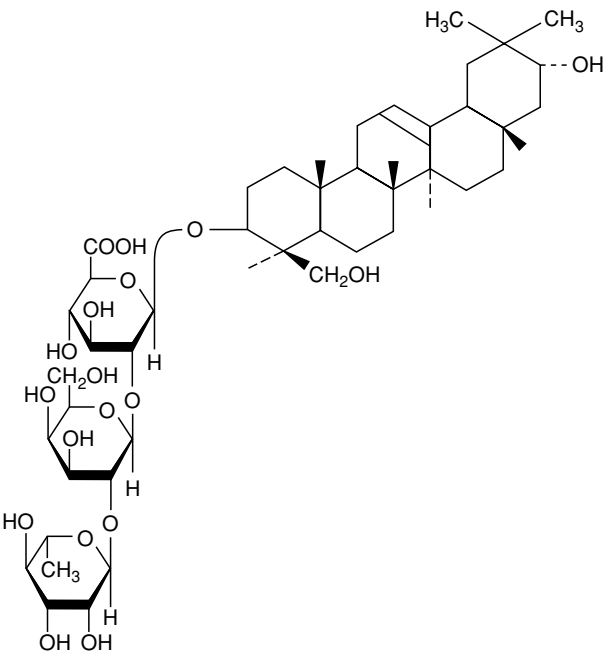

Figure 1 Structure of a typical saponin (from soya beans). Reproduced from Saponins, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 1 Saponin content of some legume seeds

Legume	*Saponin content (gram per kilogram of dry matter)*
Chickpeas (*Cicer arietinum* L.)	2.3
Green pea (*Pisum sativum*)	1.8
Haricot bean (*Phaseolus vulgaris*)	4.1
Kidney beans (*Phaseolus vulgaris*)	3.5
Lentils (*Lens culinaris* Medik.)	1.1
Mung bean (*Vigna radiata* L.)	0.5
Runner bean (*Phaseolus coccineus* L.)	3.4
Soya beans (*Glycine max* L. Merrill)	6.5
Yellow split pea (*Pisum sativum*)	1.1

intravenous injection of saponin extracts have resulted in this class of compounds being regarded as antinutritive factors in foods. Their low oral toxicity and the potentially useful nutritive value in foods have only recently been appreciated.

Effects on Blood and Tissue Cholesterol Levels

Since many legume saponins form insoluble addition complexes with cholesterol, the effects of dietary saponin on cholesterol might be expected. Saponin in the diet of chicks has been reported to reduce plasma cholesterol in cholesterol-fed animals. Saponin extracted from *Quillaia saponaria* has been shown to reduce liver (but not plasma) cholesterol levels in chicks. An exhaustive series of experiments, with a variety of saponins, have consistently demonstrated cholesterol-lowering effects. These include the feeding of lucerne saponins to rats, rabbits, and monkeys. Some workers have suggested that dietary saponins do not have any effect on plasma cholesterol concentrations. This is perhaps to be expected in experiments with rats that received no dietary cholesterol.

Experiments on humans have produced variable results. Different saponin levels in a dietary supplement based on soy flour had no effect on plasma cholesterol levels in hypercholesterolemic men. As the experiment was conducted on free-living subjects, it cannot be guaranteed that the subjects actually consumed the experimental diets.

In a more closely supervised trial, with subjects having normal blood cholesterol levels, it was found that a dietary supplement containing saponins did not have a significant effect on plasma cholesterol level. Increased fecal excretion of bile acids and neutral sterols was observed. Foods containing saponins, or diets supplemented with saponins, have been shown to reduce blood cholesterol levels in humans under conditions that would be expected to induce high levels of blood cholesterol. (*See* **Cholesterol**: Role of Cholesterol in Heart Disease; **Fats**: Digestion, Absorption, and Transport.)

Mode of Action

Saponins remain within the gastrointestinal tract. Some interact directly with dietary cholesterol, producing an insoluble complex that prevents the cholesterol being absorbed. Others appear to affect cholesterol metabolism indirectly by interacting with bile acids. An increased fecal excretion of bile acids is observed in response to feeding additional saponins in the diet. Bile acids thus diverted from the enterohepatic cycle would be replaced by hepatic synthesis from cholesterol. (*See* **Bile**.)

In the digestive tract, saponins form mixed micelles with cholesterol and bile salts; their hydrophobic triterpene or steroid groups stack together like small piles of coins. These micelles are then too large to pass through the intestinal wall.

Bile salts normally pass through the wall of the small intestine by both passive diffusion and active transport. Passive diffusion takes place along the entire length of the ileum and jejunum; active transport is confined to the terminal ileum. Saponins can interact with cell membranes, as shown by their hemolytic activity. It is possible that they may also have an effect on the cell membranes of the intestinal mucosa. However, the effects of saponins on both passive absorption and active transport can be

explained simply as being due to the reduction in the concentration of free bile acid, because bile acids assist lipid absorption. Low concentrations of free bile acids would also impair the efficiency of lipid absorption and presumably affect the absorption of fat soluble vitamins. Thus, there may also be a significant metabolic effect within the animal as well as in the digestive tract.

Effects on Growth

When lucerne saponin was added at high levels to the diets of monogastric animals, reduced feed efficiency and growth rates were observed. It is clear that species differ considerably in their responses to dietary saponin, e.g., poultry are more sensitive than rats. In contrast, soya bean saponins had little effect on the growth rate of experimental animals. Some saponins increase the permeability of the small intestinal mucosal cells, thereby inhibiting the active transport of some nutrients, but at the same time, they facilitate the uptake of materials to which the gut would normally be impermeable.

The biochemical mechanism that accounts for the growth-depressing effects has not been fully identified. In at least one experiment, the addition of 1% cholesterol to the diets of chicks completely overcame growth depression produced by 0.3% saponin.

It is possible that, in addition to their effects on lipid absorption, saponins also affect chymotrypsin and trypsin activity, which would affect the absorption of protein. (*See* **Protein**: Digestion and Absorption of Protein and Nitrogen Balance.)

In addition to observing reduced intestinal uptake of cholesterol in rat intestinal perfusates on the addition of soy saponin extracts, some workers have also observed a significant reduction in cholate and glucose uptakes. It was observed that the saponins could be washed out of the intestinal lumen, suggesting that inhibition, at least in the short term, was not caused by modification of, or damage to, the intestinal mucosa.

Metabolism of Saponins

Toxicity studies indicate that only very low levels of saponin absorption occur. Saponins are between 10 and 1000 times more toxic when administered intravenously than when given orally. Destruction of saponins in the digestive tract of both ruminants and monogastric animals has been observed. Saponins are degraded by rumen bacteria and by microflora found in the cecum of rats, mice, and chicks. Since saponin in the cecum is past the major sites of absorption, the release of sapogenins and sugars is considered to be insignificant.

Conclusions

The substantial evidence that saponins from a number of plant species can reduce plasma cholesterol levels in man is likely to encourage further interest in these plant foods. It is generally recognized that the overall nutritional value of many Western type diets would be considered improved if more legumes or legume-based products were consumed regularly. The acceptance of more legumes in Western type diets is limited by undesirable taste characteristics, some of which may be due to the higher levels of saponin found in many legume seeds.

There is little doubt that saponins can be incorporated into human diets at levels that can give a beneficial effect and would not entail a risk of acute toxicity. The fact that saponins can increase the permeability of intestinal mucosa raises the possibility of interesting nutritional and pharmacological uses.

See also: **Beans**; **Bile**; **Cereals**: Dietary Importance; **Cholesterol**: Role of Cholesterol in Heart Disease; **Chromatography**: Principles; **Fats**: Digestion, Absorption, and Transport; **Legumes**: Legumes in the Diet; **Protein**: Digestion and Absorption of Protein and Nitrogen Balance; **Pulses**; **Sensory Evaluation**: Taste **Soy (Soya) Beans**: Properties and Analysis

Further Reading

Anderson JO (1957) Effect of alfalfa saponin on the performance of chicks and laying hens. *Poultry Science* 36: 873–876.

Birk Y (1969) Saponins. In: Liener IE (ed.) *Toxic Constituents of Plant Food Stuffs*, pp. 169–210. New York: Academic Press.

Birk Y, Bondi A, Gestetner B and Ishaaya I (1963) A thermostable haemolytic factor in soybeans. *Nature* 197: 1089–1090.

Calvert GD, Bligh L, Illman RJ, Topping DL and Potter JD (1981) A trial of the effects of soya bean flour and soyabean saponins on plasma lipids, faecal bile acids and neutral sterols in hypercholesterolaemic men. *British Journal of Nutrition* 45: 277–281.

Cheeke PR (1971) Nutritional and physiological implications of saponins: a review. *Canadian Journal of Animal Science* 51: 621–632.

Fenwick DE and Oakenfull D (1983) Saponin content of food plants and some prepared foods. *Journal of the Science of Food and Agriculture* 34: 186–191.

Gestetner B, Birk Y and Tencer Y (1968) Soybean saponins. Fate of ingested soybean saponins and the physiological aspects of their hemolytic activity. *Journal of Agricultural and Food Chemistry* 16: 1031–1035.

Heaton KW (1972) *Bile Salts in Health and Disease*. Edinburgh: Churchill Livingstone.

Johnson IT, Gee JM, Price K, Curl C and Fenwick GR (1986) Influence of saponins on gut permeability and active nutrient transport *in vitro*. *Journal of Nutrition* 116: 2270–2277.

Malinow MR, McLaughlin P, Kohler GO and Livingston AL (1977) Prevention of elevated cholesterolemia in monkeys by alfalfa saponins. *Steroids* 29: 105–110.

Malinow MR, Connor WE, McLaughlin P *et al*. (1981) Cholesterol and bile acid balance in *Macaca fascicularis* – effects of alfalfa saponins. *Journal of Clinical Investigation* 67: 156–162.

Oleszek WA (2002) Chromatographic determination of plant saponins. *Journal of Chromatography A* 967(1): 147–162.

Pathirana C, Gibney MJ and Taylor TG (1980) Effects of soy protein and saponins on serum, and liver cholesterol in rats. *Atherosclerosis* 36: 595–599.

Pistelli L, Bertoli A, Lepori E, Morelli I and Panizzi L (2002) Antimicrobial and antifungal activity of crude extracts and isolated saponins from *Astragalus verrucosus*. *Fitoterapia* 73(4): 336.

Potter JD, Illman RJ, Calvert GD, Oakenfull DG and Tapping DL (1980) Soya saponins, plasma lipids, lipoprotein and fecal bile acids. *Nutrition Reports International* 22: 521–528.

Price KR, Curl CL and Fenwick GR (1986) The saponin content and sapogenol composition of the seed of 13 varieties of legume. *Journal of the Science of Food and Agriculture* 37: 1185–1191.

Sidhu GS and Oakenfull DG (1986) A mechanism for hypocholesterolaemic activity of saponins. *British Journal of Nutrition* 55: 643–649.

Sirtori CR, Agradi E, Conti F, Mantero O and Gatti E (1977) Soybean protein diet in the treatment of type II hypercholesterolaemia. *Lancet* I: 275–277.

Yayla Y, Alankus-Caliskan O, Anil H, Bates R, Stessman C and Kane V (2002) Saponins from *Styrax officinalis*. *Fitoterapia* 73(4): 320.

Sardines *See* **Fish**: Introduction; Catching and Handling; Fish as Food; Demersal Species of Temperate Climates; Pelagic Species of Temperate Climates; Tuna and Tuna-like Fish of Tropical Climates; Demersal Species of Tropical Climates; Pelagic Species of Tropical Climates; Important Elasmobranch Species; Processing; Miscellaneous Fish Products; Spoilage of Seafood; Dietary Importance of Fish and Shellfish; **Fish Farming**; **Fish Meal**

SATIETY AND APPETITE

Contents

Food, Nutrition, and Appetite

D A Booth, University of Birmingham, Edgbaston, Birmingham, UK

Introduction

Appetite for food and drink is the momentary disposition of an individual to seek and ingest edible or potable materials. The concept of having an appetite for food and drink has been widely misunderstood. Appetite has often been assumed to be a subjective phenomenon, private to a person's consciousness. The view has been held that appetite can be measured by a rating procedure, such as marking a visual analog scale (a line labeled at each end, e.g., with the phrases, 'extremely hungry' and 'not at all hungry'), or magnitude estimation (assigning numbers to the strengths of a named aspect of awareness).

A measure of appetite, whether based on ratings, behavioral patterns, or intakes, is an objective

estimate of the strength of influence of specified current determinants on the disposition to seek and consume food. Appetite is a publicly observable relationship between the tendency to ingest and the sensed food composition, the culturally interpreted context, and the bodily state of the eater or drinker.

Appetite has also been restricted to facilitatory influences that come from food, or at least from outside the skin. In such terminology, facilitation from under the skin is often called hunger (and only the inhibition of eating from truly physiological sources is called satiety). Alternatively, appetite is in the mind, and hunger in the body.

However, appetite is not one sort of influence on eating behavior – external, mental, or other. Ingestive appetite is the causal structure by which all influences are momentarily affecting eating and drinking. It is sometimes necessary to distinguish appetite for food (hunger) from appetite for drink (thirst). The subjective experience of hunger can be referred to external sources, as in a food craving or the desire to eat at the usual time, not just to bodily sensations such as the epigastric pang.

Aspects of Appetite

The immediate causes of eating and drinking might be crudely divided into three categories, which can be dubbed sensory, somatic, and social. This subdivision is not entirely coherent, however, because appetite is an integral whole. Not only do sensed food factors interact among themselves in the consumer's mind but also the sensory configuration interacts with physiological factors and economic, cultural, and interpersonal factors, in determining a momentary decision to accept an item of food or drink.

It is therefore scientifically and practically unwise to study one factor by itself, especially out of the normal context of eating (or of food purchase or preparation). Moreover, it cannot be assumed that sensory, social, and somatic groups of factors independently add into the disposition to eat or purchase a food. Thus neither theory nor application is sound if based solely on economics or physiological or sensory research. (*See* **Food Acceptability**: Affective Methods.)

Sensory Aspects of Appetite

The influences on eating or drinking that arise immediately from the sensing of the food or drink have traditionally been called palatability. However, incorrect assumptions are made in many uses of this term; one such assumption is that palatability is an invariant property of the food or drink itself. Palatability is not inherent to the food; it is an effect of the food on the eater. This effect can vary, not just between people, but within a person in different contexts of eating.

The ordinary person regards foods as having constant palatabilities, even if varying from person to person. It is natural to deny that one stopped eating because the food became less palatable; cessation of eating is generally attributed to feeling full or to having had an appropriate amount to eat.

Many scientific investigations and theories have also treated palatability as a constant sensory effect. In physiological psychology, for example, the standard account of the control of meal size is that accumulating satiety subtracts from constant palatability until insufficient facilitation of eating exists to continue the meal. It is more likely, however, that the sensory facilitation differs between the contexts of starting and ending a meal. When the sensory facilitation of eating is actually measured during or from before to after meals, it typically decreases to a minimum at the end of eating and for some while afterwards. Moreover, at the anecdotal level, a savoury course is not expected to be as palatable after a sweet dessert as it would be before the dessert.

Measurement of sensory preferences Any measure of the sensory aspect of appetite must compare the responses to two or more foodstuffs differing in known sensed characteristics, unconfounded by any other differences that might affect the responses. Whether the responses are concrete (such as selection among the foods) or symbolic (such as numerical or line rating of liking, pleasantness, or likelihood of choice), and whether the food samples are presented simultaneously (e.g., triadic test) or successively (i.e., monadically), the relative acceptances give an estimate of the sensory preference of that assessor in the context of testing.

Attention must be given to the basic psychological mechanisms operative in sensory effects on food choice. The most preferred version of a food for a consumer in a given context is always a particular physicochemical configuration of its sensed characteristics. In fact, the individual's ideal point is a more precise sensory level than a descriptive verbal anchor (such as 'extremely strong') and is often no worse than a familiar physical standard. Taken with a second sensory anchor, such as apparent absence of the characteristic or perhaps a presence so weak (or so strong) as to be unacceptable, the ideal point defines a sensory scale as objectively and precisely as any purely descriptive rating. The crucial requirements are that each tested sensory level is described by the assessor or identified by the investigator as above or below the ideal level and that the preference-anchored responses are scored according to

sensory level, not for degree of preference. (*See* **Sensory Evaluation**: Sensory Characteristics of Human Foods; Food Acceptability and Sensory Evaluation.)

The raw sensory preference scores and the slopes of plots of these scores against physical levels are as meaningful as conventional descriptive scores and slopes (or psychophysical power function exponents). They can be averaged across a panel and tested statistically for differences between samples. This amounts merely to data analysis, however; it is not scientific measurement of the strength of influences on food perception or appetite. The most direct step towards efficient design of palatable foods, rather, is to use the effect of levels on score to measure causal strength: the sensitivity of each assessor's preference to a physically measured factor should be calculated from the slope and residual variance of a linear plot of responses against sensory levels, using general psychological theory of the mechanisms by which different responses are given to different signals (Figure 1). (*See* **Sensory Evaluation**: Sensory Rating and Scoring Methods.)

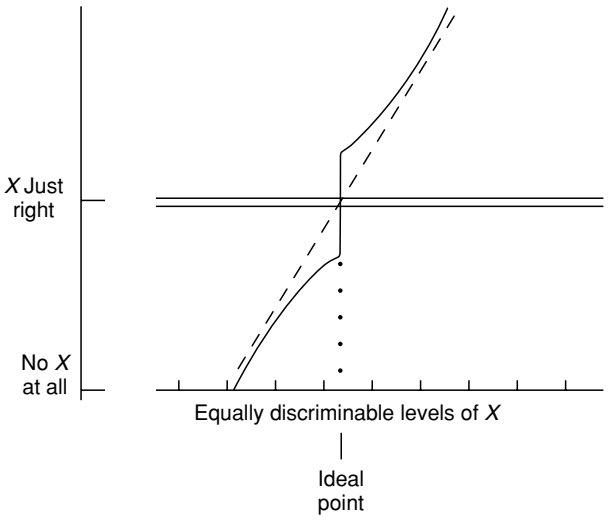

Figure 1 Effect of equally discriminable differences in level of a described sensory or other influence on appetite on the strength of that factor, rated from too weak to be recognized to ideal strength, in a context where other factors are at inappropriate levels. The response line is in the Euclidean distance of the factor level from ideal, i.e., the square root of the sum of the square of the difference of a test sample from the influence's ideal level and a constant that represents the square of the size of the contextual defect; its discontinuity around the ideal point is the truncation of the preference peak by this defect. Reproduced from Satiety and Appetite, Food, Nutrition, and Appetite, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

A great advantage of this approach is that sensitivity of perception or preference to each investigated factor can be estimated when the other foods, or the social and somatic contexts of the test, depart moderately from the combination in which the tested food is normally eaten or purchased. If tolerably low and high levels of a factor are well and equally represented among the responses, the ideal level and the sensitivity can be interpolated with fair accuracy. Moderate inappropriateness in the context of testing creates a Euclidean discontinuity which is virtually confined to the midrange of the theoretical linear function (Figure 1).

The sensitivities of a preference response to different factors can be used to identify how those factors interact in the individual's mind in the test situation, potentially enabling understanding of how a food works in dietary habits. Preference characteristics can also be aggregated across a panel of representative consumers. This could be used to give a physically and behaviorally objective market-response surface from which one or more popular products could be designed.

Somatic Aspects of Appetite

Somatic influences on appetite include physiological signals of hunger, thirst, and satiety from the digestive tract or from tissues such as the liver or the brain itself, as well as these and other postingestional effects of foods that induce sensory preference or aversion.

As already indicated, even a physiological signal of hunger or satiety (such as gastric motility or distension) is likely to operate through particular sensory configurations when the eating situation is familiar. There is evidence that appetite from an empty stomach is stronger for an entrée than a dessert, and vice versa for a partly filled stomach. Similarly, we should expect conventions about foods that are appropriate to particular times of day (such as breakfast) to affect hunger from an empty upper intestine or the satisfaction of appetite arising from stimulation of the intestinal wall or hepatic metabolism, although such hypotheses remain to be tested physiologically. Investigation of behavior at ordinary meals is necessary to elucidate the normal operation of somatic factors in appetite. Evidence for particular physiological signals has been obtained with unusual foods at imposed times but that leaves open the role of a signal in everyday eating and drinking.

Those physiological factors which alter food preferences by associative conditioning or other forms of learning are interacting with sensory factors. It must be noted that they are interacting quite differently

from the contemporaneous mixing of factors considered elsewhere in this article. Indeed, a preference acquired by the association of certain sensory characteristics with a physiological consequence might be unaffected if the conditioning physiological event recurs when the conditioned food stimuli are also present. In psychology, this distinction in mechanisms is referred to as the difference between reinforcement and motivation, or between learning and performance. The distinction is crucial to the understanding of learned behavior such as normal appetite.

The same metabolic signals that inhibit appetite may also induce preferences for food flavors and textures presented shortly beforehand. This is the main basis for the desire for familiar foods, ranging from the appetite for a staple food, such as potatoes or bread, to the craving for an indulgence item such as chocolate. Association with these reinforcers enables bodily states (such as a partly full stomach) and external situations (such as lunchtime, or a party atmosphere) to acquire a learned power to motivate or de-motivate eating or drinking.

The temporal pattern of appetite suppression following the consumption of a particular type of menu also appears to be learnt. Such expectations of satisfaction may be at least as important as physiological signals in the sensation of fullness and in the rise of appetite before the next meal.

Additional physiologically mediated psychological effects of foods are suspected to condition sensory preferences or to create functional expectations of sensorily identified foodstuffs or beverages, although conclusive evidence is lacking. These appetite-reinforcing somatic factors include the stimulant action of caffeine and the sedating action of a heavy meal or of alcohol. However, cultural stereotypes and personal experience can create stimulant, soothing, or cheering effects from consumption of a drink or food, without any specific physiological basis.

A somatic factor in appetite can only be identified if the study disconfounds variations in the physiological effect from variations in both social and sensory effects. Neglect of this point has led to ill-founded claims of measuring physiologically specific aspects of appetite, such as carbohydrate craving, protein selection, protein- and fiber-induced postingestional satiety signals, and gastric distension satiety. The technical difficulties are daunting, but there is no gain in knowledge from neglecting these variations.

All relevant sensory and social factors, as well as somatic factors, must be specifically manipulated and/or measured as perceived by the individual, and shown to be uncorrelated with the physiological effect(s) of interest. Only then can the sensitivity of appetite to that somatic factor in that context be estimated, or indeed even a qualitative effect of the factor be established.

Social Aspects of Appetite

Interpersonal, cultural, and economic factors in appetite appear to influence food choices by modulating the effects of sensory factors. It may be feasible to investigate a range of foods or drinks for general effects of factors such as price, declared sweetener type, or use by children, on appetite at the point of purchase, serving, or ingestion. However, effects may be specific to a food.

Social influences on appetite are often present in a symbolic form, e.g., the category of eating occasion or type of company at the meal, a brand logo, an advertising picture, compositional information, or the price label attached to a food item. Social factors do not differ from rated sensations generated by foods or in the body: descriptive scores that relate to physical factors can be used to model sensory and somatic aspects of appetite; hence the procedure for scaling a social factor and estimating its strength of influence on appetite is the same in principle as that for other factors.

Psychology, Society, and Nature

Appetite is a psychological phenomenon, an aspect of the mental processes organizing an individual's behavior. Hence the cognitive approach to food consumption should be capable of yielding general explanations of the anthropology of cuisine and the microeconomics of the food market. The constructs about cooking and eating held in common among individuals from a social group are evidence of their shared culinary culture. The effects of price, pack label, and shelf display on the personal disposition to purchase a food item will aggregate across consumers into sales of that brand in the shops. The performance of selective eating behavior provides challenges to scientific explanation that go beyond the physics and chemistry of foods and their biochemical and physiological effects. Each individual has his or her own causally structured behavior beyond the food materials, the nutritional physiology, and the socioeconomic systems.

Development of the quantitative behavioral science of food choice and intake is needed. With fundamental understanding of the determinants of eating habits at the level of the causal processes in the mind of the consumer, we would be able to bring together knowledge of the socioeconomic, physicochemical, and biochemical–physiological processes into a coherent

science and effective practice of food design and dietary health.

See also: **Food Acceptability**: Affective Methods; **Sensory Evaluation**: Sensory Characteristics of Human Foods; Food Acceptability and Sensory Evaluation; Sensory Rating and Scoring Methods

Further Reading

Axelson ML and Brinberg D (1989) *A Social–Psychological Perspective on Food-Related Behavior*. New York: Springer-Verlag.

Barker LM (ed.) (1982) *The Psychobiology of Human Food Selection*. Westport, Connecticut: AVI Publishing.

Bennett GA (1988) *Eating Matters: Why We Eat What We Eat*. London: Heinemann Kingswood.

Berthoud H-R and Seeley RJ (eds) (2000) *Neural and Metabolic Control of Macronutrient Intake*. Boca Raton: CRC Press.

Boakes RA, Burton MJ and Popplewell DA (eds) (1987) *Eating Habits: Food, Physiology and Learned Behaviour*. Chichester: John Wiley.

Bolles RC (1980) Historical note on the term 'appetite'. *Appetite* 1: 3–6.

Booth DA (ed.) (1978) *Hunger Models: Computable Theory of Feeding Control*. London: Academic Press.

Booth DA (1994) *Psychology of Nutrition*. Bristol, PA & London: Taylor & Francis.

Booth DA, Mobini S, Earl T and Wainwright CJ (2003) Consumer-specified instrumental quality of short-dough cookie texture using penetrometry and break force. *Journal of Food Science* (in press).

Chivers DJ, Wood BA and Bilsborough A (eds) (1984) *Food Acquisition and Processing in Primates*. New York: Plenum Press.

Dobbing J (ed.) (1987) *Sweetness*. London: Springer-Verlag.

Farb P and Armelagos G (1980) *Consuming Passions: the Anthropology of Eating*. Boston: Houghton Mifflin.

Fieldhouse P (1986) *Food and Nutrition: Customs & Culture*. London: Croom Helm.

Freeman RPJ, Richardson NJ, Kendal-Reed MS and Booth DA (1993) Bases of a cognitive technology for food quality. *British Food Journal* 95(9): 37–44.

Logue A (1991) *The Psychology of Eating and Drinking*, 2nd edn. New York: WH Freeman.

Meiselman HL (ed.) (2000) *Dimensions of the Meal. The Science Culture, Business and Art of Eating*. Aspen, CO: Aspen Publishers Inc.

Murcott A (ed.) (1983) *The Sociology of Food and Eating*. Cardiff: Gower Press.

Ramsay DJ and Booth DA (eds) (1991) *Thirst: Physiological and Psychological Aspects*. London: Springer-Verlag.

Solms J and Hall RL (eds) (1981) *Criteria of Food Acceptance: How Man Chooses What he Eats*. Zurich: Forster.

Solms J, Booth DA, Pangborn RM and Raunhardt O (eds) (1987) *Food Acceptance and Nutrition*. London: Academic Press.

The Role of Satiety in Nutrition

D A Booth, University of Birmingham, Edgbaston, Birmingham, UK

Introduction

Satiety, a transient lack of interest in further ingestion, is a dispositional state of the individual. This state may limit how much food and drink is consumed on one occasion. It may also delay the next occasion of consumption. The satiating effect of a food is thought to contribute to that food's acceptability, by serving as part of the satisfaction to be gained from eating it. However, there are few relevant hard data, and thinking on this matter is confused.

Scientific Concept of Satiety

Satiety is specifically the inhibitory effect of dietary consumption on appetite. The decrease in hunger or thirst must, by definition, have been caused by some consequence of ingestion. In studies of satiety it is not sufficient to record satiety ratings alone. These may be graded expressions of the sensation of abdominal fullness, the wish to eat only a small amount, an awareness of how much was eaten and how recently, or some other measure of the disposition to refuse food. The origins of the lack of appetite must be shown to include an effect of eating for the rating to be a true measure of satiety. (*See* **Satiety and Appetite: Food, Nutrition, and Appetite.**)

Scientific analysis of satiation requires identification of the influence(s) from eating that reduce appetite. Satiety is not a response but the operation of a type of mechanism.

What Role does Satiety have in Food Acceptability?

It is reasonable to expect that short-term satisfaction of hunger by particular foods increases the acceptability of those foods. Another attraction of a food or meal could be a longer postponement of the need to eat. Understanding satiety mechanisms could be considered to have broad relevance to the formulation and marketing of foods.

However, there has so far been little scientific investigation in humans of the role of satiety in palatability and choice of foods. A considerable body of data in experimental animals has shown that the satiating effect of eating food carbohydrate induces sensory preferences for foods eaten shortly before

absorption of the carbohydrate begins. Sensory preferences are also acquired by association with the effects of absorption of amino acids and of digested fat, under conditions where it is less clear that these nutrients are contributing to satiety. Experiments in humans on the acquisition of sensory preference by association with carbohydrate or protein also indicate that satiety may influence the learning of palatability. (*See* **Sucrose**: Dietary Importance.)

What Role does Satiety have in Weight Control?

Satiety is widely assumed to play a major role in weight control. However, there is a dearth of data on the mechanisms by which satiety can inhibit food consumption in the short term, or on how such effects influence energy intake and long-term energy balance. Research methods are not generally attuned to the concept of satiety as a disposition integrated from sensory, physiological, and cultural mechanisms, or to daily energy intake as an epiphenomenon of diverse sequences of complex dietary selection decisions. (*See* **Obesity**: Treatment.)

As a result, satiety mechanisms are not effectively exploited in established ways of attempting weight control. This is a potential explanation for the fact that obesity has not been prevented, and generally reliable procedures have not been found to reduce unhealthy overweight. There are sound reasons for expecting satiety to have some part to play in weight control. The satiating effects of foods or meals could operate within the framework of certain patterns of eating in ways that help to reduce habitual daily intake, making excess energy intake over expenditure less likely.

Understanding satiety mechanisms is crucial for extrapolating to common eating patterns in order to identify foods and drinks that support successful long-term weight control. Unfortunately, most frequent research on human satiety has attempted only to test foods and food components for effects on appetite ratings and nutrient intake, without measuring the foods' mechanisms of action or their effects on eating habits. As a result, the role of satiety in weight control remains largely a matter of speculation.

Food-Specific Satiety Mechanisms

Satiety is essentially sensory: it is a disposition to reject foods or to accept them in limited amounts. In principle, inhibition of eating could be similarly strong for all foods. Alternatively, it is conceivable that the strength of inhibition may vary with the sensed identity of the food. Such satiety with a degree of specificity to a food is called sensory-specific satiety. The Greek form of this term was applied to a taste-selective satiety that is claimed to depend on the nutrient consumed: negative alliesthesia (literally, a change in sensation – negative in the sense that the food is less attractive). The sensory impression does not necessarily change; the loss of interest is not because the taste seems weaker or different in gustatory quality. The change is in the reaction to the taste (or other sensory characteristics of the foodstuff).

There is evidence from experimental animals that hypertonic sugar solutions (which can be nauseating in people) cause a temporary loss of the sensual pleasure obtained from sweetness. In humans, however, suppression of a pleasurable reaction to foods has yet to be distinguished from a reduced disposition to consume them. Whether eating is motivated only by pleasures and pains, or by other things as well, is an empirical issue for cognitive psychology.

There are at least two sorts of food-specific satiety. One is caused by immediately preceding exposure to food having the consequently satiated sensory characteristics. The other is a learnt reaction to those sensory characteristics in conjunction with a particular physiological signal, such as a partly filled stomach: it is not necessary for the particular food to have been ingested previously in that meal; the reaction is acquired from experience at one or more earlier meals.

Habituation Satiety

Satiation for a food induced by recent exposure to that food meets the basic criteria for the behavioral change known as habituation. This is a decline in a specific response (eating) during and shortly after repeated or continuous exposure to a specific stimulus (the food). The decline is not fatigue in the sensory pathways or in the eating movement circuits but involves a decline in effectiveness of central connections between these two neural systems. At the sensory end the rated strength of the sensation of, for example, flavor does not decrease, at least not by as much as the decrease in the rated pleasantness of eating the food.

Introduction of a new stimulus disrupts habituation and the original response reemerges, an effect called dishabituation. In the case of satiety, this disruption is known as the variety effect. Presenting a food that was not in the preceding meal can reactivate eating and its rated pleasantness. Switching foods can provoke eating a larger meal than switches between samples of the same food in experiments with rats and with human subjects.

If the larger meal is compensated by omission of a subsequent snack or by a smaller following meal,

body weight should not be affected. There is no conclusive evidence that marketing a wide variety of flavor variants contributes to obesity.

The mental processes involved in habituation satiety remain to be elucidated. Boredom with a food seems to be more complex than mere fatigue in a central neuronal connection; yet it could still be based on a subconscious decline in a particular sensory facilitation of eating.

Habituation in invertebrates is mediated by a decline in effectiveness of synapses on the motor neurons from the repeatedly stimulated sensory pathway. Similarly, taste-specific satiation in monkeys involves no changes in the taste pathways through to the cerebral cortex: changes in line with the satiated behavior occur only beyond the sensory cortex, towards the food-selection pathways.

Habituation also has a long-term component. The short-term habituated response recovers completely within an hour or so of the end of repeated stimulation. When the same stimulus is repeatedly presented on a later occasion, the decline in response is more rapid. Long-term habituation (i.e., more rapid rehabituation) may contribute to boredom with the same meal presented day after day.

There are indications of the more complex processes in the attractions of long-term variety. While the main item that characterizes a course in a meal might have to be varied, the same staple food (such as potatoes or bread) can be eaten day after day. In addition, drinks seem to show less item-specific satiation than many foods. Conventions about menus or personal dietary habits may contribute to cessation of eating one food or to starting to eat a different food.

Famine relief may provide only two or three foodstuffs, albeit in ample quantities. This diet might become tedious even to the chronically hungry who are being offered culturally appropriate staples. There is some evidence that intake increases when a variety of flavorings is provided with the basic commodities.

Conditioned Satiety

The second known type of food-specific satiety has been learned from experiences in the past. Satiety learning processes have the basic form of associative conditioning – stimuli elicit a reaction that they have not previously evoked, because they have been associated with consequent stimuli. Repetition is not enough; the stimuli must be paired with conditioning consequences.

The Russian behavioral physiologist Pavlov demonstrated conditioning of appetite when the ringing of a bell in hungry dogs was paired with feeding or a food taste; after that, ringing the dinner bell when the dogs had empty stomachs elicited salivation and orientation to the food provider or food bowl before the food was presented. The bell and the empty stomach predict the taste of food. The dog associated this cue or conditioned stimulus (CS) with the consequence or unconditioned stimulus (US).

The conditioning of satiety follows the same principles. A food stimulus and a partly filled stomach (the CS) are paired with subsequent strong satiation, such as a mildly bloated sensation (the US). When the food or flavor is presented again later, while the stomach is just as full, the food flavor has become less attractive or even slightly repulsive. To the extent that this loss of interest comes to consciousness, it will be expected that particular foods or meals will be regarded as satisfying.

A key difference from habituation satiety is that conditioned satiety does not depend on the sensory characteristics of the food that has just been eaten. The loss of interest can apply to a food that had not been presented before on that particular occasion, because the satiety had been learned on earlier occasions. All foods to which satiety has been previously conditioned will become less attractive when the stomach begins to fill.

Integration of Satiety Processes

Conditioned elicitation of loss of interest in eating by cues from food and from the digestive tract suggests that satiety in familiar situations is a learnt response. If internal signals can become part of the food-specific satiating pattern, external cues such as emptied plates may also come to suppress appetite.

Experiments on satiety risk missing the normal mechanisms of satiety when they impose unfamiliar or extreme conditions. Eating would stop when visceral or social discomfort is strong enough, independently of prior learning. Therefore, to elucidate postingestional and social satieties, the role of any one satiety mechanism must be examined in its normal range and within the learnt complex a person has established for his or her dietary habits.

Postingestional Satiety Mechanisms

The effects of food arise in sequence as the ingested food passes down the digestive tract, from sight and aroma, through texture, temperature, irritation, taste, and retronasal aroma in the mouth, to stimulation of the gut wall, and the stimulation of tissues such as the liver and the brain itself as the food is absorbed. The traditional assumption in the physiology of ingestion has been that inhibitory stimuli subtract from facilitatory stimuli. Intake of sugar solutions in rats has an intestinal effect which

subtracts from the palatability of the sweetness; however, these fluids are not familiar parts of the subjects' diet.

Linear or additive interactions are not the norm in neurophysiological studies, or in experiments on the interactions of mechanisms in the ingestion or rating of ordinary foods and drinks. There is no convincing evidence for interactions among satiety mechanisms that justify the use of the term cascade. The evidence is consistent with satiety arising from recognition of the accustomed pattern of eating, postingestional stimulation, and social constraints, with reduced satiety in the absence of any one element in that learnt pattern.

Cultural and Interpersonal Satiety Mechanisms

Appropriate proportions of food items on a plate (e.g., meat and vegetables) and the total amount (e.g., for a man or at the price) are subject to cultural variation, although these in turn may be influenced by average physiological requirements.

The process of eating in company can influence food intake. For example, a dieter may inhibit eating. This social satiety for food is likely to have been integrated with physiological and sensory satiety, at least in situations about which sufficient has been learned from personal experience.

Satiety Disruption in Dieting

Reducing dietary intake may evoke desires for more food and, possibly, cravings for forbidden foods. Cognitive bases for food cravings are more likely than hormonal imbalances or nutrient deficiencies.

Small meals are less likely to create satiety and will allow the hunger pattern in the viscera and within the memory of that meal to form sooner than after a larger meal. Sufficiently severe restriction of food at meals, even when the gut and the tissues adapt, will activate innate mechanisms of distress, readily labeled hunger by the dieter. The temptation to look for food will be increased, as will the temptation to eat when food is available.

Satiety Manipulation in Weight Control

Permanently successful weight control will depend on keeping satiety as normal as possible in the new lifestyle required to maintain target weight.

Satiety Synapses and Antiobesity Drugs

The pathways in the brain over which the inhibitory effects of food on eating are transferred are poorly understood. Satiety signals, such as distension of the stomach, chemical stimulation of the upper intestine, and oxidation of energy substrates in the liver, have been investigated in isolation from stimuli from food. Reactions to these food stimuli constitute the disposition to eat that satiety suppresses. Neural signals from the viscera and sensors of the circulation at the surface of the brain have their first relay at the back of the brainstem. It is not known where or how they interact there and higher in the brain with the sensory control of eating in the inhibition of appetite.

Visceral signals are relayed to the hypothalamus at the head of the brainstem; this has complex connections with the forebrain, where incoming and past information from different sources is put together. For example, gastric distension affects neuronal activity in the ventromedial hypothalamic nucleus. Yet this is not evidence that this region mediates satiety, because the hypothalamus has many autonomic and endocrine functions to perform. These are liable to affect eating behavior indirectly (via neural circuitry elsewhere in the brain). In fact, it has been found that the ventromedial hypothalamic region is not crucial to any of the appetite-suppressant effects of food that have been tested.

These tests were carried out to evaluate the theory that the obesity seen after destruction of this region of the hypothalamus arises from a defect in satiety resulting from destruction of 'the satiety center.' Since satiety mechanisms operate normally in rats made fat by lesions in this area, the hypothalamus is unlikely to contain any such center.

The effects of drugs on the sizes of meals in experimental animals have inspired theories that particular monoamine neurotransmitters operate in the pathways on the brain that mediate satiety. A region of the hypothalamus forward from, and above, the originally supposed satiety and appetite centers appears to contain some of the synapses at which transmission alters meal sizes.

Norepinephrine (noradrenaline) appears to be the transmitter at synapses in this region that increases meal sizes and even restarts eating in a satiated animal. Analysis of the different possible mechanisms implicates a disruption of conditioned satiety.

However, it may be a localized effect of the arousal system that has norepinephrine (noradrenaline) synapses in many brain regions. This may be experimental isolation of the effect of distraction or excitement on normal moderation of eating. Conceivably, therefore, disruption of learnt satiety is related to emotional eating. This is amenable to psychological management and does not need drug treatment.

Drugs that activate synapses using serotonin (5-hydroxytryptamine; 5-HT) decrease the size of the first meal that a rat takes after it has been deprived

of food for a while. This has been interpreted as an augmentation of satiety. However, starved rats tend to eat fast and long at this initial meal. Serotonin is a transmitter in the brainstem circuits that organize chewing and other tactually guided movements. Hence these drugs may simply be slowing movement or making chewing and swallowing more difficult; they have been shown to alter the effects of food textures on preference and intakes in rats. Serotonin is also important in sleep circuits; indeed, when these drugs are having their greatest impact on people's eating, they are also having their strongest sedative effect. The classic meal-size decrease in these circumstances may have nothing to do with satiety.

Serotonergic drugs that have been used to aid rapid weight reduction may exert part of their effect peripherally. Serotonin is one of the most important transmitters in local neural control of the contractions that pass food down the digestive tract. These and other drugs therefore slow gastric emptying under some conditions, prolonging the satiating action of food and reducing the temptation to snack. They may even postpone the next meal.

Satiety in Dietary Treatment and Advice

There is a medical consensus that energy from starchy foods should not be decreased. Where energy is needed to replace that removed by reduction in fat intake, starch intake should be increased. Readily assimilated carbohydrate, particularly in the earlier stages of meals, is likely to play a major role in immediate satisfaction following the meal. This function and its training of expectations and palatability are important for acceptability of the diet.

In addition, intake of low-energy bulky foods, such as pulses, grains, and vegetables and fruit that provide nonstarch polysaccharides, is recommended. The satiating effect of normal amounts of fiber is probably mediated by learning by association with intestinal and metabolic effects such as those of glucose and fructose from the starch and sugar present in high-fiber foods. Therefore the sugar that is often required to make high-fiber foods edible also has a role in their satiating effects. Alternatively, some of the starch may be rapidly assimilated. However, much of the starch will be in slowly assimilable forms, potentially delaying the return of hunger.

Enthusiasm for a nutrient, a diet product, or a dietary regimen by a professional or in another dieter can convey itself to the dieter or at least lead to it being tried in desperation. For example, it seems obvious that high-fiber foods are filling, and these foods have acquired a nutritious image. This sensory satiety can suggest visceral satiety, and these expectations may be fulfilled as long as sufficient energy is included with the fiber. The assumption is that sufficiently high doses of fiber augment satiety by postingestional mechanisms such as slower gastric emptying and slower absorption of nutrients through soluble fiber gels in the small intestine. These theories remain to be critically tested, by studies of normal eating with sensory factors dissociated out.

Satiety in Psychological Weight Therapy

Eating slowly is a standard part of behavior modification for weight reduction, although its effectiveness has not been evaluated.

Part of the rationale has been to 'satisfy the taste buds.' Perhaps savoring all the food that is eaten helps to ward off the sense of restriction that makes forbidden eating more tempting. Another idea is that more prolonged sensing would augment food-specific satiety.

A different rationale for slow eating has been that it would fill the stomach more effectively and so satiate with less food. However, there is no evidence that gastric distension plays such a precise or dominant role in meal termination. Furthermore, if the eating were as slow as gastric emptying, the stomach would not begin to fill.

A variety of other suggestions have been made as to how to strengthen the influence of satiety but none has yet been taken up seriously, let alone tested for efficacy.

See also: **Obesity**: Treatment; **Sucrose**: Dietary Importance

Further Reading

Bennett GA (1986) Behavior therapy for obesity: a quantitative review of the effects of selected treatment characteristics on outcome. *Behavior Therapy* 17: 554–562.

Berthoud H-R and Seeley RJ (eds) (2000) *Neural and Metabolic Control of Macronutrient Intake*. Boca Raton: CRC Press.

Bolles RC (ed.) (1991) *The Hedonics of Taste*. Hillsdale, New Jersey: Erlbaum.

Booth DA (1994) *Psychology of Nutrition*. London: Taylor & Francis.

Booth DA (1998) Promoting culture supportive of the least fattening patterns of eating, drinking and physical activity. [Prevention of Obesity: Berzelius Symposium and 8th International Congress of Obesity Satellite.] *Appetite* 31: 417–420.

Booth DA, Rodin J and Blackburn GL (eds) (1988) *Sweeteners, Appetite and Obesity (supplement to Appetite)*. London: Academic Press.

Capaldi ED and Pawley TL (1990) *Taste, Experience, and Feeding*. Washington, DC: American Psychological Association.

Craighead LW, Stunkard AJ and O'Brien RM (1981) Behavior therapy and pharmacotherapy for obesity. *Archives of General Psychiatry* 38: 763–768.
Dobbing J (ed.) (1989) *Dietary Starches and Sugars in Man: A Comparison*. London: Springer-Verlag.
Legg C and Booth DA (eds) (1994) *Appetite: Neural and Behavioural Bases*. Oxford: Oxford University Press.
Stunkard AJ (ed.) (1980) *Obesity*. Philadelphia: WB Saunders.
Wadden TA, Sternberg JA, Letizia KA, Stunkard AJ and Foster GD (1989) Treatment of obesity by very low calorie diet, behaviour therapy, and their combination: a five-year perspective. *International Journal of Obesity* 13 (suppl. 2): 39–46.

Sausages *See* **Meat**: Sources; Structure; Slaughter; Preservation; Eating Quality; Sausages and Comminuted Products; Analysis; Nutritional Value; Hygiene; Extracts

Scallops *See* **Shellfish**: Characteristics of Crustacea; Commercially Important Crustacea; Characteristics of Molluscs; Commercially Important Molluscs; Contamination and Spoilage of Molluscs and Crustaceans; Aquaculture of Commercially Important Molluscs and Crustaceans

Scanning Electron Microscopy *See* **Microscopy**: Light Microscopy and Histochemical Methods; Scanning Electron Microscopy; Transmission Electron Microscopy; Image Analysis

SCURVY

T K Basu, University of Alberta, Edmonton, Alberta, Canada
D Donaldson, Chemical Pathology Department, East Surrey Hospital, Redhill, Surrey, UK

Etymology

'Scurvy' is the current word in the English language for what has at various times in history been referred to as skurvie, scurvie, skirvye, scurvey, scurby, skyrby, scorbie, and scorby. The same word, in varied forms, has also been used to refer to scurf, i.e., dandruff. Confusion has also arisen because one sign, the petechiae (small hemorrhages in the skin), that are a feature of scurvy and many hematological disorders, has been referred to loosely and quite incorrectly as scurvy, because they are confused with purpura. In addition, many of the early references to scurvy are almost certainly erroneous as only later was it possible to differentiate this clinical disorder from other diseases of close or superficial resemblance. Further confusion and lack of clarity were caused by reference to patients with mixed disorders such as rickets as a comorbidity as scurvy, when the disease might in reality have been just a portion of the whole clinical presentation. Scurvy is a clinical syndrome caused by vitamin C deficiency which is associated with a derangement of collagenous protein synthesis; this is rapidly reversed with vitamin C supplementation.

History

The Ebers Papyrus from Ancient Egypt (*c.* 1500 BC) contains a reference to what could well be scurvy. In Ancient Greece Hippocrates referred to a disease occurring in soldiers and comprising pain in the legs, gangrene of the gums, and the consequent loss of teeth. During the Middle Ages 'land scurvy' was

recognized to be both commonplace and seasonal throughout the whole of Europe, usually being seen in the winter months, long after the cessation of availability of summer fruits and vegetables.

'Land scurvy' became 'sea scurvy' during the 14th and 15th centuries at the time of a rising spirit of adventure, catalyzed by new technical developments in ship design and advancements in navigational instrumentation, and compounded by growing and powerful commercial interests in the trading of silk and spices. These factors cleared the way for longer sea voyages of weeks or months in duration, without the opportunity for the crew to land at ports where fresh vegetables and fruits would be available. The diaries of Commodore George Anson's voyage of 1740–44, during which he attempted unsuccessfully to circumnavigate the world and on which 1051 of his 1955 men died – the majority from scurvy – referred to the fact that 'the scars of old wounds, healed for many years, were forced open again' and stated 'many of our people, though confined to their hammocks, ate and drank heartily and were cheerful, yet having resolved to get out of their hammock, died before they could well reach the deck.'

Vasco da Gama sailed from Lisbon on 9 July 1497 with about 140 Portuguese sailors. He reached the south-eastern coast of Africa 7 months later; the records indicated 'many of our men fell ill here, their feet and hands swelling and their gums grown over their teeth so that they could not eat.' On 6 April 1498 came the opportunity to purchase oranges from Moorish traders. Just 6 days later it is stated 'all our sick recovered their health for the air of this place is very good.'

In the winter of 1536, on the frozen St Lawrence River, the North American Indians taught the French explorer Jacques Cartier and his men the value of boiling the bark and leaves of the white cedar tree in water and then drinking the juice and dregs. Three doses on alternate days were sufficient 'miraculously' to cure their loss of strength and their swollen and inflamed legs.

The introduction of the potato to Europe from South America, where it had been cultivated for nearly 1500 years, was effected in the second half of the 16th century by the Spaniards who invaded that country. It is now seen to have provided a ready means of combating the problem of low vitamin C status in the winter months.

In 1617 John Woodall, Surgeon-General of the British East India Company, wrote *The Surgions Mate*, which contained a 23-page chapter on the subject of scurvy. He emphasized at that time the necessity of providing the crews of ships with the juice of oranges, lemons, or limes. He added a recommendation of two to three spoonfuls of lemon juice as a medicine against scurvy, and as a preventive too, if enough could be spared. This suggestion preceded the work and publications of James Lind, which did not appear until the following century.

James Lind was born in Edinburgh in 1716 and eventually became Physician-in-charge of the 2000-bed Haslar Hospital near Portsmouth, the largest and newest of the naval hospitals, in which there was the opportunity to study 300–400 cases of scurvy at a time. The famous experiment of his earlier days, in 1747, involved six groups of two men each, treated as follows: group 1 with cider, group 2 with elixir vitriol, group 3 with vinegar, group 4 with sea water, group 5 with two oranges and one lemon each day over 6 days, and group 6 with garlic, mustard seed, balsam of Peru, dried radish root and gum myrrh, together with barley water, tamarinds and cremor tartar. All received the same diet apart from the above 'medicines.' The best response was from group 5 by the sixth day, and group 1 came second at 2 weeks; the other 'remedies' proved to be of no value. Lind published his famous book *A Treatise of the Scurvy* in 1753, in Edinburgh. Nevertheless, it was not until 1804 that the Royal Navy decreed that lemon or lime juice must be provided daily; limes were later substituted for lemons as they were cheaper and could be obtained from the new West Indian colonies, although they are now known to contain less vitamin C. Hence the origin of the term 'limeys' for British sailors. In retrospect, one can now say that James Lind made one major error in preparing his 'rob of oranges' by evaporating juice down to 10% of its original volume, as heat is now known to destroy vitamin C; moreover, further storage following the preparative procedures permits yet greater deterioration in the vitamin C content.

The history of vitamin C is strewn with anecdotes such as these. Others should be briefly mentioned; they include Captain James Cook's 1768 voyage to the South Pacific in which his crew avoided scurvy by his insistence on the consumption of various fresh vegetables. An outbreak of scurvy at the National Penitentiary at Millbank in London in February 1825, following drastic reduction of the diet of the inmates, responded to the simple prescription of three oranges a day; however, neither the kitchen staff nor the prison officers suffered!

The Great Potato Famine of 1845–48 involved not only the UK and Ireland but also France and Belgium; many cases of scurvy were being reported in the British medical literature from 1847 onwards. Scurvy, not surprisingly, was seen in the general population and in the prisons and hospitals, too. The Royal Navy's Arctic expedition, which returned in 1876, suffered

badly from scurvy in spite of the availability of lime juice. However, lime juice had not been supplied on sledging expeditions, and even when it had been taken, the bottles had frozen solid, causing them to break on thawing over the evening fire.

In 1928 Albert Szent-Györgyi isolated a sugar from cabbage and lemon juice, and also from the adrenal gland. He was not permitted by the editor of the *Biochemical Journal* to call it 'ignose' or 'godnose' and had to call it 'hexuronic acid' instead. In 1932 CG King and WA Waugh identified the antiscorbutic factor from lemon juice to be the same as hexuronic acid; the name ascorbic acid was adopted. In 1933 Sir Norman Hawarth and his colleagues in Birmingham, UK, established the chemical structure of ascorbic acid. Also in 1933 both Sir Norman Hawarth's team and the team led by Tadeus Reichstein in Switzerland succeeded in synthesizing vitamin C.

During the latter part of World War II, at the time when many of those captured in the Far East were in Japanese prisoner-of-war camps, it was accepted that deficiency of vitamins A and B was commonly seen; however, deficiency of vitamin C did not occur. The reason for this was that the prisoners themselves prepared and drank 'jungle juice,' a mixture of the green leaf tops of sweet potatoes, kang kong (a vegetable eaten by the local natives in Borneo), leaves of the tapioca tree, and onion leaves for flavoring, all boiled up quickly in water. Grass, which was in abundance, and which is now known to contain weight-for-weight more vitamin C than citrus fruits, was never used as it was thought to have no nutritional value.

Species-sensitivity

Most animals do not develop scurvy since they are able to synthesize vitamin C from glucose by the reactions in eqns (1–3).

$$\text{D-Glucose} \xrightarrow{\text{NADPH.H}^+} \text{D-Glucuronic acid} \longrightarrow \text{D-Glucuronolactone} \quad (1)$$

$$\text{D-Glucuronolactone} \longrightarrow \text{L-Gulonolactone} + O_2 \xrightarrow[\text{oxidase}]{\text{gulonolactone}} \text{2-Ketogulonolactone} + H_2O_2 \quad (2)$$

$$\text{2-Ketogulonolactone} \longrightarrow \text{L-ascorbic acid} \quad (3)$$

The only known species unable to synthesize vitamin C are the primates (including humans), guinea pigs, the red-vented bulbul, the fruit-eating bat (*Pteropus medius*), the rainbow trout and the coho salmon. They lack the enzyme gulonolactone oxidase, which is necessary for the conversion of L-gulonolactone to L-ascorbic acid. Hence, these species are dependent on dietary sources (**Table 1**) for the vitamin. (*See* **Ascorbic Acid**: Physiology.)

Table 1 Vitamin C content of foods and plants

Food or plant	*Vitamin C content (mg per 100 g or 100 ml)*
Fruits	
Apples	4
Banana with peel	7
Citrus fruits	25–60
Gooseberries, fresh	60–65
Strawberries	58
Vegetables	
Broccoli	90
Cabbage	
Fresh	45–60
Boiled	24–30
Peas	
Sprouting	25–50
Dried	0
Potatoes	
Uncooked	10–30
Baked in skin	15
Boiled	5–15
Boiled and reheated	2–8
Sauerkraut	
Stored 1 month	10–15
Tomatoes	20
Animal origin	
Adrenal glands	100–200
Eggs	Trace
Meat or fish, well cooked	Trace
Milk	
Cows', pasteurized	1
Human breast milk	30–55
Botanical	
Scurvy grass	66–100
Spruce pine needles	65–200
Gramineae (grasses)	140–173
Liquids	
Apple cider, fresh unpasteurized	4–5
Blackcurrant syrup	60
Lemon juice, fresh	45
Lime juice, fresh	30
Orange juice, fresh	48
Rosehip syrup	150–200

Pathophysiology

Vitamin C deficiency causes failure to maintain the cellular structure of the supporting tissues of mesenchymal origin, such as bone, dentine, cartilage, and connective tissues. As a consequence, vitamin C deficiency is characterized by weakness and fatigue, hyperkeratosis of the hair follicles, perifollicular hemorrhages, petechiae and ecchymoses, swollen bleeding gums, delayed wound healing, and the easy fracturing of bone. These are the characteristic clinical features of scurvy.

The clinical manifestations of scurvy are the result of complete vitamin C deprivation of 100–160 days'

Table 2 The consequences of withdrawing vitamin C from the diet

	Plasma ascorbic acid		*Buffy coat ascorbic acid – in leukocytes (WC)*		*Body pool of ascorbic acid*		
Day	*mg l^{-1}*	*μmol l^{-1}*	*μg 10^{-8} WC*	*nmol 10^{-8} WC*	*g*	*nmol*	*Clinical state*
0	8–15	45–85	21–57	119–323	0.6–1.5	3.4–8.5	Adults
20	3	17	10–38	57–216			
40	1–3	6–17	2–10	11–57	0.3–0.6	1.7–3.4	Subclinical deficiency
60	< 1	< 6	< 5	< 28	0.3–0.6	1.7–3.4	
80	< 1	< 6	< 5	< 28	0.3–0.6	1.7–3.4	
100	< 1	< 6	< 5	< 28	0.3–0.6	1.7–3.4	
120	< 1	< 6	< 2	< 11	< 0.3	< 1.7	Perifollicular hyperkeratosis
140	< 1	< 6	< 2	< 11	< 0.3	< 1.7	
160	< 1	< 6	< 2	< 11	< 0.3	< 1.7	Petechiae and ecchymoses of the skin and failure of wounds to heal
180	< 1	< 6	< 2	< 11	< 0.1–0.3	0.6–1.7	Gingival changes
200	< 1	< 6	< 2	< 11	< 0.1	< 0.6	Dyspnea, edema, and very rapid progression

duration (**Table 2**). During deficiency the vitamin C pool of the body is depleted at a daily rate of about 2.6% of the existing pool; 92% is lost after 100 days and symptoms of mild scurvy are evident when the pool is less than about 300 mg.

Fibroblasts play a central role in the wound-healing process. In vitamin C deficiency the fibroblasts proliferate but in an unimpeded manner. These proliferating cells generally remain immature, however, and fail to synthesize collagen molecules, which are the 'building blocks' of tissue repair. The impaired fibroblastic activity is essentially responsible for the delayed wound healing in vitamin C deficiency. Moreover, in vitamin C deficiency cartilage cells of the epiphyseal plate, at the end(s) of the diaphysis of long bones, continue to proliferate and line up in rows. The cartilage between these rows is calcified where osteoblasts do not migrate, resulting in the development of compressed and brittle bone.

Ultrastructural studies in scurvy show distinct alterations of the ribosomal and polyribosomal pattern. The endoplasmic reticulum shows considerable dilation and loss of ribosomal granules. The polyribosomes are disaggregated, unbeaded fibrillar material being seen in the extracellular space.

Biochemical Aspects

The characteristic feature of scurvy is an inability of the supporting tissues to maintain the synthesis of collagen, the intercellular substance. The collagen molecule consists of three linked popypeptide chains of amino acid residues. For the collagen molecule to aggregate into its triple helix configuration, the proline and lysine residues of newly synthesized collagen must be hydroxylated. Formation of the triple helix is important because it is in this configuration that the procollagen is secreted from the fibroblast. Hydroxylation of both proline and lysine occurs after the amino acids have been incorporated into the peptide chain; vitamin C is required in these reactions. The essential process mediated by vitamin C is the formation of hydroxyproline from proline and of hydroxylysine from lysine, as shown in the following reactions:

$$\text{Proline} + \text{ascorbate} + O_2 \xrightarrow{\text{proline hydroxylase}} \text{OH-proline} + \text{Dehydroascorbate} + H_2O \quad (4)$$

$$\text{Lysine} + \text{ascorbate} + O_2 \xrightarrow{\text{lysine hydroxylase}} \text{OH-lysine} + \text{Dehydroascorbate} + H_2O \quad (5)$$

Vitamin C appears to act as a cofactor for peptidyl proline and lysine hydroxylases, possibly by keeping copper and iron in a reduced state. Defective hydroxylation within the synthesizing fibroblast gives rise to the formation of an abnormal collagen precursor called protocollagen. Unlike the normal precursor (tropocollagen), this substance may not be extruded from the cell. If it is extruded it polymerizes into an abnormal collagen which lacks tensile strength. In addition to its requirement for collagen synthesis, vitamin C is believed to be an essential cofactor in forming the intercellular cement substance which binds the lining cells of blood vessels not only to their basement membrane but also to each other. (*See* **Coenzymes.**)

Occurrence of Scurvy

Scurvy is now a rare condition but isolated cases can still be seen in certain population groups. The at-risk groups are often the elderly (not so much in North America), food faddists, alcoholics, those living in institutions, and patients with psychiatric disorders.

Outbreaks of scurvy occur in poor nomadic populations in arid or semidesert districts when there is a threat of famine or a long-standing drought. There is also isolated evidence that long-term administration of large doses of vitamin C could lead to adaptation which might be responsible for developing scurvy, following sudden cessation of the extra vitamin intake. The conditioning effect is more pronounced during intrauterine life than either following birth or in adults. Overdosing with vitamin C during pregnancy may thus be contraindicated.

See also: **Ascorbic Acid**: Properties and Determination; Physiology; **Coenzymes**

Further Reading

Basu TK (1985) The conditioning effect of large doses of ascorbic acid in guinea pigs. *Canadian Journal of Physiology and Pharmacology* 63: 427–430.

Basu TK and Schorah CJ (1982) *Vitamin C in Health and Disease*. London: Routledge.

Carpenter KJ (1986) *The History of Scurvy and Vitamin C*. Cambridge: Cambridge University Press.

Irwin MA (1976) A conspectus of research on vitamin C requirements of man. *Journal of Nutrition* 106: 823–879.

Lind J (1976) *A Treatise on the Scurvy*. Republished (1953) by Edinburgh University Press, Edinburgh.

Vilter RW (1978) Nutrient deficiencies in man: vitamin C. *CRC Handbook Series (Nutrition and Food)* 3: 91–103.

Seafood *See* **Fish**: Introduction; Catching and Handling; Fish as Food; Demersal Species of Temperate Climates; Pelagic Species of Temperate Climates; Tuna and Tuna-like Fish of Tropical Climates; Pelagic Species of Tropical Climates; Demersal Species of Tropical Climates; Important Elasmobranch Species; Processing; Miscellaneous Fish Products; Spoilage of Seafood; Dietary Importance of Fish and Shellfish; **Fish Farming**; **Fish Meal**; **Shellfish**: Characteristics of Crustacea; Commercially Important Crustacea; Characteristics of Molluscs; Commercially Important Molluscs; Contamination and Spoilage of Molluscs and Crustaceans; Aquaculture of Commercially Important Molluscs and Crustaceans; **Marine Foods**: Production and Uses of Marine Algae; Edible Animals Found in the Sea; Marine Mammals as Meat Sources

Seaweed *See* **Marine Foods**: Production and Uses of Marine Algae; Edible Animals Found in the Sea; Marine Mammals as Meat Sources

SELENIUM

Contents

Properties and Determination

G F Combs Jr, Grand Forks Human Nutrition Research Center, Grand Forks, ND, USA

Introduction

The chemical properties of the trace element selenium are similar to those of sulfur; however, unlike sulfur, which tends to be oxidized in biological systems, selenium tends to undergo reduction in the tissues of microbes, plants, and animals. Selenium is present in most biological tissues at very low concentrations where it is almost exclusively bound to proteins, mostly as analogs of the sulfur-amino acids. Several methods are available for the analysis of selenium; the most commonly used ones are electrothermal atomic absorption spectrophometry, and a chemical procedure based on the formation of a fluorescent,

piazoselenol derivative. These methods are useful for the quantitation of total selenium in tissues; however, more sophisticated methods, such as those involving high-performance liquid–liquid partition chromatography linked with mass spectrometry, are required for the chemical speciation of tissue selenium.

Chemical Forms of Selenium

Selenium is in group IVa of the periodic table of elements. This group includes the nonmetallic elements sulfur and oxygen in the periods above selenium and the metallic elements tellurium and polonium in the periods below. By period, selenium lies between the group Va metal arsenic and the group VIIa nonmetal bromine. Thus, selenium is considered a metalloid, having both metallic and nonmetallic properties. Its atomic properties are listed in Table 1.

Elemental selenium shows allotropy; that is, it can exist in either an amorphous state or in one of three crystalline states. Amorphous selenium in a hard, brittle glass at temperatures below 31 °C, is vitreous at 31–230 °C, and is a free-flowing liquid above 230 °C. The increased viscosity of amorphous selenium at temperatures less than 230 °C results from the formation of polymeric chains. Crystalline selenium can take several forms: flat, polygonal (often hexagonal) crystals called γ-monoclinic or red selenium; needle-like, prismatic crystals called γ-monoclinic or dark red selenium; or spiral polyatomic chains variously called metallic, gray or black selenium. Of these forms, the most stable is the hexagonal, crystalline form to which both monoclinic forms convert at temperatures above 110 °C and to which amorphous selenium converts spontaneously at 70–120 °C.

Elemental selenium can be reduced to the −2 (selenide, Se^{-2}) oxidation state or oxidized to the +4 (selenite, Se^{+4}) or +6 (selenate, Se^{+6}) oxidation states. Hydrogen selenide (H_2Se) is a fairly strong acid with a pK_a of 3.8 in aqueous systems. It is a colorless gas with an unpleasant odor. The gas is highly toxic (the LC50 for 30 min exposure for guinea pigs is 6 p.p.m.); however, Se^{-2} is also a pivotal metabolite in animals and at least some microbes, being the obligate precursor to selenocysteine at the active centres of a number of physiologically important selenoenzymes. Hydrogen selenide can be produced by heating elemental selenium (Se^0) to temperatures above 400 °C, but it decomposes in air to form the element and water. It is fairly soluble in water (270 ml per 100 ml of water at 22.5 °C). It reacts directly with most metals to form insoluble metal selenides. Organic selenides are ready electron donors; thus, they are readily converted to higher oxidation states.

In the +4 oxidation state, selenium can exist as selenium dioxide (SeO_2), selenious acid (H_2SeO_3), or selenite (SeO_3^{2-}). Selenium dioxide is formed by burning elemental selenium in air, or by reacting it with nitric acid. It is readily reduced to the elemental state by ammonia, hydroxylamine, sulfur dioxide, or a number of organic compounds. It is soluble in water (38.4 g per 100 ml at 14 °C) and forms selenious acid when dissolved in hot water. Selenious acid is weakly dibasic (pK_a 2.6), and dissolved selenite salts exist as biselenite ions in aqueous solutions at pH 3.5–9. In contrast to the organic selenides, selenites readily accept electrons from their environments and are, therefore, easily reduced. At low pH, selenite is readily reduced to the elemental state by such

Table 1 Atomic properties of selenium

Atomic number	34
Atomic weight	78.96
Stable isotopes (natural abundance)	^{74}Se (0.87%), ^{76}Se (9.03%), ^{77}Se (7.58%), ^{78}Se (23.52%), ^{80}Se (49.82%), ^{82}Se (9.19%)
Radioisotopes (emission; [a]half-life)	^{70}Se (γ^+, EC; 44 months); ^{71}Se (γ^+, EC; 5 months), ^{72}Se (EC; 8.4 days), ^{73}Se (γ^+, EC; 7.1 h), ^{75}Se (γ-; 120.4 days), ^{77m}Se (IT; 17.5 s), ^{79}Se (γ-; 6.5 × 10^4 years), ^{81m}Se (IT; 57 months), ^{81}Se (γ-; 18.6 months), ^{83m}Se (γ-; 70 s), ^{83}Se (γ-; 23 months), ^{84}Se (γ-; 3.3 months), ^{85}Se (γ-; 39 s)
Electronic configuration	Ar3d^{10} 4$s^2$2p^4
Atomic radius	1.40 γ
Covalent radius	1.16 γ
Ionic radius	1.98 γ
Common oxidation states	−2, 0, $^+$4, $^+$6
M-M bond energy	44 kcal mol^{-1}
M-H bond energy	67 kcal mol^{-1}
Ionization potential	9.75 eV
Electron affinity	−4.21 eV
Electronegativity	2.55 (Pauling's scale)

[a]emission codes: γ^+, positron emission; $\gamma-$, negative beta emission; IT, isomeric transition; EC, orbital electron capture.

reducing agents as ascorbic acid and sulfur dioxide. Selenites in soils are strongly bound by hydrous oxides of iron, forming insoluble complexes at moderate pH.

In its highest oxidation state (+6), selenium can exist as selenic acid (H_2SeO_4) or selenate (SeO_4^{-2}) salts. Selenic acid is a strong acid. It is formed by the oxidation of selenious acid or elemental selenium with strong oxidizing agents in the presence of water. It is very soluble in water, as are most selenate salts. Selenates tend to be rather inert and are very resistant to reduction.

Six stable isotopes of selenium exist in nature and 13 radioisotopes of the element can be produced by neutron activation or radionuclear decay (Table 1). The stable isotopes ^{74}Se, ^{76}Se, and ^{80}Se have been employed in the study of the metabolism of selenite and selenomethionine by humans; these can be detected by mass spectrometry or neutron activation analysis. The short-lived radioisotope ^{77m}Se (half-life 17.5 s) has been used in the neutron activation analysis of the element in biological materials. Due to its emission of γ-radiation and to its relatively long half-life (120 days), ^{75}Se has proven widely employed in biological experimentation and in medical diagnostic work.

Selenium is a semiconductor that exhibits photoconductivity; that is, excitation with electromagnetic radiation can increase its conductivity. This property has made selenium compounds useful in photovoltaic cells and in xerography. As a result of the many practical utilities conferred by this property, the literature on the chemistry of inorganic and organic selenium compounds is large.

The selenium compounds of greatest relevance in biology are listed in Table 2. It should be noted that, whereas some dietary supplements and fortificants of foods and livestock feeds include selenium compounds of higher oxidation states (e.g., selenate [Se^{+4}]), the major metabolites are in the reduced (Se^{-2}) state.

Chemical Properties of Selenium

The chemical properties of selenium, like its physical properties, are very similar to those of sulfur. The two elements have similar outer valence shell electronic configurations and similar atomic sizes in both the covalent and ionic states. Their bond energies, ionization potentials, and polarizabilities are virtually the same. Despite these similarities, the chemistry of selenium differs from that of sulfur in two important respects which distinguish the elements in biological systems: their oxyanions are not similarly reduced, and their hydrides have different acid strengths. For these reasons, selenium behaves very differently from sulfur in biological systems.

In biological systems, selenium compounds tend to be metabolized to more reduced states while sulfur compounds tend to be metabolized to more oxidized states. For example, quadrivalent selenium (Se^{+4}) in selenite tends to undergo reduction to Se^{-2}, being metabolized first to H_2Se and, ultimately, being methylated to a variety of excretory forms. In contrast, quadrivalent sulfur in sulfite tends to undergo oxidation. This difference is demonstrated by the ability of selenious acid to oxidize sulfurous acid: $H_2SeO_3 + 2H_2SO_3 \rightarrow Se^0 + 2H_2SO_4 + H_2O$.

Although the oxyacids of selenium and sulfur have comparable acid strengths (pK_a 2.6 versus pK_a 1.9, respectively, for the quadrivalent species; pK_a 3 for both of the hexivalent species), the hydride H_2Se is much more acidic than H_2S (pK_a 3.8 versus 7.0, respectively). This difference in acid strengths is

Table 2 Biologically important selenium compounds

Oxidation state	*Compound*	*Biological relevance*
Se^{-2}	H_2Se	Obligate metabolic precursor to selenoproteins
	Methyl selenol, CH_3SeH	Excretory form (lung) shown to be anticarcinogenic
	Dimethyl selenide, $(CH_3)_2Se$	Excretory form (lung)
	Trimethyl selenonium, $(CH_3)_3Se^+$	Excretory form (kidney)
	Selenomethionine	Common food form; form in nonspecific Se-containing proteins (e.g., albumin); metabolized to selenocysteine
	Selenocysteine	Common food form; form in selenoproteins; metabolized to H_2Se
	Se-methylselenomethionine	Form found in some foods; metabolized to CH_3SeH
	Se-methylselenocysteine	Form found in some foods; metabolized to CH_3SeH
	Selenobetaine	Metabolic precursor of CH_3SeH
	Selenotaurine	Form found in some foods
Se^0	Selenodiglutathione	Reductive metabolite of selenite/selenate with anticarcinogenic activity
Se^{+4}	Na_2SeO_3	Commonly used feed supplemental form
Se^{+6}	Na_2SeO_4	Potential food/feed supplement form

reflected in the different dissociation behaviors of the selenohydryl group (–SeH) of selenocysteine (pK_a 5.24) and the sulfhydryl group (–SH) of cysteine (pK_a 8.25). This, while thiols such as cysteine are mainly protonated at physiological pH, selenols such as selenocysteine are predominantly dissociated under the same conditions.

Selenite (Se^{+4}) can readily react with nonprotein thiols and with protein-sulfhydryl groups to undergo reduction to Se^0 with the concomitant formation of 1,3-dithio-2-selane products of the form RS-Se-SR, called 'selenotrisulfides.' In this reaction, $4RSH + H_2SeO_3 \rightarrow RS\text{-}Se\text{-}SR + 3H_2O$, four sulfhydryl sulfur atoms (S^{-2}) are oxidized to the disulfide state (S^{-1}) with the concomitant reduction of a single selenium atom from the selenite state (Se^{+4}) to the zero oxidation state. However, because the electronegativities of selenium and sulfur are very similar, the –2 charge may be distributed across the selenotrisulfide bridge, yielding an effective oxidation number of –2/3 for each of its members. A similar reaction can occur between selenite and free sulfhydryl groups of proteins to yield selenodisulfide (RS-SeH) and selenotrisulfide-type oxidation products that, in some cases, can inhibit enzymatic activity. Indeed, such reactions are thought to play a role in selenium toxicity.

Analysis of Selenium

The analytical challenge for determining selenium in biological samples, i.e., plant and animal tissues, is to detect quantitatively the element over at least three orders of magnitude of concentration. The nutritional assessment of selenium status calls for analytical procedures capable of measuring plasma/serum selenium levels in individuals who may be adequately nourished (> 80 p.p.b. selenium), marginally nourished (40–80 p.p.b.), frankly deficient (< 40 p.p.b.), or potentially intoxicated (> 1000 p.p.b.). The analysis of selenium in foods and feedstuffs calls for methods capable of measuring the element in the range of 20–500 p.p.b., while the detection of selenium in nonprotein-bound fractions of animal/human tissues calls for detection capability of 1–5 p.p.b.

Several procedures are available for the quantitative determination of selenium; however, some that have industrial utility do not lend themselves to biological applications due to their relatively low sensitivities. Such methods include both gravimetric (e.g., after reduction and quantitative precipitation with acid or electrolytic deposition with copper) and colorimetric ones (e.g., titration with oxidizing agents after reduction with thiocyanate; measurement of selenium hydrosols after reduction by hydrazine, stannous chloride, or ascorbic acid; measurement of azo-compounds formed by the reaction of aromatic amines with diazonium salts produced by the oxidation of organic compounds by selenite; measurement of complexes of Se^{-2} with phenyl-substituted thicarbazides or semicarbazide after reduction of selenium). The limit of detection of none of these methods is less than 0.5 p.p.m. and, for the gravimetric methods, they can be several parts per million. None of these procedures is free of interferences that can give false-positive results; other elements can coprecipitate with selenium to affect the results of gravimetric methods, and other oxidizing agents that can affect the colorimetric methods.

Other methods have been found to be more useful for the determination of selenium in plant and animal tissues. These include: (1) fluorimetric measurement of selenium, after nitric-perchloric acid digestion and conversion to Se^{+4}, using diaminonaphthalene (DAN) to form benzopiazselenol; (2) atomic absorption spectrophotometry; (3) instrumental neutron activation analysis; (4) mass spectrometry; (5) atomic fluorescence spectroscopy; (6) inductively coupled plasma emission spectroscopy; and (7) proton-induced X-ray emission spectrometry. Each of these approaches can yield sensitivity less than 5 p.p.b. Of these, the methods most widely employed in biomedical applications have been the DAN-fluorimetric and electrothermal atomic absorption methods.

The chemical determination of selenium by the DAN-fluorimetric method is of lower cost with comparable sensitivity (2–5 p.p.b.) compared to most other methods. This method necessitates the quantitative conversion of the selenium in the sample to Se^{+4}. Because most, if not all, selenium in biological tissues occurs as Se^{-2} compounds, this is achieved using nitric-perchloric acid digestion to oxidize organic matter and form Se^{+6}. This is followed by treating the sample with hydrochloric acid to reduce any Se^{+6} that may be present to the +4 oxidation state. Se^{+4} is reacted with DAN, yielding benzopiazselenol which is extracted into cyclohexane and measured fluorimetrically (excitation 390 nm; emission 520 nm). The DAN-fluorimetric method has two points of vulnerability: effect that impairs the conversion of selenium to Se^{+4}, and factors that produce fluorescence interference. One potential source of the first type of error is the loss of volatile H_2Se formed in the transition, particularly in the presence of large amounts of organic matter such that sample charring can occur (e.g., fatty materials such as egg yolk or adipose tissue). Because H_2Se is volatilized from acid solutions by reducing agents, this loss can be avoided by maintaining strongly oxidizing conditions during digestion and by using low heat such that

the oxidation of Se^{+4} to Se^{+6} proceeds relatively slowly (this can be achieved by gradually raising the temperature of the perchloric acid solution to 210 °C). When the nitric-perchloric acid digestion is controlled and carefully attended, it produces satisfactory conversion to Se^{+4}, even of such forms as trimethylselenonium which is resistant to oxidation by nitric acid alone. The second type of problem involves fluorescence interference caused by degradation products of DAN. This can be avoided by purifying DAN prior to use either by recrystallization from water in the presence of sodium sulfite and activated charcoal, or by hexane extraction after hydrochloride stabilization.

Although conventional atomic absorption spectrophotometry (AAS) does not yield satisfactory sensitivity for most biological applications, two variant methods do so. The best results are obtained using electrothermal AAS which avoids the problems associated with wet digestions by employing high-temperature oxidation and antomization in a graphite furnace. This reduces interferences due to nonspecific absorption of organic compounds and nonselenium salts, but introduces the problem of selenium volatility under such conditions and can also be prone to 'matrix' effects, particularly due to phosphate. These sources of interference can be effectively reduced by using a palladium-based matrix modifier to reduce the volatility and, thus, permit the vaporization of nonselenium compounds at lower furnace temperatures than those used to vaporize and quantitate selenium. Also effective is background correction by the Zeeman effect, in which a magnetic field is used to split the resonance line into its Zeeman components, converting a single beam into a double one, the wings of which can be monitored with polarizers to assess the background while the central portion is monitored to assess the analyte. In practice, electrothermal AAS with the use of automatic Zeeman-effect background correction can achieve a sensitivity of 2 p.p.b. A less sensitive AAS technique involves the pretreatment of the sample with sodium borohydride and acid to convert selenium to H_2Se. This method requires selenium to be in solution as Se^{+4}, which can be achieved by nitric-perchloric acid digestion of the sample, after which the hydride generation step can be automated. This approach yields a limit of detection of only 50–100 p.p.b. and is subject to interferences due to other elements that can also form hydrides, particularly copper.

Investigators with access to research atomic reactors have found instrumental neutron activation analysis of selenium to be very useful, as it offers the advantages of applicability to small samples, ease of sample preparation, and nondestruction of sample. The greatest sensitivity, about 20 p.p.b., by this method is obtained by measuring ^{75}Se; however, the use of that radioisotope requires relatively lengthy periods of irradiation (100 h), postirradiation holding (60 days) and counting (2 h), making it relatively expensive. Greater economy and increased sample throughput are achieved through the use of the short-lived ^{77m}Se which can be produced by only 5 s of irradiation, and counted (25 s) after only 15 s of decay using an automated system. Nevertheless, this fast method is limited by its relatively low sensitivity, rendering it unsuitable for accurate quantitation of such low amounts of selenium as are found in tissues of animals chronically deficient in the element.

Selenium can be measured by mass spectrometry, which method offers the distinct advantage of being able to distinguish between the major stable isotopes of the element. There is, therefore, great interest in developing mass spectrometric techniques for use in metabolic studies in which ^{74}Se, ^{76}Se, or ^{82}Se are used as tracers. Mass spectrometric procedures have been developed for determining selenium in biological samples by isotope dilution. These involve the analysis of the naturally abundant ^{80}Se in addition to ^{82}Se which is used as an internal standard. Samples spiked with known amounts of the internal standard are treated by nitric-perchloric acid digestion to convert all of the selenium in the sample to Se^{+4}. The latter can be chelated using *o*-nitrophenylenediamine, extracted into an organic solvent, and subjected to gas–liquid partition chromatography–mass spectrometry to determine the ratios of ^{82}Se to ^{80}Se, comparing that ratio to the ratio of the one naturally occurring in the sample. Mass spectrometry can yield limits of detection for selenium as low as 1 p.p.b. The technique known as inductively coupled plasma mass spectrometry (ICP-MS) involves coupling using mass spectrometry inline after ionizing the sample in a high-temperature plasma.

Selenium can be measured by emission spectrophotometry using either a flame or high-temperature plasma to atomize selenium. A very sensitive atomic flame fluorescence spectrophotometric method uses automated hydride generation of a predigested sample. This method has not been widely used, but offers very good specificity, in as much as only a few elements can form volatile hydrides and as the fluorescence emission wavelength can be measured selectively. The method offers impressive sensitivity with limits of detection on the order of 2–5 p.p.t.; the sample dilution usually required for hydride generation reduces the working limit of detection to the 20–50 p.p.t. range, which is still superior to other methods. Selenium can be measured by inductively

coupled plasma emission spectrometry, which offers the advantage of the simultaneous determination of other elements. This method can yield limits of detection of about 1 p.p.b., but is limited by spectral interferences of the dimeric argon formed in the plasma.

The measurement of selenium by proton-induced X-ray emission (PIXE) spectroscopy offers the potential advantage of simultaneous analysis of other elements. This method involves proton bombardment of target atoms to cause loss of the latters' inner-shell electrons which are consequently replaced by electrons from the outer shell. The X-rays emitted during that transition are characteristic of the energy differences between the respective electron shells and are, therefore, identifiable and quantifiable. The sensitivity of the PIXE procedure for the determination of selenium is about 10 p.p.b., making it useful for some biological purposes but not sufficiently sensitive for the accurate determination of very low tissue levels of the element.

Chemical Speciation of Selenium

Most selenium naturally present in biological materials appears to be in the reduced (selenide) state, Se^{-2}. There is, however, little information available concerning the various chemical forms in which that Se^{-2} occurs in plant and animal tissues. Studies have shown that the selenium in wheat and corn grown on seleniferous soils is predominantly protein-bound; in wheat half or more occurs as the selenium analog of methionine, selenomethionine. Because plants cannot synthesize the cysteine analog selenocysteine, it is generally thought that selenomethionine is the predominant form of the element in plant tissues in which it acts as a methionine mimic in general protein metabolism. This is not the case for animals and at least some microorganisms, which can biosynthesize selenocysteine from H_2Se. In fact, it is as selenocysteine in specific selenoenzymes that selenium serves its nutritionally essential metabolic functions. Therefore, the tissues of selenium-fed animals and humans typically contain both selenocysteine from the biosynthesis of selenoproteins as well as selenomethionine derived from the diet and used nonspecifically in general protein synthesis. In addition to these two main forms, it is clear that plant tissues can contain smaller amounts of certain derivatives, including Se-methylselenomethionine and Se-methylselenocysteine, and that animals produce metabolically both H_2Se as well as a number of methylated excretory metabolites thereof (methylselenol, dimethylselenide, and trimethylselenonium); however, there are virtually no data currently available concerning the amounts of these metabolites in animal or human tissues.

The chemical determination of selenium species is the next frontier in the analytical chemistry of selenium. Recent efforts in this area have involved the use of gas–liquid partition chromatographic separation techniques coupled with atomic plasma emission detection methods for the simultaneous detection and quantitation of selenoamino acids; of ethylation derivatization to separate volatile organoselenium compounds; and of ion-pair, reverse-phase, high-performance liquid–liquid partition chromatography couple to ICP-MS for the determination of selenoamino acids, selenoxides, and other selenium derivatives. Future efforts in this field are likely to employ molecular mass spectrometry.

See also: **Chromatography**: High-performance Liquid Chromatography; **Mass Spectrometry**: Principles and Instrumentation; Applications

Further Reading

Alfthan G and Kumpalainen J (1982) Determination of selenium in small volumes of blood plasma and serum by electrothermal atomic absorption spectrometry. *Analytical Chemica Acta* 140: 221.

Kotrebai M, Birringer M, Tyson JF, Block E and Uden PC (2000) Selenium speciation in enriched and natural samples by HPLC-ICP-MS and HPLC-ESI-MS with perfluorinated carboxylic acid ion-pairing agents. *Analyst* 125: 71.

Kumpalainen J (1990) Summary of new analytical techniques and general discussion. In: Tomita H (ed.) *Trace Elements in Clinical Medicine*, p. 451. New York: Springer.

Kumpalainen J, Raittila AM, Lehto J and Koivisoinen P (1983) Electrothermal atomic absorption spectrometric determination of selenium in foods and diets. *Journal of the Association of Official Analytical Chemists* 66: 1129.

Olson OE, Palmer IS and Cary EE (1975) Modification of the official fluorimetric method for selenium in plants. *Journal of the Association of Official Analytical Chemists* 58: 117.

Reilly C (1996) *Selenium in Food and Health*. New York: Blackie Academic & Professional.

Spate VL, Mason MM, Reams CL *et al.* (1994) Determination of total, and bound Se in sera by INNA. *Journal of Radiological and Nuclear Chemistry* 179: 315.

Uden PC, Bird SM, Kotrebai M *et al.* (1998) Analytical selenoamino acid studies by chromatography with interfaced atomic mass spectrometry and atomic emission spectral detection. *Fresenius' Journal of Analytical Chemistry* 362: 447.

Vanhoe H (1993) A review of the capabilities of ICP-MS for trace element analysis in body fluids and tissues. *Journal of Trace Elements and Electrolytes in Health and Disease* 7: 131.

Wolfe WR, Holden JM and Schubert A (1992) Selenium content of selected foods important for improved assessment of dietary intake. *Journal of Food Composition and Analysis* 5: 2.

Physiology

C D Thomson, University of Otago, Dunedin, New Zealand

Introduction

Selenium first attracted attention in the 1930s as a toxic trace element that caused 'alkali disease' in livestock consuming high-selenium plants. In 1957, selenium was shown to be essential for animals when traces of this element prevented liver necrosis in vitamin E-deficient rats, and later to prevent a variety of economically important diseases such as white muscle disease in cattle and sheep, hepatosis dietetica in swine, and exudative diathesis in poultry. The demonstration in 1973 of a biochemical function for selenium as a constituent of the enzyme glutathione peroxidase helped to explain the interrelationship between selenium and vitamin E. The importance of selenium in human nutrition was highlighted in reports in 1979 of the selenium responsive condition, Keshan disease in China, and selenium deficiency in a patient on total parenteral nutrition in New Zealand. Considerable research during the last two decades has provided information on the molecular biology, metabolism, and importance of selenium in human nutrition, leading to a much greater understanding of the functions of selenium and the establishment of recommended dietary intakes.

Metabolism of Selenium

Comparatively little is known about the forms of selenium in foods. The major form in cereals and other plants is selenomethionine, while the most likely form in animal foods is selenocysteine, as this is the active form in functional selenoproteins in the mammalian body. Inorganic forms are often used in experimental diets and as supplements, but the extent of occurrence of inorganic or other forms of selenium in foods is not clear. The metabolism of selenium, including absorption, transport, distribution, excretion, retention, and transformation to the active form, is very much dependent on the chemical form and amount of selenium ingested, and on interacting dietary factors. There is considerable species variation in many aspects of selenium metabolism.

Absorption

Selenium is absorbed mainly from the duodenum. Selenomethionine and methionine share the same active transport mechanism, but little is known about the transport of selenocysteine. Absorption of inorganic forms of selenium such as selenite and selenate is via a passive mechanism.

While the absorption of selenium is generally high in humans, probably about 80% from food selenium, selenomethionine appears to be better absorbed than selenite. Absorption of selenium is unaffected by selenium status, and there appears to be no homeostatic regulation of absorption. Selenium is better absorbed from a high-protein diet, and the absorption of selenomethionine may be influenced by the methionine content of the diet.

Bioavailability

As well as absorption, utilization of a nutrient may also include transformation to a biochemically active form, which, for selenium, is assessed from changes in tissue glutathione peroxidase. Few studies have been made in humans of the bioavailability of selenium, but animal studies show a wide variation in the availability from different foods. In rats, the bioavailability from mushrooms, tuna, wheat, beef kidney, and Brazil nuts is 5, 57, 83, 97, and 124%, respectively, in comparison with sodium selenite. In humans, the bioavailability of selenium from fish is low in comparison with wheat. Human studies also show differences among various forms of selenium such as selenate, wheat, and yeast, but this also depends upon the criterion of measurement used for availability, indicating the need to consider several variables, including short-term changes in glutathione peroxidase activity, long-term retention of tissue selenium, and metabolic conversion to biologically active forms.

Transport

Little is known about the transport of selenium in the body, although it appears to be transported bound to plasma proteins. A selenium-containing protein called selenoprotein P, isolated from rat and human plasma,

has been suggested as a transport protein, and plasma also contains extracellular glutathione peroxidase, but lower-molecular-weight forms of selenium are more likely to act as transport proteins.

Metabolism and Distribution

An outline of selenium metabolism is shown in Figure 1. In animal tissues, selenium occurs in association with protein, and is present in two main compartments or forms. The first is selenocysteine, which is present as the active form of selenium in selenoproteins, including the selenoenzyme glutathione peroxidase. The second is selenomethionine, which is incorporated in place of methionine in a variety of proteins, unregulated by the selenium status of the animal.

Selenium levels in tissue are influenced by dietary selenium intake; this is reflected in the wide variation in blood selenium levels of residents of countries with differing soil selenium levels (Figure 2). The total amount of selenium in the body ranges from 3 to 20 mg, depending on intake of selenium. Retention of selenium is also profoundly influenced by the form administered, with selenomethionine being much more effective in raising blood selenium levels than sodium selenite or selenate. Selenomethonine appears to follow the same metabolic pathway as methionine and is incorporated nonspecifically into protein in its place. This contributes to tissue selenium, where it may in fact accumulate, but has no physiological function and is not available for synthesis of functional forms of selenium until it is catabolized. Selenate is reduced to selenite, then selenide, and in this oxidation state (−2) is introduced into selenocysteine, the active form of selenoproteins. The selenium from inorganic selenium, or from catabolism of selenomethionine or selenocysteine, is incorporated into selenocysteine, which enters regulated selenium metabolism and is then incorporated into selenoproteins. Selenium intake from selenate, selenite, or selenocysteine, unlike that from selenomethionine, will be excreted in urine if in excess.

Incorporation of Selenium into Selenoproteins

All functional mammalian selenoproteins contain selenocysteine at the active site. Selenium is incorporated by replacement of oxygen into serine to form selenocysteine, while serine is attached to a unique tRNA (transfer ribonucleic acid). Selenocysteine is encoded by a unique stop codon UGA on the mRNA specific for the selenoprotein. Regulation of selenoprotein synthesis appears to be via individual mRNAs (messenger RNAs) at the transcriptional or posttranscriptional levels in response to selenium availability, as well factors such as the chemical form of selenium and oxygen exposure. However, there is differential depression of selenoprotein synthesis in response to an inadequate selenium supply with preservation of the metabolically most important proteins. This has been termed the 'hierarchy' of importance of functional selenoproteins.

Excretion

Urine is the principal route of excretion of selenium, followed by feces, which mainly contain unabsorbed selenium. Homeostasis of selenium is achieved by regulation of its excretion. Daily urinary excretion of selenium is closely associated with plasma selenium and dietary selenium intake and has been used as an indicator of selenium status. Balance studies show

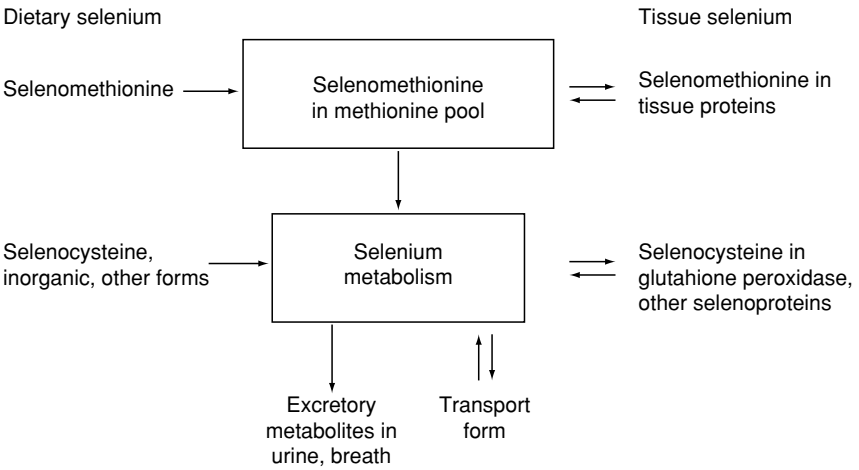

Figure 1 Outline of selenium metabolism.

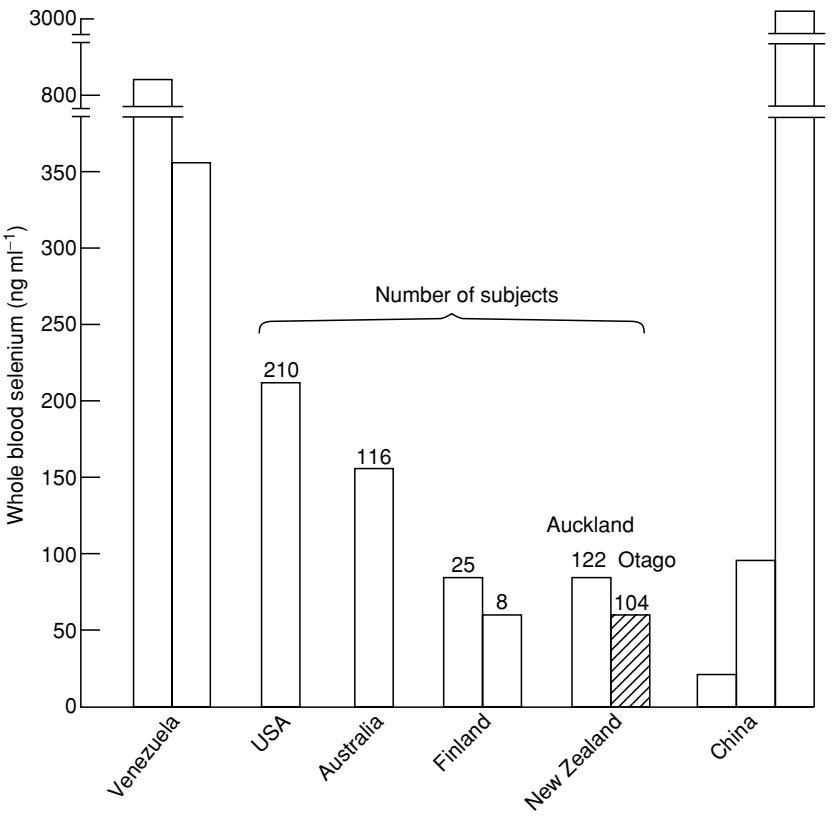

Figure 2 Blood selenium concentrations of healthy adults in Venezuela, the USA, Australia, Finland, New Zealand, and China (low- and high-selenium areas). (To convert nanograms of Se per milliliter to micromoles per liter, multiply by 0.0127.)

that over a wide range of intakes, urinary excretion accounts for 50–60% of the total amount excreted. Measurement of the plasma renal clearance of selenium, which expresses its rate of excretion in the urine in terms of the amount contained in a unit volume of plasma, shows that the kidneys of residents of the low-selenium country, New Zealand, excrete selenium more sparingly than those of North Americans, and Chinese from low-selenium areas have an even lower renal clearance. This indicates a possible adaptation to low selenium status.

Trimethylselenonium ion, one of several urinary metabolites, was once thought to be a detoxification product of selenium and appears to be a minor metabolite in humans. Another methylated metabolite, methylselenol, may constitute a larger proportion of urinary selenium. Small losses of selenium occur through the skin or hair or, at high intakes, in expired air as volatile dimethylselenide.

Excretion of Selenium in Milk

Most of the selenium in human milk is protein-bound; at least nine selenium-containing proteins have been detected in milk with glutathione peroxidase accounting for 15–30% of total selenium.

Role of Selenium in the Body

Selenoproteins

Selenium exerts its biological effect through several selenoproteins, of which there may be upwards of 30 in mammalian systems. Identification and characterization of many of these proteins in recent years have resulted in major advances in our understanding of the function of selenium. The following selenoproteins have been purified and studied.

- glutathione peroxidases;
  - cytosolic, cellular
  - gastrointestinal
  - plasma
  - phospholipid hydroperoxide
- selenoprotein P;
- iodothyronine 5′-deiodinases;
- sperm capsule selenoprotein;
- selenoprotein-W;
- thioredoxin reductase;
- selenophosphate synthetase;
- 58-, 56-, and 14-kDa Se-binding protein.

Glutathione peroxidases The first of these selenoproteins to be characterized was glutathione

peroxidase, which consists of four identical subunits, each containing one selenocysteine at the active site. Activity of this enzyme can be reduced to less than 1% in tissues of selenium-deficient animals. Glutathione peroxidase present in cells (including erythrocytes), plasma, and the gastrointestinal tract may function *in vivo* to remove hydrogen peroxide, thereby preventing the initiation of peroxidation of membranes and oxidative damage. However, the significance of this function in the body is uncertain, and it seems likely that the oxidant defense role for selenium is exerted more through other selenoproteins. Glutathione peroxidase may have more specific functions in arachidonic acid metabolism in platelets, microbiocidal activity in leukocytes, and the immune response mechanism or perhaps as a storage protein.

Another selenium-containing enzyme, phospholipid hydroperoxide glutathione peroxidase, is different from the classic glutathione peroxidase in that it can metabolize fatty acid hydroperoxides that are esterified in phospholipids in cell membranes. This enzyme can inhibit microsomal lipid peroxidation and is most likely the basis of the selenium/vitamin E interaction in the pathogenesis of several deficiency diseases in animals.

Selenoprotein P A plasma protein designated selenoprotein-P has been purified and characterized from rat and human plasma and is also associated with endothelial cells. Selenoprotein-P is a glycoprotein containing selenium as selenocysteine, and its concentration in rat plasma falls to less than 10% of the control in selenium deficiency. Its function is unknown, but it may have an extracellular oxidant defense role because its presence correlates with selenium protection of selenium-deficient rats against diquat-induced lipid peroxidation and liver necrosis. A transport role has also been suggested for this protein.

Iodothyronine deiodinases The discovery that type I iodothyronine 5′-deiodinase and, more recently, type II and type III deiodinases are selenoproteins indicates a role for selenium in the metabolism of thyroid hormones that are essential for growth and development. This enzyme catalyzes the conversion of thyroxine (T_4) to its active metabolite triiodothyronine (T_3) in the liver and kidney, and selenium deficiency results in an increase in levels of plasma T_4 and a corresponding decrease in levels of more active T_3. Selenium is preferentially supplied to this enzyme rather than glutathione peroxidase when the selenium supply is inadequate. The interactions of selenium and iodine deficiencies have implications for both human and animal health and livestock production. In humans, selenium deficiency can either exacerbate or ameliorate the effects of concurrent iodine deficiency as has been observed in the development of iodine deficiency disorders in Zaire.

Thioredoxin reductase A recent addition to the list of selenoproteins is thioredoxin reductase, an NADPH-dependent flavoenzyme that reduces the disulfide of thioredoxin. This enzyme regenerates ascorbic acid from dehydroascorbic acid. The activity of thioredoxin reductase declines in selenium deficiency, and there is evidence that selenium is present in the active site as selenocysteine.

Selenoprotein W Selenoprotein W is a selenoprotein found in muscle and other tissues. Its concentration decreases during selenium deficiency, and it may be involved in the development of muscular degeneration in selenium-deficient sheep.

Other selenoproteins Several other selenium-containing enzymes have been identified in microorganisms, and other selenoproteins have been found in animal tissues, suggesting further functions for selenium. These include a specific selenoprotein in prostate tissue, a mitochondrial selenoprotein in sperm, and a 14-kDa and 56/58-kDa selenium-binding protein. One of two selenophosphate synthetases recently identified contains selenocysteine, and it is possible that this enzyme may be involved in the regulation of selenium homeostasis.

Selenium and Host Defense Towards Viruses

The seasonal nature of Keshan disease suggested the likely involvement of an infectious agent, which led to recent work on host response to myocarditic and amyocarditic strains of Coxsackie virus B_3 in mice. Selenium deficiency potentiated cardiotoxicity of myocarditic strains, but in addition, the amyocarditic strain caused heart lesions in selenium deficient mice. This indicated that this specific nutrient deficiency allowed a benign virus to become virulent, apparently as a result of a change in the viral genome. The observation that selenium deficiency may accelerate viral evolution may help to clarify the etiology of Keshan disease and may also be applicable to other RNA viruses. Vitamin E deficiency had effects similar to those observed with selenium deficiency, which suggests the involvement of antioxidant properties of the two nutrients. This is the first reported instance of an influence of host nutritional status on the genetic composition of the pathogen.

Selenium and Immune Function

Adequate selenium is essential for a variety of aspects of the immune function. Selenium is necessary for the development of the acquired immune system, and has a role to play in the defense system of animals against bacteria and other infections. The recent finding of a role for selenium in cell-mediated immunity to Coxsackie virus in mice indicates a protective effect of selenium against viral infection. The mechanisms for the involvement of selenium in the immune system are likely to be related to its antioxidant function through one of the antioxidant selenoproteins glutathione peroxidase, phospholipid glutathione peroxidase, or selenoprotein-P.

Deficiency

Interaction between selenium and vitamin E is observed in the etiology of many deficiency diseases in animals and pure selenium deficiency is in fact rare. Thus, selenium deficiency may only occur when low selenium status is linked with an additional stress such as chemical exposure, increased oxidant stress due to vitamin E deficiency, exercise, or increased dietary intake of polyunsaturated fatty acids. Although residents in some low-selenium areas have low blood selenium, glutathione peroxidase activity, and Selenoprotein P levels, there is little evidence that these are suboptimal or have resulted in any noticeable oxidative damage or changes in other oxidative defense mechanisms. Moreover, people have not shown noticeably improved health when glutathione peroxidase activity is saturated by selenium supplementation. However, an exception to this general rule may be found in the problem of Keshan disease in China. Whether any of the newer functions of selenium are suboptimal in persons with low selenium status is being investigated.

Selenium-responsive Diseases in Humans

Keshan disease, an endemic cardiomyopathy occurring in low-selenium areas of China, was reported in 1979 to be responsive to supplementation with sodium selenite. The principal pathological finding is multifocal necrosis of the myocardium that causes cardiac enlargement, congestive heart failure, cardiogenic shock, and death. The disease is associated with low selenium intake and low blood and hair levels, and affects mainly children and women of childbearing age. It is probably the only case of naturally occurring selenium deficiency. Because some features of Keshan disease (e.g., seasonal variation) cannot be explained solely on the basis of a very low selenium status, Chinese researchers have suggested that other factors may be involved, such as a virus, mineral imbalance, or environmental toxins. Recent research on Coxsackie virus B-induced myocarditis in selenium-deficient mice supports a possible viral involvement in Keshan disease. Another disease that has been associated with poor selenium status in China and Russia is Kashin–Beck disease, an endemic osteoarthritis that occurs during preadolescent or adolescent years. Again, other etiologic factors such as contamination of grain by fungi and high concentrations of organic matter (e.g., fulvic acid) may be contributing factors. The main features of the disease are shortened stature and joint deformation, resulting from multiple focal necroses in the growth plate of tubular bones.

Selenium deficiency has been associated with long-term intravenous nutrition, because of the low levels of selenium in the fluids. Clinical symptoms of cardiomyopathy, muscle pain, and muscular weakness are responsive to selenium supplementation, but are not seen in all patients with extremely low selenium status, indicating that there may be other interacting factors. Furthermore, children on very low selenium synthetic diets for inborn errors of metabolism such as phenylketonuria do not develop selenium-deficiency syndromes.

It is not clear whether syndromes in animals such as the selenium-responsive muscular dystrophy observed in sheep are related to human diseases. Although persistent anecdotal reports from New Zealand farmers in low-selenium areas indicate their conviction that selenium relieves the farmers' muscular aches and pains, double-blind trials have failed to give a clear-cut answer.

Selenium and Cancer

Several lines of scientific enquiry suggest that an increased risk of cancer occurs as a result of low levels of selenium in the diet. An association between selenium and cancer was first proposed 30 years ago when it was observed that regional cancer mortality rates in the USA correlated with selenium exposure, as reflected by concentrations in plants. Evidence for the role of selenium as an anticarcinogenic agent comes from *in vitro* and animal studies that suggest that selenium is protective against tumorigenesis at high levels of intake and from case-controlled prospective studies in human subjects. However, epidemiological evidence linking a low selenium status with an increased incidence of cancer is conflicting. Few intervention studies in humans have been carried out. In Linxian China, a significant effect of a combination of vitamin E, β-carotene and selenium was observed on mortality, particularly from esophageal cancer. Recently a 10-year controlled clinical trial to test the efficacy of selenium in preventing skin cancer

was completed in the USA. There was no effect of daily supplementation with 200 μg of selenium on the primary endpoint skin cancer, but there was a statistically significant reduction in several additional endpoints such as total cancer mortality (50%), and incidence of cancer of the prostate (63%), lung (46%), colorectal (58%), and total cancer. Further studies are needed to confirm these observations.

Selenium and Cardiovascular Disease

Similarly, dietary deficiency of selenium has been implicated in the etiology of cardiovascular diseases, but the evidence at present is less convincing than for cancer. Although a large case-control study in Finland suggested that selenium was an independent risk factor for myocardial infarction in a low-selenium population, evaluation of prospective epidemiological studies has failed to provide sufficient evidence to implicate selenium deficiency in most aspects of cardiovascular disease. While some investigations have observed a relationship between low serum-selenium levels and risk of coronary disease, others have not. However, selenium may have some role in protection against thrombosis and low-density lipoprotein oxidation, in particular in individuals such as smokers at risk from increased oxidant stress. Further evidence must come from controlled intervention trials to clarify any possible role for selenium in atherosclerotic disease.

Toxicity

The margin between an adequate and toxic intake of selenium is quite narrow. Overexposure or selenosis may occur from consuming high-selenium foods grown in seleniferous areas in Venezuela and some areas of China. People following long-term liberal megadosing can also attain an undesirably high selenium status. The most common sign of poisoning is loss of hair and nails, but lesions of the skin, nervous system, and teeth may also be involved. Garlic odor on the breath is an indication of excessive selenium exposure resulting from expiration of dimethylselenide. Sensitive biochemical techniques are lacking for selenium toxicity, which is at present diagnosed from hair loss and nail changes. Some effects of selenium toxicity are seen in individuals with dietary intakes as low as 900 μg, and the maximum safe dietary intake has been suggested as 400 μg per day.

Assessment of Selenium Status

Blood selenium concentration is generally considered a useful measure of both selenium status and intake, but other tissues are often assessed as well. Plasma selenium reflects short-term status and erythrocyte selenium long-term status, but blood selenium concentrations are influenced by the chemical form of selenium ingested as a result of the different patterns of absorption and retention outlined previously. Toenails are often used, but selenium-containing shampoos restrict the use of hair. Urinary excretion can also be used to assess selenium status, and total dietary intake is estimated as twice the daily urinary excretion.

The close relationship between blood or red-cell glutathione peroxidase activity and selenium concentrations (**Figure 3**) is useful for assessment in people with a relatively low status, but not once the saturating activity of the enzyme is reached at blood selenium concentrations above $100\,\mu g\,l^{-1}$ ($1.27\,\mu mol\,l^{-1}$). More recently, mesurement of selenoprotein P has been used to assess selenium status, and there is the potential for measurement of other enzymes as functional markers. However, their use is limited at present by the lack of simple assay techniques. Furthermore, the conclusions drawn from measurement of one selenoprotein may not apply to all biological functions of selenium because of the differences in responses of tissues and these proteins to deficient, adequate, or high levels of selenium in terms of their 'hierarchy' of importance. Therefore, there may be no single indicator of functional selenium status, but rather a series of markers that apply to specific problems associated with suboptimal selenium status. The situation is further complicated by the large number of interacting factors, including, protein, methionine, polyunsaturated fatty acids and other oxidant stressors, vitamin E, other trace elements, and heavy metals such as mercury, cadmium, and lead.

Dietary Intake

Food is the major source of selenium; drinking water contributes little. Dietary intake varies with the geographic source of foods and eating habits of the people. Plant food concentrations reflect selenium content of soils and its availability for uptake, as plants do not require selenium for growth; cereals and grains grown in soils poor or rich in selenium may vary over 100-fold in selenium content. Animal foods vary less. Fish and organ meats are the richest sources followed by muscle meats, cereals, grains, and dairy products, with fruits and vegetables mostly poor sources. Average daily dietary intakes vary considerably, depending on the levels of selenium in soils (**Table 1**), ranging from 10–20 μg of selenium in low soil-selenium areas of China where Keshan disease is endemic, 30–60 μg in New Zealand, and up to 200 μg

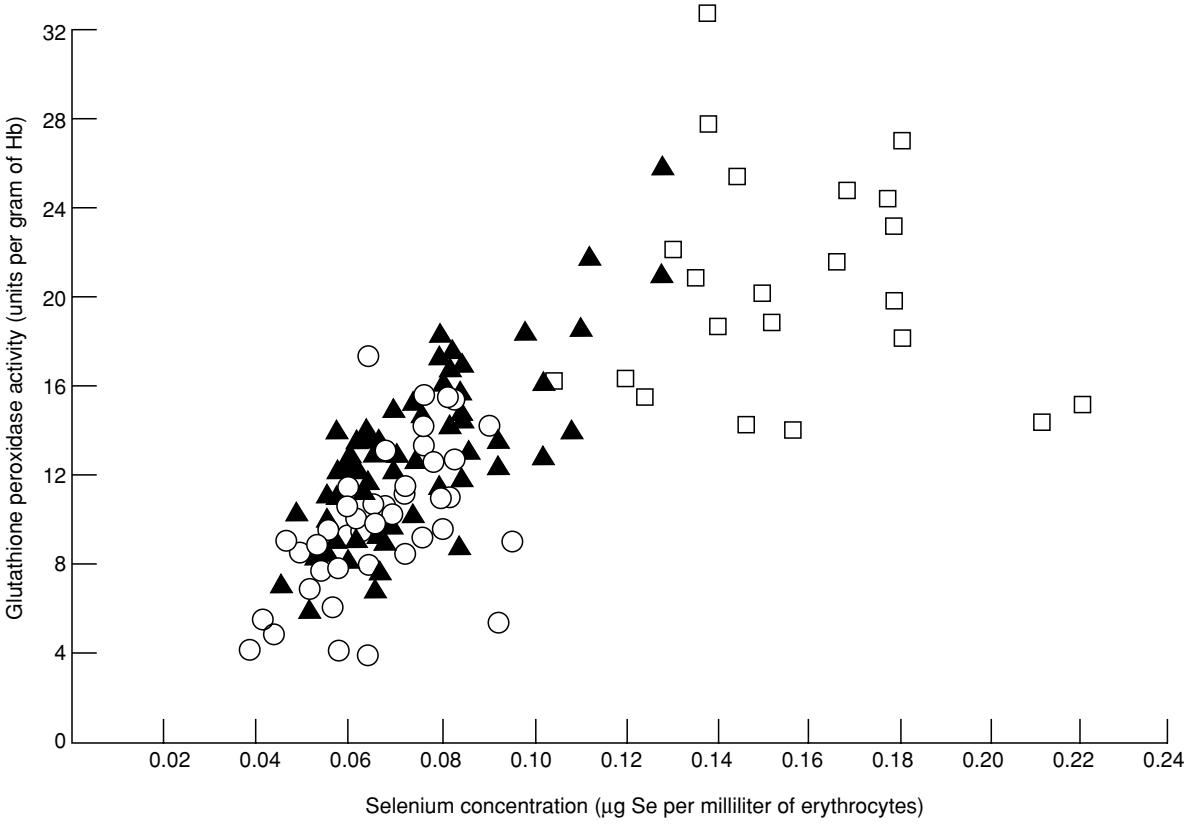

Figure 3 Relationship between selenium concentration of erythrocytes and glutathione peroxidase activities for New Zealand residents: Otago patients (○): Otago blood donors (▲) and overseas subjects (□). (To convert nanograms of Se per milliliter to micromoles per liter, multiply by 0.0127.) From Rea HM, Thomson CD, Campbell DR and Robinson MF (1979) Relation between erythrocyte selenium concentrations and glutathione peroxidase (EC 1.11.1.9) activities in New Zealand residents and visitors to New Zealand. *British Journal of Nutrition* 42: 201–208, with permission.

Table 1 Daily dietary intakes of selenium and whole blood selenium concentrations

Country	*Selenium intake (μg per day)*	*Blood selenium ($\mu g\,l^{-1}$)*
China (Keshan area)	9–11	10–20
New Zealand (before 1990)	28–32	50–100
(post 1990)	30–60	60–120
Finland (1985)	30	50–100
Finland (1989–1990)	100	120
UK	60	80–320
USA	62–216	150–400
Venezuela (Caracas)	218	350
China (seleniferous area)	5000	3200

From Thomson CD (2002) Selenium. In: Mann JI and Truswell AS (eds) *Essentials in Human Nutrition*, 2nd edn, pp. 172–181. Oxford: Oxford University Press, with permission.

in seleniferous areas in Venezuela. In 1985, selenium was added to fertilizers in Finland to increase selenium intake throughout the population, and the daily intake rose from 40 μg to close to 100 μg per day, resulting in an increase in serum selenium in healthy individuals from 65–70 $\mu g\,l^{-1}$ in 1985 to 120 $\mu g\,l^{-1}$ in 1989–1991. In New Zealand, intakes are increasing due to increasing importation of Australian wheat, and selenium status is rising accordingly. However, this increase is greater than expected from the increase in wheat selenium, indicating that other dietary factors are involved. These differing intakes in various countries are reflected in the wide range of blood selenium concentrations found in residents of these countries (Figure 2).

Requirements and Recommended Dietary Allowances (RDAs)

Many countries have proposed recommended dietary intakes based upon estimates of requirements from Chinese intakes for endemic and nonendemic Keshan disease areas, as well as intakes at which maximal levels of plasma glutathione peroxidase activity occurred. The recommended dietary intake for Australia was the first set in the world in 1986 and may have been cautiously on the high side at 85 and 70 μg

Table 2 Recommended intakes of selenium for adults (μg/day)

	Australia (1990) RDI	*USA & Canada (2000) RDA*	*UK (1991)*		*Germany, Austria, Switzerland (2000) RNI*	*WHO*[a] *(1996) NR*
			LNRI	*RNI*		
Men	85	55	40	75	30–70	40
Women	70	55	40	60	30–70	30

[a]World Health Organization (1996) *Trace Elements in Human Nutrition and Health*, pp. 105–122. Geneva: WHO.
RDI, recommended dietary intake; RDA, recommended dietary allowance; RNI, reference nutrient intake; LNRI, lower reference nutrient intake; NR, normative requirement estimate.

selenium for Australian men and women. Others, including the recommended intakes of the USA, the UK, and other European countries, are summarized in Table 2. Each set of recommended intakes can be met from habitual diets in each country. Whether optimal health depends upon saturation of glutathione peroxidase activity has yet to be resolved, leading to disagreement regarding the use of this approach for assessing selenium requirements. The US RDAs are based on the desirability of the full activity of glutathione peroxidase, whereas a WHO group concluded that only two-thirds of the maximal activity was needed, based on the observation that abnormalities in metabolism of hydrogen peroxide in blood cells is apparent only when enzyme activity falls to one-quarter or less of normal. This has led to a wide range of dietary recommendations, with the WHO recommendations approaching intakes in countries with a naturally low selenium status such as New Zealand. In the future, several of the newly discovered selenoproteins might be used as endpoints for determining selenium requirements. However, the maximal activity of some of these proteins occurs at dietary intakes of selenium less than those needed for maximal glutathione peroxidase activity. This would no doubt lead to dietary recommendations lower than current values.

See also: **Bioavailability of Nutrients**; **Cadmium**: Toxicology; **Cancer**: Diet in Cancer Prevention; **Coronary Heart Disease**: Prevention; **Dietary Requirements of Adults**; **Mercury**: Toxicology; **Tocopherols**: Physiology

Further Reading

Arthur JR, Beckett GJ and Mitchell JH (1999) The interactions between selenium and iodine deficiencies in man and animals. *Nutrition Research Reviews* 12: 55–73.

Barceloux DG (1999) Selenium. *Clinical Toxicology* 37: 145–172.

Beck MA (1998) The influence of antioxidants on viral infection. *Nutrition Reviews* 56: S140–S145.

Burk RF and Levander OA (1999) Selenium. In: Shils ME, Olson JA, Shike M and Ross CA (eds) *Modern Nutrition in Health and Disease*, 9th edn, pp. 265–276. Baltimore, MD: Williams & Wilkins.

Dreosti I (1990) Selenium. In: Truswell A, Dreosti I, English R, Ritihauser I and Palmer N (eds) *Recommended Nutrient Intakes Australian Papers*. Sydney: Australian Professional Publications.

Levander OA and Burk RF (1996) Selenium. In: Ziegler EE and Filer LJ (eds) *Present Knowledge in Nutrition*, 7th edn, pp. 320–328. Washington, DC: ILSI Press.

Reilly C (1996) *Selenium in Food and Health*. London: Blackie Academic & Professional.

Robinson MF (1989) Selenium in human nutrition in New Zealand. *Nutrition Reviews* 47: 99–107.

Thomson CD (1998) Selenium speciation in human body fluids. *Analyst* 123: 827–831.

World Health Organization (1996) *Trace Elements in Human Nutrition and Health*, pp. 105–122. Geneva: WHO.

SENSORY EVALUATION

Contents

Sensory Characteristics of Human Foods

R York, Canadian Food Inspection Agency, Winnipeg, Manitoba, Canada
M Vaisey-Genser, The University of Manitoba, Winnipeg, Manitoba, Canada

Background

Satisfaction with the sensory features of foods and beverages is a major determinant of the consumer's nutritional status. If something does not taste 'good,' it is not likely to become part of the food habit. However, the colloquial term 'good' is a reflection of an integrated response to the several food features that are considered in detail in following sections. The present discussion is an introduction to sensory evaluation that focuses on its ongoing evolution as a discipline.

Definitions

Early work in sensory evaluation was referred to as organoleptic testing, a term that was discarded in the 1960s because it failed to recognize the complexity of the sensory impression. In 1975, the US Institute of Food Technologists defined the subject as a scientific discipline used to evoke, measure, analyze, and interpret reactions to those characteristics of foods and materials as they are perceived by the senses of sight, smell, taste, touch, and hearing. This definition has since been extended to include the purpose of providing information for decision-making with economic consequences. As the field increases in sophistication, the term 'sensory science' is becoming the descriptor of choice rather than sensory evaluation. The sequence and the strategy for sensory science measurements is illustrated in Figure 1.

Justification

Sensory measurements provide a unique resource of information for the evaluation of foods. Humans are sensitive and complex instruments capable of measuring and integrating simple and complex stimuli, repeating measurements after a brief resting stage, and differentiating relevant from irrelevant information. Accordingly, most physical tests for 'ideal' product characteristics are based on human definition. The results are reliable and repeatable as long as the test conditions are controlled.

History

Among sensory features, aroma seems to have captured major interest in early times, perhaps reflecting the importance of spices for foods and of other aromatics used in perfumes and oils. In 320 BCE (Before the Common Era), a treatise on odor was put forth by Theophrastus in Greece. That was followed by the occasional classic on the nature of smell capped by Linnaeus, odor classification, which appeared in 1756. Over time, the 'expert taster' became relied upon within specific commodities, most notably teas, wines, and other alcoholic beverages.

Sensory evaluation as we know it today stems from the attention to methodology that began in the latter part of the nineteenth century. This included methods as simple as the score card to monitor commodity quality and as sophisticated as the Weber–Fechner Law, which related the sensory response quantitatively with incremental changes in a physical stimulus. This spawned the discipline of psychophysics, a

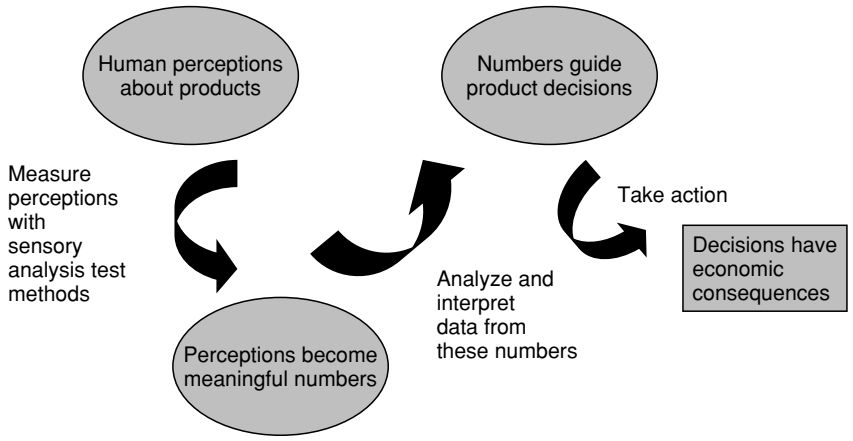

Figure 1 Sequence and strategy of sensory science measurements.

hybrid of psychology and physiology. The late nineteenth century saw the introduction of statistical methods for handling sensory data.

There were two sensory landmarks in the twentieth century: the development of statistically sturdy measurement methods and attention to the selection, training, and performance monitoring of sensory panelists. In the early 1930s, rating scales were developed to measure the intensity of specific food qualities such as the tenderness of meat, and concurrently, paired comparison and rank order methods were introduced. The early 1950s was marked by the introduction of the flavor profile system for the qualitative analysis of foods by the Arthur D. Little Company (ADL) in Massachusetts. ADL introduced training of selected members from an industry in the systematic 'descriptive analysis' of their products.

The mid-1950s saw further advances in statistical methods for both design and analysis as well as the launch in the USA of forced-choice methods such as duo–trio and the triangle test, which originated in the USA and Sweden, respectively. These tools are still widely used by industry for sensory testing in their quality control laboratories. A singular contribution from this era was the nine-point hedonic scale for measuring how much a food was liked or disliked. Developed by the food acceptance branch of the US Quartermaster Food and Container Institute in Chicago, this tool has been used far beyond its original purpose, which was to assist the development of more acceptable rations for the military. It remains a worldwide standard tool for food-preference measurement.

The momentum in sensory evaluation continued throughout the 1960s marked by three major events: (1) the 1961 formation of Committee E-18 within the American Society for Testing of Materials (ASTM), (2) the publication in 1963 of the Texture Profile Method by scientists at General Foods Corporation led by Alina Szczesniak, and (3) the 1965 Academic Press publication of *Principles of Sensory Evaluation of Food*, written by Amerine, Pangborn, and Roessler at the University of California in Davis, a book that promptly became a standard text. Further steps forward were the 1968 opening of the Monell Chemical Senses Center in Philadelphia to concentrate on basic research of the senses under the leadership of physiologist Morley Kare, and the 1969 introduction of the *Journal of Texture Studies*.

In the early 1970s, the International Organization for Standardization/Organization Internationale de Normalisation (ISO) formed a subcommittee, TC34/SC12, for Sensory Evaluation with the aim of having internationally agreed upon standards for sensory methodologies. This group currently includes representatives from 17 countries as members and 30 countries as observers.

During the 1970s and 1980s, newer methods of data handling erupted with the maturing of the computer age. Multivariate analyses became the norm, facilitating methods such as response surface and principal-component analyses. Communication worldwide was sparked by international symposia such as the Mediterranean Odor Conference in Cannes in 1970 followed by the later conferences in Switzerland, Sweden, and the UK. New journals such as 1974s *Chemical Senses and Flavor* (now *Chemical Senses*), and later, in 1986, the *Journal of Sensory Studies* emerged followed by *Food Quality and Preference* in 1988. The 1990s saw the formation of the Sensometrics Society for the development of the special methodologies and statistical methods, which were needed by the fields of sensory and consumer science.

By the 1990s, electronic mail encouraged speedy interaction among sensory scientists worldwide. Other tools continued to evolve. The beginning years of the twenty-first century are bringing intriguing tools such as the electronic nose to the challenge of sensory research.

Disciplines

Figure 2 describes the root disciplines involved in sensory evaluation together with those that have been most active in its applications. This picture confirms the interdisciplinary nature of the field. As each of these grows in sophistication, so does sensory science.

Some examples of the contributions of each of the basic areas are:

- *Psychology/Psychophysics*: the study of the laws of sensory perception, basic information on testing methodologies and the requirements for effective testing (from ballot design to sample presentation to testing protocols), sources of psychological error in sensory testing, the study of cognitive functions in the development, and use of terminology in sensory testing.
- *Physiology*: the functioning of the sensory systems, sources of physiological error in sensory testing.
- *Statistics*: the data analysis methods used to date and the development of new methods specific to sensory analysis.

Further examples from the applied areas include:

- *Home economics/food science*: applications of sensory analysis to the production of foods as basic commodities and in processed forms including product stability, product quality, consumer preference.
- *Chemistry*: the instrumental techniques for the analysis of foods and other products tested in sensory science.
- *Management/marketing*: methods for the assessment of consumer behavior.
- *Engineering*: the development of sensor-based technology in the forms of the electronic nose and the electronic tongue, and of neural networks.

Classifications of Areas of Sensory Testing

The most fundamental division of sensory measurement is determined by whether the subject is asked how they 'fit' relative to the stimulus or whether the subject is required to make an analytical judgement of the stimulus. This is given in sensory analysis by the terms subjective and objective testing (see Table 1). Answering a question using sensory analysis requires the development of a strategy for the approach, including the identification of the question that is really being asked and formulating an experimental plan to obtain appropriate data.

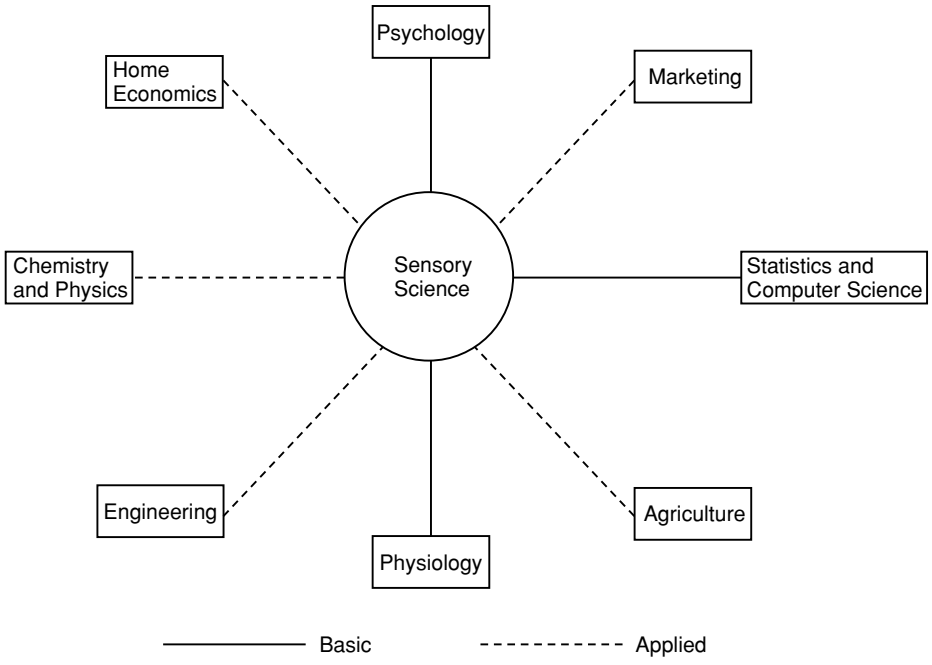

Figure 2 Disciplines contributing to sensory science.

Table 1 General characteristics of subjective *vs.* objective tests

	Subjective test	*Objective test*
Measures properties of:	Person (opinion)	Products (level or intensity of attribute)
Approach to test:	Emotional	Analytical
Assessor selected for:	Represents a population segment or a group with particular product experience	Demonstrated ability to perform sensory measurement tasks
Training level:	Untrained (naïve response)	Trained (educated response)
Number of assessors:	Many (30–50 for preliminary test, up to several hundred for final test)	Few (usually four to 10)
Performance evaluation of assessors:	No	Yes
Terms used on ballot:	Value judgements (like/dislike, good/bad, accept/reject)	Descriptive of phenomena (sweet, salty, red, tough, etc.)
Method described in early sensory evaluation literature as:	Organoleptic testing	Sensory analysis

Subjectivity and Objectivity

Subjective tests are those that evaluate the behavior of the assessor relative to the sample, i.e., they measure the opinion of a person who is using a product and are more commonly referred to as 'consumer tests.' They imply a response that is affected by personal bias, emotional background, etc., and describe the individual. These tests are not necessarily validated through demonstration of sensory ability. The term subjective is often incorrectly used to describe all sensory tests.

Objective tests are those that measure the presence and/or intensity of specific characteristics within a sample. These tests imply a response that is based on external phenomena or events, and the approach of the panelist is expected to be detached, impersonal, and unprejudiced. This term is used in sensory evaluation to define those tests that are based on demonstrated results with selected, trained judges appropriate to the test procedure.

Sensory test methodologies have been classified in different ways – by the overall purpose of sensory testing, by format of test methodologies, and by the type of primary 'tool' used, i.e., the type of sensory assessor.

For example, Michael O'Mahony in 1988 classifies the world of sensory evaluation into four types of testing, based on test function and experimental goals:

- Sensory Evaluation I: sensory evaluation used as a basic tool to study chemical and physical characteristics of food systems, e.g., a change in a flavoring would result in a change in a volatile chemical and thus a perceived change in odor. Judges are selected for sensory ability.
- Sensory Evaluation II: testing with consumers to measure the degree to which product changes can be detected by untrained assessors. Assessors can be trained in sensory methodology, but not in the product characteristic.
- Consumer testing: preference testing, etc.
- Sensory psychophysics: studies of testing methodologies and assessor sensitivities.

Several workers have used a classification system by the type of assessor chosen. In this classification, there are four basic types of assessors:

1. Naïve untrained assessors: consumer testing; used to measure subjective opinions about a sample.
2. Selected assessors trained for specific short-term tasks: laboratory tests for specific attributes; objective tests including the difference tests and the rating scales that measure specific product characteristics.
3. Descriptive analysis panels: groups selected and highly trained assessors that work as an expert group; objective procedure to evaluate specific products or product groups for complete ranges of product characteristics.
4. Expert assessors: selected, highly trained, and employed to work as individuals; objective testing of specific product or product group by an individual who has responsibility for the samples and the data about them.

Sensory Science and Psychophysics

Psychophysics defines those aspects of sensory science that include the way in which stimuli are perceived and the way in which they are measured. Some of the pioneering work in psychophysics was done by the psychologist L. L. Thurstone, mainly in the areas of psychology, but also in applications to food testing. Psychophysics itself has been defined as 'concerning the functional relationship between stimulus and response.' Work in this area includes not only the relationship between the perceived intensity and physical intensity of the stimulus, but also the ways in which judge sensitivity can change in test situations, the nature of language and terminology development

and its relationship to everyday language, and the development and processing of sensory concepts.

An example of the understanding this discipline provides is from the work by Stevens, which categorized sensory measurement as involving three parameters – the task undertaken by the assessor, the manner of the stimulus, and the statistical measure used to describe the data.

The task undertaken by the assessor was further defined by four groupings, which are generally accepted as defining the data collected in sensory testing:

1. Classification or nominal data: where the observer judges whether the stimulus is present or not. The data consist of meaningful categories that can be clearly differentiated. No numerical values are associated with the categories within this group – the data analysis includes frequency data only. One example of these data is given by the numbers on football jerseys – numerical handling of these data, e.g., a mean of the numbers, is meaningless.
2. Order or ordinal data: consisting of meaningful categories that can be clearly differentiated and placed in order of magnitude, e.g., greater or lesser of some defined characteristic, ranked data.
3. Interval data: involving judging the apparent difference between two or more perceptions and assigning values to the intensity of the perceptions using a continuum that is divided into apparently equal intervals.
4. Ratio or magnitude data: involving reporting the ratio of magnitude between two or more perceptions or the apparent magnitude of a perception.

Another example of the contribution of psychophysics is the development of a numerical expression of the relationship between the physical intensity of a stimulus to its perceived sensory intensity. This began with the Weber Fraction and Fechner's semilog relationship between these phenomena. This evolved into the development of Stevens' Power law as the basis for sensory work with the definition:

$$I = kS^n,$$

where I = perceived intensity; S = stimulus intensity; n = the exponent of the power function; k = a constant. Once this function was established, Stevens' Law offered useful predictive power for the study of sensory perception and the relation to other product attributes.

Work on the application of Thurstonian methods to sensory data analysis is currently underway through the activities of one of the ASTM E18 Subcommittees and will result in a standard for this application.

Sensory Science and Advancements in Methodology Development

Developments in data collection methodologies in the evolution of sensory science have included landmarks such as the nine-point hedonic scale, category scaling systems for descriptive analysis, the line scale method of Quantitative Descriptive Analysis, ratio scaling methods for both perception and preference, and the labeled magnitude scale, first developed for the study of pain perception and now applied to odor and taste measurements as well as to preference measurements.

Advances in sensory methodology have been enhanced by the recent advances in computer technology and the easy availability of both hardware and software for the complex analyses that are often required. This has provided sensory testing with a greater flexibility in data collection and analysis. The sensory laboratory is no longer limited to paper ballots and the more cumbersome manual handling of data. Laboratories can now be equipped with data acquisition hardware and software in the form of computer terminals that allow the ballots to be presented and recorded as data without requiring manual handling of the data. Other systems allow the collection of consumer information using telephone and internet formats.

Consumer testing has evolved greatly since its first uses in sensory science with the work of the US Quartermaster Corps and the evaluation of military rations. Part of this was the development of the nine-point hedonic scale and the application of this to preference testing. While this method is still used, the current work in this field now includes greater breadth in the considerations given to the information which is needed to make product decisions from consumer responses. This includes the interactions of preference with underlying sensory mechanisms and with various cultural experiences that cause preference to vary between different groups. These groups can be characterized by factors such as sensory acuity to particular stimuli and particular demographics such as age, cultural background, etc. This has led to the developments in methodology such as sensory segmentation, preference mapping, applications of advanced design methods such as conjoint analysis, and structural equation modeling.

The use of expert assessors has developed as well. Expert sensory panels for complex descriptive analysis have evolved in both training and application systems. The use of expert assessors in regulatory and industrial applications is well established and, in current practice, relies on the basic knowledge of sensory science methodologies.

Resources in Sensory Science

Sensory evaluation is currently recognized as an established discipline with courses and/or programs in many universities all over the world. Major international food corporations have units specializing in sensory analysis. Additionally, there are independent companies with facilities specializing in sensory applications and advice in different areas of application.

Two contemporary international events especially for sensory scientists are the Pangborn Sensory Science Symposium and meetings of the Sensometric Society, which are now scheduled for alternate years. Historical attendance at these meetings has been impressive, with 600 registrants for the 2001 Pangborn Symposium in France.

Publications are keys to communication in any discipline. *Chemoreception References* circulated by the European Chemoreception Research Organization since 1984 has proved to be a particularly valuable resource for sensory scientists. In 2002, the Sensometric Society chose *Food Quality and Preference* as its official journal. The primary criterion for acceptance for this journal is that the work include some form of human measurement. Standards for, and information on, testing methodology have been developed by the ISO and ASTM and can be ordered through their respective websites.

Conclusion

Sensory evaluation is a good example of an interdisciplinary success. Its maturation has been relatively rapid in the history of science, due to the strong leadership provided by imaginative researchers spanning a range of traditional and applied disciplines. Its evolution continues in all areas in which human perceptions are used to evaluate food and other products.

See also: **Sensory Evaluation**: Food Acceptability and Sensory Evaluation; Descriptive Analysis; Appearance; Texture; Aroma; Taste; **Spices and Flavoring (Flavouring) Crops**: Use of Spices in the Food Industry

Further Reading

Amerine MA, Pangborn RM and Roesler EB (1965) *Principles of Sensory Evaluation of Food*. New York: Academic Press.

Boring EG (1942) *Sensation and Perception in the History of Experimental Psychology*. New York: D. Appleton Century.

Chambers IVE and Wolf MB (1996) *Sensory Testing Methods, ASTM MNL 26*. West Conshohocken, PA: ASTM.

IFT Sensory Evaluation Division (1981) Sensory evaluation guide for testing food and beverage products. *Food Technology* 35: 50–59.

Lawless HT and Heymann H (1999) *Sensory Evaluation of Food – Principles and Practices*. Gaithersburg, MD: Aspen.

Lawless HT and Klein BP (1991) *Sensory Science Theory and Application in Foods*. New York: Marcel Dekker.

McBride RL and MacFie HJH (1990) *Psychological Basis of Sensory Evaluation*. London: Elsevier Applied Science.

Meilgaard MC, Civille GV and Carr BT (1999) *Sensory Evaluation Techniques*, 3rd edn. Boca Raton, FL: CRC Press.

Moskowitz HR (1988) *Applied Sensory Analysis of Foods*, vol. 1. Boca Raton, FL: CRC Press.

Munoz AM, Civille GV and Carr BT (1992) *Sensory Evaluation in Quality Control*. New York: Van Nostrand Reinhold.

O'Mahony M (1986) *Sensory Evaluation of Food. Statistical Methods and Procedures*. New York: Marcel Dekker.

O'Mahony M (1995) Sensory measurement in Food Science: Fitting methods to goals. *Food Technology* 49: 72–82.

Pangborn RM (1989) The evolution of sensory science and its interaction with IFT. *Food Technology* 43: 248–256, 307.

Peryam DR (1990) Sensory evaluation – The early days. *Food Technology* 44: 86–91.

Schutz HG (1998) Evolution of the sensory science discipline. *Food Technology* 52: 42–46.

Stone H and Sidel JL (1993) *Sensory Evaluation Practices*, 2nd edn. New York: Academic Press.

Szczesniak AS (1998) Sensory texture profiling – Historical and scientific perspectives. *Food Technology* 52: 54–57.

Yantis JE (1992) *The Role of Sensory Analysis in Quality Control*. Philadelphia, PA: American Society for Testing and Materials.

Food Acceptability and Sensory Evaluation

J M Murray, Quest International Australia, Smithfield, Australia
I A Baxter, Food Science Australia, North Ryde, Australia

Background

Food acceptability is affected by many factors, which may be related to the individual, the food, or the environment in which the food is consumed. Acceptability is a subjective measure based on hedonics

(pleasure), which in turn is influenced by the sensory properties of the food, previous exposure to it and subsequent expectations, contextual factors, an individual's culture, physiological status (i.e., hunger, thirst, and presence/absence of illness), and many other variables. The measurement of food acceptance is highly complex and relies on psychometrics (scales) and/or behavioral models (food-choice models). This article aims to discuss some innate factors that impact on foods' sensory acceptability and how these can be measured and understood using sensory evaluation.

Food Properties and Sensory Acceptability

The acceptability of a food in sensory terms is influenced by the intrinsic properties it possesses, that is, the appearance, aroma, flavor, texture, aftertaste, and auditory attributes of the food. The interaction between these variables is also an important consideration, and the manifestation of sensory attributes can be affected by factors such as temperature, portion size, and so on.

Appearance of a food provides cues regarding its acceptability, e.g., indicating freshness or lack of (the turgid appearance of lettuce, the wrinkled skin of an apple that is past its best, the appearance of green mold on cheese), ripeness (greening in tomatoes), or skill in preparation (e.g., absence of burning, absence of lumps in mashed potato). Color affects acceptability, influencing taste thresholds, taste perception, expected food pleasantness, and acceptability. For example, certain colors are associated with particular flavors so that a yellow sweet may taste 'sour' as it is associated with lemons, and foods that are intensely colored may be perceived as more intense in flavor.

Aroma suggests the flavor to be expected when eating foods and thereby influences acceptance. Volatile compounds are generally responsible for aroma and vary from one food to another; for example, brussel sprouts are rich in volatile sulfurous compounds that contribute to their characteristic odor and flavor, thus influencing acceptance among many consumers. Aroma is also a good indicator of the freshness of a product; think of yourself sniffing a carton of milk to check the freshness before using it. In this respect, aroma can also be regarded as a 'defense mechanism,' by providing the consumer with information about the safety of a food prior to consumption.

Flavor is frequently regarded as the most influential attribute for sensory acceptance, and perception of flavor is a complex process involving the sense of smell, taste, and chemesthesis. A distinction is often made between the terms 'taste' and 'flavor.' The former refers to the four basic tastes, sweet, sour, salty, and bitter (a fifth, umami has also been proposed), and the latter refers to a combination of 'taste' and 'aroma,' with the volatile components of the food mainly contributing to the overall perception. Other sensations such as 'chilli burn,' perceived by the trigeminal nerve, and the 'cooling' sensation of foods such as mints contribute to the overall flavor.

The acceptability of a 'flavor sensation' changes throughout life, and genetic components have been proposed for sensitivity to certain bitter compounds, e.g., phenylthiocarbamide (PTC) and 6-*n*-propylthiouracil (PROP), with consumption of bitter-tasting foods being inversely related to sensitivity. Humans are born with particular taste preferences, indicated by newborns whose facial expressions show a preference for sweet, indifference to salt, and negative reactions to bitter flavors. Evidence suggests that these are innate responses, whilst throughout the development of a child to adulthood, a combination of innate and learned responses determine food acceptability. Aging has a diminishing effect on our taste and aroma thresholds and the ability to differentiate particular flavors, and thus ultimately influences acceptability. Cultural differences in the perception of flavor have also been suggested but have yet to be proved. Flavor perception and release has been the subject of a vast amount of research in the last 10 years and is influenced very much by the food matrix and individual characteristics of the consumer (e.g., rate of saliva production in the mouth). The interaction between flavor and texture is vitally important, with the perception of a flavor changing, depending on the texture of the food. For example, the same amount of cheese flavor on a hard and crunchy corn chip will be released and perceived differently to that on a puffed (extruded) chip.

Texture is perceived by the senses of sight (visual texture), touch (tactile texture), and sound (auditory texture), and its contribution to overall food quality can be relatively simple (a cup of coffee, puréed soup) or complex (steak, potato chip). Consumers use texture information for classifying foods, but they do not normally volunteer comments on texture, unless it is definitely different from the consumers' expectation, it has been suggested that texture awareness is usually on the subconscious level, with awareness heightened only when expectations are violated, or nonfood associations are triggered. Auditory texture in some foods is critical for overall acceptability: dried breakfast cereals that do not snap, crackle, and pop, cola that does not fizz, or snacks that do not crunch would lose their inherent identity and acceptability. The

acoustic properties of foods are also reliable indicators of freshness; imagine an apple that did not crunch when first bitten.

Visual texture contributes significantly to overall acceptability and suggests the texture to be perceived upon consumption. Visual texture can suggest the freshness of the food; for example, a cut slice of cake may look moist or dry and thus may be perceived as stale or fresh. Tactile texture perceived upon consumption is related to the food matrix and size and shape of the food being consumed. Mouthfeel characteristics (which can be, for example, viscosity-, carbonation-, or body-related) impact significantly on acceptance depending on the food, and factors such as melting are vital attributes for categories of food such as icecream and chocolate. The perception of texture is dependent mainly on the teeth and masticatory musculature. Infants are able to consume only liquids until the musculature has developed sufficiently to allow them to ingest semiliquid (puréed) foods. The development of teeth affords children the ability to deal with foods that need biting, tearing, and chewing. In adulthood, most texture challenges can be dealt with, but ageing and the subsequent decline in the quality of dentition and musclature impact on the acceptability of certain textures.

The manifestation of sensory properties can also be affected by factors such as the temperature and portion size, and thus influence perception and acceptability. Changes in temperature can alter the viscosity of a sauce (and thus the texture sensation). The perception of the basic tastes can also be altered: a cold, sweetened beverage is perceived to be less sweet as the tongue is cooled, but as the temperature of the drink rises, the tongue is not cooled, and the drink's sweetness seems to increase. The amount of food placed in the mouth also affects perception: flavor may be perceived as more intense when a larger piece of food is eaten, and less intense with a smaller portion.

Determining Acceptability by Sensory Evaluation

Many complex variables of the food interact with the consumer to influence sensory acceptability, and as such, objective measurement of sensory acceptability proves difficult. However, there is an increasing requirement to understand and measure sensory preferences *reliably* within the food industry. As the focus of most food manufacturers' product development, quality assurance, and formulation efforts, it is vital to ensure that the results of these sensory tests are valid. Attention must therefore be paid to the type of outcomes required from the test, the design of the experiment, sources of bias in sensory evaluation, and the use of appropriate participants. Today, there are a number of approaches that can be used for eliciting both qualitative and quantitative information relating to foods sensory properties and their subsequent acceptability. These range from simple, 'I like it, I don't like it' tests, to complex modeling procedures that uncover sensory reasons for preference and relate these to the physical and chemical properties of the food. Both qualitative and quantitative techniques are described below for measurement of sensory acceptability.

Hedonic Testing

Hedonic trials are performed in order to obtain quantified information on, for example, how much people like a product or how much a new prototype is preferred over the others. They are conducted with large representative samples of the population and encompass both preference and acceptability tests.

Preference Tests

The most commonly employed preference test is the *Paired Preference test*, which is used when one wants to examine the preference of one product directly against a second, as in the case of product improvements, or determine parity or preference over a competitive brand. Participants in a paired preference test receive two samples simultaneously, and the panelist is asked to identify the sample that is preferred. The major advantage of this test is that it is simple, and participants can easily grasp the task. The *Nonforced Preference test* is identical to that described above, except that it allows a no-preference option. This can yield valuable insights into the consumer's state of mind, as there may be participants who genuinely do not have any preference; however, this option does complicate the analysis, the power of the test is decreased, and it can give some consumers an 'easy way out,' as they do not have to, and do not, make a choice. In the case of *Preference ranking tests*, consumers are asked to rank several products in either descending or ascending order of preference or liking. The advantage of this test is, again, that it is intuitively simple for the consumer and can be conducted quickly and with very little effort, but the data from different sets of overlapping products cannot be compared. Visual and tactile rankings are normally quite straightforward, but multiflavor and taste comparisons can be fatiguing.

Acceptability Tests

Acceptability tests still investigate liking for products but do not provide a direct comparison of liking over

other products. The most common acceptability scale is the nine-point hedonic scale, which has labeled categories for indicating a level of like or dislike. The advantages of this scale is that it is simple and easy to use, with a reportedly high level of reliability and stability independent of region and, to some extent, panel size. The neutral 'neither like or dislike' category is a valid response for some consumers, although some do show 'end use avoidance,' where the participants avoid using the extremes of the scale. This serves as a warning to those who may be tempted to truncate the scale further to seven or even five points. Hedonic measurements can also be achieved using scales such as smiley faces, but the pictures may be distracting to the participant, and the results often prove unreliable. Acceptability ratings may also be given on unstructured line scales, anchored by a like or dislike at each end, and magnitude estimation scales have also been proposed for the measurement of food acceptability. Variations on acceptability scaling include the just-right scale, which measures the desirability of a specific attribute, e.g., a scale may be anchored with 'not sweet enough' at the left, 'just right' in the center, and 'much too sweet' at the right. However, as consumers must have a consensus as to the understanding of the attribute in question, the use of this scale is often confined to a few simple attributes that are widely understood. The analysis of this scale can also pose some problems, and the assumption is made that the population is homogenous when in fact there may be two or more consumer segments that prefer different levels of the attribute(s) in question.

Preference Mapping: Internal

Internal preference mapping is based solely on the acceptability information gained from hedonic tests (commonly the nine-point hedonic scale) and provides a multidimensional representation of the samples and the position of consumers in relation to the samples. It provides an illustration of the consumer samples likes and dislikes, and the intensity of these, and additionally allows a visual segmentation of the samples likes and dislikes to be made. An internal preference map may be derived by multivariate statistical techniques such as principal-components analysis and is relatively straightforward to interpret. However, one potential limitation of this method is that it assumes that sensory perception is equal over all consumers and that differences in the rating assigned to products are due to varying preferences and not differences in perception. Figure 1 illustrates an internal preference map for eight cheeses; the further a consumer is from the center of the plot, the more discriminatory their evaluation was. From the plot, we can determine that

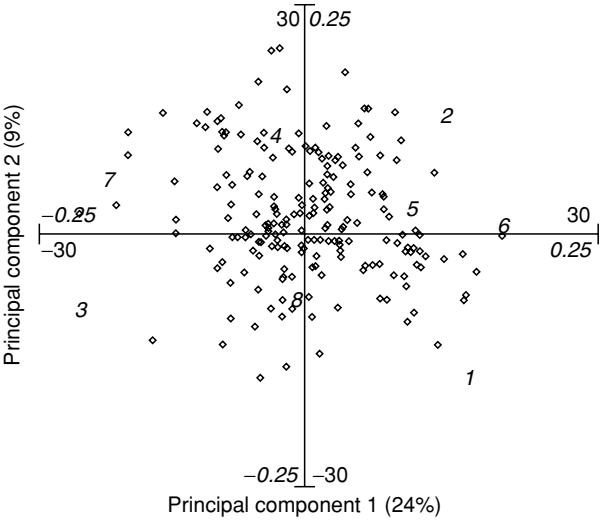

Figure 1 Internal preference map illustrating consumer preference for cheeses (consumers are represented by dots, and products are represented by numbers).

cheese 3 was the least liked, and cheese 8 was most liked.

Whilst an invaluable part of determining food sensory acceptability and preference, hedonic tests and their representation by internal preference mapping give no indication of why a product is liked and thus are sometimes combined with other sensory tests, such as a descriptive sensory analysis in a procedure known as *External Preference Mapping*, or analytical information, such as that obtained from gas chromatography–mass spectrometry experiments.

Preference Mapping: External

The technique of external preference mapping (PREFMAP) requires an objective characterization of products' sensory attributes, achieved by descriptive sensory analysis (conducted using qualified and trained panelists), which is then related to hedonic acceptability ratings for the product obtained from a representative sample of consumers. Descriptive sensory analysis has been recommended as the most appropriate sensory tool to provide the sensory attribute measures, because there is no prior knowledge concerning the most important sensory characteristics. Statistical modeling of the data, using techniques such as response surface analysis or partial least-squares regression, allows the complex nature of the relationships between consumer preference data and descriptive data to be established.

There are a number of models/approaches that can be used for external preference mapping. The most simple of these is the vectorial model, which operates on the assumption that 'more of attribute X is better'.

In practice, this model has obvious limitations as liking for a product attribute will increase until the point of maximum pleasantness (the 'bliss point'), beyond which an increasing intensity may begin to impact negatively on acceptability. The ideal point model, which is more complex, identifies the optimum combination of product attributes ('X and Y are ideal') for maximum liking and assumes that some products have too much of an attribute, and some products have too little. This technique is useful for market researchers who are involved in new product development and trying to identify gaps in the market place (i.e., the map can indicate consumer groups whose preferences are not currently met by any of the products on the market place). Less commonly used and more complex models used for external preference mapping include the elliptical ideal point model, and the elliptical ideal point model with rotation.

Consumers are inherently complex, and thus attempting to describe consumer behavior using a limited number of models is difficult, with a large number of consumers who are not explained by the resultant maps. To overcome this problem, the AUTOFIT model was developed, which selected the most acceptable model for each consumer, with some consumers fitted by the vectorial model, some by the ideal point model, and so on. However, this method results in 30–50% of consumers still not being explained in the analysis and has obvious limitations. The question of discarding consumers for the reason that they do not 'fit' the analysis has also been debated and should be considered carefully.

Preference mapping can also be combined with segmentation techniques, such as cluster analysis. These analyses are conducted prior to creating the external preference map, and the scores assigned by each of the consumer segments input into the analysis. This allows consumers with similar preference behavior to be identified. An example of an external preference map for Cheddar cheese, using clustering, is illustrated in Figure 2. In this trial, few consumers liked the light cheese, which was perceived as 'bitter' and 'rubbery.' Some consumers (cluster 4) liked premium and vegetarian cheese, whose preference appeared to be driven by a 'buttery' flavor and 'smooth' texture. The map also illustrates consumer needs not met by the products currently on the market. One group of consumers (cluster 5) would prefer sweeter and more caramel-like cheese.

In general, these statistical modeling techniques have brought a wealth of information about product liking; however, some practical considerations are worthy of mention here. At present, at least six products must be used in the external analysis, and all consumers must assess all products, as missing values

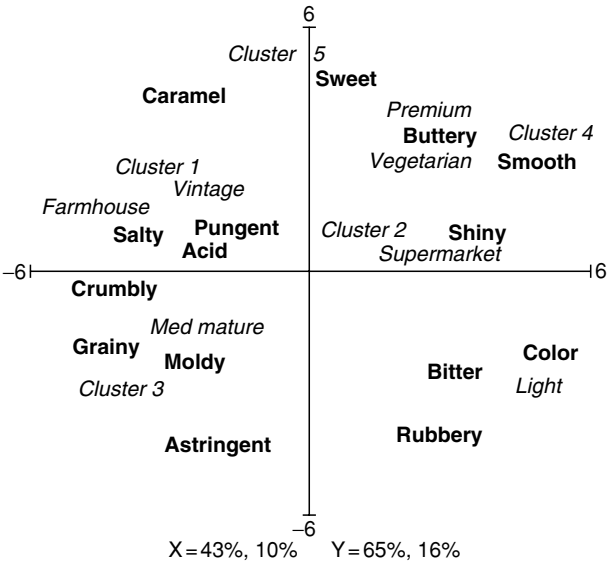

Figure 2 External preference map illustrating consumer clusters' preference for cheeses and their associated sensory attributes.

cannot be accounted for. The sample range must also be considered, and their properties in the context of each other evaluated. The reliability of the preference map result is dependent on the quality of the external (trained panel) data, and often it is necessary to weight all attributes equally (standardise), as some attributes that are highly discriminating (e.g., color) can skew the analysis. Whilst most commonly used to understand sensory preference, preference mapping can also be used to understand physical or chemical reasons for consumer preference. Finally, when choosing a modeling method, it is worthwhile to remember a commonly used rule of thumb: '80% of the modeling benefit comes from the first 20% of model complexity.'

Overall, preference mapping is increasingly being used within the food industry as a tool to reduce the risk of failure in new product development, increase liking, and secure market share for existing products.

Focus Groups

Focus groups are an increasingly common method for eliciting qualitative information about foods' sensory properties and how these may relate to food acceptability. The group should be a carefully planned session designed to obtain the maximum amount of relevant information on the product/attribute of interest. A moderator skilled in consumer research and the nuances of group dynamics should coordinate the effort to ensure that the result is a balanced and informative representation of all the participants' views. The information from focus groups is often

used as a starting point in sensory research and can be used as the basis for developing sensory questionnaires/tests in order to elicit quantitative acceptability information about products.

Repertory Grid Method

Individual or one-to-one interviews are an alternative method for eliciting qualitative information about products' sensory attributes and acceptability in situations where group dynamics, are perceived to be a problem. The repertory grid method (RGM), which has its origins in the field of psychology, is now commonly used to collect qualitative sensory data on product attributes and acceptability. The interviewer should be trained and familiar with the method in order to ensure that the maximum amount of product information is gained. Products are presented to the interviewee in triads (groups of three), and the interviewee is subsequently asked to describe the similarities and differences between the products in terms of their sensory properties. When no more attributes can be generated, the next triad of products is evaluated. The proposed advantage of RGM is that the continually changing frame of reference to which the participant is exposed allows a greater insight into the details and intricacies of the product attributes than simply asking the participant to 'describe the sensory properties of this product.' The qualitative information gained from this procedure is often used as the basis for quantitative research, for example in free-choice profiling studies (discussed below) or for further sensory questionnaires/tests.

Free-choice Profiling

Free-choice profiling (FCP) is a variation of descriptive sensory analysis and can also be combined with acceptability information to ultimately determine foods sensory acceptability. FCP allows consumers to generate their own terms to describe product perceptions, as in the case of descriptive sensory analysis and PREFMAP; the information can then be related to hedonic measurements, thus allowing an understanding of the sensory attributes driving acceptability.

Generally, for the attribute elicitation stage, a method such as RGM is used to maximize the amount of information gained. Once the vocabulary of attributes to use has been decided upon, participants score products for the attributes on line scales. In addition, consumers rate the acceptability of the product. The attribute data are then analyzed by a procedure known as generalized procrustes analysis (GPA). Through stages of translation, scaling, and rotation/reflection, the GPA adjusts for several sources of variation, such as assessors using different levels and ranges of the measurement scales. The resulting plots demonstrate consumer perceptions of products, which are then further analyzed in conjunction with the hedonic data in order to determine reasons for consumer preference. Figure 3 illustrates children's preference for the sensory attributes of vegetables. Children used FCP and generated individual vocabularies to describe vegetables and rated these attributes on line scales. In addition, they were also asked to rate the acceptability of the vegetables. Sensory attribute data were then analyzed by GPA, and the subsequently identified consensus attributes were then correlated with the first two preference dimensions of an internal preference map (produced using the acceptability data). Vegetables perceived as 'big' and 'hard,' such as cauliflower were least liked whilst those perceived as 'soft' and 'small' were most liked.

The advantage of using FCP to determine acceptability is that the descriptive sensory measurements (conducted using highly qualified panellists who can be expensive to maintain) are not required. Participants need only to be able to use a scale and be consumers of the product under evaluation. It also provides a more 'consumer orientated' opinion on the products. However, sometimes, the handling of individual ballots for each participant in an FCP study can prove time-consuming, and the interpretation of the results by the sensory analyst can also be challenging. Although, ultimately, the attribute information is not as discriminatory as that obtained from a trained panel, FCP measurements when related to hedonic information in this way provide valuable acceptability information about product sensory attributes.

Future Role of Sensory Evaluation in Determining Food Acceptability

The area of sensory evaluation and its role in determining food acceptability have increased rapidly and will continue to do so for many years to come. The use of validated statistical models will increasingly allow the results of sensory tests (such as FCP, descriptive analysis, time–intensity flavor evaluations) to be related to hedonic measurements of acceptability, uncovering not just the 'what is liked' but the 'why.' Reverse control modeling, which allows a feedback loop to be established between chemical/physical, sensory, and acceptability data will be used increasingly to understand product acceptability. The integration of the sensory analyst into both the technical and marketing departments of organizations will extend further the boundaries of the sensory science discipline. Already, we have research that

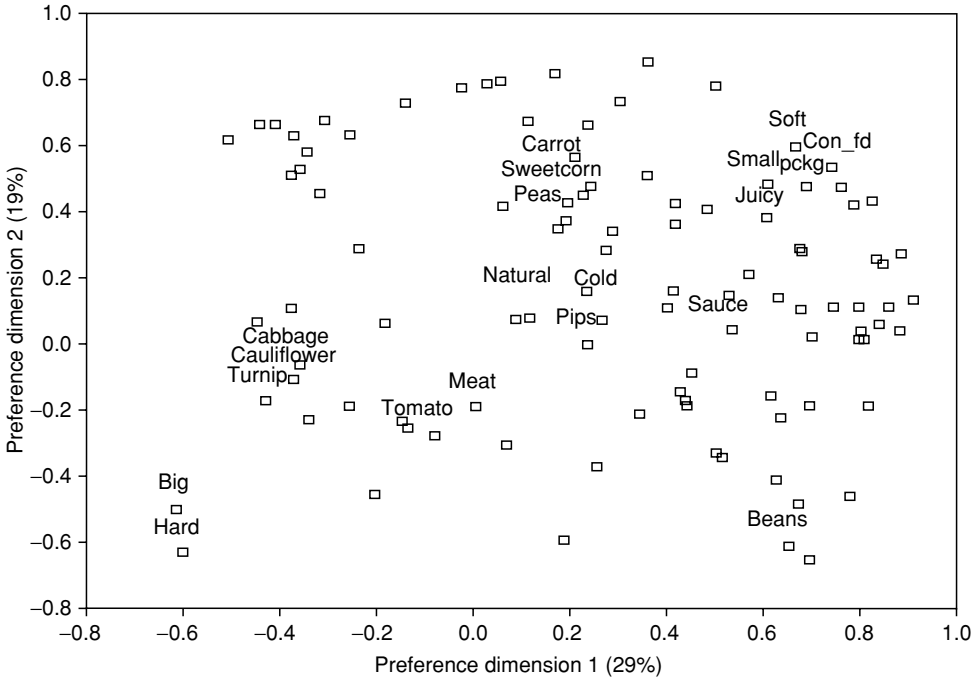

Figure 3 Plot illustrating children's preference for the sensory attributes of vegetables, obtained through correlation of consensus attributes (identified by generalized procrustes analysis) with the first two preference dimensions of an internal preference map (consumers are identified by squares).

illustrates how generated product concepts 'fit' the sensory characteristics of products, and how these may be matched better. Such techniques will be used increasingly in order to minimize risk in an increasingly competitive and global food industry.

See also: **Sensory Evaluation**: Sensory Characteristics of Human Foods; Appearance; Texture; Aroma; Taste

Further Reading

Amerine MA, Pangborn RM and Roessler EB (1965) *Principals of Sensory Evaluation of Food*. New York: Academic Press.

Lawless HT and Heymann H (1998) *Sensory Evaluation of Food: Practices and Principals*. New York: Chapman & Hall.

Lawless HT and Klein BP (eds) (1984) *Sensory Science: Theory and Applications in Foods*. New York: Marcel Dekker.

MacFie HJH and Thompson D (eds) (1986) *Measurement of Food Preferences*. London: Blackie Academic & Professional.

Meilgaard MC, Civille GV and Carr BT (1999) *Sensory Evaluation Techniques*, 3rd edn. Boca Raton, NY: CRC Press.

Meiselman HL and MacFie HJH (eds) (1996) *Food Choice, Acceptance and Consumption*. London: Blackie Academic & Professional.

Moskowitz HR (1983) *Product Testing and Sensory Evaluation of Foods: Marketing and R & D Approaches*. Westport, CT: Food and Nutrition Press.

McBride RL and MacFie HJH (eds) (1990) *Psychological Basis of Sensory Evaluation*. London: Elsevier Applied Science.

Naes T and Risvik E (eds) (1996) *Multivariate Analysis of Data in Sensory Science*. Amsterdam: Elsevier Science.

Piggott JR (ed.) (1981) *Sensory Analysis of Foods*. London: Elsevier Applied Science.

Piggott JR (ed.) (1986) *Statistical Procedures in Food Research*. London: Elsevier Applied Science.

Stone S and Sidel JL (1993) *Sensory Evaluation Practices*, 2nd edn. London: Academic Press.

Thompson DHM (ed.) (1988) *Food Acceptability*. London: Elsevier Applied Science.

Practical Considerations

E Chambers and B McGuire, Kansas State University, Manhattan, KS, USA

Background

Many of the most common applications of sensory analysis are in the consumer goods industries. For those purposes, the basic aim of sensory analysis is

to provide information about how the sensory characteristics of products, ingredients, or other related information and services are related to perceived quality characteristics and consumer liking. How do you make a peanut butter 'creamier' or a facial tissue 'softer'? How do you describe the fragrance of a fine perfume? How easy is it to push the pump on a spray deodorant? When does a potato chip start to taste 'stale'? These are just some examples of what sensory analysis can do.

Setting Objectives

Given numerous applications of sensory analysis, the practical considerations associated with effective testing must begin with the objectives. In short, what information is needed, or 'What do you want to know?'

Consider the following hypothetical situation. You own an icecream company. Your supplier of vanilla tells you to expect a substantial increase in price. You do not want to make consumers pay more for your icecream, so you find a different provider at a lower price. Your biggest concern is providing the same great flavor to your consumers. Therefore, an initial question is: 'If the new vanilla is used, can people distinguish between the products?' or 'If I use the new vanilla flavor, do I still have a great product?'

A key practical consideration is to realize that those two questions are fundamentally different. The first question is an issue of difference; the second is an issue of liking or preference. If you decide that people should not be able to detect any difference in the product, you have set a very different criteria than if you determine that the product can taste different as long as consumers like the new product as much as, or more than, current product.

Part of setting the objectives in a sensory test is risk management. Every person who will buy the product will not be asked to test the product (see section Resource Management); nor will it be tested in every situation in which people eat icecream. You have to identify an objective and select a test protocol that addresses how much risk you are willing to accept. If you want to be especially sure that a consumer will not find any differences in the icecream, you need to specify that in the objective and use a large population of people who are able to discriminate differences in icecream. If you just want to be pretty sure that you still have a good icecream that will be acceptable to the average user, specify this objective and test a population of typical users to make sure that the icecream still scores high in acceptance.

Deciding on the objective(s) of the sensory test *and* getting everyone to agree on the objective(s) is the first practical consideration in sensory testing. If this is done poorly, the rest of the test may be meaningless. If the objectives are set correctly, and everyone has agreed in advance, the test and the reporting of results will be much easier.

Resource Management

The next practical consideration in sensory testing is managing the test to provide the key information in a cost-effective, timely way. Although it is possible to conduct sensory studies in many ways, the best ways will be those that provide appropriate answers to key questions as quickly as possible using the minimum amount of resources. It is very easy to overplan or overconduct studies. Commonly, tests are conducted that either are not required or that try to answer too many questions simultaneously.

If we go back to the icecream company, an initial question might be, 'Is there an obvious difference if we change to the new vanilla?' If there is, then difference testing is a waste of time. Initial benchtop screening can help to answer those questions. Although lab personnel and other employees are not representative of the typical user population, they can be a cost-effective way to initially screen products to determine whether further testing is needed.

If the company decides, after benchtop screening, that testing is needed, an appropriate test is required. While it might be nice and interesting to find every consumer of our icecream and ask them to taste the two samples, such an enormous study would make little sense from the perspective of cost or time and would provide no information that we cannot gain from a smaller more focused study. It would be a huge waste of resources to ignore the practical issues of resource management.

Sampling

Another key practical consideration is sampling of the population. It is important that participants in sensory test are the people we really need to answer the questions. Caution must be exercised that the testers do not have biases that would affect their testing. Thus, the first step in obtaining a group of people who are representative of the product's consumers is acknowledging that employees of the company often do not meet these criteria. The potential biases, both good and bad, that employees may bring about their companies products often disqualify them from making a real assessment of the quality of the product. Although employees provide a cheap and

convenient source of people to test, and may be suitable for a screening test, their views should not be used as a final decision-maker when deciding what consumers really think about the product(s) in question.

For descriptive sensory tests, the appropriate people are those who have a good ability to sense attributes and have been trained in some way to describe what they are testing. Panelists trained to have a common vocabulary make the task of translating sensory perceptions into meaningful information for product scientists much easier.

Consumers usually are screened to help assess their potential to represent either potential users or current users of the product. Screening questions may include: age, gender, product related behaviors (e.g., use in the past 3 months), or attitudinal and other factors that help to identify people who are likely to use, or do use, the product or other products in the category.

There is no 'magic' minimum number of consumers required for any specific consumer test. The number of respondents should be large enough to be representative of the consumer population and to provide the statistical power needed to differentiate the products. This is a fine balance. The tendency is to test too few people 'to control costs' or to test too many people 'just to be sure.' Historical information on the consumer population, often obtained from marketing or purchase data, and information on statistical variability can help to determine the number of consumers for the test that balances resources and risk management.

The second major part of sampling is the product. The range of products or samples tested can have an impact on the test. For example, testing a high-fat vanilla icecream with a reduced calorie vanilla icecream will result in different sensory scores than testing the high-fat product against other similar high-fat products. In the end, the results obtained from testing are dependent on the samples evaluated.

The second major concern in product sampling is making sure that the product being tested represents the product that actually is manufactured. Too often, incorrect conclusions are drawn from a test because 'early prototypes' (that is, samples that are different from the actual final product) are used in the test, for example, testing a product from one manufacturing plant when three plants product the product using different equipment and ingredients suppliers, or when products have been mishandled. In those and similar cases, the products in the test are not representative of the product that consumers actually see when they 'purchase off the shelf.' If the product does not really represent 'the product,' the test results are meaningless.

Issues associated with sampling are critical practical considerations in sensory testing. Ensuring that the right product is tested and that the correct population of people tests the product are essential to providing actionable data.

Analysis and Reporting

Time and resources are limited, finite properties. Within this context, the use of appropriate analysis and reporting is the essence of practical considerations for sensory analysis. Proper analysis and appropriate reporting of well-conducted studies provide timely information that is actionable and concise.

The use of appropriate data analysis is an extension of the first practical consideration – clear objectives. 'How is this analysis related to the specific objectives?' is a question that should precede all statistical applications. One advantage of basing analysis and reporting on specific objectives is that it refines the scope of analysis that is appropriate. For example, analysis of results from common difference tests may require only the use of standard tables, whereas analysis of results for data that attempt to 'explain' liking data on the basis of descriptive sensory properties is more complex. Appropriate analysis concludes with obtaining all findings that are directly relevant to the stated objectives.

At the end of almost every study, people want three basic questions answered: 'What did we learn?', 'Why is this information important?', and 'How can we use this information?' In reality, after testing is complete, and the data have been analyzed, some findings are not as important as originally thought, some will require further investigation, and some may provide several possible courses of action. Within this context, quality reporting, whether written or oral, provides specific, concise answers with supporting details when needed. Reporting that leads the audience to wonder 'so what?' is misguided and will be ineffective. Further, quality reporting does not mean that everyone is happy with the answers: it means that everyone is clear on what the answers are, the relative confidence in these answers, and how they are actionable.

The basic purpose of a report is to communicate information related to specific goals. Thus, clarity and brevity are key considerations in appropriate reporting. If the relevance of the information presented cannot be clearly understood by the reader or listener, the value of the work will not be maximized. Further, acknowledgement that people have a finite attention span is key to effective communication, especially for complex sensory information. This emphasizes the importance of stating key findings early

so that the audience can focus on the meaning of the information presented.

Sensory analysis can answer a wide range of different questions provided that practical considerations such as clear objectives, resource management, appropriate sampling, and efficient analysis and reporting are used effectively. Without those key considerations, any sensory testing is likely to be misguided and wasteful. The following examples have been provided to illustrate how these common themes apply to the different areas of sensory analysis.

Do You Want to Know if Consumers can Distinguish Products?

Consider the following hypothetical situation. You own an icecream company. Your provider of sweetener tells you to expect a substantial increase in price. You do not want to make consumers pay more for your icecream, so you find a different provider at a lower price. Your biggest concern is providing the same great flavor to your consumers. Therefore, an initial question is: 'If the new sweetener is used, can people distinguish between the products?'

There are numerous tests that can be used for this purpose, collectively referred to as discrimination or difference tests. Difference tests do not answer the question 'How much consumers like a product?' or 'How are these products different?' Difference tests answer one simple question, 'Can people distinguish samples?'

Difference tests often test 30 or more known discriminators in order to determine whether differences that are found can be attributed to random variability or are product-based. Both the analysis and reporting for these tests are straightforward. Simple, test specific tables (often based on binomial distributions) are available to determine the likelihood that the difference between the samples is product-based. A concise way to report these findings is to state the percentage of people who could distinguish the products, followed by the likelihood (a statistical p-value) that this difference can be attributed to random variability. For example, results might be summarized as 'In the paired difference test, 51% (24 out of 47) selected the different sample, which is less than the number required to find a significant ($p = 0.05$) difference in the products. This suggests that consumers will not perceive icecream with the new sweetener as different.'

Are You Interested in Exploring Consumer's perceptions?

Again, let us imagine that you own the icecream company, but now you are interested in understanding more about how your product is packaged. Your main competitor uses different colors and a different shape, and some consumers have noted that their package is more 'exciting' and 'appealing.' You are interested in knowing the specific characteristics of the package that are related to how 'eye-catching' the product is. What are the key factors – the color, the material, the shape, the advertising on the product?

Focus groups or one-on-one interviews are effective for initially exploring consumers' perceptions. This research is described as qualitative because the objective is to answer the questions, 'What is important to consumers and why?', instead of quantifying 'how much.' Qualitative consumer research allows for in-depth discussion of consumers' attitudes and perceptions.

As with any technique, the first step is using qualitative methods to define the specific objective. The fundamental challenge of qualitative research is to obtain relevant information from the consumer. Given this, it is essential to develop a guide of how the conversation (discussion) with the consumer will proceed that will lead them to provide answers to issues but will not bias their response. A well-prepared guide is an important asset in managing resources and obtaining the desired information and is more than just an outline. It can provide the interviewer or moderator with transitional statements between topics, anticipate potential problem areas, and provide ideas for 'follow-up' questions.

In qualitative consumer research, acknowledging possible factors that may bias the sampling is important. Because the basic purpose of qualitative research is to better understand consumers ideas, any influence that causes an inconsistency between what consumers communicate and what they actually believe may lead to inaccurate conclusions being made. There are many potential sources of bias, including: the facility used, the attitude of the moderator, how questions are asked, and the interaction among participants. Effective focus groups, for example, generally involve a comfortable, distraction-free room, with the discussion led by a moderator, who promotes an environment in which all participants feel comfortable expressing and discussing their ideas and opinions.

Because of the nature of qualitative research, analysis and reporting of findings are different than with quantitative data. An effective way to analyze qualitative data is to begin by reviewing all discussion and finding 'common themes' that are related directly to the objectives. For each theme, specific comments from as many people as possible that illustrate these

themes are noted. For reporting, the use of bulleted points outlining the themes with some illustrative quotes can be a concise, effective way to present findings.

Do You Want to Estimate How Much Consumers Like Products?

Now, consider the situation: same icecream company, only now you would like to sell more icecream. You decide that a 'fruity raspberry' icecream is the best new product because consumers like the flavor, and your main competitor is having success with a similar product. In short, you want a 'fruity raspberry' icecream that consumers like more than your main competitor's. You produce a small amount of a new 'fruity raspberry' icecream, but how do you determine how much consumers like it?

This objective requires a quantitative consumer acceptance test. Consumer acceptance testing is a generic phrase that describes a group of different methods for estimating how much consumers like products. In acceptance tests, consumers generally evaluate multiple samples and are asked questions including: overall liking and liking of basic product characteristics (for example, sweetness, creaminess, or raspberry flavor).

One important key to using acceptance tests is to ask questions that are directly related to the specific objectives of the test and to include the products that are necessary. Because testing can be expensive, clarity of objectives is important because cost constraints likely will limit the number of consumers who can be tested or the number of questions that can be asked. For example, in addition to overall liking, is it important to obtain feedback on the overall flavor, specific flavors, color, texture, aroma, meltdown? Since one of the goals of this example is to determine whether the product is liked more than the competitor, samples that represent the competitors' product(s) are a prerequisite for testing.

Keep the questions simple. Consumers generally do not have the level of interest or attention that people who work with the product every day have. For these reasons, their willingness to think about confusing issues, or respond to complex questions, may be limited. If consumers do not understand the question, or interpret a question in different ways, the results are unlikely to be an accurate measure of the characteristics that the questions were intended to answer.

It is crucial that the participants in an acceptance test are representative of the larger consumer population. One of the best ways to ensure this is to test only people who have been identified as potential consumers of the products. In general, a consumer acceptance test will collect responses from at least 100 consumers, but the actual number is dependent largely on the decision being considered and may range into the hundreds or, on rare occasions, into the thousands.

Although the analysis of acceptance tests is more complex than difference testing, the emphasis on addressing the specific objectives and brevity is no less important. In this case, reporting the mean liking scores and comparing the company's new product with that of the competition is usually the main area of emphasis. If the new product does not score highly, a review of the other consumer data is essential to understand why.

Do You Want to Describe and Measure the Sensory Properties of a Product?

Let us use the icecream company one last time. You are receiving complaints that your vanilla icecream has a 'funny,' 'off,' or 'bad' flavor. You attempt to have consumers provide a more specific description of the flavor, but they are unable to do this. That is not surprising; consumers are great at communicating what they like or dislike, but they are not as good at describing why. Descriptive sensory analysis, using highly trained assessors, is more appropriate for this task because those assessors are able to reduce complex perceptions that often involve multiple senses into simpler senses that can be defined and measured.

In our example, you have determined that the icecream is close to, but not past, the 'best if used by' date. You are concerned that the shelf-life of the product may be too long, but you need a better description of the flavor and its intensity during storage.

Product orientation and the use of appropriate methods are important resource-management tools in descriptive sensory testing. Allowing the panelists to work briefly with products similar to what they will test allows the panelists to become oriented with the product and helps you and the panelists decide on the best way to evaluate the products. Specific methods and scales that are used for descriptive sensory analysis are described extensively in the literature and are dependent on the product being tested and the objectives of the study.

Appropriate analysis and reporting for descriptive sensory analysis again are dependent on the objectives. The objective of the sample study would be to provide information on how the sensory properties of the products are changing as storage time increases. Thus, a chart that tracks changes in specific attributes over time is a simple, effective method for evaluating the changes.

Conclusion

Sensory analysis integrates many different sciences to understand better the sensory properties of products and consumer response to these properties. Because sensory analysis has a broad range of applications, start with the basic questions: 'What do you want to know?', 'What resources are available?', 'Who needs to conduct the test?', 'What products will you use?', and finally, 'What answers do you need?' Answering these questions will lead you to a consideration of the practical aspects of conducting any sensory study.

See also: **Sensory Evaluation**: Food Acceptability and Sensory Evaluation; Sensory Difference Testing; Sensory Rating and Scoring Methods

Further Reading

Amerine MA, Pangborn RM and Roessler EB (1965) *Principles of Sensory Evaluation of Food*. New York: Academic Press.

Chambers E IV and Wolf MB (eds) (1996) *Sensory Testing Methods*, 2nd edn. West Conshohocken, PA: ASTM.

Hootman RC (1992) *Manual on Descriptive Analysis Testing for Sensory Evaluation*. Philadelphia, PA: ASTM.

Jellinek G (1985) *Sensory Evaluation of Foods Theory and Practice*. Deerfield Beach, FL: Ellis Horwood.

Lawless HT and Klein BP (1991) *Sensory Science Theory and Applications in Foods*. New York: Marcel Dekker.

Lyon DH, Francombe MA, Hasdell TA and Lawson K (1992) *Guidelines for Sensory Analysis in Food Product Development and Quality Control*. London: Chapman & Hall.

Meilgaard M, Civille GV and Carr BT (1999) *Sensory Evaluation Techniques*, 3rd edn. Boca Raton, FL: CRC Press.

Moskowitz HR (ed.) (1988) *Applied Sensory Analysis of Foods*, vol I and II. Boca Raton, FL: CRC Press.

Moskowitz HR (1983) *Product Testing and Sensory Evaluation of Foods Marketing and R&D Approaches*. Westport, CT: Food & Nutrition Press.

Munoz AM, Civille GV and Carr BT (1992) *Sensory Evaluation in Quality Control*. New York: Van Nostrand Reinhold.

Piggott JR (1988) *Sensory Analysis of Foods*, 2nd edn. New York: Elsevier Science.

Poste LM, Mackie DA, Butler G and Larmond E (1991) *Laboratory Methods for Sensory Analysis of Food*. Ottawa: Canada Communication Group-Publishing Centre.

Stone H and Sidel JL (1993) *Sensory Evaluation Practices*, 2nd edn. San Diego, CA: Academic Press.

Watts BM, Ylimaki GL, Jeffery LE and Elias LG (1989) *Basic Sensory Methods for Food Evaluation*. Ottawa: The International Research Centre.

Yantis JE (ed.) (1992) *The Role of Sensory Analysis in Quality Control, Manual 14*. Philadelphia, PA: American Society for Testing and Materials.

Sensory Difference Testing

B Rousseau, The Institute for Perception, Davis, CA, USA

Introduction

This article describes our current knowledge on the topic of sensory discrimination testing. The general objectives of the commonly called 'difference tests' are first emphasized, the main objective being the scaling of sensory magnitudes. Then, the traditionally encountered discrimination protocols are described in terms of the nature of the instructions given to the subject. The common data analysis based on binomial tables and its limitations are mentioned. An approach based on an index of perceptual differences, called d' and pronounced 'd prime,' is introduced. It is shown how the results using this index are consistent across methods. A comparison of the power of the traditional discrimination tests is presented, showing which tests are less likely to fail to detect nonnegligible sensory differences between food products.

The field of sensory evaluation has been commonly divided into three categories: sensory discrimination testing, descriptive analysis, and hedonic studies. The sensory discrimination tests are traditionally used to investigate whether two food samples can be perceived as significantly different by a group of panelists. Descriptive analysis will provide qualitative information regarding the difference among food samples, by focusing on particular sensory attributes inherent to the products under study. Hedonic testing will be used in order to study the acceptance and preference of food samples by consumers.

Discrimination testing is sometimes considered as a straightforward data collection method, with results providing useful but limited information. One will use a difference test in order to determine whether the sensory characteristics of two samples, for instance a standard product and its reformulation, can be distinguished. Assuming that a proper experimental design is used, the similarity or difference of the two products will be concluded. However, this reasoning can appear unsatisfactory. It could be easily argued that two products with different physical and chemical characteristic will never be perceived as identical. If a large enough sample size is used, any sensory difference will be detected.

This brings us to the notion which will be used in the present article. Instead of merely answering the question: 'Are my products the same or different?,' we will show how discrimination test results can be

used as a tool to answer the following question: 'How large is the sensory difference between my products?' These results will permit the scaling of sensory magnitudes and sensory differences, information thought until now to be provided only through the use of rating methods.

Traditional Protocols Used in Discrimination Testing

There is no limitation on the number of different discrimination tests that can be used to study sensory differences. The list which will be given here is not exhaustive, and alternative discrimination methods can easily be created. The issue is then to develop the proper model to conduct statistical analyses.

The discrimination protocols can be classified into two main categories: those that require the nature of the sensory difference to be specified in the instructions and those that do not.

The protocols described here will be illustrated by the comparison of two cookies *A* and *B*, *B* being slightly sweeter than *A*. The samples are presented blind, and no other difference (visual, textural, temperature) is assumed to be present. A subject will perform several tests in succession, or a group of subjects will perform one test each. From the results of the test, the similarity of the difference between the two samples will be concluded with a given level of confidence, depending on the sample size (number of tests) involved in the comparison.

Directional Difference Tests

These methods require the nature of the difference to be specified in the instructions. Table 1 gives details for each method.

The chance probability represents the probability of getting the answer right if the subject cannot discriminate between the products. For the sorting protocols, it is important to keep in mind the fatiguing nature of food samples and the memory requirements that might significantly hinder subjects' performance. These tasks are best suited to textural and visual investigations.

Nondirectional Difference Tests

Those are the most commonly used methods in sensory discrimination testing, even if their statistical power can be up to 100 times inferior to that of the directional difference tests (Table 2).

Traditional Data Analyses

The traditional way of analyzing data extracted from difference tests involves no consideration of the psychological processes occurring at the time a response is generated. This is what is called a response-based analysis. The information obtained from this rather limited statistical analysis is either 'Yes, the products are different,' or 'No, they are not different.' No information is obtained regarding the size of the sensory difference. This can yield inconsistent results and this issue will be discussed later.

Table 1 Directional difference tests

Test	*Number of samples*	*Instructions*	*Chance probability*
2-AFC	Pairs of samples: *A*: 1, *B*: 1	 Find the sweeter cookie	1/2
3-AFC	Triads of samples: *A*: 2, *B*: 1 or *A*: 1, *B*: 2	 Find the sweetest cookie or  Find the least sweet cookie	1/3
A/Not *A*	Single sample: *A* or *B*	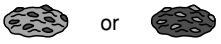 Is it *A* or is it Not *A* (*B*)?	1/2
Sorting	*n* samples: *A*: $n-p$, *B*: p	Find the p sweetest samples	$\frac{p!(n-p)!}{n!}$

AFC; alternative forced choice.

Table 2 Nondirectional difference tests

Test	*Number of samples*	*Instructions*	*Chance probability*
Triangle	Triads of samples: *A*: 2, *B*: 1 or *A*: 1, *B*: 2	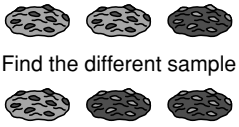 Find the different sample	1/3
Duo–trio	Triads of samples: Reference: *A* and *A*: 1, *B*: 1 or Reference: *B* and *A*: 1, *B*: 1	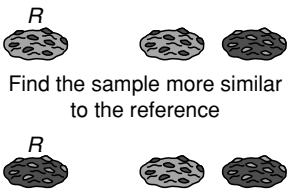 Find the sample more similar to the reference	1/2
Same–different	Pairs of sample: *AA*, *BB*, or *AB*	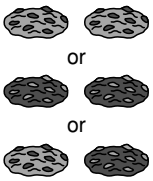 Are the samples the same or different?	1/2
p-out-of-*n*	*n* samples: *A*: $n-p$, *B*: p	Make one group of $n-p$ similar samples and one group of p similar samples	If n is odd: $\frac{p!\,(n-p)!}{n!}$ If n is even: $2\times\frac{p!\,(n-p)!}{n!}$

In discrimination testing, one is concerned with two types of error. The first one is the type I error: concluding that the products were significantly different while they were not. The second is the type II error: failing to detect an existing sensory difference. Statistical tests are conducted at a particular α level, usually 0.05, describing the type I error. An α level of 0.05 means that we have only a 5% chance of concluding that the products were different while they were not. For the type II error, described by the parameter β, the level used is usually 0.2 or 0.1. The statistical power of the test is defined by: Power $=1-\beta$. A power of 80% implies that a test has 80% chance of detecting a difference of a given size. The issue of power will be discussed later.

Binomial Analysis

For a given protocol, the number of correct tests is compared to the number that would be expected by chance. This type of model is called a guessing model. It assumes that if the subject cannot discriminate between the samples, he or she will pick a sample or an answer randomly. We are testing the H_0 hypothesis that the two products are identical. If rejected, we will accept the H_1 hypothesis that the products are different. For instance, if a study with 20 tests was run using the two-alternative forced choice (2-AFC) protocol and 14 were performed correctly, this result would be compared to an expected chance result of 10. Tables have been published indicating the minimum number of tests correct for a given sample size to be significant at a given α level, usually 0.05. In our example, to be significant at the 5% level, we would need 15 tests correct. Therefore, we cannot conclude that the two cookies were significantly different. We cannot reject H_0.

It is critical to be aware that we are not concluding that the cookies are the same. We can only conclude that they are not different. Had we used a larger sample size, we might have obtained a significant result. In order to accept H_0, it is necessary to consider the power of the test.

χ^2 Analysis

This analysis is traditionally used with tests yielding results in a contingency table, such as the *A*/Not *A* and same–different tests. An example of a contingency table is shown in Table 3. A study between

Table 3 Results of a same–different test conducted with 100 consumers

Presentation	*Response*	
	'Different'	*'Same'*
Same	20	30
Different	35	15

the two cookies *A* and *B* was conducted with 100 consumers, each getting either a different pair (*AB* or *BA*) or an identical pair (*AA* or *BB*).

A χ^2 value can be calculated and we find that $\chi^2 = 9.1$. Looking in a table for the significance of χ^2, we find that a χ^2 of 9.1 and 1 degree of freedom is significant at the 1% level.

Therefore, here we can reject the null hypothesis H_0 and conclude that the products are different.

Taking into Account the Degree of Difference Between the Products: Thurstonian Models

The data analyses discussed above only provide one type of information from discrimination testing: 'Are my products significantly different or not?' We never mentioned the magnitude of the difference between the products. Was it a large difference or a small difference? How does it compare to results from previous studies? In order to answer these questions, it is necessary to introduce new notions, such as an index of perceptual difference, called d'. We are first going to show why it is essential to consider this index when comparing products.

Insufficiency of the Guessing Model: The Paradox of Discriminatory Nondiscriminators

It has been reported numerous times in the literature that, while using the same subjects and the same products, the null hypothesis could be rejected using some protocols but not others. The most illustrious example is the result discrepancy between the 3-AFC and triangle tests. While the two tests have the same design and only very slightly different instructions, subjects' performance is significantly greater in the 3-AFC than in the triangle test. With a 3-AFC test, it might be concluded that the products are significantly different, while we would not be able to reach that conclusion with the triangle test. Which result is right? Or are the two methods giving two versions of the same information, i.e., the magnitude of the difference between the two products? With the triangle test, nonnegligible sensory differences might not be detected because of insufficient statistical power. This might result in the release of reformulated products in the marketplace that will be rejected by consumers.

These experimental results underscore the insufficiency of relying on the number of correct answers in a given test to determine how different two products are. It is essential to take into consideration two very important aspects of sensory evaluation: variability and behavior. These aspects are taken into account in Thurstonian modeling.

Thurstonian Models

These models are very similar to those encountered in the field of psychology and discussed in signal detection theory.

Thurstonian modeling assumptions

Variability Let's take our sweet cookie example again. The idea behind perception variability is that, when tasting the same cookie several times, the perceived sweetness will not always be the same. On average it will have a certain intensity, but at any given moment, the perceived intensity can be slightly stronger or slightly weaker (Figure 1). Several reasons can explain these variations: first, there is always some random noise in the nervous system due to the spontaneous firing of nerves. Second, the number of compounds binding with the sweet receptors can be slightly different each time. Third, there might be variation in the cookie itself: the sweet compounds might not be evenly distributed on the cookie, or the subject might take bites of slightly different sizes every time.

The likelihood of each sweetness magnitude occurring for the cookie can be represented by a distribution. This distribution is usually assumed to be normal, even though alternative models have been published using other types of distributions. Figure 2 illustrates this concept. The height of the distribution at each intensity level represents the likelihood of this intensity occurring in a given tasting. The intensity at the mean of the distribution is the most likely. The further away the intensity is from the mean, higher or lower, the less likely it is to occur.

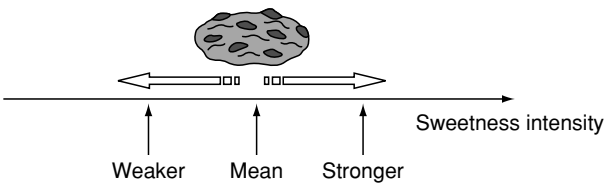

Figure 1 Representation of variable sweetness intensities for a cookie.

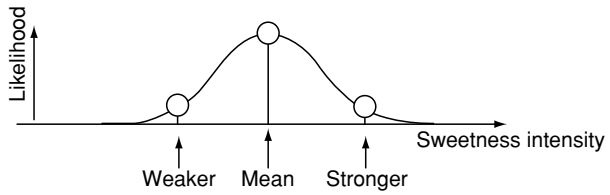

Figure 2 Normal representation of the likelihood of sweetness perception.

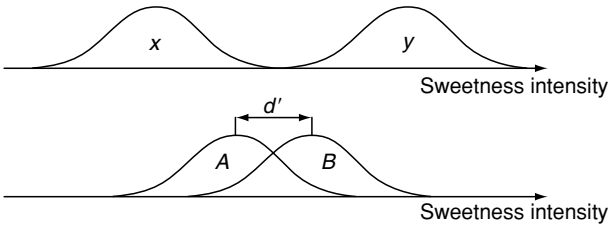

Figure 3 Thurstonian representation of two products differing in sweetness.

When performing a discrimination test, the two products to be compared can be represented by two normal distributions. This is shown in Figure 3.

In the top picture, product *Y* is always sweeter than product *X*. In this situation test, a subject will always get 100% tests correct in a difference test. Discrimination testing is not even necessary in this case; *X* and *Y* are totally discriminable. In the bottom picture, there is some overlap between the two distributions. Overall, *B* is sweeter than *A* and will be perceived as such most of the time, but on some occasions, *B* will be perceived as less sweet than *A*. This will happen when the *B* intensity is on the lower part of its distribution and the *A* intensity is on the higher part of the distribution. The measure of the extent of the overlap is called d' and is the distance between the means of the two distributions measured in terms of their standard deviation. A d' of 1 means that the two distributions are 1 standard deviation apart. The bigger the d', the more different the products. A d' of 0 means that the products are undistinguishable. N.B.: a d' of 1 corresponds to 76% tests correct in a 2-AFC test. In psychophysics, this would correspond to the definition of a detection threshold.

We now have an index measuring the magnitude of a difference, something we did not have when using the guessing model. Let's see now how we can obtain a d' value when using discrimination tests. For this, we need to take into account the behavioral aspect of discrimination testing, namely the decision rule, or cognitive strategy, used by the subject when performing a discrimination test.

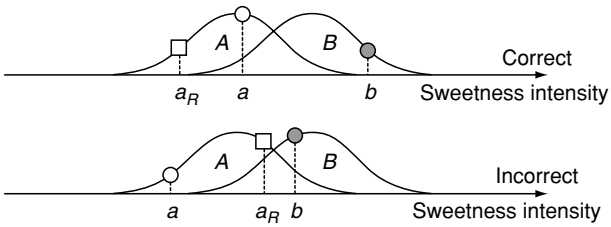

Figure 4 Correct and incorrect answers in a duo–trio test. Square, reference; circle, alternative sample.

Behavior The guessing model assumes that subjects make mistakes when they cannot tell the difference between the products and make a wrong guess about the answer. We will see that this is not how incorrect answers arise.

Duo–trio example Let's imagine that a subject performs a duo–trio test between two cookies *A* and *B*, the less sweet cookie, *A*, being the reference. A trial is described in Figure 4.

In the top picture, the subject tasted the reference and perceived a sweetness on the left-hand side of the distribution (a_R). He then tasted the two alternative samples and perceived the *a* and *b* sensations. When asked which of the two is more similar to the reference, he will choose *a*, since it is closer to a_R. His answer will be correct.

In the bottom picture, we can see that the sensations perceived will yield an incorrect answer. This time, *b* is closer to a_R than *a* is, thus the subject will pick *b* as more similar and his answer will be wrong, even if he gave the right answer based on the sensations perceived. We see that the error was not made because the subject guessed, but because the perceived sensations happened to be in such locations that a wrong answer was generated.

Over repeated trials, the top situation will occur more often than the bottom one, resulting in the measure of a d' value and the conclusion that the products are different.

The d' values can easily be calculated from discrimination test results. Tables have been published that provide the correspondence between d' and proportion of correct answers for the most traditionally used protocols.

Main cognitive strategies For each discrimination protocol, a subject will use a specific cognitive strategy. The two main cognitive strategies are the comparison of distances strategy and skimming strategy. Figure 5 illustrates these two decision rules with the triangle and 3-AFC tests.

Figure 5 represents a triadic test where two *A* and one *B* samples were presented. In the triangle test, the

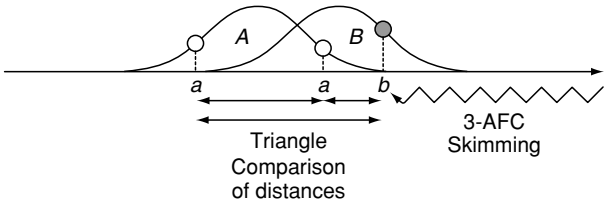

Figure 5 Comparison of distances and skimming strategies in the triangle and three-alternative forced choice (3-AFC) tests.

subject will compare the distances to determine which of the three samples is most different from the other two. We see that the subject will select the *A* sample furthest left and his answer will be recorded as incorrect. For the 3-AFC, the subject needs only to select the sample with the largest magnitude. Therefore he will pick the *B* sample (most intense) and his answer will be recorded as correct. This example illustrates why subjects' performance will be lower in the triangle than in the 3-AFC test.

We can generalize this example to the other protocols described earlier. For directional discrimination tests (*m*-AFC tests), subjects will use a skimming strategy. For the nondirectional protocols (triangle, duo–trio), subjects will use a comparison of distances strategy. The *A*/Not *A* and same–different tests use alternative strategies, since the instructions require the use of a psychological criterion (How strong does a sample need to taste before I call it '*A*'? How different do two samples need to taste before I call them 'Different'?) The *A*/Not *A* test has a strategy analogous to a skimming strategy, while the same–different test has a strategy analogous to a comparison of distances strategy.

Protocols with a skimming strategy are statistically more powerful, i.e., not as likely to fail to detect a nonnegligible difference as a test using a comparison of distances. This was illustrated earlier with the triangle and 3-AFC tests.

When using d' values, results from discrimination tests now become independent of the protocol used. This is what we would predict, as discrimination tests are merely an instrument to measure product differences. Simply calculating the proportion of tests correct and using the binomial test will provide bias and unstable information.

Practical aspects of Thurstonian modeling Now that we have ways of measuring a degree of difference between food products, we need to set up the experimental parameters that will allow us to detect it accurately. This brings us to the issue of power.

Power In order to define the power of a test, we need to decide what degree of difference we want to be able to detect. This will require some general knowledge of the magnitude of the difference that usually needs to be detected and can be determined using past data and converting them in terms of d' values. The power of a test can be illustrated by calculating the sample size needed to detect a given d' with a given confidence. The smaller the sample size required, the more powerful the test. Equations have been published providing the sample size needed for a given d' value, α level, and power. Tables can be constructed and an example is shown in Table 4.

Table 4 Sample size required to detect a given degree of difference (d') for given α and power levels and discrimination protocol

d'	α	*Power*	*2-AFC*	*Duo–trio*	*3-AFC*	*Triangle*
0.5	5%	0.8	78	3092	64	2742
1.0	5%	0.8	20	225	15	197
1.5	5%	0.8	9	55	6	47
0.5	1%	0.8	128	5021	103	4441
1.0	1%	0.8	34	366	25	318
1.5	1%	0.8	16	91	11	76
0.5	5%	0.9	108	4283	89	3810
1.0	5%	0.9	27	310	21	276
1.5	5%	0.9	12	76	9	66
0.5	1%	0.9	165	6511	135	5776
1.0	1%	0.9	43	473	32	416
1.5	1%	0.9	20	117	14	100

AFC, alternative forced choice.

Here is the way to read the table. For the 2-AFC, first line: if we run 78 tests and detect a significant difference at the 5% level, we are 95% sure that a difference exists between the products. If we do not detect a significant difference at the 5% level, we are 80% sure that the d' was less than 0.5. In order to reach the same confidence level for the duo–trio, 3-AFC, and triangle tests, we need a sample size of 3092, 64, and 2742 respectively. These variations in sample size indicate the larger power of the protocols using the skimming strategy.

The same–different and *A*/Not *A* tests cannot be included in this table because of the issue of response bias. A power value would need to be calculated for each criterion location and size. However, as a general rule, the same–different test is slightly more powerful than the triangle and duo–trio tests but less powerful than the 2- and 3-AFC tests, while the power of the *A*/Not *A* test is closer to that of the 2-AFC test. N.B.: power tables have been published using the proportion of discriminators in the population. This notion is flawed (it is directly linked to the proportion of correct answers, which we've seen is not a reliable measure of the degree of difference between products) and such tables should be used with extreme caution.

The power of a test is also shown in the calculated variance of d': the more powerful the test, the smaller the variance of d'. This variance is very useful when comparing several d' values (for instance, when comparing a standard to several reformulations) in order to determine whether they are significantly different from each other.

Replicated testing Considering the sample sizes required for the duo–trio and triangle tests, it would be more suitable to use the 2- or 3-AFC protocols. If it is impossible to use the directional difference tests, the power of the triangle and duo–trio tests can be somewhat increased by replicating the number of tests per subject. However, combining data from different subjects brings the issue of overdispersion. References in the Further Reading section at the end of this chapter should be consulted in order to insure accurate data analysis.

Experimental variables Under certain conditions, d' values obtained from discrimination tests may not correspond exactly. This may be due to the effect of experimental variables such as memory and sequence/adaptation effects which are not considered in the models. Since a larger d' value requires a smaller sample size to be detected, it may be beneficial to use protocols providing larger d' values. Experimentally, it has been found that memory requirements can significantly hinder subjects' performance and thus product discrimination. Under certain conditions, tests with only two samples (2-AFC, same–different) have been found more sensitive (higher d' value) than those with more stimuli (3-AFC, triangle).

This aspect of difference testing should also be considered when selecting a protocol to conduct a discrimination study.

Conclusion

For an unaware scientist, the topic of discrimination testing can be perceived as deceptively simple and limited. However, we showed here that proper knowledge from psychology, physiology, and statistics is necessary in order to insure proper data collection and analysis. Furthermore, discrimination testing can provide very valuable information by not limiting the information to a mere yes/no answer regarding the existence of a difference. This can be achieved by calculating d' values and their variances from widely available tables, and by determining the magnitude of the difference between the samples.

This chapter summarized the scope of this topic. The reader is advised to refer to the review articles cited at the end of this section in order to build a more general knowledge about the topic of discrimination testing.

We still need to mention that the concepts of sensory magnitudes and d' values are not limited to sensory difference tests and that they can easily be extended to ratings on category scales and consumer preference and hedonic data. This approach is extremely useful for sensory science since it allows the connection of all kinds of sensory measurements in a common structural framework.

See also: **Carbohydrates**: Sensory Properties; **Sensory Evaluation**: Sensory Characteristics of Human Foods; Food Acceptability and Sensory Evaluation; Practical Considerations; Sensory Rating and Scoring Methods; Descriptive Analysis; Appearance; Texture; Aroma; Taste

Further Reading

Bi J, Ennis DM and O'Mahony M (1997) How to estimate and use the variance of d′ from difference tests. *Journal of Sensory Studies* 12: 87–104.

Brockhoff PB and Schlich P (1998) Handling replications in discrimination tests. *Food Quality and Preference* 9: 303–312.

Ennis DM (1993) The power of sensory discrimination methods. *Journal of Sensory Studies* 8: 353–370.

Ennis DM and Bi J (1998) The beta-binomial model: accounting for inter-trial variation in replicated difference and preference tests. *Journal of Sensory Studies* 13: 389–412.

Green DM and Swets JA (1966) *Signal Detection Theory and Psychophysics*. New York: Wiley.

Kaplan HL, Macmillan NA and Creelman CD (1978) Tables of d' for variable-standard discrimination paradigms. *Behavior Research Methods and Instrumentation* 10: 796–813.

Lawless HT and Heymann H (1998) *Sensory Evaluation of Food: Principles and Practices*. New York: Chapman & Hall.

Macmillan NA and Creelman CD (1991) *Detection Theory: A User's Guide*. New York: Cambridge University Press.

Meilgaard M, Civille GV and Carr BT (1999) *Sensory Evaluation Techniques*, 3rd edn. Boca Raton: CRC Press.

O'Mahony M (1986) *Sensory Evaluation of Food: Statistical Methods and Procedures*. New York: Marcel Dekker.

O'Mahony M (1995) Who told you the triangle test was simple? *Food Quality and Preference* 6: 227–238.

O'Mahony M, Masuoka S and Ishii R (1994) A theoretical note on difference tests: models, paradoxes and cognitive strategies. *Journal of Sensory Studies* 9: 247–272.

Stone H and Sidel JL (1993) *Sensory Evaluation Practices*, 2nd edn. San Diego: Academic Press.

Sensory Rating and Scoring Methods

J A McEwan, MMR Food and Drink Research Worldwide, Wallingford, UK
D H Lyon, Campden Food and Drink Research Association, Chipping Campden, Gloucestershire, UK

Purpose of Rating and Scoring Methods

Rating and scoring methods provide the basis for quantification of sensory information. Although these two terms are sometimes used interchangeably by sensory scientists, they have different meanings.

Rating refers to the quantification of information by the use of ordinal categories, while scoring is a more defined form of rating as it uses a numerical interval or a ratio scale, of which the properties are known. A scale can be defined as a measurement continuum divided into successive units according to the properties associated with it.

There are many different rating and scoring methods used in sensory analysis, as illustrated both here and in other articles. In each case, these scales are physical measurement tools used to measure some sensory phenomenon perceived by individuals. Thus, implicit in using rating and scoring is that these scales provide meaningful representations of some psychological process or processes.

When considering rating and scoring methods, the reader should be aware that the experimental design considerations given to sensory analysis procedures should be observed. In this article, example forms are given with the illustrations of different scales to aid the reader in designing appropriate questionnaires.

Type of Scale

Four types of scale can be used to collect data: nominal scales, ordinal scales, interval scales, and ratio scales.

Nominal Scale

A nominal scale is one where data collected are categorized by a name or a number. Each observation collected using these scales must fall within one of the categories. For example, 'canned,' 'frozen,' 'dried,' 'chilled,' and 'fresh' are five categories used to describe methods of food preservation. These categories have no logical ordering and, thus, the key point about nominal scales is that the different categories have no quantitative relationship.

Ordinal Scale

An ordinal scale is one which allows observations to be ordered according to whether they have more or less of a particular attribute. Successive numbers or words are used to indicate more (or less) of the attribute being measured. Ordinal scales do not allow the amount of difference between observations to be quantified. The nine-point hedonic scale (described later) is ordinal, as are ranked data.

Interval Scale

An interval scale is one where the distance between points on the scale is quantifiable. In many instances, the distance or intervals between points on a scale will represent an equal perceptual distance. For example, if the perceptual distance between 1 and 2 on a seven-point scale of sweetness was the same perceptual distance as between 2 and 3, 3 and 4, and so on, then this scale would have interval properties.

Ratio Scale

A ratio scale is one where the observations collected can be expressed as a percentage or ratio of each other. For example, a person eating 100 g of chocolate a day eats twice as much as a person eating 50 g day^{-1}. An example of a ratio scale in sensory analysis is magnitude estimation, which will be discussed later. The main difference between interval and ratio scales is that the latter has a true zero, whereas the zero point of an interval scale is arbitrary.

Data Collection Methods and Data Analysis

Nominal Data

Nominal data can be collected in a number of ways (Figure 1), but common to all nominal data is that each observation can only fall into one category. A logical first step in analysis and interpretation of the data, therefore, is to produce a histogram indicating

Please taste the sample coded 457, and identify which of the four basic tastes you perceive.

Sweet _____
Sour _____
Salt _____
Bitter _____
Unable to Identify _____

Figure 1 Taste identification test using a nominal scale.

Please taste the sample coded 658, and indicate how firm it is by placing a tick in the appropriate box below.

Not firm — Very firm

0	1	2	3	4	5	6	7	8	9

Figure 2 Evaluation of the texture attribute firm, using a numeric category scale.

Please taste the sample coded 943, and rate the sweetness intensity by placing a tick in the appropriate box below.

Not sweet	Slightly sweet	Moderately sweet	Very sweet	Extremely sweet

Figure 3 Evaluation of the taste attribute sweet, using a verbal category scale.

the frequency of occurrence of each category. The next step is to determine whether more observations fall in one category than another, or if the distribution of counts for one sample is the same as for another. In either case, the usual method of analysis is by the χ^2 (chi-squared) test.

A more advanced method for treating nominal data is the technique of multiple correspondence analysis. This method produces a multidimensional spatial representation of the relationship between samples and attributes, which is often a useful summary of the data.

Please taste the sample coded 183, and indicate how much you like or dislike the flavor by placing a tick by the appropriate descriptor.

_____ Like extremely
_____ Like very much
_____ Like moderately
_____ Like slightly
_____ Neither like nor dislike
_____ Dislike slightly
_____ Dislike moderately
_____ Dislike very much
_____ Dislike extremely

Figure 4 Hedonic assessment of flavor using a nine-point verbal category scale.

Rating Methods

Rating methods involve the quantification of perceived sensations by the use of scales.

Category scales Category scales are widely used in sensory analysis, both for objective assessment and affective (related to preference or liking) response. Category scales for objective sensory measurement are most often unipolar, as it is the amount of a particular attribute which is being measured. An example of a unipolar scale for firmness is given in Figure 2. A category scale may also be constructed such that each category has a verbal label attached to it, as illustrated in Figure 3. Bipolar scales can be used to measure attributes such as texture, e.g., substitute the two anchors of Figure 2 with 'soft' and 'firm,' respectively. While these scales give the appearance of having interval properties, care should be taken in making this assumption without experimental evidence. Generally speaking, category scales provide ratings and not scores; thus, category scales provide ordinal data.

A well-known scale for affective measurement is the nine-point hedonic scale (Figure 4). Variations of this rating scale exist, comprising fewer categories and the absence of the middle category.

Continuous line scales Continuous line or visual analog scales take the form of an unstructured line, as illustrated in Figure 5. Such scales are usually unipolar and measure the two extremes of an attribute, the extremes of which are represented as anchor points at the left and right of the scale. The length of the line is often 100 mm, although this may be longer or shorter according to need. However, research conducted by both psychologists and market researchers has indicated that 100 mm is satisfactory. A line that is too short may decrease the discrimination achievable by the assessors, while beyond a particular length it is unlikely that an assessor will provide more precise information.

Continuous line scales are often used with trained sensory panels as they allow the assessor to be more discriminating than with the category scale. Data collected on rating scales are measured by an interval ruler; however, strictly speaking the data should be considered ordinal, as it is usually impossible to prove

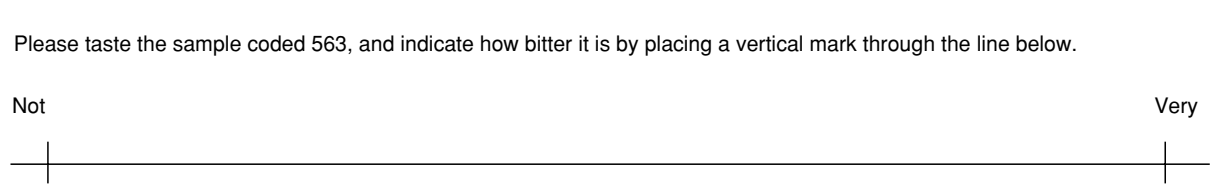

Figure 5 Evaluation of the taste attribute bitter, using a continuous line scale.

the linear (interval) properties of every continuous line scale used for sensory measurement.

With a trained sensory panel, continuous line scales are usually assumed to be linear with the understanding that the extremes of the scale may be slightly curved due to end effects. If continuous line scales were truly interval, then the data would be scores rather than ratings. It would be unwise, however, to make this assumption with an untrained panel or a consumer panel without justification.

Data analysis Data that have interval properties can be analyzed using parametric statistical methods, providing certain assumptions are satisfied. These methods assume that the data are normally distributed, which implies a symmetric distribution (i.e., the mean, median, and mode are equal) and that the data have interval or ratio properties. Parametric methods of analysis are powerful methods for sensory interval data as they enable more precise interpretation to be made of the results. There is a wide range of statistical methods available for interval data, depending on the question being asked. The simplest form of analysis is to calculate means and standard deviations, which measure the location and spread of the data, respectively.

This can be taken one step further by calculation of a standard error and confidence interval for the mean. This allows the sensory analyst to make inferences about the mean of a sample with respect to the attribute being measured, or to test a hypothesis that the mean is equal to a particular value. *t*-tests can be used to test the hypothesis that two samples have different mean values for a particular attribute, while analysis of variance will test whether more than two samples have different mean values.

When several attributes are being evaluated, multivariate analysis procedures are often adopted. Techniques that can be used include principal component analysis (PCA), generalized Procrustes analysis, factor analysis, discriminant analysis, canonical variate analysis, and cluster analysis.

Whilst data collected by rating methods are mostly ordinal, it is a common practice to use the univariate parametric statistical methods discussed. However, these methods assume that the data are derived from an underlying normal distribution, which assumes that the data are continuous, i.e., they have interval or ratio properties. Clearly, by definition, ordinal data cannot be continuous. In fact, for each of the above-mentioned parametric statistics there is a nonparametric equivalent. These are medians (means), interquartile ranges (standard deviations), Mann–Whitney *U*-test and Wilcoxon paired test (*t*-tests), and the Kruskal–Wallis and Friedman rank test (analysis of variance).

Multivariate analysis on ordinal data can be achieved through correspondence analysis (CA), which reduces the dimensionality (number of measured attributes) to a smaller number of dimensions, which effectively describes the correspondence between samples and attributes. However, there is little evidence to suggest that CA is better than PCA for sensory profile data collected on ordinal scales.

Magnitude estimation

Method Magnitude estimation is a form of ratio scaling where the perception of a specified attribute in one sample is measured as a ratio of the perception of that attribute in another sample. Stevens introduced this method in 1953, and one of its first applications in sensory analysis was to evaluate pleasantness of odors.

There are two main methods of collecting ratio data by this method. In the first method, one sample is designated the standard and allocated a whole positive (not zero) number (e.g., 100) to present the perception of a specified attribute. The perception of this attribute in subsequent samples is represented as a fraction or multiple of the standard number. For example, if the sample evaluated is twice as sweet as the standard then it is allocated the value 200, if it is a quarter as sweet than it is allocated the value 25. It is often recommended that the standard sample should be one-third of the way along in the range of samples to be used in the experiment.

In the second method, the assessor allocates a positive (not zero) number to the first sample, and evaluates the perception of that attribute in subsequent samples as a ratio of the first sample. Whichever approach is adopted, magnitude estimation usually requires more training than the other methods of

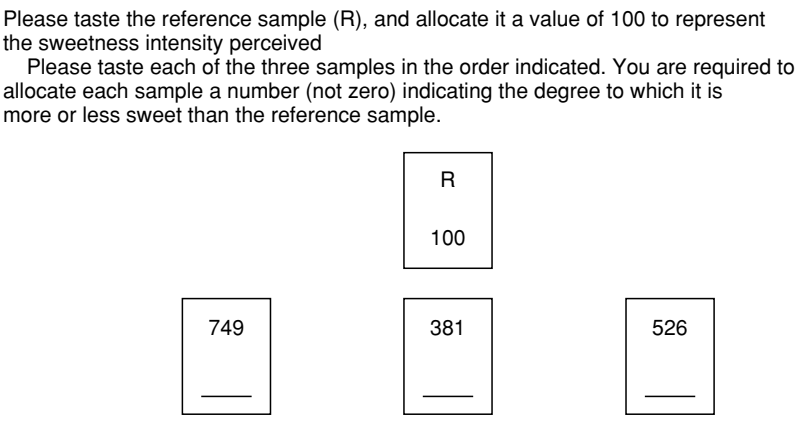

Figure 6 Flavor assessment by magnitude estimation, where 100 represents the reference against which other samples are evaluated.

data collection. An example of a magnitude estimation form is given in Figure 6.

Data analysis Data collected on a ratio scale, by magnitude estimation, require additional thought as they have to be transformed before analysis. In considering data analysis, it is important to recall that the data are ratios for each individual. Further, as each individual may have used his or her own range within the ratio scale, the distribution of the data will differ from individual to individual. It is also likely that if an individual's data are plotted, the shape of the distribution of data about the standard sample will differ for each subject. Thus, not only are there scaling problems to contend with, but there is also inconsistency in the distributional plot of the data. For these reasons, individual data cannot be assumed to come from a normal distribution, and further direct averaging of data would be misleading due to the scaling problems. Thus, data are transformed to insure that each individual has, in effect, used the same scale. Several methods can be used to rescale magnitude estimation data: modules equalization, modulus normalization, and external calibration.

Modules equalization requires that the ratings provided by each assessor are multiplied by a constant, with the aim of insuring that the geometric mean of each individual is the same as the geometric mean of the grouped data. Modulus normalization is used if many samples are evaluated over a number of sessions, and the experimenter has built in common samples that appear at each tasting session. The information from the common samples is then used to rescale the data based on the geometric mean.

Averaging magnitude estimation data is not as straightforward as interval data. However, these data usually have a log-normal distribution and hence, by taking the logarithms distribution of the data becomes symmetric and hence more like a normal distribution. It is often easier to analyze logarithmic data. Parametric statistics, as described above, can then be used to analyze transformed magnitude estimation data if it is certain that the transformed data are normally distributed; otherwise nonparametric methods must be used.

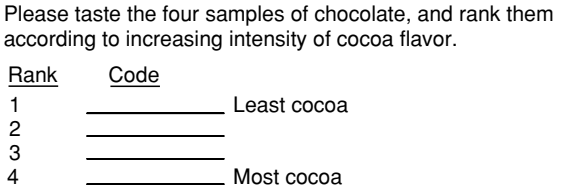

Figure 7 Flavor assessment by ranking, where the higher rank represents higher intensity of cocoa flavor.

Ranking

Ranking is neither a rating nor a scoring method, but as it is widely used in sensory analysis and provides ordinal data, it is included here for completeness. The method of ranking is used to order a number of samples according to increasing or decreasing perception of a specified attribute. Samples to be ranked need to be presented at the same time, and this limits the total number of samples that can be used in such a test, due to sensory fatigue. Ranking of five or six samples is often possible, although this number will depend largely on the nature of the samples; for example, chocolate has mouth-coating properties that make it difficult to rank many samples. An example of a ranking exercise is illustrated in Figure 7. Data are analyzed using nonparametric statistical methods, as described earlier in this article.

See also: **Sensory Evaluation**: Sensory Difference Testing; Descriptive Analysis

Further Reading

Carpenter RP, Lyon DH and Hasdell TA (2000) *Guidelines for Sensory Analysis in Food Product Development and Quality Control*, 2nd edn. Maryland: Aspen.

Chatfield C and Collins AJ (1980) *Introduction to Multivariate Analysis*. London: Chapman and Hall.

ISO (1987). *Sensory Analysis – Guidelines for the Use of Quantitative Response Scales*. ISO 4121. (new version on draft.)

Land DG and Shepherd R (1988) Scaling and ranking methods. In: Piggott JR (ed.) *Sensory Analysis of Foods*. London: Elsevier.

Lawless HT and Heymann H (1998) *Sensory Evaluation of Food: Principles and Practices*. New York: Chapman and Hall.

Lebart L, Morineau A and Warwick KM (1984) *Multivariate Descriptive Analysis: Correspondence Analysis and Related Techniques for Large Matrices*. New York: Wiley.

Moskowitz HR (1977) Magnitude estimation. Notes on what, when and why to use it. *Journal of Food Quality* 3: 195–228.

O'Mahony M (1986) *Sensory Evaluation of Foods: Statistical Methods and Procedures*. New York: Marcel Dekker.

Peryam DR and Giradot NF (1952) Advanced taste-test method. *Food Engineering* 24: 58–61, 194.

Peryam D R and Pilgrim FJ (1957) Hedonic scale method for measuring food preferences. *Food Technology* 11: 9–14.

Stevens SS (1953) On the brightness of lights and the loudness of sounds. *Science* 118: 576.

Descriptive Analysis

H Stone and J L Sidel, Tragon Corporation, Redwood City, CA, USA

Introduction

Descriptive analysis continues to attract interest and applications in more and more companies as the results from tests are discussed and used in support of product market decisions. At the same time research on new methods has been discussed, offering the sensory professional a variety of options. As a methodology, descriptive analysis is the most sophisticated source of product information available in the context of providing a complete and quantitative word description of a product's sensory characteristics. Such information can be considered in the same way as one considers chemical or biological information about products. Because it is quantitative, it provides a focus for development efforts; it provides a basis for measuring the effects of a process or of ingredients; i.e., establishing causal relationships between ingredients, technology, and perceptions; it provides a basis for linking consumer response behavior to specific product characteristics, and for business applications, identifying those product attributes that are most important to consumer preferences. Before discussing methodologies, it is necessary to define what we mean by descriptive analysis.

Descriptive analysis is a sensory methodology that provides quantitative word descriptions of products based on perceptions verbalized by a group of qualified subjects. It is a total sensory description, taking into account all the sensations that are perceived – visual, auditory, gustatory, olfactory, kinesthetic – when the product is evaluated. This evaluation could include product handling and use at home, for example, and in that sense, it is a more typical and total experience. The evaluation could focus on a single modality of a product, such as aroma or texture; however, there is considerable risk in such a narrow focus, a risk that will be discussed later in this article.

The earliest practitioners of descriptive analysis were the brewmasters, perfumers, flavorists, and other product specialists. Employers appreciated the value of their information because they not only described products and made recommendations about the purchase of specific raw materials, they also evaluated the effect of process variables on finished product quality (as they determined product quality). They also determined that a particular product met their criteria (and that of their company) for manufacture and sale to the consumer. These activities served as the basis for the foundation of sensory evaluation as a science, although at the time it was not considered within that context. It was possible for the expert to be reasonably successful as long as the marketplace was limited (i.e., less competitive and regional in character). However, this changed with the scientific and technical changes in consumer products industries and especially in the food and beverage industry. Use of more sophisticated raw materials and processes made it increasingly difficult for experts to be as effective as in the past. The increasingly competitive and international nature of the marketplace, the rapid introduction and proliferation of new products, and the sophistication of the consumer's palate further challenged the expert's ability to function as in the past. During this time, new sensory test methods and more sensitive scales were described along with improved statistical procedures for the analysis of responses. These developments enabled the sensory professional to provide more precise product information and participate more fully in product market

decisions. Formal descriptive analysis (and the separation of the expert from sensory evaluation) received its major impetus from 'flavor profile,' an approach which demonstrated that it was possible to select and train individuals to describe the sensory properties of a product in some agreed sequence, leading to actionable results without dependence on the individual expert. The method was distinctive in that no direct judgment was made concerning consumer acceptance of the product, although most investigators assumed consumer acceptance based on a result. The method attracted considerable interest as well as controversy; however, there was no question as to its importance to the development of the field. Since then, other methods have been developed, in part responding to its limitations as well as developments in measurement and human perception.

Background

We begin this discussion by first focusing on the most fundamental issues on which all descriptive methods are based, i.e., the subject selection process, the ability of the subjects to verbalize perceptions, and their ability to assign measures of strength (or intensity) to those perceptions in a reliable manner. Implicit in this statement is the recognition that each person is unique in terms of sensory skills; different individuals can be differentially sensitive to a sensation, more than a single subject (the expert n of 1) and more than a single judgment (from each person) is required for each product evaluated. People exhibit variation in their sensory skills and this variation reflects the state of those individuals at the time of that judgment – their physiology, attitudes, and motivation. Not only are people different from each other, but they also vary within themselves, from moment to moment, from day to day, and so forth. This individual variation is one of the most critical aspects of the descriptive process (and of all human response behavior) and it is often conveniently ignored. This issue will be discussed in more detail in subsequent sections; suffice to say that it must be accounted for in any descriptive test if results are to be believed, and to have any hope of external validity and credibility with the test requestor.

This discussion is particularly important because, in the course of organizing a descriptive panel, there are numerous decisions made by the sensory professional (the panel leader) that determine the usefulness of the results. These decision points and the actions taken derive from the panel leader's understanding of (1) the perceptual process in general and (2) the specific descriptive process used. Unlike discrimination and acceptance tests where subjects exhibit choice behavior in a global sense, i.e., all perceptions are taken into account to yield a single judgment, the descriptive test requires each subject to provide numerous judgments (one for each attribute) for each product. These judgments can be influenced by the instructions given to the subjects and the type and extent of the training. Too much information/instruction can yield a result that reflects the panel leader's knowledge of the products and not the subjects' perceptions. Alternatively, too little information will result in a lack of focus, as evidenced by greater than expected variability and loss of product differentiation.

Using the quantitative descriptive analysis (QDA) methodology as a reference, various methods, such as spectrum analysis, are described and contrasted. Other methods are also mentioned; however, most are modifications of the two previously mentioned. Free-choice profiling, the one different method requiring no subject selection and no training, has undergone many changes such that it more closely approximates the afore-mentioned methods and no longer warrants separate consideration.

All descriptive methods share some common characteristics as well as some differences and these will be compared and contrasted in the following sections. To understand better these similarities and differences, it is useful to think in terms of the steps involved in the development of a descriptive panel capability. These include four main activities: screening, training or language development, data collection, analysis and interpretation, and a fifth often overlooked activity, panel performance.

Screening

A descriptive test typically involves relatively few subjects (as few as 10 to as many as 20). Reliance on so few subjects is possible provided there is good evidence that the specific differences obtained are reliable and valid, and not the result of spurious responses from one or two insensitive or more variable subjects. An effective means of minimizing this latter problem is to screen the subjects for their sensory skills before training. One can show that in any population there is a wide range of sensory skills; some individuals will be very sensitive and some will be very insensitive (sometimes by as much as 100-fold or more), and this skill can be product- and attribute-specific. A goal of screening is to identify and excuse from further testing those individuals who cannot perceive product differences at better than chance – those that are insensitive or so variable as to raise doubt as to their participation in a test. Failure to carry out adequate screening raises serious questions

as to the quality of any sensory information obtained. Empirically, we observe that individuals whose sensory skills are not equal to or better than chance are usually less sensitive and more variable than the person who is sensitive and whose performance is better than chance. This enables one to use the fewest number of subjects without loss of information.

There are differences in the approaches used for the qualification of subjects, and panel leaders must decide what procedure(s) they are willing to use. For example, it has been recommended that screening for food or beverage tests be based on sensitivity to particular chemicals (e.g., sweet, sour, salt, and bitter stimuli or an identification of specific odorants) and/or on various personality tests. It is surprising that these types of sensitivity tests continue to be recommended (as used in flavor profile and related methods) in spite of earlier evidence that it had a low correlation with a subject's subsequent performance. The spectrum analysis method suggests using a combination of threshold testing followed by discrimination with actual products. Free-choice profiling claimed subject screening was not necessary and atypical responses would be dealt with by statistical means.

In QDA, the discrimination methodology is the screening method of choice. It is preceded by determining that the participant uses and likes the products that will be tested. Regular users of a product are more sensitive to differences in those products (e.g., compared to the infrequent user). The discrimination model is very effective for establishing the level of sensory skill among products that will be tested. It is also the most parsimonious in terms of time required to establish the level of skill. Based on results of 20–30 discrimination trials completed over 3 or 4 days (daily sessions last about 90 min), one can identify discriminators from amongst a group of totally naive individuals. The trials themselves encompass all the modalities that are expected to be involved, and the pairs of products are made progressively more difficult. Working with qualified subjects, extensive screening tests are not needed unless there is evidence that some subjects exhibited decreased sensitivity in a previous test, or one is using an entirely different product category, and again, there is reason to believe that such screening is warranted. Screening has several objectives: first, it identifies the less sensitive, more variable subject; second, it familiarizes the subjects with the sensory characteristics of the products; and third, it identifies any individuals who have difficulty following instructions. However, there are no guarantees, and it is only after data collection that one can empirically determine the effectiveness of the screening and subsequent training activities.

This screening procedure is intended to be product- or category-specific, as is the subsequent training effort. For example, if one were testing Cheddar cheeses, then screening would use samples of Cheddar cheeses, and preferably those that will be tested, and not other foods or unrelated ingredients or specific chemicals. This does not mean that subjects are product-specific, rather that screening should be of the products that will be tested or used during the language sessions. This should not be construed to mean that one must qualify subjects every time a different set of products is evaluated. Most subjects are capable of meeting the qualifying criteria for many different categories of products. Experienced subjects (e.g., those who have participated in at least one test within the past 1 or 2 months) will have learned how to use their senses and how to evaluate products within a category being evaluated without a significant loss of sensitivity, and are therefore very likely to be able to evaluate products beyond those for which they were screened. The panel leader makes this decision based on the existing subject performance data.

Screening subjects for more than one type of product makes sense provided the tests will follow one another with minimal delay. It will eliminate the need to initiate any additional screening before the next test. This does not guarantee that a subject will stay qualified, nor does it eliminate the need for the other stages of training. In addition, it does not excuse the panel leader from monitoring subject performance before the next test. Infrequent use of subjects, e.g., no testing for 3 or more weeks, may require that some screening is necessary, in which case much of the earlier effort will have been wasted. A potential problem with screening subjects on several products is the assumption that these individuals will be willing to participate in future tests. If used too frequently (e.g., daily) there is every likelihood that the subjects will lose interest and thus not perform satisfactorily. It may be better to develop a larger pool of qualified subjects and reduce reliance on a small subset of people for all testing, or qualify small groups for each category depending on the anticipated need for the product information. Panel leaders must develop guidelines for subject use based on their company experiences and not blindly follow an external practice.

In summary, it is necessary that subjects for a descriptive test demonstrate their ability to perceive differences at better than chance among the products that they will be testing. This will be insured if products used are representative of the products that will be tested. As a guide in selecting subjects we have used a minimum of 65% correct across all discrimination trials with greater emphasis placed on those

pairings that were more difficult. However, there have been situations in which the number of subjects meeting this criterion is not sufficient and it is necessary to lower this percent. In some instances an individual has been selected whose performance has been as low as 50%. This latter situation would be atypical, one which would be a decision of the panel leader. It is important to keep in mind that screening cannot guarantee subsequent performance; however, there is no doubt that individuals whose skill is less than chance do not perform well (in descriptive tests). Over many thousands of such tests we have observed that about 30% of those who volunteer will fail to meet these performance criteria. This particular point has implications that extend well beyond descriptive testing.

Training

Training is initiated directly after screening and it is here that one also encounters different approaches. While the training process is usually focused on the language (or attributes) used as a basis for scoring the products, there are other equally important activities. These include grouping attributes by modality, ordering them by occurrence within a modality, developing definitions for each attribute, familiarizing subjects with scoring the attributes, and identifying references that are helpful in training. Inexperienced subjects starting with a new set of products (and no list of attributes) will need about 7–10 h of training. For this same panel and an existing language, the time is reduced to about 6–7 h. Experienced subjects usually need about 5–6 h when presented with an entirely new product category. These training times are intended solely as a guide, as the products and the skill of the panel leader will also have an effect. Regardless of the situation, the subjects work, at times individually but mostly as a group, to insure that the attributes are sufficiently understood by one another, and that all of the product's characteristics have been fully accounted for. All of this is done under the direction of a panel leader who does not participate in developing the language. The QDA method was the first method and remains one of the few that excluded the panel leader from directly contributing attributes in language development. This is true whether one is working with an experienced panel or with a panel that has never participated before. The panel leader's primary responsibility is to facilitate communications among subjects and organize results of their discussion. Participating as a subject (as is allowed with the spectrum method) is not recommended because of bias related to one's awareness of product differences and the test objectives. Not only does this action communicate the wrong message to the subjects (subjects will tend to defer to the panel leader, whom they assume has the correct answer), but also the end result is more likely to be a group judgment rather than a group of judgments.

Developing a sensory language or using one that already exists is an interesting process and certainly one that is essential for the successful outcome of a test. For some, the language assumes almost mystical importance, such that a considerable body of literature has been developed in which lists of words (lexicons) are published for specific types of products. Implicit in use of an agreed list is the assumption of universality, i.e., panels in different parts of the world would yield the same result by using the same language. It appears that many of these lists were prepared by technologists and sensory professionals by consensus. Unfortunately, this is not very different from the efforts of product specialists or experts of 50 or more years ago when they developed quality scales and corresponding descriptions for products as part of an effort to establish food quality standards. Besides their interest in evaluation of their respective company's products, their technical and trade associations often formed committees for the express purpose of developing a common language for describing the flavor (or odor) of the particular category of products. For purposes of this discussion we chose the publication by Clapperton *et al.* (1975) on beer flavor terminology as an example of the efforts (of the association of brewers) to develop an internationally agreed terminology. These authors stated the purpose as follows: 'to allow flavor impressions to be described objectively and in precise terms.' Various literature and professional sources were screened for words (or descriptors) describing the flavor of beer, and, after reviews, the committee arrived at an agreed list. As the authors noted, the issue of chemical names compared with general descriptive terms was resolved by the inclusion of both.

From a technical viewpoint, the use of chemical names was appealing because it was believed that they could be related directly to specific chemicals in the beer. An example would be the term 'diacetyl,' which could be ascribed to a specific chemical, as compared with 'buttery,' a less precise term that might be ascribed to several chemicals. By including both types of terms it was stated that the terminology would be of value to the flavor specialist (the expert) and the layperson (the consumer). While the concept of an agreed terminology should have considerable appeal, careful consideration of this approach reveals significant limitations, especially in relation to its application beyond that technical panel. After all,

the words used to describe sensations are nothing more than labels to represent sensations – no causality is inferred. It is risky to decide *a priori* what word(s) subjects should use to describe a particular sensation, or still greater risk that naive consumers will understand them. Recall that about 30% of any population of consumers cannot differentiate products at least at chance. In addition, subjects will not be equally sensitive to all sensations. Some sensations are complex, and in all likelihood subjects will want to use more than a single word to represent that experience. The fact that product changes based on formulation, process, or both rarely yield a single sensation means that, for example, a flavor change would not be totally represented. To restrict subjects to specific words assumes that the language is unchanging, as are the meanings assigned to it, and that all subjects will associate a specific word to a specific sensation. The frame of reference for an attribute is unique to each subject, and attributes requiring a very different frame of reference are difficult for inexperienced and/or nontechnical subjects to understand. While chemical terminology is supposed to have specific meaning to an expert, it is unlikely to have the same meaning to a consumer, to a trained subject, or to another expert.

It should be noted that the use of technical terminology extends training time and this appears to be related to the complex nature of this terminology, especially for subjects with no technical training. The spectrum method provides the subjects with attributes, as well as references and designated intensity scores for those references, all of which are intended to enhance the testing process. The use of standard or previously specified attributes and references implies that product variables will not produce unique sensations, that references themselves are not variable, and forces subjects to limit the value of their perceptions relative to what has been perceived in the past by others. The training effort is extended over a period of 3 months; this raises questions as to its responsiveness and overlooks the inherent variability in the subjects (over time) and the references that are being used (that also change over time). Regardless of the source, a language that does not provide for subject input is unlikely to yield uncomplicated sensory responses. Subjects are influenced by the information given to them, and are much less likely to question it, because of its source. While it can be argued that such approaches are merely intended to help panel leaders and the subjects, the temptation is very strong to use this approach rather than allowing subjects to use their own terminology and, most important, to allow subjects to express what they perceive.

To the student of the history of psychology, descriptive analysis can be considered as a form of introspection, a methodology used by the school of psychology known as structuralism in its study of the human experience. Structuralism required the use of highly trained observers in a controlled situation verbalizing their conscious experience (the method of introspection). Of particular interest to us is the use of the method of introspection as an integral part of the descriptive analysis process and, specifically, the language development component.

However, the obvious difference is that products are included (in descriptive analysis) so that the subject's responses are perceptual and not conceptual. Subjects are encouraged to use any words, provided that they explain to the other members of the panel what they mean – they define the meaning of each word-sensation experience. The goal is for the panel, by consensus, to associate a particular sensation with a particular word or group of words. While each subject begins with his or her own set of words, they work as a group to come to agreement as to the meaning of those words, i.e., the definitions or explanations for each word-sensation experience, and also when they (the sensations) occur. In addition, they also develop a standard evaluation procedure. All of these activities require time; in the QDA methodology, there can be as many as five consecutive daily sessions, each lasting about 90 min. This amount of time is essential if the sensory language is to be developed and understood, and the subjects are capable of using it (and the scale) to differentiate the products. These sessions help to identify the word-attributes that could be misunderstood, and also enable the subjects to practice scoring products and discussing results, on an attribute-by-attribute basis. Of course, all this effort cannot make subjects equally sensitive to all attributes. In fact, subjects rarely, if ever, achieve complete agreement for all attributes, nor are subjects equally sensitive to all attributes, nor should it be expected (if this did occur, one could rely on the n of 1). The effectiveness of the training process can be determined only after a test, after each subject has scored products on a repeated trial basis and the data have been analyzed. The idea that there should be a specific number of attributes is, at best, questionable, as is the issue of whether or not one has the correct attributes. How does one know that all the attributes have been developed or that they are the right ones? The answer to the former is empirical (and in part answered in much the same way as one determines the number of angels that can occupy the head of a pin). The answer to the latter is also empirical, i.e., given a set of variables, do some attributes exhibit systematic changes as a function of those

variables? Other methods, including free-choice profiling, claim that no training is required; subjects can use any words they want and the results are collated for analysis. However, the 'no training' is not necessarily correct in the sense that researchers describe the time and effort required by the subjects to develop definitions for the words; in more recent publications on use of the method, as many as 10 training sessions are used. Clearly, there are wide differences in what is meant by training as concerns these different methods.

The words used to represent sensations are nothing more than labels. There is no reason to assume that these words represent anything beyond that – no causality is inferred. Without an appropriate design study any idea as to the external validity of a specific word is hypothetical, at best. Some researchers suggested that the words represent concepts, and that for a descriptive panel to be effective, concepts must be aligned, i.e., subjects must agree on all the sensations (or attributes that represent those sensations) to be included in a concept if the results are to be useful. Unfortunately this idea of concept alignment remains to be more clearly delineated before one can determine its relevance to the descriptive process. One could propose that the process by which subjects discuss their judgments for an attribute, and the definition for that attribute, represent concept alignment, which is an integral part of the QDA training process. Whether this, in fact, is concept alignment remains to be demonstrated; however, it is clear that most subjects can reach sufficient agreement on attributes and can reliably differentiate amongst products, after completion of training. How attributes are formulated in the brain and the true meaning of those attributes are issues that go well beyond descriptive analysis and sensory evaluation, in general.

One should remember that subjects are not likely to agree totally on all the sensations to be included, any more than there is agreement on all the attributes. The sensations are themselves complex and interactive, leading to multiple words to represent them. The individuality of each subject (sensitivity, motivation, and personality) further complicates or adds to the complexity of the process. As a result, a descriptive panel typically develops many more attributes (30 or 40 or more) than will be necessary to describe an array of products fully. The fact that there are many more attributes than are needed should not be unexpected or of concern. This reflects the uniqueness of the individual and the inherent imperfections of the perceptual process, or at least the ability to verbalize more precisely and/or capture those perceptions.

In addition to a descriptive language and definitions, it may be useful to have references available for training or retraining subjects. Here, too, one finds different opinions as to the types of references and how they are to be used. For example, a comprehensive list of references and how they are to be used may be presented, including their respective intensity scores for scale extremes. Unfortunately, these references are based on commercially available products, all of which are variable in the normal course of production, in addition to the intended changes based on changing technologies, ingredients, and/or market considerations. Over time, references will change which can influence subjects in unanticipated ways, further changing responses to product characteristics. What, then, is the value of such references? They have a role to play in helping subjects relate to a particular sensation that is not easily detected or not easily described. However, references should not introduce any additional sensory interaction or fatigue, or significantly increase training time. In most training (or retraining) situations, the most helpful references are usually a product's raw materials. Of course, there will be situations in which totally unrelated materials will prove helpful to some subjects, and it is the panel leader's responsibility to obtain such materials. There will also be situations in which no reference can be found within a reasonable time period. A panel leader should not delay training just because a reference cannot be found. While most professionals agree that references are helpful, there is no evidence that without them a panel cannot function or that results are unreliable and/or invalid. We have observed that so-called expert languages usually require numerous references, and subjects take considerably longer to learn (this language) than they do a language developed by themselves. This should not be surprising, if one thinks about it. After all, references are themselves a source of variability; they introduce other attributes unrelated to their purpose, they are out of context for the complete product, and increase the potential for sensory interactions. The panel leader must therefore consider their use with appreciation for their value as well as for their limitations, and must decide when and what references will be used. In our experience they are of limited value, for use in the language development and training or retraining activities. In retraining, or when adding new subjects to a panel, they are helpful in enabling these individuals to experience what the other subjects are talking about and possibly to add their comments to the language.

Some methods make great use of references, while others do so on an *ad hoc* basis, i.e., they use them only when they are helpful to the subjects.

Scoring

Most methods today are presented as quantitative, relying on various types of scales. The QDA methodology makes use of a line scale (a graphic rating scale) and the concept of functional measurement to obtain intensity judgments. The scale is a 15-cm line with two vertical lines, each placed 1.27 cm from the scale ends. Below each of the vertical lines are words that designate scale direction and intensity for that attribute. For the scale to be used effectively, the subjects must be provided with ample opportunity to practice scoring products and become comfortable with using the scale and with the range of product sensory characteristics. This latter point is especially important to the successful use of the scale. In the training no attempt is made to require subjects to use the same part of a scale, only that each subject be consistent with him- or herself. Because each product is scored on a repeated-trials basis by each subject, it makes little difference what part of the scale is used to differentiate the products. This use of a single scale is in contrast with those methods, in which more than one type of scale may be used. One type of scale is the difference from reference scale in which an identified product is provided with a designation as to its scale location, and products are scored relative to that standard. This type of scale is popular but is flawed as the reference itself will vary, requiring subjects to alter their perceptions (and responses) in relation to the variability associated with it. Most methods are now using some type of line scale but it is not clear as to whether a context is provided during training.

Design and Analysis

The design and analysis of a descriptive test are equally important for a successful outcome. A descriptive test yields a large sensory database (in comparison with a discrimination or an acceptance test), often yielding 10 000 or more data points. There are usually both univariate and multivariate components, and as such, it permits a wide range of statistical analyses. One of the main features of the QDA methodology was the use of a comprehensive statistical analysis of the data, which represented a significant development for sensory evaluation. With the availability of statistical packages and the power of personal computers, panel leaders have unlimited and low-cost resources, providing an online capability for obtaining means, variance measures, ranks and pairwise correlations, and for factor analysis, multiple regression, cluster analysis, discriminant analysis, etc.

All descriptive methods espouse the use of replication (repeated trials); but only the QDA method appears to use it for all test situations. The extent of replication is a decision that the panel leader makes based on the products, the expected degree of difficulty, and past experience of the panel. For QDA a maximum of four replicates is usually sufficient. With more replication, to six, eight, or more, there may be a decreasing return as the number of replicates increases versus the amount of time required for the subject to complete all the evaluations versus statistical power. Empirically we have not found decision changes as a result of increasing replication beyond four. As the number of products evaluated is increased beyond 10, one has the option of decreasing replication to three, taking advantage of the total number of evaluations. These are panel leader decisions; there are no rules as to how many replicates is necessary. The decision takes into account past test history with the product, expected product variation based on manufacturing experience, and so forth. Experience has shown that three to four replicates are usually sufficient.

The analysis of variance with replication model is a very powerful and robust system for descriptive data and is the core method with QDA; other parametric and nonparametric algorithms are also used. With repeated trials one is more likely to yield significant product differences. In addition, this design provides a basis for measuring and visually examining subject performance on an individual attribute basis. Other descriptive methods neither specify these kinds of analyses or make use of programs with defaults that remove some judgments without the researcher knowing what has happened. One such program, Procrustes Analysis, eliminates data based on a test for outliers; while mathematically correct, it takes no account of whether it is a true perception. Many consider its use as experimental and too risky for everyday use.

All sensory data need to be analyzed in some detail as a basis on which to reach conclusions about product differences. As noted before, the descriptive test uses a limited number of subjects and one must be very confident about their behavior before one initiates any extrapolations or investigations about the underlying structure of the data. Specifically, replication provides a basis for determining the skill of each subject and of the panel which increases the confidence of all product decisions. It is an essential cost that cannot be forfeited without serious compromises in any conclusions that are reached about results.

As noted previously, in QDA, there are parametric and nonparametric procedures to identify subjects and attributes for which sensitivity is reduced and/or

there is a significant interaction. This information has several applications. First it provides the panel leader with specific information about a subject's response behavior for each attribute relative to the panel's. Second, it identifies whether a subject is a significant source of interaction and the type of interaction. Crossover interaction is the more serious type (versus magnitude) and with the appropriate nonparametric procedure, one can easily determine a subject's rank order compared with the panel rank order. This process focuses the panel leader's attention on those subjects who exhibited this response behavior, as well as on that individual's scale use. This kind of information can be displayed numerically as well as graphically, along with other types of subject performance behavior, e.g., discriminant functions. For the panel leader, this information is very useful in determining the quality of the results. However, the value to the user rests with product differentiation, and the easiest way of demonstrating this is through the use of plots such as the one shown in Figure 1, a typical 'spider' plot for some of the mouth-feel attributes. Also included is the LSD (a black bar) for each attribute. The mean intensity score for each attribute is determined from measuring the distance from the center to that point at which the product bisects the line. With today's technology, these results can also be displayed in a variety of ways; for example, displaying only the most discriminating attributes based on a discriminant analysis, and so forth.

As with any readily available resource, statistics are often misunderstood and/or misused, particularly when responses are highly variable or when the panel leader confuses use of some of the multivariate procedures with evidence of the validity of results. In other instances, investigators use factor analysis and/or clustering techniques as a basis for excluding subjects who are not in agreement with the panel or to eliminate attributes that they believe are not being used to differentiate products or are used as

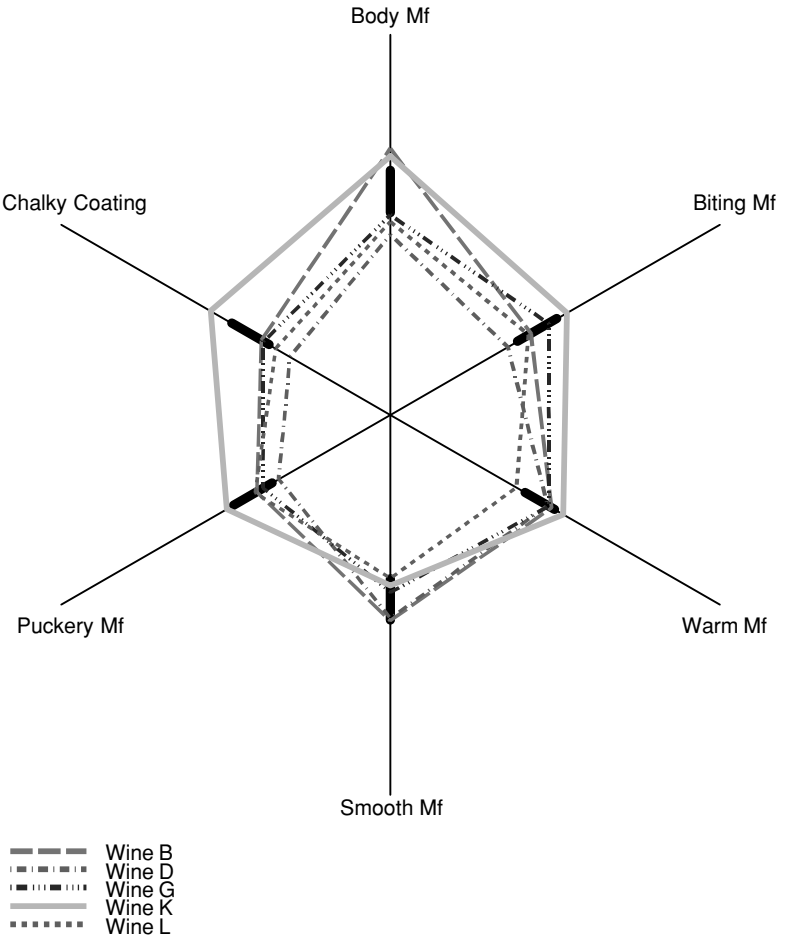

Figure 1 Typical 'spider' plot for selected mouth-feel attributes. Intensity scores are measured from the center out to that point where the product crosses. The attributes are spaced equally around the center point. Bars on the spokes represent the magnitude of difference for that attribute to be significant using the least significant difference (LSD Mf, mouth feel).

substitutes for the same sensation. These have been proposed for use during training to reduce the number of attributes on the score card (from 45 to 12) or the number of subjects. One must be careful about using procedures that, *a priori*, decide which subjects or attributes will be used. After all, the subjects are still in training and to have them score products and use the results as a basis for reducing either, or both, may be premature. This approach can lead to a group of subjects who are more likely to agree with each other, but how does one know that those who agree are correct? How does one differentiate subjects who are discriminators from nondiscriminators if, for example, the nondiscrimination is of 30% or 40% of the attributes? One of several objectives of training is to enable subjects to learn how to evaluate products using the attributes that they determined were helpful to them (at that stage of the training). To proceed from one session to another and substantially reduce the list of attributes is communicating a message that there is a correct list of words and, with a little patience, it will be provided. Using factor analysis to reduce the number of attributes during training is troublesome because it assumes that attributes highly correlated to a factor are measuring the same component of the product and, therefore, one or two attributes can represent the others. However, attributes correlated to a factor are only an indication of some common basis which assigns those attributes to that factor; it does not imply causality. It is interesting to note that one goes to considerable effort to encourage subjects to contribute words to use in a score card and then to devise procedures to eliminate about 75–80% of them before the actual test is performed, thus sending the message to subjects that there is a correct list. As mentioned previously, not all subjects will use all the attributes to the same extent in differentiating products, and eliminating too many of them substantially increases the likelihood of reducing sensitivity and overlooking product differences (an example of type 2 error).

After a test one should use univariate and multivariate procedures to help identify subjects who are experiencing difficulty with specific attributes and/or with use of the scale. This information is then used prior to the start of the next test so as to clarify further the testing process.

Conclusions

Descriptive methods are an established methodology recommended for all sensory programs. Results from a descriptive test provide precise quantitative descriptions of products, and this information has many applications that range from product development, to establishing relationships between ingredients and sensory characteristics, to advertising claims.

In the past decade much attention has been directed to the development of methods that provide the sensory professional with alternatives on how a test is organized and fielded, and how results are analyzed. There are currently two very different approaches described in the literature about the organization and operation of a trained panel. One takes an expert approach, relying primarily on thresholds, and identification of specific flavors and odors (flavor profile) while the other seeks to identify subjects based on their sensory skill with the products of interest. The former takes several weeks, while the latter takes 3–5 days.

For training, one method begins by focusing on single modalities with language obtained from a list and subjects must learn to respond to those attributes, while the QDA methodology requires that subjects develop the list of words, the order within a modality, the definitions, and the evaluation procedure. No standards are used in this training. While the former takes more than 8 weeks, the QDA completes these activities in about 7–10 h for a totally new panel and half or less this time for an experienced panel.

For data collection, QDA recommends a total of four replicates, while the spectrum method indicates that replication is appropriate without further discussion as to how much or in what way the data will be used. It is clear that those who have used these methodologies and those who will use them in the future will make their own modifications.

Data analysis is particularly important for determining the quality of the basic information. While current interest is very high in the use of clustering techniques and other multivariate procedures, such analyses should be considered only after the quality of the database has been assessed.

For the sensory professional there is much to choose from when considering a descriptive test. Whatever methodology is chosen, there are many aspects that require decisions from the professional. It is the authors' contention that these decisions should be influenced by the following considerations:

1. All subjects should be screened for their sensory skill using products from the category being tested.
2. About 30% of the people who volunteer to participate in these screening tests will not meet minimum requirements of at least 50% correct matches (in a discrimination test).
3. Subjects should use words that are derived from common, everyday language to describe products.

4. Subjects should practice scoring products using the afore-mentioned words as part of the training effort.
5. Subjects should develop definitions or, if using a previously developed score card, they should have the opportunity to modify the definitions.
6. Subjects should score each product on a repeated-trials basis with four replicates recommended.
7. Data analysis must provide measures of subject reliability on an attribute basis, means and variance measures for products, and tests for significance.
8. Multivariate procedures should be used where the data warrant the effort.

Sensory evaluation has achieved considerable recognition as a source of unique product information. Much of this derives from the use of descriptive analysis methods. This discussion is intended as a perspective of the development and use of descriptive analysis, as well as a review of the main features of methods currently in use.

See also: **Sensory Evaluation**: Practical Considerations; Sensory Difference Testing; Sensory Rating and Scoring Methods

Further Reading

Boring EG (1950) *A History of Experimental Psychology*, 2nd edn. New York: Appleton.

Cairncross WE and Sjöstrom LB (1950) Flavor profile – a new approach to flavor problems. *Food Technology* 4: 308–311.

Caul JR (1957) The profile method of flavor analysis. *Advances in Food Research* 7: 1–40.

Civille GV and Lawless HJ (1986) The importance of language in describing perceptions. *Journal of Sensory Studies* 1: 203–215.

Clapperton JR, Dalgliesh CE and Meilgaard MC (1975) Progress towards an international system of beer flavor terminology. *Master Brewers Association of America. Technical Quarterly* 12: 273–280.

Harper RM (1972) *Human Senses in Action*. Edinburgh: Churchill Livingstone.

Huitson A (1989) Problems with Procrustes analysis. *Journal of Applied Statistics* 16: 39–45.

Ishii R and O'Mahony M (1990) Group taste concept measurement: verbal and physical definition of the umami taste concept for Japanese and Americans. *Journal of Sensory Studies* 4: 215–227.

Jones FN (1958) Prerequisites for test environment. In: Little AD (ed.) *Flavor Research and Food Acceptance*, pp. 107–111. New York: Van Nostrand Reinhold.

Lawless HT and Heymann H (1998) *Sensory Evaluation: Principles and Practices*. Chapman and Hall, New York.

Lyon BG (1987) Development of chicken flavor descriptive attribute terms aided by multivariate statistical procedures. *Journal of Sensory Studies* 1: 99–104.

Mackey AO and Jones P (1954) Selection of members of a food tasting panel: discernment of primary tastes in water solution compared with judging ability for foods. *Food Technology* 8: 527–530.

Marshall RJ and Kirby SPJ (1988) Sensory measurement of food texture by free-choice profiling. *Journal of Sensory Studies* 3: 63–80.

Marx MH and Hillix WA (1963) *Systems and Theories in Psychology*. New York: McGraw-Hill.

Meilgaard M, Civille GV and Carr BT (1991) *Sensory Evaluation Techniques*. Boca Raton, Florida: CRC Press.

Muñoz AM (1986) Development and application of texture reference scales. *Journal of Sensory Studies* 1: 55–83.

O'Mahony M, Rothman L, Ellison T, Shaw D and Bureau L (1990) Taste descriptive analysis: concept formation, alignment and appropriateness. *Journal of Sensory Studies* 5: 71–103.

Powers JJ (1988) Current practices and applications of descriptive methods. In: Piggott JR (ed.) *Sensory Analysis of Foods*, 2nd edn, pp. 187–266. London: Elsevier Applied Science.

Rainey BA (1986) Importance of reference standards in training panelists. *Journal of Sensory Studies* 1: 149–154.

Smith GL (1988) Statistical analysis of sensory data. In: Piggott JR (ed.) Sensory *Analysis of Foods*, 2nd edn, pp. 335–379. Essex: Elsevier Science Publishers.

Stone H (1999) Sensory evaluation: science and mythology. *Food Technology* 53: 124.

Stone H and Sidel JL (1992) *Sensory Evaluation Practices*, 2nd edn. Orlando, Florida: Academic Press.

Stone H and Sidel JL (1998) Quantitative descriptive analysis: developments, applications, and the future. *Food Technology* 52: 48–52.

Stone H, Sidel J, Oliver S, Woolsey A and Singleton RC (1974) Sensory evaluation by quantitative descriptive analysis. *Food Technology* 28: 24.

Williams AA and Arnold GM (1985) A comparison of the aromas of six coffees characterized by conventional profiling, free-choice profiling and similarity scaling methods. *Journal of the Science of Food and Agriculture* 36: 204–214.

Williams AA and Langron SP (1984) The use of free-choice profiling for the evaluation of commercial ports. *Journal of the Science of Food and Agriculture* 35: 558–568.

Appearance

D B MacDougall, The University of Reading, Reading, UK

Introduction

Of the five human senses, vision has had probably the longest research into its mechanisms and is the best understood of the five. Human eyes have a vast ability

to detect and process external information for the brain to translate instantaneously into the perceptions of appearance and color. The appearance of any object or scene can be defined as the response of the observer to the totality of information that is perceived by the eyes. Because the mechanism of visual perception is complex, it can be divided into several interacting components, which, when taken together, are usually described as total appearance. Undoubtedly, the most important of these components is the perception of color. Most of this article will deal with the way human visual signals are perceived and how color is measured. However, a broader classification of the components that comprise total appearance would divide them into the physical and chemical nature of the object and the influences of the illumination and surroundings on the object's appearance. Additionally, for food, it must also include the social and psychological factors associated with the consumer such as culture, food habits and preference. The judgment of food appearance at the point of purchase is related to the potential satisfaction of the product coupled with its value. Such decisions are usually made quickly with the criterion that rejection may be more related to the presence of faults in the food's appearance than to what is regarded as normally acceptable. The success of the food industry in making consistent and attractive products depends on the consumer's continuing confidence that the appearance of the product is a true indicator of the subsequent acceptability of the eating quality of the product. That is, the appearance and color of the product are likely to be related in the consumer's mind with a value judgment of the other sensory attributes and their expected satisfaction.

The Visual Process

Vision is the property of the response of the eye to the stimulus of light, as interpreted by the brain. Human vision is binocular, i.e., objects are seen as solids in perspective space. The appearance of any object in a scene can be defined as the recognition of the properties of that object, i.e., size, geometry, solidity, surface characteristics, and subsurface optical properties, i.e., transmission, opacity, or degree of translucence and, most importantly, color. The visual process leading to the recognition of appearance, as with all sensory events, consists of a rapidly occurring sequence of steps such that the observer sees the response to the stimulus near instantaneously. Three factors are involved in vision: sufficient light to illuminate the object, human eyes with their complex detecting mechanism and the object itself with its interaction with the incident light and the background surroundings. Human eyes are virtually spherical, approximately 2 cm in diameter, with six muscles to control their mobility to give each eye a near circular field of view. Light is detected after entering the iris, which self-adjusts for light intensity. It is then focused through the eye's lens on to the retina with its two types of light-sensitive cells, the rods and cones. This results in a distinct image of the object being focused on to the retina. Most of the retina is sensitive to low levels of light, detected by the rods whose response is noncolored and is known as 'scotoptic vision.' The cones located in a small spot, approximately 2 mm in diameter, centrally located in the retina and known as the fovia, detect photoptic or color vision. In a person with normal color vision, i.e., one who is not color blind, there are three types of cones whose light-detecting ability is wavelength-dependent. The mechanism of light detection in the photoreceptor cells is by a conformational change in the photopigment molecule rhodopsin, producing an electrical potential for amplification and transmission via a nerve fiber to the brain. It is important to emphasize that the perception of an object's color and appearance exists in the observer's brain and not in the object itself.

The wavelengths of maximum sensitivity of the so-called blue (B) detecting cones is at 440 nm, that of the green cones (G) at 540 nm and for the red cones (R) it is at 600 nm. The G and R detecting bands overlap equally at 580 nm, producing that narrow band in the visible spectrum that appears bright yellow. The perception of any color is therefore due to the responses generated by three signals varying in absolute and relative intensity and is known as trichromatic or three-colored vision. Thus, because cone vision is trichromatic, a suitable mixture of monochromatic B, G, and R primary lights can match any colored light. Subsequent processing of the trichromatic signals by the brain gives rise to opponent red/green and blue/yellow color perception mechanisms and an achromatic lightness/darkness mechanism.

The human eye–brain relationship has the ability to adjust to changes in illumination without much of an apparent change in the color of the object. This process is called adaptation and is accompanied with the phenomena of color constancy, where colors remain essentially, but not truly identically, the same hue when the spectrum of the incident light changes. Hence, a colored product under a blue white light, such as occluded daylight with a high component of blue in the spectrum, will appear to possess a similar color as the same product under orange–yellow tungsten illumination. The reason for this is that, in most situations, white is normally a component of the

scene, and the eye adjusts to recognize white as white, whatever the light source might be.

Orderly Color Arrangement

Many attempts have been made to arrange colors in systematic order (Figure 1). Humans have the ability to remember between 1000 and 2000 colors, many of which can be identified by name, and most color atlases have more than 1000 distinctly different colors in them. The most obvious arrangement is that of hue, i.e., the difference between red, orange, and yellow, etc. to blue following the order of the colors in the spectrum. This is an insufficient separation, because colors can also be arranged in order of their lightness and saturation or intensity. This is the basis of two of the most commonly used color atlases, that of Munsell (Figure 2) and the more recent Natural Color System (NCS). They both have an achromatic lightness to darkness vertical axis around which the hues are visually separated and radiate from the central axis in increasing intensity

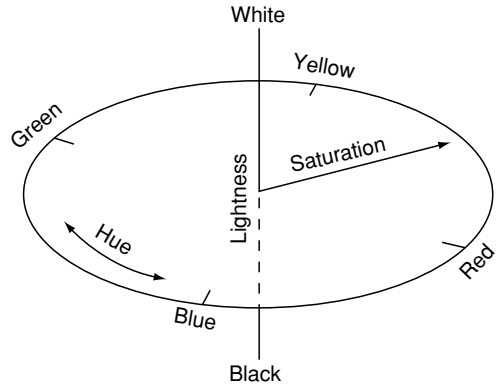

Figure 1 Three-dimensional color order system.

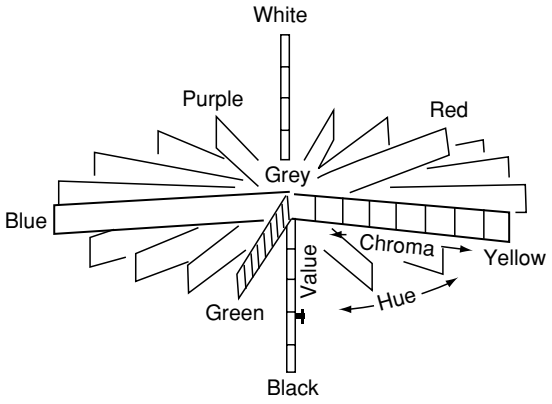

Figure 2 Munsell color system.

(saturation or chroma), but they are conceptually different. The NCS atlas is based on the fact that there are four uniquely distinct hue perceptions of red, yellow, green, and blue although derived from only the three tristimulus responses. Thus, the hue red is opposite green, yellow is opposite blue, and the four resulting quadrants of the space are divided, in the NCS atlas, into 10 steps. Thus, any color can be defined by its lightness, hue angle, and chroma values. The hue arrangement in the Munsell system was based on visual hue separation. Both atlases enjoy wide use as reference in the food industry.

CIE Color Spaces

Both the above atlases, and others like the Pantone Color Specifier, are based on visual matching of pigments on paper. For instrumental measurement, account must be taken of both the nature of the transmission or reflectance spectrum of the product and that of the incident light. The ability of an observer with normal color vision, under strictly controlled conditions, to match the color of any wavelength with a mixture of red, green, and blue wavelengths resulted in the construction of the 1931 CIE (Commission Internationale de l'Eclairage) color space. Color-matching functions of the responses to mixtures of three monochromatic wavelengths in the red, green, and blue regions were used to match the entire spectrum wavelength at a time. This resulted in what is known as the CIE 1931 2° visual field of view standard observer and is the basis of all instrumental color-measurement and color-match prediction. In 1964, a supplementary set of functions was defined for a wide field observer of 4°, because, for most practical purposes in real-life situations, samples are viewed at a wider visual angle than 2°. The functions used in the calculation of the tristimulus values X, Y, Z have the advantage that, unlike the original R, G, B matching, in which some of the matches are negative, they are all positive with Y equal to the luminance or lightness of the color. The relationship of X, Y, Z to R, G, B is:

$$X = 0.49R + 0.31G + 0.20B$$
$$Y = 0.17697R + 0.81240G + 0.0106B$$
$$Z = 0.00R + 0.01G + 0.99B.$$

Thus, the values of any set of CIE tristimulus values X, Y, and Z will define color in three-dimensional space. Although Y contains the entire lightness or luminance response, X and Z are more difficult to relate to color perception directly. The concept of chromaticity, correlating to the visual recognition of color perception of hue and color purity can be derived from the coordinates x and y where:

$$x = X/(X + Y + Z) \text{ and } y = Y/(X + Y + Z).$$

Therefore z, which equals $(1-(x + y))$, is not required in the construction of the space. The CIE Color Matching Functions and the 1931 chromaticity diagram are shown in Figures 3 and 4.

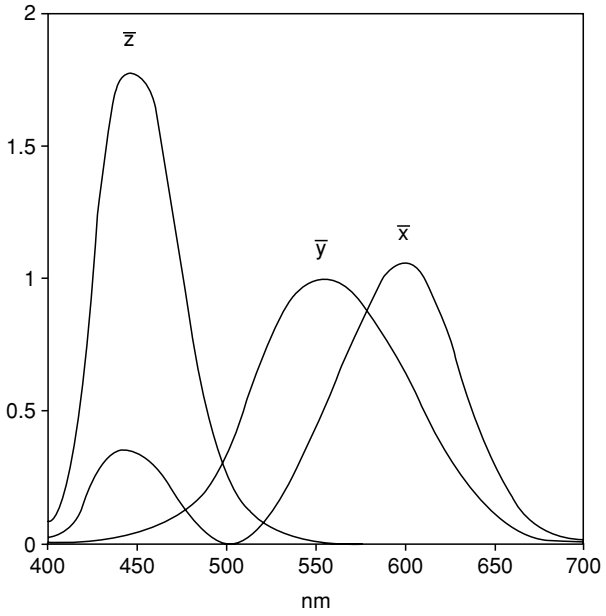

Figure 3 CIE color-matching functions for the 2° Standard Observer.

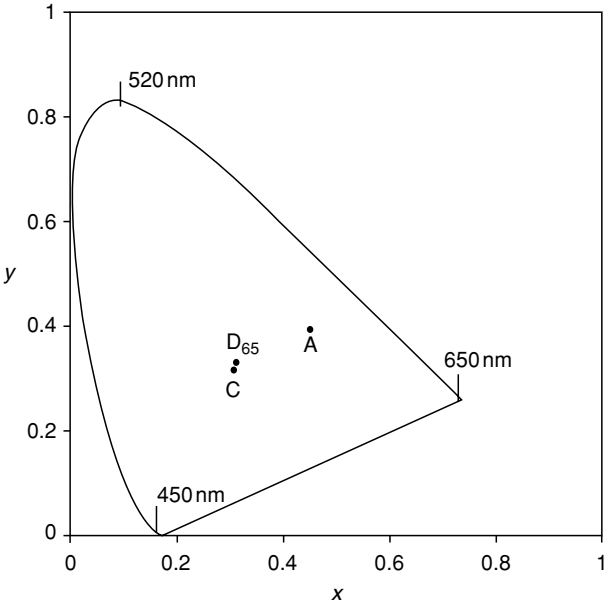

Figure 4 CIE x, y chromaticity diagram showing location of illuminants A, C, and D_{65}.

Uniform Color Spaces

Although the development of the 1931 CIE color space provided a means whereby any color could be defined unambiguously, it has the major disadvantage that the colors in the space are not equally separated from one another visually. For example, the chromaticity distance between just noticeable greens is several times greater than the equivalent distance between just noticeable blues. The objective of any color-measuring system is not only to provide an unambiguous definition of color but also to obtain a realistic relationship to visual perception of color differences. Two near-uniform color spaces that have importance to industry are the Hunter and CIELAB spaces. The Hunter L, a, b opponent color space (Figure 5) was a development of the earlier Adams chromatic value systems devised in the mid-1940s, where red is opposite to green, and yellow is opposite to blue, i.e., the concept of the four unique hues. The Hunter L, a, b was constructed in 1958 primarily for use in color-measuring instruments with analog devices. The formulae for transformation of the tristimulus values to L, a, b are:

$$L = 10Y^{1/2}$$
$$a = 17.5(1.02X - Y)/Y^{1/2}$$
$$b = 7.0(Y - 0.847Z)/Y^{1/2}.$$

The use of the square root of Y resulted in a color space that was readily interpretable and approached visual uniformity when compared with the spacing of the Munsell color atlas with its uniform and orderly arrangement of real colors. This space has had wide application in the food industry but is now being superseded by the 1976 CIELAB L^*, a^*, b^* color space, where L^*, the cube root of Y, is more uniformly spaced than L. The formulae are:

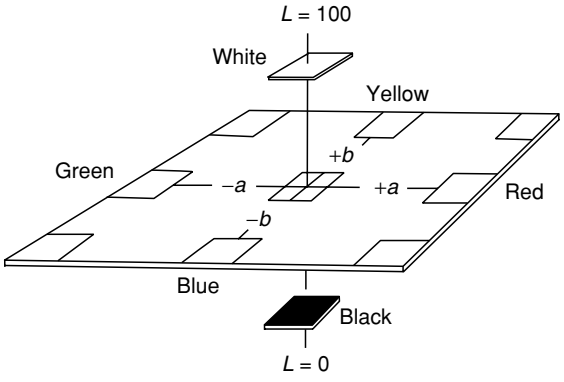

Figure 5 Hunter three-dimensional uniform L, a, b color system.

$$L^* = 116(Y/Y_n)^{1/3} - 16$$
$$a^* = 500[(X/X_n)^{1/3} - (Y/Y_n)^{1/3}]$$
$$b^* = 200[(Y/Y_n)^{1/3} - (Z/Z_n)^{1/3}],$$

where X_n, Y_n, and Z_n refer to the nominally white object color stimulus. Although this scale is similarly constructed to present lightness and opponent colors with distances that are more nearly inform, it has the additional advantage of being related to real illuminants rather than the L, a, b space, which used the theoretical light source C. The three most common illuminants used in the calculation of L^*, a^*, b^* are standards A, C and D_{65}. A is the spectrum of a tungsten filament lamp with a color temperature of 2856 K, and C is standard A with a blue filter to stimulate average daylight with a color temperature of 6774 K. Illuminant D_{65} is the spectrum of the color of daylight with a correlated color temperature of 6504 K. Currently, the tendency is to use the 10° standard observer and illuminant D_{65} rather than the 2° observer and C. Many modern color spectrophotometers also have the facility of additional illuminants in their programs, e.g., a series of typical fluorescent lamps of both wide and narrow emission spectra.

Food Appearance

The most important attribute of any food's appearance is its color, especially when it is directly associated with other food-quality attributes, for example the changes that take place during the ripening of fruit or the loss in color quality as food spoils or becomes stale. Every raw food and manufactured product has an acceptable range of color appearance that depends on factors associated with the consumer and the nature of the surroundings at time of judgment and the structure and pigmentation of the food itself. However, color specification alone is insufficient to define food appearance. The color quality of the illumination, in terms of intensity, color temperature and fidelity and the nature of the structure of the product all affect the appearance. The distribution of surface reflectance, the nature of internally scattered light, and the pigmentation of the product are all necessary for a complete specification of appearance. Most important is the interaction of the food's structure with its variable light scatter and pigmentation, which affect opacity and translucency as well as the color. Small changes in scatter can produce greater changes in visual color appearance than are attributable to the normal range in pigment concentration in some products. It is important to distinguish between these, but this is not always carried out during color measurement with the consequence that erroneous interpretation of the data commonly results.

A food's appearance therefore can be reduced to two principal factors, the physical and the psychological. The physical factors consist of the geometrical, the food's dimensions of size, shape and intrinsic characteristic variability in uniformity and mass, and the optical, surface gloss or dullness, the nature and degree of pigmentation, and the light-scattering power of the food's structure. The objective, therefore, is to convert the physical to the psychological by translating the object's reflectance or transmittance spectrum into the tristimulus values and then to a defined color space.

Food Color Measurement

As already discussed, foods have an infinite variety of appearance characteristics. Their surfaces may be diffuse, glossy, irregular, or flat. They may be transparent, hazy, translucent, or opaque, and their color may be uniform or patchy. Hence, color-measuring procedures for foods often have to be modified from the more standardized techniques used in the measurement of flat opaque surfaces such as paint and paper for which most color-measuring instruments are designed. Different instrument optical geometries will lead to difficulties in sample presentation, producing different color values for the same material. The inclusion or exclusion of surface specular reflection in the measurement depends not only on its importance as a characteristic of the food but also on the design of the detector-sensor unit in the instrument. Lateral transmittance of light through translucent materials affects their reflectance, and this should be allowed for in the assessment of such products as tomato paste and fruit juice, because the ratio of absorption to scatter varies with aperture area and the concentration of the components in the product. Colorimeters and spectrophotometers do not necessarily produce the same result, because of differences in their construction. Some spectrophotometers give the option of including or excluding the specular component of reflectance from the measurement. It has been the experience of this author to include the component unless the surface is flat when the difference between the two readings is likely to be related to the degree of gloss of the surface. Most colorimeters, unlike spectrophotometers which have measuring geometries of 8°/diffuse or diffuse/8°, exclude the specular by measuring at 45°/0°.

Measuring at a defined sample thickness is recommended for consistent data. Usually, this would be at a so-called infinite (∞) thickness, where an increase in thickness would not alter the reading. However, for

translucent samples, it is advisable to measure at more than one thickness on white and black background to distinguish between the light-scattering component of reflectance and pigment absorption. The technique of Kubelka and Munk, used by the paint and paper industry, is one method. This relates the ratio of an absorption coefficient (K) over a scatter coefficient (S) to reflectance at an infinite thickness thus:

$$K/S = (1 - R_\infty)^2/2R_\infty.$$

It is possible to calculate K and S separately and also the internal transmittance (T_i) from the differences in the readings of the white-backed and black-backed thin layers.

Measurement Procedures

Because foods are near infinitely variable in their appearance, it means that there is no consistent way of measuring them. With few exceptions, e.g., the measurement of tomato paste, there are no defined methods, although a variety of approaches have been found to be successful. For transparent liquids, there is no difficulty, but care has to be taken to define the concentration or series of concentrations that are measured. It can be advantageous to measure a series of dilutions and relate these to their change in the L^*, a^*, b^* diagram as well as the normal technique of absorbance measurement at fixed wavelengths. For solid materials, reflectance measurements are necessary, and the appearance variables encountered fall into three classes as follows:

1. The truly flat and opaque variables are easy to measure, and the procedure is to measure the sample at infinite thickness. Infinite thickness is defined as the thickness that, if it were increased, would not lead to a change in the data.
2. The translucent variables are less easy to measure, because the incident light falling on the material when viewed by the eye produces an image that is somewhat different to that presented to the instrument, because the effect of lateral light scatter may not be left out of the measurement. As well as measuring at infinite thickness, it is recommended that the sample be measured as thin sections on top of black and white from which K, S, and T_i can be calculated.
3. The particulate variables have the difficulty of holes and gaps in the material, e.g., the surface of many breakfast cereals is composed of both the cereal and spaces, which, when judged visually, are not included in the perception of color but are included when measured on an instrument. One way of alleviating the problem is to grind and filter the sample to a uniform particle size range and then measure the ground material packed into a cell.

Visual Assessment

Instrumental color measurement is only valid if it accurately describes the visual characteristics of the food. The quality of incident light alters the perception of the food's color. Light boxes used for color judgment are now equipped with several light sources fitted with color-matching lamps and also those typical of those used in industry. Foods are usually viewed in the supermarket under 'warm' appearing lighting, because the enhanced red in the spectrum aids attractiveness. In the home the most likely light source is the tungsten lamp with its yellow–orange illumination. However, most color data are reordered relative to D_{65} or C, and it should be appreciated that if accurate perceptual equivalence is required from the measured color data, it should also be measured to source A or to the requisite fluorescent lamp.

Realistic Limitations

In some industries, particularly the color-matching industries such as the paint, plastics and printing industries, the objective of color appearance measurements is to obtain data that are accurately related to existing colors, e.g., those in an atlas or some internal standard. This may not be required for specific foods. Control procedures for quality assurance may only demand that some critical aspect of appearance be measured. For some products, a single figure quality color scale might be appropriate, e.g., one of the CIE yellowness indexes. In recent years, miniaturization in instrument manufacture has led to a greater availability of hand-held colorimeters and spectrophotometers. The use of these instruments, even with their limitations, in the measurement of the color of food is likely to increase with commensurate improvements in assurance of consistent product quality to the consumer.

See also: **Analysis of Food**; **Browning**: Nonenzymatic; Enzymatic – Biochemical Aspects; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; Properties and Determinants of Synthetic Pigments; **Quality Assurance and Quality Control**; **Sensory Evaluation**: Sensory Characteristics of Human Foods; Food Acceptability and Sensory Evaluation; Appearance; **Spectroscopy**: Visible Spectroscopy and Colorimetry

Further Reading

ASTM (2000) *ASTM Standards on Color and Appearance Measurement*, 6th edn. West Conshohocken, PA: American Society for Testing and Materials.

BSI/ISO (1999) *BS 5929-10:1999, ISO 11037: 1999 Methods for Sensory Analysis of Food. General Guidance and Test Method for Assessment of the Colour of Foods*. London: British Standards Institution.

CIE (1986) *Publication Commission Internationale de l'Eclairage no. 15.2 Colorimetery*, 2nd edn. Vienna: CIE Central Bureau.

Francis FJ and Clydesdale FM (1975) *Food Colorimetry: Theory and Application*. Westport, CT: AVI.

Hunt RWG (1998) *Measuring Colour*, 3rd edn. Surbiton, UK: Fountain Press.

Hunter RS and Harold RW (1987) *The Measurement of Appearance*, 2nd edn. New York: John Wiley.

Hutchings JB (1999) *Food Colour and Appearance*, 2nd edn. London: Lippincott Williams & Wilkins.

Mackinney G and Little AC (1962) *Color of Foods*. Westport, CT: AVI.

McDonald R (1997) *Colour Physics for Industry*, 2nd edn. Bradford, UK: Society of Dyers and Colourists.

MacDougall DB (2002) *Colour in Food – Improving Quality*. Cambridge: Woodhead Publishing Co. Ltd.

Texture

M C Bourne, New York State Agricultural Experiment Station, Cornell University, Geneva, USA
A S Szczesniak, New York, NY, USA

A Quality Factor

Texture is a sensory property of foods which, together with appearance, aroma, and basic taste, has a profound effect on consumer acceptance of foods. Each of these properties is composed of a number of notes. Texture lies between taste and aroma in this respect, several dozen different texture notes being detectable in foods, which is more than the four taste notes of sweet, sour, salt, bitter, and less than the hundreds of odor notes that can be recognized.

Textural perception occurs directly through the tactile (touch) and kinesthetic (movement) senses, and indirectly through the senses of vision and hearing. In contrast to color and flavor, there are no specific sensory receptors for texture. Texture is an important quality attribute in almost all foods and most important in foods that are bland in flavor, or have the characteristics of crispness or crunchiness. It contributes to the satisfaction of chewing and the pleasure of eating. Good textures are indicators of excellent food preparation. Highly valued is textural contrast, i.e., the presence of two or more contrasting, but compatible, characteristics that can occur within a meal, on the plate, or within a food product.

Texture may be defined as 'that group of physical characteristics that arise from the structural elements of the food, are sensed primarily by the feeling of touch, are related to the deformation, disintegration, and flow of the food under a force, and are measured objectively by functions of mass, time, and length.' This definition teaches that texture has its roots in structure (molecular, microscopic, macroscopic) and the manner in which this structure reacts to applied forces. It also emphasizes that texture is a multidimensional property comprising a number of sensory characteristics.

A large number of terms are popularly used to describe textural sensations. **Table 1** organizes many of these terms into a manageable system that facilitates understanding their interrelationships. It classifies textural properties into mechanical characteristics (reaction of the food to stress), geometrical characteristics (the feeling of the size, shape, and arrangements of particles in the food, sometimes called 'particulate properties'), and other characteristics (relating to the sensations of moisture, fat, and oil in the mouth). **Table 2** provides physical and sensory definitions of the mechanical characteristics.

Sensory Evaluation

Since, by definition, texture is a sensory property, the most logical approach to its description and quantification is by sensory evaluation. In the early days, panels with various degrees of training were used to score specific textural characteristics or 'texture' in general. The scoring methods used were either numerical intensity scales (frequently 0–7, with 0 denoting absence and 7 a very high intensity of a specific characteristic) or hedonic scales (ranging from 'dislike extremely' to 'like extremely'). The latter should not be used when the objective is to describe texture in terms of the characteristics present or to quantify their intensity.

The sensory perception of texture is a dynamic process which involves the rate and magnitude of the applied forces, and also the effects of temperature, saliva, and time. The time element includes the repeated application of destructive forces in the masticatory process, and the duration of the food's contact with saliva and mouth temperature.

The multiparameter nature of texture and the process dynamics of its sensory perception during mastication form the basis for the sensory texture profile,

Table 1 Classification of textural characteristics and their relationship to popular nomenclature

Characteristics	*Primary parameters*	*Secondary parameter*	*Popular terms*
Mechanical	Hardness		Soft, firm, hard
	Cohesiveness	Brittleness	Crumbly, crunchy, brittle
		Chewiness	Tender, chewy, tough
		Gumminess	Short, mealy, pasty, gummy
	Viscosity		Thin, thick
	Springiness		Plastic, elastic
	Adhesiveness		Sticky, tacky, gooey
Geometrical	**Class**		**Examples**
	Particle size and shape		Gritty, grainy, coarse, etc.
	Particle shape and orientation		Fibrous, cellular, crystalline, etc.
Other	Moisture content		Dry, moist, wet, juicy
	Fat content	Oiliness	Oily
		Greasiness	Greasy

Adapted from Szczesniak AS (1963) Classification of textural characteristics. *Journal of Food Science* 28: 385–389, with permission.

Table 2 Definitions of the mechanical parameters of texture

	Physical	*Sensory*
Hardness	Force necessary to attain a given deformation	Force required to compress a substance between molar teeth (solids) or the tongue and palate (semisolids)
Cohesiveness	Strength of internal bonds	Amount of sample deformation before rupture when biting with molars
Fracturability	Forces necessary to fracture the material	Force with which the material crumbles, cracks, or shatters
Chewiness	Energy required to disintegrate a solid food to a state ready for swallowing	Number of chews required to masticate a sample at one chew per second and constant rate of force application to reduce it to a consistency suitable for swallowing
Gumminess	Energy required to disintegrate a semisolid food to a state ready for swallowing	Denseness that persists throughout mastication of a semisolid food
Viscosity	Rate of flow per unit force	Force required to draw a liquid from a spoon over the tongue
Springiness	Rate at which a deformed material goes back to its undeformed condition following removal of the deforming force	Degree and speed with which the material returns to its original height following partial compression with molar teeth
Adhesiveness	Work necessary to overcome the attractive forces between the surface of the food and other surfaces with which the food comes in contact	Force of the tongue required to remove the material that adheres to the mouth (generally the palate, but also lips, teeth. etc.) during the normal eating process

the schematic for which is shown in Figure 1. The method is used to define the texture characteristics present, the intensity of each, the order in which they appear, and all the changes that occur from the first bite through completion of mastication. Texture profile analysis is presently the preferred sensory method for texture characterization because it is the only method that provides a complete analysis of all the textural properties of a food. Using reference samples and standard scales for specific parameters, highly trained panels provide a descriptive and quantitative 'fingerprint' of the product's texture. The training and maintenance of the panel may be tedious and expensive, but the quality of the generated data usually compensates for it. The basic principles can be adapted to different products in different situations including untrained consumer panels. Two examples of sensory texture profiling are shown in Table 3.

Most other sensory methods can be regarded as partial texture profiles or modifications of the basic procedure. The two main methods are: Quantitative Descriptive Analysis (QDA) and Spectrum. QDA differs from texture profiling in that it uses unstructured line scales for quantifying the intensity of texture notes present, and the data are generated by averaging the ratings given by the individual panelists rather than through panel discussion and a consensus. Another difference is that the panelists generate their own sets of terms. In contrast, the Spectrum method uses an expanded set of terms within the standard lexicon of texture profiling. It employs 15-point intensity scales anchored by reference samples. The extensive panel training and the active role

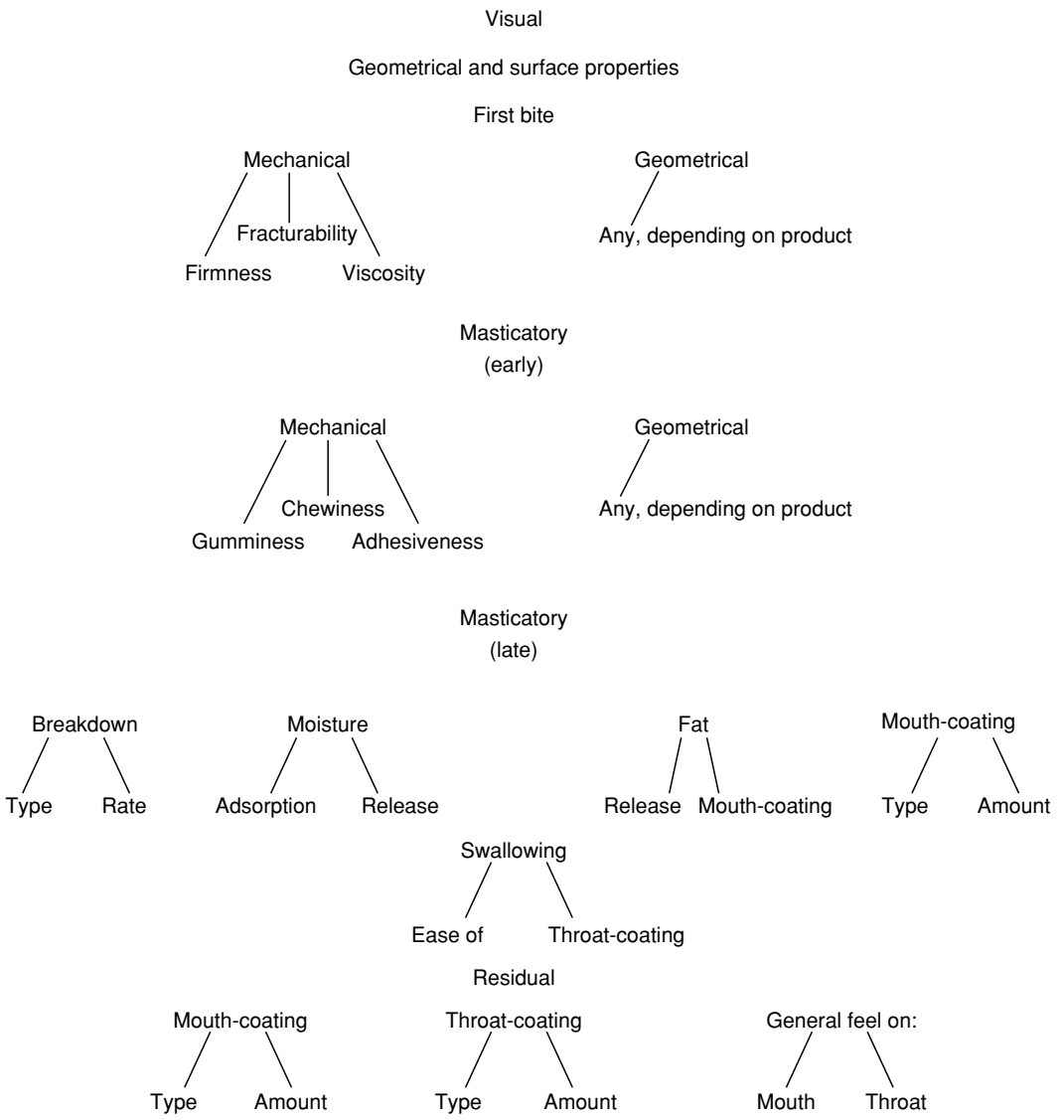

Figure 1 Sensory texture profiling technique. Reproduced from Sensory Evaluation: Texture, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press, with permission.

of the panel leader are similar to those of texture profiling.

Instrumental Evaluation

Although sensory methods are the final arbiter of textural quality, instrumental methods are the most frequently used for texture measurement because they are cheaper and less time-consuming. To be successful, any instrumental measurement must correlate highly with the sensory evaluation. It needs to be recognized that instrumental methods measure one or more textural properties of a food before it is placed in the mouth, but cannot follow the changes that occur during mastication. Most instrumental methods are 'one-point' measurements, i.e., they measure only one dimension of the texture, albeit usually a dominant textural property. Because there is a wide range of types of foods and textures, people use a wide variety of methods to assess texture sensorially. Before they put food into the mouth they use their fingers and hand to squeeze, stir, bend, or puncture the food. After the food is put into the mouth the chewing rate, the amount of lateral movement of the mandible (jaw), and the movements of the tongue are automatically adjusted to handle that particular food in the most convenient way. A large number of texture-testing instruments have been described in the literature and almost 100 are available commercially. Table 4 classifies objective methods of

Table 3 Basic sensory texture profile ballot for meatballs and soda crackers

Texture notes	*Meatballs*	*Soda crackers*
Initial		
Mechanical		
Hardness (9-point scale)	3.4	4.0
Fracturability (7-point scale)	0.7	2.5
Viscosity (8-point scale)	Not applicable	Not applicable
Geometrical	Lumps, with a grainy surface	Flaky and puffy
Other	Moist, uncut surface is slippery and cut surface is not slippery	Dry
Masticatory		
Mechanical		
Gumminess (6-point scale)	1.2	0
Chewiness	17.7 chews	16 chews
Adhesiveness (5-point scale)	1.2	0.7
Geometrical	Coarse, grainy; some fibrous particles are present	Flaky
Other	Moist	Dry
Residual		
Rate of breakdown	Large lumps break down fast; grains break down at a medium rate	High
Type of breakdown	Lumps turn into a nonhomogeneous, grainy paste, and grain size decreases. Some stringly fibrous grains are present; they become more noticeable towards the end and require more effort to chew	It breaks down into little rough sheets, then changes into a smooth dough
Moisture absorption	Initially moist. Saliva mixes easily with slurry and the bolus becomes progressively more moist. Residual grains feel dry	It absorbs a lot of saliva slowly and change into a moist dough
Mouth-coating	Slight residual oiliness. A few particles stick between the teeth and around the mouth	Little pieces stick to the mouth and gums

From Bourne MC (1982) *Food Texture and Viscosity: Concept and Measurement*. New York: Academic Press, with permission.

Table 4 Objective methods for measuring food texture

Principle	*Measured variable*	*Dimensional units*	*Examples*
Force	Force (F)	mlt^{-2}	
Puncture	F	mlt^{-2}	Fruit pressure testers
Extrusion	F	mlt^{-2}	Shear press, Tenderometer
Shear	F	mlt^{-2}	Warner– Bratzler shear
Crushing	F	mlt^{-2}	
Tensile	F	mlt^{-2}	Extensograph
Torque	F	mlt^{-2}	Rotary viscometers, Farinograph, Struct-o-Graph
Snapping	F	mlt^{-2}	
Deformation	F	mlt^{-2}	
Distance			
	Length	l	Penetrometer, Bostwick Consistometer
	Area	l^2	USDA Consistometer
	Volume	l^3	Bread volume, Succulometer
Time	Time (T)	t	Ostwald Viscometer, Biscuit Texture Meter
Energy	Work ($F \times D$)	$ml^{-2}\,t^{-2}$	Area under force–distance curves
Ratio	F or D or T measured twice	Dimensionless	Specific Gravity
Multiple	F and D and T	mlt^{-2}, l, t	Instron, Lloyd, Zwick, TA.XT2
Chemical analysis	Concentration	Dimensionless (%)	Alcohol-insoluble solids
Miscellaneous	Anything	Anything	Optical density, crushing sounds

Adapted from Bourne MC (1996) A classification of objective methods for measuring texture and consistency of foods. *Journal of Food Science* 31: 1011–1015, with permission.

food texture measurement on the basis of the variable(s) measured and the principle of the test. Force-measuring instruments are the most common, but other principles, such as distance, time, and energy, are also used. There are a few examples where a chemical analysis correlates well with a textural prop-

erty. Finally, the sounds that are generated are an important dimension of the textural quality of crisp and crunchy foods. Characteristic sounds are generated when these foods are compressed. The sounds are of low frequency and come in a series of sharp bursts, that occur in rapid succession. Crunchy foods exhibit notes of a lower frequency than crisp foods.

The key element that differentiates force-measuring instruments is the geometry of the test cell which holds the sample and applies a force to it (cutting, puncture, compression, extrusion, etc.). Recognition of the above fact has led to the widespread use of strength-of-materials-testing machines that provide a sophisticated driving mechanism, force sensor, and recording of the force–time relationship. They are more expensive than the simple instruments that have been in use for many decades, and they came into use in the late 1960s. They are called universal testing machines (UTMs) because, with appropriate test cells attached, all test procedures that require rectilinear movement can be performed on the same machine. With the advent of computer retrieval and analysis of the force–time data their use is rapidly becoming routine for quality purposes also. One UTM, the TA.XT2 texture analyzer, has been expressly developed for measuring textural properties of foods. Its operation is so simple, yet precise, that it is now widely used for both research and quality control purposes.

The following criteria are recommended as a guide to selecting a texture-testing instrument or the type of test to program into an UTM.

- Purpose – research or quality assurance.
- Nature of product – rheological type, heterogeneity.
- Required accuracy – high variability inherent in unprocessed foods often necessitates many replications.
- Cost – including operation and maintenance.
- Time – routine use requires rapid testing.
- Location – ability to withstand adverse environmental conditions when placed in factory surroundings.
- Nature of sensory evaluation method used by people (squeeze in hand, cut with incisors, crush between molars, roll with the tongue against the hard palate, etc.)

This should narrow the field to the most promising two or three test principles. The final candidates should be tested over the full range of textures normally encountered with the food and correlated with sensory evaluation. A statistical analysis of the results should identify which principle and instrument are the best for each particular application. The final step is to establish the test conditions that give the strongest resolution between different samples and then standardize them. These include sample size, test cell dimensions, force range, speed of travel of moving parts, chart speed, temperature, and perhaps other factors.

Texture Profile Analysis

The methodology of texture profile analysis (TPA) brings instrumental texture evaluation a step closer to sensory testing. It involves compressing a bite-size piece of food two or more times in a reciprocating motion that simulates the action of the jaw, and quantifying from the resulting force–time curve a number of textural parameters that correlate well with sensory evaluation. The method, originally developed for the General Foods Texturometer, has been adapted to universal testing machines. Figure 2 shows a generalized TPA force–time curve. Hardness is defined as the peak force on the first compression cycle (first bite); fracturability is defined as the first significant break in the curve on the first bite. The areas under the curve during the first bite and the second bite are a measure of the work done in the compression. The ratio of these two areas (A2/A1) is defined as cohesiveness. The negative force peak in the first decompression is defined as adhesive force and the negative areas as adhesive work. The distance that the product extends in the decompression is defined as stringiness, and the distance that the product recovers its height between the first and second compressions is defined as springiness. Two other parameters are derived by calculation: gumminess is defined as the product of hardness $\times$ cohesiveness; chewiness is defined as the product of hardness $\times$ cohesiveness $\times$ springiness. Gumminess is the energy required to disintegrate a semisolid food while chewiness is the energy required to masticate a solid food. It is incorrect to report gumminess and chewiness for the same product.

In the early days, instrumental texture profile analysis was an academic curiosity because of the long time required to extract the information from the force–time curves. With computer retrieval and analysis of the data, the time requirement has been reduced to the point where it can become a simple routine test. The TA.XT2 texture analyzer is particularly well suited to do this, and as a result measuring TPA in the TA.XT2 is now a widely used procedure.

Rheological Measurements

Historically, rheology has been defined as the study of the deformation and flow of matter or the response

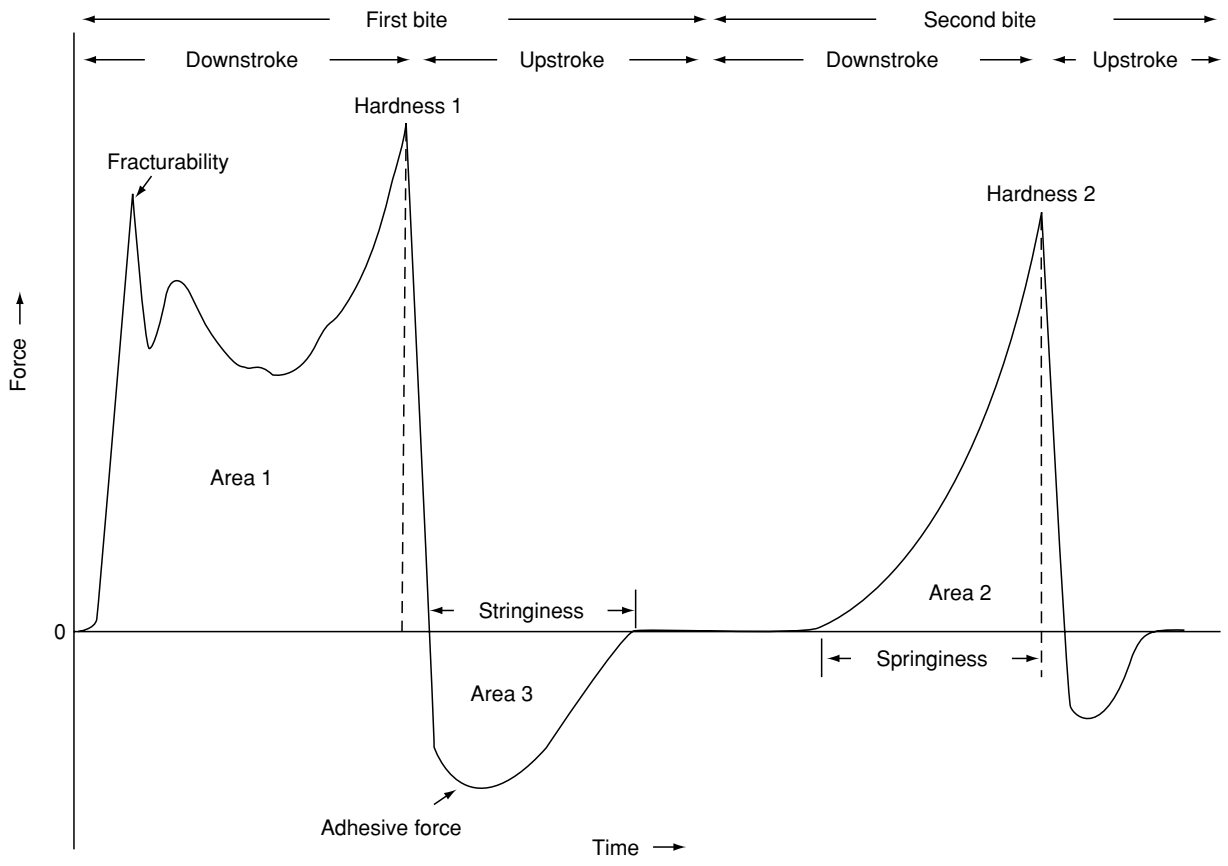

Figure 2 A generalized texture profile analysis curve. From Bourne MC (1978) Texture profile analysis. *Food Technology* 32: 63, with permission.

of materials to stress. The science of rheology, proven so useful in the field of high polymers, has many applications to food, including raw materials (e.g., grains, meats, fruits), intermediate products in the manufacturing process (e.g., bread dough, cheese curd, sausage emulsions), and finished products (all foods). However, rheology does not cover all aspects of texture. The size reduction that occurs during mastication is not rheology, neither are the sensations of oiliness, moistness, and particle size and shape.

Most people believe that there is a sharp distinction between solids (which do not flow) and liquids (which flow). In fact, the distinction between solids and liquids is far from clear because many liquids possess some of the properties of solids, and many solids possess some of the qualities of liquids. The science of rheology specializes in the study of these complex materials (of which there are many examples in foods) that are partly solid and partly liquid.

The flow of liquids may be divided into several broad classes:

1. Newtonian flow, in which the shear rate* is directly proportional to the shear stress†. Examples are edible oils, sugar syrups, milk, and honey. The measurement of the properties of these foods is straightforward, since viscosity‡ is independent of shear rate.
2. Plastic or Bingham flow, in which a minimum shear stress, known as the 'yield stress,' must be exceeded before flow begins. Examples are tomato ketchup, whipped egg white, mayonnaise, margarine, and butter.

*Shear rate (denoted by the symbol $\dot{\gamma}$ and expressed in s^{-1}) is the velocity gradient established in a fluid that results from the application of a shear stress.

†Shear stress (denoted by σ and expressed in Pa) is the force per unit area applied tangential to the plane on which the force acts.

‡Viscosity (denoted by η and expressed in Pa s) is the internal friction of a fluid or its tendency to resist flow ($\eta = \sigma/\dot{\gamma}$); it should only be used for Newtonian fluids. Apparent viscosity (denoted by η_a) is the viscosity of a non-Newtonian fluid expressed as the viscosity of a Newtonian fluid at a specified shear rate (e.g., $\eta_a = \sigma/\dot{\gamma}_{50}$ is the apparent viscosity of a non-Newtonian fluid at a shear rate of 50 s^{-1}).

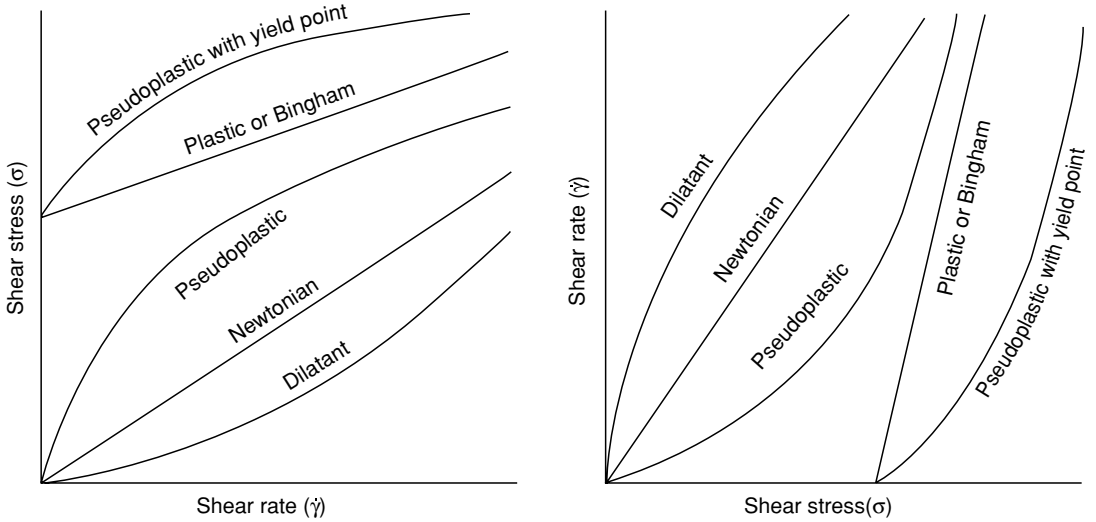

Figure 3 (a) Shear-stress versus shear-rate plots for various types of flow. (b) Shear-rate versus shear-stress plots for the same types of flow. From Bourne MC (1982), *Food Texture and Viscosity: Concept and Measurement*. New York: Academic Press, with permission.

3. Pseudoplastic flow, in which an increasing shear force gives a more than proportionate increase in shear rate, i.e., apparent viscosity decreases with increasing shear rate. Salad dressings exemplify this type of flow.
4. Dilatant flow, in which equal increments in shear stress give less than equal increments in shear rate, i.e., apparent viscosity increases with increasing shear rate. This type of flow is rare in foods but is found in high-solid suspensions of raw starch and some chocolate syrups. (Refer to individual foods.)

The differences between these types of flow are shown in Figure 3. The shear-stress versus shear-rate plot (Figure 3a) is the more common manner of presentation; another that is sometimes used is shown in Figure 3b, in which the axes are interchanged.

Time Dependency

For some fluids the shear stress is a function of both the shear rate and the time it is subjected to shear. For thixotropic products the apparent viscosity decreases with time of shearing. This condition is frequently found in food systems such as gum solutions and starch pastes. Rheopectic products increase in apparent viscosity with time of shearing. This type of behavior is rare in foods.

Viscoelasticity

As discussed above, most foods combine some of the properties of ideal liquids, which exhibit only viscosity (flow) and ideal solids which exhibit only elasticity (deformation). These are called viscoelastic foods. In characterizing these systems rheologically it is necessary to measure both the viscous component (loss modulus, G″) and the elastic component (storage modulus. G′). A food with high G′ and low G″ behaves more like a solid than a liquid, while a food with low G″ and high G′ behaves more like a liquid than a solid. An elastic solid such as rock candy will have a G″ value of zero, while a Newtonian liquid such as sucrose syrup will have a G′ value of zero.

Traditionally, viscoelasticity was measured by performing creep tests in which a weight was placed on the test material and the change in specimen height was monitored over a period of time. Presently, viscoelasticity is usually measured by an oscillation test in which a sample of known dimensions (if solid) or filled into the space between a cone and plate or parallel plates or a cup with standard dimensions (if liquid) is subjected to repeated small sinusoidal deformations that do not fracture the sample. Analysis of the resulting shear stress versus time curves yields the numerical values of G′ and G″. Modern instrumentation provides a computerized system for quantifying these moduli. The rate and degree of deformation may be varied to provide information on the internal structure and mechanical behavior of the material. (*See* **Food Acceptability**: Affective Methods.)

See also: **Food Acceptability**: Affective Methods; **Sensory Evaluation**: Sensory Characteristics of Human Foods; Food Acceptability and Sensory Evaluation; Sensory Difference Testing; Sensory Rating and Scoring Methods

Further Reading

Aguilera JM and Stanley DW (1999) *Microstructural Principles of Food Processing and Engineering*, 2nd edn. Gaithersburg, Maryland: Aspen Publishers.

Bourne MC (2002) *Food Texture and Viscosity: Concept and Measurement*, 2nd edition. New York: Academic Press.

Christensen CM (1984) Food texture perception. *Advances in Food Research* 29: 159–199.

De Man JM, Voisey PW, Rasper VF and Stanley DW (eds) (1976) *Rheology and Texture in Food Quality*. Westport, Connecticut: AVI Publishing.

Lawless HT and Heymann H (1999) *Sensory Evaluation of Food – Principles and Practices*. Gaithersburg, Maryland: Aspen Publishers.

Meilgaard M, Civille GV and Carr BT (1999) *Sensory Evaluation Techniques*, 3rd edn. Boca Raton, Florida: CRC Press.

Mohsenin NN (1970) Application of engineering techniques to evaluation of texture of solid food materials. *Journal of Texture Studies* 1: 133–154.

Muñoz AM and Civille GV (1987) Factors affecting perception and acceptance of food texture by American consumers. *Food Reviews International* 3: 285–322.

Rao MA and Steffe JF (1992) *Viscoelastic Properties of Foods*. New York: Elsevier Applied Science.

Rosenthal AJ (1999) *Food Texture – Measurement and Perception*. Gaithersburg, Maryland: Aspen Publishers.

Sherman P (ed.) (1979) *Food Texture and Rheology*. New York: Academic Press.

Szczesniak AS (1987) Correlating sensory with instrumental texture measurements – an overview of recent developments. *Journal of Texture Studies* 18: 1–15.

Szczesniak AS (1991) Textural perceptions and food quality. *Journal of Food Quality* 14: 75–85.

Szczesniak AS and Kahn E (1971) Consumer awareness of and attitudes to food texture I. Adults. *Journal of Texture Studies* 2: 280–295.

Aroma

D S Mottram and J S Elmore, The University of Reading, Whiteknights, Reading, UK

Aroma

The sense of smell (olfaction) is one of the two chemical senses in humans, the other being the sense of taste. The term 'odor' has been defined by the British Standards Institute as the quality of the sensation perceived, via the olfactory organ situated in the nasal cavity, from certain volatile compounds; aroma is an odor with a pleasant connotation. The olfactory organ is extremely sensitive, and hence the concentrations of volatile compounds in foods contributing to aroma (odorants) may be very low. These complex mixtures of volatile compounds may be analyzed by extraction, followed by separation and characterization.

Olfaction

The olfactory receptors are sited in the olfactory epithelia, two small patches of mucous membrane in the recesses of the nasal passages (**Figure 1**). The response of these receptors to airborne molecules determines the odor of a substance. Molecules entering the nose find their way to the olfactory epithelia where they come into contact with cilia (hairs) on the receptor cells. These produce a change in electrical potential and a signal travels along nerve fibers to the olfactory bulb, part of the brain. There are approximately 100 million olfactory receptor cells in the human nose.

Olfaction plays a primary role in the perception of flavor in foods and beverages. When food is eaten taste, smell and mouthfeel act in concert to produce the integrated sensation of flavor. The aroma of the food is not just perceived in inhaled air. When food is in the mouth, odorants can reach the olfactory receptors by the retronasal route, entering the nasal cavity by the same route as expired air.

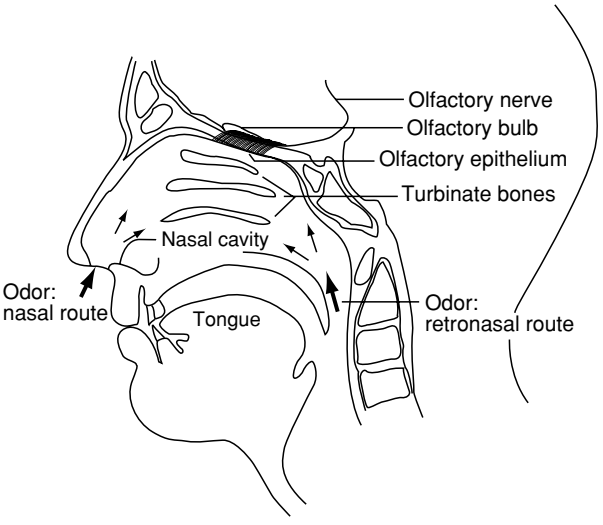

Figure 1 Representation of a section through the nasal and oral cavities showing the olfactory organ. From Thomson DMH (1986) The meaning of flavour. In: Birch GG and Lindley MG (eds.) *Developments in Food Flavours*, pp. 1–21. London: Elsevier Applied Science, with permission.

Odor Sensitivity

Odor perception is considerably more sensitive than the perception of taste, by factors of one million to one billion. As we are able to differentiate a large number of odorants, aroma plays the major role in delineating the characteristic flavor of a particular food. The inability to differentiate flavors is usually associated with anosmia (lack of a sense of smell) rather than ageusia (lack of a sense of taste).

It has been suggested that the triggering of one human olfactory neurone by a powerful odorant may require only eight molecules, and as few as 40 molecules can produce an identifiable sensation. The odor threshold of a compound is defined as the minimum concentration at which the odor of a compound can be detected. The range of thresholds exhibited by odorants extends over at least 10 orders of magnitude (Table 1). One compound with one of the lowest reported odor thresholds is bis-(2-methyl-3-furyl) disulfide, which has a meaty aroma and can be detected at concentrations of 2 parts in 10^{14} parts of water. This concentration is equivalent to about 20 μg in the water of a 50-m swimming pool of 1000 m^3 capacity! At the other end of the range, at least 1 mg of ethanol needs to be present in 10 ml of water before the smell of ethanol can be detected.

Chemical Nature of Aroma Compounds in Foods

The range of chemical classes that contribute to food aromas is diverse in both chemical composition and physical properties, and includes aliphatic, alicyclic, aromatic, and heterocyclic compounds. All aroma compounds have some volatility, but this ranges from permanent gases to compounds with very little vapor pressure and molecular weights up to 300. Approximately 10 000 different volatile compounds have been identified in foods and beverages; complex cooked foods such as coffee and meat contain in excess of 1000 aroma compounds. Many of these compounds are common to a number of different foods, and the contribution of any compound to the characteristic aroma of a particular food will depend on a number of factors, including:

- odor character;
- concentration in product;
- odor threshold;
- vapor pressure;
- adsorption on the food matrix;
- interaction with other components;
- synergism with other volatiles.

Some volatiles have little aroma and do not contribute to food flavor, whereas others define the aromas of specific foods (e.g., benzaldehyde – almonds; citral – lemons). However, for most foods, the aroma depends on contributions from a complex mixture of volatiles comprising different chemical classes. (*See* **Essential Oils**: Properties and Uses; **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Characterization of Aroma Volatiles

In attempting to identify those compounds that characterize different food aromas, flavor scientists have analyzed the volatile composition of many different foods. This usually involves the extraction of volatiles from the food matrix, their concentration, and the separation and identification of the individual components (Table 2). As far as possible, the extract should contain only the volatile components in the same relative proportions as the original food, without the introduction of artefacts. The wide range of odor thresholds can result in quantitatively minor

Table 1 Odor thresholds of some aroma compounds in aqueous solution

Compound	*Threshold (ppb, μg l^{-1})*
Ethanol	100 000
2,5-Dimethylpyrazine	1 800
Butyric acid	250
Limonene	10
Hexanal	5
Ethyl 2-methylbutyrate	1
Methanethiol	0.2
β-Ionone	0.007
2-Isobutyl-3-methoxypyrazine	0.002
Bis-(2-methyl-3-furyl)disulfide	0.00002
2,3,6-Trichloroanisole	0.0000003

Table 2 Stages in the analysis of aroma volatiles

Step	*Analytical techniques*
Isolation and concentration	Headspace analysis
	Adsorption
	Distillation
	Extraction
Separation	Gas chromatography
	Preparative liquid chromatography
Identification	Chromatographic retention time
	Mass spectrometry
	Infrared spectroscopy
	Nuclear magnetic resonance spectroscopy
	Chemical synthesis
	Reference compounds
Sensory characterization	Gas chromatography olfactometry
	Sensory panel

components making a major contribution to the aroma, whereas some large components will have no sensory significance. Successful isolation and identification of minor components with aroma significance is one of the major challenges of aroma analysis.

Isolation and Concentration of Aroma Volatiles

The components of food that are responsible for aroma are present in extremely small quantities compared with the major constituents, of which water is usually the most abundant. A number of isolation techniques exist, all based on utilizing the volatile nature of the aroma compounds to separate them from the food matrix.

Headspace analysis The concentration of volatiles in the headspace vapor above a food or beverage will be very low. However, it will contain a representative mixture of those compounds that contribute to the aroma of the product, and therefore, analysis of that headspace will provide the best method of obtaining a sample representative of the food aroma. However, direct introduction of a given volume of the vapors above a food sample on to a gas chromatographic column is rarely used, because the volatiles are not present in sufficiently high concentrations and because water from the sample will often cause interference. Purging the headspace volatiles with a stream of inert gas (nitrogen or helium) and condensing the volatiles in a series of traps cooled by ice, solid carbon dioxide, or liquid nitrogen will result in a concentrated aroma condensate. Extraction of the condensate with a small amount of a suitable solvent provides an aroma extract suitable for chromatographic analysis. Instead of using cold traps, headspace volatiles can be collected on suitable adsorbents.

Adsorption The ability of certain solid surfaces to reversibly adsorb volatile organic molecules is widely used in the analysis of aroma compounds. Activated charcoal was the first to be used, but now, certain porous polymers, in particular Tenax TA, are frequently employed. Like charcoal, all of these materials readily adsorb volatiles while having little affinity for water, making them particularly useful in the analysis of samples with a high water content.

In a typical collection, a stream of a purified inert gas (helium or nitrogen) sweeps the volatiles from the sample flask into a small tube containing 10–200 mg of the adsorbent. Adsorbed volatiles can be heat desorbed directly on to a gas chromatographic column by placing the adsorbent trap in a specially modified injection port, thus avoiding any loss of components or unnecessary dilution. Cooling the front of the column ('cryofocusing') with solid carbon dioxide or liquid nitrogen during this desorption will avoid any loss in chromatographic resolution. Alternatively, volatiles may be removed from the trap by passing a small amount of a low-boiling-point solvent through the adsorbent. Concentration of the collected solution by careful evaporation of the solvent provides an aroma extract for gas chromatography (GC).

Solid-phase microextraction This is an adaptation of traditional adsorption methods in which a small fused silica fiber, coated in an adsorbent material, is mounted in a syringe-like device (Figure 2). The needle is pushed through a septum, and the fiber is exposed to the headspace above the food or beverage sample, which is sealed in a suitable container. Volatile compounds are adsorbed on to the fibre, and at the end of the extraction, the fiber can be removed from the sample vessel and directly desorbed into the split/splitless injector of a gas chromatograph. Again, cryofocusing of the aroma volatiles on the front of the gas chromatographic column is performed to insure no loss of resolution. Solid-phase microextraction (SPME) has evolved rapidly since its introduction in the early 1990s, and fibres coated with Carboxen™ are suitable for the analysis of a wide range of aroma compounds. SPME can be automated, and its ease of use, relatively low cost, and affinity for a large range of compounds would suggest that it will become a widely used technique for the isolation of aroma volatiles.

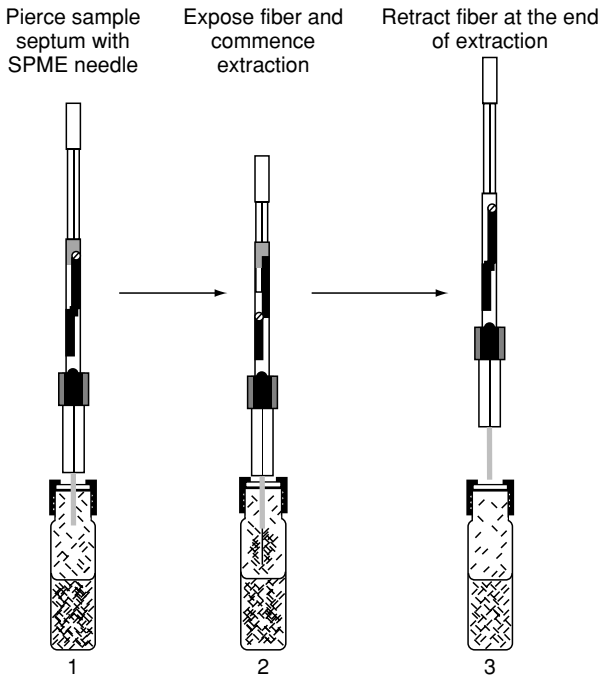

Figure 2 Principle of solid-phase microextraction.

Distillation Steam distillation finds application in the analyses of volatiles from beverages and high-water-content foods, although it is less applicable to fats and oils. It has the disadvantage that the large quantities of aqueous distillate require further extraction with a solvent, to separate the volatiles from the water. Concentration of the extract is then necessary. The formation of artefacts may also be a problem. Vacuum distillation finds wide application in the extraction of volatiles from fats and oils, and at low extraction temperatures artefact formation can be minimized. In molecular distillation, a high vacuum ($< 10^{-3}$ mmHg), is used, and the volatiles have a relatively short path to travel from the surface of a liquid oil sample to a cold surface where they condense. Efficient collection of high-boiling-point components is achieved because at high vacuum, the distance between the sample surface and the cold condenser (10–20 mm) is less than the mean free path of the molecules. All of these distillation processes incorporate cold traps for the collection of the distillate.

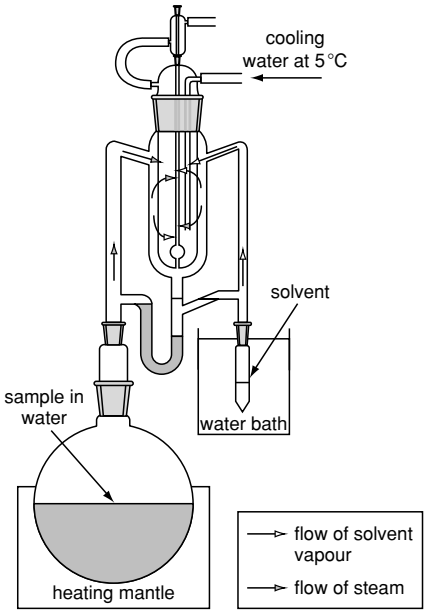

Figure 3 Likens–Nickerson apparatus for continuous steam distillation–solvent extraction.

Extraction Direct extraction of an aqueous food fraction with an organic solvent is of limited use, because the extract will contain a large amount of nonvolatile matter. However, supercritical carbon dioxide has been utilized as a solvent for the extraction of aromas on both a laboratory and a commercial scale, particularly the preparation of high-quality essential oils. Its solvating qualities can be altered by changing the pressure or temperature at which the extraction takes place, and under ideal conditions, supercritical carbon dioxide exhibits a strong affinity for most aroma compounds, whereas most nonvolatile constituents are insoluble. The ease of removal of the solvent, after extraction, to give a concentrated aroma extract is another attractive feature of supercritical carbon dioxide extraction.

One of the most widely used techniques in aroma analysis combines steam distillation with solvent extraction in a Likens–Nickerson apparatus (Figure 3), which was first reported in 1964 for the extraction of hop oil. The essential feature is the simultaneous condensation of the steam distillate with an immiscible extraction solvent. A simple U-tube, with appropriate side arms, allows the return of water to the steam distillation flask and the solvent, containing the extracted volatiles, to a reservoir flask where it redistils for further extraction. The technique provides a simple rapid method involving only small volumes of solvent. By carrying out the extraction under a reduced pressure, thermal degradation of labile components can be diminished.

The organic solvents used for extraction are chosen on the basis of their selectivity for the volatile compounds of interest and on boiling point. Commonly used solvents are diethyl ether, pentane, isopentane, dichloromethane, and some chlorofluorocarbons. After each type of solvent extraction, most of the solvent must be removed to provide an aroma concentrate suitable for GC. Removal of traces of water, using a drying agent (sodium or calcium sulfate) or by freezing out at $-20\,^{\circ}C$, is essential before careful removal of solvent by distillation. Final concentration to a volume of 0.1–1 ml is often achieved by purging with a gentle stream of nitrogen.

Separation of Aroma Components

To determine the sensorially important compounds in an aroma isolate, the complex mixture needs to be separated into its components. The amount of isolate is usually small, containing many compounds of diverse chemical structures, varying greatly in concentration, and important components are often present in extremely low amounts. The success of any aroma analysis depends mainly upon the efficiency of separation and the sensitivity of detection. GC using bonded-phase fused-silica capillary columns is almost universally used as the separation method in aroma analysis. Such columns can separate complex mixtures, and the most commonly used stationary phases are Carbowax 20M, a polar phase, and the two nonpolar phases, 100% poly(dimethylsiloxane) and poly(5% diphenylsiloxane/95% dimethylsiloxane). The stereochemistry of compounds can influence

aroma, e.g., (+)-carvone possesses a spearmint aroma, whereas (−)-carvone possesses a caraway-like aroma; capillary columns coated with chiral phases can be used to separate enantiomers. The retention times of an aroma compound on two columns with different stationary phases can be helpful in its identification, and databases containing retention data for volatile compounds are available.

Preparative liquid chromatography is sometimes performed on solvent extracts, as a first step, to simplify subsequent separations. The most widely used example of this is flash chromatography, where the extract is pipetted on to the surface of a stationary phase, such as silica gel, which is contained in a glass column. Fractions of the extract are obtained, as solvents of increasing polarity are passed through the column under positive air pressure. The fractions are analyzed subsequently by GC. (*See* **Chromatography**: Gas Chromatography.)

Detection of Components of Sensory Significance

An essential step in the analysis of aroma is to determine which volatile compounds contribute to the aroma of the food or beverage. Sensory profiling of the whole product can give some indication of the important aroma characters present in the product, and such information should be used when assessing the contribution of the individual components of an aroma isolate to the overall flavor quality. A widely used technique for determining components that contribute to aroma is GC olfactometry (GCO). The column effluent is split between a conventional GC detector and a vent to the outside of the oven, where the odors emerging can be smelled and described. Chromatogram peaks can thus be identified that correspond to specific aromas, and 'aromagrams' can be compiled to complement the chromatogram.

To estimate the relative contributions of volatile components towards the total aroma quality of a food, a set of serial dilutions of an aroma extract can be made up and analyzed by GCO. At a certain dilution, the aroma of a particular volatile compound will no longer be perceived. The flavor dilution factor for that compound is defined as the highest dilution at which that compound can be perceived by GCO. For example, if the concentration of the extract was halved at each dilution and the seventh dilution was the last at which the compound could be detected, its flavor dilution factor would be 2^7 (128). Hence, if the aroma extract is representative of the food from which it is derived, the most important contributors to the aroma of the food are those with the highest flavor dilution factors. It should be noted that components with high flavor dilution factors might not give GC peaks of any significant size.

Another way of estimating the contribution of a volatile compound to the aroma of a food is by calculating its odor unit number. The odor unit number of an aroma compound in a food is defined as the ratio between its concentration in the food and its odor threshold, as measured by sensory evaluation. Hence, the higher the odor unit number of the compound, the more likely it is to contribute to the overall aroma of the food. Several literature compilations of odor thresholds exist.

Identification of Components

Structure elucidation of the chromatographically separated components is the next step in the analysis of an aroma isolate. Coupled gas chromatography–mass spectrometry (GC–MS) allows direct analysis of the separated components and provides the most efficient means of volatile identification.

Gas chromatography–mass spectrometry The availability of coupled gas chromatograph–mass spectrometer systems in the late 1960s was one of the most important developments for flavor research, and this was followed a decade later by the introduction of efficient computerized data-handling systems. Today, sensitivity is usually such that a full spectrum can be readily obtained on 1 ng of a single component in a complex mixture injected on to the GC column, and in many cases, an identifiable spectrum can be obtained from as little as 10 pg. Quadrupole mass spectrometers, obtaining at least one spectrum per second, are ideal for low-resolution GC–MS of aroma extracts. However, double-focusing magnetic sector mass spectrometers can acquire accurate mass data, allowing the calculation of empirical formulae for ions in a spectrum, which can be a great asset in the identification of unknown compounds.

The characterization of unknown compounds is greatly facilitated by comparing their mass spectra with those of known compounds in compiled libraries incorporated into the GC–MS data system. Current databases contain the mass spectra of over 200 000 compounds, but special compilations containing only volatile compounds have also been prepared, and at least one is available in computer-readable format. Confirmation of the identity of compounds should always be carried out, preferably by comparing their mass spectra and retention times with those of authentic samples. (*See* **Mass Spectrometry**: Principles and Instrumentation.)

Spectroscopic methods Nuclear magnetic resonance and infrared (IR) spectroscopy are both extremely valuable techniques in the identification of organic compounds. However, they require larger sample

sizes than mass spectrometry and are not as widely used in aroma analysis. They are primarily used to aid structure elucidation of unknown compounds.

Fourier-transform IR instruments directly coupled to a gas chromatograph have been developed, giving adequate vapor-phase spectra on 10–100 ng of compound. IR offers functional group information and in general is complementary and supportive of MS. Because IR is nondestructive, it has been possible to develop tandem GC–IR–MS systems, providing a powerful tool for aroma analysis.

Quantification of Aroma Components

Often, quantitative information on aroma compounds in a food is needed, such as for the calculation of odor unit numbers. Quantification is rarely simple, because most extraction techniques only remove a proportion of the aroma from the food. Sometimes, the extraction technique may generate aroma compounds, and difficulties in quantification may arise when compounds are not resolved by GC. The most effective means of quantification is isotope dilution assay using GC–MS, where a known amount of a ^{13}C- or ^{2}H-labeled internal standard is added to a slurry of the food to be extracted, in order to quantify its nonlabeled equivalent. As the labeled and unlabeled aroma compounds possess similar physical properties, the proportion of each extracted will be the same. The relationship between the labeled standard and the compound of interest can be used to calculate accurately the amount of the compound of interest in the food.

Other quantification methods include the addition of an internal standard, not present in the food, of a similar chemical composition to the compound of interest, e.g., 2-methylpentanal could be used to quantify hexanal. Conversely, standards could be added to the extract, in order to calculate the amount of the compound of interest in the extract rather than the food itself. Standards can be injected into traps containing adsorbent and, in the case of SPME, injected on to the GC column immediately before desorption of the fiber.

The Electronic Nose

The electronic nose is an array of chemical sensors, connected to a pattern-recognition system that responds to odors passing over it. Different odors cause different responses in the sensors, and these responses provide a signal pattern characteristic of a particular aroma. The computer evaluates the signal pattern and can compare the aromas of different samples, using pattern recognition. For example, if a manufacturer uses coffees from around the world, the electronic nose can analyze them, and their aromas can be plotted on a multidimensional response map. The analysis of several samples of each coffee, under ideal conditions, can result in a group of points for each coffee, well resolved from any other group. As the number of samples increases, the resolution between the groups should also increase. Plotting the sensor responses of an unknown sample on the same map should permit its recognition, by its proximity to one of the known samples.

Sensors are usually made of metal oxides or organic polymers, although more recently, surface acoustic waves and piezoelectric crystals have been used. Problems may exist when samples with a high water content are analyzed, as many of the sensors respond strongly towards water, preventing any sample differences being observed.

Measurement of Aroma Release during Eating

Attempts have been made by several researchers to measure the release of aroma during the eating of food. At present, mass spectrometry using an atmospheric pressure chemical ionization (APCI) interface appears to be the most effective way of doing this. As a subject eats a food or beverage, a sampling tube draws the volatiles from the mouth or nose into the mass spectrometer interface. Released aroma compounds are ionized in air containing water vapor at atmospheric pressure in the APCI source; the ions are subsequently sampled into the high-vacuum area of the mass spectrometer. APCI does not normally induce fragmentation, and a strong protonated molecular ion is observed. By monitoring the molecular ions associated with the major aroma compounds in the sample, volatile release profiles can be measured. Pulsed release curves are usually obtained, due to respiration of the subject, and these can be smoothed to show how each of the volatiles of interest behaves during the eating process. This equipment has been used, for example, to examine how reducing the fat content of a food will affect its flavor or to study how the chewing patterns of different people affect flavor release.

See also: **Chromatography**: Gas Chromatography; Combined Chromatography and Mass Spectrometry; **Essential Oils**: Properties and Uses; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Food Acceptability**: Affective Methods; **Mass Spectrometry**: Principles and Instrumentation; Applications; **Spectroscopy**: Infrared and Raman; Nuclear Magnetic Resonance; **Taints**: Types and Causes; Analysis and Identification

Further Reading

Acree TE and Teranishi R (eds) (1993) *Flavor Science: Sensible Principles and Techniques*. Washington, DC: American Chemical Society.

Goodenough PW (1998) The molecular biology of olfactory perception. *International Journal of Food Science and Technology* 33: 63–77.

Maarse H (ed.) (1991) *Volatile Compounds in Foods and Beverages*. New York: Marcel Dekker.

Marsili R (ed.) (1997) *Techniques for Analysis of Food Aroma*. New York: Marcel Dekker.

Meilgaard M, Civille GV and Carr BT (eds) (1991) *Sensory Evaluation Techniques*, 2nd edn. Boca Raton, FL: CRC Press.

Morello M and Mussinan C (eds) (1998) *Challenges in the Extraction and Isolation of Food Flavors*. Washington, DC: American Chemical Society.

Nijssen LM, Visscher CA, Maarse H, Willemsens LC and Boelens MH (1996) *Volatile Compounds in Food – Qualitative and Quantitative Data*, 7th edn. Zeist, The Netherlands: TNO Nutrition and Food Research Institute.

Pawliszyn J (ed.) (1999) *Applications of Solid Phase Microextraction*. Cambridge: Royal Society of Chemistry.

Taylor AJ and Mottram DS (1996) *Flavour Science: Recent Developments*. Cambridge: Royal Society of Chemistry.

Taste

H Valentová and Z Panovská, Institute of Chemical Technology, Prague, Czech Republic

Taste

Taste, according to the International Standard Sensory Analysis Vocabulary ISO 5492 (1992), is defined as sensations perceived by the taste organ when stimulated by certain soluble substances. Taste is closely related to smell. The perception of odor and taste, combined with trigeminal sensations, results in the overall flavor. Flavor influences food acceptance and selection of food intake, and helps us to distinguish potentially harmful compounds. The taste sensation is a very complex process starting at the sensory receptor level and finishing in the central nervous system, where it is combined with information coming from other senses. The earliest work concentrated on anatomical studies, and so we have a good deal of information about the structural taste system on the tongue.

Anatomy of the Tongue

Mouth

An important role in the perception of taste is played by our mouth cavity and especially by the tongue. Food is placed inside the mouth, where our teeth chew it, and the saliva makes it into a solution, and therefore, the food comes into contact with the surface of the tongue. The main role for perception of taste is played by taste papillae, taste buds, and taste cells.

Papillae

Taste papillae are located on the tongue, soft and hard palate, pharynx, and epiglottis. The papillae give the tongue its bumpy appearance and can be classified, according to their shape, into four types: fungiform, foliate, circumvallate, and filiform papillae. The first papillae appear when the human fetus is 6 weeks old, and at around 10 weeks, the first primitive taste pores are found. **Table 1** shows a description of the different types of papillae.

Taste Buds

In the epithelia of the papillae, there are taste buds, 40–80 μm in size. Children have about 10 000 taste buds, but the amount decreases with age. The average number for adults is 2000–3000 buds. The taste-bud density and placement of the taste buds among people vary. For example, the density on the tongue tip ranges from 3–512 taste buds per square centimeter, with an average of around 115 buds per square centimeter. People with a higher density of buds perceive taste stimuli more intensively. The specificity of individual taste bud cells is being studied. It is still not clear at the taste bud level whether there is modality specificity. Some studies have confirmed that some fungiform papillae contain only one taste bud, but this could not be proved. It has also been demonstrated that taste buds respond differently to different

Table 1 Description of papillae[a]

Types of papillae	*Number of papillae*[a]	*Taste buds per papilla*[a]	*Location*
Fungiform	100–400	0–22	Base and edges of the tongue
Foliate	8–20	10	Each side of lingual posterior
Circumvallate	3–13	100–250	V-shaped pattern across the root of the tongue
Filiform	1000	No taste buds	All the tongue

[a]Average values are given because numbers differ among authors.

substances. It seems that all areas with taste buds respond to some degree to all tastes, but there is usually one taste to which the response is the best.

Taste Cells

The tip of the taste bud has a small opening about 2–5 μm in diameter, called a taste pore, where taste stimuli interact with taste receptors. Taste buds are aggregates of some 40–60 elongated epithelial cells. There are different types of cells in a taste bud: dark cells, light cells, cells with dark-cored and synaptic vesicles, basal cells, and perigemmal cells. It is not clear which of the cells serve as receptor cells or if the different cell types are specialized to detect different ranges of the gustatory spectrum. The taste cells continually regenerate; they live for only 7–10 days and, within this period, grow to maturity and die. The young taste receptor cells move from the periphery of the taste bud to its center, and the older taste cells in the central portion of the taste buds die away and are then pressed out from the taste bud into the mouth through taste pores. By replacing the cells in this way, the gustatory system is protected from losing its sensitivity. All taste cells from one taste bud are connected by basal synapses to about 50 afferent nerve fibers. The nerve fibers from taste buds converge into different cranial nerves.

Cranial Nerves

The sense of taste is mediated by three cranial nerves. The anterior (front) two-thirds of the tongue and palate is supplied by branches of facial nerves, namely chorda tympani branch (seventh cranial nerves), the back of the tongue by glossopharyngeal nerve (ninth cranial nerves), the base and central region, and the throat and larynx by certain branches of vagus nerve (tenth cranial). All nerves subserve not only taste but also touch, temperature, and pain sensitivity. The lingual nerve is the branch of the fifth cranial nerve and is responsible for the perception of pungency, heat, cooling, and others.

The taste fibers from all the sensory nerves from the mouth come together in the brainstem (medulla oblongata). From the brain's medulla, the gustatory fibers ascend by a pathway to a small cluster of cells in the thalamus and thence to a taste-receiving area in the anterior cerebral cortex.

Two main theories try to describe how taste information is coded in the brain. In the pattern theory, the main role is played by taste receptor cells that respond with different sensitivities to the different tastants. In the labeled line theory, the four tastes are sensed in four discrete type of receptor cells, and the information from receptors goes to the brain through independent channels. It seems that both types of mechanism are involved in taste perception.

Transmission of Taste Through Nerve Impulses

The transmission of electrochemical impulse through the nerves is called transduction. The understanding of this process has been the main focus of taste researchers.

Generally, the arrival of a chemical stimulant on the surface of a receptor changes the ionic conductance in the taste membrane and the electrochemical impulse is produced. The transduction pathway involves a number of different receptors and different cellular mechanisms.

Taste stimuli may interact:

- directly with the ion channel – this produces receptor cell depolarization either by direct permeation or by blockage of the ion channel;
- with receptors and alter the flux of ion through the ion channel (K^+, Na^+, Ca^+);
- may diffuse through lipid phase and affect G protein or the endoplasmatic reticulum.

Researchers believe that each basic taste has a particular transduction pathway (Table 2) that varies with the location of taste buds on the tongue. For

Table 2 Suggested pathways for individual tastes

Taste	*Possible pathways for transduction mechanism for different stimuli*
Bitter	Multiple mechanisms – depending on the structure of the compounds: (a) Inhibition of apical K^+ channel (mechanism for chinine, divalent salts) (b) Blockage of Na^+ amino transport system (c) G protein-mediated activity involving phospholipases
Sweet	Specific membrane receptors are required for transduction of sugars: (a) Multiple sweet receptors coupled to G protein (gastducin) and second messenger system (b) Stimulus-gated ion channel (c) Direct activation of G protein by sweet stimuli
Sour	Does not require specific taste receptor – several apically located ion channels may participate in sour transduction: (a) sodium ion channel – proton enters taste cells through amiloride-sensitive sodium channels; (b) channels are stimulated by protons that permit entry of cations into the cell
Salty	Does not require specific taste receptor – direct permeation of Na^+ ions through amiloride-sensitive sodium channel
Umami	Stimulus-gated ion channel: taste stimulus binds to a receptor site on a receptor–ion channel complex

sweet and bitter taste, there probably exist multiple pathways that can be common for substances with a similar structure. The transduction of the different taste stimuli has been studied on nonhuman species, and it is not still clear whether the same mechanism is involved in humans.

Basic Tastes

Over the centuries, the numbers of primary tastes have varied from two – sweet and bitter – to more then 10. Studies carried out to date have distinguished four taste categories: sweet, salty, bitter, and sour taste. Now, some of the scientists also include between basic tastes a less-established taste, 'umami.' International Standards ISO 5492 in paragraph 3.22, under basic taste, states: 'basic taste: any one of the distinctive tastes: acid, bitter, salty, sweet, alkaline, umami, metallic.'

Sweet Taste

Sweet taste is one of the favorite tastes and therefore the most studied. There is evidence that prenatal infants detect and like sweet *in utero* and that human infants are born with a preference for sweet taste.

Scientists tried to solve the problem of sweet taste through the chemical structure of sweet compounds, but substances that have a sweet taste are found in all chemical classes. They are usually associated with multiple hydroxyl groups, and α- but not β- or γ-amino acids. Salts such as beryllium or lead salts, simple organic compounds such as chloroform, and alcohols can have a sweet taste. The main groups are carbohydrates, with sweetness decreasing in a homologous series sugars > oligosaccharides > polysaccharides. Other sweet organic compounds include amino acids, peptides, proteins, compounds with NH_2 groups such as amides, ureas, and hydrazides. Benzene compounds with one single NO_2 group are often sweet, and those with more NO_2 groups are usually bitter. Small changes in the stereochemie of atoms within a molecule may affect its taste. The chemical structure of the sweet compound also plays an important role in the determination of the area of the tongue, where the sweet taste is perceived. For example, the taste of inorganic salts and nitrogen compounds is perceived on the front of the tongue, some dihydrochalcons on the back, and some sweet proteins such as thaumatine and nonelline on the sides.

From the time when, in 1914, the need for discrete functional groups of a sweet compound was recognized, several theories tried to explain the sweet taste. In Shallenberger's hypothesis sugar molecules must possess more hydrogen-bonded functional groups per carbon. The molecules interact with the sweet-taste receptor sites of a protein through hydrogen bonds. The question is if there is a receptor for sweet taste or if a different structural class has its own special receptor and how they interact with the sweet compound. It seems that this question will be answered soon. Recently, scientists discovered a sweet receptor gene, T1R3, thought to be the likely basis for the tongue's sweetness sensor. One of the newest theories is multipoint attachment theory. This theory assumes the presence, probably a seven-pass transmembrane receptor, of at least eight fundamental recognition sites. The multiple theory is supported also by other findings, such as synergy effects of sweeteners, the adaptation of the response to some sweeteners, but not to others, the sensitivity of different parts of the tongue to the different sweeteners, and human variability in sensitivity to different sweeteners.

Studies with a new approach to studying the mechanism of sweet taste also appeared during the last decade. Schallenberger in 1993 described the role of symmetry. Sweetness is elicited through a bilaterally symmetrical and concentrated dipolar interaction between glycophore and the receptor. Also, the role of partial molar and specific volumes has been explored in sweet taste chemoreception. Some scientists focused on the role of water that could be the vehicle by which stimulus molecules are transported to the receptor environment and oriented towards it. Recent studies suggest that there is a different transduction mechanism for natural sugars and synthetic sweeteners that could be activated by different G proteins from those activated by sucrose. Although the sweet taste may be present in hundreds of compounds, fewer than 30 are permitted for use. The sensory reference standard for sweet taste is sucrose. No other sweetener is perceived to be identical; for example, glucose produces moderate burning and a bitter taste. Humans are more sensitive to synthetic sweeteners. A comparison of the sweet taste of synthetic sweeteners with that of sucrose is given in Table 3. Some synthetic sweeteners have a bitter and metallic taste, so they are used in mixtures.

Bitter Taste

Bitter taste is closely linked to sweet taste. Some compounds can have both a sweet and bitter taste, and slight structural modifications can alter their intensities. Bitter compounds require polar (electrophilic or nucleophilic) and hydrophobic groups and must be at least slightly soluble in water. Among the best-known bitter substances are alkaloids (often toxic) such as quinine, caffeine, and strychnine.

Table 3 Relative sweetnesses for several artificial sweeteners compared with 3% sucrose solution

Sweetener	*Date of discovery*	*Relative sweetness (sucrose = 1)*
Saccharin	1879	300–500
Dulcin	1883	70–350
Cyclamate	1937	30–60
Aspartame	1965	100–200
Acesulfame-K	1967	130–200
Sucralose	1974	600
Alitame	1979	2900

Alkaloids are often found in many plants, and so bitter taste is thought to be a mechanism for poison avoidance by animals. Also, salts that contain the cesium or rubidium cation or the iodide anion are bitter. Other bitter-tasting compounds include peptides, amino acids, ureas, thioureas, terpenoids, and polyphenols. The ability to perceive bitter substances has been related to their lipid solubility, which can be altered by pH. Bitterness is perceived towards the back of the tongue, but some compounds are perceived also on the tip. Transduction of bitter tastants can involve multiple mechanisms, depending on the structure of the bitter compound. One group of bitter compounds may be detected by several different mechanisms. The existence of more bitter receptors is confirmed by a single Mendelian gene, which affects the ability to taste one particular class of bitter compounds. Recently, remarkable progress has been made in establishing the nature of bitter taste. Two groups of scientists have confirmed the existence of bitter taste receptors. They identified a large family of G-protein-linked receptors (40–80 members). These receptors are organized in clusters.

One of the most interesting bitter compounds that has been studied in the past decade is 6-n-propyl-2-thiouracil (PROP). Individuals detect this compound with very different sensitivities that could be inherited.

Salty Taste

Salty taste is the last of the basic tastes that a newborn child perceives (at approximately 4 months). It is mainly associated with the taste of NaCl. At first, it was thought that the saltiness of NaCl is caused by Cl^- anions, but later, scientists preferred the theory that saltiness is produced by Na^+ cations and that anions modulate the intensity of the saltiness (larger anions reducing saltiness more than smaller anions). The molecular weight of the cation does not influence the saltiness, but anions with a low molecular weight are predominantly salty, whereas those of a higher molecular weight tend to be bitter. The salts of heavy metals such as mercury have a metallic taste. The degree of saltiness decreases in the following order: ammonium (most salty), potassium, calcium, sodium, lithium, and magnesium salts (least salty). This order can vary for different animals. Some salts in low concentration can taste sweet, NaCl tastes sweet in concentrations below 0.04 M, KCl below 0.02 M, and in some salts, especially lead acetate and beryllium chloride, the sweet taste dominates. It was found that people are more sensitive to NaCl at the tip than on the more posterior region of the tongue and that NaCl taste thresholds are systematically related to the size and region of tongue stimulated.

Although sodium chloride is essential for life and is the second most-used additive, it is believed to contribute to hypertension and some cardiovascular diseases. Therefore, scientists are looking for possible substitutes. Other chlorides have a salty taste but cannot be used as a good substitute because of their bitter and sour taste or toxic effects. Recently, peptides with a salty taste (L-ornithyltaurine, glycine methyl, or ethylester hydrochloride) and also different dipeptides were examined. The food industry tries to avoid using salt by replacing it with herbs, spices, and organic acids.

Sour Taste

Sour taste is acceptable when mild but becomes unpleasant when strong. It helps us to avoid unripe fruit and damage our tissue with acids. In humans, acids induce salivation, which increases bicarbonate secretions, the major buffering agent in saliva.

From a chemical point of view, acids can be distinguished by their pH index (describes the activity of protons in a solution) as a weak pH 3–7 and strong pH 1–3. We cannot use this distinction when we look at sour taste. Weak organic acids (e.g., the acetic acid in vinegar) taste more sour than mineral acid at the same pH.

The perception of sour taste is influenced not only by the activity of the proton, but also by the quality and character of the anion. The role of anion has not been studied in detail at receptor level. Sour taste in food is mainly connected with citric, malic, oxalic, and tartaric acids in fruits and lactic acid in yogurt and other dairy and meat products. Propionic acid is found in Emmental cheese, and acetic acid is used in the canning industry.

Umami

In 1909, the Japanese scientist Ikeda described a special delicious or savory taste and used the Japanese word 'umami.' This taste is familiar in the East, where seaweed and mushroom are used as ingredients,

because they have a high content of glutamate. This taste is sometimes called a savory or meat-like taste, because meat also has a high content of glutamate. Extensive scientific tests carried out in the twentieth century indicated that the umami taste is different from the four classical tastes and that it is connected with the L form of monosodium glutamate (MSG). Later, nucleotides such as guanosine-5′-monophosphate (GMP) and inosine-5′-monophosphate (IMP) were also considered to have umami flavor.

The occurrence of peptides (around 30), usually dipeptides and tripeptides, with a possible umami taste has also been reported, but this has not been confirmed by other scientists. Some dipeptides have a salty or bitter taste, or their taste is described as a salty/umami or sour/umami taste.

Faurion in 1991 found that the MSG taste is recognized by more than one type of glutamate receptor. Recently, the genes for proteins that serve as the molecular receptors for umami taste were identified.

Other Tastes

Other tastes such as metallic, soapy, alkaline, and astringent are described for the use of sensory evaluation but have not been studied at the receptors level.

Astringency taste Astringency is a tactile taste felt as a dry, rough feeling in the mouth and contraction of the tongue tissue. It usually involves the formation of aggregated precipitates between tannins or polyphenols and proteins in the saliva. Tannins positively influence the taste of tea, coffee, cocoa, wine, and beer but have a negative effect in immature fruit (bananas). Lawless and co-workers developed a vocabulary for astringency taste to describe its multiple perception characteristics. They defined astringency as a combination of three tastes: drying (the lack of lubrication or moistness), roughing (the rough texture in the mouth) and puckery (the drawing or tightening sensation). There are also specific terms such as sappy, harsh, woody, powdery, and gritty that are used for describing astringency taste in wine and beer.

Metallic taste Metallic taste is elicited by some metal salts, as a standard is used iron (II) sulfate heptahydrate in concentration $0.00475\,g\,l^{-1}$. The taste threshold for copper (using copper sulfate and copper chloride) in different types of drinking water is $2.4–3.8\,mg\,l^{-1}$, depending on the type of water.

Alkaline taste An alkaline, soapy taste is associated with potassium carbonate (potash), a substance that was used for the production of soap.

Pungency The trigeminal nerve is stimulated by chemical-induced sensations such as pungency. A pungent taste is characterized by spices such as hot pepper (in which the pungency substance is the alkaloid piperine), ginger, Spanish red pepper (capsaicine), and clove (eugenol). Vegetables from the *Brassicaceae* family such as horse-radish, black and white mustard, and cabbage have a pungent taste owing to their isothiocyanates content. Pungency is also the main sensory attribute of benzoic acid (3–36 mM). Sometimes, other sensations are described, such as the bite of carbon dioxide, the sting of horseradish, and other sensations derived from spices.

Factors Influencing Perception of Taste

Color

Several studies have examined the effect of color on taste perception. The results depend on conditions, concentration of taste substances, and intensity of color. Generally, red and orange increase the threshold for bitter, sour, and sweet taste, and green and yellow increase the threshold for sour. The effect of color on salty taste has not been studied.

Viscosity

The influence of viscosity has been studied mainly for sweet compounds, but it is clear that it depends on the taste substances and thickening agents. Viscous solutions can suppress taste because of a lower diffusion rate and poorer access to the receptors, but they can enhance the taste because the time during which the compounds bind to the receptors is longer. For example, in a chocolate drink, sweetness increases with increasing concentrations of lecithin. Usually, the sweetness of sucrose is lower in liquid than in aqueous solutions. The threshold of pure substances in aqueous solutions is different from the threshold values in other liquids.

Temperature

Temperature effects must be studied separately for every taste and in a wide range of temperatures. Sweet and bitter tastes are more affected by temperature than salty or sour tastes. Increasing the temperature usually increases the sweetness of the solution. For some compounds, a V-shaped function has been described, meaning that the minimum is near the center of the temperature range. For example, the sensitivity for NaCl was greatest in the range of 22–37 °C and lower at 0 and 55 °C. The optimum temperature for the perception of bitter taste is 10 °C, and the bitter aftertaste is longer than for

other tastes. The sweetness of sugars is reduced in cool solutions and the optimum is 35–50 °C.

Odors

Smell and taste are closely related. Therefore, several studies have concentrated on the relationship between perception of an odor and the taste of the solution to which the odor is added. It has been confirmed that flavorings added to a solution of sucrose increase or suppress the perception of sweet taste. For example, the flavor of caramel, marajuca, and strawberry enhance sweetness, whereas angelica oil and damascone suppress sweetness. Also, caramel odor suppresses the sourness of citric acid.

Interactions Between Basic Compounds

Food is a complex of chemical substances, and therefore, it is important to study not only the individual responses of the basic taste but also the responses and interactions in a mixture. Although this problem has been studied extensively and in a previous review, it is difficult to summarize general conclusions. Usually, only binary mixtures in aqueous solutions are studied, and their results cannot be extrapolated directly to food.

Moreover, the interaction between specific chemicals in a binary mixture must be studied also, because compounds from one taste group, such as bitter compounds, do not behave in the same way. Also, all interactions must be studied over different concentration ranges. An important role is played not only by the type of mixture, but also by the type of experimental work and differences in human sensitivities. The interactions in mixtures can be linear – suppression (perceived as less intensive), enhancement (perceived as more intensive) – and non-linear – masking and synergism.

The results can be summarized as follows:

- salts and acids enhance each other at moderate concentrations but suppress each other at higher concentrations;
- some salts enhance the sweet taste (potassium acetate, potassium chloride, sodium chloride);
- sour substances inhibit sweetness and vice versa;
- bitter compounds and acids can either enhance or suppress each other, depending on the concentration;
- bitter taste is generally suppressed by sodium salts, and saltiness is unaffected;
- bitter and sour tastes are suppressed by nucleotides, and salty and sweet tastes are enhanced by nucleotides;
- umami substances in water solution do not influence basic taste: there is only synergism between umami substances.

Examples of Some Interesting Compounds

Caffeine and other methylxanthines can enhance some sweeteners with a bitter component (acesulfam-K, thaumatin, steviosid, sodium saccharin), but not compounds such as aspartame, sucrose, fructose, and calcium cyclamate.

Certain amino acids, including D-phenylalanine, L- and D-histidine, and L-arginine are also enhanced by caffeine. Sweetness is inhibited, for example, by sodium (+)-2-(methylphenoxy)phenoxy propionate, triterpene saponins (gymnemic acid, holodulcin, ziziphens), cyclohexyl acetic acid, indole acetic acid, and α-naphthol sulfate.

Molecules that inhibit more than one taste are called multisaphoric molecules. The glycoproteins miraculin (without taste) and curculin (sweet taste) have a special property, because they convert the taste of acids from sour to sweet.

Cause of the Differences in Feeling Taste

There are more reasons why people differ in their perception of taste. Some can be explained by differences in anatomy and psychology among people, others by some illnesses or by the process of aging. The most-studied factor that influences taste perception is aging. Among older people at about the age of 60, there is normally a decline in the sense of taste, especially sour and bitter.

Disorders associated with taste are less common than with smell, because taste is innervated by three cranial nerves compared with a single cranial nerve for olfaction. Ninety per cent of patients complaining of a reduction or loss of taste had a reduction or loss of smell. This was confirmed by findings that people who had lost their sense of smell (anosmic) can distinguish between four basic tastes. Changes in the sense of taste can be either temporary or permanent.

Decreased perception of taste is called hypogeneusia. This loss was initially thought to be due to a reduced number of taste buds, but recent data indicate that a decline in the number of receptors may be responsible. A persistent taste in the mouth in the absence of stimulation is called dysgeusia, and complete loss of the sense of taste is called ageusia (gustatory anesthesia).

Loss of the sense of taste can be due to acute and subacute diseases: seasonal allergies, head injury, acute viral illness, idiopathic causes, and also the case of hormonal disorder, tumors in the brain, cancer, disease of the gums or teeth, nerve damage,

Table 4 Values of detection thresholds for individual tastes

Sweet	*Threshold (mol dm^{-3})*	*Salt*	*Threshold (mol dm^{-3})*	*Sour*	*Threshold (mol dm^{-3})*	*Bitter*[a]	*Threshold (mol dm^{-3})*
Saccharose	0.011–0.017	Sodium chloride	0.03	Citric acid	0.00078	Caffeine	0.00002–0.002
Lactose	0.065–0.072	Calcium chloride	0.017	Acetic acid	0.00183	Quinine	0.0000001–0.00001
Maltose	0.038	Ammonium chloride	0.004	Tartaric acid	0.00053	(+)-Catechin	0.002
D-glucose	0.065–0.08	Sodium fluoride	0.005	Malic acid	0.00082	L-Leucine	0.015
Saccharin	0.000000023	Magnesium chloride	0.015	Hydrochloric acid	0.0078	Magnesium sulfate	0.005

[a]A range is given because the values for bitter taste differ between authors.

and so on. Burning mouth syndrome is an oral pain disorder occurring primarily in women who have difficulties with identifying salty and sour tastes. The intake of certain drugs (for example anesthetics such as pantocaine and novocaine) and exposure to chemical substances can also influence our perception of taste. For example, chlorhexidine, an antiseptic, can decrease the bitterness of quinine, and it is the only known blocker of the human salty taste.

Some people cannot taste the bitter substance phenyl thiocarbamide (taste blindness), and people with congenital idiopathic hypoparathyroidism cannot recognize sweetness.

Methods of Sensory Analysis

The basis of sensory analysis is the perception of food attributes through the direct use of human senses. The main tool, therefore, is an assessor, who must be selected and trained. There are standardizing methods for training, but it is necessary to check, refresh and update the performance of assessors. Sensory methods must be standardized on international, regional, or national scales. Of all the sensory methods, the most useful for studying taste are detailed below.

Methods for Investigating Sensitivity of Taste

Taste is studied by different methods of sensory analysis that require good methodological and experimental control. The International Standard ISO 3972 describes the method of investigating sensitivity of taste. First, the assessors identify the taste of reference substances (for example, the concentration for reference sucrose is 5.76 g l^{-1}, and for caffeine 0.195 g l^{-1}). It has been proven by practical tests that these concentrations are detected by 50% of novice assessors. Another approach is to determine the threshold. We recognize three types: detection threshold (stimulus threshold) is the minimum value of a sensory stimulus needed to bring about a sensation. In a recognition threshold, accessors must identify this sensation, and the difference threshold is a value of the smallest perceptible difference in the physical intensity of a stimulus. The threshold values from different studies differ between authors, but the results also vary between individuals and the results from one individual can vary, depending on their health condition and mood, and so on. Among all tastes, bitter generally has the lowest detection threshold for naturally occurring compounds (see Table 4). The threshold is usually measured by using pure substances in aqueous solutions. The temperature of solutions and aqua used must be specified, because there are differences between values obtained from distilled water, tap water, and deionized water.

Time–intensity Technique (TI)

The time–intensity technique is the only technique to measure changes in perceptual characteristics with time. The idea of measuring sensation intensity originated in the 1960s but now has become widely used with the development of computerized systems for data collection.

The concept of TI is founded on the assumption that intensity and time are two independent dimensions of sensory space. During measurement, curves are produced, with time on the horizontal axis and sensation on the vertical axis, and these curves can indicate how long it takes to reach the maximum intensity for that taste, the overall maximum intensity, the total duration of sensation, the rate of extinction, and the rate of appearance.

New Analytical Methods for Determination of the Taste

Sensory analysis is a time-consuming and expensive method that requires trained assessors or experts. Attempts have been made to replace this method by various chemical and instrumental methods. An 'electronic tongue' has recently been developed to measure taste (based on pulsed voltametry). The Japanese electric tongue is composed of several lipid/polymer membranes that transform the information of taste substances into an electric signal.

Analytical methods can be useful but still are less objective than sensory analyses, because the former measure only stimuli and not sensations.

See also: **Elderly**: Nutritionally Related Problems; **Sensory Evaluation**: Sensory Characteristics of Human Foods; Food Acceptability and Sensory Evaluation; Practical Considerations; Sensory Difference Testing; Sensory Rating and Scoring Methods; Descriptive Analysis; Appearance; Texture; Aroma; Taste; **Spices and Flavoring (Flavouring) Crops**: Properties and Analysis; **Sucrose**: Properties and Determination; **Sweeteners**: Intensive; **Taste Enhancers**

Further Reading

Breslin PAS (1996) Interaction among salty, sour and bitter compounds. *Trends in Food Science & Technology* 7(Special Issue on Flavour Perception).

Cardello AV (1996) The role of the human senses in food acceptance. In: Meiselman HL and MacFile HJH (eds) *Food Choice, Acceptance and Consumption*, pp. 1–82. London: Blackie Academic.

DuBois GE, Walters DE and Kellogg MS (1993) Mechanism of human sweet taste and implication for rational sweetener design. In: Chi-Tang Ho and Manley CHH (eds) *Flavour Measurement, Basic Symposium Series*, pp. 240–266. Chicago, IL: Institute of Food Technologists.

Erickson RP (2000) The evaluation of neural coding ideas in the chemical senses. In: *Physiology and Behavior*, vol. 69, Issues 1, 3–13. New York: ElsevierScience.

Grenby TH (1989) *Progress in Sweeteners*, pp. 1–394. London: Elsevier Applied Science.

Halpern BP (1997) Psychophysics of taste. In: Beauchamp GK and Bartoshuk L (eds) *Tasting and Smelling*, pp. 77–123. San Diego, CA: Academic Press.

Kawamura Z and Kare MR (1987) *Umami. A Basic Taste*. New York: Marcel Dekker.

Lawless HT and Lee CH (1993) *Taste*. In: Acree TE and Teranishi R (eds) *Flavour Science, Sensible Principles and Techniques*, pp. 23–131. Washington, DC: American Chemical Society.

Mistretta CM and Hill DL (1995) Development of the taste system: basic neurology. In: Doty RL (ed.) *Handbook of Oflaction and Gustation*, pp. 635–668. New York: Marcel Dekker.

Nabors LB and Gelardi RC (eds) (1991) *Alternative Sweeteners*, pp. 1–459. New York: Marcel Dekker.

Pangborn RM (1987) Selected factors influencing sensory perception of sweetness. In: Dobbing J (ed.) *Sweetness*, pp. 49–66. London: Springer.

Piggott JR (1988) *Sensory Analysis of Foods*, 2nd edn., pp. 1–425. London: Elsevier Science.

Rouseff RL (1990) *Bitterness in Foods and Beverages*, pp. 1–356. Amsterdam: Elsevier.

Shallenberger RS (1993) *Taste Chemistry*, pp. 1–570. Glasgow: Blackie Academic & Professional.

Walters DE, Orthoefer FT and Dubois GE (1991) *Sweeteners: Discovery, Molecular Design and Chemoreception*, pp. 1–333. Washington, DC: American Chemical Society.

Wendin K (1999) *Chemical Senses and Food Quality Perception*, pp. 1–52. Gothenburg, Sweden: The Swedish Institute for Food and Biotechnology.

SEPARATION AND CLARIFICATION

J E Lozano, Plapiqui (Uns-Conicet), Bahía Blanca, Argentina

Background

Every year thousands of different processed food products are introduced in supermarkets. Food processing has three major aims: (1) to make food toxically and microbiologically safe; (2) to provide products of the highest nutritional and organoleptic quality; and (3) to make foods with ease of use.

Most food-processing comprise a series of physical processes, or unit processes, that can be separated into a number of basic operations, or unit operations, which depend upon coherent physical principles.

The main unit operations usually present in a typical food-processing line, including:

1. Flow of fluid – when a fluid is moved from one point to another by pumping, gravity, etc.
2. Heat transfer – in which heat is either removed or added (heating; cooling; refrigeration and freezing).
3. Mass transfer – whether or not this requires a change in state. Processes that use mass transfer include drying, distillation, evaporation, crystallization, and membrane processes.
4. Other operations requiring energy, such as mechanical separation (filtration, centrifugation, sedimentation, and sieving); size adjustment by size reductions (slicing, dicing, cutting, grinding) or size increase (aggregation, agglomeration, gelation); and mixing, which may include solubilizing solids, preparing emulsions or foams, and dry blending of dry powders (flour, sugar, etc.).

For example, the manufacture of concentrated clarified apple juice includes several unit operations

such as size reduction, pressing, heating, pumping, and evaporating, in sequence.

Separation Processes

Separation technologies, including extraction and sedimentation, are widely used in the food industry and basic knowledge on separation science is required to design a process properly. Separation in food usually implies pumpable suspension of solids in a liquid (water solution of true solutes, like sugars) or wet solid (e.g., fruit pulp). Separations can be achieved on the basis of density or size and shape, and may be classified into settling (or flotation), filtration, and expression methods (Figure 1). Washing, a typical food process, may also be considered a separation process in which soil or some other undesirable component is eliminated by the combined use of water and mechanical methods.

Washing

Washing of foods includes soaking, brushing, and the application of chlorine or water spraying. Heavy spray and rotary brush wash applications may also be required to remove any rot. Many products such as mechanically harvested berries are air-cleaned on mesh conveyors or vibrator passing over an air jet. Washers are conveyor belts or roller conveyors with water sprays, reel (cylinder) type with internal spray (Figure 2), brushes, and/or rubber rolls with or without studs. Vibratory-type washers are very effective for berries and small fruits. Brushes are effective to eliminate rotten portions of fruits, thus preventing problems with yeast toxins.

Settling

Settling is the use of the force of gravity to cause particles to settle down the liquid food. Although settling does not give a complete separation, it is often the best way to process very large volumes of a food suspension and remove most of the liquid. The concentrated suspension can then be filtered with smaller equipment and the cloudy liquid can be clarified if necessary. Settlers can also be used for classifying particles by size or density, which is not usually possible with filtration.

Gravitational separation Separation by the action of gravity, or sedimentation, can be represented by Stoke's law:

$$V_p = \frac{g(\rho_p - \rho_1)D_p}{18\mu} \qquad (1)$$

where V_p = the velocity of particles ($m\,s^{-1}$); ρ_1 = specific gravity of liquid media ($kg\,m^{-3}$); ρ_p = specific gravity of particles ($kg\,m^{-3}$); D_p = mean diameter of particle (m); μ = viscosity of liquid media ($Pa\,s^{-1}$); and g = gravitational constant ($m\,s^{-2}$). Solid–liquid separation by screening is also possible for some food suspensions, including coarse particles (fruit juice from belt presses), but it is not widely applicable.

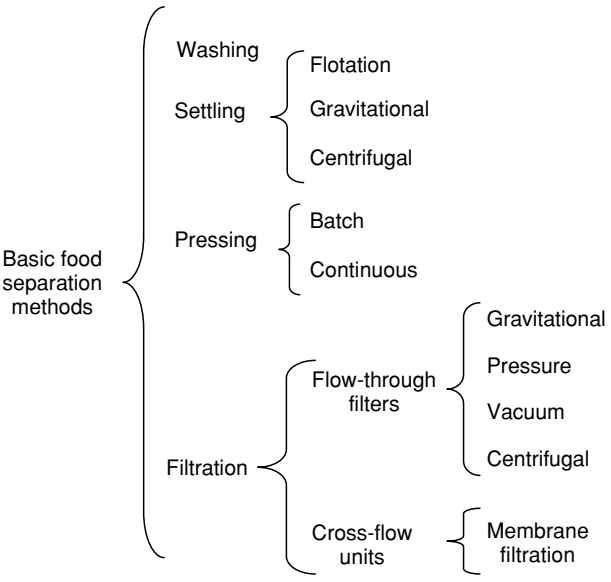

Figure 1 Separation methods commonly used in the food industry.

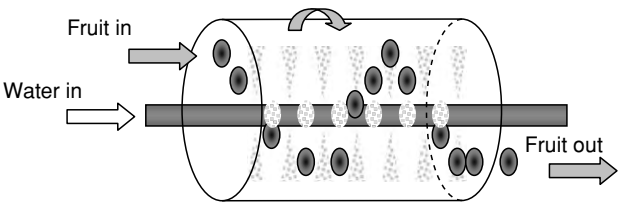

Figure 2 Reel washer with internal spray.

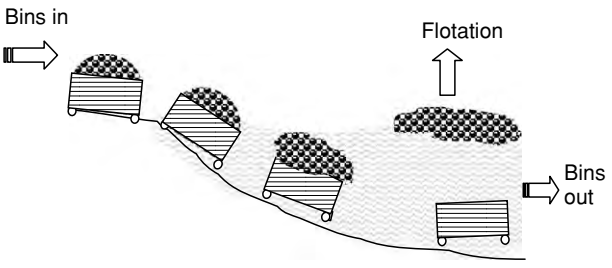

Figure 3 Example of the use of flotation for separation of fruits from container (bins).

Flotation A typical example of flotation as a separation process is unloading fruits into a silo system. Harvesting containers, known as bins, are commonly used to transport fruit from the orchard to the processing plant. Once in the plant, bin dumping unloading and soil and foreign material separation can be performed by flotation (this is good for fruits with density < 1, like apples), as Figure 3 shows.

Centrifugation Centrifugal separation is a process which is quite often used in the food industry. In the dairy industry, for example, uses of centrifugal force include butter oil purification (separation of serum phase from anhydrous milk fat); clarification (removal of solid impurities from milk); separation of fat from whey; skimming and standardizing.

Centrifugation is based on the same Stoke's law presented in eqn (**1**) by changing the force of gravity g by $(r\omega^2)$, where r = radius of axis of rotation; and ω = angular velocity. Stoke's equation (eqn (**1**)) indicates that a fat globule with a diameter of 2 μm will rise four times faster than a fat globule with a diameter of 1 μm.

Modern centrifuges are self-cleaning, allowing a continuous separation/clarification process. Centrifuges have also been used in preparing fruit juice from which sediment and microorganisms can be removed centrifugally in a clarifier, which are disk-type centrifuges that employ forces of 5000–10 000 times gravity.

Pressing

Solid–liquid separation is concerned in this article with mechanical processes for the separation in foods of water (including soluble solids) and finely divided insoluble particles. Expression is the removal of relatively small content of liquid from compressible milled food, by mechanical means. The way to separate most water and soluble solids depends on the food.

For citrus fruits there are three main types of extractors: (1) The FMC-type (FMC Corporation, Lakeland, USA) citrus juice extractor which basically consists of a pair of cups that cut two holes in the fruit, separate the peel, and move the peeled fruit into a strainer tube where the juice is separated from the seeds and the rest of the fruit; (2) the Brown-type extractor (Automatic Machinery & Electronic, Los Angeles, USA); in which fruit is cut in half, the halves placed in rubber cups and the juice removed by rotating plastic reamers; and (3) the rotary press, in which fruit is cut in halves and juice extracted in rotary cylinders.

Most systems for extracting juices from apple and similar fruit and vegetable pulps use some method of pressing juice through cloth of various thickness, in which pomace is retained. These systems are called filter presses and usually come in three main types: (1) rack-and-cloth press; (2) horizontal pack press; and (3) continuous belt press.

1. In a rack-and-cloth press the milled fruit pulp is placed in a nylon, Dacron or polypropylene cloth to form a 'cheese' with the help of a cheese framework. Layers of pulp cheeses, separated by racks made of hardwood or plastic, are stacked up to 1 m or more in height depending on the maturity of the fruit and the size of the racks (Figure 4). Rack-and-cloth presses are efficient but very labor-intensive in operation, cleaning and repairing.
2. Cage presses are horizontal presses with enclosed cages of several cubic meters in which pressing takes place. The cage is filled with a complex filter system consisting of grooved flexible rods filled with sleeves of press cloth material (Figure 5a). Juice passes from the pulp, through press cloth sleeves, along grooves in the flexible rods, and out to collecting channels in the ends of the cage

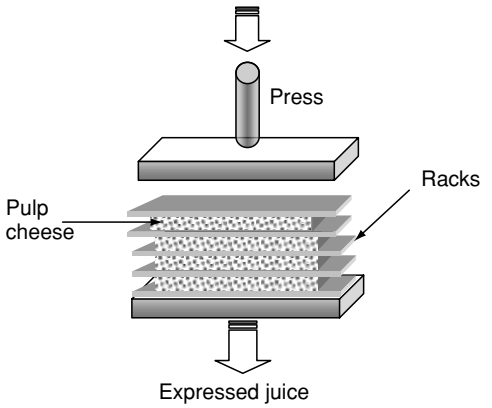

Figure 4 Rack-and-cloth press.

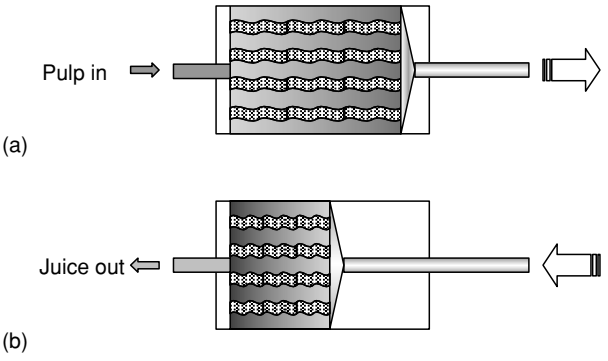

Figure 5 Hydraulic press: (a) loading; (b) pressing/extraction.

and in the piston (Figure 5b). The drum may be rotated, breaking up the pulp, and adding more water, permitting a second pressing with more juice extraction. The whole process may be automated. Although some cleaning labor is saved, rods and sleeves require a considerable amount of maintenance.

3. Continuous belt presses are based in the Ensink design for paper pulp pressing. This type of press offers a truly continuous operation (Figure 6). In belt presses, a layer of mash (pulp) is pumped on to the belt entering the machine. Press aid may be added for improved yield.
4. Screw presses. A typical screw press consists of a stainless-steel cylinder screen enclosing a large-bore screw with narrow clearance between screw and cylinder. Adjustable backpressure is usually provided at the end of the chamber. Breaker bars must be incorporated to disrupt the compressing mash.
5. Diffusion extraction. Typically used for the extraction of sugar from sugar beets (Figure 7), diffusion extraction is a typical countercurrent-type process. It is desirable to retain the same driving force, ΔC (concentration of soluble components in solids vs concentration of soluble components in liquid). In order to maintain a constant ΔC throughout the extraction process, it is necessary to carry out a continuous weighing of feed and to control the water flow to the counterflow extractor by means of a control loop. The diffusion extraction process is influenced by a number of variables (temperature, raw product size, water, etc.) Extracted products are usually passed through a conventional press system and the very dilute juice may be returned to the extractor.

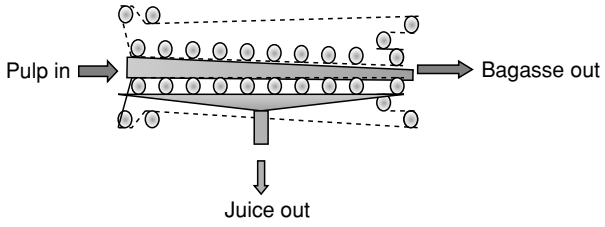

Figure 6 Typical fruit belt press.

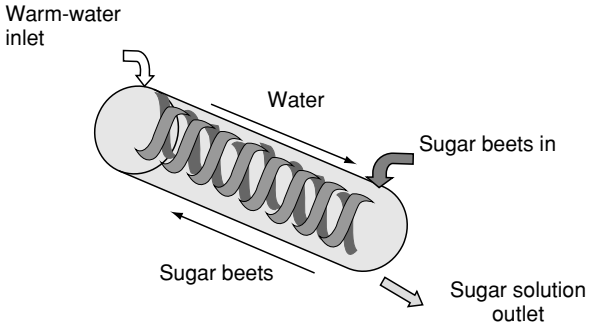

Figure 7 Diffusion extraction process.

Hydraulic pressing does not usually requires the addition of press aids, unless exceptionally overmature fruit is used. For continuous screw presses, however, it is usually necessary to add 1% (w/w) or more of cellulose. Mixing off cellulose and fruit occurs in the mill and subsequent pumping to press.

Filtration

Filtration is also a mechanical process designed for clarification by removing insoluble solids from a valuable liquid food, by the passage of most of the fluid through a porous barrier which retains most of the solid particulates contained in the food. The piece of equipment by which filtration is performed is named the filter. The filter medium, which may be a screen, cloth, paper, or bed of solids, is the barrier that lets the liquid pass while most of the solids are retained. The liquid that passes through the filter medium is called the filtrate. Filtration and filters can be classified in several ways, including:

1. Driving force: The filtrate is induced to flow through the filter medium by:
   c. Hydrostatic head (gravity)
   d. Pressure (upstream filter medium)
   e. Vacuum (downstream filter medium)
   f. Centrifugal force across the medium
2. Filtration mechanism:
   a. Cake filtration (solids are retained at the surface of a filter medium and pile upon one another)
   b. Depth or clarifying filtration (solids are trapped within the pores or body of the filter medium)
3. Operating cycle:
   a. Intermittent (batch)
   b. Continuous rate.

These methods of classification are not mutually exclusive. Thus filters are usually divided first into the two groups of cake and clarifying equipment, then into groups of machines using the same kind of driving force, then further into batch and continuous classes. Some filtering devices usually employed in the food industry are described below.

Pressure filters The main advantages of pressure filtration compared to other filtering methods are

that cakes are obtained with very low moisture content; clean filtrates may be produced by recirculating the filtrate or by precoating; and solutions may be polished (finished) to a high degree of clarity. Among the disadvantages it must be noted that cloth washing is difficult and precoat is required; the operator cannot see the forming cake and is unable to carry out an inspection while the filter is in operation; the internals are difficult to clean, and this may be a problem with food-grade applications.

With the exception of the rotary drum pressure filter, these types of filters are semicontinuous machines in which wash and cake discharge are performed at the end of the filtration cycle. Since the operation is in batches, intermediary tanks are required. The collection of filtrate depends on the operating mode of the filter, which can be constant flow rate, constant pressure, or both, with the pressure rising and flow rate reducing during filtration. The filtration rate is mainly influenced by the properties of food (particle size and distribution, presence of gelatinous solids like pectin, liquid viscosity, etc.) Although continuous-pressure filters are available, they are mechanically complex and expensive, so they are uncommon in the food industry.

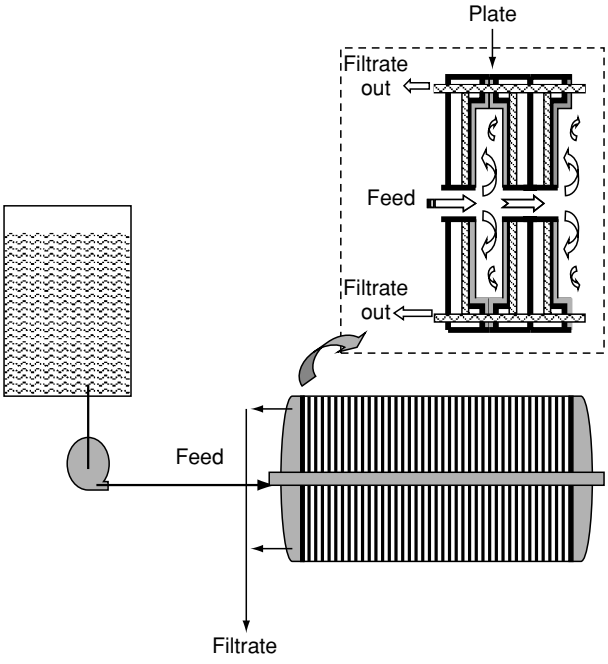

Figure 8 Filterpress operation sketch and details of filtering plate.

Filter Aid and Precoating Filter aid and precoating are often mentioned as in the case of pressing, in connection with pressure filtration. Filter aid is used when the pulp or turbid liquid food is low in solids content, with fine and muddy particles that are difficult to filter. To enhance filtration, coarse solids are used with large surface area that capture and trap in their interstices the slow-filtering particles from the suspension, producing a porous cake matrix.

On the other hand, precoating is the formation on the filter plates of a defined thick medium of a known permeability. Precoating prior to filtration serves when the particles to separate are gelatinous and sticky, forming a barrier that avoids cloth blinding. The filter aids and precoating materials most commonly employed in the food industry are: diatomaceous earth (silicaceous skeletal remains of aquatic unicellular plants); perlite (glassy crushed and heat-expanded volcanic rock); cellulose (fibrous, lightweight, and ashless paper); and special groundwood.

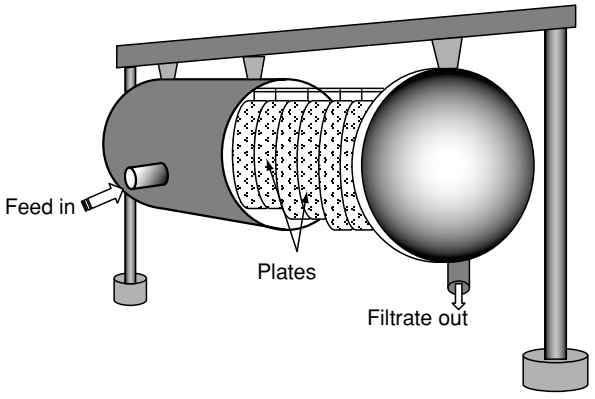

Figure 9 Horizontal plate filter.

Types of pressure filters Pressure filters usually found in the food industry are:

1. Filterpress, also named plate-and-frame, consists of a head and follower contained between a pack of vertical rectangular plates supported by side or overhead beams (Figure 8). The head serves as a fixed end to which the feed and filtrate pipes are connected and the follower moves along the beams and presses the plates together during the filtration cycle by a hydraulic or mechanical mechanism. Each plate is dressed with filter cloth on both sides and, once pressed together, they form a series of chambers that depend on the number of plates. The plates generally have a centered feed port that passes through the entire length of the filterpress, so that all the chambers of the plate pack are connected together.
2. Vertical and horizontal pressure leaf filters consist of a vessel that is fitted with a stack of vertical

(Figure 9) or horizontal leaves that serve as filter elements. The leaf is constructed with ribs on both sides to allow free flow of filtrate towards the neck and is covered with coarse mesh screens that support the finer-woven metal screens or filter cloth that retain the cake. The space between the leaves may vary from 30 to 100 mm depending on the cake formation properties and the ability of the vacuum to hold a thick and heavy cake to the leaf surface. They may be applied to polishing liquid foods with very low solids or for cake filtration with a solids concentration < 20–25%. In contrast to filterpress, pressure leaf filters may be readily jacketed for applications whenever hot or cold temperatures are to be preserved. The cloth mesh screens that cover the leaves may be accessed more easily on horizontal than on vertical tanks.

3. Candle filters may be used in applications that require efficient low-moisture cake filtration or a high degree of polishing (finishing). Candle filters are available in many sizes, containing up to 250 or more filtering elements (Figure 10). The candle filter consists of three major components: the vessel, the filtering elements (candles), and the cake discharge mechanism. Vessel configuration may be with a conical bottom for cake filtration and polishing, or with a dished bottom for slurry thickening. This last vessel is scarcely or never used in the food industry.

The filtering element generally consists of the filtrate core and the filtering medium. The core serves for filtrate passage and to support the filter medium. In the food industry the core is a bundle of perforated stainless-steel tubes. The types of filter medium available are diverse, such as porous ceramics, woven mesh screens, sintered metal tubes, or synthetic filter cloths. Finally, there are two methods of discharging the cake at the end of the cycle: snap blow or vibrating mechanism. The advantages of a candle filter are the excellent cake discharge capacity and its mechanical simplicity.

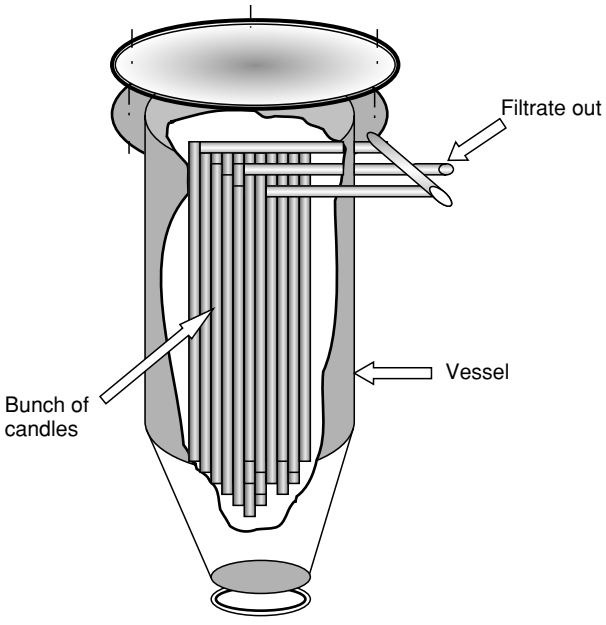

Figure 10 Candle-type filter.

Vacuum filters Vacuum filters are simple and reliable machines and therefore have gained wide acceptance in the food industries. Among the different types and sizes for standard vacuum filters, the following are included: drum; disk; horizontal belt; tilting pan; and table filters.

Drum filters are usually utilized in the food industry. Advantages and disadvantages of vacuum filtration as compared to other separation methods are as follows:

- Advantages: continuous operation; very effective polishing (finishing) of solutions (on a precoat filter); and easy control of operating parameters such as cake thickness.
- Disadvantages: higher residual moisture in the cake and difficult to clean (mainly as required for food-grade applications).

Precoat filters are used when liquid foods (e.g., clarified apple juice) require a very high degree of clarity. To polish the solution the drum deck is precoated with an appropriate medium. A scraper blade, also called the doctor blade, moves slowly towards the drum and shaves off a thin layer of the separated solids and precoating material. This movement continuously exposes a fresh layer of the precoat surface so that when the drum submerges into the tank it is ready to polish the solution. On precoat filters the entire drum deck is subjected to vacuum (Figure 11).

Membrane processes The membrane is an interface between two bulk phases. In the food industry, membrane processes include a wide range of unit operations, from sieving to reverse osmosis. However, during conventional filtration, only the separation of coarse particles to micron range was considered. This operation requires frequent cleaning due to a build-up of a cake layer which results in increased resistance to filtration. For separation of fine solids from liquids, cake filtration or cross-flow filtration should be considered. In this section only cross-flow membrane filtration will be considered. In cross-flow filtration (Figure 12) the bulk phase under pressure flows along

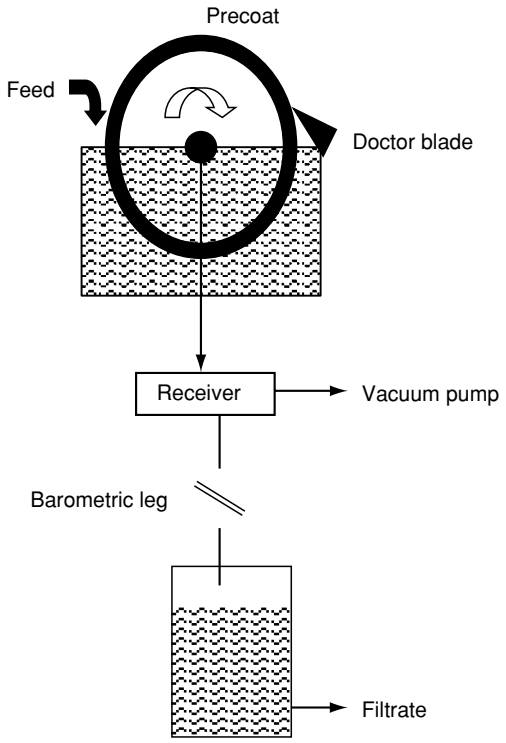

Figure 11 Vacuum drum precoat filter.

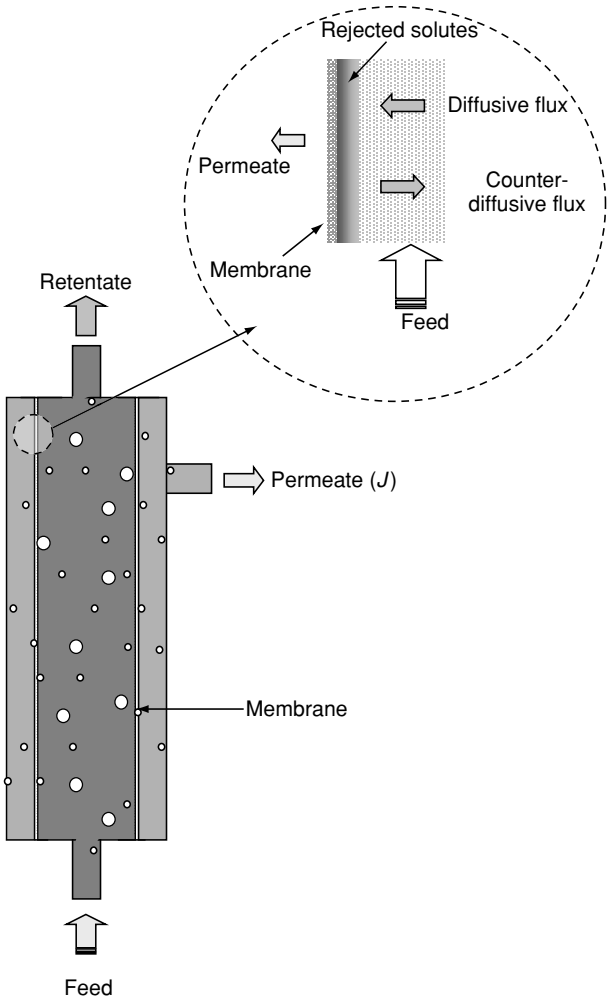

Figure 12 Typical cross-flow membrane process and details of acting fluxes.

the surface of the membrane, sweeping and at least partially removing the cake layer. This allows relatively high permeate fluxes over long periods. Reverse osmosis (RO), ultrafiltration (UF), and microfiltration (MF) are typical cross-flow filtration methods that use membranes with varying pore sizes to effect separation on the basis of size and shape. RO has membranes with the smallest pore and is used to separate water from other solutes.

High pressures (up to 10 000 kPa) are required to overcome the high osmotic pressure of solutes (sugars and salts). UF has membranes with larger pores and will retain proteins, lipids, and colloidal salts, while allowing smaller molecules to pass through to the permeate phase. Required working pressures during UF are in the range 130–1400 kPa. MF, with pores $> 0.1\,\mu m$, is used to separate fat from proteins and to reduce microorganisms from fluid food systems. This process requires the lowest operating pressure (< 350 kPa).

Table 1 lists examples of commercial uses of membrane processes in different food industries. However, there are still a number of technical obstacles that still need to be resolved, including inadequate processing rates (flux) and separation capabilities (rejection and selectivity), insufficient membrane or module life, and incompatible materials.

Clarification

Many liquid foods are colloidal suspensions which must be unstabilized to give a completely transparent liquid. This process is known as clarification (Table 2). Usually, colloidal state is only defined in terms of particle size; the lower limit is generally taken as 1–5 nm (10–50 Å) and the upper size limit as 1–5 µm. When the dispersed phase is a solid and the continuous phase is a liquid, the system is defined as a sol. A typical example of a food sol is a cloudy fruit juice, which is a colloidal system where the continuous medium is a solution of pectin and sugars, and the dispersed (colloidal) matter is mainly formed by cellular tissue comminuted during fruit processing. Pectin is basically a biopolymer made by D-galacturonic acid units.

Particles in a cloudy juice adhere together and form aggregates of increasing size (flocculation) which may

Table 1 Use of membrane processes in different food industries

	Ultrafiltration	*Reverse osmosis*
Beverage industry	Sterilization of draft beer	Juice concentration in fruit wines
	Wine clarification	Tartrate removal
	Fractionation of whey to produce protein concentrate	Concentration of whey
Dairy industry	Fractionation of milk for cheeses	Concentration of milk
Fats and oils industry		Recovery of oilseed protein and tocopherol and oil from waste water
Fruit/vegetable industry	Juice clarification	Sugar concentration
	Juice concentration	
Meat processing	Concentration of egg white, gelatin, and blood serum proteins	
Sugar/sweetener industry	Grain milling	Concentration of maple sap
	Clarification of dextrose	Recovery of sugar from rinse water

Table 2 Fruit juice clarification agents

Name	*Description*
Sparkolloid	A natural albuminous protein extracted from kelp and very finely powdered
Gelatin	Mixture of gelatins and silicon dioxide, with the active ingredient being animal collagen
Kieselsol	Suspension of small silica particles usually used in combination with gelatins. This fining aids in pulling proteins out of suspension
Bentonite	Refined clay (montmorillonite) presented as powder or coarse granules. Bentonite must be added to hot water: stir well and let it stand for 36–48 h before use, so that it swells and becomes almost gelatin-like
Isinglass	Protein produced from sturgeon swim bladders. It is presented either as a fine powder or as dry hard fragments

settle due to gravity. If flocs change to a much denser form, it is said to undergo coagulation which is an irreversible process. For a clear juice, these suspended particles have to be removed. It may seem simple merely to filter them out, but unfortunately some soluble pectin remains in the juice, making it too viscous to filter quickly. A dose of commercial enzyme is the accepted way of removing unwanted pectin. Enzyme depectinization has two effects: it degrades the viscous soluble pectin and it also causes the aggregation of cloud particles. Figure 13 shows how this may happen: pectin forms a protective coat around proteins in suspension. In an acidic environment (apple juice typically has a pH of 3.5), pectin molecules carry a negative charge. This causes them to repel one another. Pectinase degrades this pectin and exposes part of the positively charged protein beneath. The electrostatic repulsion between cloud particles is thereby reduced, so that they clump together.

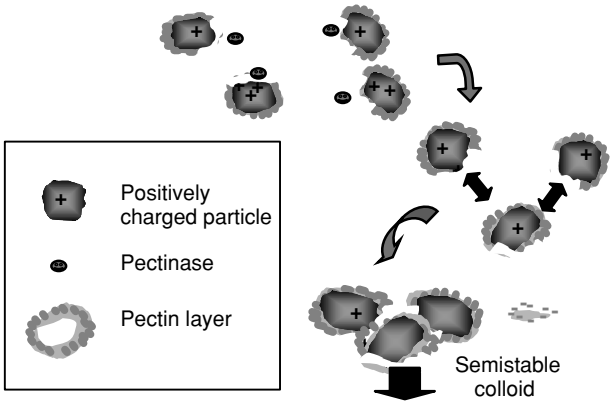

Figure 13 How the breakdown of pectin leads to the formation of aggregated cloud particles during fruit juice clarification.

These larger particles will eventually settle out, but to improve the process, flocculating, also named fining, agents (Table 1) can be added. Fining agents work either by sticking to particles, making them heavy enough to sink, or by using charged ions to cause particles to stick to each other, making them settle to the bottom. What is left is a transparent but by no means clear juice. A second centrifugation and subsequent filtration (fining) are needed to give the clear juice that many consumers prefer.

Another potential contributor to the haziness of juice is starch. Unripe apples, for example, may contain up to 15% starch. Although the first centrifugation (before the juice reaches the clarification tank) removes most of the starch, about 5% usually remains. This can be broken down using a specific enzyme (amyloglucosidase) added at the same time as

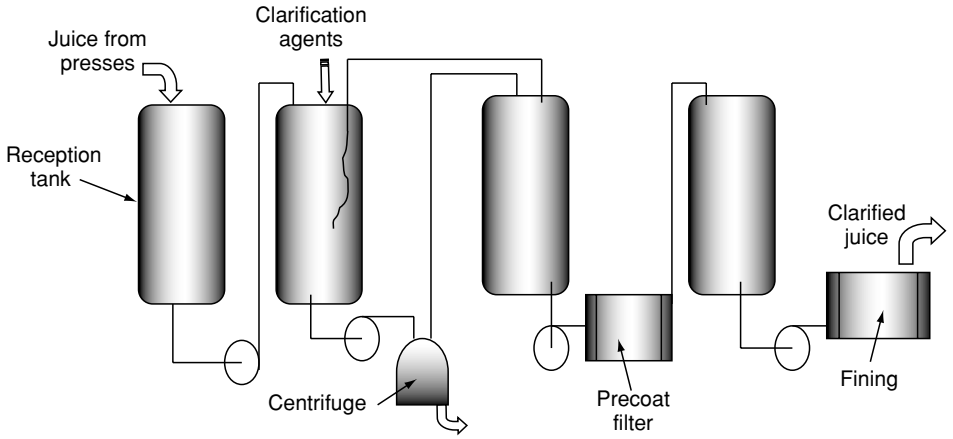

Figure 14 Fruit juice conventional clarification line.

the pectinase. Optimization of fining and ultrafiltration steps can help retard or prevent postbottling haze development. Figure 14 shows a typical fruit juice clarification process.

Prepress treatment with pectinase takes anything from 15 min to 2 h depending upon the enzyme, the reaction temperature, and the fruit chosen. Enzyme treatment is considered to be complete once the viscosity of the juice has returned to its original level or less. It is important that the pulp is not broken down too much. Other polysaccharides (like araban, a polymer of the pentose arabinose) may appear as a haze in fruit juices and other beverages a few weeks after it has been concentrated. Although commercial pectinase preparations often contain arabanase, certain fruits (like pears) are rich in araban and may require the addition of extra arabanase to the clarification tank.

Clarification by Ultrafiltration

The application of UF as an alternative to conventional processes for the clarification of fruit and vegetable juices has been demonstrated. However, the acceptance of UF in the food-processing industry is not yet complete, because there are problems with the operation and fouling of membranes. During UF, two fluid streams are generated: the ultrafiltered solids free juice (permeate), and the retentate with variable content of insoluble solids and macromolecules.

Permeate flux (J) results from the difference between a convective flux from the bulk of the retentate to the membrane and a counterdiffusive flux or outflow by which solute is transferred back into the bulk of the fluid (Figure 12). The value of J is strongly dependent on hydrodynamic conditions, membrane properties, and operating parameters. The main driving force of UF is the transmembrane pressure (ΔPTM), which, for example, in the case of hollow fiber ultrafiltration systems, can be defined as:

$$\Delta PTM = (P_i + P_o)/2 - P_{\text{ext}} \qquad (2)$$

where P_i = pressure at the inlet of the fiber, P_o = outlet pressure, and P_{ext} = pressure on the permeate side. In practice, the J-values obtained with apple juice are much less than those obtained with water only. This phenomenon is attributable to various causes, including the concentration polarization boundary layer (defined as a localized increase in concentration of rejected solutes at the membrane surface due to convective transport of solutes) and plugging of pores due to fouling. Where some of these phenomena are reversible and disappear after cleaning of the UF membranes, others are definitively irreversible.

Flux decline due to this phenomenon can be reduced by increasing flow velocity on the membrane. Traditionally, correlations of J with ΔPTM have been determined by parameter fitting of the experimental data.

As a summary, membrane processes are limited by the following factors:

1. They must be combined with other processes or steps.
2. The process is practically limited to low-viscosity liquids and foods.
3. Ceramic membranes are versatile, but expensive to install.
4. Although the membrane process is easy to understand, very careful design and trained and skilled operators are needed to insure long-term filtering capacity.

A generalized application of membrane technology in the food industry may require improvements in thermal stability and chemical resistance; a more

controlled pore structure and narrow pore size distribution; a higher permeate flux; an improved microbiological resistance; and more controlled fouling.

See also: **Colloids and Emulsions**; **Filtration of Liquids**; **Membrane Techniques**: Principles of Reverse Osmosis; Applications of Reverse Osmosis; Principles of Ultrafiltration; Applications of Ultrafiltration

Further Reading

Cain CA Jr (1995) Filter aids: use in filtration. In: McKetta J (ed.) *Encyclopedia of Chemical Processing and Design*, vol. 21, pp. 348–372. New York: Marcel Dekker.

Cheryan M (1986) *Ultrafiltration Handbook*. Lancaster, PA: Technomic.

Dahlstrom D (1995) Filtration. Vacuum systems. In: McKetta J (eds) *Encyclopedia of Chemical Processing and Design*, vol. 21, pp. 412–432. New York: Marcel Dekker.

Jahreis CA (1995) Filtration. Pressure systems. In: McKetta J (ed.) *Encyclopedia of Chemical Processing and Design*, vol. 21, pp. 380–396. New York: Marcel Dekker.

Letka A, Moll RT and Shapiro L (1997) Separation (centrifugal). In: *Kirk-Othmer Encyclopedia of Chemical Technology*, 3rd edn., vol. 21, pp. 828–875. New York: Wiley.

Loncin M and Merson RL (eds) (1979) Mechanical Operations. In: *Food Engineering. Principles and Selected Operations*, pp. 231–264. New York: Academic Press.

Lozano JE, Constenla DT and Carrín ME (2000) Ultrafiltration of apple juice. In: Lozano JE, Añón MC, Parada-Arias E and Barbora-Cánovas G (eds) *Trends in Food Engineering. Food Preservation Technology Series*, pp. 117–134. Lancaster, Basel, USA: Technomics.

Perry RH and Green DW (eds) (1984) Liquid–solid systems. In: *Chemical Engineering Handbook*. Section 19, pp. 1–109. New York: Mc Graw-Hill.

Svarovsky L (1997) Sedimentation. In: *Kirk-Othmer Encyclopedia of Chemical Technology*, 3rd edn., vol. 21, pp. 667–685. New York: Wiley.

Wallas SM (1995) Solid liquid separation. In: McKetta J (ed.) *Encyclopedia of Chemical Processing and Design*, pp. 457–472. New York: Marcel Dekker.

Shark *See* **Fish**: Introduction; Catching and Handling; Fish as Food; Demersal Species of Temperate Climates; Pelagic Species of Temperate Climates; Tuna and Tuna-like Fish of Tropical Climates; Demersal Species of Tropical Climates; Pelagic Species of Tropical Climates; Important Elasmobranch Species; Processing; Miscellaneous Fish Products; Spoilage of Seafood; Dietary Importance of Fish and Shellfish; **Fish Farming**; **Fish Meal**

SHEEP

Contents

Meat

A J Kempster, Formerly of Meat and Livestock Commission, Snowdon Drive, Milton Keynes, UK

Background

Lamb is the meat from young sheep. It should be distinguished from mutton, which is from adult animals. The two categories together are referred to as sheep meat.

Sheep, raised for both meat and wool production, form a significant part of the economy of many countries. The current world production of mutton and lamb is 6.5 million tonnes (FAO, 1990). Sheep are farmed in all the main regions of the world, but it is notable that 17% occur in Australia and New Zealand and that exports from these countries are the core of world trade.

Sheep farming is possibly the most diverse branch of animal production. There are hundreds of breeds, each with its particular environmental niche and economic function. There is a large variation between the major sheep-producing countries in climatic

conditions, in management procedures and in the breed types available. At one end of the climatic scale, there are the fat-tailed desert sheep, which survive and thrive where other sheep would perish; at the other extreme are mountain breeds like the Scottish Blackface of the UK, which can stand harsh conditions, high rainfall, snow, and indifferent herbage. Most sheep feed outside continuously with only limited supplementation of available pasture, whereas a few, mainly in Europe and North America, are housed from birth and fed carefully formulated diets. There is little mechanization or intensification.

In some countries, sheep-farming systems can cover wide ranges of climatic condition, as in the UK, where there is a stratification system involving the crossing of breeds adapted to particular environments: crosses from hill and upland areas are the female breeding stock for lowland areas where they are mated with sire breeds selected specifically for meat production.

Genetic improvement in the world's sheep population has been very uneven. There are pioneering examples of application of genetics, particularly in Norway, Australasia, and the UK. Published information on sheep breeding is much more a review of the technology about to be applied, rather than a record of achievement.

Progress is slow because environmental, technical, and economic conditions prevailing in areas where sheep are kept are frequently difficult. Breeding objectives are often not clear: indeed, there may be direct conflicts between adaptation for survival and for productivity. Relatively low reproduction rates and long generation intervals do not help. The multipurpose use of sheep for meat, wool, and milk also reduces the selection pressure for meat characteristics under some circumstances.

Sheep meat production is dominated by two factors: first, there are the biological and agricultural considerations of resource use; second, there are the marketing considerations, in particular the ability of farmers to adapt output to conform more closely with changing market demand. In developed countries, this demand is for an increasingly lean product.

Carcass Quality

The value of sheep carcasses depends on several factors, namely weight, conformation (carcass shape), proportion of the main tissues (muscle, fat, and bone), distribution of these tissues through the carcass, muscle thickness, and meat quality. (*See* **Meat**: Structure.)

The weight and size of the carcass have a major influence not only on the quantity of the various tissues but also on the size of the muscles exposed on cutting and of the individual joints prepared from it. This is of importance particularly in relation to a retailer's ability to provide cuts of a suitable size to meet customer requirements. Over generations, wholesalers and retailers from different regions of countries and different parts of the world have become accustomed to handle certain weight ranges of carcasses; abattoir practices and cutting methods have been developed accordingly. Most wholesalers state the desired weight ranges in buying schedules and apply discounts to carcasses outside these ranges. In some cases, these discounts are severe and are a constraint on the use of improved breeds and production systems. There has, for example, historically been a demand in the UK for small legs suitable for the Sunday roast, which has slowed the movement to heavier carcasses that are likely to be more efficient to produce and process. This demand has also had a major effect on the type of carcass produced in New Zealand for the British market. (*See* **Meat**: Slaughter.)

Although conformation (the shape of the carcass) provides a poor indication of carcass composition, it is still regarded by many in the meat industry as valuable in this respect. Carcasses with good conformation normally command higher prices, and most national classification and grading schemes include conformation as a factor. It is particularly important in some European markets.

Among carcasses of similar weight, the percentage formed by each tissue varies considerably depending on the breed type and level of feeding. The proportion of lean meat in the carcass is of major importance, since this is the principal determinant of yield and commercial value in many countries. Taken as a generalized ideal, the best carcasses should have an optimum level of fatness, sufficient to ensure that carcasses do not dry out and to ensure good eating quality, and minimum bone.

The general trend is now towards the production of leaner carcasses because of consumers' increasing demand for meat with a low fat content. Table 1 lists the typical composition for carcasses in different fat classes of the National Carcass Classification Scheme operated in the UK (this can be linked directly to the developing EC scheme). (*See* **Meat**: Nutritional Value.)

Definition of Retail Cuts

Traditionally, lamb is sold by butchers largely bone-in, so about 90% of the carcass is 'saleable meat' and goes across the counter.

The lamb carcass is typically separated into the sides by cutting it lengthwise through the spine. The main cuts (using British terminology) are as follows.

Table 1 Composition of carcasses in different fat classes of the British classification scheme

	Fat class[a]					
	1	*2*	*3L*	*3H*	*4*	*5*
Lean meat in carcass (%)	62.4	58.7	56.1	54.3	51.7	48.1
Separable fat in carcass (%)	16.5	21.7	25.6	28.2	32.1	37.3
Total protein in carcass (%)	14.2	13.3	12.7	12.2	11.4	10.2
Lipid in lean (%)[b]	3.6	5.1	6.2	6.9	8.1	9.6

[a]Most carcasses fall into fat class 3L. Fat classes 4 and 5 would be too fat for most requirements.
[b]Intramuscular fat.
Data from Kempster AJ, Cook GL and Grantley-Smith M (1986) National estimates of body composition of British cattle, sheep and pigs with special reference to trends in fatness. A review. *Meat Science* 17: 107–138.

The *forequarter* is a large cut which includes the neck, shoulder, and part of the breast. It makes an economical family roast but is more difficult to carve than a leg. Boned and rolled or stuffed, it is easier to carve.

The *shoulder* is smaller than the leg and is easier to carve after boning. Shoulder chops or cubed shoulder meat should be well trimmed to remove excess fat before grilling or stewing.

Neck chops may be stewed, braised, or casseroled.

Breast is an economy cut. The ends of the rib bones can be cut out to simplify carving or rolling.

The *best end of neck* or *rack* is made up of six or seven rib chops. When it is to be cooked whole, it may be boned and rolled, or chined (backbone removed) for easy carving. The rib chops are trimmed to make cutlets.

The *loin* can be cut into about seven or eight meaty chops, each with a short T bone; alternatively, it can be left in one piece and rolled or trimmed, boned, rolled, and cut into noisettes before cooking.

Chump chops (or *leg chops*) are lean, meaty chops cut from the end of the leg nearest the loin.

The *leg* is a large, lean roast which may be cooked with the bone in, or boned and stuffed. Meaty steaks can be cut from the thick part of the leg.

Consumer attitudes indicate that lamb is often regarded as fatty, not versatile, and difficult to carve. Major developments have, therefore, taken place in new cutting methods. This work produces boneless cuts which are convenient in size, easy to carve, and lead to minimal waste on the plate.

The new techniques only work satisfactorily on lambs which are quite lean at the outset, since it is difficult to trim the intermuscular fat without adopting a complete muscle seaming technique. If there is increasing use of these boneless techniques, it will create further pressure towards the production of heavier and leaner lambs.

Lamb and mutton are used less in processed form than other meats. One of the main reasons is the high bone-to-meat content of the carcass and the small size of the carcass. These two factors make it more labor-intensive for boning-out and processing and therefore more costly relative to other meats. Lamb fat is also more saturated and less suitable for processing.

The main area of development in the UK at the moment is in kitchen-ready products which offer variety, convenience, and novelty to consumers. The range of products now being introduced in the UK, particularly by independent butchers, includes lamb burgers, marinated chops and steaks, and more exotic dishes such as lamb bengali, lamb pasanda, and lamb italliene. Lamb is the basis for many spiced dishes on a worldwide basis, and quite a number of these are being introduced into Europe by the major supermarket chains.

Recently, there has been renewed interest in the curing of lamb because of its cheapness in some countries relative to pigmeat. (*See* **Meat:** Preservation.)

See also: **Meat**: Structure; Slaughter; Preservation; Nutritional Value

Further Reading

Food and Agriculture Organization of the United Nations (1990) *Production Yearbook, 1989*, vol. 43. Rome: FAO.

Kempster AJ, Cook GL and Grantley-Smith M (1986) National estimates of body composition of British cattle, sheep and pigs with special reference to trends in fatness. A review. *Meat Science* 17: 107–138.

Milk

M Juarez, Instituto del Frio (CSIC), Madrid, Spain
M Ramos, Instituto de Fermentacinos Industriales (CSIC), Madrid, Spain

Background

Although the production of sheeps' milk (approximately 8 million tonnes) is of marginal importance

compared to cows' milk in quantitative terms (2% of the total), it is of major economic relevance in a numbers of areas such as the Mediterranean and Middle East, where the climatic conditions are not favorable for raising cattle. The number of sheep does not fully reflect the amount of milk produced, since sheep are often used for other purposes, such as meat and wool. In general, sheeps' milk is used essentially for cheesemaking, but in some countries, some is made into yogurt or whey cheeses.

Chemical Composition

Milk is a characteristic secretion of mammals designed to meet the complete nutritional requirements of the neonate of the species. It is a mixture, in equilibrium, of proteins, fat, lactose, and minerals in different states of dispersion in water, such as emulsions, colloidal suspensions, and true solutions.

Sheeps' milk is more viscous than cows' milk and is more resistant to the proliferation of bacteria during the first hours after milking, owing to its greater immunological activity. It has a unique smell, characteristic of the animal that produces it, and is an opaque white in color.

The variability in the composition of sheeps' milk is due to genetic, physiological, and environmental factors. The principal factor is breed; breeds selected for a high milk yield produce milk with a lower total solids content. Fat shows more variation between breeds than proteins. Another factor to consider is the stage of lactation. Colostrum contains high levels of fat (10.7–17.2%), proteins, mainly as immunoglobulins (7.1–19.5%), and low levels of lactose (1.6–4.0%). The fat levels of protein in the milk decline during the first weeks postpartum and then increase gradually after the first month until lactation ends, and the yield falls off. At the end of 7 months' lactation, fat values may be as high as 8.1–10% and protein as high as 6.8–8.9%. The lactose content falls slightly throughout lactation and more sharply at the end (3.4–4.1%). Another factor having a notable influence on milk composition, especially the fat content, is the feeding of the livestock. Other influencing factors include the age and health of the animal, season, and milking times and procedures. Table 1 shows the gross composition of sheeps' milk. Sheeps' milk contains higher concentration of fat, proteins, mineral salts, and total solids than cows' milk. The lactose content is comparable in the two species. Fat and protein are the principal components of the, total solids, constituting 69% in sheeps' milk, compared to 56% in cows' milk. The fat-to-protein ratio is higher in sheeps' than in cows' milk, and the cheese yield expressed as kilograms of cheese per 100 l of milk is approximately 18–20 kg, compared to 10 kg for cows' milk.

Nitrogen Compounds

Sheeps' milk contains 0.7–1% nitrogen. This is distributed in fractions the importance of which varies in terms of dairy technology and nutrition. Proteins account for approximately 95% of total nitrogen, while 5% is nonprotein nitrogen. Milk proteins occur in two distinct phases. One is an unstable micellar phase composed of caseins, which is suspended as micelles averaging about 190 nm in diameter, interlinked by calcium phosphate and small amounts of magnesium, sodium, potassium, and citrate, which diffuse light and give the milk its opaque white appearance. The other is a soluble phase composed of whey proteins. The caseins precipitate at pH 4.6 at room temperature, whereas under the same conditions, the whey proteins (β-lactoglobulin, α-lactoalbumin, and serumalbumin) remain soluble.

Caseins Caseins are the principal proteins in sheep's milk (76–83% of total proteins) and are present in most types of cheese. The caseins are of four types of polypeptide chain, αs1-CN, αs2-CN, β-CN, and

Table 1 Composition of ewes' milk

Component	*Average content (%)*	*Range (%)*	*Average of dry matter (%)*
Water	81.60	79.27–83.80	
Lactose	4.61	4.10–4.95	25.00
Fat	7.09	5.10–8.70	38.50
Crude protein (total nitrogen × 6.38)	5.72	4.75–6.60	31.10
Casein	4.44		
Whey proteins	0.98		
Nonprotein nitrogen	0.05		
Ash	0.91	0.70–1.10	4.90
Total solids	18.40	16.20–20.73	
Solids–nonfat	11.31		

Average values have been calculated from the results given by different authors.

κ-caseins. The heterogeneity of caseins is due either to the presence of genetic variants or other factors such as a discrete phosphorylation level, variation in the extent of glycosylation of the κ-casein fraction, and the coexistence of protein forms having different chain lengths. It has been demonstrated that four ovine casein genes, αs1-CN, αs2-CN, β-CN, and κ-casein, are polymorphic and are linked on the both the ovine and the bovine genome. In recent years, five variants of αs1-CN have been described, designated A, B, C, D, and E, in line with the nomenclature proposed for cow or goat caseins. αs1-CN D was called the Welsh variant in view of its discovery in the milk of Welsh mountain sheep by King in 1966. This is the least phosphorylated variant. The lower number of phosphate groups explains the higher migration time of this variant by capillary electrophoresis at acid pH and the slow migration by alkaline polyacrylamide gel. Variant αs1-CN C differs from the A variant by the substitution of Ser for Pro at position 13, which explains the loss of the phosphate group on site 12 of the protein chain. Two variants of αs2-CN, A and B, have been described, differing in that Asn_{49} and Lys_{200} are replaced by Asp_{49} and Asn_{200}. In addition, a variant with a high electrophoretic mobility and a low molecular weight has been found in the milk of the Manchega breed. Ovine β-CN consists of 209 amino acid residues. There is nongenetic polymorphism due to varying degrees of phosphorylation, with six and five phosphate groups in β1 and β2, respectively. In addition, three genetic variants, designated A, B, and C, have been described recently. The only sequence difference found between A and C is the substitution of Glu at position 2 in variant A for Gln in variant C; no sequence data for the B variant are yet available. Both the stability of the casein micelle and the availability and distribution of Ca are affected by the extent of phosphorylation of the caseins. Sequencing of ovine κ-casein has shown that it consists of 171 amino acid residues. No genetic variants of κ-casein have been found, but it shows nongenetic polymorphism due to varying degrees of glycosylation at three Thr residue sites (positions 135, 137, and 138) and two phosphorylation sites (Ser P_{151} and Ser P_{168}). The casein fraction also contains γ-caseins, the products of the breakdown of β-caseins by plasmin.

Whey proteins Sheeps' milk whey proteins represent 17–22% of total protein. The principal proteins are β-lactoglobulin and α-lactalbumin. Immunoglobulins, serum albumin, and proteose peptones are present at lower concentrations. The latter are products of the breakdown of β-casein by plasmin. Another soluble protein found in small amounts and possessing antibacterial properties is lactoferrin. In the case of rennet whey, caseinomacropeptide is present also, produced by the chymosin action on the bond 105–106 of κ-casein. Serum albumin is not specific to milk, and it is considered to be the same as that found in the blood. The IgA in milk is synthesized in the mammary gland, i.e., it is milk-specific.

β-Lactoglobulin, the principal protein in the whey, consists of a polypeptide chain of 162 amino acids. Three genetic variants have been described in sheep's milk: β-Lg A, β-Lg B, and β-Lg C. Ovine β-Lg B and A differ only by a single amino acid exchange, His for Tyr at position 20. β-Lg C is a subtype of ovine β-Lg A with a single exchange, Arg for Gln at position 148. α-Lactalbumin is closely homologous to bovine α-lactalbumin. It is a metalloprotein, containing one atom of Ca per molecule and is important from a biological standpoint in that it is involved in lactose synthesis. Two genetic variants, A and B, have been described, although the B variant is very uncommon.

Nonprotein nitrogen Nonprotein nitrogen accounts for between 5 and 6.8% of total nitrogen. Nonprotein nitrogen compounds are: urea (45%), amino acids (16%), creatine (2.4%), creatinine (1.7%), ammonium (1%), uric acid (2.1%), and other unidentified compounds (32.3%). Sheeps' milk contains more urea and uric acid than cows' milk.

Lipid Fraction

Lipids are one of the most important components of milk in terms of cost, nutrition, and the physical and sensory characteristics they impart to dairy products. They are present in the form of globules and are characteristically abundant in sizes of less than 4.5 μm. No appreciable differences have been found in the mechanism of fat-globule secretion in sheep and cows; the structure and composition of the membrane are similar in both species. The phospholipid profile in both species is similar to that of the plasma membrane, which would tend to confirm their common origin. Along with triglycerides, the lipid fraction of sheeps' milk contains other simple lipids (diglycerides, monoglycerides, and cholesterol esters), complex lipids (phospholipds) and liposoluble compounds (sterols, alcohols, and hydrocarbons). Triglycerides constitute the largest group (nearly 98%), including a large number of esterified fatty acids, and so the composition is complex.

Fatty acids Fatty acids may be saturated or unsaturated (having from one to four double bonds). Most acids, from acetic to arachic acid, contain an even number of carbon atoms, of while approximately 2% are saturated with an odd number of atoms, and

roughly the same percentage of saturated fatty acids with branched methyl chains with an odd or even number of carbon atoms. Sheeps' milk contains more caproic, caprylic, and capric acid than cows' milk. These fatty acids are associated with the characteristic flavor of sheeps' milk cheeses and can also be used to detect mixtures of milk from different species (Table 2).

The most important of the factors that affect fatty acid composition is the feed. The addition of animal fats to the diet produces a decrease of C4–C14 acids and an increase of C16, C18, and C18:1 acids. If the intake of forage is reduced, the proportion of C4 to C16 fatty acids decreases accordingly.

Triglycerides The triglyceride structure of the milk fat is responsible for its rheological properties and its behavior during melting and crystallization. The triglyceride composition is useful to verify the authenticity of milk fat.

Triglycerides are almost invariably accompanied by small amounts of di- and monoglycerides, mainly at positions 1 and 2, which are therefore probably intermediates in the biosynthesis of triglycerides. The distribution of fatty acids in the triglyceride molecule, as determined by stereospecific analysis, differs slightly from cows' milk fat, but the distribution of acids in the molecule is not random. However, in both sheeps' and cows' milk, butyric acid and other short-chain fatty acids (C6–C8) are esterified mainly at position 3 of the glycerol molecule; the distribution of the other fatty acids (C10 or greater) exhibits no such marked specificity for positions 1 or 2.

Table 3 shows the average triglyceride composition of sheeps' milk fat compared with cows' milk fat. The triglycerides in sheeps' milk show a wide range of molecular weights when distributed according to the number of carbon atoms (taking into account the carbon atoms of the three acyl radicals), with two peaks at C38–C40 and C50–C52 and a minimum at C44–C46. Sheeps' milk fat has a higher percentage of short-chain triglycerides (C26–C36) than cows' milk fat (18% versus 11%). Percentages of medium-chain triglycerides (C38–C44) are also higher in sheeps' milk fat (33% versus 25%), and the percentages of unsaturated trglycerides are lower than in cows' milk fat (51% versus 55%). These differences relate to the need for a triglyceride composition with the appropriate melting point to allow the fat to be secreted. It has been established that the principal triglycerides in sheeps' milk fat are composed largely of three fatty acids (C14, C16, and C18:1), combined with the short-chain fatty acids C4 and C6.

Unsaponifiable lipids The unsaponifiable fraction of the milk fat is composed largely of sterols, with a smaller proportion of hydrocarbons, mainly squalene, and trace amounts of a number of normal- and branched-chain hydrocarbons in the range C17–C48 with odd and even numbers of carbon atoms, liposoluble vitamins and aliphatic alcohols. Ovine milk contains virtually no tocopherol or β-carotene.

Sterols are a minor fraction of total milk fat, the main component being cholesterol (270–350 mg per 100 g of fat, equivalent to approx. 20 mg per 100 ml of sheeps' milk), but small amounts of other sterols

Table 2 Fatty acid composition of ewes' milk fat

Fatty acid (long chain)	*Percentage*
Butanoic acid (C4)	3.00–5.80
Hexanoic acid (C6)	2.10–4.00
Octanoic acid (C8)	1.50–3.60
Decanoic acid (C10)	5.00–9.00
Decenoic acid (C10:1)	0.10–0.30
Dodecanoic acid (C12)	2.90–5.20
Tetradecanoic acid (C14)	7.00–13.40
9-Tetradecenoic acid (C14:1)	0.40–1.00
Pentadecanoic (C15)	0.60–1.50
9-Pentadecenoic (C15:1)	0.20–0.60
Hexadecanoic (C16)	20.00–28.50
9-Hexadecenoic (C16:1)	1.00–2.80
Heptadecanoic (C17)	0.20–1.00
9-Heptadecanoic (C17:1)	0.20–0.70
Octadecanoic (C18)	6.20–13.10
9-Octadecenoic[a] (C18:1)	16.60–27.70
9,12-Octadecadienoic[a] (C18:2)	2.80–4.30
9,12,15-Octadecatrienoic (C18:3)	0.60–2.00

[a]All isomers.
Ranges have been calculated from the results given by different authors.

Table 3 Triglyceride composition of ewes' and cows' milk fat[a]

Triglyceride	*Cow (wt%)*	*Ewe (wt%)*
C26	Trace	0.60–1.00
C28	0.30	1.30–2.50
C30	0.80	2.00–3.40
C32	1.90	3.10–5.00
C34	4.70	5.70–7.00
C36	9.80	9.20–10.30
C38	13.00	12.50–14.00
C40	10.80	11.00–12.50
C42	6.40	8.00–9.70
C44	5.80	7.30–8.60
C46	7.00	6.50–6.90
C48	9.10	6.10–7.80
C50	13.10	4.70–9.10
C52	12.10	5.00–9.40
C54	5.30	3.10–5.30
C56		< 0.50

[a]Triglycerides are identified by the number of acyl carbon atoms per glyceride molecule.
Ranges have been calculated from the results given by different authors.

implicated in cholesterol biosynthesis have also been found in ovine milk: lanosterol (5–15 mg per 100 g of fat) and, in even smaller proportions, dihydrolanosterol, desmosterol, and lathosterol.

Phospholipids The major phospholipids in both sheeps' and cows' milk are: phosphatidylethanolamine, phosphatidylcholine, phosphatidylserine, phosphatidylinositol, and sphingomyeline. These account for roughly 0.8% of the total lipid fraction.

Lactose

Lactose is the principal carbohydrate in milk. Chemically, it is a disaccharide consisting of a residue of D-glucose and D-galactose joined in a β-1, 4-glycosidic linkage. Sheeps' milk contains between 45 and 50 g of lactose per kilogram of milk, and lactose accounts for 22–27% of dry matter versus 33–40% in cows' milk. The lower lactose content does not present a problem in cheesemaking, since a sufficient amount of lactose is available to ensure lactic fermentation.

Enzymes

Enzymes are constituents of the mammary gland which pass into the milk during the secretion process. The enzymes found in milk are highly specific and are chiefly oxidoreductases, transferases, and hydrolases. Particularly important for technological purposes are proteases and lipase. Lipase, lysozyme, ribonuclease, and xanthineoxidase are less active in sheeps' milk, whereas alkaline phosphatase, despite having the same molecular weight and the same properties, is more active than in cows' milk. The most abundant enzyme is lactoperoxidase, which has proved impossible to distinguish from the bovine enzyme but appears to be more thermolabile. The indigenous proteases include aminopeptidases, thermolabile neutral acid proteinases, and serine proteinases, which are heat-resistant. The serine proteinase, plasmin, is heat-stable and hydrolyzes Lys—X and Arg—X bonds, becoming more active after thermal treatment.

Renneting Properties

The renneting properties of sheeps' milk are affected by a number of factors such as pH, physicochemical composition, micellar system, salts equilibrium, calcium concentration, and the temperature and time of heating. Sheeps' milk coagulates well and is indeed very suitable for making high-quality cheeses. It coagulates faster than cows' milk, and less rennet is needed to coagulate it in the same time as cows' milk. Curd formation is also faster than in cows' milk, but syneresis takes longer. The reason for this difference may be that sheeps' milk contains more casein, ionic calcium, and colloidal calcium than cows' milk. The renneting time and rate of firming are practically unaffected by the addition of calcium.

Physical Properties

The differences in the composition of sheeps' and cows' milk affect their physical properties. Sheeps' milk has a higher density, viscosity, refractive index, freezing point, and titrable acidity than cows' milk. Sheeps' milk is slightly more acid (lowest pH) than cows' milk, possibly due to its higher protein content.

Nutritive Value

The nutritional importance of sheeps' milk is due to its composition. Although sheeps' milk is richer in nutrients than cows' milk, it is rarely used as milk for drinking. In general, it is used for cheesemaking, but in certain countries, some is made into yogurt or other fermented milks. Sheeps' milk is an excellent source of high-quality protein, calcium, phosphorus, magnesium, zinc, and several vitamins. Only iron is in short supply, but its bioavailability is good because it is bound by lactoferrin, which is involved in the transport of iron in the body.

There is a good balance between the protein, fat, and carbohydrate components, each being present in similar amounts. The supply of nutrients is high in relation to the calorie content of the food. Sheeps' milk has a higher calorific value (500 kJ per 100 g) than cows' milk. Table 4 shows the range of concentrations of minerals and vitamins in ewes' milk.

Sheeps' milk contains mineral salts (around 0.9%). The most abundant elements are Ca, P, K, Na, and Mg. Ca and P are the most important of the macroelements in nutritional terms and for their role in the stability of casein micelles, and hence in the behavior of the caseins during milk processing. The most abundant of the trace elements are Zn, Fe, Cu, and Mn. With the exception of sodium, the concentrations of these elements are higher than those found in cows' milk. Sheeps' milk contains an average of 2 g of citrate per kilogram, which is slightly higher than that in cows' milk.

Concerning the distribution between the soluble and colloidal phases, the percentages of Ca and P in the soluble phase are, respectively, 20–25% and 35–40% of the total. These proportions are lower than in cows' milk, but the absolute concentrations in the soluble phase are comparable for the two species. Therefore, the concentrations of both elements in the colloidal phase are much higher than in cows'

Table 4 Vitamin and mineral content of sheeps' milk (per 100 g)

	Mean value	*Range*
Vitamins		
Vitamin A (μg)	50.00	
β-Carotene (μg)	5.00	2.00–7.00
Thiamin (B_1) (μg)	48.00	28.00–70.00
Riboflavin (B_2) (mg)	0.23	0.16–0.30
Nicotinamide (mg)	0.45	0.40–0.50
Pantothenic acid (mg)	0.35	
Biotin (μg)	9.00	
Vitamin B_{12} (μg)	0.51	0.30–0.71
Vitamin C (mg)	4.25	3.00–6.00
Folic acid (μg)	5.60	
Vitamin E (μg)	120.00	
Minerals		
Ca ($g\,l^{-1}$)	1.98	1.80–2.39
Mg ($g\,l^{-1}$)	0.18	0.10–0.22
Na ($g\,l^{-1}$)	0.50	0.27–0.81
K ($g\,l^{-1}$)	1.20	0.96–1.96
P ($g\,l^{-1}$)	1.30	1.17–1.70
Fe ($mg\,l^{-1}$)	0.76	0.34–1.40
Zn ($mg\,l^{-1}$)	7.50	0.47–1.20
Cu ($mg\,l^{-1}$)	0.07	0.04–0.14

[a]Values have been given by different authors.

milk, given the higher levels of casein found in sheeps' milk. Na and K are the major elements in the soluble phase. Of the trace elements Zn, Mn, Fe, and Cu, the first two are present mainly (around 90%) with the colloidal fraction of the milk, while Fe and Cu are present to a lesser extent (70 and 65%, respectively).

The vitamin content is generally higher in sheeps' milk than in cows' milk. The level of retinol activity of ewes' milk is approximately twice that of goats' and cows' milk.

Cheeses

Practically all sheeps' milk is used to make cheeses, most of which are craft products. Some have an Appellation of Origin, so that production is confined to specific regions, and only a minority are made on an industrial basis. There are six categories of sheeps' milk cheese based on the technology used: fresh, white brined (pickled), blue-veined, semihard, hard, and whey cheeses.

Roquefort cheese, is a blue-veined cheese made in France from raw sheeps' milk inoculated with *Penicillium roqueforti* spores. It undergoes extensive proteolysis, 50% water-soluble nitrogen in total nitrogen (WSN/TN) and lipolysis, 8–10% of total fatty acids after 5 months of ripening. Spain produces a blue-veined variety, Cabrales cheese, with similar characteristics. This cheese is highly appreciated, although the raw material is a mixture of cows', sheeps', and goats' milk.

Feta cheese is a white-brined cheese originally from Greece, and is one of the most popular cheeses internationally. Traditionally, it is made from raw sheeps' milk and is ripened in barrels with brine (6–8% NaCl) for around 1 month at 8–10 °C; it is then kept chilled for at least 2 months prior to consumption. Nowadays, it is made mainly from pasteurized sheeps' milk. There is a variety similar to feta called teleme, originally from Romania, which is generally made from a mixture of milk of all three species. The white-brined category also includes halloumi, a cheese produced chiefly in Cyprus from raw sheeps' milk.

Pecorino is an Italian semihard or hard cheese. There are several varieties with Appellations of Origin, including Romano, Siciliano, and Fiore Sardo. The Romano variety accounts for 50% of total production. This cheese is made with raw or pasteurized sheeps' milk, in most cases coagulated with lamb or kid rennet paste. Proteolysis is moderate (≈20% WSN/TN), but there is intense lipolysis due to the presence of pregastric esterases in the traditional rennet pastes used. Kefaloteri is a variety similar to pecorino produced in Greece, although rennet pastes is not normally used for coagulation.

Kachkaval has been produced in various European countries, particularly in the Balkans, at least since the eleventh century. This is a stretched-curd cheese which undergoes moderate proteolysis even after 3 months of ripening (≈13% WSN/TN).

The most characteristic Spanish semihard/hard sheeps' milk cheeses are Manchego, Zamorano, Roncal, and Idiazabal, traditionally made from raw milk. These are generally ripened for 3–6 months, although Manchego may be matured in olive oil for a longer period. Proteolysis is medium/high, 20–45%WSN/TN, while lipolysis is low. Idiazabal cheese is ripened for 1–2 months and then smoked.

Some sheeps' milk cheeses are made in Portugal and Spain using vegetable rennet extracted from the flower of the thistle, *Cynara cardunculus*. Examples include Serra da Estrela in Portugal and La Serena, Los Pedroches, and Torta del Casar in Spain. These are commercialized after relatively short ripening (30–60 days), but the proteolysis levels are high (40–50% WSN/TN) due to the strong proteolytic activity in the curd. Spain produces two highly popular fresh cheeses, Burgos and Villalón. These are now made from pasteurized milk using animal rennet, but without a starter culture. They must be kept in chilled storage and consumed in less than 6 days.

Finally, in various countries, the whey from sheeps' milk in cheesemaking is used on its own or mixed with milk to manufacture products of high nutritional value. The whey contains mainly water-soluble

Table 5 Average content of fat, protein, calcium, and phosphorus in cheeses made from ewes' milk

Country	*Name*	*Type of cheese*	*Fat content (percentage total solids)*	*Protein (%)*	*Ca (%)*	*P (%)*
France	Roquefort	Blue-veined	50.00	21.00	0.62	0.42
Greece	Feta	Brined	40.00	18.00	0.65	0.40
Italy	Pecorino	Hard	42.00	29.00	0.40	0.32
Portugal	Serra da Estrela	Semihard	56.00	20.00	0.65	0.53
Romania	Teleme	Brined	50.00	18.00	0.53	0.40
Spain	Manchego	Semihard	50.55	23.00	0.68	0.54

Average values have been calculated from the results given by different authors.

proteins ($\approx 1\%$), fat ($< 1\%$), lactose, minerals, nonnitrogen substances, and vitamins. These whey cheeses are the result of heat-induced coagulation of the whey protein (which occludes fat). The best-known whey cheeses are ricotta from Italy, Manouri and Myzithra from Greece, and Requesón from Spain.

Besides the cheeses mentioned here, many countries produce cheeses, generally semihard varieties, in which the raw material is a mixture of sheeps', cows', and goats' milk.

Nutritional Aspects of Cheese

Cheese, in its concentrated form, contains much of the same nutrients as milk, caseins, fat, minerales, vitamins, colloidal salts, and some whey constituents. Some (up to 90%) of the hydrosoluble vitamins may be lost in the whey, but the levels of vitamins in cheeses depend on the balance on the consumption and production of the vitamins by microorganisms. The important nutrients in cheese come mainly from the high content of biologically valuable proteins which supply essential amino acids, and the levels of minerals, mainly calcium and phosphorus. Table 5 shows the average content of fat, protein, calcium, and phosphorus of some of the principal sheeps' milk cheeses produced in different countries, indicating the type of consistency.

The transformation of lactose to lactic acid occurs during the first stage of processing, this means that people who are lactose-intolerant can consume cheese with impunity.

The protein content of different varieties of ewes' milk cheese varies between 20 and 30%. In cheese manufacture, the casein is incorporated into the curd. The more biologically valuable whey proteins are lost in the whey, but the biological value of the proteins in cheese is higher than that of casein alone. The biological value of proteins is improved when the whey proteins are incorporated into the cheese by membrane processes such as ultrafiltration. During ripening, caseins are hydrolyzed to peptides and amino acids by the action of rennet and microbial proteases. The nutritional value of the proteins in cheese is not affected by the above-mentioned processes, and the bioavailability of lysine is comparable to that of casein in milk.

The fat is hydrolyzed during ripening by microbial lypases, which produce free fatty acids. These components contribute to the development of the aroma of the cheeses. The level of fat in sheeps' milk cheeses varies from 12 to 30%, depending on the length of time and degree of ripening. The digestibility of the fat of different varieties is between 88 and 94%. The content of Ca and P in cheese is higher than that in milk, four to five times higher in fresh cheeses, seven to eight times higher in semihard cheese, and 10 times higher in hard cheeses. Only in acid-coagulated cheeses is the content of Ca and P lower than in milk. The bioavailability of calcium cheese is comparable to that from milk and is not affected by the ripening process.

See also: **Casein and Caseinates**: Methods of Manufacture; Uses in the Food Industry; **Cheeses**: Types of Cheese; Starter Cultures Employed in Cheese-making; **Enzymes**: Functions and Characteristics; **Fats**: Requirements; **Fatty Acids**: Properties; **Lactose**; **Phospholipids**: Properties and Occurrence; **Triglycerides**: Structures and Properties; **Whey and Whey Powders**: Protein Concentrates and Fractions

Further Reading

Alichanidis E and Polychroniadou A (1996) Special features of dairy products from ewe and goat milk from the physicochemical and organoleptic point of view. In: *Production and Utilization of Sheep and Goat Milk*, pp. 21–43. Brussels: International Dairy Federation.

Amigo L, Recio I and Ramos M (2000) Genetic polymorphism of ovine milk proteins: its influence on technological properties of milk. A review. *International Dairy Journal* 10: 135–149.

Anifantakis EM (1986) Comparison of the physicochemical properties of sheeps' and cows' milk. In: *Bulletin 202*, pp. 42–53. Brussels: International Dairy Federation.

Assenat L (1985) Le lait de brebis. In: Luquet FM (ed.) *Laits et Produits Laitiers. Vache, Brebis, Chévre*, pp. 279–392. Paris: Technique et documentation Lavoiser.

De la Fuente MA, Olano A and Juarez M (1997) Distribution of calcium, magnesium, phosphorus, zinc, manganese, copper and iron between the soluble and colloidal phases of sheeps' and goats' milk. *Lait* 77: 515–520.

Juarez M and Ramos M (1984) Dairy products from sheeps' and goats' milk. *Dairy Industries International* 49: 20–24.

Kalantzopoulos GC (1993) Cheeses from sheeps' and goats' milk. In: Fox PF (ed.) *Cheese: Chemistry, Physics and Microbiology Vol. 2 Major Cheeses Group*, 2nd edn, pp. 507–553. London: Chapman & Hall.

Núñez M, Medina M and Gaya P (1989) Sheeps' milk cheese: technology, microbiology and chemistry. *Journal of Dairy Research* 56: 303–321.

Rainal K and Remeuf F (1998) The effect of heating on physicochemical and renneting properties of milk: a comparison between caprine, ovine and bovine milk. *International Dairy Journal* 10: 135–149.

Ramos M and Juárez M (1981) The composition of sheep's and goats' milk. In: *Bulletin* 140, pp. 1–19. Brussels: International Dairy Federation.

Ramos M and Juárez M (1986) Chromatographic, electrophoretic and immunological methods for detecting mixtures of milk from different species. In: *Bulletin 202*, pp. 175–190. Brussels: International Dairy Federation.

Ramos M and Juarez M (in press) Sheep's milk. In: *Encyclopedia of Dairy Science*. London: Academic Press.

Renner E (1987) Nutritional aspects of cheese. In: Fox PF (ed.) *Cheese: Chemistry, Physics and Microbiology Vol. 1 General Aspects*, 1st edn, pp. 345–363. London: Elsevier Applied Science.

Scott KJ (1989) Micronutrients in milk products. In: Renner E (ed.) *Micronutrients in Milk and Milk-based Food Products*, pp. 71–123. London: Elsevier Applied Science.

Shelf-life *See* **Chilled Storage**: Principles; Attainment of Chilled Conditions; Quality and Economic Considerations; Microbiological Considerations; Use of Modified-atmosphere Packaging; Packaging Under Vacuum; **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage; Fungi in Food – An Overview; Molds in Spoilage; Yeasts in Spoilage; **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability; Shelf-life Testing; Shelf-life Testing

SHELLFISH

Contents

Characteristics of Crustacea

D A Jones, University of Wales Bangor, Gwynedd, Wales, UK

Introduction

As shellfish or shelled aquatic invertebrates, the Crustacea share important features, such as bilateral symmetry, body segmentation, and the possession of a well developed exoskeleton, with the other groups within the phylum Arthropoda. However, they may be separated at the subphylum level from the Cheliceriformes (scorpions, spiders, etc.), Trilobitomorpha (fossil trilobites) and Uniramia (insects and myriapods) by the features listed below:

1. Appendages are uniramous or biramous.
2. Brain is tripartite (with deutocerebrum).
3. Body is divided into cephalon and trunk, or latter subdivided into thorax and abdomen.

4. There are five pairs of cephalic appendages: preoral first antennae and four pairs of postoral appendages – second antennae (which migrate to preoral position in adults), mandibles, maxillules, and maxillae.
5. Mandible is gnathobasic (mandible arises from limb base); endopod and exopod are reduced in adults.

There are at least 39 000 species of crustaceans, ranging in size from less than 1 mm in length to spider crabs with leg spans of over 4 m, and these are subdivided into six classes and 38 orders. Although most of these are marine, small groups have exploited the fresh-water and terrestrial environments, so that crustaceans can be found from mountain tops and deserts to abyssal depths. Despite the large number of species, many are either too small or too dispersed to be exploited by humans. The forms most commonly exploited commercially are crabs, lobsters, and shrimps. Planktonic copepods (10 000 species), Cladocera (450 species), planktonic and benthic Ostracoda (5000 species), and mainly benthic Amphipoda and Isopoda (10 000 species) often form an essential link in food webs which support fisheries, but such forms are only rarely themselves the subject of fisheries. However, a key problem in the management of crustacean fisheries concerns euphausiids (krill) which play a central role in the Antarctic ocean food web as food for many whales, seals, and birds, yet are also increasingly being fished directly as food for humans and domestic animals.

Taxonomy

A brief outline classification of crustaceans utilized directly by humans is given in Table 1; for more detailed descriptions, refer to the Further Reading section. It is clear that, apart from larger planktonic species such as euphausiids, mysids, and sergestids, which swarm in sufficient concentration to be harvested economically, and gourmet species such as stalked barnacles (*Pollicipes*) and stomatopods, the majority of edible crustaceans are found in order Decapoda, which comprises some 10 000 species. The possession of branched or dendrobranch gills, external fertilization, and release of eggs into the sea separates the penaeid and sergestid shrimp from the rest of the Decapoda.

The Pleocyemeta have unbranched gill filaments and brood their eggs, which hatch at a later stage than the nauplius produced on hatching by the Dendrobranchiata. This group contains the majority of the familiar shrimps or prawns (the terms are now synonymous), crayfish, lobsters, squat lobsters, and crabs. Caridean shrimps (*Palaemon, Macrobrachium*) possess phyllobranchiate or flattened, plate-like gills, separating them from the larger and more robust Astacidea (which possess trichobranchiate or unbranched, tubular gill filaments). All the fresh-water crayfish (Astacidea) and marine lobsters (*Homarus* and *Nephrops*) have greatly enlarged chelae or claws on the first pair of walking legs, distinguishing them from the Palinura (spiny and slipper lobsters), which have similar gill structures, but lack the enlarged claws. This marine group tends to be more mobile and possesses a flattened abdomen and large tail fan used in swimming.

The Anomura contain forms in which the abdomen is either soft and twisted asymmetrically to fit into gastropod shells, or flexed beneath the cephalothorax as in galatheid lobsters. The coconut crabs (*Birgus*) and lithodids represent extreme groups in which the asymmetrical abdomen is closely applied to the underside of the thorax, giving the appearance of true crabs or Brachyura. This final group is distinguished by the lateral expansion of the cephalothorax, and reduction of the abdomen to form a symmetrical flap, lacking uropods, which is flexed under the thorax.

Crustacean Structure and Function

As a predominantly aquatic group, crustaceans have been able to exploit the arthropod chitinized exoskeleton fully without the weight restrictions imposed by such a system on land. In the classes under consideration (Table 1), fusion of the head and one to three thoracic segments has occurred to produce a carapace, or protective shield, which often extends forwards in the form of a rostrum and posteriorly to protect the cephalothorax (Figure 1). The abdomen usually retains flexibility by means of soft cuticular joints between each segment. Even barnacles (Figure 1) retain most of these features, although much modified for their sedentary existence.

Each segment typically bears a pair of jointed appendages, anteriorly. The two pairs of antennae are often elongated and mobile, and carry esthetascs which are chemosensory hairs. Other head appendages usually form mouthparts concerned with feeding and include the mandibles or jaws, maxillules, and maxillae, and, in malacostracans, additional maxillipeds.

The rest of the thoracic appendages are modified into pereopods (Figure 1) specialized for walking, swimming, respiration, feeding, or defense. The abdominal segments (pleonites) typically bear biramous pleopods, paddle-like appendages, used for

Table 1 Pennant (1771) classification of phylum, subphylum or superclass Crustacea utilized directly for human nutrition

Class Maxillopoda	
Subclass Cirripedia	
Order Thoracica	Stalked barnacles, e.g., *Pollicipes*
Subclass Copepoda	
Order Calanoida	Copepods, e.g., *Calanus plumchrus*
Class Malacostraca	
Subclass Hoplocarida	
Order Stomatopoda	Mantis shrimp (*Squilla mantis*)
Subclass Eumalacostraca	
Superorder Peracarida	
Order Mysidacea	Possum shrimp (*Neomysis intermedia*)
Superorder Eucarida	
Order Euphausiacea	Krill, e.g., *Euphausia superba*
Order Decapoda	
Suborder Dendrobranchiata	Penaeid and sergestid shrimps, e.g., *Penaeus, Sergestes*
Suborder Pleocyemata	
Infraorder Caridea	Caridean and procaridean shrimps, e.g., *Macrobrachium, Palaemon*
Infraorder Astacidea	Crayfish and chelate lobsters, e.g., *Astacus, Homarus, Nephrops*
Infraorder Palinura	Palinurid, spiny, and slipper lobsters, e.g., *Panulirus, Palinurus, Thenus, Scyllarides*
Infraorder Anomura	Galatheid crabs, king crabs, e.g., *Paralithodes, Pleuroncodes*
Infraorder Brachyura	Crabs, e.g., *Cancer, Scylla, Callinectes, Maia*

swimming in malacostracans. In this group the terminal somite bearing the anus forms a flattened telson, and the last pair of abdominal segments are modified to form uropods; together with the telson; these form a tail fan used in swimming.

The gut is divided into a chitin-lined fore- and hindgut and a midgut lined with endoderm. The foregut esophagus leads to a stomach which is often subdivided into cardiac and pyloric regions in malacostracans. The midgut forms an intestine of variable length and bears the digestive cecae or hepatopancreas, emptying into the pyloric chamber of the stomach. The hindgut is usually short, absorptive in function, and leads to the anus. In crabs and lobsters, mechanical breakdown of food in the gastric chambers of the stomach is performed by heavily sclerotized teeth which form a grinding gastric mill; in shrimps this structure may be absent and breakdown is solely by enzymes secreted from the hepatopancreas. The products of digestion are absorbed by cells in the fine tubules of the hepatopancreas, or cells lining the midgut trunk, where further intracellular digestion occurs. In some groups, expelled fecal material is sheathed in a peritrophic membrane.

The circulatory system is comprised of a dorsal muscular heart with ostia or pores to draw in blood from the pericardial cavity. In advanced malacostracans the heart has a series of blood vessels insuring that blood flows to body organs and to the gills (Figure 1). Return to the heart is via pools or the hemocoel, although active crustaceans may have a primitive venous system to return the blood to the pericardial cavity. Gaseous exchange in large, advanced crustaceans is via gills, which arise as branches from the base of the thoracic limbs. The gills are modified in various ways to provide a large surface area of thin, permeable cuticle, and in decapods are protected under the carapace in branchial chambers through which ventilating currents of water are drawn.

Excretion is in the form of ammonia, which is released across gill surfaces and via nephridia in maxillary or, in most malacostracans, antennal glands (Figure 1). These glands are also active in osmoregulation, as are the gill surfaces. The crustacean cuticle, unlike the waxy insect cuticle, is largely permeable, and imposes severe constraints upon ionic regulation; hence few crustaceans are found away from water.

The crustacean brain is composed of three fused ganglia – two anterior dorsal supraesophageal ganglia and a third which forms a pair of circumenteric connectives extending round the esophagus to a subesophageal ganglion linked to the ventral nerve cord. This cord bears paired segmental body ganglia. The optic and antennulary nerves run to the supraesophageal ganglia, and the subesophageal ganglia often become a large, fused mass serving the nerves to the mandibles, maxillules, maxillae, and maxillipeds. In crabs all the thoracic ganglia fuse to form a large ventral nerve plate. Sensory systems are well developed, despite the exoskeleton, and take the form of innervated setae responding to touch or currents, whilst others, such as esthetascs, detect chemicals or gradients in attractants emanating from food.

Crustaceans are well adapted to detect light, and photoreceptors range from the simple larval naupliar eye, responding to light direction and intensity, to the

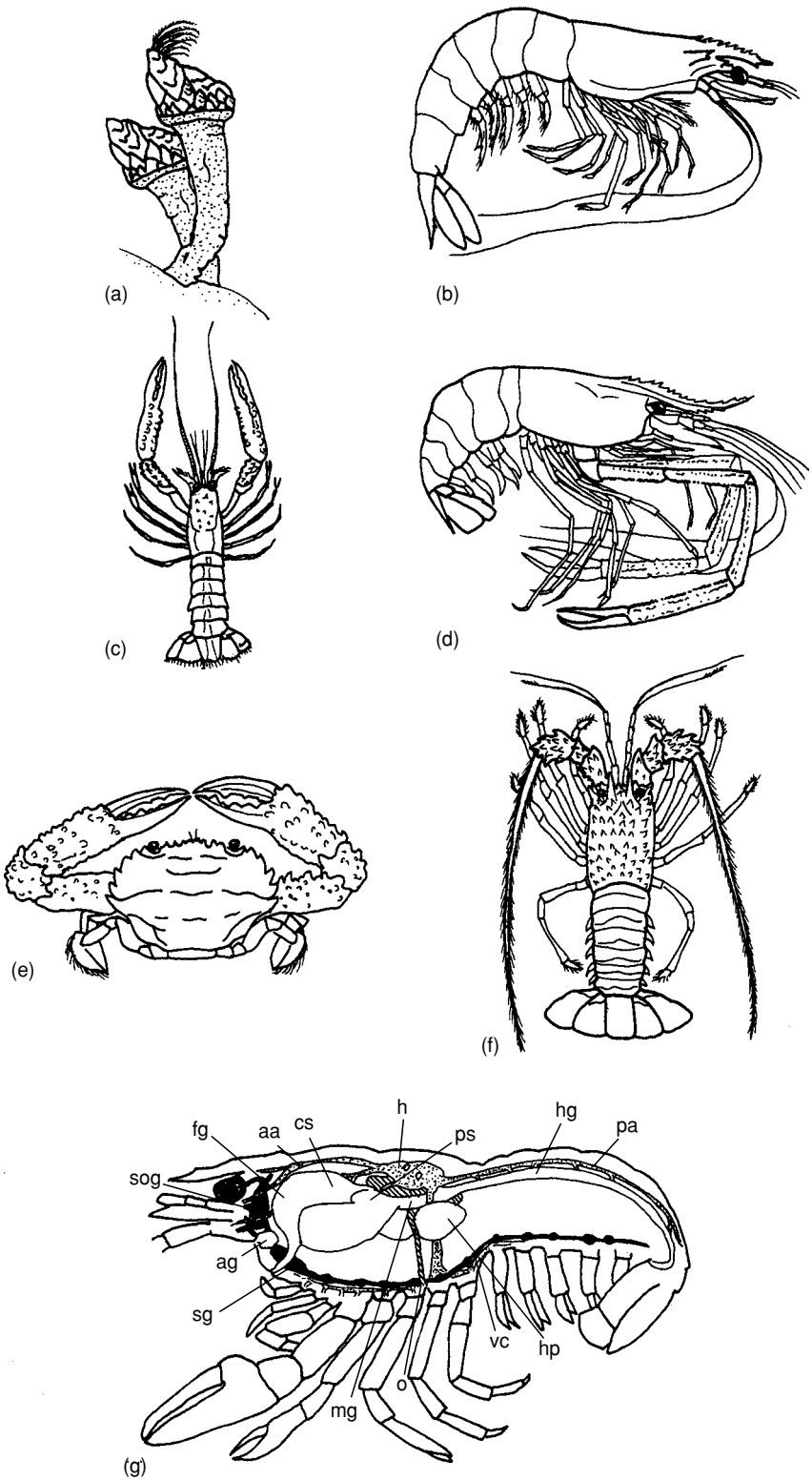

Figure 1 (a) Stalked barnacle (*Pollicipes* sp.); (b) *Penaeus* shrimp; (c) (Macrobracium Norway lobster (*Nephrops* sp.); (d) caridean shrimp (*Macrobracium* sp.); (e) swimming crab (*Charybdis* sp.); (f) spiny lobster (*Panuliris* sp.); (g) fresh-water crayfish (*Astacus* sp.). sog, supraesophageal ganglia; fg, foregut; aa, anterior aorta; cs, cardiac stomach; h, heart; ps, pyloric stomach; hg, hindgut; pa, posterior aorta; hp, hepatopancreas; vc, ventral nerve cord; o, oviduct; mg, midgut; sg, subesophageal ganglion; ag, antennal gland. Reproduced from Shellfish, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 2 Composition of commercial crustaceans (all figures are per 100 g of raw material, except *Homarus*, which is boiled)

	Energy (kcal)	*Carbohydrate (g)*	*Protein (g)*	*Total fat (g)*	*Fatty acids*				*Cholesterol (mg)*
					Saturated (g)	*Mono-unsaturated (g)*	*Poly-unsaturated (g)*	*n-3 (g)*	
Crab (mixed)	74–95	0–2.2	15–18	0.8–1.9	0.16–0.17	0.22	0.4–0.5	0.38–0.44	60–78
Penaeid shrimp	87–100	0–2.7	17–22	0.4–0.8	0.11–0.2	0.05–0.15	0.09–0.49	0.07–0.34	96
Panulirid lobster	100	1.7	19.2	1.2	0.14	0.14	0.59	0.27	106
Homarid lobster	93	5.4	20.5	0.6	0.08	0.13	0.07	0.06	72

Data modified from various sources.

Table 3 Habitats and global distribution of crustaceans utilized for human nutrition

Group	*Habitat*	*Distribution (commercial fisheries)*
	Marine	
Copepods	Pelagic	Norway, Canada, Japan
Cirripedes	Rocky, coastal	Portugal
Mysids	Pelagic, coastal, estuarine	Japan, South-east Asia, China, Korea
Euphausiids	Pelagic, offshore	Antarctica, Canada, Norway, Mediterranean
Sergestids	Pelagic, coastal, estuarine	China, South-east Asia, Japan, East Africa, India, Brazil, Surinam, Philippines
Penaeid shrimp (60 species)	Benthic, soft-substrate, nutrient-rich, estuarine, coastal	Worldwide between 40 °N and 40 °S
Plesiopenaeus	Benthic, soft-substrate, deep-water	Atlantic, Australia, South Africa
Pleoticus	Soft-substrate, deep-water	South-west Atlantic
Caridean shrimp		
Crangon	Benthic, soft-substrate, coastal	Europe, former Soviet Union, Algeria
Pandalus	Benthic, coastal to 1500 m	North Pacific, Atlantic
Palaemon	Benthic, rocky, coastal	Europe, Algeria
Stomatopoda	Benthic, rocky, coastal	Tropical to Mediterranean
Lobsters		
Homarus	Benthic, rocky-soft, coastal to 700 m	North Atlantic, Mediterranean
Nephrops	Benthic, soft substrate, 15–800 m	North-west Atlantic, Mediterranean
Panulirids	Benthic, rocky, coastal, 700 m	30 °N–50 °S worldwide
Syllaridae	Benthic, soft-rocky, coastal	Mediterranean, Japan, Indian Ocean
Anomurans		
Galatheids	Benthic, rocky, coastal	Mediterranean, Japan, western USA
Lithodes	Benthic, rocky-soft, coastal	South-west Atlantic
Paralithodes	Benthic, rocky-soft, coastal	North-west Pacific
Crabs		
Chionecetes	Benthic, rocky, coastal	North-west Atlantic, north-west Pacific, east central Atlantic
Maia	Benthic, rocky, seaweeds, coastal	Mediterranean
Cancer	Benthic, rocky-soft, coastal	Africa, north-east and central Atlantic, Mediterranean, north-east and -west Pacific, east central Pacific
Portunids	Benthic, rocky-soft, coastal	North-east Atlantic, Asia, west Pacific
Callinectes	Benthic, rocky-soft, coastal, shelf edge	West and north-west Atlantic
Scylla serratus	Benthic, mangal, coastal	Asia, India, west central Pacific
Geryon	Benthic, soft, 300–1500 m	North-west Atlantic
	Fresh water	
Caridean shrimp		
Macrobrachium	Estuaries, rivers, benthic, soft-substrate	Tropical, introduced worldwide
Palaemonids	Estuaries, rivers, lakes, soft, benthic	Tropical
Astacidea		
Crayfish	Rivers, lakes, streams, rocky-soft, vegetated	Temperate to tropical, worldwide

stalked, multifaceted compound eye found in decapods. This is capable of discerning shapes, patterns, and movement and at least some have color vision. Molting, chromatophore activity, tidal and daily locomotor rhythms, and aspects of reproduction are under hormonal and neurosecretory control, but an understanding of the mechanisms of this control is still at an early stage.

Sexes are separate in most crustaceans, but even some advanced malacostracans such as *Pandalus* may be protandrous hermaphrodites (maturing first as males, then later changing sex). Gonads are paired structures which are found in various regions of the trunk and empty via genital pores, usually on a trunk sternite. In male decapods an anterior pair of pleopods is modified for sperm transfer. Sperm is deposited directly into the oviduct or into a seminal receptacle, where it may be stored for some time. Crustaceans may brood fertilized eggs, usually in an external pouch (mysids) or attached to pleopods (most decapods), but exceptionally (penaeids) may release their eggs freely into the sea.

Typically crustacean eggs hatch into planktonic larval forms, although these are suppressed, for example, in the Mysidacea, Amphipoda, Isopoda and fresh-water Astacidea, where direct development occurs. The most primitive larva is the nauplius, with a single median simple eye and three pairs of biramous limbs; this is followed by the zoea and megalopa, gaining additional limbs and segments at each molt. A wide variation in numbers of larval stages and duration is seen; penaeids have 12 stages extending over 13 days, homarid lobsters four stages over 15 days, whilst the planktonic life of palinurids may extend up to 12 months.

Table 2 gives the composition of crustaceans commonly consumed by humans and reveals that all are low in saturated fats, high in polyunsaturates, particularly the highly unsaturated *n*-3 series, and contain medium levels of cholesterol. (*See* **Carbohydrates**: Requirements and Dietary Importance; **Energy**: Measurement of Food Energy; **Fats**: Requirements; **Protein**: Requirements.)

Several crustacean groups (copepods, branchiurans, cirripedes, and isopods) have given rise to parasites which may cause serious infestations in edible fish and shellfish.

Habitats and Distribution

Apart from the pelagic zooplanktonic copepods, mysids, euphausiids, and sergestids, which are mainly fished commercially in colder productive waters (Table 3), most other crustaceans only occur in commercial densities in shallow coastal seas. The fast-growing, high-value penaeid shrimp are restricted to warmer waters, as many have life cycles associated with estuarine mangals, and many burrow in sands and muds. In contrast, the slower-growing caridean shrimps are centered in the northern boreal region, where the shallow-sea pandalids make up the bulk of the fisheries.

The slow-growing homarid lobsters also have a cold-to-warm temperate distribution and are replaced in warmer seas by the panulirid and scyllarid lobsters. The most productive crab fisheries used to be those for king crabs (*Paralithodes*) and snow crabs (*Chionecetes*) from the North Pacific and Alaska, but blue crabs (*Callinectes*) from the western Atlantic have recently produced the highest landings. Although most lobsters and crabs show a preference for rocky habitats, both *Cancer* and *Homarus* will construct burrows in soft substrates.

Many crustacean groups have invaded fresh-water habitats, but only caridean shrimp, notably *Macrobrachium* and crayfish, attain commercial fishery sizes. The former are restricted to tropical waters, whilst the latter occur worldwide. European crayfish populations declined as a result of disease, but recent introduction of expatriate species has led to these becoming pests, particularly in Africa and southern Europe.

See also: **Carbohydrates**: Requirements and Dietary Importance; **Energy**: Measurement of Food Energy; **Fats**: Requirements; **Protein**: Requirements

Further Reading

Bowman TE and Abele LG (1982) Classification of the recent crustacea. In: Bliss DE (ed. in chief) *The Biology of Crustacea*, vol. 1, pp. 1–25. New York: Academic Press.

Brusca RC and Brusca GJ (1990) *Invertebrates*, pp. 1–922. Sunderland, Massachusetts: Sinaver Associates.

Provenzano AJ (1985) Economic aspects: fisheries and culture. In: Bliss DE (ed. in chief) *The Biology of Crustacea*, vol. 10, pp. 1–331. New York: Academic Press.

Commercially Important Crustacea

B D Paterson, Bribie Island Aquaculture Research Centre, Bribie Island, Australia

Background

Visit a fish market, supermarket, or restaurant anywhere in the world, and a variety of crustaceans are likely to be presented for sale. The particular species found will depend greatly upon the country and the region concerned. However, as crustaceans and other kinds of seafood move further afield with the

increasing international trade in food products, this means both an increasing variety in the product displayed in the major markets of Asia, USA, and Europe, and more intense competition within broad product categories. As a consequence, the names of imported product can be a hot 'branding' issue amongst local producers. But rather than perpetuating arbitrary distinctions, for example about what constitutes a 'true' lobster, in order to objectively come to grips with the varieties of crustaceans produced in significant quantities throughout the world, it is easier to consider the three broad product groups; shrimp, lobsters, and crabs. These commodity divisions nevertheless mirror broad taxonomic differences, largely because the morphology of crustaceans has a large impact on their processing and marketability.

Most of the commercially important crustaceans are decapods, the group containing the largest and most familiar crustaceans (Table 1). The sheer diversity and global distribution of fished and cultured crustacea seemingly count against any attempt at a detailed statistical compilation. But it is possible, within the limits of the data available, to find major patterns in the production figures and identify outstanding species. Of course, species that do not rate highly in terms of global production can still be commercially important. They may be high in value and on display, such as crustaceans marketed alive, or they may be less valuable but still contribute locally to nutrition, employment, and prosperity. Commercial exploitation itself is enough to warrant local regulation and management of stocks or farming operations, though, now, the viability of 'local' industries and the quality of their products have reached new horizons of concern, now that quality and sustainability are the concern of world consumers and trade regulators.

Table 1 Taxonomic relationships amongst commercially important crustacea

- Crustacea (Class)
  - Decapoda (Order)
    - Natantia (Suborder)
      - Caridea (Infraorder)
        - Crangonidae (Family) *Crangon crangon*
        - Palaemonidae (Family) *Macrobrachium rosenbergii*
        - Pandalidae (Family), e.g., *Pandalus borealis*
      - Penaeoidea (Infraorder)
        - Penaeidae (Family), e.g., *Penaeus* spp.
        - Sergestidae (Family), paste shrimp, *Acetes japonicus*
    - Reptantia (Suborder)
      - Anomura (Infraorder)
        - Galatheidae (Family), Squat lobsters, *Pleuroncodes monodon*
        - Lithodidae (Family), Stone crabs, *Paralithodes camtschaticus*
      - Astacura (Infraorder)
        - Cambaridae (Family), Freshwater crayfish, e.g., *Procambarus clarkii*
        - Nephropidae (Family), Lobsters, e.g., *Homarus americanus*, *Nephrops norvegicus*
        - Parastacidae (Family) Freshwater crayfish, e.g., *Cherax quadricarinatus*
      - Brachyura (Infraorder)
        - Cancridae (Family), Edible crabs, e.g., *Cancer* spp.
        - Majidae (Family), Spider crabs, *Chionoecetes* spp.
        - Portunidae (Family), Swimming crabs *Portunus trituberculatus*, *Callinectes sapidus*
      - Palinura (Infraorder)
        - Palinuridae (Family), Spiny lobsters, e.g., *Panulirus argus*
  - Euphausiacea (Order)
    - Euphausiidae (Family), Krill, *Euphausia superba*

The total global production of crustaceans from fisheries and aquaculture was 7.8 million tonnes in 1999, and crustaceans represented about 20.1% of global aquaculture production in that year. Shrimps and shrimp-like crustaceans account for about half the global crustacean trade, largely due to the large output of shrimp fisheries and an increasing component of farmed tropical marine shrimp. About 81% (6.3 million tonnes) of harvested crustaceans came from capture fisheries, so the taxonomic/market breakdown of this group (Figure 1) still has a significant effect on global totals. Marine shrimp currently dominate the crustacean farming sector (Figure 2), with much of the remaining farm production based in freshwater.

Beyond the level of the total contribution by various taxa, it is not always possible to categorically identify the most important species by isolating these data from the production figures. Sometimes, the compiled global statistics for particular groups include large percentages of production that cannot be assigned to anything but the broadest taxonomic categories. Nevertheless, the species that emerge from the data contain no surprises, even where relatively large proportions of production remain unspecified.

While the better known crustaceans originate from the world's oceans, it would be wrong to discount altogether the 10% of global production of crustaceans that originates from freshwater sources. While these of course are also shrimp, crabs, or indeed 'lobsters', they deserve separate treatment, because they do not necessarily compete in the same market segments as their marine relatives.

Marine Shrimp

The suborder Natantia (shrimps and prawns) are decapod crustaceans adapted for swimming, with a light cuticle and a well-developed abdomen with a

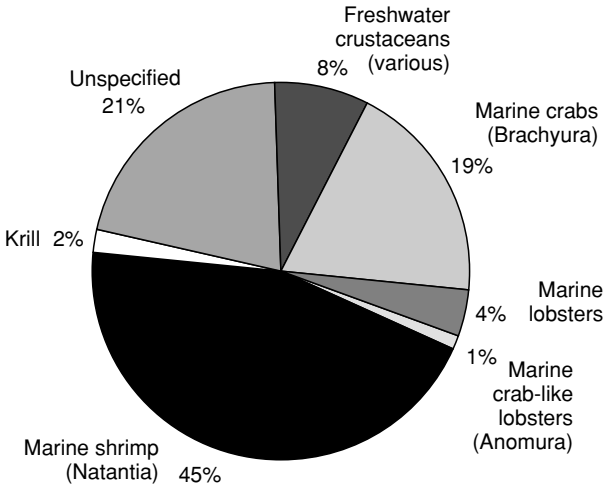

Figure 1 Global fisheries production of major product/taxonomic groups of crustaceans in 1999. From Anon (1999) *FAO Fishery Statistics: Capture Production; Aquaculture Production; Commodities*, vol. 88(1–3). Rome: FAO.

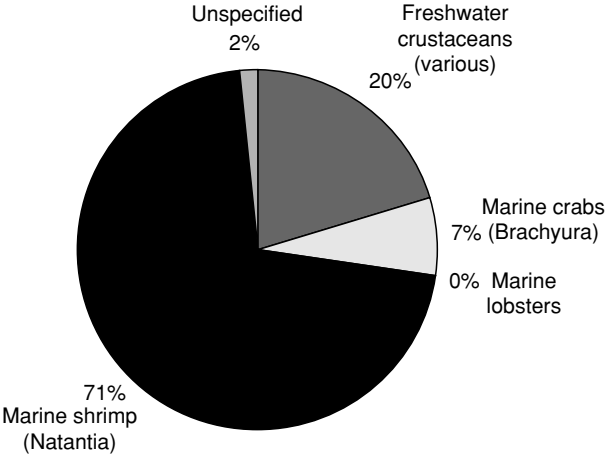

Figure 2 Global farmed production of major product/taxonomic groups of crustaceans in 1999. From Anon (1999) *FAO Fishery Statistics: Capture Production; Aquaculture Production; Commodities*, vol. 88(1–3). Rome: FAO.

full complement of swimmerets. The abdomen is also used for the powerful tail-flicks that the shrimp uses to flee danger, so it is ironic that it is this very adaptation that makes shrimps so commercially attractive.

Production of shrimp in 1999 was about 4 million tonnes. There is sufficient detail in the global statistics to be sure of the general spread of taxonomic groups, as only about 13% of recorded production is unassigned. The diversity in this group is such that a number of major taxonomic groups of shrimp and shrimp-like organisms are involved. Three quarters of the total shrimp production is represented by shrimps of the Infraorder Penaeoidea, most of that accounted for by several species of the diverse and widespread Family Penaeidae (2.4 million tonnes). The Sergestidae, a group common in deeper water, make up a major slice of the remaining penaeoids harvested, and most of that is the akiami paste shrimp *Acetes japonicus*, a species used for manufacture of paste, salted, and fermented products. While another major superfamily, the Caridea accounts for only 10% of the global production of shrimp (or 430 000 tonnes), the bulk of that (91%) is represented by a single North Atlantic species of pandalid shrimp, the Northern shrimp, *Pandalus borealis*. The remaining data for the caridean shrimp are mostly taken up by landings of common shrimp *Crangon crangon* (around 37 000 tonnes).

The diversity of the shrimp group is such that it is not possible to discuss all in detail. If we arbitrarily consider species to which landings represent individually more than 10% of total shrimp landings, four species stand out. The paste shrimp and the Northern shrimp (*Acetes japonicus* and *Pandalus borealis*) have already been mentioned, but to this group should be added the giant tiger shrimp, *Penaeus monodon* (on the basis of farmed production!), and the southern rough shrimp, *Trachypenaeus curvirostris*. But this summary perhaps unfairly overlooks the diversity of other species of the Penaeidae such as the banana shrimp *Farfantepenaeus merguiensis*, *P. chinensis*, and *P. aztecus*.

By 1999, aquaculture accounted for over a third of of global shrimp production (1.1 million tonnes). As a market segment, farmed shrimp has been floating around 20–25% of global shrimp landings since the mid-1990s, the upward trend being checked or reversed by outbreaks of shrimp viral diseases at various times around the world, with consequences for supply and pricing felt throughout the market. The farmed shrimp are also penaeids and about half of the total is represented by farmed production of giant tiger shrimp *Penaeus monodon*. The three most notable farmed shrimp in the FAO statistics (in order of decreasing quantities) are the giant tiger shrimp (*Penaeus mondon*), the Pacific white shrimp *Litopenaeus vannamei*, and the Chinese white shrimp *Penaeus chinensis*. *L. vannamei* is farmed in the Americas, while the other two species are produced in South-east Asia. These species also figure highly in wild production statistics, though the farmed production for these species now exceeds the corresponding fisheries production.

Shrimp, which probably owes its dominance to sheer biological abundance as well as simple peeling and high meat recovery (50–60% of fresh weight), lends itself to a diversity of processed fresh/frozen and

packaged, value-added forms. Live marketing of the kuruma shrimp (*Marsupenaeus japonicus*) involves relatively small quantities but is noteworthy for the high prices achieved. Chilled, raw shrimp have a limited shelf-life, so much of the product is processed at sea or on farm in various ways to allow for marketing. This might involve simply packaging and freezing the raw or cooked shrimp in a manner suitable for sea-freight, but more sophisticated processing is increasingly being practiced (Figure 3).

One concern with minimal processing is 'black spot', a form of enzymatic spoilage common to commercial crustacea. This melanosis is mediated by polyphenoloxidases in the shell or hemolymph (blood). (*See* **Browning**: Enzymatic – Biochemical Aspects.) 'Black spot' can be avoided to some extent by removing the shell or by thorough cooking (to deactivate the enzymes), or treating the product to inhibit the formation of melanin. A widespread 'black-spot' treatment is to dip the raw or cooked shrimp in a solution of sodium metabisulfite or related chemicals (which scavenge oxygen, a requirement for melanosis). However, concerns over introducing undesirable sulfite residues into the product has recently led to adoption of other methods. One of these involves an enzyme inhibitor (4-hexylresorcinol), which interrupts the enzyme cascade leading to melanin.

Given the diversity of species, it would be wrong to say that there is one homogeneous global market for these 'commodity' shrimp, though few sectors seem immune from the ups and downs of supply and demand. The meat of different species varies in appearance, flavor, and texture, and this still explains some market segmentation in countries or regions where historically consumers ate particular types or shrimp or prawns. Marine penaeid shrimp have already made inroads into the European market, where *Pandalus borealis* once held sway, but even these penaeids vary. Consider the contrast, in terms of both raw and cooked meat between the highly pigmented giant tiger shrimp and various species of 'white' shrimp marketed.

However, ultimately, national tastes may come to mean little when products are highly processed, breaded, and packaged. Raw shrimp can be headed and the tail split (or 'butterflied'), breaded, and stored frozen in a form ready for cooking. Similarly, raw or cooked shrimp can be shelled and the peeled meat frozen and once again marketed as a convenience product. The processing, freezing, and cooking equipment used varies, with the cost of labor in the country concerned having a major role. The packaging used can range from bulk packs to prepackaged individual serves.

The shrimp sector has not been without troubles. Sporadic disease episodes in the farming sector have impacted upon supply and in recent years have raised concerns about the transfer of shrimp diseases by the global trade in shrimp. In addition, major seafood importing markets such as the USA, the European Union (EU), and Japan have emphasized the quality and safety of food imports by insisting on the adoption of Hazard Analysis Critical Control Point (HACCP) and related schemes. The trade in 'commodity' shrimp has had to fall in line with this, since imports from noncompliant countries have been banned. Various segments of the shrimp industry have also found themselves squarely in the sights of environmental activists concerned with reducing trawl by-catch and stemming the impact of the farming sector. For example, the fate of turtles drowned in trawls prompted calls in the USA for shrimp imports to be banned from countries using gear considered harmful to turtles. This has seen an increased emphasis on 'turtle excluders' and other kinds of by-catch reduction devices amongst world shrimp capture fisheries. The environmental impact

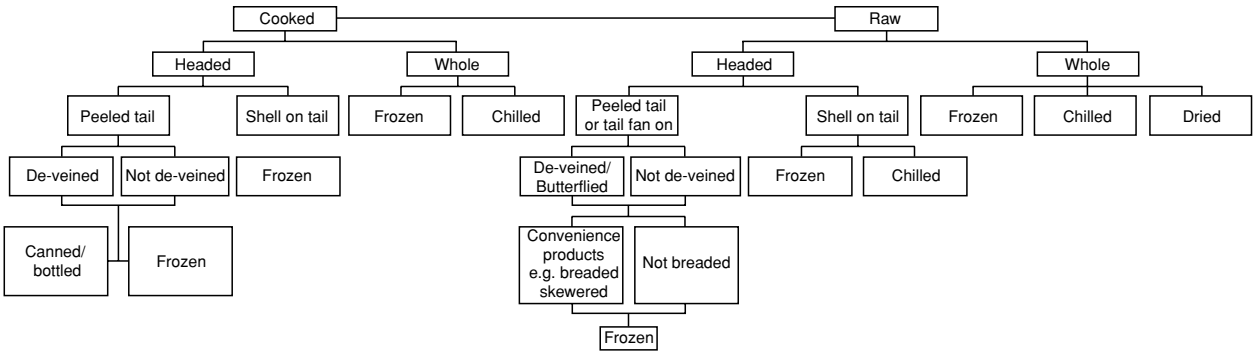

Figure 3 Summary of processing and storage methods applied to raw and cooked shrimp. Based on Lee DO and Wickins JF (1992) *Crustacean Farming*. Oxford: Blackwell Scientific.

of shrimp aquaculture is another area of controversy. All agriculture has some environmental impact, and it would be surprising if any global industry producing around a million tonnes of product annually had no local impact on the environment. Nevertheless, the recurrent problems with disease outbreaks certainly point to areas where farm management has needed to become more sustainable. Producers now appreciate the power that can be wielded by consumer campaigns, and development of more environmentally friendly farming methods is already a major priority.

Marine Crabs

The so-called 'true' crabs and, to some extent, certain 'crab-like' lobsters are the pinnacle of the tendency amongst decapod crustaceans to abandon the pelagic, swimming lifestyle. Their commercial importance no longer relies on tail meat, owing to their rudimentary, indeed vestigial, abdomen. Crabs hide from attack in burrows, buried in sand or beneath rocks, or perhaps rely for protection upon sheer size, a heavy cuticle, and threatening, sometimes well-muscled claws.

The marine production of true crabs in 1999 was about 1.3 million tonnes, or about 16.4% of global crustacean production. It is hard to be definitive as to the contribution of individual species to this total. As much as 26% of the total is represented by statistics that are only assigned to the broadest categories. Still, the available detail reveals a number of significant fisheries for swimming crabs of the Family Portunidae, particular centered in Asia. This group of crabs swim using a pair of modified, paddle-like legs. This, if anything, reiterates the capacity of evolution to reinvent a trait since lost in other crabs. About a third of estimated true crab production is represented by the swimming crabs *Portunus trituberculatus* (22%) and *P. pelagicus* (10%). Another important swimming portunid, the blue crab *Callinectes sapidus* from the east coast of the USA represents just 8% of the crab total. To contrast this approximately 40% share for portunids, the spider crabs (queen, snow, or tanner crabs, *Chionoecetes* spp.) together account for 16% of the total for true crabs. A diversity of other species of crab are harvested, and perhaps a discussion of the true crabs should not omit mention of the edible crabs, Cancridae, represented by various species of *Cancer*.

A small number of crab-like lobsters are not included in this total for true crabs. Marine crab-like lobsters such as stone crabs (Lithodidae, 57 500 tonnes) and squat lobsters (Galatheidae, around 26 500 tonnes) together account for only about 1% of global crustacean production. Taxonomically, these crustaceans belong to the diverse Infraorder Anomura. This assemblage of crustaceans has a diversity of body forms, though again with a trend for the abdomen to shrink. The major species of stone crabs and squat lobsters are, respectively, the Alaskan king crab *Paralithodes camtschaticus* and the red squat lobster *Pleuroncodes monodon*.

Crabs are harvested in a variety of ways, ranging from trawls to pots. Some farm production of portunid crabs has been achieved using *Scylla* spp. and *Portunus pelagicus* in South-east Asia, but it has some way to go before it reaches levels of production typical of shrimp farming. Probably the most unusual method of harvesting a crab or indeed any crustacean is one that takes advantage of the ability of crustaceans to regrow missing limbs. In the fishery for the stone crab *Menippe mercenaria*, only the claws are harvested, and the crab is returned to the sea, to grow new claws! The processing and marketing of newly molted blue crabs *C. sapidus* as 'soft-shell' crabs is another case where crustacean physiology is exploited commercially.

Processing options for crabs are broadly similar to those applied to shrimp, though in contrast to shrimp, many 'minor' species of crabs are often sold, transported, and marketed while alive, giving them a high level of consumer recognition. The live trade is preferred over that for raw chilled or frozen product, because the digestive gland (hepatopancreas) breaks down soon after death, discoloring the meat and releasing digestive proteases that readily soften the meat.

Unlike shrimp and lobsters, crabs cannot be conveniently 'headed' to remove the digestive gland. Instead of 'peeling' cooked crabs, more ingenuity is required to cut or saw through the thicker exoskeleton. The catch from the big tonnage fisheries is processed in large quantities and may reach the consumer in a highly refined form. Crabs taken live to a processor can be cooked (to inactivate the enzymes) and sold either whole/frozen or in a progressively broken apart form, to give packaged/frozen or canned meat. Processing the cooked product involves breaking open the shell and extracting the meat associated with the leg muscles and the claws. Though the details differ between species, picked meat production is probably as close as the crab sector comes to achieving the 'commodity' status as a source of convenience food. Conditions of strict hygiene must be employed during picking, because physically breaking the shell to extract meat presents an opportunity for the introduction of microbial contamination (e.g. *Listeria* and other pathogens), unless the product is further pasteurized. These risks can be addressed using the principles of HACCP.

Significant quantities of crabs are processed by individual factories, which translates into a considerable amount of waste. As companies in some countries are increasingly finding that waste costs money to dispose of, attention continues to be paid to economically recovering saleable material or products (e.g., chitin, astaxanthin, and other biological material) from the waste.

Marine Lobsters

While armored and strictly designed for walking, marine lobsters (and crayfish and indeed some 'shrimp' from freshwater) retain the basic body plan of pelagic crustaceans. The cephalothorax or head of lobsters is more prominent than is generally the case for shrimp, but the abdomen or 'tail' is only slightly diminished and, when danger threatens, can still generate powerful tail-flicks.

Marine lobsters account for only about 3% of total global landings, or nearly 229 000 tonnes. At least three-quarters of this is in the form of clawed Homarid and Nephropid lobsters from the North Atlantic (from the Infraorder Astacura). The American lobster *Homarus americanus*, from the USA and Canada, accounts for almost 40% of the lobster landings, followed by the Norway lobster *Nephrops norvegicus* at about 27%. Lesser quantities of European lobster *Homarus gammarus*, and a few species of other minor Nephropids (e.g., *Metanephrops* spp.) are caught elsewhere. The other major group of lobsters, the Infraorder Panulira – the spiny and rock lobsters – has a more widespread distribution. These lobsters lack claws as such, and instead have prominent spiny antennae. The Caribbean spiny lobster *Panuilrus argus* accounts for 16% of global lobster landings, but there are several other *Panulirus*, *Palinurus*, and *Jasus* species harvested commercially throughout the tropical to temperate seas of the world.

Harvesting methods for lobsters vary from species to species, but potting and trawling are practiced. Clawed lobsters can be produced in hatcheries, but currently, this is for reseeding programs rather than aquaculture *per se*. The numerous long-lived planktonic larval stages of typical panulirids is a major practical obstacle to the aquaculture of panulirids, with aquaculture ventures mostly concentrating on the collection of wild seed. Commercial production of farmed shovel-nosed lobsters *Thenus orientalis* from the hatchery is currently under development, while research continues on aquaculture of other panulirids. Like crabs, lobsters tend to be marketed whole and alive, though frozen product is also distributed cooked whole or tailed. Again, the major objective is to deal with quality loss associated with breakdown of the hepatopancreas. There is also some cutting, chopping, and meat picking from cooked lobsters, to produce both canned and frozen product. Where live transport of lobsters is logistically difficult (e.g., Pacific Islands), there is a preference for marketing frozen raw tails rather than freezing whole animals with the hepatopancreas intact.

Others

Global production of crustaceans from freshwater is estimated at about 817 000 tonnes (about 10% of the total). About half of that is fisheries data from Asia and particularly China, and is not identified to species level. To the extent that the data allow, two cultured species, the Chinese river crab *Eriocheir sinensis* and the giant river prawn *Macrobrachium rosenbergii*, together account for at least 33% of the global production of freshwater crustaceans. Another recognized aquaculture species in the USA, *Procambarus clarkii*, accounts for only 2% of the total landings, and beyond this, the landings are divided amongst numerous species that may of course be relevant in a geographical context. Freshwater crustaceans include numerous freshwater crayfish of the Astacidae, Cambaridae in Europe and the Americas, and a small representation from the Parastacidae (originally from Australia).

Product from freshwater tends to have a milder flavor, largely due to the different demands for osmotic regulation faced by freshwater-adapted crustaceans. Anatomically, commercial freshwater shrimp and crayfish also tend to have relatively smaller abdomens leading to less favorable meat recoveries than from their marine counterparts, though the variety of raw, cooked, and processed products can come close to rivalling that of marine shrimp in variety, if not in quantity. Translocation and release of Cambarid species into Europe have seen the spread of an introduced crayfish disease that has decimated native crayfish populations in some areas. Freshwater shrimp or prawns such as *M. rosenbergii* belong to the Caridea (Palaemonidae).

Another notable group, the krill are small shrimp-like crustaceans that swarm in the plankton of the world's oceans and form an importance food source for marine foods chains (and particularly many whales). These small shrimp are a different order of Crustacea altogether, the Order Euphausiacea. Global landings have varied between 80 000 and 100 000 tonnes for the past few years. One important use of dried krill meal is for aquaculture feed ingredients (*see* **Fish Meal**). Excitement about using krill as a human food source has been tempered in recent

years by a more realistic understanding of the resource and its limitations.

If imitation is considered to be the most sincerest form of flattery, then the very existence of surimi is further evidence of the commercial importance of crustaceans. Surimi is the deboned, washed, and gelled protein manufactured from white fleshed fish. Formed into sticks, flakes and chunks, the edges are coloured red to resemble picked crab meat.

See also: **Browning**: Enzymatic – Biochemical Aspects; **Fish Meal**; **Hazard Analysis Critical Control Point**

Further Reading

Anon (1999) *FAO Fishery Statistics: Capture Production; Aquaculture Production; Commodities*, vol. 88(1–3). Rome: FAO.

Ingle R (1997) *Crayfishes, Lobsters and Crabs of Europe*. London: Chapman & Hall.

Lee DO and Wickins JF (1992) *Crustacean Farming*. Oxford: Blackwell Scientific.

Ruiter A (ed.) (1995) *Fish and Fishery Products: Composition, Nutritive Properties and Stability*. Wallingford, UK: CAB International.

Characteristics of Molluscs

K J Boss, The Agassiz Museum Harvard University, Cambridge, MA, USA

Background

Molluscs constitute a unique phylum of animals, the Mollusca, which are characterized by a combination of morphological and anatomical features separating them from all other invertebrate organisms. Many common names, such as snails, clams, and squids, apply to representatives of these animals, which number fewer than 50 000 living species. Molluscs are known to have diversified early in the fossil record, namely the Cambrian of the Paleozoic over half a billion years ago; the antecedent or ancestral mollusc is presumed to have evolved in the Precambrian and many different lineages have radiated into the vast array of ecological niches of the world's biotopes.

Molluscs are widespread in marine environments, living from the shore to the greatest abyssal depths and occurring in pelagic or oceanic realms as well as benthically, both on and in all kinds of substrates. In some parts of the world, they dominate the intertidal zone, and in the recently discovered aphotic deep-water vents and seeps as well as other oxygen-depleted environments, molluscs, especially mussels and certain clams, are particularly conspicuous, frequently utilizing endosymbiotic chemoautotrophic bacteria. Groups of representative snails and bivalves have also successfully adapted to terrestrial and freshwater habitats. Finally, some species have become specialized endoparasites.

In adult size, molluscs range from tiny gastropods and bivalves less than 1 mm in diameter or length to the giant squid which may be over 15 m long and upwards of 1000 kg in weight.

Typical Morphological and Anatomical Features

Molluscs are protostomous coelomates, exhibiting spiral, determinate cleavage and schizocoely as well as having trochophore larvae and the blastopore forming the mouth of the adult. They usually possess all or a combination of the following features: (1) a reduced coelom and vestiges of metamerism; (2) a mantle or fleshy epidermis of the dorsal body wall which has glands capable of secreting calcium carbonate to form an exoskeletal shell or shelly parts, such as plates, spines, and spicules; (3) a mantle cavity or an invagination of the mantle which contains a pair or more of specialized respiratory structures, the ctenidia or gills, and into which the digestive, metanephridial excretory, and reproductive systems debouch their products; (4) the ventral body surface modified into a pedal groove or muscularized foot for progression or locomotion; (5) a special chitinized, rasp-like tongue or radula; (6) an open hemocelic circulatory system with a chambered heart having auricles and ventricles. Additionally, the nervous system has variously paired ganglionic portions, particularly cerebral, pedal, and visceral ones, as well as ventral anteroposterior cord-like connective elements; specialized sensory structures were evolved for olfaction, vision, balance, and tactile stimulation.

Primitively, these animals were gonochoristic, that is, having the male and female sexes in separate individuals; fertilization was external and the eggs developed into pelagic larvae; hermaphroditism with both sexes in the same individual, brooding of eggs, and ovoviviparity are a few among many of the modifications in the reproductive strategies of these animals.

Taxonomy of the Group

Currently seven classes of living molluscs are recognized. The entirely marine, shell-less Aplacophora,

with at least 250 species, are vermiform animals which have calcareous spicules, scales, or plates embedded in the mantle and which are separable into two principal groups, sometimes considered as independent classes. The more or less cylindrical, gonochoristic Caudofoveata (Figure 1a), having the body somewhat separated into anterior, medial, and posterior sections and possessing an anteroventral pedal shield, attain lengths of 140 mm and live mostly as infaunal burrowers. Fewer than 100 species are known, and they feed mainly on microorganisms and detritus. The free-living, more or less uniformly worm-like, elongate and laterally compressed Solenogastres (Figure 1b, c), with over 200 recognized species, are 1–300 mm in length. Hermaphroditic and predacious, they have a distinct pedal groove, posterior gill folds and live either epibiotically on sediments or epizoically on cnidarians, which usually constitute their principal food.

The Monoplacophora (Figure 1d, e), with fewer than 20 living species, have a controversial fossil history dating from the Early Cambrian; they are benthic, mainly deep-sea animals having a single cap-shaped or limpet-like shell and variously paired organ systems such as the pedal retractor muscles, gills or ctenidia, nephridia or kidneys and gonads, reflecting a primitive metamerism. In monoplacophorans, the foot is large with a flattened, creeping sole, and the pallial mantle cavity is peripheral along the sides of the animal.

The chitons or so-called coat-of-mail shells, separated as the class Polyplacophora (Figure 1f, g), also have a broadened, flat muscular foot and a mantle cavity containing numerous pairs of ctenidia. Consisting of about 600 species of marine benthic epifaunal animals with dorsoventrally depressed bodies, they are unique in having eight calcareous shell plates held together by a strong, peripheral girdle. The body outline is elongate to ovate, and adult individuals range in size from 3 to almost 400 mm in length. Although some chitons are known to occur at depths beyond 7000 m, most live in shallower waters and are mainly grazing herbivores, with a radula having 17 teeth in transverse rows and cusps reinforced with magnetite. The anterior portions of the alimentary canal have pairs of glands for the digestion of carbohydrates, and the broad, large foot is used to creep over, or to attach by suction to, the substrate. Special photosensory structures, the esthetes, are found on the dorsal surface of the valves and may, along with the iron-containing radula, facilitate homing behavior.

The Gastropoda or snails (Figure 2a–c), the largest and most diverse class with as many as 40 000 species, have representatives in most of the world's biotopes and are known from the Early Cambrian. Although most species live in the marine environment, many kinds occur both on land and in fresh water; a few taxa have become specialized internal parasites. Principally, gastropods are univalves, usually with a coiled, apically closed calcareous shell, which also exhibits a great many different shapes and forms, being sometimes merely cap-shaped, greatly reduced, or even altogether lost. The very largest of living snails has a shell over 500 mm in length, the smallest less than 1 mm.

Anatomically, snails and their relatives are basically divisible into head, foot, and visceral mass; they are all characterized by a unique process called torsion which occurs during ontogeny. Primitively, the mantle cavity, into which the alimentary canal, excretory structures, and gonads of the visceral mass empty, contains a pair of laterally disposed respiratory ctenidia or gills and is located posteriorly in the early larval gastropod (Figure 2a). During development, the entire mantle cavity and its contents are twisted or torted through an arc of 180 ° to the right, bringing the cavity itself into an anterior position above and behind the neck and head of the snail (Figure 2b). Internally, torsion causes a peculiar crossing (or streptoneury) of the laterally disposed, paired nervous connections between the anteriorly placed cerebral ganglia and the posteriorly located visceral ganglia.

The head of a gastropod usually has at least one pair of cephalic tentacles, and eyes are often associated with them; however, both structures can be lost. The foot has a creeping sole, and primitively, there are lateral grooves between it and the mantle, forming a so-called epipodium with sensory tentaculate and integumentary organs. The sole of the foot may be subdivided and variously modified for swimming and, in the case of internal parasites, it is entirely lost. The distinctive visceral mass, a dorsal hump, is covered by the mantle which secretes the shell, and contains the internal organs or viscera, including the heart, kidneys, gonads, and alimentary canal. Generally speaking, the jaws and radula assist the mouth in bringing food into the digestive tract; one or more pairs of salivary glands or specialized pouches may contribute enzymes to the buccal cavity or esophagus; the stomach leads through the intestine to the terminal anus and may have a crystalline style, style pouch, typhlosole, digestive diverticula, and a triturating gizzard associated with it. All these structures – the head, foot, visceral mass, mantle and shell – are different and variously altered in the major subdivisions of the Gastropoda.

Traditionally, the snails are subdivided into three major subclasses: the Prosobranchia, Opisthobranchia, and Pulmonata. Recently, with the advent of

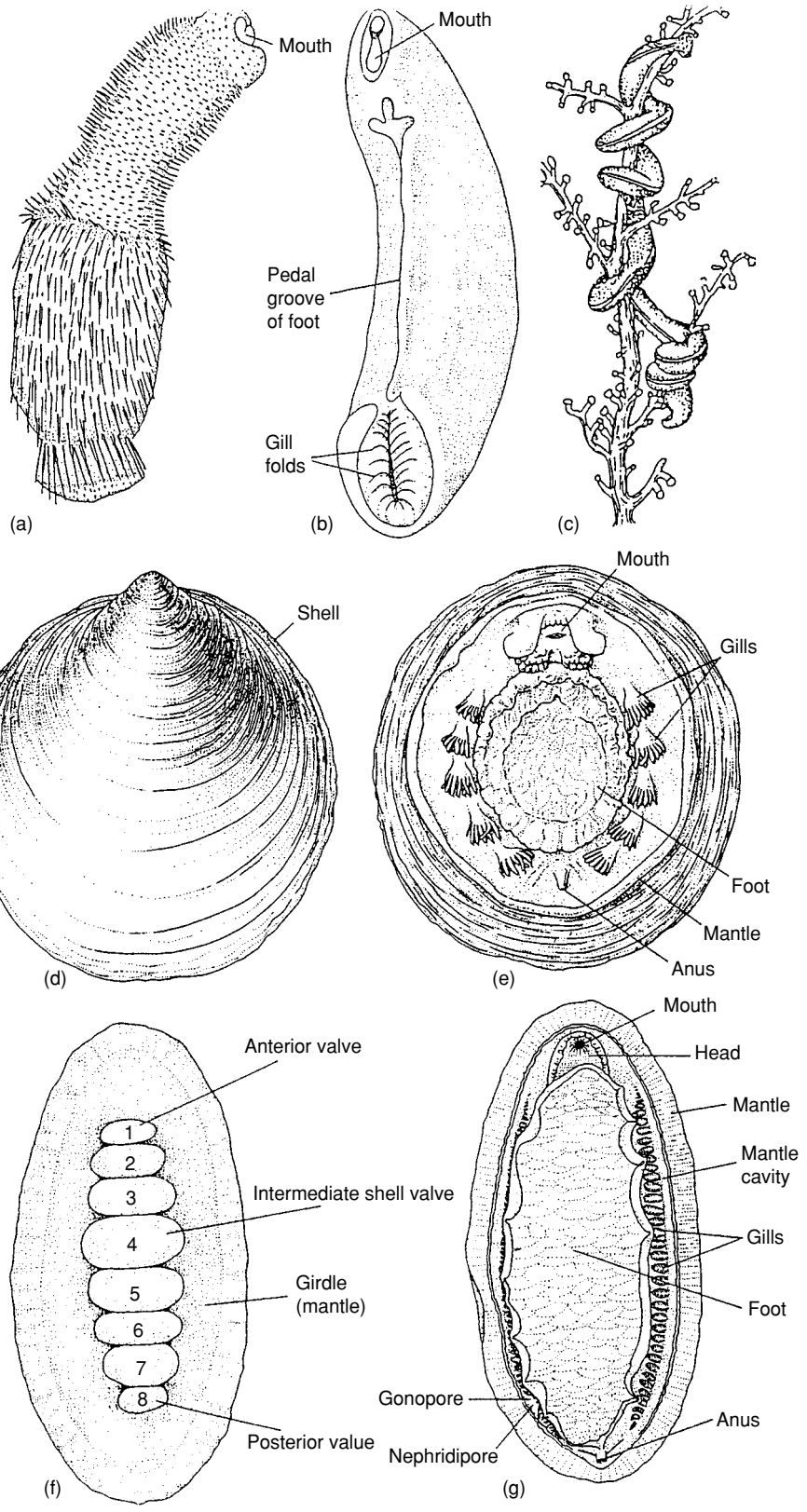

Figure 1 (a) Aplacophoran caudofoveate showing tripartite body form and spicules. (b) Aplacophoran solenogastre showing the ventral pedal groove. (c) Aplacophoran solenogastre on a cnidarian. (d) Dorsal view of monoplacophoran shell. (e) Ventral view of monoplacophoran showing gross anatomical detail. (f) Dorsal view of polyplacophoran with the shell plates and girdle. (g) Ventral view of polyplacophoran showing gross anatomical detail. Reproduced from Shellfish: Characteristics of Molluscs, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

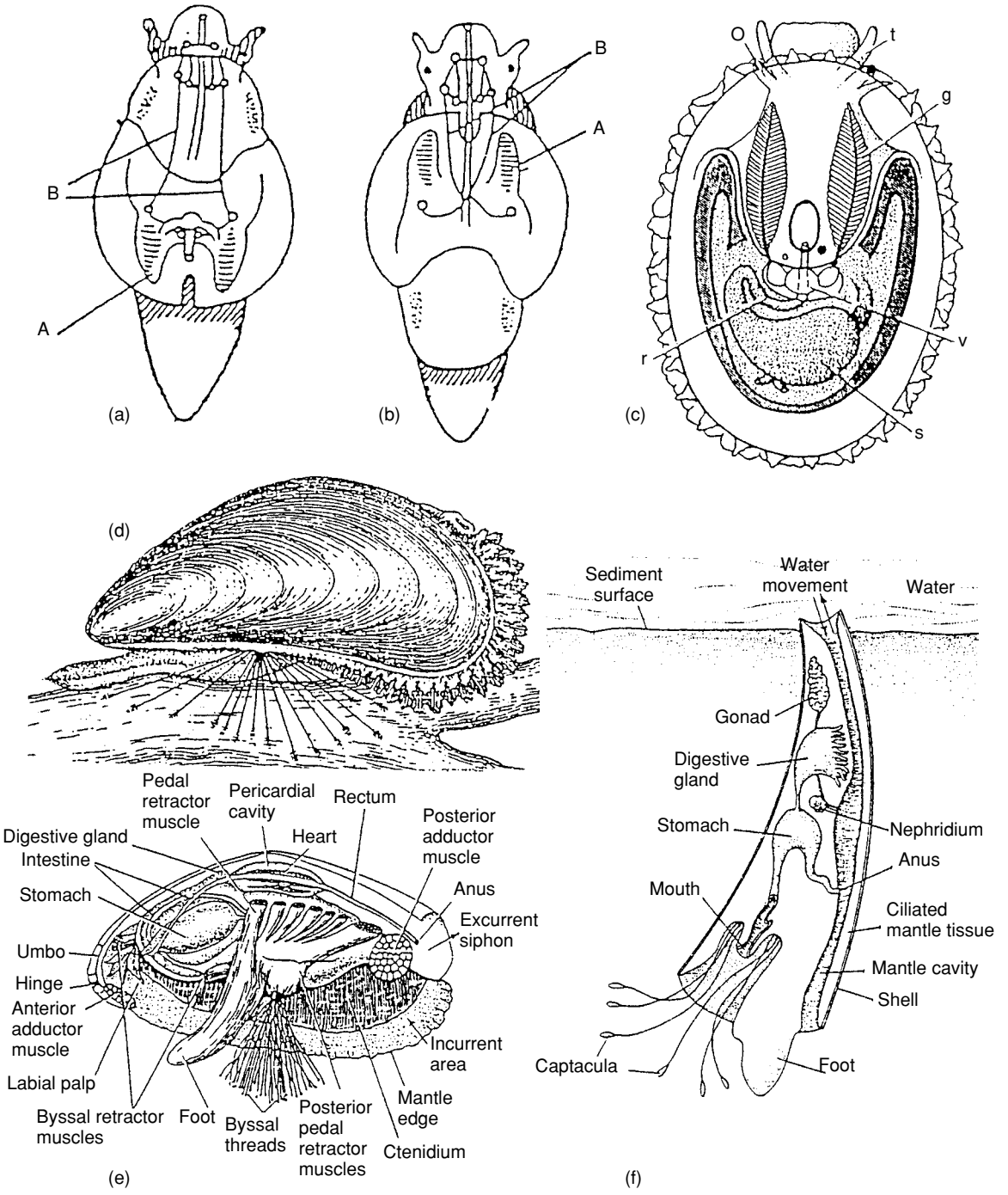

Figure 2 (a) Pretorsional gastropod showing ctenidium (A) in posterior mantle cavity and uncrossed cerebrovisceral connective nerves (B). (b) Posttorsional gastropod showing ctenidium (A) in anterior mantle cavity and crossed cerebrovisceral connective nerves (B), or streptoneury. (c) Primitive prosobranch gastropod showing: g, ctenidium; o, eye; r, rectum; s, stomach; t, tentacle; v, ventricle of heart traversed by rectum. (d) External view of a bivalve mussel with its byssal attachment. (e) Internal view of a bivalve mussel with gross anatomical detail. (f) Internal view of a scaphopod with gross anatomical detail.

cladistic analysis and the discovery of unusual forms in the deep sea, authorities have begun to dismantle the more conservative classification system long in use by elaborating complex split taxonomic systems incorporating numerous new higher-level taxa; but at least 350 families constitute the diversity of gastropods, about 150 in the prosobranchs and 100 each in pulmonates and opisthobranchs, and if one

includes the fossil taxa, there are about 8000 different genera though many more than this have been named.

Prosobranchs (Figure 2c), with the gills anterior to the heart, form the most primitive and most numerous of gastropods with more than 20 000 species; the shell is almost always present and usually spirally coiled, although it may sometimes be patelliform, tubular, or lost in the adult; usually the foot with its creeping plantar sole bears a calcareous or corneous operculum, a structure capable of covering the aperture of the shell when the animal is retracted. As a rule, the mantle cavity is anteriorly directed and it contains, in the most primitive prosobranchs, a pair of respiratory structures (the ctenidia), a pair of chemosensory organs (the osphradia), and a pair of mucus-secreting structures (the hypobranchial glands – sometimes also a source of purple pigments that are a problem in commercially important species); the alimentary canal as well as the excretory and reproductive tracts empty into the mantle cavity. The ventricle of the heart is, in the most primitive forms, traversed by the rectum; there are also paired lateral auricles. In more advanced prosobranchs, for example, the so-called neogastropods, the mantle cavity becomes somewhat detorted and shifted toward the right; the paired structures, e.g., the gills, osphradia, hypobranchial glands, auricles, and kidneys, are reduced to single ones; siphonal structures, elaborated by folds of the mantle, direct the flow of water into and out of the mantle cavity and thus facilitate respiration as well as the removal of excretory and alimentary waste products. The nervous system is streptoneurous, with the cerebrovisceral connectives crossed, and the animals are usually gonochoristic. Prosobranchs primitively are herbivorous grazers or detritus feeders, using the radula to scrape edible material off the substrate, but different evolutionary lines have culminated in many highly specialized carnivores, frequently accompanied by changes in, or even loss of, the radula. According to the kind of preferred diet, great differences obtain in the nature of the alimentary canal with its radula and various associated digestive glands.

On more than one occasion, prosobranchs have given rise to lineages which have invaded fresh-water and terrestrial habitats, and, of course, concomitant alterations, especially in the respiratory and reproductive structures have resulted. Land prosobranchs tend to be more tropical or subtropical in their distribution while the fresh-water groups are widely successful in both the tropics and temperate zones. In the marine realm, prosobranchs are the predominant snails from the intertidal areas over the continental shelf and slope down to the greatest depths.

The opisthobranchs, with the gills, if present, usually behind the heart, are a numerically much smaller group than either prosobranchs or pulmonates, with fewer than 2000 species; their hallmarks include the progressive loss of the shell and operculum in adults, further detorsion and loss of streptoneury, wherein the cerebrovisceral connectives are not crossed (euthyneury), and finally, hermaphroditism.

With only a few fresh-water representatives, this subclass of snails is virtually entirely marine and includes herbivores, vigorous, rapacious carnivores, and floating mucus-net suspension feeders. Probably their most spectacular representatives are the nudibranches, which, as adults, lack the shell and mantle cavity with its ctenidium; instead, they have sometimes elaborated separate respiratory structures, such as cerata and anal or lateral gills, which, along with the mantle, may be brilliantly colored.

The third subdivision of gastropods, the Pulmonata, with about 15 000 species, are, with a few exceptions, entirely fresh-water and terrestrial; these animals are hermaphroditic and detorted, euthyneurous, lack an operculum, ctenidium, or osphradium, and have transformed the pallial cavity into a special, highly vascularized pulmonary or lung-like cavity which facilitates gaseous respiration and opens to the exterior by means of a pneumostome, a narrowed, contractile aperture. Generally, a shell is present, spirally coiled, and variously shaped or reduced to a patelliform or flattened disk; sometimes, the shell is wholly enveloped by the mantle or totally lost, an evolutionary phenomenon adopted by several different lines of descent and leading to a slug-like condition as an adult. In contrast to the prosobranchs, the pulmonate radula shows a remarkable degree of uniformity, mostly being adaptations for herbivory, although some lineages of these snails have become carnivores.

Fresh-water pulmonates, though occurring in tropical and subtropical waters, are particularly successful in exploiting marginal, sometimes ephemeral habitats; they are widespread but more dominating in the cool and temperate regions of the Paleoarctic. Terrestrial lunged snails are cosmopolitan in distribution, with some taxa found on the most remote islands or in the most rigorous deserts. Although largely living on the ground or in leaf-litter, some representatives of these snails have become arboreal, adapting to conditions on the bark, branches, or leafy canopy of trees. In these taxa, often the shells are brilliantly colored and exhibit great polymorphism.

A sister-group of the Gastropods, the Cephalopoda, including squids and octopuses, with some 600 living species are also entirely marine organisms

which have almost exclusively reduced, lost, or internalized the shell, reorganized the body, and adopted a predatory mode of feeding either in the open ocean or along the bottom of the sea; these animals are discussed elsewhere. (*See* **Marine Foods**: Edible Animals Found in the Sea.)

The Bivalvia (Figure 2d, e), comprising familiar organisms such as clams and mussels, with fewer than 8000 species and over 100 families, are aquatic, marine and fresh-water, epifaunal or infaunal, predominantly deposit or filter-feeding animals characterized by having bilaterally disposed shells or valves which can be tightly closed by adductor muscles. In size, living bivalves are known to range from 1–2 mm up to over 1000 mm such as the giant clams of the reefs of the Pacific Ocean. Formed by specialized distal portions of the flap-like mantle, the valves are connected dorsally by a hinge which may consist of a ligament and so-called dental elements. The body is generally laterally compressed; the spatulate or lanceolate foot is variously developed, adapted for creeping, burrowing, or cleansing, and, in attached or cemented forms such as oysters, it may be greatly reduced or lost. A pedally secreted byssus normally occurs in bivalves, though sometimes only in the young embryonic stages. In the reduction of a distinct cephalic region, these animals lost the radula and evolved specialized palps on either side of the mouth which assist in sorting the filtered, particulate food. Their nervous system is decentralized and more specialized posteriorly in conjunction with the posterior development of mantle openings or siphons.

In bivalves, the mantle cavity shows marked expansion and deepening, occupying much of the space between the shells, and the ctenidia exhibit a significant elaboration in most taxa. These form greatly enlarged structures consisting of lateral pairs of plates or sheets of filaments, in which there are blood vessels for respiration and on the surfaces of which are elaborated a complex ciliation, facilitating the filtering of edible materials from plankton suspended in the water column or microorganisms from the detritus or bottom sediments. Mainly gonochoristic with external fertilization, bivalves may also be hermaphroditic and sometimes protandrous, with free-swimming or parasitic larva; retention of the young in the mantle cavity or in special incubatory marsupia in the gills occurs in many species.

Living benthically both in fresh and saline waters, clams and their relatives have exploited many of the biomes of aquatic habitats; they occur from shallow intertidal waters to the greatest abyssal depths of the oceans; they utilize both soft sediments and hard substrates, attaching, burrowing, and even boring into limestones, other shells, or wood. Fresh-water bivalves include the tiny so-called fingernail clams which almost become terrestrial, sometimes living in moist litter, and the pearly mussels, widely distributed on continental areas and even oceanic islands. One group of small marine epizoic and commensal clams has evolved as internal parasites in echinoderms.

Closely related to the bivalves, the Scaphopoda (Figure 2f), commonly known as tusk or tooth shells, with about 350 species, first occur in the Middle Ordovician over 450 million years ago. With a bilaterally symmetrical body occupying an elongate, gently curved, tubular shell which is open at both ends, scaphopods have fused their lateral mantle flaps, reduced their cephalic region, and lost their ctenidia, replacing these respiratory structures with crenulations or folds in the mantle cavity. Calcareous and composed of three layers, the shell is attenuated into a narrow posterior opening; the external shell surface is variously smooth or sculpted, and in size the animals range from 2 to 150 mm in length in adults, though one fossil species at the end of the Paleozic attained nearly 300 mm. Located on a protrusive proboscis, the mouth is surrounded by long filiform, tentacular organs called captacula which probe the sediment and collect, by adhesive glands and ciliary tracts, tiny interstitial microorganisms such as benthic foraminiferans, which constitute their food.

Entirely marine and relatively conservative in their radiation, scaphopods are shallow infaunal burrowers found throughout the world's oceans from the sublittoral to deep abyssal zones below 6000 m. Preferring soft substrates, they flourish mainly in muddy, sand bottoms with the posterior portion of the shell with its aperture projecting from the substrate.

See also: **Marine Foods**: Edible Animals Found in the Sea; **Shellfish**: Commercially Important Molluscs; Aquaculture of Commercially Important Molluscs and Crustaceans

Further Reading

Boss KJ (1982) Mollusca and classification of Mollusca. In: Parker SP (ed.) *Synopsis and Classification of Living Organisms*, vol. 1, pp. 945–1166; vol. 2, pp. 1092–1096. New York: McGraw-Hill.

Pojeta J, Runnegar B, Peel JS and Mackenzie G Jr (1987) Phylum Mollusca. In: Boardman RS, Cheetham AH and Rowell AJ (eds) *Fossil Invertebrates*, pp. 270–435. Oxford: Blackwell Scientific.

Wilbur KM (ed.-in-chief) (1983–88) *The Mollusca*, 12 vols. New York: Academic Press.

Commercially Important Molluscs

P F Duncan, The University of the Sunshine Coast, Maroochydore DC, Queensland, Australia

Background

The commercially important molluscs belong to the classes Bivalvia, Gastropoda, and Cephalopoda, with this article covering the first two groups (for cephalopods, *see* **Marine Foods:** Edible Animals Found in the Sea). Bivalves and gastropods constitute a huge fishery resource, amounting to over 12 million tonnes from marine harvest fisheries and aquaculture in 1999, and over 550 000 tonnes from freshwater. The production methods are as diverse as the species, and products range from high value, such as abalone, to low value, such as periwinkles. This article provides details on the size and diversity of the commercially important species and includes information on major production methods and countries, product types, and current issues.

Vernacular terms to describe edible molluscs vary around the world and may cause confusion, e.g., 'clam' may refer to at least three distinct taxonomic groupings. Vernacular names will be used only for important species, with scientific taxonomy used generally.

Bivalves

There are approximately 20 000 bivalve species worldwide, with the principal commercial species belonging to three taxonomic orders: Mytiloida (including mussels, scallops, and oysters), Veneroida (true clams and cockles), and Myoida (soft-shell clams and geoducks).

Bivalves are generally characterized by paired shell valves joined at a dorsal hinge and a sessile, filter-feeding way of life. Their internal body form consists of adductor muscles (usually paired) for shell closure, the mantle for shell secretion, gills for feeding and respiration, a digestive gland, and seasonally variable gonad tissue.

These features influence their commercial importance and use as food products. Generally, bivalves are consumed whole, e.g., oysters and mussels, or processed to adductor muscles (meat) and/or gonad (roe), e.g., scallops. Muscle tissue is rich in carbohydrate (glycogen) and protein, and mature gonad tissue is high in lipid.

Mussels

Production Commercially important mussels belong to the family Mytilidae and are predominantly marine (see Table 1). Freshwater mussels involved in pearl production belong to the Order Unionida. Commercial *Mytilid* species come from wild harvest fisheries (global production of 237 823 tonnes in 1999) and a greater proportion from aquaculture (1 451 032 tonnes in 1999). Global mussel aquaculture at first sale was worth US$0.52 billion in 1999.

Harvest fisheries Harvest fisheries target three main species, *Mytilus edulis* (Blue) (121 964 tonnes in 1999), *M. galloprovincialis* (Mediterranean) (55 819 tonnes) and *Perna viridis* (Green) (21 520 tonnes). The main countries harvesting *M. edulis* were Denmark (96 215 tonnes), Canada (11 565 tonnes), UK (7972 tonnes), USA (4086 tonnes), and France (2098 tonnes). Production of *M. galloprovincialis* comes from Italy (37 876 tonnes), Greece (15 860 tonnes), and Turkey (1800 tonnes). Thailand dominates production of *P. viridis* (21 500 tonnes). Mytilid mussels are principally caught by dredging flat seabed areas where the animals occur in dense aggregations, connected by byssus, which are proteinaceous threads used for attachment in many bivalves.

Aquaculture Aquaculture production is also dominated by *M. edulis* at 498 461 tonnes, worth US$0.27 billion, and ranked as the 15th species by quantity under cultivation. The principal producing countries are Spain (261 969 tonnes), The

Table 1 Principal commercial mussels, including main production sources and methods

Species	*Common name*	*Main production sources and method*
Mytilus edulis	Blue	Europe/North America/China (F/A)
Mytilus galloprovincialis	Mediterranean	Europe, South Africa (F/A)
Mytilus chilensis	Chilean	Chile (F/A)
Aulacomya ater	Cholga	Chile, Peru (F)
Perna viridis	Green	South-east Asia (F/A)
P. perna	South American rock mussel	Brazil, Venezuela (A)
P. canaliculus	New Zealand or green lipped	New Zealand (A)

A, aquaculture; F, fisheries.

Netherlands (100 800 tonnes), France (51 600 tonnes) and Germany (37 912 tonnes).

Also important is *M. galloprovincialis* at 162 179 tonnes, valued at US$100 million. Italy (130 000 tonnes), Greece (16 912 tonnes), and France (10 900 tonnes) dominate production. The green-lipped mussel (*Perna canaliculus*) (not to be confused with the green mussel, *P. viridis*) is also a significant aquaculture species at 71 000 tonnes, worth US$21.3 million, with production dominated by New Zealand. The other main species, *P. viridis*, provides 68 509 tonnes, is worth US$8.3 m, and is cultivated in Thailand (40 300 tonnes) and the Philippines (15 478 tonnes).

It should be noted that China is the dominant aquaculture producer of mussels at 608 115 tonnes, worth US$90 million. International statistics do not recognize individual species composition in Chinese production, but other sources indicate that *M. edulis* dominates, with additional supply from *M. coruscus, Musculus senhousei*, and *P. viridis*. All are local species. Mussel aquaculture is principally based on suspended longline or raft culture, although other methods exist. (*See* **Shellfish**: Aquaculture of Commercially Important Molluscs and Crustaceans.)

Product types Mussels are traditionally a chilled, live product, particularly in Europe, where a post-harvest shelf-life of around 5 days is expected. Recently, new packaging and processing methods, greater acceptance of convenience foods, and value adding have led to new products. Mussels are now frequently sold as individually quick frozen (IQF) cooked in the whole or half shell, precooked ready-to-eat meals and high value, luxury products such as smoked mussels. Traditional pickled or brined preserved products are common.

Issues As coastal filter feeders, where the whole animal is consumed, food safety is important for mussels and other bivalves. Accumulation of bacteria, viruses, toxic chemicals, elements, and biotoxins is associated with bivalves. Most edible bivalves, including mussels, are generally subject to purification, or depuration, which places the shellfish in a clean water flow for 12 h or more. This enables expulsion of many potentially harmful contaminants. Health certification of live mussels is now common, as are monitoring and certification of growout waters.

However, depuration does not remove biotoxins, produced by microalgae (more accurately phycotoxins), and their control and impact on mussel production are generally managed by long-term monitoring programs. These toxins may result in illnesses such as paralytic, diarrhoetic and amnesic shellfish poisoning. In severe biotoxin outbreaks, area closures may be necessary, and these have been well reported in recent years. The protozoan parasites *Marteilia* spp. affect *Mytilus* spp., although generally, mussels appear to little affected by disease in culture systems.

Scallops

Production Commercial scallops belong to the family Pectinidae, and the principal species are listed in Table 2. Global harvest fishery production in 1999 was 567 507 tonnes and from aquaculture 951 866 tonnes. Aquaculture production has increased from 340 807 tonnes in 1990, with an equivalent increase in value from US$0.55 billion in 1990 to US$1.35 billion in 1999.

Harvest fisheries Fisheries are dominated by *Patinopecten yessoensis* (305 510 tonnes), *Placopecten magellanicus* (131 962 tonnes), *Pecten maximus* (35 411 tonnes), and *Argopecten purpuratus* (30 141 tonnes). Japan (299 628 tonnes, 98%) is the main

Table 2 Principal commercial scallops, including main production sources and methods

Species	*Common name*	*Main production sources and method*
Pecten maximus	Great/king scallop	Europe (F/A)
Pecten novaezelandiae	New Zealand scallop	New Zealand (F)
Pecten fumatus/meridionalis	Commercial scallop	South-eastern Australia (F)
Placopecten magellanicus	Sea scallop	USA, Canada (F)
Chlamys islandica	Iceland scallop	Iceland, Canada, USA (F)
Chlamys farreri	Zhikong scallop	China (A)
Patinopecten yessoensis	Yesso scallop	Japan, Russia, China (F/A)
Patinopecten caurinus	Weathervane scallop	Canada, USA (F)
Argopecten irradians	Bay scallop	China, USA (F/A)
Argopecten purpuratus	Peruvian Calico scallop	Chile, Peru (F/A)
Argopecten ventricosus	Pacific Calico scallop	Mexico (F/A)
Aequipecten opercularis	Queen scallop	Europe (F/A)
Amusium balloti	Saucer scallop	Western and eastern Australia, Japan (F)

A, aquaculture; F, fisheries.

producer of *P. yessoensis*, with the Russian Federation (5764 tonnes, < 2%), providing the remainder. *P. magellanicus* production is from the USA (77 179 tonnes, 59%) and Canada (54 756, 41%), and *P. maximus* fishery production derives mainly from the UK (19 108 tonnes, 54%) and France (12 745 tonnes, 36%). Peru provided all production of *A. purpuratus* in 1999.

Wild harvest of all the main commercial scallop species usually involves a toothed (e.g., *Pecten* spp.) or smooth-framed dredge (e.g., *Placopecten* spp.), or occasionally conventional trawl gear (e.g., *Amusium* spp., *Argopecten gibbus*). There are several large-scale diver fisheries for some species, e.g., Chile, China, and Mexico, and some small fisheries that supply high-value, live product markets, e.g., Scotland.

Scallop fisheries are noted for catch fluctuations, as a result of exploitation and recruitment variability. Several fisheries have consequently declined or closed, although notably, the successful application of stock enhancement and management programs helped to revive Japanese and New Zealand production during the 1970s and 1990s.

Aquaculture Three countries and two species dominate scallop aquaculture production. In 1999, China was the largest producer of *P. yessoensis* at 712 330 tonnes, followed by Japan at 216 017 tonnes. Chile produced 20 668 tonnes of *A. purpuratus*, with the remainder from Peru (1585 tonnes). On a much smaller scale, and using various species, other countries are developing pectinid aquaculture. For example, in 1999, the UK produced 114 tonnes of *Aequipecten opercularis*, whereas France, Spain, Ireland and the UK produced over 700 tonnes of *P. maximus*.

Scallop production from aquaculture is based on suspended longline culture or on some form of seabed culture.

Product types Scallops are sold in processed form; fresh or frozen, meat (muscle) only or meat with gonad (roe) attached, although there is a small, valuable market for live scallops, particularly in Western Europe. Processing removes the meat (and gonad) from the remaining viscera and shell, a procedure known as shucking. This is usually a manual operation, although machines are available. Usable product may be block frozen in multi-kilogram amounts or IQF if the product is of a higher value and quality.

In scallop marketing, the term *Coquille St Jacques*, refers to the Genus *Pecten* and is specified under French law. This category is important because of France's position as one of the principal scallop markets. True *Pecten* species are sold as adductor muscle and gonad together (roe-on) and command higher prices. This premium relates partly to selling gonad also, providing a greater return per scallop.

China is unique in predominantly producing a dried scallop meat product, although small quantities are locally sold live.

Issues Because of the type of product (meat and/or gonad only), scallops are less affected by contamination and biotoxin problems, as the contaminated portions (mainly mantle and digestive gland) are discarded during processing. However, microalgae biotoxins can concentrate in meat and gonad, and have resulted in production closures in some countries (e.g., UK) and the widespread establishment of precautionary monitoring programs (e.g., New Zealand, UK, various states in the USA).

Diseases affecting scallops appear to be rare compared with other bivalves, with *Rickettsia*-like and *Chlamydia*-like organisms occasionally associated with mass mortalities of wild scallops and hatchery mortalities. Outbreaks may be associated with scallop density. The protozoan *Marteilia* spp. has been implicated in mass mortalities of wild scallops in the USA, although such events are rare.

Oysters

Production Commercial oysters belong to the family Ostreidae, and principally to the genera *Ostrea* and *Crassostrea* (see **Table 3**). The global wild harvest in 1999 was 157 538 tonnes, and aquaculture production 3 711 606 tonnes, valued at US$3.4 billion. Aquaculture production has increased from 1 251 660 tonnes since 1990 (value US$1.57 billion).

Harvest fisheries The dominant fishery species in 1999 was the American oyster, *Crassostrea virginica* (132 207 tonnes, 84% of total), followed by the Pacific (cupped) oyster, *C. gigas*, at 12 271 tonnes (8%). The genus *Ostrea* contributed an additional 5951 tonnes, the majority (2456 tonnes, 41%) from *O. edulis*, the European flat oyster, the remainder from *O. lutaria* and other minor species.

Fishery production of *C. virginica* in 1999 was from three countries: the USA (89 714 tonnes), Mexico (39 268 tonnes), and Canada (3225 tonnes). Almost all (11 609 tonnes) of *C. gigas* production comes from China, Taiwan, and South Korea. Fishery production of *O. edulis* derives from Turkey (840 tonnes), Ireland (680 tonnes), Spain (443 tonnes), and the UK (407 tonnes). New Zealand produced 994 tonnes of the dredge oyster, *O. lutaria*, and Mexico 2489 tonnes of unnamed *Ostrea* spp. European *Ostrea* fisheries were

Table 3 Principal commercial oysters, including main production sources and methods

Species	*Common name*	*Main production sources and method*
Crassostrea gigas	Pacific (cupped) oyster	China, Japan, Korea, France, USA, New Zealand (F/A)
Crassostrea virginica	American oyster	USA, Canada (F/A)
Crassostrea iredalei	Slipper oyster	Philippines (F/A)
Ostrea edulis	European flat oyster	Europe (F/A)
Ostrea lutaria	Dredge oyster	New Zealand (F)
Ostrea chilensis	Chilean flat oyster	Chile (F/A)
Saccostrea glomerata	Sydney rock oyster	Australia (A)

A, aquaculture; F, fisheries.

formerly more significant, but the protozoan, *Bonamia ostreae*, introduced from the USA, devastated production in the 1980s. Commercial fishery methods for oysters are predominantly dredge-based, although hand picking occurs in some countries, e.g., USA, Australia.

Aquaculture The dominant aquaculture oyster is the Pacific (cupped) oyster, *Crassostrea gigas*, providing the greatest quantity of any aquaculture species in 1999 at 3 600 459 tonnes. This species accounted for 97% of all oyster aquaculture production with a value of US$3.3 billion in 1999. China accounted for 2 988 613 tonnes, with other major producers including Japan 205 345 tonnes, Korea 177 259 tonnes, France 134 800 tonnes, and the USA 33 259 tonnes. Other aquaculture species are *C. virginica* 57 522 tonnes (54 037 tonnes, USA), *C. iredalei* 13 698 tonnes (Philippines), *Ostrea edulis* 6496 tonnes (Spain 3348 tonnes, France 2300 tonnes), and *Saccostrea glomerata (commercialis)* 5104 tonnes (Australia). Production methods vary, although most are based on the oyster's natural attachment to artificial structures, or seabed sowing.

Product types Oysters are best known as live seafood, and most production is live, in-shell, or fresh chilled in half-shell. Traditionally, oyster culture has produced a bottled product, with the whole shucked animal, washed and preserved in salt water. They survive low-temperature storage for several weeks, but half-shell and bottled oysters have a limited shelf-life. Maintenance of quality and rapid transport and sale are essential components of the industry. Oysters are now sold as a frozen product, either in whole or half shell. This maintains the traditional appearance while limiting spoilage and health risk. Other value-added products such as smoked oysters and pre-cooked meals are becoming common.

Issues Oysters are subject to many significant diseases affecting fishery and culture operations. These include the protozoans *Bonamia* spp. (particularly affecting *Ostrea* spp.), *Marteilia* spp. (affecting *Ostrea* and *Saccostrea* spp. (causing QX (Queensland unknown) disease in the latter), *Mikrocytos* spp. (affecting several *Crassostrea*, *Saccostrea*, and *Ostrea* spp.), *Haplosporidium* spp. (affecting *Crassostrea* spp.), and *Perkinsus* spp. (affecting *Crassostrea* and *Haliotis* spp. (abalone)). An iridovirus, causing gill necrosis, is also associated with hatchery-reared *C. gigas*.

From a human health perspective, oysters are often associated with food poisoning events, because of their intake of bacteria and viruses and subsequent consumption as a whole, raw product. Such events focus attention on the production environment and require management of local sources of pathogenic effluent. Consequently, water-quality monitoring and health certification of production areas are common, along with a certification requirement from importing countries.

Other Bivalves

Production International fisheries statistics combine true clams and cockles (order Veneroida) with arks (order Arcoida, *Anadara* spp.) and indicate total world production from fisheries at 812 501 tonnes, with a further 2 744 846 tonnes from aquaculture, alone worth US$3.15 billion. This constitutes a commercially significant group, despite including numerous species. They are treated under broad groupings, with taxonomic relationships indicated where appropriate.

Order Veneroida This includes, amongst many others, the families Veneridae (venus clams), Cardiidae, (cockles), and Tridacnidae (giant clams). Several *Ruditapes* spp. (carpet shells), support a modest harvest fishery, mainly Japanese, producing around 58 874 tonnes in 1999. The most important carpet shell is *Ruditapes philippinarum* (Manila clam), which also provided 1 820 413 tonnes (worth US$2.19b) from aquaculture, mainly from China. Other cultivated *Ruditapes* spp. added 54 421 tonnes and US$0.13 billion. The fishery for *Chamelea*

gallina (striped venus) was 45 012 tonnes in 1999, and mainly from Italy (36 462 tonnes).

Meretrix spp., known as hard clams, are fished in South-east Asia, and provided 22 100 tonnes in 1999. Indonesia is the largest supplier with 17 990 tonnes. This genus is also cultivated in South Korea (*Meretrix lusoria*, 17 tonnes), but more widely in Vietnam, where 80 000 tonnes of *M. meretrix* and *M. lyrata* are produced per annum. This product is sold into South-east Asia, Japan, and Europe, although EU regulations regarding biotoxin monitoring have limited the European market.

Arctica islandica (Ocean or black Quahog) is an important fishery species in North America, with the USA producing 147 933 tonnes in 1999. Iceland also produces a small quantity (3501 tonnes) through harvest fisheries, most of which is exported to the USA.

Mercenaria mercenaria (also known as the northern Quahog or hard-shell clam) is also principally a North American species, with Canada and the USA traditional producers. In recent years, only Canada has provided a harvest fishery product of 2536 tonnes (1999), with a further 53 050 tonnes, worth US$63.4 million, from aquaculture, equally from China and the USA. As with most clam culture, spat are sown on to preselected sand/mud bottoms at densities appropriate for good growth and survival and with well-exchanged phytoplankton rich water. Harvesting is by dredge. To reduce growout periods, some clams are cultured on trays, either suspended or stacked off the bottom. This improves growth through increased exposure to nutrients and higher temperatures.

Fishery production of *Paphia* spp. (short-necked clams) was 42 282 tonnes in 1999, from Thailand and the Philippines. The North Atlantic surf clams, *Spisula solidissima* and *S. polynyma*, are produced from harvest fisheries in the USA (142 370 tonnes in 1999) and Canada (26 722 tonnes), respectively. Razor clam (*Solen* spp.) aquaculture in China is significant with 479 252 tonnes, worth US$383 million, produced in 1999. Additional sources indicate that this species may also be reported as *Sinonovacula constricta*.

The European edible cockle, *Cerastoderma edule*, is dredge-fished in The Netherlands (50 888 tonnes) and the UK (14 123 tonnes), for a total production of 70 401 tonnes in 1999. Recent aquaculture, based in Spain (3713 tonnes) and France (1300 tonnes), also contributed 5169 tonnes to production, valued at US$8.9 million. Aquaculture produced very small quantities (2 tonnes, in 1998) of cockles (Family Cardiidae) in the USA.

Order: Myoida The soft-shell clam, *Mya arenaria*, and Geoducks, *Panopea abrupta* are produced and consumed in North America. In 1999, the USA and Canada produced 5113 and 2643 tonnes of *M. arenaria*, and 2451 and 1800 tonnes of *P. abrupta*, respectively.

Order: Arcoida There are three main genera of Ark 'clams' of commercial importance, *Arca* spp., *Scapharca* spp., and *Anadara* spp. Fishery production of *Arca* spp. and *Scapharca subcrenata* totalled 52 972 tonnes in 1999. Aquaculture of *S. broughtonii* in South Korea produced 8550 tonnes, worth US$49.8 million.

Dredge fisheries also produced 42 627 tonnes of *Anadara* spp. in 1999, with *A. granosa* (misleadingly called the blood cockle) the main species. Aquaculture of *A. granosa* was 315 811 tonnes, worth US$277.4 million, mainly from China (188 355 tonnes) and Malaysia (79 912 tonnes).

Gastropods

The class Gastropoda is commercially less important, although the marine abalone, periwinkles, conchs and whelks, and terrestrial snails are significant as food products. Other species of gastropod are valued as shell products, such as mother of pearl and specimen shells. Gastropods are generally characterized by a single spiral shell and large muscular foot, enabling greater motility than bivalves and forming the principal edible portion. Fishery production of gastropods, as abalone, winkles, and conchs was 105 047 tonnes in 1999. Equivalent aquaculture production was 2694 tonnes, worth US$32.4 million.

Abalone

Production Abalone (family Haliotidae) is mainly a temperate marine group, although they extend into tropical areas. Approximately 50 species are recognized, associated with rocky habitats off Australia, New Zealand, Japan, Pacific North America, and South Africa. The main commercial species are shown in Table 4. Abalone production is from divercollected harvest fishery or aquaculture.

Harvest fisheries The total fishery production was 10 453 tonnes in 1999, the majority from Australia (5593 tonnes of *Haliotis rubra*), followed by Japan (2109 tonnes of *H. gigantea*). The value of combined fisheries production has not been estimated, but, as an indicator, the Australian fisheries production in 1999–2000 was worth approximately US$118 million (A$235.6 million) (Australian Bureau of Agricultural Resource Economics).

Commercial harvest is by diver using hookah or scuba equipment. Abalone attach to rocky substrates

Table 4 Principal commercial abalone, including main production sources and methods

Species	*Common name*	*Main production sources and method*
Haliotis gigantea	Giant	Japan (F)
Haliotis tuberculata	Tuberculate or Ormer	Channel Islands, France (F/A)
Haliotis midae	Perlemoen	South Africa (F/A)
Haliotis rubra	Black lip	Australia, Solomon Islands (F/A)
Haliotis laevigata	Green lip	Australia (F/A)
Haliotis roei	Roe's	Australia (F)
Haliotis conicopora	Brown lip	Australia (F)
Haliotis iris	Paua	New Zealand (F)
Haliotis rufescens	Red	Pacific North America (F/A)

A, aquaculture; F, fisheries.

with a muscular foot, and they are manually removed with a blunt knife or similar implement. Owing to the slow growth rates and high value, most countries regulate harvesting. Size and catch limits and closed seasons apply in Australia, New Zealand, South Africa and the USA, although with variable management success.

Aquaculture World aquaculture production in 1999 was estimated at 1885 tonnes, valued at US$30.6 million. China dominated with 1799 tonnes (*H. discus hannai* and *H. diversicolor*) and Chile with 48 tonnes (*H. discus hannai*). Variable production since 1990 has originated from South Africa, the USA, Australia, and New Zealand. The principal culture methods include suspended net culture, raceway culture, or tidal pond culture. (*See* **Shellfish**: Aquaculture of Commercially Important Molluscs and Crustaceans.)

Product types Most Australian abalone production is canned, with 2340 tonnes (US$60 million) exported to Hong Kong (40%), Japan (27%), and Taiwan (20%) in 1999–2000. Other product types for fished and cultivated abalone include; parboiled and frozen on shell, frozen on shell, chilled, frozen, or salted meat, and vacuum packing. Live abalone are common as a value-added market product or mail order.

Issues The long growout period of most abalone, up to 5 years, may affect aquaculture viability, although their final value is significant and may offset cash flow issues. Abalone are affected by potentially significant diseases, notably the protozoan, *Perkinsus olseni*, *Rickettsia*-like infections, and the bacterial pathogen *Vibrio*.

Other Gastropod species The following gastropod molluscs are of minor commercial importance in terms of volume, but may be regionally or traditionally important.

Periwinkles Periwinkles of the family Littorinidae are small, intertidal algal grazers, associated with rocky shores. They are a traditional source of food throughout their almost-worldwide distribution, particularly in Europe, where commercial production is greatest.

Production of wild-harvested Littorinids was 4828 tonnes in 1999. The principal species is *Littorina littorea*, with 3144 tonnes, mainly derived from Ireland (3018 tonnes) and Spain (126 tonnes). Other Littorinids, without species designation, contributed 1714 tonnes, 78% from the UK, with the remainder from the USA and Canada. European production is by hand harvesting from intertidal areas and represents a source of income for both local and transient workers. Periwinkles are bagged and sold as a live product, or processed as brined or pickled meat only. There is interest in cultivation, principally in France and Scotland. French aquaculture production of *Littorina* spp. in 1999 was 800 tonnes, worth US$2.08 million.

Conchs The family Strombidae includes algal or detrital feeders, mainly associated with inshore sandy or muddy areas throughout tropical and subtropical seas. Conchs are an important marine resource in the Caribbean, where the Queen conch (*Strombus gigas*) is particularly well known. Conch fishery production was 15 487 tonnes in 1999. The largest producer was Mexico (8591 tonnes), followed by Jamaica (1366 tonnes) and the Dominican Republic (1257 tonnes). Aquaculture of *Strombus* spp. is developing with 9 tonnes, valued at US$61 000, produced by Netherlands Antilles and Turks and Caicos Islands in 1999.

Whelks Whelks are predatory or scavenging molluscs, of boreal, temperate, and tropical seas. The term 'whelk' is associated with several families, with Buccinidae and Melongenidae being commercially important. The common whelk, *Buccinum undatum*, and *Busycon* spp. are the principal species with 17 988 and 6181 tonnes, respectively, in 1999. The *B. undatum* fishery is dominated by France (7885 tonnes), the UK (4925 tonnes), and Ireland (4561 tonnes). The USA and Canada produce all *Busycon* spp., with 4624 and 1557 tonnes, respectively.

Nutritional features of commercially important molluscs Table 5 provides information on the

Table 5 Proximate nutritional composition (per 100 g fresh weight) of principal commercial molluscs (sources: Nettleton (1985), Ulieg (1988))

Species	*Protein (g)*	*Total fat (g)*	*Carbohydrate (g)*	*Energy (kJ)*	*Cholesterol (mg)*	*Fatty acids (saturated/ monounsaturated/ polyunsaturated/ω-3)*	*Moisture (%)*
Abalone *Haliotis iris*	20.8	1.0	0.9	415	111[a]	0.1/0.09/0.12/ 0.04[a]	75.7
Scallop *Pecten novaezelandiae*	15.4	1.3	2.7	349	102[b]	0.1/0.06/0.24/0.18[c]	78.7
Oyster *Crassostrea gigas*	11–13	2.7 (0.8–3.2)	5.8	360	47	0.49/0.36/0.90/0.71	79.7
Mussel *Mytilus edulis*	12.0	2.2 (1.2–2.1)	4.5	372	63	0.41/0.50/0.75/0.43	80.9[d]
Periwinkle *Littorina littorea*	18.2	3.0 (1.2–4.5)	2.3	114		0.45/0.83/0.90/0.44	
Hard-shell clam *Mercenaria mercenaria*	9.2	1.0 (0.2–2.0)	60	11.7	40[e]	0.16/0.16/0.33/0.24	

[a]Data from *Haliotis japonica*.
[b]Data from *Pecten fumatus*.
[c]Data from *Placopecten magellanicus*.
[d]Data from *Perna canaliculus*.
[e]Estimated from cooked product assuming 95% retention.

nutritional composition of edible molluscs. It should be noted that, particularly for molluscs consumed whole, there is considerable variation in content, as gonadal tissue is highly seasonal.

See also: **Marine Foods**: Edible Animals Found in the Sea; **Shellfish**: Aquaculture of Commercially Important Molluscs and Crustaceans

Further Reading

FAO (2001a) *FAO Yearbook, Fishery Statistics: Aquaculture Production 1999. Fisheries Series No. 58*, vol. 88/2. Rome: FAO.

FAO (2001b) *FAO Yearbook, Fishery Statistics: Capture Production 1999. Fisheries Series No. 57*, vol. 88/1. Rome: FAO.

Gosling E (ed.) (1992) *The Mussel Mytilus. Developments in Aquaculture and Fisheries*, vol. 25. Amsterdam Elsevier.

Guo X, Ford SE and Zhang F (1999) Molluscan aquaculture in China. *Journal of Shellfish Research* 18(1): 19–31.

Hallegraeff GM (1993) A review of harmful algal blooms and their apparent global increase. *Phycologia* 32(2): 79–99.

Kraeuter JN and Castagna M (eds) (2001) *Biology of the Hard Clam. Developments in Aquaculture and Fisheries*, vol. 31. Amsterdam: Elsevier.

Nettleton J (1985) *Seafood Nutrition: Facts, Issues and Marketing of Nutrition in Fish and Shellfish*. New York: Osprey Seafood Handbooks.

OIE (Office International des Epizooties) (2001) *International Aquatic Animal Health Code*, 4th edn. Paris: OIE.

Sheperd SA, Tegner MJ and Guzman del Proo SA (1992) *Abalone of the World*. Oxford: Blackwell.

Shumway SE (ed.) (1991) *Scallops: Biology, Ecology and Aquaculture. Developments in Aquaculture and Fisheries Science*, vol. 21. Amsterdam: Elsevier.

Vlieg P (1988) *Proximate Composition of New Zealand Marine Finfish and Shellfish*. Palmerston North, New Zealand: Biotechnology Division, Department of Scientific and Industrial Research.

Contamination and Spoilage of Molluscs and Crustaceans

D R Livingstone, Plymouth Marine Laboratory, Plymouth, Devon, UK
S Gallacher, Fisheries Research Services, Aberdeen, UK

Introduction

Seafood is of worldwide economic and nutritional importance, providing about 16% of the total animal protein consumed globally. Of this, approximately 25% is shellfish, corresponding to an annual 17 million tons taken from sea, coastal, and estuarine waters, with a total value on landing of about £10 billion. Included in the shellfish are crustaceans (e.g., crab, lobster, shrimp), and bivalve (e.g., mussel, oyster, clam, scallop), gastropod (e.g., whelk,

periwinkle), and cephalopod (octopus, squid, cuttlefish) molluscs. The nutritional value of shellfish and other seafood resides in their high-quality protein, essential amino acids, vitamins, and minerals. They are also the only natural dietary source of *n*-3 polyunsaturated fatty acids (or PUFAs), necessary for normal development of human neural, cerebral, and visual functions. The health benefits of this food source are reflected in the lower risk of coronary heart disease, hypertension, and cancer indicated in human populations eating seafood. However, spoilage of shellfish can occur through two major and distinct processes, namely the uptake of chemical pollutants, such as metals and organic compounds, and the accumulation of biological agents, principally bacteria, viruses, and specific toxins produced by phytoplankton (Table 1). In both cases, there are usually no visible signs of contamination. Chemical pollutants may take from days to years to show adverse effects towards human consumers, whereas the impact of biological agents is generally more immediate and obvious, with illness occurring within hours to weeks. (*See* **Fatty Acids**: Properties.)

The section on chemical contaminants covers the variety and levels of chemicals found in tissues from near pristine through to highly polluted environments; the processes of uptake, metabolism, accumulation, elimination, and persistence, including food chain aspects; the toxic threat to consumers; and monitoring and regulatory considerations. The section on biological agents is divided into two. The first covers the types and characteristics of shellfish toxins and the second deals with microbial infectious agents, particularly viruses and bacteria. Some information is given on monitoring procedures and the methods applied to biological contaminants of shellfish. In addition, the effectiveness of depuration and processing on these agents is discussed.

The terms 'contaminant' and 'pollutant' are often used interchangeably and arbitrarily in both the scientific and lay literature in referring to chemicals which are generally regarded as not normal to the environment and potentially harmful. The term 'pollutant' will be used in this article to describe chemicals that are potentially toxic to the consumer that may be synthetic, or natural but present at levels significantly above what is considered to be in the normal biological range. 'Contamination' in this article is taken to describe the presence of pollutants, algal toxins, bacteria, or viruses in the molluscs or crustaceans leading to the condition of spoilage, i.e., rendering them unsuitable for human consumption. The term 'xenobiotic' also appears widely in the scientific literature in relation to the metabolism of pollutants, drugs, and other organic chemicals by organisms, and refers to a chemical that is foreign to an animal, i.e., not part of its normal metabolism.

Pollutants

Types of Pollutants, Routes of Entry, and Tissue Levels

The amount, variety and number of industrial, agricultural, and other chemicals produced by humans are continually increasing. In the early 1990s, the Organization for Economic Cooperation and Development (OECD) estimated that approximately 1000–1500 new substances entered daily use each year, adding to the already existing 60 000 substances. Of these chemicals, it was estimated that about 11 000 were being manufactured in sufficient quantities to merit environmental attention and assessment. Additionally, more pollutants are produced by the process used to treat industrial and domestic waste such as chlorination. The major classes of chemical pollutant are given in Table 1, most of which are synthetic, but also include natural chemicals whose presence in the environment has been increased by human activities, e.g., metals and polycyclic aromatic hydrocarbons (PAHs). A part of

Table 1 Sources of contamination of molluscs and crustaceans

Nature of contamination	*Type of contamination*	*Major sources*
Pollutant	Metals	Industrial
	Organic compounds	Industrial, agricultural, combustion, domestic
	Organometallic compounds	Industrial, agricultural, boating
Shellfish toxins	Paralytic, diarrhetic, amnesic, and neurotoxic shellfish poisons	Phytoplankton
Bacteria	*Campylobacter* spp., *Escherichia coli*	Fecal sources
	Salmonella spp.	
	Vibrio spp.	Natural sources
	Shigella spp., *Staphylococcus aureus*	Food handling
Viruses	Norwalk-like viruses, hepatitis A	Fecal sources

this potentially toxic material enters the aquatic environment where it comprises a diverse range of chemicals of both long-standing and more recent concern. The former includes PAHs, organochlorine pesticides and industrial compounds, and metals (sometimes called heavy metals) such as arsenic (As), cadmium (Cd), copper (Cu), iron (Fe), lead (Pb), mercury (Hg), silver (Ag), and zinc (Zn). Pollutants of more recent concern include polychlorinated dibenzo-*p*-dioxins (PCDDs) and polychlorinated dibenzofurans (PCDFs), organophosphorous insecticides, organotin antifouling agents such as tributyltin (TBT), and estrogenic compounds such as alkylphenol ethoxylate surfactants. Whereas some of the metals, such as Cd and Hg, have as yet no discernible biological function, most are essential to animals and pose a toxic threat only at concentrations considerably in excess of the normal levels. (*See* **Arsenic**: Requirements and Toxicology; **Cadmium**: Toxicology; **Copper**: Properties and Determination; **Iron**: Properties and Determination; **Mercury**: Toxicology; **Zinc**: Properties and Determination.)

Pollutants enter and are dispersed in aquatic ecosystems by various routes, including direct discharge,direct use, land run-off, atmospheric deposition, *in situ* production, abiotic and biotic movement, and food chain transfer. Pollutants are generally readily taken up into the tissues of molluscs, crustaceans, and other resident animals. The uptake into the individual organism can occur via several routes, including from bottom sediments, suspended particulate material, the water column and food sources. The major routes of input depend on the particular dietary and ecological lifestyle of the species, e.g., via algae, particulates, and water column for filter-feeding bivalves; detritus and mixed diet for scavenging crabs; and the food chain for carnivorous cephalopods. All species of molluscs and crustaceans will therefore likely take up pollutants from the environment into their tissues, and any pollutant present in the environment is likely to be present in the tissues of resident organisms. Exceptions to the latter are those pollutants that are not bioavailable to the animal, e.g., those that for physicochemical reasons are tightly bound and 'locked up,' such as PAHs in sediment particles and metals in biological granules.

The extent to which pollutants taken up by shellfish remain in their tissues depends on a number of factors, including continuing input from the environment, duration of exposure, biotransformation and depuration, which are discussed in a later section. Pollutants have been measured in the tissues of many different commercial and noncommercial crustacean and molluscan species, and many different types of pollutants have been detected. The crustaceans include crabs, crayfish, lobsters, prawns, shrimps, copepods, amphipods, and barnacles. The molluscs include clams, cockles, mussels, oysters, scallops, snails, whelks, winkles, squids, chitons, and limpets. Examples of the range of pollutants detected in shellfish tissues are given in Table 2. Also, taken up into the tissues of shellfish are radionuclides associated with the nuclear fuel cycle, such as uranium and the transuranium elements plutonium (Pu) and americium (Am), i.e., ^{239,240}Pu, ^{238}Pu, and ^{241}Am.

The major factor determining the levels of pollutants in the tissues of shellfish is the levels in the immediate environment and their food sources. Thus, the more polluted the environment, the greater the chemical contamination of the tissues. This is particularly evident in pollution incidents, such as oil spills, and in comparisons of urban or industrial

Table 2 Examples of pollutants detected in crustaceans and/or molluscs in the field

Type	*Chemicals*
Metals	Cadmium, chromium, cobalt, copper, iron, lead, mercury, manganese, nickel, silver, zinc
Organic compounds	
Aliphatic hydrocarbons	Acyclic isoprenoids, *n*-alkanes (C_{15}-C_{35}), UCM
PAHs and related compounds	Many two-through to six-ring PAHs, including naphthalene, phenanthrene, anthracene, fluorene, fluoranthene, pyrene, benzo[*a*]anthracene, chrysene, benzo[*b*]fluoranthene, benzo[*k*]fluoranthene, benzo[*a*]pyrene, benzo[*e*]pyrene, dibenz[*ah*]anthracene, benzo[*ghi*]perylene indeno[1,2,3-*cd*]pyrene, dibenzothiophenes, methylated PAHs
PCBs	Many of the possible 209 congeners, including tri-, tetra, penta-, hexa-, hepta-, octa-, nona-, and decachlorobiphenyls
PCDDs, PCDFs	2,3,7,8-Tetrachlorodibenzodioxin, others
Other organochlorine compounds	Chlordanes, chlorinated cyclodienes (e.g., aldrin, dieldrin), DDTs, DDDs, DDEs, hexachlorobenzene, hexachlorocyclohexanes (e.g., lindane), polychlorinated phenols (e.g., pentachlorophenol)
Organometallic Compounds	Tributyltin, triphenyltin, methylmercury

UCM, unresolved complex mixture of alkanes; PAH, polycyclic aromatic hydrocarbon; PCB, polychlorobiphenyl; PCDD, polychlorinated dibenzo-*p*-dioxin; PCDF, polychlorinated dibenzofuran; DDT, 1,1-bis(4-chlorophenyl)-2,2,2-trichloroethane; DDD, 1,1-dichloro-2,2-bis(*p*-chlorophenyl)ethane; DDE, 1,1-dichloro-2,2-bis(*p*-chlorophenyl)ethylene.

locations with remote pristine environments. The pollutants are generally not distributed evenly through the different tissue types, but are often highest in tissues concerned with the processing of food, such as the hepatopancreas and digestive gland, and in the case of organic pollutants are highest in tissues high in fat, such as digestive tissues (hepatopancreas/digestive gland) and ripe reproductive tissues, e.g., mantle of mussels. Some pollutants, such as organophosphorous pesticides, are short-lived and readily broken down in the environment, and therefore their concentrations are often low in the tissues of animals. Examples of levels of pollutants found in different tissues of crustaceans and molluscs from different types of environment are given in Tables 3 and 4 respectively.

Because of the marked accumulation of pollutants by shellfish, the analysis of their tissues has for many years been used in pollution monitoring as a measure of environmental contamination. Mussel and oyster species in particular have been extensively used in so-called 'mussel watch' monitoring programs because of their bioaccumulating properties, sessile nature, wide geographical distribution, and resistance to general stress. Such programs have been carried out on the coasts and inland waters of North and South America, Europe, Asia, and Australia, including the North and South Atlantic, Pacific, Indian, and Mediterranean regions. Common species of marine and fresh-water mussel used include the blue mussel (*Mytilus edulis*), Mediterranean mussel (*Mytilus galloprovincialis*), American mussel (*Mytilus californianus*), horse mussel (*Modiolus modiolus*), swan mussel (*Anodonta cyngea*), and green-lipped mussel (*Perna viridis*). Commonly used oyster species include *Crassostrea virginica* and *Crassostrea gigas*.

Processes of Uptake, Depuration, Biotransformation, Bioaccumulation, and Persistence

Organic pollutants The accumulation of organic pollutants in shellfish (i.e., bioaccumulation) depends principally upon the processes of uptake and depuration, with probably lesser roles for active excretion, 'storage' and biotransformation. The uptake and depuration of pollutants by shellfish depend on a number of physicochemical and biological variables, but mainly on the environmental concentration of the chemical and its fat-solubility. Other environmental and biological factors that may affect uptake and depuration to varying degrees include temperature, tidal cycle, dissolved organic material, algae, nutritional state, feeding habits, and reproductive condition and/or seasonality.

Uptake and depuration are thought to be largely passive processes, involving movement and equilibrium of the chemical between aqueous (external environment) and biotic (animal) compartments. Rates of uptake are usually greater from the water column than from sediments and abiotic particles. Distribution of the pollutant within the animal depends on route of intake and tissue fat levels, e.g., gills are an initial site of bioaccumulation where uptake is from the water, whereas digestive tissues are important for pollutant input from food sources. If the external exposure concentration of the pollutant is constant, then the uptake usually involves an initial linear rate, followed by the eventual reaching of a maximal tissue equilibrium concentration. Uptake of the pollutant into the shellfish will thus continue for as long as pollutant is present in the environment and/or food source, and until the tissue equilibrium concentration is reached. Multiphasic patterns of uptake have also

Table 3 Examples of pollutant levels in different tissues of crustaceans from different environments[a]

Animal	*Chemical*	*Location*	*Tissue*	*Level ($\mu g\, g^{-1}$ wet wt)*
Barnacle (*Balanus amphrite*)	Copper	Hong Kong	Total tissue	12–694
	Zinc	– clean to polluted		545–1871
	Cadmium			1.1–1.5
Crab (*Carcinus maenas*)	Copper	Scotland	Total tissue	6
	Zinc			19
	Cadmium			0.2
	Total PAHs (2–5-ring)	Norwegian fjord	Total tissue	0.03–0.05
Crab (*Uca pugnax*)	Total PAHs	Oil spill	Total tissue	18.3–56
Crab (*Carcinus maenas*)	Total PCBs	Norwegian fjord	Total tissues	0.03–0.10
Crab (*Carcinus mediterraneus*)	Total PCBs	NE Mediterranean Sea	Total tissues	0.3–3.6
Blue crab (*Callinectes sapidus*)	Total PCBs	S. Carolina, USA – industrial gradient	Muscle	0.20–7.73

[a]Dry weight converted to wet weight levels using a conversion factor of $\times$ 0.2.

PAH, polycyclic aromatic hydrocarbon; PCB, polychlorobiphenyl.

Source of date: Livingstone DR (1991) Organic xenobiotic metabolism in marine invertebrates. In: Gilles R (ed.) *Advances in Comparative and Environmental Physiology*, vol. 7, pp. 1–185. Berlin: Springer-Verlag; and Langston WJ and Bebianno MJ (eds) (1998) *Metal Metabolism in Aquatic Environments*. Ecotoxicology Series, vol. 7. London: Chapman and Hall.

Table 4 Examples of pollutant levels in molluscs from different environments[a]

Animal	*Chemical*	*Location*	*Level ($\mu g\,g^{-1}$ wet wt)*
Mussel (*Mytilus edulis*)	Silver	Mediterranean, France, Italy	0.02–0.38
	Cadmium		0.08–1.18
	Cobalt		0.10–1.48
	Chromium		0.10–5.76
	Copper		0.48–30.8
	Iron		29.8–444
	Manganese		0.66–14.0
	Nickel		0.18–2.82
	Lead		0.54–23.4
	Zinc		19.4–129
	Mercury	England	0.02–0.21
Green-lipped mussel (*Perna viridis*)	Cadmium	Hong Kong	0.02–0.29
	Copper		1.7–55.6
	Mercury		0.02–0.03
Mussel (*Mytilus edulis*)	Total PAHs (2–5-ring)	Norwegian fjord – clean to industrial	0.45–3.09
	Total PAHs (2–5-ring)	Norwegian fjord – industrial gradient	0.74–29.9
Mussel (*Perna viridis*)	Total PCBs	Hong Kong coast – industrial gradient	0.01–0.38
Mussel (*Mytilus galloprovincialis*)	Total PCBs	Mediterranean coast	0.02–1.63
Mussel (*Mytilus edulis*)	Total PCDDs	Norwegian waters	0.03–0.30
	Hexachlorohexane	UK waters	2.4–15.2
Mussel (*Mytilus californianus*)	Total DDT, DDE, DDD	S. California, USA – clean to industrial	0.02–1.05
Mussel (*Mytilus edulis*)	Tributyltin	European waters	0.04–2.04
		USA, Pacific waters	0.02–2.2

[a]Values are for whole-body tissues, dry weight converted to wet weight levels using a conversion factor of $\times$ 0.2.
PAH, polycyclic aromatic hydrocarbon; PCB, polychlorobiphenyl. DDT, 1,1-bis(4-chlorophenyl)-2,2,2-trichloroethane; DDE, 1,1-dichloro-2,2-bis (*p*-chlorophenyl)ethylene; DDD, 1,1-dichloro-2,2-bis(*p*-chlorophenyl)ethane. Source of data: Livingstone DR (1991) Organic xenobiotic metabolism in marine invertebrates. In: Gilles R (ed.) *Advances in Comparative and Environmental Physiology*, vol. 7, pp. 1–185. Berlin: Springer-Verlag; Gosling E (ed.) (1992) *The Mussel Mytilus: Ecology, Physiology, Genetics and Culture*. Amsterdam: Elsevier; and Lasserre P and Marzollo A (eds) (2000) *The Venice Lagoon Ecosystem. Inputs and Interactions Between Land and Sea*. Man and the Biophere Series, vol. 25. Paris: Parthenon Publishing.

been seen, which have been interpreted in terms of equilibration of the pollutant into multiple compartments (tissue, cellular, subcellular, or molecular) within the animal. The initial rate of uptake and final tissue equilibrium concentration of the pollutant increase with increasing concentration of the chemical in the environment.

The tendency of organic pollutants to bioaccumulate increases with increasing fat-solubility of the chemical and with increasing levels of fats (lipids) in the shellfish tissue. For many organic pollutants taken up by molluscs and crustaceans, a linear relationship exists between the equilibrium bioconcentration factor (BCF; ratio of pollutant concentration in the animal to that in the environment) and fat-solubility, i.e., chemical hydrophobicity, the latter being measured in terms of either octanol/water coefficient (K_{ow}) or water-solubility. Positive linearity of BCF with K_{ow} in shellfish is seen over a log K_{ow} range of about 2–6 which encompasses most of the organic pollutants of environmental importance, i.e., aliphatic hydrocarbons, many polycyclic aromatic hydrocarbons (PAHs) and polychlorobiphenyls (PCBs), dioxins, and other organochlorine compounds. Thus, pollutants of high hydrophobicity, such as PCBs and 1,1-bis(4-chlorophenyl)-2,2,2-trichloroethanes (DDTs), show high BCF values of up to $\times$ 100 000. Deviations from these relationships also occur, particularly for very hydrophobic compounds ($K_{ow} > 6$), such as the five-ringed PAH benzo[a]pyrene and high hexachloropolybiphenyls, which show BCFs less than expected. In other cases, higher BCFs than predicted from K_{ow} have been seen, e.g., > 100-fold higher bioaccumulation of the antifouling paint TBT by mussel (*Mytilus edulis*). These enhanced BCF are probably due to strong interactions between pollutant and the biological system, particularly proteins. The time required for the uptake of the pollutant to reach the tissue equilibrium concentration is relatively quick for water-soluble compounds, e.g., hours for toluene ($\log K_{ow} < 3$), but may take weeks or even months for very fat-soluble compounds, e.g., hexachlorobiphenyl ($\log K_{ow} > 7$).

Partial or complete decrease in pollutant levels in the environment, either by cessation of input, or transfer of the shellfish to clean water, will result in depuration, i.e., loss of pollutant from the tissues back into the water. This depuration will continue

as long as the environmental pollutant level remains very low by dilution into a large body of water, replacement with fresh clean water, or removal of the pollutant. Depuration is normally exponential, with the pattern of loss being resolved into an initial faster phase followed by a longer-term slower phase. The rate of loss during the phases can be described in terms of the depuration half-life ($t_{1/2}$), i.e., the time required for the tissue concentration of contaminant to be reduced by 50%. Single-phase depuration has also been seen in some instances. Depuration can also be markedly affected by the length of time of the exposure period. Short-term exposure results in rapid and complete, or almost complete, elimination of the pollutant during subsequent depuration, whereas longer-term exposure is followed by slower and often incomplete elimination. This is particularly evident for the depuration of petroleum hydrocarbons by bivalve molluscs where the $t_{1/2}$ of depuration is seen to increase with increasing time period of previous exposure (Table 5). Therefore it may be very difficult, if not impossible, to remove all organic pollutants by depuration in clean water from shellfish that have had a long history of exposure to pollution in the field. For pollutants of similar fat-solubilities, the chemical structure of the chemical may also affect uptake and depuration, e.g., PAHs were bioaccumulated less from sediments than were PCBs of similar log K_{ow} by mussel (*M. edulis*), but PCBs were depurated more slowly than equivalent PAHs. (*See* **Polycyclic Aromatic Hydrocarbons.**)

The bioaccumulation of organic pollutants in shellfish can also be affected by active excretion processes, 'storage,' and biotransformation. Although knowledge of these processes is relatively limited, their quantitative impact is indicated not to be great, at least in the long term, because the bioaccumulation of most types of organic pollutant occurs according to a simple water/octanol partitioning model, where water is the environment and octanol represents the animal, i.e., BCF is directly related to K_{ow}. Active excretion has been indicated in some studies; e.g., naphthalene lost from gills and kidney of mussel (*M. edulis*), but not others; e.g., the anus and kidney played little or no role in the clearance of naphthalene from crab (*Hemigrapsus nudus*). Active energy-driven pumps that function to pump xenobiotics out of cells, variously known as *P*-glycoproteins, multidrug resistance (MDR) transporters or multixenobiotc (MXR) transporters, have recently been identified in a wide range of aquatic organisms, including molluscs and crustaceans. However, relatively little is yet known about their ability to pump out organic pollutants, although hydrocarbons, pentachlorophenol, DDTs, and PCBs have been indicated as potential substrates for some shellfish species, but not others.

Fecal material and urine can also make a contribution to elimination, e.g., in the loss of sulfadimethoxine from lobster (*Homarus americanus*). The term 'storage' is probably a misnomer, but organic pollutants are known to accumulate in specific subcellular organelles, namely lysosomes that are present in high

Table 5 Increase in duration of exposure period results in increased length of time for subsequent depuration of a range of hydrocarbons by various bivalve shellfish

Shellfish	*Pollutant*	*Exposure period*	*Depuration half-life*
Mussel (*Mytilus edulis*)	Naphthalene	4 h	2 h
	Heptadecane	24 h	< 24 h
	n-alkanes	2 days	0.5 days
	Naphthalenes	2.8 days	0.9 days
	Alkanes	41 days	3.1 days
	4–6-ring PAHs	40 days	12–30 days
	Benzo[a]pyrene	Field	16 days
	Alkanes/PAHs	Field	70 days
Clam (*Rangia cuneata*)	*n*-alkanes	8 h	> 3 h
	2–3-ring PAHs	8 h	3 h
	Anthracene	15 h	5 days
	Benzo[a]pyrene	1 day	8 days
Oyster (*Crassostrea virginica*)	*n*-alkanes	8 h	4.5 h
	2–3-ring PAHs	8 h	6 h
Horse mussel (*Modiolus modiolus*)	Phenanthrene	2 days	1.5 days
Clam (*Macoma balthica*)	Alkanes/PAHs	170 days	> 60 days
Clam	Alkanes/PAHs	Field	> 120 days

PAHs, polycyclic aromatic hydrocarbons.
Data source: Livingstone DR (1991) Organic xenobiotic metabolism in marine invertebrates. In: Gilles R (ed.) *Advances in Comparative and Environmental Physiology*, vol. 7, pp. 1–185. Berlin: Springer-Verlag.

numbers in tissues such as the hepatopancreas and digestive gland. They can also become chemically bound to macromolecules, including in particular proteins and DNA, to form macromolecular adducts. Such adducts are likely eventually removed and the macromolecules repaired or replaced, but nevertheless will contribute to the persistence of a pollutant in an animal. Examples of pollutants indicated, or demonstrated, to give rise to adducts in shellfish include chlorinated paraffins, vinyl chloride, PAHs, nitroaromatics, fenitrothion (organophosphate pesticide), picric acid, and phthalate ester plasticizers. In common with all other animals, shellfish also possess a suite of biotransformation enzymes, present in several tissues but highest in the hepatopancreas or digestive gland, the major function of which is to convert fat-soluble organic xenobiotics such as pollutants to water-soluble excretable products (metabolites). Modeling of biotransformation in shellfish shows that: (1) biotransformation is significantly slower than uptake, such that bioaccumulation of the original pollutant inevitably occurs; and (2) generally crustaceans have slightly higher biotransformation rates than molluscs for both hydrocarbons and other types of organic pollutants. Paradoxically, biotransformation, which is meant to speed up elimination, can contribute to persistence if the rate of loss of metabolite to clean water is slower than the rate of depuration of the original unchanged pollutant. This has been seen for mussel (*M. edulis*) for 1-naphthol and for several crustaceans for phenoxyacetic and phenylacetic acid herbicides, hydrocarbons, and sulfadimethoxine.

Because pollutant uptake from the environment is largely a passive process determined by physicochemical principles, the rates of uptake are therefore somewhat similar for different animal groups, including shellfish and vertebrates such as fish. In contrast, rates of biotransformation of pollutants and excretion of the water-soluble metabolites are intrinsic to the particular animal or group of animals, and depend on the level of biotransformation enzymes in the organism and the effectiveness of its excretory mechanisms. These are generally significantly greater in vertebrates such as finfish than in invertebrates such as shellfish. Rates of biotransformation approach more rates of uptake in finfish than shellfish, and can therefore reduce the long-term bioaccumulation of readily metabolizable pollutants in the former compared to the latter. In food chains, the differences between uptake and biotransformation therefore result in different patterns of bioaccumulation in different animal groups, with readily metabolized pollutants, such as PAHs, reaching highest tissue concentrations near the bottom of food chains in shellfish, and poorly metabolized or recalcitrant pollutants, such as many PCBs, accumulating to highest tissue concentrations in vertebrates at the top of food chains.

Metals and radionuclides The uptake and depuration of metals have some, but not all, similarities with the situation for organic pollutants. Bioaccumulation depends on chemical speciation of the metal, routes and mechanisms of uptake, intracellular compartmentation, cellular control of metal levels, excretion, and depuration. All of these processes may be strongly affected by the species' particular biology, including morphology, physiology, feeding mode, and environmental adaptations. Overall, there is probably more variability in bioaccumulation trends with respect to species and individual pollutant for metals compared to organic compounds, particularly for crustaceans. For example, below a certain threshold metal tissue level, a range of bioaccumulating properties for essential metals such as Zn and Cu may be seen in crustaceans, ranging from regulation (i.e., no accumulation where rate of uptake equals rate of excretion) to accumulation, e.g., at an exposure concentration of 100 p.p.b., Cu was bioaccumulated from sea water by the barnacle *Elminius modestus* but not by the decapod *Palaemon elegans*.

Direct uptake of metals and radionuclides from the water column is a quantitatively important process because most metals occur in highly water-soluble forms, e.g., Am, Cd, Hg, Pb, Pu, and vanadium. Rates of uptake are generally proportional to the metal concentration in the water column. Uptake is thought to occur mainly by passive transport processes, involving both diffusion of polar metal ions/complexes and partitioning of hydrophobic metal organic complexes into lipid membranes. The influence of dissolved organic matter, particulates, and food on uptake is complex. Depending on the nature of these substances and the metal, either increased or decreased bioaccumulation can occur, e.g., in one study, food contributed significantly to the uptake of Am, Pu, and Pb, but not Cd, by mussel (*M. edulis*). The gills are an important site for soluble metal uptake, whereas digestive gland is a major site for particulate-bound metal uptake. The latter occurs via an active transport process, endocytosis, which results in the fusion of the endocytotic vesicles, containing the metal, with the lysosomes. The kidney is another major site of metal accumulation, and metals can also be incorporated into the shells of, for example, bivalves. Major sites of localization of radionuclides in bivalves include byssal threads and the periostracum covering the shell. The metals are retained within the tissues by binding to various

specific and nonspecific sites. The former includes specialized proteins called metallothioneins, specialized sulfur- or phosphate-containing insoluble inorganic granules, and lysosomal residual bodies, all of which have high affinities for metals. The inorganic granules in particular can result in very high levels of metals in certain shellfish, e.g., metal/phosphate granules in scallops, winkles, and crabs.

Exposure of the shellfish to clean water will result in depuration and loss of part or all of the previously accumulated metal. Like the situation for organic pollutants, the depuration is biphasic with a fast initial loss followed a slower subsequent loss, describable in terms of $t_{1/2}$ values. Overall $t_{1/2}$ values can vary from days to several months. The observations on depuration profiles for both metals and organic pollutants indicate at least a two-compartment model (often called the 'fast' and slow' compartments) of pollutant uptake and release. The slow compartment likely includes fat stores, lysosomes, and macromolecular adducts for organic pollutants, and metallothioneins, lysosomes, and granules for metals. The phenomenon of induction, i.e., increased synthesis of metallothionein, in response to metal exposure could also contribute to the slow compartment. Active excretion of metals in molluscs can occur through elimination of residual bodies by the process of exocytosis. Such processes appear to be more important for some metals than others, resulting in very different $t_{1/2}$ for certain metals and species, e.g., in mussels (*Mytilus* spp.), the $t_{1/2}$ for Cu was 9 days compared to 7 months for Cd. Metal that is nonbioavailable in metal/phosphate granules can also be eliminated when prey are eaten by predators, e.g., the metal/phosphate granules contained in barnacles eaten by whelks are lost from the latter in its feces. The consequence of such nonbioavailability is that generally metals, like readily metabolizable organic pollutants such as PAHs, are present in highest tissue concentrations at the bottom of food chains, i.e., in shellfish, and do not bioaccumulate to higher concentrations in top predators, such as fish and other vertebrates. The exception is Hg that is bioaccumulated along food chains, but principally in the form of fat-soluble methylmercury.

Toxic Threat and Regulatory Standards

The vast range and diversity of pollutants that are produced by human activities permeate to varying degrees every part of the earth's ecosystem, from urban and industrial centers through to remote areas such as deep oceans and the Arctic and Antarctic circles. For example, polychlorinated aromatic compounds such as PCBs, PCDDs, and PCDFs appear ubiquitous, having been found in air, soot particles, sediments, water column, and all types of animal tissue from shellfish to fish and other vertebrate wildlife, to human fat, milk, and blood serum. Therefore, unlike the biological agents described later, and excepting localized pollution incidents such as the Minamata disease tragedy, the main potential health threat to humans of pollutants in shellfish is probably their contribution to the general body-burden of such chemicals. The latter can receive a wide range of inputs of pollutants from various sources, including other foodstuffs, cooking procedures, smoking, and atmospheric pollution. Depending on the fat content and which part is eaten, shellfish may make a particular contribution in the case of very fat-soluble organic pollutants, such as PAHs, PCBs, dioxins, and other chlorinated hydrocarbons. Because there are so many pollutants that may be present in shellfish, the potential deleterious effects that they may cause in the short, medium, and long term are many and diverse and encompass most known mechanisms of chemical-mediated toxicity. Examples of some of these are given in Table 6. The extent to which such illness and effects will be manifested or not will depend on the dosage of pollutant to which the consumer is exposed, of which the shellfish may contribute only a part. Some routes of exposure may be much more important for some pollutants than others, e.g., atmospheric pollution can be a major route for Pb.

Localized poisoning incidents involving shellfish have happened in the past, but largely due to major ignorance at the time. In the infamous Minamata disease incident in the 1950s, industrial waste that included mercuric chloride was discharged into Minamata Bay in Japan. The water-soluble Hg was methylated by bacteria to fat-soluble methylmercury that was then bioaccumulated and passed along the food chain. The combination of the high toxicity of Hg, and the seafood forming a major part of the diet, resulted in illness appearing within a year and eventually over 70 deaths and 700 cases of poisoning in the local population.

Chemical analysis of processed commercial seafood in the 1970s through to the 1990s revealed highest levels of PAHs in molluscs, lower levels in crustaceans, and lowest in fish, consistent with their general capacity for biotransformation. In one study, the carcinogenic PAH benzo[a]pyrene was detectable in commercial molluscan seafood from nine countries. Levels of benzo[a]pyrene were 0–36 $ng\,g^{-1}$ wet weight in molluscs compared to 0–8 $ng\,g^{-1}$ wet weight in crustaceans, and were higher in smoked seafood and in animals collected or kept near obvious environmental sources of PAHs, e.g., lobsters held in a commercial tidal pound constructed of creosoted timber had levels of benzo[a]pyrene of 2300 and

Table 6 Examples of possible deleterious effects and mechanisms of toxicity of pollutants found in shellfish to human consumers[a]

Pollutant	*Toxic effect*	*Biochemical effects*
PCBs, PCDDs, PCDFs	Cancer, developmental effects, immune dysfunction, liver damage, reproductive effects, thymic atrophy, weight loss	Induction of CYP1A, UDPGT
Certain PAHs, e.g., benzo[a]pyrene	Cancer	Induction of CYP1A, DNA adduct formation
Aromatic amines	Bladder cancer	
Organophosphorous pesticides	Neuromuscular and neurotoxic effects	Inhibition of acetyl cholinesterase
Cadmium	Kidney damage, bone damage	Induction of metallothionein
Lead	Neurotoxic effects, bone and kidney damage	Interference of heme and porphyrin synthesis
Mercury and methyl mercury	Neurotoxic effects, memory loss, loss of muscle coordination, cerebral palsy	Binds to protein sulfhydryl groups

[a]Data are for humans and/or mammalian studies.
PCB, polychlorobiphenyl; PCDD, polychlorinated dibenzo-*p*-dioxin; PCDF, polychlorinated dibenzofuran; CYP1A1, cytochrome P4501A1; UDPGT, UDP-glucuronyl transferase; PAH, polycyclic aromatic hydrocarbon.

$281\,ng\,g^{-1}$ wet weight in the hepatopancreas and edible tail meat respectively.

Various national and international bodies issue regulations and/or advise on matters relating to the health quality, production, and marketing of fishery products, including shellfish, e.g., the European Union (EU) issues directives which pertain both to member countries and to exports received within the EU. In the UK, environmental quality standards (EQS) exist for some pollutants, whereas guidelines only are identified for others. Examples of these for shellfish from UK waters are given in Table 7. In the USA, the Food and Drug Administration (FDA) issues action levels and guidelines for shellfish; guidelines are also issued by the US Environmental Protection Agency (EPA). Depending on the country, contravention of statutory EQS, action levels, or guidelines can form the basis for closure of a shellfish fishery and/or removal of products from outlets. Guidelines may also be used in conjunction with risk assessment to provide health advice on minimizing chronic effects of pollutants such as PCBs, dioxins, and Hg through limiting shellfish consumption. This advice may be targeted at particularly vulnerable groups such as pregnant and nursing women, children and women of child-bearing age. On the other hand, in the case of certain metals, higher levels may be permitted in some shellfish species if environmental or tissue levels are considered to be high naturally, e.g., Cu and Zn. Chemical analytical techniques exist for many thousands of pollutants, but routine measurement is generally limited to a select few which are of particular concern (Table 7). The bodies responsible for the measurement, and the conditions of measurement (e.g., sampling frequency), vary with the national or international organization. The depuration procedures required for treatment of shellfish prior to marketing (i.e., to remove bacteria and other biological agents) will also remove most or many of the pollutants.

Table 7 Outline of standards/guidelines for maximum acceptable pollutant levels in shellfish from UK waters[a]

Pollutant	*England and Wales*[b] *(mg kg^{-1} wet wt)*	*Scotland and Northern Ireland*[c] *(mg kg^{-1} wet wt)*
PCBs	Molluscs 0.02–0.1 Crustaceans 0.01–0.05	Mussel 0.3–1.0
DDT	Fish liver 0.5	Mussel 0.03–0.1
Dieldrin	Fish liver 0.2–0.3	Mussel 0.015–0.05
Hexachlorobenzene	Fish liver 0.1	Mussel 0.03–0.1
Mercury	Molluscs 0.12–0.2 Crustaceans 0.1–0.3	Mussel 0.2–0.6
Cadmium	Mussel 0.4–1.0	Mussel 1.0–3.0
Copper	Some molluscs 5.0 Crustaceans 20–30	Mussel 3.0–6.0
Lead	Some molluscs 4.0 Crustaceans 1.0	Mussel 3.0–10
Zinc	Some molluscs 100 Crustaceans 100	Mussel 50–100

[a]For full details see the specified reports.
[b]Data from CEFAS Science Series Aquatic Monitoring Report no. 47, Annex 2. (www.cefas.co.uk/homepage.htm)
[c]Data from Fisheries Research Services, Aberdeen: Henderson A and Davies IM (2001) Review of the regulation and monitoring of aquaculture in Scotland, with emphasis on environment and consumer protection. FRS Marine Laboratory Report 01/01.
Dry wt converted to wet wt using a conversion factor of × 0.2.
PCBs, polychlorobiphenyls; DDT, 1,1, bis(4-chlorophenyl)-2,2, 2-trichloroethane.

Shellfish Toxins

Shellfish, particularly marine bivalve molluscs, can be contaminated with naturally occurring poisons, also known as toxins, phycotoxins, or marine biotoxins. It is common to refer to the toxins under generic headings (Tables 1 and 8), with each grouping

consisting of a number of compounds. Consumption of shellfish containing the toxins may lead to illness and/or death in humans. Details of the toxin groups, their origin, and associated toxic syndrome are summarized in Table 8.

The source of toxins in bivalve molluscs is microscopic planktonic algae, termed phytoplankton, which are concentrated in bivalves by their filter feeding mechanism and which form part of the animal's natural diet. Other types of shellfish (e.g., crabs, lobsters), finfish, marine mammals, and seabirds may also be contaminated with the toxins, generally through the food chain. With the exception of shellfish, the other animals have not been associated with human food poisoning, and may themselves become ill or die due to the toxins. The toxins may also cause physiological changes in some shellfish. However, generally the changes are not obvious and contaminated shellfish cannot be differentiated from their nontoxic counterparts by their physical appearance.

Table 8 Details of shellfish toxins, their origin, and toxic syndromes

		Toxin group		
	Paralytic shellfish toxins	*Amnesic shellfish toxins*	*Diarrhetic shellfish toxins*	*Neurotoxic shellfish toxins*
Principal toxin (molecular weight)	Saxitoxin (carbamate toxin); (299.3)	Domoic acid (311.14)	Okadaic acid (OA) (804.5)	Brevetoxin PbTx 1 (867.07)
Associated toxins	Carbamate toxin (neo-STX); gonyatoxins (GTX 1–4); *N*-sulfocarbamoyl toxins (B1, B2, C1–C4); decarbamoyl toxins (dc-GTX 1–4, dc-Neo, dc-STX, do-STX, doGTX2, doGTX3)	C5' diastereomer; Isodomoic acid, mainly D, E, and F in shellfish	OA compounds: dinophysistoxins (DTXs), OA; OA-diol esters; others: YTXs; PTXs; AZTs	PbTx 1–10
Chemical type	Water-soluble tertrahydropurine	Water-soluble tricarboxylic acid	Polyethers	Polyethers
Origin in relation to shellfish	*Alexandrium* spp. *Gymnodinium catenatum* *Pyrodinium bahamense*	*Pseudo-nitzchia* spp.	OA compounds: *Prorocentrum* spp. and *Dinophysis* spp. PTXs, *Dinophysis fortii* YTXs, *Protoceratium reticulatum* AZTs, unknown	*Gymnodinium breve*
Type of phytoplankton	Dinoflagellates	Diatoms	Dinoflagellates; AZPs unknown	Dinoflagellates
Food-poisoning syndrome	Paralytic shellfish poisoning	Amnesic shellfish poisoning (ASP)	Diarrhetic shellfish poisoning (DSP)	Neurotoxic shellfish poisoning (NSP)
Mode of action	Block influx of sodium ions in neuronal and muscular voltage gated sodium channels	Glutamate agonist, binds to kainate-type glutamate receptors	OA/DTXs inhibit serine/threonine protein phosphatases; YTX hepatic and cardiac damage; PTXs hepatotoxic; AZTs necrosis of pancreas cells, liver, lymphocytes, and myocardium	Influx of sodium ions into voltage gated sodium channels
Examples of symptoms	Tingling of lips and tongue, numbness, mild headache, incoherent speech, nausea, vomiting, respiratory difficulties, paralysis, death	Nausea, vomiting, diarrhea, headache, memory loss, seizures, coma, death	Diarrhea, nausea, vomiting, abdominal pain and chills (OA, DTXs and AZTs); PTX and YTX not defined	Abdominal pain, nausea, diarrhea, chills, reversal hot/cold, headache, muscle/joint pain
Time to onset	30 min–4 h	24–48 h	30 min–12 h (OA + DTXs)	3–6 h
Life-threatening	Yes	Yes	No	No
Concentration reported to cause symptoms	144–12 400 μg 100 g^{-1}	60–295 mg	OA 25–36 μg 100 g^{-1}; AZP not defined	1600 μg 100 g^{-1}
Regulatory limit	80 μg 100 g^{-1}	20 μg g^{-1}	Not detected by mouse bioassay	20 MU 100 g^{-1}

STX, saxitoxin; YTXs, yessotoxins; PTX, pectenotoxins; AZT, azaspiracid toxins; PbTx, brevetoxin; MU, mouse units.

Approximately 40 of the estimated 5000 phytoplankton species produce toxins. These occasionally occur in large numbers causing discoloration of the water column that has led to descriptions such as 'red' and 'brown tides.' However, large numbers of algal cells need not be present for toxins to accumulate in shellfish. The toxic algae and hence shellfish containing toxins occur worldwide, and it has been suggested that they are increasing in distribution, frequency, and intensity. Reasons postulated include climate change, increased utilization of coastal waters, and transport of toxic species via ships' ballast water. The toxins are generally described as secondary metabolites and their functional significance to the algal cell is unknown. Inter- and intraspecies differences in the quantity of toxin produced and the toxin profile exist, as do nontoxic variants of toxic species. The pathways for toxin biosynthesis have not been fully elucidated for most of the toxin groups, although some steps have been characterized. The genes involved are currently not identified.

The occurrence of specific algal species in some areas, in some seasons, is fairly predictable (e.g., *Alexandrium* species, associated with paralytic shellfish toxins (PSTs), tends to occur in the summer months in parts of the UK). Although, some field data are available on environmental factors that trigger the emergence of these toxic phytoplankton and production of their toxins, it is currently not possible to predict accurately the number of toxic cells and the length of time for which these cells will remain in the water column, nor their cellular toxin concentration. All of these factors influence the total toxin burden in exposed shellfish. Other factors which may influence shellfish toxicity are rate of ingestion, degradation to nontoxic compounds, biotransformation from one toxin in a group to a less toxic compound in the same group, and excretion.

Toxin Groups

Paralytic shellfish toxins PSTs are a group of approximately 20 closely related tetrahydropurine compounds of varying potency (Table 8). The dinoflagellates involved in production of these toxins are listed in Table 8. Saxitoxin (STX) was the first of these compounds to be identified and is also considered the most potent. Some steps in the biosynthesis of STX have been determined. These involve a high-energy methyl group donor (*S*-adenosylmethionine), a two-carbon unit (acetate), and an amino acid precursor (arginine). However, the sequence of reactions and exact number of biosynthetic steps have not been fully defined. Additionally, controversy surrounds suggestions that the dinoflagellate's bacterial flora are involved in the production/biotransformation of PSTs.

Shellfish contaminated with PSTs can contain different combinations of the compounds that can differ from that of the dinoflagellate ingested. This is more pronounced in some shellfish species (e.g., *Spisula solidissima*) than others (e.g., *M. edulis*). PSTs are also produced by some species of fresh-water cyanobacteria, e.g., *Aphanizomenon flos-aquae*, although to date there are no reports of human food poisoning from fresh-water shellfish due to these toxins. The mode of action of PSTs is listed on Table 8, along with details of the toxic syndrome, i.e., paralytic shellfish poisoning (PSP) that they cause. PSP has been known for centuries and the concentration of PSTs reported to cause illness varies (Table 8). Out of all the shellfish toxins, some consider PSTs to be the most dangerous. This is due to the high concentrations of PSTs that may occur in shellfish (Table 9) and the fact that death can result within an hour.

Amnesic shellfish toxins Details of the amnesic shellfish toxins (ASTs) are summarized in Table 8. In this group, domoic acid (DA) is considered the

Table 9 Example of peak toxin concentration and depuration rates for some shellfish species contaminated with paralytic shellfish toxins

Species	*Peak toxicity (μg STX equiv. 100 g^{-1})*	*Time to regulatory limit (weeks)*	*Detoxification rate (% per day)*
Crassostrea gigas	209–710	0.6–2.0	NA
Ostrea edulis	1000	6.4	4.0
Mytilus edulis	100–19259	0.6–15.6	4.8–15.4
Mytilus californianus	240–5300	2.8–9.0	5.1–8.9
Pecten maximus	2700	6.4	7.4
Placopecten magellanicus	809–6179[a]	> 52	0.2–0.6
Patinopecten yessoensis	6000–340000[b]	> 12–30	1.2–11.7

Regulatory limit, 80 μg saxitoxin (STX) equiv. 100 g^{-1}; NA, not available.
[a]Includes data from different organs.
[b]Data from digestive glands.
Modified from Bricelj MV and Shumway SE (1998) Paralytic shellfish toxins in bivalve molluscs: occurrence, transfer kinetics and biotransformation. *Reviews in Fisheries Science* 6: 315–383.

main compound of concern. Little detailed toxicological information exists on the other ASTs, although it has been suggested that the isomers are less toxic than DA, and the diastereomer is of equivalent toxicity. DA has been detected in a range of marine macroalgae and in some species of the diatom genus *Pseudo-nitzschia*. Biosynthesis of DA is thought to involve two different precursor units. One is formed from the direct incorporation of acetate into oxaloacetate during the Krebs cycle, forming an activated derivative, e.g., 3-hydroxyglutamic acid. The other unit is thought to be an isoprenoid unit, geranyl-pyrophosphate.

Accumulation of DA-producing *Pseudo-nitzschia multiseries* in *M. edulis* led to the first recorded incidence of amnesic shellfish poisoning (ASP; Table 8) in shellfish consumers in Canada in 1987. Gastroenteritis was the milder form of the illness and tended to occur in those under 40 years of age. Severe neurological symptoms, such as short-term memory loss, were experienced by older people, reportedly lasting in some cases for a number of years. Those who died suffered from other preexisting illnesses, e.g., diabetes. Although DA has subsequently been detected in shellfish in countries other than Canada, no further confirmed outbreaks of ASP have been reported.

Diarrhetic shellfish toxins Diarrhetic shellfish toxins (DSTs) were originally defined as okadaic acid (OA) and related toxins called dinophysis toxins (DTXs; Table 8) They were first suspected of causing shellfish-related illness (Table 8) in the Netherlands in 1961, and confirmed as being responsible for food poisoning in Japan in the 1970s. The principal toxin of this group, OA, has also been detected in the sponges *Halichondria okadai* and *H. melanodocia*. The method of detecting DSTs is a mouse bioassay using a solvent extract. The nonspecific nature of this technique has led to the inclusion of other compounds in the DST group. These include fused polyether compounds named yessotoxins (YTXs) and a group of macrocyclic polyether lactones called pectenotoxins (PTXs). Some effects of these compounds in rodents are summarized in Table 8. However, little detailed toxicological and epidemiological information is available and international debate is ongoing on their inclusion in the DST group, their threat to human health, and values, if any, for regulatory limits. The biosynthesis of OA and the related YTXs and PTXs is thought to occur by the successive additions of acetate units to a growing polyketide chain, mediated by the enzyme polyketide synthase (PKS). Modification of PKS, involving for example oxidative modification, methylation, and cyclization, results in different structural types of the enzyme.

A further group of toxins, named azaspiracid toxins (AZTs), is also currently included within the DSTs. These are lipophilic, highly oxygenated polyethers containing an unusual azaspiro-ring structure. Although dinoflagellates are suspected as the source of the toxin, undisputed evidence does not yet exist. AZTs were first reported as causing food poisoning in the 1990s with symptoms similar to those for OA (Table 8). Effects of AZTs on rodents are summarized in Table 8. There is strong support to separate this group from the other DSTs and to name the illness azaspiracid shellfish poisoning. Discussions are currently ongoing on what regulatory limits should be set, given the limited extent of epidemiological data available.

Neurotoxic shellfish toxins Neurotoxic shellfish toxins (NSTs; Table 8) cause shellfish-related food poisoning. However, they are more noted for finfish and seabird kills than human illness. Produced by the dinoflagellate *Gymnodinium breve*, they are infamously known as the cause of the Florida red tides that have been recorded for centuries. *G. breve* is typically found in Florida waters, but similar forms have been detected in waters elsewhere, particularly New Zealand.

Monitoring and Methodology for Shellfish Toxins

A large number of countries worldwide monitor their waters for toxic algae, the shellfish for toxins, or a combination of both. Internationally accepted standards are recognized for PSTs and ASTs (Table 8). Regulations for DSTs are less clear, with some countries using detection by mouse bioassay as criteria for implementing harvesting closures and others suggesting a limit of 8–16 μg $100\,g^{-1}$ (Table 8). Within the EU, limits are incorporated into directives with which member states and countries exporting to the EU must comply.

Some suggest that these regulatory limits can be deemed effective as judged by the lack of shellfish toxin-related food poisoning in countries with comprehensive monitoring programs. However, detailed toxicological and epidemiological data supporting these limits are lacking for many of the toxins, hindered in some cases by lack of toxin standards and appropriate methodology for toxin detection. Additionally, questions relating to sampling procedures for collection of an appropriate representative sample from shellfish populations remain, e.g., sample size, frequency, definition of sampling area, and number of samples per area.

Most countries use a mouse bioassay for the detection of PSTs, DSTs, and NSTs in shellfish-monitoring programs. Ethical considerations are the driving force

behind the development of alternative methods, such as immunoassays, cell assays, and analytical instrumental techniques, e.g., liquid chromatography–mass spectrometry. In some cases progress is hampered by the complexity of the toxins and the lack of toxin standards, and to date, with one exception, no alternative method has been fully accepted for use in routine monitoring. The exception is ASTs where a high-performance liquid chromatographic technique was developed at the time of the Canadian ASP outbreak and is now used routinely in many countries.

Depuration and Processing of Shellfish Contaminated with Toxins

Once shellfish are contaminated with toxins, options for detoxifying or depurating the animals are limited. They mainly involve commercial canning of some shellfish species e.g., *Mya arenaria* and *Mytilus edulis* contaminated with PSTs. However, the effectiveness of these processes is dependent upon initial low levels of toxin and caution is recommended in their use.

Currently, commercial depuration systems are ineffective at cleansing shellfish of toxins, which leaves the option of allowing shellfish to depurate naturally in the environment. This can be problematic as rates of detoxification vary considerably between species, with some shellfish remaining toxic for a year or more (Table 9). Depuration in some shellfish species seems to be biphasic. It has been suggested that the initial detoxification phase represents gut evacuation of unassimilated toxin, whereas the second phase represents the release of toxins assimilated and incorporated in the tissue. The situation is made more complex by changes in toxin profile observed during depuration, particularly for PSTs. A role for biotransformation enzymes in this process has been suggested and recent work has demonstrated the induction of glutathione *S*-transferase in *M. edulis* exposed to PSTs. Nevertheless, mechanisms involved in the depuration of shellfish toxins are poorly understood and hinder the development of commercial depuration processes.

Infectious Microbes

A number of bacteria and human viruses responsible for food poisoning are transmitted via the fecal–oral route (Table 1). Infected individuals can shed large numbers of these microbes in their feces. The link between shellfish-transmitted disease and sewage pollution in the marine environment has been known since the late nineteenth century, with early documented evidence associated with the bacterial infection typhoid fever (Table 10). Of the different shellfish species, bivalve molluscs are the main source of microbial contamination that poses a health risk to human consumers. This is exacerbated by traditional consumption of some species raw or only lightly cooked. As with shellfish toxins, there are no obvious visible indications of contamination.

Ineffectively processed or untreated sewage, faulty septic tanks, and feces from farm animals and seagulls are a principal source of bacterial food-poisoning agents in the marine environment. However, human sewage is currently thought to be the only source of human enteric viruses in marine waters. Sewage treatment processes are only partially effective at removing viruses, and therefore coastal discharges are a major source of these infectious agents in some areas. Additionally, contamination of the marine environment by both bacteria and viruses is exacerbated after heavy rainfall due to increased land run-off and storm water discharges. However, not all infectious agents concentrated in shellfish are from fecal sources. Naturally occurring bacteria from the genus *Vibrio* are also responsible for shellfish-transmitted infection, particularly from raw oysters.

In recent years there have been fewer incidences of infection due to shellfish consumption and this is generally attributed to the success of the controls now implemented in many countries for reducing bacterial contamination of shellfish. When shellfish-related food poisoning does occur, evidence from the UK and USA suggests that the causative agent is not identified in the majority of cases. However, enteric viruses are often suspected because the clinical symptoms are frequently consistent with the epidemiological criteria for viral gastroenteritis.

Bacterial Contaminants

Table 10 lists bacterial contaminants reported to cause shellfish-related food poisoning. These can be split into two main groups – bacteria where the principal source of contamination is from animal or human feces, and those which occur naturally in the marine environment.

Salmonella spp. are commonly found in intestinal tracts of animals and human carriers. Over 2000 different serotypes have been reported, most of which are pathogenic to humans to varying degrees. The most serious illness is typhoid fever resulting from infection from *Salmonella typhi* (Table 10). A number of cases of this acute systemic disease occurred in the early 1900s, due to contaminated oysters from grossly polluted waters. Nowadays, modern sanitation practices, improved control systems, and a reduced number of human carriers mean this disease has not been reported in relation to shellfish from developed counties for a number of years. Gastroenteritis is a more common *Salmonella*-related illness, although again as a consequence of modern practices

Table 10 Microbes reported to cause shellfish-related food poisoning and a summary of the symptoms

Type	*Name*	*Description*	*Type of illness*	*Incubation period*	*Symptoms*	*Length of illness*
Bacteria	*Salmonella typhi*	Gram-negative	Typhoid fever	1–2 weeks	Fever, chills, headache, constipation, diarrhea (20%), muscle pain	1–2 weeks; can be fatal
Bacteria	*Salmonella* spp.	Gram-negative	Gastroenteritis	12–36 h	Diarrhea, nausea, occasionally vomiting, abdominal pain, chills, moderate fever, drowsiness	2–7 days; low number of fatalities
Bacteria	*Campylobacter* spp.	Gram-negative	Gastroenteritis	2–5 days	Bloody diarrhea, nausea, malaise, fever, and severe abdominal pain	3–5 days; low number of fatalities[a]
Bacteria	*Escherichia coli*	Gram-negative	Gastroenteritis	12–72 h	Diarrhea, abdominal cramps, nausea, and chills	2–3 days
Bacteria	*Vibrio parahaemolyticus*	Gram-negative	Gastroenteritis	4–48 h	Profuse diarrhea, abdominal pain, nausea, sometimes fever, and occasional vomiting	2–5 days
Bacteria	*Vibrio vulnificus*	Gram-negative	Primary septicemia	< 24 h	Fever, abdominal pain, vomiting, diarrhea, sepsis	50% mortality[a]
Bacteria	*Vibrio cholerae* non-O1	Gram-negative	Gastroenteritis		Diarrhea, abdominal cramps, occasionally vomiting and fever	
Virus	Norwalk-like viruses	Single-stranded RNA; 30–35 nm diameter	Gastroenteritis	1–4 days	Nausea, vomiting, abdominal pain, occasional diarrhea and low-grade fever	12–48 h
Virus	Hepatitis A	Single-stranded, RNA *c.* 27 nm diameter	Hepatitis	2–6 weeks	Fever, headache, nausea, vomiting, diarrhea, abdominal pain, jaundice	Up to several months

[a]Mortality rate highest in people with other predisposing illnesses, e.g., diabetes.
Data source: Eley AR (1992) *Microbial Food Poisoning*. Hackney CR and Pierson MD (eds) (1994) *Environmental Indicators and Shellfish Safety*. London: Chapman and Hall.

is not often caused by shellfish (Table 10). Gastroenteritis has also been reported from consumption of shellfish contaminated with species of *Campylobacter* and enterotoxigenic strains of *Escherichia coli* (ETEC: Table 10). (*See* ***Salmonella***: Salmonellosis; ***Campylobacter***: Properties and Occurrence; ***Escherichia coli***: Food Poisoning by Species other than *Escherichia coli*.)

Hemorrhagic strains of *E. coli*, particularly strain 0157, causes hemorrhagic colitis bloody diarrhea, with 10% of cases developing serious complications, such as hemolytic uremic syndrome. This may lead to death, particularly in children and the elderly. Although to date there are no reports of infection via shellfish from this pathogen, they remain of concern given their prevalence in the environment, association with cattle, and their low infective dose. Research is required to assess their survival in the marine environment, their potential to accumulate in shellfish exposed to farm run-off, and the risk to the consumer of shellfish. However, it must be borne in mind that other bacterial pathogens, including *Yersinia enterolitica* and *Clostridium perfringens*, may be isolated from molluscan shellfish, although no cases of shellfish-borne illness due to these species have been reported. Bacteria such as *Shigella* spp. and *Staphylococcus aureus* have also been associated with shellfish-borne infection, but generally by contamination from infected food handlers. (*See* ***Clostridium***: Occurrence of *Clostridium Perfringens*; ***Shigella***; ***Staphylococcus***: Properties and Occurrence.)

Vibrios, particularly species of *Vibrio parahaemolyticus* and *V. vulnificus*, are also causative agents of illness due to the consumption of raw shellfish. However, unlike the bacteria described above, these species are prevalent naturally in warm estuarine waters throughout the world. Attempts to recover these bacteria can be complicated by their reported ability to form a viable but nonculturable state. *V. parahaemolyticus* tends to cause gastroenteritis,

whereas *V. vulnificus* is a particularly invasive pathogen which may cause primarily septicemia within 24 h of exposure (Table 10). Raw oysters containing large numbers of the latter bacterial species tend to be the main source of infection, which is characterized by a mortality rate of 50% in susceptible individuals. *V. cholerae* is separated into two groups on the basis of the O1 somatic antigen. Strains possessing the O1 antigen were previously believed to derive only from human feces, but are now known to exist in low numbers in some estuarine waters. These strains are responsible for cholera, and shellfish have been implicated as a vector for the disease. *V. cholerae* non-O1 are ubiquitous in some coastal waters and are frequently identified as a cause of gastroenteritis from raw shellfish (Table 10). (*See* **Vibrios**: *Vibrio cholerae*; *Vibrio parahaemolyticus*; *Vibrio vulnificus*.)

Viruses

Viral pathogens may also cause food poisoning due to consumption of contaminated bivalve molluscs (Table 10). Epidemiological evidence suggests that the Norwalk-like viruses (NLV), also known as small round structured viruses, are the most common cause of infection. This group of viruses consists of a genetically diverse virus strain, separated into genogroups I (including the prototype Norwalk virus) and II. With an infectious dose thought to be as low as 10–100 virions, an US FDA risk assessment estimated that 100 000 cases of seafood-related NLV gastroenteritis occur per year. The immunological response to NLV infection is not well understood and appears to be short-term, with individuals becoming susceptible to reinfection within 6 months to 1 year. Patients may also suffer from a mixed infection due to viruses of both genogroups.

The most serious virus infection linked to shellfish consumption is hepatitis A. This virus is extremely stable and is currently thought to replicate in the liver. Unlike the other enteric viruses, hepatitis A has a long incubation period of about 2–6 weeks and causes a serious debilitating disease (Table 10). It is a self-limiting infection and, although it rarely causes death, infected people may be incapacitated for several months. Young children frequently experience only mild illness, whereas full symptoms develop in the majority of infected adults. Recovery is complete and leads to long-term immunity from reinfection. Occasionally, epidemics due to hepatitis A in shellfish occur. In Shanghai, China, during 1988, approximately 300 000 cases of hepatitis A were traced to clams from a sewage-polluted area. It is also interesting to note that incidences of gastroenteritis, followed by hepatitis A infection, have been reported from contaminated shellfish.

Astroviruses have also been implicated in outbreaks relating to oysters. However, detailed epidemiological data are lacking and the importance of these organisms in shellfish-related food poisoning is not clear. Other viruses, such as rotaviruses, adenoviruses, and enteroviruses, have been detected in shellfish, but to date they have not been linked with infectious disease following seafood consumption. Despite their name, enteroviruses (e.g., polioviruses) do not commonly cause gastroenteric symptoms, e.g., diarrhea and vomiting, although they do replicate in the intestinal tract and are passed via the fecal–oral routes. These viruses are often found in shellfish but have not been reported to cause infection.

Parasites

Helminth parasites have been implicated in human infection, primarily from consumption of raw or under cooked finfish and, more rarely, from crabs and bivalves. Protozoan parasites of the genera *Giardia* and *Cryptosporidium* cause human disease and are generally associated with contaminated fresh water. Both occur in detectable numbers in raw and treated sewage. There is a dearth of information regarding the survival of these organisms in sea water and their potential to contaminate shellfish.

Monitoring, Methodology, and Use of Indicators for Infectious Microbes

Bacteria Most countries have implemented sanitary controls for the production of bivalve shellfish. In the EU, these are applied via the European Directive 91/492/EEC (Table 11) and in the USA by interstate trading agreements set out in the FDA National Shellfish Sanitation Program Manual of Operations. Third-country imports into the EU and the USA have to comply with these standards. These regulations include assessment of fecal pollution in shellfish (determined by measuring bacterial indicators), processing requirements for shellfish from contaminated areas, conditions for processing and dispatch establishments, and marketing documentation.

The bacterial indicators generally involve measurement of fecal coliforms or *E. coli* in shellfish (EU) or shellfish-growing waters (USA). The concept is that if indicators are present in sufficiently high numbers, there is a high probability that specific pathogens are also present. The appropriateness of these indicators has been intensely debated over the years. Nevertheless, fecal coliform/*E. coli* indicators are used routinely worldwide and their implementation, along with specific processing methods, has resulted in a decrease in shellfish-related bacterial food poisoning.

Table 11 Regulatory limits implemented on biological contaminants of shellfish using EU regulations as an example

	Harvesting limits for bacteria		*Harvesting limits for toxins*		*End-product standard*[a]
Classification	*Microbial standard (per 100 g shellfish)*	*Treatment required*	*Toxin*	*Limit*	
A	< 230 *Escherichia coli* or < 300 FC	None	PSTs	80 μg 100 g^{-1}	Negative for *Salmonella* in 25 g of bivalve flesh
B	< 4600 *E. coli* or < 6000 FC in 90% of samples	Depuration or relayed	DSTs	Not detected in bioassay[b]	< 230 *E. coli* (or < 300 FC) 100 g^{-1}
C	< 60 000 FC 100 g^{-1}	Relayed for at least 2 months	ASTs	20 μg g^{-1}	Below harvesting limits for toxins
Ns	> 60 000 FC 100 g^{-1}	Prohibited			

[a]Microbes and toxins only.
[b]Mouse bioassay methodology not defined.
FC, fecal coliforms; PSTs, paralytic shellfish toxins; DSTs, diarrhetic shellfish toxins; ASTs, amnesic shellfish toxins; Ns, not stated but often called class D. Data source: Council Directive 91/492/EEC and 97/61/EC.

One of the main advantages of these bacterial indicators is the relative ease by which they can be detected and quantified utilizing traditional bacteria culture techniques. For example, in the UK, a five-tube, three-dilution, most probable number, two-stage technique with minerals-modified glutamate broth and chromogenic medium for detecting β-glucuronidase activity, is used for the detection of *E. coli* in shellfish.

Viruses People in a number of countries have contracted viral infections after consuming shellfish that comply with bacterial indicator regulations. This suggests bacterial standards are inappropriate controls for a viral hazard in shellfish; yet virus standards for bivalve molluscs do not currently exist. Developments of such regulatory limits are hampered largely by lack of appropriate, reliable, robust, and easy-to-use methods.

The most common methods for the detection of viruses is electron microscopy and cell culture, neither of which has proved very successful in the detection and quantification of enteric viruses, such as NLVs and hepatitis A, in shellfish. Electron microscopy is not sufficiently sensitive and many of these viruses are considered nonculturable (NLV) or difficult to culture (hepatitis A). In recent years, immunoassays have become available, although the success of application to shellfish has been limited. Molecular techniques, such as polymerase chain reaction (PCR), combined in some instances with sequencing of the amplified product or hybridization of the product with probes, seem to be the most promising approach to date. Such methods have been developed for hepatitis A and NLV detection in shellfish. Their use, however, is not straightforward because shellfish extracts contain substances inhibitory to PCR that may cause false negatives. This has resulted in the use of time-consuming, laborious extraction procedures that are problematic to apply in a routine monitoring scenario. Additionally, it is uncertain that PCR results indicate the presence of viable infectious viruses. However, some workers argue that this is not an issue because viral RNA is unlikely to survive for any length of time in shellfish without a complete protein coat. Others suggest use of antibody-capture techniques aimed at recovering the whole virion, prior to PCR. This also has the potential advantage of reducing complex extraction techniques. However, for NLVs, application of immunocapture techniques is hindered by the lack of suitable immunological agents. Although PCR-based methods have advanced research on viruses in shellfish, further work is required to simplify, improve, and standardize the methodology. It has been suggested that their routine use is a number of years in the future and that consideration should be given to utilizing viral indicators.

Bacteriophages are the most frequently suggested viral indicators because of their physical and genomic similarity to human enteric viruses, their abundance in sewage effluents, and the relative ease in which they are quantified. Methodology generally utilizes plaque assays, although it should be noted that these are not always robust. Use of general somatic bacteriophage of coliforms has also been proposed. However, standardization and reproducibility of methods utilizing this diverse group are problematic. Male specific (F+) bacteriophages are another contender as viral indicators and have been used by some researchers to study shellfish depuration. Studies have shown that F+ elimination kinetics reflects that of enteric viruses and that it is a more effective indicator of viral presence in depurated shellfish than *E. coli*. However, one criticism of F+ bacteriophage is that their distribution is not restricted to human effluents,

thereby causing doubt as to their suitability as human virus indicators in rural areas primarily impacted with animal fecal contamination. Currently, researchers are attempting to optimize systems to differentiate between animal and human F+ bacteriophage. As an alternative, the bacteriophage of the obligate anaerobe *Bacteroides fragilis* has been proposed as an indicator of human specific pollution.

An alternative to bacteriophage is the detection of other human viruses that are either widely prevalent in polluted waters and/or easier to detect than viral pathogens. Human enteroviruses have been used in this context, particularly for water. However, methods of detection are currently complex. Human adenoviruses have also recently been suggested but further work is required to determine their prevalence in relation to NLVs and hepatatis A. Overall, more work is needed to assess further prevalence of potential indicators compared to the viral pathogens, to improve methods, and to develop standards to apply before the most appropriate viral indicator approach can be developed.

Depuration and Processing of Shellfish Containing Infectious Microbes

Shellfish may be harvested and marketed directly without any further processing providing they meet specific standards, including limits on the numbers of bacterial indicators. However, if they exceed these limits, depuration and/or relaying are permissible up to certain concentrations of the indicators. To provide an example of the type of regulations imposed, the criteria set within the EU are summarized in Table 11.

Depuration procedures for treatment of microbial contaminants of shellfish were first developed during the 1920s. These involve placing bivalves in clean sea water in tanks away from the source of microbial contaminants and allowing their natural excrement process to remove contaminants within the mollusc. The tanks contain static, flow-through, or recycled sea water sterilized with ultraviolet irradiation, ozone, or chlorination. Depuration is implemented for 1–7 days, although 2 days is more common. The effectiveness is assessed generally by a reduction in fecal coliform or *E. coli* numbers to specific concentrations (e.g., to class A in the EU). However, shellfish from depuration plants meeting these limits may still contain human enteric viruses at concentrations sufficient to cause infection. Recent work suggests that during depuration viruses are eliminated at much lower rates than bacteria and sea-water temperature is critical.

Relaying requires that shellfish are removed from their original contaminated harvesting site and placed in an area with reduced microbial contaminants for at least 2 months to allow natural depuration to occur. Practical difficulties associated with this process include problems in finding suitable waters and obtaining ownership rights to these waters. One additional issue that is rarely addressed is the possibility that shellfish may be contaminated with the resting stage of toxin-producing dinoflagellates. Moving such shellfish to a 'clean' site risks seeding that site with toxic dinoflagellates. Shellfish may also be sold as a processed product which have been commercially heat-treated. These modern, regulated methods (e.g., raising internal shellfish meat temperature to 90 °C for 1.5 min) are effective at reducing both the bacterial and virus loading in contaminated shellfish.

Future Perspectives

The impact of contaminants on shellfish as a resource can be severe in terms of both the viability of the associated industries and the health of the human consumers. The highly nutritious and economic value of shellfish argues for the continuing improvement of its husbandry and the monitoring of contamination by pollutants, toxins, and microbes in order to prevent spoilage, waste, and ill health. Vigilance and awareness of existing and changing contamination threats are imperative. The selection of statutory tissue levels and contaminants to be monitored must be responsive to changing and improving information on toxicity, toxic mechanisms, production of new chemicals, occurrence of 'new' toxins, and increasing hazards from viruses. Automation, new analytical/biological techniques and the production of appropriate standards will increase the number and efficiency by which contaminants can be routinely measured. These may include biological measurements (i.e., biomarkers) of chemical exposure, such as inhibition of the shellfish cholinesterase (EC 3.1.1.8) activity as an indicator of contamination by organophosphate pesticides, and induction of 7-ethoxyresorufin *O*-deethylase (EROD) activity in cell systems as an indicator of PAHs. The development of new technology based on immunorecognition and molecular biological techniques is also likely to become increasingly important in the future. Efforts must be made to prevent contamination of this valuable resource from occurring and, in the case of toxins and pollutants, models should be produced to assist regulatory authorities to predict their accumulation and depuration in shellfish. A greater understanding of the biological processes in shellfish, once contaminated, is also required in order to develop efficient and safe treatment processes. Additionally, the standardization of statutory contaminant levels and analytical procedures between national and international bodies

will facilitate regulation and monitoring of the shellfish industry to the consumer's benefit.

See also: **Parasites**: Occurrence and Detection; ***Salmonella***: Salmonellosis; **Vibrios**: *Vibrio parahaemolyticus*; *Vibrio vulnificus*; **Viruses**

Further Reading

Anderson DM, Cembella AD and Hallegraeff GM (eds) (1998) *Physiological Ecology of Harmful Algal Blooms*. NATO ASI Series G: *Ecological Sciences*, vol. 41. Berlin: NATO.

Bricelj MV and Shumway SE (1998) Paralytic shellfish toxins in bivalve molluscs: occurrence, transfer kinetics and biotransformation. *Reviews in Fisheries Science* 6: 315–383.

Falconer IR (ed.) (1993) *Algal Toxins in Seafood and Drinking Water*. London: Academic Press.

Gosling E (ed.) (1992) *The Mussel Mytilus: Ecology, Physiology, Genetics and Culture*. Amsterdam: Elsevier.

Hackney CR and Pierson MD (eds) (1994) *Environmental Indicators and Shellfish Safety*. London: Chapman and Hall.

Hallegraff GM, Anderson DM and Cembella AD (eds) (1995) *Manual on Harmful Marine Microalgae*. IOC Manuals and Guides no. 33. Paris: UNESCO.

Langston WJ and Bebianno MJ (eds) (1998) *Metal Metabolism in Aquatic Environments*. Ecotoxicology Series, vol. 7. London: Chapman and Hall.

Lasserre P and Marzollo A (eds) (2000) *The Venice Lagoon Ecosystem. Inputs and Interactions Between Land and Sea*. Man and The Biosphere Series, vol. 25. Paris: Parthenon Publishing.

Lees D (2000) Viruses and bivalve shellfish. *International Journal of Food Microbiology* 59: 81–116.

Livingstone DR (1991) Organic xenobiotic metabolism in marine invertebrates. In: Gilles R (ed.) *Advances in Comparative and Environmental Physiology*, vol. 7, pp. 1–185. Berlin: Springer-Verlag.

Livingstone DR (1992) Persistent pollutants in marine invertebrates. In: Walker CH and Livingstone DR (eds) *Persistent Pollutants in Marine Ecosystems*. SETAC Special Publication Series, pp. 3–34. Oxford: Pergamon Press.

Livingstone DR (1998) The fate of organic xenobiotics in aquatic ecosystems: quantitative and qualitative differences in biotransformation by inverebrates and fish. *Comparative Biochemistry and Physiology* 120A: 43–49.

Reguera B, Blanco J, Fernandez ML and Wyatt T (1998) *Harmful Algae*. Xunta de Galicia and Intergovernmental Oceanographic Commission of UNESCO. Spain: Grafisant.

Thurnham DI and Roberts TA (eds) (2000) Health and the food-chain. *British Medical Bulletin* 56: 236–252.

Walker CH, Hopkin SP, Sibly RM and Peakall DB (1996) *Principles of Ecotoxicology*. London: Taylor and Francis.

Aquaculture of Commercially Important Molluscs and Crustaceans

P F Duncan, The University of the Sunshine Coast, Maroochydore DC, Queensland, Australia

Background

Aquaculture is the managed cultivation of aquatic organisms for food or profit and now contributes approximately 36% of total global fishery production. As capture fisheries have declined or plateaued over the last decade, aquaculture is the alternative supply for increasing aquatic product demand and has been growing at a rate of 10% per annum since 1990. Aquaculture products are diverse, with molluscs, crustaceans, finfish, and algae predominating. This section deals with molluscs and crustaceans whose global production volumes are considerable, for example, the pacific oyster (*Crassostrea gigas*) 3.6 million tonnes, carpet clams (*Ruditapes philippinarum*) 1.8 million tonnes and various penaeid prawn species (*Penaeus* spp.) 0.9 million tonnes. Production methods for mollusc and crustacean aquaculture vary from subsistence level, artisinal scale to high intensity managed systems. This article covers the key species, production quantities, methods and sources, and current and future issues.

General Characteristics of Aquaculture Species

The basic requirement for all aquaculture species include market acceptability, established production methods, availability of appropriate and economically viable feeds, and biological characteristics suitable for culture. These include high growth rates, captive reproduction or seed stock availability, tolerance to environmental variation (e.g., temperature), disease resistance, and behavioral characteristics amenable to high-density culture.

Molluscs

Molluscs, and particularly bivalves, are well suited to aquaculture. The principal groups include oysters (Family Ostreidae), mussels (Family Mytilidae), scallops (Family Pectinidae) and clams (various taxonomic groups), which constitute 87% of the 10.1 million tonnes (**Table 1** and *see* **Shellfish**: Commercially Important Molluscs). All of these bivalves have a high market demand and feature in the traditional diets of most human cultures. Recently, increasing affluence and health benefits associated with seafood

Table 1 Aquaculture production of principal mollusc groups and species (*see* also **Shellfish**: Commercially Important Molluscs)

Species	*Quantity (tonnes)*
Oysters	3 711 606
Crassostrea gigas	3 600 459
Scallops	951 866
Patinopecten yessoensis	928 724
Mussels	1 451 032
Mytilus edulis	498 461
Other Molluscs	4 017 574
Ruditapes philippinarum (Manila clam)	1 820 413
Solen spp. (razor clams)	479 252
Anadara granosa (blood cockle)	315 811
Total	10 132 078

From FAO (2001) *FAO Yearbook, Fishery Statistics: Aquaculture Production 1999. Fisheries Series No. 58*, vol. 88/2. Rome: FAO.

consumption have increased the demand for these products.

Mollusc Nutrition

Almost all cultivated molluscs are bivalves and therefore herbivorous or omnivorous filter feeders, consuming planktonic microalgae and organic detritus. Tridacnid (giant) clams are filter feeders with additional nutrition from symbiotic photosynthetic algae contained within their tissues. The main cultivated Gastropod group, abalone (Family Haliotidae) consists of herbivorous micro- or macroalgae grazers. All are therefore low trophic-level feeders and either obtain nutrients from their surrounding environment or have it supplied during hatchery-based culture periods. Although production of microalgae requires equipment and expertise, it is not technically difficult or expensive relative to the provision of the more complex crustacean and finfish diets. Abalone differ owing to their grazing habit. After the larval phase, juvenile abalone are provided with plates covered with bacterial and algal films. In growout systems, fresh macroalgae (e.g., *Laminaria japonica*) or recently developed artificial diets are supplied.

Production Methods (Bivalves)

Owing to the shared biological characteristics of bivalves, the basic principals of culture are similar.

Hatchery Egg production (fecundity) is high, typically several to tens of millions of eggs per spawning, e.g.,scallops and oysters, respectively. Spawning of mature adults occurs in the wild or under controlled hatchery conditions. Spawning may be induced by temperature change, emersion, exposure to UV-treated water or chemical inducers, e.g., serotonin. A planktonic larval phase follows hatching, which, in the wild, acts to disperse the juveniles from the parent stock. The larval phase, which may last from days to weeks, depending on the species, consists of several developmental stages. A ciliated trochophore larva occurs after hatching, and as shell secretion continues, the swimming larva becomes a veliger and finally, near metamorphosis to a non-swimming form, a pediveliger, which has an active foot for crawling. Under hatchery conditions, the larval phase occurs in large-volume tanks, typically between 5 and 40 m^3. For balanced nutrition, larvae are fed on mixed microalgal cultures, e.g., *Chaetoceros* spp., *Pavlova* spp., *Isochrysis* spp., and *Tetraselmis* spp.

Settlement At metamorphosis, larvae settle on to either natural or artificial substrates, and at this stage, collection of spat, or settled bivalves, occurs. In the wild, spat are collected on suspended or fixed material substrates, such as ropes (e.g., mussels), shell, stones, wooden structures (e.g., oysters), or mesh-filled suspended bags (e.g., scallops). Similar materials can be introduced into larval rearing tanks to collect spat in hatcheries.

Bivalve spat attach via proteinaceous byssal threads (e.g., mussels, scallops) or by direct cementation of the lower valve (oysters). Byssally attached spat may be easily removed manually or by mechanical stripping. Many scallop species stop byssal attachment around 15 mm and may be collected directly from settlement bags. Mussels attach throughout life. Oysters may be removed manually or retained on settlement material for further growout. All bivalves are size-graded and thinned before transfer to growout systems, producing a more consistent-sized product and a culture density to maximize growth and survival.

Growout Growout to commercial size typically takes from 12 months to 4 or 5 years (in the case of some temperate scallops and abalone). However, most cultured bivalves are an annual or 2-year crop, e.g., blue mussels (*Mytilus edulis*) 12–18 months, greenlip mussels (*Perna canaliculus*) 18 months, Manila clam (*Tapes philippinarum*) and the Australian rock oyster (*Saccostrea glomerata*) approximately 24 months.

As filter feeders, bivalves grow optimally in suspended culture system, which maximizes access to food. Consequently, most bivalves are grown in this way, although some exceptions, such as clams and some oysters, are grown on, or in, the bottom substrate.

Suspended culture systems are generally based on hanging a form of enclosure from a surface, or

subsurface floating structure. Alternatively, oysters may be held just off the bottom on a series of trays and frames. The most common systems employ a longline, a rope, up to 250 m or longer, supported by floats and anchored to the seabed at several points to ensure stability. From this main rope are hung vertical droppers that hold animals directly, e.g., byssally attached mussels, or, for ear-hung scallops, the shell is pierced and attached via nylon line.

More commonly, a cage or net holds animals. Lantern nets are a collapsible net stocking with multiple horizontal platforms holding shells, individual pearl nets have a rigid frame, and vertical panel nets contain numerous pockets in which individuals are placed. For lower-intensity operations, or for some particular species, such as the South-East Asian *Ruditapes* and *Meretrix* clam species, and the American hard clam (*Mercenaria mercenaria*), growout is based on seeding of suitable sandy/mud nearshore areas. The juveniles may be produced in the hatchery or collected in the wild.

Production Methods (Gastropods)

The main gastropods under cultivation are abalone, and, although production systems are still under development, their high value and declining wild harvest give them significant potential. Production now comes from China, Chile, South Africa, Australia, New Zealand, and the USA. The principal culture methods include a suspended lantern net system, using larger and more structurally complex cages than for bivalves, and raceway culture in long, narrow glass fiber or concrete tanks, within which flat plates are inclined to provide shelter and a surface for growth of algal feed. Tidal pond culture is also practiced in China, though less frequently than other methods.

Abalone culture relies on hatchery-produced juveniles, and broodstock are spawned using similar methods to bivalves. The larval development of abalone is similar, though shorter at 7 days or less, and nonfeeding, so provision of algae is not necessary. At settlement, the larvae metamorphose into the bottom-living form. This transitional phase from larvae to settled juveniles remains the period of highest mortality. At settlement, larvae are induced to attach to corrugated plastic or tiles, which have a surface growth of algal and bacterial cells. The juveniles feed on these films and, as growth continues, are transferred to nursery and growout systems. The options of open-water ranching or restocking depleted areas have also been attempted and have the advantage of no management input, but the disadvantage of little control over stock.

Current Issues and Future Prospects

As with most aquaculture, there is a tendency away from wild seed collection, for sustainability and supply reasons, but also for benefits from hatchery production. These include genetic selection (e.g., faster growth rates, disease resistance) and genetic manipulation to develop polyploid animals, i.e., three or four copies of genes, compared with two. Polyploid animals may be reproductively inactive, essential for translocation of non-native species to open-water systems, and often show increased somatic growth rates owing to reduced gonadal development. Continued vigilance in the control and detection of toxic algal blooms and monitoring for product safety will remain significant issues if public confidence in mollusc consumption is to be maintained.

Crustaceans

Most aquacultured crustaceans belong to the Class Malocostraca, Order decapoda, e.g. prawns/shrimps, crabs, lobsters, although brine and mysid shrimp (Classes Branchiopoda and Mysidacea) and copepods (Class Copepoda) are also produced in small quantities as live feeds for other aquaculture species, such as finfish. Decapod production is dominated by the marine penaeid prawns, with the black tiger prawn (*Penaeus monodon*) being the most widely cultivated crustacean worldwide. In 1999, production was 575 842 tonnes, predominantly from Asia, with an estimated value of US$3.65 billion. *P. vannemei* and *P. chinensis* are the aquacultured penaeids in the Americas and China, respectively, contributing 359 196 tonnes. Penaeids are grown worldwide (Table 2).

The giant river prawn (*Macrobrachium rosenbergii*), is the most widely cultured freshwater decapod crustacean, with production in 1999 of 102 124 tonnes, worth US$416 million. It has been introduced into many countries, including North and South America, Africa, Asia, and the Pacific Islands. China, Bangladesh, and Taiwan account for over 90% of production. Freshwater crayfish culture is widespread and based on six main species (in order of quantity): *Procambarus clarkii*, *Pacifastacus leniusculus*, *Astacus leptodactylus*, *Cherax destructor*, and *C. quadricarinatus*. Total production amounted to 21 171 tonnes in 1999, with US production of *P. clarkii* over 19 000 tonnes alone. Other minor species are produced locally (Table 3).

Crab aquaculture production is relatively small and dominated by the freshwater Chinese river crab (*Eriocheir sinensis*) at 171 955 tonnes in 1999.

Table 2 Principal aquacultured penaeids and main production sources

Species	*Sources*
Penaeus monodon Giant tiger prawn	Thailand, India, Vietnam, Indonesia, Philippines
Penaeus vannamei Whiteleg shrimp	Ecuador, Mexico, Brazil, Colombia
Penaeus merguiensis Banana prawn	Vietnam, Indonesia
Penaeus stylirostris Blue shrimp	Ecuador
Penaeus chinensis Fleshy prawn	China, Korea
Penaeus japonicus Kuruma prawn	Japan, China, Australia
Penaeus indicus Indian white prawn	India, South Africa

Table 3 Principal aquacultured freshwater crayfish and main production sources

Species	*Source*
Procambarus clarkii Red swamp crawfish	USA, Mexico, France (I)
Pacifastacus leniusculus Signal crayfish	France (I), Spain (I)
Astacus leptodactylus Danube crayfish	France
Astacus astacus Noble crayfish	France
Cherax destructor Yabby	Australia
Cherax quadricarinatus Redclaw crayfish	Australia, Argentina (I), Mexico (I)
Cherax tenuimanus Marron	Australia, South Africa (I)

I, introduced.

Additional species under small-scale culture, but with potential, include the mud crab (*Scylla serrata*) and various portunid swimming crabs.

Lobster culture worldwide is extremely small, 58 tonnes from four different *Panulirus* spp., a tropical/subtropical genus. The slow development of lobster culture is due to a relatively long larval cycle that is technically difficult to replicate in hatcheries.

Crustacean Nutrition

The larval stages of crustaceans vary in their nutritional requirements. Typically, under hatchery conditions, early stages feed on stored yolk, before moving to phytoplankton then zooplankton, e.g., *Artemia* (brine shrimp) larvae or rotifers. Crabs and lobsters may be carnivorous from hatching.

Penaeid prawns are fed a processed pelleted diet, consisting principally of marine fish meal, and additionally balanced for nutritional requirements. The food conversion ratio, i.e., the quantity of feed (kg) required to produce 1 kg of product is typically around 1.5–2.5:1, in semi-intensive/intensive penaeid production systems.

Feed is a major cost and may constitute around 30% of production costs. Uneaten feed is a major contributor to poor water quality, and nutrient-rich waste water contributes to external environmental impacts. Feed management is vital for successful aquaculture.

Freshwater crayfish are omnivores and obtain much of their nutritional requirements from the culture environment in the form of organic detritus. Grain-based pellets, or hay, are often added to ponds to provide additional feed and promote pond productivity by enhancing nutrient levels.

Crabs, such as *Scylla* spp., are also omnivores but with a tendency towards animal protein. Crab aquaculture is still relatively undeveloped, to the extent that no standard feed is produced or used. Most production is from South-east Asia, and much use is made of local feed sources such as trash fish and shrimp, squid, and plant meals such as soy. Pen culture in mangrove areas also provides additional nutrition in the form of fallen mangrove leaves.

A feeding behavioral characteristic of many crustaceans, and one that heavily influences production management, is cannibalism. Crustaceans are most vulnerable immediately after molting, a regular process essential for growth. When the old, hard exoskeleton is shed, the animal is soft until the new shell hardens and is vulnerable to predation, especially if stocking densities are high. Cannibalism is less common in penaeids and crayfish, where it is managed by appropriate stocking, size grading and feeding regimes, but more common in crabs and lobsters. Similar management strategies are again employed but are not wholly effective. As a consequence, crab production densities tend to be low, and separation into individual culture vessels has been employed for cultivated soft-shell crab and Hommarid lobsters. The associated cost of this makes it viable only for high-value products, or short-term culture of juveniles for restocking purposes.

Production Methods

Decapod crustaceans are characterized by complex and diverse larval cycles, which, as previously noted, have contributed to their relative ease and development of aquaculture. Differences in larval cycles also contribute to differences in production methodologies.

Hatchery The most commonly cultured crustaceans, the penaeid prawns, are unusual in having no maternal

care, and fertilized eggs are shed directly into the water. Most decapods retain the developing larvae, within the egg, on modified appendages beneath the abdomen, e.g., freshwater crayfish and crabs, and release larvae at a later developmental stage.

For aquaculture operations, mature broodstock are typically sourced from wild populations, although farm-reared broodstock are becoming more common as they enable reliable supply and the opportunity for genetic selection. Broodstock are maintained in specialized spawning tanks, under conditions of reduced density and high-quality nutrition favoring gonad development. Environmental variables such as temperature and light intensity are also controlled to reduce stress and promote gonad maturation. Mating involves transfer of a sperm package that the female retains until required, and fertilization typically occurs as the eggs are released. Spawning occurs under relatively predictable circumstances following ovary maturation, enabling hatcheries to harvest eggs as they are produced.

Penaeid eggs hatch within about 12 h into the nauplius phase, which persists through several stages, before progressing through major metamorphoses into protozoeal, mysis, and postlarval phases. The duration of the whole larval cycle, up to saleable postlarva (PL) is typically around 20–25 days. The larval phase, whether in open water or in the hatchery, is a free swimming period, becoming the more familiar bottom-living form after the postlarvae phase.

Growout Typically, the postlarval stage is transferred from the hatchery to grow out ponds, although some operations also have intermediate nursery growout or conduct growout in indoor, tank-based systems under high-density conditions. However, growout is generally earthen pond-based, ranging in size from less than 0.5 ha to over 10 ha, although larger ponds are typically harder to manage effectively. Ponds are generally 1–2 m in depth and are stocked with PL at between 1 (or less) to more than 100 PL m^{-2}, depending on the production type. For example, Australian stocking rates for *P. monodon* under relatively intensive 1-ha pond production is 30–40 PL m^{-2}.

During the production cycle of 3–4 months, water quality is managed through aeration, water exchange, and a phytoplankton bloom, which tends to stabilize physicochemical water parameters, e.g., oxygen, carbon dioxide, pH. Feed is provided three or four times per day, with consumption closely monitored and adjusted as necessary. Production from such intensive penaeid culture is typically 5–10 tonnes per hectare per crop but varies with species and management.

For other crustaceans, the principal differences in larval development and production methods are as follows.

Freshwater Crayfish Females hold 500–800 eggs under the abdomen that hatch as miniature crayfish after 2–10 weeks, depending on the temperature and species. Juveniles are collected and graded for size and gender, and stocked at around 5–15 m^{-2} (e.g.*C. quadricarinatus*), typically into earthen ponds. Production of this species seldom exceeds 1–1.5 tonnes per hectare per crop, owing to lower densities and lower intensity management.

Crabs The crab larval phase begins as a zoea, of which there are five, followed by megalopa then crab. Although still developmental, pond-based culture for *Scylla* spp. crabs appears feasible. Small-scale commercial trials have indicated that monosex cultures, low–medium stocking densities, and provision of shelters, to protect newly moulted individuals from cannibalism, provide improvements on current low-intensity culture systems. Polyculture with penaeids, as a secondary crop, also occurs.

Perhaps the greatest opportunity for value adding to crab culture is a soft-shell product. Soft-shell crabs have been a valuable secondary product of the eastern US blue crab (*Callinectes sapidus*) fishery for more than 100 years. The development of culture techniques, and the ability to control aspects of molting, should enable commercial scale production within a few years. Presently, it remains labor-intensive and confined largely to South-east Asia.

Macrobrachium spp. have even fewer distinct larval stages, with several zoea, followed by a pl, then adult. Unusually, the larvae require brackish water for development, although culture is in fresh water. This finding was critical in enabling the development of *Macrobrachium* aquaculture.

Harvest, Handling, and Marketing Harvesting is dependent, to some extent, on the market requirements. Fresh, frozen, or processed crustaceans are drain-harvested, with the complete crop being removed immediately. This applies to most penaeids, e.g., *P. monodon*, crabs (except soft-shell, which are harvested at molt), and *Macrobrachium* spp. Most live products, e.g., freshwater crayfish and some penaeids (*P. japonicus*), are harvested in small batches, typically using baited traps. Harvesting of crayfish may make use of their rheotactic behavior, in which the animals self-harvest, by following a water flow up a ramp into a collection chamber.

Crustaceans generally, and prawns in particular, are typically processed and sold as either frozen

(cooked, green, whole or tails and as block or individually quick frozen), fresh chilled, or as a prepared or preserved product, e.g., canned. Aquaculture tends to concentrate on the higher value fresh or frozen product, to offset production costs. Some species, notably, *Penaeus japonicus*, freshwater crayfish, and some crabs, e.g., *Scylla* spp., are marketed live for value-adding or spoilage reasons. Live kuruma prawn can attract prices of US$75 kg^{-1} during Japanese holiday periods, justifying the significant extra harvest, processing and packaging procedures necessary to ensure survival. By contrast, freshwater crayfish tend to be marketed live to ensure product quality at market.

Current Issues and Future Prospects

The principal issues affecting penaeid prawn culture worldwide are disease and environmental impacts. The possibility of rapid disease spread in high-intensity production systems is high, and mortality can be significant. At the individual farm level, and outwith epidemic-type outbreaks, producers manage disease risk through quarantine, exclusion of disease vectors, health testing of stock, and management of water quality and animal density.

On a larger scale, where entire regions are affected, disease impacts, particularly viruses, have been catastrophic. These outbreaks are harder to manage and appear to be due to overall pathogen levels in stock, both broodstock and pl, and the general status of the culture or external aquatic environment. For example, the passage of whitespot syndrome virus (WSSV) and yellowhead disease through Asia can be traced from the early and mid-1990s. Whitespot alone was considered to have resulted in losses of US$600 million in Thailand and in excess of US$2 billion throughout Asia in 1997.

Similarly, taura syndrome virus and WSSV affected central and South American production from around 1992.

These diseases occur naturally in the environment, but significant outbreaks occur when animals are produced under high-density, sub-optimal conditions. However, introductions of diseases, for example from Asia to the Americas, have occurred.

Disease is also a major concern for freshwater crayfish production given the impact that the fungus *Aphanomyces astaci*, or crayfish plague, had on wild and captive populations in Europe. As a result, several resistant species have been introduced from North America, now forming the basis of European production. The high susceptibility of native Australian crayfish to the fungus has ensured that strict quarantine procedures apply to importations of potential vectors.

The environmental impacts of aquaculture, and particularly prawn farming, have attracted significant debate in recent years. Largely as a result of unregulated development in Asia and South America, prawn aquaculture has been linked with coastal ecosystem destruction, disease outbreaks and introductions, and unsustainable practices. Consequently, significant environmental and planning regulations are now enforced in many countries new to aquaculture, such as Australia and the USA, and increasingly so in established aquaculture countries. Regulatory measures have included limiting farm sizes and number, control of discharge water quantity and quality, and the introduction of disease-management and health-monitoring strategies.

The reliance on fish meal and oil for the production of animal feeds is an issue facing aquaculture. Estimates suggest that fish meal use by aquaculture will increase from the 35% of total production (6.5 m tonnes in 2000) to 55% by 2010. In addition, fish-oil use by aquaculture was 55% of total production in 2000 and is expected to rise to 75% over the same period. Although mostly used in finfish diets, crustacean feeds also contain significant quantities, and usage should be reduced because of supply-reliability issues and energy-conversion inefficiency. Substitution with plant proteins, such as soy, may represent viable alternatives.

Crustacean aquaculture offers significant advantages over traditional wild-capture fisheries, which are generally in decline or static. The ability to manage most aspects of aquaculture means that, theoretically, the environmental impact can be minimized, which will be essential for public acceptance of aquaculture product and practices. Waste-nutrient management programs, including the use of plants and animals to lower nutrient concentrations (bioremediation), multiple water use for irrigation, and polyculture will reduce off-farm impacts and improve production efficiency. Improved feed management, reduced fish meal usage, and the application of biotechnology, in the form of genetic selection and management, should contribute further to cost reduction and efficiency improvement, enabling aquaculture to continue its growth.

Nutritional Features of Commercially Important Shellfish

Shellfish are high in protein and unsaturated oils, and low in cholesterol. They are particularly high in ω-3 and ω-6 polyunsaturated fatty acids, which are considered beneficial in reducing high blood pressure, coronary heart disease, arthritis, and some cancers. Consequently, nutritionists and medical authorities have promoted their consumption recently. Table 4

Table 4 Proximate nutritional composition (per 100-g fresh weight) of principal aquacultured crustaceans (sources Nettleton (1985), Yearsley *et al*. (1999))

Species	*Protein (g)*	*Total fat (g)*	*Carbohydrate (g)*	*Energy (kJ)*	*Cholesterol (mg)*	*Fatty acids (Saturated/monounsaturated/ polyunsaturated/ ω-3)*
Black tiger prawn *Penaeus monodon*	19.2	0.6 (0.4–0.7)	1.8	394	120–160[a]	0.28/0.18/0.34/0.07–0.34[b]
Freshwater crayfish *Astacus* spp.	16.0	0.05	1.0	76		0.21/0.26/0.33/-[c]
Blue crab *Calinectes sapidus*	16.2	1.0 (0.4–2.2)	0.6	81	76	0.17/0.22/0.40/0.38[d]
Rock lobster *Panulirus argus*	19.2	1.2	1.7	100	106[e]	0.14/0.14/0.59/0.27

[a]General data for penaeids.
[b]ω-3 range from other penaeids.
[c]Data for *Cherax quadricarinatus*.
[d]Data for *Cancer Magister*.
[e]Cooked product.

shows the proximate composition of the important nutritional components of key cultivated crustaceans. (*See also* Table 5 in **Shellfish**: Commercially Important Molluscs.)

See also: **Shellfish**: Commercially Important Molluscs

Further Reading

Bayne BL (ed.) (1991) The biology and cultivation of mussels. *Aquaculture* 94(2–3): 121–278.
Fallu R (1991) *Abalone Farming*. Oxford: Blackwell.
FAO (2001) *FAO Yearbook, Fishery Statistics: Aquaculture Production 1999. Fisheries Series No. 58*, vol. 88/2. Rome: FAO.
Fast AW and Lester LJ (eds) (1992) *Marine Shrimp Culture: Principles and Practices. Developments in Aquaculture and Fisheries Science*, vol. 23. Amsterdam: Elsevier.
Keenan CP and Blackshaw A (1999) Mud crab aquaculture and biology. In: *Australian Centre for International Agricultural Research, Proceedings No. 78. Canberra.*
Landau M (1992) *Introduction to Aquaculture*. New York: John Wiley.
Matthiessen G (2000) *Oyster Culture*. Oxford: Blackwell.
Nettleton J (1985) *Seafood Nutrition: Facts, Issues and Marketing of Nutrition in Fish and Shellfish*. New York: Osprey Seafood Handbooks.
New M and Valenti W (eds) (2000) *Freshwater Prawn Culture*. Oxford: Blackwell.
Philips BF and Kittaka J (2000) *Spiny Lobsters, Fisheries and Culture*. Oxford: Blackwell.
Primavera JH (1996) Environment friendly aquaculture and rehabilitation in mangrove ecosystems. In: *World Aquaculture 96. Annual Meeting of the World Aquaculture Society Jan. 29–Feb. 2, Bangkok, Thailand.*
Shumway SE (ed.) (1991) *Scallops: Biology, Ecology and Aquaculture. Developments in Aquaculture and Fisheries Science*, vol. 21. Amsterdam: Elsevier.
Yearsley GK, Last PR and Ward RD (1999) *Australian Seafood Handbook*. Canberra: CSIRO Marine Research.

Dietary Importance *See* **Fish**: Dietary Importance of Fish and Shellfish

SHERRY

Contents

The Product and its Manufacture

J Bakker, Old Town, Swindon, UK

Introduction

Sherry is the name given to a number of related fortified wines made from grapes grown in Jérez de la Frontera, in the province Cádiz in the south of Spain. The hot climate of the region would be expected to produce white grapes which would turn into rather bland white table wines. However, as a result of a unique method of maturation, the wines have individuality and style. White wine made from Palomino fino, the main grape variety, is fairly neutral, lacking in acidity and without distinct varietal character. This neutral base forms an excellent background for the delicate flavors produced as a result of the maturation and blending procedures. There are three main types of sherry, made from the base wine by three different aging techniques. Sherry can be matured under flor (a layer of yeasts growing on top of the wine), developing into fino, or matured without flor yeasts, developing into oloroso. A combination of flor maturation followed by a period of aging without flor results in amontillado. The type of wine making has also been adopted by other wine-making regions, but the discussion below will focus on sherry as made in the demarcated area around Jérez de la Frontera. The quality and style of the product depends on many aspects, such as grapes, viticulture, soil, and climate. The combination of factors is unique for this region, making sherry stand apart from these wine styles made elsewhere. The factors affecting the production of grapes, the wine making, the maturation, and final blending of sherry are described.

Grape Production

Climate

The climate in Jérez de la Frontera is generally warm, with maximum temperatures in the summer of about 40 °C. Rainfall is moderate and can be variable, but is generally about 60 cm $year^{-1}$. Most rain falls between November and February, but rain during the vintage in September is not uncommon. It causes inconvenience in the harvesting of the grapes, but has no effect on the quality of the grapes. Winters are cool, and occasionally there is light frost, which tends to occur when the wines are dormant, and therefore it does not usually cause any damage. The predominant winds are southwesterly, and come from the sea and have a cooling effect while increasing the humidity. The southeasterly winds are unpopular and are believed to dry the crop due to a drop in humidity.

Soil

The best vineyards consist of rather chalky soils, known as Albariza, and are situated in the low hills west of Jérez de la Frontera. The main merit of the soil seems to be its associated physical properties: this soil has good water-absorbing properties and is able to retain water throughout the very dry summer. However, the soil goes to a paste-like consistency when wet, making vineyard access difficult. The soil dries to a powdery structure, without cracks. The light color of the soil reflects the heat, and also helps to prevent the soil from drying out. Irrigation of the soil is not allowed, but water availability has been improved by cutting square basins in vineyard slopes after the harvest. This reduces soil erosion by diminishing the water flow during the rains.

Vineyards

All vines are grafted on rootstocks to avoid infection by the root aphid *Phylloxera*. The rootstocks are hybrids between American and European vines; the American influence is to confer resistance to the root aphid *Phylloxera*. The rootstock should not influence the flavor of the grapes, but can affect the vigor of the vine. The very high calcium carbonate in the soil requires special selection of rootstock, since not all are able to grow in this type of soil. Young vines start to produce small quantities of grapes about 2 years after grafting. A good commercial yield is obtained after 6 years, and after 25 years the vineyard gives its maximum production of fine wines. After 30–40 years the yield drops, and when the vineyard is no

longer economically viable, the vines are pulled up, and replanted once the soil has recovered. Currently only four varieties are allowed to be planted (Palomino fino, Palomino de Jérez, Muscatel and Pedro Ximénez). Palomino Fino is the main variety grown; this variety yields a much higher crop than Palomino de Jérez, which is currently not planted. Only about 2% of the planting consists of Muscatel, grown on sandy soil, and a little can be used to flavor the wines. Pedro Ximénez is used to sweeten wines, and producers are allowed to buy this outside the demarcated area. Rows of vines are now planted well apart (2.3 m), so access with machinery for mechanical harvesting is possible, but these are not yet in commercial use. Vines are generally supported by stakes with two wires 50 cm and 100 cm above the ground respectively. The vines are pruned low with two short spurs 30–40 cm above the ground. However, low pruning systems cannot be harvested mechanically. Concerns exist regarding the possible damage to grapes due to mechanical harvesting, with negative effects on wine quality due to expected oxidation problems.

Harvest

The grapes must have a potential alcohol content of at least 10.5% by volume before they are allowed to be picked. Generally, 11.5% potential alcohol by volume and a fairly low acidity (2.75 g l^{-1} total titratable acidity as tartaric acid) are considered satisfactory. Optimum ripeness is defined as maximum juice and sugar yield, when the stems start to darken and the seeds separate easily from the pulp. Since only one grape variety is grown in a small area, the grapes tend to ripen all at the same time, and harvesting needs to take place over a short period. Generally the harvest starts in the first week of September, and takes no more than 3 weeks. Since damage to the grapes leads to the release of polyphenol oxidase, which will cause browning, care is taken to cause minimal damage to the grapes. Grape bunches are picked by hand, put in small containers, and transported to the winery. Crushing within 4 h of picking will help to prevent tissue damage and the resulting negative effect on wine quality. The style of wines depends on the growing season. In hot years, with drying easterly winds, the vintage is early, and more oloroso is produced, while in cooler, more humid years the vintage is late, and more fino is produced.

Wine Making

Pressing

A number of crushing and pressing methods are used. Whole bunches are conveyed from the hopper into a roller crusher by screw conveyor. The stalks help with the draining of the juice, and are not removed. Pressing takes place immediately after crushing, to avoid juice contact with the skins and the stems. Both batch and continuous presses are used, and although both produce good-quality juices, the latter is more economical to use due to the lower cost of labor. It is at this stage that the juice is separated for the intended wine style to be produced. The free-run juice and the early pressings containing no more than 200 mg l^{-1} total phenols (expressed as gallic acid equivalent) are separated from pressed fractions which contain up to 500 mg l^{-1} total phenols, while heavily pressed juices containing more than 850 mg l^{-1} total phenols are not suitable to make quality sherry. The solids in the must are reduced to about 1% before fermentation. Therefore the juices are usually cooled (12–15 °C) and allowed to settle for 8–18 h to allow the excess solids to deposit, before being centrifuged. Since the acidity of the grapes is often low due to the warm growing conditions, traditionally calcium sulfate was added to the grapes at crushing, precipitating calcium tartrate, and apparently liberating free tartaric acid. Nowadays, an addition of tartaric acid is made to juices with low acidity to reduce the pH to about 3.45. An addition of about 100 mg l^{-1} sulfur dioxide is made to the juice; the amount is varied according to the condition of the grapes. Sulfur dioxide has a number of functions and will be more effective at a lower pH value. It inhibits browning of the juice, thus protecting its pale yellow color. It also inhibits some of the undesirable microflora and favors the growth of fermenting yeasts.

Fermentation

Sherry used to be made by fermentation in oak casks, with a capacity of 500–600 l. Due to the small volume of fermentations, temperature control was not essential. Some wines are still made in small casks, to produce blending wines with a characteristic flavor, or to season new vats before being used for the maturation of sherry, or to satisfy the demand from the Scottish whisky industry for vats seasoned by sherry. Currently most wine is fermented in large open cylindrical fermentation tanks, normally with a capacity between 500 and 1000 hl. Temperature control in these large tanks is essential, and most wine is fermented at about 25 °C. Commercial yeast inocula are not normally used, but yeast starters are often prepared by harvesting some grapes just before the vintage and allowing the naturally present fermenting yeasts to develop. This will help to ensure that the fermentation vats will start to ferment after a short time. The initial fermentation is dominated by apiculate yeasts, such as *Kloeckera* and/or

Hanseniaspora, but at the later stage *Saccharomyces cerevisiae* (the main fermenting yeast) predominates. Most of the fermentation will have completed in about a fortnight, but the fermentation continues slowly until all the sugar has been depleted. (*See* **Beers**: Biochemistry of Fermentation; **Yeasts**.)

Classification

By November the fermentation is finished, and the dry wines have an alcoholic strength of 11–12% by volume. By that time, lactic acid bacteria naturally present in the wine have completed the malolactic fermentation, which converts the fairly sharp-tasting malic acid into the softer-tasting lactic acid and carbon dioxide. The wines are initially classified on the basis of their quality, and fortified accordingly. However, at this stage it cannot always be predicted which wines will spontaneously produce flor, which is a layer of yeasts on the top of the wine and allows the wine to develop the typical characteristics of fino sherry. Generally, wines from grapes grown on soils with high calcium carbonate content in vineyards with a cool westerly exposure are likely to develop flor. Therefore in cooler years more fino-style wines are produced. Fino is usually made from the lighter dry wines, made from mainly free run juice and some light pressings. This wine has a pale yellow color, low total phenols (200 mg l^{-1}), a good pungent aroma, a volatile acidity between 0.3 and 0.5 g l^{-1} (expressed as acetic acid), and is free of any bacterial spoilage. These wines usually develop flor spontaneously. Oloroso is usually made from slightly darker wines, with higher concentrations of total phenols (up to 475 mg l^{-1}), which are less likely to develop and support flor. These base wines have a more vinous and full-bodied nose and a higher volatile acidity (0.7–0.9 g l^{-1}). Wines made from the higher pressings and subsequently with higher total phenols (over 550 mg l^{-1}) are classed as raya wines, but also belong to the oloroso style.

Fortification

After fermentation the wines usually clarify spontaneously, and just after their classification they are racked of their lees (the deposit of yeast cells, and lactic acid bacteria from the malolactic fermentation). At that time the wines are fortified with fortifying spirit, which is almost neutral in character. The fortifying spirit makes little direct contribution to the sensory characteristics of the wine, but it will affect the maturation of the wine. It is prepared by distillation of wine and wine byproducts such as lees and pomace (the left-over solids after pressing the grapes). It has a strength of at least 95% by volume. The fortifying spirit is also produced outside the sherry region and is usually produced in the La Mancha area in central Spain. Fino and oloroso wines are fortified 15.5% alcohol by volume, while raya wines are fortified to 18.5% alcohol by volume. Before adding the fortifying spirit it is mixed with an equal quantity of wine, and allowed to settle for about 3 days. The half-strength spirit is used to fortify the wines, and causes much less clouding of the wine than the undiluted spirit would have done.

Maturation

Wines are stored in bodegas, which are tall, well-ventilated buildings, designed to stay relatively cool, especially from the ground level to about 2 m above the ground. The wines are matured in seasoned oak casks, with a capacity of 500–600 l, and organized to simplify fractional blending. The wine can either be matured under flor, and develop the typical fino character, or matured without flor to develop oloroso wines. The wines will be allowed to mature unblended for about a year, and a collection of wines all from one year is referred to as añada. Young fino wines are kept in vats only 80% full, providing a relatively large wine–air interface on which flor can develop, and may need frequent topping up to maintain the wine level. They are kept at a uniform temperature in the cooler part of the bodega, and the alcohol is maintained between 14.5 and 15.5% by volume. The casks are disturbed as little as possible, to prevent any damage to the layer of flor. Young oloroso wines, which did not develop much or any flor, are fortified to 18.5%, which inactivates any flor yeast and prevents flor from being formed. The casks are kept 95% full and their storage temperature is less critical than for the fino wines. Raya wines will be matured at higher temperature, often by storing them outside in the sun, and the losses in wine as a direct result of evaporation can be substantial. They will develop into dark wines and can be blended with olorosos, or develop into oloroso-style wine themselves. In order to obtain large quantities of wines with a consistent quality, an elaborate fractional blending system is used, referred to as a solera system. The sherries from the añada will gradually be fed into a solera system.

Flor

The wines with low total phenol contents suitable to become fino-style wines will spontaneously develop a wrinkly film of yeast on the surface of the wine. The composition of the wine determines whether a vigorous flor is able to develop. There is still a debate regarding the origin of the flor yeasts and their taxonomy, although a number of the flor yeasts belong to

the *Saccharomyces* species. However, the yeasts are physiologically different from the fermenting yeasts dominating during the fermentation. The fortification of the wines to 15.5% alcohol (by volume) inhibits film-forming acetic acid bacteria, which would spoil the wine. Development of flor depends on the temperature, and wines should be stored between 15 and 20 °C. The activity of the flor depends on the season; the flor is most active between February and June, and then declines until October time, when activity increases again. The flor protects the wine from the uptake of oxygen, and prevents oxidative browning, to which the wine is very susceptible. The growing flor will consume any dissolved oxygen and protects the wine from absorbing oxygen from the head space in the cask. This lack of oxygen in the wine prevents the oxidation of phenols, hence fino maintains a pale yellow color. The flor also results in numerous biochemical changes, which greatly influence the character of the wine. There is a slow decrease in volatile acidity, a reduction in glycerol, and a reduction in alcohol, which is used as a carbon source for flor yeast growth. There is a significant increase in acetaldehyde, normally to 260–360 mg l^{-1}. Acetaldehyde contributes to the typical, somewhat apple-like nose of fino wines. Possibly autolysis (spontaneous rupture of old yeast cells, thus releasing their contents) of the yeast cells which form part of the sediment on the bottom of the cask may well contribute to the typical flavor. The film should not be disturbed; however, the blending system is crucial to maintain the flor. Small volumes of wine have to be taken out frequently, and replaced with fresh wine from the añada to renew the oxygen supply and renew nutrients for flor maintenance. Specific wine transfer techniques have been established to transfer the wine with minimal disturbance of the flor in order to maintain the flor and to achieve the blending.

Maturation without Flor

In order to produce a good oloroso, the neutral base wine needs to contain sufficient oxidizable phenols, and is therefore not suitable for fino wines. Fortification to 18.5% alcohol by volume avoids the growth of any flor. Wines are stored in warmer parts of the bodega and under oxidative conditions. The dark golden color of the oloroso can be attributed to oxidation of phenols. Although the wines are fermented to dryness, the glycerol formed during alcoholic fermentation (7–9 g l^{-1}) gives a hint of sweetness on taste. A considerable loss in volume during maturation and usually an increase in alcoholic strength have been reported in oloroso wines, resulting in a higher concentration of nonvolatiles. The volatile acidity and the concentration of ethyl acetate increase also, in part as a concentration effect and in part formed by oxidation and esterification. The higher alcohol concentration, often combined with a higher storage temperature, may lead to increased extraction of phenols from the wood during maturation of olorosos, explaining the higher concentrations of phenols. Oloroso takes a long time to develop and requires about 7–8 years of maturation.

Maturation With and Without Flor

Some wines mature initially under flor, and develop all the characteristics of a fino. However, when the wines are not refreshed with additions of younger wines in the solera system, or when the wine has reached a considerable age, the wine may start to lose the flor. Further fortification to about 17.5% alcohol by volume is usual, to protect the wine against spoilage by acetic acid bacteria and prevent any further development of flor. The wine is matured in a second solera system in casks which are 95% full, and the aging processes change, with oxidation of the wine changing the pale yellow color to amber and dark gold, together with the development of a nutty, complex flavor typical for this wine, called amontillado. It takes at least 8 years to transform a fino into an amontillado, and genuine amontillados may have gone through three fino solera systems (which may be between 18 and 21 criaderas) and an amontillado solera system (5–8 criaderas).

Solera System

The solera system contains blends of wines of different ages and vintages, and aims to mature wines to achieve a steady supply of wine of comparable and consistent quality. The blending procedure to which the solera system is subjected is very complex. Some of the oldest wine is frequently taken out, and prepared for shipping. Wine from the previous stage is used to replace the oldest wine taken out, and the topping-up of older wines with younger wines continues throughout the solera system, until the youngest stage is reached; this is then replenished with a suitable añada wine. Wine in the last stage of the solera system is usually called solera, while the other wines at various stages of blending and maturity are referred to as criaderas. Since the fino wines are most variable in composition, they tend to need the largest number of blending stages or criaderas. Careful management of the system is needed to ensure that a continuous flow of wine with the required quality characteristics is obtained. The añada wine, with which the solera system is topped up, is carefully selected based on analysis and tasting, but there will be some differences in composition from year to year. A typical solera system with flor

maturation may consist of five criaderas and a solera, each stage comprising 100 casks. The maturity of the wine in the casks increases progressively, with wines in the fifth criadera being the youngest. Typically upto 25% of the wine in a fino solera is moved every 3 months. Less frequent withdrawals of larger volumes tends to hinder the development of the flor. However, in order to meet seasonal demand, the withdrawals need to be made at somewhat irregular intervals.

Coloring Wines

Special coloring wines are prepared, which also impart flavor to the blend. Grape must is concentrated by boiling down, which heavily caramelizes the must. This concentrate is added to fermenting must made of the last pressing of Palomino fino grapes, in a sufficient quantity to stop the fermentation at about 8% alcohol by volume. After about 3 months the wine is racked, fortified to 15–17% alcohol by volume, and matured in an oxidative solera system. A number of food caramels are also allowed for use in sherry. (*See* **Caramel**: Properties and Analysis.)

Sweetening Wines

A number of different sweetening wines are prepared. One is made from fortified free-run juice, which is racked, fined, and matured before use. A wine with more character (dulce pasa) is made by drying the bunches of Palomino grapes for 2 weeks in the sun. The juice is fortified to 17.5% alcohol by volume, and the wine is matured in a solera system, where it develops a deep golden color and raisin-like flavors. This dulce pasa is increasingly used, since it can be made more economically than the traditional sweetening wine made from Pedro Ximénez grapes. Pedro Ximénez wines have flavors which contribute to the character of blended sweet sherry. Pedro Ximénez grapes are dried in the sun for up to 20 days, until they have doubled in sweetness. During this process they develop a characteristic dark color and flavor, reminiscent of a rich Christmas cake. The grapes are pressed and the juice is fortified to about 9% alcohol by volume. Pressing these very dry grapes is difficult, and occasionally fresh grapes are added to aid the pressing. The Pedro Ximénez wines are usually matured in an oxidative solera system. Pedro Ximénez wine can be imported from the Montilla district, where the hotter climate lends itself better to the production of the grapes. Rectified concentrated grape musts, which are almost colorless and odorless, are also widely used to obtain pale cream sherries, by sweetening fino wines without altering their delicate flavor.

Volatile Compounds

The volatile compounds in sherry can be derived from the grapes, formed during the fermentation or maturation. It is during this last stage, especially when maturation under flor is used, that the aroma typical for the style of sherry is formed, and hence forms the most important stage for the formation of volatiles typical for sherry. Sotolon (4,5-dimethyl-3-hydroxy-2(^{5}H)-furanone) has been identified in sherry matured under flor only; it is present in sufficient concentrations to lend its nutty odor to the characteristic aroma of flor sherry. Solerone (4-hydroxy-5-ketohexanoic acid lactone) has been reported to contribute an aroma of premium-quality Pinot noir wine. Oak lactone (*trans*-3-methyl-4-hydroxyoctanoate acid lactone) extracted during maturation in oak gives a distinct woody note to sherry, and has also been found in other alcoholic beverages matured in oak. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Commercial Styles

The end products have very complex sensory characteristics. Sherry is fermented to dryness, and fino sherries are usually marketed dry. Finer and older wines can also be sold dry, but most of the wines are sold sweetened. Fino has a pale straw color, and is very dry but without acidity. Its bouquet is delicate yet pungent and the alcoholic strength usually lies between 15.5 and 17% by volume. Manzanilla, produced in the coastal town Sanlúcar de Barrameda, is a regional variation of the fino style, which has a delicate and highly individual aroma. It is very dry, with a clean and slightly bitter aftertaste, being slightly less full-bodied than fino. Manzanilla has a pale straw color much like a fino, and an alcoholic strength between 15.5 and 16.5% by volume. Amontillado is very dry and clean, with a pungent aroma reminiscent of fino but nuttier and fuller-bodied. The color is amber, and becomes darker with increasing age. The alcoholic strength is 17–18% by volume. Oloroso has a strong bouquet, being less pungent than fino or amontillado, but with more body, sometimes referred to as fatness. The sherries are dry, but have a sweet aftertaste. Oloroso has the darkest color, best described as dark gold, and its color intensity increases with age. Wines marketed as medium dry, medium, cream, or pale cream sherry are usually based on oloroso wine, with an addition of fino to lighten the color and flavor and a small amount of amontillado. The choice of sweetening and coloring material depends on the required end product. Pedro Ximénez is often added to round the flavor of the blend.

Final Processing

Most blends are prepared a few months before shipping, which allows the final blend to be stabilized. Especially since such different wines are blended, it is important to ascertain that no instabilities occur, and stabilization and fining techniques are commonly used to ensure a clear and stable wine in the bottle.

See also: **Grapes**; **Wines**: Dietary Importance; **Yeasts**

Further Reading

Gonzalez Gordon M (1972) *Sherry, The Noble Wine*. London: Cassell.

Goswell RW (1986) Microbiology of fortified sherries. In: Robinson RK (ed.) *Developments in Food Microbiology*, 2, pp. 1–19. London: Elsevier Applied Science.

Goswell RW and Kunkee RE (1977) Fortified sherries. In: Rose AH (ed.) *Economic Microbiology*, vol. 1, *Alcoholic Beverages*, pp. 477–535. London: Academic Press.

Martin B, Etievant PX and Henry RN (1990) The chemistry of sotolon: a key parameter for the study of a key component of flor sherry wines. In: Bessiere Y and Thomas AF (eds) *Flavour Science and Technology*, pp. 53–56. Chichester: John Wiley.

Reader HP and Dominguez M (1994) Fortified wines: sherry, port and Madeira. In: Lea AGH and Piggott JR (eds) *Fermented Beverage Production*, pp. 159–207. Glasgow: Blackie.

Webb AD and Noble AC (1976) Aroma of sherry wines. *Biotechnology and Bioengineering* XVIII: 939–1052.

Composition and Analysis

M Medina, J Moreno, J Merida, M Mayen, L Zea and L Moyano, University of Cordoba, Campus de Rabanales, Cordoba, Spain

Background

Most sherry wines are produced in two southern regions of Spain, Jerez and Montilla-Moriles, where the grape varieties Palomino Fino and Pedro Ximenez, respectively, are preferentially grown. These grapes are used to obtain so-called fino, oloroso, and amontillado wines, which are three typical types of these wines. These are obtained starting from the same wine base but subjected to different aging conditions. Fino wines have a very light yellow color and an almond flavor with pungent notes. Like some Italian- and French-like sherry wines, they are obtained by biological aging under the action of flor yeasts growing on the wine surface. This aging procedure involves a complex series of wine transfers that leads to a mixture of different vintages in the same cask. Oloroso wines are obtained by oxidative aging, after fortification with ethanol, in order to prevent growth of flor yeasts. They have a very dark color that results from the oxidation of phenolic compounds, and a flavor with distinct notes of oak and walnut. Amontillado wines are produced by using the previous two aging procedures in succession (biological aging first and then oxidative aging). These wines are highly desirable, have a color in between those of fine and oloroso wines, but closer to the latter, and have a flavor with hazelnut notes that is the most complex among the three.

Musts and Wines. General Composition

The prevailing climate of the sherry-making regions of southern Spain determines to a great extent the general composition of the musts and wines they produce, and also some winemaking techniques. In summer, temperatures above 35 °C, or even above 40 °C in areas far from the Mediterranean coast, are usual during the grape-ripening period. A milder climate can be found in zones such as Sanlucar de Barrameda, on the coast, which produces a wine similar to the fino type (called Manzanilla) but with distinctive sensorial characteristics. Under these climatic conditions, grapes accumulate large amounts of reducing sugars, resulting in concentrations above $200\,g\,l^{-1}$ in the must (specific gravity $1.094\,g\,ml^{-1}$) or even more than $270\,g\,l^{-1}$ (specific gravity $1.113\,g\,ml^{-1}$) in warmer areas. As a result, the base wines obtained after alcoholic fermentation have a high ethanol content and are classified according to their suitability for producing the different types of sherry wines. Wines to be aged biologically are fortified up to 15% v/v ethanol content, if necessary, whereas those destined for the production of oloroso wines are fortified to 17–18% v/v.

The climatic conditions also result in musts with a low acidity (3–4.5 $g\,l^{-1}$ as tartaric acid) that is usually increased by addition of tartaric acid, although winemakers in some areas continue to use gypsum for this purpose. The musts are also supplied with SO_2 at concentrations of around $100\,mg\,l^{-1}$.

The special features of the process by which flor yeasts produce fino wine have aroused much interest among researchers. The yeast types most frequently used for this purpose are *Saccharomyces cerevisiae* varieties *beticus*, *cheresiensis*, and *montuliensis* in the Jerez region, and races *capensis* and *bayanus* in

the Montilla-Moriles region. These yeasts develop an aerobic metabolism, consuming ethanol (1–1.5% v/v) and glycerol (decreasing from 7–8 g to less than $0.5\,g\,l^{-1}$), in addition to some acids such as acetic (the concentration varying as the volatile acidity falls from 0.4 to $0.04\,g\,l^{-1}$) and lactic acid, which are converted to acetaldehyde and other byproducts. Quantitatively, changes during the biological aging process depend on the distribution of the yeast population and on the stability of the film they form; this results in some differences among fino wines produced in different zones. However, the development of flor yeasts is essentially influenced by the ethanol content in the wine (around 15% v/v) and the temperature during aging (below 22.5 °C). Any deviation from these conditions (e.g., higher temperatures) deteriorates the flor film and retards the aging process.

In the absence of flor yeasts, oloroso wines consume no glycerol, which remains at levels of 6.5–$8\,g\,l^{-1}$ during oxidative aging. However, their volatile acidity increases during aging (up to $0.68\,g\,l^{-1}$ as acetic acid), because of the oxidation of ethanol and acetaldehyde. Amontillado wines exhibit an intermediate behavior between those of the previous two as a result of the mixed aging procedure used.

The three types of quality sherry wines require very long aging times: at least 4–5 years for fino wines and more than 10 years for amontillado wines. In order to shorten the process, some authors recommend the use of various procedures to accelerate biological aging of the wine. Such procedures are based on an increase in the population densities of flor yeasts, which can be obtained by using containers with an increased surface-to-volume ratio or controlled aerations of the wine during aging. Oloroso wines have also been subjected to accelerated aging procedures involving higher temperatures or maceration with oak wood shavings.

Volatile Compounds and Flavor

Musts and Wines

Palomino Fino and Pedro Ximenez, the two grape varieties preferentially used to obtain sherry wines in the Jerez and Montilla-Moriles regions, are occasionally aromatic with an overall free terpene content in the musts of $0.27–0.37\,mg\,l^{-1}$ for the latter variety. In order to improve the sensory profile of the must, cold maceration with grape skin at 10 °C for 24 h has been found to increase the terpene content by a factor of 4.5, but these results probably do not offset the cost of the operations involved.

Traditionally, the musts are fermented with wild yeasts, the population consisting primarily of various *Saccharomyces cerevisiae* varieties. The content in higher alcohols (mainly isobutanol, isoamyl alcohols, and phenethyl alcohol) at the end of fermentation ranges from 207 to $405\,mg\,l^{-1}$. Compared with other types of white wines, the ester fraction is less concentrated, and so the wine has a stronger sensory profile prior to aging. This is also a result of a high fermentation temperature (frequently above 25 °C), which favors the formation of higher alcohols. Pure and mixed cultures of the yeasts have been tested with a view to softening the sensory profile of the base wine. In this sense, some authors have studied the enzymatic activity of acetyl transferase and esterase in pure cultures of the fermentation yeasts in order to determine the contribution of specific yeast varieties involved in the alcoholic fermentation process to the flavor of the base wine. However, the need to ferment large amounts of sugars, as a result of a high tolerance to ethanol, makes replacing the wild yeasts very difficult.

Wine Aging

Most of the analytical and sensory differences among sherry wines result from differences in the aging procedures to which the base wines are subjected. On the one hand, the biological aging process that yields fino wine is affected by flor yeasts, which use some compounds in the base wine as carbon sources, thereby altering its composition to a variable extent depending on the duration of the process. On the other hand, the oxidative aging process that produces oloroso wine allows for no biological transformation, so the changes involved should be preferentially ascribed to evaporation, chemical oxidation, and extraction from constituents of the wood casks. Because amontillado wine is obtained by a combination of both types of aging procedure, its analytical properties are inherited to a variable extent, even though its sensory profile is unique to this wine type.

More than 300 different compounds have been identified in sherry wine-type fino. Only a small fraction, however, can be specific to them, with most of the analytical differences observed being ascribed to changes in the concentrations of compounds also present in other types of white wines. At this point, the contents of acetaldehyde and derivatives such as 1,1-diethoxyethane and acetoin are noteworthy. The concentrations of these compounds increase during biological aging as a result of the aerobic metabolism of flor yeasts consuming the ethanol of the wine. Their concentrations in commercially available

wines are in the range of 400–500 mg l^{-1} for acetaldehyde, 40–75 mg l^{-1} for 1,1-diethoxyethane, and 8–19 mg l^{-1} for acetoin, depending on the duration of the aging process. Acetaldehyde is thus used as a measure of wine maturity, being largely responsible for the typical pungent flavor of fino wines. Some lactones and their oxygen derivatives also impart different notes to the flavor of wines obtained by biological aging. In fact, despite their low concentrations, some lactones such as sotolone (0.036–0.143 mg l^{-1}), γ-butyrolactone (27–54 mg l^{-1}) and pantolactone (3–8.5 mg l^{-1}) impart a caramel-like odor to these wines and increase their concentration during biological aging. In particular, the former is considered an important contributor to the flavor of fino wines, introducing nut-curry notes. One other distinctive feature of the biological aging process is the decrease in volatile acidity resulting largely from the consumption of acetic acid by flor yeasts, which reduces its typical odor and flavor.

Changes in other aroma compounds are not as clear as those in the aforementioned compounds. Although some authors have reported a decrease in the contents of major higher alcohols, others have found slightly increased or constant contents. These contradictory results are probably a consequence of differences in the experimental conditions used, particularly as regards the distribution of the flor yeast population and the aging temperature. In relation to the ester, changes during aging can also be observed giving contradictory results, although increased contents in ethyl lactate and diethyl succinate, and decreased contents in ethyl, isoamyl, isobutyl, and phenethyl acetates, appear to be the norm. Some authors have found that flor yeasts synthesize small amounts of terpenes such as linalool, α-terpineol and z-nerolidol.

The aroma fraction of oloroso and amontillado wines is mainly influenced by a concentrating effect during their lengthy aging; as a result, the contents in major higher alcohols and esters increase with increasing wine age. Also, chemical esterification is favored by a high ethanol content in the wine, so the concentrations of esters such as ethyl acetate, ethyl lactate, and diethyl succinate increase during aging.

Phenolic Compounds and Color

Color is the most obvious distinctive feature of sherry wines for the consumer. Thus, fino wines, which are obtained by biological aging, exhibit a very pale yellow color, whereas amontillado and, especially, oloroso wines, have a dark color.

Musts and Wines

More than 20 phenolic compounds in monomeric and simple oligomeric forms have been identified in base wines prior to aging. Although their concentrations depend markedly on the particular climatic conditions of each year and on the grape ripening status, flavan-3-ol derivatives invariably exhibit the highest concentrations of all (100–120 mg l^{-1}, most of which is contributed by catechin and epicatechin). The concentrations of hydroxycinnamic acids range from 1 to 5 mg l^{-1}, and those of their esters from 45 to 75 mg l^{-1}, in which *t*-caftaric acid is the best represented. Gallic acid exhibits the highest contents among hydroxybenzoic acids (the overall concentration ranging from 15 to 25 mg l^{-1}). Flavonols and coumarins are also present, at concentrations of 10–20 and about 50 mg l^{-1}, respectively.

Fino wine is a delicate product and it tends to get browner after bottling. This alters not only its color but also its sensory properties. Thus, fresh fino wine has an absorbance at 420 nm of 0.100–0.120 A.U. At 0.170–180 A.U., it is normally discarded. Depending on the particular production technique used, this absorbance threshold is reached within about a year after bottling. This has promoted the development of improved production techniques to balance supply and demand better in order to avoid overstocking. Because brown pigments in the wine are formed by oxidation of phenolic compounds, winemaking techniques such as must hyperoxidation have been used to decrease the contents in these compounds and increase the resistance of the resulting wines to browning. However, one should bear in mind the complexity of the aging process, by which a given cask contains mixed wines from different vintages. Reliable conclusions on this point can only be reached from a 'solera and criadera system' exclusively using hyperoxidized wines from various vintages, requiring a long-term research. Also in order to improve the resistance to browning, traditional practices such as the addition of SO_2 to grapes during the harvest period have been suppressed, because these result in an increased degree of extraction of phenolic compounds from the grape skin.

Wine Aging

Because of their differential aging processes, the three types of sherry wine differ in their contents in phenolic compounds. Thus, flavan-3-ol derivatives, predominate in fino wines (with concentrations of 60–80 mg l^{-1}), followed by hydroxycinnamic esters (15–25 mg l^{-1}), hydroxybenzoic acids (15–30 mg l^{-1}), flavonols (5–10 mg l^{-1}), and hydroxycinnamic acids

(1.5–2.5 mg l^{-1}). The contents in vanillic and ferulic acid, catechin, epicatechin, and procyanidin B2 and B4, have been found to decrease during biological aging, whereas those of syringic acid and procyanidin B1 appear to increase during the process. It should be noted that this type of wine does not brown under flor yeasts, thus preserving its pale color with slight variations for years. This has traditionally been ascribed to the flor yeasts that grow on its surface, making the diffusion of atmospheric oxygen across the flor film difficult and thus protecting wine from this. In addition, flor yeasts consume dissolved oxygen in the wine to maintain their aerobic metabolism, thereby preventing the oxidation of phenolic compounds. However, the only contribution of this protective effect to color stability in the wine subjected to biological aging has been partly questioned by some authors who believe that the wine is partially oxygenated over the years as a result of two actions. On the one hand, older wine is mixed with younger wine three or four times a year, which introduces some oxygen during the transfer operations. On the other hand, the flor film fails in some seasons because of its strong dependence on temperature. The lack of the flor film facilitates the access to oxygen. In fact, the absorbance values at 420 nm measured at the end of summer are normally higher than those recorded in late spring and autumn. Thereby, in addition to the aforementioned protective effect of the flor film, some authors think that yeasts may be able to retain brown polymers formed by oxidation of phenols.

Oloroso wines are considerably darker than fino wines as a result of their oxidative aging, showing absorbance values at 420 nm, ranging from 1.30 to 1.60 A.U.

The phenolic contents of these wines exceed those in fino wines, with levels of 35–45 mg l^{-1} for hydroxybenzoic acids, 5–20 mg l^{-1} for hydroxycinnamic acids, 20–25 mg l^{-1} for hydroxycinnamic esters, and 100–130 mg l^{-1} for flavanes. Oxidative aging usually takes place at a higher temperature than does biological aging, which results in volume losses in the aging wine, with a subsequent concentrating effect on the soluble compounds.

Amontillado wines show intermediate levels in between the previous two because of the combined aging process used in their production, so their corrected absorbances at 420 nm range from 0.950 to 1.100 A.U. However, in general, their phenol contents are similar to those of oloroso wines, with values of 45–75 mg l^{-1} for hydroxybenzoic acids, 5–15 mg l^{-1} for hydroxycinnamic acids, and 125–160 mg l^{-1} for flavanes.

Other Compounds

Amino acids are important compounds because they are used by the flor yeast as a nitrogen source. L-Proline is the predominant compound of this fraction in the wine base, accounting for about 60% of total nitrogen, with a concentration of 4.5–7.5 mmol l^{-1}. Other abundant amino acids include L-tryptophan (8%), L-phenylalanine (7.5%), L-cysteine (6%), and L-threonine (3%). L-Proline is the most heavily used nitrogen-containing compound and the principal source of nitrogen for the flor yeast, and so this compound is consumed most during biological aging. In the deficiency of L-proline, flor yeasts use other amino acids such as L-glutamic acid, L-alanine, L-arginine, and L-tryptophan. Oloroso wines show slightly increased amounts of amino acids as a result of the aforementioned effect of volume concentration during oxidative aging.

The most abundant polyalcohols quantified in sherry wines are inositol and erythrol, with variable levels depending on the method of wine-making. No significant changes have been observed during aging, because the initial variability of the wine base is greater than the contribution of the maturation processes. However, there is a general trend towards higher contents in these compounds for the oloroso and amontillado wines.

See also: **Amino Acids**: Properties and Occurrence; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Oxidation of Food Components**; **Phenolic Compounds**; **Sherry**: The Product and its Manufacture; **Yeasts**

Further Reading

Baron R, Mayen M, Merida J and Medina M (1997) Changes in phenolic compounds and browning during biological aging of Sherry type wine. *Journal of Agricultural and Food Chemistry* 45: 1682–1685.

Cortes MB, Moreno JJ, Zea L, Moyano L and Medina M (1999) Response of the aroma fraction in sherry wines subjected to accelerated biological aging. *Journal of Agricultural and Food Chemistry* 47: 3297–3302.

Domecq P (1989) Sherry: state of art on a very special fermentation product. In: *Proceedings of the International Symposium on Yeasts (Issy XIII)*, pp. 15–35. Leuven: Leuven Press.

Fabios M, Lopez-Toledano A, Mayen M, Merida J and Medina M (2000) Phenolic compounds and browning in sherry wines subjected to oxidative and biological aging. *Journal of Agricultural and Food Chemistry* 48: 2155–2159.

García-Barroso C, López Sánchez L, Otero JC, Cela-Torrijos R and Pérez-Bustamante JA (1989) Studies on the

browning of the fino sherry and its relation to polyphenolic compounds. *Zeitschrift für Lebensmittel-unterschung und -forschung* 189: 322–325.

Guichard E, Etievant P, Henry R and Mosandl A (1992) Enantiomeric ratios of pantolactone, solerone, 4-carboethoxy-4-hydroxy-butyrolactone and of sotolon, a flavor impact compound of flor-sherry and botrytized wines. *Zeitschrift für Lebensmitteluntersuchung und -forschung* 195: 540–544.

Ibeas JI, Lozano I, Perdigones F and Jimenez J (1997) Effects of ethanol and temperature on the biological aging of sherry wines. *American Journal of Enology and Viticulture* 48: 71–74.

Martin B, Etievant PX, Lequere JL and Schlich P (1992) More clues about sensory impact of sotolon in some flor sherry wines. *Journal of Agricultural and Food Chemistry* 40: 475–478.

Martinez P, Rodriguez LP and Benitez T (1997a) Evolution of flor yeast population during the biological aging of fino sherry wine. *American Journal of Enology and Viticulture* 48: 160–168.

Martinez P, Rodriguez LP and Benitez T (1997b) Velum formation by flor yeasts isolated from sherry wine. *American Journal of Enology and Viticulture* 48: 55–62.

Mauricio JC, Moreno J and Ortega JM (1997) In vitro specific activities of alcohol and aldehyde dehydrogenases from 2 flor yeasts during controlled wine aging. *Journal of Agricultural and Food Chemistry* 45: 1967–1971.

Mesa JJ, Infante JJ, Rebordinos L and Cantoral JM (1999) Characterization of yeasts involved in the biological aging of sherry wines. *Lebensmittel-Wissenschaft und Technologie* 32: 114–120.

Moyano L, Moreno J, Millan C and Medina M (1994) Flavour in Pedro Ximénez grape musts subjected to maceration processes. *Vitis* 33: 87–91.

Webb AD and Noble AC (1976) Aroma of sherry wines. *Biotechnology and Bioengineering* XVIII: 939–1052.

SHIGELLA

K A Lampel, US Food and Drug Administration, Laurel, MD, USA
R Sandlin, Towson University, Towson, MD, USA

Characteristics

Shigella spp. are the causative agents of shigellosis (bacillary dysentery). These organisms are endemic worldwide, resulting in high morbidity and mortality in third world countries. *Shigella* are small, slender, Gram-negative rod-shaped bacteria that are nonmotile and nonspore-forming. The genus is a member of the family *Enterobacteriaceae* and is closely related to *Escherichia coli*. *Shigella* spp. are divided into four major groups based on biochemical reactions and the O antigen type (**Table 1**). The four species are *S. dysenteriae* (A), *S. flexneri* (B), *S. boydii* (C), and *S. sonnei* (D), and are further subdivided into serotypes (except *S. sonnei*), based on differences in the O antigen. The severity of disease varies, depending on the particular species, from mild diarrhea to bacillary dysentery (stools containing blood, mucous, and inflammatory cells). *S. sonnei* produces the mildest form of disease, whereas *S. dysenteriae* produces the most severe. *S. flexneri* and *S. boydii* cause intermediate forms of the disease. Additionally, the geographic distribution and epidemiology produced by the different species are different (**Table 1**). For example, *S. dysenteriae* is responsible for epidemics of the disease, whereas shigellosis outbreaks in industrialized nations are usually caused by *S. sonnei*. In addition, *S. flexneri* is more frequently found in developing countries, and *S. boydii* is usually limited to the Indian subcontinent.

Shigella spp. are facultative anaerobes with relatively simple nutritional requirements. Typically, *Shigella* spp. are lysine decarboxylase negative and lactose nonfermenting (except for a slow reaction with *S. sonnei*), characteristics that separate them from closely related commensal *E. coli*. However, enteroinvasive *E. coli* (EIEC) share many properties with shigellae producing clinical symptoms indistinguishable from shigellosis. Moreover, EIEC has similar biochemical traits (e.g., nonmotile, lactose negative) and possess immunologically cross-reactive somatic antigens with several *Shigella* serotypes. Thus, similarities between EIEC and *Shigella* can complicate the accurate diagnosis of shigellosis.

The World Health Organization has estimated that worldwide, diarrheal diseases account for approximately 5 million deaths each year among children with nearly 10% caused by *Shigella* spp. *S. dysenteriae* type I is the only *Shigella* spp. known to produce shiga toxins, a common etiological agent of epidemic dysentery. *S. dysenteriae* has also been shown to persist in the poorest regions of the world, notably in parts of Africa, Central America and Asia, where malnutrition is common. In the USA, of the estimated 76 million people that contract foodborne illness

Table 1 Characteristics of *Shigella* spp.

Species	*Serogroup*	*Serotypes*	*Geographic distribution*	*Distinguishing characteristics*
S. dysenteriae	A	15	Indian subcontinent, Africa, Asia, Central America	Produce Shiga toxin, causes most severe dysentery, high mortality rate if untreated
S. flexneri	B	6	Most common isolate in developing countries	Less severe dysentery
S. boydii	C	19	Indian subcontinent, rarely isolated in developed countries	Biochemically identical to *S. flexneri*, distinguished by serology
S. sonnei	D	1[a]	Most common isolate in developed countries	Mildest form of shigellosis

[a]Forms I (smooth) and II (rough) are serotypically distinguishable.

annually, an estimated 300 000 cases are caused by *Shigella*. In a recent report from the US Center for Disease Control and Prevention, Emerging Infections Program Foodborne Diseases Active Surveillance Network (FoodNet; http://www.cdc.gov/foodnet), 2324 of the 12 631 confirmed cases of known foodborne illnesses reported (only nine diseases were followed), were caused by *Shigella* spp., which ranked *Shigella* as the third leading cause of diseases of bacterial origin. As expected, most of the reported *Shigella* cases were caused by *S. sonnei* (85%) and the remainder by *S. flexneri* (15%).

Shigella is principally a human disease, although outbreaks have been described in subhuman primates. One distinctive characteristic of the disease is its high communicability attributed to the low infectious dose required to cause illness. For example, ingestion of fewer than 100 organisms has been shown to result in clinical disease in volunteer studies. Thus, the low infectious dose is the primary reason cited for ease of *Shigella* spread in populations living under crowded and unsanitary conditions. The incubation period for *Shigella* is 1–7 days, with the onset of symptoms usually appearing in 3 days. The clinical presentation of shigellosis can vary from mild diarrhea to severe dysentery with symptoms including fever, cramps, and frequent bowel movements containing blood, pus, and mucus. Infections are associated with mucosal ulceration, rectal bleeding, and dehydration. Polymorphonuclear leukocytes in the stools is one characteristic feature of the inflammatory nature of dysentery. Mortality has been reported as high as 10–15% in some strains, but the infection is usually self-limiting and resolves without treatment within 1–2 weeks. Complications of shigellosis include reactive arthritis, and hemolytic uremic syndrome.

Asymptomatic carriers of shigellae may be partially responsible for the perpetuation of shigellosis outbreaks. Owing to socioeconomic factors that affect both living and sanitary conditions, adults, particularly those that have resided in one area for a long period of time, are thought to exacerbate the spread and maintenance of this pathogen. Furthermore, shigellosis appears to have the highest incidence in summer months, presumably because of increased contact of asymptomatic carriers with uninfected populations. Therefore, asymptomatic carriers may play a significant role in incubating this pathogen through the colder months.

Unlike most pathogens, *Shigella* spp. are not indigenous to any food. Most cases of illness are a result of close contact with infected individuals, with transmission most frequently occurring via the fecal-to-oral route. The organism is commonly found in water contaminated with human feces and can be spread via food, flies, fingers, and feces. In highly endemic areas, shigellosis is associated with children who frequently lack adequate personal hygiene habits. Children, ages 1–6, are most susceptible to infection by *Shigella* spp. The carrier state usually lasts 1–4 weeks, but long-term carriers have been described and are thought to be reservoirs of infection. It has been estimated that during the acute phase of the disease, 10^3–10^9 colony-forming units (CFU) per gram of stool are shed, whereas in convalescing patients, 10^2–10^3 CFU per gram of stool are shed. To control the spread of shigellae, infected individuals are monitored for the presence of the organism in stool until cleared.

Virulence Factors

Members of the genus *Shigella* are facultative intracellular pathogens. Following ingestion, the organisms are capable of surviving the acidic barrier of the stomach in order to reach the large intestine. Shigellosis is caused when virulent *Shigella* organisms induce their uptake into M cells and epithelial cells of the large intestine. After invasion, the bacteria are released from the phagosome into the cytoplasm and multiply intracellularly utilizing the host cell cytoplasm for nutrients. The infection is perpetuated by the spread of the organism to neighboring epithelial cells through projections from the originally infected cell. This cycle results in damage and destruction of the intestinal epithelial cell, resulting in

dysentery. *Shigella* spp. are also able to induce apoptosis or programmed cell death of macrophages. Virulence factors that contribute to these processes are both chromosomally encoded and present on a large 220-kb virulence plasmid (Table 2). The nucleotide sequence of the entire plasmid was recently reported (GenBank Accession No. AL391753). The virulence plasmid is essential for many virulence phenotypes, and in the absence of the plasmid, the organism is avirulent. A related virulence plasmid is also present in EIEC and is essential for this organism's ability to cause disease. The genetic factors involved in the virulence phenotype can be loosely grouped into genes that encode proteins involved in epithelial cell uptake, genes that encode proteins involved in the intercellular spreading phenotype, and genes involved in the regulation of these processes.

The expression of the virulence genes is regulated by a variety of factors, most notably temperature. Temperature is an important factor in the regulation of virulence genes in bacterial pathogens, including *Salmonella typhimurium, Bordetella pertussis, Yersinia* spp., and *Listeria monocytogenes*, as well as *Shigella*. Virulence gene expression in *Shigella* is repressed at 30 °C and induced at 37 °C by the action of three gene products. The virulence plasmid encoded VirF is a transcriptional activator that is a member of the AraC family of activators. VirF functions to upregulate expression of VirB. VirB is encoded on the virulence plasmid and is thought to induce the expression of a variety of essential virulence genes including the *ipa*, *spa*, and *mxi* genes. VirR (H-NS) is a chromosomally encoded global negative regulator that is involved in a variety of regulatory pathways. H-NS acts to block VirF activity at the *VirB* promoter. The exact mechanism is unknown, but several factors, including temperature, affect regulation at both the transcriptional and translational level. Mutations in H-NS result in an absence of temperature regulation and a constitutive virulent phenotype, whereas mutations in *VirF* or *VirB* result in an inability to produce the Ipa, Spa, and Mxi proteins and an avirulent phenotype.

Upon contact with a target epithelial cell, *Shigella* spp. are induced to secrete up to 25 proteins, including the Ipa B, C, D, and A proteins. Some of these proteins, including IpaC and IpaA, are translocated directly into the host cell cytoplasm. The Ipa proteins

Table 2 Virulence loci of *Shigella*

Locus	*Protein*	*Role in virulence*
Plasmid-encoded genes		
ipaA	70-kDa protein	Invasion; associates with vinculin
ipaB	62-kDa protein	Invasion; lysis of vacuole; induction of apoptosis
ipaC	43-kDa protein	Invasion; induces formation of filopodia and lamellipodial extensions
ipaD	38-kDa protein complex with IpaB	Invasion; forms antisecretion
ipgC	17-kDa protein	Chaperon for IpaB and IpaC
ipgD	59.8-kDa protein	Modulates invasion of bacteria into epithelial cells
ipaH	Family of proteins	Present in plasmid and chromosome
mxi/spa	20 proteins	Secretion of Ipa and other virulence proteins
icsA (virG)	120-kDa protein	Actin polymerization for intracellular motility and intercellular spread
sen	ShET2	Enterotoxin
virB	Transcriptional activator	Temperature regulation of virulence genes
virF	Transcriptional activator	Temperature regulation of virulence genes
osp	Outer *Shigella* proteins	Encoded proteins secreted by type III secretory system
Chromosomal-encoded virulence-associated genes		
virR (hns)	Histone-like protein	Repressor of virulence gene expression
rfa; rfb	Enzymes for LPS core and O-antigen biosynthesis	Unipolar localization of IcsA
stx[a]	Shiga toxin	Destruction of vascular tissue
vacB (rnr)	Exoribonuclease RnaseR	Posttranscriptional regulation of virulence gene expression
cpxR	Response regulator of *virF*	CpxA–CpxR two-component system
luc	Aerobactin and receptor	Acquisition of iron in the host
SodB	Superoxide dismutase	Defense against oxygen-dependent killing in host
set [b]	ShET1	Enterotoxin
dsbA	Disulfide bond catalyst	Facilitates secretion of IpaB and IpaC
sigA	139.6-kDa exported cytopathic protease	Intestinal fluid accumulation

[a]The *stx* locus and production of Shiga toxin are observed only in *S. dysenteriae* 1.
[b]The *set* locus and production of ShET1 is observed almost exclusively in *S. flexneri*.

are encoded on the virulence plasmid by the *ipaBCDA* operon and are recognized by convalescent patient sera. Ipa B, C, and D are essential for invasion and are thought to be involved in signaling the uptake of the bacteria by the host cell. Additionally, IpaB is hypothesized to be responsible for release from the phagocytic uptake vacuole and for induction of apoptosis in infected macrophages. The Ipas are secreted by a dedicated virulence protein secretion apparatus known as a Type III secretory system. Related secretion systems have been described for a variety of plant and animal pathogens and share both functional and sequence homologs. The group of proteins that compose this secretion apparatus are encoded by the *mxi/spa* loci present on the virulence plasmid. These proteins have been shown to comprise a needle complex with both cytoplasmic and transmembrane domains that span the inner and outer membrane of the bacterial cell to allow the release of the Ipa proteins as well as other secreted proteins. Mutation in any of these structural components of the secretory apparatus results in a loss of virulence, owing to the inability to secrete the Ipa proteins. The *icsA* (*virG*) locus on the virulence plasmid is another essential virulence factor. This outer membrane protein is asymmetrically localized at the old pole of the bacterial cell, where it catalyzes the polymerization of host cell actin. This results in the formation of an actin tail at one end of the bacterium that propels the organism through the host cell and allows the bacterium to enter the neighboring cell without leaving the initially infected host cell and encountering the exterior environment. Functional homologs of IcsA have been described in *L. monocytogenes, Rickettsia*, and *Vaccinia*.

In addition to the mucosal damage that results from the process of epithelial cell invasion and intracellular multiplication, shigellae also produce enterotoxins that are presumably responsible for the diarrheal component of the disease. One serotype, *S. dysenteriae*, elaborates a potent toxin (Shiga toxin) that inhibits protein synthesis and exhibits both cytotoxic and enterotoxic activity. Some *E. coli* strains also produce Shiga-like toxins and have been associated with outbreaks spread by contaminated meats and sprouts. The Shiga toxin is produced only in *S. dysenteriae*, which causes the most severe form of dysentery, and may contribute significantly to the severity of the disease.

Because the virulence plasmid-encoded genes are essential to the disease process, the ability of the pathogen to maintain the presence of the plasmid is critical for pathogenesis. It has been shown that the expression of the plasmid-encoded virulence genes destabilizes the maintenance of the plasmid, presumably owing to the high metabolic demand of synthesizing the virulence proteins. The virulence plasmid is lost from the bacterial cell at a higher frequency at 37 °C when virulence protein expression is upregulated than when the bacteria are grown under repressing conditions at 30 °C. This observation could affect the detection of the organisms from clinical, food, and environmental samples using DNA-based or serological assays.

Shigella in Foods

Regulatory Issues

Unlike other bacterial foodborne pathogens, such as *Salmonella* spp., *Shigella* spp. are not associated with any specific food. The contamination of foods with *Shigella* usually results from an infected food handler or from contaminated water, either from irrigation or after processing. In the former case, the poor personal hygiene of one or more individuals can lead to diarrheal diseases. Each year, there are a significant number of outbreaks of shigellosis from consumption of contaminated foods. Many of these outbreaks are confined to single countries, but international travel and global commerce have been influential factors in the spread of the disease. For example, in 1994, contaminated lettuce from one country was the source of an outbreak of shigellosis that occurred in several European countries.

Many shigellosis outbreaks have been traced to the consumption of raw or fresh vegetables, particularly in salads, seafood, bakery products, chicken, and hamburger. In response to the increasing number of foodborne outbreaks resulting from consumption of contaminated fresh produce, the US Food and Drug Administration published a document entitled 'Guidance for Industry-Guide to Minimize Microbial Food Safety Hazards for Fresh Fruits and Vegetables' (http://www.foodsafety.gov/~dms/prodguid.html). These voluntary guidelines address concerns in the following areas: agricultural and packinghouse water use, manure management, worker hygiene, field and packinghouse sanitation, and transportation. The aim of these recommendations is to reduce transmission of shigellae and other enteric pathogens via fresh vegetables and fruits.

Another factor in the spread of *Shigella* sp. is improper storage temperature of contaminated foods. For the consumer, steps can be taken to reduce risk in contracting shigellosis: purchase food at reliable sources, wash food adequately, store food at suitable (refrigerated) temperatures, cook food adequately, refrigerate left-overs promptly, and wash cooking utensils and equipment adequately. The transmission

of *Shigella* spp. via foods is very efficient, and because of its low infectious dose, action to remove contaminated food from commerce should be swift.

Probably the most effective means to reduce the spread of many bacterial and parasitic pathogens is proper handwashing. Many agencies, including the World Health Organization, have emphasized adequate hand washing before and during meal preparation by those that work in food establishments and the consumer as an effective means in reducing foodborne outbreaks.

To address the concerns of the introduction of pathogens during food processing, a program, Hazard Analysis and Critical Control Point (HACCP), was established to identify and control specific food-related practices. This analysis determines where tests should be implemented to assay for the presence of foodborne pathogens at critical stages of food production. HACCP is an alternative approach to the traditional method of analyzing the final food product. Although HACCP may be a good monitoring system for other pathogens commonly known to be associated with foods, such as *Salmonella* spp. and *E. coli* O157:H7, testing for the presence of shigellae in foods at this point may not be useful. For example, contamination of foods with *Shigella* spp. commonly occurs between the processing plant and the consumer. However, proper adherence to good manufacturing process, such as hand washing and temporary removal of employees with diarrheal illness, can reduce outbreaks of shigellosis as well as other forms of foodborne diseases. As an example, at a mass gathering, a large outbreak occurred with nearly half of the attendees (3000 people) developing shigellosis from consuming meals prepared where one food handler was found to be ill.

Survival

Since shigellae are not indigenous to any particular food, this pathogen may be found in any food matrix. In the past, foods containing contaminated raw vegetables have been implicated in many shigellosis outbreaks. Other foods that are common vehicles for the spread of shigellae include tossed salads, potato salads, chicken, and shellfish. In a recent study, the survivability of *Shigella* in packaged vegetables (sterile and unsterile) held at different temperatures (5 °C, 10 °C, and room temperature) was reported. The greatest reduction in number of shigellae recovered was from the first 24 h. However, the data indicate that shigellae do survive in the vegetables tested, albeit after a 3 to 7 log reduction, for at least 10 days (although the initial inoculum was high (10^{11}), the final numbers ($CFU\,g^{-1}$) were between 10^4 and 10^8).

In laboratory conditions, *Shigella* spp. can grow at temperatures between 6 °C and 48 °C and at a pH of between 4.8 and 9.3. In addition, *Shigella* can survive at room temperature for up to 50 days in foods such as milk, flour, eggs, clams, shrimp, and oysters, and only 5–10 days in acidic foods (e.g., orange juice, tomato juice, carbonated soft drinks) and 1–2 weeks in refrigerated, fermented milk. It should be noted that although shigellae do not survive well in acidic foods, these organisms are able to survive the acidic conditions of the stomach.

Detection of Shigella from Foods

Foods are not routinely examined for the presence of shigellae. An outbreak of shigellosis is usually identified initially from clinical findings and epidemiological investigations. Therefore, 7–10 days may have passed before the first individual becomes ill. During this time, the fate and condition of the contaminated food is unknown, presenting a challenge to the laboratory responsible for isolating the etiological agent. Moreover, the need to rapidly isolate and identify shigellae in foods is necessary to reduce the spread of this pathogen.

Bacteriological scheme The Bacteriological Analytical Manual describes one protocol to isolate *Shigella* spp. from foods. Twenty-five-gram sample portions are added to 225 ml of *Shigella* broth supplemented with novobiocin (0.5 $\mu g\,ml^{-1}$ for *S. sonnei*; 3 $\mu g\,ml^{-1}$ for other shigellae; this antibiotic is added to suppress the natural bacterial flora of the food) and held for at least 10 minutes at room temperature. The broth is contained in a 500-ml flask and incubated overnight at 37 °C with shaking. Alternative protocols use other enrichment broths, such as Gram-negative broth, where food samples are blended in a sterile stomacher bag, the contents poured into a flask and incubated at 37 °C with shaking for 16–20 h.

Recovery of injured cells owing to a lengthy time delay on foods and/or storage conditions before processing a sample for analysis is problematic. Bile salts and desoxycholate have an adverse affect on growth of impaired cells; therefore, selective media without these compounds are recommended. Food samples are added to 100 ml of tryptic soy broth, the pH adjusted to 7.0, and then blended. After 8 h of incubation at 37 °C, 125 ml of enrichment broth are added and the incubation extended for an additional 16–20 h.

After enrichment in broth, cultures can be plated on selective (low to high) agar medium. To optimize the recovery of shigellae from foods, two or three different ranges of selective agar plates should be used (see Table 3). MacConkey agar is a low-selectivity

Table 3 Growth characteristics of *Shigella*, *E. coli*, and *Salmonella* on selective media

Organism	*Medium*	*Colony appearance*
Shigella	MacConkey	Colorless (lactose nonfermentor); *S. sonnei* colonies are colorless to pink[a], translucent, flat with jagged edges
	Hektoen	Green and moist
	XLD	Colorless
E. coli	MacConkey	Lactose fermentor; flat, pink colonies surrounded by darker pink region (indicates sorbitol fermentors, nonsorbitol fermentors form colorless colonies)
	Hektoen	Yellow/yellow orange, salmon
	XLD	Yellow
Salmonella	MacConkey	Colorless (lactose nonfermentors)
	Hektoen	Green
	XLD	Red with black center

[a]*S. sonnei* may ferment lactose slowly (> 40 h). Form I vs. Form II; smooth to rough look as a result of the loss of a 120-MDa virulence plasmid.

medium. *Shigella* produce colonies that are translucent and slightly pink (*Shigella* are lactose-negative) with or without rough edges. Eosin methylene blue (EMB) and tergitol-7 agar are alternative low-selectivity agars containing lactose. Colorless colonies on eosin methylene blue plates or bluish colonies on the yellow–green tergitol-7 agar are indicative of *Shigella*. The preferred media for *Shigella* isolation are desoxycholate and xylose–lysine–desoxycholate (XLD) agars, intermediate selective media. *Shigella* colonies on XLD agar medium are translucent and red (alkaline), whereas on desoxycholate medium, *Shigella* produce reddish colonies. Although most *Shigella* spp. do not ferment xylose, some species, e.g., *S. boydii* (variable) may be missed, and therefore, plating on XLD and MacConkey agar plates is recommended. Highly selective media include *Salmonella–Shigella* and Hektoen agars, media designed to differentiate and isolate *Salmonella* and *Shigella* spp. from other enteric bacteria. However, some *Shigella* spp., such as *S. dysenteriae* type I, are unable to grow on highly selective *Salmonella–Shigella* medium. *Shigella* produce colorless, translucent colonies on this agar. Colonies of *Shigella* on Hektoen agar appear to be green and moist, as do colonies from *Salmonella* spp. (some are blue green with or without black centers); *E. coli* strains, and other coliforms, such as *Klebsiella* and *Enterobacter*, form yellow, yellow–orange or salmon colonies. All presumptive colonies are inoculated into semisolid (motility) test agar for further testing; *Shigella* spp. are nonmotile.

Biochemical tests are used to further identify *Shigella* spp. Suspected colonies that are Gram-negative, nonmotile rods are inoculated on to lysine iron, Kliger iron, or triple sugar iron agar. On triple sugar iron agar slants, *Shigella* typically produce acid from glucose utilization (yellow butt) with no gas production, and because they do not utilize lactose or sucrose, the slant retains its color (red); no black color is produced in the butt, indicating the absence of H_2S (Table 4). Further biochemical characterizations show that *Shigella* spp. are negative for phenylalanine deaminase, sucrose, and lactose fermentation, do not utilize citrate, acetate, KCN, malonate, inositol, adonitol, and salicin, and lack lysine decarboxylase. Shigellae are negative for the Voges–Proskauer test (*S. sonnei* and *S. boydii* serotype 13 are positive). However, all shigellae are methyl red-positive and are unable to produce acid from glucose and other carbohydrates (acid and gas production occur with *S. flexneri* serotype 6, *S. boydii* serotypes 13 and 14, and *S. dysenteriae* 3). *S. dysenteriae* type I is catalase-negative, and *S. sonnei* has ornithine decarboxylase activity.

Table 4 Tests to differentiate *Shigella* spp. from *E. coli*

Test	*Reaction*	
	Shigella	*E. coli*
Motility	−	+[a]
Gas from glucose	−[b]	+[c]
Lysine decarboxylase	−	+[d]
Christensen's citrate	−	+
Acetate	−	+
Mucate	−	+

[a]Most positive, some negative (enteroinvasive *E. coli* are nonmotile).
[b]Some strains of *S. flexneri* 6 produce small amounts of gas from glucose.
[c]Some exceptions.
[d]Enteroinvasive *E. coli* are also negative.

Table 4 and 5 show key biochemical reactions to differentiate *Shigella* from *E. coli* and also to distinguish each *Shigella* sp. from each other. Growth on Christensen citrate, sodium mucate, or acetate agar is one characteristic that discriminates between *E. coli* and *Shigella*; shigellae are unable to utilize citrate, acetate, or mucate as sole carbon source.

Other biochemical tests are used to identify the serotypes of *Shigella*. The ability to utilize mannitol, dulcitol, xylose, rhamnose, raffinose, glycerol, and indole and the presence of ornithine decarboxylase

Table 5 Biochemical differentiation of *Shigella* species

Test	*S. dysenteriae*	*S. flexneri*	*S. boydii*	*S. sonnei*
β-galactosidase	−[a]	−	−	+
Ornithine decarboxylase	−	−	−	+
Indole production	+/−[b, c]	+/−[c,d]	+/−[c]	−
Gas from glucose	−	−[e]	−[f]	−
Acid from				
Dulcitol	−[g]	−[h]	−	−
Lactose	−	−	−[i]	+[j]
Mannitol	−	+	+	+
Raffinose	−	+/−[c]	−	+[j]
Sucrose	−	−	−	+[j]
Xylose	−	−	+/−[c]	−

[a] *S. dysenteriae* type 1 are positive.
[b] *S. dysenteriae* type 1 is negative; *S. dysenteriae* type 2 is positive.
[c] Reaction is variable.
[d] *S. flexneri* 6 is negative.
[e] *S. flexneri* 6 is positive.
[f] *S. boydii* serovars 13 and 14 produce acid and gas.
[g] *S. dysenteriae* type 1 may be positive.
[h] *S. flexneri* 6 may be positive.
[i] *S. boydii* may be positive.
[j] Positive reactions may take 24 h or longer.
+/−, variable reaction.

have been used to discriminate physiologically between *Shigella* spp.

Serological testing, using polyvalent antiserum, is used to identify the *Shigella* group (A–D). A note of caution should be addressed, concerning EIEC. This pathogen causes the same disease, bacillary dysentery, as do the shigellae. In some cases, some EIEC strains share homology with antigenic structures of some *Shigella* serotypes. Several serotypes of *S. dysenteriae*, *S. flexneri*, and *S. boydii* have reciprocal cross-reactivity with *E. coli* O antigens of the Alkalescens-Dispar bioserogroup or EIEC.

DNA-based methods Initially, DNA probes were developed to detect *Shigella* spp. from clinical samples, a far easier task than from foods. In the former, shigellae are shed in large numbers (10^3–10^9 per gram of stool) in patients in the acute phase of infection, and 10^2–$10^3\,g^{-1}$ can be found in convalescent individuals. Identification and detection of shigellae from stool samples is relatively easy, owing to the high number of organisms in a relatively homogeneous sample. The analysis of foods is obviously much more complicated. The diversity of foods, such as composition (pH, fat content), microbial content of the natural flora, and the length of time from collecting food samples, as in the case of a shigellosis outbreak, to the laboratory are factors leading to a diminishing chance of isolating or detecting this pathogen.

The application of the polymerase chain reaction (PCR), as with DNA probes, was initially developed to detect *Shigella* from clinical samples. PCR primers are selected that target specific virulence genes. For detecting shigellae, primers to amplify a segment of the *ipaH* gene have an advantage over other virulence genes. The *ipaH* gene is present in multiple copies in *Shigella*, including the bacterial chromosome. A distinct advantage of this set of primers is that in cases where the large virulence plasmid is no longer present, a product can still be amplified from the *ipaH* genes present in the chromosome reducing the occurrence of a false-negative reaction. Another positive feature of this primer set is that having multiple copies as targets influences the sensitivity of the PCR-based assay. However, it is not only the quantity of template for PCR that determines the successful PCR: the quality of the DNA is as important. DNA contaminated with inhibitors of PCR can lead to false negative reactions.

Much attention has been given to the preparation of DNA template free of PCR inhibitors. Clearly, in the clinical milieu, protocols can be designed for one particular matrix; quantity of template should not be a problem. Even though convalescent patients shed 10^2–10^3 shigellae per gram stool for perhaps weeks after the onset of disease, this number is suitable for detection. The limit of detection ranges from 30 to 50 CFU for PCR-based assays for detecting shigellae. In foods, as discussed above, multiple factors will influence how to prepare template. Several commercially available kits provide procedures for the detection of bacterial pathogens from foods. These protocols involve an enrichment step in broth for 6–18 h followed by treatment, as described by the manufacturer. Since shigellae do not survive well in most foods, this step may prove futile. Recently, a method has been described in which a small volume of a food wash is applied directly to a filter. These filters have the capacity not only to bind bacterial cells, but also to lyze the cells on contact and sequester the DNA within the membrane. The advantage of these filters is that bacterial (and parasitic) cells are concentrated and lyzed simultaneously, and washing the filters removes PCR inhibitors. With a pure culture of bacterial cells or shigellae seeded in different foods, the limit of detection is usually less than 100 CFU. Unlike other methods, which may take several days to determine if *Shigella* spp. are present in foods, PCR-based assays can yield a result in less than 1 day.

A method that uses a combination of polymerase chain reaction and enzyme-linked immunabsorbent assay in lieu of identifying polymerase chain reaction products by gel electrophoresis was found to be suitable for processing a large number of clinical samples. Its application to food analysis has yet to be tested.

Treatment and Prevention

Shigellosis is a self-limiting infection in normally healthy patients, and full recovery usually occurs without the use of antibiotics. In cases where antimicrobials are administered, the antibiotic of choice is trimethoprim–sulfamethoxazole. Ampicillin and tetracycline are alternative antibiotics to use, but multiple drug resistance among isolates of *Shigella* is becoming more common. Clinical isolates resistant to sulfonamides, ampicillin, trimethoprim–sulfamethoxazole, tetracycline, chloramphenicol, and streptomycin have been found. Consequently, drug-resistant organisms emerge from extensive use of antibiotics, and therefore, many practitioners believe that antimicrobial therapy for shigellosis should be reserved only for the most severely ill patients. However, there are persuasive public health arguments for the antibiotic management of shigellosis. Effective antibiotic treatment limits the duration of the disease and shortens the period of fecal excretion of shigellae, thus limiting the carrier state. Prompt replacement of liquids and electrolytes is important to restoring the health of the patient.

Since an infected person or asymptomatic carrier can be a focal point for person-to-person and food and waterborne spread, antibiotic treatment of these individuals can have an immediate effect on public health by containing the spread of shigellae. In industrialized countries, *Shigella* are one of the leading causes of outbreaks of diarrhea in daycare centers and have a secondary attack rate of 10–30%. As noted, antibiotics are not a substitute for good hygienic practices to contain secondary spread of shigellosis. As promulgated by the WHO, the single most effective means of preventing secondary transmission is handwashing. Food handling and preparation are important processes that also deserve attention, and persons with diarrhea should be excluded from handling food.

Despite many years of intensive effort, an effective vaccine against shigellosis still has not been developed. Attenuated, oral vaccine strains of *S. flexneri* are currently being tested. However, one major drawback still to be resolved is the transient fever and mild diarrhea associated with administration of the vaccine. A persistent problem impeding the development of a safe *Shigella* vaccine is designing a strain that can induce a protective immune response without producing unacceptable side-effects.

Summary

Foodborne outbreaks caused by *Shigella* species remain a concern to public health. *Shigella* can spread quickly because of the low infectious dose required to elicit clinical symptoms. This pathogen is not indigenous to any particular food; its host is either humans or subhuman primates. Outbreaks of shigellosis, from a relatively small number of infected people in a confined environ or to a multitude of clinically ill people, can occur by the ingestion of contaminated foods. A primary source of contamination is food handlers with poor personal hygiene. The disease is self-limiting, but treatment with antibiotics can limit the extent and possibly the spread of the illness. Currently, there is no vaccine available for shigellosis.

See also: **Biosensors**; **Food Poisoning**: Classification; **Hazard Analysis Critical Control Point**; ***Salmonella***: Salmonellosis

Further Reading

Bacteriological Analytical Manual (1998) *Shigella*. Chapter 6, pp. 6.01–6.06. Gaithersburg, MD: AOAC International (http://www.cfsan.fda.gov/~ebam/bam-toc.html).

Blocker A, Jouihri N, Larquet E *et al.* (2001) Structure and composition of the *Shigella flexneri* 'needle complex', a part of its type III secretion. *Molecular Microbiology* 39: 652–663.

Buchrieser C, Glaser P, Rusniok C *et al.* (2000) The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of. *Shigella flexneri*. *Molecular Microbiology* 38: 760–771.

Dorman CJ and Porter ME (1998) The *Shigella* virulence gene regulatory cascade: a paradigm of bacterial gene control mechanisms. *Molecular Microbiology* 29: 677–684.

Dupont HL (1995) Shigella species (bacillary dysentery). In: Mandell GL, Bennet JE and Dolin R (eds) *Principles and Practice of Infectious Diseases*, pp. 2033–2039. New York: Churchill Livingstone.

Ewing WH (1986) *Edwards and Ewing's Identification of Enterobacteriaceae*, 4th edn., pp. 146–147. New York: Elsevier.

Sansonetti PJ (2000) Rupture, invasion and inflammatory destruction of the intestinal barrier by *Shigella*, making sense of prokaryote–eukaryote cross-talks. *FEMS Microbiology Reviews* 25: 3–14.

Smith JL (1987) *Shigella* as a foodborne pathogen. *Journal of Food Protection* 50: 788–801.

Smith JL and Palumbo SA (1982) Microbial injury reviewed for the sanitarian. *Dairy and Food Sanitation* 2: 57–63.

Shrimps *See* **Shellfish**: Characteristics of Crustacea; Commercially Important Crustacea; Characteristics of Molluscs; Commercially Important Molluscs; Contamination and Spoilage of Molluscs and Crustaceans; Aquaculture of Commercially Important Molluscs and Crustaceans

SINGLE-CELL PROTEIN

Contents

Algae

M García-Garibay, L Gómez-Ruiz and A E Cruz-Guerrero, Universidad Autónoma Metropolitana, Mexico City, Mexico
E Bárzana, Universidad Nacional Autónoma de México, Mexico City, Mexico

Introduction

Single-cell protein (SCP) refers to crude or refined protein of algal, bacterial, mold, or yeast origin which is used either as animal feed or human food. The production and utilization of microbial biomass as a source of food proteins gained particular interest as an alternative source for proteins of agricultural origin due to its high content of protein. In addition to proteins, SCP contains other nutrients such as lipids and vitamins. Algae as a source of SCP is a term which refers to either microscopic single-cell true algae or prokaryotic cyanobacteria, and their growth is based on use of carbon dioxide and light energy (autotrophic growth). In contrast to other SCP-producing organisms, algae are grown in many cases by processes resembling traditional agriculture, since they depend on large areas and sunlight radiation. (*See* **Mycoprotein**.)

Organisms

The term 'algae' is generically given to photosynthetic organisms, either microscopic or macroscopic, living largely in water habitats, growing as undifferentiated or little-differentiated tissues. The designation is applied to taxonomically unrelated species, and in some cases includes groups like cyanobacteria or *Euglena*. However, the cyanobacteria is a group included in the Procaryotae kingdom, therefore they are more closely related to bacteria than to 'true' algae. However, they are sometimes known as blue-green algae, and for SCP purposes they are usually considered as such. *Euglena* is a genus of microorganisms belonging to the Protozoa kingdom; it is a strange case of an unicellular animal with chlorophyl, therefore it should not be considered as algae, and it is not commonly used as a source of SCP.

'True' algae belong to the Plantae kingdom, being the simplest plants. There are unicellular and multicellular organisms, some of them reaching huge sizes.

Many algae have been used as food for a long time. These include single-cell organisms as well as multicellular seaweeds, which have an important position in the diet of coastal communities. Table 1 shows the main genera of algae used as food and their cellular characteristics.

Macroscopic algae do not easily fit the SCP definition, due to their multicellular nature, and the low protein content of the final product (6–30% on dry-weight basis). They are mostly collected from the sea and, when cultivated, the production resembles farming more than a biotechnological process. The most widely consumed seaweed is *Porphyra* (an alga belonging to the Rodophyceae – the red algae), particularly *P. tenera*, which has a widespread distribution. It is mainly consumed in Japan, and also in the Philippines, Wales, and New Zealand. Among the most important species of seaweeds used as food

is *Ulva lactuca* (sea lettuce) which is used as a salad ingredient in western Europe. *Enteromorpha* is consumed in Hawaii and the Philippines in salads or as a flavor-enhancer for fish dishes. *Caulerpa* is also consumed in the Philippines. These three later algae belong to the Chlorophyceae – the green algae.

Both true unicellular algae and cyanobacteria have been consumed for centuries. This practice dates back to the Aztecs in Central Mexico, long before the discovery of the New World, when *Spirulina maxima* was harvested from natural habitats for human consumption. A similar species, *S. platensis*, is still used in Lake Chad in Central Africa for the same purpose. Both species *S. maxima* and *S. platensis* have been classified interchangeably in either the genus *Arthrospira* or the genus *Spirulina*. Figure 1 shows a picture of *S. maxima* from Lake Texcoco in Mexico. Other ancient cultures have used microalgae as food but on a lesser scale. For instance, *Nostoc* is consumed in Mongolia, China, Thailand, and Peru, while *Oedogonium* and *Spirogyra* are consumed in Burma, Thailand, and Vietnam; *Chlorella* is produced in Japan and Taiwan and *Scenedesmus* in China.

Table 1 Main genera of algae used as food and their cellular organization

	Prokaryote organisms (cyanobacteria belonging to the Prokaryotae kingdom)	*Eukaryote organism (true algae belonging to the Plantae kingdom)*
Unicellular growth	*Anabaena* *Nostoc* *Spirulina* (*Arthrospira*) *Tolypothrix*	*Chlorella* *Oedogonium* *Scenedesmus* *Spirogyra*
Multicellular growth		*Caulerpa* *Coelastrum* *Enteromorpha* *Porphyra* *Ulva*

Production

Besides the traditional harvesting of native algae in several regions, industrial and experimental efforts are being made in many countries, with either large and shallow ponds in areas of high sunlight, or bioreactors for the production of unicellular organisms.

Macroalgae cultivation is started by providing anchorage in sheltered bays for the initial culture; once a considerable growth of young plants is obtained, they are transplanted to nitrogen-rich tidal estuaries with low salinity.

Single-cell algae are produced by a variety of methods, ranging from cultivation in lakes or earthen ditches or ponds to technically advanced fermenters or bioreactors. Intensive cultivation was initiated in the 1940s in Germany, followed by technological developments in Japan and Taiwan during the 1950s and 1960s. Later on, commercial production systems were developed in the USA, Mexico, Thailand, and Israel. Currently commercial production is carried out in many countries around the world, though at a modest scale.

Figure 1 *Spirulina maxima* harvested from Lake Texcoco in Mexico, observed by scanning electron microscopy. Courtesy of Dr Jorge Sepúlveda, Universidad Nacional Autónoma de México.

Outdoor cultivation may be in open or closed systems. Growth occurs not more than 0.5 m from the water surface, as it is controlled by light penetration. In clean-culture systems a single species is inoculated and can be maintained over extended time periods; these are closed photobioreactors operating either outdoors or indoors, ranging from large closed areas exposed to sunlight to smaller reactors illuminated with artificial light.

Substrate Requirements

Algae are autotrophic organisms with the distinct advantage of using carbon dioxide as carbon source. Carbon dioxide represents the cheapest and probably most abundant carbon source for microbial growth. Due to such inexpensive requirements, these organisms are very convenient for SCP production. Some species can also grow heterotrophically using organic carbon sources.

The process is limited by the carbon source; the concentration of dissolved carbon dioxide is rather low, owing to its low solubility in aqueous solution. Carbon dioxide can be injected from combustion gases. Some additional sources of carbon enhance cell growth: cheap organic materials such as manure, molasses, or industrial wastes can be used. The added organic compounds are rapidly degraded by bacteria to make the carbon dioxide needed for algal reproduction. Lake Texcoco in Mexico has high concentrations of carbonate and bicarbonate, which are efficiently consumed by *Spirulina maxima*.

Nitrogen sources are generally nitrates, nitrites, ammonia, or urea. Oxidized nitrogen compounds require energy to be reduced, therefore ammonia is a more convenient form. Many cyanobacteria are able to fix atmospheric nitrogen; species of the genus *Anabaena* are particularly active in doing this.

Another nutrient of importance is phosphorus that can be incorporated in inorganic form. Micronutrients are needed in minimal amounts.

The use of waters polluted with organic wastes as input for algal ponds has the advantage of using a cheap raw material and promoting decontamination while obtaining a good source of SCP. Even more, the macronutrients needed (ammonium, phosphate) are normally present in domestic sewage, animal wastes, and food industry residual waters. (*See* **Effluents from Food Processing**: On-Site Processing of Waste; Composition and Analysis.)

Mass-Culture Open Systems

For open systems, lakes, ponds, and ditches are used. They can have an earthen floor or be lined with concrete or plastic. Open systems have low cell densities with large variation in productivity, depending very much on water properties and environmental conditions. In these kinds of systems the weather conditions, especially temperature and sunlight radiation, are critical. On the other hand, the needed facilities are constructed at low cost, and large areas are available.

It is a challenge to keep a monoalgal culture in an open-air system. The propagation of mixed populations and frequent problems of contamination by bacteria, fungi, protozoa, and invertebrates usually disturbs the productivity and lowers the quality of the product. However, under such nonsterile, mixed-culture conditions some algal species tend to predominate. The surface of the water in some ponds or ditches used for mass production is covered with polyethylene or other plastic material to reduce the risk of infection. Another possibility of avoiding contamination is the common practice of seeding a large inoculum to dominate the culture, at least during the first growth phase. Other approaches to promote single-species culture have been evaluated; for instance, the use of nitrogen-fixing species of the genus *Anabaena* in media without other nitrogen source has proved useful. Similarly, the selection of thermophilic species, such as *Scenedesmus obliquus*, has shown some preliminary potential.

Gentle agitation is very important to achieve high productivity. It prevents sedimentation, allows a more homogeneous exposure of algal cells to light, and reduces nutrient and temperature gradients along the depth of the culture. To this end, several designs have been implemented in ponds and ditches, including paddlewheels, gravity flow, and pump recycle, combined with special designs of slopes in oval ponds and horizontal raceways.

Productivities rarely exceed $30\,g\,m^{-2}\,day^{-1}$ and cell densities of $2\,g\,l^{-1}$, which are much lower than the values for other industrial fermentation processes. Some experimental improvements have been achieved by optimizing the use of wastes through the addition of nitrogen sources, adding aeration ports, and inoculating selected bacteria that efficiently degrade the diluted organic materials. By such means, the dry-weight productivities reach about four and three times the average figures obtained for maize and soybean respectively on a yearly basis. With intensive modes of cultivation, algal cultures can produce up to 20–35 times more protein than soybean for the same area of land.

An example of a typical mass-culture open system is that conducted by the Sosa Texcoco Company in Mexico, mainly during the 1970s and 1980s, in which an alkaline (pH 9–11) lake with 900 hectares of surface produced up to 1000 metric tons per year

of *Spirulina maxima* (equivalent to 0.111 kg m^{-2}). Weather conditions in Central Mexico, with high and practically constant temperature and solar radiation through the year, and the high alkalinity of the water favor effortless predomination of *S. maxima* on Lake Texcoco. The water of the lake flows by gravity into a sloped spiral raceway, maintaining gentle agitation. This factory stopped its production in the mid-1990s due to a long strike that led to the bankruptcy of the company.

The most advanced system developed so far is the high-rate algal ponds (HRAP) which combines the treatment of sewage with the simultaneous massive production of algae. The project was developed by the Technion Research Center at Haifa, Israel. Basic infrastructures are shallow canals that add up to a maximum of 1000 m^2, equipped with systems for gentle agitation and aeration. The process is operated continuously with retention times varying between 2 and 6 days, depending on the season. A steady multiculture is established in the system within a few days of operation. This includes bacteria that degrade organic compounds, and well-defined algal species, with *Euglena*, *Chlorella*, and *Scenedesmus* predominating. The maximum daily productivity reported at times of maximum solar radiation is 30 g m^{-2}, with an average annual production of 7 kg of algae per m^2. To recover the cells, aluminum sulfate is added as a flocculant; the float is then dewatered by centrifugation and dried in a drum drier to reach a final moisture content of 10%. The final product has been shown to have an excellent nutritional quality, containing 57.4 g of crude protein per 100 g, and an amino acid profile superior to the average for soybean protein. It has been used to complement at least 25% of fish diet and 10% of poultry diet with no toxic effects. The resulting effluent can be used directly for crop irrigation.

Chlorella ellipsoidea is produced in Taiwan for food use in open ponds with agitation or circulation. The product is recovered by filtration. In China, pilot-scale open systems have been used for the production of *Spirulina*, mainly *S. platensis*. Ponds about 100 m^2, covered with polyvinylchloride sheets, lead to a daily production of 10 g m^{-2} (equivalent to more than 3 kg per year per m^2 if the production could be maintained constant through the year). Other microalgae produced in China in open-air systems are *Scenedesmus, Chlorella*, and *Anabaena*.

Photobioreactors

Photobioreactors are closed systems working either outdoors or indoors, in which a single species is inoculated to keep a clean-culture operation. Closed cultivation systems offer better control of contamination and cell physiology than open systems, leading to higher growth and quality of the harvested product, but manufacturing costs increase.

Large systems operating outdoors consist of tubes covering large areas exposed to sunlight and can be operated either in batch or continuously. Many designs have been constructed or proposed at pilot scale. Tubes are made of either glass or plastic, such as polyethylene. Since the tubes behave as solar collectors, overheating is a problem. Hence, the tubular solar receptors must have a temperature control system, which is usually a water pool. Alternatively, the use of thermotolerant strains has been proposed to avoid cooling facilities. Generally, tubes are grouped in several modules to facilitate control, operation, and to offer flexibility to the system in terms of culture volume. CO_2 supply systems such as carbonation towers, pumps for circulating the medium, and tanks to mix nutrients are attached to the solar receptors. These photobioreactors have been used for the production of *Chlorella, Spirulina*, and *Scenedesmus*.

Other designs constructed in Chile up to a scale of 110 m^2 of solar irradiation area consist of a pond made of cement lined with epoxy resin and covered with a polyethylene dome. The agitation system is a paddlewheel. It has been used to produce *Spirulina* biomass, reaching a growth density of 450–750 mg l^{-1}.

An innovative design, operated in the USA, is based on the use of oval plastic bags floating on thermal waters.

Photobioreactors operating indoors necessarily have smaller sizes because artificial light is needed. Designs can be either plastic tubular systems, or stainless-steel fermenter-like reactors with internal illumination to allow maximum light incidence. Their use is rather limited for SCP production because of low throughputs; however, they are quite adequate for the production of algal metabolites with high added value, such as polysaccharides, carotenes, and other pigments, polyunsaturated fatty acids, etc.

Harvesting

The recovery of microalgal biomass after production is rather difficult, particularly in large-area lakes, or when low concentrations occur. Some species, such as *Spirulina platensis, S. maxima*, and *Coelastrum probiscideum*, are easily skimmed off or harvested by filtration through cloths or screens. Filter presses can be used as well. Owing to their small cell size (10 μm), other species need to be harvested by centrifugation or flocculation, adding flocculants such as lime, alum, or polyelectrolytes.

After harvesting, the algal biomass must be dewatered by centrifugation and/or dried. Operations to dehydrate biomass are normally done by

drum-drying, sun-drying, or spray-drying; drum-drying is the most widely preferred. In some cases, the product is not dried; for example, in China, *Scenedesmus* and *Chlorella* are fed to swine as fresh algal slurry.

Nutritional Value and Human Consumption

Algal SCPs have nutritional values similar to other SCP sources. The crude protein content (N × 6.25) varies between 45 and 73%, while lipid and mineral contents are 2–20% and 5–10% respectively. The chemical composition of some algal species is shown in Table 2. The protein content of algae is higher than the value for soybean (40 g 100 g^{-1}).

The amino acid content of algal SCP shows an adequate balance except, as any other microbial biomass, for the sulfur-containing amino acids methionine and cystine. They are rich in vitamins, especially some of the water-soluble ones, and essential fatty acids. The content of some vitamins, such as thiamin, riboflavin, folic acid, and carotene, is higher in algae than in many vegetable foodstuffs. However, the content of nutrients is highly dependent on cultivation and processing conditions. Table 3 shows some indices of protein quality of some algae. The protein efficiency ratio (PER), net protein utilization (NPU), and biological value (BV) of algal proteins are somewhat lower than casein. Nutritional tests have shown promise when algae supplemented with methionine and cystine are fed to broilers. However, monogastrics have problems digesting the whole cells and some processing is therefore needed.

Algal cell wall is not readily digestible, therefore any treatment to disrupt the walls will increase

Table 2 Composition of algal species. Main components in g 100 g^{-1} dry wt. Amino acids and vitamins as specified

	Spirulina maxima	Chlorella	Scenedesmus obliquus	Scenedesmus acutus
Crude protein (N × 6.25)	55–71	40–58	50–56	46–64
True protein	48–61			44–48
Amino acids (g 16 g^{-1} N)				
Alanine	5.0–6.1	4.2–7.4		5.3–10.4
Arginine	4.5–9.3	5.8–10.2		4.6–7.1
Aspartic acid	6.0–15.2	6.9–8.8		6.5–11.1
Cystine	0.6–2.2	0.3–0.9		0.6–1.6
Glutamic acid	8.2–21.8	8.0		5.3–10.7
Glycine	3.2–4.0	4.9–5.5		3.4–7.0
Histidine	0.9–1.6	1.4–3.0		1.5–2.3
Isoleucine	3.7–4.5	3.1–6.4		2.2–4.9
Leucine	5.6–7.7	6.8–9.7		5.0–10.6
Lysine	3.0–4.5	4.9–9.4		5.0–6.4
Methionine	1.6–2.2	1.0–2.0		1.4–2.7
Phenylalanine	2.8–4.0	3.2–5.1		3.6–6.4
Proline	2.7–3.2	2.2–6.4		3.1–6.1
Serine	3.2–4.3	3.0–4.1		3.2–5.4
Threonine	3.2–4.5	3.6–4.7		3.0–5.8
Tryptophan	0.8–1.2	1.0–1.5		0.3–1.8
Tyrosine	3.9	2.6–4.1		2.0–4.6
Valine	4.2–6.0	4.8–6.0		4.7–7.4
Lipids	4–7	6–16	12–14	8–14
Carbohydrates	13–16		10–17	
Minerals	4–9	6–9	4–9	6–17
Vitamins (mg 100 g^{-1})				
Thiamin	5.5	0.6–2.3		1.2–8.2
Riboflavin	4.0	2.0–6.0		3.4–36.6
Pyridoxine	0.3	0.1–3.2		1.1–2.5
Nicotinic acid	11.8	10–22		12–16.7
Pantothenic acid	1.1	1–10		1.5
Folic acid	0.05	0.1–4.0		0.7
Biotin	0.04	0.015–0.064		0.02–0.2
Cyanocobalamin	0.02	traces		0.04–0.44
Ascorbic acid		18–370		165–181
β-Carotene	0.17			
γ-Tocopherol	19.0	26–33		14–18.5

digestibility and hence nutritional value. Many treatments have been suggested, such as mechanical disruptions, extractions with organic solvents, and treatments with alkalis and/or acids. The method of drying affects product bioavailability. In drum-dried algae compared to the air-dried product, NPU is increased to around 100%, while digestibility is increased to about 60%. This phenomenon may be due to the rupture of the algal cell walls when water is removed at controlled conditions. Figure 2 shows spray-dried cells of *Spirulina maxima*; it is clear that the drying process disrupts cells if the cellular structures shown in Figure 1 and Figure 2 are compared. The micrograph shown in Figure 2 was taken from a sample of the spray-dried commercial product (powder) shown in Figure 3.

Table 3 Parameters of protein quality of some microalgae

Product	*NPU*	*PER*	*BV*
Spirulina platensis	52.7		68
S. platensis + methionine	62.4		82.4
Spirulina spp.	65	1.80	75
Spirulina spp. + methionine	73		82
Chlorella spp.	66		72
Chlorella spp. + methionine	78		91
Scenedesmus spp.	67	1.93	81
Casein	83	2.50	88

NPU, net protein utilization; PER, protein efficiency ratio; BV, biological value.

Many uses have been made of algal SCP, including the cultivation of daphnid and similar species that thrive on plankton as a food source in aquaculture. Particularly, some algal species rich in carotenoids have been used to feed salmon and trout to enhance the color of their flesh, and the US Food and Drugs Administration (FDA) has approved the 'all natural' statement for farm-raised fish. In addition to its use as feed for chicken and swine, algal biomass has been used as food worldwide. *Chlorella* is produced in Taiwan for food use; *Spirulina* has been produced commercially in the USA, Mexico, Taiwan, Japan, Thailand, and Israel with the same purpose.

Algal SCP has mainly been used for the preparation of nutraceuticals or dietary supplements, alone or mixed with other sources of protein and other food ingredients, sold as tablets, caps, powders, and other products available mainly in health food stores or nutrition centers, but also sold in groceries, drug and discount chain stores; they are promoted as protein and vitamin supplements, or to help people lose

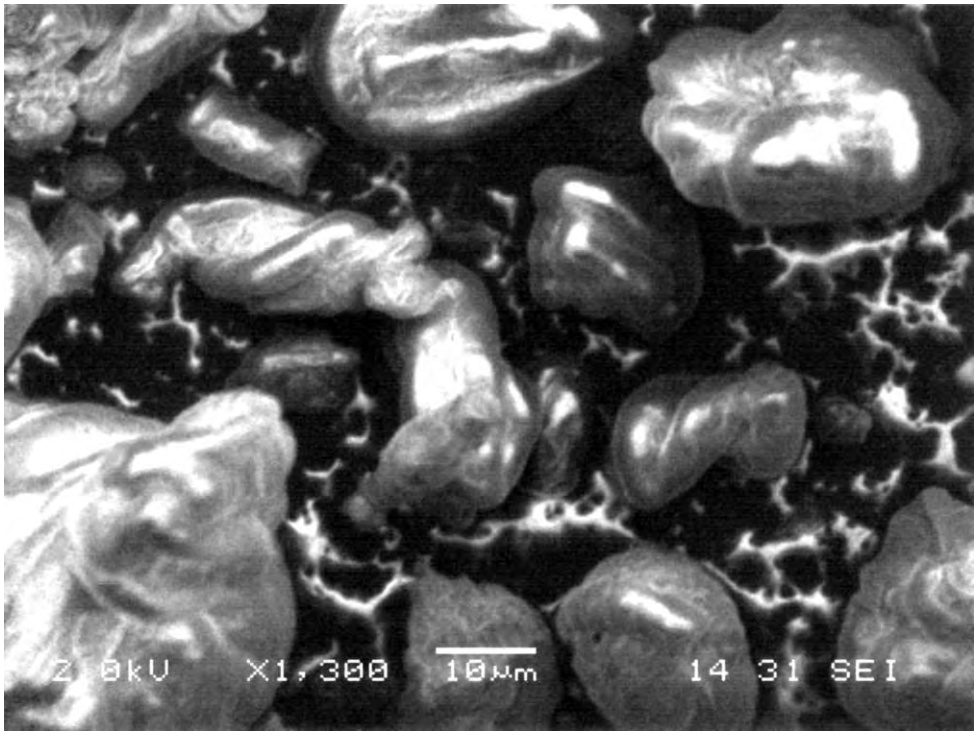

Figure 2 Cells of *Spirulina maxima* disrupted by the spray-drying process, observed by scanning electron microscopy. The sample was taken from a commercial powder.

Figure 3 (see color plate 126) Two commercial products elaborated with 100% dried *Spirulina maxima* biomass, sold as nutraceuticals in health food stores in Mexico City.

weight. However, very few indepth studies, if any, have been conducted to evaluate the nutritional or health-associated benefits of algal SCP. In countries such as the USA, Mexico, and Chile algal biomass is sold as tablets or powders, while in Japan and Taiwan it is sold either as dry powder or as pellets. Figure 3 shows two commercial products elaborated with 100% biomass of *Spirulina maxima*.

The FDA considers SCP from algae to be a dietary supplement; like any other dietary supplement, manufacturers do not need FDA approval to sell their product; this means that the FDA does not keep a list of manufacturers or products. However, claims on the product label other than nutrition content or approved health claims must be notified to the FDA. Examples of such claims are 'free radical defense' or 'helps combat free radicals,' used in some products containing algal SCP rich in astaxanthin, a carotenoid with antioxidant capability, or 'structure/function' claims such as 'promotes cellular health' or 'support immune system health' for a mixture of *Spirulina* and *Chlorella*, statements which had not been evaluated previously as health claims by the FDA. Another example is 'helps maintain healthy eyes' and 'helps support immune function,' claims used in a product sold as powder or tablets made by a strain of *S. platensis* rich in carotenoids.

A wide range of food formulations has been prepared with algal biomass. For instance, powder meals containing *Spirulina* plus soy proteins, oat bran, apple pectin and vitamins, either chocolate- or vanilla-flavored, are sold as energy supplement powders to mix with milk or juices. Algal SCP has been widely used as an additive or supplement to cereal foodstuffs, or as a garnish to salads. Experiences related to supplementation of cereal foods with algal SCP include mixtures with doughs for baked goods and pasta, such as bread, rolls, cookies, noodles, etc. In Mexico, *S. maxima* has been used as a supplement for cookies produced by a state company as part of a national breakfast program for schoolchildren. Also, the use of *S. maxima* has been suggested to enrich tortillas, increasing their protein content; however, these products have not found success in the market place.

Table 4 Functional capacities of a protein concentrate obtained from *Spirulina platensis*

Water absorption (g 100 g^{-1} protein)	145
Fat absorption (g 100 g^{-1} protein)	373
Emulsifying capacity (ml oil 100 g^{-1} protein)	113
Foam capacity (%)	205
Foam stability (1 h) (%)	27

Concerning functional properties, it has been reported that *S. platensis* flour had similar emulsion and foam capacities, slightly lower water and fat absorption capacities, and lower foam stability than soybean meal. Using the protein concentrate obtained from the flour, improved functional characteristics with the exception of water absorption ability are obtained. Some values obtained for functional capacities of a protein concentrate obtained from *S. platensis* are given in Table 4.

The major problems in acceptability of algal biomass are its unfamiliar and sometimes bitter flavor, as well as the presence of dark green pigments which are difficult to mask. In addition to chlorophyls, other pigments such as carotenes, xanthins, and phycocyanin are present in varying amounts. Flavor and color may be improved if algal biomass is treated during downstream processing to remove undesirable components.

Toxicological Problems

Algae have been consumed as food for generations, without ill effects. No toxic effects have been reported in animal evaluations. However, the following considerations must be taken into account.

A common problem for SCP from any microorganism is the high content of nucleic acids present in microbial cells. Consumption by humans of nucleic acids in amounts higher than 2 g day^{-1} can lead to accumulation of uric acid, which increases the risk of gout and/or kidney stones. The concentration of nucleic acids in algal biomass depends on several factors, such as species and growth conditions. Cyanobacteria have a nucleic acid content of 2.9–5 g

$100\,g^{-1}$ while microscopic plant algae have 1–17 g $100\,g^{-1}$. These amounts are higher than in most other foodstuffs.

In order to reduce nucleic acid content, protein concentrates or isolate can be prepared by cell disruption and protein separation. Figure 4 shows a flow diagram of a process for the preparation of protein isolate from *Spirulina*. However, this increases the cost of the product. In general, algal biomass is consumed by humans in small amounts and so the consumption of nucleic acids is below risk.

The cell wall of microalgal biomass represents about 10% of its weight. It is mainly composed of indigestible polysaccharides and some other compounds, e.g., murein in cyanobacteria. The bioavailability of protein from whole cells is therefore very low. The preparation of protein concentrates or isolates can be used to obtain products with a high nutritional value and free of undesirable pigments, although it represents a costly alternative.

Algae have the ability to remove heavy metals from polluted waters. Similar physiological phenomena account for the accumulation of pesticides and organochlorine compounds. This represents an objection when algae are intended for use as SCP. However, the causes are well identified and some steps can be implemented to maintain the final product composition within safety levels.

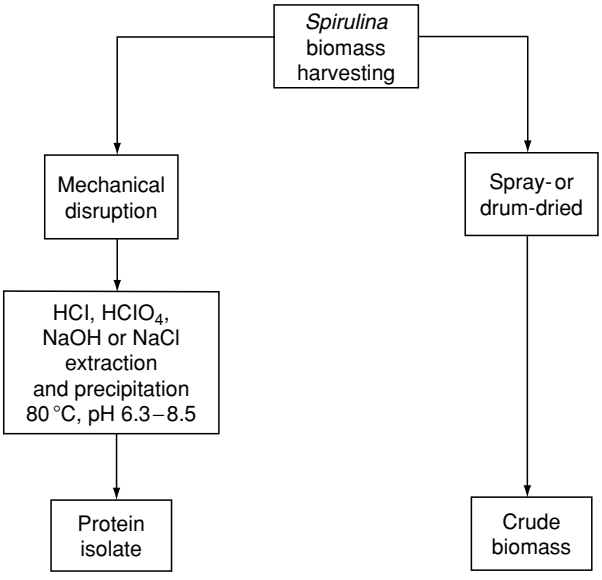

Figure 4 Process for *Spirulina* single-cell protein recovering either as crude biomass or as protein isolate.

The origin of the water used in cultivation ponds determines the need for pretreatment. In general, wastes generated in food industries carry low amounts of the contaminants mentioned. However, urban sewage and run-offs show high variability in the content of heavy metals and other toxicants. When these waste streams are subjected to standard secondary treatments, most organics are degraded, whereas metals remain associated with the activated sludge, rendering them safe as an input for algal growth. Furthermore, some studies have demonstrated that biological absorption of metals is a rather slow process, requiring more time than the usual retention time of the water within the bioreactor. It appears that fears concerning the presence of recalcitrant toxicants in algae might be excessive, although more research is needed to get a clear picture.

Another problem to consider is the possibility of contamination by pathogenic microorganisms. Certainly, the culture practices in open systems increase the risk of pathogenic infections. The downstream processes of SCP products are designed to destroy most of the viable forms present, although some could survive in the product. Recommendations have been established for microbiological standards of SCP products for use in animal feeds.

See also: **Single-cell Protein**: Yeasts and Bacteria; **Yeasts**

Further Reading

Goldberg I (1985) *Single-Cell Protein*. Berlin: Springer Verlag.

Litchfield JH (1991) Food supplements from microbial protein. In: Goldberg I and Williams R (eds) *Biotechnology and Food Ingredients*, pp. 65–109. New York: Van Nostrand Reinhold.

Moo-Young M and Gregory K (1986) *Microbial Biomass Proteins*. London: Elsevier.

Rolz C (1990) Características químicas y bioquímicas de la biomasa microbiana. *Archivos Latinoamericanos de Nutrición* 40: 147–193.

Rose AH (ed.) (1979) *Microbial Biomass. Economic Microbiology*, vol. 4. London: Academic Press.

Stadler T, Mollion J, Verdus MC, Karamanos Y, Morvan H and Christiaen D (eds) (1988) *Algal Biotechnology*. London: Elsevier Applied Sciences.

Switzer L (1980) *Spirulina. The Whole Food Revolution*. Berkeley, CA: Proteus.

Wood A, Toerien DF and Robinson RK (1991) The algae – recent developments in cultivation and utilization. In: Hudson BJF (ed.) *Developments in Food Proteins*, vol. 7, pp. 79–123. London: Elsevier Applied Science.

Yeasts and Bacteria

M García-Garibay, L Gómez-Ruiz and A E Cruz-Guerrero, Universidad Autónoma Metropolitana, Mexico City, Apartado, Mexico
E Bárzana, Universidad Nacional Autónoma de México, Mexico City, Mexico

Introduction

The single-cell protein (SCP) concept is applied to the massive growth of microorganisms for human or animal consumption. Single-cell protein is a generic term for crude or refined protein whose origin is bacteria, yeasts, molds, or algae, microorganisms that usually contain above 40% of crude protein on dry weight bases. Yeasts and bacteria have been particularly important for SCP production and easily acceptable as their biomass has been consumed by man since ancient times in the form of fermented foods. The production of SCP has important advantages over other sources of proteins, such as its considerably shorter doubling time, the small land requirement, and the fact that it is not affected by the weather conditions. Much attention was focused on the use of petroleum derivatives as substrates for the SCP production during the 1960s and 1970s when the price of this reserve was low, but currently, the production of SCP is based on renewable resources, and its interest is also kept as a means to confer value to waste materials. The organoleptic and functional properties of SCP are not always competitive, and its main drawback has been its high production costs. However, recent advances in fermentation technology and genetic engineering could reevaluate SCP production.

Historical Developments and Implementation of SCP Production

The first developments in SCP production were achieved during war times when conventional foods were in short supply. During World War I, *Saccharomyces cerevisiae* was massively produced in Germany from molasses to replace up to 60% of protein imported. Similar measures were taken during World War II for the mass production of *Candida utilis* (known formerly as Torula yeast, or *Torulopsis utilis*) on sulfite liquor from paper manufacturing wastes. After the war, several plants were built in the USA and Europe, mainly for *C. utilis* production.

Accelerated industrial development and general welfare expectancy led to a renewed interest in SCP as an alternative to alleviating food shortages due to a growing imbalance between food production and demand by the world's population, mainly in developing countries. During the 1960s and 1970s, several SCP production plants were built in the UK, France, Italy, Russia, Japan, Taiwan, etc.

An important breakthrough occurred when the production of SCP from hydrocarbons was demonstrated by several petroleum companies during the 1950s and 1960s. During the 1970s, considerable research efforts resulted in the use of methanol and ethanol, derived from petroleum, as convenient substrates. In the early 1970s, the cost of *n*-paraffin was approximately US$80 per tonne, whereas crude molasses had an approximate cost of US$82 per tonne. However, concern about substrate safety and the increase in petroleum prices shifted interest back to the utilization of renewal sources, mainly food and agriculture byproducts like molasses and whey, or industrial wastes rich in starch, cellulose, and hemicellulose. The major SCP projects based on petroleum derivatives as substrates were abandoned in the 1980s.

In recent times, among the European Communist countries, Russia has had the greatest capacity for SCP production, with at least 86 plants in operation, using different substrates. Several other countries in Eastern Europe, like the former Czechoslovakia, had similar capacities. The Russian delegation reported the production of 1 million tonnes of SCP per year in 1983 at an international symposium on SCP, and estimated that they would be producing 9 million tonnes by the end of the 1990s.

To date, an enormous number of reports about SCP production have appeared in the scientific literature. Two main approaches have been followed: the utilization of convenient substrates and the use of waste materials where SCP production brings about pollution control.

Many industrial process have been developed world-wide; the most important are shown in Tables 1 and 2. The countries with important industrial outputs are the USA, UK, France, Russia, and Cuba.

Suitable Organisms and Substrates

The most frequently used yeasts are as follows: *Saccharomyces cerevisiae* (and related synonymous species such as *S. carlsbergensis*, *S. uvarum*, etc.), *Kluyveromyces marxianus* (including synonymous subspecies and asexual forms such as *K. fragilis*, *K. lactis*, *K. bulgaricus*, *Candida kefyr* and *C. pseudotropicalis*), *C. utilis* (and its sexual form *Hansenula jadinii*) and *Yarrowia lipolytica* (formerly

Table 1 Main yeast SCP industrial and pilot developments

Substrate/yeast	*Process/product*	*Country*
Sulfite liquor		
Candida utilis	St. Regis Paper	USA
Candida utilis	Boise Cascade	USA
Candida utilis	State industry	Russia
Candida utilis	Attisholz	Switzerland
Hydrocarbons		
Yarrowia lipolytica	British Petroleum/ Toprina	UK
Candida tropicalis	British Petroleum/ Toprina	France
Candida tropicalis	Italprotein	Italy
Yarrowia lipolytica	Liquichimica/ Liquipron	Italy
Yarrowia lipolytica	State industry	Russia
Yarrowia lipolytica	Swedt	Germany
Yarrowia lipolytica	Roniprot	Romania
Ethanol		
Candida utilis	Amoco/Torutein	USA
Candida sp.	Mitsubishi	Japan
Methanol		
Pichia sp.	Mitsubishi	Japan
Pichia pastoris	Philipps Petroleum/ Provesteen	USA
Starch		
Saccharomycopsis fibuligera and *Candida utilis*	Symba	Sweden
Saccharomyces cerevisiae	Brewers	Several
Molasses		
Candida utilis	Several industrial processes	Cuba
Candida utilis	Several industrial processes	Taiwan
Candida utilis	CINVESTAV IPN	Mexico
Liquid sucrose		
Hansenula jadinii	Philipps Petroleum/ Provesta	USA
Whey		
Kluyveromyces marxianus	Wheast–Knudsen	USA
Kluyveromyces marxianus	Amber Lab–Universal Foods	USA
K. marxianus, K. lactis, and Candida pintolopesii	Fromageries Bel	France
Kluyveromyces marxianus	S.A.V.	France
Candida intermedia	Vienna	Austria
Candida utilis	Waldhof	USA
Confectionery effluent		
Candida utilis	Bassett	UK

Table 2 Main bacterial SCP industrial and pilot developments

Substrate/bacteria	*Process/product*	*Country*
Methane		
Methylococcus capsulatus	Shell Oil	UK
Methanol		
Methylophilus methylotrophus	ICI/Pruteen	UK
Methylomonas clara	Hoechst–Uhde/ Probion	Germany
Ethanol		
Acinetobacter calcoaceticus	Exxon–Nestle	Switzerland
Cellulose		
Cellulomonas sp. and *Alcaligenes* sp.	Louisiana State University	USA
Whey		
Lactobacillus bulgaricus and *Candida krusei*	Kiel	Germany

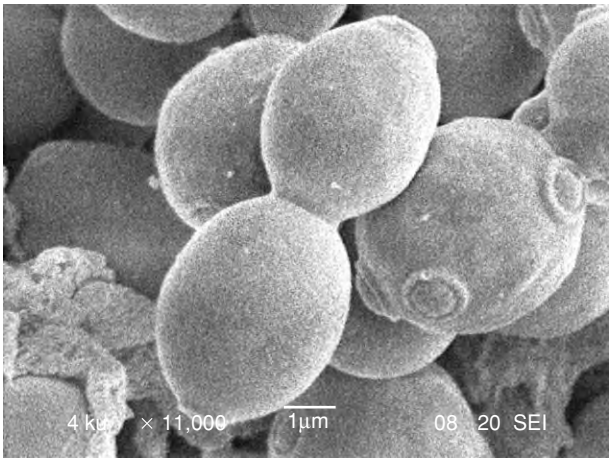

Figure 1 *Saccharomyces cerevisiae* observed under a scanning electron microscope.

known as *Candida lipolytica* and *Saccharomycopsis lipolytica*).

Some yeasts have been widely used in the manufacture of human foods: *S. cerevisiae*, *K. marxianus*, and *C. utilis* have the generally recognized as safe (GRAS) status for human consumption by the Food and Drugs Administration in the USA. *S. cerevisiae* is also available as spent yeast from breweries and alcohol industries; Figure 1 shows an electron micrograph of this species. *K. marxianus* and related species are widely used, due to their capacity to assimilate lactose, the carbohydrate present in cheese whey, but it can also grow on inulin, a fructose polymer found in some plants, and other simple sugars such as glucose, fructose, and sucrose; therefore, sometimes, it is also grown in molasses. Since it is able to grow at temperatures as high as 45 °C, it has been used to produce biomass in tropical areas. Figure 2 shows an electron micrograph of cells of *K. marxianus*. *C. utilis* is used for a wide variety of substrates such as sucrose, ethanol, and sulfite-spent liquor. It can also grow on wood hydrolysates because of its ability to assimilate pentoses. Starchy solids or water streams from potato and maize industries require previous hydrolysis for yeast growth, as in the case of *C. utilis*, or as in the Symba process, which utilizes amylolytic yeast (*Saccharomycopsis fibuligera*). Yeasts able to assimilate

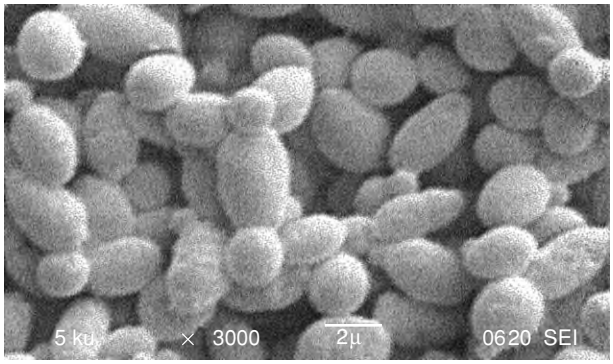

Figure 2 *Kluyveromyces marxianus* observed under a scanning electron microscope.

hydrocarbons include *Y. lipolytica*, *C. tropicalis*, *C. rugosa*, and *C. guilliermondii*, which usually can also grow on lipids. Methanol is the preferred alcohol for use as a substrate by *Pichia* species (*P. pastoris*, *P. methanolica*, etc.), *Hansenula polymorpha*, *Hansenula capsulata*, and *Candida boidinii*. The most important processes developed for yeast SCP production are shown in Table 1.

Bacteria have been used mostly for the production of animal feed. The most commercially important species utilize methane and/or methanol as substrates, as shown in Table 2. Methanol is usually preferred over methane because it is water-soluble and less explosive. The Ministry of Agriculture, Fisheries and Food in the UK allows the use of ICI's (Imperial Chemical Industries) product 'Pruteen' in animal feed. For the production of bacterial SCP from ethanol, either *Acinetobacter calcoaceticus* or *Alcaligenes* sp. have been used in laboratory or pilot-plant studies.

Other petroleum derivatives such as paraffins have been considered only to a minor degree for the production of SCP using bacteria; some studies have been carried out with *Acinetobacter calcoaceticus*.

To produce bacterial SCP from whey, several lactic acid and propionic bacteria have been investigated, frequently in mixed cultures with yeasts as in the Kiel process. In this case, *Lactobacillus delbrueckii* ssp. *bulgaricus* grows using the lactose, converting it to lactic acid; then, *Candida krusei*, which is unable to ferment lactose, uses the lactic acid as a carbon source. Both fermentations can be performed simultaneously, controlling the pH by adding ammonia, which is used as a nitrogen source by the yeast.

Lactic acid bacteria proposed for the fermentation of whey include species such as *Lactobacillus delbrueckii* subsp. *delbrueckii*, *L. delbrueckii* subsp. *bulgaricus*, *Lactobacillus casei* ssp. *casei*, *Leuconostoc* sp., etc. Fermentation of this substrate with rumen bacteria has also been proposed to produce biomass concentrated by ultrafiltration for use as feed.

Propionibacteria, such as *Propionibacterium freudenreichii* ssp. *shermanii* and *P. freudenreichii* ssp. *freudenreichii*, produce significant amounts of vitamin B_{12}, giving additional attractiveness to their utilization in the production of SCP. A process using a mixture of *P. freudenreichii* ssp. *freudenreichii* and *K. marxianus* has been proposed for the production of SCP from whey rich in vitamin B_{12} and sulfur amino acids.

Bacteria have been explored very little with regard to producing SCP from substrates such as starch and cellulose. A case in point is the project developed by Louisiana State University in USA to produce SCP from *Cellulomonas* sp. and *Alcaligenes* sp. from cane bagasse; this process was scaled up to pilot-plant level.

Generally, yeasts have been preferred for SCP production over bacteria, especially for human consumption. It seems that yeasts are more acceptable because they are more familiar to humans through ageless foods like bread or beer. However, bacteria have some advantages over yeasts such as a higher protein content (Table 3), higher yields (carbon source to protein conversion), and a faster growth rate, though a higher nucleic acid content (Table 4) limits their uptake in the diet. (*See* **Lactic Acid Bacteria**; **Yeasts**.)

Production Processes

Since the SCP must have a competitive price in the protein market, especially proteins of vegetable origin, it is essential to guarantee efficiency along all stages of the process. The carbon source accounts for up to about 60% of the operation costs, so high yields of substrate conversion are required, high productivity processes must be implemented, and the utilization of an inexpensive but easily assimilated carbon source has to be guaranteed. This explains the generalized use of molasses, whey, or industrial residues, depending on local availability, and the attempts to implement fossil fuels as substrates. An important advantage in using hydrocarbons is the yield, obtaining 1 g of dry biomass per gram of hydrocarbon, compared with carbohydrates, which typically yield around 0.5 g of dry biomass per gram of substrate.

The typical process stages for the production of SCP comprise: raw material preparation, fermentation, biomass recovery, cell disruption, and drying. Figure 3 shows a typical process for the production of yeast SCP.

Table 3 Nutritional parameters of SCP products

	Kluyveromyces marxianus	Saccharomyces cerevisiae	Candida utilis	Methylophilus methylotrophus	Methylomonas clara	*Soya meal*
Protein (grams per 100 g dry weight)						
Crude (N × 6.25)	43–58	48	42–57	72–88	80–85	44–50
True protein	40–42	36	47	64	69–73	48
Essential amino acids (grams per 16 g of N)						
Isoleucine	4.0–5.1	4.6–5.5	4.3–5.3	5.2–5.4	3.6	5.4
Leucine	7.0–8.1	7.0–8.1	7.0	8.2–8.4	6.6	7.7
Phenylalanine	3.4–5.1	4.1–4.5	3.7–4.3	4.3–6.5	5.1	5.1
Tyrosine	2.5–4.6	4.9	3.3	3.5–3.8	5.1	2.7
Threonine	4.1–5.8	4.8–5.2	4.7–5.5	5.7–6.5	4.8	4.0
Tryptophan	0.9–1.7	1.0–1.2	1.2	1.1–1.6	1.5	1.5
Valine	5.4–5.9	5.3–6.7	5.3–6.3	6.3–6.5	4.8	5.0
Arginine	4.8–7.4	5.0–5.3	5.4–7.2	4.3–5.6	3.4	7.7
Histidine	1.9–4.0	3.1–4.0	1.9–2.1	2.2–2.3	2.8	2.4
Lysine	6.9–11.1	7.7–8.4	6.7–7.2	4.1–7.3	6.2	6.5
Cystine	1.7–1.9	1.6	0.6–0.7	0.8		1.4
Methionine	1.3–1.6	1.6–2.5	1.0–1.2	1.4–3.0	2.5	1.4
PER	1.8	2.0	1.7			1.4–2.2
NPU	67			84		64
Vitamins (μg per g)						
Thiamin	24–26	104–250	8–9.5	5		9.0
Riboflavin	36–51	25–80	44–45	40		3.6
Pyridoxine	14	23–40	79–83	2		6.8
Nicotinic acid	136–280	300–627	450–550	57		24.0
Folic acid	6	19–30	4–21	15		4.1
Pantothenic acid	67	72–86	94–189	11		21.0
Biotin	2	1	0.4–0.8	3		
Vitamin B_{12}	0.015–0.05		0.0001			0

PER, protein efficiency ratio; NPU, net protein utilization.

Table 4 Nucleic acid contents of SCP products (g per 100 g of biomass in dry weight basis)

Kluyveromyces marxianus	2.7–10
Saccharomyces cerevisiae	2.5–15
Candida utilis	6.2–10
Methylophilus methylotrophus	16
Methylomonas clara	10–15
Isolated protein from *K. marxianus*	1.4–5.7

To maximize carbon assimilation, the nutrients must be balanced. Sources of nitrogen, minor elements (P, K, S, Mg, etc.), trace elements, and vitamins are adjusted according to the general composition of the carbon source. This in turn is highly dependent on the strain used. In general, simple nitrogen sources such as urea, ammonia, and nitrate are used to keep costs down. Phosphate is supplied as phosphoric acid or as soluble phosphate salts.

Fermentation variables like temperature, pH, ionic strength, level of oxygenation, and, in the continuous fermentations, dilution rate, have a strong influence on cellular yield. In particular, an abundant supply of oxygen promotes aerobic metabolism and higher growth rates. However, due to the low solubility of oxygen in aqueous media, the cost of aeration, through air sparging and agitation, increases rapidly with the scale of operation, resulting in an important technical challenge. Assimilation of *n*-paraffins requires considerably high levels of oxygenation, and the growth of microorganisms on these water-insoluble substrates takes place on the hydrocarbon–water interface. The surface area becomes the limiting factor, and heat production is about twice that using sugars as substrates.

In general, when yeast biomass is produced, alcohol accumulates due to oxygen limitation. Some alternatives proposed for SCP from whey are the production of alcohol as a byproduct from fermentation or the use of *Kluyveromyces* in a mixed culture with *Candida pintolopesii*, where the latter consumes the alcohol produced by the former. The first approach has been followed by Amber Laboratories (Universal Foods, USA), and the second approach has been adopted by the Fromageries Bel in France; *Candida valida* has also been demonstrated to prevent ethanol accumulation when it is grown

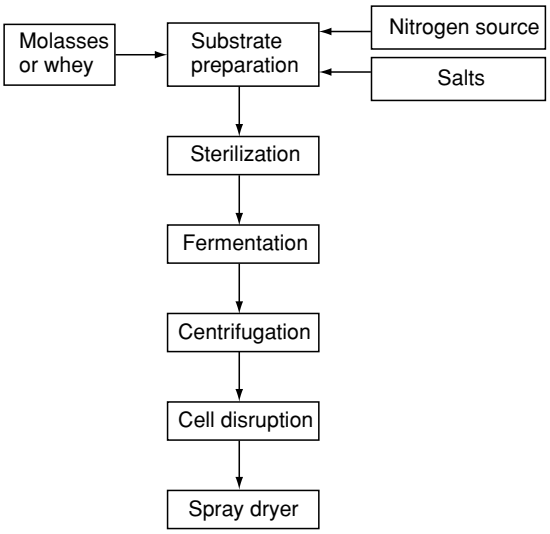

Figure 3 Typical process for the production of yeast SCP.

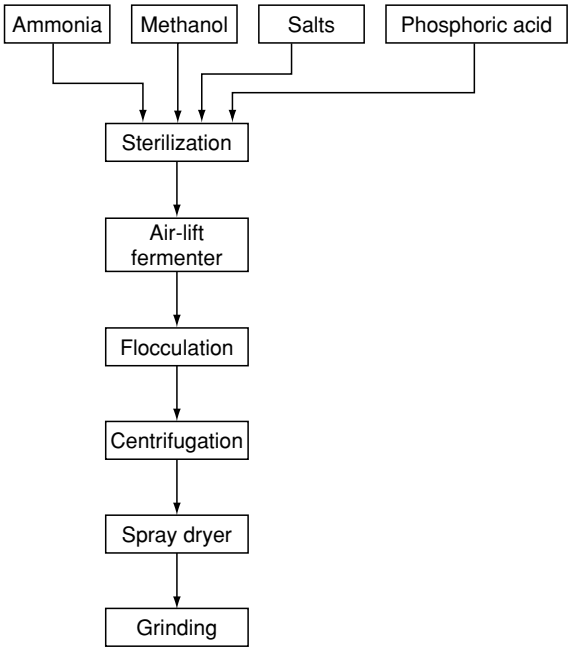

Figure 4 Industrial process developed in the UK by ICI for the production of Pruteen using *Methylophilus methylotrophus* grown in methanol.

mixed with *Candida kefyr* (the asexual form of *K. marxianus*), increasing the efficiency of whey conversion into biomass up to 20%. A limited oxygen supply, when hydrocarbons are used as substrates, has led to the production of some metabolites such as mono and dicarboxylic acids and ketones.

To sustain high oxygen transfer rates, large air volumes have to be supplied with high agitation rates. Alternative fermenter designs include the air-lift type, which leads to maximum oxygenation with minimum power requirements, diminishing aeration costs considerably. In fact, the largest fermenter ever operated is an air-lift ($3000\,m^3$) for the aerobic production of *Methylophilus methylotrophus* by ICI. Figure 4 shows a simplified diagram of the process developed by ICI, which is the most successful process developed for bacterial SCP production. Currently, the high-cell-density fermentation designs pioneered by Provesta (a company belonging to Philipps Petroleum) allows them to obtain up to $160\,g\,l^{-1}$ of yeast biomass, while traditional fermentation techniques yield at most $30\,g\,l^{-1}$. These fermenters have attached very efficient systems for heat removal and oxygen transfer.

Once obtained, the microbial biomass is recovered by filtration or centrifugation. The resulting cell suspension can be either spray-dried, or the cells can be broken to obtain extracts, hydrolysates, or autolysates. Finally, the protein can be concentrated, isolated, or texturized. In the Philipps Petroleum process, owing to the high cell density, the spent medium is fed directly to the spray-drier without prior concentration.

Nutritional Value

The main nutritional contribution of SCP either for human food or animal feed is its high protein content. Bacteria have a protein concentration ranging from 50 to 85% and yeasts from 45 to 70%. Protein quality is also quite acceptable, as compared with vegetable proteins. Generally, SCP from any source is limited in sulfur-containing amino acids; however, bacterial proteins tend to have better balances of essential amino acids than yeasts and other microorganisms, particularly with respect to sulfur amino acids. Some parameters of protein quality are given in Table 3. When methionine is added to SCP, the protein quality increases considerably and reaches values similar to those of casein. (*See* **Amino Acids**: Properties and Occurrence; Metabolism; **Protein**: Quality.)

SCP is also an important source of vitamins; actually, brewer's yeast has long been used as a vitamin supplement. Vitamin contents are also listed in Table 3.

Toxicological Aspects

Safety of both the microorganism and the substrate is an important consideration. Microorganisms used for SCP production have to be subjected to extensive toxicological clearance. Those microorganisms

normally present in fermented foods, such as *Saccharomyces cerevisiae*, are safe for human consumption. The main concern has been linked to the use of petroleum derivatives in which residual alkanes must be removed with solvents. However, some residual hydrocarbons may still be present, and several reports have stated the presence of unusually high amounts of odd-chain fatty acids and paraffins in tissues from animals fed with SCP from alkanes. These fatty acids, particularly unsaturated C17, have been suspected of having toxic effects, even though no evidence has been reported.

The high content of nucleic acids present in microbial cells also has some consequences. Human consumption of nucleic acids in amounts higher than 2 g per day could lead to the accumulation of uric acid, resulting in kidney stones and/or gout onset in susceptible people. The concentration of nucleic acids in the biomass is highly variable and depends on several factors. The first element is the nature, species, and strain of the microorganisms; normally, bacteria are present in higher concentrations than yeasts. Another factor is the growth conditions: the higher the growth rate, the higher the nucleic acids content in the cell. The nucleic acid content also changes with the growth phases. For instance, for *K. marxianus* grown on whey, the concentration of nucleic acids reached its peak at the middle of the exponential phase. Table 4 shows the nucleic acids content of several SCP products.

In order to reduce the nucleic acid content, two approaches have been followed: grow the biomass at low rates or isolate the protein by eliminating undesirable compounds. The later is usually applied, to eliminate not only nucleic acids but also the cell wall. In recent years, research has been conducted on the use of external (added) endonucleases to hydrolyze nucleic acids and so obtain protein isolates free from these polymers. Immobilized endonucleases have been proposed particularly for this purpose.

Yeasts and bacterial cell walls are difficult to digest, leading to poor bioavailability of proteins, flatulence and diarrhea, and allergic responses. Some cases of skin rashes, nausea, vomiting, and other allergic reactions have been reported, but they may be eliminated by reducing cell walls and nucleic acids.

Nucleic acid content in SCP for animal feed is not a problem, because animals have the enzyme uricase, which prevents uric acid accumulation. However, cell wall digestibility in monogastric animals is also poor.

Cell Disruption and Protein Isolation

Many processes to disrupt the cells have been developed. A common process is autolysis, in which the biomass is exposed to a heat shock or to chemical compounds, such as low-molecular-weight thiols. Yeast autolysis usually takes place when cells are heated to 45–55 °C, and it is enhanced by the presence of NaCl. Further incubation for around 24 h induces cellular enzymes, leading to complete cell lysis. The process also activates endogenous ribonucleases that reduce nucleic acids. Lysis can be facilitated by the addition of exogenous enzymes such as proteases, β-glucanases or lysozyme. The disadvantages of these techniques are the high costs involved and extensive proteolysis, which reduces the yield and functional properties of proteins. Chemical treatments with alkalis, organic solvents, or salts, which weaken cell walls, are also used. Alkaline treatment may result in undesirable side reactions, forming compounds such as lysinoalanine and off-flavors.

Hydrolysis requires the use of hydrochloric acid at temperatures as high as 100 °C, to treat a slurry with 65–85% solids content, followed by neutralization with NaOH. This process has many drawbacks such as the risk of accidents with concentrated acid, the need for anticorrosive equipment, which is cost-intensive, and the destruction of some amino acids and vitamins, reducing the nutritive value of the product.

Physical methods to break cell walls are the methods most widely used. High shear rates are achieved by means of homogenizers or colloidal mills and have been used extensively for SCP processes.

Once the cells have been broken, the protein is extracted using water or alkaline conditions; the cell wall debris is removed by centrifugation, and the protein is further precipitated with acid, salt, or heat treatment, while the nucleic acids remain in the supernatant. Usually, microbial proteins have their maximum solubility at pH 12, and the precipitation occurs at pH 4.5. The protein isolate is then obtained. Some chemical modification during protein extraction includes phosphorylation or succinylation, which facilitates protein separation from nucleic acids and improves its functional properties.

Utilization for Human Food

For food applications, besides toxicity and nutritional quality, organoleptic acceptability and functional properties are important considerations.

SCP has been used as a protein supplement in baked goods, cookies, crackers, snacks, soups, noodles, and special foods like geriatric or baby meals. Its use as an extender in sausages and other meat products has been important, mainly in Eastern European

countries. Despite the justified production of SCP in terms of the world's protein shortage and widespread malnutrition, a real demand for a protein is based on its favorability in terms of functional properties like solubility, water binding, emulsification capacity, gelation, whippability, and foam stability. The successful supplementation of existing products and the replacement of traditional proteins with SCP depend on the availability of proteins matching in functionality, price, and organoleptic acceptability.

For food applications, whole dried cells, disrupted cells, and textured protein products are useful. From disrupted cells, either protein concentrates or isolates can be obtained, which are better suited for the food industry. Moreover, SCP isolates can compete favorably with soy isolates from the functional point of view. However, isolation or concentration increases production costs dramatically.

Processes such as texturization, by spinning and extrusion, and enzymatic or chemical modification can improve the functionality of SCP. For instance, protein fibers obtained by spinning can form textured protein products such as meat extenders.

Enzymatic modification includes partial proteolysis to improve solubility, emulsification capability, and whippability, or the reverse reaction known as plastein (peptide bond formation) to improve the nutritional value through the addition of limiting amino acids. Promising chemical modifications include acetylation, which improves the thermal stability, succinylation, or phosphorylation to increase the solubility, emulsification, and foaming capacities. However, such modifications tend to reduce the nutritive value of the proteins. Experiences with phosphorylated yeast proteins have demonstrated that protein recovery can be improved, reducing nucleic acids. Functional properties such as water solubility, water-holding capacity, and thickening properties can be enhanced, whereas the emulsifying activity is better than soy protein isolate and equivalent to sodium caseinate.

Although dried whole cells have limited functional properties, they are frequently used as flavor-carrying agents and food binders. Dried yeast cells can act as oil-in-water emulsion stabilizers. (*See* **Emulsifiers:** Phosphates as Meat Emulsion Stabilizers; Uses in Processed Foods.)

The major market for microbial biomass is as a flavor enhancer for meat products, soups, gravies, barbecues, sauces, salad dressings, seasonings, and any food with savory, cheesy, or meaty flavors (flavor notes associated with the fifth basic flavor called 'umami'), including pizzas, snacks, chips, etc. In fact, yeast protein hydrolysates, autolysates, and extracts have long been used as food flavorings.

Prospects

SCP has to compete with other protein sources such as soy bean, fish meal, and milk proteins. It has been widely demonstrated that in proteins to be used as additives or to be incorporated into a food, the most important factors to be considered are their functional properties and price; therefore, these are the main challenges that SCP has to face. Unfortunately, production and isolation of protein from microbial biomass are rather expensive because they are capital- and energy-intensive. Its broad utilization has been limited for economical reasons. However, autolysates or hydrolysates prepared mainly from yeasts have gained a wide acceptability as functional food ingredients. In addition, recent biotechnological advances such as high-cell-density fermentations, more efficient downstream operations, and the possibility to genetically improve microorganisms could re-evaluate SCP.

High-cell-density fermenters have made possible a considerable reduction of equipment size, energy savings, very high productivities, and cheaper downstream processing. For instance, direct spray-drying from the fermenter is possible. These kinds of improvements could bring SCP to a competitive level. Currently, Philipps Petroleum is producing Provesta and Provesteen, trade marks for SCP from different strains, using this process (see Table 1).

The use of fed-batch fermentations has not been fully explored for SCP production. This technique is widely used for the production of bakers' yeast, increasing yields and avoiding ethanol accumulation. Some reports applying this strategy to yeast SCP production have demonstrated a yield increase up to 70%. The possibility of implementing continuous fermentation technologies to improve productivity has been explored for the production of SCP. A commercial process operated by Société des Alcohols du Vexin (SAV) in France is currently using a continuous culture for SCP production from whey.

Genetic engineering has focused on the possibility of improving substrate utilization. The first modified microorganism utilized in an industrial process and the largest-scale application for genetic engineering is a strain of *Methylophilus methylotrophus* developed by ICI in 1977. The improved strain, grown on methanol, is able to utilize the ammonium ion as a nitrogen source more efficiently than the wild strain, saving 1 mol of ATP per mole of ammonium assimilated, with an increase in efficiency of 4–7%.

Considerable research based on genetic engineering has been carried out to obtain yeasts able to utilize a broader range of carbon sources, such as lactose,

starch, cellulose, xylose, and chitin, in order to use cheaper and more available substrates.

Another interesting issue is to obtain easily and/or genetically controlled autolytic cells, for which different approaches have been followed, including the selection of mutants with weaker cell walls; also, the introduction of genes coding for lytic enzymes to facilitate cell disruption, and genes coding for nuclease activity to reduce nucleic acids content, have been explored. Recently, a system was developed that consists of a regulated promoter that controls two genes involved in cell wall biogenesis, which led to the possibility of triggering cell lysis of *S. cerevisiae* by the addition of methionine; this system can be extended to other yeast species. Also, an enhancement of the nutritional value by modification of the amino acid content, or obtaining proteins with better functional properties looks promising. All these possibilities have been investigated, but practical results will take longer to be implemented.

Another approach to improve the economics of SCP would be to produce it as a low cost byproduct in multiproduct microbial processes, such as during processing of food wastes to reduce the biological oxygen demand, or as a byproduct from high-added-value enzymes. Another possibility is the recovery of nucleic acids from bacteria and yeast biomass to produce 5′-nucleotides, which can be used as flavor enhancers.

See also: **Single-cell Protein**: Algae; **Yeasts**

Further Reading

Batt CA and Sinskey AJ (1987) Single-cell protein: production modification and utilization. In Knorr D (ed.) *Food Biotechnology*, pp. 347–362. New York: Marcel Dekker.

de la Torre M and Flores-Cotera LB (1993) Proteínas Unicelulares. In: García-Garibay M, Quintero-Ramirez R and López-Munguía A (eds) *Biotecnología Alimentaria*, pp. 383–397. Mexico City: Editorial Limusa.

Goldberg I (1985) *Single-cell Protein*. Berlin: Springer.

Guzmán-Juarez M (1983) Yeast protein. In: Hudson BJF (ed.) *Developments in Food Proteins – 2*, pp. 263–291. London: Elsevier Applied Science.

Harrison JS (1993) Food and fodder yeasts. In: Rose AH and Harrison JS (eds) *The Yeasts*, 2nd edn, vol. 5, pp. 399–433. London: Academic Press.

Litchfield JH (1983) Single-cell proteins. *Science* 219: 740–746.

Litchfield JH (1991) Food supplements from microbial protein. In: Goldberg I and Williams R (eds) *Biotechnology and Food Ingredients*, pp. 65–109. New York: Van Nostrand Reinhold.

Moo-Young M and Gregory K (1986) *Microbial Biomass Proteins*. London: Elsevier.

Reed G (1982) Microbial biomass, single cell protein, and other microbial products. In: Reed G (ed.) *Prescott & Dunn's Industrial Microbiology*, 4th edn., pp. 541–592. Westport CT: AVI.

Reed G and Nagodawithana TW (1991) *Yeast Technology*, 2nd edn. New York: Van Nostrand Reinhold.

Rolz C (1990) Características Químicas y Bioquímicas de la Biomasa Microbiana. *Archivos Latinoamericanos de Nutrición* 40(2): 147–193.

Rose AH (ed.) (1979) *Microbial Biomass. Economic Microbiology*, vol. 4. London: Academic Press.

Schlingmann M, Faust U and Scharf U (1984) Bacterial proteins. In: Hudson BJF (ed.) *Developments in Food Proteins – 3*, pp. 139–173. London: Elsevier Applied Science.

Slaughter *See* **Meat**: Sources; Structure; Slaughter; Preservation; Eating Quality; Sausages and Comminuted Products; Analysis; Nutritional Value; Hygiene; Extracts

SLIMMING

Contents

Slimming Diets

D L Miller, Central Soya Company Inc., Fort Wayne, IN, USA

Background

The prevalence of being overweight or obese in America is staggering. An accepted measure of overweight status and obesity is the body mass index (BMI). BMI is weight (kg) divided by height-squared (m^2). A BMI of 30 is generally accepted as the lower limit for obesity, whereas individuals with BMIs between 25 and 30 are considered overweight (Table 1). A recent survey has estimated that more than 33% of Americans meet the criteria for being obese, whereas 50% of American adults meet the criteria for overweight/obese.

It is not surprising, then, that a large number of adults are attempting to lose weight. Consumers are bombarded with endless commercially available weight-loss programs – many promising fast and easy weight loss. Some commercial programs recommend a low-fat diet, whereas others recommend low-carbohydrate diets, and protein intake varies among these programs. Still others advocate single foods, juices, or herbal remedies (Table 2). In order to review this topic in a concise method, weight-loss diets will be discussed, based on their macronutrient composition.

Concept of Macronutrient Composition

Foodstuffs are generally comprised of three major 'macro' nutrients: protein, carbohydrate, and fat (see Figure 1). Foods are not solely made of one macronutrient, but are composites of all macronutrients in various proportions. Macronutrients vary in their energy yield per gram: protein and carbohydrate both yield 17 kJ (4 kcal) g^{-1}, whereas fat is a much denser source of energy, yielding an average of 38 kJ (9 kcal) g^{-1}. Alcohol is also a dietary source of energy, with 29 kJ (7 kcal) g^{-1} but alcohol will not be discussed in this review. In terms of energy (kJ/kcal), the typical American diet comprises 10–15% protein, 50–55% carbohydrate, and 35–40% fat.

Energy Equation

In order to discuss the effects of alterating the macronutrient composition of a food on body weight, it is useful to discuss the basic concept of energy balance. To maintain a stable body weight, the total energy intake must equal the total energy loss or output, as shown by the following equation:

$$E_{\text{input}} - E_{\text{output}} = \text{E}_{\text{stored}}.$$

The simple interpretation of this equation, as it relates to human obesity, is that if more energy is consumed by an individual than is expended by that individual, the excess energy will be stored. This model becomes more complicated when the ingested and stored energy is subdivided into the three macronutrients. A normal-weight human being typically stores 585 760 kJ (140 000 kcal) of fat in adipose tissue, 100 416 kJ (24 000 kcal) of protein, mainly in muscle mass, and 3347 kJ (800 kcal) as carbohydrate in the form of glycogen (stored mainly in the liver, kidney, and muscles) plus glucose that circulates in the blood. In humans, stores of protein and carbohydrate remain rather constant, regardless of the level of intake, whereas theoretically, fat can be stored in an almost unlimited capacity.

Macronutrient Composition and Weight Loss

High-protein Diets

Protein makes up a small and relatively constant proportion (14–18%) of total energy intake of adults. There is much debate in the scientific community about the role of protein intake in weight loss. Some researchers theorize that, because dietary protein is essential for physiological processes (growth, tissue maintenance, enzyme production, hormonal synthesis, etc.), humans and other animals consume foods to deliver the protein needed to sustain these processes. As it relates to weight control, if an individual chooses high-carbohydrate and/or high-fat foods

Table 1 Body mass index

	Weight (lb)																														
Height (ft)	*100*	*105*	*110*	*115*	*120*	*125*	*130*	*135*	*140*	*145*	*150*	*155*	*160*	*165*	*170*	*175*	*180*	*185*	*190*	*195*	*200*	*205*	*210*	*215*	*220*	*225*	*230*	*235*	*240*	*245*	*250*
5′ 0″	20	21	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49
5′ 1″	19	20	21	22	23	24	25	26	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	43	44	45	46	47
5′ 2″	18	19	20	21	22	23	24	25	26	27	27	28	29	30	31	32	33	34	35	36	37	37	38	39	40	41	42	43	44	45	46
5′ 3″	18	19	19	20	21	22	23	24	25	26	27	27	28	29	30	31	32	33	34	35	36	37	37	38	39	40	41	42	43	43	44
5′ 4″	17	18	19	20	21	21	22	23	24	25	26	27	27	28	29	30	31	32	33	33	34	35	36	37	38	39	39	40	41	42	43
5′ 5″	17	17	18	19	20	21	22	22	23	24	25	26	27	27	28	29	30	31	32	32	33	34	35	36	37	37	38	39	40	41	42
5′ 6″	16	17	18	19	19	20	21	22	23	23	24	25	26	27	27	28	29	30	31	31	32	33	34	35	36	36	37	38	39	40	40
5′ 7″	16	16	17	18	19	20	20	21	22	23	23	24	25	26	27	27	28	29	30	31	31	32	33	34	34	35	36	37	38	39	39
5′ 8″	15	16	17	17	18	19	20	21	21	22	23	24	24	25	26	27	27	28	29	30	30	31	32	33	33	34	35	36	36	37	38
5′ 9″	15	16	16	17	18	18	19	20	21	21	22	23	24	24	25	26	27	27	28	29	30	30	31	32	32	33	34	35	35	36	37
5′ 10″	14	15	16	17	17	18	19	19	20	21	22	22	23	24	24	25	26	27	27	28	29	29	30	31	32	33	34	34	34	35	36
5′ 11″	14	15	15	16	17	17	18	19	20	20	21	22	22	23	24	24	25	26	26	27	28	29	29	30	31	31	32	33	33	34	35
6′ 0″	14	14	15	16	16	17	18	18	19	20	20	21	22	22	23	24	24	25	26	26	27	28	28	29	30	31	31	32	33	33	34
6′ 1″	13	14	15	15	16	16	17	18	18	19	20	20	21	22	22	23	24	24	25	26	26	27	28	28	29	30	30	31	31	32	33
6′ 2″	13	13	14	15	15	16	17	17	18	19	19	20	21	21	22	22	23	24	24	25	26	26	27	28	28	29	30	31	31	32	32
6′ 3″	12	13	14	14	15	16	16	17	17	18	19	19	20	21	21	22	22	23	24	24	25	26	26	27	27	28	29	29	30	30	31
6′ 4″	12	13	13	14	15	15	16	16	17	18	18	19	19	20	21	21	22	23	23	24	24	25	26	26	27	27	28	29	29	30	30

Healthy; overweight; obese

Table 2 Various diet plans

Diet plan	*Type of diet*	*Comments*
Cabbage Soup/All You can Eat Soup Diet	Single-food, low-fat, low-carbohydrate, very-low-energy density	Based on continued consumption of a buillion-based cabbage and vegetable soup, which is not highly palatable; other foods permitted include fruit, but not bananas
Dr. Atkins Diets	Low-carbohydrate	Ketogenic diet, which encourages free intake of fat and protein foods with severe carbohydrate restriction (< 20 g per day). Compliance can be checked by urine ketostix
Grapefruit Juice Diet	Single-food, high-water	Grapefruit juice is touted as a energy burning catalyst with the specific foods recommended. No evidence exists regarding the validity of these claims
Russian Airforce Diet	Low-fat, low-carbohydrate, low-energy	Diet includes red meat, eggs, fruit, and salads with no starch sidedishes
Scarsdale Diet	Low-fat, low-carbohydrate, low-energy	Lean meats, high-water fruits (melons and citrus), and salads, no starch sidedishes
The Zone Diet	Low-carbohydrate	Touted to stimulate the ideal amount of insulin by consuming macronutrients in the ratio of 40% carbohydrate, and 30% each of protein and fat
Liquid Diets	Low-fat, portion-controlled	Simple to use and may be of use for those who have difficulty with portion control. Monotonous food intake may be difficult to maintain
Dieting, Italian Style	Portion-control, increased-fat	This diet advocates determining your individual energy needs (via table of metabolic test) and consuming a diet high in pasta, bread, cheese, avocado and olive oil. Portion-control is necessary on this diet

with little protein content, the individual would need to consume a larger volume of food to obtain the necessary amount of protein.

Indeed, researchers have found that high levels of protein (> 25% of energy) have been shown to have a suppressive effect on energy intake in humans and animals. In humans, the consumption of high-protein foods has been shown to acutely reduce subsequent energy intake relative to low-protein foods. However, the protein intake of adults in Western populations is generally higher than the recommended dietary levels. The US recommended daily amount (RDA)

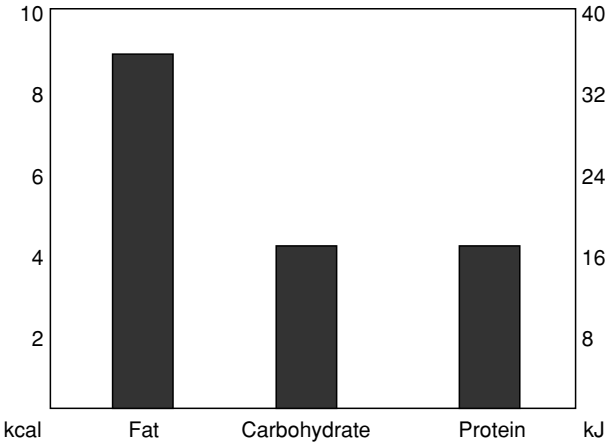

Figure 1 Energy content of macronutrients.

for protein is 63 g for males and 50 g for females; therefore, most American males eat 120% of their RDA for protein, and American females 158%. This adds confusion as to whether increases in protein intake are justified, and some have questioned whether a chronic high intake of protein could cause kidney damage or even osteoporosis. Currently, there are no compelling data to suggest a link with such damage.

The question of optimal protein intake during weight gain is controversial; however, alterations in the levels of fat and carbohydrate, which supply the bulk of total energy intake, are likely to have greater effects on weight loss. Some diets, which claim to be 'high-protein,' are actually based more on lowering or eliminating carbohydrate intake and replacing it with protein and fat energy. These diets will be discussed below.

Low-carbohydrate Diets

In the 1960s and 1970s, the 'bad nutrient' was carbohydrate or, more specifically, sugar. In his 1972 text, Craddock states, 'sugar is indeed the great enemy.' During this time, the low-carbohydrate diet grew out of the frustration of physicians who had documented that just telling patients to 'eat less' or restrict calories was not producing positive results. Thus, the notion of restricting a particular macronutrient, instead of calories, with free access to other macronutrients was proposed. Reducing protein intake was illogical, as protein represents a relatively small proportion of the total diet. At this time, before the advent of fat replacers and low-fat food technology, a diet low in fat was very unpalatable. This was further reinforced by the fact that protein and fat both have dietary requirements, whereas carbohydrate does not. Reducing carbohydrate intake from 350 g per day (average intake) to 50–60 g per day, therefore, seemed to be the logical nutrient to restrict as a reasonably palatable diet could be designed and maintained with 50–60 g of carbohydrate per day.

The low-carbohydrate diet encouraged participants to eat as much as they wanted of foods low in carbohydrate (meat, fish, cheese, butter, and green leafy vegetables and 330–670 ml of whole milk per day). It was assumed that the quantity of these foods consumed would be similar to the typical intake and, with the elimination of the carbohydrate energy, produce a deficit of 3360–4200 kJ (800–1000 kcal) per day. This energy reduction thus shifted the individual into a negative energy balance in which increased levels of fatty acids would be oxidized – or burn fat – and result in a significant weight loss.

Ketogenic diets An extreme form of the low-carbohydrate diet is a regimen almost devoid of carbohydrate. This diet encourages *ad libitum* consumption of meats, cheeses, eggs, and butter. Low-starch (mainly water-containing) vegetables (lettuce, peppers, tomatoes, onions, zucchini) are permitted in limited amounts, depending on the carbohydrate content. Sugars and other monosaccharides and complex carbohydrate sources (breads, grains, pasta, rice) or any digestible carbohydrate that would in turn provoke an insulin response are strictly restricted.

On this diet, glycogen stores are depleted within 24 h, and within 48 h the individual becomes ketotic – a metabolic state in which glucose is no longer available in adequate quantities to serve as a fuel. In ketosis, large amounts of the ketone bodies, acetoacetic acid, beta hydroxybutyrate, and acetone, are produced to be utilized as fuel. These compounds are produced in the absence of glucose/insulin as a method to spare glucose for use by the brain and other solely glucose-utilizing tissues. In the synthesis of ketone bodies, large amounts of fatty acids from adipose tissue are mobilized for use as fuel – utilizing stored fat as energy. In a starvation mode, which the body believes it is in with a ketogenic diet, the brain may obtain 30–40% of its energy from ketone bodies. Ketosis also occurs in uncontrolled diabetes and, if untreated, can lead to ketoacidosis and coma.

The evidence for the effectiveness of very low-carbohydrate diets is dated as far back as 1932, when the principle of reducing weight by a diet 'containing much fat but little carbohydrate' was introduced. However, not until the 1950s, following a series of experiments did low-carbohydrate/high-fat weight-loss diets begin to emerge in the popular press. In these experiments, the researchers fed overweight participants, isocaloric (4200 kJ or 1000 kcal) per

day diets comprising 90% protein, carbohydrate, or fat. Participants lost more weight on the low-carbohydrate diets (high in fat and protein) than the high-carbohydrate diet (low in fat). The mean weight loss on the low-carbohydrate diet was 2.33 kg over 5–9 days, but there is great intersubject variability with weight losses ranging from 0.6 to 3.2 kg. Also, the duration of the diet periods (5–9 days) is too short to adequately assess this diet composition on long-term weight loss. Further, it is well established that each gram of muscle glycogen that is converted to glucose for use as fuel yields 3 g of water. Thus, it is theorized that during the initial stages of a low-carbohydrate diet, when glycogen stores are being depleted, much of the weight loss is from water loss.

Some follow-up studies compared low-carbohydrate diets with low-fat diets and investigated the amount of weight and water loss on each regimen. One study showed that there was no significant difference in the rates of weight loss on either diet. In addition, when these diets were interchanged, deviations from the weight-loss curve occur mainly as a result of fluid balance. Another study found that a high-fat/low-carbohydrate diet resulted in weight loss, but the weight loss ceased after a few days and was explained largely by fluid loss. However, a study in 1981 measured sodium, potassium, and water losses on diets low or high in carbohydrates over a 28-day period and found conflicting results. In this study, although sodium and potassium losses were greatest on the low-carbohydrate diets (at least until day 14), water losses showed no significant differences at any time period, and the low-carbohydrate diet produced a significantly greater weight loss than the high-carbohydrate diet (12.5 ± 0.9 kg and 9.5 ± 0.7 kg).

Recent studies The 1990s saw a resurgence in the popularity of low-carbohydrate diets, and with it, new research has emerged. In a 1996 study, 43 obese participants were randomly assigned to consume diets of equal energy but which differed in macronutrient composition:

- low-carbohydrate diet: 32% protein, 15% carbohydrate, 53% fat;
- low-fat diet: 29% protein, 45% carbohydrate, 26% fat.

Similar to the studies done in the 1960s, this study found that weight, plasma glucose, cholesterol, and high-density-lipoprotein cholesterol changes were similar on both diets. Interestingly, however, the higher-protein/fat, lower-carbohydrate diet was found to be more beneficial on plasma insulin and triacylglycerides (triglycerides). It is important to note that this diet was not ketogenic, as were diets fed to the participants of the earlier studies. These results were consistent with other, more modern studies.

The current view of ketogenic low-carbohydrate diets is that at least 100 g per day or more of carbohydrate is recommended both to spare protein and to avoid large shifts in weight resulting from changes in water balance and electrolyte loss. When carbohydrate levels are below 100 g per day, insulin levels fall, and protein is catabolized to provide the glucogenic amino acids. Most diets providing under 100 g per day of carbohydrate or approximately 3360 kJ (800 kcal) per day are ketogenic. Also, dieters may have negative side-effects, including fatigue, postural hypotension, a fetid taste in their mouths (from increased acetone production), and elevated serum uric acid levels, and may be more subject to negative nitrogen balance. In addition, because consuming a low-carbohydrate diet often means consuming a diet higher in fat, cardiovascular health is of concern. A high intake of saturated fat, found mainly in meat and dairy products, has been linked to increased serum low-density lipoprotein cholesterol levels and is thus thought to be atherogenic; however, the beneficial effects of low-carbohydrate diets on insulin and triglyceride levels should not be dismissed. High triglyceride levels have begun to receive increased attention in medical research as a serious cardiovascular disease risk factor. Low carbohydrate diets which obtain the majority of protien from vegetable protein (nuts, soy, legumes) should be more healthful overall, but more research is needed.

Low-glycemic index diets Recent research focus has shifted away from low-carbohydrate diets to 'low-glycemic index' diets. The glycemic index was proposed by Jenkins to characterize the rate of carbohydrate absorption after a meal and is defined as the area under the glucose response curve after consumption of 50 g of carbohydrate from a test food divided by the area under the curve after consuming a control food (usually glucose). The glycemic index of foods became of interest as dieters following a low-fat regimen began replacing fat calories with carbohydrate calories. Many of the foods chosen to replace high-fat foods were high-glycemic index foods (Table 2). To compound this situation, ingested fat serves to slow the rate of absorption of concomitantly ingested carbohydrate. Thus, some researchers believe that low-fat diets have resulted in an increased intake of high glycemic index foods that are more rapidly absorbed, thus causing a functional hyperinsulinemic response. This hyperinsulinemic state may promote weight gain by preferentially directing nutrients

away from oxidation in muscle and toward storage in fat.

A number of studies have examined the effects of glycemic index on appetite and food intake regulation in humans. Two studies found lower blood glucose levels and a slower return of hunger after low-glycemic index meals compared with high-glycemic index meals. Another study investigated the hormonal response to high- and low-glycemic index foods. This study provided obese teenage boys low-, medium- and high-glycemic index meals. Relatively high insulin and low glucagon concentrations were observed following the high-glycemic index meal. Such hormonal changes would be expected to promote uptake of glucose in muscle, liver, and fat tissue, restrain the release of glucose, and inhibit lipolysis (release of fatty acids for fuel). In addition, a 'reactive hypoglycemic' period was observed following the high-glycemic index meal – high insulin levels effectively cleared the abundant blood glucose and lowered free fatty acid concentrations 3–5 h after the high glycemic index meal compared with the low-glycemic index meal. In this metabolic state, stimulation of appetite would be expected. Indeed participants consumed more energy after the high-glycemic index meal compared with the medium- or low-glycemic index meals. Thus, hormonal responses to a high-glycemic index diet appear to lower circulating levels of metabolic fuels, stimulate hunger, and favor storage of fat, events that may promote excessive weight gain if excess energy is consumed. A diet of low-glycemic index foods would include abundant quantities of vegetables, fruits, and legumes, moderate amounts of protein and healthful fats, and decreased intake of refined grain products, potato, and concentrated sugars.

Low-fat Diets

In contrast with the dated research done on low-carbohydrate diets, research on the effects of low-fat diets is more recent. Until the mid-1980s, there were few studies examining the effects of variations in the level of fat in the diet on energy intake and body composition. This was because, before this time, there was relatively little emphasis on the role of dietary fat in obesity and related disease states. Research from the 1980s suggested that the oxidation of protein and carbohydrate is closely tied to their intake, whereas fat oxidation is not closely correlated with intake. This suggests that fat, unlike the other macronutrients, does not promote its own oxidation, and to maintain fat balance, fat should be consumed in small amounts or coupled with behaviors that promote fat oxidation, such as aerobic exercise.

This work, coupled with several low-fat feeding trials (discussed below), effectively shifted the emphasis from sugar to fat as the 'bad' nutrient. Food manufacturers followed this trend by producing a large number of products that were low-fat or fat-free versions of traditionally higher-fat foods. The development of a number of functional, fat-replacing ingredients allowed food technology to develop palatable low-fat/no-fat foods in a range of food categories. However, although there have been a number of studies on the role of dietary fat in the etiology of obesity, there have been few well-controlled studies on diet composition and weight loss. Surprisingly, there are almost no data on the effects of fat-replaced foods on weight control.

Three large, controlled, laboratory studies from the 1980s and early 1990s investigated the effects of covert manipulations of the fat content of a diet consumed *ad libitum* on energy intake and body weight. In a study in 1983, participants were fed a low-fat diet comprising traditionally low-fat foods (fruits, vegetables, and grains) and a high-energy-density diet for 5 days. Participants could eat the foods freely or *ad libitum*. Participants reduced their energy intake by half when eating the low-fat diet compared with the high-fat diet [12 552 kJ (3000 kcal) vs. 6569 kJ (1570 kcal) per day]. However, no data were reported regarding weight loss.

Two other studies, conducted at Cornell University, are often cited in both scientific and popular literature as proof that *ad libitum* consumption of low-fat foods can reduce fat intake and lead to weight loss. In the first of these studies, overweight participants were each fed three diets: low-fat 15–20%, medium-fat 30–35%, and high-fat 45–50% of energy. Energy consumption increased as the level of fat of the diet increased, with the low-fat diet resulting in an average reduction of 2205 kJ (527 kcal) per day compared with the high-fat diet. Despite these daily differences, the diets did not produce any statistically significant weight changes over the 2-week periods. The second Cornell study was similar but extended the intervention period to 11 weeks. In this study, participants consumed an average of 1201 kJ (286 kcal) per day less on the low-fat diet than on the high-fat diet. Weight loss was significantly greater on the low-fat diet than on the high-fat diet, but weight loss occurred on both diets, making it difficult to conclude that the low-fat diet was more effective for weight loss. In addition, although statistically significant, the actual weight losses are quite small. This evidence indicates that low-fat diets consumed *ad libitum* may be somewhat useful for reducing the amount of fat consumed – important for cardiovascular health, but their effectiveness for weight loss is less impressive.

Another study examined the weight loss potential of restricting fat plus caloric intake compared with restricting just fat intake. Over the 16–20 weeks of treatment, the group with energy restriction lost significantly more weight (males 11.8 kg, females 8.2 kg) than the group only modifying fat intake (males 8.0 kg, females 3.9 kg). There was also a significantly greater loss of body fat in the energy restriction group. The authors concluded that it is more effective to instruct overweight persons to restrict both fat and energy intake than it is to restrict only fat intake as a weight loss strategy. Finally, a study from the US Department of Agriculture investigated the effects of varying the fat level of a reduced-energy diet on body weight and composition. In this study, the fat content of the diet of eight overweight males fed at 50% of maintenance was manipulated. No significant differences were found between the low-fat and high-fat diets in the extent or composition of body-weight loss. Body weight decreased by 5.2 kg for the high-fat group and 5.0 kg for the low-fat group. Subjects in the low-fat group, however, did lose more fat and less fat-free mass than did the high-fat groups. This research implies that weight loss is not dependent on the composition of the diet as long as the diet is reduced in total energy.

These studies taken together indicate that, although fat restriction may be of some use for weight loss, restricting total energy intake is the most important factor. The results regarding energy deficit and diet composition are consistent with other research reports that state that the most significant factor determining the amount and rate of weight loss is the degree of negative energy balance achieved. It may be that low-fat diets are most useful in preventing excess energy intake and obesity. There has never been a single study showing an advantage of high-fat diets over low-fat diets in reducing energy intake and preventing obesity. For the same reasons, fat restriction may also be beneficial as part of a weight-loss maintenance program.

Low-energy-density Diets

Another impediment to weight loss is the abundant availability of an energy-dense diet in Western culture. Energy density is a measure of the amount of energy in a given weight of food. It is reasonable to assume that when foods are more heavily packed with energy, even small portions can produce large energy intakes and increase the probability of occurrence of periods of positive energy balance. What may be key is reducing the overall energy density of the diet.

Energy-dense foods include those that are high in fat, those with large amounts of added sugars or processed flours, and those low in water. Some fat-replaced foods, sweet, and other snacks can have large amounts of added sugars/processed flours, making them low in fat, but also energy-dense. Because both high-fat foods and high-glycemic index foods are energy-dense; choosing foods less dense in energy may bring these philosophies together. Energy density is affected by the fat content of food, and processed flours contribute to energy density because large amounts of these nutrients can be incorporated into foods and not diluted with water, as in naturally occurring foods. Because it increases the weight of food but not the energy content, the amount of water in a food also greatly affects energy density. Fiber in foods holds a large amount of water and can provide a sensation of fullness in the gut.

The studies described in the low-fat diet section also illustrate the utility of a diet lower in energy density. Individuals in those studies tended to eat a constant weight of food. When the food consumed was lowered in energy-density, the same weight of food supplied fewer calories. It has been demonstrated that individuals tend to eat a constant weight of food, rather than eat a constant amount of any of particular macronutrient or amount of energy. Another study allowed lean women to consume *ad libitum* three different diets that varied in energy density but had constant amounts of carbohydrate, fat, and protein. The women also tended to eat the same amount or quantity of food but consumed 30% less energy on the low-energy-density diet.

In selecting foods appropriate for a low-energy-density diet, it is difficult to reduce the energy density of the diet without decreasing fat and added sugar intake and increasing dietary fiber intake, bringing together some principles of the low-fat and low-glycemic index philosophies.

Conclusions

It is clear, with the large number of individuals meeting the overweight and obese criteria at this time and attempting to diet, that weight loss is not simple or easy. It is also clear that no one dietary principle is applicable to all individuals. It may be that genetic influences such as those that influence fuel utilization, glucose metabolism, fat storage, etc., may make the restriction of one macronutrient result in a more substantial weight loss for one person while being less effective for another. The impact of genetic differences on the effectiveness of pharmaceutical products has received great attention recently, and such differences are being explored from a nutritional perspective as well. It may be that future dietary treatment for weight loss will be based (at least partially) on

one's individual genome to optimize diet manipulation effects and maintain/restore maximal health.

At present, the most beneficial aspect of altering the composition of the diet in terms of weight control may be that the alteration simply results in less energy being consumed. The data presented here, however, indicate that diet-composition changes alone are insufficient for significant weight losses. Weight loss is dependent upon creating a negative energy balance, which can be achieved in part by diet-composition changes but also must include reductions in overall energy intake.

Because eating patterns, personal habits, and food likes and dislikes vary so greatly between individuals, an individual seeking a weight-loss plan must decide which approach to reducing energy intake is best suited for their lifestyle. If an intake pattern can be established and maintained in which less energy is being consumed than oxidized, weight loss will occur. Weight loss will generally occur in proportion to the degree of energy restriction and the duration of that energy restriction.

The next level of concern regarding weight loss is the secondary effects of the weight loss diet on health and whether that diet is palatable and sufficient in nutrients so that it can be maintained long enough to produce the desired weight loss. From a health standpoint, reducing saturated fat remains supported by the literature, whereas following a diet very low in fat may not be, especially if healthy fats (such as monounsaturated fats and omega-3 fatty acids) are restricted. In addition, fat should not be replaced with high-glycemic index foods. The data for a hyperinsulinemic response following ingestion of high-glycemic index foods and the subsequent effects on energy storage and appetite are compelling. Ketogenic diets, however, are associated with few health benefits, other than reduced blood glucose and insulin – which, most likely, will not be maintained once carbohydrates are reintroduced.

So what is the healthiest method by which to reduce energy and lose weight? The concept of reducing the energy-density of the diet by decreasing consumption of high-fat/high-energy foods (cheese, bagels, butter, fried foods, some dairy products) and increasing consumption of low-energy dense foods (vegetables, fruits, legumes, and high-fiber foods) may be the best prescription for overall health. Health effects of increased intake of dietary fiber include reduced blood glucose levels and cholesterol as well as a balanced gastrointestinal function. The addition of healthy fats in moderate amounts should also be considered; however, even healthy fats are high in energy – thus, smaller portion sizes may be needed to reduce energy when such fats are included. A low-energy-density diet should result in a reduced energy intake and produce a steady weight loss unless unusually large portions are consumed.

See also: **Carbohydrates**: Classification and Properties; **Energy**: Measurement of Food Energy; Intake and Energy Requirements; Measurement of Energy Expenditure; Energy Expenditure and Energy Balance; **Fats**: Requirements; **Glucose**: Glucose Tolerance and the Glycemic (Glycaemic) Index; **Ketone Bodies**; **Snack Foods**: Range on the Market

Further Reading

Atkins RC (1998) *Dr. Atkins New Diet Revolution*. New York: Avon Books.

Burton-Freeman B (2000) Dietary fiber and energy regulation. *Journal of Nutrition* 130: 272S–275S.

Hill JO, Melanson El and Wyatt HT (2000) Dietary fat intake and regulation of energy balance: implications for obesity. *Journal of Nutrition* 130: 284S–288S.

Ludwig D (2000) Dietary glycemic index and obesity. *Journal of Nutrition* 130: 280S–283S.

McCrory M, Fuss PJ, Saltzman E and Roberts S (2000) Dietary determinants of energy intake and weight regulation in healthy adults. *Journal of Nutrition* 130: 276S–279S.

Rolls BJ (2000) The role of energy density in the overconsumption of fat. *Journal of Nutrition* 130: 268S–271S.

Rolls BJ and Barnett RJ (2000) *The Volumetrics Weight Control Plan: Feel Full on Fewer Calories*. New York: Quill.

Rolls BJ and Hill JO (1998) *Carbohydrates and Weight Management*. Washington DC: ILSI Press.

Sears B and Lawsen B. *The Zone: A Dietary Road Map to Lose Weight Permanently: Reset Your Genetic Code, Prevent Disease: Achieve Maximum Physical Performance*. New York: HarperCollins.

Simopoulos AP and Robinson J (1999) *The Omega Diet: The Lifesaving Nutritional Program Based on the Diet of the Island of Crete*. New York: HarperCollins.

Metabolic Consequences of Slimming Diets and Weight Maintenance

E Rasio, Hôpital Notre-Dame, CHUM, Montreal, Quebec, Canada

Introduction

In many societies today, a plethora of individuals pursue weight reduction diets for a variety of reasons,

ranging from social pressures to medical concerns. Dieters usually fail to perceive that the phenomena underlying weight gain and weight loss may be quite different in complexity. Weight gain, as a consequence of excess energy intake, is a relatively simple process whereby fat stores are increased. To a large extent, it is independent of the quality of food eaten. Thus, whenever calories are ingested in excess of calories expended, be they in the form of sugars, proteins, lipids, or alcohol, the balance is stored in the most compact, weight-efficient form, as fat. The only exceptions to this rule are growth, pregnancy, muscle building, and rehabilitation following malnutrition, where excess calories, with adequate amounts of protein, can also be retained as protein.

Weight loss, as a consequence of insufficient energy intake, has a greater variety of effects which depend upon nutritional status, energy content of the diet, and balance of nutrients. Thus, an excess daily intake of 1250 kJ (300 kcal) in the form of 75 g of sugars, or 75 g of protein, or 35 g of lipids, will result, in most individuals, in a gain of 35 g of body fat. A deficient daily intake of energy of 1250 kJ may induce a considerable range of changes in body weight and body composition. In appropriately chosen conditions, the subject will lose 35 g of body fat, but electrolytes and protein shifts will often alter the body weight response so that not all of the weight lost will be in the form of fat. (*See* **Energy**: Energy Expenditure and Energy Balance.)

Regimens for Weight Control

The ideal diet should deplete body fat stores, which are excessive, without reducing the protein pool out of proportion to the small increment always associated with the obese state itself. For this to occur, the chosen diet should meet stringent qualitative and quantitative requirements. Brain metabolism requires 150 g of glucose per day: with the exception of prolonged fast when ketone acids can be oxidized, the need for glucose is mandatory and cannot be met by transformation of lipids, whether supplied by the diet or mobilized from body reserves. Glucose can be generated from protein, from the glycogen pool, or provided in the diet. Consequently, diets which provide at least 2500 kJ (600 kcal) per day in the form of carbohydrate, and a substantial portion of protein, will spare the waste of endogenous protein and mobilization of the glycogen pool. Many slimming diets do not satisfy the brain's need for glucose, and therefore deplete glycogen and protein stores. These diets may be grouped into first, total fasting and certain very-low-calorie diets, or 'modified fasts,' which do not contain enough glucose and protein precursors, and second, diets which supply adequate calories but mostly in the form of fat. With both groups of diets, body proteins are mobilized and converted by the liver to glucose, which is then oxidized by the brain. The loss of body protein entails a loss of weight which is 10 times that of adipose tissue for an equivalent calorie content. A total fast will induce a protein loss not too dissimilar to that of a 5000 kJ-per-day diet, composed mainly of fat. In both instances, the weight loss will be very important: the high-weight-low-calorie protein waste will overwhelm the low-weight-high-calorie changes of fat stores. Concurrent major fluid losses will compound the phenomenon. (*See* **Carbohydrates**: Digestion, Absorption, and Metabolism; Requirements and Dietary Importance; **Fats**: Digestion, Absorption, and Transport; Requirements.)

Dietary strategies other than a reduction in calorie content have not been thoroughly evaluated for the treatment of obesity. There is some evidence that macronutrients may act directly, by their nature rather than energy, on mechanisms which regulate body weight. Furthermore, the potential of micronutrients to affect fat storage has not been explored.

Behavior modification targeted at overeating is also advocated to help reduce body weight. Some obese individuals may suffer from specific behavioral syndromes characterized by excessive consumption of certain foods, at particular times, e.g., carbohydrate cravings have been associated with seasonal affective disorders.

The major obstacle to dietary and behavioral treatment of obesity is that the voluntary control of energy intake is difficult and requires strong motivation. This is particularly true in affluent societies where food is plentiful and varied. The success rate of dieting can be disappointing: studies published in the medical literature indicate that weight loss after 1 year is usually less than 10% of entry weight. However, most people who determine that they should control their weight, or are advised to do so, are never part of medical surveys. Thus the real success rate for populations are not known.

Metabolic Consequences of Weight-reducing Diets

Negative Nitrogen Balance

Whenever a diet, irrespective of its calorie content, does not have sufficient protein to maintain required rates of protein synthesis, or a sum of protein and carbohydrate to meet the brain's energy requirements, it will induce a negative nitrogen balance, with protein tissue waste. Obese individuals have an

increased lean body mass and a higher metabolic rate than control individuals of 'ideal' body weight. Therefore, protein waste will be tolerated to some extent during dieting. Unfortunately, there is no evidence that protein is lost from those tissues where it accumulated during weight gain, i.e., from adipose tissue as a consequence of cell hyperplasia and hypertrophy, and from muscle and supporting tissues as a response to increased gravity. On the contrary, the protein loss may occur from all tissues. The digestive tract and the heart are particularly prone to functional alterations. In conditions of excessive negative nitrogen balance, thinning of the skin and loss of brain tissue have been reported.

Water and Electrolytes Losses

Elimination of excess water is a common feature of weight loss, particularly during the early stages. A negative calorie balance reduces insulin secretion and has a diuretic effect which enhances weight loss beyond that expected from endogenous fat oxidation. The diuretic effect varies with the quantity and quality of energy consumed. In general, the lower the energy content as carbohydrate, the greater the diuretic effect. Concurrent loss of sodium and potassium in the urine reflects the contraction of the extracellular and intracellular water compartments. Cardiac arrhythmia, including ventricular fibrillation and death, have been reported, probably as a consequence of electrolyte changes in the serum and cells. The use of diuretics alone or in combination with diets, especially when markedly carbohydrate-restricted, is therefore dangerous or, at least, superfluous in the treatment of uncomplicated obesity. Before counteractive adaptive mechanisms are set in motion, water loss at the initiation of low-calorie-low-carbohydrate diets may induce an extra weight loss of 0.5 kg per day, or even more. (*See* **Water**: Physiology.)

Water loss will confound changes in body fat weight. In extreme situations, such as those created by regimens very high in calories and fat, protein waste and dehydration will induce weight loss while fat stores are actually increased. Other diets, along the same principles, will also reduce weight with little fat loss.

Lowering of Blood Pressure

Overeating stimulates the release of catecholamines and insulin. Consequently, renal reabsorption of sodium is increased, and higher blood volume and cardiac output may then raise blood pressure. Hypocaloric diets will reverse this sequence of events and frequently reduce the hypertension associated with obesity, without medication or sodium restriction. (*See* **Hypertension**: Physiology.)

Weight-reducing diets, in particular those low in carbohydrate, induce relaxation of arteriolar tonicity and lowering of diastolic blood pressure. This, in combination with fluid volume contraction mentioned above, is responsible for orthostatic hypotension, a frequent side-effect, characterized by the inability to redistribute blood from the lower limbs when standing. Salt supplementation can diminish such an effect, but not abolish it, as a new steady state is reached in which the extra sodium is simply excreted. (*See* **Hypertension**: Hypertension and Diet.)

Lowering of Blood Lipids and Glucose

Although most obese individuals are free of metabolic complications such as dyslipoproteinemias and non-insulin-dependent diabetes mellitus, many patients with these metabolic anomalies are obese. When energy and carbohydrate intake are excessive, high levels of triglycerides and low-density lipoprotein cholesterol may accumulate in plasma. Obesity is also often associated with a diminution of circulating high-density lipoproteins. These changes in plasma lipid composition will contribute to the development of atherosclerosis and increase the risks of cardiovascular disease. Hypocaloric diets can rapidly blunt or eliminate hypertriglyceridemia, by the plasma-clearing action of lipoprotein lipase, and hypercholesterolemia, by reducing both the endogenous and exogenous flow of atherogenic lipoproteins. Concomitantly, the level of high-density lipoproteins in the plasma increases. Weight reduction can therefore often help to control dyslipoproteinemia without the use of pharmacological agents. (*See* **Atherosclerosis**; **Lipoproteins**.)

Blood glucose abnormalities in obese subjects are frequent. They range from mild glucose intolerance to overt noninsulin-dependent diabetes mellitus. These anomalies of glucose homeostasis are characterized by the presence of normal-to-high levels of circulating insulin which are insufficient to overcome the resistance by target tissues to its biological effects. Excess plasma insulin is a potential, albeit still controversial, risk factor for atherosclerosis. One of the most impressive effects of energy-restricted diets is the improvement or normalization of the altered glucose homeostasis associated with obesity. In a matter of days, even without significant losses of body weight, the obese subject can be switched to a normal glucose metabolism. A negative calorie balance can rapidly restore insulin sensitivity in a dieting individual. The mechanisms of this spectacular metabolic regulation are not known. The treatment of diabetes in the obese requires oral hypoglycemic agents only when

energy-restricted diets have proved unsatisfactory. Insulin may also be required to control blood glucose levels, in spite of possible atherogenic changes of the plasma lipid profile.

Reduced Resting Metabolic Rate and Postprandial Thermogenesis

With the loss of significant amounts of body weight, oxidative metabolism decreases, as does energy expenditure. This occurs both at rest and following meals. At rest, most of the energy produced is utilized to maintain body temperature and vital processes. Following the ingestion of a meal, additional oxygen is required for mechanical work associated with digestion, and for chemical reactions involved in processing the nutrients. The energy cost of disposing of food in the body varies with the caloric content of the meal, its composition, and sapidity. It represents only a small percentage of the overall resting metabolic rate. The main factors responsible for slowing of the oxidative metabolism are hormonal adaptive changes. Faced with reduced energy intake, the body responds by lowering the circulating levels of hormones which stimulate energy expenditure. Thus, the activity of the sympathetic nervous system is diminished and catecholamine production is lowered. Furthermore, the conversion of the thyroid hormone thyroxine into the more powerful hormone triiodothyronine is also reduced. These changes combine their effects in reducing mitochondrial oxidation and, therefore, energy output. This adaptive sparing of body weight varies with the energy restriction and the composition of the diet. It may represent approximately 850 kJ (200 kcal) per day after 2–3 weeks of fasting, but less if the diet has energy, particularly in the form of carbohydrates. Finally, when significant glycosuria is present in obese diabetic patients, the rapid lowering of blood glucose levels at the start of a hypocaloric diet will reduce or abolish the glucose caloric output in the urine and slow down weight loss accordingly. On the other hand, exercise, by implementing energy expenditure and by blunting the drop in metabolic rate associated with a negative energy balance, enables the weight loss to continue unmitigated. (*See* **Hormones**: Thyroid Hormones; **Thermogenesis**.)

A more chronic effect of weight-reducing diets is the loss of significant amounts of body weight when lean tissue is also wasted. The ensuing cell atrophy will contribute to lowered oxygen consumption.

Patterns of Weight Loss

After the initiation of a hypocaloric diet, a combination of behavioral and metabolic mechanisms takes place wherein the rate of weight loss decreases with time, despite a constant dietary intake.

On the behavioral side, many dieters tend to reduce their energy expenditure as a reflex in the face of scarce energy intake. Probably the most common cause for plateaus in weight loss is that, even with adherence to appropriate pattern and composition of the diet, portion sizes tend to increase.

On the metabolic side, a number of predictable adaptive reactions slow the rate of weight loss. First, water and electrolyte losses are rapidly blunted by sensitive hormonal counterregulatory responses. These may involve, depending upon the degree of water loss, the liberation of antidiuretic hormone by the hypothalamus or the stimulation of the renin–angiotensin–aldosterone axis. These responses are more decisive when diuretics are prescribed with the diet. Second, the body adapts by shifting its oxidative metabolism from glucose to fat and ketone acids. The reduced need for glucose spares the energy-diluted protein tissue at the expense of the energy-dense fat stores. Thus, the rate of body weight loss declines to very near the theoretical level expected from conversion to fat as the primary source of energy utilization. Third, the oxidative metabolism diminishes *pari passu* with the activity of the thyroid gland and sympathetic nervous system. Fourth, the loss of lean body mass, with time, lowers the resting metabolic rate, thus reducing the energy deficit.

Occasional interventions in the course of dieting may also slow down the rate of weight loss. Rotating diets, depending on their sequence, may induce rapid shifts in water and protein pools. For example, when a 3500 kJ per day ketogenic diet is followed by a diet of the same calorie content but richer in carbohydrate and protein, the weight loss may be temporarily reduced and weight may even increase as a consequence of water and protein gains. This same diet would have the opposite predictable effects if administered *de novo*. Thus, the same hypocaloric diet is capable of inducing either weight and protein gain, or weight and protein loss in the same individual, depending on the previous metabolic balance at any given total body weight. These considerations emphasize the extent to which the dependence of the dieter upon body weight measurement as a source of feedback information can be misleading.

A less common but interesting situation is created when dieting individuals embark on exercise programs. This leads to loss of body energy, without the expected concurrent loss of weight. It is possible, for example, to lose 1 kg of body fat as a consequence of a negative calorie balance, and to gain 1 kg of tissue protein as muscle is built up. The body weight has not changed, but the individual has lost in the process

30 000 kJ of body energy. Recent studies have shown that a similar protein-sparing effect, with negative calorie balance, can be achieved by administration of appropriate doses of insulin or growth hormone. The usefulness of these hormonal manipulations in the dieting individual has not yet been established.

Maintenance of Weight Loss

Whenever the stability of body weight is disrupted, the hypothalamic counterregulation will tend to restore weight to its initial value. This phenomenon applies in general to any level of weight equilibrium, whether the subject is thin, of 'ideal' weight, or obese; it is also operative for whichever direction the change of weight takes place, be it a gain or a loss. It is therefore recommended, in order to overcome the constraints of the 'ponderostat,' that slimming strategies aim for a moderate rather than rapid rate of weight loss, interspersed with periods of weight stabilization.

Physical exercise has the ability, within a wide range of intensity, to maintain and possibly reestablish the regulatory functions of the brain concerning energy balance. Exercise programs should therefore complement the slimming diet, with the reasonable expectation that they will help the dieter to maintain the weight loss.

The temporary use of anorectic drugs has been advocated to promote weight maintenance between episodes of weight loss. The administration of thyroid hormones, to offset their lowered secretion during weight loss, or to raise their plasma levels to the high range of normal values during weight stabilization, is likely to increase nitrogen loss. Drugs designed to reduce the intestinal absorption of carbohydrates or lipids may prove beneficial in weight maintenance programs, when dieters tend to regain weight. Today, there is not sufficient evidence to recommend the regular use of any of these pharmacological agents.

Finally, therapies intended to improve the will to lose weight are an essential part of slimming programs. Group support therapy, as provided by associations such as Weight Watchers, or individual psychotherapy, may be successful in some individuals, while knowledge of nutrition principles seems to improve the incentive to lose weight and maintain the loss.

Benefits of Exercise

Exercise has many beneficial effects for weight maintenance. The most acknowledged effect, albeit not the most important, is that exercise increases calorie expenditure above the resting metabolic rate. The energy cost of exercise is often overestimated by dieters and easily annulled by food self-reward. For example, a brisk 1-h walk has approximately the equivalent energy of three small slices of bread.

Exercise, of some intensity and frequency, has anorectic properties that are probably mediated by changes in hypothalamic neurotransmitters and sex hormone production. With more moderate ranges of physical activity, as previously stated, the hypothalamic regulation of energy balance seems to be better able to adjust energy intake to energy expenditure on a day-to-day basis, thereby improving the chances of weight maintenance.

Exercise has powerful effects on fuel utilization in general and glucose metabolism in particular. Acute exercise accelerates the net rate of glucose utilization through the actions of humoral and hormonal factors on adipose tissue, muscle, and liver. The blood glucose-lowering effect of exercise may be rapid and important. It mimics the effect of intravenous injections of significant amounts of insulin, with the added characteristic that it manifests itself in most instances, whether the subject is insulin-resistant or not. Furthermore, chronic exercise counteracts insulin resistance associated with obesity, hence improving glucose and lipid metabolism. The metabolic properties of acute and chronic exercise should therefore be advantageously used in the treatment of obesity. (*See* **Anorexia Nervosa; Bulimia Nervosa; Obesity**: Etiology and Diagnosis; Fat Distribution; Treatment.)

See also: **Anorexia Nervosa**; **Bulimia Nervosa**; **Carbohydrates**: Digestion, Absorption, and Metabolism; Requirements and Dietary Importance; **Energy**: Energy Expenditure and Energy Balance; **Fats**: Digestion, Absorption, and Transport; Requirements; **Hormones**: Thyroid Hormones; **Hypertension**: Physiology; Hypertension and Diet; **Obesity**: Etiology and Diagnosis; Fat Distribution; Treatment; **Protein**: Requirements; **Water**: Physiology

Further Reading

Atkinson RL (1989) Low and very low calorie diets. *Medical Clinics of North America* 73: 203–215.

Atkinson RL and Hubbard VS (1994) Report on the NIH workshop on pharmacologic treatment of obesity. *American Journal of Clinical Nutrition* 60: 153–156.

Black D, James WPI and Besser GM (1983) Obesity. *Journal of the Royal College of Physicians of London* 17: 5–65.

Bray GA and Gray DS (1988) Treatment of obesity: an overview. *Diabetes and Metabolic Review* 4: 653–679.

Cowley DK and Sizer FS (1987) Fad diets, fact and fiction? *Nutrition Clinics* 2.

Flier JS and Foster DW (1999) Eating disorders: obesity, anorexia nervosa, and bulimia nervosa. In: Wilson JD, Foster DW, Kronenberg HM and Larsen PR (eds)

Williams Textbook of Endocrinology, 9th edn., pp. 1061–1097. Philadelphia: Saunders.

Howard AN, Bray GA, Novin D and Bjorntorp P (eds) (1981) Proceedings of a symposium on obesity and hypertension. *International Journal of Obesity* 5(suppl 1).

Garrow JS (1978) *Energy Balance and Obesity in Man*. Amsterdam: Elsevier.

Olson RE (ed.) (1986) Diet and behavior: a multidisciplinary evaluation. *Nutrition Reviews* 44.

Wurtman RJ and Wurtman JJ (eds) (1987) Human obesity. *Annals of the New York Academy of Sciences* 499.

SMOKED FOODS

Contents

Principles

L Woods, Woodside Consulting, Camberley, UK

Introduction

The foods most commonly smoked are meats, fish, shellfish, and cheese, although some snack foods and nuts are smoke-flavored. Meats or fish grilled over an open fire are exposed to smoke during cooking, but are not normally considered to be smoked foods as such. In general, smoke is produced by substantially raising the temperature of wood, and limiting the air supply so as to prevent combustion but allow destructive distillation.

Smoke flavors and essences may be prepared by condensation of smoke obtained by pyrolysis of wood in a limited supply of air, when the product is known as pyroligneous acid. The initial condensate separates into an aqueous phase and a tarry phase. The smoke condensate may be separated into fractions by physical separation techniques or by solvent extraction. These fractions may be further purified to remove hazardous substances known to be present in smoke. Smoke flavorings include smoke condensates, fractions thereof, and mixtures of such fractions.

The temperature at the center of a heap of smoldering sawdust may be 700–1000 °C, but the temperature gradient is steep, with the temperature falling to 300 °C or less a short distance from the center. Above approximately 200 °C, the wood undergoes destructive distillation, and the desired decomposition products are best generated in the 200–400 °C range. This smoke is allowed to diffuse over, or more commonly is blown over the food to be smoked, with varying levels of control depending on the technology available.

The thermal decomposition of wood can be influenced by numerous factors such as temperature, wood composition, amount of oxygen present, and amount of water vapor available during pyrolysis. Temperature is the most important and the various constituents of wood react at different temperatures. The first major component to undergo thermal decomposition is hemicellulose, which decomposes between 200 and 260 °C to yield furan and its derivatives, as well as a series of aliphatic carboxylic acids. Cellulose is the next major wood component to undergo thermal decomposition, pyrolysis occurring between 260 and 310 °C, which in turn leads to the formation of carbonyls, acetic acid, and its derivatives, together with water and occasionally small amounts of furans and phenols.

The lignin fraction is the most resistant, thermal decomposition occurring at 310–500 °C, and the derived compounds include phenols and phenolic esters along with their homologs and derivatives. Thus, if a relatively low temperature is used, lignin may not be completely degraded, and the resulting smoke will have a different chemical composition from smoke produced at a higher temperature. (*See* **Cellulose**; **Hemicelluloses**; **Lignin**.)

When first formed, smoke is generated as a vapor, but as it cools, some less volatile components condense to form a disperse liquid phase, which, along with soot particles, if present, constitute visible smoke. The remaining more volatile components distribute between the gaseous and disperse phases according to their solubility and volatility at the

prevailing temperature and humidity. Studies of wood-smoke composition have shown the presence of over 400 volatile components comprising 48 acids, 22 alcohols, 131 carbonyls, 22 esters, 46 furans, 16 lactones, 75 phenols, and some 50 miscellaneous compounds. The precise composition of woodsmoke and the extent to which the food absorbs smoke constituents depend on several factors.

Factors Affecting the Composition of Woodsmoke

Type of Wood

Wood generally contains some 40–60% celluloses (β-glucans), some 20–30% hemicelluloses (heteroglycans containing pentose and hexose residues), some 20–30% lignins (complex three-dimensional phenolic polymers), and a small amount of protein. In addition, there is a quantitatively minor, but chemically complex, low-molecular-mass fraction in which phenols and terpenes often dominate, particularly in softwoods, and variable water content.

Commonly woods are classified as hard or soft, depending on their botanical origin, the former from broad-leaved angiosperms, and the latter from needle-leaved (often evergreen) gymnosperms, but the terms 'hard' and 'soft' in this context are not good guides to physical properties. Compared with softwoods, hardwoods have larger contents of hemicelluloses, those in softwoods having a greater proportion of pentosans. Softwoods have greater lignins contents, with more guaiacyl (2-methoxyphenyl) residues relative to syringyl (2, 6-di-methoxyphenyl) residues than is usual in hardwood lignins.

There is a general belief that smoking with hardwoods such as oak (*Quercus* spp.), hickory (*Carya* spp.), beech (*Fagus* spp.) and alder (*Alnus* spp.) produces superior smoked foods compared with smoking with softwoods. Hardwoods generate more syringyl derivatives relative to guaiacyl derivatives and more acid smokes (see Table 1), due to structural differences in the lignins and the greater pentosan content of hardwood hemicelluloses, respectively. These acids produce smoked foods with lower pH values and hence a greater microbiological stability, and this may explain the traditional preference.

Wood Combustion Temperature

The temperature of smoke generation influences its composition. The total carbonyls concentration increases in the temperature range 200–600 °C, total phenols concentration in the range 400–600 °C and polycyclic aromatic hydrocarbons (PAH), while virtually absent below 400 °C, increase rapidly thereafter with increased generation of methylene radicals. The PAH have implications for product safety, and modern methods of smoking are designed to keep the generation temperature below 400 °C or otherwise remove the PAH. The balance between individual compounds within these broad groups is also affected by combustion temperature, with guaiacol, acetovanillone, and acetosyringone dominating the phenols fraction at 380 °C, but with acetosyringone and propiosyringone dominating at 560 °C. Some further data are shown in Table 2.

Smoke House Humidity

It is thought that smoke components enter the food primarily by absorption into the interstitial water in the food, at least for low fat foods. Initially, most foods being smoked will contain about 80% moisture and have a high water activity, both of which decline progressively during the smoking process. The absorption of smoke components is thus more rapid at the start of smoking and declines as the water content declines.

The rate of water loss from the food is greatly influenced by the absolute humidity of the air in the kiln relative to the surface water activity in the food. More rapid loss of water due to the use of hotter and/or dryer smoke gives rapid surface drying and a firm 'skin' to the food, which may inhibit absorption of smoke components. The use of wetter smoke slows dehydration, but if insufficient drying occurs, the product may be microbiologically unstable due to its high water activity. Often, wet sawdust is used so as to keep the smoldering temperature low and prevent combustion, but excessive damping must be avoided,

Table 1 Effect of wood source on composition of smoke

Source	Aroma fraction (mg per 100 g of wood)			
	Total	Phenols	Carbonyls	Acids
Oak (*Quercus seriata*)	1800	151	117	1140
Oak (*Quercus acuta*)	1600	225	323	820
Cherry	1490	101	111	660
Pine	1180	166	53	640

Table 2 Effect of combustion temperature on smoke composition

Combustion temperature (°C)	Content in smoke (mg per 100 g of sawdust)		
	Phenol	Guaiacol	Syringol
450	20	60	85
550	30	80	150
650	55	125	260

otherwise the very wet atmosphere created will inhibit drying. Smoke composition is in any case influenced by the moisture content of the wood that is used – see Table 3. (*See* **Water Activity: Principles and Measurement.**)

Some woodsmoke phenols have antioxidative and antimicrobial properties and, together with other chemicals such as formaldehyde, can increase the stability of smoked foods relative to unsmoked foods. However, it is generally accepted that the main preservative effect of traditional smoking is through drying, especially in combination with salting. Modern smoked foods have comparatively low salt and high water contents (thus increasing the yield and palatability of the product), but due to the relatively high water activity that results, the product shelf-life must be extended by chilling, freezing, or canning.

Effect of Air Flow Rate

Air flow is essential to keep wood smoldering, to transfer smoke from the generator to the kiln, and to minimize nonuniform smoke distribution within the kiln. As discussed above, air flow rate and humidity influence the rate at which the food surface dehydrates and thus the rate of smoke absorption. It has been demonstrated that the dispersed particles of the smoke act as a reservoir for the volatile vapor constituents. Diluting the smoke with air at constant temperature, or raising the smoke temperature, drives volatiles from the dispersed droplets into the vapor phase and encourages rapid and extensive transfer of these volatiles to the food, whereas dilution of smoke with cold air has the opposite effect.

Time of Smoking

The moisture content in the surface layers of the food decreases as smoking time increases, and this affects the rate and extent of smoke absorption. The distribution of fat on the surface of the food may also change during the process, particularly where high temperatures will cause greater protein denaturation and shrinkage, and movement of fat from the interior to the surface of the food, which will modify smoke absorption. Since smoke absorption is a first-order process, the rate declines as surface saturation is approached, since inward diffusion is slow and limiting. Even after extensive smoking, the smoke concentration falls markedly over the first 5 or 10 mm below the surface.

It should be noted that factors such as kiln temperature, kiln humidity, air flow rate, and smoke density are difficult to control in traditional kilns where these parameters are strongly influenced by weather conditions. In modern mechanical kilns, such factors can be controlled within much closer limits.

Table 3 Influence of wood moisture content on smoke composition

Moisture content (% as is)	*Yield per 100 g of sawdust (as is)*			
	Condensate (g)	*Phenols (mg)*	*Acids (mg)*	*Formaldehyde (mg)*
1.8	28.2	236	3203	122
21.5	41.5	136	3288	81
24.5	43.5	100	3003	78
31.2	32.8	33	890	

Smoke-generation Technology

Conventional Smoke Generators

In this type of generator, the smoke is produced by the slow combustion of sawdust using heating coils or gas flames. The sawdust is gently agitated continuously, in the presence of sufficient oxygen. The resulting smoke is then either drawn or pushed into the actual smoking chamber and allowed to interact with the food. The amount of tar and solid particles that enter the chamber can be controlled by varying the distance of the smoldering unit from the smoking unit; with a longer distance, there is more opportunity for tar to settle out. Various baffles and filters can be placed in line to remove tar. Smoke production or generation temperature using the smoldering technique can vary over a fairly wide range, dependent upon the available air. If a large quantity of air is present, the actual smoldering temperature may be in excess of 800 °C, but if the amount of available air is controlled, lower smoldering temperatures will result. The smoke-production temperature can be lowered by controlling the moisture content of the sawdust. Low smoldering temperatures also produce smoke generally lower in polycyclic aromatic hydrocarbons, which has significant health implications.

Friction Smoke Generators

Some processors use the friction method, whereby an intact wood block is pressed against a rapidly rotating metal friction wheel. Heat produced by friction is sufficient to cause wood pyrolysis. Air enters through the center of the friction wheel, thus producing a cooling effect on the resulting smoke. The more air that is present, the lower the resulting smoke temperature. This is in direct contrast to the smoldering technique. The temperature can be regulated by adjusting the pressure on the block of wood, the

wheel speed (which also governs the amount of air used), and the shape of the friction edges. The actual temperature during wood pyrolysis using friction is in the region of 450–560 °C; since rapid cooling occurs, the actual smoke temperature is relatively low.

Moist Smoke Generator

In the wet smoke or condensate smoke method, sawdust and superheated steam containing varying amounts of heated air are blended, causing sawdust pyrolysis. The resulting smoke has a temperature of approximately 80 °C and is moist because of condensed water from steam. This represents another flameless method, and pyrolysis normally occurs in the 300–400 °C range. Advantages cited for the process include a short treatment time and improved smoke yield.

Fluidized-bed Smoke Generator

With this technique, air is heated to between 300 and 400 °C by means of an electric heater, and under high velocity, mixed with sawdust. Because of the resulting turbulence, the sawdust is 'suspended' or fluidized; pyrolysis occurs at approximately 350 °C in around 10 s. A cyclonic separator is used to yield smoke and charred sawdust.

Two-stage Production of Smoke

The two-stage procedure is a modification and extension of the fluidized technique. The first stage involves initial wood pyrolysis by the action of nitrogen or carbon dioxide at a temperature of 300–400 °C on sawdust. In the second stage, the reactants of the first stage are mixed with oxygen or air heated to 200 °C, and further pyrolysis occurs. The second stage promotes oxidation, condensation and polymerization reactions, resulting in a more complex mixture of flavor compounds in the smoke.

Production of Smoke by Carbonization

In this method, sawdust is compressed in a cylindrical pipe by a screw, so that most of the air is squeezed out. An electrical resistance heater situated at the end of the cylinder heats the sawdust to 3000–4000 °C, and because of a lack of air in the compressed sawdust, the product gives off smoke during carbonization.

Smoking Techniques

Smoking of food products may be performed in various ways; the two main methods are cold or hot smoking, but techniques such as electrostatic smoking and treatment with liquid smoke condensate are also in use. The physical processes taking place during the action of the smoke on the product are adsorption, adhesion, condensation, diffusion, and absorption, and the smoking technology uses one or more of these processes.

Cold Smoking

In cold smoking, the temperature is maintained between 15 and 25 °C. The temperature of the smoke is regulated either by regulation of air or by passing smoke through a heat exchanger. The duration of treatment can range from a few hours to several days. It is suitable for cooked, raw, fermented, or dried products. Examples of foods that are typically cold-smoked are raw ham, bacon, salami and dried sausage.

For dried products, the rate of circulation of smoke must be controlled since too high a flow rate will cause crust formation on the surface, which will retard the subsequent drying process.

Hot Smoking

Depending on the temperature at which this is performed, the product undergoes either partial or complete cooking. Hot smoking is carried out at temperatures in the range of 55–80 °C. The activity of smoke is considerably enhanced under these conditions, allowing the treatment to be greatly shortened, relative to cold smoking. Some processors practice what is called moist smoking, which involves using a smoke temperature of 24–48 °C and relative humidities in excess of 30%. Typical hot-smoked foods include cooked ham, frankfurters, salami, and cocktail sausages.

The choice of operating temperature is largely determined by the nature of the product and the desired organoleptic attributes, as constrained by microbial safety. Guidelines on safe operation have been published by MAFF, USFDA, and FAO, among others.

Electrostatic Smoking

In this type of smoking, a continuous tunnel type arrangement is used, where the product is positioned between live electrical wires that are charged between 20 and 60 kV. Smoke passing through this system is charged according to its phase (smoke is a two-phase system, particulate and vapor), and smoke components can then precipitate on the oppositely charged food surface. This stage is usually followed by infrared irradiation, which dries the product surface.

Treatment with Smoke Essences or Condensates

This technique relies on obtaining a liquid essence of the smoke produced in a conventional smoke generator. Generally, condensed smoke can be obtained

dissolved in aqueous solutions (e.g., water or vinegar), as an organic solution (e.g., in oil or alcohol) or as a dry powder (e.g., on a salt, spice, starch, or gum arabic carrier).

These liquid smokes can be introduced into the product by various means, for example by addition to the recipe or by dipping, injection, or atomization. For comminuted products such as sausages, frankfurters, and salami, smoke with a dry carrier can be added to the mix during the chopping sequence.

For products manufactured using liquid processes, smoke essence can be incorporated by dipping. This is generally for a period of 5–60 s, and, while dipping imparts a smoked appearance, the smoked taste is weak. For injected products such as bacon and ham, smoke essence can be included in the injection brine at a concentration of 0.25–1%.

There are several advantages in using liquid smoke. Flavor can be incorporated throughout a product in a uniform manner instead of having a higher concentration on the outside than on the inside, as occurs in traditional smoking. There is also a closer control on the amount of smoke flavor a product receives. The liquid smoke condensate can be fractionated to intensify the smoke flavor and to remove potentially harmful compounds before it is used in food. It is also cheaper, as wood and smoking equipment are not required as part of a food-smoking plant.

Effects of Smoking on Food

Changes in Nutritive Value

Smoking, particularly hot smoking and especially as practiced in the tropics, can lead to destruction of tryptophan, cysteine, lysine, and other basic amino acids whether free or protein-bound. The losses recorded for lysine are variable, reflecting the variations in the smoking process (and analytical procedure), but may reach 55%. Such losses are concentrated in the outer 5–10 mm where conditions are harshest. It is clear that these losses are not just a function of temperature, but are due in part also to interactions with woodsmoke constituents and ingredients such as nitrate and nitrite, which may be added during the curing stage. Among the products detected are imines from carbonyl–amino interactions, some of which contribute to the characteristic yellow or reddish brown surface coloration of traditionally smoked foods. In the case of certain amino acids, these imines react further, yielding various β-carboline carboxylic acids (from tryptophan and woodsmoke carbonyls), and at least one nitroso-thiazolidine carboxylic acid (from cysteine, formaldehyde, and nitrite). (*See* **Amino acids**: Properties and Occurrence.)

Sensory Properties of Smoked Food

Texture

European cold-smoked products will generally have a texture that is soft and tender, whereas the equivalent hot-smoked products will have a firm dry surface with softer interior, being more succulent, the higher the fat content. The main determinants of texture are:

1. the extent and rate of water loss, with a greater loss giving a firmer texture, and a more rapid loss giving a greater difference in texture between the surface and interior;
2. the fat content and its distribution, largely a function of the food, with a food with a higher fat content being more succulent;
3. the extent of denaturation of structural and connective tissue protein, denaturation being more extensive for higher temperatures and higher salt contents, with fish in general being more susceptible than meat, and with greater denaturation giving a firmer texture;
4. the extent of autolysis, particularly proteolysis, with ungutted fish being most susceptible, and greater proteolysis giving softer texture. (*See* **Sensory Evaluation**: Texture.)

Color

As referred to above, carbonyl–amino interactions are important, with model system studies suggesting that interaction of lysine with glycolic aldehyde, 2,3-butane-dione, pyruvaldehyde, coniferaldehyde, or sinapaldehyde are particularly important. Colour may derive from colored substances in the smoke being deposited directly, or from permitted colorings (such as the azo food color Brown FK (for kippers) or natural carotenoids) incorporated during processing. (*See* **Colorants (Colourants)**: Properties and Determinants of Synthetic Pigments.)

Flavor

The presence of an odoriferous or sapid compound in woodsmoke or smoked food does not necessarily equate with organoleptic importance, since the quantity present must be sufficient to exceed the odor or flavor threshold of the consumer (see Table 4). (*See* **Sensory Evaluation**: Taste.)

Many investigators consider the phenolic constituents of woodsmoke, especially 4-methyl-guaiacol, guaiacol, and eugenol to be particularly important, at least in providing those smoky flavor notes that are

Table 4 Threshold values, sensory index values and described odor for some woodsmoke phenols

Compound	In water				In oil		Odor description
	Taste threshold	Taste index	Odor threshold	Odor index	Odor threshold	Odor index	
Guaiacol	0.013	6 400	0.021	4 600	0.07	1 000	Sweet, smoky, and somewhat pungent
4-Methyl-guaiacol	0.065	90 000	0.09	58 800	0.4	18 000	Sweet, smoky
2,6-Dimethoxyphenol	1.65	1 400	1.85	1 200	0.34	7 000	Smoky

essentially unique to smoked foods. There is no doubt, however, that other classes of smoke volatiles are also very important, and attention has been directed towards the pyrazines and lactones.

Aqueous extracts of smoke condensates may be prepared as flavoring ingredients and to avoid the potentially hazardous PAH fraction. Analysis of commercially available extracts indicates that there can be considerable variations in composition, depending on the method of manufacture. Whether friction smoking or electrostatic smoking duplicates the flavor achieved by the traditional process is also a matter of some debate. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Assessment of Toxic Hazard

Concern about the possible safety of smoked foods centers primarily on the presence of PAH. Thirteen PAHs have been detected in smoked foods, at least one of which, benzo[a]pyrene, is a known carcinogen. The greatest levels appear to be associated with products such as charcoal broiled steaks, which are only incidentally smoked, rather than those products normally described of as smoked. It has been estimated that the average daily per-capita intake of benzo[a]-pyrene in the UK is less than 4 μg, and that smoked foods do not contribute significantly to this burden in the UK. (*See* **Carcinogens**: Carcinogenicity Tests; **Polycyclic Aromatic Hydrocarbons**.)

The safety of smoke condensates was reviewed by the Joint FAO/WHO Expert Committee on Food Additives (JECFA) in 1987. Their concern centered mainly on the hazard that would arise if nitrosamines and carcinogenic PAH were present. The evidence available to JECFA indicated that neither class of substance was detectable in the preparations tested. More recently, a limit of 1 μg kg^{-1} has been suggested as an upper limit for benzo(a)pyrene, although studies have shown that some foods, particularly ethnic foods and the outer layers of smoked fish, may contain levels above this.

Concern has also been expressed regarding the presence of heterocyclic aromatic amines in foods, particularly in grilled meat and fish, and presumably also in the corresponding smoked foods. These compounds are pyrolysis products derived from amino acids, such as tryptophan, glutamic acids, or creatinine and products of the Maillard reaction. Some are potent mutagens in certain *in vitro* mutagenicity tests.

Recently, varying amounts of three β-carboline-3-carboxylic acids have been detected in some smoked foods produced in Europe. Although some β-carbolines are known to be pharmacologically active, for example inhibiting monoamine oxidase, there has been no evaluation of the significance of any of those detected in smoked foods.

See also: **Amino Acids**: Properties and Occurrence; **Carcinogens**: Carcinogenicity Tests; **Cellulose**; **Colorants (Colourants)**: Properties and Determinants of Synthetic Pigments; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Hemicelluloses**; **Lignin**; **Polycyclic Aromatic Hydrocarbons**; **Sensory Evaluation**: Texture; Taste; **Water Activity**: Principles and Measurement; Effect on Food Stability

Further Reading

Gangoli SD (1986) The toxicology of smoked foods. *Proceedings of the Institute of Food Science and Technology* 19: 69–78.

Garcia Falcon MS, Gonzalez Amigo S, Lage Yusty MA and Simal Lozano J (1999) Determination of benzo(a)pyrene in some Spanish commercial smoked products by HPLC-FL. *Food Additives and Contaminants* 16: 9–14.

Maga JA (1987) The flavour chemistry of woodsmoke. *Food Reviews International* 3: 139–183.

Papavergou K and Clifford MN (1986) Chemical interactions between woodsmoke components and foods. *Proceedings of the Institute of Food Science and Technology* 19: 13–17.

Toth L (1982) *Chemie der Räucherung*. Weinheim: Verlag Chemie.

WHO (1987a) Evaluation of Certain Food Additives and Contaminants. Thirty-first Report of the Joint FAO/WHO Expert Committee on Food Additives. Technical Report 759, pp. 21–23. Geneva: WHO.

WHO (1987b) *Toxicological Evaluation of Certain Food Additives. WHO Food Additives Series*, vol. 21: pp. 41–57. Cambridge: Cambridge University Press.

Production

M Vaz-Velho, Escola Superior de Tecnologia e Gestão, Viana do Castelo, Portugal

Background

Smoking is an ancient process for preserving and flavoring food, and is a traditional part of the diet of a large section of the world population. Smoking processes are mainly applied to flesh food (meat and fish), but other foodstuffs such as cheese can also be smoked.

The original principle of preservation was due to a combination of lowered water activity of the product and the uptake by the product of bactericidal and antioxidant components of smoke. The secondary objective is to impart sensory characteristics such as smoke color and smoke flavor to the product. With the exception of some African and Asian countries, where a refrigeration chain is not widely established, nowadays, smoking is used mainly to give the product a characteristic flavor, the preservative effect being slight, and many smoked fish products keep in good condition for very little longer than the fresh product from which they were made.

Smoking fish is a very traditional and ancient process in Northern European countries. However, smoked meat products, mainly from pork meat, have a huge impact on the economy of Mediterranean countries, where a great variety of smoked meat products are industrially or traditionally produced.

In industrialized countries, despite globalization and standardization of smoking processes and products, the elements that differentiate and identify a product have an increasing importance for producers and consumers to whom it is important to link products to the region where they originated and to their specific means of production. Within each country, local resources represented by traditional smoked products, although produced on a small scale, now have a great economic impact owing to the creation of organizations and schemes to exploit diversities and complementarities and consequently allowing the achievement of considerable profits.

Smoke Production

A smoked product can be defined as a product that was prepared by subjecting the raw material to the direct action of smoke from pyrolysis of wood, wood shavings or chips, and sawdust.

In most parts of the world, wood is still preferred to sawdust. Wood leads to a hotter fire, and the products might be charred, rather than smoked. However, commercial smoked products in industrialized countries are usually smoked by a sawdust fire where, unless there is a forced rapid air flow, the sawdust smolders rather than burns. This lower smoke temperature gives more flavor and preserving substances to the product. Higher temperatures and more oxygen oxidize these substances.

Smoke is composed of two phases:

1. A particulate or dispersed phase (droplet phase) – the visible portion of smoke composed of tars, wood resins, high-boiling phenolic compounds, and lower boiling compounds that vary with temperature and smoke concentration.
2. A gaseous or dispersing phase (vapor phase) – most of the characteristic flavor associated with smoke is present in this gas phase:

Conventional smoke generation methods can be:

1. Simple – smoke is generated from pyrolysis of sawdust and woodchips in a simple fire box. These wood chips burn the sawdust, and smoldering starts.
2. Continuous – Modern smoke producers use a continuous smoldering by feeding the sawdust slowly on to a very hot surface.

In both methods, smoke generation temperatures are dependent on the rate of burning of the wood to produce heat.

Reducing the temperature to which wood is heated to minimize the production of undesirable smoke components such as polycyclic aromatic hydrocarbons (PAH) has been an important issue. (*See* **Polycyclic Aromatic Hydrocarbons**.)

Two methods of smoke generation, in which the temperature is controlled externally, have been developed for reducing PAH levels:

1. Fluidized bed generator – sawdust is fluidized by hot air at 350 °C, and smoke with a high proportion of acids is formed;
2. Steam smoke generator – high-pressure steam at 350 °C is passed through a bed of sawdust and smoke with a high proportion of water is formed.

Electrostatic precipitation has also been shown to reduce the concentration of PAH. It can be performed in a continuous tunnel-type kiln, in which the product is placed between electrical wires with a charge of 20–60 kV. The smoke moving through this system is charged according to its phase, and smoke components can precipitate on the oppositely charged product surface. Electrostatically filtered smoke was found to be lower in acids, phenol, and carbonyls than untreated smoke, giving a product with less smoky aroma and color development. However,

reducing the electrical wires spacing can lead to similar phenols levels as untreated smoke.

When using traditional smoke generators, by merely extending the distance between the smoke generator and the smoking chamber, most of the particulate phase of the smoke is allowed to settle before it makes contact with the product, thereby removing much of the PAH.

Smoking Kilns

As smoking is an ancillary process for increasing the storage life of food products, an enormous variety of smoking kilns and smoking processes are used for that purpose all over the world.

If smoking in industrialized countries is mainly a pleasurable alternative to the consumption of freshly cooked food, preservation, nevertheless, is still the prime objective of smoking in many parts of the world. In these countries, the smoking process usually takes place in simple boxes, or products are just placed over a fire place.

In European Mediterranean countries, traditional local smoked products can be produced in chimney structures many centuries old, usually coupled to a brick oven, and may still used for different purposes – every day cooking, as a heater in cold days, and for smoking meat in winter time (**Figure 1a**). These traditional kilns were built inside the house, the operation is dependent on the weather conditions, and production is seasonal, taking place in winter time. Hams, sausages, and sides of bacon were suspended and exposed to smoke from fires of hard wood (usually *Quercus* sp.). This produced a hotter fire because wood was used in preference to sawdust. However, as the flames were formed under an iron pot, they were not in direct contact with the products. During daytime, a mixture of smoke, water vapor and air reached the products. During the night, as no fire was needed for cooking, the smoking process continued by means of a wood brazier kept alight by the high brick temperature, but the process was flameless. Of course, as nowadays most of the chimneys are no longer used for cooking, the smoking process is kept flameless as it was during night.

Northern European countries still have chimney-like kilns hundreds of years old, but most of them were built even in that time with one single purpose – to smoke food, mainly fish. These traditional kilns, built outside the house, can vary in size from 1–2 m (hot smoking) to over 20 m in height (cold-smoking) (**Figure 1b**). Production takes place throughout the year, but in summer usually during the night when the ambient air is cooler. Traditional smoking involves the generation of smoke from hard woods (oak, beech, and hickory in Europe). Whole fish, fish fillets, sides of bacon, ham, and other meat products are suspended in the brick kilns and exposed to smoke from smoldering fires of wood chips or sawdust.

In these primitive techniques, smoke generation and application steps are accomplished in the same chamber. Hot air is less dense than cold air and moves upwards, carrying the smoke with it (**Figure 1a and b**). As temperatures are higher and smoke more dense lower down the kiln, the products have to be moved to different parts of the kiln so that all items receive the same smoke treatment.

These traditional kilns are notoriously difficult to control and have been replaced by a modern type of smoker in which the smoke is generated externally and blown into a metal chamber containing the products to be smoked (**Figure 1c–e**). Also, in most of these modern kilns, the air movement is across, rather than up, and smoke is drawn across the products by means of an electric fan to ensure that all products are dried/smoked to the same extent (**Figure 1d and e**). In modern smoking kilns primarily designed for hot-smoking (**Figure 1d**), the smoke is introduced by a series of ducts into the smoking chamber, but smoke distribution is less uniform than in the tunnel-type kiln (**Figure 1e**).

In recent commercial developments (**Figure 1e**), the direction of the smoke flow can be reversed automatically at frequent intervals, which obviates the need to move the product. However, there is still little chance of an even smoke or temperature treatment in all parts of the kiln because, as the smoke/air mixture passes from one rack to another, it becomes less smoky, more humid, and cooler. But, as racks are placed on a trolley, they can be easily moved into another part of the smoker.

These horizontal kilns – mechanical kilns – have different sizes for large- and small-scale productions. They can be used for cold- and hot-smoking by the use of electric heaters in the smoking chamber. In large kilns, the heaters can be placed at different points in the smoke path to keep the temperature more even throughout the smoking chamber. Humidity can also be controlled, and the consistency of smoking is high. Owing to the concern over PAH in smoke, some generators are equipped with a purification stage that involves the removal of undesirable compounds either by water sprays or by precipitation.

The Smoking Process

The preservative effect of smoking combines the effect of salting, drying, heating, and smoking. Although the general operations in all smoked processing plants are similar, the specific processing procedures vary considerably, owing to differences in equipment, national

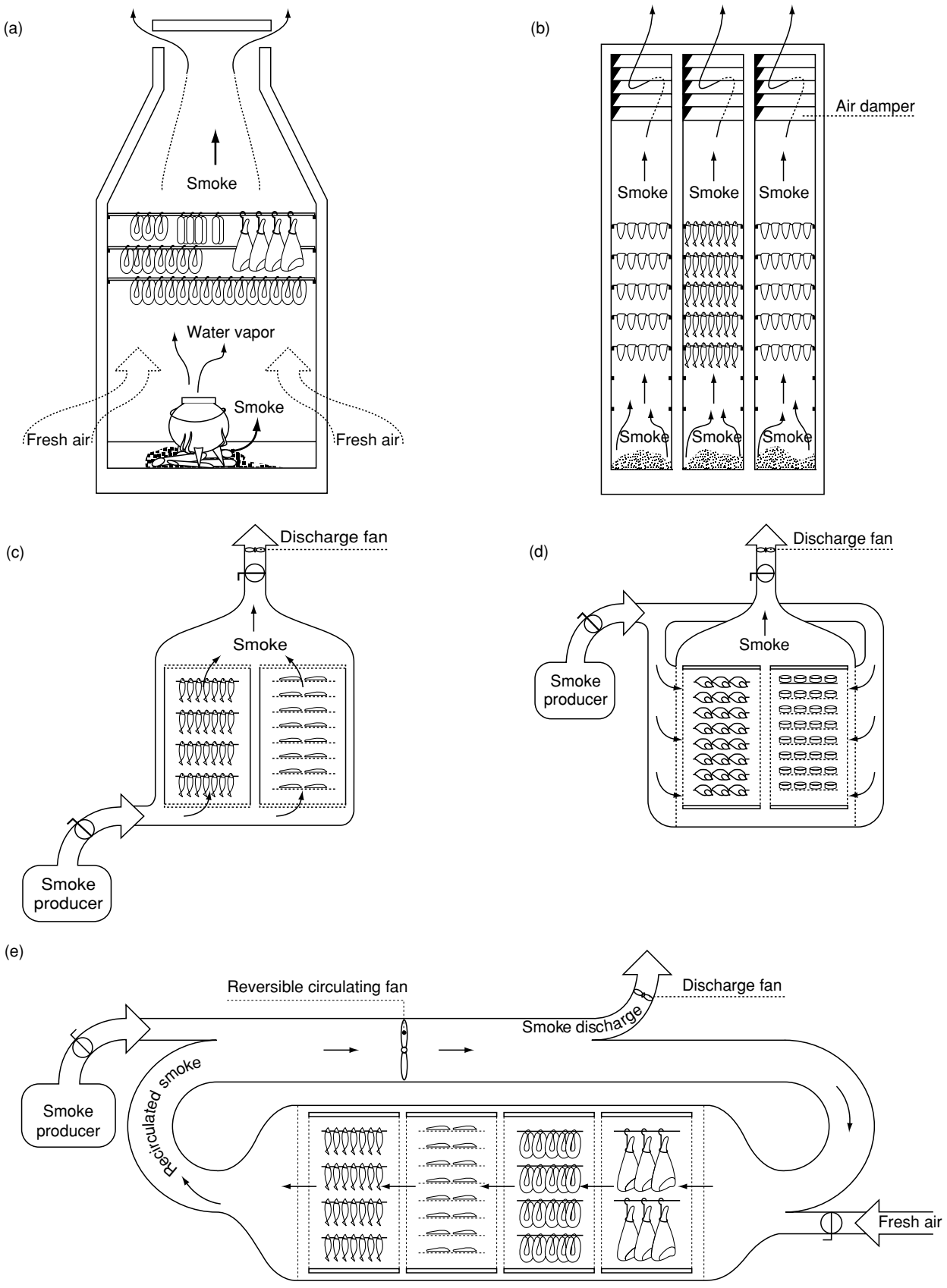

Figure 1 (a) Traditional Mediterranean smoker; (b) traditional north European smoker; (c) mechanical smoker (vertical smoke flow); (d) mechanical smoker (horizontal smoke flow); (e) mechanical smoker (horizontal smoke flow – tunnel type).

and regional consumer preferences, raw materials, and tradition.

Salting

Salting can be done by soaking or injecting the product with a prepared salt solution or by embedding or rubbing the products with salt. Excess of salt on the surface of the product, in these last methods, should be removed by showering with water. In both methods, salting nowadays, rather than for preserving, is used mainly to firm the flesh and impart flavor to the product.

In terms of safety, for cold-smoked fish, the aim is to obtain a concentration of 3–3.5% of salt in the water phase throughout the product to minimize the growth of *Clostridium botulinum*. The salt content in the water phase means the percentage of salt in the water contained in the fish. Salt in the water phase links the two conditions of moisture and saltiness. This should be distinguished from the salt content, which is the percentage of salt in the whole weight of fish, that is, water plus protein plus fat, etc.

The uptake of salt by products from brine or dry salt is a diffusion-controlled process dependent on the temperature, water content, salt concentration, and the amount of fat. The time required for a given product to absorb a given amount of salt is approximately inversely proportional to the square of the half thickness of the product. (*See* **Curing**.) The salt used should be of food-grade quality, low in calcium and magnesium, and essentially free of iron and copper.

In dry-salted products when the surface of the product dries out too quickly, owing to the small grain size of the salt, it can inhibit the diffusion of the water from the bulk of the product to the surface, which can lead to the decomposition of the center of thick muscles before it reaches the level of salt concentration sufficient to prevent spoilage.

The main objective of the brining method of salting is limitation of air penetration during processing, thus limiting the amount of rancidity that can develop. This is why it is the preferred method for fatty products. It also gives an attractive glossy appearance to the smoked fish fillets. However, as the water content of the product is increased, more drying is required after brining, than after dry salting.

A 70–80% saturated brine is commonly used. A 100% saturated brine can leave powdery salt crystals on the surface of finished product, whereas a 50% saturated brine would cause the product to swell, necessitating a longer drying period.

Making and maintaining brine quality involve the use of good-quality water and ingredients, the control of brining temperatures, and assuring high standards of equipment hygiene.

When products are brined or brine-injected, curing agents are added to the salt in the solution, e.g., for whole pieces of meat. Curing agents can be added to the ingredients together with salt, e.g., chopped or minced products. Nitrates, nitrites, ascorbic acid, sodium glutamate, polyphosphates, and sugar are the main additives used. Nitrates and nitrites are used in smoked meats to deliver the characteristic color by reacting with the heme pigments, contributing to flavor and, owing to its antimicrobial activity, inhibiting particularly *Clostridium botulinum* growth and toxin production. (*See* ***Clostridium***: Occurrence of *Clostridium botulinum*; **Curing**; **Nitrates and Nitrites**.)

Drying

In conventional smoking, drying can be performed before or after the smoke uptake or can be concomitant to smoke production. During smoking, the drying continues. It is important to balance the drying rates and quality/safety issues. The amount of moisture transfer from the product to the air is dependent on the relative humidity: the lower the humidity, the better the moisture transfer. However, if the rate of water removal is too high in comparison with the diffusion rate, the surface will form a hard case (case hardening) creating a barrier to water transfer and evaporation; consequently, the diffusion of the bulk water to the surface of the product is blocked. Thus, spoilage and possibly *Clostridium botulinum* growth might occur in the bulk of the product. To prevent this, the protein must be set or denaturated with low-temperature drying before applying higher temperatures. In conclusion, the product must be dried slowly enough to avoid case hardening, but fast enough to avoid spoilage caused by bacterial and enzymatic activities.

A large proportion of the drying of flesh food muscle usually occurs during constant-rate drying – while the surface of the product being dried remains moist, water is considered to evaporate as from a free water surface. The evaporation rate is thus controlled by the conditions of the drying air, namely speed over the surface, temperature, and humidity. As evaporation takes place at the product surface, water travels to the surface by diffusion and by capillaries formed by structural components. When the movement of this water through the product can no longer keep pace with the evaporation rate, the surface becomes dry, and this point in the drying sequence is referred as 'critical moisture content.' Thereafter, the rate of drying falls progressively and is dependent on the permeability of the already dry part of the product. During this stage, humidity and air velocity do not significantly affect the process, whereas temperature,

as it affects diffusion rate, should be adjusted during this period of falling. (*See* **Drying**: Theory of Air-drying; **Water Activity**: Principles and Measurement; Effect on Food Stability.)

Drying for cold- or hot-smoked fish usually takes place inside the smoker before smoke production starts – the diffusion distance between the center and surface of the fish or fish fillets is small, and water activity is reduced rapidly.

For whole pieces of meat (smoked hams), separate chambers are needed because the maturing/drying process can last for several days before products are exposed to smoke. The temperature and humidity of those chambers must be carefully controlled and are set according to the level of dehydration required for each particular product. It is important to evaluate the maximum temperature to which the drying air can be heated in order not to cause the product surface temperature to rise above the upper limit for maintaining specific product quality.

Modern equipment is designed to accelerate the drying process by increasing the evaporation rate or by using faster air speeds to remove the water vapor and increase the surface heat transfer rate.

Smoking

The smoking process takes anywhere from a few hours to a few days. A combined heating/drying/smoking process entails application of smoke, at temperatures below 30 °C for cold-smoked products or alternatively 60–80 °C for a hot-smoked product. In cold-smoking processes, the product is subjected to heat for a period of time that does not coagulate protein. Cold-smoking does not cook the flesh or coagulate protein, so refrigerated storage and further cooking are necessary before consumption. Cold-smoked fish like salmon, salmon-trout, swordfish, and tuna fish are not cooked before consumption. The maintenance of temperatures below 30 °C avoids protein coagulation, and the cold smoke is more dense, thus allowing the deposition of smoke on the fish surface giving the product the desirable appearance and flavor. Drying should be minimal at the start but will increase towards the end of the process. The length of the smoking process for such products is much greater than for hot-smoked fish, but a 'safety' (i.e., pasteurization) temperature is not achieved in any stage of the process. In a hot-smoking process, the product is subjected to heat for a sufficient period of time to coagulate protein throughout. The hot-smoking process adequately cooks the flesh, and so the products are ready to eat.

The uptake of wood smoke components from the particulate and vapor phase of wood smoke is dependent on the nature of the product, temperature, humidity, and velocity of smoke. The deposition of volatiles is maximum at the beginning of the smoking process whilst the surface remains wet. As the surface becomes dried, the uptake of volatiles decreases, and the proportion of smoke constituents arising from the particulate phase increases. The particulate phase acts as a reservoir of volatile and nonvolatile smoke components, releasing more volatiles as they are adsorbed from the vapor phase. Raising the smoke temperature or diluting the smoke with air at a high constant temperature drives the volatiles from the droplets into the vapor phase, and from here, rapidly to the food product, whereas dilution of smoke with cold air has the opposite effect.

Cold-smoking involves keeping the smoke temperature below 30 °C throughout the smoking process. A humidity of 70% RH seems to be a good compromise between maximum smoke adsorption and minimum case hardening in cold-smoking of fish. In a mechanical smoker, hot-smoking of fish usually undergoes a typical three-stage process: initially, cold-smoking for 1 h, then at 50 °C for 30 min, and finally at 80 °C for 2 h.

In a mechanical smoker, the temperature regime most often used for hot-smoked pieces of meat, smoked directly after curing, follows a two-stage process. The first, cold-smoking (low temperature/long time) involves smoking at 32–38 °C for 15–18 h; the second, hot-smoking (high temperature/short time), uses a temperature of at least 60 °C for 2–4 h).

When large pieces of meat are smoked after a long curing/maturation process, cold-smoking takes place for a few days.

Alternatively, in either cold- or hot-smoking methods, the product can be smoked by means of smoke-flavoring processes with liquid smoke, which imparts to a product the flavor of smoke. The classical liquid smoke is a natural aqueous condensate of wood smoke that has been aged and filtered to remove tars and particulate matter, and is dissolved in water, oil, or organic solvents.

For preparing a smoked-flavored product, after the salting step, the product can be smoked by adding smoke extracts to the mixture (smoked sausages), injecting (smoked ham) or dipping (smoked sausages, smoked ham) in a solution of liquid smoke or applying the liquid smoke by pulverizing it or using it as an atomized spray within a modern automatic smoke kiln. According to some recent research, spraying is preferable because it is easiest to control flavor and acceptability with this method.

In conventional smoking, it is expected that phenol components are at much lower concentrations towards the center of the product compared with liquid smoked products, where their levels should be very

uniform throughout the product. This is true if products are dipped or injected with the liquid smoke, but not if liquid smoke is only sprayed on to the product surface.

The composition of commercial liquid smoke can vary widely, depending on the method of manufacture. Natural smoke extracts, synthetic smoke flavors, and substances unconnected with smoke, such as yeast derivatives but with a smoky flavor or smell, are used.

The use of liquid smoke also reduces the undesirable PAH levels in smoked foods. The European Scientific Committee for Food, in its Directorate-General Industry Guidelines (1995) established that smoke flavorings (liquid smoke) added to a food product must assure a level lower than 0.03 μg of benzopyrene per kilogram and less than 0.06 μg of benzoanthracene, another PAH compound, per kilogram.

In the meat industry, liquid smoke is commonly used in place of traditional smoke. However, smoke-flavored fish represents a small percentage of the total smoked fish industry output.

Smoked Food Products Quality and Safety

Microbiological Safety

The intrinsic microbiological stability of smoked foods is dependent on the ingredients, additives, and processes. In one extreme are traditional products usually heavily cured and smoked, which can be stable for many months at ambient temperatures; at the other extreme, there are lightly cured and smoked products, whose stability is dependent on extrinsic factors such as refrigeration, vacuum, or gas packing.

The length of the smoking process for cold-smoked products is much longer than for hot-smoked products, but a pasteurization temperature is not achieved in any stage of the process. Thus, the temperatures and times used in processing cold-smoked products are very favorable for the proliferation of food spoilage and food poisoning types of microorganisms.

Most of the commercialized smoked products in Western countries are sold vacuum-packed and stored at refrigeration temperatures, and so spoilage and food poisoning flora of the finished products are mainly anaerobic or microaerophilic, and psychrotrophic or psychrophilic.

Cold-smoking, or any smoking process where temperatures in the interior of the foodstuff do not reach 50 °C for a significant period of time, will have a different residual postsmoking microflora than those hot-smoked products where a high internal temperature is reached.

The temperature achieved during hot-smoking kills vegetative microorganisms but not all spores, so the most probable spoilage agents will be spore formers or postsmoking contaminants. Refrigeration of hot-smoked fish greatly reduces the number of organisms able to grow and lengthens the shelf-life of the product.

Cold-smoked products will contain a microbial flora representative of that of the raw material. The salt in the water phase must be high enough to inhibit the growth of *Clostridium botulinum* ($\geq 3\%$), but this will not stop the growth of spoilage bacteria. Immediately after vacuum packing, bacterial counts in cold-smoked salmon, commonly range between 10^4 and $10^5\,g^{-1}$, and at the end of the shelf-life (6 weeks), the product typically contains a microflora dominated by lactic acid bacteria at levels of 10^7–$10^9\,g^{-1}$. The use of drying and/or salt to lower the water activity and smoke to add chemical preservatives limits the types of microorganisms capable of growth, and the production process affects the microbial flora present on the product; therefore, both of these affect the spoilage patterns. Spoilage needs to be defined carefully for each product.

Listeria monocytogenes has been found consistently in smoked fish products. Nitrites, usually added to smoked meat products, are known to inhibit growth of *C. botulinum* and also can inhibit *L. monocytogenes* growth. (*See* ***Listeria***: Properties and Occurrence.) Nitrites also can be added to cheese to prevent very late fermentation and gas production by *Clostridium tyrobutyricum* or *C. butyricum*.

For *C. botulinum*, levels of salt higher than 3% in the water phase together with refrigerated storage can prevent growth and toxin production in cold-smoked fish products, even when nitrites are not added. However, with respect to *L. monocytogenes*, which is able to grow at salt levels up to 10% and temperatures as low as 1 °C, when nitrites are not allowed in fish products (EU) there is no step in the cold-smoked process that is able to prevent its growth.

The International Commission for Microbiological Specifications for Foods recognized that numbers of *L. monocytogenes* not exceeding $100\,g^{-1}$ at the time of consumption are of low risk to the consumer. Therefore, for ready-to-eat, not heat-treated, foods, where the presence of *L. monocytogenes* cannot be fully avoided, the critical points are to limit the occurrence of *L. monocytogenes* in foods but, more importantly to control its growth and survival in order to keep levels below the $100\,g^{-1}$ at the point of consumption. This policy is followed by some European countries but is considered a hazard by USA, which maintains a zero-tolerance policy regarding *Listeria* incidence in ready-to-eat foods. If a zero-tolerance policy were to be established in EU countries, where nitrites or other antimicrobial additives are not allowed to be added to fish products, this

would lead to the cessation of sales of cold-smoked fish products where the presence of *L. monocytogenes* cannot be fully avoided. For cold-smoked fish products, using reliable raw material and good manufacturing practices (GMP), keeping storage temperatures under 4 °C, and reducing the shelf-life to 3 weeks, has been reported to be sufficient to maintain *L. monocytogenes* numbers below $100\,g^{-1}$.

Nisin, the only approved bacteriocin allowed to be added to foodstuffs, and commonly used in milk products to prevent *C. tyrobutyricum*, is also capable of preventing *C. botulinum* and *L. monocytogenes* growth in smoked products. Recent studies on the application of different bacteriocins from *Carnobacterium* spp. to cold-smoked fish products to inhibit *L. monocytogenes* growth have yielded encouraging results on inhibiting the growth of this pathogen.

Cold- and hot-smoked foods, like every food commodity, can potentially cause safety problems. To improve the safety of commercially smoked products, producers should use reliable raw materials, improve smoking technologies (temperature and humidity control), assure plant equipment and personnel hygiene, and implement GMP and hazard analysis and critical control points. (*See* **Hazard Analysis Critical Control Point**.)

Chemical Safety

Preventing the occurrence of carcinogens during smoking has been an issue for several years. There are two groups of chemicals of concern in smoke: PAH and *N*-nitrosamines (NNAs), both of which are considered potential carcinogens. The compound benzopyrene, the most concerning PAH, is regarded as an indicator of carcinogenicity.

As mentioned before, there are several means for reducing PAH formation. The use of a separate smoke generator, maintaining the temperatures of pyrolysis between 200 and 425 °C, electrostatic filtration of the smoke, smoke generated by superheated steam or the use of liquid smoke are some of the ways of reducing the PAH levels in smoked foods. Reports on the quantification of PAH in smoked fish, meat, and cheese have suggested that as, in this last product, the rinds are removed before consumption, the PAH intake would be much lower. However, in some traditional cheeses, where manufacture includes heating the milk on an open wood fire, or in commercial hot-smoked cheeses, the levels of PAH are considerably higher compared with liquid smoked samples.

The NNAs in smoked food are primarily formed by the reaction of nitrogen oxides (generated from nitrites) of the wood smoke with, mainly, secondary amines present in the flesh. (*See* **Nitrosamines**.)

Nitrates and nitrites used in smoked meats to impart color and flavor and due to its antimicrobial ability are an additional risk factor, as they may react with both secondary and tertiary amines of the products leading to the formation of NNA. The reaction can occur during processing, reactive substrates are amino acids, and many of the resulting NNAs have been identified as carcinogens. Reducing agents, like ascorbic acid, added together with nitrites to the brine to improve NO_2 activity, also can act as efficient factors inhibiting NNA formation. Considerable research attention has been paid to determining the levels of occurrence and formation of these substances in foods. It is currently considered that nitrite levels in smoked meat are such to ensure the stability of the product and to control *C. botulinum*, while not presenting a significant risk from NNA.

Although nitrites are legally added as preservatives to smoked products elsewhere, in EU countries, they are restricted to smoked meats and prohibited in the fish industry. Therefore, it is expected that smoked fish products from EU countries are poor sources of nitrites, and consequently of nitrosamines, compared with smoked fish produced elsewhere and compared with smoked meat products.

See also: ***Clostridium***: Occurrence of *Clostridium botulinum*; **Curing**; **Drying**: Theory of Air-drying; **Hazard Analysis Critical Control Point**; ***Listeria***: Properties and Occurrence; **Nitrates and Nitrites**; **Nitrosamines**; **Polycyclic Aromatic Hydrocarbons**; **Water Activity**: Principles and Measurement; Effect on Food Stability

Further Reading

Bosset JO, Buetikofer U, Dafflon O *et al.* (1998) Occurrence of polycyclic aromatic hydrocarbons in cheese with and without a smoked flavour. *Sciences des Aliments* 18: 347–359.

Culik J and Kellner V (1995) Toxic compounds arising by interaction of food constituents with food additives. In: Davidek J (ed.) *Natural Toxic Compounds of Foods*, pp. 229–249. London: CRC Press.

Dafflon O, Scheurer L, Koch H and Bosset JO (2000) Quantification of nitrated polycyclic aromatic hydrocarbons in fish, meat products and cheese using high performance liquid chromatography. *Mitteilungen aus Lebensmitteluntersuchung und Hygiene* 91: 158–171.

FAO Fisheries Report No. 604 (1999) *FAO Expert Consultation on the Trade Impact of* Listeria *in Fish Products*. 17–20 May 1999. Amherst, MA.

Final Report of EC FAIR Project CT95–1207 (2000) Spoilage and Safety of Cold-smoked Fish.

Foster WW and Simpson TH (1961) Studies of the smoking process for foods, I. The importance of vapours. *Journal of the Science of Food and Agriculture* 12: 363–374.

Horner WFA (1997) Preservation of fish by curing (drying, salting and smoking). In: Hall GM (ed.) *Fish Processing Technology*, pp. 32–73. London: Chapman & Hall.
Maga J (1988) Smoke generation technology. In: *Smoke in Food Processing*, pp. 39–48. Boca Raton, FL: CRC Press.
Storey RM (1986) The technology of traditional smoking. *Proceedings of the Institute of Food Science and Technology* 19: 58–62.
Varnam AH and Sutherland JP (1995) Cured meats. In: *Meat and Meat Products – Technology, Chemistry and Microbiology. Food Products Series*, vol. 3, pp. 167–223. London: Chapman & Hall.
Vaz-Velho M, Capell C and Gibbs P (1998) Cured, smoked and dried fish. In: *Fish and Seafood. Microbiology Handbook*, vol. 3, pp. 119–155. Leatherhead, UK: Leatherhead Food Research Association.

Applications of Smoking

J I Ahmad, University of Lincolnshire and Humberside, Grimsby, UK

Introduction

Smoke used for curing of foods consists of a suspension of minute particles in a vapor phase making up an aerosol. It is produced by controlled thermal combustion of sawdust, wood shavings, coffee husk, sugar cane pulp, coconut husk, and rice straw, etc. The composition of smoke varies with the choice of wood, the temperature of pyrolysis (**Table 1**), relative humidity and the moisture content of the wood (**Table 2**). The relative concentration of different components depends on the type of wood used and the controls adopted during its combustion (**Figure 1** and **Table 3**).

Table 1 Influence of the glowing temperature on the phenol, carbonyl, and acid yields in smoke (in milligrams per 100 g of sawdust)

Temperature °C	*Total phenols*	*Total carbonyls*	*Total acids*
380	998	9 996	2 506
600	4 858	14 952	6 370
760	2 632	7 574	2 996

Table 2 Mechanism of smoke formation

(a) *Pyrolysis of wood*	
Up to 170 °C	Thorough drying
200–260 °C	Pyrolysis of hemicellulose
260–310 °C	Pyrolysis of cellulose
310–500 °C	Pyrolysis of lignin
(b) *Secondary reactions of the pyrolytic products at about 200 °C*	Oxidation, polymerization, condensation, pyrolysis

Polycyclic aromatic hydrocarbons (PAH), which are believed to be carcinogenic, are a well-known group of constituents of wood smoke (**Figure 2**). These compounds are probably formed through condensation of naphthalene. Their concentration strongly increases when other organic materials are added during combustion of wood. (*See* **Polycyclic Aromatic Hydrocarbons**.)

Apart from providing desirable color and flavor to foods, wood smoke contributes to preservation by acting as an effective antioxidant and a bacteriostatic and bactericidal agent. Phenolic components of the smoke contribute to the flavor and aroma of the product. Smoke also provides a protective film on the surface of the smoked product, thus providing a barrier against development of rancidity. Combined chemical constituents of smoke together with heating and drying processes are responsible for bactericidal and bacteriostatic effects. (*See* **Antioxidants**: Natural Antioxidants; **Phenolic Compounds**.)

Methods of Smoking Foods

Basically, three methods are used in the production of smoke-flavored foods.

The Traditional Method

This is the direct incomplete thermal degradation of wood to produce smoke. It may be carried out either as *hot smoking* or as *cold smoking*, i.e., the criterion is the maximum temperature to which the product is exposed. (*See* **Drying**: Theory of Air-drying.)

Hot smoking Hot smoking involves exposing the product to at least 80 °C and probably up to 100 °C, the intention being to cook the product as well as smoke and dehydrate it. Duration of smoking depends on the level of curing required, e.g., light, mild, or strong and the thickness of the product. Techniques of smoke generation vary in different parts of the world. It may be a simple barrel-type smoke oven, inverted drums with a fire pit at the bottom or located remotely, or the very latest high-tech equipment as available in the UK with automatic smoke density, temperature, and humidity controls.

The principles of hot smoking are simple. Smoke is generated from desired wood (e.g., oak) with a controlled air supply. The product is usually prebrined for an appropriate period from a few minutes up to several hours before being placed on the racks where the smoke is allowed to circulate at a selected rate and temperature. Modern kilns yield a uniform, clean product with little waste. There are basically three

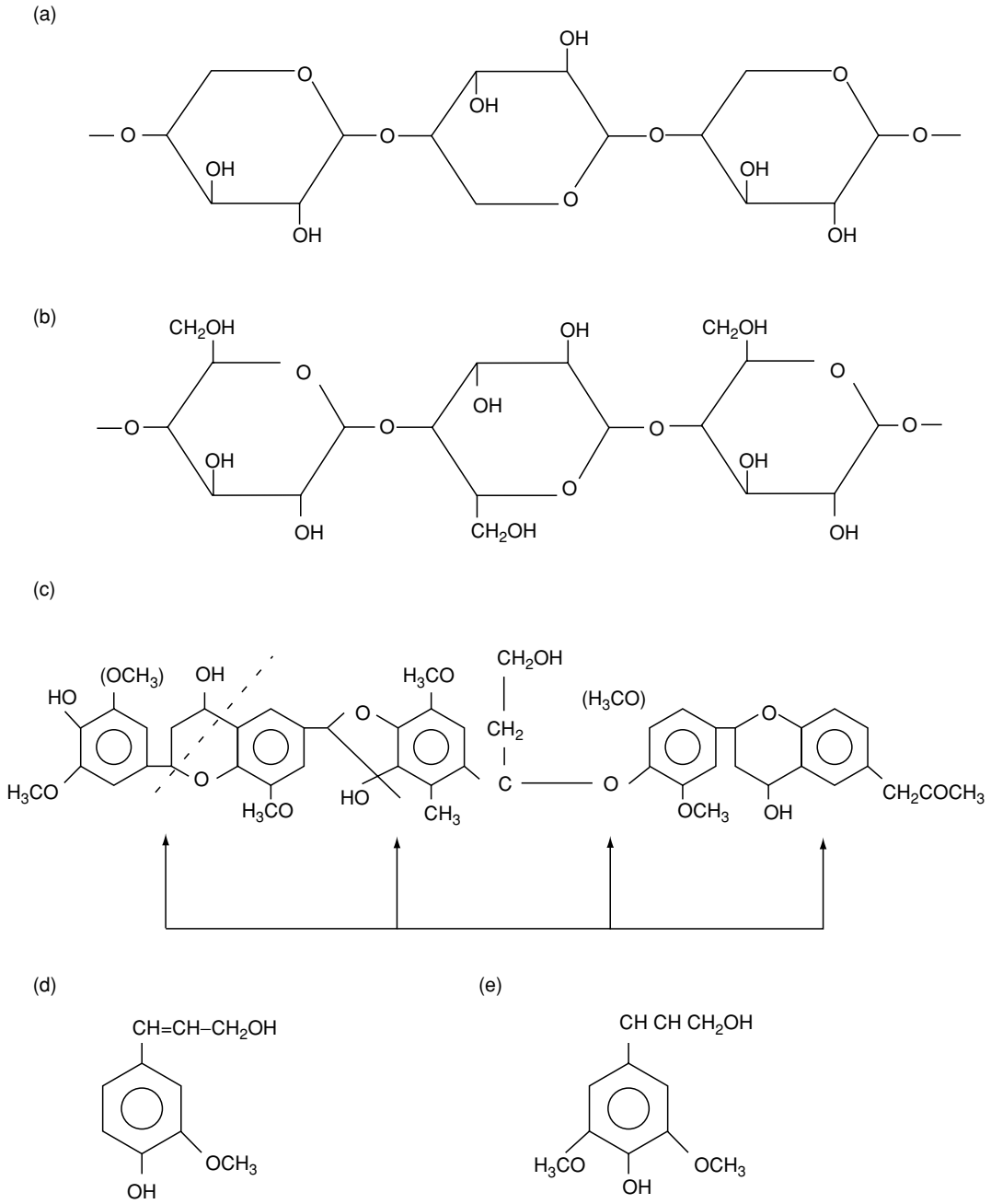

Figure 1 Main chemical constituents of 'hard' and 'soft' woods: 25% hemicellulose (a); 50% cellulose (b) and 25% lignin (c) from softwood and, with ... OCH_3, from hardwood. Arrows in (c) show splitting points during pyrolysis. Also shown are (d) coniferylalcohol and (e) sinapinalcohol.

types of smoke houses: (1) with natural air circulation, (2) air-conditioned or forced air, and (3) continuous. There are many modifications of these three types. Air-conditioned or forced-ventilation smoke houses have largely replaced the natural-air type. They allow much more precise control of smoking through uniform air movement and good control of temperature. Smoke houses with endless revolving chains have been developed specifically for frankfurter production.

Cold smoking Cold smoking was introduced as an alternative to hot smoking in the early part of the last century. Its general purpose is to impart a desirable flavor to the product, rather than preservation, which is achieved by refrigeration at low temperatures. In the kilns, the smoking temperature is maintained below 30 °C, and the time of smoking may vary from a few hours to several days, depending on the product. The shelf-life of the product can vary and

Table 3 Smoke compounds from different species of wood

Compound	*Amount of compound (grams per 100 g of sawdust)*					
	Acacia	*Beech*	*Oak*	*White beech*	*Spruce*	*Pine*
Monocarbonyls	1.70	1.77	2.40	2.50	3.06	3.10
Oxy- and diacarbonyls	1.86	3.60	1.98	4.95	6.30	6.12
Formaldehyde	0.11	0.15	0.14	0.16	0.21	0.20
Acetaldehyde	0.81	0.66	0.63	0.65	1.15	1.00
Acetone	0.64	0.74	0.37	0.69	0.69	0.54
Furfural	0.24	0.18	0.35	0.26	0.21	0.18
NaOH reducing substances	4.56	6.94	4.68	9.75	7.83	7.40
Reducants	0.13	0.16	0.17	0.32	0.21	0.29
Dehydroreductants	1.56	1.32	0.66	2.16	2.10	1.80
Total acids	2.94	4.06	3.56	6.79	3.07	3.17
pH	4.3	4.8	4.0	4.1	3.8	4.7
Esters	8.48	6.69	10.47	10.35	8.08	11.16
Phenols	0.36	0.14	0.16	0.32	0.29	0.29

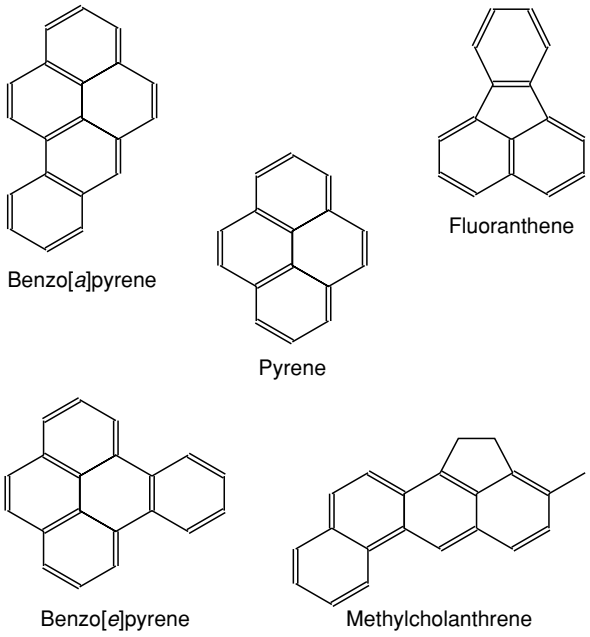

Figure 2 Structures of a few important polycyclic aromatic hydrocarbons found in smoke.

may depend on whether the product was brined or not prior to smoking.

Liquid Smoke

Manufacture Low-moisture saw dust is heated to 500–600 °C in a combustion reactor, and the resulting smoke vapor is passed to the primary tar and ash separation vessel and then on to a packed tower of ceramic beads. Here, the vapor smoke is condensed. The condensate is composed of (1) an aqueous phase made up of a mixture of low-boiling-point fatty acids, carbonyl compounds, and alcohols (pyroligneous acid, formalin, and furfurol in particular; formalin is responsible for immediate sterilization of the food surface, and the carbonyl compounds are claimed to be responsible for the formation of a desirable color) and (2) an immiscible tar or 'liquid smoke' (see Table 4). The tarry phase contains cresol, guaiacol, eugenol, methyl guaiacol, and pyrogallol to mention a few compounds (Table 5). These substances are credited with the antibacterial and antioxidant effect that smoking is supposed to have. The liquid smoke is stored in holding tanks for 2–4 weeks to allow separation of tars and other colloidal particles containing high levels of PAH, including benzo[*a*]pyrene. The remaining filtrate is a clear liquid that may be used as such, or it may be further concentrated by fractional distillation and extraction with selective solvents (see 'Smoke Flavors'). A schematic method of making PAH-free liquid smoke on a laboratory scale is shown in Figure 3.

Table 4 Composition of marketable smoke condensates from France, Germany, and the USA

Group	*Level (%)*
Water	11–92
Phenols (determined by GC)	0.2–2.9
Acids	2.8–9.5
Carbonyls	2.6–4.6

The tastes of the tarry and aqueous phases are different, and opinions differ as to which is to be preferred. When liquid smoke from different parts of the world are examined for their relative chemical constituents, it is confusing to note that some consist chiefly of acids, others of esters and carbonyl compounds, and some chiefly of phenols (Table 6).

There are four basic principles of application of liquid smoke: atomization, drenching, dipping, and direct addition. Further technical developments of these methods can be found under Application of Smoke Flavors.

Application of liquid smoke offers several advantages over the traditional smoking processes in that it offers consistent quality and a uniform and

Table 5 Sensorial judgment of a few phenols

Substance	*Best judged concentration (mg per 100 ml)*	*Smell*	*Taste*
Guaiacol	3.75	Phenolish, smoky, aromatic, sharp, sweet	Phenolish, sharp, spicy, smoked-sausage-aromatic, sweet, dry
Syringol	7.5	Smoky, spicy, aromatic, smoked-sausage-toned, phenolish, sharp, sweet	Phenolish, smoky, like freshly charred wood, like whiskey, dry, sharp
4-Methyl-guaiacol	1.9	Sweet, like vanilla, fruity, like cinnamon, somewhat smoky, pleasantly sharp, phenol tones	Sweet, like vanilla, caramel-like, aromatic, pleasant, smoke tones, burning
i-Eugenol	9.8	Sweet-fruity, like vanilla, like rhubarb, phenol tones	Sweet-fruity, mild smoke tones, dry, sharp
o-Cresol	7.5	Phenolish, sweet-fruity, aromatic, like caramel, smoked-sausage-tones	Sweet, sharp, unpleasantly smoky, burning
Dimethyl-phenol	0.9	Phenolish, like ink, aromatic, sweet	Phenolish, sharp, somewhat charred, sweet, dry

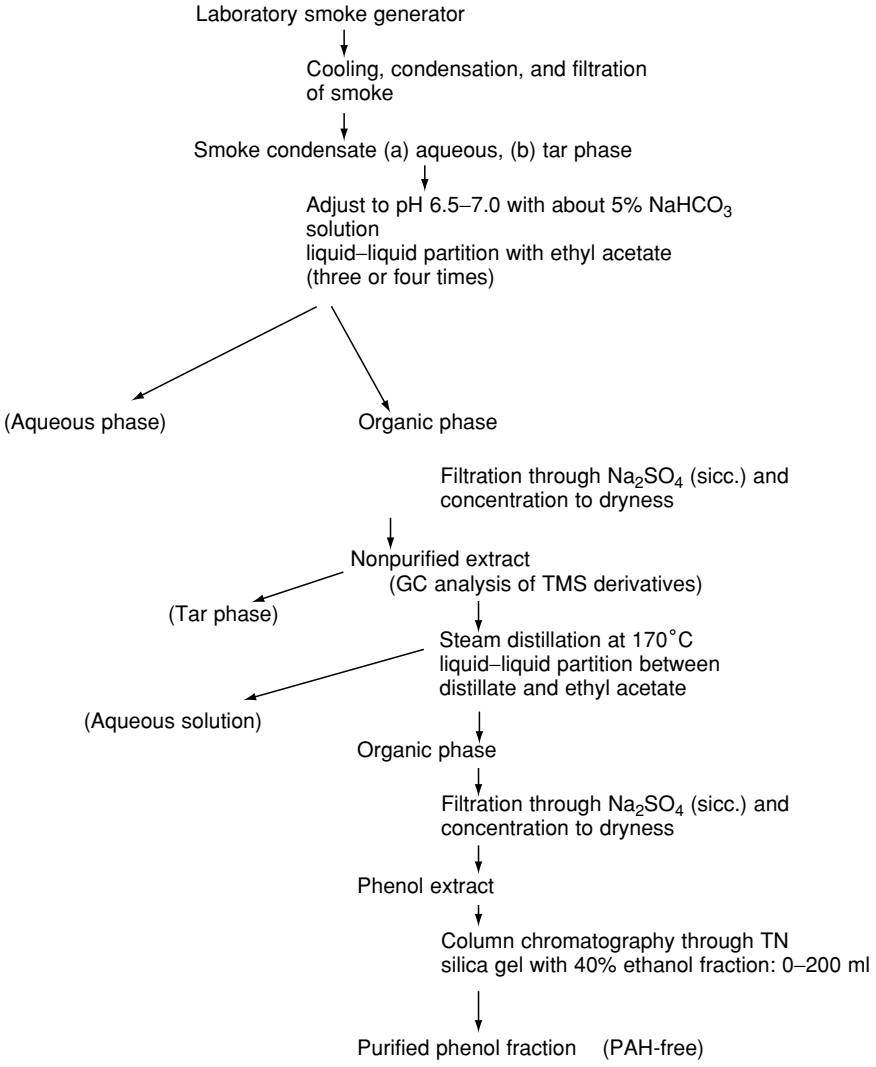

Figure 3 Outline scheme for the preparation of a PAH-free liquid smoke.

Table 6 Composition of a smoke condensate

Fractions	*Percentage of the whole condensate*
Formaldehyde	0.12
Aldehydes of higher molecular weight	0.57
Ketones	0.67
Formic acid	0.38
Acetic acid and acids of higher molecular weight	1.71
Methanol	0.96
Tar	4.81
Phenols	0.07
Residue	4.21
Water	82.42
Total	**95.92**
Extracts from activated charcoal	4.08

Table 7 Composition of a synthetic smoke aroma

	Percentage	*Level*
Furfural	0.17	0.05–0.8
Valeraldehyde	0.002	0.002–0.05
Dihydroxiaceton	0.0012	0.003–0.004
Diacetyl	0.0002	0.0001–0.001
Acetic acid	0.9	0.2–1.5
Formic acid	0.06	0.01–0.1
Propanoic acid	0.14	0.03–0.25
Butyric acid	0.016	0.001–0.03
Valeric acid	0.06	0.01–0.1
Caproic acid	0.008	0.003–0.05
Caprylic acid	0.006	0.001–0.01
Guaiacol	0.001	0.0002–0.01
o-Cresol	0.001	0.0002–0.01
Hexylamine	0.00016	0.00005–0.0008
Ethylamine	0.00016	0.00005–0.0008

reproducible product. Since the polycyclic aromatic hydrocarbons, tar and soot are removed in its manufacture, it is safer to use. It is economical in terms of time and ease of application, and there are no fire hazards.

Smoke Flavor

The toxicological problems associated with conventional smoking of foods have attracted a great deal of interest in recent years. There is evidence that the consumption of smoked meats, sausages, and fish over prolonged periods may be a causal factor in the development of cancer of the alimentary canal. For this reason, extensive efforts are being made to develop toxicologically safe smoke flavors. To achieve this, flavor technologists world-wide have attempted to split the smoke condensate into a number of subdivisions. Using a variety of countercurrent techniques, the fractions obtained are then examined to establish the elimination of PAH. Of all of the chemical constituents of smoke – alcohols, aldehydes, ketones, phenols, amines, tar, and cresol resin – it is the PAH typified by benzo[*a*]pyrene that are known to be carcinogenic and have thus attracted the most medical attention. Techniques are now known that can measure PAH down to 0.001 p.p.m. (*See* **Amines**; **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; **Flavor (Flavour) Compounds**: Production Methods.)

Opinions still differ as to which of the chemical constituents produce the characteristic smoke flavor. A study of the oil-soluble fraction of smoke condensate has been made by gas–liquid chromatography. After purification by extraction, the smoke flavor has been shown to contain some 20 different aromatic components, most of which are phenolic in nature. A few aromatic aldehydes have also been shown to be present (Table 7).

Organoleptic studies of the individual components or group of components showed *cis*-iso-eugenol, 2,6-dimethoxyphenol and 2,6-dimethoxy-4-methylphenol to be the compounds contributing most markedly to the characteristic smoky flavor. (*See* **Sensory Evaluation**: Taste.)

Optimum flavor characteristics can only be achieved by applying proper concentrations using appropriate technology. The concentration required to achieve the optimum flavor development depends on the type of product, its fat content, and the processing method. Examples of concentration of smoke flavor using one such commercially available preparation are: ham 45 p.p.m., bacon 32 p.p.m. (in both cases, added to brine), and sausages, cod roe, fish paste, and fish sausages 15–45 p.p.m. The object of conventional smoking is to increase the shelf-life of the product. It is thus imperative to ascertain what effect application of smoke flavor or liquid smoke has on the keeping qualities of the product. Recent research indicates that 2,6-dimethoxy-4-hydroxybenzaldehyde, one of the active components of smoke flavor, has marked antioxidative and antibacterial properties.

Application of Smoke Flavors

The flavor components are available in three physical forms: aqueous, oil-soluble, and dry. They may be applied to products by any of the following methods based on their suitability.

Spinning head spray system This method utilizes a variable-speed air motor that spins an atomizing disk at between 1000 and 18 000 rpm, creating a very fine atomization. The liquid is fed to the gun via an air-operated vernier-controlled pump, and an electrostatic charge is applied at the gun tip. This system can be used for spraying aqueous solutions of smoke extracts and flavor on to bacon, ham, and chicken, and in a solvent-based form on to a product such as salmon, where little or no water is required.

Airless spray system An airless pressure pump feeds the liquid through a spray nozzle. The electrostatic charge is applied by a probe at the spray tip. Various spray nozzle designs are available for achieving the desired pattern of application. Applications of this technique are in aqueous systems – extracts in brine solutions that allow smoking and curing of fish products in a single operation.

Powder spraying Powder flavors are obtained by mixing the flavor components with dextrose or starch base. The application system consists of a powder hopper and spray gun. The hopper has two regulators to control air supplied to a Venturi pump and to the air fan of the gun; a third regulator controls the degree of vibration at the base of the hopper. The gun has interchangeable heads that allow for variation of the spray pattern.

Flavor coating of savory snacks This system incorporates positive charging of the product and negative charging of the flavor medium. The plant consists of variably controlled fluidized bed vibrator feeding the flavor over a knife edge where it picks up a negative charge and projects it into a stream of positively charged product falling freely into the air. Thus, the flavor adheres all around the falling product. In addition to these high-tech systems, smoke flavors, brined or unbrined, can be injected directly to the product and absorbed during curing in massaging machines and tumblers. The advantage of this technique is that it simplifies application and improves control and hygiene.

It has been shown that water-based as well as salt-based smoke flavors can be used in the production of sausages and other comminuted meat products. They are ideal for pork, beef, horse meat, and reindeer meat. In the fish industry, smoke flavor is successfully used in the preparation of fish pastes, sausages, caviar, roe, and sprats. Smoke flavor is also applied to processed cheeses. It is reported that panelists could not differentiate between traditionally smoked and cheese to which smoke flavor had been added.

In canned products, smoke flavor can be added directly to the oil, thus enabling flavor penetration during heat processing of already sealed cans.

Technology of Smoking Fish, Meat, and Cheese

Fish

Operations necessary for production of smoked fish will depend on species, consumer preference, and product specification. However, the general procedure is as follows. Fresh grade A fish should always be used. They are washed to remove scales and surface lime, etc., and gutted and headed if required. The belly cavity is cleaned to remove traces of blood and bellywall lining. Fish are then usually brined before smoking. Brining is recommended on grounds of safety. It is recommended that to exclude *Clostridium botulinum* completely, fish should be brought to an internal temperature of 82 °C for 30 min during smoking if 3.5% brine solution is used, or 65 °C for 30 min with 5% brine. (*See* ***Clostridium***: Occurrence of *Clostridium perfringens*; **Fish**: Processing.)

After brining, fish are suspended or laid horizontally on racks in the smoking kilns. If equipment permits, smoking may be carried out in stages, for example, initially at 30 °C for 60 min, followed by 50 °C for 30 min, and finally at 80 °C for 2 h or more until the product has attained the required color and texture (Table 8).

Cold smoking of fish is commonly practiced on the leaner varieties of fish such as cod or haddock and may utilize traditional or liquid smoke. Depending upon the type of smoke, brined or unbrined fish is drenched with a suitable quantity of liquid smoke, which is then allowed to drain before being put through a drier to remove the excess moisture if long-term storage is envisaged. Nowadays, after addition of smoke, the fish may be vacuum-packed or refrigerated for transport to the point of sale. Virtually all types of fish may be smoked. The most popular varieties in the UK are smoked herrings (kippers), haddock, mackerel, and sprats. Smoked salmon is a great delicacy.

Meat

Common salt is the basic curing ingredient, although other compounds such as sugar, nitrite, and/or nitrate may also be added. (*See* **Meat**: Preservation.)

Salt in Cures

Salt is the basic ingredient necessary to all curing mixtures. It dehydrates and alters the osmotic pressure. It inhibits mold growth and subsequent bacterial spoilage. Its use tends to give a hard, dry, salty product that is not very palatable. It is therefore generally used in combination with nitrite/nitrate. A 65% 'pickle' cure is the most popular strength used for meats. (*See* **Curing**.)

Sugar in Cures

The addition of sugar is primarily for flavor. It also moderates the harshness of salt. During cooking, it forms browning products that enhance the flavor of cured meats.

Table 8 Fish smoking in a torry kiln

	Time in kiln (h)	*Temperature (°C)*	*Desired weight loss (%)*
Kippers	4	29	14–18
Kippers for canning (mild cure)	2.5	29	12–14
Kipper fillets	2.5–3	29	10–12
Cod or haddock fillets	3–5	27	10–12
Finnans	4–6	27	12–14
Pale cure	2–3	27	8–12
Golden cuts	2–3	27	10–12
Bloaters	3	32	14–16
Red herring	36*	29	20–25
Silver cure	8	29	20–25
Salmon	8–12	29	10
Cod roe	8	38	20–25
Buckling	2–3	27–71	10–12
Sprats for canning	0.25–0.5	27–71	10–12
Kielersprotten	1–1.5	27–71	10–12
Smokies	2	27–71	30
Trout	3	27–71	30
Eels	2	27–71	10–15

Nitrite and/or Nitrate in Cures

The addition of nitrates and/or nitrites offers the following advantages: (1) stabilization of the colour of lean tissue; (2) contribution to the characteristic flavor of the cured meats; (3) inhibition of growth of spoilage organisms, thus reducing the risk of food poisoning; (4) retardation of development of rancidity in the fatty constituents of the meats.

The processes of traditional smoking of meats may vary, but the general principles are the same. Prime quality meats are trimmed of most excess fat and then cured in a curing mixture (containing salt, nitrates and nitrites in water) for anything up to 7–10 days at 1–4 °C; meats may be pump-injected with curing mixtures before being smoked in a traditional wood-chip smoker for several hours. Depending upon the temperature of smoking, they may be cooked further in an oven until they are judged to be just right. Appropriately sized pieces may then be vacuum-packed and transported to the point of sale.

Nowadays, cured meats may be cooked in a mixture of liquid smoke before being packed, but evidence suggests that consumers prefer the traditionally smoked product and are prepared to pay more for it.

The most popular varieties of smoked meat products include the well-known hams, beef, horse, and reindeer meats. In addition, smoked poultry, turkey, and a variety of game birds can be found in specialist delicatessen shops. A number of different types of smoked meat sausages are available throughout the world.

Cheese

Certain varieties of cheeses, e.g., Seretpenir (Iran), Caramakase (Germany), Bandal (India), and Volgodski (Russia), are traditionally smoked. They may be smoked by hanging in a smoke-charged atmosphere without the necessity for higher temperatures. Both oak and applewood shavings may be used for smoke generation. Nowadays, incorporation of liquid smoke is becoming popular. The smoke preparation may be added to the milk or sprayed on to the curds before pressing. Occasionally, smoke flavor with salt may be applied as a spice. In another method of preparation, cheese is enclosed in a permeable membrane and immersed in a solution of smoke liquor.

Traditional smoking brings fat to the surface of the cheese, evaporates moisture, and incorporates smoke vapors containing phenolic substances that aid in giving a preservative effect as well as imparting a savory flavor to the cheese. The fat on the surface is also a deterrent to mold growth if the cheese is kept dry.

See also: **Amines**; **Antioxidants**: Natural Antioxidants; **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; ***Clostridium***: Occurrence of *Clostridium botulinum*; **Curing**; **Drying**: Theory of Air-drying; **Fish**: Processing; **Flavor (Flavour) Compounds**: Production Methods; **Meat**: Preservation; **Phenolic Compounds**; **Polycyclic Aromatic Hydrocarbons**; **Sensory Evaluation**: Taste

Further Reading

Arai I (1997) *Casing for Smoking and Storage of Food Stuffs*. European Patent Application.

Esterad-Munoz R, Boyle EAE and Marsden JL (1998) Liquid smoke effects on *Escherichia coli* 0157:H7 and its antioxidant properties in beef products. *Journal of Food Science* 63(1): 150–153.

Guillen MD and Ibargoitia ML (1998) New components with potential antioxidant and organoleptic properties,

detected for the first time in liquid smoke flavouring preparations. *Journal of Agriculture and Food Chemistry* 46(4): 1276–1285.
Guillen MD and Manzanos MJ (1997) Characterisation of the components of a salty smoke flavouring preparation. *Food Chemistry* 58(1–2): 97–102.
Korkeala H, Stengel G, Hyytia *et al*. (1998) Type E botulism associated with vacuum packaged hot smoked whitefish. *International Journal of Food Microbiology* 43(1/2): 1–5.
Moeller PW (1997) *Method of Making a Tar Depleted Liquid Smoke*, US Patent.
Olsen CZ (1976) Chemical Composition and Application of Smoke Flavour. *Proceedings of the European Meeting of Meat Research Workers*, 22, pp. F1–F8.
Ova G and Onaran S (1998) Polycyclic hydrocarbon contamination in salmon-trout and eel smoked by two different methods. *Advances in Food Sciences* 20(5/6): 168–172.
Perez-Rodrguez ML, Garcia-Mata M and Bosch-Bosch N (1998) The effect of adding smoke flavouring to frankfurters on nitrite and nitrate levels. *Food Chemistry* 62(2): 201–205.
Plahar WA, Nerquaye GA and Annana NT (1999) Development of an integrated quality assurance system for the traditional *Sardinella* sp. and anchovy fish smoking industry in Ghana. *Food Control* 10(1): 15–25.
Spink RR and Spink DWB (1998) *Apparatus for Hot Smoking of Food*. UK Patent application.
Toth L and Potthast K (1984) Chemical aspects of smoking meat and meat products. In: *Advances in Food Research*, pp. 84–158. London: Academic Press.
Underwood GL (1998) *Method of Preparing a Smoke Composition from a Tar Solution*. United States Patent.
United States of America, Association of Food and Drug Officials (1998) Cured salted and smoked fish establishments good manufacturing practices. *Journal of the Association of Food and Drug Officials* 62(1): 52–64.
Wendorff WL and Wee C (1997) Effect of smoke and spice oils on growth of molds on oil coated cheeses. *Journal of Food Protection* 60(2): 153–156.

SMOKING, DIET, AND HEALTH

M Indira, University of Kerala, Thiruvanathapuram, Kariavattom, India

Background

Smoking is the single largest avoidable cause of ill health. In addition to its effects on the physical and mental health of the individual, social, psychosocial, and economic consequences of cigarette smoking cause serious problems in the individual, the immediate family, and society at large. However, many hectares of land are used for the cultivation of tobacco plants, millions of dollars are pumped into the manufacture and advertisement of cigarettes, and huge amounts are being spent for the treatment of cigarette-related problems. All this could have been diverted for the health and nutrition of the underprivileged throughout the world.

Smoking cigarettes for as few as 5 years can cause permanent damage to the lungs, heart, eyes, throat, urinary tract, digestive organs, bones and joints, and skin. These effects may persist even if the smoker quits smoking after a few years of smoking. Many studies have proved that tobacco-related health effects decline substantially as the period of abstinence from smoking increases. Some of the benefits begin within months after quitting. Quitting improves pulmonary function by about 5%, prolongs life, and reduces the risk of tobacco-related cancers, myocardial infarction, cerebrovascular disease, and chronic obstructive pulmonary disease (COPD). Smoking directly irritates and damages the respiratory tract. This irritation and damage cause a variety of symptoms, including bad breath, cough, sputum production, wheezing, and respiratory infections such as bronchitis and pneumonia. One of the proposed mechanisms of smoke-induced toxicities is injury caused by free radicals. Hence, a diet rich in antioxidants and low in fat is recommended to counteract the injuries caused by smoking.

What is Smoking?

Smoking entails drawing the tobacco smoke from the cigarette, cigar, or pipe with the mouth, inhaling into the lungs, and puffing the smoke out.

Tobacco

Tobacco is an annual shrubby plant whose leaves are used chiefly in making cigarettes and cigars. Its leaves are large, with a color ranging from light green to dark green. The tobacco of commerce is derived almost entirely from *Nicotiana tobacum* L; Fam. Solanaceae. Certain other species, namely *N. rustia* (wild tobacco), *N. attenuata*, *N. trigonophylla*, and *N. quadrivalis* are also used, but to a lesser extent.

History

The Indians in the northern part of South America were the first to cultivate tobacco. When Christopher

Columbus and other early explorers arrived in America, they found the natives using tobacco much in the same manner as it is used today. They used it in their religious ceremonies. The Indians believed that it possessed medicinal properties, and that was the chief reason for its introduction in Europe. Indians used it as a remedy for aches and pains, snakebites, abdominal and heart pains, chills, convulsions, epilepsy, skin diseases, fatigue, hunger, wounds, sores, and numerous other ailments. From Europe, tobacco culture practically spread to the rest of the world. Jean Nicot, the French ambassador at Lisbon port, in whose honor the generic name '*Nicotiana*' was named, is said to have sent the seeds of *N. tabacum* to the queen of France, Catherine de Medici. Blessed with medical virtues and a quality to induce trance, smoking became very popular. Some of the factors that have contributed to its popularity include income from excise duty, automatic manufacturing machines, which lowered its price, advertising campaigns, and, above all, social acceptability. All this led to an increase in the number of smoking addicts. But at present, the world is turning away from cigarette smoking, following the American lead. The decline in smoking in the USA was initially triggered by the Surgeon General's reports on smoking and health. These reports were first published in 1964 and thereafter almost every year. Extension programs through mass media communications have led to a decline in smoking throughout the world. But now the challenge is to sustain this decline. Recent studies have shown that smoking is on the increase among teenagers and youths.

Composition

The cigarette burns at about 880 °C, and this burning of the cigarette produces smoke. The lit cigarette has a steep temperature gradient, 880 to 40 °C, which can be demarcated into three zones: (1) high-temperature zone (900–600 °C); (2) oxygen-depleter pyrolysis zone (600–100 °C); and (3) low-temperature zone (< 100 °C). Cigarette smoke consists of a heterogeneous mixture of gases, condensed vapor, and minute particles. Approximately 4000 chemicals have been detected in the mixture of gases and tarry droplets that comprise the cigarette smoke. The major components in the tobacco smoke are listed in Table 1. Of these, over 30 components have been suspected to be responsible for various health hazards, and of these, the major components causing or promoting various diseases are nicotine, carbon monoxide, and tar.

The gas phase accounts for about 60% of the total cigarette smoke. At least 11 of the gases have been found to be harmful to lung tissue. These are formaldehyde, carbon monoxide, acetaldehyde, acrolein, menthol, acetone, methyl ethyl ketone, ammonia, nitrogen dioxide, hydrogen cyanide, and hydrogen sulfide. Carbon monoxide makes up about 3–5% of the cigarette smoke.

Table 1 Major components in tobacco smoke

Acetaldehyde
Acetone
Acrolein
Ammonia
Anabaisine
Antabine
Arsenic
Benza(a)antracine
Benzo(a)pyrene
Cadmium
Carbon dioxide
Carbon monoxide
Copper
Cotinine
Formaldehyde
Hydrogen cyanides
Hydrogen sulfide
Menthol
Mercury
Methyl ethyl ketone
Nickel
Nicotine
Nitrogen dioxide
Nitrogen oxides
Nitrosamines
Nornicotine
Polynuclear aromatic hydrocarbons

Nicotine is a dense, oily alkaloid. Pipe tobacco contains 0.6–1.43% nicotine by weight. Cigars and cigarette tobacco have 0.75–3% and 0.9–1.96% nicotine, respectively. Nicotine has been reported to stimulate the release of adrenaline (epinephrine) and noradrenaline (norepinephrine), which have direct influences on the nerve centers controlling blood pressure and heart rate. It has been found to raise the tissue lipid levels and stimulate platelet adhesion.

The particulate phase contains a vast number of compounds, of which more than 250 have been isolated from cigarette smoke. About 30 metals are present in the tobacco smoke particulate phase. Several of these have been found to be carcinogenic. Tar is a dark brown viscous material composed of condensed particles from cigarette smoke after moisture and nicotine have been removed. Strong carcinogens like polycyclic aromatic hydrocarbons, nitrosamines, phenols, fatty acids, and their esters, etc., are some of the constituents that have been detected in tar. The tar component of cigarette smoke consists of all the particulate phase and much of the condensable components of the gas phase. The burning of an average-sized cigarette produces 3–40 mg of yellowish brown tar and about 1–2 mg of nicotine.

Why do People Smoke?

This habit stems mainly from psychological and social drives. It is a psychological tool, helping the smoker to relieve their stress, to relax, and to be at ease in social situations. Nicotine is the only psychoactive ingredient in tobacco smoke. It stimulates the production of powerful chemical messengers such as epinephrine, norepinephrine, dopamine, etc. The net effect is a temporary improvement in brain chemistry that is experienced by the smoker as enhanced pleasure, decreased anxiety, and a state of alert relaxation.

Second-hand Smoke, Environmental Tobacco Smoke (ETS), and Passive Smoking

The term 'passive smoking' describes the inhalation by a nonsmoker of the smoke of tobacco products from a smoke-filled atmosphere. The *Journal of the American Medical Association* has reported measurable levels of cotinine in the blood of 88% of all nontobacco users. Cotinine, is a metabolite of nicotine, and its presence is proof of a person's exposure to tobacco smoke.

The chemical constituents of ETS are derived from two main sources: (1) mainstream smoke and (2) side-stream smoke. The mainstream smoke emerges from the tobacco product during puffing. The side-stream smoke is produced during smoldering of the cigarette at the peak temperature inside the glowing cone. The chemical compositions of the side-stream and mainstream smoke are different. The temperature of combustion of the tobacco is different during puffing and smoldering. The substances are partially absorbed from the mainstream smoke by the smoker, depending on the characteristics of the substance, filter, additives, and depth of inhalation by the smoker. Analysis of the side-stream smoke has shown that it contains more combustion products of tobacco than mainstream smoke. In fact, it contains five times more carbon monoxide, three times more tar and nicotine, and four times more benzopyrine. Concentrations of nitrosamines can be up to 50 times higher in side-stream smoke than in mainstream smoke. The smoke in a smoke-filled atmosphere is mainly from side-stream smoke, which is a great threat to the health of the nonsmokers, especially children.

In 1992, the US Environmental Protection Agency (EPA) studied the respiratory health effects of environmental tobacco smoke. Their report, a compilation of 30 epidemiological studies, concluded that there is a strong association between ETS exposure and lung cancer. Scientists estimate that ETS is responsible for approximately 3000 lung cancer deaths per year among nonsmokers in the USA. On the basis of the health hazards of ETS, the EPA has classified second-hand smoke as a Group A carcinogen. ETS has also been linked with eye and nasal irritation in adults and middle ear infections in children, it also affects the cardiovascular system and causes acute and chronic coronary heart disease. ETS has also been linked with other cancers, including those in the nasal sinus cavity, cervix, breast, and bladder. Animal experiments support the view that second hand smoke can induce tumor formation.

In adults, ETS can worsen existing pulmonary symptoms for people with asthma and chronic bronchitis, as well as for people with allergic conditions. Even individuals who are not allergic can suffer eye irritation, sore throat, nausea, and hoarseness. Contact lens wearers can find tobacco smoke very irritating. Pregnant women, newborn children, and asthma sufferers have an increased vulnerability to cigarette smoke. It affects the development of the fetus and causes impaired growth, low birth weight, and sudden infant death syndrome (SIDS). Children are at particular risk from passive exposure to cigarette smoke, since they are more vulnerable to air pollutants and respiratory pathogens. Moreover, an infant or toddler has little opportunity to move away from the smoking parent.

Effects of Smoking on the Body

There is now an extensive body of evidence – including epidemiological, experimental, pathological, and clinical studies – to demonstrate that smoking increases the smoker's risk of death and illness from a wide variety of diseases.

Tongue

The pollution begins from the tongue. The surface of the tongue of a smoker who smokes on average 50 cigarettes per day results in a loss of taste. Inflammations are also found in the mouth cavity. The smoke causes pathological changes in the lingual papillae and adversely affects the mucous membrane of the mouth.

Ears

Smoking weakens hearing too. Even 20 cigarettes per day can weaken normal reception of speech.

Eyes

Researchers at the Johns Hopkins University have found evidence linking cigarette smoking and formation of cataracts. The eyes are quite sensitive to nicotine. Within a few seconds of the first puff, irritating gases (formaldehyde, ammonia, hydrogen sulfide and

others) begin to work on sensitive membranes of the eyes. Adverse ophthalmic effects of active smoking include glaucoma, cyanide-induced optic neuritis, corneal arcus, and vascular changes in the retina. The associated problems include a restriction to visual field, impaired reactions to dazzling light, etc. Smokers tend to have a deficiency of vitamin B_{12}. This deficiency is responsible for a disease called amblyopia. Tobacco amblyopia can produce dimming of vision in the central part of the visual field. Sufferers find it difficult also to distinguish between red and green.

Nose

Heavy smoking causes chronic inflammatory swelling of the nasal mucosa and damages the receptor cells, thus impairing the sense of smell. There is also an increase in the risk of nasal carcinoma in heavy smokers. Chronic exposure leads to damage of the olfactory cells, causing insensitivity to odors.

Lungs

Smoking directly irritates and damages the respiratory tract, and the respiratory rate increases. Irritating gases can damage the tissues of the lungs, and the airways leading to the lungs. This speeds up the production of mucus and leads to an increased tendency to cough up sputum. The excess mucus serves as a breeding ground for microbes. The lining of the bronchi begins to thicken and predisposes to cancers of the bronchi. The smoke weakens the free moving scavenger cells that remove foreign particles from the air sacs of the lungs, and continued smoke exposure leads to a variety of symptoms, including bad breath, cough, sputum production, wheezing, and respiratory infections such as bronchitis and pneumonia.

Many compounds that are inhaled are deposited as a layer of tar on the lining of the throat and bronchi and in the delicate air sacs of the lungs. This tar is rich in carcinogenic compounds. Lung cancer accounts for nearly 30% of all cancer deaths in the USA, and smoking accounts for nearly 90% of cancer deaths. The risks of dying from lung cancer are 23 times higher for male smokers and 13 times higher for female smokers than for nonsmokers. Additionally, smokers are at increased risk for cancer of the larynx and the oral cavity, and smoking causes a fivefold increase in the risk of dying from chronic bronchitis and emphysema.

Cardiovascular System

The effect of smoking on the heart and blood include an increased blood clotting rate, increased level of catacholamines, increased heart rate, decreased oxygen supply to the heart muscle, increased irritability to the electrical conducting system of the heart, and increased blood pressure. Blood pressure increases by 10–15%. Smoking increases the risk of vascular diseases of the extremities. Carbon monoxide from the blood binds competitively to the oxygen receptor sites of blood cells. Hemoglobin bound to the carbon monoxide is converted to carboxyhemoglobin, and so, less oxygen reaches the brain and other vital organs.

Cigarette smoking is now recognized as a major risk factor for atherosclerosis. A number of studies have indicated a close association between cigarette smoking and development of arterial lesions. Increased levels of tissue and serum cholesterol, triacylglycerols, and low-density lipoproteins (LDL), and low levels of high-density lipoproteins (HDL) have also been reported. A greater concentration of very-low-density lipoprotein cholesterol and a lower concentration of apoprotein 1 and HDL-C with a predominant decrease in the subfraction HDL2 have been observed in smokers. Alterations in the lipid profile have also been reported in the offspring of dams exposed to cigarette smoke. Lipid peroxidation and associated lipoprotein oxidation caused by cigarette smoke also accelerate the development of cardiovascular problems. Oxidized LDL is selectively taken up by monocytes/macrophages, unlike native LDL, resulting in the formation of cytokine-producing foam cells.

Skin

Smoking constricts the blood vessels of the skin, thus decreasing the delivery of oxygen. Hence, a smoker's skin becomes more susceptible to wrinkling. Smokers are at a particular high risk for a medical syndrome called smoker's face, which is characterized by deep lines around the corners of the mouth and eyes, a gauntness of facial features, a grayish appearance of the skin, and certain complexion abnormalities.

Male Reproductive System

Impotency has been observed in male smokers. Researchers have concluded that decreased potency might result from the detrimental effect of smoking on the blood vessels leading to the reproductive organs. Smoking can cause erection problems, decreases in sperm count and sperm motility, and impaired sperm morphology.

Female Reproductive System

Women who smoke experience gender-specific health consequences, including an increased risk of various adverse reproductive outcomes. A decline in fertility and earlier menopause have also been observed in women smokers, and smoking has been consistently associated with an increased risk for cervical cancer.

Smoking alters the level of estrogens and other hormones and interferes with hormone-related disorders. Some studies have suggested that cigarette smoking might alter menstrual function by increasing the risks for dysmenorrhea, secondary amenorrhea, and menstrual irregularity.

Fetus

Scientific literature published in the twentieth century has indicated that cigarette smoking during pregnancy adversely affects the fetal development and the development of the baby. A variety of mechanisms have been suggested, whereby maternal smoking habits could influence fetal health. These include the underlying constitution of the smoker, changes in nutrition, and impaired oxygen delivery, as well as specific effects of nicotine and polycyclic aromatic hydrocarbons upon the parents, uterus, and fetus. Placental transfer of nicotine has been observed. Nicotine has also been detected in the breast milk of female smokers. Fetal hemoglobin has a stronger affinity for carbon monoxide than maternal hemoglobin. In addition to this, nitrogen oxides oxidize hemoglobin iron in the ferrous form to the ferric form, to produce methemoglobin. Cyanide combines with methemoglobin, forming cynomethemoglobin. Cyanide also inhibits cytochrome oxidase. The combined effect of all of these leads to a hypoxia condition and affects fetal growth and delivery process. Infants born to women exposed to environmental tobacco smoke during pregnancy have a small decrement in birth weight and a slightly increased risk of intrauterine growth retardation compared with infants of nonexposed women. The risk for perinatal mortality – both stillbirth and neonatal deaths – and the risk for SIDS are increased among the offspring of women who smoke during pregnancy.

Paternal Smoking

A few studies have revealed the effects of paternal smoking and genetic damage to the fetus. The tobacco smoke containing carcinogens may reach the gonads through the blood. Since spermatogenesis continues throughout the male's reproductive period, these carcinogens may genetically alter the sperm. There is an increased risk of genetic defects to the offspring of fathers who smoked at the time the child was conceived.

Immune System

Cigarette smoking weakens the immune system by depressing the antibody response and decreasing cell-mediated reactions to foreign invaders.

Brain

A reduced level of oxygen carried to the brain can produce lethargy, confusion, and difficulty in thinking clearly.

Oxidative Damage and Pathogenesis

One of the major mechanisms for smoke-induced injury is the oxidative damage. Five per cent or more of the oxygen is converted to reactive oxygen species (ROS) such as O_2^-, H_2O_2, and $^{\cdot}OH$. This process can be represented as

$$O_2 \xrightarrow{e^-} O_2^- \xrightarrow{e^-} H_2O_2 \xrightarrow{e^-} OH^- \xrightarrow{e^-} H_2O.$$

Super oxide anion — Hydroxyl radical

The super oxide radical is formed in almost all aerobic cells by three ways:

1. Normal oxidative metabolism.
2. Microsomal cytochrome P_{450} activity that is inducible by a variety of foreign compounds.
3. Respiratory burst of stimulated phagocytes.

Free radicals have an odd electron in their outer orbital, and so they tend to be hyperactive to attain a stable form. They either abstract electrons from, or donate their odd electrons to, other molecules in their proximity. In this process, they modify and damage biomolecules. Metal ions like cadmium, iron, and copper are powerful promoters of free-radical generation.

The gaseous and tar phases of cigarette smoke contain high levels of carbon and oxygen-centered free radicals as well as $NO^{\cdot}$ and NO_2 (which are capable of attacking compounds with >C═C< bonds), which can directly initiate and propagate the process of lipid peroxidation of membrane lipids both *in vivo* and *in vitro*. Their approximate concentrations in the gaseous and tar phases are, respectively, 10^{15} per puff and 10^{17} per puff. Cigarette smoke also contains traces of metal ions – cadmium, copper and iron – that are powerful promoters of free-radical generation. They can react with H_2O_2 to form $OH^{\cdot}$, the most reactive oxygen radical known. Cellular antioxidants (vitamin C, E, glutathione, selenium, etc.) and free-radical scavengers (super oxide dismutase, glutathione peroxidase, catalase, etc.) normally protect a cell from the toxic effects of ROS, but when the balance between ROS production and antioxidant defenses is lost, oxidative stress results, which, through a series of events, deregulates the cellular activity leading to various pathological conditions. These reactions can alter the intrinsic membrane properties like fluidity, ion transport, and loss of enzyme activity, protein cross-linking, inhibition of protein synthesis, DNA damage, and,

ultimately, cell death. This is the underlying mechanism of the formation of various diseases, including cardiovascular dysfunction, neurodegenerative diseases, gastroduodenal pathogenesis, metabolic dysfunction of almost all vital organs, cancer, and premature ageing. Polyhydroxyl aromatic compounds present in tar have been proposed to bind to genetic material, leading to the generation of H_2O_2, which in turn causes DNA strand breakage via a hydroxy radical-dependent mechanism.

Other Forms of Smoking

Cigar Smoking

Cigar smoking is increasing in the USA, since it is considered to be a safer alternative to cigarettes, but this is not so. It is a risk factor for cancer of the mouth, throat, larynx, esophagus, and lungs.

Beedi Smoking

In India, Beedi (tobacco wrapped in tendu (*Diospyros melanoxylan* Roxb.; Fam: Ebenaceae) leaf) is used for smoking, mostly by people belonging to lower socioeconomic strata. Other forms of smoking are chutta, a type of homemade cigar smoked with the lit end placed inside the mouth, hookahs (or water pipes), and pipes. Contrary to the belief that traditional forms of smoking are less dangerous than smoking cigarettes, it was found that the relative risk of cancer and cardiovascular problems were higher in the users of beedi, chutta, or hookhas.

Strategies to Keep Healthier

Quitting cigarette smoking is the best measure to keep healthy, but some may find it difficult to give up the habit. A reduction in the number of cigarettes smoked, and a change in the life pattern may provide some protection from smoke-induced toxicities.

Exercise

In addition to the psychological benefits, a few studies have suggested that smokers who begin exercising regularly may also obtain a small degree of protection against the health risks of of smoking.

Nutrition

It is widely accepted that smoking tends to reduce nutrient intake, but it is not known for sure whether it acts directly upon the satiety centers, affects the sense of taste or smell, or operates less directly through psychological mechanisms. Smoking itself reduces the intake of vitamins such as ascorbate, tocopherols, and folate. Smokers have low levels of vitamin C and carotene. An *in-vitro* study showed that a wide spectrum of lipophilic micronutrients (retinal, carotene, vitamin E) undergo degradation when exposed to the gas phase of cigarette smoke. Experimental and epidemiological studies have found some deficiencies of antioxidants in smokers.

The toxicities caused by smoking are mediated through ROS. Hence, a diet rich in vegetables and fruits is especially beneficial in reducing the toxicities induced by smoking. Their beneficial effects are due to the presence of antioxidants and anticancer nutrients (vitamins A, C, and E, selenium, isothiocyanates, indoles, calcium, fiber, vitamin B_{12}, etc.). Many human studies have supported this view. There is a fivefold decrease in risk ratio for cataracts in persons consuming ≥ 1.5 servings of fruits and/or vegetables. Vitamin B_{12} helps to detoxify the cyanide found in tobacco smoke. Smokers excrete more of this vitamin. This condition is treated with large doses of vitamin B_{12}. Hence, supplementation of antioxidants, either through increased intake of vegetables and fruits or through supplements, is essential in a smoker's diet.

Smoking is an independent risk factor for atherosclerosis. It increases the lipid content of tissues and serum, and so a low-fat diet rich in fiber is ideal.

See also: **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; Synthetic Antioxidants, Characterization and Analysis; Role of Antioxidant Nutrients in Defense Systems; **Ascorbic Acid**: Properties and Determination; **Cancer**: Carcinogens in the Food Chain; Diet in Cancer Prevention; Diet in Cancer Treatment; **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; **Coronary Heart Disease**: Etiology and Risk Factor; Antioxidant Status; Intervention Studies; Prevention; **Minerals – Dietary Importance**

Further Reading

Bendich RJ and Deckelbaum J (eds) (1997) *Preventive Nutrition: The Comprehensive Guide for Health Professionals*. Totowa, NJ: Humana Press.

Ferguson T (1987) *The Smoker's Book of Health*. New York: G.P. Putnam's Sons.

Halayya M (1992) *Stop Smoking and Start Living*. Calcutta: Rupa & Co.

Kumar A (1999) Environmental tobacco smoke: A big threat. In: *Environmental Problems Protection and Control*, vol. 2. New Delhi: Anmol.

Rana Sastry BV and Janson VE (1995) Smoking, placental function and fetal growth. In: Ramasastry PV (ed.) *Placental Toxicology*. Boca Raton, FL: CRC Press.

SNACK FOODS

Contents

Range on the Market

A S Bawa and J S Sidhu, Guru Nanak Dev University, Amritsar, India

Introduction

The term 'snack' or 'snack food' is difficult to define or categorize. The dictionary meaning of snack is a 'titbit' which is a small meal in the broadest sense. Snacking can be described as the problem-free consumption of easy-to-handle, miniature-portioned, hot or cold products in solid or liquid form which need little or no preparation and are intended to satisfy the occasional pangs of hunger. Thus snacks should be in convenient manageable portions satisfying short-term hunger. The different categories of snack foods are given in Table 1.

Snacks can be sweet or savoury, light or substantial, and they may be endowed with attributes such as healthy or just for fun. In European markets the word 'food' is omitted. This article deals with various aspects of different snack foods and the quality checks involved.

Snack foods are a significant part of the food industry. The dominant leader is still potato chips (crisps), followed by extruded snacks, corn chips, nuts, meat snacks, pretzels, and popcorn.

Potato-Based Products

Potato-based snacks can be made either exclusively from fresh, sliced potatoes or from potato dough. The chips in the latter case are called simulated potato chips. The steps involved in potato chip processing include selection, procurement and receiving, storage under optimum conditions, peeling and trimming of potato tubers, slicing, frying in oil, salting, or the application of flavored powders and packaging. (*See* **Potatoes and Related Crops**: Processing Potato Tubers.)

Potato must be of high specific gravity in order to obtain superior yield, low oil absorption, and low reducing sugar content. Peeling should remove only a very thin layer of potato, leaving no eyes, blemishes or other material for later removal by hand trimming. The chips from peeled and unpeeled potatoes are similar in appearance, flavor, and shelf-life. The chips from unpeeled potatoes have higher oil contents but omission of the peeling step results in a 7% increase in potato solids and a reduction in waste.

The peeled potatoes are cut into slices 10–17 mm thick by rotary slicers. Slices must be very consistent in thickness in order to obtain uniformly colored chips. Slices with torn surfaces lose excess solubles from the ruptured cells, and absorb large amounts of fat. Starch and other materials released from cut cells are washed away prior to frying and the surface of the

Table 1 Snack food categories and examples of products in each category

Category	*Product*
Hot snacks	Minipizzas, pizza baguettes, etc., toasts au gratin, cup noodles, spring rolls, filled croissants
Cold snacks	
Milk and dairy products	Yogurt, plain or with fruit, mini cheese cubes
Bakery products	Cake bars, minitarts, biscuits
Bars	Granola or muesli bars, chocolate bars, minibreak bars, energy bars
Savory products	Chips (crisps), sticks, crackers, pretzels, salt sticks
Other products	Popcorn, puffed cereals, rice snacks, fruit sticks, dip sticks, tortilla chips, fruit snack bar
Confectionery	Plain cookies, deposit cookies, macaroons, wire-cut cookies
Extruded snacks	Potato sticks, cereal-based tubes with cheese filling, corn pellets and chips
Low-pressure extruded fried snacks	Sev, omumbadi, murukku
Nut-based snacks	Roasted peanuts, fried peanuts, coated and fried peanuts, toasted and salted pecans, roasted and salted almonds, sugared and spiced nuts, flavored nuts and nut mixtures
Legume-based Indian snacks	Fried green peas, chick peas, lentils, dhals, fried products from bengalgram meal
Meat-based snacks	Expanded pork rinds or skeen, jerky, chimni

From Tettweiler P (1991) Snack foods worldwide. *Food Technology* 45: 58.

slices dried; this step shortens the frying time and increases the capacity of the cooking unit.

The rinsed and dried potato slices are conveyed directly to the frier. The fried slices are removed from the tanks and drained on mesh belts. The temperature of oil in friers is held within a narrow range. Temperatures normally used are 176–190 °C at the receiving end and 160–173 °C at the exit end of the frier.

The oil used for deep-fat frying of potato chips has two functions: (1) it serves as a medium for transferring heat from a thermal source to the tuber slices; and (2) it becomes an ingredient of the finished product. The oil should be highly refined. The flavor and appearance of chips are affected both by the amount of oil absorbed and its characteristics as it exists in the chips. The amount of oil absorbed by potato slices is affected by: (1) the solids content of the tuber; (2) the temperature of the fat; (3) the duration of frying; and (4) the thickness of slices. Fat pick-up by potato chips is 10–15% lower when the chips are fried in liquid oil as compared to fat which is solid at room temperature. Reducing slice thickness increases the oil content of chips. A hot-air blast at the point where the slices emerge from the fat reduces fat absorption, and partial drying of sliced potatoes before frying also reduces the oil content of chips. Leaching raw slices with hot water results in an increased uptake of oil. The treatment of slices with glucose oxidase produces a chip of higher uniform color and lower oil absorption, without off-flavors.

Baked Snacks

Soda Crackers

Soda crackers and variants, suck as saltines, oyster crackers, etc., are themselves used as snacks, and also serve as a basis for combination snack foods either made in the home (crackers and cheese) or by manufacturers (cheese crackers, peanut butter and cracker sandwiches, etc.) The soda cracker is made from a lean, fermented dough. It does not contain much shortening, sugar, or milk. (*See* **Biscuits, Cookies, and Crackers**: Nature of the Products; Methods of Manufacture; Wafers.)

Flour constitutes 80% or more of finished crackers. Its qualities are important in determining the machining qualities of the dough, as well as the texture of the end product. Weak flours and lengthy fermentation times combine to yield flat, tender crackers. Flour for thick saltines should be stronger than for thin crackers. Thin saltines require a weaker flour; salt may be added at a rate of 2–5% to the dough.

Sprayed crackers are rich crackers, round in shape, and usually made from chemically leavened dough. Many of the representatives of this class could be considered a cross between crackers and cookies because they are not only leavened but also sweeter than saltines. The leavening system for these is adjusted so that the pH of the finished products is below neutrality. A pH of 6.5 is regarded as desirable. Sprayed crackers are coated with 20–25% (w/w) of a blend shortening. Coconut oil is preferred, but peanut oil or even hydrogenated vegetable shortening can be used. The oil is applied either by spraying at 65–71 °C, or by a device similar to an enrober, in which the crackers pass through a flowing curtain of oil. The oil must be applied when the crackers are still hot. The crackers may be salted very lightly on the cutting machine.

Cheese Crackers

Cheese crackers are usually made from fermented dough and they must be on the acid side of neutrality to yield a typical cheese flavor. The formulation is based on the soda cracker, except that fat and moisture added in the fresh cheese have to be compensated for. Paprika (0.25%) and a very small amount of cayenne are added to Cheddar cheese crackers to intensify the color and to add a flavor which is complementary to cheese crackers to impart certain desirable characters. Natural cheese should be finely ground before being placed in the mixer. Sometimes a premix of ground cheese and shortening, kept previously in the fermentation room for 24 h, is added at the doughing stage. This holding stage encourages the development of more flavor, and leads to a better dispersion of cheese in the dough. Artificial Cheddar and blue cheese flavors are also used whenever required.

Pretzels

Pretzels are considered to be the world's oldest snack food, and may have originated around AD 610 on the borders of France and Italy. Pretzels then become a popular snack food in Germany and Austria, where they were called 'pretzel,' later to be altered to today's pronunciation, 'pretzel.' Originally, all pretzels were soft, like unleavened bread. Accidentally, some pretzels were left in a cooling oven and the remaining heat dried them, removing the moisture and giving them a hard crisp texture and a golden coating. This led to the introduction of hard pretzels.

Soft pretzels today are baked, not dried, and have a texture resembling that of a fresh, soft roll, as well as a high moisture content. Soft pretzels are popular as a snack purchased at street vendors' stands. Hard pretzels have an extremely long shelf-life, especially when stored in airtight containers. Hard pretzels can be thin or thick, and shaped as twists, sticks, nuggets, logs, or

rings. They can be buttered, wholewheat, sesame-coated, chocolate-dipped and 'teething.' It is the drying which gives the hard pretzels their distinctive crunch, and this crunchiness is fundamental to the appeal of hard pretzels.

Pretzel doughs are made very stiff so that they can withstand the punishment of machining without becoming sticky or misshapen. The sponge is fermented for a shorter time than with cracker dough. Doughs receive a short proof stage but machinery steps, including formation of the pretzel, are handled automatically in all but a very few small plants. The characteristic gloss of the pretzel is the result of a lye dip. The lye dip solution contains about 0.5% sodium hydroxide or 2% sodium carbonate and is maintained at about 100 °C. The immersion time is about 10 s. The solution may also be applied by spraying. Immediately after the pretzels leave the caustic solution they are salted. The general aim is to achieve 2% salt in the finished product, but it is necessary to keep the initial application rate at 8–10%, to allow for losses in processing. A long drying time is required to reduce the moisture content to 2–2.5%, and temper the pretzel so that it does not break too easily during packaging. The long pretzels can be twisted or stick-type. Stick pretzels are extruded using a group of five extruding heads having 10–12 holes each. Logs and nugget-type pretzels are made in a similar manner to that used for the sticks, except that they are cut into short lengths at the extruded head. The stick pretzels can be made from almost any flour, but the choice of flour used in twisted pretzels is critical.

Filled pretzel sticks or nuggets are also available. These are produced by drilling a hole in the completely baked pretzel stick, and then inserting a paste-like filling of peanut butter and cheese, along with oil and sugar or a nonsweet agent such as lactose or dextrin.

Confectionery (Sugar-Rich Products)

Plain cookies are the cookies made in one operation, and this group does not include filled, coated, sandwiched, and other multiple-component cookies. The continuous structure of a cookie arises from the flour, and this basic framework is tenderized with sugar, invert sugar, egg yolk, ammonia, soda (or baking powder), and shortening. It is firmed or toughened with water, cocoa, egg white, whole egg, milk solids, and the leavening acids. Shortening is one of the principal agents for increasing tenderness, at least as far as rich, sweet biscuits are concerned. Too much shortening leads to a greasy smeary cookie which is susceptible to rancidity because of the free fat, while too much sugar results in hard and excessively sweet cookies.

A wide variety of flours can be used, varying from a soft cookie flour to a rather strong sponge flour. The stronger the flour, the more shortening and sugar are required to obtain an acceptable texture. High protein contents lead to hardness of texture and coarseness of internal grain and surface appearance. Chlorine-bleached flours should not be used for soft cookies where relatively large amounts of tenderizing and moisture-retaining ingredients, such as sugars, shortening, and egg yolk, are used. Cookies may become fragile if the quantity of flour is decreased too much.

Deposit Cookies

Deposit cookies are machine-made counterparts of the 'hand-bagged' cookies. They contain about 35–40% sugar, 65–70% shortening, and 15–25% liquid whole egg. The flour should be from soft wheat, unbleached with 8–8.5% protein and 0.35–0.40% ash. It should have a viscosity of 40 Ns/m^2 and a spread factor of 79–80. The flour should be able to carry sugar and shortening without too much spread, so that the top design is preserved through baking. The flour or other ingredients must contribute enough adhesive properties to the dough so that it will adhere to the band and pull away from the main column of the dough in the deposit stage.

Macaroons are made by a cold, or hot syrup cooking process, but they should not be made by the cold process unless they are to be consumed within 3 days.

Wire-Cut Cookies

The dough composition varies over a wider range for wire-cut cookies than for any other type. Thus the dough material should be sufficiently cohesive to hold together as it is extruded through an orifice, and yet be nonstick and short enough so that it separates cleanly as it is cut by wire. These cookies can be further subdivided as follows: (1) drop-type, which are used in sandwich cookies filled with a marshmallow or imitation cream, and usually containing equal amounts of sugar and flour; (2) sugar cookies, molasses cookies, coconut, raisin, date and honey varieties; (3) shortbreads, in which the shortening content is usually 50–70% of the flour content; and (4) macaroons, with little flour and large proportion of sugar.

Nut-Based Snacks

Nuts are often used to upgrade popcorn-based snacks, and are also themselves sold for snacks either as the individual variety or as mixed nuts.

Peanuts

The chemical composition of peanut is approximately 46% fat, 26% protein, 18% carbohydrate, 6% water, 2% ash and 2% fiber.

Roasted Peanuts

Peanuts are sorted, blanched, roasted, and salted. Blanching for peanuts and other nuts means removal of the skin or testa. During roasting, the kernels become dehydrated, and browning reactions occur throughout the kernel. The texture, flavor, and appearance change markedly. Toasting can be continuous or batch-type, and dry or in oil. The unique nutty flavor of roasted peanuts results from browning reactions of the Maillard type. The type of roasting affects the course of browning reactions because of variations in the rate of heat penetration and exposure to high temperature in different parts of the nut. Adhesion of salt to dry-roasted nuts is achieved by adding about 2% (w/w) coconut oil as a dressing to the warm nuts, and immediately applying the salt. The salt must have small particles of irregular configuration to obtain maximum adherence.

Fried Peanuts

Peanuts are fried at different temperatures depending on the flavor the customer wants to impart. Most peanuts are fried at temperatures around 140–160 °C but some customers prefer a slightly higher temperature in order to impart a roasted-burnt flavor into the peanut.

After the peanuts are removed from the frier, it is essential to reduce their temperature rapidly as they retain a lot of heat and will continue to darken as the Maillard reaction continues to occur with elevated temperatures. When they are cooled glazing oil can be applied to allow salt or flavoring to adhere to the peanuts. When nuts are fried in hydrogenated fat, the salt is applied before they are cooled to room temperature and while the fat on the surface is still molten. As the fat cools and sets, the salt granules are held firmly.

Coated and Fried Peanuts

A typical process would involve the peanuts being placed into a coating pan with a small quantity of starch. The peanuts will be roasted to insure they are completely coated. An adhesive solution will be placed on to the edge of the coating pan and the drum rotated. Starch is added on top of the peanuts and they are again rotated to insure the coating is smooth. This process is repeated until the desired coating thickness is achieved.

When the peanuts are deposited into the frying system, the starch contained within the coating will begin to gelatinize. This causes the exterior of the product to become sticky and the peanuts must be agitated to prevent them from sticking together. The peanuts are fried at a temperature of approximately 145 °C. The frying time varies due to different types and levels of coating on the peanuts. Peanuts that are completely coated require a longer fry time and lower temperature than peanuts that are only partially coated. The recommended frying temperature is lower than the temperature at which caramelization of sugars begins (i.e., 160 °C). This limits the color development of the coating and allows the moisture to be removed from the center of the peanut before the outside coating darkens. The moisture content of the interior of the peanut needs to be reduced to below 2.0% to produce a shelf-stable product. After frying the peanuts are cooled and can be flavored if desired. (*See* **Peanuts**.)

Pecans

Toasted and salted pecans are used directly as snacks. Blanching is never required in pecans. Pecans can either be dry-toasted or cooked in oil. Much less heat treatment is required than for peanuts, because the initial moisture content of pecans is lower, and the development of a true roasted flavor is seldom desired. Frying gives a more uniform roast and causes less mechanical damage. Roasting is performed, on an average, for 12–15 min at 190 °C. After roasting, the kernels are quickly cooled in tunnels or perforated bins, with forced air circulation to remove any undesirable odors. The cooled product is cleaned to remove any debris, foreign matter, and scorched nuts. The roasted kernels are coated with about 1% (w/w) finely ground salt. Pecans may be coated with cinnamon, sugar, barbecue-flavored powders, etc. Since toasted pecans have a very limited storage life, they are vacuum-packed.

Almonds

Almonds are a common constituent of mixed-nut packs, and are popular as the roasted and salted variety. The kernels are not generally blanched for the production of salted almonds. For blanching, the nuts are soaked in hot water until the skin loosens sufficiently to be removed by rubbing. Almonds may be dry-roasted or fried, but oil roasting is the more common method. In dry roasting, hot air is passed through rotating perforated cylinders or a fluidized bed to reduce the moisture content to 1%. Oil roasting is carried out by immersing the nuts in oil, preferably coconut oil, and heating to between 121 and 176 °C. After frying, they are cooled, salted, and

dressed with oil. They can also be coated with smoke-flavored salt, cinnamon, sugar, or similar flavored powders. (*See* **Almonds**.)

Sugared and Spiced Nuts

Sugared and spiced nuts are prepared by adding nuts and spices to water and sugar heated to 115 °C in a revolving pan. After coating, the nuts are dried for 1 h at 48 °C, or 2 h at room temperature, and packed in hermetically sealed containers, preferably under vacuum. The nuts must be dried to below 4% moisture before coating. This product should be refrigerated if not consumed within a month.

Extruded Snacks

Single- or twin-screw extruders are used to get extruded snacks. There are three types of extruded snacks: (1) direct expanded snacks, (2) coextruded snacks, and (3) indirect expanded snacks.

Direct Expanded Snacks

This category gets its name from the fact that products expand directly out of the extruder die and require no further processing, such as corn curls, both baked and fried, onion rings, and potato sticks. The puffing of direct expanded snack products is created by heating the ingredients to temperatures over 100 °C. Inside the extruder the water in the dough mass remains a liquid because the dough is under pressure. When the dough exists the extruder through the die opening, the superheated water is exposed to atmosphere. The vaporization of water into steam during this rapid pressure loss causes stretching and expansion of the starch matrix. The equipment used to make such products ranges from simple collect extruder to twin-screw extruders. Potato sticks are made from 65% potato starch and 35% corn meal. Blend is extruded to cook and puff. The high level of potato starch gives this product a unique flavor. The baked collet is then coated with oil, salt, and other seasoning.

Coextruded Snacks

Coextrusion technology became popular in the snack food industry around 1984. A coextrusion process is a technique where two materials (usually dissimilar in nature) are combined in an extrusion die. The two materials can come from two extruders or from an extruder and a pump. This process gives the manufacturer the ability to make a product which has two textures, or two colors or two different flavors. In snack food industry the most common form of coextrusion is the extrusion of a cereal-based outer tube with a cheese filling pumped into the center of the tube.

Indirect Expanded Products

Indirect expanded snack products are also referred to as 'third-generation snacks' and 'half-products.' These products are not expanded directly through the die. The additional process steps which contribute to the appearance or texture include frying or hot-air puffing to remove moisture and achieve the final texture. There are two subcategories of indirect expanded products: (1) pellets; (2) fabricated chips.

Pellets

Snack pellets are products cooked in an extruder and forced through a die at temperatures below 100 °C. The shape of the dough at low temperatures prevents the puffing forces of water turning into steam, thus intricate shapes can be made in pellets which would normally be destroyed in the direct expanded extrusion process. Snack pellets are extruded at 25–35% moisture and then dried to around 12% moisture for stability. Since the snack pellet is unexpended and dense in this stage, drying must be done at low temperatures and high humidity to allow the moisture to migrate out of the pellet. So snack pellets can be held for longer periods, shipped to long distances, or distributed to the snack food manufacturers or consumer directly. The pellets are characterized by a translucent appearance, due to above 90% level of starch gelatinization. Then pellets can be expanded by heating them rapidly to cause the water in the pellet to turn into steam in an explosive manner which puffs the pellet. This can be accomplished by frying, baking, or microwaving, then coating with seasoning and packing. Frying is the most common method because it is less expensive and works well. Frying conditions depend on the pellet thickness, shape and composition. Frying temperatures usually range between 170 and 210 °C for 10–60 s. For low-oil snack products, baking and microwaving have been explored. Due to rapid heat transfer rates required to expand a pellet uniformly, the ordinary snack dries and a baking oven will not puff a pellet. Microwaving of pellets also requires a rapid heat transfer rate, therefore high-powered continuous microwave expanders are commercially used. Expanded pellets have a moisture content of 1–2%. They are usually flavored in coating drums as for direct expanded products.

Fabricated Chips

There are many fabricated chip products (FCP) made through a wide variety of processes. One of these

processes is extrusion-based, similar to the process for making pellets. The main difference is that FCP are not dried to below 12% moisture to be stored or sold. They are also cooked less during the extrusion step. After extrusion, they are fried to remove moisture and complete the cooking process. Fabricated chips are usually sheeted out of the extruder and cut into a shape which simulates a potato or a tortilla chip.

Corn Chips and Simulated Potato Chips

Commercial simulated potato chips have assumed great importance. The corn chips and tortillas prepared from alkali-treated corn dough have similar flavors and textures. The ground, alkalized corn dough, or masa (Table 2) is conveyed to the hopper of the tortilla cutting head by means of a screw-type extruder. A thick sheet of masa is extruded between sizing rolls, and the resulting ribbon is cut into dices of appropriate size. Tortilla blanks are baked in gas-fired ovens at 315 °C for 30–32 s. Tortillas discharged at 79 °C are cooled to 29–32 °C in an atmospheric multilayered conveyer.

Corn chips are prepared from a mixture of white and yellow corn, varying in ratio depending upon the color desired in the final product. The alkalized dough is cut into pieces of desired length and shape, and transferred directly into cooking oil held at about 190 °C. After the moisture content has been reduced to a few percent, the chips are salted, cooled, and packaged.

Cereal grains other than corn can also be formed into chips and fried; the alkalizing process is omitted in these cases. Shredded wheat, formed into multiple layers, salted and baked, forms the basis of Triscuit.

The high-fiber, high water-holding capacity (WHC) and low calorific value of corn bran makes it suitable for low-calorie and variety-baked snack foods, e.g., cookies, crackers, etc.

Fabricated potato snacks can be classified into four groups: (1) dry collet processes; (2) single-screw extrusion of dry potatoes; (3) forming and frying of moist dough; and (4) forming of a high-solids dough into a thin sheet which is then cut into pieces and fried. The latter process is used in the manufacture of simulated potato chips.

Simulated potato chips may be formed by methods resembling those used in the cookie and cracker industry. It is necessary to formulate the dough with physical properties that make it suitable for rolling out into a thin sheet. This is achieved by the inclusion of substantial amounts of cereal flour. Another approach is to use ungelatinized starch in the dough, and then steam-treat the pieces after all the forming operations have been completed. After sheeting, the dough is cut into pieces of the desired shape and size, and then fried or baked. Simulated potato chips can also be prepared from reconstituted dehydrated potatoes.

Low-Pressure Extruded Fried Indian Snacks

A lot of legumes are manufactured into flours and are used as the basis for formulations that are extruded in low-pressure systems. One prominent market is in the Indian snack food area that produces a large variety of snack food-type noodles (sev, omumpodi). The number of shapes, sizes, and ingredients used is diverse, but basic principles involved in producing such products are quite similar. In extruded gram flour-based snacks, flour is produced from chickpea, which is one of the main legumes grown in India. The composition of gram

Table 2 Parameters for the major steps in the manufacture of an alkaline-processed corn snack

Operation	*Tortilla chip*	*Corn chip*
Lime content	1% of corn weight	1% of corn weight
Cooking	4–10 min > 93 °C. Moisture: 29%	14–45 min > 93 °C. Moisture 42–44%
Quenching	Quickly, < 66 °C	Quickly, < 66 °C
Steeping	8–24 h, 46–60 °C. Moisture > 45%	8–24 h, 46–60 °C. Moisture 50–52%
Washing	Wash temperature < 27 °C	Wash thoroughly and drain: < 27 °C
Grinding	Stone, masa at < 38 °C. Water added, moisture 52–54%	Stone, masa at 46–54 °C. Moisture 50–52%
Masa sheeting and forming	Promptly, 21–27 °C	Masa is extruded or sheeted and cut directly into frying oil
Baking	12–18 s, 399–454 °C. Moisture 40%	
Cooling and equilibrating	Cool to 60 °C within 15–30 s. Moisture 38% after equilibration	
Frying	2 min at 179 °C	1.5 min at 210 °C
Chip moisture	1.0–1.2%	1.0–1.5%
Oil content	25%	35%

From Gomez MH, Rooney LW and Waniska RD (1987) Dry corn masa flours for tortilla and snack food production. *Cereal Foods of the World* 35: 372.

flour is approximately 11.2% moisture, 22.5% protein, 5.2% fat, and 58.9% carbohydrate. Other flours that are commonly used as base ingredients for this type of extrusion include soybean and pea flour.

Gram flour is mixed with other ingredients to obtain a dough that is suitable for extruding and has an acceptable texture and taste. Other ingredients that are commonly used in the formulations for Indian-style extruded products include rice flour, urad dhal powder, ghee, sodium bicarbonate, chilli, cumin, paprika, coriander, turmeric, asafetida, and pepper. Gram flour is typically mixed with rice flour, other spices, and water to obtain dough that is extremely sticky. The moisture content of this dough varies depending on what the manufacturer requires but typical moisture levels vary from 40 to 60%.

The type of die that is used for extruding will determine the style of product manufactured. Common shapes include a round product with diameter varying from 1.2 to 6 mm. Other shapes include a star or a ribbon. The diameter of the die and the number of holes on the die plates will produce different production volumes.

The noodles are fried using a single-zone, indirectly fired frier. The temperature used is approximately 180 °C and the frying time varies from 30 s to 5 min, depending on the moisture content of the dough, diameter, and shape of the extruder die.

Corn Snacks

Corn-based snacks fall into three categories: extruded from corn meal, baked and fried from masa, and corn chips or simulated potato chips.

Extruded snacks are made of corn meal with, occasionally, a small amount of another grain, such as rice or wheat, added. The moistened meal is passed through extrusion cooking equipment, giving a collet, which is an expanded matrix of gelatinized starch. The dimensions are determined by the kind of die used, the cut-off knife speed, and other mechanical factors. The collet has a moisture content of 8–10% and is quite tough and chewy, an aspect which is remedied by baking or deep-frying to obtain a final moisture level of around 2%. The dried collets are flavored with cheese (applied in a powdered form), chilli, barbecue sauce, sour cream and onion, or a number of other flavors. Other extruded snacks consist of hollow extrudates which are modified to make rectangles, ovals, or ornamental shapes, and filled with all kinds of savory or sweet fillings. These snacks are quite popular, and the most popular flavors for extruded snacks are chocolate, butterscotch, and peanut butter. Delicate fruit creams, such as wild strawberry, lemon-lime, bitter orange, sour cherry, kiwi, peach, and cola, are also being used to fill extrudates.

Baked and fried snacks make up the largest share of the corn-based snack market. They are produced using centuries-old processes that have been modified by the adoption of modern technology. Cleaned corn is placed in steam kettles and covered with a dilute solution of lime ($Ca(OH)_2$). The mass is brought to near-boiling temperature, and held for a period of time depending upon the nature of the corn used. It is then cooled and allowed to steep for 8–16 h. During steeping, water penetrates the kernel and gives the desired degree of friability to the endosperm. The cooling and steeping also loosen the outer hull and germ, which are removed from the corn by washing. The washed kernels are ground to produce masa, a cohesive, plastic corn meal with a moisture content of approximately 50–60%. Masa is the raw material for corn and tortilla chips. To make corn chips, masa is extruded through a slot to give a ribbon of dough. Cut off by a knife, the raw chips fall into hot fat for deep-frying. They are transported through the frying tank by a traveling belt in 2–5 min. During this time, the starch is gelatinized and the moisture level is reduced to approximately 2% to give the crisp bite of the finished chip. After the traveling belt brings the chips out of the bath, they are given a light dusting of salt and flavoring poser, if required, and then packed. Whole kernel corn can be heat-expanded by selecting a corn variety, drying, soaking, separating the corn kernels from water, heating the soaked kernels in edible oil to expand, and cooking the interior starch. The fried kernels are separated from oil and seasoned.

Tortilla chips are produced more or less as in a traditional process. The masa is rolled out into a thin sheet, cut into the desired shape, and transported into an open-flame oven. The heat sets the structure, drives out most of the moisture, and changes the flavor slightly by creating small amounts of Maillard reaction products. The chips are held in a tempering chamber to equilibrate moisture and stresses throughout the chip. Finally, they are deep-fried to create a crisp texture, salted or seasoned, and packaged. Use of dry corn masa in the preparation of corn tortillas and other corn-based snacks is advantageous as it eliminates the capital cost of cooking, washing, and grinding equipment, avoids the problems associated with high solids, high biological oxygen demand waste water; minimizes variability, and improves consistency and quality as well as the partial size distribution requirement of the end products. (*See* **Tortillas**.)

Legume-Based Indian Snacks

Fried Green Peas

Fried peas can be eaten as a fried snack or consumed as a part of a mix of snack food products.

After the variety of pea has been chosen it is necessary to soak the peas in sodium bicarbonate solution containing green food coloring (within Asia). The purpose of adding green food colouring is to enhance the green pigmentation of the surface of the peas.

The peas are obtained in the dehydrated form and rehydrated in water at room temperature for approximately 8 h. If the temperature is too high, it will cause the starch within the peas to gelatinize and the protein to denature. The moisture content of the peas increases to approximately 48–57%. After soaking, the peas swell to approximately twice their original size. The soak time is critical to minimize the level of fines produced as peas soaked for longer than 9 h tend to produce a higher level of fines as their skins get removed more easily during the cooking process. If the peas are soaked for excessively longer periods of time, they may ferment and cannot be utilized by the manufacturer.

The addition of sodium bicarbonate produces a multitude of effects. Initially it will increase the pH of the soak water and brightness will occur to a greater extent with the presence of green food coloring in the soak water. The increase in pH of the soak water insures that the water absorbed into the peas has a higher pH than the one it contained. It acts as a form of chemical leavening. The reaction involves the acid (contained within the pea) and soda (sodium bicarbonate). One of the main purposes of using this chemical is to produce carbon dioxide when the peas are fried. This gives fried peas a lighter texture and a better mouth feel.

After soaking, the peas are fried to reduce the moisture content below 2.5%. By having the flexibility to vary the temperature of the frying oil, the product quality of the peas can be improved. If the frying oil temperature is not controlled then the moisture content in the final product may be too high or too low, the texture may be too hard, the surface of the peas may blister, there may be an overexpansion of the peas giving excess fines, and product capacity may be reduced.

Fried Chickpeas

Due to their high protein and low fat content, fried chickpeas are a healthy alternative to other snack foods. The dried chickpea contains 7–10% water, 12–31% protein, 4–10% fat, 58–68% carbohydrates, 3–5% fiber, and 2–5% ash.

Chickpeas are soaked for approximately 10 h at room temperature, increasing the moisture content to approximately 53%. The chickpeas are removed from water, rinsed, and drained to insure excess surface moisture is removed.

The chickpeas have an anatomy that resembles the green pea and therefore the problems experienced during frying are very similar. It is necessary to insure the frier has the flexibility to vary the initial and final frying temperature (as was the case with green peas). If a single zone temperature is used, either the oil has to be set at a low temperature, or if the temperature is too high this will cause problems with rupturing of the skin and increase in level of fines. If the temperature is too low then the product output will be reduced and oil absorption will increase. The slow increase in temperature during frying will increase the output obtained and decrease oil absorption.

Fried Lentils

The chemical composition of dried lentils is approximately 23% protein, 1% fat, 1% fiber, 61% carbohydrates, and 14% water.

The lentils are soaked for 3 h in water at room temperature to a final moisture content of 52%, rinsed and drained to remove excess water.

This legume performs differently when fried, as compared to green peas or chick peas. A single temperature zone can be used because the product is not as susceptible to thermal shock when it initially enters the frier. A temperature of 180 °C is used to obtain a rapid decrease in moisture content of the legume. The time required to fry this legume is very short because of the large surface area of the product and the rapid heat transfer into the product.

Meat-Based Snacks

There are a few snacks that are composed primarily of animal-derived raw materials. Expanded pork rinds, also called broken skins or skeen, have been popular as a between-meal snack in the southern USA for many years. These are pieces of hog skin which have been processed so that they puff to many times their original volume. The flavor is bland and reflects the oil in which the skin has been processed. The texture is very crisp and friable; they are not as hygroscopic as many other puffed snacks.

The raw materials are green belly skins, green fat hog skins and green ham skins from any type of hog; belly skins give the best finished product. The skins are first dipped into an air-agitated brine solution held at 100 °C for 30 s. The brine may contain 12 kg dextrose, 11 kg sucrose, and 68 kg salt in 909 l in

water. After dipping, the skins are drained and cooled at room temperature. Then the skin is diced into 127- or 254-mm squares, ready for further processing.

Chimni, a processed seafood snack popular in Japan, is gaining popularity as a main dish rather than a snack consumed with drinks, but it involves high labor.

Jerky

The name 'jerky' is derived from a dried-meat product which is popular in the USA. Many countries have their own form of dried-meat products. The product name varies, as does the form in which it is eaten. The Americans call it 'jerky' and eat it as a leathery whole tissue, or comminuted into strips. The American Indians call it 'pemmican;' they grind the diced meat and mix it with fat before eating it. In South Africa it is called 'biltong.' The French version is called 'grisons;' before being eaten, the dried meat is smoothed with hot cheese.

Meat jerky was originally prepared by the American Indians, who salted strips of muscle tissue from deer, buffaloes, and other game animals and cured the strips in sun or over smoky fires for long periods of time. Commercially, beef jerky is made by marinating strips of beef, drying the marinated meat, and finally cutting the dried strips to the desired dimensions. Beef jerky can also be prepared by subjecting a mixture of ground and chunk beef to a saline treatment for 12 h or more and then to freezing temperatures for 1–3 weeks. The meat preparation is then sliced into strips and dried.

Other Snack Foods

The future of traditional snack foods depends upon the development of new forms and flavors. Pretzels and salt sticks, crackers, and nut mixtures have been modified, and the market of these products has been increased.

Soft pretzels available in fresh, frozen, and microwavable forms have resulted in growth of the pretzel market. In most countries, the market of pretzels and salt sticks has increased with the introduction of health-oriented products that are salt-free, cholesterol-free, and possess increased fiber content. Other new pretzels and salt stick products are produced in various shapes and sizes. The new products are also available in such varieties as savory-flavored, sweet-flavored, chocolate-coated, brightly colored, and peanut-butter-filled.

Along with the salted peanuts and nut mixtures that continue to be popular all over the world, the current market for nut-based snacks is influenced by: (1) an increased demand for flavored nuts; (2) the growing popularity of high-quality exotic nut and dried-fruit mixtures; and (3) the increasing success of premium nuts such as cashew, pistachio, and macadamia nuts. The development of flavored nuts and nut mixtures has resulted in a virtually limitless list of new products, including products such as hickory-smoked and honey-roasted nuts, nut mixtures containing rice, lentils, coconut, and vegetables, and nuts of all types flavored with garlic, soy sauce, herbs, ham, barbecue, and many newly developed flavors.

Khao kriap wane (KKW) is a popular Thai glutinous rice-based snack food made from pounded, gelatinized glutinous rice which is sun-dried and puffed by direct flame roasting. Puffed KKW does not stay crunchy for more than 3 h after puffing. It can be stored for less than 2 months in low-density polyethylene (LDPE) or for greater than 6 months in aluminum/polyethylene laminate or polypropylene and retains crunchiness.

Novelose, a speciality starch, consists of at least 30% dietary fiber and functions well in direct expanded extruded products. It also has a synergistic effect with traditional cellulosic fiber and counteracts negative processing aspects. (*See* **Cashew Nuts and Cashew Apples**.) Specialty and modified starches can provide a number of functional benefits to snack foods like replacement of fats and oils, satisfying consumer demand for healthier snacks for tasty and appealing reduced-fat products.

A fruit snack bar, prepared using apricot pulp and soyabean slurry, contains 15.3% moisture, 7.8% protein and 16.5 mg ascorbic acid $100\,g^{-1}$, with good flavor. A snack product comprised of dried agglomerated, sugar-soaked fruit or vegetables like apples, carrots, and beets, followed by drying, has a low water activity, bulk density, and crisp tender eating quality.

A bagged expanded cooked snack comprises a number of individual morsels with several added vitamins, such as thiamin, riboflavin, niacin, ascorbic acid, vitamin D, and at least one added mineral other than sodium. It may be a savoury product with an oil or fat content of 5–40% by weight.

Rice flour can improve the texture and handling of multigrain products or substitute completely for another type of grain. Blending long-grain rice flour increases the crispiness in fried or baked snack chips based on wheat or corn. Waxy rice flour can reduce chip hardness and at the same time can provide a melt-in-the-mouth texture which is usually achieved with extra fat. Adding 5–10% waxy rice flour to a baked or fried chip product diminishes chip breakage during packaging.

Shelf-stable grain-based Indian snacks include fried products (paras, fried products made from bengal

gram meal, fried products made from rice and legumes, fried dhals); moist fried products, popped and puffed cereals, expanded cereals, beaten (flaked) rice, and extruded foods (extruded pellets, ready-to-eat expanded products, corn and tortilla chips). Kokoro, an indigenous corn-based snack food product prepared using deep frying of a corn, onion, and shrimp blend, is popular in Nigeria.

Quality Control for Snack Foods

The extent of quality control analysis in the snack food industry is quite variable, ranging from completely subjective testing procedures to series of elaborate instrumental analyses performed on a continuous basis. The maintenance of uniform quality in snack foods is essential for the very existence of the business. Quality control expresses the objectives of a group of people who do the sampling, testing, and other activities related to the evaluation of product conformance to predetermined standards. It is the expended responsibility of the group to correct problems which could result in a defective product. (*See* **Quality Assurance and Quality Control**.)

Quality control aims at maintenance of quality at levels and tolerances acceptable to the buyer, while minimizing costs to the vendor. It can also be defined as the application of sensory, physical, and chemical tests in industrial production to prevent undue variation in quality attributes, such as color, viscosity, and texture. In addition to satisfying the consumer, the product must meet the requirements of government regulatory agencies. A snack food which does not meet the minimum acceptability requirements of all customers may result in only a few complaints, but a product which does not meet all of the legal requirements may be confiscated and the manufacturer subjected to punitive fines and other penalties. A product may look good, taste good, have good texture, and perform well in its intended application, and still be unfit for distribution. To be suitable for sale it must also be wholesome, conform to all applicable labeling and packaging requirements, and be prepared and stored under conditions tending to prevent contamination by noxious or esthetically undesirable materials.

The quality of the finished product is inseparable from the quality of the components. Seldom can processing operations compensate for raw material inadequacies. Unless the raw materials are of good quality, it is very difficult to obtain a good-quality finished product. Standards must be developed and strictly adhered to while processing the various ingredients, during the production and processing steps, as well as during handling, packaging, and storage of the finished snacks.

During pretzel production, a number of problems, particularly in the cooking process, may arise. These include brittleness, dark or light color, wrinkling, loss of glaze, breakage, browned interior, and lack of crispness.

In the production of corn-based, baked, and fried products, the consistency of the masa is very important. If excess starch granules are gelatinized during cooking and steeping, the masa is too sticky to process smoothly. On the other hand, inadequate cooking may give a dry, crumbly mass which does not work in the equipment. Potato chips undergo a visual inspection at the end of the process, and excessively dark ones or those containing discolored portions or possessing some other defect are manually removed by picking. In a modern industry, this may be performed by computerized optical scanners. The sensitivity of the computer scanner can be adjusted to pick up all less-than-perfect chips, or selectively to remove only the least desirable. Blistering of chips on frying is quite a problem. Such chips break in conveying, filling, or handling the filled bags. Another factor affecting the quality of chips is the shift to thicker chips: it results in significant reduction in oil absorption owing to a decreased surface area relative to the volume. Ridged chips solve the blistering problem.

In general, snack foods are believed to be 'empty calories,' however, lately manufacturers have started supplementation with vitamins, minerals, or proteins. Vitamins can affect flavor and some can also cause off-flavors. While supplementing with nutrients, it is necessary to establish a quality control program to insure the claimed amounts in the product. Most snacks are poor sources of protein, and the protein, which is present, is often of poor nutritional quality. Protein supplementation is possible by increasing the protein efficiency ratio of the protein already present or increasing the protein content through addition of casein, soya protein isolates, or egg white. Certain combinations of cereals and legumes are very desirable from a nutritional point of view. The high snackers have low energy and low macronutrients from the meals; however, snacks do make an important contribution to energy and nutrients. A high snacking frequency and choosing from a wide variety of snack foods can result in a balanced intake of nutrients.

See also: **Almonds**; **Biscuits, Cookies, and Crackers**: Nature of the Products; Methods of Manufacture; Wafers; **Cashew Nuts and Cashew Apples**; **Peanuts**; **Potatoes and Related Crops**: Processing Potato Tubers; **Quality Assurance and Quality Control**; **Tortillas**

Further Reading

Anonymous (1996) Healthy starch innovations for snacks and cereals. *Food Technology Europe* 3: 46.

Arya SS (1991) Grain based snack and convenience foods. *Indian Miller* 21: 6.

Booth RG (ed.) (1990) *Snack Food*. New York: Van Nostrand Reinhold.

Burge RM and Duensing WJ (1989) Processing and dietary fibre ingredient application of corn bran. *Cereal Foods World* 34: 535.

Chauhan SK, Joshi VK and Lal BB (1993) Apricot-soy fruit bar: a new protein enriched product. *Journal of Food Science and Technology* 30: 457.

Crocombe BI (1988) The development and production of jerky at Mirinz's meat products line. *Bulletin of the Meat Industry Research Institute of New Zealand* 855: 126.

Frame ND (ed.) *The Technology of Extrusion Cooking*. New York: Blackie Academic and Professional.

Friesen SA (1995) *Heat Expanded Whole Kernel Snack Food*. United States patent.

Gomez MH, Rooney LW and Waniska RD (1987) Dry corn masa flours for tortilla and snack food production. *Cereal Foods World* 35: 372.

Gould WA (1985) Changes and trends in snack food industry. *Cereal Foods World* 30: 219.

Howlett RJ (1993) *Bagged Snacks*. UK patent application.

Huang DP (1995) New perspectives on starch and starch derivatives for snack application. *Cereal Foods World* 40: 528.

Jomduang S, Mohamad S, Muhammand K and Abu-Bakar A (1995) Extending shelf life of 'Khao Kriap Wane' – a traditional Thai glutinous rice-based snack. *Journal of Science and Food Agriculture* 67: 29.

Kano I (1993) A growing chimni market in Japan. Infofish International 4: 18.

Lasekan OO, Lasekan W, Idown MA and Oja OA (1996) Effect of extrusion conditions on the nutritional value, storage stability and sensory characteristics of a maize-based snack food. *Journal of Cereal Science* 24: 79.

Madonna MD (1983) Pretzels as low calorie snacks. *Cereal Foods World* 28: 297.

Matz SA (1976) *Snack Food Technology*. Westport, Ct: AVI Publishing.

Midson K (1999) Indian ethnic snacks. *Processed Food Industry* 3: 38.

Robert BA and Burkes AL (1989) *Crisp Fruit or Vegetable Snack Products and Process*. United States patent.

Sheng DY (1995) Rice-based ingredients in cereals and snacks. *Cereal Foods of the World* 40: 538.

Slige MR (1983) Fruit and nut ingredients. *Cereal Foods of the World* 25: 286.

Stauffer CE (1983) Corn-based snacks. *Cereal Foods World* 28: 301.

Stauffer CE (1983) Corn masa. *Baking and Snack*. October: 35.

Tettweiler P (1991) Snack foods worldwide. *Food Technology* 45: 58.

Dietary Importance

A S Bawa and J S Sidhu, Guru Nanak Dev University, Amritsar, India

Background

This article deals with the dietary importance of processed snack foods with a passing reference to fruits and fruit bars. However, these as well as fruit juices may be included under the term 'snack foods.' In general, snack foods are considered as 'junk foods' or 'empty calories.' Our diet is generally based upon three principal meals a day, and the role of snack foods is to offer a light, convenient, and enjoyable food option, when, for reasons of hunger or sociability, a food selection is required at a time between that allotted for principal meals. Thus, snack foods are not designed to be alternatives for three main meals a day. They are convenient to eat, involving no preparation time or effort by the consumer. Snacks are eaten not only to satisfy hunger or supply the nutrients to our body, but also for social reasons. On many occasions, the inclusion of snack foods in our diet is not justified in terms of nutritional demands. The nutritional value of different snack foods, extent and frequency of consumption by different people, and the need for fortification have been discussed.

Snack items and processed foods are presumed to have a high sodium content, and this should be avoided. Teenage boys are most likely to have a low intake level of calcium, iron, magnesium, and vitamin B_6, and teenage girls consume these nutrients at still lower levels. The food items most frequently consumed by both sexes between meals are cakes, cookies, pies, candy, and desserts. The most popular foods are salted snack foods, such as chips (crisps), pretzels, and popcorn. Males incorporate more breads, wafers, and other bread products into their snacks than females, whereas females eat more snacks containing condiments, fat spreads, and dips than do males. (*See* **Calcium**: Physiology; **Iron**: Physiology; **Magnesium**; **Sodium**: Physiology; **Vitamin B_6**: Properties and Determination.)

The real nutritional concerns about snack foods center on the sodium, energy, vitamin, mineral, and dietary fiber contents.

Nutritional Value of Different Types of Snack Food

Potato Chips (Crisps)

Nutritionally, 28 g (1 oz.) of potato chips provide 647 kJ (154 kcal), contributed by 2 g of protein, 14 g

of carbohydrates, and 10 g of fat. A portion of potato chips (25 g) contains 8% of the recommended daily allowance of vitamin C, 3% of thiamin, 11% of nicotinic acid, 8% of iron, 11% of vitamin B_6, 4% of phosphorus and 4% of magnesium. Potatoes are also an important source of dietary fiber. Refer to individual nutrients.

Cookies and Crackers

The principal ingredient in cookies and crackers is unbleached, soft wheat flour, and most of the nutrients in these products are derived from the flour. It can range from more than 90% for saltines and similar products to less than 25% in products such as enrobed wafers. The soft wheat flour used in cookies and crackers contains 8–10% protein, with the higher-protein flours generally being used for crackers and the lower-protein flours for more tender sweet goods. With certain exceptions, the quality does not deteriorate on baking. Most cookies and crackers are made from enriched flour. The main component present in flour is carbohydrate, which is a good source of energy. (*See* **Flour**: Dietary Importance.)

Other ingredients added to cookies and crackers influence the nutrient content of products either by their own nutritional composition or by diluting the nutrients contributed by enriched flour: for example, sugar contributes energy and carbohydrates; shortening contributes fat, essential fatty acids, and energy; salt contributes sodium; leavening agents contribute sodium, potassium, calcium, and phosphorus; peanut contributes fat and nicotinic acid; fruits and vegetables contribute vitamins A and C, fiber, and minerals; chocolate contributes iron and fiber; and other grains contribute minerals, fiber, and other nutrients similar to wheat flour. The potassium content of cookies and crackers is 25–200 mg and 30–80 mg per 28 g (1 oz.), respectively. During baking and storage, the main nutrients lost are thiamin and vitamin C. Lysine reacts with reducing sugars and becomes unavailable, especially at a high pH. Plain and sandwich cookies have similar nutrients. However, nut-containing cookies provide comparatively more protein, vitamins (such as nicotinic acid), and minerals. (*Refer to* individual nutrients.)

Pretzels

The average energy content of a 28-g of serving pretzels is 462 kJ. One pretzel rod contains 231 kJ; one Dutch-style pretzel contains 460 kJ, and a handful of twisted, three-ring pretzels contain approximately 419 kJ. Pretzels are low in energy and fat, and have no sugar, but contain a large amount of sodium. The average amount is 450–700 mg of sodium per 28-g serving of pretzels. However, low-salt pretzels, containing 300 mg per 28-g serving, and unsalted pretzels are also available for sodium-conscious consumers.

The typical ingredients of pretzels are flour, water, yeast, soda, and salt. The flour is from soft wheat with a protein content of 8.50–9.50% and an ash content of 0.40–0.50%. A 28-g serving of pretzels offers approximately 3 g of protein, 23 g of carbohydrates, and 1 g of fat. Pretzels also provide several vitamins and minerals, including thiamin, riboflavin, nicotinic acid, iron, magnesium, and copper. Although the quantities of these nutrients are moderate, the nutritional value of pretzels is significant, since they contain no chemical additives or preservatives. However, some pretzels also offer calcium, phosphorus, and other nutrients in significant amounts. (*Refer to* individual nutrients.)

Corn Snacks

Corn snacks are accused of being junk food and without nutrient content. Flour tortillas are low in moisture and high in fat, protein, and carbohydrates, compared to corn tortillas. The energy content is, on average, 25% higher than that of corn tortillas. The phosphorus and calcium contents in corn tortillas are substantive. The iron content in corn tortillas is similar to that of wheat flour tortillas. Both the flour and corn tortillas are significant sources of zinc; corn tortillas are also a significant dietary source of copper and magnesium. Tortillas contain modest amounts of sodium and potassium, but because of the significant losses during processing, they are an insignificant source of thiamin and riboflavin. All tortilla products are substantial sources of nicotinic acid. Tortillas do not contain measurable quantities of vitamin C or β-carotene. Refer to individual nutrients.

A fruit bar containing apricot pulp and soy slurry in the ratio 70:30 has been found to be most acceptable in terms of sensory properties. This bar contained 15.3% moisture, 7.8% protein, and 16.5 mg of ascorbic acid per 100 g, and had a good flavor (not beany) and texture.

In one study, consumers in Newcastle perceived cereal bars to be healthiest. However, an analysis of seven popular cereal bars revealed that they are only marginally healthier than traditional cereal snacks. The varieties containing chocolate chips were preferred but were not necessarily the healthiest.

An assessment of the perception of fruits in relation to manufactured snack foods has revealed that fresh fruit and a canned snack fruit product were both perceived as healthy and refreshing by working women in the UK, manufactured snack foods were considered more convenient and suitable for comfort eating, and bananas and apples were

perceived as more convenient than orange and kiwi fruit.

The Nigerian snack foods, robo, kulikuli, and kokoro, prepared from watermelon seeds, groundnut, or maize kernels, have 23.4–42.0% fat and only 1.5–7.3% moisture, and a maximum shelf-life of about 150 days at 30–31 °C.

An extruded snack prepared from a blend of tapioca, wheat flour, maize grits, spirulina, etc. is quite nutritious and contains: energy 411 kcal, protein 8.7 g, fat 7.6 g, carbohydrates 77.2 g, Ca 68.0 mg, Fe 2.8 mg, vitamin A less than 3 IU, thaimin 0.1 mg, riboflavin 0.2 mg and vitamin C less than 0.35 mg per 100 g. It is low in fat and relatively high in Ca and Fe, compared to other similar snacks.

Overall, corn snacks can contribute positively to the diet by providing certain nutrients and minerals necessary for nutritional balance and they have a good flavor and a pleasing texture.

In potato chips and nuts, the salt is on the surface of the foods, so the immediate taste perception is of saltiness – much more so than if the salt were uniformly distributed, as in most other foods. Many snack foods contain moderate to high amounts of salt on a percentage basis. The average serving of snack foods is much less than 100 g, ranging up to 40 g and averaging about 35 g; in comparison, many sodium-containing staple foods are eaten in quantities greater than 100 g. It is more sensible to examine the sodium content of snack foods on a per-serving basis (Table 1) because the sodium content of snack foods is much lower than that of certain commonly consumed foods, such as bread and butter, sandwiches, and other foods at the bottom of the table. The sodium content of a packet of chips, for example, is less than that of a plain sandwich, and over six packets of salted chips would have to be consumed to obtain the same amount of sodium contained in a single hamburger.

The second aspect of concern is the energy content of snack foods. It is generally believed that snack foods are both rich in energy and fattening, since they are cooked in, or blended with, oil. A comparison of energy content of foods on a per-serving basis (Table 2) shows that the consumption of lightly cooked vegetables instead of snacks can reduce the energy intake and that the energy content of snacks is not as high as is popularly perceived, since hot dogs and hamburgers have much higher energy contents. A packet of chips does not contain much more energy than a cup of milk or a plain yogurt, and somewhat less than a bread-and-butter sandwich. It is much more difficult to control the energy and fat content of snack foods than it is to control sodium content. The technology to produce fat-reduced potato chips does not exist, but the level of fat can be controlled by careful regulation of cooking parameters. Oil is essential in cereal snacks to give an acceptable mouthfeel, and with nuts, the oil is already present before cooking. The only way to control the energy consumption via snack foods is to reduce the size of the serving; Table 3 shows gives the effect of serving size on energy content.

The vitamin and mineral contents of snack foods are another area of concern; Table 4 lists the vitamin and mineral contents of some selected snack foods.

Table 1 Sodium content of some foods

Food	*Sodium content (mg per serving)*
Most fresh fruits and vegetables	Negligible
Eggs (one medium)	54
Rump steak (130 g)	121
Fried chicken leg	122
Whole milk (1 cup, 237 ml)	128
Salted peanuts (25 g)	197
Toasted CCS (25-g packet)	220
Salted chips (25-g packet)	234
Flavored chip crisp sticks (25-g packet)	280
Flavored chips (25-g packet)	306
Bread-and-butter sandwich	367
Cheese Twisties (25-g packet)	440
Scrambled egg	534
Fritz or Devon (two slices)	650
Hot dog	1102
Hamburger	1591

Data from Delroy B (1985) The role of snack foods. *Food Technology in Australia* 37(4): 154–158.

Table 2 Energy content of some foods

Food	*Energy content (kJ) per serving*
Lettuce	17
Peas	163
Boiled egg (one medium)	301
Baked fish (one fillet)	326
Salted peanuts (25 g)	584
Whole milk (1 cup, 237 ml)	644
Plain yogurt (1 cup, 237 ml)	690
Potato chips (25-g packet)	739
Twisties (25-g packet)	810
Chip crisp sticks (25-g packet)	814
Bread-and-butter sandwich	865
Hamburger patty	932
Flavored and fruit yogurt (1 cup, 237 ml)	982
Fish, buttered and fried (one fillet)	1041
Hot dog	1342
Grilled steak (130 g)	1822

[a]4.2 kJ is equivalent to 1 kcal. Data from Delroy B (1985) The role of snack foods. *Food Technology in Australia* 37(4): 154–158.

The potassium content of chips is higher than sodium in spite of the added salt, and this is significant in view of the criticism that high sodium levels in processed foods are associated with depleted potassium levels. However, since snack foods are designed to be occasional between meal snacks, their contribution to the diet may be of little significance to the average person. Cereal snacks can be fortified with added vitamins and minerals to any desired level, but it is not common practice, since snack foods are not considered to be suppliers of significant quantities of minerals and vitamins. (*See* **Food Fortification**.)

The importance of dietary fiber in snack foods is still not very clear, partly because of problems with analytical methods; Table 5 compares the dietary fiber content of snack foods with that of a bread sandwich. The figures indicate that the four snack foods contain more dietary fiber than a white-bread sandwich, and chip crisp sticks more than a brown-bread sandwich. Since bread is one of our best sources of dietary fiber, and the cereals and root vegetables from which snacks are developed are sources of the highest-quality fiber, snack foods cannot be rejected on the basis of their fiber content. Snack foods are neither an important source of micronutrients nor simply 'empty calories.' Snack foods prepared from whole cereal grains can be a good source of dietary fiber, and the addition of a few per cent of bran in place of the purified cereal fractions improves the fiber content of the finished products. Dairy and meat products have negligible fiber contents, and potatoes are rather low in this component. (*See* **Dietary Fiber**: Physiological Effects; Effects of Fiber on Absorption.)

Table 3 Effect of serving on energy content

Food	*Energy (kJ) per serving*	*Fat (g)*	*Sodium (mg)*
Small apple	222.0	0.3	Negligible
Twisties (15-g packet)	460.0	5.9	250.0
Salted chips (20-g packet)	565.0	8.8	179.0
Milk (1 cup, 237 ml)	644.0	8.7	128.0
Bread-and-butter sandwich	865.0	8.6	267.0

Data from Delroy B (1985) The role of snack foods. *Food Technology in Australia* 37(4): 154–158.

Levels of Snack Foods Eaten

The results of various previous studies (Table 6) reveal that children and teenagers have at least one snack during the day. Adults tend to snack less frequently, and the percentage of adults who 'snack' decreases with age. Snacks provide at least 20% of energy for those who eat at least one snack a day and account for 12% of the average protein intake, 16% of the fat intake, and 25% of the daily carbohydrate intake. 'Snackers' obtain, on average, 14% of their daily iron intake, 18% of their phosphorus intake, 21% of their magnesium and calcium intakes, 13–14% of their vitamin A, thiamin, preformed nicotinic acid, vitamin B_6, and vitamin B_{12} intakes, and 16–17% of their vitamin C and riboflavin intakes. As the number of snacks eaten increases, the amounts of nearly all nutrients and energy also increase. An increased level of snacking makes a significant contribution to children's intake levels of magnesium and zinc, and to adolescents' intake of vitamin B_6, calcium, iron, magnesium, and zinc. On days when no snacks are eaten, the intake of fat, sugar, and sodium is decreased, and more food is eaten during the main meals. Snacking has little impact on the nutritional quality of the diet of elderly individuals. Crisps are the most popular snack among 7–8-year-old children in Scotland and are eaten by 92% of the children. Snacks contribute 26% of energy intake and a large proportion of macronutrient intakes. Low and high snacking habits have no significant effect on the anthropometry of children. (*Refer to* individual nutrients.)

Children aged 9–13 years eat dessert items between meals followed, in descending order of frequency, by fruit and milk. The snacks most frequently eaten by 3–5-year-olds, in descending order of frequency, are bakery products, milk, soft drinks, fruit, mild desserts, candy, and bread. For 6–11-year-olds, the snacks in order of importance are bakery products, soft drinks, milk, milk desserts, candy, fruit, salty snacks,

Table 4 Vitamins and minerals in snack foods

Snack food	*Vitamin B_1 (μg)*	*Nicotinic acid (mg)*	*Vitamin C (mg)*	*Calcium (mg)*	*Iron (mg)*	*Potassium (mg)*
Potato chips (25-g packet)	37.0 (3%)	1.2 (11%)	2.4 (8%)	8.5 (1%)	0.8 (8%)	445.0
Twisties (25-g packet)	7.0 (1%)	0.2 (2%)	0.0	44.0 (6%)	0.1 (1%)	70.0
Chip crisp sticks (25-g packet)	78.0 (7%)	0.5 (4%)	0.0	46.0 (7%)	0.9 (9%)	64.0
Peanuts (25 g)				8.0 (1%)	0.3 (3%)	135.0

Figures in parentheses are the percentage of the recommended dietary allowance, as defined in the NH and MRC Food Legislation (1983). Data from Delroy B (1985) The role of snack foods. *Food Technology in Australia* 37(4): 154–158.

Table 5 Dietary fiber in various snack foods

Snack	*Insoluble dietary fiber (gram per serving)*
White-bread sandwich	0.2
Brown-bread sandwich	1.4
Wholemeal-bread sandwich	3.1
Twisties (25-g packet)	0.3
Potato chips (25-g packet)	0.6
Chip crisp sticks (25-g packet)	1.5
Peanuts (25 g)	1.6

Data from Delroy B (1985) The role of snack foods. *Food Technology in Australia* 37(4): 154–158.

Table 6 Consumption of nuts and nut mixtures

Country	*Quantity (kg per capita)*
The Netherlands	1.28
Germany	0.95
UK	0.66
France	0.65
USA	0.60
Belgium	0.56
Switzerland	0.18
Italy	0.10

Data from Tettweiler P (1990) *Snacks – Food, Fun and Fashion. Dragocco Report 3*, pp. 79–104. Holzminden, Germany: DRA-GOCCO.

and bread. Among teenage boys, the most popular snacks are soft drinks, milk, bakery products, bread, mild desserts, salty snacks, meats, and fruit, while teenage girls prefer soft drinks, bakery products, milk desserts, salty snacks, fruit, milk, candy, bread, and meat. Among adolescents, the preference for snack foods is bakery products, milk, fruit and fruit juices, milk desserts, salty snacks, and bread. Table 6 lists the per capita consumption of nuts and nut mixtures. (*See* **Adolescents**; **Children**: Nutritional Requirements.)

Nutritional Improvement of Snack Foods

In order to counter the belief of consumers and nutritionists that snack foods are 'empty calories,' a number of manufacturers of snack foods have started the development of products supplemented with vitamins, minerals, or proteins. Vitamin supplementation is somewhat simpler than mineral or protein supplementation. The cost is reasonable; a small fraction of a per cent per portion of a supplement is restricted to five vitamins, but vitamins can affect the flavor, and some can cause off-flavors. Riboflavin is highly colored, but its yellow hue is compatible with many snacks. Storage deterioration of these labeled substances is compensated for by the addition of extra amounts, so that the consumer should obtain the full claimed quantity. When supplementing with any nutritional factor, the manufacturer must establish a quality-control program that will ensure that every lot contains the claimed amount. Storage loss of minerals is not generally observed, except for iodine.

Most snacks are poor sources of protein, and the protein that is present is often of a poor nutritional quality. Protein supplementation is possible by increasing the protein-efficiency ratio of the protein already present, or increasing the total amount of protein present by adding some purified nitrogenous material, such as casein, isolated soya protein, or egg white. Untreated corn is a staple food in South America, and there is a need to develop more nutritious, corn-based snack foods. These foods could serve as vehicles for nutrients while at the same time being readily accepted by the population. Certain combinations of cereals and legumes are very desirable from a nutritional point of view. Legumes are a better source of lysine and total protein than cereals, and the latter are a better source of sulfur amino acids. Many attempts have been made to raise the protein value of tortillas by combination with legumes. Snack foods are generally preferred by children and teenagers; in these stages of life, the amount and nutritional quality of proteins are important because of their essential function in physical and mental development. Baked snack foods containing high levels of fiber and fat can be produced using specialty (100% fat-free) starches, e.g., Novelose. Fried snack formulations containing carboxymethyl cellulose absorb less oil during frying, which is advantageous for low-fat applications.

The high snackers have low energy and low macronutrients from meals, so snacks make an important contribution to energy and nutrients, even though a high intake of snacks affects the quality of children's diets. A high snacking frequency and choosing from a wider variety of snack foods results in a balanced intake of nutrients.

An extruded snack food (soyabari snack sticks) based on soybeans and maize (corn) rich in crude fiber and protein contents can be produced by adding different ingredients such as pepper, onion, salt, palm oil, plantain, and banana. This product is acceptable to Nigerians. A high-protein, shelf-stable, expanded extruded snack food resembling bread sticks has been prepared using mechanically recovered poultry meat and wheat. (*See* **Protein**: Quality.)

Consumption of olestra, a fat replacer in savory snack foods, has been found to be age- and gender-dependent, with mean chronic and acute olestra intakes for all ages and both genders of 3.1 and 10.2 g per day.

A study evaluating the nutrient density of snacks and meals, and the effect of eating frequency on mean daily nutrient intake in female students (aged 17–26) has shown that as the eating frequency increased, the number of snacks and the number of different snack items in the diet increased, while the number of meals remained constant. Snacks had lower nutrient densities than meals for nonstarch polysaccharide, minerals, and vitamins, except vitamin C.

Mango bars fortified with desiccated coconut powder or soyprotein concentrate are quite acceptable initially, but storage beyond 90 days lowers their rating, compared to plain mango.

A low-calorie, low-cholesterol, shelf-stable expanded snack product can be prepared from raw or mechanically separated, comminuted meat mixed with pregelatinized flour and processed through high-temperature, short-time screw-type cooker extruder. A gun-puffed, extruded snack food, low in fat and high in carbohydrates, prepared by incorporating fruit and vegetable extracts with oat flour, starch, and sugars, is low in fat and high in calories.

Substantial efforts have been made to reduce the 'snacking habit' of various populations, but snack-food consumption continues to increase. Consequently, there is a tendency to increase the production of more nutritious snacks, including those made from corn. Composite flours using corn, chickpea, soya bean meal, and methionine can be formulated, extruded into snack foods, and fried. These products have higher protein-efficiency ratio values than the controls, while the color, flavor, texture, and overall acceptability remain similar. These products are a good alternative to commercially available snack foods.

See also: **Adolescents**; **Calcium**: Physiology; **Children**: Nutritional Requirements; **Dietary Fiber**: Physiological Effects; Effects of Fiber on Absorption; **Flour**: Dietary Importance; **Food Fortification**; **Iron**: Physiology; **Magnesium**; **Protein**: Quality; **Sodium**: Physiology; **Vitamin B_6**: Properties and Determination

Further Reading

Anon (1993) CMC smooths production of extruded snacks, cereals. *Prepared Foods* 162(10): 81.

Anon (1996) Healthy starch innovations for snacks and cereals. *Food Tech Europe* 3(2): 46.

Bednarcyk NE (1987) Nutrient contribution of cookies and crackers. *Cereal Foods World* 38(5): 367.

Chauhan SK, Joshi VK and Lal BB (1993) Apricot–soy fruit bar: a new protein enriched product. *Journal of Food Science and Technology* 30(6): 457–458.

Conninghum FE, Neumann PE and Huber GR (1993) *Extruded Meat-based Snack Food and Method for Preparing Same*. US Patent.

Delroy B (1985) The role of snack foods. *Food Technology in Australia* 37(4): 154–158.

Henshaw FO and Ihedioha CN (1992) Shelf life studies on some Nigerian indigenous snackfoods. *Nahrung* 36(4): 405–407.

Jack FR, O'Neill J, Piacentini MG and Schroeder MJA (1997) Perception of fruit as a snack: a comparison with manufactured snack foods. *Food Quality and Preference* 8(3): 175–182.

Madonna MD (1983) Pretzels as low calorie snacks. *Cereal Foods World* 28(5): 297.

Matz SA (1976) *Nutritional Supplementation. Snack Food Technology.* Westport, CT: AVI.

Mir MA and Nirankar N (1993) Storage changes in fortified mango bars. *Journal of Food Science and Technology* 30(4): 279–282.

Mitchell V and Boustani P (1990) Cereal bars. *Nutrition and Food Science* 127: 20–21.

Morgan KJ (1983) The role of snacking in American diet. *Cereal Foods World* 28(5): 305.

Olusola-Omueti and Morton ID (1996) Development by extrusion of soyabari snack sticks; a nutritionally improved soya–maize product based on the Nigerian snack (Kokoro). *International Journal of Food Science and Nutrition* 47(1): 5.

Ranhotra GS (1985) Nutritional profile of corn and flour tortillas. *Cereal Foods World* 30(10): 703.

Ray EE (1994) *Method of Preparing Snack Food Products*. US Patent.

Ruxton CHS, Kirk TR and Belton NR (1996) The contribution of specific dietary patterns to energy intakes in 7–8 year old Scottish school children. *Journal of Human Nutrition and Dietetics* 9(1): 23–31.

Stevens MPA, van Zuilichem DJ and de Waart J (1991) Tapsea-vegetable: Development of a new snack. *Voedingsmiddelentechnologie* 24(2): 12–14.

Tettweiler P (1990) *Snacks – Food, Fun and Fashion. Dragocco Report 3*, pp. 79–104. Holzminden, Germany: DRA-GOCCO.

Warren AB, Hnat DL and Michnowski J (1983) Protein fortification of cookies, crackers, and snack bars: uses and needs. *Cereal Foods World* 28(8): 441.

Webb DR, Harrison GG, Min JL and Mei HH (1997) Estimated consumption and eating frequency of olestra from savoury snacks using menu sensus data. *Journal of Nutrition* 127: 1547S–1554S.

Whybrow S and Kirk TR (1997) Nutrient intake and snacking frequency in female students. *Journal of Human Nutrition and Dietetics* 10(4): 237–244.

Zammer CM (1995) Gun-puffed vegetable snacks: a new way to eat your veggies. *Food Technology* 49 (10): 64.

SODIUM

Contents
Properties and Determination
Physiology

Properties and Determination

R Moreno Rojas and M A Amaro López, University of Córdoba, Córdoba, Spain

Background

Sodium, prepared for the first time by the British chemist Humphry Davy in 1807, is the eleventh element of the periodic table (atomic number 11). Its symbol is Na, and it is included in group 1 (or IA), thus corresponding to the alkali metals. It is an extremely soft silvery metallic element, so much so that it can be cut with a knife (hardness of 0.4). It is a very reactive element that tends to be used extensively in the chemical industry. It rusts very quickly when exposed to air and reacts violently with water, forming sodium hydroxide and hydrogen, and producing a yellowish flame. It has a fusion point of 98 °C, a boiling point of 883 °C, an atomic weight of 22.98977, and a relative density of $0.97\,g\,cm^{-3}$.

Sodium has an electronic structure of [Ne] $3s^1$, and consequently, it always acts with valence 1. It has only one external electron, in orbital 3s, and so the degree of attraction that it suffers with respect to the nucleus rivals the repulsion of the lower orbitals, making it very easy for this electron to disappear, forming the Na^+ cation, its most common state.

Two per cent of the earth's crust comprises sodium, making it the sixth most abundant element in nature, where it is always found in a combined state. Sodium is currently manufactured by electrolytic procedures on melted sodium chloride. Its combining facility yields numerous compounds; with oxygen it forms Na_2O, Na_2O_2, and even NaO_2, an important whiting and rusting agent. However, the most abundant form and one of the main sources for man is in combination with chlorine in the form of sodium chloride (NaCl), which is usually extracted from the sea, salt lakes, or in mineral form as rock salt. Sodium can also be found in the form of sodium carbonate (Na_2CO_3), sodium bicarbonate ($NaHCO_3$), and sodium sulfate (Na_2SO_4). Also worth mentioning is sodium hydroxide, known as caustic soda, which is used throughout industry, from the manufacture of soap to rayon, paper, rubber, oilskin, in petrol refineries or in the textile industry. Sodium fluoride (NaF) is used as an antiseptic, as a poison for rats and cockroaches, and in the manufacture of ceramics. Nitrate from Chile (sodium nitrate) is used as a fertilizer. Sodium thiosulfate ($Na_2S_2O_3{\cdot}5H_2O$) is used in photography as a fixing agent and is even used in the nuclear power industry as a cooling agent.

All these specific uses of sodium are only a few examples, since this cation is one of the most common companions of different molecules, both in nature and when created by man for different purposes.

In terms of the nutritional characteristics of sodium, we could say that 90% of sodium provided in the diet is obtained from the sodium chloride that we consume. The rest of the sodium that reaches our organism may originate from other sources such as bicarbonate or sodium glutamate. Of the total sodium intake, it has been shown that only 10% originates naturally from foods, around 15% from salt added to cooking or to season food once cooked, and the remaining 75% from the processing and manufacturing stages of foods.

Given the physiological need to supply the human body with sodium (*see* **Sodium: Physiology**) in order to maintain extracellular osmotic pressure, and as a result of the obligatory losses through urine and sweat, we need to consume sodium on a regular basis, just like most animals. In ancient times, this meant a dependence on salt deposits, and a huge value was placed on what is now common salt, so much so that this gave rise to the Latin word *salarium*, corresponding to the payment in the form of salt made to soldiers in ancient Rome, and it was also used as money in Tibet and Ethiopia, as a tax in Asian countries, and as a ritual element in Greek, Roman, Hebrew, and Christian religions.

This physiological need for sodium may explain why our language contains specific receptors corresponding to the taste of salt and our appetite for this taste in food. For this reason, and because of the tradition of preserving certain foods in salt, which is

today probably unnecessary, although culturally and gastronomically unnoticeable, we have reached a situation in which salt is consumed, and with it sodium, in developed societies. Consumption is estimated to be normally more than 6 g per day, the minimum amount required for the control of blood pressure in hypertense individuals. (*See* **Hypertension**: Hypertension and Diet.)

Preparation of Samples for Sodium Analysis

The determination of sodium in foods may be achieved using a variety of sample preparations, ranging from noninvasive methods, such as those used for determining sodium by selective electrodes, to complete incineration of the sample and subsequent solubilization of the ash in acid solutions. Each analytical technique requires specific preparation of the sample, since the instrumental requirements in each are very different.

In any case, and since sodium is usually found in soluble form in foods, the simplest method of preparing samples involves washing and then analyzing the sodium from the solution sample; for this purpose, the food often only has to be crushed, solvent added, homogenized, and filtered or centrifuged. Solvents to be used may be simple water or slightly acidic or alkaline solutions.

Some methods even propose the direct use of food (particularly liquids) in the detection equipment. Although these may yield good analytical results, other food components such as fats, sugars, or proteins can also deteriorate or stain the instruments.

Given the stable character of sodium, the most aggressive techniques usually yield very good results in terms of the recovery of analyses; hence, the use of incineration at around 500 °C overnight and subsequent recovery in weak acid solutions can be used as a standard sample preparation method for a wide range of foods, modifying temperature and time according to the food in order to obtain soft ash, evidence of good interaction.

Other methods for destroying organic matter, such as acid hydrolysis, can also be used without fear of undermining the detection of sodium, and a wide range of acids may be used in variable concentrations, without significantly affecting results in terms of the determination of sodium content.

The use of methods for destroying organic matter by incineration or by acid hydrolysis is particularly recommended, because of the possibility of obtaining solutions in which other cations of interest for food can be analyzed, such as K, Ca, Fe, Cu, Zn, etc.

Methods of Analysis

Although sodium may be determined by means of a wide variety of instrumental techniques, in line with the applications recommended in scientific publications and the most suitable instruments, the technique used most often is atomic spectroscopy, either by emission or by absorption (Table 1).

This technique is based on the atomic properties of the elements, and more specifically on their capacity to react in response to specific stimuli, modifying their orbital distribution, for which purpose they absorb light at a specific wavelength for each element, passing on to a state of higher energy, or lower energy if passing from the previous stage, emitting light at the same wavelength. In the case of sodium, the recommended specific lengths are 589.6 and 589.0 nm.

In order to achieve the passage of electrons to a state of higher energy, it may be necessary to increase the temperature, usually by means of a flame, or the incidence of a beam of light of a specific wavelength (Table 2). The latter technique enables the determination of the analytical concentration by determining the difference between the energy supplied to the sample and that which remains after crossing – this is genetically known as atomic absorption spectroscopy. In contrast, if the energy of the element is increased by another method (flame), what is usually measured in order to determine the concentration of the analyte is the emission of characteristic wave light when returning to the state of lower energy; this technique is known as atomic emission spectroscopy.

Emission Spectroscopy

Atomic spectrophotometers and emission flame photometers can be used to determine sodium content. In both cases, a sample in a liquid state is introduced into a fuel carrier (acetylene or propane) and combined with atomized or nebulized N_2O_2 air or oxygen. This dispersion is led to a lighter where the fuel combusts. The heat of the flame prompts an increase in the energy of the sodium atoms, which, upon returning

Table 1 Standard conditions for sodium analysis by atomic absorption spectrometry

Spectrometer setting	*Potassium*
Wavelength/slit (nm)	589.0/0.2
Nebulizer	Spoiler
Oxidant	Air
Fuel	C_2H_2
Flame condition	Oxidizing
Optimum concentration range in solution (μg ml^{-1})	1–20
Interferences	Ionization

Table 2 Accuracy and precision of sodium analysis by flame emission

Accuracy	*Certified (g kg^{-1})*	*Found (g kg^{-1})*	*IC (95%)*[a]	*Rec (RSD)*[b]
Citrus leaves SRM-1572	0.16 ± 0.02	0.17 ± 0.03	0.11–0.24	108 (18)
Nonfat milk powder NIST-1549	4.97 ± 0.10	5.17 ± 0.14	4.90–5.46	104 (2.8)
Precision	*Relative standard deviation*			
Plant food	1.15			
Dairy food	1.2			

[a]Intervals of confidence (95%).
[b]Recovery percentages and relative standard deviation.

to their fundamental state, emit a light beam of characteristic wavelength that is quantified by a detection system. The amount of light emitted is proportional to the concentration of the analyte in the sample. In addition to the option of atomic absorption, in which a hollow cathode lamp must be used in order to generate the specific lengthwave, most atomic spectrophotometers enable emission spectroscopy to be performed, an option that requires a lamp and provides good results.

With this technique, the optimum gain is adjusted to the highest standard of sodium used, and an attempt is made to keep the lighter free of sodium impurities emitted by the actual samples and standards analyzed, since these impurities may introduce undesirable emissions or obstruct the passage of fluids through the lighter; when the flow decreases, the signal obtained will also decrease. Avoiding samples with high concentrations of sodium (such as undiluted sea waters, cheese sera, etc.) is considered good practice.

A further potential problem lies in small fluctuations that may occur in the flame or in the carried mixture; these may lead to inaccurate and variable results. To correct this problem, some devices are equipped with a second detector that enables the emission of a second element (usually lithium) to be estimated, with a different wavelength to that of sodium. This second element must be added to solutions of both samples and patterns in a stable and known concentration. When there is a variation caused by the flame or the flow, both detectors must reflect this simultaneously, distinguishing this from variations in the analyte.

Another precautionary step that must be observed when working in atomic emission with sodium involves adjustment of the calibration line, since, in most cases, this is not linear; instead, it displays a decrease in the slope when concentrations increase. This is due to the fact that the area of linearity between the analytical concentration and atomic emission is exceeded. The simplest solution involves diluting the samples until the standards that include them present a straight adjustment; however, it is feasible to make a logarithmic adjustment between the emission and the concentration of the type $E = a^*\ln[\mathrm{Na}] + b$, or by means of a simple logarithmic transformation of the concentrations prior to the calculation of the regression line, and the subsequent retrotransformation when applying the linear equation in order to calculate concentrations. This type of adjustment is usually extremely useful when the sample values are high but not very different from one another, the results of the equation being valid, provided that all samples are located between the minimum and maximum values of the patterns used. As a result of this restriction, calculations such as those corresponding to the detection or concentration limit cannot be performed reliably on the standards used, although the high concentrations used guarantee that the readings can never be confused with the fluctuation of the baseline of the device.

Atomic Absorption Methods

Although the basis of the technique is the same as that for atomic emission, and in many cases, the same equipment is used for both analyses, it is true that this technique is favored by fewer analysts, owing in part to the fact that for normal concentrations of sodium in foods, calculations may be performed for emissions with little need to perform dilutions, and because of the added cost of using a hollow cathode lamp for analysis, without this being justified by the analytical improvements achieved. However, the existence of multielement lamps provides researchers with the possibility of using this technique with their samples at no further cost. This technique tends to be recommended more than the emission technique when the sample concentrations present very low values, and lower detection limits may be obtained.

Ion-selective Electrodes

Together with the optical techniques of absorption and atomic emission, this is one of the most common

techniques used for analyzing sodium content in foods; one of the reasons is the low cost of the instruments required, their ease of use, speed, noninvasive nature, and even the possibility of using them in food production lines without causing any interruptions.

The technique is based on the capacity to enable the passage of ions through a glass membrane, each type of membrane being specific to an ion or small group of ions. In the case of sodium, the probe is capable of detecting this activity once the membrane has been crossed; this activity is proportional to concentration. Possibilities of interference with other ions exist, i.e., specificity is not as high as with spectromorphometric methods, and the formation of complexes with other substances reduces sodium activity and therefore its capacity for determination. However, its usefulness becomes clear when the expected levels of sodium are high, as is the case with salted products, such as ham, cheese, etc., in which controlling salt penetration may be recommended. (*See* **Potassium**: Properties and Determination.)

Inductively Coupled Plasma Atomic Emission Spectrometry (ICPAES)

ICPAES enables samples with a high degree of variability and minimum interferences to be used, and these can be prepared either by wet oxidation or by dry ashing. A wide range of concentrations can be used without the need for dilution or concentration, and many of the mineral elements can be quantified, thus reducing analysis time. These advantages, together with the high speed and excellent instrument stability, make the ICPAES method highly attractive in view of its analytical applications.

Other Techniques

A large number of other analytical techniques have been described that may be used to determine sodium content in foods, ranging from classical gravimetric methods, now virtually obsolete, to techniques that involve the use of instruments that are expensive and sophisticated, this factor substantially limiting their use. The most noteworthy techniques described in the Further Reading list include the following: nuclear magnetic resonance (NMR) spectroscopy, ion chromatography, capillary ion electrophoresis analysis, electron-probe microanalysis, helium glow photometry, ion-microprobe analysis, laser probe analysis, ion-scattering spectrometry, auger spectroscopy, impact radiation analysis, enzymatic determination, microwave-induced plasma atomic emission spectrometry and α-particle backscattering analysis. (*See* **Potassium**: Properties and Determination.)

Of all these less common methods, NMR spectroscopy been used increasingly in recent years for determining sodium content, mainly in living tissue or biological samples. Therefore, it can be used for foods, particularly for establishing sodium distribution in such samples. In this technique, Na^+ has a nuclear magnetic moment of 3/2, which allows it to undergo nuclear magnetic resonance; ^{23}Na is extremely easy to detect in biological samples. For more information on this technique (*see* **Spectroscopy**: Nuclear Magnetic Resonance.) Nevertheless, this technique must be used with caution, since comparative studies indicate discrepancies in terms of results obtained in connection with the quantification of sodium when using this technique, which may only detect 60% of sodium.

In contrast, other methods are merely variations on the major analytical techniques, such as emission spectrometry or selective electrodes, are normally used in microsamples such as blood serum or other body products, and are rarely used in foods.

See also: **Hypertension**: Hypertension and Diet; **Potassium**: Properties and Determination; **Sodium**: Physiology; **Spectroscopy**: Nuclear Magnetic Resonance

Further Reading

Ammon D (1986) *Ion Selective Microelectrodes*. New York: Springer.

Benton Jones J Jr. (1984) Developments in the measurement of trace metal constituents in foods. In: Gilbert J (ed.) *Analysis of Food Contaminants*, pp. 157–202, Barking, UK: Elsevier.

Commission of the European Communities (1993) *Nutrient and Energy Intakes for the European Community. Reports of the Scientific Committee for Food, Thirty-first Series*, pp. 174–178.

Gunther K and von Bohlen A (1990) Simultaneous multi-element determination in vegetable foodstuffs and their respective cell fractions by total-reflection X-ray fluorescence (TXRF). *Zeitschrift für Lebensmitteluntererersuchung und -forschung* 190(4): 331–335.

Intersalt Cooperative Research Group (1988) Intersalt: an international study of electrolyte excretion and blood pressure. Results for 24 hour urinary sodium and potassium excretion. *British Medical Journal* 297: 319–328.

Kane MR, Fregly MJ and Berbard RQ (eds) (1980) *Biological and Behavioral Aspects of Salt Intake*. New York: Academic Press.

Luft FC (1990) Sodium, chloride and potassium. In: Brown M (ed.) *Present Knowledge in Nutrition*, 6th edn, pp. 233–240. Washington, DC: International Life Sciences Institute Nutrition Foundation.

Physiology

A R Michell, University of London, Hatfield, Hertfordshire, UK

Physiological, Clinical and Nutritional Importance of Sodium

Despite the fact that the body contains more calcium and potassium, sodium is arguably the most important cation because it dictates the volume of extracellular fluid (ECF) and its concentration affects osmotic concentration of both ECF and intracellular fluid (ICF). Abnormalities of ECF sodium concentration cause movement of water into or out of cells, thus altering the osmotic concentration of ICF in parallel and causing swelling or shrinkage of cells. The main impact of this is on the brain because its cells are rigidly enclosed by the cranium.

Sodium depletion is mainly caused by enteric, renal, or adrenal disease, and sodium retention is caused by renal disease; healthy kidneys are well able to excrete excess dietary salt. However, chronic ingestion of excess salt, whether or not it increases ECF volume, is a predisposing or exacerbating factor in hypertension. Until the 1980s, knowledge of the regulation of body sodium mainly concerned defenses against depletion, whereas the last two decades have seen a rapid growth in knowledge of the mechanisms which excrete excess sodium. This seems appropriate since most species, especially humans, dogs, and laboratory rats, are exposed to dietary sodium intakes well above their nutritional requirement. (*See* **Renal Function and Disorders**: Kidney: Structure and Function.)

The nutritional requirement is a reflection of obligatory losses (maintenance) and the needs of growth, pregnancy, and lactation. Abnormal losses owing to disease, or in animals such as humans and horses which sweat extensively, raise the requirement. The impact of equine sweating is different from that in humans. Human sweat always contains sodium at concentrations well below plasma levels (and when aldosterone secretion is raised, levels of sweat sodium fall very low); horse sweat is hypertonic but this helps to offset the osmotic effect of the increased respiratory water loss during exertion, i.e., it may be a defense against hypernatremia, rather than a potential cause of sodium depletion.

Consideration of the physiology of sodium thus includes its distribution in the body, regulation of total content and concentration, causes of and responses to depletion or excess, and their nutritional implications.

Distribution

Sodium behaves physiologically as a cation, i.e., a positively charged ion; its distribution and effects are fairly independent of the negative ions (anions) which originally accompanied its ingestion, though they may affect its absorption and excretion. Most sodium is in ECF (Table 1), kept there by the sodium pump, an enzyme system, (Na^+-K^+)-ATPase, which uses substantial amounts of energy (adenosine triphosphate; ATP) in maintaining a low intracellular sodium concentration and a high intracellular potassium (K^+) concentration. Sodium transport is a central issue in the physiology of sodium for a number of reasons:

1. It helps to maintain the ionic environment of ICF and the volume of ECF.
2. It prevents cell swelling (the Na^+ efflux exceeds the K^+ influx).
3. It establishes gradients which, in various tissues, allow transport of other cations in exchange, other anions in parallel or organic solutes – these are often cotransported with sodium down concentration gradients which are secondary to the low sodium environment created by the pump.
4. It establishes the membrane voltages on which excitability and secretory activities frequently depend.
5. The energy expenditure of the pump is a substantial portion of total metabolic activity and contributes to thermogenesis.
6. Sodium transport is not only a key factor in the retention and loss of sodium in the kidney, gut, salivary, and sweat glands but also influences the excretion or retention of many other solutes. Thus, for example, diuretics intended to promote sodium excretion may also cause unintentional losses of potassium and magnesium. Similarly, when renal

Table 1 Summary of sodium (Na) distribution and requirements

Typical plasma Na concentration (mmol l^{-1})	*145 (130–160)*
Typical body Na content (mmol kg^{-1})	*50–55*
Typical proportion (%) of total Na	
Intracellular fluid	*10*
Extracellular fluid	*50*
Bone	*40*
Maintenance requirement in mammals (mmol kg^{-1} day^{-1})	
Sheep	*0·1*
Cattle and goats	*0·3–0·7*
Pigs	*0·6*
Rats	*0·6*
Dogs	*0·2–0·5*
Humans	*?·0·6*

1 mmol = 23 mg Na^+, 58.5 mg NaCl.

sodium excretion increases appropriately in response to ingestion of excess salt, there may also be unwanted losses of calcium and in postmenopausal women these may contribute to loss of bone mineral. (*See* **Electrolytes**: Analysis; **Thermogenesis**.)

Bone also contains substantial quantities of sodium but, as yet, its significance is unknown since it does not appear to be mobilized during sodium depletion. Gut fluids contain considerable amounts of sodium, mostly secretory rather than dietary, and mostly reabsorbed in more distal regions of the intestine.

Extracellular Sodium

Of the extracellular fluid, most is in interstitial fluid (ISF) in the tissue spaces, providing the transport medium between capillaries and cells. The sodium concentration in plasma is slightly above that in ISF because plasma contains more proteins, notably albumin, which do not readily escape into ISF across the capillary membranes, and the effect of their negative charges is to hold more positively charged ions, notably sodium, in circulation (Gibbs–Donnan equilibrium).

The main effects of excess ECF volume are seen as expanded ISF, visible clinically as edema (or ascites, when fluid accumulates in the abdomen rather than the tissue spaces). Mild edema is merely a cosmetic problem in itself but pulmonary and cerebral edema, or severe ascites, are potentially serious forms. Edema can result from excess ingestion or retention of sodium (overall expansion of ECF) or leakage from plasma to ISF, with plasma volume continuously replenished by renal sodium retention. Such maldistribution of ECF occurs if plasma albumin is very low (renal leakage, hepatic impairment, or severe malnutrition), or with excessive capillary blood pressure (venous blockage, inactivity, heart failure, or arteriolar dilation, e.g., from heat or allergy), capillary damage, or lymphatic blockage. Lymphatic blockage prevents the removal of proteins which have leaked into ISF. Accumulation of protein in ISF undermines the osmotic gradient which normally favors uptake of water at the venous end of the capillary, where the pressure is lower. Since edema involves the expansion of a larger compartment (ISF) from a smaller one (plasma), it is only possible as long as the latter is replenished; hence the kidney, while seldom the primary cause of edema, is always the enabling cause; the use of diuretics is therefore appropriate in the treatment of nonrenal as well as renal causes of edema.

The main effect of inadequate ECF volume is to reduce plasma volume and thus to compromise cardiovascular function, in extreme cases by causing circulatory shock.

Regulation of ECF Sodium

In a mature, nonpregnant, nonlactating, healthy animal, sodium excretion matches sodium intake and is often used to estimate it, although this is not reliable, especially when intake is low. Dietary sodium is readily available, i.e., readily absorbed; thus the traditional view of sodium regulation emphasizes renal regulation of urinary Na^+ loss. This oversimplifies the more subtle interplay seen, for example, in herbivorous animals, where salt appetite may contribute to regulation by intensifying during sodium depletion. Moreover, in many herbivores the feces, rather than urine, may be the major route of sodium excretion and the gut may therefore be an important regulator of sodium balance. Indeed, it is interesting that sodium transport mechanisms in the small intestine show considerable similarities to those of the proximal part of the renal tubules (e.g., linked transport of Na^+, glucose, and amino acids) whereas the colon, like the distal nephron, responds to the salt-retaining (and potassium-shedding) hormone of the adrenal cortex, aldosterone.

Provided that the adrenal gland is healthy, urinary and fecal sodium loss can be reduced virtually to zero. Sweat loss can also be very low, although with severe exertion in hot climates the volume of sweat may exceed the ability of aldosterone to reduce its sodium concentration and net loss of sodium can occur. Aldosterone also reduces salivary sodium (and raises $[K^+]$). (*See* **Potassium**: Physiology.)

There are two components to the regulation of ECF sodium: the total amount of sodium retained, and its concentration. The former is regulated by mechanisms which directly affect sodium, whereas the latter is essentially regulated via water balance. Thus whatever sodium is retained in ECF is 'clothed' with the appropriate amount of water to maintain the normal plasma sodium concentration within narrow limits; deviations of less than 1% (hard to measure in the laboratory) trigger corrective responses. Thus a raised plasma sodium concentration (e.g., after water loss) stimulates both thirst and renal water conservation; antidiuretic hormone (ADH) from the posterior pituitary reduces urine output through its effect on the renal collecting ducts. Even one of these mechanisms can defend body water; thus diabetes insipidus (inadequate production or effect of ADH) does not cause severe dehydration but polydipsia (increased fluid intake; thirst is a sensation).

Excess salt intake does not raise plasma sodium concentration (hypernatremia) if water is available and the patient can drink; the excess sodium is diluted. The resulting increase in ECF volume then stimulates increased sodium excretion. Sodium also enables ECF to hold water against the osmotic 'pull' of the solutes in ICF and sodium thus functions as the 'osmotic skeleton' of ECF; it is the main determinant of its volume.

Plasma sodium concentration is therefore only indirectly related to sodium balance. When ECF volume, notably circulating volume, is severely reduced, this stimulus, rather than Na^+ concentration, becomes the main drive for thirst and ADH secretion. Until ECF volume is restored, water is retained (to protect ECF volume) even though this undermines the protection of ECF Na^+ concentration and, as a result, plasma sodium falls. Thus, during sodium depletion, contraction of ECF volume precedes significant reductions of plasma Na^+ and is therefore a poor index of sodium status.

Sodium-Retaining Hormones

Sodium depletion, by reducing plasma volume and renal perfusion, stimulates the production of renin (from the kidneys) which generates angiotensin in circulation. This hormone is a vasoconstrictor (so protects blood pressure), stimulates thirst (so helps to restore ECF volume) and, above all, stimulates sodium retention both directly (renally) and indirectly (by stimulating adrenal secretion of aldosterone); it thus reduces the sodium concentration of urine, feces, saliva, and sweat, but not milk. (*See* **Hormones**: Adrenal Hormones.)

Indices of aldosterone secretion (reduced sodium or potassium concentration in urine, feces, etc.) are often taken as evidence of sodium depletion or inadequate sodium intake, but the following points apply:

1. Aldosterone secretion is also stimulated directly by hyperkalemia (elevated plasma K^+) and promotes potassium excretion.
2. Such interpretations involve a subjective judgment concerning adequate or excessive sodium intake. Because physiologists and clinicians were traditionally more concerned with sodium depletion as well as its consequences and the defenses against it, elevated aldosterone secretion was readily seen as a warning signal. However, if sodium intakes associated with increased aldosterone have no other harmful effects, and especially if excess sodium intakes cause concern, low levels of aldosterone secretion might equally indicate excessive salt intake.

While sodium reabsorption in the distal nephron, influenced by aldosterone, is particularly important because it can produce sodium-free urine and promote potassium loss, the great majority of renal sodium reabsorption occurs elsewhere: about 25% in the loop of Henle and most in the proximal tubule. The loop is also a main site of magnesium reabsorption, hence the tendency for loop diuretics to cause hypomagnesemia.

The factors controlling proximal reabsorption are incompletely understood but their effect is clear: proximal reabsorption of sodium increases or decreases according to the need to enhance or diminish plasma volume. Since the fluid in the proximal tubule is similar to plasma, having been formed from it by glomerular filtration, it has the ideal composition for this purpose.

Natriuretic Hormones

Excretion of excess sodium involves not only suppression of salt-retention mechanisms but also activation of sodium-shedding (natriuretic) mechanisms. Two types of hormones are involved: atrial natriuretic peptide (ANP), produced by the cardiac atria when they are overstretched (reduction of ECF volume being an appropriate response to cardiac overload), and active sodium transport inhibitors (ASTIs), probably produced within the brain. These were probably the original molecules associated with the receptors binding cardiac glycoside drugs and are therefore also called endogenous digitalis-like inhibitors (EDLIs); their exact identity remains uncertain. ANP has various effects which essentially oppose those of the salt retention induced by aldosterone: it increases sodium excretion, lowers arterial pressure, and promotes movement of ECF towards the interstitial compartment.

Other hormones (e.g., sex steroids, parathyroid hormone, calcitonin, thyroid hormone, prolactin) affect renal sodium retention or loss but are not thought to regulate it. (*See* **Hormones**: Thyroid Hormones; Pituitary Hormones.)

Adequate, Inadequate, and Excess Sodium

It is unlikely that adult daily maintenance requirement exceeds 0.6 mmol per kg body weight and could well be below this in many mammals. Newborn, growing, pregnant, or lactating animals have increased requirements. The appropriate sodium intake for humans remains controversial, with some cultures managing on less than 1 mmol day^{-1}, while western intakes may be in the range 200–300 mmol day^{-1} – more where processed foods are heavily

consumed. There has been insufficient awareness among physicians and human nutritionists of just how high such intakes are, compared with requirement in other animals. Granted that humans are bipeds with a stressful lifestyle quite different from those of animals, there is no real evidence that human obligatory losses or sodium requirements are significantly greater. Rather, there is an ingrained tradition of regarding sodium intake as a benign pleasure, involving a harmless and healthy dietary constituent. The main warnings against this view come from the fact that hypertension is virtually unknown in low-salt cultures and that they do not even have an age-related rise in normal blood pressure. Moreover, there are numerous studies which, when rigorously analyzed, indicate that human arterial pressure and salt intake are positively correlated – sufficiently to anticipate large reductions in the prevalence of hypertension in response to manageable reductions in dietary sodium. Unfortunately, such reductions are handicapped by inadequate food labeling and the fact that most sodium is added by the processor rather than the consumer.

Because obligatory losses of sodium are so low, dietary sodium depletion is hard to induce and sodium deficiency usually results from losses caused by renal, adrenal, or enteric disease; renal disease may cause either retention or loss of sodium. Globally, both in humans and animals, the most common cause of sodium deficits is acute diarrhea. Fortunately, sufficient gut usually remains unaffected for uptake of sodium and water to be stimulated by suitably formulated oral rehydration solutions. These essentially restore ECF volume (and acid–base balance), allowing natural defenses to overcome the underlying cause of the diarrhea. Despite some species variations, such solutions usually work best if their glucose:sodium ratio (in $mmol\,l^{-1}$) is close to unity and they are virtually isotonic (i.e., they have a similar osmotic concentration to ECF; hypertonic solutions draw water into the gut). The function of glucose in these solutions is to promote sodium uptake; its nutritional contribution is trivial. Anions such as citrate, acetate, propionate, bicarbonate, and amino acids (e.g., glycine and alanine) may further enhance the uptake of sodium and therefore water. These sodium cotransport mechanisms are very similar to those of the proximal renal tubule.

Sodium is thus central to the management of two of the most widespread clinical problems: hypertension (in humans) and diarrhea. Indeed, the World Health Organization (WHO) regards the discovery of oral rehydration, which depends on restoration of enteric sodium uptake, as the main life-saving development in twentieth-century medicine. This powerful clinical application rests on a simple physiological observation concerning an elementary but vital dietary constituent. (*See* **Hypertension**: Physiology; Hypertension and Diet; Nutrition in the Diabetic Hypertensive.)

See also: **Electrolytes**: Analysis; **Hormones**: Adrenal Hormones; Thyroid Hormones; Pituitary Hormones; **Hypertension**: Physiology; Hypertension and Diet; Nutrition in the Diabetic Hypertensive; **Potassium**: Physiology; **Renal Function and Disorders**: Kidney: Structure and Function; **Thermogenesis**

Further Reading

Avery ME and Snyder JD (1990) Oral therapy for acute diarrhea. *New England Journal of Medicine* 323: 891–894.

Denton DA (1982) *The Hunger for Salt*. Berlin: Springer-Verlag.

El-Dahr SS and Chevalier RL (1990) Special needs of the newborn infant in fluid therapy. *Pediatric Clinics of North America* 37: 323–335.

Field M, Rao MC and Chang EB (1989) Intestinal electrolyte transport and diarrheal disease. *New England Journal of Medicine* 321: 800–806, 879–883.

Hirschhorn N and Grenough WB (1991) Progress in oral rehydration therapy. *Scientific American* 264: 16–22.

Law MR, Frost CD and Wald NJ (1991) By how much does dietary salt lower blood pressure? *British Medical Journal* 302: 811–815, 815–818, 819–824.

Maxwell MH, Kleenan CR and Narins RG (1987) *Clinical Disorders of Fluid and Electrolyte Metabolism*, 4th edn. New York: McGraw-Hill.

Michell AR (1989) Physiological aspects of mammalian sodium requirement. *Nutrition Research Reviews* 2: 149–160.

Michell AR (1991) Regulation of salt and water balance. *Journal of Small Animal Practice* 32: 135–145.

Rutlen DL, Christensen G, Helgesen KG and Ilebekk A (1990) Influence of atrial natriuretic factor on intravascular volume displacement in pigs. *American Journal of Physiology* 259: H1595–H1600.

SOFT DRINKS

Contents

Chemical Composition

K Jorge, Brewtech Ltd, Jacarepagua, Rio de Janeiro, Brazil

Introduction

The forerunner of soft drinks began more than 2000 years ago when Hippocrates first suspected that mineral waters could be beneficial to our well-being, for relaxation and bathing.

More than 1000 years passed before mineral waters made the transition from therapeutic bath to refreshing beverage.

Scientists produced artificially carbonated water in the laboratory, and it was only a matter of time before soft drinks made it into the hands of the public. Pharmacists believed they could add curative properties to the water by experimenting with several interesting flavors and tastes. The soft-drink industry started as a seasonal business, operating primarily during the summer months.

Gradually, demand grew for soft drinks to be consumed at home. Bottling the products proved to be difficult, since pressure from the carbon dioxide forced corks out of the bottles. Then, in 1892, the 'crown cap' was invented, revolutionizing the soft-drink industry. Since that time, considerable transformation occurred, making the soft-drink industry more efficient and productive with a high degree of mechanization, together with the use of new containers and the addition of attractive flavors.

Product Definition

Typically, soft drinks are nonalcoholic (less than 1% alcohol by volume), carbonated, flavored beverages, which are usually drunk cold. They are available in bottles, cans, or dispensing systems. They include lemon, orange, and lime drinks, or other fruit-flavored beverages such as colas, ginger, sodas, tonics, and cocktail mixers.

Regular soft-drink ingredients are listed on the label or cap in descending order of importance. The precice recipes are valuable corporate properties that are kept secret. The basic ingredients of most soft drinks are: water purified to the exacting standard required; high-fructose corn syrup, sugar, or a combination of the two to achieve the desired sweetness; carbon dioxide, which creates the bubbles that stimulate the palate and heighten taste sensations; some sodium (in the majority); and flavors. Diet soft drinks use calorie-reduced sweeteners. The flavors used may be very complex. For example, the orange flavor contains up to 400 individual ingredients. Flavor concentrates from international manufacturers are constant throughout the world. Table 1 lists a typical soft-drink composition.

Flavors

Flavors are used to give soft drinks a pleasant taste. It is the flavoring that gives a soft drink its specific taste, smell, and appearance. Although it is only added in very small quantities, flavoring is one of the most important raw materials of a soft drink plant.

Usually, it is a balanced composition with different aromatic substances from different types of fruit. Fruit, berries, and a number of other plants contain substances that have an agreeable aroma, and are commonly called 'tastes.' 'Taste' comprises only four taste sensations – sweet, salt, sour, and bitter. All other 'tastes' are smell in combination with one or more of these basic tastes.

When eating, the smell passes not only through the nose, but also through the throat up to the area in the nose that is sensitive to smell. The brain puts together

Table 1 Illustration of a typical soft-drink formula

Ingredient	*Percentage by weight*
Carbonated water	87–89
Sugar	10–12
Acids, colorings and flavorings	< 1.0

these sensations, which are translated as 'aroma.' According to international praxis, flavors are divided into three categories – natural, nature identical, and artificial. (*See* **Sensory Evaluation**: Taste.)

Natural flavors are mainly based on essential oils and extracts where natural aromatic substances dominate. In addition to the aromatic components, they also contain constituent parts from the fruit, which change during storage and can worsen the taste of the drink. For this reason, these natural aromatic substances are sometimes strengthened with certain identical aromatic substances that contain aromatic components only and give a better stability to the taste. (*See* **Essential Oils**: Properties and Uses; Isolation and Production.)

Nature-identical flavors are based on ingredients as they occur in the natural fruit but which are produced by chemical methods. They have a good solubility, a reasonable price and a long shelf-life, which makes them very suitable for soft drinks production.

Artificial flavors contain chemicals that do not occur in nature but have a familiar taste such as vanillin, which has the taste of vanilla. This is an ever-decreasing category and in time will be replaced by nature-identical flavors.

Water

Water is the main raw material in soft drinks. In weight terms, a soft drink is 90% water. Water of high quality is therefore important. It must be pure and from a source approved by the regulatory authorities for drinking purposes. This means that it must be free from biological and chemical pollutants, such as bacteria and poisonous substances. It should also have the cleanest taste and be least interactive and microbiologically stable. The quality of water has an impact on the quality and taste of the soft drink.

Sweeteners

Sweeteners can be divided into two main categories: nutritive or caloric sweeteners and nonnutritive or low calorie sweeteners. We will discuss only the sweeteners important for the soft-drink industry.

Sugar

Sugar is a generic name for a class of sweet carbohydrates, including fructose, glucose, maltose, sucrose, and lactose, that are readily soluble in water, colorless, odorless, and usually crystallizable.

For soft drinks, the most commonly used is sucrose, which can be supplied either as a liquid or in granulated form. Liquid sugar is dissolved granulated sugar in water, normally at 67° Brix.

Sugar is an important ingredient for soft drinks, not only because of its properties as a sweetener, but also because it is highly nutritious and is an important carrier of the fruit aromas. The different fruit aromas require different quantities of sugar, depending on which fruits are in the drink. This means that the composition of the sugar solution is critical if the drink is not to be too 'thin' or taste too sugary. Two other types of sugar are found in fruit – fructose (fruit sugar) and glucose (grape sugar). Fructose is 50% sweeter than ordinary sugar, but contains the same number of calories per gram. A smaller quantity of fructose gives the same sweetness. It is therefore sometimes used to reduce the number of calories. Honey is another less commonly used sweetener that also contains glucose and fructose. (*See* **Carbohydrates**: Classification and Properties; **Glucose**: Properties and Analysis; **Honey**; **Lactose**; **Sucrose**: Properties and Determination; Dietary Importance.)

Nonnutritive Sweeteners

In the past, low-calorie drinks were only marketed as diabetic products. Now, they have become increasingly popular for both weight-watching and economic reasons. Low-calorie drinks are an important part of the total soft-drink market. When using a nonnutritive sweetener, it is important to have up-to-date information about the food regulations concerning this subject, as they differ greatly from country to country and are subject to many changes.

A sweetener has to comply with the following standards, assuming it is allowed by the food laws concerned:

- It should not be hazardous to health.
- It should give no off-taste.

Saccharin

Saccharin is a synthetic, white, crystalline powder, of formula $C_6H_{16}CONHSO_2$, which, in its pure state, is 550 times sweeter than sugar. In its commercial form, saccharin is estimated to have a sweetening power 375 stronger than that of sugar.

The only synthetic, nonnutritive sweetener presently allowed in most countries in soft drinks is saccharin. It has been in use continuously since 1900. The use of saccharin is mostly restricted to special dietary foods and beverages appropriately labeled as such.

For several years, saccharin has been under investigation as a potential cause of cancer. Warning labels were required on saccharin-containing foods. The most recent studies indicate that it is a weak carcinogen in laboratory animals but does not, in moderate use, present an increased cancer risk in humans. The

recommended acceptable daily intake (ADI) is 1 g per person per day. The tolerance level for soft drinks is 0.4 mg ml^{-1}. This is an interim tolerance level, and safety studies are continuing. (*See* **Cancer**: Epidemiology; **Carcinogens**: Carcinogenic Substances in Food: Mechanisms; Carcinogenicity Tests; **Saccharin**.)

Cyclamate

Cyclamate is an odorless, white crystalline powder. The name usually denotes either calcium cyclamate or sodium cyclamate, both of which are salts of cyclohexylsulfamic acid. They have a very sweet taste, with about 30 times the sweetening power of sucrose.

After saccharin, cyclamate was the most commonly used sweetener until it was banned in the USA in 1970, after which many countries followed suit. Much of the scientific community doubted that cyclamate was hazardous but for new FDA approval more very costly research is needed and people would rather concentrate on other sweeteners. (*See* **Cyclamates**.)

Aspartame

Aspartame is a sweetener made from the natural amino acids L-aspartic acid and L-phenylalanine. L-Aspartic acid is tasteless, and L-phenylalanine is slightly bitter, so the taste of aspartame could never have been predicted. Aspartame was discovered by accident in the laboratories of the Searle Company in 1965 and was first approved for use in food products in Canada 15 years later. In 1983, approval was granted by the FDA, followed by many authorities in other countries. The only negative aspect of aspartame is that it can be hazardous for people who suffer from phenylketonuria, an inherited disorder that can lead to brain damage in those exposed to phenylalanine. This disease is rare, however. Since aspartame is so sweet (about 200 times sweeter than saccharin), only a small quantity is needed, and it is therefore put into the same category as energy-free sweeteners. The calorie content can be reduced by 95% compared with corresponding drinks sweetened with saccharates. During digestion, aspartame is treated as protein, and so diabetics are able to eat and drink food sweetened with aspartame. (*See* **Amino Acids**: Properties and Occurrence; **Aspartame**; **Protein**: Chemistry; Food Sources; Determination and Characterization; Requirements; Functional Properties; Interactions and Reactions Involved in Food Processing; Quality; Digestion and Absorption of Protein and Nitrogen Balance; Synthesis and Turnover; Deficiency; Heat Treatment for Food Proteins; Sources of Food-grade Protein.)

Acesulfame-K

Acesulfame-K was developed by Hoechst in 1973, was approved by the British Food Regulations in 1983, and has since been adopted in countries like Germany, Belgium, and Denmark. Acesulfame-K is not metabolized at all and therefore has no calorific value. It is absorbed quickly in the intestine, and excretion is similarly fast. (*See* **Acesulfame/Acesulphame**.)

Stevia (*Stevia rebaudiana Bertoni*)

Stevia contains natural compounds, especially stevioside and rebaudioside A, that are estimated to be 150–400 times sweeter than saccharose. Used for centuries in parts of South America, stevia has been discovered in recent years by much of the calorie-conscious modern world. It is now widely and legally consumed by millions of people, from South Korea, Israel, and the People's Republic of China, but no country has done more to demonstrate stevia's dietary and economic potential than Japan, where the herb and its extract have been used since 1970s.

The Japanese, having subjected stevia extract to extensive safety testing and found it without health risk, now incorporate it in numerous food products, including soft drinks.

Sucralose

Sucralose was approved by the FDA in 1998 for use in a wide variety of food products including soft drinks. Sucralose is a low-calorie, high-intensity sweetener that is about 600 times sweeter than sugar. It is sold under the brand name of 'Splenda.' Sucralose and sucrose (sugar) have been shown to have similar taste and flavor profiles.

A number of other fascinating low-calorie sweeteners are currently undergoing safety evaluations for future use. These include alitame, a compound similar to aspartame that is remarkably 2000 times sweeter than sucrose, and various naturally occurring plant derivatives, such as stevia and thaumatin.

Acids

Acids are used as pH buffers in beverages and to impart an astringent taste that offsets the sweetness of the sugar and enhances or complements the associated flavor. Thus, the characteristic flavor of a beverage is developed in part through proper acidulation. Acids also help in preventing microbiological spoilage.

All acids used in beverages must be 'edible grade' or 'food grade.' Those commonly used are citric-, phosphoric-, and malic acid. Each has the property of being weak and nonharmful at the concentrations

used. Inherent chemical differences in the various acids impart different taste characteristics, and in substituting one acid for another, these differences have to be kept in mind. Table 2 shows the acidification strength of some acids relative to citric acid.

The kind of acid and the concentration employed are very important for adjusting the flavor of the beverage.

Citric Acid

Traces of citric acid are found in numerous plants and animals, because it is a nearly universal intermediate product of metabolism. Large amounts of the acid are found in the juice of citric fruits. Fermentation of sugar by *Aspergillus niger* is the chief commercial source of the acid. Since it is a natural ingredient of citrus fruits, it adapts itself well to beverages with such a flavor.

Citric acid may be used in the anhydrous form or containing one molecule of water. It occurs as colorless, translucent crystals or as a white, granular-to-fine crystalline powder. It is odorless with a strong acid taste, and the hydrous form is efflorescent in dry air. (*See* **Citrus Fruits**: Types on the Market; Composition and Characterization; Oranges; Processed and Derived Products of Oranges; Lemons; Grapefruits; Limes.)

Phosphoric Acid

Phosphoric acid is a very popular acidulant because of its strength and very low cost; in fact, it is the cheapest acid used in the industry. It is used mainly in cola-type beverages, in which it improves not only the acid taste but also the typical cola bite.

Malic Acid

Malic acid is found in fruits such as apples, apricots, and cherries. The food industry finds the tartness of this additive a valuable flavor booster for candies, jams, ice creams, fruit, and soft drinks. (*See* **Ice Cream**: Methods of Manufacture; Properties and Analysis; Dietary Importance; Microbiology; **Jams and Preserves**: Methods of Manufacture; Chemistry of Manufacture; **Sweets and Candies**: Sugar Confectionery.)

Table 2 Acidification strength (%) relative to citric acid

Type	*Strength (%)*
Citric acid	100
Tartaric acid	93
Malic acid	100
Ascorbic acid	43
Lactic acid	60
Phosphoric acid	125

Carbon Dioxide

Carbon dioxide is present in the air in very small quantities (approx. 0.03%) and is about 1.5 times as dense as air. It is a colorless, odorless, and tasteless gas and has a good solubility in water. When confined within a suitable pressure vessel, CO_2 can exist as a solid, liquid, or gas.

When drinking a carbonated beverage, the development of CO_2 in the mouth, because of the warmth of the throat, causes a light anesthetic effect on the taste buds, which diminishes and sometimes even stops the sensation of thirst. Carbon dioxide makes the drink more refreshing through its stimulation of the mouth's mucous membranes. This adds to the sensation that the soft drink is colder than it actually is and also brings out the aroma.

Carbon dioxide also checks microbiological growth, which means that the product lasts longer. Many bacteria die after a certain time in carbonated drinks. This applies equally to carbonated soft drinks, mineral water, and beer. There are good grounds for advising people, for hygienic reasons, to drink carbonated drinks, particularly in hot climates. (*See* **Beers**: History and Types.)

Preservatives

The low pH value of soft drinks is good for the life of the product from a microbiological point of view since bacterial growth is checked. Carbon dioxide also has a positive effect. Sometimes, however, this is not enough, and preservatives have to be used.

Preservation denotes any method that extends the shelf-life of a product. Methods such as pickling, freezing, drying, sterilizing, smoking, or the addition of a chemical agent with a bactericidal activity all have undesirable effects on the odor, taste, or digestibility of the foodstuffs. (*See* **Drying**: Theory of Air-drying; **Freezing**: Principles; **Pickling**; **Smoked Foods**: Principles; Production; **Sterilization of Foods**.)

The history of the soft-drink industry has demonstrated that beverages are good nutrient media for certain acidophilic microorganisms, particularly yeasts. It has been demonstrated by various investigations that over 90% of all cases of microbiological spoilage of soft drinks are caused by yeast. Spoilage of soft drinks by yeasts is evidenced by visible sediment, gas formation, off-odor, off-taste, and changes in beverage color and clarity.

To prevent soft drinks spoiling, it is necessary to use good manufacturing practices to control the growth of microorganisms. High hygienic standards should be applied during production, together with the creation of unfavorable environments for microorganisms, such as acidifying, cooling and pasteurizing, and, if necessary, chemical preservation. Only small quantities of preservatives are allowed in soft drinks. The preservatives allowed are subject to the food and drugs laws of the different countries involved. However, chemical preservation alone is not sufficient. Therefore, these types of products must always be pasteurized.

The most common preservatives are sodium benzoate (sodium or potassium sorbate can be used as a replacement) and sulfur dioxide, particularly if it is required to prevent the color darkening.

Benzoic Acid

Benzoic acid or benzene-carbonic-acid is a monobasic aromatic acid, moderately strong, white crystalline powder, very soluble in alcohol, ether, and benzene, but poorly soluble in water (0.3 g of benzoic acid in 100 g of water at 20 °C).

Benzoic acid has the advantage that it does not affect the odor or taste of the soft drink, if used in small quantities. The preserving quality of benzoic acid is based on its activity to delay the multiplication of several groups of microorganisms, which, however, are not killed by this product. The low solubility of benzoic acid in water complicates its application in products containing large amounts of water. Therefore, the water-soluble salt sodium benzoate is used.

This product, which is the salt of benzoic acid, has no preserving activity by itself. Therefore, after addition of sodium benzoate, the acidity of the soft drink is increased (pH < 3.5), with the result that free undissociated benzoic acid is formed, which has a preserving property. In an alkaline environment, benzoic acid is split into ions and thus loses its preserving activity.

Sodium benzoate is the sodium salt of benzoic acid used as a white crystalline or amorphous (without crystal structure) powder, very soluble in water (66 g of sodium benzoate in 100 g of water at 20 °C) but poorly soluble in alcohol.

Sorbic Acid

Sorbic acid, potassium sorbate, and calcium sorbate are novel, highly efficient, safe, and nonpoisonous food preservatives. They are the substitute for the benzoic acid as a traditional preservative. Sorbic acid, potassium sorbate, and calcium sorbate approved worldwide are often now successfully used as standard products in many branches of the food industry. As they are acidic preservatives, it is better to use them at pH 5–6.

Sorbic acid, potassium sorbate, and calcium sorbate are unsaturated fatty acids and salts of unsaturated fatty acids, which participate in the normal fat metabolism in human body and are oxidized into carbon dioxide and finally water. They do not accumulate in the human body. (*See* **Fatty Acids**: Properties; *Trans*-fatty Acids: Health Effects.)

Sulfur Dioxide

The preservative effect of sulfur dioxide, like benzoic acid, is greatly increased by a corresponding decrease in pH. Undissociated H_2SO_3 is responsible for most of the preservative action. In contrast to sodium benzoate, sulfur dioxide kills microorganisms instead of delaying their growth.

Sulfur dioxide also combines regularly with many compounds present in juices and soft drinks, and the combined SO_2 exhibits little or no preservative effect. This must be taken into account when deciding on the levels required to preserve soft drinks, but it must be remembered that the legal limit is often based on the total free and combined SO_2 in the product. Sulfur dioxide can be added as the salt of sodium or potassium metabisulfite or as a solution of sulfurous acid. This compound gives off SO_2 gas and therefore must be stored in airtight containers. A disadvantage of the use of SO_2 as a preservative is its relatively bad taste at a dosage above 10 mg l^{-1}. (*See* **Sodium**: Properties and Determination.)

Coloring Matter

Food that looks beautiful when served, e.g., nicely served and has a beautiful color, affects the consumer's experience of taste positively. Color is an important signal of identification that complements the label and is a dominant factor in consumer acceptance of a beverage. An attractive, natural-looking color tempts one to taste and consume a beverage.

Before the mid-1970s, many more colorings were used than is the case today. The soft-drinks industry, itself, has reduced the number of coloring matters used. However, the pure fruit-juice content of soft drinks makes it difficult to create the right colors. For this reason, identical coloring agents of the sort found in the fruits are used extensively. The most important is β-carotene, which is the predominant coloring agent in carrots and oranges.

Brown drinks are colored with caramel. Different drinks require caramel with different qualities. This

can be achieved by having different nitrogen compounds present during the heating phase of caramel production.

Colors can be either natural or artificial. Natural colors include: anthocyanins (from berries and grapes), caramel, xanthophyl, and carotenoids. Artificial colors include: tartrazine NS, yellow FCS, and Amaranth AS. (*See* **Caramel**: Properties and Analysis; **Carotenoids**: Occurrence, Properties, and Determination; **Colorants (Colourants)**: Properties and Determination of Natural Pigments.)

Legal requirements for colors used in soft drinks differ from country to country, as do the required declarations for each country. The overall move toward more health-oriented products has increased requests for natural colors. When using natural colors, ascorbic acid has to be added to improve the stability (better resistance to light). A positive side-effect of using ascorbic acid is a reduced incidence of can corrosion. The disadvantages of natural colors are the high cost, the complication of the manufacturing process, and the poor stability. (*See* **Ascorbic Acid**: Properties and Determination.)

Antioxidants

Antioxidants prevent reactions, which destroy aromatic substances. The most common antioxidant is ascorbic acid, i.e., vitamin C. Sulfur dioxide, named above as a preservative, is also used as an antioxidant. (*See* **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants.)

Other Additives

Emulsifying agents, stabilizing agents, and thickening agents are all used to ensure that the contents of the drinks remain evenly distributed. The fat peel oil from oranges, for instance, would otherwise form lumps and create a ring around the neck of the bottle. Since the oil is the carrier of the aroma and sometimes the coloring agents, the drink would soon become uneven in taste and look unattractive. Examples of stablizing agents and thickening agents are pectins, which are obtained from citrus fruits or apples, and alginates and carragheen, which are obtained from algae. (*See* **Emulsifiers**: Organic Emulsifiers; Phosphates as Meat Emulsion Stabilizers; Uses in Processed Foods; **Pectin**: Properties and Determination; Food Use; **Stabilizers**: Types and Function; Applications.)

Soft drinks generally contain no significant amounts of protein, fat, fiber, or vitamins. The small amounts of minerals (calcium, iron, magnesium) and trace elements (copper, manganese, zinc, fluoride) that may be naturally present will vary depending on the local water supply. However, some soft drinks now contain added vitamins (C, niacin, B_6, B_{12}, biotin, pantothenic acid, and folic acid) and/or increased levels of potassium (from added juice).

Table 3 Calorie, nutrient, and ingredient content of major types of soft drinks

Flavor types	*Calories*	*Carbohydrates ($g\,ml^{-1}$)*	*Total sugars ($g\,ml^{-1}$)*	*Sodium ($mg\,ml^{-1}$)*	*Potassium ($mg\,ml^{-1}$)*	*Phosphorus ($mg\,ml^{-1}$)*	*Caffeine ($mg\,ml^{-1}$)*	*Aspartame ($mg\,ml^{-1}$)*
Regular								
Cola	0.4–0.5	0.10–0.12	0.10–0.12	0–0.08	0–0.05	0.11–0.21	0.08–0.13	0
Caffeine-free cola	0.4–0.5	0.10–0.12	0.10–0.12	0–0.08	0–0.05	0.11–0.21	0	0
Cherry cola	0.4–0.5	0.10–0.12	0.10–0.12	0–0.04	0–0.03	0.13–0.15	0.03–0.13	0
Lemon–lime (clear)	0.4–0.5	0.10–0.12	0.10–0.12	0–0.15	0–0.01	0–0.003	0	0
Orange	0.5–0.6	0.11–0.14	0.11–0.14	0.04–0.12	0–0.05	0–0.17	0	0
Other citrus	0.3–0.5	0.08–0.14	0.08–0.14	0.03–0.14	0–0.33	0–0.003	0–0.18	0
Root beer	0.4–0.5	0.10–0.14	0.10–0.14	0.01–0.17	0–0.05	0–0.05	0	0
Ginger ale	0.3–0.4	0.08–0.11	0.08–0.11	0–0.08	0–0.01	0–trace	0	0
Tonic water	0.3–0.4	0.08–0.10	0.08–0.10	0–0.03	0–0.01	0–trace	0	0
Other regular	0.4–0.6	0.10–0.15	0.10–0.15	0–0.12	0–0.06	0–0.26	0–0.12	0
Juice added	0.4–0.6	0.10–0.14	0.10–0.14	0–0.06	0.08–0.33	0–0.21	0	0
Diet								
Diet cola	< 0.03	0–0.003	0	0–0.17	0–5.0	0.07–0.16	0–0.16	0–0.53
Caffeine-free diet cola	< 0.03	0–0.003	0	0–0.2	0–10.0	0.07–0.16	0	0–0.53
Diet cherry cola	< 0.03	0–< 0.001	0–trace	0–0.02	1.5–5.0	0.07–0.11	0–0.13	0.50–0.52
Diet lemon–lime	< 0.03	0–0.003	0	0–0.26	0–6.9	0–trace	0	0–0.53
Diet root beer	< 0.06	0–0.013	0	0.11–0.28	0–3.0	0–0.05	0	0–0.58
Other diets	< 0.2	0–0.05	0–0.05	0–0.27	0.3–10.1	0–trace	0–00.19	0–0.57
Club soda, seltzer, and sparkling water	0	0	0	0–0.27	0–0.5	0–0.003	0	0
Diet juice added	< 0.1	0.003–0.017	0.003–0.017	0–0.06	0–0.3	0–0.17	0	0.38–0.53

Table 3 shows the calorie, nutrient, and ingredient content of major types of soft drinks.

Sodium and potassium values do not include the levels contributed by water, which will vary depending on geographic location and season. Most soft drinks are very low in sodium, and some are sodium-free. (*See* **Potassium:** Properties and Determination.)

See also: **Ascorbic Acid**: Properties and Determination; **Caramel**: Properties and Analysis; **Carbohydrates**: Classification and Properties; **Citrus Fruits**: Types on the Market; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Glucose**: Properties and Analysis; **Lactose**; **Potassium**: Properties and Determination; **Preservatives**: Classifications and Properties; **Saccharin**; **Sensory Evaluation**: Aroma; Taste; **Sodium**: Properties and Determination; **Sucrose**: Properties and Determination; Dietary Importance

Further Reading

Colonna WJ, McGillivray T, Samaraweera U and Torgeson T (1996) *Proceedings Conference of The Sugar Processing Research Institute*.

Henkel J (1999) *Sugar Substitutes: Americans Opt for Sweeteners and Lite*. USA: FDA.

Production

C Matthews, NET International, Gloucester, UK

Introduction

Soft drinks have been consumed, in one form or another, for thousands of years from the first time it was discovered that the addition of acid materials such as lemon juice and/or the incorporation of only 5% of alcohol from fermentation of grape and other juices resulted in the preservation of dietary water. Although milk, wine, and beer have all been around for well over a thousand years, and tea and coffee for many centuries, carbonated soft drinks are approximately 200 years old, dating from the commercial production of seltzers and sodas in the late 1700s. Revitalization of the squashes, syrups, or concentrated still soft drinks market has been facilitated in the last few decades by the introduction of aseptic packaging techniques, enabling continuous filling of juices and ready-to-drink squash-type products into foil laminates, pouches, and cartons. Some would argue, with justification, that this is a completely new sector of soft drinks made possible by advances in production and packaging technology.

The production of soft drinks falls into three main areas of activity:

1. Supply, handling and treatment of ingredients (including the water supply).
2. Blending of these ingredients and the processing of the blends/combinations so achieved.
3. Packing or packaging of these blended ingredients into the finished product for distribution and sale (this may involve a further processing step, as in the case of aseptic filling or dilution, and carbonation, as with carbonated soft drinks).

Supply, Handling, and Treatment of Ingredients

Most ingredients will be supplied in a user-friendly and stable form by the ingredient manufacturers or suppliers, with detailed specifications and usage instructions. During the time these ingredients are held at the site of the soft drinks manufacturer it will simply be necessary for them to be stored according to the storage instructions of the supplier and used within shelf-life. Some of the storage conditions and shelf-lives of usually encountered raw materials will be found in **Table 1**. (*See* **Storage Stability:** Parameters Affecting Storage Stability.)

It is normally the responsibility of the quality control department to approve the quality of all incoming materials and insure correct rotation for use prior to expiry of shelf-life.

This leaves the main ingredient in soft drinks – water – which requires special treatment to insure sensory, microbiological, and physical acceptable quality. The chemical quality of water may be characterized as follows.

1. Appearance free from sediment, color, and cloud.
2. Taste free from taint or materials such as chlorine, hypochlorite, or nitrates which are capable of reacting with other components or materials (e.g., cans).
3. Free from toxins of any description.
4. Free of water hardness which can destabilize fruit colloidal suspensions via the calcium and magnesium which cause water hardness. (*See* **Quality Assurance and Quality Control.**)

This is achieved in a soft drinks production plant in a number of ways and the following list comprises the main parameters covered in the specification (and

Table 1 Storage conditions of soft drink raw materials

Material	*Storage conditions*	*Recommended shelf-life*
Fruit		
Frozen	Below −18 °C	2–3 years
Preserved	Ambient	3–9 months
Aseptic	*c.* 4 °C	Up to 6 months
Sugar syrup	Dependent on concentration	
	Some may be stored at ambient	48 h
Glucose syrup (GS)/high-fructose GS	35 °C	48 h–1 week
Citric acid solution	Ambient	1 week
Volatile and natural flavoring materials	4 °C	3–6 months
Nature-identical and artificial flavoring materials	Dark UK; ambient	6–12 months
Water	Ambient	On demand

hence treatment, where necessary) of water: appearance; pH; total dissolved solids; total hardness; alkalinity; nitrogenous compounds; chloride; organic content; microorganisms; phosphate; silicates; trace metals; chlorinated compounds.

Water Treatment

Chemical coagulation is widely used in soft drink production plants as a means of removing unwanted impurities. The incoming water passes into a reaction vessel where coagulating chemicals are added. The reaction products and impurities are precipitated and form a gelatinous sludge through which the treated water flows to a clear zone at the top of the vessel. Since small particles of the flocculant will travel with the treated water, it is normal practice to pass water treated in this way through a sand-bed filter. A typical system of this type is illustrated in Figure 1. (*See* **Water Supplies**: Water Treatment.)

The types of coagulants used are ferrous sulfate, aluminum sulfate, and Lapofloc PAC (a solution of polyaluminum chloride).

Chlorination is the next stage in the treatment of water. Free chlorine is usually used as the sterilizing agent at a level of about 6–8 $mg\,l^{-1}$ of chlorine after the sand filter to insure completion of the reaction. Either chlorine gas or hypochlorite solution may be used. Where ferrous sulfate is employed as the coagulant, free chlorine oxidizes the ferrous salt to ferric salt and is usually added to the coagulation tank where the residence time may be included as part of the contact time necessary for sterilization. Where aluminum sulfate is used, the chlorine is usually added after the sand filter and residence time in the coagulant tank is not part of the contact time.

Organic matter, sulfites, nitrites, and ammonia (or its compounds) absorb or react with chlorine and the bactericidal action of chlorine will not be effected until after these reactions have taken place. Chlorine is best added to the coagulation tank or immediately after, since contact with raw, untreated water containing humic organic substances produces trihalomethanes (THMs) which would result in an unpleasant taint to the water if present in quantities in excess of 100 $\mu g\,l^{-1}$. Via the chlorination process all impurities will be oxidized, removing all taste and odor, including those due to phenol and its derivatives.

The next stage in the process is the dechlorination of the water after sterilization by passing the water through an activated carbon filter which is housed in a vessel similar to that of the sand filter (Figure 1). The quantity of carbon will be selected to achieve a contact time of approximately 5 min with the water, removing all of the chlorine and any remaining organic molecules. Carbon filters are fitted with steam injection at the base to allow sterilization when required to reduce the gradually increasing microbiological contamination from its use.

Quality control on a continuous basis of all aspects of the process must be effected to insure optimum performance of the process. In particular it must always be able to indicate an imminent breakdown of the carbon bed when it requires regeneration.

The above process is the classical procedure for water treatment. There are, however, other methods of either total treatment or treatment for specific requirements of the water. They are as follows:

- Ion exchange will perform softening; dealkalization; nitrate removal; organic compound removal.
- Reverse osmosis will effect removal of high levels of dissolved solids, resulting in a water which may be used directly or further treated for carbonated soft drinks production.
- Ultrafiltration removes colloidal substances and may be substituted for the polishing filter in a conventional process.
- Alternative sterilization processes include ultraviolet light, ozonization, and micropore filtration.

Blending and Processing of Key Ingredients in Soft Drinks

There are two basic methods of manufacture of soft drinks; these are the single batch production and continuous production. Batch manufacture is the process where a tank with a capacity of up to 25 000 may be employed to manufacture the product and/or intermediate syrup to feed the filling line over a period of time. Continuous production utilizes two tanks of only 50–100 l each, with a computer, continuously making one batch after another in a type of tandem operation. The main advantages of the latter system are associated with line problems and interuptions, particularly where a heat treatment is part of the process, since only the 50 or 100 l of product already manufactured will be affected, and may easily be 'ditched' without significant economic consequences. In the case of batch manufacture, however, the 25 000 l of product or intermediate syrup will either remain under nonoptimum conditions or, at worst, continue to be heat-processed (with a return to the batch tank) for the duration of the line downtime.

In each case the steps and safeguards it is necessary to take in the production process are the same and will be covered in the same way. The schematic representation of the production of a soft drink is shown in Figure 2. It can be seen that the processing required for the production of a soft drink involves blending the individual ingredients with heat treatment and/or homogenization where necessary. In reality all ingredients are not added separately to the main blending vessel, since this would cause significant logistical problems for items that are difficult to handle, such as solid sugar (dissolution problems) or thickeners or stabilizers (localized viscosity build-up, lack of powder wetting). There are basically four streams of ingredients:

1. Colors, flavorings, etc.
2. Sugars (i.e., all carbohydrates).
3. Fruit materials.
4. Citric acid, stabilizers, emulsifiers, etc. (this stream utilizes a premix blending vessel with a significantly increased blending power-to-volume ratio for difficult-to-dissolve/disperse ingredients).

Many soft drink production sites are now converting to tanks located on load cells so that the vessel may be tared and ingredients weighed directly into them.

If liquid carbohydrates are being used, these may be added directly from their optimum storage position to the main blending vessel via an inline sieve or

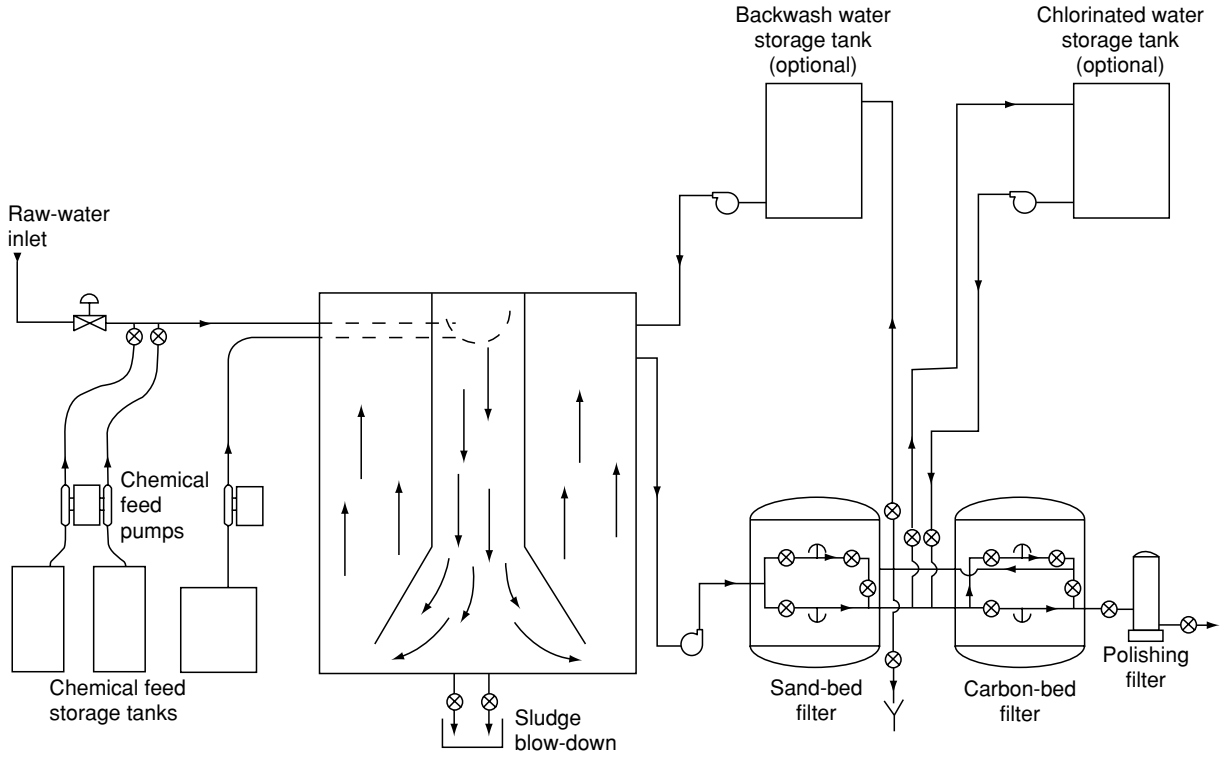

Figure 1 Typical water treatment process in the production of soft drinks.

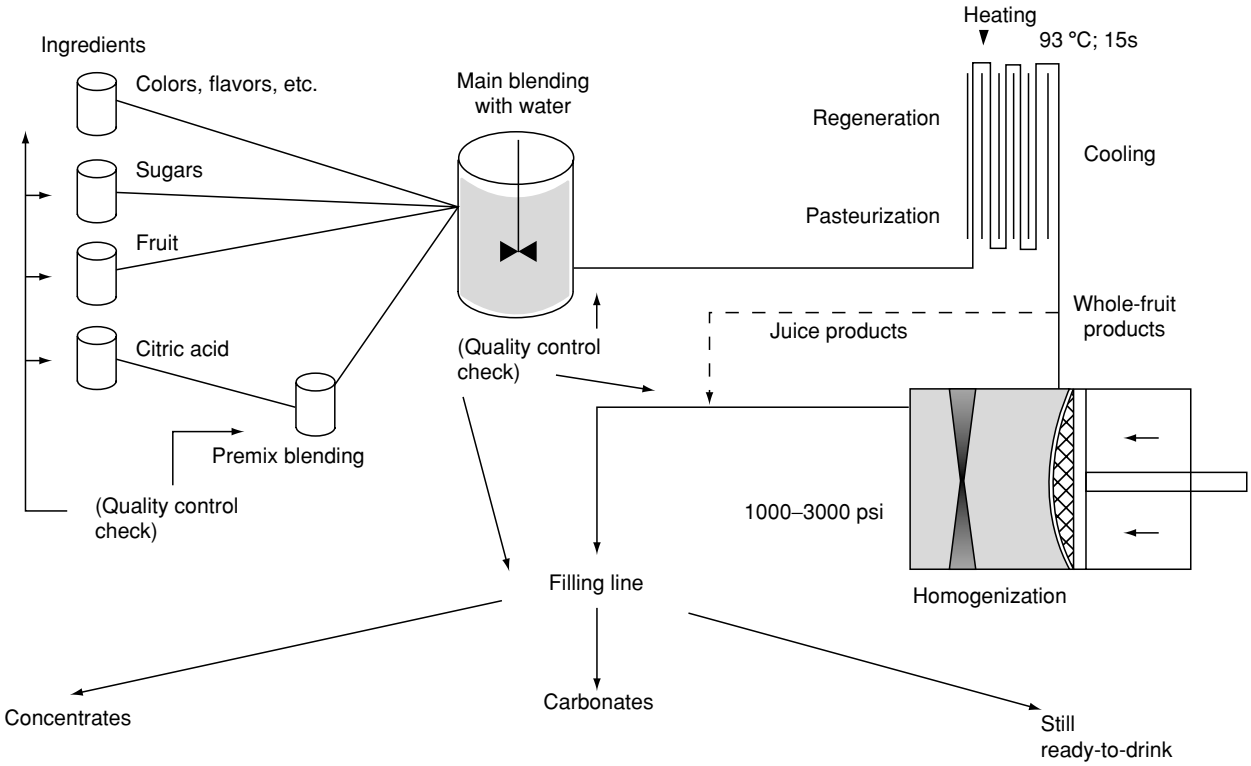

Figure 2 Schematic representation of the production of soft drinks. See Figure 3 for individual key stages.

filtration system to remove any small particles. If solid sugar is used, a simple syrup must be prepared of the sugar and water at approximately 50 °C to aid dissolution; this is then cooled before addition to the main blending vessel. All other ingredients are added to the carbohydrate syrup in the blending vessel in an order which takes into account their chemical reactivity with other ingredients. For example, sodium benzoate must always be added prior to the citric acid, since benzoic acid is only sparingly soluble in acid solution. Certain clouding agents and stabilizers must not be added in close proximity to a flavoring that contains a high level of an alcohol as solvent, since they will interreact.

With a high level of carbohydrate or thickening agent in the syrup in the main blending vessel, it is very easy to entrain air with too much vigorous agitation, and this would result in fobbing during filling, particularly of carbonated products. As long as agitation has not been excessive, then 2 h standing after mixing, with gentle agitation to keep all ingredients evenly dispersed, will be sufficient to allow entrained air to escape. This gentle agitation should be continued for the duration of any pumping or direct bottling so that particulate materials, such as fruit comminute, are evenly dispersed throughout the liquid.

Homogenization

This process is used for many fruit-containing concentrates as part of the finished beverage manufacture enroute to the bottling line and for intermediate syrup processing for some fruit-containing carbonates and ready-to-drink still beverages. The process is a means of reducing the particle size of particulate matter (e.g., fruit comminute) by means of energy input resulting in a more stable suspension and hence better appearance. The conditions which are normally used are 1000–3000 psi at ambient temperature.

Carbonation

The simplest type of carbonator is a pressure vessel in which the liquid and the carbon dioxide are allowed to stay in contact with each other. There are three basic ways to achieve this: (1) vessel partly filled with liquid and pressurized with carbon dioxide; (2) liquid falling through carbon dioxide in a pressurized container; (3) carbon dioxide bubbling through liquid in a pressurized vessel or pipeline. All carbonators operate on one or more of these principles.

One of the essential requirements of a production carbonation system is control over the degree of saturation of the carbonated product, i.e., how close to

the maximum possible volume of carbon dioxide is allowed to dissolve. An uncontrolled release of gas from solution is known as fobbing. Partially saturated solutions are more stable than saturated ones owing to the additional pressure available over the equilibrium pressure. This additional pressure is known as the overpressure and is used in the control of the process. When the necessary degree of overpressure may be maintained over a range of temperature and carbonation, this is known as variable saturation. The ready-to-drink, dilute product may be carbonated by a reduction in temperature and injection of carbon dioxide via one or more of the above techniques, or the concentrated intermediate syrup may be blended with a stream of carbonated water on line to the filling apparatus, resulting in the correctly diluted, carbonated, finished product in the pack.

Hygiene

Throughout the production of all types of soft drinks it is essential that plant hygiene is maintained at the optimum level, whether chemical preservatives and/or heat processing are used.

Good manufacturing practice is a vital and integral part of insuring minimal microbiological contamination of the product and will only be achieved if attention is paid to certain areas of operation:

- Hygienic practices in receiving, handling, and processing materials.
- Tight specifications for clean raw materials.
- Hygienic attitudes and understanding in all personnel.
- Hygienic design of plant and equipment.
- Hygienic operations of suppliers (especially of fruit and carbohydrate materials).

Heat Processing

An examination of the different types of processing for concentrates, carbonates, and ready-to-drink soft drinks is shown in Figure 3. This shows clearly that, in addition to chemical preservation, blending, carbonation, and basic hygiene and quality control requirements, heat processing is a vital part of the production of soft drinks.

Although a manufacturer's main protection against microorganisms is the hygienic operation of the manufacturing plant and that of suppliers, it may still be necessary with certain microbiologically sensitive formulations to enlist the help of both chemical preservatives and heat processing. The use of heat during the processing of soft drinks is covered in outline here as it relates to the microbiological integrity of the product formulation. (*See* **Preservation of Food**.)

Heat processing, or pasteurization, may be subdivided into four main categories for soft drinks and/or material processing: (1) hot-fill; (2) in-pack (or tunnel) pasteurization; (3) flash pasteurization: (a) nonaseptic conditions or (b) aseptic conditions; and (4) microfiltration. The selection of the required/desired heat-processing conditions for any given soft drink, intermediate syrup/blend, or raw material is determined by a number of factors relating to the following:

- Presence of fruit materials and their origin, quality, and microbiological status.
- Presence of chemical preservatives – qualitative and quantitative factors apply.
- pH of the product.
- Solids content and hence water activity.
- Any other relevant processing conditions – holding of syrup at ambient/raised temperature, use of homogenization, filtration, etc.
- Type of packaging to be employed.
- Desired shelf-life.
- Microbiological conditions (and history) of the filling line and ancillary equipment to be employed. (*See* **Pasteurization**: Principles.)

Hot-fill This can only be used for finished still products or intermediate raw materials when being packed for storage and shipment. The product/material is heated in a heat exchanger to a temperature of 85–90 °C; filled into the pack (glass, cans, etc.) so that the temperature of the contents is in excess of 85 °C; sealed via capping or seaming; inverted for a minimum of 1 min; and cooled as quickly as possible. Any flaws in the capping/seaming will result in contaminated cooling water or air entering the package and this will cause spoilage.

The disadvantage of this system is the large amount of heat that is put into the product/material during the whole process, which results in flavor, color, and clouding degradation on storage, producing a reduction in total shelf-life on the product/material.

In-pack (or tunnel) pasteurization In its simplest form, this type of pasteurization involves immersing the filled, capped, or seamed pack of either still or carbonated product in a tank filled with hot water at 60–90 °C in a removable basket and holding at the selected temperature for the required period of pasteurization holding time. In reality the packs are placed in water at a temperature lower than that required and the temperature of the water is raised until the specified temperature has been reached.

Examples of pasteurization conditions for this type of processing are:

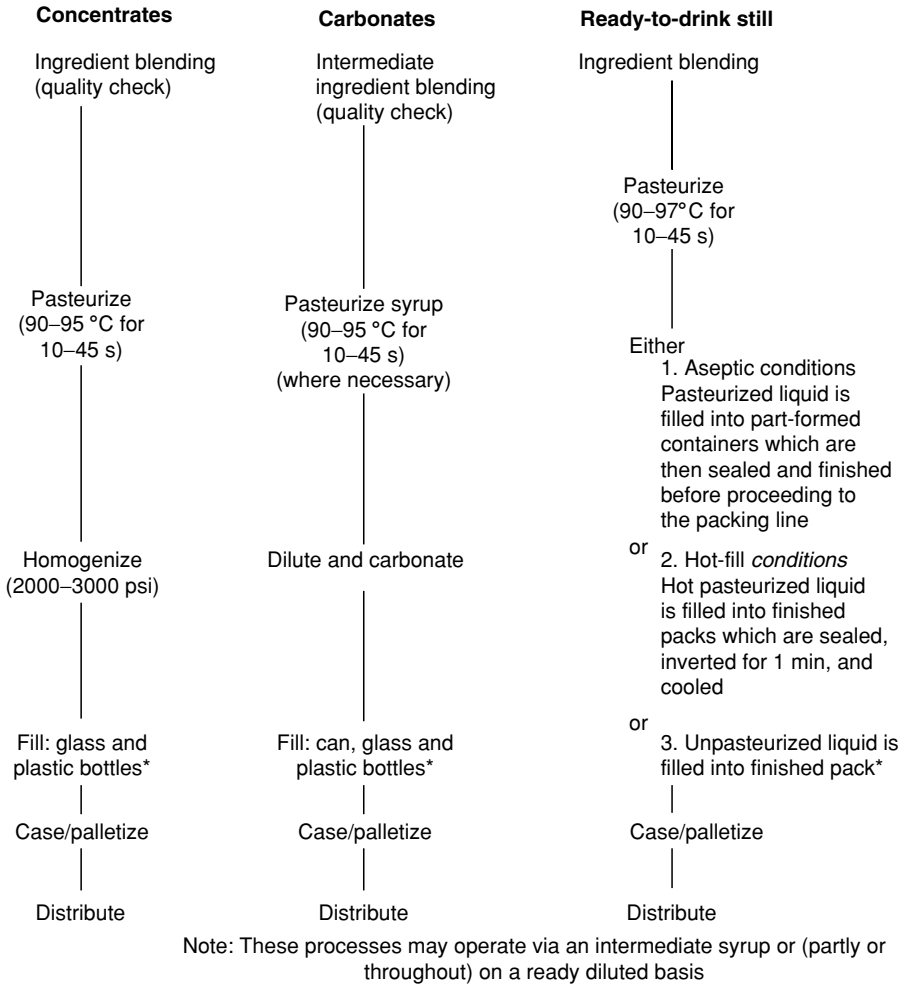

Figure 3 Key stages in the processing of soft drinks.

- 63 °C for 10 min: typically used for shandies
- 70 °C for 10 min: typically used for squashes
- 75 °C for 15 min: typically used for susceptible products such as high-blackcurrant-juice products.

Tunnel pasteurization may be considered as a continuous version of in-pack pasteurization.

After the product has been filled, packed, and seamed, the packages are transported through a tunnel made up of different heat zones by means of water of different temperatures being sprayed over the packages, which gradually increase in temperature to the selected pasteurization level where they are then held for the desired time (holding time) before being cooled slowly to ambient temperature. Processing time is approximately 40–50 min, resulting in very effective pasteurization at relatively high cost (it is energy-intensive and filling has to proceed at the speed of the pasteurizer).

Flash pasteurization

Nonaseptic conditions This system may be used for carton, polyvinylchloride, polyethylenetetraphthalate (PET), and all other nonglass/metal drinks packaging materials, and is suitable for both still and carbonated beverages. The heat exchanger is connected in line with the product mixing tank in the case of still products and the intermediate syrup tank in the case of carbonated products. The syrup or finished product is heated very quickly to the specified pasteurization temperature, held for the specified time (usually quoted as seconds), and cooled rapidly to the filling temperature. The final drink may also be passed through the heat exchanger in this way *en route* to the carbonator if required.

The advantage of this type of pasteurization is that the total heat input to the product is only a fraction of that required for in-pack pasteurization and results in considerably reduced flavor, color, and shelf-life

effects. The disadvantage is that contamination may occur during filling and/or capping and rigorous standards of hygiene and microbiological control must be adhered to by fully trained and committed staff. It may even be necessary to fill some products/packs in enclosed areas with overpressure sterile air rooms.

Aseptic conditions Before a discussion on the types and methods of aseptic flash pasteurization, it is necessary to consider exactly what this means in real terms.

Asepticity is sometimes referred to as commercial sterility, but it is necessary to recognize the important microbiological differences between the two types of processing employed.

With sterility it is realistic to expect that all measurable organisms have been killed by the process, which is almost always carried out at temperature conditions well in excess of 100 °C in hermetically sealed containers, thus insuring the microbiological integrity of the whole pack (product plus packaging).

In aseptic packing the product is heat-treated separately from the container and maintained in positive air pressure conditions (thus excluding any potentially contaminating organisms) until filled and sealed into the selected pack, which has been separately treated to reduce any microbiological contamination on its surface. Examples of aseptic conditions are those which apply to the Tetra Pak process where the product is flash-pasteurized at 96 °C for 13 s and filled into cardboard previously treated with hydrogen peroxide; this is formed into a tube and the individual pack seams are carried out through the product.

With aseptic processing/packaging it is generally recognized that, independently of any pack integrity problems, the normal failure rate will be of the order of one pack in every 10 000 produced. A better rate than this is rarely achieved and aseptic conditions are sometimes quoted with failure rates as high as one in 3000, as in the case of carbonated products packed in PET containers.

See also: **Pasteurization**: Principles; **Preservation of Food**; **Quality Assurance and Quality Control**; **Storage Stability**: Parameters Affecting Storage Stability; **Water Supplies**: Water Treatment

Further Reading

Cullen KW, Ash DM, Warneke C and de Moor C (2002) Intake of soft drinks, fruit-flavored beverages, and fruits and vegetables by children in grades 4 through 6. *American Journal of Public Health* 92(9): 1475–1478.

Fried EJ and Nestle M (2002) The growing political movement against soft drinks in schools. *Journal of the American Medical Association* 288(17): 2181.

Heath HB (1981) *Source Book of Flavours*. Westport: AVI.

Hicks D (1990) *Non-Carbonated Fruit Juices and Fruit Beverages*. Glasgow: Blackie.

Matthews AC (1991) *Food Flavourings*. Glasgow: Blackie.

O'Donnell K (1990) *Formulation and Production of Carbonated Soft Drinks*. Glasgow: Blackie.

Microbiology

M Stratford and C J Capell, Unilever Research, Sharnbrook, UK

Background

'Soft drinks' is a term often used to describe all beverages with the exception of alcoholic or hot beverages. This includes a vast array of products ranging in composition from mineral waters, sports drinks, diet formulations, colas, mixers, and tonics to fruit juices. Some soft drinks are sparkling (carbonated), whereas others are still. Some are manufactured with a designated shelf-life greater than a year, whereas others are sold for immediate consumption (freshly squeezed juices). As a consequence of this variation in composition and packaging, the microbiology of soft drinks also shows considerable variation. Some soft drinks form an environment so hostile, or so deficient in nutrients, that the microflora is nearly nonexistent, and spoilage is consequently rare. Other soft drinks, such as fruit juices of a higher pH (**Table 1**) form an environment ideally suited to rapid proliferation of yeasts, molds, and bacteria. For the purposes of this article, the average soft drink will be considered to have a pH between 2.5 and 3.5 (**Table 1**), sugar at 5–10%, 1–3 $g\,l^{-1}$ citric acid (or titratable acidity equivalent), and fruit juices or vegetable extracts sufficient to comprise a source of minerals and vitamins for microbial growth.

In theory, the presence of microbes in a food product can have any one of three possible effects. It can be a positive attribute, growth of specific microbes creating desirable organoleptic changes, as in

Table 1 pH range and acids prevalent in various soft drinks and fruit juices

Beverage	*pH range*	*Acids*
Apple juice	2.9–4.5	Malic, citric
Orange juice	2.6–4.3	Citric, malic
Blackcurrant juice	3.0–3.8	Citric
Grapefruit juice	2.9–3.6	Citric
Lemonade	2.8–3.2	Citric
Cola	2.5–2.8	Phosphoric

fermented foods or alcoholic beverages. It can be a negative attribute, causing illness in the consumer or spoilage of the food, or it can be entirely neutral, the presence of microbes having no effect whatsoever on the food or on the consumer. In soft drinks, microbes are almost never beneficial, a possible exception being the oriental acidic tea, Kombucha, fermented by yeasts, molds and acetic acid bacteria. In the great majority of instances, the presence of microbes in soft drinks is entirely neutral. Most bacteria are sensitive to the acidity of soft drinks, whereas the majority of yeasts and molds may be suppressed by a lack of oxygen within the soft drinks packaging (Table 2). Any such organisms tend to arrive in bottles of soft drinks in low numbers and have no impact whatsoever on the beverage. Most will die in this environment within a few hours to a few weeks, depending on the organism.

A small minority of microbes, however, find the soft drink environment suitable for proliferation and growth to substantial populations. These are the soft-drinks spoilage flora (Table 2). Although the occasional microbe per liter of beverage has little impact, growth to 10^5 cells per milliliter results in visible spoilage, and 10^6–10^8 cells per milliliter are sufficient to cause very considerable changes to the chemistry of the beverage. Such changes are almost always organoleptically unacceptable and in themselves constitute spoilage.

Manufacturing Process of Soft Drinks

Process

Fruit-juice manufacture typically occurs close to the area of production. Fruit is processed to remove leaves, stems, etc., washed, milled, and crushed. The juice is usually extracted from the crushed pulp by pressure. Single-strength juice is bulky and is often concentrated at the site of production. Juices are heated to near 100 °C, a vacuum is applied, and water is removed by evaporation. Volatile components are recovered and added back after cooling. To ensure microbiological stability, concentrates may then be frozen, preserved, or aseptically filled into drums.

Soft-drinks manufacture (Figure 1) essentially involves dissolving weighed amounts of flavors, juices, acids, antioxidants, and sugars into water. This is then dispensed into the appropriate packaging. Sugar enters the syrup room of the factory either as granules, which are dissolved in hot water, typically 85 °C/30 s or as 67 °Brix syrup, known as 'simple sugar syrup.' Other components of the soft drink, acids, flavors, fruit juices, vegetable extracts, antioxidants, etc., are weighed out and dissolved in water. These are added to the simple sugar syrup, forming the final syrup at 40–50 °Brix. The final syrup is piped from the syrup room to the proportioning pump in the filling hall (Figure 1). Here, it is proportioned with water to form the product, typically, 7–10 °Brix. This is filled into bottles/cans/tetrapaks at the filler, which are then sealed at the capper.

Sources of Infection

Microbes may enter soft drinks and their production facilities from a number of sources. These may directly infect products or may colonize various sites within the factory and subsequently cause infection from these sites. Entry to the factory is most common on raw materials, returned bottles, via insects or aerial vectors.

Aerial contamination obviously depends on the microbial loading and composition in the air. Nearby sources of microbes greatly increase the microbial load, e.g., breweries, dust sources (harvesting or building work), fruit storage, or processing of fruit or vegetables. Fresh fruit are routinely heavily surface-contaminated with yeasts and molds, 10^3–10^6 per gram. Factories processing fruits will inevitably become contaminated with the molds associated with those fruits' spoilage flora. Ascomycetous molds produce conidiospores for dispersal and ascospores for survival. Conidiospores are often airborne; ascospores often are not, making the presence of the heat-sensitive conidiospores generally more common as airborne contaminants. Aerial contamination is highly seasonal, with mold spores predominating,

Table 2 Summary of the growth, spoilage, and safety significance of microbes in soft drinks

Microbe	*Growth*	*Inhibition*	*Spoilage*	*Safety*
Spoilage yeasts	+	−	Excess gas	Exploding bottles
Other yeasts	−	Low O_2	−	−
Spoilage molds	+	−	Mycelium	Possible mycotoxins
Other molds	−	Low O_2	−	−
Vegetative bacterial pathogens, e.g. *Salmonella* spp., *E. coli* VTEC (enteropathogenic *E. coli*)	−	Low pH	−	Food poisoning, in fresh, unpasteurized, fruit juices
Acid-tolerant bacteria	±	Low O_2	Taints, slime	−
Most bacteria	−	Low pH	−	−

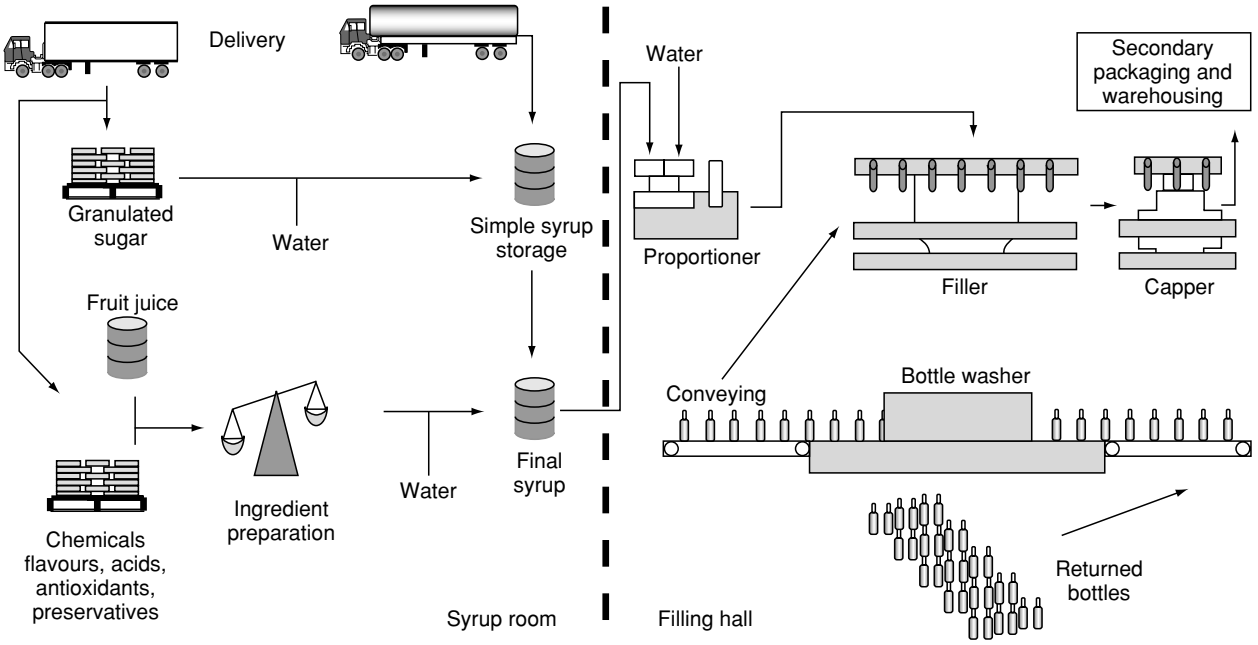

Figure 1 Schematic diagram of a soft-drinks production facility. In the syrup room, sugars are dissolved in water to form simple syrup, and juices and flavors are added to form the final syrup. This is pumped into the filling hall, diluted with water by the proportioning pump, and filled into bottles. Bottles are then capped, packaged, and distributed. Tunnel pasteurization is applied to capped bottles or 'in-line' pasteurization between the proportioner and filler.

and is probably the most significant cause of mold spoilage. (*See* **Microbiology**: Classification of Microorganisms.)

Insects are becoming increasingly recognized as a vector for yeasts. Many insects carry yeasts and acid-tolerant bacteria between flowers, actively consume yeasts as part of their diet, and benefit from yeast colonization of their digestive systems. Insect frass, notably from fruit flies (*Drosophila* spp.), is particularly rich in soft-drink spoilage yeasts.

Returned bottles, when brought into the factory, form a potent source of microbial infection of soft drinks. Returned bottles are frequently cap-less and unwashed, often containing a residue of soft drink. Storage at ambient temperature allows selective growth of spoilage microbes in the residue up to high levels and formation of stationary-phase survival mechanisms, such as heat-resistant ascospores by yeasts and molds. Stringent bottle-washing removes almost all residues, but fungal debris and viable ascospores can persist through the whole process. Fruit-flies, attracted to the sugary residues in returned bottles, can also spread infection around the factory.

The microbial loading of raw materials is normally low where materials have been selected from good quality sources. Sugar granules, 67 °Brix syrups, or fruit-juice concentrates are a potential source of contamination. Fruit juices are often heavily contaminated; citrus fruits are especially associated with mold spoilage, as the fruit defense is most susceptible to pectinolytic molds. Pasteurization of the single-strength juice or the concentration process should result in a near-sterile product, except for rare heat-resistant molds. *Byssochlamys* spp. and *Talaromyces* spp. have been the cause of spoilage in pasteurized fruit juices. The ascospores of these molds are relatively uncommon, but significant growth of the mold on the raw material must have occurred for the presence of problematic numbers of ascospores. Similarly, mold spores in small numbers may survive the sugar refining process. Both fruit juice concentrates and sugar syrups are more likely to acquire an osmophilic flora through mishandling after cooling. Owing to the low water activity, only osmophilic yeasts and, more rarely, molds can grow, albeit very slowly in sugar syrups. Other ingredients, such as water, flavors, acidulants, colorings, antioxidants, and preservatives, are also possible sources of spoilage microorganisms.

Build-up or colonization of a spoilage flora within the factory is probably the most significant source of infection. It has been estimated that 95% of yeast infections of soft drinks are caused by poor plant hygiene. Soft-drink manufacture is generally a very high-volume throughput operation, with up to $36\,000\,l\,h^{-1}$ passing through each filling line. Little time is therefore allowed for growth of microbes within the bulk volume of liquid, before pasteurization or addition of preservatives. Microbial build-up is therefore most significant in dead spaces or on

surfaces proximate to the product stream. Areas most at risk include the proportioning pump, filler and filler heads, and capper. Other areas of concern include old, dried splashes of fruit juice or soft drink, the syrup room, particularly where ingredients have been spilled or allowed to become damp, and drains into which sugars are washed. (*See* **Contamination of Food.**)

Microbial Stability

Microbial stability can be achieved by preventing viable microbes entering the product, (aseptic filling), sterilization (pasteurization), or by factors preventing microbial growth in soft drinks (preservation). Pasteurization is the most effective method of eliminating microbes in soft drinks (Table 3). Vegetative yeast and mold cells are generally heat-sensitive, almost all being killed in 10 min at 60 °C. Pasteurization of fruit juices, such as orange juice, which are typically heavily contaminated with yeasts and molds, range from 10 to 15 min at 70–75 °C. Such pasteurization will not eliminate heat-resistant forms of molds, such as ascospores.

The microbial stability of soft drinks can be achieved most simply and effectively by 'in-pack' sterilization, downstream from the capper using a tunnel pasteurizer (Table 3). This insures absence of microbes in the bottle and cap as well as the product. Alternatively, products may be sterilized 'in-line' using a plate pasteurizer and aseptically filled into sterile bottles, or products may be given a flash in-line pasteurization to lower the microbial loading and preservatives added to prevent the growth of surviving microbes. Fresh-pressed fruit juices that would normally spoil within 2 days may be preserved by chilled storage for a limited period of time, e.g. 2–3 weeks.

Some factors suppressing microbial growth are intrinsic to the composition of soft drinks, such as the phosphoric acid in cola or the low pH that inhibits or kills the majority of bacteria. Pressurized CO_2 in carbonated soft drinks has a pronounced antimicrobial effect, particularly against molds and bacteria. Fermentative yeasts are more resistant, particularly *Brettanomyces* spp., but growth can be inhibited at high carbonation. The lack of nutrients in many soft drinks, such as vitamins, phosphate, or a usable nitrogen source, is also an aid to preservation. It has been noted that soft drinks improved by the addition of real fruit juices often became increasingly prone to microbial spoilage.

Table 3 Common methods of achieving microbial stability in soft drinks

Method	*Comment*
Tunnel pasteurization	Sterilizes product and packaging together
In line, plate pasteurization	Sterilizes product only, then requires aseptic or hot filling
Flash pasteurization + preservatives	Flash lowers the microbial load, and preservatives inhibit survivors
Frozen/chill storage	Inhibition or delay of microbial growth

Preservatives are added to soft drinks to kill or prevent the growth of microbes. The preservatives commonly encountered in soft drinks are sorbic acid and benzoic acid. Less frequently found are sulfites (SO_2), parabens, and dimethyldicarbonate (DMDC). Preservative usage and permitted concentrations vary greatly, depending on the legislation in force where soft drinks are produced and consumed. Sorbic and benzoic acids function far better as preservatives at lower pH. Resistance to preservatives has been reported to the greatest extent in *Zygosaccharomyces* spoilage yeasts, but also occurs to a lesser extent in certain molds and bacteria, such as *Gluconobacter* (*Acetomonas*).

Characteristic of Soft-drinks Spoilage

Soft drinks may be spoiled by yeasts, molds, or bacteria. Yeast and mold spoilage predominates, being favored by the low pH and high sugar/assimilable nitrogen ratio. Spoilage of soft drinks may be clear and easily detectable by a number of visual or organoleptic symptoms. However, it is by no means uncommon not to isolate viable microbes from a spoiled product. It appears that spoilage organisms often grow and cause the symptoms of spoilage before dying. This may be due to a lack of nutrients or oxygen, or to the formation of toxic byproducts such as ethanol. Consequentially, it is not always possible to identify the causative agent of spoilage.

Yeast growth in excess of 10^5 cells per milliliter results in a visible cloud or haze. Some spoilage yeasts secrete proteolytic or pectolytic enzymes that can destroy the natural hazes of fruit juices. Spoilage yeasts may be flocculent, forming particulates, whereas others form a surface skin or vellum, notably *Pichia membranaefaciens*. The most significant spoilage yeasts are highly fermentative, producing CO_2 up to high pressure, and consequently distorting and bursting the packaging (Figure 2). Fermentation also results in ethanol production, and yeast spoilage is often associated with a sweet alcoholic taste.

Mold spoilage of soft drinks is usually visual and obvious. Molds given more oxygen via loose seals, large headspace, or oxygen-permeable packaging tend to form surface pellicles and may sporulate. Mold growth in suboptimum conditions, low oxygen or preservatives, tends to form submerged mycelial

masses, often resembling small wispy balls of cotton wool. Mold growth can also eliminate clouds in citrus fruit juices. Taints and musty flavors may be produced.

Bacterial spoilage is usually associated with fruit or vegetable juices and the presence of oxygen. Spoilage is often associated with loss of astringency, souring, slime or ropiness in texture, turbidity, and off-flavors.

Figure 2 Ruptured can of a soft drink caused by yeast spoilage. The can was infected with *Zygosaccharomyces bailii* and incubated for 2 weeks at 25 °C. Cans were visibly swollen within 7 days.

Spoilage Microorganisms of Soft Drinks

Currently, some 800 yeast species, thousands of molds, and vast numbers of bacteria are recognized. The ability of microorganisms to exchange DNA often makes species identification of an organism causing spoilage to be of little value. In the 'forensic approach' to spoilage microbiology, Professor Davenport proposed a simplified taxonomy based on the behavior of microbes, rather than on their species names. Group 1 were the spoilage microoganisms, well suited to proliferate in soft drinks with the potential to cause spoilage from as few as one cell per container. Group 1 spoilage microbes for soft drinks were limited to a number of yeasts, listed in Table 4. All are generally osmotolerant, highly fermentative, vitamin-requiring and resistant to preservatives, such as acetic, sorbic, or benzoic acids. These yeasts are fortunately rare in soft-drink environments and do not form any part of the flora of a well-run soft-drink production facility.

Group 2 microbes, described as spoilage/hygiene, are capable of causing spoilage of soft drinks but only as a result of mistakes occurring during manufacture. This group, examples of which are listed in Table 4, includes yeast and mold species and acid-tolerant bacteria. These can be regarded as opportunistic spoilage organisms, ready to exploit errors such as under-dosing of preservatives, failure of pasteurization, hygiene failures, or ingress of oxygen. Such organisms are often present in soft-drink factories, usually in low numbers and restricted by good hygiene procedures.

Group 3 microbes are hygiene indicators and will not cause spoilage. Most yeasts found in soft drinks production and products are Group 3 organisms (Table 4). Their numbers can be used as a measure of the general hygiene of the facility, and their origin may indicate the origin of the contamination. For

Table 4 Yeasts associated with soft drinks and fruit juices

Group 1	*Group 2*	*Group 3*
Zygosaccharomyces bailii	***Candida parapsilosis***	***Aureobasidium pullulans***
Zygosaccharomyces bisporus	***Debaryomyces hansenii***	***Cryptococcus albidus***
Zygosaccharomyces lentus	***Pichia membranaefaciens***	***Cryptococcus laurentii***
Zygosaccharomyces rouxii	*Saccharomyces exiguus*	***Rhodotorula glutinis***
Saccharomyces exiguus	*Saccharomyces cerevisiae*	***Rhodotorula mucilaginosa***
Saccharomyces cerevisiae	*Candida krusei*	*Candida sake*
Brettanomyces bruxellensis	*Pichia anomala*	*Candida intermedia*
Schizosaccharomyces pombe	*Kloeckera apiculata*	*Candida zeylandoides*
Torulaspora delbrueckii		
Zygosaccharomyces fermentati		
Zygosaccharomyces florentinus		
Zygosaccharomyces microellipsoides		

Group 1 are spoilage yeasts, Group 2 are opportunistic spoilage microbes, and Group 3 are hygiene indicators. The most significant species are in bold. *Saccharomyces exiguus* and *S. cerevisiae* are listed in two groups in respect of their strain variability, abnormal strains causing spoilage.

example, the yeast-like mold *Aureobasidium pullulans*, or red-pigmented yeasts, *Rhodotorula glutinis* or *R. mucilaginosa*, can be used to indicate airborne dust contamination. The presence of such organisms in beverages is relatively common and not a cause for concern. Such organisms are usually present in very low numbers, are harmless to the consumer, and are unable to grow or cause spoilage. It has been estimated that up to a third of all commercial fruit juices contain a few viable yeasts.

Spoilage Yeasts

The most widely reported yeasts in soft drinks are the Group 1 spoilage yeasts, which include *Zygosaccharomyces bailii, Z. rouxii, Z. lentus*, and *Z. bisporus*. These are rarely found in soft drinks or production facilities, but when they do occur, spoilage is usually widespread and devastating. These yeasts share many physiological characteristics, being osmotolerant, preservative-resistant, and highly fermentative, and most are taxonomically closely related.

Zygosaccharomyces bailii (Figure 3) is the most preservative-resistant organism known, able to resist high concentrations of acetic acid and ethanol. Strains of *Z. bailii* can tolerate sorbic and benzoic acids at pH 4.0 at up to 1000 p.p.m. (Figure 4). The European statutory limit for these preservatives is currently 300 p.p.m.! Adaptation to preservatives by prior exposure can increase tolerance by some 50–100%. *Z. bailii* readily forms ascospores, although these are not markedly heat-resistant. *Z. bailii* grows well in sugar syrups and juice concentrates, containing 50–60% sugar, thus allowing easy infection of soft drinks via the raw materials. The infective dose of *Z. bailii* is low; this yeast is reputed to cause spoilage from as little as one cell per container. *Z. bailii* can cause spoilage of fruit juices, cordials and concentrates, and carbonated and still soft drinks. *Z. bailii* requires B-group vitamins and is less likely to spoil products lacking fruit juices or similar vitamin sources.

Zygosaccharomyces bisporus shares most of the characteristics of its close relative *Z. bailii*. *Z. bisporus* is much less frequently encountered than *Z. bailii*. Examination of the characteristics of several strains of *Z. bisporus* show much diversity; some strains being comparable with *Z. bailii* in terms of preservative resistance and osmotolerance, whereas other strains are more sensitive. Overall, *Z. bisporus* strains are less preservative-resistant than *Z. bailii* (Figure 4).

Zygosaccharomyces lentus is a newly discovered species, members of which had previously been mistakenly identified as *Z. bailii*. *Z. lentus* strains are often more preservative-resistant than *Z. bisporus*.

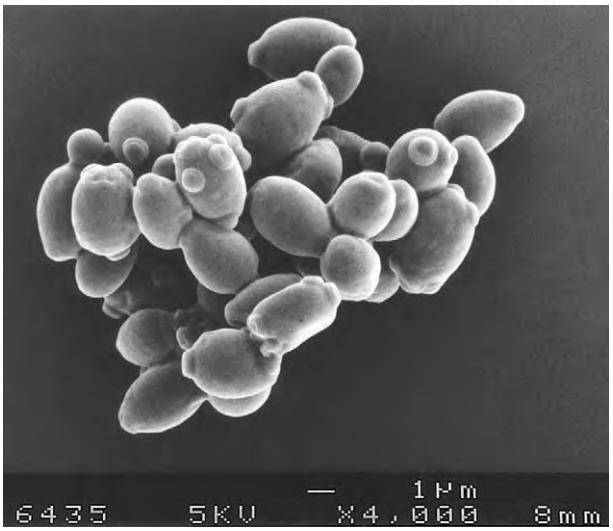

Figure 3 Scanning electron micrograph of the spoilage yeast *Zygosaccharomyces bailii* NCYC 1766. Cells are ovoid, 4 µm in length, with prominent surface bud scars.

All of the *Z. lentus* strains isolated to date are from spoiled foods, predominantly from orange and tomato products. *Z. lentus* strains are unusual in growing very slowly and having a low temperature range. Growth is poor above 25 °C, but these yeasts will grow at 4 °C and cause spoilage of refrigerated soft drinks and fruit juices.

Zygosaccharomyces rouxii is extremely osmotolerant being able to tolerate water activity down to 0.62 in fructose. Consequently, this yeast is probably the most significant spoilage organism in syrups and fruit-juice cordials and concentrates, but also spoils fruit juices and soft drinks. Like the other *Zygosaccharomyces* spoilage yeasts, *Z. rouxii* is highly fermentative, spoilage of concentrates often being detectable by bulging containers. *Z. rouxii* is also reported capable of growth from very low contamination levels. Direct comparisons show that *Z. rouxii* is significantly less preservative-resistant than *Z. bailii* (Figure 4) and shows significantly less resistance to pasteurization than most other yeasts. The rarer honey spoilage yeast, *Z. mellis*, shows very similar characteristics to *Z. rouxii*. Many *Z. mellis* strains have been shown by 26S rDNA sequencing to have been misidentified, in reality, *Z. rouxii* strains.

Saccharomyces exiguus, sometimes known as *Candida holmii* or *Torulopsis holmii*, is also closely related to the central *Zygosaccharomyces* spoilage family. *S. exiguus* strains are vigorous-growing, moderately osmotolerant, and highly fermentative, generally resembling *Saccharomyces cerevisiae*. Some strains of *S. exiguus* are capable of growth in extreme acid conditions and high levels of preservatives.

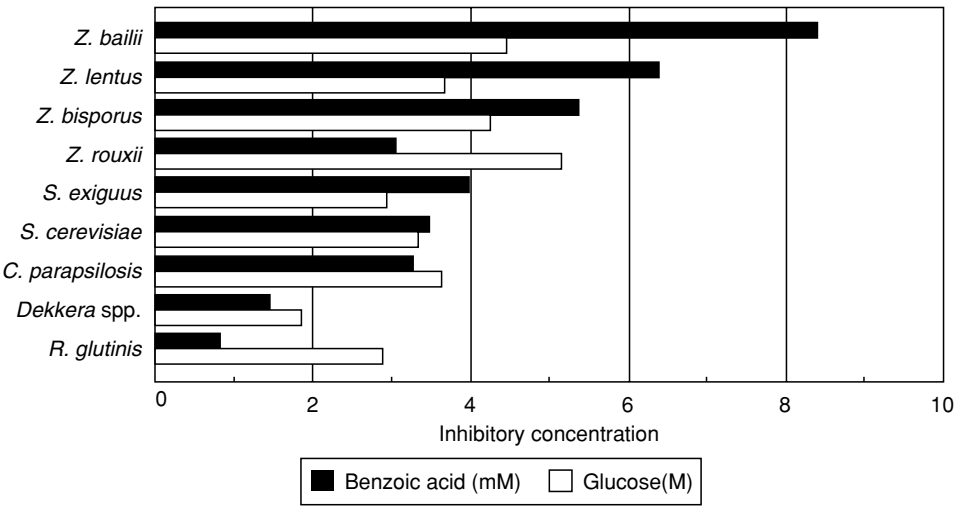

Figure 4 Bar chart showing the relative preservative resistance and osmotolerance of a number of spoilage yeasts (unpublished data from H. Steels, with permission). Inhibitory concentrations were measured in YEPD (yeast extract, peptone, glucose) broth, pH 4.0, at 25 °C. Data are the mean values of 28 strains of *Z. bailli*, 10 strains of *Z. lentus*, six *Z. bisporus*, nine *Z. rouxii*, five *S. exiguus*, 11 *S. cerevisiae*, seven *C. parapsilosis*, five *Dekkera/Brettanomyces* spp., and five *R. glutinis*.

Brettanomyces, and its ascospore-forming teleomorph, *Dekkera*, is a genus containing species noted for spoilage of low nutrient, carbonated beverages such as soda water, tonic, cola, or clear lemon. The most commonly encountered species are *Dekkera bruxellensis* and *D. naardenensis*, causing sediments and characteristic off-flavors and odors. These yeast species are characteristically sensitive to preservatives (Figure 4), high sugar, and heat, but are regarded as spoilage Group 1 yeasts owing their its resistance to CO_2 and spoilage of carbonated beverages.

Other spoilage yeast species are less commonly encountered, including other near-relatives of the *Zygosaccharomyces* spoilage family, *Z. fermentati. Z. florentinus, Z. microellipsoides, Torulaspora delbrueckii*, and the fission yeast *Schizosaccharomyces pombe*. These are all fermentative yeasts with moderate osmotolerance. Their resistance to preservatives is generally moderate to low. (*See* **Yeasts**.)

Spoilage Molds

Molds are, by and large, aerobic multicellular hyphal fungi, but a spectrum of phenotypes exist. Food spoilage molds are generally fast-growing opportunistic organisms. Fungi have a greater tolerance of low pH and low water availability than bacteria and are therefore most closely associated with foods with these characteristics. Their requirement for oxygen and sensitivity to carbon dioxide mean that mold spoilage may be effectively controlled by carbonation. Some molds can produce heat-resistant stages, including sclerotia, chlamydiospores, and ascospores, and are potentially capable of surviving normal pasteurization (Table 5). Although molds produce hyphal growth, they are able to grow well in liquid culture, either growing on the surface or completely submerged. Although molds are generally regarded as obligate aerobes, some do ferment sugars to ethanol and carbon dioxide; those best studied, the dimorphic yeast-like molds such as *Mucor* spp., are also associated with the spoilage of soft drinks. As these molds do not produce exceptionally heat-resistant stages, they should be eliminated by pasteurization, and such spoilage is an indication of hygiene failure. Mold resistance to the common food preservatives is generally reported to be less than spoilage yeasts. However, many molds will degrade sorbic acid to strong-smelling 1,3-pentadiene, making mold spoilage of preserved beverages particularly distinctive.

Many molds have the potential to cause spoilage of soft drinks. Those most commonly encountered include *Aspergillus niger, A. ochraceus, A. fischeri*, and *A. tamarii, Byssoclamys nivea, B. fulva, Paecilomyces variotii, Neosartorya fischeri, Eupenicillium brefeldianum, Phialophora mustea, Taleromyces flavus, T. trachyspermus*, and *Thermoascus aurantiacum*.

Spoilage Bacteria

The ecological conditions of soft drinks limit the growth of bacteria, with formulation pH having the clearest effect on the development of potential spoilage organisms. Bacterial soft-drink spoilage organisms fall into three main categories: spore-formers,

lactic acid bacteria, and acetic acid bacteria. As most soft drinks are pasteurized, acidoduric/philic spore-forming bacterial species are of greatest threat to fruit-juice manufacturers. High-pH fruit or vegetable juices, tomato, apple, pear, or carrot juices have been reported spoiled by *Clostridium* spp. or *Bacillus* spp. including *C. butyricum, C. pasteurianum, Bacillus coagulans, B. licheniformis, B. macerans, B. subtilis*, and *B. polymyxa*, the thermophilic *B. coagulans* being the causative agent of 'flat-souring' in tomato juices.

Alicyclobacillus acidoterrestris is an obligately aerobic spore-forming thermophilic bacteria, associated with soils and acidic fruit juices principally from warm climates. Although not common at present, this organism has the potential to become a major cause of spoilage of fruit-containing soft drinks (Tables 5 and 6). It will grow at pH 2.5, and its spores have been reported to be very resistant or immune to pasteurization ($D_{95\,^\circ C}$ of 2–12 min; $D_{95\,^\circ C}$, time at temperature leaving 10 % survivors in the population). It may grow and produce tarry off-odors, including guaiacol and 2,6-dibromophenol, through the reduction of taint precursors. It does not produce gas, and bacterial haze is often not an important consideration as fruit juice may be naturally cloudy. Potentially, removal of taint precursors, limitation of storage temperature to < 20 °C, for example, prevention of ingress of oxygen or super-pasteurization will eliminate spoilage, though practicalities may limit these solutions.

Unpasteurized fruit juices with pHs higher than about 3.2 may be spoiled by the growth of lactic acid bacteria. *Lactobacillus* or *Leuconostoc* spp. can cause slime, ropiness, and off-flavors in fruit juices, particularly where adapted strains have been allowed to build up in production facilities. Lactic acid bacteria are heat-sensitive and lose viability in chilled juices. (*See* ***Bacillus***: Occurrence; **Lactic Acid Bacteria**.)

Gluconobacter spp. (*Acetomonas*) is the most frequently encountered cause of bacterial spoilage at low pH. Spoilage changes flavor characteristics and may lead to pack swelling and haze. *Gluconobacter* spp. are resistant to preservatives such as sorbic acid, benzoic acid, and DMDC, but are heat-sensitive and absolutely dependent on the presence of free oxygen. Spoilage is a problem only in gas-permeable packaging, such as still soft drinks in plastic beakers (Table 6).

Table 5 Heat resistance of yeasts, molds, mold ascospores, and bacterial spores associated with soft drinks

Microbe	*Form*	*D-values (min)*
Byssochlamys fulva	Ascospores	$D_{80\,^\circ C}$ 300, $D_{90\,^\circ C}$ 90–900
Byssochlamys nivea	Ascospores	$D_{88\,^\circ C}$ 45–48
Neosartorya fischeri	Ascospores	$D_{88\,^\circ C}$ 84, $D_{88\,^\circ C}$ 72–450
Aspergillus flavus	Conidiospores	$D_{55\,^\circ C}$ 186
Talaromyces flavus	Ascospores	$D_{88\,^\circ C}$ 468, $D_{90\,^\circ C}$ 60–420
Talaromyces macrosporus	Ascospores	$D_{88\,^\circ C}$ 420–1320, $D_{90\,^\circ C}$ 132–360
Alicylobacillus acidoterrestris	Spores	$D_{90\,^\circ C}$ 960–1380
Saccharomyces cerevisiae	Ascospores	$D_{55\,^\circ C}$ 6360
Saccharomyces cerevisiae	Vegetative cells	$D_{55\,^\circ C}$ 55
Zygosaccharomyces bailii	Ascospores	$D_{50\,^\circ C}$ 600–2220
Zygosaccharomyces bailii	Vegetative cells	$D_{50\,^\circ C}$ 120–240

Safety and Microbial Hazards in Soft Drinks

The hazards posed by the presence of microorganisms on soft drinks are much lower than for the majority of foods. Owing to the low pH of soft drinks, bacterial spores will not germinate, and vegetative cells progressively die off. Pathogenic bacteria are therefore of little consequence in acidic soft drinks. Fruit juices, notably apple juices prepared from ripe fruit, are generally less acidic than the majority of soft drinks and allow a longer survival of pathogens. Recent, much-publicized cases of illnesses caused by *Escherichia coli* 0157:H7 in unpasteurized apple juice appeared to be caused by animal fecal contamination of the fruit. Pathogens, such as *E. coli* or *Salmonella* spp., die in acidic apple juices, but viability can be maintained for several days and progressively lost over 3–5 weeks, respectively. Occurrences of salmonellosis in fruit juices are rare but finite. Pathogens in fruit juices are easily inactivated by pasteurization or by the addition of preservatives. Soft drinks may also very rarely contain viruses or the parasitic protozoan,

Table 6 Niche habitats formed by different varieties of soft drinks, and spoilage microbes associated with these habitats

Habitat	*Spoilage microbe*	*Spoilage indications*
High carbonation	*Brettanomyces/Dekkera* spp.	Clouds, off-flavors, odors
High preservatives	*Z. bailii, Z. bisporus, Z. lentus*	Excess gas, blown cans
Juices in thin plastic	*Gluconobacter (Acetomonas)*	Off-flavors, haze
Tomato juices	*Bacillus coagulans*	Flat souring
Syrups or concentrates	*Z. rouxii, Z. bailii*	Excess gas, blown cans
Refrigerated juices	*Z. lentus*	Clouds, off-flavors
Heat-treated juices	*Alicyclobacillus acidoterrestris*	Tarry taints, guaiacol
Heat-treated juices	Ascospore-forming molds	Mycelium, off-flavors

Cryptosporidium parvum, infection occurring via contaminated water supplies. Illnesses have been reported, caused by small, round-structured viruses in orange juice or *Cryptosporidium parvum* in apple juices and flavored fruit drinks.

The presence of fermentative yeasts in soft drinks also poses a physical hazard to the consumer and is caused by fermentation to high pressure within packages. Pressures comparable with that in Champagne bottles will cause metal cans and kegs to split or rupture (Figure 2), plastic bottles to distort and effervesce violently when opened, and glass bottles to shatter. Exploding bottles caused by *Zygosaccharomyces*. yeasts are a recognized cause of eye injuries. Molds do not generate gas to the same extent but, in higher pH products, could pose a threat from the formation of mycotoxins. Some concern has been raised regarding formation of patulin in stored apples by *Penicillium expansum*. Apple juices prepared from moldy fruit have been found to contain 1130 $\mu g\,l^{-1}$. Consequently, European and FDA limits have been set for patulin in apple juices.

See also: **Acids**: Natural Acids and Acidulants; **Pasteurization**: Pasteurization of Liquid Products; **Preservatives**: Classifications and Properties; Food Uses; **Soft Drinks**: Chemical Composition; Production; **Spoilage**: Bacterial Spoilage; Molds in Spoilage; Yeasts in Spoilage; **Yeasts**

Further Reading

Barnett JA, Payne RW and Yarrow D (2000) *Yeasts: Characteristics and Identification*, 3rd edn. Cambridge: Cambridge University Press.

Davenport RR (1998) Microbiology of soft drinks. In: Ashurst PR (ed.) *Chemistry and Technology of Soft Drinks and Fruit Juices*, pp. 197–216. Sheffield, UK: Sheffield Academic Press; Boca Raton, FL: CRC Press.

Deak T and Beuchat LR (1996) *Handbook of Food Spoilage Yeasts*. Boca Raton, FL: CRC Press.

ICMSF (International Commission on Microbiological Specifications for Foods) (1998) *Micro-organisms in Foods, 6, Microbial Ecology of Food Commodities*. London: Blackie Academic & Professional (Chapman & Hall).

Pitt JI and Hocking AD (1997) *Fungi and Food Spoilage*, 2nd edn. London: Blackie Academic & Professional (Chapman & Hall).

Samson RA and van Reenen-Hoekstra ES (1988) *Introduction to Food-borne Fungi*, 3rd edn. Baarn, The Netherlands: Centraalbureau voor Schimmelcultures.

Stratford M, Hofman PD and Cole MB (2000) Fruit juices, fruit drinks, and soft drinks. In: Lund BM, Baird-Parker AC and Gould GW (eds) *The Microbiological Safety and Quality of Food* pp. 836–869.Gaithersberg, MD: Aspen.

Varnam AH and Sutherland JP (1994) *Beverages: Technology, Chemistry and Microbiology*. London: Chapman & Hall.

Dietary Importance

S J Gatenby, Reading Laboratory, Reading, UK

Nutritional Composition

Soft drinks may make a valuable contribution to fluid intake and have become, to some extent, established as part of the daily diet, particularly of young children and adolescents. The nutritional value of a number of readily available soft drinks is shown in **Table 1**. Squashes, crushes, cordials, and carbonated drinks are, however, of little nutritional value (apart from their energy content) as their main ingredients are water and sugar. These soft drinks can be a valuable source of vitamin C, although it is unlikely that they will contain significant quantities unless the vitamin is added. Soft drinks do not contain fat or fiber but may contain nutritionally insignificant traces of protein. (*See* **Ascorbic Acid**: Physiology.)

Energy Content

The energy content of soft drinks varies greatly and is derived exclusively from the sweetening agents, which are principally sugars (**Table 1**). Soft drinks sweetened with mixtures of sugar and intense sweeteners are less caloric than drinks sweetened entirely with sugar, and drinks labeled as being low-calorie are required by UK law to contribute a maximum of 22 kJ (5 kcal) per 100 ml.

Sugar Content

The added sugar content of soft drinks varies from 6 to 10% and is mostly glucose and fructose, with small quantities of sucrose and perhaps maltose (**Table 1**). The sugar content of soft drinks is regulated by the UK 1964 Soft Drinks Regulations (amended 1969, 1970, and 1976). The nutritional value of a soft drink as consumed is dependent upon the dilution factor, which must now be stated on the label of all dilutable drinks. Soft drinks are a major market for intense sweeteners, particularly in the UK where, unlike many other countries, they can be used in conjunction with nutritive sweeteners and are not therefore limited to use in dietetic beverages. There are technical reasons, economic objectives, and health considerations supporting the need for intense sweeteners in soft drinks. No single sweetener is ideally suited to meet all the requirements of soft drinks, but by using them in combination the limitations of one sweetener can be offset by the strengths of another. The sweetening of soft drinks with

Table 1 Nutritional composition of soft drinks per 100 ml

Beverage	*Moisture (g)*	*Energy (kJ)*	*Carbohydrate (g)*	*Glucose (g)*	*Fructose (g)*	*Sucrose (g)*	*Maltose (g)*
Citrus squash	72.0	398	24.9	10.5	10.5	3.8	
Citrus drink	74.1	341	21.3	11.6	9.1	0.5	3.2
Citrus crush (nonconcentrated)	80.1	158	9.9	4.1	4.1	1.6	
Low-calorie citrus squash	95.4	37	2.3	1.2	1.1		
Low-calorie citrus drink	94.6	86	5.5	1.4	1.4	0.4	
Low-calorie citrus crush (nonconcentrated)	98.7	13	0.8	0.2	0.6		
Carbonates	90.7	152	9.5	4.0	4.8	1.7	
Low-calorie carbonates	99.6	0	0	0.03	0.03	0.06	
Ginger ale citrus drink	95.9	62	3.9	1.7	1.6	0.5	

Unpublished data with kind permission from the Ministry of Agriculture, Fisheries and Food, UK.

artificial sweeteners reduces the energy content and encourages fluid consumption without decreasing the nutrient density of the diet. Those drinks sweetened with aspartame may contain the amino acid phenylalanine in amounts which may be undesirable for individuals with phenylketonuria, a rare genetic disorder or metabolism that requires the patient to follow a special diet. The more liberal use of artificial sweeteners, the establishment of acceptable daily intake levels, and the increased number of sweeteners available has resulted in a wide range of low-calorie beverages which are acceptable to many consumers. Consumers have been advised to insure that they consume a mixture of sweeteners to avoid large intakes of any one type. (*See* **Aspartame; Carbohydrates**: Sensory Properties; **Sweeteners**: Intensive.)

Consumption Patterns

Information on consumption levels of soft drinks within the UK is mostly derived from three sources: the National Food Survey, the Dietary and Nutritional Survey of British Adults, and individual dietary surveys which are reported in the scientific literature. The UK National Food Survey omits foods which are bought for consumption outside the home. Information on soft drinks brought into the home is recorded but this information is not included in the major analysis of the survey data and is reported separately. Together with alcoholic drinks, sweets, and chocolate, for which no information is collected, soft drinks are often bought without the knowledge of the householder and are therefore liable to be underreported. Individual surveys have focused little on soft drink consumption in adults, although there is more information available on children's eating habits in this respect. The consumption of diet and other soft drinks was reported in the Dietary and Nutritional Survey of British Adults (Table 2).

The consumption of soft drinks in all age groups continues to increase and in 1989 was reported to be

Table 2 Consumption (grams per week) of soft drinks by adults by age

Drink	*Age (years)*			
	16–24	*25–34*	*35–49*	*50–65*
Low-calorie	778 (18)	1039 (17)	743 (16)	837 (12)
Other	1689 (82)	1073 (68)	889 (56)	667 (47)

Figures in parentheses indicate percentage consuming the drink. Data from Gregory J, Foster K, Tyler H and Wiseman M (1990) *The Dietary and Nutritional Survey of British Adults*, pp. 43, 44, 46. London: HMSO.

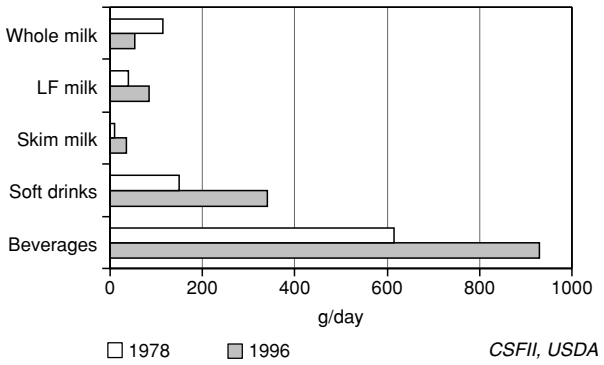

Figure 1 Trends in liquid food and beverage intakes in the US, 1978 and 1996. Data from the US Department of Agriculture, continuing Survey of Food Intake by individuals.

7524 million liters per year, with a total market value of £4.7 billion. This figure excludes the consumption of the increasingly popular fresh fruit juices. Consumption levels as reported in the UK National Food Survey can be seen in Table 3 and Figure 1.

While meal patterns still conform to what could be considered traditional, there is nevertheless a shift towards greater emphasis on snacking, particularly amongst younger age groups. Concern has been voiced regarding the nutritional quality of the diets of schoolchildren and snack eating in particular has been singled out as providing 'empty calories' but little else in the way of nutrients. Soft drinks are very popular amongst children of all age groups and

are frequently consumed between meals. (*See* **Snack Foods**: Dietary Importance.)

Contribution of Soft Drinks to Nutrient Intake in Schoolchildren

As many soft drinks are a source of energy, there is concern that children may obtain a significant proportion of their total daily energy requirement from the sugar contained within soft drinks. However, there is little published evidence to support this. In the scientific literature soft drinks have been reported to provide between 0.7% and 4.8% of total daily energy intake. With the increasing popularity of soft drinks, the wider variety of those available, and powerful advertising, it may be possible that children will reduce the nutrient density of their diet and be nutritionally at risk by consuming large quantities of soft drinks, replacing perhaps more nutrient-dense foods and liquids such as milk. A dietary study of 143 schoolchildren aged 11–12 years examined food consumption and nutrient intake using the 7-day dietary record technique. Soft drinks, confectionery, and table sugar contributed 10% of dietary energy and 40% of the sugar in the diet. Soft drinks were consumed at 1800 g per head each week, providing 4.7% of total energy intake and contributing 21.4% to total sugar intake. These values compare well with values obtained in the survey of British schoolchildren, where the mean consumption of soft drinks was 1775 g per head per week. There was no evidence from the anthropometric data that a high percentage of energy from sugar was associated with obesity. In fact, the average body mass index was lowest in the groups with the highest percentage of sugar intake, hence the putative effect of percentage of energy from sugar on body weight was not demonstrated in this study.

In another study investigating 11–14-year-old Northumbrian children, added sugars contributed on average 15% of dietary energy and 69% of total sugar intake. Soft drinks contributed 3% of energy and 17% of total sugar intake.

It has been estimated that if half the soft drinks consumed were sweetened with artificial sweeteeners alone, sugar consumption would fall by about 10%. Soft drinks are very popular with young children and are universally consumed, with 85–95% of those children surveyed in the UK study consuming these items during the study period. Soft drinks are without doubt one of the most commonly selected snacks amongst schoolchildren and it is reported that children find it difficult to rank fizzy drinks as either healthy or unhealthy and that they make their choice according to taste preference. (*See* **Children**: Nutritional Requirements; **Obesity**: Etiology and Diagnosis.)

In the British study of 15–18-year-olds there was an apparent decline in the consumption of soft drinks with increasing age. Average consumption fell from 200 to 155 g and 220 to 145 g soft drinks per day for female and male subjects, respectively. By adulthood the consumption of soft drinks has fallen markedly (Table 2).

Further studies are required to examine changing dietary patterns and the influence of increased soft drink consumption on total energy intake and nutrient density.

Product for Specific Uses

Diabetic Drinks

There are a number of soft drinks which are sold specifically for use by individuals with diabetes mellitus. By law these must not contain any sugars other than fructose, which is metabolized mainly in the liver, the initial steps in its metabolism being independent of insulin. Some products are sweetened with sugar alcohols (sorbitol, xylitol, mannitol, lactitol) but these offer no advantage over fructose and do not result in a reduced-energy drink. It has been suggested that the daily intake of sugar alcohols should not exceed 25 g as this may result in osmotic diarrhea. There is also an increasing number of low-calorie drinks in which the sugar has been replaced either wholly or in part by an intense sweetener, reducing the energy content significantly. The nature of these drinks enables those on reduced-energy diets to consume them without adding significantly to their energy intake.

The need for dietetic products of this nature is recognized by the provisions of the UK Soft Drinks Regulations (1964, as amended), which exempt low-calorie soft drinks from limitations imposed on the saccharin content of soft drinks in general. The

Table 3 Purchase quantity of soft drinks

Drink	*Quantity purchased (ml (fl. oz) per person per week)*				
	1984	*1987*	*1988*	*1989*	*1990*
Low-calorie	26 (0.91)	58 (2.04)	78 (2.74)	109 (3.85)	129 (4.55)
All soft drinks	844 (29.7)	855 (30.1)	886 (31.2)	1020 (35.9)	1071 (37.7)

Data from *Household Food Consumption and Expenditure* 1990 (with a study of trends over the period 1940–1990), Ministry of Agriculture, Fisheries and Food, UK; *Annual Report of the National Food Survey Committee*. London: HMSO, 1991.

relatively low caloric limits (22 kJ (5 kcal) per 100 ml) imposed on soft drinks compared with other foods make it virtually impossible to include carbohydrate sweeteners in low-calorie soft drinks. (*See* **Saccharin; Sugar Alcohols**.)

Sports Drinks

A range of foods, drinks, and supplements has been launched to fill the requirements of the 'get-fit' craze for consumption before, during, and after exercise. A relatively new phenomenon, these drinks are promoted as being designed to replace water and electrolytes lost through sweating during exercise. Most of the drinks are mixtures of sugar, salt, potassium, and water with a little vitamin C. Some of these drinks claim to be isotonic and some hypotonic. Dehydration does impair performance and therefore anything which decreases dehydration will maintain optimum performance for longer, although rehydration does not improve performance *per se*. In many circumstances water would be an acceptable alternative.

Infant Drinks

Infant drinks, including baby herbal drinks, have been developed to provide a thirst-quenching drink to coincide with the reduced consumption of milk that occurs in the first year of life. Although manufacturers are obliged to recognize that cooled, boiled water is the best refreshment to offer infants, they suggest that many will refuse this and offer their products as an alternative. The infant drinks are lower in sugar than baby fruit juices and unsweetened orange juice and usually provide the recommended daily intake of vitamin C. Potassium citrate is added as an acidity regulator, reducing the intensity of any acid, and infant drinks are of lower osmolality than conventional baby fruit juices. Health educators suggest that after the age of 6 months babies can be offered regular cows' milk or diluted natural fruit juice or water as a refreshment. (*See* **Infants**: Nutritional Requirements.)

Soft Drinks and Health

Dental Caries

Extensive evidence suggests that sugars are the most important dietary factor in the etiology of dental caries. Caries experience is positively related to the amount of nonmilk extrinsic sugars in the diet and the frequency of their consumption. Dental caries is dependent upon bacterial growth on the tooth surface, metabolism of sugars in the mouth by these bacteria, and the formation of acid which attacks the teeth. It is well known that one of the main etiological factors is the length of time sugar is in the mouth. (*See* **Dental Disease**: Role of Diet.)

In younger age groups there has been a shift away from traditional eating patterns of three or four meals per day towards an increasing reliance on snack foods, fast foods, and convenience foods accompanied by a range of sweetened or naturally sweet soft drinks consumed at frequent intervals throughout the day. There is therefore particular concern about the effect of soft drink consumption on dental health. Excessive use of soft drinks has been attacked on two accounts: first, almost all of them are fruit-based or carbonated or both and may therefore be acidic enough to erode the surfaces of the teeth not covered by dental plaque. Second, those which contain fermentable carbohydrate may serve as a source of substrate diffusing into the dental plaque, from which microorganisms inhabiting the plaque can generate acid which, in turn, brings about the destruction process of dental caries.

Carbonated drinks appear to be less cariogenic than pure orange juice and apple juice drinks and erosiveness may be more important than cariogenicity. Many soft drinks do contain sugars and, if allowed to reside in the mouth over long periods of time, as part of a frequent, protracted sugar intake pattern, then it is likely that they will be able to contribute to the caries process in which sugars serve as a substrate for acid formation.

Protective factors such as calcium and phosphorus may help to limit demineralization of teeth. In studies where iced lollies were fortified with calcium and phosphorus, this supplement substantially improved dental properties. Use of intense sweeteners in soft drinks reduces their viscosity and may assist in shortening exposure time.

See also: **Ascorbic Acid**: Physiology; **Aspartame**; **Carbohydrates**: Sensory Properties; **Children**: Nutritional Requirements; **Dental Disease**: Role of Diet; **Infants**: Nutritional Requirements; **Obesity**: Etiology and Diagnosis; **Saccharin**; **Snack Foods**: Dietary Importance; **Sugar Alcohols**; **Sweeteners**: Intensive

Further Reading

Bull NL (1985) Dietary habits of 15–18 year-olds. *Human Nutrition Applied Nutrition* 39A (suppl. 1): 1–68.

Department of Health (1989) *Dietary Sugars and Human Disease*. Reports on Health and Social Subjects, 37. London: HMSO.

Department of Health (1989) *The Diets of British Schoolchildren*. Report on Health and Social Subjects, 36. London: HMSO.

Department of Health (1991) *Dietary Reference Values for Food Energy and Nutrients for the United Kingdom*. Reports on Health and Social Subjects, 41. London: HMSO.

Gregory J, Foster K, Tyler H and Wiseman M (1990) *The Dietary and Nutritional Survey of British Adults*, pp. 43, 44, 46. London: HMSO.

Hackett AF, Rugg Gunn AJ, Appleton DR, Eastoe JE and Jenkins GN (1984) A 2-year longitudinal nutritional survey of 405 Northumberland children initially aged 11.5 years. *British Journal of Nutrition* 51: 67–75.

Ministry of Agriculture, Fisheries and Food (1964) *Soft Drinks Regulations 1964* (SI No. 760) as amended.

Nelson M (1991) Food vitamins and IQ. *Proceedings of the Nutrition Society* 50: 29–35.

Rugg Gunn AJ, Hackett AF, Appleton DR and Moynihan PJ (1986) The dietary intake of added and natural sugars in 405 English adolescents. *Human Nutrition Applied Nutrition* 40A: 115–124.

Sorbic Acid *See* **Preservatives**: Classifications and Properties; Food Uses; Analysis

SORGHUM

L W Rooney, Texas A&M University, TX, USA
S O Serna Saldivar, Instituto Tecnologico y de Estudios Superiores de Monterrey, Monterrey N. L., Mexico

Introduction

In terms of total world production, sorghum (*Sorghum bicolor L* Moench) is ranked fifth among cereals, with 62.7×10^6 tonnes in 1999. Production of sorghum is greatest in the USA, Nigeria, Mexico, India, China, Sudan, Ethiopia, and Argentina, but also occurs in other countries. Most African and Asian countries use sorghum for food, feed, forage, fuel, and building material. In the western hemisphere and Australia, sorghum is used mainly for feed and forage. Speciality types are used for syrup, sugar and alcohol. (*See* **Cereals**: Contribution to the Diet.)

The sorghum plant originated in the northeast quadrant of Africa. It belongs to the Gramineae family, Panicoideae subfamily, and Andropogeneae tribe. It is a warm-season, annual crop, favored by high day and night temperatures and intolerant to low temperatures. Improved types germinate, mature, and yield grain in an average of 120–140 days. Sorghum plants range in height from 0.6 to 6 m and possess an imperfect complete flower that generally self-pollinates. The grain develops on a branched terminal panicle that can be compact or very open. Flowering proceeds from the top of the panicle downward, with each panicle containing 800–3000 caryopses.

An enormous range in morphological diversity exists in the sorghum crop. The World Sorghum Collection in India has more than 29 000 entries. Sorghum is classed as grain sorghum, forage sorghum, grass or Sudan sorghums, and broomcorn. The latter is grown for its long, fibrous panicle branches that are used to manufacture brooms. The grain of sorghum is classed according to pericarp color (white, yellow, or red), presence or absence of a pigmented testa (with or without tannins), pericarp thickness (thin or thick pericarp), endosperm color (white, heteroyellow, or yellow), and endosperm type (normal, heterowaxy, or waxy). These kernel characteristics are genetically controlled. Brown sorghums have a pigmented testa and contain tannins; they may have any color pericarp.

Grain Structure and Physical Properties

The sorghum kernel is considered a naked caryopsis, although some African types retain their glumes after threshing. The kernel weight varies from 3 to 80 mg. The size and shape of the grain vary widely among sorghum races. Commercial sorghum grain has a flattened spherical shape, 4 mm long, 2 mm wide and 2.5 mm thick, with a kernel weight of 25–35 mg. The volumetric weight and grain density range from 708 to 760 kg m^{-3} and from 1.26 to 1.38 g cm^{-3}, respectively.

The sorghum caryopsis is composed of three anatomical parts: pericarp, endosperm, and germ (**Figure 1a**). The relative proportion of these structures varies but in most cases is 6, 84, and 10%, respectively. The pericarp (**Figure 1b**) is the fruit coat and is fused to the sorghum seed. It originates from the ovary wall and is subdivided into three distinctive parts: epicarp, mesocarp, and endocarp. The epicarp is the outermost layer and is generally covered with a waxy film. The mesocarp varies in thickness and can contain starch granules. The endocarp is composed of cross and tube cells and plays a major role during germination. (*See* **Wheat**: Grain Structure of Wheat and Wheat-based Products.)

The true seed consists of the seed coat or testa, endosperm, and germ. The endosperm tissue is triploid, resulting from the fusion of a male gamete with two female polar nuclei. The testa, or seed coat, is derived from the ovule integuments; in brown

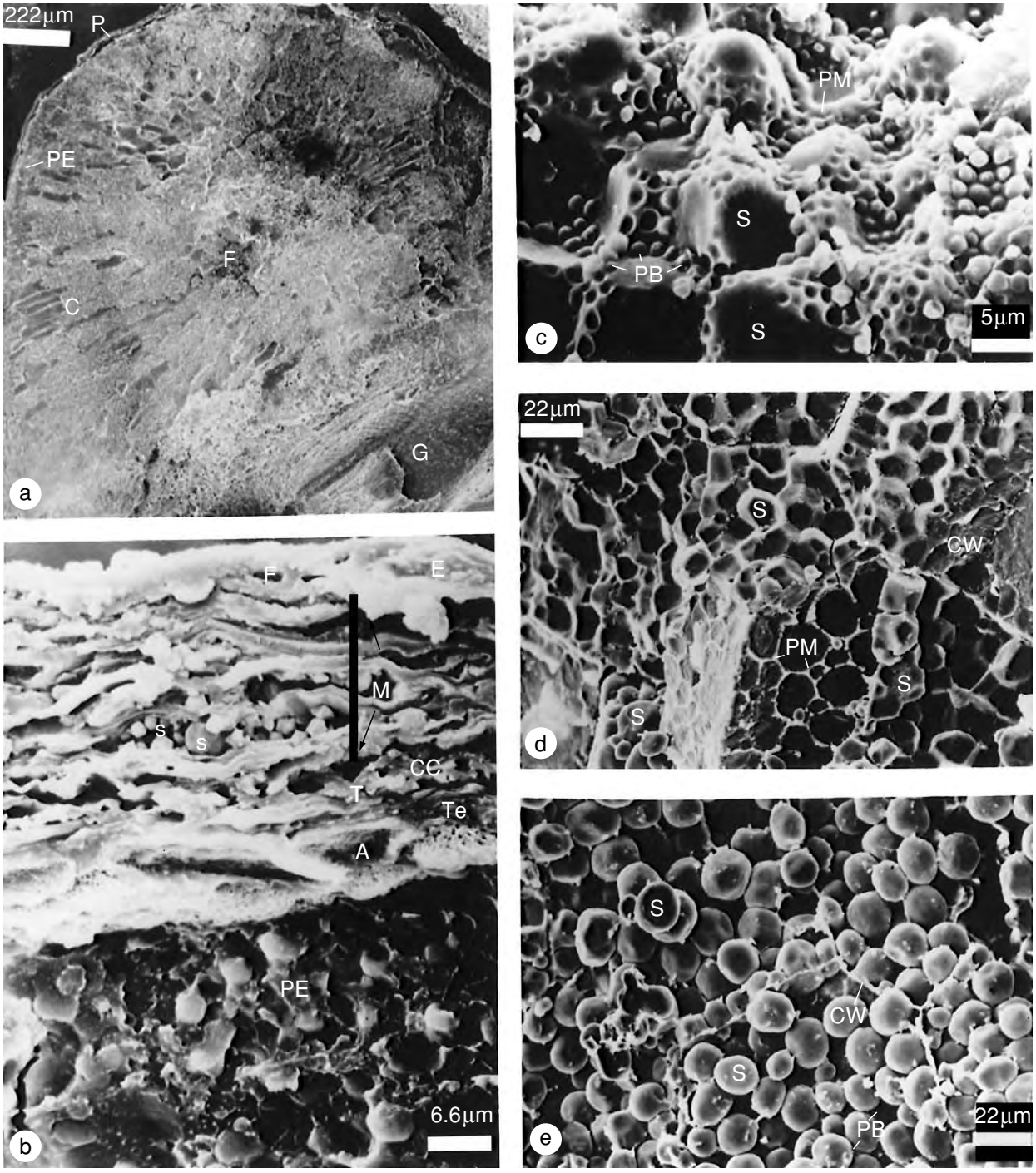

Figure 1 Structure of the mature sorghum kernel viewed with the scanning electron microscope. (a) Cross-section, (b) pericarp, (c) peripheral endosperm, (d) corneous endosperm, and (e) floury endosperm. P, pericarp; PE, peripheral endosperm; C, corneous endosperm; F, floury endosperm; G, germ; E, epicarp; M, mesocarp; CC, cross cells; T, tube cells; Te, testa; A, aleurone; S, starch granule; PB, protein body; PM, protein matrix; CW, cell wall. From Rooney LW and Serna-Saldivar SO (2000).

sorghums, it is thick and contains condensed tannins. The testa is pigmented when the genetics are B_1-B_2-. The endosperm is composed of the aleurone layer, peripheral, corneous, and floury areas. The aleurone consists of a single layer of rectangular cells adjacent to the tube cells or testa. Aleurone cells contain a thick cell wall, large amounts of proteins (protein bodies) and enzymes, ash (phytic acid bodies), and oil bodies (spherosomes). The peripheral endosperm (Figure 1c) adjacent to the aleurone layer is composed

of dense cells containing large quantities of protein and small starch granules. These layers affect the processing and nutrient digestibilities of sorghum. Processing of sorghum by steam flaking, micronizing, popping, and reconstitution is designed to disrupt the endosperm structure to improve the digestibility.

The corneous and floury endosperm cells are composed of starch granules, protein matrix, protein bodies and a thin cell wall rich in β-glucans and hemicellulose. In the corneous endosperm, the protein matrix has a continuous interphase with the starch granules, with protein bodies embedded in the matrix (Figure 1d). The starch granules are shaped polygonally and often contain dents from protein bodies. The appearance is translucent or vitreous. The opaque-floury endosperm is located around the geometric center of the kernel. It has a discontinuous protein phase, air voids, and loosely packaged, round-lenticular starch granules, and is opaque to transmitted light (Figure 1e).

The germ is diploid owing to the sexual union of one male and one female gamete. It is divided into two major parts: the embryonic axis and scutellum. The embryonic axis forms the new plant and is subdivided into a radicle and plumule. The radicle forms primary roots, whereas the plumule forms leaves and stems. The scutellum is the single cotyledon of the sorghum seed. It contains large amounts of oil (spherosomes), protein, enzymes, and minerals, and serves as the connection between the endosperm and germ.

Table 1 Composition (%, unless otherwise stated) of sorghum grain[a]

Component	*Value*
Protein (N × 6.25)	11.6
Ether extract	3.4
Crude fiber	2.7
Ash	2.2
Nitrogen-free extract[b]	79.5
Starch	74.1
Fiber	
Dietary, insoluble	7.2
Dietary, soluble	1.1
Acid detergent	3.3
Soluble sugars	2.1
Pentosans	1.3
Prolamine[c]	52.7
Glutelins[c]	34.4
Albumins[c]	5.7
Globulins[c]	7.1
Essential amino acids[c]	
Lysine	2.1
Leucine	14.2
Phenylalanine	5.1
Valine	5.4
Tryptophan	1.0
Methionine	1.0
Threonine	3.3
Histidine	2.1
Isoleucine	4.1

[a]All values are expressed on a dry-matter basis. Adapted from Rooney LW and Serna-Saldivar SO (2000).
[b]Calculated by difference.
[c]FAO/WHO suggested pattern (g amino acid per 100 g protein): lysine, 5.44; leucine, 7.04; phenylalanine plus tyrosine, 6.08; valine, 4.96; tryptophan, 0.96; methionine plus cysteine, 3.52; threonine, 4.0; isoleucine, 4.0.

Composition

Sorghum grain composition (Table 1) varies significantly, owing to genetic and environmental influences, and is similar to that of maize (*Zea mays* L). Starch (75–79%) is the major component, followed by protein (9.0–14.1%) and oil (2.1–5.0%). The protein content (N × 6.25) of sorghum is more variable and usually 1–2% higher than maize. Approximately 80, 16, and 3% of the protein is in the endosperm, germ, and pericarp, respectively. Sorghum generally contains 1% less oil and significantly more waxes than maize. (*See* **Cereals**: Dietary Importance.)

Sorghum starch is composed of 70–80% amylopectin and 20–30% amylose. Waxy sorghums contain starch with 100% amylopectin; their properties and uses are similar to those of waxy maize. The gelatinization temperature of sorghum starch is slightly higher than that of maize starch. (*See* **Starch**: Structure, Properties, and Determination.)

The main protein fraction in the kernel is the prolamines (kafrins), followed by glutelins (Table 1). The alcohol-soluble prolamine fraction comprises 50% of the protein. These proteins are hydrophobic, rich in proline, aspartic, and glutamic acids, and contain little lysine. They are mainly found in protein bodies and are affected by nitrogen fertilization. Glutelins are high-molecular-weight proteins, mainly located in the protein matrix. The lysine-rich protein fractions – albumins and globulins – predominate in the germ. High-lysine sorghums, such as P-721 and certain Ethiopian types, contain lower levels of kafrins and higher levels of albumins and globulins. The higher-lysine sorghum varieties are soft or dented, and are not produced commercially. (*See* **Protein**: Chemistry.)

Most of the fiber is present in the pericarp and cell walls. Aleurone and endosperm cell walls are associated with ferulic and caffeic acid. Around 85% of the dietary fiber is insoluble and mainly composed of hemicellulose and cellulose. The soluble fraction is rich in pentosans and β-glucans. Sorghum contains approximately 1.3% pentosans, located mainly in the pericarp. Approximately 70% of the pentosans are alkali-soluble, and some 30% are water-soluble. (*See* **Dietary Fiber**: Properties and Sources.)

The germ and the aleurone layer are the main contributors to the lipid fraction. The germ provides about

80% of the oil. The fatty acid composition consists mainly of linoleic (49%), oleic (31%) and palmitic (14.3%) acids. Refined sorghum oil is very similar to maize oil in quality. (*See* **Fatty Acids**: Properties.)

Most of the minerals are concentrated in the pericarp, aleurone, and germ. The ash fraction is rich in phosphorus and potassium and low in calcium and sodium. Most of the phosphorus is bound to phytic acid. The germ and aleurone are rich in fat-soluble and B vitamins. Precursors of vitamin A or carotenes are found only in yellow and heteroyellow endosperm cultivars. (*Refer to* individual vitamins and minerals.)

All sorghums contain phenolic acids, and most contain flavonoids, but only brown sorghums contain condensed tannins. Tannins protect the kernel against preharvest germination and attack by insects, birds, and molds. Brown sorghums always have a pigmented testa (genes B_1-B_2-ss) and some have tannins in the pericarp (B_1-B_2S-). Sorghums without a pigmented testa do not contain any condensed tannins. Birds can and do consume brown sorghums when other food is unavailable. (*See* **Tannins and Polyphenols**.)

Uses of Grain

Milling

Thirty percent of world sorghum production is for human consumption. For production of most traditional foods, sorghum is first dehulled with a wooden mortar and pestle. For decortication, the grain is usually washed, placed in the mortar, and pounded vigorously with the pestle. The abrasive action frees the pericarp from the kernel. The pericarp usually detaches at the mesocarp. Thick pericarp cultivars, with hard endosperm and round kernels, are preferred for decortication. The bran or pericarp is separated from the grain by washing with water or by winnowing the sun-dried grain. Most sorghums are decorticated to remove 10–30% of the original grain weight. Mechanical decortication with rice milling equipment or abrasive disks is becoming more popular. The decorticated kernels are reduced into flour by pounding in the mortar and pestle, with stone mills or by diesel-powered attrition mills. Flour is sieved to obtain fractions with an acceptable particle size for specific products.

Commercial milling of sorghum is practiced in several countries. The sorghum is decorticated using abrasive pearlers, followed by degermination and subsequent sieving, milling, gravity separation, and sieving to produce low-fat grits, meal, and flour. Milling of sorghum with wheat roller mills produces acceptable flour and other products.

Traditional Food Uses

The major categories of traditional foods are as follows: fermented and unfermented flat breads; fermented and unfermented thin and thick porridges; steamed and boiled products; snack foods and alcoholic and nonalcoholic beverages. World-wide, the most popular unfermented flat breads are roti in India and tortillas in Central America. For roti, a portion of the flour is gelatinized, mixed with more flour, and warm water and kneaded into a dough, which is shaped or rolled into a circle and baked on a hot griddle. For tortilla production, whole sorghum is limecooked, steeped overnight, washed, stone-ground into a dough, shaped into thin circles, and baked on a hot griddle. (*See* **Chapatis and Related Products**; **Tortillas**.)

The most popular fermented breads are injera, kisra, and dosa, consumed in Ethiopia, Sudan, and India, respectively. About 80% of the Ethiopian sorghum is used for the production of injera. The sorghum flour is mixed with water and a yeast starter from a previous batch of injera. After fermentation for 24–48 h, the batter is poured onto a greased pan for baking. The resulting product is a flexible, brown pancake containing uniformly distributed fish eyes or air bubbles. Dosa is consumed in India and is produced from a mixture of black gram, sorghum, and rice flour.

Porridges can be fermented or cooked with acid or alkali. Tô is an unfermented stiff porridge, very popular in Mali and Burkina Faso. Decorticated sorghum flour is cooked in plain water or water acidified with tamarind juice or made alkaline with wood ashes (potash). The most popular fermented porridges are ogi and nasha, widely consumed in West and East Africa, respectively. Whole sorghum is soaked in water and allowed to ferment for 2–3 days. The wet grain is crushed in a slurry of water and sieved to remove the bran. The fine particles are allowed to ferment longer. Excess water is decanted and the resulting slurry cooked in water or milk.

For couscous production, sorghum flour is kneaded with enough water to form agglomerates. The particles are forced through a coarse screen and steamed. The cooked product is consumed with a sauce or milk.

Decorticated sorghums are often cooked like rice. Special types of small-seeded, very hard sorghums are used as a substitute for rice.

Two major kinds of alcoholic beverages are produced from malted sorghum. The most common type, called opaque beer, undergoes souring and yeast fermentation and is very popular in southern Africa. The high-solids beer is sour, alcoholic, pinkish and

effervescent. (*See* **Fermented Foods**: Beverages from Sorghum and Millet.)

Industrial Uses

Sorghum grain or sweet sorghum biomass is used for ethanol production. Yields of alcohol (182° proof) per tonne of sorghum grain are comparable with maize (387 vs 372 l). The commercial technology required to ferment sweet sorghum biomass into alcohol has been perfected in Brazil. One tonne of sweet sorghum biomass has the potential to yield 74 l of 200° proof alcohol.

In Southern and West Africa, sorghum malt is used for alcoholic and nonalcoholic beverages, weaning foods, and breakfast foods. Sour-opaque beers are produced commercially in Southern Africa. Opaque beer is produced following the basic steps of the traditional process. Breweries in Mexico, Africa, and Asia use sorghum grits as an adjunct in brewing lager beer. In Nigeria, sorghum and maize are used to produce lager or clear beer without barley malt. Nigerian breweries produce clear beer from a combination of malted sorghum, sorghum, and/or maize grits with commercial enzymes to convert the starch to fermentable sugars because sorghum malt has a low diastatic power. The quality of clear beer is good, but the taste differs from barley malt beer.

Sorghum grits, meal, and flour can be used alone or mixed with wheat flour to produce an array of baked goods. Sorghum does not contain gluten. Thus, the amounts of sorghum flour in the blends depend on the quality of the wheat flour, baking procedure, formulation, and quality of the baked products desired. New food-type sorghums that produce excellent yields of flour with a bland flavor and light color are available. They can be used to extend wheat-based products without affecting flavor. (*See* **Flour**: Dietary Importance.)

Sorghum can be puffed, popped, shredded, and flaked to produce ready-to-eat breakfast cereals. Extrusion of sorghum produces acceptable snacks and precooked products. (*See* **Cereals**: Breakfast Cereals.)

In the Sudan, sorghum is wet-milled to produce starch with properties and uses similar to those of maize starch. Commercial wet milling of sorghum in the USA was discontinued in the 1970s because of poor economics. The enzymatic conversion of starch to liquid glucose syrup is possible.

Nutritional Value

Sorghum is an excellent feed for livestock and companion animals. The feeding value of sorghum for livestock species is generally considered 95% or more of the feeding value of yellow dent maize. Brown sorghums are considered to have 80–85% the feeding value of maize. Sorghum must be properly processed to enhance its digestibility. Popping, steam flaking, and reconstitution are used to prepare sorghum grain in beef cattle feedlots in the USA. Grinding, rolling, crushing, and pelleting are used for poultry and swine feeds.

Sorghum has a proximate composition, amino acid contents, and nutritional value similar to those of maize. However, due to its lower fat content, sorghum usually has a slightly lower gross, digestible, and metabolizable energy. Lysine and threonine are the first and second limiting amino acids. The tryptophan content is higher than that in maize. High-lysine cultivars contain approximately 50% more lysine and promote better weight gains in weaning rats. Fermentation, malting, and other processing methods improve the nutritional value significantly. (*See* **Amino Acids**: Properties and Occurrence.)

Most sorghum hybrids do not contain condensed tannins in contrast to impressions given in some publications. Kernels that contain condensed tannins have a clearly pigmented, thick testa. These sorghums are classed as brown or tannin sorghums. The brown sorghums are grown in some areas where birds and molds are major problems. Birds will consume brown sorghums when other foods are scarce. Animals fed brown sorghum rations eat more feed and produce about the same amount of gain so feed efficiency is reduced significantly.

See also: **Cereals**: Contribution to the Diet; Breakfast Cereals; **Chapatis and Related Products**; **Dietary Fiber**: Properties and Sources; **Fatty Acids**: Properties; **Fermented Foods**: Beverages from Sorghum and Millet; **Flour**: Dietary Importance; **Protein**: Chemistry; **Starch**: Structure, Properties, and Determination; **Tannins and Polyphenols**; **Wheat**: Grain Structure of Wheat and Wheat-based Products

Further Reading

FAO (1999) *FAOSTAT Database Results*, Web Site http://apps1.fao.org, Yearly Production Database.

Rooney LW and Miller F (1982) Variation in the structure and kernel characteristics of sorghum. In: Rooney LW and Murty DS (eds) *International Symposium on Sorghum Grain Quality*, pp. 143–162. Patancheru, India: ICRISAT.

Rooney LW, Kirleis AW and Murty DS (1986) Traditional foods from sorghum: their production, evaluation and nutritional value. In: Pomeranz Y (ed.) *Advances of Cereal Science and Technology*, vol. VIII, pp. 317–353. St Paul, MN. American Association of Cereal Chemists.

Rooney LW and Serna-Saldivar SO (2000) Sorghum. In: Kulp K and Ponte J (eds) *Handbook of Cereal Science*

and Technology, 2nd edn, pp. 149–175. New York: Marcel Dekker.

Rooney LW and Waniska RD (2000) The sorghum kernel: structure and chemistry. In: Smith C and Frederiksen R (eds) *Sorghum: Origin, History, Technology and Production*, pp. 689–729. New York: Wiley.

Serna-Saldivar SO and Rooney LW (1995) Structure and chemistry of sorghum and millets. In: Dendy D (ed.) *Sorghum and Millets: Chemistry and Technology*, pp. 69–124. St Paul, MN: American Association of Cereal Chemists.

Stoskofp NC (1980) *Cereal Grain Crops*. Reston, VA: Reston.

Waniska RD and Rooney LW (2000) Structure and chemistry of the sorghum caryopsis. In: Smith C and Frederiksen R (eds) *Sorghum: Origin, History, Technology and Production*, pp. 649–688. New York: Wiley.

Sourdough Bread *See* **Bread**: Dough Mixing and Testing Operations; Breadmaking Processes; Chemistry of Baking; Sourdough Bread; Dietary Importance; Dough Fermentation

SOY (SOYA) BEAN OIL

F D Gunstone, Scottish Crop Research Institute, Invergowrie, Dundee, UK

Background

Soybean is the oilseed produced in the largest amounts. The extracted oil is also produced in the largest amounts, though in the next 10–15 years, the supply of palm oil is expected to exceed that of soybean oil. Soybean meal is the dominant oilseed source of high-quality protein for both human and animal use. Other entries in this Encyclopedia deal with the bean and its protein product: this section is concerned with soybean oil. Some advantages and disadvantages have been claimed for this oil. It is widely used on the basis of its advantages, and much effort is devoted to overcoming its alleged disadvantages. The advantages are its high unsaturation, its richness in essential fatty acids, its liquid nature over a wide temperature range, its extensive use as a pourable semisolid oil after partial hydrogenation, and its value as a rich source of phospholipids, antioxidants, and sterols. Its disadvantages include the need to process it to remove phospholipids before use and the presence of linolenic acid, which is considered to promote oxidative instability and to be the cause of the rapid development of off-flavors, which limit shelf-life.

Oil Extraction and Refining

Beans are cleaned, dried to about 10% moisture, dehulled, and flaked before extraction with solvent (isohexane or hexane) by percolation or immersion. Oil/solvent is separated from meal/solvent, and the solvent is removed and recovered for reuse from both fractions. The crude oil is then refined by degumming to remove and recover phospholipids, neutralizing to remove free acids, bleaching to remove undesirable colours, and deodorizing to remove unwanted odor and flavor compounds. These processes, and particularly the last, should be carried out at as low a temperature as possible. Formation of *trans* isomers is negligible below 220 °C, becomes significant between 220 and 240 °C, and rises rapidly with respect to temperature above 240 °C. Some 3–24% of the linolenic present may be isomerized, but linoleic acid is less reactive, and isomerization does not usually exceed 2%.

$$9c12c15c \rightarrow 9t12c15c \text{ and } 9c12c15t \text{ and } 9t12c15t.$$

These processes produce a high-quality bland oil, though for several uses, it may also be necessary to effect some degree of hydrogenation. Very light hydrogenation (brush hydrogenation) is used to reduce the level of linolenic acid to about 3%. This has the advantage of extending the shelf-life of food products containing this material, as linolenic acid is oxidized about twice as quickly as linoleic acid and gives rise to oxidation products with a greater off-taste and off-odor than those resulting from linoleic acid. More extensive hydrogenation is designed to increase the melting point of the oil, so that it can be used to make plastic semisolid fats. This is achieved by raising the levels of acids with melting points above ambient so that the triacylglycerols in which they are present are also solid. These acids are either saturated (stearic)

or monounsaturated acids with *trans* configuration. These changes also have nutritional consequences. The ratio of total dietary *n*-6 (linoleic acid and its metabolites) to total dietary *n*-3 acids (linolenic acid and its metabolites) is considered by many to be too high, with an excess of the former and a deficiency of the latter. This is exacerbated when valuable dietary linolenic acid is reduced through partial hydrogenation. Also, there is now nutritional concern about *trans* acids, and serious attempts are being made to reduce the dietary intake of such acids. (*See* **Vegetable Oils**: Oil Production and Processing.)

Composition of Soybean Oil

The following composition figures for soybean oil refer to the the crude and refined oils, respectively: triacylglycerols (95–97 and >99%), phospholipids (1.5–2.5 and 0.003–0.045%), free fatty acids (0.3–0.7 and <0.05%), plant sterols (0.33 and 0.13%), tocopherols (0.15–0.21 and 0.11–0.18%), iron (1.3 and 0.1–0.3 p.p.m.), and copper (0.03–0.05 and 0.02–0.06 p.p.m.). It is apparent that refining removes most of the phospholipids and free acids but only some of the sterols, tocopherols, and metals. The level of tocopherols remaining in the refined oil is sufficient to provide a measure of oxidative stability. Material removed during the refining process contains several products of commercial value to the food, cosmetics, and pharmaceutical industries. These include the phospholipids (lecithin), tocopherols (vitamin E), and sterols with these industries relying on the supply of each of these.

The following figures for refined soybean oil have been cited as target values: FFA (0.05 max), *trans* acids (1% max), moisture (0.05% max), impurities (nil), peroxide value (max 0.5 mEq O_2 kg^{-1}), phosphorus (max 1 p.p.m.), tocopherols (min 750 p.p.m.), total metal (max 0.1 p.p.m.), PAH (polycyclic aromatic hydrocarbons max 25 p.p.b.), colour 5.25 inches max (1.5R–10Y), cold test (min 48 h per °C), smoke point (min) 220 °C.

Fatty Acid Composition

Typical fatty acid composition data for soybean oil are listed given in Table 1. The oil is rich in linoleic acid (54%, Codex range 50–57%) and oleic acid (24%, Codex range 18–25%) but also contains palmitic (11%, Codex range 10–13%), stearic (4%), and linolenic acids (7%, Codex range 5–10%), with only traces of acids of other chain lengths. Attempts are being pursued both by traditional mutation seed-breeding methods and by genetic engineering to produce soybeans with different fatty acid compositions, including oil rich in lauric acid as an alternative to coconut and palmkernel oils, oil with less palmitic and stearic acid on dietary grounds, oil with more palmitic and/or stearic acid for use in solid (plastic) fats without hydrogenation (which adds cost and produces undesirable *trans* acids), oil with more oleic acid on dietary grounds, and oil with less linolenic acid to enhance the shelf-life without brush hydrogenation. Many such modified oils have been described, and a typical selection is reported in Table 1. Several of these are in an advanced stage of development, and it is likely that these, or materials akin to them, will become available soon. It is expected that they will carry a premium price, but details are not yet known.

Table 1 Fatty acid composition of commodity soybean oil and some modified soybean oils

	16:0	*18:0*	*18:1*	*18:2*	*18:3*
Commodity oil	11	4	24	54	7
High-oleic	11		81	4	4
Low-saturated	4	3	31	59	3
High-saturated	25	4	16	44	10
	24	19	9	38	10
	12	21	63	1	3
Low-linolenic	10	3	44	40	3

Soybean oil can be converted to a semisolid material by partial hydrogenation. For example, soybean oil of iodine value 133 can be converted to fats of iodine value ~80 and mp 30–37 °C depending on the conditions of hydrogenation. These materials contain 18:0 8–12% compared with 4% in the original oil, 18:1 63–72% (23%), 18:2 9–13% (53%), and 18:3 0–0.3% (7.4%). But these numerical changes in fatty acid composition do not fully convey the changes that have taken place for the partially hydrogenated oils contain 27–40% of *trans* acids, which will be mainly 18:1 with the *trans* double bond ranging from Δ6 to Δ16 but mainly Δ8 to Δ13. The *trans* acids have higher melting points than their *cis* isomers and thus contribute to the hardening process. (*See* **Fats**: Classification; **Fatty Acids**: Properties.)

Triacylglycerol Composition

Although most oils are traded and used according to their fatty acid composition, it is known that some properties are related to triacylglycerol composition. The following data are taken from a paper describing triacylglycerol composition for soybean oil and for several modified oils, but only the former is discussed here. The original paper must be consulted for other details.

Ignoring the minor components, the soybean oil under study had a fatty composition of palmitic (10.2%), stearic (4.1%), oleic (23.4%), linoleic (52.8%), linolenic (8.3%), and other acids (1.2%)

and was analyzed by appropriate methods to give 'individual' molecular species and also the distribution of fatty acids at the *sn*-1, 2, and 3 positions.

In terms of molecular species it is to be expected that in view of the high level of linoleic acid, the dominant triacylglycerol molecules will contain at least two linoleic chains, thus: LLLn (7.9%), LLL (17.6%), LLO (15.3%). LLSt (4.2%), and LLP (10.2%), making a total for this group of 55.2%. Other triacylglycerols include LnLnL (3.1%), LnLO (4.8%), LnLP (3.7%), LOO (6.3%), LOP (6.9%), PLL (3.1%), OOO (3.3%), LOSt (3.7%), OOP (3.4%), PLSt (3.1), others (3.4%) where Ln=linolenic, L=linoleic, O=oleic. St=stearic, and P=palmitic. These three-letter symbols indicate only the acids present, and not their position on the glycerol backbone, so that, for example, LLO represents three different isomers (LLO, LOL, and OLL), and LOP represents the sum of six isomers. These results were obtained by reverse-phase HPLC.

The stereospecific analysis of the soybean oil gives the results indicated in Table 2. As would be expected, the *sn*-2 position has enhanced levels of oleic and linolenic acid and lower levels of palmitic and stearic acid. The results for linolenic acid are somewhat unexpected, since all unsaturated acids are usually enhanced in the *sn*-2 position. The two sets of results for the *sn*-2 position agree reasonably well, except for oleic and linolenic acid, when allowance is made for the difficulties of these analytical procedures.

Other Components

The AOCS publication on 'Physical and Chemical Characteristics of Oils, Fats and Waxes' reports that crude soybean oil contains 1.8–4.1 mg per kilogram of sterols, which are mainly campesterol (16–24%), stigmasterol (16–19%), and β-sitosterol (52–58%). Tocopherol levels are given as 601–3363 mg kg^{-1} and include the α (9–352), β (0–36), γ (89–2307), and δ (154–932) forms. (*See* **Tocopherols**: Properties and Determination; **Vitamins**: Overview.)

Table 2 Stereospecific distribution of fatty acids in soybean oil

	Oil	*sn-1*	*sn-2*	*sn-3*	*sn-2*[a]
16:0	10.2	18.2	1.9	12.0	3.7
18:0	4.1	5.4	2.7	4.5	3.4
18:1	23.4	15.9	28.4	25.0	21.4
18:2	52.8	47.8	64.1	48.3	65.0
18:3	8.3	12.9	2.6	10.3	6.4
minors	1.2				

[a]Results obtained by pancreatic lipase hydrolysis; other results obtained by the method involving naphthylurethanes of diacylglycerols. The original paper includes more detail and similar data for modified soybean oils (Reske J, Siebrecht J and Hazebroek J (1997) Triacylglycerol composition and structure in genetically modified sunflower and soybean oils. *Journal of the American Oil Chemists' Society* 74: 989–998).

Physical Properties

The same publication lists the following physical and chemical properties properties: specific gravity at 20/20° 0.919–0.925, refractive index at 40 °C 1.466–1.470, iodine value 124–139, saponification value 189–195, unsaponifiable not above 1.5%.

Major Food Uses

Oils and fats are significant components of the diet contributing 30–40% of caloric intake. They are consumed in many different forms as: spreading or yellow fats (butter, margarine, industrial margarine, low-fat spreads), frying fats, salad oils, and mayonnaise, and as an essential component of baked goods. Refined soybean oil, before or after hydrogenation, is used in these. As indicated later (Table 3), its contribution is very high in the USA but somewhat lower in other countries. In many cases, blends of two or more oils are required to achieve the desired properties. When considering the relative nutritional merits of these preparations, it is important to note the extent to which the various components have been hydrogenated, that is to know the level of *trans* acids that they contain.

Mayonnaise must contain at least 65% oil and salad dressings at least 30%, though products with lower levels of fat are also produced. These products are generally refrigerated, so it is important that the oil does not crystallize and break the emulsion. It is therefore usual to use soybean oil that has been hydrogenated to IV 110–115 to reduce the level of linolenic acid and then winterized to remove components that would crystallize under refrigeration. Oils used for mayonnaise or salad dressings must remain clear when cooled in ice water (0 °C) for 5.5 h. Soybean oil treated as above remains clear, even after 24 h.

Table 3 Disappearance (million tonnes) of selected oils in USA, EU-15, China, India, and Japan in 1999/2000

	sbo	*sun*	*rape*	*palm*	*pko*	*coco*
USA	7.42	0.13	0.68	0.15	0.20	0.48
EU-15	1.88	2.16	3.23	2.24	0.48	0.60
China	2.87	0.25	4.52	1.37	0.02	0.05
India	1.55	0.82	1.94	3.38	0.02	0.44
Japan	0.69	0.01	0.93	0.36	0.05	0.06

sbo, soybean oil; sun, sunflower oil; pko, palmkernel oil; coco, coconut oil. Source: *Oil World Annual* (2000) Hamburg: ISTA Mielke.

Frying oils, particularly for the preparation of fried goods that will be stored, should not contain linolenic as this will shorten the shelf-life. For such products, it is usual to use a partially hydrogenated oil, and the hydrogenated–winterized oil developed for mayonnaise is often used for frying purposes also. For fast-food purposes, where long-term storage is not necessary, this additional processing of the oil is less important.

Margarines have to meet certain physical specifications, particularly in respect of their appearance and melting behavior between refrigerator (4 °C) and mouth temperature (37 °C) as well as nutritional guidelines. A variety of binary and ternary mixtures of soybean oil and of soybean oils hydrogenated to various iodine values have been used. The hydrogenated oils contain *trans* acids, and attempts are being made to reduce the level of these by developing procedures of hydrogenation that give less *trans* acids or by blending nonhydrogenated oil with 'hardstock,' which may be a fully hydrogenated oil or palm stearin. These products are improved by interesterification of the blended oils. (*See* **Margarine**: Types and Properties.)

Industrial margarines used in the baking industry need specific functionality. This requirement may take precedence over nutritional factors since labeling of baked goods does not include the content of *trans* fatty acids, and blends used for this purpose frequently contain hydrogenated soybean oils.

Shortenings or baking fats used for biscuits, biscuit fillings, and various kinds of pastry and bread have to vary in their fat composition depending on their end use and generally consist of blends of two or more hard and soft fats, some or all of which may be hydrogenated. Soybean oil and hydrogenated soybean oil can be components of these blends.

Market Data

Soybeans are used as beans in some animal feeds, but for the most part, they are crushed and extracted, either in the country of origin or after export to another country. Production of soybean oil is therefore not confined to those countries where the beans grow. Extraction of soybeans gives ~18% of oil, and the balance is seed meal, which is also an important commodity of world trade. The main countries that produce, export, import, and use soybean oil (for food or nonfood purposes) are shown in Table 4. The figures cited are millions of tonnes for the year 1999/2000. In respect of production, USA, Brazil, Argentina, and China grow and extract their own beans, though China is also a large importer of beans for local crushing. The other countries listed in Table 4 crush imported beans. Exports from some of the producing countries/regions go to a wide range of geographical locations, of which only the larger importing countries/regions are listed. Soybean oil is used in virtually every country in the world. The major consumers include the producing countries, except that Argentina has only a small population so that only limited quantities of oil are consumed locally.

As the major grower of soybeans, the major producer of soybean oil, and the major consumer of the oil, the USA is a special case. The following information is taken from SoyStats published by The United Soybean Board (America). Edible oil consumed in the USA is almost entirely soybean oil, as is apparent from the following figures for 1999: total oil 6.93 million tonnes, including soybean (5.73), corn (0.23), rapeseed (0.22), cottonseed (0.18), lard (0.13), other (0.31). Thus, soybean oil represents 82.6% of US consumption, and no other oil exceeds 3.3%. The oil is consumed in a refined state, but much of it is hydrogenated to a greater or lesser extent before use. With brush hydrogenation, the level of linolenic acid is reduced to around 3%, but otherwise, the oil is hardly changed. With more extensive hydrogenation, linoleic and linolenic acids decline in level, and saturated acids and *trans* monoene acids increase. Soybean oil is consumed in the USA mainly in salad and cooking oils (~47%), for baking and frying (~36%), and in margarine production (~12%).

Table 4 Production, exports, imports, and consumption (million tonnes) in 1999/2000 by country/region

	Total	*Major countries/regions*
Production	25.21	USA 8.15, Brazil 4.24, Argentina 3.10, EU-15 2.87, China 2.20, Japan 0.69. Taiwan 0.41
Exports	7.15	EU-15 1.00, Argentina 3.00, Brazil 1.49, USA 0.58
Imports	7.21	Iran 0.82, China 0.80, India 0.74, Bangladesh 0.47, Pakistan 0.28, former USSR 0.24, Venezuela 0.25, Morocco 0.23
Consumption (food and nonfood use)	25.02	USA 7.42, Brazil 2.88, China 2.87, EU-15 1.88, India 1.55, Mexico 0.71, Japan 0.69, Iran 0.67, Taiwan 0.47, Bangladesh 0.46, South Korea 0.33, former USSR 0.33, Pakistan 0.32, Venezuela 0.29, Canada 0.29, Mexico 0.26

Source: *Oil World Annual* (2000) Hamburg: ISTA Mielke.

Soybean oil does not hold the same dominant position in other countries where there is a greater use of palm oil, rapeseed/canola oil, sunflower oil, and lauric oils.

See also: **Fatty Acids**: Properties; **Margarine**: Types and Properties; **Palm Oil**; **Rape Seed Oil/Canola**; **Soy (Soya) Beans**: The Crop; **Sunflower Oil**; **Vegetable Oils**: Types and Properties

Further Reading

Hamm W and Hamilton RJ (eds) (2000) *Edible Oil Processing*. Sheffield, UK: Sheffield Academic Press.

Official Methods and Recommended Practices of the American Oil Chemists' Society, Physical and Chemical Characteristics of Oils, Fats and Waxes (1996) Champagne, IL: AOCS Press.

Oil World Annual (2000) Hamburg: ISTA Mielke.

Reske J, Siebrecht J and Hazebroek J (1997) Triacylglycerol composition and structure in genetically modified sunflower and soybean oils. *Journal of the American Oil Chemists' Society* 74: 989–998.

Sipos EF and Szuhaj BF (1996) Soybean oil. In: Hui YH (ed.) *Baileys Industrial Oil and Fat Products*, 5th edn. vol. 2, pp. 497–602. New York: Wiley.

2000 SoyStats, United Soybean Board www.unitedsoybean.org

SOY (SOYA) BEANS

Contents

The Crop

H E Snyder, University of Arkansas, Fayetteville, AR, USA

Production

The soya bean originated in the Far East and has provided food for that part of the world for thousands of years, but only during the past few decades have soya beans become an important food crop in the West. **Table 1** shows how the production of soya beans has increased since 1940. The newly recognized value of the soya bean is in both the extracted oil and the residual high-protein meal. The oil is widely used as a food in margarines, shortenings, and salad oils. The meal is valuable for its high-protein content as an animal feed, particularly for poultry and swine.

Table 1 Soya bean production ($\times 10^6$ t) by major countries during the past 50 years

	1940	*1950*	*1960*	*1970*	*1980*	*1990*
USA	2.1	7.8	16.3	30.7	48.7	52.4
Brazil				2.1	15.5	20.0
China	9.7	4.9[a]	7.0	9.7	7.9	10.8
Argentina					3.4	10.5
World	11.9	17.8	28.1	45.2	81.2	107.1

[a]Data for 1948.

The USA, Brazil, China, and Argentina produce 87% of the world's soya beans. China has been a major producer for a long time, but the USA reached that status only in the 1940s, and Brazil and Argentina only since the 1970s.

In comparison with grain production, annual soya bean production of about 100×10^6 t is not large. World production of coarse grains (maize, sorghum, etc.) is about 800×10^6 t, while rice and wheat amount to about 450×10^6t each. Soya beans per unit weight are worth about twice as much as grains. (*See* **Cereals:** Contribution to the Diet.)

Soya beans are the predominant oilseed in the world. About half of the world oilseed production comes from soya beans, and this is more than the combined production of the next four oilseeds: cottonseed, peanut, sunflower, and rapeseed (canola). Oilseeds vary widely in total oil content; the annual vegetable oil production therefore differs from oilseed production. **Table 2** shows vegetable oil production for the world during the 1980s. Palm oil comes from palm fruit and is not classified as an oilseed, but palm oil is produced in the largest quantities next to soya

Table 2 World vegetable oil production ($\times 10^6$ t)

	1980	*1982*	*1984*	*1986*	*1988*
Soya bean	12.8	13.8	13.4	14.8	15.0
Palm	5.2	5.6	6.9	7.9	9.0
Sunflower	4.7	5.8	6.1	6.6	7.0
Rapeseed	4.1	5.3	5.3	6.6	7.3
Cottonseed	3.2	3.4	4.2	3.2	3.6
Peanut	2.7	3.0	3.4	3.1	3.8
Coconut	2.9	2.7	2.8	2.8	2.6
Olive	1.9	1.9	1.7	1.9	1.8
Palm Kernel	0.6	0.8	0.9	1.0	1.3

bean oil. Most of the vegetable oils are increasing in production to keep up with world demand, but cottonseed, coconut, and olive oil production are relatively stable. (*See* **Ground Nut Oil; Palm Oil.**)

The increased use of vegetable oils for foods parallels the growth of soya bean production. In the 1920s, animal fats were predominant in human diets, but since the 1940s, vegetable oils have steadily replaced animal fats. The particular vegetable oil that is used depends on the country. Soya bean oil makes up about 75% of vegetable oil use for food in the USA. Peanut oil is the oil of choice in India, and rapeseed oil is important in Canada.

Soya beans are annual legumes and were used at first in the USA as a hay crop. Since legumes are nitrogen-fixing plants, the hay crop could be used in rotation with corn to replenish soil nitrogen. With increased use of nitrogen fertilizers, the value of soya beans for their nitrogen fixation decreased.

Most of the spectacular recent increase in production in the USA and South America (Table 1) is due to increased area devoted to growing soya beans. There has been a slow but definite increase in yield per unit area, with an approximate doubling since 1940 to 2 t ha^{-1} in 1990 in the USA.

The problems that producers face in trying to increase yields of soya beans are the same as with any agronomic crop. Pests such as birds and rodents may eat planted seeds; diseases may slow down the growth of roots, stems, or leaves; insects can cause damage as well as transmit diseases; weeds compete for moisture, soil nutrients, and sunlight.

The flowers of soya beans are small and inconspicuous, making it difficult to produce hybrid seed. Under normal conditions, soya beans are self-fertilizing. The time of flowering is triggered by lengthening nights, and different soya cultivars (cultivated varieties) respond to different night lengths. Thus, it is important to select the proper cultivar for the latitude in which it will be grown. If a cultivar from southern latitudes is grown farther north, it will not flower until late in the season and may freeze before the seed matures. Conversely, northern cultivars grown in southern latitudes may flower before the plant has reached full growth, thus greatly reducing yield.

Morphology

Soya beans can range in weight from about 100 to 300 mg each, with diameters of 4–8 mm. They are roughly spherical in shape when dry and swell to a definite bean or kidney shape when wet. The morphological features of the whole bean (Figure 1) are the hilum (point of attachment to the pod), the micropyle (a small opening through which the germ tube grew), and the chalaza (a small groove opposite the hilum from the micropyle).

The seed coat makes up about 9% of the weight of soya beans and tightly encloses the two cotyledons and embryo. The seed coat is not easily separable from the intact dry seed, but if the seed is broken or if it imbibes water, the seed coat separates readily. The morphological features noted above (hilum, micropyle, and chalaza) are part of the seed coat, which

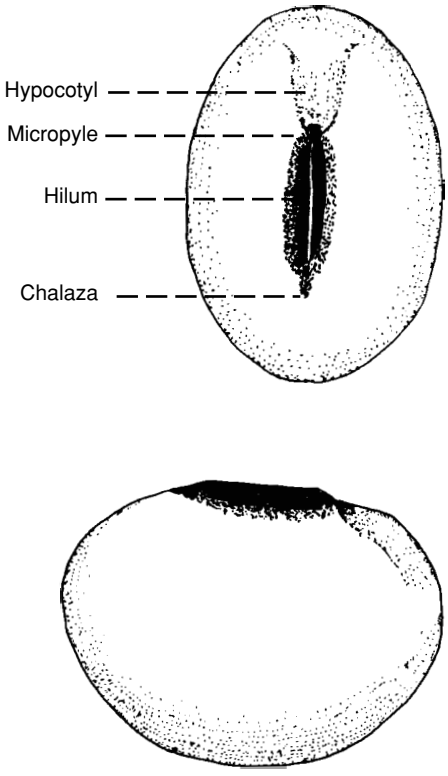

Figure 1 Morphological features of whole, mature soya beans. The features shown are associated with the seed coat or hull. Reproduced from Soya Beans: The Crop, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

may be yellow, green, brown, or black, but the cotyledons are green or yellow. United States grading standards discriminate against brown and black soya beans. Since very little oil is present in the seed coat, it is removed before oil extraction and may or may not be added back to the defatted meal used as animal feed.

Several layers of cells make up the seed coat, and cells of a distinctive type, termed hourglass cells because of their shape, are readily distinguishable by light microscopy. The hourglass cells can be used as qualitative markers to determine whether or not soya bean meal has been added to other foods.

If soya beans are soaked in cold water, a few, perhaps 1%, fail to imbibe water and are called 'hard seeded.' The failure to imbibe water is a property of the seed coat being completely intact and is probably due to the outermost waxy layer, termed the cuticle.

The two cotyledons make up 90% of the soya bean. The cotyledon cells are packed with protein bodies and lipid bodies, the two principal organelles (Figure 2). Early in the maturity of the cotyledons, starch granules are prominent, but they decrease to less than 1% of weight as beans mature.

Protein bodies are 2–20 μm in diameter and contain principally the glycinin and conglycinin storage proteins. Protein bodies are osmotically fragile but can be isolated by using high-osmotic-strength buffers at pH 5. In addition, heating (soaking soya beans for a few minutes in boiling water) causes protein bodies to be heat-fixed and not to disperse in water.

Lipid bodies (0.2–0.5 μm) are much smaller than protein bodies and are the sites of oil storage. Centrifugation of broken cotyledon cells separates lipid bodies as a floating layer, but because of their phospholipid-protein membrane, the smallest lipid bodies sediment in a centrifugal field.

The hypocotyl of the soya bean has a different composition to that of the cotyledon (less oil and more carbohydrate), but since it makes up only about 2.5% of the weight of the bean, the different composition has little influence. The hypocotyl is processed into oil and meal along with the cotyledons after the seed coat is removed.

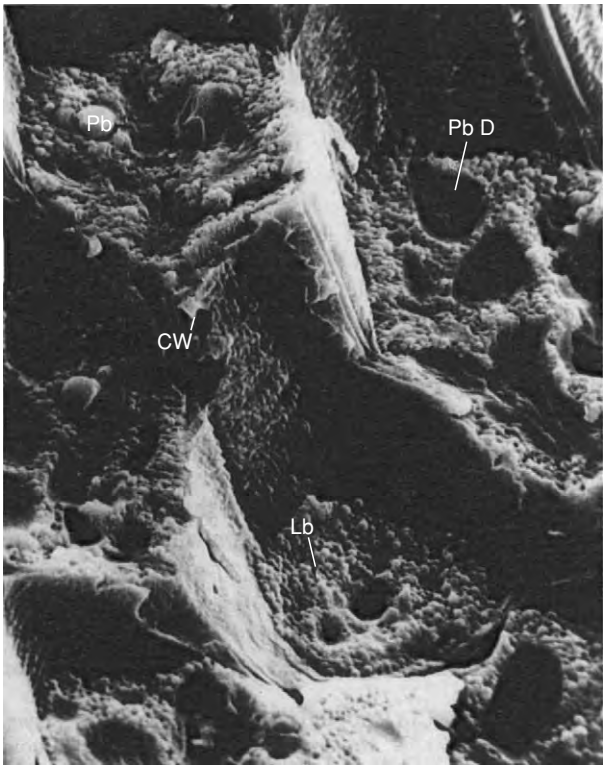

Figure 2 Scanning electron micrograph of freeze-fractured soya bean cotyledon. CW, cell wall; Lb, lipid body; Pb, protein body; PbD, protein body depression. Reproduced from Soya Beans: The Crop, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Composition

As a first approximation, soya beans contain 20% lipid, 40% protein, 35% carbohydrate, and 5% ash on a dry-weight basis. Considerable variability exists in these numbers, depending on the cultivar and the growing conditions. In the northwestern part of the soya bean production area in the USA, protein content of soya beans is generally lower than in other production areas. Protein content is known to be inversely related to yield and to lipid content.

Lipids

The crude oil extracted from soybeans contains about 96% triglycerides. The fatty acid composition of the triglycerides is approximately 11% palmitic (C16), 4% stearic (C18), 25% oleic (C18:1), 51% linoleic (C18:2), and 8% linolenic (C18:3), with other minor fatty acids also present. The fatty acid composition is not fixed and can be changed by breeding new cultivars. (*See* **Fatty Acids**: Properties.)

Minor components in the crude oil include the plant pigments lutein (3,3′ dihydroxy α-carotene) and chlorophyl. Also present are phospholipids, free fatty acids, triglyceride hydroperoxides, sterols, saponins, squalene, tocopherols (vitamin E), and flavor compounds. During refining of the crude oil, phospholipids, free fatty acids, pigments, and flavor compounds are removed to produce an oil that contains more than 99% triglyceride. (*See* **Chlorophyl**; **Colorants (Colourants)**: Properties and Determination of

Natural Pigments; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Phospholipids**: Properties and Occurrence; **Triglycerides**: Structures and Properties.)

Protein

Soya proteins have been separated in the ultracentrifuge and characterized as 2*s*, 7*s*, 11*s*, and 15*s* fractions (*s* is a sedimentation constant, and larger numbers indicate larger proteins). The 2*s* fraction contains low-molecular-weight proteins, including the Bowman–Birk and Kunitz trypsin inhibitors that inhibit growth in young animals. Heating of soya bean meal is necessary to denature trypsin inhibitors and allow full utilization of the protein. (*See* **Protein**: Chemistry.)

The 7*s* fraction includes a storage protein, β-conglycinin, with a molecular weight (M) of about 180 000. β-Conglycinin is a trimer with α, α', and β subunits, which, in different combinations, give six distinct β-conglycinins. Also present in the 7*s* fraction are hemagglutinins (lectins) and lipoxygenases. Hemagglutinins are glycoproteins that can cause agglutination of red blood cells and can be toxic, but are probably heat-denatured along with trypsin inhibitors and have no toxic effects when fed. Several lipoxygenases have been characterized that catalyze the removal of hydrogen and addition of molecular oxygen to linoleic and linolenic fatty acids to yield hydroperoxides. The hydroperoxides can break down to yield off-flavor compounds associated with oxidative rancidity. (*See* **Oxidation of Food Components**.)

The 11*s* fraction is glycinin, a hexamer (M, 360 000). Each of the six subunits consists of acidic polypeptides (M, 32 000–36 000) and basic polypeptides (M, 19 000–20 000). The 15*s* fraction is a dimer of glycinin.

Associated with protein in soya beans is phytin, the calcium and magnesium salt of hexaphosphoryl inositol. Phytin strongly chelates divalent cations and has been implicated in decreased availability of iron in diets with soya protein. (*See* **Phytic Acid**: Properties and Determination.)

Carbohydrate

Insoluble carbohydrate includes the pectin, cellulose, and hemicellulose that are associated with cell walls in soya bean cotyledons. This is also the dietary fiber component, which is increasingly recognized as an important part of the human diet. The fiber is more prevalent in soya bean seed coats (hulls) than in cotyledons. (*See* **Carbohydrates**: Classification and Properties; **Cellulose**; **Hemicelluloses**.)

The soluble carbohydrate in soya beans is about 10% by weight of the dry bean and includes about 5% sucrose, about 4% stachyose, and about 1% raffinose. Raffinose is a trisaccharide, and stachyose is a tetrasaccharide, neither of which can be digested by humans. As a result, microbial fermentation of these sugars in the intestine produces gas (flatus), which leads to gastrointestinal distress.

Handling and Storage

As soya beans are harvested and moved from the field into storage, conditions have to be controlled to minimize deterioration. The moisture content of the beans should be 14% or lower to prevent microbial growth. Also, the cleanliness of the beans is important to avoid insect or other contaminants, which may provide moisture and be a focus for microbial growth. If deterioration does start, bacterial or mold growth will generate higher temperatures and moisture so that the process becomes autocatalytic. The best control procedure in storage is to have temperature sensors spread throughout, and quickly mix beans that show generation of hot spots. (*See* **Spoilage**: Bacterial Spoilage; Molds in Spoilage.)

Much of the crop from the USA and Argentina is exported as beans. As the beans are transported by truck, train, barge and ship, and they are transferred into and out of storage, they tend to break, and the pieces are known as splits. Splits are a factor in grading, and the oil extracted from splits tends to have greater amounts of free fatty acids and phospholipids than oil from intact beans.

Much of Brazil's crop is crushed in the country, and oil and meal are exported. In handling oil, it is necessary to minimize exposure to oxygen of the air during transfer. Any outlet spouts should be below the surface to prevent air incorporation in the oil, and blanketing the surface with nitrogen is a good protective measure. Storage tanks must be kept clean. This is particularly important for the tanks in ship's holds or for tankers, both of which may carry many different liquid cargoes. Cross-contamination between cargoes is avoided by careful cleaning of the tanks.

Grading

In the USA, there are four numerical grades and sample grades for those beans that have the lowest quality. The factors used for grading are test weight, splits, heat damage, foreign material, and color. These are similar criteria to those used for grains, and the grading systems in other countries rely on the same criteria, although the actual grades and maximum limits may differ. The economically valuable constituents, protein and oil, are not part of the grading system.

Until recently, there have been difficulties in measuring protein and oil quickly and accurately enough to be part of the grading system. Infrared spectrophotometers now perform excellent analyses for protein, oil, and moisture of either whole or ground samples. Consequently, there is a renewed interest in using the protein and oil content of soya beans to provide premiums or discounts in pricing. Aside from the analysis problem, other problems exist in keeping high-oil or high-protein beans segregated as they move through marketing channels.

See also: **Carbohydrates**: Classification and Properties; **Cellulose**; **Chlorophyl**; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Fatty Acids**: Properties; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Hemicelluloses**; **Oxidation of Food Components**; **Phospholipids**: Properties and Occurrence; **Phytic Acid**: Properties and Determination; **Protein**: Chemistry; **Spoilage**: Bacterial Spoilage; Molds in Spoilage; **Triglycerides**: Structures and Properties

Further Reading

Anonymous (1990) *Soya Bluebook*. Bar Harbor, ME: Soyatech.

Erickson DR (ed.) (1990) *Edible Fats and Oils Processing: Basic Principles and Modern Practices*. Champaign, IL: American Oil Chemists' Society.

Erickson DR, Pryde EH, Brekke OL, Mounts TL and Falb RA (eds) (1980) *Handbook of Soy Oil Processing and Utilization*. St Louis, MO: American Soybean Association. Champaign, IL: American Oil Chemists's Society.

Heatherly LG and Hodges HF (eds) (1999) *Soybean Production in the Midsouth*. New York: CRC Press.

Snyder HE and Kwon TW (1987) *Soybean Utilization*. New York: Van Nostrand Reinhold.

Wilcox JR (ed.) (1987) *Soybeans: Improvement, Production and Uses*, 2nd edn. Madison, WI: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America.

Processing for the Food Industry

H E Snyder, University of Arkansas, Fayetteville, AR, USA
L A Wilson, Iowa State University, Ames, IA, USA

Background

Soya beans are seldom simply cooked and eaten. They are traditionally processed into soyafoods or into food ingredients. This article covers soyfood processing and the initial separation of soybeans into oil and meal by solvent extraction and then the further processing of oil and meal. Oil needs to be refined to remove minor components and is then processed to make products such as margarines, shortenings, and salad dressings. The protein-rich meal is mainly used as animal feed but can be processed into food ingredients such as flours, concentrates, and isolates. The protein products may also be textured in various ways to simulate meats or cheeses.

Oil Separation

The first step in processing most soya beans is to separate the oil, either by solvent extraction or by expelling. The processing companies that do the separation are called crushers, and each of the 100 or so crushers in the USA processes on average about 1000 tonnes of soya beans daily. (*See* **Vegetable Oils**: Oil Production and Processing.)

Preparation

Before solvent extraction, soya beans are cleaned and cracked into several pieces (meats). The hulls are removed by aspiration, and the meats are conditioned by warming and by adding moisture. Conditioning is necessary to make a cohesive flake. The conditioned meats are put through smooth rollers that make flakes of approximately 0.025 cm thickness. Making flakes is advantageous for uniform penetration of solvent in deep beds (minimal channeling) and for disruption of the soya bean tissue so that solvent can penetrate and dissolve the oil.

A recent development is to put the flakes through a cooking extruder (expander) to yield collets. The collets are pieces of the 'rope' exiting the outlet of the expander. This processing gives a porous but still high-density collet that extracts more readily than flakes. Also, the collet holds less solvent that the flake, thereby minimizing the energy needed for removing the solvent.

Solvents

The solvent of choice is commercial hexane, a petroleum fraction with a boiling range of 65–70 °C. Commercial hexane is a mixture of *n*-hexane, cyclohexane, and methylpentanes. It has low viscosity and a low heat of vaporization. The main disadvantages of hexane are its flammability and the safety precautions needed for safe handling.

Chlorinated hydrocarbons are equally good solvents for soya bean triglycerides, but the use of trichloethylene led to toxic soya bean meal during the 1940s, and this has discouraged further use of these solvents.

Ethanol and isopropyl alcohol are effective triglyceride solvents when hot, and cooling can be used to separate the dissolved triglyceride from the solvent. Isoflavones are also extracted by alcohol. However, alcohol solvents are not being used in commercial extractions.

Extraction

The full-fat flakes (or collets) are loaded into the extractor to make beds over which the solvent flows. Extractors may have deep (1.2 m) or shallow (0.6 m) beds, and are always arranged so that solvent flows countercurrent to the movement of the beds. Thus, the fully extracted flakes are contacted by fresh hexane entering the extractor, and full-fat flakes are contacted by the full miscella (solution of crude oil in hexane) just before it leaves the extractor. The temperature of extraction is about 60 °C to speed up diffusion of solvent and to lower the miscella viscosity, both of which enhance the extraction of oil. Solvent extraction is capable of reducing the residual oil in the soya flakes to less than 1%.

Removal of Solvent

Upon completion of extraction, the solvent has to be removed from both the oil and flakes. The full-fat miscella contains 25–30% oil, and solvent is removed by two stages of rising-film vacuum evaporators followed by a third-stage stripping column. At the end of this process, the crude oil contains about 1000 p.p.m. of hexane, which corresponds to a flash point for the oil of 121 °C.

The flakes are treated in a desolventizer-toaster by direct contact with steam first to remove the hexane and second to heat treat the flakes for trypsin inhibitor destruction. After leaving the desolventizer-toaster, the flakes are cooled and ground to a meal for use as a high-protein feed ingredient. The protein content is 44% with hulls added or 47.5–49% protein without hulls.

If flakes are to be used to produce soya bean protein isolates or soluble soya bean concentrates, the solvent has to be removed with minimal heat to maintain protein solubility. Flash desolventizers are available in which superheated hexane is used as the heat-transfer medium to evaporate hexane. The flakes are kept dry and, with this system, protein solubility is preserved.

Expelling

The original methods for recovering oil from oil seeds were designed to exert pressure on the seeds and express the oil. The modern technique for exerting pressure is an expeller: a motor-driven auger rotating within a cage of narrowly spaced metal bars. As the oil seed flakes move through the expeller, pressure exerted by the auger forces oil out through the bars, while the defatted flakes are pushed by the auger to the opening at the end. Considerable heat is generated by expellers, and often the protein meal is of a lower quality than the solvent-extracted meal because of excess heat. Also, the residual oil in expeller meal is 3–4%. Nevertheless, expellers are widely used when it is necessary to handle a variety of oil seed in one facility. Expellers can handle as much as 50–80 tonnes of oil seeds per day. The soya bean preparation for expelling is similar to that for solvent extraction. Recent interest in extruding or expressing has increased the number of small operations that can produce 'crude' oil and reduced-fat flour.

Soya Oil Refining

After extraction and removal of solvent, the crude soya bean oil needs to be refined to convert it to edible products. (*See* **Vegetable Oils**: Oil Production and Processing.)

Degumming

The crude oil contains about 1–3% phospholipids, known in the industry as gums. The phospholipids act in conjunction with tocopherols as antioxidants, but they will also precipitate out of the oil, making a sludge that is difficult to remove from storage tanks, road tankers, or ship holds. Therefore, crude oil is often degummed by the crusher before shipping or storing.

The degumming process involves adding 1–3% water to hydrate the phospholipids, followed by centrifuging to remove the hydrated material. The phospholipids are removed in the water phase and either recovered to be sold as a food ingredient, lecithin, or added to the meal portion to be sold as animal feed. Phospholipids remaining in the oil are said to be nonhydratable, and are largely calcium and magnesium salts of phosphatidic acid. An acid treatment of the crude oil makes these phospholipids removable by water washing. About half of the soya bean oil in the USA is degummed before going to the refiner.

Alkali Refining

The purpose of alkali refining is to remove free fatty acids from the crude oil (if the oil has not already been

degummed, phosopholipids are also removed by alkali refining). Free fatty acids are undesirable because they lower the temperature at which the oil begins to smoke and flash (ignite) when heated, as in frying. The fatty acid content of crude soya bean oil is 0–0.7%, and alkali refining yields oil with < 0.05% free fatty acids.

The process washes the oil with a 12% sodium hydroxide solution followed by centrifugation. Fatty acids are converted to sodium soaps and separate with the water phase. Invariably, some oil is lost with water phase (neutral oil loss), and the greater the amount of free fatty acids, the greater the neutral oil loss. After washing with dilute alkali, the oil is washed once with water to remove residual soaps and vacuum-dried to a moisture content of 0.1%.

The fatty acids may be recovered by acidifying and centrifuging the water phase. Fatty acids can be used for soap manufacture but are often added to the soya bean meal to serve as additional feed energy.

Bleaching

Crude soya bean oil is darker in color than desired by consumers, and bleaching is used to lighten the color. Bleaching is achieved by the addition of bleaching earth (adsorbent clays usually acid-treated) to the hot oil under vacuum. The main pigments in soya bean oil are carotenoids and chlorophyll, and these are adsorbed on bleaching earths, which are separated from the oil by filtration.

In addition to removal of pigments, the bleaching process adsorbs oxidation products and the prooxidant metals, iron and copper. Thus, bleaching plays an important role in stabilizing the oil regarding oxidation as well as lightening the color.

Hydrogenation

Soya bean oil is highly unsaturated and, to make products such as shortening and margarines, hydrogen is added to double bonds in the triglycerides to increase their melting points. Hydrogenation is also used to saturate partially double bonds in soya bean oil, so that the oxidative stability is increased. (*See* **Vegetable Oils**: Oil Production and Processing.)

Hydrogenation is usually done in batches (although continuous equipment is available), and the requirements are a nickel catalyst, hydrogen gas under pressure, vigorous agitation, and soya bean oil at 120–200 °C. The temperature, pressure, degree of agitation, and amount of catalyst are variables that affect the rate of hydrogenation and the degree of selectivity. Selectivity refers to the increase in hydrogenation rate of linolenic acid to linoleic acid to avoid off-flavor generation from oxidation of linolenic acid.

During hydrogenation, isomerization of fatty acids takes place to give double bonds in new positions and to shift double bonds from *cis* to *trans*. This process reduces the amount of ω-3 fatty acids (linolenic). Melting point increases result both from saturation of double bonds and from isomerization of *cis* to *trans* double bonds.

Deodorization

The development of hydrogenation and deodorization in the early 1900s made it possible to substitute vegetable oils for animal fats in human diets. Hydrogenation controlled the texture, and deodorization controlled the flavor of vegetable oils.

Deodorization is steam distillation at a high temperature under a vacuum. By this process, unwanted flavors are removed, and some bleaching of remaining carotenoid pigments occurs. The distillation temperatures range from 204 to 275 °C at 6 mmHg vacuum. Under these conditions, not only flavor compounds but free fatty acids and about half of the tocopherols are distilled out of the soya bean oil.

If the oil to be deodorized is sufficiently devoid of phospholipids (less than 5 p.p.m. phospholipid phosphorus), free fatty acids can be removed by distillation rather than by alkali refining. Such removal is known as physical refining. The phospholipids must be present in low concentrations because they brown severely at high temperatures.

The distillate remaining from deodorization is a valuable byproduct because of the relatively large content of tocopherols (vitamin E). Batch, semicontinuous, and continuous equipment is available for deodorization.

Soya Oil Products

After sufficient refining, soya bean oil is used to produce shortening, margarines, salad dressings, and cooking oils of various types. (*See* **Dressings and Mayonnaise**: The Products and Their Manufacture; **Fats**: Uses in the Food Industry; **Margarine**: Methods of Manufacture.)

For plastic products such as shortenings and margarines, the texture is achieved by a mixture of triglyceride crystals plus oil. To avoid a large crystal size and grainy texture, it is necessary to have a mixture of fatty acids. For soya oil products, this means adding oils that contribute palmitic acid to achieve small crystals. In addition to the familiar plastic shortenings produced for home use, dry shortenings are manufactured for incorporation into various bakery mixes,

and liquid shortenings are used for ease of metering in automated baking processes.

For margarines, the solids content at various temperatures (solid fat index) is an important quality criterion. If the margarine is to be used in the home, manufacturers want a product that spreads easily at refrigerator temperatures and melts completely in the mouth (body temperature). These texture characteristics are achieved by blending hard and soft fats and oil. The crystal formation for both shortenings and margarines is achieved by cooling mixtures rapidly while mixing in scraped-surface heat exchangers.

Cooking oils based on soya oil are usually partially hydrogenated to minimize the linolenic acid content and to minimize oxidative rancidity in the hot oil.

Soya Bean Protein Products

Although an estimated 90% of the soya bean defatted protein meal goes to animal feed, a variety of food products and ingredients are produced from the remainder. These include soya bean flours, ranging from full fat to defatted, soya bean protein concentrates, soya bean protein isolates, and various textured products simulating meats, seafoods, and cheeses.

Soya Bean Flours

By grinding defatted soya bean flakes to flours (or grits, if the particle size is larger), a product is produced that is used in the baking industry for its moisture sorption capability. Soya bean flours and grits are also used in meat emulsions for the emulsifying ability in addition to moisture sorption. (*See* **Emulsifiers**: Uses in Processed Foods.)

Full-fat flours are produced by grinding dehulled soya beans. The unheated flours have active lipoxygenase and are used to bleach wheat flours used in the baking industry. The oxidative action of lipoxygenase both bleaches the carotenoids in wheat flour and oxidizes the proteins to improve machinability. Full-fat flours can be heat-treated to produce enzyme inactive flours where less 'beany' flavors with more cereal/nutty flavors are desired.

A range of soya bean flours between defatted and full fat can also be produced by adding soya bean oil and lecithin in varying amounts to defatted flours. These products are used mainly in the baking industry.

If the soya bean flours are used in substantial concentration in food products, soy beany flavors can become a problem. The presence of indigestible oligosaccharides, raffinose and stachyose, that cause flatus can also be a problem. To avoid the flavors, heated soyflours and lipoxygenase-null soybeans have been developed. Reduction of the oligosaccharides has been accomplished by the development of soya bean protein concentrates. (*See* **Soy (Soya) Beans**: Processing for the Food Industry.)

Soya Bean Protein Concentrates

The processing of defatted soya bean flakes into concentrates requires the protein to be rendered insoluble and extracting the soluble sugars (sucrose, raffinose, and stachyose) with water. The protein can be made insoluble by heat treatment, by aqueous ethanol treatment, or by adjusting the pH to the isolectric point of 4.5. The water extraction leaves complex carbohydrates and protein, and the protein concentration has to be 70% to be classified as a concentrate. The use of water extraction retains more of the isoflavones than traditional ethanol extract concentrates. Both types of concentrates are currently in the market place.

In some food applications (a high-protein drink, for example), soluble protein may be needed, and this can be achieved by using a pH to render the protein insoluble followed by pH increase after extracting the sugars. Protein solubility is measured by the nitrogen solubility index (NSI) or protein dispersability index. The indices give the percentage protein remaining in solution or suspension after centrifugation. (*See* **Protein**: Chemistry.)

Soya bean protein concentrates are used in the baking and meat industries for their moisture sorption and emulsifying properties.

Soya Bean Protein Isolates

For those food applications in which only soya bean protein is suitable, soya bean isolates are available. Isolates contain 90% soya bean protein and are prepared from defatted flakes with a high NSI. The protein is dissolved with dilute alkali, separated from the remaining cellular material, and precipitated by adjusting the pH to 4.5. The protein can be recovered as spray-dried isolectric protein (no pH adjustment and low solubility) or as soluble proteinate after raising the pH. About one-third of the original flake is recovered as protein isolate.

Soya bean protein isolates can be used for moisture sorption and emulsification in baked goods and meat emulsions. They can also provide textural improvement through gelation and adhesion properties or serve as primary nutrients in infant formulae.

Soya bean flours, concentrates, and isolates may modify texture in foods by increasing viscosity, by holding moisture or fat, or by adhesion and cohesion.

Furthermore, texture in foods can be modified by changing the texture of the soya bean materials.

Texturing

Numerous ingenious methods have been patented to add chewy textures into flours, concentrates, and isolates. The most commonly used procedures are extrusion and spinning.

Extrusion is used mainly with flours and concentrates and is a method of both heating and forming. The extruder is a screw or auger rotating within a jacketed barrel. As the flour or concentrate is moved through the auger, it is heated by friction and/or by steam jacketing. High pressures are developed in the nose of the barrel just before the exit. As the material is forced through the exit port, it is shaped and puffed due to the sudden drop in pressure.

Spinning applies only to soya bean isolates. A concentrated solution of isolate is prepared and forced through spinnerets (platinum plates with small holes) into an acid bath. The protein precipitates in the acid as a continuous fiber, and the spun fibers can be wound together into a bigger 'tow.'

With both extruded and spun soya bean products, flavors, colors, fats, etc., can be added to simulate meats or seafoods.

Soyfood Manufacture

Traditional soybean foods are classified into two groups: nonfermented foods, including regular, deep-fried, frozen-dried, roasted tofu, soybean protein film (yuba), and soybean sprouts; and fermented soybean foods such as miso, soy sauce (shoyu), and fermented whole soybeans (natto) and Tempeh.

Soymilk

Soymilk is a very popular beverage with the Chinese, but considerably less so with the Japanese and Western populations. However, soymilk production has increased in the USA due to lipoxygenase-null soya beans and improved processing methods that remove or limit the 'beany' flavor of the milk. The health benefit of soy protein to reduce cholesterol (heart disease) has also caused an increase in soymilk sales (and soya used in foods) as the US FDA now allows a health claim if the food contains at least 6.25 g of soy protein with low levels of cholesterol, saturated fat, and total fat. Soymilk is very important to tofu producers because it is the intermediate product in the manufacture of tofu.

Many Japanese, like their Western counterparts, find the odor of soymilk undesirable. This odor is formed by the oxidation of specific unsaturated fatty acids by the enzyme lipoxygenase during seed grinding. Understanding how off-flavors and undesirable flavors interact with soy proteins and similar seed constituents may lead to improved processing systems. Breeding programs at Purdue University (West Lafayette, IN), Iowa State University (Ames, IA), and the Japanese Ministry of Agriculture, Forestry, and Fisheries (Tsukuba, Japan) have developed lipoxygenase-null varieties.

Additionally, odor formation may be circumvented by heat inactivation of the enzyme before the beans are ground (significantly lowering yields), or by masking the flavor with additives. (*See* **Soy (Soya) Beans: Properties and Analysis**; **Soy (Soya) Milk.**)

Soymilk is traditionally made by soaking soybeans in water overnight, then grinding the beans with water added during grinding. Alternatively, full fat flakes, grits, or flour can be used to produce the soymilk slurry. The resulting slurry is boiled and stirred for 1–30 min (depending on the temperature). This heating step improves the nutritional value of the milk (by inactivating trypsin inhibitors) and improves the flavor (by inactivating lipoxygenase and volatizing some of the off-flavor compounds that result during grinding). Heating also increases the shelf-life of the milk by reducing its microbial load. The heated slurry is then filtered through a cloth or nylon bag to separate the undispersible fiber residue (okara) from the soymilk. The resulting soymilk may then be flavored, if desired. It may also be pasteurized, homogenized, or sterilized before being bottled, aseptically packaged, or retorted.

Typically, high-protein, clear or yellow hilum, large-seeded soybeans are preferred for soymilk production. Two hundred grams of soybeans will yield about 1 l of soymilk. Soymilk can be made more shelf-stable by spray-drying or roller-drying it into a dry powder (as is done with cows' milk). Spray-dried soymilk can be instant in beverages, confections, and meat fillers.

Although soymilk has a composition similar to cow's milk, the oil-to-protein ratio is lower. For this reason, reconstituted soymilk is often supplemented with oil. In addition, soymilk may be supplemented with sugar to enhance its palatability. These soymilk beverages often contain fruit juice, cocoa, flavors (artificial or natural), stabilizers, and other ingredients to enhance customer acceptance by masking soy flavors.

Tofu

Tofu has been produced in Japan for more than 2000 years. It was introduced from China along with the agronomic introduction of soybeans. The shelf-life of tofu can be quite variable, ranging from 1 to 5 days

for fresh tofu, 3 weeks to 1 month for pasteurized tofu in the package, and 6 months to 2 years for tofu processed aseptically. The initial microbial load and storage temperature largely govern the shelf-life. Some products (such as silken tofu) are more difficult to transport because they are susceptible to temperature abuse and physical damage.

Large-scale factories in Japan consume 2–3 tonnes of soybeans per day. These large manufacturers have developed integrated production and marketing systems. From the factory to the supermarket showcase, superior product quality is ensured by sophisticated, timely distribution using refrigerated transport and display systems. Likewise, in Japan, some supermarket chains are now producing their own tofu in-house, where it is kept in refrigerated display cases. Aseptically packaged tofu gained popularity in the USA because it has a longer shelf-life and can be stored and transported easily.

Different types of tofu can be made from the soymilk by altering the starting materials (beans, full fat flakes, etc.), solids content, temperature of coagulation, type of coagulant, breaking of the curds, and pressing. Firm tofu (called momen in Japan) is more cheese-like in texture, whereas Kinugoshi (silken or soft) tofu has a much more homogenous, delicate texture, is softer, and has a smoother feeling than regular momen tofu. Reduced fat, grilled, backed, and flavored tofus can be made by altering the starting material or by grilling, baking, and flavoring. Packed tofu differs from other forms because it is processed within its sealed retail container. Aseptically packaged tofus are packaged tofu products made from all ingredients commercially sterilized before aseptic packaging. Deep-fried tofu includes three main types: namage (single-fried tofu), aburage, and gan-modoki (double-fried tofus). Kori tofu is a frozen, dehydrated tofu. Upon rehydration, its texture is very different from the original tofu in that it is more meat-like in texture, and it acts as a sponge in soups, etc.

Yuba

Yuba is a soy protein film produced from the surface of cooling soymilk. Some consumers use fresh film for cooking, but others prefer dried yuba. It is packaged in sheet form, small rolls, or pieces of various shapes and sizes, and it can be used in a variety of food applications.

Texturized Soy Protein Foods

Texturized soy protein products, made by texturizing whole soybeans, soy flour, concentrates, or isolates with single and twin-screw extruders are used to make meat analogs and snacks.

Fermented Soyfoods

Fermented soyfoods usually contain salt and the byproduct of a desirable fermentation. They have a relatively long shelf-life compared with fresh soy products like traditionally prepared momen tofu.

Miso

Miso is made in Japan by mixing cooked soybeans with rice koji (fermented rice started culture) and salt water. This material is then fermented for several months. There are several miso products that differ in the type of koji starter (a high ratio of rice to soybeans results in a more lightly colored and sweeter miso) used for the fermentation. Miso is used in soups, as a spice, and in coatings for vegetables and meat products.

Soya Sauce (Shoyu)

Soy sauce (shoyu) manufacturing was modernized much earlier than other soy foods. Defatted soya flakes are used instead of whole soyabeans because the soybean oil separates and collects on the surface of the sauce. Both fermentation and acid hydrolysis are used to produce different types of soy sauce that are now commonly available around the world.

Natto

Natto is cooked whole soybean product fermented with *Bacillus natto*. Small soybeans are often preferred for natto production. The soybeans are fermented until a white, sticky glutamic acid polymer is formed, which gives the natto a distinctive flavor and sticky and stringy characteristics.

Tempeh

Tempeh is a fermented whole soybean product that originated in Indonesia but is equally popular in Malaysia. Tempeh is usually cooked before it is eaten. Preparation usually involves frying, deep-fat frying, or baking the product. It is also added to soups and fast foods, and is used as a meat replacement in main dishes.

See also: **Dressings and Mayonnaise**: The Products and Their Manufacture; **Emulsifiers**: Uses in Processed Foods; **Fats**: Uses in the Food Industry; **Margarine**: Methods of Manufacture; **Protein**: Chemistry; **Soy (Soya)**

Beans: Processing for the Food Industry; Properties and Analysis; **Soy (Soya) Milk**; **Vegetable Oils**: Oil Production and Processing

Further Reading

Erickson DR (ed.) (1990) *Edible Fats and Oils Processing: Basic Principles and Modern Practices*. Champaign, IL: American Oil Chemists' Society.

Erickson DR (ed.) (1995) *Practical Handbook of Soybean Processing and Utilization*. Champaign, IL: AOCS Press and St. Louis, MO: United Soybean Board.

Erickson DR, Pryde EH, Brekke OL, Mounts TL and Falb RA (eds) (1980) *Handbook of Soy Oil Processing and Utilization*. St. Louis, MO: American Soybean Association and Champaign, IL: American Oil Chemists' Society.

Kinsella JE and Soucie WG (eds) (1989) *Food Proteins*. Champaign, IL: American Oil Chemists' Society.

Liu, KeShun (1997) *Soybeans: Chemistry, Technology, and Utilization*. New York: Chapman & Hall, ITP Publishing.

Snyder HE and Kwon TW (1987) *Soybean Utilization*. New York: Van Nostrand Reinhold.

Weiss TJ (1983) *Food Oils and their Uses*, 2nd edn. Westport, CT: AVI.

Properties and Analysis

E G Hammond, P A Murphy and L A Johnson, Iowa State University, Ames, IA, USA

Background

Soya beans vary widely in their appearance and composition. The main constituents of soya beans, in descending order, are protein, oil, complex carbohydrates, oligosaccharides, simple sugars, and minerals. Phytochemicals in soybeans have gained prominence in the last 10 years, especially isoflavones, saponins, and phenolic acids, because of their potential health-protective effects. Soya isoflavones in particular have been associated with a variety of other improved physiological states, including reduced cholesterol levels, anticarcinogenicity, and improved bone health.

Most legumes contain 20–25% protein, but soya beans typically contain 30–45% protein (moisture-free basis), and average 35.35% at 13% moisture. Levels as high as 55% protein (moisture-free basis) have been observed. The oil content typically ranges from 15 to 24% and averages 19.16% on a 13% moisture basis. The composition varies with growing area; for example, soya beans from northern and midwestern areas of the USA typically contain 1.5–2% less protein and 0.2% more oil than beans grown in southern states. One Iowa soybean processor has begun offering premiums for soybeans containing higher-than-normal oil contents.

The total crude fiber content of soya beans is about 4.4% at 13% moisture. These materials are predominantly cellulose, hemicellulose, and pectin. The outer hull, typically 8% by weight of the bean, is especially rich in crude fiber (35%). Fiber is hard to digest and contributes little to the nutrition of swine and poultry, the primary markets for soya bean meal.

The total sugar content of soya beans is 4.9–9.5% at 13% moisture. Of the total sugar content, about 60% is sucrose, 10% raffinose, and 30% stachyose. Raffinose and stachyose are oligosaccharides, causing flatulence (intestinal gas) in humans and reduced feed efficiency in livestock.

Isoflavone content of soya beans varies from 800 to 4000 $\mu g\,g^{-1}$ on an aglycon basis. Soya contains 12 isoflavones, namely genistein, daidzein and glycitein, as β-glucosides, 6″-*O*-malonylglucosides, 6″-O-acetylglucosides and aglycons. Soya saponin concentrations vary in a similar manner to isoflavones ranging from 2 to 12 $\mu mol\,g^{-1}$.

Because of the variety of glucoside forms of saponins and isoflavones, their concentrations should be expressed in micromoles. The isoflavones and saponins fractionate with the seed protein and are not lipid-soluble. As soya beans are fractionated into food ingredients, the forms of the isoflavones and saponins are rearranged depending on the extent of heat processing and the method of protein fractionation.

Generally soya bean meal is sold on a 44% protein basis and 48% when soya beans are dehulled (Hi-Pro meal). The average price of soybeans during the 1997–98 crop year was $6.47 per bushel (0.259 kg). Oil sells for about twice the value of meal. Typical selling prices for soya bean oil were 25.7 cents per pound (0.566 kg) and meal (44% protein) for $197 per ton or 9.9 cents per pound (0.566 kg) during 1998. The 2000 crop year prices were considerably lower.

Recent developments in breeding and genetic modification of soya beans have achieved: new fatty acid compositions, e.g., $<2\%$ linolenic acid and $>80\%$ oleic acid to improve oxidative stability; removal of all lipoxygenase isozymes to improve the flavor of soya protein ingredients; elimination of oligosaccharides to eliminate intestinal gas; altered protein and amino acid composition to improve functional and nutritional properties; and inserted

vaccines for human and animal health. While genetic engineering holds great promise for the future (healthier foods, reduced environmental impacts, increased yield, and lower input costs), consumers have begun to question the safety of this new technology.

Proteins

Glycinin and β-conglycinin constitute 65–80% of the protein fraction or 25–35% of the seed weight. Glycinin is classified as a legumin, characterized by molecular weights of 300–400 kDa and sedimentation coefficients of 11S ± 1S. β-Conglycinin is a vicilin, which is in the range of 150–250 kDa, is glycosylated, and has sedimentation coefficients of 7S ± 0.5S. In soya beans, the major proteins, glycinin and β-conglycinin, are frequently described by their respective sedimentation values, 11S and 7S, but such fractions are often impure. The 7S fraction of soya protein contains, in addition to β-conglycinin, lectins, lipoxygenase, and β-amylase.

β-Conglycinin is a trimer and/or hexamer in solution and probably occurs in both forms in the seed. Two similar peptides, α and α′ (57 kDa), and a glycosylated β peptide (42 kDa) are assembled in the mature protein in a nonrandom set of seven forms, $\alpha'\beta_2$, $\alpha\beta_2$, $\alpha\alpha'\beta$, $\alpha_2\beta$, $\alpha_2\alpha'$, α_3, and β_3 with molecular weights of 125–171 kDa. The α and α′ subunits have 1–2 mol of cysteine per mole of peptide, while the β peptide has no cysteine.

Glycinin is a hexamer, although in older literature, it is referred to as a dodecamer. It is composed of six nonrandomly paired acidic and basic peptides. The acidic peptides have molecular weights of 44, 37, and 10 kDa; the basic peptides are 20 kDa. The acidic–basic (AB) pairs have been identified in the experimental line, CX635-1-1-1, and are shown in Table 1. Seven acidic and eight basic peptides have been identified in 18 cultivars. There appear to be sulfur-rich and sulfur-poor AB pairs.

Soya beans contain two main classes of protease inhibitors or trypsin inhibitors (TI), although there appear to be many isogenic variants. The two principal classes are the Kunitz inhibitor with a molecular weight of 21 500 and the Bowman–Birk inhibitor, with a molecular weight of 8000. The Kunitz inhibitor acts only on trypsin, while the Bowman–Birk protein inhibits both trypsin and chymotrypsin. Moist heat treatment denatures about 90% of the TI activity with the residual being heat-stable. TI affects animals of guinea-pig size and smaller but has little effect on larger animals, except for weanling pigs. Recently, the Bowman–Birk TI has been demonstrated to be an anticarcinogen and is currently in Phase II cancer trials. There are reports that TIs exert a carcinogenic effect on rodents. There has been interest in recovering and purifying soya bean protease inhibitors to treat AIDS patients.

The United States Food and Drug Administration has approved a health claim for soya protein that, as part of a heart-healthy diet, 25 g per day of soya protein will contribute to improved cardiovascular health. (*See* **Trypsin Inhibitors**; **Protein**: Chemistry.)

Lipid

Soya bean lipids contain about 2–5% phospholipids, depending on the growing conditions, and 1.6% unsaponifiables. The balance is chiefly triacylglycerols.

Oleic (18:1), linoleic (18:2), palmitic (16:0), stearic (18:0), and linolenic (18:3) acids are present in soya-bean oil, along with traces (less than 1%) of myristic (14:0), palmitoleic (16:1), heptadecanoic (17:0), eicosenoic (20:1), arachidic (20:0), behenic (22:0), and erucic (22:1) acids. The range of acyl groups present in soya bean oil has been extended by mutation breeding and selection to the values reported in Table 2.

The acyl groups are distributed asymmetrically in the triacylglycerols with all the saturates on the *sn*-1 and -3 positions and linoleate concentrated on the *sn*-2 position. Generally, the *sn*-1 contains more palmitate and stearate than *sn*-3, and the oleate is enriched on *sn*-3. Phospholipids contain the same acyl groups found in the triacylglycerols, but the

Table 1 Glycinin acidic–basic complexes from CX635-1-1-1

Acidic–basic complex	*S-amino acids*	*Molecular weight*
A_{1a} B_2	14	57
A_{1b} B_{1b}	12	57
A_2 B_{1a}	14	57
A_3 B_4	9	62
A_5 A_4 B_3	3	67

From Nielsen NC (1985) Structure of soy proteins. In: Altschul AM and Wilcke HL (eds) *New Protein Foods Vol. 5. Seed Storage Proteins*, pp. 27–64. New York: Academic Press.

Table 2 Range of acyl group percentages produced in soya bean lipids by mutation breeding and genetic engineering, and the composition of a typical unselected variety

	Range	*Typical value*
Palmitate	3.5–30	10
Stearate	2.5–32	4
Oleate	8.0–85	26
Linoleate	2.0–60	52
Linolenate	1.8–13	8

concentration of palmitate is generally higher and that of oleate lower. Phosphatidyl choline (50%), ethanolamine (26%), and inositol (18%) are the chief phospholipid components along with lower concentrations of phosphatidic acid and phosphatidyl serine.

The unsaponifiables contain sterols, hydrocarbons, and tocopherols. The chief sterols (3.5 mg pergram of oil) are β-sitosterol, campesterol, and stigmasterol. The tocopherols (about 1.25 mg per gram of oil) are typically more than 70% gamma with smaller amounts of delta and alpha. (*See* **Fats**: Classification; **Fatty Acids**: Properties.)

Methods to Measure Composition

Proximate analyses for moisture, protein, crude free fat, crude fiber, ash, and total carbohydrate have been adopted by the American Oil Chemists' Society, the American Association of Cereal Chemists, and the Association of Official Analytical Chemists. All composition values for soya beans are reported either on a moisture-free basis or at 13% moisture.

Several methods are acceptable for moisture determination, but the most widely used procedure involves measuring the weight loss when drying the ground sample for 3 h at 130 °C. In South America a Karl Fischer titration is frequently used.

Protein is estimated from Kjeldahl nitrogen. The nitrogen content is multiplied by a factor of 6.25 to convert to protein values despite the major soya protein, glycinin containing only 17.5% nitrogen (equivalent to a N conversion factor of 5.7).

Oil content is determined as crude fat by continuously extracting dried ground samples with petroleum ether for 5 h. Total fat, which includes bound fat as well as free fat, requires acid hydrolysis of the sample prior to extraction. In recent years, the use of near-infrared reflectance and near-infrared transmittance has become widespread for rapid estimation of grain composition, especially moisture, protein, and oil. These spectrophotometers must be calibrated against the standard wet chemical methods described above. Moisture is also routinely measured by electrical capacitance.

Crude fiber is measured as the weight loss on incineration of the oven-dried residue remaining after digestion of the sample with boiling dilute sulfuric acid followed by boiling dilute sodium hydroxide.

Ash is primarily composed of noncombustible minerals and is determined by heating the ground sample in a muffle furnace for 2 h at 600 °C. Soya beans contain about 4.7% ash at 13% moisture. Total carbohydrate is often estimated as the difference after subtracting other constituents. This method does not discriminate between oligosaccharides and other sugars, and it often gives inflated values. Sugars can be extracted with hot aqueous ethanol and quantified by gas or liquid chromatography.

Acyl group composition of soybean lipids generally is determined by gas chromatography. Lipid classes are separated by thin-layer or liquid chromatograms.

Qualitative and quantitative analyses of the individual proteins have been performed by immunoelectrophoresis and/or sodium dodecyl sulfate polyacrylamide electrophoresis.

Isoflavones are assayed by reverse-phase high-performance liquid chromatography (RP-HPLC) of 80% acetonitrile extracts and absorbance detection at 250–260 nm. Methanol and ethanol tend to underextract some isoflavone forms. Saponin analysis is typically by RP-HPLC with absorbance detection at 290 and 205 nm, depending on the form, although older methods used silica-based thin-layer chromatography. The major impediment to accuracy has been the lack of suitable standards, although most of the isoflavone forms are now available. (*See* **Analysis of Food**; **Chromatography**: High-Performance Liquid Chromatography; Gas Chromatography.)

Grading Standards

Each exporting country has unique standards, but the grade standards established by the Federal Grain Inspection Service (FGIS) for the USA listed in Table 3

Table 3 Official grades and grade requirements of the Federal Grain Inspection Service, United States Department of Agriculture

Grade	*Minimum test weight kg/L (lbs/bushel)*	*Maximum limits*				
		Damaged kernels		*Foreign material (%)*	*Splits (%)*	*Soyabeans of other colors (%)*
		Heat-damaged (%)	*Total (%)*			
US No. 1	0·721 (56·0)	0.2	2.0	1.0	10.0	1.0
US No. 2	0·695 (54·0)	0.5	3.0	2.0	20.0	2.0
US No. 3	0·669 (52·0)	1.0	5.0	3.0	30.0	5.0
US No. 4	0·631 (49·0)	3.0	8.0	5.0	40.0	10.0
US sample grade						

are typical. Soya beans are divided into two classes based on color: yellow soya beans and mixed soya beans. Each class is divided into four numerical grades and a US sample grade. Special grades (e.g., garlicky, infested) are provided to emphasize special qualities affecting the value, and are added to, and made part of, the grade designation. Six factors are considered in assessing a grade designation: test weight, heat damage, total damage, foreign material, splits, and soya beans of other colors. Although protein and oil contents are not part of the official grading standards, they may be specified in some markets. Near-infrared transmission is used for rapid estimation of moisture, protein, and oil contents.

Test weight is determined on a 1·36 L sample before removing foreign material using an Official Test Weight Apparatus. Foreign material is determined by sieving. Splits are determined by sieving a portion of the grain after removing foreign material. Damaged kernels are determined by hand-picking after removal of foreign material.

See also: **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography; **Dietary Fiber**: Properties and Sources; **Fats**: Occurrence; **Protein**: Food Sources; **Soy (Soya) Bean Oil**; **Soy (Soya) Beans**: The Crop; Processing for the Food Industry; Dietary Importance

Further Reading

Brumm TJ and Hurburgh CR (1990) Estimating the processed value of soybeans. *Journal of the American Oil Chemists' Society* 67: 302–307.

Erickson DR (ed) (1995) *Practical Handbook of Soybean Processing and Utilization*. Champaign, IL: American Oil Chemists Society.

Erickson DR, Pryde EH, Brekke OL, Mounts T and Falb RA (eds) (1980) *Handbook of Soy Oil Processing and Utilization*. Champaign, IL: American Oil Chemists' Society.

FGIS (1990) *Federal Grain Inspection Handbook, Book II*. Washington DC: Federal Grain Inspection Service.

Hammond EG (2000) Genetic alteration of food fats and oils. In: Chow CK (ed.) *Fatty Acids in Foods and their Health Implications*, 2nd edn. New York: Marcel Dekker.

Hurburgh CR, Brumm TJ, Quin JM and Hartwig RA (1990) Protein and oil patterns in the U.S. and world soybean markets. *Journal of the American Oil Chemists' Society* 67: 966–973.

Murphy PA, Song TT, Buseman G *et al.* (1999) Isoflavones in retail and institutional soy foods. *Journal of Agricultural and Food Chemistry* 47: 2697–2704.

Nielsen NC (1985) Structure of soy proteins. In: Altschul AM and Wilcke HL (eds) *New Protein Foods* Vol. 5. *Seed Storage Proteins*, pp. 27–64. New York: Academic Press.

Smith AK and Circle SJ (1978) Chemical composition of the seed. In: Smith AK and Circle SJ (eds) *Soybeans: Chemistry and Technology, Vol. 1, Proteins*, pp. 61–92. Westport, CT: AVI.

USDA (1999) U.S. Department of Agriculture, Economic Research Service. In: Golbitz P (ed.) *Soya & Oilseed Bluebook*. Bar Harbor, ME: Soytech.

Wang HJ and Murphy PA (1996) Mass balance of isoflavones study in soybean processing. *Journal of Agricultural and Food Chemistry* 44: 2377–2383.

Wu SW, Murphy PA, Johnson LA, Fratzke AR and Reuber MA (1999) Pilot plant fractionation of soybean glycinin and β-conglycinin. *Journal of the American Oil Chemists' Society* 76: 285–293.

Dietary Importance

A M Fehily, Tinuviel Software, Warrington, UK

Background

The soya bean (*Glycine max*) is a native of China, where it has been part of the diet for several thousand years. It has also been important in Japan and Korea for almost as long. These beans are used to produce foods that are important components of the diet in the Far East, e.g., fermented soya products such as tofu, miso, tempeh, and soy sauce. More recently, soya beans have become widely used in the Western diet, being used to produce a range of ingredients: soya flour and soya protein, which are used to make foodstuffs such as bread, other cereal products, meat products, and vegetarian alternatives; soya oil, which is widely used as a vegetable oil both in domestic cooking and by the food industry and is also used in the manufacture of spreadable fats; soya lecithin, which is used as an emulsifier in many foods; and soya milk, a vegan alternative to cow's milk. Soya is also used to produce infant formulas for those infants who are intolerant to lactose or cows' milk protein.

Soya can provide a valuable contribution to nutrient intakes and may also have a number of potential health benefits. This article discusses the composition of soya beans, the role of soya in the diet, and its potential health benefits.

Composition

Soya beans contain a wide range of nutrients. They contain some 'antinutrients' and also contain other components, e.g., isoflavones, which may have beneficial effects on health.

Nutrients

Soya beans contain protein, fat, carbohydrate, vitamins, minerals, and fiber.

Protein Soya beans are a good source of protein, containing 14 g per 100 g. This is more than that in most other vegetables and more than that in most other legumes (typically 5–9 g per 100 g), the exception being peanuts, which contain 26 g of protein per 100 g. In terms of protein quality, soya contains substantial amounts of most essential amino acids. When compared with a reference protein (egg), soya protein, in common with other legumes, is deficient in sulfur-containing amino acids (the essential amino acid methionine and the nonessential amino acid cystine) and has a higher content of lysine. Combining legumes with cereals provides a meal with a high protein quality, as cereals have adequate amounts of sulfur-containing amino acids but are deficient in lysine. (*See* **Protein**: Food Sources.)

Carbohydrate Soya beans contain 5 g of carbohydrate per 100 g. This is lower than that of many other legumes. Of the total carbohydrate, 37% is starch, 41% is sugars, and 22% is oligosaccharides. The sugars present are sucrose, fructose, and glucose. (*See* **Carbohydrates**: Classification and Properties.)

Fat Soya beans contain 7 g of fat per 100 g. This is more than that in other legumes, most of which are low in fat. The exception to this is peanuts, which contain 46 g of fat per 100 g. The fatty acid composition of soya has a high ratio of unsaturates to saturates. Of the fatty acids, 16% are saturates, 24% are monounsaturates, and 60% are polyunsaturates. The polyunsaturated fatty acids present are the essential fatty acids linoleic acid (18:2*n*-6) and α-linolenic acid (18:3*n*-3). (*See* **Fats**: Classification.)

Minerals Soya beans contain significant amounts of several minerals: calcium, iron, magnesium, potassium, phosphorus, and zinc (see Table 1). One hundred grams of soya beans provide about 12% of the daily reference nutrient intake (UK) for calcium, 20% of that for iron, 23% of that for magnesium, 15% of that for potassium, 45% of that for phosphorus, and 13% of that for zinc (see Figure 1).

Vitamins Soya beans contain a number of water-soluble vitamins: thiamin, riboflavin, niacin, vitamin B_6 and folate (see Table 1). One hundred grams of soya beans provide about 15% of the daily reference nutrient intake (UK) for thiamin, 8% of that for riboflavin, 21% of that for niacin, 19% of that for vitamin B_6 and 27% of that for folate (see Figure 1). Soya beans also contain vitamin E: 1 mg per 100 g. (*See* **Vitamins**: Overview.)

Fiber Soya beans provide a valuable source of fiber in the diet: 6 g of nonstarch polysaccharides (NSP) per 100 g. Of this, 44% is soluble fiber, and 56% is insoluble fiber. One hundred grams of soya beans provide

Table 1 Vitamin and mineral composition of soya beans

	Amount per 100 g as consumed
Sodium	1 mg
Potassium	510 mg
Calcium	83 mg
Magnesium	63 mg
Phosphorus	250 mg
Iron	3.0 mg
Copper	0.32 mg
Zinc	0.9 mg
Manganese	0.7 mg
Selenium	5 μg
Iodine	2 μg
Vitamin A (retinol equivalents)	1 μg
Vitamin D	0 μg
Vitamin E	1.13 mg
Thiamin	0.12 mg
Riboflavin	0.09 mg
Niacin equivalents	2.7 mg
Vitamin B_6	0.23 μg
Vitamin B_{12}	0 μg
Folate	54 μg
Pantothenate	0.18 mg
Biotin	25.0 μg
Vitamin C	Trace

Source: *McCance and Widdowson's The Composition of Foods* (1991) 5th edn. Royal Society of Chemistry and Ministry of Agriculture Fisheries and Food. Crown copyright is reproduced with the permission of the Controller of Her Majesty's Stationery Office. HMSO: London.

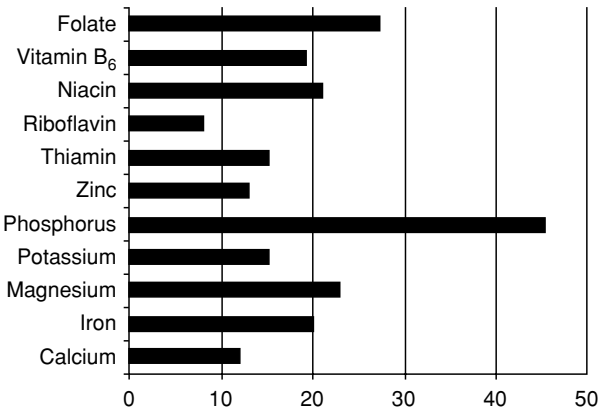

Figure 1 Vitamin and mineral composition of soya beans in relation to reference nutrient intakes (percentage of reference nutrient intake provided by 100 g).

33% of the daily dietary reference value (UK) for fiber. (*See* **Dietary Fiber**: Properties and Sources.)

Antinutrients

Raw soya beans contain a number of substances that may have detrimental effects on digestion and other metabolic processes (lectins, goitrogens, and digestive enzyme inhibitors). These substances are, however, inactivated by appropriate cooking of the beans or, in the case of goitrogens, counteracted by adequate iodine intake. (*See* **Plant Antinutritional Factors**: Characteristics.)

Lectins Lectins, also known as hemagglutinins, are present in raw soya beans and in other legumes. They are, however, heat-labile and are inactivated when the beans are properly cooked. In experimental animals, these polymeric proteins have been shown to cause damage to red blood cells and intestinal mucosa, and thereby impaired nutrient utilization and loss of body weight. In humans, these compounds can result in nausea, vomiting, diarrhea, and abdominal pain.

Goitrogens Enlargement of the thyroid gland has been shown to occur in rats fed soya bean meal. The goitrogenic agent in soya beans is unknown. However, the effect is counteracted by adequate iodine intake.

Digestive enzyme inhibitors Raw soya beans contain certain proteins (protease inhibitors) that react with digestive enzymes (trypsin, chymotrypsin, or salivary and pancreatic α-amylase), thereby interfering with the digestion of protein and starch. In humans, raw soya or isolated protease inhibitors increases levels of cholecystokinin (CCK) and pancreatic secretion. It is thought that chronic pancreatic stimulation may lead to pancreatic hypertrophy, hyperplasia, and possibly cancer. In rats, raw soya has been shown to increase the risk of pancreatic cancer. However, protease inhibitors present in raw soya are inactivated by heat and are therefore not a problem in cooked beans. For example, in countries such as Japan, where soya foods are widely consumed, the incidence of pancreatic cancer is similar to, or less than, that for Western countries, where soya forms a relatively small component of the total diet.

Phytic acid Phytic acid is present in soya beans and also in other fiber-containing foods. It has been suggested to reduce the absorption of calcium, iron, zinc, and vitamin D from the diet. However, these effects have been observed in *in vitro* experiments and not in *in vivo* studies, even at fiber intakes at the upper limit of the normal human consumption range. In addition, levels of micronutrients tend to be higher in fiber-rich foods than in fiber-poor foods. Thus, adverse effects of phytic acid on nutrient absorption are likely to be unimportant at the levels of fiber normally consumed in the human diet. (*See* **Phytic Acid**: Nutritional Impact.)

Isoflavones

Soya beans contain isoflavones, a group of compounds that are structurally similar to estradiol (phytoestrogens). Soya beans contain high amounts of the isoflavones genistein and daidzein (see **Figure 2**) and smaller amounts of glycitein, primarily in the form of glycosides: genistin, daidzin, and glycitin (total isoflavone content of 180 mg per 100 g). The glycosides are hydrolyzed in the gut by a bacterial enzyme, glucose being removed to produce the aglycone forms (genistein, daidzein, and glycitein). Daidzein and genistein have been shown to have a weak estrogenic activity and are able to bind with a low affinity to estrogen receptors, the affinity being greater for β-receptors than for α-receptors. Compared with 17β-estradiol, daidzein and genistein have been reported to have much less potency in producing estrogenic effects, by about 1000-fold or more.

In the 1940s, it was reported that phytoestrogens may exert adverse effects on uterine and ovarian function. This was because sheep grazing on pastures containing a particular type of clover (*Trifolium* sp.) rich in formononetin, which is converted to daidzein in the rumen, developed a widespread infertility. An infertility syndrome has also been described in captive cheetah, as a result of soya bean use, the syndrome

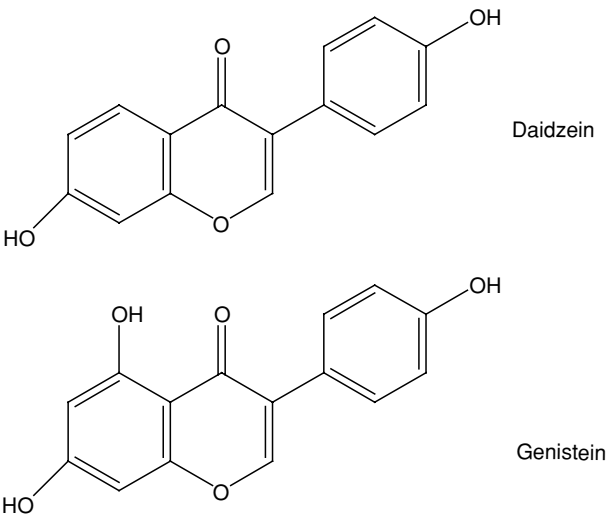

Figure 2 Chemical structure of the main isoflavones present in soya (daidzein and genistein).

being reversed by removal of soya from the food. No adverse effects of soya beans on human fertility have been reported.

More recently, it has been recognized that isoflavones may act either as weak estrogens or as antiestrogens, competing for estradiol at the receptor complex, yet failing to stimulate a full estrogenic response after binding to the nucleus. This fact has stimulated much research into whether isoflavones may have a protective role in hormone-related diseases such as breast cancer.

Role of Soya in the Diet

Soya has been a staple in the diet of many Far Eastern countries for centuries, e.g., foods such as tofu, tempeh, and miso. In Western countries such as the USA and UK, the amount of soya consumed is much less than in countries such as Japan. In Western countries, these traditional soya foods are not very widely consumed but are sometimes used as an alternative to meat. Soya beans are, however, extremely versatile, providing a wide range of ingredients that can be used in foods for most meal occasions, e.g., soya flour, soya protein, soya oil, soy sauce, and soya lecithin. For this reason, consumption of foods containing soya has increased in Western countries over the last 50 years. Nevertheless, soya still forms only a small part of the total diet. It is noteworthy that of the soya beans grown world-wide, only about 10% of the total is used in foods for humans, the majority being used as feed for animals.

Soya beans can be an important component of a normal healthy diet, as they are low in saturates, provide a source of fiber, and contain a wide range of other nutrients. Soya beans are also the main source of phytoestrogens in the diet. In addition, soya beans may have a number of potential health benefits, and these are discussed in the following sections.

For vegetarians, and particularly for vegans who eat no animal products, soya beans play a very important role in the diet. They provide a valuable source of protein that can be of high quality if soya beans are combined with cereals, since the amino acid compositions are complementary. They also provide an important source of iron. Low iron intake can lead to iron-deficiency anemia. Ensuring adequate iron status can be difficult for vegetarians as, in general, iron from vegetable sources is less well absorbed than that from animal sources. However, soya beans provide a source of iron that can be absorbed as well as that in meat.

For infants intolerant to cows' milk protein or to lactose, soya-based infant formulae can be an invaluable alternative to cows' milk-based formulae. For older children and adults who are allergic to cows' milk or intolerant to lactose, drinks and other foods made from soya can be very important components of the diet, helping to ensure that nutrient intakes are adequate for the maintenance of health.

In common with other legumes, soya beans are also of agricultural importance, as they increase the nitrogen content of the soil. Although green plants cannot utilize nitrogen in the atmosphere, there are several species of bacteria, fungi, and blue–green algae that are able to transform nitrogen in the air into a form that can be used by plants. An important genus of nitrogen-fixing bacteria is *Rhizobium*, which forms nodules in the roots of legumes. These bacteria live symbiotically with the legumes, the bacteria obtaining food from the green plant, and the legumes obtaining abundant usable nitrogen compounds from the bacteria.

Soya and Coronary Heart Disease

Coronary heart disease (CHD) is a major cause of death in many countries. For example, in the UK, CHD accounts for about 30% of male deaths and 23% of female deaths.

The level of cholesterol in the blood is a major risk factor for CHD. Soya protein substituted for animal protein in the diet results in reductions in total plasma cholesterol (by 9% on average), low-density lipoprotein cholesterol (LDL-C) (by 13% on average) and triacylglycerols (by 11% on average). High-density lipoprotein cholesterol (HDL-C) is unchanged or may be slightly increased. The amount of reduction in blood lipids tends to be greater among those with the highest plasma cholesterol levels at baseline.

A great deal of research has been done to identify the component(s) of soya protein responsible for its effect on blood lipids. Studies using a mixture of amino acids that duplicates the amino acid profile of soya protein have found that this does not have the same effect on blood lipids as the intact protein. However, studies of soya protein with isoflavones intact have shown that this has a lipid-lowering effect, whereas soya protein with isoflavones removed has no significant effect. It has been suggested, therefore, that the isoflavones present in soya protein are largely responsible for the lipid-lowering effect of soya protein. The benefits of isoflavones present in soya protein on cholesterol lowering have been suggested to be mediated through upregulation of LDL-receptor activity. Nevertheless, studies of isolated and purified isoflavones have failed to show that these have a lipid-lowering effect.

Soya may reduce the risk of CHD through several mechanisms in addition to its ability to lower blood

lipids. For example, soya isoflavones are known to act as antioxidants and have been suggested to reduce oxidative damage to LDL-C. A decrease in oxidized LDL particles, which are considered atherogenic, may reduce the risk of atherosclerosis.

A study of young cynomolgus monkeys showed that the size of atherosclerotic lesions was 70% less in those fed on a diet containing soya protein with isoflavones present compared with those fed on a diet containing soya protein with isoflavones removed. This suggests that isoflavones may inhibit atherogenesis. In another study of surgically postmenopausal cynomolgus monkeys, soya protein with isoflavones intact was found to reduce the progression of atherosclerosis. The magnitude of the effect was reported to be comparable with that of postmenopausal estrogen therapy.

Soya bean isoflavones, in particular genistein, impact beneficially on vascular reactivity. Genistein has also been shown, *in vitro*, to have a number of other effects whereby the development of atherogenesis may be delayed: inhibition of the migration and proliferation of smooth muscle cells; inhibition of platelet activation and aggregation; and reduction in platelet serotonin uptake.

In 1999, in response to the available evidence on the benefits of soya protein, the USA Food and Drug Administration (FDA) announced that it would allow food manufacturers to label products containing 6.25 g of soya protein per serving as helping to reduce risk of heart disease, as part of a balanced diet low in fat and saturates.

Soya and Cancer

Cancer is a major cause of morbidity and mortality. For example, in the UK, about one in three people will develop cancer at some time during their life, and cancer accounts for about one in four of all deaths. There are marked differences in death rates from several types of cancer between Asian countries and many Western countries. The differences are more striking for hormone-dependent cancers of the prostate, breast, and colon/rectum. For example, the risk of dying from prostate cancer among Japanese men is only one-fifth that of men in the USA, and Japanese women have a breast cancer mortality rate that is only one-quarter of that of women in the USA.

Countries where mortality from prostate and breast cancers is low have considerably higher intakes of soya than those countries where mortality from these cancers is high. In animal models, studies investigating the effects of soya beans on prostate cancer or breast cancer have shown reduced tumorigenesis. In humans, however, evidence is inconclusive. Epidemiological studies of prostate cancer conducted in Japan and in Japanese migrants to Hawaii have shown no significant effect of soya consumption. However, a recent study in the USA reported that men who drank one cup of soya milk per day had a risk of developing prostate cancer that was 70% lower than that of controls. For breast cancer, although some studies have shown that soya can help prevent breast cancer, others have found no significant effect. An increase in menstrual cycle length has been observed in some studies, in response to soya bean consumption, although no change was observed in others. Such an effect is of potential relevance to the hypothesis that soya may be protective against breast cancer, because some data indicate that longer cycles, which are typical of Asian women, are associated with a reduced risk of breast cancer.

For colorectal cancer, data are also inconclusive: some studies showed no significant effect, some reported a protective effect of soya, and others observed an increased risk with soya consumption. For cancers at other sites, data tend to show that increased soya bean consumption may be protective against lung and stomach cancer. Much more research is needed to confirm whether soya has a protective effect against cancer in humans. (*See* **Cancer**: Diet in Cancer Prevention.)

Isoflavones appear to be the components of soya that would be most likely to account for any protective action of soya against cancer, especially in inhibiting the initiation stage of carcinogenesis. Genistein has been shown to suppress the growth of a wide range of cancer cells. It inhibits certain enzymes that could affect the onset of cancer or the growth of tumors: tyrosine protein kinase, mitogen-activated protein (MAP) kinase and ribosomal S6 kinase. Genistein also inhibits the DNA repair enzyme topoisomerase II and acts as an antioxidant, thus potentially preventing oxidative DNA damage. It increases *in vitro* concentrations of transforming growth factor β, which is thought to inhibit the growth of cancer cells. In addition, genistein has also been shown to inhibit angiogenesis, the formation of new blood vessels, an abnormal event that occurs as part of the growth and expansion of malignant tumors. However, many of these effects have been shown with very high concentrations of genistein and not in cells treated with levels likely to be achieved in plasma of human subjects eating soya foods.

Soya and Bone Health

Osteoporosis is a clinical condition in which there is a reduced amount of bone per unit volume and an

increased susceptibility to fractures, particularly fractures of the vertebrae, distal forearm (Colles fracture), and hip. Of these, hip fracture is the most severe, since patients require a lengthy hospital stay, there is a high mortality rate (about 20% within 6 months of the fracture), and of those who survive, many suffer permanent disability and dependency. Osteoporosis is a major health problem in many Western countries, but not in Asian countries. A number of factors are known to increase the risk of osteoporosis, including insufficient dietary calcium, low physical activity, and lack of estrogen.

In women, bone mass reduces at a rapid rate in the first few years after the menopause, owing to the reduction in estrogen. Hormone-replacement therapy (HRT) is well known to reduce bone loss in postmenopausal women. However, HRT is not appropriate for all women. Since soya may have estrogenic effects, it has been hypothesized that it may provide an alternative to HRT.

Soya protein containing isoflavones has been reported to reduce bone loss due to estrogen deficiency in ovariectomized rats. Soya bean isoflavones have also been shown to have a protective effect on bone loss in ovariectomized rats, suggesting that the beneficial effect of soya protein is due to isoflavones. The effect of isoflavones on bone loss in rats has been reported to be similar to that for estrogen. The bone-sparing effect of soya protein isoflavones has been suggested to be due to a reduction in bone resorption and/or to an increased osteoblast activity (increasing bone formation).

In humans, most but not all studies indicate that soya protein containing isoflavones favorably affects bone turnover and bone mineral density in the lumbar spine of perimenopausal and postmenopausal women. Treatment with ipriflavone, a synthetic isoflavone, has also been reported to have bone-conserving effects in postmenopausal women with a low bone mass, thus confirming that the beneficial effect of soya protein is likely to be due to isoflavones.

In addition to the effect of soya protein on bone mineral density, it has also been suggested that soya protein, when substituted for animal protein, may indirectly enhance bone strength. Another effect of soya protein is that it helps to conserve calcium by reducing urinary calcium excretion. This is due to the lower sulfur amino acid content of soya protein.

Soya protein therefore appears to have modest beneficial effects on bone density. However, studies to date have been short term and have involved only small numbers of subjects. In addition, no study has investigated whether soya protein containing isoflavones has an effect on fracture risk.

Soya and Menopausal Symptoms

Hot flushes (also known as hot flashes) are a common symptom among menopausal women in Western societies but are reported to be much less common in Japan. Whether this difference is due to the higher consumption of soya in Japan is not clear. HRT generally alleviates hot flushes and other menopausal symptoms such as vaginitis. There is much research interest in the possibility that soya may provide an alternative to HRT in this regard.

Evidence of benefit of soya in menopausal women is conflicting. Some researchers have reported an improvement in the frequency of hot flushes in women taking soya protein daily. Others have reported no effect of soya protein on the number of hot flushes experienced but found a reduction in the severity of symptoms. Others have reported a reduction in both the incidence and severity of hot flushes with soya protein or an isoflavone extract compared with a control group. Others have reported no difference in either the frequency or severity of hot flushes in the intervention group compared with the control group.

Two studies have investigated the effects of phytoestrogen supplements on vaginal cytology and found an increase in cell proliferation (an indication of estrogenic activity) and reversal of menopausal atrophy. Others have reported no significant effect.

There are difficulties in interpreting the results of these studies. This is because there are differences in the amounts of soya protein and isoflavones used and differences in the duration of the studies. Interpretation is also complicated by the fact that the reported frequency and severity of symptoms tend to decrease in the control group as well as the intervention group. Thus, much more research needs to be done before soya protein can be proposed as a potential alternative to HRT for the control of menopausal symptoms.

See also: **Amino Acids**: Properties and Occurrence; **Anemia (Anaemia)**: Iron-deficiency Anemia; **Atherosclerosis**; **Cancer**: Epidemiology; Diet in Cancer Prevention; **Cholesterol**: Factors Determining Blood Cholesterol Levels; Role of Cholesterol in Heart Disease; **Colon**: Cancer of the Colon; **Coronary Heart Disease**: Etiology and Risk Factor; Prevention; **Dietary Fiber**: Physiological Effects; Bran; **Osteoporosis**; **Plant Antinutritional Factors**: Characteristics; **Vegetarian Diets**

Further Reading

Anderson JJB, Anthony M, Messina M and Garner SC (1999) Effects of phyto-oestrogens on tissues. *Nutrition Research Reviews* 12: 75–116.

Anderson JW, Johnstone BM and Cook-Newell ME (1995) Meta-analysis of the effects of soy protein intake on

serum lipids. *New England Journal of Medicine* 333: 276–282.
Bingham SA, Atkinson C, Liggins J, Bluck L and Coward A (1998) Phyto-oestrogens: where are we now? *British Journal of Nutrition* 79: 393–406.
Lissin LW and Cooke JP (2000) Phytoestrogens and cardiovascular health. *Journal of the American College of Cardiology* 35: 1403–1410.
Messina MJ (1999) Legumes and soybeans: overview of their nutritional profiles and health effects. *American Journal of Clinical Nutrition* 70(Supplement): 439S–450S.
Messina M and Messina V (2000) Soyfoods, soybean isoflavones and bone health: a brief overview. *Journal of Renal Nutrition* 10: 63–68.
Messina M, Persky V, Setchell KDR and Barnes S (1994) Soy intake and cancer risk: a review of the *in vitro* and *in vivo* data. *Nutrition and Cancer* 21: 113–131.
Potter SM (2000) Soy – new health benefits associated with an ancient food. *Nutrition Today* 35: 53–60.

SOY (SOYA) CHEESES

T Ono, Iwate University, Morioka, Iwate, Japan

Introduction

Soy cheeses (soya cheeses) are a curd made from soymilk (soybean milk) and its processed foods. The soybean curd is a popular food used in daily meals in eastern Asia. It is usually called tofu in English. Tofu is made in a similar process to making cheese from milk. The soymilk is curdled by adding coagulants ($CaSO_4$, $MgCl_2$, glucono-delta-lacton (GDL)). Curds are then pressed into cakes of various types. There are three types of tofu: soft tofu, medium-soft tofu, and firm or extra-firm tofu. Soft tofu is not pressed and is eaten as it is. Medium-soft tofu is the most popular in everyday use. Firm tofu is used in processed foods which are fried, grilled, fermented, and dried. Tofu derivatives are also familiar in eastern Asia.

Tofu has a long history of 2000 years and has held an important position in the eating habits as popular foodstuffs in eastern Asia. The composition and nutrient contents of tofu are similar to those of cheese made from milk. Soybean products, however, contain phytoestrogens, such as isofravones, which may lead to a low incidence of estrogen-related cancers and low prevalence of climacteric syndrome and osteoporosis. The rate of these diseases is lower in Asian peoples who consume a lot of soybean products.

Soy cheeses (tofu and its derivatives) are now spreading throughout the world as healthy foods and a low-cost supply of nutrients.

History of Tofu

It is believed that tofu was invented by Lui An, King of Hui Nang (178–122 BC) of the western Han dynasty (202 BC–8) in China. A wall relief showing tofu making was found in Dahuting tomb no. 1 (Mi, Henan) in the eastern Han era (25–220). It became popular among the public in the Song (960–1271). The technology for making tofu was spread to neighboring countries (Japan, Korea, Vietnam) in the early stage of the Tang dynasty (618–907). The Chinese character dou-fu means soybean curd; dou (to) means soybean and fu means a soft elastic object, although the object is a solid. Doufu was first written in the early stages of the Song dynasty (960–1271). Perhaps the idea for the name derives from nyu-fu (yogurt); nyu means milk. This nyu-fu was written in historical documents of the Tang dynasty (618–907).

In Japan, the oldest literature about tofu is found in the record of Kasuga-shrine in 1183. When tofu was introduced into Japan in the Nara period (761–793), it was called okabe, and the old cookbook referred to it as kabe (wall) or shirakabe (white wall). Tofu became a common food in the eating habits of Japanese people in the Edo era (seventeenth century) and cookbooks on its use were published one after another.

The soy curd is called doufu in China, tahu in Indonesia, tofu in Japan, tobu in Korea, tau fu and tau kuan in Malaysia, tokwa in the Philippines, and tau pho and tau hu in Vietnam. The Chinese character (dou-fu) was spread to eastern Asia through trading with China at 1000 years ago, and called by the pronunciation of each country. The name tofu has now entered into the English language. Tofu is gradually becoming a popular food for health and as a low-cost supply of proteins in the American continent, Europe, and Africa.

Manufacturing Process

Soybeans are soaked in water overnight, then drained. The beans are ground as a small quantity of water is poured over them. The resultant mash is adjusted by adding 10 times the amount of water as there are

soybeans for medium-soft and firm tofu, or seven times the quantity of water for soft tofu. The slurry is allowed to boil gently for about 5 min by steaming or heating and is then filtered. Alternatively, after the slurry is filtered, the filtrate (raw soymilk) is then heated (in the Chinese old style). The heated filtrate is soymilk and the residue is okara. A small amount of either calcium sulfate or magnesium chloride is introduced to coagulate the soymilk. The use of the calcium salt (gypsum) is in the original Chinese style and using magnesium chloride (bittern) from sea salt is the Japanese style. After the milk has curdled (above 80 °C), a part of the whey is squeezed to make medium-soft and firm tofu. The curds are ladled into a rectangular forming container lined with cheesecloth. A lid is placed on the container, and the curds are pressed at 70–80 °C to make a block and to squeeze the whey. The water content of medium-soft tofu and firm tofu is about 85% and 75–80%, respectively. The tofu block is sometimes soaked into a tub of cold water to replace the whey with water. On the other hand, to make soft tofu, the soymilk is curdled in the container and is packed without pressing. Soft tofu (packaged tofu) is also made from a packaged soymilk containing GDL by heating above 90 °C for 1 h. GDL changes to gluconic acid by heating and pH decreases to 5.8. Soymilk coagulates by decreasing pH.

Bean Curd (Tofu)

The types of soft tofu products include silken, doufang (China), oboro (Japan) and packaged. Doufang (meaning bean flower) and oboro (meaning hazy) are eaten like yogurt as they are. Silken and packaged are packaged, have a longer shelf-life, and are used in salads and various cooked foods. Types of medium-soft tofu include momen (meaning cotton cloth) tofu (Figure 1), which is used in various cooked foods. It is the most popular type of tofu in eastern Asia. Firm tofu (pressed/firm curd (Figure 1) and curd sheet) is used as processed foods which are fried, grilled, fermented, dried, and so on. These tofus are available in grocery stores and supermarkets.

Tofu is like a dairy product in that it must be refrigerated and has a short shelf-life (1–7 days). Packaged tofu (silken) is often ultrapasteurized and so has a very long shelf-life (1–2 months). These tofus are always offered as fresh.

Processed Foods (Derivatives of Tofu)

Tofu Cheese (Cheese Alternative)

Firm tofu is scrambled and mixed with casein, soybean oil, modified food starch, salt, citric acid, sodium citrate, natural flavor, vitamins and minerals, and the mixture is heated and molded. It is sold as a cheese alternative containing no cholesterol and no lactose. There are Cheddar, Mozzarella, garlic and herb, and other types.

Fried Bean Curd (Aburage)

Firm tofu is cut into boards or cubes and deep-fried. A board-shape one is then swollen with air, and becomes something like the skin of fried chicken. It

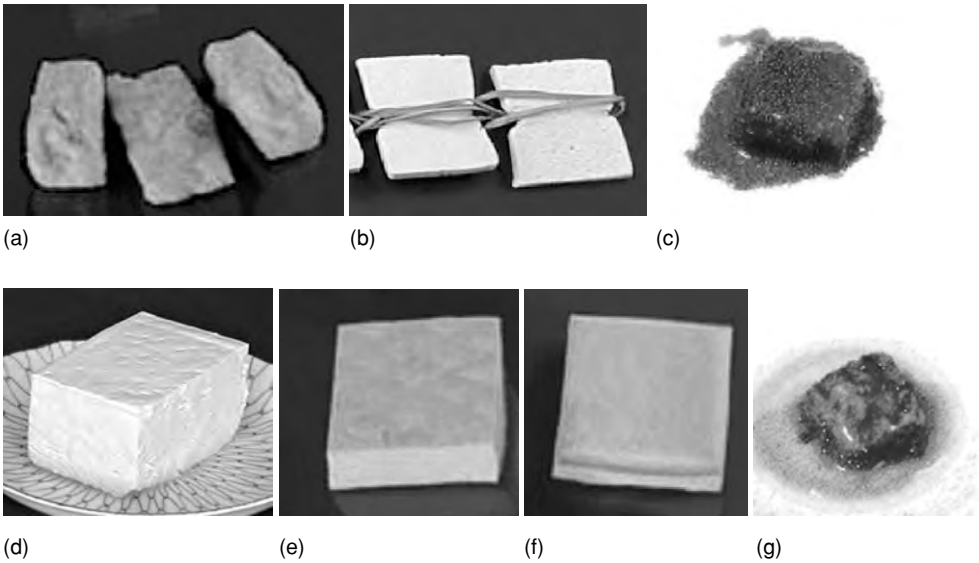

Figure 1 (see color plate 127) Photographs of various tofu and its derivatives. a, fried tofu; b, dried frozen tofu; c, furu; d, momen tofu; e, firm tofu; f, dried tofu; g, choutofu.

is often used for wrapping boiled rice or other foods. Cube and block types (Figure 1) are eaten like fried tofu.

Grilled Bean Curd (Yaki-Tofu)

The top and bottom surfaces of a rectangular firm tofu are roasted till a part is scorched. This process gives it durability against damage in cooking.

Ganmodoki

Firm tofu is scrambled, mixed with vegetables and seasonings, molded into a disk shape and deep-fried. Sometimes beans and seaweeds are mixed. These are eaten intact or after roasting and cooking.

Dried Bean Curd

Firm tofu is pickled and dried. This is an amber hard block (Figure 1) and has a very long shelf-life (6 months). When cooked, it is cut into a chip shape.

Dried Frozen Tofu (Kohri-Tofu)

Firm tofu is cut into rectangular boards, frozen between −3 and −5 °C for 1–7 days, thawed, and dried. The tofu becomes a porous plate (Figure 1) in this process and has a very long shelf-life (1 year). This porous texture is made by tofu proteins denaturing during freezing. When cooked, its texture is different from tofu and it is like a sponge with many fine holes. It is able to contain large quantities of soup in cooking.

Smoked Bean Curd

Firm tofu is dipped into soy sauce, smoked, and aged for 1–2 weeks at a low temperature. Its texture is like smoked cream cheese. It is excellent as a side dish with alcohol.

Tofu-Kamaboko

Firm tofu is scrambled, packed into a semicylindrical shape, and steamed. Sometimes beans and vegetables are mixed in. It is flavored with sugar and salt.

Bean Curd Stick

Curd sheet is cut into stick shapes and deep-fried. This is used for Chinese cooking.

Fermented Tofu (Furu, Choutofu, Tofuyoh)

There is an original fermented tofu in each district, particularly in China. It is similar to fermented cheese, like blue and Brie cheese.

Furu (Figure 1)

Firm tofu is cut into small cubes and these are inoculated over their surfaces with appropriate molds (*Actinomucor elegans*, *Mucor hiemalis*, *M. silvaticus*, *Rhizopus chinensis*, and the like). The molded cubes are salted in brine containing alcohol (red or white), moldy rice (*Monascus purpureus, M. anka*, and the like) and spices (hot pepper, star anise, bay laurel, sesame oil, and rose essence), then packed and aged for 1–6 months. It has a creamy, salty, hot taste and a strong flavor. There are many synonyms for furu: sufu, doufuru, fooyue, fuju, and toefuru. Furu is the standardized name in China. Furu is broadly divided into three categories by color: red, white, and yellow. It is eaten with rice gruel or mantou (steamed bread) at a morning meal, and cooked with meat or vegetables as a seasoning in China.

Choutofu (Figure 1)

The making process is similar to that of furu. However, the moldy rice is not added to the brine and bacterial fermentation occurs in the brine. The brining period is about 1 month. Choutofu is a kind of furu, but is differentiated from other furu in China. It is usually fried before serving and eaten as a relish.

Tofuyoh

Tofuyoh is like furu, and is produced in Okinawa, Japan. Cubes of firm tofu are fermented in brine of moldy rice and awamori (rice liquor). The mold inoculation is neglected when making tofuyoh, and red moldy rice (*Monascucs*) and yellow moldy rice (*Aspergillus*) are added to the brining solution. It is eaten as a relish with liquor.

Bean Curd Skin (Yuba)

Fresh yuba is the layer of skin that is formed by simmering soymilk at near boiling point. After drying, the film becomes yuba. Therefore, it is not a tofu derivative but rather a product of soymilk. It can be eaten as a snack, or used in cooking. Yuba is used in traditional cooking in Japan.

Composition and Nutrients

The composition of medium-soft tofu (momen-tofu), being the most popular, is shown in Table 1, together with that of cheese (Edam) for comparison. The equivalent value of momen-tofu is also shown to compare with cheese composition. The content of protein, lipid, and carbohydrate is similar in both. The calories in 100 g are also similar. As concerns minerals, the sodium content of tofu is 13 times lower than that of cheese, while the potassium content is 10 times higher. These mineral proportions reduce high pressure in blood vessels and may reduce the risk of a cerebral hemorrhage. The iron content of

tofu is 13 times higher than that of cheese. It is effective for healing anemia. Tofu is rich in vitamins D, K, and the B group.

The composition of the main tofu derivatives is shown in Table 2. The lipid content increases in deep-fried products. Each content changes proportionally with water content, except for that of fried products.

The lipid content of tofu is 20% less than that of cheese, and less in saturated fat and cholesterol. It is important to eat less saturated fat and cholesterol to reduce the risk of heart disease. Tofu and its derivatives may play a role in keeping the heart healthy. (*See* **Cholesterol**: Factors Determining Blood Cholesterol Levels.)

Asian people have a low incidence of estrogen-related cancers and a low prevalence of climacteric syndrome and osteoporosis. It is suggested that the cause is due to consumption of a lot of soybean products containing phytoestrogens, such as isofravones. Tofu comprises 40–60 mg isofravones 100 g^{-1}. It is reported that people in Japan and Shanghai consume 20–30 mg of isoflavones per day on average.

Tofu comprises 50% protein and 35% fat in a dry-matter base. The fat in the curd is very stable against ooze and oxidation by cooking and storing. Therefore, the tofu curd must support a stable fixation of lipids. Soymilk for tofu has astringent and soybean flavors and is not suited for drinking unless it is already part of your culture. But these flavors are restrained in tofu and many people can eat tofu. These characters can be explained by the mechanism of curd formation.

Table 1 The compositions of tofu and cheese

Contents	*Tofu (momen)*	*(Equivalent)*	*Cheese (Edam)*
Water (kcal/100g)	86.8	(41)	41.0
Calories (kcal/100g)	72	(322)	356
Protein (g/100g)	6	(29.5)	28.9
Lipid (g/100g)	4.2	(18.8)	25.0
Carbohydrates (g/100g)	1.6	(7.2)	1.4
Ash (g/100g)	0.8	(3.6)	3.7
Calcium	120	(536)	660
Sodium	13	(58)	780
Potassium	140	(626)	65
Phosphorus	110	(492)	470
Iron	0.9	(4.0)	0.3
Thiamin (B_1)	0.07	(0.31)	0.04
Riboflavin (B_2)	0.03	(0.13)	0.42
Niacin	0.1	(0.45)	0.1
Tocopherol (E)	0.6	(2.7)	1.3

Table 2 The compositions of tofu derivatives

Contents	*Ganmodoki*	*Fried bean curd*	*Dried frozen tofu*
Water (g)	63.5	75.9	8.1
Calories (kcal)	228	150	529
Protein	15.3	10.7	49.4
Lipid	17.8	11.3	33.2
Carbohydrates	1.6	0.9	5.7
Ash	1.8	1.2	3.6

Mechanism of Curd Formation

Tofu curd is made from soymilk. Therefore we need information about soymilk before we can understand curd formation. The following is known about soymilk. Soymilk is a turbid solution like cows' milk and contains particulates (40–100 nm) and soluble proteins and oil globule fractions. The particulate protein has a higher binding ability with calcium than soluble protein. The oil globules contain almost all lipids in soymilk, and exist as a globule (200–400 nm) like oil body, as shown in a scanning electron micrograph (Figure 2). The globules disperse in a stable form, packed with oil body protein and soy proteins. The oil globules also have a higher binding ability with calcium than the soluble protein. It is found that in a mixed solution of oil globules and particulate proteins, both fractions aggregate at the same concentration of coagulants, such as $CaCl_2$. Therefore, these results show that a mechanism of curd formation can be illustrated by the behavior of protein particles, oil globules, and soluble proteins when the coagulants are added into soymilk, as shown in Figure 3. At first, protein particles (80 nm) combine with the oil globule (350 nm) by adding magnesium or calcium ions, and then the bound globules aggregate with each other while bringing round water. The curd becomes firm,

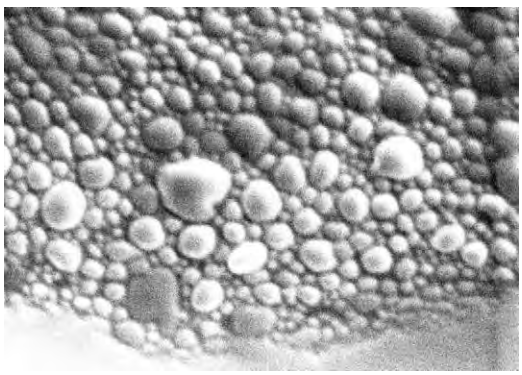

Figure 2 Scanning electron micrograph of oil globules in soymilk.

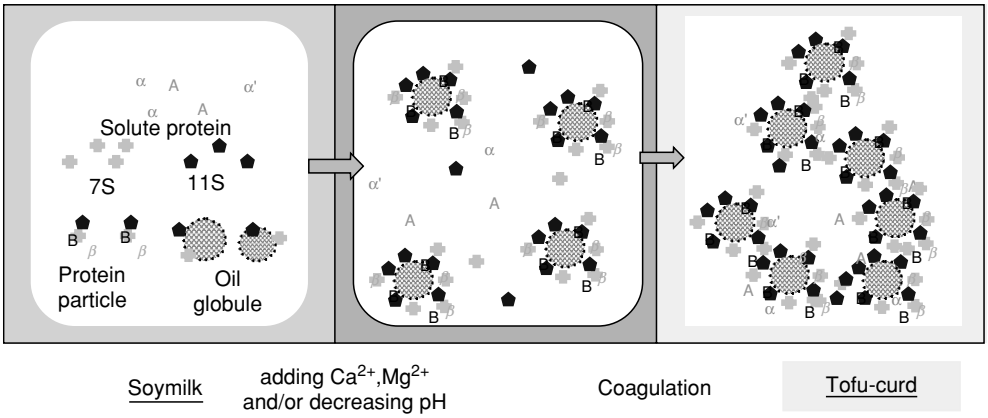

Figure 3 (see color plate 128) Formation of tofu-curd from soymilk (speculation). Oil globule is packed with triple layers of the proteins; oil-body's (oleosin), particulate and solute proteins.

as combining soluble proteins with the pH decrease caused by calcium binding. This assumes that the oil is packed with triple layers of proteins, that is, oil body protein, protein particles, and soluble proteins. Therefore, the oil in the curd is very stable against ooze and oxidation during cooking and storing. The astringent and soybean flavors are also packed in the oil with triple layers of proteins. When we eat tofu and its derivatives, there is no oily and astringent flavor.

The protein particles bind at first on to the oil globule and become a trigger for curd formation. Therefore, the protein particulate content in soymilk is important when making tofu. The particulate content is influenced by the various species and genetic variants of soybean. It is known that the soymilk from glycinin-rich (major soybean protein) soybean contains more particles and the curd shows greater braking stress. The combination of protein particles with oil globules decides the quality of tofu. (*See* **Protein**: Functional Properties.)

See also: **Soy (Soya) Beans**: Processing for the Food Industry; **Soy (Soya) Milk**

Further Reading

Guo W (1993) *The History of Soybean Cultivation in China*. Hehai, People's Republic of China: Hehai University Publishing.

Guo ST, Ono T and Mikami M (1999) Incorporation of soymilk lipid into protein coagulam by addition of calcium chloride. *Journal of Agriculture and Food Chemistry* 47: 901–905.

Honda K (1998) *Tofu and Soybean Cooking: The Japanese Healthy Way*. Tokyo, Japan: Japan Publications.

Li L and Li S (2000) Reviews on Chinese fermented soybean foods. *Proceedings of the Third International Soybean Processing and Utilization Conference*, pp. 701–706. Japan: Korin Publishing.

Messina M, Messina V and Setchell K (1994) *The Simple Soybean and Your Health: How Soy Foods Can Lower Your Cholesterol and Reduce Your Risk*. USA: Avery.

Ono T (2000) The mechanism of curd formation from soybean milk to make a stable lipid food. *Proceedings of the Third International Soybean Processing and Utilization Conference*, pp. 51–52. Tokyo, Japan: Korin Publishing.

Paino J and Messinger L (1990) *The Tofu Book. The New American Cuisine*. USA: Avery.

Resources Council, Science and Technology Agency (2000) *Standard Tables of Food Composition in Japan*, 5th edn. Tokyo, Japan.

Shurtleff W and Aoyagi A (1970) *The Book of Tofu*. USA: Autumn.

Watanabe T (1997) *Science of Tofu – Easy to Understand*. Kyoto, Japan: Food Journal.

Watanabe S, Uesugi S, Zhou XG *et al*. (2000) Effects of soy rich diet and isoflavone supplement on healthy women. *Proceedings of the Third International Soybean Processing and Utilization Conference*, pp. 203–206. Japan: Korin Publishing.

Yamori Y (2000) Soybeans for health in the world – lessons from Okinawan diets and healthy longevity by WHO – cardiac and immigrant studies. *Proceedings of the Third International Soybean Processing and Utilization Conference*, pp. 195–198. Japan: Korin Publishing.

SOY (SOYA) MILK

T Odo, Iwate University, Morioka, Iwate, Japan

Background

Soymilk (soya milk) is a milky liquid made from the soybean. Its composition and nutrient and contents are similar to those of dairy milk. Soymilk has been a traditional breakfast drink for centuries throughout South-east Asia. It is made from soaked soybeans by grinding, heating, and filtering. There are three types of soymilk: regular (traditional), reconstituted (modern), and derivatives. Regular soymilk is prepared from whole soybeans by the aforementioned method. Reconstituted soymilk is made from soybeans by a new processing method and adjusted by adding soybean meal, vegetable oils, and seasonings. Derivatives of soymilk are made from regular or reconstituted soymilk and include drinks, yogurt, soy cheese (tofu), and so on. There are two manufacturing processes for regular (traditional) and reconstituted (modern) soymilk. The traditional process involves a heating process after grinding soaked soybeans. To Chinese people, both in China and overseas, traditional (Chinese-style) soymilk is an everyday regular drink, just like cows' milk is in the West. There are many soymilk producers in Hong Kong, Singapore, Malaysia, Thailand, Taiwan, and China. However, some people are uncomfortable with the natural bean (or raw grass)-like flavor of soymilk. In modern processing, grinding and heating are carried out, reducing the bean-like flavor, and thus increasing the popularity for this type of soymilk for many in Eastern Asia, including Korea and Japan. Soybean products contain phytoestrogens, such as isoflavones, which may reduce the incidence of estrogen-related cancers and prevalence of climacteric syndrome and osteoporosis. The rate of these diseases is lower in Asian people consuming a diet high in soybean products. Soy products are now spreading across the world as healthy foods and as a low-cost source of nutrients.

History of Soymilk

A stone mill for soaked soybeans, found at Huainam in China, is thought to have been made in the Western Han dynasty (202 BC–AD 8). A wall relief showing milling of soaked soybeans was found in Dahuting tomb No. 1 (Mi, Henan) from the Eastern Han era (AD 25–220). These remains indicate that soymilk was originally produced in China over 2000 years ago. Tofu made from soymilk became popular in the Song dynasty (AD 960–1271). The technology for tofu manufacture eventually spread to neighboring countries (Japan, Korea, Vietnam, etc.) in the early years of the Tang dynasty (AD 618–907). Traditional soymilk is called 'Doujiang' in China ('dou-jian' means soybean-serous; 'dou' means soybean, and 'jiang' means a serous liquid). Chinese people went abroad and lived overseas from the song era (AD 960–1271), and in the nineteenth century, many emigrated to South-east Asia in particular (Singapore, Malaysia, Thailand, and so on). Some people are uncomfortable with the natural bean flavor of soymilk, and so soymilk has not become popular in some countries. A new processing method (modern style) for soymilk was developed in the USA in the twentieth century. In this method, grinding and heating are carried out simultaneously. The bean flavor is reduced, thus increasing its popularity. Modern-style soymilk gradually spread to north-eastern countries (northern China, Korea, and Japan). There are many soymilk producers in East Asia, just as there are many dairy producers in Western countries. Nowadays, modern-style soymilk has become a very popular beverage across the world.

Manufacturing Process

There are three types of soymilk: traditional, modern style, and derivatives.

1. Traditional (Chinese-style) soymilk (regular soymilk): Soy beans are washed with water and soaked overnight in cold water. The soaked soybeans are then ground, with a small amount of water being poured over them. The resultant mash is adjusted by adding 10 times more water than soybeans. The slurry is filtered and then boiled gently for 30–60 min. The boil filtrate is traditional (Chinese-style) soymilk (bean-flavored soymilk).
2. Modern-style soymilk (reconstituted soymilk): There are three methods for producing modern-style soymilk: the Cornell, USDA, and Illinois methods, after the organizations involved in the development, i.e., the University of Cornell, United States Department of Agriculture, and University of Illinois, respectively. In these methods, enzymes of the soybean are inactivated by heating before being dissolved in water, because the bean flavor develops by enzymatic reactions. There are several modified

methods for producing modern-style soymilk. These are as follows:

a. Cornell method: Soybeans are soaked in 0.05 N NaOH at 50–65 °C for 2 h and washed with water. The washed beans are ground as hot water (80–100 °C) is poured over them. The resultant, mash is filtered, and the filtrate is digested to soymilk by adding water, emulsifier, sugar, and seasonings, and then homogenized.
b. USDA method: Dry soybeans are dehulled and then mashed at a high temperature with an extruder. The resultant soybean powder is mixed with water, emulsifier, oils, and seasonings, and vigorously homogenized. This method has no filtration process, and whole soybeans (minus their hull) are used for soymilks.
c. Illinois method: Cleaned soybeans are selected by removing dirt, and cracked, damaged, and discolored soybeans. The soybeans are blanched for 5 min in boiling water containing 0.25% baking soda, and blanched again for 5 min in boiling water containing 0.05% baking soda. The blanched soybeans are then ground with hot water (near boiling), and the resulting mash is adjusted to soymilk with water, emulsifier, oils, and seasonings, and then vigorously homogenized. This method has no filtration process, and whole soybeans are used for soymilk. There is also a variation of this method in which the ground mush is filtered and adjusted to soymilk.
d. Other methods: Dried soybeans are dehulled and ground as hot water is poured over them. Alternatively, cleaned soybeans can be soaked and ground with hot water. The resultant mashes are then filtered. The filtrate is adjusted to soymilk by adding water, oils, and seasonings, and then homogenized. Another method uses lipoxygenase-deficient mutant soybeans for soymilk processing. The soymilk from this soybean has no bean flavor, even using the traditional method.

3. Soymilk derivatives
   a. Soymilk soft drinks: Soymilk soft drinks are made from a mixture of reconstituted soymilk and fruit or vegetable juice, coffee, cocoa, or cows' milk. Emulsifier and seasonings are also added and homogenized. The product is then ultrapasteurized and packed in aseptic Tetra brick or retort standing pouches.
   b. Soy yogurt: Soy yogurt is made from reconstituted soymilk by using a starter culture (*L. acidophilus, St. thermophilus*, etc.) and flavors. Soymilk is mixed with sugar, glucono-delta-lacton (GDL), starter, flavors, and seasonings. The mixture is then poured into cups and fermented for 1–2 days. Soy yogurt is sold in chilled lines.
   c. Soy cheese (tofu and its derivatives): Soy cheese (tofu and its derivatives) is made from traditional soymilk. The soymilk is curdled by adding coagulants ($CaSO_4$, $MgCl_2$, GDL). The curds are then pressed into cakes of various types. There are four types of tofu: soft, medium, firm, and extra firm. These are discussed in detail later.

Products

Traditional (Chinese-style) Soymilk (Regular Soymilk)

Regular soymilk has a mellow, bean-like flavor, and is acceptable to the Chinese people. Tofu curd can be made from this soymilk. This soymilk is used for cooking in North-east Asia and is used as a traditional breakfast drink in South-east Asia.

Modern-style Soymilk (Reconstituted Soymilk)

There are two types of reconstituted soymilk. One has only a slight bean-like flavor and is prepared from soaked soybeans by grinding together with hot water. People who often eat soy products prefer the slight bean-like flavor of soymilk. The bean-like flavor is an important factor. Some flavors are often added (masking) to reduce the bean-like flavor in soymilk. This type of soymilk is drunk in North-east Asia. Another type of reconstituted soymilk is the Western-style soymilk (characterized by an off-bean-like flavor) produced by modern methods (Cornell, USDA, and Illinois). This white soymilk has a dairy-like taste and is often ultrapasteurized. It is packed into aseptic Tetra brik packs or retort standing pouches and is distributed and sold without the need for refrigeration. This type of soymilk is becoming a very popular beverage in North America and Europe.

Composition and Nutrients

Table 1 lists the compositions of traditional soymilk, bovine milk, and human milk. The protein content of soymilk is slightly higher than that of bovine and human milks, the lipid and carbohydrate contents are 30% lower, and the calories per 100 g are 30% lower. Therefore, the constituents of soymilk can be adjusted to be the same as those of bovine milk. In terms of minerals, the calcium content of soymilk is 50% lower than that of human male, whereas the potassium content is twice as high. The iron content of soymilk is 12 times higher than that of bovine and human milk, and so it is effective in treating anemia.

Table 1 Compositions of soymilk, bovine milk, and human milk

Contents	*Soymilk*	*Bovine milk*	*Human milk*
Water (g/100 g)	90.8	88.7	88.0
Calories (g/100 g)	192	247	272
Protein (g/100 g)	3.6	2.9	1.1
Lipid (g/100 g)	2.0	3.2	3.5
Carbohydrates (g/100 g)	2.9	4.5	7.2
Ash (g/100 g)	0.5	0.7	0.2
Calcium (mg/100 g)	15	100	27
Sodium (mg/100 g)	2	50	15
Potassium (mg/100 g)	90	150	48
Phosphorus (mg/100 g)	49	90	14
Iron (mg/100 g)	1.2	0.1	0.1
Thiamine (B_1) (mg/100 g)	0.03	0.03	0.01
Riboflavin (B_2) (mg/100 g)	0.02	0.15	0.03
Niacin (mg/100 g)	0.05	0.1	0.2

Soymilk contains small amounts of the the lipid-soluble vitamins (A, D, and E) but slightly lower levels of the vitamin B groups (B_1, B_2, niacin) than the other types of milk. (*See* **Protein**: Food Sources; Quality; **Vitamins**: Overview.)

Soymilk contains less saturated fat and cholesterol than bovine and human milk. Nutritionists recommend that less saturated fat and cholesterol be consumed so as to reduce the risk of heart disease. It has been reported that a soy protein intake of about 25 g per day (roughly the equivalent of four glasses of soymilk per day) reduces the risk of coronary heart disease and lowers blood cholesterol levels. Evidence for these two claims is so compelling that since October 1999, the US Food and Drug Administration (FDA) has allowed manufacturers of soy products to cite these claims on the packaging and in advertising. Soymilk may play a more important role in keeping the heart healthy than bovine milk.

Soymilk does not contain any lactose, and it is an ideal milk for lactose-intolerant people. Its popularity is now increasing in North America and Europe. Asians have a low incidence of estrogen-related cancers and a low prevalence of climacteric syndrome and osteoporosis, and it has been suggested that this is because they consume a large number of soybean products, which are high in phytoestrogens such as isoflavones. Soymilk contains 40–60 mg of isoflavones per 100 g. It has been reported that consumers in Japan and Shanghai consume 20–30 mg per day on average. (*See* **Lactose**.)

Structure of Soymilk

Soy milk is a milky liquid and contains mainly protein, lipid, and glucide. bovine milk is also an milky liquid because it contains micellar proteins (100 nm mean diameter) and Lippitt globules of (5 μm mean diameter). Soymilk is known to contain a particular protein (18 nm mean diameter) and oil globules (350 nm mean diameter). Half of the protein in soymilk exists in a particular form, and half is in soluble form. The glucide in soymilk consists of sugar and oligosaccharides, which are soluble in solution. The colloidal solution forms a curd like bovine milk. Soymilk can be made to curdle by adding coagulants to make soy cheese (tofu).

See also: **Milk**: Physical and Chemical Properties; **Soy (Soya) Beans**: The Crop; Processing for the Food Industry; Properties and Analysis; Dietary Importance; **Soy (Soya) Cheeses**

Further Reading

Fournier DB (1998) Soy, its components, and cancer prevention: A review of the *in vitro*, animal, and human data. *Cancer Epidemiology, Biomarkers & Prevention* 7: 1055–1065.

Geetha P (2000) *Proceedings of the Third International Soybean Processing and Utilization Conference*. Japan: Korin Publishing Co.

Guo W (1993) *The History of Soybean Cultivation in China*. People's Republic of China: Hehai University Publishing Co.

Kawase S, Nomura M, Ohnishi C *et al*. (2000) *Proceedings of the Third International Soybean Processing and Utilization Conference*. Japan: Korin Publishing Co.

Messina M, Persky V, Setchell KDR and Barnes S (1994) *The Simple Soybean and Your Health: How Soy Foods can Lower Your Cholesterol and Reduce Your Risk*. USA: Avery.

Mindell E (1995) *Earl Mindell's Soy Miracle*. New York: Simon & Schuster.

Ono T (2000) *Proceedings of the Third International Soybean Processing and Utilization Conference*. Japan: Korin Publishing Co.

Potter SM, Baum JA, Teng H *et al*. (1998) Soy protein and isoflavones: their effects on blood lipids and bone density in postmenopausal women. *American Journal of Clinical Nutrition* 68: 1375–1379.

Resources Council, Science and Technology Agency (2000) *Standard Tables of Food Composition in Japan*, 5th rev. edn. Tokyo, Japan: Publishing in the Ministry of Finance.

Setchell KDR and Cassidy A (1999) Dietary isoflavones: Biological effects and relevance to human health. *Journal of Nutrition* 129: 758–767.

Stein K (2000) FDA approves health claim labeling for foods containing soy protein. *Journal of the American Dietetic Association* 100: 292.

Tsukamoto C, Abe K, Yoshida T and Ono T (2000) *Proceedings of the Third International Soybean Processing and Utilization Conference*. Japan: Korin Publishing Co.

Watanabe S, Uesugi S, Zhuo XG *et al*. (2000) *Proceedings of the Third International Soybean Processing and Utilization Conference*. Japan: Korin Publishing Co.
Watanabe T (1997) *Science of Tofu – Easy to Understand*. Kyoto, Japan: Food Journal Co.
Yamori Y (2000) *Proceedings of the Third International Soybean Processing and Utilization Conference*. Tokyo, Japan: Korin Publishing Co.
Yamori Y, Ohta S and Watanabe S (2001) *Soybean Isoflavones*. Tokyo, Japan: Saiwaishobo.

Soy (Soya) Sauce *See* **Fermented Foods**: Origins and Applications; Fermented Meat Products; Fermentations of the Far East; Beverages from Sorghum and Millet

SPECTROSCOPY

Contents

Overview

R H Wilson, Institute of Food Research, Colney Lane, Norwich, UK

Background

The term 'spectroscopy' encompasses a range of techniques for acquiring information on atomic and molecular structure. In all cases there is absorption or emission of electromagnetic radiation, but the basic processes by which this occurs vary considerably, as does the nature of the information that can be obtained. This article provides a broad overview of the important processes that give rise to the most important of the chemical analytical spectroscopies in food and nutrition.

The Electromagnetic Spectrum

When light interacts with matter it may stimulate transitions between quantized energy levels. Light is considered to travel in discrete packets (quanta) of energy called photons. The exact type of transition stimulated depends upon the energy (E) of the photon which in turn is related to the frequency (υ) by:

$$E = h\upsilon \qquad (1)$$

where h is Planck's constant.

The energy of the quanta is highest at the γ-ray end (10^6 eV) and lowest at the radiofrequency end (10^{-8} eV) of the electromagnetic spectrum. The full electromagnetic spectrum is shown in **Figure 1**. In this discussion a more detailed examination will concentrate on the region between the ultraviolet and the radiofrequency end of the spectrum, as this is where most routine chemical spectroscopy is carried out. Spectroscopy is usually concerned with the measurement of a specific spectrum, which is a measure of energy absorbed or emitted as a function of wavelength (or frequency) across a limited region. The objective is to use the spectrum to obtain information on molecular structure or for quantitative analysis. The types of transition that are normally stimulated range from electronic to vibrational through rotational to the low-energy nuclear transitions that are examined in nuclear magnetic resonance (NMR). There are also high-energy nuclear transitions that can be stimulated by γ-radiation and are used for Mössbauer spectroscopy, but this will not be considered in this article.

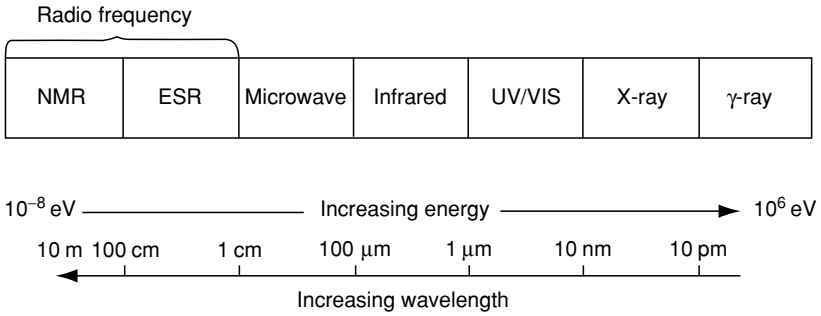

Figure 1 The electromagnetic spectrum. NMR, nuclear magnetic resonance; ESR, electron spin resonance; UV/VIS, ultraviolet and visible. Reproduced from Spectroscopy: Overview, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Since the energy required to stimulate an electronic transition is greater than that for a vibrational transition, which in turn is greater than for rotational transition, it is found that more than one transition is usually stimulated so that, for example, pure vibrational spectra are not seen and are nearly always complicated by rotational transitions.

The normal way for interaction of the radiation to occur is through one of four processes: (1) absorption, (2) emission, (3) elastic scatttering, although there is not net energy absorption, and (4) inelastic scattering. These mechanisms are described in more detail under the appropriate headings as the spectroscopies are discussed in more detail. The first to be discussed are ultraviolet and visible (UV/VIS) spectroscopies, both of which involve electronic transitions.

Electronic Transitions

In UV/VIS spectroscopies, absorption of radiation is the result of the excitation of bonding electrons. The types of bonds that give rise to absorption are known as chromophores and in the UV the electrons of the chromophore are either directly used in bond formation or are nonbonding or unshared outer electrons of an electronegative atom, such as oxygen, nitrogen, or sulfur. The general mechanism in a chromophore such as C═C, in which orbitals are used for bonding, involves the promotion of an electron in a bonding π orbital into a nonbonding π^* orbital (a so-called $\pi-\pi^*$ transition), which typically requires about 7 eV, corresponding to a wavelength of 180 nm. It is also possible for a nonbonding electron to be promoted to a π^* (an $n-\pi^*$ transition). These two are the most common transitions, although similar ones exist for single (σ) bonds ($n-\sigma^*$ or $\sigma-\sigma^*$). However, because the latter required much higher energies, they are seen in the vacuum-UV and are harder to observe. The frequencies of the absorption can be influenced by solvents and by delocalization in conjugated systems.

Transition-metal ions absorb in the UV/VIS region and the transitions responsible involve 4f and 5d electrons. Alternatively, in some inorganic complexes the process of charge-transfer absorption occurs.

Most UV/VIS spectroscopy involves absorption processes and normally a spectrophotometer is used to measure a spectrum. The major components are a source, a dispersing system, and a detector. Normally, light from a suitable source is passed to a prism or grating where it is dispersed into its component frequencies. The dispersing element may be rotated so that each frequency is passed in turn through a narrow slit. This light may be divided so that half passes through a channel containing the sample and half through a reference channel. The emerging beams can be directed in ratios at a detector and the absorbance of the sample as a function of frequency (or wavelength), i.e., the spectrum, can be plotted.

It should be noted that UV/VIS spectra do not consist of discrete lines. The reason is that the high energy of the UV/VIS region can be transferred into the vibrational and rotational substates so that both types of transition are simultaneously stimulated. In Figure 2 the energy level diagram for a chromophore is shown. E_0 and E_1 represent the ground and excited electronic energy levels of a molecule. Each electronic level has associated with it various vibrational sublevels, $\nu_{0,1}$ $\nu_{0,2}$, etc., which in turn have rotational sublevels. An electron may be promoted from the E_0 to E_1 electronic state but may go from the $\nu_{0,0}$ to $\nu_{1,0}$ or $\nu_{1,1}$ state; i.e., there is a simultaneous vibrational transition. The range of vibration subtransitions possible, combined with rotational transitions, means that there is no discrete frequency at which transition occurs.

Furthermore, if an electron is promoted from the $\nu_{0,0}$ state of E_0 to the $\nu_{1,1}$ state of E_1 it may lose energy by collision, for example, and may become lowered into the $\nu_{1,0}$ state. During relaxation to the ground electronic state a photon is emitted of

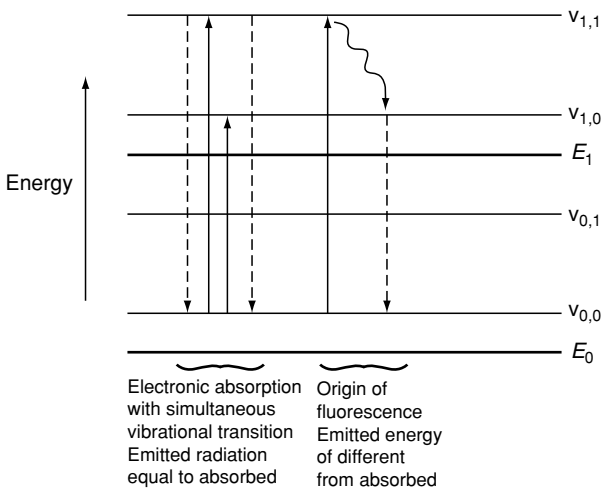

Figure 2 Energy levels for a chromophore, showing electronic and vibrational levels. Reproduced from Spectroscopy: Overview, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

different energy from that absorbed, and this process is called fluorescence.

Vibrational Spectroscopy

The transitions between the vibrational energy levels are the basis of infrared and Raman spectroscopies. The infrared region is divided into the near, middle (or mid) and far infrared. This division is on the basis of instrumental factors as well as the types of vibration that occur in each region. It is easiest to consider first the middle infrared, which is usually considered to lie between 2.5 and 25 μm in wavelength. It is common practice, however, for vibrational spectroscopists to use the unit wavenumber (reciprocal of the wavelength in centimeters) rather than wavelength; this has units of cm^{-1}, which is a frequency term. The middle infrared then stretches from 4000 to 400 cm^{-1}.

The bond between two atoms can be considered to be rather like a spring that has a certain strength of force constant (k). The bond, or spring, can be stretched and caused to oscillate. It will do so at some natural frequency, f, that depends upon k and the masses of the atoms according to Hooke's law:

$$f = (1/2\pi)\sqrt{k/\mu} \qquad (2)$$

where μ is the reduced mass given, defined as:

$$\mu = (m_1 m_2)/(m_1 + m_2) \qquad (3)$$

m_1 and m_2 being the masses of the individual atoms constituting the bond. Equation (2) shows that a particular bond will give rise to a characteristic frequency that depends upon the masses of the atoms and the strength of the bond. Therefore, a C═O bond, which has a greater force constant than a C—O bond, will have a vibrational frequency which is larger. In practice, different functional groups give rise to characteristic vibrational frequencies. This is the major use for vibrational spectroscopy; it is a highly useful probe for the identification of functional groups and for structural determination. It has greater selectivity than UV/VIS in this respect.

Equation (2) is derived from classical physics, but of course the actual process is quantized and (2) should be written as:

$$F = (h/2\pi)\sqrt{k/\mu} \qquad (4)$$

where h is Planck's constant.

In the main, an infrared spectrum is generated by absorption using a similar arrangement to that used for UV/VIS but with different source, detector, and dispersing optics. The process is illustrated in Figure 3. The energy level diagram shows the ground ($v_{0,0}$) and excited ($v_{1,0}$) vibrational states. Also shown are the various rotational substates (J' and J''). Excitation can occur from $v_{0,0}$ ($J''=0$) to $v_{1,0}$ ($J'=0$), corresponding to the band center of the absorption band. However, excitation from the $J''=1$ to $J'=1$ rotational substates (i.e., $\Delta J=0$) will produce a slightly different frequency as the rotational sublevels are not equally spaced. It is also possible for ΔJ to be ± 1, giving rise now to a complicated absorption

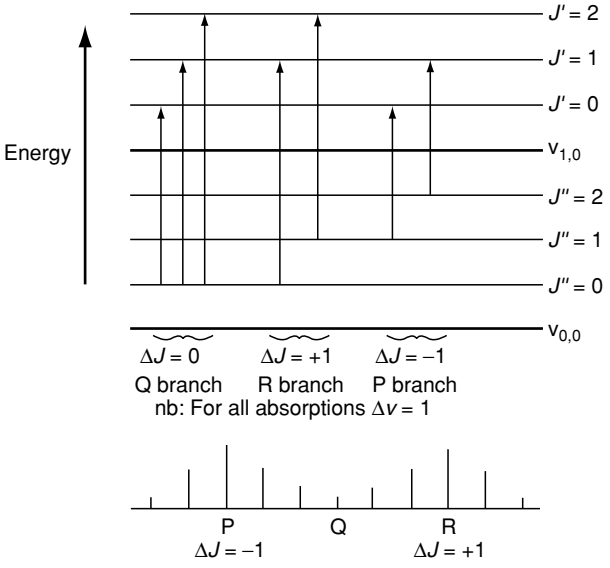

Figure 3 Energy levels for infrared transitions, showing vibrational and rotational levels. Reproduced from Spectroscopy: Overview, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

band comprising of a central absorption (Q branch) with equally spaced lines either side called the P branch ($\Delta J = -1$), and the R branch ($\Delta J = +1$). This structure is only seen as such in the gas phase. In solid or solution state the result is that a broad absorption rather than a sharp line is seen. However, no absorption will be seen at all unless the selection rule is applied. This states that for absorption to occur there must be a change in dipole moment during the vibration. Consequently, homonuclear bonds do not absorb.

At normal room temperature most molecules will be in the ground vibrational state. However, as the temperature is increased, a more significant population will develop in the excited state. As a result transitions from the ν_1 to ν_0 state can occur with the emission of a photon. This is the process of infrared emission which is, albeit rare, an alternative to absorption spectroscopy. In this case the (heated) sample acts as the infrared source.

Infrared spectroscopy has, until recently, been of little use for industrial, biological, and food use owing to the difficulties of sample handling and the time of data acquisition. However, the recent development of Fourier transform methods involving the replacement of the dispersing element with an interferometer has benefits of increased speed, throughput, and frequency reproducibility. Coupled with new methods of sample presentation, this has led to a reawakening of interest in the middle infrared.

The absorptions in the middle infrared are known as the fundamentals. However, various overtones and combinations of the fundamentals can arise. For example, a molecule with two fundamentals at frequencies ν_1 and ν_2 may give overtones at $2\nu_1$, $3\nu_1$, $4\nu_1$, or $2\nu_2$, etc., or combinations at, say, $\nu_1 + \nu_2$ or $2\nu_1 + \nu_2$. In practice, not all fundamentals give rise to overtones, usually only bonds in which a heavy atom such as N or O is coupled to hydrogen. The overtone and combinations constitute the near infrared ($2.5-0.7\,\mu$m) which, despite the apparent complexity of the spectra, has found considerable application to food problems.

Raman Spectroscopy

If a sample is illuminated with monochromatic visible light, it is found that much of the light is scattered and that the scattered light is of the same frequency as the illuminating light. This is elastic or Raleigh scattering. However, analysis shows that a small amount ($< 10^{-6}$) of the incident radiation is scattered with a different frequency. A series of lines is found with frequencies less than the incident light. A weaker series is found with higher frequencies. When the former set of lines are presented as a spectrum of intensity versus frequency shift, the result is something similar to an infrared spectrum with the shift scale from about 4000 to 20 cm^{-1}. This effect is the Raman effect and the spectrum is called the Raman spectrum. The lines comprising the spectrum are called Stokes lines. Those of higher frequency than the exciting line are called anti-Stokes lines and consist of the same peaks with the same shift, but there may be different intensity ratios.

The electrical field of the incident radiation interacts with the electrons in the sample and causes periodic polarization and depolarization so that energy is momentarily absorbed in a distorted, polarized state or virtual state. Most molecules relax by the emission of energy of the same frequency to that absorbed. In a few cases some of the energy will be dissipated amongst the vibrational energy levels, causing vibrational excitation and giving rise to the Raman spectrum. Even fewer molecules will not be in the ground vibrational state before excitation but in the virtual state may relax and the emitted photon will be of higher energy than that incident, leading to the anti-Stokes lines. In contrast to infrared spectroscopy, the selection rule for absorption is that, during vibration of the bond, there must be a change in the electronic polarizability. There is thus a distinct difference in the two spectra and vibrations that may be weak or absent from infrared spectra, e.g., C—C are present and perhaps strong in the Raman. The two spectroscopies are thus complementary and together provide a complete picture of the vibrational states of a molecule.

Far Infrared/Microwave

To complete the picture, at lower energy there is the far infrared ($400-10\,\text{cm}^{-1}$), which has major applications in inorganic chemistry as bonds between metals and organic ligands appear here as well as skeletal vibrations of molecular backbones. This region is of limited application in food and nutritional studies.

In the microwave region, at even lower energy, pure rotational spectra can be produced. However, they will not be addressed here as this is also of limited application.

At the radiofrequency end of the spectrum is NMR spectroscopy which involves transitions between magnetic quantum levels of atomic nuclei. Nuclei have properties of spin and magnetic moment. Splitting of the energy levels can be induced by placement in a magnetic field and transitions can be induced by the application of radiofrequency radiation. Today, this is usually achieved by irradiating the sample exposed to a high magnetic field with a pulse of broadband radiation. After excitation the nuclei reemit

energy at their resonance frequencies and the observed signal is a combination of these frequencies, and this decays with time. A spectrum can be produced by Fourier transformation of this decaying signal. The usefulness of the technique lies in the fact that the resonance frequency of a given nucleus depends upon its chemical environment. However, the range of NMR experiments possible is very large indeed and it is a very powerful method for structural analysis.

In the food industry the use of NMR spectra as such is increasing but relaxation time measurements are still more important, particularly in the determination of solid/liquid ratios. The relaxation rate from the excited state depends on environmental factors and molecular mobility.

Absorption Laws

In UV, near infrared, and mid-infrared adsorption spectroscopy, the fundamental law governing adsorption is the Beer–Lambert relationship. For a sample illuminated by radiation of intensity I_0 the amount transmitted, I, is given by:

$$I = I_0 e^{-\varepsilon c l} \tag{5}$$

where c is the concentration of absorbing species, l is the pathlength through which the light passes, and ε is the molar absorptivity.

For quantitative analysis, spectra are usually presented in absorbance units, where absorbance, A, is defined as

$$A = -\log(I/I_0) = \varepsilon c l \tag{6}$$

so that A is directly proportional to the concentration at constant pathlength.

Practically, optical spectroscopy requires that ε be determined for any absorbing species. This is achieved by calibration and their absorbance is measured. Solutions of the sample to be determined are prepared at known concentration and their absorbances are measured. When the latter are plotted against the concentration, a linear plot results of slope ε. Unknown concentrations can be calculated by measuring absorbance and interpolating from the calibration curve.

Deviations from the Beer–Lambert relationship can occur if too wide a range of concentration is chosen so that solute–solute interactions occur, or where there is chemical interaction between components.

A particular problem that exists in the near and mid-infrared is where significant overlap of absorbance peaks occurs. Clearly, the absorbance at a given wavelength may then depend upon more than one concentration, so that:

$$A = \varepsilon_1 c_1 = \varepsilon_2 c_2 + \varepsilon_3 c_3 \ldots \tag{7}$$

Hence, more complicated solutions to the Beer–Lambert relationship may be required for multicomponent analysis. Such methods include p and k matrix, partial least-squares, or principal components regression.

In NMR single-pulse experiments the signal observed is directly proportional to the number of nuclei, provided sufficient time is allowed between pulses for the reestablishment of equilibrium. Under such circumstances the NMR experiment is quantitative and requires no calibration. Double-resonance experiments can, however, lead to enhanced signals for certain nuclei (nuclear Overhauser effect) so that some form of calibration is necessary. In relaxation measurements the magnetization decay can be broken down into components from fast (solid) and slow (liquid) components, the relative magnitude of each reflecting the relative concentrations.

Practical details and applications of the most relevant spectroscopies in food analysis can be found in following chapters.

See also: **Spectroscopy**: Infrared and Raman; Near-infrared; Fluorescence; Atomic Emission and Absorption; Nuclear Magnetic Resonance; Visible Spectroscopy and Colorimetry

Further Reading

Andrews DL (ed.) (1990) *Perspectives in Modern Chemical Spectroscopy*. Berlin: Springer-Verlag.

Banwell CN and McCash EE (1994) *Fundamentals of Molecular Spectroscopy*, 4th edn. London: McGraw Hill.

Colquhoun IJ and Goodfellow BJ (1994) Nuclear magnetic resonance spectroscopy. In: Wilson RH (ed.) *Spectroscopic Techniques for Food Analysis*. New York: VCH.

Wilson RH (ed.) (1994) *Spectroscopic Techniques for Food Analysis*. New York: VCH.

Infrared and Raman

B Schrader, Essen, Germany
S Keller, Hamburg, Germany

Principles

A molecule can be identified uniquely by its vibrational spectrum, since this depends on every structural feature. The vibrational spectrum consists of the frequencies and intensities of the vibrations

observed by infrared and Raman spectroscopy. Both are nondestructive methods. This means that the samples can be investigated afterwards by other methods. The intensity of the bands in the infrared and Raman spectra gives complementary information about the nature of the vibrating bonds. Therefore, it is advisable to evaluate both the infrared and Raman spectra in order to obtain the maximum amount of analytical information.

Now, the advantages of both methods, IR and Raman spectroscopy, can be employed by powerful spectrometers, thus making vibrational spectroscopy a very useful tool in food analysis. State-of the-art instruments for vibrational spectroscopy are equipped with interferometers coupled to computers for the transformation of interferograms into spectra, so-called Fourier transform infrared (FTIR), Fourier transform near-infrared (FTNIR), or FT Raman spectrometers. Compared with grating instruments, interferometers have several advantages, in particular the fact that the radiant flux analyzed is larger by about two orders of magnitude. Interferometers therefore permit reliable routine analyses of small samples and of low concentrations.

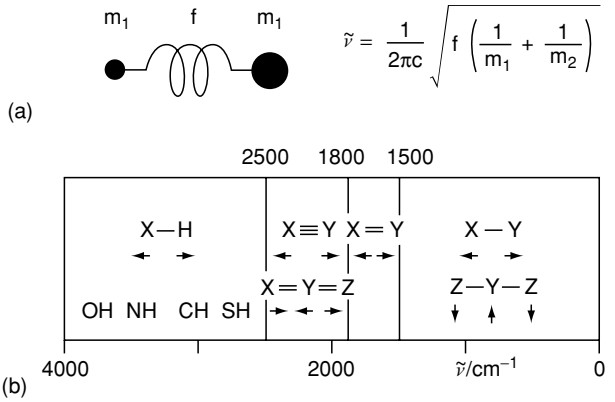

Figure 1 Principles of infrared and Raman spectroscopy: (a) model and equation describing frequency of diatomic molecule; (b) frequency ranges of different small molecules.

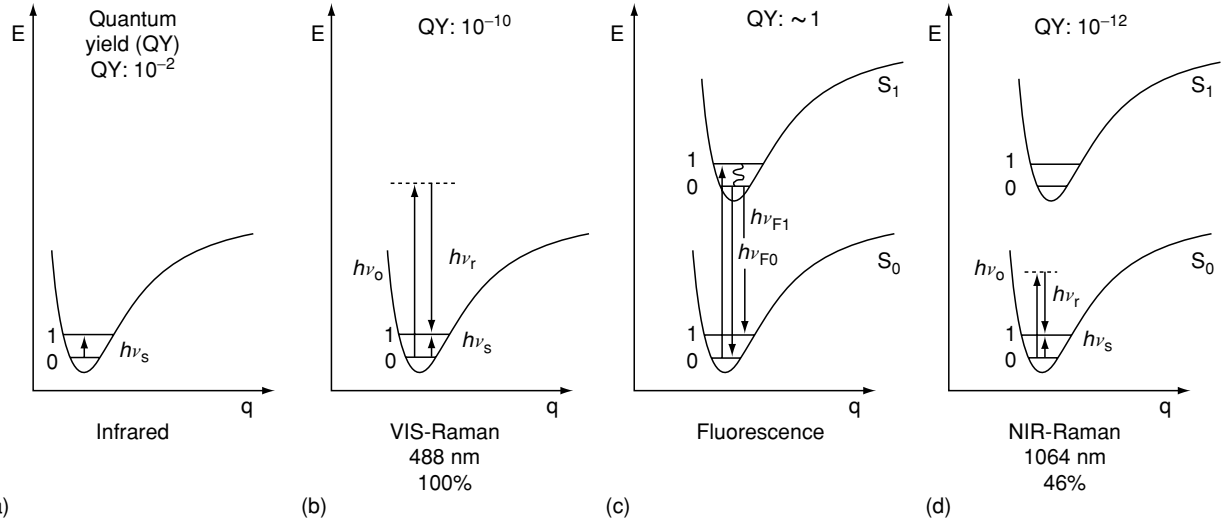

Figure 2 Term scheme of spectra: (a) infrared; (b) Raman (excitation in the VIS); (c) fluorescence; (d) Raman (excitation in the NIR).

Table 1 Comparison of infrared, near-infrared and Raman spectrometry

	Infrared spectrometry	*Raman spectrometry*	*Near-infrared spectrometry*
Sample preparation	Partly expensive, moisture-sensitive	Simple	Simple
Sample container	Cells of KBr, NaCl, ZnSe, Si, CsI, TlBr/TlJ	Glass, quartz	Glass, quartz
Bands of water	Strongly disturbing	Low intensity	Strong intensity
Observed vibrations	Antisymmetric polar groups, substituents	Symmetric unpolar groups, skeletons	Overtones and combinations of C—H, N—H and O—H
Intensity, I, ⇔ concentration, c	$\log(I_0/I) \sim c$	$I \sim c$	$\log(I_0/I) \sim c$
Spectral range	400–4000 cm^{-1} standard, 10–400 cm^{-1} with FIR-optics	10–4000 cm^{-1}	4000–12 500 cm^{-1}

Table 2 Characteristic frequencies and Raman and infrared intensities of food components

Assignment	*Infrared (cm^{-1})*	*Raman (cm^{-1})*
δ(C—O), sugar		479
ν(S—S)		520–540
Phenylalanine (skeletal)		622
Tyrosine (skeletal)		642
Adenine		725
Thymine		750
Breathing vib. (α-anomers, Type III, sugar)	756–776	
Tryptophan		759
Breathing vib. (β-anomers, Type III, sugar)	765–783	
Cytosine, uracil (ring, str)		785
ν(—O—P—O—)$_s$, A-helix		810–814
Tyrosine		829
ν(—O—P—O—)$_s$, B-helix		830
C_1—H bending (α-anomers, Type II, sugar)	836–852	840–850
Tyrosine		852
C_1—H bending (β-anomers, Type II, sugar)	884–898	890–910
Ring vibration (α-anomers, Type I, sugar)	904–930	
Ring vibration (β-anomers, Type I, sugar)	915–925	
$\nu(PO_4)_{s,as}$		960
ν_{as}(C—N^+—C), choline	970	
Phenylalanine		1004
ν(C—N) and ν(C—C)		1061
Deoxyribose – Z-DNA	1065	
ν(O—P═O)$_s$	1070–1100 (m)	1080–1100 (s)
ν(C—N) and ν(C—C)		1093
ν(C—N), ν(C—C)		1129
ν(C—C), carotenoids		1157
$\nu(PO_4^{3-})_{as}$ – Z-DNA	1215	
ν(O—P═O)$_{as}$	1225 (s)	1225 (vw)
Amide III, (β-structure)		1225–1245 (s)
$\nu(PO_4^{3-})_{as}$ – B-DNA	1225	
$\nu(PO_4^{3-})_{as}$ – A-DNA	1240	
Amide III (random coil)		1241–1251
$\nu_{as}(PO_2^-)$	1250	
ν(═C—H)$_{i.p.b.}$ unsaturated fatty acids		1267
Amide III (α-helix)		1270–1300 (vw, br)
$\delta(CH_2)$, twisting		1295
γ_w $(CH_2)_n$	1342–1180	
Tryptophan		1358
$\nu(CH_3)$		1374
Aspartic acid, $\nu(COO^-)$		1408
Deoxyribose – A-DNA	1418	
Deoxyribose – B-DNA	1425	
Glutamic acid, $\nu(COO^-)$		1437
δ(C—H)		1445
$\delta(CH_2)$	1470	
Amide II, δ(N—H) (antiparallel β-sheet)	1510–1540	
ν(C═C), aromate (tyrosine), carotinoids	1515	1519
δ(N—H), amide II (β-sheet structure)	1530–1545	
Amide II, δ(N—H) (α-helix)	1540–1560	
Amide II, δ(N—H) (random coil)	1535	
Amide II, δ(N—H) (parallel β-sheet)	1550	
Guanine, adenine (ring, str)		1575
Tryptophan		1578
Phenylalanine		1606
$\nu(COO^-)$, carboxylate	1610	
Tyrosine		1614
Tryptophane		1621
Amide I ν(C═O) (parallel β-sheet structure)	1632–1648	
Amide I, ν(C═O), (α-helical structure)	1655	1658
Amide I, ν(C═O) (random coil)	1655–1660	1665
ν(C═C) unsaturated fatty lipids		1657
Amide I, ν(C═O) (antiparallel β-sheet and β-turn structure)	1685–1700	1665–1675
– Z-DNA	1695	
– A-DNA	1705	
– B-DNA	1715	
ν(C═O), ester	1720–1750	1746
ν(C═O), carboxylic acid	1730	
$\nu(CH_2)_s$ lipid	2849	2847
Methyl	2872	
$\nu(CH_2)_s$, protein, $\nu(CH_2)_{as}$, lipid		2890
$\nu_{as}(CH_2)$	2918	
$\nu(CH_3)_s$, lipid, $\nu(CH_2)_a$, protein		2935
$\nu_{as}(CH_3)$	2956	
$\nu(CH_3)_{as}$, protein, $\nu(CH_3)_{as}$, lipid		2975
$\nu_{as}(CH_3)\ {}_3N^+$	3028	
ν(C═C—H)$_s$, lipid		3010
ν(NH), ν(OH), base sugar	3300–3600 (m)	

Basic Theory

In Figure 1, the theory necessary to understand infrared and Raman spectroscopy is outlined. Every molecule can be considered as being built up from atoms with a definite mass, connected by chemical bonds behaving like elastic springs (Figure 1a). The frequency, $\tilde{\nu}$, is usually measured in wavelengths per centimeter, called the wavenumber (cm^{-1}). For a diatomic molecule of two bound atoms, the frequency of the vibration is proportional to the square root of the force constant (f), – which is approximately proportional to the bond order – and the sum of the reciprocal masses.

For larger molecules, the vibrations of the fragments couple and produce complicated 'fingerprint' patterns. However, most of the structural units keep their typical frequency range. Thus, X—H bonds, with X any element, show their typical vibrations between 4000 and 2500 cm^{-1} (Figure 1b), followed by groups with triple bonds or cumulated double bonds in the range 2500–1800 cm^{-1}. Groups with 'double' bonds show their characteristic vibrations between 1800 and 1500 cm^{-1}, and groups

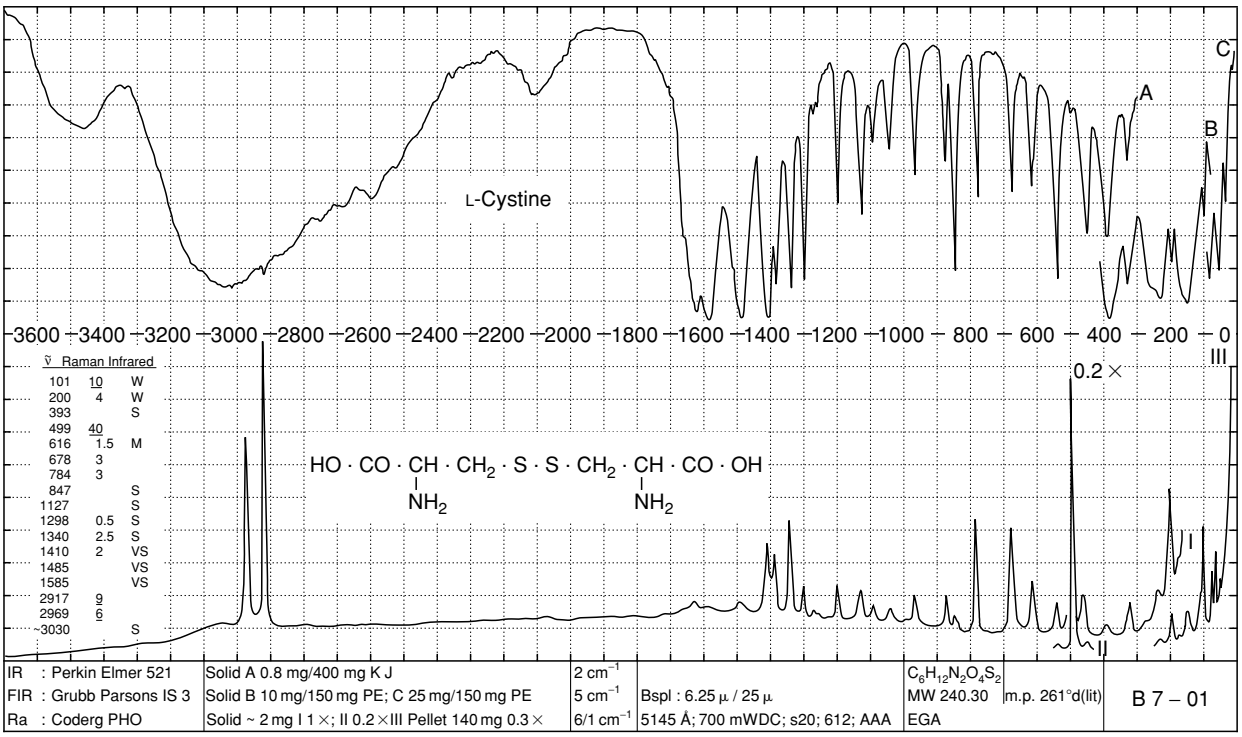

(a)

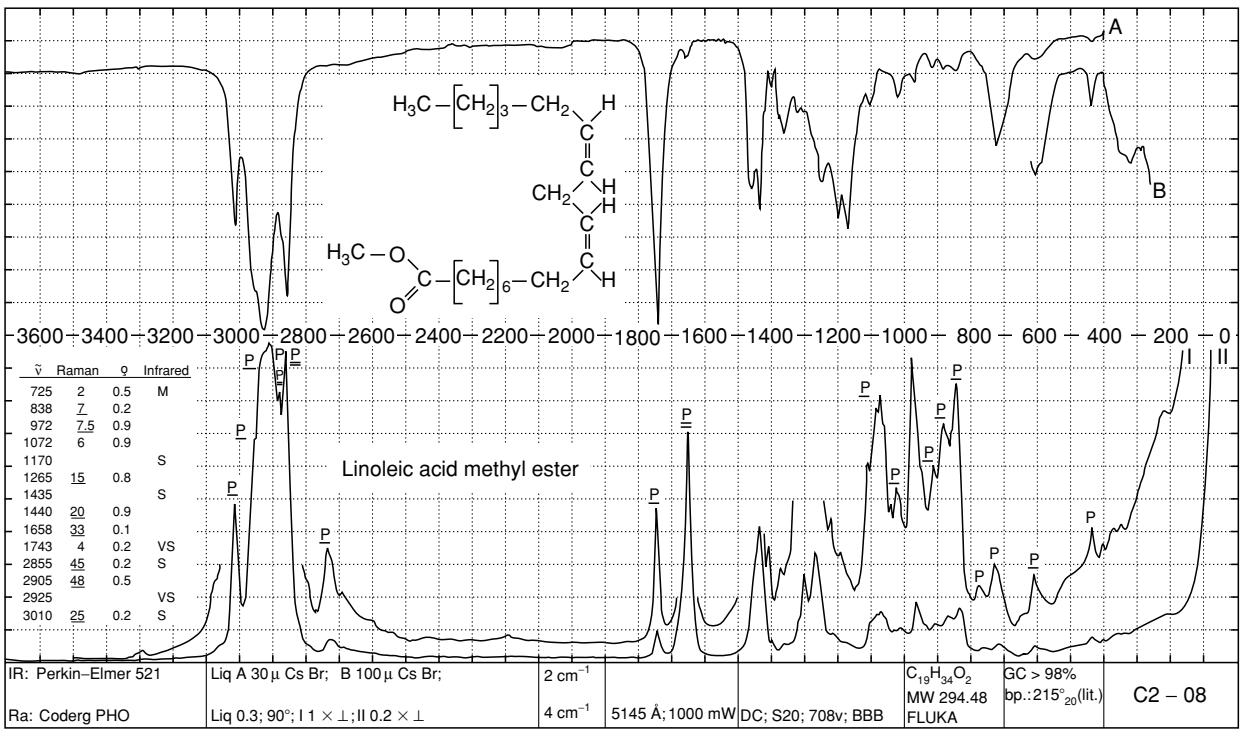

(b)

Figure 3 Infrared transmission (above) and Raman spectrum (below) of: (a) cystine; (b) linolenic acid methyl ester; (c) retinol, vitamin A; (d) caffeine; (e) thiamine hydrochloride, vitamin B_1; (f) L-ascorbic acid, vitamin C; (g) α-D-glucose; (h) sucrose. From Schrader B (1989) *Raman/Infrared Atlas of Organic Compounds*, 2nd edn. Weinheim: VCH Verlagsgesellschaft with permission.

Figure 3 Continued

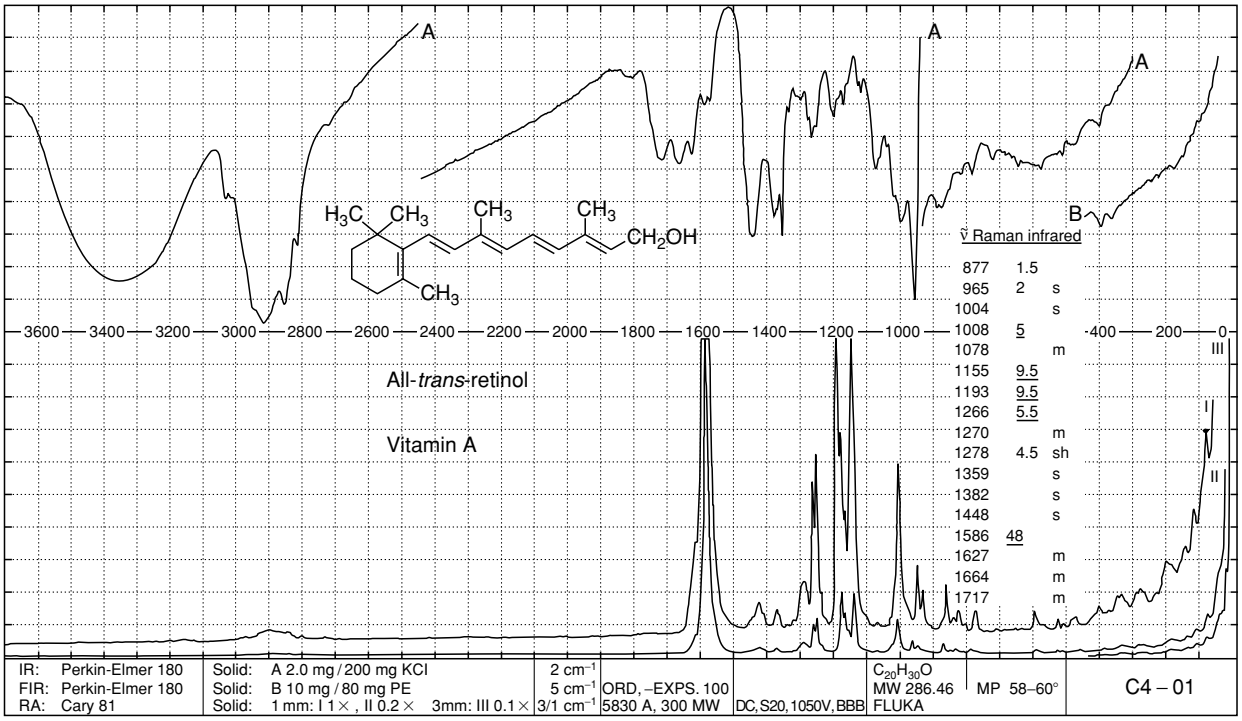

(c)

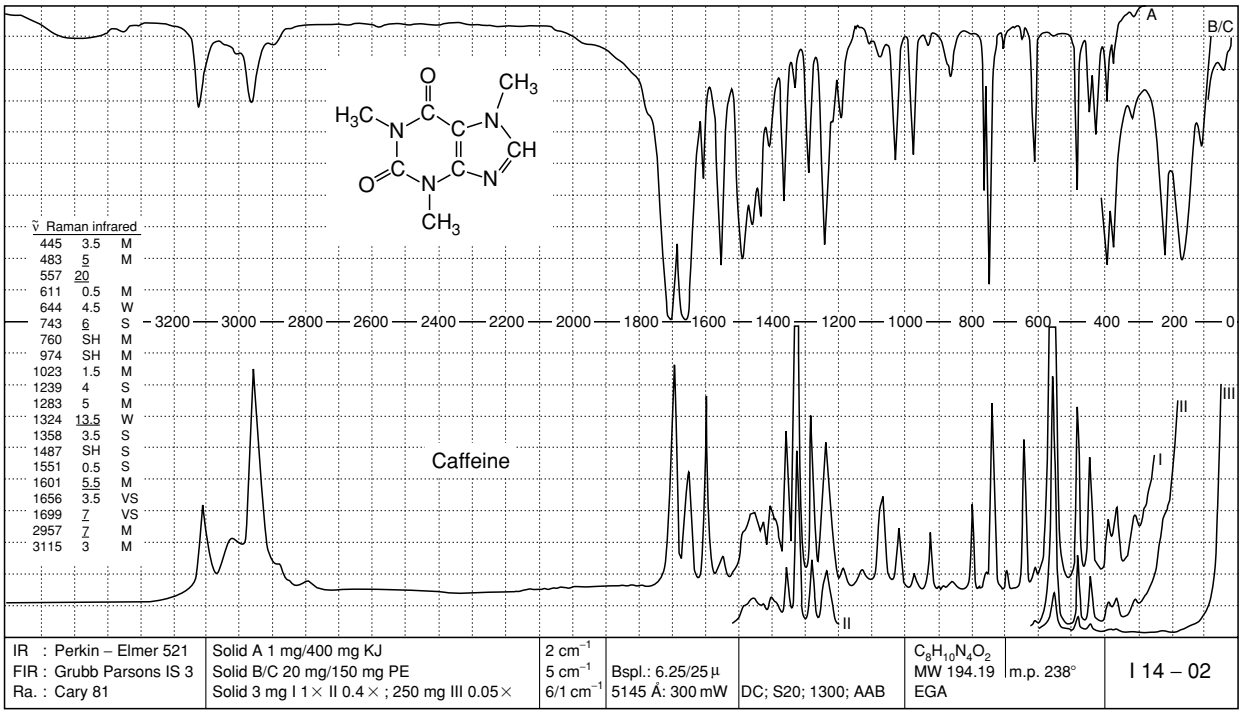

(d)

Continued

Figure 3 Continued

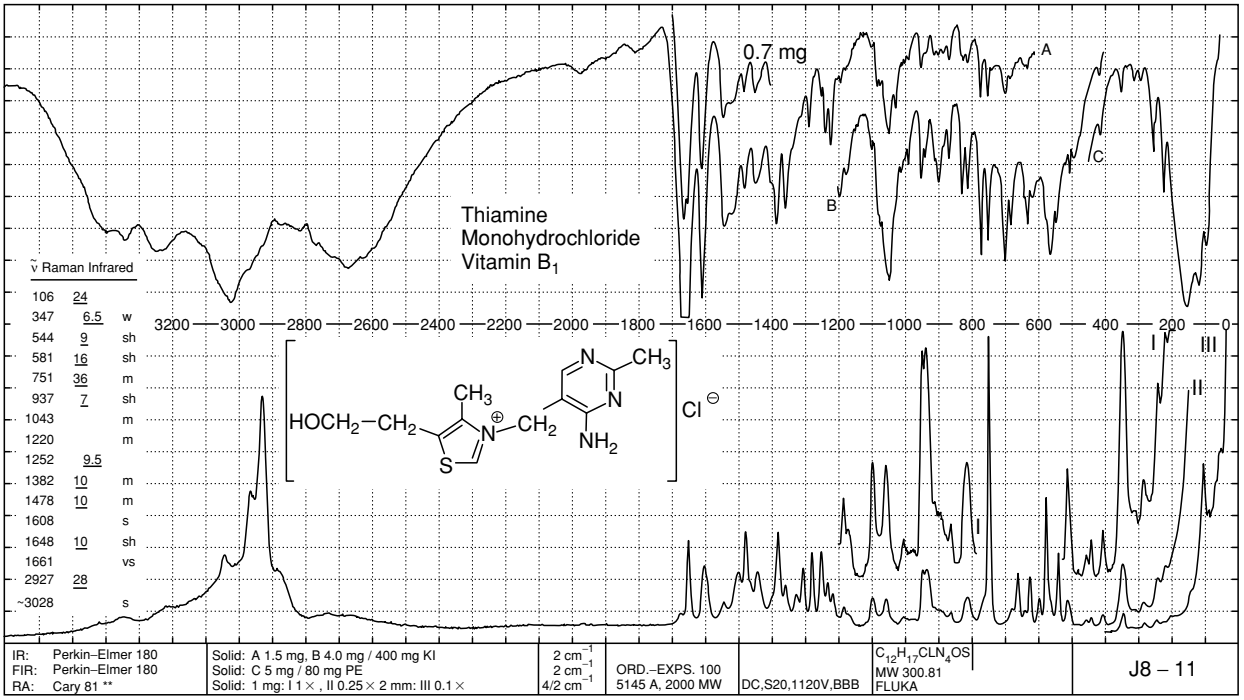

(e)

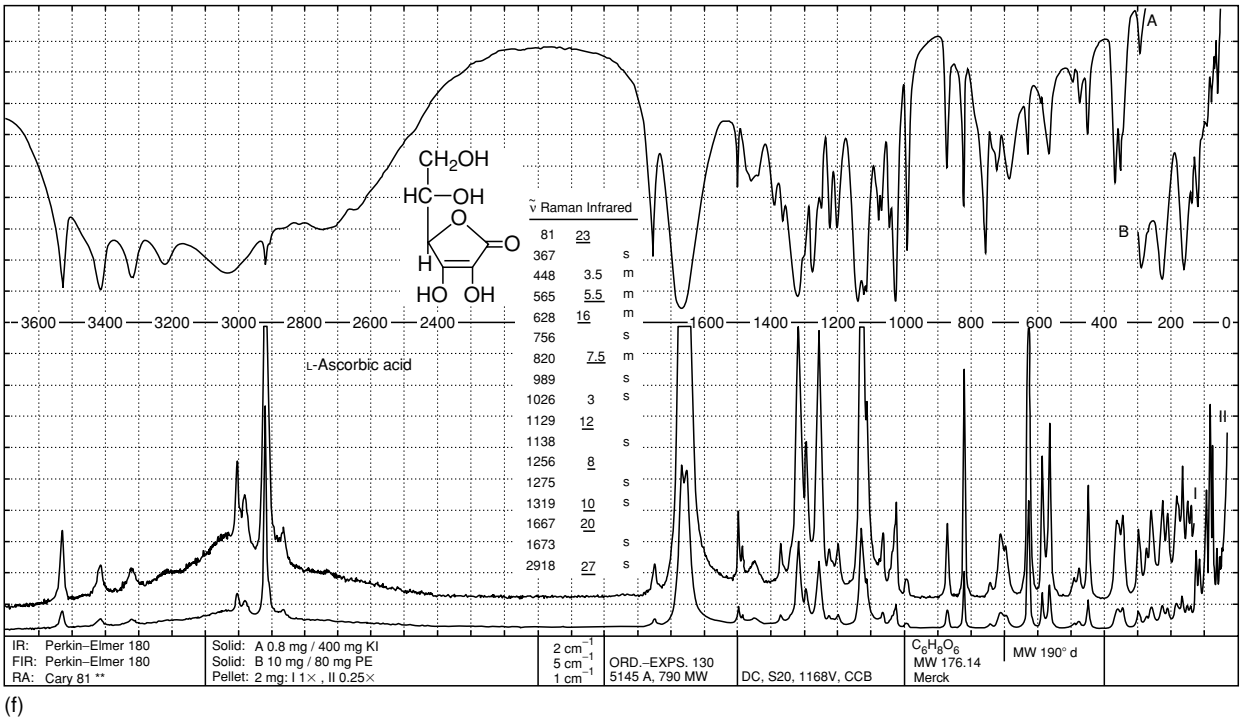

(f)

Continued

Figure 3 Continued

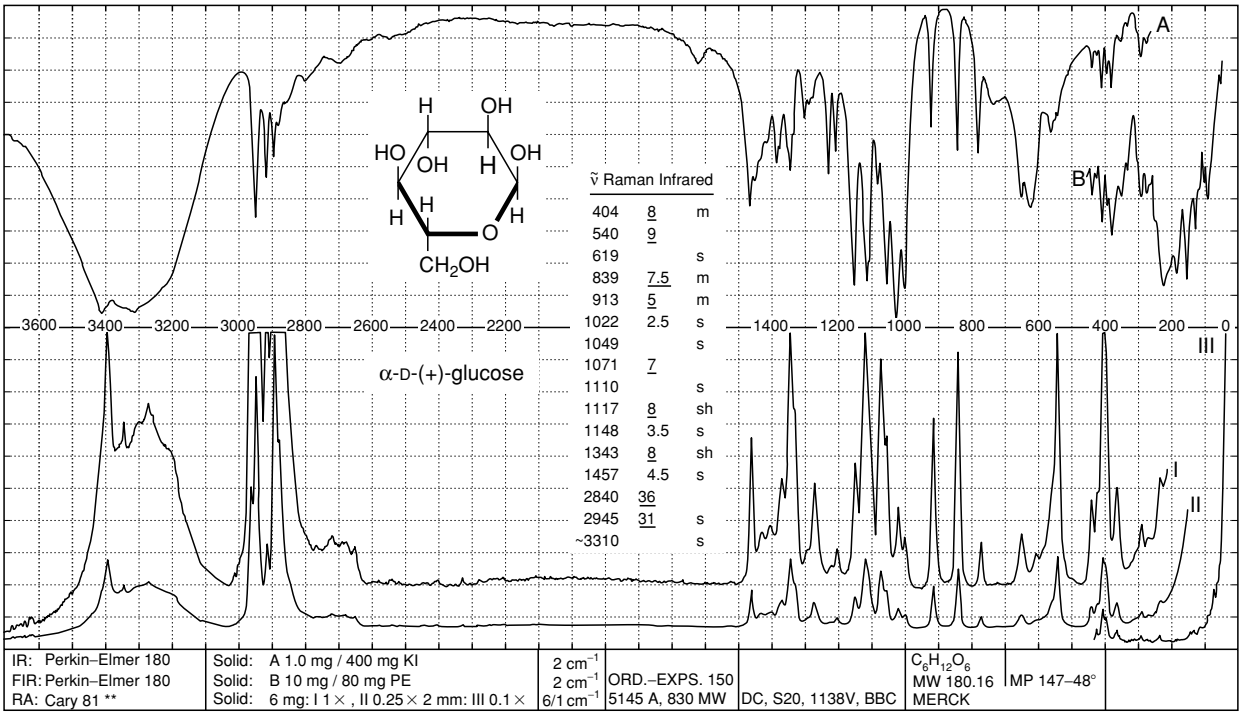

(g)

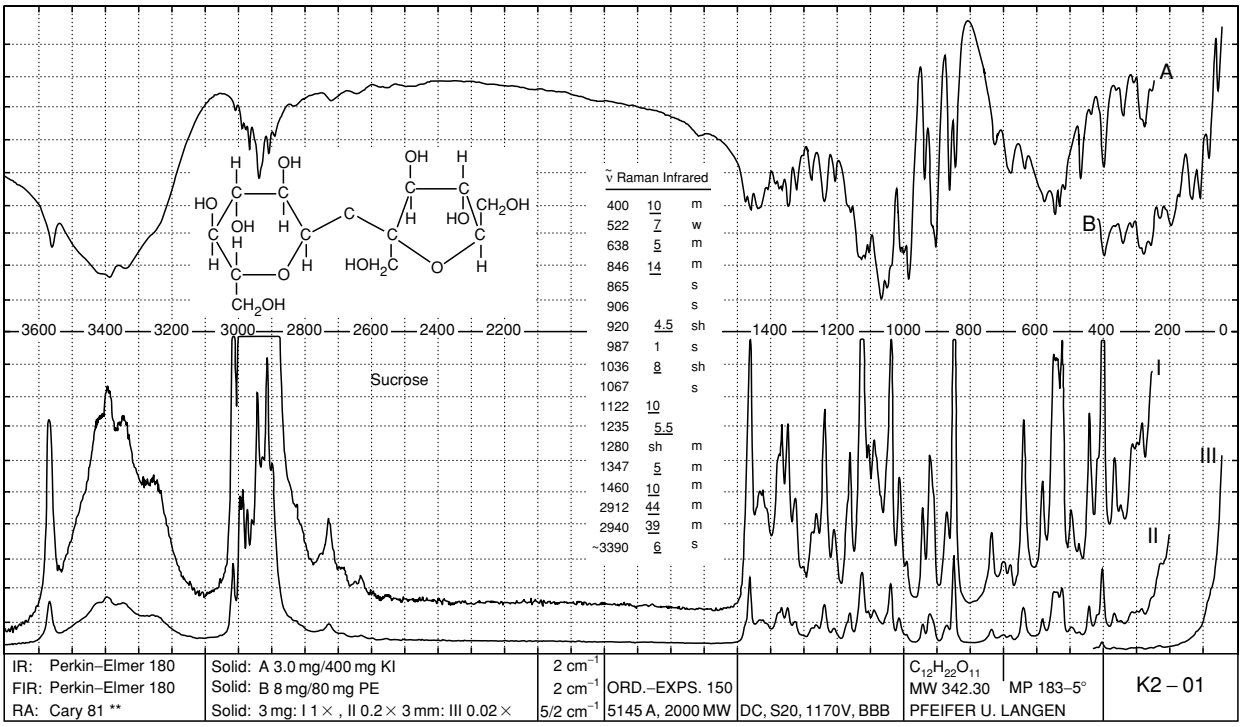

(h)

with single bonds vibrate below 1500 cm^{-1}. Stretching vibrations change bond lengths, and deformation vibrations change bond angles. They occur at approximately half the stretching frequency.

According to quantum mechanics, vibrations are excited by energy quanta, $h\nu_s$, where h is Planck's constant, and ν_s is the vibrational frequency. Radiation quanta in the infrared range, 2.5–1000 μm, equivalent to 4000–10 cm^{-1}, can be absorbed by the molecule and excite it from the vibrational ground state (v=0) to the vibrational excited state (v=1). Thus, absorption bands are produced in the infrared spectrum (Figure 2a).

If the molecules are irradiated with quanta of the near-infrared range, 0.8–2.5 μm, equivalent to 12 500–4000 cm^{-1}, the overtones and combination bands of the infrared spectrum, especially, of the CH, NH, and OH bonds are excited, giving rise to the near-infrared spectrum.

Near infrared

When the molecules are irradiated with quanta of large energy, $h\nu_0$ (by monochromatic, usually laser, radiation from the ultraviolet, visible or near infrared region), a scattering process may occur, by which vibrations with energy $h\nu_s$ are excited and quanta of lower energy, $h\nu_R = h\nu_0 - h\nu_s$, are emitted simultaneously. These constitute the so-called Raman spectrum (Figure 2b). The quantum yield of Raman scattering is quite low (10^{-12}), this means that Raman spectra are of a low intensity. If there are samples with absorption bands that can be excited by the quanta $h\nu_0$, then fluorescence may occur, a strong emission spectrum with broad bands (Figure 2c). Since the quantum yield for this process is nearly 1, Raman spectra can be completely concealed, even when the concentration of the fluorescing impurities is very low. Until 1986, Raman spectroscopy with excitation in the visible region was only rarely applied to food analysis, since the spectra were heavily masked by the fluorescence spectra of some natural compounds: cells can only live with the help of an enzymatic machinery with absorption bands in the visible range of the spectrum – consequently, it shows fluorescence and is photochemically endangered. Raman spectra of most food products therefore cannot be investigated with 'classical' excitation in the visible range. However, when Raman spectra are excited in the near-infrared range, e.g., with 1064 nm, the energy of the quanta, $h\nu_0$, is too small (only about half of the quanta at 488 nm) to excite fluorescence (Figure 2c). Excitation with the Nd:YAG laser at 1064 nm marks an optimum for non-destructive recording of the Raman spectra of living cells and all kinds of food.

The intensities of the vibrations revealed by the infrared spectrum are proportional to the square of the change of the molecular dipole moment, μ, during the vibration, described by the 'normal coordinate,' q. The intensity of the Raman spectrum is proportional to the square of the change of the molecular polarizability, α, during the vibration:

$$\text{Infrared intensity} \sim \left(\frac{\partial\mu}{\partial q}\right)^2.$$

$$\text{Raman intensity} \sim \left(\frac{\partial\alpha}{\partial q}\right)^2.$$

The polarizability characterizes the elastic deformation of the charge distribution of a molecule by an electric field. Therefore, infrared spectra usually show strong bands for vibrations of polar groups. Raman spectra, however, show strong bands for vibrations of nonpolar and multiple bonds as well

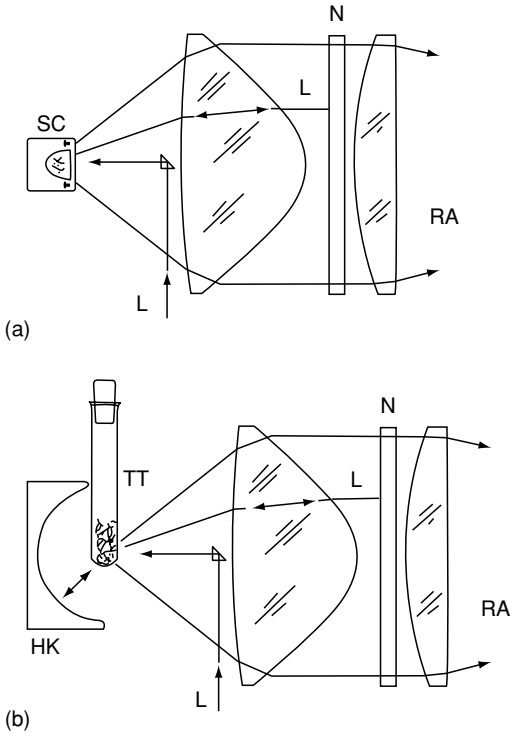

Figure 4 Entrance optics for an NIR-FT-Raman spectrometer: (a) SC sample cup with CaF_2 window and a VITON O-ring, L beam of a Nd:YAG laser; (b) the sample in a small test tube, TT, in the center of a hemispherical mirror HK. The exciting radiation, diffusely scattered by the sample is reflected by the notch filter N; together with HK this multiple reflection arrangement enhances the intensity of the observed Raman spectrum and integrates the spectra of the whole sample.

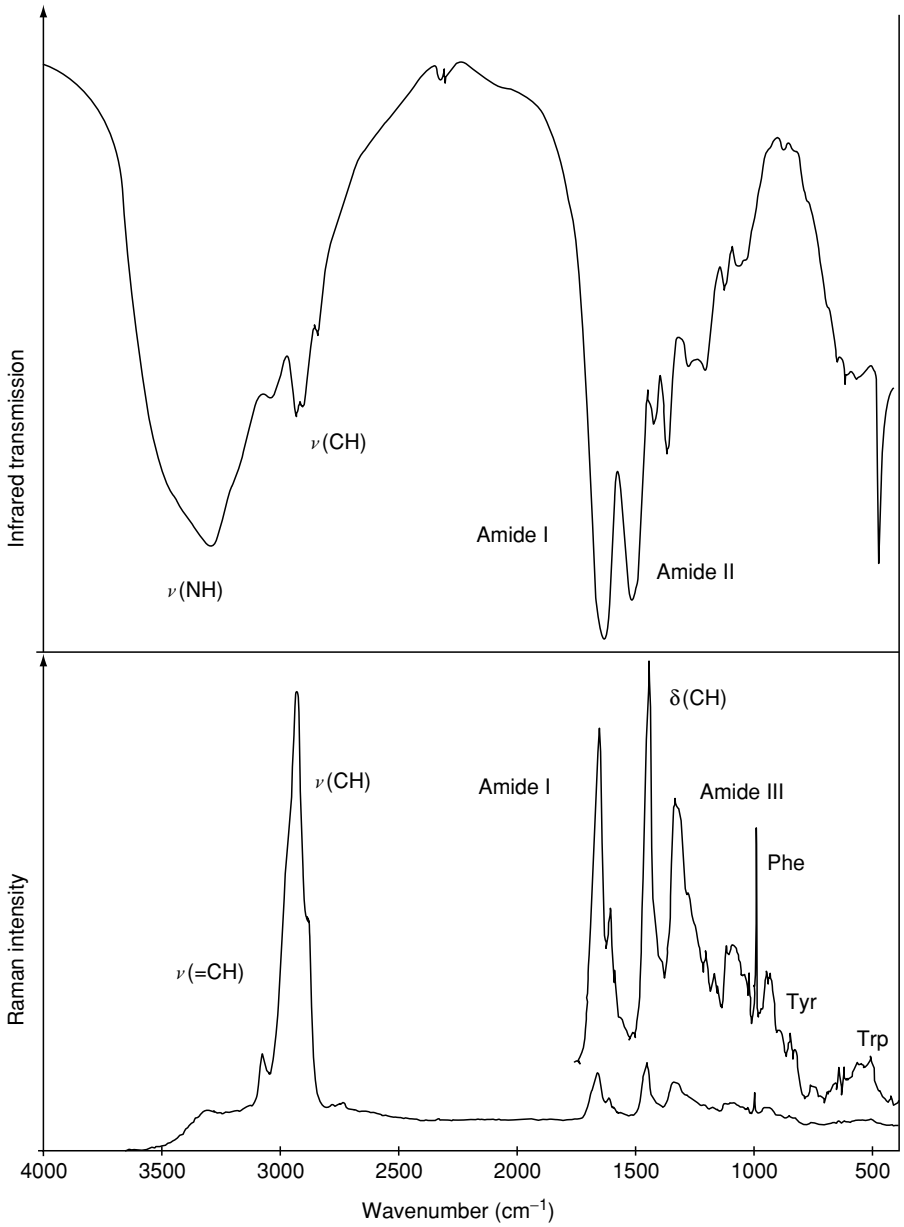

Figure 5 Infrared and Raman spectra of bovine serum albumin. ν, mean stretching vibration; Phe, Tyr and Trp mark typical vibrations of amino acids.

as for symmetric vibrations of collections of similar bonds, as in ring systems.

Typical Features of Infrared and Raman Spectra

In Table 1, typical properties of the methods of vibration spectroscopy are collected. In several aspects, Raman spectroscopy has several advantages compared with infrared spectroscopy: it needs only very simple or no preparation, uses nonproblematic sample containers, and needs only one spectrometric set-up for the whole spectrum. In particular, the intensity of the bands is proportional to the concentration. Unfortunately, Raman spectrometers are at present more expensive than the infrared instruments. The properties of the near-infrared spectrometry are also listed, for comparison.

Table 2 lists the assignment of the bands that can be observed using both methods. Vibrations of polar groups are more intense in the infrared spectrum, whereas multiple bonds and symmetric

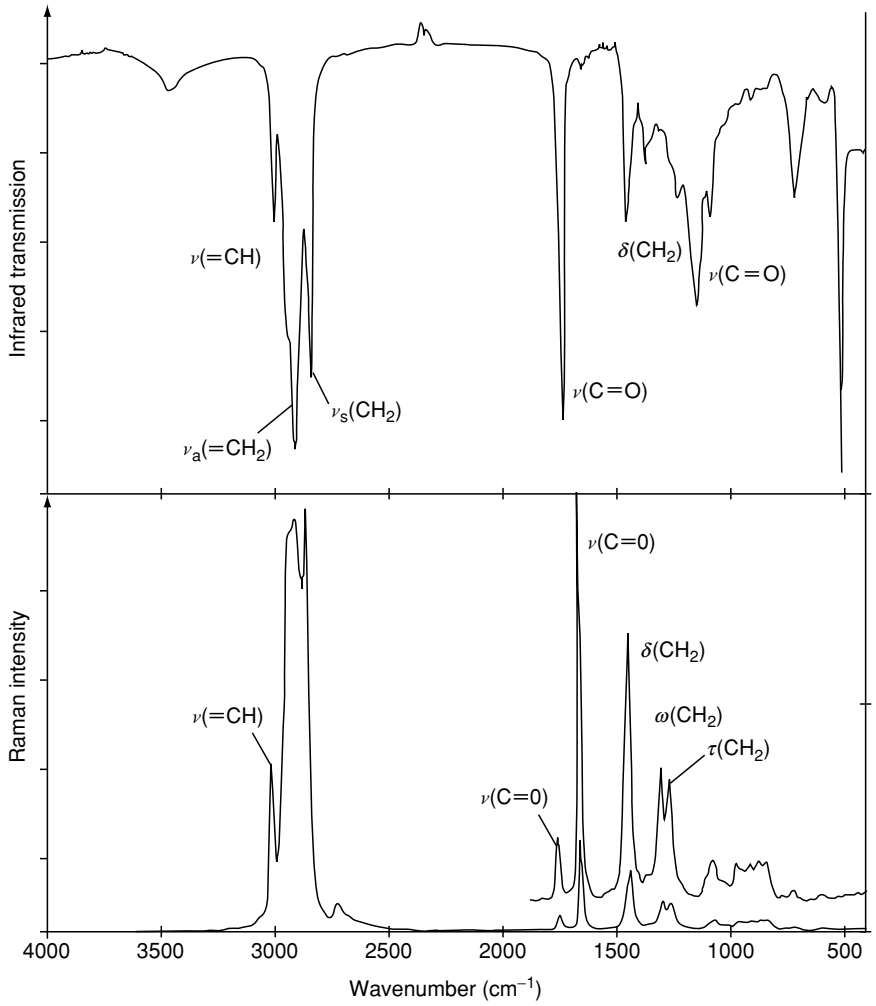

Figure 6 Infrared and Raman spectra of safflower oil. ν, δ, ω and τ mark stretching, deformation, wagging and twisting vibrations. Note the high intensity of the ν (C = C) in the Raman spectrum.

vibrations are usually more intense in the Raman spectrum. Therefore, it is often useful to evaluate both spectra.

Every molecule is a vibrating unit. If it contains n atoms, it has $3n - 6$ 'normal vibrations.' Together, they represent a unique fingerprint, allowing the identification of the molecule. For example, Figure 3 shows the infrared spectra (above) and the Raman spectra (below) of eight important food components.

The spectra of the cystine molecule (Figure 3a) shows many bands with the same frequencies in the infrared and the Raman spectrum. However, the intensities clearly demonstrate the nature of the groups responsible: vibrations of polar groups are strong in the infrared. The bands due to the COOH and NH_2 groups with participation of zwitterionic COO^- and NH_3^+ groups are strong and broad in the infrared spectrum at 3000 and 1600–1400 cm^{-1}, but they are weak in the Raman spectrum. Conversely, the stretching vibrations of the CH_2 group at 2917 and 2969 cm^{-1} and especially the S—S stretching vibration at 499 cm^{-1} are very strong in the Raman spectrum.

In Figure 3b, the spectra of linolenic acid methyl ester are shown. The ester C═O group (1743 cm^{-1}) produces a strong infrared and a weak Raman band; the C═C group (1654 cm^{-1}), however, shows a very strong Raman and a very weak infrared band. These are useful for the determination of the unsaturated fatty acids in fats (see below).

The spectra of retinol, vitamin A (Figure 3c) show the typical vibrations of the conjugated C═C bonds at 1586, 1193, and 1155 cm^{-1}, strong in the Raman and weak in the infrared spectrum.

In Figure 3d and 3e, the spectra of caffeine and vitamins B_1 show many strong and sharp bands, typical for nonflexible ring systems.

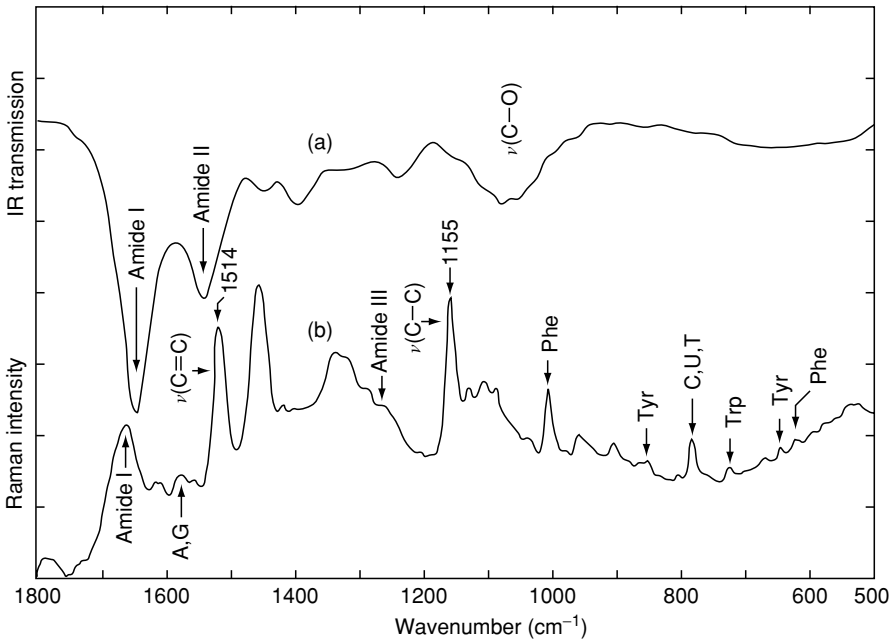

Figure 7 (a) Infrared and (b) Raman spectrum of *Micrococcus roseus*. C, cytidine; U, uridine; T, thymidine; A, adenosine; G, guanosine.

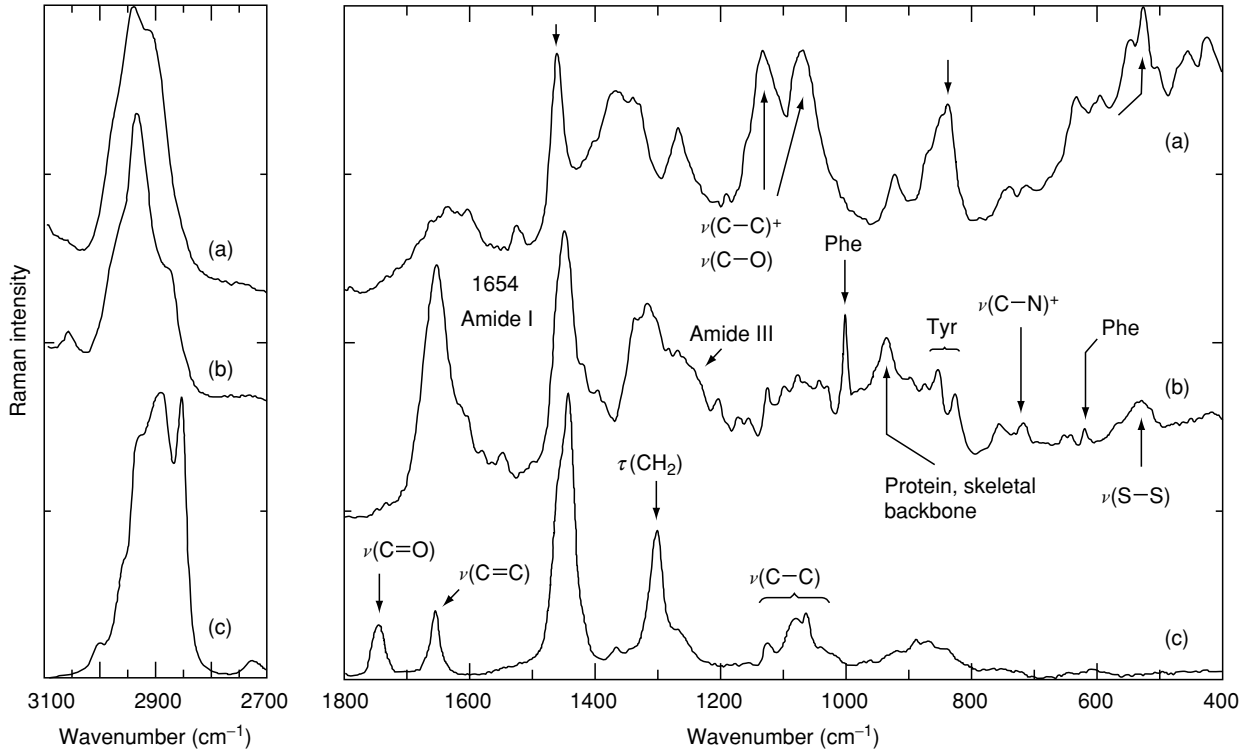

Figure 8 Raman spectra of typical components of food: (a) banana (carbohydrate); (b) turkey breast (protein); (c) butter (lipid).

Figure 3f, g, and h show the spectra of carbohydrates: vitamin C, glucose, and sucrose. All show strong bands in the range 1200–1000 cm^{-1}, due to the C—O groups. Hydrogen-bonded O—H groups show bands around 3400, and 'free' OH bonds above 3500 cm^{-1}.

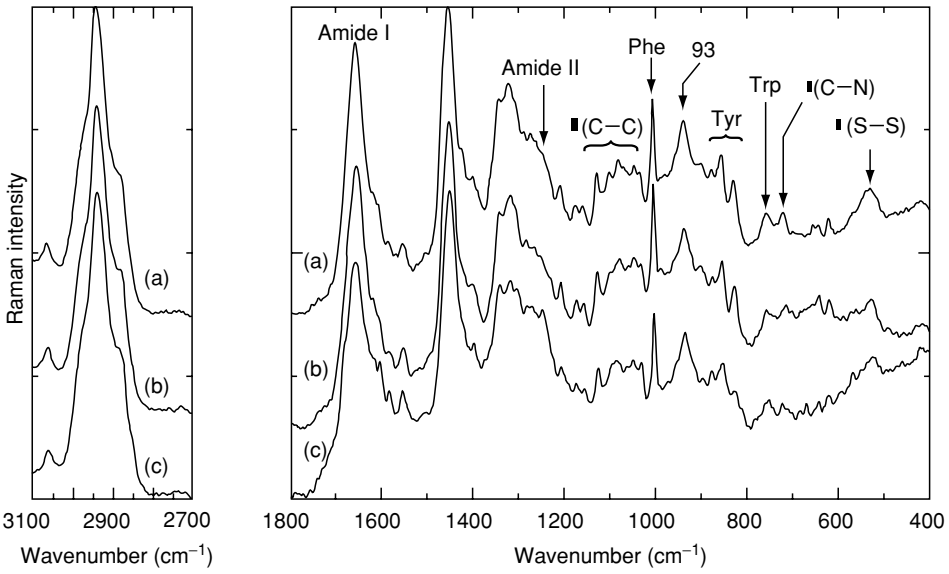

Figure 9 Raman spectra of fresh meat of (a) turkey, (b) pork, and (c) beef.

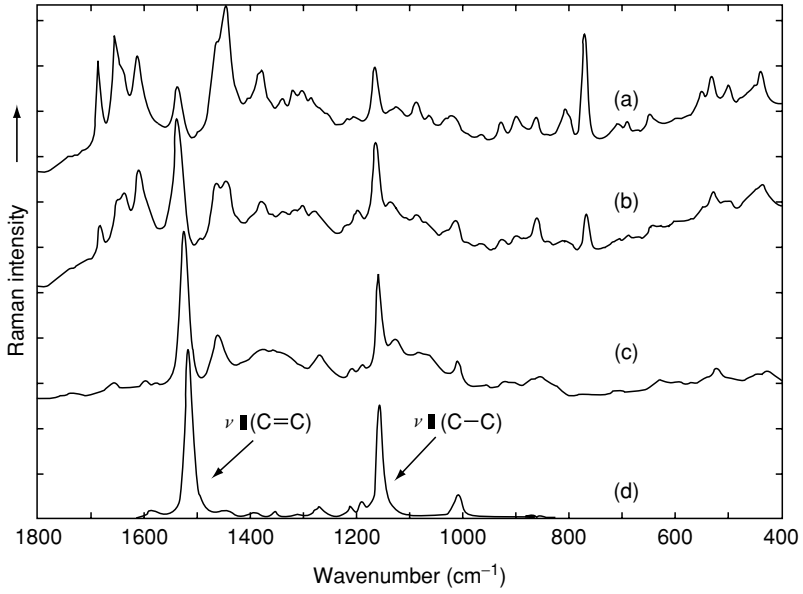

Figure 10 Raman spectra of food components containing carotinoids: (a) peel of untreated orange; (b) peel of orange, treated with wax and fungicides; (c) carrot; (d) β-carotene.

Identification is made easy by comparison of the observed spectra with those in atlases of Raman and infrared spectra of organic and inorganic compounds and by collections of digitized spectra supplied by the instrument manufacturers.

Quantitative Analyses

The basic relation for quantitative analyses by infrared absorption spectroscopy is the Beer–Lambert law:

$$\Phi = \Phi_0\ 10^{-\varepsilon c l} = \Phi_0\ 10^{-A}.$$

Here, Φ_0 is the radiant power illuminating the sample, and Φ is the transmitted power. The transmittance of the sample, $\tau = \Phi/\Phi_0$, is usually recorded, ε is the molar absorption coefficient, c is the concentration and l is the absorbing path length. Most instruments allow quantitative analyses by recording the absorbance, A, which is proportional to the concentration.

The intensity of Raman lines is proportional to the concentration over several orders of magnitude. Non-linearities due to intermolecular interactions can be taken into account by calibration curves determined under conditions similar to those of the practical application. Depending on the circumstances, 'intensities' are measured at the maximum of a band or by integration of the band contour. The analysis of spectra is supported by the use of sophisticated multicomponent analysis, e.g., cluster analysis or neuronal networks.

Special Techniques

Infrared Spectroscopy

For infrared spectroscopy, the common technique uses sample cells with a lamella of a liquid sample of a path length between 5 and 20 μm or a solution of about 100–500 μm, the transmittance of which is measured as a function of the wavenumber. The lamella is supported by infrared-transparent windows, mostly of (hygroscopic(!)) potassium bromide (KBr). However, solid samples need special preparation. They can be mixed with KBr powder and pressed into disks, or they can be ground together with an infrared-transparent oil.

The so-called 'attenuated total reflection' (ATR) or 'frustrated multiple internal reflection' (FMIR) technique is especially useful for food analyses by infrared spectroscopy. An infrared beam coming from a material with a large refractive index is totally reflected at a surface adjacent to a sample having a smaller refractive index. Under the conditions of total reflection the beam emerges into the adjacent medium as 'evanescent wave', thus being partially absorbed by infrared active vibrations. The ATR technique can be used for samples which are difficult to prepare such as meat, cheese, vegetables or dressings. Especially as 'circle cell' this technique is specially designed for continuous analyses of flowing liquids such as alcoholic beverages, syrups or milk products.

Powdered materials can be investigated by the 'diffuse reflection' method. By making use of multiple reflections of the infrared beam by the powder, components in even very low concentrations may be detected. Analysis of samples on thin-layer chromatography (TLC) plates is possible directly after extraction. (*See* **Chromatography**: Thin-layer Chromatography.)

Investigation of volatile components is possible by coupling of an infrared spectrometer to a gas chromatograph either by making use of internally reflecting glass capillaries as sample cells or by condensing the fractions on a cooled infrared-transparent support. Aroma components are determined in this way by headspace analyses. Trace analyses of toxic components in gases, e.g., in human breath, employ cells with mirror systems (White cells) allowing large effective path lengths by multiple pass of the beam. (*See* **Chromatography**: Gas Chromatography.)

Infrared microscopes are now available, which employ the properties of reflecting objectives having no chromatic aberration. The samples can therefore be adjusted in the beam by inspection with visible light, and afterwards, the infrared spectra in either transmission or reflection can be run from exactly the same spot. The spatial resolution of infrared microscopes is limited by the wavelengths of the radiation employed and is of the order of 50 μm.

Raman Spectroscopy

The new FT Raman spectrometers usually observe the spectra of samples with a 180° geometry. This means that the exciting radiation is directed on to the sample perpendicularly to its surface. The Raman radiation emerging from the surface is collected with

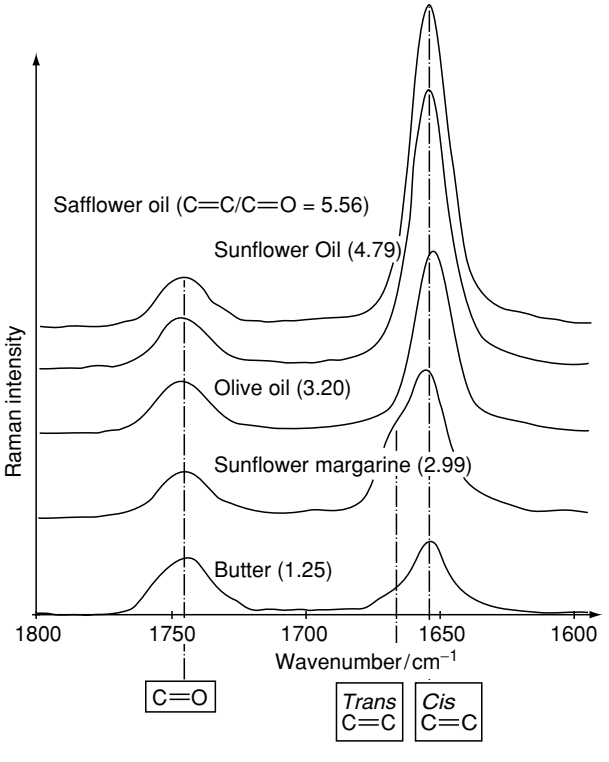

Figure 11 Raman spectra of different fats: (a) safflower oil; (b) sunflower oil; (c) olive oil; (d) sunflower margarine; (e) butter. All spectra are normalized to the intensity of the C=O vibration at 1746 cm^{-1}. The *trans*-C=C is at 1662, and *cis*-C=C at 1653 cm^{-1}. Intensity ratio of *cis*-C=C to C=O for (a) 5.56, (b) 4.79, (c) 3.20, (d) 2.99 and (e) 1.25.

a lens system with a large aperture. The samples are analyzed in cells, allowing multiple reflection (Figure 4). The amount of sample is about 100 mg, but 1 mg of a liquid or solid sample in a melting-point tube is often sufficient; much less is needed when a Raman microscope is applied.

The Raman spectra were usually recorded with the Bruker FT-Raman spectrometer (IFS 66 with FRA 106) and a resolution of 4 cm^{-1}, 350 mW of laser power at 1064 nm with unfocused beam, and a recording time of 15 min.

Fiber-optical systems are especially useful for food analyses by Raman spectroscopy. They connect the spectrometer with a remote sample. Optical fibers may have a very high transmittance in the near-infrared range (about 80% per kilometer), and therefore an analysis of remote process streams of powders or fluids is possible, even of the content of reactors.

A microscope for recording of FT-Raman spectra of small particles or inclusion is commercially available as an accessory to the instruments. It is coupled by fiber optics to the spectrometer, which may be placed at some distance. Raman spectra of tissues, microorganisms and fibers, often complementary to infrared microscopy, can be investigated. One advantage of Raman microscopy is its greater spatial resolution of the order of 5 μm.

Selected Applications

Figure 5 shows the infrared and Raman spectra of bovine serum albumin. The typical vibrations of the amide groups of proteins can be seen in the infrared as well as in the Raman spectra (Table 2). (*See* **Amino Acids**: Determination.)

Figure 6 shows the spectra of safflower oil. As mentioned in the discussion of Figure 3b, the esters of unsaturated fatty acids show complementary intensities of the C═O and the C═C stretching vibrations in the IR and Raman spectra. (*See* **Fatty Acids**: Analysis.)

Microorganisms can be characterized by Raman as well as infrared spectroscopy. Employing multivariate statistical analysis of the second derivative of the spectra, different strains of bacteria can be classified. With FTIR and especially near infrared FT Raman microscopes the identification of bacteria, viruses, fungi, and even mammalian cells is possible (Figure 7).

As already discussed with Table 1 for several applications, Raman spectroscopy is better than infrared spectroscopy. This is partially due to the fact that the water content of food shows a very strong absorption in the infrared range and also, since, for Raman spectroscopy, only minimal sample preparation is necessary. Therefore, the examples discussed below concentrate on NIR-FT-Raman spectra.

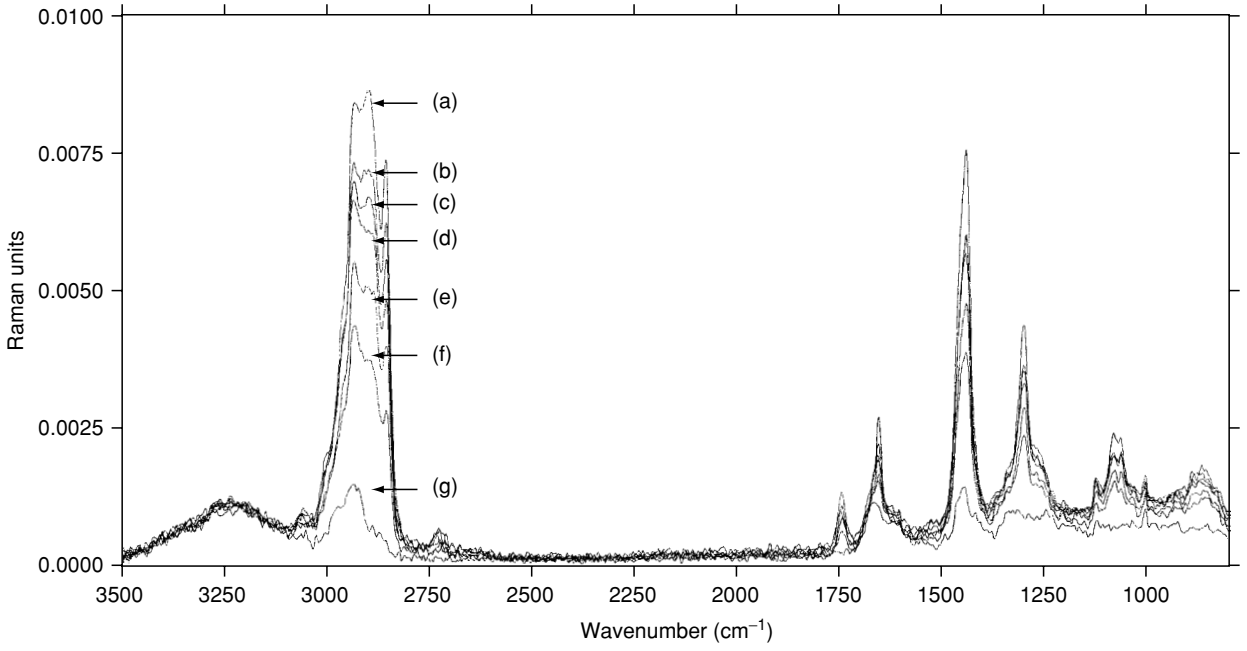

Figure 12 Raman spectra of cheeses with different concentration of fat in solid matter (%): (a) Wörrishofer (60), (b) Pikantje (48), (c) Gouda (48), (d) Edam I (40), (e) Edam II (40), (f) Westlite (30), and (g) Harzer Roller (*c.* 3). Spectra are not processed.

Figure 8 represents typical spectra of important food components: carbohydrates (banana), proteins (turkey breast), and lipids (butter).

Carbohydrates are poor Raman scatterers, revealing vibrations at 2902 and 2941 cm^{-1}, based on their large quantity of methine groups. The spectral range from 1200 to 800 cm^{-1} is dominated by stretching vibrations resulting from the carbon skeleton strongly coupled with C—O bonding.

Protein spectra demonstrate a stretching vibration at 2939 cm^{-1}. The amide I-band, amide III-band, and the vibrations originating from the protein backbone, allow an analysis of the secondary protein structure (see Table 2). The Raman spectra show specific vibrations of amino acids with ring systems: phenylalanine; 1607, 1003, 621 cm^{-1}; tyrosine 1611, 854, 829 cm^{-1}; tryptophan; 1556, 1424, 1356, 1009, 873, 755 cm^{-1}. The very strong S—S vibration in the Raman spectrum between 540 and 520 cm^{-1} in cystine is shown in Figure 3a. Cysteine shows a strong S—H vibration at 2543 cm^{-1}.

Lipids are dominated by strong stretching vibrations of the methylene group at 2850 and 2884 cm^{-1}, indicating a large quantity of coupled methylene groups. Several typical spectral features are: the vibration of the ester group at 1744 cm^{-1}, an intense bending vibration at 1440 cm^{-1}, a twisting vibration at 1299 cm^{-1}, and vibrations of the carbon chain between 1150 and 800 cm^{-1}. If unsaturated fatty acids are present, the spectra show an additional band, a sensitive marker of conjugation, ranging from 1655 cm^{-1} (C═C, cisoid form) to 1670 cm^{-1} (C═C, transoid form) (Figure 9).

Fresh meat can be distinguished by general differences of the spectra (Figure 10).

Carotenes show strong vibrations in the Raman spectrum at about 1150 and 1500 cm^{-1}, due to the C═C and the C—C vibrations of the whole chain. Since the absorption bands of carotenes are not far from the position of the exciting line, the observed Raman spectrum show a 'pre-resonance' enhancement. They allow the detection of carotenes, even in low concentrations in plant and animal cells.

As demonstrated in Figure 11, the Raman spectra of fats, when the spectra are normalized to the intensity of the C═O vibration, show the relative amount of unsaturation. The spectra demonstrate the high number of *cis*-substituted double bonds in safflower and sunflower oil but demonstrate that the fat hardening for preparing sunflower margarine reduces the concentration of the *cis* bonds and enhances the *trans*-substituted double bonds.

Figure 12 shows spectra of cheeses with different concentration of fat in solid matter, recorded using the original samples. The bands due to the fat and protein content reduce their intensity with a lower concentration, whereas the band at 3200 cm^{-1}, due to water, is nearly constant.

Typical noodle spectra reveal the spectral feature of starch (amylose and amylopectin) (Figure 13 and 14). The intense band at 479 cm^{-1} is assigned to a C—O bending mode. The vibrations at 845 $\pm$ 5 cm^{-1}

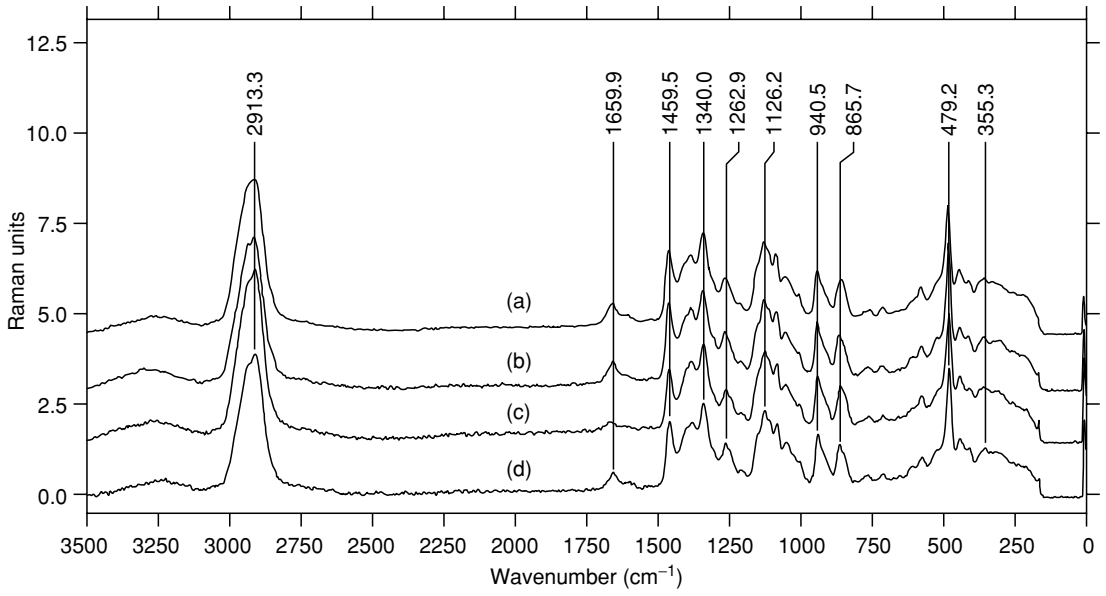

Figure 13 Raman spectra of noodles: (a) Spätzle; (b) Gold-Ei Landnudeln (durum wheat semolina, whole eggs 15%); (c) Bamboo Garden Chinese Rice Noodles; (d) Reine Hartweizen-Schnittnudeln (100 % durum wheat semolina, no egg). All spectra are normalized to the carbohydrate band at 1162 cm^{-1}.

(α-anomers) and at $900 \pm 10\,cm^{-1}$ (β-anomers) are marker bands indicating the type of configuration at the anomeric C_1 position. However, the spectra structure correlations show that the band at $900\,cm^{-1}$ is not attributed exclusively to the β-anomeric configuration. The variation in protein content can be monitored by the band at $1660\,cm^{-1}$ (amide I) in relation to the band at $1459\,cm^{-1}$ (δ(C—H)).

The spectra of noodles are an example of a nondestructive investigation of packaged food. There are only small reductions of the intensity of all bands visible. Food, even in glass containers, can be investigated by Raman spectroscopy.

Finally, Figure 15 gives examples of food containing fat, carbohydrates and protein in different relative concentration with the typical bands already

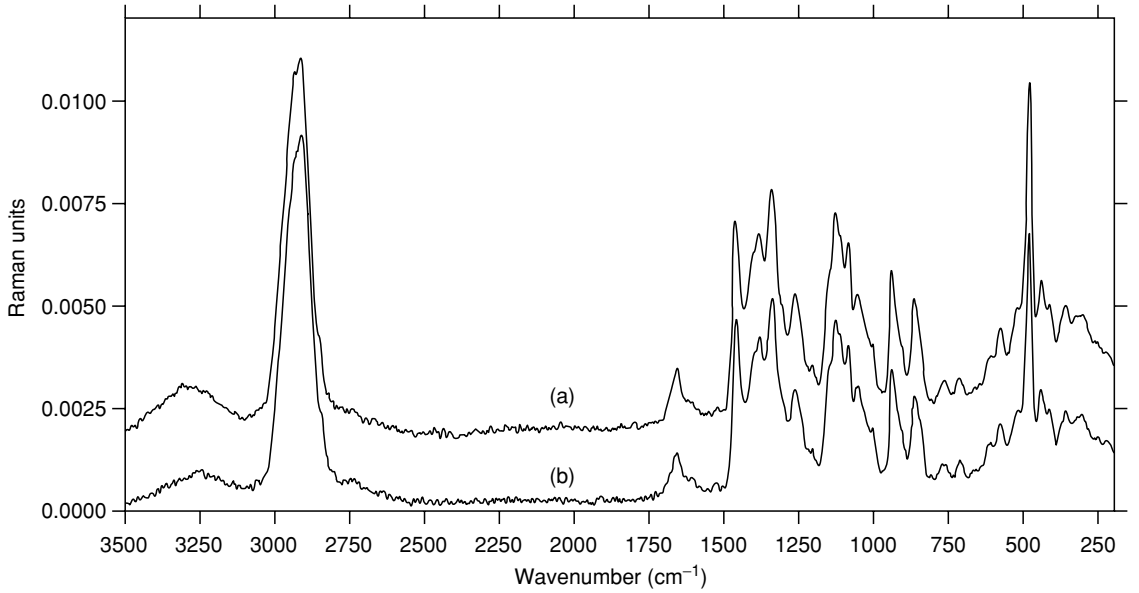

Figure 14 Raman spectra of noodles as in Figure 10b: (a) in test tube; (b) directly through the original plastic bag.

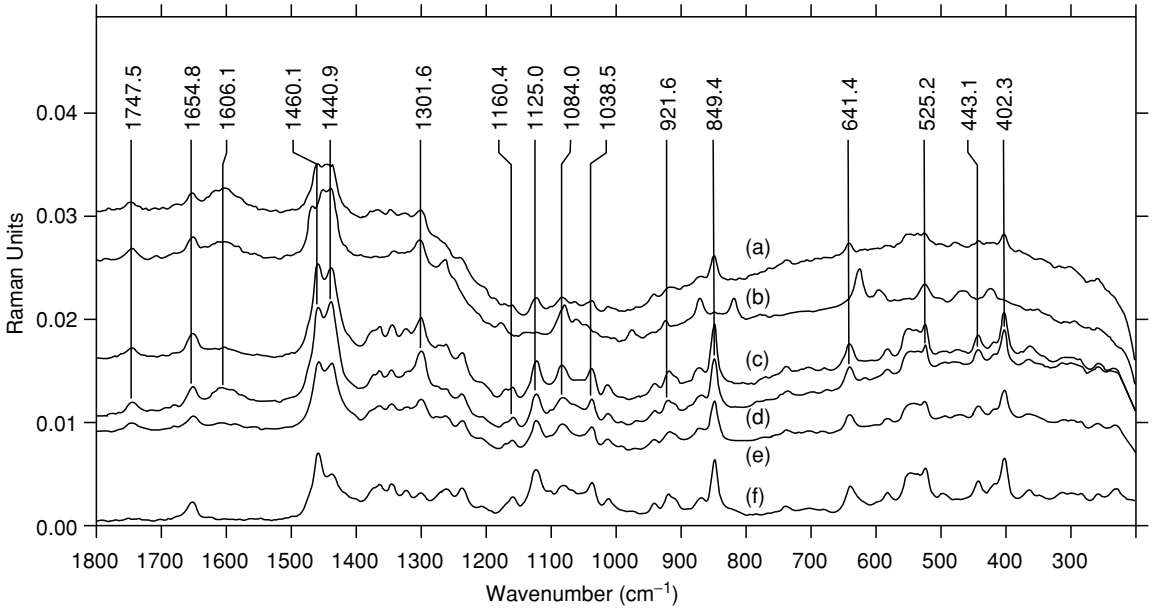

Figure 15 Raman spectra of chocolates and marzipan: (a) Schwarze Herrenschokolade (Chocolate puro extrafino), 60% cocoa; (b) chocolate for diabetics, $\geq$50% cocoa; (c) nougat chocolate, 42% cocoa; (d) superior semibitter chocolate, $\geq$55% cocoa; (e) milk chocolate, $\geq$30% cocoa; (d) marzipan.

discussed. Of special interest is chocolate for diabetics (Figure 15b with the typical bands of sucrose with very low intensity).

See also: **Amino Acids**: Determination; **Chromatography**: Thin-layer Chromatography; Gas Chromatography; **Fatty Acids**: Analysis

Further Reading

Andreev GN, Schrader B, Schulz H, Fuchs R, Popow S and Handjeva N (2001) Non-destructive NIR-FT-Raman analyses in practice. Part I. Analyses of plants and historic textiles. *Fresenius Journal of Analytical Chemistry* 371: 1009–1017.

Cao A, Liquier J and Taillandier E (1995) Infrared and Raman spectroscopy of biomolecules. In: Schrader B (ed.) *Infrared and Raman Spectroscopy*. Weinheim: VCH Verlagsgesellschaft.

Colthup NB, Daly LH and Wiberley SE (1990) *Introduction to Infrared and Raman Spectroscopy*, 3rd edn. Boston, MA: Academic Press.

Keller S, Loechte T, Dippel B and Schrader B (1993) Quality control of food with near-infrared-excited Raman spectroscopy. *Fresenius Journal of Analytical Chemistry* 346: 863.

Kuptsov AH and Zhizhin GN (1998) *Handbook of Fourier Transform Raman and Infrared Spectra of Polymers*. Amsterdam: Elsevier.

Li-Chan ECY (1996) The applications of Raman spectroscopy in food science. *Trends in Food Science and Technology* 7: 361.

Lin-Vien D, Colthup NB, Fateley WG and Grasselli JG (1991) *The Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules*. New York: Academic Press.

Nakanishi K and Solomon PH (1977) *Infrared Absorption Spectroscopy*, 2nd edn. Düsseldorf: Holden-Day.

Sadeghi-Jorabchi H, Wilson RH, Belton PS, Edwards-Webb JD and Coxon DT (1991) Quantitative Analysis of Oils and Fats by Fourier Transform Raman spectroscopy. *Spectrochimica Acta* 47A: 1449.

Schrader B (1989) *Raman/Infrared Atlas of Organic Compounds*, 2nd edn. Weinheim: VCH Verlagsgesellschaft.

Schrader B (ed.) (1995) *Infrared and Raman Spectroscopy. Methods and Applications*. Weinheim: VCH Verlagsgesellschaft.

Schrader B, Dippel B, Fendel S *et al*. (1998) Medical diagnostics with NIR-FT-Raman spectroscopy. In: Mantsch HH and Jackson M (eds) *Infrared Spectroscopy: New Tool in Medicine*. SPIE 3257, 66–71.

Schrader B, Dippel B, Erb I, Keller S, Löchte T, Schulz H, Tatsch E and Wessel S (1999) NIR Raman spectroscopy in Medicine and Biology. *Journal of Molecular Structures* 480–481: 21.

Schrader B, Schulz H, Andreev GN, Klump HH and Sawatzki J (2000) Non-destructive NIR-FT-Raman spectroscopy of plants and animal tissues, of food and works of art. *Talanta* 53: 35–45.

Weitkamp H and Barth R (1976) *Einführung in die quantitative Infrarot-Spektrophotometrie*. Stuttgart: Georg Thieme Verlag.

Near-infrared

R L Wehling, University of Nebraska, Lincoln, NE, USA

Background

Near infrared (NIR) spectroscopy is a technique that is widely used throughout the food-processing industry for both quantitative and qualitative analysis. Typically, NIR spectroscopy measures a food's absorption of electromagnetic energy over a wavelength range of 800–2500 nm (12 500–4000 cm^{-1}). For most foods, this results in spectra with broad, overlapping absorptions that involve C—H, O—H, or N—H bonds. Since at least one of these functional groups is found in many of the major components of foods, information about a wide range of constituents and quality attributes can be extracted from NIR spectra.

Near infrared spectroscopy began in 1800 with an experiment by Herschel. When he used a prism to create a spectrum from white light and placed a thermometer at a point just beyond the red region of the spectrum, he noted an increase in temperature. This was the first observation of the effects of NIR radiation. Between 1920 and 1945, many of the absorption bands observed in the NIR spectra of organic compounds were assigned to specific chemical functional groups. By the 1940s, NIR spectroscopy was being used in the chemical industry for applications such as the determination of water in organic liquids.

Modern NIR spectroscopy applied to foods and agricultural commodities established its foundations in the 1960s and 1970s. Karl Norris and colleagues at the Beltsville, MD laboratory of the United States Department of Agriculture (USDA) developed successful methods for obtaining NIR reflectance spectra from solid biological samples such as leaves, fruit, or cereal grains. In their work developing a rapid method for moisture determination, they also coupled NIR spectroscopy with computer-based multivariate statistical techniques that helped to eliminate interference from absorptions of other constituents (protein, fat, carbohydrate) on the moisture determination. Ultimately, the use of such correlational techniques allowed the measurement of these other constituents, as well.

The first commercial NIR instruments specifically designed for analysis of foods and agricultural

commodities were introduced in the early 1970s. These instruments were dedicated primarily to the determination of moisture, protein, and oil in grains such as wheat, barley, and soybeans. In the intervening years, rapid advances in optics, digital electronics, and computer technology have greatly improved the reliability of the technique, as well as making it possible to analyze a much larger range of constituents in a wide array of products.

Origins of Spectra

A molecule can absorb infrared radiation if it vibrates in such a way that its charge distribution, and therefore its electric dipole moment, changes during the vibration. A vibrating molecule, or molecular functional group, can absorb radiant energy to move from the lowest (ground) state to its first excited state. The frequency of radiation that will make this occur will be identical to the initial frequency of vibration, and is referred to as the fundamental absorption. Typically, fundamental absorptions have frequencies that fall in the midinfrared region of the electromagnetic spectrum ($4000–650\,cm^{-1}$, wavelength = 2.5–15 μm). Molecules can also absorb radiation to move to a higher (second or third) excited state, such that the frequency of the radiation absorbed is two or three times that of the fundamental frequency. Such absorptions are referred to as 'overtones,' and the intensities of these absorptions are much lower than the corresponding fundamental, since these transitions are less favored. Combination bands can also occur if two or more different vibrations interact to give bands with frequencies that are sums of their fundamental frequencies. Combination and overtone bands typically have frequencies that fall in the NIR spectral region.

Both the overtone and combination bands observed in the NIR tend to be weak in intensity. However, this is actually an advantage, since absorption bands that have sufficient intensity to be observed in the NIR region arise primarily from functional groups that have a hydrogen attached to a carbon, nitrogen, or oxygen atom. These are common functional groups in the major constituents of food, including water, proteins, lipids, and carbohydrates. The NIR absorption bands associated with a number of important food constituents are listed in Table 1. The NIR absorption bands of many foods frequently overlap, yielding spectra that are quite complex. This can be illustrated by observing the spectra of the starch and zein (protein) components isolated from corn (maize) shown in Figure 1. The protein fraction exhibits absorption bands centered at about 2050 and 2180 nm, whereas starch has a strong absorption band centered at 2100 nm. Note, however, that there is significant overlap of the starch band into the regions where protein absorbs. Because of such overlaps, quantitative analysis of most foods requires measurements to be made at multiple wavelengths in order to relate the amount of a constituent present to the spectral measurement.

Principles of Diffuse Reflectance Measurements

One of the advantages of NIR spectroscopy is its ability to take measurements directly on solid food

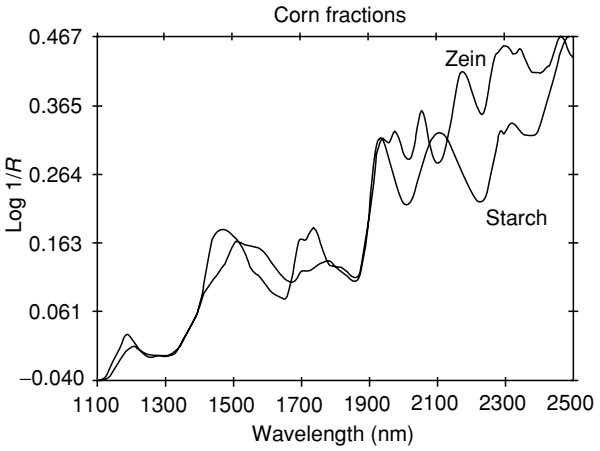

Figure 1 Near infrared reflectance spectra of the starch and zein (protein) components of corn (maize).

Table 1 Absorption bands for important food constituents

Constituent	*Absorber*	*Wavelength (nm)*
Protein, peptides	Amide I, II, and III combinations	2050–2180
	—N—H stretch first overtone	1520–1570
Carbohydrate	C—O, O—H stretching combination	2060–2150
Lipid	Methylene —CH stretch + deformation combination	2300–2350
	$—CH_2$ and $—CH_3$ C—H stretch first overtone	1690–1765
Water	—O—H stretch first overtone	1400–1450
	O—H stretch + O—H deformation combination	1920–1950

samples with little or no sample preparation. This is typically accomplished by measuring the diffusely reflected radiation as a beam of NIR energy is directed on to the sample. When radiation strikes a solid or granular material, part of the radiation is reflected from the sample surface. This mirrorlike reflectance is referred to as specular reflectance, and gives little useful information about the sample. Most of this specularly reflected radiation is directed back toward the energy source. Another portion of the radiation will penetrate through the sample surface and be reflected off several sample particles before it exits the sample. This is referred to as 'diffuse reflectance,' and the diffusely reflected radiation will emerge from the surface of the sample at various angles through 180°. Each time the radiation interacts with a sample particle, the chemical constituents in the sample can absorb a portion of the radiation. Therefore, the diffusely reflected radiation contains information about the composition of the sample, indicated by the amount of energy absorbed at specific wavelengths. It should be noted that the size and shape of the sample particles can affect the amount of radiation that penetrates and escapes from the sample surface.

In NIR spectroscopy, the intensity of the diffuse reflectance from the sample is quantitatively measured at each wavelength of interest. The intensity of radiation reflected from the sample is compared with the intensity reflected from a nonabsorbing reference, such as a ceramic or fluorocarbon material. Reflectance (R) is calculated by the following formula:

$$R = I/I_0,$$

where: I = the intensity of radiation reflected from the sample at a given wavelength and I_0 = the intensity of radiation reflected from the reference at the same wavelength. Reflectance data are expressed most commonly as $\log(1/R)$, an expression analogous to absorbance in transmission spectroscopy. Various mathematical transformations, such as derivatization and scatter correction, can also be applied to the spectra. Such treatments may help to remove spectral differences that can occur due to variations in sample particle size.

Transmission measurements can also be made in the NIR region, and this is often the method of choice for liquid samples. A liquid can be placed in a quartz cuvette, and the absorbance measured at the wavelengths of interest. Transmission measurements can also be taken from solid samples, but generally only in the 800–1100 nm range. In this wavelength region, the absorption bands are higher overtones that are very weak, allowing the radiation to penetrate through several millimeters of a solid sample. The use of transmission measurements can minimize errors that could occur with reflectance measurements taken from samples having heterogeneous surfaces.

Instrumentation

The radiation source in most NIR spectrometers is a tungsten–halogen lamp with a quartz envelope, similar to a projector lamp. Such lamps emit large amounts of energy in both the NIR and visible regions. Wavelength selection in many NIR instruments is accomplished with a grating monochromator. These may be scanning monochromator instruments where a grating is used to disperse the radiation by wavelength, and is then rotated to impinge a single wavelength, or, more appropriately, a narrow band of wavelengths, on to a sample at any given time. With this arrangement, it may take several seconds to collect a spectrum from a sample over the entire NIR region. Some rapid scanning instruments impinge light covering the entire NIR region on to the sample, with the reflected or transmitted light then directed on to a fixed grating that disperses the light by wavelength, and also focuses it on to a multichannel array detector that measures all wavelengths at once. These instruments allow a spectrum to be collected from a sample in less than 1 s.

Instruments dedicated to specific applications often use optical filters to select from 6–20 discrete wavelengths that can be impinged on a sample. The filters are selected to obtain wavelengths that are known to be absorbed by the sample constituents. The instrument inserts filters one at a time into the light beam to direct individual wavelengths (or narrow bands of wavelengths) of radiation on to the sample.

Recently, commercial NIR instruments based on interferometry and using Fourier transform (FT) mathematics have become available. In these FT instruments, the radiation is not dispersed as with a grating monochromator, but rather all wavelengths arrive at the detector simultaneously, and a Fourier transform mathematical treatment is used to convert the results into a typical NIR spectrum. Because all wavelengths are measured simultaneously with an FT-NIR instrument, these instruments can acquire spectra very rapidly with a very high signal-to-noise ratio. Also, such instruments have very high resolution, indicating that they can discriminate differences in reflectance or absorbance occurring between very small wavelength (or wavenumber) intervals. While this may not be much of an advantage for food and agricultural materials whose absorption bands are very broad, it can be important when working with mixtures of pure chemicals that have narrow absorption bands, such as in the pharmaceutical industry. More complete descriptions of

interferometric principles and Fourier transform mathematics can be found in the Further Reading section.

Semiconductor detectors are most commonly used to measure the NIR radiation reflected from or transmitted through a sample. Silicon detectors are used primarily in the 800–1100 nm range, whereas lead sulfide (PbS) detectors are used for wavelengths from 1100 to 2500 nm. Solid-state detectors with a more rapid response, such as indium–gallium arsenide (InGaAs), are being used more frequently in FT and rapid scan dispersive instruments because of their rapid response to changes in light intensity, but they are limited to a maximum wavelength of around 1700 nm.

When taking diffuse reflectance measurements, arrangement of the sample and detectors is extremely important. Most commonly, solid or granular food samples are packed tightly into a cell against a quartz window. Quartz does not absorb in the NIR region, and packing the samples against a window helps to provide a smooth, uniform surface from which reflection can occur. For diffuse reflectance measurements, it is important that the specular reflectance not be measured. In some instruments, this is accomplished by positioning the detectors at a 45° angle with respect to the incoming infrared beam, such that the mirrorlike radiation reflected back toward the infrared source is not measured. Other instruments use an integrating sphere, which is a gold-coated metallic sphere with two holes cut on opposite sides of the sphere, one positioned directly over the sample, and the other through which the radiation beam is introduced. The detectors are mounted inside the sphere. The sphere collects the diffusely reflected radiation coming at various angles from the sample and focuses it on to the detectors. The specular component exits from the sphere through the same port by which the incident radiation beam enters and strikes the sample.

Recent advances in fiber optic technology allow NIR energy to be transmitted through a fiber optic cable some distance from the monochromator or interferometer. This allows either reflectance or transmission measurements to be made remotely from the instrument. Various commercial sampling probes are available that can be inserted directly into bulk granular materials, or can be inserted in a pipe carrying a flowing liquid.

Quantitative and Qualitative Analysis Using NIR

Because of the overlapping nature of NIR absorption bands arising from various food constituents, it is usually necessary to take measurements at two or more wavelengths to quantitate a food component reliably. Most commonly, some type of multivariate regression technique is used to develop mathematical models, or calibrations, that relate NIR spectral measurements to the amount of a particular constituent in a food. The most widely used regression techniques include multiple linear regression (MLR), partial least squares (PLS) regression, and principal-components regression (PCR).

The first step in calibrating an NIR instrument for quantitative analysis is to select a set of calibration, or training, samples. The samples should be representative of the products that will be analyzed, contain the constituent of interest at levels covering the range that is expected to be encountered, and have a relatively uniform distribution of concentrations across that range. The calibration samples are analyzed by the classical analytical method normally used for that constituent, and spectral data are also obtained on each sample using the NIR instrument at all available wavelengths. All data are stored in computer memory.

When multiple linear regression is applied for calibration development, it is used to select the optimum wavelengths for measurement, along with an intercept and associated coefficients for each wavelength, such that an equation with the following form is developed:

$$\%\ \text{constituent} = z + a\log(1/R_{\lambda 1}) + b\log(1/R_{\lambda 2}) + c\log(1/R_{\lambda 3}) + \ldots$$

Derivatized or other transformed spectral measurements can be used in lieu of the raw log $(1/R)$ values. Wavelengths are selected based on statistical significance by using a step forward or reverse stepwise regression procedure, or by using a computer algorithm that tests regressions using all possible combinations of two, three, or four wavelengths to determine the combination that provides the best results. Most calibrations will use between two and six wavelengths, and one should always check to make sure that the wavelengths chosen on the basis of statistical significance also make sense from a spectroscopic standpoint. The use of too many wavelengths can result in 'overfitting' of the data, decreasing the predictive capability of the technique. Calibration results are evaluated by comparing the multiple correlation coefficients, Fs of regression, and standard errors for the various equations developed. It is desirable to maximize the correlation coefficient (generally, R should be > 0.9) and minimize the standard error.

PLS and PCR regression techniques can make use of the entire NIR spectrum, rather than a few selected wavelengths, to predict sample composition. PLS and PCR both use data reduction techniques to extract

from a large number of variables (i.e., reflectance or absorbance measurements at many wavelengths) a much smaller number of new variables that account for most of the variability in the samples. These new variables then can be used to develop a regression equation to predict the amount of a constituent in samples of a food. When using PLS and PCR methods, it is not necessary to eliminate spectral information, as it is when measurements at only a limited number of wavelengths are used. PLS and PCR methods have been found to yield improved results for some analyses when compared with the MLR technique.

Artificial neural networks (ANN) have also recently been used to predict composition from NIR spectra. Neural networks may have some advantages over the linear regression techniques for dealing with highly complex samples, samples from diverse geographic regions, or samples where the relationships between composition and spectral properties do not change in a linear manner.

No matter which calibration technique is used, the models developed should always be tested by using the instrument to predict the composition of a set of test samples that are independent of the calibration set, and comparing the results obtained with those from the classical analytical method.

NIR spectroscopy can also be used for qualitative analysis, classifying a sample into one of two or more groups, rather than providing quantitative measurements. Discriminant analysis techniques can be used to compare the NIR spectrum of an unknown sample with the spectra of samples from different groups. The unknown sample is then classified into the group to which its spectrum is most similar. This technique has been used most widely in the chemical and pharmaceutical industries for raw material identification, but it is beginning to be used more widely for food applications.

Advantages and Limitations of NIR Spectroscopy

NIR spectroscopy provides a number of advantages over traditional wet chemical methods of analysis. Foremost, once an instrument has been calibrated, samples can be analyzed very rapidly, usually in less than 1 min. Also, sample preparation is minimal, and in some cases unnecessary. The technique requires no chemical reagents, so no hazardous wastes are generated. The need for weighing of samples is eliminated as well, thereby eliminating a traditional laboratory bottleneck.

NIR instruments are available that can be used in a production plant environment either in an at-line mode or as on-line monitors. Again, once an instrument has been calibrated, it can be used routinely by production personnel without extensive training. The ability of this technology to provide real-time process control has tremendous potential.

Limitations of the technology include the need for a specific calibration for each product to be analyzed. This is generally not a problem in a production environment where only a few products are likely to be manufactured. However, this limits the usefulness of the technique for situations such as contract laboratories, where a few of many different types of samples may need to be tested each day. Also, early instruments required frequent recalibration. For example, when measuring protein in wheat, it was usually necessary to recalibrate an instrument when grain from a new crop year arrived. However, the use of large calibration sets containing samples from wide geographic areas and time frames, more stable instruments, and more robust calibration techniques such as PLS has made the need for frequent recalibration much less. In fact, recent results have shown that surprisingly diverse sample types can be analyzed with a single calibration model. Research conducted by Kays *et al*. at the USDA Russell Research Center in Athens, GA found that dietary fiber could be measured successfully in products as diverse as snack crackers, granola bars, sugar-coated cornflakes and cookies, all with a single calibration.

Another limitation for users with low sample throughput is the capital cost of commercial NIR instruments. Instrument cost can vary widely, depending on the level of sophistication and automation needed. However, capital expenditures can often be recovered rapidly through labor and reagent

Table 2 Applications of NIR spectroscopy in food analysis

Product	*Constituents measured*
Cereal grains and flours	Moisture, protein, starch, oil, dietary fiber
Oilseeds, flours, meals	Moisture, oil, protein, fiber
Bread	Moisture, protein, fat, fiber
Cereal breakfast foods	Moisture, dietary fiber, sugar
Pasta	Moisture
Cheese	Moisture, fat, protein
Butter	Moisture, fat
Milk and whey powders	Moisture, fat, protein, lactose
Red meat, poultry, fish	Moisture, fat, protein
Processed meats	Moisture, fat, protein
Dehydrated eggs	Moisture, fat, protein
Fresh fruits and vegetables	Total sugar, soluble solids
Dehydrated fruits	Moisture, total sugar, fiber
Potato chips	Moisture, fat
Beer	Alcohol
Corn sweeteners	Sugar

savings, and by minimizing the production of out-of-specification product.

Applications

Applications of NIR spectroscopy to food analysis have greatly expanded since its first use for measuring protein and moisture in wheat and barley. An in-depth discussion of specific applications is beyond the scope of this chapter; however, a number of representative applications are listed in Table 2. A number of NIR procedures have been adopted as approved methods of analysis by the American Association of Cereal Chemists. Typically, if a substance or constituent absorbs in the NIR region, and is present at a level of a few tenths of a per cent or greater, it has potential for being measured by this technique.

In addition to performing compositional analyses of foods and ingredients, NIR has also been used to measure specific chemical constituents in foods that affect their end-use qualities, for directly predicting processing characteristics of commodities that are related to chemical composition, and for monitoring changes that occur during processing. Examples of such applications include determining the amylose content in rice, an important determinant of rice starch quality; predicting corn-processing quality for dry and wet milling; estimating the baking quality of hard wheat; monitoring the degree of cook during extrusion processing; and monitoring the coagulation of milk during cheese-making. In some of these cases, correlations between NIR measurements and measurements from the conventional technique may be < 0.9, but the NIR technique is still good enough for purposes of screening out very good and very poor materials.

Examples of qualitative applications of NIR spectroscopy include the classification of milk powders based on the level of heat treatment, the classification of wheat as hard red spring or hard red winter, the detection of internally infesting insects inside kernels of stored wheat, and the identification of orange juice samples from different sources.

See also: **Spectroscopy**: Overview

Further Reading

AACC (2000) *Approved Methods of Analysis*, 10th edn. St. Paul, MN: American Association of Cereal Chemists.

Burns DA and Ciurczak EW (eds) (2001) *Handbook of Near-Infrared Analysis*. New York: Marcel Dekker.

Griffiths PR and De Haseth JA (1986) *Fourier Transform Infrared Spectrometry*. New York: John Wiley.

Kays SE, Windham WR and Barton FE (1998) Prediction of total dietary fiber by near-infrared reflectance spectroscopy in high-fat- and high-sugar-containing cereal products. *Journal of Agricultural and Food Chemistry* 46: 854–861.

Kramer R (1998) *Chemometric Techniques for Quantitative Analysis*. New York: Marcel Dekker.

Mark H (1991) *Principles and Practice of Spectroscopic Calibration*. New York: John Wiley.

Martens H and Naes T (1989) *Multivariate Calibration*. New York: John Wiley.

Osborne BG, Fearn T and Hindle PH (1993) *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*. Harlow, UK: Longman.

Wehling RL (1998) Infrared spectroscopy. In: Nielsen SS (ed.) *Food Analysis*, 2nd edn. Gaithersburg, MD: Aspen.

Williams PC and Norris KH (2001) *Near Infrared Technologies in the Agricultural and Food Industries*, 2nd edn. St. Paul, MN: American Association of Cereal Chemists.

Fluorescence

L Munck, Royal Veterinary and Agricultural University, Frederiksberg, Denmark

Introduction

Fluorescence (the name originates from the fluorescent mineral fluorspar) refers to cold light emission (luminescence) by electron transfer in the singlet state when molecules are excited by photons. The afterglow lasts less than 10^{-9} s and it is independent of temperature, in contrast to phosphorescence, which is light from the triplet state defined as an afterglow lasting longer than 10^{-6} s, and which is temperature-dependent.

Thus, the fluorescence phenomenon is a form of energy transfer mediated in each case by a specific electronic arrangement, which upon absorbing excitation light of a specific wavelength transforms this energy to another rather specific emission peak, always at a longer wavelength. With suitable filters and mirrors in the fluorescence microscope, fluorescent components are visualized against a black background after filtering out the shorter-wavelength light used for excitation. Similarly, the grating monochromator of a spectrofluorometer enables a sample to be illuminated with light of a specific excitation wavelength. The light emitted at 90° to the incident light passes through another grating that allows only the specific emitted light of interest to pass through to reach the photodetector. Just as in the microscope, absence of fluorescence gives a black background, so in the spectrofluorometer the absence of fluorescence

gives total darkness, i.e., no emitted light. This leads to one of the main problems in fluorescence analysis – how to calibrate against zero light and a zero-signal baseline. This, together with differences in equipment design, confers individual characteristics on spectrofluorometers to a much greater extent than with spectrophotometers. Spectrofluorometers are up to 1000 times more sensitive than most spectrophotometers, but while this may confer considerable advantages, it may also cause additional problems, e.g., with fluorescent impurities. Moreover, the fluorescence phenomenon is sensitive to environmental conditions, such as temperature, viscosity, pH, and interacting substances, which can diminish the intensity of fluorescence by quenching.

Due to the unique structure of the fluorescence emission landscape from a sample, new data techniques may deconvolute the excitation and emission spectra of the pure underlying fluorophores as a form of mathematical chromatography. With chemometric methods, whole fluorescence spectra may be calibrated, e.g., to predict oxidation in vegetable oils, thus complementing current predictive near infrared and infrared spectroscopic screening methods with methods which are both more sensitive and specific.

Fluorescence analysis can be used to follow separation processes of plant and animal chemical components and tissues in food production because of naturally present fluorescence markers (autofluorescence or primary fluorescence). Thus, through primary fluorescence, and without using chemical reagents, we can study the refinement process of sugar, the health of chloroplasts in plants, respiration enzymes in microorganisms and fish meat, and one state of lipid oxidation in meat. The value of fluorescence techniques resides in the fact that specific changes in food components during processing can be followed in the fluorescence microscope and then these changes can be quantified using homogenized samples, and even with a turbid suspension, in a spectrofluorometer. Specific fluorochromes, prepared by organic synthesis or tagged on to specific immunoglobulins, can also be added with different affinities for chemical structures and enzymes (Table 1) and thus biochemical changes in foods can be followed. These are termed secondary fluorescence techniques. Here again we can combine the structural visualization in the fluorescence microscope with direct spectroscopic measurements of the fluorescence intensity.

Physics and Chemistry

In general, the vast majority of intensely fluorescent organic compounds (fluorophors) are aromatic in nature, such as benzene and naphthalene. Some nonaromatic but highly conjugated compounds, such as β-carotene and vitamin A, are also fluorescent. In exploiting secondary fluorescence, nonfluorescent derivatives such as fluorescein dibutyrate (FDB: Figure 1) and calcofluor white Mr2 new (Table 1) are used, which can be rendered strongly fluorescent by enzymic modification or by binding, respectively. (*See* **Carotenoids**: Occurrence, Properties, and Determination; **Polycyclic Aromatic Hydrocarbons**; **Retinol**: Properties and Determination.)

Any fluorescent molecule has two characteristic spectra (Figure 2): the excitation spectrum (the

Table 1 Examples of the use of fluorochromes

Compound	*Excitation wavelength (nm)*	*Emission wavelength (nm)*	*Application*
Acridine orange	470	530–650	Starch, nuclei, pectin, microorganisms, DNA, RNA, acidic nuclear proteins
Aniline blue	420–490	520	Cell wall, (1,3)-β-D glucans, callose
Auramine	400–440	470	Bacteria, cutin
Benzoylarginyl-β-naphthylamide	338	410	Protease activity (trypsin)
Calcofluor white CFW	350	435	Cellulose, chitin, collagen, elastine creatine, bacteria, fungal mycelium/spores
Calcofluor white Mr2 new	360	415	β-D-Glucans in cell walls from, e.g., barley, oats, and wheat
Dansyl fluoride	495	525	Primary amines, A-T bases of DNA
Fluorescein dibutyrate	495	525	Esterase/lipase activity
Fluorescein isothiocyanate (FITC)	490	520	Used as a marker of immunoproteins; in yeast viability test and in cell culture
Phenylalanyl-β-naphthylamide peptides	335	410	Amino peptidase assay
Resorufin butyrate	540	560	Cellulase

The conditions under which the fluorochromes are used largely determine their specificity and fluorescence properties. See Guilbault GG (ed.) (1990) *Practical Fluorescence – Theory, Methods and Techniques*. New York: Marcel Dekker and Francisco A (1989). In Munck L (ed.) *Fluorescence Analysis in Foods*, pp. 266–276. Harlow: Longman.

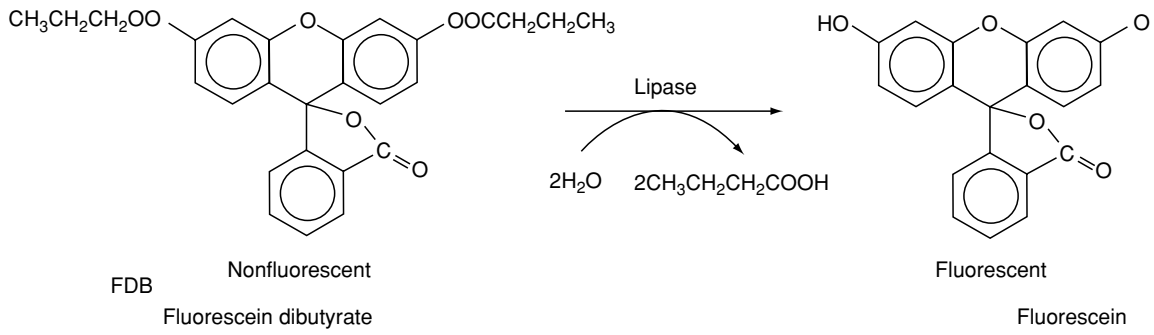

Figure 1 Use of secondary fluorescence. Fluorescein dibutyrate used to detect lipase (esterase) activity by enzymic cleavage to the yellow fluorescent substance fluorescein (excitation 495 nm, emission 525 nm). Reproduced from Spectroscopy: Fluorescence, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

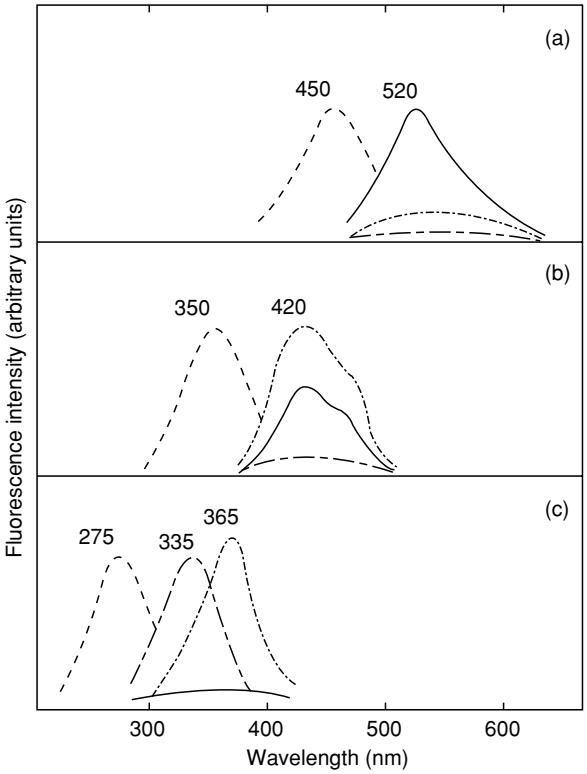

Figure 2 Excitation and emission spectra in the detection of the botanical components of wheat by their primary fluorescence. Excitation (- - -) for (a) pericarp; (b) aleurone; (c) endosperm. Emission spectra of manually dissected pericarp (——), aleurone (–·–), and endosperm (- –) obtained at maximum excitation wavelength for these components: (a) 450 nm; (b) 350 nm; (c) 275 nm. The primary fluorescence probably originates from carotenoids (a), ferulic acid (b), and tryptophan (c). Measurements were made on suspensions of ball-milled flour in glycerin. Reproduced from Jensen SAA, Munck L and Martens H (1982) The botanical constituents of wheat and wheat milling fractions. I. Qualification by fluorescence. *Cereal Chemistry* 59: 477–484, with permission.

relative efficiency of different wavelengths in causing fluorescence) and the emission spectrum (the relative intensity of radiation which is emitted at various wavelengths at a given, e.g., maximal, excitation wavelength). Theoretically, the shape of the excitation spectrum of a fluorescent substance, discounting instrumental artifacts, is identical with that of the absorption spectrum as measured by spectrophotometry. The longest-wavelength peak should be chosen for excitation of the sample, thereby minimizing photodecomposition of the fluorophore by ultraviolet (UV) light. Rayleigh and Tyndal scattering can be observed in the emission spectrum at the same wavelength as the excitation wavelength, and also at twice this value (second-order grating effect) in a grating spectrofluorometer.

The basic equation defining the relationship of fluorescence of a substance to its concentration is:

$$F = \Phi I_0(1 - e^{-\varepsilon bc})$$

For very dilute solutions, this equation is reduced to Beer's law in spectrophotometry:

$$F = K\Phi I_0 \varepsilon bc$$

where Φ is the quantum efficiency of yield (number of quanta emitted divided by number or quanta absorbed), I_0 is the incident radiant power, ε is the molar absorptivity, b is the path length of the measurement cell, and c is the molar concentration of the fluorophore. K designates an arbitrary constant with no unit.

The sensitivity of fluorescence is large because a signal over a zero background is measured. Its specificity is high because, although all compounds absorb radiation, only a few emit, and those that do can be differentiated on the basis that they very seldom have the same excitation and emission maxima. The main disadvantage of fluorescence is that the output is strongly dependent on temperature although the afterglow time is independent. It is also dependent on pH, ionic strength, and viscosity, making it necessary to standardize analysis conditions rigorously. Fluorescence is affected by quenching, which reduces

fluorescence by a competing deactivation process that might be enhanced by oxygen, impurities (e.g., dichromate and iodide), high temperature and high concentrations of the fluorophore.

Concentration quenching may be substantial in transmission mode (inner-cell effect), whereas it is negligible in front-face (reflectance) fluorescence with solid samples.

Instrumentation and Sample Preparation

Light Sources

Spectra for two conventional light sources for microscopy (HBO, mercury lamp; XBO, xenon lamp) and spectrofluorometry (XBO) are presented in Figure 3. Xenon lamps for spectrofluorometry need a certain burning time in order to obtain a constant spectral output, and they have a limited lifetime of 200–1000 h. When working with the information from complete spectra, lamp spectral stability is crucial, and spectral instability cannot be totally compensated for across complete spectra by using the rhodamine standard in a ratio spectrofluorometer, as explained below. Recently, monochromatic laser techniques have increased light intensity and wavelength precision, but at high cost. The aim of increased light intensity is often counteracted by photodecomposition of the fluorophore which necessitates short illumination periods.

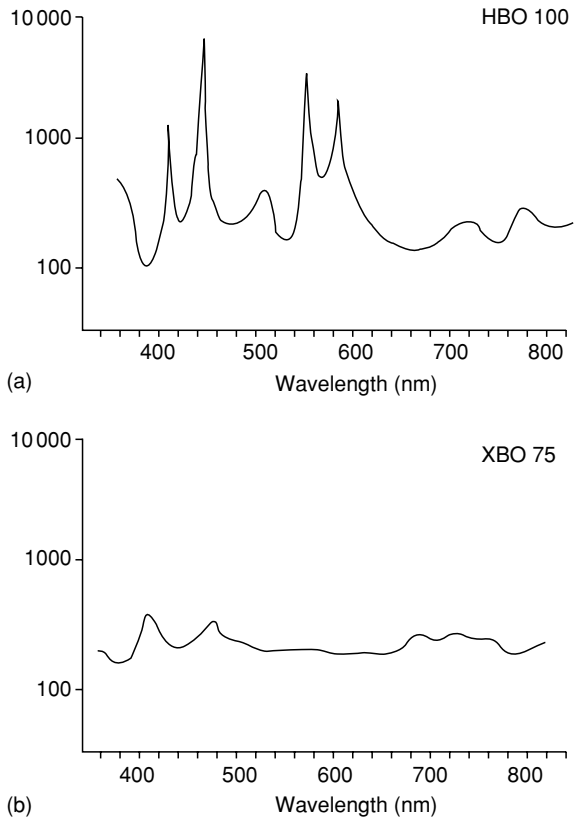

Figure 3 Light sources. Examples of spectral outputs typical of common illuminators in fluorescence microscopy and spectrofluorometry. (a) Mercury lamps (e.g., HBO series), mostly used in microscopy, generate very intense peaks at certain wavelengths (e.g., 365 nm (not shown), 450, 435, and 546 nm), while (b) xenon (e.g., XBO series) lamps produce a more uniform spectrum at lower intensity, and these are used in both fluorescence microscopy and spectrofluorometry. Reproduced from Fulcher RG, Irwing DW and Francisco A de (1989) Fluorescence microscopy: Applications in food analysis. In: Munck L (ed.) *Fluorescence Analysis in Foods*, pp. 59–109. Harlow: Longman.

Fluorescence Microscopy

Ideally a minimum of preparation should be used so that the material can be examined in as near a native state as possible. A suitable technique is cryosectioning at freezing temperatures (−20 °C) with and without additives to prevent ice crystal formation. Simpler, but also effective, is hand-sectioning with a razor blade at room temperature. The sections or macerated material are transferred to a microscope slide, stained, and covered with a cover slip. Permanent preparations need fixation of samples, e.g., with glutaraldehyde, followed by embedding in a resin such as glycol methacrylate (GMA), which is especially suitable for fluorescence, including immunofluorescence; sections are then cut with a glass-knife microtome and stained appropriately. GMA is easy to handle at room temperature and need not (and cannot) be removed from tissues prior to staining. Tissue sections do not require hydration or dehydration prior to staining. GMA fluoresces minimally; the polymerized block is much harder than paraffin and it can thus be sectioned more thinly.

The most frequently used fluorescence microscopy technique is epiillumination (reflectance), which is most light-efficient owing to its specific filter/optics arrangement (Figure 4), but the older transmission mode (diaillumination) can also be used. There is an inverse relationship between the number and volume of glass pieces in an objective lens and its usefulness in fluorescence techniques because a more sophisticated optical construction compromises effective light transmission. It is difficult to find light-efficient, low-magnification objectives for fluorescence work because of their low numerical aperture. Modern sensitive films and video cameras can compensate for the limiting factor of low light intensity in fluorescence microscopy.

Successful fluorescence microscopy is critically dependent on the use of filters that have transmittance characteristics such that they can excite maximum fluorescence in a specimen (excitation filter)

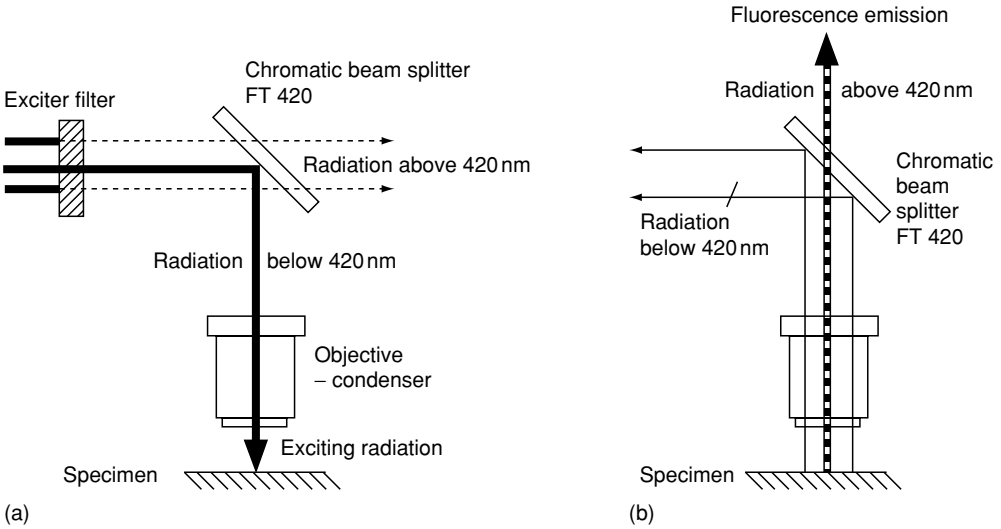

Figure 4 Diagrammatic representation of a typical fluorescence microscope epiillumination system equipped with a filter combination, as shown in **Figure 5(a)**. (a) Broad-spectrum illumination first passes through a colored glass filter that transmits maximally at 365 nm. Upon striking the chromatic beam splitter in this example all wavelengths below 420 nm are reflected to the top surface of the specimen through the objective, causing it to fluoresce at a longer wavelength. Both the excitation (by reflection) and fluorescent wavelength are generated at the specimen surface and are again collected by the objective (b). However, only fluorescent wavelengths longer than 420 nm are now transmitted to the detector (e.g., the eye) by the beam splitter, which acts as a barrier filter. In most instruments, an additional barrier filter is placed between the beam splitter and the detector to define further the range of fluorescence wavelengths to be examined. Reproduced from Fulcher RG, Irwing DW and Francisco A de (1989) Fluorescence microscopy: Applications in food analysis. In: Munck L (ed.) *Fluorescence Analysis in Foods*. Harlow: Longman, with permission.

and subsequently filter out all but the fluorescent wavelengths (barrier filter) from the final image (Figure 5). Fluorescence microscopes can be combined with sophisticated monochromators and detectors which can extend into the short-wave UV region (220 nm) with the use of quartz objectives.

An attractive refinement of the fluorescence microscopy is the laser confocal microscope (such as the Leica Lasertechnik TCS NT confocal microscope). A laser beam is scanned over a fixed specimen, restricting illumination and detection to the same small volume. Detection, via a computer-controlled video system that controls the scanning function, is limited to what is in focus, allowing the study of crude samples in three dimensions and eliminating glare. Up to four lasers can be attached to the microscope from UV to near infrared. The specimens studied are usually stained by fluorochromes responsive to the specific wavelength of the laser.

Spectrofluorometry

Recent research has shown that spectrofluorometers may be used in combination with computers and suitable software to predict important quality factors in foods analogous with the established near infrared spectroscopic techniques. (*See* **Spectroscopy**: Near-infared.)

Applications are, for example, to monitor formation of colored substances (Maillard compounds) in sugar processes, oxidation products in frying oils, and purity of pectins. However, the application of these techniques in industry has been delayed due to lack of suitable instruments which are able to measure a sample area large enough to obtain correct sampling of the often nonhomogeneous food products.

Sample holders delivered with the instrument are for front-surface (reflected) or transmitted light and employ quartz cuvettes. Fluorescence of insoluble particles such as wheat flour can be studied after fine ball-milling and suspension (about 1 mg ml^{-1}) in fluorescence-free glycerin/water which slows down sedimentation.

A typical laboratory spectrofluorometer with a measurement area of only 1×9 mm is shown in Figure 6. It is a convenient tool for obtaining complete excitation and emission spectra of homogeneous fluorescent samples, for identification and quantification through data processing, and for measuring fluorescence at a defined pair of excitation and emission wavelengths. For some applications in which maximum excitation and emission wavelengths are known, a simple and inexpensive filter fluorometer is superior to a spectrofluorometer because the superior accuracy in wavelength definition obtainable with

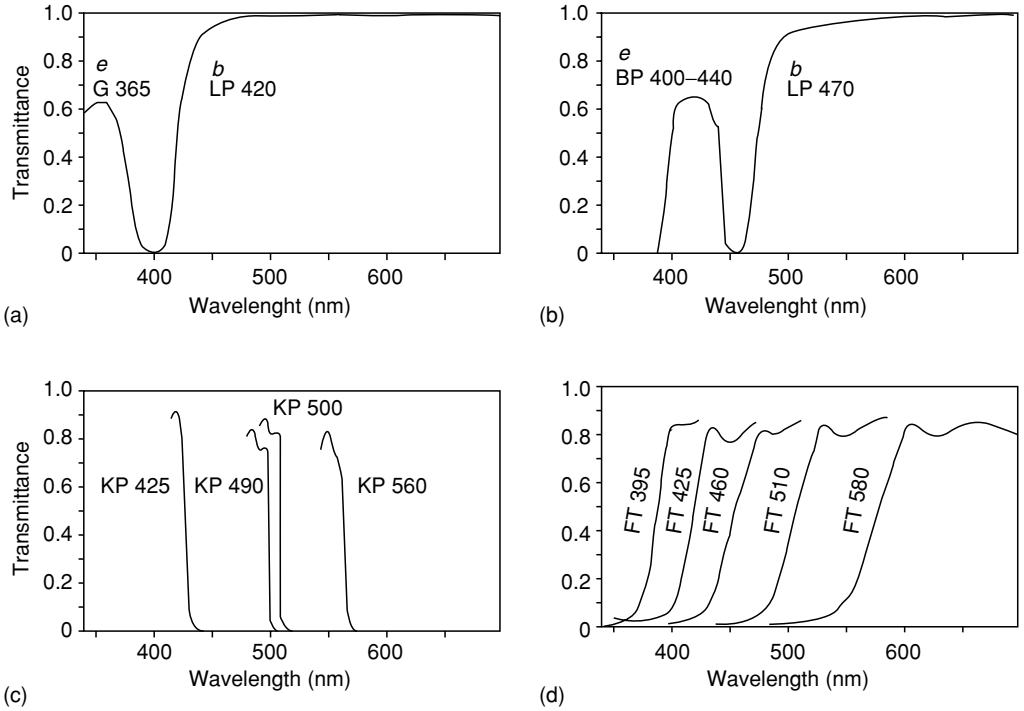

Figure 5 (a, b) Filters. Typical spectral transmission curves of paired exciter (*e*) and barrier (*b*) filters designed to optimize fluorescence by an HBO illumination. G 365 is a colored glass filter while the other filters displayed in the figure are interference filters coated by vacuum evaporation of metallic oxides. The latter can be specified more flexibly, but are also more expensive. A bandpass filter (BP 400–440) shows sharp cut-offs on each side of the curve (b). Short-pass (KP) interference filters (c) are used as exciting filters. They suppress long-wavelength light, and show a steep cut-off at the high limit of transmittance. Barrier filters may also be of the BP type but are mostly long-pass (LP) filters (d) showing steep cut-offs at the lower limit of transmittance and uniform transmittance throughout the rest of the visible range, making it possible to detect a range of emission wavelengths with the same barrier filter. Filters are examined in a high-quality spectrophotometer. Reproduced from Fulcher RG, Irwing DW and Francisco A de (1989) Fluorescence microscopy: Applications in food analysis. In: Munck L (ed.) *Fluorescence Analysis in Foods*. Harlow: Longman, with permission.

the latter instrument may be compromised by the greater amount of light that can pass through the filters of a filter fluorometer. Fluorescence thin-layer chromatography (TLC) scanners, which are specialized (epi)spectrofluorometers, are often convenient for studying, for example, bones in fish fillets and connective tissue in minced meat. Filter and spectrofluorometry may be exploited to provide highly sensitive detection of specific substances in high-pressure liquid chromatography (HPLC) and flow-injection analysis (FIA). (*See* **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography.)

Applications

Monitoring Structure and Function in Food Systems

In many applications, the evaluation of structural (physical) properties of a food material accompanies determination of its (bio)chemical characteristics. In the case of wheat grain, it becomes readily apparent that the bran tissues (pericarp and aleurone; Figure 7:2:1 and 3) have yellow and blue fluorescence, respectively, at specific excitation wavelengths. A method can then be developed to quantify fluorescence on the basis of excitation and emission spectra obtained with suspensions of finely milled hand-dissected samples of the pure botanical tissue components (Figure 2). An additional finding will be that the floury endosperm part of the seed autofluoresces in the UV region, which cannot be detected by the eye in the fluorescence microscope. This information could be used tentatively to monitor the fractionation process in a wheat mill by spectrofluorometry much more accurately than with the present ash and color analyses. It is also possible, for example, to optimize the milling process to obtain an increased content of the vitamin- and mineral-rich bran component, aleurone, in the flour, while still obtaining a relatively white bread with good baking volume despite a high ash content.

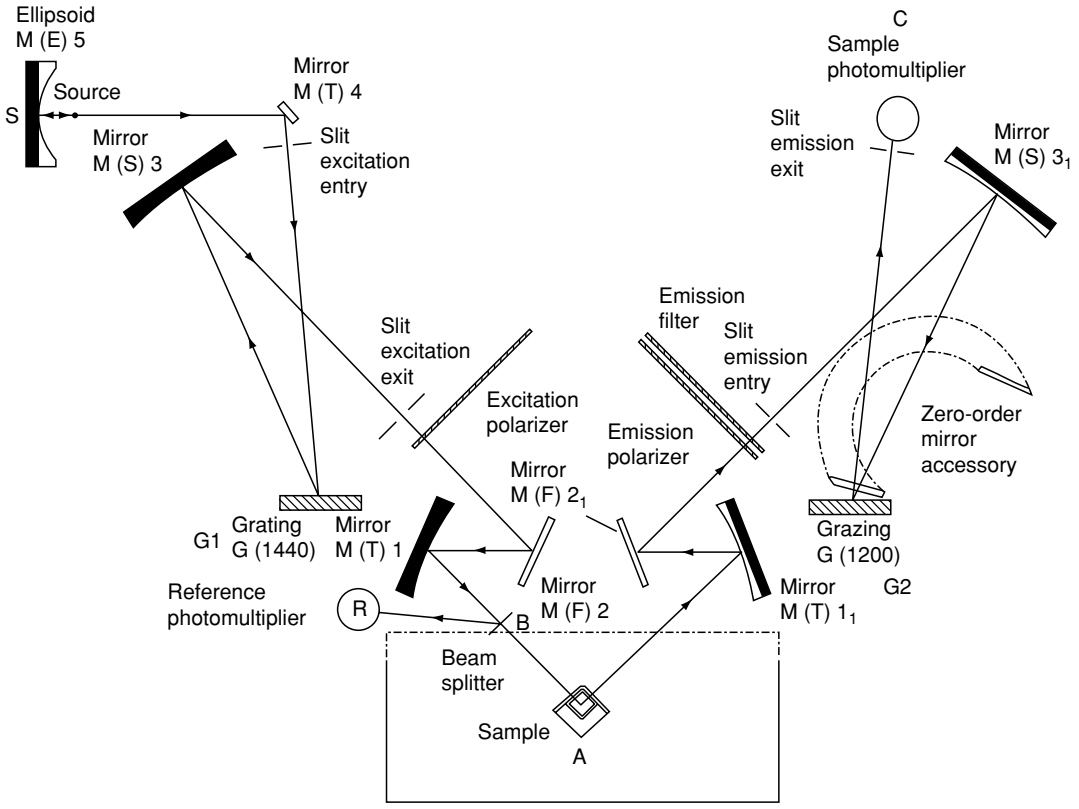

Figure 6 Fluorescence spectrophotometer (Perkin Elmer LS-50). The source (S) is a pulse xenon flashtube (less than 0·02 s for a complete pulse). Light from S is reflected by a mirror through a slit to the diffraction grating (G1) which has 1440 lines per millimeter. The resulting narrow-wavelength band is transferred via another system of mirrors and slit to the sample (A). The effective wavelength of the excitation beam is determined by the setting of the grating, the angle of which is controlled by means of a stepping motor. A very small proportion of the excitation light is reflected by a beam splitter (B) to a reference photomultiplier. To correct for the response of the reference multiplier, a rhodamine correction curve is stored within the software in the instrument computer. Light emitted by the sample is transferred via slits and mirrors to a second diffraction grating (G2), which has 1200 lines per millimeter and is controlled by another stepping motor, and on to the sample photo multiplier (C). A microprocessor controls the xenon tube, grating stepping motors, slit apertures, and photo multipliers. The digitized signal is accumulated for about 0·08 s, i.e., four flashes of the source. The sample accumulation is divided by the corrected reference accumulation to obtain a ratio that is unaffected by variation in source intensity. Output is given by the microprocessor on a data screen and a recorder. The computer is programmed so as to allow calibration spectra to be subtracted from sample spectra to generate difference spectra. The wavelengths of the instrument can be calibrated by the characteristic fluorescence spectrum of benzene vapors, for example, or from fluorescent standards cast into plastic blocks furnished by the instrument manufacturers. It is important in some applications to be able to record accurately several of the internal settings of the instrument, including spectral correction and different scanning modes, and this can be done with a high-specification spectrofluorometer. The practicable wavelength ranges for the standard version of the instrument are about 230–600 nm for excitation and about 250–650 nm for emission. Reproduced from Spectroscopy: Fluorescence, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

The problem of calibrating and handling such complex fluorescence information will be discussed below. Likewise, autofluorescence of animal tissue components and parasites, detected in the fluorescence microscope and defined in the spectrofluorometer, can be exploited in digital video techniques coupled to image analyzers and robots, e.g., to facilitate automatic butchering operations and cutting out bones from cod fillets (Figure 7:3A–D). In the laboratory, naturally fluorescent compounds such as vitamin E components can be separated and detected in an HPLC system coupled to a fluorometer. In commercial-scale fermentation processes, the fluorescing reduced forms (NADH and NADPH) of the respiratory coenzymes nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP) can be used online to follow the rate of the fermentation process. (*See* **Tocopherols:** Properties and Determination.)

Similarly, a food chemist in need of a certain assay, e.g., to monitor the breakdown of mixed-linkage β-glucan in the malting and brewing industry or to

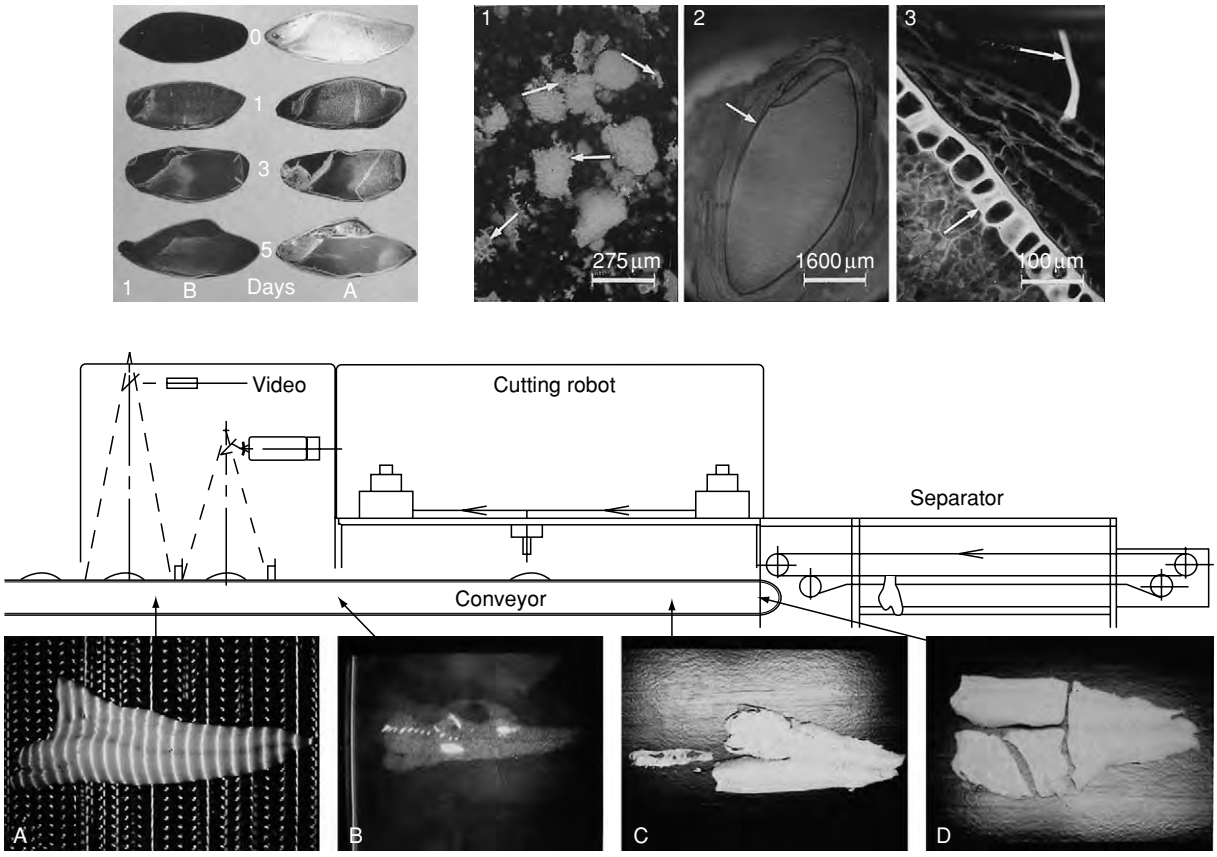

Figure 7 (see color plate 129) Examples of fluorescence microscopy and imaging reproduced from articles referred to below in Munck L. (ed.) (1989) Fluorescence Analysis in Foods, Harlow: Longman. pp. 289. 7:1. Cryostat sections of 0, 1-, 3- and 5-day malted barley viewed through a 4× microscope objective. The sections were stained by (A) Calcofluor/Fast green photographed at 400 nm excitation/418 nm emission in order to show the patterns of cell wall (β-D-glucan) breakdown and by (B) FITC-labelled antibodies against α-amylase viewed with BP 455-490 nm exciter/LWP515 nm emission filter. Sections are identical pairwise – first stained with method B followed by A. It is seen that α-amylase is spread from the germ (the structure on the left of each section) and from the aleurone layer enclosing the starchy endosperm. The enzyme invades only the endosperm areas where cell walls are absent. (Compare B with A on each day of malting.) From Munck L. Practical experiences in the development of fluorescence analyses in an applied food research laboratory, ibid. p. 1-32. 7:2:1-3. represent a wheat and barley material with natural and added fluorochromes. 7:2:1. A view of ground whole wheat flattened slightly with cover glass, excited at 365 nm with a barrier filter from 420 nm. Bran fragments are readily detected by the natural bright blue fluorescence characteristic of the aleurone layer (arrows). See also **Figure 2b**. 7:2:2. A low-power view of a longitudinally halved barley kernel treated with a multiple fluorescent stain (Safranin, Basic Fuchsin, Alcian Blue) to show several grain components simultaneously. The starchy endosperm (central region of the kernel) fluoresces bright yellow due to starch-safranin interaction under 450-490 nm excitation, while the outer bran layers are deep red as a result of Basic Fuchsin interaction with lignified and other phenolic-enriched structures such as cell walls (arrows). In this case, safranin is used to detect starch, while the other stains are added as counterstains to surpress non-starch specific fluorescence. 7:2:3. As in 7:2:2, but showing the bran tissues at higher magnification and 365 nm excitation. In this case, the procedure clearly differentiates between I. ferulic acid-enriched aleurone cell walls (lower left) detected due to their natural fluorescence (as in 7:2:1), II. surface trichomes (arrows) which do not bind significant amounts of the red fuchsin stain and III. the pericarp cell walls (right diagonal) which are intensely stained deep red. Starch does not fluoresce significantly at these wavelengths. From Fulcher RG, Irwing DW and de Francisco A. Fluorescence microscopy: Applications in food analysis, ibid. 59-109. 7:3:A-D. Operation of an automatic Fish Fillet Deboning Line based on fluorescent technology. Skinned cod fillets are fed onto a steel conveyor while moving at one m/s. The video detector (see drawing above the photographs) detects a fillet (A). The detector consists of two cameras and two light sources for visible and UV light for fluorescence, respectively. The visual light is projected in stripes (A) in order to detect (I) the boundaries of the filet and (II) the topography of the fillet for later partition (D) into desired weight portions. The video UV fluorescence system detects the naturally fluorescent bones (ex. max 340/em. max 390) seen on a TV monitor (B). The information from the two cameras is directed to a computer which directs a water jet nozzle (diameter 0.1 mm, pressure 300 bar) robot with an operating space of 1 m. The cutting of the fillet bones (C) and further partition of the fillet (D) is performed while the steel conveyor is moving and completed within the time span of one second. Bones and fillet pieces are separated and graded in the separator unit drawn above. This principle may also be used to remove bones from chicken breast fillets. From Jensen SAa, Reenberg S and Munck L. Fluorescence analysis in fish and meat technology, ibid. 171–180.

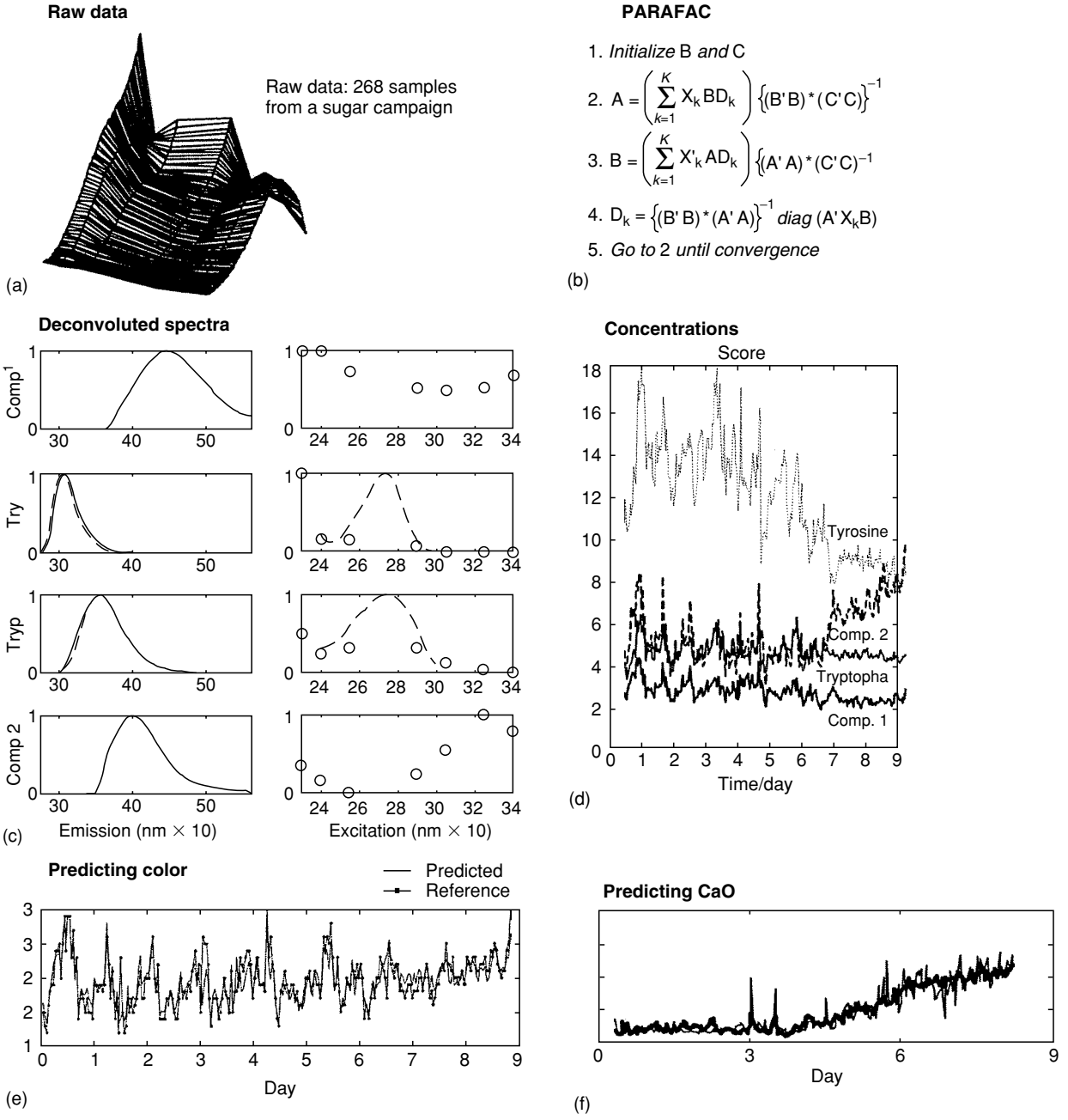

Figure 8 An example of mathematical chromatography is multiway modeling of fluorescence landscapes of sugar into the pure excitation and emission spectra of the underlying fluorophores. A total of 268 samples of sugar have been dissolved in water and measured spectrofluorometrically, spanning the sugar campaign with three samples representing three 8-h periods per 24 h. Each such measurement gives a characteristic landscape of a sample (a). By using the so-called parallel factor analysis (PARAFAC) model (b), it is possible to describe these landscapes by the variation in four estimated fluorophores (c). Thus, at any time the sugar landscape of any sugar sample can be described by having different amounts expressed in scores of these four fluorophores, defined by their estimated excitation and emission spectra. The former is described less accurately than the latter due to sampling (20 nm). The amounts of the four fluorophores in the 268 samples thoughout the season are shown in (d). Furthermore, it is possible to qualify which chemical analytes these four fluorophores represent, just as in ordinary chemical chromatography. In this case, two of the analytes have been identified as the uncolored amino acids tyrosine and tryptophan which have the potential for color formation through reaction with reducing sugars. (Estimated and true emission spectra are shown.) The other two fluorophores represent colored fluorescent high-molecular Maillard products. These interpretations have been checked with high-performance liquid chromatography. Thus, a chemical understanding of the fluctuations during the campaign is provided. In (e) and (f) it is demonstrated that this spectrofluorometric chemically based variation can be used for creating cheap online screening analyses for prediction of, e.g., important process variables such as CaO in sugar juice or quality parameters such as color in sugar by using a multiple linear regression model based on scores of the four fluorophores.

screen for lipases in oat products, may select from potentially useful fluorochomes (Table 1) to solve the analytical problems by employing secondary fluorescence techniques: for example, calcofluor white Mr2 new and fluorescein dibutyrate can be used to determine mixed-linkage β-glucan and esterase/lipase contents, respectively, of cereal grains. In the case of the calcofluor/β-glucan assay, two official European brewing convention methods, one visual – malt modification analysis (Figure 7:1A) – and one chemical – FIA for mixed-linkage β-glucan in wort and beer – have been developed. These methods have helped to strengthen the insight of brewing chemists into the strong relationship between the physical cell wall structure of barley and malt and the chemistry of wort and beer related to viscosity and filterability due to the influence of β-glucans. Mg-ANS (1-aniline-8-naphthalene sulfonic acid) can be used to detect dead cells in pitching brewer's yeast, and fluorescein isothiocyanate (FITC)-immunofluorescent visual and chemical methods can also be applied successfully, for example, to monitor the distribution of α-amylase activity in malting barley seeds (Figure 7:1B). A range of components such as aleurone, starch, β-glucans, protein bodies, fat globules, protein, and crystalline inclusions can be convincingly demonstrated by employing a range of fluorochromes and treatments one at a time or in combination as demonstrated in an example with wheat in Figure 7:2:2 and 3. Several of these methods can also be used quantitatively to determine components in homogenized samples by spectrofluorometry.

Fluorescence Spectrofluorometry for Prediction of Food Quality Calibrated with Chemometrics

Due to its specificity and sensitivity, spectrofluorometric analysis allows a unique characterization of a fluorescent food sample, thus complementing near infrared spectrophotometry. Fluorescence measurement with fluorochromes *in vitro*, e.g., to detect lipase (esterase) activity with fluorescein dibutyrate at 495 nm excitation and 525 nm emission (Figure 1) is a rather straightforward analysis using the equations given above. However, spectra from complex fluorescence systems, such as that from wheat flour (Figure 2), need advanced data analysis techniques (chemometrics) similar to those developed for protein analysis of seeds by near infrared spectroscopy. These use complete spectra, and involve principal component analysis for discrimination between normal and deviating samples and partial least-squares analysis (PLS) for calibration to chemical analyses for prediction. Precise calibration of the spectrofluorometer that is stable from day to day is obligatory here, if data from whole spectra are to be utilized. Spectrofluorometric data in the form of excitation/emission landscapes are ideally resolved by multiway chemometric algorithms such as parallel factor analysis (PARAFAC: Figure 8) which in the fluorescence example from process control in a beet sugar factory covers three dimensions: excitation, emission, and sample. In a PLS evaluation of a complex fluorescence landscape the latent factors, or loadings, are difficult to interpret due to orthogonality. In contrast, PARAFAC loadings are interpretable and expressed in the form of estimates of the underlying excitation and emission spectra of the corresponding fluorophores. If stable computerized instruments with enlarged measurement area for fluorescence monitoring (mathematical chromatography) of inhomogeneous samples in the food industry can be developed, the chemometric software already developed will be able to create a measurement system with unique potential with regard to sensitivity and specificity to measure those quality characteristics which depend on fluorophores.

See also: **Carotenoids**: Occurrence, Properties, and Determination; **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography; **Polycyclic Aromatic Hydrocarbons**; **Retinol**: Properties and Determination; **Spectroscopy**: Near-infrared; **Tocopherols**: Properties and Determination

Further Reading

Bro R (1999) Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis. *Journal of the Chemometrics and Intelligent Laboratory Systems* 46: 133–147.

Engelsen SB (1997) Explorative spectrometric evaluations of frying oil deterioration. *Journal of the American Oil Chemist Society* 74: 1495–1508.

Engelsen SB, Mikkelsen E and Munck L (1998) New approaches to rapid spectroscopic evaluation of properties in pectic polymers. *Progress in Colloid Polymer Science* 108: 166–174.

Guilbault GG (ed.) (1990) *Practical Fluorescence – Theory, Methods and Techniques*. New York: Marcel Dekker.

Hurtubuise RJ (1984) *Solid Surface Luminescence Analysis*. New York: Marcel Dekker.

Lakowicz JR (1999) *Principles of Fluorescence Spectroscopy*. New York: Kluwer Academic/Plenum Publishers.

Morgan LC, Gormally J, Hubbard AR, d'Lacey C and Ockleford CD (1999) Confocal microscopy: theory and applications. *Methods in Molecular Biology* 114: 51–74.

Munck L (ed.) (1989) *Fluorescence Analysis in Foods*. Harlow: Longman.

Munck L, Nørgaard L, Engelsen SB, Bro R and Andersson CA (1998) Chemometrics in food science – a demonstration of the feasibility of a highly exploratory, inductive

evaluation strategy of fundamental scientific significance. *Journal of the Chemometrics and Intelligent Laboratory Systems* 44: 31–60.
Rost FWD (1992) *Fluorescence Microscopy*, vols 1 and 2. Cambridge: Cambridge University Press.
Von Sengbusch G and Thaer AA (1973) In: Thaer AA and Sernetz M (eds) *Fluorescence Techniques in Cell Biology*, p. 31. New York: Springer.
Wang XF and Herman B (eds) (1996) *Fluorescence Imaging Spectroscopy and Microscopy*. New York: Wiley.
Wolfbeis OS (ed.) (1993) *Fluorescence Spectroscopy: New Methods and Applications*. Berlin: Springer.

Atomic Emission and Absorption

N Ulrich, University of Hannover, Hannover, Germany

Summary

Atomic spectrometry is suitable for the determination of trace concentrations of most elements. The principle is the absorption or emission of light of certain wavelengths after the atomization and excitation of the sample. The wavelength gives qualitative information about the element, whereas the intensity of the emission or the quotient of the intensity before and after the absorption is proportional to the concentration of the analyte. Several spectral and non-spectral interferences must be considered.

Background

Atomic emission and absorption spectrometry are used for the qualitative and quantitative determination of chemical elements, mainly metals and semimetals in a huge variety of different matrices. Both techniques are based on the interaction of atoms and electromagnetic energy. Kirchhoff and Bunsen grounded the atomic spectrometry in the middle of the nineteenth century. They observed that free, gaseous atoms of an element can absorb or emit radiation of discrete and specific wavelengths. In atomic spectrometry, wavelengths between 170 and 800 nm are investigated.

The Atomizing Process – Important for Both Techniques

Prior to the excitation or absorption step, free atoms of the elements must be formed. In classic techniques, the samples are introduced into the system as aerosols, e.g., solutions of sodium chloride. These aerosols are desolvated and evaporated rapidly by adding thermal energy to form microparticles. By interactions with the environment in the flame or the plasma, free atoms are formed that are capable of absorbing light of specific wavelengths (atomic absorption) or, if excited by thermal energy, can emit light of specific wavelengths (atomic emission). The wavelength is equal for both processes. **Figure 1** shows the atomization process for aerosols using sodium chloride (NaCl) as an example.

Excitation and Relaxation of Valence Electrons

After the atomization, the excitation of the electrons (absorption) and relaxation (emission) take place. Only the valence electrons are capable of the radiation used in atomic spectrometry. In their basic form, these electrons are in certain energy levels, representing the ground state for an atom of a given element. The electrons can absorb energy and form energy-enriched, excited states. Two forms of excitation are of importance for atomic spectrometry:

1. Thermal excitation: The electrons can be excited by thermal energy. At temperatures of 2000–3000 °C, most of the atoms are in the ground state, following the Boltzmann distribution. More states can be populated at higher temperatures, e.g., in the Ar-plasma of an inductively coupled plasma (ICP). The relaxation takes place as electromagnetic radiation, which is measured in atomic emission spectrometry (AES).
2. Electromagnetic excitation: The electrons can absorb light of a certain wavelength according to Einstein's law of absorption and emission, which is measured in atomic absorption spectrometry (AAS). The wavelength depends on the energy difference between the ground and excited electronic state of the atom. The relaxation may take place as thermal or electromagnetic radiation.

There are many emission lines in different series starting from the ground state or an excited state, but

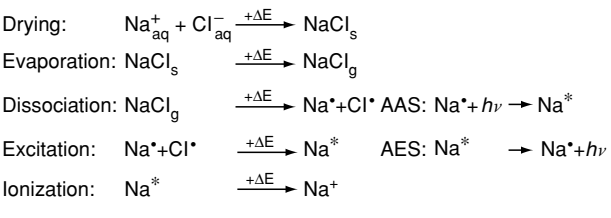

Figure 1 Processes in the atomizer of an atomic spectrometer.

not every emission line is also a detectable absorption line in AAS. The intensity of the absorbance line depends on the population of that state (,) which absorbs the light. In most cases, only the ground state is populated to such an extent that intensive radiation is emitted.

Atomic Emission Spectrometry

In AES, the atoms are excited by thermal energy and emit light of specific wavelengths. The frequency of the light is proportional to the energy difference of both states. For optical emission spectrometry, the wavelengths are in the ultraviolet/visible region. This light is measured and can be correlated by the wavelength to the element (qualitative information) and by the intensity to the concentration (quantitative information).

The intensity of the emitted light depends on the number of atoms in the excited state. This number will be larger for higher temperatures given by the Boltzmann distribution. At temperatures below 2500 °C, only transitions between the ground state and one of the first excited states are possible for most of the elements. The alkali metals can be excited in sufficient rates at temperatures of *c*. 2000 °C. The concentration of the analyte in the sample is directly proportional to the intensity of the emitted light.

Principal Components of an Atomic Emission Spectrometer

The main components of an atomic emission spectrometer are the sample introduction device, which is often a pneumatic nebulizer system, the atomization and excitation source, one or more monochromators, and a detector, mostly a photomultiplier. The principal set-up of an AES is shown in Figure 2. AES can be used as single element method or as a simultaneous measurement device, as detailed below.

Atomizing and Excitation Techniques

There are many techniques for atomizing and exciting samples, like bow and spark excitation or glow discharge excitation. The two most important methods are flame photometry, which is suitable primarily for alkaline and some earth alkaline elements, and ICP, which is used widely as an energy source in atomic emission spectrometry.

Flame Photometry (Atomic Emission Spectrometry)

Flame photometry was established by Bunsen. It is called photometry, although it is an emission technique. The correlation between the intensity and concentration is linear, and the Lambert–Beer law cannot be applied. The atoms are excited in a flame, and the intensity of the emitted light is measured.

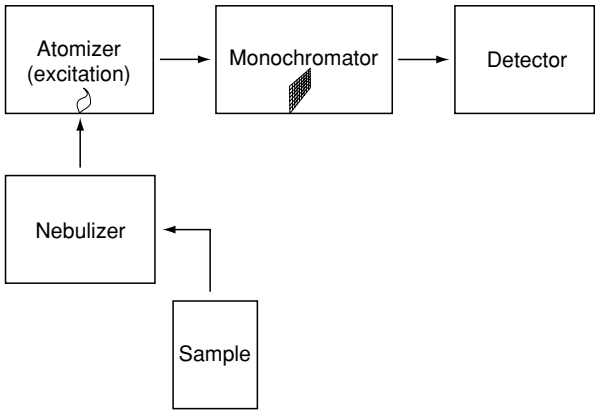

Figure 2 Principal set-up of an AES apparatus.

The apparatus of FAES contains a nebulizer system, a burner, suitable for different gases like nitrous oxide or methane, a monochromator, which is important for the quality of the measurement, and a detector. Often, flame atomization absorption spectrometry systems with the lamps turned off are used as the FAES device.

Applications and analytical performance In the early years, FAES was used for all metals. However, the exciting efficiency of the flames was poor, and the detection limits were quite high. Today, FAES is mainly used for the determination of alkali metals in solutions. The sensitivity for these elements, especially Na and K, can be better than for the FAAS. Typical detection limits are $0.01\,\mathrm{mg\,l^{-1}}$ (Na) and $0.1\,\mathrm{mg\,l^{-1}}$ (K) in liquid biological samples. For other elements, AES with plasma excitation, atomic absorption spectrometry, or mass spectrometry is used nowadays.

Problems and interferences Spectrometric determinations are relative methods, so calibration is necessary. Differences between the samples and the standard solutions can lead to systematic errors, called interferences. In spectrometry, there are two types of interferences. The spectral interferences are problems caused by unspecific radiation or absorption, whereas nonspectral interferences affect the concentration of the analyte in the atomizer and exciting zone.

Spectral interferences in flames include the spectral lines of matrix compounds and the high continuous background of the flame, emitting light of all wavelengths. In relatively cold flames (2000 °C), emission

spectra with few lines are observed. However, excesses of one element, e.g., 1000-fold excess of Na for Li determination can lead to an overlapping of the signals. In addition, Ca and Sr can lead to interferences.

Nonspectral interferences can be classified as chemical interferences, transport interferences, ionization interferences, and self-absorption. If stable compounds of the analyte and matrix components are formed, these will hinder or even completely suppress the atomization and excitation. Organic substances can alter the flame temperature, which is important for the FAES, and can emit continuous background radiation. In addition, the analytes can react with the flame gases, e.g., Ca can form CaO with oxygen.

Transport interferences occur when the amount of sample introduced into the atomizer changes over time. The main reason is a difference in viscosity between water solution and organic solvents. In addition, larger concentrations of salts and matrix compounds can disturb the nebulizing process.

Ionization interferences occur, when part of the analyte is ionized within the flame. It only takes place to a small extent, but causes a reduction of the population in the excited state. One problem with a high concentration of the analyte is self-absorption. Because emission and absorption take place at the same wavelength, analyte atoms can absorb light, which is emitted by other analytes. This leads to a significant curving of the calibration graph.

Inductively Coupled Plasma–Atomic Emission Spectrometry (ICP-AES)

The inductively coupled plasma was introduced into atomic emission spectrometry by Greenfield in 1964. ICP-AES is today one of the most widely used techniques for trace metal and semimetal determination in a huge variety of different samples.

An argon plasma with a gas consumption of *c.* 17 l min^{-1} is used as atomizer. The gas is delivered through three concentric quartz tubes (the torch), which are surrounded by an induction coil. This coil is operated at a frequency of 27.1 MHz and a power of 1000–2000 W. Gas temperatures of 6000–8000 °C are observed in the plasma, which, in combination with the relative long residence time of the analytes in the plasma, lead to an effective energy transfer on to the analytes. Even refractive metals and oxides can be atomized to a great extent. The spectra obtained are often complicated and contain many lines, especially for the transition metals. The plasma can be observed either axially or radially.

One of the main advantages with respect to AAS is the ability of multielement determinations, because sequential and simultaneous set-ups can be used. Analytical wavelengths range from 170 to 800 nm. For sequential investigations, the monochromator, often of the Czerny–Turner type, is tuned for each element. Several monochromators of unchangeable wavelengths are fixed in the detector, mostly forming a Rowland circle, for simultaneous measurements.

Applications and analytical performance The ICP-AES is suitable for a wide range of elements, because the high temperatures provide very good atomizing conditions. All metals and semimetals and even some of the nonmetals (e.g., sulfur, phosphorus, and iodine) can be detected. The simultaneous spectrometers are limited to six to 30 elements, depending on the number of monochromators in the detector. Quantitative results for these elements can be obtained in *c.* 1 min. Typical detection limits range from 0.01 $\mu g\,l^{-1}$ for Ca, Mn, Mg, and Mo to 20 $\mu g\,l^{-1}$ for As, Bi, and Sn. The dynamic range reaches usually over five to six decades, so that matrix compounds and trace elements often can be detected in one analysis.

Gaseous, liquid, and solid samples can be detected using ICP-AES. Liquids can be nebulized with any of the standard nebulizers. Flow injection and on-line coupling to chromatographic devices are other options. Gases, e.g., metal hydrides, can be brought directly into the plasma. The dry plasma and the nearly complete sample introduction lead to a significant enhancement of the detection limits. For solid samples, mostly slurry techniques with particles less than 2 μm in diameter are used.

Problems and interferences The drawbacks of ICP-AES are the high costs of the apparatus and maintenance, and the high spectroscopic background of the plasma.

In addition, there are numerous interferences observed in the spectra. Spectral interferences arise from the plasma itself, which is an intensive source of unspecific radiation and argon emission lines. Both depend on the temperature and the composition of the plasma gas and vary, e.g., with the introduction of water from the sample. Further, there are emission lines of hydrogen, nitrogen, and oxygen. Different molecular species are formed in the plasma, e.g., N_2^+, OH, NH, or NO, which show a large number of rotation and vibration bands.

Nearly all elements in the sample emit characteristic spectral lines not only from the ground state but from several excited states, because the high temperatures in the plasma lead to a remarkable population of numerous states. These interferences can be resolved to some extent by the use of high-efficiency

monochromators, but many lines are too close together for separation.

Chemical interference is not important in most cases. Transport interference can occur, as described for FAES.

Other Plasma Sources for Atomization

Microwave-induced plasmas (MIP) are formed by microwaves (frequency 2.45 GHz, power 50–200 W). This plasma is mainly used as a detector for gas chromatography. Mostly, He is used as the plasma gas, providing very high electron temperatures and excitation energies, e.g., for the nonmetals.

Several other plasma sources have been applied for AES. Glow discharges, sparks, and bows are mostly used for metallic samples. In addition, capacity coupled plasmas can be utilized to excite metal atoms.

Atomic Absorption Spectrometry

This technique is based on the selective absorption of light by the gaseous free atoms. The absorbance is the logarithm of the quotient of the radiation intensity before the absorption, I_0, and after the absorption, I. According to the Lambert–Beer law, it is directly proportional to the concentration of the absorbing atoms in the atomizer, c, the length of the atomizer, d, and the absorption coefficient, a.

$$A = 1^{\circ}\mathrm{g}\frac{I_0}{I} = adc. \qquad (1)$$

The coefficient, a, depends on the element and the wavelength, γ, and can be related to the Einstein transition probability of an absorption process, B and the Planck constant, h.

$$a = B\frac{h}{\lambda}. \qquad (2)$$

Principal Components of an Atomic Absorption Spectrometer

The main component of an analytical absorption spectrometer is the light source, providing monochromatic light for the absorption process. Two types of light sources are mostly used. Hollow-cathode lamps contain a cathode of the analyte element and an anode, and are filled with a noble gas. There is a glow discharge between the cathode and the anode, in which positive gas ions are formed, which sputter element atoms of the cathode at relatively low temperatures. These lamps emit mostly lines excited from the ground state and show only slight line broadening. The other type of lamps are electrodeless discharge lamps that contain the element in a small quartz tube filled with a noble gas. A high-frequency field (*c*. 27 MHz) leads to a plasma within the tube, in which the element is excited and emits specific light. The advantages are the better stability of the lamp, especially for elements like antimony, mercury and tin, and the higher intensity. A problem is the formation of ions, which reduce the effective concentration of the excited atoms. Both lamp types generate narrow emission lines and are available for all elements that can be determined by AAS. AAS is mostly a single element or sequential analytical technique, because a specific light source is used for each analyte.

The light beam is sent through the atomized sample and then via a monochromator to the detection system similar to that of AES. Pneumatic nebulizers are used as well for AAS. In addition, slurry and solid samples have been investigated. The atomizer in the light path can be of different types. The principal set-up of an AAS is shown in Figure 3.

Atomizing Techniques for Atomic Absorption Spectrometry

The purpose of the atomizer is the reproducible formation of gaseous atoms of the analyte in the ground state. There are three main atomizing techniques in atomic absorption spectrometry. The classical method is flame atomization. Electrothermal atomization provides better detection limits for most elements, and chemical vapor atomization is well suited for gaseous samples like the metal and semimetal hydrides and gaseous mercury.

Flame Atomic Absorption Spectrometry

In FAAS, the liquid sample is introduced as an aerosol in a flame with temperatures between 2000 and 3000 °C. The flame should have a low self-absorption at the analyte wavelength and be a low source of

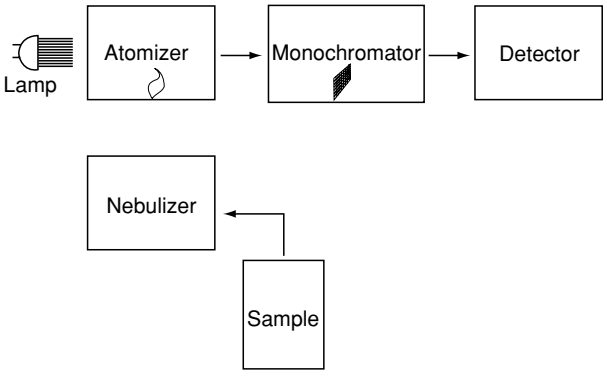

Figure 3 Principal set-up of an AAS apparatus.

white light to provide adequate signal-to-noise ratios. In addition, the residence time of the analyte atoms in the flame should be sufficient for the absorbing process, and therefore, lower flow rates for the flame gases should be used.

The most important is the acetylene/air flame, which is rather transparent over a broad region of wavelengths and, at temperatures of *c*. 2500 °C, is suitable for most analytes. Higher temperatures are needed for some elements like Be, Ca, Ti, V, Mo, and the rare earths, so acetylene/nitrous oxide flames with temperatures of *c*. 3000 °C are used. Other flame types include hydrogen/air and methane/air.

Applications and Analytical Performance

FAAS is suitable for samples with analyte concentrations of mg l^{-1} to the higher μg l^{-1}. The sensitivity is often given as a characteristic concentration, c_0, which depends on the parameter of the instrument (e.g., the geometry of the atomizer) and the element. Linear calibration functions are obtained up to 100 c_0, and the precision is frequently better than 0.5%.

FAAS can be used to determine the macronutrient elements in human body fluids and other biological matrices. Whereas sodium and potassium are mainly investigated by flame photometry, calcium and magnesium are analyzed by FAAS by direct measurement after dilution.

Problems and Interferences

The matrix in the samples can influence the drying, crystal formation, atomization, and distribution of elements in the flame. Mostly, the kinetic of the atomization is changed by analyte compounds with high boiling points. The energy of the flame can lead to an ionization of certain elements like the alkali or earth alkali metals and to curved calibration functions. The addition of electron donators like Cs shifts the equilibrium to the atoms and provides a better linearity of the calibration. Chemical reactions and spectral effects are less important for the FAAS, because of the dilution of the analytes and the matrix by the flame gases. Problems can arise from organic solvents and high salt concentrations, which alter the viscosity of the solution.

Electrothermal Atomic Absorption Spectrometry (ETAAS)

One major problem of the FAAS is the residence time in the atomizer. In order to improve this time, the ETAAS has been invented by L'vov in the middle of the twentieth century and is now one of the most widely applied techniques for trace and ultratrace metal determinations.

In the ETAAS, the atomizer is a small quartz tube mounted between two electrodes. Samples containing a few microliters of liquid or micrograms of solid are injected through a small hole on top of the tube. The oven is heated electrically following a temperature program:

1. The drying step, in which the water or the solvent is evaporated.
2. The sample preparation step, in which matrix components are evaporated or altered.
3. The atomization step, in which the analytes are atomized at constant temperatures, 1500–2500 °C, depending on the analyte.
4. The cleaning step, in which the temperature is brought to a maximum, and the graphite furnace is cleaned.

In order to improve the sensitivity of the ETAAS and to inhibit the formation of carbides by elements like Mo, V, and Ti, the furnace can be coated by different materials including Os and Pd. In contrast to the continuous signal of the FAAS, in ETAAS, a transient signal, mostly a heavy tailing peak, is obtained.

Analytical Performance

The ETAAS has several advantages in respect to the FAAS. Slurry and solid samples can be used without dissolution, and the analytes can be separated from the matrix using a program of different temperatures. The method provides a high sensitivity and very low sample consumption, because nearly the whole sample is introduced into the atomizer. Drawbacks are the complicated apparatus and handling, the low sample throughput, and the large number of interferences. The ETAAS is now widely used for the determination of trace elements in biological samples.

The detection limits of the ETAAS are from 0.5 μg l^{-1} for elements like V and Li to 0.005 μg l^{-1} for elements like K, Cd, and Zn (10–1000-fold better than for the FAAS). The absolute sensitivity depends on the dimensions and geometry of the quartz tube and further experimental parameters. It is given as characteristic mass, m_0. The precision is *c*. 1% standard deviation at 10 m_0, and the linear range ends by *c*. 100 m_0, similar to the FAAS.

Problems and Interferences

In the ETAAS, numerous spectral interferences have been determined. These interferences are based on the insufficient separation of the analytes and matrix. The radiation from the light source can be absorbed unspecifically by atoms and molecules, leading to overlapping atomic and molecular bands or can be scattered on particles like undissociated metal salts

in the light beam. At high atomizing temperatures, sublimating graphite particles can lead to problems.

The background depends strongly on the matrix, so that background correction is much more important than for FAAS. Modern apparatus use either a deuterium lamp or the Zeeman effect for correction of the data. The standard addition method should be used for complex matrices.

Chemical Vapor Atomic Absorption Spectrometry (CVAAS)

CVAAS is used for the determination of gaseous metal and semimetal compounds. It is an important detector for the speciation of hydride-forming elements. There are two sectors of the CVAAS: the cold vapor technique, which is only suitable for mercury, and the hydride generation process, which can be applied to Se, Te, As, Sb, Bi, Sn, and Pb. In the latter case, the elements are reduced to the corresponding hydrides mostly with sodium tetrahydridoborate in acidic solution. A quartz furnace is used as the atomization unit, which is typically heated electrically to 900–1000 °C. The samples are introduced as gases to nearly 100% by an inert carrier gas stream. The residence time in the atomizer depends on this gas stream.

Mercury compounds can be reduced by sodium tetrahydridoborate to elemental mercury, which can be introduced by inert gas into the atomizer nearly without heating. The quartz furnace is used at 120 °C in order to inhibit the condensation of mercury and to reduce the scattering on water particles in the oven.

Analytical Performance

The absolute performance of the hydride generation is *c*. 100-fold less than the ETAAS, but the larger sample volumes (up to 100 ml in one batch reactor) lead to an increase in the detection ability usually at a factor of 10. The determination of 0.01 μg per liter of arsenic or selenium is possible. In addition, the oxidation states and organic species of the hydride formers can be detected separately, so that more information about the species distribution of these elements in different matrices can be gathered. In addition, the matrix is nearly completely separated from the analytes. The drawbacks are the complicated experimental set-up, relatively long analysis times, the instability of the formed hydrides, and the high risk of contaminations by the addition of high excesses of solid and difficult-to-clean reagents (e.g., sodium tetrahydridoborate, sodium hydroxide).

The CVAAS for the mercury determination provides very good detection limits (0.001 μg l^{-1}). Even enrichment is possible by using gold foils for amalgam formation.

Problems and Interferences

In the CVAAS, the separation of the analytes of the matrix is nearly complete for nonhydride formers. However, the hydride-forming elements can strongly interfere with the reduction of one another (e.g., As interfering with the reduction of Sb). In addition, the reduction process can be inhibited by oxidants or complexing agents in the matrix or even by some transition metals. The platinum group elements adsorb large amounts of hydrogen, so that the effective reduction ability of the solution is lowered. Standard addition for calibration is therefore strongly recommended.

The volume of the sample must be constant. Foam formation can influence the degassing of the analytes, especially in biological matrices. This should be suppressed.

Conclusion

Atomic spectrometry is suitable for trace concentrations of most elements. The principle is the absorption and emission of light of a certain wavelength after atomization of the sample. The wavelength provides qualitative information about the element, whereas the intensity of the emission (AES) or the quotient of the intensity before and after the absorption (AAS) is proportional to the concentration of the analyte. Several spectral and nonspectral interferences must be considered.

See also: **Aluminum (Aluminium)**: Properties and Determination; **Copper**: Properties and Determination; **Electrolytes**: Analysis; **Iron**: Properties and Determination; **Lead**: Properties and Determination; **Lithium**; **Mass Spectrometry**: Principles and Instrumentation; **Mercury**: Properties and Determination; **Mineral Water**: Types of Mineral Water; **Potassium**: Properties and Determination; **Selenium**: Properties and Determination; **Sodium**: Properties and Determination; **Spectroscopy**: Overview; **Tin**; **Trace Elements**; **Water Supplies**: Chemical Analysis

Further Reading

Dean JR (1997) *Atomic Absorption and Plasma Spectroscopy*. Chichester, UK: John Wiley.

Dedina J and Tsalev D (1995) *Hydride Generation Atomic Absorption Spectrometry*. Chichester, UK: John Wiley.

Ebdon L and Hywel E (1998) *An Introduction to Analytical Atomic Spectrometry*. Chichester, UK: John Wiley.

Hill SJ (1999) *Inductively Coupled Plasma Spectrometry and its Applications*. Sheffield, UK: Sheffield Academic Press.

Jackson KW (1999) *Electrothermal Atomization for Analytical Atomic Spectrometry*. Chichester, UK: John Wiley.

Lajunen LHJ (1992) *Spectrochemical Analysis by Atomic Absorption and Emission*. Cambridge: Royal Society of Chemistry.
Radziuk B and Schlemmer G (1999) *Analytical Graphite Furnace Atomic Absorption Spectroscopy: A Laboratory Guide*. Basel: Birkhauser Verlag.
Sanz-Medel A (1999) *Flow Analysis with Atomic Spectrometric Detectors*. New York: Elsevier.
Sneddon J (2000) *Advances in Atomic Spectroscopy*. Amsterdam: Elsevier.
Sperling M and Welz B (1999) *Atomic Absorption Spectrometry* Weinheim: Wiley-VCH.
Wilson RH (1994) *Spectroscopic Techniques for Food Analysis*. New York: VCH.

Nuclear Magnetic Resonance

A M Gil, University of Aveiro, Aveiro, Portugal

Introduction

The nuclear magnetic resonance (NMR) phenomenon was discovered at Stanford University in 1945 by Bloch, Hansen, and Packard, who detected a proton signal from water and at Harvard University by Purcell, Torrey, and Pound, who observed a signal from the protons of paraffin wax. Bloch and Purcell were jointly awarded the Nobel Prize for physics in 1952 for this discovery.

The observation of an NMR signal from a nucleus requires that the nuclear spin (I) is described by $I \neq 0$. For instance, ^{12}C and ^{13}C have respectively $I = 0$ and $I = 1/2$, hence only ^{13}C is observable by NMR. Associated with the nuclear spin, there is a nuclear magnetic moment (μ), which will interact with an applied magnetic field B_0. This interaction causes the magnetic moment of a nucleus of spin I to take $2I + 1$ possible orientations, each corresponding to an energy level. In the case of $I = 1/2$ (e.g., ^{1}H, ^{13}C, ^{31}P, ^{15}N), two energy levels separated by $\Delta E = h\gamma B_0/2\pi$ are generated, where γ is the nuclear magnetogyric ratio. However, the exact value of ΔE absorbed by a nucleus is also very sensitive to its surrounding molecular environment; hence, by registering the absorbed energy, NMR probes molecular structure and molecular dynamics. Chemical shifts (δ), scalar coupling constants (J) and relaxation times (T_1, T_2) are NMR parameters which carry both structural and dynamic molecular information.

One of the earliest applications of NMR in food science dates from 1957 and concerned the measurement of moisture in foods. However, it is only from the early 1980s that a consistent and widespread growth in the use of NMR in food science has been observed. This resulted from: (1) the increasing sophistication and relative ease of use of NMR instrumentation; (2) the increasing need of the food industry to understand and innovate its products and processes; and (3) the increasing pressure for new methods to enforce legislation and control quality. Many applications have grown in parallel to applications in the biological and biomedical fields, in which the noninvasiveness of the technique is a requirement. In foods, such requirement stems not from ethical considerations but from the need to preserve and examine food structure.

The first review of applications of NMR in food science dates from 1993; the first international conference on the subject was held in 1992 and has been followed by meetings every 2 years since then.

The NMR Techniques for Food Analysis

Foods are generally very complex systems both chemically (containing an enormous variety of compounds differing in nature, size, and quantity) and physically (often being multiphase systems). The choice of NMR method depends primarily on the chemical and physical complexity of the food. For liquid foods like juices, wine, and oil, the method of choice is usually liquid-state high-resolution one-dimensional (1D) and two-dimensional (2D) NMR spectroscopy. Total correlation spectroscopy (TOCSY), heteronuclear correlation spectroscopy (e.g., $^{13}C/^{1}H$ or $^{15}N/^{1}H$) and J-resolved spectroscopy are 2D experiments which give, respectively, information on homonuclear couplings, heteronuclear couplings, and J-values/multiplicity. This information enables chemical identification and structural determination to be carried out, provided the spectra are free from serious signal overlap. In recent years, automated methods have been developed to allow the handling of high numbers of samples with little or no user intervention. Hyphenated techniques have also recently been developed, to handle the high chemical complexity of many foods. These couple NMR with liquid chromatography (LC) and mass spectrometry (MS) in the forms of LC-NMR and LC-NMR/MS.

If the above NMR techniques are applied to solid or semisolid foods, spectra with very broad lines will result, from which structural information can be very difficult to extract. Solid-like foods call for specific methods which enable resolved spectra to be obtained. Magic angle spinning (MAS) aids the observation of dilute nuclei, e.g., ^{13}C, ^{15}N, when coupled with cross-polarization (CP) and high-power decoupling. The CPMAS experiment selects the nuclei present in the solid phase of the food.

Conversely, direct excitation methods select the nuclei in the mobile phase of the food, therefore both CPMAS and direct excitation methods can give complementary information on multiphase foods. But for solid and heterogeneous foods, magnetic resonance imaging (MRI) has been one of the most recent and popular methods. In this technique, the frequency of the NMR signal is dependent on position and, therefore, putting together the information arising from different parts of the sample gives rise to a 2D or 3D representation. The contrast in such image may result from changes in spin density (related to concentration), relaxation times, or diffusion coefficients. A recent novel idea is the attempt to correlate molecular properties, viewed by NMR, with macroscopic properties, e.g., rheological behavior. The method of Rheo-NMR aims precisely at achieving this correlation and its use for foods is now at its infancy. The method registers NMR information (either using imaging or spectroscopy) while stress (shear or extensional) is applied to the sample.

Water in Foods

Water profoundly affects many aspects of food quality: texture, microbiological safety, nutritional status, and digestibility. NMR has been used to probe water activity and water translational mobility in foods. Water relaxation times provide information on the amount and mobility of water. Of the three NMR-active nuclei available in water (^{1}H, ^{2}H in HOD, and ^{17}O in $H_2{}^{17}O$), ^{17}O is the more direct probe since it is not affected by the chemical exchange of protons between solute and water. The different degrees of 'water binding' to solute or biopolymer molecules have been measured in systems as varied as caseinate solutions, potato and wheat starch suspensions, and gelatin gels. Glass transitions in food ingredients like casein, gluten, amylopectin, protein/sugar, and polysaccharide/sugar mixtures are strongly dependent on water activity/content and have important consequences in product quality. Again, the understanding of molecular mobility of water and of the solute, by NMR relaxation, has enabled structural models to be proposed. For instance, relaxation data have suggested a model of a maltose glass in which water molecules undergo rapid rotational and translational motion inside more rigid 'cages' or 'channels' formed by maltose molecules. (*See* **Water**: Structures, Properties, and Determination.)

The effect of food microstructure on water relaxation is important since a distribution of relaxation times occurs and may be used to determine pore size distributions in the food. Many studies have involved model systems like water-saturated porous rocks, sandstones, chalks, and cement pastes. Some studies of actual food materials have been done, trying to handle the additional complexity due to the chemical nature of food components.

At a macroscopic scale, water located in different types of tissue is characterized by different relaxation properties and diffusion coefficients, which result from distinct cell types and sizes. Water is distributed among different subcellular compartments such as vacuoles, starch granules, and the cytoplasm, and again each compartment is characterized by a certain distribution of water relaxation times. These NMR measurements may be used to give information about cell morphology and membrane water permeabilities. Applied to fruits and vegetables, this approach is useful to determine water/air/ice distribution, related to overall quality changes, as the food is ripened, dried, rehydrated, frozen, and stored.

The dependence of relaxation properties on tissue type is the origin of contrast in relaxation or diffusion-weighted MRI of intact fruits and vegetables. MRI has long established its value and potential in food science, not only enabling the qualitative imaging of 'static' food structure but also, and more importantly, following mass and heat transport in foods, noninvasively, and in real time during processing or storage. MRI may be used for temperature mapping of foods, after a suitable calibration of the relaxation times or diffusion coefficients against temperature. The temperature profile of a whole food may be used to deduce thermal diffusivity and surface heat transfer coefficients. Subzero temperature MRI is very useful for the detection of freezing effects, detection of unfrozen water, measurement of freezing times, and of freeze-drying kinetics. MRI is also used to monitor changes at food surfaces, an important application since the efficiency of many processing operations (e.g., baking, drying) depends on the magnitude of heat and moisture transfer rates at the food surface.

Besides providing the moisture and temperature profile of a food, MRI is sensitive to several other quality factors: lipid distribution, solute concentration, pH, solid–liquid ratios, protein and polysaccharide aggregation. Table 1 gives some examples of the applicability of MRI in foods, and a more comprehensive list can be found in the specialized manuals indicated at the end of this text.

Food Biopolymers

Proteins

Both liquid-state and solid-state NMR methods have been of value to elucidate aspects of structure and

behavior of food proteins, such as the milk proteins casein and α-lactalbumin and many cereal proteins. Whereas the wide range of liquid-state NMR methods has been applied to small proteins like α-lactalbumin, seed and storage proteins generally have higher molecular weight and higher heterogeneity and are often insoluble, which renders the usual high-resolution methods rather ineffective. Hence, ^{1}H and ^{13}C solid-state NMR methods have been valuable in characterizing the structure and behavior of large cereal proteins of wheat, barley, maize, and sorghum. These proteins have important functional and nutritional properties which require an understanding at the molecular level, in order to enable suitable quality control. The Rheo-NMR method has already proved promising in the attempt to correlate the molecular structure of hydrated gluten (wheat protein) with its characteristic viscoelasticity. Multinuclear NMR (^{25}Mg, ^{31}P, ^{43}Ca) has also been used to study ion-binding proteins, e.g., β-casein and α-lactalbumin, in order to determine the characteristics of the binding site and of metal competitivity under physiological conditions (*See* **Protein**: Chemistry.)

Table 1 Some examples of magnetic resonance imaging (MRI) applications in foods

Food	*Study*
Apple	Water relaxation in parenchyma tissue; drying; bruising; deterioration during storage
Potato	Water relaxation at subzero temperatures; freeze-drying; drying
Carrot, potato	MRI temperature mapping
Peach	Freezing; bruising
Extruded pasta	Rehydration, radial imaging
Cod	Freezing (texture changes)
Peanut butter/bread	Fat transport
Milk	Cream separation
Food gels	Temperature mapping; spatial heterogeneity; drying
Cereal seeds	Maltose and water distribution during germination
Tomato	Water distribution during ripening
Foams	Water drainage
Biscuits	Water and fat distribution during baking
Meat/fish	Fat and water distribution

Polysaccharides

The interest of the food scientist in polysaccharides includes the issues of dietary components (starch, plant cell wall polysaccharides) and functional ingredients like gelling, thickening, and stabilizing agents. Naturally, high-resolution NMR is a powerful technique for carbohydrate analysis in solution, whereas CPMAS and related solid-state techniques have been applied to solid polysaccharides, in gels and in heterogeneous samples. In solution, NMR can not only provide a measurement of overall composition and detailed conformational structure, but also be sensitive to fine details such as degree of esterification, degree and type of substitution, residue sequence, and average block length. Table 2 shows some examples of studies of several food polysaccharides in solution (*See* **Carbohydrates**: Classification and Properties.)

The knowledge of the physical properties of complex carbohydrates is extremely useful for food scientists. Such information is available in several databases, including an NMR database. Processes like gelation, crystallization, starch gelatinization, and retrogradation are extremely important in determining the quality of a polysaccharide-containing food. Cellulose and starch provide excellent examples of solid polysaccharides from which extensive structural information may be obtained by solid-state NMR. ^{13}C CPMAS NMR clearly distinguishes the A (from cereals) and B (from tubers) starch polymorphs and can also measure the proportions of crystalline and amorphous components. Starch

Table 2 Examples of food polysaccharides characterized by high-resolution nuclear magnetic resolution (NMR)

Polysaccharide		*Study*
Pectins	^{1}H and ^{13}C NMR	Degree of esterification; characterization of monomer sequences and average chain lengths; galacturonic acid content; conformational analysis
Guar, locust bean gum	^{1}H and ^{13}C NMR	Detection of substituted mannose residues; characterization of monomer sequences
Alginates	^{1}H and ^{13}C NMR	Determination of average block lengths; effects of calcium binding
Amylopectins	^{1}H and ^{13}C NMR	Ratio of α(1-4) and α(1-6) linkages
Wheat arabinoxylans	^{1}H and ^{13}C NMR	Degree and type of substituted xylose rings; correlation between structure and gelling capacity/thermal stability
Agarose, κ, ι, λ, β-carrageenans	^{1}H, ^{13}C, ^{87}Rb, ^{23}Na, ^{39}K, ^{133}Cs NMR	Study of coil–helix transitions;strength of metal binding
Xanthan	^{1}H and ^{13}C NMR	Acetate and pyruvate contents; study of coil–helix transition
Gellan gum	^{1}H and ^{13}C NMR	Structural determination
Starch	^{1}H and ^{13}C NMR	Amylose content; degree of amylopectin branching; interaction with metals

gelatinization occurs upon heating up to 70 °C, causing starch granules to absorb water and burst, leading to the formation of a gel. On subsequent cooling, starch crystallization or retrogradation occurs. Retrogradation is the main cause for bread staling and, therefore, it has attracted a lot of interest from the food scientific community. A variety of ^{13}C CPMAS, ^{1}H MAS, and relaxation times measurements have been coupled to follow and understand both the gelatinization and the retrogradation processes, including the understanding of the role of each component, amylose and amylopectin, in the process. ^{31}P NMR can also be used to view phosphate groups which are found in starch molecules, believed to contribute to the clarity and viscosity of starch pastes.

Many gelling processes of a variety of food carbohydrates (starch maltodextrins, carrageenans, xanthan, pectins, scleroglucan, konjac glucomannan, methyl and hydroxymethyl cellulose) have been characterized, generally by ^{13}C, ^{1}H, and in specific cases by ^{127}I, ^{23}Na, ^{87}Rb, ^{133}Cs NMR. Such studies give information as varied as solid-to-liquid ratios, identification of different water populations, degrees of metal binding, characteristics of coil–helix transitions, and degree of aggregation. Plant cell wall polysaccharides, e.g., pectins, are a major source of fiber and determine the texture of plant foods. Much NMR work has been applied to acquire knowledge of the structure, composition, and function of these polymers, both in their natural environment – the plant – and as food additives.

Analysis and Authentication of Foods by NMR

The high cost of NMR spectrometers and the relatively low sensitivity of the technique compared to competing methods has prevented the adoption of the technique as a routine analytical tool in industry. A very important exception to this is the low-resolution 'benchtop' NMR spectrometers which have found numerous uses in food industry as well as in other industries. However analytical applications of high-resolution NMR to lipids, sugars, and many other compounds still remain nonroutine. In the last few years, high-resolution NMR has been coupled to LC and to MS techniques, making way for a new generation of hyphenated techniques with great potential in rapidity, sensitivity, and precision. But, one high-resolution NMR technique has alone justified the costs of sophisticated equipment enabling the achievement of information of a unique nature: the site-specific natural isotope fractionation NMR (SNIF-NMR) technique.

Low-resolution Pulsed NMR

Low-resolution NMR methods involve the extraction of information directly from the magnetization decay, and not from the NMR spectrum. Usually the nucleus observed is ^{1}H but others may be used, e.g., ^{2}H, ^{13}C, ^{31}P. The rates of transverse magnetization decay (or T_2 relaxation rates) following a 90° pulse are very different for solid and liquid phases and a typical determination which has been extensively explored is the determination of solid fat content (SFC) of semisolid fats, not requiring any sample preparation. Absolute measurements of oil or moisture content are based on the same principle but require sample weighing and a calibration of the method with an independent method. When species with similar relaxation times are present, e.g., oil and water in some oil-seeds, the Carr–Purcell–Meiboom–Gill (CPMG) experiment is used and appropriate software used to obtain relaxation times and relative proportions of the components of interest. These low-resolution methods are noninvasive, involve short measuring times (a few seconds), and require little sample preparation. They are therefore ideally suited for quality control of large numbers of samples and this has led to their extensive industrial use, supported by the development of dedicated, low-cost benchtop instruments. Existing online applications include determination of oil and moisture contents in seeds, determination of SFC of fats, use of relaxation times to follow cake cooking, nondestructive detection of fruit quality (as sugar content), and the determination of phospholipid content in peas.

High-resolution NMR and Hyphenated Techniques

The applications of high-resolution NMR in food science have been reviewed in 1996 and, since then, the use of 2D NMR for structural analysis of complex natural products, including minor food components, has become widespread. The qualitative applications are numerous and many quantitative determinations have been explored.

One of the recognized applications of ^{1}H, ^{13}C, and ^{31}P high-resolution NMR has been the study of lipid structure and composition. The quantitative determination of lipids by NMR has been carried out extensively. For example, for palm oil, composed of triacylglycerol mainly of palmitic (P), oleic (O) and linoleic (L) acids, the mole fractions of P, O, and L may be obtained from integration of particular signals in the ^{13}C NMR spectrum and their positional distribution may be determined. Applications have included determination of saturated and unsaturated fatty acids in edible oils and spreads, differentiation of different grades of olive oils, detection of synthetic

oils in mixtures of virgin olive oil, determination of ω-3 acids in fish samples, the study of the oxidative deterioration of several oils, and the effects of storage and thermal stressing, structural determination of pigment compounds and antioxidants in oils. Interestingly, the determination of oil in intact seeds has been made possible by high-resolution MAS NMR, and sunflower and rapeseeds are examples.

Quantitative ^{13}C NMR has been applied to many other systems, e.g., quantification of disaccharides in honey and quantification of sugars and sugar derivatives in wine. However, ^{1}H high-resolution NMR has gradually proved its usefulness too. Up to the early 1990s ^{1}H NMR had received little attention as a tool for food analysis, mainly because of the great complexity of the spectra and reduced spectral dispersion, as well as the need for efficient suppression of the very high water signal. In recent years, water suppression efficiency has improved largely and the use of higher field instruments (> 400 MHz) has caused a renewed interest for the analysis of complex food mixtures like fruit juices (Figure 1), coffee, and wine by ^{1}H NMR. Indeed, at higher fields a wealth of signals may be detected in all regions of such spectra, besides the main solute signals (sucrose, glucose, fructose, citrate, etc.). Such spectra often show great spectral complexity and signal overlap, even in 2D (Figure 2) and, in some cases, multivariate analysis has proved useful for spectral interpretation. It is in the disentanglement of the information present in high-field ^{1}H NMR spectra that hyphenated techniques play a determinant role. Methods such as LC-NMR or LC-NMR/MS enable NMR, LC, and MS data to be collected for the same sample in relatively short experimental times (min–1 h). The usual set-up of such experiments is represented in Figure 3. These techniques present a valuable aid in cases of metabolites which, because of their low abundance or particular chemical nature, cannot be identified using solely 2D NMR techniques. Along with NMR parameters, the coupled techniques add information about absolute or relative retention times and molecular mass. These developments show the potential for rapid detection and characterization of minor compounds in foods, many with important roles in aroma/flavor chemistry or with potential interesting properties such as antioxidant or preservative properties.

Site-Specific Natural Isotope Fractionation Nuclear Magnetic Resonance

SNIF-NMR is certainly a success story in the use of NMR for the practical needs of the food industry. The technique has been developed based on the fact that the percentage of heavy isotopes occurring for a given element (hydrogen, carbon, oxygen) in natural products is not a fixed quantity but depends on the history of the sample, e.g., geographical, climatic, and biochemical factors. The differences arise

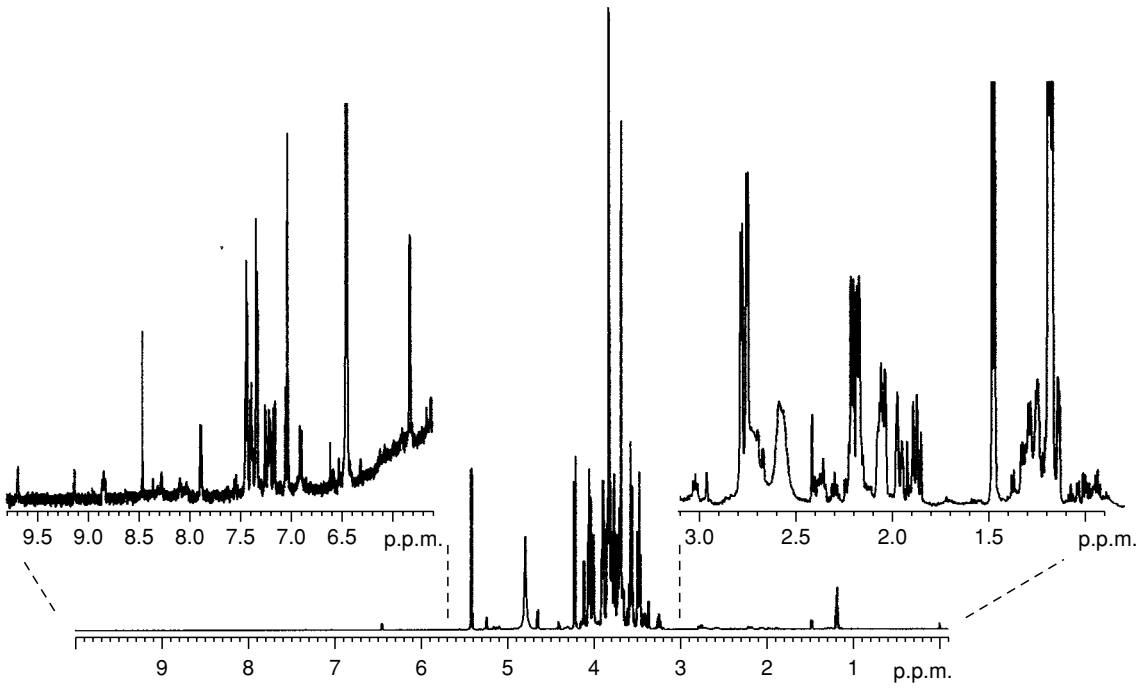

Figure 1 ^{1}H nuclear magnetic resonance (NMR) spectrum of mango juice, at 600 MHz; 64 scans.

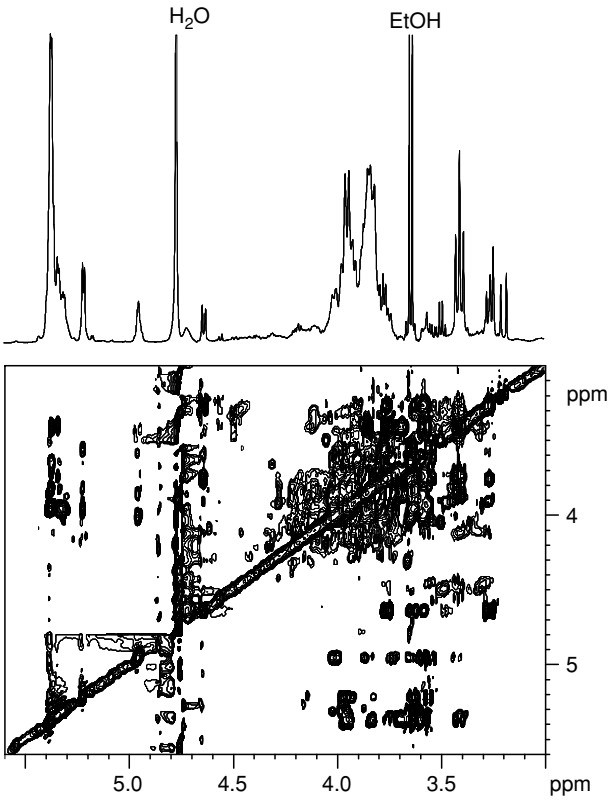

Figure 2 Total correlation spectroscopy (TOCSY) spectrum of a beer sample, at 500 MHz.

from distinct isotopic composition of the starting materials (e.g., rainwater at different latitudes) and from fractionation effects occurring during bioconversions. Measurement of isotope ratios by MS is widely used for assessment of the authenticity of foods but this technique gives an average isotope ratio and no information on the intramolecular distribution of the isotopes. SNIF-NMR uses ^{2}H NMR to measure the deuterium-to-hydrogen isotope ratios at specific sites (i) in the molecule, $(D/H)_i$. The analysis and characterization of wines and other alcoholic drinks were among the first applications of SNIF-NMR. Such samples are examined as distillates with about 95% ethanol. The proton-decoupled ^{2}H NMR spectrum of an ethanolic distillate consists of four resonances: from CH_2DCH_2OH (1), CH_3CHDOH (2), CH_3CH_2OD (3) and HDO (the latter two collapse into one single resonance under conditions of fast exchange). From the intensities (or areas) of the peaks of 1 and 2 it is possible to obtain the ratio $(D/H)_2/(D/H)_1$ which varies for alcohols from different sources, reflecting the occurrence of strong site-specific fractionation effects. $(D/H)_3$ is related to $(D/H)_{water}$ and may be obtained through the measurement of $(D/H)_{water}$ by NMR or by MS. Furthermore, absolute D/H ratios may be obtained using a working standard with a known isotope ratio.

Ethanols from different plants (sugar cane, corn, wheat, barley, apple, grape, sugar beet, and potato) have been distinguished using differences in the $(D/H)_1$ ratio. These results have been used to detect the origin of raw materials used to make commercial spirits (whiskey, vodka, gin, rum, and fruit brandies). The characterization of wines has involved additional systematic experiments to relate the (D/H) ratios of the starting materials (starting water and glucose) with those of the end products, ethanol and water. By showing the connection between isotope ratios of starting and end products, it has been shown that the ratios for ethanol and water can be used as 'fingerprints' in the characterization of wines. SNIF-NMR parameters in tandem with statistical analysis have performed remarkably in distinguishing wines of different origins, according to varieties, geographical origins, and year of production. Distinction in terms of country of origin reflects differences in temperature and rainfall figures. SNIF-NMR became the most efficient technique to detect enrichment of wines with exogenous sugars and watering of wines. Other applications relate to the determination of origin of flavors and fragrances (natural or synthetic),

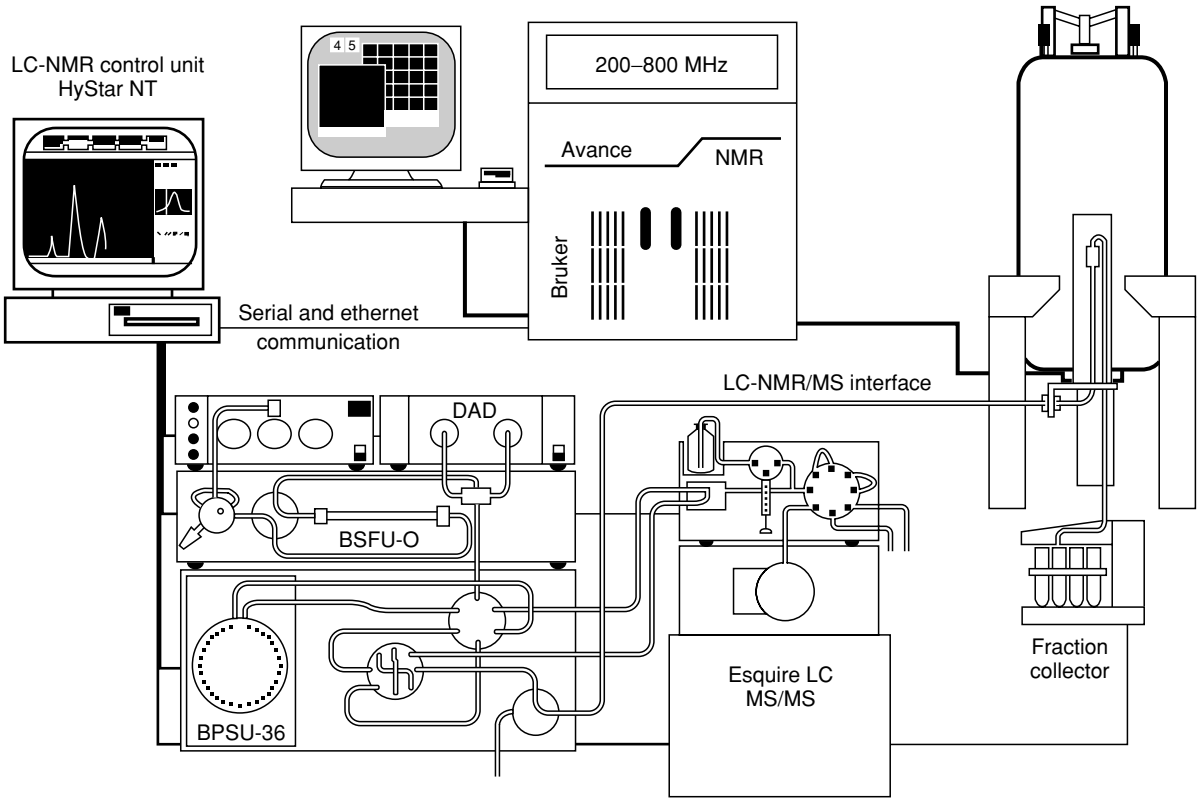

Figure 3 Set-up for the LC-NMR/MS technique. Courtesy of Bruker BioSpin, Germany.

vinegars, fruit juices and jams (detection of beet sugar addition), almond and cinnamon oils (detection of synthetic benzaldehyde).

Other Foods

Many additional studies have been devoted to other foods, e.g., milk, meat, spices and phytochemicals, tea, coffee, wine, fruits, and vegetables. These studies often involve an interplay of high-resolution NMR, solid-state NMR, and imaging methods. Additional applications to those already mentioned are as varied as predicting sugar content in intact fruits, analyzing pesticide residues in vegetables, determining the fluorine and aluminum contents in intact tea leaves, and analysis of the insoluble deposits in bottled wine. A comprehensive description of the enormous variety of studies of foods by NMR may be found in many reviews, some of which are indicated below. It was the aim of this text to give the reader a flavor of the versatility of NMR spectroscopy, adaptable to different physical states and to the chemical complexities of foods. Its applications now extend in new directions, such as online applications and the study of nutritional issues. Scientific growth is continuing in this field and, increasingly, industrialists and scientists are coming together to share practical problems and scientific knowledge.

See also: **Adulteration of Foods**: Detection; **Analysis of Food**; **Carbohydrates**: Interactions with Other Food Components; **Drying**: Physical and Structural Changes; **Freeze-drying**: Structural and Flavor (Flavour) Changes; **Freezing**: Structural and Flavor (Flavour) Changes; **Protein**: Determination and Characterization; Functional Properties; **Spectroscopy**: Near-infrared; Infrared and Raman; **Starch**: Structure, Properties, and Determination; **Water**: Structures, Properties, and Determination; **Water Activity**: Principles and Measurement

Further Reading

Ashurst PR and Dennis MJ (1998) *Analytical Methods of Food Authentication*. London: Blackie.

Belton PS, Colquhoun IJ and Hills BP (1993) Applications of NMR to food science. *Annual Reports on NMR Spectroscopy* 26: 1–53.

Callaghan PT (1991) *Principles of Nuclear Magnetic Resonance Microscopy*. Oxford: Oxford Science Publications.

Gidley MJ, McArthur AJ, Darke AH and Ablett S (1995) High resolution nuclear magnetic resonance spectroscopy of solid and semi-solid food components. In:

Dickson E (ed.) *New Physico-chemical Techniques for the Characterization of Complex Food Systems*, pp. 296–318. Glasgow: Blackie.

Gil AM, Belton PS and Hills BP (1996) Applications of NMR to food science. *Annual Reports on NMR Spectroscopy* 32: 1–49.

Gonzalez J, Jamin E, Remaud G, Martin YL, Martin GG and Martin ML (1998) Authentication of lemon juices and concentrates by a combined multi-isotope approach using SNIF-NMR and IRMS. *Journal of Agricultural and Food Chemistry* 46: 2200–2205.

Harris RK (1986) *Nuclear Magnetic Resonance Spectroscopy – A Physicochemical View*. Essex: Longman Scientific & Technical.

Hills B (1998) *Magnetic Resonance Imaging in Food Science*. New York: Wiley.

Hills BP, Manning CE and Godward J (1999) A multistate theory of water relations in biopolymer systems. In: Belton PS, Hills BP and Webb GA (eds) *Advances in Magnetic Resonance in Food Science: Proceedings of the Fourth International Conference on Applications of Magnetic Resonance in Food Science*. Cambridge: RSC.

Kauten R and McCarthy M (1993) Applications of NMR imaging in processing of foods. In: *Food Processing: Recent Developments*, pp. 1–22. Amsterdam: Elsevier.

Linden G (1996) *Analytical Techniques for Foods and Agricultural Products*, Cambridge: VCH.

McCarthy MJ (1994) *Magnetic Resonance Imaging in Foods*. New York: Chapman and Hall.

McDonald PJ (1995) The use of nuclear magnetic resonance for on line process control and quality assurance. In: Gaonkar AG (ed.) *Food Processing: Recent Developments*, pp. 23–36. Amsterdam: Elsevier.

Mossoba MM (1998) *Spectral Methods in Food Analysis: Instrumentation and Applications*. New York: Marcel Dekker.

Ruan RR and Chen PL (1998) *Water in Foods and Biological Materials: A Nuclear Magnetic Resonance Approach*. Lancaster, PA: Technomic.

Sanders JKM and Hunter BK (1993) *Modern NMR Spectroscopy – A Guide for Chemists*. Oxford: Oxford University Press.

Visible Spectroscopy and Colorimetry

T Chiba, Iwate Prefectural University, Takizawa, Japan

Background

Visible spectroscopy and colorimetry are forms of spectroscopic analytical methods. They are based on the measurement of the absorption of visible light and can provide qualitative and quantitative information about a sample.

Visible light consists of electromagnetic radiation in range wavelengths from 380 to 780 nm, to which the human eye is sensitive. In this part of the spectrum, radiation with different wavelengths gives rise to light with different colors, while a mixture of light from these wavelengths forms white light. The color of a substance depends on the wavelength of light it absorbs. Thus, when a sample is exposed to white light, certain wavelengths are absorbed, and the remaining wavelengths are transmitted to the eye. The color perceived by the eye is determined only by the wavelength transmitted. In simple terms, the color seen is the complementary color to the color(s) absorbed (**Table 1**). The intensity of color relates to the amount of light absorbed by a substance. The wavelength and efficiency of absorption by a substance depend on the structure of the substance and its environment, making it possible to measure the presence or concentration of the substance.

Identification or determination of a substance based on its color is probably one of the oldest methods of examination in analytical chemistry. Earlier work was entirely empirical and led to qualitative or semiquantitative estimation rather than quantitative determination. The hues and intensity of colors were measured by visually comparing the color of a sample with that of several permanent color standards, which may have been samples containing the same substance at known concentrations or colored glass. These methods, with the human eye as a detector, inherently include a subjective perception of color and are dependent on the reproducibility of the observer's eye. Moreover, even the best observer has difficulty comparing the intensity of two colors with slightly different hues.

The development of detectors of radiation together with the general advance of instrumentation has resulted in improvements in measurement techniques. Photometric techniques with a high level of sensitivity and reliability are available. With these techniques, an instrument called a photometer is used as a photoelectric detector instead of the human eye and converts the intensity of light to an electrical current. Recent photometers are spectrophotometers capable of splitting the incident light into different wavelengths and measuring the intensity of that radiation. Spectrophotometric measurements are an important tool for field testing, industrial applications, and research departments. The term colorimetry describes the measurement by the human eye, and colorimetric methods have been largely replaced by spectrophotometric methods.

Table 1 Colors of light

Absorbed wavelength (nm)	*Absorbed color*	*Transmitted color (complement)*
400–435	Violet	Yellow–green
435–480	Blue	Yellow
480–490	Blue–green	Orange
490–500	Green–blue	Red
500–560	Green	Purple
560–580	Yellow–green	Violet
580–595	Yellow	Blue
595–620	Orange	Blue–green
620–750	Red	Green–blue

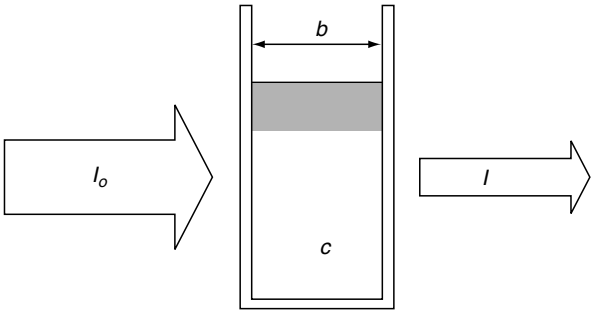

Figure 1 Absorption of radiation by a sample. I_o, intensity of incident light; I, intensity of transmitted light; b, path length of a sample; c, concentration of a sample.

Colorimetric methods (visual methods) are still utilized for fieldwork and simple determinations such as monitoring the pH and chlorine levels of swimming pools. The apparatus required for these methods is simple and inexpensive. The sensitivity and precision are often satisfactory for particular determinations.

Principles

Beer–Lambert Law

Electromagnetic radiation can be characterized as a wave with frequency, v, and wavelength, λ. This wave has energy, E, is proportional to its frequency and is given by the equation:

$$E = hv = hc/\lambda,$$

where h is Plank's constant, and c is the velocity of light. If such a wave of radiation encounters a substance, and the energy is absorbed, a molecule of the substance is promoted to an excited state. The absorption of visible light generally excites electrons of a molecule from a ground electronic state to an excited electronic state.

The amount of light absorbed by a substance in a solution is quantitatively related to its concentration. This relationship is expressed in two fundamental laws: the Bouguer–Lambert law, which is also known as Lambert's law, relates the amount of light absorbed to the length of the path it travels through a medium containing an absorbing substance; Beer's law relates light absorption to the concentration of the absorbing substance. These two laws are combined as the Beer–Lambert law, or simply called Beer's law, and are written as the equations (Figure 1):

$$\log_{10}(I_0/I) = -\log_{10} T = abc = A$$

or

$$I_0/I = 1/T = 10^{abc},$$

where I_0 is the intensity of incident light, I is the intensity of transmitted light, T is the transmittance, b is the path length through an absorbing solution, c is the concentration of the absorbing substance, and A is the absorbance (extinction or optical density in older literature).

The constant a, which is called the absorptivity or the extinction coefficient, specifies a characteristic property of the absorbing substance and is a function of its wavelength. Its units depend on the units of concentration and length of the path employed. When c is expressed in moles per liter, and b is expressed in centimeters, the constant is known as the molar absorptivity, formerly called the molar absorption or molar extinction coefficient. The molar absorptivity is used as a physical constant for absorbing species under standard conditions. It is in liters per mole per centimeter ($\mathrm{l\,mol^{-1}\,cm^{-1}}$) and is given a symbol, ε.

These expressions show that there is a linear relationship between the absorbance and the concentration of a given solution, if the length of the path and the wavelength of light are kept constant. By measuring the transmittance or absorbance, the concentration of a substance in a solution can be calculated.

Deviation from the Beer–Lambert Law and Errors in Determination of Concentration

The Beer–Lambert law is not necessarily obeyed because the law applies strictly to instances when incident radiation is monochromatic, and when the absorbing centers act independently of one another regardless of the number and kind. In practice, there is usually a deviation from a linear relationship between the concentration and absorbance as with a concentrated solution (> 0.01 M) of most substances. These deviations occur as the result of chemical and instrumental factors. Chemical factors can be caused by chemical equilibria with the solution, intermolecular

interaction, reagent and chromophore stability, and impurities that fluoresce or absorb at the absorption wavelength of incident light. Instrumental factors are associated with instrumental limitations determined by the quality of the instrument's optical, mechanical, and electrical systems. These deviations can produce errors in application of the law for practical purposes, but this problem can be overcome by preparing calibration curves that show the experimental relationship between absorbance and simple concentration used for preparing a calibration curve, provided this relationship is always followed under analytical conditions.

There are also several sources of error in measuring transmittance and absorbance. These sources of error include a faulty readout from a detector, incorrect wavelength settings, improper handling of cells, and faulty sample preparation. Any measurement of transmittance or absorbance is more or less subject to error.

The magnitude of instrumental errors, in relation to the real value of the concentration, changes with the transmittance measured. The relative error is high at both very high and very low values of transmittance and lowest over the transmittance range of 20–65%. It is advisable to work in the transmittance range 10–90% (absorbance 1.0–0.1) to prevent large errors.

Methods of Operation and Instrumentation

It may be suitable to use a classification where the two methods are distinguished by the means of detecting light: the visual method, as judged by the human eye, and the spectrophotometric method, as detected by a photoelectric detector.

Visual Method

This method is a visual comparison of samples with standards of known concentration observed in ambient light by the human eye. Therefore, a successful result requires that the sample be the same color as the standard, with only the intensity of the color differing.

Dilution method (duplication method) This is performed by matching color intensity by adding water or a preferred diluent to a darker solution until the two match in terms of color intensity. A standard and a sample solution are placed in a colorless glass tube of equal thickness with a flat bottom, called a Nessler tube. Light from a source passes through each tube and is compared. The more concentrated solution is progressively diluted until the light emerging from both tubes is equal in intensity when viewed crosswise and protected from side illumination. At this point, the concentrations of the two solutions are equal, and the contents of the solutions are related to each other by volume.

Series-of-standards method A sample solution is compared with a series of standards consisting of colored glasses or solutions of known concentration prepared in the same way as the sample solution. Nessler tubes or test tubes are used. A series of tubes of the standard exhibiting a regular gradient of color are placed in a rack. The tube of a sample solution is compared side by side with the standard until a match is found or until the concentration is determined to lie between that of two standards.

Spectrophotometric Method

This is a more sophisticated method and uses a spectrophotometer, which is an instrument for measuring the intensity of light of various wavelengths transmitted by a sample. The intensity of light is determined by photoelectric detectors, which convert radiation energy to electrical energy and can eliminate the need for subjective measurement by the human eye. This method has a number of advantages over visual methods: the limit of detection is lowered by measuring the absorption of a solution at the maximal wavelength; it is possible to avoid or minimize the effect of foreign colored substances by working at a suitable wavelength; a greater precision can be obtained with spectrophotometric methods than with visual methods.

A spectrophotometer consists essentially of (1) a light source; (2) a monochromator for furnishing monochromatic light; (3) one or more sample containers; (4) a detector for measuring the intensity of transmitted light; and (5) a signal processing and readout device (Figure 2).

A common light source is a tungsten filament lamp. This lamp produces continuous and intense radiation in the range of 320–3000 nm and is adequate for measurement over the visible region. Devices used as monochromators include diffraction gratings and glass prisms. A diffraction grating is a superior form of general device, giving a constant and narrow radiation bandwidth. Various types of optical filters may be used as a monochromator, although they provide only a limited wavelength while a different filter is necessary for each wavelength required. Such an instrument is known as a filter photometer. Sample containers are usually called cells or cuvets. They must be transparent in the visible region. General cells are commonly rectangular in shape with a path length of 10.0 mm and are made of optical-quality

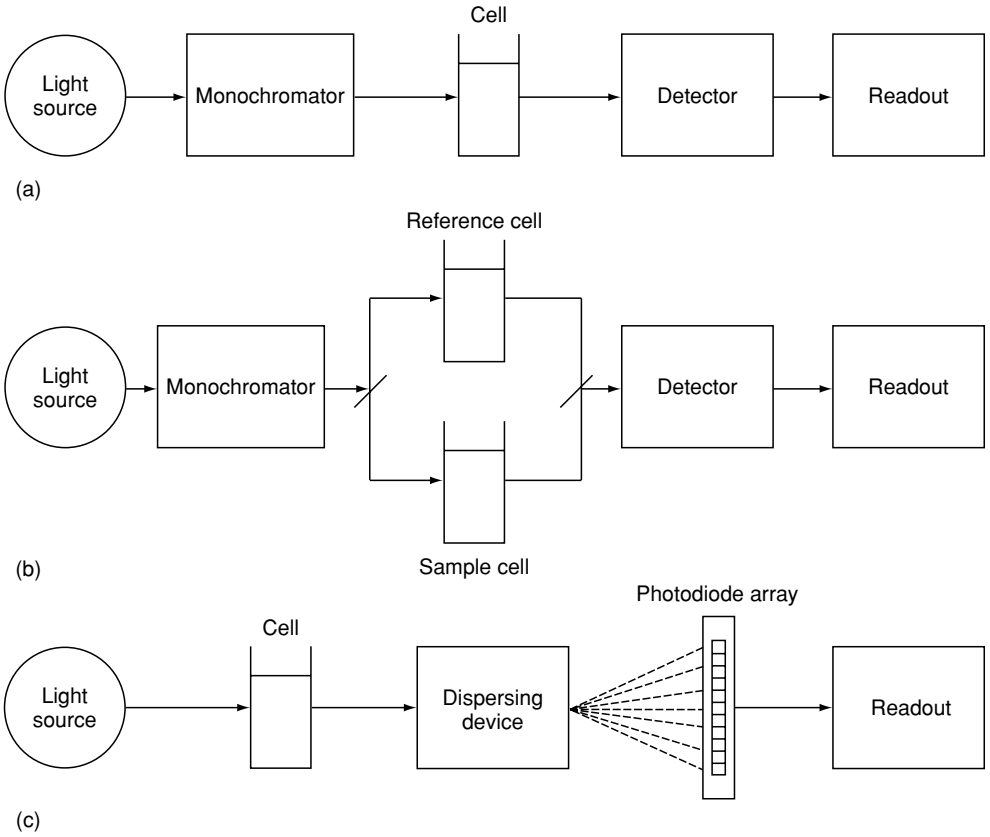

Figure 2 Designs of various types of instruments for visible spectroscopy: (a) single-beam instrument; (b) double-beam instrument; and (c) multichannel instrument with a photodiode array detector.

glass or plastic. There are several types of cells available from commercial sources: a microcell for small volumes of solutions, a stoppered cell for volatile solutions, and a flow cell for measuring the absorbance of flowing solutions. The most common detectors are photomultiplier tubes or phototubes. Photodiode arrays, which are solid-state detectors, are also used in newer spectrophotometers. The electrical current produced in these detectors is amplified and plotted on a recorder or displayed on the spectrophotometer as either absorbance or percentage transmittance.

Measurements are made by inserting a cell containing a sample into the light path of the spectrophotometer. The concentration of a substance can be determined either from a calibration curve or from a calculation based on the Beer–Lambert law through the use of an instrumental readout.

Commerical spectrophotometers are classified into two broad categories, single- and double-beam systems, depending on their optics. The individual components of the system have the same function.

A basic spectrophotometer is a single-beam instrument, which has only one light path for radiation (Figure 2a). This instrument is simple and inexpensive, but is somewhat troublesome to measure with because a sample and a reference (or a reagent blank) must be placed alternatively in the light path, and for measurement at different wavelengths, set-up procedures are required at each wavelength. It is well suited for quantitative measurements at a single wavelength.

The double-beam system provides two equivalent paths, both originating with the same light source (Figure 2b). One of them travels through a sample cell, and the other passes through a reference cell. The two beams are measured separately, either by two matched detectors or by rapidly alternating use of the same detector. The ratio of the intensity of the two beams can be obtained by this system. Thus, the double-beam instrument automatically compensates for absorbance by the reference, as well as for any variation of source intensity with time and wavelength. This instrument provides more accurate quantitative measurement of transmittance or absorbance than a single-beam instrument. There are also spectrophotometers equipped with an automatic drive for scanning the wavelength. An automatic recording

of absorbance as a function of wavelength permits an absorption curve (an absorption spectrum) to be obtained. The spectrum is useful for the qualitative identification of a substance and selecting a suitable wavelength for quantitative analysis. It may also reveal the presence of impurities that are interfering in determination.

A conventional spectrophotometer simultaneously measures transmittance or absorbance only at a single selected wavelength. However, a modern spectrophotometer is a multichannel instrument, which can monitor any selected wavelength range in the visible region. This spectrophotometer is usually a single-beam system with a photodiode array as a detector (Figure 2c). The photodiode array is designed with a linear array that has a series of photodiodes arranged side by side. Each photodiode responds to visible light. If dispersed light in its component wavelength is focused on the surface of the array, the individual photodiode measures the intensity of radiation from a particular section of the spectrum. The data from photodiodes are then accumulated and processed by a computer. This instrument allows the immediate and continuous recording of an entire spectrum without mechanical wavelength scanning. This feature is very useful for the analysis of mixtures of absorbing species and kinetic studies.

Spectrophotometers with a flow cell are often used as detectors in analytical systems and subsequently have extended the application of visual spectroscopy. When linked with a turnable type of sampler and a recorder, they expedite manual procedures and read large batches quickly. By directly observing eluate from a chromatographic column with this detector or, if the substance to be measured is colorless, after a color-developing reaction, several constituents can be concurrently determined by high-performance liquid chromatography. The photodiode array spectrophotometer is an effective detector for this purpose, since it can obtain essentially continuous spectra of the eluate and produce an absorption spectrum for any point on the chromatogram.

There are several types of automated analyzers equipped with spectrophotometric detectors in which analytical procedures are entirely automated to handle large numbers of analyses and routine work, especially for clinical specimens.

A modern spectrophotometer, incorporating a computer for instrument control, data collection, and data processing, provides derivative determinations using derivative absorption spectra and dual-wavelength, absorption difference methods by simultaneous measurements at two different wavelengths. These methods lead to analytical approaches that increase both the sensitivity and selectivity.

New absorptiometric techniques, photoacoustic and thermal-lens spectroscopy, now allow trace amounts of light absorption to be measured. Photoacoustic spectroscopy is based on the detection of thermal modulations from a light-absorbing sample as pressure waves in gas surrounding a sample. Thermal-lens spectroscopy utilizes a thermal-lens effect where, when a laser beam is introduced into a sample, local heating along the beam leads to a change in the refractive index in the vicinity of the beam. These highly sensitive techniques can accurately detect an extremely small absorption, even lower than 0.001, by increasing the laser's power.

Analytical Applications

A visible spectroscopic method is one of the most powerful and widely used procedures for quantitative analysis. There are many applications for inorganic, organic, biochemical, and food constituents. However, qualitative applications of this method are restricted, since the spectra of most substances in the visible region are simple and consist of one or a few broad peaks with no fine structure. Usually, this method is combined with a more useful quantitative technique like nuclear magnetic resonance or mass spectrometry.

Derivatization

Naturally colored substances can be directly analyzed, while most other substances, which are colorless or weakly colored, must be converted to an intensely colored derivative. The reaction for color formation may include enzymatic conversion or chemical modification of a substance to produce colored products and a complex formation between a substance and a color-forming reagent. Color-forming reagents have been developed and are now available for analysis of many substances including inorganic and organic constituents.

Specific Examples

Protein (Folin–Lowry method) This is widely used for the measurement of protein and consists of two steps: the reaction of a protein with copper ions in an alkaline solution and the reduction of a Folin–Ciocalteu reagent by copper-treated protein containing tyrosine and tryptophan residues.

A carbonate-copper solution (0.1 ml) is added to a sample solution (0.2 ml) in a test tube containing 5–100 μg of protein. The carbonate–copper solution is prepared by combining 2% sodium carbonate in 0.1 N sodium hydroxide (50.0 ml) and 0.5% hydrated copper sulfate in a 1% potassium tartrate

solution (1.0 ml). The sample solution is mixed vigorously and allowed to stand for 10 min at room temperature. Folin–Ciocalteu reagent (0.1 ml) (commercially available) is added and mixed. After 30 min, the absorbance of the sample solution is measured at 750 nm. The protein content is calculated from a standard curve prepared from bovine serum or egg white albumin.

Glucose (glucose oxidase method) This procedure, taking advantage of the specificity of glucose oxidase, is based upon the conversion of glucose to gluconic acid and hydrogen peroxide by glucose oxidase and the subsequent oxidation of *o*-dianisidine to its oxidized form, measurable at 420 nm, by hydrogen peroxide with peroxidase.

A sample solution (0.5 ml), containing 0–50 μg of glucose, is mixed with glucose oxidase reagent (3.0 ml) in a test tube and incubated in a water bath at 37 °C for 60 min. The glucose oxidase reagent is prepared by creating a mixture of glucose oxidase (125 000units), horse-radish peroxidase (0.5 mg), and 1% *o*-dianisidine in 95% ethanol (0.5 ml) to 100 ml with 0.5 M sodium phosphate buffer, pH 7.0, and, if necessary, filtering the solution. After incubation, the absorbance at 420 nm is measured against a blank. The sample content is then calculated from a standard curve.

Iron (di(2-pyridyl)-*N*,*N*-di[(8-quinolyl)-amino]-methane method) Di(2-Pyridyl)-*N*,*N*-di[(8-quinolyl)amino]methane (DPQAM) is a selective chelating reagent for iron(II), and under suitable conditions, the green 1:2 metal ligand chelate formed can be used to determine iron in an aqueous solution and, by solvent extraction, in the presence of large amounts of foreign metal ions.

An aliquot of the sample solution (10–150 ml), containing 2.5–25 μg of iron(II) and iron(III), is placed in a separating funnel. A 5% ascorbic acid solution (1.0 ml) and 0.84 M acetic acid buffer, pH 4.7 (5.0 ml), are added and extracted with one 10-ml aliquot of 0.1% DPQAM in chloroform. The mixture is then shaken vigorously and transferred into a 25-ml flask containing anhydrous sodium sulfate. The absorbance of the chloroform extract is measured against distilled water at 693 nm, and the calibration graph is prepared using standard solutions of iron (II) treated in the same way.

See also: **Chromatography**: High-performance Liquid Chromatography; **Spectroscopy**: Overview

Further Reading

Dahlqvist A (1961) Determination of maltase and isomaltase activities with a glucose-oxidase reagent. *Biochemical Journal* 80: 547.

Escobar R and Cano Pavon JM (1983) Selective spectrophotometric determination of trace amounts of iron with di(2-pyridyl)-*N*,*N*-di[(8-quinolyl)amino]methane: Determination of iron in blood serum. *Analyst* 108: 821.

Harris DC (1998) *Quantitative Chemical Analysis*, 5th edn, p. 511. New York: W.H. Freeman.

Hargis LG, Howell JA and Sutton RE (1996) Ultraviolet and light absorption spectrometry. *Analytical Chemistry* 68: 169R.

Holme DJ and Peck H (1998) *Analytical Biochemistry*, 3rd edn, p. 36. New York: Longman.

Howell JA (1997) Ultraviolet and visible molecular absorption spectrometry. In: Settle FA (ed.) *Handbook of Instrumental Techniques for Analytical Chemistry*, p. 481. Englewood Cliffs, NJ: Prentice Hall.

Howell JA and Hargis LG (1994) Ultraviolet and light absorption spectrometry. *Analytical Chemistry* 66: 445R.

Howell JA and Sutton RE (1998) Ultraviolet and light absorption spectrometry. *Analytical Chemistry* 70: 107R.

Long ME, Swofford RL and Albrecht AC (1976) Thermal lens technique: A new method of absorption spectroscopy. *Science* 191: 183.

Lowry OH, Rosebrough NJ, Farr AL and Randall RJ (1951) Protein measurement with the folin phenol reagent. *Journal of Biological Chemistry* 193: 265.

Thomas MJK (1996) *Ultraviolet and Visible Spectroscopy*, 2nd edn. Chichester, UK: John Wiley.

SPICES AND FLAVORING (FLAVOURING) CROPS

Contents

Use of Spices in the Food Industry

L A Wilson, Iowa State University, Ames, IA, USA

Introduction

Spices and herbs are valuable commodities that have played a major role in world civilization; wars were fought, governments and their economies altered, cultures enriched, legends created, and new worlds discovered. Many of the great chronologers and explorers – Hippocrates, Theophrastus, Marco Polo, Christopher Columbus, Vasco da Gama, Ferdinand Magellan – were associated with the use or trade of spices. Throughout history, the most powerful and richest countries were those countries that controlled the spice trade. Mankind has used spices and herbs to mask spoiled meat and body odor, preserve the remains of their dead, heal afflictions of the flesh and mind, and add enjoyment to their food.

Spices and herbs have historically been used to improved color (turmeric, saffron, paprika, chillies) and flavor (cinnamon, cloves, cumin, rosemary, sage, pepper) and, in some cases, improve the shelf-life (antioxidant and antimicrobial activity) of foods. Without spices, our diet would be bland, boring, and unappetizing. The aromatic and pungent principles that make spices so valuable are contained in their essential oils and oleoresins. (*See* **Antioxidants:** Natural Antioxidants.)

Classification

The definition of spices, herbs, and condiments has created considerable debate over the years. The International Standards Organization (ISO) defines spices and condiments as the natural vegetative products or mixture thereof, without extraneous matter, that are used for flavorings, seasonings, and imparting aroma to food. Spices and condiments are usually grown in tropical semitropical climates and comprise one portion of the plant. Spices are the whole or ground seeds, fruits, bark, or roots of a plant (Table 1). Herbs are generally defined as the leafy parts of a soft-stemmed plant whose main stem dies down to the roots and regrows each year (Table 1). Herbs are generally found in the more temperate regions of the world. Herbs generally contain a lower percentage of volatile (essential) oil than spices (Table 2). The US Food and Drug Administration (FDA) further specifies that paprika, saffron, and turmeric are colors, and that they must be labeled with their common names on food labels. While the legal definition of spices and herbs may vary between countries and trade organizations, this article will use the term spice to describe both spices and herbs collectively.

Economics

Since spices are an agricultural commodity, their prices are influenced by climatic and political conditions. Wars and national disasters also limit the exporting of spices.

While more than seventy recognized spices are used to season food, approximately 30% of the total international trade is accounted for by black and white pepper. Cloves, cardamon, cinnamon, nutmeg, and pepper represent about 70% of the world spice trade.

In 2000, the USA consumed approximately 1 billion pounds (453 600 tonnes) of spice per year. Per capita consumption increased to 3.7 pounds (1.7 kg). The average consumption has increased by over 50% from a decade ago due in part to the increased consumption of spices used to make foods less bland (e.g., capsicums, oregano, basil). Imported spices represent more than 70 herbs and spices grown and processed worldwide under a variety of environmental conditions. Black, white, and red peppers, mustard seed, capsicum, cassia, paprika, coriander,

Table 1 Geographic origin of common spices and herbs and plant parts used

Name	*Latin name*	*Plant part*	*Country of origin*
Allspice	*Pimenta dioica* L. (formerly *Pimenta officinalis*)	Unripe berry	Jamaica, Guatemala, Mexico, Honduras,
Anise seed	*Pimpinella anisum* L.	Seed	Spain, Egypt, Turkey
Basil	*Ocimun basilicum* L.	Leaves	France, Egypt, USA
Bay leaves	*Laurus nobilis* L.	Leaves	Turkey, Greece
Capsicum peppers	*Capsicum frutescens* L.	Includes red pepper, chilli pepper and sweet bell peppers	Hottest varieties: Africa, China Milder varieties: China, Japan, the Middle East, India, USA, Mexico Chilli pepper: California, New Mexico, Mexico
Caraway seed	*Carum carvi* L.	Fruit	Native to: Europe, Asia, North Africa Primarily exported from the Netherlands
Cardamon	*Elettaria cardamomum*	Fruit	India, Ceylon
Celery seed	*Apium graveolens* L.	Fruit	France, India, USA
Cinnamon	*Cinnamomum zeylanicum* Nees, *Cinnamomum cassia*	Bark	Indonesia (Korintyi), China
Cloves	*Eugenia caryophyllata* Thunb	Unopened flower buds	Madagascar and Brazil
Coriander seed	*Coriandrum sativum*	Ripe fruit. Leaves known as cilantro	Coriander: Morocco, Romania
Cumin	*Cuminum cyminum* L.	Fruit	Middle East, India, Turkey
Dill	*Anethum graveolens* L.	Dillweed, seeds	Egypt, India, USA
Fennel seed	*Foeniculum vulgare* Mill.	Dried fruit	India, Egypt, Turkey
Fenugreek	*Trigonella foenumgraecum* L.	Pod	India, Morocco
Ginger	*Zingiber officinale*	Rhizomes	China, India, Jamaica
Marjoram	*Majorana hortensis*, Moench	Leaves	France, Egypt, Europe, South America, USA
Mustard seed	*Brassica hirta*	Seeds	Canada, Northern plains of USA
Nutmeg and mace	*Myristica fragrans*, Houtt	Seed. Mace is a lacey netlike orange covering over the seed	Indonesia, East Indian Islands of Grenada and Trinidad
Oregano	*Origanum vulgare* L.	Leaves	Greece, Turkey, Mexico
Parsley	*Petroselinum crispum* L.	Curly leaf variety	California
Paprika	*Capsicum annum* L.	Ground fruit pods	California, Spain, Hungary
Pepper	*Piper nigrum* L.	Berries, Green (immature), White (ripe with black hull removed), Black.	Malabar Coast of India, Lampong district of Sumatra in Indonesia and Brazil
Rosemary	*Rosmarinus officinalis*	Leaves	Former Yugoslavia, France, Spain, Portugal
Saffron	*Crocus sativus* L.	Dried stigmas of the flower	Spain
Sage	*Salvia officinalis* L.	Leaves	Former Yugoslavia (Dalmatia), Albania, Turkey
Savory	*Satureja hortensis* L.	Leaves	Former Yugoslavia, France
Tarragon	*Artemisia dracunculus* L.	Leaves	California, France
Thyme	*Thymus vulgaris* L.	Leaves	Spain
Turmeric	*Curcuma longa* L.	Rhizomes	India (alleppey)

ginger, and oregano accounted for approximately 75% of the weight and 63% of the dollar value of these imported spices. The USA domestically produces 35–40% of its spice needs, including onions, shallots, chives, garlic, parsley, dill, basil, tarragon, mustard seed, mild and chilli peppers, cayenne (capsicum), spearmint, and peppermint leaves each year.

In the marketplace, similar spices from different geographical locations do not necessarily demand the same prices. European oregano (Greek and Turkish) demands a higher price than Mexican oregano.

Spices obtained from different plant parts have different volatile characteristics owing to differences in spices, environmental conditions, geographical origin, and processing. Milling also significantly influences the characteristic volatile oil content, composition, aroma, flavor, and color.

Spice Quality

Spice quality typically includes its degree of excellence and uniformity, as measured against specific

Table 2 Volatile oil content (% v/w)

Spice or herb	*Volatile oil content (% v/w)*
Anise	1.5–3.5
Allspice	2–4.5
Basil	0.1–1.1
Bay leaves	1.5–2.5
Caraway	1.5–3.5
Cardamon	3–8
Cassia (cinnamon)	0.5–4.0
Celery seed	1.5–3
Cloves	15–20
Coriander	0.1–1.5
Cumin	1.5–5
Dill seed	2–5
Fennel seed	1–3
Ginger	1.5–3
Mace	5–25
Marjoram	0.7–3
Nutmeg	5–16
Oregano	2.0–4.4
Pepper, black	0.6–5.0
Pepper, white	1.0–3.0
Rosemary	0.5–2.0
Saffron	< 1
Sage	1.5–2.5
Tarragon	0.2–1.5
Thyme	0.8–2.0

standards and specification (sensory properties, cleanliness, and safety). The quality attributes of spices can be divided into specifications concerning microbiological quality and safety, filth/contamination, foreign material, granulation, color, flavor, total ash and acid-soluble ash, moisture level, and volatile oil content. Specific tests (American Spice Trade Association (ASTA) color, Scoville heat units, curcumin content, piperine content, phenolic content) may also be used for certain spices. These specifications are set by the processor on the basis of international, federal, and trade standards.

Microbiology

A large percentage of the spices supplied to the world are produced in countries where sanitary conditions and hygienic handling and storage of spices are poor. Bacteria, insects, and bird and rodent droppings are common sources of contamination. Often, spices are laid out in fields or in backyard areas around houses to dry where they are subject to a high degree of contamination from animals and humans alike. It is not uncommon to find standard plate counts on black pepper of up to 10 million per gram. Spices with high microbial counts include allspice, caraway, celery seed, cumin, pepper, paprika, and onion powder. Standard plate count, yeast mold, coliform (*Escherichia coli*), and *Salmonella* tests are typically used to evaluate the microbiological quality of raw and processed spices. It is not uncommon to find metal wires and scrapings, dirt, sand, glass, bird feather, whole insects, and insect fragments in raw spices. Therefore, ISO, ASTA, and the US FDA have set cleanliness specifications (or guidelines) that all imported spices should meet. These are usually referred to as Defect Action Levels (DALs) and ASTA Cleanliness Specifications. Both of these specifications are designed to take into account that spices are a natural agricultural commodity that cannot be totally free of insects and foreign material. These numbers reflect the maximum amount of material that can be found under good manufacturing practices. Heavy and light filth testing procedures are used to evaluate the amount of filth and foreign material present in the spice (whole and ground spices). Total ash after ignition of the spice is a measure of metal, sand, rocks, and minerals naturally occurring in the spice. Acid-insoluble ash is the residue remaining after the total ash has been treated with hydrochloric acid, leaving only the sand. (*See* **Contamination of Food**.)

Moisture

High moisture levels in a dried spice are indicative of improper storage. High moisture levels lead to microbiological growth of molds and bacteria. Both degrade the quality of the spice and may lead to the presence of toxins (such as aflatoxin). Moisture testing is usually determined by vacuum oven drying for capsicums and dehydrated vegetables with low volatile oils. Toluene distillation of spices is used to determine the moisture of spices with high volatile oil contents (the procedure minimizes the loss of volatile oils that would contribute to a higher moisture determination). (*See* **Mycotoxins**: Occurrence and Determination; **Water Activity**: Effect on Food Stability.)

Volatile Oil

The terms volatile oil and essential oil are synonymous and are used interchangeably in the flavoring industry. These oils volatize at low temperatures and are the main contributors to the aroma and flavor of spices. Volatile oils are a complex mixture of organic compounds, hydrocarbons (terpenes, sesquiterpenes), oxygenated compounds (alcohols, esters, aldehydes, ketones), aromatic benzenoids, and compounds containing nitrogen or sulfur, present in droplets of cell sap and in glandular hairs. In plants, essential oils are functional enigmas; some may act as insect attractants and so aid in pollination, whereas others may protect plant tissue against parasites and animal depredation. For example, oregano and thyme essential

oils have been shown to inhibit molds and foodborne-illness bacteria. (*See* **Essential Oils**: Properties and Uses.)

The percentage of volatile oil in spices usually ranges from 0.5 to 3.0%, with the seed spices typically having higher values than the herbs (Table 2). The volatile oil content of spices is influenced by crop year, origin, milling technique, age, and extraction technique. Owing to these variations, a standard method and a minimum level of volatile oil are specified for the determination of volatile oils. ASTA Method 5 Modified Clevenger Trap Method) is typically used to steam-distill off the volatile oils from the spices and collect them to be measured.

The percentage of volatile oil has long been the method of choice for defining spice quality, but the validity and sensitivity of this method are questionable, as research has shown that consumers and analytical methods are able to detect differences not identified by the percentage of volatile oil. The steam distillation process often produces conditions that may distort volatile oil composition. The composition of the oil can be more important than the total amount of oil present in the spice. On-column cryogenic focused gas chromatography of the spice headspace and/or heat gas-chromatography injections of the volatile oil can be used as a supplemental technique for the determination of spice flavor quality. (*See* **Chromatography**: Gas Chromatography.) Electronic nose technology is also being used as quality-control instrument to differentiate different spices based on their volatile patterns. Typically, one to 32 sensors, of different types, are exposed to the headspace over the spice. If the volatiles bind with a sensor, they change its electrical resistance, which can be measured. The patterns produced from this measurement can be related to spices from different geographical locations and aroma quality. Many of these units have neural networks, which can be trained (using sensory or chemical data) to recognize these differences.

The relative heat of red peppers is measured by Scoville heat units. This procedure uses five trained taste panelists to determine the threshold hotness of ethanol extracts of red pepper: the higher the number, the hotter the red pepper. Unfortunately, the variation in Scoville numbers between laboratories can be as high as 50%. Newer instrumental methods using high-performance liquid chromatography have been developed to determine the amount of specific capsaicinoids present to try to reduce the variation found with taste panels. (*See* **Chromatography**: High-performance Liquid Chromatography.)

Solvent extraction methods are typically used to determine the color of paprika (carotenoid pigments) and turmeric (curcumin content). These procedures utilize acetone to extract the color pigments, which are then measured spectrophotometrically. The results are then expressed as ASTA color units. (*See* **Spectroscopy**: Overview.)

Spice Processing

Spices are treated as an agricultural commodity and shipped in bags, bales, or totes. Upon arrival, they must meet the cleanliness specifications of that country. Processors first check the whole spice for cleanliness by pulling a representative sample from multiple containers and determining the amount of foreign material (stones, metal, hairs, excreta, glass, insects) and microbiological load. If a product is found to have these items present in excess of the DALs, the spices must be cleaned or treated. Infestations due to insects and high microbial loads can be reduced by ethylene oxide, propylene oxide or irradiation treatment. Steam sterilization procedures are also used by a few spice companies, especially in Europe, to reduce microbiological loads. Wet steam sterilization requires drying of the spice after treatment. The use of ethylene oxide in sealed chambers usually reduces the microbiological load by 90%. However, ethylene oxide has been classified as a carcinogen, and the FDA has set a maximum residue tolerance of 50 p.p.m. in spice. Owing to the undesirable toxicological effects of ethylene and propylene oxide, alternative methods for sterilizing spices are being investigated. Irradiation of spices using ionizing radiation up to 30 kGy is allowed in the USA. The FDA requires that an irradiated spice must be labeled with the irradiation logo and the statement 'treated with/by irradiation.' However, if the irradiated spice is used in a food as an ingredient, the finished product does not have to be labeled as being irradiated. Irradiation of spices is very effective (killing 95–100% of microorganisms); however, many consumer groups are opposed to this procedure, owing to concerns about radiation, nuclear wastes, and unknown byproducts in the food. After the sterilization procedures (if required), the spices can be cleaned to remove foreign material. Magnets are used to remove ferrous metal during the cleaning and grinding operations. Nonmetallic foreign material can be removed from spices using a variety of methods, depending upon the foreign material present and the type of spice. Three basic types of cleaning techniques are typically used, alone or in combination with each other: air, gravity, and centrifugal separators. Each of these methods uses the difference in weight or specific gravity of the spice to separate it from the lighter or heavier foreign material. Gravity separators or air tables are

the most commonly used pieces of cleaning equipment. This technique uses air forced through a porous moving plate (deck) to suspend the particles into layers of different densities. The lighter material moves downhill, and the heavier material moves uphill. Proper adjustment of this system allows the processor to separate the product into many different density fractions (dust, lights, heavies, stones, etc.). (*See* **Irradiation of Foods**: Basic Principles; **Sterilization of Foods**.)

After cleaning, the spice is checked for cleanliness and other quality attributes. If the spice meets these specifications, it can be packaged and sold as whole spice or ground into smaller particles.

Spice Milling

Since ancient times, milling has been used to release the natural aromatic compounds contained by the plant's protective cell wall. Milling shatters the cellular material, starting the slow release of volatile aromatic compounds. During milling, not all the plant cells are disrupted; consequently, ground spices retain a portion of their original aromatic components during processing and storage.

Using a variety of milling techniques and equipment, millers encounter a wide range of different physiological material. The use of high-speed mills, which generate large quantities of heat with rapid air movement, results in the loss and/or deterioration of volatile components. Therefore, milling is carried out using as little grinding as possible with high-speed disintegrators, various types of hammer mills, and slow-speed vertical and horizontal burr stone mills in a wide range of sizes and capacities. During traditional milling, the frictional heat of grinding can increase to 93 °C, causing the loss of volatile oil and the development of off-aromas and flavors. Volatile oil losses of up to 1.5% during the milling of different spices between 42 and 95 °C are common. (*See* **Milling**: Characteristics of Milled Products.)

Cryogenic milling has been utilized to reduce these problems. In this procedure, spices are ground with liquid nitrogen to reduce the loss of volatile oil. The liquid nitrogen cools both the spice and grinding zone by instantaneous evaporation, absorbing any frictional heat. Grinding zone temperatures well below −73 °C with theoretical grinding zone temperatures approaching −195 °C have been reported.

There are several advantages to the 'Cryomill' process of spice milling.

1. It reduces the oxidation of spice oil, thereby increasing stability and flavor intensity by expelling air from the grinding zone.
2. It permits production of exceedingly fine particle size because the low temperatures solidify spice oil, making the spice very brittle.
3. The fine spice particles disperse uniformly throughout the finished product to give a uniform flavor, eliminate speckling, and reduce settling in liquid products.
4. It reduces the loss of volatile aromatic compounds, increasing flavor intensity.
5. The low-temperature treatment may reduce the microbial load.
6. The milling rate is increased, by reducing spice stickiness.
7. Cryogenically milled spices show a reduced cost when the increased flavor strength is factored into the cost.

Spice Oleoresins

Spices can also be solvent-extracted to produce oleoresins or steam-distilled to produce essential oils. Oleoresins are liquid, semisolid, or a solid residue composed of volatile oils, fixed oils (nonvolatile plant oils), pigments, pungent constituents, and natural antioxidants. Because oleoresins contain all of the volatile and nonvolatile flavor components of the spice, they provide a more rounded, closer duplication of the original spice. For example, the oleoresin of ginger contains pungent (hot sensation on the tip of the tongue or back of the throat) gingerol, shogal, paradol, and zingerone; white and black pepper contain piperine, and hot chillies (capsicum) contain capsaicinoids such as capsaicin, dihydrocapsaicin, and nordihydrocapsaicin. The oleoresins also contain natural antioxidants such as romaridiphenol and romariquinone in rosemary. Mace, sage, thyme, cloves, oregano, allspice, and cinnamon have also been reported to have antioxidant properties.

Both oleoresins and essential oils are very concentrated flavors and can be difficult to use in food processing. Extractives are typically used at levels of less than 0.01% of the finished food product. Standardization of the extractive is necessary, owing to crop year, geographical origin, age and lot-to-lot variation. In order to overcome these difficulties, extractives are usually incorporated into food products by being (1) dispersed on a dry carrier (salt, dextrose, flour), (2) solubilized with Polysorbate 80, (3) encapsulated inside vegetable gums or food starches (spray-drying), and (4) made into a water-dispersible liquid using emulsifiers. The amount of residual solvent in the spice oleoresin is also of concern, and levels are monitored by federal agencies. For example, in the USA, Title 21 of the Code of the Federal Regulations, parts 173 and 182.20, gives essential oils and

oleoresin generally recognized as safe (GRAS) status and defines the maximum solvent residue for hexane, acetone, isopropyl alcohol, and ethylene dichloride (25, 30, 50, 30 p.p.m., respectively) allowed in the spice extract. The label of a food containing oleoresins or essential oils need only state 'natural flavors.'

See also: **Antioxidants**: Natural Antioxidants; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Contamination of Food**; **Essential Oils**: Properties and Uses; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Irradiation of Foods**: Basic Principles; **Milling**: Characteristics of Milled Products; **Mycotoxins**: Occurrence and Determination; **Sensory Evaluation**: Aroma; **Spectroscopy**: Overview; **Sterilization of Foods**; **Storage Stability**: Mechanisms of Degradation; **Water Activity**: Effect on Food Stability

Further Reading

American Spice Trade Association (1976) *Official Microbiological Methods of the American Spice Trade Association*. Englewood Cliffs, NJ: ASTA.

American Spice Trade Association (1985) *Official Analytical Methods of the American Spice Trade Association*, 3rd edn. Englewood Cliffs, NJ: ASTA.

Boelan MH (1991) Spices and Condiments II. In: Maarse H (ed.) *Volatile Compounds in Food and Beverages*, pp. 449–482. New York: Marcel Dekker.

Dziezak JD (1989) Spices. *Food Technology* 43: 102–116.

Farrell KT (1990) *Spices, Condiments and Seasonings*. New York: Van Nostrand Reinhold.

Heath HB (1987) The quality control of flavoring materials. In: Herschdoerfer SM (ed.) *Quality Control in the Food Industry*, vol. 4, pp. 193–287. New York: Academic Press.

Pesek CA and Wilson LA (1986) Spice quality: effect of cryogenic and ambient grinding on color. *Journal of Food Science* 51: 1386, 1388.

Pesek CA, Wilson LA and Hammond EG (1985) Spice quality: effects of cryogenic and ambient grinding on volatiles. *Journal of Food Science* 50: 599–601.

Pruthi JS (1980) *Spice and Condiments: Chemistry, Microbiology, Technology. Advances in Food Science, Supplement 4*. New York: Academic Press.

Richard HMJ (1991) Spices and Condiments I. In: Maarse H (ed.) *Volatile Compounds in Food and Beverages*, pp. 411–447. New York: Marcel Dekker.

Risch SJ and Ho C-T (ed.) (1996) *Spices: ACS Symposium Series 660. Flavor Chemistry and Antioxidant Properties*. Washington, DC: American Chemical Society.

Rosengarten F (1973) *The Book of Spices*. New York: Pyramid Books.

Tainter DR and Grenis AT (2001) *Spices and Seasonings. A Food Technology Handbook*, 2nd edn. New York: John Wiley.

US Department of Health and Human Services (1989) *The Food Defect Action Levels. HHS Publication No. (FDA) 89-2199*. Washington, DC: USDHHS.

Fruits and Seeds

G D Bagchi and G N Srivastava, Central Institute of Medicinal and Aromatic Plants, Lucknow, India

Introduction

There are about 30 types of fruits and seeds that are commonly used as spices and for flavoring purposes throughout the world. Their frequency of use varies in different parts of the world. In this article, some of the most commonly used spices and flavoring crops are discussed in detail, and the remaining equally important ones are summarized in **Table 1**: each section deals with an individual spice in terms of its common name, botanical name, family, and parts used, along with diagnostic, morphological and anatomical characters; details of handling, drying, and storage have also been given wherever available.

Allspice or Pimento

Allspice consists of the dried fruits of *Pimenta officinalis* Lindl. (Myrtaceae). Dried berries are globular, 4–7 mm in diameter, hard, and dark reddish brown; the surface of the berry is rough with protuberances, and the apex bears the remains of the calyx and the style; there is a short stalk, or its scar at the base. The berries are bilocular each with hard, dark brown, reniform seeds (**Figure 1a** and **b**).

The epidermis of the exocarp bears stomata and nonglandular hairs up to 150 μm in length. The mesocarp, towards the outside, consists of numerous oval or rounded oil cavities up to 210 μm in diameter; numerous scattered, variable, pitted stone cells occur singly or in groups; beneath, there is a zone of fibrovascular bundles; ground tissue is parenchymatous, and contains 'rosette' calcium oxalate crystals. Group of stone cells, 30–60 μm in diameter, are present towards the inner side. The endocarp consists of a few layers of compressed parenchyma (**Figure 1c**).

Allspice is aromatic and pungent, and possesses the flavor and aroma of clove, nutmeg, cinnamon, and black pepper, hence the name 'allspice.' Mature berries are removed from the twig when still green by hand or flail, sorted, and dried in the sun for 6–8 days; when dull reddish brown, the berries are packed in jute bags. At present, sun drying of berries is done on a concrete barbeque. Because of frequent shifting of the berries in and out of the sheds during rainy days, many berries break. Hence, mechanical drying is preferable, and it is also more economical than

Table 1 Some other commonly used spices from fruits and seeds

Botanical name	*Common name*	*Family*	*Part used*	*Characters*
Aframomum korarima (Pereira) Engler	Korarima cardamom	Zingiberaceae	Fruit and seed	Fruits ovate, about 4 cm by 2 cm pointed seeds reddish brown, striated
Aframomum melequeta (Rosc.) K. Schum.	Grains of paradise	Zingiberaceae	Fruit and seed	Fruits ovoid 10 cm long, pointed. Seeds inverted subpyramidal, somewhat four-sided, 2–3 mm by 2.5 mm, testa rich reddish brown, papillose
Amomum subulatum Roxb.	Greater or Nepal Cardamom	Zingiberaceae	Fruit and seed	Fruit is dark red–brown, globose capsule, 2.5 cm long; several cardamom seeds, similar to true cardamom, embedded in a vicid pulp
Bunium persicum (Boiss.) Fedts	Black caraway	Apiaceae	Fruit	Mericarp oblong (3–4 mm by 0.7–1.0 mm), yellowish dark brown with ridges
Dipteryx odorata Willd.	Tonka beans	Fabaceae	Fruit	Fruit egg-shaped with hard shell and pulpy flesh surrounding a single seed with a black wrinkled surface
Garcinia indica Choisy	Kokam	Clusiaceae	Fruit	Fruit globose or spherical 2.0–3.75 cm in diameter, dark purple when ripe
Juniperus communis L.	Juniper	Pinaceae	Fruit	Fruit subglobose, 5–9 mm in diamter externally smooth, shining, purplish black to dusky red purple internally yellowish brown to dusky yellow flesh, containing many large, schizogenous cavities; ripe fruits are collected, and dried flavor diminishes after storing fruits for more than a year
Mangifera indica L.	Amchoor	Anacardiaceae	Fruit	Fruit a drupe, variable in form and size, leathery, dotted with glands; mesocarp of unripe fruit hard, acidic; ripe fruits juicy; endocarp stony; skin of unripe fruit removed, mesocarp sliced and dried in sun; the dried product is packed in wooden casks and may also be seasoned before drying
Nigella sativa L.	Black cumin	Ranunculaceae	Seed	Seeds trigonous, black, rugulose and tuburculate; seeds are collected mostly from wild plants
Piper cubeba Linn. f.	Cubeb	Piperaceae	Fruit	Fruit globular, 3–6 mm in diameter with a stem-like portion about 7 mm long, apex with four remains of stigma; dusky red, brown or, rarely, grayish; coarsely reticulate; nearly full grown fruits are collected and dried in sun
Piper longum L.	Long pepper	Piperaceae	Fruit	Catkin is subcylindrical, up to 40 mm by 5–7 mm; greyish black and composed of numerous small immature drupes and bracts embedded spirally in the axis; immature catkins are harvested and immediately dried in the sun or oven to prevent formation of mold
Pistacia vera L.	Pistachio	Anacardiaceae	Fruit	Drupe oblong, linear to globose, laterally compressed, 10–20 mm by 6–12 mm; outer husk variously colored, readily separating from the dehiscent or indehiscent grayish white, bony keeled nutshell (endocarp) enclosing yellow to green kernel
Punica granatum L.	Anardana (pomegranate)	Punicaceae	Seed	Seeds hard, angular, covered with a juicy red, pink or yellowish white pulp; seeds of ripe fruits are taken out and sun-dried or dehydrated
Sisymbrium irio L.	London rocket (khub-kalan)	Brassicaceae	Seed	Seeds ovoid (1.3 mm long), reddish or yellowish brown, mucilaginous

Continued

Table 1 Continued

Botanical name	*Common name*	*Family*	*Part used*	*Characters*
Trachyspermum ammi (L.) Sprague	Bishops weed (Ajowan)	Apiaceae	Fruit	Cremocarp ovoid (2–3 mm long), muricate, grayish brown; mericarps compressed with distinct ridges and tuberculate surface harvested when fruits turn brown; plants are pulled out and dried completely on threshing floor; fruits are separated by rubbing.
Tamarindus indica L.	Tamarind	Fabaceae	Fruit	Fruits 5–20 cm by 2 cm in size; epicarp surface rough, brownish hard and brittle; mesocarp pulpy light brown changing with age to dark brown or black; containing some branched fibers and numerous seeds
Vanilla pompona Schneider	Pompon or West Indian vanilla	Orchidaceae	Fruit	Pods are 10–17.5 cm long and 2.5–3.3 cm in diameter
Vanilla tahitensis J.W. Moore	Tahitian vanilla	Orchidaceae	Fruit	Pods 12–14 cm long and about 9–10 mm wide with tapering ends; pods are harvested when tips of the pods turn brown
Xylopia aethiopia (Dunal) A. Richard	Xylopia, Netawu	Annonanaceae	Fruit	Mature carpels cylindrical, incurved, about 5 cm long, embedded in the center of the receptacle; seeds glabrous, ellipsoid (about 0.5 mm by 0.25 mm), brown

sun-drying. (*See* **Drying**: Drying Using Natural Radiation; **Sensory Evaluation**: Aroma; Taste.)

Anise

Anise is the dried fruit of *Pimpinella anisum* L. (Apiaceae). The schizocarps are 2–5 mm, ovoid–conical, grayish brown, and pubescent, with a pedicel attached; they are crowned with a short, bifurcate stylopod. The mericarps, attached to the carpophore, are broadly ovoid; they are flat or concave on the commissural side, convex and pubescent on the dorsal side (Figure 1d and e).

The exocarp has an outer striated cuticle with stomata and rectangular epidermal cells bearing surface hairs up to 150 µm in length. The mesocarp consists of five fibrovascular bundles underlying ridges, two to four vittae on the commissural side and 20–40 vittae on the dorsal side. The endosperm of the seed is composed of thick-walled parenchymatous cells containing oil and aleurone grains and one or two small 'rosette' calcium oxalate crystals (Figure 1f and g).

Anise has a characteristic and agreeable aromatic odor and a pleasant taste. Fruits are harvested when their color changes from green to gray; shattering is minimized by avoiding overripening and harvesting during the morning with scythes or reaping machines. The crop is piled into small stacks and allowed to wither for 2–3 days; it is then threshed and dried in partial shade. After winnowing, sieving, and cleaning, the crop is dried naturally to 18% moisture followed by artificial drying at 80–90 °C; it is then packed and stored in a cool, dry room. For keeping spices belonging to family Apiaceae, the maximum tolerance limit of moisture is 10%. Reducing the moisture below 8% can generally have a bad effect on the quality due to loss of volatile oils. Crushed or ground spice is stored in airtight containers. (*See* **Drying**: Theory of Air-drying.)

Black Pepper

Black pepper consists of the dried, unripe fruit of *Piper nigrum* L. (Piperaceae). It is the oldest and most important spice and commonly known as the 'King of spices.' There are about a hundred cultivars of black pepper under cultivation. There is considerable variability among the accessions with regard to various morphological as well as floral characters like leaf shape, size, spike length, density, composition of male, female and bisexual flowers in the spike, fruit set, shape, weight and volume of the fruit, and time taken for maturity. The fruit, though known as a berry, is a sessile globose drupe with a pulpy pericarp. The berry is globular, 6 mm or less in diameter, brownish black, with a coarsely reticulate surface, and single-seeded. The seed is almost white hollow, and adheres to the pericarp (Figure 1h and i).

The exocarp is cutinized and darkly pigmented, with rectangular cells beneath containing calcium oxalate crystals. The hypodermis consists of stone cells interspersed with parenchyma. The mesocarp is comparatively broad: the outer seven to eight layers

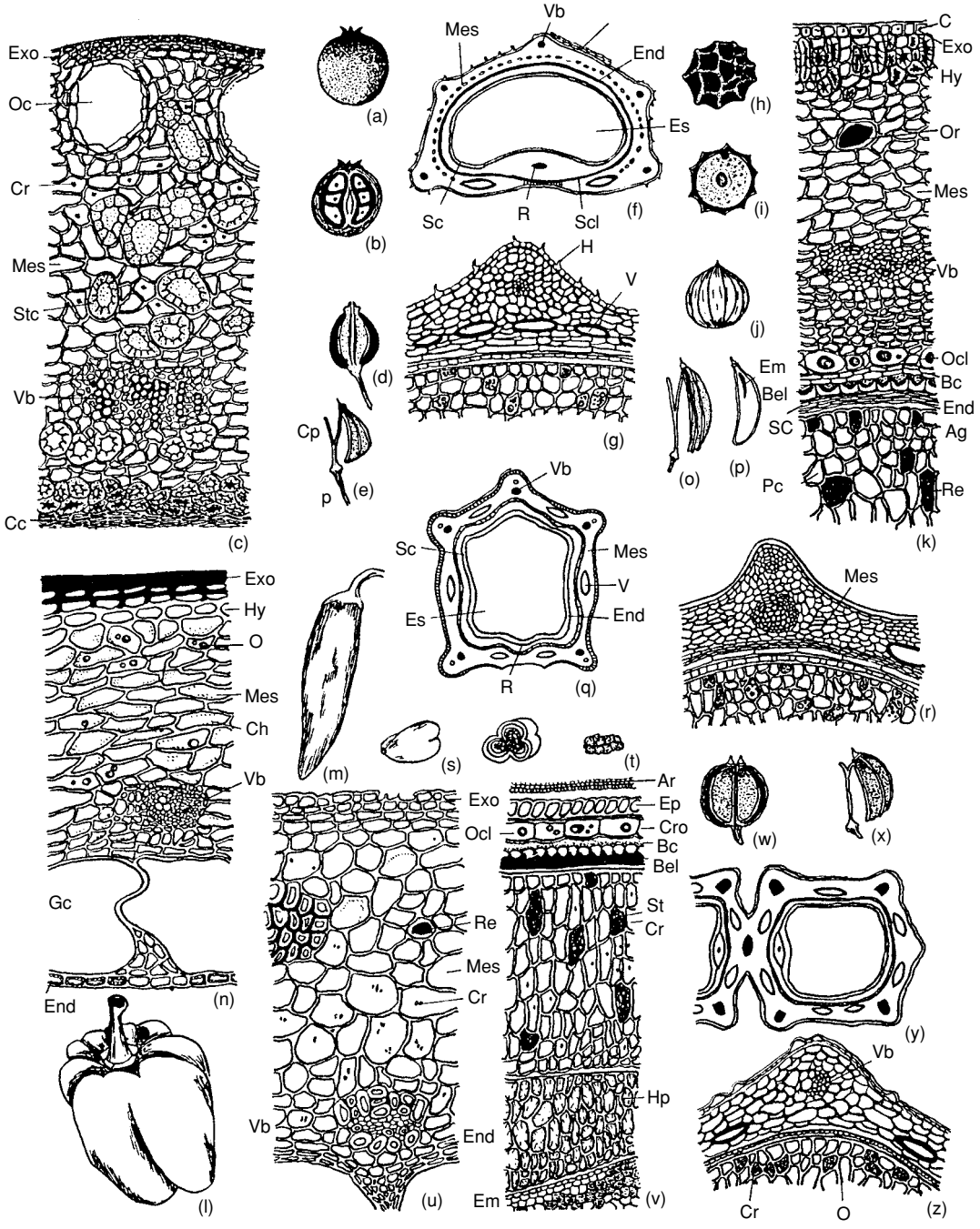

Figure 1 Spices from fruits. **Allspice:** (a) fruit; (b) longitudinal section of fruit; (c) cross-section of pericarp. **Anise:** (d) schizocarp; (e) mericarp; (f) cross-section of mericarp; (g) cross-section through ridge of pericarp. **Black pepper:** (h) fruit; (i) cross-section of fruit; (j) pericarp, seed coat and portion of perisperm. (k) **White pepper. Capsicum:** (l) *Capsicum annum* fruit; (m) *C. frutescens* fruit; (n) cross-section of pericarp of *C. frutescens*. **Caraway:** (o, p) mericarp; (q) cross-section of mericarp; (r) cross-section through ridge of pericarp. **Cardamom:** (s) fruit; (t) seeds; (u) cross-section of pericarp; (v) cross-section of seed. **Celery:** (w) schizocarp; (x) mericarp; (y) cross-section of mericarp; (z) cross-section through ridge of pericarp. Reproduced from Spices and Flavoring (Flavouring) Crops: Fruits and Seeds, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press. Abbreviations: Ad, amylodextrin granules; Ag, aleurone grains; Ar, aril; Bc, beaded cells; Bel, beaker cells; C, cuticle; Cc, compressed cells; Ch, chloroplast, Col, colenchyma; Cot, cotyledon; Cp, carpophore; Cr, crystal; Crc, crystal cells; Cro, cross cells; Dl, dehiscence line; Em, embryo; End, endocarp; Ep, epidermis; Es, endosperm; Exo, exocarp; Fl, fiber layer; Fo, fixed oil; Gc, giant cell; H, hair; Hc, hourglass cells; Hp, horny protein; Hy, hypodermis; Li, light line; Mc, mucilage cells; Mes, mesocarp; Mp, mesophyl; O, oil; Oc, oil cavity; Ocl, oil cells; Or, oleoresin; P, pedicel; Pa, parenchyma; Pal, palisade; Pe, perisperm; Pl, placenta; R, raphe; Rc, reticulated cells; Re, resin; Rub, rudimentary vascular bundle; Sc, seed coat; Scl, sclerenchyma; Sd, seed; Se, secretory cells; St, starch; Stc, stone cells; V, vittae; Vb, vascular bundle.

are parenchymatous with scattered oleoresin cells, and beneath, there is a zone of fibrovascular bundles, a region of compressed cells, a layer of large oil cells, and two layers of small parenchyma cells. The endocarp consists of a layer of 'beaker cells,' the radial and inner walls of which are strongly lignified. The seed coat is differentiated into the distinct layers, and whereas the outer perisperm cells contain aleurone grains, the inner cells contain starch grains (Figure 1k).

The fruits are aromatic and pungent. The quality of black pepper depends upon piperine and essential oil present in the oleoresin. Piperine contributes to the pungency, and essential oil is responsible for the characteristic flavor. The harvested green berried spikes are heaped for a day and spread on a mat or floor. The berries are separated from spikes by rubbing. After separation, the berries are cleaned and put out for drying. The fresh harvested berries will have a moisture content of 80%, and this has to be dried to 8% moisture content for safe storage. A blanching treatment by dipping the berries for 1 min in boiling water gives a glossy appearance to the dried berries and reduces the microbial load on the product. Due to enzymic oxidation of color compounds present in the skin, the colour of pepper turns black during drying, which effectively masks the green color that is left in the berries after drying. The berries are dried under sun for 4–7 days until the outer skin of the berries becomes black in color with a wrinkled appearance. Small and light berries are removed from the dried produce. Berries can be dried in a solar cabinet dryer in 4 days with a spreading density of $18\,kg\,m^{-2}$ and resulting in a black dried produce. By contrast, sun-dried berries take 5 days to dry with a spreading density of $5\,kg\,m^{-2}$ and resulting in a dull black color. Dark black berries obtained from solar cabinet drying are preferred in trade over the dull black berries obtained from open sun drying. Sixty kilograms of fresh pepper can be dried in the indirect heating type of dryers in about 30 h. After drying, the pepper has to be garbled. Ungarbled pepper contains fractions, dust, stalk, pinheads, and immature, overmature, and large berries. The grading is done by a combination of size sieving and weight classification by air blast. The major grade is the average-sized black pepper known as 'Malabar Garbled,' which constitutes 95% of India's export. The recovery of black pepper from fresh berries is 35–36%. For black pepper powdering, a normal pin type of hammer mill is used so that there is no temperature rise during grinding. The ground pepper is only used as a raw material for oleoresin and in the oil industry. White pepper (Figure 1j) is usually prepared from overripe berries. In this state, the skin comes off easily, and the resultant product is dried. White pepper fetches double the price of black pepper. The piperine content also is found to be apparently higher in white pepper due to its concentration in the endosperm of the berries. In India, the berries are dipped in water and the skin removed by rubbing with hand. The skinless berries are dried in open sun for 7 days to obtain white pepper. Skin is also removed in a pulping machine after steam cooking. At the National Research Centre for spices in Calicut, a process of steaming and boiling the despiked pepper for 13 min and rolling it in a container in running water to remove outer skin has been developed.

After sun-drying, the berries are cleaned and polished with a piece of cloth. The recovery of white pepper is about 28% of the fresh ripe berries. In European Countries, the green or fresh flavor of pepper is desired in certain food preparations. Green pepper is made in canned and bottled forms. In the canned form, it is prepared by heat-sterilizing green berries and packing them in 2% brine solution for protection. However, the rupturing of berries and making the covering liquid turbid are problems associated with this method. These problems have been overcome with proper canning methods and the introduction of a trace of pepper oil into the can before sealing. In bottled green pepper, heat processing is avoided, but a higher salt strength of 15–20% of the covering liquid is essential for good preservation. At CFRI, undermature berries are dried by arresting the enzyme action responsible for oxidative blackening and drying by mechanical means at low temperatures. The product has an excellent green flavor.

Capsicum

Capsicum is composed of dried fruits of the genus *Capsicum* L. (Solanaceae). Cayenne pepper and red pepper are two of the various synonyms of *C. frutescens* L., whereas paprika is *C. annuum* L. The fruits of *C. frutescens* are 6–150 mm long, laterally compressed, musky brownish to deep red, smooth or wrinkled, and bi- to tetralocular with central placentae united at the base-apical placentae with longitudinal ridges; the seeds are numerous, discoid, flat, hard, smooth, yellow, and about 2–5 mm long (Figure 1l and m).

The exocarp consists of five to seven rows of subrectangular cells, 25–60 μm by 15–20 μm; the outer epidermis is cuticularized and the mesocarp is parenchymatous, 30–85 μm by 15–20 μm. Many cells contain oil droplets and chromatophores, or sandy crystals. The innermost layer consists of larger cells, 0.8–1.7 mm by 0.2–0.4 mm. The endocarp is single–layered, consisting of parenchymatous and

sclerenchymatous cells, the latter directly beneath a layer of giant cells possessing thick, lignified, porous walls. The dissepiment consists of a zone of tangentially elongated, collapsed cells (some containing microcrystals) and epidermal cells on either side, radially elongated, containing capsaicin (Figure 1n).

The odor of the spice is characteristic, sternutatory but agreeable, mildly to intensely pungent. The stem of the entire plant is cut, hung, and exposed to the sun to remove 80% moisture; it is then dried further in a dryer to a critical moisture level; pricking the skin longitudinally reduces the drying time. Completely dried fruits are removed by a dressing machine and packed in polythene bags.

Caraway

Caraway is the dried fruit of *Carum carvi* L. (Apiaceae). The mericarps are separated, crescent-shaped, 3–7 mm by 1–2.3 mm, and brown, with five lighter-colored, filiform primary ribs, between which, on the dorsal surface, there is a secondary rib (Figure 1o and p).

In transverse section, the epicarp consists of slightly tangentially elongated epidermal cells, and its outer surface is cuticularized; the mesocarp consists of a zone of several rows of collapsed, tangentially elongated parenchymatous cells, with fibrovascular bundles located in the rib portion; large vittae, two and four in number, are present on the commissural and dorsal side, respectively; the endocarp contains inner epidermal 'cross cells' that are closely adherent to the seed coat (Figure 1q and r).

The odor and taste of caraway are characteristically aromatic. The endosperm structure of the seeds, and handling, drying, and storage of the spice are the same as for anise.

Cardamom

Cardamom is the dried fruit of *Elettaria cardamomum* Maton (Zingiberaceae). It is popularly known as the 'Queen of spices.' There are many varieties, which are classified on the basis of their capsule size and color. Some of the important varieties are Alleppy, Mysore, Malabar, and Mangalore cardamoms. The capsules are trilocular, containing seeds on an axile placenta; they are 1–2 cm long, oblong-ovoid, three-sided, greenish to pale buff or yellowish in color; each locule contains five to eight seeds, 3 mm by 2–3 mm, dark reddish brown, adhering together with a membranous aril to form a single mass (Figure 1s and t). The pericarp in transverse section shows an outer exocarp with cells up to 30 μm long. Parenchymatous cells of the mesocarp are up to 180 μm long, containing one or more calcium oxalate crystals. The inner mesocarp contains fibrovascular bundles and resin cells. Endocarp cells are compressed and invaginate into locular partitions (Figure 1u). Seeds, in transverse section, show a three-cell-layered aril containing a pale-colored substance and calcium oxalate crystals. It has an outer epidermis followed by a layer of pigmented, large, clear cells containing volatile oil, a one- to two-layered parenchyma, and a layer of sclerenchymatous, thick-walled, dark brown 'beaker cells,' each with a nodule of silica; the perisperm consists of starch-containing polygonal cells, the endosperm cells containing proteinaceous substance having a waxy or horny nature, and the embryo cells containing aleurone grains and fixed oil (Figure 1u and v).

The seeds are much more aromatic than the pericarp and possess a characteristically warm and slightly pungent taste. Fruits on the same raceme ripen at different times, and ripe fruits are collected before splitting. A green curing method is common in commerce. Fully mature green capsules, not having excessive shrinkage and splitting, which are not fully ripe yet, retain a good color after drying. The temperature and relative humidity of drying air inside the drying chamber are the two important factors to be controlled in the cardamom drying system. The recommended temperature for cardamom is 45–55 °C. Cardamom are dried by different systems by kiln, driers, electric driers, or infrared driers. Cardamom drying has two phases; in the first phase, the surface moisture and moisture in fleshy parts of the mesocarp have to be removed, and in the second phase, the moisture in the seeds inside has to be removed. Controlled temperature, quick removal of the moisture, and quick flushing out of moisture-laden air from inside the drying chamber to prevent condensation are basic criteria in the retention of a green color. The dried cardamom capsules are placed into a wiremesh cylinder. The wiremesh cylinder is swung to separate the flower stalks, and the capsules are cleaned by winnowing. The cleaned capsules are packed in gunny bags lined with black polythene sheet for better retention of green color during storage. Grading of green cardamom is done according to the standard specifications at the exporter level. Extra Bold and Bold are the superior qualities.

Bleached Cardamom or White Cardamom

This processing is resorted to, especially in varieties that do not retain the green color during drying or lose the color during storage. This product fetches lower prices. Freshly harvested ripe capsules are pretreated by soaking in a 1% solution of bleaching powder for 30–60 min followed by dipping the lot

in dilute HCl solution (0.2–0.5%) for 2–3 min. The treated capsules are washed and sun-dried to 8% moisture level. But the sulfur dioxide content was found more in the capsule treated by the above method, a simple treatment with hydrogen peroxide (solution of 0.5% strength) has been standardized to overcome this problem. Complete drying using a solar cabinet dryer is obtained in 3 days in comparision with 5 days under open sun-drying, and the color of the capsules was white, the required bleaching obtained without sulfur bleaching. The volatile oil in cardamom is responsible for cardamom's aroma and therapeutic properties. Green cardamom yields more oil than bleached cardamom.

Celery

Celery is the dried fruit of *Apium graveolens* L. (Apiaceae). The schizocarps are ovoid and laterally compressed and occasionally possess a pedicel. The mericarps, usually separated, are ovoid, slightly curved, 1–2 mm by 0.5–1 mm, and light to olive brown; the commissural surface is flat and the dorsal surface convex and smooth, with five longitudinal ridges (**Figure 1w** and **x**).

In transverse section, the mericarps are equilaterally pentagonal; the exocarp has a striated cuticle, and the epidermis has stomata – some of the cells are papillose; the mericarp is parenchymatous, has four to 15 vittae, of which two are on the commissural side, and fibrovascular bundles occur in each primary rib; the endocarp consists of a layer of tangentially elongated cells with yellowish orange to greenish yellow walls, and the spermoderm consists of collapsed cells. The structure of the endosperm is similar to that of anise (**Figure 1y** and **z**).

The fruit bear a characteristic celery aroma and taste. Celery yields 2–3% of pale yellow volatile oil, which is valued both as a fixative and as an ingredient of perfumes. Handling, drying, and storage of the spice are also similar to those of anise. The quality of celery depends upon its volatile oil content, and hence, the higher the oil content, the better the quality. The seed yield shows a significant and positive correlation with plant height, total umbels per plant, number of umbellets per umbel, weight of the main umbel, and number of seeds per umbel. All these characters have a positive effect on celery seed yield.

Coriander

Coriander spice is derived from the dried, ripe fruits of *Coriandrum sativum* L. (Apiaceae). It was one of the earliest spices used by mankind. The schizocarp is globular, yellowish brown, 2–5 mm in diameter, with five calyx teeth and a short, conical stylopodium at the apex. Each mericarp has five prominent, straight, longitudinal primary ribs and four undulate, secondary ribs, which are easily separated, being deeply concave on the commissural side and convex on the dorsal side (**Figure 2a** and **b**). In transverse section, the exocarp consists of a layer of small, thick-walled epidermal cells, and the mesocarp consists of several rows of thin-walled, somewhat collapsed parenchymatous cells, separated from a broad continuous zone of thick-walled, strongly lignified sclerenchyma fibers. The commissural side has two large elliptical vittae. The endocarp is a layer of tabular cells, and the cells of the seed coat are tangentially elongated and brownish. The structure of the endosperm is similar to that of anise (**Figure 2c** and **d**).

The spice has a characteristically pleasant and fragrant odor and an aromatic taste. Significant variations for oil content in coriander have been observed, ranging from 0.17 to 0.26%. The preferred cultivars of the plant for essential oil production are the small fruit varieties with a variable oil content of 4% or more. For spice use, however, it is the larger fruits from Morocco and India that are more commonly preferred. It has been reported that large thick skinned fruits are less aromatic and have a lower oil content than the smaller thin skinned fruits. There is no correlation between the number of fruits in the main umbel and yield. The flowers on the branches in the middle of the stalk are the most productive. It has been found that the number of umbels per plant is the chief contributing character and that the plant height determines the yield. Handling, drying, and storage are similar to those for anise.

Cumin

Cumin consists of the dried, ripe fruit of *Cuminum cyminum* L. (Apiaceae). The schizocarps are about 6 mm long, attached to short pedicels, and separated into mericarps. The fruit resembles caraway, but can be distinguished by its straighter mericarps and bristly hairs along the ribs (**Figure 2e–h**).

In the transverse section, the exocarp has an outer epidermis bearing stomata, and a thin and striated cuticle; the hairs on the secondary ridges are larger than those on the primary ridges, being up to 2 mm long, and multicellular with broad bases and rounded apices; the mesocarp is parenchymatous, consisting of five fibrovascular bundles beneath the primary ridges, two vittae on the commissural side, and four vittae on the dorsal side. The parenchymatous cells of the endocarp have pale yellowish walls, and the seed coat consists of a layer of epidermal cells, each containing a prismatic crystal of calcium oxalate.

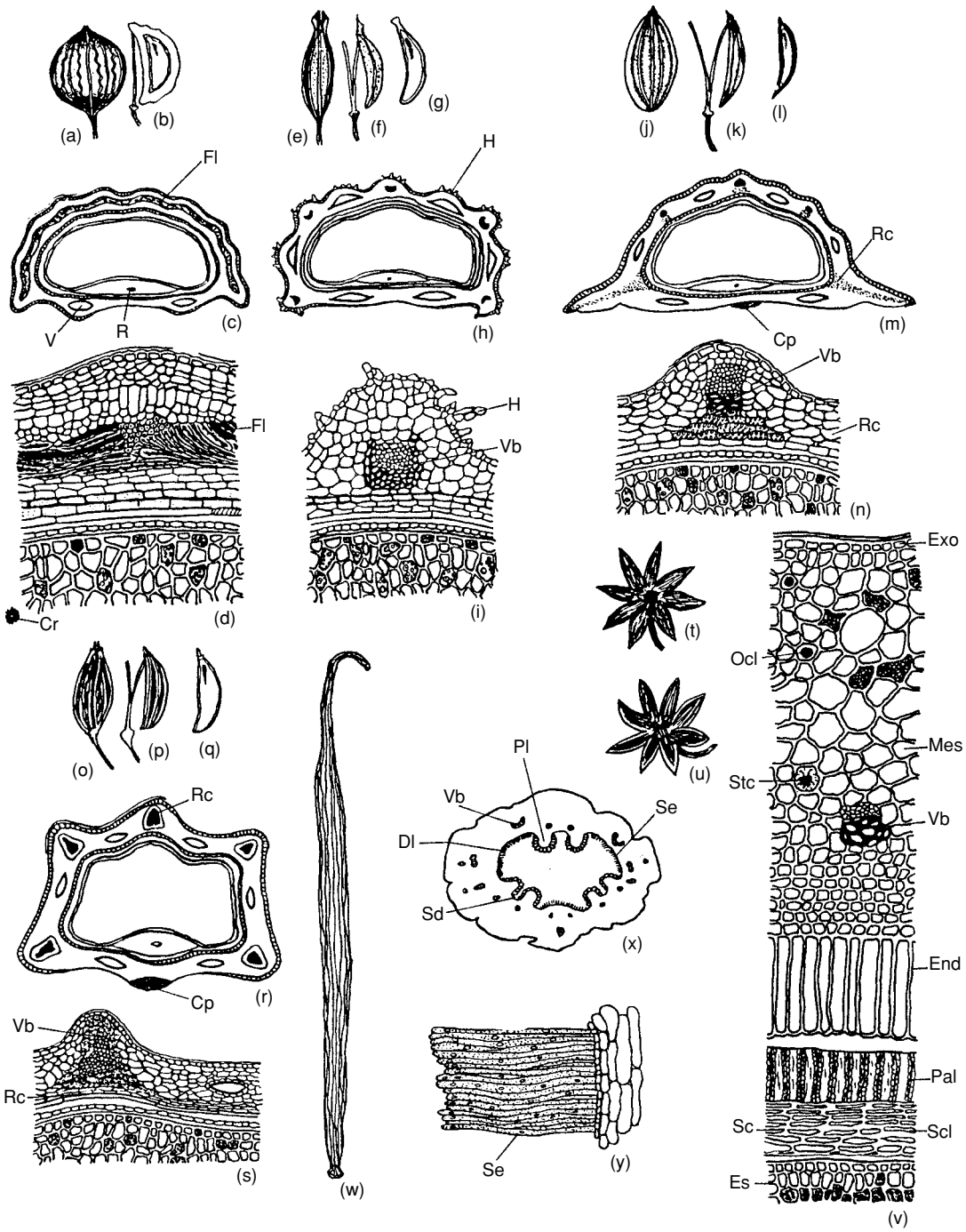

Figure 2 Spices from fruits. **Coriander:** (a) schizocarp; (b) mericarp; (c) cross-section of mericarp; (d) cross-section through primary ridge of pericarp. **Cumin:** (e) schizocarp: (f, g) mericarp; (h) cross-section of mericarp; (i) cross-section through ridge of pericarp. **Dill:** (j) dorsal view of mericarp; (k, l) side views of mericarp; (m) cross-section of mericarp; (n) cross-section through ridge of pericarp. **Fennel:** (o) schizocarp; (p, q) mericarp; (r) cross-section of mericarp; (s) cross-section through ridge of pericarp. **Star anise:** (t, u) fruit; (v) cross-section of fruit and seed. **Vanilla:** (w) fruit; (x) transverse section of fruit; (y) secretory hairs. Reproduced from Spices and Flavoring (Flavouring) Crops: Fruits and Seeds, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press. Abbreviations: Ad, amylodextrin granules; Ag, aleurone grains; Ar, aril; Bc, beaded cells; Bel, beaker cells; C, cuticle; Cc, compressed cells; Ch, chloroplast, Col, colenchyma; Cot, cotyledon; Cp, carpophore; Cr, crystal; Crc, crystal cells; Cro, cross cells; Dl, dehiscence line; Em, embryo; End, endocarp; Ep, epidermis; Es, endosperm; Exo, exocarp; Fl, fiber layer; Fo, fixed oil; Gc, giant cell; H, hair; Hc, hourglass cells; Hp, horny protein; Hy, hypodermis; Li, light line; Mc, mucilage cells; Mes, mesocarp; Mp, mesophyl; O, oil; Oc, oil cavity; Ocl, oil cells; Or, oleoresin; P, pedicel; Pa, parenchyma; Pal, palisade; Pe, perisperm; Pl, placenta; R, raphe; Rc, reticulated cells; Re, resin; Rub, rudimentary vascular bundle; Sc, seed coat; Scl, sclerenchyma; Sd, seed; Se, secretory cells; St, starch; Stc, stone cells; V, vittae; Vb, vascular bundle.

The endosperm is similar to that of anise fruits (Figure 2h and i).

The spice has a peculiar, strong odor, and the flavor is warm and slightly bitter. The volatile oil content varies from 3 to 6%. Handling, drying, and storage of cumin are the same as those for anise.

Dill

Dill consists of the dried and ripe fruits of *Anethum graveolens* L. (Apiaceae), usually composed of separated mericarps, which are brown, glabrous, broadly oval, and compressed (about 4 by 2.5 mm) with five ribs – three short dorsal, and two lateral, yellowish and wing-like (Figure 2j–m).

The exocarp has a striated cuticle and an epidermal layer of rectangular cells bearing stomata. The parenchymatous mesocarp has three fibrovascular bundles below the dorsal ridges and two in lateral ridges. The inner side of the vascular bundles consists of several tangentially elliptical, sclerenchymatous cells having reticulate thickenings. The mesocarp has two vittae on the commissural side and four vittae on the dorsal side. The seed coat consists of a layer of epidermal cells, and the endosperm is similar to that of the anise fruits (Figure 2m and n).

The aromatic odor and taste of dill are similar to those of caraway. Handling, drying, and storage are as described for anise.

Fennel

The spice is the dried fruit of *Foeniculum vulgare* Miller (Apiaceae). The schizocarp is oblong-oval or elliptical, usually with long pedicels (4–5 mm by 1–3.5 mm); at the end is a short, conical stylopodium. The mericarps are attached to a divided carpophore and separate easily. Each is broadly elliptical, more or less curved (dorsal surface is convex), and light brown to light olive in color, with five prominent, longitudinal, primary ribs (Figure 2o–q). Based on the appearance of the plant and seed and the compositional data of the volatiles from the fruit, fennel is classified into two subspecies and further classified into numerous varieties. The important varieties are:

- *F. vulgare* ssp. *piperitm* wild fennel
- *F. vulgare* ssp. *capillaceum* with three varieties
  1. var. *vulgare* – bitter fennel
  2. var *dulce* – sweet fennels
  3. var. *azoricum* – culinary florence fennel.

In transverse section, the epidermis of the exocarp shows polygonal tabular cells, having occasional stomata and a smooth cuticle; the mesocarp is parenchymatous with two to four vittae on the commissural side and four vittae on the dorsal side, whereas the endocarp consists of narrow, elongated cells in a parquetry arrangement in a transverse section as long, narrow, rectangular cells with groups of short cells. The endosperm has a structure similar to that of anise (Figure 2r and s).

Fennel has a pleasant, aromatic odor and a taste resembling that of anise. The volatile oil content in fennel ranges from 0.7 to 6.0%. The main constituent of oil is anethole. Handling, drying, and storage of this spice are also similar to those of anise.

Star Anise

Star anise is the ripe, dried fruit of *Illicium verum* Hook. Filius (Magnoliaceae). The fruit whorl (2.5–4.5 cm in diameter) is formed of six to eight (rarely seven to nine), one-seeded, boat-shaped, woody, wrinkled, ridged, reddish brown follicles that are internally smooth, lustrous, and light brown, about 9–19 mm in length and surmounted on a curved peduncle. The seed is light brown, compressed-ovoid, shiny, and smooth with a conspicuous hilum and raphe (Figure 2t and u).

In transverse section, the exocarp consists of a striated cuticle and epidermal cells bearing stomata. The mesocarp consists of parenchymatous cells with brown contents. The cells towards the central zone are larger (up to 220 μm in length), and the inner mesocarp cells have thicker walls, the thickness increasing towards the dehiscent side of the carpel. Resin and stone cells occur scattered throughout the mesocarp, and vascular bundles occur in between the central and inner regions. The endocarp consists of thin-walled, sclerenchymatous palisade cells up to 440 μm in length. The seed coat consists of an outer, thick-walled epidermis of sclerenchymatous, pitted palisade cells up to 200 μm in length, followed by five layers of sclerenchyma and two or more layers of parenchyma containing numerous calcium oxalate crystals (Figure 2v). The odor and taste of star anise are anise-like. The plant starts fruiting at the age of 6 years and continues until it is over 100 years old. Fruits are collected before full maturity and dried in the sun.

Vanilla

Vanilla is the cured, fully grown, unripe fruit of *Vanilla planifolia* Andr. (Orchidaceae). The fruits are linear and flattened (120–350 mm by 5–9 mm), and both ends taper; the fruit occasionally splits into three parts in the tip; fruits are externally dusky to moderate brown, longitudinally wrinkled and covered

with minute crystals of vanillin; many minute seeds are embedded in a brownish black, balsamic fluid (Figure 2w and x).

In transverse section, the outer epidermis of the epicarp shows red to brown bodies, some with calcium oxalate and vanillin crystals. The hypodermis is composed of collenchymatous cells with dark-colored contents, and the sarcocarp consists of loosely arranged, large deeply undulate parenchyma, becoming smaller on the inner side; most cells contain brownish contents and oily globules, and some have raphides of calcium oxalate up to 400 μm long. There is an interrupted circle of leptocentric fibrovascular bundles surrounded by a few spiral or pitted tracheae. The cells of the inner epidermis of the endocarp are elongated into numerous, thin-walled, nearly straight, glandular hairs containing a gummy, resinous secretion (Figure 2x and y).

At the time of harvest, in the vanilla fruit or pods, the normal aromatic components forming the characteristic vanilla flavor do not exist in their final form, but appear only following a curing (fermentation) process involving enzymatic actions on glucosides, the most important of which is glucovanillin, which produces vanillin (the main aromatic component of vanilla) and sugar as a result of the action of glucosidase. Similarly, aldehydes, protocatechoic acid, benzoic acid, vanillic acids, anisic alcohol, etc. are also formed. These different compounds impart subtlety to the fragrance of natural vanillin.

The different stages of the curing methods are as follows:

1. Cessation of the vegetative life of the bean to allow the onset of enzymatic reactions.
2. Raising of the temperature to promote this action and to stimulate at first rapid drying, thereby preventing the appearance of harmful fermentation.
3. Slower drying followed by the production of different fragrance components.
4. Conditioning of the product, during which the procedures for commercial presentation and to obtain a good preservation are carried out.

The improved method involves sorting or grading, scalding or blanching, autoclaving, loading, and mechanically drying in a tunnel dryer, followed by drying in the shade. After scalding and autoclaving, the vanilla are chopped into 20- or 30-cm pieces. This leads to an acceleration of the oxidative and enzymatic reactions. The cuts are put through the dryer at a temperature of 65 °C each day. As soon as they come out, they are put into an isothermal chest for the next 24 h, with the temperature being maintained at around 50 °C. These operations are repeated for 12 days, after which a product containing 20–25% moisture is obtained, with a yield of about 4.5 to 1.0. The transfer to the isothermal chests after the tunnel drying is vital, in addition to equilibrate their moisture content. The process requires a series of sheds or buildings for 'sunnings' and 'sweatings' and 'mahogany sweating boxes' or 'aging boxes.' The total process takes about 5–6 months, including aging. Finally, the beans are graded, bundled, and packed in special tins lined with waxed paper.

The best-grade beans are very dark, chocolate brown in color, and oily in appearance with a strong vanilla odor. Vanilla should be stored in a cool, dry place in an airtight container to stop the beans becoming brittle.

Fenugreek

Fenugreek is the dried seed of *Trigonella foenum-graecum* L. (Fabaceae). The seeds are yellowish brown, oblong–rhomboidal (2.5–5 mm by 2–3.5 mm) with an oblique furrow on one side, smooth, hard, and mucilaginous. The hilum is in a notch, the raphe short, the endosperm translucent, and the embryo waxy and yellow (Figure 3a–d).

In cross-section, the spermoderm consists of an outer cuticle followed by a layer of palisade cells (measuring 65–70 μm by 8–20 μm) with flat bases and pointed apices, a layer of hourglass cells and four to five layers of parenchyma. The endosperm consists of an outer layer of cells containing aleurone grains, followed by several layers of thin-walled mucilaginous cells; the embryo cells contain aleurone grains and fixed oil (Figure 3d).

Fenugreek possesses a strong, pleasant, sweetish odor and tastes mucilaginous, farinaceous, and slightly bitter. The pods split easily and are threshed on the floor with sticks to free the seeds. The seeds are then winnowed, cleaned, and further dried in the sun. Dried fenugreek seeds can be stored for about 2 years.

Mustard

Mustard is composed of the dried, mature seeds of the following *Brassica* L. species (Brassicaceae): *B. nigra* (L.) Koch (black mustard), *B. alba* Boiss. (white or yellow mustard), and *B. juncea* (L.) Czern & Coss. (Rye or Indian mustard).

Black mustard seeds are globular, 1–1.6 mm in diameter, dark brown to nearly black, minutely reticulate, and mucilaginous. The kernel is greenish yellow and oily (Figure 3e and f). White mustard seeds are spherical, about 2 mm in diameter, minutely

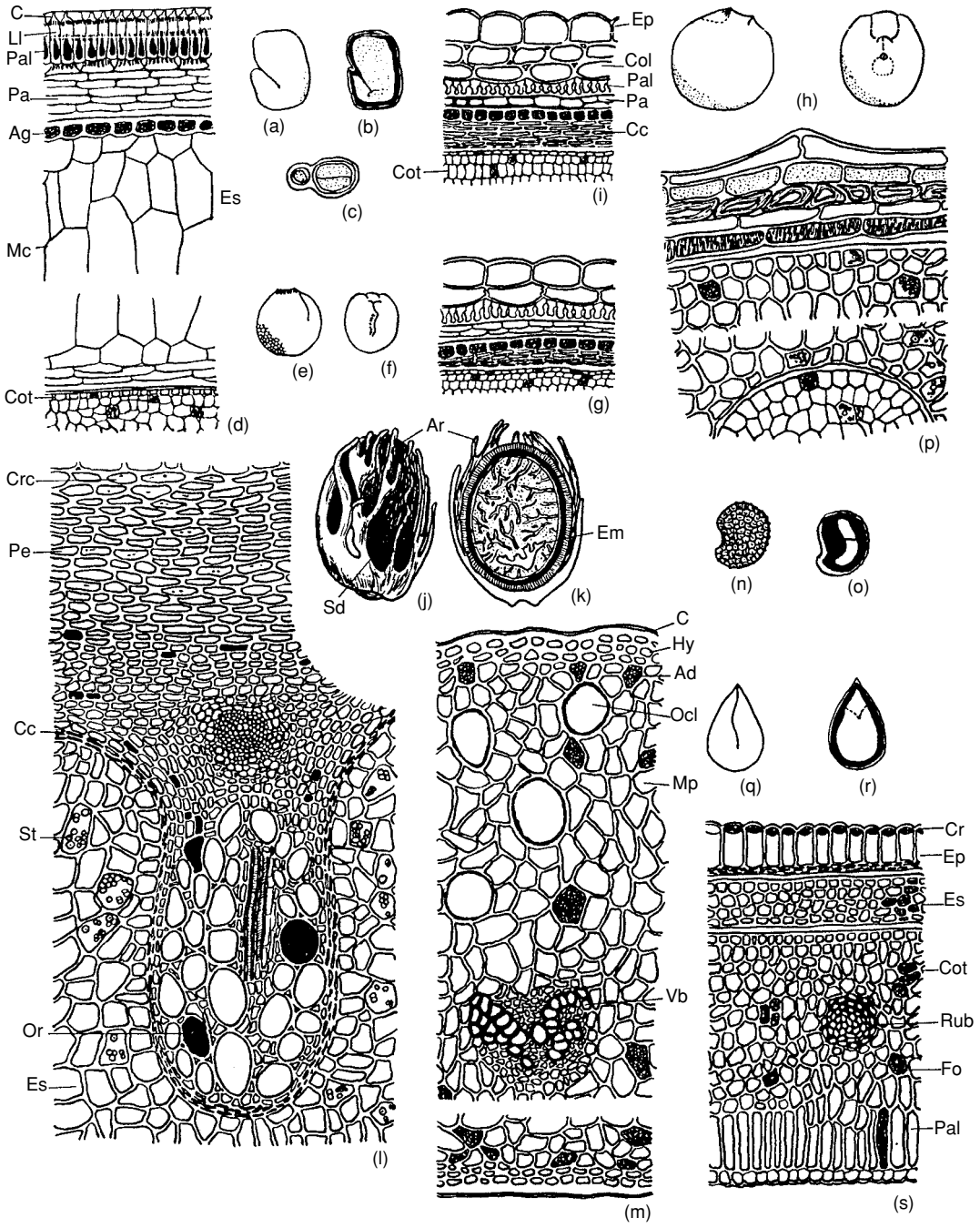

Figure 3 Spices from seeds. **Fenugreek:** (a) seed; (b) longitudinal section of seed; (c) transverse section of seed; (d) cross-section through seed coat, endosperm and part of cotyledon. **Mustard black:** (e, f) seed; (g) cross-section through seed coat, endosperm and part of cotyledon; (h) **Mustard, white or yellow:** (i) cross-section through seed coat, endosperm and part of cotyledon. **Nutmeg and mace:** (j) nutmeg surrounded by mace; (k) same as (j), cut longitudinally; (l) cross-section through perisperm and endosperm of nutmeg; (m) cross-section through mace. **Poppy:** (n, o) seed; (p) cross-section through seed coat, endosperm and part of cotyledon. **Sesame:** (a, r) seed; (s) cross-section of seed. Reproduced from Spices and Flavoring (Flavouring) Crops: Fruits and Seeds, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press. Abbreviations: Ad, amylodextrin granules; Ag, aleurone grains; Ar, aril; Bc, beaded cells; Bel, beaker cells; C, cuticle; Cc, compressed cells; Ch, chloroplast, Col, colenchyma; Cot, cotyledon; Cp, carpophore; Cr, crystal; Crc, crystal cells; Cro, cross cells; Dl, dehiscence line; Em, embryo; End, endocarp; Ep, epidermis; Es, endosperm; Exo, exocarp; Fl, fiber layer; Fo, fixed oil; Gc, giant cell; H, hair; Hc, hourglass cells; Hp, horny protein; Hy, hypodermis; Li, light line; Mc, mucilage cells; Mes, mesocarp; Mp, mesophyl; O, oil; Oc, oil cavity; Ocl, oil cells; Or, oleoresin; P, pedicel; Pa, parenchyma; Pal, palisade; Pe, perisperm; Pl, placenta; R, raphe; Rc, reticulated cells; Re, resin; Rub, rudimentary vascular bundle; Sc, seed coat; Scl, sclerenchyma; Sd, seed; Se, secretory cells; St, starch; Stc, stone cells; V, vittae; Vb, vascular bundle.

pitted, and mucilaginous (Figure 3h). Indian mustard seeds resemble black mustard but are larger and brown in colour.

The testa of black mustard in transverse section consists of an outer epidermis of polygonal tubular cells, 40–50 μm long, containing mucilage; hypodermis consists of empty, polygonal cells, 10–105 μm by 60–90 μm; the inner epidermis of testa consists of slightly lignified, palisade cells (unequal in height) followed by cells of tegmen containing tannin. The endosperm has an outer aleurone layer of thick-walled cells containing fixed oil and aleurone grains and an underlying layer of collapsed cells. Embryo cells are polyhedral and thin-walled, containing fixed oil and aleurone grains (Figure 3g). White mustard in transverse section resembles black mustard but differs in that it has two layers of somewhat collenchymatous cells beneath the epidermis of the testa; the palisade cells have no tannin (Figure 3i).

Black mustard is slightly odorous, and tastes acrid and pungent when crushed and moistened. White mustard seed is not pungent. The fruits are harvested when they begin to turn yellow. Then, they are threshed by beating with wooden mallets. The threshed seeds are separated from the husks by winnowing and then dried in the sun for a couple of days. Seeds are stored in seed bins or gunny bags. Storage of the whole seed for 12 months reduces the volatile oil content by 54%, but for ground seeds, the loss is 84%.

Nutmeg and Mace

Nutmeg is the dried kernel, and mace is the dried aril of the seed of *Myristica fragrans* Hout. (Myristicaceae). The kernel constitutes 16–17% and mace constitutes 2.0–2.5% by weight of the fruit. Nutmegs are ovoid (2–3.5 cm by 1.5–3 cm) and the kernels are grayish brown externally and reticulately marked with small furrows. Numerous minute, dark reddish brown points and lines, constituting a thin layer of perisperm, grow inwards into the endosperm at the furrows; the hilum lies in a small, circular depression at one end (Figure 3j and k). Mace consists of flattened lobed pieces about 25 mm or more in length, somewhat less in breadth, 1 mm thick, buff or light brown, translucent, smooth, horny and brittle (Figure 3j and k).

In longitudinal section, nutmeg has a perisperm about 0.12 mm thick. The cells of the peripheral perisperm are flattened and polyhedral; some contain prisms of an unidentified crystalline substance. The cells of the inner perisperm are parenchymatous, and grow into endosperm, blunt-ended ruminations, each containing a vascular strand and numerous volatile oil-containing cells. The endosperm contains abundant bunches of needle-shaped, fat crystals, small, irregular aleurone grains with large crystalloid and rounded (up to 22 μm in diameter) simple or compound starch grains and scattered tannoid cells (Figure 3l).

In transverse section, mace has the following: anouter epidermis of irregularly shaped, thick-walled cells covered by cuticle; a hypodermis consisting of a layer of thick-walled cells and one or more additional layers of comparatively thin-walled cells; a parenchymatous mesophyl consisting of densely packed amylodextrin granules, numerous large oleoresin cells and vascular tissues (Figure 3m).

Nutmeg and mace have a characteristic nut-like aroma and taste. The volatile oil content of nutmeg and mace is 15 and 20%, respectively. The fruits are harvested when split on the tree or, more usually, gathered after they have fallen to the ground. Mace is separated from the pericarp by hand or between boards, without breakage (which reduces its quality), and dried in the sun or artificially. Drying of mace and its storage is a delicate process. Mace has to be dried to a 5% moisture level, for which it takes about 16 h under open sun, 24–28 h in the shade, and 5–6 h in mechanical drying. A mild blanching followed by sulfiting helps in the retention of color and quality of mace. Nutmegs are dried in their shells on large wooden trays and turned daily for uniform drying and to avoid fermentation. Artificial drying between 35 and 40 °C prevents cracking of the shells and any melting of fat in the kernels. Various traditional methods are practiced to crack the nuts, which are stored and graded according to size and quality. This involves immersing the nutmegs in water and separating unhealthy or floating ones; the sound ones are drained, air-dried, and stored in bins before packing. Introducing a little carbon disulfide prevents insect attack.

Poppy Seeds

Poppy seeds are obtained from *Papaver somniferum* L. (Papaveraceae) capsules. The seeds are white, yellow, brown, bluish or nearly black, and subreniform (1–1.25 mm long), with a reticulate surface; the hilum and micropyle are situated in a depression near one end (Figure 3n and o).

In cross-section, the seed consists of an outer epidermis of tangentially elongated, thin-walled cells, a layer of parenchymatous cells containing numerous very small prismatic crystals of calcium oxalate, a layer of thick walled cells, a layer of parenchymatous cells, and a layer of thick-walled reticulate cells. The cells of the endosperm and embryo are polygonal and

contain aleurone grains and globules of fixed oil (Figure 3p).

The odor and taste of the seeds are pleasant and nut-like. For the collection of the seeds, the capsules are dried on the plant for about 15 days. After plucking by hand, these are threshed, and the seeds are separated and stored.

Sesame

Sesame is the seed of *Sesamum indicum* L. (Pedaliaceae). The seeds are yellowish white, brown, gray to black, flat pyriform (2.5–3.5 mm by 1.2–2 mm by 1 mm), and finely punctate, with four delicate, longitudinal ridges at the edges of the flat faces. The raphe extends from the hilum at the pointed end as a line along the center of one flat face to a broader end, and the scanty endosperm surrounds a large embryo. The hulled seeds are cream or pearly white in color (Figure 3q and r).

In cross-section, the seed coat consists of an epidermis of radially elongated cells (30–60 μm on sides and 45–130 μm at ridges), with thin, cutinized outer walls; each cell contains a 'rosette' calcium oxalate crystal; the rest of the testa consists of collapsed cells containing numerous, prismatic calcium oxalate crystals, and the endosperm and cotyledons consist of polygonal parenchymatous cells containing fixed oil and small aleurone grains measuring 2–10 μm (Figure 3s).

The seeds have a faint nutty odor and taste, and are ripe at the end of about 5 months. The whole plants are then cut down and made into stooks to dry, and the seeds are shaken out from the capsules and stored.

See also: **Drying**: Theory of Air-drying; Drying Using Natural Radiation; **Sensory Evaluation**: Aroma; Taste

Further Reading

Bagchi GD and Srivastava GN (1989) SEM of epicarp surfaces of some medicinally important Apiaceae. *International Journal of Crude Drug Research* 27: 171–177.

Chadha KL and Rethinam P (1994) *Advances in Horticulture: Plantation and Spice Crops Part 182*, vols. 9–10. New Delhi: Malhotra Publishing.

Parry JW (1969) *Spices*, vols. I and II. New York: Chemical Publishing.

Purseglove JW, Brown EG, Green CL and Robbins SRJ (1981) *Spices*, vols. I and II. London: Longman.

Vaughan JG (1970) *The Structure and Utilization of Oil Seeds*. London: Chapman & Hall.

Wallis TE (1960) *Textbook of Pharmacognosy*. London: Churchill.

Leaf and Floral Structures

G D Bagchi and G N Srivastava, Central Institute of Medicinal and Aromatic Plants, Lucknow, India

Introduction

The aromatic leaves and floral structures of more than 30 plants are commonly used as spices and for flavoring purposes. Leaves and floral structures have delicate flavors, and these spices require proper care, particularly during the drying process. Leaf spices are also used fresh. The diagnostic morphological and anatomical characters of the most commonly used spices are dealt with in detail, along with their common names, botanical names, family and parts used. Handling, drying, and storage details are given wherever available. Other spices, which have specific, but equally important, uses are summarized in **Table 1**.

Chervil

Chervil consists of the dried leaves of *Anthriscus cerefolium* Hoffm. (Apiaceae). The leaves are decompound and oval, and the leaflets alternate, thin pale green 2–5 × 2–4 cm and pinnate; a few hairs occur on the mid-vein and the margins of leaflets (**Figure 1a**).

In surface view, the lower epidermal cells have more wavy walls than those on the upper epidermis; stomata are numerous on the lower epidermis, but absent on the upper; hairs, usually unicellular and pointed, are on both surfaces of the main vein and the margins, but more arise on the lower surface; the mesophyll consists of a row of palisade cells, 18–75 × 16–35 μm in size, followed by spongy parenchyma containing vascular bundles; an oil duct is found below the mid-vein (**Figure 1b** and **c**).

The odor of chervil somewhat resembles that of caraway and anise. Leaves, after collection, are then dehydrated by circulating hot air (50 °C) and are then fragmented mechanically and packed in plastic–lined bags. (*See* **Drying**: Theory of Air-drying.)

Laurel or Sweet Bay

The leaves are derived from *Laurus nobilis* L. (Lauraceae). They are lanceolate-acuminate, 10 cm or more long, coriaceous, pellucid–punctate, and with revolute, entire margins; the upper surface is glabrous and shiny, olive green to brown; the lower surface is dull olive to brown with a prominent mid-rib and veins (**Figure 1d**).

Table 1 Some other commonly used spices from leaf, flower, and flower parts

Botanical name	*Common name*	*Family*	*Part used*	*Characters*
Allium ascalonicum L.	Shallot	Liliaceae	Bulb	Bulbs about the size of a nut, white, leaves basal; bulbs may be taken up and dried or allowed to remain in the ground and dug up when needed; after lifting, they are skinned, washed and tied in bunches for marketing
Allium cepa L.	Onion	Liliaceae	Bulb	Bulbs are variable in size: small bulbs are silvery, and large bulbs are red or yellow in color; bulbs are harvested when mature, which is indicated by the tops drooping just above the bulb, while the leaves are still green; onions are stored in crates, before storage, they are thoroughly cured
Allium fistulosum L.	Leek	Liliaceae	Leaf	Leaves terete and hollow, usually clustered
Allium sativum L.	Garlic	Liliaceae	Bulb	Compound bulb, subglobose, has eight to 15 bulblets, bulblets ovoid, three- to four-sided, covered with whitish scale leaf which encloses whitish flesh, one or two translucent scales, in the center, two or three yellowish to yellowish green leaves present; harvested when tops turn yellow, exposed to sun for 4–5 days and winnowed; after a fortnight of curing, they are packed in bags and stored
Allium schoenoprasum L.	Chive	Liliaceae	Leaf	Leaves 10–25 cm, terete or grooved above, scape 15–35 cm; the cut leaves are gathered into bunches of about 2.5 cm diameter, secured and packed in layers in trays
Angelica archangelica L.	Angelica	Apiaceae	Leaf and young stem	Leaves ovate, 30–90 cm, two- to three-pinnate, ultimate pinnae toothed; leaves cut for drying and freezing in early summer; the stems should be harvested before the plant flowers
Cinnamomum tamala Nees & Eberm	Tejpat (Indian cassia lignea leaf)	Lauraceae	Leaf	Leaves subcoriaceous, opposite, triple nerved rarely alternate, elliptic oblong, acute base, acuminate, glabrous, shining, leaves are collected and dried in sun for 3–4 days and packed in a cylindrical bamboo nets for marketing
Gaultheria fragrantissima Wall.	Wintergreen	Ericaceae	Leaf	Leaves alternate, oblong-lanceolate to elliptic rhomboid, 7–13 × 2–6 cm, serrate, stiffly coriaceous, glands dotted
Hyosopus officinalis L.	Hyssop	Lamiaceae	Leaf	Leaves sessile, 13 × 2–8 mm, opposite, linear or lanceolate, entire, hairy on both surfaces; harvested just as flowers begin to bloom
Levisticum officinale Koch	Lovage	Apiaceae	Leaf and flowering top	Lower leaves large with long petioles, up to three pinnate, about 70 cm long, uppermost leaves entire and lanceolate; flowers yellow in stiff umbels; fresh leaves are generally used; drying of thick leaves takes 4–5 days
Marrubium vulgare L.	Farasiyum	Lamiaceae	Leaf	Leaves opposite, petiolate, downy, oval, up to 5 cm long, surface wrinkled and has protuberances, margin dentate; harvested before full flowering and dried in the shade below 35 °C
Melissa officinalis L.	Balm	Lamiaceae	Leaf	Leaves opposite, ovate or triangular; up to 3.5 cm long, hairy, margin broadly dentate; usually, fresh leaves are used; for drying, stems are cut and tied in small bunches and hung in airy space.
Mentha pulegium L.	Pennyroyal	Lamiaceae	Leaf	Leaves shortly petioled, up to 1.25 cm long, round to oval, entire or slightly crenate and pubescent; leaves are used fresh or in dried form
Murraya koenigii L. Spreng.	Curry leaf	Rutaceae	Leaf	Leaves imparipinnate, leaflets nine to 25, ovate-lanceolate, dentate, acuminate, base oblique, glabrous above, pubescent beneath, gland dotted; fresh or shade dried leaves are used
Nepeta cataria L.	Catnip	Lamiaceae	Leaf	Leaves ovate, coarsely crenate; harvested when in full bloom

Continued

Table 1 Continued

Botanical name	*Common name*	*Family*	*Part used*	*Characters*
Pimenta racemosa (P. Miller) J. W. Moore	West Indian bay	Myrtaceae	Leaf	Leaves obovate or elliptic, leathery, shining above; during harvesting, either leaves are stripped from the trees or both shoots and leaves are removed and tied in bundles for drying
Ruta graveolens L.	Rue	Rutaceae	Leaf	Leaves alternate, 15 cm long, two to three pinnate, oblong to spathulate, with many small oil glands; leaves are usually used fresh
Salvia sclarea L.	Clary, clary sage	Lamiaceae	Leaf	Leaves petiolate, opposite, ovate, apex obtuse, base cordate, crenate margin, thinly hairy above
Tanacetum vulgare L.	Tansy	Asteraceae	Leaf and flowering top	Leaves downy, one to two pinnate, margin dentate, capitula in flat umbels, golden yellow, about 1 cm in diameter; collected when in flower and dried in the shade at below 35 °C

In cross–section, the leaf has a thick cuticle, and the epidermal cells, in surface view, have sinuous, pitted and thick walls; the lower epidermal walls are more curvilinear and distinctly beaded; the lower epidermis alone bears stomata arranged singly or in pairs; the mesophyll is differentiated into two layers of palisade parenchyma, a region of spongy parenchyma containing scattered spheroidal oil reservoirs, fibrovascular, and collenchyma tissues; the latter occurs above and below the vascular bundle of mid-rib, but not in the veins (Figure 1e–h).

The leaves are characteristically fragrant when crushed but taste bitter and aromatic. (*See* **Sensory Evaluation**: Aroma; Taste.)

Marjoram (Sweet)

Marjoram consists of dried leaves, with or without flowering tops, of *Majorana hortensis* (L.) Moench (Lamiaceae). The pinnate leaves are obovate to broadly elliptical, 21×11 mm in size with entire margins and short petioles; both surfaces are gray–green, glandular punctate, and pubescent. The floral parts are: bract – light grayish yellow, obovate, thin, up to 5.5×3.5 mm; calyx – gamosepalous, about 3 mm long with five prominent teeth; corolla – gamopetalous, white to pink or pale lilac, about 3.5 mm long (Figure 1i).

In a transverse section and/or surface view, it is possible to see an epidermis with a thin cuticle, one to two palisade layers and upper and lower epidermal cells with slightly wavy, knotty, vertical walls. Glandular and nonglandular hairs occur on both surfaces, and the latter are uniseriate, two to five cells in length and more or less bent inwards at apex; some are papillose. Glandular hairs are of two types, one with a one- to two-celled stalk and a one- to two-celled glandular head, the other with no stalk and a large, eight-celled head; the latter is sunk in a depression in the epidermis; stomata occur on the lower epidermis with two 'neighboring' cells (Figure 1j and k).

The odor and taste of sweet marjoram are aromatic, characteristic, and slightly pungent; the plants are harvested as soon as they flower. The plants are cut and tied in bunches and dried in the open air, or spread on wire trays in ventilated rooms and dried by circulating warm air. (*See* **Drying**: Drying Using Natural Radiation.)

Mint

Mint consists of the dried leaves and flowering tops of *Mentha piperita* L. (peppermint) and *M. spicata* L. (spearmint) belonging to the family Lamiaceae. *M. piperita* is a hybrid between *M. spicata* and *M. aquatica* since *M. spicata* itself is a hybrid. Therefore, *M. piperita* is a triple hybrid. It is an allohexaploid, $2n = 6x = 72$. However, this species has a variable chromosome number, which ranges from $2n = 36$ to $2n = 134$ with polyploid and anueploids. *M. spicata* ($2n = 48$) has originated by hybridization between *M. longifolia* and *M. suaveolens* and doubling of chromosomes. The resultant polyploid behaves as allotetraploid. Peppermint leaves are dark green, about 4.5×2.3 cm, petiolate, ovate and with serrate margins. The apex is acute and the midrib prominent, and dot-like oil glands are visible with a hand lens. Nonglandular hairs are present on the lower epidermis over the mid-rib and veins at margins, and a few arise on the mid-rib of the upper epidermis. Verticillaster inflorescences are usually arranged in compact, oblong or oval terminal spikes: calyx – gamosepalous, purplish, about 3 mm long, glandular punctate and five-toothed; corolla – short, tubular, light purple and four-cleft; stamens – four; style with bilobed stigma (Figure 1l). Spearmint differs from peppermint in having a sharply pointed, long, narrow spike, an absence of nonglandular hairs in the upper epidermis, fewer and shorter hairs on the

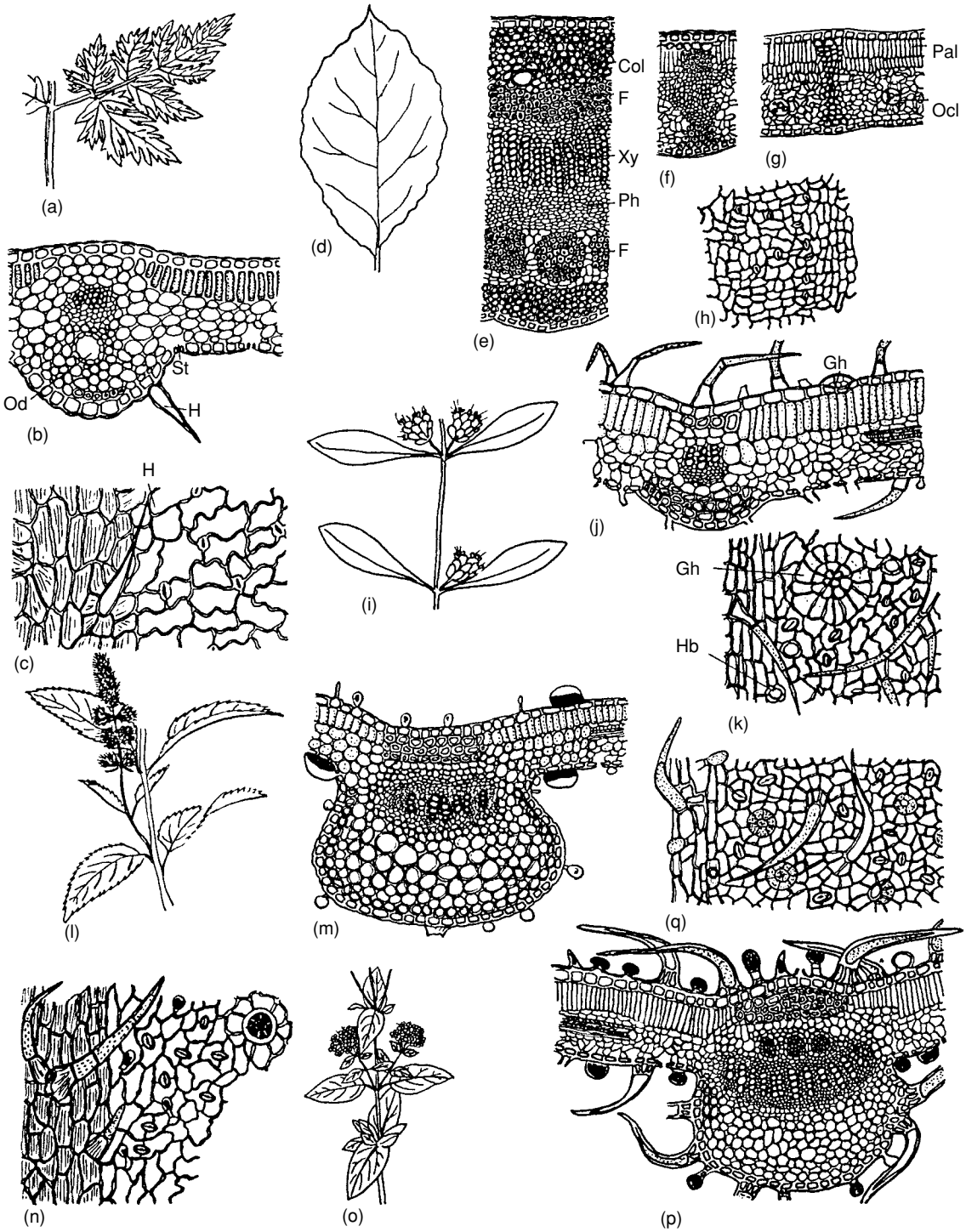

Figure 1 Spices from leaves and flower tops. **Chervil**: (a) leaf; (b) cross-section of leaflet through mid-rib; (c) lower epidermis in surface view. **Laurel**: (d) leaf; (e) cross-section of leaf through mid-rib; (f) vein; (g) leaf blade; (h) lower epidermis in surface view. **Marjoram**: (i) a portion of flowering twig; (j) cross-section of leaf through mid-rib; (k) surface view of lower epidermis; **Peppermint**: (l) a portion of flowering twig: (m) cross-section through mid-rib of leaf; (n) surface view of lower epidermis. **Origanum**: (o) a portion of flowering twig: (p) cross-section through mid-rib of leaf; (q) lower epidermis in surface view. Reproduced from Spices and Flavoring (Flavouring) Crops: Leaf and Floral Structures, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press. Abbreviations: C, calyx; Col, collenchyma; Cor, corolla; Cp, compressed parenchyma; E, endocarp; F, fiber; Gh, glandular hair; H, hair; Hb, hair base; Hy, hypodermis; Mes, mesophyl; O, ovary; Oc, oil cavity; Ocl, oil cell; Od, oil duct; P, pericarp; Pal, palisade; Pap, papilla; Pg, pollen grain; Ph, phloem; Pp, pigmented parenchyma; R, receptacle; Rc, reticulate cells; S, stigma; Sc, stone cells; St, stomata; Sta, stamen; Sty, style; Vb, vascular bundle; Xy, xylem.

lower surface over the mid-rib, with few, if any, over the veins, and the absence of crystals in the glandular hairs.

In cross-section, peppermint leaves show a cuticle covering both the upper and lower epidermis; upper epidermal cells having moderately thicker outer walls than the lower epidermal cells; glandular hairs containing groups of crystals; a mesophyll consisting of a layer of palisade cells and spongy tissue; between the upper epidermis and the vascular tissue of the mid-rib and veins is a conspicuous group of collenchyma cells. In the surface view, the cuticle is striated over the mid-rib, veins and margins, and upper epidermal cells are irregular with wavy walls, except in the latter regions; numerous small crystals are present in the cells of upper and lower epidermis (**Figure 1m** and **n**).

Peppermint has a characteristic aromatic odor and taste, with a cooling sensation on the breath, whereas the odor and taste of spearmint are characteristic and aromatic but not followed by a cooling sensation. Mints are harvested, when in bloom, in bright sunny weather, by cutting with a sickle 2–3 cm above the ground. The harvested herb is allowed to dry partially in the open. Drying is completed by circulating warm air, which avoids overheating and obviates drying for too long. After attaining the desired brittleness, the leaves are stripped from the stem and completely dried. Afterwards, they are fragmented mechanically or through wire sieves.

Origanum or Oregano

This consists of the dried leaves and flowering tops of *Origanum vulgare* L. (Lamiaceae). The leaves are light green to greenish brown, with crenate to crenate–serrate margins and are about 3 cm long and glandular punctate. The flowers are in spike-like clusters: bracts – broadly ovate, 2.5–3 × 2 mm with prominent hairs on the outer surface; calyx – gamosepalous four-toothed and deeply clefted on two sides; corolla – 5 mm long, gamopetalous; stamens – four, epipetalous; style – plumose (**Figure 1o**).

In surface view and/or transverse section, the epidermal cells of both surfaces of the leaf are irregular in shape with uneven and thickened walls; both surfaces have numerous curved, pointed, unicellular, nonglandular hairs up to 900 μm long; the surfaces of some hairs are smooth, whereas many have papillose surfaces. Uni- to multicellular glandular hairs are present on both surfaces (**Figure 1p** and **q**).

Origanum has a strong camphoraceous odor and a warm, pungent, bitter, camphoraceous taste. The plant is collected when in flower and dried in the shade at less than 35 °C.

Parsley

Parsley consists of the dried leaves of *Carum petroselinum* Hoffm. (Apiaceae). The leaves are pinnately compound of three-lobed leaflets; each lobe is divided into smaller, tooth-like segments that are usually curly (**Figure 2a**).

The upper and lower epidermal cell walls are, in surface view, sinuate, and stomata are few in the upper and numerous in the lower epidermis. In cross-section, beneath the upper epidermis, the mesophyll consists of a layer of palisade cells about 40–55 μm long followed by spongy parenchyma; below the vascular bundle of the mid-rib is an oil duct (**Figure 2b** and **c**).

The dried leaves have a characteristic, agreeable, aromatic odor and taste. Fresh leaves are dried rapidly under controlled conditions of heat and air flow to avoid loss of essential oils, and to preserve a good, green color. (*See* **Essential Oils**: Properties and Uses.)

Rosemary

Rosemary consists of the dried leaves of *Romarinus officinalis* L. (Lamiaceae). The leaves are about 3.5 cm long and 2–4 mm wide, numerous, opposite sessile, linear, leathery, entire and slightly glossy with revolute margins. The upper surface is dark green, whilst numerous branched trichomes on the lower surface make it gray and wooly and glandular punctate with a prominent mid-rib (**Figure 2d**).

In surface view, the cells of the upper epidermis are polygonal with unevenly thickened and beaded walls, whilst the lower epidermal cells are larger and thin-walled. In cross-section, the collenchyma tissue present below the upper epidermis forms a thick-walled hypodermal layer; collenchymatous cells also surround the vascular bundle of mid-rib. One to three layers of palisade cells are up to about 80 μm in length and are followed by a zone of spongy parenchyma. A dense, tangled mass of branched, multicellular, smooth hairs covers the lower epidermis, and the tips of the hairs are pointed and curved; glandular hairs are uni- to multicellular. In the mid-rib vascular bundle, the vessels are arranged in radial rows (**Figure 2e** and **f**).

The odor of rosemary is characteristically aromatic, having a slight camphoraceous note; the taste is pungent, aromatic, bitter, and camphoraceous. Rosemary is harvested when flowering has started. At that time, 30–50 cm of the top along with leaves and flowers are harvested with sickle or by forage harvester. Care should be taken not to cut the woody portions of the stem. The first harvest is obtained approximately 6 months after planting, and

Figure 2 Spices from leaves and flower tops. **Parsley**: (a) a portion of flowering twig; (b) cross-section leaflet through main vein; (c) surface view of lower epidermis. **Rosemary**: (d) a portion of flowering twig; (e) cross-section through mid-rib of leaf; (f) lower epidermis in surface view. **Sage**: (g) a portion of flowering twig; (h) cross-section through mid-rib of leaf; (i) cross-section of a portion of leaf blade; (j) lower epidermis in surface view. **Savory**: (k) a portion of flowering twig; (l) cross-section through mid-rib of leaf; (m) cross-section through leaf blade; (n) lower epidermis in surface view. **Sweet basil**: (o) a portion of flowering twig; (p) cross-section of leaf through mid-rib; (q) cross-section through side vein; (r) surface view of upper epidermis. Reproduced from Spices and Flavoring (Flavouring) Crops: Leaf and Floral Structures, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press. Abbreviations: C, calyx; Col, collenchyma; Cor, corolla; Cp, compressed parenchyma; E, endocarp; F, fiber; Gh, glandular hair; H, hair; Hb, hair base; Hy, hypodermis; Mes, mesophyl; O, ovary; Oc, oil cavity; Ocl, oil cell; Od, oil duct; P, pericarp; Pal, palisade; Pap, papilla; Pg, pollen grain; Ph, phloem; Pp, pigmented parenchyma; R, receptacle; Rc, reticulate cells; S, stigma; Sc, stone cells; St, stomata; Sta, stamen; Sty, style; Vb, vascular bundle; Xy, xylem.

subsequent harvests can be obtained every 4 months. Harvested leaves are dried in the shade.

Sage

Sage consists of the dried leaves of *Salvia officinalis* L. (Lamiaceae). The leaves are grayish green, oblong–lanceolate, about 8.5 × 2.5 cm, petiolate, apex acute, margins finely crenate, densely pubescent and glandular punctate with glands obscured by hairs (Figure 2g).

The upper epidermal cells, in surface view, have few stomata and straight, beaded walls, whereas the lower epidermis bears numerous stomata, and the cells have wavy, beaded walls; numerous nonglandular and glandular hairs occur on both the surfaces. The nonglandular hairs are uni- or multicellular, and the multicellular hairs are more numerous, very long, narrow pointed and mostly smooth, but some have a papillose surface; glandular hairs are also uni- or multicellular. The mid-rib vascular bundle is surrounded by parenchymatous and collenchymatous cells, and the mesophyll consists of one to three layers of palisade cells followed by spongy tissues (Figure 2h–j).

Sage is aromatic in odor and somewhat bitter in taste. The leaves are gathered for use while the plants are in flower and are dried in the shade or indoors by circulating warm air.

Savory

Savory consists of dried leaves and flowering tops of *Satureja hortensis* L. (Lamiaceae). The leaves are green to grayish green, nearly sessile, entire, linear oblong to spathulate when unfolded, glandular punctate, acute to rounded at the apex, tapering at the base, 12–35 × 1–2 mm and pubescent. The flowers are tubular–bilabiate, purple and up to 5 mm long in leaf axils; calyx – campanulate, hispid on nerves and with five subulate teeth; corolla – about 4–5 mm long; stamens – four and didynamous; style with bilobed stigma (Figure 2k).

In the surface view, epidermal cell walls of the upper and lower surfaces are sinuate and beaded; the stomata are elliptical. The nonglandular hairs are uniseriate and numerous, with up to six cells and a pointed, distal cell; the glandular hairs are numerous and have an up to 12-celled head and one-celled stalk (Figure 2l–n).

The odor of savory is characteristically aromatic, and the taste is aromatic and warm. The leaves are collected when the plants are in flower and are dried in the shade, or indoors by circulating warm air, before packaging.

Basil (Sweet)

Sweet basil consists of the dried leaves of the *Ocimum basilicum* L. (Lamiaceae). The leaves are ovate, tip acute, petiolate, green, up to 5 cm long and finely serrate. Hairs arise on the veins, especially on the underside of the leaf, which is also glandular punctate (Figure 2o)

In surface view, the upper and lower epidermis have wavy cell walls, and the nonglandular hairs are one- or multicelled, uniseriate, with pointed apices; the surfaces of the hairs are finely warty. Glandular hairs have a two- to four-celled head, and occur on both surfaces but more on the upper surface. In transverse section, a single palisade layer lies beneath the upper epidermis, followed by a zone of spongy parenchyma; the vascular tissue of the mid-rib and large veins is separated from the upper and lower epidermis by parenchymatous and collenchymatous tissues (Figure 2p–r).

The odor of basil is fragrantly aromatic, and the taste is warmly aromatic and pungent. Fresh leaves can be dried in the open air but more efficiently indoors by controlled artificial heat and circulating air. Dried leaves are fragmented by machines and graded fine, medium, or coarse before packaging.

Tarragon

Tarragon consists of the dried leaves of *Artemisia dracunculus* L. (Asteraceae). The leaves are lanceolate, entire, sessile, glabrous, and bright green. The edges of dried leaves curve towards the lower surface, and are finely revolute, 4 × 3.5 mm in size (Figure 3a).

The cuticle on both surfaces is striated over the mid-rib and veins. The epidermal cell walls of both surfaces are undulate; occasionally, small, inconspicuous glandular hairs of 35–40 μm diameter are present on the epidermis. The mesophyll consists of two rows of palisade cells towards the upper surface and a single row towards the lower. Below the mid-rib vascular bundle are a few collenchymatous cells, and on the upper side, two secretory ducts are present (Figure 3b and c).

The odor and taste of tarragon are agreeable, aromatic, and like that of anise. The leaves are dried indoors by artificial heat and circulating air to insure retention of aroma and color. After being crushed by machine into particles of a suitable size, the product is packed in bags and kraft paper.

Thyme

Thyme consists of the dried leaves and flowering tops of *Thymus vulgaris* L. (Lamiaceae). The leaves are linear/linear–lanceolate, ovate or oblong, up to 6 × 0.5–2 mm, apex acute, base obtuse tapering into

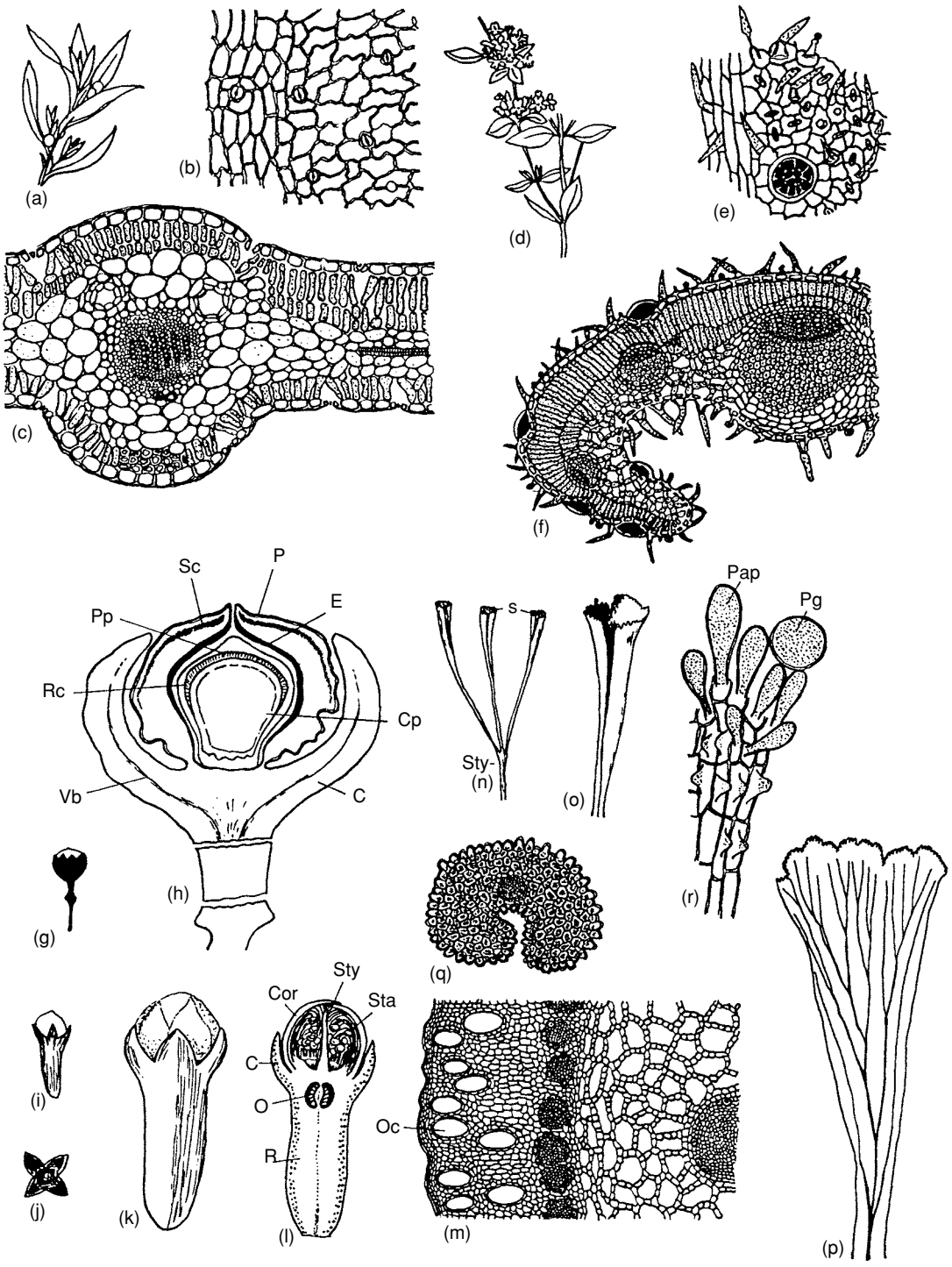

Figure 3 Spices from leaves, flower tops and flowering parts. **Tarragon**: (a) a portion of flowering top; (b) surface view of upper epidermis; (c) cross-section of leaf through mid-rib. **Thyme**: (d) a portion of flowering twig: (e) surface view of lower epidermis; (f) cross-section of leaf blade. **Cassia bud**: (g) bud; (h) longitudinal section through bud. **Clove**: (i,k) whole clove; (j) same viewed from above; (l) longitudinal section of same; (m) transverse section through the receptacle; **Saffron**: (n) style with three stigmas: (o) a stigma; (p) same showing vascular supply; (q) transverse section through lower portion of stigma; (r) upper end of stigma showing papillae and pollen grain. Reproduced from Spices and Flavoring (Flavouring) Crops: Leaf and Floral Structures, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press. Abbreviations: C, calyx; Col, collenchyma; Cor, corolla; Cp, compressed parenchyma; E, endocarp; F, fiber; Gh, glandular hair; H, hair; Hb, hair base; Hy, hypodermis; Mes, mesophyl; O, ovary; Oc, oil cavity; Ocl, oil cell; Od, oil duct; P, pericarp; Pal, palisade; Pap, papilla; Pg, pollen grain; Ph, phloem; Pp, pigmented parenchyma; R, receptacle; Rc, reticulate cells; S, stigma; Sc, stone cells; St, stomata; Sta, stamen; Sty, style; Vb, vascular bundle; Xy, xylem.

a petiole and with revolute margins. The upper surface is light gray or light brownish gray to weak olive green with numerous hairs; the lower surface is grayish, pubescent, and glandular punctuate. About 12 flowers arise in axillary whorls: calyx – tubular, bilabiate, about 4 mm in length, pubescent, nine to 12-nerved with the upper lip three-toothed; the lower lip has two hairy, ascending, attenuate divisions, and the throat is hairy; the corolla is about twice as long as the calyx, purplish and bilabiate; the stamens are four and didynamous, style with bi-lobed stigma (Figure 3d).

The epidermis of both leaf surfaces in the surface view shows wavy walls and numerous nonglandular and glandular hairs; the nonglandular hairs are uni- or multicellular with papillose walls, and are stiff and pointed. The glandular hairs are of two types: one with a short stalk and unicellular head, and the other with an eight- to 12-celled head, which measures up to 18 μm in diameter, and no stalk. In cross-section, the mesophyll consists of one or two layers of palisade cells, variable in size, followed by spongy tissues; the vascular bundle of the mid-rib has numerous fibers; the lamina contains two to three small vascular bundles on both sides of the mid-rib (Figure 3e and f).

Thyme has a fragrant and aromatic odor when crushed, and the taste is aromatic, warm, and pungent. About 15 cm from the tops of the plants is harvested, when in bloom, and the clipped stems are dried in the sun or in well-ventilated shed or room.

Cassia Buds

Cassia buds consist of the dried flowers of *Cinnamomum cassia* Blume (Lauraceae) gathered shortly after opening. They are dark brown, club, urn- or top-shaped bodies, enclosed within a grayish brown, cup-shaped calyx with a dome-like visible top of the pericarp. The calyx tube is rough and hard, and frequently has the pedicel attached; the exposed pericarp is smooth and lustrous with the small protuberant base of the style visible (Figure 3g and h).

In longitudinal section, the cortex of the calyx contains parenchyma cells up to 75 μm long, secretion cells up to 66 μm long and stone cells up to 80 μm long. The pericarp consists of a sclerenchymatous epicarp about 45–58 μm in radial thickness; the parenchymatous mesocarp contains prismatic calcium oxalate crystals, stone cells, secretion cells, and vascular tissues. The endocarp sclerenchymatous cells of the ovary canal are thick-walled and without a lumen, whilst the cells of the remaining region are rectangular, moderately thick-walled, and contain large prismatic crystals of calcium oxalate (Figure 3h).

Cassia buds have a cinnamon-like odor and a sweet, warm pungent taste akin to that of cassia bark.

Cloves

Cloves are dried flower buds of *Syzygium aromaticum* (L.) Merril & Perry (Myrtaceae). Cloves are like small, round-headed nails, about 10–17.5 mm long and blackish brown in color. The stalk consists of a cylindrical hypanthium, above which is a bilocular ovary containing numerous ovules on axile placentae; the 'head' consists of four, slightly projecting, calyx teeth, four membranous petals and numerous incurved stamens surrounding a large style (Figure 3i–l).

In transverse section, the hypanthium (receptacle), cut below the ovary, shows an epidermis with a thick cuticle and numerous, scattered, ranunculaceous stomata on the surface. The outer cortex contains two or three rows of large, ovoid to elliptical oil reservoirs (up to 200 μm in length) lined by a secretory epithelium and embedded in parenchyma; beneath is a zone consisting of an interrupted circle of bicollateral, fibrovascular bundles accompanied by a few sclerenchyma fibers and crystal fibers containing rosette calcium oxalate crystals. The bundles are embedded in parenchymatous cells with sinuate walls, and beneath these is a zone of parenchyma, in which some cells possess thickened angles and wavy walls, whereas others form chains around intercellular air spaces. Next is a zone of small-celled parenchyma followed by a second or inner zone of fibrovascular bundles, along with a few thick-walled, pericyclic fibers (Figure 3m).

The sepals have an epidermis similar to the hypanthium, with numerous stomata on the outer surface, a mesophyl with a few slender, vascular strands and numerous, ovoid oil glands and rosette crystals of calcium oxalate. The petals have straight-walled epidermal cells, an undifferentiated mesophyll containing oil glands, rosette calcium oxalate crystals, and small vascular strands. The filament of the stamen bears oil glands beneath the epidermis, and the connectives bear a large oil gland at the apex.

Cloves have a strong, pungent and spicy odor and a pungent, aromatic taste. Cloves are propagated by seeds. Hulled washed seeds produce better seedlings than unhulled fruits. The seeds are planted in shaded nurseries, and seedlings are raised. Shade and wind breaks are necessary in the early stages of growth. Trees begin to flower at about 4–5 years and continue to produce until 70 years of age. The crimson buds are picked before the corolla expands and are evenly spread under the sun on a concrete floor or on mats and gently stirred to bring about uniform drying. In good weather, drying is completed in 4–7 days. Quick drying produces the best-quality spice. Some cloves are dried on zinc sheets or in kilns. Cloves are apt to absorb moisture if, while drying, they are left in the

open overnight. The average annual yield of dried cloves per tree is about 3 kg but yields of 18 kg or more per tree are also not uncommon. Finally, they are packed in sacks known as 'mats' made of coconut leaves and stored in a dry room. Prolonged storage may cause a loss of volatile oils.

Saffron

Saffron is made from the dried stigmas of *Crocus sativus* L. (Iridaceae). The three stigmas are attached to the top of the style, and are 25 mm long, cornucopia-shaped, dark red and with fimbriate margins. The styles are about 10 mm long, cylindrical, solid and yellowish brown to pale yellowish orange (Figure 3n–p).

The stigma consists of thin-walled parenchyma containing coloring matter and covered by a thin-walled epidermis. The distal end of the stigma shows numerous bladder-like, cylindrical papillae up to 150 μm in length, among which occur a few smooth, spherical, pollen grains from 40 to 120 μm in diameter (Figure 3q–r).

The odor of saffron is strong and characteristically aromatic, and the taste is bitter/aromatic. The flowers are collected every morning, as they open, and the stigmas are then collected, either by pulling them out or cutting them off with the fingernails, after which the flowers are discarded. The value of saffron depends heavily on the methods by which the stigmas are processed. The stigmas are placed in sieves in layers 2–3 cm thick. These sieves are placed 15 cm above an almost spent fire for drying. By stacking them and by changing their order or position, the product is carefully dried. Stigmas are protected against dampness as well as light, because light bleaches saffron to a dull yellow color. Saffron should be preserved in amber-colored bottles or tin boxes.

See also: **Drying**: Theory of Air-drying; Drying Using Natural Radiation; **Essential Fatty Acids**; **Sensory Evaluation**: Aroma; Taste

Further Reading

Dampney J and Pomeroy E (1977) *All About Herbs*. London: Hamlyn.

Fliick H (1976) *Medicinal Plants and Their Uses*. Slough, UK: Foulsham.

Harrop R (1977) *Encyclopedia of Herbs*. London: Marshall Cavendish.

Husain A (1994) *Essential Oil Plants and Their Cultivation*. Lucknow, India: CIMAP.

Parry JW (1969) *Spices*, vols. 1 and 2. New York: Chemical Publishing.

Youngken HW (1948) *Text Book of Pharmacognosy*. New York: McGraw-Hill.

Tubers and Roots

K K Sinha, T. M. Bhagalpur University, Bhagalpur, India

Background

The use and cultivation of spice plants date back to the beginning of our history. Most of the spices that we value today were known to the ancient civilizations of China, Egypt, Greece, Rome, the East Indies, and the rest of the Old World. Spices have also been responsible for the rise and fall of empires and the great sea voyages to explore the distant corners of the globe. The search for spices by early European explorers also led to the discovery of new continents and waterways.

Spices are derived from different plant parts, including roots, stem, leaves, flowers, fruits and seeds. This article, however, deals only with those spices and flavoring crops which are obtained from tubers and roots. These are categorized into two groups, major and minor crops, depending upon their distribution and usage in different parts of the world.

Major Crops

Ginger

Ginger is the most important of all the spices obtained from underground plant parts. It has been in use in India and China since very ancient times. However, it is difficult to trace with certainty the exact place of its origin. Ginger was mentioned by the Chinese philosopher Confucius (551–479 BC) in his *Analects*. Ginger was also one of the first oriental spices to be known to the Europeans. The Greeks and Romans first obtained ginger from the Arab traders. It was known in Germany and France in the ninth century and in England during the tenth century.

Ginger is derived from the rhizomes of *Zingiber officinale* Rosc., a plant belonging to family Zingiberaceae. It is a slender perennial herb, mostly grown as an annual, 30–100 cm tall with a robust rhizome borne horizontally just below the soil surface. The fleshy sympodial rhizome is hard and thick, somewhat laterally compressed, often palmately branched and varying in shape and size in different cultivated types. It is covered with small, distichous scales with an encircling insertion, and with fine fibrous roots in the top layers of the soil. It is pale yellow in color externally and greenish yellow within.

The crop is always propagated vegetatively using small fragments of the rhizome known as 'seeds' or

'setts.' It is also grown both as a pure crop and as an intercrop. It may be grown in rotation with paddy-rice, samai (*Panicum miliare*). If irrigation facilities are available, ginger may be rotated with vegetables, chilli, groundnut, ragi (*Eleucine coracana*) and maize. Ginger is sometimes grown mixed with turmeric, which grows rapidly early on and provides partial shade to the developing ginger. It is harvested at different stages depending upon the requirements of the user. Green ginger is harvested between 5 and 7 months after sowing, and the rhizomes are used for preservation as *preserved ginger*. Final harvesting is carried out after 8–10 months from planting for the production of *dried ginger*. Shriveling, yellowing, and withering of the leaves, accompanied by drying and lodging of the aerial shoots, indicate maturity of the rhizomes.

The maturity of the rhizome has a significant influence on both its quality characteristics and its suitability for processing into preserved or dried ginger. The optimum time for harvesting ginger in India is 245–260 days from planting, after which the rhizomes become more fibrous. Harvesting of the mature ginger is mostly done by hand using a spade, hoe, or digging fork. In advanced countries, the late crop is also harvested by mechanical diggers. Extra care is required for the harvesting of healthy rhizomes. The soil, roots and shoots are subsequently removed, and the rhizomes are carefully washed and dried for storage. *See* Plates 1 and 2.

Several commercial types of ginger are recognized in cultivation, which are generally named after the localities where they are grown. The types with less fiber, which varies from 1.7 to 9.0%, have a higher demand. Some of the prized types of ginger in international trade include Jamaican ginger, Indian ginger, African ginger, and Chinese ginger, and these vary in taste and aroma.

Since the rhizome is perishable, it is susceptible to attack by the soilborne fungi, insects, and white ants. Rhizome rot or soft rot is the most serious disease found all over the ginger-growing areas of the world. Healthy and disease-free rhizomes are thus selected for storage. The use of specific fungicides for the treatment of rhizomes before storage is recommended on a regional basis. The treated rhizomes are air-dried under shade to remove excess moisture. The optimum conditions for the storage of fresh ginger have been investigated in detail in Hawaii. Harvested rhizomes should be washed immediately and then air-dried in the shade for 1–2 days. The rhizomes can be stored safely for up to 6 months by maintaining the temperature at 12 °C and the relative humidity at 65%.

In India, the ginger is loosely stored in pits, which may be up to 1 m deep. These pits are cool and provide protection against sunlight. A layer of sand or sawdust is spread at the bottom of the pit before storing the rhizomes. The pit is left uncovered or covered with a wooden plank, which in turn is plastered with mud, with a hole in the center for aeration.

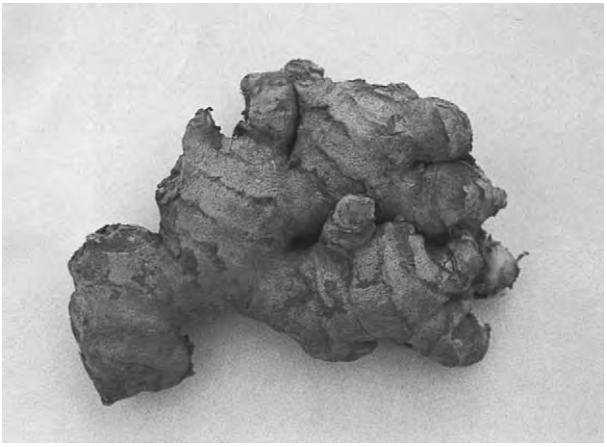

Plate 1 (see color plate 130) Spices and flavouring crops: Rhizome of green ginger (*Zingiber officinale*).

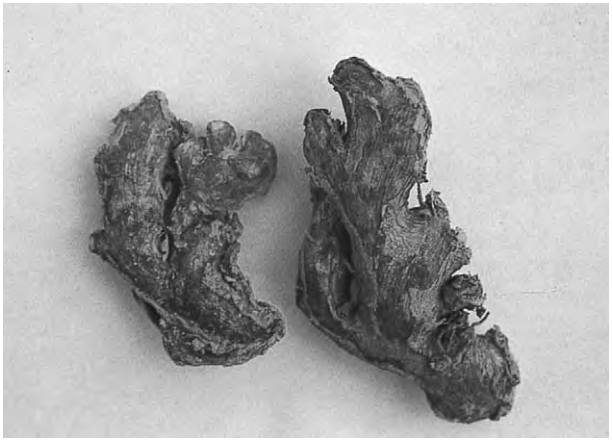

Plate 2 Spices and flavouring crops: Rhizome of dried ginger (*Zingiber officinale*).

As mentioned earlier, there are three primary products of ginger rhizomes: fresh or green ginger, preserved ginger in syrup or brine, and dried ginger. Green ginger is consumed mostly in the areas of production and nearby places. Preserved and dried products are the major forms in which ginger is traded internationally. Dried ginger is used as a spice and also for the preparation of its extracts.

Commercial preparation of preserved ginger is mostly practiced in China, Hong Kong, Australia, and India. Young juicy rhizomes are dried, cleaned, and boiled in water until they become tender. They

are then peeled, scraped, and boiled several times in a sugar solution. Sometimes, preserved ginger is also prepared in the dry state by dusting the drying rhizomes with crystalline sugar.

The preparation of dried ginger involves processing the cleaned rhizomes by peeling, splitting them into pieces, immersing them into boiling water for about 10 min, and, finally, drying. The appearance of processed ginger (black or white), the contents of volatile oil and fiber, the pungency level, and an assessment of the aroma and flavour are important in the quality evaluation of dried ginger.

The characteristic pleasant and aromatic odor of ginger is due to essential oil, which can be separated from the dried rhizome by steam distillation. The essential oil, known in the trade as 'oil of ginger,' is greenish to yellow in color, mobile (viscous on aging), with the characteristic warm and aromatic odor, but not the pungent flavor of the spice. The pungent principles of ginger are nonvolatile oleoresin, commonly known in the trade as 'gingerin.' This can be extracted from dried ginger by percolation with a suitable solvent. It is a viscous and dark brown oil and retains the full pungency of the spice. Besides the pungent principles, the oleoresin also contains gingerol, shogaol, and zingerone.

In medicine, ginger is used as a carminative and digestive stimulant. It also dilates the blood vessels in the skin, causing a feeling of warmth. Its preparations are also useful in the treatment of coughs and colds. Ginger is extensively used in culinary preparations such as soups, puddings, pickles, gingerbread, and cookies, and is an ingredient of all curries, except those used with fish. Ginger is also popular for flavoring beverages such as ginger ale and ginger beer.

Turmeric

Turmeric, another important spice, is also obtained from the underground plant part, i.e., the stem. The plant, a native of south-east Asia, is widely cultivated in India, Sri Lanka, Indonesia, China, Taiwan, Peru, Haiti, and Jamaica, of which India is by far the largest exporter.

Turmeric is obtained from the rhizomes of *Curcuma domestica* Val. (syn. *C. longa* L.), also a member of the family Zingiberaceae. The plant is a robust perennial herb with a short stem and tufted leaves. The rhizomes, brownish yellow in color, consist of a central bulbous portion bearing a number of finger-like lateral offshoots. Although the characteristic yellow matter is distributed throughout the plant, it is highly concentrated in the rhizome. There are several varieties of turmeric distinguished by the names of the localities in which they are grown. Turmeric cultivated in the hills is said to be of a better quality than that raised in the plains. Even the same variety which is grown in plains and on the hills shows distinct differences in quality and yield. *See* Plate 3.

Cultivation of turmeric requires a hot and moist climate, a liberal water supply, and a well drained soil. It can be grown successfully in rice fields on raised beds beyond the reach of stagnant water. It is usually rotated with ragi (a cereal grass cultivated in Africa and Asia), paddy, and sugarcane. It is rarely cultivated as a pure crop and is usually grown mixed with castor bean, maize, ragi, onions, brinjal, and tomato. The plants thrive well in partially shaded areas, but shade adversely affects the yield.

Turmeric is a good host for many pathogenic microorganisms and insectpests. Rhizome rot and leaf spot caused mainly by *Taphrina maculans* Butler are the important diseases of turmeric. The use of Bordeaux mixture or any other nonmetallic fungicide has been suggested as a remedy.

The crop is ready for harvesting in about 9–10 months, when the lower leaves turn yellow. The rhizomes are carefully dug up with hand-picks. The leafy tops are then cut off, and the roots are removed. The adhering soil particles are removed by brisk shaking or by rubbing, and the rhizomes are then thoroughly washed with water, placed in small heaps, covered with turmeric leaves, and left to dehydrate. The fingers and bulbs (or mother rhizomes) are then separated and cured separately.

The curing of turmeric involves boiling or steaming the fresh rhizomes in water, followed by drying and polishing. The rhizomes are boiled for 45–60 min

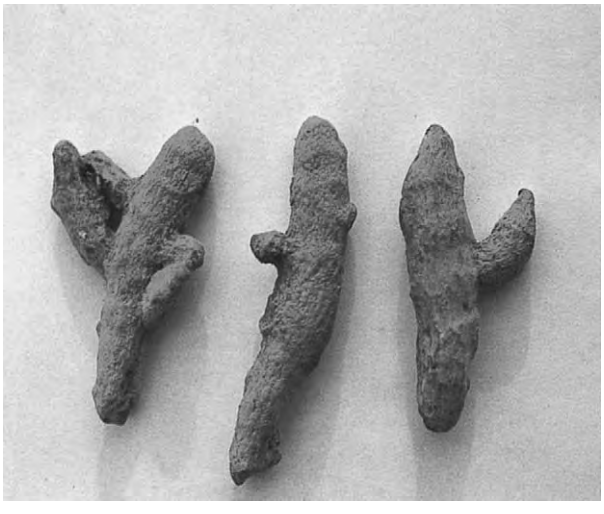

Plate 3 (see color plate 131) Spices and flavouring crops: Rhizome of turmeric (*Curcuma domestica*).

over a slow fire until they become soft. The rhizomes are then cooled and spread out to dry in the sun. Boiling of the rhizomes in very dilute solutions (0.05–0.1%) of mild alkali modifies the color of the core of the rhizome to the desired yellow–orange. The duration of boiling markedly affects the color and aroma of the final product. After boiling, the rhizomes are dried in sunlight for 10–15 days. Sun-drying gives rise to a surface-bleaching effect. After drying, the rhizomes become quite hard and brittle, and break with a metallic sound. Drying is followed by polishing, which includes cleaning off the outer skin, rootlets and adhering particles of soil, and making the rhizomes smooth. Polishing is performed either manually or in power-operated drums. The dried turmeric is coated with a slurry of turmeric powder in the course of polishing to impart a better color to the rhizomes.

The cured rhizomes are stored in pits dug at a raised site. The bottom and sides of these pits are lined with *Saccharum spontaneum* (commonly known as Rellu grass) and palmyrah mats. The pits, after being filled with cured turmeric, are covered with mats and grass and, finally, with earth. The produce may be stored for a year. The turmeric is also stored in 'godowns' or warehouses in gunny bags.

The coloring matter of turmeric is yellowish crystalline curcumin, which can be extracted from the powder either by direct solvent extraction or by extraction with alkali and subsequent precipitation with acid.

Turmeric is mainly used for domestic culinary preparations and is an important constituent of curry powder. Turmeric is also used as coloring agent in the textile, food, and pharmaceutical industries. In the Indian system of medicine, turmeric is used to some extent as a stomachic (i.e., promoting the appetite or assisting digestion), tonic, and blood purifier. The juice of the fresh rhizome is used as an antiparasitic for many skin diseases. Oil of turmeric, distilled from the dried rhizome, shows antiseptic properties.

Minor crops

Spices obtained from Roots

Angelica Angelica has been in use as a spice since AD 1500. It is obtained from the roots of *Angelica archangelica* L., a member of the family Apiceae. It is native to Syria but is known to occur in many parts of Europe and western Asia. In India, it is grown in the Kashmir valley.

The plant is a stout perennial herb with large, pinnately compound leaves and small, greenish white flowers. The dry rootstocks yield 0.35–1% of essential oil, the main constituent of which is β-phellandrene. The roots contain several furocoumarins such as angelicin, bergapten, xanthotoxin, etc. in addition to umbelliprenin and several phenols. The dried roots are frequently used for flavoring cakes, candy, and beverages. Angelica root is the main flavoring ingredient of gin. It is also used in perfumery and medicine. In medicine, the roots have been shown to possess stimulant, expectorant, and diaphorectic (i.e., inducing perspiration) properties.

Horse-radish The horse-radish (*Armoracia lapathifolia* Gilib.) is a member of the family Brassicaceae and has been in use as a condiment since ancient times. The plant is a native of southeastern Europe and is widely grown in Europe, America, and the hilly regions of India.

The plant is a tall, hardy herb with glossy green, toothed leaves. The roots are fleshy, yellowish white, and cylindrical. Initially, the root is about 2.5 cm in diameter, but it may reach a length of 1 m. The English refer to the plant as 'red cole,' perhaps due to the hot sensation experienced when ingesting even small pieces of root. The roots are dug up and are scraped or grated before use. Both fresh and dried roots are used as a condiment, which aids digestion and prevents scurvy. It is similar to mustard in most of its properties. The smell and flavor of horse-radish come from the glycoside sinigrin, which decomposes in the presence of water by enzyme action to form mustard oil.

Asafoetida Asafoetida is the dried latex obtained mainly from living rootstocks or tap roots of several species of *Ferula*, namely *F. foetida* Regel., *F. alliacea* Boiss., *F. rubricaulis* Boiss., *F. asafoetida* Linn., and *F. narthex* Boiss, of the family Apiaceae. The plants are perennial herbs which are mainly distributed from the Mediterranean region to Central Asia. The main regions of asafoetida production are eastern Iran and Western Afghanistan. Some species are also found in the Punjab and Kashmir regions of India.

The plants bear large, carrot-shaped roots, 10–15 cm in diameter at the crown after 4–5 years of growth. At the time of flowering, the upper part of the root is laid bare, and the stem is cut off close to the crown. The exposed surface is covered by a dome-shaped structure made up of twigs and mud. The exudate (milky juice) oozing out from the cut surfaces is scraped off after a few days. The collection of the resin and the slicing of the root are repeated until exudation ceases.

Asafoetida occurs in three forms in commercial trade: tears, mass, and paste. The tears make the

purest form of resin, it is bitter and acrid in taste and emits a strong and peculiar odor.

Asafoetida contains mainly resin (40–64%), gum (25%), and volatile oil (10–17% on a dry-weight basis). The resin portion consists mainly of asaresinotannol, free or combined with ferulic acid. Umbelliferone seems to be present in combined state. Oil of asafoetida is obtained by steam distillation of the gum resin.

Asafoetida is generally used for flavoring curries, sauces, and pickles. Medicinally, it stimulates the intestinal and respiratory tracts and the nervous system. It is useful in asthma, whooping cough, and chronic bronchitis.

Sarsaparilla Sarsaparilla is obtained from the dried roots of several tropical species of *Smilax*, a member of the family Liliaceae. Some of the important spice-yielding species include *S. aristolochiaefolia* Mill. (Mexico), *S. officinalis* H.B. & K. (Honduras), and *S. regelii* Killip & Morton (Jamaica). The plant thrives best in hot and humid climates and on light, well-drained sandy loam rich in humus.

The plants are climbing or trailing vines with prickly stems. The rhizomes are short and thick, with very long, thin roots. The plants are harvested after attaining 2–3 years of age. They are carefully dug up and cut off near the stock, which is then covered up again with surface soil. The harvested roots are then washed well, dried in the sun, and tied up in bundles. The roots contain a bitter principle which is used as a flavoring agent. It is chiefly used as a beverage condiment for the preparation of soft drinks. Sarsaparilla is mostly used in combination with winter green and other aromatic plants.

Medicinally, sarsaparilla has been used to treat syphilis and other skin diseases.

Spices obtained from Rhizomes or Tubers

Zedoary Zedoary (*Curcuma zedoaria* Rosc. syn. *C. zerumbet* Roxb.) closely resembles turmeric (*C. longa*) in appearance. It is a native of northeast India and is widely cultivated in many parts of China, Sri Lanka, and India.

The plant bears green leaves with brownish purple veins and grows up to a height of 50 cm. The rhizomes, which are large and fleshy, are cut into small pieces and dried. Dried slices have a bitter and strong, camphoraceous taste and are used for commercial purposes.

Steam distillation of the rhizome yields a light yellow oil. The odoriferous consitutent is said to be a sesquiterpene alcohol belonging to the tricyclic group.

Zedoary was an important spice in the past, but these days, it is usually used only for flavoring liqueurs and curries. The rhizomes have aromatic, stimulant, and carminative properties. Besides its use as spices, the rhizomes are also used for the preparation of shoti starch (flour), a substitute for arrowroot and barley. It is highly valued as a dietary item, especially for infants and convalescents. A red powder, 'abir,' is also prepared from powdered rhizomes by treatment with a decoction of sappan wood.

Mango ginger Mango ginger (*Curcuma amada* Roxb.) is a member of the family Zingiberaceae. The plant is native to India and occurs in the wild state in parts of Bengal, Konkan, and Madras. *See* **Plate 4**.

The plants are herbaceous, and the rhizomes have the characteristic odor of green mangoes. At the time of harvesting, entire plants are lifted up with a crowbar, and the rhizomes are collected, cleaned, and dried in the shade. On distillation, the rhizome yields essential oil, which consists mainly of pinene, ocimene, and linalool. The rhizome is mostly used in pickles and is considered stomachic and carminative. It is also applied over contusions and sprains.

Galangal Galangal is of three types: lesser galangal, greater galangal, and light galangal. The reddish-brown rhizomes are used as condiment and have an aromatic spicy odor and a pungent taste.

The lesser galangal (*Alpinia officinarum* Hance.), a member of the family Zingiberaceae, is a native of southern China. The plant is a perennial herb with a raceme of showy flowers and attractive foliage. The rhizome is smaller and has a stronger odor and taste. The lesser galangal is used as a spice in cooking, and also in medicine and for flavoring liqueurs. It is also used to impart a pungent flavor to vinegar.

The greater galangal (*Alpinia galanga* (L.) Willd.) is also a perennial herb with showy flowers and

Plate 4 (see color plate 132) Spices and flavouring crops: Rhizome of mango ginger (*Curcuma amada*).

beautiful foliage. It is commonly found in Indonesia and Malaysia, and is also cultivated in Bengal and southern parts of India. The green rhizomes contain essential oil consisting mainly of methyl cinnamate and cineol. It gives a pungent taste like a mixture of pepper and ginger. Medicinally, it is useful in respiratory troubles, especially in children. The rhizomes are also carminative and stomachic.

The light galangal (*Alpinia speciosa* (Wendl.) K. Schum) is a native of the Eastern Archipelago, off the Coromandel Coast of south-east India. Its rhizome is much larger and is generally used as substitute for greater galangal, and even as a substitute for ginger in many preparations.

See also: **Spices and Flavoring (Flavouring) Crops**: Use of Spices in the Food Industry; Properties and Analysis

Further Reading

Akamine EK (1962) *Storage of Fresh Ginger Rhizomes. Bulletin No. 130*. University of Hawaii: Hawaii Agriculture Experimental Station, Hawaii.

Anon. (1976) *The Wealth of India. Raw Materials*. New Delhi: Council of Scientific and Industrial Research.

Hirasa K and Takemasa M (eds) (1998) *Spice Science and Technology*. New York: Marcel Dekker.

Kochhar SL (1981) *Economic Botany in the Tropics*. Madras: Macmillan.

Parry JW (1969) *Spices: The Story of Spices*, vol. I. New York: Chemical Publishing Company.

Pruthi JS (1976) *Spices and Condiments*. New Delhi: National Book Trust.

Purseglove JV, Brown EG, Green CL and Robbins SRJ (1981) *Spices*, vols. I and II. London: Longman.

Rosengarten F Jr. (1969) *The Book of Spices*. Philadelphia, PA: Livingston.

Simpson BB and Connez-Ogorzaly M (1986) *Economic Botany: Plants in our World*. New York: McGraw Hill.

Watt G (1972) *Dictionary of the Economic Products of India*, vol. VI. New Delhi: Cosmo.

Properties and Analysis

C Jian Qin, Perfumery Research and Development Center of Yunnan, Kunming, Yunnan, China

Background

In the vegetable kingdom, there are many plants that are odorous. A range of spices used in foods is shown in **Table 1**. This article reviews the properties of these odor-bearing molecules as well as the technology of extraction. In addition to their odorant properties, many herbs and spices have an important effect on taste, as well as heating/cooling in the mouth. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics; **Sensory Evaluation**: Aroma; Taste.)

Range of Compounds Present

The odorants are chemically mainly of three types, and they can be present in any organ of the plant, such as the seed, fruit, flower, leaf, root, rhizome, and/or bark, even in moss. Some of the more important compounds found are shown in **Table 2**, with their structures in **Table 3**. (*See* **Essential Oils**: Properties and Uses.)

Terpenes

Most odorous compounds present in plants are monoterpenes, sesquiterpenes, and their derivatives. This kind of compound is derived from secondary metabolites of biosynthesis via coenzyme A. Their chemical structure is essentially formed by a number of units of 2-methylbutadiene (trivially called isoprene), so that their general formula is $(C_5\ H_8)_n$. If n equals 2, the compound is called monoterpene; if n equals 3, it is called a sesquiterpene, and so on.

Derivatives of Phenylpropane

Although this type of compound is less widely distributed than the terpenes, it is very important from the fragrance point of view, for example, *trans*-anethole in anise and star aniseed oil and cinnamic aldehyde in cinnamon bark oil. These compounds are also products of secondary metabolites; in comparison with terpenes, biosynthesis is via shikimic acid, which is an intermediate in the synthesis of lignin.

Aliphatic, Heterocyclic, and Phenolic Compounds

The third type of flavoring compounds consist of derivatives of aliphatic heterocyclic and phenolic compounds. These compounds can be associated with the two preceding types of odorants as minor compounds in essential oils. However, they are frequently encountered; for example, 6-methyl-5-hepten-2-one is found in most citrus essential oils. Some plants contain only aliphatic and heterocyclic compounds without either terpenes or derivatives of phenylpropane; for example, rose, jasmine, and violet. They are found in the plant in the flowers, leaves, and/or fruits. (*See* **Phenolic Compounds**.)

These compounds can belong to a wide variety of chemical groups:

Table 1 Spices used in food preparation/processing

Type and number	*Common name*	*Latin name*	*Class/order family (bot)*	*Cultivation region*
Fruits				
1	Pepper, black	*Piper nigrum*	Piperaceae	Tropical and subtropical regions
2	Vanilla	*Vanilla planifolia*	Orchidaceae	Madagascar, Comore Island, Mexico, Uganda
3	Allspice	*Pimenta dioica*	Myrtaceae	Caribbean Islands, Central America
4	Paprika (red pepper)	*Capsicum annuum*	Solanaceae	Mediterranean and Balkan region
5	Bay tree[a]	*Laurus nobilis*	Lauraceae	Mediterranean region
6	Juniper berries	*Juniperus communis*	Cupressacceae	Temperate climate region
7	Chilli	*Capsicum frutescens*	Solanaceae	Tropical region
8	Aniseed	*Pimpinella anisum*	Apiaceae	Temperate climate region
9	Caraway	*Carum carvi*	Apiaceae	
10	Coriander	*Coriandrum sativum*	Apiaceae	
11	Dill[a]	*Anethum graveolens*	Apiaceae	
Seeds				
12	Mustard	*Sinapis alba*[b] / *Brassica nigra*[c]	Brassicaceae / Brassicaceae	Temperate climate region
13	Nutmeg	*Myristica fragrans*	Myristicaceae	Indonesia, Sri Lanka, India
14	Cardamom	*Elettaria cardamomum*	Zingiberaceae	India, Sri Lanka
Flowers				
15	Clove	*Syzygium aromaticum*	Myrtaceae	Indonesia, Sri Lanka, Madagascar
Rhizomes				
16	Ginger	*Zingiber officinale*	Zingiberaceae	South China, India, Japan, Caribbean Islands, Africa
17	Turmeric	*Curcuma longa*	Zingiberaceae	India, China, Indonesia
Barks				
18	Cinnamon	*Cinnamomum zeylanicum, C. aromaticum, C. burmanii*	Lauraceae	China, Sri Lanka, Indonesia, Caribbean Islands
Roots				
19	Horse-radish	*Armoracia rusticana*	Brassicaceae	Temperate climate region
Leaves				
20	Parsley	*Carum petroselinum*	Apiaceae	Temperate climate region
21	Marjoram	*Origanum majorana*	Lamiaceae	Temperate climate region
22	Oregano	*Origanum heracleoticum, O. onites*	Lamiaceae	Temperate climate region
23	Rosemary	*Rosmarinus officinalis*	Lamiaceae	Mediterranean region
24	Sage	*Salvia officinalis*	Lamiaceae	Mediterranean region
25	Thyme	*Thymus vulgaris*	Lamiaceae	Temperate climate region

[a]Fruits and leaves.
[b]White mustard.
[c]Black mustard.
From Belitz H-D and Grosch W (1987) *Food Chemistry*. Berlin: Springer.

- Aldehydes, alcohols, esters, ketones, and lactones, which are formed by the metabolism of lipids and glycosides.
- Phenols, which result from the metabolism of proteins and glycosides.
- Sulfides and nitrogenous compounds (mercaptans, thiophenes, thiazoles, pyrazines and pyridines), which arise from the metabolism of proteins.
- Carboxylic acids, which come from the metabolism of sugars and glycerides.

They have a delicate fragrance and are generally responsible for floral notes and fruit flavor; their odor, therefore, is very different from that of terpenes. It must be noted that the heterocyclic compounds can be formed not only by biosynthetic processes but also by fermentation by microorganisms or thermal reactions, including the Maillard reaction. (*See* **Browning**: Nonenzymatic.)

Extraction and Isolation

Process Using Water Vapor

Essential oils can be produced by use of water vapor. The term 'oil' refers to their lipophilic nature, which causes them to separate from aqueous distillate. The adjective 'essential' refers to its origin in a living plant. (*See* **Essential Oils**: Isolation and Production.)

Table 2 Major components of the essential oils of spices

Spice[a]	*Components*[b]
Pepper (1)	22% α-Pinene [**29**], 21% sabinene [**25**], 17% β-caryophyllene [**49**], Δ^3-carene [**32**], limonene [**9**], β-pinene [**30**]
Allspice (3)	70% Eugenol [**50**], β-caryophyllene [**49**], methyleugenol, 1,8-cineole [**23**], α-phellandrene [**11**]
Bay leaf (5)	50–70% 1,8-Cineole [**23**], α-pinene [**29**], β-pinene [**30**], α-phellandrene [**11**], linalool [**4**]
Juniper berries (6)	36% α-Pinene [**29**], 13% myrcene [**1**], β-pinene [**30**], Δ^3-carene [**32**]
Aniseed (8)	80–90% Anethole [**54**]
Caraway (9)	55% Carvone [**21**], 44% limonene [**9**]
Coriander (10)	Linaool [**4**], linalyl acetate, citral[c]
Dill fruit (11)	35% Carvone [**21**], 12% dihydrocarvone, 10% limonene [**9**], carveol [**16**], α-terpinene [**10**]
Nutmeg (13)	27% α-Pinene [**29**], 21% β-pinene [**30**], 15% sabinene [**25**], 9% limonene [**9**], safrole [**55**], myristicin [**56**]
Cardamom (14)	20–40% 1,8-Cineole [**23**], 28–34% α-terpinyl acetate, 2–14% limonene [**9**], 3–5% sabinene [**25**]
Clove (15)	80–90% Eugenol [**50**], 9% caryophyllene, eugenol acetate
Ginger (16)	30% (−)-Zingiberene [**42**], 10–15% β-bisabolene [**41**], 15–20% (−)-sesquiphellandrene [**43**], (+)-arcurcumene [**60**], citronellyl acetate
Turmeric (17)	30% Turmerone [**62**], 25% arturmerone [**63**], 25% zingiberene [**42**]
Cinnamon (18)	50–80% Cinnamaldehyde [**57**], 10% eugenol [**50**], 0–11% safrole [**55**], 10–15% linalool [**4**], camphor [**33**]
Parsley (20)	1,3,8,*p*-Menthatriene [**58**], 1-methyl-4-isopropenylbenzene [**59**], β-phellandrene [**12**], myrcene [**1**]
Marjoram (21)	49–65% 1,8-Cineole [**23**], 25% estragole [**53**], 15% α-terpineol [**17**], 11% eugenol [**50**], linalool [**4**], geranyl acetate, ocimene [**2**]
Oregano (22)	Carvacrol [**51**], thymol [**52**], *p*-cymene [**61**], carvacrol methyl ether
Rosemary (23)	1,8-Cineole [**23**], camphor [**33**], β-pinene [**30**], camphene [**31**]
Sage (24)	1,8-Cineole [**23**], camphor [**33**], thujone [**26**]
Thyme (25)	Thymol [**52**], *p*-cymene [**61**], carvacrol [**51**], linalool [**4**]

[a]The numbers in parentheses refer to the numbers in Table 1 .
[b]The bold numbers in square brackets refer to the structures shown in Table 3.
[c]A mixture of neral and geranial.
From Belitz H-D and Grosch W (1987) *Food Chemistry*. Berlin: Springer.

Steam distillation This process, which is an indirect distillation, involves putting raw materials above the vapour current without previous maceration. The saturated vapors bearing volatile organic compounds are condensed on a cold surface, and the essential oil is recovered by decantation. This process allows treatment of raw materials that are heat-sensitive.

Hydrodistillation In this case, the raw materials are immersed in boiling water. The volatile compounds are carried away with the water vapor and collected after condensation and decantation, as in steam distillation.

Hydrodistillation is generally carried out at atmospheric pressure. Nevertheless, we can improve the relationship between the weight of material carried and that of evaporated water, which represents an improvement in energetic efficiency. Such material as vetiver rhizome can be treated with this technique, because it bears many compounds of high molecular mass, which cannot be evaporated easily at atmospheric pressure.

Microwave hydrodistillation Recently, a microwave technique has been developed for treating aromatic plants. The advantage of this technique is that it allows the recovery of odorant compounds present in smaller quantities in plants. In addition, the time of extraction is very short in comparison with traditional steam distillation. However, it seems as if this method will not be commercially available because of difficulties in the choice of microwave frequency.

Hydrodiffusion In contrast to the usual direct steam distillation, this process allows steam to enter from the top of the vessel. The resulting solution is condensed and separated by decantation, as mentioned above. The main advantages of this technique are in saving energy owing to reduced steam consumption and in the elimination of hydrolysis, because the raw material is never in contact with water, only with steam.

Vacuum distillation In many cases, the essential oils produced either by solvent extraction (see below) or by steam distillation do not exactly fit the perfumers' and the flavorists' needs. It is sometimes necessary to eliminate certain constituents, such as terpenes or heavy colorings, or to increase the percentage of major aroma compounds. Figure 1 shows a distillation column on which different compounds can be separated according to their boiling points. The temperature gradually increases during the distillation, the most volatile compounds being isolated first. The better the vacuum, the lower the distillation temperature, which in turn protects the odorants from degradation.

Table 3 Terpenes and phenolic and other compounds in food

Monoterpenes

Acyclic (including cyclic derivatives)

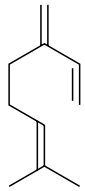

[1] Myrcene

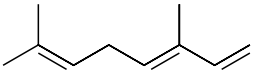

[2] *trans*-Ocimene

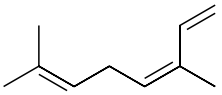

[3] *cis*-Ocimene

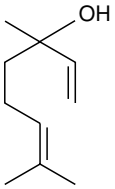

[4] Linalool

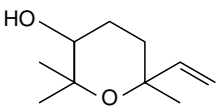

[4a] 2,6,6-Trimethyl-2-vinyl 5-hydroxytetrahydropyran[a]

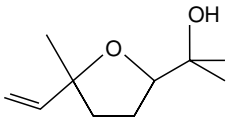

[4b] 2-Methyl-2-vinyl-5-hydroxyisopropyltetrahydrofuran[a]

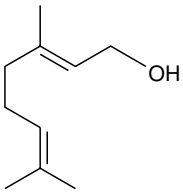

[5] Geraniol[b,c]

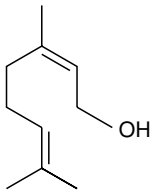

[6] Nerol[b]

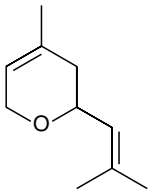

[6a] Neroloxide

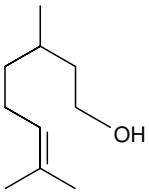

[7] Citronellol[b]

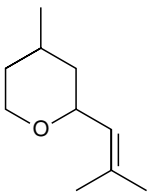

[7a] Rosenoxide

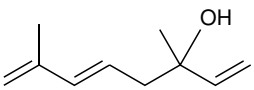

[8] Hotrienol[d]

Monocyclic

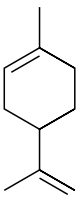

[9] Limonene

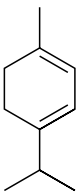

[10] α-Terpinene

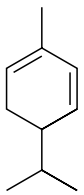

[11] α-Phellandrene

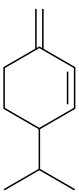

[12] β-Phellandrene

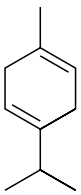

[13] γ-Terpinene

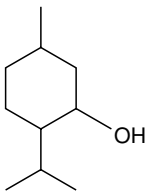

[14] Menthol

Continued

Table 3 Continued

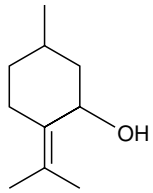

[15] Pulegol

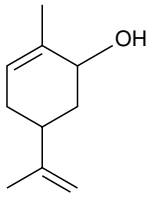

[16] Carveol

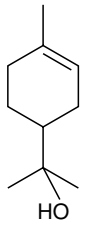

[17] α-Terpineol

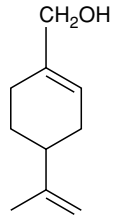

[18] Perilla alcohol

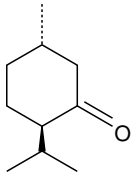

[19] Menthone

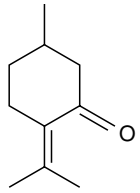

[20] Pulegone

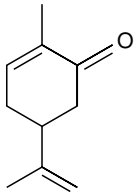

[21] Carvone

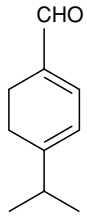

[22] 1,3-*p*-Menthadien-7-al

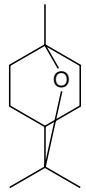

[23] 1,8-Cineole

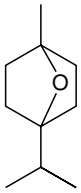

[24] 1,4-Cineole

Bicyclic

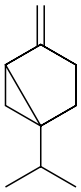

[25] Sabinene

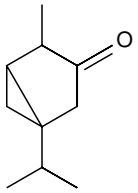

[26] α-Thujone

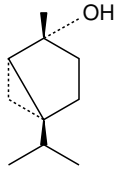

[27] (+)-*cis*-Sabinene hydrate

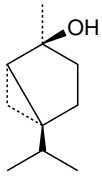

[28] (+)-*trans*-Sabinene hydrate

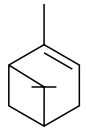

[29] α-Pinene

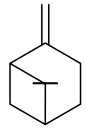

[30] β-Pinene

Continued

Table 3 Continued

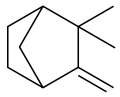

[31] Camphene

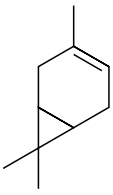

[32] Δ^3-Carene

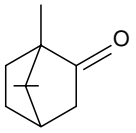

[33] Camphor

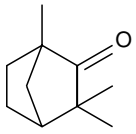

[34] Fenchone

Sesquiterpenes

Acyclic

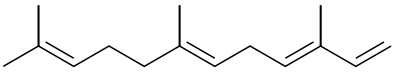

[35] (3*E*,6*E*)-α-Farnesene

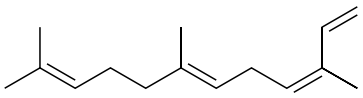

[36] (3*Z*,6*E*)-α-Farnesene

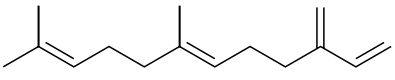

[37] β-Farnesene

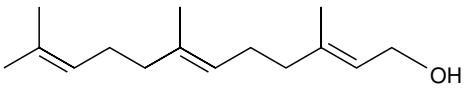

[38] Farnesol

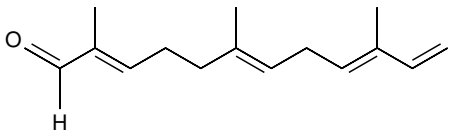

[39] (All-*trans*)-α-Sinensal

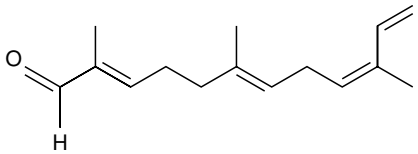

[40] (*trans,trans,cis*)α-Sinensal

Monocyclic

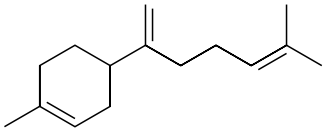

[41] β-Bisabolene

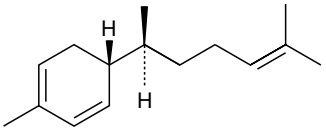

[42] (−)-Zingiberene

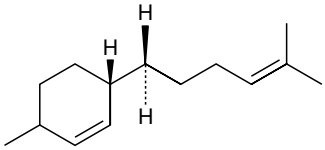

[43] (−)-Sesquiphellandrene

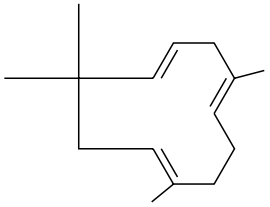

[44] Humulene

Bicyclic

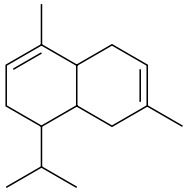

[45] β-Cadinene

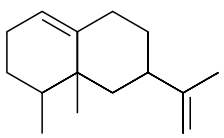

[46] Valencene

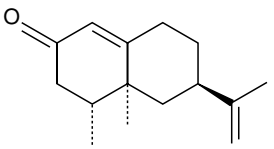

[47] (+)-Nootkatone

Continued

Table 3 Continued

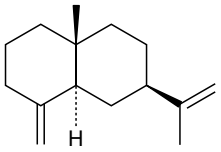

[48] β-Selinene

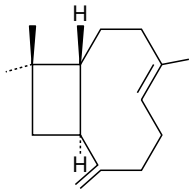

[49] β-Caryophyllene

Other

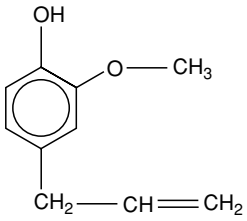

[50] Eugenol

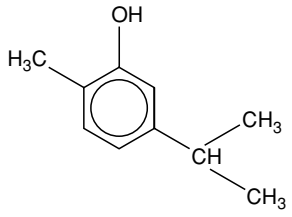

[51] Carvacrol

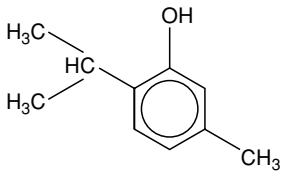

[52] Thymol

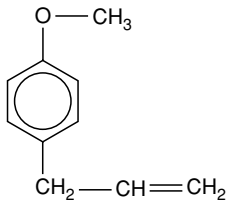

[53] Estragole

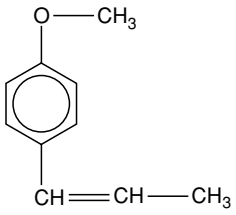

[54] Anethole

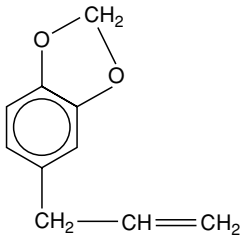

[55] Safrole

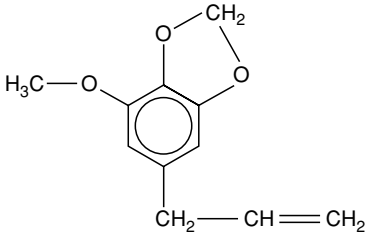

[56] Myristicin

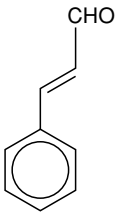

[57] Cinnamaldehyde

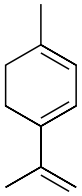

[58] 1,3,8,*p*-Menthatriene

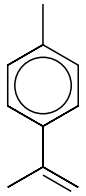

[59] 1-Methyl-4-isopropenylbenzene

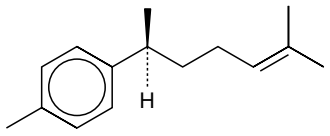

[60] (+)-ar-Curcumene

Continued

Table 3 Continued

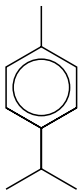

[61] *p*-Cymene

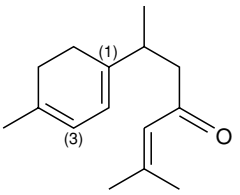

[62] Turmerone

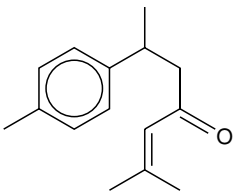

[63] ar-Turmerone

[a]Compounds [4a] and [4b], which occur as *cis* and *trans* isomers in wines, are also known as linalooloxides.
[b]Corresponding aldehydes genanial, neral, and citronellal also occur in foods. Citral is a mixture of neral and geranial.
[c]The corresponding acid is an important aroma constituent of wine cultivars 'Traminer' and 'Scheurebe.'
[d](−)-3,7-Dimethyl-1,5,7-octatrien-3-ol (hotrienol) is found in grape, wine, and tea aromas.
From Belitz H-D and Grosch W (1987) *Food Chemistry*. Berlin: Springer.

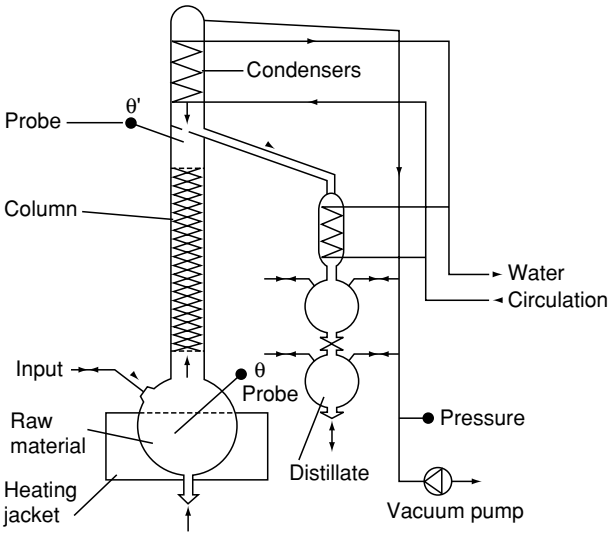

Figure 1 Vacuum distillation column. Reproduced from Spices and Flavouring Crops: Properties and Analysis, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Extraction Techniques

Extraction with organic solvents This technique is frequently used, like steam distillation. It involves putting aromatic plants in a static extraction vessel and covering them with an organic solvent, such as hexane, benzene, toluene, ethanol, or light petroleum or a binary mixture. Once equilibrium has been reached (or at least nearly so), the solution is separated, and the solvent is evaporated in a vacuum to yield the extract. Two particular methods are extraction with water and extraction with aqueous alcohol.

The choice of the solvent depends on many technical and economic parameters, in particular: selectivity; boiling point; diffusivity; miscibility with water; facility for recycling; safety – in general, the solvent chosen must be of as low a toxicity as possible both for operation of the extraction process and for consumption of the product.

The process of extraction used also depends on the nature of vegetable matter, such as its thermal lability, the operating temperature ranging from ambient to the boiling point.

According to the technique and solvent used, the products are known by one of the following names:

- *Tinctures*: obtained either by treating natural raw materials with ethanol or ethanol–water mixtures or by dissolving an extract in these solvents.
- *Oleoresins*: concentrated extracts with a characteristic odor and/or flavor, obtained from a natural raw material.
- *Resinoids*: concentrated nonaqueous extracts with a characteristic odor obtained from a dried natural raw material.
- *Concretes*: concentrated nonaqueous extracts with a characteristic odor and/or flavor obtained from a fresh natural raw material.

Besides odorant compounds, the organic solvent also extracts such undesirable substances as waxes and lipids, which are responsible for the nature of 'concretes.' In fact, a transformation is needed from concrete to absolute. Because waxes are not soluble in alcohol at low temperatures (about $-10\,^{\circ}C$), the concrete is diluted with alcohol at 30–40 °C and strongly stirred. When this alcohol solution is refrigerated at -5 to $-10\,^{\circ}C$, waxes precipitate and are filtered off. The filtrate is concentrated under vacuum, and after elimination of alcohol, the absolute remains. Generally absolutes are in liquid form, though they may be viscous or 'pasty.'

CO_2 extraction This process (see **Figure 2**) uses high pressure and, depending on the characteristics desired in the product, employs either liquid carbon dioxide (300 bar, 30 °C) or supercritical carbon dioxide (350

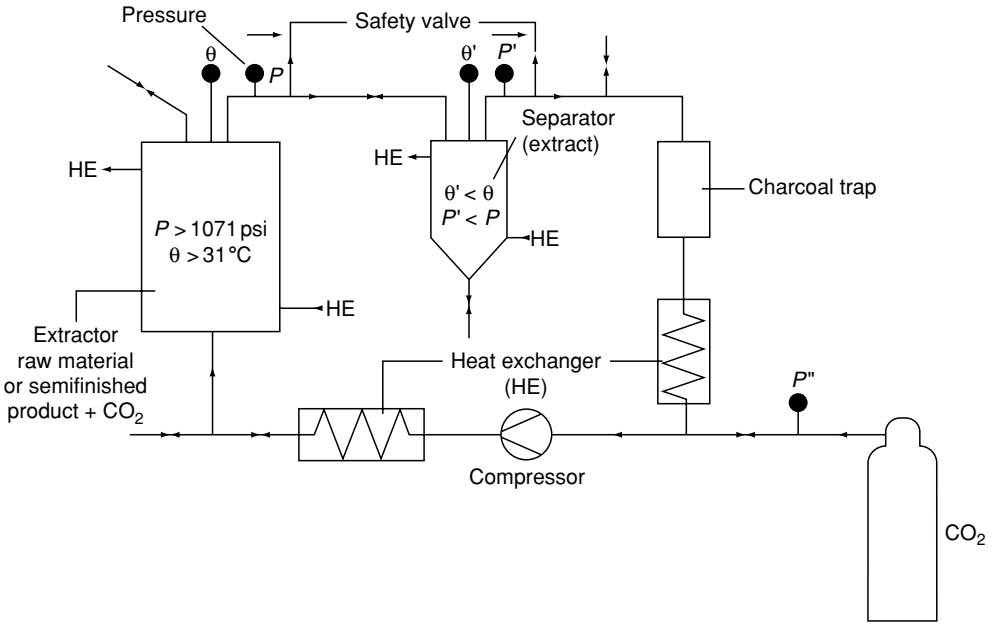

Figure 2 CO_2 extraction system. Reproduced from Spices and Flavouring Crops: Properties and Analysis, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

bar, 50 °C). The separation of extract takes place in the gas phase by a simple separation of gas and liquid.

The principle of the method is based on good solubility in carbon dioxide of most odorants of vegetable matter and semifinished products, such as concrete or even essences. The technique has the following advantages:

- No residual toxic solvent, particularly important for materials used in flavors.
- Low process temperature, important when processing unstable and heat-sensitive products.
- Selectivity (e.g., in caffeine extraction).
- Lack of fire hazards.
- Energy efficiency (no loss of organic solvents such as hexane or benzene, which have to be redistilled for use again).

Carbon dioxide as a solvent is relatively nonpolar, so its solubilizing powers are comparable to, but slightly more polar than, those of hexane. The latter is used frequently in the production of aromatic vegetable matter.

Extraction using ultrasound This process consists of vibrating the material by means of supersonic waves, which, owing to resonance, disintegrates the product matrix. More particularly, when the process is carried out in a liquid medium, the vibrational energy applied produces pressure variations that greatly facilitate the dispersion of the desired products in the solvent after they have been separated from their natural matrices.

Expression

Expression of oil at low temperature is a response to two particular features of the essential oils of citrus plants: (1) these oils contain terpene peroxides and polymerizable terpenes, so that they are heat-sensitive to hydrodistillation temperatures; (2) the oils are located in the porous pericarp, which is easily torn. It is therefore preferable to recover the essential oils by simple mechanical action.

It is the only process for treatment of aromatic plants without using a fluid extraction phase and involves either expressing the pericarp by water pressure or crushing whole fruits in a metallic cylinder. The essences so produced have low densities and can be separated from the aqueous phase by centrifugation. They are also designated 'essential oil.'

Specialized Extraction Techniques

Molecular distillation Molecular or short-path distillation is used to obtain colorless products; more stable products, because of the elimination of constituents with higher molecular weight (acids, pigments); and more delicate notes, because of the increase in the proportion of odorant in the oil.

Because of the high vacuum (0.05–0.0005 torr), the boiling point and extraction time are greatly

reduced, thus preventing the loss of some heat-sensitive notes in the oil. The product to be processed is combined with a heavy, and a light solvent and is then passed through the first short-path evaporator; the most volatile components condense together with the light solvent on a so-called 'finger' in the middle of the evaporator and are recovered as the first distillate (see **Figure 3**). The other components condense on the walls of the evaporator and are pumped into a second short-path evaporator. Here again, the next most volatile components of the oil condense on the inner 'finger' and are recovered as the second distillate. The residues and the heavy solvent condense on the walls and are recovered in the residue tank.

Simultaneous distillation/extraction This original process was designated by Likens and Nickerson. It uses two fluids; the volatile compounds are removed by steam distillation, and a nonpolar solvent continuously extracts them from the aqueous distillate.

The geometry of the apparatus enables the steam carrying the volatiles and the solvent vapor to meet in the condenser. The condensate separates into two phases, which are directed to their respective vapor generators. Eventually, the evaporation of the solvent allows recovery of a concentrated solution of odorant.

The relative position of two vapor generators differs according to the density of the solvent used. The method is particularly applicable in the study of rare vegetable materials.

Capture of headspace Three methods have been used as follows:

- Adsorption in liquid, which allows capture of the volatile components by passing the head space vapors through a solvent, and then concentrating the solution by evaporation.
- The cryogenic method, which condenses the volatile components in a series of refrigerated traps.

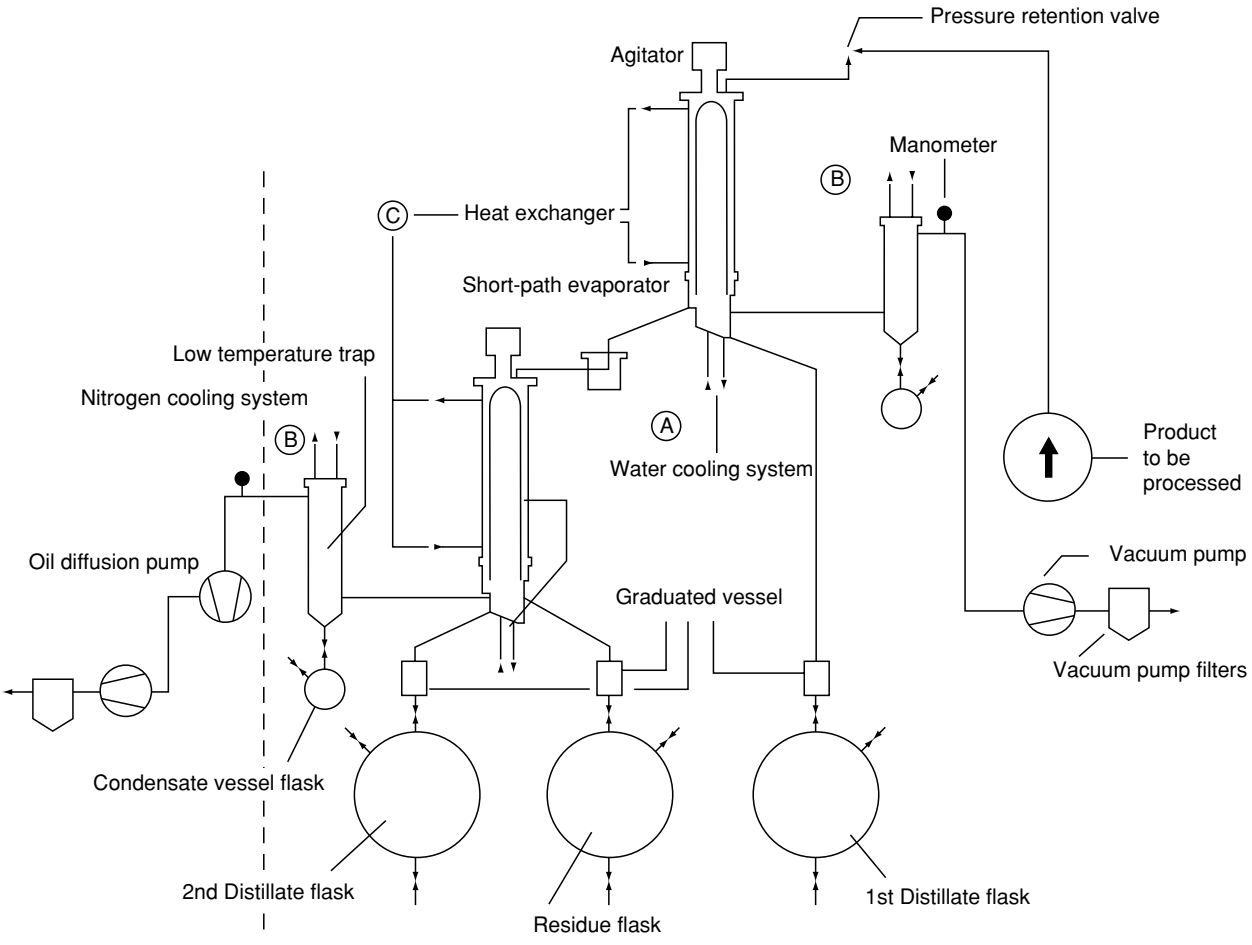

Figure 3 Molecular distillation unit. Reproduced from Spices and Flavouring Crops: Properties and Analysis, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

This method has been used particularly for the capture of the volatile components of fruit.

- Adsorption of solids, which involves passing the headspace over an adsorbent, such as XAD-4 (polystyrene) or TENAX [poly(2,6-diphenyl-*p*-phenylene oxide)]. The volatiles are recovered by thermal desorption or by elution with an organic solvent.

Volatiles

Odorants must be volatile to some extent, and this property is common in those molecules that contain fewer than 10 carbon atoms, for example, monoterpenes and their derivatives. It must be noted that aldehydes, ethers, and esters generally have a stronger odor than their parent hydrocarbons. This is why they are used for head notes in the formation of perfumes, cosmetics, and deodorants.

Volatiles of higher molecular mass evaporate relatively slowly at room temperature; because of this property, such compounds provide the more persistent odors.

Nonvolatiles

Nonvolatiles such as sugar, salt, piperine, and caffeine, may affect taste, although they cannot be detected by smell at room temperature.

Mechanism of Degradation and Staling

During steam distillation of an essential oil, terpenoid compounds are susceptible to degradation. Terpene alcohols and esters, especially, undergo a variety of well-known chemical reactions (hydrolysis, hydration, and cyclization), in particular under acidic conditions. For example, low pH values induce considerable hydrolysis of linalyl acetate and extensive rearrangement reactions. It is well known that tertiary alcohols and their esters are protonated in aqueous media to yield oxonium ions, which split to form a tertiary carbocation; thus, degradation reactions and rearrangements can be observed under the conditions of steam distillation.

In the mechanism of linalyl acetate thermal degradation, the generation of a primary allylic carbocation – the linalyl carbocation, which creates a *p*-menthane skeleton by cyclization – explains the presence of nerol and geraniol among the degradation products.

Some degradation can take place during the storage of essential oils, especially for oxygenated components. For example, citronellal can be resinified in a basic medium or oxidized in contact with air in the presence of light.

Since flavoring compounds are volatile, they can be lost easily by evaporation, thus leading to staling if the conditions of storage are not appropriate.

Microbiological Contamination

A number of essential oils, as well as some spices and aroma extracts, e.g., thyme and cloves, which are rich in phenolic compounds, possess antifungal and antibacterial properties. In general, essential oils have a good bacteriological quality, provided humidity is not too high. In contrast, the chemical stability of some essential oils is limited on account of oxidation, for example, those that are rich in limonene. Such essential oils must be protected against oxidation during storage. (*See* **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage.)

Bacteriological contamination is a hazard for spices and aroma extracts, particularly if they are aqueous extracts, for example, alliaceous extracts. It is necessary to ensure their bacteriological quality. The chemical stability for certain extracts, such as pepper and rosemary, is protected by the presence of natural antioxidants. (*See* **Antioxidants**: Natural Antioxidants.)

See also: **Antioxidants**: Natural Antioxidants; **Browning**: Nonenzymatic; **Essential Oils**: Properties and Uses; Isolation and Production; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Phenolic Compounds**; **Sensory Evaluation**: Aroma; Taste; **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage

Further Reading

Association Française de Normalisation (1988) *Contrôle de la qualité des produits alimentaires – Epices et Aromates*, 2nd edn. Paris: AFN.

Bicchi C and Sandra P (1987) *Capillary Gas Chromatography in Essential Oils Analysis*. Heidelberg: Springer.

Cu Jian Qin (1990) *Extraction de Compositions Odorantes Végétales*. PhD thesis, Institut Polytechnique de Toulouse No. 393.

International Standards Organization (1984) *Essential Oils – Terminology*. Murcia: ISO.

Meyer-Warnod B (1984) Natural essential oils – extraction processes and application to some major oils. *Perfumer & Flavorist* 9(2): 93–104.

Morin Ph and Richard H (1984) Thermal degradation of linalyl acetate during steam distillation. In: *Proceedings of The 4th Weurman Flavor Research Symposium, Dourdan, France*, pp. 563–576. Amsterdam: Elsevier.

SPOILAGE

Contents

Chemical and Enzymatic Spoilage

N F Haard, University of California, Davis, CA, USA

Introduction

The storage life of foodstuff is limited by the occurrence of chemical reactions which alter edible quality, including deterioration in optimal color, appearance, texture, aroma, flavor, nutrition, safety, and functional properties. When food undergoes deterioration in quality, it is not always clear whether the reaction causing the problem is of nonenzymatic, enzymatic, or microbial origin. For example, development of bad odor in stored coleslaw prepared with mayonnaise and sour cream was first attributed to microbial spoilage. However, investigation of the problem revealed that endogenous enzymes in the cabbage were activated by the anaerobic conditions in the package and were responsible for the formation of offensive- smelling volatiles. Likewise, enzymatic reactions rather than microbial spoilage was shown to be the primary cause of early spoilage in chilled, properly handled fish, by studying the initial quality loss of the flesh in sterile and nonsterile fish. In most cases, chemical and microbial spoilage acts in concert. For example, in cold-smoked salmon microbial activity causes production of characteristic spoilage off-odors and off-flavors while autolytic enzymes from the fish tissue are responsible for the characteristic texture deterioration. Moreover, there are also many instances where chemical, enzymatic, and microbial spoilage reactions are interactive processes. For example, products of nonenzymatic browning have been shown to inhibit other enzymatic, nonenzymatic, or microbial spoilage reactions.

Chemical Reactions that Contribute to Spoilage

Chemical spoilage includes enzyme-catalyzed reactions as well as nonenzymatic reactions. In general, enzyme-catalyzed reactions are of primary concern in untreated plant and animal tissues and nonenzymatic reactions predominate in properly processed foodstuffs. There are, however, some exceptions to this general rule. In addition, enzymatic and nonenzymatic spoilage reactions may act in concert. For example, discoloration of the surface of red meats is caused by a nonenzymatic reaction, oxidation of myoglobin (Fe^{2+}) to form metmyoglobin (Fe^{3+}). However, enzymatic reactions involved with respiration at the tissue surface can lower oxygen concentration and indirectly promote nonenzymatic oxidation of myoglobin. Some muscle tissues also contain the enzyme metmyoglobin reductase, which catalyzes the reduction of metmyoglobin back to myoglobin. Likewise, the nonenzymatic denaturation and aggregation of myosin primarily cause texture deterioration during frozen storage of certain species of fish. However, these reactions may be accelerated by the action of trimethylamine oxide demethylase, which forms formaldehyde, or by phospholipase, which forms free fatty acids. It is normally possible to distinguish between enzymatic and nonenzymatic catalyzed spoilage reactions by conducting simple experiments (Table 1).

Physiological Processes

Some spoilage reactions are of biological origin, i.e., are 'of or pertaining to living organisms,' and these include sets of enzyme-catalyzed reactions. For example, a set of biological reactions described under the general heading 'postharvest physiology of fruits and vegetables' are important because they may contribute to deteriorative processes such as plant senescence, wound response, chilling injury, and other types of stress response. Physiological processes in fruits and vegetables also lead to desirable changes such as fruit ripening, wound healing after harvest, and the reduction of reducing sugars during the 'conditioning' of potato tubers. Likewise, 'postmortem physiology of food myosystems' describes a set of reactions that normally cause a desired conversion of muscle to meat.

Table 1 Some general characteristics that can distinguish enzymatic and nonenzymatic catalyzed reactions

Treatment	*Characteristics*
Thermal inactivation	As proteins, most enzymes are inactivated by a brief heat treatment, while catalysts of nonenzymatic reactions are normally heat-stable
Dialysis	As macromolecules, enzymes are nondialyzable, while catalysts of nonenzymatic reactions are normally low-molecular-weight and removed by this treatment
Specificity	The number of reactants and products in enzymatic reactions are normally less than in the corresponding nonenzymatic reaction. Enzyme catalysis and inhibition are characterized by stereospecificity
Proteolysis	Enzyme catalysts, as proteins, are often inactivated by treatment with proteolytic enzymes
pH	While nonenzymatic reactions are often acid- or base-catalyzed, enzymatic reactions are characterized by a relatively narrow pH optimum
Temperature coefficient	The Q_{10} of nonenzymatic reactions is normally much lower than that of enzymatic reactions

However, in some cases postmortem reactions may contribute to poor-quality meat, e.g., 'burnt tuna,' 'pale, soft, exudative pork,' 'cold shortening,' and 'thaw rigor.'

Spoilage reactions in the case of harvested and stored plant and animal tissues are, for the most part, catalyzed by enzymes. Myriad enzymes occur naturally in biological tissues used by humans for food. There is remarkable similarity in the basic metabolism of biological cells. However, the amounts and different kinds of specific enzymes present in a given cell may be quite distinctive because of heritable characteristics, intrinsic factors such as age of the source organism, and extrinsic factors such as preharvest stress, growing conditions, and diet. Only a small set of enzyme-catalyzed reactions has thus far been clearly identified with food spoilage. Many important deteriorative processes in food are caused by families of enzymes operating in sequence and catalyzing a net change in the tissue, such as synthesis of lignin in asparagus. Some other examples of multistep reactions, which contribute to losses in food quality, are summarized in Table 2. The detailed mechanisms and biochemical controls for many of these multistep reactions are not completely understood. For example, until the advent of rDNA technology it was generally believed that the increased solubility of pectin and the associated texture softening of fruit are caused by hydrolysis of α-1,4 galacturonate linkages in pectin, a reaction catalyzed by polygalacturonases (EC 3.2.1.15; EC 3.2.1.67). However, blocking the synthesis of polygalacturonases in ripening tomato fruit was not completely effective in preventing the softening of transgenic fruit. Since pectin contains small amounts of other sugar residues in addition to α-1,4 galacturonate linkages, it now appears that other, not yet identified, mechanisms play a role in tomato fruit softening.

Other Endogenous Enzymes

One-step enzyme-catalyzed processes may also contribute to quality deterioration, particularly in processed foods. For example, for orange juice, in which a stable colloidal suspension is desired, the action of endogenous pectin methylesterase (PME; EC 3.1.1.11) is undesirable since demethylation of pectin, catalyzed by this enzyme, leads to separation of serum from particulates in juice. Heat processing of orange juice to inactivate PME is undesirable since a nonenzymatic chemical reaction adversely affects the delicate flavor of the product. Hydrolases, like PME, as well as oxidoreductases are the most studied enzymes in connection with food spoilage. Examples of enzymes and their contribution to quality loss in specific commodities are summarized in Table 3.

Nonenzymatic Reactions

In the case of heat-treated or other processed foodstuffs where enzymes have been destroyed or their activity has otherwise been arrested, nonenzymatic chemical reactions play a more important role in food spoilage than enzyme-catalyzed reactions. As with enzyme-catalyzed reactions some, but not all, nonenzymatic changes that occur during food processing and storage will result in loss of quality. Examples of nonenzymatic reactions that can lead to deterioration in food quality are listed in Table 4. The two nonenzymatic reactions in foods which appear to be of most widespread importance are Maillard browning and lipid oxidation. These reactions are discussed in detail in other sections.

Factors Influencing Chemical Reactions in Food

The rate of chemical reactions in foodstuffs is a function of one or more variables, including pH, temperature, ionic strength, concentration of reactants, presence of catalysts, mobility of reactants, oxidation

Table 2 Some multistep enzyme-catalyzed reactions that contribute to food spoilage

Net reaction	*Contribution to food spoilage*
Glucosestarch	Loss in sweet taste of some vegetables, e.g., sweetcorn
Glucose$CO_2 + H_2O$	Respiration rate is directly related to spoilage rate of fruits, vegetables, and seed crops
Starchglucose + fructose	Cold sweetening of some vegetables leads to excessive Maillard browning, e.g., potato tuber
Glycogenlactic acid + H_2O	Glycogenolysis and glycolysis are of central importance in postmortem myosystems. The rate and extent of associated pH decline directly (e.g., water-holding capacity, protein denaturation) and indirectly (e.g., rate of other enzymatic and nonenzymatic reactions) to influence quality
ATPhypoxanthine	The rate and extent of adenosine triphosphate catabolism in myosystems influences meat quality in several ways, e.g., the acceptability of fish is directly related to the accumulation of hypoxanthine
Methionineethylene	The plant hormone ethylene is synthesized by the 'methionine cycle' during specific developmental change or as a consequence of wound injury. In turn, increased ethylene can initiate biosynthesis of enzyme cascades, e.g., chlorophyll destruction or lignin biosynthesis
Insoluble pectinsoluble pectin	Texture softening and increased vulnerability to saprophytic microorganisms and physical damage in some fruits and vegetables, e.g., ripe tomato
Phospholipidaldehydes, ketones, free radicals	Lipoxygenase cascades can decrease the amounts of essential nutrients, and cause off-flavors
Collagenpeptides, amino acids	Postmortem degradation of collagen is catalyzed by a family of enzymes and can influence physical integrity, appearance, and yield of fish fillets

Table 3 Some endogenous enzyme-catalyzed reactions that contribute to food spoilage

Enzyme	*EC number*	*Food*	*Importance*
Lipoprotein lipase	3.1.1.34	Milk	Releases short-chain fatty acids from milk fat leading to hydrolytic rancidity
Alkaline protease	3.4..24.40	Milk	Since this enzyme is stable to heat it may contribute to gelation in products processed at ultra high temperatures
Thiaminase	2.5.1.2, 3.5.99.2	Shellfish	Loss of thiamin in fermented products
Phospholipase	3.1.1.4, etc.	Fish	Releases fatty acids in frozen product causing denaturation of muscle proteins and texture deterioration
Trimethylamine oxide demethylase	4.1.2.32	Fish	Releases formaldehyde in frozen gadoid fish that contributes to protein aggregation and texture deterioration
Lipoxygenase	1.13.11.12, etc.	Legume seeds	Formation of specific hydroperoxides can lead to bleaching of pigments, offensive flavor formation, as well as texture change and loss of nutrients
Peroxidase	1.11.1.1, etc.	Vegetables	Decomposition of hydroperoxides with generation of free radicals appears to cause bleaching of pigments, off-flavors, etc.
Ascorbic acid oxidase	1.11.1.11	Citrus	Results in loss of vitamin C activity in orange juice
Chlorophyllase	3.1.1.14	Green vegetables	Removal of the phytol side chain from chlorophyll appears to be part of the degreening process during plant senescence
Polyphenol oxidase	1.10.3.1	Fruits, shellfish	Enzymatic browning

reduction potential, competing reactions, decompartmentation of reactants, and the physical state of the reaction milieu. One of the most important variables influencing chemical changes in food is temperature. The Arrhenius relation gives the influence of temperature on reaction rate:

$$k = k_o^{-E_a/RT}$$

where k is the reaction rate constant, k_o is the preexponential constant, E_a is the activation energy in kilojoules per mole, R is the gas constant, and T is the absolute temperature in Kelvin. E_a is normally much lower for enzymatic reactions than for nonenzymatic reactions. Thus, in general, the rate of a nonenzymatic reaction is more sensitive to temperature change than is an enzymatic reaction. Enzymes, like other proteins, are subject to thermal denaturation. Accordingly, the range of temperature in which catalytic rate increases with a rise in temperature is limited by the enzyme's thermal inactivation temperature. Because of these two considerations, the 'master' chemical reaction, i.e., that which limits the storage

Table 4 Some nonenzymatic reactions that can lead to loss in food quality

Reactant(s)	*Product or result*	*Importance*
Chlorophyll, H^+	Pheophytin, Mg^{2+}	Loss of Mg^{2+} from chlorophyll results in olive-brown discoloration of green vegetables. Reaction occurs rapidly at low pH and high temperature
All-*trans* β-carotene	*Trans–cis* rearrangement	Isomerization of carotenoids is promoted by light or heat and results in isomers, which absorb light in shorter wavelength and with lower extinction coefficients, and loss of provitamin A activity
Cysteine	H_2S, NH_3, acetaldehyde	Thermal degradation of sulfur-containing amino acids can directly produce aroma and also lead to other reactions involving the products
RCHO, RNH_2	Maillard browning	The Maillard reaction is one of the most important reactions in foods, affecting color, flavour, nutrition, and possibly safety
Amylose	Crystallization	The alignment of linear starch chains by hydrogen bonding to form insoluble precipitates, a process called retrogradation, is important in processed products containing gelatinized starch, e.g., bread staling
Anthocyanin, SO_2	Decolorization	The addition of SO_2 to the 4-position of anthocyanins to form a bisulfite addition product results in loss of color
Organic acid, Ca^{2+}	Ca-chelate	Organic acids like phytic acid or citric acid can destabilize polygalacturonate in the middle lamella by sequestering Ca^{2+} and thereby influence texture
Ascorbic acid	2-Furaldehyde, CO_2	The nonenzymatic degradation of vitamin C occurs by an acid/metal ion-catalyzed reaction or by an oxidative mechanism. In addition to loss in nutritive value, carbonyl degradation products, such as diketogulonic acid, can contribute to Maillard browning
Fatty acid, O_2	Autooxidation	Autooxidation is a free radical reaction which can be catalyzed by metal ions; the reaction rate increases with degree of unsaturation; a primary effect on quality is formation of off-flavors; however, all quality indices, including nutrition, color, texture and safety may be influenced under appropriate conditions
Protein	Denaturation	Loss of protein native structure can lead to protein aggregation and loss in functional properties with the possibility of influence on all quality indices under appropriate reaction conditions
Myoglobin, thiol compound, TMAO	Green tuna	Precooking tuna containing a high content of trimethylamine oxide (TMAO) can lead to green discoloration of tuna by an oxidation reduction reaction
Aspartame, peptides, nitrite	Mutagenic compounds	Naturally occurring dipeptides and the artificial sweetener aspartame can form mutagenic compounds after nitrosation. Aspartame is also unstable under alkaline conditions and heat
Arginine	1-Methylguanidine	High-temperature (e.g., 150–210 °C) treatment of arginine results in formation of mutagenic compounds, such as those found in cooked grain-based foods
Thiamine, H^+ or OH^-	2-Methyl-4-amino 5-aminethyl pyrimidine, etc.	Vitamin B_1 is destroyed by various reactions, including acid- or base-catalyzed reactions, redox reactions, photolysis, and reaction with bisulfite

life of a given product, normally differs with the temperature at which the product is stored (Figure 1). It is also important to recognize that E_a for a given reaction in a particular temperature range is also a function of other parameters (e.g., pH, glass transition). The extent to which other parameters influence thermal response depends on the particular reaction under consideration.

The extent and direction to which chemical transformations are influenced by parameters other than temperature depend on the reaction under consideration. For example, removal of oxygen from fish tissue decreases the rate of enzymatic or nonenzymatic lipid oxidation and the development of off-flavors, but at the same time it increases the rate of enzymatic trimethylamine oxide (TMAO) demethylation and associated protein aggregation. Likewise, reducing the water concentration of food by dehydration may decrease the mobility of water-soluble reactants, but it does not serve the same function for fat-soluble reactants and may even promote lipid oxidation by exposing lipid substrates to oxygen and increasing the availability of a metal catalyst.

It is also important to consider that, food-processing and storage conditions, designed to minimize one detrimental reaction, may inadvertently alter other

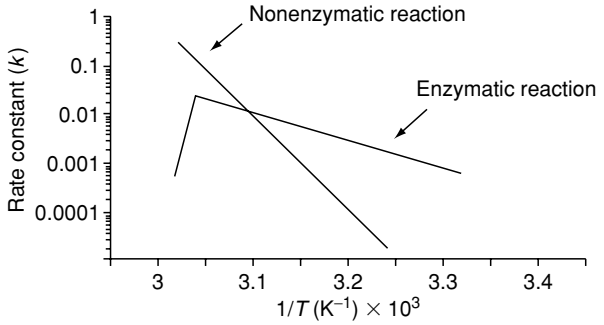

Figure 1 Typical temperature dependence of enzymatic and nonenzymatic reactions. Reproduced from Spoilage: Chemical and Enzymatic Spoilage, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

parameters which, in turn, may function to promote the same detrimental reaction. For example, while freezing of food lowers temperature and water activity, and thus is expected to lower the rate of reaction, it may serve otherwise to accelerate the rate by unexpected changes in pH, ionic strength, or reactant availability.

Moreover, storage or processing strategy designed to minimize a reaction that adversely affects storage life may accelerate other reactions that cause a different, but equally unacceptable, quality. For example, pasteurization of kiwi fruit juice (pH 3.5) inactivates enzymes and spoilage organisms that contribute to spoilage; but at the same time, heat treatment causes rapid pheophytinization of chlorophyll and loss in the characteristic green color of the product. Because of the complexity of food systems it is often difficult to predict the consequences of processing and storage conditions on quality. Accordingly, efforts to minimize the contribution of chemical reactions to food spoilage require an empirical approach, at least in part.

See also: **Browning**: Nonenzymatic; Enzymatic – Biochemical Aspects; Enzymatic – Technical Aspects and Assays; **Enzymes**: Functions and Characteristics; Uses in Food Processing; **Spoilage**: Bacterial Spoilage

Further Reading

Bradley DG and Min DB (1992) Singlet oxygen oxidation of foods. *CRC Critical Reviews in Food Science and Nutrition* 31: 211–236.

Chien-Yi W (1994) Chilling injury of tropical horticultural commodities. *HortScience* 29: 986–988.

Eskin NAM (1990) *Biochemistry of Foods*, 2nd edn. New York: Academic Press.

Fennema OR (ed.) (1996) *Food Chemistry*, 3rd edn. New York: Marcel Dekker.

Haard NF (1994) Technological aspects of extending prime quality of seafood: a review. *Journal of Aquatic Food Product Technology* 1: 9–27.

Haard NF (1997) Food as cellular systems: impact on quality and preservation. In: Taub IA and Singh P (eds) *Food Storage Stability*, pp. 39–74. Boca Raton: CRC Press.

Haard N (1997) Product composition and the quality of frozen foods. In: Erickson MC and Hung Y-C (eds) *Quality in Frozen Foods*, pp. 275–295. New York: Chapman and Hall.

Haard NF and Simpson BK (eds) (1999) *Seafood Enzymes*. New York: Marcel Dekker.

Huis INT and Veld JHJ (1996) Microbial and biochemical spoilage of foods: an overview. *International Journal of Food Microbiology* 33: 1–18.

Jayaprakasha HM, Jayaraj-Rao K and Lokesh-Kumar WA (1997) Studies on the influence of water activity (A_w) on the stability of foods – a critical appraisal. *Journal of Food Science and Technology, India* 34: 273–285.

Richardson T and Finley JW (eds) (1985) *Chemical Changes in Food During Processing*. Westport: AVI.

Rizzi GP (1997) Chemical structure of colored Maillard reaction products. *Food Reviews International* 13: 1–28.

Wong DWS (1989) *Mechanism and Theory in Food Chemistry*. Westport: AVI.

Bacterial Spoilage

E A Zottola, (retired) University of Minnesota, St. Paul, MN, USA

Introduction

Food for humans is obtained from plants and animals. The nutrients in these foods serve not only human needs but are also vital to the growth and survival of bacteria. Thus, if food is not immediately utilized or preserved after harvest, it spoils. Bacterial spoilage is inherent in foods. The spoilage process begins as soon as the food is harvested or slaughtered, and if steps are not immediately taken to prevent spoilage, the food will become unacceptable for human use. Spoilage of food can be caused by many factors: enzyme activity, physical damage, chemicals, rodents, insects, and metabolic activities of microorganisms, bacteria, yeasts, and molds. Foremost among the causes of food spoilage is the growth and utilization of the food by bacteria. Food is spoiled by bacteria when it becomes esthetically unacceptable to the user. What is spoiled to one may be a delicacy to another. It may have a foul odor or off-flavor, or be discolored, slime may be apparent, or there may be visible microbial growth on the food. These obvious

defects are revolting to the consumer and make it unacceptable for consumption. Food is also considered spoiled if the potential for the spread of foodborne disease exists. For example, food contaminated with high numbers of *Staphylococcus aureus* and possibly staphylococcal enterotoxin would not have any obvious defects that would make it apparently unacceptable, but it could make the consumer very ill and from this point of view it is spoiled. Food contaminated with industrial chemicals, pesticides, herbicides, or any material that should not be in the food would be considered spoiled. There are many facets of food spoilage. In this article, food spoilage caused by bacteria is reviewed. Foodborne illness is discussed in other sections of the *Encyclopedia*. (*See* **Wastage of Food**.)

Food as a Habitat for Bacteria

In its natural state, food, whether derived from animals or plants, is very susceptible to the action of bacteria and is easily spoiled by their metabolic activities. The processing of food is designed, among other things, to control the activity of bacteria and other microbes to preserve the food for later use. As food is processed, the type of bacterial spoilage that may occur will change. The major thrust of this article will be to describe how varying process methods affect food spoilage by bacteria. The growth of bacteria in foods is affected by several factors inherent in the food or by conditions external to the food. These include the acidity or pH of the food, the availability of water (water activity, a_w), the oxidation reduction potential (redox or E_h) of the food and the environment, availability of the nutrients in the food, the temperature of the food and the environment, and time available for bacterial growth to occur. (*See* **pH – Principles and Measurement**; **Water Activity**: Effect on Food Stability.)

Processing of Food and Bacterial Spoilage

The extent of processing of foods varies. Some foods are not processed prior to use and are very perishable. Others are lightly processed and are easily spoiled by bacterial action. The use of two or more preservation methods will give a food product additional stability, but does not make it immune to bacterial spoilage. This process of using one or more minimal preservation processes on a food to extend its quality retention time is referred to as the 'hurdles concept.' By adding another process to the food a hurdle is created, that is, one more factor that the bacteria have to overcome before they will grow and spoil the food. Thus the use of several hurdles should have a very positive effect on the quality retention time. There are a significant number of high-quality minimally processed foods that utilize the hurdles concept to control bacterial spoilage and effectively increase quality retention.

Foods given a more severe preservation process, such as drying or canning (high-temperature thermal process), are not as easily spoiled by bacterial action. With these concepts in mind it is possible to classify foods into spoilage potential categories based on the type of preservation process applied to it:

1. Highly perishable, no preservation process (such as fresh poultry) with a quality retention time of 1–3 days
2. Semiperishable, lightly processed (such as pasteurized milk) with a refrigerated quality retention time of 1–3 weeks
3. Semishelf-stable, two or more preservation processes used (the hurdles concept), such as vacuum-packaged cooked meat products, with a quality retention time from 1 to 4 months
4. Shelf-stable, highly processed foods, such as canned vegetables or dried milk products with a quality retention time exceeding 1 year

Thus, the processing or lack of processing that the food receives significantly influences the type of bacterial spoilage that occurs in foods. By developing and understanding the factors affecting the growth of bacteria, such as the chemical, physical, and compositional properties of food and the processing applied to the food, it is possible to predict with some reliability the type of bacterial spoilage that may occur in food. (*See* **Canning**: Principles; **Drying**: Theory of Air-drying.)

The Spoilage of Food by Bacteria

Bacterial spoilage of food occurs when a number of factors are in place:

1. The food must be contaminated with spoilage bacteria
2. The food must be suitable for growth of the contaminating bacteria
3. The environment associated with the food must support microbial growth
4. The bacteria must grow and produce metabolites which 'spoil' the food
5. There must be sufficient time for the bacteria to grow

Control of these five factors will markedly influence the growth of bacteria in food and the resultant spoilage.

Source of Bacteria in Food

Bacteria come into contact with or contaminate food from several sources. Food obtained from plants that is consumed by humans is in direct contact with soil. Fertile soil will contain up to 10^9 microbes per gram of soil. Consequently, foods that grow in or on the soil will be contaminated with significant numbers of bacteria. Animals used as food for humans are contaminated on the outside by soil and manure bacteria and internally by bacteria associated with the digestive system of the animal. When slaughtered, the flesh becomes contaminated with the internal bacteria as well as the external ones. Foods obtained from marine sources are contaminated by bacteria from the water in which they live and by bacteria in their digestive system.

Animal, plant and marine foods have native or indigenous microflora associated with them. In general, bacteria on the exterior of these foods are usually aerobic and frequently psychrotrophic, whereas, bacteria in the interior of the animal, plant, or marine food will be obligate or facultative anaerobes. The interior of these food sources will have reduced E_h values or a negative oxidation reduction potential, creating anaerobic conditions, which encourages the development of these organisms. Regardless of the source of the organisms, under the right conditions all are capable of causing food spoilage.

Food is also contaminated with bacteria from people who handle the food. These bacteria are usually mesophilic and may occasionally be pathogenic. A primary source of contamination of food with spoilage bacteria is the equipment used to handle and process the food. If the equipment is not clean, microcolonies, called biofilms, composed of surface dirt and bacteria will develop and consequently millions of bacteria will be present on the surface: when the food comes in contact with the biofilms on the surface they can be transferred to the food. This concept of bacteria transfer to foods from biofilms is referred to as the biotransfer potential. Control of direct human contact with the food and proper cleaning and sanitizing of food contact surfaces and nonfood contact surfaces are critical in preventing bacterial contamination of foods. (*See* **Cleaning Procedures in the Factory**: Overall Approach.)

Bacterial Food Spoilage Based on Preservation Processes

Highly Perishable Foods

Highly perishable foods are those that have been harvested or slaughtered and are utilized with little or no further processing to prevent bacterial spoilage. This category includes, but is not limited to, the following fresh foods: raw vegetables, raw fruits, raw milk, fish, poultry, and red meats. Of these food groups, fresh raw fruits are the least susceptible to bacterial spoilage since the pH of the fruits retards bacterial growth. The environment in which the food is maintained affects bacterial spoilage of these foods. Under low-temperature storage, between 0 °C and 7 °C, the primary bacterial spoilage is caused by psychrotrophic bacteria. These organisms are able to grow slowly at low temperatures; they have optimum growth temperatures in the mesophilic range and could also cause spoilage of these foods at higher temperatures. The majority of these bacteria are in the family Pseudomonaceae and are species in the genera *Pseudomonas, Xanthomonas, Gluconobacteria, Achromobacter, Flavorbacterium*, and *Alcaligenes*. Growth of these bacteria in highly perishable foods results in a variety of sensory defects (off-flavors – putrid, rancid, bitter, etc. – formation of slime, color changes, or strong odors).

The temperature of raw unprocessed foods, in essence, selects the type of bacteria that will grow and spoil the food. If these foods are stored at ambient temperatures of 18–43 °C, mesophilic bacteria of many types will grow and spoil the food. Species in the genera *Clostridium, Bacillus, Erwina, Streptococcus, Lactobacillus, Escherichia, Proteus, Enterobacter*, and many others, will grow and cause spoilage of the food.

It should be clear from the discussion up to this point that many different types of bacteria are able to grow on raw, unprocessed foods. Even refrigeration will not stop bacterial spoilage of these foods. The identification of genus and species of the bacteria causing food spoilage is not critical to preventing food spoilage; however, it may be useful to identify the bacteria if the source of the spoilage organisms is important. What is critical is that bacterial spoilage of food is recognized and that appropriate steps are taken to control or prevent the spoilage. (*See* **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Fish**: Spoilage of Seafood; **Meat**: Preservation.)

Food is processed to prevent spoilage. As the intensity of the preservation process increases, the types of bacterial spoilage are changed or eliminated. Spoilage of the other three categories of processed foods reflects these changes.

Semiperishable or Lightly Processed Foods

Semiperishable or lightly processed foods are subject to much the same type of spoilage as the unprocessed or perishable food described above. The preservation

processes used with this type of food are usually a mild heat treatment coupled with refrigerated storage, below 7 °C. Pasteurized refrigerated dairy products or pasteurized refrigerated egg products are good examples of lightly processed foods. The heat treatment used with these foods will usually destroy psychrotrophic bacteria and nonspore-forming mesophilic bacteria, but the handling of the foods after heat processing usually results in contamination of these products with spoilage-type bacteria. Such spoilage bacteria have been found in unclean equipment, moisture droplets on the equipment, and in liquid drain traps. During subsequent refrigerated storage, the psychrotrophic bacteria will grow and, if storage time is long enough, spoilage will result. The same genera given above for the raw or unprocessed foods can be involved in the spoilage of the lightly processed foods. Foods that have received this type of processing will usually maintain desired quality attributes for 1–3 weeks. (*See* **Heat Treatment**: Chemical and Microbiological Changes; **Pasteurization**: Principles.)

Semishelf-stable Foods

Semishelf-stable foods are preserved utilizing more than one preservation process. This process has been referred to as the hurdles concept. The hypothesis of the hurdles concept is that each less than lethal preservation process becomes one more hurdle for the bacteria to overcome before they can initiate growth in the food. These types of foods are becoming popular as they best reflect the quality attributes of fresh foods. They are often referred to as 'fresh-like' or 'chef-like,' implying a freshly prepared product, which they are not. Foods that have been processed using the hurdles concept will usually maintain desired quality attributes for anywhere from 1 to 6 months.

This concept might best be explained by following a product through such a series of processes or hurdles. For example, in the processing of cooked beef products, such as roast beef, several steps are involved. Usually the processor will receive 'box beef' which would be containers of the particular cut of meat to be used. These cuts were initially processed in another facility and shipped by refrigerated transport. The first step in the processing of the beef roasts involves trimming and shaping the meat piece. This removes some surface contamination but does little to extend the shelf-life. The second step might involve massaging the meat and adding salt and other seasonings. The salt and seasoning will have minimal effect on extension of shelf-life, unless the salt concentration is above 4 or 5%. If sodium nitrite is used in the seasonings, some inhibition of spore germination may result. This would be the first hurdle.

The seasoned, massaged meat is next put into heavy polyethylene cook bags and slowly cooked in water to the desired internal temperature. This process usually takes several hours and should result in the destruction of psychrotrophic and mesophilic spoilage bacteria, the most common types of bacteria that might spoil this product. The cooking process is a major hurdle as it effectively destroys most spoilage organisms.

The cooked meat is cooled to less than 7 °C and the cook bag removed. It is at this point in the process that recontamination with psychrotrophic and mesophilic spoilage bacteria can occur. The meat is cut into pieces of a size suitable for resale, rebagged in oxygen impermeable polyethylene bags and vacuum-sealed. The vacuum package changes the oxidation reduction potential in the package and prevents the growth of aerobic psychrotrophic bacteria. The finished package should be maintained at low temperature – less than 2 °C – until utilized.

A cooked beef product, processed as described, should have a shelf-life of at least 60 days, if not 120 days. It is not shelf-stable, as spoilage will occur – usually, thermoduric, psychrotrophic, facultative; heterofermentative lactobacilli will cause spoilage which becomes apparent by the swelling of the bag caused by the gas produced and acid flavor of the product.

A further look at each step will help to form an understanding of how the hurdle process achieves better maintenance of quality attributes in the food. (*See* **Preservation of Food; Chilled Storage**: Packaging Under Vacuum.)

The cooking step will kill the potential psychrotrophic spoilage bacteria, but others, such as thermoduric bacteria and spore-forming bacteria, will survive the heat process. The addition of salt and nitrite to the meat will deter the growth of spore-forming bacteria, but will have little effect on nonspore-forming bacteria. Vacuum packaging changes the E_h from positive to negative, which inhibits the outgrowth of psychrotrophic aerobic spoilers but allows the growth of anaerobic and facultative anaerobic bacteria. The endresult is that the meat will eventually spoil. In other words, it becomes esthetically unacceptable as a result of the growth of thermoduric, psychrotrophic, facultative anaerobic bacteria that grow slowly at refrigeration temperatures.

Shelf-stable Foods

Shelf-stable foods have undergone preservation processes that have produced foods that are considered

to be 'commercially sterile,' i.e., will not spoil or cause disease under normal conditions of handling and distribution. These foods are not completely sterile and are subject to some type of bacterial spoilage. For example, aciduric and thermoduric bacteria may spoil foods rendered shelf-stable by reduction of the pH to less than 4.6 and application of a pasteurization heat treatment. Certain species of *Lactobacillus*, *Clostridium*, and *Bacillus* are capable of surviving a pasteurization heat process and, under the right conditions, will grow in high-acid foods. It is generally recognized that high-acid foods are more susceptible to spoilage by yeasts and molds but, as indicated above, some bacteria will grow and spoil even heat-processed acid foods.

Canned foods are given a severe heat process, which is primarily designed to destroy spores of *Clostridium botulinum*. Other spore-forming bacteria such as *C. sporogenes* PA 3679 and *Bacillus stearothermophilus* will not be destroyed by a thermal process designed to control *C. botulinum* and subsequently may grow in canned foods. Canned food spoilage, referred to as 'flat-sour' spoilage, is caused by the outgrowth of *B. stearothermophilus*, which produces acid but no gas in the canned food. Other *Bacillus* may also survive minimal heat process and cause spoilage of canned foods.

The outgrowth of spores of *C. sporogenes* in canned foods produces considerable amounts of noxious gas and may lead to rupture of the cans. Canned food spoilage will also occur if the food has been underprocessed, i.e., not given sufficient heat treatment to produce a commercially sterile product. Spore-forming bacteria which have survived the process usually cause canned food spoilage of this type. This type of spoilage is apparent from the sour or putrid odor and frothy or viscous appearance of the spoiled canned food. (*See* **Canning**: Quality Changes During Canning.)

Failure of the double seam in canned foods results in what is termed 'leaker spoilage.' This type of spoilage is characterized by the presence of several different types of bacteria in the food. These bacteria may cross the defective double seam from the cooling water, dirty conveyor system, can unscramblers, or other can-handling equipment. Foods thermally processed in glass containers will spoil in much the same manner if the closure mechanism used on the glass jar is defective or does not seal properly.

Dried foods are also considered to be shelf-stable or 'commercially sterile.' The available water in dry food systems is less than 0.80 and, as a result, they are not subject to spoilage by bacteria as long as the available water remains low. Yeasts and molds are more tolerant to low available water and may spoil these type of foods if the a_w is more than 0.60 but less than 0.80. Foods preserved by controlling water activity are more frequently 'spoiled' by the action of enzymes than by direct bacterial growth in the foods. (*See* **Drying**: Theory of Air-drying.)

The spoilage of food by bacteria is preventable if the food is processed and handled properly. The preservation processes used affect the probability and type of spoilage that will occur. The greater the number of hurdles that the bacteria have to overcome to grow, the longer the time for spoilage to occur. Contamination with spoilage bacteria can occur at any point in the system that transfers food from the point of initial production to final consumption. Constant attention to the prevention of bacterial contamination of food at each of these process steps will insure a continued safe and wholesome food supply.

See also: **Canning**: Principles; Quality Changes During Canning; **Chilled Storage**: Packaging Under Vacuum; **Cleaning Procedures in the Factory**: Overall Approach; **Contamination of Food**; **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Drying**: Theory of Air-drying; **Fish**: Spoilage of Seafood; **Heat Treatment**: Chemical and Microbiological Changes; **Meat**: Preservation; **Pasteurization**: Principles; **pH – Principles and Measurement**; **Preservation of Food**; **Wastage of Food**; **Water Activity**: Effect on Food Stability

Further Reading

Hood SK and Zottola EA (1995) Biofilms in food processing. *Food Control* 6: 9–18.

Jay JJ (1996) *Modern Food Microbiology*, 5th edn. New York: Chapman and Hall.

Potter NH and Hotchkiss JH (1995) *Food Science*, 5th edn. New York: Chapman and Hall.

Ray B (1996) *Fundamental Food Microbiology*. CRC Press

Zottola EA (1994) Microbes, hurdles, food safety and process optimization. In: Singh RP and Olivereira FAR (eds) *Minimal Processing of Foods and Process Optimization: An Interface*. Boca Raton: CRC Press.

Zottola EA and Sasahara KC (1994) Microbial biofilms in the food processing industry – should they be a concern? *International Journal of Food Microbiology* 23: 125–148.

Zottola EA and Smith LB (1990) Pathogenic bacteria in meat and meat products. In: Pearson AM and Dutson TR (eds) *Meat and Health Advances in Meat Research*, vol. 6. London: Elsevier Science Publishers.

Fungi in Food – An Overview

L B Bullerman, University of Nebraska-Lincoln, Lincoln, NE, USA

Introduction

The kingdom Fungi includes eukaryotic organisms which range from microscopic, unicellular yeasts, to multicellular molds, to the macroscopic mushrooms. Fungi are heterotrophic organisms which obtain food by absorption and require organic compounds for energy and carbon. In terms of food safety and food spoilage by fungi, we are concerned primarily with molds (microfungi) and yeasts. (*See* **Mushrooms and Truffles:** Classification and Morphology.)

Molds

Molds are eukaryotic, multicellular, multinucleate, filamentous organisms that grow as compact masses of intertwining, branching, hairlike filaments. The total mold organism can frequently grow to macroscopic size and be visible to the unaided eye. Thus, the molds are not true microorganisms, but parts of mold structures are microscopic in size. Molds are sometimes referred to as microfungi because they are much smaller than the so-called macrofungi, such as mushrooms. The individual mold filaments are called hyphae (singular: hypha) and a mass of these branching filaments, forming a mold colony or portion of mold growth, is referred to as a mycelium or (plural) mycelia. The hyphae may be submerged, or growing in a food substrate. The submerged hyphae, also called vegetative hyphae, serve to anchor the mold to the substrate and take up nutrients and water by absorption. In this respect they are not unlike plant roots. Hyphae that grow above the substrate, the visible part of the mold, are called aerial or fertile hyphae. They are called fertile because they give rise to reproductive structures known as conidiophores or sporangiophores which produce millions of spores – conidiospores, or conidia, and sporangiospores, or sporangia.

Conidiospores are produced free by special cells on the ends of conidiophores and are not enclosed in any type of structure. Conidia are microscopic in size, very light-weight, and very dry. The conidia are not easily wetted and are hydrostatic, tending to associate with and behave like dust particles. Thus, these spores are readily spread through the air and are disseminated on air currents to new surfaces and habitats. If the spores settle where conditions are favorable for growth, they rapidly germinate and begin to form a new mold colony. Sporangiospores are spores that are produced in an enclosed, sac-like structure, called a sporangium, at the end of the sporangiophores. These spores are released into the air when the sporangium ruptures. Conidia and sporangiospores are common in the molds found in foods. In addition, individual molds may form other types of spores, such as arthrospores, resulting from fragmentation of septate mycelia, and chlamydospores which result from a thick wall that develops around individual mycelial cells. These spores are all referred to as asexual spores, i.e., they are formed by nonsexual means and without exchange of any genetic material by two different molds. Many of the molds that are important in foods reproduce in this way, without sexual stages in their life cycles, and are known as anamorphs. In older literature these are placed in a group known as the Fungi Imperfecti. These are molds for which the sexual cycles are unknown.

Some of the molds that are important in foods also reproduce by sexual means in addition to asexual means. They form, by sexual processes, other types of spores, including ascospores and zygospores. These are referred to as teleomorphs (perfect molds, or the perfect state in older literature), or the teleomorphic state. Ascospores are usually formed in some type of enclosed, sac-like structure located down in the mycelial mass. These structures are known as asci (singular: ascus). The asci are enclosed in a fruiting body called an ascocarp. An ascocarp that is spherical or flask-shaped and with no opening is known as a cleistothecium. A spherical or flask-shaped ascocarp with an opening (ostiole) is called a perithecium. A saucer or cup-shaped ascocarp that is open is known as an apothecium. A zygospore is formed when the tips of two hyphae come together and their contents fuse. The zygospore develops as a thick-walled structure between the two hyphal tips. All sexual spores result from the fusion of two haploid nuclei.

Occurrence

Molds are ubiquitous organisms, which literally means they are found everywhere. The natural habitat of most molds is the soil, where they grow on and break down decaying vegetable matter. Molds rot wood, leaves, and other organic debris to form humus in the soil. During growth on organic debris, molds form their spores which are readily picked up by air currents and widely disseminated. When there is decaying organic matter in an area, there are often high numbers of mold spores in the atmosphere of the area. Thus most plant materials, including grains and seeds, will have a microflora that includes mold spores. In addition, some grains and seeds

will be colonized by mold while yet in the field, and these too will become a part of the microflora of these products. Many of these molds are capable of growth within the seed under proper conditions. Molds are therefore common contaminants of food and animal feed materials, and occur throughout the environment.

Growth Requirements

Molds can tolerate relatively harsh environments and adapt to more severe stresses than most microorganisms. Molds require less available moisture for growth than bacteria or yeasts and can grow on substrates containing concentrations of sugar or salt that bacteria cannot tolerate. Molds can also grow on drier substrates than bacteria and can survive in dehydrated environments. Molds can tolerate and grow in high concentrations of acid and over a wide pH range (2.0–9.0). Since molds are slower-growing than bacteria and yeasts, the dry or acidic conditions that inhibit the growth of these organisms, especially bacteria, flavor the growth of molds. If competitive growth of bacteria is eliminated by either lower moisture or low pH, mold growth will be enhanced. Thus mold growth in grains and animal feeds with lowered moisture content may sometimes occur if the moisture content is not too low. Some molds can obtain moisture from the atmosphere, as well as the respiration of other organisms, such as insects, and begin to grow at even low moisture levels. Once mold growth begins it will probably continue, being fed by moisture released through its own respiration.

Most molds are highly aerobic, i.e., they require oxygen for growth, and an abundant supply of oxygen enhances growth. Carbon dioxide, on the other hand, inhibits mold growth, and if the carbon dioxide concentration becomes high enough it can totally prevent mold growth. Some molds are better adapted than others to growth in elevated concentrations of carbon dioxide. Molds also grow over a wide temperature range, but most have their optimum growth temperatures between 25 °C and 35 °C. Mold growth is especially rapid under conditions of high temperatures and high humidity. Some molds can grow at temperatures as low as 0–5 °C and a few species can even carry on growth processes just below freezing. A few molds can grow at extremely high temperatures, even above 60 °C.

Molds have simple nutritional requirements. They can utilize a range of organic substrates, from simple to complex, requiring primarily a source of carbon and simple inorganic nitrogen. Molds are capable of synthesizing their own vitamins and growth factors. Molds can readily utilize simple carbon sources such as glucose and other sugars, as well as complex carbohydrates such as starch and cellulose. Molds can also utilize inorganic nitrogen in the form of nitrate and ammonium salts, and organic nitrogen such as that found in proteins, amino acids, and nucleic acids. Because molds possess a vast array of hydrolytic enzymes, they are capable of dissimilating and utilizing a very wide array of substrates. Thus most organic materials are subject to deterioration by molds if the conditions of moisture and temperature are not limiting.

Molds and Food Spoilage

Molds are very efficient in converting nutrients into cell material and mycelial biomass. If a substrate is composed of nutrients in moderate amounts, most of the substrate will be converted to cell biomass and products of primary metabolism for essential life processes. If nutrients are available in large or excessive amounts, a variety of breakdown products may be excreted into the medium, and storage reserves of carbohydrates and lipids may accumulate in the mycelia. At some point in the life cycle of the mold, when growth slows and conditions are appropriate, the mold may convert these carbohydrates and lipids to alcohols, organic acids, and complex heterocyclic biochemical compounds. This process is known as secondary metabolism since the metabolites or compounds produced have no apparent purpose in essential life processes. These 'secondary metabolites' may thus accumulate in the substrate, causing off-flavors and other problems, including toxicity. The filamentous hyphae of the molds are well adapted to growth over surfaces and through or into porous and solid substrates. The hyphae provide a large surface area relative to the mold biomass to excrete enzymes that break the substrate down into available nutrients, which in turn are absorbed back into the hyphae. The nutrients may be transported to the actively growing hyphal tips, where they are utilized for energy and to produce primary metabolites and form new active cytoplasm. The nutrients may also be used for maintenance of cellular activities, or be converted to cellular storage reserves and secondary metabolites.

As a result of the metabolic activity of molds in a substrate, a number of consequences – desirable or undesirable – can occur. If the substrate is a food or an animal feed, the biochemical activities of the mold may result in deterioration and spoilage as the substrate is broken down, and undesirable byproducts accumulate, causing off-flavors, loss of dry matter and nutrients, as well as toxicity and other problems. Particularly in drier substrates, such as grains and

animal feeds, such deterioration will result in spoilage. Some of the secondary metabolites of certain molds are compounds that are toxic to humans and animals. These toxic substances are collectively known as 'mycotoxins.' (*See* **Mycotoxins**: Occurrence and Determination.)

Control of Mold Through Processing and Storage

Mold growth is affected by a number of factors, including atmosphere, moisture content, relative humidity, temperature, microbial competition, and chemicals in the substrate. Since molds are very tolerant of acidic conditions, and have few nutritional requirements, the pH and nutrient content of a substrate cannot be used to affect to any significant extent the ability of molds to grow.

Moisture and temperature, however, are probably the two most critical factors affecting mold growth that can be used as control factors, and it is difficult to discuss one without the other. The moisture content of a substrate is less meaningful, in terms of understanding the effect of water on mold growth, than water activity (a_w). Water activity has taken the place of moisture as the most useful expression of the availability of water for microbial growth. The a_w of a substrate is defined as the ratio of the vapor pressure of the substrate to that of pure water. (*See* **Water Activity**: Effect on Food Stability.)

The a_w of pure water is 1.0; therefore the a_w of any substrate will be less than 1.0. The a_w is a measure of the amount of water not bound by the substrate, and which is available to microorganisms for their growth. The lower the a_w, the less water is available to molds for growth. Moreover, the a_w of the substrate is affected by the relative humidity (RH) of the environment in which the substrate is found. RH refers to the moisture in the atmosphere surrounding the substrate, and a_w is a property of the substrate and its moisture content. In a closed system, the a_w of a substrate and the RH of the atmosphere surrounding it will be in equilibrium. Under equilibrium conditions, the a_w of a substrate will be equal to the RH of the atmosphere surrounding it divided by 100. Thus the final moisture content of a substrate, i.e., food or feed, will depend upon the substrate reaching equilibrium with the atmospheric RH to which it is exposed.

The concept of a_w is important in understanding how molds may grow in a seemingly dry substrate. Conditions such as high humidity or a microenvironment affected by insect respiration can cause the a_w of a small portion of a substrate to rise to a level that may permit mold growth. Once growth has been initiated, respiration of the mold will also contribute to increasing a_w of the surrounding substrate. In this way the mold growth may become a spreading, self-perpetuating process, with the pocket of growth becoming ever larger. This is what causes 'hot spots' in a mass of stored grain or animal feed.

The minimum a_w at which microorganisms associated with foods and feeds will grow have been studied extensively. The minimum a_w for the general groups of microorganisms are given in Table 1. Halophilic bacteria, xerophilic molds, and osmophilic yeasts are adapted to growth at very low a_w. Most spoilage and toxigenic molds grow over the a_w range 0.72–0.94. Mold growth is completely inhibited at a_w values below 0.65. This would be equivalent to moisture contents well below 20%.

Temperature and a_w interact to affect microbial growth. If temperature is near the optimum growth temperature of a mold, the range of a_w over which the mold can grow will be greatest, and at any temperature the ability of a mold to grow will be reduced as the a_w is reduced. Conversely, if the a_w of a substrate is high, molds will be able to grow over a wider range of temperatures and will be able to grow at lower temperatures.

Atmospheric gases, other than moisture, also affect mold growth. Molds require oxygen to grow, and are inhibited by increasing concentrations of carbon dioxide or decreasing concentrations of oxygen from those found in air. A concentration of 40% carbon dioxide in air depresses fungal growth, but the carbon dioxide level must be increased to more than 90% to inhibit growth completely. Likewise, reducing the oxygen content of air to less than 2.0% depresses mold growth, but to prevent growth the oxygen level must be reduced to 0.2%. Complete replacement of air with a nitrogen atmosphere will also inhibit mold growth. Controlled-atmosphere (CA) storage of commodities has been studied as a measure of preventing mold deterioration. Both the carbon dioxide and oxygen concentrations are usually controlled. Controlled atmosphere with 10% carbon dioxide and 2.0% oxygen greatly increases the time

Table 1 Minimum water activity (a_w) requirements of microorganisms

Microorganism	*Minimum* a_w
Most spoilage bacteria	0.90
Most spoilage yeasts	0.88
Most spoilage molds	0.80
Halophilic bacteria	0.75
Xerophilic molds	0.65
Osmophilic yeasts	0.60

for mold growth to begin and reduces the amount of growth. (*See* **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs.)

The presence of other microorganisms tends to restrict fungal growth if conditions are favorable for the growth of the other microorganisms. Bacteria and yeasts are capable of more rapid growth than molds and tend to overgrow molds. The rapid growth of bacteria on fresh meat, for example, is the probable reason for the fact that molds are rarely seen growing on these substrates. Lactic acid bacteria have been shown to be competitive with molds, and to suppress mycotoxin production. Molds are also competitive with each other, and under certain conditions one mold may prevent the growth of another mold or may alter its growth patterns and metabolism. Competition by bacteria and yeasts and between mold species is affected by the microenvironment of the substrate. The a_w, RH, and temperature all have an impact on competition and growth and will determine which organism or group of organisms will predominate. (*See* **Lactic Acid Bacteria**.)

Mold growth is also affected by chemicals in the substrate which have antimicrobial or antifungal properties. These chemicals may be naturally occurring compounds in the substrate, or they may be added for the purpose of preservation. Naturally occurring substrates, such as benzoic acid in cranberries, and components of essential oils in herbs and spices, may restrict or prevent fungal growth. Mold growth may also be prevented in foods and feeds by the addition of antifungal or antimycotic chemicals. These substances may be organic acids such as sorbic, propionic, and benzoic acids, among others, or salts of these acids, antibiotics such as natamycin, chemical dyes such as gentian violet in the case of poultry feeds, antioxidants, or combinations of these and various other chemicals. In most cases the levels of the chemicals used are such that they are fungistatic, i.e., they prevent or delay mold growth, but do not kill or completely inhibit growth for an indefinite period of time. Molds are not very heat-resistant and are readily killed by most thermal processes, such as cooking, baking, etc. (*See* **Essential Oils**: Properties and Uses; **Fungicides** ; **Preservation of Food**.)

Classification and Specific Genera of Molds

There are about 100 000 different species of fungi, although relatively few of these are directly involved in the deterioration of food and agricultural products, and/or mycotoxin production in these products. Genera that are of particular interest because of their involvement in deterioration of foods and agricultural commodities, and potential mycotoxin production are as follows: *Aspergillus, Penicillium, Fusarium, Alternaria, Trichothecium* and *Trichoderma*. Other genera, important primarily as spoilage organisms, include *Rhizopus, Mucor*, and *Cladosporium*. The most significant mycotoxin-producing molds are considered to be numerous species in the genera *Aspergillus, Penicillium*, and *Fusarium. Alternaria, Trichothecium*, and *Trichoderma* contain a few individually toxic species that are important. Descriptions of nine of the most common genera of molds found in foods are given below. The first eight are septate molds, i.e., the hyphae have cross-walls, while the last one is representative of nonseptate molds, in which the hyphae do not have cross-walls.

Aspergillus

Aspergillus is a very important genus of molds, very widespread, with species involved in food spoilage, mycotoxin production, and fermentations. The *A. flavus–oryzae* group contains the species *A. flavus* and *A. parasiticus*, which can produce aflatoxins. This group also contains *A. oryzae*, which is nontoxic and is used in oriental food fermentations to produce items such as soya sauce and miso. The *A. ochraceus* group includes *A. ochraceus* and other species that are capable of producing ochratoxins and penicillic acid. The *A. niger* group is widespread, and often involved in food spoilage. *A. niger* is now known also to be a producer of ochratoxin. What used to be called the *A. glaucus* group is now known as the anamorph of the genus *Eurotium*, which includes *E. glaucus* and *E. repens*. These are xerotolerant molds that can grow on foods that are very dry or contain high concentrations of sugar or salt. The aspergilli reproduce by producing conidia, which are produced on a conidiophore that arises from a special cell in the mycelium called a foot cell (**Figure 1**). The entire conidiophore appears to be a single cell that grows upright and terminates in a globose, elliptical, or clavate swelling known as a vesicle. Arising from the vesicle are bottle-shaped structures called sterigma (phialides) in which the conidia are produced. The conidia are pushed out of the end of the sterigma and remain loosely attached to one another, forming chains. The colors of the conidia of the aspergilli are characteristic for different groups and species, and this is helpful in recognizing the different groups. Spores of the *A. flavus–oryzae* group are different shades of olive-green to yellow-green; spores of the *A. niger* group are jet-black to brownish-black to purple-brown; spores of the *A. ochraceus* group range from buff-tan to yellow in color. *E. glaucus* and *E. repens* possess conidia that are some shade of

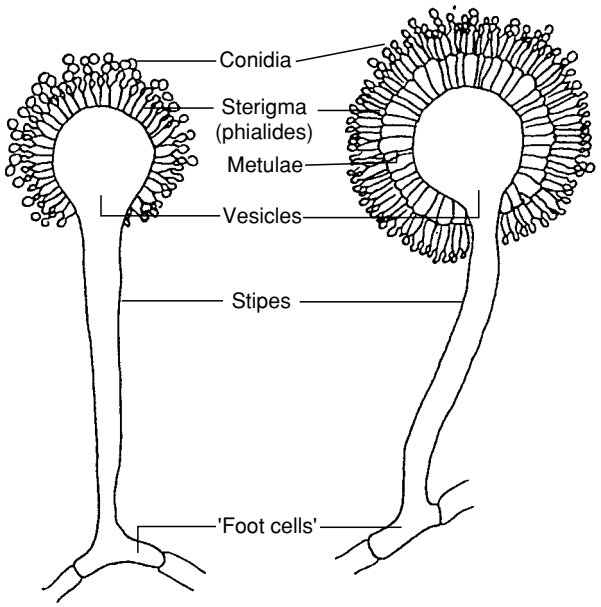

Figure 1 Conidiophores characteristic of *Aspergillus* species, showing a single layer of phialides or sterigma (uniseriate) and double layer of cells, phialides and metulae (biseriate). Adapted from Klich MA and Pitt JI (1988) *A Laboratory Guide to the Common Aspergillus species and Their Telemorphs*. North Ryde, Australia: Commonwealth Scientific and Industrial Research Organization.

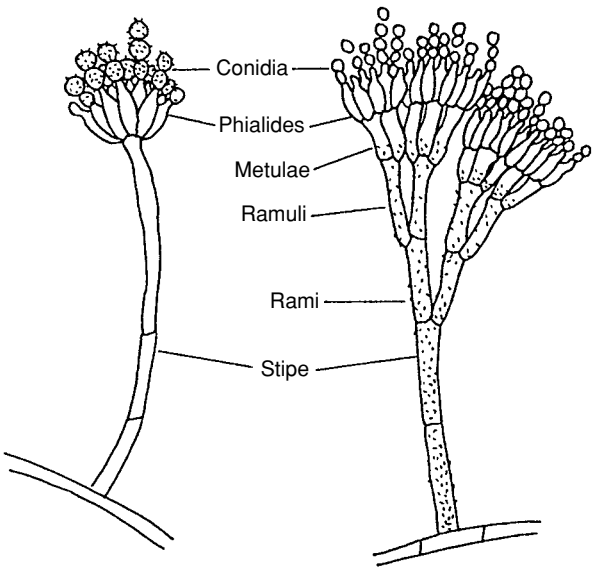

Figure 2 Examples of the simplest and most complex conidiophores produced by *Penicillium* species. Adapted from Pitt JI (1988) *A Laboratory Guide to Common Penicillium Species*, 2nd edn. North Ryde, Australia: Commonwealth Scientific and Industrial Research Organization.

green, but also produce brightly colored yellow to reddish cleistothecia. Many *Aspergillus* species also produce sclerotia, which are macroscopic, hard, dense masses of hyphae that appear as small, dark-colored masses in the mycelia. *Aspergillus* species often occur in grains, nuts, oil seeds, and on certain types of dry cured meat.

Penicillium

Penicillium is also a widespread genus that is important in foods. *Penicillium* species contaminate a wide variety of foods and are capable of growing at refrigeration temperatures. Thus they often spoil refrigerated foods, especially cheese. They are also common on grains, breads, cakes, fruits, preserves, cured and aged hams and sausages, and in the spoilage of certain fruits. The penicillia produce conidia from conidiophores that branch near the apex, forming a brush-like structure or penicillus (Figure 2). At the apex of the conidiophore are somewhat enlarged cells known as metulae. From the metulae arise the sterigma or phialides. It is in these structures that the conidia are produced and pushed out in chains. The conidia of the penicillia are colored, but mostly in shades of gray to blue to blue-green. The colors are not as distinctive for the various species as for the aspergilli, and are therefore not as helpful in the identification of species. Some species form ascospores in cleistothecia and are also placed in the teleomorphic genera of *Talaromyces* or *Eupenicillium*. There are a number of important *Penicillium* species. *P. verrucosum*, *P. viridicatum*, and *P. aurantiogriseum* are common in grains and some can also occur on cheese. They can produce a number of mycotoxins, including ochratoxin, penicillic acid, and others. *P. martensii*, a synonym for *P. aurantiogriseum*, has been found growing in high-moisture corn and can produce penicillic acid. *P. expansum* causes rots of fruits, especially apples, and produces patulin. *P. digitatum*, with green-colored conidia, causes soft rot of citrus fruit, usually at ambient temperatures, whereas *P. italicum*, which has blue spores, causes soft rot of citrus at refrigerated temperatures. *P. roqueforti* has blueish-colored conidia and is used in the ripening of blue-veined cheeses. However, wild types of *P. roqueforti* often occur in dairy environments and also contaminate other types of cheeses, such as Cheddar and Swiss, and grow and cause spoilage under refrigerated storage. *P. camemberti* produces grayish spores, and is used for surface ripening of Camembert and Brie cheeses. Many other *Penicillium* species are known to produce various toxic substances, many of which also have antibiotic properties, but appear to be too toxic for therapeutic use. The most famous antibiotic, penicillin, which has been used to cure countless bacterial infections, is produced by *P. notatum*. (*See* **Cheeses: Chemistry and Microbiology of Maturation.**)

Fusarium

Fusarium species are capable of growing on stored foods, but more commonly occur in the field, where they infect grain plants and may also invade the grain itself. *Fusarium* colonies on agar appear to be cottony owing to dense growth of white hyphae. They produce various pigments, with colors ranging from white, through pink, salmon-pink and carmine red, to purple. Some species also produce yellow and brown pigments. *Fusarium* species produce septate, curved to crescent or sickle-shaped conidia, called macroconidia which are, to some extent, characteristic of species (Figure 3). The macroconidia are slightly curved and tapered towards each end (fusiform) and are sometimes also referred to as boat-shaped. In some species the apical cell of the macroconidium is elongated and the basal cell, where the conidium was attached, is shaped like a foot cell. In addition, some species also produce smaller, one- or two-celled conidia known as microconidia. The microconidia may be pear-shaped (pyriform), fusiform or ovoid, and straight or curved.

Fusarium species are widely distributed throughout the world, in both temperate and tropical regions. They are found in the soil, especially cultivated soils, and actively degrade organic matter in the soil. Many species are also considered to be plant pathogens, capable of causing a range of plant diseases such as root and stem rots, blights, and wilts. Two plant diseases that occur in the field, but which affect grain quality and safety, are wheat scab, also called *Fusarium* head blight, and ear rot in corn (maize). One of the species involved in both of these conditions is *F. graminearum*, which invades the grain (wheat or corn) in the field and produces deoxynivalenol (DON), also called vomitoxin. *F. graminearum* is most common in North America, whereas a closely related species, *F. culmorum*, is more common in Europe. *F. graminearum* and *F. culmorum* also produce zearalenone. *F. verticillioides (moniliforme)* produces fumonisins and other metabolites in corn, as does *F. proliferatum*. *F. proliferatum* and *F. subglutinans* also produce moniliformin and other compounds. Other species can also invade grains and produce other toxins. *Fusarium* species can also invade fresh fruits and vegetables during storage, causing various rots and spoilage.

Some of the *Fusarium* species have a sexual stage and are placed in the teleomorphic genera of *Gibberella, Nectria, Calonectria*, and *Plectoshaerella*. For example, the telemorph of *F. graminearum* is *Gibberella zeae*. *Fusarium* species are very important as plant pathogens and as potential mycotoxin-producing organisms. They can therefore cause both spoilage and safety problems in foods, especially cereal grains and foods containing cereal products. (*See* **Cereals:** Dietary Importance.)

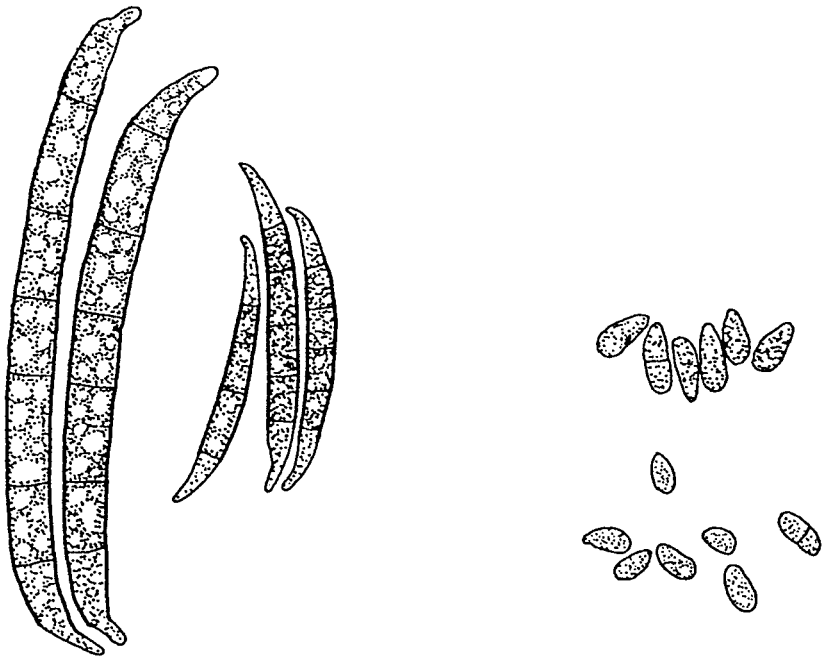

Figure 3 Macro- and microconidia typical of the genus *Fusarium*. Reproduced from Spoilage: Fungi in Food – An Overview. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Alternaria

Species of *Alternaria* occur worldwide on many foods and plant materials. They have septate hyphae and appear to be dark, gray-green in color, and almost black on the reverse side of colonies growing on agar. They produce septate conidia which are also dark-colored (Figure 4). The septa are both transverse and longitudinal, and spores are oblong in shape, sometimes with an elongated apical cell. The spores are produced in chains. *Alternaria* can cause spoilage of a variety of foods, including tomatoes and fresh vegetables, such as bell peppers. *Alternaria* species have been found in the core area of apples, in nuts, including peanuts, hazelnuts, and pecans, in grains such as wheat and sorghum, and in cold-stored meat and spices. *Alternaria* are capable of growing at low temperatures on refrigerated foods. Common species include *A. alternata*, which is considered synonymous with *A. tenuis*, and *A. tenuissima*. *Alternaria* species are capable of producing several mycotoxins and thus present both food spoilage and potential safety hazards.

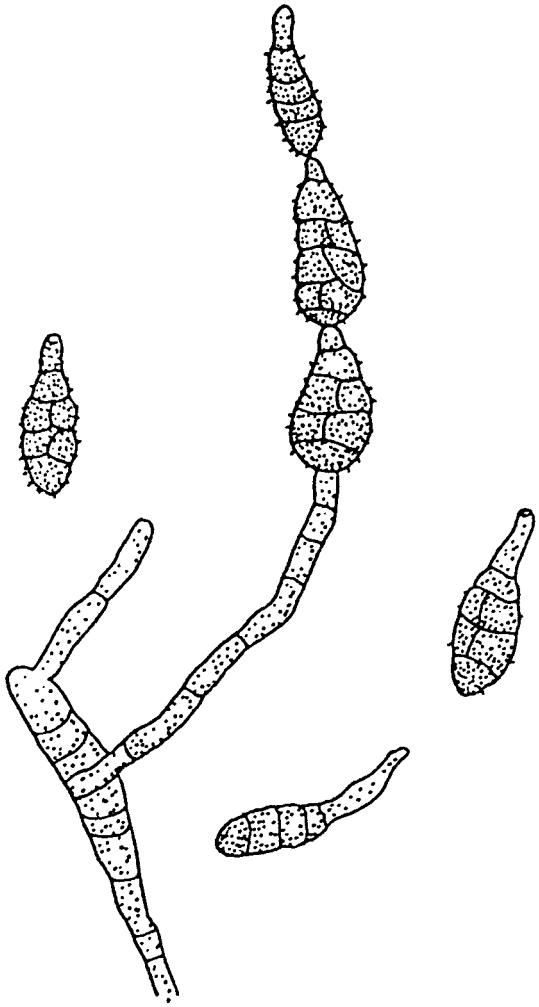

Figure 4 Multicelled conidia of *Alternaria* with transverse and longitudinal septa. Reproduced from Spoilage: Fungi in Food – An Overview. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Trichothecium

Trichothecium is a common genus of mold which occurs throughout the world. A common species is *T. roseum*, which is pink in color, and has been found growing on various fruits and vegetables, as well as cereal grains, such as barley, wheat, and corn, and cereal products such as flour. It has also been found on beans, pecans, filberts, and meat. *Trichothecium* species produce clusters of two-celled or single-septate conidia that are elliptical to pear-shaped. Some strains of *Trichothecium* have been shown to produce trichothecenes.

Trichoderma

Two of the most common species of *Trichoderma* are *T. viride* and *T. harzianum*. *Trichoderma* colonies are bright green. *Trichoderma* contaminate many food products, including stored grains such as barley, wheat, and oats, nuts such as peanuts and pecans, and various fruits and vegetables, such as tomatoes, sweet potatoes, and citrus fruit.

Cladosporium

Cladosporium has dark mycelia which may be brown to blackish-brown or gray-green in color. The reverse side of a colony of *Cladosporium* on agar is very dark greenish-black or blue-black. *Cladosporium* species produce many one-celled conidia, but two- and three-celled forms are also common. Two common species are *C. cladosporioides* and *C. herbarum*. *Cladosporium* species, which can grow at low temperatures, tend to be slow-growing and form black spots on foods. *Cladosporium* species have been isolated from cereal grains, peanuts, fruits, and refrigerated meat, particularly beef. *Cladosporium* is very common and can grow under refrigerated storage conditions. The main consequence of *Cladosporium* species in foods is spoilage and discoloration.

Geotrichum

Geotrichum has septate mycelia which readily fragment into arthrospores (Figure 5), which are the organism's primary means of reproduction. The only species that is significant in foods is *G. candidum*; other names for this organism are *Oidium lactis* and *Oospora lactis*; it produces ascospores, and the

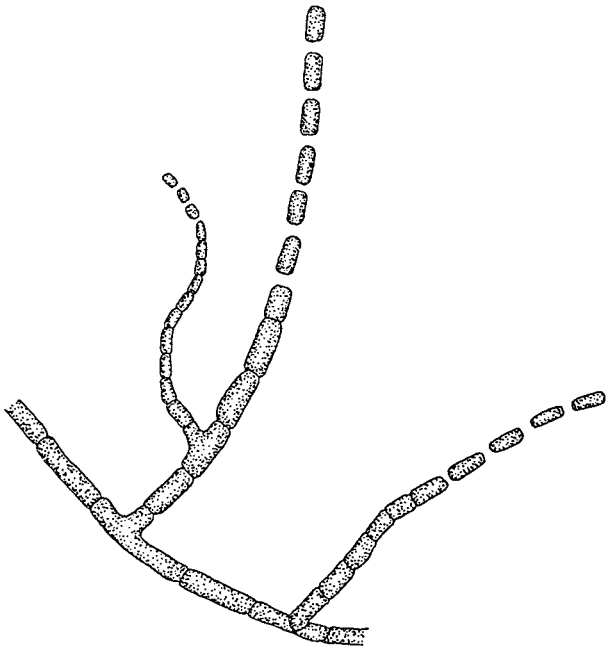

Figure 5 Arthrospores typical of *Geotrichum*. Reproduced from Spoilage: Fungi in Food – An Overview. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

telemorphic state is known as *Endomyces geotrichum*. *G. candidum* is referred to as a yeast-like fungus because it exhibits a dimorphic form and readily fragments into arthrospores, which appear to be single-celled. In addition, colonies of this organism can be white, soft, creamy, and yeast-like in appearance. *G. candidum* is a problem on food-processing equipment, particularly in vegetable-processing plants, and is referred to as 'machinery mold.' Unclean equipment provides a favorable environment for rapid growth of this organism. It attacks citrus fruits, as well as other fruits. It is able to infect fruits mainly through injuries to the skin, as well as growing on overripe fruit. *G. candidum* has also been obtained from cheese, meats, and frozen foods, particularly vegetables.

Rhizopus

Rhizopus differs from the molds previously described, because it is nonseptate and produces sporangiospores rather than conidia. It is a very-fast-growing, spreading type of mold which has white mycelia and black sporangia. *Rhizopus* species form rhizoids at the base of the sporangiophores, and columella in the sporangium (Figure 6). Young sporangia are white before turning black with age. Probably the most common of the *Rhizopus* species is *R. stolonifer*, the so-called bread mold. Besides bread, *Rhizopus* causes spoilage of strawberries, other berries, fruits, and vegetables. *Rhizopus* species have also been isolated from cereal grains, nuts, and meat. *R. oligosporus* is used in making tempeh and certain other mold-fermented foods. *Rhizopus* is placed in the taxonomic order Mucorales and is somewhat typical, and probably the most common genus of that order to contaminate foods. Other genera in the order Mucorales which are nonseptate and produce sporangiospores are *Absidia*, *Mucor*, *Rhizomucor*, *Syncephalastrum*, and *Thamnidium*. All of the Mucorales genera that contaminate food are found in the family Mucoraceae. These genera are sometimes referred to as mucoraceous fungi. They are a part of the

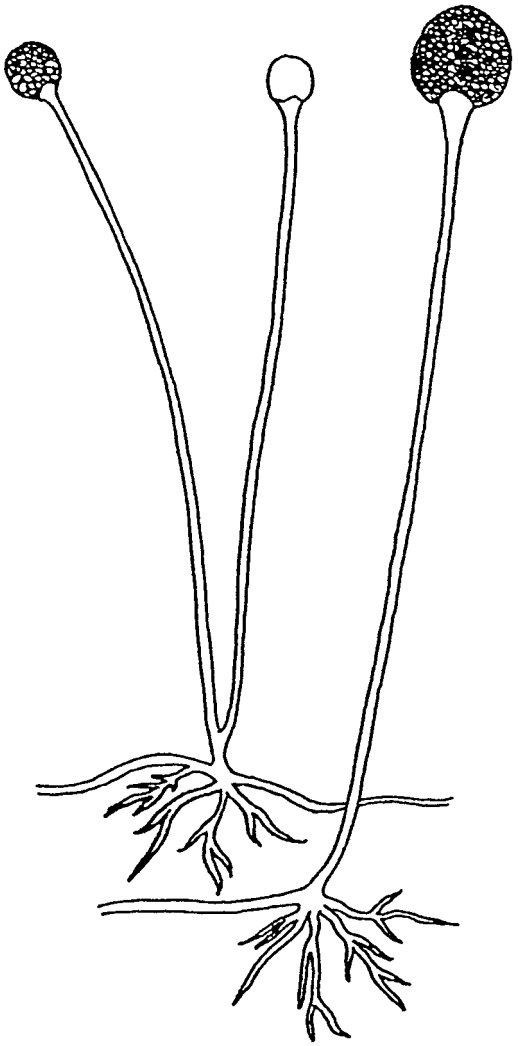

Figure 6 Sporangiospore typical of *Rhizopus* and other mucoraceous fungi. Reproduced from Spoilage: Fungi in Food – An Overview. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

class Zygomycetes and form zygospores by sexual processes and chlamydospores (Figure 7). They require high-moisture conditions for growth.

For more detailed descriptions of these and other genera of molds, refer to the references given in the Further Reading Section at the end of this article.

Recognizing Molds

Molds are recognized primarily as visible fuzzy, cottony, or colored growth on the surfaces of substrates such as foods or other organic materials. In grains, seeds, and animal feeds, however, such growth may not always be visible in this manner. Internal mold contamination and growth in nuts, seeds, and grains may not always be visible as filamentous growth. However, mold growth in nuts, whole grains or seeds often results in discoloration and altered appearance. In corn (including popcorn), for example, invasion and growth of mold in the germ area may produce darkened, greenish- or blueish-colored germ areas in affected kernels. When this growth is attributable to certain blue-colored *Aspergillus* or *Penicillium* species, and a blueish line appears in the germ area, the condition is known as 'blue eye.' In other grains, such as wheat, white to pinkish discolored and shrivelled kernels may indicate mold invasion. Likewise, with nuts, shrivelled or misshapen kernels may be caused by mold, although that is not the only cause of such problems.

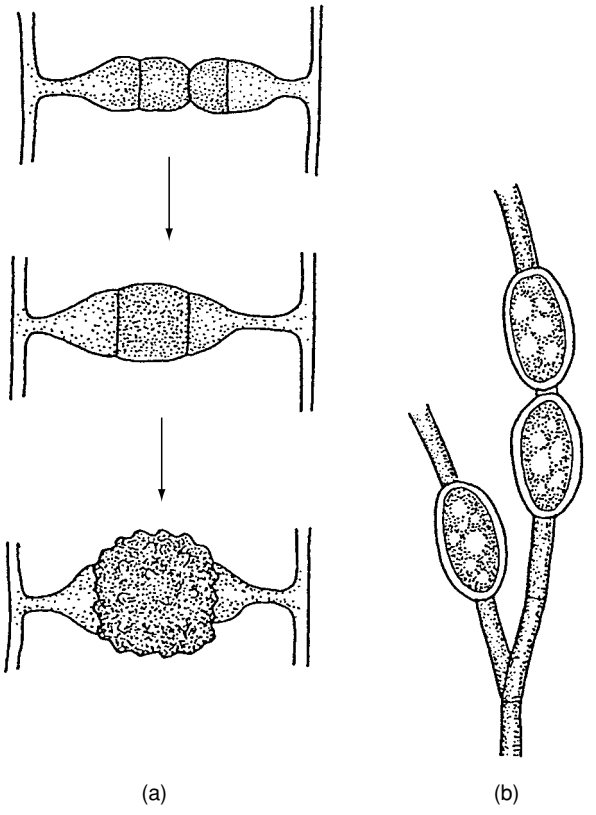

Figure 7 (*a*) Zygospore and (*b*) chlamydospores. Reproduced from Spoilage: Fungi in Food – An Overview. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Health Implications

Mold deterioration of foods and agricultural commodities is a serious problem worldwide. Besides deterioration of food and feed products by causing damage that reduces quality, grades, and prices, resulting in economic loss, the presence of molds in foods and commodities poses hazards to human and animal health. A serious hazard associated with mold growth in foods and feeds is the possible production of mycotoxins – substances which are toxic and some of which are carcinogenic. (*See* **Aflatoxins**; **Mycotoxins**: Classifications.) Additional hazards arise from the presence of the molds themselves. A wide spectrum of human and animal diseases are attributable to molds, ranging from mycotic abortion to aflatoxin poisoning in animals, and from allergic reactions to life-threatening systemic infections in humans. In recent years, invasive mycoses have become a more frequent and important cause of morbidity and mortality in humans, particularly in individuals who are immunosuppressed as a result of organ transplantation, underlying disease (acquired immune deficiency syndrome or AIDS), chemotherapy, radiation therapy, or age. The environment, including foods, is the source of these molds. Certain occupational niches, such as grain handling and storage, pose additional hazards to agricultural workers through exposure to high levels of mold spores and endotoxins in grain dusts. These dusts may also be a significant source of exposure to mycotoxins. Various lung diseases in grain handlers, farmers, and food-producing animals are associated with molds and grain dust.

Yeasts

Yeasts are single-celled microscopic organisms that are larger than bacteria, and are oval, elliptical, spherical, or elongated into rod shapes. Most yeasts reproduce by a process called budding (Figure 8). Budding is an asexual process in which a bulge (bud) develops on the cell wall, and protoplasm, including replicated nuclear material, fills the bulge. The bulge grows in size and eventually a wall forms between the bulge and the parent cell, forming a new cell. The new cell then separates from the parent cell. Buds may form in

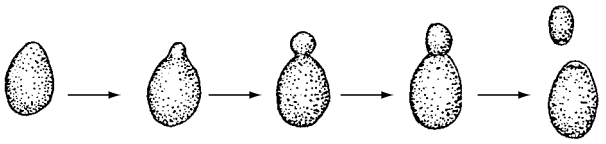

Figure 8 Diagrammatic representation of budding in a typical yeast cell. Reproduced from Spoilage: Fungi in Food – An Overview. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

a polar position or multilaterally. Some yeasts reproduce by binary fission, similar to fission in bacteria, but these are few in number. One yeast reproduces by both fission and budding. So-called true yeasts also reproduce by a sexual process involving conjugation of two cells and the resultant formation of ascospores within the cells. The cell containing ascospores is then known as an ascus. A few yeasts produce ascospores without conjugation, but the resulting ascospores then undergo conjugation. True yeasts are referred to as ascosporogenous yeasts and are classified in the subdivision Ascomycotina, which used to be called Ascomycetes. Asporogenous yeasts, also called false yeasts, do not produce ascospores or other sexual spores, and are placed in the subdivision Deuteromycotina, also called Fungi Imperfecti.

Another type of spore formed by some yeasts is called a chlamydospore. A chlamydospore forms when a thick wall surrounds the yeast cell, and is similar to chlamydospores formed by molds. These appear to be resting or survival structures, and are asexual. Yeasts that form pseudomycelia can also produce by fragmentation spores known as arthrospores and blastospores.

Growth Requirements

Individual yeasts vary in their growth requirements, but some generalizations can be made. Yeasts are aerobic and facultatively anaerobic. Some yeasts are extremely aerobic and oxidative in their metabolism. These oxidative yeasts grow on the surface of liquids and are known as film yeasts. Yeasts that are aerobic and facultatively anaerobic can grow both ways. When growing aerobically, yeasts produce mainly biomass, i.e., more yeast cells. When growing anaerobically, yeasts have a fermentative metabolism and produce carbon dioxide and ethanol. For the most part, sugars are the best source of energy for yeasts, but the film yeasts can oxidize organic acids and alcohol for energy. Yeasts can utilize simple nitrogen compounds, such as ammonia, urea, and amino acids. Some yeasts can also hydrolyze proteins and polypeptides.

Yeasts require somewhat less available moisture than bacteria, but more than molds. In terms of a_w, the lower limits for most yeasts are 0.88–0.94. Some yeasts, known as osmophilic yeasts, can grow in the presence of high concentrations of sugars or salt at a_w as low as 0.62–0.65. Each yeast will have its own characteristic optimal a_w and a range of a_w over which it can grow. The minimum and optimum a_w for a given yeast may change depending on outside factors, such as nutrition, pH, temperature, oxygen, and presence of inhibitors.

The optimum temperature for growth of yeasts is in the mesophilic range of 25–30 °C. Yeasts in general can grow over a range of temperatures from 0 °C to 47 °C. Yeasts grow well under acid conditions, at pH 4.0–4.5. They can grow at lower pH than most bacteria, but do not grow well under alkaline conditions. Yeasts are widely distributed in nature and commonly occur on grapes and other fruits. They are also found in dust and water, and may be found on the skin and in the intestinal tract of humans and animals.

Food Spoilage and Effects of Processing on Yeasts

Some yeasts, mostly in the genus *Saccharomyces*, are very important industrially for leavening of bread and production of wine, beer, alcohol, glycerol, and invertase. Other yeasts, referred to as wild yeasts, can cause spoilage problems in foods. Film yeasts can oxidize enough lactic acid to allow spoilage in sauerkraut and pickle fermentations by nonacid-tolerant organisms. Yeasts can grow in vacuum-packaged wieners and other cured meats, forming visible colonies or slime. Other salt-tolerant yeasts can grow in curing brines, salted meats and fish, soya sauce, tamari sauce, and miso. Osmophilic yeasts can grow well in foods with high concentration of solutes, especially sugars and salts, and cause spoilage of salad dressings, honey, syrups, concentrated fruit juices, and even dried fruits. Certain other yeasts, known as apiculate or lemon-shaped yeasts, can contaminate wine fermentations, causing off-flavors, low yields of alcohol, and production of volatile acids. Yeasts can also reach high numbers in frozen vegetables prior to freezing or after thawing, especially if the vegetables are mishandled.

Yeasts are not heat-resistant and are readily destroyed in most heat processes. Other food processes, such as freezing, refrigeration, vacuum packaging, and acidification, will not necessarily destroy yeasts. Vacuum packaging will prevent growth of highly aerobic film yeasts, but not of fermentative types. Lower pH will not necessarily prevent the growth of yeasts, but organic acids may prevent growth of

some yeasts. Sorbic acid and sorbates are particularly effective in preventing yeast growth. Sulfur dioxide can be used to kill wild-type yeasts in dried fruits and in grapes for wine-making. Heating of dried fruits in the package can also be used to kill yeasts and pasteurize the fruit.

Specific Genera of Yeasts

There are a number of genera of yeasts that are important in foods. These include both beneficial and harmful types. The following is a synopsis of 10 genera of yeasts that include the most important industrial yeast, and various spoilage yeasts, including osmophilic yeasts, film yeasts, and asporogeneous yeasts.

Saccharomyces

Saccharomyces is the most important genus from an industrial standpoint. The main species is *S. cerevisiae* and different strains of it are used in many food industries for leavening bread and making ales, beer, wine, alcohol, and other products. It has also been used to produce single-celled protein, and food ingredients such as autolyzed yeast. It is probably the single most important yeast known. *S. cerevisiae* cells are elliptical to ovoid in shape, with some spherical and elongated cells as well. They reproduce by budding, and form ascospores. *S. cerevisiae* can also cause spoilage under certain conditions where its fermentative activity is not desired.

Zygosaccharomyces

Zygosaccharomyces is considered to be a subgenus of *Saccharomyces*, and some authors have reclassified this genus as *Saccharomyces*. The yeasts in this genus are osmophilic and can grow in high concentrations of sugar. They are responsible for spoilage of molasses, syrups, honey, salad dressing, soya sauce, and wine. They reproduce by budding, and form ascospores.

Schizosaccharomyces

Schizosaccharomyces species are found in sugar, molasses, honey, tropical fruit, and soil. They reproduce by fission, forming arthrospores, or by sexual means, forming ascospores. There are either four or eight ascospores per ascus. The spores are oval, spherical, or kidney-shaped.

Debaryomyces

Yeasts in the genus *Debaryomyces* are film yeasts, and they form pellicles of surface growth on meat brines. They also grow on cheese and sausage products. The cells of *Debaryomyces* are round to oval-shaped and they form ascospores. They may also form pseudomycelia.

Hansenula

Yeasts in the genus *Hansenula* are also film yeasts. They are common in olive brines, fruit juice concentrates, citrus fruits, and grapes. Although they are film yeasts that form pellicles, they are somewhat fermentative. They produce spherical to oval elongated cells. They reproduce by budding and by forming ascospores. The ascospores have a characteristic shape, similar to a bowler hat. They may form pseudomycelia.

Pichia

Yeasts in the genus *Pichia* are also film yeasts that form a pellicle on liquids, such as beer and wines. Cells are oval to cylindrical-shaped and form ascospores which are either round or hat-shaped.

Torulopsis

The *Torulopsis* genus may also be called *Torula* and it consists of asporogenous yeasts that are round to oval in shape. They reproduce by budding. These yeasts are common in nature, and contaminate many foods, including refrigerated foods. Species of this genus can ferment lactose and can spoil milk products, such as sweetened condensed milk, as well as fruit juice concentrates and acid foods.

Candida

Candida species are asporogenous yeasts that form pseudomycelia or true mycelia. They are sometimes categorized as Fungi Imperfecti in the family Moniliaceae with other yeast-like genera, particularly *Trichothecium* and *Geotrichum*. They reproduce by budding and fragmentation of the mycelia into blastospores. Members of this genus are common in foods and occur in fresh meat, such as ground beef, cured meats, and butter and margarine. They can spoil foods high in acid and salt, and the lipolytic species *C. lipolytica* can cause rancidity of butter and margarine. Other species have been grown for food and feed. At least one species, *C. albicans*, is responsible for infections in humans and animals.

Trichosporon

Trichosporon species are asporogenous yeasts that reproduce by budding and form arthrospores. They grow at lower temperatures and can be found in many foods, including chilled beef, beer, and maple sap.

Rhodotorula

Yeasts in the genus *Rhodotorula* are sporogenous and reproduce by budding. They sometimes also from pseudomycelia. These organisms are widespread in nature, often found in air and dust, and contaminate many foods. They produce red, yellow, and pink pigments and often discolor foods with red and yellow spots on meats, particularly cured meats, and pink areas in sauerkraut.

Health Implications

Yeasts are generally not associated with foodborne illnesses, but certain yeasts can cause infections in humans and animals. Probably the most common infection caused by a yeast is candidiasis, caused by *C. albicans*. This organism is a common inhabitant of the mucous membranes of the mouth, vaginal canal, and intestinal tract. Yeast infections can occur most readily in persons receiving broad-spectrum antibiotic therapy and those who are immunocompromised. In the immunocompromised subjects, *Candida* species and other yeasts can cause endocarditis, urinary infections, kidney infections, esophageal infections, and septicemia or systemic candidiasis. As the number of immunocompromised persons increases from radiation and chemotherapy, AIDS, and simple aging, it remains to be seen if foodborne yeasts may become a problem for these individuals as opportunistic infectious agents.

See also: **Aflatoxins**; **Antibiotics and Drugs**: Uses in Food Production; **Cereals**: Dietary Importance; **Cheeses**: Chemistry and Microbiology of Maturation; **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Essential Oils**: Properties and Uses; **Fungicides**; **Lactic Acid Bacteria**; **Mushrooms and Truffles**: Classification and Morphology; **Mycotoxins**: Classifications; Occurrence and Determination; Toxicology; **Water Activity**: Effect on Food Stability; **Yeasts**

Further Reading

Bullerman LB (2000) *Mold Growth in Bakery Products and its Prevention*. AIB Technical Bulletin, vol. XXII, issue 5. Manhattan, KS: American Institute of Baking.

Doyle MP, Beuchat LR and Montville JT (2000) *Food Microbiology: Fundamentals and Frontiers*, 2nd edn. Washington, DC: ASM Press.

Klich MA and Pitt JI (1988) *A Laboratory Guide to the Common Aspergillus Species and Their Telemorphs*. North Ryde, Australia: Commonwealth Scientific and Industrial Research Organization.

Nelson PE, Toussoun TA and Marasas WFO (1983) Fusarium *Species. An Illustrated Manual for Identification*. University Park, Pennsylvania: The Pennsylvania State University Press.

Northolt MD and Bullerman LB (1982) Prevention of mold growth and toxin production through control of environmental conditions. *Journal of Food Protection* 45: 519–526.

Pitt JI (1988) *A Laboratory Guide to Common Penicillium Species*, 2nd edn. North Ryde, Australia: Commonwealth Scientific and Industrial Research Organization.

Pitt JI and Hocking AD (1999) *Fungi and Food Spoilage*, 2nd edn. Gaithersburg, MD: Aspen Publishers.

Rhodes M (ed.) (1979) *Food Mycology*. Boston: GK Hall.

Samson RA, van Reenen-Hoekstra ES, Frisvad JC and Filtenborg O (1996) *Introduction to Food-borne Fungi*, 5th edn. Baarn, The Netherlands: Centraalbureau voor Schimmelcultures.

Molds in Spoilage

F Ottaviani, State University of Bologna, Italy
M G Ottaviani, CONSAL – Advising in Food Microbiology, Sermide, Italy

Basic Biology of Molds

Molds are eukaryotic organisms widely distributed in nature, and are common contaminants of agricultural commodities, food, beverages, and feed.

In contrast to the simple (prokaryotic) bacteria, molds are complex organisms whose mycelium is composed of individual filaments termed hyphae. The hypha is essentially a tube, consisting of a rigid wall and containing a moving mass of protoplasm; its interior is frequently divided into compartments by cross-walls which, being incomplete or perforated, do not separate the content into individual cells. This coenocytic condition gives molds an advantage over the single-celled bacteria, permitting both a more efficient distribution of nutrients and a better dilution of external stresses over a large number of cells. Organelles inside the hypha are of the eukaryotic type, i.e., more sophisticated than those of the bacterial cell. Nuclei and energy-yielding mitochondria are present.

The hyphal wall is predominantly composed of polysaccharides which give the molds a more defined barrier against environmental stress.

Chitin, a polymer of *N*-acetyl-D-glucosamine, is the more frequently occurring polymer in the walls of fungal spores and mycelia and it does not exist in bacteria or most uncontaminated food and feed products. Glycoproteins such as galactomannan are either

extracellular or bound to the fungal cell wall. Ergosterol is considered the principal sterol of fungi and plays an important role in cell membranes.

All these specific components of fungi are used to estimate fungal activity by rapid chemical, biochemical, and immunological methods even when fungal biomass may have been killed, given that chitin and fungal polysaccharidic antigens are heat-resistant and they can survive after the death of fungal species.

Because of their peculiar structure, molds behave in a different way from bacteria, with regard both to food preservation technologies and to laboratory analytical techniques – they have better control of ionic homeostasis (with regard to pH and water activity (a_w)), and insensitivity to antibiotics commonly used to repress bacterial growth.

With generation times ranging between 2 and 5 h, fungal hyphae extend at their apices with their oldest parts producing branches which fuse between adjacent hyphae. Branched vegetative hyphae form a vegetative mycelium which appears as a light, ill-defined spot on the food or culture medium surface at the beginning of the spoilage or during the initial stage of the colony development. Later, specialized cells emerge from the flat hyphal mass bearing at their ends the staminal components of the molds. These constitute the reproductive mycelium of the mold, whose cells are usually called conidiophores and staminal components either conidia, if asexually formed, or spores, if sexually generated. Sexual formation of spores, which involves genetic rearrangements, greatly contributes to the mold's survival in nature, but it is seldom observed in food and culture media where asexual conidia are more likely to appear. Sexual spores are frequently thick-walled and able to withstand severe environmental conditions, including heat processing, as in the case of the *Byssochlamys* sexual spores which may spoil fruit juices. In common with conidia, which possess a thinner wall, spores differ from vegetative cells in having a thicker tegument, a poorly differentiated cytoplasm, and a low water content, resulting in the ability to survive when environmental conditions become lethal to living cells.

Resting cells, which can withstand many adverse conditions such as drying or oxygen deprivation, may later restore their normal metabolic pathways, eventually giving rise to the first hyphal cell and so commencing the mold life cycle when availability of water, or air, or any other condition suitable to growth, occurs.

Nutrient requirements for mold growth are very simple. A suitable organic carbon source is necessary (this can be simple, such as glucose, or very complex, like cellulose and hemicellulose composing the plant cell wall). Many molds grow on poor substrates containing only inorganic nutrients, e.g., inorganic nitrogen or phosphorus. This explains why molds may actively grow where bacteria cannot proliferate, such as on the apparently hostile substrates of walls, wood, fabrics, etc.

Many molds produce energy by reoxidation of nicotinamide adenine dinucleotide/nicotinamide adenine dinucleotide phosphate (NAD/NADP) coenzymes along an electron transport chain where an inorganic compound serves as a terminal electron acceptor (respiration); usually oxygen is the final acceptor, so molds need an aerobic environment for their growth, but many common strains (e.g., aspergilli) use anaerobic respiration with nitrogen ions as electron acceptors. *Byssochlamys* and a very few other molds may break down sugars by fermentation, in the complete absence of oxygen, initially producing carbon dioxide, which can cause visible swelling of a package. Growth of proteolytic and lipolytic strains of molds is responsible for both the ripening and distinctive flavor of the internally or surface mold-ripened cheese and off-flavors and off-odors resulting from fungal lipase activity, with an increase in the free fatty acid (FFA) values.

Spoilage refers to any change that occurs in food resulting in reduction in sensory quality, i.e., changes in color, flavor, texture, or aroma brought about by the growth of microorganisms in food and feed. Degradative enzymes, such as pectinases and cellulases, are primarily responsible for degradation of the main structural components of vegetables, fruits, and legumes and related transformed commodities, pectin, and cellulose.

Flavor-volatile metabolites produced by molds, such as 3-methyl-butanol, 3-octanone, 3-octanol, and many others, cause adverse changes in flavor and aroma.

Besides the primary metabolic pathways supplying the fungal cells with energy and cellular metabolites necessary for growth, molds may show a range of unusual metabolic processes (grouped as secondary metabolism). These are not concerned with survival and growth and take place when normal development is restricted or the substrate composition is unbalanced (e.g., a very high carbon to nitrogen ratio, as in the case of plant tissues). Some secondary metabolites are of great commercial value (e.g., antibiotics like penicillins) but others, such as mycotoxins, pose a health risk to the consumer in crops and other plant products. (*See* **Mycotoxins**: Occurrence and Determination.) Mold spoilage is so undesirable not solely because of the resulting unesthetic changes, but also because it may constitute a hazard to human and animal health.

Growth Requirements of Molds and Effects of Processing

The metabolism of molds, like that of other organisms, is affected by environmental conditions which may increase or decrease the rate of enzymatic reactions. Physical and chemical conditions which play a major role in controlling the mold's metabolism and growth are temperature, water availability, and ion content.

The nutritional, physical, and chemical attributes of a given food substrate may select molds for predominant growth, these being the most highly adapted organisms, e.g., to reduced water content in high ionic strength (high acidity). Moreover, organisms can survive to some degree in conditions that do not allow growth. Spores and, to a lesser extent, conidia may tolerate extreme conditions in food and environment, retaining for extended times their ability to develop a mold when the substrate becomes favorable again. Mold propagula held on a frozen food may later undergo germination when mixed at room temperature with other ingredients. Although usually repressed by the high osmotic pressure (low a_w) in sugar-preserved or dried commodities, such as in preserves in tanks for industrial use or in pasta in its packaging, spores may give rise to a visible spoilage when water condenses at the mass surface.

Most fungi are mesophilic, so they grow at moderate temperatures, in the range 10–40 °C. Lowering the temperature decreases the rate of metabolism and is often used to slow deterioration. A few fungi grow at very low temperature. *Cladosporium* and *Thamnidium*, sometimes colonizing animal hairs, are important as spoilage agents of improperly dressed meat that is held for long enough (usually more than a month) at subzero temperatures above − 10 °C.

Important intrinsic factors which influence the microflora that develops in food and feed include pH and a_w.

Fungi show a wide range of behavior with respect to water availability, generally being more tolerant of water stress than other organisms, including the majority of yeasts. The lower limit for growth of molds, expressed as a_w, vary from 0.93 for destructive rots of fruits and vegetables like *Rhizopus* and *Mucor*; to 0.87 for toxigenic fusaria; to 0.83 for agents of surface discoloration in fresh cheese and sausage like *Penicillum expansum* or *P. viridicatum*. Xerophilic fungi are characterized as being capable of growth below an a_w of 0.85 and are most commonly associated with intermediate-moisture food such as dried meats, seafood, fruit concentrates, jellies, jams, preserves, and syrups, spices, nuts, and cereals, where their effective growth range can be as low as 0.65. Conditions of reduced a_w lower the final fungal biomass produced because of reduction of the growth rate and of increase of the time required for spore germination up to 3 months or more. Properly dried and stored commodities are inherently resistant to spoilage due to their low a_w; however, temperature differentials, such as in bulk-stored grains, may lead to moisture condensation and possible growth of xerophilic molds. The majority of xerotolerant fungi belong to genera *Penicillium* and *Aspergillus*, including its perfect forms such as *Eurotium* (*Aspergillus glaucus*), *Emericella*, and *Wallemia sebi*. (*See* **Water Activity**: Effect on Food Stability.)

Media and diluents with reduced a_w are necessary for enumerating xerophilic fungi, otherwise vegetative cells and spores may be lysed by osmotic shock or remain dormant when exposed to the relatively high a_w associated with the conventional diluents and media used in microbiological analysis. Addition of humectants, such as 50% glucose to the diluent, and 18% glycerol or 50% glucose to traditional media, especially malt extract agar, lowers a_w levels to about 0.94 and 0.92 respectively. Extended incubation periods, such as 4–6 weeks at room temperature instead of the traditional 5 days time, are required for recovery.

Molds, like yeasts, tolerate a greater pH range than bacteria do. Many fungi can grow over a wide pH range of at least 3–8, while pH values below about 5 negatively affect the metabolism and growth of bacterial cells. Where the food substrate shows a high a_w, it is the acidic environment which plays a decisive role in the competition between bacteria and filamentous fungi. As fruits show pH values of 4.6 or lower, mold growth is more likely. Spoilage thus arises from the intense activity of pectic and proteolytic fungal enzymes that have pH optima between 2 and 5. (*See* **pH – Principles and Measurement**.)

Because of the the acido-tolerant behavior of molds, originally food mycologists relied on low-pH media, such as acidified potato dextrose, Sabouraud, or similar agars to detect fungi in clinical and food specimens. However, in food samples acid pH may adversely affect mold spore germination and give rise to protein coagulation with the formation of artifacts, and, therefore, it is more effective to use media of neutral pH in combination with antibiotics. Antibiotics, such as chloramphenicol, gentamicin, and others, are well-known biomolecules perturbing the cell wall of bacteria which is different from that of fungi.

Molds may alter the pH of the environment in which they grow, by producing ammonium ions that neutralize the substrate acidity and thus allowing

(re)growth of normally repressed, dangerous organisms. Such an example of metabiotic association is given by the persistence of *Listeria* in cheese ripened by selected molds. (*See* ***Listeria***: Properties and Occurrence.)

Molds are aerobic organisms in that they require oxygen and/or a food substrate with a high oxidation-reduction potential (E_h) for growth. Oxidized foods, such as fruits with E_h values betwee0n $+300$ and $+400$ mV, support the growth of molds well, while reduced ones, like yogurt with its E_h value of about -100 mV, do not. In the latter case, molds may develop a mycelium only at the surface in contact with the air. Some fungi (e.g., strains of *Paecylomyces* and *Penicillium*) are microaerophilic, growing in an atmosphere containing only 0.1–0.5% oxygen. A few species, such as *Fusarium oxysporum*, may show a very delayed growth in the complete absence of oxygen. It would appear, therefore, that the degree of anaerobiosis attained under commercial processing conditions is not always adequate to prevent completely the growth of molds.

There is evidence that raising the level of carbon dioxide in the atmosphere above 70–80% greatly reduces the growth rate of most molds, with some notable exceptions, such as *Penicillium roquefortii* or *Byssochlamyis nivea*.

Growth of most molds is restricted – sometimes inhibited – by the use of preservatives at the legally permitted concentrations (between 0.1 and 0.2%) The acids and salts commonly used are propionic, sorbic, and benzoic. Propionates are added to bakery products and to certain processed cheeses. Benzoic acid and its sodium salt are most widely used in highly acidic foods such as fruit products and carbonated beverages. The undissociated acid is the active molecule and therefore the effectiveness increases with decrease in pH while the organic acids exhibit little fungistatic activity in neutral conditions. Inhibition of growth by weak acid preservatives is due to a number of actions, including membrane disruption, inhibition of essential metabolic reactions, stress on intracellular pH homeostasis, and accumulation of toxic anions. Natamycin (pymaricin), a natural antimycotic compound produced by *Streptomyces natalensis*, is commercially available for the surface treatment of cheese. In recent years, strains of preservative-resistant molds (particularly aspergilli) have been reported which tolerate high (0.5–1.5%) concentrations of organic acids due the induction of an active transport system that pumps undissociated acid molecules out of cells. The enzymatic degradation of sorbic acid to pentadien by certain fungal species is well documented. (*See* **Preservation of Food**.)

Most molds possess little heat resistance, and are easily destroyed when food with a_w values above 0.93 are heated at pasteurization temperatures. Vegetative cells and asexual propagules behave in the same manner. Heat resistance of molds is enhanced in the presence of high levels of solutes such as salt and sugar. The decimal reduction time for *Aspergillus niger* conidia heated at 100 °C increases from about 1 ms in a moist substrate ($a_w = 0.98$) to about 1 s in intermediate-moisture substrates (a_w between 0.85 and 0.90, such as cakes and most oven-processed sweet products), and to 10 s when a_w falls below 0.85 because of the addition of sugar (e.g., in fruit preserves). In the latter products, the low pH, usually between 3.2 and 3.5, partially counteracts the protective effect of sugar, thus decreasing the resistance of mold to heat.

The responses of molds to the presence of sodium chloride (salt) is somewhat different from that observed with sucrose. Protection against heat inactivation of molds is increased even at lower a_w values, usually below 0.97, i.e., a salt concentration of about 3%. This should be kept in mind when setting up pasteurization procedures for brines used in cheese and sausage manufacture.

A few molds produce heat-resistant ascospores or sclerotia. Ascomycetous species are widely distributed in soil and cause spoilage in heat-processed food products containing fruits which have been extensively contaminated with soil. Ascospores produced by the two species of *Byssochlamys, B. fulva* and *B. nivea*, and by *Neosartorya fischeri* are the most resistant of all and they are typically associated with spoilage of thermally processed fruit products, mainly canned, where the can center temperature is raised up to 80–90 °C. The decimal reduction time of *Byssochlamys nivea* spores at 80 °C increases from about 35 min ($a_w = 0.95$) to about 4 h in neutral foods with a_w values between 0.85 and 0.86. *Neosartorya fischeri* var *fischeri* has a D_{85} (decimal reduction time at 85 °C) value of 6–10 min with a Z value (Thermal Death Time curve's slope) of 5–7 °C while the *spinosa* variety is more heat-resistant (D_{85} value of 10–96 min with a Z value of 5–7 °C). Symptoms observed with spoiled canned fruits include breakdown of fruit texture, off-odors, gas, and visible mold growth. (*See* **Heat Treatment**: Chemical and Microbiological Changes; **Pasteurization**: Principles.)

Classification of Molds

Only a small proportion of the thousands of genera and species of molds identified in nature are of any significance in foods.

Ranked taxonomically in the separate kingdom of Fungi, molds are subdivided into further taxonomic categories on the basis of ecological behavior, and gross colony and microscopic morphology.

The most important diagnostic factor relies on the morphology of the asexual reproduction system (the so-called anamorphic stage of the mold), best observed under the microscope in slide-mounted cultures. Spores formed through sexual reproduction (the so-called telemorphic stage) are rarely used in identifying foodborne molds, given the difficulty of inducing gametogenesis and sexual fusion *in vitro*.

Molecular tools based on the polymerase chain reaction (PCR) may be applied in food mycology for diagnostic purposes. The noncoding ITS (Internal Transcribed Spacer) region of the rDNA unit, and other genes coding for metabolic and structural functions (e.g., the β-tubulin gene), have been investigated to clarify the genetic structure of some species.

Three subkingdoms include genera of significance in food spoilage: Zygomycotina, Ascomycotina, and Deuteromycotina. Most fungi, within the subkingdom Zygomycotina, belong to the class Zygomycetes, which includes, in the order Mucorales, three phytopathogenic genera frequently involved in the spoilage of moist wheat, flour, and cereal products: *Absidia, Mucor*, and *Rhizopus*. Distinctive features of these molds are (1) rapid growth when water availability is above the limiting value of 0.92–0.93, including filling the whole Petri dish with fine white to greyish to blackish aerial mycelium, and (2) asexual spore formation by cytoplasmic cleavage within a large globose cell termed sporangium, from where sporangiospores are liberated by breakage or deliquescence of the wall. Sporangia are located at the end of a supporting hypha (sporangiophore), from where they are separated by a septum forming a protrusion called columella (**Figure 1**).

A distinctive feature of the subkingdom Ascomycotina is the development of sexual spores in a cell (sporangium), more specifically termed the ascus (bag). Asexual spores, called conidia, develop in various ways directly from reproductive hyphae. Ascospores of *Byssochlamys* species show a very distinctive heat resistance.

The subkingdom Deuteromycotina is used mainly for convenience to accommodate all those fungi that seldom, if ever, produce sexual stages. In all other respects, however, these fungi are similar to the Ascomycotina, and where sexual stages are discovered they are usually found to belong to the Ascomycotina. In Deuteromycotina systematics the teleomorph (perfect, sexual stage) name should preferentially be used to indicate the species. Unfortunately, the anamorphic name is firmly rooted in the literature and in common parlance. The teleomorphic stage cannot usually be seen in routine examination of fungi unless special cumbersome techniques are used, so the anamorph name is commonly retained: *Aspergillus*, instead of its teleomorphs *Eurotium* and *Emericella*; *Penicillium*, not the teleomorphs *Eupenicillium* and *Talaromyces*, and so on.

The morphology of the fertile hypha is again the most important factor in taxonomic characterization. During development of the mold, asexual propagation occurs which starts with conidiogenesis: asexual propagula, named conidia, are produced from the mother cell (conidiogenous cell). The system composed of a conidiogenous cell and its conidia is known as a conidiophore. Conidiogenous cells are sometimes supported by mono- or pluricellular stipes whose shape may help in mold identification. Differentiation of the conidiophore and conidial arrangements serve as diagnostic features, and we may distinguish the following:

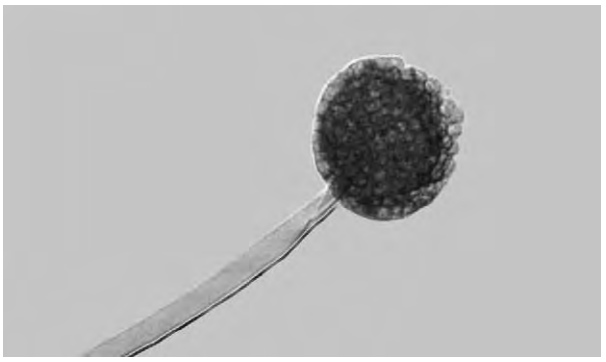

(a)

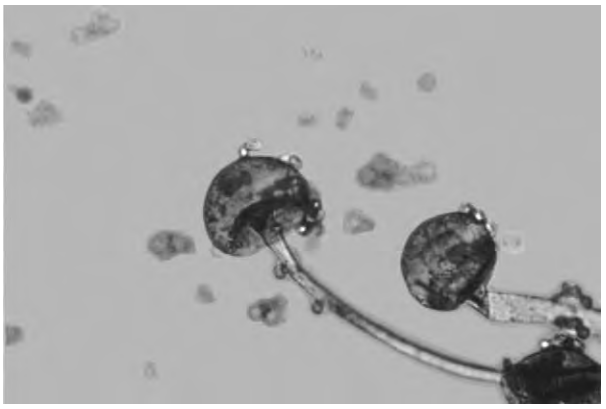

(b)

Figure 1 (a) Large, round sporangium of *Rhizopus oryzae*. (b) After disintegration of the sporangial wall, small clusters of sporangiospores are released, leaving a naked umbrella-shaped columella. Reproduced from Spoilage, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds) 1993, Academic Press.

1. Arthroconidia: usually square-shaped conidia originating directly by fragmentation of a hypha (e.g., *Geotrichum candidum*, or 'machinery mold' because of its characteristic enrichment in samples from improperly sanitized machinery).
2. Blastoconidia: unbranched chains with dark hila (basal scars) that arise directly from a distinct conidiophore; those of *Cladosporium* are small and typically apiculate (lemon-shaped).
3. Poroconidia: dark-walled (dematiaceous) conidia in chains, each of them developing through a pore in the apex of the preceding conidium. The youngest conidia (smaller) are at the tip of the chain, while the most mature conidia are closer to the conidiogenous cell. The claviform shape, large dimensions, and both longitudinal and transversal septa give a typical appearance to the conidia of *Alternaria* (Figure 2).
4. Phialoconidia: produced in a long chain from an opening of conidiogenous cell called a phialide because of the distinctive amphora (flask) shape.

Both *Penicillium* and *Aspergillus* produce phialoconidia in structures which look different under the microscope because of (1) the presence in the former mold of the typical 'penicillus' (brush), a conidiophore with several terminal branches each bearing a phialide (Figure 3), while (2) in the latter the conidiophore is formed by an unbranched stipe, which enlarges at its end into a vesicle from which phialides arise with their chains of very small conidia (Figure 4). The sickle-shaped, septate phialoconidia of *Fusarium*, a powerful mycotoxigenic mold, are called macroconidia (Figure 5), in contrast to the smaller and usually nonseptate chains of microconidia.

Figure 2 *Alternaria tenuis*. Chains of large brown clavate conidia with transverse septa, blowing out from a conidiophore. Reproduced from Spoilage, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds) 1993, Academic Press.

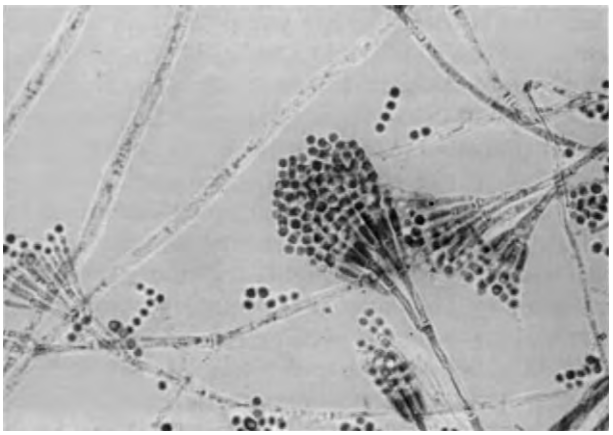

Figure 3 *Penicillium*. Branched conidiophore bearing conidial chains, in the form of a brush ('penicillus'). Reproduced from Spoilage, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds) 1993, Academic Press.

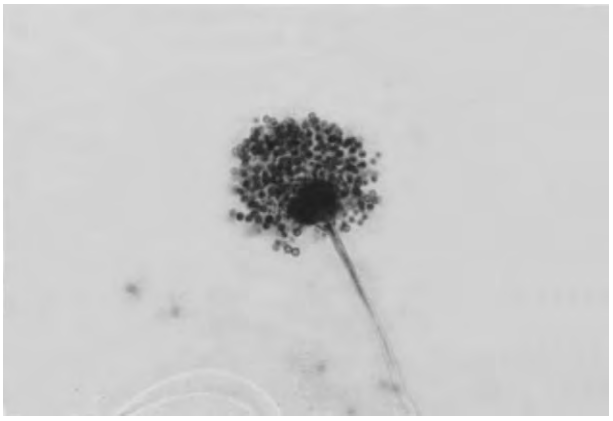

Figure 4 Conidiophore of *Aspergillus* enlarged in a round vesicle covered with a mass of round conidia. Reproduced from Spoilage, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds) 1993, Academic Press.

Contamination and spoilage of food

Factors contributing to the wide distribution of molds in the natural environment are their capacity to utilize almost all carbonaceous substrates, their ability to develop in the presence of very low concentrations of nutrients, and their tolerance to many adverse conditions. Fungal spores are present in dust and dead materials where, because of their ability to utilize cellulose and lignin, they may grow extensively during those materials' decomposition.

During cultivation, grains, and to a lesser extent vegetables and fruits, are exposed to fungal spore contamination from many sources, such as dust,

Figure 5 Sickle-shaped macroconidia of *Fusarium* with three transverse septa. Reproduced from spoilage, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds) 1993, Academic Press.

Table 1 Risk relationships for moisture content against deterioration

Grain	*Minimum moisture level to prevent spoilage (20–30 °C)*
Barley (*Hordeum sativum*)	15–14.5
Corn (*Zea mays*)	14.5–13.5
Cotton (*Gossypium barbadense*)	8.5–8
Rice (*Oryza sativa*)	13.5–13
Soybean (*Glycine soja*)	12–11.5
Wheat (*Triticum vulgare*)	14.5–13.5

Table 2 Overview of molds occurring in field and stored grain

Mold genus/species	*Minimum moisture level for growth*	*Minimum water activity level for growth*
Field molds		
Fusarium spp.	20–22%	0.87–0.89
Fusarium moniliforme		0.87
Fusarium graminearum		0.89
Alternaria alternata	18–20%	0.85–0.87
Cladosporium spp.	18–20%	0.85–0.87
Storage molds		
Penicillium spp.	16–18%	0.80–0.85
Penicillium viridicatum		0.81
Penicillium griseofulvum		0.83
Penicillium citrinum		0.82
Others		
Aspergillus spp.	14–20%	0.72–0.87
Aspergillus flavus		0.85
Aspergillus ochraceus		0.80
Aspergillus glaucus		0.74

insects, and rainfall, so that it is not surprising to find propagule counts on their surface between 10 and 10^2 colony-forming units (CFU) per square centimeter, i.e., between 100 and 10^5 CFU g^{-1}. Invasion of internal tissues is usually prevented by the integrity of the outer skin. During storage, control of relative humidity and temperature and the use of modified gaseous atmospheres (high levels of carbon dioxide) retard germination and growth rate of molds, thus preventing spoilage. Moisture content–deterioration risk relationships for given kinds of grain are shown in Table 1.

Handling and processing may either give mold propagules the chance to invade the body of injured fruits or vegetables or remove them when the grain pericarp is separated from endosperm during milling. Mold counts in flour may therefore be reduced to 10–10^3 CFU g^{-1}.

'Field fungi' include molds that grow on grains in the field, where the a_w is in the range 0.95–1.00, while 'storage fungi' can grow at the lower a_w levels found in grains during storage (0.70–0.90). Most species of the genus *Aspergillus* grow well when the moisture content is in the range of growing grains (i.e., > 18%, or $a_w > 0.85$) as well as when the moisture is that typical of stored grains (i.e., < 18% or $a_w < 0.85$). An overview of field and stored grain molds is shown in Table 2. (*See* **Cereals: Handling of Grain for Storage**; **Controlled-atmosphere Storage: Effects on Fruit and Vegetables.**)

Other dried agricultural commodities, such as spices, may retain their original fungal bioload (i.e., counts between 100 and 10^4 CFU g^{-1}) if sufficient attention is given to protection from dust and water condensation during processing. Unfortunately, most important producers are located in the Third and Fourth World areas, where conditions of harvesting and processing are sometimes primitive, so that mold counts may reach levels as high as 10^5–10^6 CFU g^{-1}, posing a risk of spoilage in foods with which spices are mixed.

Well-slaughtered and dressed meat is contaminated with only a few mold spores, usually less than 10 CFU cm^{-2}, which increase when skin and hide are not completely removed. If the surface of meat becomes dry, because of too long storage in the chiller, yeasts and molds (e.g., *Cladosporium cladosporioides*, agent of the 'black spot') may develop and give visible spoilage. (*See* **Meat: Preservation.**)

Molds are sometimes isolated from egg shells. Their presence is mainly as a result of dust and moisture in the hatchery and the cold store, but their growth is prevented by the low E_h value of the egg.

Molds entering raw milk from soil, bedding, feed residues, manure, and so forth, during milking are so greatly outnumbered by bacteria, and so effectively inhibited by natural antimicrobial substances such as lactoperoxidase and lactoferritin, that their number

never exceeds a few tens of spores per liter. (*See* **Milk**: Processing of Liquid Milk.)

Molds pose a real problem to the food industry mainly as secondary invaders of heat-treated (not aseptically packed) products. These treatments are usually sufficient to inactivate all but the most heat-resistant microorganisms. Airborne mold spores may contaminate products after heating (e.g., baking) and, when chemical and physical properties of the substrate are unsuitable to bacteria, may invade the food and become the main agent of spoilage. The main sources of airborne contamination are (1) crops and trees in close proximity of the factory; (2) areas where dusty raw materials, such as flour, cardboard packages, and woody materials such as pallets are stored; and (3) aerosol-generating operations, such as in waste treatment plants. Molds carried by air movements and/or by the clothes and shoes of people coming from these areas may penetrate processing rooms, where they fall directly on to the product before packaging, and are also deposited over the floor and wall surface. In this case the spores may later recirculate and contaminate the product.

When large volumes of air are needed, e.g., for dehydration or forced cooling of foods, the number of airborne organisms needs to be reduced by suitable filtration. High-efficiency nonabsolute filters, which remove 90–99% of the 1-μm or larger particles, are the minimum requirement for drying and cooling operations. Ultrahigh-efficiency (removing at least 99–99.9% of all particles of 0.1–0.2 μm or larger) or absolute filters (efficiency > 99%), even in the form of hoods or ceilings, are suggested to insure trouble-free production in aseptic or semiaseptic (such as yogurt) packaging lines.

To prevent condensation and subsequent mold growth on surfaces, good ventilation is required, which removes the excess moisture released during processing and storage of certain products, e.g., during cooking, salting, aging, and other common operations.

Possible fallout of airborne mold spores on to the surface of exposed products during high-risk operations such as cooling after baking or during packaging may be prevented by a careful separation between these environments and sources of molds such as storage of raw commodities and packaging materials.

The microbiological condition of packaging materials may directly contribute to the fungal contamination of packed foods, owing to the presence of spores on the material. The high temperatures used in the fabrication of plastic articles cause most of them to be virtually sterile (usually less than 1 mold spore per 100 cm^2). The maintenance of hygiene during handling and storage is therefore important. Materials and containers should be stored in clean and dust-free warehouses, well sealed in their original bags, until the moment of use. At that time, packaging materials should be taken out of their wrapper before entering the processing room, in order to limit the entry of dust and associated fungal spores.

Unless a food has been previously heated and subsequently aseptically placed in a hermetically sealed container, the presence of mold spores arising from raw materials or processing environments has to be considered as 'normal'. Fortunately, the growth of mold colonies on incubated Petri dish does not necessarily imply their development in the food substrate, because food is usually protected through suitable preservation techniques that prevent or delay the mold spores' germination and growth. Cooling and lowering the E_h value of cheese and sausages and drying and storing dry wheat and flour are examples of appropriate means for avoiding overt spoilage by molds.

None the less, molds are involved in the spoilage of both fresh, perishable foods and stored or processed foods. They are the main cause of market disease of fresh fruits and vegetables. Some of the spoilage organisms are true plant pathogens in that they have the ability to invade healthy, viable plant tissue. An example is *Phytophthora infestans*, the causal organism of late blight of the potato. Other fungi are saprophytes in that their development is restricted to dead or damaged plant tissue. Species of penicillia and aspergilli fall into this category. Molds produce enzymes which digest the main fruit and vegetable skin components, i.e., pectinases and esterases for the pectic substance and glucanases and glycosidases (cellulases) for the soluble cellulose. Growth of fungi on fruits and vegetables therefore results in tissue disintegration, i.e., rotting.

The ubiquity of molds in the environments where foods are handled, stored, and processed is such that these organisms will always be found in, and on, items which have not been heat-treated in a sealed container or aseptically packed after heating. Notwithstanding this, molds are involved in the spoilage of foods only under special circumstances. They dominate in spoiled food whose intrinsic properties are suited for their physiology, i.e., moist acidic foods rich in carbohydrates (fruits), those having a tendency to dry out during storage (vegetables), or with added salt (cheese, sausages), and dry foods rich in carbohydrates, such as cereals and flour. Development of grain spoilage molds is usually limited to high-moisture, temperature-abused stored cereals. The water content of flour is usually just below the critical level for mold growth (i.e., under 16–17%), hence flour carries fungal propagules to cereal products rather than undergoing spoilage. When growth is present flour gives a very pungent musty smell.

Infection of the surface of bakery products, particularly bread and cakes, may arise from dispersal of fungal propagules by air when products are cooled after baking. Enough water and a long storage time are required for fungal growth, which appears in the form of colored spots. French-style high-moisture breads, although wrapped and treated with preservatives, spoil very rapidly (i.e., within 2–7 days). Cakes have a longer mold-free life, depending on the individual a_w, processing technology, sanitation, and selection of ingredients.

Herbs and spices often contain large numbers of fungal spores and, even if the proportion of spices in recipes is usually low, under some conditions they contribute to an overt spoilage of the final product. The most important factors in determining the quality of dried herbs and spices are the initial bioload of the source crop and subsequent storage conditions. The desired properties of many spices are lost if storage is too extended or under poor conditions. (*See* **Herbs**: Herbs and Their Uses.)

Under special circumstances, molds may spoil even protein-rich foods, e.g., fresh meat whose surface dries following prolonged cool storage, and long-term ripened cheese whose rind becomes too dry to support the growth of competing bacteria. Penetration of mold mycelium through cracks in the rind and spreading of fungal enzymes across the rind may result in localized off-colored and off-flavored areas in the body of the cheese. Growth of spoilage molds is a problem on all cheeses but it is particularly rapid on soft cheese where cutting for prepacking is involved – even more so where untreated ingredients, pepper, nuts, herbs, and mushrooms are used.

Conversion of an oil-in-water emulsion, i.e., cream, to a water-in-oil emulsion, i.e., butter, makes butter a substrate with such a low a_w that surface molding is seldom a major problem. When present it is manifested by discolored areas, appearing more frequently on the surface of items manufactured from unpasteurized cream (farm-related contamination) or under unsanitary processing conditions (process-related contamination). (*See* **Butter**: The Product and its Manufacture.)

Both fresh and fermented sausages represent an ecological niche where inherent (pH, a_w) and environmental (temperature and moisture in the cellar) factors promote mold growth. Traditionally some salami-type sausages in Europe are mold-ripened. The organisms involved, mainly belonging to the genus *Penicillium*, are chosen on the basis of their ability to develop pleasant flavors during growth with no adverse side-reactions such as the production of mycotoxins. (*See* **Meat**: Sausages and Comminuted Products.)

See also: **Butter**: The Product and its Manufacture; **Cereals**: Handling of Grain for Storage; **Controlled-atmosphere Storage**: Effects on Fruit and Vegetables; **Heat Treatment**: Chemical and Microbiological Changes; **Herbs**: Herbs and Their Uses; ***Listeria***: Properties and Occurrence; **Meat**: Preservation; Sausages and Comminuted Products; **Milk**: Processing of Liquid Milk; **Mycotoxins**: Occurrence and Determination; **Pasteurization**: Principles; **pH – Principles and Measurement**; **Preservation of Food**; **Water Activity**: Effect on Food Stability

Further Reading

Doyle MP, Beuchat LR and Montville TJ (eds) (1997) *Food Microbiology. Fundamentals and Frontiers*. Washington, DC: ASM Press.

Gams W, van der Aa HA, van der Plaats-Niterink AJ, Samson RA and Staplers JA (1987) *CBS Course of Mycology*. Baarn: Centraalbureau voor Schimmelcultures.

Moss MO, Jarvis B and Skinner FA (eds) (1989) *Filamentous Fungi in Foods and Feeds*. Society for Applied Bacteriology Symposium Series no. 18. Oxford: Blackwell Scientific.

Ottaviani F and Disegna L (1987) Moulds and yeasts in dairy products and environment. In: Cominazzini C (ed.) *Microbial Recontamination in Dairy Processing*, pp. 22–54. Thiene: Food and Dairy Institute of Thiene.

Pedersen LH, Skouboe P, Boysen M, Soule J and Rossen L (1997) Detection of *Penicillium* species in complex food samples using the polymerase chain reaction. *International Journal of Food Microbiology* 35: 169–177.

Pitt JI and Hocking AD (1985) *Fungi and Food Spoilage*. North Ryde: Academic Press.

Samson RA, Hocking AD, Pitt JI and King AD (eds) (1992) *Modern Methods in Food Mycology*, Developments in Food Science 31. Amsterdam: Elsevier Science.

Samson RA, Hoekstra ES, Frisvard JC and Filtenborg O (1995) *Introduction to Food-borne Fungi*. Barn: Centraalbureau voor Schimmelcultures.

Yeasts in Spoilage

V Loureiro and M Malfeito-Ferreira, Universidade Técnica de Lisboa, Tapada da Ajuda, Portugal

Background

Yeasts are microorganisms generally perceived as favorable components in food systems, since their roles are well known in the production of foods and fermented beverages – bread, beer, wine, cider, sake,

distilled spirits, etc. More recently, this image has been strengthened by the fact that yeasts can be obtained in capsules, as sources of B-complex vitamins, and are medically prescribed for several conditions. However, yeasts are also responsible for undesirable effects such as food spoilage. This article summarizes the main aspects of the occurrence, effects, and monitoring of yeasts in the spoilage of foods and beverages.

Competition among Bacteria, Molds, and Yeasts

The contaminant microflora of food is constituted, in most cases, of mixed populations of bacteria, molds, and yeasts. Microbial alteration becomes a competitive process among the different groups of microorganisms, prevailing over those which show a better adaptation to the environmental conditions in each product.

In food commodities, under optimal growth conditions, the populations of most bacteria, owing to their faster growth, overtake those of molds and yeasts. While bacteria have typical generation times of less than 1 h and, in some cases, around 10–15 min, mold and yeast populations may take several hours (or days) to duplicate. Thus, molds or yeasts can compete with bacteria only when environmental conditions severely affect bacterial activity. Low pH and low water activity (a_w) favor the growth of molds and yeasts, as they are more acid-resistant and xerotolerant than bacteria. The presence of chemical preservatives, mainly those of an acidic nature, is also less inhibitory towards molds and yeasts.

Yeast activity can lead to subsequent bacterial activity, as in acid-fermented foods; in this case, organic acids can be consumed by film-forming yeasts, leading to deacidification and promotion of bacterial activity. The competition between molds and yeasts depends, to a large extent, on oxygen availability. While the former are obligate aerobes, the latter may be facultative, growing either under aerobic or anaerobic conditions. For this reason, yeasts are found much more frequently on liquids, where good aerobic conditions exist only at the air–liquid interface, whereas filamentous fungi predominate in solid foods.

Concept of Spoilage Yeasts

Of all yeasts isolated from nature, about 500 species are presently recognized by taxonomists. Of these, about a quarter may be readily isolated from foods, but only a few play a significant role in food alteration. Those which do not affect foods are generally called adventitious or innocent yeasts; those responsible for undesirable changes in foods and beverages are called spoilage yeasts. For food technologists, the concept of spoilage yeast generally has a stricter sense. It takes into account only whether a particular species is able to cause deterioration in foods and beverages which have been processed and packaged according to the standards of good manufacturing practice (Table 1). It should be emphasized that the concept of spoilage yeast is conditioned by the type of product and by the moment when the yeasts exert their activity. For instance, in the production of wines, the yeast *Saccharomyces cerevisiae* is

Table 1 Principal yeast species causing spoilage of foods and beverages processed according to the standards of good manufacturing practices

Type of food	*Food product*	*Yeast species*
Fresh fruits and vegetables	Strawberries, figs, tomatoes	*Hanseniaspora*/*Kloeckera* spp.
Refrigerated and frozen foods	Icecreams, frozen peas	*Rhodotorula* spp.
Pasteurized foods	Yogurts, fruit juices, ketchup	*Kluyveromyces marxianus*, *Saccharomyces cerevisiae*, *Zygosaccharomyces bailii*
Preserved foods	Mayonnaise, salad dressings, sauces, soft drinks, chutneys	*Brettanomyces*/*Dekkera* spp., *Issatchenkia orientalis*, *Pichia membranifaciens*, *S. cerevisiae*, *Saccharomycodes ludwigii*, *Schizosaccharomyces pombe*, *Z. bailii*
Fermented foods	Pickles, olive brines, sauerkraut, cheeses, sausages	*Debaryomyces hansenii*, *I. orientalis*, *P. membranifaciens*, *Yarrowia lipolytica*, *Z. bailii*
Alcoholic beverages	Wine, beer, cider	*Brettanomyces*/*Dekkera* spp., *Pichia anomala*, *P. membranifaciens*, *S. cerevisiae*, *S. cerevisiae* var. *diastaticus*, *Sac. ludwigii*, *Sc. pombe*, *Torulaspora delbrueckii*, *Z. bailii*
Concentrated products	Dry fruits, jams, jellies, honey, fruit concentrates, filled chocolates, marzipan	*Candida versatilis*, *D. hansenii*, *P. anomala*, *T. delbrueckii*, *S. cerevisiae*, *Sc. pombe*, *Z. bailii*, *Z. rouxii*, *Z. bisporus*

recommended for its good fermentation characteristics. However, later in the production process, this species is considered a dangerous spoiler in sweet bottled wines, because it is able to ferment residual sugar. (*See* **Wines**: Production of Table Wines; Production of Sparkling Wines.)

Importance of Yeasts in Food Spoilage

The presence of yeasts in food is generally believed to be harmless to public health, although occasional allergies have recently been attributed to them. The absence of pathogenic foodborne yeasts diminished interest compared with bacteria and molds, and so, for a long time, their importance as food contaminant microorganisms has been underestimated. More recently, however, yeasts have been established as important factors in food spoilage, which involves the visible or detectable deterioration of physical and sensorial properties of foods. These undesirable activities are responsible for serious economical losses, which remain mostly unreported due to the confidentiality assumed by the companies involved (e.g., manufacturers, suppliers of raw materials and packaging, retailers).

Among the factors that have contributed to this increasing importance are the use of modern technologies in food processing, the great variety of new formulations of foods and beverages, and the tendency to reduce the use of preservatives, particularly those effective against yeasts (e.g., sulfur dioxide and benzoic acid). Modern technologies tend to utilize less severe processing conditions so as to preserve as much as possible their flavors, tastes, and natural colors. Thus, for instance, in the thermal processing of products of vegetable origin, it is becoming more commonplace to use cold filling, after a gentle pasteurization treatment, which significantly increases the risk of yeast contamination and, consequently, the risk of yeast spoiling. New food and beverage formulations, using juices and concentrated fruits, sugar syrups, sliced fruits, etc., such as the different types of yogurts, have also contributed significantly to the importance of yeasts as spoilage agents. In the same way, low-calorie foods, where preservation no longer depends on the effect of high sugar concentrations, providing low water activities, are good habitats for yeast activity.

Effects and Extent of Yeast Spoilage

The most common spoiling effects occur in sweet acid drinks and are characterized by abundant gas production, which may deform or explode the packages, haziness, sediment or film formation, off-flavors, and off-tastings. Deterioration of food and drinks by yeasts may present other effects, more or less evident, according to the type of food (Table 2). However, in certain cases, yeast spoilage is not so easily defined, mainly in fermented foods and drinks (e.g., wines, artisanal beers, black olives, soy sauce, cheeses), where the metabolites produced contribute to the flavor and aroma. In fact, the distinction between detrimental and beneficial activities is not always clear-cut. For instance, the 4-ethylphenol production by *Brettanomyces/Dekkera* yeasts in red wines is only regarded as spoilage when this secondary metabolite is present at levels above about 700 $\mu g\,l^{-1}$. At levels below 400 $\mu g\,l^{-1}$ it contributes favorably to the complexity of wine aroma.

The level of contamination in foods is not a clear indicator of product deterioration. For instance, in many cheeses, cured meats, and cloudy beers, yeast counts as high as 10^5–10^8 cells g^{-1} do not cause any detectable depreciation; on the contrary, they are supposed to improve aromatic complexity. The reverse

Table 2 Principal spoilage effects of food products resulting from yeast activity

Food product	*Spoilage effect*						
	Color spots/discoloration	*Gas production*[a]	*Haze/ cloudiness*	*Sediments*	*Films*	*Off-flavors/ off-tastings*	*Texture changes*
Fresh fruits and vegetables	X	X				X	X
Brined vegetables	X	X	X	X	X	X	X
Fruit juices, wine, beer		X	X	X	X	X	
Mayonnaise, salad dressings	X	X			X	X	X
Confectionery	X	X			X	X	X
Syrups, honey, fruit concentrates		X	X		X	X	
Butter, cream	X					X	
Cheeses, yogurts	X	X				X	
Sliced bread, unbaked bread dough	X					X	
Sausages, meat products	X	X				X	

[a]Gas production may lead to distortion or explosion of the container or product.

situation also occurs, when counts below 10^5 cells g^{-1} induce spoilage. This is the case of the formation of visible cell clusters, in bottled white wines, or of the development of surface color spots due to yeast colonies, in certain solid foods (e.g., skinless sausages, sliced bread, etc.).

Factors Affecting Yeast Activity

The microbiological stability of foods and beverages can be achieved by complete removal of microorganisms before packaging of the final product, as in sterile filtration of beverages without suspended solids. It can also be achieved by total destruction of contaminant microorganisms, as in heat sterilization, pascalization, or γ-irradiation, or by inhibiting metabolic activity by freezing. However, in most cases, food microbiological stability is dependent upon the inhibition (or reduction) of microbial activity by the creation of one or more adverse environmental factors, such as low water activity (dried, concentrated, and salted foods), low pH (acid foods), low temperatures (refrigerated foods), and the presence of inhibitors (preserved foods). In the case of sterile filtration, heat, pascalization, or γ-irradiation, microbial contamination is possible only when there is process failure or when there is contamination after treatment. When thermal processes, pascalization, and γ-radiation are used, heat-, high-pressure-, or radioresistant yeasts predominate as contaminant organisms. For food preservation based on microbial inhibition through the action of adverse environmental conditions, the spoilage microflora is restricted to organisms resistant to the factor(s) used in the preservation process, e.g., low water activity, organic acids, and so on. (*See* **Freezing**: Structural and Flavor (Flavour) Changes; **Irradiation of Foods**: Basic Principles; **Sterilization of Foods**.)

Water Activity (a_w)

Water activity is, probably, the environmental factor which most strongly affects living organisms, given the narrow range of values allowing growth. Like other organisms, yeasts are also markedly affected by water availability, although, together with molds, they are the least sensitive. Thus, it is not surprising that molds and yeasts are the dominant flora of low-a_w foods. (*See* **Water Activity**: Effect on Food Stability.)

For identical a_w values, different solutes have different effects on yeast growth, salts (e.g., NaCl) being generally less well tolerated than sugars (e.g., sucrose). The most resistant yeasts to low a_w, in the presence of either sugar or salt, belong to the genus *Zygosaccharomyces, Z. rouxii* being the most resistant species (with an a_w of approximately 0.65 in the presence of sugar). Other species resistant to high sugar and salt levels are *Debaryomyces hansenii, Candida etchelsii, Candida parapsilosis, Torulaspora delbrueckii, Pichia membranifaciens*, and *Pichia anomala*.

pH

Yeasts and molds are generally more resistant than bacteria to low pH values and constitute the typical flora of acid food. Acetic acid bacteria, and particularly lactic acid bacteria, can compete with yeasts in acid foods, although they are more sensitive to low pH, low a_w, and some preservatives, such as sulfur dioxide. (*See* **Lactic Acid Bacteria; pH – Principles and Measurement**.)

Yeasts usually grow in the pH range 2.0–8.0, with an optimum at 4.0–4.5. The minimum pH for yeast growth is not easy to define, owing to the strong influence of the particular acid, other inhibitors, and environmental factors. The nature of acids present in foods significantly affects yeast activity, since their inhibitory effect is dependent on the concentration of undissociated form of the acid. Among other species resistant to low pH, *Zygosaccharomyces bailii* is considered the most resistant to organic acids. This feature, associated with resistance to low a_w, makes this species one of the most feared yeasts by food technologists, particularly in the production of fruit juices, fruit concentrates, marzipan, and acid-preserved foods like mayonnaises and salad dressings.

Temperature

The influence of temperature on yeast activity should be considered in two ways: as a lethal agent, during thermal processing, and as an environmental factor, during storage. Concerning the first, the thermal resistance of yeasts is considered to be low, as yeasts are destroyed at pasteurization temperatures. The higher resistance of some species is related to the formation of ascospores, which are slightly more heat-resistant than vegetative cells. Ascospores are not, however, thermally resistant forms in the same sense as bacterial endospores. (*See* **Pasteurization**: Principles.)

The yeasts more frequently referred to as contaminants of pasteurized foods belong to the species *Kluyveromyces marxianus, S. cerevisiae*, and *Z. bailli*. Thermal resistance varies with the nature of the media, the protective effect of high sugar concentrations being well known. In concentrated foods, therefore, the pasteurization conditions need to be more severe to obtain absolute safety. The following treatments are recommended: 20 s at 85 °C for concentrated orange juices of 50 °Brix, 30 s at 90 °C for chocolate syrups of 75 °Brix.

As an environmental factor, temperature does not favor spoilage yeasts in relation to other groups of microorganisms. Food yeasts are essentially mesophiles, capable of growing in the temperature range 0–40 °C, with an optimum temperature of about 25–27 °C. Some species capable of growing at low temperatures – psychrotrophic yeasts – are frequently isolated from refrigerated products, suggesting that, in this case, temperature plays an important selective role. These yeasts are nearly always pigmented, and consequently, colored spots may develop on foods. Storage temperature has a strong effect on spoilage yeast activity, often accelerating the food-alteration process. For this reason, some foods that are very susceptible to dangerous spoilage yeasts (e.g., *Z. bailii*), such as fruit concentrates, are generally stored under refrigeration.

Oxygen Availability

The presence of oxygen always enhances yeast activity, because it allows oxidative metabolism, which is energetically more efficient, as well as enabling the utilization of a wider range of energy-rich substrates (such as ethanol and organic acids). For this reason, in all foods susceptible to yeast activity, it is necessary to exclude oxygen so as to prevent film-yeast growth and reduce the yeast metabolism.

Nutrients

In general, the range of chemical compounds found in foods is adequate to supply the nutrient demands for the majority of spoilage yeasts. Essentially all food yeasts can utilize glucose and fructose as carbon and energy sources; less frequent is the utilization of organic acids (e.g., acetic and lactic acid) and alcohols (e.g., ethanol and glycerol); the utilization of starch and lipids is rare.

Organic nitrogen is not necessary for yeast metabolism, all yeasts being able to utilize ammonia as the sole nitrogen source. They can also utilize some amino acids owing to their ability to deaminate or transminate acids. Few yeasts have proteolytic activity. Consequently, foods with a low carbon-to-nitrogen ratio tend to be spoiled by bacteria, while those with a high carbon-to-nitrogen ratio are preferentially affected by yeasts. (*See* **Amino Acids**: Metabolism.)

Some yeasts grow very slowly, or do not grow at all, in certain foods as a result of difficulties in synthesizing one or more vitamins. *Z. bailii*, for example, requires group B vitamins, and *Brettanomyces/Dekkera* require biotin and thiamin (vitamin B_1). In addition, the inability of *Z. bailii* to utilize sucrose as an energy source is of some significance. Extreme care must be taken when sucrose syrups are substituted by fructose syrups.

Preservatives and Antimicrobials

Most preservatives used in the food industry are weak organic acids – sorbic, benzoic, acetic, sulfurous, and propionic acids. Their activity increases with reduced pH, ceasing to be effective at values much higher than their pK_a, i.e., for pH values where the undissociated form is insignificant. The acid resistance of yeasts is the major factor in their predominance over bacteria in the alteration of preserved food. Certain species, such as *Z. bailii*, may grow on foods with levels of sorbic acid (*c.* 800 mg/kg) much higher than those permitted by law. Other spoilage yeast species known for their high resistance to preservatives are *Saccharomycodes ludwigii, Schizosaccharomyces pombe, Issatchenkia orientalis, P. membranifaciens*, and *S. cerevisiae*. The contact of such yeasts with preservatives contributes to an adaptive process which can significantly increase their resistance and subsequent growth. (*See* **Preservation of Food**.)

The presence of antimicrobials in foods also strongly affects the contaminating flora. Among other compounds, ethanol exerts a strong selective effect, particularly on molds and most bacteria. Lactic and acetic bacteria are the most resistant to ethanol. The former may resist up to 20% (v/v) ethanol and the latter to about 14% (v/v). Thus, bacteria frequently compete with yeasts in the alteration process of alcoholic beverages. The yeasts' behavior in the presence of ethanol is quite variable. There are several very sensitive species such as *Hanseniaspora uvarum* and *Rhodotorula glutinis* and several quite resistant species, tolerating up to 18–20% (v/v) ethanol (e.g., some strains of *S. cerevisiae* and *Z. bailii*).

Prevention and Monitoring of Yeast Spoilage

When a food commodity is spoiled by yeasts, the yeasts are present as contaminants and environmental conditions make them stronger and quicker than other microorganisms. The prevention of spoilage, with fewer chemicals and milder preservation processes, requires a sound understanding of the problem, and so the food microbiologist should have available tools to evaluate the whole microbiological ecosystem. In a code of good manufacturing and distribution practices, several conditions should be met to avoid or to control the activity of yeast spoilers: (1) good-quality raw materials, i.e., with low yeast counts; (2) adequate and efficient hygiene;

(3) adequate and efficient processing operations; and (4) careful monitoring of spoilage yeasts.

Good-quality Raw Materials

The microbiological quality of raw materials is essential in the prevention and control of spoilage yeasts because higher initial contaminations increase the risk of failure of technological control processes. As a rule, natural raw materials (e.g., fruits and sugar cane) have an adventitious microbial flora which is not active in foods, after processing, being easily controlled through good manufacturing practices. However, when raw materials have already been processed (e.g., fruit concentrates, fruit juices, and sugar syrups), as in many foods of complex formulation (e.g., dessert yogurts, reconstituted fruit juices, and filled chocolates), the type of contaminant yeast is much more dangerous. This situation results from the enrichment of the contaminating flora in yeasts highly resistant to the environmental stress present during the processing of natural raw materials.

Processing Operations

One of the most efficient measures for controlling spoilage yeast activity is thermal processing of packaged foods and beverages. Unfortunately, there are few products where this is possible without any adverse effects on the quality of the food commodity. Among other measures that might be taken to prevent yeast spoilage, the following should be noted:

- preservatives should be added immediately before packaging so as to avoid contact with yeasts, which might lead to adaptation and increased preservative resistance;
- oxygen should be excluded through air evacuation or purging with inert gases before packaging to inhibit yeast metabolism;
- cross-contaminations should be avoided in plants processing more than one kind of product (e.g., jams and fruit juices);
- food products should be stored at low temperatures, particularly for foods made with ingredient mixtures with a high susceptibility to yeast activity.

Hygiene

It is generally accepted that the main source of contamination by yeasts is food residues remaining in processing equipment after inadequate cleaning and disinfection. It is essential, therefore, to ensure high standards of hygiene, for which several conditions should be met:

- suitable plant layout and easily accessible equipment;
- awareness of the most likely sites for spoilage yeasts (e.g., filler heads, diaphragm valves, pressure gauges, dead ends of sterilizing filters, pumps, stirrer bearings, and glands);
- cleaning and disinfection programs should include dismantling of pieces of equipment that are difficult to sterilize, and adequate contact time and concentrations of sterilants must be used. (*See* **Sanitization**.)

Monitoring of Spoilage Yeasts

Microbiological control Generally, food microbiological control does not provide satisfactory information for an effective prevention against spoilage yeasts. The main reasons for this are:

- there is a lack of specificity of culture media used for yeast identification, which does not allow a distinction between adventitious and spoilage species (often, the same media are used for yeast and mold counts);
- sampling procedures are usually statistically inadequate owing to time and cost restrictions;
- sampling size is often insufficient to detect low contamination levels;
- there is a lack of reference values needed for adequate interpretation of yeast counts.

Accordingly, it should be noted that for yeasts such as those from the genus *Zygosaccharomyces*, which are potentially dangerous even in very low numbers, the microbiological control information is often rather inadequate.

Microbiological control allows determination of the critical points of processing lines and estimation of their relative importance as origins of outbreak of spoilage yeasts. These critical points are dependent on the quality of the equipment. For example, in wine-bottling lines with poorly designed filling machines, the main critical points of yeast contamination are as shown in Figure 1.

Besides yeast detection and counting, the microbiological control may be further aimed to the typing or identification of contaminant yeasts. Classical identification is based on physiological, biochemical, or sexual characteristics and cannot be routinely utilized in the food industry. As a consequence, various miniaturized and simplified identification methods have been developed. However, they are based on the same approach of the classical method of yeast identification and are time-consuming, even when procedures are automated and computerized, and often lead to false or equivocal identifications. To overcome these difficulties, alternative faster typing methods have been developed, based, among others, on analysis of

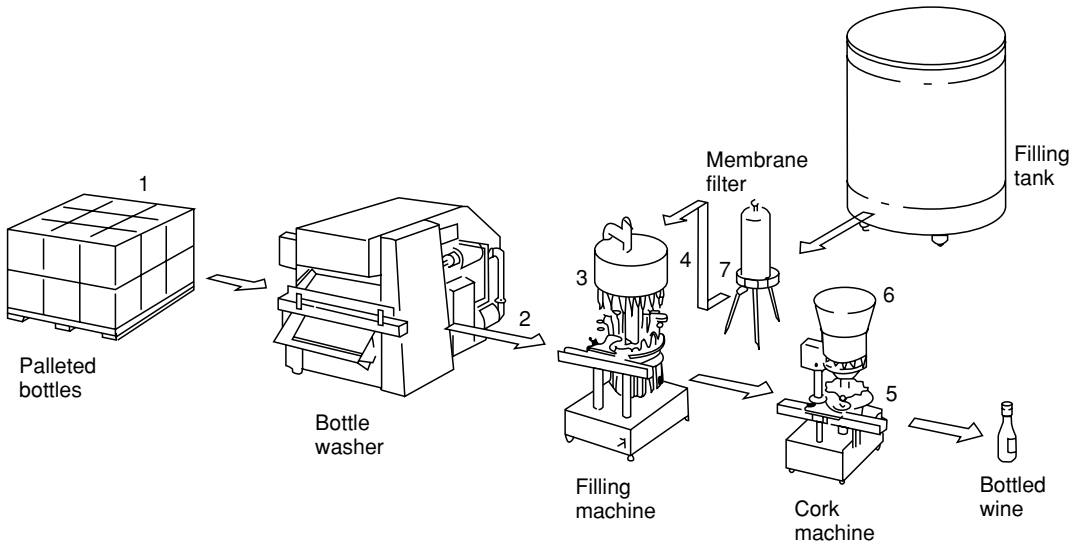

Figure 1 Main sampling points in spoilage-yeast monitoring of a cold-sterile wine bottling line. Empty bottles (1) are frequently infected by fungi, some bacteria (generally spore-forming), and adventitious yeasts; after washing/rinsing (2), an increase in the microbial load is frequently seen, and spoilage yeast infections may occur (about 5% of all outbreaks). The filling machine is usually the main cause of contamination (c. 50% of cases), particularly owing to deficient disinfection of the filling heads (3) and self-levelling system (4). The cork machine is responsible for about 30% of outbreaks, mainly owing to contamination of cork jaws (5) and also the corks (6); membrane filters occasionally (c. 10% of cases) may be a serious infection point when a membrane rupture occurs.

total proteins, total long-chain fatty acids, and isoenzymes. In addition, recent progress in molecular biology has led to the development of highly specific techniques (e.g., restriction-fragment length polymorphism of mitochondrial DNA, chromosomal DNA electrophoresis, restriction-enzyme analysis of polymerase chain reaction-amplified ribosomal DNA, and random amplified polymorphic DNA assay).

Zymological indicators The microbiological quality of foods is currently evaluated through microbiological indicators. When yeasts are the most dangerous microorganisms for the stability of foodstuffs, the microbiological indicator used is, generally, the 'yeast and molds,' resulting from the utilization of a general-purpose plating medium to enumerate microfungi. However, this indicator does not distinguish between dangerous and adventitious yeasts, and so other more meaningful indicators may be used. For instance, yeasts tolerant to environmental stress are detected/enumerated by general-purpose media modified according to the stress factor involved. Thus, 'acid-resistant yeasts' is a zymological (zymo = yeast) indicator used to characterize dangerous yeasts in acid food industries, 'xerotolerant (osmophilic) yeasts' is used for grouping yeast particularly dangerous in industries that process low-available-water foods. Occasionally, it may be of interest to detect the presence of yeasts with specific hydrolytic capacities, namely 'lipolytic yeasts,' 'pectinolytic yeasts,' and 'proteolytic yeasts,' using the appropriate culture media. These indicators may also be directed to detect individual spoilage yeast species provided the selective or differential media to quantify them in a rapid way. Examples of this approach include media for *Debaryomyces hansenii* (e.g., intermediate moisture foods and cheeses), *Dekkera/Brettanomyces* spp. (e.g., wines, beers, and ciders), *Kluyveromyces marxianus* and *K. lactis* (e.g., dairy industry), *Yarrowia lipolytica* (e.g., butter and cheeses) and *Z. bailii* and *Z. bisporus* (e.g., wines, beverages, and fruit concentrates).

An alternative approach to the zymological indicators, determined by microbiological methods, is to examine food samples for chemical (or sensorial) evidence of past microbial activity. The use of metabolites as indicators of spoilage is often more convenient and faster than using microbiological methods of examination. Examples are given by determination of ethanol and acetoin (e.g., fresh fruits), carbon dioxide (e.g., detection of fermentative yeasts in foods), 4-ethyl-phenol (chemical marker for *Dekkera/Brettanomyces* in wines) and 1,3-pentadiene (chemical marker for *Z. rouxii* and *D. hansenii* in marzipan based products preserved with sorbate). The two last chemical indicators may be regarded as 'spoilage predictors,' because at low concentrations, they may reflect future product deterioration.

See also: **Amino Acids**: Metabolism; **Freezing**: Structural and Flavor (Flavour) Changes; **Irradiation of Foods**: Basic Principles; **Lactic Acid Bacteria**; **Pasteurization**: Principles; **pH – Principles and Measurement**; **Preservation of Food**; **Sanitization**; **Sterilization of Foods**; **Water Activity**: Effect on Food Stability; **Wines**: Production of Table Wines; Production of Sparkling Wines

Further Reading

Deak T (1995) Methods for the detection and identification of yeasts in foods. *Trends in Food Science and Technology* 6: 287–292.

Deak T and Beuchat LR (1996) *Handbook of Food Spoilage Yeasts*. Boca Raton, FL: CRC Press.

Fleet GH (1992) Spoilage yeasts. *Critical Reviews in Biotechnology* 12: 1–44.

King AD, Pitt JI, Beuchat LR and Corry JE (eds) (1986) *Methods for the Mycological Examination of Food*. New York: Plenum Press.

Kurtzman CP and Fell JW (eds) (1998) *The Yeasts. A Taxonomic Study*, 4th edn. Amsterdam: Elsevier.

Pitt JI and Hocking AD (1985) *Fungi and Food Spoilage*. Sydney: Academic Press.

Samson RA, Hocking AD, Pitt JI and King AD (eds) (1992) *Modern Methods in Food Mycology*. London: Academic Press.

Skinner FA, Passmore SM and Davenport RR (eds) (1980) *Biology and Activities of Yeasts*. London: Academic Press.

Thomas DS (1993) Yeasts as spoilage organisms in beverages. In: Rose AH and Harrison JS (eds) *The Yeasts*, 2nd edn. vol. 5, pp. 517–561. London: Academic Press.

Tudor EA and Board RG (1993) Food spoilage yeasts. In: Rose AH and Harrison JS (eds) *The Yeasts*, 2nd edn. vol. 5, pp. 435–516. London: Academic Press.

Spores *See* **Allergens**; **Food Intolerance**: Types; Food Allergies; Milk Allergy; Lactose Intolerance; Elimination Diets; **Microbiology**: Classification of Microorganisms; Detection of Foodborne Pathogens and their Toxins; **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage; Fungi in Food – An Overview; Molds in Spoilage; Yeasts in Spoilage

Spray Drying *See* **Drying**: Theory of Air-drying; Drying Using Natural Radiation; Fluidized-bed Drying; Spray Drying; Dielectric and Osmotic Drying; Physical and Structural Changes; Chemical Changes; Hygiene; Equipment Used in Drying Foods

Squashes *See* **Melons, Squashes, and Gourds**

Squid *See* **Marine Foods**: Production and Uses of Marine Algae; Edible Animals Found in the Sea; Marine Mammals as Meat Sources

STABILIZERS

Contents
Types and Function
Applications

Types and Function

J E Fox, P Ingenpass and S Zachow, GC Hahn and Company, Lubeck, Germany

Introduction

That familiarity breeds contempt is an oft-quoted maxim of political life, yet it might also be applied to many of our foodstuffs, which have acquired the epithet of common, basic, staple, or simple. The majority of these foods are, in a physicochemical sense, far from simple. Almost all are complex multiphase systems, in which gases, liquids, or solids are dispersed in a liquid continuous phase, typically aqueous, which itself contains a multitude of solutes. Examples of such natural foods are milk and eggs, whilst icecream and mayonnaise are representatives from manufactured foods.

Disperse systems are inherently instable, and this is true for many of these complex foods. Thus, full-fat milk spontaneously creams on standing, and acidified milk irreversibly curdles. This instability can be resolved into two principal processes:

1. Flocculation of fine or colloidally dispersed particles to aggregates which, if liquid or gaseous, may coalesce to form larger droplets or bubbles.
2. Displacement of particles under gravity as a result of density differences between the phases.

These processes can influence the appearance and rheology of the food and may detract from its perceived quality and acceptance. The precise effects, however, will depend on the volume fraction of the disperse phase, and will naturally be greater where this is higher. Particularly in such foods, it is important to control or inhibit flocculation and gravitationally induced phase separation. (*See* **Flocculation**.)

The word 'control' is included here by virtue of the fact that, in some foods, it may be advantageous to promote these processes. Thus, in fat-reduced mayonnaise, an aggregation of fat droplets may be required to give the product a structure and plasticity. Aggregation and structure formation may also be utilized to prevent phase separation. In fat-rich fresh cheese products, a casein network prevents the creaming of the fat, and a similar mechanism is operative in the stabilization of foams by fat crystals. Although contradictory in terms, a stabilizer may achieve stability by controlled and partial destabilization. (*See* **Casein and Caseinates**: Uses in the Food Industry.)

Control of flocculation and separation can be achieved by the judicious use of polysaccharide gums, proteins, organic salts (and emulsifiers), alone or more often in combination. In this role these substances have become known as stabilizers, a suitable definition for which would be as follows: food additives which prevent or control phase separation in foods consisting of two or more phases.

Because stabilizers influence the structure of a food, and because many of the additives used have, in their own right, a thickening or gelling function, the term stabilizer has become loosely used to describe any additive that influences the rheology of the products. This is a somewhat dangerous practice as, in many countries, food-labeling law has led to a legal definition of the term, and this usually restricts its use to multiphase systems. It should also be remembered that the term refers to a functionality; therefore one and the same additive may not always be properly described as a stabilizer. Thus, when carrageenans are used to prepare a clear jelly – a single-phase system – it should be declared not as a stabilizer but as a gelling agent. On the other hand, in chocolate milk, where the gel serves to hold the cocoa particles in suspension, the former declaration would be correct.

Flocculation

In order to illustrate how stabilizers influence flocculation, it is necessary to review the forces which are believed to be acting on dispersed particles. These were first set out in the theory of Derjaguin Landau Verwey Overbeck, which recognizes two types of interparticle interactions:

1. Van der Waals forces.
2. Electrostatic interaction.

Van der Waals forces are strong attractive forces which are principally attributable to fluctuating polarization of the electron distribution in the molecule. These forces decrease rapidly with distance, being inversely proportional to the sixth power of the separating distance.

Dispersed particles may also carry surface charges, with the resulting formation of an electrical double layer of counterions in the proximal continuous phase. When two similarly charged particles approach each other, these ion layers begin to overlap and interact, giving rise to repulsive forces. Typically, these forces are stronger than Van der Waals forces at larger separations, so resulting in a net repulsive force. As separating distance decreases these repulsive forces attain a maximum before Van der Waals attraction causes flocculation (Figure 1). The height of this barrier will naturally be dependent upon the magnitude of electric double layer but, in order to limit flocculation, it should be considerably larger than the thermal energy of the system. The absolute magnitude of these forces is generally related to the surface area of the particle. They are therefore important where the ratio of surface area to weight of the particles is large, and therefore increase in importance with diminishing particle size.

From these considerations it follows that a colloidal dispersion may be stabilized by increasing the surface charge and this, in turn, can be achieved by the adsorption of polyelectrolytes. This approach is particularly important in the stabilization of protein dispersions, which tend to flocculation when the pH of the continuous phase approaches their isoelectric point. Here sodium carboxymethylcellulose (CMC) has proved effective down to a pH of 4.2, below which the polymer itself becomes increasingly insoluble. CMC offers the additional advantages that: (1) its calcium salts are soluble, and there is therefore no restriction on its use in calcium-rich systems, in particular milk, and (2) it is in itself a good thickening agent, and when used in excess, it can retard gravitationally induced phase separation. (*See* **Colloids and Emulsions**.)

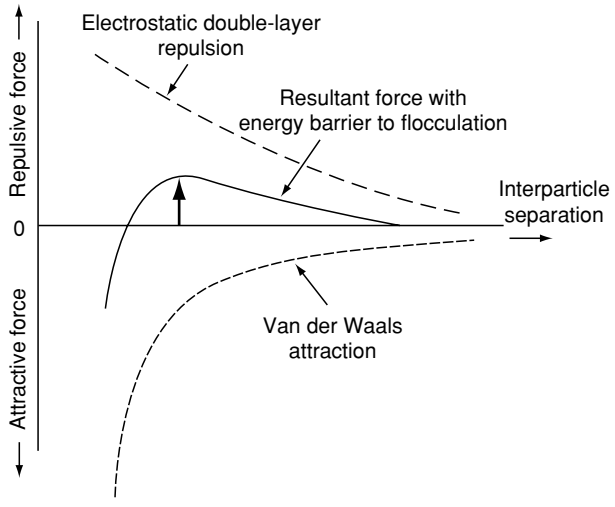

Figure 1 Schematic representation of particle interaction. Reproduced from Stabilizers/Types and Function, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Below pH 4.2, pectin and, to a lesser degree, alginate esters can be used. For stabilization, a high proportion of the carboxylic acid residues present in these molecules must be blocked, thus only high-methoxy pectin, with a degree of esterification (DE) above 70%, and propyleneglycol alginate (DE > 60%) are effective. Polymers with a higher proportion of free acid groups tend to facilitate flocculation, presumably by bridging the dispersed particles, and are calcium-sensitive. The esterification of the acid also promotes the cleavage of the glycosidic bond by β-elimination. These hydrocolloids are therefore susceptible to thermal degradation, particularly in neutral systems, and this can be a handicap for their use in pasteurized and sterilized products.

With these pectins and alginate derivates, a further mechanism of stabilization may be operative. The steric effects of the long molecules adsorbed on to the surface of the particles may prevent their mutual approach. In order to function in this way, the stabilizing polymer must have an appreciable portion of its length as a fully solvated chain in the proximal solvent (Figure 2). This requirement can rarely be met by homopolymers and usually requires block polymer species, such as high-methoxy pectin, containing both lyophilic and lyophobic segments. This stabilizing mechanism will be affected by the solvent quality with respect to the lyophilic portion; reducing the quality will decrease stability. The stabilization also requires that the surface be fully coated with polymer, which should therefore be used in excess. If added sparingly, lyophobic portions of the same molecule may cause bridging between particles and thereby induce flocculation.

Gravitationally Induced Phase Separation

Neglecting interparticle forces, the movement of particulate phase under gravity is governed ideally by Stokes law. This law relates the terminal velocity v of a spherical body to the viscosity of continuous

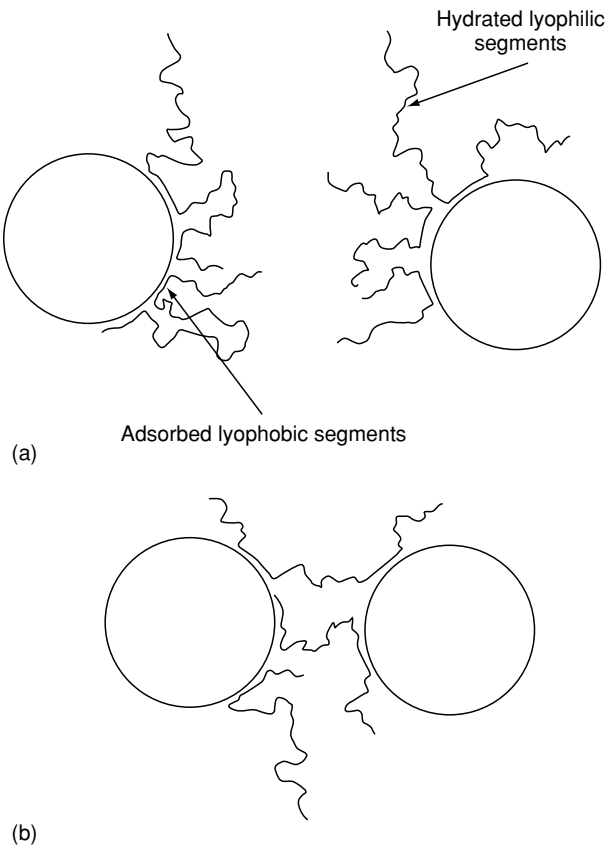

Figure 2 (a) Steric stabilization and (b) flocculation. Reproduced from Stabilizers/Types and Function, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

phase, η, the radius of the particle, r, and the force acting on the particle, f:

$$\nu = (6\pi f)/\eta r$$

$$f = (\rho_p - \rho_s)\tfrac{4}{3}\pi r^3$$

The density of the particle is ρ_p, and that of the continuous phase is ρ_s.

It follows from these equations that the speed of creaming or sedimentation increases with particle size and decreases with the viscosity of the continuous phase. This is, of course, the rationality behind the use of homogenization to delay creaming in milk, or the use of thickening agents to delay sedimentation.

In order to have appreciable influence on stability, the viscosity must be drastically increased and this can best be achieved by the use of polysaccharide gums. These substances are extremely efficient thickening agents, being effective at concentrations typically below 1.0%. (*See* **Gums**: Properties of Individual Gums.)

The effect on viscosity is believed to be the result of nonspecific entanglement of the randomly coiled polysaccharide chains. It follows from this model that the more extended the polysaccharide molecules are in solution, the more entanglement would be likely and the higher the viscosity obtained. The extension, and thereby the viscosity, is governed by size and topology of the molecule, its flexibility, and its interaction with the solvent. Given the same molecular weight, a linear molecule will be more extended and more prone to entanglement than a highly branched species. For linear molecules the extension should be directly related to the molecular weight. (*See* **Carbohydrates**: Classification and Properties.)

It is therefore not surprising to find that all commercially important gum thickeners are linear polymers, and that they are available in a wide range of viscosities which reflect their degree of polymerization.

The fact that these substances show a much larger effect on viscosity as compared with equally large synthetic polymers or proteins has been ascribed to the bulky monomer sugar which restricts chain flexibility. None of these polymers are rigid rods, however, so that the quality of the solvent will influence the degree of extension. In thermodynamically poor solvents, and in the presence of salts which shield electrostatic repulsive forces in polyelectrolytes, the molecules tend to assume more compact conformations and the efficiency of thickening will be reduced. It also follows from the model of molecular entanglement that the concentration must be such that physical contact is possible. This condition is usually satisfied when the concentration of hydrocolloid gums, c, is such that:

$$c.[\eta] \geq 4$$

In this equation, $[\eta]$ is the intrinsic viscosity of the polysaccharide; $c.[\eta]$ is known as the overlap factor. The intrinsic viscosity is a measure of the extension of the molecule in solution and can be related to its radius of gyration.

The flow properties of hydrocolloid-thickened solutions are markedly pseudoplastic (Figure 3). Under the influence of shear, the dynamic viscosity falls. This is thought to be the result of orientation of the extended molecules under the shear gradient, with a concomitant reduction in entanglement. This property can be used to great advantage in stabilizers. The shear stresses operative in gravitational sedimentation or creaming are in the order of 10^0–10^1 mPa. At these extremely low shear stresses, hydrocolloids exhibit extremely high dynamic viscosities. A 1% xanthan solution shows a viscosity of 10^5 mPa s^{-1} under a shear stress of 10 mPa. At shear rates of between 10 and 50 s^{-1}, believed to be present in the mouth during mastication, the viscosity of the same

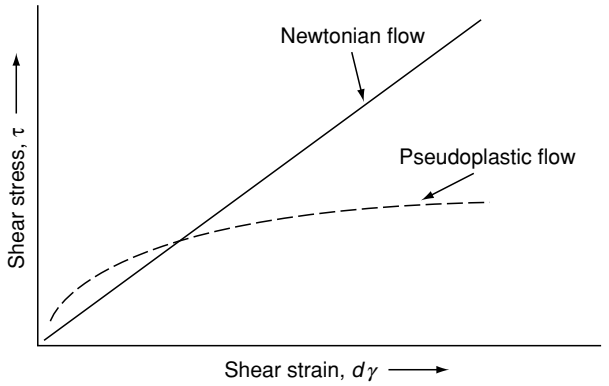

Figure 3 Schematic representation of pseudoplastic and Newtonian flow. $\eta = \tau/d\gamma$, where η is dynamic viscosity. Reproduced from Stabilizers/Type and Function, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

xanthan solution falls to $10^2-10^3\,\text{mPa}\,\text{s}^{-1}$. Thus, foods thickened in this manner exhibit a light, pleasant mouth feel.

The same stabilization effect cannot be obtained with a thickener showing Newtonic flow characteristics in which viscosity is independent of shear. If the high viscosities could be obtained at low shear then the product would be unacceptably thick on eating. This can be illustrated by honey, a Newtonic fluid thickened by mono- and disaccharides, which has a viscosity of only $10^4\,\text{mPa}\,\text{s}^{-1}$, independent of shear rate. Organoleptically, such Newtonically thickened solutions are perceived as long and slimy.

This approach to stabilization, and the use of non-absorbed polysaccharides in food in general, is limited by thermodynamic incompatibility with other colloidally dispersed polymers. This prevents the polysaccharide and its incompatible neighbor from occupying the same space, the so-called excluded volume, and can drive the excluded colloid to flocculate or lead to phase separation. Such thermodynamic incompatibility is known to exist between proteins and polysaccharides and the effect is therefore important in foods containing colloidally dispersed proteins such as milk or gelatin solutions. Thus, in milk, the addition of a low concentration of hydrocolloid with a large hydrodynamic volume (i.e., extended conformations) will cause curdling, while gelatin solution, in the presence of larger concentrations, will exhibit phase separation. Indeed this effect may be used to advantage in the gelatin encapsulation of fats or, possibly, in the stabilization of foams. (*See* **Carbohydrates**: Interactions with Other Food Components; **Protein**: Interactions and Reactions Involved in Food Processing.)

A further explanation for the flocculation of colloidal particles has been proposed; this relies on the driving force of the osmotic gradient, caused by the exclusion of bulky polymer molecules from the narrow gap between two approaching particles. This phenomenon, christened depletion flocculation, has yet to be definitively observed in food systems.

The seed gums, guar and locust bean gum (LBG), which are both galactomannans, together with the microbiological slime, xanthan, are the principal hydrocolloids used commercially for thickening foods. Of these, xanthan shows the most pronounced and desirable pseudoplastic flow and, in its purified form, is colorless and tasteless. Although in many respects an ideal stabilizer, xanthan is considerably more expensive than the seed gums and, presumably for this reason, its use has been somewhat restricted. Guar suffers from the problem of taste and, although it can be reduced by steam treatment, this erodes the price advantage. For LBG the critical quality characteristic is the presence of dark specks, which originate from the seed coat and which can prohibit its use in white products. (*See* **Gums**: Nutritional Role of Guar Gum.)

These hydrocolloid thickeners can be used individually, but it is well known that galactomannans interact with xanthan and other hydrocolloids. This so-called synergism results, in the case of guar and xanthan, in the mixture having a higher viscosity than predicted on the basis of the individual concentrations. In the case of LBG and xanthan, the mixture gels. These effects are believed to be the result of ordered associations forming cross-links between the chains. The naked mannan segments of the galactomannans are implicated in the formation of these junction zones with xanthan, although their exact form is at present unknown.

Generally speaking, viscosity increase should only be used to stabilize suspensions in which the particle size is small. With macroparticular suspension, such as herbs in oil-free dressings or chocolate particles in milk, greater resistance to movement is usually required and here the elastic properties of the continuous phase should be augmented, i.e., the gel character of the continuous phase increased.

The origin of gelling in polysaccharide solution is believed to lie in formation of intermolecular cross-links between the individual chains. With the onset of cross-linking, the solution becomes progressively more viscous until a distinct gel begins to form. With increasing cross-linking the gel loses its elasticity, becomes more brittle, and exhibits syneresis. Eventually, the hydrocolloids may precipitate out of solution. The cross-links, or junction zones, are

complex, involving at least two polysaccharide strands, and are believed to be crystalline in nature. The driving force in their formation is hydrogen bonding between hydroxyl groups and/or ionic bonding where polyelectrolytes are involved.

For the stabilization of liquid suspensions generally, a weak and, preferably, thixotropic gel should be used, and the breaking strength should be chosen so that the weakly gelled aqueous phase flows when subject to the high shear stresses on pouring. In principle, any gelling polysaccharide can be employed but, in practice, carrageenans enjoy the widest commercial use. At dosage levels of 0.5%, Jota carrageenans give a weak thixotropic gel suitable for water-based dressings. It is a requirement of their use, however, that the dressing is filled at a temperature of around 20 °C. This may be bacteriologically undesirable and, in such cases, a weakly gelling xanthan galactomannan mixture may be used, hot-filled. Where milk is the suspending medium, κ-carrageenans at dosage levels of 0.02% provide a satisfactory gel.

Ostwald Ripening

One further mechanism of destabilization is known.

In polydisperse systems it can be observed that larger particles grow at the expense of smaller ones. This disproportionation phenomenon, which was first described by Ostwald in crystallization processes, is only important in the context of food systems in the destabilization of foams. Gas trapped in a foam bubble is subject to a pressure, the Laplace pressure, due to the surface tension of the surrounding, typically aqueous, interface. The magnitude of this internal pressure is inversely proportional to the diameter of the bubble, i.e., the smaller the bubble, the higher the pressure. If the gas is soluble in the continuous phase then an equilibrium will be established between the pressure of the gas in the bubble and the concentration of dissolved gas in the continuous phase immediately surrounding the interface. This dissolved gas may diffuse away from the interface or the continuous phase itself may be transported in bulk through the foam. If this becomes adjacent to a larger bubble with a lower Laplace pressure then gas will be released from solution to reestablish the equilibrium.

The speed of Ostwald ripening in foams may be reduced by decreasing the surface tension of the aqueous interface or decreasing the solubility of the gas in the aqueous continuous phase. In practice, however, longer-term stability can be conferred on the foam by gelling the continuous phase so that bubble disproportionation cannot take place.

See also: **Carbohydrates**: Classification and Properties; Interactions with Other Food Components; **Casein and Caseinates**: Uses in the Food Industry; **Colloids and Emulsions**; **Flocculation**; **Gums**: Properties of Individual Gums; Nutritional Role of Guar Gum; **Protein**: Interactions and Reactions Involved in Food Processing

Further Reading

Dickinson E and Stainsby G (1988) *Advances in Food Emulsions and Foams*. Essex: Elsevier Science Publishers.

Larsson K and Friberg SE (eds) (1990) *Food Emulsions*. New York: Marcel Dekker.

Mitchell JR and Ledward DA (1986) *Functional Properties of Food Macromolecules*. Essex: Elsevier Science Publishers.

Philips GO, Williams PA and Wedlock DJ (eds) (1989) *Gums and Stabilisers for the Food Industry*, vol. 5. Oxford: IRL.

Tanford C (1961) *Physical Chemistry of Macromolecules*. Chichester: John Wiley.

Applications

J E Fox, P Ingenpass and S Zachow, GC Hahn and Company, Lubeck, Germany

Neutral Milk Products

The inhabitants of Normandy and Ireland's west coast were historically the first to use hydrocolloids as food additives. With the help of calcium-sensitive, gelling polysaccharides which were extracted from red seaweeds, and which later became known as carrageenans, a gelled milk pudding or blancmange was prepared.

Carrageenans are still used in food in this capacity. κ-Carrageenan dosages between 0.2% and 0.3% produce a brittle-gelled pudding when hot-filled above the gel melting point of about 70 °C, but should the product be subject to shear whilst gelling and cold-filled, then a smooth, creamy, custard-type product results. The consistency and mouth feel of both the hot- and cold-filled products can be modified by the incorporation of other polysaccharides. Starch, for example, will improve the body and locust bean gum (LBG) the elasticity. (*See* **Gums**: Properties of Individual Gums.)

The classification of carrageenans as stabilizers is, in these products, a somewhat moot point. There are, however, two well-established applications of carrageenans in which their stabilizing functionality is undisputed.

Ultra-heat-treated (UHT) Whipping Cream

One aspect of the instability of milk, and particularly cream, arises from the flocculation and creaming of the milk fat. In short-life products, this can be adequately controlled by homogenizing the product to reduce the oil droplet size. However, this cannot be done with long-life UHT whipping cream because it would impair the whipping properties. The problem can be solved by the incorporation of stabilizers which form weak gels. (*See* **Cream**: Types of Cream.)

Carrageenans have proved efficient here when added at levels between 0.01% and 0.02%. The stabilizer should be dissolved at 70 °C in the skimmed milk used to standardize the fat content, before UHT treatment. In these systems, the casein micelles appear to play an active role in the gel structure. The carrageenan becomes attached to the micelle surface by a process other than that requiring the intermediacy of calcium ions, and probably involving the cationically charged section of the casein molecules. It is therefore not surprising to find that the strength of the gel is dependent on casein quality or, more precisely, on the thermal history of the cream. Where the heat treatment was prolonged or repeated, absorption or binding is poor and a higher dosage of carrageenan will be required to obtain the required stability. Carrageenans are also amongst the most thermal-instable hydrocolloids. Prolonged storage at elevated temperatures is therefore to be avoided.

Cocoa Drinks

A similar problem arises with chocolate milk. The cocoa particles tend to sediment to the bottom of the bottle, giving a deep brown, compact layer below a somewhat pale (milk) fluid. This can be corrected in the same way as the creaming in UHT cream by the formation of a weak gel using about 0.02% of carrageenans. Other hydrocolloids may also be included to improve the mouth feel of the drink. The dosage of carrageenan is critical. Too much carrageenan results in the formation of a gel which is too strong and shows syneresis on standing, leaving an almost clear serum at the top of the bottle. Too low a dosage results in a gel which is insufficiently strong to support the cocoa particles which sediment in the bottle. The quality of the cocoa can also affect the stability. Cocoa which has been treated with alkali to obtain a good color is difficult to stabilize. (*See* **Cocoa**: Production, Products, and Use.)

To obtain stabilization, the UHT or sterilized drink must also be bottled at a temperature below the gelling point of the system, and ideally 24 °C should not be exceeded.

Foamed Milk Products

Two groups of foamed milk products may be recognized: those containing fat in which the fat plays a functional role in stabilization, and those based on skimmed milk.

In fat-free systems, the milk proteins themselves are sufficient to lower surface tension and form a stabilizing layer around the air cells. This can be reinforced by the addition of other hydrolyzed proteins, and thickening agents can be added to reduce the speed of drainage of the aqueous phase from the lamellae between the air cells, which would otherwise lead to collapse of the foam. Thickening agents will not prevent Ostwald ripening, so that to insure complete stability the aqueous continuous phase must be gelled as in a mousse-type product.

An entirely different situation exists in fat-containing products. Here the fat displaces the protein from the air surface and prevents its stabilization of the foam. In such products, the foam is stabilized by an entirely different mechanism. An emulsifier, typically an acetic acid or lactic acid ester of a mono- or diglyceride, which is capable of displacing the protein from the oil-water interface, is included in the stabilizing system. On whipping or aeration, the emulsifiers promote coalescense of the fat globules, and the fat crystals, which must be present in the fat phase at the aeration temperature, form a supporting network around the air cell. Such a mechanism is operative in artificial cream, icecream, aerosol cream, cream toppings, milkshakes, and other products. (*See* **Emulsifiers**: Uses in Processed Foods.)

In addition to emulsifiers, a variety of gums are added to these stabilizer systems to modify the rheology of the aqueous phase. This is particularly important where the fat content is low or the crystallization poor. In cream toppings for desserts and cakes, 1.0–1.5% gelatin is added directly before aeration to stabilize the foam. In addition, carrageenans, guar, and starches may be added to influence the eating quality of the topping. In aerosol cream in spray cans, the cream must remain fluid in the can, and this system therefore cannot be used. Here the protein content of the cream is enhanced and a low dosage of gelling polysaccharide, carrageenan, or pectin added.

Surface-active polysaccharides, such as propylene glycol alginate or methylcellulose, may be used to enhance foaming and foam stability. Where the foam must be stable at elevated ambient temperature,

the gelatin may be totally or partially replaced by xanthan or gelling mixtures of xanthan and LBG.

Icecream

Hydrocolloids and gelatin are also widely used in icecream, and are customarily described and declared as stabilizers, although their funtionality here has little in common with stabilization in its strict sense. Depending upon the type of icecream to be produced and its fat content, typically between 0.5% and 0.8% of a stabilizer mixture is used. Of this, about 50–60% is a monoglyceride emulsifier to destabilize the fat globules; the remaining 0.2–0.4 % is hydrocolloid, of which about 80% is a thickening gum such as LBG, guar, or sodium carboxy-methylcellulose (CMC), and 20% a gelling agent, such as alginate or carrageenan (0.01–0.03%). The gums serve to improve mouth feel and limit ice crystal growth on freezing and storage. The gelling agents are added to prevent whey separation on thawing and give a slow meltdown. (*See* **Ice Cream**: Methods of Manufacture; Properties and Analysis.)

Sour Milk Products

Yogurt

On approaching the isoelectric point, the casein micelles in milk flocculate to give a delicate gel structure typified by natural yogurt. With the natural solid content of milk, the gel formed is extremely weak and fragile and unsuitable for commercial exploitation. The mouth feel is also poor and watery. The addition of 2–5% milk solids as milk powder to the base milk before incubation can remedy this situation. However, the problem remains that on aging or exposure to high ambient temperature the casein gel shows syneresis or wheying-off. To inhibit this, the milk powder may be totally or partially replaced by modified starches, gelatin, isolated milk protein, or combinations thereof. Although from an economic point of view the use of chemically modified starches is very attractive, too high a dosage gives the product a heavy, pasty and, for yogurt, very uncharacteristic mouth feel. Gelatin typically endows the product with a smooth, light texture. (*See* **Starch**: Modified Starches; **Yogurt**: The Product and its Manufacture.)

Polysaccharide hydrocolloids are not generally employed because of their flocculation of milk proteins at the isoelectric point, owing to the excluded volume effects when added at effective concentrations. The exception is the use of low-methoxy pectin, at low dosages of about 0.1%, to promote flocculation and thereby strengthen the gel.

Set yogurts, fermented in the tub, are inefficient to produce on an industrial scale. Much more suitable to the restraints of an industrial process are stirred yogurts, in which the fermentation can take place in larger vessels under carefully controlled conditions. Unfortunately, however, once the yogurt gel is broken by stirring, pumping, and filling, it will only partially reform in the tub. This situation can be remedied by the use of stabilizers which, if they are to be added to the milk before culture, are limited to those used in set yogurt, although they differ qualitatively from these.

With yogurt as a pumpable liquid, it becomes possible to pasteurize or even sterilize it in a plate heat exchanger. This removes one of the major hurdles to wide-scale distribution, i.e., the limited shelf-life. At a pH not far below the isoelectric point of the proteins, the stability of the dispersed protein is delicate and thermal energy can easily lead to flocculation. This flocculation is manifested as an unpleasant, sandy texture in thermized yogurt. This phenomenon can be delayed by the addition of high-methoxy pectin, typically at levels between 0.5% and 0.7%, before pasteurization. Pectin is believed to function by forming a protective layer on the surface of the casein micelles. A finer dispersion with a larger surface will therefore require a higher dosage. The dosage is also related to the thermal stress which the product must survive. Indeed, under ideal technology and recipe, a commercially satisfactory thermized product may be manufactured with almost no stabilizer. However, if the product is kept at elevated temperature over hours, no amount of protective colloid can prevent unacceptable sandiness. At extremely high dosage, pectin also causes flocculation owing to excluded volume effects. As an alternative to pectin, propylene glycol alginate can be, but seldom is, used commercially. CMC may be used above pH 4.2. (*See* **Pectin**: Food Use.)

Pasteurization also presents the opportunity to add new sterile ingredients to the sour product before thermization. Thus low concentrations of gums, typically guar and LBG at 0.1% dosage levels, can be added to improve the mouth feel of the product. (*See* **Pasteurization**: Principles.)

Quarg and Fresh Cheese Products

The problems associated with the production of fresh cheese products have their roots in the same phenomena as those that destabilize yogurt. (*See* **Cheeses**: Quarg and Fromage Frais.)

Collapse and contraction of the casein gel lead to syneresis or wheying-off, and heating of the casein enhances flocculation and causes sandiness in the product. Both these effects can be eliminated or

reduced by the use of stabilizers; the type and dosage depend greatly on the recipe, on the technology, and the type of product to be manufactured. A typical thermized quarg stabilizer contains 0.3% gelatin, 0.3% starch, and small amounts of gum, including CMC, guar, and LBG. Gelatin gives the quarg a pleasant, smooth, and light mouth feel and appears to function, at least partially, as a protective colloid for underefficient thermization processes – such as those occurring in a scraped-surface heat exchanger – and in the presence of moderate amounts of fat, no sandiness occurs.

When the product is subject to high thermal stress, such as prevails in a batch process with a long cooling time or when the product is fat-free, other protective colloids, such as CMC, may have to be included. Starches are usually included to give the product body. Gums confer viscosity on the product. This is of special importance in hot-whipped products, where the viscosity of the hot quarg would be insufficient to sustain a foam. In whipped products a higher dosage of gelatin is also used to stabilize the foam when cold.

Sour Milk Drinks

Compared to yogurt or quarg, sour milk drinks contain appreciably less protein, and the problem is not so much one of stabilizing a casein gel network as maintaining the individual casein flocculates in suspension. To this end, almost all sour milk drinks are homogenized under high pressure to break down the casein flocculates and give a stabilizable particle density. The resulting high surface area dictates that a higher than normal concentration of protective colloids must be employed. Typically, this is high-methoxy pectin for drinks below pH 4.2, and CMC for a pH above 4.2. Below pH 4.2, CMC becomes increasingly insoluble, and chain association leads to flocculation enhancement. Above this pH, however, CMC has the added advantage of increasing viscosity and thereby slowing sedimentation. With pectin, other gums may be added to increase viscosity.

Emulsified Foods

Two types of emulsified foods can be recognized: those such as mayonnaise and dressings, in which the oil is dispersed in a continuous aqueous phase (o/w), and those typified by such products as butter and margarine, which are water-in-oil (w/o) emulsions. Both these systems are inherently unstable owing to their high surface energy content and density differences between the phases.

Oil-in-water Emulsion

Traditional mayonnaise is an o/w emulsion containing 80% oil, in which egg yolk serves as an emulsifier. The short plastic structure of this product is attributable to three factors:

1. The fine oil droplet size, 1–10 μm, results in a relatively high Laplace pressure which resists droplet deformation.
2. The high volume of oil is somewhat in excess of the optimal packing condition, thus forcing the droplets to be in intimate contact.
3. Attractive forces between proteins on the aqueous face of the droplet surface cause adhesion and restrict movement. (*See* **Dressings and Mayonnaise**: The Products and Their Manufacture; Chemistry of the Products.)

With no room for gravitational separation and an adequate protein layer around the droplets, mayonnaise represents a perfectly stable system. The modern consumer, aware of the dangers of a high-lipid diet, has shied away from such products and demanded surrogate products showing identical rheological behavior, but with 50%, 30%, and even 15% oil contents. Although milk powder or caseinate, which promotes droplet adhesion, is used as the emulsifier, the volume of oil is too low either to form a continuous agglomerated network and to endow the product with stability against creaming, or to give the desired rheological properties. This problem may be solved by including a starch paste (10–12% solids) to replace the oil in a 1:1 ratio. This, especially in very-low-oil emulsions, gives the product a cheap and pasty mouth feel. A better and more acceptable salad or low-energy mayonnaise can be stabilized by replacing some of the starches by hydrocolloids. For example, for the cold preparation of a 30% emulsion, a blend of either LBG or guar gum and xanthan at a total dosage of 0.4–0.6% can be used in combination with starch. Guar gum or CMC can be used in addition where the fat content is very low in order to achieve a mayonnaise-like consistency. The stabilizers can be used with emulsifiers such as milk protein or egg yolk.

For the stabilization of hot-prepared dressings and mayonnaises, guar gum and xanthan in the ratio of 3:1 at an addition of about 0.50% can be employed. The viscosity can be adjusted by varying the combination and quality of raw material and stabilizer(s).

Mayonnaises are widely used for the preparation of salads. There are three special applications of hydrocolloid stabilized products. When marinated fish or fresh vegetables are included they tend to lose appreciable amounts of cell fluids which can

form unsightly pools in the mayonnaise. The incorporation of gums can help bind this water. For example, guar gum is used in combination with xanthan, LBG, and propylene glycol alginate. For dressings with high fat or cream contents, milk proteins, especially whey protein and caseinates, are widely used.

Fresh fruit and vegetables often contain amylases which attack and hydrolyze the starch in the mayonnaise with the concomitant loss of viscosity and stability. For such applications, special mayonnaises have been developed which are wholly stabilized by enzyme-resistant hydrocolloids. Typical are blends of guar gum and xanthan in the ratio of 3:1 at an addition rate of 0.50%. The water-binding ability and 'cling' of the dressing will be increased by adding CMC to this blend. Where the oil content of the mayonnaise has been reduced in the formulation of a calorie-reduced or 'light' product, the consequent reduction in opacity can be compensated by the addition of milk proteins.

In some salads, where the proportion of mayonnaise is low, it must be prevented from running-off the other components. This feature, known as cling, is important in such products as coleslaw and may be achieved by improving the pseudoplastic flow characteristics of the emulsion by the use of stabilizers. Good cling to salad ingredients, also to hot salads, will be obtained by using balanced combination of water-binding and gellifying hydrocolloids.

Water-in-Oil Emulsions

Those arguments which led to the introduction of low-oil mayonnaises have also promoted the development of fat-reduced spreads. In butter or traditional margarine, the distribution of the finely dispersed aqueous phase is stabilized by the crystal matrix in the fat. With half-fat products and spreads containing more than 60% aqueous phase, the crystal matrix is depleted and insufficient to insure stability. Hydrocolloid stabilizers are, of course, unable to influence the rheology of the fat phase. The usual approach is therefore to prevent coalescence of the droplets by increasing their viscosity or gelling. The viscosity can be raised by the addition of proteins or gums. The use of 2% caseinate in combination with other milk proteins at a total dosage of about 4% is necessary for the stability of the emulsion, possibly in combination with fatty acid emulsifiers. The quality of the product with regard to melting and spreadability can be improved when adding hydrocolloids such as carrageenan and xanthan. For gelling, gelating is customarily used by virtue of its low melting temperature. (*See* **Margarine**: Types and Properties; Methods of Manufacture.)

Deep-Frozen Foods

The crystallization of both oil and water phases in deep-frozen products presents a special stabilization problem. In addition to damaging and weakening cell structures, crystal growth also physically concentrates the dispersed phase and thereby promotes agglomeration. Icecream is typical of the foods in which this can occur. In an analogous manner, ice crystal growth within polysaccharide gels forces the chains together; these may then form macroscopic, solid-phase regions between the crystals which will not rehydrate on thawing. The effect is most pronounced in native starch gels where, after freezing, the gel resembles a sponge from which the water can be removed by simply squeezing. This situation can be remedied by the addition of low levels of a hydrocolloidal gum to the paste. Typically, 0.15% xanthan and CMC is used. Alternatively, chemically derivatized starches which show little or no retrogration, such as hydroxypropyl derivatives, can be used. The addition of emulsifiers and special starch-degradation products counteracts crystallization. (*See* **Freezing**: Principles.)

This behavior is not restricted to starch. Where the chemical nature of the gelling polysaccharide is conducive to ordered association, considerable syneresis results from a freeze–thaw cycle. Such gels are, by virtue of their propensity to cross-linking, naturally strong and brittle. As with starch, syneresis can be limited by the addition of small amounts of nongelling gums. With κ-carrageenan – a typical brittle, freeze–thaw-unstable gel – small additions of LBG may be used, but this will also influence gel rheology. As would be expected, weak, elastic, and preferably thixotropic gels may survive freeze–thaw cycles undamaged. Commercially, ι-carrageenans are widely used in this role.

Crystallization of the fat in an o/w emulsion usually results in a penetration of the protective layer around the oil droplet with subsequent coalescence on thawing. Provided that the oil content remains low, preferably $< 20\%$, emulsions which do not break can be prepared by using propylene glycol alginate as the emulsifier in combination with xanthan as the stabilizer at levels of 0.5–1.0%.

Miscellaneous Applications

More than 10 years ago, the first oil-free dressing appeared on the market in which the herbs were suspended in a clear aqueous phase. In these products the herbs are held in a suspension by a weak, preferably thixotropic gel, usually based on carrageenans. This technique may be applied to a wide range of

novelty drinks, including those that contain suspensions of fruit or fruit pieces.

Many fruits do not retain their physical integrity on cooking and turn into a nondescript mash. Where cooking stability is required, the raw fruit purée can be reformed into a fruit shape with an alginate gel. Such gels do not melt at normal cooking or baking temperatures. The technique is widely used, though for a different reason, to reform paprika in cocktail olives. Currently its use in reformed meat is being advocated.

Natural meats and fish show a considerable weight loss on cooking. To counteract this, mixtures of salts and hydrocolloids, typically carrageenans, are injected or pumped into the raw meat and the mixture gels after cooking. In meat pies and pastries, the escaping cell fluid can also disfigure the pastry surface. This may be obviated by incorporation of cellulose ethers, usually methylcellulose, into the filling. In contrast to almost all other polysaccharide food additives, colloidal solutions of methylcellulose form a gel when heated and thereby prevent the escape or boil-out of juices. The technique is also applicable to fruit-filled bakery products. Cellulose ethers have also been employed as film-building agents to reduce the penetration and uptake of fats in deep-fried foods.

See also: **Cheeses**: Quarg and Fromage Frais; **Cocoa**: Production, Products, and Use; **Cream**: Types of Cream; **Dressings and Mayonnaise**: The Products and Their Manufacture; **Emulsifiers**: Uses in Processed Foods; **Freezing**: Principles; **Gums**: Properties of Individual Gums; **Ice Cream**: Methods of Manufacture; Properties and Analysis; **Margarine**: Types and Properties; Methods of Manufacture; **Pasteurization**: Principles; **Pectin**: Food Use; **Starch**: Modified Starches; **Yogurt**: The Product and its Manufacture

Further Reading

Glicksman M (1969) *Gum Technology in the Food Industry*. London: Academic Press.

Tamime AY and Robinson RK (1985) *Yoghurt – Science and Technology*. Oxford: Pergamon Press.

Webb BH, Johnson AH and Alford JA (1978) *Fundamentals of Dairy Chemistry*. Westport, Connecticut: AVI Publishing.

STAPHYLOCOCCUS

Contents

Properties and Occurrence

A L Wong and M S Bergdoll[†], University of Wisconsin, Madison, WI, USA

Classification

In 1880, Ogston first demonstrated that a cluster-forming coccus caused many types of purulent infections in humans. He named the organism 'staphylococcus'. In 1884, Rosenbach isolated staphylococci from pus and determined that the organisms were identical to those described by Ogston. He proposed the genus *Staphylococcus* to contain these organisms. Rosenbach observed yellow and white colonies formed by the staphylococci and named the yellow colony-forming organism *Staphylococcus aureus*, and the white colony-forming organism *S. albus*. It was later determined that pigmentation was a variable characteristic and therefore not a valid criterion for classifying the staphylococci, and *S. albus* was reclassified as *S. aureus*. *Staphylococcus* was placed in the family Micrococcaceae in 1920.

In 1908, Winslow and Winslow proposed a second species, *S. epidermidis*. It was not until 1974 that a third species, *S. saprophyticus*, was added. By 1980, the number of species had increased to 13 and by 1984 to 20. Currently, at least 32 species have been described, based on DNA homology and immunochemical and biochemical characteristics.

[†]Deceased.

Staphylococci are Gram-positive, catalase-positive, nonmotile cocci that characteristically divide in more than one plane to form irregular clusters. The cell wall typically contains teichoic acid, and the peptide subunits of the peptidoglycan are cross-linked by pentapeptide bridges containing solely or primarily glycine. Staphylococci are facultative anaerobes but grow more rapidly and abundantly under aerobic conditions. A variety of carbohydrates and/or amino acids can be utilized as carbon and energy sources. The staphylococci can be differentiated from members of the strictly aerobic genus *Micrococcus* by their ability to grow and produce acid from glucose anaerobically. In addition, the two genera have different cell wall structures, cytochrome and fatty acid profiles, menaquinone content, and aliphatic hydrocarbons.

Some of the characteristics used to differentiate staphylococcal species are given in Table 1. *S. aureus* is the predominant species involved in staphylococcal food poisoning outbreaks. The illness is caused by the ingestion of preformed staphylococcal enterotoxins in foods. Thirteen enterotoxins have been identified: enterotoxins A (SEA), B (SEB), C1 (SEC1), C2 (SEC2), C3 (SEC3), D (SED), E (SEE), G (SEG), H (SEH), I (SEI), J (SEJ), K (SEK), and L (SEL). Two of the newer species, *S. intermedius* and *S. hyicus*, were formerly considered to be *S. aureus*. Although they differ in some of their characteristics, they can produce coagulase and thermonuclease (TNase), the two characteristics used most frequently to distinguish *S. aureus* from other staphylococci. All the other staphylococcal species are coagulase- and TNase-negative. The use of the name *S. aureus* herein includes both *S. intermedius* and *S. hyicus* because, essentially, all of the information presented was obtained before these two species were separated from *S. aureus*. Occasional strains of a few coagulase-negative species have been reported to produce enterotoxin. The degree of involvement of these organisms in food poisoning is unknown.

Ecology

Staphylococci are ubiquitous. Their main habitats are the noses, skin, and throats of humans and warm-blooded animals. Staphylococci can also be found in the air, soil, water, sewage, plant surfaces and products, meats, poultry, and dairy products.

Certain staphylococcal species show host preferences. The primary host for *S. aureus* is the human; however, the organism can be found also in a variety of animals and birds, and can cause many types of infections and diseases. *S. epidermidis* is the most prevalent on human skin, whereas *S. hyicus* commonly inhabits the nares and skin of chickens and pigs. Dogs are the preferred host for *S. intermedius*, which can also be found on pigeons.

About 30–50% of healthy individuals are carriers of *S. aureus*, with 40–50% of the isolates being enterotoxin producers. The incidence of enterotoxigenic strains among individuals with staphylococcal infections is higher than among healthy people. In general, the incidence of enterotoxin-producing staphylococci in healthy animals is relatively low. *S. aureus* causes many types of infections and diseases in animals. The incidence of enterotoxigenic staphylococci from mastitic animals depends on the animal species. A high percentage (60–80%) of staphylococcal isolates from mastitic sheep and goats are enterotoxigenic, whereas less than 15% of those isolated from mastitic cows produce enterotoxin.

Essentially all raw foods, especially raw meats and poultry, can be contaminated with *S. aureus*, by humans or animals, or both. Raw milk often contains staphylococci, and the organism may be isolated from 30–50% of poultry carcasses. Staphylococci isolated from goats and sheep usually produce SEC, and those

Table 1 Characteristics of *Staphylococcus* species

Property	S. aureus	S. intermedius	S. hyicus	S. epidermidis
Pigment	+[a]	−	−	−
Coagulase	+	+	±	−
Thermonuclease	+	+	+	−
Hemolysins	+	+	−	±
Mannitol[b]	+	−	−	−
Acetoin	+	−	−	+
Clumping factor	+	+	±	−
Hyaluronidase	+	−	+	−
Lysostaphin	HS[c]	HS	HS	SS[d]

[a] +, over 90% positive; −, over, 90% negative; ±, 11–89% positive.
[b] Anaerobic conditions.
[c] HS, high sensitivity.
[d] SS, slight sensitivity.

from cows produce either SEC or SED, whereas strains isolated from humans produce primarily SEA, the enterotoxin type most commonly involved in staphylococcal food poisoning. (*See* **Meat**: Preservation; **Milk**: Processing of Liquid Milk; **Poultry**: Ducks and Geese; Turkey.)

Growth and Enterotoxin Production

Growth and enterotoxin production by *S. aureus* are influenced by a variety of environmental and nutritional factors including temperature, pH, water activity (a_w), inoculum size, atmospheric composition, carbon and nitrogen sources, salt levels, and competing microflora. Generally, growth is necessary for enterotoxin production, although toxin production does not always accompany growth, especially in foods. Experimental nongrowing cell cultures have been observed to produce enterotoxin. Production of SEB and SEC is affected more by the culture conditions than is SEA production, which is more closely related to the growth of *S. aureus*.

Temperature

S. aureus can grow between 7 and 47.8 °C, with an optimum of 37 °C. Enterotoxin is produced at 10–46 °C, with an optimum temperature range of 37–45 °C. The temperature for supporting growth of staphylococci in a variety of foods ranges from 6.7 to 45.6 °C, with no growth occurring below 5.6 °C. Vanilla pudding inoculated with *S. aureus* supported enterotoxin production over a temperature range of 10–45 °C, whereas pasteurized milk supported growth and enterotoxin production at 20–35 °C, but not at 10 °C. Detectable levels of SEA were produced in 12 h when incubated at 35 °C. Longer incubation times were necessary at lower temperatures for enterotoxin production. SEB was produced in cured hams that were incubated for 2 weeks at 10 °C.

pH

The optimum pH for growth is between 6.0 and 7.0, but the organism can grow over a pH range of 4.0–9.8. The pH at which a strain will grow depends on other cultural parameters such as the atmosphere, a_w value, the type of medium and the salt concentration. Generally, the less optimal the other parameters, the narrower the pH range tolerated by *S. aureus*. For example, the lowest pH at which *S. aureus* grew and produced enterotoxin in aerobic cultures was 4.0, whereas the lowest pH values that supported growth and enterotoxin formation anaerobically were 4.6 and 5.3, respectively. (*See* **pH – Principles and Measurement**.)

Enterotoxin can be produced over a pH range of 4.0–9.0, the optimum being 6.5–7.5. As with staphylococcal growth, whether enterotoxin will be produced at a certain pH depends on other cultural parameters. The acid used for pH adjustment is also an influencing factor. When milk was acidified with hydrochloric acid, SEA was produced at pH levels of 4.5, 5.0, 6.0, and 6.4. However, when lactic acid was used, growth and enterotoxin production were observed at the higher pH levels, but not at pH 4.5. Foods having pH values below 5.0 or above 9.0 do not support enterotoxin production.

Water Activity

S. aureus can grow over a much wider a_w range than many other food-associated pathogens, with growth of some strains occurring at an a_w of 0.86; the optimum is > 0.99. The minimum a_w for anaerobic growth is 0.90. The minimum a_w for enterotoxin production is 0.86, and the optimum is > 0.99. (*See* **Water Activity**: Effect on Food Stability.)

The humectant used for a_w adjustment has a significant effect. For example, when sodium chloride was used, the minimum a_w for SEB production was 0.90–0.92. In a mixture of sodium chloride, potassium chloride and sodium sulfate, SEB production occurred at an $a_w < 0.90$. When glycerol was used, the minimum a_w was 0.98–0.99.

Temperature and pH also affect the a_w at which *S. aureus* will grow and produce enterotoxin. When these parameters deviate from their optimum levels, the minimum a_w tolerated by *S. aureus* is elevated.

Atmospheric Conditions

Staphylococci are facultative anaerobes, but the amount and rate of growth and enterotoxin production are considerably less under anaerobic conditions than under aerobic conditions. Excessive aeration or oxygen levels can decrease the amount of SEB produced. At a dissolved oxygen (DO) level of 100% growth of *S. aureus* at 37 °C was maximal, but no SEB was produced. The optimal DO level for SEB production was 10%. In contrast, SEA production was related more to growth and not affected by the level of DO.

Staphylococcal growth and enterotoxin production have been observed in Canadian bacon and ham, and sausage, turkey, and hamburger sandwiches stored anaerobically. As with culture media, enterotoxin yields were higher under aerobic conditions.

Nutritional Factors

Most species of staphylococci require an organic nitrogen source and one or more B vitamins for

aerobic growth. Media containing protein digests generally enhance growth and enterotoxin production. For anaerobic growth, a fermentable carbon source and uracil are required.

Addition of readily fermentable carbon sources such as glucose or pyruvate to aerobic cultures repressed enterotoxin synthesis by *S. aureus*. The repression is partially attributed to decreased pH due to acid production, and partly to catabolic repression.

Sodium Chloride

One characteristic of *S. aureus* is its ability to survive and grow in relatively high salt concentrations. The organism can grow in up to 20% sodium chloride, whereas enterotoxin production occurs with up to 10% sodium chloride. Growth and enterotoxin production are generally retarded at increased salt concentrations.

Competing Microorganisms

Staphylococci are not good competitors, especially when the inoculum is small compared with that of the other organisms present. Many common food bacteria have been shown to inhibit growth of *S. aureus* and/or its ability to produce enterotoxin. In raw foods, *S. aureus* will not grow appreciably because of the presence of other organisms. An exception would be in the case of raw milk from a cow with staphylococcal mastitis. The high numbers of staphylococci present would be able to overcome other microorganisms. In heat-processed foods, contaminating *S. aureus* would have a competitive edge, especially if the food contains salt and has a reduced a_w.

Survival in Foods

Staphylococcal food poisoning results from the ingestion of enterotoxin produced by *S. aureus* growing in the food. Raw foods such as meat and milk are frequently contaminated with staphylococci. However, they are seldom involved in food poisoning because other contaminating microorganisms inhibit their growth and enterotoxin production. In addition, not all staphylococcal isolates are enterotoxigenic.

Heat is the most effective way to inactivate *S. aureus* in food. Heating meat to an internal temperature of 73.9–76.7 °C will kill any staphylococci present. Temperatures normally used for cooking meats should be sufficient to inactivate *S. aureus*. The time–temperature treatments used to pasteurize milk are adequate to destroy the organisms. The *D* values of *S. aureus* in skim milk at 60.0 and 65.5 °C are 3.44 and 0.28 min, respectively. The enterotoxins, however, are very heat-resistant, and are not inactivated by pasteurization. An outbreak occurred in 1985 from chocolate milk served to schoolchildren. The milk had been held inadvertently for several hours at a warm temperature before pasteurization. *S. aureus* present in the milk was able to grow and produce SEA. No viable staphylococci were present in the pasteurized milk, but SEA was detectable. (*See* **Heat Treatment**: Chemical and Microbiological Changes.)

Temperatures used in commercial canning are sufficient to destroy staphylococci and the amount of enterotoxin usually present in foods involved in food poisoning outbreaks (1 ng to > 50 ng per gram of food). There have been outbreaks associated with the consumption of canned corned beef, but these were traced to recontamination of improperly sealed cans after processing. Two people were reported to become ill after eating lobster bisque that had been heated at 118 °C for 86 min, but this was due to inadequate processing of some of the cans. In general, the heat stability of the enterotoxins is greater in foods than in buffer. The degree of inactivation depends on a variety of factors, including the nature of the food, pH, and concentration and type of enterotoxin. (*See* **Canning**: Principles.)

S. aureus inoculated on to frankfurters was inactivated when they were heated to an internal temperature of 71.1 °C in the smoking procedure. *S. aureus* in the interior of ham that survived the curing process was slowly inactivated during heating in a smokehouse at 48.9 °C for 48 h. (*See* **Curing**.)

Freezing and thawing have no significant effects on the viability of *S. aureus*. However, prolonged storage at subfreezing temperatures reduces the number of *S. aureus* in meats. The population of *S. aureus* in raw minced beef was reduced by 91% after storage at −22 °C for 4 months.

S. aureus is relatively resistant to drying. Nonfat dry milk and foods containing nonfat dry milk have been implicated in several staphylococcal food poisoning outbreaks. Staphylococci present in the milk may survive spray drying, depending on the temperature, the moisture content of the product, and the strain of *S. aureus*.

See also: **Canning**: Principles; **Curing**; **Heat Treatment**: Chemical and Microbiological Changes; **Meat**: Preservation; **Milk**: Processing of Liquid Milk; **pH – Principles and Measurement**; **Poultry**: Ducks and Geese; Turkey; **Water Activity**: Effect on Food Stability

Further Reading

Bergdoll MS (1989) *Staphylococcus aureus*. In: Doyle MP (ed.) *Foodborne Bacterial Pathogens*, pp. 463–523. New York: Marcel Dekker.

Jablonski LM and Bohach GA (2001) *Staphylococcus aureus*. In: Doyle MP, Beuchat LR and Montville TJ (eds) *Food Microbiology*, pp. 411–434. Washington, DC: American Society for Microbiology.

Martin SE and Myers ER (1994) *Staphylococcus aureus*. In: Hui YH, Gorham JR, Murrell KD and Cliver DO (eds) *Foodborne Disease Handbook*, vol. 1, pp. 345–394. New York: Marcel Dekker.

Detection

M S Bergdoll[†] **and A L Wong**, University of Wisconsin, Madison, WI, USA

Introduction

The fact that staphylococcal food poisoning is caused by the ingestion of enterotoxins produced by staphylococci requires that methods be available for detection of both the organisms and the enterotoxins. Although it is essential that staphylococci are present and grow in the food to produce enterotoxin at some point in the preparation or processing of the food, their presence in the food is not proof that enterotoxin is present. Not all staphylococci are capable of producing enterotoxin. On the other hand, the absence of staphylococci in food does not prove that enterotoxins are absent because these organisms are easily destroyed by heat, whereas the enterotoxins are not. Detection of enterotoxins is most important because, if they are present at detectable levels in a food, ingestion of the food most likely will result in food poisoning.

Detection of Staphylococci in Foods

The isolation of coagulase-positive staphylococci from foods is of greatest concern because these are the types most often involved in food poisoning, although occasionally coagulase-negative staphylococci have been implicated. Coagulase-positive staphylococci include not only *Staphylococcus aureus* but also *S. intermedius* and *S. hyicus*; some strains from each of these species are known to produce enterotoxin. Hence, coagulase production by suspected food-poisoning isolates is one of the most important properties tested for. Thermonuclease (TNase) production is also commonly associated with food-poisoning staphylococci and is another important characteristic of these species.

Detection in Raw Foods

Normally, a 50-g sample of the food is used for detection of the presence of staphylococci. The sample should be as representative of the food from which it is taken as possible. Frozen samples should be thawed under refrigeration just before testing. All sampling should be done under sterile conditions and samples should remain under sterile conditions and kept refrigerated until tested.

The food samples are suspended in or mixed with a suitable diluent, plated on Baird–Parker agar and incubated. One or more colonies of each colony type are selected from a plate containing 20–200 colonies for coagulase determination. Any colonies testing negative for coagulase should be tested for TNase. Any colonies positive for either coagulase or TNase can be considered candidates for enterotoxin production.

Detection in Processed Foods

Collection of processed foods and preparation of the dilutions are the same as for the raw foods. The diluted samples are incubated in double-strength trypticase soya broth before single-strength trypticase soya broth containing 20% sodium chloride is added and the incubation continued. The cultures are plated on Baird–Parker agar and the same procedures as described above are used for testing colonies. This procedure can be used for raw or nonprocessed foods suspected of containing < 100 *S. aureus* organisms per gram in the presence of large numbers of competing organisms.

Detection in Cases of Food Poisoning

In most cases of food poisoning the food will have been contaminated after it was cooked or heated. In these cases the same procedures used for isolation of staphylococci from raw foods are employed. In those cases where the staphylococci have grown in the foods and produced enterotoxin before the food was processed and in those cases where the food history is not known, the procedure for processed foods is used. In some cases the staphylococci have either been killed or else they died during storage. Because TNase is heat-stable and survives in food for long periods of storage, testing for it can be undertaken to indicate whether sufficient staphylococcal growth had occurred to produce enterotoxin.

If staphylococci are isolated from the food they should be tested for enterotoxin production. If they

[†]Deceased.

produce enterotoxin, the food can be tested for the type of enterotoxin produced by them. In the case where staphylococci cannot be isolated but the TNase test is positive, the foods must be tested for enterotoxin because it is not known whether the staphylococci that grew in the food were capable of producing enterotoxin.

Phage typing of the staphylococci isolated from the food is done in order to aid in the identification of the source of contamination. Any staphylococci isolated from an individual who handled food that was implicated in a food-poisoning outbreak should be phage-typed. If the isolates from the food and the individual have the same phage pattern, it can be concluded that the individual was the source of the contamination. (*See* **Food Poisoning:** Tracing Origins and Testing.)

Detection of the Enterotoxins

All of the methods for the detection of the enterotoxins are based on the use of specific antibodies to the enterotoxins. Detection of the enterotoxins is complicated by the fact that seven enterotoxins have been identified – enterotoxins A (SEA), B (SEB), C_1 (SEC_1), C_2 (SEC_2), C_3 (SEC_3), D (SED), and E (SEE). Five specific antibodies are needed for their analysis; the SECs can be detected by one antibody. Unidentified enterotoxins do exist, but their involvement in food poisoning is minor. Antibodies prepared in animals such as rabbits are polyclonal and react with enterotoxins in gels to give precipitin reactions. Monoclonal antibodies cannot be used in gels because their reactions with the enterotoxins do not result in the formation of precipitates.

The detection of enterotoxin in foods requires methods that are much more sensitive than those required for determination of the enterotoxigenicity of strains. The quantity of enterotoxin present in foods involved in food-poisoning outbreaks may vary considerably, from less than 1 ng g^{-1} to greater than 50 ng g^{-1}. Usually, little difficulty is encountered in detecting the enterotoxin in foods involved in staphylococcal food-poisoning outbreaks. However, outbreaks do occur in which the amount of enterotoxin is less than 1 ng per gram of food. In such cases, the enterotoxin can only be detected by the most sensitive methods. Another situation in which it is essential to use a very sensitive method is in determining the safety of a food for consumption.

Gel Diffusion Methods

Many types of gel reactions have been used in the detection of the enterotoxins; the most common ones are some form of either the Ouchterlony gel plate or the microslide (Figure 1). These methods have been used widely in the determination of the enterotoxigenicity of staphylococcal strains. The modification of the Ouchterlony gel plate test that is used in the Food Research Institute, University of Wisconsin, and recommended to others is the optimum sensitivity plate (OSP) method (Figure 2). It is easy to use and, in conjunction with production of the enterotoxins by the membrane-over-agar method (Figure 3) or the sac culture method, is of adequate sensitivity to detect most enterotoxigenic staphylococci. In the sac culture method, the medium is inside a dialysis sac that is placed in an Erlenmeyer flask with the inoculum in buffer outside the bag; incubation is with shaking. The normal sensitivity of the OSP method is 0.5 μg ml^{-1}, but can be increased to 0.1 μg ml^{-1} by a fivefold concentration of the staphylococcal culture supernatant fluids. The microslide is used by some investigators, but care is needed in preparing the slides; even so, the results are often difficult to interpret. Experience is important in achieving its maximum sensitivity (50–100 ng ml^{-1}), but even with experience many individuals are unable to achieve this sensitivity.

The original methods for the detection of enterotoxin in foods employed the microslide as the test method for detection of the enterotoxins. These methods required the use of 100 g of food and extraction and concentration procedures to reduce the extract to 0.1–0.2 ml. These were cumbersome and time-consuming procedures which have been outdated by more sensitive methods for the detection of the enterotoxin in the extracts. The US Food and Drug Administration personnel still use the original method developed in 1965 and consider it the official method for the detection of enterotoxins in foods.

Sensitive Detection Methods

Development of sensitive methods for the detection of proteins at less than 1 ng per milliliter of fluid greatly simplified the detection of enterotoxins in foods. Thus, it was possible to use a simple procedure for extracting the enterotoxin from foods. The one used in the Food Research Institute, University of Wisconsin, is an example: (1) grind the food to a homogeneous slurry with 1–1.5 ml of fluid per gram of food; (2) adjust the pH to 4.5 and centrifuge; (3) adjust the supernatant fluid to pH 7.5 and centrifuge if necessary; (4) extract with chloroform if fat interferes, centrifuge, and filter. Use of simple extraction procedures not only greatly shortens the time required to prepare the sample for enterotoxin detection (from 2–3 days to 1–2 h), but also improves the recovery of the enterotoxin from the food. This is particularly important in cases where very low amounts of enterotoxin ($\leq$ 1 ng g^{-1}) are present. In these cases one can

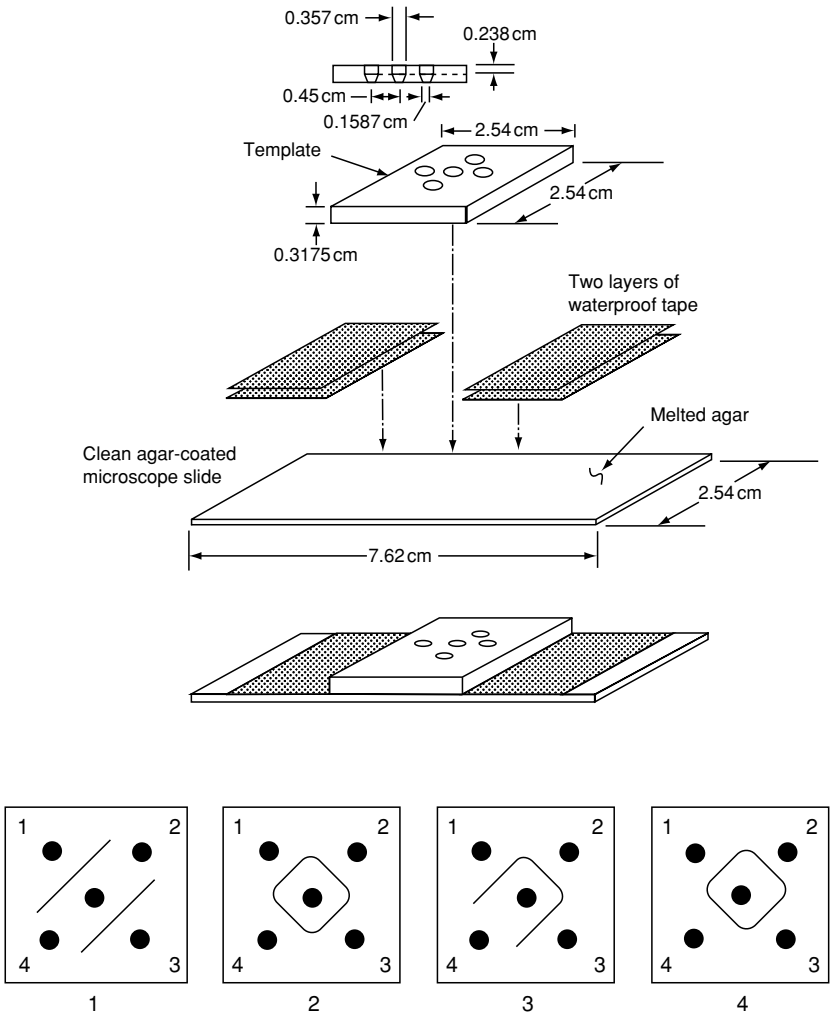

Figure 1 Microslide gel diffusion assembly showing typical results with the microslide test. The antibody is placed in the center well and the control enterotoxin is placed in wells 1 and 3. Slide 1, positive reactions with control toxin; slide 2, positive reactions with unknowns in wells 2 and 4; slide 3, positive reaction with unknown in well 2; reaction when unknown in well 4 contained a larger amount of toxin and unknown in well 2 contained less toxin than the unknown in slide 3. Reproduced from *Staphylococcus*: Detection, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

expect recovery of at least 50% of the enterotoxin but, in the long extraction and concentration procedures, less than 10% will be recovered.

Enzyme-Linked Immunosorbent Assay (ELISA)

The ELISA methods were applied to the detection of the enterotoxins in foods soon after they were originally developed for the detection of other proteins. Essentially, all of the ELISA methods used are of the sandwich type; in this type the enzyme is coupled to the antibody instead of the enterotoxin. The amount of enzyme and, thus, the color developed from the enzyme–substrate reaction, is directly proportional to the amount of enterotoxin present in the unknown sample. (*See* **Immunoassays:** Radioimmunoassay and Enzyme Immunoassay.)

Most users of ELISA methods employ microtiter plates to which the antibodies are attached. The large number of wells in a microtiter plate allow examination of several samples at one time, although there may not be uniformity in all of the wells, particularly those around the edge of the plate. A plate reader is necessary for recording the results, which adds to the expense of the method.

An alternative method is to use polystyrene balls to which the antibodies are attached. The ball method is more cumbersome because each ball must be handled separately. A relatively large volume of the unknown sample can be used, thus increasing the amount of enterotoxin adsorbed, and hence the sensitivity of the assay. The use of 1-ml volumes of substrate allows the color developed to be read in a simple colorimeter,

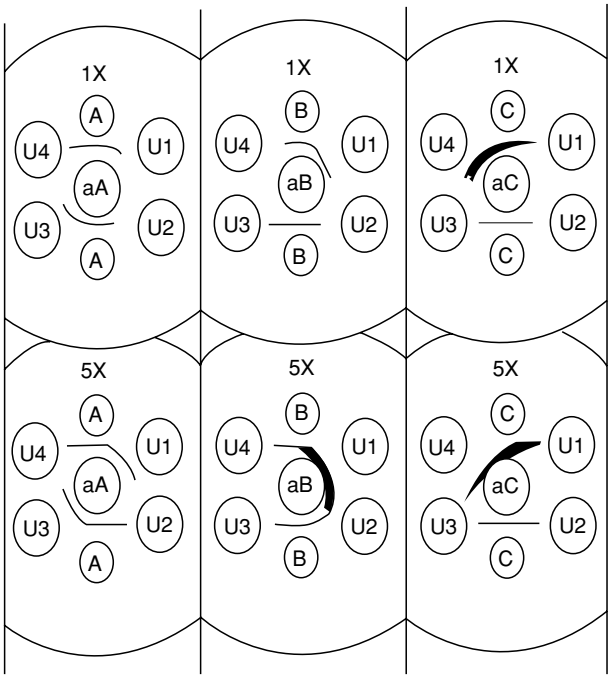

Figure 2 Optimum sensitivity plate (OSP) with typical results. The wells containing the control toxins are one-half the cross-sectional area of the antibody wells (middle) and the wells containing the unknowns (U1–U4). This results in a doubling of the sensitivity. Unknowns in wells U1 and U3 in the first place, U1 in plate 2, and U4 in plate 3 are enterotoxin-positive. The hook in U1 in plate 1 is confirmed by fivefold concentration of the unknown sample. The unknown in well 2 in plate 2 is not positive; the observed result is affected by the large amount of toxin in the unknown in well 1. Reproduced from *Staphylococcus*: Detection, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

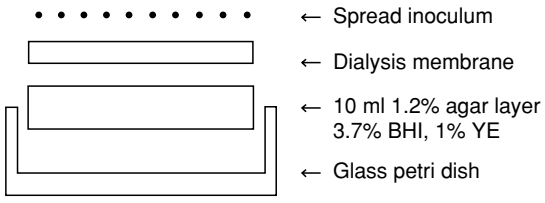

Figure 3 Membrane-over-agar enterotoxin production method. The organisms grow on the surface of the dialysis membrane as the nutrients in the medium (brain–heart infusion (BHI) plus yeast extract (YE)) in the agar can diffuse through the membrane, but the enterotoxin cannot pass through the membrane. The culture is removed from the membrane with buffer and centrifuged before analysis. Reproduced from *Staphylococcus*: Detection, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

an instrument that most laboratories would have available. The sensitivity of the ELISA methods is between 0.1 and 1.0 ng per gram of food (Table 1).

Table 1 Detection of staphylococcal enterotoxins (SEs) in foods by enzyme-linked immunosorbent assay

Food	*SE*	*Amount of SE added (ng g^{-1})*	*Amount of SE detected (ng g^{-1})*
Milk	A	0.63	0.63
Ham	A	0.63	0.34
		1.25	0.55
Genoa sausage	A	0.63	0.36
Cheese	A	0.63	0.59
	D	0.63	0.15
		1.25	0.64
	E	0.63	0.38
Cheese food	A	0.63	0.36
Potato salad	B	0.63	0.18
		1.25	0.44
Spaghetti	A	0.63	0.54

Reproduced from Freed RC, Evenson ML, Reiser RF and Bergdoll MS (1982) Enzyme-linked immunosorbent assay for the detection of staphylococcal enterotoxins in foods. *Applied and Environmental Microbiology* 44: 1349–1355.

Although polyclonal antibodies are used in most of the enterotoxin detection methods, the development of monoclonal antibodies to the enterotoxins has made possible their use in enterotoxin analysis. Two monoclonal antibodies are required for each enterotoxin in the sandwich ELISA as only one site exists on the molecule for each monoclonal antibody. Not all monoclonal antibodies are usable, particularly as the coating antibody, because the reacting site may be inactivated in the coating reaction.

In some instances it may not be necessary to determine the type of enterotoxin, only if enterotoxin is present; for example, in examining the marketability of suspect foods. The food would not be marketable if any enterotoxin were present. For this purpose, including all of the enterotoxins in one test saves time as only one test is needed. This would not save time in examining foods implicated in food-poisoning outbreaks, because identification of the type of enterotoxin is valuable in tracing the source of the contamination. However, if more than one enterotoxin was present, the amount of each could be below detectable levels for the individual enterotoxins. In this case it would not be possible to detect them by the methods used to identify individual enterotoxins.

Reversed Passive Latex Agglutination (RPLA)

In the RPLA method the specific antibodies are attached to latex particles. When these particles are added to a solution containing enterotoxin, the latex particles agglutinate. The sensitivity is less than 1 ng per gram of food. The method is adequately sensitive for the detection of enterotoxin in most foods implicated in food-poisoning outbreaks; however, it may be inadequate for detection of the small amounts

Table 2 Detection of staphylococcal enterotoxin (SE) in foods from outbreaks

Food	S. aureus *count* g^{-1}	*SE detected in food*			
		SE strain	*ELISA plate*	*ELISA kit*	*RPLA kit*
Ham	1.5×10^9	A	A	A	A
Lasagne, dried	2.0×10^8	A	A	A	A
Salmon, canned	3.5×10^6	A	A	A	A
Sheep's milk cheese	ND		A	A	NSA
Corned beef	4.0×10^7	A, B	A, B	A, B	A, B
Smoked bacon spread	1.0×10^9	A, D	A[a]	A, D	A, D
Beef rolls	4.0×10^6	A, D	A[a]	A	ND
Pork	6.0×10^9	B	B	B	B
Chicken	1.0×10^6	C		C	ND
Meat pies	1.0×10^6	A	ND	ND	ND

Reproduced from Wieneke AA and Gilbert RJ (1987) Comparison of four methods for the detection of staphylococcal enterotoxin in foods from outbreaks of food poisoning. *International Journal of Food Microbiology* 4: 135–143.
ND, not detectable; NSA, nonspecific agglutination; ELISA, enzyme-linked immunosorbent assay; RPLA, reversed passive latex agglutination.
[a]Extract not tested for SED.

of enterotoxin that are sometimes present (Table 2). One problem with the RPLA method is that the food extract from an occasional food may give a nonspecific agglutination.

Use of Sensitive Methods for Strain Testing

More recently, a question has arisen regarding the sensitivity of the gel diffusion methods for determining the enterotoxigenicity of staphylococcal strains. It has been reported that enterotoxin production by strains was observed by the RPLA method that was not detectable by the OSP method. The amounts produced were approximately 10–20 ng ml^{-1}. This was confirmed by concentrating the culture supernatant fluid from five strains that tested positive for SEA by ELISA about 100-fold for testing by the OSP method. The importance of this low production may be questioned; however, we have found strains that were implicated in food-poisoning outbreaks to be negative by OSP but positive by ELISA. This resulted from examination of strains by ELISA that were negative by OSP but positive by the monkey feeding test. A number of these strains were positive for one or more of the identified enterotoxins by the ELISA method, particularly for SED (23 strains). Some of these strains had been isolated from food-poisoning outbreaks, which is significant because SED has been implicated as the second most important enterotoxin in food poisoning. It should be pointed out that, of all of the enterotoxins, SED is produced in the smallest amounts. Only three of the strains produced low amounts of SEA, the enterotoxin implicated in 75% of staphylococcal food-poisoning outbreaks. The production of 10–20 ng of enterotoxin per milliliter is probably of significance because only 100–200 ng of SEA has been shown to be necessary to produce food poisoning. The amount present in the vehicle, 2% chocolate milk, was 0.50–0.75 ng ml^{-1}. Admittedly, the amounts of enterotoxin produced by the membrane-over-agar or sac culture methods are 5–10 times those produced in shake flasks or even possibly in food, yet, if growth is sufficient, 10^7 colony-forming units (CFU) and 1–2 ng of enterotoxin per gram of food can be produced. This would be enough to result in food poisoning in sensitive individuals.

Kits Available for Enterotoxin Detection

Several commercial kits are available for enterotoxin detection:

1. An ELISA ball kit, available from Labor Dr W. Bommeli, Länggassstrasse 7, CH-3012 Bern, Switzerland. In this kit the specific antibodies to SEA, SEB, SEC, and SED are absorbed on to separate polystyrene beads and the enzyme alkaline phosphatase is coupled to the specific antibodies. The substrate, *p*-nitrophenyl phosphate, gives a yellow color with the enzyme. Although the test can be completed in 1 day, recommendations are that the antibody-coated balls are shaken with 20 ml of extract overnight to obtain the highest sensitivity (0.1–1.0 ng ml^{-1}). The color can be measured in a colorimeter because 1 ml of substrate is used. The method can be used for quantitative measurements.
2. An ELISA dipstick kit, available from Transia, 8 Rue Saint Jean de Dieu, 69007 Lyon, France. Monoclonal antibodies to SEA, SEB, SEC, SED, and SEE are coated on nitrocellulose paper in wells in the dipstick. The conjugate consists of different monoclonal antibodies coupled to horseradish peroxidase, which reacts with the substrate

to give a blue color. The method is sensitive to 0.5–1.0 ng per gram of food; the test can be completed in 6–7 h. The results are read visually; only one test is needed per sample.

3. An RPLA kit, available from Oxoid Ltd, Wade Road, Basingstoke, Hampshire RG24 0PW, UK. Latex particles are coated with specific antibodies to SEA, SEB, SEC, and SED. More than 24 h is required for completion of the test.
4. An ELISA tube screening kit, available from Transia. Monoclonal antibodies to SEA, SEB, SEC, SED, and SEE are coated on to the bottom of a small flanged tube. The conjugate is a mixture of monoclonal antibodies coupled to horseradish peroxidase. The enzyme reacts with the substrate to give a blue color. The sensitivity of the test is less than 0.2 ng ml^{-1}. The time required for an assay is 1 h.
5. An ELISA microtiter plate screening kit, available from Biotechnology Australia Pty Ltd, PO Box 20, Roseville, NSW 2069, Australia. A mixture of specific antibodies to SEA, SEB, SEC, SED, and SEE is adsorbed to the wells of microtiter plates. The enzyme used is horseradish peroxidase and the substrate is ABTS (a sulfonic acid), which gives a green color. The sensitivity of the method is less than 1 ng ml^{-1} and the time required is 4 h. An additional step of treating the food extracts with urea followed by concentrating 20-fold is included in the procedures in this kit for recovery of enterotoxin from heated foods. Supposedly, the purpose of the urea is to renature any denatured enterotoxin in the heated food, but it is doubtful that this is possible. In any case, the denatured enterotoxin would be digestible by pepsin in the stomach and would not cause food poisoning. However, the concentration procedure does increase the sensitivity of the method, but whether this additional sensitivity is necessary is questionable.

See also: **Food Poisoning**: Tracing Origins and Testing; **Immunoassays**: Radioimmunoassay and Enzyme Immunoassay

Further Reading

Apfalter P, Assadian O, Kalczyk A *et al.* (2002) Performance of a new chromogenic oxacillin resistance screen medium (Oxoid) in the detection and presumptive identification of methicillin-resistant *Staphylococcus aureus*. *Diagnostic Microbiology and Infectious Disease* 44(2): 209–211.

Bergdoll MS (1979) Staphylococcal intoxications. In: Riemann H and Bryan FL (eds) *Food-Borne Infections and Intoxications*, 2nd edn, pp. 443–494. New York: Academic Press.

Bergdoll MS (1989) *Staphylococcus aureus*. In: Doyle MP (ed.) *Foodborne Bacterial Pathogens*, pp. 463–523. New York: Marcel Dekker.

Bergdoll MS (1990) Analytical methods for *Staphylococcus aureus*. *International Journal of Food Microbiology* 10: 91–100.

Gurran C, Holliday MG, Perry JD, Ford M, Morgan S and Orr KE (2002) A novel selective medium for the detection of methicillin-resistant *Staphylococcus aureus* enabling result reporting in under 24 h. *Journal of Hospital Infection* 52(2): 148–151.

Tatini SR, Hoover DG and Lachica RVF (1984) Methods for the isolation and enumeration of *Staphylococcus aureus*. In: Speck ML (ed.) *Compendium of Methods for the Microbiological Examination of Foods*, pp. 411–427. Washington, DC:American Public Health Association.

Food Poisoning

M S Bergdoll[†] **and A L Wong**, University of Wisconsin, Madison, WI, USA

Introduction

Staphylococcal food poisoning is one of the most common types of foodborne disease and probably occurs in every country of the world, although records of the milder foodborne diseases are not kept in most countries. Even when the illnesses are reported, inadequate investigation or laboratory examination frequently prevents proper diagnosis. The illness is due to the ingestion of enterotoxins present in the food as a result of enterotoxigenic staphylococci growing in the food. Recovery is rapid; hence, in most cases individuals do not consult a doctor, but conclude it was something they ate. It is not a reportable disease in the USA; however, enough identifiable outbreaks are reported to the Centers for Disease Control (CDC) to result in its listing as one of the leading foodborne diseases. During the period 1983–1987, 600 confirmed bacterial foodborne disease outbreaks were reported with *Staphylococcus aureus*, responsible for 47 (7.8%) of the outbreaks and 3181 (6.3%) of the cases, which probably represents only a small percentage of the actual number of staphylococcal outbreaks. Many cases of foodborne

[†]Deceased.

illness with symptoms similar to those of staphylococcal food poisoning are reported to the CDC, but are not classified because of inadequate information about the presence of staphylococci or enterotoxin in the foods involved. (*See* **Food Poisoning**: Tracing Origins and Testing.)

Staphylococcal foodborne outbreaks vary in size, and can involve one or two persons to as many as over 1000 people. One outbreak in the USA involved 1300 people attending a picnic in Indiana. Baked ham was the implicated food. An even larger outbreak occurred in Japan, involving 1500 people who became ill from eating contaminated rice balls contained in lunch boxes prepared at one location. (*See* **Food Poisoning**: Statistics.)

Symptoms of the Disease

The major symptoms of staphylococcal food poisoning, i.e., vomiting and diarrhea, occur within 1–6 h after ingestion of food containing enterotoxin. A representative set of symptoms observed in 122 cases of illness that occurred among high-school students after they had eaten cream puffs is given in Table 1.

Ninety-five (57%) of the 165 students who ate the cream puffs became ill, whereas none of those who did not eat them became ill. This is typical of outbreaks involving a large number of people. The average onset time in the outbreak was also normal for large outbreaks, with a few having earlier onset times, usually because of consumption of large amounts of toxin or increased sensitivity of some individuals, and a few that have delayed onset times of 7–8 h. Although vomiting is the symptom most frequently observed, some individuals may have other symptoms, such as diarrhea without vomiting. Diarrhea may not be observed in milder cases, whereas vomiting usually is. Changes in temperature and blood pressure are not common, particularly with milder cases, although it is known that in experimental cases a change in temperature can occur, either up or down; a drastic lowering of blood pressure may also occur.

Death is uncommon with this type of food poisoning, but deaths have occurred in children or older individuals, often with other complications. Two children, 3 and 4 years of age, died within 24 h after each drank 125 ml of milk from a goat with staphylococcal mastitis. The deaths of the children and the results of the autopsy indicated that the children received a relatively large amount of enterotoxin in the milk they drank.

Several deaths of older people who consumed food containing enterotoxin have been reported; the deaths usually occurred after severe vomiting and diarrhea. One elderly woman died of acute vascular collapse secondary to fulminating gastroenteritis in an outbreak of staphylococcal food poisoning after eating baked ham. Of 28 other people who were known to have become ill with symptoms characterized by vomiting, diarrhea, and severe prostration, six individuals required hospitalization.

A 57-year-old woman reportedly in excellent health died of shock within a few hours of the onset of illness after eating a cold plate lunch at a restaurant. Three other individuals developed severe gastroenteritis after eating plate lunches at the same restaurant.

Table 1 Symptoms of staphylococcal food poisoning from 122 cases[a]

Symptom	*Reaction*			
	Cases[b]	*None*	*Mild*	*Severe*
Vomiting	122	15	12	95
Pains in abdomen	122	6	40	76
Diarrhea	103	13	75	15
Headache	101	29	59	13
Muscular cramping	113	41	58	14

[a]Includes 94 students, eight lunchroom supervisors who took cream puffs home, and 20 cases resulting from the consumption of cream puffs sold at three cafés and from a bakery truck.
[b]Cases from which this information was available.
Data from Dennison GA (1936) Epidemiology and symptomatology of a *Staphylococcus* food poisoning. *American Journal of Public Health* 26: 1168–1175.

Requirements Necessary for Food Poisoning

Certain conditions are necessary for staphylococcal food poisoning to occur: (1) the food must be a good medium for the growth of staphylococci and the production of enterotoxin; (2) enterotoxigenic staphylococci must be present in the food; (3) the food must be held at a warm temperature for several hours for the staphylococci to grow to sufficient numbers to produce enterotoxin.

Foods Involved in Outbreaks

A variety of foods can support growth of enterotoxigenic staphylococci. Foods involved in outbreaks are usually rich in protein, and include meat and meat products, poultry, milk and dairy products, salads (tuna, chicken, ham, and potato), custards, pudding, and cream-filled bakery products. (*See* **Cream**: Types of Cream; **Milk**: Processing of Liquid Milk; **Poultry**: Chicken; Ducks and Geese; Turkey.)

Source of Contamination

Staphylococci can be found in most raw foods. However, the most frequent source of contamination of foods implicated in staphylococcal food-poisoning outbreaks is the food handler. Most foods involved in such incidents are cooked, and are recontaminated in the final preparation for serving. A food handler with any type of staphylococcal infection or one who is carrying the organism in his or her nose is a likely source of contamination.

An outbreak occurred in California in 1983 involving several hundred children in an Easter-egg hunt. The source of the contamination was an infected lesion on the hands of the person who prepared the eggs. Another outbreak involved packaged sliced ham that was contaminated with staphylococci from the nose of the person who packaged it. The ham was packaged on a Saturday morning and sold throughout the day without any of the consumers getting sick. The unsold ham was stored in a relatively deep pan and refrigerated overnight. Cooling might have been quite slow because of the size of the pan. On Sunday morning, the ham was warmed before being sold around 11:00 a.m. People who bought the ham on Sunday became ill with staphylococcal food poisoning. Enterotoxin A (SEA) was isolated from the leftover ham.

Animals carry staphylococci in their noses and certain animal diseases, such as mastitis, are caused by staphylococci. However, animals are less important sources of contamination because, even though raw meats and milk are usually contaminated with staphylococci, cooking and pasteurization temperatures will destroy the organisms. In addition, the presence of other competing microorganisms in the raw foods inhibits staphylococcal growth. An exception is when staphylococci are present in much higher numbers than other organisms, as in the case of milk from a mastitic animal, e.g., the two children who died after drinking milk from a goat with mastitis.

Equipment can sometimes be a source of contamination. An outbreak associated with baked ham resulted from contamination of the ham with the meat slicer. The same strain of enterotoxigenic *S. aureus* was isolated from both the ham and the slicer.

Conditions Necessary for Growth and Toxin Production

In order for food poisoning to occur, the food containing an enterotoxigenic staphylococcal strain must be held at a warm temperature, usually room temperature or above, for several hours (at least 4 h) for the staphylococci to grow to sufficient numbers to produce enterotoxin. Growth of staphylococci to a population of 1 million or more organisms per gram of food is assumed necessary to produce enough enterotoxin to cause illness.

Care needs to be taken in interpreting the significance of staphylococcal numbers in food when investigating an outbreak. Not all staphylococcal strains produce enterotoxin. A large number of staphylococci in a food may indicate poor hygienic practices but, to incriminate a food definitely as the cause of food poisoning, it is necessary to demonstrate that the staphylococcal isolate produces enterotoxin. Alternatively, the lack of staphylococci or the presence of low numbers does not indicate the absence of enterotoxin in the suspect food. This is especially important in the case of heat-processed foods. Absence of viable staphylococci does not necessarily indicate a safe product because preformed enterotoxin may not be inactivated.

Data from outbreaks indicate that less than 1 μg of SEA can result in illness. In a chocolate milk outbreak in 1985, it was found that 100–200 ng of SEA was sufficient to cause staphylococcal food poisoning in children. SEA is the most toxic enterotoxin and the one most commonly involved in staphylococcal food-poisoning outbreaks.

Site of Action

Emetic Action

The site of the emetic action of enterotoxin in the monkey is in the abdominal viscera, with the sensory stimulus for emesis reaching the vomiting center of the brain via the vagus and sympathetic nerves. This is supported by experiments in which transthoracic vagotomy partially protected the monkey from the emetic action of enterotoxin given intragastrically and, when coupled with abdominal sympathectomy, the monkey was completely resistant to the emetic action of enterotoxin. The monkeys with deafferentated abdominal viscera still vomited as normal animals when the emetic compound, Veriloid, was administered. Bilateral destruction of the area postrema on the floor of the fourth ventricle (chemoreceptor trigger zone) made the monkeys completely refractory to the emetic action of enterotoxin given intravenously or intragastrically.

Attempts to determine the site of action of enterotoxin in the intestinal tract have been unsuccessful. Specific binding of enterotoxin to cells closely related to intestinal epithelial cells that have been cited as potential target cells for direct cytopathic effects of the staphylococcal enterotoxins was not detectable. Binding of SEA in the rat intestinal tract was not

detectable, but rather the enterotoxin passed through the gastrointestinal mucous membrane within 15 min after administration and was detected in the kidney. Neuronal binding of SEA in the intestinal tract was not demonstrated; this binding might have been expected based on the finding that the sensory stimulus for vomiting reaches the vomiting center by way of the vagus and sympathetic nerves.

Diarrheal Action

Diarrhea is the second most common symptom seen in staphylococcal food poisoning. In outbreaks involving a large number of people, as many individuals develop diarrhea as have an emetic reaction. Stimulation of this reaction is not understood because the mechanism of action is apparently quite different from that of the so-called diarrheal diseases such as cholera and *Clostridium perfringens* and *Bacillus cereus* food poisoning. The toxins responsible for the diarrheal diseases elicit an ileal loop reaction in rabbits, whereas the staphylococcal enterotoxins do not. Enterotoxin B (SEB) had no effect on the absorptive mechanism of the rat small intestine, but the enterotoxin somehow triggered a secretory mechanism that resulted in a net secretion of fluids. The actual mechanism responsible for stimulating excretion has not been elucidated. (*See* ***Bacillus***: Food Poisoning; ***Clostridium***: Food Poisoning by *Clostridium perfringens*.)

Enteritis

Enteritis is another effect of enterotoxins taken orally, but it is not easily observed except by special examination. Pseudomembranous enterocolitis was observed to be associated with staphylococcal infections in patients who had been administered antibiotics to control infection during surgery. Many of these patients died before the cause of their illness was diagnosed. The staphylococci had become resistant to the antibiotics used and grew prolifically in the intestinal tracts of these patients because the normal flora had been eliminated by the antibiotics. These staphylococci were potent producers of enterotoxin, later identified as SEA and SEB. Experiments in animals with SEB revealed that enteritis can be produced in monkeys, dogs, and chinchillas, with the severity of enteritis depending on the amount of enterotoxin given intragastrically. Autopsy of dogs 24 h after ingesting SEB revealed acute enteritis with edema, hyperemia, and ulcerations of the mucosa, and destruction of the surface of the epithelium. Severe round cell infiltration of the mucosa and destruction of the surface of the epithelium were observed by microscopic examination. These observations in animals were similar to changes characteristic of acute exogenous gastritis in humans resulting from staphylococcal food poisoning. These findings, determined by gastroscopic examination, include patchy mucosal hyperemia, regional edema, muscular irritation, erosions, petechiae and purulent exudate. Examination 48 h later revealed the stomachs had returned to normal. Acute gastroenteritis was observed in monkeys 2 h after intragastric administration of SEB, with the most severe changes occurring at 4–8 h postinoculation. The mucosa was again normal after 72 h postinoculation.

Death

Death is uncommon in staphylococcal food poisoning, hence there is little opportunity to observe the pathological effects of enterotoxin. However, autopsy of the two children who died after drinking milk from a mastitic goat revealed a moderate amount of pulmonary edema marked with congestion of the alveolar vessels and, in a few areas, there appeared to be hemorrhages into the alveolae. There was a small amount of leukocytic infiltration in the periportal areas of the liver. These findings are similar to those observed in monkeys that received a lethal dose of enterotoxins by intravenous injection; in these animals there was a decrease in intravascular fluid volume with an increase in lung weight. The fluid was confined to the perivascular and peribronchial interstitial space; the lymphatics in these areas were gorged. The primary pathological change was degeneration and necrosis of capillary endothelial cells, with some but less frequent damage to the endothelium of venules. Also noted were interstitial hemorrhage and edema, a histolytic infiltrate, and a striking herniation of capillary endothelium into the vascular lumen. Apparently these changes were brought about by the enterotoxin binding to leukocytes with subsequent sequestering of toxin-bound leukocytes in the lungs.

Treatment and Course of Illness

The rapid development of the illness makes it impossible to prevent the symptoms from developing after the enterotoxin has been ingested; once the symptoms have occurred there is no treatment that will counteract them. In most cases recovery is within a few hours and no treatment is necessary; however, in severe cases where vomiting and diarrhea are excessive, treatment intravenously with fluids is beneficial, particularly to restore the salt balance.

A misunderstanding of the disease by the medical attendant can lead to unfortunate results. Such was the case when a young man afflicted with the illness was given antibiotics, with death as a result. If the disease was due to the ingestion of the organisms,

treatment with the proper antibiotic may be desirable but, because it is due to a toxin, antibiotic treatment is of no benefit. However, it could have an adverse effect, particularly if the staphylococci were resistant to the antibiotic. In this case the natural flora of the intestinal tract would be eliminated and this would allow the staphylococci to grow uninhibited and produce enterotoxin. Staphylococci are not good competitors and usually do not grow in the intestinal tract in the presence of the natural flora.

Prevention

Temperature Control

Most staphylococcal food-poisoning outbreaks could be prevented if proper temperature controls were observed during holding of the food after its preparation and before being served. One method is proper refrigeration of susceptible foods after they are prepared. The lower the temperature, the slower the growth of the staphylococci, with little or no growth at refrigeration temperatures. An excellent example of the effect of refrigeration is an outbreak from cream-filled coffee cake that occurred in Wisconsin in 1970. Approximately 50 people became ill after eating the cake, with differing severity of symptoms, depending on the length of time the cakes stood unrefrigerated before being eaten. The cakes were made by a bakery and delivered to several stores at 5:30 a.m. on a warm Saturday morning; the cakes were not refrigerated at the stores. Purchases were made all day Saturday and on Sunday morning. Many who bought the cakes on Saturday did not eat them until Sunday morning, but in most cases the cakes were refrigerated by the purchasers. Most of those who bought the cakes early Saturday morning and refrigerated them did not become ill. No one who ate cakes purchased at the bakery became ill because the cakes were refrigerated at the bakery. The source of the staphylococci was from the throat of the food handler who had placed the cream filling into the cakes with his hand. The organisms isolated from his throat and the cakes produced enterotoxin E; not many staphylococci produce this enterotoxin and its involvement in food poisoning is rare.

Some of the largest outbreaks of staphylococcal food poisoning have occurred at picnics, a number of these from baked ham sandwiches. Baked ham is an excellent medium for the growth of staphylococci, which is not always appreciated because it is a cooked food. However, if the ham is recontaminated after the baking, the results can be tragic. Providing refrigeration for such a large quantity of food is difficult at a large picnic and normally the food will be at the picnic for a number of hours before being consumed. In cases such as this, every effort should be made to provide refrigeration for the food at the picnic grounds or not to transport it to the grounds until shortly before it is to be served. (*See* **Pork**.)

One error that is frequently made is to place the food in large containers to conserve space before it is placed in the refrigerator. Under these conditions, cooling is very slow and allows sufficient time for the staphylococci to grow and produce enterotoxin. A staphylococcal food-poisoning outbreak, in which 104 of 231 inmates in a country jail in New York became ill, resulted from inadequate refrigeration of macaroni salad. It had been prepared the day before it was served and had been stored overnight in two large, deep containers in a walk-in cooler. Isolation of over 10^7 staphylococci per gram from the macaroni salad indicated inadequate cooling.

Food-Handling Practices

In the New York jail outbreak described in the previous section, it was ascertained that the food was prepared by inmates who were inadequately trained and not well supervised. Many food-poisoning outbreaks are a result of inadequate training and supervision of the individuals preparing the food. The major conclusion from both the First and Second National Conferences for Food Protection held in Denver, Colorado, in 1971 and in Washington, DC, in 1984 was the need to educate the food handler as well as the consumer on food safety. Unfortunately, very little has been done in this area.

Food-Handler's Infections

One method of preventing staphylococcal food-poisoning outbreaks is to prohibit food handlers with infections from handling foods. In some countries food handlers who are staphylococcal carriers are not allowed to handle foods as long as they are carriers. In the UK an assistant cook, who carried the same staphylococcal organism in her nose as the one isolated from lemon snow pudding involved in the illness of six adults and 101 children in a primary school, was excluded from work until the organisms were eradicated by treatment with gentamicin cream. The pudding contained large numbers of staphylococci ($1.5 \times 10^8\,g^{-1}$) that produced SEA and enterotoxin D, as did the staphylococci isolated from the nose of the assistant cook. Although it was feasible to exclude the cook in this case, usually it is difficult and impractical to bar workers who are carriers of staphylococci in their noses or throats from handling foods because approximately 50% of the population are carriers. On the other hand, food handlers with

any type of infection should not be permitted to handle foods, particularly if the food is for mass feeding. The only viable solutions to this problem are proper refrigeration of vulnerable foods between preparation and consumption, or maintaining the temperature above 45 °C.

See also: ***Bacillus***: Food Poisoning; **Cadmium**: Toxicology; ***Clostridium***: Food Poisoning by *Clostridium perfringens*; **Cream**: Types of Cream; **Food Poisoning**: Tracing Origins and Testing; Statistics; **Milk**: Processing of Liquid Milk; **Poultry**: Chicken; Ducks and Geese; Turkey

Further Reading

Bergdoll MS (1979) Staphylococcal intoxications. In: Riemann H and Bryan FL (eds) *Food-Borne Infections and Intoxications*, pp. 443–494. New York: Academic Press.

Bergdoll MS (1983) Enterotoxins. In: Adlam C and Easmon CSF (eds) *Staphylococci and Staphylococcal Infections*, vol. 2, pp. 559–597. London: Academic Press.

Bergdoll MS (1989) *Staphylococcus aureus*. In: Doyle MP (ed.) *Foodborne Bacterial Pathogens*, pp. 463–523. New York: Marcel Decker.

Dack GM (1956) Staphylococcus food poisoning. In: Dack GM (ed.) *Food Poisoning*, 3rd edn. Chicago: University of Chicago Press.

STARCH

Contents

Structure, Properties, and Determination

D S Jackson, University of Nebraska, Lincoln, NE, USA

Starch Structure

Amylose and Amylopectin Polymers

Starch is a mixture of two glucose polymers. These polymers are initially enclosed within a semicrystalline granule formed inside starch-synthesizing plant organelles. Amylopectin, the larger of the polymers, is an alpha-1, 4-linked, alpha-1,6-branched (4–6% branching) polymer with an average molecular weight near 10^8. Amylose, a smaller, mostly linear polymer, is also composed of alpha-1,4-linked glucose units; long chains are sometimes connected with alpha-1,6 branches, although branching probably accounts for < 1 % of the glucose unit connections. The molecular weight of amylose is approximately 10^5. The molecular weight of starch polymers varies depending upon its plant source. Amylopectins have been reported to vary from 50×10^6 to 500×10^6, and amyloses from 1600 to 1×10^6. The degree of branching also varies. Common starches typically have 25–35% amylose and 65–75% amylopectin.

Genetic Differences

Several genetically modified starches are available. The most common forms are the waxy starches and the high-amylose starches. Waxy starch, which contains 100% amylopectin, is found in grains (i.e.,maize, sorghum, and wheat) that have three recessive waxy genes. Combinations of recessive (*wx*) and dominant (*Wx*) waxy genes result in starches with varying percentages of amylopectin from 75 (*WxWxWx*) to nearly 100% (*wxwxwx*). High-amylose (*ae*) starches are available with approximately 53 or 70% amylose.

Microscopy of Granules

Starch granules have unique microscopic structures (**Figures 1** and **2**), and the source of starch can frequently be determined by a trained observer. The shape and size of starch granules differ according to the botanical source and the environmental condition under which a crop was grown. Typically, rice and maize starch have angular (polyhedral) granules; potato starch has oval-shaped granules. Wheat starch

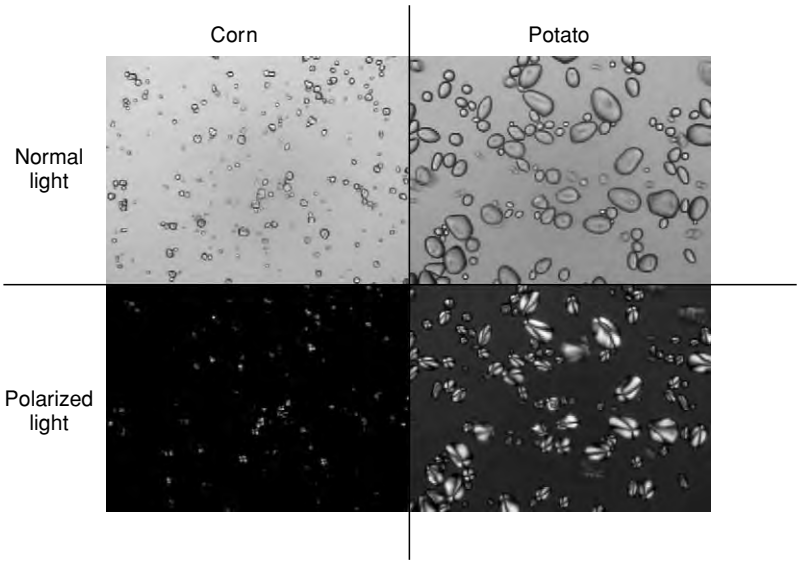

Figure 1 Light micrographs (× 100) of corn (left) and potato starch (right) shown using normal (top) and polarized (bottom) light microscopy.

consists of spherical and flat circular (lens)-shaped granules. Sizes also vary widely. Rice starch granules are very small (6 μm diameter), while potato starch granules can exceed 100 μm. The mean diameter of maize starch granules is 35 μm; wheat starch has a bimodal distribution of granular sizes – small (B) granules average 4 μm, while the large (A) granules average 14 μm. Granules can also have compound structures (i.e., oats, rice) which appear as separate starch grains in a single amyloplast enclosure.

Several microscopic techniques have found widespread use. When observed under polarized light, undamaged starch granules (as defined by the retention of crystalline structure) exhibit a phenomenon called birefringence (Figure 1). This birefringence is the result of polarized light being bent as it traverses a region of high molecular order. Scanning electron microscopy (Figure 2), as well as other higher-magnification techniques, have proven extremely useful in identifying the three-dimensional structure of starch granules, as well as showing the interrelationships of starch granules and other chemical components in cereal grains, tubers, and other food systems.

Changes during Hydration and Heating

Starch granules, when exposed to water, undergo swelling and upon heating undergo a process of irreversible swelling and crystalline melting called gelatinization. If the crystalline region is destroyed, granular birefringent properties disappear. The temperature at which birefringence is lost has been defined by some as the gelatinization point. Since loss of crystallinity causes loss of birefringence, additional thermal disruptions of the granule still occur after the gelatinization point as defined using microscopy techniques. When starch granules are exposed to liquid water below 40–50 °C, the amorphous gel-like portions of the starch granule absorb water, causing the granule to swell. When the temperature is > 40–50 °C, and water is in excess, granules begin to undergo reversible swelling and around 60–80 °C lose birefringence and undergo irreversible swelling. During the transition from reversible swelling to loss of birefringence, the amorphous regions have a rubber-like structure and the specific volume of the amorphous areas increases, as does the mobility of molecular segments which are thermally softened and plasticized by water. At the gelatinization temperature, the mobility of molecules increases rapidly. Granule components are in a more flowable state. The mobility of molecules in the amorphous regions and the mobility of molecular regions which traverse both crystalline and amorphous regions cooperatively contribute to intergranular stresses. The continued hydration of amorphous areas places additional stress on those molecular regions (including regions within the same molecule) which have remained crystalline. These molecular movements have a cooperative effect on granule swelling and the overall melting of starch crystallites.

Starch Sources: Grain and Tuber/Root Structure

The majority of starch in cereal grains is located in the endosperm. Many grains can be referred to as having 'soft or floury starch' or 'hard or vitreous starch.'

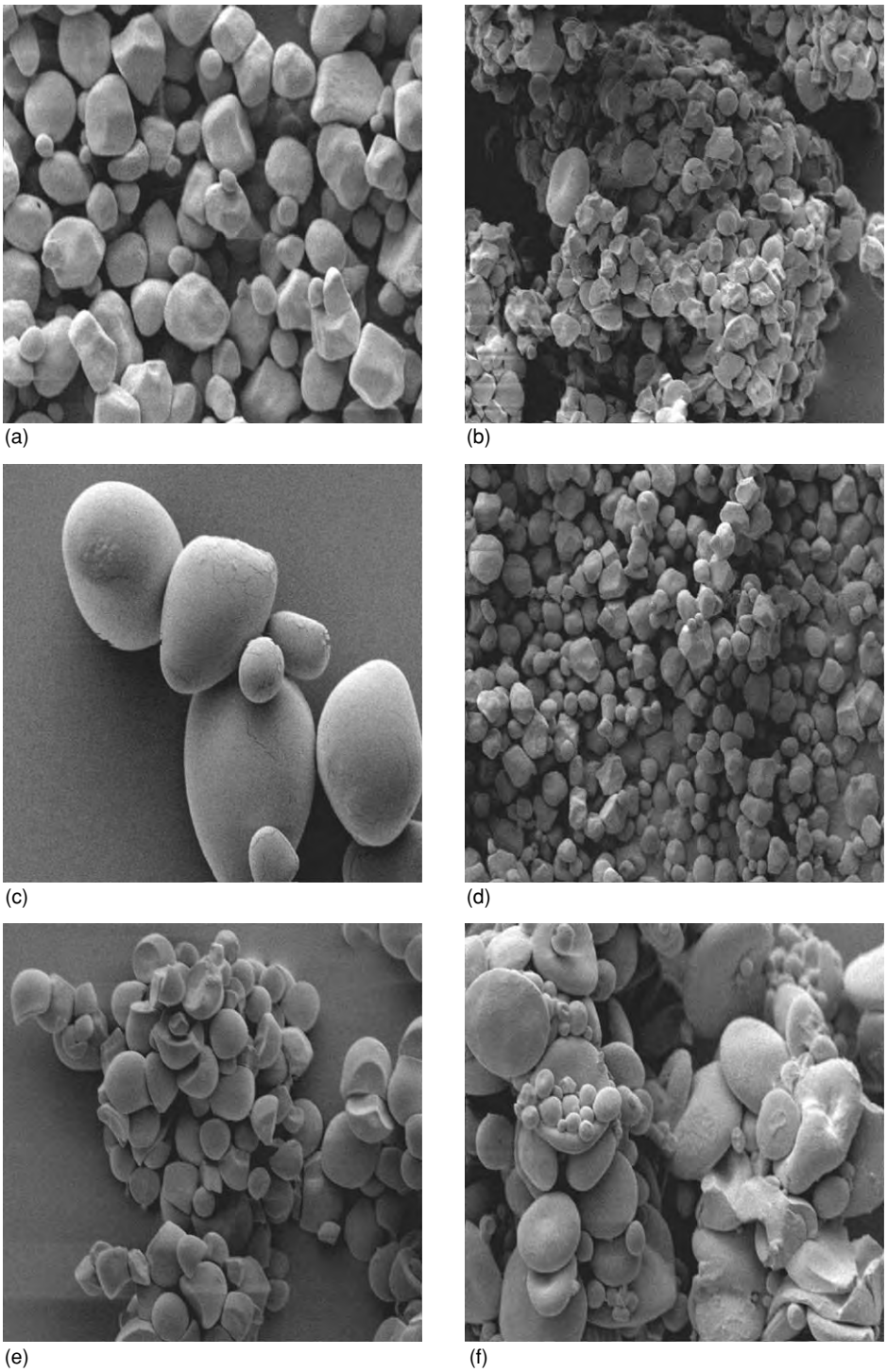

Figure 2 Scanning electron micrographs (× 1000) of (a) corn, (b) oat, (c) potato, (d) rice, (e) tapioca, and (f) wheat starches.

There is little evidence that the molecular components of their starches are any different. The packing of the starch granules within a kernel, however, is different. Floury maize kernels have less protein binding the starch granules together, and starch granules within more flinty maize kernels are encased more tightly within a protein matrix. Cereal starch is obtained via wet milling; after the endosperm is disrupted, the starch is collected by washing and centrifugation.

Cassava root, from which tapioca starch is produced, typically contains 25% starch. The roots have a hard or cork-like outer layer which is usually

removed before starch processing. Starch is collected by disrupting the cells, screening, and subsequent collection by washing and centrifugation.

Potatoes are also another important starch source; they typically contain about 22% starch. Raw potatoes are ground to disrupt cell walls; the pulp is washed, and screened. Starch is later collected via centrifugation. Starch can also be obtained, as a by-product, from commercial production of value-added potato products such as french fries, instant potatoes and potato chips (crisps).

Viscosity Measures

Amylograph/Rapid Viscoanalyzer

The viscosity of cooked starches is important to the food industry. One of the most important industrial measures of starch properties uses amylograph or Rapid Viscoanalyzer (RVA) instruments to measure a starch's thermal behavior. In the standard amylograph, a 3–5% starch dispersion is mixed in a cylindrical bowl. Paddles mix the starch dispersion, and the force these paddles encounter is recorded with a computer. The test begins with the gradual heating (1.5 °C min^{-1}) of the starch dispersion. The starch granules swell and the viscosity of the mixture increases as the temperature increases because the swollen granules occupy more area within the mixing vessel. When the temperature reaches 95 °C, the starch mixture continues to be stirred while the temperature is maintained for 60 min. When swollen unmodified granules are exposed to shear, the viscosity of the mixture decreases because the granules break apart into smaller pieces. After the holding cycle, the starch is subsequently cooled to 50 °C (also at 1–5 °C min^{-1}) and a highly viscous gel structure can form by hydrogen binding and water entrapment between starch polymers. The RVA is a similar device that uses smaller sample sizes and has programmable heating, cooling, and stir speeds (the manufacturers of the amylograph also make a instrument suitable for analyzing small amounts of starch). Typical RVA curves are shown in Figure 3.

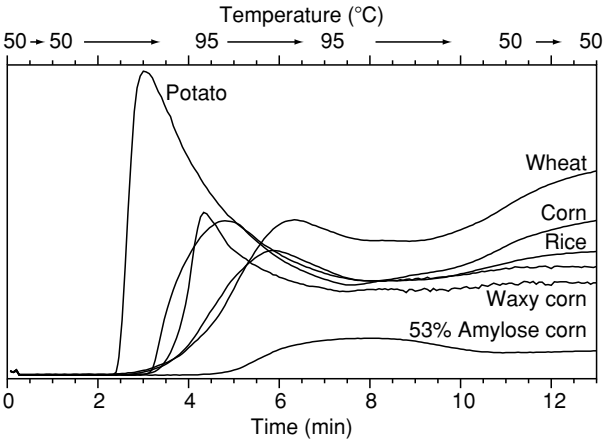

Figure 3 Representative Rapid Viscoanalyzer curves for potato, wheat, corn rice, waxy corn, and 53% amylose corn starches.

Chemical Methods

Determination of Total Starch

There are several chemical methods available to measure the total amount of starch in a sample. For samples high in starch content (cereal grains, tubers, and their products), starch's optical rotation in solution can be determined using a polarimeter. Samples are initially extracted with ethanol to remove optically active soluble materials and the remaining starch is suspended in a calcium chloride solution and placed in a boiling water bath. Proteins are removed in the procedure with stannic chloride-phosphotungstic acid. The resulting filtered solution's optical rotation is measured in the polarimeter and used to calculate percent starch; a similar optical procedure uses hydrochloric acid and a saturated lead solution. More recently developed methods of starch analysis, however, usually rely on the ability of the enzymes alpha-amylase and amyloglucosidase to degrade cooked (or chemically treated) starch to glucose. These methods often have the advantage of being applicable to a wider range of starch-containing products. For these procedures, it is usually essential to gelatinize fully (cook/disperse) the starch. For common materials without extensive resistant starch (fiber-like starch), the sample can be gelatinized in an autoclave. For more difficult-to-gelatinize samples, starch can be dispersed in methyl sulfoxide (also known as dimethyl sulfoxide or DMSO) prior to enzymatic digestion.

After the starch is digested with enzymes to glucose, most methods subsequently convert the glucose solution into a colored product that can be measured colorimetrically in a spectrophotometer. In fact, however, many of the glucose measurement techniques can be used at this stage. The percentage of starch is calculated by multiplying the measured amount of glucose evolved from the enzymatic treatment. Total glucose is multiplied by 0.9 to determine total starch; this calculation accounts for the difference in molecular weight between individual glucose units and glucose polymers. In addition to these chemical total starch methods, indirect procedures using near

infrared spectroscopy (NIR) technology (especially for grain/cereal crops) have proven to be very useful. While NIR methods must rely on these chemical procedures for method development and subsequent machine calibration, when used appropriately they are extremely rapid and can be performed by personnel not necessarily trained in complex chemical analyses.

Determination of Percentage Amylose

Since amylose and amylopectin have unique physical and chemical properties, and their relative proportions influence the overall properties of starches, the determination of the ratio of these polymers becomes important. Potentiometric titration is considered one of the better techniques for determining the percentage of linear components (i.e., amylose) in starch. It relies on the affinity between amylose and iodine. Potato and wheat starches have typical iodine affinities of approximately 19.9%, whereas maize starch has an iodine affinity of 19.0%.

Colorimetric tests for the percentage of amylose are also widely used. Solutions of amylose and iodine result in a deep blue-purple color. Amylopectin and iodine show only a slight affinity towards each other; their solutions have a reddish color. The blue value test relies on these properties; starch samples are dispersed in a sodium hydroxide solution and, after neutralization, the samples are allowed to react with an iodine solution. The absorbance (usually at 620–680 nm) of the solutions is measured to calculate blue value, and the apparent percentage amylose content can be determined by developing a blue value standard curve with 'purified' amylose. Unfortunately, commercially available amyloses are usually not exactingly pure, nor is such a pure amylose available for every kind of starch. Accurate chemical determinations of amylose content from different starch sources is made more difficult because not only do amyloses from different botanical sources have different iodine affinities, but different amylopectins will also bind a small amount of iodine and thus cannot be discounted in amylose determinations. A newer procedure for amylose determination relies upon the complexing action of concanavalin A (A) with amylopectin. The procedures precipitate amylopectin from starch solutions, leaving only amylose. The amylose is thus subsequently analyzed in much the same way as for total starch determinations.

Instrumental Methods

Differential Scanning Calorimetry

The differential scanning calorimeter (DSC) is widely used to determine the crystalline melting temperature of various starches (Figure 4). Small samples of starches are heated at a constant rate and the endothermic heat flow is recorded. Melting of crystalline sites (or some starch chemists postulate melting of crystalline sites and unfolding of double helices) results in a thermal event typically recorded as an endotherm. Usually, a broad crystalline melting range is recorded; sites of crystallinity melt at different temperatures depending upon their individual structure, the action of solvent (usually water), and the interaction of individual polymers which are contained and traverse crystalline sites. DSC can also be used to examine the recrystallization of gelatinized starches. Most notably, DSC is used to study the retrogradation process under various environmental conditions of time and temperature, and the interactions of starch with complexing agents and crystallization promoting or inhibiting ingredients.

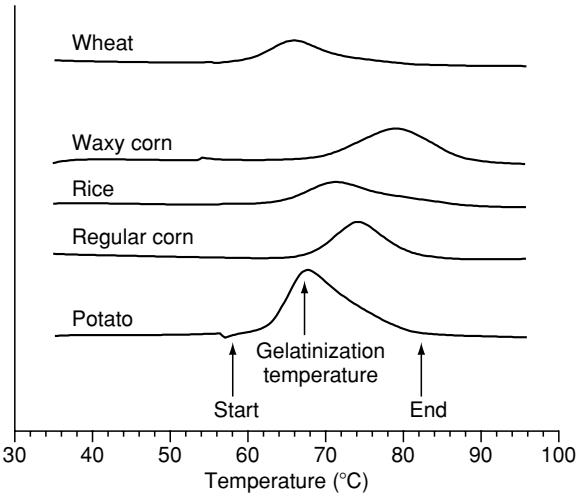

Figure 4 Differential scanning calorimetry (DSC) endotherms of wheat, waxy corn, rice, corn, and potato starches (scanned at 10 °C min^{-1}, > 60% water). The figure shows the start of crystalline melting, point of maximum endothermic energy (gelatinization temperature), and temperature at which all crystals have melted.

X-Ray Analysis

X-ray analysis of starch is useful in determining the extent of crystallinity and crystallite structure, the botanical origin of starches, the formation of starch complexes, and can be used in structural studies to determine spacing within starch helices. Unlike DSC, X-ray diffraction is less subject to scientific debate regarding the 'true' measurement of starch crystallinity. The diffraction patterns (groupings of spectral lines) of native starches have been broadly classified as A, B, C and V (Figure 5). Researchers have also proposed variants on these patterns. Cereal grain

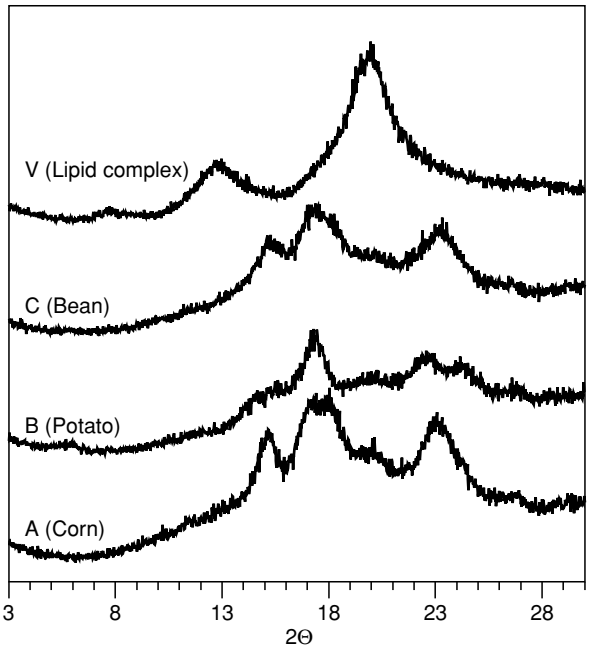

Figure 5 X-ray diffraction patterns (40 kV, 30 mA, Cu K-alpha radiation, 2-theta diffraction angle) of a lipid complexed corn (V-pattern), bean (C-pattern), potato (B-pattern), and corn (A-pattern) starches.

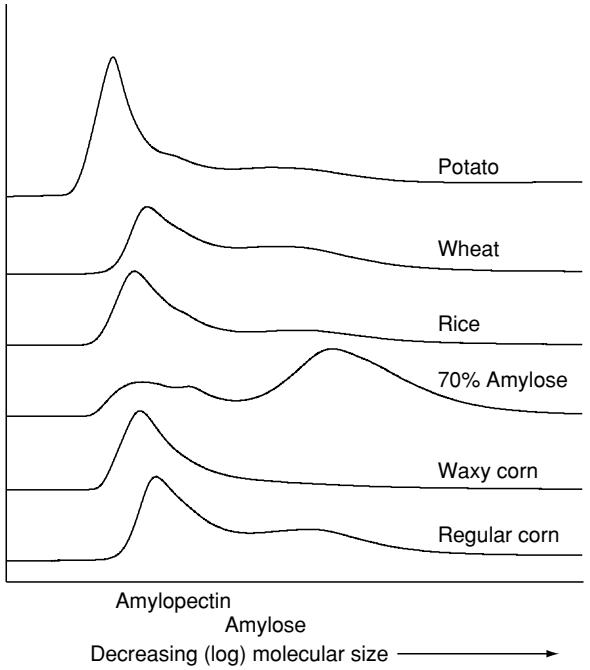

Figure 6 Typical high-performance size exclusion chromatograms (HPSEC) of potato, wheat, rice, 70% amylose, corn, waxy corn, and corn starches. The figure shows amylopectin and amylose peaks. Starch was dispersed in methyl sulfoxide (DMSO) before injection into the HPSEC system consisting of four columns linked in series. Carbohydrate was detected using a refractive index detector.

starches generally have the A pattern, although defatted high-amylose starches have a B pattern. Bean and cassava starches have a C pattern and potato starch has a B pattern. Complexes between fatty acids and amyloses, seen especially strongly in some high-amylose maize starches, exhibit a characteristic V pattern.

Chromatography

Size exclusion chromatography has been important in determining the fine structure of amylose and amylopectin molecules, as well as the relative proportion of those molecules in food materials. In addition, chromatographic techniques have proven useful in determining the relative solubility of starch polymers in solution and the percentage of amylose in starches. Chromatography can be conducted using conventional gel filtration techniques (soft gels), or more rapid high-performance liquid chromatography methods with rigid stationary-phase material (Figure 6). Starch samples are first dispersed (usually in methyl sulfoxide, also known as DMSO, NaOH, or water) and injected into a chromatographic system. The amylopectin usually elutes at or extremely close to the void volume of most modern column systems, and the smaller amylose elutes later. The molecular weight of these materials can be determined by coupling the chromatographic system to a laser light-scattering detector, or can be estimated by employing known-molecular-weight standards of pure carbohydrate polymers. The relative molecular weights or sizes of starch polymers can be compared, as well as the relative effect of various food-processing unit operations on starch structure. The extent of starch depolymerization caused by food or industrial processing is also easily measured using size exclusion liquid chromatography. Amylases can be used, coupled with chromatography and postchromatographic chemical analysis, to determine the extent and length of polymer branches, and other details about the fine structure of particular starches.

Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) is a powerful tool that has been used to understand starch–water interactions and the molecular relationship between chemical groups in starch molecules. For example, wide-line ^{1}H NMR has been used to characterize the mobility of water associated with starch pre-, post-, and during gelatinization; these results have confirmed the notion that gelatinization is a melting phenomenon. In addition, both ^{1}H and ^{13}C NMR have been used to study the degree of branching

found in starches from different botanical sources. ^{1}H NMR can distinguish between anomeric protons involved in alpha-1,4 and alpha-1,6 branching, and small peaks found in ^{13}C NMR scans have been associated with nonreducing ends. NMR has also proven useful in understanding interactions between starch and other chemicals such as sugars, proteins, and fatty acids.

See also: **Chromatography**: High-performance Liquid Chromatography; **Spectroscopy**: Nuclear Magnetic Resonance; **Starch**: Sources and Processing; Functional Properties

Further Reading

Hill RD and Munck L (eds) (1985) *New Approaches to Research on Cereal Carbohydrates*. Amsterdam: Elsevier.

Kulp K and Ponte JG (eds) (2000) *Handbook of Cereal Science and Technology*. New York: Marcel Dekker.

Manners DJ (1985) Some aspects of the structure of starch. *Cereal Foods World* 30: 461–467.

Radley JA (ed.) (1976) *Examination and Analysis of Starch and Starch Products*. London: Applied Science.

Thomas DJ and Atwell WA (1999) *Starches*. St Paul: Eagen Press/American Association of Cereal Chemists.

Whistler RL (ed.) (1984) *Starch Chemistry and Technology*, 2nd edn. Orlando: Academic Press.

Whistler RL and BeMiller JN (1997) *Carbohydrate Chemistry for Food Scientists*. St. Paul: Eagan Press.

Whistler RL and Paschall EF (eds) (1965–1967) *Fundamental Aspects*, vol. 1. *Industrial Aspects*, vol. 2. *Starch: Chemistry and Technology*. London: Academic Press.

Zobel HF (1988) Molecules to granules: a comprehensive starch review. *Starch/Starke* 40: 44–50.

Sources and Processing

W S Ratnayake and D S Jackson, University of Nebraska, Lincoln, NE, USA

Background

Starch is found in almost all tissues (leaves, roots, tubers, seeds, stems, flowers, etc.) in green plants. Only a small number of plants, however, are commercially grown for starch; they include cereal (such as wheat, corn, sorghum, and rice), tuber (mainly potato), root (tapioca and arrowroot), stem (sago), and legume (mainly pea) crops. All starches occur as minute granules in nature, but the size, shape, and nature of molecular arrangement inside the granule depends on the species and sometimes on crop cultivar or variety. The processing methods employed for starch extraction mainly depend on the nature and composition of raw material, i.e., source of starch.

Cereal Starches

Corn and Sorghum

Corn (maize, *Zea mays* L.) has been grown for 5000–7000 years and was probably first grown in Central Mexico. That first corn was very different from today's varieties, which were evolved through mutation, hybridization, and random and conscious selection. Among the types of corn available today are popcorn, sweet corn, dent corn, flint corn, and flour corn. The availability of dent corn at low prices, its storability, and its high starch content have led to its increasing use for starch production.

Grain sorghum (*Sorghum bicolor*), also known as milo, was probably initially cultivated 5000–7000 years ago in eastern Africa. Although sorghum is a major cereal crop and resembles corn in general composition, it is sometimes considered inferior to corn for food, feed, and industrial uses. Compared with corn, it can be grown effectively in more arid regions and can be milled in a similar way. Sorghum starch has properties similar to those of corn starch, and the extraction procedure (wet milling) is similar to that applied in corn starch extraction.

The corn kernel is a caryopsis (a berry), borne on an inflorescence called 'an ear.' The caryopsis matures approximately 60 days after pollination, and corn is harvested in late summer or early autumn. In sorghum, the seeds are borne in a terminal, bisexual rachis (head). Waxy sorghum or corn starches are produced by the varieties containing three recessive *wx* genes.

The corn (and sorghum) wet milling process consists of: (1) cleaning the grain, (2) steeping, (3) milling and fraction separation, (4) starch processing, and (5) product drying. Grain is initially cleaned by screening to remove foreign material and broken kernels. The grain is then steeped to soften the kernel. Sorghum typically needs longer steeping times than yellow dent corn. Steeping involves maintaining the correct balance of steeping water with regard to flow, temperature, sulfur dioxide concentration, and pH. Normal steeping conditions are 30–40 h at 48–52 °C. Sulfur dioxide (*c*. 0.1%) was initially used in steep water to prevent the growth of spoilage organisms but was later recognized as an indispensable reagent to maximize starch yield. The effect of sulfur dioxide on the protein matrix is apparently such that it disperses the protein matrix by breaking disulfide bonds, thus

releasing more starch. The use of unusually high temperatures and/or sulfites during steeping reduces the starch yield. Proper steep temperature is also important, as steeping between 45 and 55 °C promotes the growth of indigenous lactic acid bacteria. Growth of these bacteria lowers the pH of the steep, making the antimicrobial action of SO_2 more effective, and is helpful in softening the kernel structure. Lactic acid also increases the absorption rate of sulfur dioxide into the kernel. Steep-water pH is normally maintained at about 4.0 and is highly buffered.

After steeping, the grain is milled to obtain as complete a separation of components as possible. Milling is usually done in an attrition mill (Bauer-type) in preparation for germ separation. Once the grain integrity is disrupted, it is a fairly simple process to separate the lighter germ, an oil-rich fraction, from the heavier starch and protein fractions by use of a centrifugal hydroclone. Separated germ is cleaned by washing and dewatered by mechanical squeezers in preparation for oil recovery by hexane extraction. In milling grain sorghum, starch and protein are more difficult to separate, because the sorghum pericarp is more fragile than that of corn. Separation is enhanced if the kernels are dehulled prior to milling. This gives a sharper separation of starch and gluten with starch of whiter color and lower protein content. Sorghum dehulling prior to milling also improves germ separation.

Free starch and protein from the first pass through the mill are separated from unmilled endosperm and fiber by screening. The unmilled endosperm–fiber fraction is re-milled in a grinding mill to maximize starch separation. Screening after this process separates the starch–protein fraction from the remaining fiber fraction. The fiber fraction is dewatered and further processed and often incorporated with other wet milling end products (evaporated steepwater, spent germ cake and protein) to produce feed fractions.

The starch–protein mixture (mill starch) is separated by centrifugation. Protein (sometimes called 'gluten' in the corn processing industry) is less dense ($1.1\,g\,cm^{-3}$) than starch ($1.5\,g\,cm^{-3}$), so that these fractions may be easily separated. Starch obtained has 1–2% protein, which is mainly removed by further washing in a second centrifugal separator or set of hydroclones to obtain a final protein content of less than 0.38% (dry-weight basis). The pigments of sorghum sometimes discolor (light pink or gray color) starch. Bleaching with $NaClO_2$ or rinsing with NaOH or methanol is done to reduce this discoloration in isolated starch.

The final starch slurry may be dried and sold as unmodified starch, or the starch may be modified and dried, gelatinized and dried, or hydrolyzed to a syrup, which may be used directly or fermented to produce ethanol or other chemicals.

Starch is usually dried by flash drying. The starch slurry is dewatered by centrifugation or filtration, and the moist starch is introduced into the bottom of a stream of rapidly moving air (93–127 °C). Drying is rapid, and the product is collected in cyclones. Drying conditions can be altered to control bulk density and particle size. Wet milling is water- and energy-intensive; the wet-milling industry is the second most energy-intensive food industry in the USA.

Wheat

Wheat has been grown by man for 8000–9000 years, its cultivation probably beginning in the Middle East. The first commercial manufacture of wheat starch was in England in the 1500s. Wheat is mainly harvested by mechanical means, and starch is separated by one of several processing methods, such as the Martin process, the Batter process, the Fesca process, the ammonia process, the acid process, wet wheat milling, whole wheat fractionation, the Rasio process, or the hydroclone process. The first two processes are the most commonly used. Wheat has 60–70% starch (on a whole-seed basis).

At present, low-grade flour (rather than wheat) is used as the starting material for most wheat starch production. There are two, somewhat similar processes commonly used, i.e., Martin process and batter process.

The Martin (dough ball) process was first used in 1835 and employs wheat flour as its raw material. It consists of making a flour dough, washing out the starch, drying the gluten protein, purifying the starch, and drying the starch. Wheat flour and water (2:1) are blended to form a stiff dough, and the dough is then allowed to rest, hydrate, and strengthen the gluten. The dough is then washed with water while being kneaded or rolled, to remove the starch. Separated gluten is dewatered by roller compression and dried in a flash-drier. Starch slurry overflow from the dough washer is sieved to remove gluten particles and then fine-screened to remove bran particles. Purification and dewatering of the starch slurry are performed by centrifugation to produce a high-solids starch cake, which is then flash-dried. The final product is dried to 10–12% moisture and contains approximately 0.3% protein.

The Batter process is a derivative of the Martin process and differs in both the way the dough is formed and the way it is treated. Wheat flour and water (approximately 1:1) are mixed and the resultant dough allowed to rest for 30 min. This dough is then vigorously mixed with additional water equal to twice the weight of the original flour. Protein is

separated from this mixture by screening, and the starch further purified from bran by additional screenings. Centrifugal separation of the filtrate separates the large, lenticular starch granules (prime or type A starch) from the type B starch (a mixture of small granules, damaged granules, and pentosan complexes – also called tailings or squeegee starch), that is about 20% of the total starch. Type A starch is essentially pure and is subsequently dewatered and flash-dried. Type B starch is likewise processed but is of low purity. Sulfur dioxide is not used in wheat starch processing, as (1) it is not required to soften the wheat flour particles because water alone can soften the particles to an extent that protein is effectively separated from starch, and (2) it would destroy the desirable characteristics of the gluten.

Most wheat starch manufactured is sold unmodified and is mainly used in paper adhesives, laundry sizing, and cotton finishing, rather than in food. In foods, the thickening power of wheat starch is inferior to corn starch, but paste texture and strength, as well as paste clarity, are about equal to corn starch. Wheat starch may be preferred for use in certain baked goods owing to the lack of chemicals used in its production. Some modified wheat starches show emulsification properties in selected foods.

Rice

Rice (*Oryza sativa*) cultivation probably originated in Asia, and rice is still a staple food. In the USA, more than 90% of the rice crop is produced in four states: Arkansas, Louisiana, Texas, and California. The rice grain consists of a hull and an edible berry (caryopsis). The outer fibrous layers cover the starchy endosperm and the germ. The endosperm contains mainly starch granules and protein bodies, with little matrix protein in evidence.

To isolate rice starch, broken rice is treated with dilute (0.3–0.5%) sodium hydroxide for about 12 h to soften the grain and aid protein removal. It is wet hammer-milled, the cell walls are removed by screening, and the protein is extracted. The rice starch slurry is purified by further centrifugal washing and then dried.

Rice starch is little used, because of its high cost of production relative to other commercial starches but may be used as a custard or pudding thickener in foods, as a dusting powder in cosmetics, and as a stiffening agent in the cold-starching of fabrics. Rice starch is used in Europe in baby foods, and in the laundry and paper industries.

Potato

Potato starch is extracted from the tubers of *Solanum tuberosum*, which was first cultivated around AD 200 in Peru. Potato starch is mainly produced in Europe. Starch is typically isolated from cull potatoes, surplus potatoes, and waste streams from potato processing. However, there are special cultivars developed for starch manufacture. The tubers generally contain 65–80% starch (w/w).

Culled or surplus potatoes are washed with water in a flume, to remove dirt and foreign matter. They may be washed again on arrival at the processing plant. They are then disintegrated by a saw blade rasp or hammer mill, and the mashed product is screened to remove skins and fiber. During the disintegration process, sulfur dioxide is usually added to preserve the color and inhibit oxidation. Water-soluble impurities are separated by washing, and other impurities are separated by gravity separation.

Screening, by using a screening battery, separates the starch from the potato pulp, which can be reground, and a second starch extraction made to obtain a total yield of 12% starch based on raw potatoes. Starch isolated from the screens or sieves is reslurried in water to remove soluble material, and then dewatered in a continuous centrifuge. Addition of water, and concentration in a hydroclone is followed by further batch centrifuge or tabling operations for final purification of the potato starch. Starch slurry isolated from these operations is dewatered by vacuum filtration and flash-dried ($< 175\,^{\circ}C$ inlet temperature). Potato starch, dried to 17–18% or less moisture, is then screened and packaged. Unmodified potato starch has a relatively low temperature of gelatinization, and exhibits high viscosity on gelatinization, but breaks down on further heating and stirring. It can form useful films and has a high binding power.

Potato starch is used unmodified, or it may be pregelatinized, converted to a cationic form, dextrinized, or derivatized by hydroxyethylation, hydroxypropylation, or carboxymethylation. Large amounts of potato starch are used in thickening commercial soups, and pregelatinized potato starch is useful in instant puddings. Other uses of potato starch include pie fillings, sweets, chewing gums, and extrusion cooking, and as a filtering medium in breweries, where it is employed as a filter precoat when filtering yeast from the wort.

Tapioca

Tapioca starch is obtained from the large, tuberous roots of the cassava plant (*Manihot utilissima*, *Manihot esculenta*), which grows in many equatorial regions. Cassava roots may be either 'sweet' (containing less than 50 mg of potential hydrogen cyanide per kilogram of fresh root) or 'bitter' (containing 250 mg

or more of hydrogen cyanide per kilogram of fresh root). Sweet root varieties are grown for food purposes and bitter varieties for other industrial uses. In either case, the hydrogen cyanide is lowered to acceptable levels during processing.

Harvesting of roots is often manual, and the yield is usually 13–50 t ha^{-1}. Typically, roots contain 70% water, 24% starch, 2% fiber, 1% protein, and 3% fats, minerals, and sugars. Tubers should be processed within 24 h of harvest, since deterioration sets in from the time of root extraction and proceeds throughout the process.

The roots are initially washed to remove dirt and other field debris, and then the outer skin is removed. Breakage of root cell walls is accomplished by chopping the root into small pieces and passing them through a rasp disintegrator. During this process, hydrogen cyanide is released and washed away. Obtained pulp is washed on screens through which starch passes, but fiber is held back. Fiber is generally used in fertilizer or cattle feed. The starch slurry (also called starch milk), after screening, is put through a continuous centrifuge to separate the starch from fine fiber and soluble material. This can also be achieved by sedimentation. Starch thus collected may be reslurried and put through a centrifugal purification process as desired. Typically, sulfur dioxide (0.05%) is added to water used in the centrifugation purification processes to prevent microbial growth.

Starch slurry from the purification process is dewatered by centrifugation or vacuum filtration and then dried by drum, belt, tunnel, or flash methods. Flash-drying is most common. The final moisture content of the dried starch is in the range of 12–14%. Dried starch aggregates are pulverized to obtain a free-flowing powder. Tapioca starch is sold unmodified or is modified by derivatization, pregelatinization, or cross-linking. The viscosity of tapioca starch depends on plant variety, geography, harvest time, age of roots, soil fertility, rainfall, and manufacturing practices during production of the starch.

In some Asian countries, there is a commercial product called 'sago' made from tapioca starch. This 'sago' has no relation to actual sago starch. It contains starch pearls ($\sim$0.5–1.0 mm diameter) made from the extracted tapioca starch cake. During the production, the starch cake is mechanically globulated (or 'pearled') and then roasted to obtain a final moisture content of 9–11% moisture.

Arrowroot

Arrowroot starch is obtained from the root of the tropical perennial, *Maranta arundinacea*. Roots are harvestable after 6–12 months of growing and may contain more than 20% starch, most of which is extractable in the same manner, as previously described for tapioca starch. The difference in processing between the two roots is that arrowroot requires more washing than cassava. In addition, the outer skin of arrowroot root must be completely removed to prevent the starch from having an off-color and off-flavor. Arrowroot starch is principally produced in China, Brazil, and St. Vincent in the West Indies.

Sago

Sago starch is derived from the stem of palms (principally *Metroxylon* spp., *Arenga* spp., and *Maurilia* spp.) that are eight or more years old. The main production areas are Sarawak and Papua New Guinea; *Metroxylon sagu* is the popular species in these areas. Starch production is primarily done at the household level by hand. Cut palm trunks are split, and the pith (which contains about 40% starch) is scooped out. Kneading of the pith with water releases the starch, which is sieved to remove fiber, isolated by filtration, and dried. In commercial production, starch extraction is conducted in a manner similar to that of the household method. The pith of the trunk is rasped out, and the kneading is done mechanically. The crude, extracted starch is further purified in factories by water washing and sieving, and then dried using hot air. One palm trunk can yield 90–180 kg of sago starch, the granules of which are large (20–60 μm in diameter). Sago starch is used in foodstuffs or in textile sizings and adhesives.

Pea

Pea starch is extracted from the species *Pisum sativum*. It is also identified by many other names and produced throughout the world (field pea, garden pea, green pea, yellow pea, smooth pea, and wrinkled pea, etc.). The largest producer ($\sim$25% of the total) and the largest exporter (40% of the total) of peas in the world is Canada. France produces about 17% of the total.

Pea starch is mainly available as a byproduct of protein extraction. Therefore, it is considered to be a relatively cheap source of starch compared with corn, wheat, and potato starches. Pea starch is primarily used in industrial applications; its use in foods is limited, owing to its poor functional properties.

The isolation of starches from peas is difficult, owing to the presence of insoluble flocculent proteins and fine fiber, which decreases sedimentation and cosettles with the starch to give a brownish deposit. Pea starches are isolated using aqueous techniques as well as dry methods (pin milling and air classification). Air classification is the most commonly used

Table 1 Comparative properties of some commercial starches

Source of starch	*Amylose (%, w/w)*	*Amylopectin (%, w/w)*	*Granule diameter (μm)*	*Granule shape*	*GTR[a] (°C)*
Normal corn	28	72	3–26	Round, polyhedral	62–72
Waxy corn	0	100	3–26	Round, polyhedral	63–72
High amylose corn (amylomaize)	50–70	30–50	3–24	Round, deformed	67–92
Sorghum	28	72	15	Round (compound granules)	71–80
Wheat	30	70	2–38	Lenticular, round	58–64
Rice	17	83	3–8[b]	Polygonal, angular	68–78
Oat	16–29[c]	71–84[c]	3–10[b]	Polyhedral	53–59
Potato	20	80	5–100	Oval, spherical	50–68
Tapioca	16	84	4–35	Oval, truncated	49–65
Arrowroot	21	79	15–70	Oval, truncated	62–70
Sago	26	74	5–65	Oval, truncated	60–77
Peas (normal)[d]	33–49	51–77	5–40	Round, oval, spherical	55–61

[a]GTR, gelatinization temperature range.
[b]Individual (not compound) granules.
[c]Depends on the variety.
[d]Wrinkled peas and genetically modified varieties have different properties.

commercial method for pea starch isolation. The process requires a very high degree of particle size reduction (achieved by pin milling) in order to separate the starch granules from the protein matrix. The major fraction from the air classification process is the low-protein-containing starch fraction, which is separated from the fine protein fraction during the process. The crude starch concentrate contains about 65% starch. The residual protein associated with air-classified field pea starch granules is derived from protein bodies, agglomerates, chloroplast membrane remnants (which enclose the starch granule), and a water-soluble fraction that is presumably derived from the dehydrated starch. Remilling and reclassifying the starch fraction remove most of the protein bodies and agglomerates, and water washing removes most of the remaining attached protein. Purified, air-classified starch contains 0.25% protein.

The purity of pea starch obtained by wet milling is higher than that obtained by air classification. Repeated filtration through polypropylene screens (20 and 70 μm) combined with alkaline treatment (0.02% NaOH) causes substantial reduction in the protein content of wet process extracted pea starch. However, wet processing is not used to any great extent at commercial levels.

There is a slightly different procedure used for wrinkled pea extraction. Wrinkled pea seeds are steeped in warm water, followed by dehulling using rubber rollers (to separate hulls from cotyledons), and then the cotyledons are disintegrated by applying high pressure. The slurry is suspended in water, and the starch fraction is screened out. This process enables up to 89% of the starch present in the wrinkled peas to be extracted with a residual protein content of around 1%.

Comparative Properties and Applications

Some comparative properties of various commercial starches are given in Table 1. The general uses and applications of major commercial starches are discussed under the specific sections above.

In general, the physicochemical properties of a given starch vary depending on the source of starch and often on the starch extraction method. Commercially, wet milling techniques result in higher-quality and purer starches compared with dry techniques, as: (1) there is high starch granular damage in dry techniques, and (2) most impurities are easily removed by water washing in wet processing. Ultimately, however, the process chosen for starch extraction is usually determined by consumer/industrial demand, availability of utilities, yield/quality considerations, and various other economic factors.

Acknowledgement

This article is a revision of the first edition article by J. R. Daniel and R. L. Whistler, Purdue University, West Lafayette, USA from the Encyclopaedia of Food Science, Food Technology and Nutrition, Copyright 1993, Academic Press.

See also: **Cassava**: Uses as a Raw Material; **Flour**: Analysis of Wheat Flours; **Oats**; **Potatoes and Related Crops**: The Root Crop and its Uses; **Rice**; **Sago Palm**; **Starch**: Structure, Properties, and Determination; Sources and Processing; **Wheat**: Grain Structure of Wheat and Wheat-based Products

Further Reading

Balagopalan C, Padmaja G, Nanda SK and Moorty SN (1988) *Cassava in Food, Feed and Industry*. Boca Raton, FL: CRC Press.

Inglett GE (ed.) (1970) *Corn, Culture, Processing, Products*. Westport, CT: AVI.
Hoseney RC (1986) *Principles of Cereal Science and Technology*. St. Paul, MN: American Association of Cereal Chemists.
Kulp K and Ponte JG (eds) (2000) *Handbook of Cereal Science and Technology*. New York: Marcel Dekker.
Pomeranz Y and Munck L (eds) (1981) *Cereals: A Renewable Resource*. St. Paul, MN: American Association of Cereal Chemists.
Ratnayake WS, Hoover R and Warkentin T (2002) Pea Starch: Composition, Structure and Properties – A Review. *Starch/Starke* 54(6): 217–234.
Smith CW and Fredriksen RA (eds) (2000) *Sorghum: Origin, History, Technology and Production*. New York: John Wiley.
Talburt WF and Smith O (1987) *Potato Processing*, 4th edn. New York: Van Nostrand Reinhold.
Van Beynum GMA and Roels JA (eds) (1985) *Starch Conversion Technology*. New York: Marcel Dekker.
Watson SA and Ramstad PE (eds) (1987) *Corn: Chemistry and Technology*. St. Paul, MN: American Association of Cereal Chemists.
Whistler RL and Daniel JR (1985) Carbohydrates. In: *Food Chemistry*, 2nd edn. New York: Marcel Dekker.
Whistler RL, BeMiller JN and Paschall EF (1984) *Starch: Chemistry and Technology*. New York: Academic Press.

Functional Properties

D S Jackson, University of Nebraska, Lincoln, NE, USA

Introduction

Starches are used in a wide range of food products. They are sold to consumers as ingredients for thickening, and are used by food companies as thickeners, binding agents, stabilizers, texturizers, fat replacers/enhancers, encapsulation agents, gelling agents, film formers, moisture barriers, and crispness enhancers.

Solution Properties

Starches can be found in many physical forms. In water they can be found as raw (uncooked) granules, swollen (solvated) granules, granule pieces, entangled polymer masses of amylose and/or amylopectin, or as individual molecules of amylose and amylopectin. The presence and proportion of the components contribute to the unique functional properties of starch in specific food systems. Undamaged (partially crystalline) starch granules are insoluble in cold water and cannot be permanently dispersed in cold water without the addition of solvents such as methyl sulfoxide (DMSO) or alkali. If granules are suspended in water at 50–60 °C, however, they swell considerably and are partially dispersed. Upon prolonged gentle stirring of swollen starch granules at temperatures just below gelatinization, amylose is leached from the granules. After 4 h of aqueous leaching at 85 °C, approximately 29% of the total amylose in normal dent corn can be solubilized. This fraction of the amylose, however, is generally that which is exclusively within the amorphous regions of the starch granule. Specifically, without crystalline melting or gelatinization, amylose that traverses crystalline regions or whose movement is restricted within the granule by crystalline regions would be unable to leach from the granule. Upon vigorous agitation or some other induction of shear, more amylose and a small amount of amylopectin can also be solubilized at this temperature.

At temperature above the point of complete crystalline melting, starches can be thoroughly dispersed in water. Such a dispersion typically consists of soluble amylose and amylopectin, entangled polymer masses, and granule pieces consisting of mostly amylopectin. Complete dispersion of most normal starches (solubilized individual amylose and amylopectin molecules, and entangled polymers) occurs at autoclave temperatures without exposure to shear. Solubilization of 80–90% of individual molecules of amylose and amylopectin (no granule pieces or entangled polymers masses) can be achieved when corn starches (with amylose contents ranging from < 1 to 70%) are autoclaved (120 °C) and subsequently exposed to shear.

The degree of water solubility and dispersibility of starch depend largely on the temperature at which crystalline melting takes place. In the presence of excess water, the melting of crystalline areas takes place at relatively low temperatures (60–110 °C) because water helps to plasticize these regions. In water-limiting systems (including many foods), the temperature of melting (and hence gelatinization) is increased. Typical crystalline melting data, as determined by differential scanning calorimetry (DSC), are shown in **Table 1**.

Individual crystalline regions within starch molecules melt within a specific narrow temperature range, and the melting of some crystalline regions probably facilitates the melting of adjacent regions as amylopectin molecules that traverse the crystalline regions become more mobile. The temperature at which the first crystalline regions melt is referred to as the endotherm peak start and the temperature at which a tangent drawn along the steepest point of the endotherm curve intersects the baseline is the onset

Table 1 Typical differential scanning colorimetry starch crystalline melting (gelatinization) data when starch is exposed to excess water

Starch	*Endotherm peak start (°C)*	*Endotherm peak*[a] *(°C)*	*Endotherm peak end (°C)*
Waxy corn (< 1% amylose)	61.8	72.7	82.4
Normal dent corn (25% amylose)	61.8	71.4	80.1
53% Amylose corn	62.6	74.8	104.6
70% Amylose corn	62.3	92.2	108.4
Waxy sorghum (< 1% amylose)	70.8	76.1	82.7
Normal sorghum (24% amylose)	69.0	74.5	83.6
Waxy rice (< 1% amylose)	63.6	72.0	79.0
28% Amylose rice	68.6	77.5	86.2
Potato starch (16% amylose)	56.0	61.5	68.2

[a]Endotherms peaks, depending upon convention, are displayed as either going upward from a baseline or as dropping below the baseline.

temperature. The temperature at which the largest quantity of crystalline regions is melting is at the endotherm peak (often referred to as the gelatinization point or gelatinization temperature), and the temperature at which all crystalline regions have melted is the endotherm peak end. Since complete molecular solubility cannot be achieved without complete crystalline melting, it is apparent that high-amylose corn starches, whose endotherm peaks are wider and end at higher temperatures than other starches, are more difficult to solubilize completely. Even after the temperature of crystalline melting is exceeded, however, polymer entanglements must be dislodged by either increased temperatures and/or introduction of shear.

When concentrated suspensions of starch granules (i.e., 5%) are heated in water, complete solubilization of individual molecules of amylose and amylopectin is unlikely. When the granules swell, complete disruption of their original molecular structure may not take place, even if all of the crystalline areas are melted. The concentrated, dispersed polymers do not have complete freedom of movement. They are easily entangled with polymers and polymer segments within their originating starch granule, from other starch granules, and with molecules that leached from granules during the initial stages of swelling and gelatinization.

Gel Systems

Upon cooling, hydrated cooked starches form complex gel systems. The degree of association and the final rigidity of the gel system are largely dependent on the amylose/amylopectin ratio and the molecular characteristics (chain lengths and degree of branching) of both polymers. The initial gel formation, however, is largely dependent upon amylose properties; not only is linear amylose the primary component solubilized but it is that component which forms associations with other starch polymers most easily. Amylose is usually the component that forms the initial gel matrix which 'entraps' amylopectin and incompletely solubilized granule pieces. Waxy starch gels are far less rigid than normal starches or high-amylose starch. The final rigidity of an unmodified starch gel is largely dependent upon how closely solubilized starch polymers can associate and hydrogen-bond with each other. Solubilized amylopectin molecules, because of their highly branched nature, cannot associate as closely as linear amylose molecules. Therefore, the polymers in high-amylopectin-containing starches are prevented from establishing long, close associations of parallel hydrogen-bonded chains. The hydrogen-bonding that does take place between amylopectin molecules and amylopectin–amylose molecules results in water entrapment between the polymer chains. Linear amylose polymers, however, can form rigid and closely hydrogen-bonded associations (researchers often use a zipper analogy). Nevertheless, in the initial stages of cooling (when hydrogen bonds are being formed between amylose molecules), water is also entrapped between the polymers.

After prolonged cooling, the amylose–amylose, amylose–amylopectin, and amylopectin–amylopectin hydrogen bonding results in the squeezing out of water between associated polymers. This weeping process is referred to as syneresis. The process frequently manifests itself as a layer of water on a food product when unmodified starches are used to prepare fruit pie fillings or starch puddings and these gel systems are allowed to cool for several days (or shorter times depending upon the amylose–amylopectin ratio of the starches).

Because of differences in the closeness of association and the amount of water entrapment, the appearance of starch gels differs depending upon the amount of amylose and amylopectin. High-amylose starches are more white and opaque, while

high-amylopectin gels are more clear and have a slightly gray-white appearance. Potato starch, because its amylose component has a significantly higher molecular weight than that of other starches (corn amylose 100 000–200 000; potato amylose 400 000), has gels whose physical/chemical characteristics are closer to waxy starches than to cereal grain starches with equivalent amylose contents. Some amylose within potato starch associates by folding back on itself. Because of intramolecular hindrance, the association is not as tight as two separate amylose polymers lined up parallel to each other.

Starch Gel Characteristics

In addition to differences in starch gelatinization properties discussed elsewhere, native starches have widely differing gel-forming properties. These properties are summarized in Table 2.

The characteristics that starch imparts during gelatinization can be substantially influenced by chemical or physical modification of the starches. Much of the starch used in commercial food processing has been modified in some manner to impart desirable swelling, thickening, gelatinization, and retrogradation properties. Chemical modification of native starches can result in decreased swelling, changes in gelatinization temperature, and increased or decreased gel and retrograded starch-forming capacity. Most modifications, however, take advantage of the inherent properties of the native starch that is being modified.

Starch Inclusion Complexes

Starches form inclusion complexes with several chemical constituents, including fatty acids and mono- and diglycerides, emulsifiers, and many small flavor compounds. In general, starch complexes are formed when molecules bind within the alpha-helical structures found in amylose and portions of amylopectin polymers. Starch inclusion compounds, however, also encompass those instances when a nonstarch chemical constituent is entrapped within the three-dimensional network of starch polymers.

Starch–Iodine Binding

One of the simplest tests for identifying the presence of starches is by the application of iodine solutions to the sample in question. Since iodine can readily penetrate raw starch granules, gelatinization is not a prerequisite for staining. Starch will typically stain a vivid purple-blue color; waxy starches stain red. A central core is formed by amylose's alpha-helical structure; polyiodide ions are typically found within this core. Similar binding also takes place with outer chains of amylopectin, as well as smaller carbohydrates. The affinity of starch for iodine decreases as the starch chain-length is decreased. Since linear portions of starches are the primary binder of iodine, isoamylase- or pullalanase (breaks alpha-1,6 branch points)-treated starches show a far smaller decrease in iodine affinity versus starches treated with acid or alpha-amylases. Iodine complexing is also temperature-sensitive. Binding is reduced as temperatures increase above 40 °C; this corresponds to the change in starch structure from helical to coil-shaped. If allowed to cool, the iodine–starch complex reforms.

Starch–Lipid Complexes

Inclusion complexes similar to those involving iodine also form between starch and lipid components. This effect was first noticed when it was discovered that

Table 2 Starch properties during cooking in water (5% starch concentration)

Starch property during cooking	*Corn*	*Waxy corn and sorghum*	*High-amylose corn*[a]	*Potato*	*Tapioca*	*Rice*	*Wheat*
Rate of granular swelling (onset of thickening)	Slower	Slower	Very slow	Fast	Fast	Fast	Slower
Extent of granule swelling and solubility (viscosity during cooking)	Moderate	High	Low to moderate	Very high	Moderate	Moderate	Low
Swollen granule fragility (cooked starch susceptibility to shear)	Moderate	High	Moderate	High	Moderate	Moderate	Very low
Retrogradation of linear polymers (gel viscosity)	Very high	Very low	Extremely high	Low	Low	Low	Moderate

[a]High-amylose starches require more extensive cooking (to gelatinize) than regular starches (see Table 1). **Table 2** data assume extensive cooking. When not fully cooked, high-amylose corn starch paste viscosities are substantially lower than for cooked regular starches.
Certain data adapted, in part, from Zobel HF (1984) Gelatinization of starch and mechanical properties of starch pastes. *Starch Chemistry and Technology*, 2nd edn. Orlando: Academic Press.

lipids interfere with amylose–iodine complex formation. The carbon-chain segment of a lipid is located in the alpha-helical structure of amylose molecules and in long, linear segments of amylopectin (or amylopectin-like 'intermediate material') molecules. Hence, the binding of lipid is somewhat dependent on the molecular availability or solubility of starch polymers, especially amylose. If starch is fully gelatinized (amylose-solubilized), as the chain-length of saturated fatty acids increases, the amount of starch-lipid binding also increases. Also, there is less binding of lipids (given equivalent lipid solubility) to previously gelatinized starch as the temperature of starch–lipid mixtures increases. This is in contrast to an increase in binding that occurs between starches and lipids during the gelatinization process. Since starch–lipid binding is not a surface effect, amylose molecules are not readily available for binding with lipid if they are not gelatinized. Therefore, as the starch granule swells during gelatinization, more amylose is molecularly available for binding with lipid.

Monoglycerides generally bind less to starch than their corresponding fatty acid. A combination of lower lipid solubility and a slightly increased three-dimensional incompatibility between the monoglyceride and central binding core found within helical amylose probably account for this difference. Incompatible molecular shape also accounts for the decrease in lipid binding of unsaturated lipids as compared to their saturated counterparts; double bonds found in unsaturated lipids decrease subtle molecular movements in solution and form a molecular shape that may partially interfere with lipid inclusion within the amylose core.

The binding between starch and lipids has practical implications in the formulation of several starch-containing food products. For example, various surfactants (mono- and diglycerides) and fats are widely used to reduce bread staling. During the breadmaking process, starch granules become swollen in the presence of amylose-complexing lipid components. Since, during baking, starch–lipid interactions take place while the granule is swelling (50–60 °C), but before gelatinization, lipids can complex with amylose prior to its becoming soluble. As a result, starch granule swelling and gelatinization are reduced. Thus, less free amylose is subsequently available to hydrogen-bond (retrograde) with other amylose molecules. In addition, because of the reduced swelling, the moisture distribution between the wheat gluten and starch is altered, contributing to a softer crumb structure. Interestingly, the softening of crumb structure by amylose–surfactant complexing does not significantly affect the mechanism by which bread firms, i.e., the retrogradation of amylopectin. Surfactants have only a low affinity for complexing with amylopectin, although they may act via other mechanisms to retard staling.

Other Starch Inclusion Compounds

Flavor compounds and similar low-molecular-weight organics also readily complex with starch. The primary binding mechanism, as with iodine and lipids, is the formation of an inclusion complex within the helical structure of amylose. In addition, some flavor molecules may become physically entrapped by the long polymer chains of starches, although these molecules would be less likely to be retained during storage. The commercial application of flavor inclusion complexes is the preparation of dry flavor compounds using starch and liquid flavor ingredients. Inclusion compounds are most likely to form when starch is molecularly dispersed in the presence of flavor molecules. When hot water is used as a dispersant, the reformation of helical segments (from a coil) in amylose forms inclusion complexes in the presence of flavor compounds. Commercial application of this property includes the encapsulation of pollen flavor for Honey Bee food, and the encapsulation of thousands of flavoring substances for dry flavoring mixes for direct consumer use and as an aid in commercial food processing. The inherent property of starches to form inclusion complexes can be enhanced or reduced by chemical modification.

See also: **Iodine**: Properties and Determination; **Starch**: Structure, Properties, and Determination; Sources and Processing; Modified Starches; Resistant Starch

Further Reading

Hahn DE and Hood LF (1986) Factors influencing corn starch-lipid complexing. *Cereal Chemistry* 64: 81–85.

Kulp K and Ponte JG (eds) (2000) *Handbook of Cereal Science and Technology*. New York: Marcel Dekker.

Manners DJ (1985) Some aspects of the structure of starch. *Cereal Foods World* 30: 461–467.

Osman-Ismail F and Soms J (1973) The formation of inclusion compounds of starches with flavor substances. *Lebensmittel-Wissenschaft und Technologie* 6: 147–150.

Thomas DJ and Atwell WA (1999) *Starches*. St. Paul: Eagen Press/American Association of Cereal Chemists.

Whistler RL (ed.) (1984) *Starch Chemistry and Technology*, 2nd edn. Orlando: Academic Press.

Whistler RL and BeMiller JN (1997) *Carbohydrate Chemistry for Food Scientists*. St Paul: Eagan Press.

Whistler RL and Paschall EF (eds) (1965–67) *Starch. Chemistry and Technology*, vol. 1, *Fundamental Aspects*. London: Academic Press.

Zobel HF (1988) Molecules to granules: A comprehensive starch review. *Starch/Starke* 40: 44–50.

Modified Starches

J N BeMiller, Purdue University, West Lafayette, IN, USA

Background

Food processors generally require starches with better behavioral characteristics than provided by native starches. Cereal starches produce particularly weak-bodied, cohesive, rubbery pastes and undesirable gels when cooked. However, via modification, the functional properties of starches can be improved. Modification is done to introduce specific functionalities and to make resultant cooked products better able to withstand the conditions of heat, shear, and pH (acid) associated with processing conditions. The final products, modified food starches, are abundant, functional, and useful food ingredients, generally macroingredients. (*See* **Starch**: Functional Properties.)

Modifications can be chemical or physical. Chemical modifications are oxidation, cross-linking, stabilization, and depolymerization. Physical modifications make pregelatinized and cold-water-swelling products. Chemical modifications have the greatest effects on functionalities. Modifications can be single modifications, but modified starches are often prepared by combinations of two, three, and sometimes four processes. Chemical derivatives found in modified food starches (in the USA) are the following:

1. Stabilized starches
   a. Hydroxypropyl starches (starch ether)
   b. Starch acetates (starch ester)
   c. Starch octenylsuccinates (monostarch ester)
   d. Monostarch phosphate (ester)
2. Cross-linked starches
   a. Distarch phosphate
   b. Distarch adipate
3. Cross-linked and stabilized starches
   a. Hydroxypropylated distarch phosphate
   b. Phosphorylated distarch phosphate
   c. Acetylated distarch phosphate
   d. Acetylated distarch adipate

Property improvements that can be obtained by chemical modifications include the following:

1. Hypochlorite-oxidized starches
   a. Whiter
   b. Lower gelatinization and pasting temperature
   c. Decreased maximum paste viscosity
   d. Softer, clearer gels
2. Stabilized (hydroxypropylated or acetylated) starches
   a. Lower gelatinization and pasting temperatures
   b. Improved freeze–thaw stability of pastes and gels
   c. Decreased setback of pastes and gels (improved paste stability)
   d. Easier redispersibility when pregelatinized
   e. Greater clarity of pastes and gels
3. Cross-linked (phosphorylated) starches
   a. Increased gelatinization and pasting temperatures
   b. Increased shear resistance
   c. Increased acid stability
   d. Decreased setback of pastes and gels (improved paste stability)
   e. Increased viscosity of pastes
4. Cross-linked and stabilized starches
   a. Lower gelatinization and pasting temperatures, but increased paste viscosity
   b. Other attributes of stabilized and cross-linked products
5. Thinned (depolymerized) starches
   a. Decreased viscosity of pastes
   b. Lower gelatinization and pasting temperatures
   c. Increased solubility

Any starch (corn, potato, tapioca/cassava, wheat, rice, etc.) can be modified, but modification is practiced significantly only on corn (both common corn and waxy maize) and potato starches and, to a much lesser extent, on tapioca and wheat starches. This article is written primarily from the point of view of corn and waxy maize starches. (*See* **Starch**: Sources and Processing.)

Methods of Production and Applications

Cross-linked and/or stabilized starch products are prepared by chemical derivatization of a starch, most often in an aqueous slurry in a batch process. In such a process, a slurry of 30–45% solids (starch) as obtained from the mill is introduced into a stirred reaction tank. Sodium chloride or sodium sulfate is added to inhibit granule swelling. The pH is adjusted with sodium hydroxide (up to values of ≈ 11.5, depending on the reaction). Chemical reagents are added. Temperature is controlled. Reactions may be done at temperatures up to 50 °C, but gelatinization must be avoided to allow recovery of the modified starch in granule form by filtration or centrifugation. Because the gelatinization temperature may be lowered by the modification, there may be, and often is, a limit to the degree of substitution that can be made in this manner. (The degree of substitution is the average number of hydroxyl groups per α-D-glucopyranosyl unit (the monomeric unit of starch)

that have been derivatized, the maximum being 3.) In some reactions, the pH needs to be controlled by the metered addition of dilute sodium hydroxide solutions. Following modification to the desired level, the starch is recovered by centrifugation or filtration, washed, and dried.

Chemical reactions currently both allowed and used to prepare modified food starches in the USA are as follows:

- Esterification with acetic anhydride, succinic anhydride, the mixed anhydride of acetic and adipic acids, 1-octenylsuccinic anhydride, phosphoryl chloride, sodium trimetaphosphate, sodium tripolyphosphate, or monosodium orthophosphate
- Etherification with propylene oxide
- Acid modification with hydrochloric or sulfuric acids
- Bleaching with hydrogen peroxide, peracetic acid, potassium permanganate, or sodium hypochlorite
- Oxidation with sodium hypochlorite
- Various combinations of these reactions

Other reagents may be used in other countries.

Waxy maize starch modifications are especially popular in the US food industry because the inherent properties of waxy maize starch make them preferred over modifications to common corn starch.

Cross-Linking

Cross-linking is the most important modification of a food starch. Cross-linking occurs when starch granules are reacted with difunctional reagents to connect hydroxyl groups on two different molecules within the granule. Cross-links reinforce the granule and reduce both the rate and the degree of granule swelling and subsequent disintegration, i.e., reduce sensitivity to processing conditions (high temperature; extended cooking times; low pH; high shear during mixing, milling, homogenization, and/or pumping). Cooked pastes of cross-linked starches are more viscous, heavier-bodied, shorter-textured, and less likely to break down with extended cooking times, greater acidity, or severe agitation than are pastes of the native starches from which they are prepared. Only a small amount of cross-linking is required to produce a noticeable effect; for example, one cross-link for every approximately 1200 α-D glucopyranosyl units greatly reduces both the rate and the degree of granule swelling, greatly increases paste stability, and changes dramatically both the viscosity profile as the starch is cooked and the textural characteristics of its paste. Three times that much crosslinking, for example, produces a product in which granule swelling is restricted to the point that a peak viscosity is never reached in a slurry heated to 95 °C and held at that temperature with moderate stirring. As the number of cross-links increases, the granules become more and more tolerant to physical conditions and acidity, and swell and disintegrate (solubilize) upon cooking less and less. Energy requirements to reach maximum swelling and viscosity are also increased.

By far the most common cross-links are distarch phosphate esters. These distarch phosphates are prepared with either phosphoryl chloride or sodium trimetaphosphate. Phosphoryl chloride is very reactive and undoubtedly reacts near granule surfaces. To prepare cross-linked starches with phosphoryl chloride, the reagent is added to an aqueous starch suspension of pH 8–12. To cross-link a starch with sodium trimetaphosphate, it is slurried in a solution of the reagent at pH 5.0–8.5; the suspension is filtered, and the starch is dried. In this case, the cross-links are undoubtedly more evenly distributed throughout the granule.

A relatively small amount of cross-linked starch is made by reaction of corn starch with the mixed anhydride of adipic and acetic acids in aqueous alkaline suspension.

Stabilization

Derivatization of a starch with monofunctional reagents reduces the intermolecular associations which result in gelation of its paste and/or precipitation of the starch polymers (combined processes termed retrogradation or setback). Pastes of unmodified starches generally will gel, and the gels will usually be cohesive, rubbery, long-textured, and prone to syneresis. (Waxy maize starch pastes gel to a very limited extent at room temperature, but will become cloudy and chunky and exhibit syneresis when stored under refrigerator or freezing conditions.)

The most common derivatives employed for starch stabilization are the hydroxypropyl ether and acetate and monostarch phosphate esters. Acetylation is accomplished by treating a starch slurry with acetic anhydride at pH 7–11, the optimum pH depending on the reaction temperature. Acetylation of starch lowers the gelatinization temperature, an indication of a weakening of granules. Upon cooking, a higher peak viscosity is obtained due to greater granule swelling. Upon cooling of the resulting paste, the viscosity becomes lower than that obtained from the unmodified starch, an indication of improved stability, i.e., less retrogradation. Acetylated starches with an acetyl content of up to 2.5% (degree of substitution, DS, 0.09) can be used in food products (USA). (A DS of 0.09 indicates an average of nine acetyl groups per 100 α-D-glucopyranosyl units.)

Sodium phosphate monoesters are prepared by impregnating the starch with a solution of sodium

tripolyphosphate. After adjustment of the pH to 5.0–8.5, the slurry is mixed, then filtered, and the filter cake is dried and heated. Sodium tripolyphosphate is used to make products of up to 0.002 DS (one phosphate group per 500 α-D-glucopyranosyl units), the maximum allowed in the USA. Monosodium orthophosphate in the pH range 5.0–6.5 is also used to produce monostarch phosphates in the same way.

Monostarch phosphates produce stable pastes that are clear and have a long, cohesive texture. Paste viscosity can be controlled by varying the concentrations of phosphate salt, time of reaction, temperature, and pH. Increasing substitution lowers the gelatinization temperature; products become cold-water-swelling at DS 0.07. Corn starch phosphates of DS 0.01–0.03 produce pastes with hot viscosity, clarity, stability, and texture more like those of potato starch. Starch phosphates are good emulsion stabilizers and produce pastes with improved freeze–thaw stability.

Hydroxypropyl ether derivatives of starches are prepared by reacting an alkaline slurry with propylene oxide. To a starch slurry is added sodium sulfate and sodium hydroxide. The reactor is charged with propylene oxide and sealed. Reaction is continued for about 24 h at about 49 °C. The maximum allowable moles of substitution in the USA is 0.2 (7.0% of hydroxypropyl groups). (Moles of substitution, MS, is essentially the same as DS but is used in place of DS because each hydroxypropyl group contains a hydroxyl group that can itself be etherified, so that the maximum number of substituent groups per glucosyl unit can be more than 3.)

Low-MS hydroxylpropylstarches behave much like low-DS starch acetates and are used because of similar improvements in texture and appearance. The hydroxypropyl ether linkage is, however, much more stable than an ester linkage.

Starch succinate half-esters are prepared by reacting starch with succinic anhydride.

Starches with Hydrophobic Groups

Reaction of starch with 1-octenylsuccinic anhydride introduces hydrophobic substituent groups. Such derivatives can be used as emulsifiers and emulsion stabilizers in products based on oil-in-water emulsions, such as pourable dressings and flavored beverages. Flavor oil emulsions containing a thin-boiling starch or dextrin (see below) derivatized with 1-octenylsuccinate ester groups may be spray-dried. The flavor oil in the resulting powder is protected against oxidation, and the emulsion will reform when the powder is stirred into an aqueous medium. Gum arabic is, however, usually the material of choice for this application. Higher-DS products are nonwetting and are used as release agents for dusting on dough sheets and as processing aids. The maximum DS level allowed in the USA is 0.02.

Acid Modification

Thin-boiling starches are prepared by treating a suspension of a native or derivatized starch with dilute mineral acid at a temperature below the gelatinization temperature. When a product that gives the desired paste viscosity is produced, the acid is neutralized, and the product is recovered by centrifugation or filtration, washed, and dried. Even though only a few glycosidic bonds are hydrolyzed, granules disintegrate more easily and after only a small degree of swelling. Acid-modified starches form gels with improved clarity and increased strength, even though their pastes are less viscous. Thin-boiling starches are used as film formers and adhesives in products such as pan-coated nuts and candies, whenever a strong gel is desired, e.g., in gum candies such as jelly beans, jujubes, orange slices, and spearmint leaves, and in processed cheese loaves. To prepare especially strong and fast-setting gels, a high-amylose corn starch is used. More extensive modification with acid produces dextrins. (*See* **Dextrins**.)

Oxidation

Depolymerization, viscosity reduction, and decreased pasting temperature can also be achieved by oxidation with sodium hypochlorite (chlorine in an alkaline solution). Oxidation also reduces association of amylose molecules, i.e., results in some stabilization via introduction of small amounts of carboxylate and carbonyl groups. Oxidized starches produce intermediate-viscosity and soft gels and are used when these properties are needed. They are also used to improve adhesion of starch batters to fish and meat and in breadings. Mild treatment with sodium hypochlorite, hydrogen peroxide, or potassium permanganate simply bleaches the starch and reduces the count of viable microbes.

Pregelatinization

Pregelatinized starches are precooked starches that can be dispersed (dissolved) in water at temperatures below the gelatinization temperatures of the parent starches; thus, these 'instant' starches need no cooking. To prepare a pregelatinized starch, a slurry is simultaneously cooked and dried on hot drums. Because pregelatinized starch products are powders prepared from dried pastes, generally no granules are present, although granule fragments may be. Both chemically modified and unmodified starches can be used. The resulting products contain no intact starch

granules. If chemically modified starches are used, the properties introduced by the modification(s) are found in the pregelatinized products; thus, paste properties such as freeze–thaw stability can be characteristics of pregelatinized starches. Several physical forms of pregelatinized starches are produced. For example, some will produce smooth solutions; others will produce pulpy or grainy dispersions and find use in fruit drinks and tomato products. Pregelatinized starches are often used in dry mixes, as are maltodextrins, because they disperse readily, even when mixed with other ingredients. Starches that are not pregelatinized are known as cook-up starches.

Cold-Water-Swelling Starches

Starch products that are gelatinized starches, i.e., starches that have lost their crystallinity, but which retain their granular form, in contrast to standard pregelatinized starches, are called cold-water-swelling starches. There are several ways that such products can be prepared; one way is to heat a starch in an aqueous alcohol solution with sufficient water to allow gelatinization and sufficient alcohol that granule integrity is maintained. Cold-water-swelling products swell rapidly and thicken unheated aqueous systems. (A granular, cook-up starch requires heating a slurry to the pasting temperature before thickening occurs.)

Multiple Modifications

Modified food starches are tailor-made for specific applications. Most modified food starches are made by cross-linking, introduction of monosubstituent groups (stabilization), or a combination of these two approaches. Many products, in fact, have received two or more modifications. For example, a modified food starch may be a cross-linked and stabilized waxy maize starch; another may be a stabilized, acid-thinned, and pregelatinized common corn starch. Characteristics that can be controlled/improved by multiple modifications include, but are not limited to, one or more of the following:

- Adhesion
- Clarity of solutions/pastes
- Color
- Emulsion stabilization
- Film formation
- Flavor release
- Hydration rate
- Moisture retention and control in product
- Mouth feel of product
- Oil migration control in product
- Paste texture/consistency
- Product form (liquid, semisolid, solid)
- Sheen of product
- Shelf-stability of product
- Stability to acids
- Stability to heat
- Stability to shear
- Tackiness
- Temperature required to cook
- Viscosity (hot paste and cold paste)

Digestion and Metabolism

Various regulations concerning reagents that may be used and the maximum allowable modification of a starch for food use, alone or in combination with another modification, are in effect around the world. Generally, the level of substitution in a derivatized food starch is below DS 0.1 and in the range DS 0.002–0.2. Because of this low level of modification, the digestion, metabolism, and caloric values of modified food starches are reduced only to a minor, usually unmeasurable, extent as compared to native starches. Because only monosaccharides (D-glucose in this case) are absorbed, fragments containing esterified, etherified, or oxidized α-D-glucopyranosyl units should not be absorbed from the small intestine. (*See* **Carbohydrates:** Digestion, Absorption, and Metabolism.)

See also: **Carbohydrates**: Digestion, Absorption, and Metabolism; **Dextrins**; **Starch**: Sources and Processing; Functional Properties

Further Reading

Eliasson A-C (ed.) (1996) *Carbohydrates in Food*. New York: Marcel Dekker.

Light JM (1990) Modified food starches. *Cereal Foods World* 35: 1081–1092.

Whistler RL, BeMiller JN and Paschall EF (eds) (1984) *Starch: Chemistry and Technology*, chapters 10, 17, 19. Orlando, Florida: Academic Press.

Wurzburg OB (ed.) (1986) *Modified Starches: Properties and Uses*, chapters 2–4, 6, 7, 9, 12. Boca Raton, Florida: CRC Press.

Resistant Starch

R M Faulks, Institute of Food Research, Norwich, UK

Introduction

The term 'resistant starch' was first coined by Englyst, Wiggins, and Cummings in 1982 to describe a small

fraction of starch that was resistant to hydrolysis by exhaustive α-amylase and pullulanase treatment *in vitro*. This starch fraction was found during the destarching of food samples for dietary fiber analysis and consists of retrograded amylose, B-type starch granules, starch encapsulated within plant or tissue structures, and some modified starch structures.

Recognition that this small fraction of starch could be a significant contaminant of the dietary fiber led to calls to have it included in the definition of dietary fiber. More fundamentally, it also led to a reappraisal of the way starches behave *in vivo* and, in particular, whether or not resistant starches possessed any of the properties usually associated with dietary fiber, e.g., fecal bulking, laxation, shorter transit times. However, because starch reaching the large intestine may be more or less fermented by the gut microflora, some workers prefer to consider resistant starch as that portion of dietary starch which escapes digestion in the small intestine. This is an important distinction because resistant starch measured as a dietary fiber contaminant can only account for a fraction of the starch lost from the terminal ileum. (*See* **Dietary Fiber**: Physiological Effects.)

Because of the possible nutritional significance of unavailable dietary starch, it is now widely accepted that resistant starch is best defined as that fraction of dietary starch which escapes digestion in the small intestine. Resistant starch is therefore a variable fraction of starch, defined by physical, chemical, and physiological factors, but not necessarily resistant to hydrolysis by starch-degrading enzymes *in vitro*. This definition permits the fractionation of resistant starch into different categories according to the reason for its nondigestion (e.g., enclosed in food matrix, slowly digested, enzymatically resistant), and allows the determination of any functional or physiological role in the small intestine. Furthermore, this classification may permit a clearer understanding of how much starch is lost from the small intestine, and of its function and fate in the large intestine.

Formation

The formation of resistant starch, or, more accurately, the change in the degree of resistance of starch to hydrolysis by α-amylase and glucoamylases, is a function of the structure and composition of the starch, its processing environment, and subsequent storage conditions.

The well-known granular structure of starches is lost if starches are heated above about 70 °C in an aqueous environment, although some starches, particularly those from some legumes or high in amylose, may not gel unless heated to a much higher temperature, e.g., 120 °C. In conditions of limited or no moisture, heating may not result in significant changes in starch granules, which may retain their native crystalline form. Formation of resistance normally occurs most extensively in high-amylose starches that have been thoroughly dispersed by cooking (100–120 °C) in water and subsequently allowed to cool to produce a set gel containing 10–30% starch. Prolonged storage of gels under cool conditions, freezing and thawing, and reheating and cooling may also increase the degree of resistance.

As the dispersed gel cools, sections of the linear amylose molecules rapidly associate by hydrogen bonding to produce retrograded crystalline structures. Starch gels can therefore be considered as gelatinized granules embedded in an amylose matrix. If the amylopectin in the gelatinized granules is not too highly branched, parts of its structure may also crystallize. This retrogradation of amylopectin is a much slower process. The result is therefore the production of a gel system containing two phases: a highly structured phase, containing predominantly retrograded amylose, and a less structured amorphous phase, containing some retrograded amylopectin.

Examination of retrograded starch gels by differential scanning calorimetry clearly shows two endotherm peaks relating to the melting of retrograded amylopectin (*c.* 70 °C) and retrograded amylose (*c.* 160 °C).

It is the formation of physically ordered structures within the gel which is believed to confer resistance to amylolysis. The degree of association (retrogradation) of cooked starches may be influenced by a range of factors, including amylose-to-amylopectin ratio, starch concentration, pH, ions, temperature, age, and the presence of lipids.

Physical and Chemical Properties

Resistant starch is not a chemically defined entity. Its properties *in vivo* or *in vitro* can therefore be described only in general terms.

Resistant Starch Granules

Raw starch granules embedded in the cellular structure of food fragments may escape digestion because they are inaccessible to α-amylase and glucoamylases *in vivo*. This occurs mainly with the consumption of 'whole-foods', in which the particle size is large, or in tissues that have thick cell walls, or in some processed foods, in which the food structure restricts access by amylolytic enzymes. In these cases, both the chemical and physical properties of the starch are normal for the type of starch.

Although most native starches undergo quite rapid hydrolysis *in vivo*, in that they are found only in trace amounts in the terminal ileum, some native starches are very resistant to digestion. The main dietary starches to show marked resistance to amylolysis *in vivo* are raw banana and potato. It has been suggested that resistance to digestion is caused by the B form of crystallinity of the granule, unlike the A forms (wheat, maize) and C forms (sweet potato, tapioca) more commonly found. Retrograded starches are also known to develop the B-form structure. (*See* **Bananas and Plantains**; **Potatoes and Related Crops**: Processing Potato Tubers.)

Resistant Cooked Starch

The amorphous phase of retrograded starch gels is rapidly hydrolyzed (in minutes) by α-amylase, leaving the structured phase which is only slowly attacked (days). The hydrolysis curve of cooked retrograded starch gels takes the general form shown in Figure 1. The initial rate and extent of hydrolysis are dependent on the botanical origin of the starch and factors influencing the degree of retrogradation and the α-amylase activity. Prolonged treatment with α-amylase may lead to almost 100% hydrolysis of the starch, except in the case of pure amylose, when the degree of hydrolysis is much lower. *In vivo*, therefore, resistant starch from cooked foods is probably composed of a variable proportion of highly α-amylase-resistant starch (depending on the efficiency of digestion) and any residual starch that is truly α-amylase-resistant (retrograded amylose).

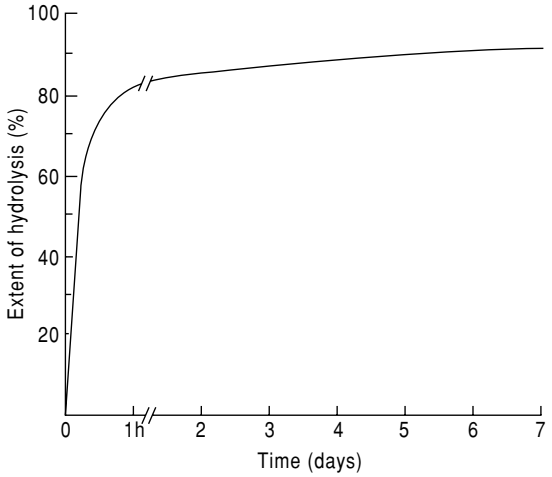

Figure 1 Typical hydrolysis curve of retrograded starch gel treated with α-amylase *in vitro*. Reproduced from Starch, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Enzymatic treatment of starch gels may be used to isolate the more resistant fractions for further study.

Isolated starch with a high degree of resistance to α-amylase is a water-insoluble white solid which has none of the functional characteristics normally associated with water-soluble polysaccharides (e.g., viscosity) or with the heterogeneous mixture of polysaccharides that make up the cell wall of plants.

Retrograded amylose isolated by α-amylase treatment has been found to consist mainly of linear glucose polymers with a degree of polymerization of 60. Careful acid hydrolysis of retrograded amylopectin yields the associated regions which have been shown to contain branched fragments in which the individual chains have a degree of polymerization of 15.

Resistance to α-amylase can be effectively abolished by melting the retrograded structures (*c.* 70 °C for amylopectin and *c.* 160 °C for amylose).

Disruption of the H bonds that hold the crystalline structures together may be achieved chemically by treatment with cold dilute alkali (potassium hydroxide, 2 mol l^{-1}) or with hot (100 °C) anhydrous dimethyl sulfoxide (DMSO).

Some of the resistant structures from high-amylose starch gels (retrograded amylose) may also show considerable resistance to hydrolysis in dilute mineral acid (sulfuric acid, 1 mol l^{-1}) and may need a Seaman-type hydrolysis (cold sulfuric acid, 12 mol l^{-1}, followed by 1 mol l^{-1} at 100 °C) to achieve 100% yield of glucose.

Physiological Effects

When resistant starch residues were first described in 'destarched' dietary fiber residues, it was considered appropriate to allocate them to the dietary fiber complex, despite the fact that they are predominantly α-1,4-linked glucans and potentially digestible in the small intestine. It was later realized that the amount of resistant starch could be manipulated by processing and would therefore be variable as well as chemically distinct from the nonstarch polysaccharides (NSP). More recently, it has been demonstrated that highly α-amylase-resistant starch fractions that have been isolated from different sources differ with respect to the extent of digestion in the small intestine and fermentation in the large intestine. Resistant starch is therefore a variable fraction of the starch, and the fact that it escapes digestion either *in vivo* or *in vitro* is insufficient reason to allocate it as dietary fiber. All the current main methods for the measurement of dietary fiber (Englyst, Uppsala, Association of Official Analytical Chemists) have modifications that allow the inclusion or exclusion of residual starch from the dietary fiber value. (*See* **Dietary**

Fiber: Properties and Sources; **Dietary Fiber**: Determination.)

There is little experimental evidence for the functional properties or physiological effects of resistant starch *in vivo*. This is because (1) it is difficult to define resistant starch other than in physiological terms; (2) there are inadequate *in vitro* methods for resistant starch measurement; and (3) feeding trials have tended to use foods in which there are overlying effects of the food matrix and a great excess of readily digestible starch.

It is known that when isolated, highly α-amylase-resistant starch is fed to rats, there is some (20–60%) digestion of the carbohydrate in the small intestine, the remainder being lost to the cecum. This starch is digested and absorbed more distally in the ileum than rapidly digestible starch. More distal absorption of glucose might be expected to have some effects on gut physiology and gut hormones. However, no differences were found in glucagon, enteroglucagon, and cholesterol between rats fed isolated resistant starch from high-amylose maize starch as the only complex carbohydrate, and those fed corn starch. However, in the resistant-starch-fed rats, the crypt cell production rate in the ileum was significantly greater.

The presence of resistant starch in the diets of unadapted rats appears to cause a slowing in both the rate and velocity of ileal digesta transport. It may be that this constitutes some evidence for an 'ileal brake' mechanism which slows digesta movement as a result of presence of absorbable carbohydrate in the distal ileum. This effect is lost during adaptation.

Starch that is not digested in the small intestine is lost to the cecum or large intestine, where it may undergo fermentation by the microflora. Evidence from rat-feeding experiments has shown that the amount of fermentation is dependent on the source of the resistant starch and the extent of adaptation of the animal.

The ceca of rats fed resistant starch are generally much enlarged, and the pH of the cecal contents is lower than normal, indicating fermentative production of fatty acids. By analogy with fermentable NSP, the main fatty acids produced from the fermentation of resistant starch are believed to be acetate, propionate, and butyrate.

Initially, a large proportion of resistant starch is voided in the feces but this falls quite quickly as the cecal microflora adapts. The rate of adaptation also depends on the origin of the resistant starch. Rats fed resistant starch from normal maize adapted quickly (5 days), losing only a few percent of the ingested resistant starch whereas those fed resistant starch from wrinkled peas or high-amylose maize adapted more slowly (28 days) and were still losing a high percentage of the ingested resistant starch. Resistant starch may therefore contribute significantly to fecal bulk but this is variable depending on the state of adaptation of the animal. (*See* **Microflora of the Intestine**: Role and Effects.)

Significance

At this stage it is difficult to estimate the significance of resistant starch. It is not well defined, and proposed methods of analysis may determine fractions that possess different structures and functions or that may be physiologically irrelevant.

What can be said relates mainly to the properties of resistant starch by analogy to dietary fiber and, by extension, its deliberate production, either within a food or as a food additive. Where resistant starch is present in a food it will help to flatten the glycemic response curve by reducing the amount of rapidly digestible starch and by moderating the rate of digestion. The resistant starch that reaches the more distal part of the small intestine may exert an influence on the rate of transport along the ileum, and help the more uniform absorption of carbohydrate. This type of 'slow-release' carbohydrate may find application in improving the carbohydrate tolerance of diabetics, and it has implications for the utilization of energy. (*See* **Glucose**: Glucose Tolerance and the Glycemic(-Glycaemic) Index.)

Any starch that is lost from the small to the large intestine represents a loss of dietary energy since, even if it is fully fermented, the energy salvaged is only about 50% of that in the original carbohydrate. Some additional losses from the small intestine of other energy sources may occur as a result of any 'carry-over' effect.

The energy value of resistant starch is therefore somewhat less than $16.8\,\mathrm{kJ\,g^{-1}}$; its actual value depends upon how much can be salvaged by fermentation in the large intestine, the pattern of short-chain fatty acids produced, and any energy lost to hydrogen or methane.

The fermentation of dietary fiber in the large intestine is believed to be a protective factor against large-bowel cancer because the short-chain fatty acids produced lower the pH and inhibit 7α-dehydroxylase. This prevents the formation of secondary bile acids that are suspected of a promotional role in colonic carcinogenesis. In addition, butyrate, produced by fermentation, is known to be a major energy source for the colonic tissue and is believed to be a protective factor in the development of colon cancer.

Carbohydrate fermentation in the large intestine may therefore be a major protective factor against the development of colon cancer. Because resistant

starch is at least partially fermentable, it may play a crucial role in these mechanisms. (*See* **Colon**: Cancer of the Colon.)

Resistant starch has also been demonstrated to have fecal-bulking properties and this could also play a role in the dilution of fecal carcinogens by reducing the time and level of exposure of the colonic mucosa.

From the energy point of view, there may be some health benefits from consuming foods high in resistant starch. Foods in which resistant starch may be created *in situ* would have a lower energy density but otherwise maintain their nutrient density. Starches high in resistant starch or isolated resistant starch fractions may be created with suitable functional characteristics to find application in the food industry as an ingredient or simply as a lower-energy bulking agent.

See also: **Bananas and Plantains**; **Colon**: Cancer of the Colon; **Dietary Fiber**: Properties and Sources; Determination; Physiological Effects; **Glucose**: Glucose Tolerance and the Glycemic (Glycaemic) Index; **Microflora of the Intestine**: Role and Effects; **Potatoes and Related Crops**: Processing Potato Tubers

Further Reading

Asp N-G, van Amelsvoorth and Hautvast JGAJ (1996) Nutritional implications of resistant starch. *Nutrition Research Reviews* 9: 1–31.

Englyst HN and Kingman SM (1990) Dietary fibre and resistant starch. A nutritional classification of plant polysaccharides. In: Kritchevsky D, Bonfield C and Anderson JW (eds) *Dietary Fibre*, pp. 49–65. New York: Plenum Press.

Englyst HN, Wiggins HS and Cummings JH (1982) Determination of the non-starch polysaccharides in plant foods by gas–liquid chromatography of constituent sugars as alditol acetates. *Analyst* 107: 307–318.

Faulks RM, Southon S and Livesey G (1989) Utilisation of α-amylase (EC.3.2.1.1) resistant maize and pea (*Pisum sativum*) starch in the rat. *British Journal of Nutrition* 61: 291–300.

Gee JM, Faulks RM and Johnson IT (1991) Physiological effects of retrograded α-amylase resistant cornstarch in rats. *Journal of Nutrition* 121: 44–49.

Livesey G, Davies IR, Brown JL, Faulks RM and Southon S (1990) Energy balance and energy value of α-amylase resistant maize and pea (*Pisum sativum*) starches in the rat. *British Journal of Nutrition* 63: 467–480.

Miles MJ, Morris VJ, Orford PD and Ring SG (1985a) The roles of amylose and amylopectin in the gelation and retrogradation of starch. *Carbohydrate Research* 135: 271–281.

Miles MJ, Morris VJ and Ring SG (1985b) Gelation of amylose. *Carbohydrate Research* 135: 257–269.

Orford PD, Ring SG, Carrol V, Miles MJ and Morris VJ (1987) The effect of concentration and botanical source on the gelation and retrogradation of starch. *Journal of the Science of Food and Agriculture* 39: 169–177.

Ring SG, Colonna P, l'Anson KJ *et al*. (1987) The gelation and crystallisation of amylopectin. *Carbohydrate Research* 162: 277–293.

Ring SG, Gee JM, Whittam M, Orford P and Johnson IT (1988) Resistant starch: its chemical form in foodstuffs and effect on digestibility *in-vitro*. *Food Chemistry* 28: 97–109.

Zobel HF (1988) Molecules to granules: a comprehensive starch review. *Starch* 40(2): 44–50.

STARTER CULTURES

L Durso and R Hutkins, University of Nebraska, Lincoln, NE, USA

Introduction and History

Starter cultures are an essential component of nearly all commercially produced fermented foods. Simply defined, starter cultures consist of microorganisms that are inoculated directly into food materials in order to bring about desired and predictable changes in the finished product. These changes may include enhanced preservation, improved nutritional value, modified sensory qualities, and increased economic value. Although many fermented foods can be made without a starter culture, the addition of concentrated microorganisms, in the form of a starter culture, provides a basis for insuring that products are manufactured on a consistent schedule, with consistent product qualities.

Fermented foods and beverages have long been manufactured without the use of commercial starter cultures. Traditional methods of production include backslopping, or using a small amount of the finished specifically preserved product to inoculate a new batch, the use of microorganisms found naturally on the product, and the use of special containers that allow for the survival of the starter culture microorganisms within cracks and pores. These traditional

methods allow for the development of individual varieties of fermented foods and beverages, and they are still practiced today for small- to mid-scale production facilities, as well as in less developed countries and in homemade-type products. Traditional methods, however, are prone to slow or failed fermentations, contamination, and inconsistent quality. In contrast, modern large-scale industrial production of fermented foods and beverages demands consistent product quality and predictable production schedules, as well as stringent quality control to insure food safety.

Given that pure culture techniques in microbiology were not developed until Pasteur in the 1860s and that lactic acid bacteria (LAB), in particular, were identified by Lister in the 1870s, it is noteworthy that an industry devoted to producing pure cultures had its beginnings only a short time later. In the late 1880s, Storch in Denmark, Weigman in Germany, and Conn in the USA showed that pure cultures could be used to ripen cream, and soon the role of flavor-producing bacteria (i.e., citrate-fermenting diacetyl-producers) was established. By 1878 Christian Hansen began a culture business that continues even today to be a major supplier of starter cultures for the dairy, meat, brewing, baking, and wine industries.

Initially, starter strains were prepared by the manufacturer by growing pure strains in heat-sterilized milk. Calcium carbonate was often added as a buffer in order to maintain a neutral pH. These liquid cultures remained popular until relatively recently, even though they had a relatively short shelf-life due to the loss of cell viability and fermentative activity. Eventually, rather crude dry culture preparations were produced which required several transfers in milk to revive the culture to an active state. Freeze-dried cultures also became available, but the early product also required growth in intermediate or mother cultures. Frozen cultures, now the most common form for dairy cultures, were not introduced until the 1960s. Significant improvements in freezing and freeze-drying technologies have led these types of cultures to dominate the starter culture market.

The modern starter culture industry provides cultures for nearly every type of fermented food and beverage. Most culture houses also produce and sell the media used to propagate starter cultures, the enzymes used to coagulate milk for cheese production, and other ancillary products. In addition, the starter culture companies maintain a staff of highly trained microbiologists who provide expert technical service and support when issues or problems related to culture performance arise. And although there are many small culture manufacturers throughout the world, specializing in cultures for specific products or applications, the industry is dominated by a small number of large companies.

This article will review the types of microorganisms used as starter cultures, how starter cultures are produced and preserved, criteria used by manufacturers to assess culture performance, and ultimately how these cultures are used by the fermented foods industry. Issues facing the starter culture industry and the new technologies being developed to address these issues will also be examined. Throughout this chapter, dairy starter cultures will be emphasized, since the cheese and cultured dairy products industry represents the largest user of commercial starter cultures.

Starter Culture Microorganisms

Given that fermented foods have historically been made using various species of bacteria, yeasts, and molds, it follows that modern starter cultures also contain these organisms. In general, bacteria are used in the production of cheese, cultured dairy products, fermented sausages, and fermented vegetables. Yeasts are used in the manufacture of bread and alcoholic beverages, and molds are used for cheeses and sausages as well as for soy-derived products, such as soy sauce and tempeh. In most cases, the organisms contained within a starter culture preparation are well-defined, often to the species or even strain level, and are carefully selected based on criteria relevant for a particular product. Some starter culture organisms, however, are not well defined, but instead are selected based on a history of successful use.

Bacterial Starter Cultures

LAB are clearly the most important group of bacteria used as starter organisms (*See* **Lactic Acid Bacteria**). The LAB consist of a cluster of low G + C Gram-positive cocci and rods from at least 11 genera; however, only a few of these are used as starter cultures (Table 1). *Lactococcus lactis* subsp. *lactis* has long been considered the most important LAB, due to its widespread use in cheese manufacture, but other species, such as *Streptococcus thermophilus*, which has gained in importance due to its use in yogurt and Mozzarella cheese, are also widely used (Figure 1).

In general, the LAB are catalase-negative heterotrophs having complex nutritional requirements. They have a fermentative metabolism, deriving energy, in the form of adenosine triphosphate (ATP), via substrate-level phosphorylation. Sugar metabolism is either homofermentative or heterofermentative. In homofermentation, nearly all (> 90%) of the

sugar substrate is converted directly to lactic acid, whereas in heterofermentation, lactic acid, acetic acid, CO_2, and ethanol are produced. Although most genera are either one or the other, some species of *Lactobacillus* have the biochemical capacity for both pathways. Both hetero- and homofermentative LAB are used as starter cultures. Their optimum temperature for growth varies; however, most LAB used as dairy starter cultures are either mesophilic (*Lactococcus lactis*) or are moderate thermophiles (*S. thermophilus* and *Lactobacillus* spp.).

Table 1 Lactic acid bacteria used as starter cultures

Organism	*Application*
Lactococcus lactis subsp. *lactis*	Cheese, cultured dairy products
Lactococcus lactis subsp. *cremoris*	Cheese, cultured dairy products
Lactococcus lactis subsp. *lactis* biovar. *diacetylous*	Cheese, cultured dairy products
Lactobacillus helveticus	Cheese, cultured dairy products
Lactobacillus delbrueckii subsp. *bulgaricus*	Cheese, yogurt
Lactobacillus sanfranciscensis	Sourdough bread
Lactobacillus casei	Cheese, cultured dairy products
Lactobacillus sakei	Sausage
Lactobacillus plantarum	Sausage, fermented vegetables
Lactobacillus curvatus	Sausage
Lactobacillus sanfrancisco	Sourdough bread
Streptococcus thermophilus	Cheese, yogurt
Pediococcus acidilactici	Sausage, fermented vegetables
Pediococcus pentosaceus	Sausage
Pediococcus halophilus	Soy sauce
Oenococcus oeni	Wine
Leuconsotoc lactis	Cheese, cultured dairy products
Leuconsotoc mesenteroided subsp. *cremoris*	Cheese, cultured dairy products, fermented vegetables

Since the primary function of starter culture bacteria is to ferment sugars and to produce acids, the ability of LAB to metabolize carbohydrates is of critical importance. Indeed, the specific substrate range and rate at which metabolism occurs may be used as criteria for strain selection. In dairy fermentations, for example, rapid lactose fermentation is required, whereas in sourdough fermentation, maltose and glucose metabolism are relevant and in sausage fermentation glucose or sucrose metabolism is most important. In addition, these bacteria are often selected based on their ability to perform other metabolic functions. In cheese, degradation of proteins via proteolytic and peptidolytic enzymes is essential to develop proper aged cheese flavor and texture. In sour cream and cultured buttermilk manufacture, the ability of the starter culture bacteria to utilize citrate and to form the aroma compound diacetyl is required. Production of exopolysaccharides by strains of *S. thermophilus* is desirable in yogurt in order to impart greater viscosity. In wine-making, the conversion of malic acid to lactic acid by starter malolactic

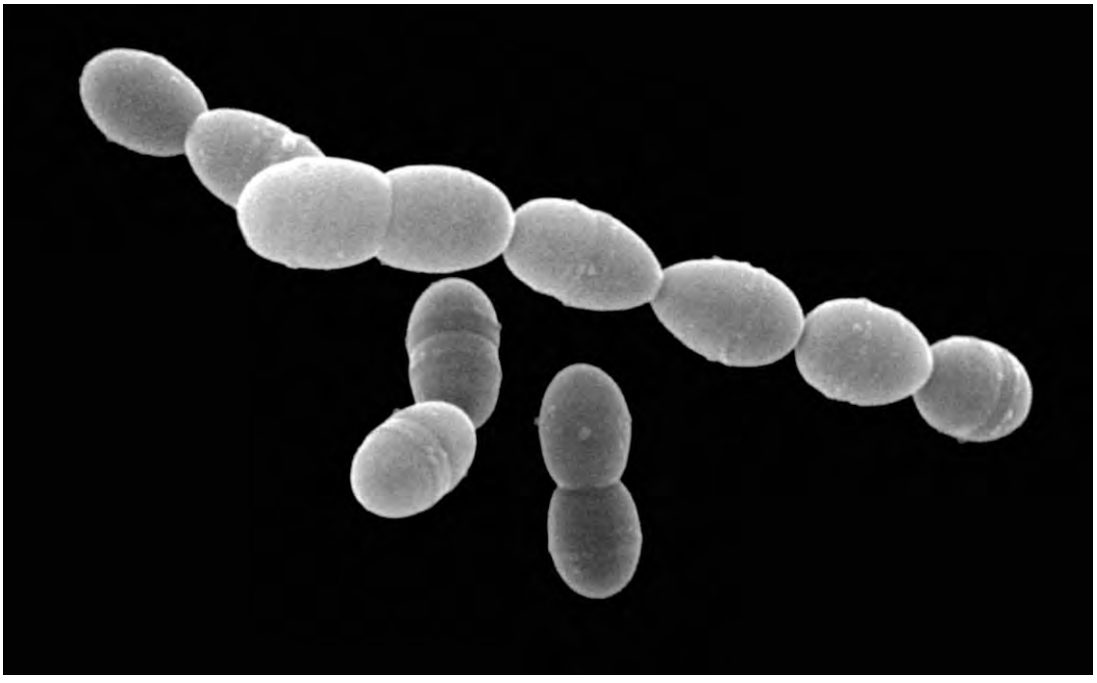

Figure 1 Electron micrograph of *Streptococcus thermophilus*.

bacteria is necessary to reduce the acidity of certain grapes.

Although LAB are clearly the most widely used and most important group of bacteria used as starter cultures in fermented foods, other bacteria are also used (Table 2). In the cheese industry, *Propionibacterium shermanni* and *Brevibacterium linens* are used in the Manufacture of Swiss (and related varieties) and Limburger, Muenster, and Brick cheeses, respectively (*See* **Cheeses**: Types of Cheese; Starter Cultures Employed in Cheese-making; Chemistry of Gel Formation; Chemistry and Microbiology of Maturation; Manufacture of Extra-hard Cheeses; Manufacture of Hard and Semi-hard Varieties of Cheese; Cheeses with 'Eyes'; Soft and Special Varieties; White Brined Varieties; Quarg and Fromage Frais; Processed Cheese; Dietary Importance; Mold-ripened Cheeses: Stilton and Related Varieties; Surface Mold-ripened Cheese Varieties). Some fermented sausage manufacturers add *Micrococcus* spp. to meat batters in order to develop flavor and color attributes (*See* **Fermented Foods**: Fermented Meat Products). The manufacture of vinegar requires *Acetobacter aceti*, which oxidizes the ethanol generated from alcohol fermentations. (*See* **Vinegar.**)

Yeast Starter Cultures

Nearly all alcoholic beverages are made using one or more strains of yeast (Table 2). Although many of the wines produced around the world are made via natural fermentations, relying on yeasts ordinarily present on the processing equipment or on the grape surface, the trend for most wine manufacturers has been to use starter culture yeasts, mainly strains of *Saccharomyces cerevisiae* (*See* **Wines**: Types of Table Wine; Production of Table Wines; Production of Sparkling Wines; Dietary Importance; Wine Tasting). Like LAB, strain selection is based primarily on the desired attributes of the final product. In addition, traits such as the ability to flocculate (and, hence, be easily separated from the wine), to grow at high sugar concentrations, and to produce adequate ethanol levels, are all considered. Yeast starter cultures are also used by the brewing industry, and although many large breweries maintain their own proprietary cultures, commercial ale (*S. cerevisiae*) and lager (*S. carlsbergensis*) cultures are widely used (*See* **Beers**: History and Types; Wort Production; Raw Materials; Chemistry of Brewing; Biochemistry of Fermentation; Microbreweries). Again, strain selection is based on characteristics relevant to the needs of the particular brewer, and these include flocculation, flavor development, and ethanol production rates. Bread manufacturers represent the other large user of yeast starter cultures (*See* **Bread**: Dough Mixing and Testing Operations; Breadmaking Processes; Chemistry of Baking; Sourdough Bread; Dietary Importance; Dough Fermentation). Yeast starter cultures used for bread-making are optimized for the production of CO_2 from sugar or starch, whereas brewing strains used in the production of alcohol are generally optimized for the production of ethanol.

Table 2 Other organisms used as starter cultures

Organism	*Application*
Bacteria	
Brevibacterium linens	Cheese: pigment, surface
Propionibacterium freudenreichii subsp. *shermanii*	Cheese: eyes in Swiss
Staphylococcus carnosus subsp. *carnosus*	Meat: acid, flavor, color
Mold	
Penicillium camemberti	Cheese: surface ripens white
Penicillium roqueforti	Cheese: blue veins, protease, lipase
Penicillium chrysogenum	Sausage
Aspergillus oryzae	Soy sauce, miso
Rhizopus microsporus subsp. *oligosporus*	Tempeh
Yeast	
Saccharomyces cerevisiae	Bread: carbon dioxide production
Saccharomyces cerevisiae	Ale beers
Saccharomyces carlsbergensis	Lager beers
Saccharomyces cerevisiae	Wine

Mold Starter Cultures

Mold cultures are used for several types of cheeses, including all varieties of blue cheese, as well as the white surface mold-ripened cheeses typified by Brie and Camembert (Table 2). Blue mold cheeses are made using spore suspensions of *Penicillium roqueforti* and white mold cheeses are made using *P. camemberti*. Mold cultures are also used in the production of so-called Asian or soy-derived fermented foods. Examples include tempeh, made using *Rhizopus microsporus* subsp. *oligosporus*, and miso and soy sauce, made using *Aspergillus oryzae*. Fungal starter cultures are also occasionally used for the production of European-style sausages and hams.

Types of Starter Cultures

One of the first requirements of a starter culture is that it rapidly initiates a fermentation. Thus, the initial inoculum must contain large numbers of microorganisms. As the volume of the food (or liquid) increases, then either larger starter culture volumes or greater

starter culture cell concentrations are required. Assuming that 1% inocula is ordinarily used, 10 kg (or 10 l) of starter culture would be sufficient to inoculate 1000 kg (or 1000 l) of substrate. Such an inoculum could easily be produced from a colony or culture slant, simply by successive transfers through intermediate cultures (i.e., 1 ml into 100 ml, and 100 ml into 10 l). Indeed, when pure cultures, rather than backslopping techniques, were introduced nearly a century ago in the cheese industry, the propagation of a 'mother' culture and intermediate cultures was the normal practice. However, as the size of the cheese industry increased, such that starter culture volumes of over 10 000 kg could be required for a single day's production, it no longer became feasible for cheese manufacturers to prepare cultures in this manner.

Currently, there are two general types of cultures that are manufactured and sold to the fermented food and beverage industries. The first type, often referred to as 'bulk cultures,' is used to inoculate a bulk tank, which is then used to inoculate the production vat (Figure 2). The bulk culture essentially represents the intermediate cultures that have traditionally been required. After a suitable incubation period in the appropriate culture medium, the fully grown bulk culture is used to inoculate the raw material. Provided that the starter culture organisms are protected against acid damage, oxygen, hydrogen peroxide, or other inhibitory end products, bulk cultures will remain viable for many hours. Thus, in the cheese industry, bulk cultures can be used to inoculate multiple cheese vats throughout a manufacturing day. Maintenance of culture viability is an important issue, however, and methods to insure that cultures remain active will be discussed below.

The second type of starter is designed to be inoculated directly into the food material, thus eliminating the need to grow bulk cultures. In the cheese industry, these are often referred to as 'direct-to-vat set' or DVS cultures. Of course, such cultures are also widely used in the sausage, baking, and other industries, where product volumes are usually small, compared to cheese manufacture. These cultures have the obvious advantage of eliminating the labor, hardware, and accessory problems that accompany the preparation and maintenance of bulk starter cultures. In order to provide a sufficient inoculum into large volumes of raw material, however, DVS cultures must be highly concentrated, especially those used for cheese manufacture. The added concentration steps may reduce culture viability, causing slow-starting fermentations. Direct-to-vat cultures are also more expensive to purchase than bulk cultures, and may not be economical for large-scale operations.

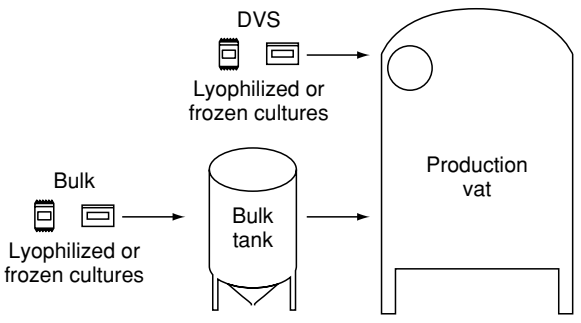

Figure 2 Bulk versus direct-to-vat set (DVS) culture preparation.

Mixed versus Defined Cultures

Mixed cultures contain historically tested blends of starter culture organisms. They can include several genera, species, or even strains of organisms. Often the actual identities of the organisms present in a mixed culture are not known, and the individual species may not have been characterized microbiologically or biochemically. Moreover, the proportion of different organisms in a mixed culture may not be constant from one product lot to another. Yet, mixed cultures containing undefined strains are still used as starter cultures for many applications because they have a proven history of successful use. When used in the manufacture of cheese, strains that are sensitive to a specific bacteriophage may quickly die out. However, other strains present in the mixed culture may be resistant to that particular phage and can then complete the fermentation. Frequent exposure to different bacteriophages, therefore, provides an effective mechanism for insuring that phage-resistant strains will be present in repeatedly propagated mixed cultures. These so-called P starter cultures are commonly used in the Netherlands for the manufacture of Gouda, Edam, and related cheese varieties.

The main disadvantage of mixed cultures is that they may yield finished products of inconsistent quality. In addition, fermentation rates may vary from day to day, affecting production schedules. For applications where time is flexible, and product quality variations are more tolerable, such as in small-scale cheese factories, mixed cultures are still widely used. However, in large production facilities where precise schedules are critical and consistent product quality is expected, mixed cultures have become less common. Instead, defined starter cultures have become prevalent.

Defined cultures simply refer to cultures that contain microbiologically characterized strains, which

are used individually or as blends. Most defined strains have been isolated from wild or mixed cultures, and characterized for metabolic properties, phage-resistance, and other desirable traits. When assembling a culture blend (containing multiple defined strains), all strains must be compatible. That is, one strain must not dominate over the others or produce inhibitory agents that would affect the growth of other organisms. Blends are often preferred over single or paired strains for their predictability and consistency. They are formulated to contain a broad spectrum of strains having resistance to different phage types, so that if one strain is affected, the other strains can complete the fermentation.

Manufacture of Starter Cultures

Bacterial starter cultures, including those produced for the dairy industry, are mass-produced in modern fermenters, under aseptic conditions, not unlike those used in the pharmaceutical industry for the production of biomedical products. Although the size of the fermentors may vary from as little as 10 l to over 300 l, the basic operational parameters and control features are consistent. For production of dairy starter cultures, milk- or whey-based media are most often used; however, molasses or corn syrup can also be used as the basal medium. Nutrients are usually added to the medium. Vitamins, especially water-soluble B-vitamins, are required for optimum growth of lactococci and lactobacilli, and some species of *Streptococcus, Leuconostoc*, and *Lactobacillus* also require specific amino acids. In addition, stability of some lactobacilli species during frozen storage may be enhanced when the growth medium contains Tween 80, a surfactant thought to stabilize cell membranes. Acids produced during culture growth can reduce final cell densities and decrease cell viability; thus the pH is controlled by the addition of alkali, usually gaseous NH_3, NH_4OH, Na_2CO_3, or KOH. Catalase is often added to degrade hydrogen peroxide that is generated by growth of some LAB, and oxygen incorporation is minimized. Cells are typically grown to achieve cell densities of 10^9 to 10^{10} cells per ml, and then packaged in a liquid form, concentrated or lyophilized (freeze-dried). LAB are ideally harvested either at late log phase or early stationary phase; however, the optimum harvest time depends on the specific organism. Liquid cultures are packaged in cans having volumes ranging from 100 to 500 ml, then rapidly frozen in liquid nitrogen ($< -196\,^{\circ}C$). It is important that the cultures remain frozen ($< -40\,^{\circ}C$) throughout storage, as well as during transportation, since freezing and thawing lead to the formation of ice crystals, which can puncture and kill cells. Thawing can occur even when the cells are held below $0\,^{\circ}C$, and so it is important to put frozen cultures promptly into a $-40\,^{\circ}C$ or $-70\,^{\circ}C$ freezer as soon as they arrive at the plant.

To produce frozen direct-to-vat cultures, cells are prepared via high-cell-density fermentations (as above), and then must be concentrated prior to freezing. Either continuous centrifugation or cross-flow membrane filtration is used. A washing step may be included to remove the spent medium. Cell densities of concentrated cultures can be as high as 10^{11} to 10^{12} cells per gram. Typically, a single can (usually 360 ml) can be used to inoculate 2000 kg of milk.

For lyophilization, cells are grown and harvested, as above, then freeze-dried under vacuum. Freeze drying is simply a process whereby water is removed from the frozen material by sublimation (*See* **Freeze-drying**: The Basic Process; Structural and Flavor (Flavour) Changes). It provides a means not only for concentrating cultures, but also preserves them due to the low water activity that is achieved. Freeze drying is a far gentler process compared to other drying technologies, such as spray drying. Freeze-dried cultures can contain from 10^9 to 10^{12} cells per gram. The cells are usually packaged in foil pouches or other oxygen-impermeable material. Lyophilized cells are best maintained at $-20\,^{\circ}C$, but are somewhat stable even at room temperature. Although lyophilized cultures have become popular as DVS starter cultures for cheese, yogurt, and cultured milk products, cells may be sluggish after inoculation, with a longer lag phase. To maintain cell viability for both frozen and lyophilized cells, cryoprotectant agents are usually added. Common cryoprotectant agents include glycerol, lactose, sucrose, trehalose, ascorbate, and glutamate. It is important to note that freezing and lyophilization parameters vary for different lactic acid starter bacteria. For example, *L. lactis* subsp. *lactis* is far more tolerant to freezing and lyophilization than *L. lactis* subsp. *cremoris*.

Criteria for Evaluating Culture Performance

In general, the job of most starter cultures is to carry out the desired fermentation, promptly and consistently, and to produce products having the appropriate level of fermentation end products and desired flavor and texture. The particular requirements for a given strain, however, depend entirely on the application for that culture (Table 3). For example, LAB used as dairy starter cultures are selected based not only on lactic acid production rates, but also on flavor- and texture-producing properties, salt-sensitivity, phage resistance, and durability during production and

Table 3 Desirable properties of starter cultures

Culture	*Property*
Dairy cultures	Lactic acid production rates
	Lag phase times
	Phage resistance
	Ease of manufacture
	Stability and consistency
	Produce desired flavor and texture
	Preservation tolerance
	Produce
	Lack of off-flavors
Meat cultures	Fast acidification
	Produce desired flavor
	Antimicrobial activity
Beer cultures	Rapid fermentation
	Produce desired flavor
	Preservation tolerance and stability
	Flocculation
	Lack of off-flavors
	Proper attentuation
	Growth at wide temperature range
	Tolerant to osmotic, temperature, and handling stresses
Wine cultures	Osmotolerant
	Ethanol-tolerant
	Flocculation, sedimentation
	Growth at low temperature
	Produce consistent flavor
	Malolactic fermentation
Bread cultures	Freeze-tolerant
	Produce desired flavor
	Produce adequate leavening

storage. The strains of *Streptococcus thermophilus* and *Lactobacillus delbrueckii* subsp. *bulgaricus* used as yogurt starter cultures, for example, are selected based, in part, on acid development rates, acetaldehyde production, exopolysaccharide synthesis, and lack of postfermentation acidity. However, different strains of these bacteria may be used for Mozzarella and Swiss cheese manufacture, since the latter products have their own requirements.

To assess functional properties of dairy cultures, several tests are routinely performed. Acid-production rates are easily determined by inoculating heat-treated milk with a suitable (and standardized) inoculum (e.g., 0.1–3.0%), incubating the milk at a given temperature, and then either measuring the decrease in pH as a function of time or simply by determining the pH or titratable acidity after a specific incubation period. Lag times can also be estimated and are useful in determining the effects of frozen or lyophilized storage on culture activity. Another critical test is determination of bacteriophage resistance, especially under conditions that simulate cheese making. In the Heap–Lawrence challenge test, cells are inoculated into culture tubes of milk, representative or factory phage mixtures are added, and the tubes are incubated according to a cheese production time–temperature profile. After incubation, the entire process is repeated up to six times. An increase in pH (less acid produced) in successive cycles indicates that the culture has become sensitive to bacteriophage and should be removed from the rotation.

Yeast cultures used for wine, beer, and bread should ferment rapidly and have good sensory properties and good storage stability. In addition, wine yeasts and beer yeasts should sediment or flocculate to enhance their separation and to reduce autolysis. Flocculation is a heritable trait, requiring several genes that encode for proteins involved in cell aggregation. Flocculated cells sediment much faster relative to the free cells. Other performance characteristics will depend on the specific product. Many wine yeasts, for example, are selected based on growth at low temperatures (10–14 °C), resistance to sulfiting agents, and high ethanol tolerance. Beer yeasts should also be able to grow at low temperature, produce flavorful end products, and deplete fermentable carbohydrates (good attenuation). Bread yeasts should withstand freezing (so that frozen bread doughs will rise after thawing).

Using Starter Cultures

As described above, many starter culture preparations, whether in a frozen or lyophilized form, are simply added directly into the food material. Frozen canned cultures are routinely thawed in cold chlorinated water immediately prior to use. Lyophilized cultures may require additional stirring or mixing in the vat to facilitate hydration. For bulk cultures, additional steps are required to insure that the organisms have reached sufficient levels and viability. In the dairy industry, specialized culture media are almost always used for bulk starter preparation, not only to provide optimum growth conditions, but also to protect the cells from acid damage and from infection by lytic bacteriophage. The bulk culture media is not much different from the media used by the culture supplier to mass-produce starter culture cells. A basal medium containing a fermentable carbohydrate (lactose, glucose, or sucrose) is supplemented with additional sources of nitrogen, vitamins, minerals, and other nutrients. Commercially available yeast extract is an excellent source of many of these materials. Phosphate salts are frequently added to provide buffering as well as chelating properties.

Controlling pH during bulk culture preparation is especially important since most LAB grow best above pH 6.0, and can lose viability if held at pH values below 5.0. The use of acid-injured starter cultures in cheese manufacture will result in a long lag phase and

prolonged fermentation times. Several notable approaches have been developed to minimize or prevent acid damage to lactic starter cultures. The first approach, known in the industry as 'external pH control,' relies on the addition of alkaline solutions, such as ammonium hydroxide, or ammonia gas, into the bulk culture tank during fermentative growth. The tank is fitted with a pH electrode, and a pH-monitoring device that sends a signal to a pump when the pH has decreased below a critical threshold. Alkali is then added until an upper pH limit is reached. Although the range at which the pH is controlled depends on the specific organisms, most external control systems maintain a pH between 5.8 and 6.2. Thus, as cells grow and produce acid, the pH never falls below the set threshold. Such systems yield starter cultures having high cell densities (as much as 10 times more cells) with enhanced viability. Lactate salts that accumulate as a byproduct result in higher media osmolality and ionic strength, but these effects do not appear to have any significant negative effects on cell viability. Due to the hardware expenses (pH electrodes, pumps, pH-monitoring devices) inherent in external pH control systems, however, a less expensive alternative is simply to manually neutralize the bulk culture tank. This is done by a one- or two-time addition of neutralizing agent (sodium or potassium hydroxide) that raises the pH from around 5.0 to near 7.0.

Although most culture media ordinarily contain phosphate or citrate salts that provide some buffering, it is not possible to add enough of these salts to maintain a near-neutral pH during growth by LAB. This is because high phosphate concentrations are inhibitory to some LAB used as starter cultures. Media can, however, still be formulated to contain buffer agents provided that the buffers are either in an insoluble or encapsulated form. These so-called internal pH control systems typically contain carbonate or phosphate salts (e.g., sodium carbonate or trimagnesium phosphate) that are released into the medium as a function of low pH, such that the pH of the culture medium is maintained above 5.0. Internal pH control systems have the added advantage of not requiring the expensive external pH control apparatus. Agitation is required, however, to prevent the buffer agents from settling out during bulk culture growth.

Although equipment and labor costs for external pH control systems and for internal pH control media are higher than those for conventional culture preparation systems, several advantages are realized. Since higher cell densities are achieved, and cell viability is enhanced, less culture is needed. In addition, because the cells are maintained in a viable state, they can be used for a longer period of time.

Bacteriophage

Bacteriophage are viruses that can infect and eventually lyse bacteria (**Figure 3**). They can multiply rapidly and decimate the starter culture. Bacteriophage problems during bulk culture production are relatively rare because culture media are heated to a high temperature and aseptic conditions (i.e., closed, sterilizable vats) are usually employed. However, milk used for cheese manufacture receives only a modest heat treatment, and cheesemaking, in general, is an 'open' process, providing ready access by air- or whey-borne phage. Thus, in most modern cheese production factories, where vats are filled and refilled during the production day, phage may have ample opportunity to propagate and to reach high levels, resulting in slow or arrested fermentations and poor-quality product. Until recently, lactococcal bacteriophage were the main concern; however, increased production of Mozzarella cheese, yogurt, and other products that rely on thermophilic starter cultures has led to the proliferation of phages against *S. thermophilus* and *L. bulgaricus*.

For nearly 50 years, problems caused by bacteriophage have been the driving force for much of the

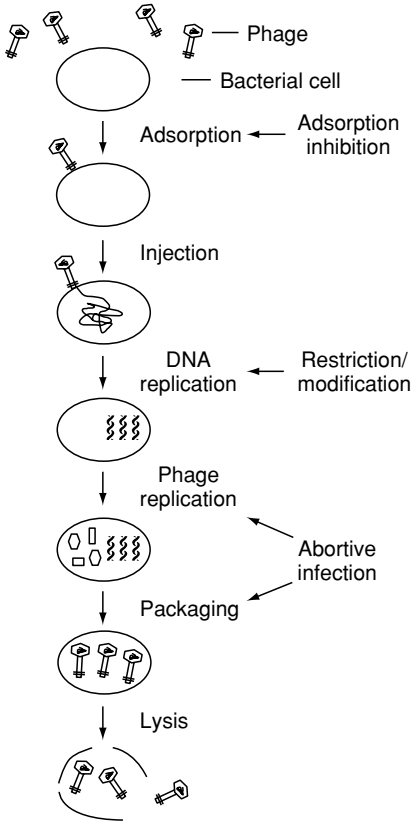

Figure 3 Lytic cycle of lactic bacteriophage and steps at which lactococcal phage resistance mechanisms operate.

research on starter cultures. The dairy starter culture industry and the cheese industry have adopted several strategies to prevent or reduce the impact of bacteriophage problems. Strict sanitation and attention to plant design are the first lines of defense against phage. Ideally, rooms used to 'step up' starter cultures to the concentrations needed for production are isolated from the rest of the processing facility. Since phage are commonly transmitted via air, these contained areas should be held under positive pressure, and filtered air should be used both for the intake into the starter room, as well as into the starter tanks. The production rooms should be designed such that 'downstream' product and waste flow do not contaminate the starter preparation room.

Phage-inhibitory media have been embraced as another means of phage control. Phage require calcium ions to attach to, and subsequently invade, their bacterial hosts. In 1956, the US Department of Agriculture introduced the idea of using phosphates in culture media to bind calcium, thus making the calcium ions unavailable for the phage. In addition to their role in binding calcium, the phosphate salts in phage-inhibitory media provide buffering capacity. Citrate salts provide similar protection. Phage-control strategies are summarized in Table 4.

Most starter culture strains are only sensitive to a specific bacteriophage, and some strains are naturally resistant, due to one or more of several types of phage-resistant mechanisms (Figure 3). In addition, sensitive strains can become resistant by frequent exposure to lytic phage (bacteriophage-insensitive mutants or BIM) or by genetic manipulation. If the phage-resistant variants still have good cheese-making properties (which must be tested), then these strains can be introduced as a starter culture, usually as one of several other defined phage-resistant strains. As described above, multiple defined strain starters contain as many as five different strains, selected largely on the basis of being phage-unrelated. These cultures can be used on a continuous basis providing consistent product quality. It is important, however, to monitor phage levels in the cheese whey on a regular daily basis, so that strains that become phage-sensitive can be removed from the mixture and replaced by a resistant strain.

Additional protection against phage can be achieved by rotating cultures, meaning that a single culture (containing as many as five or six different strains) is used on a given day and then followed by successive cultures containing strains whose phage-sensitivity pattern is different from the preceding culture. Thus, since phage proliferation requires susceptible host strains, removing those strains from the cheese production environment will reduce the background level and accumulation of phages such that normal fermentation rates can be achieved. However, this practice is constrained by the limited availability of phage-unrelated strains, as well as by less consistent product quality. Many of the defined, phage-resistant cultures now contain only two or three strains.

New Developments in Starter Culture Technology

The pioneering work of McKay and coworkers led to the discovery that most of the phenotypic traits necessary for lactic starter cultures, including lactose fermentation, casein hydrolysis, and diacetyl formation, were encoded by plasmid DNA. The Klaenhammer group subsequently demonstrated that phage-resistant determinants were also plasmid-encoded. These discoveries, and the development of gene-exchange and gene-transfer techniques, made it possible to manipulate the genetic properties of lactic starter cultures (*See* **Biotechnology in Food Production**). However, because of regulatory and consumer concerns regarding genetic modification of foods, the applications of these technologies have generally been restricted to noninvasive approaches. (*See* **Genetically Modified Foods.**)

Engineering Phage Resistance

Of particular importance has been the construction of genetically modified, phage-resistant lactococcal strains. In the early 1980s, the Klaenhammer group identified a 46-kb plasmid in *Lactococcus lactis* subsp. *lactis* that encoded for a heat-sensitive abortive infection mechanism and that conferred resistance to specific bacteriophage. This self-transmissible

Table 4 Phage-control strategies

Method	*Purpose*
Sanitation	Kill and remove phage in plant environment
Plant design	Keep phage out of production area, prevent cross-contamination
Phage-inhibitory media	Prevent phage from attaching to and infecting culture cells
Phage-resistant cultures	Design starter cultures cells that will grow and perform well even in presence of phage
Culture rotation	Prevent proliferation of phage by limiting access to suitable host

plasmid, named p TR2030, was later found to contain other regions, including genes coding for a restriction and modification system, that also conferred bacteriophage resistance. Transfer of the plasmid into a phage-sensitive, cheesemaking strain via conjugal mating resulted in fast lactose-fermenting transconjugants having resistance to a broad range of lytic industrial phages. Although this approach was successful for a time, phages eventually appeared that could overcome the host resistance. In fact, due to the dynamic nature of phage–host interactions and the nonsterile manner in which cheese manufacturing occurs, the propensity of lytic phages to adapt to and ultimately appear against formerly resistant strains is nearly inevitable.

Several recent strategies have since been described to provide more long-term phage protection. One approach is simply to introduce or 'stack' two plasmids, each carrying phage resistance mechanisms against different phage types, into a single strain. In contrast, different phage-resistant genes can be introduced individually into a single strain generating several isogenic phage-resistant derivatives that can then be used in a rotation scheme. Since each strain harbors genes effective against different phage types, the same parental strain can be used on a continuous basis. Another system, called Per, relies on the integration of a phage DNA fragment (containing the origin of replication region or *ori* site) into the genome of the host strain. Ordinarily, when a bacterium is infected by bacteriophage, phage DNA replication begins at the *ori* site. However, in the Per strains, phage development is stalled, since the chromosomal copy of the *ori* site acts as a decoy for the phage replication machinery. Antisense RNA strategies function in a similar manner, in that the modified cell is directed to transcribe an antisense messenger RNA that is complementary and capable of hybridizing to the normally transcribed RNA. Recently, a triggered-suicide system was described that, when activated by the presence of a lytic phage, causes the infected cells to lyse and die before phage replication has occurred.

Strain Identification

Advances in nucleic acid-based bacterial fingerprinting methods have led to new identification tools and renewed interest in bacterial taxonomy. The ability to distinguish between strains of the same species is important not only for identification purposes, but also because it provides a basis for understanding the phylogenetic and evolutionary relationships between LAB. Although morphological, biochemical, and other phenotypic characteristics remain useful for genus and species identification, molecular approaches that rely on nucleotide sequences have proven to be more reliable, more reproducible, and more robust. Several techniques, in particular, are widely used, including pulsed-field gel electrophoresis (PFGE), restriction fragment length polymorphism (RFLP), ribotyping, 16S ribosomal RNA sequence analysis, and various polymerase chain reaction-based protocols. More precise methods of strain identification have also become important as proprietary organisms are developed with unique characteristics and production capabilities.

Encapsulated and Immobilized Cells

Another manner in which starter cultures can be produced and used is to encapsulate or surround cells in inert materials, such as alginate beads. Alternatively, starter culture cells can be adsorbed to a carrier surface or covalently attached to inert support materials. In theory, encapsulated cells may be less sensitive to phage infections, be more stable during storage, and be recovered after fermentation for reuse in subsequent fermentations. However, immobilization procedures can be costly and may alter starter culture metabolism. For these reasons, immobilized culture technology has not yet been adopted in the dairy industry.

It is worth noting that a natural form of encapsulated starter culture has long existed in the form of kefir grains. Kefir, a fluid-fermented product widely consumed in Russia and Eastern Europe, is traditionally made by inoculating milk with kefir grains. Kefir grains are irregularly shaped, polysaccharide-containing particles up to 15 mm in diameter that harbor a complex microflora, including homo- and heterofermentative LAB and yeasts. Kefir grains can be removed from the completed fermentation by filtration or sieving, washed, and reused in subsequent fermentations. Pure cultures, usually containing strains of *Lactobacillus* spp. and *Lactococcus* spp., are now available commercially, and are commonly used in the USA for kefir manufacture.

Probiotics and Antimicrobials

Although starter cultures usually consist of those organisms required to perform a specific fermentation, the same technology, for the most part, is used to produce bifidobacteria and other probiotic cultures, even though they are used differently (*See* ***Bifidobacteria* in Foods; Probiotics**). Additionally, it has long been known that some strains of LAB produce compounds that inhibit other microorganisms. One particular class of inhibitory compounds are the

bacteriocins, proteinaceous materials produced by an organism that inhibits other closely related organisms. Nisin, a bacteriocin produced by strains of *Lactococcus lactis*, is approved for use in processed cheese and other foods as a direct food additive (*See* **Nisin**). LAB that produce nisin and other bacteriocins can be used as part of a lactic acid-producing starter culture or as an adjunct in dairy, meat, and other foods to inhibit pathogens as well as spoilage organisms. In some cases, the organism produces bacteriocin in the food, but does not produce acids or other fermentation end products that would alter the sensory characteristics of the product. Alternatively, the organism can be grown in a dairy- or nondairy-based medium, and the fermented medium containing organic acids and bacteriocin is subsequently pasteurized and concentrated. Viable cells are absent, but the material, when added to foods, would still contribute antimicrobial activity. These 'bioprotective' products have been marketed as shelf-life extenders, although pathogens may also be inhibited. Most of these products are effective only against Gram-positive bacteria; however, some products also inhibit Gram-negative spoilage bacteria, including psychrotrophs.

See also: **Beers**: History and Types; Raw Materials; Wort Production; Biochemistry of Fermentation; Chemistry of Brewing; ***Bifidobacteria* in Foods**; **Biotechnology in Food Production**; **Bread**: Dough Mixing and Testing Operations; Dough Fermentation; **Cheeses**: Types of Cheese; **Genetically Modified Foods**; **Lactic Acid Bacteria**; **Probiotics**; **Vinegar**; **Wines**: Types of Table Wine

Further Reading

Allison GE and Klaenhammer TR (1998) Phage resistance mechanisms in lactic acid bacteria. *International Dairy Journal* 8: 207–226.

Caplice E and Fitzgerald GF (1999) Food fermentations: role of microorganisms in food production and preservation. *International Journal of Food Microbiology* 50: 131–149.

Champagne CP, Lacroix C and Sodini-Gallot I (1994) Immobilized cell technologies for the dairy industry. *Critical Reviews in Biotechnology* 14: 109–134.

Ford A and Fitzgerald GF (1999) Bacteriophage defence systems in lactic acid bacteria. *Antonie van Leeuwenhoek* 76: 89–113.

Hammes WP and Hertel C (1998) New developments in meat starter cultures. *Meat Science* 49(suppl.): S125–S138.

Hutkins RW (2001) Metabolism of starter cultures. In: Marth EH and Steele JL (eds) *Applied Dairy Microbiology*, pp. 209–241. New York: Marcel Dekker.

McKay LL and Baldwin KA (1990) Applications for biotechnology: present and future improvements in lactic acid bacteria. *FEMS Microbiology Reviews* 7: 3–14.

Mäyrä-Mäkinen A and Bigret M (1999) Industrial use and production of lactic acid bacteria. In: Salminen S and von Wright A (eds) *Lactic Acid Bacteria*, pp. 73–103. New York: Marcel Dekker.

Ross RP, Stanton C, Hill C, Fitzgerald GF and Coffey A (2000) Novel cultures for cheese improvement. *Trends in Food Science and Technology* 11: 96–104.

Tannock GW (1999) Identification of lactobacilli and bifidobacteria. In: Tannock GW (ed.) *Probiotics: A Critical Review*, Wymondham, pp. 45–56. England: Horizon Press.

Starvation *See* **Famine, Starvation, and Fasting**

STERILIZATION OF FOODS

M N Ramesh, Central Food Technological Research Institute, Mysore, India

Background

Sterilization is the complete destruction or elimination of all viable organisms in/on an object being sterilized. Sterilization destroys yeasts, molds, vegetative bacteria, and spore formers, and allows the food processor to store and distribute the products at ambient temperatures, thus extending the shelf-life. Sterilization procedures involve the use of heat, radiation, or chemicals, or physical removal of cells. The sterilization process involves four distinct stages. First, the product must be heated at 110–125 °C to ensure sterilization. After this, the product requires a few minutes to equilibrate, since the surface will be hotter, and the central portion of the container will still be cool. The equilibration stage allows a reduction in the temperature gradient. Next, the product must be held

at this temperature for a certain period of time to ensure a predetermined sterilization value designated by the F_0 value. Finally, the product has to be cooled down mainly to arrest further heat treatment and to avoid bursting of the container at hot conditions.

The basic principles of sterilization technology as applied to food processing are as follows:

- The processed product must be free from microorganisms capable of producing food poisoning toxins and those microorganisms that cause food spoilage during the product's shelf-life, until it is consumed.
- *Clostridium botulinum* spores are capable of growing in low acid (pH > 4.6) products during storage and hence must be heat-treated to the equivalent of at least 121.1 °C for 3 min (an F_0 value of 3) to achieve a 12-decimal reduction in the numbers of the microorganism.
- The processing conditions should be applied to the slowest-heating point, referred to as the 'cold point.' This facilitates the assumption that, when the slowest heating part is sterilized, by exposing it to the required time–temperature profile, the rest of the product will be sterilized.

Practically, complete sterilization leads to a deterioration in product quality and nutrient retention (Figure 1). Hence, in practice, commercial sterility is defined as a product that has been processed so that, under normal conditions, the product will neither spoil nor endanger the health of the consumer. The pH of the product is an important factor in determining the severity of the sterilization process.

Theory of Sterilization

Thermal treatment of food products to render them free of pathogenic microorganisms has been practiced for several years. However, a method to quantify the microbial destruction that takes place during thermal treatment has only been understood for the last 75 years. In order to determine the amount of microbial destruction that a thermal treatment delivers to a process, an understanding of the amount of heat delivered to every portion of the food product and the destruction kinetics of the microorganisms of interest is required. The amount of heat delivered by a food process is dependent on the way in which the product is heated and its physical nature. Process-dependent factors can include: processing equipment design, type of heating media, container, food size and shape, product composition, and viscosity (conduction- or convection-heated). The thermal destruction kinetics of microorganisms or their ability to be killed within the food matrix is dependent on a number of factors, including: pH of the product, levels and types of preservatives, water activity, the previous growth conditions of the microorganisms of concern, product composition, and competitive microorganisms.

The two types of bacteria of concern in food preservation are organisms of public health significance and spoilage-causing bacteria. In low-acid foods with a pH greater than 4.6, the organism of public health significance is *Clostridium botulinum*. Canned foods are processed, based on the survival probability for *C. botulinum* of 10^{-12}, or one survivor in 10^{12} cans. The organism most frequently used to characterize low-acid food spoilage by mesophilic spore-formers is PA 3679, a strain of *C. sporogenes*. Most food companies accept thermal inactivation of 10^{-5} for mesophilic spore-formers and 10^{-2} for thermophilic spore-formers. The processing time depends on the bioburden of the most resistant bacteria in a particular food, the spoilage risk involved, and whether food can support the growth of potential contaminating bacteria. Although a large amount of research work has been carried out on the influence of different factors on the processing time and the corresponding sterilization value, a number of uncertainties still exist on the application of these factors to arrive scientifically at the exact processing conditions. In order to avoid any risk resulting from these uncertainties, a safety factor is added to increase the processing time to sterilize the food product completely, which invariably reduces the nutrient content and the increase in energy cost.

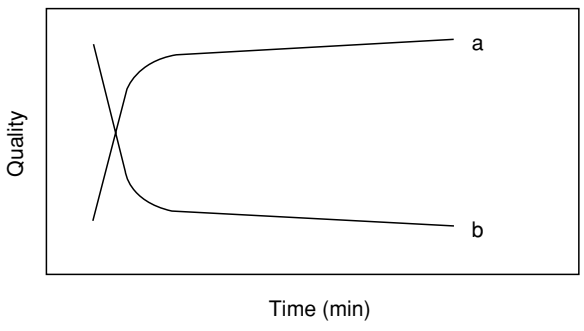

Figure 1 Effect of sterilization on quality. (a) Degree of sterilization; (b) degree of nutrient loss. From Ramesh MN (1995) Optimum sterilization of foods by thermal processing – a review. *Food Science and Technology Today* 9(4): 217–227 with permission.

Decimal Reduction Time or *D* value

The decimal reduction time, or D value, is the time at any temperature to destroy 90% of the spores or vegetative cells of a given organism. It is equal to the number of minutes for the survivor curve to traverse

one log cycle and can be calculated from the reciprocal of the slope of the survivors, assuming a logarithmic death rate:

$$N = N_0 e^{-kt},$$

where N is the final number of survivors after heat treatment, N_0 is the initial number of organisms, e is the exponential function, k is the thermal death rate coefficient (traditionally measured in s^{-1}), and t is the processing time. Rearranging gives

$$\log\ N = \log\ N_0 - (60kt/2.3).$$

Plotting log N against t in minutes gives the slope $-60k/2.3\,\text{min}^{-1}$. The inverse of the slope is

$$D = 2.3/60k\ \text{min}.$$

Thermal Death Time Constant or *z* Value

The D value may be calculated from the inverse of the slope of the survivor curve by regression analysis of the data points. If a presence/absence (growth/no growth) method is used to estimate the number of survivors, N_u can be calculated using the equation

$$N_u = 2.303 \log(n/r),$$

where n is the number of units heated, r is the number of sterile units, and N_u is the most probable number of survivors. The graphical representation of the D-value concept is indicated in Figure 2.

The change in D value with temperature can be obtained by plotting log D against temperature. The z value is the number of degrees for the thermal resistance curve to transverse one log cycle and is equal to the reciprocal of the slope of the curve. The equation applied to the thermal resistance curve is given by

$$\log\ D_{\text{Ref}} = \log\ D_T = 1/[z(T - T_{\text{Ref}})],$$

where D_{Ref} is the D value at the reference temperature (T_{Ref} of 121.1 °C), D_T is the D value at any other temperature, T, and z is the thermal death time constant.

A graphical representation of the D-value concept is shown in Figure 3.

Heat Resistance of Microorganisms

The most important bacterial spore former with respect to heat processing is *Clostridium botulinum*, because of the potent neurotoxin that it produces. The organism occurs in seven distinct stereotypes A–G, which are further subdivided into proteolytic and nonproteolytic strains. The most heat-resistant spores are those produced by type A and the proteolytic B strains. The D value at 121.1 °C of the most resistant strains is generally considered to be 0.21 min. *Clostridium sporogenes*, which is closely related to the proteolytic strains, produces spores that are more heat-resistant. The D value of the spores at 121.1 C° can be up to 1.5 min. *C. sporogenes* has been reported to be the most common mesophilic putrefactive anaerobe in the spoilage of low-acid (pH $\geq$ 4.6) canned foods.

The butyric anaerobes (*C. butyricum*, *C. beijerinckii*, and *C. pasteurianum*) are usually associated with spoilage of products with pH values between 3.9 and 4.6 (e.g., tomatoes and pears), producing blown cans and a butyric odor. Another organism that causes spoilage of products down to pH 4.2, *Bacillus coagulans*, has been reported to have a D value at 98.9 °C of 3.1 min and a z value of 16.1C°. Other mesophilic *Bacillus* species that produce heat-resistant spores and have been implicated in food poisoning outbreaks are *B. subtilis* and *B. licheniformis*. Spores of

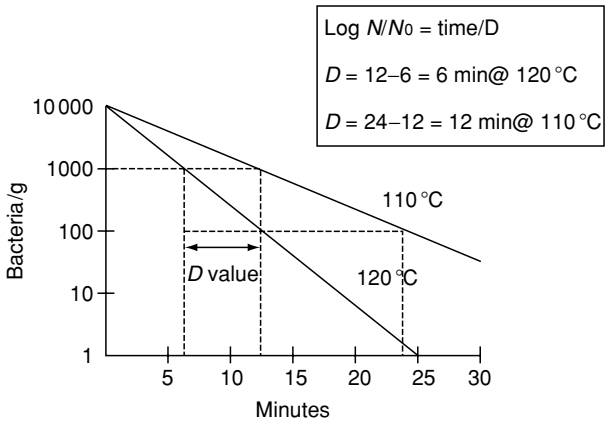

Figure 2 *D*-value of a microorganism. From http://courses.che.umn.edu/00fscn1102-1s/pdf_files/topic_11_Past_and_-Ster.pdf with permission.

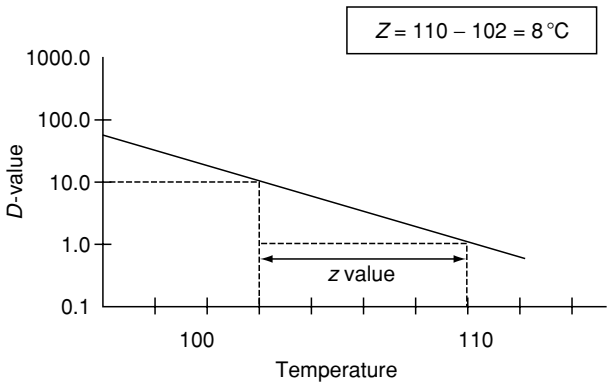

Figure 3 *z*-value of a microorganism. From http://courses.che.umn.edu/00fscn1102-1s/pdf_files/topic_11_Past_and_-Ster.pdf with permission.

the thermophilic organisms that have an optimum growth temperature of ≈55 °C are usually much more resistant to wet heat than mesophilic organisms.

The most resistant thermophile is *Clostridium thermosaccharolyticum*, which produces blown cans and a butyric or 'cheesy' odor, with D values as high as 68–195 min at 121.1 °C. The spores of *Bacillus stearothermophilus* have often been used in process evaluation studies because of their high heat resistance. The organism can cause spoilage of products with pH values > 5.3. Another unusual bacterial spore former is *Sporolactobacillus inulinus*. This organism has been isolated from a number of food and environmental sources and has been reported to have a D value at 90 °C of 5.1 min and a z value of 13 C°. Quality-control personnel should be aware, therefore, that not all isolates of Gram–positive spore-forming rods belong to the genera *Bacillus* or *Clostridium*. Typical D and z values for some of the spore formers are listed in Table 1.

Lethal Rate

Using the z value, the lethal rate, L, can be calculated from

$$L = \log^{-1}[((T - T_{\text{Ref}})/z)].$$

The lethal rate is a measure of the lethality of any temperature, T, relative to the reference temperature T_{Ref}. For example, for a reference temperature of 121.1 °C and a z of 10C °, the lethal rate at 111.1 °C will be

$$\log^{-1}[(111.1 - 121.1)/10)] = 0.1.$$

Thus, 1 min at 111.1 °C is worth 0.1 min at 121.1 °C in terms of lethality. Note that z is written C ° and not °C, since z represents a change of 10 Centigrade degrees in the above example and not a temperature of 10 degrees Centigrade.

F Value

The symbol F designates the equivalent in minutes at 121.1 °C of the combined lethalities of all time–temperature relationships at the point of slowest heating for a product during heat processing. Thus, the F value is a measure of the killing power of a heat process. The term F_c denotes the F value at the center of a pack, F_0 denotes the equivalent F value in minutes at 121.1 °C, and F_s denotes the integrated lethality of heat received by all points in a container. F_s can be related to the D value by the equation

$$F_s = D_{\text{Ref}}(\log\ N_0 - \log\ N),$$

where D_{Ref} is the D value at 121.1 °C, log N_0 is log of the initial number of organisms, and log N is the log of the final number of surviving organisms. In the above equation, F_s can be considered as equivalent to F_0 or F_c in rapid heating systems such as the thermoresistometer. The F_0 value of a process can be obtained in practice by summing the lethal rates at 1-min intervals from the heating and cooling curve of a product during a heat process. This is a simplified view, and process calculations in the industry involve rather more complicated mathematics and usually do not include the cooling curve contribution to the process. The process time can be calculated using the equation

Table 1 Typical heat resistances of selected spore formers

Organism	*Medium*	*Temperature (°C)*	D *value (min)*	z *(C °)*
Clostriduim sporogenes (PA 3679)	Several substrates including pea purée	104.4–143.3	0.75–2.03	9.0–14.7
	Phosphate buffer pH 7	112.8–148.9	1.06	9.3
Clostriduim botulinum	Several substrates	104.0–132.2	0.051–0.58	8.2–10.4
	Phosphate buffer pH 7	120.0–140.0	0.13	11.0
Other *Clostridium* sp.	Several substrates	85.0–121.0	0.2–195	6.9–11
Bacillus sp.	Several substrates including milk	100.0–121.0	0.3–16.0	4.1–7.7
Desulfotomaculum nigrificans		121	55.0	9.5
Escherichia coli	Nutrient broth	70	0.006	4.9
	Milk	70	0.04	6.5
Lactobacillus sp.	Tomato juice	70	4.0–11.0	11.5–12.5
Listeria monocytogenes	Beef, chicken, and carrot homogenates	70	0.14–0.27	5.98–7.39
Salmonella typhimurium	Aqueous sucrose/glucose (a_w 0.995)	70	0.03–816	6.8–19.0
	Milk chocolate (51% milk)			
Staphylococcus aureus	Milk	70	0.30	5.1
Streptococcus sp.	Broth and ham	70	0.015–2.84	3.5–17.0
Microbacterium lacticum	Skim milk	70	4.0	

From Brown KL (1991) Principles of heat preservation. In: Rees JAG and Bettison J (eds) *Processing and Packaging of Heat Preserved Foods*, pp. 15–49. New York: Van Nostrand Reinhold.

$$B = f_h(\log\ J_{ih}I_h - \log\ g_c),$$

where B is the thermal process time corrected for the time to bring the retort to the process temperature, f_h is the time in minutes for the semilog heating curve to traverse one log cycle, J_{ih} is the heating lag factor, I_h is difference in temperature between the retort and the food at the start of process and g_c is the difference between the retort temperature and the maximum temperature reached by the food at the center. In the processing of low-acid foods with a pH value of 4.6 or above, a process equivalent in lethality to at least F_0 of 3 min must be applied to minimize the risk of spores of *C. botulinum*. An F value determined on the basis of one z value to an F value determined on the basis of a second z value can be converted using the appropriate formula.

Cook Value or *C* Value

Mansfield introduced the concept of a lethality-like value for sensory degradation, termed the cook value or C value. This has a reference temperature of 100 °C and a z value typically in the range 20–40C°.

$$C_{100} = 10^{(T-100)/z}\text{c min.}$$

The processing conditions mainly depend on the pH value of the product.

Acid Products (pH < 4.6)

At a pH below 4.6, the risk of growth and toxin production by *C. botulinum* is extremely low, and for products with pH values between 4.0 and 4.6, processes are aimed at controlling the survival and growth of spore forming organisms such as *Bacillus coagulans, Bacillus polymyxa, Bacillus macerans*, and the butyric anaerobes such as *C. butyricum* and *C. pasteurianum*. A heat process of $F^{10}_{121} = 0.7$ is regarded as adequate for this purpose, and a process equivalent to 10 min at 93.3 °C ($F^{8.3}_{93.3} = 5$) when the pH is between 4.0 and 4.3. However, more severe processes may be required to control excessive contamination. In a previous study of nine cases of botulism from products with pH below 4.6, including canned pears, apricots, tomato ketchup, tomato–onion chili sauce, and green tomatoes, it was suggested that growth of spoilage organisms raised the pH to a level at which dormant spores of *C. botulinum* could germinate and grow. Below pH 3.7, the processor is concerned with the control of nonsporing bacteria, yeasts, and molds. These may be generally controlled by heat processes at temperatures below 100 °C.

Methods of Sterilization

The food sterilization methods are divided into two categories: sterilization by heating (thermal processing) and sterilization without heating (nonthermal processing). Thermal processing is widely practiced these days, in spite of some problems such as that the process of heating might reduce nutrition or deteriorate the quality of foods, and that it is ineffective against certain types of bacteria. Nonthermal processing is considered an effective method that does not cause any deterioration of quality, in contrast with thermal processing. However, no reports have shown the effect of sterilization without heating. Research on the evaluation of the technology of nonthermal sterilization without heating is widely being pursued internationally.

Thermal processing is divided further into two categories: in-container sterilization (processing). The principles involved in thermal sterilization of foods remain the same whether attempting to sterilize products in containers or sterilize products prior to filling in the final container (aseptic processing). Thus, it is necessary to know the thermal destruction rate for the microorganisms of consequence in the food being processed. The procedures necessary for acquisition of such data are available. It is important to use this information properly so that the appropriate time and temperature for destruction of the organisms can be achieved. Sterilization procedures for products in containers usually require longer times, since the heat transfer to the product is relatively slow. Sterilization prior to filling in the container, as accomplished in aseptic processing, requires relatively short heating periods. This sterilization process is usually accomplished by heating the product rapidly to 130–145 °C, holding for an appropriate time, then rapidly cooling the product. The specific product will determine the actual combination of temperature and time required for sterilization.

Product heating is accomplished by either indirect or direct heat exchange. In liquid homogeneous products, indirect heating occurs when the product to be heated and the heating medium are separated by the heating surface. This type of heating is accomplished using a scraped surface, tubular, or plate heat exchanger. Direct heat exchange is achieved by placing the heating medium directly into the product or vice versa. In either case, the added water from the steam must be either removed or accounted for in the formulation. The specific type of heating used is usually dictated by the nature of the product and the economics of operation. Following heating to the sterilization temperature, the product must be held at this temperature for a specific length of time to

accomplish sterilization. The sterilization time in continuous-flow aseptic systems is obtained by conveying the product through a nonheating pipe attached to the heating system. This pipe is called the holding tube and is of a specified uniform diameter and length. Its capacity is such that the fastest particle is held at the maximum time required to sterilize the product. The length of time for which the product remains in the tube is dependent not only on the holding capacity of the tube and the rate at which the product is pumped through the tube, but also on the manner in which the product flows through the tube.

Thermal Processing

During sterilization, the type of heat, time of application, and temperature to ensure destruction of all microorganisms are very important. Endospores of bacteria are considered the most thermoduric of all cells, so their destruction guarantees sterility. The lethal temperature varies for microorganisms. The time required to kill the microorganisms depends on the number of organisms, species, nature of the product being heated, pH, and temperature. Whenever heat is used to control microbial growth, inevitably both time and temperature are considered.

Sterilization (boiling, autoclaving, hot air oven) kills all microorganisms with heat and is commonly employed in canning, bottling, and other sterile packaging procedures. Heat sterilization is the unit operation in which foods are heated at a sufficiently high temperature and for a sufficiently long time to destroy microbial and enzyme activity.

Sterilized foods have a longer shelf-life but undergo substantial changes in quality. Developments in processing technology therefore aim to reduce the overall processing time. The effects of microbial heat resistance on the design of heat-sterilization procedures and equipment are an important aspect of sterilization in both in-container heat sterilization and aseptic processing.

In order to determine the process time for a given food, it is necessary to obtain information on both the heat resistance of microorganisms and the rate of heat penetration into the food. A process that reduces cell numbers by eight decimal reductions (an 8D process), applied to a raw material containing 10^5 spores per container would reduce microbial numbers to 10^{-3} per container, or one microorganism in every thousand containers. Commercial sterility, therefore, means that the vast majority of containers are sterile, but there is a probability that nonpathogenic cells survive the heat treatment in a predetermined number of containers. A 12D process is used when *C. botulinum* is likely to be present but in foods that contain less heat-resistant spoilage microorganisms. In addition to information on heat resistance, it is necessary to collect data describing the rate of heat penetration into the food in order to calculate the processing time needed for commercial sterility.

Retorting (Heat Processing)

The most widely used system for sterilization uses overpressure retorts or agitating sterilizers. Generally, the overpressure retorts are of the batch type with steam or pressurized hot water as the heating media. The pressure in the retort is maintained during the entire processing cycle, with steam during the heating cycle and with compressed air during the cooling cycle. Agitating sterilizers gently move the product within the container, which facilitates more uniform heating and cooling owing to better heat transfer. This reduces the come-up time, delivering an improved product.

Heating by Saturated Steam

Latent heat is transferred to food when saturated steam condenses on the outside of the container. If air is trapped inside the retort, it forms an insulating boundary film around the cans, which prevents the steam from condensing and causes under processing of the food. It also produces a lower temperature than that obtained with saturated steam. It is therefore important that all air be removed. Sterilization requires a temperature of 121.1 °C. This creates pressure inside the container, and if the pressure on the outside of the container is less than that inside, the container will expand and subsequently break. This is more critical in glass jars and flexible containers like pouches and plastic cans. To avoid this situation, an overhead pressure of 230–250 kPa is applied to equalize the internal pressure.

Aseptic Processing

Conventional retorting of A2 cans of vegetable soup requires 70 min at 121 °C to achieve an F_0 value of 7 min, followed by cooling for 50 min. Aseptic processing in a scraped-surface heat exchanger at 140 °C for 5 s gives an F_0 value of 9 min. Increasing the can size to A10 increases the processing time to 218 min, whereas with aseptic processing, the sterilization time is the same. In aseptic processing, containers are not required to withstand sterilization conditions. Cartons are presterilized with hydrogen peroxide, and filling machines are enclosed and maintained in a sterile condition by ultraviolet light and filtered air.

A positive air pressure is maintained in the filling machine to prevent entry of contaminants. The process is successfully applied to liquid and small-particulate foods, but problems remain with larger pieces of solid food.

Nonthermal Processing

Nonthermal methods for the preservation of foods are under intense research to evaluate their potential as an alternative or complementary process to traditional methods of food preservation. Traditionally, most preserved foods are thermally processed by subjecting the food to a temperature of 60–100 °C for a few seconds to minutes. During this period, a large amount of energy is transferred to the food. This energy may trigger unwanted reactions in the food, leading to undesirable changes or formation of byproducts. For example, thermally processed milk may have a cooked flavor accompanied by a loss of vitamins, essential nutrients, and flavors. The fact that the shelf-life and the quality of food are important to consumers gave birth to the concept of preserving foods using nonthermal methods. Nonthermal methods of food preservation are currently being developed to eliminate, or at least minimize, the quality degradation of foods that results from thermal processing.

During nonthermal processing, the temperature of the food is held below the temperature normally used in thermal processing. Therefore, the quality degradation expected from high temperatures is minimal in nonthermal processing. The vitamins, essential nutrients, and flavors are expected to undergo minimal, or no, changes during nonthermal processing. In addition, nonthermal processes use less energy than thermal processes. Foods can be processed nonthermally using high hydrostatic pressure, oscillating magnetic fields, high-intensity pulsed electric fields, intense light pulses, irradiation, chemical, biochemical, and hurdle technology. Although these technologies have been used for a long time to inactivate microorganisms and/or preserve food, they have gained recognition as nonthermal methods of food preservation only in the recent past. Nonthermal processes are expected to induce only minimal degradation of food.

High-pressure Processing

High-pressure processing (HPP), also described as high hydrostatic pressure (HHP) or ultrahigh pressure (UHP) processing, subjects liquid and solid foods, with or without packaging, to pressures between 100 and 900 MPa. A high hydrostatic pressure is used for the inactivation of microorganisms and certain enzymes and for shelf-life extension of foods. Spores can be inactivated by combining high pressure with temperature. The germination of spores is an important step in spore inactivation. The process temperature during pressure treatment can be specified from less than 0 °C to more than 100 °C. Commercial exposure times can range from a millisecond pulse to more than 20 min. HPP acts instantaneously and uniformly throughout a mass of food, independent of the size, shape, and food composition. Compression uniformly increases the temperature of foods by approximately 3 °C per 100 MPa. Compression of foods may shift the pH of the food as a function of imposed pressure and must be determined for each food treatment process. Water activity and pH are critical process factors in the inactivation of microbes by HPP. An increase in food temperature above room temperature and, to a lesser extent, a decrease below room temperature increase the inactivation rate of microorganisms during HPP treatment. Temperatures in the range of 45–50 °C appear to increase the rate of inactivation of food pathogens and spoilage microbes. Temperatures from 90–110 °C in conjunction with pressures include batch and semicontinuous systems, but no commercial continuous HPP systems are operating.

The critical process factors in HPP include pressure, time at pressure, time to achieve treatment pressure, decompression time, treatment temperature (including adiabatic heating), initial product temperature, vessel temperature distribution at pressure, product pH, product composition, product water activity, packaging material integrity, and concurrent processing aids. Chemical changes in the food generally will be a function of the process and treatment time. Because some types of spores of *C. botulinum* are capable of surviving even the extreme pressures and temperatures of HPP, there is no absolute microbial indicator for sterility by HPP. For vegetative bacteria, nonpathogenic *Listeria innocua* is a useful surrogate for the foodborne pathogen, *Listeria monocytogenes* a nonpathogenic strain of *Bacillus* may be useful as a surrogate for HPP-resistant *E. coli* 0157:H7 isolates.

Subjecting foods to pressures of 100–800 MPa inactivates vegetative bacteria, yeasts, molds, and parasites in products such as jams, orange juice, and meat products. Factors that affect the rate of microbial inactivation include pressure and magnitude, microbial type and growth stage, temperature, pH, water activity, and food composition. Application of a pressure of 680 MPa on grape juice for 10 min would arrest the growth of the microbes and thereby stops further fermentation of the grape juice. Peaches and pears subjected to a pressure of 410 MPa for 30 min exhibit a shelf-life of 5 years. One of the important

challenges in using HPP is the fabrication of pressure vessels and seals that can withstand the high pressures during the cycles of pressurization and depressurization.

Pulsed Electric Fields

High-intensity pulsed electric field (PEF) processing involves the application of pulses of high voltages (typically 20–80 kV cm^{-1}) of exponentially decaying, square wave, bipolar, or oscillatory pulses and at ambient, subambient, or slightly above ambient temperatures for less than 1 s to foods. High-intensity electric fields applied to a food in the form of short-duration pulses can inactivate the microorganisms and certain enzymes. Energy loss due to heating of foods is minimized, reducing the detrimental changes of the sensory and physical properties of foods. Some important aspects in pulsed electric field technology are the generation of high electric-field intensities, the design of chambers that impart uniform treatment to foods with minimal increase in temperature, and the design of electrodes that minimize the effect of electrolysis.

Although different laboratory- and pilot-scale treatment chambers have been designed and used for PEF treatment of foods, only two industrial scale PEF systems are available. The systems including treatment chambers and power supply equipments need to be scaled up to commercial levels. To date, PEF has been applied mainly to improve the quality of foods. Application of PEF is restricted to food products that can withstand high electric fields, have a low electrical conductivity, and do not contain or form bubbles. The particle size of the liquid food in both static and flow treatment modes is a limitation. Several theories have been proposed to explain microbial inactivation by PEF. The most-studied theories are electrical breakdown and electroporation. Factors that affect the microbial inactivation with PEF are process factors (electric field intensity, pulse width, treatment time and temperature, and pulse wave shapes), microbial entity factors (type, concentration, and growth stage of microorganism) and media factors (pH, antimicrobes and ionic compounds, conductivity, and medium ionic strength). Microbial inactivation increases with increasing electric field intensity, exposure time, and temperature of the food. However, it is desirable to maintain the temperature below 30–40 °C by providing a cooling system. Different bacteria have different sensitivities to electric field treatment. In general, Gram-positive bacteria and yeasts are more resistant to electric fields than Gram-negative bacteria. The optimum conditions for maximum inactivation of a specific microorganism can be determined after preliminary research. Although PEF has potential as a technology for food preservation, existing PEF systems and experimental conditions are diverse. The effects of critical process factors on pathogens of concern and kinetics of inactivation need to be studied further. An electric pulse process for the treatment of fresh citrus juices capable of reducing target pathogens without alteration of the juice was granted an FDA letter of no objection for its use in April 1999.

High-voltage Arc Discharge

Arc discharge is an early application of electricity to pasteurize fluids by applying rapid discharge voltages through an electrode gap below the surface of aqueous suspensions of microorganisms. A multitude of physical effects (intense wave) and chemical compounds (electrolysis) are generated, inactivating the microorganisms. The use of arc discharge for liquid foods may be unsuitable, largely because of electrolysis and discharge, but more recent designs have shown some promise for use in food preservation.

Pulsed Light Technology

Pulsed light is a method of food preservation that involves the use of intense and short-duration pulses of broad spectrum 'white light' (ultraviolet to the near-infrared region). The use of pulsed high-intensity light to inactivate microorganisms is a relatively new technology. For most applications, a few flashes applied in a fraction of a second provide a high level of microbial inactivation. This technology is applicable mainly in sterilizing or reducing the microbial population on packaging or food surfaces. Extensive independent research on the inactivation kinetics under a full spectrum of representative variables of food systems and surfaces is needed. Application of light pulses involves exposure of foods to short duration pulses (1 μs to 0.1 s) of intense incoherent light. Light with an energy density of about 0.01–50 J cm^{-2} and a wavelength in the range of 170–2600 nm is used. Such incoherent intense pulses of light may be generated using gas–filled flash lamps or spark gap discharge apparatus. Full – or filtered – spectrum light may be used, depending on the degree of sterilization expected. The filtered–spectrum light is devoid of wavelengths known to cause undesirable reactions in foods. Glass or liquid filters are used to obtain the filtered spectrum. In general, filtered light is more effective for microbial inactivation than full–spectrum light.

Ultraviolet Light

There is particular interest in using ultraviolet (UV) light to treat fruit juices, especially apple juice and

cider. Other applications include disinfection of water supplies and food contact surfaces. Ultraviolet processing involves the use of radiation from the ultraviolet region of the electromagnetic spectrum. The germicidal properties of UV irradiation (200–280 nm) are a result of DNA absorption of the UV light. This mechanism of inactivation results in a sigmoidal curve of microbial population reduction. To achieve microbial inactivation, the UV radiant exposure must be at least 400 J m^{-2} in all parts of the product. Critical factors include the transmissivity of the product, the geometric configuration of the reactor, the power, wavelength and physical arrangement of the UV source(s), the product flow profile and the radiation path length. UV may be used in combination with other alternative process technologies, including various powerful oxidizing agents such as ozone and hydrogen peroxide, among others.

Oscillating Magnetic Fields

Static (SMFs) and oscillating magnetic fields (OMFs) have been explored for their potential to inactivate microorganisms. For SMFs, the magnetic field intensity is constant with time, while an oscillating magnetic field is applied in the form of constant amplitude or decaying amplitude sinusoidal waves. An OMF applied in the form of pulses reverses the charge for each pulse. The intensity of each pulse decreases with time to about 10% of the intensity. Preservation of foods with OMFs involves sealing food in a plastic bag and subjecting it to one to 100 pulses in an OMF with a frequency of 50–500 kHz at a temperature of 0–50 °C for a total exposure time ranging from 25–100 ms. The effects of magnetic fields on microbial populations have produced controversial results. Consistent results concerning the efficacy of this method are needed before considering this technology for food-preservation purposes.

Oscillating magnetic fields with a magnetic flux of 5–50 T and a frequency of 5–500 kHz have been reported to inactivate microorganisms. Foods with an electrical resistivity of 10–25 $\Omega \cdot$ cm may be sealed in a plastic bag and subjected to OMFs. One of the attractive features of using magnetic fields for food preservation is that the food can be packaged prior to processing, reducing the possibility of cross-contamination during packaging. Studies on the effects of static and pulsed magnetic fields as an alternative to conventional thermal treatments on the inactivation of *Saccharomyces cerevisiae* have been reported. The potential advantages of food preservation by magnetic fields, the proposed interaction mechanisms, and some of the results that have been obtained by exposing living systems to low- and high-frequency, high-intensity magnetic fields were discussed in the IFT Annual Meeting in 2000.

Pulsed X-rays

A number of studies have compared the effects of electron beam, gamma rays, and X-rays, but comparison between these technologies is inconclusive owing to differences in the doses applied. Electrons have a limited penetration depth of about 5 cm in food, whereas X rays have significantly greater penetration depths (60–400 cm) depending on the energy used. The use of pulsed X-rays is a new alternative technology that utilizes a solid-state opening switch to generate an electron beam 30 ns down to a few nanoseconds; repetition rates of up to 1000 pulses per second. The practical application of food irradiation by X-rays in conjunction with existing food-processing equipment is further facilitated by:

- the possibility of controlling the direction of the electrically produced radiation;
- the possibility of shaping the geometry of radiation field to accommodate different package sizes;
- its high reproducibility and versatility.

Potentially, the negative effects of irradiation on the food quality can be reduced. Irradiation of foods was one of the earliest nonthermal food preservation technologies. Irradiation is the exposure of food to radiation with an energy of 5–10 kGy and wavelengths of 2000 Å or less. Ultraviolet, beta, gamma, and X-, and microwaves are included in this range. One of the attractions of irradiation is its ability to pasteurize foods in the frozen state. The World Health Organization (WHO) approved a radiation dosage of up to 10 kGy as being 'unconditionally safe for human consumption.' Irradiation has the potential to replace the use of many hazardous chemical pesticides and preservatives.

Ultrasound

Ultrasound is the energy generated by sound waves of 20 000 or more vibrations per second. Although ultrasound technology has a wide range of current and future applications in the food industry, including inactivation of microorganisms and enzymes, presently, most developments for food applications are nonmicrobial. Data on inactivation of food microorganisms by ultrasound in the food industry are scarce, and most applications use combinations with other preservation methods. The bactericidal effect of ultrasound is attributed to intracellular cavitations, a phenomenon in which mechanical high-frequency vibrations cause alternate compressions of millions of microscopic bubbles containing gas and vapor. The

bubbles expand then implode violently, releasing large amounts of energy and generating very high temperatures and pressures within the bubbles. The molecules of the vaporized reaction mixture are fractured, forming highly reactive free radicals. Cavitations occur as a result of micromechanical shocks that disrupt cellular structural and functional components up to the point of cell lysis.

The heterogeneous and protective nature of food with the inclusion of particulates and other interfering substances severely curtails the singular use of ultrasound as a preservation method. Although, at present, these limitations make commercial development unlikely, the combination of ultrasound with other preservation process (e.g., heat and mild pressure) appears to have the greatest potential for industrial applications. Critical processing factors are assumed to be the amplitude of the ultrasound waves, the exposure/contact time with the microorganisms, the type of microorganisms, the volume of food to be processed, the composition of the food, and the temperature of treatment.

Microwave and Radio-frequency Processing

Microwave and radio-frequency heating refers to the use of electromagnetic waves of certain frequencies to generate heat in a material by two mechanisms – dielectric and ionic. Microwave and radio-frequency heating for pasteurization and sterilization is preferred to conventional heating, because they require less time to come up to the desired process temperature, particularly for solid and semisolid foods. Industrial microwave pasteurization and sterilization systems have been reported for over 30 years, but commercial radio-frequency heating systems for the purpose of food pasteurization or sterilization are not known to be in use. For a microwave sterilization process, unlike conventional heating, the design of the equipment can dramatically influence the critical process parameters – the location and temperature of the coldest point. This uncertainty makes it more difficult to make general conclusions about processes, process deviations, and how to handle deviations.

Many techniques have attempted to improve the uniformity of heating. The critical process factor when combining conventional heating and microwave or any other novel processes would most likely remain the temperature of the food at the cold point, primarily due to the complexity of the energy absorption and heat-transfer processes. Since the thermal effect is presumably the sole lethal mechanism, the time–temperature history at the coldest location will determine the safety of the process and is a function of the composition, shape, and size of the food, the microwave frequency, and the applicator (oven) design. Time is also a factor in the sense that, as the food heats up, its microwave absorption properties can change significantly, and the location of cold points can shift.

Ohmic and Inductive Heating

Ohmic heating (sometimes also referred to as Joule heating, electrical resistance heating, direct electrical resistance heating, electroheating, and electroconductive heating) is defined as the process of passing electric currents through foods or other materials to heat them. Ohmic heating is distinguished from other electrical heating methods by the presence of electrodes in contact with the food, frequency, and waveform. The principal advantage claimed for ohmic heating is its ability to heat materials rapidly and uniformly, including products containing particulates. The principal mechanisms of microbial inactivation in ohmic heating are thermal. While some evidence exists for nonthermal effects of ohmic processes, which rely on heat, it may be unnecessary for processors to claim this effect in their process fillings.

Inductive heating is a process wherein electric currents are induced within the food owing to oscillating electromagnetic fields generated by electric coils. No data on microbial death kinetics under inductive heating have been published.

Hurdle Technology

Hurdle technology combines nonthermal food processing with traditional or other emerging technologies. The most promising combinations include nonthermal methods, such as high hydrostatic pressure, ultrasound, and pulsed electric fields. In the inactivation of spores, it is necessary to use a combined methods approach using 'hurdles.' Hurdles are physical or chemical parameters that can be adjusted to ensure the microbial stability and safety of foods. The physical parameters include the processing and storage temperatures, water activity, pH, and redox potential at levels that inhibit or inactivate the microorganisms and thus render the food safe. Hurdle technology is used in the preservation of meat and seasonal or regional fruits and vegetables.

Besides preserving food quality, new nonthermal technologies have to achieve an equivalent or, preferably, a better enhanced safety level than that for procedures that they replace. Most nonthermal preservation techniques are highly effective in inactivating vegetative forms of bacteria, yeast, and molds, but bacterial spores and most enzymes remain difficult to inactivate.

See also: ***Escherichia coli***: Food Poisoning by Species other than *Escherichia coli*; **Food Poisoning**: Classification; Tracing Origins and Testing; Statistics; Economic Implications; **Food Safety**; **Food Security**; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; Electrical Process Heating; **Pasteurization**: Pasteurization of Liquid Products; Pasteurization of Viscous and Particulate Products; Other Pasteurization Processes; **Preservation of Food**; **Spoilage**: Bacterial Spoilage; Molds in Spoilage; Yeasts in Spoilage

Further Reading

Ball CO (1923) *Bulletin of the National Research Council 7, Part 1, No. 37*.

Barbosa-Canovas GV, Pothakamury UR, Palou E and Swanson BG (1998) Emerging technologies in food preservation. In: *Nonthermal Preservation of Foods*. New York: Marcel Dekker.

Barbosa-Canovas GV, Pothakamury UR and Swanson BG (1995) State of the art technologies for the stabilization of foods by non thermal processes: physical methods. In: Barbosa-Canovas and Welti-Chanes J (eds) *Food Preservation by Moisture Control. Fundamentals and Applications*, pp. 493–532. Lancaster, PA: Technomic.

Brown KL (1988) Use of computer data file for storage of heat resistance data on bacterial spores. *Journal of Applied Bacteriology* 65: 49–54.

Brown KL (1991) Principles of heat preservation. In: Rees JAG and Bettison J (eds) *Processing and Packaging of Heat Preserved Foods*, pp. 15–49. New York: Van Nostrand Reinhold.

Cash JN and Sinha NK (1997) Canning of vegetables. In: Smith DS, Cash JN, Nip WK and Hui YH (eds) *Processing Vegetables: Science and Technology*. Lancaster, PA: Technomic.

Doores S and Westhoff D (1981) Heat resistance of *Sporolactobacilus inulins*. *International Journal of Food Science* 46: 810–812.

Gaze JE and Brown KL (1988) The heat resistance of spores of *C. botulinum* 213B over the temperature range of 120 °C to 140 °C. *International Journal of Food Science & Technology* 23: 373–378.

Hersom AC and Hulland ED (1980) *Canned Foods. Thermal Processing and Microbiology*, 7th edn. Edinburgh: Churchill Livingstone.

Hofmann GA (1985) Deactivation of Microorganisms by an Oscillating Magnetic Field. US Patent 4,524,079.

Holdsworth SD (1985) Optimization of the thermal processing – A review. *Journal of Food Engineering* 4: 89–116.

Hartfinger L (1992) *Microwave Sterilization, Food Technology* 46(12): 57–61.

Hulsheger H, Potel J and Niemann EG (1981) Killing of bacteria with electric pulses of high field strength. *Radiation and Environmental Biophysics* 20: 53–65.

IFT (2001) Kinetics of microbial inactivation for alternative food processing technologies *Journal of Food Science* Special Supplement.

Jay JM (1992) *Modern Food Microbiology*, 4th edn. New York: Van Nostrand Reinhold.

Leistner L (1994) Introduction to hurdle technology. In: Leistner L and Morris LGM (eds) *Food Preservation by Combined Processes, Final Report of Flair Concerted Action No. 7, Subgroup B*, pp. 1–6.

Matsuda N, Komaki N, Ichikawa R and Gotoh S (1985) Facultative anaerobic spore-forming bacteria isolated from spoiled canned foods. *Nippon Shokuhin Kogyo Gakkaishi* 32(16): 391–398.

National Canners Association (1968) *Laboratory Manual for Food Canners and Processors*, vol. 1. *Microbiology and Processing*. Westport, CT: AVI.

Parrott DL (1992) Use of ohmic heating for aseptic processing of food particulates. *Food Technology* 46(12): 68–72.

Pflug IJ and Christensen R (1980) Converting an *F*-value determined on the basis of a second *Z* value. *Journal of Food Science* 45: 35–40.

Pothakamury UR, Barbosa-Canovas GV, Swanson BG and Meyer RS (1995) The pressure builds for better food processing. *Chemical Engineering Progress* March: 45–55.

Quin B, Pothakamury UR, Barbosa-Canovas GV and Swanson (1996) Nonthermal pasteurization of liquid foods using high intensity pulsed electric fields. *Critical Reviews in Food Science and Nutrition* 36(6): 603–627.

Quin B, Vea-Mercado H, Pothakamury UR, Barbosa-Canovas GV and Swanson BG (1995) Application of pulsed electric fields for inactivation of bacteria and enzymes. *Journal of the Franklin Institute* 332A: 209–220.

Ramesh MN (1995) Optimum sterilization of foods by thermal processing – a review. *Food Science and Technology Today* 9(4): 217–227.

Ramesh MN, Prapulla SG, Kumar MA and Mahadevaiah M (1997) Thermal processing of Foods: A retrospective Part I, Uncertainties in thermal processing and statistical analysis. *Advances in Applied Microbiology*. 44: 287–314. New York: Academic Press.

Russell AD (1982) *The Destruction of Bacterial Spores*. London: Academic Press.

Stevenson KE and Ito KA (1991) Aseptic processing and packaging of heat preserved foods. In: Rees JAG and Bettison J (eds) *Processing and Packaging of Heat Preserved Foods*, pp. 72–91. New York: Van Nostrand Reinhold.

Stumbo CR (1973) *Thermobacteriology in Food Processing*. New York: Academic Press.

Teixeiria A (1992) Thermal process calculations. In: Heldman DR and Lund DB (eds) *Handbook of Food Engineering*. New York: Marcel Dekker.

Steroid Hormones *See* **Hormones**: Adrenal Hormones; Thyroid Hormones; Gut Hormones; Pancreatic Hormones; Pituitary Hormones; Steroid Hormones

Stilton *See* **Cheeses**: Mold-ripened Cheeses: Stilton and Related Varieties

STORAGE STABILITY

Contents

Mechanisms of Degradation

M Hole, University of Lincolnshire and Humberside, Lincoln, Grimsby, UK

Background

The storage of a food is limited by changes to the food, which render it unacceptable to the consumer for reasons of safety and of quality, whether organoleptic or nutritional.

The cause of these undesirable changes can be related to physical, microbiological, enzymatic, and chemical reasons. (*See* **Spoilage:** Chemical and Enzymatic Spoilage; Bacterial Spoilage; Fungi in Food – An Overview; Molds in Spoilage; Yeasts in Spoilage.) Which mechanism of degradation operates depends on the type of food and the storage conditions (see Figure 2).

Postharvest storage of fresh commodities (whether of plant or animal origin) is mainly limited by microbial action and the natural metabolic (i.e., enzymatic) processes occurring after harvest or death. The modern food industry is primarily concerned with minimizing the degradation reactions occuring during storage. (*See* **Preservation of Food.**) Thus, storage at a reduced temperature is a successful strategy (*See* **Chilled Storage:** Principles; Attainment of Chilled Conditions; Quality and Economic Considerations; Microbiological Considerations; **Freezing:** Storage of Frozen Foods) as is modified atmosphere packaging (MAP) (*See* **Packaging:** Packaging of Solids) for the control of microbial growth in foods. High-temperature processing aims to eliminate the microbial mechanism of degradation, and also to inactivate enzymes. This would leave chemical mechanisms as the main possible degradation pathway, and examples of such changes during the extended ambient storage periods of canned foods are given below. More generally, the relatively high reactivity of oxygen with food components, especially lipids, is a major chemical degradation route and is discussed further below. (*See* **Oxidation of Food Components.**)

Physical changes in foods result in undesirable textural changes, especially for fresh commodities where the texture relates to the cellular nature of the food and the way in which water is held within these cells. This 'cellular water' can be adversely affected by frozen storage/thawing cycles.

The classification of foods as cellular (i.e., plants, animal commodities) or dispersive (i.e., colloidal) is relevant to physical mechanisms of degradation. For dispersive foods, which are usually fabricated (e.g., margarine, icecream, salad cream, gelled products), the inherent physical instability of colloidal dispersions must be considered in relation to their mode of storage. (*See* **Colloids and Emulsions.**) Water in foods (*See* **Water Activity:** Effect on Food Stability) is a major factor for the possible mechanisms of degradation. It is essential for microbial growth and

reactions involving enzymes (hence the success of traditional drying methods for food preservation). The presence of water, as indicated above, is important for food texture, but the absence of water can be desirable for the crispness of snack foods. Storage that allows snack foods to absorb water, and hence lose their crispness, is clearly undesirable – as is the opposite effect of drying out for foods (e.g., cakes, cheese). (*See* **Intermediate-moisture Foods**.) From a chemical point of view, water can react with food components in a hydrolysis reaction, and examples are discussed below.

Physical Degradation Mechanisms

Water binding is especially important in relation to the texture of fresh foods, whether of plant or animal origin. For fresh leafy vegetables, e.g., lettuce, the desirable firmness is dependent upon the turgor pressure of the cellular water – excessive transpiration owing to storage in a warm or dry environment will cause a reduction in turgor pressure with a resultant undesirable loss of firmness, and reduced consumer appeal for the 'limp' result. Chilling can be beneficial, but if frozen, undesirable texture changes can occur resulting from the loss of cellular water. The mechanism involves differential freezing of water in the intercellular space, with a resultant, irreversible, osmotically induced flow of water from inside the cell into the intercellular space. On thawing, this water is expelled, leading to the unattractive 'drip loss' and leaving a less desirable texture. This is a general problem for fresh meat and fish, and is especially severe for fruits such as strawberries.

The instability of water in disperse foods (i.e., colloidal systems) is illustrated by the separation of water from oil in emulsions (e.g., low fat spreads) or by water being lost from gels (i.e., syneresis), especially during low-temperature storage. Syneresis of gelled dairy products, e.g., set yogurts, can be a problem, as can gels or thickened foods based on starch. The linear amylose polymers in starch tend to associate (a process referred to as 'starch retrogradation'), to cause gels to undergo syneresis or pastes to go lumpy, especially on long-term frozen/chilled storage. This has led to specialized modified starches being developed for food products designed to be stored frozen. The starch in wheat flour is responsible for the staling of bread, but this is due to the highly branched amylopectin component undergoing retrogradation (i.e., a form of crystallization). Staling refers to the interior of the bread becoming firmer, in contrast to crust firming if the crust dries out. Such staling can be slowed down by storing bread at 30 °C; in fact, stale bread can be 'freshened' by heating it and so reversing, to some extent, the crystallization of the amylopectin.

Pectin-based gels, e.g., in traditional jams, can also show syneresis, and the increasing tendency to store jams in the refrigerator will hasten such an effect – the suggestion is made in order to slow down microbial spoilage, which is discussed in the next section.

Microbiological Degradation Mechanisms

Microorganisms are the major cause of deterioration in foods, not least because foods generally have a high water activity, which microorganisms require in order to grow. (*See* **Microbiology**: Detection of Foodborne Pathogens and their Toxins.) In particular, storage of plant and animal products that are unprocessed is limited by microbial action. Fabricated foods can have water activities that are too low for microbial growth (e.g., margarine, high-sugar-content preserves), and frozen storage (e.g., − 21 °C) can effectively prevent microorganisms growing.

The use of foods as a source of organic compounds by microorganisms is a natural decay process, and the initial stages can give rise to undesirable organoleptic effects, ranging from visual discolorations to off-odors and slime formation. When the process produces toxic effects without the accompanying spoilage indicators, consumers are unable to recognize that the food is unsafe to eat, as in cases of salmonellosis and botulism, and this is a most dangerous situation. (*See* **Food Poisoning**: Classification.)

The food degradation involving microorganisms depends upon the type of microorganism present. Food spoilage arises from the presence of bacteria or fungi, fungi being subdivided into molds and yeasts (*See* **Microbiology**: Classification of Microorganisms.) With a water activity above 0.90, the faster growing bacteria tend to dominate, but with a water activity of about 0.90–0.80, molds and yeasts can be dominant, especially if the pH is below about 5. In fact, yeasts and molds can grow at a pH of around 2. An important difference between yeasts and molds is that molds require oxygen (i.e., are aerobic), whereas yeasts can grow in air (oxygen) or in the absence of air; thus, molds are found on the surface of solid foods (e.g., bread mold), whereas yeasts can be found in liquids (e.g., in acetic acid (vinegar) preserves).

For the large number of different bacteria that can be found in foods, some require air (oxygen) for growth (i.e., aerobic), whereas others can grow in the absence of oxygen (i.e., anaerobic). Thus, if a food is stored in a vacuum, or in a modified atmosphere with no oxygen, the bacteria that flourish will be different from those present in the food when

stored in air. This change to an anaerobic bacterial population may result in a danger to health on consumption of the food, since many anaerobic bacteria are pathogenic, e.g., *Clostridium botulinum*, which has been the cause of deaths from botulism from its occurence in canned meats (e.g., corned beef). (*See* **Canning**: Quality Changes During Canning; **Emerging Foodborne Enteric Pathogens**; **Meat**: Preservation.)

Foods that are especially subject to deterioration resulting from surface mold growth include fruit, cheese, sausage, bread, and cereals. With fruit, the molds produce enzymes (see below) that catalyze reactions leading to the breakdown of pectins and cellulose, with resultant rotting of the fruit. Molds that produce toxins, e.g., aflatoxin, ochratoxin, and patuline, can be particularly dangerous to health – a particular problem can exist with stored grain and nuts, especially peanuts when used as animal feed. (*See* **Aflatoxins**; **Mycotoxins**: Toxicology.)

Yeasts are very useful in producing foods such as bread and alcoholic drinks, but they can also cause food spoilage, especially where food is less severely processed in order to maximize organoleptic properties, or fruit pieces are added to dairy products, or in low-calorie (and hence low-sugar) formulations. The undesirable results of yeast activity include, especially, off-flavor formation and gas production in foods ranging from jams and mayonnaise to yogurts (*See* **Yeasts**.) Gas production as a problem, e.g., in jams, is a result of alcoholic fermentation (i.e.,the desirable process in ethanol production), producing carbon dioxide, i.e.:

$$\underset{C_6H_{12}O_6}{\text{Sugar}} \longrightarrow \underset{2C_2H_5OH + 2CO_2}{\text{Ethanol} + \text{Carbon dioxide}}$$

Bacteria constitute the major microbial mechanism of food degradation, especially for raw foods with a high amount of protein such as raw meats and fish, and also for pasteurized milk and dairy products.

Lactic acid bacteria are able to grow in mildly acidic foods, e.g., milk, and the lactic acid formed is the cause of the 'souring' observed. This is an example of a sugar being used by a bacterium – with resultant formation of a possibly organoleptically undesirable compound. (*See* **Lactic Acid Bacteria**.) Some bacteria, e.g., *Clostridium butyricum* convert sugars to butyric acid, which is particularly foul-smelling with an odor of rancid butter. It can occur in fermented cabbage (sauerkraut). Bacterial decomposition of organic nitrogen compounds in foods, i.e., proteins and amino acids, leads to very offensive odors and potentially dangerous toxins – a process referred to as putrefaction. The major initial change in this process is the decarboxylation of the α-amino acids, i.e.:

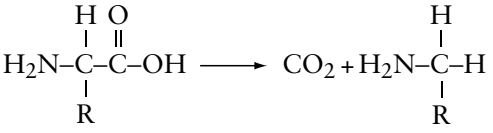

In histamine poisoning, especially relevant for scombroids such as mackerel and tuna, the amino acid histidine is decarboxylated to histamine in this particular type of mild but alarming enteric poisoning. The enteric bacteria responsible do not grow above 10 °C, and hence carefully controlled chilled storage can be used. Histamine is referred to as a biogenic amine; others with very undesirable odors include cadaverine and putrescine. More complex decomposition of the protein, especially using the sulfur groups present, gives rise to the foul-smelling hydrogen sulfide (rotten egg smell) and the thiols, e.g., methanethiol, H_3CSH.

Putrefaction is caused by several bacteria, both aerobic and anaerobic types, and salmonellosis and botulism, as indicated above, represent the dangerous initial stages before putrefaction is organoleptically detectable. With canned foods that are weakly acidic and high in protein, improper sterilization can lead to sulfide spoilage resulting from the growth of the anaerobic bacterium *Clostridium nigrificans*. This results in the production of hydrogen sulfide and blackening of the can interior. Foods that are high in lipid content and consequently low in water activity do not favor bacterial growth. However, bacteria such as the *Pseudomonas* species can cause hydrolysis of lipids, e.g., triacylglycerols, to fatty acids and glycerols (see below also) – and further oxidation to methylketones. This type of rancidity can occur with 'fatty fish' such as herrings and salmon.

The range of bacterial effects on food is enormous – not least because of the wide range of organic compounds that bacteria use for energy, and the possibility of producing a vast array of new compounds of questionable safety or organoleptic quality. Raw fish are particularly prone to bacterial degradation – being examples of high-protein foods as illustrated above. The characteristic fishy smell is due to the production of trimethylamine (TMA), resulting from the bacterial reduction of the naturally occuring trimethylamine oxide (in marine but not fresh water fish), i.e.:

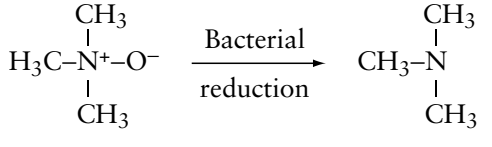

The level of TMA produced can be used as an 'index of freshness,' as the fish decays during storage on ice. This illustrates how a volatile compound can be classed as desirable, or as an off-odor, dependent on its concentration.

Enzymatic and Chemical Degradation Mechanisms

The metabolic processes in plants and animals, which are critical in relation to their storage stability after harvesting or after slaughter, involve a complex series of enzyme-catalyzed chemical reactions (i.e., biochemical reactions). These lead to deterioration of food quality, and the process is normally referred to as autolysis (i.e., self-destruction). The endogenous enzymes causing autolysis are an integral part of the plant or animal food – and any treatment to affect the enzyme activity cannot be done without the possibility of affecting the food itself.

Enzymatic and chemical reactions overlap in the sense that the enzyme speeds up the chemical reaction, but to such an extent that the reaction rate is not significant in the absence of the enzyme. A 'chemical catalyst,' e.g., a transition metal ion such as copper, iron, or nickel, can cause a large increase in the rate, but the reaction will normally proceed at a measurably rate in the absence of the 'chemical catalyst.' Thus, the oxidation in air of fish oils is catalyzed by copper ions, but the oxidation still proceeds measurably in the absence of copper ions. In contrast, the hydrolysis of fish oil occurs readily at room temperature when catalyzed by the enzyme lipase, but in the absence of lipase, the hydrolysis does not occur significantly, not even at 100 °C (at pH 7).

Enzymes are complex proteins, and their catalytic action is related to the shapes taken up by the protein chains. This shape is held together mainly by numerous weak hydrogen bonds, and these can be weakened and changed by processes such as heating, freezing, or changing the pH. The result is a loss of catalytic activity by the enzyme; however, for endogenous enzymes, the effect of such treatments on the organoleptic qualities of the food must be considered. More specific methods of inhibiting enzyme activity include the use of specific enzyme inhibitors such as sulfur dioxide and bisulfites, the use of competitive substrates, or chelation of metals necessary for the activity of the enzyme.

The undesirable effects of enzyme-catalyzed reactions (i.e., autolysis) can be prevented in principle, but often, the effect on the food's organoleptic quality may not be acceptable.

The major chemical components of foods, i.e., lipids, proteins, and carbohydrates, can undergo chemical changes, whether enzyme-catalyzed or not, which can influence the safety, organoleptic quality, or nutritional quality of the food. The major reactions involved are oxidation and hydrolysis, at least initially, but more complex chemical changes may occur subsequently. These can result in the production of rancidity and browning (i.e., discoloration) which are often undesirable (see Figure 1).

In general terms, the chemical changes can:

- produce new compounds – possibly with undesirable toxic or organoleptic properties
- destroy food components of relevance to its nutritional quality (e.g., vitamins) or organoleptic quality (e.g., loss of color).

Some examples of enzymatic and chemical degradation mechanisms are discussed below, but reference to Figure 1 will emphasize how these mechanisms are interconnected, especially in cellular foods.

Enzymatic Degradation Mechanisms

As previously stated, fresh meat and fish deteriorate in quality mainly for microbiological reasons when stored at ambient or chill temperatures. Frozen storage (e.g.,– 30 °C) will prevent microbiological growth, but enzyme activity may still proceed slowly.

Coldwater fish often show rapid autolysis when stored chilled or on ice, and in this case, enzymatic degradation mechanisms are as important as microbiological degradation mechanisms. This arises because coldwater fish live at temperatures near zero, and their enzymes work efficiently in this temperature range – hence, storage in ice, or at refrigeration temperature, affords little reduction in temperature. This contrasts with tropical fish where ice storage is effective in slowing down autolysis. (*See* **Fish**: Spoilage of Seafood.)

Thus, for coldwater fish (e.g., cod, haddock), enzyme activity proceeds noticeably during chilled storage, especially protease activity leading to protein hydrolysis and resultant excessive softening of fish tissue. This tissue breakdown encourages microbial growth, especially involving the breakdown of sulfur containing amino acids to volatile, foul smelling thiols, and disulfides. The fish gut is the source of very active proteases, and gutting of the fish is effective in slowing down this autolysis and so extending the chill storage life. (*See* **Fish**: Processing.) Frozen storage also slows down the enzyme activity, but for gadoid fish, such as cod, haddock, and hake, enzyme activity can convert trimethylamine oxide to dimethylamine and formaldehyde, i.e.:

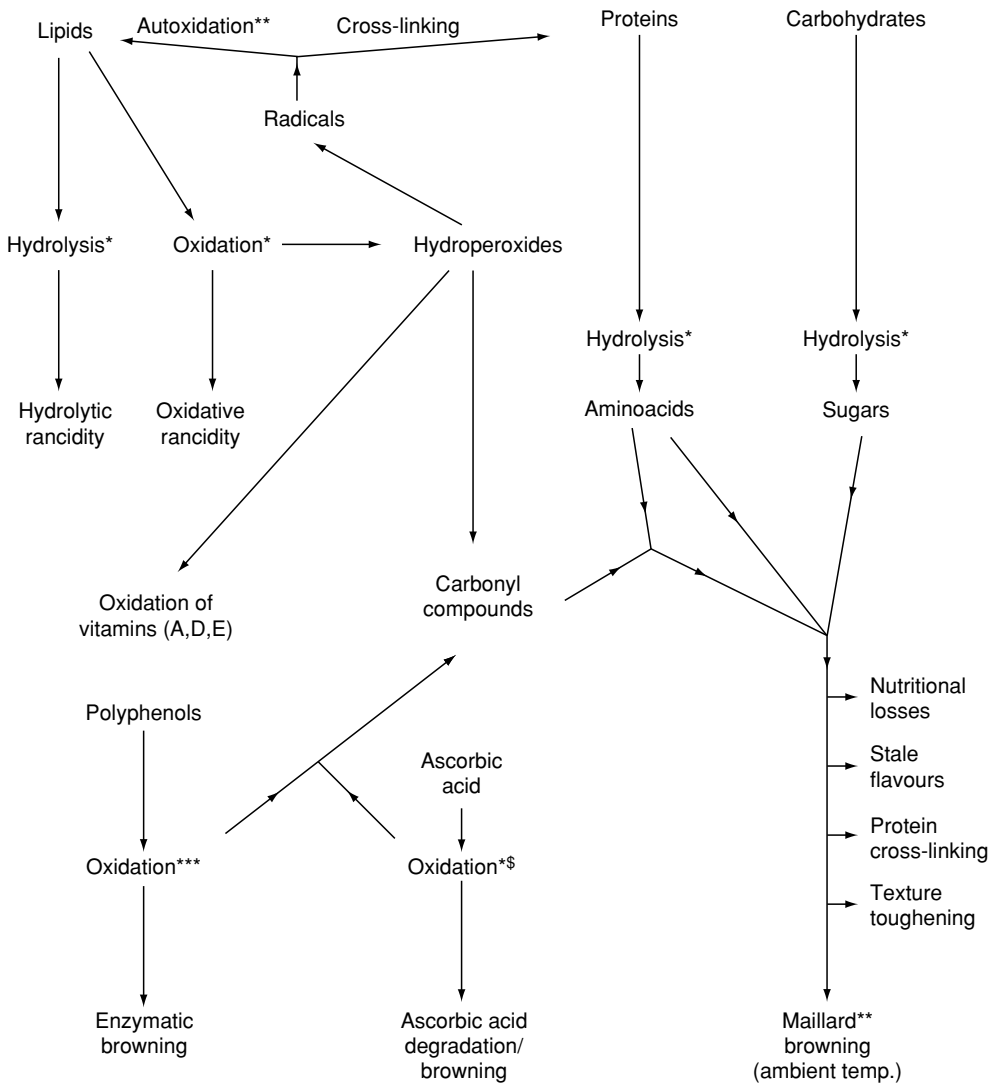

Figure 1 Enzymatic and chemical degradation mechanisms – rancidity and browning. *Enzymatic or chemical mechanism; **chemical mechanism; ***enzymatic mechanism; *$anaerobic also, involving hydrolysis.

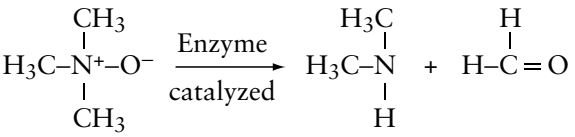

Trimethylamine oxide — Dimethylamine — Formaldehyde

The formation of formaldehyde has been related to the toughening of such fish in frozen storage by its ability to cross-link proteins.

Meat and fish may be high in lipid content, and for fish, the lipid is highly unsaturated and susceptible to oxidation (see Figure 1). Lipoxygenases catalyze oxidation in air of lipids with at least two double bonds (*See* **Oxidation of Food Components**), and specific lipoxygenases have been identified in the gill tissue of fish. The initially formed hydroperoxides are decomposed to volatile carbonyl compounds (aldehydes) with undesirable odors (see Figure 1).

Even though plant foods have a low lipid content, they have various hydrolytic and oxidative enzymes that catalyze hydrolysis and oxidation of the lipids to produce rancidity. The initial action of lipases in producing fatty acids from the hydrolysis of triacylglycerols may not produce the necessary volatile fatty acids for rancidity, but the lipoxygenase enzymes effectively catalyze the oxidation of the unsaturated fatty acids to give the carbonyl compounds, which have the rancid odor, i.e.:

$$\text{Triacylglycerols (unsaturated)} + H_2O \xrightarrow{\text{Lipases}} \text{Fatty acids (unsaturated)}$$

$$\xrightarrow[O_2]{\text{Lipoxygenases}} \text{Carbonyl compounds}$$

This process occurs more readily if the plant food is crushed or ground, allowing the substrates and enzymes to mix. Thus, raw peanuts only go rancid rapidly if they are ground, but if they are dry-roasted, the enzymes are inactivated, and no oxidation occurs.

The blanching process used for unprocessed plant foods (e.g., peas, beans) prior to freezing is necessary to prevent flavor deterioration resulting from enzyme-catalyzed oxidation, especially lipoxygenase. (*See* **Freezing**: Operations.) Additionally, for plant foods, the oxidation (in air) of phenolic substrates catalyzed by a group of enzymes referred to as the polyphenol oxidases (PPO) is important if they are to be processed and stored under ambient or chilled conditions. This is referred to as enzymatic browning (see Figure 1) and is typified by the surface browning of peeled fruit and vegetables (e.g., apples and potatoes). Processing involving chopping, slicing, and peeling allows the substrate, air, and enzymes to mix and promotes enzymatic oxidation of the phenolic substrate to quinones. These undergo further complex nonenzymatic reactions to form melanin pigments, which may be pink, brown, or black.

$$\text{Polyphenol substrate} + O_2 \xrightarrow{\text{PPO enzymes}}$$

$$\text{Quinone} \longrightarrow \longrightarrow \longrightarrow \text{Melanins}$$

The process may be controlled by inactivating the polyphenol oxidase enzyme system – but care must be taken not to cause undesirable changes in texture (e.g., by blanching) or flavor (if the enzyme inhibitor sulfur dioxide is used). Addition of a competitive substrate (e.g., 4-hexylresorcinol) for the enzyme may be more acceptable, or simply the addition of ascorbic acid, which is able to reduce the quinone back to the phenolic compound, hence preventing the conversion of the quinone to the melanins. (*See* **Browning**: Enzymatic – Biochemical Aspects; Enzymatic – Technical Aspects and Assays.)

Plants after harvesting are still respiring, and some plants, fruits in particular, show a large increase in respiration during storage. This is linked to the rapid ripening of some fruits, e.g., bananas and pears, which can limit their desired shelf-life. The enzyme systems involved in the ripening process can be slowed by chilling the fruit (though for bananas, the temperature must not be below 10 °C) and also by decreasing the oxygen concentration and increasing the carbon dioxide concentration (i.e., the reverse of the respiration process). However, the oxygen concentration must not drop below about 3–5% in order to prevent anaerobic respiration, where alternative enzyme systems operate to convert sugar to ethanol (i.e., fermentation) rather than to carbon dioxide and water. (*See* **Ripening of Fruit**.)

During the ripening process of plant foods, the taste and texture develop to give an acceptable sweetness and softness. In relation to the taste, the process involved is the starch – sugar conversion, and this can be temperature-dependent in a complex manner.

The overall starch – sugar (sucrose, glucose, fructose) conversion is as follows:

$$\text{Starch} \xrightarrow[H_3PO_4]{\text{Phosphorylase}} \text{Glucose-1-phosphate}$$

$$\downarrow \text{Several steps involving glycolytic enzymes}$$

$$\text{Sucrose phosphate}$$

$$\downarrow H_3PO_4$$

$$\text{Glucose} + \text{Fructose} \xleftarrow[H_2O]{\text{Invertase}} \text{Sucrose}$$

The starch may be converted mainly to sucrose (a nonreducing sugar), or the sucrose may undergo hydrolysis catalyzed by the enzyme invertase (sucrase) to the reducing sugars glucose and fructose. The mixture of glucose and fructose (so-called 'invert sugar') is sweeter than sucrose – hence, the actual sweetness achieved can vary in relation to the sucrose – 'invert sugar' ratio. An important commercial aspect of the starch–sugar conversion is the storage of potatoes to be used for making crisps (i.e., fried potato slices, called 'chips' in the USA). If the potato has too much of the reducing sugar, the crisp may become too dark as a result of the Maillard browning reaction (see Figure 1). Storage of the potatoes at 4 °C (to reduce respiration) results in 'cold-induced' sweetening, i.e., the reducing sugar level increases, and the crisps produced are too dark. Storage at about 10 °C results in less starch–sugar conversion, and the crisps produced are of an acceptable light brown color. The higher storage temperature (10 °C) appears to inactivate the glycolytic enzymes, hence inhibiting the sugar production. (*See* **Potatoes and Related Crops**: Processing Potato Tubers.)

Chemical Degradation Mechanisms

Storage conditions under which microorganisms and enzymes are inactive (e.g., after heat processing) can

still enable chemical degradation mechanisms to operate, especially where radicals are involved in oxidation reactions (see below). In general terms, however, the conditions, especially temperature, required for rapid chemical reactions are more extreme than used for food storage, where ambient and subambient temperatures are used. However, the long storage times required for processed foods (often in terms of months) can allow time for products from a slow chemical reaction to become noticeable. One example is the Maillard reaction, an example of nonenzymatic browning (see Figure 1) but better considered as a 'carbonyl-amino' reaction. (*See* **Browning**: Nonenzymatic; Toxicology of Nonenzymatic Browning.) The Maillard reaction occurs quickly on heating foods containing amino compounds (e.g., proteins, amino acids) and reducing sugars (e.g., glucose, lactose) and is responsible for the desirable colors and flavors of baked, fried, and toasted foods. The Maillard reaction is slow at normal storage temperatures but, over extended storage periods, can produce undesirable darkening and a stale flavor (e.g., in dried fruits and milk powder) and toughening of texture (through protein cross-linking) in frozen fish.

The Maillard reaction involves a reaction between a polar carbonyl compound (aldehyde or ketone) and a polar amino compound, as outlined below:

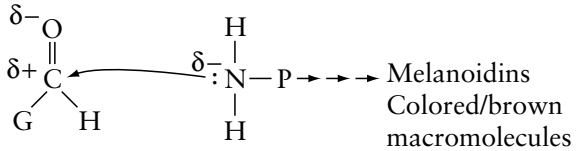

Polar carbonyl compound (e.g., glucose) — Polar amino compound (e.g., protein)

The opposite charges initiate a linking (as indicated by the arrow) between the glucose and the protein, which ultimately produces the brown melanoidins.

The degradation aspects arise from the fact that nutritionally important proteins and essential amino acids react and so become unavailable for digestion and assimilation. Additionally, the complex series of reactions leading to the undesirable brown melanoidins involve side-reactions leading to volatile off-odors (stale odors) for the stored foods.

The broader description of the Maillard reaction as a carbonyl-amino reaction indicates how the carbonyls produced from the oxidation of lipids, polyphenols, and ascorbic acid can also give the Maillard reaction with proteins/amino acids in the absence of reducing sugars (see Figure 1). Ascorbic acid is oxidized readily in air, and the dehydroascorbic acid produced is a diketone, which can give the carbonyl-amino reaction – this explains why some foods brown quickly, even though their reducing sugar content is low. The Maillard reaction (carbonyl-amino reaction) is very slow at pH 3–3.5, but ascorbic acid can oxidize at this pH and still produce darkening as a result of its own specific degradation. Such ascorbic acid degradation is the form of nonenzymatic browning occuring in stored citrus fruit concentrates and spray-dried powders, especially noticeable for lemon and grapefruit. Obviously, the nutritional value as a source of vitamin C is reduced. (*See* **Citrus Fruits**: Lemons; Grapefruits.)

The oxidation of unsaturated lipids giving rise to oxidative rancidity is perhaps the most important chemical degradation mechanism for stored foods, and especially so for refined edible oils and derived products such as margarine and salad creams. *See* **Oxidation of Food Components**. On exposure to air, and especially in the presence of sunlight and transition metals (e.g., copper, iron), unsaturated lipids are oxidized to hydroperoxides by a radical chain reaction normally referred to as autoxidation (self-oxidation). This process is important since the autoxidation, once started, is self-perpetuating, and the hydroperoxides produced not only decompose to rancid odors but are good oxidizing agents, which can cause the destruction of fat-soluble vitamins (A, D, E) and the oxidation of sulfur side groups in proteins leading to their cross-linking, with consequent loss of nutritional quality and toughening of texture. The autoxidation of lipids requires a radical to be formed to initiate the radical chain reaction, this may take a long time, and this induction period may be followed by a surprisingly rapid oxidation, making prediction of storage life difficult (*See* **Storage Stability**: Shelf-life Testing). The initial radical may be derived from enzyme (lipoxygenase)-catalyzed lipid oxidation (during the extraction/processing procedures) producing a trace of hydroperoxide. The subsequent breakdown of this hydroperoxide to radicals gives the autoxidation mechanism its required initial radical (i.e., an unpaired electron system):

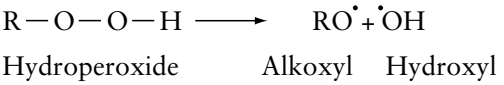

Hydroperoxide — Alkoxyl radical — Hydroxyl radical

For other possible sources of this first radical, (*See* **Oxidation of Food Components**.)

Once a radical has been produced (e.g., OH), the following chain reaction can occur with the lipid (L — H):

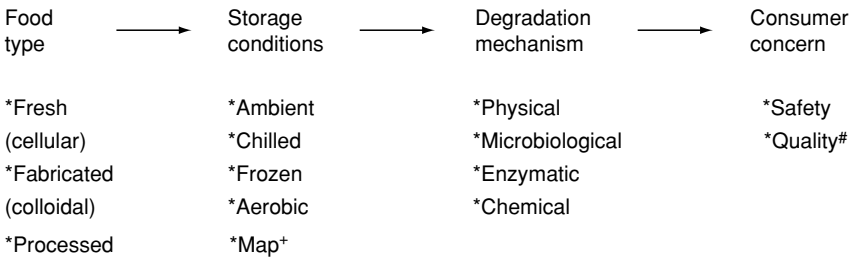

Figure 2 Factors of relevance to mechanisms of degradation of foods. +Modified atmosphere packaging/vacuum packaging; #Both organoleptic and nutritional.

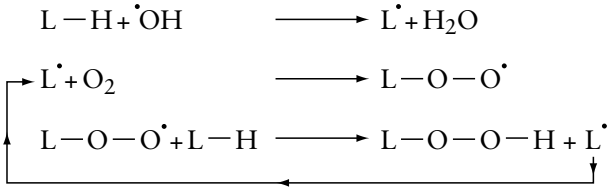

The last two steps provide the repeating chain reaction, and as long as air (oxygen) is present, the conversion of the lipid to hydroperoxide will continue. In addition to the points made above concerning the reactions and decomposition of the lipid hydroperoxide, it has been observed that one particular dialdehyde (malondialdehyde) is produced, which, with protein, gives the Maillard reaction in such a way that protein cross-linking occurs.

The role of transition metals in catalyzing chemical reactions has been referred to above, e.g., the decomposition of the lipid hydroperoxide is catalyzed by transition metals and especially by the heme iron pigments in meat. Other reactions of metals are examples of chemical degradation mechanisms when undesirable precipitates or colors are produced. The heme pigments in meat react in the presence of oxygen to produce either the desirable bright red oxymyoglobin or the undesirable metmyoglobin. (*See* **Meat:** Preservation.) Anthocyanins (phenolic compounds) in fruits can form colored salts with metals, one example being the formation of a pink tin complex on the surface of pears when they are canned. The blue–black discoloration that can occur with some canned peaches is of a similar origin An example of undesirable precipitate formation is the formation of magnesium ammonium phosphate (struvite) in canned crutaceans (e.g., crab). (*See* **Shellfish:** Contamination and Spoilage of Molluscs and Crustaceans.) The crystals formed resemble glass and hence lead to consumer rejection. However, this can be prevented by adding calcium disodium ethylenediaminetetraacetic acid, which preferentially binds with the magnesium. For acidic foods, the reaction of metals with acids to produce hydrogen and dissolve the metal can be a problem during storage of canned foods. (*See* **Acids:** Properties and Determination; and **Canning:** Quality Changes During Canning.) The tin plate used in steel cans can be dissolved by acidic foods, and the dissolved tin concentration could be a health problem – the use of protective lacquers to prevent this is well established. However, the recent recall (November, 2000) of a batch of canned tomato soup because of its high dissolved tin concentration shows that this chemical reaction can still be a problem!

The complex nature of foods makes it difficult to be sure of the degradation mechanisms operating, as reference to Figure 2 will confirm. In some cases, all the mechanisms proposed could be operating, as is found for milk. Milk is highly susceptible to degradation by mechanisms that are:

- physical (e.g., cream (fat) separation);
- microbiological (e.g., both spoilage and pathogenic bacterial growth);
- enzymic (e.g., lipase-catalyzed lipolysis to fatty acids);
- chemical (e.g., 'light-induced flavor' from lipid oxidation and methional formation catalyzed by riboflavin in the presence of light)

(*See* **Milk:** Physical and Chemical Properties.)

The above list of food degradation mechanisms provides a framework that can be used as a starting point for appreciating the degradation mechanisms operating in specific food systems.

See also: **Canning**: Quality Changes During Canning; **Emerging Foodborne Enteric Pathogens**; **Food Poisoning**: Classification; **Freezing**: Operations; Storage of Frozen Foods; **Meat**: Preservation; **Microbiology**: Classification of Microorganisms; Detection of Foodborne Pathogens and their Toxins; **Packaging**: Packaging of Solids; **Preservation of Food**; **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage; Fungi in Food – An Overview; Molds in Spoilage; Yeasts in Spoilage; **Storage Stability**: Shelf-life Testing

Further Reading

Adams MR and Moss MO (2000) *Modern Food Microbiology*, 2nd edn. London: Royal Society of Chemistry.
Eskin NAM (1990) *Biochemistry of Foods*, 2nd edn. New York: Academic Press.
Frankel EN (1998) *Lipid Oxidation*. Dundee, UK: The Oily Press.
Ho C-T and Hartman TG (1994) *Lipids in Food Flavours*. Washington, DC: American Chemical Society.
Jay JM (2000) *Modern Food Microbiology*, 6th edn. New York: Aspen.
Jen JJ (1989) *Quality Factors of Fruit and Vegetables*. Washington, DC: American Chemical Society.
Karel M (1984) Chemical effects in foods stored at room temperature. *Journal of Chemical Education* 64: 335–339.
O'Brien J, Nursten HE, Crabbe MJC and Ames JM (1998) *The Maillard Reaction in Foods and Medicine*. London: Royal Society of Chemistry.
Powrie WD (1984) Chemical effects during storage of frozen foods. *Journal of Chemical Education* 61: 340–347.
Snowdon AL (1990) *A Colour Atlas of Post Harvest Diseases and Disorders of Fruit and Vegetables*. London: Wolfe Scientific.
Wong DWS (1989) *Mechanisms and Theory in Food Chemistry*. New York: Van Nostrand Reinhold.
Wong DWS (1995) *Food Enzymes – Structure and Mechanisms*. New York: Chapman & Hall.

Parameters Affecting Storage Stability

M Hole, University of Lincolnshire and Humberside, Lincoln, Grimsby, UK,

Introduction

In developed countries modern attitudes to food retailing include the demand for even longer shelf-lives at the point of sale and within the home. Together with the increasing demand for convenience foods and complete meals (which can readily be reheated if required, even from the frozen state) the parameters influencing the necessary storage stability must be increasingly better understood. This is especially so considering the trend towards short-shelf-life, chilled foods, which have the desired 'fresh' qualities.

The nature of the changes which influence have been characterized in part (*See* **Storage Stability: Mechanisms of Degradation**) as physical, chemical, enzymatic, and microbial.

The relevance of such changes in influencing storage stability will depend on the treatment of the food prior to storage. It is useful to consider foods as either fresh or processed. Processed foods (e.g., heated, frozen, dried, salted, smoked) generally will have no active microbial population and minimal enzymatic activity prior to storage. In contrast, fresh foods (i.e., not processed and stored at ambient or chill (typically 0–5 °C) temperatures) will have active microbial populations, and some residual enzymatic activity. There is also an increasing demand, both by the catering sector and the individual consumer, for minimally processed fruit and vegetables (e.g., sliced, peeled). Additional problems exist for this form of fresh food in relation to its storage, due to increased enzymatic reactions, e.g., discolorations, and increased susceptibility to microbial growth.

The parameters of major importance for storage stability are temperature, moisture content, packaging and food additives (i.e., added chemicals). Each of these parameters will be considered in turn but it should be appreciated at the outset that the combined use of the above parameters is usually beneficial in practice.

The methods used to extend the storage life of foods are increasingly subject to critical examination by consumers, and this can influence their implementation by the major supermarket chains. Thus the use of irradiation as a means of extending shelf-life is not presently used in the UK, except for some spices. A recent (1999) statement by the Institute of Food Science and Technology (UK) commends irradiation as a safe and effective method for food preservation. Applications identified include:

- certain fruits and vegetables: to reduce spoilage by microorganisms, e.g. molds in strawberries
- poultry: to reduce levels of *Salmonella* and other food-poisoning organisms
- red meats: to reduce levels of *Escherichia coli* 0157:H7 and other food-poisoning organisms
- imported sea foods, e.g., warm-water prawns: to improve their safety during storage
- dried herbs and spices: to reduce food-poisoning bacteria

The use of irradiation allows for extended storage if foods are suitably packed to prevent contamination, although frozen storage may also be beneficial.

The use of irradiation in this way is gaining acceptance in the USA but not in the UK. In trials in the USA irradiated fruit (e.g., strawberries) outsold the non-irradiated form – not least because of their extended mold-free storage period. (*See* **Irradiation of Foods: Basic Principles.**) The major parameters for storage stability described reflect those widely used in present

commercial practice. Consumer opinion is very important when supermarkets consider new methods of storage stability and this is especially the case with genetically modified (GM) foods. Genetic modification of foods can allow extended storage, and a genetically modified tomato with a reduced ripening rate (and hence longer storage period) was used to produce a tomato paste which was successfully sold in UK supermarkets until quite recently, when the supermarkets made a decision not to stock GM foods in general. (*See* **Biotechnology in Food Production; Gene Expression and Nutrition.**)

Temperature

The temperature at which food is stored is the most important parameter which influences storage stability – especially for fresh foods. Biological processes, such as the respiration of fruit and vegetables, enzymatic activity, and microbial growth, such as the mold on cheese or bread, can all be greatly reduced in rate by the use of chill or refrigerated conditions (0–5 °C): whilst in the frozen state (−20 °C typically) they effectively cease. Chemical reactions are typically reduced by 50% of their ambient (20–25 °C) rate by a drop of 10 °C but some reactions can still proceed even during frozen storage if this exceeds a few months. Physical effects resulting in textural changes are one potential disadvantage of frozen storage, especially for intact tissue systems, such as vegetables, fruit, meat, and fish. This has led to the increased use of chilled storage where the 'fresh' image is maintained for such tissue systems. The use of chilled storage combined with modified-atmosphere packaging can extend the storage period significantly. However some fruits and vegetables suffer chilling injury, notably bananas below 12 °C, tomatoes below 7 °C, apples below 2 °C, and cucumbers below 7 °C. (*See* **Chilled Storage**: Use of Modified-atmosphere Packaging; **Chill Foods**: Effect of Modified-atmosphere Packaging on Food Quality.)

Reference to Table 1 shows the much greater storage periods possible with frozen storage but possible texture changes must be considered. An extreme example concerns the strawberry which loses all its firmness on thawing with excessive loss of water ('drip loss'). Frozen storage of eggs leads to the yolk gelling and fish texture can toughen appreciably after frozen storage. Frozen stored milk acquires a gel-like consistency and frozen cream has large aggregates of demulsified fat after thawing. Icecream can acquire a 'sandiness' due to lactose crystallization. The use of cryoprotectants to reduce such detrimental effects is described below. (*See* **Freezing**: Principles.)

Table 1 Temperature effects on shelf-life stability

Commodity	*Storage temperature and shelf-life days (D), months (M) at (°C)*		
	Ambient	*Chill*	*Frozen[a] (−20 °C)*
Fish (lean)	< 1D	5–7D (0)	10–12M
Fish (oily)	< 1D	3–5D (0)	6–8M
Meat (red)	< 1D	4D (2)	6–12M
Pork	< 1D	4D (2)	3–6M
Milk (pasteurized)	0.5D	4D (5)	NS
Bread	2–4D	NS	3M
Peas	1–2D	4–8D (5)	12M
Lettuce (iceberg)	< 1D	10D (5)	NS
Strawberries	1–2D	3D (5)	NS
Cakes (nondairy)	7D	NS	3M

[a]Wrapped to prevent dehydration: see text.
NS, not suitable: see text.

When water freezes to form ice there is a 9% increase in volume and since ice is 'pure H_2O,' the remaining water will have a higher concentration of solutes (so-called 'freeze concentration'). These two aspects can cause undesirable changes, especially textural, during frozen storage. In addition temperature fluctuations during frozen storage will allow small ice crystals to melt and reform as larger ice crystals which can result in cellular damage. Sublimation of ice crystals from the surface of the food due to the low humidity at frozen storage temperatures (−10 °C to −30 °C) can result in 'freeze-drying' of the surface, with the resultant discoloration. (*See* **Freeze-drying**: The Basic Process.)

Despite the above limitations, frozen storage is perhaps the most widely accepted method (in developed countries) of long-term storage of tissue foods, such as vegetables (with prior blanching to inactivate enzymes), meat, and fish. For processed foods such as bakery products and increasingly for complete meals, frozen storage is used and can be combined with microwave treatment for rapid regeneration.

Bread is an example of a processed food where there is no cellular structure but where texture is important – rejection due to 'staling' is an important contribution to waste. Chill storage accelerates such staling but frozen storage below −20 °C will prevent staling for up to 3 months.

More generally, processed foods containing starch can suffer textural changes due to the starch polymer chains associating. Thus starch gels will be lumpy on thawing after frozen storage. An array of structurally modified starches has been produced which can prevent these undesirable changes occurring during frozen storage of such processed convenience foods. (*See* **Starch**: Modified Starches.)

In general the minimal changes in organoleptic properties of appropriate frozen foods, along with

low nutritional changes (e.g., vitamin losses) and the inherent safety from microbial growth at such low temperatures (−20 °C or below) suggest the continued popularity of frozen storage where extended storage times (months) are required. With chill storage the knowledge that some pathogens can grow at 0 °C, e.g., *Listeria monocytogenes*, restricts storage times and necessitates strict time–temperature regimes. However the increasing demands for 'fresh-like' produce has led to improved storage periods at chill conditions by utilizing specific packaging (see below). (*See* ***Listeria***: Properties and Occurrence; **Spoilage**: Bacterial Spoilage.)

Moisture

The moisture (water) content of fresh tissue foods can be very high (above 70%) and this allows microbial spoilage to limit storage stability, if the temperature is not controlled. (*See* **Water Activity**: Effect on Food Stability.)

Foods can be made stable towards microbial changes, and also enzymatic and chemical changes, by reducing the water content. Figure 1 indicates how water activity, a_w (which is related to moisture content but attempts to measure how available the water is), is related to possible changes during storage. Fresh tissue foods have a_w above 0.95, jams (30% moisture) have a_w 0.80–0.75, dried fruits (20% moisture) have a_w 0.65–0.60, pasta (12% moisture) has a_w 0.50, whilst dry foods such as biscuits and cornflakes with 5% moisture or below have a_w below 0.30. Figure 1 would suggest that storage stability is assured if a_w is kept below about 0.3. For processed foods of such low a_w which are consumed directly (rather than rehydrated), their texture can be of critical importance. Thus the crispness of breakfast cereals, biscuits, and potato crisps (chips in the USA) is perhaps the major consumer expectation. Such products are hygroscopic due to their high carbohydrate content and can absorb water with resultant loss of the desirable crisp or brittle texture. Figure 1 also shows that dry foods, if lipid containing, can readily undergo oxidative rancidity – hence a potato crisp with oil content of 35% and a moisture content of 2–3% would be prone to both off-flavor production and loss of crispness during storage. (*See* **Drying**: Theory of Air-drying.)

Dehydrated foods intended for rehydration prior to consumption, e.g., dried potato mash, will absorb moisture with resultant clumping, which adversely affects the texture on rehydration.

In contrast to dried foods, intermediate-moisture foods (a_w 0.6–0.9) represent processed foods storable without refrigeration but which can lose moisture and acquire a dry rather than the desired moist texture – cakes and pastry products are typical examples. (*See* **Intermediate-moisture Foods**.)

The undesirable ingress or loss of moisture depends on the humidity of the surrounding atmosphere and can be controlled by packaging. Where intermediate-moisture foods and fresh (tissue) foods are concerned,

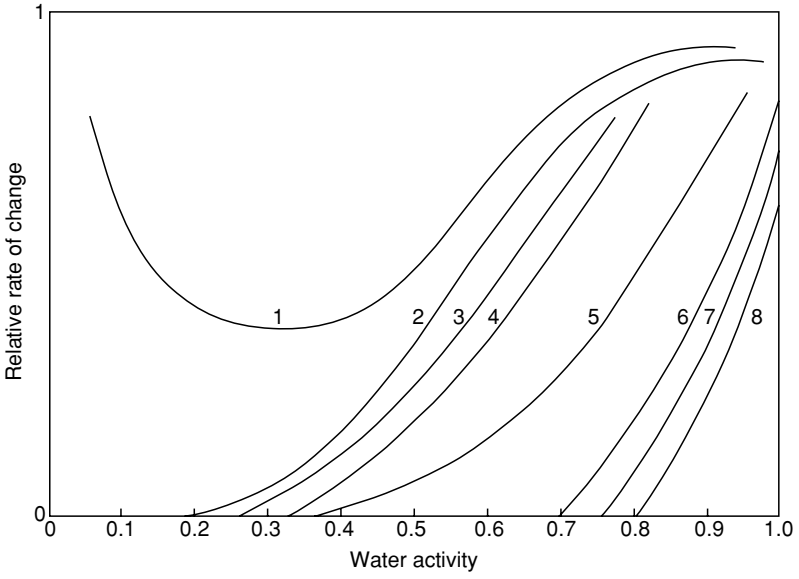

Figure 1 Stability of food as a function of water activity (a_w) at ambient temperature. $a_w = p/p_0$, where p and p_0 are the equilibrium partial vapor pressure of water in the food and of pure water respectively. Key: 1 Lipid oxidation; 2 nonenzymic browning; 3 hydrolytic reactions; 4 loss of ascorbic acid; 5 enzymatic activity; 6 mold growth; 7 yeast growth; 8 bacterial growth.

the high a_w could allow microbial – especially mold – growth. The composition of intermediate-moisture foods is controlled by use of humectants (e.g., sugar), pH, and preservatives in order to prevent migration of moisture and microbial growth. (*See* **Preservation of Food**; **Spoilage**: Molds in Spoilage.)

For fresh fruit and vegetables, where respiration continues, water is produced and will encourage mold growth if not removed, e.g. by storage in a dry atmosphere. However, such water loss must be minimized if undesirable texture changes are not to occur (e.g., loss of firmness of green vegetables and potatoes) – hence a reduction in temperature in order to slow respiration is often used. Bread has a relatively high moisture content (35–40%) and to maintain the desirable firm crust (which has a moisture content of 3–5%) the moisture which migrates from the interior towards the crust must be allowed to escape.

Packaging

Along with the general function of containing the food and preventing contamination, packaging can be an effective barrier against transfer of microorganisms, oxygen and, as already considered, moisture.

Additionally, exclusion of light and increasingly the control of the gaseous contents of the pack are important.

For sterilized foods prevention of ingress of microorganisms is the major requirement and metal containers (cans) have been so used for over 100 years. Flexible pouches and rigid plastic containers – the latter allowing direct reheating by microwave – are increasingly used. With oxygen removal (or reduction) such sterilized foods are storable for up to 12 months at ambient temperature. Detrimental oxidation of lipids, color, and flavors is accelerated by light and hence opaque packaging, e.g., metallized films, is beneficial. Discoloration in (metal) cans, e.g., with high-pH protein-rich foods (meat, fish), is avoided by the use of special lacquers on the inside surface of the can. (*See* **Canning**: Quality Changes During Canning; **Sterilization of Foods**.)

Where lipid oxidation is the major limitation to storage stability, vacuum or inert gas (nitrogen) packaging is beneficial. Vacuum packaging of cured meats and nitrogen flushing of dry roasted nuts and muesli are examples.

Carbon dioxide (CO_2) at -1 to $+2\,^{\circ}C$ has an inhibitory effect on microbial growth and this allows modified-atmosphere packaging (MAP), coupled with chill storage, to give 'fresh' food an extended storage life, which is commercially important for foods such as fresh meat, fish, fruit, and vegetables, and for fresh baked products but not for dairy products (Table 2). For fresh red meat, storage with high levels of oxygen (70–85%) is needed to maintain the desirable red color (otherwise the brown color typical of frozen stored meat is formed). Other examples are shown in Table 2, where the avoidance of oxygen for fatty fish should be noted, along with the high nitrogen concentrations to prevent pack collapse due to the high solubility of carbon dioxide in fish flesh. This latter aspect gives rise to increased 'drip loss' which is absorbed into a pad placed beneath the fish.

For fresh fruit and vegetables, their continuous respiration results in changes in the oxygen and carbon dioxide concentrations in the pack and the permeability of the packaging film to these gases must be considered.

The continuing and rapid respiration after harvesting of fruits, especially climacteric fruits such as bananas and apples, presents particular problems in extending their storage lives. The use of relatively low oxygen concentrations (2–3%) and high carbon dioxide concentrations (approximately 5%), allied with chill temperatures (0–5 °C), can reduce the rate of respiration and give a useful extension of the storage period. The required gas mixture can be achieved

Table 2 Examples of shelf-life extension using modified-atmosphere packaging

Commodity	*Gas mixture (%)*			*Shelf-life days (°C)*
	Oxygen	*Carbon dioxide*	*Nitrogen*	
Fish (lean)	30	40	30	10 (0) 5(+2)
Fish (oily)		60	40	10 (0) 5(+2)
Meat (red)	75–85	25–15		21 (0) 7(+2)
	70	20	10	21 (0) 7(+2)
Poultry		90	10	11 (+2)
Meat (cooked)			100	10 (+2)
Baked products (nondairy)		20–70	30–80	3 months (ambient)
Pasta (fresh)		50	50	3–4 weeks (0 to 5)
Strawberries	1–3	4–6	95–91	7 (+5)
Lettuce (iceberg)	2–3	5–6	93–91	21 (+5)

rapidly by gas flushing, and with permeable plastic packaging it can be maintained if the permeability to oxygen (from outside) and to carbon dioxide (generated inside the pack) is correctly allowed for. It is possible to achieve the required carbon dioxide and oxygen concentrations by simply packaging in air but equilibration then occurs more slowly. Recognition of the role of ethene (ethylene) in initiating the ripening process has suggested the use of 'active packaging' where ethene is removed by oxidation with potassium permanganate, normally included in a sachet within the pack. Such an approach has been used in Japan with kiwi fruit but is not presently (2002) used in the UK.

Genetic modification to prevent ethene formation, and hence ripening, is an alternative possibility and has been demonstrated with tomatoes. Ripening can then be initiated as required by the introduction of exogenous ethene. Such an approach would reduce dependence on MAP and refrigeration, resulting in a simpler and cheaper storage system. However, the present adverse consumer opinion concerning GM foods would be a major problem in applying this approach.

Some novel gas mixtures for use in MAP have recently been shown to have additional beneficial effects for fruit and vegetables. These include the use of very high oxygen levels (70–100%) and the gases argon and nitrous oxide. High-oxygen MAP can inhibit enzymatic discolorations, prevent anaerobic fermentations, and inhibit microbial growth, including aerobic microbial growth. The inhibition of enzymatic discolorations is particularly useful for minimally processed fruit and vegetables (e.g., peeled, sliced), such as apples and potatoes, susceptible to enzymatic browning, and offers an alternative to the use of sulfites.

The use of argon and nitrous oxide MAP is effective in reducing fungal growth, reducing ethene formation, and generally maintaining a higher-quality product, especially by comparison to the use of nitrogen.

The application of edible semipermeable coatings (e.g., semper-fresh) to fruits, especially exotics, has doubled shelf-lives at ambient temperatures. The coating, which can be applied to the ripe fruit at harvest, is edible, invisible, tasteless, and odorless and seals the fruit in its own modified atmosphere, which, while inhibiting oxygen inflow (and hence respiration) allows carbon dioxide to escape and hence prevents anaerobic effects which lead to unpleasant tastes. This approach could usefully be applied to minimally processed fresh produce, e.g., peeled and sliced fruit.

Food Additives

The technical functions of such chemicals, as distinct from any cosmetic, nutritive, or processing-aid functions, can be categorized as:

1. control of microbial changes, e.g., preservatives, a_w control
2. control of enzymatic changes, e.g., pH control, sulfites (SO_2)
3. control of chemical changes, e.g., antioxidants, sulfites
4. control of physical changes, e.g., emulsifiers, stabilizers, cryoprotectants

Traditional ingredients such as salt, sugar, and vinegar also serve as preservatives but are not legally defined as food additives. Hence the traditional use of salt was to reduce a_w and so restrict microbial growth. However the food additive preservatives are effective at very low concentrations, as shown in Table 3. Although high-acid foods (pH below 4) will restrict microbial growth, the preservatives in Table 3 allow control of microbial growth at up to pH 6.5. The majority are weak acids for which the presence of the undissociated acid is necessary for antimicrobial activity. Such preservatives allow storage at ambient temperatures for extended periods, e.g., benzoic acid in fruit squashes and nitrite in cured (canned) meat. Nisin is unusual in being a polypeptide and an

Table 3 Examples of chemical food preservatives

Preservative	*Typical food application*	*Maximum level (mg kg^{-1})*	*Organism affected*
Benzoic acid and benzoates (E210–213)	Low-sugar jams, marmalades	800	Yeasts and molds
Sorbic acid and sorbates (E200–203)	Cheese	1000	Yeasts and molds
	Low-fat spreads	2000	
Propionic acid and propionates (E280–283)	Bread	3000 (mg kg^{-1} flour)	Molds
Sulfur dioxide, sulfites, hydrogen sulfites	Fruit juice	350	
and metabisulfites[a] (E220–228)	Jams	100	Microorganisms
	Dried fruit	2000	
Nitrites and nitrates (E249–252)	Cured meats	150 (nitrite < 50)	Bacteria
Nisin (E234)	Cheese (processed)	12.5	Bacteria

[a]Enzymes are inhibited also.

antibiotic. It occurs in milk and is effective against some bacteria, but not yeasts or molds, and is used in processed cheese. (*See* **Curing; Nisin.**)

The development of low-sugar preserves necessitates the addition of a preservative (often sorbic acid) to prevent mold growth and a similar situation exists for low-fat spreads where the usual 80% fat of margarine can be reduced to below 30%, with corresponding increase in a_w and susceptibility to mold growth.

Table 3 indicates that propionic acid, usually in the form of calcium propionate, is used to control mold growth in bread. However some bakers have chosen to replace this preservative by vinegar, a solution of acetic acid in water. This change was due to adverse consumer perceptions of food additives or 'chemicals' (which could be identified by their E-numbers) in contrast to vinegar which is classed as an ingredient and does not carry an E-number. However the preservative action of vinegar is due to the acetic acid present – a weak acid like propionic acid.

Enzymatic activity varies widely with pH and is reduced by lowering the pH to below 4 but this acidic environment (e.g., use of vinegar) must be acceptable in terms of flavor – otherwise blanching may be preferred. Treatment with sulfur dioxide or sulfites is also effective – a particular application is prevention of darkening of potato slices/chips during chill or frozen storage. Such use for prevention of enzymatic browning of apple slices is restricted due to flavor taints but acid dipping (e.g., citric acid, lemon juice) is effective. Peroxidases in general cause oxidative changes to color and flavor and their control by heat treatment, pH control, or sulfites must be adopted to the particular food system. (*See* **Enzymes**: Functions and Characteristics.)

Chemical changes limiting storage stability are principally lipid oxidation and nonenzymatic (Maillard) browning. The latter is particularly relevant for dried milk, dried egg white, and dried fruits. Oxidative rancidity of lipids is a radical oxidation process and phenolic antioxidants (e.g., tocopherols, butylated hydroxyanisole (BHA), butylated hydroxytoluene (BHT)) are effective at concentrations below 200 p.p.m. In practice combinations of a phenolic antioxidant and ascorbic acid are often used since they have a synergistic effect. (*See* **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; **Browning**: Nonenzymatic; **Oxidation of Food Components**.)

The maintenance of the correct physical state of a food can include maintaining the free-flow properties of a powder (e.g., table salt or icing sugar) and the stabilization of oil–water emulsions (e.g., salad cream), whilst humectants (sugar, glycerol) help to prevent moisture loss and hence dry textures. The texture of a food often depends on it being a colloidal system (emulsion, gel, foam) and additives are necessary to stabilize such systems, especially if long storage times at low temperatures are involved. (*See* **Stabilizers**: Types and Function; Applications.)

Emulsifiers, like the naturally occurring lecithin in egg yolk, help to form stable mixtures (emulsions) of oil and water which are found in salad creams, mayonnaise, margarine, icecream – the last has air incorporated whilst freezing to form a solid foam. To aid emulsion stabilization, e.g., in salad creams, thickening agents (referred to as stabilizers) are used to increase viscosity and so hinder oil and water separation. Such stabilizers are natural polysaccharides like starch, seaweed extracts such as alginates and carrageenans, plant extracts such as gum arabic, carob gum (locust bean gum) and pectins, and microbially produced xanthan gum. Blends of these so-called hydrocolloids are widely used to create and to maintain texture in chilled and frozen processed foods, especially dairy products ranging from icecream to gelled milk products and yogurt. (*See* **Emulsifiers**: Organic Emulsifiers; Phosphates as Meat Emulsion Stabilizers; Uses in Processed Foods.)

Less obvious applications are the use of the emulsifiers to retard bread staling, the use of a suspending agent, often carboxymethyl cellulose, to maintain the water-insoluble carotene pigment as a fine dispersion in orange squashes (and so prevent the 'orange scum' at the top of the bottle!), and the use of glyceryl monostearate in dried potato mash to insure rehydration without stickiness developing.

Additives can usefully be incorporated into products to minimize textural changes during their frozen storage. The use of hydrocolloids in this connection has been referred to above, and they are often referred to as cryoprotectants in this connection. Thus the addition of cryoprotective alginates to icecream helps to prevent ice-crystal formation (and hence 'grittiness') during storage. The manufacture of surimi (frozen, stabilized washed fish mince) depends on the incorporation of a cryoprotectant mixture consisting of 4% sucrose, 4% sorbitol, and 0.3% polyphosphate. This prevents protein denaturation during frozen storage, and so maintains the heat gelation properties of the fish proteins required for the production of products such as shellfish analogs (crab sticks), fish sausage, noodles, and burgers. Recently so-called 'antifreeze proteins' have been isolated from cold-water (e.g., Antarctic) fish and shown strongly to inhibit ice-crystal formation in icecream products and to reduce ice-crystal formation in frozen meat, with resultant reduced drip loss on thawing. Presently the high cost of these cryoprotectants precludes their commercial application.

Shelf-Life Extension

It is clear from the examples mentioned that methods adopted to insure shelf-life extension must consider the state of the food at the point of storage and select the best combinations of the relevant parameters.

Organoleptic reasons and concepts of naturalness are increasingly important factors in determining the methods used to insure the required storage stability. Thus the increased use of MAP allied with the use of chill rather than frozen storage is related to its 'fresh' appeal, in contrast to frozen food which has a 'processed' image. Heat sterilization and sealing under reduced air content, as typified by traditional canning, or frozen storage at −20 °C or below allows safe storage for many months but flavor/texture changes may arise.

The sous-vide or nouvelle carte processes have been promoted as maximizing flavor, texture, and nutrient content, as well as being bacteria-free. These processes involve vacuum packing in heat-resistant packaging, cooking the vacuum-packed product under vacuum in a moist steam oven or pressure cooker at less than 110 °C, rapidly cooling, and storing at 1–4 °C for up to 3 weeks. This represents the combination of chill storage, relevant packaging, and exclusion of oxygen to extend shelf-life. Where low-acid foods are concerned, the additional use of preservatives is suggested and overall a maximum storage time of 2 weeks is recommended.

When chill storage is used, the shelf-life is measured in days and there is less of a margin for storage abuse than with frozen storage. For foods which are chill-stored and which are not reheated (above 70 °C) before consumption, the recommended chill storage temperature and time must be carefully observed. This places more emphasis on accurate sell- and use-by dates by the processor, careful stock control by the retailer, and increased awareness by the consumer.

See also: **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; **Biotechnology in Food Production**; **Canning**: Quality Changes During Canning; **Chilled Storage**: Use of Modified-atmosphere Packaging; **Curing**; **Drying**: Theory of Air-drying; **Freeze-drying**: The Basic Process; **Freezing**: Principles; **Irradiation of Foods**: Basic Principles; **Oxidation of Food Components**; **Preservation of Food**; **Spoilage**: Bacterial Spoilage; Molds in Spoilage; **Stabilizers**: Types and Function; Applications; **Sterilization of Foods**; **Storage Stability**: Mechanisms of Degradation

Further Reading

Ahuenainen R (1996) New approaches to improving the shelf life of minimally processed fruit and vegetables. *Trends in Food Science and Technology* 7: 179–187.

Coultate TP (1996) *Food – The Chemistry of its Components*, 3rd edn. London: Royal Society of Chemistry.

Debeaufort F, Quezada-Gallo JA and Voilley A (1998) Edible films and coatings: tomorrow's packaging: a review. *Critical Reviews in Food Science* 38: 299–313.

Feeney RE and Yeh Y (1998) Antifreeze proteins, current status and possible food uses. *Trends in Food Science and Technology* 9: 102–106.

Fennema OR (ed.) (1996) *Food Chemistry*, 3rd edn. New York: Marcel Dekker.

Institute of Food Science and Technology (UK) (1999) Position statement: the use of irradiation for food quality and safety. *Food Science and Technology Today* 13: 177–179.

IFST (UK) Current Hot Topics (1999) *Position Statement: Genetic Modification of Food*. http//ifst.org/hottop10.htm.

Jeremiah LE (1996) *Freezing Effects on Food Quality*. New York: Marcel Dekker.

King GA and O'Donaghue EM (1995) Unravelling senescence: new opportunities for delaying the inevitable in harvested fruit and vegetables. *Trends in Food Science and Technology* 6: 385–389.

Rockland LB and Beuchat LR (1987) *Water Activity: Theory and Application to Food*. New York: Marcel Dekker.

Taub IA and Singh RP (1998) *Food Storage Stability*. Boca Raton, Florida: CRC Press.

Vesmeiren L, Deulieghere F, van Beest M, de Kruijf N and Debeuere J (1999) Development in active packaging of foods. *Trends in Food Science and Technology* 10: 77–86.

Shelf-life Testing

M Hole, University of Lincolnshire and Humberside, Lincoln, Grimsby, UK

Introduction

The shelf-life of a food is perhaps the most important detail a consumer needs to know in relation to their purchase and home storage of foods. This information is embodied in the 'best before' and 'use by' dates indicated for the food, and which are required by law to be clearly shown on the food packaging. The 'use by' date is particularly critical in relation to the safe consumption of a purchased food. Such foods have relatively short shelf-lives, often in terms of days, and include fresh and chilled meats and dairy products. In contrast, 'best before' dates are used for foods where quality, rather than safety, is the practical concern – thus, a potato crisp, which has absorbed moisture so that it has lost its crispness, is of undesirable quality but does not pose a food safety issue.

If the manufacturer should propose an unrealistic shelf-life, leading to a 'use by' date that is too long, then the likely consumer complaints due to adverse health and/or quality aspects, possibly requiring product recalls, would do great damage to the manufacturer's reputation. The subsequent commercial consequences for the food manufacturer would be serious, and it is obvious that the correct determination of the shelf-life is most important – for both the manufacturer and the consumer. The consumer can obtain very useful information on the purpose of 'use by' dates from the Foodsense leaflets provided by the Ministry of Agriculture, Fisheries and Food.

In unprocessed and unpreserved foods, the factors that limit the shelf-life are mostly related to the growth of spoilage microorganisms, whilst, in non-perishables, chemical and physical factors become more important, e.g., development of off-odors and off-flavors due to rancidity in fried snack foods, moisture migration in biscuits, and staling of bread and cakes. The factors that limit the shelf-life of a selection of foods are indicated in Table 1.

In order to give an indication of the shelf-life of a product, some type of test must be available to the manufacturer. Storage studies would normally be performed as new products are developed, with regular testing for the shelf-life limiting factors. However, with some long-term preservation methods, it may be more convenient to have accelerated shelf-life studies. This would enable production to commence quite rapidly after the initial development of the product.

Table 1 Factors that limit the shelf-life of foods

Type of food	*Acceptability criteria*
Bread	Mold growth
	Loss of moisture, staling
Breakfast cereals	Rancidity development
	Moisture gain/loss of crispness
	Vitamin loss
Dried pasta	Moisture gain (or loss)
	Color loss
	Adsorption of undesirable flavors
Fried snack foods	Oxidative and hydrolytic rancidity
Fresh poultry	Pathogen growth/microbial decay
	Bruising
Frozen poultry	Sensory quality changes
	Color changes/rancidity
	Syneresis, moisture loss
Fresh meat	Bacterial growth
	Loss of color
Fresh fish	Bacterial decomposition
Frozen fish	Lipid oxidation
	Protein denaturation (toughening)
Frozen vegetables	Color
Dairy products and milk	Bacterial growth
	Lipid hydrolysis/flavor changes
Ice cream	Textural changes
	Oxidation of lipid
Evaporated milk	Loss of vitamins
Salad dressing	Breakdown of emulsion
	Color/flavor changes
Canned foods	Taste/flavor changes
Chilled foods	Bacterial growth
	Flavor changes

Experimental Design

With any experimental procedure, results will only be valid if the initial sampling and testing stages are carefully planned and controlled. If any statistical analysis of the result is required, then the testing procedure should be designed to be appropriate to the statistical method being applied. The samples taken should be representative of the whole batch and, in careful shelf-life studies, samples should include any extremes of product storage conditions. It is also important to test at regular intervals during production, especially after any change in ingredients or methods of production. The design of any procedure for testing shelf-life should include consideration of the following factors:

- elucidation of the major types of quality loss for the product (see Table 1 for examples);
- knowledge of the factors that control the initial quality during manufacture, e.g., use of food additives;
- the environmental conditions of storage, e.g., temperature, equilibrium relative humidity (ERH), and light;
- the type and properties of the packaging to be used, e.g., permeability to oxygen, light and moisture;
- the kinetics of the reaction or reactions leading to loss of quality in the actual food (stored in the usual way).

Acceptability Criteria

The acceptance of a food product depends to some extent on the individual preferences of the consumer; however, all consumers have a right to expect wholesome foods that will not cause disease, are of attractive appearance, and are of a good (unchanged) nutritional value. The major attributes of acceptability can be described as: absence of pathogenic microorganisms; low levels of spoilage organisms; absence of off-flavors; no deterioration in color or overall appearance; no loss of desired texture; and high (unchanged) nutritional value. It is the function of the manufacturer to decide upon suitable acceptability limits for each type of product, and this should form part of the specification of that product.

Methods of Testing

It is vital that the tests carried out need to relate specifically to the safety aspect concerned or, in quality terms, relate to the limiting quality criteria.

The main methods of testing the shelf-life of a food include chemical/physical tests, microbiological enumerations and sensory evaluations. Due to the complex nature of most composite foods, many chemical tests, e.g., to establish nutrient levels, are very lengthy, often involving complex extraction procedures and use of sophisticated and expensive equipment. Microbiological and sensory tests are generally regarded as easier to perform and are more adaptable for use in nonspecialized laboratories.

Although the ultimate test of a product's acceptability or attractiveness lies with the consumer, the consumer cannot necessarily judge whether foods are safe.

Microorganisms

The first criterion of any testing is to insure that the product is safe – both from food-poisoning microorganisms and from any toxins. Food-poisoning organisms, e.g., *Salmonella typhimurium*, may be detected by growth of the organism on selected media, but growth is usually relatively slow, and a result would not be obtained until 3–4 days after the test had been initiated. Rapid methods of measuring microbial concentrations are now available, and a range of automated instrumental methods can be used, e.g., the 'Bactometer'. (*See* **Food Poisoning**: Tracing Origins and Testing; **Microbiology**: Detection of Foodborne Pathogens and their Toxins.)

Chemical Toxins

A variety of tests can be utilized to test for the range of toxins found in foods. For example, the presence of toxic metals, e.g., mercury or lead, may be determined by the use of atomic absorption spectrometry, and the highly toxic aflatoxins, produced as secondary metabolites from the growth of the mold *Aspergillus flavus* on cereals, may be detected by using the fluorescent properties of the compounds, e.g., use of high-performance liquid chromatography with fluorescence detection. (*See* **Heavy Metal Toxicology**; **Mycotoxins**: Occurrence and Determination.)

Flavor Changes

From Table 1, it is apparent that the shelf-life of 'microbiologically safe' foods is often limited by either changes in flavor or development of off-odors and off-flavors, e.g., production of volatile carbonyl compounds from lipid oxidation leading to rancid flavors. These changes in flavor may be monitored by sensory means or from utilization of analytical techniques such as headspace gas chromatography. With this technique, the volatiles arising from a foodstuff are swept into a chromatography column contained in a thermostatically controlled oven. A carrier gas moves the flavor compounds through the column where selective separation occurs, depending on the polarity of the packing material (stationary phase) of the column and the molecular structure of the volatiles. Eventually, the separated compounds are detected and are registered as a peak on a trace. The identity of the peaks may be established by comparison of their retention times with standards or from coupling the gas chromatography with a mass spectrometer. The contribution of a specific volatile to any off- or undesirable odor may be found by smelling the individual components as they emerge from the column. Increasing use is made of 'electronic noses' (e.g., 'Aromascan') for the instrumental characterization of aroma. In contrast to gas chromatography, these electronic noses detect the overall aroma, giving rise to an aroma fingerprint that can be used to indicate the shelf life limit. (*See* **Sensory Evaluation**: Aroma.)

It is well established that a wide range of analytical tools are available to the food analyst for determination of the (molecular) components of foods that may be responsible for any shelf-life deterioration, e.g., nutrients (essential amino acids and vitamins). In some foods, determination of the factors that actually limit shelf-life may involve very complex analyses; thus an associated factor, which changes at the same rate but does not affect shelf life directly, may be used. The reader is referred to the bibliography for a more thorough review of methods used in the chemical analysis of foods.

Color

Changes in the color of foods are often major factors limiting shelf-life. Instrumental testing of surface color, or of the color of a well-homogenized sample, may be achieved with a colorimeter, which gives results in terms of the Hunter L, a and b values. That is, a measure of the degree of 'whiteness,' 'redness,' and 'blueness' of a sample. Otherwise, comparisons of the food color with acceptable colors are useful, but this is essentially a sensory method of analysis. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments; Properties and Determinants of Synthetic Pigments; **Sensory Evaluation**: Appearance.)

Texture

Sensory methods are also used to provide information on the texture of foods. Attributes such as firmness, crispness, juiciness, and chewiness are fairly typical of those used. In addition, the texture of certain foods, e.g., biscuits and apples may be measured using physical (instrumental) tests and equipment such as the Instron. This particular piece of equipment may be modified to undertake a variety of penetration, extensibility, or shearing tests. (*See* **Sensory Evaluation**: Texture.)

Sensory Methods

A wide variety of sensory evaluation tests are available for use in testing the acceptability of foods. The tests may be classified into those that are affective (i.e., ask for an opinion or preference) and those that are nonaffective (i.e., require a score to be assigned for a particular attribute). However, for shelf-life studies, it is usually more appropriate to use scoring tests. For the results to be meaningful, it is important to have a suitably designed taste panel/sensory test giving results that can be statistically validated. This also requires the establishment of a null hypothesis, e.g., no difference between the samples presented, in order to establish the significance of the results. It is usually recommended that at least 10 persons are required to act as panellists if the results obtained are to be valid, the actual number depending on their level of training (if untrained, then more panellists will be required). Sensory panels are sometimes regarded as time-consuming and expensive to set up, but they do have an advantage over instrumental methods in that several attributes of the food can be measured in one session. (*See* **Sensory Evaluation**: Sensory Characteristics of Human Foods; Food Acceptability and Sensory Evaluation; Practical Considerations; Descriptive Analysis.)

Accelerated Shelf-life Testing

Ideally, a food would be tested under the conditions of intended storage in order to measure the length of time over which it is acceptable, whether from a safety or quality aspect. For a food with a shelf-life of a few days or weeks (e.g., chilled meat), this is no problem, but for foods with shelf-lives of months, or years, the manufacturer would like a quicker method, i.e., one that gives a result in a shorter time than the shelf-life under the intended storage conditions.

Since storage temperature is usually the main factor influencing the shelf-life, most accelerated methods simply use a higher temperature than the intended storage temperature and hope to relate the shorter shelf-life at the higher temperature to the shelf-life at the intended storage temperature. The theoretical basis is related to the Arrhenius relationship, one form of which is:

$$\ln k = cst. \frac{-Ea}{RT},$$

where *cst.*, *Ea* (the activation energy), and *R* (universal gas constant) can be taken as constants. Hence:

$$\ln k \propto \frac{1}{T},$$

where T is the temperature (degrees Kelvin), and k is the rate constant, which, in this context, can be related to the shelf-life time, i.e.:

$$\ln(\text{shelf-life}) \propto \frac{1}{T}.$$

A graph of ln(shelf-life) against $1/T$ will give a straight line (see Figure 1). Thus, if the shelf-life is measured

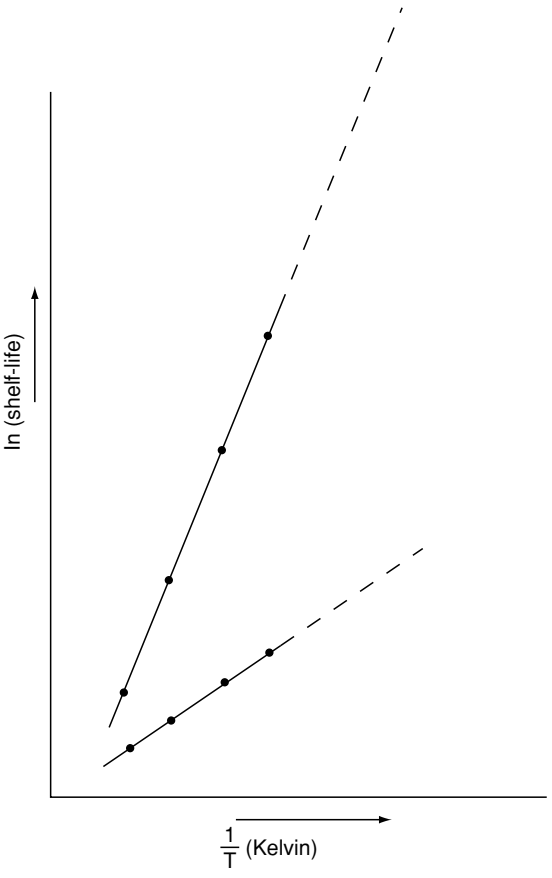

Figure 1 Arrhenius plot showing extrapolated shelf-lives for two foods with different rates of quality deterioration.

for a series of temperatures higher than the intended storage temperature, extrapolation will give the shelf-life at the lower temperature. The method has limitations and can be costly, but this approach is often used. Some examples are given below; other examples will be found in the bibliography. Frozen foods often have shelf-lives of many months, or years, and accelerated shelf-life testing can be usefully applied. However, the elevated temperature in this case must be realistic – a temperature at which the food was not frozen would not be realistic. Thus, the shelf-life at an intended storage temperature of $-30\,^{\circ}C$ might be usefully indicated from shelf-life measurements made at up to $-5\,^{\circ}C$. This would ensure that the food matrix was similar (i.e., frozen) at the elevated temperatures. This illustrates an important point; the nearer the elevated temperatures are to the intended storage temperature, the more realistic the extrapolated shelf-life will be.

Where foods are being reformulated and the effect of different formulations on shelf-life is needed, accelerated methods are very useful in indicating relative shelf-lives. This allows a new shelf-life to be proposed from a knowledge of the original shelf-life. It has been stressed that the shelf-life determinations are only relevant if based on measurements of the actual limiting factor, for example, odor. This aspect can become critical when accelerated shelf-life testing is carried out, if the limiting factor changes at the higher temperature. An example (discussed by Labuza – see Further Reading) is provided by the storage of dehydrated mashed potato. At ambient temperatures, the limiting factor is off-odors from lipid oxidation, but above about $40\,^{\circ}C$, the limiting factor for organoleptic acceptance is discoloration due to Maillard browning (*See* **Browning**: Nonenzymatic). Hence, if accelerated shelf-life testing is carried out above $40\,^{\circ}C$, sensory analysis will give results that extrapolate to an incorrect shelf-life for ambient temperature storage. More specifically, results obtained above $40\,^{\circ}C$ extrapolate to a shelf-life above 1000 days at $20\,^{\circ}C$, whilst measurements made at $20\,^{\circ}C$ (where odor is the limiting factor) indicate a shelf-life of about 500 days.

Thus, the food system must be carefully studied to insure that the accelerated shelf-life testing is done as realistically as possible. The sensitivity of a food to temperature changes is indicated by the slope of the 'Arrhenius graph' (*See* Figure 1) – the more sensitive food having the larger gradient. An alternative approach is the use of the Q_{10} value, where

$$Q_{10} = \frac{\text{rate at } (T+10)^{\circ}C}{\text{rate at } T\,^{\circ}C}$$

or

$$Q_{10} = \frac{\text{shelf-life at } T\,^{\circ}C}{\text{shelf-life at } (T+10)^{\circ}C}.$$

The more sensitive the food is to temperature change, the higher the Q_{10} value, as illustrated in Table 2. For an initial shelf-life of 100 days at $T\,^{\circ}C$, a food with $Q_{10}=2$ would have a half-life of 50 days at $(T+10)\,^{\circ}C$, whilst if $Q_{10}=4$, the half-life would be 25 days at $(T+10)^{\circ}$.

Table 2 Q_{10} values for selected foods

Type of food	*Q_{10} value*	*Criteria for end of shelf-life*
Fresh cod	4.4	Bacterial growth
Sterilized milk	1.71	Flavor change
Pasteurized milk	2.64	$>10^6$ CFU[a] ml^{-1}
Pasteurized egg	5.37	Flavor change
Spray-dried egg	1.21	60% loss of vitamin A
Margarine	1.91	25% loss of vitamin A
Canned kidney beans	1.7	20% loss of thiamin

[a]CFU, colony-forming units.

Prediction of Shelf-life

The use of accelerated shelf-life testing avoids the need for prolonged storage studies – and in reality is a predictive method based on measurements made at higher temperatures than the actual storage temperature. The possible problems with accelerated shelf-life testing can be minimized if predictive methods using data from the actual storage conditions can be used. Thus, if the rate of deterioration (i.e., the kinetics) is known (at the actual storage temperature), initial values of the loss of quality, for example, can be used to indicate the quality loss, and hence shelf-life, for extended storage times.

The loss of quality for most foods can be represented by:

$$-\frac{dA}{dt} = kA^n,$$

where A is the quality factor measured, e.g., extent of nonenzymatic browning, t is the time, k is a constant that depends on the temperature and water activity, n is a power factor giving the order of the reaction and dA/dt is the rate of change of A with time (negative, loss of A; positive, production of undesirable end products). In practice, zero-order ($n=0$) or first-order ($n=1$) kinetics are relevant for quality changes in foods.

Zero-order Kinetics

The order of the reaction, n, defines whether the rate is dependent on the value of A. Many food systems

are assumed to behave as zero-order reactions ($n = 0$). Thus, the rate of loss is constant under constant storage conditions. Thus:

$$-\frac{dA}{dt} = k,$$

which may be integrated to give

$$A = A_o - kt \quad \text{or} \quad A_e = A_o - kt_s,$$

where A_o is the initial quality value at time t, A_e is the value of A at the end of the shelf-life and t_s is the shelf-life in days, months, or years.

A may be defined analytically or by taste-panel evaluation. If A_o is assumed to be 100% quality and A_e just acceptable quality, then:

$$k = \frac{A_o - A_e}{t_s} = \frac{100\%}{t_s},$$

Thus, based on this knowledge, it is possible to predict the shelf-life of a food at a single given temperature, if the value of the quality change at any given time is known.

For example, if it is known that a certain food has lost 25% of its quality in 50 days when held under constant conditions, then the rate of loss:

$$k = \frac{A_o - A}{t} = \frac{100 - 75}{50} = 0.5\% \text{ per day}.$$

The amount of shelf-life left as a function of time may also be represented by a graphical plot (Figure 2).

Reaction types in foods that are thought to give zero-order kinetics include enzymatic degradation, nonenzymatic browning, and lipid oxidation (rancidity). Loss of quality of frozen foods in general follows zero-order kinetics.

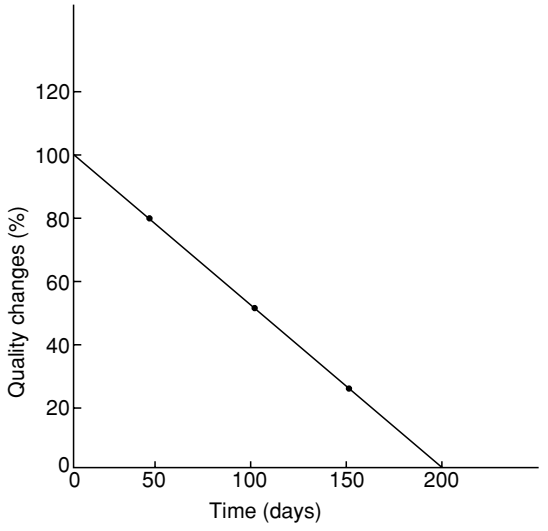

Figure 2 Zero-order reaction plot showing residual shelf-life against time ($k = 0.5\%$ per day).

First-order Kinetics

Many foods deteriorate by first-order kinetics ($n = 1$), which results in an exponential decrease in the rate of change as the quality decreases. Thus, the rate of quality loss is directly dependent on the amount left:

$$-\frac{dA}{dt} = kA.$$

Integration gives:

$$\ln (A_c/A_o) = -kt_s.$$

A semi-logarithmic plot of A_c/A_o against time (t) gives a straight line with slope k.

The types of deterioration that follow first-order kinetics include microbial growth (fresh meat and fish), microbial production of off-flavors, vitamin losses (canned and dried foods), and loss of protein quality (dried foods). (*See* **Canning**: Quality Changes During Canning; **Fish**: Spoilage of Seafood; **Meat**: Preservation.)

A knowledge of the kinetics of food deterioration has been incorporated into computer packages, which can be applied to rapidly predict shelf-lives for more complex situations.

Such computer-based predictive models were developed initially to predict the growth of pathogens in foods in order to predict shelf-lives where the food would be safe to eat. The 'Food Micromodel', developed by the Ministry of Agriculture, Fisheries and Food in the early 1990s is such a model and can be applied to predict the growth of a wide range of microorganisms, both pathogenic and spoilage organisms, under a wide range of storage conditions. The approach is particularly relevant for chilled foods, and the 'Forecast' service provided by Campden and Chorleywood Food Research Association (CCFRA) allows the prediction of the growth of common spoilage bacteria for a range of storage conditions (pH, temperature, salt content, etc.). Another example of a computer-based predictive model (also from CCFRA) is the prediction of mold-free shelf-life using the ERH CALC™ package. This model allows rapid calculation of the ERH (or water activity) and hence the growth of molds from a knowledge of the ingredients and storage conditions of bakery products.

Time–Temperature Indicators (TTIs)

Since temperature is such an important factor in determining shelf-life, it would be useful to know the full thermal history of a food after processing. Any thermal abuse during distribution and storage would alter the shelf-life, and hence the 'use by' date of the food.

Time–temperature indicators are available that give a full thermal history of the food, and such sensors attached directly to the packaged food would be most useful in this connection. If they were linked to the computer-based models for shelf-life prediction, any temperature abuse would result in a new shelf-life prediction.

Home Storage

There is an obvious need for the manufacturer's storage instructions to be followed in order that the 'use by' date is valid. The food industry can arrange the correct storage conditions up to the point of sale, but thereafter, it is necessary for the consumer to insure correct home storage, especially for chilled foods. The possible lack of correct temperature control during home storage indicates the need for some margin of error to be incorporated, especially into the 'use by' date.

See also: **Fish**: Spoilage of Seafood; **Food Poisoning**: Classification; Tracing Origins and Testing; **Heavy Metal Toxicology**; **Meat**: Preservation; **Microbiology**: Detection of Foodborne Pathogens and their Toxins; **Mycotoxins**: Occurrence and Determination; **Preservation of Food**; **Water Activity**: Effect on Food Stability; **Sensory Evaluation**: Sensory Characteristics of Human Foods; Food Acceptability and Sensory Evaluation

Further Reading

Campden and Chorleywood Food Research Association (CCFRA) (1997) *Evaluation of Shelf Life for Chilled Foods, Technical Manual No. 28*. Chipping Campden, UK: CCFRA.

Campden and Chorleywood Food Research Association (CCFRA) (1999) *ERH CALC™ – Version 2*. Chipping Campden UK: CCFRA.

Fennema OR (ed.) (1996) *Food Chemistry*, 3rd edn. New York: Marcel Dekker.

Institute of Food Science and Technology (IFST) (1993) *Shelf Life of Food – Guidelines for its Determination and Prediction*. London: IFST.

Labuza TP (1984) Application of chemical kinetics to deterioration of foods. *Journal of Chemical Education* 61: 348–358.

Ministry of Agriculture, Fisheries and Food (MAFF) (1998) *Understanding Food Labels*. London: Foodsense, Food Safety Directorate.

Man CMD and Jones AA (1994) *Shelf Life Evaluation of Foods*. London: Blackie Academic and Professional.

Nielson SS (1998) *Food Analysis*, 2nd edn. Gaithersburg, MD: Aspen.

Singhal RS, Kulkarni PR and Rege DV (1997) *Handbook of Indices of Food Quality and Authenticity*. Cambridge, UK: Woodhead.

Smith G (1993) *Encyclopaedia of Food Science, Food Technology and Nutrition*. London: Academic Press.

Thorne S (1992) *Mathematical Modelling of Food Processing Operations*. London: Elsevier Applied Science.

Stouts *See* **Beers**: History and Types; Wort Production; Raw Materials; Chemistry of Brewing; Biochemistry of Fermentation; Microbreweries

STRAWBERRIES

R B Smith and L J Skog, University of Guelph, Vineland Station, Ontario, Canada
A Dale, University of Guelph, Simcoe, Ontario, Canada

Background

The strawberry belongs to the genus *Fragaria*, and is a member of the rose family, Rosaceae. The cultivated strawberry, *F.* × *ananassa* L. Duchesne, is a hybrid between *F. chiloensis* Duch and *F. virginiana* Duch. The *ananassa* designation is due to the resemblance to the pineapple in flavor, odor, and shape. *F. chiloensis* was introduced into Europe from Chile and *F. virginiana* from eastern North America. The fruit of the strawberry is grown and enjoyed worldwide. This article discusses the global distribution, varieties, commercial importance, morphology, nutritional composition, handling, storage, and industrial uses of this crop.

Global Distribution

The strawberry plant is highly adapted to a wide range of climatic and soil conditions. Commercial plantings of strawberries are found on every continent (except Antarctica) and from the subarctic (Finland) to the tropics (Ecuador). However, most of the world's production is in the northern hemisphere (Table 1). Strawberries are planted on soil types ranging from desert sands requiring extensive irrigation to heavy clay soils requiring underdrainage and the addition of organic matter. In subarctic areas, the plants may be exposed to temperatures as low as $-50\,^{\circ}C$ in winter and as high as $40\,^{\circ}C$ in summer. In desert areas, temperatures may be above $50\,^{\circ}C$ in the day and near $0\,^{\circ}C$ at night. Where harsh winter conditions prevail, plants may be protected from extreme temperatures by covering with straw or other organic mulch. Mist irrigation is sometimes used to cool the fruit and leaves under conditions of high summer heat.

The geographical areas with the highest total production and the highest production per hectare are located where there is a Mediterranean-type climate. These areas include California, Spain, southern Italy, and southern France. In these areas strawberries can be produced virtually all year and day-neutral varieties, i.e., those in which fruit bud initiation is independent of day length, are grown more extensively. The June-bearing varieties, which predominate in most other growing areas, have a requirement for a short day (some as little as 8 h).

Table 1 Global distribution of production by geographical area, including a listing of countries producing 5.0×10^7 kg or more

Global distribution by geographical area and countries within any one area producing 5.0×10^7 kg or more
World total
Africa
North and Central America
Mexico
USA
South America
Asia
Japan
Korea, Republic of Russian Federation
Turkey
Europe
France
Germany
Italy
Poland
Spain
Oceania (includes Australia and New Zealand)

From Food and Agriculture Organization of the United Nations (1998) Production Statistics. FAOSTAT, Available at http://apps.Fao.org (February 2000).

Varieties

The commercial strawberry is a heterozygous octoploid with characteristics inherited mainly from two genetically variable wild species. The first crosses between *F. chiloensis* and *F. virginiana* were made in Europe in the early 1760s. Breeding work with other species has not been successful, primarily owing to genetic incompatibility.

Variability in the wild species has been used by breeders who use parents with desired traits when crosses are made. When seeds from a cross are planted, the resultant plant is a clone. Thus, breeders select individual plants with desired characteristics rather than selecting a population of plants.

The general objectives of strawberry-breeding programs are as follows: (1) to enhance yields; (2) to improve fruit quality, which includes factors such as flavor, firmness, keeping quality or other characteristics; (3) to produce plants adapted to specific environmental or cultural conditions; (4) to develop varieties resistant to diseases and insects; and (5) to develop varieties with specific requirements, such as suitability for mechanical harvesting and processing, or extra firmness for long-distance shipping. New strawberry varieties are continuously being introduced, thus worldwide they number in the hundreds. Some of the more important varieties are listed in Table 2.

Commercial Importance

The Food and Agriculture Organization (FAO) lists strawberry production statistics for 64 countries. Seven countries – Spain, Japan, Korea, Turkey, Poland, Mexico, and the USA – account for more than 66% of the world's 2.7×10^9 kg annual production (Table 1). This worldwide production would have a farm-gate value of over $\$5.4 \times 10^9$ (US) based on their value in North America. This return would represent $\$2\ kg^{-1}$, averaged for fresh-market berries.

In countries such as Poland and Mexico, a large portion of the crop is processed and sold for foreign currency. In the European Community (EC), strawberries are an important cash crop as this trading block produces over 7.0×10^8 kg annually. Much of the fresh crop produced in the USA and the EC is shipped to northern areas for fresh-market consumption. Growers harvesting for fresh-market consumption receive a higher price per unit of fruit than those harvesting for processing, but fresh market harvesting costs are higher.

Yields in the state of California are over 40 000 kg ha^{-1} – a level of production two to six times that of other areas in the USA or in other countries. California produces 80% of the US crop, and 65% of the

Table 2 Strawberry varieties grown in quantity in different geographical areas of the world

Geographical area	*Varieties*
California (some of the same varieties in Mexico)	Selva, Chandler, Commander, Oso Grande, Camerosa, Seascape, Lido, Key Largo, E26, Coronado, San Miguel, Catalina
North America (except California and Mexico)	Totem, Bounty, Kent, Earliglow, Honeoye, Governor Simcoe, Glooscap, Selva, Camerosa, Sweet Charlie, Annapolis, Jewel, Cavandish, Delmarvel, Veestar, Redcrest, Hood, Chandler
South America	Camerosa, Chandler, Pajaro, Selva, Oso Grande, Sweet Charlie, Tudla
South Asia	Hokowase, Reiko, Toyonoko, Nyoho, Sachinoka, Tochiotome, Suhong, Rachel, Douglas, Pajaro, Cruz, Senga Sengana, Chandler, Oso Grande, Selva, Sequoia
Western Europe (excluding Spain, Italy, France)	Senga Sengana, Dania, Elsanta, Elvira, Karola, Symphony, Pegasus, Honeoye, Evita, Bolero, Tango, Selva, Zefyr, Bounty
Spain, Italy, France	Pajaro, Camerosa, Chandler, Oso Grande, Tudla, Eris, Tethis, Clea, Miranda, Marmolada, Idea, Addie, Miss, Selva, Seascape, Elsanta, Gariguette, Dar Select, Eros
Eastern Europe/northern Asia	Senga Sengana, Zarya, Gorella, Elsanta, Kokinskaja Rannaja, Festivalnaja Romashka, Zenit, Nadezhda, Rannyaya Plotnaya, Lurck Vira, Istochnik, Desna
Oceania	Chandler, Pajaro, Parker, Selva, Camerosa

Data from Hancock JF (1999) *Strawberries*. Wallingford, UK: CABI Publishing.

North and Central American crop, making it an important cash crop in that state.

In some countries, particularly around large cities, consumers are invited by growers to pick their own fruit. This has become a very important means of harvesting and marketing the strawberry crop. Pick-your-own farm operators do not have problems finding labor to harvest their berries. However, there are other problems, such as in-field consumption, crowd control, incomplete harvesting, and grower–public interaction.

Morphology

The strawberry is an accessory fruit with an aggregate of achenes (the true fruits which contain the seed), attached in an orderly fashion to the epidermis of the receptacle (the floral axis to which the various flower parts are attached). The large, fleshy receptacle (consisting of pith, cortex, and a vascular system) is the succulent, edible portion of the fruit. It expands in response to hormones produced by the ovules after fertilization has occurred.

The fruit of the commercial strawberry has a red-colored exterior and an interior color ranging from white to dark red. Fruit shape depends on such factors as variety, environmental conditions under which it was grown, and planting location. Fruit size, which varies considerably even within cultivars, depends on environmental factors and position of the fruit within a cluster. The primary berry is the largest and has the most achenes. Secondary and tertiary berries become progressively smaller and have fewer achenes. The fruit ranges between 2 and 5 cm in length depending on variety and other factors. Descriptors used to identify the shapes of strawberries include globose, globose conic, conic, long conic, oblate, necked (a long neck), long wedge, and short wedge.

Table 3 Nutrient composition (per 100 g fresh weight) of strawberries compared to that of apples and oranges

Nutrients and units	*Strawberry*	*Apple*	*Orange*
Water (g)	91.57	83.93	86.75
Food energy (kcal)	30	59	47
Food energy (kJ)	127	245	197
Protein (N × 6.25: g)	0.61	0.19	0.94
Total lipid (g)	0.37	0.36	0.12
Cholesterol (mg)	0	0	0
Carbohydrate (g)	7.02	15.25	11.75
Fiber (g)	0.53	0.77	0.43
Calcium (mg)	14	7	40
Magnesium (mg)	10	5	10
Phosphorus (mg)	19	7	14
Potassium (mg)	166	115	181
Ascorbic acid (mg)	56.7	5.7	53.2
Folic Acid (μg)	17.7	2.8	30.3
Vitamin A (μg)	8.1	1.5	6.3

Data from USDA (1982) *Composition of Foods: Fruits and Fruit Juices, Raw, Processed, Prepared*. US Department of Agriculture, Human Nutrition Information Services, Agriculture Hand-book, pp. 8–9. Washington, DC: US Government Printing Office.

Chemical and Nutritional Composition

A list of some of the nutrients in strawberries is presented in Table 3, with those for apples and oranges included for comparison. The strawberry contains the

highest percentage of water and the lowest level of food energy of these fruits. An intake of 100 g of strawberries would provide the US Department of Agriculture adult daily requirement for ascorbic acid (vitamin C). The ascorbic acid content of strawberries is slightly higher than that of oranges, and much higher than that of apples. Although not a rich source of potassium, a 100-g portion would provide about 10% of the suggested minimum estimated safe daily intake. This is more than apples, less than oranges, and about 42% of that provided by bananas.

Strawberries contain several vitamins and minerals (Table 3) in concentrations that would provide less than 5% of the minimum daily requirement. There are also 18 amino acids in strawberries, but all occur in extremely low concentrations. (Refer to individual nutrients.)

Handling and Storage

Strawberries are sold by volume or by weight. The volume containers hold approximately 0.25, 0.5 or 1 l. A small container would hold approximately 225 g; some of the larger containers may hold up to 3 kg. Fruit for the fresh market is generally harvested directly into the retail container. These are placed in fiberboard or corrugated trays, or masters, each holding a convenient multiple of the smaller harvesting containers.

Fresh-market strawberries must be harvested in a manner which eliminates, or keeps to a minimum, bruising, skin disruptions, and the bleeding of fruit juices. Berries should be harvested by pinching off the stem at the cap, as clasping the fruit between the fingers and then pulling causes excessive bruising. Good commercial practice is to hold harvested fruit out of the sun and protected from warm winds and blowing dirt. Growers following recommended practices move their fruit to cold storage within 1 or 2 h of harvest. Warm fruit bruises easily and must be transported gently.

Field heat must be removed rapidly to reduce losses from rot, caused by fungi such as *Botrytis cinerea* (grey mold), *Phytophthora cactorum* (leather rot), and *Rhizopus nigricans* (black whisker rot). These fungi either grow very slowly, or not at all, at fruit temperatures of 0–2 °C. Changes in the fruit associated with senescence, e.g., skin darkening, softening of the flesh, and loss of flavor, occur more slowly at lower temperatures. (*See* **Ripening of Fruit**.)

Forced-air cooling is the preferred method of removing field heat in many strawberry-producing areas. In these coolers, the fruit can be cooled from 25–35 °C to 2–4 °C within 1–2 h. In these coolers, trays of fruit are placed so as to form the side walls of a plenum, open to a wall with an exhaust fan. With the open top and end covered with fabric, chilled air is drawn from the cold room through the trays of fruit. To retain the best quality, the chilled fruit is held at 0–2 °C and 90–95% relative humidity. Fruit treated in this manner can be stored for 5–7 days and retains its quality during 3–5 days of marketing after storage. (*See* **Controlled-atmosphere Storage: Applications for Bulk Storage of Foodstuffs.**) In some growing areas, fruit is forced-air-cooled on pallets and then covered with an airtight plastic bag. Carbon dioxide is added to an approximate level of 15% to help maintain quality during shipping to market. Some varieties of strawberries become 25–40% firmer when they are exposed to 10–20% carbon dioxide; with others, the carbon dioxide serves only to maintain firmness. There are fewer losses from decay among these berries, and their respiration rates are reduced.

Although the vast majority of strawberries are harvested by hand for the fresh market, a small number are mechanically harvested (predominantly in Europe). Mechanically harvested berries are not suitable for the fresh-market trade because of extensive bruising during the harvesting process. However, mechanically harvested fruit is suitable for processing. The berries from the harvester are collected in large totes, or bulk bins, for transporting to the processing facility. This fruit is processed within a few hours of harvest and thus may not be cooled before processing. Recovery of finished product is generally higher when the fruit is cooled before processing. The quality of mechanically harvested berries does not deteriorate excessively for 2–3 days, possibly longer when they are cooled to 0–2 °C immediately after harvesting. Mechanically harvested fruit does contain a moderate quantity of green berries. Some of this green fruit is incorporated into the end product without affecting quality.

Industrial Uses

Although most strawberries are destined for the fresh market, there are certain production areas that concentrate on processed fruit. Inferior fruit from the fresh market, and a small quantity of mechanically harvested fruit with green berries included, are used for making purée, juice, and concentrate. The purée and the juice are preserved by freezing or by adding sulfur dioxide. Hand-harvested fruit is processed as frozen whole pack with or without sugar, individually quick-frozen (IQF), brined in sulfur dioxide, or heat-processed in lacquered cans. Since fruit is needed throughout the year, products for further processing are usually stored in bulk containers. During

processing, calcium salts are often added to the fruit to minimize structural changes. (*See* **Freezing**: Operations.)

Brined fruit, used in the manufacture of some strawberry jams, can be preserved in more than 2000 p.p.m. sulfur dioxide. Much of the sulfur dioxide dissipates during the cooking process. Fast freezing of strawberries is essential for retention of firmness and quality. Fast freezing is either accomplished in an air-blast freezer where jets of cold air blow up through berries on a mesh belt, or the berries are frozen in liquid nitrogen; both methods provide IQF fruit. Berries can also be fast-frozen in small packages in a blast freezer or more slowly in larger containers on pallets in a freezing room. When berries are frozen in a sugar pack, the ratio of sugar to fruit, by weight, can range from 1:3 to 1:27, depending on the secondary manufacturer's specifications. (*See* **Freezing**: Blast and Plate Freezing.)

Strawberries or strawberry-based products are present as the main fruit ingredients or as a mixture with other fruit in many consumer products. Large volumes are used in the production of jams, jellies, icecream and fruit yogurts. Strawberries in moderate quantities are used in spreads, preserves, icecream toppings, strawberry syrups, juices, juice blends, nectars, drinks, milkshakes, and various bakery products, such as pies, pastries, turnovers, doughnuts, and cheesecakes. Lesser quantities of fruit are used in the production of fruit beverage mixes, alcoholic beverages, flakes, glazes, powders, and freeze-dried whole fruit used in cereals and other products.

See also: **Controlled-atmosphere Storage**: Applications for Bulk Storage of Foodstuffs; **Freezing**: Operations; Blast and Plate Freezing; **Ripening of Fruit**

Further Reading

Chandler CK (1991) North American strawberry cultivars. In: Dale A and Luby JJ (eds) *The Strawberry into the 21st Century*, pp. 60–65. Portland, Oregon: Timber Press.

Committee on Dietary Allowances (1989) *Recommended Dietary Allowances*. Washington, DC: National Academy of Sciences.

Darrow GM (1966) *The Strawberry*. New York: Holt, Rinehart and Winston.

Food and Agriculture Organization (1998) *Production Statistics*. *FAOSTAT*. Available at http://apps.fao.org (February 2000).

Galletta GH and Himelrick DG (1990) *Small Fruit Crop Management*. New Jersey: Prentice Hall.

Hancock JF (1999) *Strawberries*. Wallingford, UK: CABI Publishing.

Hancock JF and Scott DH (1988) Strawberry cultivars and worldwide patterns of strawberry production. *Fruit Varieties Journal* 42: 102–108.

Janick J and Moore JN (1996) *Fruit Breeding*, vol. 3. *Vine and Small Fruits*. New York: Wiley.

Rosati P (1991) The strawberry in Europe. In: Dale A and Luby JJ (eds) *The Strawberry into the 21st Century*, pp. 27–35. Portland, Oregon: Timber Press.

USDA (1982) *Composition of Foods: Fruits and Fruit Juices, Raw, Processed, Prepared*. US Department of Agriculture, Human Nutrition Information Services, Agriculture Hand-book, pp. 8–9. Washington, DC: US Government Printing Office.

STRESS AND NUTRITION

A C Brown and C I Waslien, University of Hawaii at Manoa, Honolulu, HI, USA

Introduction

Hans Selye first described the stress phenomenon as the body's response to a wide range of stressors. Others have defined stress as anything interfering with a person's optimal mental and physical health. Sources of stress can be exogenous (external) or endogenous (internal) and include anything that is 'perceived,' psychologically or physiologically, to be a threat to the organism. Examples of physical stressors include trauma, injury, or violence, illness (heart attack, cancer, infection, hemorrhage), surgery, climate (cold, heat), muscular exercise, fasting, weather changes, chemical exposure, drugs, and biochemical changes (postnatal depression, schizophrenia). Some psychological stressors include driving to work, phone calls, having lunch with a client, giving a speech, unstable relationships, work conflicts, purpose-in-life conflicts, and significant life events, as shown in **Table 1**.

Some people are more susceptible to stress by their exaggerated perception of the stressor, lack of necessary skills to handle minimum daily life stressors, and/or lack of an ability to adapt or cope in such a way to restore homeostasis. A person's reaction to stress, whether psychological or physical, orchestrates a neuroendocrine response that can in turn cause

Table 1 Part of the Holmes–Rahe Social Adjustment Scale that assigns values to various stressful events (good or bad) (the higher the score, the greater the risk of a stress-related health condition)

Events	*Value*
Death of a spouse	100
Divorce	73
Marital separation	65
Personal injury or illness	53
Marriage	50
Retirement	45
Pregnancy	40
Change in work responsibilities	29
Outstanding personal achievement	28
Trouble with boss	23
Change in residence or schools	20
Change in working hours or conditions	20
Change in social activity	18
Vacation	13
Christmas	12

Table 2 Partial list of conditions related to stress

Physical	*Psychological*
Gastrointestinal:	*Anxiety disorder*
Indigestion	*Phobic states*
Gastrointestinal reflux disease	*Depression*
Peptic ulcers	*Post traumatic stress disorder*
Ulcerative colitis	*Chronic fatigue syndrome*
Constipation/diarrhea	*Eating disorders*:
Irritable bowl syndrome	Binge eating
Diverticulosis	Bulimia
Cardiovascular:	Anorexia nervosa
Heart disease	*Alcoholism*
Hypertension	*Drug addiction*
Palpitations	*Panic disorders*
Arrhythmias	*Chronic pain*
Cancer	*Family dysfunction*
Diabetes mellitus	*Sleep disorders*
Osteoporosis	
Reproductive:	
Amenorrhea	
Infertility	
Immune function (suppressed):	
Common cold	
Autoimmune diseases	
Rheumatoid arthritis	
Systemic lupus erythematosus	
Neurological:	
Migraine headaches	

more stress by resulting in illness or diminished psychological or physical health (Table 2). The question arises, however, as to why the response to stress becomes so severe and prolonged in some individuals that it develops into an incapacitating clinical disorder. This individual susceptibility may be attributable to factors in the person's genetic background, personality, and environment. (*See* **Atherosclerosis; Colon:** Diseases and Disorders; **Coronary Heart Disease:** Intervention Studies; **Inflammatory Bowel Disease; Migraine and Diet.**)

Stress Reaction

The stress reaction is defined by Hans Selye in his general adaptation syndrome model (Figure 1). Each of the stages in this model – alarm, resistance, and exhaustion – involves the body's neuroendocrine system, which is influenced by psychological and physical stimuli.

Alarm, fight or flight, is the immediate response of the body to 'perceived' stress. Physiologically, this starts at the brain's hypothalamus, which acts as the central computer chip of the body regulating such functions as heart rate, blood pressure, respiration, body temperature, digestion, hunger, thirst, and libido. Surrounding the hypothalamus is the limbic system, one of the oldest parts of the brain, which houses the emotions. Powerful feelings such as fear and rage trigger the hypothalamus, which sends messages to the autonomic nervous system by the hypothalamic–pituitary–adrenal (HPA) axis and separately through the autonomic nervous system. The nerves of the autonomic nervous system are split into either parasympathetic or sympathetic branches, the latter of which causes an immediate stress reaction by releasing catecholamines (adrenaline-like chemicals). These chemicals released from the adrenal gland increase heart rate, respiration, and blood pressure.

The activated hypothalamus also secretes its own hormones and stimulates the pituitary gland to secrete hormones, which produce some of the same effects as those of the catecholamines but which last some 10 times longer and have a far wider reach. One of these hormones, corticotropin-releasing hormone (CRH), is sent to the pituitary to trigger the release of adrenocorticotropin hormone (ACTH). ACTH travels through the bloodstream to the adrenal glands, on top of the kidneys, to produce glucocorticoids (such as cortisol), which start a cascade of events, including increased blood glucose concentrations, elevated blood pressure, and slowed digestion (Figure 2). Specifically, insulin's ability to facilitate glucose uptake by the cells is reduced, while gluconeogenesis, the synthesis of new glucose (from glycerol and amino acids) is increased. Blood pressure rises as the kidneys are signaled to retain more sodium, which raises water volume in the blood vessels. Also, digestion slows as hormonal changes cause muscles to become engorged with oxygen and glucose-rich blood that is shunted away from the digestive tract. The body in this conditional response

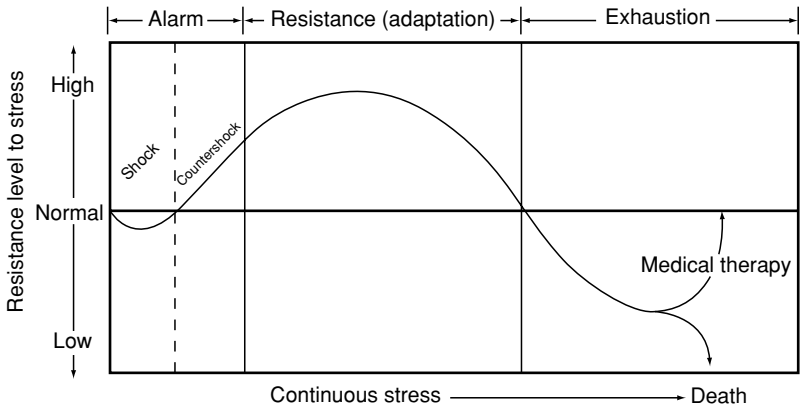

Figure 1 General adaption syndrome, consisting of three stages: (1) alarm, (2) resistance, and (3) exhaustion.

to stressors is ready for fight or flight. (*See* **Hormones**: Pituitary Hormones.)

Resistance (adaptation), the second stage of the stress response, is to achieve optimal adaptation in resisting the stressor. Everyday stressors (eustressors) are beneficial in maintaining the psychophysiological balance that results when the stressor is successfully removed, adapted, or coped with by the person. Stress is actually a necessary component in life because it contributes to survival and, ultimately, growth. Optimal stress fuels maximum performance, but excess stress results when the demands on a person exceed or fall far below their capabilities. Hans Selye said the only time an individual is free from stress is death – the ultimate flight. However, too much stress, and failure to adapt and reach a healthy homeostasis, can also result in illness or death.

Exhaustion, the last stage of the stress response, a continued, chronic response to stress, can be a risk factor for many multifactorial disorders. These in turn may lead to a downward spiral of more stress, exhaustion, and possibly extinction. Eustress becomes distress if not adequately handled by the body and mind (Table 3). Physical and psychological well-being become 'ill,' decreasing the quality of life, if not its very presence.

While the changes produced by stress-initiated hormones are beneficial in the short term, they can be detrimental when prolonged. It is now generally believed that most signs and symptoms of stress-related disease are the result of stress hormones marshaled by the hypothalamus in response to an alarm for which their particular actions are no longer appropriate.

Stress and Nutritional Needs

The hormonal concert of changes occurring in the stress response creates a cascade of biochemical and physiological events that require energy and nutrients. The main physical stressors to the body requiring an increase in energy and nutrients are injuries, surgeries, burns, illnesses, and infections. These conditions inflict greater requirements than emotional stressors whose exact nutrient needs have not yet been clearly documented. However, it is well established that physical stressors do increase requirements for energy and protein. Less information is available for the vitamins, minerals, and electrolytes. The degree to which these physical stressors foster a need for more nutrients is based on the severity/duration of the condition, the person's age and nutritional status, and individual ability to overcome stressors in life. If not corrected, malnutrition during these states can result in an increased recovery time, length of hospitalization time, and/or morbidity and mortality.

Energy

With acute stress, such as surgery, energy requirements rise only about 10%; severe infections raise requirements by 20–40%, bone fractures by 10–30%, and burns by 40–125%. The extra energy requirements caused by psychological stresses such as fear, anger, or rage are difficult to determine but presumably fall within the range error and are not given a special allocation. Acute stress of short duration is usually self-corrected by an increase in appetite after the stress has abated, but special enteral or parental formulae may be needed for treating severe stress. (*See* **Burns Patients – Nutritional Management**.)

Protein

Additional energy intake not only supplies necessary energy, but spares the body's protein reserves – primarily the muscle. If glycogen reserves from the liver

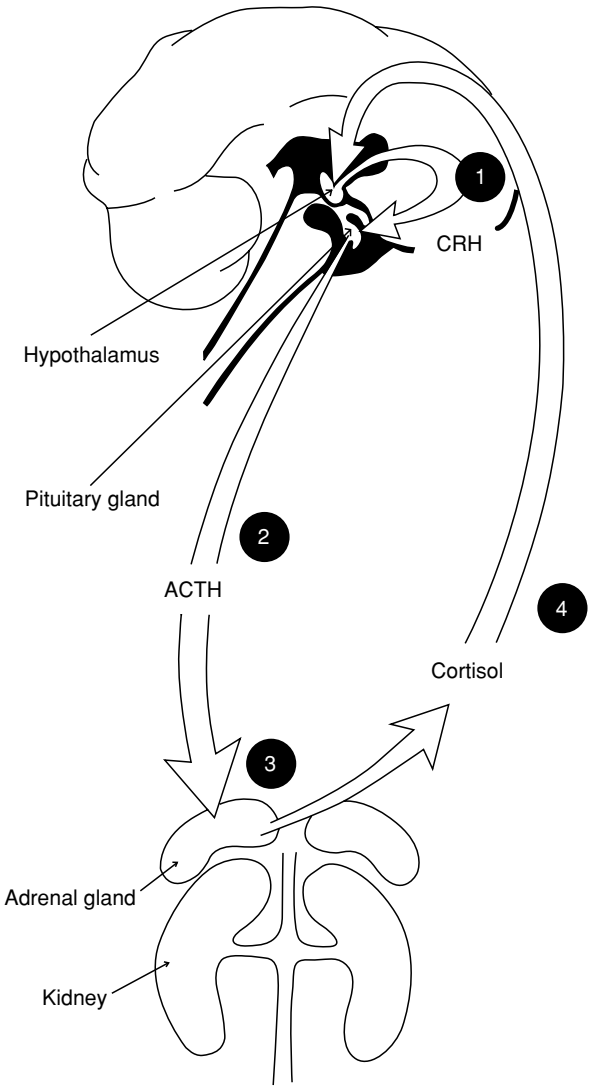

1. Stress triggers cells in the brain (hypothalamus) to release corticotropin-releasing hormone (CRH).
2. CRH signals the brain's pituitary gland to release adrenocorticotropin hormone (ACTH).
3. ACTH travels through the bloodstream to the adrenal glands – just above the kidneys – where they produce glucocorticoids such as cortisol.
4. Excess cortisol levels are prevented by feedback to the hypothalamus that occurs when elevated cortisol levels signal the brain to stop releasing CRH.

Figure 2 Hypothalamic–pituitary–adrenal (HPA) axis that sets in motion a myriad of stress-related psychophysiological events.

and muscles are depleted and additional energy is not provided during the catabolic state of acute stress, the body turns toward its fat and protein resources. Although the loss of fat is usually not a problem initially, it is detrimental to the skeletal muscles if they are broken down to serve as a source of amino acids

Table 3 Factors and feelings associated with eustress and distress

Eustress	*Distress*
Increased mental acuity	Forgetfulness
	Diminished attention to detail
	Poor work performance
Pleasure/happiness	Sadness
	Emotional outbursts
Euphoria	Lethargy, apathy

Modified from Clancy J and McVicar A (1993) *British Journal of Nursing* 2(8): 410–417.

to make glucose (gluconeogenesis). Physical stressors also stimulate glucocorticoid secretion, which induces protein breakdown. If prolonged, this leads to massive tissue wasting. Adequate protein of a high biological value is necessary to avoid a negative nitrogen balance and to prevent the following: diminution in immunocompetence, hypoalbuminemia, decreased wound healing, decreased synthesis of enzymes and plasma proteins, infection, and pressure sores associated with poor protein status.

Moderate stress requires 1.0–1.5 g of protein per kilogram or bodyweight, whereas severe stress requires 1.5–2.0 g protein per kilogram. This is only slightly higher than the 0.8 g of protein per kilogram required in most healthy adults. During severe stress, the body may be unable to convert essential amino acids to nonessential amino acids, so they must be obtained from the diet. Research shows promising results for the use of branched-chain amino acids, glutamine, and arginine during periods of hypermetabolic states. Blood levels of branched-chain amino acids (BCAA), leucine, isoleucine, and valine, are low during stress. Mixed results have been reported with BCAA formulas for improving morbidity or mortality, reducing the length of hospital stay, or maintaining tissue protein reserves of hospital patients. As a precaution, a 35–50% BCAA formula is recommended for highly stressed patients who cannot achieve nitrogen balance within a reasonable amount of time after receiving a standard formula. A similar recommendation exists for glutamine, the principal free amino acid in muscle. Catabolic patients receiving a glutamine-peptide-supplemented solution were in a more positive nitrogen balance than control patients given isonitrogenous, isoenergetic, or conventional total parenteral nutrition amino acid mixtures. Another amino acid, arginine, is beneficial during critical illness. Supplementation during this time with arginine may increase nitrogen retention, accelerate wound healing, and enhance the immune system by increasing lymphocyte response. (*See* **Glucose**: Function and Metabolism; **Glycogen**; **Protein**: Requirements.)

Fatty Acids

ω-3 fatty acids appear to improve survival rates of burn patients, and reduce infections and wound complications of postoperative patients. Perhaps this is related to the fact that certain polyunsaturated fatty acids are used in the biosynthesis of eicosanoid mediators influencing the systemic inflammatory response. ω-3 fatty acids lessen inflammation, whereas ω-6 fatty acids increase the inflammatory response. Sources of ω-3 fatty acids include canola oil, fish oils, and green vegetables, whereas ω-6 fatty acids are derived from corn, safflower, soybean, and sunflower oils. It has been postulated that patients with inflammatory conditions (systemic lupus erythematosus, rheumatoid arthritis, and multiple sclerosis) on hospital tube feedings (enteral and parenteral formulas, intravenous emulsions) would fare better if the lipid source were ω-3 instead of ω-6 fatty acids.

Vitamins, Minerals, and Electrolytes

The adrenal glands have the highest concentrations of vitamin C of any organ in the body, and because this vitamin is needed for the synthesis of catecholamines, massive doses of the vitamin have been advocated to help cope with modern-day stress. However, clinical evidence at this time does not support this supposition. Other nutrients that are lost in the urine with stress, such as zinc, copper, magnesium, and calcium, do not present problems for people with good dietary intakes. Potassium, which is lost along with protein catabolism of muscle and organ tissue, may be an exception but can be replaced with increased consumption of fruits and vegetables.

In severe stress, the requirements for vitamins A, B_1, B_2, B_6, C, D, E, K, nicotinic acid, folate, and pantothenic acid, as well as for the minerals iron, zinc, copper, selenium, and magnesium are probably increased, although actual requirements are unknown. Vitamin C is known to be required for wound healing and collagen formation, and urinary losses are increased during stress; thus, supplements may be prescribed to promote extensive tissue regeneration. Studies have shown that 1000 mg of vitamin C per day appears to speed the healing of bone fractures. Tissue regeneration is also dependent on vitamin A; low levels of plasma retinol have been observed in fracture patients after surgery. Pyridoxine, potassium, magnesium, phosphate requirements, and, to a lesser extent, calcium requirements seem to parallel the level of protein consumption; hence, supplements of these nutrients may be needed when protein and carbohydrate intakes are greatly increased. (Refer to individual nutrients.)

Actually, the immediate major concern following a physical stress is correcting fluid and electrolyte losses. Intravenous infusions may be needed, but even hydration may be restricted if stress leads to renal shutdown. A diet high in protein and energy can be initiated within approximately 5 days. Excess delay in realimentation is also not warranted as it is critical to prevent weight loss that may approach 20–30% of body weight if nutritional support is not promptly implemented. A weight loss of 40–50% is generally fatal.

Stress and Appetite

Stress, depending on the individual, can either increase or decrease appetite. Some people under pressure eat less, whereas others overeat in an effort to relieve stress. According to one study, snacking behavior in 212 college students exposed to stress increased by 73%. However, the total daily food intake decreased in 42% of the students, whereas 38% reported an increase. This study, along with other research, revealed that dieters are much more likely to overeat during stress than nondieters. The increased food consumption in dieters is especially pronounced when the stress source is a threat to the ego or self-image. Results from a study revealed that psychosocial stress induces a large cortisol response in glucose-treated subjects, compared with fasting subjects subjected to the same stressor but without the resulting cortisol peak. The researchers concluded that low glucose levels appear to inhibit the HPA axis response resulting in the secretion of cortisol and its influence on appetite. Another appetite theory is that stress-prone subjects are prone to serotonin deficiency, which may be related to 'carbohydrate craving' and subsequent binging and/or obesity. It is interesting to note that the weight gains of genetically obese rats are abolished by adrenalectomy, while cortisol (corticosterone) treatment restores their diminished food intake to normal levels.

Stress-inducing Foods

Refined Carbohydrates

Refined cereals and sugars have been made scapegoats, responsible for the stress of Western diets. Claims are made that these foods, low in B vitamins, rob the body of vitamins. Some authors go on to advocate up to 100 times the recommended intake of vitamins as compensation. Certainly, when chronic stress increases nutrient requirements, 'empty calories' are inappropriate, but there is no rationale for labeling moderate amounts of refined cereals and

sugars as 'stress producing.' Sugar has come under particular fire because of its supposed role in hypoglycemia. Although it is true that hypoglycemia can affect behavior in a number or ways, e.g., alternating 'highs' and 'lows,' there are two major misunderstandings. Contrary to popular opinion, diagnosable hypoglycemia is quite rare; the Mayo Clinic finds a rate of approximately one case per 2500 persons. Although a person with hypoglycemia should avoid sugar in any form, sugar does not cause the disease. Both the low blood sugar reaction and the hormonal response described as the alarm reaction are symptoms of hypoglycemia. (*See* **Glucose**: Maintenance of Blood Glucose Level.)

Fasting

Researchers report that a good portion of the psychological syndrome of anorexia nervosa is in response to starvation. Emotional instability is a classic sign of the ketosis that accompanies fasting. One does not have to endure long-term starvation to feel less than able to cope with stress. Just skipping a meal or two can have this effect. Steady eating of moderate, balanced meals is likely to be more beneficial to stress-prone individuals than a lifetime supply of various supplements. (*See* **Anorexia Nervosa**.)

Caffeine

Certain commercial advertisements insist that stress can be reduced by drinking decaffeinated coffee or caffeine-free soft drinks. While some research reports indicate that caffeine induces stress, others show that caffeine reduces it. A recent study clarified the situation by demonstrating that persons habituated to the use of caffeine had increased anxiety when deprived of it but were calmed by a dose of caffeine. In contrast, low-caffeine users showed a slight increase in tension when given caffeine. Apparently, the kind of response depends on previous use and individual differences.

Sodium

For many people, excessive fluid retention, and other premenstrual changes in women, may be related to emotional stability and resistance to stress. Stress itself induces sodium retention and water accumulation. Any dietary factors that further contribute to fluid retention, in particular excessive intake of sodium, may therefore increase the stress effect. (*See* **Premenstrual Syndrome : Nutritional Aspects**.)

Stress-relieving Foods

Can specific foods be used to reduce stress? When under stress, most people instinctively turn to certain foods – foods that create a feeling of security. The most common security food is milk, perhaps as a symbol of the more protected days of childhood and a desire to return to them. In fact, it has been speculated that the use of milk to treat ulcers may be successful partly because it stimulates this 'return to the womb' association. Our choice of security foods seems to be based on our childhood experiences, and those that trigger comfort in one person may trigger anxiety in another.

Among women, chocolate is the most craved food. Researchers believe that this may be due to a combination of chocolate's sensory characteristics, nutrient composition, and psychoactive ingredients. It has been suggested that chocolate may be used by some women as a form of self-medication to balance low levels of neurotransmitters involved in mood regulation. The desire for chocolate by women fluctuates with hormonal changes and is higher just before and during their menses. This hormonal link in turn might be related to endogenous opiates (endorphins) that influence appetite. Women tend to eat when they feel emotionally low, and the affinity to chocolate may be due to its 'drug-like' constituents that remain elusive.

Dietary Supplements

Alternative/complementary treatments are used by 20–30% of the general North American population. Among these are plants that have been used medicinally throughout history for thousands of years. About 25% of pharmaceutical drugs used today are derived from botanical sources. Among psychotropic herbs, two of the earliest were the opium poppy and the coca plant, producing two of the most notorious street drugs, heroin and cocaine. Drugs are often taken to treat the symptoms rather than the cause, and many people opt to select dietary supplements rather than prescriptions to reduce the side-effects of stress such as anxiety, depression, and insomnia. A few of the more popular, but not necessarily proven effective, psychotropic herbs are discussed below.

California Poppy (*Eschscholtzia californica*)

This flowering plant has a long history of treating insomnia. It does not contain the narcotic morphine found in its cousin, the opium poppy.

Chamomile (*Matricaria recutita*)

This flower, usually in the form of a tea, has been used medicinally for thousands of years. Its primary use is for anxiety and mild insomnia as well as an antispasmodic and antiinflammatory agent, especially for ulcers. The active ingredient is attributed to chrysin, a flavonoid component. Potential rare side-effects include allergies and contact dermatitis.

Kava (*Piper methysticum*)

The roots of this plant have long been used by people in the South Pacific, where it is grown, for ceremonial purposes in much the same way that some cultures use wine. Kava differs from wine in that it is nonaddictive. Its sedative-like quality may be due to kavalactones (kavapyrones), which appear to act as monoamine oxidase inhibitors. Side-effects include dizziness, a drunk-like state, and gastrointestinal upset. Excessive use is associated with Kava dermopathy (scaly, shiny skin), seizures, impaired vision and hearing, yellow discoloration of skin/nails, and liver dysfunction. Reports linking Kava to hepatitis have resulted in some European countries banning this supplement. Kava has also been known to potentiate the effects of alcohol, barbiturates, and Xanax.

Hops (*Humulus lupulus L.*)

Although hops are traditionally used by the brewing industry, the flowers of this plant have long been utilized as a mild sedative. Side-effects include allergy and disrupted menstrual cycles.

Passion Flower (*Passiflora incarnata L.*)

Native to the Americas, its vine leaves are used as a sedative, which is supported by some animal studies.

Skullcap (*Scutellaria laterifolia*)

Rooted in traditional Chinese medicine, this herb has a long history of being used as a sedative and anticonvulsant. Scutellarin is the plant flavonoid glycoside thought to be responsible for the sedative effects of skullcap. Reported adverse reactions include giddiness, confusion, sedation, and seizures. The most serious possible side-effect is hepatitis, which was reported in a some people taking herbal stress remedies that contained skullcap and valerian. It is important to read the label of such products to determine whether skullcap is an ingredient.

St. John's Wort (*Hypericum perforatum*)

Long-term stress is occasionally a risk-factor for mild depression, which has been reported in some cases to respond to St. John's Wort. An analysis of 23 controlled studies concluded that it is significantly more effective than a placebo in treating mild to moderate depression. The effective dose in clinical trials is 300 mg three times a day from an extract standardized at 0.3% hypericin. This active ingredient in St. John's Wort appears to act as a weak monoamine oxidase inhibitor. The main side-effect associated with this herbal antidepressant is photosensitivity to the sun, resulting in rashes. Other less commonly reported side-effects include gastrointestinal upset, dizziness, dry mouth, sedation, and constipation.

Valerian (*Valeriana officialis*)

The root of the valerian plant has been used as a calming and sleep-promoter for centuries. German health officials have approved it as a mild sedative and sleep-aid. Two double-blind, placebo-controlled cross-over studies (400–450 mg before bedtime) resulted in significantly improved sleep quality. There was also no reported residual sedation in the morning or addiction associated with valerian use. Valerian appears to work by inhibiting the enzyme-induced breakdown of γ-aminobutyric acid (GABA) in the brain. Symptoms are rare, but may include headaches, palpitations, insomnia, and possibly hepatitis.

Amino Acids

High intakes of tryptophan – the amino acid precursor of the neurotransmitter serotonin – have been popularized as a treatment for stress. Although taken off the market because of a contaminant that elicited neurological side-effects, tryptophan still has many proponents. A high-carbohydrate meal can have some of the same benefits by causing insulin secretion and favoring the uptake of amino acids other than tryptophan into muscle. This increases the concentration of tryptophan present in the blood and increases its uptake by the brain and its subsequent conversion to serotonin. Supplementation with tyrosine, the precursor for the catecholamine neurotransmitters, has also been shown to reduce psychological and physiological effects of stress. Tyrosine has also been recently demonstrated to ameliorate stress-induced increases in ventricular vulnerability to arrhythmias in conscious animals.

Limited research suggests that subclinical deficiencies of other nutrients involved in the synthesis of neurotransmitters (e.g., nicotinic acid, riboflavin, folic acid, vitamin C, and copper) may contribute to the ability to withstand stress, and that others (e.g., ω-3 fatty acids) may alter the stress response, making it less hazardous for the arteries. In most cases, the alteration in the level of these nutrients needed to make a detectable difference in stress response is of such a magnitude as to preclude normal diets. However, differences in an individual's stress-related illness may in part reflect lifelong differences in intakes of some of these nutrients.

Melatonin

This hormone secreted by the pineal gland is used by some people as a sleep-aid taken 30 min to 2 h before bedtime. Melatonin acts as a mild sedative, but its affect on insomnia remains controversial. It is known that melatonin released nightly by the pineal gland to signal the body to sleep does subside as people age.

However, supplements often contain 10 times the amount secreted by the body (3 mg or more), and it is recommended that less than one-third of that be taken. Melatonin is also not appropriate in children younger than 12. Side-effects include fatigue, headache, confusion, and depression.

Dietary supplements are categorized as neither a food nor a drug, so they are exempt from FDA regulation. As a result, no proof of efficacy or warnings about side-effects (known or testing for the unknown) are required. There are also currently no standardization or quality assurance methods for herbal formulations. Even if there were, it would be very difficult to identify the active ingredient(s) of a herbal product for the often more than one ailment it is claimed to benefit. Added to this is the possibility of contamination, mislabeling, and misidentification of herbs. Assuming that herbs are 'natural,' and therefore safe, can be a wrong assumption. The long-term hepatotoxic effect and other side-effects of many herbs are unknown and may not be elucidated for a long time. As with many compounds, herbal supplementation is not recommended during pregnancy or lactation.

Summary

A moderate, well-balanced diet with emphasis on the stress-related nutrients and regular but not excessively large meals is probably one approach that can be recommended. There is no evidence that any one special dietary preparation or plan is more effective than any other in coping with stress, but specific recommendations concerning energy and nutrients exist in the literature that may benefit certain conditions.

See also: **Anorexia Nervosa**; **Atherosclerosis**; **Burns Patients – Nutritional Management**; **Colon**: Diseases and Disorders; **Coronary Heart Disease**: Intervention Studies; **Glucose**: Function and Metabolism; Maintenance of Blood Glucose Level; **Glycogen**; **Hormones**: Pituitary Hormones; **Inflammatory Bowel Disease**; **Migraine and Diet**; **Premenstrual Syndrome: Nutritional Aspects**; **Protein**: Requirements

Further Reading

Benton D and Donohoe RT (1999) The effects of nutrients on mood. *Public Health Nutrition* 2(3A): 403–409.

Borins M (1998) The dangers of using herbs. What your patients need to know. *Postgraduate Medicine* 104(1): 91–95; 99–100.

Bruch H (1976) The treatment of eating disorders. *Mayo Clinic Disorders* 51: 266–272.

Bruinsma K and Taren DL (1999) Chocolate: food or drug? *Journal of the American Dietetic Association* 99(10): 1249–1256.

Calderon R, Jr. Schneider RH, Alexander CN *et al.* (1999) Stress, stress reduction and hypercholesterolemia in African Americans: a review. *Ethn Dis* 9(3): 451–462.

Clancy J and McVicar A (1993) Subjectivity of Stress. *British Journal of Nursing* 2(8): 410–417.

Cohen S, Tyrrell D and Smith AP (1991) Psychological stress and susceptibility to the common cold. *New England Journal of Medicine* 325(9): 606–612.

DerMarderosian A (ed.) (1999) *The Review of Natural Products*. Facts and Comparisons, 2003.

Ernst E (1999) Harmless herbs? A review of the recent literature. *American Journal of Medicine* 104: 170–178.

Fitzsimmons L and Hadley SA (1994) Nutritional management of the metabolically stressed patient. *Critical Care Nurse Quarterly* 17(3): 1–13.

Houghton PJ (1999) The scientific basis for the reputed activity of Valerian. *Journal of Pharmacy and Pharmacology* 51(5): 505–512.

Kelly GS (1999) Nutritional and botanical interventions to assist with the adaptation to stress. *Alternative Medicine Review* 4(4): 249–265.

Kirschbaum C, Gonzalez B, Rohledes N *et al.* (1997) Effects of fasting glucose load on free cortisol responses to stress and nicotine. *Journal of Clinical Endocrinology and Metabolism* 82(4): 1101–1105.

Markus CR, Panhuysen G, Tuiten A *et al.* (1998) Does carbohydrate-rich, protein-poor food prevent a deterioration of mood and cognitive performance of stress-prone subjects when subjected to a stressful task? *Appetite* 31(1): 49–65.

O'Hara M, Keifer D, Farrell K *et al.* (1998) A review of 12 commonly used medicinal herbs. *Archives of Family Medicine* 7(6): 523–536.

Oliver G and Wardle J (1999) Perceived effects of stress on food choice. *Physiology and Behavior* 66(3): 511–515.

Polivy J, Herman C and McFarlane T (1994) Effects of anxiety on eating: does palatability moderate distress-induced overeating in dieters? *Journal of Abnormal Psychology* 103(3): 505–510.

Rajput V and Bromley SM (1999) Chronic insomnia: a practical review. *American Family Physician* 60(5): 1431–1438; 1441–1442.

Rutledge T and Linden W (1998) To eat or not to eat: affective and physiological mechanisms in the stress-eating relationship. *Journal of Behavioral Medicine* 21(3): 221–240.

Schmicker R (1995) Nutritional treatment of hemodialysis and peritoneal dialysis patients. *Artificial Organs* 19(8): 837–841.

Selye H (1936) A syndrome produced by diverse nocuous agents. *Nature* 4 July.

Sinclair S (1999) Migraine headaches: nutritional, botanical and other alternative approaches. *Alternative Medicine Review* 4(2): 86–95.

Tapp WN and Natelson BH (1988) Consequences of stress: a multiplicative function of health status. *Journal of*

the American Society of Experimental Biology 2: 2268–2271.

Tinsley JA (1999) The hazards of psychotropic herbs. *Minnesota Medicine* 82(5): 29–31.

Van Meyenfeldt MF, Soeters PB, Vente JP *et al.* (1990) Effect of branched chain amino acid enrichment of total parenteral nutrition on nitrogen sparing and clinical outcome of sepsis and trauma: a prospective randomized double blind trial. *British Journal of Surgery* 77(8): 924–929.

White BC, Lincoln CA, Pearce WW, Reef R and Vaida C (1990) Anxiety and muscle tension as consequences of caffeine withdrawal. *Science* 209: 1547–1548.

Wong AH, Smith M and Boon HS (1998) Herbal remedies in psychiatric practice. *Archives of General Psychiatry* 55(11): 1033–1044.

SUCROSE

Contents

Properties and Determination

J Huberlant, Waremme, Belgium

Occurrence in Foods

Among the sugars found in nature, sucrose (also named saccharose, beet sugar, or cane sugar) plays a dominant role in food processing and agriculture, as well as in international economics and politics.

It is naturally present in many plants, in varying quantities. In some plants, it serves as a storage product, although, in other plants, the storage product comprises different sugars such as fructose (fruit), or complex carbohydrates such as starch (corn, potatoes) or inulin (Jerusalem artichoke, dahlia, chicory). (*See* **Carbohydrates**: Classification and Properties.)

There are many plants from which sucrose is commercially extracted, including the date palm, sorghum (*Sorghum vulgare*), sugar maple (*Acer saccharum*), carob or locust bean, and of course, the two most important worldwide sources: sugar cane (*Saccharum officinarum*) and sugar beet (*Beta vulgaris*). (*See* **Date Palms**; **Sorghum**; **Sugar**: Sugarcane.)

Nowadays, because of the improvements in selecting varieties, sugar cane may have a sucrose content of 17–20% and produce as much as 10–12 tonnes of sucrose per hectare, and sometimes even more, depending on numerous factors such as soil, climate, irrigation, etc. The sugar content of the sugar beet varies normally between 14 and 18% with a mean yield of more than 8 tonnes of sucrose per hectare in good conditions (western Europe).

The use of sucrose is extensive, both for homemade preparations and for processed foods, including chocolate and sugar confectionery, baked goods, soft drinks, canned, frozen and dairy products, various preserves (jams, jellies), beers and wines, and so on.

The apparent total sugar consumption based on 'disappearance data' (i.e., the total supply to food industries, wholesalers, and retailers, without considering losses, wastes, or other uses) varies considerably from country to country, according to the standard of living, the climate and dietary habits, and may amount to more than 40 kg per capita per annum (high-income countries) to less than 8 kg (developing countries).

Chemical and Physical Properties

Chemical Properties

Sucrose is a carbohydrate of the general formula $C_{12}H_{22}O_{11}$ and has a molecular weight of 342.30. Its systematic chemical name is α-D-glucopyranosyl-β-D-fructofuranoside. Sucrose is a disaccharide and is easily split by hydrolysis into the two monosaccharides (hexoses) glucose and fructose (eqn (1)). It is a nonreducing sugar, in contrast to the two sugars produced by the inversion reaction. (*See* **Fructose**.)

$$C_{12}H_{22}O_{11} + H_2O \rightarrow 2C_6H_{12}O_6. \qquad (1)$$

In solution, it is present mainly in the cyclic form, which may be represented in various ways, e.g., the Fisher projection (**Figure 1a**) or as a Haworth-type structure (**Figure 1b**).

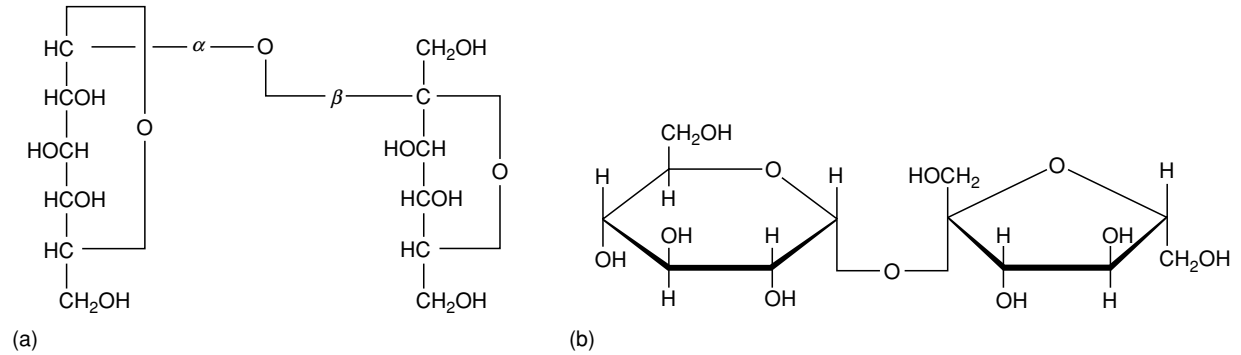

Figure 1 Structural representations of sucrose: (a) Fisher projection; (b) Haworth-type structure. Reproduced from Sucrose: Properties and Determination, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Physical Properties

Crystalline sucrose Sucrose in a pure state crystallizes in a monoclinic system, forming transparent, colorless, and odorless sphenoidal prismatic crystals, the form of which may be strongly affected by impurities (needle-like shapes).

Sucrose has a characteristic sweet taste without any other flavor or aftertaste; it is the standard reference for sweetness. If sucrose is given a sweetness rating of 100, the relative sweetness of fructose is between 105 and 125 and that of glucose between 65 and 75, depending on the conditions (acidity, pH, temperature, etc.).

The crystal density (D_4^{15}) is 1.5879, but the apparent specific gravity of high-grade white sugar varies between 0.72 and 0.88 kg dm^{-3}, depending on the crystal size and humidity, the storage duration, and the thickness of the layer in the storage silo.

Sucrose decomposes at about 186–188 °C to form brown products (caramel) and, finally, chars, but impurities and the products of thermal decomposition significantly lower the temperature of decomposition.

The heat of combustion is 1351.3 kcal mol^{-1} or 3.95 kcal g^{-1}. Sucrose is thus less caloric than fat (9.3 kcal g^{-1}) and protein (4.1 kcal g^{-1}).

Sucrose solutions Sucrose is readily soluble in water, and its solubility increases with temperature. By cooling or evaporating a saturated sucrose solution, a metastable supersaturated solution is obtained. Sucrose is sparingly soluble in ethanol and practically insoluble in ether.

The solubility of sucrose has been calculated at different temperatures. In 1978, the International Commission for Uniform Methods of Sugar Analysis (ICUMSA) officially adopted the tables of Vavrinecz and of Charles. Table 1 gives example solubilities at various temperatures, which have been taken from the Charles tables.

Table 1 Solubility of sucrose at four example temperatures according to the ICUMSA

T (°C)	*Solubility of sucrose (g) in*	
	1 g of water	*100 g of solution*
10	1.884	65.32
20	1.994	66.60
50	2.576	72.04
90	4.262	81.00

Sucrose in aqueous solution has the property of rotating the plane of linear polarized light, and the angle of rotation is proportional to the concentration and to the path length of the solution traversed by the light beam. This angle is measured by means of polarimeters. In the sugar industry, special polarimeters, called 'saccharimeters,' are directly calibrated in 'degrees sugar' (°Z). The ICUMSA adopted a new definition of the international sugar scale as from 1 July 1988. The basis of the 100 °Z point of the international sugar scale is the optical rotation of a 'normal sugar solution' at the wavelength (λ) of the green line of the mercury isotope ^{198}Hg ($\lambda = 546.2271$ nm *in vacuo*) measured at 20.000 °C in a 200.000-mm polarimeter tube. The 'normal sugar solution' corresponds to a concentration of 26.000 g sucrose weighed with brass weights in air under normal conditions (1013 hPa pressure, 20 °C, 50% relative humidity) in 100.000 cm^3 of solution at 20 °C. On this new international sugar scale, the optical rotation of the 100 °Z point is 40.777 ± 0.001 °.

The density of sucrose solutions is a function of the mass concentration and of the temperature. The density values universally used are those of Plato (first published in 1900) giving D_4^{20} for the solutions

between 0 and 95 degrees Brix (grams per 100 g of solution).

The refractive index of sucrose solutions is a function of the amount of dissolved material and also of the temperature. The ICUMSA has published refractive index tables for sucrose solutions from 0 to 85 degrees Brix at 20 and 27 °C, as well as temperature correction tables.

There are many other properties that have been measured, such as viscosity, osmotic pressure, specific heat, boiling point elevation, and equilibrium relative humidity, the knowledge of which is very useful in food technology.

Structural Properties

In addition to its use in food processing as a sweetener and an energy source, sucrose has other functions. Thus, it is not sufficient to replace sucrose by artificial noncaloric sweeteners and expect simply to obtain similar products with the same sweetness but without the calories. (*See* **Sweeteners**: Intensive.)

Sucrose acts as a preservative by its ability to reduce water activity and to increase osmotic pressure to a level where the growth of even the most osmophilic microorganisms is no longer possible (as in jams, jellies, fruit syrups, concentrated milk).

Sucrose acts as a bulking agent and a texturizer in confectionery, baked goods, and soft drinks. Owing to its high solubility and viscosity, it gives body and mouth-feel to various preparations. It influences the pore size distribution, softness, and structure of pastry.

Sucrose also has properties of humectancy, i.e., it assists the ability of products to withstand changes in moisture content, and so extends the shelf-life of goods such as cakes.

Sucrose is an antioxidant, preventing the oxidation of flavors in fruit preserves. (*See* **Antioxidants**: Natural Antioxidants.)

Sucrose can fulfill the role of a diluent or carrier for flavoring or coloring additives and, at the same time, act as enhancer for natural food flavors, balancing sweetness, sourness, and bitterness. Sucrose alone can also give a caramel flavor and a browning of the products in baking, as a result of partial degradation.

Inversion on Storage and Processing

Sucrose, in a solid crystallized form, is one of the most chemically pure food products available in the world with a purity of $99.97 \pm 0.05\%$. It can be stored unchanged almost indefinitely under ideal conditions. To avoid caking or, conversely, attraction of moisture or even liquefaction in the silo, the sugar produced, prior to storage, must be dry and cold and must have undergone a conditioning process.

Freshly produced crystals are surrounded by a thin film of supersaturated mother liquor and possibly contain syrup inclusions. It takes a few days to achieve crystallization and to release free moisture, and, if this water is not allowed to escape, serious storage problems can occur, such as chemical or microbiological inversion, and, with the lowering of the pH, an accelerated degradation with formation of colored products. It is thus useful to store sugar initially in a holding bin aerated with cold dry air before storage in the final silo.

Fermentation of Sucrose

Sucrose may be readily fermented by several microorganisms and is a good starting material for the microbial production of many chemical products such as ethanol, butanol, glycerol, citric acid, levulinic acid, dextran, and many others. Molasses, a byproduct of the sugar industry that no longer contains crystallizable sucrose, is often used as a cheap source for these processes.

The most important fermentation process known from time immemorial is probably alcoholic fermentation by yeasts used to produce beverages such as beer and wine. The first stage of alcoholic fermentation is an enzymatic hydrolysis of sucrose to glucose and fructose by invertase.

Analysis, Extraction, and Isolation

The term 'sucrose analysis' can include numerous determinations depending on the purpose of the analysis.

Sucrose in the Solid State (Commercial Sugars)

The purity of sucrose can be determined by polarimetry or, indirectly, by determining the amount of nonsucrose material present, i.e., water content and other compounds (organic or inorganic).

The water content can be determined by direct methods, such as the Karl–Fischer titration or the conventional 'loss on drying' method.

Inorganic compounds can be evaluated gravimetrically by incineration of the sugar. The result is the sum of the water-soluble and water-insoluble residues or 'ash' in the form of 'carbonate ash' or 'sulfated ash,' according to the method chosen. Ionized soluble salts ('conductivity ash') can be determined conductimetrically. This method is easier and less time-consuming than incineration, but does not have precisely the same significance as 'gravimetric ash.'

Among the so-called organic nonsugars, there may be small amounts of 'invert sugar,' a mixture of reducing sugars. There are numerous methods to

determine reducing sugars in the presence of sucrose, based on the reduction of a copper(II) complex with tartaric acid in alkaline solutions, e.g., Lane and Eynon titration, in addition to the Emmerich method for the determination of traces of reducing sugars in pure sucrose.

It is also possible, when required, to determine individual ions such as copper, lead, iron, calcium, chloride, arsenic, sulfite, and organic nonsugars such as aminonitrogen, betaine, lactic acid, citric acid, and so on.

In addition to the above purity evaluations, buyers require other quality criteria according to the specific uses intended for the sucrose. These may include particle size and uniformity, visual appearance, color and turbidity in aqueous solution or in alcoholic solution, insoluble matter or the presence of visible extraneous impurities, filterability, foaming properties, and floc formation in soft drinks. Some industries have special requirements concerning microbiological quality. For instance, the canning industry has a procedure for determining spore-forming thermophilic bacteria, and in the soft drinks industry, procedures are carried out for the detection of mesophilic bacteria, yeasts, and molds.

Sucrose Solutions and Technical Sugar Juices

Polarimetry may be used to determine the sucrose content of pure sucrose solutions and of intermediate liquid sugar products. In the latter case, it will often be necessary to eliminate many impurities to obtain a clear solution to allow passage of light through the polarimeter tube. This clarification is normally achieved with basic lead acetate solution, but the present trend is to replace this toxic reagent with nontoxic clarifying agents such as aluminum salts.

It is also possible to measure the densities or refractive indices of pure sucrose solutions to determine the sucrose content, since these values are functions of the sucrose concentration. For impure sugar juices, the influence of the soluble nonsucrose compounds on these parameters is almost the same as that of the sucrose. Consequently, the measurement represents the total soluble dry matter of the solution and allows the determination of the degree Brix (or degree Balling), i.e., the percentage of dry matter in the solution. The ratio of sucrose content (Pol) to total soluble dry matter (Brix) is called the quotient of purity (Q):

$$Q = \frac{\text{Pol}}{\text{Brix}} \times 100 \qquad (2)$$

The density can be measured by pycnometers or hydrometers, possibly calibrated in degrees Brix (formerly in degrees Baumé). However, these densitometric methods require many precautions and are time-consuming. It is much easier to measure the refractive index, but because the influence of the nonsugars is not the same, a distinction must be made between densitometric Brix and refractometric Brix, giving a slightly different purity quotient. (*See* **Rheological Properties of Food Materials**.)

Sugar-containing Products (Sweet Foods, Molassed Feeds)

Considering the high solubility of sucrose in water, the first step in analysis will often be an aqueous extraction followed possibly by clarification (or sometimes an aqueous alcoholic extraction). As glucose or invert sugar will also be frequently present in this extract, a double polarization or a double reducing sugars determination should be made, one before and one after inversion of the sucrose present in the extract. By calculation, the total sugars and sucrose content of the product can be determined.

In the sugar industry, the sugar content of beet or cane has to be determined, as this is the basis of payment for the raw material and allows calculation of the amount of sugar entering the factory. For this purpose, there are conventional methods specific to the sugar industry.

Traces of Sugars

It can be important to detect traces of sugar in some products, such as the presence of sugar in the feed water of boilers as a result of the use of condensed water from the sugar juice evaporation, and to control the losses of sugar in the waste water. This is carried out by qualitative or semiquantitative methods (α-naphthol with sulfuric acid) or by automatic apparatus based on chemical reactions or physical properties (conductivity or fluorescence).

Methods of Analysis

Polarization This method is based on measurement of the rotation of polarized light of a solution containing the normal weight (i.e., 26 g in 100 cm^3) of a sugar sample or of the unknown sugar solution to be examined (after clarification if necessary) by means of a saccharimeter calibrated to give 100 °Z with pure sucrose. Calibration of the saccharimeter is carried out with standard quartz plates. The result is given as the percentage of sucrose in the sugar sample or of the unknown sugar solution to be analyzed.

Loss on drying Instead of speaking of water content in this indirect method, it is more correct to speak of 'loss on drying.' The principle of this method is the determination of the loss in weight of a sugar sample placed in an oven at 105 °C for 3 h and cooled

afterwards in a desiccator under uniform conditions. The loss in weight for a high-grade white sugar is normally between 0.02 and 0.05%.

Conductivity ash The specific conductivity of a solution containing 28 g of sugar per 100 g is measured at 20 °C. The water used will have a conductivity of less than 2 $\mu S\,cm^{-1}$. The concentration of 28 degrees Brix has been chosen because the conductivity curve shows a flattened peak in this concentration region. The measured conductivity corrected for the water used is converted by a conventional factor into the percentage of conductivity ash in the sugar.

Visual appearance The visual grade of crystalline white sugar is assessed by comparing the sample with standard color types numbered from 0 to 6, made up of sugar artificially colored and of a defined particle size. The comparison is made by eye under well-defined conditions. Some instruments are now on the market that are intended to replace the eye and so yield a better reproducibility. (*See* **Sensory Evaluation**: Appearance.)

Color and turbidity in solution A solution of 50 g of sugar in 50 g of a buffer solution at pH 7 is filtered through a 0.45-μm membrane filter, and the absorbancy of the deaerated filtered solution is measured by means of a spectrophotometer at a wavelength of 420 nm. The final result is expressed in ICUMSA color units as a percentage of the dry sample.

Turbidity is measured by means of a turbidimeter calibrated with a formazin turbidity standard. The result is expressed in nephelometric turbidity units.

Floc test The soft-drink industry uses various tests to check the appearance of a floc during the shelf-life of a bottled beverage. This floc formation, in acid conditions, may be due to the presence of polysaccharides (gums).

Sulfur dioxide The level of sulfur dioxide in sugar is often limited by national legislation. A method of determining low levels of sulfur dioxide in white sugar is based on a colorimetric determination of the sulfite/rosaniline complex, as measured by spectrophotometry at a wavelength of 560 nm after reaction with formaldehyde.

Particle size The particle size of crystalline sugar can be characterized by the mean aperture (MA) and coefficient of variation (CV). The MA is the theoretical mesh aperture, in millimeters, of a sieve that retains (and also lets through) 50% of the sugar. The CV is the standard deviation expressed as a percentage of the MA. It is an evaluation of the dispersion of the crystal size around the MA.

Chromatographic methods High-performance liquid chromatography (HPLC) and gas–liquid chromatography (GLC) techniques for the analysis of commercial sugar products have been developed in recent years. (*See* **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography.)

Sucrose in beet and cane molasses can be determined by GLC. Molasses samples together with trehalose as an internal standard are silylated using a trimethylsilyl (TMS) reagent to convert the sugars into a form in which they are volatile. The TMS ethers of sucrose and trehalose are separated from each other and from other volatile components using a low- to medium-polarity column. Detection is carried out using a flame ionization detector, and accurate estimation of the sucrose content can be achieved by measurement of the peak areas with the use of an electronic integration system. Sucrose, glucose, and fructose in cane molasses can be determined by GLC according to the method described above by introducing an oxime formation step prior to the silylation.

Sucrose in beet and cane molasses can also be determined by HPLC. The separation column is a cation-exchange resin converted to a metal ion form. The diluted sample is filtered prior to injection on the column. The hydroxyl groups of the different sugars interact with the cations to varying extents, resulting in different elution times. Detection is achieved by differential refractometry of the column eluant. The peak area or height for sucrose is obtained by an electronic integration system, which is compared to that obtained for standards.

White Sugar or Refined Sugar

Irrespective of whether a sugar is a product of the 'first strike' or has actually been refined by remelt, decoloration, and recrystallization, the EC has defined a number of criteria to which a sugar must correspond, to be called 'white sugar' or 'refined sugar.' The criteria are based on the results of some of the analyses described above – polarization, loss on drying, conductivity ash, visual appearance, color in solution, invert sugar, and sulfur dioxide content.

See also: **Antioxidants**: Natural Antioxidants; **Carbohydrates**: Classification and Properties; **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **Date Palms**; **Fructose**; **Rheological Properties of Food Materials**; **Sensory Evaluation**: Appearance; **Sorghum**; **Sugar**: Sugarcane; **Sweeteners**: Intensive

Further Reading

Birch GG and Parker KJ (1982) *Nutritive Sweeteners*. London: Applied Science.

Clarke MA and Godschall MA (1988) Chemistry and processing of sugarbeet and sugarcane. In: *Sugar Series*, vol. 9. Amsterdam: Elsevier.

International Commission for Uniform Methods of Sugar Analysis (ICUMSA) (1978) *Report of the Proceedings of the 17th Session, Montreal*.

International Commission for Uniform Methods of Sugar Analysis (ICUMSA) (1982) *Report of the Proceedings of the 18th Session, Dublin*.

International Commission for Uniform Methods of Sugar Analysis (ICUMSA) (1986) *Report of the Proceedings of the 19th Session, Cannes*.

International Commission for Uniform Methods of Sugar Analysis (ICUMSA) (1990) *Report of the Proceedings of the 20th Session, Colorado Springs*.

Mauch W (1971) The chemical properties of sucrose. *Sugar Technology Reviews* 1(3): 239–290.

Mauch W and Farhoudi E (1980) Quality factors in commercial white granulated sugar. *Sugar Technology Reviews* 7(2): 87–171.

Norrish RS (1967) *Selected Tables of Physical Properties of Sugar Solutions. Scientific and Technical Surveys, No. 51*. Leatherhead: The British Food Manufacturing Industries Research Association.

Schneider F (1979) *Sugar Analysis, ICUMSA Methods*. Peterborough: International Commission for Uniform Methods of Sugar Analysis.

Dietary Importance

G H Anderson, University of Toronto, Toronto, Ontario, Canada

Introduction

For centuries, sucrose has played an important role both in nutrition and in making foods palatable. Sucrose, in many natural and refined forms, has been widely used in the diet since its first recorded use in India in 500 BC, when a rough sugar was produced from boiled cane. This product was called gur, meaning 'nice, sweet, sticky stuff.' Sugar began to be used in the Middle East in 600–700 AD and reached widespread usage in Europe by 1000 AD. From the eighteenth century onward, relatively inexpensive brown sugar was available for the average family. The Industrial Revolution and large-scale food processing brought refined white sugar to everyone's table, and consumption increased rapidly. Since the mid-1920s sugar consumption has remained relatively constant, except during the Second World War, when rationing led to a decrease in consumption.

Sucrose consumption has declined since the 1970s when nutritive sweeteners from corn became available. Corn sweeteners are available as corn syrups, other saccharides, or as high-fructose corn syrup, which is produced by enzymatic conversion of some of the dextrose to fructose. To avoid confusion in this presentation, 'sucrose' will be used when sugar from cane and beet sugar is being discussed and the word 'sugars' will be used when all added free sugars, including corn sweeteners, are discussed.

The facts about sugars today are often distorted, producing incorrect public perceptions; therefore a discussion of the metabolism, health implications, and current consumption levels of dietary sugars follows.

Digestion and Metabolism of Sucrose

Sucrose is hydrolyzed by the enzyme sucrase, an α-glucosidase in the human small intestine, to its component monosaccharides fructose and glucose. About 10–25% of the fructose is converted to glucose in the brush border of the upper gastrointestinal tract. The monosaccharides are absorbed and transported to the liver via the portal vein and subsequently transported to all tissues.

Within the liver fructose may be converted to glucose; it may enter the glycolysis pathway, or provide acetate units for the synthesis of fatty acids. Glucose can be stored in the liver as the glucose polymer glycogen; it can be used in glycolysis to produce energy; or it can provide acetate units for the synthesis of fatty acids.

Glucose is made available to other body tissues via the blood stream and is the primary form of carbohydrate used as energy by the tissues. Under circumstances reflected by normal consumption of sucrose or high-fructose corn syrups, little fructose appears in the blood because it is rapidly metabolized by the liver for synthesis of glucose and primarily fatty acids and triglycerides, and possibly by metabolism in the kidney. The utilization of fructose by peripheral tissues seems to be negligible, perhaps because glucose, which is always present in the blood, inhibits both its uptake and entry into the metabolic pathway of glycolysis.

As a result of insulin release in response to sucrose ingestion, blood glucose is maintained within a closely regulated range, a requirement for normal functioning of the central nervous system. Sugar does not cause abnormal insulin and blood glucose responses. In fact, the effect of ingesting sucrose on

insulin release and blood glucose is less than when an equivalent amount of carbohydrate is consumed from bread. The excess of glucose above the energy requirements of the tissues can be stored to some extent as glycogen in muscle. The remainder is used to synthesize fatty acids and glycerol, which allow efficient storage of energy in the form of triglycerides in adipose tissue.

The effect of sugars on blood glucose can be captured by measuring the total glycemic response over time. To compare the response produced by different carbohydrates, the glycemic index (GI) was developed. The GI compares the incremental area under the blood glucose curve produced by 50 g of a carbohydrate test food to that produced by 50 g of a standard food (white bread or glucose). The glycemic indices of sugars depends on their composition. For example, sucrose has a GI which is only 87% of the white bread standard. This glycemic response is explained by its components glucose and fructose. Glucose alone has a GI relative to white bread of 138%, whereas that of fructose is only 32%. Relative to many common starchy foods, both sucrose and fructose have lower GI values (Table 1).

Nutritional Properties of Sucrose

Sucrose has two important nutritional properties. It provides energy at 3.8 kcal g^{-1} and, through its sweet taste, it increases the palatability of foods.

Table 1 Glycemic index of selected foods[a]

Foods	*Glycemic index (GI)*
White bread	100
All Bran	60
Cornflakes	119
Oat bran	78
Apple	52
Banana	83
Cornmeal	98
Rice, white	81
Rice, brown	79
Rice, instant	128
Rice, parboiled	68
Baked potato	121
Boiled white potato	80
Milk, whole	39
Spaghetti, white	59
Chocolate (various)	84
Sucrose	87
Fructose	32
Glucose	138
Soya beans	23

[a]Selected from FAO/WHO (1997) *Carbohydrates in Human Nutrition. FAO/WHO Expert Consultation on Carbohydrates in Human Nutrition*. Rome, Italy: FAO/WHO.

Sucrose, in the form that it reaches the table of the consumer, is the purest of common foods and free of contaminating chemical residues. As a result, it provides only one nutrient, carbohydrate, which is readily converted to energy by the body. It is this aspect of sucrose that has led to the concept of 'empty calories.' Clearly, sucrose can be used to manufacture foods that are low in essential nutrients and, if consumed in excess, may lead to nutrient-deficiency syndrome. However, there is no objective evidence that sugar intake leads to nutrient deficiencies at current consumption levels, nor that its widespread use has compromised the food supply so that nutritionally adequate meals and diets cannot be obtained from the marketplace.

Diets containing too much sugar are theoretically possible, and examples based on readily available foods can be easily constructed. For the majority of people, however, consumption of sugar falls well within a safe range of intake.

The probable reason why sugar does not lead to nutrient-imbalanced diets is because it is used to enhance the intake of many nutrient-rich foods which may not otherwise be readily eaten. For example, small amounts of sugar are added to many fruits, vegetables, and cereals. Increasing the palatability of these foods with sugar makes them more likely to be consumed.

Many reviews have examined the role of sugar in health and indicate that, other than contributing to dental caries, which is a consideration for all carbohydrates, the amounts consumed in the average diet produce no deleterious health effects. However, perceptions that sugar is the cause of hypertriglyceridemia, diabetes, and hyperactivity in children still persist.

Hypertriglyceridemia is the elevation of triglycerides in the blood and is believed to play a role in heart disease. It has been recognized for years that large amounts of dietary sugars, particularly fructose, can raise triglycerides. Thus, sugars that contain fructose, such as sucrose and high-fructose corn syrups, also have this potential. The potential effect of fructose consumption on blood lipids is explained by its unique metabolic pathway in the liver and evidence that it is a better substrate than glucose for lipid synthesis. Therefore, it should not be a surprise to find an elevation of blood triglyceride concentrations in some subjects given excessive quantities of fructose or sucrose. However in studies in which amounts of sugars typical of the western diet were provided, such responses are not usually observed.

Diabetes mellitus is a condition in which an insulin deficiency or decreased insulin sensitivity results in

hyperglycemia. The most prevalent form of diabetes is noninsulin-dependent diabetes mellitus (NIDDM), which most often has its onset in adulthood and is associated with obesity. All reviews on the subject have concluded that there is no evidence that sugar intake is related to the incidence of diabetes or mortality. In fact, as with obesity, an inverse relationship between sugar intake and diabetes has been found.

The belief that sugar causes hyperactivity in children is one of the most persistent and widespread myths about sugar. However, the link between sugar and hyperactivity has been based largely on subjective observations made by both parents and teachers. Recent well-controlled challenge studies clearly indicate that sugar does not affect hyperactivity, attention span, or cognition in children. In fact, studies have shown that children are calmer and more sedate following sugar preloads.

Sugar Consumption versus Sugar Disappearance Data

A great deal of confusion surrounds sugar consumption statistics. This is because there are two types of consumption being interchanged in discussion of the statistics. One consumption statistic is based on economic concepts and refers to national usage or availability of sugar for use in the food supply, whereas the other consumption statistic is an estimate of the amount eaten by humans and is a physiological concept.

Estimates of the total amount of nutritive sweeteners available for use within a country are assembled on an annual basis by statistics or economic branches of the government. The figures are reported as per capita consumption, but they do not represent per capita intake by individuals. They indicate availability of added sugars to the marketplace, after consideration of production, imports, and exports, and beginning and ending stocks. No allowance is made for losses or wastage in food manufacturing or in other aspects of usage. Thus, while these consumption statistics have proved useful in describing trends in national availability of added sugars, they have proved to be misleading in predicting actual intake by humans. For example, in 1980–85, total caloric sweeteners available for consumption in the USA averaged 155 g (620 kcal) per person per day and 17% of the energy content of the available food supply. However, much of the sugar available is destroyed in production processes such as making bread, pickles and alcoholic beverages, or wasted during distribution, storage, and usage, or consumed by animals.

Added sugar consumption in the USA was estimated to average only 53 g day^{-1}, or 30% of estimated availability (155 g) derived from disappearance data in 1977–78. Consumption data for sugars added to the food supply were derived from the 1977–78 (US Department of Agriculture (USDA)) Nationwide Food Consumption Survey. Average daily intake of sugars for this population sample of 30 677 subjects was 53 g day^{-1}, or 11% of the daily intake of calories. Younger age groups had a higher intake than older age groups. The highest consumers, those in the 90th percentile of daily intake of added sugar, ranged from 1.5 to 2.5 times the mean value for the 14 age/sex groups. For the total population the 90th percentile of daily intake of added sugar was 104 g day^{-1} or 20% of calories.

More recent (1999) economic estimates of per capita consumption show that availability of caloric sweeteners averages 196 g (745 kcal) per person per day. This increase of 29%, since the early 1980s, is somewhat less than the increased per capita availability of total flour and cereal products (36%) and is greater than the increase in total fat availability (19%). When total added sugars availability is related to the availability of calories in the food supply it now approximates 20%, slightly up from 17% in the early 1980s. In contrast, more recent dietary surveys estimate consumption of sweeteners to be close to 80 g day^{-1} or nearly 18% of the energy consumed. These data suggest that the amount of added sugars eaten has risen by almost 50% from 20 years ago and is 42% of the added sugar availability. However the majority of this reported increase in intake can be accounted for by a change in the approach and definition of sugars used by USDA in reporting dietary survey data.

By definition, sugars are mono- and dissaccharides and this definition was used in the earlier survey estimate of sugar intakes. In more recent surveys all high-fructose corn syrups are included as added sugar, which does not adjust for the 70% of these syrups that are made up of polysaccharides. This fact, plus the failure to consider loss of sugars in baking and other processes, and the inclusion of high-intensity sweeteners in the calculation, account for most of the calculated increase in added sugar consumption. Thus at present it is uncertain the extent to which the modest increase in availability of added sugars as a portion of the energy content of the available food supply has influenced dietary intakes.

Since the mid-1970s there has been a considerable replacement of sucrose with corn sweeteners, including high-fructose corn syrup, glucose syrups, and dextrose. At the present time, corn sweeteners account for about 50% of the total added

sweetener that is available in the USA. Because the most commonly used high-fructose corn syrups contain 52% or less of fructose, their availability has done little to affect fructose consumption compared with the consumption of sucrose which is 50% fructose.

Due to changing lifestyles since the early 1900s, the major use of sugars has shifted from direct use by consumers in households to use by the food industry in baked goods, processed foods, and beverages. In 1925, consumer use accounted for about two-thirds of total sugar disappearance, whereas the food industry used only one-third. Today these figures are reversed. In addition to the use of sugar as a sweetener, it is widely used for its functional properties, such as providing bulk in baked goods and inhibiting spoilage microorganisms in jams and jellies. These functional uses are the same whether they occur in the home or the industry, and are essential to achieving a safe and tasty food supply. Therefore the shift to use by the food industry of a greater proportion of sugar disappearance does not suggest that the consumer is eating more sugar than in previous decades.

Dietary Goals and Guidelines

Primarily because of concern with chronic disease, developed countries have produced statements of dietary goals and guidelines aimed at improving the dietary pattern of their populations. The assumption is that nutritional factors are important in the development of chronic disease and that corrective actions in terms of diet will prove to be beneficial.

Sucrose has appeared prominently in dietary guidelines. Recent nutrition recommendations of the governments of Canada and the USA have taken a more liberal view of the nutritional role of sugar in the diet. The American and Canadian Diabetes Associations recommend that sugars and sugar-containing foods do not need to be restricted by people with diabetes and should be considered in the context of the total amount of carbohydrates consumed. The Canadian dietary guidelines for healthy people are silent on the subject, based on the conclusion of the scientific advisory committee that there was no basis on which to advise reduction of sugar intake to the population. Similarly, the guidelines for the USA are no longer punitive and have shifted from 'avoid too much sugar' to 'choose beverages and foods to moderate your intake of sugars.' This advice emphasizes the requirement of individuals to select diets that are based on variety and moderation. Avoidance of refined sugars is no guarantee of a healthy diet.

Sucrose and Food Intake

Sucrose, through its sweetness and other functional properties, makes foods more appealing and raises the probability that they will be selected over foods that are not as palatable. It is well known that the hedonic value of a food is a major determinant of its consumption. However, there is a natural response to sweetness and palatability of foods that tends to be self-limiting. Sensory-specific satiety, which refers to the decreased pleasantness of a food or beverage as it is consumed, occurs for all foods and tastes. Thus increasing the pleasantness of a food may influence its choice but does not necessarily lead to excess consumption.

The role of a varied and palatable diet in the etiology of obesity is currently of great interest because of the high prevalence of obesity in western populations. In the past 40 years, the incidence of obesity has increased fivefold. It is unlikely that sucrose can be rationalized as a causative factor, for several reasons. First, on a strictly correlative basis, the increase in obesity does not associate with the constant intake of sugars occurring during this same period. Second, obese subjects do not show a disproportionate intake of sweet foods compared with normal-weight individuals.

In contrast to hypotheses that sugars bypass appetite-regulatory systems, experimental studies indicate that consumption of sugars suppresses subsequent energy intake. Both children and adults compensate for energy provided by sucrose preloads when given within 1 h before a test meal. Similarly, compensation is also observed for other sugars, which produce higher glycemic responses, including glucose and polycose. Therefore, a rapid increase in blood glucose following ingestion of rapidly digestible carbohydrates does not negatively impact food intake regulation. In general, ≥ 50 g of sugar within 20–60 min of a meal is sufficient to reduce meal intake.

Mechanisms controlling the regulation of food intake remain poorly understood. It seems clear, however, that the cause of obesity is an excess of energy intake over expenditure. It is far too simple to attribute its origin to the composition of the food supply or of a particular food. Many foods and many dietary habits probably contribute to this imbalance, but for most the causative factor of greatest impact is reduced energy expenditure in the form of physical activity.

See also: **Diabetes Mellitus**: Etiology; **Glucose**: Maintenance of Blood Glucose Level; **Obesity**: Epidemiology; Etiology and Diagnosis; **Sugar**: Sugarcane; Refining of Sugarbeet and Sugarcane; **Sweeteners**: Intensive

Further Reading

American Diabetes Association Position Statement (2002) Evidence-based nutrition principles and recommendations for the treatment and prevention of diabetes and related complications. *Diabetes Care* 25: 202–212.

Anderson GH (1989) Sugar consumption. Are dietary guidelines needed? *Journal of the Canadian Dietetic Association* 50: 229–232.

Anderson GH (1998) Sugar consumption and health. *Food Science Biotechnology* 7(4): 229–235.

Duffy VD and Anderson GH (1998) Position of the American Dietetic Association: use of nutritive and nonnutritive sweeteners. *Journal of American Dietetic Association* 98: 580–587.

FAO/WHO (1997) *Carbohydrates in Human Nutrition. FAO/WHO Expert Consultation on Carbohydrates in Human Nutrition*. Rome, Italy: FAO/WHO.

Glinsmann WH, Irausquin H and Park YK (1986) Evaluation of health aspects of sugars contained in carbohydrate sweeteners: report of the Sugars Task Force. *Journal of Nutrition* 116(11S): S1–216.

Vettorazzi G and Macdonald I (1988) *Sucrose: Nutritional and Safety Aspects*. London: Springer-Verlag.

WHO (1997) *Obesity: Preventing and Managing the Global Epidemic*. Report of a WHO Consultation on Obesity. Geneva, Switzerland: WHO.

Woodend DM and Anderson GH (2001) Effect of sucrose and safflower oil preloads on short term appetite and food intake in young men. *Appetite* 37: 185–195.

USDA/Economic Research Service (1998) *US Food Consumption Sources of Data and Trends*. Washington, DC: USDA/Economic Research Service.

US Department of Agriculture, US Department of Health and Human Services (2000) Nutrition and Your Health: Dietary Guidelines for Americans. Home and Garden Bulletin No. 232. Washington, DC: Government Printing Office.

SUGAR

Contents

Sugarcane

M A Godshall, Sugar Processing Research Institute, New Orleans, LA, USA
B L Legendre, US Dept of Agriculture, Houma, LA, USA

Introduction

During the decade of the 1980s, word sugar production held steady at around 100×10^6 t, with cane sugar accounting for 60–64% of the total, and beet sugar for the remainder. The decade of the 1990s was characterized by a steadily increasing output of sugar, so that total world sugar production was forecast to reach 121.5×10^6 t in 1999–2000. Cane sugar now accounts for about 70% of total world sugar production. Although most tropical and subtropical countries produce cane sugar, 10 countries account for 74% of global cane sugar production, as shown in **Table 1**.

Taxonomy and Origin

Sugarcane is a large perennial tropical grass belonging to the tribe Andropogoneae of the family Gramineae and the genus *Saccharum*. The Andropogoneae

Table 1 Raw cane sugar production for the 10 leading producing countries (1997–98 crop)[a]

Country	*Raw sugar ($\times 10^6$ t)*	*Harvest season*
Brazil	14.557	June–May
India[b]	12.523	Oct–Sep
China	6.705	Sep–Jan
Australia	5.350	June–Dec
Mexico	4.997	Nov–Sep
Thailand	3.913	Nov–May
Pakistan	3.400	Oct–July
Cuba	2.951	Nov–June
USA	2.692	Oct–Mar
South Africa	2.334	May–Feb

[a]Total world cane raw sugar production in 1997–98 was 30.212×10^6 t.
[b]Includes khandsari, a cottage-industry-produced, semiwhite centrifugal sugar.

are characteristically tropical or subtropical with a high concentration of genera in two geographical areas: India and Indonesia. The genera *Saccharum, Erianthus* (sect. *Ripidium*), *Sclerostachya*, and *Narenga*, most cited in the origin of sugarcane, constitute an interbreeding group that, along with three species of *Saccharum* (*S. officinarum* L., *S. barberi* Jeswiet, and *S. sinense* Roxb) were used for commercial sugar production. *Saccharum officinarum* is a progenitor of all modern sugarcane cultivars. However, the presence of the interbreeding *Saccharum* complex of the three sugar species as well as its wild relatives, *S. spontaneum* L. and *S. robustum* Brandes and Jeswiet ex Grassl, has provided a genetic pool of unparalleled diversity, allowing for the development of thousands of cultivars that are adapted to the areas where sugarcane is grown. Today, most cultivars of sugarcane are interspecific hybrids of two or more of the five *Saccharum* species.

Description

The sugarcane plant is a large, jointed grass 2.5–4 m in height, with robust stems up to 5 cm in diameter and a tendency to tiller profusely in clumps called stools (Figure 1). The stem is made up of 10–40 internodes with leaves borne at the nodes. The inflorescence (Figure 2) is a large, terminal panicle, feathery in appearance, white to purplish in color.

History

The evolution of sugarcane is postulated to have had three foci – Polynesia, China, and India – and must have started many thousands of years ago. Sugarcane is a readily transported food plant and capable of germination several weeks after cutting from the parent stool. The 'noble' cultivars (clones of *S. officinarum*) were undoubtedly grown in native gardens of Polynesia as a food plant since earliest times. They were also carried throughout the Pacific Island archipelago by Polynesian navigators more than 1000 years ago.

There are well-documented records of sugarcane in both China and India several centuries before the Christian era. Sugarcane was carried from India to Persia in the sixth century, after which the Arabs took it to the Mediterranean littoral. The Egyptians were the first to refine sugar.

Sugarcane reached Spain and Portugal in the eighth century and was subsequently brought to the Canary Islands, the Azores, and East Africa. The first

Figure 1 Example of sugarcane growth habit. Courtesy of Mary An Godshall and Alfred D French.

Figure 2 Flowering sugarcane in Cairns, Australia. Courtesy of Mary An Godshall and Alfred D French.

introduction to the New World was made by Columbus, who carried the cane from the Canary Islands to what is now the Dominican Republic on his second voyage in 1493. By 1550, sugarcane was introduced to Mexico, Brazil, and Peru.

Because of its scarcity and expense, for much of human history, sugar was considered a valuable medicine and later a spice. It was not until the mid to late eighteenth century that practical methods for sugar extraction were developed, making cane sugar widely available and less expensive.

Breeding Programs

The productivity of any sugarcane industry is constantly threatened by diseases, insect pests, and climate. To insure consistent, economical production, new and improved, high-yielding sugarcane cultivars are constantly needed to replace those removed from production. Further, most sugarcane cultivars have a finite life span of useful productivity. Prior to the early 1990s sugarcane was improved mostly by cultivar substitution, making use of naturally occurring clones (mainly *S. officinarum*). This gave improved yields owing to better adaptation or greater inherent disease resistance. In 1888, the fertility of sugarcane seed was discovered in Java, Barbados, and Brazil, and before the end of the century, organized breeding programs were established in six countries. Initially, improvements were realized using a procedure termed 'nobilization,' whereby the hardy and disease-resistant, but otherwise inferior wild cane types, i.e., *S. spontaneum* and *S. barberi*, were crossed and back-crossed with the noble canes, *S. officinarum*, which are characteristically higher in juice quality.

Today, most sugarcane cultivars grown around the world are trispecies hybrids that can be traced to only a relatively few clones of the three species mentioned above. The major objectives of most breeding programs are to combine high-yield potential and improved juice quality with disease resistance and other important agronomic characteristics. Recurrent breeding and selection programs based on yield potential have generally been highly successful. Most of the economically important characteristics in sugarcane are quantitative ones controlled by multiple genes. Therefore, large seedling populations are required when attempting to develop new, high-yielding sugarcane cultivars when selecting for as many as 20 or more individual characteristics.

In an effort to broaden the narrow genetic base, a basic or prebreeding program has been established in many countries with its objective to broaden the genetic base upon which contemporary cultivars are bred. This base-broadening program makes most cultivars less vulnerable to disease and insect pests and increases their agronomic worth. However, the basic breeding program is long-term and high-risk, and has had success in only a few countries. It has had a dramatic impact upon the sugar industry of Louisiana in recent years with the release of three superior, high-yielding cultivars for commercial planting. One of

these cultivars, a BC_4 progeny of the *S. spontaneum* clone US 56-15-8, now occupies approximately 70% of the planted area. This cultivar alone has increased productivity by over 30%, leading to a record yield of sugar per unit area during the 1999–2000 harvest season that approached 8.16 t of sugar per hectare for a crop season which averages only 7–9 months.

To augment traditional selection strategies, researchers in various countries have developed species-specific markers in recent years that allow for the accurate selection of true hybrids of these different species at the early seedling stage. In the past, it was only through morphological traits that apparent true hybrids were selected in the basic breeding programs. However, in many cases it was later found that what was thought to be a true hybrid was, in fact, a self or a contaminant. This new molecular approach has given plant breeders a powerful tool to select only those genotypes that carry the desired markers.

Each program is tailored for the needs of its specific location, which can be quite different from country to country. For example, in Louisiana, breeding goals include development of early maturing cultivars with cold tolerance suited for the very short growing season of 9 months. These goals would be meaningless in a more tropical climate where other needs predominate. Further, rarely will a cultivar selected for one environment be adapted to the environment of another area.

Augmenting traditional approaches with new molecular techniques fills a critical need in developing these improved cultivars through wide hybridization and genetic transformation. In recent years, researchers in several countries have been successful in inserting genes into sugarcane plants that express resistance to various diseases, insects, and herbicides. Because of the great potential gains that can be realized from molecular approaches in the development of superior new cultivars, this long-term, high-risk research also offers opportunities for the twenty-first century. However, it is not envisioned that these molecular approaches will replace traditional breeding and selection in the development of future high-yielding cultivars.

Diseases and Pests

Sugarcane plants are subject to a number of bacterial, fungal, and viral diseases. At any one time in any given location, there are usually three or four prevalent diseases of major concern. The severity of infestations increases and decreases in various parts of the world depending on factors such as types of cultivars grown and other control measures. The most recent diseases to appear in the western hemisphere are: leaf scald, caused by the bacterium *Xanthomonas albilineans* (Ashby) Dowson, which arrived in USA (Florida) in 1967; smut, caused by the fungus *Ustilago scitaminea* Sydow, in 1978; rust, caused by the fungus *Puccinia melanocephala* H. & P., in 1979; and yellow-leaf syndrome, caused by a virus, in 1993.

Other important diseases include sugarcane mosaic, a viral disease which caused severe losses throughout the world in the earlier part of the century, and ratoon-stunting disease, caused by the bacterium *Clavibacterium xyli*. Red rot, caused by the fungus *Collectotrichum falcatum*, was once severe in many parts of the world but now appears to be a problem mainly in India.

Pests include rats – a severe problem in some areas – nematodes, and a number of insects. The most severe insect pests are the various types of borers, i.e., the sugarcane borer, *Diatrea saccharalis* (F.), and the eldana borer, *Eldana saccharina*, which cause damage first by boring into the cane stalk, then by providing entry points for other diseases and, finally, by reducing cane and juice quality.

Various weeds cause problems in sugarcane culture by competing for nutrients and crowding or overgrowing the young plants. Perennial grasses are the most serious weeds, harboring insects and diseases. Preemergent herbicides are commonly used for control.

Control Measures

In recent years, detection of several of these diseases has been made easier with the development of DNA-based molecular diagnostic protocols. Timeliness of diagnosis is important to the management of all diseases; however, cultivar resistance through traditional and molecular approaches still remains as the primary control mechanism.

Very little chemical treatment with fungicides or other agents is used. Such measures are largely ineffective because the leafy canopy of the cane plant prevents proper application, and such large amounts would need to be applied as to be prohibitively expensive and environmentally unsound. (*See* **Fungicides**.)

For some diseases, especially ratoon-stunting disease, the seed cane is treated with hot water.

In recent years, experiments have progressed on the use of biological control agents for long-term control of insects. Examples include the *Cordyceps* fungus to control white grubs, parasitization with the wasp *Apanteles flavipes* to control borer, and the fungus *Metarhizium anisopliae* for leaf hopper control.

The sugarcane borer, *Diatraea saccharalis* (F.), and the Mexican rice borer, *Eureuma loftini* (Dyar), are

major insect pests of the mainland USA sugarcane industry. However, in 1986 a recurrent selection program was initiated in Louisiana to develop sugarcane cultivars with resistance to these two stalk borers. Plant resistance as well as biological control have become increasingly important as tactics used in integrated pest management (IPM), thus reducing the need for chemical pesticides and increasing profit margins for farmers. During the past 10–15 years, the need for chemical pesticides has been reduced from four applications per unit area per year to less than one application per unit area per year, with the use of more resistant cultivars and IPM.

Culture

Sugarcane grows in the warm, humid tropics and subtropics. It needs at least 600 mm of annual rainfall but responds well to irrigation where rainfall is limited.

Propagation is asexual, accomplished by cuttings known as sets or 'seed cane' which contain at least one bud. It is also propagated sexually by pollination, but this is usually seen only in breeding programs since sugarcane is a polyploid hybrid and does not breed true.

Sugarcane normally requires large inputs of chemical fertilizer, especially nitrogen, in order to maximize yields; however, excellent yields can be obtained with minimum cultivation.

Types of cane harvesting are determined by terrain, soil types, and the availability and cost of labor. Methods range from manual cutting to full mechanical harvesting, which includes whole-stalk and billet-cane (combine) harvesting. In some parts of the world, the cane is briefly burned to remove the dry, trashy leaf residue. Burning also helps to reduce the number of field pests, such as snails, rats, and insects. The cane may be topped or not, burned standing or lying cut in the field, or taken green (unburned) to the mill.

Harvest seasons vary with location, ranging from mid-June to December in Queensland, from mid-October to March in Texas and Florida, and early October to the end of December in Louisiana, which is one of the shortest harvest seasons in the world. In a few tropical areas, the cane may be harvested all year around. Table 1 shows the harvest dates for several countries.

Maturation and Use of Ripeners

In nature, the cane ripens, or matures, when growth slows and the synthesis and storage of sucrose occur. Ripening is influenced by many factors, with the most important being the age of the cane, incident sunlight, temperature, and rainfall. Cultivars differ in their optimum ripening time, some being considered early-maturing and some late-maturing.

The use of chemical ripeners has increased since the introduction of glyphosine in 1972. Glyphosine, glyphosate, mefluidide, and ethepon are the compounds most widely used. The harvest is timed to coincide with the maximum level of sucrose in the mature stalks, and chemical ripeners are used to maximize the sucrose-producing potential of the cane stalks, especially under adverse ripening conditions. Ripeners are not able to induce the production of more sucrose than is inherent to the cultivar, nor are they economically effective when natural ripening conditions are optimum.

Production of Raw Sugar

The harvested sugarcane cannot be stored for very long without deteriorating, so it is converted to raw sugar with as little delay as possible. The sugar is contained in the stalk juice, ranging in concentration from 15% to 20% sucrose. The juice also contains a number of other constituents, such as starch and other polysaccharides, bits of fiber, flavonoid and anthocyanin pigments, protein, amino acids, aconitic, and other organic acids and salts. The purpose of sugar manufacture is to separate the sugar from these other components and to crystallize it in as pure a form as economically feasible.

The first stage of sugar manufacture is the production of raw sugar, which occurs in the sugar mill or factory. Juice is extracted from the stalks by either of two processes: milling, which is the more common, or diffusion. In either process, the cane tissue is sufficiently disrupted by cutting, shredding, grinding or 'fiberizing' to achieve 95–97% extraction of available sucrose. (*See* **Sucrose**: Properties and Determination.)

The raw cane juice is turbid and acidic, with pH in the range of 5.3–5.7. The first stage of sugar production is clarification (or purification) of the juice, designed to remove soluble and insoluble impurities and to inactivate enzymes that hydrolyze sucrose (invertase). Clarification is carried out by the addition of lime and flocculents to the heated cane juice. This coagulates and precipitates insoluble and colloidal material and raises the pH to near neutral. Increasing the pH of the juice helps to stabilize the sucrose against acid hydrolysis. After filtration, the clarified juice is concentrated by evaporation into crystalline raw sugar. In recent years, several variations on the above process have been developed to improve yield and product quality. Among these are the use of

enzymes to reduce problematic polysaccharides. For example, dextranase enzyme is used to reduce the amount of dextran, which is caused by infection with the bacterium *Leuconostoc mesenterioides*, and amylase is sometimes used to decrease the amount of starch in the juice from some high-starch cultivars. Expensive enzymes such as these are used only on an 'as-needed' or emergency basis.

During processing, the cane juice changes color from dark green to golden brown owing to various color-forming reactions, the most important ones being enzymatic and nonenzymatic browning and polymerization of polyphenolic acids, the Maillard browning reaction between reducing sugars and amino acids, and the caramelization reaction caused by thermal degradation of sugars. Manufacturers try to maximize sucrose yield and minimize the reactions that produce color or cause sucrose loss. (*See* **Browning**: Nonenzymatic; **Caramel**: Properties and Analysis.)

In the USA and some European countries, raw sugar is considered inedible because it is not made according to the strict sanitary codes established in those countries. However, in many other parts of the world, raw sugar is sold directly to the consumer. The composition of raw sugar is shown in Table 2. A special type of raw sugar, called turbinado sugar, is produced for the edible market in Europe and North America.

In terms of bulk, the major coproduct of milling is bagasse, the fibrous residue of the stalk. Although some of it is used in paper-making, as a chemical feedstock and in cattle feed, its principal use is as an energy source for the mills, allowing most cane factories to be partially or totally energy-self-sufficient. In some areas cogeneration of electricity is also possible.

Raw sugar is a commodity sold on the world market. Many rules and regulations control its sale, its cost, and its transport around the world. Raw sugar is exported from producing areas to refineries in industrialized countries for conversion into highly purified white sugar.

Table 2 Composition of raw sugar

Component	*Range of concentration*
Sucrose	96–99%
Glucose	0.3–0.6%
Fructose	0.3–0.6%
Moisture	0.1–0.5%
Ash	0.1–0.5%
Starch	50–400 p.p.m.
Other polysaccharides	800–1500 p.p.m.
Insoluble matter (soil and vegetable particles)	200–500 p.p.m.
Color	800–3000 ICUMSA units[a]

[a]ICUMSA, International Commission for Uniform Methods of Sugar Analysis. It is the absorbancy index of the solution × 1000:

$$\frac{A_{420} \times 1000}{\text{cell length (cm)} \times \text{Conc. (g ml}^{-1})}$$

Specialty Products from Sugarcane

While the major product from sugarcane is white, granulated sugar, the process also lends itself to the production of several specialty sweeteners. Sugarcane is uniquely suited, by virtue of trace components present in the juice and the reactions that occur during processing, to produce sugars with desirable flavors and colors. The latter range from light yellow to dark brown, and flavors include a range of molasses and browned/caramelized types.

The specialty sweeteners from sugarcane include factory (mill) products such as turbinado sugar, edible molasses, Demerara sugar, muscovado sugar, plantation white sugar, sugarcane syrup, and dried cane juice. In the last 5 years or so, organic sugar has come on the market in various locations. Specialties from refineries include brown sugar, yellow sugar (popular in Canada), golden syrups, refinery syrups, and molasses. Less common specialties include canned or packaged, pasteurized cane juice, canned cane stalks, dehydrated molasses, and liquid brown sugar.

Brown sugar, which is produced in refineries, is a very fine grain sugar which is enveloped in a thin film of dark syrup and may range in color from light yellow to very dark brown. It is produced in either of two ways. The traditional method is to crystallize it from a refinery syrup selected for its color and flavor. The second, more energy-efficient way is to coat white sugar crystals with cane syrup or molasses. The latter method can be used to produce brown sugar from beet sugar.

The last decade has also seen great changes in the production of plantation white sugar, which is a white sugar produced directly at the mill, in lieu of raw sugar. This sugar is not of as high quality as refined sugar, having slightly higher color, ash, moisture, and reducing sugar content, but it is acceptable for almost all domestic applications. It has the advantage that its production is much less expensive than that of refined sugar, and it can be made in existing raw sugar mills with a minimum of redesign of equipment. Currently, most plantation white sugar is used within the country of production, but several countries, notably Brazil, have in recent years developed a worldwide trade in plantation white sugar.

See also: **Browning**: Nonenzymatic; **Caramel**: Properties and Analysis; **Fungicides**; **Sucrose**: Properties and Determination

Further Reading

Chen JCP and Chou CC (eds) (1993) *Cane Sugar Handbook*, 12th edn. New York: John Wiley.

Clarke MA and Godshall MA (eds) (1988) *Chemistry and Processing of Sugarbeet and Sugarcane*. Amsterdam: Elsevier Science.

Heinz DJ (ed.) (1987) *Sugarcane Improvement through Breeding*. Amsterdam: Elsevier Science.

F.O. Lichts World Sugar and Sweetener Yearbook 1998/99 (1998) (F.O. Lichts World Sugar Statistics 1998/99). Ratzeburg, Germany: F.O. Licht.

Pan Y-B, Grisham MP, Burner DM, Legendre BL and Wei Q (2000) DNA-based molecular diagnostic protocols for sugarcane bacterial diseases. In: Solomon S (eds) *Sugar: Agro-Industrial Imperatives (Vision for 21st Century)*. New Delhi: Oxford/IBH Publishing.

USDA (1999) *Sugar and Sweeteners: Situation and Outlook Yearbook*. Rockville, Maryland: US Department of Agriculture, Commodity Economics Division, Economic Research Service.

The internet is now a source of information about sugar and sugar statistics. A few recommended sites are:

US Department of Agriculture, Foreign Agriculture Service http://www.fas.usda.gov/htp2/sugar/1999/November/toc.html

Sugar and sweetener summary reports http://usda.mannlib.cornell.edu/reports/erssor/specialty/sss-bb/

Sugar Information http://www.sugarinfo.co.uk/index.htm

Louisiana State University http://www.lib.lsu.edu/special/sugar/intro.html

Sugarbeet

M A Clarke, Sugar Processing Research Institute Incorporated, New Orleans, LA, USA
M F G Cleary, Holly Sugar Corporation, Colorado Springs, CO, USA

Introduction

Total world sugar production in 1990–91 was 112.6 $\times 10^6$ t, of which 41.2 $\times 10^6$ t were produced from sugarbeet, and the rest from sugarcane. Some 35% of world beet sugar is produced in the EC, and 9% in the USA. **Table 1** lists areas, yields, and sugar yield for the 10 major sugarbeet-producing countries based on (1989 to 1991 data). Sugarbeet is grown in many climates, from temperate (southern California, Spain, Italy) to relatively cold climates (North Dakota, Finland). In general, sugarbeet is converted directly to white sugar, refined quality, although some eastern European areas still make an intermediate raw sugar because of lack of processing facilities. Costs of production vary widely with climate, yield, and socioeconomic factors. Average production costs (grower cost) and factory cost (processor cost) are shown in **Table 2** for the USA, an intermediate-cost producer.

Table 1 World sugarbeet statistics: area harvested, yield, and production for leading producing countries

Country and year	*Area harvested ($\times 10^3$ ha)*	*Sugarbeet yield (t ha^{-1})*	*Sugar yield (t ha^{-1})*	*Share of world beet sugar production (%)*
Former Soviet Union				
1989–90	3344	29.1	2.85	24.4
1990–91	3267	24.9	2.80	22.2
Germany				
1989–90	609	44.2	6.71	10.4
1990–91	619	49.0	7.54	11.3
USA				
1989–90	524	43.5	6.00	8.0
1990–91	558	44.8	6.27	8.5
France				
1989–90	427	56.0	9.85	10.7
1990–91	474	52.4	10.01	11.5
Turkey				
1989–90	350	31.2	3.94	3.5
1990–91	370	35.1	5.14	4.6
Poland				
1989–90	423	34.0	4.41	4.8
1990–91	440	38.0	4.94	5.3
Italy				
1989–90	290	57.2	6.48	4.8
1990–91	262	44.3	6.06	3.9
China				
1989–90	569	16.2	1.35	2.0
1990–91	730	19.9	1.78	3.2
UK				
1989–90	194	41.2	7.10	3.5
1990–91	192	41.7	6.02	3.3
The Netherlands				
1989–90	128	56.3	9.70	3.2
1990–91	125	69.7	10.71	3.3
World				
1989–90	8620	35.9	4.54	100.0
1990–91	8726	34.7	4.72	100.0

Source: Foreign Agricultural Service, US Department of Agriculture.

Composition

The composition of sugarbeet is shown in **Table 3**. Sugarbeet, being a root, has a much higher level of nitrogen compounds than does sugarcane, and beet processing is geared to removing these, and preventing them from contributing to color and odor compounds. Composition is of vital importance to sugar yield because soluble nonsucrose constituents impede crystallization of sucrose by complexing with sucrose and carrying it into residual molasses. Each kilogram

Table 2 Sugarbeets: US production and processing costs per net ton of sugarbeets and pound of refined sugar (1989 crop)

Item	*Dollars per ton*	*Cents per pound*
Production costs		
Variable cash expenses	20.27	7.881
Fixed cash expenses	8.00	3.110
Operating capital	0.93	0.362
Nonland capital	1.02	0.397
Net land return	8.18	3.180
Unpaid labor	1.78	0.692
Total	40.19	15.626
Processing costs		
Variable cash expenses	22.47	8.735
Fixed expenses	3.45	1.342
General and administrative	1.93	0.750
Pulp drying and marketing	3.43	1.335
Total	31.28	12.162
Total production and processing costs	71.47	27.788
Credits		
Dried pulp	6.92	2.691
Molasses	2.31	0.898
Other	0.51	0.198
Total	9.74	3.787
Net production and processing costs	61.73	24.001
Yield per planted acre (ton)		19.4
Recovery per net ton of beets (lb)		257.2

Table 3 Composition of sugarbeet

Component	*Content (%)*
Juice	*c*.9
Insoluble matter (marc)	*c*.5
Water, chemically bound	*c*.3
Soluble solids	11–25
Sucrose in solids	87.5
Sucrose in beet	10–22
Nonsucrose substances soluble in juice	
Organic nitrogen compounds	*c*.44
Nitrogen-free organic compounds	36
Inorganic compounds	20

of nonsucrose constituents in sugarbeet juice can carry 0.65–0.85 kg of sucrose to molasses, removing it from the yield of crystalline sugar. Plant breeders have bred for high sucrose in beet to the point where selection for lower specific nonsucrose components (sodium, potassium, amino nitrogen) can generate greater total sucrose yields. (*See* **Sucrose**: Properties and Determination.)

Origin, Plant Breeding, and Cultivation

Sugarbeet has been grown for sugar (sucrose) production only since the late eighteenth century, when Achard in Europe identified the 'white Silesian beet' as a source of sugar, and Napoleon encouraged breeding and processing research, to provide a home-produced alternative to cane sugar, which required shipment from the West Indies.

Sugarbeet, grown from seed, is a biennial plant which accomplishes vegetative growth in its first year and seed production in the second. For sugar production it is harvested at the end of the first year of growth, after a frost-free (preferably) period of 5–6 months, in areas with annual rainfall of some 20 cm. Irrigation is used in some areas, e.g., north-western USA. Sugarbeets are often used in crop rotation, as their deep roots bring up nutrients from lower levels of soil, to become available for alternate crops, and because they cannot be planted in the same field more than once every 3–4 years without risk of attack by nematodes. Beets for seed are grown in nonsugar-producing areas, notably in Oregon, where most US and European companies operate joint-venture seed companies with the purpose of maximizing germplasm resources.

Seed is, and has been since the 1940s, all monogerm hybrid, which permits mechanical planting and harvesting. The selected monogerm, male-sterile, inbred line and the multigerm pollinator (chosen for other desirable characteristics) are initially multiplied separately; the former is planted in strips together with the fertile monogerm complementary line. Foundation seed produced by the male-sterile inbred furnishes the seed parent for the hybrid variety, which again is grown in strips with foundation seed of the pollinator. Monogerm hybrid seed is harvested only from the male-sterile parent, while the pollinator is destroyed soon after flowering.

The major varieties currently grown in the USA are shown in Table 4. Varieties must be tested in semicommercial or commercial trials for several years after development to receive test market approval.

Harvest Practice

Clean, unbroken beets, free of leaves, trash, and rocks, are the primary objective. Grower contracts can penalize noncompliance and usually include a right of refusal for beet not fit for processing. This option is rarely exercised in North America, except when beets are so severely degraded as to stop the process completely.

Harvest dates are normally agreed upon prior to planting, with premiums for early and/or late harvest deliveries to compensate the grower for beet low in sugar content and tonnes per hectare (immature), or at risk from weather, respectively. Harvesting normally

Table 4 Sugarbeet varieties grown in specific areas of the USA

Variety	*Area*						
	RRV	*MI-OH*	*CO-NE*	*WY-MT*	*ID-OR*	*TX*	*CA*
Mono-Hy E4		X					
Mono-Hy E9		X					
ACH 185		X					
KW-1745	X						
Hilleshog 5135	X						
Hilleshog 8277	X						
Hilleshog 8351	X						
Maribo Ultramono	X						
Maribo 410	X						
Maribo 403	X						
Maribo 875	X						
Betaseed 2988	X						
Betaseed 2398	X						
Betaseed 3145	X						
Betaseed 3265	X						
Monohikari	X		X				
Hilleshog 1605			X				
Hilleshog 55			X				
Betaseed 3778			X				
Hilleshog R2				X			
Hilleshog D2				X			
HH 50				X			
ACH 191				X			
PM 9					X		
WS 88					X		
MHR 2					X		
MH 55					X		
HM1 TX-9						X	
HH 57						X	
HM1 1803						X	
SS-NB3							X
HH 37							X
HH 45							X
SS-NB2							X
Betaseed 4757							X

RRV, Red River Valley; MI–OH, Michigan–Ohio; CO–NE, Colorado–Nebraska; WY–MT, Wyoming–Montana; ID–OR, Idaho–Oregon; TX, Texas; CA, California.
Data courtesy of Beet Sugar Development Foundation, Denver, Colorado, USA.

takes place in September to November, except in some areas of California where a second crop is harvested in the spring. The combined harvest and processing strategy accounts for both the 'optimum harvest window' for the grower, and the beneficial impact of increased throughput on processor profitability.

Commonly used machines can harvest four to six rows simultaneously, and up to 1000 t per day. None the less, smaller machines (one and two rows) are abundant, and there are still areas where beets are harvested by hand at a rate of less than 10 t per worker per day. The harvester is normally pulled behind a tractor which dumps the harvested beet into an open lorry alongside. Three basic operations are performed: removal of the beet from the ground; removal of the leaves and, perhaps, a portion of the top of the root; and a modest separation of dirt and trash from the root.

Complete removal of leaves is essential, whether before the beet is taken from the ground, or after the beet is lifted. Further removal of a portion of the top of the root is arguably of benefit. The portion that is removed typically displays 67% of the sugar concentration of the main root and has a higher impurity load; its inclusion would reduce both grower payment and processing profitability. On the other hand, a sizable wound is inflicted, which may lead to sugar losses between harvest and processing.

Effect of Ambient Temperature

The technical value of the beet decreases exponentially if the root is subjected to repeated freeze–thaw

cycles, with ambient temperature falling below −5 °C. Irreversible tissue damage occurs, the vascular system collapses, and secondary infections consume the sugar. The first hard freeze will normally kill the leaves while the root remains undamaged, and, if the soil temperatures are reasonably high and a variety has been chosen which 'sits down' in the soil, the crop can withstand several nights of −5 °C to −10 °C temperatures. Some cold-temperature growing areas target this period to accelerate the daily delivery rate to the factory, from about 1% of the crop to up to 10% of the crop per day.

Similar phenomena occur in beets which have been subjected to temperatures in excess of 40–45 °C for a number of days. Since the respiration rate has already been accelerated by heat, and soilborne infections and diseases thrive in warm soils, tissue degradation and technical losses occur much more rapidly. Autumn harvests around the Mediterranean and spring harvests in California are occasional victims of this plague.

Sampling

As beets are delivered and received by the processor, a representative sample is taken in order to determine, ultimately, the amount that will be paid to the grower. The sampling procedure varies from elaborate cross-sectioning of the lorry and subsequent sample-splitting, to a simple 'grab sample' of the falling stream of beet. Samples are evaluated in a laboratory usually operated by the processor but open to inspection by the grower. Parameters normally include sugar content, percentage dirt and trash (tare) and, occasionally, other analyses.

The tare is the portion of the delivery which, by contract, the processor is not obliged to pay for. It always includes rocks, leaves, and trash, but may also include a portion of the top root which the grower had agreed to 'scalp' from the beet. Once the percentage is determined in the laboratory, the delivered weight is adjusted accordingly. Dirt and trash typically contribute 1–2% to the tare value, but difficult harvest conditions may drive this contribution to over 10%. 'Top tare,' if applicable, will normally fall into the 3–5% range.

Sugar content of mature, healthy beet is between 10% and 22% (Table 3) depending on agricultural inputs: variety, disease pressure, climate, and harvest conditions. Grower payment is linearly tied to sugar content, with a multiplier based on the market value of sugar. The typical participating payment scale in the USA returns 60% of the net market value of the recovered sugar to the grower. Some contracts include a premium for exceptionally high sugar concentrations based on the assumption that recovery will increase accordingly.

Analyses for nonsucrose components may be run to estimate the amount of impurities which affect yield on crystallization. Most common are sodium, potassium, and amino nitrogen; they are used ubiquitously in breeding beet varieties, and occasionally in both Europe and North America to adjust the grower payment.

Storage of Beet

After beets are removed from the ground, they can be processed immediately, stored for a few weeks, during which time normal respiration losses are tolerated, or put into long-term storage which requires special equipment and/or techniques to maintain the beet for several months. Respiration losses during the storage of healthy beet are typically 0.15% sugar per week. Another strategy is to increase the beet-processing capacity of the factory and store an intermediate syrup.

When beets are stored in piles for up to 160 days, special precautions may range from simply orienting the pile to minimize exposure to sunlight, to the use of fully enclosed ventilated sheds which can contain 90 000 t of beet. Between these extremes are the use of straw or plastic to cover the piles, and the use of half-round ducts to ventilate beet piles. Typical piles contain 50 000–100 000 t of beet, with a trapezoidal cross-section 50–65 m wide at the base and a height of 6–9 m, depending on storage conditions.

Clean, unbroken beets, with a root temperature below 10 °C, are required to assure the proper air flow and cooling which prevent the natural respiration of the beet from heating the pile. Piles are monitored to detect such hot spots. Ideally, storage should minimize respiration rate without freezing the tissue but, in practice, the outer layer of the pile often suffers severe deterioration to a depth of about 50 cm. Forced-air ventilation may be used intentionally to freeze the entire pile where the climate allows this to be accomplished continuously over a period of a few days: this requires daily temperatures below −10 °C for 4 or 5 days. Such beets are not subject to respiration losses, but must remain frozen until they are processed.

Regardless of storage conditions, beets enter the factory process in a flow of water through the flume during which, allowing for the natural buoyancy of the root, rocks and trash are separated. Some washing is effected during this process, but all processes are supplemented with a beet washer, prior to slicing the beet. Water from the washer is returned to the flume loop, which typically includes a system to remove

suspended solids, lower biochemical oxygen demand (BOD) and chemical oxygen demand (COD), and reuse the water.

Factory Processing

Sugarbeet processing is illustrated in Figure 1. The complete sugarbeet-processing operation involves a number of unit operations, including the following:

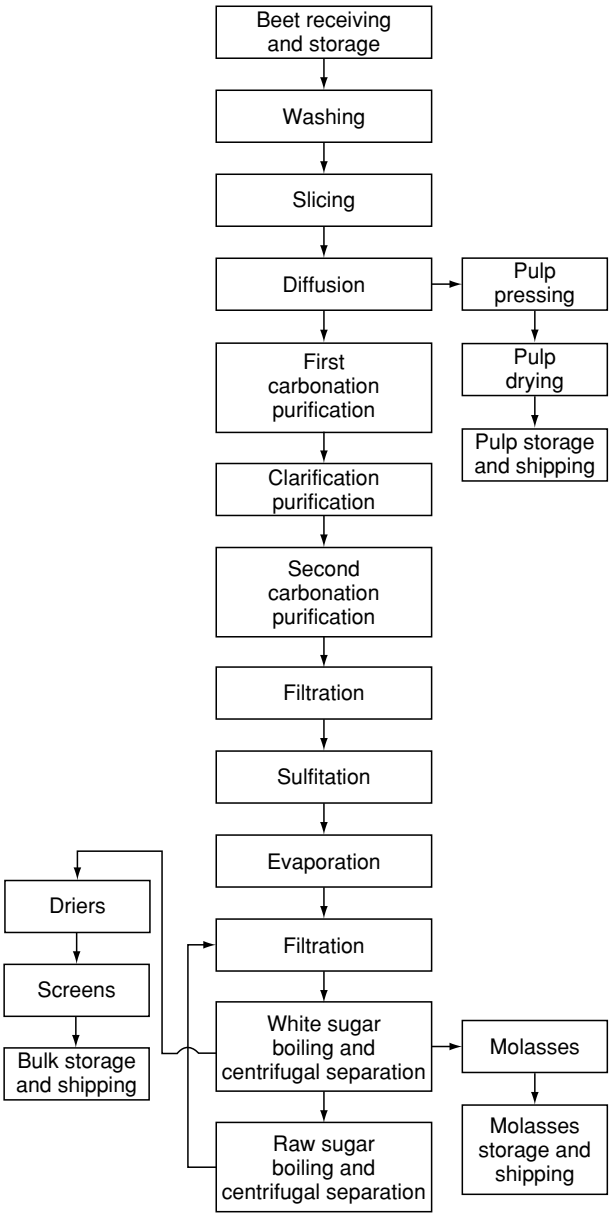

Figure 1 Outline of sugarbeet factory process. Source: Holly Sugar Corporation, a subsidiary of Imperial Holly Corporation, Colorado Springs, USA. Reproduced from Sugar: Sugar Beet, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

1. Diffusion (dried beet pulp – animal feed coproduct).
2. Carbonation (purification).
3. Evaporation (concentration).
4. Crystallization (saleable white sugar; also molasses coproduct used as animal feed and fermentation feedstock).

After washing, beets are sliced into V-shaped cossettes, about 3 cm by 4.5–7.2 cm. This shape ensures maximum surface area for sugar extraction. Cossettes are introduced into a countercurrent continuous diffuser that may be one of several designs: horizontal drum, sloped vat, or vertical-cylindrical tower. All designs involve a mechanism to transport sugarbeet cossettes in a direction against the flow of hot water extractant, and extract some 98% of the sugar (sucrose) in the beet. Sugarbeet cell walls are heat-denaturedto enhance diffusion of sugar from the plant cells to the lower concentration of sugar in the aqueous extractant. Wet beet pulp, discharged from the diffuser at approximately 92% moisture, is pressed mechanically to reduce beet pulp moisture to approximately 76%, and either air-dried under ambient temperatures (in California) or dried in direct gas, oil, or coal-fired rotary drum driers to a moisture content of approximately 10%. Pulp has long been sold as an animal feed, but in recent years it has been purified and sold as a food fiber, for breakfast cereals. Juice from the diffuser (diffuser juice or raw juice) contains approximately 12% sucrose by weight, together with 2% soluble impurities, soluble and semisoluble colloidal proteins, pectins, and saponins extracted from the cossettes; it is heated to 85 °C prior to purification with lime and carbon dioxide, called carbonation.

In this carbonation clarification, semisoluble colloidal materials are coagulated and precipitated as insoluble salts. Potassium and sodium salts of oxalic, malic, and citric acids, and phosphates and sulfates are precipitated as insoluble calcium salts. Milk of lime (a suspension of calcium hydroxide) and carbon dioxide are added continuously and concurrently to a diffusion juice in a carbonation tank, with a retention time of 12–15 min at 85–88 °C. Treated juice is then sent to a clarifier for separation of the precipitated impurities. Thickened underflow mud from the clarifier is filtered on vacuum, rotary drum filters and washed to minimize sugar loss.

Lime and carbon dioxide are produced from limerock (calcium carbonate) at each factory in a coke or natural-gas-fired, vertical-shaft Belgium-type lime kiln, or in a gas-fired, horizontal, rotary drum-type kiln where waste lime is reburned. Carbon dioxide is reclaimed from the kiln for direct use in the

lime–carbon-dioxide first carbonation purification tank. Approximately 2% of lime on sugarbeets is used for purification and clarification.

Clarified overflow juice (thin juice) is then subjected to a second treatment with carbon dioxide to reduce residual lime salts concentration. Calcium carbonate precipitate is filtered from the juice by use of pressure filters. Sulfur dioxide is added to the filtered juice to minimize color formation during subsequent processing steps. The thin juice, after all colloidal impurities and approximately 35% of the soluble impurities have been removed, is concentrated from about 13% dissolved solids (13° Brix) to approximately 60–65% dissolved solids in energy-efficient quintuple-effect evaporators.

Steam at all beet sugar factories is produced in relatively low-pressure (17.5–28 $kg\,cm^{-2}$) boilers. This steam is passed through a turbine to produce electricity for factory use. Exhaust steam from the turbine is used for heating, evaporation, and crystallization of sugar in the latter stages of the process.

Thick juice from the evaporators is concentrated and sugar is crystallized in a three-stage process. The first vacuum-pan crystallization produces white sugar, separated from mother liquor (massecuite) in basket-type batch centrifuges, in which crystals are washed with hot water. Second and third vacuum-pan crystallizers produce lower-purity raw sugars, which are redissolved after separation of the massecuite in continuous conic-basket centrifuges developed specifically for the sugar-processing industry.

White sugar, discharged wet from the batch centrifuges, is dried in rotary hot-air driers (granulators) to a moisture content of approximately 0.02%. The white refined sugar produced is 99.96% pure as a result of successive purification by diffusion, lime–carbon-dioxide purification, and triple crystallization in the batch vacuum-pan crystallizers.

The mother liquor recovered from the last three vacuum-pan crystallizers is exhausted of further raw sugar crystals under standard atmospheric conditions. The massecuite reclaimed from this process is a molasses coproduct, and approximately 60% of the solid material is sucrose. The ratio of nonsugar impurities to sugar prevents further crystallization of sugar using conventional processing technology.

Sugarbeet processing is a seasonal operation, beginning with beet harvest, which must be completed before beets freeze in the ground, and continuing through processing of stored beets. Some factories process additional beets to syrup, or thick juice, which is then stored in tanks and processed after beet input is complete. A longer operating season reduces the fixed costs of a factory.

Molasses Desugarization

In the last decade, a new process has been developed and widely applied in the North American, Japanese, and Finnish beet sugar industries. Molasses desugarization, accomplished by ion exclusion in which the final molasses is passed over ion exchange resins which separate sucrose from other molasses components, allows continuous production of a high-sucrose product stream and a desugarized molasses stream, and can increase a factory's sugar production by 10%, with no increase in incoming beet.

Problems of Waste Disposal

Leaf and root material which have been removed in the field during the harvest are normally reincorporated into the soil as green manure. Occasionally, livestock are allowed to forage directly on this material. Trash and beet pieces recovered from the flumes are also used as cattle feed, either directly or chopped with pulp for drying.

Most factories operate a self-contained system to handle the vast amounts of water used for beet transport and cleaning. Two to four million liters of water per day for a factory processing 5000 $t\,day^{-1}$ are settled to remove suspended solids and to reduce the COD and BOD, usually through an extensive lagoon system, and reused. (*See* **Effluents from Food Processing**: Composition and Analysis.)

Precipitated calcium carbonate, which contains 10–15% organic and phosphate impurities, is produced at a rate of about 4–5% (w/w) of beet. The most up-to-date processes yield material which can be directly used for soil amendment or as an animal feed supplement. Less modern factories will allow this material to dry outside before sale.

See also: **Effluents from Food Processing**: Composition and Analysis; **Sucrose**: Properties and Determination

Further Reading

Clarke MA and Godshall MA (eds) (1988) *Chemistry and Processing of Sugarbeet and Sugarcane*. Amsterdam: Elsevier Science.

Licht FO (1989) *World Sugar and Sweetener Yearbook*. Ratzeburg, Germany: Licht.

McGinnis RA (ed.) (1982) *Beet-Sugar Technology*. Fort Collins, Colorado: Beet and Sugar Development Foundation.

Pennington NL and Baker CW (eds) (1990) *Sugar: A User's Guide to Sucrose*. New York: Van Nostrand Reinhold.

USDA (1990) *Sugar and Sweeteners: Situation and Outlook Yearbook*, SRRV15N2. Rockville, Maryland: US Department of Agriculture, Commodity Economics Division, Economic Research Service.

Palms and Maples

L W Doner, Eastern Regional Research Center, Wyndmoor, PA, USA

Introduction

The sucrose content of sap from palm trees is extensively variable, but it often exceeds 10%. This is a higher level than can be obtained from any other tree family. Sucrose is produced in parts of tropical Africa and Asia by evaporating water from the sap of a variety of palm species. Sap from the North American sugar maple contains a lower level of sucrose (1–3%). Its natural range includes south-eastern Canada and north-eastern USA, where maple syrup and maple sugar are produced by evaporation of water from the sap. (*See* **Sucrose**: Properties and Determination.)

Palm Species which Serve as Sucrose Sources

The palm family (Palmae) consists of over 2700 species. The coconut palm (*Cocos nucifera*) is widely distributed, but others have a limited natural range. Most palm species are trees, and in tropical areas they are second only to grasses in economic importance. Among the many food and fiber products that are obtained is palm sugar, also known as jaggery. This is unrefined sucrose, obtained by boiling water from, in some cases, tree sap, and in others the sweet juice obtained from unexpanded blossoms.

A small number of the palm species account for most of the sucrose production. The major sources and some productivity parameters are listed in Table 1. Some less utilized Asian sources of sucrose include the buri palm (*Corypha elata*; the Philippines), the nipa palm (*Nipa fructicans*; Malaysia), and the gomuti palm (*Arenga pinnata*; Malaysia and Java). In tropical Africa, sap from the oil palm (*Elaeis guineensis*) and from palms of the genus *Raphia* contains sucrose at about the 10% level.

Production and Utilization of Palm Sucrose

Potential exists for expanding the palm sugar industry in all growth areas. Palm sugar production remains largely a village industry because of the difficulty in mechanizing both the collection and the processing of the sap. Even so, the annual production of palm sucrose in India approaches 70 000 t, in Cambodia 35 000 t, and in Burma 20 000 t. Collection vessels must be effectively sterilized in order to prevent microbial contamination of the sap and conversion of sucrose to glucose and fructose, its constituent monosaccharides. This is often accomplished by elevating the pH by the addition of lime. Collapsed sap is most often filtered before lime addition to remove debris, and is then evaporated over a wood fire while being stirred in open pans.

Palm sap, which is produced at many Asian and African locations, is often consumed as such or as fermented beverages, which are sometimes distilled. In India, for example, although raw sugar would be the most valuable product from palm sap (neera), it is not usually manufactured on a large scale. Instead, most is allowed to ferment and is then consumed as a beverage called toddy. People in India, Sri Lanka, and other parts of South-east Asia have consumed both toddy and distilled toddy for centuries.

Burma has become self-sufficient in cane sugar. In central Burma, however, a large segment of the population depends on palm jaggery production for its livelihood, and about 50% of the sucrose consumed is derived from palm. It is refined to white, crystalline sucrose in order to compete with cane sugar. However, since unrefined palm jaggery possesses flavor characteristics preferred by many people, a portion should perhaps be marketed in this form. This would

Table 1 The major sucrose-producing palm species

	Sugar palm (Borassus flabellifer)	*Date palm* (Phoenix sylvestris)	*Coconut palm* (Cocos nucifera)	*Sago palm* (Caryota urens)
Where cultivated	South-east Asia	India	Throughout the tropics	India, Malaysia
Tappable years	25–100	25–90	25–35	15–20
Trees per acre	500	500	80	100
Tappable months per year	4–6	4–6	6	6
Tapping frequency	2 times per day	Once every 3 days	2–3 times per day	2–3 times per day
Sap yield (l per day)	23–15	2–5	2	10
Average sucrose content of sap (%)	12	10	13	10

be analogous to the relationship in North America between maple sugar and cane sugar, with the former being desired for certain specialty uses because of its distinctive flavor.

In many palm sugar production areas, forests are being destroyed in order to provide firewood as an energy source for evaporation of the palm sap. More efficient means for evaporation are being tested, including the use of solar energy.

Maple Trees as Sucrose Sources

Maple sugar and syrup were high-value barter items among Native Americans. The production area was then, and remains, from Maine west to Minnesota and from Quebec south to Indiana and West Virginia. This is the natural range of the sugar maple (*Acer saccharum* Marsh.), the maple species which accounts for about 75% of the production. The sugar maple and the black maple (*A. nigrum* Michx. F.) are the only two of the 13 maple species native to North America which are used to produce maple syrup. These species are favored because their sap is much sweeter (higher in sucrose) than the sap from the other species.

The early North American colonists found cane sucrose to be both expensive and difficult to obtain. As a result, they learned how to produce maple sugar. This was increasingly relied upon during the eighteenth century, when coffee and tea were being consumed in greater quantities. Later, when both cane and beet sucrose became more readily available, production of maple syrup steadily declined, especially in the USA. At the present time Canadian production exceeds $(11 \times 10^6\,\text{l})$ $(3 \times 10^6\,\text{gal})$ per year, nearly double that of the USA.

Production of Maple Syrup

Optimal flow of sap from maple trees occurs during late winter and early spring, when consecutive days with freezing temperatures at night and thawing temperatures during the day are most common. Tapholes of 95 mm diameter are drilled into the tree at a distance of 60–90 cm from the ground. The tapholes are drilled to a depth of about 8 cm, and each tree can accommodate up to four tapholes, depending on its diameter.

Traditionally, sap was collected by hanging 15-quart (14.19 l) buckets on the tap spout. This continues to be a common practice among many hobbyists, but since about 1970, commercial operations have used networks of plastic tubing to transport maple sap to evaporating plants. A germicidal pellet is often inserted into tapholes in order to inhibit microbial growth and to insure that the tubing remains clean and sterile. Much of the hard labor has been eliminated by using the tubing collection system, and production costs have been reduced by about 40%. A typical taphole yields 5–15 gallons (19–57 l) of sap each season. Forty gallons (151 l) of typical sap is required to yield a gallon (3.785 l) of maple syrup; thus an average taphole yields a quart (0.946 l) of syrup. The syrup is produced by boiling water from the sap until the solids (mostly sucrose) content has increased to 67%.

Table 2 Composition of maple syrup

Component	*Amount (%)*
Sucrose	65–66
Water	32–33
Hexoses	0–7.9
Malic acid	0.093
Citric acid	0.010
Succinic acid	0.008
Fumaric acid	0.004
Soluble ash	0.30–0.81
Insoluble ash	0.08–0.67
Calcium	0.07
Silica	0.02
Manganese	0.005
Sodium	0.003

Sucrose in Maple Sap and Syrup

Maple sap typically contains about 2% solids, of which sucrose constitutes about 97%. The remainder includes a variety of other organic compounds and some inorganic salts. The composition of a typical maple syrup is given in Table 2. Numerous other compounds have been identified in trace amounts, and some of these provide maple syrup with its unique color and flavor. These compounds are generated during the sap evaporation process, which is conducted by atmospheric boiling in open-pan evaporators, using fuel oil as the energy source. Maple sugar is produced by continuing boiling until the sucrose level exceeds 68%. Cooling this mixture results in the rapid crystallization of maple sugar. Maple sugar is heavily flavored and consists of sucrose and all the other substances which were present in the syrup.

Considerable research effort has been aimed at increasing maple syrup production. Recently, this has included the application of high-vacuum pumping to increase sap flow, and the genetic modification of the sugar maple with the aim of increasing the sucrose concentration in the sap.

See also: **Sucrose**: Properties and Determination

Further Reading

Ali R (1960) Process development on production of sugar from palm juice. *Union of Burma Journal of Science and Technology* 2: 173–180.

Doner LW and Hicks KB (1982) Lactose and the sugars of honey and maple. In: Lineback DR and Inglett GE (eds) *Food Carbohydrates*, pp. 74–112. Westport, Connecticut: AVI.

Willits CO and Hills CH (1976) *Maple Syrup Producers Manual*. US Department of Agriculture, Agriculture Handbook 134. Washington, DC: US Government Printing Office.

Refining of Sugarbeet and Sugarcane

M Donovan, Tate and Lyle Sugars, Reading, UK

Introduction

Sugars, such as sucrose, glucose, and fructose, are produced naturally in a number of fruit-bearing plants. Sucrose is only produced commercially, however, from sugarcane or sugarbeet. Sugarcane is a grass that grows in tropical regions; sugarbeet is a root crop grown in temperate climates, such as Europe and North America. Sugar refining is the process by which white granulated sugar is produced from these crops. (*See* **Fructose**; **Sucrose**: Properties and Determination.)

The production processes are in principle the same, but differ quite considerably in detail. The sugar from the two sources is virtually identical. Both processes will be described here, with separate sections where the two processes differ.

The main demand for white refined sugar has traditionally been in the developed world. Cane is grown in tropical regions, typically in Third World countries and remote from major markets. This has resulted in a two-stage process. The first-stage product, raw sugar, is a part-purified, brown, crystalline material produced in the factory. This is then transported to the refinery, which is sited close to the consumer and from which the product is typically pure crystalline white sucrose, henceforth called sugar. Sugarcane is a perishable material and must be processed almost immediately it is cut, whereas raw sugar can be stored and transported relatively easily.

Sugarbeet is grown in temperate climates, usually close to the consumer, and beet sugar-processing factories are conveniently close to the farms. These factories usually produce refined white sugar from beet without the intermediate raw sugar stage.

The principal steps in both cases are the extraction of sugar from the cane or beet by water to form an impure sugar solution, followed by purification by a number of techniques to remove suspended solids and many of the dissolved impurities. The purified sugar solution is then evaporated and crystallized to give granulated white refined sugar. These key steps are shown in **Figure 1**.

Preparation of Sugar Solution

SugarBeet Juice

Beet sugar processing is an operation in which sugarbeets are the raw material and white sugar is the primary product, and the complete process takes place in the one factory operation.

Sugarbeet processing is a seasonal operation, usually operating in Europe and North America for 3 or 4 months during the period August to February; this is the harvesting period for beet. About 7.5 tonnes of sugarbeets are required to produce 1 tonne of refined sugar, and the logistics and equipment to handle these large quantities of beets are quite complex. Even a medium-sized beet factory can process 7000 tonnes of beet a day.

The sugarbeets are transferred into the plant using a water flume system, where various devices remove most of the foreign materials such as weeds, straw, stones, and rocks. The final removal of foreign material and soil is carried out in beet washers.

Diffusion The beets are passed through a slicer, which produces long, thin strips called cossettes. These are passed into a diffuser, where sugar is continuously extracted in a stream of water. A common type of diffuser is a vertical drum. In this type the cossettes are transported upwards by a scroll, while the water passes down the drum, leaching sugar out of the cossettes as it passes through them. The temperature used in the diffuser is usually about 65 °C, and the pH is maintained at about 6.5. The juice from the diffuser is around 15% solids, and most of the sugar in the beets is extracted. Effort is made to keep the extraction level as high as possible, around 98%. The cossettes after sugar extraction are called pulp, and this is processed for animal feed.

Cane Sugar Liquor

The cane is shredded by knives and passes into a series of mills where water is used to extract sugar from the crushed cane. The sugar juice is clarified using lime to

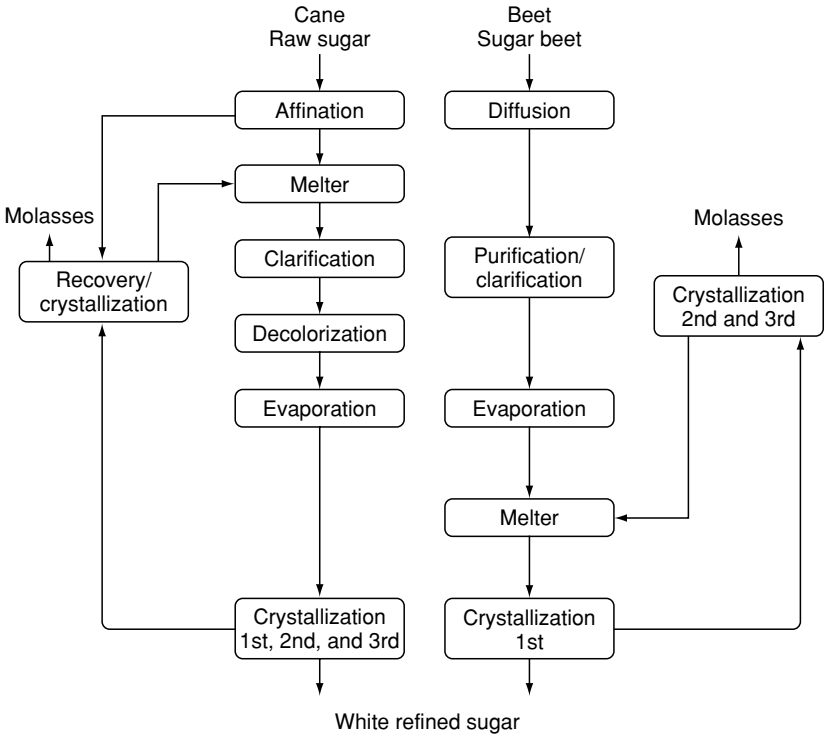

Figure 1 Process flow diagram for refining of cane and beet sugar. Reproduced from Sugar: Refining of Sugar Beet and Sugarcane, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

precipitate impurities, concentrated, and then crystallized. The product is termed raw sugar and is shipped to form the raw material for the cane refinery.

Raw sugar as received and processed in a cane refinery is really quite a pure compound, although not pure enough, and generally not prepared and transported in sufficiently hygienic conditions to be acceptable as a food ingredient. Raw sugar delivered to the UK is on average 98.0% sucrose, on a dry basis. Refined sugar, on the other hand, is 99.95% pure. At first sight the difference between the purity of raw and refined sugar would suggest a relatively straightforward process, but the impurities, including color which in weight percentage terms is very small, are quite difficult to remove. This is especially true at the throughputs required for economic, modern processing.

Affination The first step in refining raw cane sugar is affination. Raw sugar is a crystalline material, and crystals are generally quite pure. In fact, most of the impurities in raw sugar are in a syrup layer on the crystal surface. The raw sugar is mixed with a hot syrup that is at about its saturation point, in a piece of equipment called a mingler. This is a large trough with a horizontal helical screw, that mingles the raw sugar and the syrup. This mixture is then centrifuged. The most common machine is a batch basket centrifuge, centrifuging from 750 to 1500 kg in one batch, each batch being processed in about 2 min. During centrifugation a small amount of wash water is sprayed on to the cake of sugar in the basket to wash off some of the final layers of syrup. The amount is kept to a minimum to prevent dissolution of too much sugar.

This centrifugation produces two streams, a washed raw sugar, and a syrup stream containing most of the impurities from the surface of the raw sugar. This low-purity syrup is sent to a process called recovery. As the name implies, most of the sugar in this stream is recovered and sent back to the main stream for refining.

The washed raw sugar is about 99.6% sucrose, showing that affination removes the bulk of the impurities. This sugar is then sent to the melter where it is dissolved in water to about 65% solids, called 65° Brix.

Comparison of Beet Sugar Juice and Cane Sugar Liquor

The terminology used is that a low-concentration solution is called a juice, and a high-concentration solution is called a liquor. At this stage in the process,

beet sugar is in a dilute form and relatively impure. It normally contains 85% sucrose on a dry basis, called 85% purity, and is at a concentration of about 14% solids. The cane liquor on the other hand is at a higher solids, usually about 64° Brix. It has a higher purity than beet juice, that is, 99.65. This is because it has already been crystallized once. Both solutions are fairly highly colored; both solutions need to be subjected to a purification process before they can be crystallized to white sugar.

Purification of Beet Juice and Cane Liquor

Beet Juice Purification

The juice from the diffuser contains a number of impurities, both dissolved and as suspended solids, and the sugar represents about 85% of the total solids present. The purification is carried out with lime and carbon dioxide in a series of operations which increase the purity of the juice. The first step in some factories is a pre-limer, where a portion of the lime is added with control of time and temperature to precipitate much of the colloidal matter. This step is not always used, but in all cases the balance of the lime is added and then carbon dioxide is introduced through the juice to precipitate calcium carbonate crystals. Many of the impurities become absorbed on or react with the calcium carbonate, creating precipitates, and are thus removed. In the next step the calcium carbonate with the impurities is sedimented and removed by filtration. This is termed a first carbonatation stage. More carbon dioxide is then introduced through the juice to remove the remainder of the lime in a second carbonation, and the solution is filtered. A number of different types of filter can be used, but the filter leaf type is the most common. The calcium carbonate 'mud' from the two carbonatations is desweetened with water and filtered again on a rotary vacuum filter. The mud is then discarded.

Cane Liquor Purification

This is carried out in two stages: clarification to remove suspended solids and some color, followed by decolorization.

Clarification Cane liquor contains the impurities that were present in the raw sugar crystal, and the clarification step is designed to remove the suspended matter that is present. There are three processes than can be used for clarification, and two of them are also effective systems for the removal of about half of the color in the liquor. The three different systems are:

- Carbonation, combined with pressure filtration
- Phosphatation, with a clarifier using air flotation of the scum
- Pressure filtration using a filter aid

Carbonatation uses the addition of lime and then bubbling of carbon dioxide through the limed sugar liquor to precipitate calcium carbonate. In principle it is the same as beet carbonatation, but carried out at higher concentration and using less lime. The calcium carbonate crystals agglomerate into a form that is an effective filter aid and the suspended solids can be removed on pressure filters. During the precipitation of the calcium carbonate, the suspended solids, such as waxes and gums, are combined into the agglomerates and readily removed by filtration on pressure leaf filters. Some of the color also reacts with the calcium, and over 50% is removed.

Phosphatation also uses lime, but in this case the precipitate is formed with phosphoric acid to make calcium phosphate. Calcium phosphate is very difficult to filter, unlike calcium carbonate, but can be floated off as a scum. This allows the use of a clarifier, where air is introduced into the suction of a pump, and the tiny air bubbles attach themselves to the flocs of calcium phosphate, which rise to the surface. The scum can then be removed using a scraper arrangement at the top of the clarifier tank. When using a clarifier it is not possible to obtain 100% removal of solid, and a clarifier is often followed with a polish filter to prevent small quantities of solid fouling the next step, the decolorizing system.

Some refineries use pressure filtration with filter aid. This is a relatively straightforward system, but gives little color removal.

Decolorization Cane sugar liquor at this stage is actually less colored than beet sugar liquor, but because of the difference in the types of color, cane is always passed through a decolorizing step, whereas beet is only rarely decolorized.

Color can be removed from cane sugar streams by the use of carbon, or by special resins. Often a high degree of color removal is required, sometimes in excess of 80%. There are several forms of carbon decolorizer. A traditional method is to use bone char. This is a granular material prepared by grinding cattle bones, and roasting them in a kiln. Bone char is 90% hydroxyapatite, and 10% carbon.

The bone char is loaded into large vertical cylinders called cisterns. The clarified syrup is passed downwards through one, and sometimes two of these cisterns, with the second being a polish. When the char is exhausted, after perhaps 60 h, the flow of syrup to the cistern is stopped. The char is washed clear of

sugar, emptied from the cistern, and sent to the drier and then a kiln, where it is regenerated by heating to 550 °C for a few minutes with a limited quantity of air. The regenerated char is returned to a cistern and the process is repeated.

Another type of carbon decolorizer is granular carbon. This is usually manufactured from coal by kilning in the presence of steam. It contains 60% carbon and it has 10 times more decolorizing capacity than bone char. This allows the use of smaller and generally more automated plant. A granular carbon cistern may be run for 30 days; when it is exhausted it is flushed with water and the carbon is kilned at about 900 °C with a limited amount of oxygen.

A more recent method of decolorizing sugar is by the use of resins. These are in the form of small beads, and the liquor is passed through cells containing these beads. The color in sugar liquors is largely negatively charged; thus by using an anionic resin in the chloride form, color can be absorbed on to the pore walls of the resin beads. The color replaces the chloride. Once the ability of the resin to remove color has been exhausted, it can be regenerated with common salt. Two types of resin are commonly used, acrylic or styrene. A resin cell can be used for up to 60 h before needing regeneration. It is then taken offline to be desweetened, washed, treated with salt, and put back online.

Evaporation

Both liquors need to be evaporated prior to crystallization. The beet sugar juice is at 14 ° Brix and thus more water needs to be removed from beet than cane liquor.

Beet Sugar Juice Evaporation

The thin juice, as it is now called, is evaporated in a multiple-effect evaporator. This can often have as many as five effects or stages and raises the solids from about 14% to over 60%. A multiple effect is used to give economic use of steam. Live steam is used on only one of the effects, with vapor produced in that effect boiling the liquor in another effect at reduced pressure, and so on. In this way the amount of steam used is reduced by a factor of up to four. Some of the vapor is used for other heating jobs around the factory, such as heating the juice prior to carbonatation, and boiling vacuum pans used to crystallize sugar. The use of this vapor often dictates the temperatures and pressures used in the evaporator.

The product from the evaporator is about 62% solids and 89% sucrose on a solids basis, and is called thick juice. This is sent to a melter where crystal sugar from second and third boilings is dissolved in it. This is then filtered and is known as standard liquor, and is about 74% solids and 92% purity. This is the liquor that is sent to the vacuum pans to be crystallized to white refined beet sugar.

Cane Sugar Liquor Evaporation

After decolorization, the cane sugar liquor is at 60–64 ° Brix, and at a color low enough to be crystallized to white granulated sugar. However, it first needs to be evaporated at about 75 ° Brix for crystallization. The traditional evaporator used in the sugar industry has been a calandria type, i.e., a vessel with a submerged bundle of tubes. More recently both falling-film and plate-type evaporators have been used effectively. Cane liquor evaporators are usually run as double- or triple-effect, and vapor from the evaporator can be used to provide the heat for the melter, where the sugar is dissolved in water.

Crystallization

The evaporated sugar liquor is sent for crystallization, although some cane liquor can be sold in this form as liquid sugar, as it is very pure. (*See* **Crystallization**: Basic Principles.)

Crystallization is essentially the same for beet and cane. The purity of the syrup used to crystallize beet sugar is lower and this, combined, with the use of lower pressure steam, means that beet crystallizations are slower than cane crystallizations. In the sugar industry the term used for crystallization is boiling.

Sugar is crystallized in batch vessels called vacuum pans. These can crystallize up to 70 t of sugar in one batch, and a batch cycle can take from 2 to 4 h. In vacuum crystallization the sugar liquor is boiled under a vacuum until it just exceeds its saturation point. It is then supersaturated, and crystals will grow provided this supersaturation is maintained by continual boiling of the liquor. In order to have some control over the size of the crystals, the liquor is usually seeded at a predetermined supersaturation with a small amount of milled sugar crystals in alcohol. These crystals are about 10 μm in size, and 0.5–1 l is generally added. As these crystals grow, the deposition of sugar on them from the solution will reduce the supersaturation. However, by heating and evaporating the solution, water is removed and the supersaturation is carefully controlled. If this supersaturation becomes too high, nucleation can occur, and the extra crystals formed will cause the size of the final sugar to be too small. If the supersaturation is allowed to become too low, crystal growth will slow down or stop. Maintaining the growth of crystals at the correct rate, and obtaining a batch of sugar at the correct crystal size, and size dispersion has been the

skill of the pansman. Nowadays, modern measurement techniques, and computer control can boil sugar pans as well as a skilled pansman.

The key features of a vacuum pan are shown in Figure 2. The pan has a heating surface consisting of vertical tubes or plates called a calandria, and is heated by steam. Most pans have an agitator to assist in circulation of the massecuite (see below). A vacuum is maintained in the pan during the boiling to keep the sugar in the temperature range 65–85 °C. This is to prevent color formation during the crystallization. The yield of sugar is in the range 50–60%, and the final mixture of crystals and concentrated sugar liquor is called massecuite, or masse. This masse is extremely viscous, and at the end of the crystallization period the vacuum is released and the masse is allowed to drop by opening a large valve at the bottom of the vaccum pan. It is dropped into a receiving vessel, and from there it is fed to centrifuges to separate the sugar crystals from the syrup.

Continuous pans have been developed for white refined sugar, and a few are used. However, difficulties in obtaining a narrow size distribution, cost, and some other problems mean that the vast majority of refined sugar is boiled in batch pans.

At this point there is a difference between beet and cane. The high purity of the cane syrup at this point means that the mother liquor from the first boiling can be used again for another crop of white refined sugar. In fact a total of three or four batches, including the first batch described above, can be boiled. With beet, only one batch can be boiled and used directly for refined sugar. The subsequent batches, as described below, are dissolved and put back into the process.

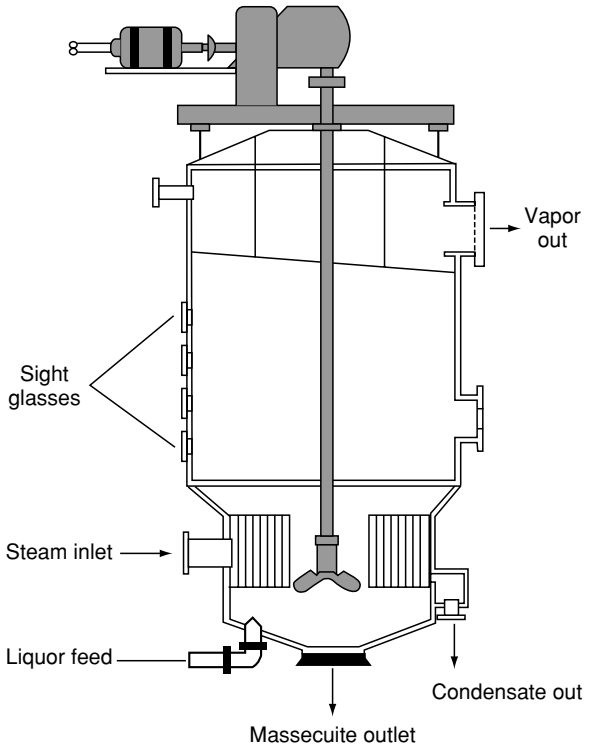

Figure 2 Vacuum pan for crystallizing sugar.

The steam used to heat a beet vacuum pan is generated in one of the effects of the evaporator, and is often around 100 °C. In recent years a lot of effort and capital have been spent in factories to reuse the steam vapors generated in a number of applications around the plant, and pans are now being boiled with steam or vapor at quite low temperatures. Pressures less than 1 bar abs and less than 100 °C are quite commonly used to boil beet sugar. This has required pans to have a larger heat-transfer surface and to have a lower hydrostatic head of liquor above the calandria.

Vacuum pans for boiling cane sugar take steam from the boiler or steam turbine backpressure at pressures of 1–2 bar gage, and sometimes up to 4 bar gage.

Centrifugation

White sugar is separated from mother syrup in batch centrifuges. The massecuite is fed into a basket which is then accelerated up to speeds in excess of 1000 rpm, where it forms a wall of sugar that is 150–200 mm thick. The sugar is then washed with a spray of hot water, which dissolves a small amount of sugar from the surface of the crystals, but assists in removing most of the impurities such as color, ash, and invert from the surface of the sugar crystals. The first two need to be removed to reach purity specifications, and removal of invert is important to prevent the sugar from caking in storage. The centrifuge is then slowed down, and a plough is inserted into the cake, which discharges from the bottom of the centrifuge.

Drying

The moisture content of the sugar crystals from the centrifuge is quite low, at about 1%, but this moisture is concentrated on the surface of the crystals. It is removed in a rotary drier, where the sugar is gently tumbled in a stream of warm air for a residence time of about 20 min. The most common type of sugar drier is an inclined rotary drum, fitted with blades or paddles to lift the sugar. The air flow can be either cocurrent or concurrent. The final moisture is extremely low, at 0.02%. This sugar is scalped to remove oversized lumps, usually agglomerates of

crystals that have built up. (*See* **Drying**: Theory of Air-drying.)

Other Boilings

Beet Sugar

In a beet factory white sugar is only produced from the first crystallization. A second and third crystallization are carried out; the crystal sugar from both of these boilings is sent to the melter to be mixed with thick juice (as mentioned earlier). The crystals are not pure enough to be sold as white sugar, but are considerably purer than thick juice, and thus raise the purity of this to provide standard liquor which can be crystallized to produce white sugar. A variety of names are used for these boilings, such as raw and after product or high and low raw.

The raw boiling is carried out with the mother liquor from the white sugar boiling, i.e., the liquor separated from the white sugar crystals in the centrifuge. The afterproduct boiling is the mother liquor of the raw boiling, although the liquor that is seeded is usually of a higher purity. This is because it can be difficult to get small seed to grow at low purity. The masse from this afterproduct boiling is very viscous, but it is still possible to crystallize more sugar from it. The complete masse is sent to a cooling crystallizer where it is cooled over a period of 48 h. This long period of time extracts as much of the sugar as is economically possible. This masse is centrifuged and the liquor is molasses.

There is still 60% sugar in beet molasses, but owing to the solubility of sugar it is not possible to separate any more by crystallization. New large-scale chromatographic techniques have been developed to separate sugar from molasses; these are not used extensively in Europe, although there are a number of operations in the USA.

Cane Sugar Recovery

The recovery process consists of a number of crystallization and centrifuging steps, and is designed to extract as much sugar as possible from low-purity syrups. The final products are molasses and a low-purity sugar, the latter being remelted and returned to the main stream of the refining process, following affination. In cane processing there are two feed streams of low-purity syrup; the main stream is from affination where the syrup from the surface of the raw sugar is removed, and the other stream is from the white crystallization process, where the syrup that, after three or four boilings, is too impure for extraction of white sugar is sent to recovery. The recovery process is a series of crystallizations, usually three, and each with a yield of 40–55%. As these crystallizations are carried out at low purity, they can be slow, and quite difficult. There are a number of ways of carrying out these boilings and the schemes can become quite complex. In one scheme the crystals grown in the third boiling are used as the seed for the second boiling, and the crystals from this boiling are used as the seed for the first. However, there are almost as many schemes as there are refineries.

Brown sugar can be boiled from the first, or purest of the recovery boilings. Adjustments are made to the purity and color of the syrup in order to obtain the appropriate flavor and color.

Beet and cane molasses are sold for animal feed or fermentation. The beet sugar industry blends a portion of its molasses with the dried pulp, and makes pellets for animal feed.

Storage

The drier produces a very dry sugar, with a moisture level of about 0.02%. Nevertheless, even at this low level there can be problems in storage, and if sugar has to be stored for an extended period of time then it is usually conditioned. This is carried out to prevent the sugar caking. Conditioning is essentially a holding period of perhaps 24 h in which any moisture released from the crystal can pass into the surrounding air, and be blown away from the sugar. It is usually carried out in a silo. In some cases dehumidified air is passed through the sugar in the silo, and in others the sugar is passed into the silo, and out again in perhaps a 24-h period; the movement of the sugar allows moisture to escape without the crystals sticking together and caking. The reason that even such a small amount of residual moisture can cause a problem is that it is all concentrated on the surface of the crystal. The water is in the form of a saturated syrup, and as this syrup crystallizes it releases moisture. If the surrounding air is not kept moving, this moisture can humidify the air to such an extent that any cooling can cause it to condense. This condensed moisture can dissolve the surface of the crystals and allow them to stick together. (*See* **Storage Stability**: Parameters Affecting Storage Stability.)

Packaging

White sugar is quite free-flowing and packaging in the standard 1-kg pack can be carried out at very high speeds in modern equipment. This equipment takes preprinted rolls of paper, folds and glues this paper to form bags, fills and seals them. Brown sugars are more difficult to package, owing to their sticky nature, and are packaged at much lower speeds. Bulk deliveries

are typically made in tankers, 1-tonne containers and 25-kg or 50-kg bags. (*See* **Packaging**: Packaging of Solids.)

Analysis

Analysis is required of intermediate products to ensure that the process is running correctly, and of final products to ensure that they meet customers' specifications.

During the process a close watch is kept on concentration of sugar. This is measured as Brix, which is weight of sugar solids per weight of syrup, and in the laboratory is measured by a refractometer. Color is also closely monitored, measured by an absorbance of light at 420 nm in a spectrophotometer. The ash content is monitored in cane, and in beet the purity is followed. Purity is the amount of sucrose per unit dry weight of substance. In both cane and beet the pH is also measured at key points of the process. Too high or low a pH leads to degradation of sucrose to color byproducts or invert.

The product may need several other measurements to ensure quality and conformance to customer requirements. The color and ash measurements are routine requirements. For liquid sugar a Brix value would be required. Others include a sieve analysis to measure both the size and size distribution of the sugar crystals, expressed as mean aperture (MA) and coefficient of variation (CV). A test is also carried out to measure suspended solids, which in white sugar are extremely low.

See also: **Crystallization**: Basic Principles; **Drying**: Theory of Air-drying; **Fructose**; **Packaging**: Packaging of Solids; **Storage Stability**: Parameters Affecting Storage Stability; **Sucrose**: Properties and Determination

Further Reading

Abram JC and Ramage JT (1979) In: Birch GG and Parker KJ (eds) *Sugar: Science and Technology*, pp. 49–96. London: Applied Science.

McGinnis RA (1971) *Beet-Sugar Technology*, 2nd edn. Fort Collins, CO: Beet Sugar Development Foundation.

Plews RW (1970) *Analytical Methods Used in Sugar Refining*. London: Elsevier.

Pennington NL (ed.) (1990) *Sugar; A User's Guide*, pp. 1–35. New York: Van Nostrand Reinhold.

SUGAR ALCOHOLS

A Rapaille and J Goosens, Eridania Béghin-Say, Brussels, Belgium
M Heume, Eridania Béghin-Say, Krefeld, Germany

Introduction

Sugar alcohols are white crystalline, non- to slightly hygroscopic powders and clear syrups. They are generally stable to heat, and do not take part in Maillard-type browning reactions. Chemically, sugar alcohols are hydrogenated mono-, di-, and oligosaccharides. The more important compounds in these groups have the structures shown in **Figure 1**. and their main physical properties are summarized in **Table 1**. Erythritol is discribed separataly further in the text.

Sensory and Nutritional Properties

Compared to sucrose, sugar alcohols have slightly reduced sweetness values, as shown in **Table 2**. Sugar alcohols produce a characteristic mouth-cooling effect which is directly related to their heats of solution and their solubility in water (**Table 1**). Sugar alcohols, also known as polyols, may give particular advantages with regard to diet control and dental health (**Table 3**). The benefits that polyols can give in reducing caloric intake is due to their different metabolic behavior after oral ingestion compared to other carbohydrates. Because polyols may have a mono-, di-, or oligosaccharide carbohydrate structure, small differences can occur with regard to their metabolic fate in the human body.

Except for maltitol-containing syrups (55–75% maltitol), in which a small fraction of oligomeric structures occurs, polyols are not degraded by saliva enzymes (amylases) upon oral ingestion. In the stomach slight acid hydrolysis of the di- and oligosaccharide polyols may occur. However, once the small intestine is reached, extensive hydrolysis by α-amylases of the mucosa occurs and up to 60–80% of the maltitol and isomalt is hydrolyzed. A 50% glucose/50% sorbitol or a 50% glucose/25% sorbitol/25% mannitol mixture respectively is obtained. The other disaccharide polyol lactitol is almost unchanged because the required β-galactosidase is only present in trace amounts. Sorbitol, mannitol, and

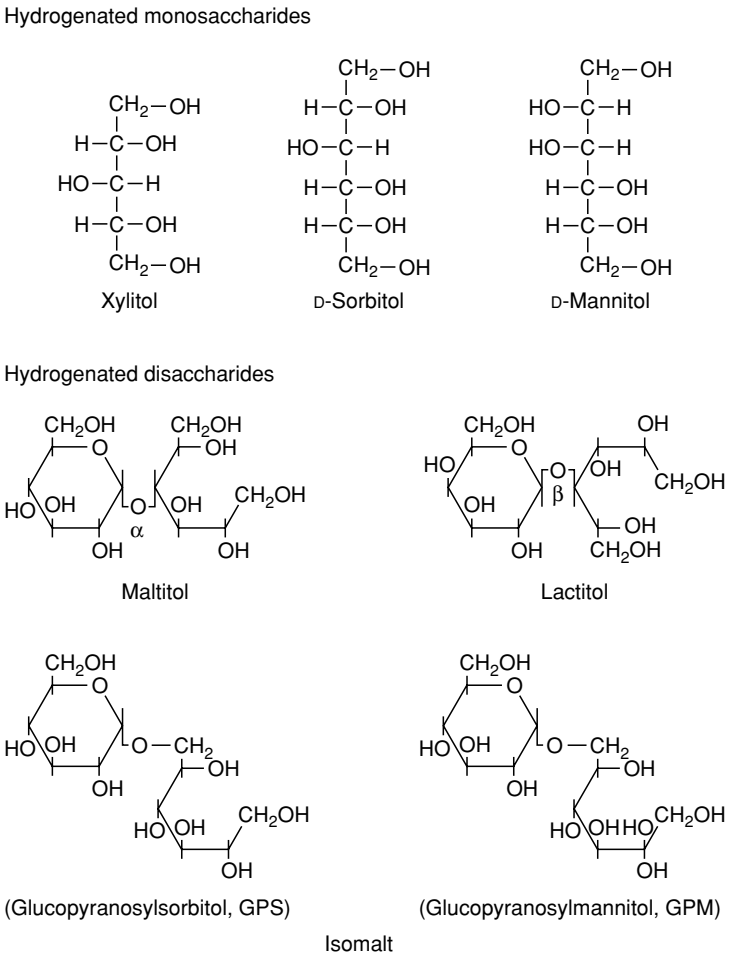

Figure 1 Structures of the most important sugar alcohols. Reproduced from sugar alcohols, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 1 Physical properties of sugar alcohols

	Xylitol	*Sorbitol*	*Mannitol*	*Maltitol*	*Lactitol*	*Isomalt*
M	152	182	182	344	344	344
$[\alpha]b^{20}$ (°) [c = 0.1; H_2O]	0[a]	−2	−0.49	+106.5	+11.5	−[b]
Melting point (°C)	92–96	93–97	165–168	148–152	94–97	145–150
Solubility: (g in 100 g water at 20 °C)	169	235	18	160	125	34
Heat of solution (kJ g^{-1})	−153	−111	−121	−69	−53	−39
Viscosity (mPa, 20 °C)	60%, 21	70%, 180	–	75%, 2500	60%, 68	30%, 4

[a]Meso compound.
[b]Varying composition.

xylitol reach the small intestine unchanged. (*See* **Isomalt**.)

In the small intestine the monosaccharide polyols are partly absorbed into the blood stream by a passive diffusion process at rates depending upon their molecular weight (Table 3). Also, the applied dose and the actual intestinal transit time affect the rate of absorption. The absorbed fraction of sorbitol and xylitol is channeled into the normal liver metabolism and contributes energy to the same extent as glucose.

The unabsorbed polyols finally reach the large intestine where the microflora ferments them almost completely to volatile fatty acids (VFAs). From the energy contained in the fermented fraction, 20% is used by the bacteria for growth. The remainder is absorbed from the gut as VFAs which are then

Table 2 Sweetness values of sugar alcohols

Compound	*Relative sweetness*
Polyol[a]	
Xylitol	80–100
Sorbitol	50–60
Mannitol	50–60
Maltitol	80–90
Lactitol	30–40
Isomalt	40–50
Erythritol	60–70
Sucrose[a]	100

[a]10% in H_2O: 20 °C.

Table 3 Absorption characteristics of sugar alcohols

Polyol	*Absorbed from the small intestine (%)*	*Excreted (%)*		*Fermented (%)*	*Calorific value (kcal g^{-1})*
		Urine	*Feces*		
Xylitol	25–40	0	< 1	60–75	2.9
Sorbitol	15–20	0	< 1	80–85	2.6
Mannitol	15–20	15–20	< 1	80–85	1.9
Maltitol	45–60	0	< 1	40–55	3.2
Lactitol	< 1	0	< 1	99	2.3
Isomalt	35–45	0	< 1	55–65	3.0

transported in the blood to be metabolized by the liver into fat constituents. In total, about 50% of the energy of the fermented polyols is available to the human body.

Based on the above metabolic behavior, the calorific value of polyols can approximately be calculated (Table 3). However, depending upon the methodologies used, values vary over a wide range. Therefore, and because many polyols are used in combination with each other, a European Union Directive has laid down an average calorific value of 2,4 kcal g^{-4} for all polyols.

Because of the reduced digestibility of polyols, their intake level is recommended to be restricted to 40–50 g day^{-1} (adults) and 30 g day^{-1} for children in order to avoid gastrointestinal discomfort. Higher doses may cause an osmotic diarrhea or increased flatulence due to slow absorption and extensive fermentation, respectively, but tolerance levels improve upon adaptation. On the other hand polyols are very suitable for diabetics, because rapid glycemic or insulinemic response is avoided. For the same reason sorbitol and xylitol are often applied in parenteral nutrition to provide a more controllable carbohydrate energy source compared to glucose or fructose.

Concerning dental health, polyols have long been known to have no adverse effect on teeth. This is due to the fact that mouth bacteria are unable to ferment polyols, thereby avoiding the production of enamel-solubilizing acids. With the intraplaque pH telemetry test, it was shown that eating polyol-based sweets did not decrease the plaque pH below the critical value of 5.7. Additionally, for xylitol, there is ample proof of a cariostatic effect. In several studies, of which the Turku study (Finland) is most famous, the incorporation of xylitol in a normal diet was proved to reduce the occurrence of dental carics. Most probably, xylitol selectively discriminates against the growth of *Streptococcus mulans*, the major organism involved in dental caries. (*See* **Dental Disease**: Role of Diet.)

Occurrence and Methods of Production

Polyols like sorbitol, mannitol, and xylitol have been well known over the last 30 years due to their occurrence in a number of fruits, vegetables, and cereals. Although they occur in nature, their concentration in plant material is too low for a simple, economic commercial extraction, as practiced with sucrose. A more complicated process is therefore needed for the production of sugar alcohols.

Several techniques have been described to produce polyols but only a few of them are of industrial importance. These are mainly based on catalytic hydrogenation of easily accessible carbohydrates, e.g., dextrose, glucose syrups, maltose syrups, invert sugar, isomaltulose, xylose, and lactose. These carbohydrates are obtained either by hydrolysis or isomerization of natural raw materials like starch, sucrose, milk (whey), and straw (Figure 2).

The carbohydrate raw material is hydrogenated in an aqueous solution at elevated temperatures (120–190 °C) and hydrogen pressure (20–200 bar) using an activated nickel catalyst. Hydrogenation can be performed either in a batch or a continuous process, and results in the addition of hydrogen to the carbohydrate precursor. Hydrogen pressure, temperature, pH value, and type of catalyst are the most important parameters influencing reaction time and selectivity. They have to be well adjusted to obtain a high-purity product in high yield. Purity of the raw material is also of importance in this respect. The hydrogenated carbohydrate (polyol-containing) solution is subsequently refined by filtration, carbon, and/or ion exchange resin treatment to remove salts as well as colored byproducts formed during hydrogenation.

Sorbitol and maltitol are available as aqncous solutions and also as crystalline material. Mannitol, xylitol, lactitol, and isomalt are available as crystalline material. All crystalline polyols are produced by crystallization, solidification, or spray-drying of an aqueous solution at an appropriate concentration.

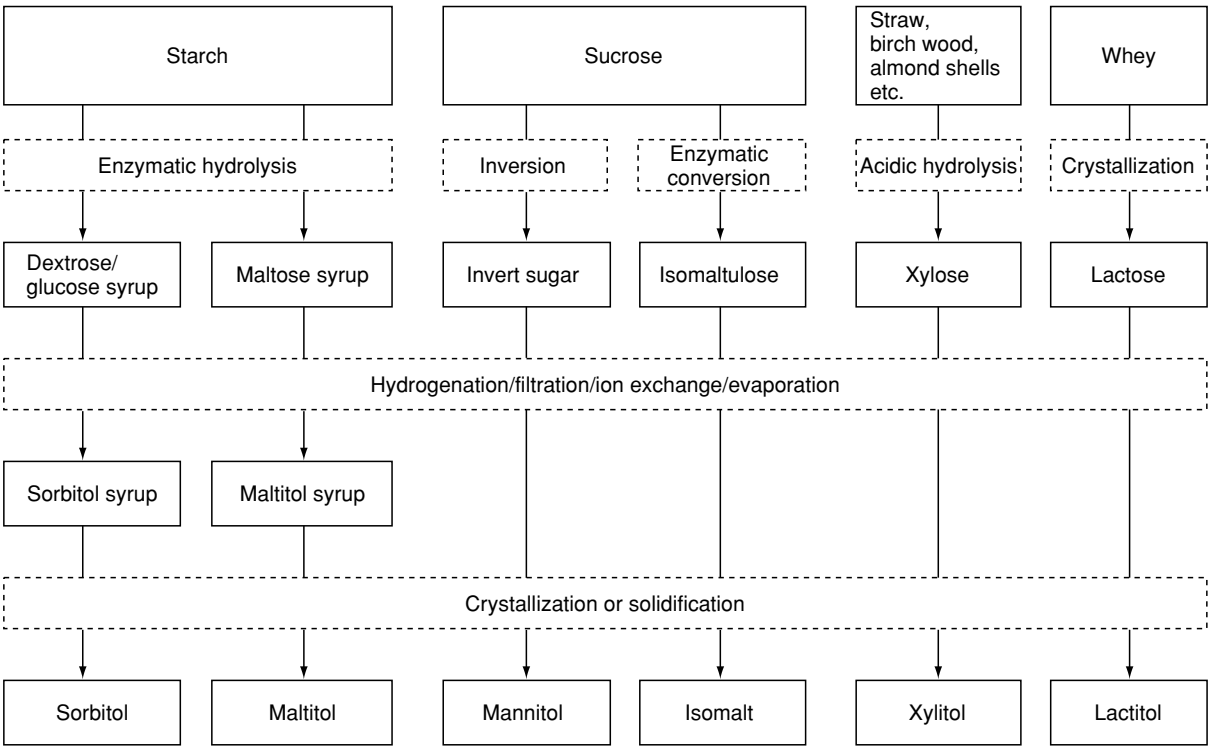

Figure 2 Commercial production of sugar alcohols. Reproduced from Sugar Alcohols, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Uses in Foods

Substitution of sugars in food to achieve specific nutritional objectives is not a new practice. Initially, polyols like sorbitol and mannitol were used to sweeten food for diabetics. To cope with sensorial and technological requirements for noncariogenic food, the range of polyols has been extended to include products such as maltitol, xylitol, isomalt, and lactitol.

For the production of polyol-based foods, most conventional types of food-manufacturing equipment are used. However, manufacturing conditions (temperature, vacuum, etc.) very often need to be adapted. As some polyols are moderately hygroscopic, the relative humidity of air during production and storage conditions require careful control. It is also necessary to assure good protection by careful wrapping and/or packaging. When neccessary, sweetness of polyol-based foods can be enhanced by adding the neccessary amounts of intense sweetners like aspartame, accsulphame K, cyclamate, and saccharin. (*See* **Sweeteners: Intensive.**)

In addition to the production of sugarless foods, polyols, e.g., sorbitol, can also be used in conventionally sweetened food products in order to give specific sensorial effects such as texture and reduced color. There are also technological advantages such as plasticity, workability, and shelf-life. In particular, sorbitol is widely used for its humectancy power to extend food products shelf-life.

The main uses of polyols in food products are outlined below.

Sugar-Free Confectionery

This is the most important application field for polyols.

Hard Confectionery

Hard confectionery is produced by boiling polyol solutions or syrups to a very high dry-matter content (98–99.4%) followed by cooling, molding, and wrapping. The boiling temperature for such polyol confectionery varies between 155 and 170 °C depending on the type of polyol. Severe vacuum conditions are needed. With the exception of mannitol, all the polyols can be used alone. However, for technical and economical reasons very often combinations of polyols are applied. It should be kept in mind that the solubility of polyols strongly influences crystallization and graining.

Soft and Chewable Confectionery

The main component in this application is a polyol or a polyol blend (65–70% dry matter) which is

presolubilized and cooked together with vegetable fat (about 6%), emulsifier (about 0.8%), gelatin (about 1.4%), and other minor components, e.g., food-grade acid, flavor, and color. After cooling, the mass is formed and wrapped. Chewability of the sweets can be obtained by acration of the cooked mass after partial cooling using conventional acration techniques such as pulling and continuous foaming.

Gums and Pastilles

Maltitol syrups, alone or in combinations with either lactitol or isomalt, are very suitable for the production of gelatin gums. For some compositions it is necessary to adapt the gelatin type and level as well as the cooking conditions in order to obtain gum properties equivalent to those obtained with conventional sweetened gums.

In pastilles a gum arabic/polyol ratio of 50/50 on a dry-weight basis is advised. This composition is cooked to 70–72% dry-matter content, molded in starch, and dried to 90% dry matter. After demolding, the gums are waxed. Maltitol syrups give good pastilles with optimal shelf-life properties.

Tubletting

Although all powdered polyols can be compressed on rotary machines to give tablets, sorbitol is the most common polyol used in this field.

Powdered sorbitol has excellent flow properties for tabletting. Good flowability is very important to provide a uniform tablet weight. Compared to other directly compressible materials, powdered sorbitol gives a high increase of tensile strength as a function of compression force and is therefore very suitable for direct compression. Due to their rather high hygroscopicity, sorbitol-based tablets should be protected or stored under relatively dry conditions in order to avoid softening due to moisture pick-up. The priability of sorbitol tablets, even those compressed at low pressure, is extremely low and is improved further by storage at close to equilibrium relative humidity (ERH). The overall performance of sorbitol powder in tabletting depends upon different factors, such as particle size and distribution, crystal morphology, and compression conditions.

Chewing Gum

The main components of sugar-free chewing gums are gum base (25–30%), polyol powder as solid phase (45–63%), and sorbitol or maltitol syrups as liquid phase (10–30%). The solid phase can be sorbitol powder, but more often a combination of sorbitol and xylitol is used. The main advantages of xylitol are its excellent cariostatic propertics, high sweetness, and a pleasant cool taste due to its negative heat of solution. Addition of a small quantity of glycerol improves the plastifying properties; mannitol addition prevents stickiness.

Chocolate

Total replacement of sucrose by polyol powder both in bitter or milk chocolate is possible providing the powder has an optimum purity, crystal morphology and granulometry. A bitter chocolate consists of 40–50% cocoa mass, 40–50% polyol powder, and 5–10% cocoa butter.

Originally only sorbitol was used for sugar-free chocolate. Recent trends are towards the use of maltitol, isomalt, and lactitol.

Dragee Coating

Coating consists in covering cores, e.g., chewing gum with a syrup which is close to its saturation point, then inducing crystallization by evaporation. Generally a lot of layers (40–50) are needed. The temperature of polyol syrups, drying air, centers, and core bed are very important processing parameters. Sorbitol is the most common sweetener for sugarless coating, but maltitol, xylitol, isomalt, and lactitol can also be used.

Bakery Products

Polyols can be used in bakery products due to their specific properties which are lacking in many other bulk sweeteners. The sugar alcohols contribute to softness and storage capacity. They also regulate moisture, taste, and sweetness. A combination of polyols or a polyol combined with another sweetener is preferred. If browning is too light, the baking temperature can be raised or a small amount of fructose can be added.

The baking industry uses crystalline sorbitol and sorbitol solution in products for special dietary use. It is also employed as a humectant to extend the shelf-life of baked goods. The stability of baked products can be improved by partial substitution of sucrose by sorbitol, e.g., in cakes. Sorbitol is used in sponges and cakes at between 5 and 8% of flour weight. In biscuits and cream fillings it is used up to 10%. The freshness of gingerbread is improved by adding only 5% sorbitol. It is also used to substitute sugar in classical formulations for low-sugar products. (*See* **Cakes**: Nature of Cakes.)

The use of the other polyols in bakery is rapidly developing. The level of sugar replacement with these sweeteners depends on the type of polyol and the application objective. While, for some polyols, rather

high usage levels can be achieved, e.g., for maltitol, lactitol, and isomalt, for others, e.g., xylitol and mannitol, usage levels are generally low. These differences are also related to the specific sensory and physicochemical properties of the polyols used. An important contribution of the other polyols is, as for sorbitol, the humectant properties which control the moisture level in bakery products when stored for long periods. In some cases, combinations of polyols can give interesting complementary effects. The main applications in the bakery field are cakes, biscuits, pastry, fillings, and icings. In the last two products, xylitol and mannitol can be an asset due to their perceptible cooling effect. (*See* **Biscuits, Cookies, and Crackers**: Nature of the Products; **Pastry Products**: Types and Production.)

Other Areas of Application

Polyols can also be used in sugarless jams and ice-creams. Sorbitol-containing jams and preserves have a clear taste and good texture. A sufficiently high solids content can be obtained in order to inhibit microbiological activity, thereby protecting the jams against mold development. In order to avoid crystallization during storage, the solubility of the polyol(s) should be checked before formulating the food.

Erythritol

Chemical and Physical Properties

Erythritol is a linear carbohydrate molecule of four carbon atoms, each carrying one hydroxyl group. The molecular structure is shown in Figure 3.

Chemically, erythritol therefore belongs to the class of monosaccharide polyols like sorbitol, mannitol, xylitol, and glycerol. Erythritol is also a symmetrical molecule and therefore it exists in only one form, the mesoform. It forms anhydrous crystals with a moderately sweet taste without off-taste or odors. The powder has a transparent white brilliant appearance and dissolves in water to give a colorless nonviscous solution. Crystals melt at 122 °C to form a colorless and brilliant nonviscous melt.

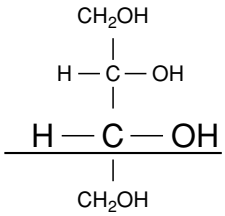

1,2,3,4-butanetetrol

Figure 3 Molecular structure and chemical formula of erythritol.

Erythritol's chemical properties are similar to those of other polyols in that it has no reducing end groups and thus has excellent heat- and acid-stability. It differs in having a low solubility (37% ww at 25 °C), a property which among these simple polyols it shares with mannitol and isomalt. Heat of solution is very low ($-181\,J\,g^{-1}$), a property shared with xylitol. However, compared to the group of polyols presently used as sugar-replacers, erythritol has the lowest molecular weight (122) which of course provides different properties, such as higher osmotic pressure and lower water activity in solution. It has also a high freezing-point depression and boiling-point elevation.

Compared to sucrose, erythritol has a sweetness value of 60–70%. It has a clean, sweet taste, similar to sucrose, a strong cooling effect, and shows many synergies with intense sweeteners.

Occurrence and Methods of Production

Erythritol occurs in several foods at levels up to 0.13%. It is also present in human and animal tissue and body fluids.

Briefly, erythritol is produced from wheat or corn starch by enzymatic hydrolysis yielding glucose, which is fermented by a safe and suitable yeast-like fungi, *Moniliella pollinis*. Once erythritol is separated from the sterilized fermentation broth, it is purified by ion exchange resin, active carbon, ultrafiltration, and crystallization. The final crystalline product, which looks very similar to sugar, is more than 99.5% pure.

Uses in Foods

The main uses of erythritol in food products are outlined below.

Tabletop Sweeteners

In tabletop applications, erythritol will be used at levels up to 99.9% as a noncaloric, noncariogenic carrier for intense sweeteners. In these applications, the sensorial profile-modifying properties of erythritol are of great importance, resulting in sweetness synergy, improved mouth feel, and masking of off-flavors. In addition, due to erythritol's crystalline structure and nonhygroscopic property, it offers excellent flowability and stability as carrier.

Beverages

The synergies which erythritol shows when used in combination with intense sweeteners are also very useful in low-calorie beverages. Also, at use-levels of 3.5% or less, erythritol can mask certain off-flavors in beverages like coffee, tea, and grapefruit juice.

Chewing gum

Good-quality noncaloric and noncariogenic chewing gum can be formulated by using solely erythritol (use-level up to 60%).

Most of the sugar-free chewing gum produced today has a tendency to become brittle and harder during aging. This can be avoided and a sugar-free chewing gum with softer texture, greater flexibility, and improved shelf-life can be obtained by partial replacement of the other sugar alcohol(s) using erythritol.

Chocolate

The use of erythritol in chocolate compositions allows a dry conching process at high temperatures. Due to the good heat-stability and low-moisture pick-up of erythritol, it is even possible to work at higher temperatures than traditionally used. This results in enhanced flavor development.

Erythritol-based chocolate is not only easy to produce, it can also be considered a true 'reduced-calorie' product.

Candies

Sugar-free fudge with texture and shelf-life properties equivalent to conventionally sweetened fudge can be produced using erythritol (up to 40%) in combination with maltitol syrup (75% maltitol) to control crystallization.

Depending on the cooking temperature and seeding level with erythritol, texture can be varied from soft to hard. In hard candies, erythritol in its crystalline form is used up to 50% for its cooling effect as sherbet filling in die-formed hard candies, or 'sandwiched' in between two layers in deposited hard candies.

Fondant

Using erythritol, it is possible to obtain sugar-free low-calorie noncariogenic fondant with identical technical properties to classical products. Previously this was impossible with the other low-calorie bulking agents currently available. Pure erythritol induces too high a crystallization, but 60% erythritol in combination with maltitol syrup as a liquid phase helps to control this phenomenon.

Lozenges

Research has shown that lactitol, crystalline maltitol, and erythritol are the only suitable materials to produce a lozenge with a lower calorie content than the conventional sugar-based lozenges, but which are otherwise identical. In addition to the lower calorie content and pronounced cooling effect, erythritol has the added advantage over lactitol and crystalline maltitol of requiring a shorter drying time to obtain the required texture at low residual moisture content. Lozenges based on solely erythritol (up to 99%) have excellent shelf-life properties, even when stored under high humidity conditions.

Bakery (Pastry) Products

In fat cream, a very good texture is obtained with erythritol at a use-level of 60% and a particle size distribution below 300 μm. The manufacturing process and shelf-life properties of the erythritol-based fat cream are similar to those for conventional fat/sucrose compositions. Erythritol has an effect of masking the fatty mouth feel, giving a more refreshing and attractive product. In addition, the calorie content of the composition is also reduced.

Reduction of calories in bakery products is particularly difficult if only the sugar moiety is replaced. Erythritol can reduce the caloric content by more than 30% (dependent on the application) without introducing undesirable side-effects. Additionally, erythritol improves the shelf-life of bakery products. Erythritol can be applied successfully in cookies, biscuits, and cakes, where it improves baking stability and shelf-life at an addition level of about 7%.

See also: **Biscuits, Cookies, and Crackers**: Nature of the Products; **Cakes**: Nature of Cakes; **Dental Disease**: Role of Diet; **Isomalt**; **Pastry Products**: Types and Production; **Sweeteners**: Intensive

Further Reading

Tinfeld T and Mühlemann HR (1978) Cariogenicity and acidogenicity of food, confectionery and beverages. *Pharmacology and Therupeutics in Dentistry* 3: 53–68.

Sheimin A and Maekinen KK (1975) Turku sugar studies-I–XXI. *Acta Odontologica Scandinavica* 33 (suppl. 70): 1350.

SUNFLOWER OIL

F J Sánchez-Muniz and C Cuesta, Universidad Complutense de Madrid, Madrid, Spain

Origin and Systematics

The cultivated sunflower (*Helianthus annus* L.) is a plant of the family Asteraceae (Compositae). The term *Helianthus* is derived from the Greek *helios*, meaning sun, and *anthos*, meaning flower. The property of this flower 'turning with the sun' is clearly expressed by the Spanish and French languages when naming this plant.

Hypothesis suggests that the wild *H. annus* of the western USA became a camp-following weed of North American Indians. In time it was introduced into the central part of the country where it was domesticated for use primarily as food. Archeological evidence indicates the occurrence of heads and achenes of domesticated types of sunflower as early as the first millennium BC, and it has been suggested that cultivation began as early as 3000 BC. During the early sixteenth century European visitors to America introduced the sunflower to Europe, principally to Spain, from where it spread throughout the continent. It was grown initially as an ornamental plant and later for food and medical purposes. The sunflower was developed as an important source of oil during the early 1800s in Russia. More recently, wild species have been used in breeding programs. They have been found useful in improving disease and insect resistance of cultivars. The success of breeders (most from the Former Soviet Union) in improving the oil content of seeds from less than 30% to over 50% was a major factor in the development of the sunflower as a major world oilseed crop.

The plant is generally grown to obtain oil, but meal and hulls are also used or present potential uses (see below).

World Sunflower Crop

Sunflower ranks among the most important annual field crops grown in the world for production of edible oil. During the last decade, the total production of sunflower has increased. Thus, in 1989–91 it was estimated at about 22.3 million metric tons. In 1998–99 world production was about 26.7 million tons (Table 1).

Sunflower grows well in most regions of the world, with significant production occurring in each of the six crop-producing areas (Figure 1). Europe and America (North, Central, and South America) are the leading producers of sunflower seed, accounting for about 50% of the world's total production. In Figure 1 production efficiency (kg ha^{-1}) in the six world crop-producing areas is also presented. Recent data show that the ex-USSR and Argentina are the largest producers (Table 1). Other leading producers are also presented in Table 1. However, because production depends on cultivars and in many countries efficiency is determined by the weather, the leading order can change. For example, in Spain, one of the leading producers, a high percentage of sunflower comes from dry-ground cultivars, thus sunflower seed production may change dramatically from a wet year to a dry year (by about half a million tons).

Although sunflower is grown primarily for its oil, there is limited production of certain cultivars grown for nonoilseed or confection purposes.

Table 1 World sunflower crop and sunflower oil production (Million Tons)

Country	*1998*	*1999*	*2000*	*2001*	*Estimated 2002 (as available at 6th of Sep 2002)*
Ex-USSR	5.76	7.39	7.88	5.4	6.97
Central Europe	2.8	3.3	1.9	1.9	2.4
Argentina	7.2	5.7	3.1	3.7	3.9
EU-15	3.4	3.1	3.3	2.9	2.6
South Africa	0.5	0.6	0.8	0.8	0.9
USA	2.4	2.0	1.6	1.6	1.5
China	1.5	1.8	2.0	1.8	1.9
India	0.9	0.8	0.8	0.95	1.0
Turkey	0.9	0.8	0.6	0.5	0.85
Others	1.5	1.7	1.6	1.7	1.7
Total world	26.9	27.2	23.6	21.2	23.8
Oil production (42% yield)	11.3	11.4	9.9	8.9	10

Source: Anuario de Estadistica Agroalimentaria (2000) Madrid: Ministerio de Agricultura, Pesca y Alimentación; Unilever Food España (2002) Statistics. Personal communication, Bilbao, Spain: Agra.

World Sunflower Oil Production

World edible oil and fats production has increased in the last years with a total seed production in the year 2001 of 306.9 million metric tons. The sunflower oil production since 1998 is presented in Table 1. World production of sunflowerseed in the season 2002/03 is currently forecast at 23.8 million metric tons. World

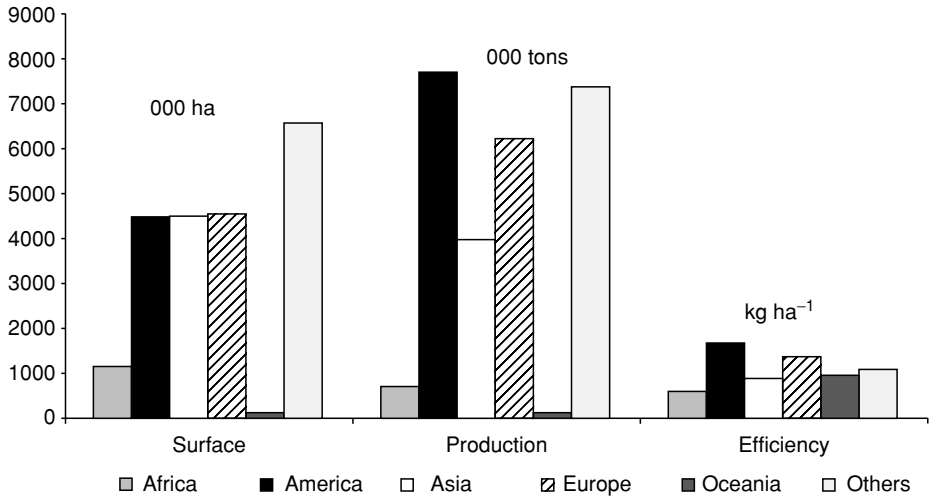

Figure 1 World crop production of sunflower. Data from Anuario de Estadística Agraria (1997) pp. 148–150. Madrid: Ministerio de Agricultura, Pesca y Alimentación; Unilever Food. España (2000) Statistics. Personal Communication. Bilbao, Spain: Agra.

production is thus seen recovering by about 2.6 million tons from last season but still this season's global sunseed crop is expected to remain approximately 3.4 million tons below the harvested amount in 1998/99. Oil production is highly dependent on the crop and thus is changeable from season to season.

Although sunflower oil is considered a premium oil, with a correct equilibrium between linoleic acid and tocopherols the benefits of monounsaturated fatty acids to health has encouraged the production of new oils with a high level of oleic acid, such as high-oleic-acid sunflower oil, high-oleic-acid rapeseed oil, high-oleic-sunflower oil (HOSO), etc. Thus, recently several new variants of sunflower have appeared with the aim of catering for the demand for oils richer in oleic acid and with a lower price than olive oil. Because they have greater oxidative stability than conventional sunflower oil, high-oleic-oil sunflower oils are expected significantly to expand food and industrial uses of sunflower oil all over the world.

Market and Prices

Sunflower derives most of its economic value from the oil extracted from the seeds; the remaining value is from the meal. The oil is considered to be of very high quality and generally sells for a premium in world markets, over soybean, palm, and rapeseed oils. In the period 1999–2001 sunflower oil price range between US $400 and 600. The year 2002 estimate price for crude and refined sunflower oils are 699 E/metric ton and 733 E/metric ton, respectively. However, due to unfavorable weather conditions further rise of prices for the sunflower oil can be expected for the next 9–10 months.

Reduced crushing of sunflower seed, as crushing margins remained inadequate, forced most crushers to process rapeseed as an alternative. However, increased arrivals of Argentina sunflower oil prevented the oil price from rising sharply.

Sunflower oil initially maintained a high and justified premium over soybean and rapeseed oil but as the lower Spanish crop was more or less compensated by higher yields in Eastern Europe and France, the market could not sustain the high premium. Currently the premium has eroded to US $80 from US $100. Sunflower seed supplies from Russia and the Ukraine have recently become uncertain as governments announced export taxes.

Oil Extraction

Before oil extraction, sunflower seed (*c.* 30% hulls) is usually dehulled to 8–12% because of the high wax content in the hulls. An outline of the extraction process is shown in Figure 2. Before pressing, other manipulations are necessary, such as delinting, cracking, breaking, conditioning, and structuring. Nowadays, the only presses used on an industrial scale are screw presses or expellers. The pressure applied in pressing can be as high as 300 000 kPa but is commonly around 150 000 kPa. Capacities of expellers used for prepressing are usually 100–200 tons per day, but equipment exists that can handle up to 500 tons per day. The solvent extraction is mainly carried out by percolation through the platelets of prepared seeds using nontoxic solvents that are powerful solvents for oils and nonsoluble in

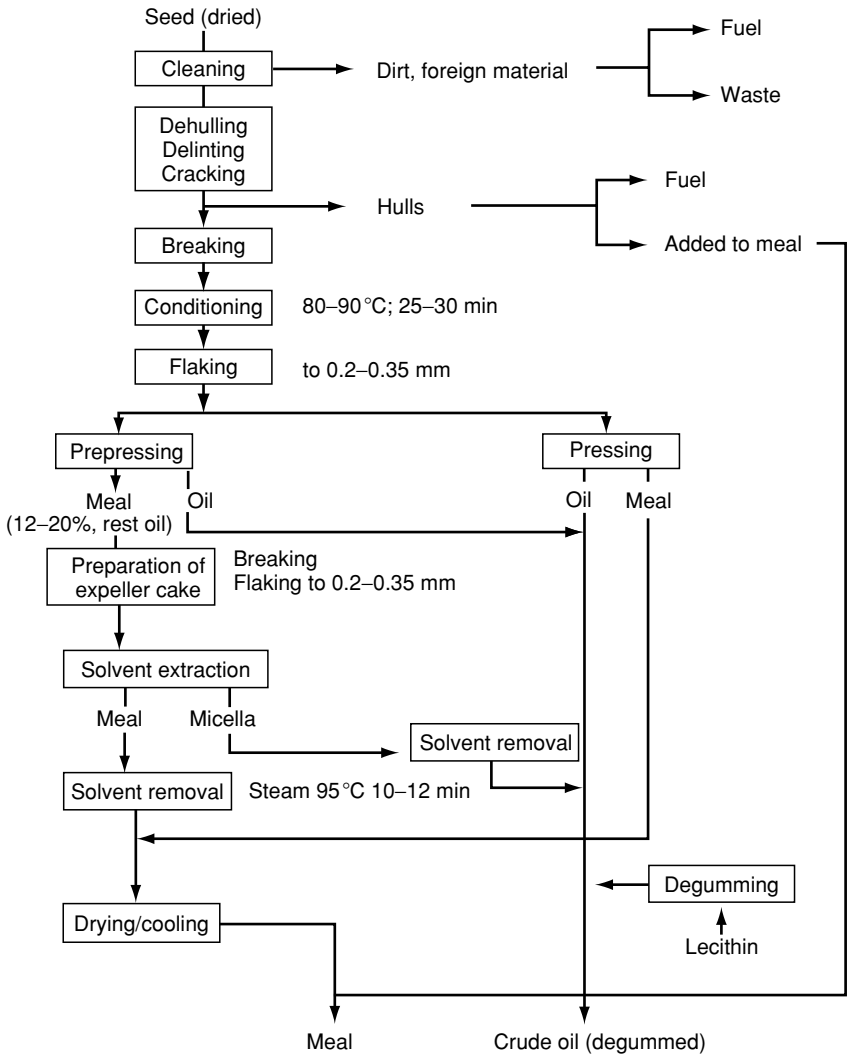

Figure 2 Sunflower oil extraction scheme. Modified from Dupont J (1999) Vegetable oils. In: Macrae R, Robinson RK and Sadler MJ (eds) *Encyclopedia of Food Science, Food Technology and Nutrition*, pp. 4686–4714. London: Academic Press.

water, low in price, and also nonflammable and nonexplosive.

Treatment of Crude Oil

In some countries the oil from the first pressing of seed is used directly. However, to refine the oil to change its color, odor, flavor, and physical properties, it has become a common procedure to satisfy consumer demands.

Refining

Fats and oils contain minor compounds which must be removed because they confer taste and smell that can diminish consumer acceptance. Steps in the chemical and physical refining of sunflower oil are similar to the ones processed in other vegetable oils.

Degumming

The degumming process is necessary to eliminate gums from the oils of some seeds. The oil is heated, the phosphoric acid (85%) is added and, after a reaction time, the gum-water sludge is filtered off or separated by centrifugation.

Neutralization

After harvesting or during ripening, lipolytic enzymes start to split the triacylglycerols contained in the oil and the free fatty acids are originated; this increases rapidly the acidity value. Because of this, it must be removed. Neutralization of these fatty acids can be performed by extraction, distillation, and chemical processes. Another action is reesterification of the free fatty acids with glycerol. Neutralization is carried out by alkalis or ammonia.

Bleaching

Among others, chlorophyll and carotenoids are colored substances. They are removed in the bleaching process by employing bleaching earth that adsorbs and removes these particles.

Deodorization

In this process odorous substances, mainly oxidation products, are eliminated. Deodorization can be considered to be a water vapor distillation, and must be carried out under vacuum, because the vapor pressure of ketones and aldehydes at 200 °C is of the order of 30 kPa, about 300 times higher than that of fatty acids. The process is generally carried out for seed oils at temperatures of about 240 °C for deodorization or 270 °C if distillative neutralization includes centrifugation.

However, the deodorization temperature for sunflower seed oil is 220 °C under ~400 kPa. Nevertheless, depending on the application, it may be necessary to modify the sunflower oil by means of hydrogenation, interesterification, or fractionation and winterization in order to change the melting point or improve clarity.

Composition

Triacylglycerol Species and Fatty Acid Composition

The chemistry and reactivity of natural fats and oils are mainly dependent on their overall composition and especially on the relative composition of the fatty acid constituents. Furthermore, the structure of the fats in terms of the relative proportions of different molecular species or of the sterospecific distribution of fatty acids on the glycerol moieties have a bearing on their properties. Table 2 presents the fatty acid profile of these two sunflower oil variants. The HOSO shows a fatty acid composition similar to other monoenoic oils, such as olive oil. The triacylglycerol composition of sunflower and HOSO is shown in Table 3. The former is rich in trilinolein, while the latter is rich in triolein.

Table 4 shows the characteristics that sunflower oil has to have in order to be an acceptable edible oil according to European Community regulations. The different fatty acid fraction contained in this oil may be quite variable, as shown in Table 2. Nevertheless, polyunsaturated fatty acids predominate. The percentage of oleic acid (C_{18}:1, *n*-9) plus linoleic acid (C_{18}:2, *n*-6) is rather constant (85–87%) because both fatty acids originated in the same biosynthesis chain. The linoleic acid is produced from oleic acid by desaturation carried out by the Δ^9 desaturase enzyme. The major *n*-6 fatty acid is linoleic acid, ranking 50–72%. This edible oil also contains about 11–12% of saturated fatty acids (mainly palmitic and stearic acids). Oleic acid represents about 16–25% and only minor amounts of α-linolenic acid are present (Table 4).

Sunflower oils from cold climates are poor in oleic acid and very rich in linoleic acid (about 15% and 75%, respectively) with an iodine value of 125–136, and do not require winterization. Southern oils have been found to contain over 50% of oleic acid and as little as 35% of linoleic acid with an iodine value of ~106.

Table 2 Composition of sunflower oil and high-oleic-acid sunflower oil

	Sunflower (%)	*High-oleic-acid sunflower oil (%)*
Trilinolein	26.1%	4.25%
Myristic acid	0.06%	0.05%
Palmitic acid	6.7%	4.29%
Palmitoleic acid	0.15%	0.13%
Margaric acid	0.07%	0.04%
Margaroleic acid	0.03%	0.05%
Stearic acid	4.9%	3.99%
Oleic acid	24.3%	75.7%
Linoleic acid	62.1%	13.7%
Linolenic acid	0.08%	0.06%
Arachidic acid	0.35%	0.35%
Gadoleic acid	0.14%	0.23%
Behenic acid	0.78%	0.97%
Lignoceric acid	0.14%	0.34%
Elaidic acid	0%	0.03%
Linoelaidic acid	0.14%	0.1%
Squalene	0.04%	0.05%
Total sterols	3618 p.p.m.	3100 p.p.m.
Cholesterol	0.09%	0.12%
Brasicasterol	0.07%	0.07%
24-metylencholesterol	0.14%	0.09%
Campesterol	8.5%	8.2%
Campestenol	0.24%	0.33%
Stigmasterol	8.0%	8.9%
Δ-7-campesterol	3.0%	3.2%
Δ-5-23-stigmastadienol	0.29%	0.38%
Clerosterol	0.87%	1.0%
β-Sitosterol	53.4%	53.8%
Sitostanol	0.36%	0.53%
Δ 5-avenasterol	2.7%	1.6%
Δ-5, 24-stigmastadienol	1.7%	2.0%
Δ-7-stigmasterol	15.6%	15.4%
Δ-7-avenasterol	5.0%	4.3%
Apparent β-sitosterol	59.3%	59.4%
Polyphenols	2 p.p.m.	2 p.p.m.
Alifatic alcohols	63 p.p.m.	59 p.p.m.
Polar content	3.7 % oil	3.6 % oil
Polymers of triacylglycerol	0.06 % oil	0.03 % oil
Dimers of triacylglycerol	0.62 % oil	0.18 % oil
Nonoxidized triacyglycerols	96.3 % oil	96.4 % oil
Oxidized triacylglycerols	1.57 % oil	1.13 % oil
Diacylglycerols	1.00 % oil	1.89 % oil
Monoacylglycerols	0.00 % oil	0.00 % oil
Free fatty acids	0.45 % oil	0.37 % oil

Table 3 Triacylglycerol composition of sunflower oil and high-oleic-acid sunflower oil

	Sunflower oil	*High-oleic-acid sunflower oil*
POP	0.3	0.6
POS		0.9
PLP	1.0	
POO + PLS	3.9	12.7
PLO	7.8	2.1
PLL	11	
SOO	1.2	10.3
OOO + SOL	10.6	60.2
OOL + SLL	16.4	6.2
OLL	22.8	3.3
LLL	21	1.1

POP, palmitic, oleic, palmitic; POS, palmitic, oleic, stearic; PLP, palmitic, linoleic, palmitic; POO, palmitic, oleic, oleic; PLS, palmitic, linoleic, stearic; PLO, palmitic, linoleic, oleic; PLL, palmitic, linoleic, linoleic; SOO, stearic, oleic, oleic; OOO, oleic, oleic, oleic; SOL, stearic, oleic, linoleic; OOL, oleic, oleic, linoleic; SLL, stearic, linoleic, linoleic; OLL, oleic, linoleic, linoleic; LLL, linoleic, linoleic, linoleic.
Source: Instituto de la Grasa (Sevilla, Spain), personal communication.

Minor components

In the last decade certain minor constituents of oils have been extensively studied, because of their antioxidant properties, and also because they can modulate the dietary cholesterol absorption, the fatty acid metabolism and thus the serum cholesterol levels. Table 2 shows a typical composition of sunflower oil's minor components. Large amounts of phytosterols and tocopherols, and relatively high amounts of squalene are present. Table 4 shows some admitted limits for tocopherols and phytosterols in sunflower oil.

Applications

Catering

Contrasting with other vegetable oils, such as soybean, corn, and cotton, the oil extracted from the sunflower seed contributes about three-fourths of the total value of the crop. Sunflower oil is appreciated among vegetable oils because of its bland flavor, light color, high smoke point, high level of linoleic acid, and very poor linolenic acid content.

Sunflower oil is primarily used as a salad dressing and cooking oil. Furthermore, it is extensively used as a major ingredient in some margarine and shortening products. The use of sunflower oil in salad is less frequent than olive oil in Mediterranean countries. Nevertheless, sunflower oil is prominently used in deep-fat frying throughout the world. At present, Spaniards and others, particularly those who live near to the Mediterranean coast, continue buying and consuming a relatively large quantity of olive oil, mainly for frying purposes. However, other oils, mostly sunflower oil, have largely substituted the use of olive oil. This situation is particularly acute in the Spanish institutional sector (hospitals and welfare centers, prisons, military bases, teaching institutions, company dining facilities, government establishments, etc.). Economic reasons have promoted the production and consumption of sunflower oil.

Table 4 Sunflower oil characteristics

Minimum quality characteristics	
Aspect	Clear, transparent after 24 h at 20 ± 2 °C
Odor and taste	Normal
Color	M 25 YU and 2.5 RU
Moisture and volatile compounds	M 0.1%
Insoluble impurities in petroleum ether	M 0.05%
Free acidity (% oleic acid)	M 0.2%
Peroxide index (mmol l^{-1} active O kg^{-1} oil)	M 10
Soap residues	Negative
Purity test	
Vizern test	Negative
Saponification index	188–194
Halphen reaction	Negative
Iodine index (Hanus method)	80–145
Refraction index (at 25 °C)	1.767–1.474
Nonglyceric esters	Negative
Saturated fatty acids in triglyceride β-position	M 1%
Fatty acid composition	
Lauric acid (12:0)	M 0.1%
Myristic acid (14:0)	M 0.1%
Palmitic acid (16:0)	3.5–8.0%
Palmitoleic acid (16:1)	M 0.2%
Stearic acid (18:0)	3–7%
Oleic acid (18:1)	15–85%
Linoleic acid (18:2)	50–72%
Linolenic acid (18:3)	M 0.2%
Arachidic acid (20:0)	M 0.6%
Gadoleic acid (20:1)	M 0.3%
Behenic acid (22:0)	M 1%
Insaponificable	0.5–1.5%
Sterol composition	
Cholesterol	M 0.5%
Campesterol	6.5–10.5%
Stigmasterol[a]	6.0–12.5%
β-Sitosterol	55.0–67.5%
Δ 7 stigmasterol	16.0–26.0%
Tocopherols[b]	
α-tocopherols	403–935 (p.p.m.)
β-tocopherols	0–45 (p.p.m.)
γ-tocopherols	0–34 (p.p.m.)
δ-tocopherols	0–7 p.p.m.)

[a]Δ 5,23 stigmasterol + clerosterol + sitosterol + Δ 5-avenasterol + Δ5,24 stigmastadienol.
[b]Alinorm 97/17. Conjoint program FAO/OMS.
Source: last modification to European Community Regulation CE no. 656/95 to CE no. 2568/91.
UA and YA refer to Lovibond test.
YU, yellow units; RU, red units.
M, Maximum.

Deep-Fat Frying

Deep-fat frying, a quick elaboration procedure of food, improves the palatability of food as well as providing flavor and nutritive value. This culinary technology can be defined as a complex process of controlled dehydration and browning with hot oil as the heat transfer medium.

Many factors are at work:

1. The process itself: temperature, time, frying method, and vessel material.
2. The type of fat: chemical and physical properties of the fat, additives, and contaminants.
3. The food size and form: rate surface/volume, battered, floured, direct action (on thermolabile nutrients improving nutrient interactions, nutrient interchange between the foodstuff and the frying fat).

During deep-fat frying a wide variety of chemical reactions take place. The steam from a moist food will cause hydrolysis of triacylglycerols, resulting mainly in the formation of free fatty acids and diacylglycerols (monoacylglycerols and glycerol are produced in small quantities during controlled frying), while the air released into the frying system will initiate a cycle of oxidation involving the formation of free radicals. These oxidation processes will involve fatty acids in intact triacylglycerols as well as the products of triacylglycerol hydrolysis. The fatty acid residues of triacylglycerols formed can react to form polymers and other complex reaction products. On the other hand, the food being fried can influence the type and quantity of breakdown products in the frying medium.

The substrate can influence the process in three ways:

1. By releasing natural antioxidants or prooxidants into the frying oil. These products can also be absorbed by the frying substrate.
2. By the absorption of lipid oxidation products on to the substrates
3. By the catalytic effect of various functional groups present in the food.

Frying with sunflower oil Sunflower oil is one of the most important vegetable oils employed for deep-frying. It has been used in the cooking of food such as French fries and frozen prefried foods at home, in fast-food restaurants, and in the industry. The frying process can also be used to produce cardio-health foods because during frying the exchange of fat between the frying medium and the food occurs. Thus, the level of linoleic acid in the food can be deeply increased when frying is performed in sunflower oil. Moreover, after frying some minor components from the oils, such as tocopherols, are present in the fried food.

The degradation must be ascribed to a lower turnover of fresh oil during frying, but in industrial cooking, where a high turnover of fresh oil is used, sunflower oil can be used successfully. The thermal oxidation of sunflower oil used in frying (as with other oils or fats) can be evaluated by measuring the percentage of total polar content by column chromatography. The level of 25% polar content has been accepted as critical for oil discarding. Frying potatoes in sunflower oil following domestic conditions shows that polar content increases in oil and fried potatoes. However, when frying is performed with frequent oil turnover, the amount of polar material in sunflower oil is substantially below the critical level (**Figure 3**).

The amount of polymers and dimers of triacylglycerols formed in sunflower oil through deep-fat frying is also lower in the frequent turnover of fresh oil than in the null turnover one. The benefit of the high turnover of sunflower oil has also been found in the frying of frozen foods (potatoes, fish fingers, croquettes, pastries, etc.) under domestic conditions.

HOSO seems to be a rather stable oil and has been recently studied. Experiments using HOSO to fry potatoes following a domestic frying model suggests

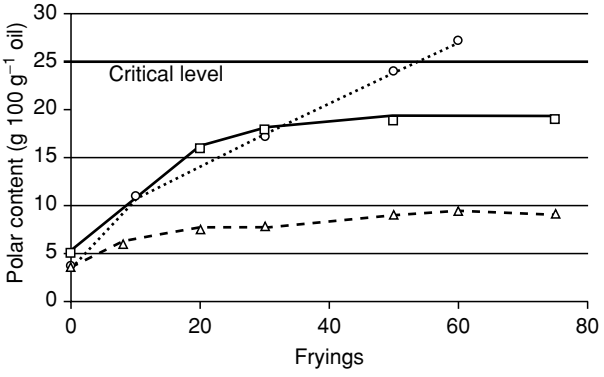

Figure 3 Sunflower oil and high-oleic-acid sunflower oil behavior in frying potato. Polar fraction (mg 100 mg^{-1} oil) change. Open circles, sunflower oil used in discontinuous frying of French fries without replenishment with fresh oil. Open squares, sunflower oil in discontinuous frying of French fries with frequent replenishment with fresh oil. Open triangles, high-oleic-acid sunflower oil in discontinuous frying with frequent replenishment of French fries with fresh oil. Modified from Cuesta C and Sánchez-Muniz FJ (1998) Quality control during repeated fryings, *Grasas y Aceites* 49: 310–318 and Sánchez-Muniz FJ, Cuesta C, López-Várela S, Garrido-Polonio MC, and Arroyo R (1993) Evaluation of the thermal oxidation rate of sunflower oil using various frying treatments. In: Applewhite TH (ed.) *Proceedings of the World Conference on Oilseed Technology and Utilization*, pp. 448–452. Champaign, IL: AOCS Press.

that frequent replenishment of used HOSO with fresh HOSO permits the frying of sets of fresh potatoes at a very high rate (Figure 3). Moreover, HOSO performs much better than conventional sunflower oil in the frying of prefried frozen foods. This occurred using the frequent and the null turnover techniques, commented on before. The triacylglycerol and fatty acid composition and the ratio of antioxidant to unsaturated fatty acids explain the different stability and frying behavior of these two variants of sunflower oil.

Among natural sunflower oil antioxidants, tocopherols seem to be more relevant. In this way, it is interesting to note that, in continuous frying operations, a regular turnover of fresh oil maintains the level of antioxidant in the oil.

However, data from some publications suggest that addition of the antioxidant exerts little influence in extending frying life of the oil used in frying but it can extend the shelf-life of the food by reducing the rate of oxidation of the absorbed oil. Nevertheless, data from the AIR (The Agriculture and Agro Industry Research) project have shown that peroxidative degradation of frying oils occurred when there was a lack of tocopherol content.

Oleochemistry

Margarines Margarines were originally obtained from lard and other fat sources, and were widely consumed through the world until the 1970s and 1980s. However the association of saturated fats with coronary heart disease led to the development of new margarine products of vegetable origin. These new margarines, with high-polyunsaturated-fatty-acid content, include soft margarines and diet spreads. Sunflower oil is one of the oils employed in the formulation of fat blends to achieve an adequate product. Typical fat blends for margarines are shown in Table 5.

In margarines the level of polyunsaturated fatty acids should not be less than 45% of *cis, cis* linoleic acid or less than 50% of total polyunsaturated fatty acids; the sum of saturated fatty acids and *trans* fatty acids should not exceed 25% and the level of cholesterol should not exceed 15 mg kg^{-1}. Considering these facts, in order to reach 45–50% in polyunsaturated fatty acids, it will be necessary to use 70–80% sunflower seed or soybean oils.

Table 5 Fat blends for polyunsaturated fatty acid margarines

	PUFA 1(%)	*PUFA 2(%)*
Hydrogenated vegetable oil (melting point 55 °C)	5	
Hydrogenated vegetable oil (melting point 46–48 °C)		20
Hydrogenated vegetable oil (melting point 40 °C)	15	
Sunflower oil	80	80

PUFA, polyunsaturated fatty acids.

Other Uses

The sunflower meal, obtained after the oil is extracted from the seed, has a high protein percentage (28% for meal from hulled seeds to 42% for dehulled seed meals) and is used primarily in food rations for livestock and poultry. Compared to soybean, sunflower meal contains lower energy and lysine but higher fiber and methionine content.

Variants of sunflower seeds with low oil content (nonoilseed sunflower) are primarily used as a snack food, an ingredient in baked foods, salads, and candies, and for feeding birds and small animal pets. During the last decades a flowering market has grown round these seeds.

Finally, sunflower hulls present a potential use as litter for livestock or poultry, and as a source of roughage in animal feeding. Building or insulation boards and, fireplace logs are other interesting potential applications.

However, the relatively higher price of sunflower oil in respect to other seed oils explains why it is not widely used for industrial purposes. Nevertheless, some paints, varnishes, and plastics take advantage of its semidrying properties without the yellowing problems associated with oils rich in linolenic acids. In some eastern European countries the oil is also employed by detergent and soap manufacturers. Like other vegetable oils, it has potential value for use in the production of adhesives, agrochemicals, surfactants, additional plastics and plastics additives, fabric softeners, synthetic lubricants, coatings, and other products.

HOSO has potential value in industrial products where oleic acid and related chemicals are currently utilized. This new oil is of initial interest in food use (primarily in frying of snack foods to enhance the shelf-life of retail products), and as an ingredient in infant food formulas which require stability, bland flavor, and specific fatty acid composition. The high level of oleic acid together with the high oxidative stability is valuable for cosmetic and pharmaceutical manufacture.

Dietary Importance and Health Effects

Sunflower Oil as Energy Source

While protein and carbohydrates provide 17 kJ g^{-1} (4 kcal g^{-1}), long-chain triacylglycerols provide 38 kJ g^{-1} (9 kcal g^{-1}). The fatty acids are used for

energy production in the cell mitochondria and peroxisomes. Present recommendations for fat are about 30% of total energy intake (80–90 g day^{-1} in a 2200–2450 kcal/diet). Due to a large proportion (more than 50% of total fat, in Mediterranean countries) that comes from cooking oil, the fatty acid composition of a diet depends largely on the culinary oil used.

Recent knowledge suggests that a unit of energy from fat may be more fattening than a unit of energy from carbohydrate. This fact can be related to the limited body capacity to convert carbohydrate into fat and the faster adaptation to oxidize excess carbohydrate than excess fat. Moreover, body capacity for storing carbohydrate as glycogen is limited, whereas in adipose tissue, fat storage is virtually unlimited.

Sunflower Oil as Essential Fatty Acid Source

Linoleic and linolenic acids are the two first (parent) members of ω-6 (*n*-6) and ω-3 (*n*-3) fatty acid families, respectively. Both are essential and must be supplied by the diet because humans and many animals have lost the ability to synthesize them. Moreover, other fatty acids of the *n*-6 family such as γ-linolenic acid (18:3, *n*-6), dihomo-γ-linolenic acid (20:3, *n*-6), and arachidonic acid (20:4, *n*-6) cannot be formed.

The major *n*-6 fatty acid in the western diet is linoleic acid. Although other oils are very rich in linoleic acid, sunflower oil appears to be the primary source of this fatty acid. The importance of linoleic acid consumption in humans and animals is detailed elsewhere. (*See* **Essential Fatty Acids**.)

Deficiency of this fatty acid is called essential fatty acid (EFA) deficiency; it is well documented in the rat and can be produced in several animals, including humans. The disease is characterized by skin symptoms such as dermatosis. Growth is retarded, reproduction is impaired, and there is degeneration or impairment of function in many organs of the body. EFA deficiency is characterized by changes in the fatty acid composition of many biological membranes, depending on how much vegetable oil, such as sunflower oil, is consumed in the diet; the intake of this fatty acid generally varies from 4–10%. Overt EFA deficiency is only seen when it provides less than 1–2% of dietary energy (about 2–5 g d^{-1} for adults).

Sunflower Oil as Antioxidant Source

As commented on before, sunflower oil contains valuable amounts of minor compounds such as α-tocopherols, γ-tocopherols, and phytosterols. The correct equilibrium between polyunsaturated fatty acids and antioxidants protects membranes from oxidative stress. Among other properties, phytosterols act not only as free radical scavengers but also as antipolymerization agents in frying oils.

Influence of Polyunsaturated Fatty Acids *n*-6 on Cholesterolemia and Lipoproteinemia

It is currently recommended that dietary total fat, saturated fatty acids, and cholesterol should be reduced while the intake of unsaturated fats should be increased. As a result of these considerations the consumption of unsaturated oils has been recommended and margarines rich in polyunsaturated and monounsaturated fatty acids, called health margarines, have been developed. Linoleic acid has been considered for many years to be the unique fatty acid that had a hypocholesterolemic effect at serum and lipoprotein particle level.

Recent investigations suggest that oleic acid also presents such an effect on low-density lipoprotein particles. However, a large intake of linoleic acid decreases high-density-lipoprotein cholesterol while oleic acid has no effect on high-density-lipoprotein cholesterol levels.

Influence of Polyunsaturated Fatty Acids *n*-6 on Other Physiological Processes

n-6 fatty acids have been proven to exert other important roles in cytokine production and monocyte quimiotactism, and in eicosanoid production.

Platelet aggregation, inflammation, and cellular immunity are processes modulated by metabolic products derived from *n*-6 polyunsaturated fatty acids. Because sunflower oil contains large amounts of linoleic acid (the fatty acid mother of other *n*-6 fatty acids), the intake of this oil should be adequate, avoiding low and very high intakes that would produce an EFA deficiency or *n*-6 fatty acid saturation.

Like other fats, sunflower oil may modulate all these previous cited processes because of its ability to change the fatty acid composition of membrane phospholipids. As a consequence, membrane fluidity may change and also, when cytokines act on target cells, the amount and types of prostaglandins and leukotriens originated are modified. Dihomo-γ-linolenic and arachidonic acids derive from linoleic acid, and are precursors of eicosanoids. These fatty acids can be metabolized via different enzymatic pathways, giving rise to thromboxanes, prostacyclins, prostaglandins, leukotriens, and hydroxy-fatty acids and oxygenated metabolites, which have been proved to modulate inflammation, cellular immunology, thrombosis, cardiovascular disease, and cancer.

The recommendation for the polyunsaturated fatty acid contribution to total energy is 7.5%, with a total of 6% as linoleic acid. Other scientific societies recommend that linoleic acid should contribute 3–6% of

total energy in the diet. However, because cardiovascular disease and cancer have been defined as a consequence of oxidative damage, vegetable oils consumed in correct amounts would add benefits in reducing such degenerative diseases, because they contain antioxidant and flavonoids and an adequate linoleic-acid-to-vitamin-E ratio. Immune response can be improved by consuming unsaturated fatty acids; however large quantities of *n*-6 fatty acids could suppress cellular immunity.

See also: **Antioxidants**: Natural Antioxidants; **Essential Fatty Acids**; **Fatty Acids**: Metabolism; Analysis; *Trans*-fatty Acids: Health Effects; **Gums**: Properties of Individual Gums; Food Uses; Dietary Importance; **Lipoproteins**; **Margarine**: Types and Properties; Methods of Manufacture; Composition and Analysis; Dietary Importance; **Marine Foods**: Production and Uses of Marine Algae

Further Reading

Abraham G and Hron RJ (1992) Oilseeds and their oils. In: Hui YH (ed.) *Encyclopedia of Food Science and Technology*, vol. 3, pp. 1901–1910. New York: John Wiley.

Anonymous. Article on high-oleic oils posted on the internet (netlink: www.high-oliec.de/ and www.High-Oleic.-com)

Anuario de EstadÍstica Agraria (1997) pp. 148–150. Madrid: Ministerio de Agricultura, Pesca y Alimentación.

Boskou D (1999) Non-nutrient antioxidants and stability of frying oils. In: Boskou D and Elmadfa I (eds) *Frying of Food. Oxidation, Nutrient and Non-Nutrient Antioxidants, Biologically Active Compounds and High Temperatures*. Lancaster: Technomic, pp. 183–204.

British Nutrition Foundation (1992) *Unsaturated Fatty Acids: Nutritional and Physiological Significance. The Report of British Nutrition Foundation's Task Force*. London: Chapman & Hall.

Cuesta C and Sánchez-Muniz FJ (1998) Quality control during repeated fryings. *Grasas y Aceites* 49: 310–318.

Dupont J (1999) Vegetable oils. In: Macrae R, Robinson RK and Salder MJ (eds) *Encyclopedia of Food Science, Food Technology and Nutrition*, vol. 7, pp. 4686–4714.London: Academic Press.

Eierdanz H (1993) Oleochemistry: processes and products. In: Applewhite TH (ed.) *Proceedings of the World Conference on Oilseed Technology and Utilization*, pp. 217–220. Champaign, IL: AOCS Press.

European Community Regulation. CE no 656/95 and CE no 2568/91.

Gundstone F (2002) Sunflower seed and its products. *Inform* 13: 159–163.

Gupta MK and NuSun (2000) Healthy oil at a commodity price. *Lipid Technology* 12: 29–33.

Fick GN (1989) Sunflower. In: Röbbelen G, Downey RK and Ashri A (eds) *Oil Crops of the World. Their Breeding and Utilization*, pp. 301–318. New York: McGraw-Hill.

Firestone D (ed.) (1998) In: *Physical and chemical characteristics of oils, fats and waxes*, 5th edition, Champaign, IL: AOCS Press.

Gray J (1998) Nuts and seeds. In: Sadler MJ, Strain JJ, and Caballero B (eds) *Encyclopeia of Human Nutrition*, vol. 3, pp. 1422–1429. London: Academic Press.

Macrae R, Robinson RK and Sadler MJ (eds) (1993) *Encyclopaedia of Food Sciences, Food Technology and Nutrition*. London: Academic Press.

Oil World Annual 2001, vol 44, number 45. (netlink: www.oilworld.de/).

Sánchez-Muniz FJ, Cuesta C, López-Varela S, Garrido-Polonio MC and Arroyo R (1993) Evaluation of the thermal oxidation rate of sunflower oil using various frying treatments. In: Applewhite TH (ed.) *Processing of the World Conference on Oilseed Technology and Utilization*, pp. 448–452. Champaign, IL: AOCS Press.

Unilever Food España (2000) Statistics. Personal communication. Bilbao, Spain: Agra.

Weiss TJ (ed.) (1983) Sunflower oil. In: *Food Oils and Their Uses*, 2nd edn, pp. 43–44. Chichester: Ellis Horwood.

SUPERCRITICAL FLUID EXTRACTION

E Reverchon, University of Salerno, Via Ponte Don Melillo, Italy

Background

Many food and nutrition ingredients are traditionally produced by organic solvents extraction. This process is relatively simple to perform and is firmly established in industrial practice; however, it has several drawbacks: it is a polluting process, since organic solvents can be released to the atmosphere, they remain in starting material, and they have to be eliminated from the extracted products. Owing to the increasing demand for nonpolluting processes and product purity, alternative processes are intensively coming under scrutiny. The most promising alternative to organic solvent processing is supercritical fluids extraction (SCFE); several SCFE produced compounds are already commercialized.

A supercritical fluid (SCF) is, as a rule, a gas under room conditions that is heated and compressed above its critical temperature and pressure. When it is under supercritical conditions, the SCF is in the gas state and maintains its related properties, but, depending on the operating pressure, it can assume variable densities from gas-like to liquid-like. Therefore, SCFs are sometimes called 'variable geometry' solvents. This characteristic makes them suitable for selective extraction processes. Indeed, it is possible to modulate the solvent power of a SCF to perform intensive or selective extraction processes. After the extraction, owing to pressure reduction, the SCF is immediately and completely eliminated from the raw material and from the extracts.

Table 1 lists some SCFs proposed for extraction processes, together with their critical temperatures and pressures. It is evident that only some of them can be applied to food extraction, since their critical temperatures are often too high for the stability of thermolabile compounds that are characteristic of food.

The large majority of the extraction applications proposed for SCFE use carbon dioxide (CO_2) as the SC solvent because CO_2 is cheap, nonflammable, nonpolluting, and available in large quantities. Its critical temperature and pressure are readily obtainable in an industrial process, and the SCF process temperatures that can be set (for example from 35 to 50 °C) are compatible with the thermal stability of most food-related ingredients. The general behavior of SC-CO_2 is similar to a nonpolar solvent, i.e., it has strong affinity with lipophylic molecules. Also, since supercritical CO_2 is used during the process, but not produced during extraction, it does not contribute to the so-called 'greenhouse effect.' Thus, the only drawback of SC-CO_2 extraction is the requirement of a pressurized plant to perform the process.

SCF extraction has been successfully performed in many cases on the laboratory scale, but only a few industrial-scale applications have been developed until now. However, the potential of SCFE is very high, and many industrial applications are expected to be developed in the next years.

Table 1 List of supercritical fluids proposed for the extraction process

Supercritical fluid	T_c(°C)	p_c(bar)
Carbon dioxide	31.1	73.8
Nitrogen protoxide	36.5	70.6
Propane	96.7	42.4
Ethanol	243.4	72.0
Isopropyl alcohol	235.3	47.6
Trifluoromethane	26.0	46.9

Extraction Processes

From the point of view of the process arrangement, SCFE can be divided into three major areas, according to the physical state of the compounds to be extracted or of the raw material to be treated:

- *Extraction from solid matrices* is used to extract one or more compounds from a solid matrix, for example, to extract oil from seeds, essential oils, etc.
- *Extraction from liquid mixtures* is used when liquid compounds are to be extracted from a liquid mixture. It is used, for example, to eliminate hexane from seed oil and alcohol from wine, and for milk fat fractionation.
- *Spray extraction* is used to extract solid compounds from a liquid mixture. In this case, the previous process arrangements cannot be used. It can be used to extract lecithin from seed oils.

This classification underlines the similarities among the various extraction processes and can be useful for choosing the process arrangement to apply SCFE to products different from those described in the scientific and technical literature.

Extraction from Solid Matrices

This is the most studied SCFE application, since the most frequently required process is the extraction/elimination of one or more compound families from a solid (natural) matrix. The base extraction scheme is represented in Figure 1. It consists of an extraction vessel filled with the raw matter to be extracted. As a rule, the raw matter is dried and finely ground to facilitate the extraction process. It is loaded in a basket located inside the extractor to allow rapid charge–discharge of the extraction vessel. Metallic frits at both ends of the extractor allow the SCF to go through the bed of solid material to extract the required compounds. The SCF at the exit of the extractor flows through a depressurization valve to a

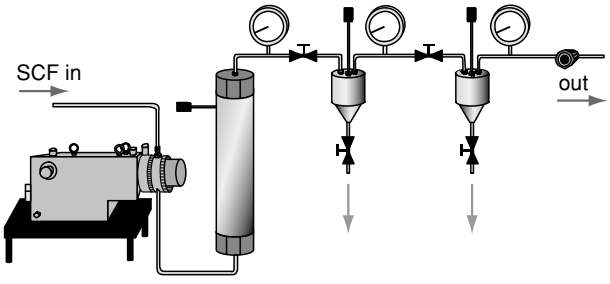

Figure 1 Extraction scheme for extraction from solid matrices.

separator in which, because of the lower pressure, the extracts are released from the gaseous medium and collected. More sophisticated extraction schemes (such as that shown in Figure 1) contain two or more separators. In this way, it is possible to fractionate the extract in two or more fractions of different composition by setting suitable temperature and pressure values in the separators.

Other possible variations of the SCFE scheme are multistage extraction and cosolvent addition. The multistage extraction strategy can be used when several compound families are to be extracted from the same matrix, and these are characterized by their different solubilities in SC-CO_2. It takes advantage of the fact that the solvent power of SC-CO_2 strongly increases with pressure at a fixed temperature. Therefore, it is useful to perform a first extraction operating at a low CO_2 density (e.g., 0.29 g cm^{-3}, 90 bar, 50 °C) followed by a second extraction step at a high CO_2 density (e.g., 0.87 g cm^{-3}, 300 bar, 50 °C). The most soluble compounds are extracted during the first stage (e.g., essential oil) and the less soluble compounds are extracted in the second one (e.g., antioxidant compounds).

A liquid cosolvent can be added to SC-CO_2 to increase its solvent power towards polar molecules. Indeed, as previously mentioned, SC-CO_2 is a good solvent for lipophilic (nonpolar) compounds, whereas it has a low affinity with polar compounds. Various authors have added small quantities of liquid polar solvents (for example, ethyl alcohol) that are readily solubilized by SC-CO_2 owing to their low molecular weight but, when in solution, modify the solvent power of the SC solvent. The only drawback with this strategy is that the cosolvent is a liquid and is collected in the separator together with the extracted compounds that have to be processed later for solvent elimination. Therefore, one of the advantages of the SCFE, i.e., the complete solvent elimination, is lost.

Selection of the Operative Conditions

The selection of the operating conditions depends of course on the specific compound or compound family to be extracted. Molecular weight and polarity have to be taken into account case by case, but some general rules can be applied. First, the SCFE temperature for food compounds is fixed between 40 and 50 °C, e.g., in the vicinity of the critical point and as low as possible to avoid thermal degradation. The increase of temperature reduces the density of SC-CO_2 (for a fixed pressure), thus reducing the solvent power of the SC solvent, but it increases the vapor pressure of the compounds to be extracted and, therefore, their tendency to pass in the fluid phase. However, the most relevant process parameter is the extraction pressure that can be used to tune the selectivity of the SCF. The general rule is: the higher the pressure, the greater the solvent power and the lower the extraction selectivity. Frequently, the solvent power is described in terms of the SC-CO_2 density under the operating conditions. CO_2 density can vary from about 0.15 to 1.0 g cm^{-3} at the operating conditions useful for SCFE and is connected to both pressure and temperature. Its variation is strongly nonlinear, so the proper selection requires the use of accurate tables of CO_2 properties (for example, the International Thermodynamic Tables of the Fluid State CO_2).

The other crucial parameters in SCFE are CO_2 flow rate, the particle size of the matrix, and the duration of the process (extraction time). The appropriate selection of these parameters leads to complete extraction of the desired compounds in a shorter time. These are associated with the thermodynamics (solubility) and the kinetics of the extraction process in the specific raw matter (mass-transfer resistance).

The proper selection depends on the mechanism that controls the process (the controlling mechanism is the slower mechanism and determines the overall process velocity). For example, particle size plays an important role in extraction processes controlled by the mass-transfer resistance, since a reduction in the mean particle size reduces the length of diffusion of the solvent. However, if particles are too small (for example, smaller than 0.2 mm), they can cause problems of channeling inside the extraction bed: part of the solvent flows through channels formed inside the extraction bed and does not contact the material to be extracted, thus causing a loss of efficiency and yield in the process. Moreover, the production of very small particles by grinding could lead to a loss of volatile compounds.

Applications of Solid Matter Extraction

Many solid matrices have been processed by SCFE. Some relevant categories of products are listed in Table 2. Some well-established industrial processes

Table 2 Some food and food ingredients processed by solid matter SCFE

Essential oils (spices, citrus, etc.)
Seed oils, fats, and waxes
Fatty acids, fatty acid methyl esters
Natural colorants, pigments
Vitamins
Antioxidants (licopene, tocopherols, etc.)
Alkaloids (caffeine)

use SCFE to produce hops extracts and several food nutritional substances that also offer some aspects of therapeutic protection to the human body (nutraceuticals). One of the largest industrial applications is coffee decaffeination, which is performed on green coffee beans presoaked in water. Green coffee beans are used to prevent coextraction of aroma compounds that are formed during roasting. The use of water in the preliminary treatment allows the release of caffeine from the natural substrate where it is linked in the form of sodium salts. Very large plants that operate at pressures between 280 and 300 bar at 40 °C are used in this process with internal volumes larger than 10 m^3. Caffeine is recovered during the SCFE with a purity of around 96% by weight; the final caffeine content in coffee beans can be lower than 0.05% w/w.

Two examples of SCFE are now described in detail: essential oil isolation, which represents a good example of extraction, followed by fractional separation, and seed oil extraction.

Essential oil isolation Essential oils are mainly formed by hydrocarbon and oxygenated terpenes and by hydrocarbon and oxygenated sesquiterpenes. They are commonly extracted from seeds, roots, flowers, herbs and leaves using the so-called hydrodistillation. This is a very simple process but has many drawbacks: thermal degradation, hydrolysis, and solubilization in water of some compounds that alter the flavor and fragrance profile of many essential oils extracted by this technique. (*See* **Essential Oils**: Isolation and Production.)

From the point of view of SCFE, essential oil isolation is an example of extraction followed by fractional separation. Indeed, this extraction process can be optimally performed at mild pressures (from 90 to 100 bar) and from 40 to 50 °C, which correspond to relatively low CO_2 densities. However, even under these operating conditions, essential oil components are extracted together with cuticular waxes, i.e., paraffinic compounds located on the surface of vegetable matter with the aim of controlling its perspiration. Paraffins have a relatively low solubility in SC-CO_2 but are located on the surface of the vegetable matter, whereas essential oil compounds are readily soluble in SC-CO_2 but are partly located inside the vegetable structure. Therefore, extraction of waxes is controlled by their solubility, and essential oil extraction is controlled by mass-transfer resistance through the vegetable structure. As a result of these different interactions, the two compound families are coextracted even under the optimum extraction conditions. Since coextraction is unavoidable, it is possible to take advantage of the fact that at the same pressures, but at low temperatures (from −5 to +5 °C), waxes are practically insoluble in CO_2, whereas essential oil compounds are still very soluble. Therefore, fractionation of the extract is possible by operating a first separation, for example at 0 °C and 90 bar and operating a second separation at 20 °C and 30 bar. In the first separator, the selective precipitation of waxes is obtained, and in the second, essential oil is recovered. This concept has been applied up to the industrial scale, and to date, two companies operating in Germany and in Italy use this process to isolate essential oils.

It is worth noting that it is not possible to perform the CO_2 extraction directly at 0 °C and 90 bar, since the vegetable matter contains many other compound families (antioxidants, colors, etc.) that are very soluble under these process conditions.

XSeed oil extraction Vegetable oil from seeds is traditionally produced by liquid-solvent extraction; as a rule, hexane is used to extract the oil from ground seeds. The process is very efficient, but its major problem is hexane elimination after extraction. Three distillation units in series have to be used, operated under vacuum. The drawbacks of this process are the high costs, the possible thermal degradation of the oil, and the incomplete hexane elimination (from 500 to 1000 p.p.m. residue).

Therefore, several authors have proposed substituting the traditional process by SC-CO_2 extraction of oil from seeds. Indeed, tryglicerides forming seed oils are readily soluble in SC-CO_2 at 40 °C and at pressures higher than about 280 bar. The main parameters to be taken into account for this process are particle size, pressure and residence time. Small particles (1 mm mean diameter or less) and high pressures (300–500 bar) can strongly reduce the extraction time. After the extraction, the SC-CO_2-tryglicerides solution is sent to a separator working at a subcritical pressure. This operation reduces the solvent power of CO_2 to almost zero and allows the recovery of oil. The complete elimination of gaseous CO_2 from oil is also carried out in the separator. The SCFE of several seed oils has been successfully performed up to the pilot scale. An alternative process has also been proposed, in which the operation is performed at a fixed pressure, and only temperature variations are used to recover the oil.

A comprehensive mathematical model of oil extraction has also been validated in which the oil-bearing seed structures, their characteristics, and the experimental data from several authors have been considered to obtain a general validity representation of the process. (*See* **Vegetable Oils**: Oil Production and Processing.)

Extraction from Liquid Mixtures

When liquid compounds have to be extracted from a liquid mixture, it is not possible to use the fixed-bed extractor. In this case, a packed tower operating in continuous mode has to be used. The packing is formed by an inert material characterized by a large surface area that facilitates the contact between the liquid and the supercritical fluid that are sent to the tower in the countercurrent. Typical packings are Rashig rings, perforate saddles, etc. Note that while the extraction from solids is a discontinuous operation, the packed tower is capable of a continuous steady-state operation, which allows large quantities of liquid mixtures to be processed in a relatively small apparatus and in a short time.

Selection of Operating Conditions

The typical extraction apparatus used in this case is shown in Figure 2. Two pumps deliver the liquid solution and SC-CO_2 to the packed column. SC-CO_2 flows along the column from the bottom to the top, whereas the liquid solution is usually added to the top, though it is also possible to feed it at an intermediate position along the column.

The process is based on the different solubilities of liquids to be treated in SC-CO_2. The ideal case is obtained when only the compounds to be extracted are soluble in SC-CO_2, whereas all the other liquid components are completely insoluble. However, this case is rare, and as a rule, the limited solubility of the other liquid compounds forming the mixture has to be taken into account. For this reason, the appropriate pressure and temperature of the process have to be chosen to select the conditions under which there is a maximum difference in solubility among the compounds to be extracted and all the other compounds in the mixture. Also, in this case, CO_2 density is frequently used as a criterion to find the conditions of maximum selectivity.

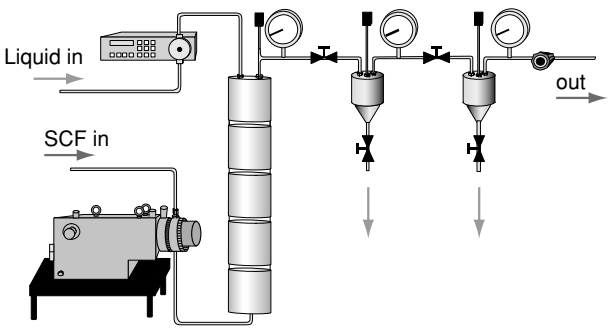

Figure 2 Extraction scheme for extraction from liquid mixtures.

The difference in density between the liquid and SCF is in this case another parameter to be taken into account: to allow the countercurrent operation, the SCF density has to be lower than that of the liquid.

The traditional operation of packed columns requires that the liquid flow rate will be greater than the minimum amount that ensures complete wetting of the packing. The feed ratio is also selected to avoid the massive entrapment of the liquid in the fluid phase (flooding). Of course, these conditions have to be respected also when a supercritical fluid is used as the fluid processing medium. The classical calculation in terms of the number of theoretical equilibrium stages required for separation can also be applied.

A possible variation of this processing scheme can involve the adoption of a temperature profile along the column with the aim of optimizing the separation temperature with respect to the composition of the mixtures at different levels inside the column.

The extraction of liquid mixtures is controlled by the relative solubilities in SC-CO_2 of the various compounds forming the mixture, that is, the thermodynamic limitation of the process. Mass transfer between the two phases represents the kinetic limitation. The distance from the equilibrium condition represents the driving force for the separation along the column.

Applications of Liquid Extraction

The fractionation of a liquid mixture by SCFE has been proposed for various applications. A selection of the most relevant examples is given in Table 3. Two examples will be discussed: hexane elimination from vegetable oils, representing a new approach to seed oil extraction, and citrus essential oil deterpenation, used to extend the shelf-life of these products.

Hexane elimination from traditionally extracted seed oils This process represents a different approach to seed oil extraction by SC-CO_2. Indeed, the supercritical extraction applied directly to seeds has the drawback of high costs owing to the use of large pressurized vessels and the discontinuous operation.

If the approach is to treat the liquid–hexane oil mixture after traditional hexane extraction, it is

Table 3 Applications of liquid fractionation by SC-CO_2

Oil deacidification
Solvent elimination (hexane from vegetable oil)
Wine dealcoholation
Milk fat fractionation
Deterpenation of citrus oils
Recycling of frying oils (fatty acid methyl esters and peroxide elimination)

possible to use a smaller apparatus and a continuous operation that allows large quantities of raw mixture to be processed.

Hexane and triglycerides are both soluble in SC-CO_2, but at a temperature of 40 °C, hexane is completely miscible at pressures higher than about 100 bar, whereas, at the same temperature, tryglicerides show a nonnegligible solubility in SC-CO_2 only for pressures higher than 250 bar. Therefore, these compounds can be separated if the operation pressure is considerably lower than 200 bar.

For example, the fractionation of soybean oil–hexane mixture has been successfully performed using a 1.8-m-tall packed tower, charged with stainless steel perforated saddles with a very large specific surface. The extraction has been performed at 120 bar and 40 °C using different starting concentrations of soybean oil in hexane up to 70:30 hexane–oil concentration by weight, which corresponds to the composition of the mixture at the exit of the traditional liquid extraction vessels. With a single passage in the packed tower, it has been possible to obtain a bottom product formed by soybean oil containing down to 20 p.p.m. of hexane residue and to recover pure hexane at the top of the column.

Citrus essential oils deterpenation Citrus essential oils are produced by cold pressing the peel of citrus fruits. These contain very large quantities of hydrocarbon terpenes (mainly limonene) up to 99% in some cases, and small quantities of oxygenated terpenes. Whilst limonene undergoes degradation and produces an unpleasant flavor, oxygenated terpenes make a large contribution to the characteristic citrus flavor and fragrance. Therefore, in order to prolong the shelf-life of these products and concentrate their fragrance, hydrocarbon terpene elimination (deterpenation) processes have been proposed. The most commonly used processes to deterpenate peel oils are vacuum distillation, extraction with alcohol, and partitioning between two solvents. The disadvantages of these processes are contamination with organic solvents and thermal degradation of the products.

The alternative to traditional processes is SCF deterpenation in a packed tower operating in continuous mode, taking advantage of the different solubilities of hydrocarbon and oxygenated terpenes in SC-CO_2 to fractionate the mixture. This separation is more difficult to perform than that described in the previous paragraph for soybean oil, since both hydrocarbon and oxygenated terpenes are very soluble in SC-CO_2, and their solubilities are very similar.

Successful processing has been proposed, operating at 40 °C and at pressures near the critical value of CO_2 (from 75 to 90 bar) to obtain the maximum possible selectivity (the maximum difference in solubility between two compound families can be obtained in this range). Nevertheless, no operating conditions exist at which the less soluble oxygenated terpenes are not soluble in SC-CO_2. Thus, fractionation is possible only using very tall columns and with a fractionation efficiency lower than 100%.

For these reasons, some authors have proposed another SC processing scheme, based on the adsorption of the citrus oil on an adsorbent (silica gel) and the fractional desorption by SC CO_2. The process is very similar to SCFE; it has the advantage that it is an enhanced separation owing to the different strength of the adsorption of hydrocarbon and oxygenated terpenes on silica-gel active sites.

Spray extraction

A liquid mixture can also be formed by a liquid that contains one or more solid solutes. The extraction of these compounds from the liquid cannot be performed using a packed tower, since the solid matter will precipitate inside the tower. It cannot be recovered, and the internal packing is made dirty by this material. Therefore, in the case of solid-containing liquid mixtures, a supercritical spray extraction process has to be adopted. The preconditions to apply spray extraction are that the liquid solvent has to be very soluble in SC-CO_2, whereas the solids have to be insoluble in the SC fluid. These conditions can be frequently obtained, since many organic solvents are readily soluble in SC-CO_2, even under mild operating conditions (e.g., a CO_2 density lower than 0.6 g cm^{-3}), and many high-molecular-weight solids have negligible solubilities in SC-CO_2, especially at low CO_2 densities.

The supercritical spray extraction process consists of a precipitation vessel previously charged with SC-CO_2 (and in which the supercritical fluid continues to flow), in which a spray of the liquid solution is continuously injected. If the process conditions have been properly selected, the liquid is rapidly dissolved in the SC fluid, whereas the solid precipitates at the bottom of the precipitation vessel. A possible representation of the process is shown in Figure 3; two pumps deliver the liquid solution and SC fluid, respectively. A precipitation vessel is used to collect the solid, and a liquid solvent recovery vessel located downstream the precipitator and operated at lower pressure (for example, 30 bar, 25 °C) is used to recover the liquid.

Selection of the Operative Conditions

One of the key steps in this process is the spray formation. The aim of this operation is to produce a

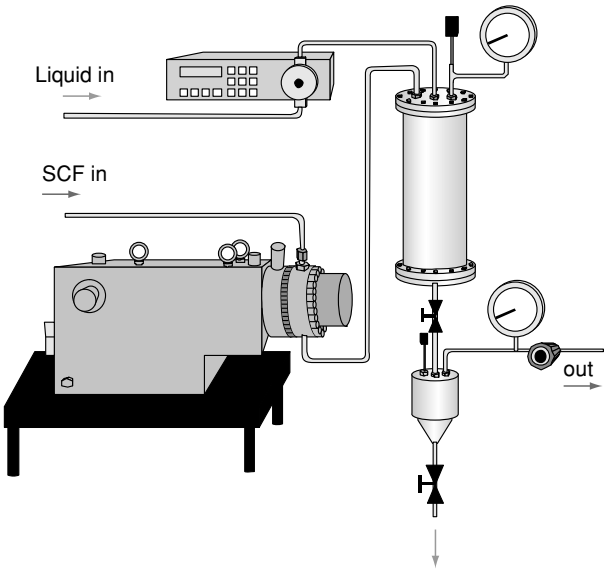

Figure 3 Apparatus used in spray extraction.

very large liquid surface forming small liquid droplets. The large surface area between the liquid and SCF strongly enhances the solubilization of the liquid phase in the supercritical medium and reduces the time required for the complete elimination of the liquid phase to a few seconds or less. For this reason, the process is performed under operating conditions at which the liquid is completely soluble in SC-CO_2. Knowledge of the solubility data on the liquid solvents and of the solids in SC-CO_2 is mandatory for this process for the appropriate selection of process temperature and pressure.

Also, in the case of SC spray extraction, there is the interaction between thermodynamics constraints and mass-transfer mechanisms that control the process performance. The enhanced mass transfer that characterized SC fluids is again a distinctive advantage in their adoption as extraction media, together with the rapid and complete separation by simple depressurization between the SC solvent and the liquid.

A limitation of this process is the possible formation of a ternary liquid–solid–SC-CO_2 mixture. Indeed, the liquid can induce a nonnegligible solubility of the solid compounds in SC-CO_2. In this case, the liquid can act as a cosolvent from the point of view of the solid solubilization. When this phenomenon occurs, the part of solid retained in the fluid phase obviously does not precipitate and is lost in the liquid recovered in the separation vessel. However, this liquid can be used to solubilize the solid compound again before the SC spray extraction. The limit case is represented by the complete entrapment of the solid in the fluid phase that produces the process failure.

Applications of Spray Extraction

Until now, spray extraction has been used in a limited number of processes, but it has a great potential for future applications. Several possible applications are listed in Table 4.

Two examples will be described in detail: lecithin extraction from seed oil and citric acid purification. The first example is a classical spray extraction with the solid solute that is completely insoluble in SC-CO_2. Citric acid purification introduces a multistep spray extraction in which different compounds are precipitated in sequence.

Lecithin extraction from soybean oil Lecithin is formed by a mixture of phospholopids and is a valuable nutraceutical compound. Its major source is soybean (1.1–3.2 wt%). It is an important natural emulsifier used in food and pharmaceutical industries and can be extracted from soybean oil. However, as long as the oil is extracted, the liquid becomes very viscous because of the concentration of lecithin. (*See* **Phospholipids:** Properties and Occurrence.)

Some authors have proposed using a spray extraction process and using SC-CO_2 as the spray extraction solvent, since seed oils are soluble in SC-CO_2 (though at pressures higher than about 280 bar, as previously specified), and phospholipids are completely insoluble in this medium.

Thus, the proposed process consists of a precipitation vessel pressurized with SC-CO_2 (at 400 bar and 40 °C, for example), in which CO_2 flows continuously and an injection system that sprays the lecithin-containing oil into the precipitator in the form of small droplets. The oil solubilizes very rapidly in SC-CO_2 and is extracted and recovered in a separator operating at low pressure. Lecithin precipitates in the precipitation vessel as a solid powder and is collected when the vessel is discharged.

Citric acid purification Citric acid is an important compound that is used in food and beverages as an acidulant. It is generally produced by fungal fermentation, and the purification process includes several treatments with chemical reagents. The spray extraction process has been proposed to purify citric acid after extraction from the fermentation broth using an organic solvent (e.g., ethanol or acetone). Two kinds

Table 4 Applications of SC-CO_2 spray extraction

Citric acid extraction and purification
Phospholipids from soybean oil and egg yolk
β-Carotene from oxides
Protein purification
Di- and triglyceride fractionation

of impurities are coextracted by acetone: sugars and by product organic acids like oxalic and malic acid. Therefore, the process involves the fractional precipitation of sugars that precipitate at low pressures (about 20 bar), the pressure is then increased, and citric and malic acid precipitate in the same range of pressures. Oxalic acid is retained in the solution and precipitated only at very high pressures (higher than 120 bar). Therefore, the purification of citric acid is not complete: the only residual impurity being malic acid, which has a very low concentration in the starting solution. (*See* **Acids:** Natural Acids and Acidulants.)

See also: **Citrus Fruits**: Composition and Characterization; **Essential Oils**: Properties and Uses; Isolation and Production; **Soy (Soya) Bean Oil**; **Vegetable Oils**: Types and Properties; Oil Production and Processing

Further Reading

Brunner G (1994) *Gas Extraction*. New York: Steinkopff Darmstadt Springer.

Bruno TJ and Ely JF (1991) *Supercritical Fluid Technology*. Boca Raton, FL: CRC Press.

Clifford T (1999) *Fundamentals of Supercritical Fluids*. Oxford: Oxford Science Publications.

King MB and Bott TR (1993) *Extraction of Natural Products using Near-critical Solvents*. Glasgow, UK: Blackie Academic & Professional.

Kiran E, Debenedetti PG and Peters CJ (eds) (2000) *Supercritical Fluids: Fundamentals and Applications*. Dordrecht, The Netherlands: Kluwer Academic.

Luque de Castro MD, Valcarcel M and Tena MT (1994) *Analytical Supercritical Fluid Extraction*. Berlin: Springer Verlag.

McHugh MA and Krukonis VJ (1986) *Supercritical Fluids Extraction: Principles and Practice*. Boston, MA: Butterworths.

Reverchon E (1997) Review: Supercritical fluid extraction of essential oils and related materials. *Journal of Supercritical Fluids* 10: 1–38.

Reverchon E and Marrone C (2001) Modelling and simulation of the supercritical CO_2 extraction of vegetable oils. *Journal of Supercritical Fluids* 19: 161–175.

Smith RM and Hawthorne SB (1997) *Supercritical Fluids in Chromatography and Extraction*. Amsterdam: Elsevier Science.

Sulfur (Sulphur) Dioxide *See* **Preservatives**: Classifications and Properties; Food Uses; Analysis

Surface-ripened Cheeses *See* **Cheeses**: Mold-ripened Cheeses: Stilton and Related Varieties

Surfactants *See* **Emulsifiers**: Organic Emulsifiers; Phosphates as Meat Emulsion Stabilizers; Uses in Processed Foods

Surgical Nutrition *See* **Burns Patients – Nutritional Management**

Swedes *See* **Vegetables of Temperate Climates**: Commercial and Dietary Importance; Cabbage and Related Vegetables; Leaf Vegetables; Oriental Brassicas; Carrot, Parsnip, and Beetroot; Swede, Turnip, and Radish; Miscellaneous Root Crops; Stem and Other Vegetables

Sweetbread *See* **Offal**: Types of Offal

SWEETENERS

Contents
Intensive
Others

Intensive

A Bassoli and L Merlini, Università di Milano, Milan, Italy

Background

The sensation derived from tasting a sweet substance is a pleasant one for most of the contemporary mankind. It is debatable whether this is an innate character or built on the millenary use of honey and sugar as foods. Sugar plays an important role in the human diet, being widely distributed in nature and accounting for a large portion of total nutrient intake, at least in the developed countries. The necessity, or interest, in substituting the whole, or part of, sugar in foods while maintaining the sweet taste, derived first from the requirement of reduction of sucrose in the diet of diabetics, who comprise up to 2% of the world population. Later on, from the attempt to avoid risks connected with obesity or overweight caused by the excessive caloric intake in developed countries, and to prevent dental caries. Sweeteners can be used to produce foods that are sweet yet have no sugar.

A sweetener is a substance that can be added to foods to confer a sweet taste. Sweeteners belong to two main classes: intensive and bulk sweeteners. Intensive sweeteners are substances so sweet that minute amounts can substitute substantial amounts of sucrose, with a greatly reduced caloric intake. However, this change can create difficulties in the preparation of traditional sweet foods, such as cakes or confectionery, in which sugar, often used in large amounts, confers very important properties besides sweetness, such as structure and absorbance of water. This problem can be solved using so-called bulk sweeteners, which have a chemical structure, sweetening power, and mechanical and physical properties very similar to those of sucrose.

From a nutritional point of view, sweeteners are often classified as caloric, low-caloric, and noncaloric. However, there are no great differences in the amount of calories produced by the complete consumption of similar amounts of different sweeteners. Low-caloric sweeteners owe their definition to the fact that they are not absorbed as much as sugar, and therefore, while their sweetening potency is maintained, only a part of their caloric value is utilized by the consumer. Similarly, noncaloric sweeteners are so called because they are so much sweeter than sugar, that only a minute amount is required to impart the sweetness required, and the energy produced is negligible. It should be noted that these definitions are not strict and are used randomly by different authors.

Intensive Sweeteners

The first sweetener to be used was saccharin (q.v.), the production of which began in 1884, and in the first half of the last century, production was directed almost exclusively to diabetics. In the 1950s, the demand for sweeteners, mostly for soft drinks, led to the discovery of other sweet compounds, first of all cyclamates (q.v.). These were used mainly in nonalcoholic drinks, and in the USA, in 1968, production rose to 9500 tonnes corresponding to about 300 000 tonnes of sugar. In the meantime, concern about the possible dangerous consequences of the systematic use of food additives became increasingly widespread, with calls for more accurate toxicological tests. Indeed, the results of studies aiming to demonstrate the innocuity of cyclamates were used by the Food and Drug Administration (FDA) to ban cyclamates from soft drinks in the USA in the 1970s. However, the controversy surrounding the interpretation of such studies meant that their use has continued, with repeated applications to reapprove the use of cyclamates, and they are still being used in many other countries, albeit with restricted daily intake doses.

In the last two decades, the market has called for new sugar substitutes with sensorial properties as similar as possible to those of sugar, and with unquestionable safety. At present, the introduction of a new sweetener on to the market is regulated by laws or directives of the authorities (FDA, European Community), who require, among other data, a dossier with acute and long-term toxicity studies. As a consequence of the different regulatory laws or procedures in different countries, it is possible that the use of a sweetener may not be allowed in all countries, or that the

procedures to put it into the market are much longer in one country than in others. For those approved for use, regulatory authorities, such as the FDA, Scientific Committee for Food of the European Commission (SCF), or Joint Expert Committee on Food Additives (JECFA) of the FAO/WHO, have established an acceptable daily intake (Table 1).

Design of a New Sweetener

Research on new sweet compounds for use as sugar substitutes has been constantly growing.

According to Hough, a new sweetener:

1. should be at least as sweet as sugar, colorless, odorless, with a pleasant sweet taste, and as similar to sugar as possible;
2. should be soluble in water, and chemically and thermally stable;
3. should have no toxic effect whatsoever, an expected metabolism, or be excreted unmodified;
4. should be easy to produce; if it is a synthetic compound, its purity must be guaranteed; if it is a natural compound, its sources of supply must be guaranteed;
5. should be compatible with production or application technologies; and
6. should be cheaper than those already in use, even if a better taste or other advantages can counterbalance a higher cost.

One of the major problems in the research of new sweet compounds is a lack of information on the mechanism of sweet-taste perception. Experimental evidence from biology, chemistry, physiology, psychology, and neurology supports the hypothesis of the existence of one or more receptor proteins on the taste buds that should mediate the chemoreception mechanism. The first results in cloning candidate genes for sweet taste receptors appeared in 2001, but the molecular details of the entire biological process are still unknown, and so the question of 'how to design a new intensive sweetener' is still open. The rational design of new molecules generally starts from known sweet natural compounds or synthetic sweet molecules. Most of the commercially successful intensive sweeteners (such as aspartame and saccharin) have been discovered by chance and/or designed from a rational and systematic modification of known molecules. In this case, the first step is the identification of those molecular fragments that are important in the receptor–sweetener interaction, the so-called glucophores. This aim is obtained *via* a systematic process of structural modification by synthesis and a critical comparison of the structure–activity relationships of the derivatives obtained. Successful examples of this procedure are sucralose and neotame.

Perceptual Characteristics of Sweeteners

Perceptual characteristics, which include sensory and hedonic, i.e., pleasant aspects of sweeteners, play a major role in food selection and intake. Beyond intensity of sweetness, other characteristics, such as the time–intensity profile, bitterness, other aftertastes (metallic, sour, etc.), fragrance, 'body' (viscosity), and freshness, influence the perception and therefore the acceptability and/or the preference for a sweet material. These characteristics are evaluated by sensory analysis (q.v.), performed by a trained panel of tasters, following established procedures, and with statistical treatment of data. In tasting a sweetener, the form in which it is presented, concentration, temperature, and pH of the solution, are important factors. Important characteristics are those related to mouth feel. A sweetener with a negative heat of solution gives a sensation of cool if tasted as a solid, that is not given by its aqueous solution. A diluted solution of an intensive sweetener has a low 'body,' which, on the contrary, is perceived when tasting a more concentrated (and more viscous) solution of a sugar alcohol, with a sweetness comparable with that of sucrose. The overall description of the taste quality of a sweetener is usually reported as a 'spider web' diagram (Figure 1), representing the mean scores for different attributes as determined by the sensory analysis.

The onset of the sweet taste and the change of its intensity over time, i.e., the temporary profile of a sweetener, is very important in determining the overall quality of the substance. The taste perception is a time-dependent phenomenon related to the time after which the sweetness is perceived (lag time), duration, and speed of the increase and decrease of the stimulus itself. These parameters are described in a time–intensity profile (Figure 2), which is characteristic and different for each sweet substance.

In the example, the time–intensity profile of sucrose and an intensive sweetener are compared. The two substances have similar lag times, and the sweet taste is immediately perceived; but in sucrose, the rate of increase and especially the decrease of the stimulus are higher than in the intensive sweetener. An intensive sweetener usually shows a longer duration of the stimulus, i.e., a prolonged onset of the sweet sensation (or lingering) that can sometimes be an undesirable side-effect. The maximum intensity of the sweet taste (I_{max}) is similar, but the overall effect on the taste of the two substances is quite different. Each sweetener has its own taste profile, which depends also on the presence of other sweeteners, taste modifiers, or flavors in the same mixture.

Table 1 Characteristics of various sweeteners

Sweetener	*RS*[a]	*Limitations*	*Status (updated to spring 2001)*	*ADI*[b] *(mg per kilogram of body weight)*	*Notes*
Acesulfame K	130–200	At high concentrations, may have a slight aftertaste	Approved by SCF[c] and JECFA[d]. Approved in more than 90 countries world-wide	15 (JECFA); 9 (SCF)	Flavor enhancer. Good shelf-life; thermal resistance. Synergistic with other sweeteners
Alitame	2000–3000	On long-term storage, can impart an off-taste	Approved in Australia, Chile, Colombia, Indonesia, New Zealand, Mexico, and the People's Republic of China. Approval is also being sought in the USA, Europe, and other countries	0.1–1	Clean sweet taste. Sweetness profile close to that of sugar. Heat stability; synergistic
Aspartame (q.v.)	200	Limited stability to acidic pH and high temperatures	Approved by SCF[c] and JECFA[d]. Approved by the US FDA	40	Produces phenylalanine during metabolism: consumption limited for people suffering from phenylketonuria. Synergistic, flavor enhancer for citrus and other fruits
Cyclamate (q.v.)	30–50		Approved by SCF[c] and JECFA. A petition for the reapproval of cyclamate is under review by the US FDA	11 (JECFA); 7 (SCF)	Good shelf-life; synergistic; economic
Neotane	7000–13 000	Slow degradation at very low pH	Approved by FDA and ANZA[e]	6	High quality sweetness, flavor enhancer, excellent stability
NHDC	400–600	At high concentrations, exhibits a long-lasting sweetness associated with a menthol- or licorice-like aftertaste	Approved by SCF[c]. Approval for food use in countries outside the European Union has been granted or is being sought	0–5 (SCF)	Obtained from bitter flavonoids in orange peels. Remarkable synergy with acesulfame and aspartame. Flavor enhancer, bitter taste inhibitor
Saccharin (q.v.)	300–500	Metallic aftertaste	Approved by SCF[c] and JECFA[d]. Approved in more than 90 countries world-wide	5 (JECFA, SCF)	The first calorie-free sweetener discovered in 1879
Stevioside	100–150	Licorice aftertaste	*Stevia* extracts are approved for food use in several South American and Asian countries, but lack approval in Europe and North America and at the international level	Not determined	Extracted from the leaves of the *Stevia rebaudiana* plant
Sucralose	600	Sucralose can hydrolyze slowly in solution, under extreme conditions of acidity and temperature	Sucralose is currently approved for use in foodstuffs in more than 35 countries, including the USA and Japan	0–15	High-quality sweetness, good water solubility and excellent stability in a wide range of processed foods and beverages
Thaumatin	2000–3000	Delayed perception of sweetness; perception lasts a long time leaving a licorice-like aftertaste at high usage levels	Approved as a sweetener by SCF[c] and JECFA[d]. Approved as a flavor enhancer in Europe. Classified as generally recognized as safe by the FDA in the USA	Not specified (JECFA)	A mixture of three proteins of molecular weight 22 000, extracted from the fruit of West African plant *Thaumatococcus daniellii*

[a]RS, relative sweetness with respect to sucrose. This value is determined by a sensory analysis of aqueous solutions of sucrose (usually at 3, 5, or 10% by weight) taken as standard references compared with solutions of the sweetener at different known concentrations.
[b]ADI, acceptable daily intake. The ADI is the amount of a food additive, expressed on a body-weight basis, that can be consumed in the diet every day throughout life without any appreciable health risks. It is in fact a safe intake level.
[c]SCF, the Scientific Committee for Food of the European Commission. Reference legislation is contained in the *European Parliament and Council Directive 94/35/EC of 30 June 1994 on sweeteners for use in foodstuffs*.
[d]JECFA, the Joint Expert Committee on Food Additives of the FAO/WHO.
[e]ANZA, Food Standards Australia New Zealand.

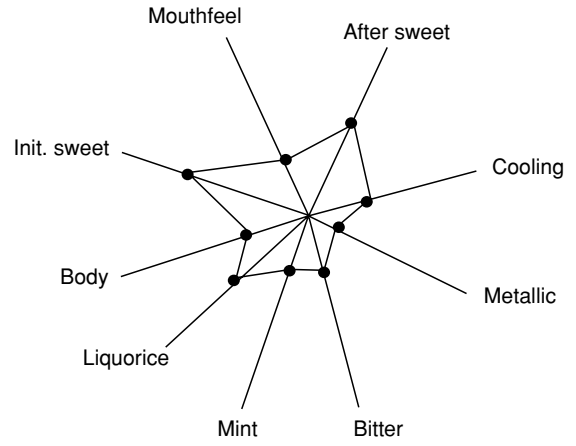

Figure 1 A 'spider web' representation of some perceptual characteristics of an intensive sweetener as determined by the sensory analysis.

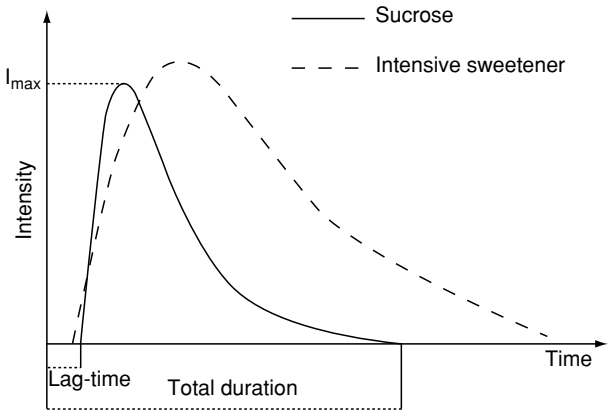

Figure 2 Schematic time–intensity profiles of sucrose and of an intensive sweetener.

Synergy

Many sweeteners show a synergistic effect when used in mixtures, i.e., the taste intensity of the mixture is higher than the sum of the intensities of the single components. This phenomenon is of practical importance, since it offers several advantages. First, the use of mixtures of intensive sweeteners helps in approaching the optimal sucrose taste, which can be mimicked by varying the components' concentrations in order to approach the desired time–intensity profile of the mixture. Sweetener mixtures can also have lower costs, especially if synergy is at play, resulting in a lower daily consumption of the individual additives in food. Manufacturers can overcome the limitations of individual sweeteners by using them in blends; mixtures of saccharin and cyclamate have been used in a number of commercial products.

Table 1 and Figure 3 provide details and structures, respectively, of the intensive sweeteners currently used in many countries.

Acesulfame K

Discovered in 1967 by Hoechst, acesulfame K is structurally related to saccharin. When used in high concentrations above normal use, it may have a slight aftertaste.

Applications Acesulfame K has a good shelf-life and is very stable with normal preparation and processing of foods; it is heat-resistant and therefore suitable for cooking and baking.

Safety Acesulfame K is not metabolized by the body and is excreted by the kidneys unchanged. A large number of safety studies have been conducted, and no adverse effects have been reported.

Status Acesulfame K has been approved for a variety of uses in more than 90 countries. In 1998, the FDA broadened the US approval of acesulfame K to allow its use in nonalcoholic beverages.

Alitame

Discovered by Pfizer Inc., alitame is a high-intensity sweetener formed from the amino acids L-aspartic acid and D-alanine, and an amine derived from thietane. The aspartic acid component is metabolized normally, and alanine amide is not hydrolyzed any further.

Applications Alitame has the potential to be used in almost all areas where sweeteners are presently used. It has an excellent stability at high temperatures, and so it can be used in cooking and baking. During long-term storage, some soft drinks sweetened with alitame can develop an off-taste.

Safety Extensive animal and human studies have supported the safety of alitame.

Status Alitame has been approved for use in a range of foods and beverages in Australia, Chile, Columbia, Indonesia, New Zealand, Mexico, and the People's Republic of China. Approval is also being sought (2000) in the USA, UK, Canada, Brazil, Europe, and other countries.

Aspartame (q.v.)

Aspartame, discovered by Nutrasweet in 1965, is the most commonly used intensive sweetener of the new generation. The compound is made by coupling two

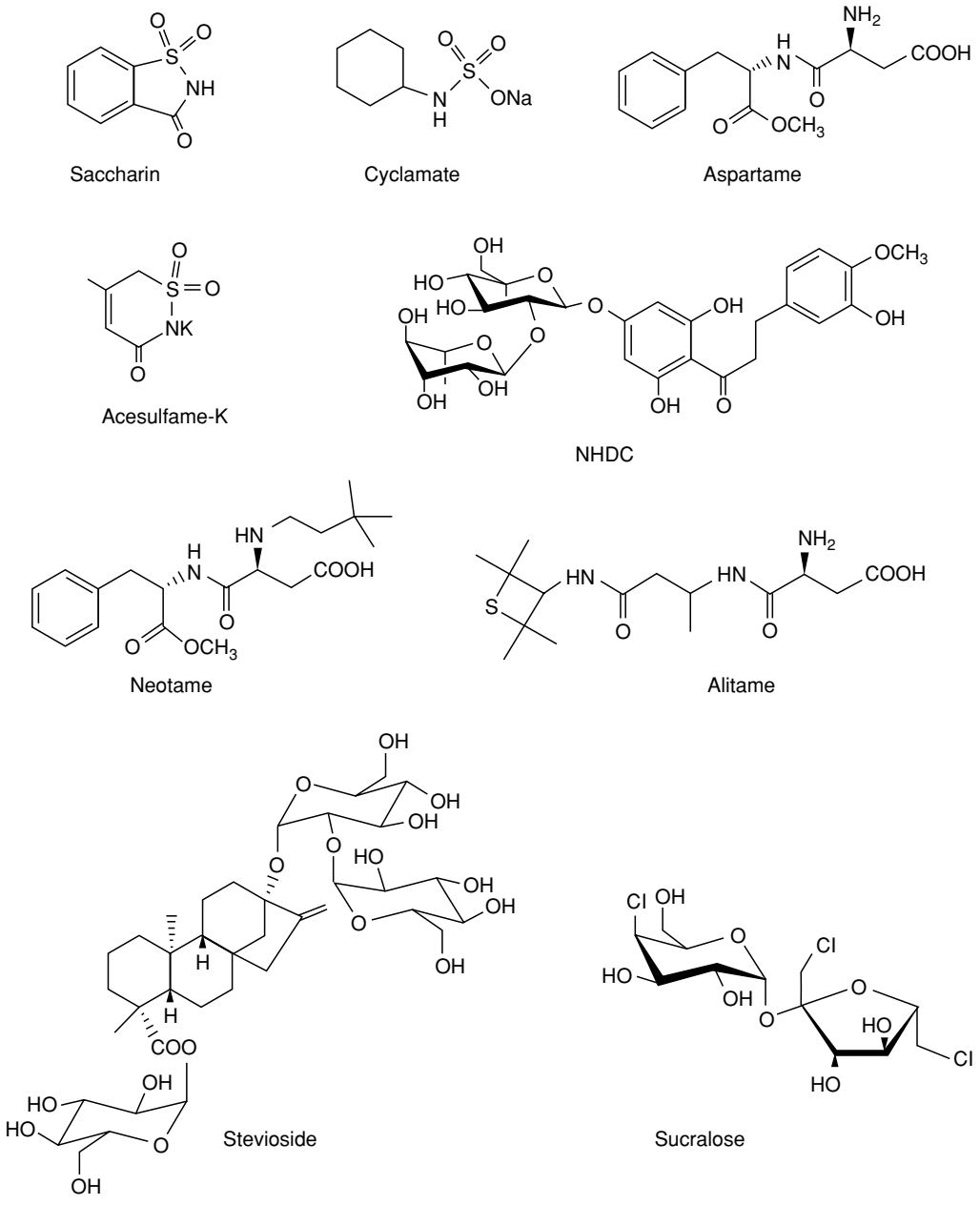

Figure 3 Structure of the most common sweeteners.

essential amino acids, aspartic acid and phenylalanine, found naturally in most protein-containing foods, including meats, dairy products, and vegetables. Upon digestion, aspartame breaks down to phenylalanine, aspartic acid, and a small amount of methanol, this last in levels that are insignificant compared with those of many natural foods. Aspartame enhances and intensifies flavors, particularly citrus and other fruits.

Applications Aspartame is used to sweeten a variety of foods and beverages, and as a table-top sweetener. It is currently used in well-known brands of a great variety of foods and drinks.

Safety Aspartame is one of the most thoroughly tested food ingredients. Aspartame is safe and approved for people with diabetes, pregnant and nursing women, and children.

Status Aspartame has been approved in more than 90 countries and is widely used throughout Eastern and Western Europe, the USA, Canada, South America, Australia, and Japan.

Cyclamate (q.v.)

Cyclamate was discovered in 1937. It is metabolized to a limited extent in the gut by some individuals, shows limited absorption by the body, is excreted unchanged by the kidneys, is stable at high and low temperatures, has a good shelf-life and pleasant taste profile, and is suitable for cooking and baking.

Applications Cyclamate, particularly in combination with one or more other low-calorie sweeteners, has a wide range of applications in foods and beverages.

Safety The SCF confirmed the safety of cyclamate in 1994, as did the Cancer Assessment Committee of the FDA in 1984 and the US National Academy of Sciences in 1985.

Status Cyclamate has been approved in more than 50 countries world-wide. A petition for the reapproval of cyclamate is still (as at July 2001) held in abeyance by the FDA.

Neohesperidinedihydrochalcone (NHDC)

NHDC is a low-calorie sweetener and flavor enhancer that can be produced by hydrogenation of neohesperidine, a flavonoid occurring naturally in bitter oranges. At high concentrations, NHDC exhibits a long-lasting sweetness associated with a menthol- or licorice-like aftertaste. NHDC is not absorbed to any significant extent, but it is metabolized by the intestinal flora, yielding the same or similar breakdown products to its naturally occurring analogs.

Applications NHDC is typically used in combination with other sweeteners, with remarkable synergistic effects. Even at very low concentrations (5 p.p.m.), NHDC can still improve the overall flavor profile and mouth feel of certain foods, acting as a flavor enhancer and modifier rather than as a sweetener. It also has bitterness-reducing properties. NHDC is stable in solid form and in aqueous solutions of pH 1–7. It is heat-stable and therefore can be used in foods requiring pasteurization or UHT processes.

Status NHDC is a permitted sweetener in Europe. Approval for food use outside the European Union has been granted, or is being sought.

Neotame

It is a derivative of a dipeptide, structurally related to aspartame. Its relative sweetness is much higher, being 7000–13 000 times that of sucrose and c.40 times that of aspartame. Neotame has been developed by Monsanto.

Applications Neotame can be used across many food categories, including beverages, dairy products, frozen desserts, baked goods, and gums. It also has flavor-enhancing properties, especially for mint.

Status In July, 2002, FDA approved neotame for use as a general-purpose sweetener. Its use is also permitted in Australia and New Zealand.

Saccharin (q.v.)

Discovered in 1879, saccharin has been used commercially to sweeten foods and beverages since the turn of the twentieth century. It is absorbed slowly, not metabolized, rapidly excreted unchanged by the kidneys, highly stable with a long shelf-life, suitable for cooking and baking, does not promote tooth decay, and is suitable for diabetics.

Applications Saccharin has the widest range of applications and is used in a great variety of foods.

Safety Saccharin has a history of a century of safe human use and is probably the most thoroughly researched of all food additives. Its safety was questioned in a 1977 Canadian study that found bladder tumors in male rats, albeit given unrealistically high doses. All scientific research conducted since then has shown that this effect is only seen in male rats at extremely high doses and has supported the safety of saccharin for human use at the levels currently consumed. Several human studies have shown no overall association between saccharin consumption and cancer incidence.

Status Saccharin has been approved in more than 90 countries. In the USA, the FDA proposed a ban on saccharin in 1977, on the basis of the aforementioned high-dose rat studies. The proposed ban was formally withdrawn in 1991, but saccharin and foods and drinks containing saccharin were still required to carry a warning label. In 2000, legislation gave saccharin a clean bill of health and removed the label.

Stevioside

Stevioside is extracted from the leaves of *Stevia rebaudiana* Bertoni, a plant growing in South America and several Asian countries. Stevioside is a glycoside consisting of the aglycone steviol (ent-13-hydroxykaur-16-en-19-oic acid) and three glucose molecules. The sweetness of stevioside is accompanied by a licorice-like aftertaste.

Applications Leaves of the stevia plant have been used for centuries in Brazil and Paraguay to sweeten

foods and beverages. Stevioside can be used in soft drinks, Japanese-style vegetable products, table-top sweeteners, confectionery, fruit products and seafood, and in some countries as stevioside-rich *Stevia* extracts.

Safety and status Stevia extracts have been approved for food use in several South American and Asian countries but lack approval in Europe and North America and on an international level. Safety studies on stevioside and *Stevia rebaudiana* have not been accepted internationally, owing to the lack of generally accepted specifications. In 1999, the SCF reiterated the opinion that "stevioside is not acceptable as a sweetener on the presently available data." The JECFA reviewed stevioside in 1998 but could not quantify an acceptable daily intake (ADI) because of inadequate data on the composition and safety of stevioside. In 2000, the European Commission refused a request for marketing authorization for *Stevia rebaudiana* plants and dried leaves. Stevioside, as a sweetener, is not permitted in the USA and may not be used or marketed in Europe.

Sucralose

Sucralose is the common name for 4,1′,6′-trichlorogalactosucrose, a high-intensity sweetener derived from ordinary sugar, developed jointly by McNeil Specialty Products and Tate & Lyle. Sucralose is not metabolized.

Applications Sucralose has a high-quality sweetness, good water solubility, and excellent stability in a wide range of processed foods and beverages. Like sugar, sucralose hydrolyzes in solution, but only over an extended period of time under extreme conditions of acidity and temperature.

Safety Extensive studies have shown that it is safe for human consumption.

Status Sucralose is currently approved for use in foodstuffs in more than 35 countries.

Thaumatin

Thaumatin is a low-calorie protein sweetener and flavor modifier extracted from Katemfe, the fruit of the West African plant *Thaumatococcus daniellii*, and is totally natural with an intense sweetness. A limitation of thaumatin is its delayed perception of sweetness, but perception lasts a long time, leaving a licorice-like aftertaste at high usage levels. It is metabolized by the body like any other dietary protein.

Applications Thaumatin is stable in freeze-dried form and is soluble in water and aqueous alcohol. It is heat- and pH-stable and synergistic when combined with other low-calorie sweeteners. Thaumatin has a wide range of applications in food and drinks and is particularly effective for its flavoring properties and because it adds mouth feel.

Safety A large number of animal and human studies have been conducted, showing no adverse reactions. Thaumatin is classified as generally recognized as safe by the FDA. The JECFA gave thaumatin an ADI of 'not specified.'

Status Thaumatin is a permitted sweetener and has been approved in all applications in the European Community as a 'flavor preparation.' Similar approval exists in Switzerland, the USA, Canada, Israel, Mexico, Japan, Hong Kong, Korea, Taiwan, Vietnam, Australia, New Zealand, and South Africa, and further approval is being sought elsewhere.

Uses

Food Intensive sweeteners have several practical applications. Their main use is in food, where they are used as additives in several products: table-top sweeteners, carbonated beverages (soft drinks), noncarbonated beverages (e.g., coffee drinks, alcoholic drinks, cider, juices, fruit nectars, shakes, ice tea, instant beverages), dairy products (yogurt, icecream), desserts (puddings, jellies, gelatins, flans), marmalade and jams, chocolate, fruit preserves, fruit spreads, baked goods (confectionery, biscuits, breakfast cereals), chewing gum, pickled vegetables, marinated fish, savories, sauces, and salad dressings.

They are also used in the preparation of dietetic products or functional foods such as vitamin- and mineral-fortified products, sport drinks, and low-fat products.

Pharmaceuticals and dental decay prevention In pharmaceuticals, sweeteners are used as additives in drugs for oral administration, especially to mask the bitter or unpleasant taste of certain active ingredients. They are widely used in the preparation of toothpaste and mouthwash, for improving taste and preventing dental decay. In fact, dental caries is the result of the interaction of sucrose or other carbohydrates with oral bacteria, leading primarily to dental plaque formation and secondarily to tooth decay, as a result of the alterations of dentin caused by acidic fermentation end products and inflammatory processes induced by bacterial toxins.

The potentiality of alternative sweeteners to reduce dental caries has attracted much interest and many

studies. All the intensive sweeteners and sugar alcohols (q.v.), such as xylitol, sorbitol, and mannitol, are noncariogenic, since they are not fermented by oral bacteria.

Economics

The world market for sweeteners has tended to increase broadly in line with the world economy. The average growth rate in the 1990s was around 2%. The world consumption of intensive sweeteners reached 12 million tonnes of sugar equivalents in 1997, with a share of the total sweeteners market of *c*. 9%. Asia accounts for 50–55% of the total, with the rest being shared almost equally between the Americas and Europe. More than 70% of the market by quantity (sugar equivalents) is represented by saccharin, aspartame having 18% and cyclamates 5%, whereas if the consumption is measured by value, aspartame accounts for 62%, saccharin 17%, cyclamates 5%, and others 16%. The high share of saccharin is due to its cheapness, making it suitable for use in many applications where only a generic sweetener is required, and for nonfood use, such as pharmaceuticals, animal feed, and toothpaste. Another reason is its stability under a wide variety of conditions, together with a high-quality sweetness. All this explains why it has a high share (90% or more) of the Asian market, whereas aspartame is mostly used in North America, followed by Europe. As for other, recently introduced, intensive sweeteners, the use of acesulfame-K is increasing, especially in blends for soft drinks, whereas the natural compound stevioside is mostly used in Asia.

Bulk Sweeteners

In sweet foods, sucrose provides not only sweetness but also, in many cases, structure, weight, and volume. This 'bulk' effect must be maintained in sugarless products to insure that those properties make the product acceptable to the consumer and to maximize the shelf-life. Moreover, the substitution of sugar must minimize the changes in the technological processes of production. In products where this bulk effect cannot be provided only by water (drinks) or partially by air (e.g., in icecreams), such as confectionery, chocolates, hard and soft candies, jellies, chewing gums, dragées, and jams, the so-called bulk sweeteners can be used. The most important of these are glucose, fructose, lactose, galactose, maltose (q.v.), and sugar alcohols (q.v.), these last obtained by hydrogenation of carbohydrates. They should have a relative sweetness and rheological properties at least as good as that of sugar. (*See* **Sugar Alcohols**.)

See also: **Acesulfame/Acesulphame**; **Aspartame**; **Cyclamates**; **Isomalt**; **Saccharin**; **Sensory Evaluation**: Sensory Characteristics of Human Foods; Taste; **Sucrose**: Properties and Determination; **Sugar**: Refining of Sugarbeet and Sugarcane; **Sugar Alcohols**

Further Reading

Corti A (ed.) (1999) *Low-calorie Sweeteners Present and Future: World Conference on Low-calorie Sweeteners*. Basel: Karger.

Gilbertson TA and Kinnamon SC (1996) Making sense of chemicals. *Chemistry and Biology* 3: 233.

Grenby TH (ed.) (1987) *Development in Sweeteners*, vols. 1–3. London: Elsevier Applied Science.

Grenby TH (ed.) (1989) *Progress in Sweeteners*. London: Elsevier Applied Science.

Guesry PR and Secrétin M (1991) *Sugars and non-nutritive sweeteners*. In: Gracey M, Kretchmer N and Rossi E (eds) *Sugar in Nutrition*. Nestlé Nutrition Workshop Series, vol. 25, pp. 33–53. New York: Raven Press.

Hough L (ed.) (1997) Plenary Lectures from the Jerusalem International Symposium on Sweeteners. *Pure & Applied Chemistry* 69: 655–727.

Laing GD and Jinks A (1996) Flavour perception mechanisms. *Trends in Food Science and Technology* 7: 387–388.

Mathlouthi M, Kanters A and Birch GG (eds) (1993) *Sweet Taste Chemoreception*. London: Elsevier Applied Science.

O'Brien L and Gelardi RC (eds) (1991) *Alternative Sweeteners*. New York: Marcel Dekker.

Schiffman SS (1991) *Receptors and transduction mechanisms for sweet taste: an overview*. In: Gracey M, Kretchmer N and Rossi E (eds) *Sugar in Nutrition*. Nestlé Nutrition Workshop Series, vol. 25, pp. 55–67. New York: Raven Press.

Shallenberger RS (1994) *Taste Chemistry*. London: Blackie & Hall.

Van der Wel H, Van der Heijden A and Peer HG (1997) Sweeteners. *Food Reviews International* 3: 193–268.

Walters DE, Orthoefer FT and DuBois G (eds) (1991) *Sweeteners: Discovery, Molecular Design and Chemoreception*. Washington, DC: American Chemical Society.

Others

M B A Glória, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

Background

Sweetness is one of the most important taste sensations for humans. Sucrose has been widely used for its

sweetness, functional properties such as texture and mouth feel, and as a bulking agent and preservative. However, the special dietary needs of diabetics and the health concerns about obesity and dental caries have prompted considerable efforts for the development of alternative sweeteners.

Sweeteners are widely used in the food, beverage, confectionery, and pharmaceutical industries throughout the world. Because of high consumer demand and acceptance of low-calorie products, the market of artificially sweetened foods has increased significantly and will continue to grow.

Sweeteners can be classified into two categories, bulk and intense. The bulk sweeteners are used as sweeteners and as bulking agents and can have a preservative and bodying effect. They are metabolized by the body and provide calories and include glucose, fructose, maltose, hydrolyzed products from starch and sugar alcohols. They vary in sweetness over a narrow range, from 0.3 to 1.2 times the sweetness of sucrose. Bulk sweeteners are permitted in a number of specified foodstuffs at *quantum satis* – as much as needed.

The intense sweeteners have a sweet taste but are effectively noncaloric. They can be natural or synthetic compounds, and they are intensely sweet, ranging from 30 to 3000 times the sweetness of sucrose, so the concentrations used for normal sweetness are very low. This category includes saccharin, cyclamate, acesulfame-K, aspartame, alitame, dulcin, sucralose, neohesperidin dihydrochalcone, glycyrrhizin, stevia sweeteners, and thaumatin. The first four compounds are the main sweeteners consumed and are presented in other sections. Dulcin (4-ethoxyphenylurea, $C_9H_{12}N_2O_2$, MW 180.20) was synthesized by Berlinblau in 1884. It is 70–350 times sweeter than sucrose. However, it should not be used as a food additive, since it has been reported to cause cancer in laboratory animals. The purpose of this chapter is to provide general information on the other intense sweeteners.

The sweeteners have different chemical structures (Figures 1–5) and, therefore, diverse properties (Table 1). Maximum levels have been established for intense sweeteners in different food categories, depending on the legislation in a particular country.

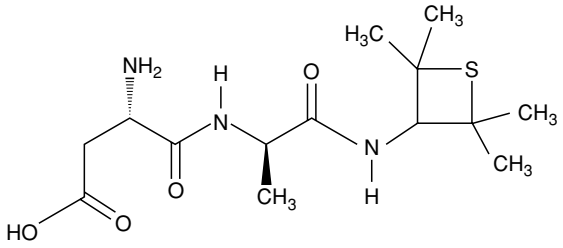

Figure 1 Chemical structure of alitame.

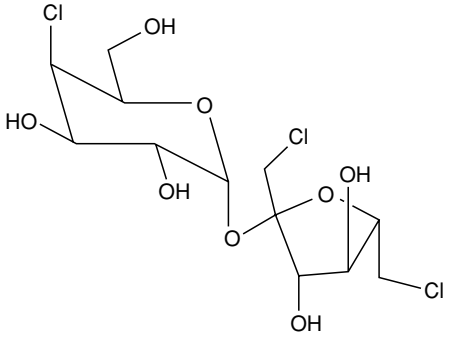

Figure 2 Chemical structure of sucralose.

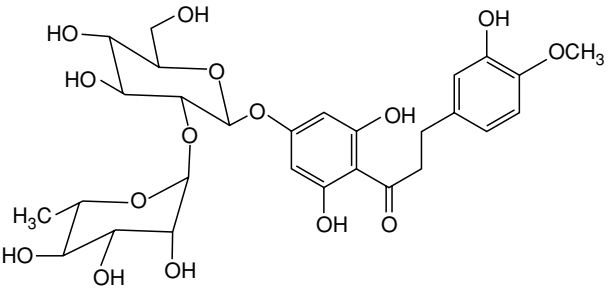

Figure 3 Chemical structure of neohesperidin dihydrochalcone.

Alitame

Alitame [L-α-aspartyl-*N*-(2,2,4,4-tetramethyl-3-thioethanyl)-D-alaninamide] is an amino acid-based sweetener (Figure 1) developed by Pfizer Central Research from L-aspartic acid, D-alanine, and 2,2,4,4-tetraethylthioethanyl amine. A terminal amide group instead of the methyl ester constituent of aspartame was used to improve the hydrolytic stability. The incorporation of D-alanine as a second amino acid in place of L-phenylalanine has resulted in optimum sweetness. The increased steric and lipophilic bulk on a small ring with a sulfur derivative has provided a very sweet product and good taste qualities.

The formula of alitame is $C_{14}H_{25}O_4N_3S$, with a molecular weight of 331.06. It is produced under the brand name Aclame®. It is a crystalline, odorless, and nonhygroscopic powder, with a good solubility in most polar solvents such as water (130 g l^{-1} at pH 5.6) and alcohol (Table 1). Alitame is 2000 times sweeter than sucrose, 12 times sweeter than aspartame and six times sweeter than saccharin. It has a

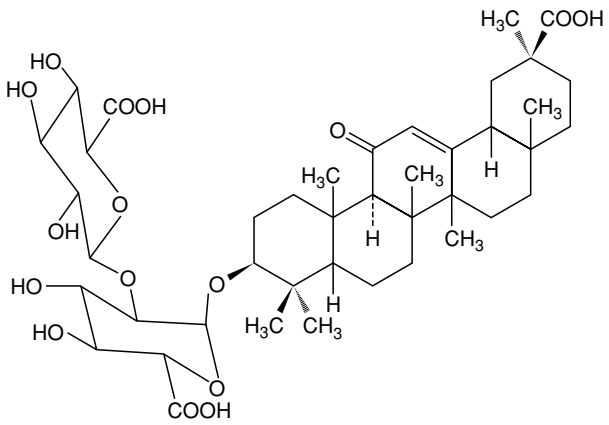

Figure 4 Chemical structure of glycyrrhizin.

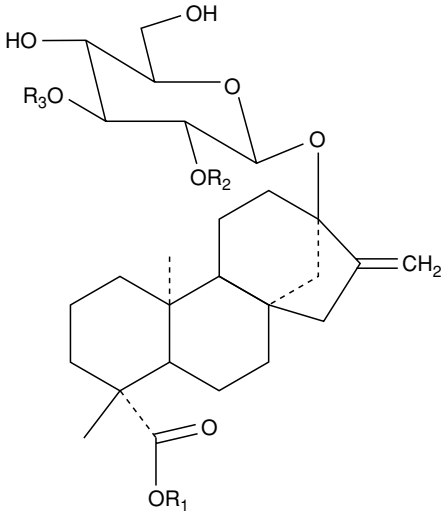

Figure 5 Chemical structure of the sweet glycoside of *Stevia rebaudiana* Bert.

clean sweet taste with no unpleasant aftertaste. It is blended with other sweeteners such as, saccharin, cyclamate, and acesulfame-K to maximize the quality of sweetness.

Alitame offers several benefits such as stability at high temperatures and a broader pH range. For instance, it is stable for over a year at pH 6–8 and room temperature and withstands pasteurization. However, prolonged storage of acidic solutions at high temperatures or in combination with certain ingredients (hydrogen peroxide or sodium bisulfite) may produce off-flavors. In the presence of high levels of reducing sugars, alitame can undergo Maillard reactions.

Alitame is noncariogenic. From an oral intake, 7–22% is unabsorbed and excreted in the feces. The remainder is hydrolyzed to aspartic acid and alanine amide. The aspartic acid is normally metabolized, and the alanine amide is excreted in the urine as a sulfoxide isomer, sulfone, or conjugated with glucuronic acid. The incomplete absorption and metabolism result in a core value of 1.4 kcal g^{-1}.

The Joint Expert Committee on Food Additives (JECFA) concluded that alitame was not carcinogenic and did not show reproductive toxicity. In 1996, an acceptable daily intake (ADI) of 0–1 mg per kilogram of body weight was allocated. It is approved for use in Australia, New Zealand, Mexico, and China. A food additive petition was submitted to the US Food and Drug Administration (FDA) in 1986, and approval is awaited. In the petition, the estimated daily intake is 0.34 mg per kilogram of body weight, which represents the amount if alitame is the only sweetener in the diet. The level at which no observed adverse effects occur in animals is 100 mg kg^{-1}. Potential uses include baked goods, baking mixes, hot and cold beverages, dry beverage mixes, table-top sweeteners, chewing gum, candies, frozen desserts, and pharmaceuticals.

Sucralose

Sucralose, 1,6-dichloro-1, 6-dideoxy-β-D-fructofuranosyl 4-chloro-4-deoxy-α-D-galacto-pyranoside or 4,1′,6′-trichloro-4,1′,6′-trideoxy-*galacto*-sucrose (Figure 2), is a chlorinated derivative of sucrose, discovered in 1976 by carbohydrate research chemists at Queen Elizabeth College and Tate and Lyle, UK. It is derived from a patented multistep process, involving selective chlorination of sugar at the 4, 1′, and 6′ positions substituting three hydroxyl groups on the sucrose molecule. It is the result of a study on a large number of related compounds, carefully synthesized and evaluated to determine the spatial structure and molecular configuration required for sweetness perception.

Ingredients and tabletop forms of sucralose are being marketed under the brand name Splenda®. Its chemical formula is $C_{12}H_{19}O_8Cl_3$ (MW 397.35). Sucralose is a white, odorless crystalline powder and is readily dispersible and soluble in water, methanol, and ethanol. At 20 °C, a 280 g l^{-1} solution of sucralose in water is possible. Sucralose presents Newtonian viscosity characteristics, a negligible lowering of surface tension, and no pH effects, and its solubility increases with increasing temperature. In ethanol, the solubility ranges from approximately 110 g l^{-1} at 20 °C to 220 g l^{-1} at 60 °C, and its solubility in ethanol facilitates in formulating alcoholic beverages and flavor systems.

Table 1 Properties of some intense sweeteners

Sweetener (INS)[a]	*Sweetness*[b]	*Sweetness characteristics*	*Synergism*[c]	*Solubility in water*	*Stability*		*ADI*[d] *(mg per kilogram of body weight)*
					In solution	*During heating*	
Alitame	2000	Clean, no unpleasant aftertaste	sac, cyc, aces	Good, 130 g l^{-1}, pH 5.6	Good, pH 6–8	Very good	0–1 (1996)
Sucralose (955)	400–800	slow onset, clean sweet sugar-like, prolonged sweetness	cyc, aces, nhdc	Good, 280 g l^{-1}, 20 °C	Good, pH > 3, loss of sweetness, pH < 3	Very good	0–15 (1990)
Neohesperidin dihydrochalcon (959)	250–2000	delayed onset, lingering licorice–menthol-like aftertaste	sac, asp, aces, cyc, sucral, sugar alcohols	Low, 0.5 g l^{-1}, 20 °C, 650 g l^{-1}, 80 °C	Stable, pH 2–6	Good	Not evaluated by JECFA, 0–5 (1987–SCF[d], EC)
Glycyrrhizin (958)	50–100	Slow onset, long aftertaste, licorice flavor	stev, thau, asp	Good	Stable, pH > 4.5	Good	Not evaluated by JECFA
Stevioside	100–300	Slow onset, menthol at high levels, bitter	asp, cyc, aces, glyc	Low, 0.8 g l^{-1}	Stable, pH 3–10	Good	Not evaluated by JECFA
Thaumatin–Talin (957)	2000–3000	Slow onset, persistent, licorice-like	sac, aces, asp, cyc, stev, glyc	Good, 600 g l^{-1}	Stable, pH 2.7–6.0	Stable, neutral to low pH	Not specified (1985)

[a]INS, International Numbering System.
[b]In relation to sucrose.
[c]sac, saccharin; cyc, cyclamate; aces, acesulfame-K; nhdc, neohesperidin dihydrochalcone; asp, aspartame; sucral, sucralose; stev, stevioside; thau, thaumatin; glyc, glycyrrhizin.
[d]SCF, Scientific Committee for Food.
From Glória MBA (2000) Intense sweeteners and synthetic colorants. In: Nollet LML (ed.) *Food Analysis by HPLC*, pp. 523–573. New York: Marcel Dekker.

Sucralose is 400–800 times sweeter than sucrose (Table 1). Although its sweetness varies with pH, sucralose has a clean sugar-like taste and a time–intensity profile like that of sucrose, albeit more persistent. It has an excellent taste profile and no bitter or objectionable aftertaste. It is a flavor enhancer and shows sweetness synergism with cyclamate, acesulfame-K, and neohesperidin dihydrochalcone.

Sucralose offers a broad pH, aqueous, thermal processing, and shelf stability. It does not interact with food ingredients and is stable in the dry form (4 years at 20 °C). It withstands high temperatures, thus making it well suited for use in pasteurized, aseptic processing, sterilized, cooked, and baked foods. However, under extreme conditions of pH, temperature, and time, sucralose may be hydrolyzed, producing 4-chloro-deoxy-D-galactose and 1,6-dideoxy-1,6-dichloro-D-fructose, or degraded with elimination of hydrogen chloride in basic medium.

Sucralose is noncariogenic. It resists hydrolysis in the human digestive tract, being excreted unchanged in the feces, and the very small portion absorbed is rapidly eliminated in the urine. Therefore, it produces no glycemic response and is virtually noncaloric. Following safety testing and toxicological studies in humans and animals, the FDA concluded that sucralose does not pose any carcinogenic, reproductive, or neurological risk. The JECFA reviewed it favorably and in 1990 recommended an ADI of 0–15 mg per kilogram of body weight. Sucralose is approved for use in a wide range of food products in Canada, USA, Australia, Mexico, Russia, Romania, China, the European Union, and Mercosur. It has been used as a table-top sweetener and in carbonated, still, and alcoholic beverages, frozen desserts, confectionery, bakery products, canned fruits and vegetables, fruit spreads, chewing gum, dry-mix products, dairy products, condiments, dressings, and breakfast cereals.

Neohesperidin Dihydrochalcone

Citrus fruits contain bitter flavanone glycosides, all derivatives of the disaccharide 2-*O*-α-L-rhamnopyranosyl-β-D-glucopyranose, neohesperidose. In 1963, Horowitz and Gentili found that catalytic hydrogenation of the chalcone form gave dihydrochalcone neohesperidosides, several of which were intensely sweet. Numerous dihydrochalcone derivatives were synthesized for taste and toxicity trials, from which neohesperidin dihydrochalcone (NHDC) emerged as a promising sweetener. It is prepared by alkaline hydrogenation of the biflavanoid neohesperidin present in Seville (bitter) oranges (*Citrus aurantium*).

NHDC is a semisynthetic nonnutritive intense sweetener. Chemically, it is 1-[4-[[2-*O*-(6-deoxy-α-L-mannopyranosyl)-β-D-glucopyranosyl]oxy]-2,6-dihydroxyphenyl]-3-(3-hydroxy-4-methoxyphenyl)-1-propanone with the molecular formula $C_{28}H_{36}O_{15}$ and a molecular weight of 612.60 (Figure 3). NHDC shows a slow buildup of sweetness, rising from 250 to 2000 times that of a 50 g l^{-1} sucrose solution, but more persistent (Table 1). It has a pleasant taste, flavor-modifying properties, ability to improve the sweetness quality and profile, and remarkable synergistic effects. Its flavor enhancement has been perceived in several products, specially fruit flavors. NHDC has the ability to decrease the perception of bitterness, saltiness, sharp, and spicy attributes. The sweetness intensity of NHDC depends on many factors such as concentration, pH, and the product to which it is added. As the concentration increases, the sweetness of NHDC decreases relative to the level of sucrose. Caffeine significantly enhances the sweetness of NHDC in certain soft drinks. At higher concentrations, NHDC has a lingering menthol or licorice-like aftertaste and a cooling sensation, which distinguishes it from other sweeteners. However, modifications of the sensorial properties of NHDC are possible by the admixture of bulk sweeteners, certain flavors, and other taste-modifying food additives such as gluconates, amino acids, or nucleotides. It also shows synergism with saccharin, aspartame, cyclamate, sucralose, acesulfame-K, and sugar alcohol.

NHDC is a nonhygroscopic colorless crystalline solid. It is sparingly soluble in water (0.50 g l^{-1}) at 20 °C but is highly soluble at 80 °C (650 g l^{-1}). It is also soluble in alcohol and aqueous alkali, but a higher solubility is achieved in ethanol–water mixtures than in water or ethanol alone. Where a higher solubility is required, monobasic salts may be used that are freely soluble in water and exhibit a shorter duration of sweetness than the parent compound. The solubility of NHDC may also be enhanced by dissolving it in glycerol and propylene glycol, as well as by using it in mixtures with readily water-soluble polyols such as sorbitol. Interestingly, these bulk sweeteners also act as taste modifiers of NHDC by reducing its menthol-like aftertaste.

NHDC presents high stability at pH 2–6. It is stable under most food-processing and storage conditions and withstands pasteurization, UHT processes, and the normal shelf-life of soft drinks. It is stable during fermentation of yogurt, but undergoes hydrolysis at high acidity and elevated temperatures, yielding hesperetin dihydrochalcone, hesperetin dihydrochalcone-4′-β-D-glucoside, rhamnose, and glucose.

NHDC is noncariogenic and has a caloric value of 2 kcal g^{-1}. Little of the compound is absorbed

unchanged from the small intestine. After cleavage of the glycosidic side-chain by intestinal mucosal or bacterial glycosidases, the residual primary metabolites are partly excreted unchanged in the bile and partly metabolized further. Standard toxicity tests have suggested its safety. In 1987, the Scientific Committee for Food of the Commission of the European Communities allocated an ADI of 0–5 mg per kilogram of body weight. It is currently approved for use in the European Communities, Sweden, Switzerland, Morocco, and Tunisia.

Owing to its highly intense and long-lasting sweetness, NHDC is normally used at concentrations of less than 100 mg kg^{-1}. Only in chewing gum are higher levels required because of its slow release from the gum base. For use in soft drinks, NHDC has been recommended in combination with other sweeteners at a concentration of 20 mg kg^{-1}. Owing to its ability to reduce bitterness and to its flavor enhancing properties, NHDC is an ideal sweetener for grapefruit or orange juice. Promising results have been reported from its use in fruit-flavored yogurts. In table-top products, the addition of small amounts of NHDC may result in significant savings because of its synergistic, sweetness enhancing effect. It has been used in juice, soft drinks, dairy products, desserts, confectionery, spreads, jams, chewing gum, chocolate based products, and icecream.

Glycyrrhizin

Glycyrrhizin is found in licorice root of a small leguminous shrub, *Glycyrrhiza glabra* L. from Europe and Central Asia. Glycyrrhizin or 20-β-carboxy-11-oxo-30-norolean-12-en-3β-yl-2-*O*-β-D-glucopyranurosyl-α-D-glucopyranosiduronic acid ($C_{42}H_{62}O_{16}$, MW 822.92), is a triterpenoide glycoside (saponin) with glycyrrhetinic acid, which is condensed with *O*-β-D-glucuronosyl-(1′→2)-β-D-glucuronic acid (Figure 4). After harvest, the roots are dried to 10% moisture, shredded, extracted with aqueous ammonia, concentrated in vacuum evaporators, precipitated with sulfuric acid, and crystallized with 95% alcohol providing a crude ammonium glycyrrhizin (AG). Further treatment yields a white, crystalline mono-ammonium glycyrrhizin (MAG). Both derivatives have the same sweetness but differ in solubility and sensitivity to pH. AG is relatively stable and highly soluble in hot or cold water and in alcohol. It withstands temperatures above 105 °C for a short period of time, and precipitates at pH values below 4.5. MAG is used in applications where low pH and color rule out AG.

Glycyrrhizin is 50–100 times sweeter than sucrose and has a slow onset of sweetness followed by a lingering licorice-like aftertaste (Table 1). It exhibits a sweet woody flavor, which limits its use as a pure sweetener. Glycyrrhizin enhances food flavors, masks bitter flavors, and increases the perceived sweetness level of sucrose. It has the potential for providing functional characteristics, including foaming, viscosity control, gel formation, and possibly antioxidant characteristics.

Studies have focused on the pharmacological effects of glycyrrhizin as antiulcer, antiinflammatory, antiviral, anticariogenic, and antispasmodic. It also has corticoid activity, influencing steroid metabolism to maintain blood pressure and volume and to regulate glucose/glycogen balance. Glycyrrhizin can be hydrolyzed by human intestinal microflora to 18-β-glycyrrhetinic acid and two molecules of glucuronic acid. After release of the acids, the compound binds to plasma protein, enters the enterohepatic circulation, and is almost completely metabolized. However, side-effects, typically involving cardiac dysfunction, edema, and hypertension have been reported among subjects receiving high doses of glycyrrhizin-based pharmaceuticals or consuming large amounts of licorice-containing confectionery or health products over a prolonged period.

The ammonium salt of glycyrrhizin is approved as a flavoring and flavor enhancer in the USA. It is on the FDA 'generally recognized as safe' (GRAS) list. The use of glycyrrhizin is permitted in Japan and Taiwan. In Japan, it is used as a flavoring for hydrolyzed vegetable protein, soy sauce, and bean paste to control saltiness. At levels of 30–300 mg kg^{-1}, it enhances the flavor of cocoa and chocolate-flavored products, flavors and sweetens candy, confectionery, and beverages, and masks the bitter taste of pharmaceuticals. Because of its pharmacological action, it should be used in moderate amounts as a sweetener.

Stevia Sweeteners

Stevia rebaudiana Bertoni (Compositae) is a herb native to Paraguay but cultivated in South-east Asia, Japan, Paraguay, Brazil, Israel, and USA. Stevia sweeteners are extracted from the dried leaves, clarified and crystallized. The sweet constituents of stevia include eight diterpene glycosides: stevioside, steviolbioside, rebaudiosides A, B, C, D, and E, and dulcoside A, which collectively are 100–300 times sweeter than sucrose. They are similar in structure in that a steviol aglycon is connected at C-4 and C-13 to mono-, di-, or trisaccharides consisting of glucose and/or rhamnose residues, as shown in Figure 5. Stevioside is the major constituent, whereas dulcoside A and rebaudiosides A and C are the other main components. Stevioside is 300 times sweeter than

sucrose (Table 1). It shows a sweetness profile similar to that of sucrose, except that it has an unpleasant, persistent, menthol, and bitter aftertaste. However, the development of new cultivars, derivatization and incorporation of cyclodextrin, L-histidine, potassium phosphate, glucono-δ-lactone and maltose in formulations has eliminated undesirable aftertastes. Rebaudioside A is more stable, sweeter, and has a better taste profile than stevioside. The remaining diterpene glycosides are not as sweet as stevioside.

Stevioside is a white powder, highly soluble in water, ethanol, and methanol, and is nonfermentable. When heated at 100 °C for 1 h, solutions of stevioside at pH 3–9 show little loss in sweetness and no change at 22 °C for 5 months. However, considerable decomposition occurs at pH 10. Some degradation of stevioside and rebaudioside has been observed in carbonated beverages acidified with phosphoric and citric acids during storage at 37 °C. Heating at 60 °C for 6 days has resulted in 0–6% loss of sweetness. Exposure to 1 week of sunlight does not affect stevioside, but results in 20% loss of rebaudioside A. The high stability of stevioside makes it a suitable sweetener for cooked and baked foods and for beverages.

Stevioside suppresses the growth of oral microorganisms, and both stevioside and rebaudioside A provide very few calories. In Paraguay, *S. rebaudiana* is used for the treatment of diabetes because of its hypoglycemic activity. Studies have suggested that stevioside is not toxic, mutagenic, or teratogenic in a number of animal species. In addition, the product has been used for more than 10 years in South America and Japan, but there are contradictory reports on the *in-vivo* metabolism to steviol, which is mutagenic for *Salmonella typhimurium* TM677.

Stevioside is available in three purity ranges: crude extract, and 50% and 90% purity. Since 1970, stevia sweeteners have been used in a wide range of food and beverage applications in Japan, including soft drinks, candies, chocolate, chewing gum, icecream, yogurt, jam, pudding, and table-top sweeteners. It is commonly used in combination with sucrose and fructose and also with other sweeteners such as aspartame, cyclamate, and acesulfame-K, but not with saccharin. It is currently approved for use in Japan, Taiwan, and Mercosur. An acceptable daily intake of stevioside for humans of 7.94 mg per kilogram of body weight has been suggested.

Thaumatin

Thaumatin is a mixture of sweet proteins originally isolated from the fruit of the West African plant katemfe (*Thaumatococcus danielli* Benth.). There are at least five thaumatins, which are obtained by extraction with water, concentration, and ultrafiltration. Tate and Lyle Ltd. manufactures a mixture of two as talin. These are proteins with isoeletric points of 11.5–12.5. Thaumatin consists of a single chain of 207 normal amino acid residues with eight disulfide bonds and has a molecular weight of about 22 000. It is very soluble in water ($600\,g\,l^{-1}$) and is stable at pH 2.7–6.0 and under pasteurization conditions (Table 1). At higher pH values, the protein becomes more heat-labile, despite its stability at pH values up to 8 under ambient conditions. The stability of solubilized thaumatin requires careful monitoring of pH, time, temperature, and other processing parameters. In addition, protection of the protein against yeasts and molds in solutions is also critical for stability. Thaumatin can associate with negatively charged compounds such as synthetic colors and acidic gums like xanthan, pectin, carrageenan, alginate, and carboxymethylcellulose, resulting in a loss of sweetness. Association with synthetic colorants may also cause color loss.

Thaumatin is 1600–3000 times sweeter than sucrose. However, it has unusual taste profile, slow in onset, followed by intensification to lingering sweetness with a licorice-type aftertaste. It masks metallic or bitter tastes. To achieve a taste closer to that of sucrose, thaumatin must be blended with other intense sweeteners or with sugars. Synergism has been observed with saccharin, acesulfame-K, and stevioside, but not with aspartame and cyclamate. By combining thaumatin with alanine and organic acids, there is a doubling in sweetness and a reduction in aftertaste and in the delay in sweetness. At subsweetness levels, thaumatin functions as a flavor enhances. It has the ability to enhance certain flavors and aromas, such as those in peppermint, spearmint, coffee, and ginger. The synergistic effect noted with monosodium glutamate enhances the aroma and improves the flavors of processed meats and fish products when used at levels of $0.5–2.0\,g\,l^{-1}$.

Thaumatin can be metabolized to its constituent amino acids, contributing the same calories as protein. However, because of its high sweetness, it has a low-calorie value per unit of sweetness, less than 0.002 kcal. It is noncariogenic and has undergone safety tests indicating that it is not allergenic, mutagenic, or teratogenic. Furthermore, it has a long history of use without any adverse effects. In 1985, a 'not specified' ADI was allocated by the JECFA. It has been approved for use in Japan, the UK, Australia, Canada, South Africa, Mexico, the European Union, Switzerland, Taiwan, Morocco, and Tunisia. In the USA and Switzerland, it is permitted as a flavor enhancer in chewing gum, and in the USA, it has been classified as GRAS.

The major uses include chewing gum, beverages, coffee, savory flavor, dairy products, dental and pharmaceuticals, animal and pet foods. However, because of its high cost, it is used on a small scale. Biotechnological alternatives to eliminate uncertainties and variability associated with agricultural production and to lower costs are being investigated.

Blends of Sweeteners

Today, there is an increasing tendency to explore the use of blends of sweeteners. When two or more sweeteners are combined, blends with an increased stability, longer shelf-life, lowered production costs, improved taste and flavor, and decreased side- and aftertastes result, as well as a positive impact on taste quality, flavor-enhancing, and flavor-modifying properties. Also, mixtures of sweeteners can exhibit additive and synergistic effects, and since lower amounts of each sweetener can be used, the average daily intake of each sweetener is low, thus minimizing the health risk from any one sweetener.

Since food laws in most countries regulate the use of intense sweeteners, analytical control for the presence and levels of sweeteners in food is essential. According to the JECFA, it is also important to know the level of additives in food products in order to estimate the actual consumption by the population. This information shows the average intake in relation to the ADI over a period of time. Based on this knowledge, regulatory authorities can propose regulations to insure intakes below the ADI. Furthermore, some sweeteners undergo decomposition during processing and storage of food products, forming a variety of degradation products that may have sensory and toxicological significance.

Several types of methods have been described for the analysis of intense sweeteners, among them, spectrophotometric, enzymatic, titrimetric, capillary isotachophoretic, thin-layer chromatography, liquid and gas chromatographic methods. High-performance liquid chromatography (HPLC) is by far the most frequently used technique. However, with the increased number of sweeteners available and their use being approved in specified food products and beverages by different countries, methods capable of separating several sweeteners simultaneously are still required.

See also: **Sweeteners**: Intensive

Further Reading

Borrego F, Canales I and Lindley MG (1995) Neohesperidin dihydrochalcone: state of knowledge review. *Zeitschrift für Lehensmitte/untersuchung und -forschung* 200: 32–37.

Duffy VB and Anderson GH (1998) Position of the American Dietetic Association: use of nutritive and nonnutritive sweeteners. *Journal of the American Dietetic Association* 98(5): 580–587.

Dwivedi BK (1993) Artificially sweetened low-calorie foods other than drinks. In: Altschul AM (ed.) *Low-calorie Foods Handbook*, pp. 331–342. New York: Marcel Dekker.

Fenwick GR, Lutomski J and Nieman C (1995) Liquorice, *Glycyrrhiza glabra* L. – composition, uses and analysis. *Food Chemistry* 38(2): 119–143.

Giese JH (1993) Alternative sweeteners and bulking agents: an overview of their properties, function and regulatory status. *Food Technology* 47(1): 114–126.

Glória MBA (2000) Intense sweeteners and synthetic colorants. In: Nollet LML (ed.) *Food Analysis by HPLC*, pp. 523–573. New York: Marcel Dekker.

Grenby TH (1991) Intense sweeteners for the food industry: an overview. *Trends in Food Science and Technology* 2(1): 2–6.

Hollingsworth P (1998) Sucralose approval sweetens low-cal market. *Food Technology* 52(5): 34.

Hough L (1993) High-intensity, low calorie sweeteners. In: Khan R (ed.) *Low Calorie Foods and Food Ingredients*, pp. 138–164. New York: Blackie.

Liu J, Ong CP and Li SFY (1997) Subcritical fluid extraction of stevia sweeteners from *Stevia rebaudiana*. *Journal of Chromatographic Science* 35(9): 446–450.

Renwick AG (1990) Acceptable daily intake and the regulation of intense sweeteners. *Food Additives and Contaminants* 7(4): 463–475.

Sardesai VM and Waldshan TH (1991) Natural and synthetic intense sweeteners. *Journal of Nutritional Biochemistry* 2(5): 236–244.

Schaller I (1997) Is Codex the answer? *Food Ingredients Analysis International* 19(3): 24–28.

Staff Report (1996) Thaumatin – the sweetest substance known to man has a wide range of food applications. *Food Technology* 50(1): 74–75.

Verdi RJ and Hood LL (1993) Advantages of alternative sweeteners blends. *Food Technology* 47(6): 94–101.

Wallis KJ (1993) Sucralose: features and benefits. *Food Australia* 45(12): 578–580.

SWEETS AND CANDIES

Sugar Confectionery

W P Edwards, Bardfield Consultants, Gt. Bardfield, UK

Background

The confectionery industry is not a science-based industry: it is an industry that has been built on the confectioner's craft. Confectionery is normally divided into three classes: flour confectionery, chocolate confectionery, and sugar confectionery. Flour confectionery covers products that are made from flour and are baked. Chocolate confectionery consists of chocolate. Sugar confectionery covers the rest of confectionery.

Confectionery making does share some technologies with the pharmaceutical industry, specifically tabletting, panning, and lozenge making. Most of the knowledge available to early confectioners was empirical rather than scientific.

Sugar confectionery making is an international industry. Some of the terms in use are clearly French in origin such as fondant (from *fondre* to dissolve), dragees, and pastilles. Unlike chocolate confectionery, sugar confectionery has few legal definitions. In some jurisdictions, caramel has a higher milk content than toffee, but in others, the two terms are used interchangeably. The types of product known as gums were undoubtedly originally made from the gum from *Acacia senegal* but are now made from any suitable ingredient. Because sugar confectionery has a long shelf-life and does not need controlled temperature storage, an international trade in confectionery had developed by the nineteenth century.

In general, the important concepts in sugar confectionery are water activity, colligative properties, solubility, and the need to use a mixture of sugars. Toffees depend on the Maillard reaction. High boilings require a sugar glass to form. Gums, jellies, and licorice need the right rheology. Panned coatings depend on crystallization. Chewing gum is a chewable piece of polymer chemistry. Aerated products are based on stabilizing and setting a foam.

Stability

Sugar confectionery products keep well compared with most food products. The long life of confectionery products occurs because spoilage organisms are unable to multiply rather than because the product is kept sterile.

Water Activity

The relevant parameter is not only the water content but the water activity. Water activity is a thermodynamic concept that was invented to explain why materials with different levels of water content did not behave in the same way chemically or biologically. The water activity reflects the ability of the water to be used in chemical or biological reactions. It is the concentration corrected for the differences in the ability of the water to undertake chemical reactions. If a nonvolatile solute were dissolved in water, the vapor pressure should decrease in a certain way for a perfect mixture. A thermodynamically ideal substance would always have a water activity of 1. If a product is held at its own water activity, it will neither gain nor lose weight.

Equilibrium Relative Humidity (ERH)

This term is normally abbreviated to ERH. The ERH is the relative humidity that matches the water activity of the product. The ERH has practical importance, since it is an indication of the conditions under which the product can be stored without deterioration.

Colligative Properties

Boiling Points

Colligative properties are defined as those properties that depend on the number of particles present rather than the nature of the particles. In sugar confectionery, the most important of these is the elevation of the boiling point. Because sugars are very soluble, very large boiling point elevations are produced, e.g., as large as 50 °C. Remembering that the elevation of the boiling point is proportional to the concentration of the solute, it is not surprising that the boiling point is used as a measure of the concentration and hence as a process control.

The boiling point of a liquid is the temperature at which the vapor pressure is equal to the atmospheric pressure. If the pressure is increased, the boiling point will increase, whereas reducing the pressure will reduce the boiling point. Most sugar confectionery is made by boiling a mixture of sugars to concentrate them. The use of a vacuum has several advantages.

The energy consumption is reduced, browning is reduced, and the process is speeded up. A common practice is to boil a mixture of sugars under atmospheric pressure to a given boiling point. A vacuum is then applied. This causes the mixture to boil under reduced pressure. This not only concentrates the mixture while the latent heat of evaporation cools it rapidly but also speeds up the production process, since the product will ultimately have to be cooled to ambient temperature.

Solubility

Sucrose is insufficiently soluble, at ambient temperatures, to lower the water activity sufficiently to give a microbiologically stable solution. A stable product can only be made by mixing the sucrose with other sugars. The ingredient that is most used as a source of nonsucrose sugars is the starch hydrolysate known variously as glucose syrup or corn syrup. Neither name is entirely accurate as the major ingredient is normally maltose rather than dextrose, but the product can be, and is, manufactured from wheat or potato starch as an alternative to maize starch. The degree of hydrolysis is characterized by the dextrose equivalent (DE), which is a measure of the equivalence of the solids in the syrup to dextrose in the Fehlings titration. The type of syrup normally used is referred to as confectioners' glucose and has a DE of around 40.

In this work to avoid confusion between glucose syrup and chemical dextrose, glucose will only be used to refer to the syrup, and the pure sugar will be referred to as dextrose.

Originally, the starch was hydrolyzed by using acid, which is a random process. Now, hydrolysis can be carried out by enzymes or a mixture of enzymes and acid. Using enzymes, it is possible to control which bonds are broken to allow the composition of the product to be varied as needed. In principle, the starch could be hydrolyzed to produce any combination of dextrose oligomers from dextrose through maltose and maltotriose to higher oligomers.

The proportion of glucose syrup used in different types of sugar confectionery varies for a number of reasons. Commercial pressure would encourage the use of a maximum amount of glucose syrup solids, but technical considerations restrict the proportion to be used. As glucose syrup inhibits crystallization, the highest proportion is used where crystallization is undesirable, e.g., in boiled sweets, and the lowest where crystallization is desirable, e.g., fondant. Figure 1 shows the ratio of sugar to glucose used for different products.

Maillard Reaction

A major feature of sugar confectionery manufacture is ensuring that the Maillard reactions occur where they are wanted in products like toffees and do not occur in products like high boilings.

Maillard reactions are nonenzymic browning reactions. In practice, any browning in foods is a Maillard reaction except where it is enzymic, e.g., the browning of a cut apple is enzymic; hence, it is not a Maillard reaction. The Maillard reaction is not a name reaction where all the details can be found in a text book. The term covers a whole range of reactions that occur in systems ranging from food to life sciences. The name of the reaction goes back to Louis Camille Maillard, who heated amino acids in a solution with high levels of dextrose. The chemistry of the Maillard reaction is easily described as complex. It is complex not only because the reaction can give complex products but because the starting materials are themselves complex. Most model systems involve studies of one reducing sugar being heated

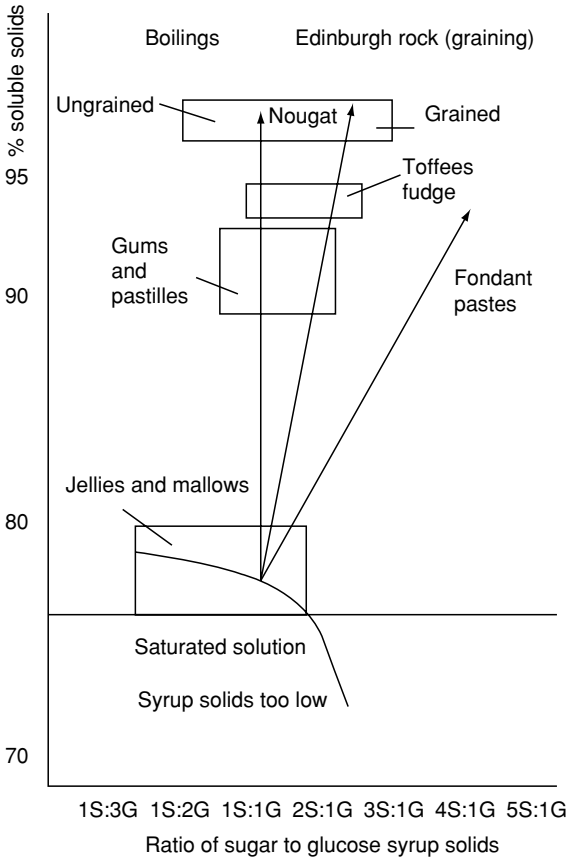

Figure 1 Sugar and glucose composition of confectionery.

with one amino acid. A typical confectionery system like a toffee would involve heating a mixture of proteins, usually from milk with a mixture of reducing sugars and fats. In sugar confectionery, the conditions of the reaction are likely to be a high temperature but a low water activity. In the early stages of the reaction, the free amino group of an amino acid, usually in a protein, condenses with the carbonyl group of a reducing sugar. The resulting Schiff bases can rearrange by Amadori (Figure 2) or Heyns rearrangement (Figure 3), the products being an N-substituted glycosylamine (if dextrose is the sugar) or an N-substituted fructosylamine for a ketose such as fructose. The *N*-glycosylamine can degrade to fission products by a free-radical mechanism. In the advanced stages of the reaction, the Amadori and Heyns rearrangement products degrade by one of three possible routes. They break down via deoxysones, fission, or Strecker degradation (Figures 4 and 5). The 1-deoxyglycosones and 3-deoxyglycosones can form reactive α-dicarbonyl compounds such as pyruvaldehyde and diacetyl by a retro-aldization reaction. These reactive intermediates are then available to react with ammonia and hydrogen sulfide.

At the end of the reaction, brown nitrogenous polymers and copolymers form. It is known that protein gels form when proteins are heated with carbonyl compounds. This could occur in toffees.

Sulfur-containing Amino Acids

Whereas sulfur free amino acids are broken down to amines via decarboxylation, the sulfur containing amino acids such as cysteine can undergo more complex reactions.

Products from Proline

Various schemes have been proposed to explain the production of nitrogen-containing heterocyclic compounds such as pyrrolidines and piperidines from proline. Nitrogen heterocyclic compounds are commonly found to be potent flavoring chemicals.

Strecker Aldehydes

These chemicals are produced by the Strecker degradation of the initial Schiff base (Figure 5). An α-amino carbonyl compound and Strecker aldehyde are generated by rearrangement, decarboxylation, and hydrolysis.

Sugar Glasses: The Chemistry of Boiled Sweets

In these products, which are variously called high boilings, hard boilings, hard candy, or boiled sweets, the sucrose is in the glassy state. The glassy state of matter is not a thermodynamic phase but a super-cooled liquid. Glassy materials are common in a number of products, whereas both natural and

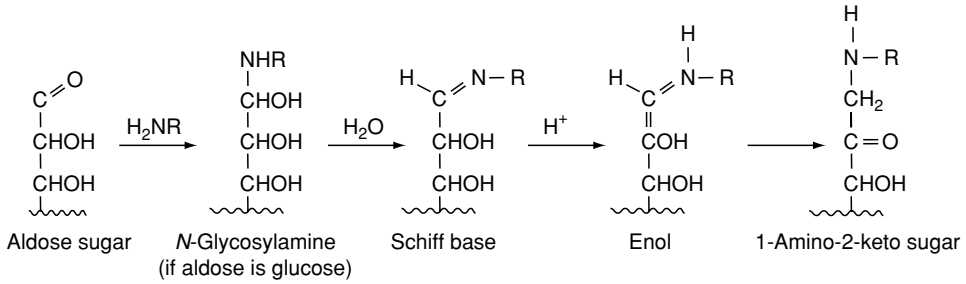

Figure 2 Amadori rearrangement.

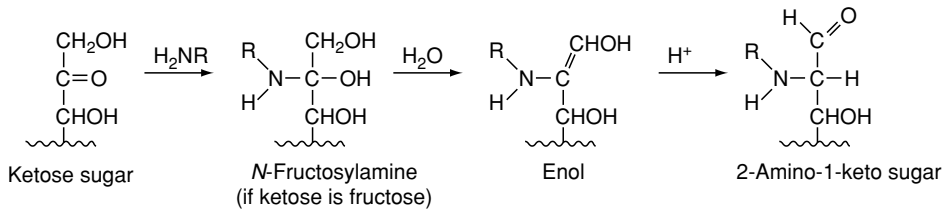

Figure 3 Heyns rearrangement.

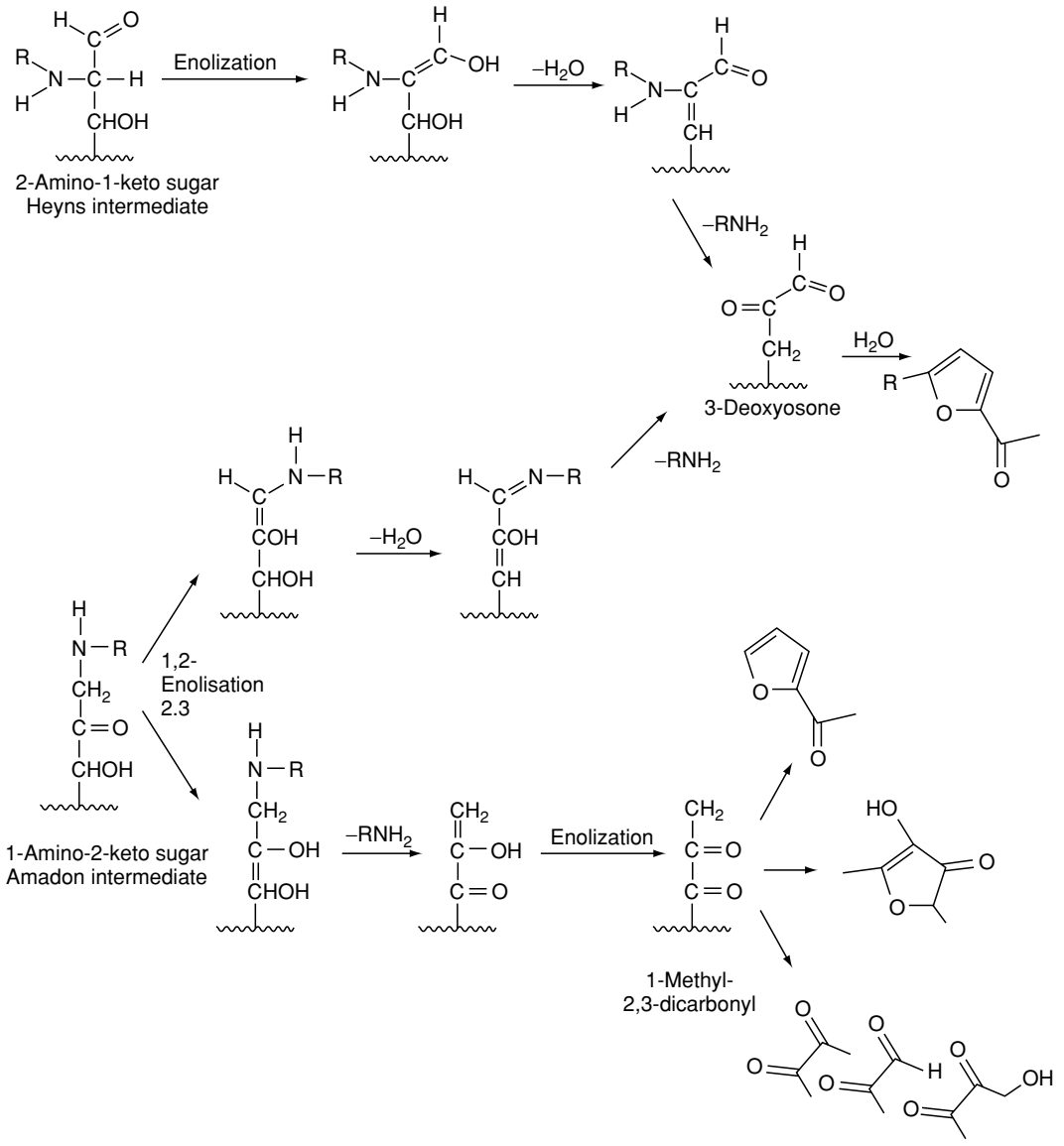

Figure 4 Formation of furans and dicarbonyls.

man-made substances can form glasses. Confusingly, although a glass is not a thermodynamic state of matter, glasses do exhibit a sharp transition temperature between the glassy and the rubbery state. Many methods have been used to determine the transition temperature between the rubbery and the glassy state. One popular method is differential scanning calorimetry (DSC). In these instruments, the sample and a blank are subjected to a change of temperature up or down at a controlled rate. The instrument measures the difference energy input or energy extracted between the sample and the control. A plot of difference gives the variation of heat capacity, C_p, with temperature. The glass transition is associated with a discontinuity in C_p. Unfortunately, there are a number of different ways in which the output from the DSC can be analyzed. Some workers would advocate integrating the result from the DSC, i.e., measuring the area under the curve. This then gives the variation of the enthalpy with temperature. The variation in several other physical parameters with temperature such as the refractive index and the dielectric constant have been used to measure glass transitions. In general, the glass transition temperature obtained depends on the method of measurement. In confectionery, the important point is that a product that is intended not to crystallize should be in the glassy state at ambient temperatures. Most sugars will form a

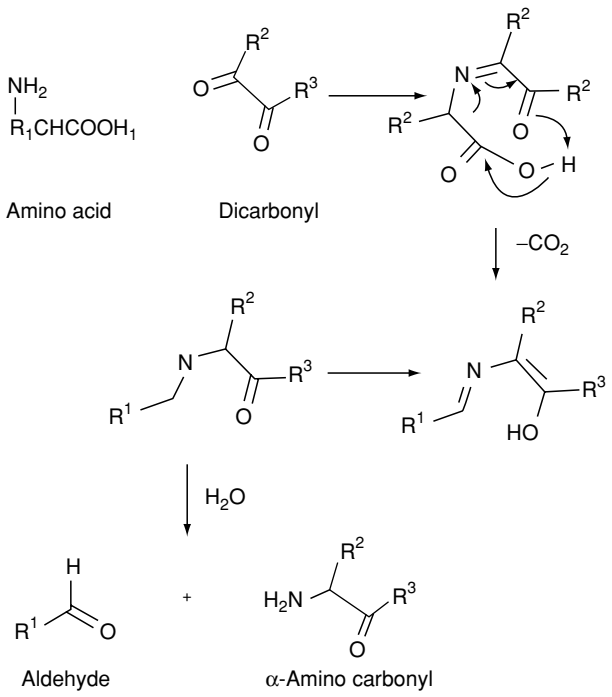

Figure 5 Strecker degradation.

glass, but pure sucrose will not form a glass. Commercial sugar glasses are always made from sucrose and some other sugar. Initially, invert sugar was used but has now been largely replaced by glucose syrup. Glucose syrup is a much better stabilizer of sugar glasses than invert sugar. In practice, boiled sweets become unstable by absorbing water. Initially, the product becomes sticky then soft, followed ultimately by crystallization. The rate at which high boilings can absorb water is diffusion-limited. The high-molecular-weight fraction in glucose syrup inhibits the migration of water into the sweet. This gives a markedly more stable product than invert sugar does. Glucose syrup also gives the boiled mass a plastic consistency when warm, which makes it much easier to handle.

Formulation of Boiled Sweets

The manufacture of stable boiled sweets depends on obtaining a product with a low moisture content. The higher the glass transition temperature, the more stable the product is. As water can be regarded as a plasticizer, it reduces the viscosity of the system and thus reduces the glass transition temperature. The water content of the finished product depends on the cooking temperature. The important parameter at the formulation stage is the ratio of sugar to glucose syrup. The situation to avoid is ending up with a product that is excessively hygroscopic, which will reduce the shelf-life by making the product sticky. Increasing the proportion of glucose syrup increases the proportion of dextrose and the high-molecular-weight dextrose oligomers. Increasing the proportion of dextrose makes the product more hygroscopic, which increases the tendency of the product to become sticky. As the proportion of high-molecular-weight material is increased, the viscosity of the liquid mass increases. These viscosity and hygroscopicity problems restrict the proportion of glucose syrup that can be used. If a conventional acid converted syrup is used, using a different DE only makes one problem or the other worse. Boiled sweets are made with ratios of sucrose to glucose of from 70:30 to 45:55. The common proportions are 60:40 and 50:50.

The high-maltose syrups break away from these restrictions, as the spectrum of carbohydrates present is modified to increase the maltose while reducing the dextrose and the high-molecular-weight fraction. Using these syrups, a high boiling can be made with approximately sucrose to glucose 35:75 with the properties of a product of a 55:45 product.

Science of Gums, Jellies, and Licorice

In these products, a colloid is incorporated to obtain the desired texture. Most of these colloids are gelling agents, e.g., gelatine, but some such as gum acacia and certain modified starches do not gel. The substances used (see Table 1) are chemically quite disparate, but they all have the ability to self-associate. An important difference is between those like gelatine, which gel thermoreversibly, and other gelling agents, which gel irreversibly, e.g., pectin. Licorice is slightly different from the other sweets in that it is made from wheat flour directly rather than with a purified starch. The important science is the texture of the product, which can be studied as rheology. If a sample is placed between the two plates (Figure 6), and one plate is oscillated, the structure of the sample will not be destroyed, but a transducer on the other element will produce an output wave form. This wave form can be analyzed in terms of an elastic and a viscous component.

Grained Products

In some products, the sucrose is intended to grain or crystallize. Examples are fondant, fudge, and montelimar. In these products, there is a high proportion of sucrose, as opposed to other sugars.

Fondant

Confectionery fondants are normally made by boiling a mixed sucrose and glucose syrup to concentrate it.

Table 1 Properties, chemistry, and sources for gums and gelling agents

Agent	*Properties*	*Chemistry*	*Source*
Gelatine	Thermoreversible gelling agent	Protein	Bovine or porcine hides or bones
Starch	Irreversible gelling agent	Carbohydrate	Maize or wheat or potatoes
High-amylopectin starch	Nongelling starch	Carbohydrate	Waxy maize
Gum acacia	Gum	Polysaccharide	Trees of the species *Acacia senegal*
Agar agar	Thermoreversible gelling agent	Polysaccharide	Red seaweeds
Alginate	Irreversible gelling agent	Polysaccharide	Brown seaweed
Carrageenan	Thermoreversible gelling agent	Sulfated polysaccharide	Red seaweeds
Gellan gum	Thermoreversible or irreversible gelling agent	Polysaccharide	*Pseudomonas elodea*
Guar gum	Thickener exhibits synergy with some gelling agents	Galactomannan	Seeds of *cyamopsis tetragonolobus*
Pectin – high methoxyl	Irreversible gelling agent	Polygalacturonic acid	Citrus peel or cider apple pomace
Pectin – low methoxyl	Thermoreversible gelling agent	Demethoxylated pectin	Citrus peel or cider apple pomace
Gum tragacanth	Gum or mucilage	Polysaccharide	*Astragalus* shrub
Locust bean or carob gum	Thickener exhibits synergy with some gelling agents	Galactomannan	Endosperm of locust beans from *Ceratonia siliqua*
Xanthan gum	Thickener exhibits synergy with locust bean gum	Polysaccharide	Aerobic fermentation of *Xanthomonas campestris*
Egg albumen	Whipping agent and irreversible gelling agent	Protein	Egg white
Enzyme-modified soy protein	Whipping agent	Protein	Soy beans

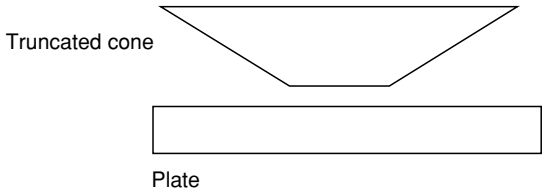

Figure 6 Cone and plate geometry.

The mixture is then cooled with beating to produce fine sucrose crystals. It is possible to make a fondant using dextrose instead of sucrose. Crystalline dextrose has a positive heat of solution, i.e., heat is absorbed. This produces a marked cooling effect, which goes well with a peppermint flavor, but it is not appreciated in other flavors. After a fondant has been made, the crystals undergo Ostwald ripening.

Fudge

A fudge-like product can be made by adding either a fondant or milled sugar to a toffee; alternatively, a sugar-rich caramel can be induced to grain.

Panned Products

In these products, one or more layers of sugars have been laid down. There are two different processes used, hard panning and soft panning. These processes involve building up a coating of sugar and sometimes other things layer by layer. This is one of those processes that the confectionery and pharmaceutical industries have in common. The normal modern small-scale method is to use a rotating dragee pan (Figure 7). These pans are traditionally made of copper but now more commonly stainless steel. The pan is equipped to be rotated, and usually, there will be a system for supplying and extracting air and possibly with a system for spraying in sugar syrup.

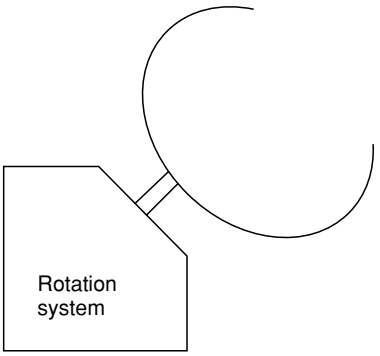

Figure 7 Dragee pan.

Hard panning

Hard-panned coatings are purely sucrose, whereas soft pan coatings are a mixture of sucrose and glucose syrup. Sugared almonds, mint imperials, nonpareils (hundreds and thousands), as well as sugar-coated chocolate lentils or eggs are normally hard panned. It is possible to soft pan some of these products, e.g., almonds. In hard panning, the centers are tumbled in the pan, and a sugar syrup is applied. The rotation

of the pan and the tumbling of the centers spread the syrup out into a thin layer. The water is evaporated, which causes the sugar to crystallize. The process is inevitably slow. The water can evaporate only at a certain rate, and the rate of crystallization of the sugar cannot be accelerated. The rate of evaporation can be increased by increasing the temperature and the air flow as well as reducing the humidity of the air. Increasing the temperature will reduce the rate of crystallization. The layers applied are only 10–14 μm thick. As the layers are so thin, they follow the contours of the product. If the center can be damaged by heat, heating the pan may not be possible. An air supply and extraction speed up the panning process considerably. In an unventilated pan, the process would slow down as the air became saturated with water vapor. In general, hard panning is done in large pans, whereas soft panning is done in small pans. Small products such as nonpareils tend to be made in small pans, whereas large items are produced in larger pans.

Soft panning

Soft panning involves applying a syrup to the centers, as does hard panning, but the soft panning syrup is intended not to crystallize. Instead of evaporating the water, as in hard panning, milled sugar is added, which dissolves in the water in the syrup. The syrup used in soft panning would be either all glucose syrup or a 50:50 mixture of sucrose and glucose syrup. This process is carried out in the cold and without using drying air. Dust extraction is needed for reasons of health and safety. Soft panning is a much more rapid process than hard panning. It can be applied to soft centers that would be unsuitable for hard panning. As the coating is thick, the shape of the center tends to be lost. Typical soft panned products are jelly beans and dolly mixture components.

Process Control Systems

Attempts to automate panning initially concentrated on the time and temperature. A more modern approach is to measure the absolute humidity to control the drying cycle. This is claimed to be an improvement on systems that rely on temperature and the elapsed time. In particular, because the actual humidity is measured, process parameters can be optimized to maximize throughput. In a time–temperature system, in order to avoid under-drying, the time set would have to be the maximum necessary.

Toffees and Caramels

In these products, the Maillard browning reaction between reducing sugars and proteins is essential to the taste and color of the finished product. The proteins used are from skim milk solids. Inefficient cooking pans promote this reaction at the heating surface to give a good product. If the same ingredients are cooked in an efficient continuous system, the reaction is suppressed to such an extent that the product is unacceptable. In practice, holding vessels known as caramelizers are used to promote the Maillard reaction.

In addition to skimmed milk solids and sugars, toffees contain fat. Originally, toffees were made from full cream milk solids, and the fat was milk fat. Now, a total or partial replacement with vegetable fat is possible. If sufficient milk protein is present, the fat disperses easily in the toffee, leaving a fatty layer on the surface that facilitates cutting. If insufficient milk protein is present, other emulsifiers will have to be added.

Chewing Gum

Chewing gum is a flavored and sweetened rubber, originally natural, but now usually synthetic. It is a product where confectionery becomes involved with polymer chemistry. The gum base in chewing gum is technically a rubber. The original ingredients were the natural products chicle and jelutong-pontianak, which were used alone or in a mixed gum base. The synthetic polymers used are mainly vinyl esters in particular polyvinyl acetate. Butadiene styrene rubber, isobutylene–isoprene copolymer, and polyisobutylene can be used. As in other fields of polymer chemistry, plasticizers such as glycerine esters of hydrogenated or polymerized resin, lanolin, and potassium, and sodium stearates are used. Different gum bases are used for chewing gum and bubble gum. Bubble gum base will contain either higher levels of polymers or polymers with a higher molecular weight. The higher level or molecular weight of the polymer makes the gum base more extensible and hence able to form bubbles.

Sugars

In order for a product not to have a gritty feel in the mouth, the sugar has to have a very fine particle size. Individuals differ in their response to particle size, but particles above 20–40 μm are normally found to be gritty.

Dextrose

Dextrose monohydrate is sometimes used as an alternative to sucrose in chewing gum. The endothermic heat of solution of dextrose gives a cooling sensation in the mouth that goes well with mint flavors but not with others.

Aerated Products

Technically, the problems in producing aerated products are in making and stabilizing a foam. Typically, aerated confectionery products involve making a foam and then causing it to set. A popular ingredient is egg albumen, which is a whipping agent and can be set irreversibly by heat. Some products are made with the combination of egg albumen and a gelling agent, e.g., gelatine. The only aerated confectionery products that work differently are those that are a high boiling that is chemically aerated and relies on the product passing through the glass transition before the bubbles collapse. These products sometimes contain material like gelatine to enhance foaming. Examples are bonfire toffee, honeycomb crunch, and peanut brittle.

Tabletting

Tablets are made by compressing a powder in a die. Under the pressure in the die, the powder fuses and is ejected as tablet. Tabletting requires a free-flowing powder that will flow freely when poured into the die but bonds satisfactorily when compressed.

Lozenges

Lozenges are one of the oldest forms of sugar confectionery. Lozenges are cut from a sheet of dough and then dried. Lozenge manufacture is one of the areas in which sugar confectionery is related to pharmaceutical products.

Lozenge manufacture

Lozenges are made by taking a milled sugar and making it into a dough. The best textures are achieved by using the smallest possible sizes of sugar. Lozenges are normally made with a binder. The traditional binder is gum tragacanth, which is probably the best material. Other materials that are used are gum acacia, gelatine, and xanthan gum. A mixture of gum acacia and gum tragacanth is sometimes used. Gum arabic used alone gives very brittle sweets that break easily.

Drying The stamped pieces are then spread on to trays where a slight crust of dry sugar forms on the surface. The lozenges are then dried at 35–40 °C until the moisture content is around 1.5%. If the lozenges are dried too rapidly, the lozenges shrink and crack. Extended drying times lead to loss of flavor and yellowing. A trace of blue color is often added to give a blue–gray color rather than a yellow–gray color. Flavor is a problem in lozenges, as the flavor must be volatile to be perceived in the finished product, but volatile flavors are often lost during drying. The reason why drying cannot be accelerated is that the moisture has to migrate from the center of the sweet to the surface.

Conclusion

Sugar confectionery is an area where the product came first and the science arrived afterwards. The science, however, is challenging.

See also: **Amino Acids**: Properties and Occurrence; **Browning**: Nonenzymatic; Toxicology of Nonenzymatic Browning; **Glucose**: Properties and Analysis; **Gums**: Properties of Individual Gums; Food Uses; Dietary Importance; **Jams and Preserves**: Methods of Manufacture; Chemistry of Manufacture; **Sucrose**: Properties and Determination

Further Reading

Edwards WP (2000) *The Science of Sugar Confectionery*. Cambridge: The Royal Society of Chemistry.

Hill S and Easa AM (1998) *The Maillard Reaction in Foods and Medicine*, pp. 113–138. Cambridge: The Royal Society of Chemistry.

Fisher C and Scott TR (1997) *Food Flavours: Biology and Chemistry*. Cambridge: The Royal Society of Chemistry.

Jackson EB (ed.) (1995) *Sugar Confectionery Manufacture*, 2nd edn. London: Chapman & Hall.

Lees R and Jackson B (1973) *Sugar Confectionery and Chocolate Manufacture*. Glasgow: Blackie Academic & Professional.

SYRUPS

M A Clarke, Sugar Processing Research Institute Incorporated, New Orleans, LA, USA

Introduction

Syrups, which are sweeteners in liquid form, usually of high viscosity, have been used as food sweeteners since the early days of humanity, e.g., honey. This article includes information on the source and production, composition and properties, and use of cane and golden syrups, molasses, sorghum syrups, corn-, glucose-, fructose- and starch-derived syrups, fruit-based syrups, and maple syrups. (*See* **Honey**.)

Cane Syrups, Golden Syrups, and Molasses

Cane syrups are usually made from sugarcane juice, concentrated by evaporation, after clarification, before any sugar is crystallized out or otherwise removed. (*See* **Sugar**: Sugarcane.)

Cane Syrup

The term 'cane syrup' as applied to consumer products has a regional orientation to southern USA and other sugarcane-processing areas, where cane syrups and blends are sold for use on pancakes, biscuits, and cereals and in cooking. For food industry use, cane syrups are produced at sugarcane factories in Louisiana and Hawaii, or at refineries, where a blend of brown and golden-colored streams are combined to produce syrup. Cane syrups are dark golden brown in color, with medium flavor intensity (caramel, butterscotch, cane and green flavors; no heavy molasses flavor), and are partially inverted. (Inversion is the process in which sucrose is converted by hydrolysis to an equimolar mixture of glucose and fructose.) A factory evaporator syrup is often completely inverted, and mixed with uninverted syrup, to give a product of approximately 85° Brix, 25–30% sucrose and 50–52% invert. This is a relatively clear and low-ash material. (*See* **Carbohydrates**: Classification and Properties; **Fructose**; **Sucrose**: Properties and Determination.)

Cane sugar refineries make liquid sucrose, a watery, colorless solution of 67–70° Brix (percentage on a solids basis) sucrose, generally prepared by redissolving granulated sugar, but in some special refineries, e.g., in Brazil and Canada, by extensive ion exchange resin decolorization and carbon treatments. This product is often sold to beverage companies and canners because the high quality of color and clarity are immediately apparent.

Liquid invert sugar, which includes a range of liquid products of sucrose, glucose, and fructose of 75–77° Brix, was, until the late 1970s, a major product (15% of market) in the USA. It has been replaced by the cheaper starch-based syrups, most notably high-fructose corn syrup. Liquid products are still made in the UK, other parts of Europe, South America, and other areas.

Golden Syrup

Golden syrup is most popular in the UK, Canada, South Africa, and Australia (countries in which it is manufactured). It is a high-Brix (77–82°), partially inverted syrup, filtered several times over bone charcoal to give it a special golden color, very mild flavor, and high clarity. The syrup is generally inverted with sulfuric acid and neutralized with calcium carbonate, so that no soluble salt will remain as a reaction product to affect the special flavor. Typical analysis (courtesy of Tate and Lyle Sugars) is as follows: invert, 50%; sucrose, 32%; ash, 1.4%, and solids, 82.6%. The syrup is used directly on cereals, breads, and baked goods, and in home baking, e.g., in syrup pudding or tart. Golden syrup tends to crystallize on storage and is therefore usually sold in cans.

The obvious advantages, in many applications, of handling sugar in dissolved form have led to the extensive distribution and use in bakery products of sucrose and invert sugar syrups. Liquid sugars can be roughly classified into sucrose types, invert (or mixed) types, and refinery syrups or liquid brown sugars. Sucrose is available at 66.5–68% solids content. This is the limit of solubility of sucrose at ordinary temperatures. Two or three grades, varying for the main part only in color, are usually available.

If part of the sugar is inverted, the resulting syrup will retain higher concentrations of solids in solution. Common commercial types are syrups of 73% or 76% solids with 30% or 60% invert. Totally inverted syrups contain 72–73% solids, of which perhaps 5% is sucrose.

Because of the low water content, all of these syrups are quite resistant to microbiological spoilage. The invert syrups are probably somewhat superior in this regard because of their lower water activity. (*See* **Water Activity**: Effect on Food Stability.)

Molasses

Molasses (or treacle in the UK consumer market) is a general term for concentrated juice from sugarcane or sugarbeet, or raw cane sugar in concentrated solution after varying amounts of sucrose have been removed. Sugarcane molasses is the major food molasses. Both sugarbeet and sugarcane molasses are used for animal feed and as fermentation sources for ethyl alcohol and other chemicals. These uses are amply described in the literature and are not discussed here. Recent developments in technology have made possible some sugarbeet molasses food products; these are at present available in limited quantity. (*See* **Sugar**: Refining of Sugarbeet and Sugarcane.)

Several common terms for molasses are defined as follows. Blackstrap molasses is the byproduct from a sugarcane factory or raw sugar refinery; it is the heavy, dark viscous liquid remaining after the final stage of sugar crystallization from which no further sugar can be crystallized economically by the usual methods. Types of blackstrap are further defined by the US Department of Agriculture (USDA) as superior, normal, or utility, but these are ready definitions for feed-grade material.

High-test molasses is the product obtained by concentrating clarified cane juice to approximately 85° Brix; it is partially inverted with either acid or invertase enzyme. High-test molasses is produced from cane juice instead of sugar, not as a byproduct of sugar production. High-test molasses, also known as fancy molasses, cane invert syrup, or cane juice molasses, is a premium product, higher in sugars content and of a more aromatic flavor than blackstrap. It has been subjected to less heat than blackstrap, and so contains relatively fewer sugar decomposition products, which can add bitter flavor.

Sulfured molasses is the byproduct of raw sugar manufacture in which sulfur dioxide has been added to the molasses to bleach color. Sulfured molasses may be lighter in color, but it is higher in ash of the insoluble sulfate type. The term 'unsulfured' is more common. The approximate composition of cane molasses (blackstrap) is given in Table 1. (*See* **Preservation of Food**; individual constituents.)

Most commercially available molasses, for consumer purchase or food industry use, is made by blending various cane factory and refinery molasses and syrups for desirable – and constant – flavor and quality.

Physical properties of molasses vary with composition. Viscosity can vary over several orders of magnitude depending on inorganic and polysaccharide composition and temperature. Cane molasses has an acid pH, usually between 5 and 7. The salts content (2–8%) can contribute buffering capacity, to stabilize flavors and prevent hydrolysis, and can also provide flavor for feed use.

Color and flavor are the major properties besides nutrition that molasses contribute to food processing. Molasses always contributes some sweetness, to a degree which usually decreases as color darkens. The range of flavors is broad, ranging from caramel, cane flavor in light high-test molasses, to heavy, bitter notes, sometimes with licorice characteristics. Extensive research has shown that molasses flavor is a complex mixture. Because of this wide range of flavors, molasses can be used to mask or disguise other, less pleasant flavors, e.g., bitterness of bran in wholewheat (wholemeal) products; as an enhancer in sauces and licorice products. Molasses can also be used as a coloring agent, for golden to dark brown colors, especially in baked goods, and as a color enhancer, or masking agent, to disguise gray or gray-brown tones. Molasses displays humectancy, colligative, and water absorption (lowering of water activity) properties that would be expected of a sucrose syrup of similar solids composition; these properties lead to wide application in intermediate-moisture foods and in baked goods with extended shelf-life. However, nonsugars in molasses apparently have some effect on humectancy: molasses has been shown to be more effective in retaining a high moisture content than either sucrose or corn syrups. Molasses nonsugars also exhibit an antioxidant effect, significant when this fraction is used at some 3% of the fat level in the food product, and important for health foods. Inclusion in whole-grain products is one potential role for this fraction. (*See* **Antioxidants**: Natural Antioxidants; **Colorants (Colourants)**: Properties and Determination of Natural Pigments.)

Molasses is a food product that has often been claimed to be a secret of youthful behavior and appearance (e.g., to inhibit hair graying). Technological investigations, such as the isolation and characterization of the antioxidant fraction, may provide a scientific basis for what formerly appeared to be myths.

A dried molasses product, usually mixed with corn syrup solids to absorb water, is also available for use in dry mixes.

Syrups and molasses products sold to the food industry represent the result of blending various molasses and syrups to produce products of consistent color, flavor, and functional properties. Table 2 lists several typical blends, their composition, flavor characteristics, and potential applications. These blends are typical of material sold directly to the consumer as molasses in the USA, and as treacles, syrups, or blends in, other countries.

Table 1 Approximate composition of cane molasses

Main constituents	*Components*	*Normal range*
Water		17–25%
Sugars	Sucrose	30–40%
	Glucose (dextrose)	4–9%
	Fructose (levulose)	5–12%
	Other reducing substances (as invert)	1–4%
	Total reducing substances (as invert)	10–25%
Other carbohydrates	Gums, starch, pentosans, also traces of hexitols; myoinositol, D-mannitol, and uronic acids	2–5%
Ash	As carbonates[a]	7–15%
	Bases:	
	Potassium oxide (30–50%)	
	Calcium oxide (7–15%)	
	Magnesium oxide (2–14%)	
	Sodium oxide (0.3–9%)	
	Metal oxides (as ferric) (0.4–2.7%)	
	Acids:	
	Sulfur trioxide (7–27%)	
	Chloride (12–20%)	
	Phosphorus pentoxide (0.5–2.5%)	
	Silicates and insolubles (1–7%)	
Nitrogenous compounds	Crude protein (as N × 6.25)	2.5–4.5%
	True protein	0.5–1.5%
	Amino acids, principally aspartic and glutamic acids, including some pyrrolidine carboxylic acids	0.3–0.5%
	Unidentified nitrogenous compounds	1.5–3.0%
Nonnitrogenous	Aconitic acid (1–5%), citric, malic, oxalic, glycolic	1.5–6.0%
	Mesaconic, succinic, fumaric, tartaric	0.5–1.5%
Wax, sterols, and phosphatides		0.1–1.0%
Vitamins	Thiamin (B_1)	2–10 p.p.m.
	Riboflavin (B_2)	1–6 p.p.m.
	Pyridoxine (B_6)	1–10 p.p.m.
	Nicotinamide	1–25 p.p.m.
	Pantothenic acid	2–25 p.p.m.
	Folic acid	10–50 p.p.m.
	Biotin	0.1–2 p.p.m.

[a]Percentage of ash given in parentheses.
Source: United Molasses Company, London, UK. By courtesy of the Technology Division of Crompton and Knowles Corporation, Mahwah, NJ, USA.

Sorghum Syrup

Sweet sorghum (*Sorghum bicolor*) is a giant grass, somewhat similar in appearance to sugarcane, but able to withstand cooler climates. In the Midwest and parts of the southern states of the USA, the juice of sweet sorghum is heated, clarified by skimming, and concentrated into a syrup. Sorghum juice tends to be higher in invert sugars than cane juice; it is therefore difficult to crystallize sweet sorghum sugar, and syrup is the product of choice. Sorghum syrup is produced by small plants in Tennessee, and northern Mississippi and Alabama (further information is available from the National Sweet Sorghum Producers Association, PO Box 1071, Knoxville, TN 37901-1071, USA). (*See* **Sorghum**.)

The light-brown-colored syrup has a distinctive, pungent odor and flavor, in addition to its sweetness and molasses-like flavor. It is often blended with other (sugar- or starch-based) syrups, but is seldom sold industrially. Production is a cottage industry at present.

Starch-Based Sweeteners and Syrups

Starch is a polymer of glucose and, upon hydrolysis by acid or by amylase enzyme, it yields glucose and a variety of maltooligosaccharides. The process has been employed in the USA to produce sweetener syrups since the 1840s, at first on potato starch and some 20 years later on corn, or maize, starch. Crystalline glucose, more often called dextrose in the corn refinery industry, and glucose syrup remained the major starch-based sweetener products until the 1960s. Enzyme-catalyzed isomerization of glucose to fructose, a sweeter sugar, then made possible

Table 2 Molasses and syrup blends

Characteristic description (%)	*Unsulfured molasses (%)*	*Bakery molasses (%)*	*Confectionery and all-purpose molasses (%)*	*Condiment molasses (%)*	*Robust molasses*
Composition					
Sucrose	35	32–36	33–37	30–36	33–37
Invert	37	36–40	28–32	21–27	16–20
Total sugars	72	70–74	63–67	54–60	51–55
Ash	2.5	1.2–2.5	4.5–5.5	6.5–8.5	8.0–9.0
Color	Golden brown	Light brown	Medium brown	Dark brown	Dark brown
Flavor	Sweet, mild aroma, syrup flavor	Sweet mild, distinctive	Moderate sweet, strong aroma	Strong, pungent flavor,good background	Strong flavor heat-resistant
Humectancy	Some	Good	Good	Some	Some
Buffer effect			Yes	Yes	Yes
Applications	Table syrups, toppings, peanut butter, fruit purées, confectionery, and alcohol products	Fruit cakes, brownies, muffins, spiced baked goods	Barbecue sauce, extended products, candies (hard and caramel), toasted foods, gingerbread	Fermented products, condiments, and sauces	Leavened and fermented goods, soy sauce, tobacco, licorice, baked beans, caramel, snacks

high-fructose syrups, with ratio of sweetening power to energy equivalent to that of sucrose. The corn-refining industry in the USA now makes available a wide range of sweeteners, with crystalline fructose as the newest addition. Corn-based starch hydrolysis products are also manufactured in Argentina, Canada, Japan, and eastern Europe. In the EC, the starch base for these sweeteners is usually potato, increasingly wheat, and, in at least one factory in Finland, barley. Products are similar in composition, properties, and use to those made from corn; thus cornstarch-based products are used as examples here. (*See* **Dextrins**; **Starch**: Structure, Properties, and Determination; **Sweeteners**: Intensive.)

In general, corn-based sweeteners are classified into the following seven categories:

1. Corn syrup (glucose syrup) is the purified, concentrated, aqueous solution of nutritive saccharides obtained from edible starch and has a dextrose equivalent of 20 or more.
2. Dried corn syrup (dried glucose syrup) is a corn syrup from which the water has been partially removed.
3. Dextrose monohydrate is purified and crystallized D-glucose containing one molecule of water of crystallization with each molecule of D-glucose.
4. Dextrose anhydrous is purified and crystallized D-glucose without water of crystallization.
5. Maltodextrin is a purified concentrated aqueous solution of nutritive saccharides obtained from edible starch, or the dried product derived from the solution and having a dextrose equivalent less than 20.
6. High-fructose corn syrup is a purified, concentrated, aqueous solution of nutritive saccharides obtained from edible starch in which a portion of the dextrose has been isomerized to fructose.
7. Crystalline fructose is a purified and concentrated D-fructose obtained from edible starch in either dry crystal or remelted liquid form.

Products in 2, 6, and 7 are of direct interest as syrups; the others are listed for explanation of the processes involved.

Composition of the typical range of syrups is outlined in Table 3.

Starch-Based Syrups Manufacture

Common methods employed in the commercial production of corn syrups are the acid process, the acid–enzyme process, and multiple enzyme process. In the acid conversion process, a starch slurry of the appropriate dry substance is acidified to pH of about 2 and pumped to the converter. After neutralization, liquor is clarified and concentrated by evaporation to an intermediate density. The resulting syrup is further clarified, decolorized, and finally concentrated in evaporators to the final required density. Some syrups are treated with ion exchange resins for further refinement.

The acid–enzyme process is similar except that the starch slurry is only partially converted by acid to a given dextrose equivalent (DE), then treated with an appropriate enzyme or contribution of enzymes to complete the conversion.

In multiple enzyme processes, starch granules are gelatinized and the preliminary starch-splitting or

Table 3 Composition of corn syrups

Designation	Ash	Saccharides, carbohydrate basis			
		DP_1	DP_2	DP_3	DP_4
28 DE	0.3	8	8	11	73
36 DE	0.3	14	11	10	65
34 HM	0.3	9	34	24	33
43 HM	0.3	9	43	18	30
43 DE	0.3	19	14	12	55
43 DE (ion-exchanged)	0.03	19	14	12	55
53 DE	0.3	28	18	13	41
63 DE	0.3	36	31	13	20
63 DE (ion-exchanged)	0.03	36	31	13	20
66 DE	0.3	40	35	8	17
95 DE	0.3	95	3	0.5	1.5
95 DE (ion-exchanged)	0.03	95	3	0.5	1.5
HFCS 42	0.03	95	3	0.7	1.3
HFCS 55	0.05	95.7	3	0.4	0.9
Crystalline fructose	0.05	100.0			

DE, dextrose equivalent, or total reducing sugars reported as dextrose percentage solids; HM, high-maltose; HFCS, high-fructose corn syrup; DP, degree of polymerization; DP_1 dextrose (dextrose plus fructose for HFCS; fructose for crystalline fructose); DP_2, maltose; DP_3, maltotriose; DP_4, sum of saccharides DP_4 and above.
Courtesy of the Corn Refiners Association, Washington, DC, USA.

depolymerization is brought about by an α-amylase enzyme, rather than by means of acid.

Various intermediate syrups of differing composition may be further converted with enzymes having specific modes of action or providing particular types of end product, such as high-maltose syrups, high-fermentable syrups, and others.

High-Fructose Corn Syrup

Dextrose or glucose solutions or high-DE substrates produced from starch by acid–enzyme or dual enzyme processes are refined by carbon and ion exchange systems and treated enzymatically with isomerase in an immobilized enzyme system. Isomerization is usually carried to a point where the substrate contains 42% fructose. Following this step, the product is refined again through carbon and ion exchange systems and is evaporated to a dry solids level of 71%. (*See* **Enzymes**: Uses in Food Processing.)

In the production of syrups with a fructose level above 50%, the original 42% fructose feedstock is passed through separation columns of cationic ion exchange resins which retain fructose and dextrose. Fructose is removed; dextrose is recirculated for further isomerization. In a standard moving-bed system, the fructose fraction is generally recovered at an 80–90% concentration, and blended with 42% fructose feedstock to produce 55% fructose content, at a dry solids level of 77%. The high-level feedstock may also be refined and evaporated to produce 90% high-fructose syrup.

High-Maltose Syrups

Enzyme technology has been developed to produce syrups that are higher in maltose than an acid-converted syrup of similar DE. These syrups are used for fermentation (brewing) and for applications in icecream and confectionery where physical properties are more important than sweetness. The low dextrose concentration allows control of color development in a heated product. The 65% maltose syrup is a fine material for catalytic reduction.

Chemical and Physical Properties of Corn Syrups

Corn syrups all have pH values on the acid side, minimum 3.5–5.5, to minimize color and flavor development. Ash (mainly sodium chloride) levels in corn syrups are low. Although sulfur dioxide is used extensively in the preparation of starch from corn, processing, especially those processes using ion exchange resins, can remove residual sulfur dioxide so that products contain acceptably low levels of sulfite and its oxidation product, sulfate. Fermentability of corn syrups is another useful chemical property; fermentable extract is defined as the percentage of carbohydrates, on a dry-weight basis, fermented by bakers' yeast under controlled conditions. The higher the DE, the higher the fermentability.

The reducing sugars characteristics of glucose, fructose, and maltose allow participation in Maillard, or browning, reactions, for brown crust formation in baking and production of caramel color. (*See* **Browning**: Nonenzymatic; **Caramel**: Methods of Manufacture.)

Density of corn syrups is usually measured in degrees Baumé instead of Brix. Syrups content (e.g., 55% high-fructose corn syrup) is always expressed on a solids basis.

Corn syrups, because of their solids content, have colligative properties similar to those of sucrose invert syrups as regards boiling point elevation and freezing point depression: the latter is an important factor in icecream hardness and melting. Osmotic pressure increase is a measure of control against spoilage in jams and preserves. The syrups' very high viscosities are an important property in food processing, controlling handling characteristics, and air retention.

The humectant, or hygroscopic, properties of corn syrups vary with their carbohydrate composition. Lower-DE syrups are less hygroscopic, while maltotriose and maltotetraose may be the most hygroscopic sugars. Variation in this property may be

used to control the texture and keeping quality of baked goods, confectionery, and intermediate-moisture foods, where corn syrups are employed as moisture conditioners, food plasticizers, crystallization inhibitors, and stabilizers.

Applications and uses of starch-derived syrups include the following: baby and geriatric foods; bakery products; beverages, brewed, alcoholic, carbonated, and still; breakfast foods; cheese spreads and cheese-based foods; coffee whiteners; condensed sweetened milk; confectionery; eggs (frozen or dried); extracts and flavors; frosting and icings; fruits and vegetables; icecreams; industrial products (adhesives, chemicals, dyes and inks, explosives, paper, textiles, tobacco); jams, jellies, marmalades, and preserves; meat products (sausages, etc.); peanut butter; pharmaceutical and medical; pickles and pickle products; pork and beans; prepared mixes; seafood (frozen); syrups (table, chocolate, cocoa, fruit, medicinal, soda fountain, cordials, etc.); soups; toppings.

High-fructose corn syrup is also widely used in processed foods, bakery products, beverages (colas and other carbonated soft drinks, and still drinks), canned juices, canned fruits, condiments, confectionery products, frozen desserts, jams, jellies, and preserves, pickles, and wine.

Lycasin Lycasin, originally manufactured in Sweden by the Lykeby Starch Co., is a hydrogenated glucose syrup, which does not crystallize out, even at low temperatures, and has a viscosity similar to that of high-fructose corn syrups. It is a mixture of hydrogenated homologs of glucose and glucose polymers, said to prevent dental cavities, although not suitable for diabetics, and used in jams, preserves, drinks, and hard-boiled candies.

Maple Syrup

Maple sweetners, made by concentration of the sap of the sugar maple tree (*Acer saccharum*), are produced in eastern Canada and north-eastern USA. Heating of the sap causes flavor and color development, as well as concentration, and the characteristic maple flavor of the syrup and sugar (sold in granulated or in block form) is a major attraction. The pure product is more expensive, by an order of magnitude, than sucrose or starch-based sweetners, and, as a result, many blends of maple syrup with other syrups, or of other syrups with maple flavoring, are available, and nowadays are generally labeled as such; false labeling of maple products was common some years ago.

The composition of maple syrup is shown in Table 4. The carbohydrate make-up of fresh maple syrup is almost entirely sucrose. This will invert with time, but provides a means of detecting adulteration or replacement of maple syrup with cheaper, starch-based syrups: genuine maple product does not contain maltose, while starch-based products do.

Table 4 Composition of maple syrup

Component	*Amount (%)*
Water	34.0
Sucrose	58.2–65.5
Hexoses	0.0–7.9
Malic acid	0.093
Citric acid	0.010
Succinic acid	0.008
Fumaric acid	0.004
Soluble ash	0.30–0.81
Insoluble ash	0.08–0.67
Calcium	0.07
Silica	0.02
Manganese	0.005
Sodium	0.003

Physical properties of maple syrup are similar to those of sucrose syrups, making the product suitable for baking and confectionery, but, because of its cost, most maple syrup is sold directly, and used as pancake syrup or in home baking. Characteristic maple flavor has been shown to be a complex mixture of plant components (guiacyl acetone, vanillin) and products of thermal degradation of carbohydrates (isomaltol, acetol, α-furanone). (*See* **Carbohydrates**: Interactions with Other Food Components.)

Fruit Syrups

A new product has entered the sweetener syrup field in the USA, Europe, and South America: fruit syrup or fruit juice concentrate. Excess fruit crops, or damaged fruit unsuitable for its original destination as fruit, fruit products, or wine, is prepared in juice form. The juice is decolorized and the flavors removed, through proprietary processes, and the remaining material – usually a solution of invert sugar, because fruit sucrose is hydrolyzed to invert in the process – is concentrated to about 75 ° Brix at pH 4, and sold as a natural fruit sweetener. The product is designed for the natural foods market, so that the terms 'fruit sweetener', 'sweetened only with fruit' and 'no sugar added' can be displayed on the label, although the last statement is questionable. These fruit products, made in different places from apple, peach, pear, citrus, and grape juices, sell, on solids basis, at five to six times the price of sucrose.

See also: **Antioxidants**: Natural Antioxidants; **Caramel**: Methods of Manufacture; **Carbohydrates**: Classification and Properties; Interactions with Other Food

Components; **Dextrins**; **Enzymes**: Uses in Food Processing; **Fructose**; **Honey**; **Sorghum**; **Starch**: Structure, Properties, and Determination; **Sucrose**: Properties and Determination; **Sugar**: Sugarcane; Refining of Sugarbeet and Sugarcane; **Sweeteners**: Intensive; **Water Activity**: Effect on Food Stability

Further Reading

Clarke MA and Godshall MA (eds) (1988) *Chemistry and Processing of Sugarbeet and Sugarcane*. Amsterdam: Elsevier Science.

Marie S and Piggott JR (eds) (1991) *Handbook of Sweeteners*. London: Blackie.

Pancoast HM and Junk WR (eds) (1980) *Handbook of Sugars*, 2nd edn. Westport, Connecticut: AVI.

T

TAINTS

Contents

Types and Causes

B D Baigrie, The Lord Zuckerman Research Centre, University of Reading, Whiteknights, UK

Background

British Standard 5098, 1974 defines a taint as a taste or odor foreign to the product, whereas the dictionary definition of an off-flavor is a flavor that is not natural or up to standard owing to deterioration or contamination. Often, the term taint is used specifically to describe a flavor defect that has arisen from extrinsic sources. Where a flavor defect has developed internally, e.g., via enzymically induced deterioration and not by external contamination, the term off-flavor is used. In practical terms, this differentiation is not particularly helpful or necessary, and hereafter, the term 'taint' will be used to describe any deviation of flavor from the accepted norm, irrespective of the cause.

The sources of taints in foodstuffs are manifold. They can arise through microbial spoilage (bacteria, yeasts, and molds), and many of the metabolites produced by these species will often remain in the food, even after extensive processing. They may also be formed by internal deterioration of food components through the intermediacy of enzymes, by other non-enzymic pathways (Maillard reactions, Strecker degradation, etc.), or by oxidative degradation. In addition, many food taints or precursors thereof arise from some source of external contamination. For example, airborne sources of taint are common. Water used in food processing, cleansing/sanitizing agents, and packaging materials are major vectors also for the transmission of taints to foods. This article discusses some of the more commonly occurring taints and their origins.

Microbial Metabolites

Microorganisms are a common source of taint in foods. Microbial activity can produce taints in several ways: by production of primary metabolites; by conversion of innocuous food components into species with a low odor threshold, or through residual enzyme activity after death of the organism. Fish is a food that is susceptible to tainting by all of these mechanisms. Fresh fish has little odor, but on storage, spoilage odors develop, primarily through bacterial action. *Achromobacter*, *Vibrio*, and *Pseudomonas* bacteria have been shown to be capable of producing sulfidic taint in fish by the production of volatiles such as hydrogen sulfide, methyl mercaptan, dimethyl sulfide, dimethyl disulfide, and dimethyl trisulfide. A metabolite of *Pseudomonas perolens*, 2-methoxy-3-isopropylpyrazine, was found to cause a potato-like taint in fish, whereas a fruity taint in fish muscle caused by ethyl esters of acetic, butyric, and hexanoic acids was produced by *Pseudomonas fragi*. The classic 'fishy' odor of aged fish has been attributed to the presence of trimethylamine, and it has been established that this volatile is produced by the action of bacterial enzymes on trimethylamine oxide, a natural constituent of fish muscle.

Metabolites from microbial species, including *Pseudomonas*, have also been implicated as the source of specific taints exhibited by various crustacea. Dimethyl trisulfide was responsible for an onion-like flavor, and indole a fecal-like flavor that had developed in red prawns. The compounds responsible for a distinctive garlic-like taint also in red prawns, namely bis-(methylthio)-methane and

trimethylarsine, were believed to arise from microbial activity. The former chemical was also the principal cause of an objectionable taint in sand lobster. 2,6-Dibromophenol was identified as the cause of an iodoform-like taint in prawns. Marine algae and bryozoa were found to contain this compound in concentrations comparable with that found in the prawns and were implicated as the primary source of the taint.

Freshwater crustacea (peneid shrimp) and fish (rainbow trout, catfish, bream) can often acquire earthy/musty taint via contaminated water. The compounds held to be primarily responsible for musty taint from these sources are geosmin ([*E*]-1, 10-dimethyl-[*E*]-9-decalol), 2-methylisoborneol (1,2,7,7-tetramethyl-exo-bicyclo [2.2.1] heptan-2-ol) and, to a lesser extent, 2-methoxy-3-isopropylpyrazine. In most cases, geosmin and methylisoborneol are found in eutrophic and hypertrophic waters, and it has been established unequivocally that the causal agents of these chemicals are actinomycetes (several species), cyanobacteria (blue–green algae), or certain species of fungi. The presence of geosmin or 2-methylisoborneol in water supplies, particularly that used for food processing, is the probable cause of earthy taints in some foods canned in brine (e.g., champignons) and syrups.

Food production systems that utilize microorganisms to produce flavor, e.g., dairy products, alcoholic beverages, and cocoa, are particularly susceptible to the development of taints via excessive microbial activity. A selection of some taints developed by these types of foods is shown in Table 1. Of course, many other foods can develop fermentation-like taints through adventitious infection by the appropriate microorganisms.

Enzymic Deterioration of Food

Foods containing lipids are particularly prone to the development of taints by enzyme-catalyzed degradation of lipid. Two main types of enzymes are involved in lipid degradation: lipoxygenase and lipase. The former are particularly common in plant tissues such as the legumes. For example, the beany flavor of soya beans and the reversion flavor of soya bean oil are due to lipoxygenase activity.

Lipoxygenase catalyzes fatty acid hydroperoxide formation in a manner similar to nonenzymic systems. Polyunsaturated fatty acids such as linoleic and linolenic acid are more susceptible to attack than monounsaturated or unsaturated acids. After initial formation of the hydroperoxides, decomposition by pathways similar to those occurring in autoxidation takes place. This is believed to involve homolytic cleavage of the oxygen–oxygen bond to yield an alkoxy radical that decomposes via carbon–carbon cleavage to give aldehyde or hydrocarbon products. Other pathways lead to the formation of hydrocarbons, alcohols, aldehydes, ketones, esters, and acids. There is a wide variation between these oxidation products with respect to flavor threshold values, concentration in the food, and flavor character. However, the *D*-value parameter, a measure of the relative intensity of flavor compounds extracted from a food, reveals that unsaturated aldehydes and ketones, namely (*Z*)-3-hexenal, 1-octen-3-one, (*Z*)-1,5-octadien-3-one, (*Z*)-2-nonenal, (*E*, *Z*)-2,6-nonadienal,

Table 1 Microbiologically produced taints in fermented foods

Food	*Microorganism*	*Tainting species*
Butter	Bacteria (*Streptococcus*)	Acetaldehyde
Cheddar cheese	Bacteria (Nonlactic)	Ethyl butanoate and ethyl hexanoate
Cheddar cheese	Yeast (*Candida*)	Ethyl acetate and ethyl butanoate
Munster cheese	Bacteria (*Pseudomonas*)	2-Methoxy-3-isopropyl-pyrazine
Yogurt	Fungi	1,3-Pentadiene
Wine	Yeast (*Brettanomyces*)	2-Acetyltetrahydro-pyridine
	Bacteria (*Lactobacillus*)	
Wine	Yeast (*Schizosaccharomyces*)	Organic sulfur compounds, e.g., 2-mercaptoethanol
Wine	Bacteria (*Pediococcus*)	Diacetyl
Cider	Bacteria (*Lactobacillus*)	Indole
Malt distillate	Bacteria (Actinomycetes)	Geosmin
Rum	Yeast (*Saccharomyces*)	3-Ethoxypropanal and 1,1,3-triethoxypropane
Beer	Yeast (*Saccharomyces*)	Diacetyl and 2,3-pentanedione
Beer	Yeast (*Saccharomyces*)	Isoeugenol and other phenols
Beer	Yeast (*Saccharomyces*)	Methional
Beer	Bacteria (*Clostridium*)	Butanoic acid
		2-Methoxy-3-isopropylpyrazine and methylisoborneol
Coffee	Molds (*Aspergillus*)	2,4,6-Trichloroanisole

and 3-methyl-2,4-nonandione, are the most potent taints. Flavor descriptors for lipid oxidation volatiles range from rancid, fishy, painty, cardboard, and metallic to mushroom, green, and fruity.

Lipase enzymes can produce taint by hydrolysis of triglycerides and release of fatty acids. The longer-chain fatty acids have higher flavor thresholds, and so lipolytic taints are more common in products containing short-chain fatty acids. Dairy products such as butter, cheese, and milk powder regularly suffer from taints described variously as rancid, butyric, goaty, and unclean, which are caused by excessive concentrations of free fatty acids of chain lengths C_4–C_{10}. Products containing coconut or hydrogenated palm kernel oil (HPKO) are prone to soapy and bitter taints caused by hydrolytic release of lauric (C_{12}) and myristic (C_{14}) acids from the triglycerides of coconut oil or HPKO.

Nonenzymic Deterioration of Food

Autoxidation

Autoxidation of lipid is initiated by the formation of lipid radicals and, subsequently, lipid peroxy radicals and lipid hydroperoxides. Volatile taints are generated from the lipid hydroperoxides via the secondary reactions described in the previous section. High-energy (short-wavelength) light is an effective initiator of lipid autoxidation. However, many other factors also influence the rate of oxidation: the number and type of double bonds; the presence of trace metals such as copper, cobalt, and iron; water activity (low-moisture foods oxidize at higher rates), temperature, and the presence of antioxidants. While dry foods such as milk powder, crisps, and breakfast cereals are particularly susceptible, fish and meat products are also prone to lipid autoxidation. (*See* **Fatty Acids**: Properties.)

Nonenzymic Browning

The importance of nonenzymic browning, also known as the Maillard reaction, in the development of desirable food flavors is well established. However, it is also recognized that the Maillard reaction is a major route for the development of taints in food.

The formation of both desirable flavors and taints via the Maillard reaction starts with a condensation reaction between a carbonyl group, usually from reducing sugars, and an amino group, most commonly from peptides or proteins. A complex sequence of chemical rearrangement, fragmentation, cyclization, and polymerization reactions occurs, which generates the flavor volatiles and the characteristic brown coloration. The generation of specific flavor chemicals depends on factors such as time, temperature, water activity, pH, and sources of carbonyl and amino groups.

Any food that undergoes substantial heat treatment during processing can develop a Maillard reaction-type taint. Maillard chemistry can also proceed significantly at room temperature and often causes taint in low-moisture foods that have been stored for long periods. For example, staleness in stored dry milk powder was thought to be caused by 2-aminoacetophenone, a Maillard degradation product of tryptophan. The same chemical, together with some furanoid species, was thought to contribute to the gluey flavor of aged casein. The concentration of furfural, a classic Maillard reaction product, in canned citrus juices was found to increase during storage and could be correlated with the development of taint in these products. Similarly, heat-treated orange juice and blackcurrants were found to contain flavor compounds thought to be derived by Maillard reactions. (*See* **Browning**: Nonenzymatic.)

Direct External Contamination

Airborne

Airborne contamination is one of the most frequent causes of direct tainting of food, and many such cases have been documented. Growing crops have been contaminated by airborne industrial discharges. For example, grapes inadvertently contaminated in this way produced wine exhibiting a taint. Exposure of food during processing and packaging operations to contaminated air either from sources external to the factory or from localized sources within a factory has resulted in many incidents of taint. Baked goods and chocolate confectionery, products that are often cooled or set using chilled air, are particularly susceptible to tainting by this route.

In one example, chocolate confectionery was found to be contaminated with a wide range of alkyl phenols. The source of the phenols was traced to an area of flooring some distance from the cooling confectionery. The flooring was found to contain a phenolic-based bonding resin that was slowly releasing alkyl phenols to the atmosphere. These chemicals were subsequently adsorbed by the confectionery. Air diffusion within warehouses or transit containers contaminated by other commodities stored either contemporaneously or at some stage in the past often results in tainting. A good illustration of this source of taint is afforded by the case of white beans that acquired a disinfectant-like taint after being transported in containers. The primary taint was identified as 6-chloro-2-methylphenol. This

compound was an impurity in a herbicide, which, during a previous voyage, had contaminated the wooden floor of the container and subsequently transferred to the beans. Examples of taints arising by these routes are outlined in Table 2.

Water

Water is used extensively in the production or processing of many foods and can be a major source of taint. Waterborne, earthy/musty taints have been discussed in the section on microbial taints. However, another common type of waterborne taint is the disinfectant-like taint caused by phenolic compounds, particularly chlorophenols. Mono-, di-, and trichlorophenols can be formed by spontaneous reaction of phenol with free chlorine present in chlorinated mains water. Plastic fittings and hoses or phenol-based resins and paints used as protective coatings on processing plant often function as ready sources of phenol.

Boiler treatment chemicals have been documented as the source of free phenol that caused chlorophenol-like taints when the boiler water was used for direct steam injection of brines and syrups. Another documented example of chlorophenol contamination of food by water involved the accidental introduction of algaecide-treated water into the processing water supply. The algaecide was found to contain tri-, tetra-, and pentachlorophenol. Disinfectant solutions used as plant-sanitizing agents have been implicated in many cases as sources of chlorophenolic taints. Several documented cases of tainting by chlorophenols are listed in Table 2.

Packaging Materials

One of the primary functions of food packaging is to prevent the ingress of contaminants and taints into the product. Unfortunately, the reverse often occurs with packaging materials frequently functioning as abundant sources of tainting chemicals. This is due in part to the ever-expanding range of materials used to package food and the increasing sophistication of packaging construction, e.g., laminates, metallized films, etc. The most common examples of direct contamination from packaging involve solvent residues from printing inks, lacquers, resins, and adhesives.

Synthetic polymers frequently used in packaging such as polythene, polystyrene, polypropylene, polyvinyl chloride, and poly(ethyleneterephthalate) often impart chemical or plastic-like odors to foods. Residual levels of monomer are often the cause of the taint. Styrene-based polymers, for example, are known to transfer styrene monomer, which has a relatively low odor threshold and imparts an objectionable taste to foods. Other polymeric packaging materials have been found to contain a wide range of volatile organic compounds, many of which could cause taint.

Table 2 Taints from external sources

Food contaminated	*Compounds involved*	*Origin*
Biscuits	Chlorophenols	Airborne from nearby herbicide factory
Chocolate confectionery	Xylenols	Airborne from adjacent area of composite flooring
Chocolate crumb	2,4- and 2,6-dichlorophenol	Airborne from walls of metal container
White beans	6-Chloro-2-methyl-phenol	Airborne from wooden floor of metal container
Bakery and confectionery products	Petroleum hydrocarbons	Airborne from petrol spillages in warehouse
Canned vegetables	Tri-, tetra-, and pentachlorophenol	Algaecide in process water
Canned fruit	Di- and trichlorophenols	In-plant reaction of steam contaminated with phenol and chlorinated water
Fruit juice	2,4- and 2,6-dichlorophenol	Phenol from in-plant fitments reacting with chlorinated water
Distilled spirit	Di- and trichlorophenol	Phenol in spirit chlorinated by sanitizing agent
Jam filling	6-Chloro-2-methylphenol	Tanker with residual traces of disinfectant
Channel catfish	2-Methylisoborneol	Pondwater contaminated with phytoplankton
Water clams	Geosmin	Marshwater
Cocoa powder	Chlorophenols and chloroanisoles	Multiwalled paper sacks with glued seams
Milled flour	Various	Jute sacks
Chocolate assortments	Aliphatic hydrocarbons	Lithographic printed cartonboard
Icecream	2-Butoxyethanol	UV-cured printed cartonboard
Cereal	Terpenes	Resin in laminated liner
Sugar confectionery	Aliphatic aldehydes	Printed cellophane bag
Soft drink	2-Methylphenol	Can lacquer
Milk beverage	Isophorone	Can lacquer
Biscuits	Styrene	Polystyrene trays
Spring water	2-Ethylhexanol and hexanal	Liner of bottle closure

Another major route for contamination of food products with chlorophenols, apart from water as discussed previously, is via packaging materials and materials encountered during transport of foods. Cork closures, Kraft bleached paper sacks, cardboard boxes, jute sacks, fiberboard containers, and wooden pallets have all been shown to contain chlorophenols, which, in many cases, were the primary tainting species and, in others, precursors of the eventual tainting species.

Paper-based packaging materials, particularly those made from recycled paper and pulp, have recently been found to contain 2,4,6-tribromophenol in high concentrations. Given that it is a powerful taint with an odor threshold in water of $0.6\,\mu g\,l^{-1}$, this compound will also impart a medicinal taint if transferred to foods.

Tables 2 and 3 show that nearly all food packaging materials, except perhaps glass, can cause tainting of food products. Invariably, the tainting chemicals will be present in packaging materials at substantially higher levels than are found in the food. Therefore, monitoring of packaging for tainting chemicals by sensory evaluation or instrumental analytical methods may prevent any potential problems.

Indirect External Contamination

Polyhaloanisoles

The polyhaloanisole family contains some of the most potent odor compounds known and are notorious as the source of musty or moldy taints in foods and beverages. Unlike many of the other tainting chemicals discussed previously, the presence of polyhaloanisoles in food or food-associated materials is completely adventitious. As chemicals, they perform no known useful function, and their presence in food systems invariably occurs indirectly, since their genesis requires both a microbiological vector (to effect the methylation stage) and a source of halophenols. Timber-based packaging materials have been implicated in many of the documented cases of haloanisole taint. The first examples of a chloroanisole taint in food occurred in eggs and broiler chickens. It was shown that both tetra- and pentachloroanisole had been formed by microbial methylation of the corresponding chlorophenols present in wood shavings used as poultry litter.

A more recent occurrence has been the emergence of polybromoanisoles, and particularly 2,4,6-tribromoanisole, as powerful taints rivalling the polychloroanisoles in terms of potency. As with the chloroanisoles, timber-based packaging has been shown to be the primary source of the bromoanisole precursor, i.e., the corresponding bromophenol. Thus, the biocidal agent 2,4,6-tribromophenol can be readily converted by microbial methylation to 2,4,6-tribromoanisole.

As the sensitivity and selectivity of analytical techniques for the detection of polyhaloanisoles in food have improved, so the number of documented cases of taint by these chemicals has increased. Some of the more recent examples of foods affected by haloanisole taint are listed in Table 3.

Table 3 Foods tainted by haloanisoles

Food contaminated	*Haloanisoles detected*	*Origin*
Wine	2,4,6-TCA[a]	Cork closures containing the chlorophenols
	2,3,4,6-TeCA[a]	
Cognac	2,4,6-TCA	Cork closures
Dried fruit	2,4,6-TCA	Packaging adhesive containing chlorophenols
	2,3,4,6-TeCA	
	PCA[a]	
Raisins	2,3,4,6-TeCA	Paper/timer packaging materials
Cocoa powder	2,4- and 2,6 DCA[a]	Multiwalled paper sacks containing chlorophenols
	2,4,6-TCA	
	2,3,4,6-TeCA	
Desiccated coconut	2,4,6-TCA	Multiwalled paper sacks containing chloroanisoles
	2,3,4,6-TeCA	
Cheese	2,4,6-TCA	Wooden shelves containing chlorophenols and chloroanisoles
	2,3,4,6-TeCA	
Spring water	2,3,4,6-TeCA	Cardboard boxes used to store empty bottles
Canned beer	2,4,6-TBA; 2,3,6-TBA	Can lids
Skimmed milk	2,4,6-TBA	Unknown
Flour	2,4-DCA; 2,4,6-TCA	Jute sacks containing chloroanisoles and chlorophenols
	2,3,4,6-TeCA; PCA	
Wheat/wheat flour	2,4,6-TCA	Unknown – but not packaging materials

[a]DCA, dichloroanisole; TCA, trichloroanisole; TeCA, tetrachloroanisole; PCA, pentachloroanisole; TBA, tribromoanisole.

Catty Taint

The development of offensive catty odors in foods such as meat, canned vegetables, and cheeses is another example of indirect tainting. In these cases, a relatively innocuous substance was transferred to the food, where it was transformed into a powerful taint by reaction with a minor component of the food. The compound responsible for the catty taint was 4-mercapto-4-methylpentan-2-one, formed by the addition of hydrogen sulfide across the double bond of mesityl oxide. The latter was present in polyurethane paint, solvents used to dilute can lacquers and plastic packaging, and was presumed to have been transferred to the foodstuff from these materials.

Geranium Taint

The development of geranium-like taint in German wines occurred, as above, by reaction of two innocuous components to form the tainting substance. In this case, one of the reactants had been formed by acid-catalyzed rearrangement of a specific microbial degradation product of the preservative used in the fermentation. Sorbic acid (the preservative) was enzymically reduced to 2,4-hexadien-1-ol, which underwent rearrangement to 3,5-hexadien-2-ol. This product reacted with ethanol to form 2-ethoxyhexa-3,5-diene, the compound held to be responsible for the taint.

Medicinal Taint

Another good example of taint arising indirectly through a cascade of chemical reactions between ostensibly innocuous food components is afforded by the example of marinated herring that developed an offensive, medicinal taint. The taint was caused by 2-bromophenol. This chemical was formed by the reaction of bromine with phenol, the latter being present as a minor impurity in the acetic acid used for the marinade. The bromine was thought to be formed via the reaction of the hydrogen peroxide used to bleach the herring with bromide ion present as an impurity in the brine component of the marinade.

See also: **Browning**: Nonenzymatic; Toxicology of Nonenzymatic Browning; **Contamination of Food**; **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage; Fungi in Food – An Overview; Molds in Spoilage; Yeasts in Spoilage; **Taints**: Analysis and Identification

Further Reading

Heath H and Reineccius GA (1986) *Flavour Chemistry and Technology*. New York: Van Nostrand Reinhold.

Nijssen B (1991) Off-flavours. In: Maarse H (ed.) *Volatile Compounds in Foods and Beverages*, pp. 689–735. New York: Marcel Dekker.

Reineccius G (1991) Off-flavours in foods. *Critical Reviews in Food Science and Nutrition* 29: 381–402.

Saxby MJ (1996) *Food Taints and Off-flavours*, 2nd edn. Glasgow: Blackie Academic & Professional.

Whitfield FB and Last JH (1986) Off-flavours encountered in packaged foods. In: Charalambous G (ed.) *The Shelf-life of Foods and Beverages*, pp. 483–524. Amsterdam: Elsevier.

Whitfield FB and Tindale CR (1984) The role of microbial metabolites in food off-flavours. *Food Technology in Australia* 36: 204–232.

Whitfield FB, Hill JL and Shaw KJ (1997) 2,4,6-Tribromoanisole: a potential cause of mustiness in packaged food. *Journal of Agricultural and Food Chemistry* 45: 889–893.

Analysis and Identification

B D Baigrie, The Lord Zuckerman Research Centre, University of Reading, Whiteknights, UK

Background

The wide diversity of chemical compounds that can cause taints and off-flavors in fresh and processed food, the various mechanisms and routes by which they may be imparted to foods, and the inherent complexity of food flavors all conspire to make the analysis and identification of taints and off-flavors a challenging and often time-consuming exercise requiring the application of a unique combination of scientific skills and analytical expertise.

Sensory Evaluation

Food manufacturers who operate to high standards of good manufacturing practice with a well-established quality assurance program and a proactive quality control scheme will normally detect grossly tainted product, usually by sensory evaluation, before it leaves the factory. However, some taints may occur intermittently within a production batch or only arise during the shelf-life of the product as a consequence of storage conditions or some other parameter. The manufacturer may then only be alerted to a taint problem indirectly through merchandise returns from retailers or directly via consumer complaints.

The first clues to the identity of the taint may come from the complainant's comments. Nevertheless, it is unwise to rely solely on untrained consumers for

an accurate description of a taint, even if large numbers of complaints about taint in a product have been received. Invariably, untrained consumers will generate a lengthy list of taint descriptors, which are often confusing, sometimes contradictory, and therefore of limited value to the analyst. The first stage in the analysis of a taint, therefore, is sensory evaluation of the foodstuff using a panel of trained experts.

It is important to use a panel of experts specifically trained to detect and describe taints, because, to any particular chemical, there will exist within a population of consumers a wide range of sensitivities. Taints are more likely to be detected by those in the most sensitive region of the distribution, and hence panels should comprise individuals with sensitivities considerably greater than that of the population mean.

For well-characterized types of taint, e.g., TCP, medicinal, antiseptic, panelists can be selected for their sensitivity to the taint species in question (usually halogenated phenols or halogenated alkylphenols). However, it should be recognized that sensitivity to a specific taint does not necessarily imply high sensitivity to taints in general.

Selection of a panel normally involves a measurement of their threshold to that taint. This requires the panelists to assess, in a specified medium, a range of stimulus concentrations ranging from zero to concentrations that can be clearly detected by all the panel. A threshold is usually defined as the concentration of taint that 50% of the panelists can perceive. For taints, two types of threshold are typically used, namely detection threshold and recognition threshold.

The ability of panelists to describe accurately the sensory properties of a taint and, if possible, to relate it to known standards is particularly useful, since it can provide vital information to the analyst or flavor chemist about the chemical structure of the taint. In practice, therefore, greater credence is placed on the comments of individual panelists who are known to be highly sensitive to certain taints, even if they constitute a minority opinion among the panelists.

Since most compounds responsible for taint are volatile, aroma assessment of the suspect food is mandatory. However, in some cases, for example taint caused by polychloroanisoles, the taint may only manifest itself when the food is held in the mouth or as a characteristic aftertaste after swallowing.

Once sensory evaluation has confirmed that a product is tainted and perhaps provided some clues to its nature, the next requirement is to establish its chemical structure and then to identify the source of the taint. This involves isolation of a concentrated flavor extract and separation of the individual flavor components. If the extract is particularly complex, fractionation into functional group classes followed by separation of individual components in each class may be necessary. Finally, identification of the taint is usually achieved using an appropriate spectroscopic technique. (*See* **Sensory Evaluation**: Aroma.)

Extraction of Taint

The first major challenge for the flavor analyst is to obtain a volatile flavor extract that represents the characteristic flavor of a food. There can be no absolute certainty that a valid extract (i.e., one characteristic of the food flavor) will be obtained. Nor is there any assurance that subsequently identified compounds were actually present in the food since artefacts can be introduced by many routes. However, with tainted foods, it is much easier to ascertain if an extract is valid, simply by assessing the odor or taste of the extract with a panel of assessors sensitive to the taint.

A wide variety of techniques exists for extracting taints from foods. Some of the isolation methods most frequently used are summarized in Table 1.

Where taints are volatile and readily detected by odor in a food or beverage, headspace sampling is often the technique of choice. Static headspace sampling is simple to perform and allows a truly valid flavor extract to be obtained. However, the quantity of the tainting species delivered to the analytical instrument is often too low to enable detection and identification to be achieved.

Table 1 Methods for isolation and/or concentration of taints

1. Headspace sampling
   - Static headspace
   - Dynamic headspace
     - Purge and trap
     - Closed-loop stripping
     - Solid-phase microextraction
2. Distillation
   - Steam distillation
   - Simultaneous steam distillation–extraction with Danish–Kuderna apparatus
   - Vacuum distillation
3. Liquid–liquid extraction
   - Discontinuous extraction with immiscible solvent
   - Continuous extraction with immiscible solvent and Kutscher–Steudel apparatus
4. Liquid–solid extraction
   - Solvent extract of solid in Soxhlet apparatus
   - Solid-phase microextraction
   - Stir bar sorptive extraction

Dynamic headspace methods such as purge and trap, closed loop stripping, and solid-phase microextraction (SPME) are used instead to overcome the problem of adequate detection limits inherent to static headspace sampling. In dynamic sampling, headspace volatiles are removed from the gas phase above the food and concentrated on a suitable adsorbent (activated charcoal, porous polymers, macroreticular resins, polymeric gas chromatography phases) and then desorbed either thermally or with a solvent.

As with static headspace, these dynamic collection techniques are simple and cause minimal artefact formation. SPME, which is a recent technological innovation, was introduced by Janusz Pawliszyn, University of Waterloo, Ontario as a solvent-free sample preparation technique. The method utilizes a fused silica fiber coated with an adsorbent polymeric phase. The fiber is mounted for protection in a syringe-like device, and by depressing the plunger of the syringe the fiber can be exposed to the headspace above the sample matrix. After partition of the headspace volatiles into the coating, the fiber is retracted into the needle and the device interfaced with the appropriate analytical instrument. SPME has been utilized extensively in the analysis of taints and other contaminants in foods. For example, it is used to monitor the presence of 2,4,6-trichloroanisole, the species responsible for cork taint in wines and distilled beverages.

Where the tainting species is only moderately or poorly volatile, headspace methods usually provide insufficient quantities for detection or identification. In these circumstances, quantitative removal of the taint from a larger sample size is a prerequisite, and the method most frequently chosen, particularly if the food matrix is solid, is simultaneous steam distillation-extraction (SSDE). This technique has the added advantage of being able to be operated under reduced pressure, thus minimizing possible thermal decomposition of the tainting species. SSDE has been used extensively to extract chlorophenol, chloroalkylphenol, and chloroanisole-type taints from various food matrices.

If the taint is not steam volatile, techniques based on direct extraction with solvents such as liquid–liquid, Soxhlet extraction, or extraction using the sorptive properties of gas chromatography (GC) polymeric phases, e.g., SPME in matrix immersion mode or stir bar sorptive extraction (SBSE), can be used.

Liquid–liquid and Soxhlet extraction are traditional methods with a long history of usage, but extracts obtained by these techniques may contain variable amounts of nonvolatile material, and additional separation techniques may need to be applied. SBSE is a recently developed extraction technique, similar in principle to SPME, but offering potentially much lower detection limits than SPME. In this technique, a glass magnetic stir bar coated with a polydimethylsiloxane (PDMS) stationary phase is stirred in the sample for several minutes. The analytes of interest come into contact with the PDMS phase and are extracted. Thermal desorption and online gas chromatography-linked mass spectrometry (GC–MS) analysis complete the analytical methodology.

Fractionation of Flavor Extracts

Fractionation of a flavor extract is best achieved using high-resolution GC on fused silica columns. However, despite recent advances in column technology in terms of stability, inertness, resolution, and sample capacity, major difficulties can still be encountered due to the inherent complexity of flavor extracts from many foods and the presence within the extract of only minute quantities of the tainting chemical relative to the other flavor components.

To minimize these difficulties, flavor extracts can be fractionated prior to analysis by GC. For example, flavor extracts, particularly those from SSDE, can be readily fractionated on the basis of chemical functionality into acids, phenols, bases, and neutrals. Fractionation of flavor extracts by conventional open column chromatography on silica gel is also well established. More recently, fractionation of food extracts by high-performance liquid chromatography (HPLC) has been utilized, but to date, this approach has found limited applicability.

An elegant way to optimize the separating power of modern GC capillary columns is fractionation of flavor extracts on two different columns – so-called multidimensional gas chromatography (MDGC). In this technique, initial separation is effected on a precolumn, and then selected fractions are introduced on to a second analytical column, where they are further separated into individual components. The columns can be located in the same oven or in two different ovens: the precolumn can be a packed column, a wide-bore capillary column, or even a high-resolution analytical capillary column coated with a different stationary phase from the second column. The selection of fractions of interest can be optimized by the introduction of a splitting device between the precolumn and detector so that GC effluent from the precolumn can be assessed at a sniff

port. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Detection and Identification of Taints

In addition to the superiority enjoyed by GC over other analytical techniques as a means of fractionating complex flavor extracts, it also possesses major advantages in terms of detection and identification of taints.

GC Detectors

The ubiquitous flame ionization detector (FID) is sensitive and is virtually a universal detector for volatile taints. It responds to most organic compounds but not to air, water, or light fixed gases. Complementing this detector are a variety of GC detectors, which are more specific and, in many cases, more sensitive as well.

The thermionic ionization detector (NPD) is specific for compounds containing nitrogen and/or phosphorus. Known fecal-like tainting compounds such as indole or skatole can be readily detected with the NPD, as can 2-acetyltetrahydropyridine, the cause of mousy taint in wines and cider. More recently, a nitrogen-specific detector based on chemiluminescence, which has a linear, equimolar response and is not quenched by hydrocarbons, has been developed. It has been used to monitor nitrogen-containing flavor volatiles in wine extracts.

The flame photometric detector (FPD) is a very selective and sensitive detector that measures chemiluminescence above a hydrogen-rich flame. By the use of filters and a photomultiplier, it can readily measure nanogram quantities of phosphorus or sulfur-containing compounds. It has been used successfully in many taint investigations and notably in the identification of taints in beer caused by sulfur-containing species.

Recent advances in the development of sulfur-selective detectors have culminated in a new sulfur chemiluminescence detector, which provides a linear and equimolar response to sulfur-containing compounds without interference from most sample matrices.

The photoionization detector (PID) can be used to analyze a wide variety of compounds, including aliphatic and aromatic hydrocarbons, amines, and organic sulfur compounds. Key features of the PID that make it suitable for detecting taints are its sensitivity, generally lower than an FID or FPD, and its wide linear dynamic range. It is also nondestructive and can be coupled in series with other detectors.

Several common tainting species containhalogen atoms, e.g., bromo/chlorophenols and bromo/chloroanisoles, and are therefore readily detected using an electron capture detector. This detector has probably the lowest detection limits of all the common GC detectors and is highly sensitive and selective for molecules containing electronegative atoms; it is, however, very limited in a linear range and requires extensive calibration for quantitative results.

The atomic emission detector is perhaps the most versatile of the selective detectors, since it can be tuned to detect virtually all of the elements in compounds eluting from a gas chromatograph (except helium). With an excellent sensitivity and selectivity, it can be used as a screen for elements to locate target or unknown peaks of interest in complex food matrices.

Despite the many advantages offered by this variety of electronic detectors, biological detectors such as the nose and tongue are probably best suited for detecting the presence of taints in food/flavor extracts. The technique of sniffing the effluent gas as it emerges from the GC – so called odor port gas chromatography (GC-O) – is commonly used in aroma analysis and is particularly useful for locating the position of tainting odorants in a gas chromatogram. The area(s) of the chromatogram that correspond to the taint can be identified, and the number and nature of the tainting species can often be ascertained. This greatly simplifies the primary task of the spectroscopist, which is to establish a chemical structure for each taint.

The usefulness of GC-O was readily illustrated in a recent study of a distilled beverage with an earthy, musty taint. Preliminary analysis of the flavor extract from the tainted beverage identified three members of the polychloroanisole family of taints, but all at subthreshold levels. Analysis by GC-O identified a region in the chromatogram with a strong musty aroma similar to that possessed by the beverage. The GC retention time was different from that of any of the three characteristically musty polychloroanisoles but identical to that of 2-methylisoborneol, a known powerful taint with a musty/camphoraceous odor and taste. Mass spectroscopic data were consistent with the presence of MIB in the tainted beverage. When the untainted beverage was spiked with MIB at $5\,\mu g\,kg^{-1}$, the taint was reproduced. (*See* **Chromatography**: Gas Chromatography.)

Mass Spectrometry

In view of the extremely low levels at which taint in food can be detected by the human sensory organs

and of the limitations on the quantity of taint that can be isolated for identification, the analytical technique best suited for detection and identification of taints is high-resolution GC-MS. Detection and identification of tainting compounds can usually be achieved using GC-MS operated in full scan mode. However, GC-MS operated in either the mass chromatography mode or selected ion monitoring mode is unparalleled in its ability to deliver both sensitivity and selectivity of detection.

Mass chromatography (MC) allows reconstruction of the ion current profiles of selected ions as a function of time from continuously collected data. One of the simplest uses of MC in taint studies is the location of certain classes of compounds in complex mixtures by looking for the presence of characteristic ions from the total ion current trace. For example, by examining at a resolution of 5000 the fragment ion *m/z* 45 (CHS^+) that occurs widely in both aliphatic and aromatic sulfur compounds, it is possible to separate this species from others of the same nominal mass (e.g., $C_2H_5O^+$) and, moreover, to detect those regions in the chromatogram where the sulfur-containing species occur.

With selected ion-monitoring (SIM), the peak selector of the mass spectrometer can record either single or multiple ion characteristics of any specific compound being searched for to provide a series of 'fragmentograms' from separated peaks during a GC run. By monitoring ions of one or just a few specific masses instead of the whole spectrum, a thousand-fold increase in sensitivity (picogram quantities) can be attained.

The successful application of GC-MS operated in the SIM mode to the identification of taints has been readily demonstrated by several groups of workers, and many case histories utilizing this technique have been extensively documented in the scientific literature. (*See* **Chromatography**: Combined Chromatography and Mass Spectrometry.)

Quantification of Taints

In many taint studies where only a limited quantity of tainted food is available for analysis, quantification of the taints by GC-MS is achieved by comparing the SIM responses for each ion monitored with those obtained with standard solutions of the tainting chemical. In these cases, it is usually assumed that complete extraction of the taint from the foodstuff has been achieved. This highlights the major difficulty in obtaining accurate quantification of odorants in general and taints in particular, namely the need to extract, enrich, and separate the taint without losses.

The isotope dilution assay is probably the most precise method for quantifying minor constituents of food such as taints. In this technique, the tainting chemical now labeled with deuterium or carbon-13 is used as an internal standard. Since the labeled standards have virtually the same chemical and physical properties as the taint, the overall efficiency of the extraction and enrichment stages does not influence the result. However, for a solid food, usually only the extract containing the volatile fraction, and not the food itself, is spiked with the labeled internal standard, and 100% efficiency of extraction is again assumed.

Despite the difficulties inherent in these methodologies, quantitative measurements are an essential part of taint studies. In tandem with sensory evaluation, they provide complementary data proving that the compound identified is responsible for the taint.

It is important to remember that taint in foods is a matter of perception: it is the taste/odor threshold level of an odorant, not its presence *per se*, that determines if it will cause taint.

See also: **Chromatography**: Combined Chromatography and Mass Spectrometry; Gas Chromatography; **Quality Assurance and Quality Control**; **Sensory Evaluation**: Food Acceptability and Sensory Evaluation; Aroma; Taste; **Taints**: Types and Causes

Further Reading

Grosch W and Schieberle P (1991) Bread. In: Maarse H (ed.) *Volatile Compounds in Foods and Beverages*, pp. 41–77. New York: Marcel Dekker.

Hirvi T and Honkanen E (1984) Selected ion monitoring technique and sensory analysis in the evaluation of the aroma of berries. In: Nykanen L and Lehtonen P (eds) *Flavor Research of Alcoholic Beverages*, pp. 275–278. Helsinki: Foundation for Biotechnical and Industrial Fermentation Research.

Kilcast D (1996) Sensory evaluation of taints and off-flavours. In: Saxby MJ (ed.) *Food Taints and Off-flavours*, pp. 1–40. Glasgow: Blackie Academic & Professional.

Leland JV, Schieberle P, Buettner A and Acree TE (2001) *Gas Chromatography–Olfactometry. The State of the Art. ACS Symposium Series 782*. Washington, DC: American Chemical Society.

Maarse H and Belz R (eds) (1981) *Isolation, Separation and Identification of Volatile Compounds in Aroma Research*. Berlin: Akademie-Verlag.

Pawliszyn J (1997) *Solid Phase Microextraction. Theory and Practice*. New York: Wiley-VCH.

Pawliszyn J (ed.) (1999) *Applications of Solid Phase Microextraction*. Cambridge: Royal Society of Chemistry.

Schreier P (ed.) (1984) *Analysis of Volatiles – Methods. Applications*. Amsterdam: de Gruyter.

TANNINS AND POLYPHENOLS

B G Swanson, Washington State University, Pullman, WA, USA

Background

Tannins are generally defined as soluble, astringent complex phenolic substances of plant origin used in tanning of animal skins or precipitation of proteins. Tannins are chemically defined as phenylpropanoid compounds often condensed to polymers of variable length. Phenolic compounds are chemically defined as compounds containing hydroxylated aromatic rings, the hydroxy group being attached directly to the phenyl, substituted phenyl, or other aryl group. Tannins and phenolic compounds are widely distributed secondary metabolites in plants, and play a prominent role in general defense strategies of plants, as well as contributing to food quality. The terms 'hydrolyzable' and 'condensed' tannins are used to distinguish between the two important classes of vegetable tannins, namely gallic acid-derived and flavan-3,4-diol-derived tannins, respectively. (*See* **Phenolic Compounds**; **Vegetables of Temperate Climates**: Commercial and Dietary Importance.)

Phenolic Compounds

Phenolic compounds are widely distributed in plant tissues, particularly contributing color, flavor, and astringency to fruits. The concentration of phenolic compounds may vary from 0.5 to 5.0 g per 100 g dry weight of plant tissues. Phenolic compounds are often considered secondary metabolites of plant metabolism that contribute little to the physiological or ecological functions of the plant. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Most phenolic compounds are believed to be byproducts of the metabolism of the aromatic amino acid phenylalanine. A major group of water-soluble phenolic compounds, the anthocyanins, contribute colors to fruits. Two identified anthocyanins, phloridzin in apples and arbutin in pears, occur only in plant tissues and are absent in associated fruits.

The greatest concentrations of phenolic compounds in plants are normally present as cinnamic acid derivatives and flavan monomers, dimers, and polymers.

Cinnamic Acid Derivatives

Cinnamic acids play key roles in the formation of other more complex phenolic compounds. The cinnamic acids (Figure 1) are rarely present in uncombined forms, occurring primarily as esters of quinic acid, but may also be esterified to malic or tartaric acids, or sugars. Chlorogenic acid (5-caffeoylquinicacid) (Figure 2) is perhaps the most important cinnamic acid observed in fruits, contributing 25% of the dry weight of the bilberry (*Vaccinium*) fruit. Chlorogenic acid can be isolated from green coffee beans, and forms a black compound with iron, believed to be responsible for the blackening of cut or cooked potatoes. Anthocyanin and flavonoid glycosides are also acylated by cinnamic acids through sugar hydroxyl groups, with *p*-coumaric acid the most common acylating agent. In addition to forming esters, hydroxylated cinnamic acids also form glycosides with sugars. (*See* **Coffee**: Green Coffee; **Fruits of Temperate Climates**: Fruits of the Ericacae.)

Cinnamic Acids in Fruits

The concentration of cinnamic acid derivatives declines as fruits mature. However, the total quantity

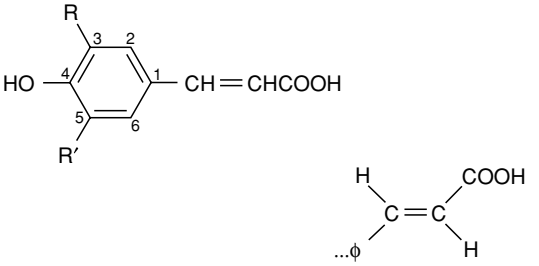

Figure 1 *Trans*-cinnamic acids: *p*-coumaric, R=R′=H; caffeic, R=OH, R′=H; ferulic, R=OCH_3, R′=H; sinapic R=R′=OCH_3.

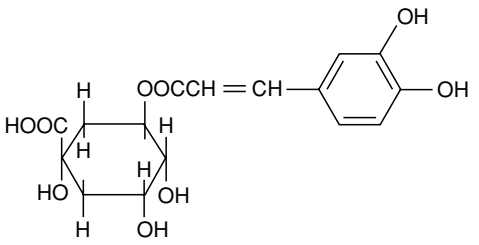

Figure 2 Chlorogenic acid.

of cinnamic acids increases as the fruits enlarge during the growing season. Cinnamic acids accumulate in the skin or peel of fruits. Cinnamic acids, especially chlorogenic acid, which is extensively destroyed during roasting, are important to the acceptability of coffee. (*See* **Ripening of Fruit**.)

Flavans

Flavans are widely distributed in nature and result from a double reduction of a flavanone. The general biochemical derivations of flavonoids are presented in Figure 3. Many natural flavans are lipid-soluble and are prominent in the skin or peel of fruits and in the cutin of leaf surfaces. A number of flavans are phytoalexins, which impart fungi or insect resistance to the plant tissues. Flavans are generally found in greater concentrations in immature fruits compared with mature fruits.

Flavans are not generally observed in plant tissues as monomers, glycosides, esterified, methylated, or in any heterogenous combination, except for galloyl esters in grapes and tea. Flavans generally are condensed to form 'condensed tannins,' more correctly referred to as 'proanthocyanidins.' Flavan-3-ols, catechin and epicatechin, are the largest class of monomeric flavans. A considerable number of flavan-3-ol glycosides can be isolated from plant tissues. (*See* **Tea: Chemistry**.)

The nomenclature of flavonoid and proanthocyanidin structures changes as the number and complexity of this group of plant metabolites expand. Leucoanthocyanidins are defined as monomeric flavonoids that produce anthocyanidins by cleavage of a carbon–oxygen bond on heating with mineral acid. Condensed proanthocyanidins are defined as flavan-3-oligomers, or more generally as proanthocyanidins that produce anthocyanidins by cleavage of a carbon–carbon bond. The isolation, structure, and properties of 'insoluble' proanthocyanins constitute a major unresolved problem for phytochemists, and enologists in particular.

Anthocyanidins and Anthocyanins

Anthocyanins are glycosides composed of the anthocyanidin aglycone plus one or more glycosidically bonded mono- or oligosaccharidic units. Anthocyanins are identified structurally by name (Figure 4). Six anthocyanins are water-soluble pigments common in flowers and plant tissues; although many other anthocyanins exist, they are rather limited among common food plants.

Color variations among the anthocyanins reflect structural differences in the number of hydroxyl groups, the presence or absence of methylation and glycosylation, as well as the distribution of positive charge about the aryl-substituted chroman ring system. Anthocyanins undergo structural alterations, depending upon the pH and ionic strength of the aqueous environment.

Tannins

Tannins are generally defined as polyphenolic compounds, with a molecular weight greater than 500, that precipitate proteins from an aqueous solution. Tannins precipitate proteins owing to functional groups that interact with two or more protein molecules, to form hydrophilic or hydrophobic complexes, building large cross-linked insoluble protein-tannin complexes. Hydrolyzable tannins are classified as gallotannins or ellagitannins on the basis of tannic acid structure relative to gallic or ellagic acids. Hydrolyzable tannins, including tannic acid, can be acid- or enzymatically hydrolyzed to yield glucose and gallic acid.

A second category of polymeric flavonoids, the condensed tannins are not hydrolyzed under physiological digestive conditions, but upon severe acid or alkaline treatment, yield less-soluble polymeric phlobaphanes or monomeric flavonoids such as catechin or epicatechin. Condensed tannins notably contain linkages between the 4-position of one catechin residue and the 6- or 8-position of another flavonoid. The condensed tannin structures are not to be confused with tannic acid, the representative structure of the hydrolyzable tannins.

Analysis of Polyphenolic Compounds

Extraction and separation of anthocyanin compounds generally involve maceration of plant tissue in methanol containing 1% HCl. Qualitative assays with thin-layer chromatography (TLC) have evolved to quantitative assays with high-performance liquid chromatography (HPLC), often utilizing selective absorption wavelengths to determine the identity and concentration. (*See* **Chromatography: Thin-layer Chromatography; High-performance Liquid Chromatography**.)

Much of the emphasis in contemporary natural products chemistry lies in the elucidation of the biological activity of plant metabolites, rather than in the intrinsic interest of the structures themselves. The rationale for the study of proanthocyanidins is to develop a knowledge of proanthocyanidins in plant

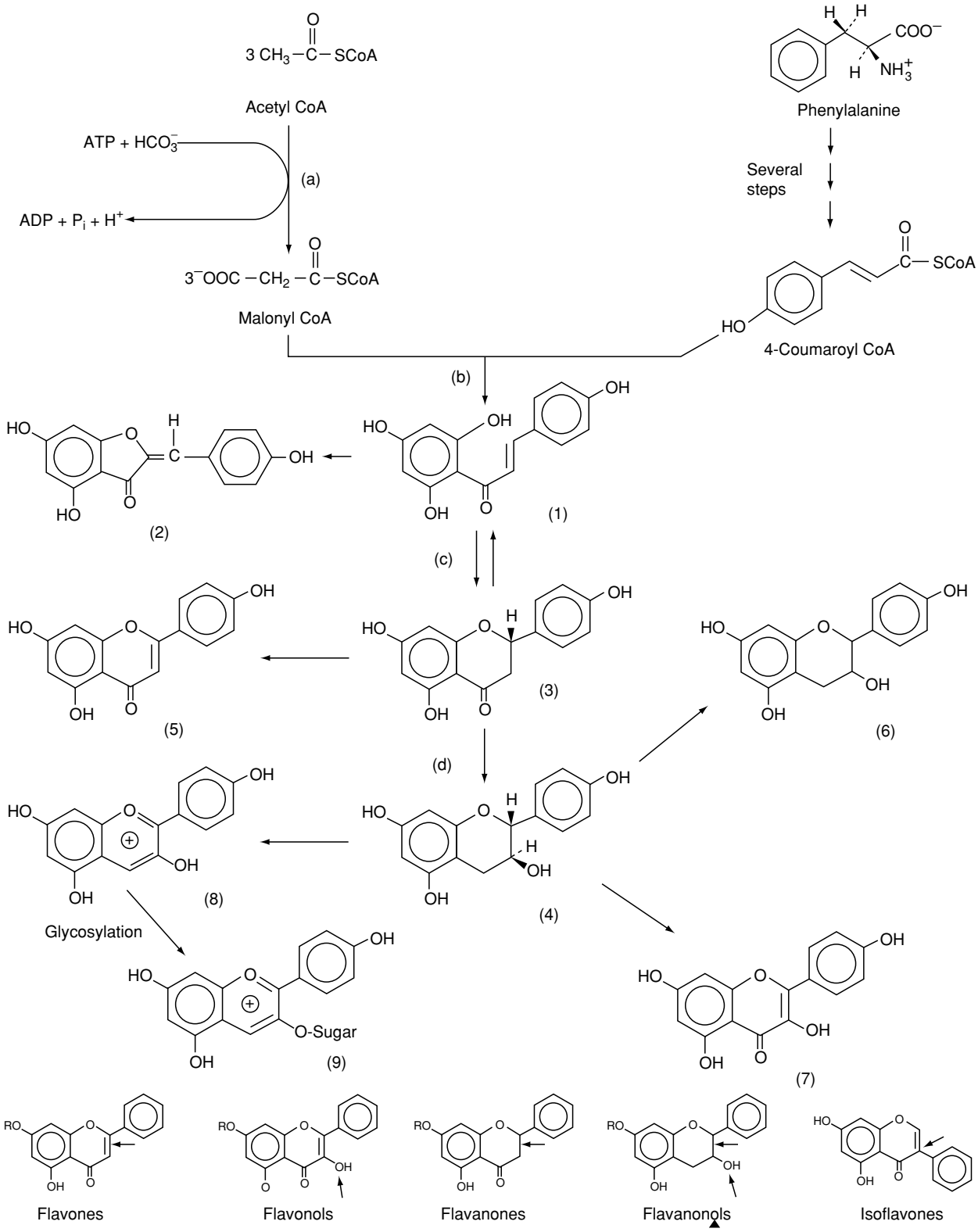

Figure 3 General biochemical derivations of flavonoids. A structural key for the group of compounds is indicated at the bottom of the figure: flavones, position 2,3 is unsaturated; flavonols, C-3 hydroxyl is present; flavanones, 2,3-site is saturated; flavanols, C-3 hydroxyl is present and site 2,3 is saturated; and isoflavones, C-3 phenyl ring substituent is displayed. Pathway key: (a) acetyl-CoA carboxylase; (b) chalcone synthase; (c) chalcone isomerase; (d) flavanone 3-hydroxylase; (1) chalcone; (2) aurone; (3) flavanone; (4) dihydroflavonol; (5) flavone; (6) catechin; (7) flavonol; (8) anthocyanidin; (9) anthocyanin. Isoflavone in the interim formation of chalcones (1) to flavanones (2).

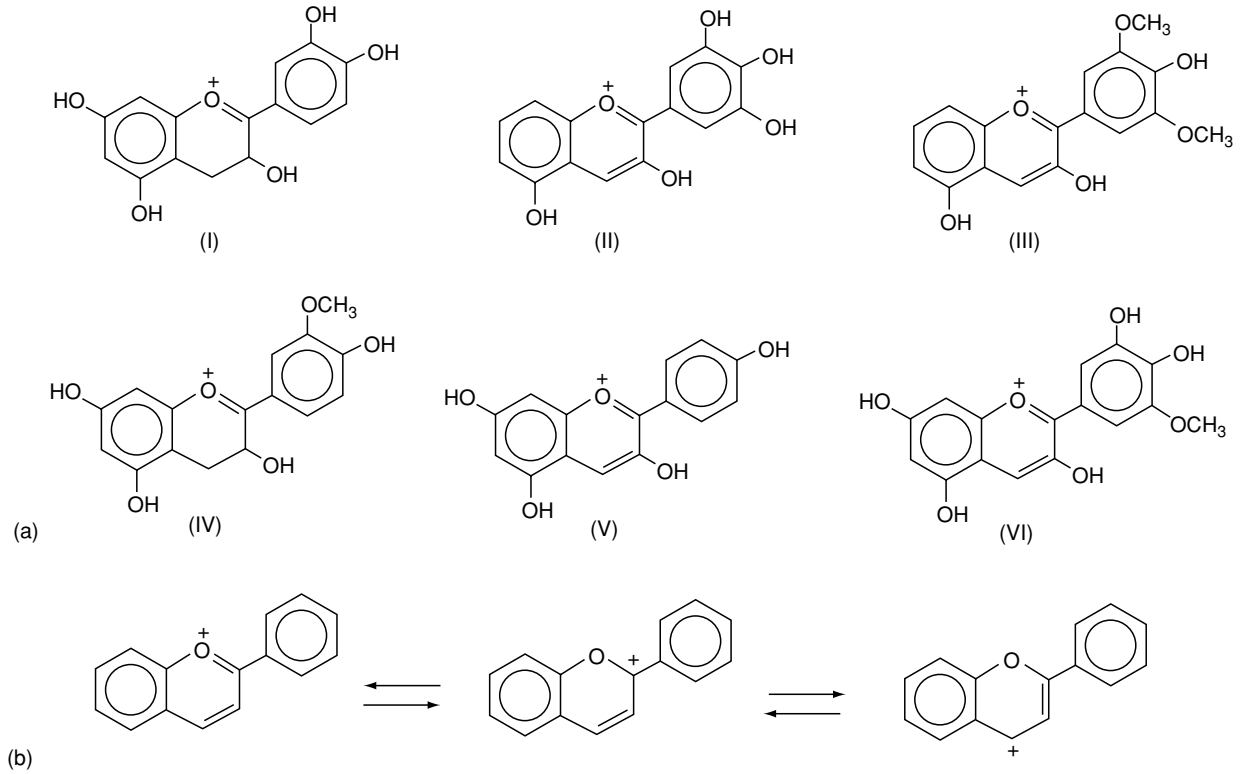

Figure 4 (a) Selected structures for six of the most common anthocyanidins. All anthocyanidins have hydroxyl groups in the 3-, 5- and 7-positions, although each structure may have its own characteristic hydroxyl or methoxyl groups on the so-called B ring: (I) cyanidin, (II) delphinidin, (III) malvidin, (IV) peonidin, (V) pelargonidin, and (VI) petunidin. These compounds occur in many species of fruits and vegetables: apple (I); blackcurrant (I, II); blueberry (I, II, III, IV, VI); cherry (I, IV); grape (I–VI); orange (I, II); peach (I); plum (I, IV); radish (V); raspberry (I); red cabbage (I); strawberry (V-major,I). (b) Possible positive charge distribution over a chroman ring system found in the structure of anthocyanidins.

biochemistry relative to structure and chemistry. To understand the biochemistry of physiologically active proanthocyanidins, the development of strategies to separate and purify proanthocyanidins in an unmodified state is necessary. Using alcoholic solvents, aqueous methanol or ethanol, mono-, di- and trimeric flavon-3-ols are separated from plant tissue with Sephadex LH-20 or a high-porosity polystyrene gel (CHP 20P). More recently, column chromatography on reversed-phase C8, C18 or CN supports has provided alternative separation techniques. HPLC on reversed-phase columns is used extensively for analytical and preparative separation of proanthocyanidins. High-resolution separation is achieved with solvent programming and dilute acid suppression of hydroxyl group ionization. Structural elucidation of proanthocyanidins may be readily determined with nuclear magnetic resonance and circular dichroism spectroscopy. (*See* **Spectroscopy**: Nuclear Magnetic Resonance.)

The condensed tannin content of plant tissues is generally determined by the traditional vanillin-HCl method or protein precipitation.

Physiological and Health Effects

Tannins are generally considered nutritionally undesirable, because they precipitate proteins, inhibit digestive enzymes, and have a detrimental effect on the absorption and utilization of vitamins and minerals. Tannin components have been implicated in incidences of cheek and esophageal cancers in specific regions of the world, and cited as antimutagenic and anticarcinogenic in other studies. Tannins are antimicrobial, inhibiting the invasion and growth of fungi, bacteria, and viruses in food plants. Tannins can theoretically serve as natural regulators of the microbial population in many habitats including the gastrointestinal tract. Tannins are also reported to produce other physiological effects such as immune response, hepatotoxicity, and lipid metabolism. Ingestion of large quantities of tannins may result in adverse health effects. However, the ingestion of small quantities of selected tannins may be beneficial to human health.

Polyphenols in herbal medicines, fruits, and vegetables are often presumed to play an important

physiological role in the maintenance of good health. Polyphenols generally exert their effects by associating with metals, their antioxidant and radical-scavenging activities, and interaction with other molecules, including proteins and polysaccharides. The antioxidant capacity of flavanoids is determined by a combination of the *O*-dihydroxy structure in the B-ring, the 2,3-double bond in conjugation with a 4-oxo function, and the presence of both hydroxyl groups in positions 3 and 5. Recent research efforts are highlighting the exciting and exceptionally dynamic role of polyphenolic components actions as antioxidants or agents of other mechanisms that contribute to their anticarcinogenic, antiviral, antiinflammatory, and cardioprotective effects derived from consumption of plant foods.

Sensory Properties

Flavonoids and anthocyanins contribute yellow, red, and blue colors to flowers, leaves, and fruits. Polyphenolic compounds also contribute sweet, bitter, or astringent flavors, depending on solubility and structure. Flavonoids such as rutin, a 3-rhamnosyl-D-glycosyl derivative of quercetin, contribute undesirable precipitates to processed foods. Tannins and tannin–protein complexes may give fruit juices a desirable body or consistency, or may contribute to undesirable hazing or loss of clarity or sparkle. (*See* **Sensory Evaluation**: Appearance; Taste.)

Pharmacological effects related to estrogen production and induction of infertility, enzyme cofactors beneficial to health, and counteracting of glucose resorption are attributed to consumption of specific polyphenolic compounds. Tannins are often related to astringency and bitterness in wines or foods, and are thought to play a significant role in the reduction of dietary protein digestibility by complexing with either dietary protein or digestive enzymes. (*See* **Coenzymes; Glucose**: Function and Metabolism.)

See also: **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography; **Coenzymes**; **Coffee**: Green Coffee; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Fruits of Temperate Climates**: Fruits of the Ericacae; **Glucose**: Function and Metabolism; **Phenolic Compounds**; **Ripening of Fruit**; **Sensory Evaluation**: Appearance; Taste; **Spectroscopy**: Nuclear Magnetic Resonance; **Tea**: Chemistry; **Vegetables of Temperate Climates**: Commercial and Dietary Importance

Further Reading

Benavente-Garcia O, Castillo J, Marin FR, Ortuno A, and Del Rio JA (1997) Uses and properties of *Citrus* flavonoids. *Journal of Agricultural and Food Chemistry* 45: 4505–4515.

Chung K-T, Wei C-I, and Johnson MG (1998) Are tannins a double-edged sword in biology and health? *Trends in Food Science and Technology* 9: 168–175.

Earp CF, Akingbala JO, Ring SH, and Rooney LW (1981) Evaluation of several methods to determine tannins in sorghums with varying kernel characteristics. *Cereal Chemistry* 58: 234.

Hagerman AE and Butler LG (1978) Protein precipitation method for the quantitative determination of tannins. *Journal of Agricultural and Food Chemistry* 26: 809.

Harbourne JB and Mabry TJ (eds) (1988) *The Flavonoids – Advances in Research*, 621pp. London: Chapman & Hall.

Haslam E (1982) Proanthocyanidins. In: Harborne JB and Mabry TJ (eds) *The Flavonoids – Advances in Research*, pp. 417–447. London: Chapman & Hall.

Haslam E (1989) *Plant Polyphenols*, 230pp. Cambridge: Cambridge University Press.

Haslam E (1998) *Practical Polyphenolics: From Structure to Molecular Recognition and Physiological Action*, 422 pp. Cambridge: Cambridge University Press.

Kapsalis C and Beck RA (1985) Phenolic Plant Pigments. In: Kapsalis C and Beak RA (eds) *Food Chemistry and Nutritional Biochemistry*, pp. 554–563. New York: Wiley.

Lewis NG and Yamamoto E (1989) Tannins – their place in plant metabolism. In: Hemingway RW and Karchesy JJ (eds) *Chemistry and Significance of Condensed Tannins*, pp. 23–46. New York: Plenum.

Taste *See* **Sensory Evaluation**: Sensory Characteristics of Human Foods; Food Acceptability and Sensory Evaluation; Practical Considerations; Sensory Difference Testing; Sensory Rating and Scoring Methods; Descriptive Analysis; Appearance; Texture; Aroma; Taste

TASTE ENHANCERS

S S Schiffman, Duke University Medical School, Durham, NC, USA

Introduction

The term 'taste enhancer' is used in the food industry to describe a substance that enhances the sensations of food (or food ingredients) when introduced into the mouth. The use of the term 'taste' is colloquial and actually refers to flavor (both taste and smell) because chemicals from food activate receptors in the nose as well as the mouth. Enhancement of the taste and smell of food is desirable to improve palatability, increase total intensity, potentially reduce the cost of ingredients, and compensate for chemosensory (taste and smell) losses in vulnerable populations such as the elderly. Enhancement can be achieved in two ways: (1) by simply adding more molecules to the food or (2) by potentiating the intensity through synergism and/or alteration of receptor mechanisms without altering the total number of molecules.

Enhancement by Addition of Molecules

Many food ingredients, including monosodium glutamate (MSG), NaCl, and sweeteners have been termed 'taste enhancers' but their main effect is simply to add more molecules that generate additional taste or smell sensations. Tastants such as MSG, salt, and sweeteners don't actually boost other chemosensory properties but rather contribute additional meaty/savory, salty, or sweet properties respectively.

Experimental studies have shown that MSG at concentrations up to $0.005\,\mathrm{mol\,l^{-1}}$ does not alter the intensities of other food ingredients including salts, sweeteners, amino acids, acids, or bitter compounds. MSG is a taste enhancer strictly from the standpoint that it adds another taste quality to the food (called umami in Japanese) which improves palatability rather than altering the intensity of other ingredients. Similar conclusions pertain to the 'enhancing' effects of NaCl and sweeteners that add saltiness or sweetness to food, respectively; they also improve palatability by reducing bitter components of some food substances. Thus MSG, salt, and sweeteners are taste enhancers from the standpoint that they add additional tastes to the food and improve palatability rather than potentiate the taste intensity of other ingredients.

Analogous reasoning can be applied to flavors. Commercial flavors are often added to food for the elderly (called 'flavor enhancement') to compensate for losses in smell perception that occur in elderly individuals. Flavors consist predominantly of mixtures of odorous molecules that can be extracted directly from natural foods or can be synthesized after chromatographic and mass spectrographic analysis of natural products. Simulated flavors are flavor enhancers only from the standpoint that they increase the total number of molecules that interact with receptors on chemosensory membranes in the nose and mouth. Intensification of odor can improve palatability, induce more salivation, produce greater stimulation of the olfactory and limbic system of the brain, and promote immune function via mechanisms described below under positive health benefits of chemosensory enhancement.

Although odorants and tastants are perceived by two different senses, the perceived intensity of a tastant may also increase with the addition of odorant molecules. For example, perceived sweetness intensity is enhanced by the addition of odorants, especially for congruent taste/odor mixtures such as sucrose/strawberry and sucrose/lemon. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Enhancement by Synergism

Taste enhancement also occurs via synergism in which the total intensity of a mixture is greater than the expected intensity based on the concentration of the individual components. Two methods of calculating the degree of synergy have been used: one which simply compares the perceived intensity of a binary mixture (for example) with the sum of the intensities of the unmixed components, and a second which compares the intensity of binary blends to the average of the two pure components of the blends. The most common practical application of synergism occurs with the use of certain binary or ternary sweetener combinations in beverages. That is, when one sweetener is combined with certain other sweetener(s), it produces a synergistic sweetening effect. Synergism permits the blending of low concentrations of sweeteners such as saccharin or acesulfame-K, that are bitter at higher concentrations, with sweeteners without salient bitter components such as aspartame and fructose to achieve a desired level of sweetness. Synergistic effects are idiosyncratic and vary with the chemical structure of the individual ingredients. Synergistic effects can occur between two or more

high-potency sweeteners (such as aspartame and acesulfame-K), between a high-potency sweetener and sugars (such as acesulfame-K and fructose), and between combinations of sugars such as fructose and sucrose. The maximum synergy found for binary mixtures of sweeteners is about 74%; the maximum synergy found for ternary mixtures of sweeteners is as much as 99%. The precise mechanisms that produce synergism among sweeteners are not known at the present time. However, it is probable that multiple receptors as well as biochemical transduction mechanisms in taste cells play a role.

Enhancement by Modifying Receptor Mechanisms

A variety of compounds have been shown in a laboratory setting to enhance taste perception including: (1) caffeine (and other methyl xanthines); (2) 5′ ribonucleotides; (3) inosine; and (4) bretylium tosylate. However, only 5' ribonucleotides are commercially practical enhancers because the other compounds require pretreatment of the tongue to achieve potentiation.

Methyl Xanthines

Pretreatment (adaptation) of the tongue with methyl xanthines has been shown to enhance other taste stimuli. Methyl xanthines, including caffeine, theophylline, and theobromine, are found in coffee, tea, and chocolate, respectively. When the tongue was adapted to caffeine, theophylline, or theobromine at concentrations ranging from 10 μmol l^{-1} to 10 mmol l^{-1}, the taste intensities of acesulfame-K (a sweetener), sodium chloride, potassium chloride, and quinine hydrochloride were enhanced. Furthermore, 10 μmol l^{-1} caffeine potentiated the taste of moderate concentrations of other sweeteners, including neohesperidin dihydrochalcone, D-tryptophan, thaumatin, stevioside, and sodium saccharin. Sweeteners potentiated by caffeine had a bitter component; sweeteners without prominent bitter components, including aspartame, sucrose, fructose, and calcium cyclamate, were not potentiated by caffeine. It has been hypothesized that this enhancement by pretreatment with methyl xanthines may be caused by the antagonistic effect of methyl xanthines on adenosine receptors on the tongue surface. (*See* **Alkaloids**: Properties and Determination; **Caffeine**.)

Inosine-5′-Monophosphate (IMP)

The 5′-ribonucleotides, inosine-5′-monophosphate (IMP) and guanosine-5′-monophosphate (GMP), have been shown to enhance certain tastes. A mixture of 5′-ribonucleotides with MSG can potentiate the taste of suprathreshold concentrations of MSG. The nucleotides with the best enhancement capacity appear to be those having a purine nucleus with a hydroxy group in the 6-position and a ribose moiety esterified in the 5′-position with phosphoric acid, e.g., IMP and GMP. Addition of 0.1 and 1 mmol l^{-1} IMP to MSG also reduces the MSG taste threshold so that it can be perceived at lower concentrations (fewer molecules are needed for detection). Pretreatment of the tongue with 1 mmol l^{-1} IMP enhanced the taste of both sucrose and aspartame by approximately 60 and 40%, respectively. Biochemical studies indicate that the synergism between MSG and certain 5′-ribonucleotides is a peripheral event involving alterations in glutamate receptors on the taste bud. Inosine, a breakdown product of both IMP and of adenosine, has also been found to enhance moderate tastes of sucrose, aspartame, and sodium chloride when the tongue is pretreated with inosine.

Bretylium Tosylate

Pretreatment of the tongue surface with bretylium tosylate, a quaternary ammonium compound that is used as an antifibrillary drug, enhanced the taste of sodium chloride. Bretylium tosylate has been shown to open amiloride-sensitive sodium channels in other biological systems. Bretylium tosylate may act in a similar fashion on the human tongue by increasing sodium transport through the amiloride-sensitive channels on the tongue that are known to be involved with salty taste.

Positive Health Benefits of Chemosensory Enhancement

Enhancement of the sensory properties of meats, soups, vegetables, and other nutritious foods with food flavors and MSG not only compensates for taste and smell losses that occur in an elderly population but can also improve health status. Four clinical studies of frail elderly described here have shown that amplification of the flavor levels of foods to preferred levels is associated with increased total numbers of lymphocytes (including T cells and B cells), increased secretion rate of salivary immunoglobin A (IgA), and improved functional status. Furthermore, flavor enhancement resulted in improved immunity and functional status even when macro- and micronutrient intakes were uneffected.

Study 1: Flavor Enhancement Increases T- and B-Cell Levels in the Elderly

Flavor enhancement of table food for independent living residents at a retirement home has been

shown to improve immune status as determined by T- and B-cell levels as well as grip strength. In one study, 39 elderly residents were divided into two groups. Group 1 received food that was unenhanced by flavor for the first 3 weeks, and food that was enhanced by flavor for the second 3-week period. For group 2, the order was reversed; they received enhanced food for the first 3-week period and unenhanced food for the second 3-week period. The menu plan during the 3 weeks of flavor enhancement was identical to the menu plan during the unenhanced 3-week period. During the 3 weeks of flavor enhancement, six different flavors were added to selected nutrient-dense foods: roast beef, ham, natural bacon, prime beef, maple, and cheese. Flavors were added to some, but not all, foods at a meal in the flavor-enhanced condition. Analysis of the data indicated that the elderly subjects consumed the same macro- and micronutrients on the two arms of the study. This occurred because not all foods at a meal were enhanced during the 3-week experimental period of flavor enhancement; subjects simply ate less of foods that were not enhanced. This study was repeated with 4-week (rather than 3-week) food rotations in which MSG as well as flavors were added on the 4-week enhancement arm of the study. In both studies there was an increase in T- and B-cell levels and improved grip strength after flavor enhancement, even though macro- and micronutrient intakes were not affected.

Study 2: Flavor Enhancement Improves Secretion Rate of Salivary IgA

Taste and odor stimuli have also been shown to increase the secretion rate of IgA in the saliva of both young and elderly individuals. IgA is an antibody that protects mucosal surfaces. Flavors can either be dropped on the tongue from an eye dropper or delivered in food. The main finding of such studies is that chemosensory stimulation improves mucosal immunity in two ways: (1) by increasing saliva production and (2) by increasing the absolute concentrations of secretory IgA in the saliva. These results have important implications for the elderly, who often suffer from dry mouth and reduced salivary flow (and hence reduced mucosal immunity) due to normal aging, diseases, and medications they are taking.

Study 3: Sensory Enhancement of Foods Increases Intake in Sick Elderly

Enhancement of hospital food with a combination of flavors and MSG has been shown to improve intake in over 90% of patients. In addition, when sensory enhancement was performed for a week or more, there was improvement in plasma protein levels (including somatomedin-C/insulin-like growth factor I, albumin, and transferrin) and T lymphocytes in some patients.

Study 4: Flavor Enhancement of the Entrée at Dinner can Reduce Sodium Intake by 500 mg

Enhancement of an unsalted entrée with salt-free flavors can reduce the sodium levels in a meal without compromising ratings of satisfaction. This finding was obtained in a study at a retirement home performed over an 8-week period during which salt shakers were removed from the tables. During the study, two entrées (beef steak and chicken breast) were each served once a week. For the first 2 weeks (control period), the entrées (beef or chicken) were salted with 500 mg sodium (the minimum preferred level of table salt for this population). For the subsequent 6 weeks (experimental period), the flavor of the entrées was enhanced by marination in sodium-free beef or chicken flavor prior to cooking; no table salt was added to the beef or chicken. Two vegetables which were lightly salted in the kitchen accompanied the entrée. During the 6 weeks of flavor enhancement, the sodium content of the meal was reduced by 500 mg. Throughout the study, residents were asked to rate their satisfaction with the sensory properties of each meal after eating. Analysis of the data indicated no significant difference in the degree of satisfaction between meals with the salted version of the entrée and the flavor-enhanced (sodium-free) version. This suggests that enhancement of odor can replace salt in an entrée with no adverse affects on acceptability when two lightly salted vegetables accompany the meal. Thus, providing more sensory input in the form of odor can reduce the need for taste stimulation by salt, and could make it easier to comply with recommended daily intake of sodium (3000 mg or less).

Conclusion

Enhancement of the taste and smell of food can be achieved in two ways: (1) by adding more molecules to the food or (2) by potentiating the intensity through synergism and/or alteration of receptor mechanisms without increasing the total number of molecules. Taste (and smell) enhancement can improve palatability, increase total intensity, potentially reduce the cost of ingredients, compensate for chemosensory (taste and smell) losses in vulnerable populations such as the elderly, and improve a number of nutritional and health parameters.

See also: **Acesulfame/Acesulphame**; **Aspartame**; **Sensory Evaluation**: Aroma; Taste; **Sodium**: Properties and Determination; Physiology

Further Reading

Bellisle F, Monneuse MO, Chabert M *et al*. (1991) Monosodium glutamate as a palatability enhancer in the European diet. *Physiology and Behavior* 49(5): 869–873.

Schifferstein HN and Verlegh PW (1996) The role of congruency and pleasantness in odor-induced taste enhancement. *Acta Psychologica* 94(1): 87–105.

Schiffman SS (1987) Recent developments in taste enhancement. *Food Technology* 41(6): 72–73,124.

Schiffman SS (1997) Taste and smell losses in normal aging and disease. *Journal of the American Medical Association* 278: 1357–1362.

Schiffman SS (1998) Sensory enhancement of foods for the elderly with monosodium glutamate and flavors. *Food Reviews International* 14: 321–333.

Schiffman SS (2000) Intensification of sensory properties of foods for the elderly. *Journal of Nutrition* 130(4): 927–930S.

Schiffman SS and Miletic ID (1999) Effect of taste and smell on secretion rate of salivary IgA in elderly and young persons. *Journal of Nutrition, Health, and Aging* 3: 158–164.

Schiffman SS and Warwick ZS (1993) Effect of flavor enhancement of foods for the elderly on nutritional status: food intake, biochemical indices and anthropometric measures. *Physiology and Behavior* 53: 395–402.

Schiffman SS, Frey AE, Luboski JA, Foster MA and Erickson RP (1991) Taste of glutamate salts in young and elderly subjects: role of inosine 5′-monophosphate and ions. *Physiology and Behavior* 49: 843–854.

Schiffman SS, Booth BJ, Carr BT *et al*. (1995) Investigation of synergism in binary mixtures of sweeteners. *Brain Research Bulletin* 38: 105–120.

Schiffman SS, Sattely-Miller EA, Graham BG, Booth BJ and Gibes KM (2000) Synergism among ternary mixtures of 14 sweeteners. *Chemical Senses* 25: 131–140.

Yamaguchi S and Kimizuka A (1979) Psychometric studies on the taste of monosodium glutamate. In: Filer LJ, Garattini S, Kare MR, Reynolds WA and Wurtman RJ *Glutamic Acid: Advances in Biochemistry and Physiology*, pp. 35–54. New York: Raven Press.

TEA

Contents

Types, Production, and Trade

S Taylor, Natural Resources Institute, Chatham Maritime, Kent, UK

Introduction

Tea is one of the most popular beverages in the world, providing a valuable source of income to many producer countries. There are a number of factors accounting for the popularity of tea, including its sensory properties, relatively low retail price, and apparent health benefits. This series of articles on tea presents information on the types, trade, agronomy, processing, chemistry, analysis, and quality of tea.

Origin and Distribution

The probable center of origin of tea is in south-east China near the source of the Irrawaddy river. From there it spread to the southern portion of China, parts of India, Burma, Thailand, Laos, and Vietnam. From these main centers in South-east Asia, tea has spread into many tropical and subtropical countries. It is thought that tea was introduced into Japan around AD 600. In the early part of the nineteenth century, unsuccessful attempts were made to establish Chinese tea in India. Only when the native 'wild' tea plants found in Assam were used did tea production in India become successful. The main areas of production in India are the north-east and later the south. Tea was introduced into Indonesia as early as 1690, but commercial production began much later (1878), with the cultivation of Assam varieties. Extensive

plantings were started in Ceylon (now Sri Lanka) during the 1870s. The cultivation of tea has since spread to other Asian countries, such as Turkey, Iran, Taiwan, Bangladesh, and Malaysia. In the late nineteenth century, cultivation of tea spread beyond Asia to Africa and South America. Tea was first introduced into Malawi in 1886 and into East Africa at the beginning of the twentieth century. Production of tea in the former Soviet Union began on a large scale in 1892. Today, tea is grown in over 30 countries. The most northerly region in which tea is grown today is Georgia, in south Russia, and the southernmost is South Africa, and the Argentine in South America.

Classification

The tea plant, *Camellia sinensis* (L.) O. Kuntze, is the only important economic species of the family Theaceae. Two botanical varieties have been recognized.

China Teas (*Camellia sinensis* var. *sinensis*)

These teas are produced from slow-growing, dwarf trees (which grow to a height of 4–6 m if unattended), with small, erect, comparatively narrow, markedly serrated, dark green leaves which are smooth with a matt surface; the flowers are borne singly. The plant is relatively resistant to cold and, as a result, the variety is found in the more temperate producing regions such as China, Japan, the former Soviet Union, Turkey, Iran, and the northern, higher-altitude growing areas of India. It is known to produce delicately flavored tea when grown at high altitudes; it is, however, low-yielding, especially at very high altitudes, such as Darjeeling.

Assam Teas (*Camellia sinensis* var. *assamica*)

This variety is faster-growing and taller (12–15 m unattended) than the China variety with larger glossy elliptical leaves; flowers are found in clusters of 2–4. This variety is less resistant to cold than the China type and, as such, can only survive at high altitudes near the Equator. In general, it is a much higher-yielding plant than the China type, and produces a less delicately flavored beverage.

Numerous hybrids between China and Assam types are known. There is evidence that *C. irrawadiensis* was hybridized with *C. sinensis* var. *assamica* to produce the uniquely flavored Darjeeling tea. As tea is largely cross-pollinated, and much of the commercial crop is raised from seed, the crop is very heterogeneous.

The term 'jat' is used to indicate seed derived from different districts or plantations or to separate types on the basis of foliar characteristics.

Types of Processed Tea

Fermented or Black Tea

Black teas form the major proportion of tea consumed in the western hemisphere. They are produced by full fermentation and roughly classified as plain or flavory.

Plain black teas are produced and sold on the basis of the taste characteristics associated with the phenolic substances produced during fermentation. These teas are mainly produced and exported from Africa (e.g., Kenya and Malawi) and India (Assam). High-quality plain teas are often known as breakfast teas. (*See* **Phenolic Compounds**; **Tea**: Chemistry.)

Flavory black teas are sold on the basis of their aroma characteristics and should not be confused with flavored teas, to which fragrances have been added. They are primarily manufactured in areas of relatively high altitude in Sri Lanka and India (Darjeeling). In Sri Lanka, the very flavory teas are produced during January and February in the Dimbula district, and during August and September in the Uva district. In these areas, the desired aroma is produced during cool, clear, dry windy weather with day and night temperatures of about 20 °C and 6–10 °C, respectively. These conditions must last for at least two consecutive weeks and even slight rain will set the process back to the beginning. The famed first- and second-flush Darjeeling flavor is produced under similar conditions when growth is slow and the plants are under climatic stress.

Green Tea

Green tea is entirely different from black tea in its preparation. There are no fermentation products and the leaf remains green. It is generally produced from var. *sinensis*. When infused, the liquors are greenish, pale primrose, or lemon-yellow in color with no trace of red or brown. Green tea is produced and consumed mainly in China and Japan. Some is also consumed in northern Africa. Often, green tea production is a cottage industry, not only in the countries mentioned above, but also in other countries such as Malaysia, Thailand, Vietnam, and Indonesia. This processing is often undertaken by experienced Chinese tea makers. (*See* **Tea**: Processing.)

There are many different kinds of green tea produced, differing in style and cup characters. Some varieties of green tea include the following:

- Sencha, the most widely drunk grade of green tea in Japan.
- Kamaira-cha, made from young leaf, but differing from sencha in that the first process is parching and not steaming.

- Bancha, made from coarse leaf and stalk.
- Gyokuro, one of the finest green teas; it is grown under complete shade, hand-plucked and hand-sorted. It has a twisted dark green leaf.
- Matcha, a powdered tea also made from shade-grown leaf.
- Tencha, the ceremonial green tea. The beverage is a suspension of finely ground leaf.
- Gunpowder, made in the form of small compact pellets of Chinese green tea.
- Pilo chun, a small-leaved Chinese tea made from the first spring flush.

Partially Fermented Tea

Typical partially fermented teas include oolong and pouchong teas. These are partially oxidized so that their appearance is somewhat intermediate between that of green and black tea. Pouchong teas have a shorter fermentation time. They are manufactured primarily in China and Taiwan. Clones selected for oolong and pouchong tea production are characterized by the ability to impart a strong flowery aroma to the product. (*See* **Tea**: Processing.)

Instant Tea

Instant tea is the water-soluble extract of tea leaf, usually marketed as a powder, flake, or granule, either pure or as a part of flavored mixes. Iced lemon teas are the most popular examples of instant teas, particularly in the USA. Most instant tea is made from black tea, but some is made from green tea. It is manufactured in a number of tea-growing countries, e.g., Kenya and China, where the starting material may be fermented leaf that has not been dried.

Flavored Teas

Teas are sometimes scented with various plant essential oils such as lemon, bergamot, rose, and fragrant olive which impart sweet floral attributes to enhance the natural flavor of the tea. Other teas are blended with flower petals, spices, or dried leaf such as chrysanthemum, rosemary, camomile, and peppermint. Floral flavored teas are mainly consumed in the USA while herbal teas are popular in continental Europe. Many of these products are sold through health food stores and may not contain *C. sinensis*: the word 'tea' refers to their method of preparation. These products are often promoted on health grounds, particularly if they are caffeine-free. (*See* **Essential Oils**: Properties and Uses.)

Decaffeinated Tea

Decaffeinated tea has been popular in the USA since its introduction in 1983; however, in the UK the picture has been slightly different. Until the late 1980s it was primarily retailed through health food outlets; however more recently a number of specialist tea companies have launched decaffeinated tea bags into the mainstream grocery trade with considerable success.

Caffeine is removed from tea by treating it with one of three organic solvents: methylene chloride, ethyl acetate, or supercritical carbon dioxide. The caffeine is dissolved in the solvent and the tea is then treated to remove most of the solvent. (*See* **Coffee**: Decaffeination.)

Organically Grown Tea

Organically grown teas from Tanzania first entered the UK market in 1989 under the Natureland brand name from the London Herb & Spice Company. The tea bags were sold at a 20% premium price over other speciality teas and at more than double the price of regular tea bags. Since this time other organically grown tea products have appeared on the market. The major producers are India and Sri Lanka.

The attraction to the grower is substantially increased price for the crop and the saving of hard currency on chemical fertilizers and pesticides. It is believed that the average transition period from conventional to organic culture is 3 years, allowing time for chemicals to disperse.

World Production

According to statistics compiled by the International Tea Committee, tea production over the past decade (1988–98) increased from 2.48 million metric tons (mt) to 2.96 million mt. This represents an average annual rate of increase of 1.9%. The major producers are listed in Figure 1 along with their average production for the 1995–97 period and for 1998, the last year for which figures are currently available. The 3-year averages have been included to reveal the trend of changes in production through the smoothing of year-to-year fluctuations.

India is currently ranked as the world's largest producer of tea, followed by China, Kenya, and Sri Lanka. Kenya in particular had a record production for 1998 that represents an increase of over 33.3% from the 1997 drought-affected crop. Increases in both yields and planted area are likely to continue to support strong growth in tea production in African countries where the Kenyan output is expected to increase at an average annual rate of 2.8%. Malawi, Tanzania, and Zimbabwe are also expected

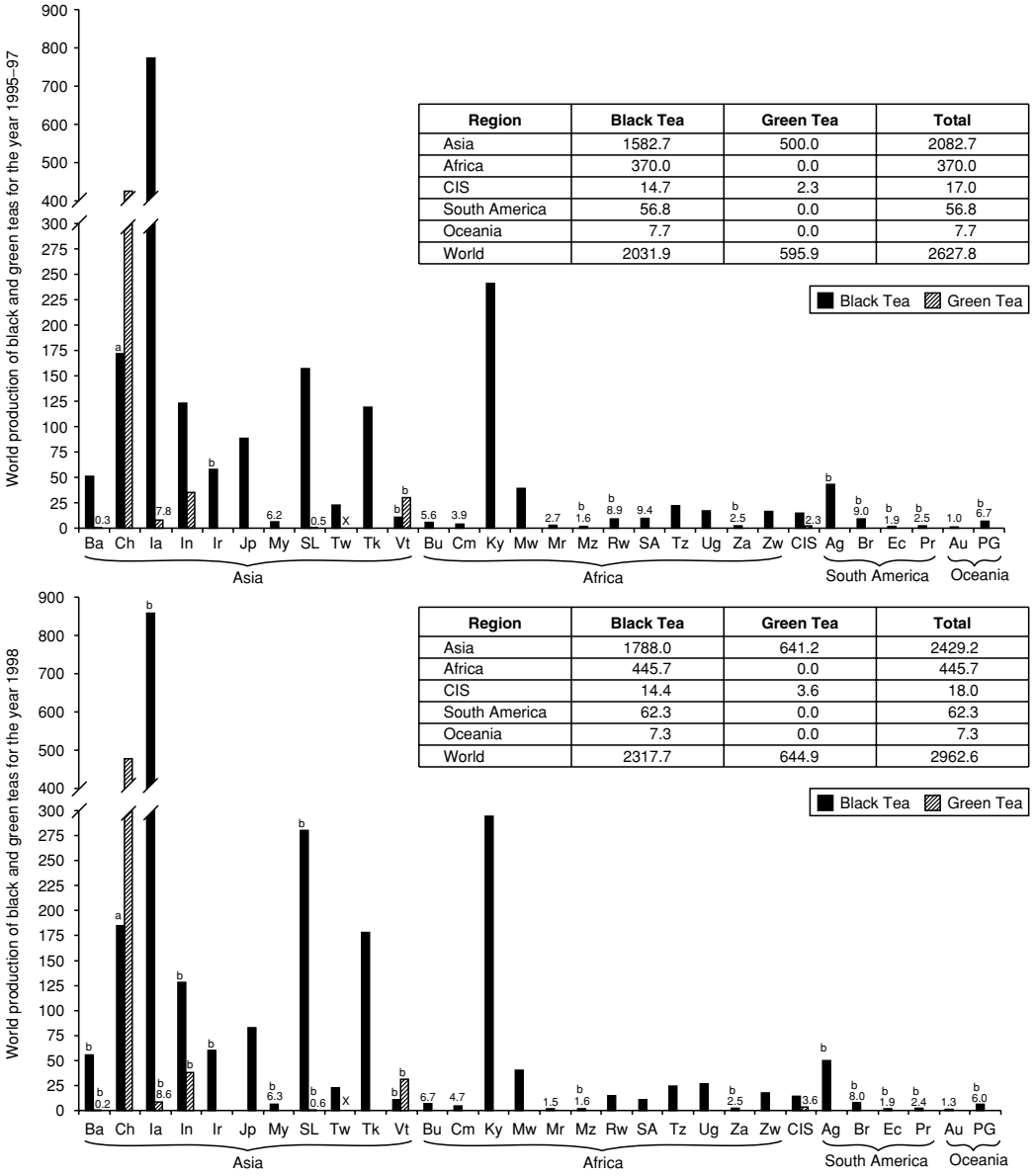

Figure 1 World production of black and green teas for the 1995–97 period and 1998 (× 1000 Mt). [a]Black and other teas, but not green teas. [b]Figures shown are provisional or estimated. CIS: Commonwealth of independent states (former Union of Soviet Socialist Republics), Ba: Bangladesh, Ch: China, Ia: India, In: Indonesia, Ir: Iran, Jp: Japan, My: Malaysia, SL: Sri Lanka, Tw: Taiwan, Tk: Turkey, Vt: Vietnam, Bu: Burundi, Cm: Cameroon, Ky: Kenya, Mw: Malawi, Mr: Mauritius, Mz: Mozambique, Rw: Rwanda, SA: South Africa, Tz: Tanzania, Ug: Uganda, Za: Zaire, Zw: Zimbabwe, Ag: Argentina, Br: Brazil, Ec: Ecuador, Pr: Peru, Au: Australia, PG: Papua New Guinea. Data from the Annual Bulletin of Statistics (2000).

to increase production significantly. The Intergovernmental Group (IGG) on Tea of the Food and Agriculture Organization (FAO) of the United Nations predicts that production in India will reach 1.02 mt in 2005. Economic reforms and the national plan to expand tea production in Sri Lanka could boost production to 285 000 mt, compared to 157 800 mt during the 1995–97 period. Other major tea-producing countries, including China and Indonesia, should also see significant growth in production, while in Bangladesh production is forecast to grow less rapidly.

Production expansion programs initiated by major exporting countries have contributed to previous increases in output of black tea in recent years, and the impact of these programs, particularly as bushes reach optimum production age, may continue.

The total tea area under production has increased from 1.4 million hectares in 1961 to 2.3 million

hectares in 1996. The average yield per hectare has increased by over 50% during the same period. However, it is questionable whether any large-scale increases will ensue from now on.

World Trade

Tea Retained in Producing Countries

In 1998, over 57% of the world's production of tea was retained in producer countries. In some countries it was much higher: Over 76% of the Indian and 74% of the Chinese production is retained for internal consumption. For India, in particular, the home market has increased greatly from 499 053 mt (1988) to 667 565 mt (1998). For a number of other countries, e.g., Iran, Japan, Malaysia, South Africa, Taiwan, and the Russian Federation, the majority, if not all, of their tea produced is consumed internally. African producer countries generally export the majority of their production, whereas most Asian producers do not.

Exports

World exports of tea, both black and green, for the 1995–97 period and for 1998 are reported in Figure 2. All major exporting countries maintained significant export volumes for 1998. The FAO projects that net export availabilities will reach 1.292 million mt in 2005 and predicts that China, India, Indonesia, Sri Lanka, and Kenya will account for 78% of this increased volume. Bangladesh, Malawi, Tanzania, Turkey, and Zimbabwe are also expected to increase their export capabilities significantly by 2005. Exports from India, the world's largest tea-producing and -consuming country, are expected to recover from the recent disappointing performance due to weather, while satisfying growing domestic demand.

The largest and most significant exporter of green tea is China. Major destinations for green tea include Morocco, the Russian Federation, Algeria, the USA, and Afghanistan. India exports small amounts of green tea to the USA, Afghanistan, and Japan. China also produces a number of other teas, e.g., partially fermented, and the major export destinations for these are Hong Kong and Japan. Taiwan produces and exports oolong, pouchong, and green tea, with Japan being the largest importer of these. The Japanese, in turn, export a small proportion of their green teas, primarily to the USA.

Imports

Imports for Consumption Figure 3 shows tea imports for consumption of major importing countries for the 1995–97 period and for 1998. The UK and the Confederation of Independent States (CIS) countries – particularly the Russian Federation – were the largest importers of tea for internal consumption. Other major destinations for tea include Pakistan, the USA, Egypt, Japan, Iraq, and Morocco. Growth in developed countries is expected to be largely concentrated in the CIS countries, the USA, and, to a lesser extent, western Europe. Consumption in the UK has been declining since 1961 when consumption per head was 4.5 kg, and reached its lowest point of 2.4 kg in 1995. The FAO reports that the apparent consumption per head has since stabilized and rose to 2.49 kg in 1996 and 2.58 kg in 1997. This is reflected in consecutive increases in net imports for these years.

Reexports of tea A proportion of the tea imported into the UK is reexported. The largest recipients of this reexported tea are Canada, Germany, Saudi Arabia, and France. The Netherlands also reexports a high proportion of its imports.

Marketing

Most of the world's black tea production is traded in either auction or forward markets. The former accounts for the largest proportion of the total tea traded. In 1997, 84% of the total tea traded was sold through auctions held in producer countries, namely: Mombasa (Kenya), Guwahati, (India), Siliguri (India), Calcutta (India), Cochin (India), Coimbatore (India), Chittagong (Bangladesh), Colombo (Sri Lanka), Jakarta (Indonesia), and Limbe (Malawi). The auctions take place weekly through the growing season.

Tea sold through forward contracts is gaining popularity. A significant advantage of this market over the auction market is that the participants can lock on prices they wish to trade in the future. In the forward market, a buyer and a seller buy/sell an agreed amount of goods for delivery on an agreed future date at a predetermined price. Prices in this market tend to be less transparent than auction prices (spot or cash markets).

Tea Consumption

According to the FAO, black tea consumption will increase from 1.97 million mt in 1993–95 to 2.67 million Mt by 2005 – an annual growth rate of 2.8%. Statistics collated by the International Tea Committee indicate that this is an underestimation, as the total world production for 1998 exceeds the FAO estimate for 2005. Developing countries are forecast to account for the largest part of the

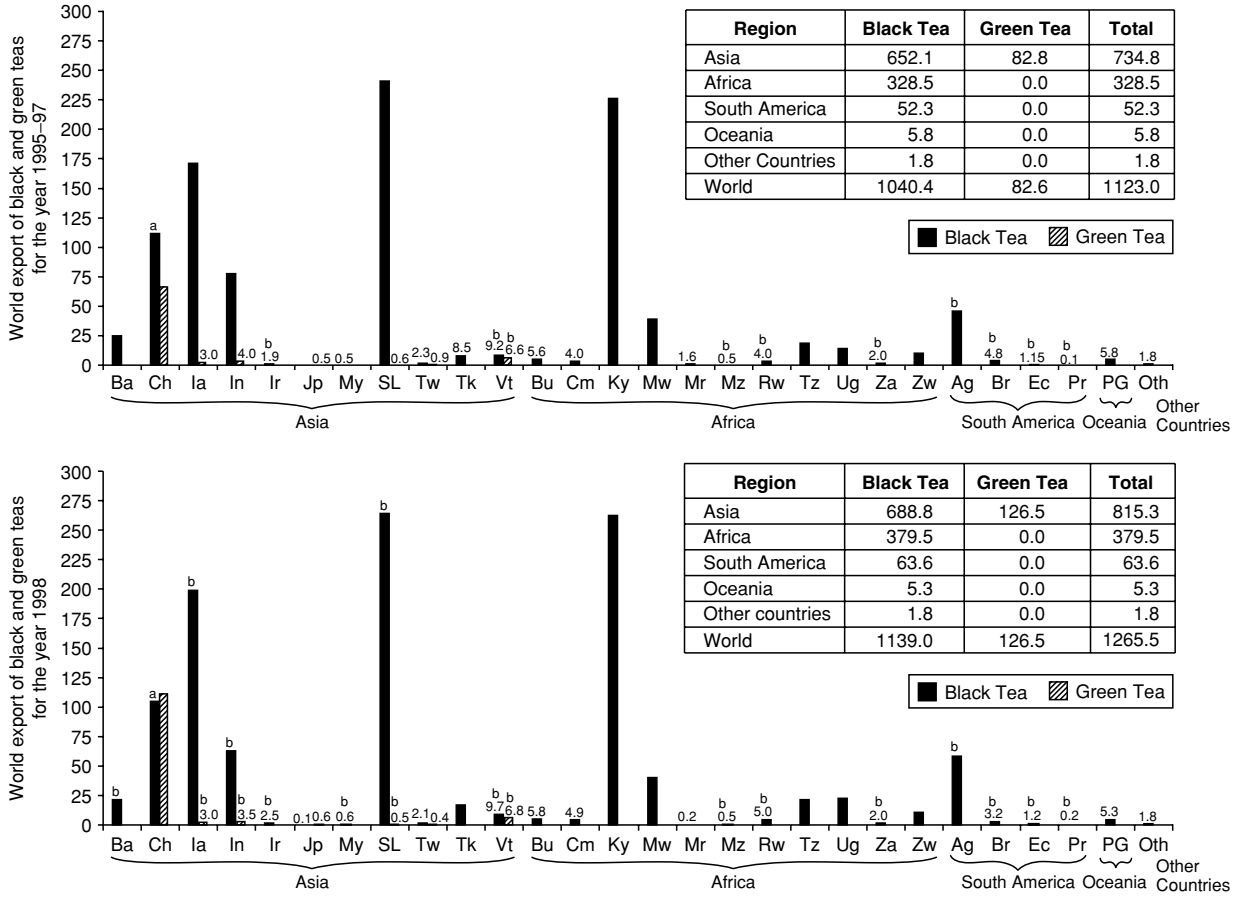

Figure 2 World export of tea (black and green) for the 1995–97 period and 1998 (× 1000 Mt). [a]Black and other teas, but not green teas. [b]Figures shown are provisional or estimated. Ba: Bangladesh, Ch: China, Ia: India, In: Indonesia, Ir: Iran, Jp: Japan, My: Malaysia, SL: Sri Lanka, Tw: Taiwan, Tk: Turkey, Vt: Vietnam, Bu: Burundi, Cm: Cameroon, Ky: Kenya, Mw: Malawi, Mr: Mauritius, Mz: Mozambique, Rw: Rwanda, Tz: Tanzania, Ug: Uganda, Za: Zaire, Zw: Zimbabwe, Ag: Argentina, Br: Brazil, Ec: Ecuador, Pr: Peru, PG: Papua New Guinea. Data from the Annual Bulletin of Statistics (2000).

prospective increase, with consumption rising from the 1993–95 average of 1.41 million mt to 1.95 million Mt by 2005 – an annual growth rate of 3%.

The FAO reports that black tea consumption in India will rise rapidly, reaching 832 000 Mt by 2005, annual growth of 3.2% from the 1993–95 average. In other major markets for black teas, such as Pakistan, the Islamic Republic of Iran, and Egypt, consumption is projected at 160 000 Mt, 122 000 Mt, and 90 000 mt by 2005, respectively. The reduction of import tariffs and declining prices could have a more pronounced effect on consumption in these countries.

Consumption in the European Community is projected to increase only slightly in the next decade since higher purchases by France, Germany, Italy, and the Netherlands would be largely counterbalanced by the decline in consumption in the UK that began in 1961. Consumption in the USA is projected to increase, though at a relatively slow rate of less than 1%. In the countries of the former Soviet Union, black tea consumption is projected to increase from 154 000 mt in 1993–95 to 250 000 mt in 2005, equivalent to an annual growth rate of 4.5%.

Trends in World Supply, Demand, and Prices

The FAO projections indicate that, by 2005, world black tea production and consumption could almost be in balance. Production gains are expected to come largely from higher yields, while the rise in consumption is from population and income growth. Developing countries are expected to account for most of the growth and their share in world consumption is expected to rise by 2% 2005. However, much will depend on the economic development of these countries.

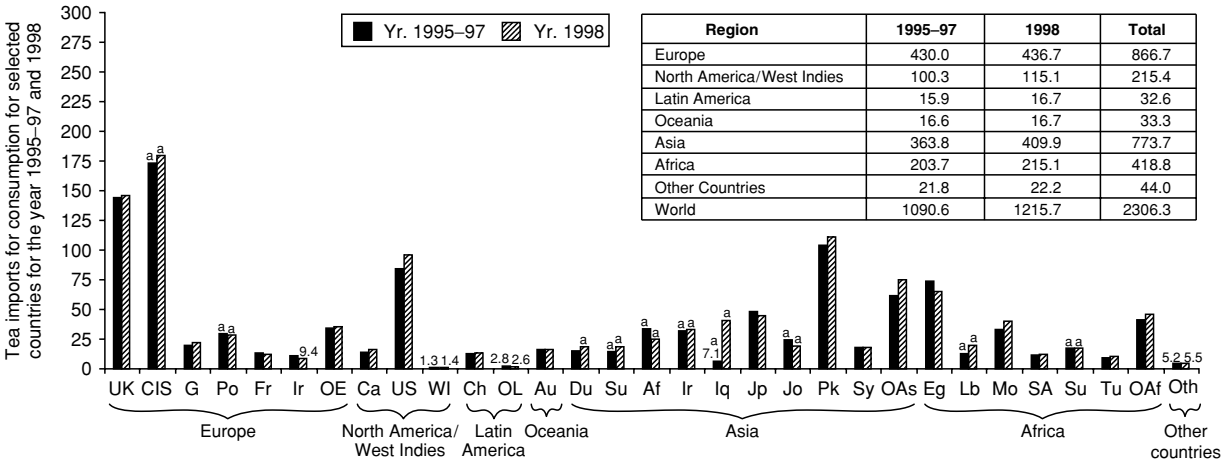

Region	1995–97	1998	Total
Europe	430.0	436.7	866.7
North America/West Indies	100.3	115.1	215.4
Latin America	15.9	16.7	32.6
Oceania	16.6	16.7	33.3
Asia	363.8	409.9	773.7
Africa	203.7	215.1	418.8
Other Countries	21.8	22.2	44.0
World	1090.6	1215.7	2306.3

Figure 3 Tea imports for consumption for selected countries for the 1995–97 period and 1998 (× 1000 Mt) (imports adjusted for re-exports). [a]Figures shown are provisional or estimated. UK: United Kingdom, CIS: Commonwealth of independent states (former Union of Soviet Socialist Republics), G: Germany, Po: Poland, Fr: France, Ir: Ireland, OE: Other European countries, Ca: Canada, US: United States of America, WI: West Indies, Ch: Chile, OL: Other Latin American countries, Au: Australia, Du: Dubai, Su: Saudi Arabia, Af: Afghanistan, Ir: Iran, Iq: Iraq, Jp: Japan, Jo: Jordan, Pk: Pakistan, Sy: Syria, OAs: Other Asian countries, Eg: Egypt, Lb: Libya, Mo: Morrocco, SA: South Africa, Su: Sudan, Tu: Tunisia, OAf: Other African countries. Data from the Annual Bulletin of Statistics (2000).

The FAO projections suggest an imbalance in the international market, with a surplus of export availabilities over import requirements, reaching 24 000 Mt by 2005 from an almost balanced market in 1993–95. This possible imbalance implies that world market prices would be under downward pressure if there were no additional increase in demand and/or downward adjustments in production. One of the ways to narrow the trade deficit and improve prices is to expand consumption through promotion. Over the last few years, the IGG on Tea has made major efforts to create awareness of the health benefits of tea drinking and to work out a generic promotion program that would stimulate consumption of tea.

The FAO reports that auction prices of black tea rose significantly during the first quarter of 1998; however they failed to maintain this vigor during the remainder of the year. Tea prices were boosted in 1997 and early 1998 by reports of drought-induced damage to the crops in Kenya and Indonesia, as well as strong import demand in the former Soviet Union. However, following the production recovery in Kenya, and to a lesser extent in Indonesia, coupled with a sharp drop in import demand from the Russian Federation due to deterioration of economic conditions, prices weakened significantly during 1998.

See also: **Coffee**: Decaffeination; **Essential Oils**: Properties and Uses; **Phenolic Compounds**; **Tea**: Chemistry; Processing

Further Reading

Anonymous (1999) *Committee on Commodity Problems. Intergovernmental Group on Tea. Thirteenth Session. Food and Agriculture Organization of the United Nations*. Ottawa, Canada, 27–29 September 1999.

Anonymous (2000) *Annual Bulletin of Statistics*. London: International Tea Committee.

Purseglove JW (1986) *Tropical Crops. Dicotyledons*. London: Longman.

Willson KC and Clifford MN (1992) *Tea*. London: Chapman & Hall.

Chemistry

P O Owuor, Tea Research Foundation of Kenya, Kericho, Kenya

Tea

Tea beverages are processed from the young tender shoots of *Camellia sinensis* (L) O. Kuntze. The plant biosynthesizes several chemicals during growth. **Table 1** summarizes the approximate chemical composition of dry tea leaves. The polyphenols, dominated by catechins (or flavan-3-ols) constitute up to 32% of the dry weight of young tender shoots. The tender tea plant shoots also contain both saturated and unsaturated fatty acids, either as glycosides or as

Table 1 Approximate chemical composition of young shoots of *Camellia sinensis* var. *Assamica*

Type	*Components*	*Dry weight (%)*
Water-soluble	Phenolic compounds	40
	Flavanols	18–30
	Epigallocatechin gallate	9–13
	Epigallocatechin	3–6
	Epicatechin gallate	3–6
	Epicatechin	1–3
	Gallocatechin	1–2
	Catechin	1–2
	Flavonol glycosides	3–4
	Proanthocyanidins	2–3
	Phenolic acids	4
	Caffeine	3–4
	Amino acids	4
	Theanine	2
	Others	2
	Carbohydrates	4
	Organic acids	0/5
Partially water-soluble	Starch	2–5
	Other polysaccharides	12
	Proteins	15
	Ash	5
Water-insoluble	Cellulose	7
	Lignin	6
	Lipids	4–9
	Pigments	0.5
	Volatiles	0.01–0.02

free acids. There are also methlyxanthines (mainly caffeine), amino acids and proteins, terpenes, and terpene glycosides, and many endogenous primary volatile flavor compounds. Although the biosynthetic pathways of these compounds are interesting, only the chemical and biochemical processes (after the young tender shoots have been plucked to make various kinds of tea beverages) are discussed here.

Of the many tea beverages, the most extensive chemical transformations occur during black tea processing. The article shall therefore reflect the chemistry of tea as represented by black tea manufacture, although, these chemical reactions partially occur to different degrees in the processing of other tea beverages.

Polyphenols

Polyphenols play a key role in the chemistry of the formation of the nonvolatile components of black tea. Although attempts to understand the chemistry of the formation of the nonvolatile black tea components started over half a century ago, to date, the chemistry is only partially understood. The young tender shoots of the tea plant used to process tea beverages contain high amounts of polyphenols, comprising flavanols (catechins), flavanol glycosides, leucoanthocyanins, and phenolic acids, etc.

The dominant polyphenols in green leaf are flavanol (Figure 1) comprising (+)-catechin (C), (+)-gallocatechin (GC), (−)-epicatechin (EC), (−)-epigallocatechin (EGC) and (−)-epigallocatechin gallate (EGCG). These compounds dominate the chemistry of nonvolatile compounds in tea. Several factors, including climate, genetic make-up, age of shoots and agronomic practices, cause variations in the total amounts and ratio of these compounds in young tea shoots. Although, recently, (−)-epicatechin-3,5-digallate and 3-methylgallates of (−)-epicatechin and epigallocatechin have been isolated in green tea leaves, their chemistry with respect to tea processing remains unknown. (+)-Catechin-3-gallate (CG) and (+)-gallocatechin-3-gallate (GCG) have also been isolated in manufactured teas, but not from fresh leaves. They are likely products of epimerization or racemization caused by firing (drying).

The other important component in the chemistry of nonvolatile components of tea is polyphenol oxidase. This is an *o*-dihidroxyphenolic oxygen reductase enzyme leading to the production of *o*-quinones. The enzyme is abundant in young tender shoots of green tea leaves, especially those cultivars suitable for making black teas.

The most noticeable chemical transformations/changes occur during the fermentation phase of black tea processing. Although this stage is called fermentation, in the true sense, the reactions that occur are those of oxidation. Following cell matrix destruction brought about by maceration, the catechins (flavanols) and other polyphenols undergo polyphenol oxidase initiated reactions, forming brown-colored products. The 1,2-dihydoxyphenols are oxidized to the quinones in the presence of oxygen, reacting further to form various brownish compounds. A simple catechin and a gallocatechin, for example, undergo oxidative reactions, which involve a loss of carbon dioxide to form a benzotropolone ring system. The compound formed is known as 'theaflavin.' Figure 2 outlines the formation of theaflavin.

From various combinations of the flavanols, it is possible to produce several theaflavins, as shown in Table 2. However, the dominant theaflavins found in black tea are theaflavin, theaflavin-3-gallate, theaflavin-3′-gallate and theaflavin-3,3′-digallate. The proportions of these four theaflavins in black tea vary with processing conditions and the composition of the individual precursor flavanols in green leaf.

Traces of epitheaflavic acid, theaflavin acid, and epitheaflavic acid-3′-gallate have also been found in black tea. Normally, there is very little gallic acid in green leaf. However, during fermentation, its levels

Table 2 Synthesis of theaflavins from pairs of flavanols

Parent flavanols		*Theaflavin*
Dihydroxy	*Trihydoxy*	
(−)-Epicatechin	(−)-Epigallocatechin	Theaflavin
(−)-Epicatechin	(−)-Epigallocatechin gallate	Theaflavin-3-gallate
(−)-Epicatechin gallate	(−)-Epigallocatechin	Theaflavin-3′-gallate
(−)-Epicatechin gallate	(−)-Epigallocatechin gallate	Theaflavin-3,3′-digallate
(−)-Epicatechin	(+)-Gallocatechin	Isotheaflavin
(+)-Catechin	(+)-Gallocatechin	Neotheaflavin
(−)-Epicatechin	(+)-Gallic acid	Epitheaflavic acid
(−)-Epicatechin gallate	(+)-Gallic acid	Epitheaflavic acid-3′-gallate
(+)-Catechin	(+)-Gallic acid	Theaflavic acid
(+)-Catechin gallate	(+)-Gallic acid	Theaflavic acid-3′-gallate

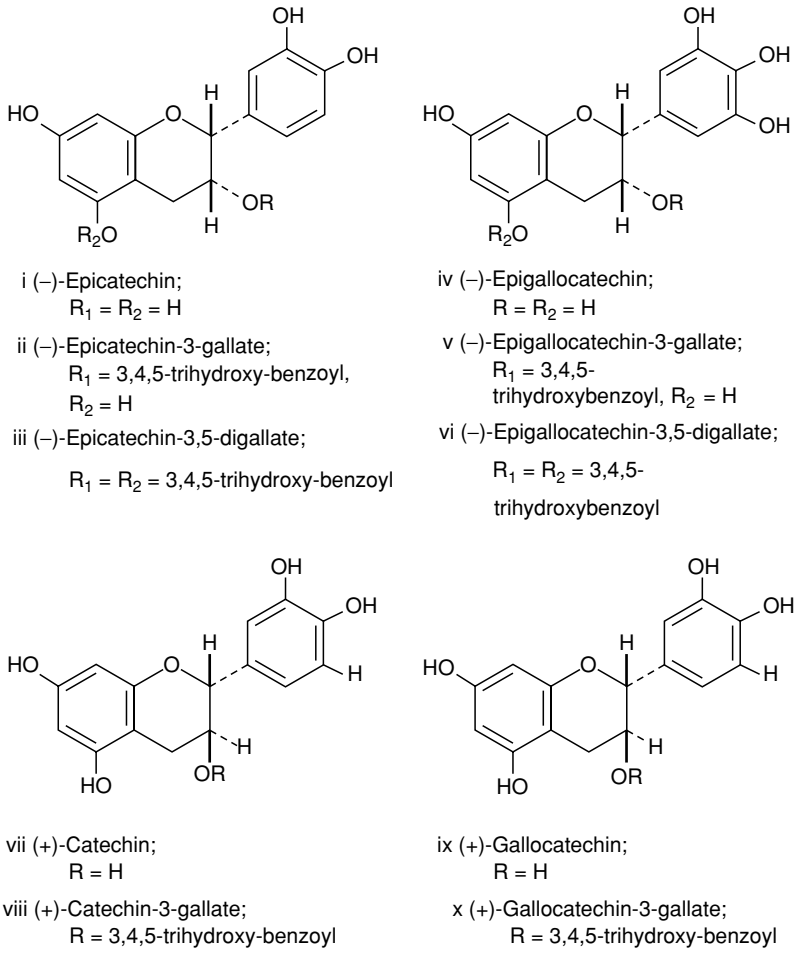

Figure 1 Structures of dominant flavanols in young tender shoots of tea plants.

increase, probably from the hydrolysis of galloyl esters from the catechins or gallated theaflavins. Although the rate of oxidation of gallic acid to the quinone acid is low, the presence of theaflavic acids is suspected to originate from gallic acid. Additional isomers of the theaflavins can also arise from racemization and epimerization of the flavanols. However, since racemization and/or epimerization of flavanols occur mainly during firing, very few isomers of theaflavins are normally found in black tea.

Theaflavins account for 0.3–1.8% dry weight of black tea and 1–6% of total solids in tea. They are

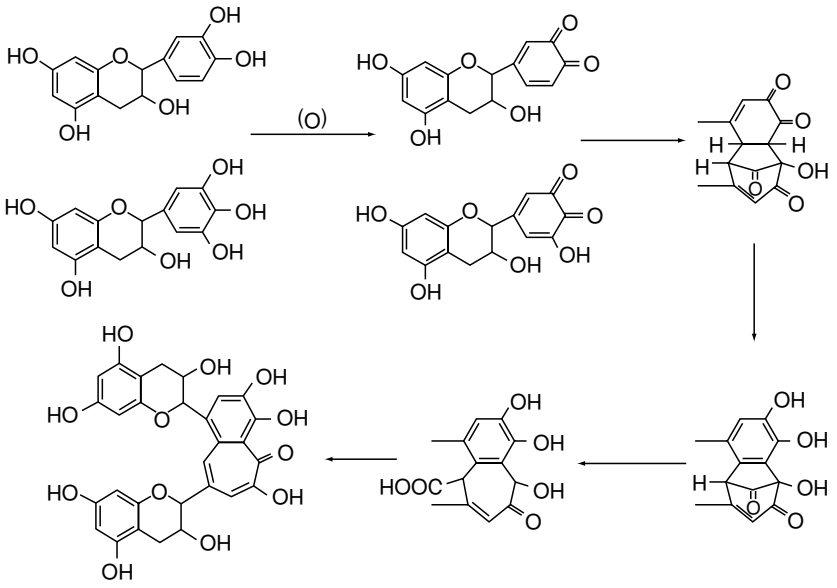

Figure 2 Mechanism of theaflavin formation. From Takino Y, Imagawa H, Horikawa H, and Tanaka A (1964) Studies on the mechanism of oxidation of tea leaf catechins – formation of a reddish/orange pigment and its spectral relationship to some benzotropolone derivatives. *Agricultural Biology and Chemistry* 28: 64–71, with permission.

bright red pigments that give the tea liquor its characteristic 'brightness' and 'briskness'. The contributions of the individual theaflavins to quality differ, with the gallated theaflavins being more astringent.

The other set of brownish compounds formed during black tea processing are called 'thearubigins.' Despite over 60 years' research, a full chemical characterization of thearubigins has yet to be achieved. Indeed, even their contribution to tea quality is not fully understood. The thearubigins are a group of compounds, of which some fractions have a large molecular weight and are nondialyzable compounds. The mixture of groups of compounds normally called thearubigins has been identified to have fractions of polysaccharides, proteins, nucleic acid anthocyanidins, cyanidin, and delphinidin. Indeed, the thearubigin structure could be a polymeric proanthocyanidin. The thearubigins are thought to be polymeric products of the various polyphenols. Indeed, a 4–8 interflavanoid bonding has been considered possible after oxidation of catechin. Over 40 peaks representing pigmented components have been separated by HPLC from black tea liquor, in the range of expected thearubigins. With prolonged fermentation, theaflavins undergo oxidative degradation. Thus, apart from thearubigins being formed directly from the catechins, and breakdown of thearubigins, theaflavin intermediates may also be involved in the coupled oxidation leading to more thearubigins. Thearubigins may not be a group of compounds with a common basic structure. They may be colored compounds resulting from catechin oxidation, catechin coupling without benzotropolone ring formation and catechin–anthocyanidin interactions. Other high-molecular-weight compounds formed by the interactions between flavanol quinones with other macromolecules such as proteins, carbohydrates, and nucleic acids are also classified as thearubigins.

The proposed strategy for the formation of theaflavins and thearubigins from flavanols is summarized in Figure 3. The formation of one theaflavin requires one molecule of a simple catechin and a gallocatechin. Tea shoots normally have higher amounts of the gallocatechins than of the simple catechins. This imbalance in the ratio affects the ability of particular leaves to make theaflavins, and may direct most gallocatechins to make thearubigins. The redox potentials of the individual flavanols to quinones also affect the amounts of the particular quinone available for reaction to produce theaflavin. The availability of a high polyphenol oxidase activity and oxygen are therefore critical in the direction of theaflavins to thearubigins formed under ideal conditions. However, even under ideal conditions, after some time, the formed theaflavins start degrading to thearubigins. Resultant thearubigins also degrade further or polymerize to other thearubigins.

Oxygen is required in three key steps: the oxidation of catechins to quinone, benzotropolone ring formation, and oxidative degradation of theaflavins. The first two are necessary for theaflavin formation. For the formation of high amounts of theaflavins, excess

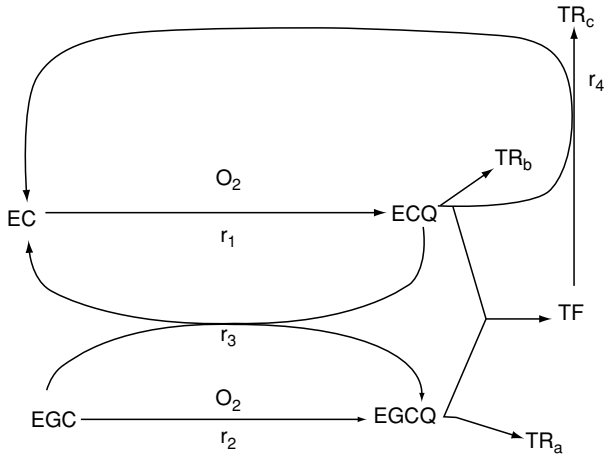

Figure 3 The formation of theaflavin and thearubigin from the catechin and possible role played by the simple catechins in coupled oxidations. EC, epicatechin; EGC, epigallocatechin; ECQ, epicatechin quinones; EGCQ, epigallocatechin quinone; TF, theaflavin; TR, thearubigin.

oxygen should therefore be available. The flavanols have a high affinity for oxygen. When oxygen is limiting, the benzotropolone ring system formation becomes inhibited, and fewer theaflavins are formed. However, during the theaflavin formation, there is also competitive degradation of theaflavins. With excess oxygen and/or when the catechin levels have been depleted, this reaction can be significant, reducing the amounts of theaflavins, but increasing thearubigins. However, this may not be the only way the thearubigins are formed. As in any chemical reaction system, the reaction temperature is critical. Fermentation of temperatures between 15 and 30 °C is beneficial to theaflavin formation. However, a higher temperature fermentation facilitates thearubigin formation at the expense of theaflavins.

Firing terminates the chemical reactions associated with the formation of the nonvolatile black tea components. Firing heat denatures polyphenol oxidase, reduces the moisture content to a product that can be stored, and enhances reactions responsible for black tea aroma. Firing black teas at temperatures above 100 °C increase the blackness, which may be due to pyrolytic reactions.

Thearubigins comprise 9–19% black tea dry weight and 30–60% soluble solids in black tea. They contribute to the 'liquor color' and 'thickness' or 'body.'

Methyl Xanthines

The caffeine is the major purine alkaloid in tea leaves, but theobromine and theophylline are also found, albeit in low quantities. Dry fresh tea shoots contain about 3–4% caffeine. After the shoots are plucked, the levels of caffeine increase slightly during withering. However, the levels do not change during fermentation and may decrease slightly during firing. A good ratio of caffeine and theaflavins in black tea imparts a brisk character to the liquors. High levels of caffeine in tea lead to the so-called 'creaming down' in the liquors, an indication of high quality.

Other Tea Components

Many aroma compounds have been identified in various teas. Their biogenetic pathways in green leaf have been worked out. However, here, only the chemistry of formation of volatile flavor compounds during tea processing shall be highlighted. The aroma compounds in tea can be classified broadly into primary or secondary products. The primary products are biosynthesized by the plant, whilst the secondary products are produced during tea manufacture via enzymatic, redox, or pyrolytic reactions of carotenes, amino acids, unsaturated fatty acids plus other lipids, and terpene glycosides.

Some aroma compounds constitute both primary and secondary products. The primary compounds that have been identified in fresh green teas include *Z*-2-penten-1-ol, *n*-hexanol, *Z*-3-hexen-1-ol, *E*-2-hexen-1-ol, linalool plus its oxides, nerol, geraniol, benzyl alcohol, 2-phenylethanol, and nerolidol. Their quantities change after plucking and during tea processing. In the course of manufacture, the concentrations of some of the alcohols increase, possibly due to hydrolysis of their glycosides, whereas for others, there is a decrease due to oxidative reactions and glycosidation.

Amino Acids

After a tea leaf has been plucked from the plant, the concentration of amino acids in the leaf increases as the proteins in the fresh green tea leaf break down in a process catalyzed by peptidase. The dominant amino acid in tea is theanine (5-*N*-ethylglutamine), accounting for 2% dry weight or 50% of total amino acids in tea. In green tea, theanine is associated with a 'brothy' taste. High levels improve the green tea quality. In black tea, however, high levels reduce the quality. This amino acid does not undergo the chemical transformations that the α-amino acids undergo to make volatile compounds. In black tea processing, the amino acid levels, decrease during fermentation, and this is accompanied by the production of aldehydes, as outlined in Figure 4. Valine, leucine, isoleucine, and phenylalanine are converted to 2-methylpropanal, 2-methylbutanal, pentanal, and phenyl acetaldehyde, respectively. These reactions

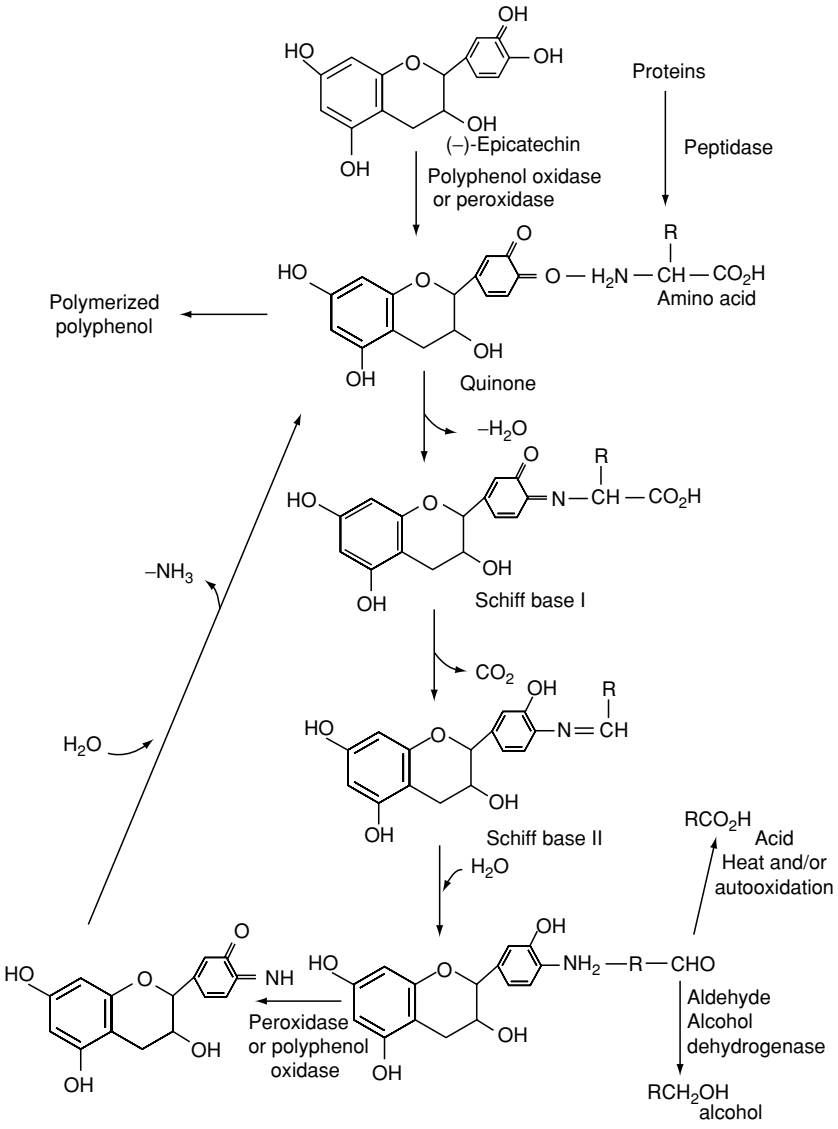

Figure 4 Formation of aldehydes, alcohols, and carboxylic acids from amino acids.

are catalyzed by polyphenol oxidase and/or peroxidase in the presence of oxygen and catechins. Some of the formed aldehydes are further oxidized to carboxylic acids during firing (as the concentration of acids is higher in fired teas) or storage, although some of the aldehydes are certainly reduced to their respective primary alcohols.

Lipids

Lipids make up between 4 and 9% dry weight of the fresh tea leaf and are composed mainly of free fatty acids and fatty acid esters. Linolenic acid is the major fatty acid in tea, but variations in this observation have been noted. The fatty acid profile changes with the geographical area of tea production, agronomic practices, and variety. The levels of fatty acids change throughout tea manufacture. During withering, the fatty acid esters are hydrolyzed to free fatty acids. The unsaturated free fatty acids degrade to form aroma compounds, but the fate of saturated fatty acids during tea processing is unknown.

The mechanism for the degradation of unsaturated fatty acids to aroma compounds is outlined in Figure 5.

Linoleic and linolenic acids produce hexanal and *E*-2-hexenal, respectively, when the acids are added to tea leaf extracts. *Z*-3-Hexenal formed from linolenic acid easily isomerizes to *E*-2-hexenal, and also is the precursor of *Z*-3-hexenol in macerated tea leaves. Alcohol dehydrogenase reduces the aldehydes to alcohols.

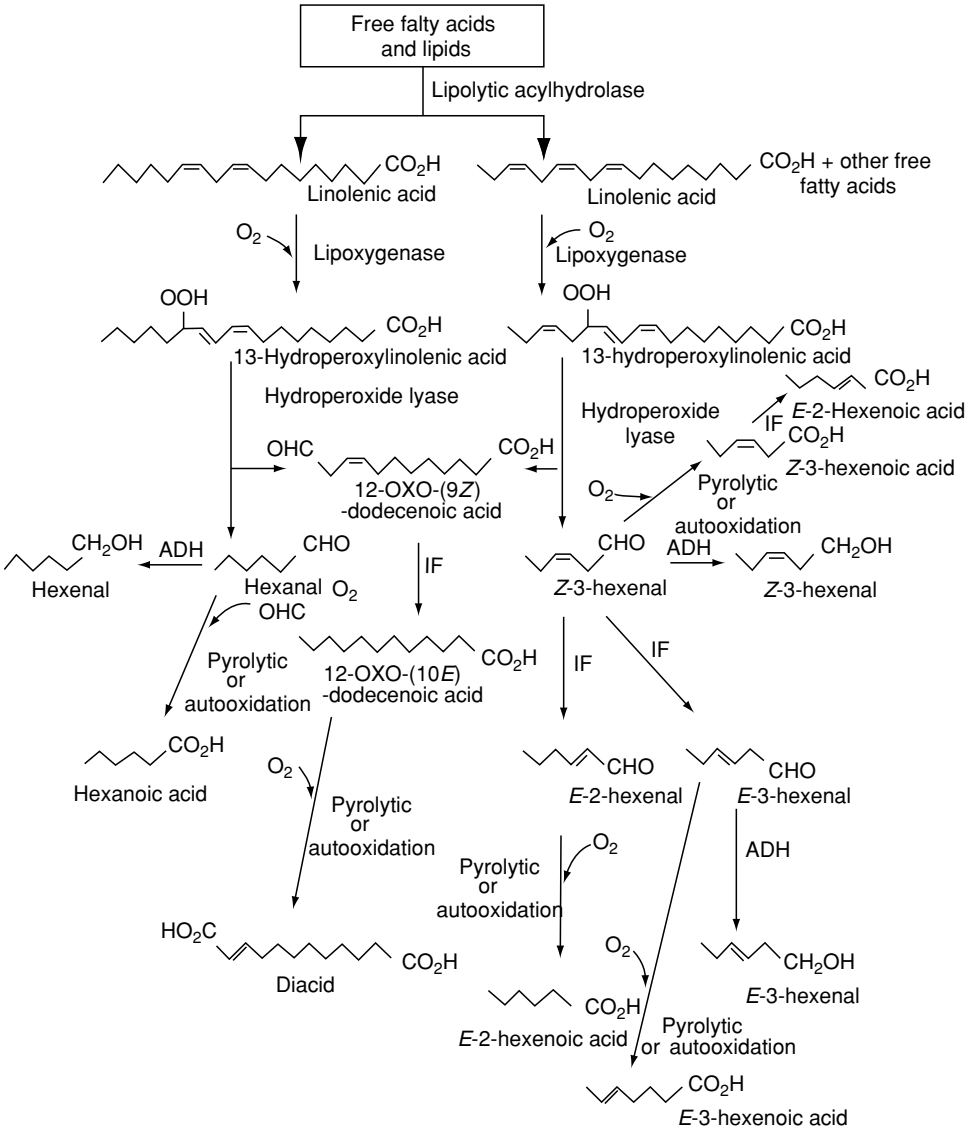

Figure 5 Production of volatile flavor compounds from linoleic and linolenic acids.

The linoleic acid forms 13-hydroperoxy acid, which is an intermediate in the production of C_6 aldehydes and alcohols in tea leaves. The hydroperoxidation of the acid occurs in the presence of lipoxygenase enzyme in a highly stereospecific manner forming only L-hydroperoxy acid. Hydroperoxide lyase breaks down the 13-hydroperoxide acid to C_6 aldehyde and 12-oxo-acid. The action of this enzyme is enentioselective, breaking down only the L-hydroperoxide acids.

The formation of 9-oxo-nonanoic acid from linolenic acid in tea chloroplasts, by cleavage at C-10, suggests that *Z*-3,*Z*-6-nonadienal, *Z*-3,*Z*-6-nonadienol, *E*-2,*Z*-6-nonadienal, and *E*-2,*Z*-6-nonadienol may also be derived from linolenic acid via a similar intermediate. Similarly, cleavage at the C-10 carbon of linoleic acid might be expected to produce *Z*-3-nonenal, *E*-2-nonenal and *E*-2-nonenol. However, only minor amounts of *E*-2,*Z*-6-nonadienal and *E*-2 nonenal have been detected in tea, implying that the hydroperoxidation of linoleic and linolenic acids occurs predominantly at the C-13 carbon to produce the C6 aldehydes and alcohols.

Low levels of palmitoleic and oleic acids have been detected in fresh tea leaves. These fatty acids break down to form heptanal and heptanol, nonanal, and nonanol, respectively, during tea processing. The relationship between precursor fatty acid in fresh leaf

and derived aroma compound in the processed product is rarely linear due to the various interactions that take place during processing. Linolenic acid and 13-hydroperoxylinolenic acid, for example, inhibit the formation of *n*-hexanal from linoleic acid during tea manufacture. In addition, the aroma compounds formed have different boiling points, and more of the lower boiling compounds are lost by volatilization during processing.

Terpene Glycosides

There has been considerable speculation on the mechanism of formation of monoterpene alcohols during tea manufacture. It was originally thought that linalool was a product of carotene degradation. Later, it was suggested that terpene alcohols were produced from oxygenated isoprenoid hydrocarbons. However, it has now been demonstrated that linalool and geraniol are hydrolytic breakdown products of β-D-terpene glycosides during tea manufacture. Indeed, many alcohols in the tea aroma are products of glycoside hydrolysis. In recent studies, several alcohol glycosides have been isolated from tea leaves. These include identified glycosides of 2-phenylethanol, all the four isomers of linalool, geraniol, benzyl alcohol, nerolidol, etc. These glycosides are hydrolyzed during tea manufacture to form their respective alcohols.

Pigments

Fresh tea leaves contain appreciable amounts of pigments, mainly chlorophyls and carotenes. Fresh tea leaves contain about $1.4\,mg\,g^{-1}$ dry weight chlorophyls a and b. During tea processing, the chlorophyls degrade to pheophytins and pheophorbides. These compounds play an important role in giving black tea its shade of color. A number of breakdown products from the phytol side-chain contribute to the aroma complex of tea.

More than 15 carotenoid pigments, dominated by neoxanthin, violaxanthin, lutein and β-carotene, have been identified in fresh tea leaf. These carotenoid compounds account for about 0.5% dry weight of tea leaves. The carotenes decrease during tea processing with the resultant production of various aroma compounds. β-Carotene degrades to β-ionone, whilst β-ionone, α-ionone, 3-hydroxy-β-ionone, 3-hydroxy-5,6-epoxyionone, 3,5-dihydroxy-4,5-dihydro-6,7-didehydro-α-ionone, and other terpenoid aldehydes and ketones are degradation products of other carotenes present in tea leaves. Dihydroactinidiolide, 2,2,6-trimethylcyclohexanone, 5,6-epoxyionone, 2,2,6-trimethyl-6-hydroxycylohexanone, and theaspirone and possibly formed form the primary oxidation products of carotenes, i.e., β-ionone. β-Damascenone, α-damascone, β-damascone, 3-oxo-β-ionone, 1,2-epoxy-1′,2′dihydro-β-ionone, loliolide, dehydrovomifoliol, and 3, 7-dimethyl-1,5-octadien-3,7-diol are speculated to be derived from carotenes via oxidative enzymatic reactions that take place during withering and fermentation, and pyrolytic reactions during firing. The mechanisms of these reactions have, however, not been fully worked out. The formation of these compounds is affected by the amounts of catechins present, oxidase activity, degree of mixing of the cell contents, and concentrations of the reactants. These factors change with degree of wither. Loss of carotenes has been demonstrated to increase with physical withering and fermentation process. Further pyrolytic and photo- and/or autooxidative reactions of carotenes occur during firing to produce more aroma compounds. The compounds produced from carotenes have a major effect on the aroma of tea. Flavory teas are normally produced from green leaf with high carotene contents.

As research on tea aroma continues, it is inevitable that more mechanisms and pathways for the formation of tea aroma compounds will be identified. These will likely involve nonvolatile precursors, which currently are largely ignored with respect to tea aroma and quality. For example, it is known that chlorophyl degrades to phytol and other products, but the contribution of chlorophyl degradation products to tea aroma is not known.

Considerable research has been directed into determining how the aroma complex changes with variations in agronomic, cultural, and manufacturing practices. Many studies have indicated the changes that occur in aroma composition by varying one parameter or the other without any attempts to quantify and classify the contribution of the aroma compounds to quality. Generally, the aroma compounds can be classified into two groups, i.e., those although important for the characteristic black tea smell, are deleterious to black tea quality when present at higher concentrations (group I compounds), and those that impart a sweet flowery aroma to tea, the presence of which is considered to be highly desirable (group II compounds). The classification of aroma compounds in group I and group II compounds has been based on either the odor characteristics or the retention time of the aroma compounds during gas chromatographic analysis. The ratio of group II to group I aroma compounds has been used to classify teas in order of flavor quality. In other studies, the ratio of terpenoid to nonterpenoid compounds has been used.

Although these ratios provide the basis of a semiquantitative method for classifying teas in order of their aroma quality, the ratios must be used with

caution as the olfactory perception limits of the aroma compounds differ widely. Some compounds may be present at very low levels yet have a large impact on aroma and vice versa. For example, methyl epijasmonate has an aroma that is 400 times stronger than that of methyl jasmonate at the same concentration. In addition, some of the compounds considered deleterious to black tea quality are important for green tea quality.

Chemistry of Tea Manufacture

Several tea beverages exist. These beverages include green teas, several semifermented teas, and black teas. The chemistry occurring during their processing varies depending on the desired final product. In green tea processing, oxidative reactions, discussed above, are completely discouraged. In black tea processing, there is more extensive oxidation of the catechins, other polyphenols, amino acids, and unsaturated fatty acids. Fewer oxidative reactions occur in the processing of the semifermented teas.

Black Tea Processing

Withering

Processing of tea beverages starts as soon as the leaf is detached (plucked or harvested) from the plant. The polyphenol oxidase activity decreases, while the catechins levels vary. (−)-Epicatechin, (−)-epigallocatechin gallate, and (−)-epicatechin gallate levels decline. This decline is associated with oxidative transformations. Caffeine levels rise, while protein levels decline. This decline is caused by an increase in the activity of proteolytic enzyme activity, which hydrolyzes the proteins to amino acids, with a concomitant increase in the level of free amino acids. Carotenoid compounds degrade due to photoisomerization to volatile flavor compounds. Fatty acid esters are hydrolyzed to free fatty acids that oxidize during fermentation through a lipoxygenase-initiated reaction to volatile flavor compounds associated with the green notes in tea. Terpene and other alcohol glycosides hydrolyze to simple alcohols that contribute to tea aroma. These transformations, which continue up to the point at which the leaf is macerated, are collectively called 'chemical wither.' Usually, the leaf is subjected to moisture loss to make it more flaccid, so that maceration is easy. Moisture loss, the most visible change in the leaf before maceration, is referred to as 'physical wither.' Chemical wither benefits mostly flavory black teas, as it improves the black tea aroma and to some extent benefits plain black tea as it reduces the level of green taste. Physical wither benefits both flavory and plain black tea quality. Hard physically withered leaves are easier to macerate and make more aromatic black teas. Withering therefore plays an important role in black tea processing.

Maceration

Maceration ruptures the leaf cell structure, exposing the chemical constituents of the cells, mainly polyphenols, oxidative and degradative enzymes, lipids, amino acids, etc., to oxygen. Most importantly, catechins come into contact with the polyphenol oxidase enzyme, initiating 'tea oxidation,' which is erroneously referred to as 'fermentation.' Several methods are used, but the most common are the orthodox rolling, crush, tear, and curl (CTC), and Laurie tea processor (LTP) methods. The method used has a significant effect on the resultant black tea. There is less cell matrix destruction with the use of orthodox rollers than the other two. The orthodox maceration therefore leads to fewer oxidative reactions, and fermentation is slow. Cell matrix destruction is greater in the CTC and LTP maceration methods, leading to more extensive oxidation, but these teas are less aromatic, with higher plain tea quality parameters. LTP manufacture requires a softer physical wither than CTC processing.

Fermentation

Most chemical transformations occur during the fermentation phase of black tea processing. These transformations are responsible for the characteristic taste and aroma products of black tea. As illustrated in Figures 2 and 3, the catechins and polyphenol oxidase form the theaflavins and thearubigins. The amino acids and fatty acids also oxidize to various volatile flavor components. For LTP and CTC manufacture, the process is complete after 60–120 min.

Firing (Drying)

Firing is necessary to terminate fermentation and to dry tea for storage and transport. As a result of the temperature rise, some reactions are accelerated until the rise is adequate to denature the enzymes or moisture has been adequately removed to prevent reactions occurring, but a lot of changes occur to give black tea its character.

As a result of firing, the color changes as a result of the transformation of chlorophyls to pheophytins and pheophorbides. Some caffeine is lost while the amount of volatile flavor compounds is reduced. The volatile flavor compounds that result from various pyrolytic reactions are formed during firing.

Sorting

The fired black tea is sorted first by removal of fiber then by separation into different particle sizes. The various particle sizes define the various grades. Generally, although the grades have different chemical compositions, quality is not solely dictated by grade.

Other Tea

Green and Semifermented Teas

There are fewer chemical transformations in the processing of the other teas compared with black teas. In green tea processing, attempts are made to insure that there is no oxidation, especially that of the catechins. The process therefore starts with steaming or roasting to deactivate polyphenol oxidase activity. The green teas are made from tea plant varieties with a lower catechin content than those for black tea, but this level is sufficient to create astringency. The volatile components of green tea are basically those of the primary products.

Partially fermented teas undergo incomplete fermentation. Several types exist. Some are processed by roasting, whereas others are subjected to high-temperature rolling.

See also: **Amino Acids**: Properties and Occurrence; **Caffeine**; **Carotenoids**: Occurrence, Properties, and Determination; **Chlorophyl**; **Sensory Evaluation**: Aroma; **Tannins and Polyphenols**

Further Reading

Owuor PO (1992) A comparison of gas chromatographic volatile profiling methods for assessing the flavour quality of Kenyan black teas. *Journal of the Science of Food and Agriculture* 59: 189–197.

Owuor PO (1995) Results from factory processing and black tea quality research in Kenya (1982–1994). *Tea* 16: 62–69.

Robertson A (1992) The chemistry and biochemistry of black tea production – the non-volatiles. In: Willson KC and Clifford MN (eds) *Tea: Cultivation to Consumption*, 1st edn, pp. 555–601. London: Chapman & Hall.

Robinson JM and Owuor PO (1992) Tea aroma. In: Willson KC and Clifford MN (eds) *Tea: Cultivation to Consumption*, 1st edn, pp. 603–647. London: Chapman & Hall.

Sakata K, Gou W and Moon JH (1999) Tea chemistry with special reference to tea aroma precursors. In: Jain NK (ed.) *Global Advances in Tea Science*, pp. 693–706. New Delhi: Aravali Books International.

Takino Y, Imagawa H, Horikawa H and Tanaka A (1964) Studies on the mechanism of oxidation of tea leaf catechins – formation of a reddish/orange pigment and its spectral relationship to some benzotropolone derivatives. *Agriculture and Biological Chemistry* 28: 64–71.

Yamanishi T (1999) Tea flavour. In: Jain NK (ed.) *Global Advances in Tea Science*, pp. 707–722. New Delhi: Aravali Books.

Processing

P O Owuor, Tea Research Foundation of Kenya, Kericho, Kenya

Tea

Beverages produced from tea leaves include black tea, green tea, and various partially fermented teas such as oolong and pouchong. All of these products are cultivated and harvested using similar procedures, but variations in manufacturing methods determine the final product. This article deals with tea processing from cultivation through to packaging.

Cultivation

Tea was introduced into many countries of the world from South-east Asia and grows in climates ranging from the Mediterranean to the hot humid tropics. Commercially viable plantations have been established between as far north as Turkey and Georgia (24 °N), and as far south as Argentina (27 °S), at altitudes ranging from sea level to 2700 m.

Successful commercial cultivation of tea requires a minimum annual rainfall of about 1400 mm when irrigation is not carried out. Rainfall needs to be well distributed with at least 120 mm per month. Prolonged drought adversely affects tea growth, and in such conditions, irrigation is advocated. The interaction of soil texture and rainfall distribution is an important factor to be considered when assessing the suitability of an area for tea. Tea does not tolerate water-logged conditions.

Ambient temperatures of 12–30 °C are considered ideal for growing tea. Temperatures above 30 °C, accompanied by low humidity, have been shown to inhibit active growth. The optimum soil temperature for active growth within the feeder root depth is 20–25 °C. Using long-term average yield data from Kenyan tea estates, situated at altitudes between 1500 and 2250 m, it has been shown that annual tea production falls by 200 kg of black tea per hectare for every 100 m rise in altitude. This has been attributed to decreases in air and soil temperature. The tea plant cannot withstand frost conditions. Night

temperatures below 6 °C require that measures be put in place to mitigate frost, which could lead to the wilting of tea leaves.

Tea grows most successfully in acidic soils at pH values ranging from 4.0 to 5.8. A wide range of soil types are suitable for cultivation, including old sedimentary soils derived from gneiss or granite, soils formed from the wash of old sandstone, alluvial, peat, and soils derived from volcanic rock.

Propagation is carried out either from stumps generated from seeds or from cuttings. Propagation from seeds takes at least 3 years from seed planting to the production of viable plants capable of transplanting to the field, whereas plants generated vegetatively require a minimum of 6 months before they are ready for transplanting. However, when the plants are vegetatively produced under very cold conditions, the period needs to be lengthened. Vegetatively produced plants have the additional asset that they are usually of predictable yield and quality. Grafting plants with other desirable characteristics can further enhance these attributes.

After transplanting, the young plants are left to establish for at least 18 months in the field before commercial plucking (harvesting) can commence. During this time, the plant is manipulated through a process described as 'bringing into bearing' to form a plucking table. This is achieved by tipping and pegging. The tea is subsequently harvested regularly, with pruning every 3–5 years, thus maintaining a height at which the leaves and buds can be plucked easily. During this time, the tea is treated with regular applications of fertilizers, particularly nitrogen.

Harvesting

Harvesting (plucking) of tea involves manual or mechanical removal of the terminal young tender portions of peripheral shoots. In most countries, the recommended plucking standard is two leaves and a bud. However, under commercial production, it is difficult to select leaves of one plucking standard exclusively, and it is usual to find varying proportions of the more mature leaves in the harvested crop. The quality of the final product deteriorates with increase in mature leaf content.

After plucking, depending on climatic conditions, new shoots take between 40 and 100 days from bud break to develop into pluckable shoots. In general, the colder or drier the production area, the slower is the shoot growth. The intervals between plucking rounds vary accordingly. The appropriate plucking interval is determined by optimizing between the extent of new shoot development and the level of overgrowth of shoots that have been left after the previous plucking. Hand plucking is the preferred method of harvesting tea as it enables the selection of the quality requirements of two leaves and a bud. Compared with mechanical harvesting, however, hand plucking is slow. Mechanical harvesting is carried out in some countries where labor is either expensive or in short supply. This can be carried out using modified hedge trimmers, motorized machine pluckers operated by one or two people, or a self-propelled machine capable of negotiating the lines of tea bushes.

Black Tea Manufacture

Withering

The changes that occur in the green leaf from the time it is detached from the plant to the time of maceration or rolling are collectively known as 'withering.' These changes can be categorized into physical and chemical processes and are thought to be important in the manufacture of tea.

The main physical process that occurs during withering is moisture loss, leading to changes in cell membrane permeability. These changes are indispensable in orthodox black tea manufacture as they precondition the leaf for maceration or rolling. For teas made by alternate methods of maceration, such as those using the Legg-cut, the Laurie Tea Processor (LTP), or Crush Tear and Curl (CTC) (see section on Maceration), physical withering may not be obligatory. During physical withering, the stomata on the lower surface of the leaf begin to shut gradually, but continue to influence the rate at which water is lost. Air temperature, atmospheric vapor pressure, and air velocity and direction all affect the rate and degree of physical wither. The many biochemical changes occurring during withering are referred to as 'chemical wither.'

Several methods of withering have been developed. Traditionally, withering was carried out on banks of trays in thin layers so that the largest possible leaf surface would be exposed to air. Alternatively withering facilities consisted of open lofts built some distance from the factory, or on the upper, open storeys. These methods have largely become obsolete, as they were inefficient, time-consuming, and incapable of coping with increased crop production.

In most modern factories, withering is now carried out in troughs that can hold between 2000 and 3000 kg of green leaf, loaded to a depth of 25–40 cm. The troughs have dual-direction fans that either drive air into the withering troughs or draw air through them. Under low humidity, physical withering can be achieved using ambient air, even if the troughs are slightly overloaded. However, at high

humidity and/or when the troughs are overloaded because of a greater quantity of crop, stream-heated air is necessary to assist physical withering. Excessive use of heat during the withering process impairs black tea quality. Chemical withering is time-dependent. It requires a minimum of 6 h from the time the leaf is detached from the plant during plucking. Chemical withering beyond 20 h impairs plain black tea quality.

Despite these improvements, withering still takes up considerable factory space, and newer methods continue to be developed. One such method involves storing the leaf in a holding tank with minimal moisture loss for about 6 h to achieve chemical wither. The leaf is then spread on withering troughs, or a moving belt witherer, and the moisture is rapidly reduced by the use of warm air. This process is known as 'two-stage withering.' Other methods of withering used in black tea manufacture include the use of drum withering, tunnel withering, Tocklai continuous withering machines, Russian withering machines, and automatic continuous operating installations. Recent studies indicate that in the unorthodox methods of producing plain black tea, physical withering is not essential to produce high-quality black tea. Some factories are therefore now processing plain black tea after achieving chemical withering but without physical withering.

Factors that affect withering include the leaf type, leaf conditions, standard of plucking, thickness of spread, length of wither, drying capacity of air, and temperature of air used in drying.

Maceration and/or Rolling

Black teas are often referred to as 'orthodox' or 'unorthodox'. These descriptions are derived from the method of leaf distortion employed during manufacture.

In orthodox manufacture, leaves that have achieved physical withering are subjected to rolling. During conventional rolling, the leaf is damaged in such a way that it becomes twisted, and the semi-permeable membrane of the leaf is distorted, allowing the cell juices to be expelled to cover the leaf surface. This allows the juices to mix with the cellular enzymes in the presence of oxygen, and the chemical reactions necessary for fermentation will commence. The orthodox method of manufacture is particularly desirable in the processing of flavory black teas, but this method is slow and requires a large amount of factory space to process high volumes of leaves.

Although some factories still make tea using the orthodox (rolling) system, other methods of maceration have been developed and introduced into many factories. Such methods include Legg-cut, CTC, Rotorvane, LTP, Triturator, and other miscellaneous methods. These methods are particularly desirable for processing plain to medium-quality black teas. They are generally faster than the orthodox methods and can handle high volumes of leaves more easily.

The CTC method is widely employed and consists of two engraved metal rollers close together running in opposite directions. The machines work like a mangle, with one roller rotating approximately 70 rpm and the other 700 rpm. The leaf is cut, torn, and rolled in the small gap between the serrated surfaces of the rollers. Using this method, leaf distortion is much greater than with the most efficient orthodox rollers, fermentation is faster, and liquoring properties are improved.

Another popular method of maceration involves use of the Rotorvane. The rotor, consisting of segments, revolves in a cylinder of 15.20 or 37.5 cm diameter and is fitted with vanes propelling the leaf towards the discharge against resistors projecting from the casing. Maceration is achieved by rubbing and shearing inside the cylinder, and fermentation occurs simultaneously.

The LTP is another successful machine for maceration. This machine resembles a hammer mill and employs a centrifugal fan to induce and discharge leaves.

Most modern factories use a Rotorvane plus three CTC machines in series. However, there are factories that still carry out orthodox manufacture, and a few use the LTP. In general, orthodox teas have a superior aroma to CTC or LTP teas. However, CTC and LTP teas have higher levels of theaflavins and thearubigins and therefore have more color and are brighter and brisker than orthodox teas. (*See* **Phenolic Compounds**.)

Fermentation

In black tea processing, fermentation is defined as the chemical transformations occurring as a result of breakup of the cell membrane due to maceration.

Factories manufacturing tea by orthodox methods generally ferment on tables or trays. In factories that employ CTC, LTP, and Rotorvane systems, fermentation is carried out in batches using troughs or trolleys, or continuously on moving machines. In the case of batch fermentation, the troughs or trolleys are connected to an air supply by a duct, which can be humidified if necessary to reduce the temperature of the fermenting tea (dhool). A trolley can contain 110–130 kg of macerated leaf. The advantage of a trolley system is that the temperature of the dhool can be controlled more precisely than other systems. However, it has the disadvantage that the timing

of fermentation duration is manual and is subject to error.

A number of systems have now been developed, in which fermenting is carried out on a continuous fermenting machine. In most of these systems, the fermenting tea moves on a perforated belt through which air is passed. The speed of the belt determines the throughput and fermentation time. In other continuous methods, the dhool is fed into an open-topped, semicircular tank with two rows of rotating paddles that push the dhool forward in a mechanism similar to that of a screw. The speed of the rotation governs the rate of throughput and duration of fermentation. The temperature of the dhool is normally controlled by external fans. The paddles that continuously turn expose the dhool to air. Although, in most continuous fermenting systems, temperature control can be difficult, the timing of fermentation duration is usually very precise.

The liquor characteristics of the black tea can be determined by controlling the temperature and lime of fermentation. Generally, the lower the fermentation temperature, the better the black tea. Usually, the ambient air blown through fermenting dhool supplies adequate oxygen. Fermenting under enriched oxygen atmosphere therefore does not improve the liquor quality provided that the fermentation duration is timed correctly.

Firing (Drying)

Firing of tea is the process that reduces the moisture content of fermented tea from about 60% to below 4% and renders the product in a form suitable for storing. Firing terminates fermentation by deactivating the enzymes by subjecting the dhool to high temperatures. Inlet temperatures in the driers usually range from 140 to 98 °C with outlet temperatures ranging from 45 to 82 °C. During firing, considerable amounts of volatile aromatic compounds are lost.

Firing can be carried out using conventional driers in which the dhool is fed on to a perforated moving belt and is discharged off after the tea is dry. In most modern factories, fluidized bed driers are used. In these driers, hot air is blown into the drier, and this moves the dhool by the process of fluidization. Generally, fluidized bed driers have a higher throughput than conventional driers.

Grading and Sorting

Prior to grading, stalks are removed by the use of electrostatic separators. The process is effective because of the higher moisture content of stalks as compared with leaves after emergence from the drier.

Grading of leaves is generally carried out using mechanically oscillated sieves fitted with meshes of many different sizes. In some machines, the sieves are arranged in banks of diminishing mesh size, such that the outfall of the upper sieve falls on to the lower.

The products of the various siftings constitute the different grades (Table 1). The grade specification is entirely artificial. However, they are generally recognizable by their appearance; for instance, broken orange pekoe contains a high proportion of buds, orange pekoe is characterized by the abundance of twisted tender stalk, and pekoe and souchong grades tend to be more compact and dense. In recent years, the pattern of grading has been altering in the direction of smaller teas.

The sieve standards adopted for regulating grades differ in the various countries and even in different districts in the same country. However, there are efforts to standardize the sieve sizes.

Winnowing in one form or another is routinely employed and, according to the size and density of the particles, separates fannings and dust, carrying away the fibrous residue that is of no commercial value as a grade.

Packaging and Storage

Tea is hygroscopic, and if not adequately packed and stored, it can readily absorb considerable amounts of moisture, leading to a deterioration in quality.

Most tea is transported in bulk, and the common forms of packaging employed are either tea chests or multiwalled paper sacks. Tea chests are normally made out of plywood, lined on the inside with aluminum foil, which acts as a moisture barrier.

Table 1 Black tea grades

Leaf grades	
TGFOP	Tippy golden flowery orange pekoe
TFOP	Tippy flowery orange pekoe
FOP	Golden flowery orange pekoe
OP	Orange pekoe
FP	Flowery pekoe
Broken grades	
TGFBOP	Tippy golden flowery broken orange pekoe
TGBOP	Tippy golden broken orange pekoe
GFBOP	Golden flowery broken orange pekoe
TBOP	Tippy broken orange pekoe
GBOP	Golden orange
FBOP	Flowery orange pekoe
BOP	Broken orange pekoe
BP	Broken pekoe
BPS	Broken pekoe souchong
PS	Pekoe souchong
S	Souchong
BM	Broken mixed
BT	Broken tea

Multiwalled paper sacks have flat hexagonal ends to facilitate stacking and consist of a minimum of two plies of Kraft paper with an additional layer of aluminum foil on the inside of the sack to prevent moisture migration. Materials used for the construction of the sacks should be free from taint or odor. Both sacks and tea chests are designed to hold 40–60 kg of manufactured tea depending on grade.

Retail packaging of tea is wide and varied, but the common consideration is the requirement to keep out moisture. Tea should be stored in conditions that minimize the absorption of moisture, preferably at a low humidity. If teas are stored for long periods, however, moisture absorption and subsequent loss of quality are inevitable. (*See* **Storage Stability**: Parameters Affecting Storage Stability.)

Green Tea Manufacture

Green tea is manufactured from fresh leaves that have not been fermented. There are various methods of green tea manufacture, but they all depend on stopping enzyme activity in the green leaf. Withering is not mandatory in green tea manufacture.

In China, green tea manufacture begins by roasting the leaf in a hot iron pan for a few minutes, followed by hand rolling on a table. The leaf is then subjected to two or more further roastings and rollings.

In Japan, the leaf is steamed for 15–20 s in a revolving cylinder provided with an agitator. The steamed material is cooled by a fan or by air on a belt conveyor and then subjected to primary heating and rolling. The leaf may undergo further heating and drying before passing through secondary (final) rollers. The green tea is then dried to a moisture content of about 3–4%.

Partially Fermented Tea Manufacture

Oolong tea is manufactured in a similar manner to green tea, but with the following variations. The fresh leaf is withered at room temperature for approximately 16 h, or at 40 °C for 2 h followed by another 4 h at room temperature. In both cases, during the last 4 h, the leaf is rolled by hand for a period of 30 min every hour. This is followed by roasting (parching or pan frying) at about 160 °C for about 20 min. The tea is rolled further and then fired.

The manufacturing process for pouchong tea is slightly different than for oolong. The leaf is withered in the sun (solar withering) for 15 min, during which time it is turned over once. This is followed by indoor withering for 3 h, during which time the leaf is turned over three times. The leaf is then pan-fried at 160 °C for 20 min, rolled by hand for about 20 min, and dried at 80–85 °C for 40 min. Low-grade pouchong tea is often scented by blending with jasmine flowers to enhance the flavor (jasmine tea). Other partially fermented teas such as teekwang-yin and pan-fried longjing are also manufactured in a similar manner.

Other Tea Products

Earl Grey is flavored with the peel oil of bergamot, a citrus fruit, which is added by spraying on to black tea before final packaging. Jasmine flowers are sometimes added to manufactured black tea in the country of origin, and these impart a characteristic floral note. Lapsang souchong tea is a Chinese black tea flavored with natural smoke.

Instant teas are prepared in the producer countries by infusing the undried leaf and then evaporating the liquor by either freeze-drying, spray-drying, or vacuum-drying. All these drying procedures avoid using excessive heat and therefore reduce the loss of flavor components. Instant tea is also produced in the USA and UK, but details of the processes are secret or protected by patents.

In recent years, decaffeinated teas have appeared on the market. The tea is decaffeinated with methylene chloride or other chlorinated solvents and supercritical carbon dioxide. (*See* **Coffee**: Decaffeination.)

See also: **Coffee**: Decaffeination; **Phenolic Compounds**; **Storage Stability**: Parameters Affecting Storage Stability

Further Reading

Anonymous (1986) *The Tea Research Foundation of Kenya Tea Growers Handbook*, 4th edn. Nairobi: Eleza Services.

BSI (1982) *British Standard Glossary of Terms Relating to Black Tea. BS 6325*. Milton Keynes, UK: British Standards Institute.

Eden T (1976) Tea. In: *Tropical Agricultural Series*, 3rd edn. London: Longman.

Orchard JE and Owuor PO (1989) Black tea quality. Effects of manufacturing practices on tea quality. *Tea* 10: 117–160.

Owuor PO (1990) Some advances in the improvement of tea quality through chemical and biochemical research in Kenya. *Tropical Science* 30: 307–319.

Owuor PO and Obanda MA (1996) The impact of withering temperature on black tea quality. *Journal of the Science of Food and Agriculture* 70: 288–292.

Owuor PO and Obanda M (1997) Advances in withering technologies and future strategies to cope with high tea production in Kenya. *Tea* 18: 184–193.

Owuor PO and Obanda M (1998a) Influence of enriched oxygen atmosphere during fermentation on plain clonal black tea quality parameters development. *Tropical Science* 38: 165–170.

Owuor PO and Obanda M (1998b) The changes in black tea quality parameters due to plucking standards and fermentation time. *Food Chemistry* 61: 435–441.
Owuor PO and Orchard JE (1989) Changes in the biochemical constituents of green leaf and black tea due to withering. A review. *Tea* 10: 117–160.
Owuor PO, Orchard JE and McDowell IJ (1994) Changes in the quality parameters of clonal tea due to fermentation time. *Journal of the Science of Food and Agriculture* 64: 319–326.
Takeo T, Tsushida T, Mahanta PK, Tashiro M and Imamura Y (1985) Food chemical investigations on the aroma of oolong tea and black tea. *Bulletin of the Natural Research Institute of Tea, Japan* 20: 91–180.
Werkhoven J (1974) *Tea Processing. Agricultural Services Bulletin, No. 26*. Geneva: Food and Agriculture Organization of the United Nations.
Willson KC and Clifford MN (eds) (1991) *Tea: Cultivation to Consumption*. London: Chapman & Hall.

Analysis and Tasting

P O Owuor, Tea Research Foundation of Kenya, Kericho, Kenya

Tea Quality

The term 'quality,' as used in the tea trade is sometimes subjective and somewhat illusory. Largely, the tea tasters set the quality standards, and most scientific methodologies are based on what tea tasters consider as important. Tea tasters have set a large vocabulary of descriptive terms with which to assess the different parameters considered to contribute to tea quality. The most significant quality parameters are associated with smell, taste, and appearance. Considerable effort has been made to correlate various constituents of the fresh leaf, processed tea, or infusions with tea sensory evaluation. However, a particular characteristic that may be highly desirable in tea destined for a specific market could well be considered detrimental in a tea required by another market. Requirements also vary between the different processed products, e.g., the green, fresh aroma that is essential for green tea quality would be considered a defect in a black tea. Furthermore, a tea that may be described by a tea taster as an overall 'quality' tea may not satisfy the target market demands: therefore, quality does not necessarily equate with profit. Indeed, whereas producers tend to manufacture teas with desirable quality characteristics, in many instances, tea prices are determined by supply and demand. This does not, however, imply a lack of quality tea. Producers of teas with most desirable characteristics usually realize better values, especially when tea supply outstrips demand.

With the exception of a few specialty brands, most black tea is bought in bulk by tea buyers and then blended to suit target market demands. After blending, the tea loses its identity, and the producer has no control over the quality of the final product. A tea buyer is well aware of the quality of the teas required for blending for their own particular market, and the availability of teas of such quality will determine whether they buy and the prices they are prepared to pay. Occasionally, this can result in a high demand for lower-quality teas, whereas the premium-quality grades remain unsold or fetch relatively low prices. A producer might thus be tempted to concentrate on production of lower-quality teas, but the vagaries of the market are such that the emphasis may well have been reversed by the time their product reaches the market, and, more seriously, such a policy results in a general decline in quality, which damages the image of tea as a beverage. It is essential for producers to maximize the quality of all their grades, particularly as tea is invariably overproduced.

Tea Tasting

As with many products, tea requires constant quality evaluation during its processing, marketing, packaging and storage. Tea tasting has evolved, over centuries, as the most appropriate method to the needs of the trade. This is particularly so as instrumental methods of analysis tend to be slow and laborious, whereas tea tasting is fast and can evaluate a large number of samples within a short duration. The following method is specifically for black teas, but similar sensory and visual appraisals are also employed for other processed tea beverages.

The technique involves the description and evaluation of teas both before and after infusion in boiling water. A batch of tea samples is laid out in a row, and each sample is assigned an earthware or China bowl and a cup fitted with a lid, both of which are of a standard capacity. Tea equivalent to the weight of a silver sixpenny piece, i.e., 2.83 g (0.1 oz or 44 grains), is placed into each cup. Boiling water is added to the cups to produce a 2% brew, and the lid applied. After infusing for 5 or 6 min, the liquor is poured into the bowl, and the infused leaf is shaken from the cup, pressed free of excess moisture, and placed in the inverted lid, which is then placed on top of the cup.

The dry leaf is examined and evaluated for color grade or uniformity of particle size and form, degree

of manufacture and presence of leaf tip (for orthodox manufactured leaf) and/or undesirable stalk and fiber, feel, and aroma. Attention is then drawn to the infused leaf, which, disconcertingly, is often referred to as the 'infusion.' Ideally, this should be of a bright copper color and virtually free of the green tinge of chlorophyl. From the color and evenness of the infused leaf, the taster is able to make an assessment of the degree and quality of the fermentation.

On removing the lid, the aroma of the vapor present in the cup is evaluated. The liquor is then judged whilst it is still warm, first for appearance, which should be bright and clear with a distinct reddish tinge and a faint pink meniscus where the liquor touches the cup, then for taste. When cool, the better-quality liquors become opaque due to a process called 'creaming down,' which is essentially the precipitation of very finely divided colloidal material derived from caffeine, protein, and polyphenols, particularly the theaflavins.

The taste characteristics of the liquor are of particular importance. In nonscientific vocabulary, the terms 'taste' and 'flavor' are often considered synonymous, but in scientific nomenclature, these words are used to describe two different and distinct characteristics. The flavor of an aqueous solution comprises:

1. The basic taste sensations of sweet, sour, bitter, and salt, which are detected by the tongue and produced by nonvolatile substances present in appreciable amounts.
2. The odor, or aroma, detected by the nose and produced by vapors of volatile components, many of which are present only in minute amounts.

'Taste' refers only to those sensations produced by nonvolatile components. In tea tasting, the term 'taste' is used in its wider sense and therefore includes the aroma. The actual tasting (**Figure 1**) is carried out by sucking, rather than sipping, so that the liquor is drawn to the back of the mouth, on an inward breath, and up to the olfactory nerve in the nose. The liquor is then swished backwards and forwards and brought into contact with the tongue, palate, and other areas of the mouth where sensory receptors are located. Using this skillful, if somewhat noisy, technique, the taster is able to feel, taste, and smell the liquor virtually simultaneously and is thus able to determine its briskness, strength, body, and flavor. The liquor is either taken directly from the cup or from a special spoon. After tasting, the liquor is not swallowed, but discarded into a mobile spittoon.

The tea tasters' report concentrates on appearance, color, strength, pungency, and flavor. For each parameter or character, a large number of specialized descriptive terms are available for use by the tea taster. A tea taster can taste over 50 samples per session. The procedure is strictly standardized, and tea tasters require considerable individual training and experience in order to attain the necessary standards of competence.

Figure 1 (see color plate 133) Tea-tasting session (courtesy of the Tea Board of Kenya, Nairobi).

Although it is clearly a subjective method, prone to influence by variables such as supply and demand, the tasters' state of health, and personal preferences and prejudices, tea tasting is still considered to be the best method available for assessing tea quality as far as the tea trade is concerned. The success of the method lies in the fact that it is quick and inexpensive, and requires little equipment. From a scientific point of view, the method is less acceptable, and several chemical methods are currently being developed, and employed, as replacements or supplements to the tea tasters' evaluation of tea quality. Despite the advanced analytical equipment and procedures that are currently available, it is still not possible to replace tea tasting with a single, reliable, and objective scientific method.

Chemical Analysis of Black Tea

The International Organization for Standardization specification *Black Tea – Definition and Basic Requirements* (ISO 3720) specifies certain chemical requirements for black tea, which, when met, are considered to be indicative of a tea produced by recognized good production practice. These chemical requirements can be found in **Table 1** together with the individual ISO specification numbers for each parameter. ISO 3720 defines black tea as that tea derived solely from the leaves, buds and tender stems of varieties of the species *Camellia sinensis*. With the exception of ISO 6078 (*Black Tea*

Table 1 Chemical requirements for black tea

Characteristic	*Requirement (%)*	*Test method*
Water extract		
Minimum	32	ISO 1574
Total ash		
Maximum	8	ISO 1575
Minimum	4	
Water-soluble ash of total ash		
Minimum	45	ISO 1576
Alkalinity of water-soluble ash (as KOH)		
Maximum	1.0[a]	ISO 1578
Minimum	3.0[a]	
Acid-insoluble ash		
Maximum	1.0	ISO 1577
Maximum	16.5	ISO 5498

[a]When the alkalinity of water-soluble ash is expressed in terms of mm KOH per 100 g of ground sample, the limits shall be 178 (minimum) and 53 mm (maximum).
Courtesy of the Kenya Bureau of Standards.

Vocabulary, a glossary of terms relating to black tea), the International Organization for Standardization has not dealt with a methodology for pricing black teas. More recently, the ISO has embarked on the development of chemical methods for classifying tea based on polyphenol content. However, efforts have not reached the standards stage. Tea prices, and quality, continue to be based on tea tasters' evaluations. In fact, it is generally accepted that expert tea tasters have the ability to assess whether a tea would be unlikely to comply with ISO 3720, and in practice, chemical analysis tends to be carried out only if the tea taster considers a tea to be 'suspect.'

Several chemical methods have been developed to determine specific quality parameters of tea. The volatile components of tea, often referred to as the 'aroma complex,' have been studied in depth by gas chromatography and mass spectrometry, and over 600 compounds have been identified. Routine methods for determining aroma have been developed. In one such method, the ratio of the gas chromatographic peak areas of the volatile compounds eluting before linalool to the sum of peak area of linalool plus all volatile compounds eluting after linalool was used as the aroma quantifying index. In another method, the ratio of the sum of the gas chromatographic peak areas of terpenoid compounds to nonterpenoid compounds was used as the aroma index. In yet another method, the ratio of the gas chromatographic peak area of linalool to that of 2-*E*-hexenal was used. For the last aroma index development, the major volatiles are divided into two distinct groups; substances that have a pleasant aroma and are highly desirable (group II) and substances that, although essential constituents of tea, are considered to have a deleterious effect on flavor when present in large amounts (group I). Lists of typical compounds in groups II and I are shown in Table 2. The ratio of the sum of the group II to group I volatile compounds gas chromatographic peak areas called the flavor index (FI) was used to quantify the black tea aroma quality. Using Kenya clonal black teas, the FI was recently shown to be a better measure for aroma quality compared with the other indices. The volatiles are usually extracted by simultaneous steam distillation/solvent extraction, separated by gas chromatography, and identified by mass spectrometry. The volatile FI is the ratio of group II to group I volatiles. Flavory teas have larger amounts of group II compounds in relation to group I.

Phenolic substances dominate the nonvolatile flavor components. In the case of black tea, most of the taste is believed to be due to the presence of the unique phenolic fermentation products theaflavins and thearubigins plus unoxidized catechins. The sensory qualities of the theaflavins, thearubigins and catechins are usually described in terms of mouth feel and color characteristics. The theaflavins are responsible for briskness and brightness, the thearubigins are considered responsible for body and color, and the catechins contribute to astringency. In general terms, teas with a high level of thearubigins relative to theaflavins and low catechin levels are considered to impart a soft mouth feel and dull color, and in the reverse situation, the teas are often described as lacking body and color.

The catechins, theaflavins, and thearubigins are determined by spectrophotometric methods, which gives a total value for each group. However, high-performance liquid chromatography (HPLC) is now

Table 2 Some volatile flavor compounds used to determine the flavor index of black teas

Group I volatiles	*Group II volatiles*
Isovaleraldehyde	Linalool oxide, (*Z*)-furanoid
2-Ethylfuran	Linalool oxide, (*E*)-furanoid
Hexanal	Benzaldehyde
1-Penten-3-ol	Linalool
Heptanal	β-Cyclocitral
(*Z*)-3-Hexenal	Ho-trienol
(*E*)-2-Hexenal	1-Ethylformylpyrrole
Penten-1-ol	Phenylacetaldehyde
(*Z*)-2-Penten-1-ol	α-Terpineol
Hexan-1-ol	Linalool oxide, (*Z*)-pyranoid
(*Z*)-3-Hexen-1-ol	Linalool oxide, (*E*)-pyranoid
Nonanal	Methylsalicylate
(*E*)-2-Hexen-1-ol	Nerol
(*E*)-2,4-Beptadienal	Geraniol
(*E,E*)-2,4-Heptadienal	Benzy alcohol
	Nerolidol
	Bovolide
	Dihydroactimdiolide

applied to these analyses as it allows the catechins and theaflavins to be determined individually. Black teas have four dominant theaflavins which have different astringencies and have variable contribution to tea taste. Their contribution to taste has been normalized to theaflavin digallate equivalent, which relates better to sensory evaluations. The chemical structures of thearubigins have not been determined.

Contaminants: Heavy Metals and Pesticide Residues

Several producer and consumer countries have passed legislation stipulating maximum permissible levels of heavy metal contaminants in tea. For example, in Kenya, according to the Food, Drugs and Chemical Substances Act (1978), limits were set for heavy metal and fluoride contaminants in black tea, based on the dry weight (Table 3). In the UK, statutory limits have been set for black tea, on a dry-weight basis, at 1 mg kg^{-1} for arsenic and 5 mg kg^{-1} for lead, and guideline limits have been set at 50 mg kg^{-1} for zinc and 150 mg kg^{-1} for copper.

High levels of lead have often been associated with tea, particularly in the past when lead linings were used in tea chests and lead-coated drier trays were used in tea factories. Lower concentrations were achieved by stopping such practices, but even so, levels of about 2 mg kg^{-1} are considered to be unavoidable. Evidence shows, however, that most of the lead remains in the tea leaves on brewing and is not consumed. Furthermore, tea leaves have been shown to absorb lead from water.

The Codex Alimentarius of the Food and Agriculture Organization of the United Nations World Health Organization has set down maximum permissible levels of various pesticide residues in black tea on a dry-weight basis (Table 4).

Most of the heavy metal and pesticide residue levels that have been set are based on the dry weight of processed tea. However, tea is drunk as an infusion using water of widely variable quality and at different temperatures. The solubility of the contaminants therefore varies. This has been largely ignored in the setting of the various limits.

Table 3 Maximum limits of heavy metals and fluoride contaminants permitted in black tea in Kenya

Contaminant	*Limit (mg kg^{-1})*
Arsenic	1.0
Lead	10.0
Copper	150.0
Zinc	50.0
Fluoride	100.0

Table 4 Maximum permissible levels of pesticide residues in tea

Pesticide	*Level (mg kg^{-1})*
Ethion	6.0
Fenitrothion	0.5
Methidathion	0.1
Methylparathion	0.2
Cyhexatin	2.0
Bromopropylate	5.0
Methylechlorpyrifos	0.1
Cartap	20.0
Permethrin	20.0
Deltamethrin	10.0

Off-flavors and Taints

Causes of Off-flavors and Taints

Off-flavor, or taint, may be defined as any taste or odor that is considered to be abnormal or foreign for tea. Off-flavors originate from a myriad of sources and can be attributed to either single or multiple chemical substances that are not normally present in tea or occasionally to unusually high concentrations of compounds that are generally found to be present only in trace amounts.

The chemicals responsible for taints are mostly volatile organic compounds with very low odor thresholds. Taint detection is an extremely time-consuming and difficult process as the compounds responsible are usually present at very low concentrations. Once tea is tainted, the condition is irreversible, and the subsequent loss in revenue is very often substantial. It is therefore imperative to take all possible measures to avoid the formation of such undesirable compounds in the first place.

Good factory hygiene is of paramount importance, and strict control of each stage of the manufacturing process is essential in order to insure an evenly processed product of the best possible quality. If hygiene standards or manufacturing controls are allowed to slip, an inferior product results, which is very often tainted or off-flavored. Additional common causes of taints in tea are described below.

Moisture Due to its hygroscopic nature, processed tea is particularly liable to generate or pick up taints; therefore, strict control of the moisture content is crucial. Moisture contents in excess of 6.5% are extremely detrimental as this encourages rapid fungal and bacterial growth. Volatile compounds produced

during the course of proliferation of fungi and bacteria are a major source of taints in tea. Taints from this source are variously described by tea tasters as 'musty,' 'fusty,' 'moldy,' 'gone-off,' 'sweaty,' and 'sour.' The presence of high levels of moisture in the tea at the time the taint is first perceived is indicative of the cause of such taints, and the identities of the individual compounds responsible are thus rarely pursued. However, such fungal and bacterial infections can also occur during the course of processing, particularly during the withering and fermentation stages, if the leaf is excessively damp or the conditions unclean. In such cases, the moisture content of the finished product would not necessarily be elevated, and the source of the contamination would be much more difficult to locate.

Conversely, teas that come out of the drier with moisture contents of 2.5%, or less, generally have a 'smoky' taint, if firing has been carried out using a conventional dryer. This renders them equally unacceptable.

Chlorophenols and chloroanisoles Chlorophenols impart disinfectant-like taints and are responsible for numerous incidences of tainting throughout the food industry. A major cause for concern is that many fungi and bacteria can readily convert chlorophenols to chloroanisoles by a simple process of methylation. Chloroanisoles possess intense musty/moldy odors, often described as 'old cellars' or 'damp sacks,' and have much lower odor thresholds than the parent chlorophenols. 2,4,6-Trichloroanisole and 2,3,4,6-tetrachloroanisole, both of which have been found in tainted teas, have odor thresholds in water of 2×10^{-5} and $4 \times 10^{-3}\,\mu g\,kg^{-1}$, respectively.

Due to the widespread usage of chlorophenols and their derivatives, considerable vigilance is required in order to minimize the chances of contamination. Chlorophenols have been detected in multiwalled paper sacks (and in the adhesives and printing inks used in their manufacture), wooden pallets, and containers used for the storage and transportation of tea. Chlorophenols from such sources have been deemed responsible for chloroanisole contamination in teas.

The use of certain chlorinated herbicides, fungicides, and pesticides in the field, and algaecides or disinfectants in the factory, may give rise to chlorophenol contamination. Storage of tea in the vicinity of such chemicals is also liable to cause tainting. In fact, tea must not be stored in close proximity to any commodity with a strong smell due to its remarkable ability to pick up odors. In addition, factory water supplies should not be chlorinated, as chlorophenols can also be formed by the reaction of chlorine with phenols in the tea or in the environment.

Detection and Analysis of Compounds Responsible for Taints

Tea tasters usually detect taints, in the first instance. This may take place at the factory, and in such cases, depending on the description of the taint, the source can be malfunctioning equipment or bad handling practices. Once the tea has been transported or stored, it can be extremely difficult to identify the source of the contamination, and it is only considered cost-effective to pursue the source of such taints in the events that they recur.

The techniques used for isolating and eradicating persistent, recurring taints are many and varied, as, indeed, are the agents responsible for such taints. Techniques employed are required to be meticulous and methodical, and involve two distinct processes. Firstly, the movements of the product from the time the taint was first perceived are traced back, even as far as the field, if necessary, in an attempt to pinpoint the association with any agent, or identify any deviation from normal practices, capable of causing the described taint. This requires the cooperation of all concerned and a knowledgable and skillful coordinator. Circumstantial evidence collected in this way can be most successful when confirmed by chemical analysis. Secondly, analysis of the tea is undertaken to identify the compounds responsible for the taint. The flavor analyst initially discusses the description of the taint with the tea tasters and then proceeds to search for specific compounds known to cause such taints. Analytical methods used include headspace gas chromatography, continuous steam distillation, and solvent extraction of the volatiles, followed by separation by gas chromatography (including splitting of the gaseous mixture exiting the gas chromatography column in order to enable 'sniffing' of the individual sample components), preparative gas chromatography to concentrate relevant fractions, specific detector gas chromatography, gas chromatography–mass spectrometry (including multiple ion monitoring) and HPLC.

See also: **Contamination of Food**; **Pesticides and Herbicides**: Residue Determination; **Sensory Evaluation**: Taste; **Taints**: Types and Causes; Analysis and Identification

Further Reading

Owuor PO, Othieno CO and Reeves SG (1987) The Kenyan position on theaflavins in tea. *Journal of Food and Agriculture* 1: 185–186.

Owuor PO (1992) A comparison of gas chromatographic volatile profiling methods for assessing the flavour

quality of Kenyan black teas. *Journal of the Science of Food and Agriculture*. 59: 189–197.
Robinson J and Owuor PO (1991) Tea aroma. In: Willson KC and Clifford MN (eds) *Tea: Cultivation to Consumption*. London: Chapman & Hall.
Rothe M (1988) *Introduction to Aroma Research*, 1st English edn. Berlin: Akademie Verlag.
Werkhoven J (1974) Tea processing. In: *Agricultural Services Bulletin, No. 26*. Geneva: Food and Agricultural Organization.
Whitfield FB (1983) Some flavours which industry could well do without. Case studies of industrial problems. *CSIRO Food Research Quarterly* 43: 96–106.

Texture *See* **Rheological Properties of Food Materials**; **Sensory Evaluation**: Sensory Characteristics of Human Foods; Food Acceptability and Sensory Evaluation; Practical Considerations; Sensory Difference Testing; Sensory Rating and Scoring Methods; Descriptive Analysis; Appearance; Texture; Aroma; Taste

Therapeutic Diets *See* **Elderly**: Nutritional Management of Geriatric Patients; **Food Intolerance**: Types; Food Allergies; Milk Allergy; Lactose Intolerance; **Hypertension**: Nutrition in the Diabetic Hypertensive; **HIV Disease and Nutrition**; **Infection, Fever, and Nutrition**; **Multiple Sclerosis – Nutritional Management**; **Pregnancy**: Nutrition in Diabetic Pregnancy

Thermal Processing *See* **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; Chemical and Microbiological Changes; Electrical Process Heating

THERMOGENESIS

P Trayhurn, University of Liverpool, Liverpool, UK

Background

The term 'thermogenesis' comes from the Greek word *thermos* for heat. All metabolic processes in animals (and plants) produce heat, as a reflection of thermodynamic inefficiency, and are therefore thermogenic. In general, however, thermogenesis is used to describe a facultative, or adaptive, process of heat generation, i.e., a process in which heat is the primary product of metabolism. It is also used to describe the heat produced in direct response to a meal.

The concept of facultative thermogenesis is well-established in relation to the generation of heat to maintain body temperature (thermoregulatory thermogenesis). It is also widely used in nutritional science to describe the adaptive production of heat in response to overfeeding (diet-induced thermogenesis), particularly as a component of the regulation of whole-body energy balance.

Thermogenesis and Thermoregulation

General

In homeotherms, heat is generated to maintain body temperature (at $\sim 37\,^{\circ}C$ in mammals) when an

animal is in a cold environment. In this context, 'cold' is defined as a temperature below the thermoneutral zone, i.e., below the range of temperatures at which metabolic rate is at a minimum. The thermoneutral zone varies widely between species, being for example, approximately 32–33 °C in mice, 26–28 °C in adult humans, and of the order of 0–25 °C in mature sheep (depending on factors such as the plane of nutrition and the length of the coat). At thermoneutrality body temperature is maintained above the ambient temperature by the heat generated by obligatory thermogenesis. This is the heat produced as a byproduct of all the basal metabolic processes of an animal, such as the pumping of the heart, the synthesis of proteins, and ion pumping across membranes.

The lowest point of the thermoneutral zone is referred to as the lower critical temperature. Once the environmental temperature has fallen below this point, then the rate of heat loss exceeds the rate of obligatory, or basal, heat production and body temperature will fall. Factors such as age, nutritional status, and the degree of thermal insulation all affect the thermoneutral zone and the lower critical temperature of an animal.

Two distinct mechanisms are available for thermoregulatory heat production – shivering and nonshivering thermogenesis. Shivering, which involves small contraction of the skeletal muscles, is a widely employed mechanism for the generation of heat in the cold. It is, however, behaviorally disadvantageous, in contrast to nonshivering thermogenesis where heat is produced without muscle contraction. The term 'thermogenesis' when applied to thermoregulation is often used with exclusive reference to nonshivering thermogenesis.

Adaptation of Adult Animals to Cold Environments

Exposure of mature homeotherms to a temperature below, or colder, than the lower critical temperature imposes an immediate requirement for thermoregulatory thermogenesis in order to maintain body temperature. The generation of heat by shivering is a major part of the initial response to cold. However, during long-term (days to weeks) cold exposure, heat generated by shivering is gradually replaced by nonshivering thermogenesis. In laboratory animals full adaptation to cold is considered to require a period of up to 3–4 weeks.

Newborn

The newborn of a species has a higher surface area to body mass ratio than the mature or adult animal, and insulation (hair, fur) is often absent or underdeveloped. These two factors result in a higher rate of heat loss at a given environmental temperature in the neonate than in the adult homeotherm. As a consequence, neonates are particularly vulnerable to cold and a high capacity for nonshivering thermogenesis is generally evident. The transition at birth from the intrauterine environment to the external world can present a major thermal challenge to the mammalian neonate, particularly in the case of precocial species. Lambs, for example, can experience a fall in ambient temperature in excess of 50 °C over a few minutes when being being born in cold northern climates in late winter. In some species, such as large ruminants (e.g., sheep, reindeer) the capacity for nonshivering thermogenesis falls rapidly over the first days or weeks of postnatal life.

Arousal from Hibernation

A limited number of mammalian species undergo hibernation during winter. Examples include hedgehogs, the European hamster (*Cricetus cricetus*) and various species of ground squirrel, e.g., Richardson's ground squirrel (*Spermophilus richardsonii*). Arousal from the deep torpor of hibernation, when body temperature may be as low as 4–6 °C, requires intense thermogenesis. Both shivering and nonshivering mechanisms are invoked in order rapidly (within a few hours) to raise the body temperature during arousal.

Thermogenic Effect of Food – Diet-Induced Thermogenesis

Nomenclature

Discussion of facultative thermogenesis in the context of nutrition is compounded by some terminological complexity and confusion. Feeding leads to a rapid increase in metabolic rate and different expressions have been used to describe this phenomenon. These include specific dynamic action (originally used in relation to the presumed particular effects of protein), thermic effect of food, postprandial thermogenesis, diet-induced thermogenesis, and the heat increment of feeding (Table 1). The term 'heat increment of feeding' is used primarily by those who work on farm animals, while the other expressions are employed in human nutrition and in connection with energy metabolism in laboratory animals.

In human and small-animal nutrition, diet-induced thermogenesis has become the customary expression. Even this, however, includes two distinct processes – obligatory and adaptive (or facultative) diet-induced thermogenesis – and involves both the short-term

Table 1 Synonyms for the thermogenic response to food

Name
Short-term
Specific dynamic action
Thermic effect of food
Postprandial thermogenesis
Heat increment of feeding
Diet-induced thermogenesis (obligatory and adaptive)
Chronic (overfeeding)
Luxoskonsumption
Diet-induced thermogenesis (adaptive)

response to a meal and the long-term response to overfeeding. It is emphasized that there are close parallels between adaptive diet-induced thermogenesis and nonshivering thermogenesis, particularly at a mechanistic level, and high levels of the former will reduce the need for the latter.

Short-Term Responses to Food

The consumption of food leads to a rapid increase in energy expenditure. Although this was originally considered to be a particular property of dietary protein, as noted above, it is now recognized that it occurs with the consumption of each of the macronutrients. The digestion, absorption, and initial metabolism of nutrients impose an energy cost on an organism, and include the cost of processes such as the synthesis of digestive enzymes.

The main component of obligatory diet-induced thermogenesis is the energy costs associated with the initial metabolic processing of substrates, once absorbed. Thus, glucose derived from dietary carbohydrate is initially stored as glycogen (in liver or muscle) or triacylglycerols (in liver or adipose tissue), following transport into tissues by either insulin-dependent or insulin-independent glucose transport. Obligatory diet-induced thermogenesis occurs in humans and in all other animals.

The energy costs associated with the storage of dietary carbohydrate as lipid amounts to approximately 25% of the energy potential of the substrate. In contrast, the deposition of dietary fat as triacylglycerol costs only some 5% of the energy associated with the substrate. Thus, calorie for calorie, the obligatory energy expenditure consequent upon the initial handling and storage of substrates as lipid will be lower with dietary fat than with carbohydrate (or protein).

The extent to which the acute response to food includes a facultative, energy-dissipative component is uncertain; there has, however, been much discussion of this possibility.

Responses to Chronic Overfeeding

In the early part of the twentieth century, German physiologists suggested that overfeeding leads to the stimulation of energy expenditure over and above obligatory diet-induced thermogenesis. The expression *Luxoskonsumption* was coined to describe the phenomenon, but adaptive diet-induced thermogenesis is now a more commonly used term. The basis of the concept is that overfeeding, particularly long-term, leads to the activation of mechanisms for the dissipation of the excess energy intake as heat. Thus, diet-induced thermogenesis is considered to represent a control mechanism in the regulation of energy balance.

Investigation of the metabolic responses to overfeeding in experimental animals has been greatly aided by the introduction of the cafeteria or supermarket diet. This diet, which provides a variety of human food items in addition to the standard laboratory animal chow, may induce up to twofold increases in voluntary food intake; the degree of hyperphagia depends, however, primarily on the variety and palatability of the foods. Energy balance measurements in rats and mice, in particular, have demonstrated that adaptive diet-induced thermogenesis is markedly activated in response to chronic overfeeding with a cafeteria diet. This reduces the extent to which fat is deposited in response to the overfeeding.

The phenomenon of adaptive diet-induced thermogenesis has at times been controversial. The controversy in part reflects the variability of the metabolic responses to overfeeding. In laboratory animals adaptive diet-induced thermogenesis appears to be strain-dependent, and to vary with other factors such as age, sex, and environmental temperature. Thus it is generally more substantial in young than in old animals, and in females than in males, and it is inhibited at thermoneutral temperatures. The process does not necessarily depend on the precise macronutrient content of the diet – it is the overfeeding rather than the nature of the macronutrients that is important. None the less, adaptive diet-induced thermogenesis has been widely noted in animals fed low-protein diets, and it is suggested that high-fat diets rich in polyunsaturated fatty acids are more thermogenic than those based on saturated and monounsaturated fatty acids.

The extent to which adaptive diet-induced thermogenesis is present, and a component of the regulation of energy balance, in humans is a matter of continuing debate. While some evidence in support of the concept is available, none is currently considered to be convincing.

Obesity

Obesity has been a major focus in studies on the regulation of energy balance, and some emphasis has been given to the hypothesis that obesity results, at least in part, from thermogenic abnormalities. Some animal models of obesity (e.g., the obese *ob/ob* mouse and the diabetic *db/db* mouse) exhibit a reduced capacity for nonshivering thermogenesis and under normal environmental conditions have a decreased energy expenditure on this process. Reduced adaptive diet-induced thermogenesis is a general characteristic of obesity in laboratory animals, and there is little doubt that this abnormality is significant in the energetics of the development of the obese condition, involving the recently discovered hormone leptin.

The situation is much more problematic in humans, and there is little convincing evidence to implicate reduced facultative thermogenesis (dietary or thermoregulatory) as an important factor in the development of human obesity. A reduction in diet-induced thermogenesis in response to nutrients is, however, frequently observed in the obese, and this appears to relate, at least in part, to insulin resistance.

Despite the absence of substantive evidence for major thermogenic abnormalities in human obesity, drugs which stimulate this component of energy expenditure have been the focus of continuing interest to the pharmaceutical industry as potential anti-obesity agents. Such drugs are primarily β-adrenoceptor agonists, selective for the β_3-adrenoceptor; however, none has yet progressed beyond the research and development phase.

Recently, the concept of nonexercise activity thermogenesis (NEAT) has been introduced and this proposes that small movements such as fidgeting and the maintenance of posture may contribute significantly to energy expenditure in humans. The absence of NEAT in response to overfeeding has been implicated as a factor in the etiology of obesity.

Biochemical Basis of Facultative Thermogenesis

A number of biochemical mechanisms have been proposed for the generation of heat by faculative thermogenesis – both nonshivering and adaptive diet-induced forms. These mechanisms are associated with a variety of tissues. Shivering is, of course, localized in skeletal muscle. Mechanisms for facultative thermogenesis can be divided into those in which there is a direct hydrolysis of adenosine triphosphate (ATP), and those in which the synthesis and hydrolysis of ATP are bypassed.

Thermogenic mechanisms where ATP is hydrolyzed include protein turnover, the pumping of Na^+ across the plasma membrane, and various substrate (or 'futile') cycles in specific metabolic pathways (Table 2). Examples of substrate cycles include the triacylglycerol/free fatty acid reesterification cycle which can occur both within and between tissues, and the fructose-6-phosphate/fructose-1,6-bisphosphate cycle in the glycolytic and gluconeogenic pathways (when both occur within the same tissue). Although there is good evidence that these mechanisms are thermogenic, individually their contribution to overall heat production is small – with the exception of highly specific situations such as in the warming of bumble bee flight muscle in cold environments. The simultaneous activation of a number of substrate cycles may, however, lead to the production of considerable amounts of heat.

Brown Adipose Tissue

The best documented and most widely recognized thermogenic mechanism is one in which the synthesis and hydrolysis of ATP are bypassed, namely the proton conductance pathway in brown adipose tissue. In this process the proton gradient that is normally established across the mitochondrial inner membrane during the oxidation of substrates is dissipated as heat rather than being coupled to the synthesis of ATP, in contrast to what generally occurs in mitochondria from other tissues.

The proton conductance pathway is regulated by a tissue-specific mitochondrial protein, termed uncoupling protein or now uncoupling protein-1 (it has also been called thermogenin), with a molecular weight of

Table 2 Putative biochemical mechanisms of facultative thermogenesis

Mechanism
Pumping of Na^+ across plasma membranes
Protein turnover
Mitochondrial Ca^+ cycle
α-Glycerophosphate shuttle
Substrate (or 'futile' cycles), e.g., fructose-6-phosphate/fructose-1,6-bisphosphate; glucose/glucose-6-phosphate; triacylglycerol/free fatty acid reesterification
'Hot pipes' in the microvasculature[a]
Mitochondrial proton conductance pathway in brown adipose tissue[b]
Mitochondrial uncoupling through UCP2 and UCP3[c]

[a]The precise biochemical mechanism involved in this process is unclear.
[b]Exclusive to brown adipose tissue; the other mechanisms may occur in a number of tissues, particularly skeletal muscle.
[c]Whether uncoupling protein-2 (UCP2) and UCP3 'uncouple' is a matter of debate.

32 kDa. The amount of this protein, and its activity, varies according to the requirements for thermogenesis in an animal. Brown adipose tissue is particularly well-developed in small animals, in the newborn of a number of mammalian species (including humans and ruminants), and in hibernators. It is also prominent in small, cold-adapted animals and in rodents exhibiting adaptive diet-induced thermogenesis where a major activation takes place.

The proton conductance pathway through uncoupling protein-1 is unique to brown adipose tissue, and this mechanism, together with the tissue itself, has been the major focus of studies on thermogenic systems in recent years. A recent major development is the discovery that there are a family of mitochondrial uncoupling proteins, not just uncoupling protein-1. Uncoupling protein-2 exhibits a wide tissue distribution, being found even in the brain, but it is particularly evident in white adipose tissue. Uncoupling protein-3 is located primarily in skeletal muscle.

The initial view was that these new uncoupling proteins uncouple mitochondria in a manner generally similar to uncoupling protein-1. However, it is far from clear that this is the case and it is argued that their role is more concerned with fat oxidation than energy dissipation *per se*. This would be consistent with the fact that brown adipose tissue expresses the genes encoding uncoupling protein 1–3; a thermogenic role for 2 and 3 in the presence of uncoupling protein-1 would seem superfluous.

Neural and Hormonal Control of Thermogenesis

Catecholamines are the main stimulus for facultative thermogenesis, both the dietary and nonshivering forms. Heat production can be stimulated by norepinephrine (noradrenaline) or epinephrine (adrenaline). Brown adipose tissue is highly innervated by sympathetic nerves, and norepinephrine secreted by them plays a central role in both the acute and chronic regulation of the thermogenic process.

The thyroid hormones, thyroxine (T_4) and triiodothyronine (T_3), have long been recognized to play a permissive role in thermogenesis. They can themselves also lead to a direct stimulation of heat production, and have in the past been used for the treatment of obesity. In brown adipose tissue, T_3 is generated endogenously through the presence of a type II iodothyronine 5′-deiodinase enzyme, and is involved in the regulation of the expression of the gene coding for the mitochondrial uncoupling protein.

Insulin also plays a significant role in facultative heat production. Diabetic animals have an impaired capacity for both nonshivering thermogenesis and adaptive diet-induced thermogenesis. Insulin is considered to stimulate thermogenesis through a central activation of the sympathetic nervous system, and by a direct effect on peripheral tissues. The uncoupling protein content of brown adipose tissue is regulated by insulin, and this in part involves an interaction with the sympathetic nervous system. Insulin resistance impairs the thermogenic response to food in both animals and in humans. The acute activation of brown adipose tissue in response to cold is also inhibited by insulin resistance in the tissue.

Glossary

Brown Adipose Tissue Form of adipose tissue specialized for the generation of heat, heat being the primary product of its metabolism. Heat is generated in brown adipose tissue through the presence of the tissue-specific uncoupling protein-1.

Diet-Induced Thermogenesis Currently preferred term to describe the increase in heat production associated with feeding. Relates to both short-term responses to a meal, and to chronic overfeeding; divided into obligatory and adaptive components.

Nonshivering Thermogenesis Heat generated specifically for thermoregulatory purposes, by mechanisms that do not involve muscle contraction.

Thermogenesis Generic term generally used to describe facultative heat production.

Uncoupling Proteins Family of mitochondrial proteins of molecular weight 32 kDa; the first such protein, uncoupling protein-1, is responsible for the generation of heat in brown adipose tissue. Uncoupling protein-2 is expressed particularly in white adipose tissue while uncoupling protein-3 is found particularly in skeletal muscle.

See also: **Adipose Tissue**: Structure and Function of Brown Adipose Tissue; **Diabetes Mellitus**: Etiology; Chemical Pathology; **Energy**: Energy Expenditure and Energy Balance; **Hormones**: Adrenal Hormones; Thyroid Hormones; **Obesity**: Etiology and Diagnosis

Further Reading

Girardier L and Stock MJ (eds) (1983) *Mammalian Thermogenesis*. London: Chapman and Hall.

Guy-Grand B and Ailhaud G (1999) *Progress in Obesity Research: 8*. London: John Libbey.

Himms-Hagen J and Ricquier D (1998) Brown adipose tissue. In: Bray GA, Bouchard C and James WPT (eds) *Handbook of Obesity*, pp. 415–441. New York: Marcel Dekker.

Levine JA, Eberhardt NL and Jensen MD (1999) Role of nonexercise activity thermogenesis in resistance to fat gain in humans. *Science* 283: 212–214.

Milton AS (ed.) (1994) *Temperature Regulation: Recent Physiological and Pharmacological Advances*. Basel: Birkhäuser.

Stock MJ and Rothwell NJ (1982) *Obesity and Leanness – Basic Aspects*. London: John Libbey.

Trayhurn P and Nicholls DG (eds) (1986) *Brown Adipose Tissue*. London: Edward Arnold.

THIAMIN

Contents
Properties and Determination
Physiology

Properties and Determination

Y Egi, Suzugamine Women's College, Hiroshima, Japan
T Kawasaki, Hiroshima University, School of Medicine, Hiroshima, Japan

Introduction

Thiamin (vitamin B_1), the antineuritic or antiberiberi factor, was first isolated from rice bran. Thiamin is distributed widely in natural materials in the forms of free thiamin and its mono- (TMP), di- (TDP), and triphosphate (TTP) esters. The ubiquitous distribution of thiamin in the plant and animal kingdoms is due to the presence of the coenzyme form, TDP, which plays an important role in carbohydrate metabolism throughout biological organisms. (*See* **Coenzymes**.)

This article concentrates on the physical and chemical properties of the vitamin, its occurrence in foods, its use as a food additive, and the problems associated with its analytical determination in biological and food samples.

The structures of thiamin and its phosphate esters found in living organisms are shown in **Figure 1**.

Physical and Chemical Properties

The physical properties of thiamin hydrochloride are given in **Table 1**. Thiamin hydrochloride crystallizes, usually as the hemihydrate. On exposure to air of average humidity, thiamin hydrochloride absorbs water in the ratio of 1 mol of water to 1 mol of thiamin. The water can be removed by heating at 100 °C or in a vacuum over sulfuric acid or phosphorus pentoxide. In the dry form the compound is stable at 100 °C.

Thiamin phosphoric acid esters have the following empirical formulae: TMP chloride, $C_{12}H_{18}N_4O_4PSC1$ (mol. wt 380.78); TDP chloride, $C_{12}H_{19}N_4\ O_7P_2SC1$ (mol. wt 460.76); and TTP, $C_{12}H_{19}N_4\ O_{10}P_3S$ (mol. wt 504.28). All these phosphate esters are freely soluble in water and are stable in the dry state for several months when stored at a low temperature in the dark.

Table 1 Physical properties of thiamin hydrochloride

Property	*Characteristics*
Formula	$C_{12}H_{17}N_4OSC1HC1$ Mol. wt 337.28
Appearance	White, slight thiazole order, crystalline solid
Crystal form	Monoclinic plates in rosette-like clusters
Melting point	248 °C
pH	3.58 (1 mg ml^{-1}); 3.13 (10 mg ml^{-1})
Ultraviolet maxima	247 nm (pH 3.0); 235 and 267 nm (pH 5.0)
Solubility	1 g in 1 ml water, 18 ml glycerol, 100 ml 95% alcohol, 315 ml absolute alcohol; more soluble in methanol; soluble in propylene glycol. Practically insoluble in ether, benzene, hexane, and chloroform

From Budavari S (ed.) (1996) *The Merck Index. An Encyclopedia of Chemicals, Drugs, and Biologicals*, 12th edn. Rahway, NJ: Merck.

Thiamin is readily split into the pyrimidine and thiazole moieties by sulfite treatment at pH 6.0 or above.

In strong alkaline solution thiamin is oxidized by oxidants (potassium ferricyanide, cyanogen bromide, mercuric chloride, and others) to thiochrome (Figure 1). Thiochrome is a highly fluorescent compound above pH 8, and thiamin phosphate esters are also quantitatively converted to thiochrome phosphate esters without affecting the phosphate bonds. All these compounds have identical excitation maxima, at 375 nm, and very similar fluorescence maxima, at 432–435 nm. This is the principle for the most common and sensitive assay procedures for thiamin as well as its phosphate esters.

Hydroxyethylthiamin (Figure 1) exists in living organisms in the form of hydroxyethyl-TDP. It is converted to the corresponding thiochrome derivative by alkaline ferricyanide oxidation but not by cyanogen bromide oxidation. This difference in the oxidation property is used for the assay of hydroxyethylthiamin.

In aqueous solution below pH 5, thiamin is quite stable to heat or even sterilization at 110 °C. At a pH of 5.5 or higher, it is rapidly destroyed by autoclaving, and at pH 7 or higher by boiling or even storing at room temperature.

In solution, TDP is unstable and partially decomposes to TMP and/or thiamin when stored for several months at pH 5 and 37 °C, but stable at pH 2–5 and 0 °C. TTP in aqueous solution is stable for at least 6 months when stored at − 80 °C. In 0.1 mol l^{-1} hydrochloric acid, TTP is decomposed quantitatively to TMP by boiling for 7 min at 100 °C, indicating

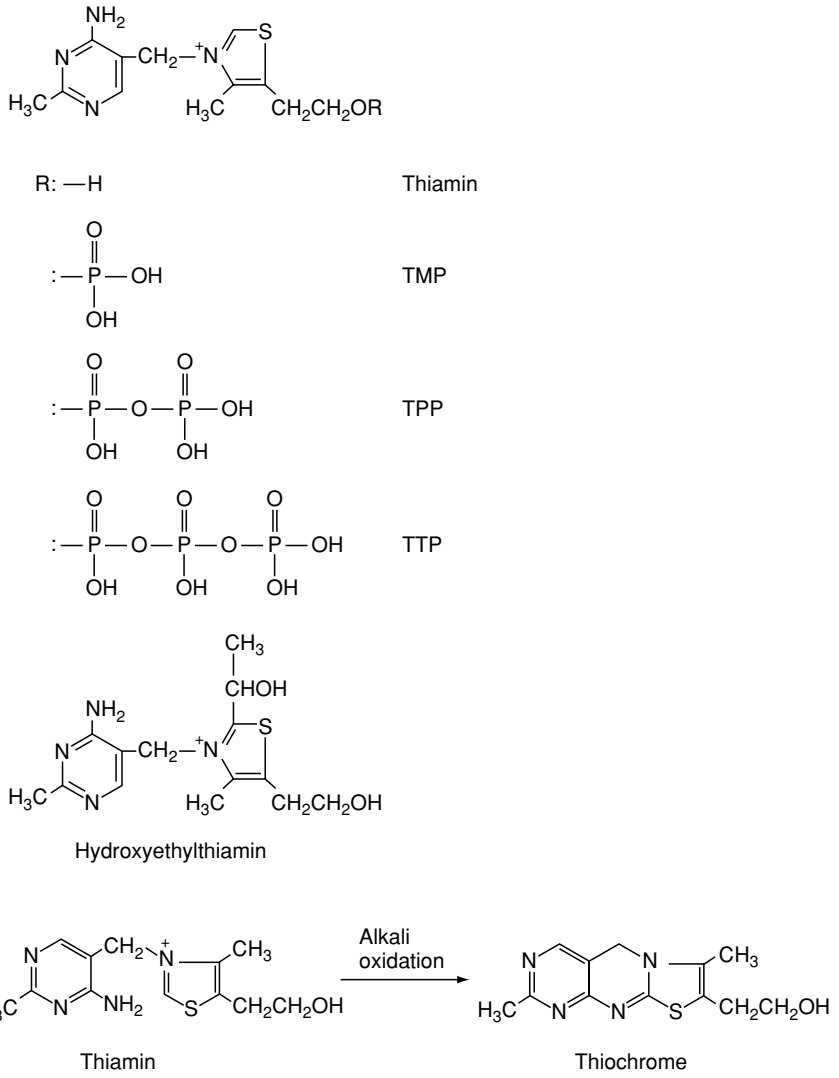

Figure 1 Structures of thiamin (vitamin B_1) and its related compounds.

the nature of the high-energy bond of its γ- and β-phosphates.

Thiamin-destroying or inactivating enzymes have been discovered in a variety of foods. Two thiaminases are involved. Thiaminase I catalyzes a base exchange between the thiazole moiety and another base and is found in fresh fish, shellfish, ferns, and some bacteria. Thiaminase II hydrolyzes thiamin to its pyrimidine and thiazole moieties and is found in certain bacteria.

In addition to thiaminases, certain plants have been shown to contain compounds which react with thiamin *in vitro* to produce thiochrome-negative products. One such product is thiamin disulfide. Such compounds in foods, which result in so-called thiamin inactivation, include caffeic acid, tannic acid, a variety of polyphenols, and some flavonoids (quercetin and rutin). The exact mechanisms of the inactivation reaction have not yet been elucidated. It appears that oxidative processes are involved, since the presence of oxygen increases the amount of thiamin inactivated.

It has been discovered that the treatment of thiamin with an extract of garlic containing allicin forms thiamin allyl disulfide, which is thiochrome-negative, but very active biologically. A variety of thiamin alkyl disulfides has since been synthesized and studied, and is now available commercially. These forms of thiamin are more fat-soluble and are therefore more rapidly absorbed than thiamin hydrochloride, and result in higher blood thiamin levels when therapeutically administered.

Occurrence in Foods

In most animal tissues, over 90% of the thiamin occurs in TMP, TDP, and TTP. The predominant form (80–85%) is TDP, the active coenzyme form. The exceptions are pig skeletal muscle and chicken skeletal white muscle, in which TTP exists in 70–80% of total thiamin (Figure 2). The most abundant form in plant tissues is free thiamin.

The content of thiamin in foods is relatively low. Practically no thiamin is contained in high-fat products (e.g., vegetable oil) and refined products (e.g., sugar). It is also relatively low in green vegetables, fruits, and seafoods. The thiamin content in a large variety of natural and processed foods has been listed in publications from the US Department of Agriculture (USA), Medical Research Council (UK) and Resources Council, Science and Technology Agency (Japan).

Table 2 lists the expected concentration in thiamin-rich foods. Refer to individual foods.

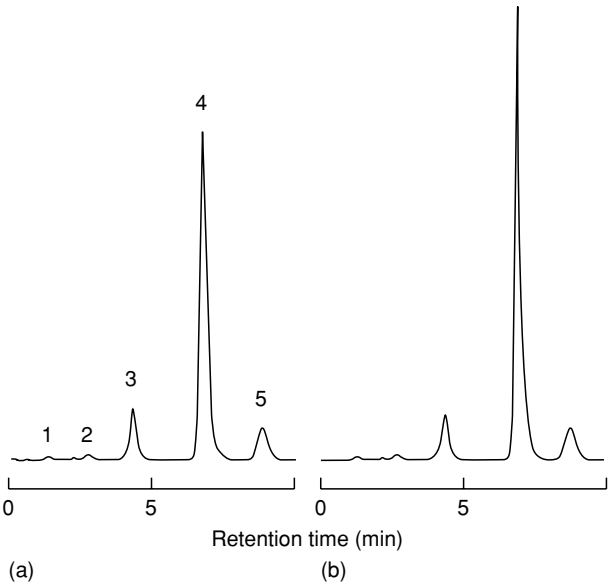

Figure 2 A typical chromatogram of fresh pig skeletal muscle extract by precolumn derivatization high-performance liquid chromatography (HPLC). (a) Skeletal muscle extract; (b) as (a), plus 1 pmol of authentic thiamin triphosphate (TTP) which was converted into thiochrome triphosphate and then added to the oxidized sample. The recovery rate of added TTP was 100%. 1, Thiamin; 2, thiamin monophosphate (TMP); 3, thiamin diphosphate (TDP); 4, TTP; 5, unknown, but possibly thiamin tetraphosphate.

In addition, average thiamin contents (μg per 100 g) in the following categorized groups of foods are as follows: dried beans, 680; nuts, 560; whole-grain cereals, 370; organ meats (liver, heart, kidney), 100; leaf and stem vegetables, 70; milk, 40; fruits, 30; pork muscle, 600–800; common white fish (cod, flounder, haddock, halibut), 50–90; hen's whole egg, 170; egg yolk, 500. Interesting foods of high thiamin content (μg per 100 g) are dried brewers' yeast (1800), wheat germ (2000), and tampala (a spinach-like vegetable) leaves (1600).

Losses During Food Processing and Storage

Factors affecting the survival of thiamin in the final food products after harvesting, handling, and processing the foods include pH, temperature, solubility, oxidation, and radiation. Among these factors, the effect of processing on the thiamin content has been extensively reviewed by several authors. A brief summary of the findings is given here.

Thiamin losses in dehydrated products, such as cereal grains, are very low if the materials are kept dry, and thiamin undergoes only limited destruction in the commercial refrigeration of meats and

Table 2 Top sources of thiamin

Top sources	*Thiamin (mg 100 g^{-1})*
Yeast, brewers', debittered	15.61
Yeast, torula	14.01
Yeast, brewers'	12.12
Sunflower seed, flour, partially defatted	3.60
Yeast, bakers', dry	2.33
Rice bran	2.26
Wheat–soy blend (WSB)/straight grade, wheat flour	2.02
Wheat germ	2.02
Sunflower seed kernels, dry, hulled	1.96
Rice polish	1.84
Wheat germ, toasted	1.65
Wheat–soy blend (WSB)/bulgur, flour	1.49
Pinenuts, piñon	1.28
Coriander leaf, dried	1.25
Cottonseed flour	1.21
Cornflakes, with added nutrients	1.20
Peanuts, raw with skins	1.14
Pork, fresh, loin, lean, boiled	1.13
Safflower seed, flour, partially defatted	1.12
Soybean flour, defatted	1.09
Alfalfa seeds	1.08
Sesame seed	0.98
Bacon, Canadian, broiled or fried, drained	0.92
Sausage, links or bulk, cooked	0.79
Wheat bran	0.72
Kidneys, beef, braised	0.67
Wheat flour, all-purpose, enriched	0.64
Rye flour, dark	0.61
Nuts, mixed, shelled	0.59
Wheat flour, whole (from hard wheats)	0.55
Pork, cured, canned, ham	0.53
Cornmeal, degermed, enriched	0.44
Rice, white, enriched, raw	0.44
Soybean sprouts, cooked	0.42
Bread, white, enriched	0.40
Peas, green, immature, boiled, drained	0.28
Turkey, hamloaf	0.27
Beef liver	0.25
Luncheon meat, salami, cooked	0.25

From Ensminger AH (ed.) (1994) *Foods and Nutrition Encyclopedia*, 2nd edn, pp. 2102–2108. Boca Raton: CRC Press.

vegetables. Foods can be stored frozen for long periods without significant loss of thiamin, but losses are significant on thawing. Losses from frozen muscle meats are 20–40% during storage for 2–8 months.

On cooking and canning of meats, losses of thiamin (25–85%) strongly depend on the processing stresses encountered and type of meat product. Cooking by microwaves leads to much lower losses. Losses on cooking of vegetables vary from 0% to 60%. Vigorous boiling or other types of agitation result in greater losses of thiamin than simmering at the same temperature or pressure and steam cooking. Losses are caused by degradation and by leaching. (*See* **Canning**: Quality Changes During Canning; **Cooking**: Domestic Techniques.)

Roasting of beef and pork involves thiamin losses of 36–53%. Ionizing radiation used during sterilization destroys 53–88% of thiamin in meats, yet, in certain poultry, fish, and vegetables, losses are only 5–37%. Baking of bread leads to losses of 5–35%. In the processing of milk, losses of thiamin are 9–20% by pasteurization, 30–50% by sterilization, 10–15% by drying, and 40% by condensing and subsequent canning. (*See* **Irradiation of Foods**: Applications.)

The milling of cereal grains, wheat flour extraction of less than 85%, and the degerming of corn cause a marked drop in thiamin. Parboiling of rice and drying before milling conserve the majority of the thiamin. (*See* **Milling**: Characteristics of Milled Products.)

Uses in Food Fortification

Thiamin hydrochloride is the most common additive. (*See* **Food Fortification**.)

Thiamin (125 mg per 100 g) is added to polished rice and (1.5 mg per 100 g) to pressed barley in Japan, mostly in special cases, such as for use in school lunches.

The recommended daily allowance of thiamin is 0.5 mg per 1000 kcal. The continuing need for thiamin fortification of processed cereals has been proposed in the UK (1986), in order to meet the adult thiamin requirement, which has been determined biochemically.

Human thiamin deficiency, called beriberi, has been largely eliminated in those countries where fortification of staple foods (rice, cereal, dairy products) is practiced, but it is still prevalent in those countries where unfortified foods are used. However, even in western countries it is becoming evident that subclinical thiamin deficiency is more widespread than previously realized, especially among infants and the elderly, and other groups who consume snacks and highly refined, processed foods. Thiamin fortification of rice, dairy products and cereals is also recommended for these 'at-risk' groups and for athletes whose metabolic demands are greater.

Analytical Procedures

Alkaline oxidation of thiamin by either ferricyanide or cyanogen bromide gives thiochrome, an intense blue fluorescence compound. This principle is the basis for the official standard method of the Association of Official Analytical Chemists (AOAC). However, high-performance liquid chromatography (HPLC) is most frequently used at the present time for the assay of thiamin in foods. Microbiological

assays are still also widely used for both foods and animal feeds. (*See* **Chromatography**: High-performance Liquid Chromatography.)

An outline of the manual fluorimetric AOAC method follows. Thiamin in foods is protein-bound in both animal and plant tissues, although it occurs principally as the diphosphate in the former and as the free form in the latter. It is therefore necessary to extract thiamin and its phosphates from food proteins.

This is carried out by boiling the finely ground or homogenized sample in 0.1 mol l^{-1} hydrochloric acid or 0.1 mol l^{-1} sulfuric acid. The extract is then neutralized to pH 4–4.5 and treated with an enzyme preparation exhibiting phosphatase activity (Taka-diastase, Mylase P, or Clarase). After filtration or centrifugation, free thiamin thus obtained is purified through a column of cation exchange-type silica (Permutit T, Decalso F, or Zepolite S/E). After washing the column several times with almost boiling water, the thiamin is eluted with almost boiling acid potassium chloride solution.

An aliquot of the above eluate is oxidized by alkaline ferricyanide or alkaline cyanogen bromide solution and another aliquot is mixed with an alkali solution (blank). Thiochrome thus obtained is extracted into isobutanol. Thiochrome fluorescence intensity is measured spectrofluormetrically at an excitation wavelength of 375 nm and an emission wavelength of 435 nm. The amount of thiamin extracted is calculated from the standard curve obtained with thiamin standard solution treated in the same manner as above. Fortified foods and pharmaceuticals can be analysed for thiamine with reduced sample preparation.

Chromatographic Methods

Recent advances in HPLC techniques allow the determination of thiamin content in foods more rapidly, accurately, sensitively, and reproducibly than the standard AOAC method. The purification step of the extracts can usually be omitted in the HPLC procedure. In addition, thiamin and its phosphate esters can be determined separately by the HPLC method when the appropriate extraction procedure is employed. Furthermore, thiamin, riboflavin, pyridoxine, or niacin in foods can be simultaneously assayed when appropriate detection systems are used. This procedure has been successfully used for quantifying these vitamins in a wide range of foods, including cereal products, raw meat, processed meats, fruits, and vegetables.

Samples are treated, for example, by the AOAC method and then subjected to enzymatic hydrolysis. The filtrate or the supernatant after centrifugation can be directly analyzed by a reversed-phase column with ion-pairing chromatography. A wide range of mobile phases has been employed, but those based on octane sulfonate (as ion pair) in aqueous methanol or acetonitrile have proved to be generally successful. Detection is performed spectrophotometrically at 254 nm, which gives a detection limit on injection of 30 ng (90 pmol) of thiamin. This procedure is suitable for food containing a relatively large amount of thiamin, especially for fortified rice or cereals. (*See* **Spectroscopy**: Overview.)

An HPLC system equipped with fluorometric detection is suitable to quantify much smaller amounts of thiamin in foods with a detection limit of 5–50 fmol thiamin. Both precolumn derivatization and postcolumn derivatization procedures of thiamin into thiochrome are used.

When the extraction procedure is carried out with cold trichloroacetic acid or perchloric acid, thiamin and its phosphate esters are extracted intact and can be individually measured. Both straight-phase and reversed-phase columns are used in the HPLC method. In the former system, thiamin, TMP, TDP, and TTP are eluted and detected in that order, and in the latter system the elution order is reversed. Total thiamin is then calculated as a sum of thiamin and its phosphates. This procedure can avoid the enzymatic hydrolysis step. This HPLC method of determining thiamin and its phosphates has been successfully used to quantify the thiamin level not only in animal tissues, but also in human blood or serum to assess thiamin status in humans.

The chromatographic profile of thiamin and its phosphates in fresh pig skeletal muscle, obtained by the precolumn derivatization method, is shown in Figure 2. The analytical system consisted of LiChrosorb NH_2 (150 × 4.6 mm inner diameter (ID), 5 μm particle size) as the stationary phase, and acetonitrile–90 mmol l^{-1} potassium-phosphate buffer (pH 8.4) (60:40, v/v) as the mobile phase. Detection was carried out fluorometrically (excitation, 375 nm; emission, 430 nm) at room temperature.

Microbiological Methods

Microbiological methods are based on the nutritional requirement of a particular microorganism for the vitamin in question, in this case thiamin. A nutrient medium which is complete in all respects except for thiamin is prepared for the microorganism. The growth response for standard vitamin solutions is then compared with that achieved with extracts of the food sample, and hence the concentration of

thiamin in the sample is calculated. In most instances it is necessary to extract thiamin from foods under hydrolytic conditions and it may also be necessary to release bound forms with an enzymatic digestion, as in the standard thiochrome method.

Over the years, many microorganisms have been used in the assay of thiamin. Assays based on the use of *Lactobacillus fermentum* and *L. viridescens* are now widely employed as these microorganisms respond only to intact thiamin. Measurement of rate of growth is usually followed by nephelometry or measurement of turbidity, although production of acidity may also be monitored.

Microbiological methods are slowly being replaced by more rapid instrumental techniques, e.g., HPLC. However, they do provide simple, sensitive analyses, especially where a high degree of precision is not required.

See also: **Canning**: Quality Changes During Canning; **Chromatography**: High-performance Liquid Chromatography; **Coenzymes**; **Cooking**: Domestic Techniques; **Food Fortification**; **Irradiation of Foods**: Applications; **Milling**: Characteristics of Milled Products; **Spectroscopy**: Overview

Further Reading

Budavari S (ed.) (1996) *The Merck Index. An Encyclopedia of Chemicals, Drugs, and Biologicals*, 12th edn. Rahway. NJ: Merck.

Ensminger AH (ed.) (1994) *Foods and Nutrition Encyclopedia*, 2nd edn, pp. 2102–2108. Boca Raton: CRC Press.

Fayol V (1997) High-performance liquid chromatography determination of total thiamin in biological and food products. In: McCormik DB, Suttie JW and Wagner C (eds) *Methods in Enzymology*, vol. 279, pp. 57–66. San Diego: Academic Press.

Gregory JF III (1996) Vitamins. In: Fennema OR (ed.) *Food Chemistry*, 3rd edn, pp. 531–616. New York: Marcel Dekker.

Helrich K (1990) *Official Methods of Analysis of the Association of Official Analytical Chemists*, 15th edn, pp. 1049–1052. Washington: Association of Official Analytical Chemists.

Kamas E and Harris RS (1988) *Nutritional Evaluation of Food Processing*, 3rd edn. New York: Van Nostrand Reinhold.

Kawasaki T and Egi Y (2000) Vitamin B_1: thiamin. In: DeLeenheer A, Lambert W and Van Bocxlaer J (eds) *Modern Chromatographic Analysis of the Vitamins*, 3rd edn, pp. 365–389. New York: Marcel Dekker.

Machlin LJ (ed.) (1991) *Handbook of Vitamins*, 2nd edn. New York: Marcel Dekker.

Paul AA, Southgate DAT and Buss DH (1986) *McCance and Widdowson's The Composition of Foods: Supplementary Information*. London: Her Majesty's Stationery Office.

Resources Council. Science and Technology Agency. Japan (1998) *Standard Tables of Food Composition in Japan*, 5th edn. Tokyo: Ministry of Finance Printing Office.

US Department of Agriculture (1995) *Composition of Foods*. US Department of Agriculture. Human Nutrition Information Services, Agriculture Handbook 8. Washington, DC: US Government Printing Office.

Physiology

R Bitsch, Friedrich-Schiller-University, Jena, Germany

Background

The discovery of thiamin or vitamin B_1 as an essential nutrient has opened a new area in the field of nutrition research in two respects. First, the nomenclature 'vitamin,' used for the whole class of the later on discovered essential micronutrients of organic origin, was derived from a functional group (NH_2 group) of the molecule. Second, thiamine was the first nutritional factor that could be identified as a curative agent against a deficiency disease in animals.

At the end of the nineteenth century, the Dutch medical officer Ch. Eijkman discovered the beneficial effect of rice bran or unpolished rice, respectively, for the treatment of polyneuritis in chickens resembling beriberi in man, which had been previously induced by feeding polished rice. In 1911, the German-Polish chemist C. Funk isolated a crystalline substance from rice bran and regarded the amino function as the essential principle of the molecule. The exact chemical structure of the thiamin molecule was elucidated by R. Williams and A. Windaus. Because of the sulfur content of the molecule or its thiazol moiety combined with the amine function, the compound was designated as thiamin instead of the former name aneurin. Nevertheless, the classical form of thiamin deficiency, beriberi, mainly affects human beings and has been well known in the rice-eating countries of East Asia for centuries. Even 25 years after the ingenious work of Eijkman, who clarified the cause of this deleterious disease, the mortality rate from beriberi in industrialized countries such as Japan was still more than 25 000 per year.

Physiological Function

Like other water-soluble B-vitamins, thiamin is in its primary function coenzymatically active. The thiamin-dependent enzymes are involved in carbohydrate and energy metabolism (**Figure 1**). Key enzymes in

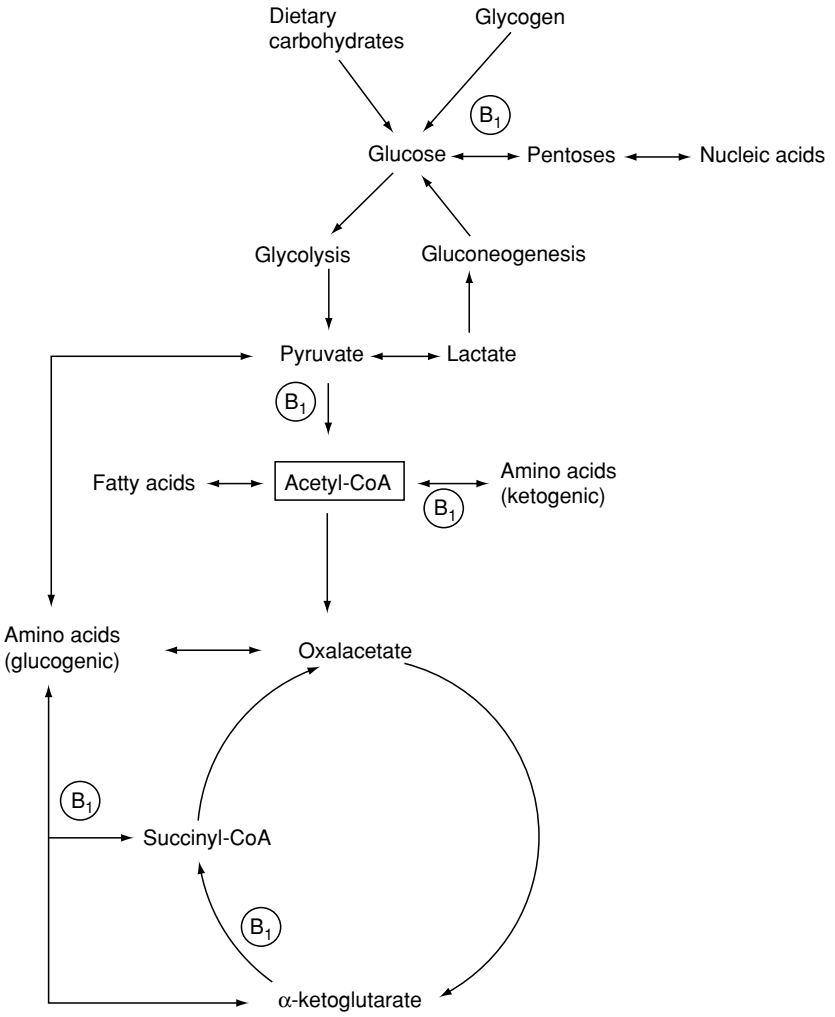

Figure 1 Metabolic pathways involving thiamin.

these metabolic pathways are mitochondrial pyruvate dehydrogenase (EC 1.2.4.1) linking glycolysis with the citric acid cycle, α-ketoglutarate dehydrogenase (KDH) (EC 1.2.4.2) within the citric acid cycle, and the cytosolic transketolase (EC 2.2.1.1) in the pentose phosphate shunt. Hence, it follows that the energy turnover of the organism is related to thiamin demand. As dietary carbohydrates are the predominant fuel for the rapid metabolizable energy of the organism, the thiamin requirement is often related merely to the carbohydrates ingested. KDH was identified as the rate-limiting enzyme of the cerebral glucose utilization. A low enzyme activity caused by a reduced intracellular TDP concentration results in lactacidosis, cerebral dysfunction of the energy metabolism, and neuronal cell degeneration. Additional thiamin-dependent enzymes are involved in the metabolism of branched-chain amino acids, e.g., leucin, isoleucin, and vallin. All enzyme reactions are characterized by transfer of an 'active aldehyde' from α-keto acids formed after previous decarboxylation to coenzyme A or an aldopentose.

The coenzymatically active form is thiamin diphosphate (TDP), also called thiamin pyrophosphate (TPP) or cocarboxylase, which is formed in the mucosal cell of the intestinal barrier with the aid of thiamin pyrophosphokinase (EC 2.7.6.2) in the presence of ATP. The latter enzyme is also found in mammalian liver and other tissues such as brain, nerves, and heart of mammals and birds. Further phosphorylated forms of thiamin, which are found in varying quantities in tissue, are thiamin monophosphate (TMP) and thiamin triphosphate (TTP). These derivatives are synthesized and metabolized according to the scheme in Figure 2.

Thiamin phosphorylation represents a partitioning process in a deeper compartment. Cell membranes are impermeable for the intracellularly active

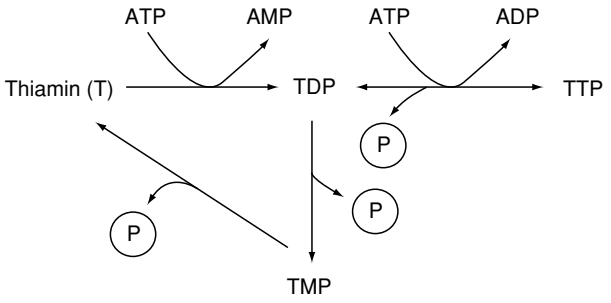

Figure 2 Interconversion of thiamin phosphates (according to Rindi, modified).

coenzyme TDP, which is released only after hydrolysis via TMP to thiamin. The total thiamin concentration of the blood cells as well as of the cerebrospinal fluid (CSF) has a nonlinear relation to the corresponding plasma concentrations. A linear increase in blood cells and CSF has been observed only with thiamin plasma levels above 10 nM ($3\,\mu g\,l^{-1}$). Otherwise, with extremely low plasma and CSF concentrations, a disproportionately high TDP content of erythrocytes and TMP content of the CSF is maintained for a longer period. Thus, the intracellular phosphorylation should be seen as a protecting mechanism, inhibiting rapid thiamin depletion by decreasing the membrane permeability.

After TDP, TTP is the next most phosphorylated metabolite, albeit the steady state concentrations in tissues are rather low. No coenzymatic function has been found for this metabolite. Several investigators have demonstrated in a variety of animals (rats, rabbits, cows, frogs, pigs) that TTP, together with a distinct part of TDP, is neurophysiologically active in the central and peripheral nervous system independent of the coenzyme function of the latter. After electrical stimulation of nerves or treatment with neuroactive drugs, e.g., acetylcholine, tetrodotoxin, or LSD, a release of the phosphorylated esters TDP and TTP concomitant with a shift to TMP and free thiamin has been observed. Furthermore, in patients with the deleterious 'subacute necrotizing encephalopathia,' or Leigh's disease, the TTP level in the brain was low and tended toward zero. It has been known for a long time that severe thiamin depletion of the mammalian central nervous system leads to encephalopathias, which are accompanied by regionally selective changes in neurotransmitter function. Though the exact function of TTP is not yet known, it plays an essential role in nerve excitation and transmission, possibly in gating the ion transport. From recent investigations, it has been argued that TDP might also be involved in the inhibition of nonenzymatic glycosylation processes. It is suggested that the emerging advanced glycosylation end products (AGE) are the reason for the segmental demyelinization observed in neurodegenerative diseases. Studies on cultivated cells have demonstrated that thiamin and TDP could prevent the AGE formation. In this regard, TDP is more effective than the standard inhibitor aminoguanidine that is normally used.

Absorption, Storage, and Excretion

Thiamin is absorbed from the gastrointestinal tract by a dual system, as has been shown for humans and rats. At low concentrations, it follows energy-dependent active transport with a saturation kinetics up to a maximal concentration of 2 μM. At higher concentrations, a simple diffusion process is predominant. However, the relative absorption decreases with increasing vitamin doses, so that with gram doses, only about 5% are incorporated.

A close association was assumed between thiamin transport and phosphorylation as well as Na^+ dependency. However, the coupling of thiamin transport with phosphorylation is now doubtful, and there is much evidence of a specific protein carrier being involved in the absorption, but the following intracellular phosphorylation seems to be a rate-limiting step in the active transport of thiamin across the epithelial cells of intestine. In contrast, in the guinea-pig intestine, thiamin seems to cross the brush-border membrane by simple diffusion throughout the concentration range from 0.06 to 10 μM.

The jejunum and ileum have been found to be the main sites of thiamin absorption in rats with a maximal rate in the proximal 22 cm of the small intestine. This specificity seems to resemble that of other animals as well as humans.

After uptake from the intestinal tract, thiamin is distributed rapidly into the organs and tissues according to the requirements. In humans and animals, the heart contains the highest thiamin concentration (3–8 $\mu g\,g^{-1}$ fresh weight), followed by the kidney (2.4–6 $\mu g\,g^{-1}$ fresh weight) and liver (2–7.5 $\mu g\ g^{-1}$ fresh weight), and the brain (1.4–4.4 $\mu g\,g^{-1}$ fresh weight). Lesser amounts are found in other tissues. Whole blood contains 0.05–0.12 μg of thiamin per milliliter, about 90% of this being found in the corpuscular constituents. The total storage capacity of man is estimated to be 25–30 mg of thiamin, 40% of this being in muscles. Interestingly, pig muscle has a relatively high thiamin content (8–12 $\mu g\,g^{-1}$) because of feeding conditions, higher than that in muscles of other mammals (1–2 $\mu g\,g^{-1}$) and pig liver (2–4 $\mu g\,g^{-1}$). The thiamin contents of several plant and animal foods are given in Table 1.

Table 1 Thiamin content of plant and animal foods (mg per 100 g fresh weight)[a]

Content	Foods	
1.0–2.0	Whole wheat and rye	Plant foods
0.9	Para nuts	
0.7–0.8	Soy flour	
0.6–0.7	White bran	
0.3–0.5	Whole cereal products, lentils, oats, unpolished rice, lima beans, whole corn meal, cashew nuts, pea nuts, hazel and walnuts, mung beans	
0.1–0.2	Parboiled rice, sorghum, beans and peas cooked, cauliflower, potatoes, fennel, broccoli cooked, asparagus, soy sprouts, mushroom	
< 0.1	Fruits, milk products, white flour products	
0.9	Loin pork	Animal foods
0.3–0.5	Liver, kidney (calves, beef, pork), duck	
0.1–0.2	Lean meat (calves, lamb, beef), eel, tuna, salmon, mackerel, plaice and flounder, oysters	

[a]Modified from Souci SW, Fachmann W and Kraut H (2000) *Food Composition and Nutrition Tables*, 6th revised edn. Boca Raton, FL: CRC Press/Medpharm Scientific.

As with other water-soluble vitamins, thiamin excretion is dependent on intake. Within the physiological range, most of it is excreted by urine. Besides free thiamin and small amounts of TMP, numerous metabolites can be detected as minor excretion products such as thiazol, pyramin, and methylthiazol acetic acid.

After ingestion of therapeutic dosages, the amount of thiamin excreted via bile increases proportionally, as does the unabsorbed part in feces.

Lipid-soluble Thiamin Derivatives

Under relatively mild conditions (e.g., in the gastrointestinal lumen) thiamin reacts with allicin, the active principle of garlic and onions, to form a lipid-soluble compound, called allithiamin, that is able to penetrate through membranes more easily than the water-soluble salts (Figure 3). Additional *S*-alkyl derivatives have been identified in cruciferous plants, and several lipid-soluble derivatives have been synthesized with a better stability than that of the spontaneously formed allithiamine (*S*-benzoyl-, *S*-tetrahydrofurfuryl analogs being the most important).

Oral administration of these lipid-soluble analogs results in higher thiamin blood levels than with equimolar thiamin doses, because the allithiamins from the gut lumen are preferentially absorbed by passive diffusion and rapidly converted to the active vitamin by reducing SH-compounds. These allithiamins are because of their improved bioavailability suitable transfer substances with which to establish high thiamin levels in target tissues comparable with drug targeting. The elevated blood level results in an improved vitamin status, as measured by enzymatic criteria.

Thiamin Antagonists

Synthetic Antivitamins

Like most of the B vitamins, thiamin has a rather high structural specificity, i.e., minor modifications of the molecule lead to a decrease or a total loss of

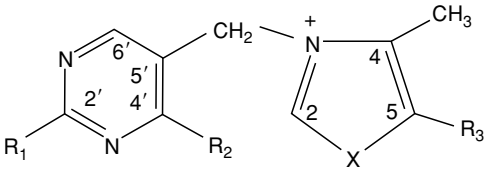

	R_1	R_2	R_3	X
Thiaminium ion	Methyl	Amino	Ethanol	Sulfur
Ethylthiamin	Ethyl	Amino	Ethanol	Sulfur
N-propythiamin	N-propyl	Amino	Ethanol	Sulfur
Oxythiamin	Methyl	Hydroxy	Ethanol	Sulfur
Pyrithiamin	Methyl	Amino	Ethanol	Ethene
Amprolium	N-propyl	Amino	Hydrogen	Ethene
Allithiamin	Methyl	Amino	Ethanol	S-alkyl (thiazol oxidatively splitted at C_2)

Figure 3 Thiamin derivatives and antagonists.

physiological activity or even to antagonistic effects. Of the numerous analogs synthesized so far, only the 2-ethyl- and the 2-*n*-propyl homologs in the pyrimidine moiety of the molecule show any biological activity comparable with that of thiamin. For example, when the 4-amino group is replaced by a hydroxyl function, oxythiamin, one of the most potent thiamin antagonists, is obtained. Another prominent antivitamin is pyrithiamin, which contains an ethylene group instead of sulfur in the thiazolium ring. Some thiamin antagonists have been found to exhibit antiprotozoal activity, as in the case of amprolium, which is effective as a prophylactic agent for coccidiosis of fowl. Some of the modified thiamin derivatives are shown in Figure 3.

The use of antagonists results in rapid and severe deficiency symptoms in humans and animals. Amprolium is already known to affect the absorption process already, and this has been suggested as the mechanism for its anticoccidial action in fowls. Other antivitamins appear to function at the conzyme level, i.e., they bind to the apoenzyme, but inhibit the following enzymatic reaction. Nevertheless, there are several differences in the action on neurological symptoms. Only pyrithiamin (and not oxythiamin) can block the action potential of a single myelinated nerve fiber. It is also well known that only pyrithiamin produces polyneuritis. In addition, differences in neurophysiological action are caused by the different abilities of both antagonists to cross the blood–brain barrier, which has only been verified in the case of pyrithiamin. However, it can be shown that the amino group of the thiamin molecule is essential for the active transport of the vitamin from the rat small intestine and the formation of thiamin pyrophosphate.

Thiaminases and other Antithiamin Factors

Two thiamin-cleaving enzymes have been identified, called thiaminase I and thiaminase II.

Thiaminase I is found in shellfish, clams (but not oysters), some freshwater fish viscera, crustacea, and certain ferns, but very few higher plants. Also, certain species of *Bacillus* and *Clostridium*, which are components of the human and animal intestinal flora, have been found to produce this enzyme. The enzyme catalyzes an exchange reaction, in which the thiazol moiety of the molecule is displaced by another N-containing base or a SH-compound. The effect of this compound was first observed when silver foxes were fed raw fish waste, resulting in a deleterious thiamin deficiency disease called Chastek paralysis. Thiaminase I may also be produced by the rumen microflora of ruminants or by plants, and, in the presence of suitable cosubstrates, e.g., niacin, or pyridoxine, and certain antihelmintics, seems to be responsible for the ruminant CNS disorder, polioencephalomalacia.

Thiaminase II is of bacterial origin (predominantly *Bacillus, Candida*, and *Oospora*) and breaks down the free vitamin, but not the thiamin pyrophosphate, into pyrimidin and thiazol components. More prevalent are thermostable thiamin-inactivating factors of plant origin, e.g., polyphenolic substances such as flavonoids and catechol derivatives in fermented tea, ferns, sweet potatoes, and betel nuts, and, in small quantities, in other leaves, fruits, and roots. These can decompose the thiazol component of the vitamin, accelerate the oxidation to the disulfide form, or form unabsorbable adducts with thiamin. Rats fed on food high in polyphenolics have been shown to develop deficiency symptoms as a result of a marked decrease in levels of cerebral thiamin and thiamin-dependent enzymes. Likewise in man, high tea consumption leads to a deficient thiamin status, but ascorbic acid, if present, completely inhibits the thiamin-inactivating processes.

Status Assessment

In terms of biochemical status, urinary vitamin excretion and blood enzymes are suitable criteria. Urinary thiamin excretion is, as with most B vitamins, largely dependent on intake and therefore provides an indication of the recent dietary intake but does not adequately reflect the body stores. At intake levels below 0.5 mg of thiamin per 1000 kcal per day (0.1 mg MJ^{-1}) thiamin excretion varies only slightly, and the proportion of metabolites in the urine exceeds the vitamin content itself. Because urinary excretion follows a diurnal rhythm, 24 h urine or the creatinine excreted is taken as the basis for status assessment. Occasionally, the status can be interpreted more accurately by a thiamin load test. After administration of a 5-mg dose, vitamin excretion is measured for the following 4 h. A deficient status can be assumed if less than 20 μg is excreted during this period.

Erythrocytic transketolase activity (ETKA), the rate-limiting enzyme in the pentose phosphate pathway, and its activation coefficient, α_{ETK} are often used as an indication of functional status as they give a better idea of body stores than urinary excretion. According to the saturation deficit of the apoenzyme, the transketolase activity can be stimulated *in vitro* by the addition of TDP. The ratio of stimulated to basal activity, which increases with the degree of deficiency, indicated by the activation coefficient (AC) or α_{ETK}, is commonly taken as a prognostic status index. The AC is better as a comparison of status in groups, because

of the large interindividual variation in basal enzyme activity. Moreover, it can be assumed that apoenzyme levels are not affected by the vitamin deficiency. Nevertheless, in alcoholics, either the apoenzyme synthesis or the linkage to the coenzyme TDP seems to be impaired, and the transketolase assay may indicate misleading results about the true status. Furthermore, confounding factors in the enzymatic assay, irrespective of the thiamin status, may include certain cancers, uremic neuropathy, diabetes, and treatment with certain anticancer drugs (fluorouracil, acytoxin) or diuretics, such as furosemide.

Recently, in view of improved high-performance liquid chromatography techniques, the TDP content in erythrocytes has been suggested to be a better indicator of the body stores compared with ETKA and independent of confounding factors. The criteria of the thiamin status indices are listed in Table 2.

Thiamin Requirement

Regarding the essential role of thiamin-dependent enzymes in carbohydrate and energy metabolism, the requirement for this vitamin is usually related to energy intake. Irrespective of the catalytic function of enzymes emerging basically unchanged from a biochemical or metabolic reaction, controlled studies in man have demonstrated that static as well as functional status criteria (transketolase activity and urinary excretion) deteriorate with increasing energy uptake even when the thiamin supply remains unchanged, thus indicating an increased vitamin turnover. The dependence on energy metabolism seems to be particularly evident in a marginal vitamin status.

Previously dietary fat was postulated to have a 'sparing effect' on thiamin requirements. This effect is, however, comparatively small, even with major increases in fat in the diet. Furthermore, in a starvation or semistarvation status when the vitamin intake is zero and increasing amounts of fatty acids are metabolized to satisfy the body's energy needs, the tissue stores of thiamin are rapidly depleted.

International committees recommend safe levels of intake amounting to 100–130 μg of thiamin per megajoule, equivalent to 0.4–0.54 mg of B_1 per 1000 kcal. A critical intake point appears to be approximately 0.05 mg of B_1 per megajoule (0.2 mg per 1000 kcal), below which urinary excretion is low, and the first clinical signs of deficiency may appear. A minimum intake of 1 mg per day is recommended to provide an adequate margin, even for those consuming less than 8 MJ (2000 kcal) daily. The additional values for pregnancy and lactation mostly exceed the calculated additional requirement of pregnant and lactating women. These extra values insure optimal pre- and postnatal development. In the past, clinical deficiency symptoms of infants were sporadically observed in spite of an inconspicuous appearance of the mother (infantile beri beri; see Table 4).

Owing to the low storage capacity, regular thiamin intake is important. Although, within the margin of physiological requirements, the organism attempts to prevent considerable thiamin losses by partial tubular reabsorption and intracellular fixation as TDP, the biological half-life time of this vitamin is assumed to be 9–18 days. After this period, when thiamin uptake is restricted, early nonspecific deficiency symptoms occur.

The recommended daily allowances for several European countries together with the USA's dietary reference intakes are listed in Table 3.

Deficiency Diseases

In thiamin deficiency, primarily, organs and tissues, such as the central and peripheral nerve system and the heart are affected showing an increased energy turnover or close carbohydrate dependency. Clinical signs described in most animals refer to lesions in these tissues.

Since Eijkman's observations in fowls, the degeneration of the peripheral nerves was the first pathological symptom of thiamin deficiency noted in

Table 2 Reference values of thiamin status parameters[a]

Indicator	*Adequate*	*Deficiency*	
		Marginal	*Severe*
Thiamin excretion in urine (μg per gram of creatinine)	> 66	27–66	< 27
ETKA–AC	< 1.15	1.15–1.25	> 1.25
TDP content in erythrocytes (ng per milliliter of sediment)	> 38.5	29.5–38.5	< 29.5

[a]From Finglas PM (1994) Thiamin. *International Journal for Vitamin and Nutrition Research* 63: 270–274 and Frank T, Bitsch R, Maiwald J and Stein G (2000) High thiamine diphosphate concentrations in erythrocytes can be achieved in dialysis patients by oral administration of benfotiame. *European Journal of Clinical Pharmacology* 56: 251–257.

Table 3 National RDAs for vitamins

Country	*B_1 (mg per day)*
Poland	1.9–2.0
Belgium, Denmark, Finland, France, Hungary, Luxemburg, Portugal, Spain	1.4
Greece, Norway	1.1–1.4
Czech Republic, Sweden	1.2
Austria, Germany, Ireland, Switzerland	1.0–1.3
The Netherlands, USA	1.0–1.2
Italy, UK	0.8–1.2

animals, leading to spastic pareses and paralysis of legs. Later on, the role of thiamin in the central nervous system, especially in pyruvate metabolism of the brain, could be demonstrated, too.

Neurological signs are also characteristic of the deficiency state in human beings. The classical manifestation of deficiency can be seen in beriberi. This disease is endemic in several regions of East Asia. Depending on how far the disease has progressed and additional nutrients are involved (e.g., protein deficiency), it is characterized by either polyneuropathias and degenerative muscle wasting (dry or atrophic form) or cardiac arrhythmias and edema (wet or exudative form). The development of exudative beriberi predominates in cases of a high carbohydrate intake and increased physical activity. Protein deficiency is probably involved, too. The Wernicke–Korsakoff syndrome, the cerebral form of deficiency, is characterized by mental confusion, memory disturbances, ataxia, ophthalmoplegia, and nystagmus. This syndrome can be interpreted as neuropathia of the central nervous system and is most commonly observed in malnourished alcoholics. The thiamin-dependent PDH and KDH complexes are key enzymes in the metabolic pathways of glucose and other energy-delivery substrates. The activities of both enzymes were found to play a rate-limiting role in mammalian brain mitochondrial preparations. Thus, lactic acid accumulation, as a result of the impaired glucose oxidation and disruption of oxidative phosphorylation utilizing pyruvate, α-ketoglutarate, and succinate as substrates has been demonstrated in brain mitochondria from deficient rats. There is some evidence that localized lactic acid accumulation may contribute to the brain lesions in the Wernicke–Korsakoff disease. Futhermore, it is suggested that neurological symptoms may involve an impairment of the cholinergic neurotransmitter function. Acetylcholine and γ-aminobutyrate are primarily produced through the oxidative metabolism of glucose or via the citric acid cycle with acetyl-coenzyme A (acetyl-CoA) and α-ketoglutarate as precursors. Although no change in regional brain acetylcholine levels was found in thiamin-deficient rats, acetylcholine utilization and turnover were reduced, and the central cholinergic mechanisms appear to be depressed in a deficient state. These alterations may be due to a decreased PDH and thereby activity limited acetyl-CoA production.

Infantile beriberi occurring in breast-fed infants between 2 and 6 months has particularly devastating consequences. Symptoms include cyanosis, dyspnea,

Table 4 Symptoms of clinical thiamin deficiency[a]

Deficiency form	*Neurological signs*	*Cardiovascular signs*	*Other signs*	*Nutrient interactions*
Atrophic beriberi (dry or polyneuritic form)	Degenerative polyneuropathia of lower extremities, paresthesia, reflex disorders, dropfoot, muscle waste, convulsions			(Caloric restriction), anorexia
Exudative beriberi (edematous or cardiovascular form)		Sinus tachycardia, cardiac arrhythmias, heart dilatation (beriberi heart), pericardial effusion, cardiac insufficiency	Lung and peripheral edemas, ascites, dyspnea, rarely lactic acidosis without edemas	(High carbohydrate intake, protein deficiency), intense physical activity
Wernicke–Korsakoff Syndrome (cerebral form)	Nystagmus, ophthalmoplegia, cerebellar ataxia, paralysis of the sixth brain nerve, polyneuropathia (burning feet syndrome), reflex disorders		Psychoses, hallucinations, memory loss, irritability	Alcohol abuse
Infantile beriberi (2–6 month)	Convulsions (increased intracranial pressure)	Tachycardia, cardiac insufficiency	Vomiting, diarrhea, abdominal distension, anorexia, cyanosis, dyspnea	

[a]Modified from Bitsch R (1997) Vitamin B_1 (thiamin). In: Biesalski HK, Schrezenmeier J, Weber P and Weiß H (eds) *Vitamine, Physiologie, Pathophysiologie, Therapie*, pp. 67–74. Stuttgart: G. Thieme.

tachycardia, abdominal distension, and convulsions. Sudden death often occurs as a result of cardiac failure. Nursing mothers, however, often do not show any signs of clinical deficiency. The thiamin deficiency symptoms are summarized in Table 4.

Therapeutic Administration and Toxicity

The importance of thiamin for the metabolism of the central and peripheral nerve system, as is becoming clear from the deficiency symptoms, implicates its therapeutic application in all neurological system disorders. The recommended doses are 100–300 mg per day, often combined with pyridoxine and cobalamin. Because of the limited gastrointestinal absorption, the dose should be split up over the day. In progressive polyneuropathia and clinical beriberi, the same dosage may be initially administered i.m. or i.v., for a rapid tissue restoration, followed by oral therapy.

Lipid-soluble allithiamin derivatives can be given in smaller doses, because of the improved absorption. The passage of those derivatives across the blood–brain barrier is alleviated, too, as has been shown in animals. A significant improvement in the neuropathia score by oral application of the lipophilic benfotiamin has been observed in diabetic polyneuropathia patients as well.

As with other water-soluble vitamins, thiamin toxicity is very small when given orally, because of the limited absorption. In rare cases, with parenteral dosages above 100 mg, impaired sensations or even anaphylactic reactions have been observed, but it is not clear as to what extent additional components of the injection solution may interfere.

Thiamin-dependent Enzymopathies

Thiamin-dependent enzymes are involved in four rare hereditary enzymopathies: leucinosis, Leigh syndrome, congenital lactacidosis, and thiamin-dependent megaloblastic anemia.

In leucinosis or maple syrup disease, the degradation of branched chain amino acids is inhibited with following accumulation of ketoanalogs. In the alleviated or intermittent form, this disorder can be improved by an intake of 10–150 mg of B_1 per day under simultaneous protein restriction. If untreated, severe neurological complications, physical, and mental dysfunctions develop, followed by death.

Leigh syndrome is characterized by a necrotizing encephalomyelopathia and is also observed in ruminants. The exact biochemical defect is poorly understood but is thought to be a hereditary malfunction of thiamin triphosphate transferase, leading to impaired TTP formation, occasionally combined with a decrease in the amount of ketoacid dehydrogenase. Neurological symptoms develop, progressing from nystagmus and paralysis of the external eye muscles to convulsions, ataxias, and disorientation resembling the Wernicke disease. Thiamin therapy may be successful, particularly when lipophilic derivatives (fursultiamin) are ingested up to gram dosages. Lactacidosis (which occurs at the same time) can be alleviated by oral intake of hydrogencarbonate combined with a low-carbohydrate diet.

The clinical picture of congenital lactacidosis resembles Leigh syndrome, and it is often impossible to distinguish between the two. The primary cause is a defect of the PDH or of partial enzymes of this complex. Generally, the prognosis is deleterious. The lactacidosis and the concomitant neurological symptoms could only in rare cases be affected by therapeutic thiamin doses.

Thiamin-responsive megaloblastic anemia has until now been very rare. The characteristic symptoms are a very strange combination of anemia, labyrinthine deafness, and insulin-dependent diabetes. The anemia can be improved only by oral doses of

Table 5 Congenital thiamin-responsive enzymopathies

Metabolic disorder	*Biochemical defect*	*Clinical symptoms*	*Therapy*
Leucinosis (maple syrup disease)	Ketoacid dehydrogenase for metabolism of branched-chain amino acids	Acidosis, ketosis, cerebral degeneration	Leucin-, isoleucin-, and valin-poor diet, thiamin (orally) 10–150 mg daily
Leigh syndrome (necrotizing encephalomyelopathia)	TTP formation, keto acid dehydrogenase	Lactacidosis, nystagmus, ataxia, convulsions, mental and psychic abnormal development	Thiamin (orally) 2.0 g daily (lipid-soluble derivatives)
Congenital lactacidoses	PDH	Lactacidosis, neurological symptoms, mental retardation	Thiamin lipid-soluble (?)
Thiamin-responsive megaloblastic anemia	Thiamin transport in tissue compartments	Megaloblastic anemia, diabetes, labyrinthine deafness	Thiamin (orally) 20–100 mg daily (lipid-soluble derivatives)

20–100 mg of thiamin. Folate and vitamin B_{12} administration is ineffective. In one case, the diabetes was improved, and insulin therapy was no longer needed.

The primary metabolic defect is unclear. Thiamin-dependent enzymes remain unchanged, and there may be an impairment in thiamin transport in certain tissues. The enzymopathies are summarized in Table 5.

See also: **Plant Antinutritional Factors**: Characteristics; **Beriberi**; **Bioavailability of Nutrients**; **Nutritional Assessment**: Biochemical Tests for Vitamins and Minerals; Functional Tests; **Vitamins**: Overview

Further Reading

Bitsch R (1997) Vitamin B_1 (Thiamin). In: Biesalski HK, Schrezenmeier J, Weber P and Weiß H (eds) *Vitamine, Physiologie, Pathophysiologie, Therapie*, pp. 67–74. Stuttgart: G. Thieme.

Duran M and Wadman SK (1985) Thiamine-responsive inborn errors of metabolism. *Journal of Inherited Metabolic Diseases* 8(supplement 1): 70–75.

Finglas PM (1994) Thiamin. *International Journal for Vitamin and Nutrition Research* 63: 270–274.

Food and Nutrition Board, Institute of Medicine (1998) *Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B_6, Folate, Vitamin B_{12}, Pantothenic Acid, Biotin and Choline*. Washington DC: National Academy Press.

Frank T, Bitsch R, Maiwald J and Stein G (2000) High thiamine diphosphate concentrations in erythrocytes can be achieved in dialysis patients by oral administration of benfotiame. *European Journal of Clinical Pharmacology* 56: 251–257.

Gregory JF (1997) Bioavailability of thiamin. *European Journal of Clinical Nutrition* 51(supplement): 34–37.

Gubler CJ (1991) Thiamin. In: Machlin LJ (ed.) *Handbook of Vitamins*, pp. 233–281. New York: Marcel Dekker.

Rindi G (1996) Thiamin. In: Ziegler EE and Filer Jr. LJ (eds) *Present Knowledge in Nutrition*, pp. 160–166. Washington, DC: ILSI Press.

Rindi G, Patrini C, Laforenza U *et al*. (1994) Further studies on erythrocyte thiamin transport and phosphorylation in seven patients with thiamin responsive megaloblastic anaemia. *Journal of Inherited Metabolic Diseases* 17: 667–677.

Souci SW, Fachmann W and Kraut H (2000) *Food Composition and Nutrition Tables*, 6th revised edn. Boca Raton, FL: CRC Press/Medpharm Scientific.

Thin-layer Chromatography *See* **Chromatography**: Principles; Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; Supercritical Fluid Chromatography; Combined Chromatography and Mass Spectrometry

THIRST

D J Ramsay, University of Maryland at Baltimore, Baltimore, MD, USA

Introduction

Thirst is a term which evokes many meanings. In most nonhuman studies, it is equated with measurements of water intake. In human studies, analog rating scales are often used to evaluate answers to questions such as: 'How thirsty are you?' 'How dry is your mouth?' or 'How unpleasant does your throat feel?' Its use is not restricted to an appetite for water. Depending on culture, experience, and availability, a variety of fluids such as soft drinks and nonalcoholic beverages, in addition to water itself, may be consumed. It has been suggested that the definition is best left in general terms such as 'the disposition to drink' or 'the tendency to drink,' and those who wish to use the word in a more specific sense should define it.

Role of Thirst

Water loss from terrestrial animals is continuous. Ingestion of food results in the addition of osmotically active solutes to the body fluids either because they are contained in the foods (e.g., sodium chloride) or because the process of metabolism produces them (e.g., oxidation of sulfur-containing amino acids to sulfate). These solutes are usually excreted in the

urine accompanied by water in amounts determined by the concentrating capacity of the kidney. Thus, if the kidney can produce urine four times more concentrated than plasma – the situation in humans – 25% of the volume of water will be lost compared to the situation in which the kidneys could not concentrate urine over plasma. Nevertheless, even kidneys producing low volumes of urine at maximal concentrations result in loss of body fluid. Water is also lost from the skin by insensible perspiration and sweating. These amounts are highly variable, depending on the level of activity, body temperature, and ambient temperature and humidity. Breathing also results in fluid lost as expired air is fully saturated with water at body temperature, whereas inspired air is usually drier and at a temperature lower than that of the body. Small amounts of water are lost in the feces, although in diseases accompanied by severe diarrhea, such as cholera, large amounts of fluid can be lost by this route. All these mechanisms cause a continuous inevitable loss of fluid. Thirst resulting in the ingestion of an appropriate volume of an appropriate fluid is the only way to repair the fluid deficit. Without thirst, animals would die of dehydration. Human subjects who lose the sensation of thirst as a result of disease are very difficult to treat. Even the most stringently prescribed timetable of water intake regime can fail. Irreversible dehydration is the inevitable outcome of the lack of thirst. (*See* **Water**: Physiology.)

Regulation of Thirst

Water is the most abundant component of the body, making up approximately 65% of body weight. The transport of solutes throughout the body takes place in physical solution, as do most of the physicochemical processes which maintain life. The total osmotic concentration of body fluids, or osmolality, is maintained within very narrow limits. In humans, as in other animals, it rarely varies by more than 1% or 2%. This is important for two reasons. First, as water can move freely through cell membranes, the osmolality of intra- and extracellular compartments is equalized by the movement of water down concentration gradients or osmotic gradients. If plasma and extracellular fluid osmolality increase, water will be withdrawn from cells, resulting in cell shrinkage. Thus, in situations of increased plasma osmolality, extracellular fluid volume will expand and intracellular fluid volume will decrease. Acutely, cells behave as perfect osmometers. It is therefore not surprising that osmolality is maintained constant in order to maintain both blood volume and cell volume within normal limits. Second, many bio-chemical reactions and cell transport processes depend critically on the concentrations of solutes. Thus accurate control of water balance to insure constancy of the volume and composition of body fluids is found in all mammals.

Cellular Dehydration Drinking

The correlation between total plasma solute concentration or osmolality and drinking has been known since the work of Mayer in 1900. More recently, the relationship between plasma osmolality and drinking has been described by a number of workers in a number of species, including humans. The relationship is generally investigated by determining the effects of intravenous infusions of hypertonic sodium chloride on water intake or, in the case of humans, increased indications of thirst, as noted on analog rating scales. In general, there is a linear increase in thirst associated with increased plasma osmolality similar to that seen in plasma vasopressin levels, the major determinant of urinary concentration. Thus increasing plasma osmolality leads to renal water conservation and the stimulation of thirst, and hence water-seeking behavior (Figure 1).

Extracellular Dehydration Drinking

Reduction of blood or extracellular fluid volume without a change in its osmolality also stimulates thirst. A number of maneuvres have been used to investigate this phenomenon. Hemorrhage, restriction of venous return by occluding the thoracic part of the vena cava, or sequestration of fluid using subcutaneous or intraperitoneal injections of colloid all

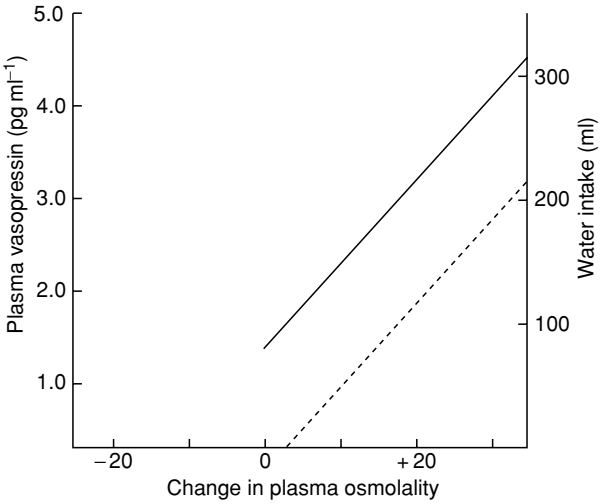

Figure 1 Increase in plasma osmolality stimulates vasopressin secretion (——) and water intake (- - -) in dogs. Reproduced from Thirst, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

result in stimulation of drinking. In general, drinking to extracellular fluid dehydration is less sensitive than that to cellular dehydration. The reduction of approximately 10% in blood volume is usually required to stimulate water intake, as is the case with stimulation of vasopressin secretion. With larger reductions in extracellular fluid volume, however, stimulation of thirst and vasopressin secretion is marked. Presumably, at the point where a reduction in extracellular fluid volumes severely compromises circulatory homeostasis, rapid restoration of blood volume becomes essential to insure survival.

Dehydration

During periods of water deprivation, plasma osmolality increases and extracellular fluid volume is reduced. It has already been emphasized that, under normal circumstances, water is lost continually, whereas drinking is a discontinuous process. Thus animals are always in a situation of potentially having to correct for dehydration by taking in fluids. In these situations, raised plasma osmolality, or cellular dehydration, is a more important input to the stimulation of thirst than is extracellular fluid dehydration. In primates, including humans, cellular dehydration may make up 90% of the stimulus.

Satiety

The stimuli responsible for the cessation of drinking are different from those which cause it. For example, dogs deprived of water for 24 h will make up fluid deficits accurately within 5 min of being allowed access to water. During this 5-min period, there is no correction of the raised plasma osmolality or decreased extracellular fluid volume as water would not have been absorbed from the gastrointestinal tract in sufficient quantities. Indeed, dehydrated dogs with an open gastric fistula will drink the same amount as intact dogs. In studies on water-deprived humans, subjects drank 65% of their total intake within 2.5 min. The remainder of the fluid deficit was consumed over a longer period of time. This satiety effect has been noted in a large variety of species; it is associated with a rapid inhibition of vasopressin secretion, and is based on oropharyngeal and gastric factors. However, it is only a temporary phenomenon and, unless drinking is followed by absorption of fluid and correction of the cellular and extracellular dehydration stimuli, drinking will be reinitiated. (*See* **Satiety and Appetite**: Food, Nutrition, and Appetite.)

Normal Drinking

Not all water intake depends on this simple deficit–satiety model. The water content of food, and that provided by its metabolism, can supply significant amounts of intake. The amount of water consumed on each occasion can depend heavily on learning and may be greatly influenced by other behaviors and by social factors. In humans the situation is even more complex and fluid may be taken in in association with caffeine or alcohol, and depend on specific tastes, such as sweetness, and on flavor. In spite of these factors, however, it should be noted than on average the kidney produces urine more concentrated than plasma. Thus renal-concentrating mechanisms are responding to a potentially dehydrating situation. In light of this information it might be asked why animals are not continually drinking. The answer is that drinking is only one of a large number of specific behaviors which may be expressed by an animal; if it were continually engaged in a search for water, to the exclusion of other behaviors such as feeding or reproduction, the animal would be incapacitated. Presumably, the inputs to thirst become more severe as dehydration proceeds, and water-seeking behavior moves to the top of the behavioral hierarchy and is expressed, resulting in water intake. In the long term, physiological control mechanisms must underpin drinking behavior. If inadequate quantities of fluid are taken in, particularly when urinary concentration becomes maximal, negative fluid balance and dehydration are the inevitable outcome.

Mechanisms of Thirst

Since the pioneering work of Verney on the location of osmoreceptors controlling vasopressin secretion, it has been realized that these are located in the hypothalamus. It now seems clear that receptors responding to physiological changes in extracellular fluid osmolality are located in circumventricular organs on the anterior wall of the third cerebral ventricle. In dogs, lesions of the organum vasculosum laminae terminalis result in the permanent loss of drinking behavior to increases in plasma osmolality of up to 15 mosmol kg^{-1}. In other species, another forebrain circumventricular organ, the subfornical organ may be involved. Circumventricular organs lack the normal blood–brain barrier, and thus can be more easily influenced by the composition of the blood that perfuses them. The connectivity between the circumventricular organs and other parts of the brain concerned with drinking behavior and secretion of vasopressin are shown in Figure 2. The nucleus medianus plays a central role in the integration of inputs which stimulate thirst. Lesions of this region result in the permanent cessation of drinking to any stimulus.

When extracellular fluid volume is reduced, information may reach the hypothalamus in two

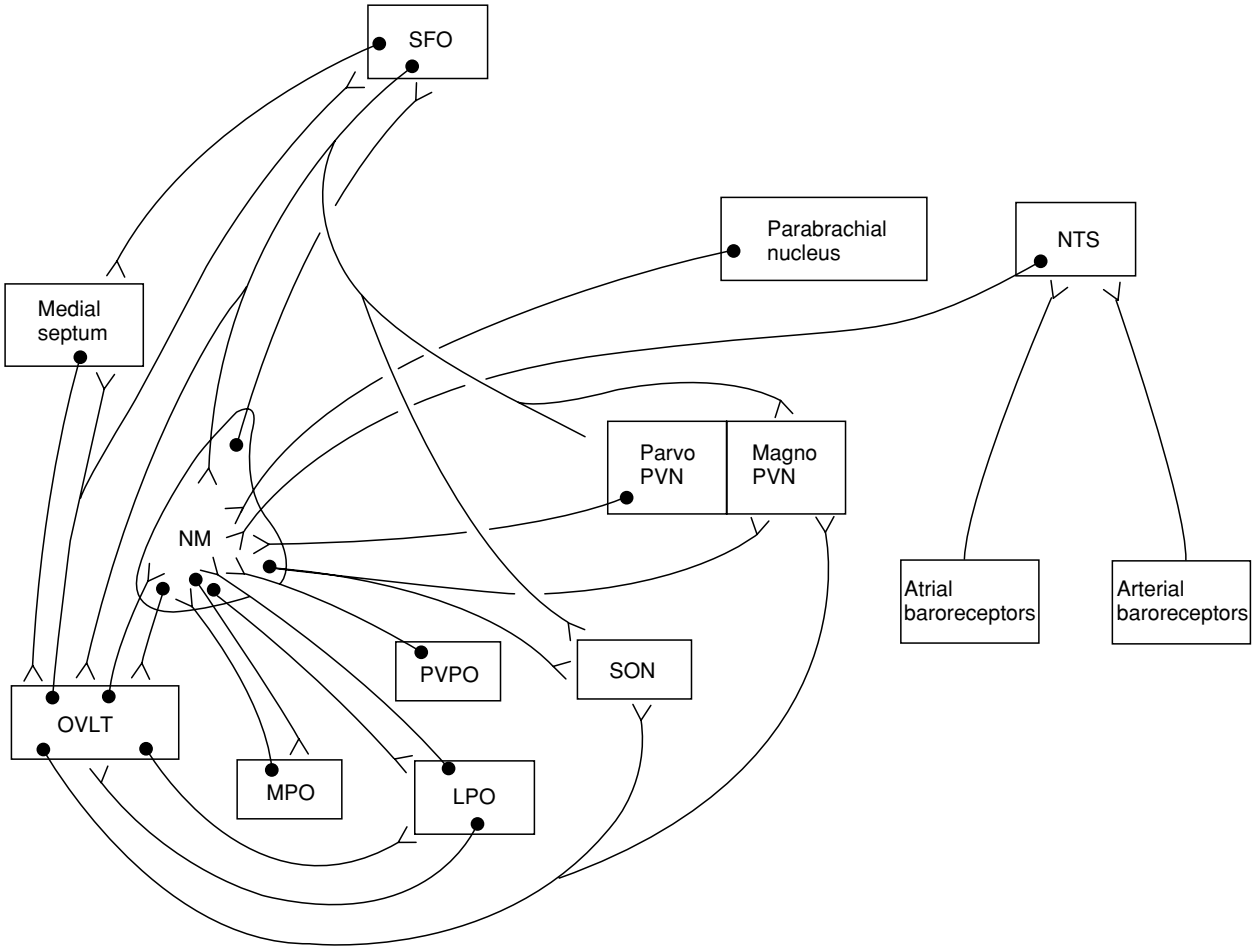

Figure 2 Neuroanatomical interconnections between forebrain circumventricular organs and the nucleus medianus in the control of water balance. SFO, subfornical organ; NM, nucleus medianus; OVLT, organum vasculosum laminae terminalis; MPO, medial preoptic area; LPO, lateral preoptic area; PVPO, periventricular preoptic area; SON, supraoptic nucleus; PVN, paraventricular nucleus; NTS, nucleus tractus solitarius. Reproduced from Thirst, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

main ways. Afferents from atrial and arterial baroreceptors enter the hindbrain via the ninth and tenth cranial nerves and synapse in the nucleus tractus solitarius. From this region, there are a number of inputs to the hypothalamus, including the nucleus medianus. Reduction in extracellular fluid volume also results in the stimulation of the secretion of renin from the kidneys. Through the renin–angiotensin cascade mechanism this results in raised plasma angiotensin II levels. Angiotensin II generated in the blood interacts with receptors in forebrain circumventricular organs, particularly the subfornical organ. Angiotensin II is a potent dipsogen, particularly when injected directly into the third cerebral ventricle. Thus extracellular fluid volume reduction can stimulate thirst directly through vagal afferents or more indirectly through the renin–angiotensin system.

The physiological inputs to thirst and the connectivity of the hypothalamic nuclei associated with these inputs are beginning to be understood. However, there is little information which links these with the complex neuro-physiological mechanisms associated with water-seeking and drinking behavior. The mechanisms which allow animals to choose drinking from the vast repertoire of behaviors available to them are also not clear. There is much to be learned of the higher-nervous-system functions involved in the link between a fluid deficit and the stimulation of thirst and fluid intake.

Requirements for Fluid

It is obvious from the foregoing that the requirement for fluid in humans must vary widely. In a temperate environment, the average loss of water

from the skin by evaporation is 900 ml day^{-1}, in feces 100 ml day^{-1} and in urine 1500 ml day^{-1}. The average figure for the amount of water contained in foods is 1000 ml day^{-1}; that produced by metabolism is 300 ml day^{-1}, leaving 1200 ml to be taken in in the form of fluid. However, these amounts may vary widely. For example, during moderate to severe exercise, particularly in a warm environment, many liters of fluid may be lost through sweating in order to maintain normal body temperatures. Under such circumstances, there is a marked stimulation of thirst with the resulting intake of large amounts of fluid. In the final analysis, in order to prevent dehydration the amount of fluids taken in must equal that which is lost.

The nature of the fluid ingested depends largely on cultural factors. For example, consumption of soft drinks appears to be the most characteristic feature of the USA. Beer, accompanied by milk (in the UK) and coffee (in Germany) are typical of middle Europe, and coffee and milk are the most popular beverages in northern Europe. Thus, as with food, culture is the major determinant of drink preference. It is important also to note the link between food intake and fluid intake. Under normal circumstances much of the daily intake of fluid is associated with eating.

A number of other factors will also influence fluid intake. During the luteal phase of the menstrual cycle, there is downward resetting of osmotic thresholds for both thirst and vasopressin secretion. These changes may be even more marked during pregnancy, when plasma osmolality may fall by about 10 mosmol kg^{-1} during the first few weeks of gestation. During this period the osmotic thresholds of both thirst and vasopressin are again reduced.

There is a reduction of thirst in the elderly. This is evidenced by the tendency for spontaneous dehydration to occur in the elderly, and their reduced sensitivity to infusions of hypertonic saline and water deprivation. The situation in the elderly is also complicated by their diminished ability to excrete water loads. For example, if an elderly individual with diminished thirst is encouraged to drink more fluid in warm weather, overcorrection can result in overhydration. Thus in the elderly, there may be problems of over- or underhydration. (*See* **Elderly**: Nutritionally Related Problems.)

See also: **Elderly**: Nutritionally Related Problems; **Satiety and Appetite**: Food, Nutrition, and Appetite; **Water**: Physiology

Further Reading

Denton D (1982) *The Hunger for Salt*. Berlin: Springer-Verlag.

Fitzsimons JT (1979) The physiology of thirst and sodium appetite. *Monographs of the Physiological Society* 35.

Ramsay DJ and Booth DA (eds) (1991) *Thirst: Physiological and Psychological Aspects*. Berlin: Springer-Verlag.

Rolls BJ and Rolls ET (1982) *Thirst*. Cambridge: Cambridge University Press.

Rundgren M and Hjelmqvist H (eds) (1989) Hypothalamus and homeostasis *Acta Physiologica Scandinavica Supplementum* 136 (583).

Thyroid Disease *See* **Iodine**: Properties and Determination; Physiology; Iodine-deficiency Disorders

Thyroid Hormones *See* **Hormones**: Adrenal Hormones; Thyroid Hormones; Gut Hormones; Pancreatic Hormones; Pituitary Hormones; Steroid Hormones

TIN

T Y Toribara, University of Rochester, Rochester, NY, USA

Background

Tin does not occur freely in nature but occurs mostly in the form of the dioxide and sometimes as the sulfide. It is found combined in silicic rocks such as granite and in small amounts in feldspar. The chemistry of tin has been examined in great detail, and much of it has been published recently in a single book. It is the intention to cover only those topics that have some relevance to the occurrence and properties of tin in food.

Occurrence in Food

The presence of tin in food is primarily the result of man's use of the element. The greatest exposure is the result of the use in cans in which food is preserved. Rather strong acid solutions are necessary for dissolution of the metal, so usually, very little of the element found its way into food kept in metallic containers which were either lined with tin or sealed with solders, leaving tin as one of the ingredients. (*See* **Canning**: Principles.)

Its metallic form has been used from very early times as a foil and in alloys such as Babbitt (Sn, Cu, Sb), bronze (Sn, Cu), pewter (Sn, Pb), solder (Sn, Pb), and type metal (Sn, Pb, Sb). Contact with these alloys may allow tin to leach into foods.

Although not in food, exposure to tin chemicals may occur through their use in dentistry. Tin(II) fluoride is used as an aid in preventing tooth decay through its presence in toothpastes, topical solutions, and mouthwashes.

More recently, a number of organotin compounds have been used as a stabilizer for the plastic PVC, as a fungicide in paints, and for wood treatment. In most uses, there is no direct contact of the organotin compounds with food products, and whatever appears in the food chain is the result of extraction from the material containing the organotin compound. These compounds will be discussed in greater detail in separate sections.

Properties of Tin Compounds

The compounds of tin may be classified as inorganic or organic. The properties of the various forms and the reactions that would cause the element to appear in the food chain are considered.

Inorganic

Tin usually occurs as the metallic element in pure or alloyed state, or as compounds in the divalent stannous form or the tetravalent stannic form. The use of the metal to coat cans and other containers of food has been common for many years. The occurrence of compounds of inorganic tin most likely to be found in food result from dissolution of the metal.

Initially, tin would be dissolved by an acid and would be in the divalent form. To keep the metal in this form, the solution must be quite acidic, and the element forms complexes of the form $[SnX_3]^-$, where X may be a halide or one of several other anions. Also, to keep the tin in the divalent form, it must be kept under an inert atmosphere such as carbon dioxide. In dilute solution, the Sn(II) compounds undergo hydrolysis, giving hydrated tin(II) oxide, which is not very soluble. The extent of hydrolysis depends upon the pH of the solution. In the usual pH range naturally encountered in foods, rather extensive hydrolysis occurs. In concentrated solution in the presence of oxygen (air), the divalent form is oxidized to the tetravalent Sn(IV) state, the ease of oxidation increasing with pH. In very strong alkaline solutions, spontaneous disproportionation to the metal Sn and Sn(IV) can occur. The divalent stannous form is a strong reducing agent, and it functions as such in many reactions.

Most of the inorganic forms of tin found in solution or combined with some solid material are in the tetravalent form. Tin(IV) oxide, the form most abundant in nature, dissolves in aqueous solutions of both acids and alkalis. The predominant species present in acid solutions containing complexing anions is $[SnX_6]^{2-}$, where X may be a halide or one of several other anions. In strongly alkaline solutions, the main species is $[Sn(OH)_6]^{2-}$. In dilute solution in the absence of complexing agents, all aqucous solutions of Sn(IV) tend to hydrolyze to give a precipitate of hydrated tin(IV) oxide. The halide $SnCl_4$, is a liquid with a boiling point of 114 °C.

Organic

Organic derivatives of both divalent and tetravalent tin are produced by man in significant quantities for various purposes. The preparations and structures of many of these compounds have been studied extensively. The divalent compounds are mostly of

academic interest and would not be encountered ecologically. Tetra-, tri-, and disubstituted tin(IV) compounds have been used extensively for a number of purposes in recent years. Tetramethyltin is a very good alkylating agent for the transfer of a single methyl group. It has been used to convert inorganic mercury to the monomethylmercury chloride for measurement of small quantities by gas chromatography. (*See* **Mercury**: Properties and Determination.)

The property of most interest in the treatment of samples for speciation is solubility. The dimethyltin and trimethyltin chlorides and bromides are water-soluble. Similar ethyl compounds are somewhat less soluble in water. The tin compounds with larger alkyl and phenyl groups are very slightly soluble or insoluble in water. The tetra-substituted compounds are all insoluble in water.

When the various compounds enter the environment, their fate is of importance. At near neutral pH and normal environmental temperatures, alkyl and aryl tin compounds are stable in the dark, but breakdown occurs in ultraviolet light. This has been observed with compounds of trimethyltin, tributyltin, and triphenyltin, in which sequential removal of the alkyl groups occurs to yield finally inorganic Sn^{4+}. Other mechanisms of breakdown by soil microorganisms and fungi have been observed.

The only significant formation of tin–carbon bonds in the environment appears to be via methylation. Monomethyltin was produced by an inoculum of mixed microorganisms from Chesapeake Bay sediments with stannic chloride. An acrobic strain of *Pseudomonas* (Ps 244) produced a variety of methyltins from inorganic tin(IV) but very little from tin(II). Formation of methyltins by other organisms has also been reported.

Analytical Methods

Sample Preparation

In general, the analysis of the tin content of a sample measures the total amount of the element. Most methods used require a solution of the sample. To extract the tin from an insoluble substance such as a tin ore, an initial extraction with dilute nitric acid is followed by a fusion of the insoluble residue with sodium peroxide or an alkaline carbonate mixture. Wet ashing procedures for samples such as tissues or other biological material can be carried out using concentrated nitric acid together with sulfuric acid. Such a procedure will destroy the organic matter, so that only the total tin content can be measured.

For the successful analysis of an organotin compound, some knowledge of the chemistry of the compound is required. The solvent used for the first extraction depends upon the compound of interest. Aqueous combinations with tropolone with some organic solvent are the most common methods. The method of analysis may be GC (gas–liquid chromatography) or HPLC (high-performance liquid chromatography), again depending upon the compound of interest. These methods will be discussed in a later section. Sampling and the treatment of the sample are of utmost importance. Tissue samples are usually homogenized and then extracted. Complete extraction is difficult because of a relatively large emulsified solids layer between the aqueous and organic layers. One procedure to correct for this is to add a known quantity of an organotin compound such as a tripropyltin salt known not to be in the sample. The recovery of this compound in the subsequent analysis will give a correction factor for the compounds being determined. For a number of compounds, the addition of tropolone to the nonpolar solvent improves the recovery in the extraction.

Instrumental Methods

In most laboratories, emission or absorption atomic spectroscopy is usually used for measuring the tin content of a sample. Tin can be analyzed with a standard emission spectrometer. The newer high-temperature DC (direct current) arc or ICP (inductively coupled plasma) spectrometers are best.

Atomic absorption spectrometry is the most sensitive method available. For the analysis of tin, a nitrous oxide–acetylene flame must be used. Modern instruments with a graphite furnace are best. Here, only a small sample (*c.* 20 µl) is required. Using the programmed temperature facility, the sample is dried, then volatilized and decomposed.

The detection limits for the atomic spectroscopic methods are:

Emission spectroscopy	$30\ \mu g\ l^{-1}$
Flame atomic absorption (AA)	$100\ \mu g\ l^{-1}$
Hydride generation AA	$4\ \mu g\ l^{-1}$
Graphite furnace AA	$0.2\ \mu g\ l^{-1}$

Chemical Methods

The early methods for the analysis of tin were gravimetric and titrimetric, depending on the quantity to be measured. These techniques required the separation of the element prior to the final measurement. In the gravimetric method, tetravalent tin is precipitated as metastannic acid, as the sulfide, or with a reagent such as cupferron after first separating it from interfering ions. The insoluble tin compound is filtered off, ignited to SnO_2, and weighed.

Titrimetric methods involve reducing the tin to tin(II) and then oxidizing it to tin(IV) with an iodine or iodate solution. The stannous form is readily oxidized in the presence of air, and for accurate work, the reaction is carried out under an atmosphere of CO_2. A method using tin(IV) makes use of the oxalatostannate complex, which will incorporate one molecule of sulfide to form the oxalatothiometastannate complex. The amount of sulfide may be measured by titration with iodine.

Chromatographic Methods

The early analyses of organotin compounds were by GC, both conventional and, more recently, with capillary columns being used. The most common organotin compounds are those that are used as fungicides in paints or sprays. These include tetra-, tri-, and dimethyltin compounds, and the homologous ethyl compounds as well as some phenyl derivatives. GC, which would seem to be an obvious choice to distinguish among the various substituted organotin compounds, must be used with caution. The problem arises from the fact that elevated temperatures must be used to volatilize some of the compounds, and at elevated temperatures, some of the compounds will undergo transformation. There is a further complication if a mixture contains both tetramethyltin and a dimethyltin compound. Tetramethyltin is a good methylating agent, and the dimethyl compound will be methylated to a trimethyltin compound. (*See* **Chromatography**: Thin-layer Chromatography.)

In order to be able to operate at lower temperatures, some of the organotin compounds, such as certain halides, are converted to more volatile species. Among the more common methods used are the replacement of the halide by a methyl or ethyl group, or the formation of the organotin hydride.

HPLC has become the method of choice for many analyses, as decomposition of the compounds being analyzed does not occur in the liquid phase at the lower temperatures used. For the separation of methyltin compounds in HPLC, a mobile phase of 60 parts acetone and 40 parts *n*-pentane is used. For ethyltin compounds, a mixture of 70 parts acetone and 30 parts *n*-pentane is used. As elevated temperatures are not used, the redistribution of species that occurs during GC is not encountered. (*See* **Chromatography**: High-performance Liquid Chromatography.)

Effects of Dietary Tin

Most of the tin ingested from canned food has been inorganic and has been considered low because of its poor absorption and retention in tissues. Many studies have been conducted at cellular levels using *Escherichia coli* and yeast cells to determine the effect of tin. It is not certain whether these studies may be applicable to humans.

By using controlled diets fed to weanling rats over prolonged periods of time, Rader and coworkers have made an extensive study of the effects of 100–1100 μg per gram of dietary tin on tissue and bone minerals of these weanling rats. The level of tin used in these studies was within the range of values reported for tin present in unlacquered cans. The total food consumption was restricted to the amount consumed by the rats fed 1100 μg of tin per gram. The amount of tin in the food of any rat was constant for a period of 4 weeks Food consumption by rats fed diets containing 0, 100, and 330 μg of tin per gram was measured weekly. After 4 weeks, the rats were sacrificed, and serum was prepared from the blood. Soft tissues were taken from liver, kidney, and duodenum. Femurs were also taken for examination. Serum ceruplasmin and cholesterol were then measured.

The results of these studies indicate that anemia and growth depression are common effects of chronic ingestion of inorganic tin. Low levels of dietary tin primarily affect copper and zinc status. Pronounced copper depletion in all tissues examined was observed at tin levels as low as $30\,\mu g\,g^{-1}$. Growth depression generally occurs when dietary tin levels exceed $500\,\mu g\,g^{-1}$. Diets high in copper and iron (50 and $250\,\mu g\,g^{-1}$) reduced the signs of anemia in rats fed tin at levels up to 1500 μg of tin per gram.

In normal animals, about 90% of copper in plasma is contained in ceruloplasmin. The finding of dietary decreased ceruloplasmin in tin-treated rats is consistent with the observation of decreased copper in plasma of rats fed diets containing more than 200 μg of tin per gram. Copper plays a major role in bone metabolism through its involvement in collagen cross-linking. The levels of copper and iron have been shown to affect anemia in rats.

Reduced growth is a common symptom of zinc deficiency. Low dietary tin adversely affects the growth of rats fed zinc-adequate diets. These studies on weanling rats indicate that many effects of inorganic tin have effects that are not directly attributed to tin, as in the case of reduced growth, even in zinc adequate diets.

The effects of continuous ingestion of small amounts of tin that reduce the quantity of essential elements to the levels that slow down growth and cause anemia have emphasized the importance of reducing the possible intake of tin. Since the principal source of tin is in containers used for food including cans, and items such as cans, trays, and other items

that may hold food, efforts are being made to reduce the possible contact to any tin that may be in the items. The primary target is the food can, and some efforts are included.

Food Cans

Early food cans were made of tin because of their permanent shiny appearance and the ease with which they could be manufactured. However, the lack of strength to maintain their shape led to the use of steel, which was much sturdier and less expensive. However, steel corrodes easily when exposed to air.

The findings that levels of tin that were at the levels of those originally eluted by contact with food in the tin cans has effects on the levels of copper, iron, and zinc, which can affect health, has made the coating of steel for food cans a very important issue. This led to various means to prevent corrosion, which occurred inside the cans with contents of juices of various foodstuffs. The early method used to prevent this corrosion (rusting) was to plate the surface with tin. However, any small deformations or pinholes would permit rusting of the steel. This led to the use of various plastics.

The literature is full of procedures for protecting steel, and most are coatings, for which patents describe in detail the ingredients and the procedure for preparing for them. An early simple procedure was to plate the outer coat with tin and apply an inner thermoplastic layer with particles of tin in them. The contents of the can would then be exposed to only a small amount of tin, and the outer surface of the can would be protected by the tin coating.

One Japanese patent describes a corrosion-resistant metal–polyester laminate formed by laminating a hot metal sheet with a polyester film coated with a thin aluminum layer. A number of layered coatings without plastic have also been patented. One such example describes a steel strip flash-electroplated with nickel, electroplated with tin, rinsed and dried, and then chromated.

Electric-resistance seam-welded cans are painted internally and have a tin coating with double chromate layers of chromium oxide layer. They then have a final coating of epoxy-phenolic paint and are baked at 205 °C. The patents all state that the resulting steel is suitable for food cans.

Summary

Long-term exposure of weanling rats to levels of tin expected in foods contained in tin cans has provided quantitative data on their effects. Copper and zinc levels were the primary targets of exposure through the diet. Since these elements are a vital part of many functions, tin has an indirect effect.

General exposure to tin is in the inorganic form in foods contained within cans containing tin. Originally, this exposure was through the use of tin cans, because tin seemed to be inert to general exposure and was an easy metal to mold into cans. However, the relative softness resulted in many dents in handling the cans, and many cans used to be rejected because of the dents. For this reason, the industry turned to the use of steel, a less expensive and much sturdier material.

However, the surface of steel is easily oxidized (rust), forming an unpleasant appearance. Initially, the steel was plated with tin to coat both the inside and outside surfaces. The tin coat did not provide complete protection, because small holes would allow exposure to the air, resulting in rust spots. What happened inside the can was not of general concern, because the amounts of soluble tin were not considered to be harmful. The first solution to maintain the external appearance was to coat the external surface with some type of plastic.

There have been many patents on the use of various plastics to coat the steel, and many elements including tin are included in the plastics. Many of the resultant steel sheets are satisfactorily electrowelded. The patents all state that the steel sheets are suitable for food cans. This interest is good evidence that the health of the consumers of canned goods is very important to showing the concern to produce safe materials for canning.

See also: **Canning**: Principles; Cans and their Manufacture; Recent Developments in Can Design; **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography; **Copper**: Physiology; **Mercury**: Properties and Determination; **Zinc**: Deficiency

Further Reading

Boyer IJ (1989) Review paper. Toxicity of dibutyltin and other organotin compounds to humans and experimental animals. *Toxicology* 55: 253–298.

Cappon CJ (1987) GLC speciation of selected trace elements. *LG-GC* 5: 400–418.

Cappon CJ (1990) Speciation of selected trace elements in edible seafood. In: Nriagu JE and Simons MS (eds) *Food Contamination from Environmental Sources*. New York: Wiley.

Chang LW (1985) Neuropathological effects of toxic metal ions. In: Gabay S, Harris J and Ho BT (eds) *Metal Ions in Neurology and Psychiatry*, vol. 15, pp. 207–230. New York: Alan R. Liss.

Donaldson JD and Grimes SM (1989) The inorganic chemistry of tin. In: Harrison PG (ed.) *Chemistry of Tin*, pp. 118–144. New York: Chapman & Hall.

Evans CJ (1989) Industrial uses of tin chemicals. In: Harrison PG (ed.) *Chemistry of Tin*, pp. 421–453. New York: Chapman & Hall.

Furman NH (ed) (1962) *Scott's Standard Methods of Chemical Analysis*, vol. I, pp. 1073–1093. Princeton, NJ: Van Nostrand.

Lickiss PD (1989) Organometallic compounds of bivalent tin. In: Harrison PG (ed.) *Chemistry of Tin*, pp. 221–244. New York: Chapman & Hall.

Molloy KC (1989) Organometallic compounds of tetravalent tin. In: Harrison PG (ed.) *Chemistry of Tin*, pp. 187–220. New York: Chapman & Hall.

Nielson FH (1976) Newer trace elements and possible application in man. In: Prasad A (ed.) *Trace Elements in Human Health and Disease, II. Essential and Toxic Elements*, p. 395. New York: Academic Press.

Rader JI (1991) Anti-nutritive effects of dietary tin. *Advances in Experimental Medicine & Biology* 289: 509–524

Selwyn MJ (1989) Biological chemistry of tin. In: Harrison PG (ed.) *Chemistry of Tin*, pp. 359–396. New York: Chapman & Hall.

Toribara TY (1985) Preparation of $CH_3{}^{203}HgCl$ of high specific activity. *International Journal of Applied Radiation and Isotopes* 36: 903–904.

Toribara TY, Jackson DA, French WR, Thompson AC and Jaklevic JM (1982) Nondestructive X-ray fluorescence spectrometry for determination of trace elements along a single strand of hair. *Analytical Chemistry* 54: 1844–1849.

Underwood EJ (1977) *Trace Elements in Human and Animal Nutrition*, 4th edn. pp. 449–458. New York: Academic Press.

Walker JQ, Jackson MT Jr. and Maynard JB (1977) *Chromatographic Systems. Maintenance and Troubleshooting*, 2nd edn. New York: Academic Press.

Willard HH and Toribara TY (1942) A volumetric method for determining tin based on the formation of a dioxalatothiometastannate. *Industrial and Engineering Chemistry Analytical Edition* 14: 716–721.

Willard HH and Toribara TY (1942) I. The preparation and properties of potassium oxalatostannate. II. A study of complex dioxalatothiometastannates. *Journal of the American Chemical Society* 64: 1759–1765.

TOCOPHEROLS

Contents
Properties and Determination
Physiology

Properties and Determination

D C Woollard, AgriQuality NZ, Auckland 1, New Zealand
H E Indyk, Anchor Products, NZWP, Waitoa, New Zealand

Physical Properties

The generic term 'vitamin E' is used to describe a group of eight structurally related tocopherol and tocotrienol compounds (**Figure 1**). The parent form is commonly referred to as α-tocopherol and systematically named 2,5,7,8-tetramethyl-2-(4′,8′,12′-trimethyltridecyl)chroman-6-ol, as a derivative of the benzene-pyran ring structure.

The tocopherols are insoluble in water, freely soluble in fats and oils, and soluble in most organic solvents. They exist as pale yellow viscous oils in purified form but can be crystallized at subzero temperatures, the melting point of α-tocopherol being 2.5–3.5 °C. Each tocopherol contains three asymmetric carbon atoms, making possible a total of eight optical isomers, although the epimeric configuration at carbon-2 dominates biological activity. **Table 1** lists the major physical properties of the naturally occurring *RRR*-tocopherols, as well as the synthetic *all*-rac-α-tocopheryl acetate. Spectral properties and specific optical rotations are highly dependent on the solvent used during measurement.

All vitamin E congeners show weak UV absorption (290–300 nm) and unusually narrow bandwidths. However, they exhibit strong native fluorescence, emitting radiation at 324–330 nm, the practical use of which has increased with advances in detector design. The esters of vitamin E have

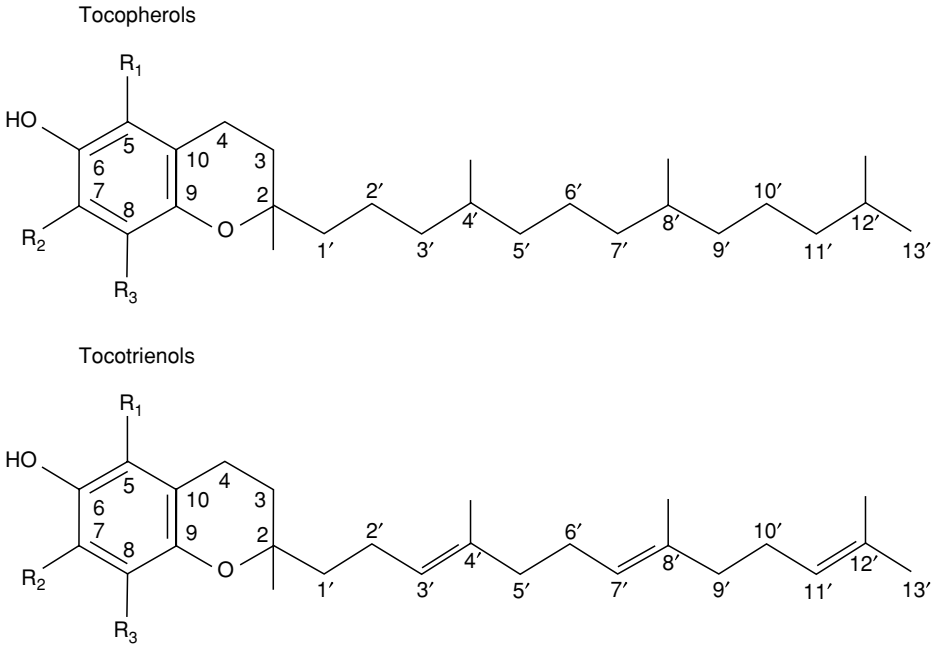

	R_1	R_2	R_3
α-Tocopherol and α-Tocotrienol	CH_3	CH_3	CH_3
β-Tocopherol and β-Tocotrienol	CH_3	H	CH_3
γ-Tocopherol and γ-Tocotrienol	H	CH_3	CH_3
δ-Tocopherol and δ-Tocotrienol	H	H	CH_3

Figure 1 Structure of the tocopherols and tocotrienols.

Table 1 Physical properties of the tocopherols

	α-*T*	β-*T*	γ-*T*	δ-*T*	α-*T acetate*
Formula	$C_{29}H_{50}O_2$	$C_{28}H_{48}O_2$	$C_{28}H_{48}O_2$	$C_{27}H_{46}O_2$	$C_{31}H_{52}O_3$
Molecular weight	430.69	416.66	416.66	402.62	472.73
Boiling point (°C, 10 kPa (0.1 atm))	210–220	200–210	200–210		184
Optical rotation (α_{546} in ethanol, *RRR*-isomers)	+0.32°	+2.9°	+3.2°	+3.4°	+3.2°
λ_{max} (nm) in ethanol	294	297	298	298	285.5
Extinction coefficient, $E^{1\%}_{1\,cm}$, in ethanol	71.0	86.4	92.8	91.2	43.6

See text for the abbreviated names used.

bathochromically modified spectral properties as a consequence of electronic interactions with the chromanol chromophore, and reduced extinction coefficients and quantum yields relative to the parent alcohols. The tocopherols and tocotrienols show a strong infrared band at 8.6 μm and O–H and C–H stretching frequencies at 2.8–3.0 and 3.4–3.5 μm, respectively.

Chemical Properties

The tocopherols and tocotrienols share a common chemical structure based upon tocol. Tocol consists of a chromanol nucleus with substitution at the 2- and 6-positions. A methyl group and a phytyl side-chain are bonded to C-2, while the hydroxy functionality on C-6 is available for esterification. Tocol itself is not

widely regarded as an E-vitamin but shares some of the properties of the group, particularly regarding its antioxidant role, and occurs naturally in low concentration.

The tocopherols are derived from tocol by the addition of methyl groups on the three available carbons (5,7,8) of the benzene ring. Trimethylation produces α-tocopherol, the most active E-vitamin as defined by rat fertility and other biological activity studies. In general, any structural modifications from α-tocopherol markedly reduces biological activity, including side-chain unsaturation, side-chain length, oxidation of the chromanol ring, and loss of any methyl group. While C-8 methylation is essential, β- and γ-tocopherol isomers (5,8- and 7,8-dimethyltocol, respectively) have reduced bioactivities of ~30% and ~12%, and δ-tocopherol (8-methyltocol) exhibits only 1% biopotency relative to α-tocopherol. Other mono- and dimethyltocols are also known to occur, as either minor natural components or synthetic byproducts, but are not classified as E-vitamins owing to their lack of any measurable physiological response.

Each of the four tocopherols has a related tocotrienol (abbreviated T3) through phytyl side-chain unsaturation at C-3′,7′ and 11′, resulting in reduced biopotency: α-T3 (25%), β-T3 (5%), γ-T3 (<1%) and δ-T3 (<1%) relative to α-tocopherol.

Optical asymmetry at C-2 yields stereoisomers, of which the D-form is biologically more active by a factor of 2.38 compared with the L-isomer. Thus, naturally occurring α-tocopherol, which exists exclusively as the D-isomer has a higher efficacy than the synthetic racemic equivalent.

The tocopherols have two additional chiral carbon atoms (4′ and 8′), and naturally occurring D-α-tocopherol is more correctly described as 2*R*,4′*R*,8′*R*-α-tocopherol (abbreviated *RRR*-α-tocopherol). The stereochemistry of the phytyl 'tail' is known to govern biological properties, believed to be a consequence of *in vivo* interaction with chiral biological membranes. Thus, the natural *RRR*-α-T isomer is approximately three times more effective than its stereochemically inverted relative *SRR*-α-T (referred to as 2-*epi*-α-tocopherol). While tocotrienols possess no side-chain asymmetric centers, geometrical isomerism becomes a possibility, although only the *trans*-isomers have been reported. The influence of *cis* geometry on biopotency is therefore unknown.

The tocopherols are vulnerable to aggressive forms of oxygen, leading to the formation of biologically inactive monomeric and dimeric quinones. These reactions are accelerated by UV light, heat, high pH, and certain transition elements, whereas in the absence of oxygen, the E-vitamins are relatively stable. Lability is decreased by esterification at the 6-hydroxy substituent, presumably through stabilization of the phenolic ring structure. Consequently, most nutritional uses of vitamin E involve α-tocopherol esters.

To account for the variation in biological activities of the isomers, the concentration of vitamin E is often defined in international units rather than by weight. The international unit (IU) is expressed relative to synthetic *all-rac*-α-tocopheryl acetate, where 1 IU is equivalent to 1 mg, based upon fetal resorption after oral administration to rats. The information agrees fairly well with rat hemolytic anemia studies. On this basis, 1 unit is provided by 0.91 mg of *all-rac*-α-tocopherol and 0.67 mg of *RRR*-α-tocopherol. The bioactivities of other tocopherols and tocotrienols are then calculated against these data by applying the appropriate conversion factors.

As a consequence of their phenolic structure, the tocopherols and tocotrienols are potent antioxidants, conferring protection to vulnerable lipids in biological tissues and foods. Such antioxidant behavior is presumed to depend on pyran ring cleavage to form the hydroquinone intermediate, followed by oxidation to tocopherol quinone. These and other oxidation products (tocopheroxides, epoxides, and polymers) are receiving increased attention.

Occurrence and Forms in Food

The dietary intake of vitamin E in the general population appears to be sufficient to meet the US Recommended Daily Allowances of 3 mg (infants) to 12 mg (adult males). Avitaminosis E is comparatively rare in well-nourished communities because of large physiological reserves in tissues and organs, although neuromuscular and other disorders have been associated with dietary deficiencies, particularly in children. (*See* **Cells; Dietary Requirements of Adults**.)

The tocopherols are well distributed in nature, the richest and most varied isomeric sources being vegetable oils and products derived from them, such as margarine. It has been common practice with these products to speciate the various vitamers during analysis (Table 2). Other valuable sources of vitamin E include nuts, cereals, green vegetables, fruits, and foods of animal origin such as eggs, dairy products, and meats (Table 3). In practice, the contribution of the non-α vitamers in most foods is sufficiently small after biopotency factors are applied that they are regularly excluded from estimations of total vitamin E status. Animal tissues contain vitamin E almost exclusively as α-tocopherol, with the exception of egg yolk, where γ-tocopherol is found in significant concentrations. There is a fairly wide disparity in literature data, probably as a consequence of genetic, geographical, seasonal, and processing factors as well

Table 2 Vitamin E content of selected vegetable oils

	Tocopherol (mg per 100 g)				*Tocotrienol (mg per 100 g)*				
Oil	*α-T*	*β-T*	*γ-T*	*δ-T*	*α-T3*	*β-T3*	*γ-T3*	*δ-T3*	*Total*
Coconut	0–2	0–1	0–1	0–1	0–1		0–4		1–5
Soya bean	1–35	0–4	40–240	15–95					60–330
Corn	2–55	0–30	25–250	2–8	0–24	0–5	0–40	0–2	30–340
Wheat germ	120–150	40–70	0–25	0–25	0–2	6–18			150–200
Sunflower	40–80	1–4	0–5	0–1					45–90
Palm	0–20		1–4		0–30	0–3	4–70	0–10	10–130
Olive	10–25		Trace						10–27
Safflower	20–45		3–10						25–50

Blanks indicate a lack of reported data, implying concentration below limits of detection.

Table 3 Vitamin E content of selected foods[a]

	Tocopherol (mg per 100 g)			
Food	*α-T*[b]	*β-T*	*γ-T*	*δ-T*
Peanut	7.0–8.3		8.0	
Spinach	1.8–2.5		0.1	1.0
Lettuce	0.2–0.4	0.04	0.1	
Parsley	0.8–1.2		0.1	
Celery	0.3–0.5		0.1	0.4
Peas	0.2–0.5	0.05	0.2	
Carrot	0.4–0.5	0.02		0.01
Tomato	0.4–0.9	0.03	0.1	
Potato	0.05–0.1			
Apple	0.3–0.6	0.02	0.07	
Blackberry	3.0–3.5		4.7	4.5
Banana	0.2–0.3			
Wheat flour	0.7–0.9	0.6		
Beef, raw	0.4–0.8		0.02	
Chicken, cooked	0.2–0.4		Trace	
Halibut	0.4–1.3		Trace	Trace
Milk	0.05–0.13		0.01	
Butter	1.0–2.0		0.05	
Margarine	10.0–15.0		25.0	3.0
Egg white, raw	0.07–1.95	0.04	0.2	0.01

[a]α-Tocopherol values are expressed as typical ranges, whereas representative values are reported for the less biologically active congeners. Blanks indicate a lack of data or the absence of traceable levels. Tocotrienol contributions are not reported as they are predominantly absent from these foods.
[b]To convert to IU per 100 g, multiply by 1.49 (as *RRR*-α-tocopherol).

as variations in analytical techniques. Refer to individual foods.

Tocopherols and tocotrienols are labile materials, and their stability during food processing and storage has been subject to intensive investigations. Storage of foods under anaerobic conditions and the incorporation of other synthetic or natural antioxidants are common strategies to optimize the preservation of their vitamin E status.

Role as Antioxidants

Oxidation of polyunsaturated lipids operates through a free-radical pathway that not only facilitates food spoilage but creates substances of potential toxicological significance that can have damaging consequences upon ingestion. The tocopherols and tocotrienols combat this process both *in vitro* and *in vivo* by trapping the hydroperoxide intermediates and preventing their propagation. The fate of the stable vitamin E phenoxy radicals so produced is variable, with formation of monomeric quinone, which can react further with a secondary radical forming a dimer. Under certain conditions, ascorbate is considered capable of synergistically repairing the phenoxy radical and regenerating the tocopherols. (*See* **Antioxidants**: Natural Antioxidants; **Oxidation of Food Components**.)

It has been confirmed that antioxidant activity of the tocopherols follows the same order as their biological potencies (α>β>γ>δ) and that this may be rationalized on a stereoelectronic basis. Thus, a higher resonance stabilization of the phenoxy radical is achieved in α-tocopherol where ring planarity is optimal relative to the other congeners.

Vegetable oils and margarines are particularly prone to oxidation, owing to their high content of polyunsaturated fatty acids but fortuitously also possess substantial native vitamin E, which functions protectively. However, additional stabilization is often considered necessary, particularly for refined oils where the tocopherols become depleted during the harsh manufacturing conditions. Although animal fats and foods rich in them are predominantly saturated and less sensitive to oxidative processes, the addition of tocopherols is usually of benefit since natural vitamin E contents are low. (*See* **Margarine**: Methods of Manufacture; **Vegetable Oils**: Types and Properties.)

The ratio of tocopherol to polyunsaturated fat is an important criterion in determining optimum oxidative protection offered by vitamin E and also the toxicological safety of the fat. Excessive concentrations of vitamin E are reported to accelerate lipid oxidation, although there is debate about the detail

and extent of this process. As expected, antioxidant and prooxidant properties differ between the various tocopherols and tocotrienols. Vitamin E acetate cannot be used as an antioxidant because this property is lost upon esterification.

Use in Food Fortification

As indicated previously, the tocopherols are among the most important fat-soluble antioxidants used in the prevention of food spoilage. However, vitamin E has other recognized biological functions and consequently is added to a variety of foods for the management of community health. This has particular relevance for infant nutrition and for sections of the population where inadequate nutrition is suspected. To maintain vitamin efficacy, the free phenols are not used during food fortification unless an antioxidant function is also required. (*See* **Food Fortification**.)

The most common nutritional additive is synthetic *all-rac*-α-tocopheryl acetate owing largely to its greatly enhanced oxidative stability. Technological problems during the fortification process can be overcome by utilizing the oily vitamin in a water-dispersible 'beadlet' form, contained within a gelatin coating. The vitamin can then be blended as an emulsion into liquid foods, usually prior to spray drying. Milk powders, dietetic, and infant formulae are often supplemented in this way, facilitating a uniform distribution of the additive. In other products, the ingredients may be mixed in the dry state, either because the nature of the food is not conducive to wet-blending procedures or because the vitamin E has a diminished stability owing to low fat content. Cereals and skim-milk powder products are regularly supplemented in this manner. The major disadvantage with this protocol is the potential for vitamin heterogeneity. Particle-size control and correct mixing, packaging, and storage conditions are therefore necessary to minimize uneven distribution of the vitamin beadlets.

Vitamin E succinate can be used if severe heterogeneity problems are encountered since this ester is a solid at room temperature and more easily dry-blended into foods and pharmaceuticals. A disadvantage with the use of the succinate ester is its lower bioavailability. The third commercially available vitamin E ester, tocopheryl nicotinate, is used in cosmetics rather than as a food additive.

Isolation and Analysis

Vitamin E poses analytical challenges because of its multiplicity of forms and its tendency to oxidize. Traditional bioassays (e.g., resorption gestation, encephalomalacia, myopathy, and erythrocyte hemolysis tests) do not discriminate between the various vitamers (vitamin isomers) in foods. They are tedious and relatively imprecise techniques, although they do exhibit true, host-specific, activity data. Lipid peroxidation tests for the accumulation of hydroperoxides or aldehydes in animals have also been used to indicate vitamin E status. (*See* **Vitamins**: Determination.)

Physicochemical tests have become more popular because of their ease of use, speed, and precision compared with bioassays. The traditional chemical test has been the Emmerie–Engel reaction, which detects the presence of fat-soluble reducing agents through conversion of trivalent iron to the red ferrous bipyridyl complex. The accuracy of this and related colorimetric procedures is, however, compromised by spectral interferences of other antioxidant materials such as biologically inactive tocols, retinoids, carotenoids, nonvitamin chromanol compounds, and synthetic food antioxidants. The success of the method is therefore dependent upon the degree of purity of the vitamin E analytes prior to colorimetry. The problem remains that the eight E-vitamins are not differentiated and will be quantified as a single analyte. These same difficulties remain, albeit to different degrees, whether detection is by UV absorption, fluorescence, or electrochemical techniques. In each case, a chromatographic step is required in the analytical scheme in order to remove interferences and, ideally, to separate the vitamers. Traditional column or thin-layer chromatographic (TLC) techniques have been used extensively for this purpose. (*See* **Chromatography**: Principles; Thin-layer Chromatography; **Spectroscopy**: Visible Spectroscopy and Colorimetry.)

Instrumental chromatographic techniques have now become the methods of choice, with separation and detection being performed concurrently. Gas–liquid chromatography (GLC) using various packed- or open-tubular capillary columns, coupled with flame ionization detection, allows a quantitative analysis of individual tocopherols and tocotrienols. A unique attribute of capillary GLC is its recent success in separating the diastereoisomers of *all-rac*-α-tocopherol and the potential of stereoactive chiral columns to resolve all eight stereoisomers. (*See* **Chromatography**: Gas Chromatography.)

The free phenols can be analyzed by GLC, where elution occurs in order of increasing methylation ($\delta < \gamma < \beta < \alpha$), although it is preferable to derivatize the analytes to their more volatile trimethylsilyl ethers or alkyl esters to improve the chromatographic quality. However, coelution of the β- and γ-isomers is often a problem unless very long capillary columns are used. In the analysis of foods, a more significant

problem exists from interferences by other lipid constituents, predominantly sterols. It is therefore usual to fractionate the crude extracts by TLC or column chromatography prior to the assay in order to facilitate resolution and interpretation.

High-performance liquid chromatography (HPLC) offers several advantages over GLC by avoiding the need for multistage clean-up or derivatization techniques. Either normal- or reversed-phase modes can be used to achieve a successful separation. Important advantages of normal-phase chromatography are its ability to separate the β- and γ-isomers and its solvent compatibility with oil-rich vitamin E extracts. Nevertheless, reversed-phase analysis is technically more robust and remains in common use, as illustrated in Figure 2 for an infant formula analysis. It is also amenable to gradient techniques, allowing concurrent determination of tocopherols and tocopheryl acetate as well as degradation products such as tocopherol quinone. Recently, variants of micellar capillary electrophoresis (CE) have been applied to fat-soluble vitamin separations. Although representing an alternative to HPLC techniques, the application of CE to the hydrophobic vitamins is currently in its infancy. (*See* **Chromatography**: High-performance Liquid Chromatography.)

Direct UV detection is commonly employed for samples with reasonable vitamin E concentrations. The difficulty with this mode of detection is the potential for other materials to interfere and also the low extinction coefficients of the tocopherols and tocotrienols. There is consequently a regular need to increase the sophistication of sample preparation, which can cause problems with such labile analytes.

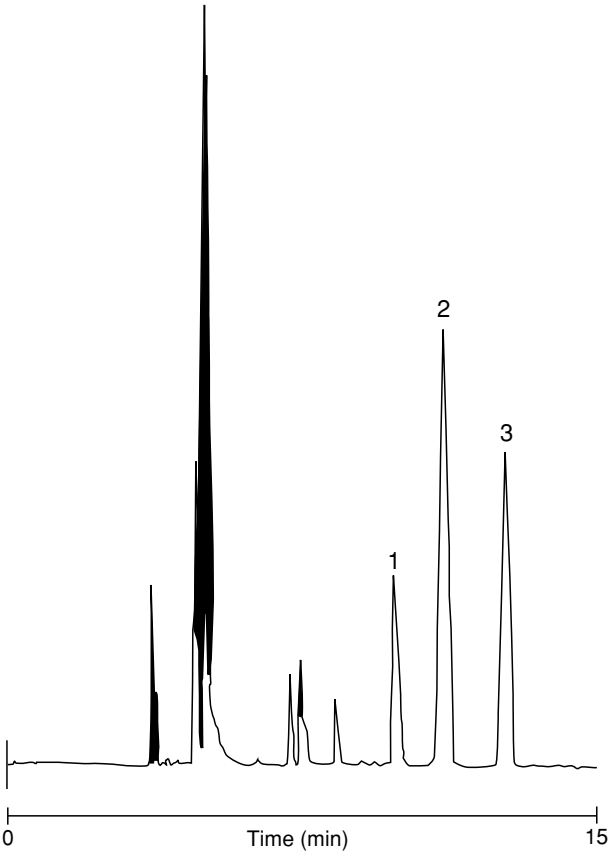

Figure 2 Reversed-phase HPLC chromatogram of a saponified fully oil-filled infant formulation. Waters Radial-PAK 5-μm C_{18} column with a mobile phase of methanol (100%). Flow rate, 1.0 ml min^{-1}. Detection by fluorescence at 295 nm (excitation) and 330 nm (emission). Peak identification: 1 = δ-tocopherol; 2 = β- and γ-tocopherol; 3 = α-tocopherol.

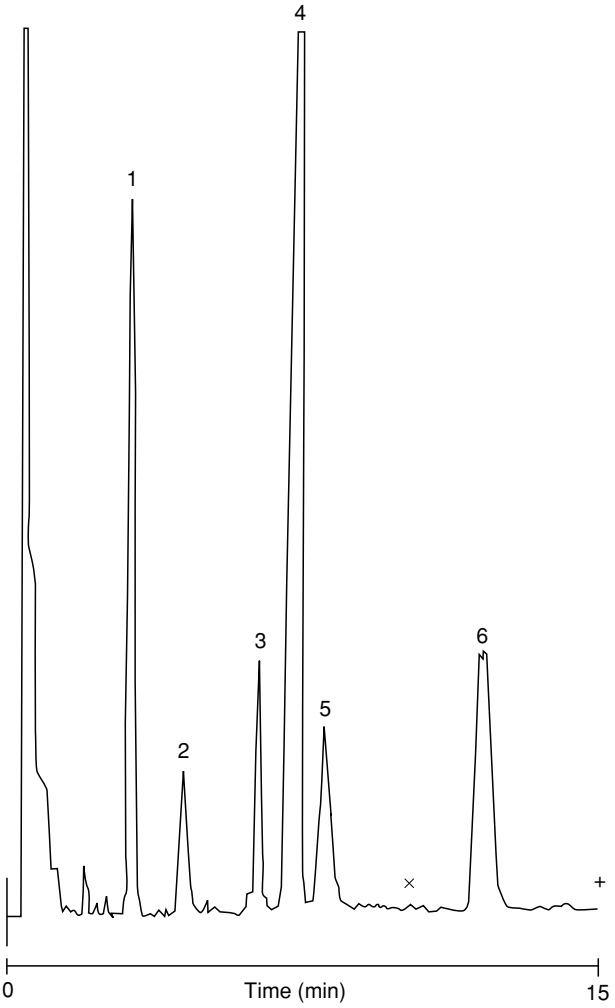

Figure 3 Normal-phase HPLC chromatogram of a saponified dietary formulation. Waters Radial-PAK 5-μm silica column with a mobile phase of 10% diisopropylether in hexane. Flow, 2.0 ml min^{-1}. Detection by fluorescence at 290 nm (excitation) and 325 nm (emission). Peak identification: 1 = α-tocopherol; 2 = α-tocotrienol; 3 = β-tocopherol; 4 = γ tocopherol; 5 = β-tocotrienol; 6 = δ-tocopherol; × = elution position of γ-tocotrienol; + = elution position of δ-tocotrienol.

Through the use of more specific detection modes, sample preparation can be further simplified together with substantial increases in sensitivity. Fluorescence detection at 292 nm (excitation) and 325 nm (emission) is a simple procedure with modern detectors, and encounters very few interference problems providing the separation is optimized. Excitation can be performed at 210 nm, but this is less common, except where spectral overlaps in the detector are unavoidable, as with filter-based instruments. Oxidative electrochemical detection (at 0.7–0.8 V) provides a further order of magnitude improvement in sensitivity compared with fluorescence but is less commonly used owing to operational difficulties and the need for a supporting electrolyte, incompatible with normal-phase separations. Similarly, recently reported LC-MS techniques have been restricted to research laboratories. (*See* **Spectroscopy**: Fluorescence.)

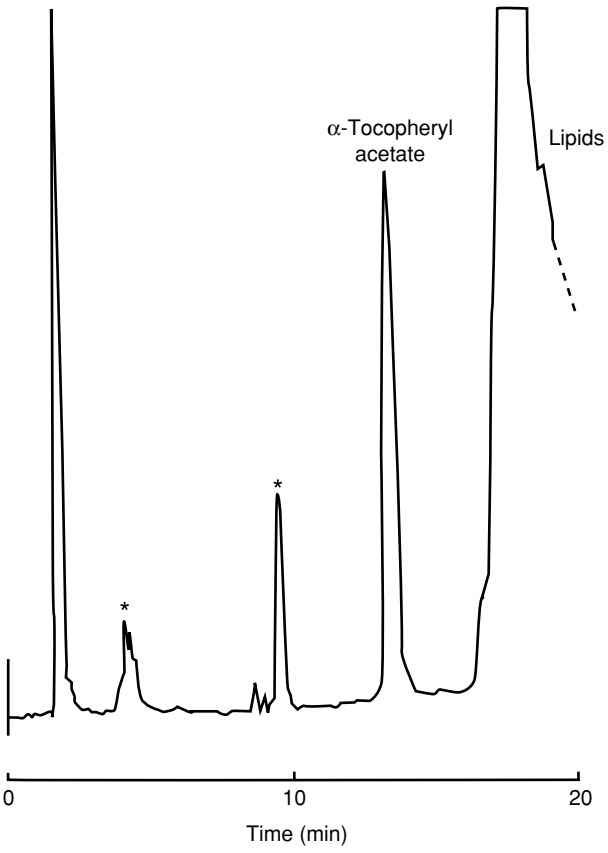

Figure 4 Normal-phase HPLC chromatogram of a fortified UHT milk without the use of saponification. Waters Radial-PAK 5-μm silica column with a mobile phase of 0.08% isopropanol in hexane. Flow rate of 2.0 ml min^{-1} is programed to 8 ml min^{-1} after 15 min to elute lipid components. Detection by UV at 280 nm at 0.01 aufs. The identification of α-tocopherol acetate is indicated. The retinyl esters (endogenous and supplemental) are also visible (*).

For most food products, it is generally necessary to precede any analysis, whether traditional or instrumental, with a clean-up stage aimed largely at removing bulk lipids, rupturing lipoprotein interactions and releasing supplemental vitamin E. Alkaline digestion (saponification), followed by solvent extraction of unsaponifiable material, is the usual technique employed where the free tocopherols are measured. If necessary, the vitamin-rich extract can then be concentrated by evaporation and subjected to further purification techniques. Figure 3 illustrates a complex food product analyzed under these conditions, using normal-phase HPLC. Six of the eight vitamin E congeners are visible, the elution order being α-T < α-T3 < β-T < γ-T < β-T3 < γ-T3 < δ-T < δ-T3.

Some simpler matrices, particularly oils, can alternatively be assayed by sample dilution, clarification and direct injection, whereas other high-fat products (e.g., dairy products) may lend themselves to simplified fat-extraction procedures without the need for a saponification step (Figure 4). This approach has the advantage of leaving any supplemental esters intact and eliminating most of the oxidative difficulties associated with the parent alcohols. Unfortunately, α-tocopheryl acetate has poorer UV absorption than α-tocopherol and greatly reduced fluorescence properties, whereas normal-phase chromatography is necessary in order to tolerate the high triglyceride loading that would quickly foul C_{18} columns. The alternative vitamin E additive, α-tocopheryl succinate, has no known fluorescence and is susceptible to peak distortion, both of which encumber chromatographic analysis, unless the sample is subjected to ion-suppression techniques or measured as α-tocopherol after saponification.

See also: **Antioxidants**: Natural Antioxidants; **Cells**; **Chromatography**: Principles; Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; **Dietary Requirements of Adults**; **Food Fortification**; **Margarine**: Methods of Manufacture; **Oxidation of Food Components**; **Spectroscopy**: Fluorescence; Visible Spectroscopy and Colorimetry; **Vegetable Oils**: Types and Properties; **Vitamins**: Determination

Further Reading

Ball GFM (1988) *Fat-soluble Vitamin Assays in Food Analyses*. London: Elsevier.

Bates CJ (2000) Vitamins: fat and water soluble: analysis. In: Meyers RA (ed.) *Encyclopedia of Analytical Chemistry*, pp. 1–35. Chichester, UK: Wiley.

Diplock AT (1985) Vitamin E. In: *The Fat-soluble Vitamins*, pp. 154–224. London: Heinemann.

Eitenmiller RR and Landen WO (1999) *Vitamin Analyses for the Health and Food Sciences*, pp. 109–148. Boca Raton: CRC Press.
Frankel EN (1998) *Lipid Oxidation*. Dundee: The Oily Press.
Friedrich W (1988) Vitamin E. In: *Vitamins*, pp. 217–283. New York: Walter de Gruyter.
Machlin LJ (1991) Vitamin E. In: Machlin LJ (ed.) *Handbook of Vitamins*, 2nd edn, pp. 99–144. New York: Marcel Dekker.
Nelis HJ, De Bevere VORC and De Leenheer AP (1985) Vitamin E tocopherols and tocotrienols. In: De Leenheer AP, Lambert WE and De Ruyter MGM (eds) *Modern Chromatographic Analysis of the Vitamins*, pp. 129–200. New York: Marcel Dekker.
Sheppard AJ, Prosser AR and Hubbard WD (1972) Gas chromatography of the fat-soluble vitamins: A review. *Journal of the American Oil Chemists Society* 49: 619–633.
Wagner AF and Folkers K (1975) *Vitamins and Coenzymes*, pp. 363–388. New York: Robert E. Krieger.

Physiology

W A Pryor, Louisiana State University, Baton Rouge, LA, USA

Background

Vitamin E is the most important fat-soluble antioxidant present in human tissues. It is found in the lipid-rich areas of cells, such as cell membranes and fat depots. Its role is thought to be the protection of polyunsaturated fatty acids (PUFAs) such as linoleic, linolenic, and arachidonic acids, which are present in lipids, against the types of oxidative damage to which these highly unsaturated fatty acids are so readily susceptible. (*See* **Antioxidants**: Natural Antioxidants; **Fatty Acids**: Properties.)

History

In 1923, Evans and Bishop reported that rats failed to reproduce if given a diet containing rancid lard, unless it was supplemented with a factor, called factor X, found in lettuce or whole wheat. Gradually, over the next 20 years, it became clear that factor X was associated with more than just reproductive failure in the rat. For example, a series of observations were reported linking this factor to encephalomalacia in the chick and nutritional muscular dystrophy in guinea-pigs and rabbits. In 1937, Olcott and Emerson reported that the tocopherols are antioxidants.

The first 15 years of the study of vitamin E were made difficult by the absence of a concentrated source of the vitamin or a knowledge of its structure and chemical make-up. However, in 1936, Evans and his group reported the isolation from wheatgerm of an 'alcohol' which had biological activity consistent with that of a tocopherol, and they suggested the correct chemical empirical formula for α-tocopherol $C_{29}H_{50}O_2$ (The free, unesterified tocopherols continue to be referred to as 'alcohols,' whereas they are actually phenols; in fact, alcohols are not, in general, antioxidants, whereas phenols are.) In 1938, the structure of α-tocopherol was deduced by Fernholz. The need for a pure, concentrated, and available source of the vitamin for nutritional studies was satisfied in 1939 when Karrer, of the Basel (Switzerland) laboratories of Hoffmann–La Roche, and then Smith and his group in America, reported the synthesis of the vitamin. (*See* **Phenolic Compounds**.)

The first international symposium on vitamin E was held in London in April 1939, and 15 papers were presented. In the current era, there have been major international symposia on the vitamin in various countries almost every year, each with 50 or 100 papers that appreciably add to our expanding knowledge of this fascinating vitamin.

Natural and Synthetic Forms of Tocopherol

The side-chain of the vitamin (see **Figure 1**) contains three carbons that are asymmetric; therefore, there are eight possible isomers of each of the tocopherols. Several forms of α-tocopherol are sold. One is a material that is derived from natural sources, called RRR-α-tocopherol, which has the stereochemical arrangement of methyl groups shown in **Figure 1**; another is the synthetic vitamin called all-*rac*-α-tocopherol, which is a mixture of all eight stereoisomers. Both the naturally derived material and all-*rac*-tocopherol are marketed as the free alcohol, the acetate ester, and the somewhat less bioavailable succinate ester. Some vegetable sources of tocopherols contain as much or more γ-tocopherol as α-tocopherol. The relative biopotencies of the various forms of tocopherol are shown in **Table 1**. Amounts of the tocopherols are expressed as either international units (IU) or equivalents of α-tocopherol (α-TE). The relationship between these terms is as follows: 1 mg of D-α-tocopherol (the naturally derived material) equals 1 α-TE; this is the unit used in *Recommended Dietary Allowances* (10^{th} edn.) by the Food and Nutrition Board of the National Research Council (National Academy Press, 1989). The IU is used in commerce for labeling

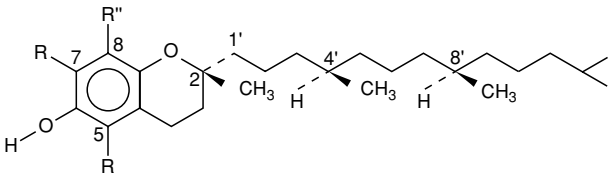

Tocopherol	Substitution pattern			Stereochemistry
	R	*R′*	*R″*	
α–	CH_3	CH_3	CH_3	2R,4′R,8′R
β–	CH_3	H	CH_3	2R,4′R,8′R
γ–	H	CH_3	CH_3	2R,4′R,8′R
δ–	H	H	CH_3	2R,4′R,8′R

Figure 1 Structures and stereochemistry of the four tocopherols. Reproduced from Tocopherols: Physiology, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 1 Biopotency of various forms of tocopherols

	IU mg^{-1}	α-*TE mg*$^{-1}$
RRR-α-Tocopherol	1.46	1.00
RRR-α-Tocopherol acetate	1.36	0.91
RRR-α-Tocopherol succinate	1.21	0.81
All-*rac*-α-Tocopherol	1.10	0.74
All-*rac*-α-tocopherol acetate	1.00	0.67
All-*rac*-α-tocopherol succinate	0.89	0.59
RRR-*β*-Tocopherol	0.30	0.20
RRR-γ-Tocopherol	0.15	0.10
RRR-δ-Tocopherol	0.01	0.01

Adapted from Machlin LJ (ed.) (1991) *Handbook of Vitamins*. New York: Marcel Dekker.

vitamin preparations and food, and is defined as follows: 1 mg of D-α-tocopherol equals 1.46 IU. (*See* **Dietary Requirements of Adults.**)

Vitamin E Requirements

A recommended dietary allowance (RDA) for various food factors was first recognized in the USA with the publication of the first edition of the book on RDA requirements in 1946. A specific RDA for vitamin E was first recommended in the sixth edition (1964), in which the concept of the RDA was broadened and redefined from the 1946 idea – enough 'to insure good nutrition' – to an amount necessary 'to permit full realization of ... potential.' Thus, in the 1964 edition, the concept was expressed that vitamins might have a pharmacological use beyond that necessary merely to prevent vitamin deficiency diseases. The 1997 edition sets the need for vitamin E as 10 α-TE per day for males and 8 mg for females, which is equal to 10 mg of the natural RRR-α-tocopherol (see Table 1).

The vitamin E requirement for humans increases as their diet includes more PUFAs because of the susceptibility of PUFAs to undergo autoxidation. A value of 0.4 for the ratio of the intake of RRR-α-tocopherol (in mg) to the intake of PUFA (in g) in the diet has been suggested to be adequate for adult humans. Thus, if a person eating a 2000-cal per day diet consumed 35% of their calories as fat, of which 40% was unsaturated (typical of today's American diet), they would require a daily intake of about 13 mg of α-tocopherol.

The intake of vitamin E for Americans on typical diets has been estimated to vary between 5 and 20 α-TE. The requirement of 10 α-TE for men and 8 α-TE for women (who are generally smaller) was established as the RDA because it is believed that very few Americans have overt vitamin E deficiency symptoms, and the amount ingested in a 'normal' diet must therefore be sufficient. This rather unsatisfactory, backward and *ad hoc* reasoning is thought to be the only approach possible for establishing an RDA for vitamin E, for which deficiency symptoms are only obvious in humans who have severe malabsorption. There is an increasing body of evidence, however, that vitamin E may provide benefits (such as protection from chemically induced cancers, cataracts, and ischemic heart disease) that require higher daily intakes; scientific studies of the usefulness of vitamin E in these contexts generally use amounts of vitamin E up to about 800 mg per day, which is generally agreed to be safe for humans. (*See* **Cancer:** Diet in Cancer Prevention.)

Metabolism

As more tocopherol is ingested, a smaller fraction is absorbed, such that the uptake of tocopherol by tissues varies with the logarithm of the tocopherol intake. Thus, ingesting four times the amount of the vitamin will raise tissue levels by only twofold. Triglycerides enhance tocopherol absorption, and PUFAs inhibit it. If the acetate or succinate is ingested, it is hydrolyzed prior to absorption through the gut wall. Since the hydrogen atom of the alcohol group is intimately involved in the antioxidant properties of the vitamin, it is not unexpected that only the free alcohol is biologically active. For high doses, a higher blood level is achieved using the free alcohol than the acetate. Absorption of vitamin E is maximal in the median portion of the small intestine, and none is absorbed in the large intestine. Vitamin E is thought to be absorbed through the gut epithelial cells as a lipid–bile micelle, together with fatty acids,

monoglycerides, and other fat-soluble substances. The retention of α-tocopherol is much greater than that of γ-tocopherol.

The vitamin circulates in the lymph and blood bound to lipoproteins, but, unlike some fat-soluble substances such as vitamin A, no specific carrier protein for vitamin E has been discovered. There is a high correlation between the total fat and the tocopherol concentrations in blood serum. Thus, diseases associated with high serum lipids (hypothyroidism, diabetes, hypercholesterolaemia) produce high plasma vitamin E levels, whereas those with low serum lipids (abetalipoproteinemia, malnutrition, cystic fibrosis) produce low vitamin E levels. (*See* **Cystic Fibrosis**; **Hyperlipidemia (Hyperlipidaemia)**; **Malnutrition**: The Problem of Malnutrition; **Retinol**: Physiology.)

Different tissues have widely differing vitamin E contents, as shown in Table 2. The vitamin is most concentrated in cellular fractions that are rich in membrane lipids such as the mitochondria. The vitamin E content of red blood cells is about 20% of that in plasma, and there is a rapid exchange between these two pools.

Less than 1% of vitamin E is metabolized; however, there are reports of oxidation products of the vitamin, such as the quinone and hydroquinone, being found in tissues such as the liver and the lung. Tocopherol is mainly excreted unchanged in feces. Very little is excreted in the urine.

Biochemical Function

Unlike most vitamins, there is no specific, enzymatically controlled reaction for which vitamin E is known to be a required cofactor. Instead, it is generally agreed that vitamin E acts as an antioxidant, protecting tissue against the nonenzymatic process of air oxidation of PUFA, as suggested by Olcott in 1941 and extended and elaborated by Dam in 1957 and by Tappel in 1961. (*See* **Coenzymes**.)

Evidence for the antioxidant theory of the function of vitamin E can be found in the following observations. The more PUFA in the diet, the more vitamin E is required. Animals on vitamin-E-deficient diets and subjected to oxidative stress conditions (such as high ozone levels) develop indirect evidence of the presence of peroxidic products in their tissues. Animals exposed to stressful oxidizing conditions are protected by increased intakes of vitamin E; examples of oxidative stress include smog, cigarette smoke, and some types of chemical carcinogens. Synthetic antioxidants that are structurally unrelated to vitamin E (e.g., ethoxyquin and diphenyl-*p*-phenylene diamine) can be used to replace it in diets. Finally, there is some evidence that other dietary antioxidants (such as vitamin C and selenium) can partially ameliorate low intake levels of vitamin E. (*See* **Ascorbic Acid**: Physiology; **Selenium**: Physiology.)

Table 2 α-Tocopherol content of human tissues

Tissue	$\mu g\, g^{-1}$
Plasma	9.5
Red blood cell	2.3
Fat	150
Adrenal	132
Heart	20
Testis	40
Liver	13

Adapted from Machlin LJ (ed.) (1991) *Handbook of Vitamins*. New York: Marcel Dekker.

Vitamin E Deficiency

Overt deficiency of vitamin E in the absence of nutritional deprivation is rare in humans, although animals can be made vitamin-E-deficient by feeding them synthetic diets that contain fats from which vitamin E has been removed. A number of diseases are known in various animals to result from a deficiency of vitamin E; perhaps the most notable feature of these deficiency states is the very great species specificity of the manifestations of vitamin E deficiency. Notable among the pathologies found are muscular degeneration, encephalomalacia, and, as discussed above, problems with gestation in some animals (e.g., rats but not mice).

Vitamin E deficiency in humans is generally measured by the time to observe hemolysis of red blood cells when the blood is exposed to hydrogen peroxide. In this test, the ability of vitamin E to protect red blood cell membrane lipids from undergoing destructive oxidation is measured. Normal levels of vitamin E in adult human blood serum average $1.05\ mg\ dl^{-1}$, but the range is very wide.

The first documented cases of vitamin E deficiency in humans were in patients with abetalipoproteinemia. These patients develop severe neurological abnormalities during their first two decades of life. Supplements of vitamin E (generally 100 mg per kg of body weight per day) can prevent the disease if started early enough in life. Children with cholestatic liver disease also develop neurological abnormalities due to inadequate delivery of vitamin E to the affected tissues. Studies in hyperlipidemic rats demonstrate that the transfer of vitamin E to tissues is impaired. (*See* **Liver**: Nutritional Management of Liver and Biliary Disorders.)

The ratio of maternal and fetal blood tocopherol levels is about 5, suggesting poor placental transport of vitamin E. For this reason, premature infants have

low vitamin E levels. Premature infants are often given elevated oxygen therapy, since their lungs are immature, and this places an added oxidative stress on their systems. Since modern medicine allows smaller premature infants to be saved than in the past, there has been an increase in diseases in these infants that may be ameliorated by vitamin E (such as intraventricular hemorrhage). Premature infants are now routinely given supplemental vitamin E.

PUFA Autoxidation and the Role of Vitamin E

Polyunsaturated fatty acids have methylene groups (—CH_2— groups) located between two double bonds, as shown in Figure 2. This type of chemical functionality makes these compounds particularly sensitive to the flameless oxidation in air that is called 'autoxidation.' Figure 2 shows the steps involved in

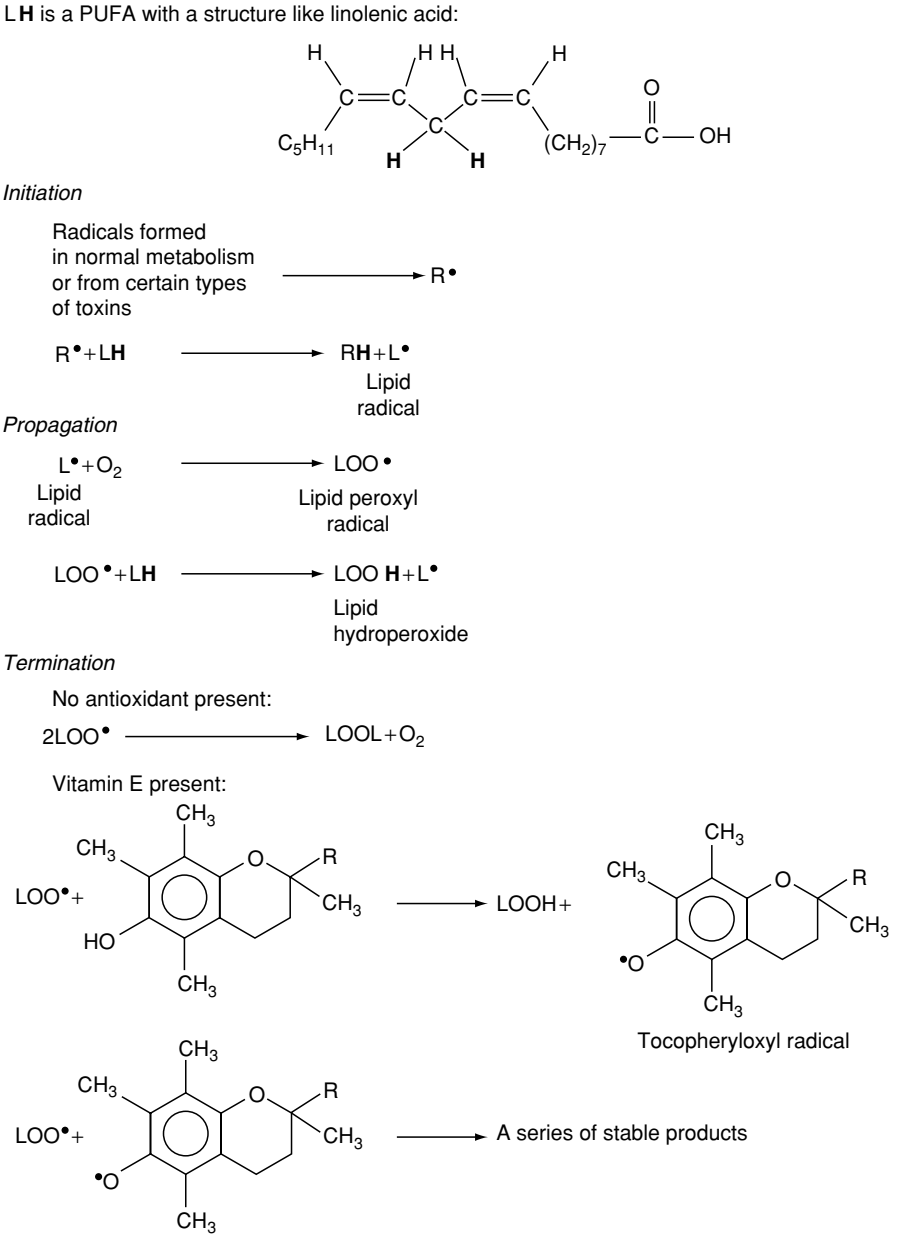

Figure 2 Chemical reactions involved in the autoxidation of a particular polyunsaturated fatty acid, linoleic acid. The autoxidation sequence involves three different types of reactions: initiation, propagation, and termination. Initiation is the process in which radicals are produced; propagation steps are those that occur over and over in the chain reaction and where products are formed; and termination steps are those in which radicals are destroyed. Reproduced from Tocopherols: Physiology, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

this process. Free radicals are produced both in normal metabolism and by the effects of toxins on tissue. These primordial free radicals can cause the formation of lipid-free radicals, as shown in the *initiation sequence*. The lipid radical, $L^{\cdot}$, then undergoes the *propagation sequence* shown, leading to the formation of lipid hydroperoxides, LOOH. These LOOH compounds also are formed in enzyme-catalyzed reactions; for example, the enzymes in the arachidonic acid cascade serve to convert arachidonic acid (the acid with 20 carbon atoms and four double bonds) to several types of hydroperoxides that have potent biological properties, such as 15-hydroperoxyeicosatetraenoic acid (15-HPETE). If vitamin E is not present, two of the peroxyl radicals ($LOO^{\cdot}$) involved in the propagation sequence must find each other and collide in order to convert radicals to stable, nonradical products. This reaction is shown in the *termination sequence*. If vitamin E is present, it reacts extremely rapidly with peroxyl radicals, converting them to nonradical products and stopping the autoxidation chain reaction. One tocopherol molecule can protect 100 or more PUFA molecules from autoxidative damage.

See also: **Antioxidants**: Natural Antioxidants; **Ascorbic Acid**: Physiology; **Cancer**: Diet in Cancer Prevention; **Coenzymes**; **Cystic Fibrosis**; **Dietary Requirements of Adults**; **Fatty Acids**: Properties; **Hyperlipidemia (Hyperlipidaemia)**; **Liver**: Nutritional Management of Liver and Biliary Disorders; **Malnutrition**: The Problem of Malnutrition; **Phenolic Compounds**; **Retinol**: Physiology; **Selenium**: Physiology

Further Reading

Ciba Foundation (1983) *Biology of Vitamin E*. London: Pitman Books.

Combs GF (1998) *The Vitamins: Fundamental Aspects in Nutrition and Health*. London: Academic Press.

Diplock A, Machlin L, Packer L and Pryor W (eds) (1989) *Vitamin E: Biochemistry and Health Implications*. New York: New York Academy of Sciences.

Frei B (ed.) (1994) *Natural Antioxidants in Human Health and Disease*. San Diego, CA: Academic Press.

Hayaishi O and Mino M (eds) (1987) *Clinical and Nutritional Aspects of Vitamin E*. Amsterdam: Elsevier.

Lubin B and Machlin L (eds) (1982) *Vitamin E: Biochemical, Hematological and Clinical Aspects*. New York: New York Academy of Sciences.

Machlin L (ed.) (1980) *Vitamin E: A Comprehensive Treatise*. New York: Marcel Dekker.

Machlin LJ (ed.) (1991) *Handbook of Vitamins*. New York: Marcel Dekker.

Packer L and Fuchs J (eds) (1993) *Vitamin E in Health and Disease*. New York: Marcel Decker.

Pryor WA (1984) Free radicals in autoxidation and aging. In: Armstrong D, Sohal RS, Cutler RG and Slater TF (eds) *Free Radicals in Molecular Biology, Aging and Disease*, pp. 193–201. New York: Raven Press.

Pryor WA (2000) Vitamin E and heart disease: Biochemistry to human trials. *Free Radical Biology & Medicine* 28(1): 141–164.

Sauberlich HE and Machlin LJ (1992) *Beyond Deficiency: New Views on the Function and Health Effects of Vitamins*. New York: New York Academy of Sciences.

Tocotrienols *See* **Tocopherols**: Properties and Determination; Physiology

TOMATOES

B McGlasson, University of Western Sydney, Penrith South DC, New South Wales, Australia

Introduction

Tomatoes are an integral part of human diet worldwide and production in 1999 was estimated at 91.7×10^6 t. The tomato is a member of the Solanaceae – the potato family. Tomato fruit, often described as a vegetable fruit, is consumed fresh and in several processed forms, including dried, purée, paste, ketchup, sauce, soup, and canned whole-peeled. It is a watery fruit containing 5–7% dry matter. It contains relatively low concentrations of vitamin C, provitamin A and minerals, compared to other commercially important fruit species. However, because it is consumed in quite large quantities, it is a major source of these nutrients, especially in western societies. There is renewed interest in the dietary properties of tomato because it is the principal source of lycopene.

Lycopene is an antioxidant that is thought to be a factor in the prevention of several chronic diseases, including cancer and coronary heart disease. There is evidence that lycopene is absorbed more efficiently from processed tomato products compared to fresh fruit. Genetically, the tomato is probably the most highly defined and best mapped of all domesticated soft-fruited crop plants. An enormous range of commercial cultivars is available. There are cultivars bred specifically for the fresh market that produce fruit ranging in diameter from 15 mm to more than 90 mm, and processing types especially suited to machine harvesting. All modern cultivars incorporate multiple resistances to many pathogenic diseases and pests.

Genetic Origins

It is thought that the original domestication of the tomato took place in Mexico and probably resulted from selection over many generations from the ancestral form now known as *Lycopersicon esculentum* var. *cerasiforme*, the cherry tomato. The genetic center of origin of the tomato is believed to be the western coastal strip of South America which extends about 80 km inland and to altitudes of 3000 m. This coastal strip extends from Ecuador to Peru and Chile (latitude 0–25 °S). Geneticists, notably the late Professor Charles Rick of the University of California, Davis, USA, have collected, classified, and documented many wild species of *Lycopersicon* and other closely related genera. Some of these wild species have been sources of resistance against a wide range of diseases and pests, which have been incorporated into modern cultivars. The wild species include ecotypes that are adapted to low and high growth temperatures, arid and humid environments, and saline soils. However, some of these species cannot easily be crossed with the domesticated tomato (Table 1).

Genetic Mutants

Many natural and some induced mutants have been documented, and some of these have been shown to have commercial application. The tomato is typically diploid, comprising 12 pairs of chromosomes ($2n = 24$). Excellent gene linkage maps have been developed by classical plant breeding using defined genetic marker stocks and isozyme markers. Molecular genomic maps have been derived from restriction length polymorphism analysis. Extensive physiological and biochemical research, especially on the ripening-impaired mutants *rin*, *nor* and *Nr*, during the 1970s, provided the basis and the stimulus for the considerable advances in knowledge of the molecular biology of the genetic regulation of fruit ripening and cell wall softening that have been achieved during the last 20 years.

Since about 1980, the tomato has become a favorite model system for gene technologists. The tomato is an excellent experimental subject because it is easy to

Table 1 The species of *Lycopersicon*

Subgroup	*Characteristics*
esculentum complex	Hybridize readily
esculentum	Normal tomato
esculentum var. *cerasiforme* (cherry tomato)	Adapted to wet, tropical conditions
cheesemanii	Yellow to orange fruit, diameter 6–9 mm Source of jointless gene, aid to machine harvesting
hirsutum and its forms	Green fruit with purplish stripes, diameter 1.5–2.5 cm Source of cold tolerance, and resistance to several insect pests, root knot nematodes, some fungi and bacteria
parviflorum and *chmielewski*	Yellow-green fruit, diameter 1–1.4 cm *chmielewski* is a source of high solids
pennellii	Small, green fruit Initially classified in the genus *Solanum* Resistant to drought and some sucking insects
pimpinellifolium (currant tomato)	Resistant to *Fusarium* and bacterial speck
peruvianum complex	Do not hybridize readily with *esculentum* complex
chilense	Green fruit, diameter 1–2 cm Source of the $Tm2^a$ gene Resistance to the tobacco mosaic virus
peruvianum	Green fruit, diameter 1–3 cm Very diverse species which includes coastal and mountain races Potentially valuable source of genes for resistance to many fungal and viral diseases and insect pests

Adapted from Taylor IB (1986) Biosystematics of the tomato. In: Atherton JG and Rudich J (eds) *The Tomato Crop*, pp. 1–34. London: Chapman and Hall.

transform and regenerate in tissue culture. The plants can be grown easily under protected growing conditions year-round and fruit mature within about 3 months from sowing. Genetic transformation and gene technology have been used to reduce the amounts of cell wall-softening enzymes and ethylene production to slow ripening, and to introduce genes for resistance to glyphosate herbicide. (*See* **Pesticides and Herbicides**: Types, Uses, and Determination of Herbicides.)

Fruit Composition

Botanically, the tomato fruit is classified as a berry. The fruit can be divided into skin, pericarp, and locular contents. Jelly-like parenchyma cells which fill the locular cavities surround the relatively large number of hard, small seeds borne per fruit. Small-fruited cultivars, such as those grown for the fresh trade in cool climates or for machine harvesting for processing, generally contain few locules, whereas large-fruited cultivars tend to be multilocular. The locular juices contain higher concentrations of organic acids than the outer locular walls but the reverse is true for the reducing sugars. Blended tissue has a pH of 4.0–4.7.

Table 2 shows data for a fresh-market cultivar representative of modern determinate types grown outdoors in countries such as the USA and Australia. There is considerable variability in the concentration of nutrients, which are affected by genetic differences and especially by irradiance, growth temperatures, irrigation, and salinity. The fresh tomato fruit contains a comparatively low concentration of dry matter. The edible portion of the fruit comprises 99% of the fresh weight. The inedible portion includes the seeds and the skin which are removed during the production of the purée, paste, and soup.

Because of the relatively large amounts of fresh and processed products consumed, the tomato is an important source of vitamin C, β-carotene, and minerals such as potassium in the human diet. Standard commercial cultivars contain predominantly the reducing sugars glucose and fructose. Although sucrose is the main metabolite translocated in the tomato plant, acid invertase, present in the vacuoles of the fruit cells, insures that all but trace amounts of sucrose are hydrolyzed. However, some species other than *L. esculentum* are known to accumulate a significant proportion of sucrose. *L. chmielewski*, which lacks vacuolar acid invertase, has been crossed with *L. esculentum* to produce high-sugar lines. However, breeding of commercial cultivars with high soluble solids has proven difficult because sugar metabolism appears to be controlled by several quantitative traits. Green tomatoes store a small quantity of starch, which reaches a maximum when the fruit are about half-grown. Starch is hydrolyzed rapidly once ripening begins.

Table 2 Nutrient composition of tomato

Nutrient	*Content (per 100 g edible portion[a])*
Energy[b] (kJ)	56
Gross constituents (g)	
Water	94.7
Protein[c]	1.0
Fat	0.1
Dietary fiber	1.6
Carbohydrates (g)	
Glucose	0.9
Fructose	1.0
Sucrose	0
Starch	0
Organic acids (g)	
Citric	0.43
Malic	0.08
Oxalic	0
Other	0
Vitamins (mg)	
Vitamin C	18
Thiamin	0.04
Riboflavin	0.02
Nicotinic acid	0.7
β-Carotene (equivalent)[d]	0.34
Minerals (mg)	
Potassium	200
Sodium	6
Calcium	8
Magnesium	10
Iron	0.3
Zinc	0.2

[a]Edible portion is 99% of the fruit.
[b]Energy was calculated as g((protein × 17) + (fat × 37) + (monosaccharides × 16) + (disaccharides × 16.8) + (starch × 17.6) + (organic acids × 10)).
[c]Protein was calculated as %N × 6.25.
[d]β-Carotene was calculated as mg β-carotene + 0.5 × mg (β-carotene + cryptoxanthin).
Adapted from Wills RBH (1987) Composition of Australian fresh fruit and vegetables. *Food Technology in Australia* 39: 523–526 with permission.

Detailed information is available on the concentrations of (1) organic acids, in addition to citric and malic, that are involved in the tricarboxylic acid cycle of aerobic respiration, and (2) the enzymes which catalyze this cycle and related metabolic pathways. The principal phenolic compound found in tomato fruit is chlorogenic acid. Alkaloid glycosides are common throughout the Solanaceae. These compounds taste bitter. α-Tomatine is commonly found in the tomato. Quite low concentrations are found in unripe fruit and these decrease by about half during ripening. In contrast to solanine in potato, α-tomatine appears to pose a low risk of toxicity to people.

Detailed information on the free amino acid composition of fresh fruit is available. Glutamic acid is the major component. Glutamic, aspartic, and γ-aminobutyric acids, plus glutamine, comprise

about 80% of the total free amino-nitrogen-containing compounds in fruit. These compounds contribute to tomato taste.

Pigments

Traditionally, tomatoes are thought to be red, but pink, orange, yellow, white and even black strains occur naturally. The typical red color is caused by Lycopene, which is usually the predominant pigment. Lycopene is a carotenoid pigment, but it is not converted to vitamin A when consumed by people. Lycopene is more concentrated in the outer fruit wall (pericarp) than in the internal locular tissue. The distribution of β-carotene follows a reverse pattern.

During ripening, the chloroplasts are transformed into chromoplasts. Chlorophyll disappears completely within about 4 days at 20 °C, and lycopene reaches a maximum by about 6 days.

There are relatively small increases in the carotene pigments during ripening. Lycopene synthesis is optimum at 13–22 °C. Lycopene concentrations range from 2 to 7 mg per 100 g edible portion. At higher temperatures, synthesis is inhibited and is completely stopped above 30 °C. The fruit color is then yellow. There are many mutants that affect pigment synthesis, including some which retain chlorophyll when the fruit is ripe, e.g., Evergreen and Greenflesh. The crimson gene (og^c) increases lycopene concentrations at the expense of β-carotene, and the high-pigment gene (*hp*) increases total carotenoids, thus enhancing the color of ripe fruit. (*See* **Carotenoids**: Occurrence, Properties, and Determination.)

Production Systems

Domesticated cultivars display two kinds of growth habit – indeterminate and determinate. Indeterminate or vine types typically continue to grow, flower, and set fruit as long as the plants remain healthy, and growing conditions are suitable. These types require support. They may be trained on stakes or trellises, or suspended on twine from overhead wires. In general, a considerable amount of labor is required to prune and train the plants. Indeterminate cultivars for the fresh market are used in commercial glass or plastic houses, and yields of 250 t ha^{-1} have been reported in a full growing season in the UK. A high proportion of fresh market tomatoes produced in glass or plastic houses are grown hydroponically. Plants are either grown in a solid medium with the nutrients being allowed to run to waste or in a recirculating aqueous medium. This system enables the mineral composition and concentration of salts in the medium to be precisely controlled.

Most of the tomatoes commercially grown outdoors have a determinate habit. This habit is controlled by *sp*, the self-pruning gene. The plants have a compact bush habit and set fruit over a brief period of about 1 month. Determinate fresh-market tomatoes may be grown as a ground crop in dry climates or may be staked or trellised in regions with relatively high rainfall. Production on stakes or trellises facilitates spraying to control foliage diseases and pests. Little or no pruning or training is required. Recently some outdoor growers have replaced determinate cultivars with indeterminate or 'gourmet' cultivars. This has been made possible by the advent of improved cultivars incorporating the *rin* gene. Fruit of these cultivars remain firm when ripe and can be allowed to ripen to an advanced color stage (vine-ripe) before they are harvested. Gourmet tomatoes are expected to have better flavor because they are harvested when ripening is well underway and they may have higher sugar levels compared to determinate cultivars because they have higher leaf–fruit ratios. Tomatoes for processing are universally grown as a ground crop on raised beds. Irrigation by trickle is well suited to outdoor tomato production. Water and nutrients can be applied efficiently to the root zone according to the demands of the crop.

Most fresh-market tomatoes are harvested by hand. Fruit at early color stages is harvested at frequent intervals from trellised plants. Some harvest machines have been developed but they have not been widely adopted because of the risk of excessive fruit damage and limited cost savings. Mechanical harvest aids have been developed and are more widely used. These machines transport pickers along the rows and convey the harvested fruit into bulk bins carried on the machine. On-ground plantings are harvested up to four times, and a high proportion of the fruit is harvested green. Green fruit is usually held in controlled-temperature ripening rooms with the continuous addition of ethylene to reduce the time to the commencement of ripening.

In countries where labor costs are high, processing tomatoes are harvested by machine when most of the fruit is fully ripe. These machines cut the bushes at ground level, elevate the bushes into a shaker section to separate the fruit from the bushes, and then present the fruit either to an automatic color sorter, or to people riding on the machine who remove clods of soil and unripe or overripe fruit by hand. The sorted fruit is accumulated in large bins, traveling with the harvest machines, and then carried by road to the factory.

The development of the cultivars and the machine harvesters was a revolutionary step pioneered by the late GC Hanna of the University of California, Davis,

USA, in the 1950s. Cultivars with the following characteristics are required: plants are compact; fruit is set and ripened over a confined period; fruit stores well on the bush when ripe; fruit detaches easily and withstands the physical rigours of machine harvesting. Processors pay growers on the basis of total solids per tonne of fruit. The incorporation of the j^2 jointless gene in modern cultivars enables detachment of the fruit free from calyces and pedicels.

Fruit Growth and Ripening

Tomatoes take 6–7 weeks from flowering, depending on temperature, to reach full size. Cell division continues for about 2 weeks after flowering, but the bulk of the increase in fruit size is the result of cell expansion. In normal cultivars, the first appearance of red or pink color at the blossom end of the fruit signals the completion of growth and the beginning of ripening. Laboratory studies with fruit harvested at a mature green stage have shown that ripening actually begins about 2 days before the external color change. An early indication of ripening is a small increase in ethylene production, which can be measured with a sensitive gas chromatograph.

The tomato is a climacteric fruit in which ripening is accompanied by an increase in both respiration and ethylene production (Figure 1). A natural increase in endogenous ethylene production initiates ripening and regulates or integrates many of the biochemical events in ripening. The disappearance of starch, the destruction of chlorophyll, and the synthesis of lycopene, aroma, and polygalacturonase, a cell wall-hydrolyzing enzyme, are highly integrated with the changes in respiration and ethylene production. These events either do not occur or are diminished and protracted in the ripening-impaired mutants *rin, nor, Nr*, and *alcobaca*. Fruit produced by F_1 hybrids of mutant lines and normal cultivars has intermediate traits. The fruit ripens but may not develop acceptable color and flavor. The severity of these effects depends on the mutant gene and the genotypic background of the parent lines. However, many cultivars incorporating *rin* are being grown for the fresh market. The benefits include a slower rate of ripening and softening, so that the fruit has a longer shelf-life. Incorporation of ripening-impaired mutant genes in cultivars for processing appears to have no benefits.

Storage Temperatures

Tomatoes are chilling-sensitive. Storage of green fruit at less than 10 °C for more than a few days inhibits ripening. Severely chilled fruit will not ripen when returned to normal ripening temperatures. In less severely chilled fruit, ripening may be retarded but the fruit is much more susceptible to pathogenic decay. The recommended storage and transport temperature for tomatoes is 12 °C. Carefully handled fruit, harvested at an early color stage, may be stored for up to 3 weeks. Green and partly colored fruit should be ripened at 13–22 °C to the desired stage. Ripe fruit may be stored satisfactorily for about 4 days at about 5 °C. (*See* **Chilled Storage: Principles.**)

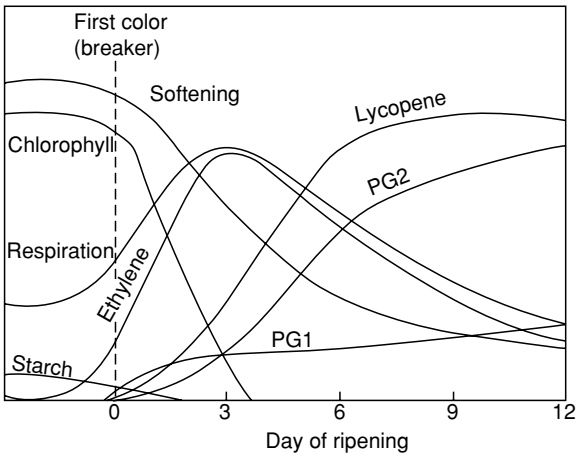

Figure 1 Changes in metabolism and composition during ripening. PG, endopolygalacturonase. Reproduced from Tomatoes, Encyclopaedia of Food Science, Food Technology and Nutrition, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Sensory Quality

Many countries have quality standards for fresh-market tomatoes. The criteria include size, color, and freedom from blemishes, decay, and insect pests. There are no dramatic changes in gross components, such as sugars and acidity, that enable the establishment of objective quality standards. The pressure of competition, on growers in countries such as the USA and Australia, to reduce production costs led to the adoption of large-scale production of determinate cultivars which soften slowly and have a long shelf-life (Figures 2–5). Fruit of the first introductions of these new cultivars sometimes retained a tough woody texture, even when fully colored. Even if tomatoes have optimum levels of sugars and acids, a tough texture greatly reduces their acceptability to consumers. The universal criticism of fresh tomatoes is tough texture and lack of flavor. This has led to an increased production of cherry and gourmet cultivars that are harvested at advanced color stages and generally sold at a premium price compared to standard cultivars. (*See* **Sensory Evaluation: Practical Considerations.**)

Figure 2 (see color plate 134) Harvest aid for fresh market tomatoes. Fruit are sorted and accumulated on a hopper on the machine. The fruit are periodically transferred to half-tonne bins for transport to the packing house.

Figure 3 (see color plate 135) Mechanical harvesting of processing tomatoes. Plants are cut at ground level, and elevated on to a shaker which separates the fruit from the plants.

Flavor of tomatoes is a product of the interaction of taste and aroma. Research has shown that taste is the more important factor. Taste principally involves an interaction between total sugars and titratable acidity. Most modern cultivars appear to have adequate acidity, but sugar levels are frequently too low. Other taste enhancers include the amino-nitrogen compounds, especially glutamate, and other unidentified compounds which contribute to 'tang.' Measurement of total soluble solids by refractometer gives a rough indication of the level of taste components because sugars comprise more than half of the soluble solids. However, refractometric values are often not well correlated with taste. Incorporating titratable acidity (mmol per 100 ml of juice) in a multiple regression equation improves the correlation, but it remains

Figure 4 In Australia fresh-market tomatoes are mostly volume-filled into 10-kg cartons for marketing.

Figure 5 (see color plate 136) Tray packs of glasshouse tomatoes, harvested once ripe.

unsatisfactory. Other components that contribute to taste remain to be identified. Improved analytical procedures allow direct measurements of the major taste components but are too expensive for routine use.

Despite these limitations, practical observations show that raising total soluble solids from the common level of 4.0–4.5% to over 6.0% gives a highly significant improvement in flavor scores. Cultivars with the genetic potential to accumulate high sugar concentrations and, consequently, higher total soluble solids are essential, but it is necessary to optimize growing conditions to enable full expression of this genetic potential. Research in several countries has examined the feasibility of raising the electrical conductivity of the irrigation solution, especially

during the fruit growth phase, either by the addition of salts, such as potassium chloride, in outdoor production, or by the addition of sodium chloride to hydroponic systems in glass or plastic houses. Ultimately the best way to increase sugar levels will be by plant breeding but this goal remains elusive because the regulation of sugar concentrations involves several quantitative genetic traits.

The volatile compounds that contribute to aroma have received much attention. Green tomato fruit produces low levels of volatile compounds, but more than 400 volatile compounds have been detected in ripe fruit. While a considerable number of these compounds contribute to tomato aroma, it is probable that only about 15 compounds have an influence on flavor quality. They include alcohols, aldehydes, carbonyls, and sulfur compounds. A few compounds may produce off-flavors but most evince a positive response. Tomatoes that have high levels of soluble solids tend to produce larger amounts of aroma.

The flavor of tomatoes is changed considerably by processing. An essential step in processing is denaturation of enzymes and sterilization by heating to about 100 °C. Although the major taste components remain, volatile aroma compounds associated with fresh fruit are lost. Additional flavor compounds are generated nonenzymatically during processing. A large proportion of processing tomatoes are concentrated in the form of paste which contains a minimum of 24% total solids. Paste can be stored for long periods before conversion to sauce, ketchup, and soup.

Cell-Wall Softening

Cell walls comprise α-cellulose, pectin, hemicellulose, and some protein. During ripening, the fruit softens, and softening is accompanied by changes in the pectin and hemicelluloses. Cellulose remains intact. Several hydrolases that attack cell walls are present in tomato fruit. Endopolygalacturonase (PG) has received particular attention and it is produced *de novo* during ripening (Figure 1). However, the extent of softening is not closely correlated with the amount of enzyme produced, although cultivars with firm fruit generally accumulate less enzyme than do cultivars with softer fruit. It appears that there are several factors that limit the action of PG *in vivo*. These factors may include the extent of methyl esterification of polygalacturonate, the amount of calcium bound to pectin, the distribution of PG in the cell walls, and the activity of other hydrolytic enzymes that act on cell wall components.

There are three closely related isozyme forms: PG1, PG2A and PG2B. PG1 comprises one subunit of PG2A or PG2B, plus a nonenzymatically active ancillary subunit. PG1 is the first form produced during ripening and may be the active form of the enzyme within the fruit. It has been suggested that the ancillary subunit plays a key role in binding the active PG2A or PG2B to its substrate. Normal cultivars continue to accumulate relatively large amounts of the small PG forms for several days after the fruit become fully ripe. Fruit of *rin* and *nor* produce little or no PG and soften very slowly. The amount of PG present in ripe fruit is especially important to processors. An important factor determining the yield of products such as ketchup, which is manufactured per tonne of raw fruit, is the viscosity (consistency) of the product, and the extent of depolymerization of pectin has a major influence on viscosity. When tomato tissue is comminuted, the vacuolar acids (principally citric acid) chelate calcium bound to pectin. The pH of the blended tissue is optimum for PG activity, and the protective calcium is removed from the pectins, which are rapidly solubilized unless the enzyme is denatured. A key step in paste manufacture is the 'hotbreak.' The tomato tissue is heated as rapidly as possible to 100 °C because PG remains highly active up to about 80 °C. Tomatoes transformed with DNA for PG in an antisense configuration have greatly reduced PG activity and yield products with higher viscosity.

Future Developments

Tomato breeders, in cooperation with plant physiologists, biochemists, and gene technologists, will continue to develop the commercial tomato. The tomato is an excellent model system for research on the genetic regulation of the biochemical events associated with ripening. Large catalogued international collections of wild tomato species are being maintained and researched. There is also enormous potential to improve resistance to pests and diseases, yield of solids for processing, and sensory quality. If further medical research confirms a role for lycopene in reducing chronic diseases in people, increased consumption of fresh and processed tomatoes selected for high levels of lycopene can be expected. A large amount of germplasm remains to be elucidated in species such as *L. peruvianum*, in which there are significant compatibility barriers with *L. esculentum*. Molecular biology provides the tools for overcoming these barriers.

See also: **Carotenoids**: Occurrence, Properties, and Determination; **Chilled Storage**: Principles; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Fruits of Temperate Climates**: Commercial and Dietary Importance; **Ripening of Fruit**

Further Reading

Atherton JG and Rudich J (eds) (1986) *The Tomato Crop*. London: Chapman and Hall.

Davies JN and Hobson GE (1981) The constituents of tomato fruit, the influence of environment, nutrition, and genotype. *CRC Critical Reviews in Food Science and Nutrition* 15(3): 205–280.

FAO www.fao.org Statistical data bases. Food and Agriculture Organization of the United Nations. Rome: Food and Agriculture Organization.

Fulton TM (ed.) (1998) *Report of the Tomato Genetics Cooperative*, vol. 48. Ithaca, NY: Department of Plant Breeding, Cornell University.

Giovannucci E (1999) Tomatoes, tomato-based products, lycopene and cancer: review of the epidemiologic literature. *Journal of the National Cancer Institute* 91(4): 317–331.

Gould WA (1983) *Tomato Production, Processing and Quality Evaluation*, 2nd edn. Westport, Connecticut: AVI.

Hobson G and Grierson D (1993) Tomato. In: Seymour GB, Taylor JE and Tucker GA (eds) *Biochemistry of Fruit Ripening*, pp. 405–442. London: Chapman and Hall.

Nevins DJ and Jones RA (eds) (1987) *Tomato Biotechnology. Plant Biology*. New York: Alan R Liss.

Seymour GB, Colquhoun IJ, DuPont MS, Parsley KR and Selvendrant RR (1990) Composition and structural features of cell wall polysaccharides from tomato fruits. *Phytochemistry* 29: 725–731.

Tigchelaar EC (1986) Tomato breeding. In: Bassett MJ (ed.) *Breeding Vegetable Crops*, pp. 135–171. Westport, Connecticut: AVI.

Wills RBH (1987) Composition of Australian fresh fruit and vegetables. *Food Technology in Australia* 39: 523–526.

Tongue *See* **Offal**: Types of Offal

Tooth Decay *See* **Dental Disease**: Structure of Teeth; Etiology of Dental Caries; Role of Diet; Fluoride in the Prevention of Dental Decay

TORTILLAS

S O Serna Saldivar, Instituto Tecnologico y de Estudios Superiores de Monterrey, Monterrey N.L., Mexico
L W Rooney, Texas A&M University, TX, USA

Introduction

A tortilla can be defined as a flat, round, unfermented bread produced from wheat (*Triticum aestivum* L.) flour or lime $Ca(OH)_2$ cooked maize (*Zea mays* L). Processes and characteristics of wheat flour and maize tortilla differ considerably. Both types of tortillas originated in Mexico, where they are considered the national bread and are generally consumed with others foods such as beans, meat, and vegetables. Tortillas are increasing in popularity throughout the world. Global sales were estimated at 6 billion dollars in 1998. Processing plants have been started in the UK, Spain, France, Australia, Brazil, India, China, Korea, and other countries around the world.

For wheat tortilla production, refined or wholemeal flour is mixed with shortening or lard, salt, baking powder, and other ingredients. The tortilla blend is mixed and kneaded with water to yield a gluten-developed dough that is divided into balls and hand-rolled or hot-pressed into a flat disk. The disk is baked on a hot griddle to yield the tortilla (**Figure 1**). Like bread dough, wheat flour tortilla dough is gluten-structured.

For maize tortillas, mature grain is cooked in water containing lime to produce nixtamal, which is ground

Figure 1 Wheat tortillas and lime-cooked maize products. Clockwise, from center top: corn on the cob; wheat tortillas; corn chips; a wheat burrito and corn taco; fried taco shells filled with meat and vegetables; tortilla chips; a stack of maize table tortillas. Courtesy of Electra Food Machinery Inc, El Monte, California, USA. Reproduced from Tortillas, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

into a moist dough called masa. Masa is the base for the production of a wide array of indigenous foods, ranging from table tortillas and chips (Figure 1) to the popular tamales and breakfast gruels. Maize-based, lime-cooked snack foods are competing strongly with potato chips and other salty snack foods in North America.

Maize, lime, and water are three basic ingredients needed for the production of masa. Lime softens the pericarp and endosperm and protein hydration and a small amount of starch is gelatinized. Due to its high pH, the starch chains are charged, which helps to retard retrogradation and staling. Other ingredients, such as preservatives, acidulants, emulsifiers, and gums, are used to improve the shelf-life of table tortillas. Frying oil is the second major ingredient in fried snacks. Snacks are usually salted and flavored.

Wheat Tortillas

Historically, tortillas were homemade, prepared on a daily basis, and consumed fresh. The fastest-growing tortilla area in the USA is wheat flour tortillas. Flour tortillas represents about two-thirds of the approximately 3 billion dollars of tortillas sold in the USA in 1998. Most wheat tortillas are industrially manufactured by hot-press, die-cut, or hand-stretch procedures. Each operation requires different flour specifications, dough preparation, and baking conditions, which result in various tortilla characteristics.

Hot-press tortillas are slightly off-round, elastic, resistant to tearing, have a smooth surface texture, and resist moisture absorption from fillings. Hot-press tortillas are consumed as gourmet table tortillas, fajitas, and soft tacos. Die-cut tortillas have a lower moisture content, are circular, less resistant to cracking, and often have dusting flour on the surface. Die-cut tortillas are mainly used in burritos, frozen Mexican foods, and fried products, i.e., taco salad bowls, taco shells, chimichangas, and buñuelos. Hand-stretch tortillas are irregular in shape, elastic, moderately resistant to tearing, and have dusting flour on the surface. Hand-stretch tortillas are consumed as table tortillas, burritos, and some fried products.

All wheat tortillas contain flour, water, fat, and salt. However, in the USA, tortillas may contain several other ingredients to improve flavor, softness, rollability, and shelf-life (1–4 weeks). These ingredients include chemical leavening agents, emulsifiers, antimicrobial agents, acidulants, gums or hydrocolloids, reducing agents, sugar, and yeast.

Wheat flour, the most important ingredient, is usually enriched, bleached, hard wheat flour with a protein content ranging from 9.5 to 14%. Equipment limitations and processing conditions determine the functionality of the flour. Flours for hot-press and hand stretch tortillas generally require less protein and gluten strength than flours for die-cut tortillas. Dough mixing time and dough properties are modified by reducing agents (sulfite, cystine), emulsifiers (lecithin, mono- and diglycerides, sodium stearoyl-2-lactylate), salt, gums, fat, water content, and dough temperature. (*See* **Emulsifiers**: Organic Emulsifiers; **Gums**: Properties of Individual Gums.)

Water (45–55% of flour weight) is needed to form the gluten complex. Solid or liquid fats (5–15% of flour weight) are added to improve dough properties, retard staling, and produce a softer and more flexible tortilla. Salt (1–2%) is added for taste and to strengthen the gluten complex. Baking powder (1.0–2.5%) gives a whiter, less dense, spongy product. Various natural and modified cellulose gums are added at 0.1–0.5% levels to improve dough machinability and decrease the stickiness of baked tortillas. Antimicrobial agents (propionates, sorbates) and acidulants (citric, fumaric, and phosphoric acids) limit fungal growth and extend shelf-life. Optimum pH for propionate activity is 5.5, and for sorbate activity is 6.0 when used at 0.2% of flour weight; however, dough mixing is more difficult below pH 5.8.

Wheat tortilla dough is mixed to incorporate the dry ingredients, fat and water, and to form a pliable, viscous dough. Tortilla dough is optimally mixed to slightly overmixed and varies in temperature from 26 to 38 °C, depending upon subsequent operations.

The dough is divided and rounded into dough balls in the hot-press and hand-stretch procedures. The dough balls are rested in a warm, moist environment for 5–20 min to relax the gluten complex. Rested dough balls machine easier and form better tortillas. In die-cut operations, the dough is pumped and shaped into a sheet that is further thinned by a series of cross-rollers on a moving belt. The thin sheet of dough (about 0.5 mm) is cut by a circular die which forms the shape. The scrap dough is returned to the dough pump and processed. Regardless of the process, the formed tortilla are baked (190–260 °C for 30–50 s) in gas-fired ovens that generally have three tiers. Oven conditions vary depending upon tortilla thickness, type of conveyor (slat or wire), and forming operation. Puffing of tortilla occurs near the end of baking and is more common in hot-press and hand-stretch tortillas. Tortillas are cooled to < 32 °C on cooling conveyors before placing into plastic bags for distribution. Improper cooling causes the tortilla to stick together and increases microbiological problems. Wheat flour tortilla producers have recently developed low-fat and fat-free products. The USA market for fat-free tortillas is between 35 and 40 million dollars. Kosher tortillas are another market under development.

Lime-Cooked Maize Products

Three basic types of product are industrially produced from lime-cooked maize: table or soft tortillas, corn chips, and tortilla chips (Figure 2). Corn and tortilla chips are primarily produced and consumed in developed countries, where they have an important share within the salted snack-food market. In 1998, sales of US-made corn and tortilla chips totaled 3.57 and 0.74 billion dollars, respectively. In contrast with sales in 1989, the market of these salted snacks increased 70% and 15.6%, respectively.

Table Tortillas

The technology for maize tortilla production was developed by early Mesoamerican civilizations. Tortillas and masa products constitute the staple food for large groups of people in Mexico and Central America. In Mexico, annual per capita intake in some

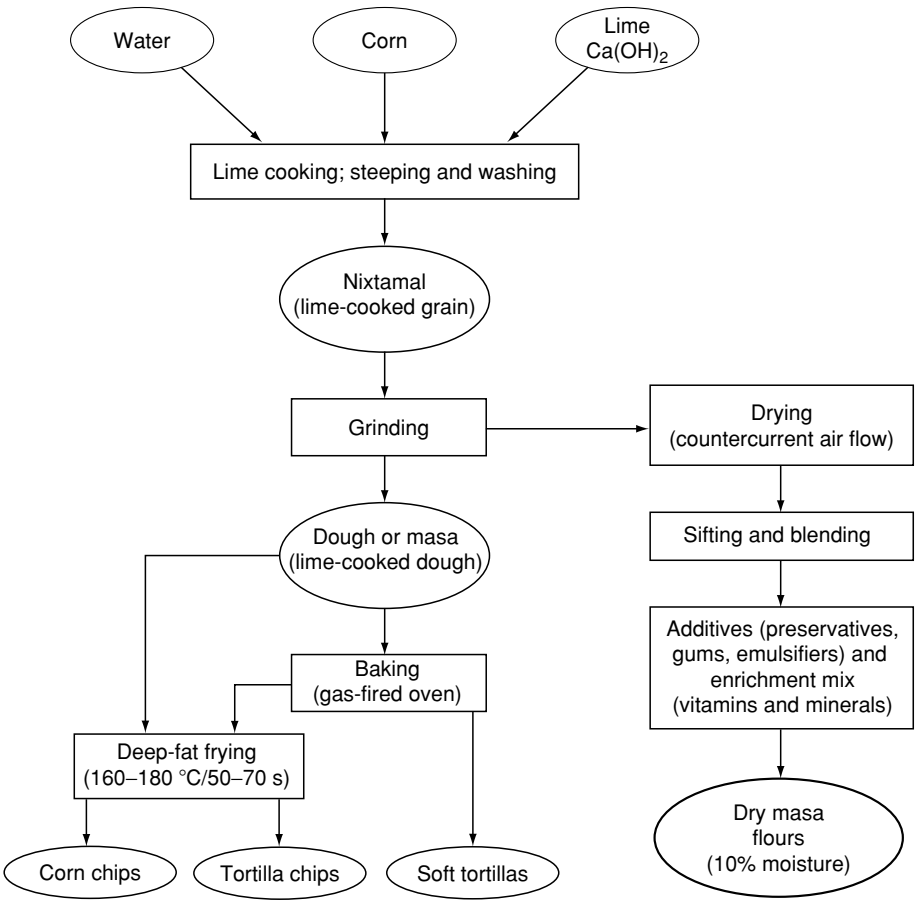

Figure 2 General scheme for the production of table tortillas, corn chips, tortilla chips, and dry masa flour.

groups of people is higher than 120 kg. In the USA, sales of table tortillas was estimated at 1 billion dollars in 1998. Tortillas are produced using traditional and industrial processes. In the traditional process maize is lime-cooked in clay pots over a fire, followed by steeping for 8–16 h (generally overnight). The cooking liquor, called nejayote, is discarded and then the nixtamal is hand-washed. Nixtamal is ground into a fine masa with a stone grinder called metate, or with hand-operated grinders. Masa is hand-molded or pressed into disks which are baked on a hot griddle or comal. Traditional tortillas are generally produced on a daily basis and are usually thicker and heavier than their industrially produced counterparts. Tortillas are the main source of energy, protein, calcium, and other important nutrients in Mexico and Central America. Lime-cooking considerably increases calcium and the bioavailability of niacin, and significantly decreases the amount of aflatoxins in contaminated maize. (*See* **Aflatoxins**; **Niacin**: Properties and Determination; **Pellagra**.)

The industrial production of maize tortillas is labor-intensive and requires considerable equipment. Quality of end products depends on the nature of raw material and manufacturing practices. The ideal grain should be sound, free of fissures, uniform, brightly colored, and with intermediate-to-hard endosperm. In addition, the kernels should have a rounded crown and a shallow, unwrinkled dent, and the pericarp should be easily removed after lime-cooking. Yellow and white maize, or mixtures therefrom, are commercially manufactured into alkaline products.

The industrial tortilla process starts when the maize is lime-cooked in agitated open baths, vertical cookers, or steam kettles. The grain is generally mixed with three parts water and 1% lime, based on grain weight, and cooked for 15–45 min at temperatures ranging from 85 to 100 °C. The nixtamal is then steeped for 8–16 h in the hot lime solution. After steeping, the maize is pumped with the steep liquor or dropped by gravity to washers. The cooking liquor is drained and the nixtamal washed with pressurized water. Most of the pericarp and excess lime is removed during this step. The cleaned nixtamal is discharged into a stone grinder, where masa is produced. The grinder consists of two volcanic or synthetic (aluminum oxide) stones carved radially, one stationary and the other rotating at 500–700 rpm. Masa particle size is directly related to the gap or pressure between the stones and the size and depth of the grooves. During grinding, the nixtamal is disrupted into a plastic and cohesive masa. Masa is then kneaded by mixers or extruders that feed the forming machine or sheeter rolls. During forming, the masa is rolled into a sheet which is cut by a rotating cutter positioned underneath the rolls. The formed pieces of masa are fed into a three-tier, gas-fired oven for baking. Tortillas are baked at temperatures ranging from 280 to 302 °C for 30–45 s, then cooled through a series of open tiers and packaged. Tortillas are generally treated with acidulants and antimold agents such as sorbates and propionates to improve their shelf-life. These additives are incorporated during grinding or masa kneading.

Dry Masa Flour

The use of dry masa flour is growing rapidly because of its convenience. More than 2 million metric tons of dry masa flour for table tortilla production are annually produced in Mexico. These flours are transformed into approximately 3.4 million tons of table tortillas. The dry masa flour industry supplies about 40% of the tortillas consumed in Mexico. Dry masa flour is produced by drying and grinding lime-cooked, coarsely ground masa (Figure 2). The masa is dried in large tunnels, or drying towers, in which warm air flows countercurrently to the masa. A wide array of products can be manufactured by selecting and blending streams with different particle size and color. In most cases, resulting flours are enriched with vitamins thiamin, riboflavin, niacin, and folic acid and minerals iron and zinc. Coarser flours with lighter colors are required for snacks. The flour with less than 10% moisture is shelf-stable and only requires water to form masa. Many manufacturers use dry masa flour because it does not require much labor, equipment, or space, and processors do not have to worry about effluent disposal and control of scheduling and manufacturing practices. Dry masa flours are generally mixed with 1.1 to 1.2 parts water for about 3–5 min to produce a suitable dough for further processing.

Fried/Snack Products

Frying has expanded the market for masa-based foods because the final product has excellent organoleptic properties and a long shelf-life. The two most popular snacks, tortilla and corn chips, are usually made from coarsely ground fresh masa or masa flours. Corn chips are produced directly from masa (Figure 2) and contain more oil than tortilla chips (Table 1). Tortilla chips are baked similarly to tortillas before frying. Tortilla chips absorb less oil, and have a firmer texture and a stronger corn flavor than corn chips. Nixtamal for these snack foods is generally cooked less than nixtamal for table tortillas, and it is ground into coarse masa which allows steam to escape through the many small pores during baking and frying. This prevents the formation of serious quality defects, such as oily appearance and pillowing or blistering. Masa for corn chips is extruded through a die, and cut by rotating knives before frying. Masa

Table 1 Nutritional profile (per 100 g serving) of flour tortillas, lime-cooked corn products, and table bread[a]

Nutrient	*Wheat flour tortilla*	*Lime-cooked products*			*Table bread*
		Table tortilla	*Corn chips*[b]	*Tortilla chips*[c]	
Water (g)	29.3	41.9	0.9	1.6	35.8
Energy (kcal)	351.3	238.3	573.1	514.3	274.0
Energy (kJ)	1471	998	2399	2153	1147
Digestible energy (kcal)	331.2	223.5	544.4	487.0	
Digestible energy (kJ)	1387	9360	2280	2039	
Protein (g)	7.2	6.5	6.3	7.6	8.7
Digestible protein (g)	6.4	5.38		6.2	7.7
Fat (g)	9.8	2.53	36.6	23.9	3.9
Ash (g)	2.2	0.90	1.9	1.1	1.9
Total dietary fiber (g)		7.4		12.4	2.7
Dietary insoluble fiber (g)		6.3		9.0	
Dietary soluble fiber (g)		1.1		3.1	
Starch (g)		44.9		59.9	
Calcium (mg)	42.2	92.8	105.0	124.0	126.0
Phosphorus (mg)	77.0	162.6	177.5	208.6	87.0
Magnesium (mg)	84.8	69.7	76.1	89.4	21.7
Sodium (mg)	573.0	13.3	1091.1		494.8
Potassium (mg)	99.0	205.3	211.2		103.0
Iron (mg)	1.5	2.5	2.9	3.5	2.6
Zinc (mg)	0.6	2.5	2.5	3.0	0.6
Copper (mg)	0.8	0.07	0.09	0.11	0.14

[a]Data from Betschart AA (1988) Nutritional quality of wheat and wheat products. In: Pomeranz Y (ed.) *Wheat Chemistry and Technology*. St Paul, Minnesota: American Association of Cereal Chemists, Bressani R (1990) Chemistry, technology, and nutritive value of maize tortillas. *Foods Reviews International* 62(2):225–264, Gonzalez Agramon MM and Serna Saldivar SO (1988) Effect of defatted soybean meal and soybean isolate on the nutritional, physical, chemical and organoleptic properties of wheat flour tortillas. *Journal of Food Science* 53: 793–797. Serna Saldivar SO, Knabe DA, Rooney LW, Tanksley TD and Sproule A (1988a) Nutritional value of sorghum and maize tortillas. *Journal of Cereal Science* 7:83–94, and Sproule AM, Serna Saldivar SO, Bockholt A, Rooney LW and Knabe DA (1988) Nutritional evaluation of tortillas and tortilla chips from quality protein maize. *Cereal Foods World* 33: 233–236.
[b]Salted corn chips.
[c]Unsalted tortilla chips. Salted tortilla chips contain approximately 774 mg of sodium per 100 g.

for tortilla chips is formed into triangles, strips, or circles before baking, equilibrating, and frying. Masa from yellow maize requires a lower frying temperature and longer residence time than masa from white maize. Frying temperatures and times range from 165 to 195 °C and 50 to 90 s.

Corn and tortilla chips are often salted and flavored immediately after frying. The hot chips are conveyed into an inclined rotating cylinder on to the product. Most popular flavorings include nacho cheese, hot/spicy, barbecue, lemon salt, and jalapeño. Corn and tortilla chips are packaged in moisture-proof or aluminized bags filled with an inert gas to protect the product physically and to prevent rancidity. Blue maize tortilla chips are often served in specialty restaurants and are available as organic and regular products. Blue maize has a pigmented aleurone that imparts an intense blue color. It has high levels of flavanoids and other phenolics that may have nutraceutical properties. The low-fat market of these products is increasing in the USA and other developed countries. Special processes to produce baked low-fat tortilla chips that combine air impingement, infrared, and microwaves have been developed. Olestra is currently being utilized to produce reduced-calorie chips. (*See* **Organically Farmed Food**.)

Nutritional Value

The nutrient composition of some lime-cooked maize foods is compared with white pan bread and wheat tortillas in Table 1. Tortillas, especially wheat flour tortillas, are commonly used as a substitute for pan bread by many people. Wheat flour tortillas have nutritional attributes similar to bread. Flour tortillas are higher in gross and digestible energy because their formula contains more shortening (5–15% based on flour weight). Wholemeal flour tortillas have higher amounts of fiber, protein, and ash than do white flour tortillas.

Lime-cooked maize products are an important source of energy, protein, dietary fiber, and calcium for people who depend on these items as their staple foods. Lime-cooking significantly improves the bioavailability of niacin. The caloric density of corn chips and tortilla chips is significantly higher than table tortillas. This is due to oil absorption during frying and their low moisture content. Tortillas and snacks produced from enriched dry masa flour contain higher levels of B vitamins and Fe and Zn than counterparts produced from fresh masa. In Mexico, dry masa flours are industrially enriched by a federal regulation and some of these flours are

optionally fortified with soybean protein. Human nutritional studies have demonstrated that fortified and enriched flours upgrade the nutritional status of low-income and marginal people, especially children. (*See* **Dietary Fiber**: Properties and Sources; **Niacin**: Properties and Determination.)

See also: **Aflatoxins**; **Dietary Fiber**: Properties and Sources; **Emulsifiers**: Organic Emulsifiers; **Gums**: Properties of Individual Gums; **Maize**; **Niacin**: Properties and Determination; **Organically Farmed Food**; **Pellagra**

Further Reading

Almeida Dominguez HD, Cepeda M and Rooney LW (1996) Properties of commercial nixtamalized corn flours. *Cereal Foods World* 41: 624–630.

Anonymous (1999) Snack Food Association state of the industry report. *Snack World* June.

Betschart AA (1988) Nutritional quality of wheat and wheat products. In: Pomeranz Y (ed.) *Wheat Chemistry and Technology*. St Paul, Minnesota: American Association of Cereal Chemists.

Bressani R (1990) Chemistry, technology, and nutritive value of maize tortillas. *Foods Reviews International* 62(2): 225–264.

Dally V and Navarro L (1999) Flour tortillas: a growing sector in the US food industry. *Cereal Foods World* 44: 457–459.

Gonzalez Agramon MM and Serna Saldivar SO (1988) Effect of defatted soybean meal and soybean isolate on the nutritional, physical, chemical and organoleptic properties of wheat flour tortillas. *Journal of Food Science* 53: 793–797.

Rooney LW and Serna Saldivar SO (1987) Food uses of corn and dry milled fractions. In: Waston SA and Ramstad PE (eds) *Corn Chemistry and Technology*. St Paul, MN: American Association of Cereal Chemists.

Rooney LW and Suhendro EL (1999) Perspectives on nixtamalization (alkaline cooking) of maize for tortillas and snacks. *Cereal Foods World* 44: 466–470.

Saldana G and Brown HE (1984) Nutritional composition of corn and flour tortillas. *Journal of Food Science* 49: 1202–1203.

Serna Saldivar SO (1996) *Química, Almacenamiento e Industrialización de los Cereales*. Mexico, DF: AGT.

Serna Saldivar SO, Knabe DA, Rooney LW, Tanksley TD and Sproule A (1988a) Nutritional value of sorghum and maize tortillas. *Journal of Cereal Science* 7: 83–94.

Serna Saldivar SO, Rooney LW and Waniska RD (1988b) Wheat flour tortilla production. *Cereal Foods World* 23: 855–864.

Serna Saldivar SO, Gomez MH and Rooney LW (1990) The chemistry, technology and nutritional value of alkaline cooked corn products. In: Pomeranz Y (ed.) *Advances of Cereal Science and Technology*, vol. X. St. Paul, MN: American Association of Cereal Chemists.

Serna Saldivar SO and Rooney LW (1994) Quality protein maize processing and perspectives for industrial utilization. In: Larkins B, Mertz ET (eds) *Proceedings of the International Symposium on Quality Protein Maize*, pp. 89–120. Sete Lagoas, MG, Brazil: EMBRAPA/CNPMS.

Sproule AM, Serna Saldivar SO, Bockholt A, Rooney LW and Knabe DA (1988) Nutritional evaluation of tortillas and tortilla chips from quality protein maize. *Cereal Foods World* 33: 233–236.

Waniska RD (1999) Prespectives on flour tortillas. *Cereal Foods World* 44: 471–473.

Yau JC, Waniska RD and Rooney LW (1994) Effects of food additives on storage stability of corn tortillas. *Cereal Foods World* 39: 396–402.

Toxicology *See* **Toxins in Food – Naturally Occurring**; **Mycotoxins**: Classifications; Occurrence and Determination; Toxicology

TOXINS IN FOODS – NATURALLY OCCURRING

S L Taylor, University of Nebraska, Lincoln, NE, USA

Background

The central axiom of toxicology is that all chemicals are toxic at some dose. Thus, all chemicals in foods, both synthetic and naturally occurring, are toxic. Though theoreticazlly correct, it is impractical to view all food constituents as toxicants. For the purposes of this review, naturally occurring toxicants are those chemicals of natural origin in foods that may present a hazard to consumers when eaten under some reasonable circumstances of exposure. Although toxicity is an intrinsic property of all chemicals, hazard is the capacity of a substance to produce

injury under the circumstances of exposure, taking into account the dose and frequency of exposure as well as the toxicity of the particular chemical. Illnesses caused by exposure to hazardous levels of chemical toxicants are called intoxications.

Definition of Natural

No legal definition exists for the word 'natural' in US food laws or the food laws of most other countries. Thus, there are no legal limitations on the use of the term. Webster's dictionary defines natural as "planted or growing by itself: not cultivated or introduced artificially (e.g., grass); existing in or produced by nature: consisting of objects so existing or produced: not artificial." Most consumers would likely define natural as being anything that is not artificial. Of course, artificial is another term that defies definition. Foods derived from animal or plant materials, whether processed or raw, are considered to be natural. The addition of additives to foods or the synthesis of a food from chemical sources would create a food that is perceived as not natural. To the food industry, additive-free or preservative-free foods are often described as natural. Naturally occurring chemicals from foods would be those chemicals that exist in foods in the raw, unprocessed state, although many of those chemicals would not be altered by processing or preparation practices.

Sources of Naturally Occurring Toxicants in Foods

Naturally occurring toxicants in foods can be either naturally occurring constituents of foods or contaminants that are produced in foods by various natural processes. Whereas naturally occurring constituents can be considered to be normal and unavoidable, naturally occurring contaminants are not always present and can be avoided, if contamination is prevented.

Naturally occurring toxicants are present in foods of animal, plant, or fungal origin. The majority of foods of animal origin, including meat, milk, cheese, eggs, fish, mollusks, and crustacea, are not hazardous. The most hazardous chemicals occurring in this category are found in poisonous seafoods. Foods of plant origin include vegetables, fruits, grains, seeds, nuts, and spices. Most plant-derived chemicals are also not especially hazardous, although some have the potential to be hazardous under some conditions of exposure. The major food of fungal origin is mushrooms, and many mushroom species are considered as poisonous and inedible.

Naturally occurring contaminants can be formed in foods as the result of contamination by bacteria, molds, algae, and insects. The chemicals produced from these biological sources can remain in the food even after the living organism has been removed or destroyed. Such contaminants represent the most important and potentially hazardous chemicals of natural origin existing in foods.

Additives Versus Natural

Consumers are often concerned about the chemicals added to their foods. These concerns likely arise from government regulatory actions against certain food additives or incidental contaminants such as pesticides. In contrast, natural products are widely viewed as beneficial to health.

Food laws are often predicated on the assumption that natural foods are safe, whereas most additives, especially all newly developed additives, must be proven to be safe. Although foods naturally contain many toxicants, including many chemicals that have been shown to be carcinogens in animal studies, government regulatory agencies have rarely taken any action to limit the availability of natural foods as a result of the presence of such toxicants. Meanwhile, the government has banned several food additives (cyclamate, FD&C Red No. 2 in the USA) and pesticides (DDT, ethylene dibromide in the USA) because of concerns about their safety, especially with respect to their possible carcinogenicity. Given this contrast, it is not surprising that consumers have developed a suspicious attitude regarding chemicals added to foods while remaining naïve about the potential hazards presented by naturally occurring chemicals in foods. (*See* **Food Additives**: Safety.)

Although unrecognized by many consumers, naturally occurring chemicals in foods present the greatest risk of any of the chemicals in foods. Thankfully, most of the potentially hazardous naturally occurring toxicants are ingested at doses that are unlikely to cause widespread illness, although the margin of safety is often lower for naturally occurring toxicants than it is for food additives. For example, the majority of the carcinogens in our diets are of natural origin. However, our diets also contain many naturally occurring compounds that have been shown to have anticarcinogenic activity in experimental animals and may also demonstrate similar activity in humans.

Hazards from Naturally Occurring Toxicants

Many naturally occurring, potentially hazardous chemicals exist in foods. These naturally occurring toxicants include acute toxicants (those capable of eliciting symptoms within a few minutes to hours after ingestion) and chronic toxicants (those that may

cause symptoms after life-long exposure or many years after exposure). Their ingestion does not invariably cause illness, because the dose and other circumstances of exposure are key factors. Chronic illnesses, such as cancer, are particularly difficult to correlate with foods or specific foodborne toxicants, since many confounding variables exist that affect the onset and course of such chronic diseases. Many naturally occurring carcinogens are known to exist in foods, but their relationship to cancer in humans remains unclear. A few of these carcinogens will be discussed. With acute toxicants, the relationship between ingestion of the toxicant and the development of symptoms is much stronger. Thus, many of the best examples of naturally occurring toxicants in foods involve acute illnesses.

The US Centers for Disease Control keep records on the incidence of foodborne disease. These statistics focus on acute illnesses, because the role of foods in chronic illnesses is uncertain. The most common foodborne illnesses are bacterial infections such as salmonellosis. Toxins produced by bacteria such as *Staphylococcus* enterotoxins and the botulinal toxins are important in the pathogenesis of some foodborne bacteriological illnesses. Illnesses of chemical etiology, other than bacterial toxins, account for 20–30% of all foodborne disease outbreaks in the USA. Fish and shellfish toxins account for the majority of these outbreaks, although other natural toxicants such as poisonous plants and mushrooms are also involved. The total number of foodborne disease outbreaks, including those of chemical etiology, is unknown because the reporting of such illnesses is very incomplete.

As noted earlier, the circumstances of exposure affect the likelihood of intoxication resulting from naturally occurring chemicals in foods. Table 1 describes some of the circumstances in which naturally occurring chemicals in foods can cause foodborne intoxications.

Table 1 Intoxications from consumption of naturally occurring foodborne toxicants

1. Abnormal though natural contaminants that adversely affect normal consumers eating normal amounts of the food
Algal toxins in seafood
Staphylococcal enterotoxins in various foods
Mycotoxins in various foods
2. Unusual 'foods' that adversely affect normal consumers eating normal amounts of this 'food'
Poisonous mushrooms
Poisonous plants
Poisonous fish such as pufferfish
3. Normal constituents of food that can cause illness if ingested by normal consumers in abnormally large amounts
Cyanogenic glycosides in lima beans, cassava, and fruit pits
Glycoalkaloids in potatoes and tomatoes
Estrogens in soybeans

Naturally Occurring Contaminants

As noted in Table 1, the sources of naturally occurring contaminants in foods can include algae, molds, bacteria, and insects.

Algal toxins in seafoods Seafoods, both fish and shellfish, can be occasionally contaminated with toxic substances that cause acute illnesses, and even death, within minutes after ingestion. These toxins are acquired by fish and shellfish through the food chain after being produced by certain species of microscopic, dinoflagellate algae. The fish and shellfish are hazardous only when feeding on toxin-producing algae. Seafood poisonings were once considered a public health problem only in coastal areas, but modern practices involving the shipment of fresh and frozen seafoods over long distances have made these a cosmopolitan concern.

Ciguatera poisoning is probably the most common cause of foodborne disease of chemical etiology on a world-wide basis. Ciguatera poisoning is common throughout the Caribbean and much of the Pacific. Fish that inhabit reef and shore areas in temperate regions, such as barracudas, groupers, sea basses, red snappers, and eels, are most commonly implicated in ciguatera poisoning. The identity of all the species of dinoflagellate algae associated with ciguatera poisoning remains unknown, although *Gambierdiscus toxicus* is one of them. Small reef fishes feed on the toxic dinoflagellate algae and are, in turn, consumed by the larger reef fishes. The toxins accumulate in the liver and viscera, but enough can enter the muscle tissues to result in ciguatera poisoning among humans ingesting these fish. Ciguatera poisoning is a neurologic illness initially characterized by tingling of the lips, tongue, and throat followed by numbness. Nausea, vomiting, a metallic taste, dryness of the mouth, abdominal pain, diarrhea, headache, and general muscular pain will ensue in many cases. Weakness may progress until the person is unable to walk. No antidotes are known, but many affected individuals recover within a few weeks, although deaths from cardiovascular collapse have occurred in a few instances.

Paralytic shellfish poisoning occurs in many coastal areas world-wide, including the Pacific and North Atlantic coasts of North America, Japan, and southern Chile. Many species of shellfish, such as clams and mussels, become poisonous through the consumption of toxic dinoflagellate algae. Two of the

most common algal species incriminated in paralytic shellfish poisoning are *Gonyaulax catenella* and *G. tamarensis*. The blooms of the toxic dinoflagellates are quite sporadic, so shellfish are hazardous only at certain times. Most shellfish clear the toxins from their system within a few weeks after the end of a dinoflagellate bloom, but some shellfish species, especially the Alaskan butter clam, appear to retain the toxins for long periods. The toxins involved in paralytic shellfish poisoning are known as saxitoxins. These are neurotoxins that act by blocking the passage of sodium ions into the nerve cells, an essential process in nerve transmission. The symptoms of paralytic shellfish poisoning include a tingling sensation and numbness of the lips, tongue, and fingertips, followed by numbness in the legs, arms, and neck with general muscular incoordination. Respiratory distress and muscular paralysis often occur. Death from respiratory failure can occur within 2–12 h depending on the dose. If the victim survives 24 h, the prognosis for a full recovery is good. No antidote exists for paralytic shellfish poisoning.

Pufferfish poisoning occurs more rarely because pufferfish are not frequently consumed except in Japan and China. About 30 species are found worldwide, but most are not poisonous. The most poisonous species occur along the coasts of Japan and China. The most choice, edible species belonging to the genus *Fugu* are the most poisonous and the most commonly consumed. The toxin in pufferfish accumulates in the ovaries, liver, intestine, skin, and roe (egg sac), so a carefully cleaned and eviscerated pufferfish may be safe to eat. Pufferfish likely become toxic through the ingestion of toxic dinoflagellate algae, although the food-chain relationships have not been completely elucidated. The toxin involved in pufferfish poisoning is a potent neurotoxin called tetrodotoxin, which also acts by blocking the sodium channel in nerve cells. The symptoms resemble those of paralytic shellfish poisoning. Death can result if a sufficient dose of tetrodotoxin is ingested.

Mycotoxins Mycotoxins are produced by a wide variety of molds that can grow on a wide variety of foods. The toxicity of many of the mycotoxins has been recognized by observing domestic animals fed moldy animal feeds. Their effects on humans are not so clearly established, although they are potentially hazardous to humans as well. (*See* **Mycotoxins**: Toxicology.)

Historically, ergotism was the first mycotoxin-associated illness recognized in humans. *Claviceps purpurea* is the responsible mold, which infects the heads of rye and sometimes wheat, barley, and oats. The shriveled, purplish grain kernel contains the mycotoxin. Ergotism is caused by a group of toxins known collectively as the ergot alkaloids and can be manifested in two forms: gangrenous ergotism and convulsive ergotism. In the former, also known as St. Anthony's fire, a burning sensation in the feet and hands is followed by a progressive restriction of blood flow to the hands and feet. This results in gangrene, which can lead to the loss of limbs. Convulsive ergotism involves hallucinations leading to convulsive seizures and sometimes death. No outbreaks of ergotism have been recorded since the early 1950s.

Aspergillus molds are known to produce several types of mycotoxins, most notably the aflatoxins and ochratoxin. The aflatoxins are produced primarily by *Aspergillus flavus* and *A. parasiticus* and often contaminate moldy peanuts and corn. Feeding of aflatoxin-contaminated grains or oilseeds to dairy cows can result in aflatoxin contamination of milk. The aflatoxins are potent carcinogens, especially affecting the liver. The role of aflatoxins in human carcinogenesis remains uncertain, but they are among the most potent animal carcinogens known. (*See* **Aflatoxins**.) Ochratoxin is produced primarily by *A. ochraceus*, which can contaminate cereals, peanuts, and tree nuts. Ochratoxin affects both the livers and kidneys of domestic animals, with the effects on the kidney being particularly damaging. The toxicity of ochratoxin to humans is unknown.

Fusarium molds produce a number of different mycotoxins, including the trichothecenes, fumonisins, and zearalenone. The trichothecenes are known to cause human illness and can be found as contaminants of grains primarily. Alimentary toxic aleukia (ATA) was observed in the former Soviet Union, owing to consumption of grains containing trichothecenes. ATA has four stages. In the first stage, a person experiences a burning sensation in the mouth and throat, which proceeds down the esophagus to the stomach. This is followed by diarrhea, nausea, and vomiting, occurring 1–3 days later and ceasing after about 9 days. The second stage occurs from about 2 weeks to 2 months and involves bone-marrow destruction, leukemia, agranulocytosis, anemia, and loss of platelets. At the end of this stage, small hemorrhages may occur on the skin. The third stage of ATA lasts from 5 to 20 days and involves total loss of bone marrow with necrotic angina, sepsis, total agranulocytosis, and moderate fever. The hemorrhages become larger, and necrotic lesions appear on the skin. Bronchial pneumonia appears, along with abscesses and hemorrhages in the lungs. The fourth stage is death, with the mortality rate approaching 80% and dependent on the dose and frequency of exposure to the trichothecenes.

Fumonisins are produced primarily by *Fusarium moniliforme*, which contaminates various grains and soybeans. The fumonisins have been implicated in equine leukoencephalomalacia, a fatal neurotoxic syndrome in horses characterized by extensive necrosis of the white matter in the brain. The fumonisins may also be carcinogenic, although this remains an active area of scientific scrutiny. Their effects on humans remain unknown, although low-level contamination of grains with fumonisins seems to be fairly common. Zearalenone is an estrogenic substance produced in grains by *F. graminearum*. It causes spontaneous abortions in pigs, but its effects on humans are unknown.

Penicillium molds produce a wide variety of mycotoxins, including rubratoxin, patulin, and citrinin. These mycotoxins cause various effects in domestic animals, but their effects, if any, on humans are quite uncertain. Many other mycotoxins have been identified, though most of these others have received little scientific study.

Bacterial toxins Most pathogenic bacteria exert their effects by the infectious route. They invade cells and tissues, multiply, and cause a variety of symptoms. A few bacteria produce exogenous toxins in foods before they are eaten. The ingestion of the toxins initiates the disease process even if the bacteria are destroyed in processing or preparation. The best examples of bacterial intoxications are the staphylococcal enterotoxins and botulinal toxins.

The staphylococcal enterotoxins can be produced in foods by certain strains of *Staphylococcus aureus*. When ingested, these protein enterotoxins cause nausea and vomiting within 1–6 h. Staphylococcal food poisoning is one of the most common forms of foodborne disease. (*See* ***Staphylococcus***: Food Poisoning.)

The botulinal toxins can be produced in foods under anaerobic conditions by *Clostridium botulinum*. Toxin formation often occurs in canned foods subjected to improper thermal processing. The commercial canning process is predicated on the destruction of this organism and its spores so that the spores will not germinate, grow, and produce toxin on storage of the canned product. Botulinal toxins are among the most potent toxins known to humans and frequently result in respiratory paralysis and death. (*See* ***Clostridium***: Botulism.)

Insect-produced toxins Insects can also produce and secrete toxic chemicals into foods, yet insect infestation of foods is often considered only in esthetic terms. Natural toxicants produced by insects in foods have received very little study. Flour beetles are known to secrete benzoquinones into flour, which are mutagenic and carcinogenic. The hazards posed by insect-produced toxicants in foods to humans remain unknown.

Naturally Occurring Constituents Eaten in Normal Amounts

Some plants and animals that should not be eaten are consumed intentionally or accidentally on occasion, resulting in foodborne chemical intoxications. Many plants and some animals contain levels of naturally occurring toxicants that are probably not hazardous to humans ingesting typical amounts of these foods.

Animals Very few animal species are poisonous. Several species of poisonous fish exist, the best example being the pufferfish, although recent evidence suggests that the toxin in pufferfish arises from an algal contaminant.

Plants Many plant species are decidedly poisonous. Plants such as hemlock and nightshade were used in ancient times to poison enemies. Consumers can easily avoid poisonous plants by purchasing foods from commercial sources. Intoxication from the ingestion of poisonous plants occurs primarily from misidentification of plants by individuals harvesting their own foods in the wild. The most frequent examples occur when harvesting herbs for herbal tea. For example, an elderly couple succumbed after mistaking foxglove for comfrey while harvesting herbs for tea.

Very rarely, intoxications from poisonous plants occur with products purchased from retail outlets. An example was the contamination of an herbal tea by *Senecio*, a well-known poisonous plant. The tea was fed to infants as an herbal remedy for viral infections. The infants suffered acute symptoms, involving impaired liver function, from the presence of pyrrolizidine alkaloids in the *Senecio* leaves. Several infants died in this unfortunate incident. (*See* **Alkaloids**: Toxicology.)

Occasionally, intoxications from poisonous plants occur from the intentional addition of such materials to foods. The most frequently encountered example would be the preparation of brownies laced with marijuana, a potent hallucinogenic plant.

In times of severe food shortages, humans may resort to eating large quantities of plants that contain hazardous components. Lathyrism is an unusual disease that has occurred on such occasions in India and Italy when seeds of *Lathyrus sativus* or chickling vetch are ingested as a principal part of the diet.

These seeds contain several lathyrogens, which cause both neurological and skeletal abnormalities when they are eaten in large amounts for an extended period.

Fungi Many species of poisonous mushrooms are known. Harvesting mushrooms in the wild can be a hazardous practice, even for those skilled in mushroom identification. Several different types of naturally occurring toxicants exist in poisonous mushrooms.

The most hazardous of the mushroom toxins are the Group I toxins. Amatoxin is the best example and is produced by *Amanita phalloides*, the so-called death cap mushroom. The symptoms of amatoxin poisoning begin 6–24 h after ingestion of the mushrooms. The first stage of the intoxication involves the gastrointestinal tract with abdominal pain, nausea, vomiting, diarrhea, and hyperglycemia. A short period of remission follows. The third and often fatal stage of the illness involves severe liver and kidney dysfunction with symptoms including abdominal pain, jaundice, renal failure, hypoglycemia, convulsions, coma, and death. Death from hypoglycemic shock occurs 4–7 days after the onset of symptoms. Recovery is possible but requires at least 2 weeks of intensive therapy.

The Group II toxins are hydrazines, with gyromitrin being the premier example, which is produced by *Gyromitra esculenta* mushrooms. Usually, the symptoms include a bloated feeling, nausea, vomiting, watery or bloody diarrhea, abdominal pain, muscle cramps, faintness, and a loss of motor coordination occurring 6–12 h after consumption.

The classic example of a Group III toxin is coprine, which causes symptoms only when ingested with alcoholic beverages. Symptoms begin about 30 min after consumption of alcohol along with mushrooms containing Group III toxins such as *Corpinus atramentarius*. Symptoms include flushing of the face and neck, distension of the veins in the neck, swelling and tingling of the hands, metallic taste, tachycardia, and hypotension, progressing to nausea and vomiting. Symptoms can last for up to 5 days.

The Group IV toxins, characterized by muscarine, affect the autonomic nervous system. Symptoms include perspiration, salivation, and lacrimation with blurred vision, abdominal cramps, watery diarrhea, constriction of the pupils, hypotension, and a slowed pulse. Death does not usually occur unless the Group III toxins are present in a mushroom such as fly agaric (*Amanita muscarina*), which also contains Group I toxins. With fly agaric, a fatal combination of symptoms can occur.

The Group V and VI toxins act primarily on the central nervous system to cause hallucinations. The Group V toxins include ibotenic acid and muscimol, which cause dizziness, incoordination, staggering, muscular jerking and spasms, hyperkinetic activity, a coma-like sleep, and hallucinations beginning 30 min to 2 h after ingestion. Fly agaric, in addition to its content of Group I and Group IV toxins, is also a good source of Group V toxins. The Group VI toxins include psilocybin and psilocin, whose symptoms include pleasant or aggressive mood, unmotivated laughter and hilarity, compulsive movements, muscle weakness, drowsiness, hallucinations, and sleep. Symptoms usually begin 30–60 min after ingestion of the mushrooms, and recovery is often spontaneous. *Psilocybe mexicana*, the so-called Mexican mushroom, is a well-known source of the Group VI toxins. Although Mexican mushrooms (also known as magic mushrooms or simply 'shrooms') have been used as recreational drugs for their hallucinogenic effects, it must be recognized that the dose of the Group VI toxins in these mushrooms varies widely (more than 100-fold). Exposure to high amounts of the Group VI toxins has led to prolonged and severe side-effects, even death. Often, patients experience persistent sequelae and are admitted to mental institutions. (*See* **Mushrooms and Truffles**: Classification and Morphology.)

Naturally Occurring Constituents Eaten in Abnormally Large Amounts

The dose of a toxic chemical is the major determinant of the degree of hazard that it poses in the diet. Many naturally occurring foodborne toxicants are present at levels that do not constitute a hazard to normal consumers eating a normal diet. However, if unusually large amounts of foods containing these toxicants are consumed, intoxication may occur. One classic example is the presence of cyanogenic glycosides in lima beans and cassava.

Many plants contain cyanogenic glycosides, which are sugars that can release cyanide on exposure to certain enzymes present in the plant tissues or by acid in the stomach of the consumer. Linamarin in lima beans is one example of a cyanogenic glycoside. Commercial varieties of lima beans contain very little linamarin, but wild lima bean varieties can have substantial amounts. Cyanide is a classic toxicant that binds to heme proteins in the mitochondria and hemoglobin in the blood. Cyanide prevents oxygen binding to hemoglobin, thus causing cyanosis, noticeable as a bluish discoloration of the skin and mucous membranes. It also inhibits cellular respiration by binding to the mitochondrial heme proteins. The

symptoms of cyanide poisoning include the rapid onset of peripheral numbness and dizziness, mental confusion, stupor, cyanosis, twitching, convulsions, coma, and death. Cyanide is rapidly absorbed, and the lethal dose is 0.5–3.5 mg of cyanide per kilogram of body weight. Commercial lima bean varieties release about 10 mg of cyanide per 100 g of beans. Assuming that the lethal dose is 0.5 mg kg^{-1}, a 70-kg adult would need to ingest 35 mg of cyanide or 350 g of lima beans. Though this is not impossible, it is certainly very unlikely.

Cassava can release as much as 50 mg of cyanide per 100 g of food. Cassava intake is substantial in Africa and South America in certain locales when other foods are scarce. The major source of cyanide intoxication remains the intentional ingestion of fruit pits that contain amygdalin, also known as laetrile, under the mistaken belief that laetrile counteracts cancer. The level of cyanogenic glycosides in fruit pits is considerable, and several deaths have occurred from their ingestion. (*See* **Cassava**: The Nature of the Tuber.)

In addition to the cyanogenic glycosides, plants contain an enormous variety of other potentially hazardous chemicals that are not typically evident unless large quantities of these foods are eaten. Examples include glycoalkaloids in potatoes and tomatoes, goitrogens in cruciferous vegetables, nitrate in spinach, celery, and lettuce, oxalates in rhubarb and spinach, estrogens and saponins in soybeans, and tannins and phytoalexins in many plant foods. Symptoms do not occur after ingestion of these foods in typical amounts.

Naturally Occurring Constituents Eaten by Consumers with Genetic Abnormalities

Some naturally occurring constituents of foods are hazardous only for consumers with allergies or intolerance to that food. Food allergies and intolerances are the subject of a separate review in this volume. (*See* **Food Intolerance**: Food Allergies; Milk Allergy.)

Naturally Occurring Constituents from Foods Processed or Prepared in an Unusual Manner

Many naturally occurring foodborne toxicants are removed or inactivated during processing or preparation. For example, raw soybeans contain trypsin inhibitors, lectins, allergens, amylase inhibitors, saponins, and antivitamins along with other potentially hazardous factors. However, humans never ingest raw soybeans. The trypsin inhibitors and lectins are inactivated by heating the soybeans. Fermentation in the preparation of tofu destroys certain toxic factors in soybeans. (*See* **Trypsin Inhibitors**.)

Intoxications have occurred from the ingestion of undercooked or raw kidney beans. Kidney beans, like soybeans, are legumes. They contain lectins, which can bind to sugar residues on the surfaces of cell membranes, causing hemolysis of red blood cells and intestinal damage. The lectins from kidney beans can be inactivated by thorough cooking. If not inactivated, the lectins will cause nausea, abdominal pain, vomiting, and bloody diarrhea. In England, intoxications have occurred among immigrants unfamiliar with proper cooking practices for this common item in the British diet. These consumers simply soaked the raw beans and ate them with little or no cooking, resulting in the prompt onset of gastrointestinal symptoms. (*See* **Hemagglutinins (Haemagglutinins)**.)

See also: **Aflatoxins**; **Alkaloids**: Toxicology; **Cassava**: The Nature of the Tuber; ***Clostridium***: Botulism; **Food Additives**: Safety; **Food Intolerance**: Food Allergies; Milk Allergy; **Hemagglutinins (Haemagglutinins)**; **Mushrooms and Truffles**: Classification and Morphology; **Mycotoxins**: Toxicology; **Trypsin Inhibitors**; ***Staphylococcus***: Food Poisoning

Further Reading

Beier RC and Nigg HN (1994) Toxicology of naturally occurring chemicals in foods. In: Hui YH, Jr., Gorham R, Murrell KD and Cliver DO (eds) *Foodborne Disease Handbook, Volume 3: Diseases Caused by Hazardous Substances*. New York: Marcel Dekker.

D'Mello JPF (1997) *Handbook of Plant and Fungal Toxicants*. Boca Raton, FL: CRC Press.

National Research Council (1996) *Carcinogens and Anticarcinogens in the Human Diet*. Washington, DC: National Academy Press.

Shibamoto T and Bjeldanes LF (1993) *Introduction to Food Toxicology*. San Diego, CA: Academic Press.

Spoerke DG, Jr. (1994) Mushrooms: Epidemiology and medical management. In: Hui YH, Jr., Gorham R, Murrell KD and Cliver DO (eds) *Foodborne Disease Handbook, Volume 3: Diseases Caused by Hazardous Substances*. New York: Marcel Dekker.

Taylor SL and Schantz EJ (1990) Naturally occurring toxicants in foods. In: Cliver DO (ed.) *Foodborne Diseases*. San Diego, CA: Academic Press.

Whittle K and Gallacher S (2000) Marine toxins. *British Medical Journal* 56: 236–253.

TRACE ELEMENTS

F H Nielsen, USDA, ARS, Grand Forks Human Nutrition Research Center, Grand Forks, ND, USA

Introduction

In the earlier part of the 20th century, scientists could qualitatively detect small amounts of several mineral elements in living organisms. In reports, these elements were described as being present in 'traces' or 'trace amounts.' Not surprisingly, these elements became known as trace elements. The trace elements found in living organisms may be essential, i.e., indispensable for performing functions necessary for life, or they may be nonessential, fortuitous reminders of our geochemical origins or indicators of environmental exposure. Some of the nonessential trace elements can be beneficial to health through pharmacological action. All the trace elements are toxic when intake is excessive.

Definition of Trace Elements

Presently, mineral elements considered to be trace elements are those that occur in the body in $mg\,kg^{-1}$ of body weight or less amounts. Essential trace elements are generally required by humans in $mg\,day^{-1}$ amounts. Since 1980, the term 'ultratrace element' has appeared in the nutritional literature. Ultratrace elements have been defined as those elements with estimated dietary requirements usually less than $1\,mg\,kg^{-1}$, and often less than $50\,\mu g\,kg^{-1}$ of diet for laboratory animals. For humans, the term is often used to indicate an element with an established, estimated, or suspected requirement of less than $1\,mg\,day^{-1}$, or generally indicated by $\mu g\,day^{-1}$.

Essential and Possibly Essential Trace and Ultratrace Elements

In the 1960s and 1970s, the standard to establish a nutrient as essential was liberalized for mineral elements that could not be fed in amounts low enough to cause death or interrupt the life cycle (interfere with growth, development, or maturation such that procreation is prevented). Thus, an essential element during this time period was defined as one whose dietary deficiency consistently and adversely changed a biological function from optimal, and this change was preventable or reversible by physiological amounts of the element. This definition of essentiality became less acceptable when a large number of elements was suggested to be essential based on small changes in physiological or biochemical variables. Many of these changes were questioned as to whether they were necessarily the result of a suboptimal function, and sometimes were suggested to be the consequence of a pharmacological or toxic action in the body, including an effect on intestinal microorganisms. As a result, if the lack of an element cannot be shown to cause death or to interrupt the life cycle, many scientists, perhaps a majority, now do not consider an element essential unless it has a defined biochemical function. However, there are still scientists who base essentiality on the older definition. Thus, there is no universally accepted list of essential trace and ultratrace elements.

Copper, iron, manganese, zinc, cobalt, iodine, molybdenum, and selenium are elements with evidence for essentiality that is substantial and noncontroversial. Specific biochemical functions have been defined for all of them. Boron and chromium also belong in the list of trace elements accepted as essential. Boron has recently been found to be required to complete the life cycle of fish and frogs. A defined biochemical function for chromium has recently been identified.

Circumstantial evidence is often used to contend that an element is essential. This evidence generally fits into four categories. These are:

1. A dietary deprivation in some animal model consistently results in changed biological function, body structure, or tissue composition that is preventable or reversible by an intake of an apparent physiological amount of the element in question.
2. The element fills the need at physiological concentrations for a known *in vivo* biochemical action to proceed *in vitro*.
3. The element is a component of known biologically important molecules in some life forms.
4. The element has an essential function in lower forms of life.

An element is considered to have strong circumstantial support for essentiality if all four types of evidence exist for it. There is strong circumstantial evidence for the essentiality of arsenic, nickel, silicon, and vanadium; thus, they are considered possibly essential elements.

If an element has only one or two types of circumstantial evidence to support essentiality, it generally does not receive widespread support for being a

possibly essential element. However, some of these elements have beneficial pharmacological actions (fluoride prevents caries; lithium is an antimanic agent), and some may eventually be found to be of some nutritional importance. Elements that fit in this category include aluminum, bromine, cadmium, fluorine, germanium, lead, lithium, rubidium, and tin.

Importance in the Diet

Trace and ultratrace elements have five known roles in living organisms:

1. In close association with enzymes, some elements are integral parts of catalytic centers at which the biochemical reactions necessary for life occur. Working in concert with a protein, and frequently with another organic coenzyme, an element attracts a substrate molecule and facilitates its conversion into a specific end product.
2. Some elements donate or accept electrons in reactions of reduction or oxidation. In addition to the generation and utilization of metabolic energy, redox reactions frequently involve the chemical transformation of molecules.
3. Some elements, particularly iron, bind, transport and release oxygen within living organisms.
4. Some elements have structural roles, imparting stability and three-dimensional structure to important biological molecules.
5. Some elements have regulatory roles. They control important biological processes through such actions as inhibiting enzymatic reactions, facilitating the binding of molecules to receptor sites on cell membranes, altering the structure or ionic nature of membranes to prevent or allow specific molecules or ions to enter a cell, and inducing genes to express themselves, resulting in the formation of proteins involved in life processes.

Although trace and ultratrace elements play key roles in a variety of processes necessary for life, except for iodine and iron, the occurrence of overt, simple, or uncomplicated deficiency of any trace element in humans is not common. The reasons for this are the powerful homeostatic mechanisms involved in absorption, storage, and excretion that work to maintain a rather constant amount of trace elements in the body despite varying intakes; and the consumption of diets with different types of foods from different sources which increases the opportunity to achieve adequate intakes. Nonetheless, reductions in health and well-being because of suboptimal status in some trace elements are probably not uncommon because of other factors. That is, most important in making trace and ultratrace elements of nutritional concern is that their metabolism or utilization can be impaired or their need can be increased by nutritional, metabolic, hormonal, or physiological stressors. This is exemplified by selenium, for which it is difficult to produce signs of pathology caused by a simple dietary deficiency in animals and humans. A stressor such as vitamin E deficiency or a viral infection is needed with selenium deficiency to obtain marked pathology such as that seen with Keshan disease, a cardiomyopathy that primarily affects children and women of child-bearing age in some areas of China.

Nutritional and Physiological Aspects of Specific Elements

Nutritional and physiological aspects of the essential elements copper, iron, zinc, iodine, selenium, and chromium are described in separate sections in this encyclopedia, and thus will not be described further here. For the other elements mentioned above, Table 1 and the following give a brief summary of findings that support their essentiality, possible physiological and biochemical functions, and conditions that might enhance susceptibility to pathological consequence because of low dietary intake.

Aluminum

The suggestion that aluminum might be essential is relatively new. It was recently reported that depriving goats of aluminum results in increased abortions, depressed growth, incoordination and weakness in hind legs and decreased life expectancy. Aluminum deficiency has also been reported to depress growth in chicks. *In vitro* biochemical actions which add circumstantial support to aluminum essentiality include the activation of the enzyme adenylate cyclase, enhancement of calmodulin activity, stimulation of DNA synthesis in cell cultures, and stimulation of osteoblasts to form bone. A concern about the evidence for essentiality is that the only deficiency sign found in more than one species is depressed growth. More of the other reported signs need to be confirmed by additional research groups or in other species. Moreover, in the deficiency studies to date, rather high amounts of aluminum were fed to supplemented controls; this gives the possibility that aluminum may have acted pharmacologically rather than nutritionally in these studies. The reported signs of deficiency do not give much insight into a possible biological function for aluminum. However, the *in vitro* findings suggest that it could possibly function as an enzyme activator.

Table 1 Human adult body content, adequate or typical daily dietary intake, and dietary sources of some trace and ultratrace elements

Element (symbol)	*Total adult body content*	*Adequate daily intake (ADI) or typical daily intake (TDI)*	*Dietary sources*
Aluminum (Al)	30–50 mg	2–10 mg (TDI)	Baked goods containing baking powder, grains, vegetables, tea
Arsenic (As)	1–2 mg	12–15 μg[a] (ADI)	Fish, shellfish, grains, cereal products
Boron (B)	10–20 mg	0.5–1.0 mg[b] (ADI)	Noncitrus fruits, nuts, leafy vegetables, pulses, legumes
Bromine (Br)	100–350 mg	2–8 mg (TDI)	Nuts, grains, fish
Cadmium (Cd)	5–20 mg	10–20 μg (TDI)	Shellfish, grains, leafy vegetables, organ meats
Cobalt (Co)	1.5 mg	2.4 μg as vitamin B_{12}[c] (ADI)	Green leafy vegetables, organ meats
Fluorine (F)	3 g[d]	3.0–4.0 mg[c] (ADI)	Seafood, fluoridated water, tea
Germanium (Ge)	3 mg	0.4–3.4 mg (TDI)	Wheat bran, vegetables, leguminous seeds
Lead (Pb)	120 mg	15–100 μg (TDI)	Shellfish, plant foodstuffs from high-lead soils
Lithium (Li)	350 μg	200–600 μg (TDI)	Eggs, meat, fish, milk products, potatoes, leafy vegetables
Manganese (Mn)	10–20 mg	1.8–2.3 mg[c] (ADI)	Whole grains, nuts, legumes
Molybdenum (Mo)	10 mg	45 μg[c] (ADI)	Milk products, vegetables, organ meats, pulses, grains, legumes
Nickel (Ni)	1–2 mg	100 μg[a] (ADI)	Nuts, chocolate, pulses, grains, legumes
Rubidium (Rb)	360 mg	1–5 mg (TDI)	Fruits, vegetables (especially asparagus), poultry, fish
Silicon (Si)	2–3 g	5–20 mg[a] (ADI)	Whole grains, cereal products, root vegetables
Tin (Sn)	7–14 mg	1–40 mg (TDI)	Nuts, canned foods
Vanadium (V)	100 μg	6–10 μg[a] (ADI)	Shellfish, mushrooms, condiments, prepared foods

[a]Estimated by author through extrapolation of data from animals.
[b]Estimated by author using data from both animals and humans.
[c]Recommended dietary intake (Co, Mo) and adequate intake (F, Mn) for adults 19 and older established by The Food and Nutrition Board, Institute of Medicine, National Academy of Sciences in *Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D and Fluoride* (1997); *Dietary Reference Intakes for Thiamin, Riboflavin, Niacin, Vitamin B_6, Folate, Vitamin B_{12}, Pantothenic acid, Biotin and Choline (1998)*; and *Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium and Zinc* (2001). Washington, DC: National Academy Press. For Flourine and Manganese, the first value is for women, the second value for men.
[d]Mostly in bone; about 30 mg in soft tissue.

Arsenic

Signs of arsenic deprivation have been described for several animal species, including the chicken, goat, hamster, rat, and pig. In the goat, rat, and miniature pig, the most consistent signs of arsenic deficiency have been depressed growth and abnormal reproduction characterized by impaired fertility and elevated perinatal mortality. Other notable signs of deprivation described for goats include myocardial damage and death during lactation. Some biochemical changes that have been found in arsenic-deficient animals include depressed serum triglycerides in goats, depressed plasma taurine concentrations in hamsters, depressed hepatic *S*-adenosylmethionine and elevated hepatic *S*-adenosylhomocystine in hamsters and rats, and depressed hepatic putrescine, spermidine, and spermine concentrations and *S*-adenosylmethionine decarboxylase activity in rats. Factors enhancing the response to low dietary arsenic are stressors that affect sulfur amino acid or labile methyl group metabolism, including high dietary arginine and selenium, low dietary methionine, zinc, selenium and choline, and taurine and guanidoacetic acid supplementation.

In vitro findings that support arsenic essentiality include its ability to activate some enzymes, enhance DNA synthesis in unsensitized and stimulated human lymphocytes, and induce the isolated production of certain proteins known as heat shock or stress proteins. The control of production of stress proteins is apparently at the transcriptional level, and may involve changes in methylation of core histones. Arsenic can increase the methylation of the p53 promoter in human lung cells. Interestingly, although arsenic is thought to be carcinogenic by many, it has recently been found to be effective in the treatment of some forms of leukemia. Moreover, it has been suggested that arsenic can play an important role in human health because it was found that injuries of the central nervous system, vascular diseases, and cancer were correlated to markedly decreased serum arsenic concentrations.

A biochemical function for arsenic has not been identified in lower forms of life, although a bacterium, *Chrysiogenes arsenatis*, reduces As^{5+} to As^{3+} to gain energy for growth. However, there are enzymes in higher animals and humans that methylate arsenic with *S*-adenosylmethionine as the methyl donor. Arsenite methyltransferase methylates arsenite to monomethylarsenic acid, which is methylated by monomethylarsenic acid methyltransferase to yield dimethylarsinic acid, the major form of arsenic in urine. Findings to date indicate that arsenic might

have a function that affects the formation and utilization of labile methyl groups arising from methionine. Through this effect on methyl group metabolism, arsenic probably affects the methylation of important molecules such as DNA, which can affect the susceptibility to certain types of cancer.

Boron

Knowledge about the role and clinical aspects of boron in nutrition is just emerging. The biological function of boron has not been identified, but boron apparently has a role that influences the metabolism and utilization of several other nutrients. Thus, described deficiency signs and the pathological consequences of inadequate boron intake, because they can be affected by the intake of several other dietary substances, are numerous and variable. In humans, boron apparently affects calcium metabolism and hormone action. Among the various responses to low dietary boron are decreased serum 25-hydroxycholecalciferol, triglycerides, and ceruloplasmin; decreased erythrocyte superoxide dismutase; increased serum calcitonin, glucose, and creatinine; increased blood urea concentration; and decreased urinary hydroxyproline excretion. Boron deprivation also depresses the elevation in serum 17β-estradiol and plasma copper caused by estrogen ingestion; alters electroencephalograms such that they suggest impaired behavioral activation (e.g., more drowsiness) and mental alertness; and impairs psychomotor skills and the cognitive processes of attention and memory. A number of signs of deficiency have been found in animals that may eventually have some counterparts in humans. Signs of deficiency in the frog include increased necrotic eggs; high frequency of abnormal gastrulation in embryos; abnormal development of the gut, craniofacial region, and eye; visceral edema and kinking of the tail during organogenesis; and delayed tail absorption during metamorphosis. Signs of boron deficiency in the zebrafish include embryo death during the zygote and cleavage periods before the formation of a blastula during embryo development, and photophobia characterized by photoreceptor dystrophy in adults. Signs of boron deficiency in rats include the exacerbation of distortion of marrow sprouts and delay in the initiation of cartilage calcification in bones during marginal vitamin D deficiency, and decreased circulating concentrations of natural killer cells and $CD8a^{+}/CD4^{-}$ cells during antigen-induced arthritis. Factors enhancing the response to low dietary boron are stressors that affect calcium metabolism and utilization, especially low dietary intakes of vitamin D or calcium, and stressors that affect cell membrane function or signal transduction, including low dietary intakes of magnesium, potassium, or copper.

Over 70 years after being found essential for plants to complete their life cycle, a biochemical function apparently has been identified for boron in plants. Boron has a structural role in the cell wall, and also apparently a functional role in the cell membrane. These roles support the hypothesis that in humans and higher animals boron has a role in cell membrane function that influences the response to hormone action, transmembrane signaling, or transmembrane movement of regulatory cations or anions. It has also been hypothesized that boron is a metabolic regulator through its ability to complex with a variety of substrate or reactant compounds in which there are hydroxyl groups in favorable positions; the regulation is mainly negative.

Bromine

Evidence supporting bromine essentiality includes findings showing that bromide can substitute for part of the chloride requirement for chicks; bromide can alleviate growth retardation induced in chicks and mice fed iodinated casein; and insomnia exhibited by some hemodialysis patients has been associated with bromide deficit. The insomnia finding is tempered by the knowledge that, before barbiturates were used, doctors prescribed bromide for sleep. Recently it was reported that, when compared to goats fed 20 mg bromide kg^{-1} diet, goats fed 0.8 mg bromide kg^{-1} diet exhibited depressed growth, fertility, hemoglobin, hematocrit, and life expectancy, and increased spontaneous abortions. Although these latest findings give more credibility to the concept of bromine essentiality, the findings are still too limited. In other words, the deficiency signs of bromine need to be confirmed in additional species, preferably in another research setting. The evidence presented to support essentiality does not provide much insight into a possible biochemical role for bromine. Because it can partially substitute for chloride, bromide could possibly have a function similar to chloride.

Cadmium

Cadmium is best known for its intoxicating properties. However, in a study of goats, it was found that 9 kids from goats fed 15 μg cadmium kg^{-1} diet, when compared to kids from goats fed 300 μg cadmium kg^{-1} diet, exhibited muscular weakness; they were stiff and clumsy when they moved. Six of the 9 kids finally could not raise their heads and died; the other 3 were fed the 300 μg cadmium kg^{-1} diet which restored their mobility, feed intake, and growth. Muscle weakness that led to death also occurred in some

lactating goats fed the low-cadmium diet. It should be emphasized that the amounts of dietary cadmium in this study were much lower than those used in toxicological studies; these usually involve quantities measured in mg kg^{-1} diet and larger. Thus, the possibility that cadmium is an essential element in ultratrace quantities cannot be dismissed. Other findings supporting cadmium essentiality are that cadmium slightly stimulated the growth of suboptimally growing rats, and cadmium showed transforming growth factor activity or stimulated the growth of cells in soft agar. Because it is consistently associated with metallothionein, cadmium may have some biochemical effects via this biosubstance. Further study is needed before any substantial hypothesis about a biochemical function or physiological role, and conditions that might enhance the need for cadmium, can be presented.

Cobalt

Among the elements required for human health, cobalt is unique because its need cannot be satisfied by the mineral form. The only known function for cobalt is as the essential component at the center of the corrin ring of cyanocobalamin. Cyanocobalamin, also known as vitamin B_{12}, is a cofactor for three mammalian enzymes: methylmalonyl-coenzyme A mutase, methionine synthase and leucine 2,3-aminomutase. Biochemical signs of vitamin B_{12} deficiency include increased plasma homocysteine and urinary formiminoglutamate and methylmalonate; physiological signs include megaloblastic anemia, spinal cord demyelination, and peripheral neuropathy. Pathological consequences of vitamin B_{12} deficiency include pernicious anemia, memory loss, dementia, irreversible neurological disease, called subacute combined degeneration of the spinal cord, and death. It has been suggested that cardiovascular disease associated with elevated plasma homocysteine is another consequence of inadequate vitamin B_{12} intake. Factors enhancing the susceptibility to vitamin B_{12} deficiency include atrophic gastritis, *Helicobacter pylori* infection, gastrointestinal bacteria overgrowth caused by achlorhydria, celiac disease, drugs such as proton pump inhibitors and oral biguanides (used in the treatment of type II diabetes), vegetarian diets, and nitrous oxide anesthesia.

Cyanocobalamin is made in nature only by microorganisms; some of these are found in the rumen of ruminants. Thus, ruminants are the only animals that can have their vitamin B_{12} requirement fulfilled by the dietary intake of cobalt. Like other nonruminant animals, humans must get their requirement for cobalt by consuming foods rich in vitamin B_{12}.

Fluorine

It has been reported that goats fed less than 0.3 mg fluoride kg^{-1} diet, when compared to goats fed 1.5–2.5 mg fluoridekg^{-1} diet, exhibited decreased feed efficiency, depressed growth, histological changes in kidney and endocrine organs, and reduced life expectancy. These rather general effects have not been well accepted as evidence that fluorine is essential, because it is known that high or pharmacologic doses of fluoride have apparently beneficial effects such as the prevention of tooth caries; improvement of back pain and bone density in patients suffering from osteoporosis; improvement in iron absorption or utilization in mice and rats fed a diet marginally sufficient in iron; and slight improvement in the growth of suboptimally growing rats and chicks. High or pharmacologic amounts of fluoride also have been found to depress lipid absorption, alleviate nephrocalcinosis induced by high dietary phosphorus, and alter soft-tissue calcification caused by magnesium deprivation.

Although fluoride is not generally considered an essential in the classical sense for humans, it is still considered a beneficial element because of its ability to protect against pathological demineralization of calcified tissues (especially teeth). However, the possibility that fluoride is an essential nutrient should not be dismissed. It seems possible that fluoride could have a role in biological mineralization.

Germanium

Of all the elements discussed here, the evidence for essentiality is probably the weakest for germanium. Nonetheless, it is included here because of the attention germanium is receiving as an over-the-counter supplement. Germanium is touted as having anticancer properties because some organic complexes of this element have been found to inhibit tumor formation in animal models. It has been suggested that these compounds have such an effect because they influence immune function. A low germanium intake has also been found to alter bone and liver mineral composition and decrease tibial DNA in the rat. Additionally, germanium has been found to reverse changes in rats caused by silicon deficiency. However, more extensive and specific deprivation signs by at least one more research group are required before germanium can be considered a serious candidate for essentiality.

Lead

Lead is best known for its intoxicating properties. Nonetheless, some scientists strongly contend that lead is an essential element. One research group has reported that rats and/or pigs fed about 30 μg lead

kg^{-1} diet, when compared to those fed about 0.8 mg lead kg^{-1} diet, exhibited depressed growth; anemia; elevated serum cholesterol, phospholipids, and bile acids; disturbed iron metabolism; decreased liver glucose, triglycerides, low-density-lipoprotein cholesterol and phospholipids; increased liver cholesterol; and altered blood and liver enzymes. Another research group reported findings that indicated lead enhanced growth and improved hematocrits and hemoglobin concentrations of iron-deficient rat pups through a pharmacological action. Confirmation of the substantial animal deprivation findings by additional research groups is not likely to produce any widespread acceptance of lead having properties other than toxic. A defined biochemical function for lead will probably be required before this element is accepted as essential. Many of the reported apparent signs of lead deficiency are associated with iron metabolism; thus, lead possibly has a function that affects the absorption or utilization of iron. Furthermore, if lead is required, its need most likely would be enhanced by the stressor of suboptimal iron status.

Lithium

Like fluoride, lithium is another element with beneficial pharmacological properties; it has been used effectively in the treatment of manic-depressive psychosis. Lithium deficiency (<1.5 mg kg^{-1} diet) reportedly results in depressed fertility, birth weight, and life span, and altered activity in several liver and blood enzymes in goats. In rats, lithium deprivation (5–15 $\mu g\,kg^{-1}$ diet) results in depressed fertility, birth weight, litter size, and weaning weight. Additionally, these lithium-deficient rats had a decreased lithium content in testes, seminal vesicles, and epididymis, and relatively high concentrations of lithium in the pituitary and adrenal glands even after two generations on a low-lithium diet. These findings suggest that lithium may have a role in the regulation of some endocrine function. Other *in vitro* biochemical actions providing circumstantial support for lithium essentially include the stimulation of growth of some cultured cells and insulinomimetic action. Its ability to affect mental function perhaps explains the report that the incidence of violent crimes is less in areas with drinking water containing relatively high amounts of lithium.

Although two research groups using two animal species have reported similar signs of lithium deficiency (depressed growth and fertility), the signs were not very marked or specific; this has inhibited the widespread recognition that lithium is possibly essential. Additionally, the wide difference between apparent deficient intakes (based on rat experiments) and typical daily intakes inhibits the suggestion that lithium could be of practical nutritional importance. Thus, although more credence should probably be given to lithium essentiality, this most likely will not occur until a specific biochemical function is identified for this element.

Manganese

The essentiality of manganese for animals has been known for over 50 years. Deficiency causes testicular degeneration (rats), slipped tendons (chicks), osteodystrophy, severe glucose intolerance (guinea-pigs), ataxia (mice, mink), depigmentation of hair, and seizures. Manganese is a cofactor for enzymes involved in protein and energy metabolism, antioxidant action, and mucopolysaccharide synthesis. These enzymes include the metalloenzymes manganese-dependent superoxide dismutase, pyruvate carboxylase and arginase, and the manganese-activated enzymes phosphoenolpyruvate carboxykinase, glycosyl transferases, glutamine synthetase, and farnesyl pyrophosphate synthetase.

With such extensive knowledge about the essential nature of manganese, it is surprising that only a few descriptions of possible human manganese deficiency have come forth. The only description of an unequivocal case of human manganese deficiency was of a child with postoperative short bowel receiving over 90% of her nutrition parenterally, which was low in manganese; she developed short stature and brittle bones. In one attempt to induce manganese deficiency, men were fed a purified diet providing only 0.11 mg of manganese per day. They exhibited decreased serum cholesterol and a fleeting dermatitis. Calcium, phosphorus, and alkaline phosphatase activity increased in blood. However, short-term manganese supplementation (10 days) did not reverse these changes; perhaps a longer period of supplementation would have. Regardless of the lack of descriptions of human manganese deficiency, manganese is thought to be of nutritional importance by some scientists because low dietary manganese or low blood and tissue manganese has been associated with osteoporosis, epilepsy, atherosclerosis, impaired wound healing, and increased susceptibility to cancer. Thus conditions under which a low manganese status results in pathological consequences for humans needs to be determined. Perhaps exposure to a stressor that enhances the need for one of the manganese metalloenzymes, or high intakes of inhibitors of manganese absorption such as calcium, phosphorus, iron, fiber, phytate and polyphenolic compounds, will result in changes that demonstrate the nutritional importance of manganese for humans.

Molybdenum

Descriptions of human molybdenum deficiency are few. A patient receiving prolonged parenteral nutrition acquired a syndrome described as 'acquired molybdenum deficiency.' This syndrome, exacerbated by methionine administration, was characterized by high blood methionine, low blood uric acid, and low urinary uric acid and sulfate concentrations. The patient suffered mental disturbances that progressed to a coma. Pathological changes occurring in individuals with a genetic disease that results in a sulfite oxidase (a molybdoenzyme) deficiency include increased plasma and urine sulfite, sulfate, thiosulfate, *S*-sulfocysteine and taurine; seizures, brain atrophy/lesions, and mental retardation; dislocated lenses; and death at an early age.

Molybdenum functions as an enzyme cofactor. In humans, three molybdoenzymes have been identified; these are aldehyde oxidase, xanthine oxidase/dehydrogenase, and sulfite oxidase in which molybdenum exists as a small nonprotein factor containing a pterin nucleus. Molybdoenzymes oxidize and detoxify various pyrimidines, purines, and pteridines; catalyze the transformation of hypoxanthine to xanthine and xanthine to uric acid; and catalyze the conversion of sulfite to sulfate. A high sulfur amino acid intake could possibly be a stressor that would help induce signs of molybdenum deficiency in humans. In addition, because all of the molybdoenzymes can be involved in the detoxification of xenobiotic compounds, perhaps humans stressed by an exposure to certain xenobiotics have an enhanced need for molybdenum.

Nickel

The reported signs of nickel deprivation are extensive and are from several animal species. However, many of the described signs, especially those from early studies of nickel essentiality, may be of questionable validity because of the possibility that they were manifestations of pharmacological actions of nickel. More nutritionally designed subsequent studies indicate that signs of nickel deprivation include increased blood pressure, kidney damage, decreased sperm number and motility, and altered cyanocobalamin (vitamin B_{12}), glucose, triglyceride and sulfur amino acid metabolism. As with other ultratrace elements, the nature and severity of nickel deprivation signs are affected by diet composition or nutritional stressors; these include low dietary protein, inadequate dietary iron, high dietary simple sugars, and stressors that alter sulfur amino acid or labile methyl metabolism such as cyanocobalamin, folic acid or pyridoxine deficiency, or homocysteine supplementation.

Nickel has an essential function as an enzyme cofactor in lower forms of life. Nickel is a component of urease from bacteria, mycoplasma, fungi, yeast, algae, higher plants, and invertebrates. Nickel is present in hydrogenases from over 35 species of bacteria; it may be a common constituent of hydrogenases that function physiologically to oxidize rather than evolve H_2. Nickel is a component of carbon monoxide:(acceptor) oxidoreductase found in acetogenic, methanogenic, phototrophic, and sulfate-reducing anaerobic bacteria, and of a tetrapyrrole known as factor F_{430} found in methyl-*S*-coenzyme-M reductase that converts CO_2 to methane in methanogenic bacteria. Nickel is required for the hydrogenase gene to be expressed in *Bradyrhizobium japonicum*. These essential functions suggest that nickel may have a role in higher animals, including humans as a cofactor or structural component in an enzyme. It is also possible that nickel has a very basic function at the molecular level because *in vitro* studies show that nickel is a calcium channel blocker and can activate the Ca^{2+} 'receptor' on the osteoclast to elicit cytosolic Ca^{2+} signals.

The findings showing that nickel is so dynamic in lower forms of life, and that experimental animals are affected by nickel deprivation, provide almost overwhelming circumstantial evidence that nickel is essential for humans. Only the lack of a defined biochemical function in some higher form of life inhibits the general acceptance of nickel essentiality.

Rubidium

Rubidium is another recent addition to the list of possibly essential elements. Compared to goats fed 1 or 10 mg rubidium kg^{-1} diet, goats fed less than 0.28 mg rubidium kg^{-1} diet exhibited decreased food intake, growth, and life expectancy, and increased spontaneous abortions. These rather general deprivation signs need to be confirmed in additional species, preferably in another research setting, before strong consideration can be given to the possibility of rubidium essentiality. The reported signs of deficiency do not give insight into a possible biochemical function for rubidium. Because rubidium resembles potassium, it may have a role similar to potassium. Also, some findings have been obtained suggesting that rubidium possibly has a neurophysiological function. The rubidium content of brain differs significantly between defined functional regions, and rubidium can apparently enhance the turnover of brain norepinephrine (noradrenaline) and cause electroencephalogram activation in monkeys and rats.

Silicon

Most of the signs of silicon deficiency indicate an aberrant metabolism of connective tissue and bone. Silicon-deprived chicks exhibit structural abnormalities of the skull, depressed collagen content in bone, and decreased tibial articular cartilage, water, hexosamine, and collagen. Silicon-deprived rats exhibit skull structural changes, increased humerus hexose, decreased humerus hydroxyproline and alkaline and acid phosphatase activity, and decreased plasma ornithine aminotransferase activity (a key enzyme in collagen synthesis). The response to silicon deprivation can be modified by low dietary calcium and high dietary aluminum. For example, rats fed a diet low in silicon and calcium and high in aluminum accumulated high amounts of aluminum in the brain; silicon supplements prevented the accumulation.

Silicon essentiality in higher animals is supported by its essentiality for lower forms of life; it has a structural role in diatoms, radiolarians, and some sponges and affects gene expression in diatoms. It is required by some higher plants, including rice. Other findings supporting essentiality include findings that silicon is localized in the active growth area, or osteoid layer, and within the osteoblasts in young bone of mice and rats; silicon is chemically combined with the glycosaminoglycan fraction in several types of connective tissue; and silicon is required for maximal bone prolylhydroxylase activity in bone tissue culture.

The distribution of silicon in the body and the biochemical changes in bone caused by silicon deprivation indicate that silicon has an essential function that influences bone formation by affecting cartilage composition, and ultimately cartilage calcification. It is also likely that silicon has a function important for collagen formation; thus, it may be needed for proper wound healing.

Tin

The evidence for tin essentiality is quite weak. One research group has reported that tin deficiency results in impaired growth, decreased feed efficiency, alopecia, depressed response to sound, and changes in mineral concentrations in various organs. Other findings supporting tin essentiality include findings that it influences heme oxygenase activity, and has been associated with thymus immune and homeostatic functions. The oxidation reduction potential for $Sn^{2+} \Leftrightarrow Sn^{4+}$ is 0.13 V, which is near the potential of flavin enzymes. Thus, tin possibly has a role in some redox reaction.

Vanadium

The most substantive evidence for vanadium essentiality has come from deficiency studies with goats and rats. Compared to goats fed 500 μg vanadium kg^{-1} diet, goats fed 1–9 μg vanadium kg^{-1} diet exhibited an elevated abortion rate and depressed milk production. About 40% of kids from deficient goats died between days 7 and 91 of life, with some deaths preceded by convulsions; only 8% of kids from vanadium-supplemented goats died during the same time. Also, skeletal deformations were seen in the fore legs, and fore foot tarsal joints were thickened. Vanadium deprivation in the rat reportedly increases thyroid weight and decreases growth. The rat studies also showed that stressors that change thyroid status or iodine metabolism enhanced the response to vanadium deprivation. Factors that reduce vanadium absorption, including high dietary iron, aluminum hydroxide and chromium, may also enhance the response to a low dietary vanadium intake.

Vanadium has an essential function as an enzyme cofactor in lower forms of life. Vanadium is a component of some nitrogenases that reduce nitrogen gas to ammonia in bacteria, and for bromoperoxidase, iodoperoxidase, and chloroperoxidase in algae, lichens, and fungi, respectively. The haloperoxidases catalyze the oxidation of halide ions by hydrogen peroxide, thus facilitating the formation of a carbon–halogen bond. Other findings supporting vanadium essentiality have come from *in vitro* studies with cells and pharmacologic studies with animals. These studies show that vanadium has insulin-mimetic properties, stimulates cell proliferation and differentiation, affects cell phosphorylation–dephosphorylation, and affects oxidation reduction processes.

Because vanadium is so pharmacologically active, identification of a specific biochemical function for vanadium in higher animals and humans will be necessary to define the nutritional importance of this element. Based on the circumstantial evidence for essentiality, vanadium possibly has an essential function in some enzyme reaction, perhaps one that is involved in thyroid metabolism.

See also: **Arsenic**: Properties and Determination; Requirements and Toxicology; **Boron**; **Cadmium**: Properties and Determination; Toxicology; **Cobalamins**: Physiology; **Copper**: Properties and Determination; Physiology; **Fluoride**; **Lead**: Properties and Determination; Toxicology; **Lithium**; **Manganese**; **Tin**

Further Reading

Bowman BA and Russell RM (2001) *Present Knowledge in Nutrition*, 8th edn. Washington, DC: ILSI Press.

Nielsen FH (1998) Ultratrace elements in nutrition: current knowledge and speculation. *Journal of Trace Elements in Experimental Medicine* 11: 251–274.

Nielsen FH (2000) Importance of making dietary recommendations for elements designated as nutritionally beneficial, pharmacologically beneficial, or conditionally essential. *Journal of Trace Elements in Experimental Medicine* 13: 113–129.

Nielsen FH, Johnson WT and Milne DB (eds) (1996) Proceedings of the workshop: approaches, endpoints and paradigms for RDAs of mineral elements. *Journal of Nutrition* 126 (suppl) 9S.

O'Dell BL and Sunde RA (eds) (1997) *Handbook of Nutritionally Essential Mineral Elements*. New York: Marcel Dekker.

Shils ME, Olson JA, Shike M and Ross AC (eds) (1999) *Modern Nutrition in Health and Disease*, 9th edn. Baltimore: Williams and Wilkins.

World Health Organization (1996) *Trace Elements in Human Nutrition and Health*. Geneva: World Health Organization.

TRADITIONAL FOOD TECHNOLOGY

S Sefa-Dedeh, University of Ghana, Legon, Accra, Ghana

Introduction

The types of foods consumed by any population, methods of handling, marketing, distribution, processing, and utilization are deeply rooted in tradition and experience. In Africa, the conversion of the bulk of agricultural produce into forms readily utilizable by the population is achieved using simple indigenous methods. The technologies applied may be classified as traditional because they have strong links with the rural traditional environment. These technologies are also used in the towns and cities. The scientific principles behind the various unit operations may be the same as found in modern food technologies, but the mode of application may be different.

In general, traditional food technology may be conceived as encompassing the methods and tools used to transform agricultural commodities into intermediate or finished products. Many innovative techniques are applied to achieve various objectives and provide a wide variety of products which contribute to the curtailment of postharvest losses and national food delivery. Traditional food technology in many cases affects the economic and social life of the operators and may contribute to family nutritional status. Examples of these technologies found in Africa are described below.

Cereal Processing

The processing of cereals is very important with respect to the quantity, type, and quality of products. **Table 1** is a summary of processing methods for selected cereal foods. Important unit operations include soaking in water, size reduction, aflatization, aging or fermentation, cooking, and packaging. (*See* **Cereals**: Contribution to the Diet; Dietary Importance.)

Soaking allows imbibition of water into the grain and also facilitates size reduction. The soaking process may be prolonged for 2–3 days when dealing with very dry grains and this may initiate fermentation.

The production of cereal flour may be achieved by using manual methods (e.g., earthenware mashing bowl, stone grinding, or mortar and pestle), or mechanical size reduction. Traditional mashing bowls are made from clay and these have grooves on the mashing surface to facilitate size reduction. Many designs of mortar and pestles are available for different purposes. The disk attrition mill is very popular and can be found in many villages. (*See* **Flour**: Roller Milling Operations.)

In the processing of some cereal products, such as komi, maasa, and aboloo, approximately one-third of the flour is cooked into a stiff gelatinized paste. This paste is mixed with the remaining flour to form aflata. The process referred to as aflatization is critical for the development of texture in the finished product.

Three types of materials may be fermented in cereal processing: the grain (soaking of dry grain), dough or slurry (ogi, ga kenkey, fante kenkey), and wort extracts (alcoholic beverages, e.g., pito, burukutu, kaffir beer, amgba, busaa, and kibuku). In most cases, souring accompanies fermentation, and the process allows the material to develop desired flavor, texture, and processing qualities. Lactic acid fermentation is known to play a significant role. The fermentation of wort extracts in the production of alcoholic beverages is achieved through the activities of yeasts. In Ghana, some traditional processors apply an innovative technique to initiate fermentation in the wort extracts. A woven belt, about 3–4 cm wide and made from

Table 1 Unit operations in cereal processing

Product	Unit operation														
	Soak	Dehull	Sprout	Mill	Slurry	Dough	Aflata	Age or ferment	Dry	Roast	Sieve	Fry	Cook	Boil	Package
Aboloo															
Baked[d]	+[1]	–	–	+[2]	–	–	+[3]	+[4]	–	–	–	–	+[6,a]	–	+[5]
Steamed[d]	+[2]	+[1]	–	+[3]	–	–	+[4]	+[5]	–	–	–	–	+[7,b]	–	+[6]
Agidi[d]	+[2]	+[1]	–	+[3]	+[4]	–	–	+[6]	–	–	–	–	+[7,c]	–	+[8]
Aliha[d]	+[1]	–	+[2]	+[4]	–	–	–	+[7]	+[3]	–	+[6]	–	–	+[5]	–
Apapransa[d]	–	–	–	+[2]	–	–	–	–	–	+[1]	–	–	–	+[3]	–
Apiti[d]	+[1]	–	–	+[2]	–	–	+[3]	–	–	–	–	+[4]	–	–	–
Dzowe[d]	–	+[2]	–	+[3]	–	–	–	–	–	+[1]	–	–	–	–	–
Fante kenkey[d]	+[1]	–	–	+[2]	–	+[3]	+[5]	+[4]	–	–	–	–	–	+[7]	+[6]
Fula[e]	+[1]	–	–	+[2]	–	–	–	–	–	–	–	–	+[4,c]	–	–
Hausa koko[d,f]	+[1]	–	–	+[2]	+[3]	–	–	–	–	–	–	–	+[4,c]	–	–
Jolly kakro[d]	+[1,2]	–	–	+[1,2]	–	–	–	–	–	–	–	+[3]	–	–	–
Komi akpa[d]	+[1]	–	–	+[2]	–	–	+[3]	+[4]	–	–	–	–	+[6,b]	–	+[5]
Maasa[d,e,f]	+[1]	–	–	+[2]	+[6]	+[4]	+[3]	+[5]	–	–	–	+[7]	–	–	–
Nmeda[d]	+[1]	–	+[2]	+[4]	–	–	–	–	+[3]	–	+[6]	–	–	+[5]	–
Ahei[d]	+[1]	–	+[2]	+[4]	+[5]	–	–	–	+[3]	–	+[7]	–	–	+[6]	–
Pito[f]	+[1]	–	+[2]	+[4]	+[5]	–	–	+[7]	+[4]	–	–	–	–	+[6]	+[8]
Ogi[d]	+[1]	–	–	+[2]	+[3]	–	–	+[5]	–	–	+[4]	–	–	+[6]	–
Injera[d]	–	–	–	+[1]	+[2]	–	–	+[3]	–	–	–	–	+[4]	–	–
Komi[d]	+[1]	–	–	+[1]	–	+[3]	+[5]	+[4]	–	–	–	–	–	+[7]	+[6]
Egblem[d]	+[1]	–	–	+[2]	–	+[3]	+[6]	+[4]	–	–	+[5]	–	+[8,a]	–	+[7]
Esofoo dokon[d]	+[1]	–	–	+[2]	–	+[3]	+[5]	+[4]	–	–	–	–	–	+[7]	+[6]
Nsiho[d]	+[1]	–	–	+[2]	–	–	+[3]	–	–	–	–	–	–	+[5]	+[4]

[1–8], Sequence of unit operations; + unit operation used; – unit operation not used. [a]Baking; [b]Steaming; [c]Slurry cooked to thick paste; [d]Maize. [e]Millet; [f]Sorghum; [g]Rice; [h]Teff.

Table 2 Packaging materials for some cereal products

Product	Name	Portion	Condition
Abifo	*Musa paradisiaca*	Leaf	Fresh, green
Agidi	*Marathocloa purpurea*	Leaf	Fresh, green
Apiti	*Megaphrynium macrostachyum*	Leaf	Fresh, green
Aboloo			
Baked	*Megaphrynium macrostachyum*	Leaf	Fresh, green
Steamed	*Baphia nitida*	Leaf	Fresh, green
Domebla	*Zea mays*	Sheath	Dry
Egblem	*Sterculia trigachanta*	Leaf	Fresh, green
Esofoo dokun	*Sterculia trigachanta*	Leaf	Fresh, green
Etsew	*Musa paradisiaca*	Leaf	Dry, brown
Ewule bolo	*Musa paradisiaca*	Leaf	Fresh, green
Fante kenkey	*Musa paradisiaca*	Leaf	Dry, brown
Fonfom	*Musa paradisiaca*	Leaf	Fresh, green
Komi	*Zea mays*	Sheath	Dry
Ga kenkey			
Nkyekyerewa	*Zea mays*	Sheath	Dry
Nsiho	*Musa paradisiaca*	Leaf	Dry, brown
Osino dokun	*Musa paradisiaca*	Leaf	Fresh, green

strands of fiber, is used to trap yeast cells in the fermenting wort. The belt, which is usually covered with yeast cells, is dried in the sun after fermentation. It is used repeatedly for inoculating fresh wort. (*See* **Lactic Acid Bacteria**.)

Clarification of the wort is also achieved by adding an aqueous extract from the fresh stems of okra (*Hibiscus esculenta*). This allows the sedimentation of colloidal and suspended particles.

Food packaging is an important operation in traditional food processing. The packaging materials are derived from plant parts (Table 2). The material in some cases imparts color and flavor to the food (Figure 1). Methods of packaging vary.

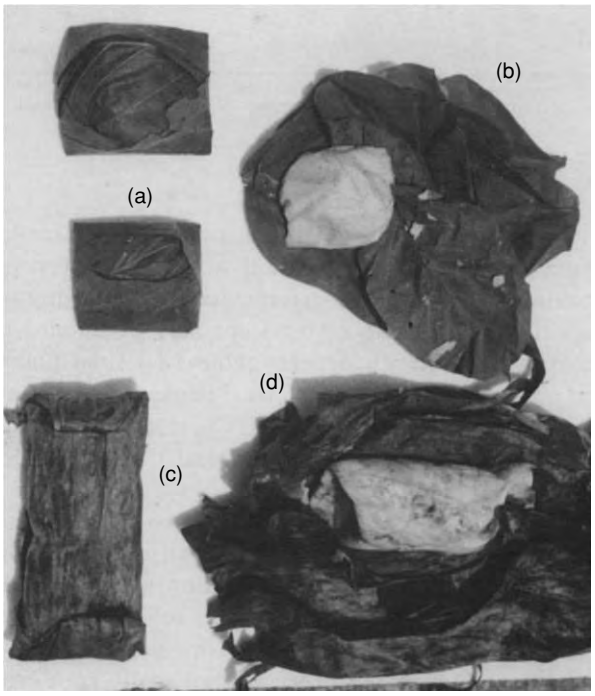

Figure 1 Packaged cereal products: (a) esofoo dokon wrapped in the fresh leaves of *Sterculia trigachanta*; (b) esofoo dokon, unwrapped; (c) nsiho wrapped in dry leaves of *Musa para-disiaca*; (d) nsiho, unwrapped. Reproduced from Traditional Food Technology, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 3 Some products from root crops and some from root crops and other commodities

Product	*Commodity*	*Country*
From root crops		
Agbelima	Cassava	Ghana
Agbeli kakro	Cassava	Ghana
Ampesi	Yam	Ghana
	Cocoyam	
	Cassava	
	Sweet potato	
Fufu	Cassava	Nigeria
Kokonte	Cassava	Ghana
Lafun	Cassava	Nigeria
Mogo oledo	Cassava	Uganda
Tapioca	Cassava	Ghana and Togo
Akyeke	Cassava	Ghana
Attieke	Cassava	Ivory Coast
Eputu	Cassava	Ghana
Fufu	Yam	Ghana
Gari	Cassava	West Africa
Kwonmogo	Cassava	Uganda
Mogo obulo	Cassava	Uganda
Yakeyake	Cassava	Ghana and Togo
From root crops plus other commodities		
Aboloo	Maize and sweet potato	Ghana
Akple	Cassava and maize	Ghana
Fufu	Plaintain and cassava	Ghana
	Cocoyam and cassava	
Kwonkal	Cassava and millet	Uganda
Akankye	Plaintain and cassava	Ghana
Banku	Cassava and maize	Ghana
Komi akpa	Maize and sweet potato	Ghana
Mutama	Cassava and sorghum	Uganda

Root Crop Processing

Root crops, such as cassava (*Manihot esculenta* Crantz), yam (*Dioscorea* spp.) and cocoyams (*Xanthosoma* spp. and *Colocassia* spp.), are important in the African diet. They form the bulk of the starchy staples utilized by the majority of the population. A variety of products are made from root crops. These may be classified into two groups: those processed exclusively from root crops, and those processed from root crops and other commodities (Table 3).

For most products, the processing technology may be at a rudimentary stage using simple techniques and implements; most are home-based operations. Table 4 is a summary of processing methods applied to root crops. Soaking, milling, or grating, fermentation, and drying are important. These processing steps may be applied to detoxify and modify in order to achieve a desired quality and preserve the commodity. (*See* **Cassava**: Uses as a Raw Material; **Vegetables of Tropical Climates**: Root Crops of Lowlands; Root Crops of Uplands.)

Soaking of root crops, especially cassava, is widely practiced. It serves as a means of detoxifying, softening, and initiating fermentation. In addition, soaking aids in manual size reduction. Size reduction achieved by either milling or grating is an important operation (Table 4), and is required in the processing of many root crop products. The operation initiates many biochemical (e.g., cyanide breakdown) and microbiological changes in the products.

Inoculants are added to grated cassava in the processing of agbelima in Ghana and attieke in the Ivory Coast. The inoculant for agbelima is called kudeme. It is made from cassava using different methods. The peeled cassava is blanched, toasted, or soaked in water. It is then wrapped in cloth and allowed to age for 2–3 days, during which it develops a disintegrated appearance. The inoculant is added as the raw cassava is grated, and this is known to affect the texture (smoothness) and promote fermentation.

Fermentation of root crops, especially cassava, is an important unit operation in processing (Table 4). The process contributes to flavor (souring) and texture development, as well as the detoxification (cyanide reduction) of the product. The detoxification is achieved through the activity of microorganisms or endogenous enzymes in the root crops.

Table 4 Unit operations in the processing of some root crop products

Product	*Soak*	*Grind*	*Inoculant*	*Ferment*	*Dewater*	*Roast*	*Steam*	*Cook*	*Dry*	*Pound*
Agbelima	$+^{1}$	$+^{1}$	$+^{2}$	$+^{3}$	$+^{4}$	−	−	−	−	−
Akyeke	−	$+^{1}$	−	$+^{2}$	$+^{3}$	−	$+^{4}$	−	−	−
Attieke	−	$+^{1}$	$+^{2}$	$+^{3}$	$+^{4}$	−	$+^{6}$	−	$+^{5}$	−
Eputu	−	$+^{1}$	−	$+^{2}$	$+^{3}$	−	$+^{4}$	−	−	−
Gari	−	$+^{1}$	$+^{2}$	$+^{3}$	$+^{4}$	+	−	−	−	−
Kokonte	−	$+^{4}$	−	$+^{2}$	−	−	−	$+^{5}$	$+^{1}$	$+^{3}$
Kwanmogo	−	$+^{3}$	−	$+^{2}$	−	−	−	$+^{4}$	$+^{1}$	−
Tapioca	−	$+^{1}$	−	−	−	$+^{2}$	−	−	−	−
Yakeyake	−	$+^{1}$	−	$+^{2}$	$+^{3}$	−	$+^{4}$	−	−	−
Lafun	$+^{1}$	$+^{3}$	−	$+^{2}$	$+^{4}$	−	−	$+^{6}$	$+^{5}$	−
Fufu (Nigeria)	$+^{1}$	$+^{3}$	−	$+^{2}$	$+^{4,a}$	−	−	−	$+^{5}$	−
Fufu (Ghana)	−	−	−	−	−	−	−	$+^{1}$	−	$+^{2}$

$^{1–6}$, Sequence of operations (in some instances steps with consecutive number may occur together); + unit operation used; − unit operation not used.
^{a}Sediment after slurry overnight.

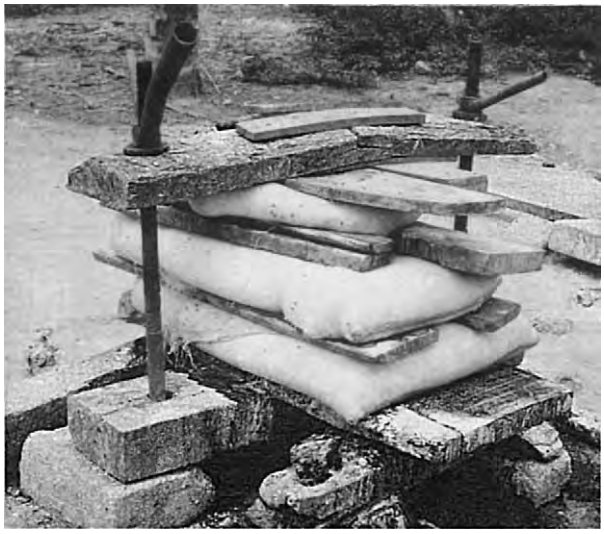

Figure 2 Dewatering of grated cassava for gari processing. Reproduced from Traditional Food Technology, *Encyclo-paedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Simple methods are used at the village level to dewater grated root crop products. These include the placing of heavy stones on the fermenting mash packed in a sack, or pressing the liquid with a manually operated press (Figure 2).

Drying of root crops is widely practiced in Africa. Peeled cassava chips are placed in the sun to dry for periods of 3–10 days. The conditions for drying vary and the product may develop characteristic color and flavor. (*See* **Drying**: Drying Using Natural Radiation.)

Fish Processing

Fish processing is carried out mainly by artisanal fish processors. The major products are dried, salted, fermented, and smoked fish. Table 5 is a summary of unit operations that may be applied in fish processing. (*See* **Fish**: Processing.)

The high perishability of fish demands that it is handled properly immediately after catching and until it is preserved for human consumption. The traditional technologies available for fish processing contribute to the reduction of postharvest losses.

The operations of importance in fish processing are washing, degutting, salting, fermentation, drying, and smoking. These operations contribute to the development of flavor, texture, color, and improved storage characteristics of the products. Different salting systems are used in traditional fish processing, e.g., dry salting and brine salting.

Salting serves to impart flavor and color to fish and has a profound effect on texture. Salted, fermented, or fresh fish may be dried in the sun to reduce the moisture content. In addition, fish may be dried without salting; examples are trigger fish (*Balistes forcipatus*), anchovy (*Engranlis encrasicholus*), and herring (*Sardinella aurita*). Drying of fish in the open sun (Figure 3) may introduce various contaminants into the product.

Smoking is an important operation used to give the combined effects of preservation, drying, and cooking to fish. Different smoking ovens are used. They are generally constructed with mud (Figure 4) and may be cylindrical or rectangular in shape. Recently, the Chorkor smoking oven has been introduced. This oven is reported to have a high fuel efficiency and give increased yield. (*See* **Smoked Foods**: Principles.)

Oil Seed Processing

Oil processing is an important activity in many rural areas in Africa. The major oils processed are palm oil (*Elaeis guineensis*), palm kernel oil (*Elaeis guineensis*),

Table 5 Traditional fish-processing operations – some examples

Product	*Wash*	*Soak*	*Scale*	*Cut*	*Degut*	*Skew*	*Salt*	*Age or ferment*	*Dry*	*Smoke*	*Pack*	*Drain*
Salted fish	$+^{5}$	–	$+^{1}$	$+^{2}$	–	–	$+^{3}$	$+^{4}$	$+^{6}$	–	–	–
Salted fish	$+^{6}$	–	$+^{1}$	$+^{3}$	$+^{2}$	–	$+^{5}$	$+^{4}$	$+^{7}$	–	–	–
Salted fish	$+^{2}$	–	–	–	$+^{1}$	–	$+^{3}$	–	$+^{7}$	–	–	–
Fermented	$+^{2}/+^{7}$	–	$+^{1}$	–	–	–	$+^{4}$	$+^{5}$	$+^{9}$	–	$+^{3}$	$+^{6}/+^{8}$
Fermented	$+^{1}/+^{5}$	$+^{4}$	$+^{2}$	$+^{3}$	–	–	$+^{6}$	–	$+^{7}$	–	–	–
Fermented	–	–	$+^{1}$	–	$+^{2}$	–	$+^{3}/+^{5}$	$+^{4}$	–	–	–	–
Fermented	$+^{1}/+^{3}$	$+^{2}$	–	–	–	–	$+^{4}$	–	$+^{5}$	–	–	–
Smoked	–	–	$+^{1}$	–	$+^{2}$	–	$+^{3}$	–	–	$+^{4}$	–	–
Smoked	$+^{2}$	–	$+^{1}$	–	–	–	$+^{3}$	–	–	$+^{4}$	–	–
Smoked	$+^{1}$	–	–	–	–	–	–	–	$+^{2}$	$+^{3}$		
Smoked	$+^{3}$	–	–	$+^{2}$	$+^{1}$	–	–	–	$+^{4}$	$+^{5}$	–	–
Dried	$+^{1}$	–	–	–	–	–	–	–	$+^{2}$	–	–	–

$^{1–9}$, Sequence of unit operations; + unit operation used; – unit operation not used.

Figure 3 Sun-drying of salted trigger fish (*Balistes forcipatus*) on straw. Reproduced from Traditional Food Technology, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Figure 4 Smoking of fish. Metal sheet (A) is used to cover the oven during smoking. Reproduced from Traditional Food Technology, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

coconut oil (*Cocos nucifera*), groundnut oil (*Arachis hypogaea*), and shea butter (*Butyrospermum parkii*). Although some large-scale modern mills have been set up in some areas, traditional technologies play a significant role in oil processing. (*See* **Ground Nut Oil; Palm Oil.**)

Table 6 is a summary of the unit operations associated with traditional oil processing. The general objective is to extract as much oil as possible from the tissues of the raw materials. Roasting is an important operation in coconut, palm kernel, groundnut oil, and shea butter processing. Simple systems are used for roasting the raw materials. Other operations of importance include aging, which is sometimes associated with fermentation. Boiling of extracts to release oil is a popular operation, normally performed after the extraction treatments, and salt may be added to aid extraction. The major problem in traditional oil extraction technology is the low yield of oil. (*See* **Vegetable Oils:** Oil Production and Processing.)

In addition to oil, the palm (*Elaeis guineensis*) yields a sap which is fermented to give palm wine. Fermentation is achieved as a result of the activity of

Table 6 Unit operations in traditional oil processing

Product	*Material*	*Dry*	*Boil*	*Roast*	*Mill*	*Slurry*	*Age*	*Extract*	*Filter*	*Salt*	*Pound*	*Heat*
Coconut oil	Copra	$+^{1}$	−	$+^{2}$	$+^{3}$	$+^{4}$	−	$+^{5}$	−	−	−	−
Coconut oil	Coconut flesh	−	−	−	$+^{1}$	$+^{2}$	$+^{5}$	$+^{6}$	$+^{3}$	$+^{4}$	−	−
Palm kernel	Kernels	−	−	$+^{1}$	$+^{2}$	$+^{3}$	$+^{4}$	$+^{5}$	−	−	−	−
	Kernels	$+^{2}$	−	$+^{1}$	$+^{3}$	$+^{4}$	−	$+^{5}$	−	−	−	−
Palm oil	Fruits	−	$+^{1}$	−	−	$+^{3}$	−	$+^{5}$	$+^{4}$	−	$+^{2}$	−
	Fruits	−	$+^{1}/+^{3}$	$+^{6}$	−	−	$+^{2,a}$	$+^{7}$	$+^{5}$	−	$+^{4}$	−
	Fruits	−	$+^{1}$	−	−	$+^{4}$	$+^{2}$	$+^{7}$	$+^{5}$	$+^{6}$	$+^{3}$	−
Groundnut oil	Nuts	−	−	$+^{1}$	$+^{2}$	$+^{3}$	−	$+^{4}$	−	−	−	−
Shea butter	Nuts	$+^{2}$	$+^{1}$	$+^{9}$	$+^{6}$	$+^{7}$	−	$+^{8,10}$	−	−	$+^{3,5}$	$+^{4}$
	Nuts	$+^{1}$	$+^{5}$	$+^{2}$	−	$+^{4}$	−	$+^{6}$	−	−	$+^{3}$	−

$^{1–10}$, Sequence of operations; + unit operation used; − unit operation not used; ^{a}Fermentation for 72 h.

yeasts, bacteria, and other agents. Palm wine is drunk in a state of active fermentation. At the village level the product is tapped from the tree daily and is not preserved. The quality may deteriorate with time. In Ghana, palm wine is normally produced from palm trees after they have ceased to give economic yields of fresh fruits. The trees are felled before tapping. In Nigeria, palm wine production involves palm trees that are not felled.

Legume Processing

Grain legumes consumed in Africa are processed using traditional technologies. The products may be treated in the following ways: soaked in water and boiled; roasted; milled into flour; fermented; fried in oil; or steamed. (*See* **Legumes**: Dietary Importance.)

Dehulling of legumes is achieved using manual or mechanical methods. When performed manually it is very laborious, but the mechanical methods are not widespread. This limits the use of legumes in products requiring dehulled flour or meal.

Milling to produce flour is carried out for many products. At the village level this may be achieved by pounding using a mortar and pestle. Multipurpose mechanical grinders may be found in some communities and these are used for flour and paste production.

Fermentation of grain legumes is not widely practiced. In Ghana, two fermented legume products are dawadawa and koligo. Dawadawa is made from the seeds of the African locust bean (*Parkia biglobosa*). It is a food condiment which is popular in northern Ghana and Nigeria. Koligo is also made from the locust bean.

Legumes constitute an important source of protein in Africa. The processing of foods containing blends of legumes and cereals is practiced. These blends may be roasted, boiled, or milled together. Nutritionally, this is a good practice.

Meat Processing

Traditional technologies for meat processing are simple. They generally involve salting, spicing, aging, fermentation, drying, and smoking. Salting, drying, and smoking of meat are very popular among traditional hunters. Their products have a characteristic taste and flavor and may be stored for long periods. (*See* **Meat**: Preservation.)

The treatments applied to meat contribute to the development of color, flavor, and texture. Traditionally processed meats, usually game and wildlife, are highly priced and have wide acceptability. Examples of animals used for meat include giant African snail (*Achatina*), grasscutter (*Thryonomys swinderianus*), squirrel (*Atlantoxerus getulus*), and giant rat (*Cricetomys gambianus*).

Starchy Fruit Processing

Plaintains (*Musa paradisiaca*) and banana (*Musa sapientum*) are starchy fruits which may be processed in the ripe or unripe state. Ripe plantains, with their characteristic sweetness, find wide applications in the processing of various products. These may be boiled, steamed, baked, or roasted after pounding and mixing with other ingredients. Products such as ablongo, apiti, kakro, etc, ukpo gede, and mri jinoko are made from pounded ripe plantains. (*See* **Bananas and Plantains**.)

Dairy Processing

Traditional technology is applied in areas where animal husbandry is part of the social, economic, and agricultural activities. Milk products are processed for local consumption. (*See* **Milk**: Liquid Milk for the Consumer.)

Dairy products are generally consumed sour or curdled. In Kenya, for example, cows' milk is kept

in gourds for 2–3 days to allow fermentation, which leads to souring and the development of a coagulum. The fermented milk, called mala or 'milk that has slept,' has a sour taste and viscous consistency. (*See* **Fermented Milks:** Types of Fermented Milks.)

A product made from cows' milk in Ghana is referred to as wagashi. This is made by pasteurizing fresh milk. The stems of the Sodom apple, *Calotropis procera*, are cut and a milky exudate squeezed out. Salt is added to the plant extract and this is mixed with the pasteurized milk. The milk coagulates and the resulting curd is collected after squeezing the mixture in a cloth. A similar product, referred to as wara, is made in Nigeria.

General Characteristics

The major executors of traditional food technology are women. Most of them have a low educational level, operate from family houses, and have limited access to credit and improved technology. They need to be assisted to contribute better to the provision of healthy nutritious foods to the populations.

The equipment and implements used for traditional food technologies are simple. The labor input is high and efficiency is generally low. Progress has been made in upgrading some of the equipment to make them efficient. In a bid to increase efficiency, it is important that attempts are made to maintain the characteristic quality of traditional foods; otherwise, the improved technologies would not be readily accepted.

The products of traditional food technology are very important in Africa. They provide the nutrients required by the population. Some of the products, however, have very short shelf-lives and there is the need to improve shelf-life and build in convenience.

Other characteristics of traditional foods are the variability in the process and generally poor quality control. Some unit operations are inefficient, laborious, and time-consuming. Despite these disadvantages, the products will continue to play a critical role in Africa's search for solutions to the food problem. For this reason, the indigenous traditional technologies need to be studied and improved, and food scientists and technologists have an important role to play in this work.

See also: **Bananas and Plantains**; **Cassava**: Uses as a Raw Material; **Cereals**: Contribution to the Diet; Dietary Importance; **Drying**: Drying Using Natural Radiation; **Fermented Milks**: Types of Fermented Milks; **Fish**: Processing; **Flour**: Roller Milling Operations; **Ground Nut Oil**; **Lactic Acid Bacteria**; **Legumes**: Dietary Importance; **Meat**: Preservation; **Milk**: Liquid Milk for the Consumer; **Palm Oil**; **Smoked Foods**: Principles; **Vegetable Oils**: Oil Production and Processing; **Vegetables of Tropical Climates**: Root Crops of Uplands; Root Crops of Lowlands

Further Reading

Campbell-Platt G (1980) African locust bean (*Parkia* sp.) and its West African fermented food product, dawadawa. *Ecology of Food and Nutrition* 9: 123–132.

Hulse JH, Laing E and Pearson OE (1980) *Sorghum and Millets: their Composition and Nutritive Value*, pp. 396–478. London: Academic Press.

Oyewole OB and Odunfa SA (1990) Characterization and distribution of lactic acid bacteria in cassava fermentation during fufu production. *Journal of Applied Bacteriology* 68: 145–152.

Sefa-Dedeh S (1990) Harnessing traditional food technology for development. In: Sefa-Dedeh S and Orraca-Tetteh R (eds) *Harnessing Traditional Food Technology for Development*, pp. 7–14. Legon: University of Ghana.

Transmission Electron Microscopy *See* **Microscopy**: Light Microscopy and Histochemical Methods; Scanning Electron Microscopy; Transmission Electron Microscopy; Image Analysis

TRANSPORT LOGISTICS OF FOOD

P van Beek, K Koelemeijer and D J van Zuilichem, Agricultural University, Wageningen, The Netherlands
M P Reinders, Agrotechnological Institute (ATO-DCO), Wageningen, The Netherlands
H F Th Meffert, Veluviaweg 3, Wageningen, The Netherlands

Introduction

Until the 1980s, logistics was synonymous with inventory control, production, and transport planning. In later years, the coordination between departments directly concerned with the flow of goods has attained more importance, covered by the term integral logistics.

The task of integral logistics is to develop control concepts for chain processes which avoid sub-optimization of links and result in win–win structures.

These chain logistics have to be based on knowledge of the processes involved, added value and cost per link, and a framework of obligatory and voluntary limitations. The flow of goods through the consecutive links is defined by:

- The composition of intermediate and/or final products.
- The production processes, and the protection required.
- The allocation of added value and cost throughout the links of the chain.
- The allocation of logistic processes in the chain with respect to the final quality of the product.
- The structure of the flow of information between the links.

The principal question is how to control the flow of goods to reach predetermined criteria of performance: flow time, cost, quality, level of service, flexibility, and added value.

The application of logistic principles in the agrofood industry has revealed specific characteristics of the chains of activities. This has resulted in the formulation of agro (food) logistics.

The specific characteristics relate to:

1. The variation of basic material in quality and composition, within lots of primary products, over the seasons. The final product must still comply with firm specifications.
2. The quality of intermediate and final products is strongly influenced by time and physical parameters (temperature, humidity, and atmospheric composition) in the chain between producer and consumer. Hence, there is a special need for protection.
3. Packaging technology plays an important role in the quality development of, as well as protection of, foods. It also influences the design of processes, size of units, and equipment for physical transformation and preservation by heating, chilling, or freezing.
4. The agrorhythm of the primary production must be obeyed for biological and technical reasons. Nevertheless, a constant flow of final products of constant quality must be achieved.
5. The production of linked products in the case of many intermediate and final products from one primary product.
6. The high expenditure on logistic efforts to bring the primary product in a suitable state to the consumer, which can amount to more than 300% of the value of the primary product.

Agrologistic Chains

While focusing on the agrofood industries, it should be recognized that other segments of agroindustrial activity form comparable logistic systems, including fisheries.

Based on these considerations, agrofood logistics can be defined as 'the integration of organization, planning, and control of acquisition, transport, storage, processing, and packaging related to the movement of foods from the origin of the primary product to the final destination at the point of consumption.'

Information technology, storage and transportation technology, and process technology must be considered to play a most important role in the flow of goods of agricultural origin.

Table 1 gives some idea of the mass flow in the food chain to consumers in Europe.

The following considerations are relevant for the control of chain processes for the distribution of food (and feed) products.

Perishability of Products

The quality loss of primary, intermediate, and final products requires the integration of quality models into the logistic concepts. The final product has to comply with requirements of warranted quality

Table 1 The food basket of European consumers, 1987/1988 (kg per capita per year)

Category	*EC*	*UK*	*Italy*	*Netherlands*
Deep-frozen and frozen[a]	16	27	6	16
Chilled	207	218	177	232
Fresh and chilled	205	124	276	231
Nonconditioned	211	228	195	186
Total	639	597	654	665
Beer	64	81	16	86
Wine	43	10	72	14
Population: EC *c*. 340 million				
Europe *c*. 675 million				

[a]Excluding icecream.
Adapted from: *Agricultural Figures* (NL) (1990) The Hague: LEI.

and availability. These requirements have to be translated into actions within the links. Two main parameters can be recognized per link: residence or throughput time – a logistic parameter – and physical condition, e.g., temperature – a technical parameter. Generally, both fall within certain distribution boundaries.

The translation of requirements into conditions or actions in the chain is of crucial importance for the final result, in terms of cost and quality. Based on a fundamental knowledge of products and transformation processes, the cost-sensitivity of alternatives which comply with the given quality requirements of the market can be studied.

Autonomy of Actors

Agroindustrial chains consist, in most cases, of a large number of links, acting autonomously, and this hampers control and optimization. Concepts have to be developed which allow for effective control and avoid suboptimization; realization of such control concepts needs information technology. Effective and efficient systems and procedures for the exchange of information between the actors in the chain have to be developed.

Autonomy of actors and product perishability require special attention for the following items:

Modeling The impact of actions in the fields of process, storage, and transport technology can be evaluated with models that couple process parameters to product qualities. Technological models describe the development of quality (of a lot) of a product as result of flow time–environmental factor distributions along the chain. By adding economic information, such models are expanded to logistic models, which allow quality–cost optimization by comparison of various logistic channels. Such integrated models can take account of economic factors explicitly in terms of fixed and variable costs, for equipment, labor and energy.

Telematics Rationalizing the exchange and processing of communications by telematics allows improvement of the competitive edge of enterprises by increasing reach and/or flexibility. The investment in telematics has to be balanced against logistic advantages or savings in the total chain.

Logistic services In recent years logistic services to the customer have become an important part of the marketing strategy of companies. Many enterprises concentrate on core activities, thus creating new opportunities for service industries. Although not specific to the agroindustry, this development increases the number of links in the chain and therefore adds to the number of actors, and consequently can hamper control.

Audits, *ad hoc* or on a regular basis, of all actors in the chain, or commodity boards of a permanent character, are important instruments in accomplishing the necessary cooperation. Strong consumer preferences for high quality and freshness strengthen the need for effective control of logistic operations.

Operations research Agrologistics based on biological, physical, and economic sciences leads to complicated problem solutions. Operations research presents a common mathematical base whereby quantitative models are combined and developed, and algorithms are designed for planning logistic structures and procedures.

These models are integrated together in decision-support systems (DSS) for use in the development of logistic structures in enterprises and chain operations. Such models can be focused on the minimization of specific or general expenditure, with fixed values of performance criteria. The balance of fixed, location-dependent costs, and costs of transportation is one of the instruments of decision.

Logistic structures In comparison with general industry, the logistic structure of an agroindustrial enterprise depends more strongly on the primary product, its properties and availability, and the market for the final product.

Product design can be considered to be a most important variable for the logistic process. The composition of the product, based on primary and intermediate products, and production processes determines the location for production units for components, intermediates, and final products. In addition, telematics, technology and requirements for

production, processing, storage and transportation determine the logistic structure of an enterprise.

Models for the quantitative evaluation of the interdependency between these variables and their combined influence can be a powerful support for decisions on strategic, tactical, and operational levels for the structure and operation of agroindustrial systems and enterprises. The fresh market, in particular, presents examples of the distribution of effort, by translating the aim of a high quality level at the consumer into actions in the links of the logistic chain, e.g., to improve temperature management or to reduce throughput time, or both.

A Systems Approach

The characteristic of agrologistic chains appears to be the rapid quality degradation that results from suboptimal environmental factors and extended lead times. Logistical methods and techniques commonly being used in other industries, such as manufacturing resources planning (MRP) and optimized production technology (OPT) cannot be applied directly to distribution networks for perishable foodstuffs. Control of environmental factors and lead time in a complex system is essential for the success of the total operation. The complexity of the distribution of food products can be demonstrated by an aggregated picture of the logistic system (Figure 1).

Logistics in Organization

The management of an organization defines the goals with respect to product quality (e.g., minimum quality level at delivery, customer service level). In practice, it proves difficult to implement this policy at the operational level. The actors on the workfloor are specialists in their field: sales managers, technicians, operators, who all have different views of the target of the operation. With conflicting aims in the different echelons, the goals defined and agreed at management level cannot be implemented properly in the daily operation. In such cases, the creation of a Logistics Department between administration and operations departments can offer the necessary coordination and integration of effort.

The management must be able to communicate with, to coordinate, and to inspire the different areas of science and technology, contributing to agrofood logistics at the strategic, tactical, and operational levels (Figure 2).

Logistic Processes

A systematic approach to logistic chain operations reveals the basic tasks of conveyance of mass, energy, and information from origin to destination. Transformation and transport can be recognized as basic characteristic processes: product transformation to bring the material to be conveyed into a certain condition, goods transformation to bring it into a shape to comply with the requirements of the transport processes. For agrofood logistics in distribution networks, preservation processes and unitization processes are important links to food and transport technology. Packaging plays a role as well in both goods and product transformation, for unitization as well as conditioning (Figure 3).

Customer Service

Customer service links logistics to marketing. As the output of logistic operations is perceived by the customer in terms of products and services performance, a marketing approach is needed to determine the optimal level of logistic efforts throughout the supply chain.

A logistic system contributes to distribution channel profits by enhancing the efficiency of movement of goods, insuring quality maintenance, and providing time and place utility. In addition to a cost-oriented

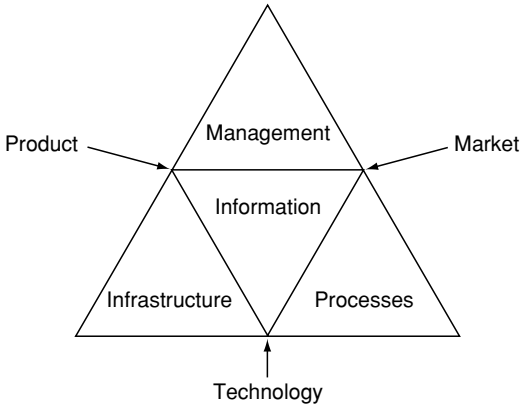

Figure 1 The agrologistic triangle. © ATO-DLO.

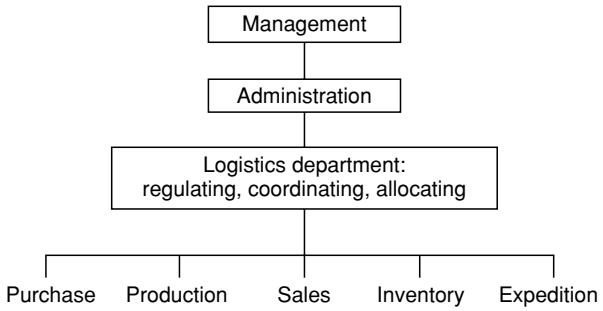

Figure 2 Logistics in the organization of an enterprise. © CIVO-TNO.

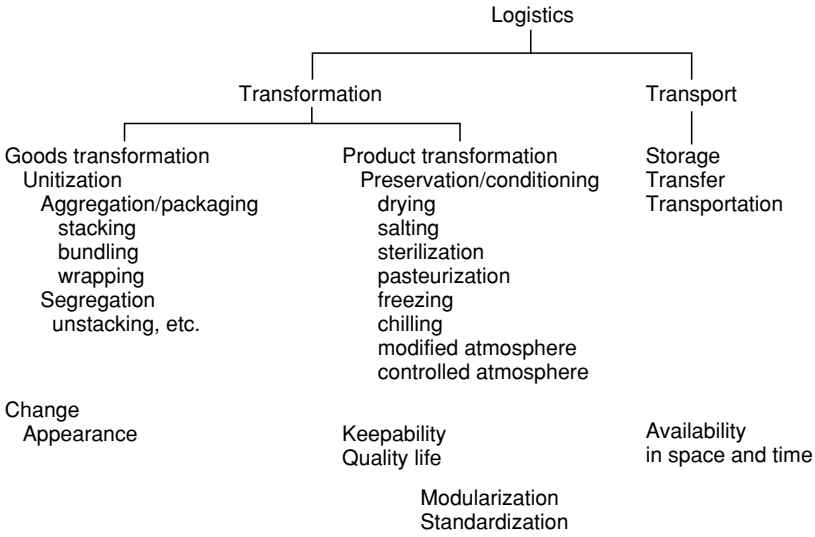

Figure 3 Physical processes relevant for agrofood logistics.

approach, the revenue-generating role of customer service and its use as a tool of competition have increasingly received attention. Logistics defines customer service as 'a process for providing significant value-added benefits to the supply chain in a cost-efficient way.'

In models of logistic operations, customer service is generally considered as a constraint which must be met while minimizing costs. More attention, however, should be given to the maximization of customer service within cost constraints. In this perspective, customer service becomes an instrument of competition.

Relevance of time and place utilities, and quality maintenance, which add value to the product, shifts along the consecutive links of logistic channels from mainly technological to mainly consumer-oriented aspects.

At the central storage and distribution center, mass and volume of units, protection against abuse (mechanical and thermal), and information are important.

On the retail level, reliability, speed, and punctuality of delivery and accurate compliance with the ordered quality and quantity are most important. Services of the preceding link in terms of continuity, communication, flexibility, and claims handling become more important.

Consumers do not explicitly consider technology and logistics, but instead focus on subjective attributes, like sensorial aspects, convenience, nutritive value, keepability, and waste disposal. Retailers translate consumer desires into their own demands.

Marketing activities are traditionally described in terms of the four 'big Ps' of product, place, price, and promotion:

- Product: the commodity that the market is looking for, including purchase of the raw material, processing and packaging, pollution, and politics.
- Place: availability of the product in space and time at consigned points of destination at the convenience of the consignee.
- Price ultimately expresses the monetary value of the product in the marketplace for the consumer, but determines at the same time the span between cost and revenue, the possible profit.
- Promotion is a means of dealing with nonuniform supply and demand. Customer service may be used as an argument in promotion. On the whole, promotion is strongly interrelated with the flexibility of the logistic system to respond to discontinuities in product flow.

Circumstantial factors can be designated the six 'small Ps':

- Purchase reaches forward to the raw material in order to meet the requirements for processing in the chain.
- Processes are necessary to bring the primary product into a state and shape to satisfy the different requirements of handling along the logistic chain, and of usage at the ultimate customer.
- Package indicates not only mass or number, volume, and shape in which the product satisfies the various requirements in the logistic chain, but also the protecting, preserving, and presenting function, especially in the downstream links of the chain.
- Pollution: environmental concern is becoming an important aspect concerning the use of resources, raw materials and energy, contributing to ozone

depletion and the greenhouse effect. Also, product losses and waste disposal (product and package material) must be considered. Product pollution by production, processing, and packaging has to be excluded for reasons of public health.

- Politics interfere on different levels in the distribution of perishable foodstuffs. Public health authorities have developed a strong interest in food processing and packaging, which has recently been paralleled in the field of pollution. Also, prices, conditions, and quality life (date marking) are under political control all along the chain. Regulations, guidelines, and standards for equipment, processes, and labor conditions can greatly affect the logic of logistic chains and, consequently, the cost/quality balance, which demonstrates the need for anticipating research and information.
- Public opinion: apart from political pressure, attitudes and opinions determine the actions of the consumer. Attentive retailers are aware of these marketing aspects, and try to follow trends of consumer preference.

Integration of Logistics and Marketing

The need for efficient distribution of perishable products, i.e., to meet quality requirements at minimum costs, has strengthened the need for integral planning of logistic activities. Customer service, considered as a goal variable and a constraint at the same time, must be analyzed from a total distribution channel perspective. This implies an integration of logistic and marketing theories.

Once the way is open to performing logistic operations in order to maximize customer service, the issue of allocation of channel cost–benefits emerges. This may give rise to alterations in distribution channel structures, resulting in forward or backward integration along the chain, and changes in distribution channel power.

Budgeting and Accounting

The growing need for the quantification of logistic efforts has induced new developments in budgeting and accounting, such as direct product costing (DPC) and direct product profitability (DPP). In view of stricter energy management and rising energy costs, energy accountancy has been used to compare different products with respect to their energy consumption, including processing, handling, and packaging. The same can be done for other cost elements, as long as they are accounted separately.

Integral Cost Accountancy The integral revenue-expenditure concept can be recognized in cost optimum production planning technique (COPPT), which requires an accurate integral cost-accounting system per unit of sales, including fixed and variable, labor, energy and maintenance costs of processes and equipment through all relevant links of the logistic system.

Total cost analysis The formulation of a total cost analysis is not without practical problems; the appreciation of quality development and its balance against capital costs represents a specific problem.

Modeling

Multiechelon models have been developed that integrate the links of the distribution channel. Although customer service is considered, in a narrow sense, as out-of-stock and/or delivery time, these models offer possibilities for modeling logistic operations, and enable users to consider customer service as goal, variable, and constraint.

The Physics of Physical Distribution

Logistics and Technological Processes

Two types of processes play important roles in logistic chains – transformation and transport. Transformation brings the products into a state and shape in which they can endure the rigors of the transport processes.

Goods transformation brings the products into a form convenient for handling in the links of the logistic chain.

Product transformation conditions the product into a state in which it can comply with the requirements of quality at the point of destination.

Transformation

Goods transformation Agricultural products become available in a given form, which is generally not the most convenient for handling in a transport chain. Small products have to be aggregated into larger units; large products have to be segregated into smaller units. Portioning, packing, and utilization are common physical processes in agrofood operations. For fluid products, the implementation of these processes is simple compared to those for solid products which cannot be fluidized. Mechanization of goods transformation is easy to accomplish for fluidizables. Solid products are more difficult to handle. Mechanization closely follows human actions. Sorting and grading are prerequisites for the formation of commercial units for the trade and, as such, are part of customer service.

Goods transformation processes affect the appearance of a product. Generally, an appearance has to

aggregate, to protect, and to present both the product and information about the product. It must also fit into the systems of transport, distribution, information and, most important for perishables, of conditioning.

These requirements can lead to contradictions with conventional solutions, and require special measures with respect to conditioning, design of appearances, and creation of the necessary external conditions. The importance of the functions of an appearance varies with the size of the unit. For the smaller units and articles, the presentation function often outweighs the protection, whereas the larger preparations mainly serve aggregation and protection purposes. With the larger size, the protective function includes facilitation of conditioning, which switches from outside to inside, e.g., pallet versus container (Table 2).

Unitization The formation of units by aggregation to larger appearances, or by segregation to smaller ones, is summarized as unitization. In more detail, stacking, bundling, wrapping, and their counterprocesses are concerned.

Unitization is a necessary step in rationalizing transport processes. Its interference with transformation and conditioning processes should be fully recognized. Methods and materials for unitization, such as glueing, bundling by straps, wrapping in foils or nets, by pallets, slip-sheets, slip-pallets, shipping cases, and ballet boxes, isolate the product from ambient conditions by the dimensions and properties of the material. Thus, more or less protection, e.g., resistance against external conditioning, is built up by insulation or permeability. Perforation of the separating covers and application of pressure differences across the units are means of controlling external conditioning.

Modularization It is obvious that a coordinated relationship between the dimensions of appearances greatly rationalizes handling, utilization of storage, and transportation space and use of equipment. An early example is the unicube concept, developed in The Netherlands, *c*. 1965. In later years, the National Materials Handling Bureau of Australia (*c*. 1980), the International Organization for Standardization (ISO) (in the early 1980s) and the US–Canadian Modularization-Unitization-Metrification Project pursued the same line.

A problem rises from the choice of the module. The early adoption of the manload as the module (1968) has led to diverging developments in unit loads, leading to the continuation of the problem in the domain of transport units. The result is a very unsatisfactory situation of competing unit loads and incompatible transport units.

Standardization Standardization of mass and dimensions is necessary for the operation of an efficient logistic system. Good examples are the Australian and the North American approach. Unfortunately, Europe cannot adopt a uniform, simple system as long as industries adhere to two standard unit loads. In spite of the recommendation of the ISO to prefer the 1200×1000 mm unit as maximum plan view size (MPVS) or gross unit load size (GULS) and central modular standard, some industries stick to 1200×800 mm units, which makes the manload (600×400 mm) the central module on a lower level of aggregation and means a loss of flexibility of configuration within the unit load.

Modeling Goods transformation process modeling has to start from the dimensions of the articles that have to be accommodated in the subsequent aggregated unit, taking account of the necessary tolerances for the design, stacking, deformation, and conditioning.

Design systems The design of modular packaging is greatly facilitated by the existence of transport package and loading pattern selection tables and palettizing charts, as well as graphical and graphical-plus-calculator design systems. Comprehensive computer programs are available to evaluate all factors influencing the basic package design. They are consulted to determine all three dimensions of the unit

Table 2 Thermophysical properties of appearances; results of goods transformation

Domain	*Mass (kg)*	*Diameter (m)*	*Cooling time (h)*[a]	*Overtemperature (°C)*[b]
Product	0.1	0.05	0.2	0.2
Article	0.5	0.1	2.0	1.0
Manload	15	0.3	18	6.0
Unit load	500	1.0	180	Thermal explosion
Transport unit	7500	2.5	900	Thermal explosion
Storage unit	30 000	5.0	3600	Thermal explosion

[a]With external conditioning.
[b]Heat generation 100 W t^{-1}.

load and the manloads which form it. Thus a means is provided to produce an extensive list of options and to evaluate the various compromises rated against the factors contained in the input data. Computer programs also allow for the use of average industry allowances for many factors involved, which reduces the preparation work. With regard to conditioning, the thermophysical properties of product and package have to be considered, as well as penetration by the conditioning medium, which is generally air.

Product transformation By various processes, agrofood products can be transformed into a state in which they can meet the requirements of a given transport process within given limits of quality. The choice is between chilling, freezing, pasteurization, sterilization, salting, and drying, in order of increasing keeping times and decreasing approximation to the fresh product. Food preservation technology is thus linked directly to agrofood logistics. Irradiation, modified-atmosphere (MA) and controlled-atmosphere (CA) treatments also transform the product with respect to its keepability. Generally, these techniques require close temperature control as well. There is also a link to packaging with respect to heat and mass transfer. Refer to individual processes.

Quality life Keeping time or quality life, as dependent on the environmental conditions around the product, limits the throughput time of products in the logistic chain. Parts of the total quality life are often considered separately, as storage life (the first part) or as shelf-life (the last part) of the total economic quality life.

The quality life of a product is limited by the rate of quality loss or deterioration and the imposed limitations. A difference can be made between high-quality life and practical quality life; these differ in the degree of change of characteristic properties of the product. Development of quality with time can be attributed to different principal processes of a physical, chemical, or biological nature, the rates of which are dependent on environmental factors such as temperature, water vapor concentration or pressure deficit, and atmosphere composition. (*See* **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability; Shelf-life Testing.)

These quality-affecting processes act in different ways on two main groups of foodstuffs:

1. Dead products: subject to (bio)chemical and microbiological processes.
2. Living produce: in addition exhibiting biological processes, dissimilation with O_2-CO_2 conversion and heat production.

N.B: fermentation and microbiological growth can also cause CO_2 and heat production. Both groups can be divided into two subgroups:

A. Those in need of a process of maturation to reach maximum consumer quality: red meats, cheeses, fresh fruits (banana, peach, pear, plum).
B. Those which start commercial life at maximum consumer quality: sterilized, pasteurized, and frozen products, white meats, fish, fresh vegetables, mushrooms, bakery products, prepared meals.

Group A products, represented by *c.* 16% of the food basket, need controlled holding times and conditions to reach maximum quality. From then onwards, they require limited throughflow times in the logistic chains, adjusted to the environmental conditions for tolerable quality loss, as do group B products.

Quality models The necessity for quality models to be incorporated into logistic models results from the great number of alternative logistic processes which have to be evaluated in the design of distribution chains. Quality loss of food is largely determined by the cumulative effect of the environmental conditions throughout the handling history of the product. The simple rules of additivity and exchangeability are obeyed better the less the product is processed.

It is impractical and expensive to conduct product spoilage tests for all possible scenarios. Mathematical models can be chosen after verification by simple experiments on the basis of statistical tests on the lack of fit against experimental error. By measuring and modeling deterioration reactions, quantitative data are collected, and models as well as model parameters can be generated. For integration into process or logistic models, quality models for the influence of environmental factors should comply with the following, in the order presented:

1. The broadest possible application on the processes considered.
2. Appropriate consideration of the mechanism of the reaction.
3. Simplicity of determination of parameters, and application.
4. The smallest deviations from experimental values.
5. No discontinuities.

These models may be based on the time-course of specific components or properties from physical, chemical, or sensorial analysis or, in the absence of objective information, on expert judgment, always keeping in mind the possibility of adjustment of the quality scale of the chosen model.

Zero or first-order reaction mechanisms can cover practically all deterioration processes regardless of their nature. Arrhenius-type relationships satisfactorily describe the influences of temperature; linear, hyperbolic, or squareroot relationships require the same type of parameters. For other environmental factors, more direct relationships to the driving force exist. For the combined influence of temperature and gas concentration, special models have to be derived from specific experimental research.

Combined models For temperature-dependent deterioration processes, combined heat-transfer models are most useful. They allow prediction of the effects of management decisions on quality life of conditioned products, especially on the shelf-life. Better protection during transport processes, by reducing the time–temperature load, can be balanced against lower temperature levels or shorter flow-times in one or more links of the chain.

It is obvious that the performance of deterioration tests for all possible scenarios is highly impractical and expensive. Mathematical modeling, computerized on the basis of experimentally established parameters, and a few validation experiments can perform the necessary sensitivity analyses and assist in finding the optimal operational conditions.

Chain reactions A product passing through a logistic chain generally encounters a sequence of environmental conditions. The final state when leaving the *n*th link of the chain is decisive. If the prerequisites of additivity and exchangeability are fulfilled, the result of the chain reaction can be described on the basis of average time and temperature conditions by a sequence of ideal reactors.

Temperature and time distributions Average time and temperature provide only a crude way of describing the flow of goods and the influence of environmental conditions in the chain. Statistical mathematics offers a way of dealing with the time and temperature distributions encountered in practice. The Weibull distribution appears to yield the most general model for time and temperature distributions in the links of logistic chains. The background for the ready modeling of logistic distributions by the Weibull function is given by the fact that the minimum value is given by a solid physical condition, whereas the maximum is open to stochastic influences. Two parameters for scale and shape allow for the modeling of a wide range of practical situations, from plug flow to ideal mixing, with all possible intermediate situations.

Arbitrary distributions A more direct approach can be based on the fractional distribution of time and conditions in the links of the chain. The stepwise calculation of the final quality distribution after a number of links, from the distributions of time and temperature in the links, and the initial quality distribution, can easily be computerized. This approach presents a means of balancing profits of quality retention against expenditure for throughflow time and improvement of conditions, if the cost–revenue analysis is done in appropriate figures.

Standardization As in goods transformation, product transformation also presents reasons for standardization in terms of quality life, temperature sensitivity and, as a consequence, keeping equipment, storage facilities, as well as transport vehicles and transfer processes. Products can be categorized according to keeping temperatures and quality life at

Table 3 Categories of refrigerated foodstuffs[a]

Code	*Reference temperature T_{ref} (°C)*	*Temperature increment B (1/K)*	*Examples*
DF	−18	Positive/negative	Regulations
F	−12		Deep-frozen, frozen products
C04	+4	Negative	Small meat portions, ready-to-eat vegetables, pasteurized meals, salads, large meat portions, sterilized meals, dairy products
C07	+7		
F00	0	Positive	Leafy vegetables, berries, pears
F04	+4		Apples, oranges, potatoes
F07	+7		Green beans, subtropical fruits
F10	+10		Avocado, mango, subtropical, and tropical produce
F13	+13		Bananas, pineapple, tropical produce

Quality life at temperature T: $t = t_{ref} \exp[B(T - T_{ref})]$.

[a]Temperatures and commodities quoted are arbitrary examples. They can vary by national regulation or specific product requirements. From Meffert HFTh (1990b) Chilled foods in the market place. In: *Processing and Quality of Foods*, vol. 3. London: Elsevier Applied Science, with permission.

Table 4 Time–temperature matrix of keepability characteristics[a]

Denomination	*Time, t_{ref} (days)*	*Temperature increment, B(1/K)*			
		< 0.05 Low	*0.05–0.15 Medium*	*0.15–0.25 High*	*> 0.25 Extra sensitive*
Chilled product (category CO4)					
Extra short	< 3	Kitchen-ready vegetables			
		Leeks	Cabbage	Carrots	
Very short	3–7		Minced beef	Poultry	Pork
Short	7–14		Beef slices	Pasteurized milk	
Medium	14–30		Beef joints (CO_2-packed)		
Long	> 30		Lamb carcass (cryo-vac)		
Frozen product (category DF)					
Extra short	< 30				
Very short	30–100	Cooked meals	Cured pork		
Short	100–200		Seafood		
Medium	200–400	Cream	Carcass meat	Vegetables	
Long	> 400	Butter, lamb	Raw dough, poultry	Green peas, fruit	

[a]Examples are given for two categories from **Table 2**.

From Meffert HFTh (1990c) Quality development of foodstuffs under time–temperature conditions in cold chains. In: *Progress in the Science and Technology of Refrigeration in Food Engineering*. Paris: International Institute of Refrigeration, with permission.

a reference temperature and temperature sensitivity, as shown in Tables 3 and 4.

Standardization in such a way facilitates the design of distribution processes in existing chains and of new operations, especially if paralleled by a standardization of storage and transport equipment, and transfer processes, in terms of operational performance, e.g., temperature maintenance by a degree of thermal protection derived from the nonaccomplished temperature fraction. Existing classification and standardization of these types of equipment and processes do not attempt to enter this problem area. ISO, ATP, and Lloyd's classification are based on technical rather than technological performance characteristics.

Recent European Community and national legislation concerning the temperature of frozen and chilled products may create a need for a more operational approach to product, equipment, and process classification, or even standardization, as explained above.

Transport = Storage–Transfer–Transportation

The second, more genuine logistic process is that of transport. Generally, the transport in a logistic chain involves mass, energy, and information (M–E–I). Mass flow downstream and information flow up- and downstream are easily recognized. Less obvious is the transport of energy. For refrigerated products, this aspect is closely related to product transformation and conditioning. Attention is required for reasons of economy in quality protection.

Mass The most apparent transport process is that of mass. However, as agrofood products are mostly in the category of high-volume goods (load density lower than $500\,kg\,m^{-3}$), it is more appropriate to treat transport in agrofood logistics in terms of volume. Figure 4 depicts a scheme of the mass flow in distribution chains for refrigerated foodstuffs.

All activities of moving goods between origin and destination can be understood as transport, storage being transportation with zero speed. Generally, the goods pass a number of storage and transportation links on their way through the chain. All links are interfaced by transfer or transshipment processes.

As a rule, a reduction of the links in the logistic chain reduces cost and losses; in addition, it saves time. This pleads for careful consideration of the necessity of the links, which can lead to integration of processes, such as product transformation by chilling or freezing, into the transportation link. In-transit treatments, thermal or by gas injection, are also practiced for disinfection or disinfestation.

The second important rule is the rationalization, for time and cost reasons, of the transfer process, which links transport closely to goods transformation.

The maintenance of the established condition of the product is a specific task in the transport chain of agrofood products, in addition to the general task of conveyance.

Interrelationships with transformation and conditioning Modular coordination, i.e., dimensional capability of appearances and equipment as elements of a system, is a prerequisite for rationalization of transport processes, and as such subject to standardization. Elements of a covering system for modular coordination are represented by ISO standards on manloads, unit loads, and transport units in ISO 668–1968, 3676–1983, 6780–1988, 3394–1984,

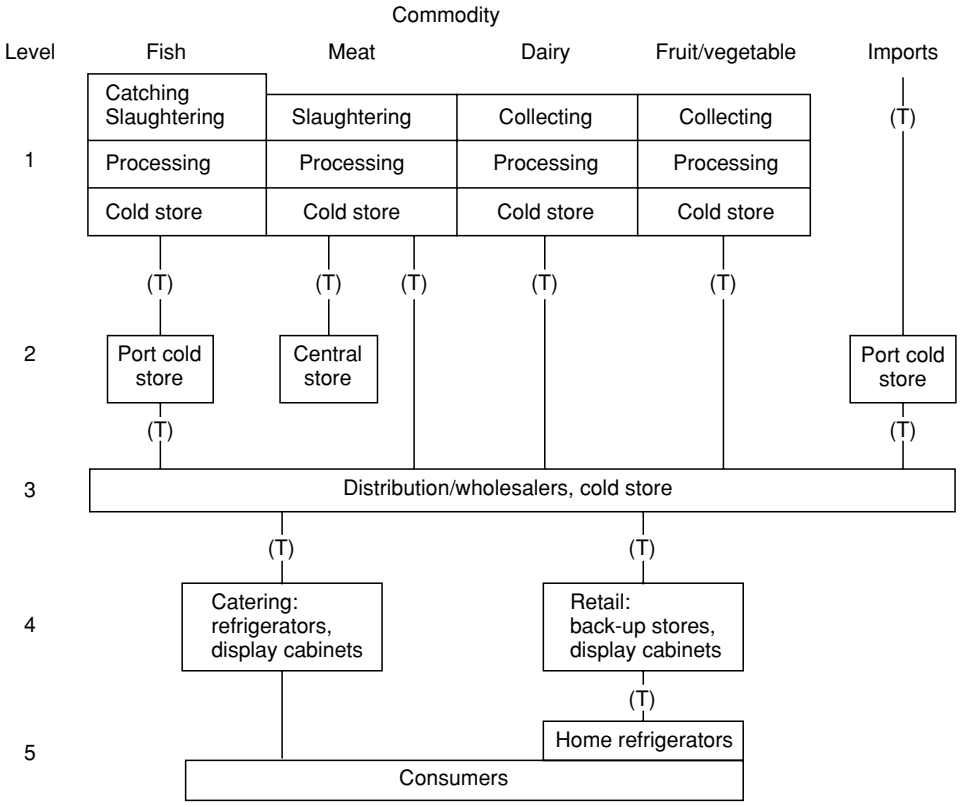

Figure 4 Principal structure of distribution chains for refrigerated foodstuffs (T), transportation and transfers. Note that for products with extremely or very short quality life, the distribution link is cut short by direct links between processing and retail/catering. Adapted from Nowotny S (1988) Computer-modeling of cold chain systems. In: *Cold Chains in Economic Perspective*, Paris: International Institute of Refrigeration.

1496/2–1988. Unfortunately, the system is not practical for conditioned units, because of the space requirements for insulation, air circulation, and the necessary equipment for conditioning. The Australian Standards represent a more logical coordination. Revision of the dimensions of the transport unit for better compatibility with unit loads has so far met much opposition. Elements for a better-fitting system appear in European and US traffic regulations, and in the standards of the European Normalization Centre (CEN) for removable vehicle bodies (swap-bodies), allowing extra width for conditioned road vehicles.

For perishable products, conditioning is necessary to keep the quality loss in logistic chains within a preset span. Distribution of time and conditions, the most important being the temperature, are the governing process variables.

Methods of unitization generally interfere with conditioning processes. Pallets, slip-sheets, slip-pallet, etc., and wrappings influence the distribution of conditioning air. Perforations and build-up of pressure differences around the unit promote air penetration. This must be considered in the choice of tolerances which make the modular system work. The larger the unit handled, the shorter are the transfer times per mass between the links of the transport chain. Shorter times of exposure, together with a small surface-to-volume ratio, improve the thermal protection.

Storage

Facilities for storage in logistic chains are destined to buffer discontinuities of the product flow between supply and demand. They can be subdivided into types differing in task, size, and operational conditions:

1. Primary or production stores. Storage space at the production site serves as a buffer volume between the discontinuities of the inflow from the production process and the outflow of products to secondary stores. Residence times are short. The band of conditions cannot be kept small because of the high turnover. For unitized products, row storage on pallets is preferred.
2. Secondary or central stores. In the case of seasonal production, central stores are loaded within a

short period, but emptied slowly. Volume-saving block storage of unit loads or transport units, if necessary with supporting frames, minimizes investment in conditioned space. The principle of first-in–first-out (FIFO) cannot be fully realized because of the limited accessibility of the units.

Nonseasonal products are supposed to pass central stores rapidly to keep investment in space and product low. Owing to the high turnover rate, the FIFO principle cannot be strictly observed, but through-flow racks, gravity or mechanically driven, can help.

Storage conditions are optimized within a narrow band of environmental factors. Residence times vary according to the type of product, but tend to be a substantial part of quality life.

3. Tertiary or distribution stores. Distribution stores are mere turnover facilities designed for rapid handling of flows of various products. Again the FIFO principle cannot be applied; mixing of products of different age is usual. Incomplete unit loads remain longer in store than complete ones.

   Rack storage, manually or automatically operated for separate accessibility of units, is frequently used, in spite of the large loss of space. Arrival and departure areas are separated, and commissioning areas are provided, mostly with only partial protection against external conditions. Environmental conditions are not optimal, but residence times are comparably short. In partially protected areas, a high flow rate of products must keep the ambient load low.

4. Quarternary stores, retail cabinets. At the retailers, display cabinets of fractions of a cubic meter shelf volume per meter present the product to the consumer. Because of the great influence of the ambient conditions in the shop, and the fact that the product is handled in articles rather than manloads, conditioning is possible only within a broad band. This restriction must be compensated for by extremely short, carefully watched residence times, which also keep at bay the comparatively high costs per volume per time in this link.

   For the reception of a discontinuous flow of goods at the retailer, a conditioned back-up store can form a buffer. This is kept as small as possible. Row storage is used. Conditioning is problematic, owing to the high turnover rate; thus residence times have to be kept at a minimum, depending on consumers' buying attitudes.

Standardization Generally, storage facilities up to tertiary level are built to customer specifications, including layout for unit loads and, in some cases, even whole transport units. Standards apply for technical equipment. Exceptions are the standard cold stores in East Europe, which are classified for layout, size, and storage conditions.

The usage of space by different stacking methods can be given approximately as:

- Block storage 100%
- Movable racks 80–90%
- Row storage 60–80%
- Rack storage 50–60%

depending on the height. Additional space has to be provided, according to the equipment, for the conveyance of goods.

Heat load from ambient and air circulation performance, which determine the quality of conditioning, are chosen for economic considerations, including losses of mass and quality. Small storage and retail cabinets are built serially. Detailed general standards exist only for refrigerated display cabinets (ISO 1992/1–8; ISO 5160/1–2).

Transfer

Processes of cargo transfer or transshipment are characterized by transfer times and degree of protection against ambient conditions. Cargoes are handled in units from manloads to transport units, with various types of equipment from manhandling to large forklift trucks or cranes for whole transport units. Fortunately, the rate of cargo transfer and the degree of protection increase with the size of the unit, and costs decrease per unit of volume (Table 5).

Table 5 Cargo transfer in large and small units: time, protection, cost

Operation	*Transfer time (min)*	*Relative cost (%)*
Long-haul truck,		
18 pallets of 720 kg, with lift truck	Load/unload, 15	100
560 manloads of 28 kg, belt conveyor	Load/unload, 84	140
Delivery truck, 16 deliveries of 20 manloads of 15 kg		
On pallet	Load, 13; unload, 16×5	100
Per manload	37 16×18	250

Flow times in the transfer links are determined by the rate of transfer and the volume to be transferred. The need to condition the transfer area depends on the external climate. The degree of protection in this case is a function of time.

Standardization The standardization of appearances in transport chains has led to a standardization of transfer equipment in terms of speed, volume, and mass that can be handled (conveyors, trucks, cranes).

Transportation

Means of transportation exist for the land, road and rail, sea and air modes. The intermodal container has united these and closed the protected-chain operation. On a regional scale, the removable body or swap-body serves the same purpose.

Road vehicles account for the largest part of the transportation of agrofood products. On a national basis in the EC, up to 90% of the flow of foodstuffs is by road transport that must conform to traffic regulations; this has resulted, practically, in standardization according to size and mass. Containers and removable bodies (swap-bodies) comply with stringent ISO standards for external and internal dimensions. Conditioned road vehicles and swap-bodies can comply better with ISO unit loads than do ISO containers. For both types of vehicle, performance is regulated. Residence times are close to transit times, and almost uniform, except for delivery runs, which are mostly short.

Railcars are larger and comply with standards of the railroad administrations, not necessarily with unit load standards. Ships are built to customers' or builders' standards, and aircraft largely to builders' standards, which include compliance with unit load, pallet, and/or container regulations. Unfortunately, air cargo unit load devices other than air/surface containers are not part of the ISO modular system.

Conditioning depends on the specifications of the customer and the performance of the builder. In sea carriage, narrow bands of conditions are normally realized, much less in air transportation, which leans on the advantage of short transit times.

Residence times under transit conditions are mostly close to transit times, because waiting times are under pressure in all links of the chain. On short distances by rail, waterways, and air, waiting times can be considerable, depending on schedules. The speeds of land and sea transportation are practically the same, between 30 and 60 km h^{-1} (considerably less on inland waterways), but transit times are in fact restricted by the frequency of departures for intermodal combined transportation, and by regulations concerning the active time of drivers for the road leg.

Standardization Standardization of means of transportation is traditionally concerned with external aspects. Only recently has the principle of vehicle design around the cargo been introduced by allowing double-pallet-wide, conditioned road vehicles in the EC and the USA. ISO is also paying attention to this aspect by standardizing the minimum internal dimensions of thermal containers (ISO 1496/2–1988). Modular coordination with unit loads has not been attained yet.

Standards for containers (ISO) and regulations for road vehicles and railcars (Economic Commission for Europe) exist for thermal insulation and refrigeration power, but not for operational characteristics, such as the temperature band.

Standardization of road vehicles, railcars, ships, and aircrafts as carriers is linked to the external dimensions of unit loads or transport units to be accommodated within the limits given by the traffic system, by law or by convention.

Degree of protection For all conditioning processes in the transport chain, the degree of protection against temperature rise can be defined by means of the nonaccomplished temperature rise of the product as the fraction of the difference from the target or reference temperature to ambient.

For refrigerated spaces, the quotient-heat load factor over air circulation is a specific operational criterion of performance. For the evaluation of the conditioning process, the influence of air distribution and product heat load can be introduced by an empirical coefficient.

For transient processes, such as the temperature rise during cargo transfer, transient heat transfer theory offers suitable models for the estimation of the degree of thermal protection, which now depends on time.

Energy If the accumulated energy per conveyed unit of mass by the processes in the chain is recognized as transported energy, the concept of energy transport in distribution chains becomes meaningful, especially if the energy input is accounted for. Particularly for perishable products which have been subjected to a product transformation, the energy accumulated by the transformation process and by the subsequent conditioning for protection against quality loss makes the quality of energy management visible. A very basic approach considers the total energy input for primary production, processing, and handling, as well as for auxiliary materials. More chain-relevant information takes account of chain processes only (Tables 6 and 7).

Low degrees of protection in the links of a chain mean great deviations from the target temperature

Table 6 Energy consumption in distribution chains of refrigerated foodstuffs

Link	Frozen			Chilled		
	Energy rate ($W\,t^{-1}$)	Time (days)	Energy accumulated ($kJ\,kg^{-1}$)	Energy rate ($W\,t^{-1}$)	Time (days)	Energy accumulated ($kJ\,kg^{-1}$)
Product transfer			400			40–120
Central store	13–60	165	100–850	50–125		5–12
Transportation[a]		1	600			570
Distribution store	30–120	46	120–480	100–250	1	9–22
Transportation[b]		1	850		1	800
Retail cabinet	*c.* 2000	10	*c.* 1700	*c.* 10 000	1	*c.* 860
Domestic refrigerator	*c.* 500	23	*c.* 1000	1500–3000	1	125–250
Total		246	4850–5850		4	1834–1979

[a]20.5 t × 500 km. Frozen, +15%; chilled, +10%.
[b]5.6 t × 200 km. Frozen, +15%; chilled, +10%.
Adapted from IIF/IIR (1980).

Table 7 Fuel energy inputs for transportation

Means of transportation	Load capacity (tonnes)	$MJ\,km^{-1}$		$MJ\,t^{-1}\,km^{-1}$ Fully loaded
		Empty	Full	
Ship	49 000			0.127
Rail	1000			0.36
Truck and trailer				
Rural	20.6	12.1	15.3	0.74
Urban	20.6	17.1	26.9	1.31
Truck				
Rural	5.6	8.3	9.8	1.75
Urban	5.6	12.1	14.5	2.59
Van				
Rural	0.4	4.3	4.8	12.0
Urban	0.4	6.7	7.3	18.3

From IIF/IIR (1980).

and bad energy management, leading to additional costs of energy to restore conditions and quality loss.

Finally, the economic consequences of distribution processes boil down to cost accounts (Table 8).

Information Transport processes are controlled by the demands of the market. In order to realize just-in-time (JIT) delivery, the signal for replenishment has to be sent in due time, allowing for the necessary time for the order cycle. A scheme of the information flow in a cold-chain system with computer-assisted guidance, management, and supervision is shown in Figure 5.

Data acquisition and processing Upstream communication along logistic chains mostly concerns the order procedure, which has to be initiated by reliable information from the outlets. Because the signal has to be given in advance of stock depletion, a model is needed to predict the necessary JIT delivery. Further information upstream can be required to redirect the goods flow in case of unforeseen developments of the market. Downstream communication should contain information on the product and its peculiarities, quality life and reference conditions (date marking). More relevant are remaining quality life, required conditions, and handling recommendations. These should be communicated in advance of the goods. They must be matched by information on equipment and conditions in the links. Standardization of information is helpful to reduce variations and mistakes, and to speed up the information flow. Bar-coding and scanning at the interfaces backed by a computerized inventory management system is a powerful means of quality management of an agrologistical chain of operations. Its introduction again requires standardization of information and procedures, and must be the subject of a careful cost–benefit analysis.

Processing of these data for operational purposes is only possible by powerful computer systems with specific software. In the preparatory phase, for the establishment of the necessary tools, even more computing capacity is needed to evaluate possible alternative chains and networks by modeling calculations and simulation in a reasonable time. It is in this application that large computer capacities and high calculation speeds are required.

Inventory management In many cases, the conventional FIFO concept is not able to satisfy the justified quality requirements of the customer. Limitation of throughflow time deals with one aspect of quality management, limitation of temperature with another, whereas only the combination of the two can yield the appropriate measures. Inventory management, therefore, should be based on information about the time–temperature load on the product rather than on time-based information alone (shortest remaining shelf-life instead of FIFO as a management criterion).

Table 8 Average production and distribution costs for three main food chains (Netherlands, Fl kg^{-1})

	Meat and meat products	*Milk and dairy products*	*Fruit and vegetables*	
			Preserved	*Fresh*
Primary product	5.0	0.75	1.43	
Processing and packaging	0.92	0.27	2.71	0.39
Wholesale trade	0.62	0.06	0.54	0.17
Retail trade	3.10	0.27	1.31	0.55
Consumer price	9.64	1.35	5.99	2.54

From Meffert HFTh (1990a) Economic developments pertinent to chilling foods. In: *Chilled Foods, the State of the Art*. London: Elsevier Applied Science, with permission.

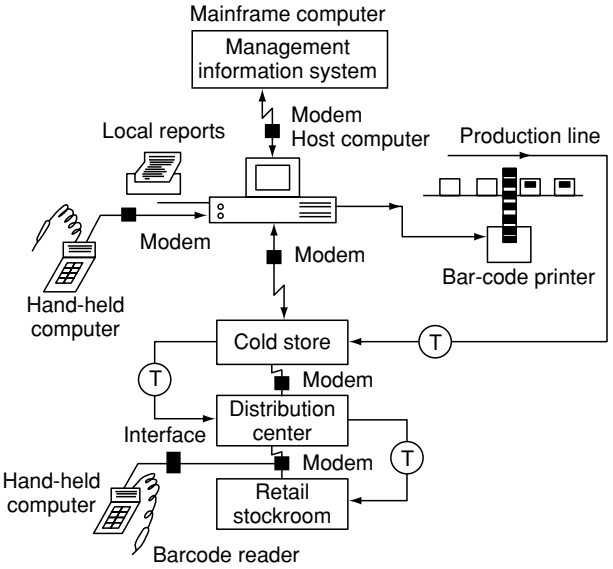

Figure 5 Schematic presentation of a system for a computer-assisted guidance, management, and supervision system of a cold-chain system. From Nowotny S (1988) Computer-modeling of cold chain systems. In: *Cold Chains in Economic Perspective*. Paris: International Institute of Refrigeration.

This requires a new system of control which, in its simplest form, is represented by cheap, physicochemical, time–temperature integrators for each lot of product. A more suitable approach can be expected from the application of microelectronic sensors backed by a computer-assisted inventory management system.

Modern miniaturized, self-contained, time–temperature data-logging sensors, with readable and/or barcode readout, and computer-assisted processing of the information, allow a system to control the flow-time down the chain according to the remaining quality life of the product.

Monitoring product conditions During recent years, a great number of devices for temperature monitoring in conditioned distribution chains have been developed, on mechanical, physicochemical, microbiological, or electronic bases and with indicating or integrating functions, or both.

As it is impossible to keep track of the quality development of every product by time–temperature integration with one or even a handful of built-in deterioration models, solutions can be found in the classification of products and corresponding sensors with direct indication, or sensors readable by a computer-based system which compares the monitored time–temperature load with stored keepability data, and converts the result into information on the remaining quality life expectation of the product.

For a wider use of such instruments, standardization of sensors and information processing equipment is necessary.

Simple time–temperature integrating devices can monitor distribution chain operations for the general purposes of quality management. But even highly sophisticated monitoring systems remain in need of expert supervision, because of the general rule that the sensor is not identical with the product.

Effects of Scale

Costs per unit of storage and transport rise approximately with the square root of volume. Therefore, large stores and vehicles are greatly favored by logistic enterprises. Whereas production stores can afford uniformly conditioned facilities of maximum size, distribution storage is in need of compartmentation and separate conditioning to meet the requirements of different product groups. Towards the end of logistic chains, at the retail outlets, storage facilities and deliveries become necessarily smaller, dictated by high surface charges and problems of accessibility in cities. At the same time, compartmentation and separate conditioning becomes more problematic.

The same reason, low cost linked to a high degree of protection, favors large appearances for transfer operations. Warehouses exist which handle whole transport units, i.e., containers, porthole or integral reefers, but also flats with superstructures to be covered by an insulated hood when in transit. Mostly,

however, unit loads (on pool-pallets) are the normal appearances for transshipment. Smaller units, carts with supporting superstructures, are favored at the retail level for closed-circuit operations.

As the transportation link tries to maximize deliveries, a trend emerges leading to maximum size, multicompartment, multitemperature, expensive vehicles. This tendency towards large vehicles is counteracted by traffic situations and regulations, which restrict the accessibility of cities and, because of the need for supplementary deliveries JIT, add to the logistic costs.

Also, the degree of protection is generally greater at large facilities and operations, owing to their more sophisticated technical equipment and better supervision of operations. For this reason, large-scale operations in logistic chains for conditioned products serve two purposes – low cost and good protection of quality. Costs per unit per time increase remarkably towards the end of the chain, where smaller facilities are in use, together with decreasing protection. Longer lead-times for storage at large facilities lead to cost and quality savings. Shorter lead-times at small facilities are favored for the same reason, and reduce inventory costs. An effective information system is necessary to guarantee the continuity of the product flow to the customer.

Location of Facilities along the Logistic Chain

The general model of a logistic chain for agrofood products is the tree of Pythagoras, showing the collecting and the distributing function on both sides of the trunk. Small facilities are located at the beginning and at the end of the logistic system, implying high flow speeds for cost and quality reasons. Large facilities mark the middle of the trunk, the center of the logistic system.

For production with a quality life longer than the minimum lead-time through all the links of the chain, a buffer station can most economically, in terms of both cost and quality, be situated at the point of maximal aggregation.

If quality life and lead-time are of the same magnitude, the buffer function tends to move towards the end of the chain. The storage function of all the preceding links is minimized for a maximization of quality at the retail outlets. The retailers, however, favor daily shopping, which generates greater sales, and reduces quality problems that result from long residence times at suboptimal conditions at the retail end of the chain. The buffer function is thus located at the wholesalers and distribution centers, which often have to take care of the delivery service as well, governed by effective order processing.

Combination of Product Flows

A powerful means of reducing logistic costs is given by the combination of goods in consecutive links of the distribution chain; Table 9 gives an example.

The possibilities for combination depend on the required conditions and lead-times, but can also be enlarged by multicompartment, multicondition facilities for storage and transportation. In this field, protective packaging for specific products has its special merits, allowing for combinations which are otherwise impossible.

Considerable reductions of transportation costs are reported through savings on labor, equipment, and energy costs. The total savings percentages are much less spectactular, but interesting enough with respect to price competition. In storage links, the advantages of combination are more indirectly realized by simplified, uniform inventory administration and order processing.

Table 9 Integration of product flows for delivery at retail level; developments in Europe (Netherlands) and USA, 1984–1990

Product group	*Conditioning*	*Average volume flow (%)*					
		Europe (NL)	*USA*[a]				
			1	*2.1*	*2.2*	*2.3*	*2.4*
General grocery	n	52 D	53 D	42 D	42 D	67 D	76 D
Meat products	ch		8 R	6 R	23 D		
Cheese/eggs	ch/n		2 R	2 R	2 R		
Fruits and vegetables	ch/n	18 R	11 R	17 R			
Frozen food	df/f	2 R	2 R	2 R	9 R	7 D	
Fresh meat	ch	5 R	5 R	7 R			
Milk products	ch	13 R	9 R	13 R	13 R	13 R	13 R
Bakery products	n	10 R	7 R	11 R	11 R	11 R	11 R
		100	100	100	100	100	100

[a]First digit (1,2) represents different stores; second digit (1–4) represents steps of development.
n, nonconditioned; ch, chilled; f, frozen; df, deep-frozen; D, from distribution center; R, from regional depot/wholesaler.

Logistical Policy

In agrofood logistic systems, storage and transportation – the main activities in all logistic systems – are, in addition, characterized by critical design considerations originating from product quality development in time, and conditioning. The two activities compete for minimum costs, by aiming for the largest possible facilities in the relevant links of the transport chain, but under the limiting influences of differentiated product requirements and restrictions imposed by high housing costs and traffic regulations in cities.

Total cost analysis provides a methodology for the integrated design of activities along the whole chain of operation. For a complete analysis in the planning phase, a wide variation of design alternatives has to be considered: alternative location of primary (production), secondary (distribution), and tertiary (retail) storage facilities, type of transportation, shipment size, appearances, equipment for processing, conditioning, and handling.

As decision criteria, maximum service, maximum profit, maximum competitive advantage or minimum investment (hired services) may be used. Regardless of the chosen strategy, the desired level of customer service should be accomplished at minimum total cost. The number of possible alternatives to reach this aim can only be tested by the use of mathematical models. The formulation of a total cost analysis is not without practical problems, especially the appreciation of quality degradation.

Limitations of throughflow time deal only with the general logistic aspect of product flow. The limitation of conditions, the most important of which is temperature, is considered the second most important aspect for perishable products, whereas only the combination of the two can yield the appropriate criteria in terms of quality maintenance in the chain.

Supervision of chain processes, therefore, should be based on information on time–temperature load on the product rather than only time-based information (shortest remaining shelf-life instead of FIFO as the management criterion). This requires a new system of control, which in its simplest form is presented by cheap physicochemical, directly readable, time–temperature integrators for each lot of product. A more suitable approach can be expected from the application of microelectronic sensors backed up by a computer-assisted inventory management.

Standards and Recommendations

ANSI, American National Standards Institute. MH 10.2–1973.

AUF, Australian United Fruit and Vegetable Organization. 1980, *Product Manual*.

EC, European Community. *Directive Deep Frozen Food for Human Consumption*, 1990, Brussels.

ECE, Economic Commission for Europe. *Agreement on the international carriage of perishable foodstuffs and on the special equipment to be used for such carriage* (ATP), Geneva, 1970, and amendments, revised 1991.

Food Marketing Institute. MUM-project. *Trade Practice Recommendations for the Fresh Fruit and Vegetable Industry*, Washington, DC (1982).

IIR/IIF, International Institute of Referigeration, *Saving of energy in refrigeration*, Paris, 1980. *Refrigeration of perishable products for distant markets*, 1982. *Cold chains in economic perspective*, 1988.

ISO, International Organization for Standardization. 3674-1984, *Packaging – Unit Load Sizes – Dimensions*. 3394-1984, *Dimensions of rigid rectangular packages – Transport packages*. 6780-1988, *General-purpose pallets for through transit of goods – Part 1: Principal dimensions and tolerances*. 1496/2-1988, *Series 1 freight containers – Specifications and testing – Part 2: Thermal containers*.

SAA, Standards Association of Australia. AS 2348-1980, *Modular Unit Loads*.

UK, *Food Hygiene Regulations*, 1990, London.

See also: **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability; Shelf-life Testing

Further Reading

Broetz W (1958) *Grundriss der Chemischen Reaktionstechnik*. Bergheim/Weinstrasse: Verlag Chemie.

Derens E, Canourges EM, Bennahmias R and Billiard F (1990) Les enregistreurs de temperature. *Revue General du Froid*, July/August: 63–68.

Jordan JL, Shewfelt RL, Prussia SE and Hurst WC (1985) Estimating the price of quality characteristics for tomatoes. *Horticultural Science* 20(2): 203–205.

Karel M (1982) Prediction of shelflife of stored food. In: *Refrigeration of Perishable Products for Distant Markets*, pp. 202–209. Paris: IIF/IIR.

Kast FE and Rosenzweig JE (1985) *Organisation and Management*. New York: McGraw-Hill.

Koelemeijer K (1991) A conceptual model of customer service evaluation. In: Bradley F (ed.) *Marketing Thought around the World*. Dublin: University College.

Labuza ThP (1982) *Shelf Life Dating of Foods*. Westport, CT: Food and Nutrition Press.

LaLonde BJ, Cooper MC and Noordewier TG (1988) *Customer Service, a Management Perspective*. Oak Brook: Council of Logistic Management.

LaLonde BJ and Zinszer PH (1976) *Customer Service: Meaning and Measurement*. Chicago, IL: National Council of Physical Distribution Management.

Levenspiel O (1972) *Chemical Reaction Engineering*, 2nd edn. New York: Wiley.
Meffert HFTh (1990a) Economic developments pertinent to chilling foods. In: *Chilled Foods, the State of the Art*. London: Elsevier Applied Science.
Meffert HFTh (1990b) Chilled foods in the market place. In: *Processing and Quality of Foods*, vol. 3. London: Elsevier Applied Science.
Meffert HFTh (1990c) Quality development of foodstuffs under time–temperature conditions in cold chains. In: *Progress in the Science and Technology of Refrigeration in Food Engineering*. Paris: International Institute of Refrigeration.
Nowotny S (1988) Computer-modelling of cold chain systems. In: *Cold Chains in Economic Perspective*. Paris: International Institute of Refrigeration.
Rich E (1988) *Artificial Intelligence*. New York: McGraw-Hill.
Saguy I and Karel M (1980) Modelling quality determination during food processing and storage. *Food Technology* 34(2): 78–85.
Schoorl D and Holt JE (1981) Fresh fruit and vegetables distribution management of quality. *Scientia Horticulturae* 17: 1–8.
Shewfelt RL, Prussia SE, Hurst WL and Jordan JL (1985) A system approach to the evaluation of changes in quality during postharvest handling of southern peas. *Journal of Food Science* 5: 169–772.
Simpson R, Lee Ken-Youn and Torres A (1989) A management tool to improve microbial quality of refrigerated foods. In: *Technical Innovations in Freezing and Refrigeration of Fruits and Vegetables*. Paris: International Institute of Refrigeration.
Spiess WEL (1988) Changes in quality of deep frozen food on the frozen food chain. In: *Cold Chains in Economic Perspective*. Paris: International Institute of Refrigeration.
Westerterp KR, van Swaail WPM and Benneckers AACM (1984) *Chemical Reactor Design and Operation*. Chichester: John Wiley.

TRICARBOXYLIC ACID CYCLE

D A Bender, University College London, London, UK

Background

The tricarboxylic acid cycle is the major energy-yielding metabolic pathway in cells, providing the greater part of the reduced coenzymes that will be oxidized by the electron transport chain to yield adenosine triphosphate (ATP). The pathway is sometimes known as the citric acid cycle, or the Krebs' cycle, after its discoverer, Sir Hans Krebs. In addition to its role in energy-yielding metabolism, and the oxidation of 2-carbon units, the cycle is also the major pathway for interconversion of 4- and 5-carbon compounds in the cell, many of which arise from, or are intermediates in the synthesis of, amino acids. Oxaloacetate, a key intermediate in the cycle, is the main precursor for gluconeogenesis in the fasting state.

Reactions of the Tricarboxylic Acid Cycle

As shown in **Figure 1**, the tricarboxylic acid cycle provides a pathway for the oxidation to carbon dioxide and water of the acetate moiety of acetyl coenzyme A (CoA) arising from the oxidative decarboxylation of pyruvate (the end product of glycolysis), the β-oxidation of fatty acids, the metabolism of ethanol, ketone bodies, and a number of amino acids. For each mole of acetyl CoA oxidized in this pathway, there is a yield of 12 × ATP equivalents, arising from:

- 3 × nicotinamide adenine dinucleotide (NAD^+) reduced to NADH, equivalent to 9 × ATP when reoxidized in the electron transport chain;
- 1 × flavoprotein reduced, leading to reduction of ubiquinone, equivalent to 2 × adenosine diphosphate (ADP) when reoxidized in the electron transport chain;
- 1 × guanosine diphosphate (GDP) phosphorylated to guanosine triphosphate (GTP), equivalent to 1 × ATP.

The four-carbon intermediate, oxaloacetate, reacts with acetyl CoA to form a six-carbon compound, citric acid. The cycle is then a series of steps in which two carbon atoms are lost as carbon dioxide, followed by a series of oxidation and other reactions, eventually reforming oxaloacetate. The CoA of acetyl CoA is released, and is available for further formation of acetyl CoA from pyruvate, fatty acids, or ketone bodies.

Although oxaloacetate is the precursor for gluconeogenesis, fatty acids and other compounds that give rise to acetyl CoA or acetoacetate cannot be used for net synthesis of glucose. As can be seen from **Figure 1**,

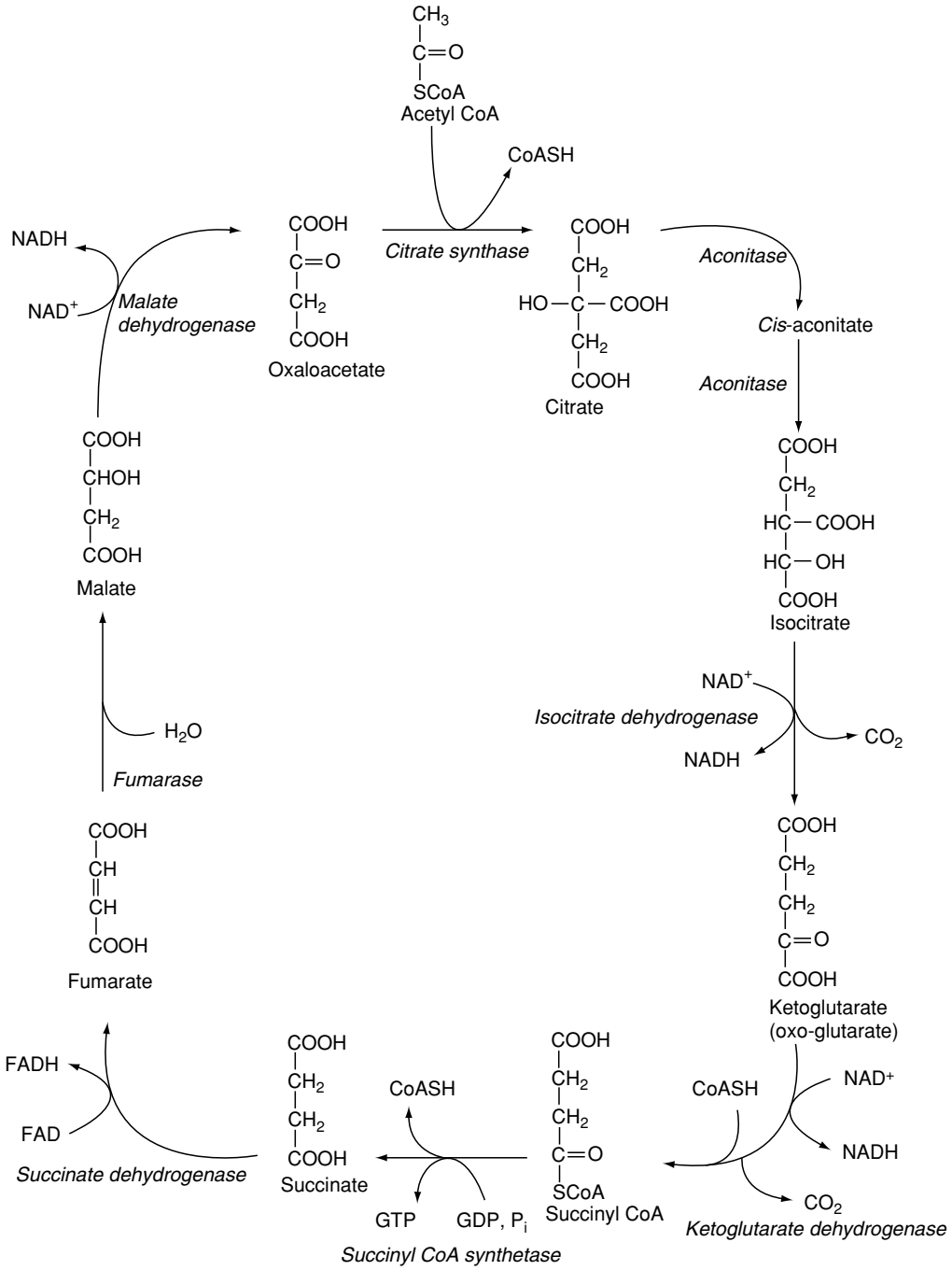

Figure 1 Tricarboxylic acid cycle. Citrate synthase EC 4.1.3.7, aconitase EC 4.2.1.3, isocitrate dehydrogenase EC 1.1.1.41, ketoglutarate dehydrogenase EC 1.2.4.2, succinyl CoA synthase EC 6.2.1.4, succinate dehydrogenase EC 1.3.5.1, fumarase EC 4.2.1.2, malate dehydrogenase EC 1.1.1.37.

although two carbons are added to the cycle by acetyl CoA, two carbons are lost as carbon dioxide in each turn of the cycle. Therefore, when acetyl CoA is the substrate, there is no increase in the pool of tricarboxylic acid cycle intermediates, and therefore oxaloacetate cannot be withdrawn for gluconeogenesis when acetyl CoA is being metabolized.

Although citrate is a symmetrical molecule (carbons 1 and 2 are equivalent to carbons 5 and 6), it behaves asymmetrically in the cycle. When tissue is incubated with [^{14}C]acetyl CoA in the presence of malonate, an inhibitor of succinate dehydrogenase, there is no release of $^{14}CO_2$, despite the fact that 2 mol of CO_2 are released in the sequence of reactions between citrate

synthetase and succinate dehydrogenase. The two carbon atoms that are lost in the first turn of the cycle are not the same two atoms added from acetyl CoA. This is because citrate does not diffuse from the active site of citrate synthetase into free solution, but is channeled directly from the active site of citrate synthetase to that of aconitase.

The main regulation of the rate of tricarboxylic acid cycle activity is by control of the activity of isocitrate dehydrogenase, which is regulated by both phosphorylation/dephosphorylation, and also by the intramitochondrial ratio of NADH:NAD^+, it is also inhibited by NADH, and activated by ADP. When the rate of entry of acetyl CoA into the cycle, and hence the rate of formation of citrate, is greater than the rate at which isocitrate undergoes dehydrogenation, aconitase is inhibited by its product. The result of this is that citrate can no longer be channeled from the active site of citrate synthetase to aconitase, and now leaves the active site of citrate synthetase into free solution in the mitochondrial matrix. In most tissues, free citrate can be transported out of the mitochondria into the cytosol. Here, it can act both to inhibit phosphofructokinase, so reducing the rate of glycolysis and formation of acetyl CoA, and the cytosolic as the source of acetyl CoA for fatty acid synthesis (see below).

α-Ketoglutarate dehydrogenase catalyzes a reaction similar to that of pyruvate dehydrogenase – oxidative decarboxylation and formation of an acyl CoA derivative. Like pyruvate dehydrogenase, it is a thiamin diphosphate-dependent enzyme. However, thiamin deficiency does not have a significant effect on the tricarboxylic acid cycle, because α-ketoglutarate can undergo transamination to yield glutamate, which is decarboxylated to γ-aminobutyric acid (GABA). In turn, GABA can undergo further metabolism to yield succinate. This pathway, shown in Figure 2, is sometimes called the GABA shunt; it provides an alternative to α-ketoglutarate dehydrogenase in thiamin deficiency, so that oxidation of acetyl CoA and formation of ATP can continue.

The reaction of succinyl CoA synthetase, the formation of free succinate and CoA from succinyl CoA, is linked to a substrate-level phosphorylation of GDP to GTP, which is energetically equivalent to the formation of ATP.

In tissues that synthesize glucose in the fasting state (liver, kidney, and intestinal mucosa), this reaction also provides an important link between the rate of tricarboxylic acid cycle activity and the withdrawal of oxaloacetate for gluconeogenesis. The formation of phospho-enolpyruvate from oxaloacetate, catalyzed by phospho-enolpyruvate carboxykinase (see Figure 4) requires GTP as the phosphate donor; the

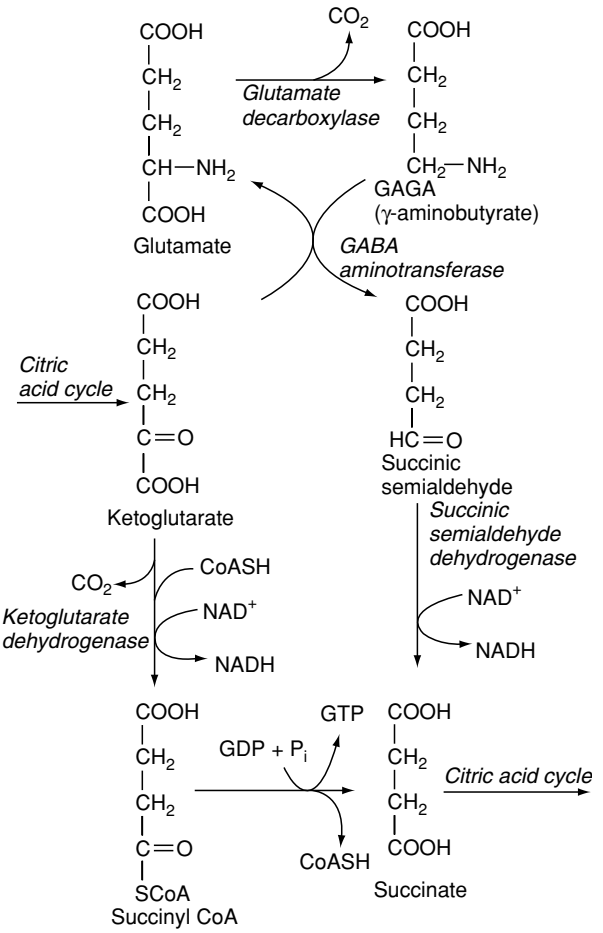

Figure 2 GABA shunt. Glutamate decarboxylase EC 4.1.1.15, GABA aminotransferase EC 2.6.1.19, succinic semialdehyde dehydrogenase EC 1.2.1.24, ketoglutarate dehydrogenase EC 1.2.4.2.

major, if not only, intramitochondrial source of GTP is the reaction of succinyl CoA synthetase. This means that it is not possible to withdraw oxaloacetate from the cycle for gluconeogenesis if this would result in a slowing of the rate of cycle activity, and hence the rate of formation of GTP.

Under conditions where ketone bodies are being metabolized (especially in muscle and brain in the fasting state), there is an alternative reaction that bypasses succinyl CoA synthetase – direct transfer of CoA from succinyl CoA on to acetoacetate to form acetoacetyl CoA. This provides a link between the rate of tricarboxylic acid cycle activity and the ability to take up and utilize acetoacetate; if there is little succinyl CoA available, little acetoacetate can be utilized.

The sequence of reactions between succinate and oxaloacetate is chemically the same as that involved in the β-oxidation of fatty acids:

- Oxidation to yield a carbon–carbon double bond. Although this reaction is shown in Figure 1 as being linked to reduction of FAD, the coenzyme is tightly enzyme bound, and succinate dehydrogenase reacts directly with ubiquinone in the electron transport chain.
- Addition of water across the carbon–carbon double bond, to yield a hydroxyl group.
- Oxidation of the hydroxyl group, linked to reduction of NAD^+, to yield an oxo-group, so reforming oxaloacetate to undergo reaction with a further molecule of acetyl CoA.

Citrate as the Source of Acetyl CoA for Lipogenesis

The substrate for fatty acid synthesis is acetyl CoA. However, the only source of acetyl CoA is intramitochondrial, whereas fatty acid synthesis is a cytosolic process. Under conditions where fatty acids are

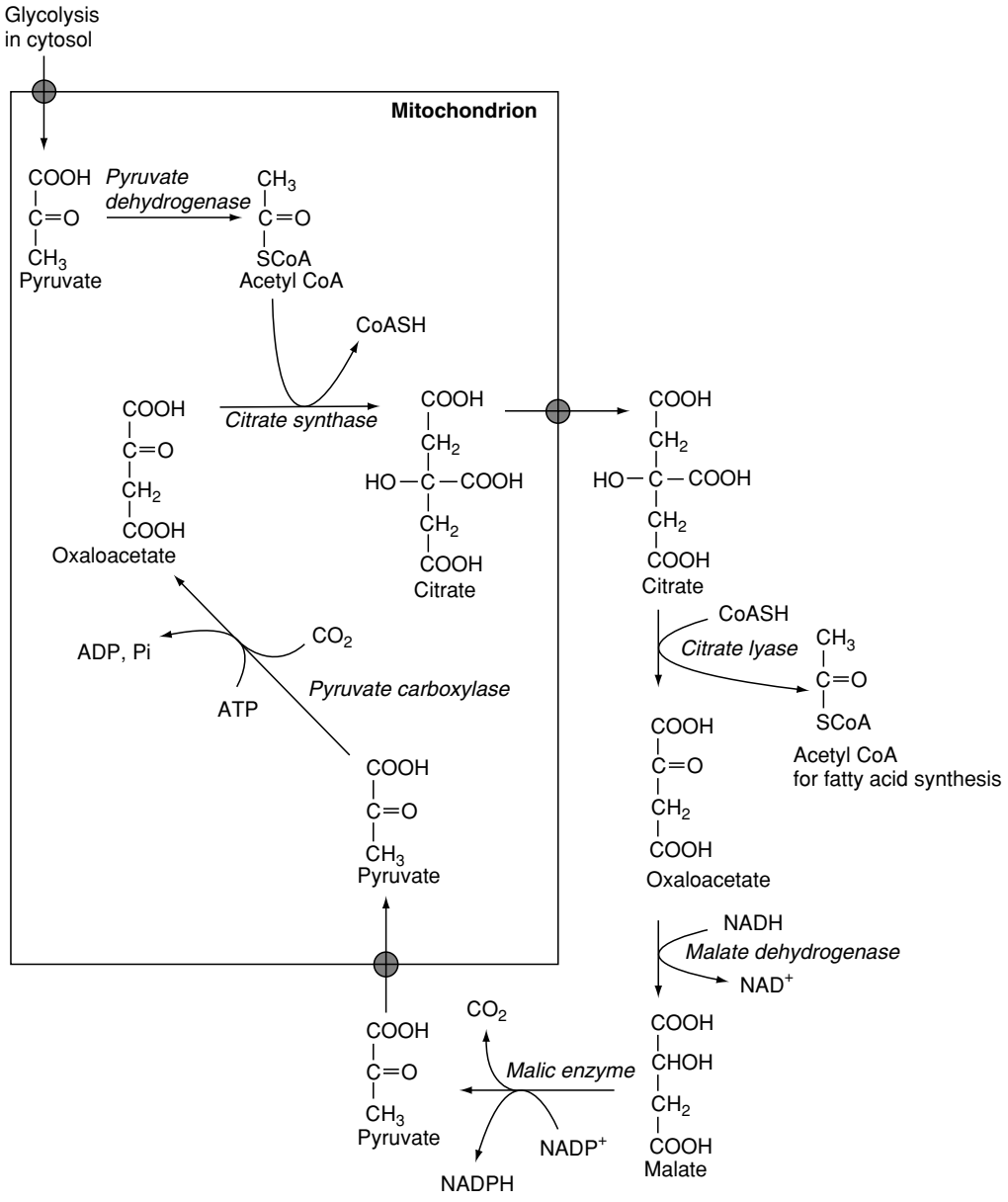

Figure 3 Citrate as the source of acetyl CoA for fatty acid synthesis. Pyruvate dehydrogenase EC 1.2.4.1, citrate synthase EC 4.1.3.7, citrate lyase EC 4.1.3.6, malate dehydrogenase EC 1.1.1.37, malic enzyme EC 1.1.1.38, pyruvate carboxylase EC 6.4.1.1, phosphoenolpyruvate carboxykinase EC 4.1.1.32.

to be synthesized, and the rate of entry of acetyl CoA into the cycle is greater than that required for maintenance of ATP formation, free citrate is transported from the mitochondrial matrix into the cytosol. Here, it undergoes cleavage to acetyl CoA and oxaloacetate.

Oxaloacetate cannot reenter the mitochondrion, but undergoes the sequence of reactions shown in Figure 3: reduction to malate, followed by oxidative decarboxylation to pyruvate, which enters the mitochondrion, and is carboxylated to oxaloacetate. The oxidative decarboxylation of malate to pyruvate, catalyzed by the malic enzyme, is linked to the reduction of $NADP^+$ to NADPH. This reaction provides half the NADPH required for fatty acid synthesis; the remainder comes from the pentose phosphate pathway of glucose metabolism.

Anaplerotic Reactions – Replenishing the Supply of Tricarboxylic Acid Cycle Intermediates

If oxaloacetate is removed from the cycle for glucose synthesis, it must be replaced, since if there is not enough oxaloacetate available to form citrate, the rate of acetyl CoA metabolism, and hence the rate of formation of ATP, will slow down. Similarly, under conditions of hyperammonemia, the reaction of glutamate dehydrogenase leads to withdrawal of a considerable amount of α-ketoglutarate, to the extent that ATP formation is impaired, leading to coma and convulsions.

As shown in Figure 4, a variety of amino acids give rise to tricarboxylic cycle intermediates, so permitting removal of oxaloacetate for gluconeogenesis. In addition, the reaction of pyruvate carboxylase is a major source of oxaloacetate to maintain tricarboxylic acid cycle activity.

The metabolic fate of pyruvate, oxidative decarboxylation to yield acetyl CoA or carboxylation to yield oxaloacetate, is determined by the relative availability of acetyl CoA (which also arises from β-oxidation of fatty acids and metabolism of ketone bodies) and the need for oxaloacetate to maintain tricarboxylic acid cycle activity. Pyruvate dehydrogenase is inhibited by both NADH and acetyl CoA, whereas pyruvate carboxylase has an absolute requirement for acetyl CoA for activity. Thus, at times when there is an abundance of acetyl CoA, pyruvate will not undergo decarboxylation and oxidation in the tricarboxylic acid cycle, but rather will be carboxylated to oxaloacetate. Once the pool of tricarboxylic acid cycle intermediates is adequate, further oxaloacetate formed by carboxylation of pyruvate can be used for gluconeogenesis.

Metabolism of Amino Acid Carbon Skeletons – Gluconeogenesis From Amino Acids

Acetyl CoA and acetoacetate arising from the carbon skeletons of amino acids may be used for fatty acid synthesis or can be oxidized as metabolic fuel, but cannot be utilized for the synthesis of glucose (gluconeogenesis, see glucose metabolism). Amino acids that yield acetyl CoA or acetoacetate are termed ketogenic. Only two amino acids, leucine and lysine, are solely ketogenic; isoleucine, phenylalanine, and tryptophan yield both ketogenic and glucogenic fragments. All the other amino acids yield carbon skeletons that can be used for gluconeogenesis.

The principal substrate for gluconeogenesis is oxaloacetate, which undergoes the reaction catalyzed by phospho-enolpyruvate carboxykinase to yield phospho-enolpyruvate, as shown in Figure 4. The points of entry of amino acid carbon skeletons into central metabolic pathways are shown in Figure 4. Those that give rise to ketoglutarate, succinyl CoA, fumarate, or oxaloacetate can be regarded as directly increasing the tissue pool of tricarboxylic acid cycle intermediates, and hence permitting the withdrawal of oxaloacetate for gluconeogenesis.

Those amino acids that give rise to pyruvate also increase the tissue pool of oxaloacetate, since pyruvate is carboxylated to oxaloacetate in the reaction catalyzed by pyruvate carboxylase.

Gluconeogenesis is an important fate of amino acid carbon skeletons in the fasting state, when the metabolic imperative is to maintain a supply of glucose for the central nervous system and red blood cells. However, in the fed state, the carbon skeletons of amino acids in excess of requirements for protein synthesis will be used mainly for formation of acetyl CoA for fatty acid synthesis, and storage as adipose tissue triacylglycerol.

Complete Oxidation of 4- and 5-carbon Compounds

Although the tricarboxylic acid cycle is generally regarded as a pathway for the oxidation of 4- and 5-carbon compounds such as fumarate, oxaloacetate, α-ketoglutarate and succinate arising from amino acids, it does not, alone, permit complete oxidation of these compounds. Four-carbon intermediates are not overall consumed in the cycle, since oxaloacetate is reformed. Addition of 4- and 5-carbon intermediates may increase the rate of cycle activity (subject to control by the requirement for ATP), but once the pool of intermediates is saturated, no more can enter.

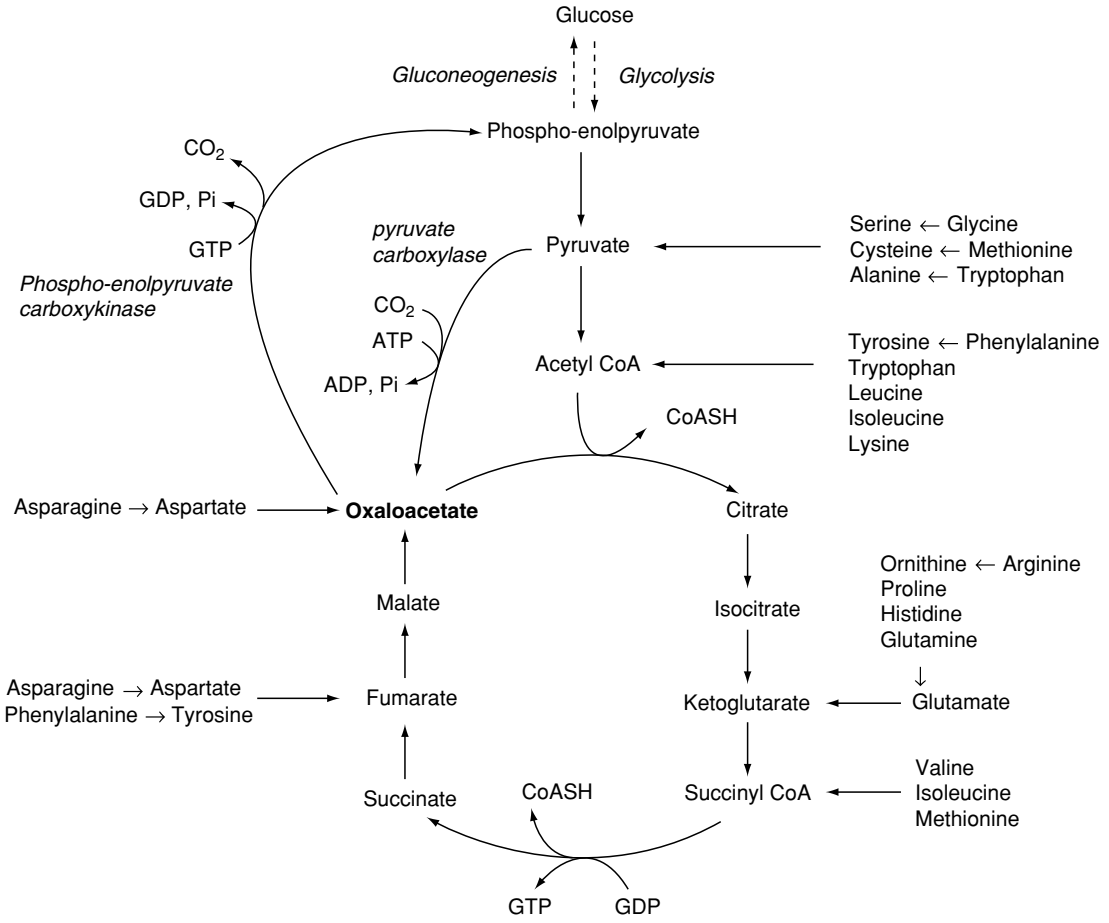

Figure 4 Entry of amino acid carbon skeletons into the tricarboxylic acid cycle.

Complete oxidation of 4- and 5-carbon intermediates requires removal of oxaloacetate from the cycle, and conversion to pyruvate, as shown in Figure 4. This pyruvate may undergo decarboxylation to acetyl CoA, which may be oxidized in the cycle, or can be used for fatty acid synthesis or as a substrate for gluconeogenesis.

See also: **Amino Acids**: Metabolism; **Energy Metabolism**; **Fatty Acids**: Metabolism; **Glucose**: Function and Metabolism; **Ketone Bodies**; **Thiamin**: Physiology; **Oxidative Phosphorylation**

Further Reading

Bender DA (2002) *Introduction to Nutrition and Metabolism*, 3rd edn. London: Taylor & Francis.

Bender DA and Bender AE (1997) *Nutrition: A Reference Handbook*. Oxford: Oxford University Press.

Murray RK, Granner DK, Mayes PA and Rodwell VW (2000) *Harper's Biochemistry*, 25th edn. New York: McGraw-Hill.

TRIGLYCERIDES

Contents
Structures and Properties
Characterization and Determination

Structures and Properties

P Kalo and A Kemppinen, University of Helsinki, Finland

Isomerism and Structure

Triglycerides are the most abundant lipid class of fats and oils, in which glycerol is esterified with three fatty acids. The glycerol molecule is prochiral. If the primary hydroxyls of glycerol are esterified with different fatty acids, the molecule is chiral and thus may have two enantiomers. Instead of the *R*/*S* convention or D/L convention, the triacylglycerol (and other glycerolipids) enantiomers are designated using 'stereospecific numbering' (*sn*) according to the recommendations of the International Union of Pure and Applied Chemistry – International Union of Biochemistry (1967) commission on the nomenclature of glycerolipids. When a Fischer projection formula is drawn for a glycerol derivative so that oxygen is shown to the left of secondary carbon 2, which is designated *sn*-2 (Figure 1). The carbon above it is *sn*-1, and the carbon below it is *sn*-3. Owing to regio-(positional) and stereoisomerism, a great number of different triacylglycerols exist. Three different fatty acids can form 10 different molecular species, i.e., triacylglycerols having different fatty acid compositions: three monoacid triacylglycerols (AAA, BBB, CCC), six diacid triacylglycerols, each composed of a pair of enantiomers and a symmetric isomer (AAB/BAA; ABA, ABB/BBA; BAB, AAC/CAA; ACA, ACC/CCA; CAC, BBC/CBB; BCB, BCC/CCB, CBC), and three pairs of enantiomers of three-acid triacylglycerols (ABC/CBA, ACB/BCA, BAC/CAB), altogether $n^3 = 27$ different triacylglycerol molecules. In natural fats such as milk fat, in which more than 400 different fatty acids have been identified, a large number of different triacylglycerol molecules may theoretically exist. In addition to positional and stereoisomers, natural fats include acyl chain isomers, which have the same molecular formula, but are composed of different fatty acids.

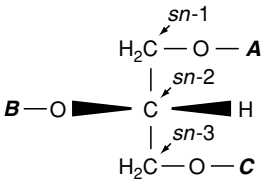

Figure 1 Fischer projection formula of a triacyl-*sn*-glycerol. ***A***, ***B***, and ***C*** are acyl groups.

Although, in recent years, positional isomers of some molecular species have been chromatographically determined, the information on triacylglycerol structure is based on studies, in which the distribution of fatty acids between the *sn*-1, *sn*-2, and *sn*-3 positions (stereospecific analysis) or between the *sn*-1(3) and *sn*-2 positions (regiospecific analysis) is determined in an isolated triacylglycerol fraction of fat or oil. A variety of methods have been developed. In these, pancreatic lipase (Methods A and B shown in Figure 2) and Grignard degradation (Methods A–E) has been used to produce partial acylglycerols. The former hydrolyzes fatty acids from the *sn*-1 and *sn*-3 position producing free fatty acids and 2-monoacyl-*sn*-glycerols and 1,2- and 2,3-diacyl-*sn*-glycerols. Owing to the fatty acid specificity of lipase, representative diacylglycerols are not formed: long-chain polyunsaturated fatty acids are hydrolyzed more slowly, and small triacylglycerol molecules containing short-chain fatty acids are hydrolyzed more rapidly than large triacylglycerol molecules. Grignard reagent degrades primary and secondary ester bonds equally and has no fatty acid specificity. The products are 1,2-, 2,3-, and 1,3-diacyl-*sn*-glycerols, 1-, 2-, and 3-monoacyl-*sn*-glycerols, and tertiary alcohols. Special attention has to be paid to the spontaneous acyl migration in both hydrolysis and Grignard degradation products. Methods A and B (Figure 2) are based on the specificity of phospholipase A_2 (EC 3.1.1.4) to hydrolyze the *sn*-2 ester bond of *sn*-1,2-diacylglycerol phosphatidylcholine and selectivity of diacylglycerol kinase (EC 2.7.1.107) to phosphorylate 1,2-diacyl-*sn*-glycerols, respectively. Methods C and E (Figure 2) are based in the chiral high-performanc liquid chromatography (HPLC) separation of

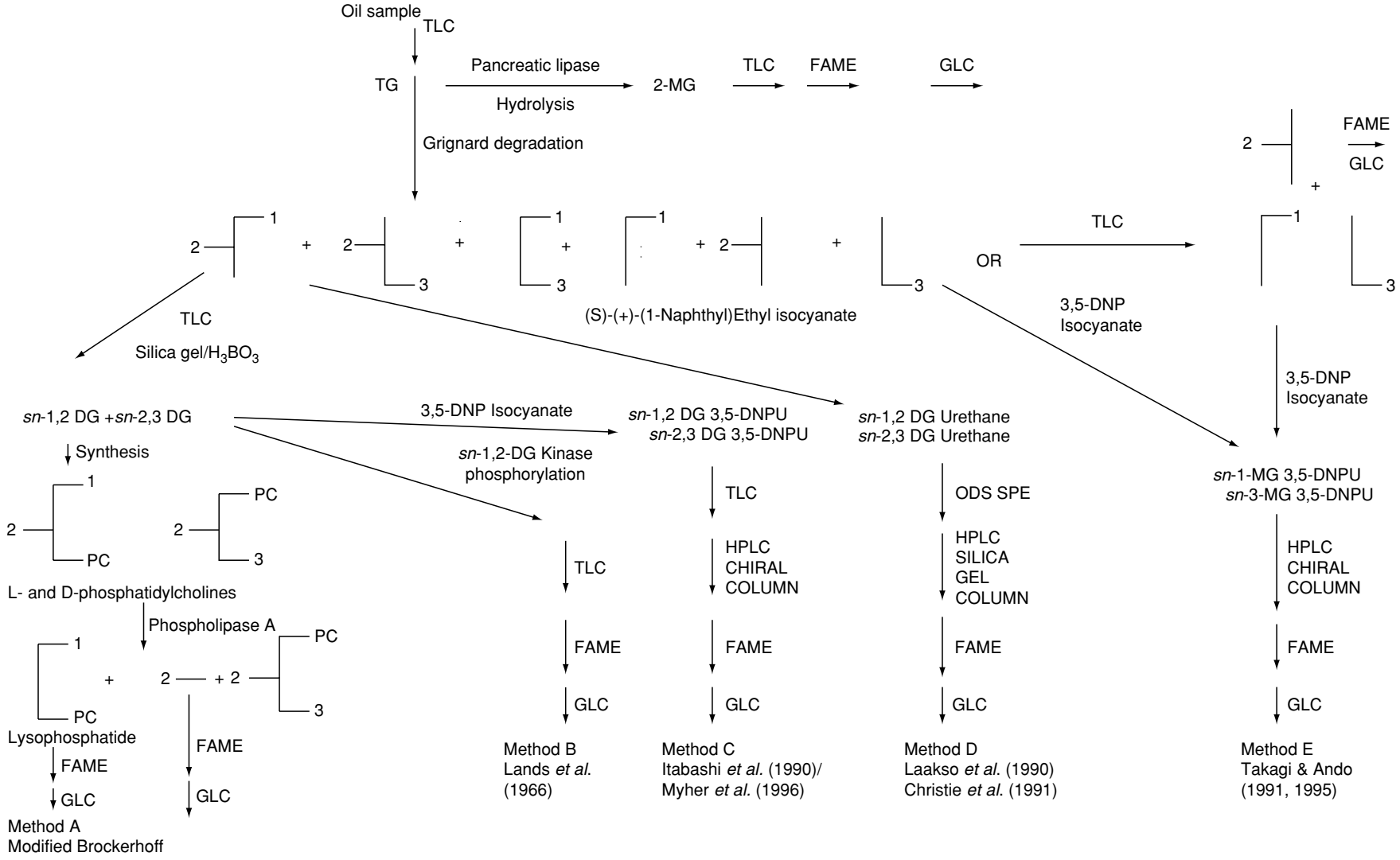

Figure 2 Basic features of the methods used in the determination of the fatty acid distribution between the *sn*-1, *sn*-2, and *sn*-3 positions in triacyl-*sn*-glycerols. TLC, thin-layer chromatography; GLC, gas–liquid chromatography; HPLC, high-performance liquid chromatography; TG, triacylglycerol; 2-MG, 2-monoglyceride; FAME, fatty acid methyl ester.

3,5-dinitrophenyl urethanes of 1,2- and 2,3-diacyl-*sn*-glycerols and 1- and 3-monoacyl-*sn*-glycerols, respectively. The method D (Figure 2) is based on the separation of diastereomeric derivatives of 1,2- and 2,3-diacyl-*sn*-glycerols on a silica-gel HPLC column.

Triacylglycerol Structure of Food Fats and Oils

Fruit Pulp Oils

In olive oil, the saturated fatty acids are esterified mainly in the *sn*-1 and *sn*-3 positions (Table 1). A little more than one-third of oleic acid is esterified in the *sn*-2 position and a little less in the *sn*-1 and *sn*-3 positions. The ratio of both primary positions is close to 1. Nearly half of linoleic acid is esterified in the *sn*-2 position and approximately one-third in the *sn*-1 position. Table 1 shows a high degree of similarity between the distributions determined by two enzymatic methods. A study by Damiani *et al*. did not reveal any statistical difference between the methods B and D (Figure 2) by Student's *t*-test. Also, in palm oil, saturated fatty acids are preferentially esterified in the *sn*-1 or *sn*-3 positions. Unsaturated fatty acids are esterified to a great extent in the *sn*-2 position.

High Lauric and Myristic Acid Seed Oils

Regiospecific analysis has shown that in coconut oil, lauric and linoleic acids are esterified mainly in the *sn*-2 position and oleic acid and saturated acids in the *sn*-1(3) positions (Table 2). A similar trend is in the distribution of lauric and oleic acids in palm kernel oil. Also, linoleic acid shows a slight preference

Table 1 Distribution of fatty acids between the *sn*-1, *sn*-2, and *sn*-3 positions in vegetable/fruit oil triacylglycerols

Fatty acid	*Total (mol%)*	*Percentage*			*Total (mol%)*	*Percentage*		
		sn-*1*	sn-*2*	sn-*3*		sn-*1*	sn-*2*	sn-*3*
Olive oil					*Palm oil*[d]			
14:0					1.3	10.3	15.4	74.4
16:0	10.5	41.7	4.5	53.8	48.4	41.4	9.2	49.5
16:1	0.8	37.5	29.2	33.3	0.2	0.0	83.3	16.7
18:0	2.3	38.2	0.0	61.8	3.7	30.4	1.8	67.9
18:1	76.2	31.4	36.3	32.3	36.4	24.6	62.2	13.2
18:2	9.6	33.9	48.4	17.6	10.0	31.0	58.3	10.7
18:3	0.9	22.2	29.6	48.1				
Olive oil[b]					*Olive oil*[c]			
16:0	12.0	46.4	1.4	52.2	12.0	41.1	3.9	55.0
16:1	0.7	33.3	23.8	42.9	0.7	52.4	4.8	42.9
18:0	2.0	40.0	0.0	60.0	2.0	30.0	10.0	60.0
18:1	78.9	30.9	37.9	31.2	78.9	31.9	37.4	30.7
18:2	6.3	39.2	49.7	11.1	6.3	34.4	51.3	14.3

[a]Enzymatic stereospecific analysis. Adapted from Brockerhoff and Yurkowski (1966).
[b]Enzymatic stereospecific analysis using *sn*-1,2-diacylglycerol kinase. Adapted from Damiani *et al*. (1994).
[c]Method based on separation of diastereomeric derivatives of diacylglycerols by silica gel HPLC. Adapted from Damiani *et al*. (1994).
[d]Method based on separation of diastereomeric derivatives of diacylglycerols by silica-gel HPLC. Adapted from Christie *et al*. (1991).

Table 2 Distribution of fatty acids between the *sn*-1(3) and *sn*-2 positions in high lauric and myristic acid seed oil triacylglycerols

Fatty acid	*Total (mol%)*	*Percentage*		*Total (mol%)*	*Percentage*	
		sn-*1(3)*	sn-*2*		sn-*1(3)*	sn-*2*
Coconut oil[a]				*Palm kernel oil*[a]		
6:0	0.6	50.0	0.0	0.4	50.0	0.0
8:0	9.4	49.5	1.1	5.7	42.1	15.8
10:0	7.0	42.9	14.3	4.7	41.1	17.7
12:0	50.0	23.2	53.6	52.8	30.2	39.6
14:0	17.4	41.8	16.5	16.2	33.7	32.5
16:0	7.6	46.5	7.0	7.6	43.6	12.7
18:0	2.0	45.0	10.0	2.2	47.7	4.5
18:1	4.7	37.6	24.8	9.2	28.6	42.8
18:2	1.2	29.2	41.7	1.2	31.9	36.1

[a]Regiospecific analysis using pancreas lipase. Adapted from Mattson and Volpenheim (1963).

for the *sn*-2 position. Myristic acid is distributed equally between the *sn*-2 and *sn*-1(3) positions, whereas the rest of saturated fatty acids are esterified principally in the primary positions.

High Palmitic Acid Seed Oils

In cocoa butter, palmitic and stearic acids are esterified to a great extent in the *sn*-1 and *sn*-3 positions, distributing nearly equally between them (Table 3). Arachidic acid is esterified totally in the *sn*-1 and *sn*-3 positions, the *sn*-3 position being more favored. Oleic and linoleic acids are esterified mainly in the *sn*-2 position, and palmito-oleic acid in all three positions. The distributions determined by enzymatic stereospecific analysis and by chiral phase HPLC separation of enantiomers of monoacylglycerol derivatives (method E, Figure 2) were very similar. In maize oil, palmitic and stearic acids are principally in the *sn*-1 and *sn*-3 positions (Table 3). Oleic acid is distributed quite evenly between the three positions with a slight excess in the *sn*-3 position. Linoleic acid is esterified predominantly in the *sn*-2 position.

High Oleic and Linoleic Acid Seed Oils

In soy bean oil, high-erucic acid rape seed oil, sunflower oil, and peanut oil, palmitic, and stearic acids are esterified preferentially in the primary positions (Table 4). Also arachidic, behenic, and lingnoceric acids are esterified mainly in the primary positions, the *sn*-3 position clearly being preferred in peanut oil. Oleic acid (18:1*n*-9) is relatively evenly distributed between the three positions in soy bean oil, sunflower oil, and peanut oil, but is predominantly at the *sn*-2 position in high-erucic rape seed oil. Vaccenic acid (18:1*n*-7) is esterified preferentially in the primary positions in soy bean oil and high-erucic rape seed oil. Eicosenoic and docosenoic acids are in the primary positions in high-erucic rape seed oil and peanut oil. In sunflower oil, linoleic acid is esterified in all positions, with a slight excess in the *sn*-2 position. In soy bean oil, high-erucic rape seed oil, and peanut oil, linoleic acid is preferentially in the *sn*-2 position. Linolenic acid is esterified preferentially in the *sn*-2 position in high-erucic rape seed oil and relatively evenly in all positions in soy bean oil. The distributions of soy bean oil fatty acids, determined by enzymatic stereospecific analysis and by HPLC separation of monoacylglycerol derivatives on a chiral column (method E, Figure 2), show a high degree of accordance. Harp and Hammond applied the method of Takagi and Ando in the stereospecific analysis of soy bean mutants representing a wide distribution of fatty acid compositions. In low-erucic-acid rape seed oil, the *sn*-2 position has reported to be occupied by unsaturated C18 fatty acids in the order of linoleic > linolenic > oleic acids.

Milk Fat

Even-numbered triacylglycerols with 26–54 acyl carbons dominate the triacylglycerol composition of bovine milk fat, but odd-numbered triacylglycerols are also present. Only very few triacylglycerols with more than 54 or fewer than 26 acyl carbons exist, indicating a scarcity of the molecular species with very-long-chain fatty acids (> 20) or those with two short-chain fatty acids, respectively. A high number of mono-butyryl triacylglycerols is unique among food fats.

Table 3 Distribution of fatty acids between the *sn*-1, *sn*-2, and *sn*-3 positions in high palmitic acid seed oil triacylglycerols

Fatty acid	*Total (mol%)*	*Percentage*			*Total (mol%)*	*Percentage*		
		sn-*1*	sn-*2*	*sn-3*		sn-*1*	sn-*2*	sn-*3*
Cocoa butter[a]					*Cocoa butter*[b]			
16:0	24.1	47.1	2.4	50.6	26.5	46.4	3.1	50.5
16:1	0.4	54.5	18.2	27.3	0.2	28.6	28.6	42.9
18:0	35.1	47.9	2.0	50.1	35.7	49.3	1.8	48.9
18:1	36.1	11.4	80.7	7.9	33.5	9.1	87.1	3.9
18:2	3.4	12.6	83.5	3.9	2.9	8.0	85.1	6.9
18:3					0.1	0.0	100.0	0.0
20:0	1.1	30.3	0.0	69.7	1.0	12.9	0.0	87.1
Maize oil[a]								
16:0	11.2	53.1	6.8	40.1				
16:1	0.2	60.0	20.0	20.0				
18:0	2.1	51.6	3.2	45.2				
18:1	28.2	32.5	31.3	36.2				
18:2	57.2	29.0	40.9	30.1				
18:3	1.0	41.4	24.1	34.5				

[a]Enzymatic stereospecific analysis. Adapted from Brockerhoff and Yurkowski (1966).
[b]Method based on separation of enantiomers of monoacylglycerol derivatives by chiral-phase HPLC. Adapted from Takagi and Ando (1995).

The asymmetric distribution of fatty acids in bovine milk fat between the three *sn*-positions has already been demonstrated in the late 1960s by Pitas *et al*. and by Breckenridge and Kuksis. Short-chain fatty acids (4:0, 6:0) in mono-butyryl and mono-caproyl triacylglycerols are almost exclusively in the *sn*-3 position (Table 5). However, a small amount of triacylglycerols with two short-chain acyl groups has been detected, and thus traces of butyric and caproic acids have to be esterified in the *sn*-1 and/or *sn*-2 positions. Medium-chain fatty acids (8:0, 10:0) are also most prominently in the *sn*-3 position. More than 50% of lauric and myristic acids is esterified in the secondary position in triacylglycerols. Fatty acids

Table 4 Distribution of fatty acids between the *sn*-1, *sn*-2, and *sn*-3 positions in high oleic and linoleic acid seed oil triacylglycerols

Fatty acid	*Total (mol%)*	*Percentage*			*Fatty acid*	*Total (mol%)*	*Percentage*		
		sn-*1*	sn-*2*	sn-*3*			sn-*1*	sn-*2*	sn-*3*
Soy bean oil[a]					*Soy bean oil*[b]				
14:0	0.1	75.0	75.0	−50.0	14:0	0.1	33.3	33.3	33.3
16:0	11.6	48.4	4.0	47.6	16:0	11.7	47.7	6.3	46.0
16:1	0.1	50.0	50.0	0.0	16:1	0.1	33.3	33.3	33.3
17:0	0.0	100.0	0.0	0.0	17:0	0.1	50.0	0.0	50.0
18:0	3.4	51.0	2.9	46.1	18:0	3.4	52.4	2.9	44.7
18:1*n*-9	24.1	29.9	33.2	36.9	18:1*n*-9	22.7	32.6	33.8	33.6
18:1*n*-7	1.4	47.6	9.5	42.9	18:1*n*-7	1.4	50.0	9.5	40.5
18:2	53.9	29.3	42.8	27.9	18:2	53.8	28.7	42.2	29.1
18:3	6.4	39.3	29.8	30.9	18:3	6.4	33.5	29.8	36.6
20:0	0.3	22.2	0.0	77.8	20:0	0.3	55.6	0.0	44.4
Rape seed oil (high erucic acid)[c]					*Rape seed oil (high erucic acid)*[b]				
					14:0	0.1	25.0	50.0	25.0
16:0	3.0	45.6	6.7	47.8	16:0	3.9	47.9	13.7	38.5
16:1	0.3	37.5	25.0	37.5	16:1*n*-9	0.1	0.0	14.3	85.7
					16:1*n*-7	0.3	37.5	37.5	25.0
					17:1*n*-8	0.1	0.0	100.0	0.0
18:0	1.7	42.3	0.0	57.7	18:0	1.3	47.5	10.0	42.5
18:1	25.7	30.0	48.4	21.6	18:1*n*-9	17.2	26.8	63.3	10.0
					18:1*n*-7	1.4	65.1	23.3	11.6
18:2	17.1	21.7	70.5	7.8	18:2*n*-6	16.1	2.5	82.0	15.4
18:3	9.8	21.8	69.3	8.9	18:3*n*-3	8.7	12.7	82.2	5.0
20:0	0.0				20:0	0.8	47.1	35.3	17.6
20:1	11.9	45.9	5.6	48.5	20:1*n*-9	8.9	57.4	5.6	37.0
					20:1*n*-7	1.5	44.2	9.3	46.5
					20:2*n*-6	0.5	80.0	6.7	13.3
22:0	0.9	53.8	0.0	46.2	22:0	0.7	0.0	0.0	100.0
22:1	29.8	39.0	4.0	57.0	22:1*n*-9	37.4	44.7	0.7	54.6
					22:1*n*-7	0.6	47.4	0.0	52.6
					22:2*n*-6	0.5	56.3	0.0	43.8
Sunflower oil[d]									
16:0	7.2	49.1	6.0	44.9					
18:0	4.6	24.3	8.1	67.6					
18:1	21.9	25.3	32.7	42.0					
18:2	66.3	34.9	38.2	26.9					
Peanut oil[c]					*Peanut oil*[e]				
16:0	8.7	51.9	6.1	42.0	16:0	11.6	57.0	6.3	36.7
16:1	0.2	42.9	14.3	42.9	16:1	0.1	18.2	21.2	60.6
18:0	3.3	46.0	3.0	51.0	18:0	2.8	42.7	12.2	45.1
18:1	58.3	33.8	33.4	32.7	18:1	50.8	31.2	33.4	35.5
18:2	22.4	27.6	57.5	14.9	18:2	27.8	30.5	56.3	13.2
20:0	1.6	14.9	0.0	85.1	20:0	1.4	7.0	0.0	93.0
20:1	1.4	26.8	7.3	65.9	20:1	1.4	19.4	0.0	80.6
22:0	2.4	17.8	4.1	78.1	22:0	2.6	1.3	1.3	97.3
24:0	1.3	17.5	12.5	70.0	24:0	1.1	0.0	0.0	100.0

[a]Enzymatic stereospecific analysis using modified Brockerhoff method. Adapted from Takagi and Ando (1991).
[b]Method based on separation of enantiomers of monoacylglycerol derivatives on a chiral HPLC column. Adapted from Takagi and Ando (1991).
[c]Enzymatic stereospecific analysis. Adapted from Brockerhoff and Yurkowski (1966).
[d]Method based on separation of diastereomers of diacylglycerol derivatives on a silica-gel HPLC column. Adapted from Christie *et al*. (1991).
[e]Enzymatic stereospecific analysis using Brockerhoff method. Adapted from Myher *et al*. (1978).

Table 5 Distribution of fatty acids between the *sn*-1, *sn*-2, and *sn*-3 positions in mammals' milk triacylglycerols

Fatty acid	*Total (mol%)*	*Percentage*			*Total (mol%)*	*Percentage*		
		sn-*1*	sn-*2*	sn-*3*		sn-*1*	sn-*2*	sn-*3*
Bovine milk fat[a]					*Human milk fat*[a]			
4:0	11.8			100.0				
6:0	4.6		6.5	93.5				
8:0	1.9	24.5	12.3	63.2				
10:0	3.7	17.1	27.0	55.9	2.9	12.8	18.6	68.6
12:0	3.9	41.9	53.0	5.1	7.3	20.6	31.7	47.7
14:0	11.2	28.9	52.1	19.0	9.4	23.0	54.4	22.6
15:0	2.1	31.7	46.0	22.2	0.8	24.0	36.0	40.0
16:0	23.9	47.4	45.1	7.5	27.0	23.1	70.4	6.5
16:1	2.6	35.9	46.1	18.0	3.6	31.5	14.8	53.7
17:0	0.8	54.1	41.7	4.2				
18:0	7.0	49.1	45.2	5.7	7.1	66.7	23.0	10.3
18:1	24.0	41.7	26.2	32.1	34.2	42.9	7.9	49.2
18:2	2.5	22.7	46.6	30.7	7.9	30.5	15.7	53.8
18:3	tr.				tr.			
Goat milk fat[b]					*Sheep milk fat*[b]			
4:0	4.4			100.0	3.6			100.0
6:0	3.5			100.0	3.5			100.0
8:0	2.5	22.7	16.0	61.3	2.2	4.5	29.8	65.7
10:0	7.5	14.7	30.8	54.5	5.6	8.3	30.8	60.9
10:1	0.2		16.7	83.3	0.2		20.0	80.0
12:0	3.3	40.8	47.0	12.2	3.5	21.2	45.2	33.6
14:0	10.4	26.7	64.7	8.6	10.4	26.4	56.9	17.0
14:1	0.6		44.4	55.6	0.7	9.1	40.9	50.0
15:0	2.0	31.1	44.3	24.6	2.7	31.7	50.0	18.3
16:0	26.9	53.9	41.9	4.2	21.5	59.1	37.0	3.9
16:1	2.2	44.6	30.8	24.6	2.0	37.3	37.3	25.4
17:0	1.2	36.1	13.9	50.0	1.1	51.5	27.3	21.2
18:0	9.8	52.2	21.5	26.3	13.6	46.8	30.9	22.3
18:1	20.8	25.8	25.8	48.4	21.8	28.7	29.6	41.7
18:2	2.4	4.1	34.3	61.6	4.3	20.9	32.6	46.5
18:3					2.8	26.5	20.5	53.0
20:1	1.8	17.0	32.1	50.9	0.4	45.4	18.2	36.4
20:2	0.5	28.6	7.1	64.3	0.1			100.0
Horses' milk fat[c]					*Pigs' milk fat*[c]			
4:0								
6:0	0.3			100.0				
8:0	2.3	2.9	44.1	53.0				
10:0	3.6	6.5	42.6	50.9				
12:0	4.3	23.1	48.4	28.5	0.1	50.0	100.0	−50.0
14:0	5.0	34.2	60.4	5.4	2.4	24.7	69.8	5.5
14:1	0.9	40.7	40.7	18.6				
15:0	0.3	54.6	63.6	−18.2	0.2	50.0	83.3	−33.3
15:1	0.2	40.0	40.0	20.0				
16:0	16.4	47.9	51.7	0.4	26.9	32.6	59.6	7.7
16:1	9.0	31.2	40.5	28.3	4.0	29.2	42.5	28.3
17:0	0.1	100.0	25.0	−25.0	0.6	44.4	27.8	27.8
17:1	0.5	35.7	35.7	28.6				
18:0	1.0	90.0	20.0	−10.0	6.8	54.6	8.8	36.6
18:1	15.2	45.3	16.5	38.2	30.3	36.0	17.0	47.0
18:2	9.3	25.7	25.7	48.6	27.3	26.7	26.8	46.5
18:3	29.8	28.6	22.5	48.9	1.4	33.3	28.6	38.1
Other	1.8	3.6	50.0	46.4				

[a]Enzymatic stereospecific analysis using modified Brockerhoff method. Data adapted from Christie and Clapperton (1982).
[b]Enzymatic stereospecific analysis using Brockerhoff method (1965). Data adapted from Kuksis *et al*. (1973).
[c]Enzymatic stereospecific analysis using modified Brockerhoff method. Data adapted from Parodi (1982).
tr., trace.

with 16–18 carbons are more evenly distributed between the *sn*-1 and *sn*-2 positions. According to Gresti *et al*. the most common individual triacylglycerol species in bovine milk fat are 4:0–16:0–18:1 (4.2 mol%), 4:0–16:0–16:0 (3.2 mol%), and 4:0–14:0–16:0 (3.1 mol%), all of which are representatives of the triacylglycerol class consisting of one butyryl group at the *sn*-3 position and two long-chain acyl groups at the *sn*-1 and *sn*-2 positions.

The triacylglycerol composition of other ruminants' (e.g., goat and sheep) milk fat resembles that of cows' milk. A high proportion of even-numbered saturated short-chain fatty acids and fatty acids with 14–18 carbons, as well as the dominating amount 18:1 fatty acids among unsaturated fatty acids of triacylglycerols, is obvious (Table 5). Distribution of fatty acids between the *sn*-positions is also similar to bovine milk fat. However, unsaturated fatty acids (18:1–18:3) and saturated 16:0 and 18:0 fatty acids are more evidently esterified in the *sn*-3 and *sn*-1 positions, respectively, than in bovine milk fat.

One of the main differences between monogastric mammals (e.g., human, pig, horse) and ruminants' milk fats is the absence or scarcity of short- and medium-chain fatty acids in the milk fat of monogastric mammals (Table 5). Also, the amount of diene and polyene fatty acids is much higher in their milk fat. Palmitic and oleic acids are the most prominent saturated and unsaturated fatty acid, respectively, in monogastric mammals' milk fat as they are also in ruminants' milk fat. High concentration (70%) of palmitic acid at the secondary position is typical for human milk fat, as is the esterification of stearic acid in the *sn*-1 position. Unsaturated fatty acids are preferentially acylated in the primary positions of triacylglycerol molecules.

Depot Fats

Triacyglycerol compositions of animal depot fats are much simpler than those of milk fats, and almost all fatty acids are very unevenly distributed between the three *sn*-positions. In triacylglycerols of lard, *c*. 85% of palmitic acid and almost 80% of stearic acid are in the *sn*-2 and *sn*-1 position, respectively. However, all unsaturated fatty acids are mainly esterified in the *sn*-3 position (Table 6).

The most characteristic feature of tallow is the high content (> 25 mol%) of stearic acid, which is equally distributed between the two primary positions (Table 6). Palmitic and myristic acids are also most prominent in the *sn*-1 and *sn*-3 position, respectively. In contrast, the main proportion of unsaturated fatty acids is in the secondary position, *c*. 90% of linoleic acid and over 50% of oleic acid.

Egg Yolk

Oleic acid, palmitic acid, and linoleic acid, in that order, consist of nearly 90 mol% of all fatty acids in triacylglycerols of egg yolk (Table 6). According to Gornall and Kuksis, the most common triacylglycerol

Table 6 Distribution of fatty acids between the *sn*-1, *sn*-2, and *sn*-3 positions in egg yolk, lard, and tallow triacylglycerols

Fatty acid	*Total (mol%)*	*Percentage*			*Total (mol%)*	*Percentage*		
		sn-*1*	sn-*2*	sn-*3*		sn-*1*	sn-*2*	sn-*3*
Egg yolk[a]					*Lard*[b]			
14:0	1.0	33.3	33.3	33.3	1.7	25.0	59.6	15.4
16:0	27.0	77.5	6.3	16.2	28.8	12.0	85.1	2.9
16:1	5.0	41.2	17.6	41.2	2.8	24.3	48.2	26.5
18:0	6.0	35.3	17.6	47.1	12.1	78.8	6.3	14.9
18:1	48.0	14.8	40.1	45.1	43.6	37.5	10.0	52.5
18:2	13.0	7.3	75.6	17.2	10.1	23.7	13.2	63.1
18:3					0.9	51.9		48.1
Tallow[c]								
14:0	3.4	28.7	14.9	56.4				
15:0	0.6	33.3	22.2	44.5				
16:0	30.5	47.7	28.2	24.1				
16:1	1.9	36.2	13.8	50.0				
17:0	1.2	33.3	27.9	38.9				
18:0	26.0	44.2	14.5	41.4				
18:1	34.9	14.2	52.8	33.0				
18:2	1.5	4.6	88.6	6.8				

[a]Stereospecific enzymatic analysis. Data adapted from Gornall and Kuksis (1971).
[b]Stereospecific enzymatic analysis using kinetic resolution with phospholipase C. Data adapted from Myher and Kuksis (1979).
[c]Method based on separation of diastereomers of diacylglycerol derivatives on a silica-gel HPLC column. Adapted from Christie *et al*. (1991).

species of egg yolk are 16:0–18:1–18:1 (29 mol%), 16:0–18:2–18:1 (15 mol%), 18:1–18:2–18:1 (7 mol%), and 16:0–18:1–18:0 (7 mol%). Palmitic and linoleic acids are very unevenly distributed between the *sn*-positions: over 70% of 16:0 and 18:2 is in the *sn*-1 and *sn*-2 position, respectively. Stearic acid is evenly distributed between the primary positions and oleic acid between the *sn*-2 and *sn*-3 positions.

Fish Oils

In fish oils, myristic and palmitic acids are esterified preferentially in the *sn*-2 position or in the *sn*-1 position (Table 7). Stearic and palmito-oleic acid are esterified to a great extent in the *sn*-1 position or distributed evenly between the three positions with a slight preference for *sn*-1 position. Oleic and

Table 7 Distribution of fatty acids between the *sn*-1, *sn*-2, and *sn*-3 positions in fish oil triacylglycerols

Fatty acid	*Total (mol%)*	*Percentage*			*Fatty acid*	*Total (mol%)*	*Percentage*		
		sn-*1*	sn-*2*	sn-*3*			sn-*1*	sn-*2*	sn-*3*
Herring[a]					*Herring*[b]				
14:0	6.7	30.0	50.0	20.0	14:0	9.7	42.0	23.7	34.3
16:0	12.0	33.3	47.2	19.4	16:0	15.7	33.3	39.3	27.4
					16:1*n*-11	0.3	63.0	37.5	0.0
					16:1*n*-9	0.2	14.3	71.4	14.3
16:1	9.3	46.4	35.7	17.9	16:1*n*-7	5.0	44.1	26.2	29.7
18:0	1.0	33.3	33.3	33.3	18:0	0.9	34.6	34.6	30.8
18:1	11.3	47.1	29.4	23.5	18:1*n*-9	9.9	39.0	30.0	31.0
					18:1*n*-7	1.5	55.8	18.6	25.6
					18:1*n*-5	0.5	42.9	35.7	21.4
18:2	2.3	42.9	42.9	14.3	18:2*n*-6	1.2	33.3	38.9	27.8
					18:3*n*-3	1.4	30.0	40.0	30.0
					18:4*n*-3	3.8	42.2	35.8	22.0
					20:1*n*-11	1.2	38.9	16.7	44.4
20:1	17.0	49.0	11.8	39.2	20:1*n*-9	11.8	46.7	12.0	41.3
20:5	8.3	12.0	72.0	16.0	20:5*n*-3	5.9	17.9	62.5	19.6
22:1	23.0	20.3	7.2	72.5	22:1*n*-11. *n*-13	16.0	25.1	10.6	64.4
					22:1*n*-9	0.8	9.1	68.2	22.7
22:5	1.7	20.0	60.0	20.0	22:5*n*-3	0.5	14.3	85.7	0.0
22:6	1.7	20.0	60.0	20.0	22:6*n*-3	6.1	12.9	87.1	0.0
					24:1*n*-9	0.7	22.7	13.6	63.6
Mackerel[a]					*Tuna*[c]				
14:0	6.0	33.3	55.6	11.1	12:0	0.0	92.9	92.9	−85.7
16:0	13.7	36.6	51.2	12.2	14:0	3.7	34.4	46.0	19.6
16:1	7.0	52.4	28.6	19.0	15:0	1.3	43.5	41.2	15.3
18:0	2.0	50.0	16.7	33.3	16:0	22.3	53.0	31.5	15.4
18:1	17.0	41.2	17.6	41.2	16:1*n*-7	5.0	47.8	25.4	26.8
18:2	1.7	40.0	20.0	40.0	17:0	2.3	59.3	33.4	7.3
20:1	10.7	25.0	15.6	59.4	18:0	5.8	69.9	14.3	15.8
20:5	9.0	18.5	44.4	37.0	18:1*n*-9	15.5	40.1	15.0	44.9
22:1	15.7	38.3	10.6	51.1	18:2*n*-6	1.6	33.7	27.5	38.7
22:5	1.7	20.0	60.0	20.0	18:3*n*-6	0.1	4.8	52.4	42.9
22:6	2.3	28.6	0.0	71.4	20:0	0.4	30.1	10.5	59.4
Cod[a]					18:3*n*-3	0.5	21.9	24.0	54.1
14:0	3.0	22.2	55.6	22.2	20:1*n*-9	1.1	24.6	18.8	56.6
16:0	10.7	37.5	43.8	18.8	20:2*n*-6	0.3	24.4	27.8	47.8
16:1	9.3	39.3	28.6	32.1	20:3*n*-6	0.1	38.5	23.1	38.5
18:0	2.3	71.4	14.3	14.3	22:0	0.3	33.3	12.8	53.8
18:1	23.0	50.7	13.0	36.2	20:4*n*-6	2.1	12.9	44.2	42.9
18:2	1.3	25.0	25.0	50.0	20:5*n*-3	7.1	15.3	31.2	53.5
20:1	12.0	38.9	22.2	38.9	24:1*n*-9	0.4	31.5	8.7	59.8
20:5	8.3	12.0	44.0	44.0	22:6	29.9	9.9	47.4	42.7
22:1	10.3	38.7	25.8	35.5	Other	0.2	82.7	15.0	2.3
22:5	1.7	20.0	60.0	20.0					
22:6	2.3	42.9	28.6	28.6					

[a]Enzymatic stereospecific analysis. Adapted from Brockerhoff *et al*. (1968).
[b]Method based on separation of enantiomers of monoacylglycerol derivatives on a chiral HPLC column. Adapted from Takagi and Ando (1991).
[c]Method based on separation of enantiomers of diacylglycerol derivatives on a chiral HPLC column. Adapted from Myher *et al*. (1996).

eicosenoic acids are esterified predominantly in the primary positions with a higher proportion in the *sn*-1 position. Also, docosenoic acid is esterified abundantly in the primary position, but with a higher proportion in the *sn*-3 position in most fish species. Linoleic acid is esterified preferentially in the *sn*-2 or *sn*-3 positions. Eicosapentaenoic acid is preferably in the *sn*-2 or *sn*-3 positions, docosapentaenoic acid in the *sn*-2 position, and docosahexaenoic acid in one of the three positions.

Triacylglycerol Structure of Modified Fats and Structured Lipids

Modification of fats and oils has been studied extensively in order to improve their physical, chemical, and nutritional properties, and to make them more suitable for manufacture of fat products with desired properties. Generally used methods of modification that change the fatty acid composition of fat or oil, are hydrogenation, fractionation, and blending. In the last two decades, interest has focused on the lipase-catalyzed interesterification reactions. The change in the distribution of fatty acids between the three positions of glycerol backbone depends on the regiospecificity of the lipase catalyst. When a nonspecific lipase is used, the specific distribution of fatty acids of a natural fat changes to a random distribution. A similar product is obtained to that in chemical randomization. The change in triacylglycerol composition depends on the type of interesterification reaction. Ester–ester exchange is a reaction between two or more esters, for example triacylglycerols of plain fat or oil (intramolecular interesterification). The acidolysis reaction is a reaction between triacylglycerol and free fatty acid, in which an exchange of acyl groups takes place. In alcoholysis reaction, an acyl group is changed between an acylglycerol and an alcohol. Lipase-catalyzed modification of milk fat has been reviewed, including industrial applications for the manufacture of lipolyzed milk and cheese flavor products and the research efforts with the aim to improve nutritional and technological properties based on ester–ester exchange, acidolysis, and alcoholysis reactions.

The following definitions have been used for structured lipids: Structured lipids are modified or synthetic oils and fats containing both long-chain (mainly essential) fatty acids and medium- or short-chain fatty acids. Specific structured lipids are modified or synthetic oils and fats containing long-chain (mainly essential fatty acids) and medium- or short-chain fatty acids, in which each group is located specifically at the *sn*-2 position or *sn*-1,3 positions of the glycerol backbone. Specific structured lipids can be produced only by interesterification reactions catalyzed by regiospecific lipases.

As examples of industrial applications of lipase-catalyzed reactions, the manufacture of cocoa butter equivalent and human milk substitute are briefly described here. In cocoa butter, 87% of oleic acid is esterified in the *sn*-2 position, and 97% of palmitic and 98% of stearic acid is esterified in the primary positions (Table 3), and the three major molecular species are 16:0–18:1–16:0, 18:0–18:1–18:0, and 16:0–18:1–18:0. This specific distribution of fatty acids is reflected in the unique melting characteristics with a narrow melting range. Using 1,3-specific lipase to catalyze the reaction between oils, in which oleic acid is in the *sn*-2 position, and saturated acyl donors, a cocoa butter substitute, can be produced, with triacylglycerol composition and melting properties very similar to those of cocoa butter. In human milk, 70% of palmitic acid and in cows' milk 45% of palmitic acid is esterified in the *sn*-2 position (Table 5). The unique fatty acid distribution plays an important role in the nutrition of newborn. During ingestion, lipases hydrolyze the fatty acids esterified in the *sn*-1 and *sn*-3 position, and 2-monoacylglycerols and free fatty acids are absorbed. However, a part of the long-chain fatty acids is precipitated in the gut as calcium salts, leading to the loss of long-chain fatty acids and calcium. To overcome the problems of palmitic acid absorption, a human milk fat substitute, which closely resembles the specific structure of human milk fat, is produced from

Table 8 Properties of the main polymorphic forms of triacylglycerols

Property	*α-Polymorph*	*β′-Polymorph*	*β-Polymorph*
Molecular packing in the crystal	Hexagonal	Orthorhombic	Triclinic
X-ray diffraction, short spacing	4.15Å	4.2 and 3.8Å	4.6Å
Infrared spectroscopy	Single band at 720 cm^{-1}	Double band at 727 and 719 cm^{-1}	Single band at 717 cm^{-1}
Melting point (°C)			
TG 24:0 (tricaprylin)	−51.0	−18.0	+10.0
TG 30:0 (tricaprin)	−10.5	+17.0	+32.0
TG 48:0 (tripalmitin)	+45.0	+56.5	+66.0
TG 54:0 (tristearin)	+54.7	+64.0	+73.3

vegetable oil in a reaction catalyzed by a 1,3-specific lipase.

Properties of Triacylglycerols

Physicochemical Properties

The melting and crystallization properties of triacylglycerols of food fats and oils are of paramount interest in the food industry in manufacturing fatty foods. They are influenced by fatty acid composition, acyl distribution, molecular size, and polymorphism of triacylglycerols. Most triacylglycerols of natural fats and oils consist of straight-chain fatty acids with an even number of carbon atoms, 18:1, 18:2, and 16:0 being the most abundant. Double bonds are almost exclusively in *cis*-configuration, but ruminants' milk fats with a significant amount of *trans*-fatty acids are a rare exception. In polyene fatty acids, methylene-interrupted double bonds are most common, but conjugated double bonds are found in significant amounts in milk fat and other fats of animal origin. Substituted fatty acids are uncommon among natural fats and oils.

The crystallization properties of triacylglycerol mixtures at ambient temperatures are strongly associated with the amount of trisaturated triacylglycerols, but also, triacylglycerol species consisting of two saturated, long-chain fatty acids and one *trans*-monoene fatty acid influence the solidification properties. A strong positive correlation between softening point and content of trisaturated triacylglycerols has been detected in relatively simple fats and oils, but the situation is more complicated with more complex triacylglycerol mixtures. For example, no significant correlation between the softening point and the amount of trisaturated triacylglycerol in milk fat was observed, indicating the importance of the composition of the trisaturated fraction. In milk fat, a decrease in softening point is associated with an increase in the proportion of short-chain trisaturated triacylglycerols (TG 26–32) and a decrease in the proportion of long-chain trisaturated triacylglycerols (TG 44–54). In the 1960s, deMan and colleagues demonstrated the effect of the distribution of fatty acids between the *sn*-positions on the melting and solidification properties of triacylglycerols by randomizing the highly specific fatty acid distribution of milk fat using chemical interesterification. Although the treatment did not affect the fatty acid composition or the proportion of trisaturated triacylglycerols of milk fat, it resulted in an increased softening point, hardness, and content of high-melting-point triacylglycerols in milk fat.

Polymorphism is a common phenomenon for organic compounds consisting of long aliphatic carbon chains resulting in various packing of molecules in the crystal lattices, hence the multiple melting points of polymorphic forms of both pure triacylglycerols crystals and compound crystals of various triacylglycerols in natural fats and oils. The main polymorphic forms of triacylglycerols are named α, β′, and β. Generally, the α polymorph is the least stable, and the β polymorph is the most stable, but all of them can be found in crystallized food fats in appropriate conditions. Some of their properties and examples of differences between melting points of polymorphs of pure medium-chain and long-chain triacylglycerols are presented in Table 8. A schematic presentation of possible phase transitions from liquid to solid, and from solid to solid is shown in Figure 3.

Technological Properties

The importance of fatty acid composition of triacylglycerols is clearly demonstrated in oxidative and hydrolytic deterioration (lipolysis). Oxidative stability increases in the order polyunsaturated > monounsaturated < saturated fatty acids. The methylene-group ($-CH_2-$) between double bonds in polyunsaturated fatty acids is especially sensitive for radical formation, and thus initiation of autooxidation of fats and oils.

In addition to fatty acid composition, the distribution of fatty acids in triacylglycerols together with molecular size is important in hydrolytic deterioration catalyzed by lipases (EC 3.1.1.3). Several technologically meaningful lipases hydrolyze fatty acids exclusively from the primary positions and attack more rapidly at small triacylglycerols. Because of the specific fatty acid distribution in most food fats and oils, the composition of free fatty acids differs markedly from that of intact fat. This is most pronounced in lipolysis of bovine milk fat. Most triacylglycerols in bovine milk fat (over 30 mol%) are relatively small molecules containing at least one short-chain

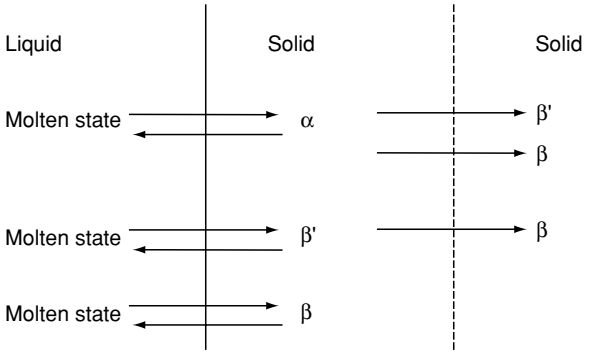

Figure 3 Possible transitions between liquid and solid, and between different polymorphic forms.

fatty acid esterified in the *sn*-3 position, resulting in rapid liberation of highly unpleasant-tasting and technologically problematic butyric and caproic acids into the media.

Rheological properties of spreads are greatly affected by the composition of raw materials, crystallization of emulsion (initial and final temperature, cooling rate, duration of crystallization), mechanical treatments, and configuration of manufacturing equipment. However, the polymorphism of triacylglycerols also has a substantial effect on the consistency and organoleptic acceptability of spreads. A rapid cooling rate of fats with complex fatty acid and triacylglycerol composition, e.g., milk fat, results in the formation of α polymorphs with rather fragile small crystals. Because of a larger mobility of molecules within the crystal lattice of the α polymorph, triacylglycerol molecules, even with a large difference between the chain length of acyl groups, are able to fit into the crystal lattice, and so a large amount of fat will crystallize initially. α crystals are very unstable and will usually transform into more stable polymorphs very rapidly. A slower cooling rate of fats with a simpler and more coherent triacylglycerol composition, e.g, cocoa butter, favor the formation of more stable β′ and β polymorphs. Small and needlelike crystals of β′ polymorph are desirable in spreads, because they increase plasticity of the products. In baked products, they aid creaming by incorporating large amounts of air in the form of small air bubbles into the structure. A decrease in the chain length variety enables the formation of a highly ordered crystal structure, increasing the probabiblity of crystallization of triacylglycerols straight into the most stable β polymorph. β crystals are larger than those of β′ and may lead to undesirable sandiness in spreads. The influence of technological modification of food fats on polymorphism has been demonstrated. For example, an increase in the proportion of liquid canola oil in butterfat decreased the amount of β′ polymorph in the mixture. Further, interesterification of butter resulted in crystallization of butterfat in the β′ polymorphic form.

Marangoni and Rousseau emphasized that, although the triacylglycerol structure, solid–liquid ratio, and polymorphism greatly affect the rheological properties of spreads, the microstructure of the fat crystals network, *i.e*, aggregates of fat crystals and their interaction, is of paramount importance in the viscoelastic properties of spreads. The elastic properties of crystal network and of individual aggregates influence the overall rheology of spreads. The formation of large crystals and further aggregation of the crystals in the form of spherulites (Ostwald ripening) are energetically favorable. The differences in aggregate formation could not be explained sufficiently by changes in polymorphism. Rousseau *et al*. suggested that the viscosity of the liquid phase, composition, and tempering have a more pronounced effect on the aggregation of spherulites than polymorphism.

The addition of milk fat or its fractions in chocolates or in imitation chocolates (compound coatings) is another important application in which modification of the composition of triacylglycerol mixture, and hence polymorphism of triacylglycerols, has an important role. Cocoa butter, the main component of chocolate, with three major triacylglycerol species, i.e., 16:0–18:1–16:0, 18:0–18:1–18:0, and 16:0–18:1–18:0, crystallizes mainly in the β polymorphic form and has six subforms named I–VI. The V polymorph is a very stable and desirable form, giving an excellent appearance (gloss) in chocolates. Polymorphic transitions during storage are known to result in bloom development, i.e., loss of the bright glossy appearance of chocolates. Milk fat is added to chocolate in order to prevent blooming, soften the chocolate, and improve the flavor. The addition of anhydrous milk fat or high-melting-point fractions of milk fat in dark chocolate inhibits both the rate of bloom development and the time for the onset of visible blooming, which is most likely related to changes in the phase behavior and crystallization kinetics. The softening effect of milk-fat addition is due to the increase in the amount of stable β′ polymorph resulting from the more complex triacylglycerol structure and the increase in the proportion of low-melting-point triacylglycerols. Up to 30% of milk fat can be added to cocoa butter without resulting in extreme softening.

Triacylglycerol Structure and Fat Metabolism

When a chiral triacylglycerol with three different esterified fatty acids enters the upper intestinal tract, lingual and gastric lipases catalyze the hydrolysis of some of the fatty acids in the *sn*-3 position, producing 1,2-diacyl-*sn*-glycerol and free fatty acid. The short-chain acyls at the *sn*-3 position are easily hydrolyzed by gastric lipases. A mixture of triacylglycerol proceeds to the small intestine. Pancreatic lipase and its colipase hydrolyze the fatty acid from the *sn*-1 position of the 1,2-diacyl-*sn*-glycerol, and 2-monoacylglycerol and free fatty acid are formed. Pancreatic lipase also hydrolyzes the triacylglycerol that was not hydrolyzed by the gastric lipase, slightly preferentially in the *sn*-1 position producing 2,3-diacyl-*sn*-glycerol and free fatty acid. The 2,3-diacyl-*sn*-glycerol is hydrolyzed further by carboxyl ester hydrolase or pancreatic lipase, and 2-monoacylglycerol and free fatty acid

are formed. A very small fraction of the 1,2- or 2,3-diacylglycerols is expected to undergo acyl migration to the 1,3-diacylglycerol, which is probably hydrolyzed by lipases. Polyunsaturated fatty acids are poor substrates for pancreatic lipase, but carboxyl ester hydrolase hydrolyzes them effectively. When the fatty acids and 2-monoacylglycerol are absorbed into the enterocyte, the 2-monoacylglycerol is reacylated by monoacylglycerol transferase to 1,2-diacyl-*sn*-glycerol and 2,3-diacyl-*sn*-glycerol. The fatty acyl-CoAs are similar for acylation of the position 1 or the position 3. The activated fatty acids used in the acylation come from newly synthesized fatty acids in the cell and the fatty acids absorbed from lumen after bile phospholipid and fat digestion. The formed 1,2-diacyl-*sn*-glycerol and 2,3-diacyl-*sn*-glycerol are acylated by diacylgycerol acyl transferase to triacylglycerols. Since the acyl migration is expected be slow at 37 °C, the synthesized triacylglycerols retain the original fatty acid at the *sn*-2 position. The reacylation of absorbed 2-monoacylglycerols is the predominant pathway during active fat absorption. When the absorption 2-monoacylglycerols from the gut is minimal, the acylation of free glycerol through the 3-glycerophosphate by activated fatty acids proceeds to form 1,2 diacyl-*sn*-glycerol phosphate. The unsaturated fatty acids acylate mainly the *sn*-2 position and saturate the *sn*-1 position. Triacylglycerol is formed through dephosphorylation and acylation at the *sn*-3 position. Triacylglycerols formed by the 3-glycerophosphate pathway do not maintain the original fatty acid of the diet at the *sn*-2 position. However, during fat absorption, the 2-monoacylglycerol pathway dominates.

The synthesized mucosal triacylglycerols are packaged to chylomicrons and intestinal very-low-density lipoproteins, and secreted into the lymph. They are hydrolyzed in the capillary bed. Finally, the fatty acids and partial acylglycerols are reesterified in adipose or muscle cells. If the monoacylglycerol pathway is active in the tissue, where the triacylglycerols are stored, the original composition of 2-monoacylglycerol is retained. However, it is more likely that a different distribution of fatty acids occurs in adipose tissue triacylglycerols.

See also: **Chromatography**: Principles; **Cocoa**: Production, Products, and Use; **Coconut Palm**; **Eggs**: The Use of Fresh Eggs; **Fish Oils**: Production; Composition and Properties; Dietary Importance; **Infants**: Nutritional Requirements; **Peanuts**; **Triglycerides**: Structures and Properties; Characterization and Determination; **Vegetable Oils**: Types and Properties; Oil Production and Processing; Composition and Analysis; Dietary Importance

Further Reading

Balcão VM and Malcata FX (1998) Lipase catalyzed modification of milkfat. *Biotechnology Advances* 16: 309–341.

Christie WW (1986) The positional distribution of fatty acids in triglycerides. In: Hamilton RJ and Rossel JB (eds) *Analysis of Oils and Fats*. London: Elsevier.

Fox PF (ed.) (1995) *Advanced Dairy Chemistry*, *Volume* 2: *Lipids*. London: Chapman & Hall.

Gunstone FD, Harwood JL and Padley FB (eds) (1994) *The Lipid Handbook*, 2nd edn. London: Chapman & Hall.

Hartel RW (1996) Applications of milk-fat fractions in confectionery products. *Journal of American Oil Chemists' Society* 73: 945–953.

Jensen RG, Ferris AM and Lammi-Keefe CJ (1991) Symposium: Milk fat – composition, function, and potential for change. The composition of milk fat. *Journal of Dairy Science* 74: 3228–3243.

Kuksis A (1992) Yolk lipids. *Biochimica et Biophysica Acta* 1124: 205–222.

Laakso P (1996) Analysis of triacylglycerols – Approaching the molecular composition of natural mixtures. *Food Reviews International* 12: 199–250.

Marangoni AG and Narine SS (eds) (2002) *Physical Properties of Lipids*. New York: Marcel Dekker, Inc.

Moran DPJ and Rajah KK (eds) (1994) *Fats in Food Products*. Glasgow: Blackie Academic & Professional.

Small DM (1991) The effects of glyceride structure on absorption and metabolism. *Annual Review of Nutrition* 11: 413–434.

Xu X (2000) Production of specific-structured triacylglycerols by lipase-catalyzed reactions: a review. *European Journal of Lipid Science and Technology* 2000: 287–303.

Characterization and Determination

E W Hammond, Greenisle-Consulting, Kettering, UK

Fatty Acid Composition

The fatty acid (FA) composition of triacylglycerides (TAG) is dependent upon the origin of the TAGs, but the FAs will normally be of even carbon number for vegetable fats and oils. Exceptions are TAG containing FA originating from bacterial (and some other microflora) origins. Thus, the TAG from ruminants such as cattle and sheep contain odd-chain and branched-chain fatty acids in small amounts. Typical fatty acid chain lengths for all TAGs are generally between C14 and C22, with the major fatty acids being of the C16 and C18 types. There are exceptions

here also. Some plants major in other chain lengths such as palm kernel and coconut fats (C8–C14), as well as cuphea, which has mostly C8 and C10, and wild-type rapeseed, which has half its FAs as a C22 chain length. Fish and marine animals have longer-chain FAs in the C18–C22 range. Amongst these FAs are the unsaturated series. These have one to six double bonds at specific positions in the chain because of their biochemical origin and species origin. There is also a range of FAs which have other chemically functional groups such as hydroxyl (e.g., ricinoleic acid from castor bean), epoxide (e.g., epoxyoleic acid from *Vernonia anthelmintica*) and cyclopropene (e.g., sterculic acid from *Sterculia foetida*).

The reason for mentioning the wide range of FAs is to point out that confident and accurate analysis of a sample of TAGs requires some knowledge of its origin. However, for most purposes, the analysis of the FA composition of TAGs from food is relatively simple, since the fats used have fairly uncomplicated fatty acid compositions. (*See* **Fats**: Classification; **Fatty Acids**: Properties.)

FA analysis of TAGs is done by gas chromatography after derivatization to methyl esters. (*See* **Fatty Acids**: Analysis.)

Positional Analysis

The FA composition of the three positions of the glycerol in TAGs is not entirely random: there is positional order. Lipase enzymes associated with TAG synthesis and hydrolysis are positionally specific in their action. This specificity is partly responsible for the differences in TAG composition between the many natural vegetable and animal fats and also for the consistency of fat structure within a fat species. The specific TAG structure of some fats is very important to their use in foods. For instance, cocoa butter has a high proportion of symmetrical TAGs, where position 2 is occupied by oleic acid, while positions 1 and 3 are occupied by a saturated fatty acid, typically palmitic and stearic acids. This structure produces a fat with very specific melting and crystallisation properties that are responsible for the particular physical qualities of chocolate products.

During research on such fats, it can be necessary to determine the positional composition of the FAs in the TAGs. This is achieved very conveniently using the specificity of some lipases. Usually, pig pancreatic lipase is used, which, in its purified form, is 1:3-specific. It hydrolyzes the esters on the 1 and 3 positions of TAGs, leaving the 2 position unchanged. The 2-monoacylglyceride obtained yields the composition of position 2. By examining the diacylglycerides (DAGs), product information on positions 1 and 3 can be obtained. If the DAG is chemically phosphorylated and then reacted with phospholipase A, information that is more specific is obtained about the *sn*-1 and *sn*-3 positions. (*See* **Cocoa**: Production, Products, and Use; **Enzymes**: Uses in Analysis; **Vegetable Oils**: Composition and Analysis.)

Recent significant advances in the interpretation of ^{13}C nuclear magnetic resonance spectra of pure TAGs show a more powerful way forward in determining TAG structure. However, at this time, the interpretation of spectra from natural vegetable oils is complex, because these oils are complex mixtures of different TAGs. In addition, the equipment used is not generally available to all laboratories that might be doing such an analysis. Alternative approaches using enzymes combined with different stereospecific derivatives and other column chromatographic techniques have been used with some success. These new approaches increase the possibility that such stereospecific positional analysis of natural TAG mixtures can become routine.

Methods

Isolation of TAGs

The sample must be greater than 99% TAGs for these methods to produce accurate results. If it is not, seven drops (100–120 mg) of melted and well-mixed fat are added to 2 ml of diethyl ether. This solution is applied to a small column of basic alumina (2.0 g, Brockmann Grade 1 basic alumina) in a glass filter tube, the column is washed with 2×5 ml diethyl ether, and the eluate is then collected in a screw-cap vial. The sample in the vial is then evaporated with a stream of nitrogen. This procedure normally provides 80–100 mg of pure TAGs. The method can be scaled up, but it is better to do multiple preparations in parallel.

Lipolysis – 2 Position Only

Approximately 100 mg of melted and well-mixed TAG is weighed into a screw cap vial (10 ml size). Two milliliters of tris(hydroxymethyl)aminomethane (Tris) buffer (1.0 M, pH 8.0) is added to the vial together with 0.2 ml of 22% aqueous calcium chloride (made up from the hexahydrate) and 10 drops of 10% aqueous Biosolve detergent/emulsifier (Beckmann solubilizer for scintillation counting). (Note: 0.5 ml of 0.1% sodium cholate may be substituted for the Beckmann biosolve, but emulsification will be less efficient, and yields are variable in our experience.) Other emulsifiers may be substituted, but require testing. The vial is placed in a water bath (set at 40 °C) for 15 min, and a solution of pure

monoheptadecanoin internal standard in diethyl ether (5 mg ml^{-1}) is prepared and stored in a closed vessel. The sample is sonicated for 1 min using an MSE ultrasonic sample preparator (MSE Soniprep 150) and then replaced in the water bath. This step needs to be done efficiently with a probe-type sonicator, particularly when the sample has a high melting point. The sample temperature is stabilized in the water bath for 2 min. An amount of the pancreatic lipase enzyme (Sigma catalogue number L3126), approximately equal to the weight of TAGs used, is added to the sample (the amount of enzyme is important). The vial is immediately capped and shaken vigorously in a wrist-action shaker for exactly 3.5 min (or the time determined for each batch of enzyme using standard TAG). Without delay, approximately 1.5 ml of cold, 2 M hydrochloric acid is added, and the vial transferred to an ice bath and shaken. Exactly 2.0 ml of the monoheptadecanoin solution and a further 2.5 ml diethyl ether are added and the sample quickly inverted a few times but without vigorous shaking. The vial is replaced in the ice bath, and the separated upper ether layer is transferred to another vial in the ice bath. The reaction mixture is extracted with a further 5 ml of ether and combined with the first on the ice bath. The combined ether extracts are washed twice with 5-ml portions of water, and then the ether extract is transferred to a vial containing approximately 3.0 g of anhydrous sodium sulfate. This is then capped and shaken to dehydrate the extract.

Isolation of monoacylglyceride (MAG) for FA analysis is carried out using the following procedure. Using either a pipette or a thin-layer chromatography (TLC) sample applicator, the ether solution is applied as a continuous streak 1 cm above the bottom of a TLC plate. The plate should be 20 cm × 20 cm glass coated with a 0.5-mm-thick layer of kieselgel 'G' (plates supplied by Anachem). This streak is focused by developing the plate twice to 3 cm above the base in 100% diethyl ether and air-drying the plate between developments. The plate is fully developed in a solvent of hexane/diethyl ether/formic acid (60:40:1, by v/v/v). The plate is air-dried and lightly oversprayed with 0.05% methanolic dichlorofluorescein. When the plate is dry, it is examined under ultraviolet light (350 nm), and the MAG, which will be at or just above the focused origin, is marked. (*See* **Chromatography**: Thin-layer Chromatography.)

The marked MAG band is scraped off and placed in a 20 ml Q/Q tube, to which 7 ml of acid methanolysis reagent is added. After thorough mixing, this is heated to 60 °C for 1 h with occasional mixing. Refluxing temperatures are discouraged here as vapor locks can form in the silica bulk, causing localized hot spots and possible decomposition of the fatty acid methyl esters (FAMEs). Water (7 ml) and diethyl ether (5 ml) are added, and the mixture is shaken after cooling. The top layer is transferred to a vial containing anhydrous sodium sulfate. The sample is then analyzed by gas chromatography. The MAG yield can then be estimated from the internal standard peak. The yield should be between 20 and 30% for reliable results that will reflect the true composition of FA on the 2 position of the TAG. (*See* **Chromatography**: Gas Chromatography.)

Lipolysis – 1:2 and 2:3 Positions

The DAG is required for this procedure. It is rare for good results for both MAG and DAG to be achieved from a single lipolysis. This is because the technique has to be optimized for the best yield of either MAG or DAG with zero or the least possible amount of acyl migration. Therefore, the lipolysis technique reported above for 2-MAG is repeated with the following changes:

1. The weight of lipase enzyme is reduced to one-quarter the weight of sample TAG (e.g., 25 mg of enzyme to 100 mg of TAG).
2. The reaction time is increased to 6 min or that time found for each batch of enzyme using standard TAG.
3. The MAG internal standard is not used.
4. After reaction, the sample is cooled on ice and carefully acidified to pH 2.0, then rapidly extracted into diethyl ether. The extract is washed twice with 5-ml aliquots of cold water, keeping it on ice as much as possible. The ether extract is then transferred to a vial containing anhydrous sodium sulfate, prior to TLC.
5. TLC is done as above, except that the formic acid is replaced by 2% ammonia. The 1,2/2,3 DAG will be found as a single band about 3 cm above the MAG. If lipolysis has been too protracted, acyl migration may have taken place, and two bands will be apparent (1,2/2,3 (lower) and 1,3 (upper) DAG). If the 1,3 DAG level is seen to be major component, the sample lipolysis must be repeated. The 1,2(2,3) DAG is recovered by elution from the silica with diethyl ether and concentrated with a stream of nitrogen.

The pure DAGs have to be chemically phosphorylated prior to reaction with phospholipase A, this is done according to the Lands *et al*. procedure using DAG kinase.

Twenty-five milligrams of the *sn*-1,2(2,3) DAG is weighted into a tube; the following are then added in sequence: 100 µl of bile salts solution (200 mg ml^{-1} in

water); 600 µl of 0.05 M adenosine triphosphate; 250 µl of 1.0 M magnesium chloride; 500 µl of 0.05 M sodium phosphate buffer, pH 7.95; 5.0 mg of DAG kinase in 500 µl of cysteine phosphate buffer.

After mixing and incubating at 37 °C with continuous shaking for 60 min, 1 ml of 1 M hydrochloric acid is then added.

The lipids are extracted from the reaction mixture with 15 ml of chloroform/methanol (2:1, v/v), shaking carefully, and then adding 3 ml of 1% sodium chloride solution. The lower chloroform layer is then transferred to another tube, and a second extraction is made with a further 6 ml of chloroform. Two drops of triethylamine are added to the combined chloroform layers, and the solvents are then evaporated. The phosphatidate products may be purified by TLC in a developing solvent of chloroform/methanol/water (65:36:8, v/v/v).

Stereospecific hydrolysis of the phosphorylated DAG is the final stage and is achieved with phospholipase A (*Crotalus adamanteus* venom – Sigma catalogue No. P0790). This enzyme is specific for the 2 position FA in an *sn*-1,2-diacyl 3-phosphatide but does not attack the *sn*-2,3-diacyl 1-phosphatide.

Ten milligrams of the phosphatidate is dissolved in 3 ml diethyl ether, to which 0.1 ml of 0.5 M Tris buffer (pH 8.5) containing 0.002 M calcium chloride and 0.5 mg of *Ophiophagus hannah* venom is added. This mixture is shaken gently for 16–20 h (overnight). Then, 5 ml of isobutanol and 20 µl of acetic acid are added and the whole reduced to dryness. The lipid in the residue is dissolved in a small amount of chloroform/methanol (2:1, v/v) and applied as a streak to a TLC plate and the application area thoroughly dried. The plate is first developed in hexane/diethyl ether/formic acid (50:50:1, v/v/v) to the top. The top third is sprayed with dichlorofluorescein (0.1%) in methanol. The free FA band is located under ultraviolet light (320 nm) and recovered. The TLC plate is then redeveloped in chloroform/methanol/14 M ammonium hydroxide (90:8:2, v/v/v), dried, and sprayed with rhodamine 6 G (0.1%) in toluene. The lysophosphatide and unhydrolyzed phosphatidate are located as above and recovered. All three of the recovered materials are prepared for analysis of their component FA by gas–liquid chromatography (GLC).

Results

The initial pancreatic lipase technique gives a reliable composition of the FAs present at the 2 position of a TAG. The phosphorylation and phospholipase A techniques produce a free FA fraction, a lysophosphatide, and unreacted phosphatidate. The FA fraction is typical of the *sn*-2 position and should confirm the result from the 2-position lipolysis, although there are often discrepancies in this comparison, and the primary (pancreatic lipase) result should be regarded as more reliable. The lysophosphatide gives the FA composition of the *sn*-1 position of the TAG. The unreacted phosphatidate gives the FA composition of the *sn*-2 and *sn*-3 positions, combined, of the TAG. The difference provides a complete TAG positional composition.

Cis/trans Ratios

The analysis of the total *trans* content by infrared (IR) techniques is covered in detail elsewhere in this Encyclopedia, which includes a full IR method. A *cis/trans* ratio is obtained in this case by the ratio of the percentage total unsaturates (determined by GLC) minus the percentage *trans* value (determined by IR analysis). (*See* **Fatty Acids**: Analysis; **Spectroscopy**: Infared and Raman.)

Trans unsaturated fats have been a part of the human diet at least since meat and fat derived from ruminant animals began to be consumed. However, the amount of such *trans* unsaturated FAs consumed was relatively small. The advent of chemically hardened fats for the manufacture of margarine and shortenings dramatically changed the level of *trans* unsaturated FAs consumed. The hardening process involves catalytic addition of hydrogen across the unsaturated centers. However, during the process, a proportion of the double bonds isomerize to the *trans* geometry, typically 30–70% depending on the type of catalyst and the conditions of the hardening process. The *trans* isomerization process is accompanied by double bond movement along the FA chain, so that a complex mixture is formed. This mixture is very difficult to analyze with which to obtain accurate *cis/trans* ratios. GLC of FAMEs derived from the fat has proved the best technique, but even here, there are considerable problems. (*See* **Margarine**: Methods of Manufacture.)

For GLC of FAMEs, the official AOCS method prescribes the cyanopropylpolysiloxane phase SP2340. The column should be 60 m × 0.25 mm fused silica with a 1-m retention gap of silanized fused silica with a film thickness of 0.2 µm. Such columns provide adequate separation of *cis* and *trans* FAME for most situations involving food fats (except fats containing hydrogenated fish oils). However, there will be peak overlap, and for exacting analyses, where information is required on the isomeric nature of the *trans* FAME, even these columns are not adequate. For separation that is more complete, a column of SP2560, 0.2-µm film on 100 m × 0.25 mm fused silica should be used. In both column cases, the

carrier gas can be helium (linear velocity 20 cm s^{-1}) or hydrogen (40 cm s^{-1}). The column temperature should be optimized, using standards, between 175 and 200 °C isothermal. Otherwise, faster results can be obtained by programming the temperature between 100 and 225 °C with some sacrifice to optimum resolution. The samples should be injected using the on-column technique in preference to split injection.

Cis/trans ratios calculated from this type of chromatography are generally more accurate than those obtained via IR techniques. However, this is true only below 10% *trans* content. Above 10% *trans* content, the IR technique gives more accurate results.

TAG Analysis in Foods

TAG is readily analyzed by chromatography after extraction from the food matrix. The extraction may be the most difficult part of the whole method and should be given careful attention. Many foods lend themselves to Soxhlet extraction with petroleum or chloroform solvents. Other foods containing spray-dried, or encapsulated fats have to be hydrated first, or fat yields will be low. Cereal flours require extraction with water-saturated *n*-butanol followed by a mixture of chloroform/methanol in order to obtain all their lipids. Once a total lipid extract is obtained, it can be analyzed complete for class composition.

If the TAG is less than 95% of the total lipid, it should be 'cleaned up' as above. This TAG can be analyzed via three routes; high-temperature gas chromatography (HTGC), nonaqueous reversed-phase high-performance liquid chromatography, or argentation high-performance liquid chromatography. (*See* **Chromatography**: High-performance Liquid Chromatography.)

HTGC can be applied in two forms: nonpolar-phase- and polar-phase gas chromatography (GC). Nonpolar phase GC is the choice for accurate basic information about composition, which, when considered together with FAME analysis, will normally confirm the type of fat. The data provided are a separation based upon 'carbon number' (essentially molecular weight), but it does not discriminate between saturated and unsaturated species of TAGs. A 7–10 m × 0.53 mm fused-silica column with a bonded phase of SP2100, OV1, or equivalent at a thickness of 0.1 μm gives excellent results (a 1-m silanized retention gap must be fitted). The injection technique should be on-column (Carlo Erba) with helium or hydrogen as the carrier gas. The carrier gas flow is high at 10 ml min^{-1}. For nonlauric fats, the initial column temperature is set at 100 °C, held for 2 min, and then programmed at 25 °C min^{-1} to 310 °C. For lauric fats or fats containing butter, the initial temperature is set to 50 °C with a hold of 4 min, and then programmed at 25 °C min^{-1} to 310 °C. The TAG sample is dissolved in iso-octane (or, if solubility is a problem, chloroform) to a concentration of about 5 mg ml^{-1} (5 μg in 5 μl). The sample volume loaded is 0.5 μl, equivalent to about 2 μg of TAGs. Figure 1 shows typical data.

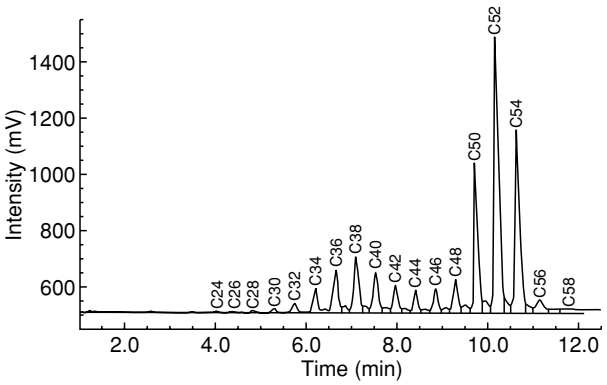

Figure 1 Triacylglyceride carbon number analysis by HTGC using cold on-column injection and a Carlo Erba gas chromatograph. See text for further details. The sample is a mixture of cocoa butter and butter fat. The peak names are the sum of the carbons in the FA chains and exclude those in glycerol (e.g., C54 might equal 3 × C18 or C16 + C18 + C20).

The polar-phase GC technique provides more information on the isomeric composition of TAGs and discriminates between the saturated and unsaturated TAGs. However, there are quantitative problems in the application of this method, requiring trained analysts with good interpretive skills.

See also: **Chromatography**: Thin-layer Chromatography; High-performance Liquid Chromatography; Gas Chromatography; **Cocoa**: Production, Products, and Use; **Enzymes**: Uses in Analysis; **Fats**: Classification; **Fatty Acids**: Properties; Analysis; **Margarine**: Methods of Manufacture; **Spectroscopy**: Infrared and Raman; **Vegetable Oils**: Composition and Analysis

Further Reading

Buteau GH, Jr. and Fairbairn D (1969) *Experimental Parasitology* 25: 265.

Damiani P, Santinelli F, Simonetti MS, Castellini M and Rosi M (1994) Comparison between two procedures for stereospecific analysis of triacylglycerols from vegetable oils – 1: Olive Oil. *Journal of the American Oil Chemists' Society* 71(10): 1157–1162.

Duchateau GSMJE, van Oosten HJ and Vasconcellos MA (1996) Analysis of *cis*- and *trans*-fatty acid isomers in hydrogenated and refined vegetable oils by capillary gas–liquid chromatography. *Journal of the American Oil Chemists' Society* 73(3): 275–282.

Hammond E (1989) Chromatographic techniques for lipid analysis. *Trends in Analytical Chemistry* 8(8): 308–313.
Hammond EW (1993) *Chromatography for the Analysis of Lipids*. Boca Raton, FL: CRC Press.
Lands WEM, Pieringer RA, Slakey PM and Zschocke A (1966) A micro method for the stereospecific determination of triglyceride structure. *Lipids* 1: 444–448.
Litchfield C (1972) *Analysis of Triglycerides*. New York: Academic Press.
Mannina L, Luchinat C, Emanuele MC and Segre A (1999) Acyl positional distribution of glycerol tri-esters in vegetable oils: a ^{13}C NMR study. *Chemistry and Physics of Lipids* 103: 47–55.
Ratnayake WMN and Beare-Rogers JL (1990) Problems of analysing *c9 cis* and *trans* fatty acids of margarine on the SP2340 capillary column. *Journal of Chromatographic Science* 28: 633–639.

Tripe *See* **Offal**: Types of Offal

TRITICALE

K Lorenz, Colorado State University, Fort Collins, CO, USA

Introduction

Triticale, the first synthetic cereal, is the product of a cross between wheat (*Triticum* spp.) and rye (*Secale cereale* L.) and has been of both practical and theoretical interest. This cross is an attempt to combine in one strain the high yield and good bread-baking quality of wheat and the winter hardiness, high lysine and protein, drought tolerance, and disease resistance of rye. In all soils presently unsatisfactory for wheat but acceptable for rye, it would be desirable to be able to grow a cereal which is equal to wheat in both yielding capacity and in nutritional, and bread-baking quality.

While it was simple to outline this task, many difficulties had to be overcome before successful crosses and fertile hybrids could actually be produced. From the 1930s to the 1950s, research on triticale never really progressed past the botanical curiosity stage, despite considerable efforts. Many of the goals that had been set eluded the plant breeders. Recent research efforts, however, have resulted in triticales of acceptable quality and characteristics which are grown commercially in many countries around the world.

Breeding

Triticales are produced by crossing either tetraploid wheat (*Triticum durum* L.: genomes AABB) or hexaploid wheat (*Triticum aestivum* L.: genomes AABBDD) with diploid rye (*Secale cereale* L.: genomes RR) and doubling the resulting haploid by treatment with colchicine, which produces a fertile amphidiploid. These crosses produce primary hexaploid (genomes AABBRR) and octoploid (genomes AABBDDRR) triticales, respectively. The production of triticales is outlined in **Table 1**. Intercrossing primary triticales within their own ploidy levels was practised from the 1930s until the 1950s and resulted in improvements in plant type, but yields and stabilities of the best lines remained significantly below those of wheat. In the mid-1960s, hexaploid triticales and hexaploid wheat were intercrossed, producing significant improvement. Dwarfing and disease resistance genes from hexaploid wheat were added to the hexaploid triticale gene pool. In the 1980s, crossing of octoploid triticales with hexaploid triticales, which allows substitution of D-genome chromosomes and chromosome segments from wheat for those of the A and B genomes, became the main approach to triticale germ plasm expansion. Perhaps the most important aspect of triticale breeding for the future, especially for hybrid production, is the use of doubled-haploid technology.

World Production and Yields

Triticale production figures, given in **Table 2**, are obtained from communications with plant breeders. Areas of production have been increasing annually for several years: the increase is occurring mainly on marginal soils, such as the sandy soils of Poland and Russia. In many countries, triticale offers farmers an alternative crop which is not controlled by any

Table 1 Development of primary and secondary triticales and examples of substitutional triticales

Primary triticale		
Common wheat	×Rye	[a]→Octoploid triticale
Triticum aestivum L.	*Secale cereale* L.	
AABBDD ($2n = 42$)	RR ($2n = 14$)	AABBDDRR ($2n = 56$)
Durum wheat	× Rye	[a]→Hexaploid triticale
T. turgidum L.	*S. cereale* L.	
AABB ($2n = 28$)	× RR ($2n = 14$)	AABBRR ($2n = 42$)
Secondary triticale		
Triticale	× Triticale	
AABBDDRR	× AABBDDRR	→AABBDDRR
AABBDDRR	× AABBRR	→AABBDDRR or AABBRR
AABBRR	× AABBRR	→AABBRR
Triticale	× Wheat	
AABBDDRR or AABBRR	× AABBDD	
AABBDDRR or AABBRR	× AABB	
Triticale	× Rye	
AABBDDRR or AABBRR	× RR	
Substitutional triticales[b] (examples with $2n = 42$ only)		
Example 1 A_7 A_7 B_7 D_2 D_2 R_5 R_5		
Example 2 A_6 A_6 B_6 B_6 D_6 D_6 R_3 R_3		
Example 3 A_7 A_7 B_6 B_6 D_7 D_7 R_1 R_1		
Example 4 A_7 A_7 B_7 B_7 D_7 D_7 $(B/R)_1$ $(B/R)_1$		
Example 5 A_7 A_7 B_7 B_7 D_7 D_7 (single gene transfer from rye)		

[a]Colchicine treatment is given to hybrid plants to double chromosome number.
[b]Subscripts indicate the number of chromosomes present from each genome. B/R indicates a translocation of B and R chromosomes.

Table 2 Estimates of world production of triticale (area under cultivation: ha)

Country	*1986*	*1996*
Poland	309 000	659 300
Russia	250 000	500 000
Australia	140 000	300 000
Germany	36 000	300 000
US	100 000	180 000
France	200 000	162 000
Bulgaria	10 000	100 000
South Africa	15 000	95 000
Brazil	16 000	90 000
Portugal	7 000	90 000
Spain	35 000	80 000
Sweden	n.a.	45 000
Italy	15 000	30 000
Romania	n.a.	20 000
Argentina	25 000	16 000
Tunisia	5 000	16 000
UK	16 000	16 000

n.a. not available.
1 acre = 0.405 ha.
Data from Gustafson JP, Bushuk W and Dera AR (1990) Triticale: production and ulitization. In: Lorenz K and Kulp K (eds) *Handbook of Cereal Science and Technology*, pp. 373–399. New York: Marcel Dekker; and Varughese G (1997) Triticale: an overview. In: *Proceedings of Satellite Meeting of the International Triticale Association on Triticale Quality*. All Africa Crop Science Congress, Pretoria, South Africa, pp. 10–15.

marketing board nor required to meet quality standards of other grain crops. Recent cultivars also have a dual purpose by providing green feed for animals before developing into a grain crop.

The yield potential of triticales under optimum crop production environments has reached nearly the same level as that of wheat while outperforming wheat under marginal environments. Yield capacity has improved because of the addition of genes for tillering, additional dwarfing, and better grain test weight. Yields will vary, however, depending on the triticale variety, location at which it is grown, the date of seeding, and the amount of available water during the growing season.

Machinery designed for the harvest of wheat can be used for triticale. No special harvest equipment is needed.

Agronomic Factors

The influence of agronomic variables on yield, composition, and other triticale characteristics has been described in many scientific publications. In general, the ability to adapt still seems to be limited for some varieties of triticale. Their performance is influenced by such factors as latitude, day length, elevation, temperature, availability of moisture, and nutrients in the soil. Some triticales still have the tendency to produce few tillers which limits yield. Triticale still lacks the benefits of many generations of natural selection which wheat and rye have gone through in order to perform well under many different agronomic conditions.

Any disease that attacks wheat and rye also attacks triticale. The cross does not exhibit improved

resistance to many wheat diseases, as was hoped. Some genes for resistance in either parental species are not fully expressed in triticale, which resulted in susceptibility to disease for which wheat or rye show resistance. There are no explanations. Obviously, complex reactions between genes of wheat and rye take place.

Kernel Development and Sprouting

Kernels of some triticales have a tendency to be shriveled. Triticales developed from the late 1960s up to the mid-1970s almost always showed shriveled grain. The causes still need to be elucidated today. Many potential causes for grain shriveling have been examined and rejected. There seems to be a genetic component influencing this condition. Shriveled kernels have a higher production of aberrant nuclei in the coenocytic endosperm and higher levels of acid phosphatase. Application of selection pressure for plump grain between 1977 and 1992 produced triticale advanced lines with high test weights and high flour yield. These lines have plump to only slightly shriveled grain.

Preharvest sprouting associated with late maturity has been a major problem, inhibiting the commercialization of triticale in several countries. Development of early-maturing varieties, capable of escaping the high-moisture period during harvest, has decreased chances of preharvest sprouting. New sources of genetic resistance to this condition in wheat parents are being utilized in present breeding efforts.

Composition

Average percentage composition of triticales is shown in Table 3. Slightly higher protein values compared to wheat are frequently cited as an advantage of triticale over wheat. However, high protein values are often misleading. Nitrogen values for shrivelled kernels, in which ratios of endosperm to bran plus germ are low compared to normal kernels, are higher than those for plump grains. Germ, scutellum, and aleurone are proportionately higher in nitrogen than the endosperm. Furthermore, triticale proteins are often calculated using a conversion factor of $N \times 6.25$, expressed on a dry-weight basis, which gives 30% higher values than $N \times 5.7$ at 14% moisture. Decreased protein values have accompanied the improvement of triticale's agronomic qualities. This is not surprising, since improvements in grain size result in a decrease in the percentage of protein in cereals owing to a disproportionate increase in starch content. (*See* **Cereals**: Contribution to the Diet.)

Lysine, which is the first limiting amino acid in wheat, is often, but not always, present in larger amounts in triticale than in wheat. Lysine is still the first limiting amino acid in triticale. Essential amino acid content of triticale compared to wheat and rye is shown in Table 4. Expressed as yield per hectare, the differences in lysine between wheat and triticale do not seem significant since lysine depends on the productivity of the variety. (*See* **Amino Acids**: Properties and Occurrence.)

The starch and fat (ether extract) contents of triticale are not significantly different from those of the parental species.

Except for lower amounts of nicotinic acid in triticales compared to wheat, the vitamin composition of triticales is comparable to that of wheat and better that that of rye, as illustrated in Figure 1. Average mineral composition of triticales compared to that of wheat shows no major differences, with the possible exception of iron, which appears to be present in higher amounts in triticales than in wheats, as shown in Table 5.

Toxic substances and nutritional inhibitors in triticale, which may interfere with the digestibility and availability of certain nutrients to humans and/or animals, include ergot, resorcinols, phytates, and enzyme inhibitors.

Ergot results from an infection of cereal grains by the fungus *Claviceps purpurea* which produces

Table 3 Average percentage composition of triticales

	Average composition (%) (ranges in parentheses)
Dry matter	87.0 (86.7–90.0)
Moisture	12.0 (10.0–13.3)
Crude protein (N × 6.25)	13.2 (11.3–18.5)
Ether extract	1.8 (1.3–2.7)
Crude fibre	2.4 (1.4–3.8)
N-free extract	67.8 (65.0–80.0)
Ash	1.8 (1.4–2.4)
Starch	50.7 (46.5–54.4)

Table 4 Essential amino acid content (g per 100 g protein) of triticale compared to wheat and rye[a]

Amino acid	*Triticale (Yoreme)*	*Wheat (INIA)*	*Rye (Snoopy)*
Lysine	3.44	2.83	4.02
Threonine	3.55	2.98	4.06
Methionine	1.28	1.42	1.35
Isoleucine	3.45	2.68	3.70
Leucine	7.20	7.22	7·75
Phenylalanine	4.94	3.77	4.74
Valine	4.48	3.73	5.10

[a]Cultivar name in parentheses. Data adapted from Gustafson JP, Bushuk W and Dera AR (1990) Triticale: production and utilization. In: Lorenz K and Kulp K (eds) *Handbook of Cereal Science and Technology*, pp. 373–399. New York: Marcel Dekker.

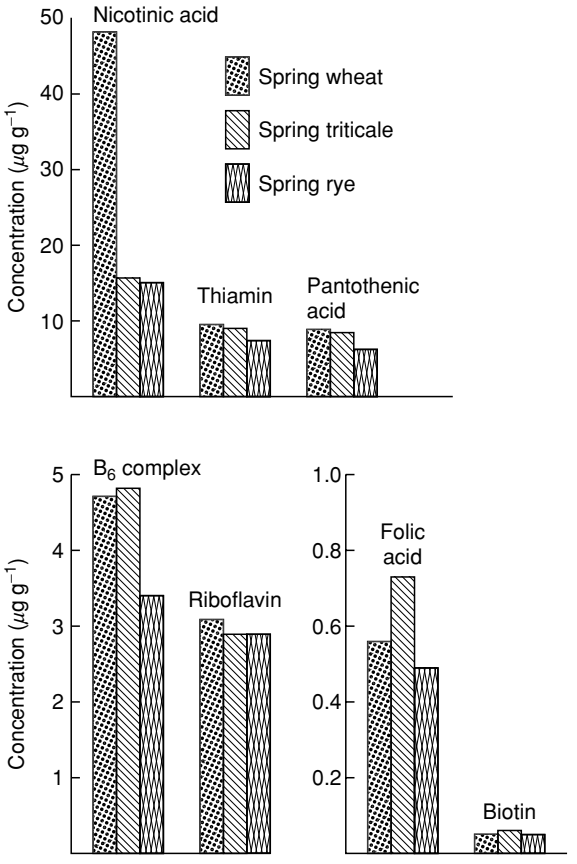

Figure 1 Vitamins in triticale, wheat, and rye. Reproduced from Triticale. *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Table 5 Mineral composition of triticale and wheat

Mineral	*Triticale*	*Wheat*
K (mg 100 g^{-1})	400–500	330–494
Mg (mg 100 g^{-1})	110–190	180–190
Ca (mg 100 g^{-1})	30–150	34–40
P (mg 100 g^{-1})	240–487	230–370
Na (ppm)	45–200	30–50
Zn (ppm)	13–26	17–30
Fe (ppm)	51–100	40–49
Cu (ppm)	3–7	4–5
Mn (ppm)	10–55	12–38

Data from Lorenz K (1982) Triticale processing and utilization: comparison with other cereal grains. In: Wolff IA (ed.) *Handbook of Processing and Utilization in Agriculture*, pp. 277–327. Boca Raton, FL: CRC Press; and Kulshrestha K and Usha MS (1992) Biochemical composition and nutritional quality of triticale. *Journal of Food Science and Technology (India)* 29: 109–115.

alkaloid-containing sclerotia. Relatively high levels of infections of ergot on triticales have occurred in some years. However, no statistics are available about the extent of infections in the various triticale-growing areas of the world.

Resorcinols seem to be partially responsible for reduced food intake and rate of weight increase of cattle, sheep, pigs, poultry, and horses. Triticales contain higher amounts of resorcinols than wheat. Milling of triticales will channel most of these resorcinols into the feed fractions. Phytic acid forms insoluble compounds with minerals, making these elements unavailable. Phytic acid, trypsin, and chymotrypsin inhibitors have been extracted from triticale in amounts intermediate between wheat and rye. (*See* **Phytic Acid**: Nutritional Impact.)

Food Uses

The prediction of the 1970s of a rapid expansion of the acceptance of triticale by millers and bakers for cereal-based products has not been realized. It appears that the success of triticale today depends on its ability to substitute to a greater or lesser extent for traditional wheat products.

Triticale grain can be milled into flour by the same milling process as used for wheat or rye. Because of the wide variation in kernel hardness and the degree of kernel shriveling, there is no standard procedure for triticale. Research has shown that breads can be baked from triticale flours provided that adjustments in formulation, mixing, and fermentation are made from those used in the production of white bread. Yeast level is increased, the fermentation temperature is lowered, and fermentation and proofing times are shortened. The high α-amylase activity of most triticale flours requires fermentation adjustments. Dough mixing is critical, since triticale flours do not have the same quality of gluten as wheat flour. The production of breads of low specific volume from triticale flours, as they are produced in many developing countries, creates less of a problem than production of the high-volume white bread.

The overall bread-making quality of newer triticale cultivars is considerably better than that of earlier ones, but it is still somewhat inferior to that of bread wheat of the same protein content due to a deficiency in protein quality, as reflected by a lower percentage of gluten compared to bread wheat.

Triticale flour can be used to some extent in the production of cakes, biscuits, tortillas, and other soft wheat products since triticale basically performs like a soft wheat. Layer cakes of acceptable quality can be produced from 100% triticale flour after proper chlorine treatment of the flour. Formulations of layer cakes from blends of triticale–wheat flour, ranging from 20% to 50% triticale, and additional emulsifier in the formulation produced cakes equal to or significantly larger than the soft wheat control cakes without additional emulsifier.

Triticale flours gave significantly smaller biscuit diameters and top-grain scores than biscuits baked with soft wheat flours. The biscuit-baking performance of flours from certain triticale cultivars may be improved, however, by increasing emulsification in the dough system, to equal or exceed soft wheat standards without additives.

Triticale pancake and waffle mixes have appeared on supermarket shelves. They are indistinguishable in appearance from those made with wheat flour, but differ in flavor and taste. Protein concentrates and starch have been prepared from triticale. Whole-grain triticale has been used to make bulgur.

Malting and Brewing

Compared to barley, triticales show relatively large malting losses, but yields of malt extracts are generally higher with triticales than with barley. Triticale malts are high in nitrogenous material, diastatic power, and α- and β-amylase. (*See* **Malt**: Malt Types and Products; Chemistry of Malting.)

Beers from triticale worts are darker in color and have a higher pH than beers from barley. Most triticales produced beers with satisfactory clarity–stability and gas stability. Triticale beers generally contain less alcohol and more nitrogenous compounds than barley beers. (*See* **Beers**: Wort Production.)

Feed Grain

Most triticale grown in the world today is used as feed grain. Triticale feeding trials in general have been encouraging. The overall conclusion from tests on weaned piglets, growing-finishing pigs, and broiler chicken was that triticale rated equal to wheat and corn. Protein digestibility was significantly higher with triticale than with sorghum in rations for cattle and sheep. Triticale was found to be comparable to durum wheat fed to turkeys. There are, however, also triticale feeding studies which reported poor feeding efficiency and reduction in weight gain. Reasons for such inconsistencies are as varied as the triticale varieties themselves. Variations in protein content, amino acid composition, nutritional availability, and antinutritional factors can all have an effect upon triticale's performance as a feed grain.

It appears, however, that ergot-free triticale is at least equal to wheat when used as a partial replacement in animal rations.

Forage

Triticale has been used as a forage from Palouse silt loam of the US Pacific Northwest, to loam soils of Guelph, Ontario, Canada, to the fine sandy loam of Texas; although not originally intended for its use, it has gained acceptance in most areas as an alternative forage. Both winter and spring varieties may be utilized as a nutritious food supply for livestock. Forage yields have been reported from as low as 4.98 tons ha^{-1} to 12.55 tons ha^{-1} with different varieties.

In southern New South Wales, Australia, dual-purpose triticales have almost completely replaced oats as a forage crop. In China, triticale has been identified as a promising silage crop compared to barley. In many parts of the world triticale is being used successfully as a forage or in mixtures with legumes for grazing, cut forage, hay, or as whole crop silage.

See also: **Bread**: Breadmaking Processes; Chemistry of Baking; Dietary Importance; **Cakes**: Nature of Cakes; Methods of Manufacture; **Cereals**: Contribution to the Diet; Dietary Importance; **Malt**: Chemistry of Malting

Further Reading

Bushuk W and Larter EN (1980) Triticale: production, chemistry and technology. In: Pomeranz Y (ed.) *Advances in Cereal Science and Technology*, pp. 115–157. St. Paul, Minnesota: American Association of Cereal Chemists.

Darvey NL, Naeem H and Gustafson JP (2000) Triticale: production and utilization. In: Kulp K and Ponte J (eds) *Handbook of Cereal Science and Technology*, pp. 255–272. New York: Marcel Dekker.

Gustafson JP, Bushuk W and Dera AR (1990) Triticale: production and utilization. In: Lorenz K and Kulp K (eds) *Handbook of Cereal Science and Technology*, pp. 373–399. New York: Marcel Dekker.

Kulshrestha K and Usha MS (1992) Biochemical composition and nutritional quality of triticale. *Journal of Food Science and Technology (India)* 29: 109–115.

Lorenz K (1974) The history, development and utilization of triticale. *CRC Critical Reviews in Food Technology* 5: 175–280.

Lorenz K (1982) Triticale processing and utilization: comparison with other cereal grains. In: Wolff IA (ed.) *Handbook of Processing and Utilization in Agriculture*, pp. 277–327. Boca Raton, FL: CRC Press.

Muentzing A (1979) Triticale: results and problems. *Zeitschrift fuer Pfanzenzuechtung* 10(suppl): 1–103.

Tsen CC (ed.) (1974) *Triticale: First Man-Made Cereal*. St Paul, Minnesota: American Association of Cereal Chemists.

Varughese G (1997) Triticale: an overview. In: *Proceedings of Satellite Meeting of the International Triticale Association on Triticale Quality*. All Africa Crop Science Congress, Pretoria, South Africa, pp. 10–15.

Varughese G, Pfeiffer WH and Pena RJ (1996) Triticale: a successful alternative crop. *Cereal Foods World* 41: 474–482.

TRYPSIN INHIBITORS

G P Savage and S C Morrison, Lincoln University, Canterbury, New Zealand

Introduction

Trypsin inhibitors occur in a wide range of foods. They have been studied extensively in plants belonging to the Leguminoseae, Solanaceae, and Gramineae families, as the majority of species in these families are considered as important food sources; the most important species are the grain legumes such as chickpeas, mung beans, and soy beans. Grain legumes are a major source of protein in the diets of many people throughout the world and are thought to provide about 10% of the world's dietary protein. They are used extensively as a protein source for animal feeds. The existence of a number of antinutritive factors in legume plants, especially a group of protease inhibitors, has limited the use of legume seed protein. Several different types of protease inhibitors may be present in the same tissue.

Structure and Mechanism of Action

The most common inhibitors in legumes act on serine proteases, a group of proteolytic enzymes, including trypsin and chymotrypsin. Serine protease inhibitors are proteins that form very stable complexes with these digestive enzymes, reducing their activity to very low levels. Some nonprotein inhibitors have also been identified.

Protease Inhibitors

Protease inhibitors have been classified into families based on the homologous sequences of amino acids at the reactive inhibitory sites. The molecular structure of the inhibitor affects both strength and specificity of the inhibitor. The two main groups of protease inhibitors found in legumes are the Kunitz soy bean trypsin inhibitor and the Bowman–Birk soy bean protease inhibitor.

Protease inhibitors exist in several different isoforms, each with differing isoelectric points and stability towards heat. In New Zealand peas, for instance, between six and 10 isoinhibitors in each cultivar have been observed, with isoelectric points ranging from 4.6 to 7.6.

Kunitz Soy Bean Trypsin Inhibitor (KSTI)

The first protease inhibitor to be isolated and characterized was KSTI. Molecular weights (mol wt) ranging from 18 to 24 kDa and isoelectric points of 3.5–4.4 have been reported. This wide variation may be attributed in part to isoforms of the inhibitor. Taking into account all the variations, it appears that the mol wt of most KSTIs is approximately 21 kDa. KSTI consists of 181 amino acid residues linked by two disulfide bonds; the reactive site is located at residues 63 and 64, as shown in Figure 1.

A competitive inhibitor, KSTI binds to the reactive site of trypsin, in a similar manner as it does to the substrate protein, causing hydrolysis of the peptide bonds between reactive site residues of the inhibitor or substrate. Inhibitors differ from substrate proteins in that the reactive site residues are held between disulfide bonds. After hydrolysis, the modified inhibitor is held together with the same conformation, because of the disulfide bond. This forms a stable enzyme–inhibitor complex.

Bowman–Birk Soy Bean Proteinase Inhibitor (BBI)

The second proteinase inhibitor to be isolated and characterized, BBI, differs from KSTI in four ways. It is a relatively small molecule of molecular weight 7–9 kDa, and is a 'double-headed' inhibitor with independent binding sites for chymotrypsin and trypsin. As it contains seven disulfide bonds, BBI is rich in cysteine (20%). There have been conflicting reports about the stability of this inhibitor. Bowman and Birk commented that BBI displayed marked stability towards heat, acid, alkali, and proteolytic enzymes such as pepsin. It was initially thought that increased heat stability was attributed to the stabilizing effect of the disulfide bonds on the whole structure (Figure 2), thus BBI would be more stable than KSTI. However, some chymotrypsin inhibitors from potato and barley, which display thermostability, contain no or only one disulfide bridge. Other reports found BBI to be more heat-labile than the Kunitz inhibitor. This might be attributed to the many disulfide bonds, which do not give BBI stability but rather make the inhibitor susceptible to bind with other sulfhydryl (-SH) groups other than those of the inhibitor. The heat stability of BBI, however, has been shown to be dependent on concentration.

Limited proteolysis results in separation of the inhibitor into two active fragments, one with trypsin inhibitor activity and the other with chymotrypsin inhibitor activity. Further study of protease inhibitors

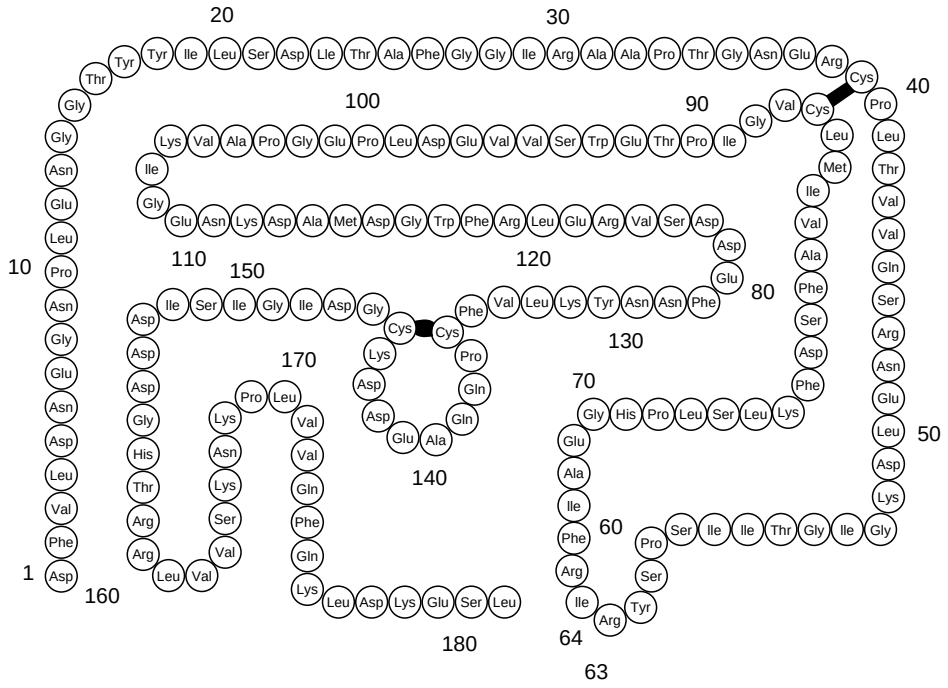

Figure 1 Structure of the Kunitz soy bean trypsin inhibitor. From Koide T and Ikenaka T (1973) Studies on soybean trypsin inhibitors: 3. Amino acid sequence of the carboxyl-terminal region and the complete amino-acid sequence of soybean trypsin inhibitor (Kunitz). *European Journal of Biochemistry* 32: 417–431, with permission.

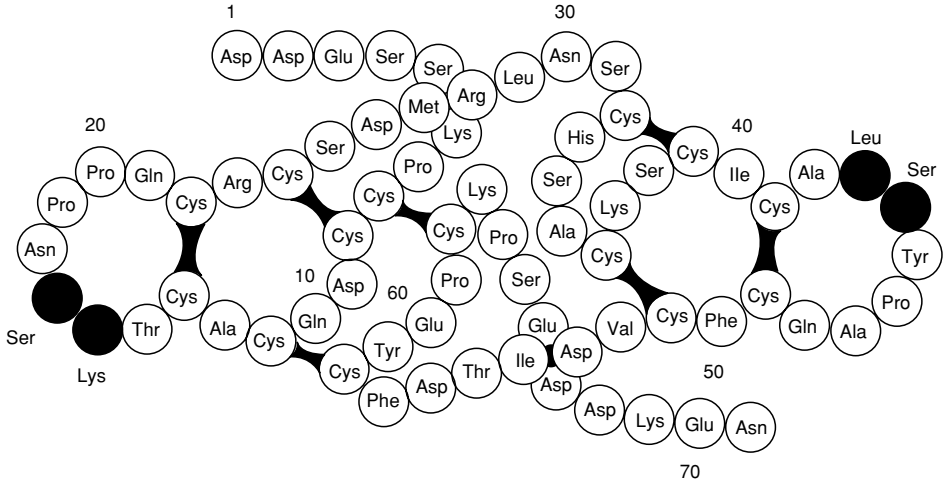

Figure 2 Structure of the Bowman–Birk inhibitor. From Odani S and Ikenaka T (1973) Studies on soybean trypsin inhibitors: VIII. Disulfide bridges in soybean. Bowman–Birk protease inhibitors. *Journal of Biochemistry* 74: 697–715 with permission.

has revealed that the BBI is a prototype for a whole family of homologous inhibitors found in grain legumes and in some other plant families. The sequences of amino acids surrounding the two reactive sites of the BBI are remarkably similar, not only to each other but also to homologous active sites of legume inhibitors isolated from other common grain legumes. Comparison of the BBI with inhibitors isolated from pea, groundnut, and chickpea show that these are double-headed inhibitors, although the mechanism by which they inhibit trypsin may differ slightly owing to the spatial arrangement of

the reactive sites. As many as five protease inhibitors that have properties similar to the BBI may be present in soy beans.

Determination of Activity

The basic method of quantifying trypsin inhibitors has been improved to increase its accuracy and sensitivity and has been further simplified and miniaturized. Many modifications to the basic method have been made: the most common are the use of different substrates, e.g., casein instead of the synthetic substrate benzoyl-DL-arginine-*p*-nitroanilide, or the use of other types of trypsin, e.g., porcine rather than bovine. These modifications make it difficult to compare the results published by different authors. Assessing the variations in inhibitor activity between legumes is made very difficult by the fact that trypsin inhibitor activity (TIA) is not determined using a standard procedure.

The trypsin inhibitor affinity assay only determines protein-type inhibitors forming complexes with trypsin, but is highly sensitive at low concentrations.

Immunological methods developed for the detection and characterization of specific types of protease inhibitors involve the use of specific protease antibodies. Recently, a specific immuno-(chymo)trypsin inhibitor assay has been developed, which discriminates between the two main protease inhibitors in soybeans, KSTI and BBI, unlike the Kakade enzymic assay which does not.

Several different units have been used when measuring TIA, which also make comparisons between data difficult. TIA is most commonly expressed as trypsin inhibitor units (TIU) per milligram of sample ($TIU\,mg^{-1}$ DM) where one TIU is defined as a decrease of 0.01 absorbance units at 410 nm per 10 ml assay solution under specified assay conditions. These units make it possible to measure protease inhibitors of any material without having to find a suitable trypsin inhibitor standard.

Interfering Substances

With samples which are low in TIA, the accuracy of the values obtained can be distorted by nonspecific interference from other proteins and compounds in the reaction mixture. Interfering substances include tannins in colored legume seed coats; free fatty acids released in fermented products such as miso and tempeh and indigestible polysaccharides can complex with trypsin and contribute to the total TIA under assay conditions.

However, the presence of tannins and TIA in field peas was found to be significantly negatively correlated, indicating that tannins were unlikely to be responsible for the additional inhibition of trypsin in the Kakade assay. Maple and wrinkled peas have high tannin levels but low TIA, irrespective of testa or flower color. It would seem that those cultivars exhibiting high tannin content have low levels of TIA or vice versa, possibly to compensate for the other compound, as both of these antinutritional factors have been linked with disease resistance in plants.

Distribution within Plants

Trypsin inhibitors are found most often in the seed but their location is not necessarily restricted to this part of the plant. In some legumes, such as the mung bean and the field bean, high levels of TIA are found in the leaves as well. They are present in many legumes in varying amounts. They are also found in the leaves and tubers of tuberous plants, such as potatoes (*Solanum tuberosum*) and sweet potatoes (*Ipomoea batatas*).

A higher level of TIA has been found in the outer part of the cotyledon of soy beans, kidney beans, and chickpeas, and in five different cultivars of peas four times as much TIA was found in the cotyledon than in the seed coat. Similar findings have been reported for soy beans. In contrast, *Vicia faba* (faba bean) seeds were found to have twice as much TIA in the hull than the cotyledon.

Levels in Grain Legumes

Trypsin inhibitors are widely distributed across many genera and species in the Leguminoseae family and many other plant families; TIA has also been found in a range of legumes, including red gram, kidney beans, navy beans, black-eyed peas, peanuts, field beans, French beans, and sweet peas, and in all varieties tested of cowpeas, mung beans, lima beans, winged beans, chickpeas, and rice beans. In addition, TIA was found in lentils, but there was almost no TIA in lupins.

Although trypsin inhibitors are found in most legumes, the levels present tend to vary considerably. Most legume species contain less than 50% of the TIA of soy beans. Particularly low levels are present in broad beans, peas, mung beans, lupins, and a few varieties of kidney beans. Those species with at least 75% of the TIA in soy beans include cowpeas, pinto beans, pigeon peas, kidney beans, moth beans, and navy beans. Common legumes that contain levels higher than those of soy bean include lima beans, winged beans, and black beans.

Differences in TIA between cultivars have been observed, suggesting the possibility of breeding for low TIA. For example, it was found that winter cultivars of peas were twice as active as summer cultivars and that wrinkled cultivars were more active than smooth cultivars. Considerable variability has also been reported amongst different strains of cowpeas.

Stage of maturity has also been shown to influence TIA. As soy beans mature, the amount of TIA increases, although the magnitude of increase differs between varieties. The germination of a number of grain legumes appears to increase their nutritive value, although the effect on TIA seems to be quite variable. There have been no reports of a significant decrease in TIA on germination of lentils, chickpeas, and navy beans but there was a decrease in TIA when field peas were germinated. Other studies have found that after germination of *Vicia faba* for 5 days TIA decreased by 63%; TIA decreased by 50% in red kidney beans after 10 days' germination, and TIA in three soy bean varieties decreased by an average of 13% over 3 days' germination. The loss of TIA in some cases may be accounted for by leaching during soaking and washing procedures.

Effect of Processing

Processing can generally be divided into methods that are commonly used to prepare legumes for human consumption, and those that are used in the production of animal meals. The popularity of grain legumes in human diets is no doubt attributable to the relative ease with which TIA can be destroyed by many of these methods. However, destruction of TIA inevitably results in loss of some of the nutritive value of the legume.

Boiling

Cooking presoaked (24-h) winged beans in boiling water for 30 min was found to destroy most TIA. It was also reported that mature soy beans required a presoak of 24 h and cooking for at least 20 min to eliminate TIA completely. Immature seeds do not require presoaking to eliminate TIA. This, and the fact that oven heating has very little effect on TIA, suggests that the moisture content of the seed plays an important role in the destruction of trypsin inhibitors. Soaking chickpeas, broad beans, and mung beans alone reduced TIA levels by 58–92%. Combined soaking (18 h) and boiling (20 min) led to a 9.9–56.7% reduction in peas, depending on cultivar.

It has been suggested that the extent to which TIA is destroyed by heating is a function of the temperature, duration of heating, particle size, and moisture content – variables that are closely controlled in the industrial processing of soy bean meals in order to obtain a product maximum nutritional value. The nonprotein TIA content would also determine the reduction in TIA achieved by the heating process. Researchers have also found that the percentage of TIA remaining in the cooked peas was found to be negatively correlated to increased TIA levels in the raw form, therefore it appears that the effect of cooking is more effective in destroying TIA when the original inhibitory activity (raw seed) was higher than when it was low. A summary of the effects of various heat treatments on TIA is given in Table 1.

Table 1 Effect (percentages of original activity) of various types of heat treatment on the trypsin inhibitor activity (TIA) of legumes

Legume	*TIA*		*Wet heat*		*Autoclaved extract: 103 kPa for 15 min*	*Dry heat*			
	(TI units per mg)	*(%: soy TIA = 100%)*	*Heating extract in boiling-water bath*			*Roasting for 15 min at:*			*Roasting for 2 min at 200 °C*
			30 min	*60 min*		*75 °C*	*100 °C*	*125 °C*	
Moth bean (*Phaseolus aconitifolus*)	1.44	27	57.8	11.6	2.4	80.8	42.5	Nil	7.6
Cow pea *(Vigna catjung)*	4.17	79	59.9	49.3	4.8	72.2	72.2	65.0	Nil
Cow pea *(Vigna sinensis)*	3.39	64	96.7	36.0	3.7	94.2	91.2	71.1	8.1
Red gram *(Cajanus cajan)*	3.19	60	71.5	64.5	12.3	80.5	54.5	17.9	11.7
French bean *(Phaseolus vulgaris)*	4.24	80	61.6	43.2	6.2	76.1	73.5	72.0	10.6
Pea *(Pisum sativum)*	1.33	25	20.8	12.5	Nil	42.1	27.6	4.3	Nil
Lentil *(Lens culinaris)*	1.31	25	8.5	Nil	Nil	73.1	69.4	65.2	20.4
Green gram *(Phaseolus aureus)*	1.98	37	11.2	5.6	Nil	97.4	69.5	24.7	Nil
Black gram *(Phaseolus mungo)*	2.74	52	28.4	16.2	Nil	69.4	57.6	39.4	11.9
Chickpea *(Cicer arietinum)*	3.47	66	16.0	3.2	Nil	62.0	40.9	14.2	7.6
Soy bean *(Glycine max)*	5.30	100	61.9	22.1		58.9			

TI, trypsin inhibitor.

From Rackis JJ, Wolf WJ and Baker EC (1986) Protease inhibitors in plant foods: content and inactivation. In: Friedman M (ed.) *Advances in Experimental Medicine and Biology: Nutritional and Toxicological Significance of Enzyme Inhibitors in Foods*, pp. 299–337. New York: Plenum Press, with permission.

Other Cooking Methods

Heating at temperatures below 75 °C appears to have no effect on winged beans. Increasing the temperature to 100 °C resulted in a 5% decrease in TIA of winged beans. By increasing the temperature to 80 °C, it was found that TIA in white winged beans decreased by 25%, in 5 min and 45% after 30 min. Dry heating (177 °C for 20 min) peanuts and soy beans decreased TIA by 7% and 20% respectively.

Autoclaving has been shown to decrease TIA significantly. In peanuts and soy beans autoclaved at 121 °C for 20 min TIA decreased by 80% and 86% respectively Autoclaving whole winged beans and winged bean meal decreased TIA by 80%.

Infrared treatment for 30 s on winged beans was reported to have destroyed most TIA in the seeds.

Microwave radiation of broad beans for 30 min was found to destroy 90% of their TIA. It was found that microwave treatment (time not known) had no effect on TIA in winged beans. Microwave treatment for 1.5–3.0 min of presoaked soy beans destroyed 85–90% of the TIA.

Physiological Effects

The majority of experiments which have considered the physiological effects of grain legume TIA have been carried out on animals and inferences made about the effects on humans. The effects of TIA found in soy beans have been predominantly studied as the availability of pure forms of KSTI and BBI have made the task of elucidating the inhibitor actions much easier. The efficient digestion of proteins in the digestive system requires the action of a number of proteolytic enzymes, each hydrolyzing peptide chains at specific points. These enzymes are produced and stored in the pancreas in a precursor form known as zymogens. The efficient digestion of proteins requires simultaneous activation of all the zymogens. Trypsin is thought to act as the common activator of pancreatic zymogens.

Proteolytic Activity and Growth

Because trypsin activates proteolytic enzymes, an overall decrease in proteolytic activity would be expected from the action of trypsin inhibitors. *In vitro* studies have shown that KSTI, BBI, and lima bean inhibitor completely inhibited trypsin and chymotrypsin activity of human and rat pancreatic juice. The TIA caused up to 50–60% reductions in total proteolytic activity, and the residual activity observed was attributed to carboxypeptidase activity.

Reduction in growth owing to loss of proteolytic activity appears to vary between species fed raw soy bean diets. Growth was reduced in the rat, mouse, chick, and young guinea pig, whereas the adult guinea pig, dog, pig, calf, monkey and, probably, the human appeared to grow normally when fed these diets.

There are a number of factors that might contribute to the differences in proteolytic activity and growth between different species. One is that gastric juices can inactivate protease inhibitors. In one study human gastric juice almost eliminated the protease inhibitor activity of KSTI when incubated for 24 h. It was found that lima bean inhibitor activity was only slightly affected by incubation in gastric juice, suggesting that inhibitors homologous to lima bean inhibitor are nutritionally more important than inhibitors from the KSTI family.

Another reason for the varying activity of protease inhibitors between species may be differences in the specific activity of species-specific trypsin. In humans, proteolytic enzymes hydrolyze casein at a lower rate than in many animals, and the effect of inhibitors may therefore be weaker in humans. Soy bean trypsin inhibitor was shown to have a much greater effect on trout trypsin than bovine trypsin, probably owing to a higher specific activity of trout trypsin.

Human trypsin is known to exist in two forms. The cationic form constitutes two-thirds of the total trypsin secreted and is only very weakly inhibited, whereas the anionic form is inhibited stoichiometrically. Infusion of raw soy bean extract into the duodenum of rats and humans caused the secretion of a modified inhibitor-resistant trypsin, which was resistant to typical serine protease inhibitors. The BBI was more potent than KSTI in producing TIA-resistant enzymes. Whether or not this was just the cationic form of trypsin is unclear. The ability of the pancreas to adapt to trypsin inhibitors may determine the effects that they have on different species.

Toxicological Effects

The main toxicological effect observed from the ingestion of trypsin inhibitors is a marked hypertrophy of the pancreas. Pancreatic hypertrophy is thought to be caused by TIA stimulating excess enzyme secretion, which draws essential amino acids away from other body functions; in some cases this results in death of the animal.

Negative feedback control of pancreatic secretion is mediated by cholecystokinin (CCK). CCK is released from the intestinal mucosa when levels of trypsin or chymotrypsin are lowered, such as in the formation of the trypsin–trypsin inhibitor complex. The action of this complex is mediated by a

trypsin-sensitive CCK-releasing peptide, which interacts with the luminal surface of the small intestine to stimulate the release of CCK into the circulation. CCK stimulates the secretion of the pancreozymins, trypsinogen and chymotrypsinogen, the precursors to trypsin and chymotrypsin respectively. Recently, complexed as well as free inhibitors have been found to stimulate the discharge of CCK in rats and as a result have significantly increased the trypsinogen and chymotrypsinogen secretion rates. In normal circumstances, when trypsin levels are adequate, trypsin exerts a negative effect on this monitor peptide in the gastrointestinal tract suppressing its pancreatic secretion rate.

The pancreas may therefore function abnormally when these inhibitors are present as it produces more enzymes to compensate for the loss of the enzyme via the enzyme–inhibitor complex. Dietary amino acids may be directed from other protein synthesis sites in the body tissues for the synthesis of these additional pancreatic enzymes. As pancreatic enzymes are rich in the essential sulfur-containing amino acids, removal of these amino acids from other sites of protein synthesis involved in growth and maintenance is critical. Continued ingestion of legumes would compound these effects and, in small animals such as rats, have been found to result in pancreatic hypertrophy, growth depression, and failure to thrive.

This negative-feedback control is known to occur in the rat, pig, and calf, as well as in humans, suggesting that other factors are also involved in pancreatic hypertrophy.

See also: **Plant Antinutritional Factors**: Characteristics; **Fermented Foods**: Fermentations of the Far East; Soy (Soya) Sauce; **Legumes**: Legumes in the Diet; Dietary Importance; **Peanuts**; **Peas and Lentils**; **Protein**: Interactions and Reactions Involved in Food Processing; Digestion and Absorption of Protein and Nitrogen Balance; Heat Treatment for Food Proteins; **Pulses**; **Soy (Soya) Beans**: Properties and Analysis; Dietary Importance; **Tannins and Polyphenols**; **Toxins in Food – Naturally Occurring**

Further Reading

Birk Y (1961) Purification and some properties of a highly active inhibitor of trypsin and α-chymotrypsin from soybeans. *Biochimica et Biophysica Acta* 54: 378–381.

Bowman DE (1946) Fractions derived from soy beans and navy beans which retard tryptic digestion of casein. *Proceedings of the Society of Experimental Biology and Medicine* 63: 547–550.

Brandon DL, Bates AH and Friedman M (1988) Enzyme-linked immunoassay for soybean Kunitz trypsin inhibitor using monoclonal antibodies. *Journal of Food Science* 53: 102–106.

Gaborit T, Quillien L and Guéguen J (1993) Determination of trypsin inhibitor activity in seeds by a microtitre plate method. In: van der Poel AFB, Huisman J and Saini HS (eds) *Recent Advances of Research in Antinutritional Factors in Legume Seeds*, pp. 37–40. Wageningen: Wageningen Pers.

Kadam SS and Smithard RR (1987) Effects of heat treatments on trypsin inhibitors and haemagglutinating-activities in winged beans. *Plant Foods for Human Nutrition* 37: 151–159.

Kakade ML, Rackis JJ, McGhee JE and Puski G (1974) Determination of trypsin inhibitor activity of soy products: a collaborative analysis of an improved procedure. *Cereal Chemistry* 51: 376–382.

Koide T and Ikenaka T (1973) Studies on soybean trypsin inhibitors: 3. Amino acid sequence of the carboxyl-terminal region and the complete amino-acid sequence of soybean trypsin inhibitor (Kunitz). *European Journal of Biochemistry* 32: 417–431.

Krogdahl A and Holm H (1981) Soybean proteinase inhibitors and human proteolytic enzymes: selective inactivation of inhibitors by treatment with human gastric juice. *Journal of Nutrition* 111: 2045–2051.

Liener IF (1980) Protease inhibitors. In: Liener IE (ed.) *Toxic Constituents of Plant Foodstuffs*, pp. 7–57. New York: Academic Press.

Noonan SC (1999) The nature and stability of trypsin inhibitor isoforms in New Zealand grown pea (*Pisum sativum* L.) cultivars. MSc thesis. New Zealand: Lincoln University.

Odani S and Ikenaka T (1973) Studies on soybean trypsin inhibitors: VIII. Disulfide bridges in soyabean. Bowman Birk proteinase inhibitors. *Journal of Biochemistry* 74: 697–715.

Rackis JJ, Wolf WJ and Baker EC (1986) Protease inhibitors in plant foods: content and inactivation. In: Friedman M (ed.) *Advances in Experimental Medicine and Biology: Nutritional and Toxicological Significance of Enzyme Inhibitors in Foods*, pp. 299–337. New York: Plenum Press.

Roozen JP and de Groot J (1988) Analysis of residual inhibitor in feed flour. In: Huisman J, van der Poel TFB and Liener IE (eds) *Recent Advances of Research in Antinutritional Factors in Legume Seeds*, pp. 114–117. Wageningen: Pudoc.

Savage GP (1988) The composition and nutritive value of lentils *(Lens culinaris)*. *Nutrition Abstracts and Reviews (Series A)* 58: 319–343.

Valdebouze P, Bergeron E, Gaborit T and Delort-Laval J (1980) Content and distribution of trypsin inhibitors in some legume seeds. *Canadian Journal of Plant Science* 60: 695–701.

van Amerongen A, Ostafe V, Meijer MMT *et al.* (1998) Specific-immuno-(chymo)trypsin-inhibitor assays for determination of (residual) activity of Bowman–Birk or Kunitz soybean trypsin inhibitors. In: Jansman AJM, Hill GD, Huisman J and van der Poel AFB (eds) *Recent Advances of Research in Antinutritional Factors in Legume Seeds and Rapeseed*, pp. 33–37. Wageningen: Wageningen Pers.

van Oort MG, Hamer RJ and Slager EA (1988) The trypsin inhibitor assay: improvement of an existing method. In: Huisman J, van der Poel TFB and Liener IE (eds) *Recent Advances of Research in Antinutritional Factors in Legume Seeds*, pp. 110–113. Wageningen: Pudoc.

Tuna *See* **Fish**: Introduction; Catching and Handling; Fish as Food; Demersal Species of Temperate Climates; Pelagic Species of Temperate Climates; Tuna and Tuna-like Fish of Tropical Climates; Demersal Species of Tropical Climates; Pelagic Species of Tropical Climates; Important Elasmobranch Species; Processing; Miscellaneous Fish Products; Spoilage of Seafood; Dietary Importance of Fish and Shellfish; **Fish Farming**; **Fish Meal**

Turkey *See* **Poultry**: Chicken; Ducks and Geese; Turkey

U

UHT *See* **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; Chemical and Microbiological Changes; Electrical Process Heating

Ultrafiltration *See* **Membrane Techniques**: Principles of Reverse Osmosis; Applications of Reverse Osmosis; Principles of Ultrafiltration; Applications of Ultrafiltration

ULTRAVIOLET LIGHT

F Thompson, Halma PLC, London, UK

Introduction

Since the turn of the century, when its existence was proved experimentally, ultraviolet (UV) light has been used to reduce microbiological contamination in drinking water, process water, fluids, effluent, and air, and on surfaces. The first municipal UV water disinfection system was built in 1910 in Marseilles, and since then, the technology has been refined and developed, providing today's reliable, computer-controlled systems.

This article reviews the technology of UV disinfection and outlines its industrial applications, especially in food and drink manufacturing.

The Nature of UV Light

What is UV?

The term 'UV' is applied to wavelengths of radiation found in the electromagnetic spectrum between visible light and X-rays (see Figure 1), from 100 to 400 nm. The region between 190 and 400 nm is divided into three sections:

1. U1V-A, 320–400 nm. The longer wavelength limit corresponds roughly to the start of visible light, whereas the lower limit corresponds to the point at which proteins and genetic material begin to absorb significant amounts of UV energy.
2. UV-B, 290–320 nm. Ultraviolet light of these wavelengths begins to damage biological molecules and is the region of the spectrum responsible for sunburn and skin cancers.
3. UV-C, 200–290 nm. Far UV wavelengths are highly destructive to biological molecules. In nature, these wavelengths in sunlight are almost totally absorbed by the ozone layer. UV-C wavelengths are mainly responsible for the germicidal properties of UV light.

Commercial UV disinfection systems generate a high proportion of UV-C, particularly at 254 nm, which is very close to the most biologically damaging wavelength of 260 nm (see Figure 2).

How UV Works

In sufficient doses, UV light of the wavelengths mentioned is lethal to all microorganisms, including bacteria, viruses, protozoa, molds, yeasts, fungal spores, and algae.

The effectiveness of UV results from its ability to cause molecular rearrangements in the genetic

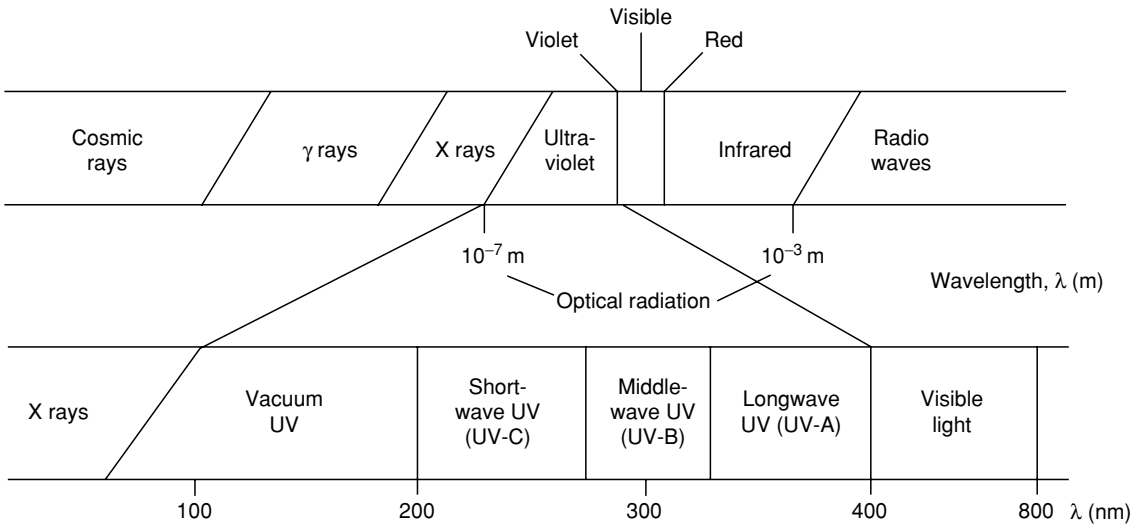

Figure 1 Electromagnetic spectrum with expanded scale of UV radiation. Reproduced from Ultraviolet Light, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

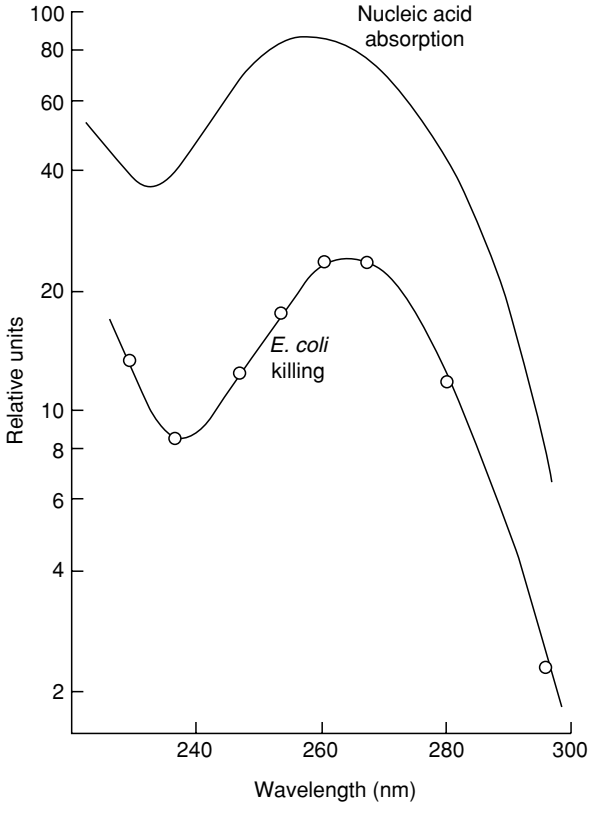

Figure 2 Similarity of the action spectrum for inactivation of *Escherichia coli* cells, determined by FL Gates, to the absorption spectrum of nucleic acids. Redrawn from Rupert CS (1960) In: Burton M, Kirby-Smith JS and Magee JL (eds) *Comparative Effects of Radiation*, pp. 49–61. New York: John Wiley with permission.

material of microorganisms. This cripples the organism's normal metabolism and prevents it from reproducing. If a microorganism cannot reproduce, it is considered dead.

The detailed mechanisms by which UV deactivates microorganisms have only been fully understood recently through advances in microbiological techniques.

Deoxyribonucleic acid (DNA) is a helical macromolecule responsible for the storage, expression and replication of a cell's genetic information. (In some cases, especially viruses, ribonucleic acid (RNA) is the main carrier of genetic information, but as the principle of UV damage is the same, we will limit the description to DNA). Molecules of DNA are particularly susceptible to damage by UV-C energy, and are the primary site of absorption of UV photons.

Deoxyribonucleic acid is composed of four types of deoxyribonucleotides, which differ only in their base components. These are adenine, guanine, cytosine, and thymine (which is replaced by uracil in RNA). The main photochemical change to occur when DNA is exposed to UV light is the dimerization of thymine (and, less often, cytosine), where two of these molecules lie adjacent on a DNA strand. The effect of this dimerization is to block the normal replication of DNA, causing an irreversible breakdown in the cell's metabolism and replication.

Quantifying the Effectiveness of UV

The germicidal effectiveness of UV is expressed as the number of organisms that survive after exposure.

Table 1 Dose–effect relationship for *Escherichia coli* (D_{10} = 5.4 mJ cm^{-2} in water)

Dose (mJ cm^{-2})	*Percentage of organisms remaining*	*Log reduction*
5.4	10	Log 1
10.8	1	Log 2
16.2	0.1	Log 3
21.6	0.01	Log 4

This depends on the intensity of UV (measured in mW cm^{-2}), the number of microorganisms present before disinfection, and the duration of exposure to the UV source. These parameters are combined to give the UV dose, the standard indicator of UV effectiveness. In its simplest terms, UV dose is a function of intensity and time and can be illustrated as follows:

$$\text{UV dose} = \text{UV intensity} \times \text{time}$$
$$(\text{mJ cm}^{-2}) = (\text{mW cm}^{-2}) \times (\text{s})$$
$$1\,\text{J} = 1\,\text{W s}.$$

For water disinfection systems, the presence of molecules or suspended solids in the water, which may absorb UV energy, must also be taken into account in any calculations.

The required UV dose differs for each type of microorganism and is measured as a D_{10} value, i.e., the UV dose necessary for 1 log reduction in concentration of any given microorganism. For example, the D_{10} value for *Escherichia coli* in water is 5.4 mJ cm^{-2}. This means that a UV dose of 5.4 mJ cm^{-2} will reduce the population of *E. coli* to 10% of the original number, i.e., a 1 log reduction. A dose of 10.8 mJ cm^{-2} (twice the D_{10}) will give a 2 log reduction, and so on (see Table 1).

Types of UV Systems

Lamp Type

Ultraviolet energy for disinfection is produced using mercury vapor lamps. These are similar in operation and appearance to a fluorescent lamp. Ultraviolet light is emitted as a current flows through vaporized mercury. A protective quartz sleeve, which is transparent to UV light, surrounds the lamp. Two types of UV lamp are available – low-pressure lamps and medium-pressure lamps.

Low-pressure Lamps For low flow rates and surface disinfection applications, low-pressure lamps are most often used. These are highly efficient at producing UV energy but are limited to a relatively low UV output.

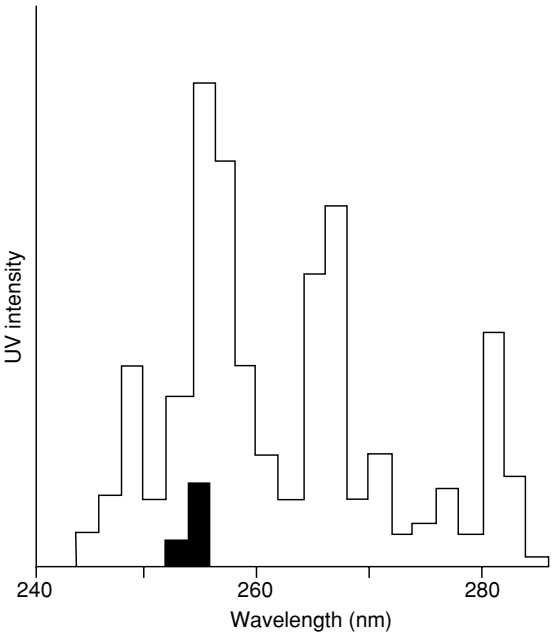

Figure 3 Energy distribution of medium- (□) and low-pressure (■) UV lamps. Reproduced from Ultraviolet Light, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Medium-pressure Lamps Medium-pressure lamps draw a much greater current than low-pressure lamps, leading to a rapid rise in the mercury vapor pressure. This gives a typical diffuse medium-pressure UV spectrum (see Figure 3). Per unit area, the effective germicidal output of a medium-pressure lamp is much greater (up to 30 times) than a low-pressure lamp, allowing high flow rates to be treated using fewer lamps. Medium-pressure lamps are used for most industrial water and fluid disinfection systems owing to their greater compactness and disinfection capability. Unlike low-pressure lamps, which only function efficiently within a limited temperature range, medium-pressure systems are temperature-independent, and therefore suitable for low-temperature and chilled-water applications.

System Design

The following parameters must be considered when specifying a UV disinfection system:

1. The energy output of the UV source.
2. The flow rate of the fluid or air through the treatment chamber, i.e., the time period of exposure to UV.
3. The transmission value, i.e., the ability to transmit UV light, of the fluid or air being treated.
4. The geometry of the irradiation chamber.

By optimizing these criteria, a UV system can be tailored to treat virtually any water- or airflow, or

transport fluid containing high levels of dissolved solids, starch or sugar compounds.

The ideal geometry of a fluid disinfection system for maximum utilization of the UV energy is a single lamp running along the axis of a cylindrical treatment chamber.

Practical Applications of UV

Microbiological contamination can occur at any stage of a production process from the incoming process water to the surfaces of packaging. In many cases, UV disinfection can offer a safe and effective method of providing microbe-free process water, fluids and air. It is effective against food pathogens, including viruses and fungal spores (see Table 2).

Unlike chemical biocides, UV does not introduce toxins or residues into process water and does not alter the chemical composition, taste, odor, or pH of the fluid being disinfected. This feature is especially important in food and drink processing plants, where the chemical dosing of incoming process water can cause off-flavors and alter the chemical properties of the product.

Ultraviolet treatment systems can be used as the primary disinfection system or as a back-up for other water purification methods, such as activated carbon filtration, reverse osmosis or pasteurization. As UV has no residual effect, the best position for a treatment system is immediately prior to the point of use. This ensures that incoming microbiological contaminants are deactivated and there is minimum chance of posttreatment contamination. Many users install UV systems after filter beds and storage tank outlet valves, to reduce the likelihood of contamination from these sources.

Ultraviolet systems can also be installed as an alternative to, or in conjunction with, air filtration systems. The UV source will deactivate any microorganisms present and prevent colonization of air-filtration systems.

Table 2 Experimentally determined UV dose, at 254 nm, required for a 90% kill rate (D_{10}) of various organisms

Organism	D_{10} *(mJ cm*$^{-2}$*)*
Bacteria	
Escherichia coli	5.4
Bacillus subtilis	7.1
Clostridium botulinum	12.0
Micrococcus candidus	6.05
Pseudomonas aeruginosa	5.5
Mold spores	
Aspergillus niger	100
Cladosporium herbarum	30–70
Mucor mucedo	50–70
Penicillium roqueforti	13
Yeasts	
Saccharomyces cerevisiae	6.0

Clean-in-place (CIP) Rinse Systems

It is essential that CIP final rinse water, which is used to flush out foreign matter and disinfecting solutions, is microbiologically safe. Fully automated UV disinfection systems can be integrated with CIP rinse cycles to insure that final rinse water does not reintroduce microbiological contaminants. Although most town supplies are free from coliforms, they are rarely sterile. Resistant to the effects of acid, detergents, steam and chemical sterilants, UV systems are an effective method of providing disinfected water for final rinse systems.

Sugar Syrups

Specially modified UV treatment chambers are routinely used in the brewing and soft drinks industry to disinfect concentrated sugar solutions and syrup. Although high-Brix syrups will not support microbial growth, any dormant spores present may become active after the syrup has been diluted. Treating the syrup and dilution water with UV prior to use will insure that any dormant microorganisms are deactivated.

Systems for syrup treatment are designed with a chamber of relatively small diameter. This helps to insure that sufficient UV light penetrates the liquid and deactivates spores, which generally require higher doses of UV than other microorganisms for complete deactivation. In food and dairy processing plants, the same equipment can be used to disinfect injection brine and cheese whey.

Air Disinfection

For machine blanketing, and in sanitary and food preparation areas, the exclusion of airborne contaminants, such as bacterial spores, yeasts and viruses, can be a major problem. Ultraviolet disinfection systems are now available for treating airflows entering a sterile environment. Air-disinfection systems are fitted into ductwork, and any microorganisms present are deactivated as they are exposed to the UV source. (*See* **Plant Design**: Designing for Hygienic Operation.)

Ultraviolet systems can also be used to disinfect displacement air for pressurizing tanks, or pipelines holding perishable fluids or free-flowing solids. Storage tanks are particularly susceptible to bacterial colonization and contamination by airborne spores. To prevent this, immersible UV treatment systems have been designed to fit in the tank head airspace and disinfect the air present.

Packaging Applications

Another aspect of contamination control that can benefit from UV technology is the disinfection of finished products and their packaging. A typical system incorporates a UV light source, shielded on three sides and mounted over a conveyor or production line. As foodstuffs or packaging pass under the light, any microbes present on the surface are deactivated.

After treatment by UV, the risk of microbial contamination in filling and packaging lines is reduced, and the shelf-life of treated products is extended. This surface sterilization equipment can also be used to disinfect bottle crowns, foil lids and plastic films. (*See* **Packaging**: Packaging of Solids.)

Recent Advances in UV

Modern UV treatment systems are microprocessor-controlled with automatic UV intensity and lamp status monitors. This facilitates the interfacing of UV systems with other process control systems to coordinate their operation.

Integral Wiping Systems

The development of integral wiping systems has increased the applications of UV to include treatment of turbid fluids containing processing residues or high solids concentrations. The wiping system is fitted to the outside of the quartz sleeve. At preset intervals during operation, a mechanical wiper moves along the lamp surface, displacing any processing residues. These would quickly foul conventional UV systems, inhibiting the transmission of UV light and reducing germicidal effectiveness. Wiped systems can constantly reduce contamination to acceptable levels in poultry, egg and meat processing, in the transport and cooling of vegetables, and in the disinfection of effluent flows. (*See* **Effluents from Food Processing**: Disposal of Waste Water.)

Lamp Output

High-output, single-lamp chambers emit the high energy levels required to penetrate the turbid waters that can be found in processing plants. This includes virtually opaque fluids, such as transport water for flour- and pasta-based products, and industrial effluent or recirculation water. Currently, the minimum single lamp power available is 8 W (low-pressure), increasing to a maximum of 5 kW (medium-pressure).

Limitations of UV Technology

System Sizing

In theory, UV systems can be built to treat any flow rate, air volume, or surface area. As UV demand increases in terms of size, degree of contamination and quality of water, the number and power output of the UV lamps are increased proportionally to maintain UV dose above the required minimum.

Obviously, practical constraints, such as the size of the UV system, pressure drop through the system, and power consumption, vary with each installation. Single-chamber UV systems treating up to 250 $m^3 h^{-1}$ are often more than adequate for process-water applications. Duty stand-by units, programmed to come on-line when flow exceeds preset levels, can also be installed as a back-up measure.

Ultraviolet systems have been successfully installed to treat water flows over 1×10^5 m^3 per day. At these flow rates, medium-pressure systems have the advantage of fewer lamps, compared with low-pressure systems, and therefore have lower maintenance requirements.

Residual Disinfection

The only potential drawback with UV water disinfection, compared with chemical treatment, is that UV does not leave a residual disinfectant effect. In food and drink manufacturing, this is a positive advantage as the presence of chlorine could adversely affect product quality. In industrial applications, the lack of residual is rarely a problem, provided that downstream pipework is maintained in a hygienic condition, and no 'dead legs' or leaks are allowed where recolonization could occur.

Ultraviolet systems are recommended for installation as close to the point-of-use as possible to reduce the likelihood of recontamination. It is mainly in the drinking-water treatment industry, where there may be miles of distribution pipework following disinfection, that a small chlorine residual is often added to the water after UV disinfection. A combination of UV disinfection and electrochlorination is most common, both systems requiring little maintenance and able to operate almost unsupervised.

With air and surface disinfection, there is a remote risk that an operator may be exposed to the UV light source. Prolonged exposure to high-intensity UV light an damage the eyes, but this kind of accident is very rare and can be avoided by carefully designed shielding around the UV source. In addition, access to the UV source can be interlocked with the power supply, ensuring that access can be obtained only when the power supply has been turned off.

Conclusion

The unique ability of UV light to deactivate microorganisms in water and air, and on surfaces, without creating by-products or residual effects, enables

UV disinfection to be used for a wide variety of applications. Whether installed as the main disinfection system, as a back-up to alternative methods, or in specific clean-room environments where high standards of sterility must be maintained, UV treatment systems can ensure that levels of microbial contaminants are effectively controlled.

See also: **Effluents from Food Processing**: Disposal of Waste Water; **Packaging**: Packaging of Solids; **Plant Design**: Designing for Hygienic Operation

Further Reading

Anonymous (1988) High UV output disinfects high solids process water. *Prepared Foods* 157: 114.

Mans J (1987) Disinfect water with ultraviolet light. *Prepared Foods* 156: 72–74.

Maunder DT (1977) Possible use of ultraviolet sterilisation of containers for aseptic packaging. *Food Technology* 31: 36–370.

Phillips R (1983) *Sources and Applications of Ultraviolet Radiation*. London: Academic Press.

United States Food and Drug Administration *See* **Food and Drug Administration**

URONIC ACIDS

F G Huffman, Florida International University, Miami, FL, USA

Chemistry

Uronic acids are produced by the oxidation of the alcohol group of monosaccharides. These compounds are named by substituting -ose with uronic acid. The structures of the most common uronic acids are shown in **Figure 1**. D-Mannuronic acid is the 2-epimer of D-glucuronic acid, and L-iduronic acid is the 5-epimer of D-glucuronic acid.

The uronic acids are important constituents of certain natural heteropolysaccharides. They play a significant role in the detoxification of substances such as drugs. Glucuronic acids are found in human urine bound with glycosidic linkages to hydroxylated compounds such as menthol, borneol, and estrogens. Owing to the increased solubility of hydroxylated compounds (when conjugated with glucuronic acid), they are readily disposed by the body. Glucuronic acid also forms a conjugate with bilirubin, a bile pigment.

The most important uronic acid in humans is the D-glucuronic acid. It, or its epimer, is a constituent of many glycosaminoglycans (GAG) as well as glucuronide derivatives of drugs and hormones. D-Glucuronic acid is also the precursor of L-ascorbic acid in animals. The schematic utilization of D-glucuronate in microorganisms, animals, and plants is shown in **Figure 2**. Glucuronic acid is synthesized from glucose in the uronic pathway, an alternative oxidative pathway for glucose without the production of adenosine triphosphate (ATP). In the uronic pathway, glucose 6-phosphate is converted to glucose 1-phosphate which subsequently reacts with uridine triphosphate to form uridine diphosphate glucose (UDPGlc). This compound is then oxidized at the six-carbon position in a two-step process by the

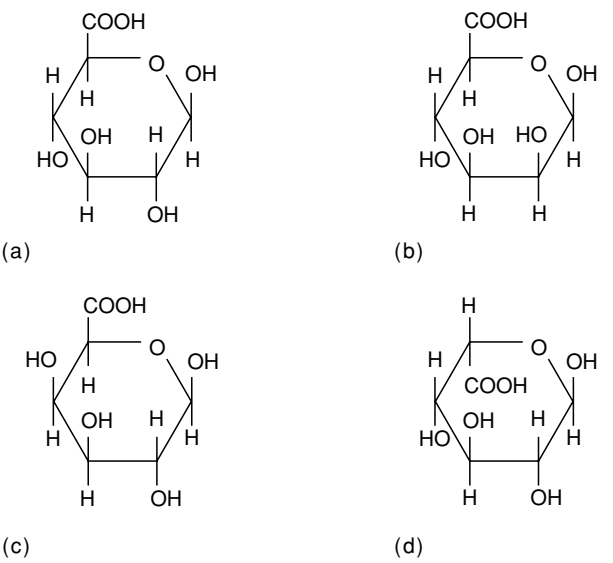

Figure 1 Structures of uronic acids: (a) D-glucuronic acid (Glu U);(b) D-mannuronic acid (Man U); (c) D-galaturonic acid (Gal U); (d) L-iduronic acid (L-guluronic acid; Gul U).

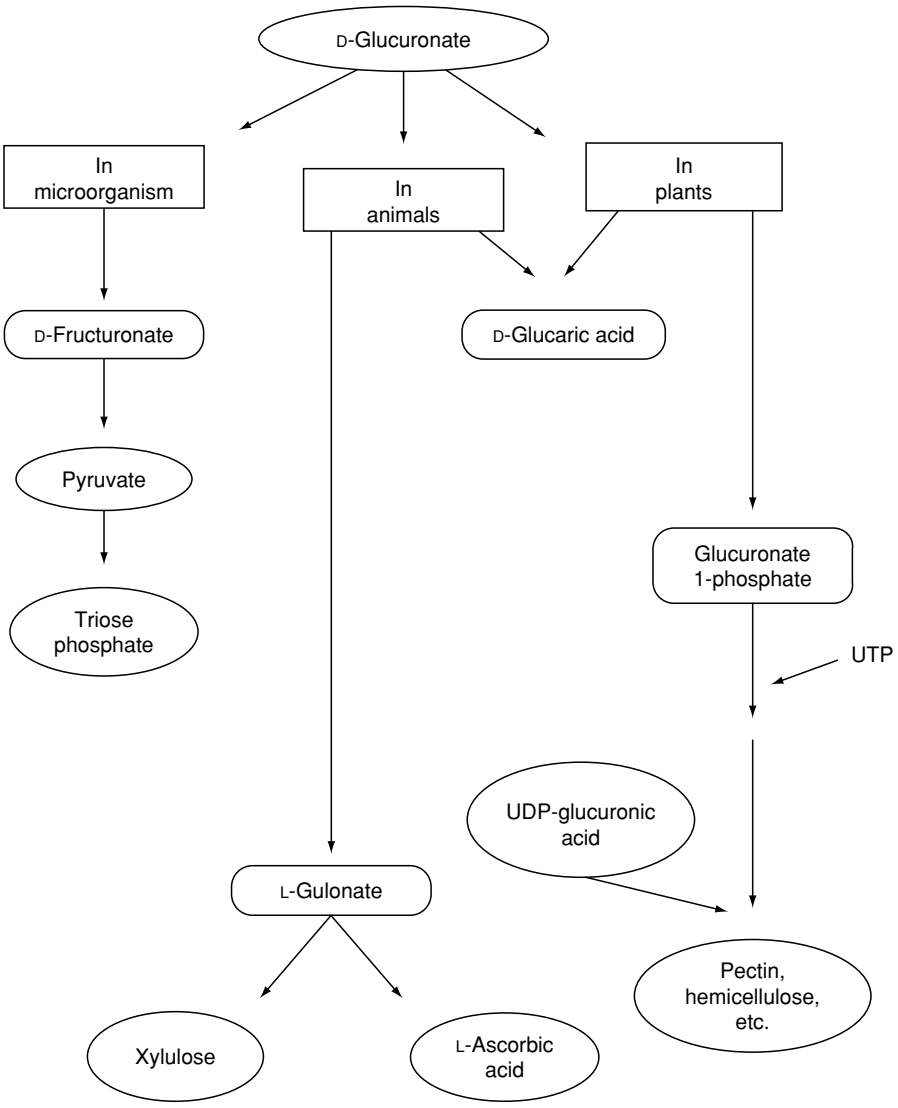

Figure 2 The fate of glucuronate in plants, animals and microorganisms. UTP, uridine triphosphate; UDP, uridine diphosphate.

nicotinamide adenine dinucleotide (NAD)-dependent enzyme (UDPGlc dehydrogenase) to form UDP-glucuronate.

UDP glucuronate is the form of glucuronic acid which can be incorporated into proteoglycans or conjugated with steroid hormones, certain drugs, or bilirubin. In bilirubin, two molecules of glucuronic acid are attached with ester linkages to the two propionic acid groups to form an acylglucuronide. This conjugation process increases the water solubility of bilirubin, thus allowing its secretion into the bile, ultimately to be excreted via the gastrointestinal tract.

A second product of the uronic pathway is L-ascorbic acid, which is produced in mammals, with the exception of humans, primates, and the guinea-pig. In a series of reactions glucuronate is reduced to L-gulonate, which is subsequently converted to L-ascorbic acid. The glucuronic acid pathway is most active in the liver, kidneys, and intestines. The primary function of the pathway is to produce UDP-glucuronic acid, which is needed for detoxification of various compounds by elimination of glucuronides in the urine or bile. Many carcinogens, and drugs, including antipyretics, hypnotics, and antimalarial drugs, may be eliminated by this mechanism. There may be certain organs which may have an active uronic pathway specific for a drug. Glucuronic acid synthesis may be stimulated when the consumption of substances that are excreted as glucuronides is increased, or when steroids and barbiturates, which induce the microsomal P-450 system, are consumed.

Table 1 Composition of glycosaminoglycans

Glycosaminoglycan	*Uronic acid composition*
Hyaluronic acid	β-glucuronic acid
Chondroitin 4-sulfate	β-glucuronic acid
Heparin	Sulfated iduronic acid
Dermatan	L-iduronic acid

Uronic Acids in Animal Tissue

Uronic acids are an integral part of GAGs, formerly known as the mucopolysaccharides. They have structural importance in vertebrate animals. Some important examples of GAGs are hyaluronic acid, the chondroitin sulfates of connective tissue, the dermatan sulfates of skin, and heparin (Table 1 and Figure 3).

Hyaluronic acid

Hyaluronic acid functions in the body as an agent which increases the viscosity of body fluids and acts as a lubricant. It is present in the joints and in the vitreous humor of the eye. Hyaluronic acid is composed of equal molecules of D-glucuronic acid and 2-acetamido-2-deoxy-D-glucose, which alternate in the heteropolysaccharide. The linkages from the amino sugar to the acid are β-1,4; those from the acid to the amino sugar are β-1,3. Hyaluronic acid is also present in skin, aorta, heart valve fibroblast, and the umbilical cord.

Chondroitin Sulfates

Chondroitin sulfates are the major constituents of hyaline cartilage. They are located in cartilage and also at the calcification site of the bone. The disaccharide unit of the chondroitin sulfates contains glucuronic acid and *N*-acetylgalactosamine. Glucuronic acids are connected by β-1,3 linkages. The two major chondroitin sulfates, A and C, differ from one another in the position of the sulfates. Chondroitin sulfate A is sulfated at position 4, and chondroitin sulfate C is sulfated at position 6. Thus the names chondroitin 4-sulfate and chondroitin 6-sulfate are used for these two polysaccharides.

The number of repeating uronic acids in the chondroitin chain varies depending on the source. The number may even differ within the same tissue. The ratio of chondroitin 6-sulfate to chondroitin 4-sulfate units increases progressively with age. In humans, a plateau is reached around the age of 50.

Dermatan Sulfate

Dermatan sulfate contains L-iduronic acid as the major uronic acid. Glucuronic acid is also present in smaller amounts. Dermatan sulfate is found in the cornea and the sclera of the eye, which helps to maintain corneal transparency and the shape of the eye. Many other tissues in animals and humans also contain dermatan, such as blood vessel walls, heart valve, and the umbilical cord. Dermatan sulfates may occur at the C4 or C6 positions of L-iduronic acid. They are found to increase normally with age, except in diseased conditions.

Heparin

Heparin is mainly stored intracellularly in the granules of the mast cells. It may be released in response to a specific stimulus. The main uronic acid of heparin is the L-iduronic acid, although D-glucuronic acid is also present. Heparin is a well-known anticoagulant that binds to factors IX and XI but, most importantly, it interacts with plasma antithrombin III.

Heparan

Heparan sulfate differs from heparin in many ways. The dominant uronic acid in heparan sulfate is D-glucuronic acid. The degree of sulfation is reduced in heparan sulfate because there are fewer L-iduronic acids. Heparan sulfates are located in the extracellular medium of mast cells. They differ in their physiological functions in that they may become receptors and participate in cell growth and communication.

Alterations in Glycosaminoglycans Metabolism

The preceding discussion shows that GAG has specific functions in vertebrate animals. Recent research in this field suggests that certain disease conditions may alter the functions, synthesis, and composition of GAGs. For example, increased chondroitin sulfate and uronic acids and decreased dermatan were observed in the breast biopsies of women with carcinoma, when compared with controls. Some tumor cells have less heparan sulfate, which may reduce their adhesiveness. Hyaluronic acid may facilitate tumor cell migration through the extracellular matrix and tumor cells are capable of synthesizing increased amounts of GAG. Samples of degenerated cartilage from human patellas had increased dermatan sulfate and hyaluronate and decreased chondroitin 6-sulfate and uronic acid content. Dermatan sulfate binds to low-density lipoproteins (LDL) in plasma and is believed to play a significant role in plaque development in arteriosclerosis. Significantly increased urinary excretion of glycosamines was reported in subjects with hypothyroidism. The levels of heparan sulfate and chondroitin sulfate were higher in subjects with hypothyroidism than the controls, indicating alteration in the metabolism of connective tissue in

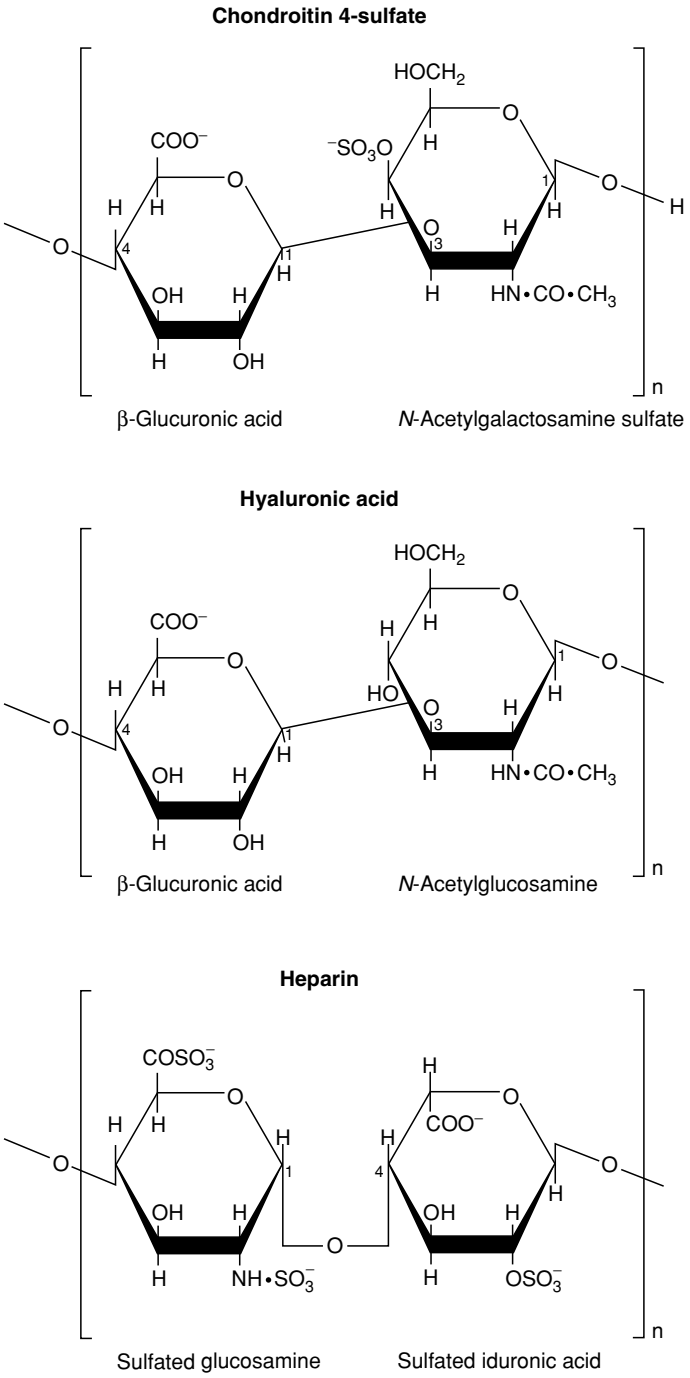

Figure 3 Structures of glycosaminoglycans.

hypothyroidism. Specific reactions responsible for these changes need to be studied.

In different types of arthritis proteins of GAG may act as autoantigens, causing additional symptoms of the disease. Recent research indicates that anti-DNA antibodies cross-reacting with GAG are present in patients with lupus erythematosus. Increased production of hyaluronic acid was observed with increased severity of the disease in patients with autoimmune thyroid disease. It appears that GAG not only plays a role in the pathogenesis of autoimmune diseases but also may be used as an activity marker of the disease.

Increases in hyaluronic acid and keratan and decreases in chondroitin sulfate may be contributing factors in the development of osteoarthritis. A recent surge in the use of chondroitin sulfate and

glucosamine by sufferers of osteoarthritis is currently being investigated.

Heparan and dermatan sulfates are found to inhibit calcium oxalate crystallization and thus prevent kidney stones. Recently a uronic acid-rich protein has been isolated from the urine of normal and stone-forming individuals. This protein appears to be a most efficient inhibitor of calcium oxalate nephrolithiasis.

Uronic Acids in Plant Tissue

In plants, uronic acids are associated with the fiber component of the plant cell. The fibers which have significance in human and animal life are called dietary fibers. Not all dietary fibers contain uronic acids.

Dietary fibers have three major components:

1. Structural polysaccharides and derivatives, such as cellulose and cellulose derivatives (microcrystalline cellulose, semisynthetic gums, and noncellulosic polysaccharides in land and sea plants).
2. Structural nonpolysaccharides, such as lignin.
3. Nonstructural polysaccharides, such as mucilages, seed gums, plant exudates, and microbial gums.

The uronic acids are found in the noncellulosic, structural polysaccharides, and derivatives – hemicellulose, pectic substances (land plants), and alginate (sea plant) – and in the nonstructural polysaccharides – mucilages, seed gums, plant exudes, and microbial gums (xanthan). These substances are also referred to as soluble dietary fiber. The most common uronic acids in the plants are D-galacturonic and D-glucuronic acids (Table 2). The uronic acid content and physiochemical properties of dietary fibers are as follows.

Hemicellulose

Hemicellulose is a branched polymer of pentose and hexose sugars, found in the plant cell wall. The uronic acid composition is mainly D-glucuronic acid and 4-*O*-methyl-D-glucuronic acid. There are two distinct hemicelluloses in plants: the acidic and the neutral. Acidic hemicelluloses contain a larger number of uronic acids than neutral hemicelluloses. Hemicelluloses are partially fermented by the microorganisms of the colon, producing some volatile fatty acids. Hemicelluloses are insoluble in water but soluble in alkaline solutions. They, along with other insoluble dietary fibers, decrease the intestinal transit time; hemicelluloses also increase fecal weight and slow down starch hydrolysis. Acidic hemicelluloses may bind to cations. These characteristics of hemicellulose may be responsible for its physiological effects.

Table 2 Uronic acid composition of fiber

Fiber	*Uronic acid composition*
Hemicellulose	D-glucuronic acid and/or 4-methyl D-glucuronic acid
Pectin	D-galacturonic acid
Gums, mucilages	D-glucuronic acid and/or D-galacturonic acid
Alginate	D-mannuronic acid and/or iduronic acid

Pectins

Pectins are a complex mixture of polysaccharides containing D-galacturonic acid as the main constituent. They are found in the primary cell walls and intercellular layers in land plants. Citrus fruits, apples, and pears contain large amounts of pectin. In the intestinal tract they are almost completely fermented by the microflora. Pectins are water-soluble and form gels under specific conditions. They are hydrophilic and have ion-binding and water-holding capacities. Pectins delay gastric emptying and increase bile acid excretion.

Gums and Mucilages

Gums and mucilages are water-soluble, viscous, and highly fermentable by the microorganisms of the intestinal tract. The uronic acids that dominate gums are D-glucuronic and D-galacturonic acids. The most important gums used in the food industry are gum arabic, gum karaya, tragacanth, carob, and guar, which are obtained as exudates from trees or shrubs. Xanthan gum, which is synthetically produced, contains D-glucuronic acid. Because of their gel-forming, water-holding ability, gums delay gastric emptying.

Alginate

Alginate is a noncell-wall component of seaweed which contains D-mannuronic and iduronic acids. It forms gels and is highly fermentable by intestinal microorganisms. Alginates are used in food processing and they enter into the human food chain as food additives.

Physiological Effects of Dietary Fiber

The physical and chemical properties of dietary fiber create physiological effects in animals. Because dietary fiber cannot be hydrolyzed by the intestinal enzymes it is unavailable for absorption and therefore continues to exert effects in the gastrointestinal tract. Soluble dietary fibers have been shown to reduce plasma levels of glucose and cholesterol. These physiological effects of soluble dietary fibers are attributed to their ability to form gels and delay gastric emptying. Viscous fibers bind to bile acids and

prevent their reabsorption, causing body cholesterol to be excreted and thus reducing plasma cholesterol. Second, fiber interferes with digestive enzymes by sequestering lipids, carbohydrates, and proteins, preventing their absorption. Fiber may also interfere with micelle formation, mixing of intestinal contents, and inhibition of cholesterol synthesis.

The viscous fibers are highly fermentable by the intestinal microorganisms, producing volatile fatty acids, which, by being absorbed into the portal blood, suppress hepatic cholesterol synthesis by inhibiting human menopausal gonadotropin coenzyme reductase activity. Some diabetics and hypercholesterolemics have experienced alleviation of their condition by increasing intakes of dietary fiber.

Isolated dietary fiber components have been used to elicit a specific response. However, It has become clear that fiber exerts its primary effect as a component of whole foods rather than as an isolated entity.

The low incidence of colon cancer in populations consuming high levels of fibrous foods has prompted scientists to study the effects of fiber on preventing colon cancer. The roles of specific dietary fiber are unclear, but diets high in fiber and low in fat are increasingly recommended.

Research indicates that the effect of fiber is not always favorable. Reduced bioavailability of certain vitamins and minerals has been reported. Among the vitamins studied, availability of vitamins B_{12}, B_6, A, and E was reduced as a result of high fiber intake. Among the minerals, the absorption of sodium, potassium, magnesium, calcium, zinc, and iron was reduced by dietary fiber, especially by the fiber fraction containing uronic acids. Research to determine the effect of fiber on nutrient bioavailability continues with vigor.

Methods of Analysis for Uronic Acids

Since uronic acids are found as integral parts of animal and plant tissues, they must be separated from their native materials.

GAGs which contain mammalian uronic acids may be separated by density gradient centrifugation and chromatography (ion exchange or gel). Hydrolysis by acid or by specific enzymes is used. Uronic acids may be measured either by colorimetry or by decarboxylation techniques.

There are two primary methods for quantification of dietary fiber. The gravimetric method measures total fiber content and uses enzymes or detergents to solubilize the nonfiber components, such as starch and protein. Defatting using organic solvents is usually performed prior to sample analysis. Acid and base solutions are used to separate acid- or base-soluble fractions.

The fractionation method, developed by Southgate in 1969, has gone through many modifications over the years. This method allows measurements of total dietary fiber as well as the fiber fractions. It is necessary to use the fractionation method for fiber analysis to be able to free uronic acids for quantification. Most fiber analyses have three steps:

1. Preparation of an extractive-free residue (alcohol-insoluble residue)
2. Removal of starch and protein from the residue (enzyme or detergent hydrolysis)
3. Analysis of destarched, deproteinated residue for neutral sugars and uronic acids

In recent years highly specific enzyme preparations have become available, improving recovery of various fiber fractions. The detergent method of fiber analysis is simple and fast, but the soluble fiber component, which contains the uronic acids, is lost during extraction. The method of choice for the quantitative analysis of uronic acids appears to be the enzymatic method followed by either colorimetry or decarboxylation.

Quantification of the Uronic Acids

Uronic acids may be determined in the hydrolysate of food samples following enzyme and/or acid hydrolysis. Dilutions of the hydrolysate to give 25–100 $\mu g\ ml^{-1}$ may be necessary and can be achieved with a mixture of sodium chloride and boric acid. Diluted samples are heat-treated in the presence of concentrated sulfuric acid. After cooling to room temperature, a 3,5-dimethylphenol solution is added and 10–15 min later the absorbance is read at 400 and 459 nm. Appropriate glucuronic acid standards are used to develop a standard curve. Differences in absorbance readings of the sample are plotted on the standard curve and sample concentrations are read.

The colorimetric methods using carbazole reagent may also be used in determining uronic acids. The sample hydrolysate is placed in a test-tube containing cold acid borate. The tubes are placed in a boiling water bath, followed by cooling to room temperature. Carbazole reagent is added and tubes are again placed in the boiling water bath. After cooling to room temperature, the intensity of the color is measured by reading absorbance at 530 nm. Sample concentrations may be calculated using values from standard curve. Recently a colorimetric method using 1,9-dimethylmethylene blue was used to determine GAG in partial urine from pediatric patients. Results indicated the efficiency and sensitivity of this method and it is recommended for widespread use. Decarboxylation of uronic acids with hydroiodic acid seems

to improve the accuracy of measurements over colorimetry procedures.

The gas–liquid chromatography (GLC) method has improved the accuracy and specificity of uronic acid determinations. Although free, low-molecular uronic acids are readily analyzed by GLC, it is impossible to measure the specific uronic acids by this method.

Recently a microtiter plate assay for the determination of uronic acids in biological samples has been validated. Modification of a commonly used procedure promises to have less risk in handling strong hot acid and increases accuracy of measurement.

See also: **Ascorbic Acid**: Properties and Determination; **Carbohydrates**: Classification and Properties; Digestion, Absorption, and Metabolism; Metabolism of Sugars; **Cellulose**; **Dietary Fiber**: Properties and Sources; Determination; Physiological Effects; **Glucose**: Function and Metabolism; **Gums**: Food Uses; **Pectin**: Properties and Determination; **Single-cell Protein**: Algae

Further Reading

Atkins EOT (1985) *Polysaccharides: Topics in Structure and Morphology*. Deerfield Beach, FL: VCH Publishers

Boons GJ (ed.) (1998) *Carbohydrate Chemistry*. New York: Blackie Academic & Professional.

David S (1997) *The Molecular and Supramolecular Chemistry of Carbohydrates: A Chemical Introduction to the Glycosciences*. New York: Oxford University Press.

Dreher ML (1987) *Handbook of Dietary Fiber: An Applied Approach*. New York: Marcel Dekker.

Hecht SM (ed.) (1999) *Bioorganic Chemistry: Carbohydrates*. New York: Oxford University Press.

Horowitz I and Pigman W (eds) (1978) *The Glycoconjugates. II – Mammalian Glycoproteins, Glycolipids, Proteoglycans*. New York: Academic Press.

Horwitz W (ed.) (2000) *Official Methods of Analysis of AOAC International*. Gaithersburg: AOAC International.

Leeds R and Avenell A (eds) (1985) *Dietary Fiber Perspectives: Reviews and Bibliography*, vol. 1. London: John Libbey.

Leeds AR and Burley VJ (eds) (1990) *Dietary Fiber Perspectives: Reviews and Bibliography*, vol. 2. London: John Libbey.

Prosky L and Harland B (1985) Dietary fibre methodology. In: Trowell H, Burkitt D and Heaton K (eds) *Dietary Depleted Foods and Disease*. London: Academic Press.

Robyt JF (1998) *Essentials of Carbohydrate Chemistry*. New York: Springer-Verlag.

Southgate DAT, Waldron K, Johnson IT and Fenwick GR (eds) (1990) *Dietary Fibre: Chemical and Biological Aspects*. RSC Special Publications no. 83. Cambridge: Royal Society of Chemists.

Spiller GA (ed.) (1993) *Handbook of Dietary Fiber in Human Nutrition*. Boca Raton, Florida: CRC Press.

Stipanuk MH (2000) *Biochemical and Physiological Aspects of Human Nutrition*. Philadelphia, PA: W.B. Saunders.

V

VEAL

R G Cassens, University of Wisconsin-Madison, West Madison, WI, USA

Introduction

Veal is the meat from a young calf. It is tender, pale in color, and has a high moisture and low fat content. It is a delicate, almost bland, meat. Consumption of veal comprises a small proportion of total meat consumption, and demand for veal is often concentrated in geographic areas or within certain ethnic populations. Veal calves may be slaughtered soon after birth or they may be specially fed for several weeks.

Methods of Production

The traditional production of veal was a segment of the dairy industry. The calves were normally fed only their mother's milk and slaughtered at a few days of age in an effort to recover some value. The meat from such immature animals was pale, delicate and tender. It is pointed out in *Larousse Gastronomique* that the ancient method of feeding the calf exclusively on the mother's milk results 'in a very pale pink meat smelling of milk, with satiny white fat having no tinge of red.'

At present, there are three methods of veal production:

1. The calf may be fed only on milk and slaughtered at 3 weeks or less of age. This produces so-called bob veal. Another alternative is to feed the calves up to about 20 weeks of age.
2. Of greatest importance now is the production of special-fed veal, a process in which the calf is fed on milk replacers.
3. When some pasture, roughage, or grain is incorporated into the diet, the result is so-called pink veal.

In the USA, the current types of production and importance of the veal industry to a dairying state such as Wisconsin can be assessed from the findings of Piwoni and Kliebenstein. They reported that the approximately 1.8 million dairy cows in Wisconsin produce about 800 000 bull calves annually. About 20% are sold as feedlot or herd replacements. About 30% are sold and slaughtered as bob veal – milk-fed calves of 3 weeks or less age. The production of bob veal continues to decline. The remaining 50% are used in the production of special-fed, fancy, or prime veal. These calves are raised in confinement, receive only milk replacer, and are slaughtered at 14–17 weeks of age when they weight 145–170 kg.

The special feeding programs have evolved into a highly intensified industry. The producers follow a detailed plan which specifies all aspects such as purchase of replacement calves, housing, feeding, management, health, and marketing. These production systems have been criticized as being factory farms in which the animals are mistreated. Such is not the case. In fact, most of these production units are indeed family operations because of the close and frequent supervision required to raise the veal calves successfully. Adequate nutritional status and comfort of the animal are prerequisites to successful production.

Wholesalers characterize veal as follows. Baby veal (bob veal) is produced from calves of only a few days' age, but they may range up to a month. They are usually male calves from the dairy industry, and the carcasses range from 10 to 25 kg in weight. Vealers range in age from 1 to 3 months, and the carcasses weigh from 35 to 70 kg. They are also raised primarily on milk. Another category comes from calves of 3–8 months of age with carcasses weighing 55–135 kg. They have been grown primarily on feeds other than milk, and their meat is maturing to resemble beef more than veal. Nature veal is produced from calves approximately 16 weeks old, and the carcass weight varies from 80 to 110 kg. They are fed a controlled and scientifically designed diet, and they have limited activity, thereby minimizing muscular development. This controlled regimen of diet and activity produces a pinkish-white color of meat, and

this is the most expensive type of veal on the wholesale market.

Marketing

The US Department of Agriculture (USDA) has official standards for grades of veal. The final grade is based on evaluation of conformation and quality, with a consideration for maturity and color of lean meat. The grades are prime, choice, good, standard, and utility. In practice, nearly 90% of the veal carcasses in commerce are not graded by the USDA.

Veal is a delicate meat which is marketed fresh, and it is not transported or stored in the frozen state. In the USA, veal is produced in the dairy production areas and is then transported to the population centers, mainly along the East Coast. In Europe, it is produced primarily in Germany, France, and The Netherlands, and much of the production is then sold to other countries.

Characteristics of the Meat

The characteristics of veal arise from the fact that it is from an immature animal. It is tender, moist, and lean. It is tender because the connective tissue is not cross-linked to the extent it is in more mature animals. (*See* **Meat**: Structure.)

A distinguishing property of veal is the pale color. Color is due primarily to the iron-containing pigment myoglobin. The muscle of newborn or immature animals has a low content of myoglobin, and as the animal matures the myoglobin increases. In raising veal, the iron content of the diet generally does not exceed the amount required for adequate nutrition. This minimizes accumulation of myoglobin in the muscle and consequently keeps the color pale.

As an animal matures, the proportions of bone and muscle decrease while fat increases. The carcass from a newborn calf is approximately 25% bone, 68% muscle, and 7% fat. The carcass from a 12-week-old animal is about 20% bone, 65% muscle, and 15% fat, while that for pink veal (animal age 16–20 weeks) is about 17% bone, 64% muscle, and 19% fat.

Differences also exist in the composition of the muscle retail cuts, as has been documented in the USDA handbooks. Veal top round in the raw state contains about 74.8% water, 21% protein, 3.1% lipid, and 1.1% ash. When the top round is cooked, the values are 55.5% water, 36.2% protein, 6.3% lipid, and 1.6% ash. Obviously, the muscle from immature animals has more water and less fat. (*See* **Protein**: Food Sources.)

Further Processing

The vast majority of veal is consumed as fresh product. Some specialty sausages, such as bockworst, are made with veal as the major ingredient. As is the case with fresh veal, these types of sausages are also noted for their rather delicate flavor, and for the fact that they are highly perishable. (*See* **Meat**: Sausages and Comminuted Products.)

Consumption

The consumption of veal in the USA, although never high, has declined markedly during the past one-third century. In the early 1960s the per capita disappearance of veal on a retail weight basis was about 2.5 kg but by the early 1990s the value had fallen to about 0.5 kg. Veal accounts for only about 1% of the total consumption of red meat in the USA.

Such average values are difficult to interpret because of regional and ethnic differences. For grocery stores overall in the USA, the percentage of sales from veal is about 1.5%, but in the north-east and Middle Atlantic regions the percentages are 2.1% and 2.4%, respectively.

Annual per capita consumption values for veal in some of the European countries, expressed in kilograms and for the year 1987, are 6.49 for France, 4 for Italy, 2.1 for The Netherlands, and 0.1 for the UK.

The probable explanations for low consumptions of veal in some countries, e.g., the USA and the UK, are that it is not readily available at market and the cost is expensive. The complications of ethnic, regional, and historical patterns must be considered in such interpretations and conclusions.

Even though veal is a naturally low-fat meat, it is unlikely that demand will increase in the future.

See also: **Meat**: Structure; Sausages and Comminuted Products; **Protein**: Food Sources

Further Reading

Anonymous (1998) *Meat Facts*. Washington, DC American Meat Institute.

Anonymous (1989) *Animal Statistical Report*. Rijswijk: PUV.

Anonymous (1988) *Larousse Gastronomique*. New York: Crown Publishers.

Fabbricante T and Sultan WJ (1975) *Practical Meat Cutting and Merchandising. Pork, Lamb, Veal*, vol. 2. Westport, CT: AVI.

Kinsman DM (1988) *Veal: Meat for Modern Menus*. Chicago, IL NLMB.

Metz JHM and Groenestein CM (eds) (1991) *New Trends in Veal Calf Production*. Proceedings of the International Symposium on Veal Calf Production, Wangeningen, The Netherlands, 14–16 March 1990. EAAP publication no. 52. Wageningen: Pudoc.

Piwoni R and Kliebenstein J (1983) *Marketing Strategies for Veal Calves*. Extension Bulletin. Madison: University of Wisconsin.

USDA (1972) *Official United States Standards for Grades of Veal and Calf Carcasses*. Washington, DC: Government Printing Office.

USDA (1989) *Agriculture Handbook 8–17 on Veal*. Washington, DC: Government Printing Office.

USDA (1990) *Agriculture Handbook 8–13 on Beef*. Washington, DC: Government Printing Office.

Vegetable Fruits *See* **Breadfruit**

VEGETABLE OILS

Contents

Types and Properties

E W Hammond, Greenisle-Consulting, Kettering, UK

Introduction

Vegetable oils are a group of fats that are derived from some seeds, nuts, cereal grains, and fruits. It is important to understand that not all of these vegetable oils are liquid oils at ambient temperatures. In addition, not all of the vegetable oils are produced in commercial quantities, and of those that are, not all are considered to be edible as in the sense of being a typical dietary component. This treatise will cover only edible vegetable oils.

Vegetable oils, as used in foods, are comprised of complex mixtures of triacylglycerols (TAGs; usually $>95\%$) with some minor amounts of diacylglycerols (usually $<5\%$). Other minor components are tocopherols/tocotrienols (up to $900\,mg\,kg^{-1}$) and phytosterol esters/phytosterols (up to 1%). The vegetable oil may be characterized confidently with expert chromatographic analysis by determining its TAG composition, together with the fatty acid composition and the minor components. **Tables 1–3** list some vegetable oils that are found in foods and give typical component analysis. The chemical and physical properties of such oils will affect how they can be used in the formulation and manufacture of foods. The main and important properties are described below.

Properties

General

When foods containing significant amounts of added edible fats are formulated, the added fat will affect three main characteristics of the food. The first is processability during preparation; the second is sensory quality (as related to taste and flavor); and the third is shelf-life, which may be physical or chemical in nature. **Figure 1** shows the relationship between some oil types, melting point, rancimat value, and total amount of both unsaturated and polyunsaturated fatty acids. Rancimat value is a measure of the oxidation potential of a fat or oil. It is the time in hours, at a stated temperature (typically 100°C) for

Table 1 Some edible vegetable oils and their typical fatty acid composition (%)

Type of oil	*8:0*	*10:0*	*12:0*	*14:0*	*16:0*	*16:1*	*18:0*	*18:1*	*18:2*	*18:3*	*20:0*	*20:1*	*22:0*	*22:1*	*24:0*
Seeds															
Babassu	5.4	6.0	44.3	15.8	8.6		2.9	15.2	1.7		0.1				
Coconut	8.1	6.5	48.6	17.7	8.5		2.5	6.5	1.5		0.1				
Palm kernel	4.0	4.1	49.7	16.0	8.0		2.4	13.7	2.0		0.1				
Sunflower			0.1	0.2	6.8	0.1	4.7	18.6	68.6	0.5	0.4				
Sunflower HOL[a]			0.1	0.2	3.0	0.4	5.9	82.4	7.4		0.3	0.3			
Rapeseed (leara)[b]				0.1	4.7	0.3	1.7	59.0	21.4	9.9	0.6	1.4	0.4	0.3	0.2
Soybean					10.0	0.2	3.5	21.0	55.3	9.2	0.5		0.3		
Cottonseed				1.0	23.9	0.5	2.9	18.5	52.5	0.3	0.4				
Peanut					10.1	0.2	3.5	51.4	27.3	0.1	1.6	1.3	3.1		1.4
Cocoa butter				0.1	26.2	0.3	34.4	34.8	2.9	1.1	0.2				
Nuts															
Almond				0.1	8.5	1.1	1.0	57.0	31.4	0.6	0.1	0.1	0.1		
Brazil			0.1	0.2	14.4	0.5	7.9	31.2	45.1	0.1	0.3	0.1	0.1		
Cereals															
Corn					10.7	0.2	1.5	30.5	55.9	0.8	0.4				
Rice bran					13.9	1.9	2.7	41.1	36.4	2.3	1.8	0.2			
Wheatgerm				0.2	18.5	0.6	0.5	18.1	55.9	5.3	0.1	0.8			
Fruits															
Palm				1.0	43.8	0.5	5.0	38.5	10.5	0.3	0.4				
Olive					10.8	0.5	3.0	75.5	8.5	0.9	0.4	0.4			

[a]High oleic sunflower oil.
[b]Low erucic acid rapeseed oil.
It should be noted that the above values are single sample analyses.
Most origins will offer a distribution of compositions depending upon variety, climate, soil, etc.

Table 2 Some edible vegetable oils and their typical triacyglycerol composition (%)

Type of oil	*C30*	*C32*	*C34*	*C36*	*C38*	*C40*	*C42*	*C44*	*C46*	*C48*	*C50*	*C52*	*C54*	*C56*	*C58*	*C60*
Seeds																
Babassu	2.8	10.1	12.4	16.3	14.0	10.0	8.1	5.3	5.1	6.0	4.0	2.9	3.0			
Coconut	3.7	12.7	15.9	18.9	17.2	10.6	7.6	4.1	2.4	2.7	1.6	1.5	1.1			
Palm kernel	1.7	6.9	9.2	21.6	16.5	9.9	8.8	6.5	5.3	5.7	2.4	2.5	2.8	0.2		
Sunflower											1.4	18.2	76.3	1.7	1.5	0.6
Sunflower HOL[a]										0.2	0.8	12.1	80.3	3.0	2.6	1.0
Rapeseed (leara)[b]										0.3	1.1	15.1	74.2	6.8	1.3	0.6
Soybean									0.2	0.7	5.1	27.1	61.3	4.0	1.2	0.3
Cottonseed										0.6	11.7	44.2	42.3	1.0	0.2	
Peanut[c]											4.2	26.4	52.3	7.2	6.4	2.9
Cocoa butter										0.5	19.3	46.8	30.9	1.9	0.4	0.2
Nuts																
Almond										0.2	4.5	19.4	75.3	0.5	0.1	
Brazil									0.2	0.5	6.7	24.6	68	0.5	0.2	
Cereals																
Corn									0.2	1.7	3.0	25.1	65.2	3.4	0.9	0.4
Fruits																
Palm									0.4	9.2	42.8	38.9	8.4	0.3		
Olive											3.4	27.2	68.2	1.0	0.2	

The numbers quoted as, for example, C54 are carbon numbers and are equal to the sum of the carbon atoms in the fatty acids of the triacylglycerol, for example: C54 = 3 × 18, or C16 + C18 + C20.
[a]high oleic sunflower oil.
[b]low erucic acid rapeseed oil − C62 = 0.4% and C64 = 0.2%.
[c]Peanut − C62 = 0.5% and C64 = 0.1%.
It should be noted that the above values are single sample analyses.
Most origins will offer a distribution of compositions depending upon variety, climate, soil, etc.

Table 3 Some edible vegetable oils and their typical tocopherol composition (mg kg^{-1})

Type of oil	*α-tocopherol*	*α-tocopherol*	*β-tocopherol*	*β-tocopherol*	*γ-tocopherol*	*γ-tocotrienol*	*δ-tocopherol*	*δ-tocotrienol*
Seeds								
Coconut	20	5	10		2	1		
Sunflower	850		25		20		2	
Rapeseed (leara)[a]	220		60		620		16	
Soybean	110		10		1250		460	
Cottonseed	550		45		330		10	
Peanut	250		15		270		10	
Cocoa butter	15	5			190			
Cereals								
Corn	150	10	20	1	720	70	120	
Wheatgerm	1900	30	710	180	250		280	
Fruits								
Palm	240	270		5	40	360	2	55
Olive	110				10			

[a]Low erucic acid rapeseed oil.
It should be noted that the above values are single sample analyses.
Most origins will offer a distribution of compositions depending upon variety, climate, soil, etc.

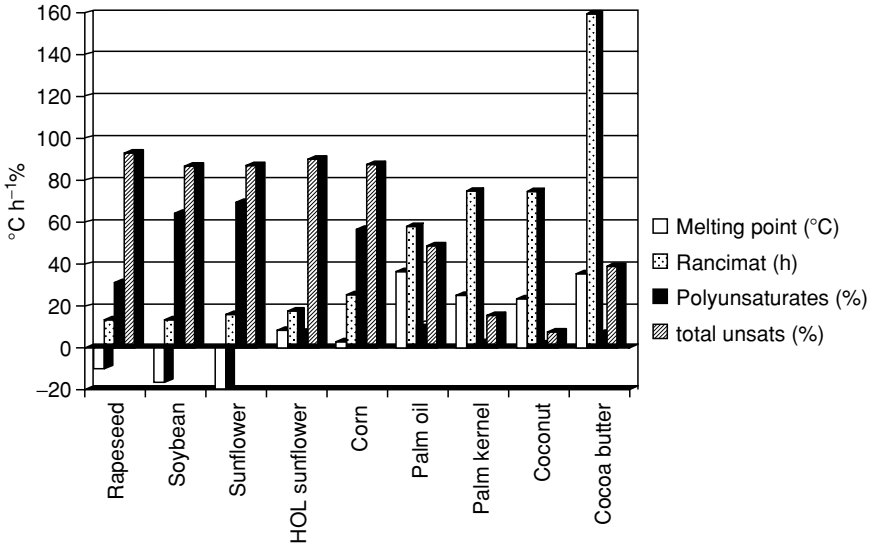

Figure 1 Relationship for different oils, between melting point, rancimat value and amount of total unsaturated and polyunsaturated fatty acids.

the fat or oil to pass through induction to a rapid oxidation rate. It can be seen that, in general, the amount of polyunsaturates has the greater effect on melting point: as polyunsaturates rise, the melting point reduces. This rise is accompanied by a reduction in the rancimat time. The rancimat time is a measure of the ease with which the fat oxidizes (becomes rancid) and therefore has a relationship to the ultimate shelf-life of a product made with that fat. Generally, as the level of total unsaturated fatty acids rises (particularly polyunsaturated fatty acids), so does the potential of the fat to oxidize and form rancid off-flavors. As an indicator, linoleic acid (two double bonds) will have an oxidation rate that is some 50 times greater than oleic acid (one double bond). Thus, it would be expected that normal sunflower oil will become rancid more quickly than high-oleic sunflower oil (Table 1).

These comparisons of properties are expressed as simple links here, although in reality the links can be quite complex. For example, comparing the data given for palm oil (PO) and palm kernel oil (PKO), both incidentally obtained from the palm fruits, shows that PKO has a lower melting point than PO, despite the fact that PO has a much higher level of polyunsaturated fatty acids. However, PKO has a

higher amount of shorter-chain fatty acids that create lower-carbon-number TAGs with reduced melting point temperatures (Tables 1 and 2). The rancimat value for PKO is higher than that for PO and this fully reflects the lower total amount of unsaturated fatty acids present in PKO (Table 1).

If corn oil is compared with sunflower oil, there appear to be anomalies also. The total unsaturates and polyunsaturates are similar and yet sunflower oil has the lowest melting point in the group shown. This is caused by the fact that sunflower oil has a high content of TAGs, with three linoleic acids in the structure.

From these points, it will be seen that the final physical properties of fats are dependent upon the 'finer' chemistry of TAG structure and this even includes the positional distribution of the fatty acids on the glycerol of TAGs.

Chemical

The main chemical property related to the use of fats in foods is the potential for oxidation and the development of rancidity. This chemical process is normally termed autooxidation and is an autocatalytic reaction with oxygen in the air. As stated above, this potential is closely related to the amount of polyunsaturated fatty acids present in the fat. Potent rancid taints can be obtained from soybean and rapeseed oils that have a descriptor of 'putty-like.' The oxidation of linolenic acid (C18:3) results in volatile compounds found in the process of setting linseed oil putty. Linseed oil is referred to as a 'drying' oil as it oxidizes rapidly in air, forming polymers that harden the putty. Sunflower oil oxidizes less rapidly, but it is linoleic acid-rich and produces potent taints with descriptors such as 'green bean' or 'cut grass.' As sunflower oil oxidizes it forms jelly-like polymers. Such taints (aromas) are found when green beans or grass are cut, although these are a result of the action of the lipoxygenase enzyme as it catalyzes the oxidation of linoleic acid.

The ultimate chemical shelf-life of a fatty food (or fat) is dependent on the unsaturation of the fat. However, the rate of oxidation of the fat will be modified by the amount and type of tocopherols present (Table 3). The tocopherols act as natural antioxidants. For example, the tocotrienols in PO are very important in controlling oxidation of the oil when it is being used as a frying medium. This activity is most important when the oil is being heated up to and used above 150 °C. However, the ambient shelf-life of fatty products is better protected by the presence of γ and δ tocopherols. Oils rich in α tocopherol appear to be less well protected.

A second chemical property that sometimes relates to shelf-life and taints is the possibility of oil hydrolysis. The glycerol ester group can be broken either with water, by straight reaction under slightly alkaline pH conditions, or through the action of lipase enzyme. The lipase can arise as a residue in the food materials or via a microbial action on the food. This latter process takes place naturally on the fats in, for example, unheat-treated wheat flours, producing free fatty acids (FFA). This process does not cause flavor taint, but may reduce the baking quality of the flour.

FFA do not normally cause flavor change at levels below 2% of the fat. However, they may cause surface activity changes in oils used, for example, in deep-frying. The result can be foaming of the oil, a deepening of the fried brown color, and a higher fat content in the product. The FFA changes the heat and moisture transfer properties at the product surface, producing an overfried effect. Where lauric oils (high content of C12:0, Table 1) are used, the production of FFA at $< 0.01\%$ of the fat can be tasted as a soapy taint. Thus, where coconut and palm kernel oils are used, FFA must be guarded against. However, FFA can be a positive property when it is present in fermented products such as soy sauce or yogurt-type products based on vegetable oils. (*See* **Fats**: Classification.)

Physical

The melting points given in Figure 1 are actually slip melting points. The slip melting point is the temperature at which 5% solid fat remains. To explain this, it is important to understand that fats are not pure compounds but are complex mixtures of TAGs. Each pure TAG type has its own melting point but when mixed with others the result will be different. Natural fats therefore do not have sharp melting points but will have a ratio of solid and liquid fat at any given temperature. The solid fat content (SFC) is an important physical property and it is measured using pulsed nuclear magnetic resonance (pNMR). The SFC may be described as a graphic curve and is used as a standard physical property when considering the function of fats in making certain foods. Figure 2 shows this type of data for a range of important food vegetable fats.

When making foods with fats that have solid properties at normal ambient temperatures, a further important physical property is the crystal form of the fat. Fats are polymorphic, that is, they have more than one crystal form, and this can be determined using X-ray diffraction. Typically, fats can crystallize in three different forms following the sequence:

$$\text{Alpha } (\alpha) \Rightarrow \text{Beta prime } (\beta') \Rightarrow \text{Beta } (\beta)$$

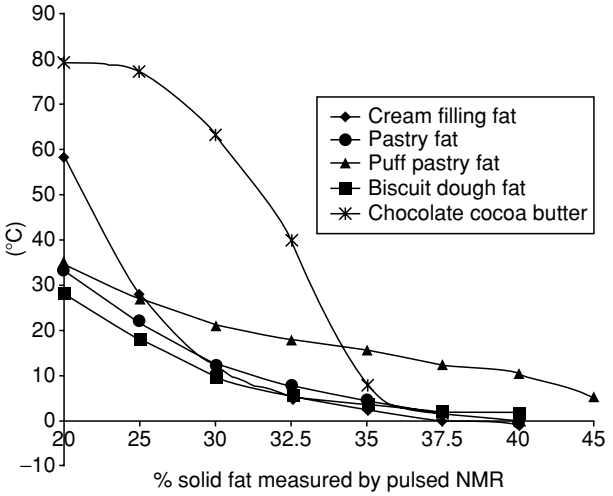

Figure 2 Comparison for different fats of percentage solid fat content over a range of temperature.

The rate of movement from one form to another is dependent upon the structure of the TAGs together with the time and temperature. It is a thermodynamic process. The α form is usually physically unstable above about 5 °C and rapidly transforms to β′. The β′ form is predominant in most fats used for food manufacture but may slowly transform over a period of weeks or months to the β form.

Cocoa butter (chocolate fat) is a special case, has a more complex crystalline nature, and can exist in six different forms. Cocoa butter has to be 'tempered' for use in chocolate so as to obtain specifically the crystal form five (form V). In this form, it retains high surface gloss and good structure. The six crystal forms and the shape of the SFC curve for cocoa butter (Figure 2) are a consequence of its special TAG structure. Cocoa butter has a predominance of symmetrical TAG, where positions 1 and 3 of the TAG are occupied by either 16:0 or 18:0, while position 2 is occupied by 18:1.

In all fats, the melting point increases as they attain the higher crystal forms. However, with the exception of cocoa butter and other fats associated with chocolate manufacture, the preferred form for use as a functional ingredient in foods is β′. The β′ and β forms are quite different and have different consequences in food products containing them. The β′ form can be likened to a bundle of needles. The crystals knit together to form a three-dimensional structure that can hold liquids such as oil or water droplets within it. The β form is more plate-like with larger crystals and these do not form a good stable three-dimensional structure.

The consequences of these physical properties are very important in food manufacture, the shelf-life of the food, and its sensory properties. It can be seen (Figure 2) that the SFC curves for dough fats are relatively flat. Most dough fats are plasticized before use, that is, they are crystallized using a scraped surface heat exchanger or rotator. This insures that the solid fraction of the fat at, say, 25 °C, is a fine distribution of β′ crystals in which is retained approximately 75% of liquid fat. This 'texturized' fat is plastic and cream-like and will also retain at least 20% of its volume of water if added at the plasticization stage. The resultant plastic fat material is easily mixed with other dough recipe components to make a cohesive and homogeneous dough. The relatively low solid fat content provides a luxurious lubricant effect in the mouth while eating the product.

Cream filling fats (Figure 2) have much higher solids at ambient temperatures, but the solids fall rapidly between 20 and 30 °C. The β′ crystal structure acts in the same way as for dough fats, retaining the liquid fat and allowing easy homogeneous blending. But in addition, the crystal network of these fats can be gas-filled to create a lower-density foamed cream filling. The steep decline in solids at the low-temperature end of the SFC curve provides a 'clean' palate, excellent flavor release, and a luxuriant cooling sensation. This latter point is caused by the rapid melting of fat and removal of heat from the tongue and palate of the mouth. At the mid-range temperature, cream fats are designed to have a particular slip melting temperature. This suits whether they are to be deposited or stencilled on to such materials as biscuit shells. The physical properties are very important for these processes to be successful.

Cocoa butter (Figure 2) shows a convex SFC curve. There is considerable resistance to melting up to 30 °C and this accounts for the ability of chocolate to stay hard and glossy over a wide ambient temperature range. Above 30 °C (mouth temperature) the fat begins to melt very rapidly. This provides a 'rush' of flavor release, a luxuriant creamy sensation, and (for high-quality chocolate) a distinct cooling of the palate. By the time body temperature is reached (37.5 °C) there is no solid fat remaining. (*See* **Cocoa: Chemistry of Processing.**)

The physical properties described above provide for a good eating product with a relatively long shelf-life. However, foods are dynamic materials and thermodynamic processes strive to obtain the least energy or lowest enthalpy in, for example, the fat. The appearance of fat bloom, softening or shrinkage of fat cream fillings, and softening and bloom of chocolate coatings can limit shelf-life. The transformation of a β′fat system to a β fat crystal in a product will almost certainly be accompanied by

the development of fat bloom or graining of the fat phase. In these cases the liquid fat is released and the product may become 'oily.' Fat bloom is shown as either excessive crystal growth (graining) or a frosted or moldy appearance on, for example, chocolate coatings or biscuit surfaces.

While β′ fat systems do retain the liquid fat efficiently it is inevitable that, with time, the liquid phase will migrate out. This is seen particularly where two fat-containing phases are in contact. Fat migration changes the composition of the fat and therefore the SFC. The outcome is usually a softening of, for example, a chocolate coating. Often the properties of vegetable fats are not optimum for an application; in this case the physical properties can be changed. Blending different fats or chemically modifying them by hydrogenation or interesterification will do this. In all cases, the chemical make-up is changed and this has a direct effect on the physical properties.

See also: **Chromatography**: Gas Chromatography; **Cocoa**: Chemistry of Processing; Production, Products, and Use

Further Reading

Ching Kuang Chow (ed.) (2000) *Fatty Acids in Foods and their Health Implications*. New York: Marcel Dekker.

Gunstone FD, Harwood JL and Padley FB (eds) (1986) *The Lipid Handbook*. New York: Chapman & Hall.

Hammond EW (1993) *Chromatography for the Analysis of Lipids*. Boca Raton, FL: CRC Press.

Timms RE and Stewart IM (1999) Cocoa butter, a unique vegetable fat. *Lipid Technology Newsletter* 5: 101–107.

Oil Production and Processing

W Hamm, Harpenden, UK

Background

Vegetable oils are principally recovered from oilseeds, with oil-rich fruit such as the fruit of the oil palm and of the olive tree providing important additional sources. Solvent extraction is used for oil recovery from oilseeds, but in the case of palm and olive oil, the oil is recovered by separating it from the aqueous phase present in the fruit after crushing. Vegetable oils are for the most part refined before consumption, refining comprising a series of steps designed to produce a bland, stable oil. Refined oils may be modified in order to change their physical properties. Virgin, i.e., unrefined, olive oil plays an important part in the market for vegetable oils.

Both refining and oil modification processes are increasingly being improved in order to conform to modern standards of a healthy food product. This entails avoiding the formation of undesirable artefacts as well as minimizing the removal of valuable minor components.

Oilseeds and Oil-rich Fruits Used in the Production of Edible Oils

Vigorous growth in the last half-century of production and consumption of vegetable oils has meant that the world-wide per capita consumption of oils and fats has risen significantly, with many parts of the developing world sharing this growth with that in the industrialized regions. A direct consequence of this increase in availability has been a substantial increase in international trade in oils and oilseeds. At the same time, vegetable oils also find substantial use in the oleochemicals industry, which therefore serves as an additional outlet for a limited number of vegetable oils.

The oils used for edible purposes fall into two categories. The seed oils are produced from annually planted seeds, and this category includes, besides soyabeans, rapeseed, sunflower seed, cottonseed, groundnuts, and many other oil-rich seeds. The oil content of the major seeds covers a wide range, the soyabean normally containing less than 20% oil, whereas many of the others contain in excess of 40% oil after dehulling. The second category of vegetable oil sources is a smaller group of tree crops, primarily comprising the fruit of the coconut and oil palms as well as that of the olive tree. Being tree crops, these do not have to be replanted annually, but in consequence, the fruit only becomes available after some years of cultivation. The fruit of the oil palm shows a unique characteristic in that the kernel of the fruit contains an oil (palm kernel oil) that has a composition that differs radically from the oil in the fleshy part of the fruit, the mesocarp.

From the point of view of applications, it is also useful to consider vegetable oils on the basis of the chain length of their major fatty acid components. Oils containing large proportions of shorter-chain fatty acids (C12 and C14 chain length), of which coconut oil and palm kernel oil are the principal representatives, comprise one group, whereas the group of oils where the major fatty acids are of C18 chain length includes the majority of the seed oils (see **Table 1**).

Table 1 Oil production: oil yield and quality

Seed/fruit	*Seed crushed 1999–2000 (million tonnes)*	*Oil production 1999–2000 (million tonnes)*	*Oil yield (%)*	*Oil yield calculated on*	*FFA of oil*
Soya	137.92	25.21	18–19	Dehulled beans	0.3–1.2
Rapeseed	37.46	14.41	38–42	Undehulled seed	1.0–2.0
Sunflower	23.72	9.57	40	Dehulled seed	1.0–2.0
Cottonseed	26.67	3.92	23	Delinted, dehulled seed	1.0–3.0
Groundnut	11.34	4.54	47	Kernels	1.0–3.0
Palm	na	20.54	22–28	Oil yield on fresh fruit bunch	3.0–5.0[a]
Coconut	4.85	3.02	60–70	Copra	2.0–4.0[b]
Olive	na	2.27	20	Yield excludes kernel oil	See section on Olive oil

[a]Calculated as palmitic acid.
[b]Calculated as lauric acid.
na, not applicable.
From Oil World Annual 2000.

Oil Recovery

Oilseeds

The extraction of oil from the various oilseeds requires the application of heat and moisture to prepare the oil-containing cells for rupture and pressure to press the oil from the seed after size reduction, size reduction generally being achieved by flaking. Screw-press expellers have made it possible to operate virtually continuously but leave 3–5% of oil in the protein-rich cake. The development of the solvent extraction process has increased oil recovery and at the same time has made it economically attractive to process seeds relatively low in oil content such as soyabeans. However, the oil obtained using a screw press contains fewer of the minor components present in the oil, e.g., phospholipids, and is consequently easier to refine.

Industrial hexane is the solvent used universally for the solvent extraction of oil from oilseed, although research into extraction with other solvents, including supercritical carbon dioxide, has been reported. Solvent extraction is carried out using countercurrent percolation of solvent through a bed of seed flakes. The oil-rich solvent leaving the extractor is then desolventized in a multiple-effect evaporator. The solvent recovery sections of the process generally account for a large share of the capital investment in a solvent extraction plant, and despite the use of sophisticated energy-recovery techniques, the energy required for solvent recovery makes the operating costs for the process considerably higher than the costs of screw-press oil recovery. This additional cost of production must be offset by the value of the increased oil recovery.

In the case of seeds having a high oil content, the oil recovery process in many cases comprises a screw-press operation followed by solvent extraction of the partially deoiled cake. Oils containing high levels of phospholipids, e.g., soyabean and rapeseed (canola) oil, are water-degummed after extraction and desolventization as the phospholipids precipitate during storage and thus interfere with oil handling and downstream processing. Degumming with water reduces the phospholipid content of an oil of good quality to approximately 200 p.p.m., expressed as phosphorus, equivalent to 0.5–0.6% phospholipids (see Figure 1). (*See* **Phosphorus**: Properties and Determination.)

Cold-pressed oils, i.e., oils recovered from seed with minimum application of heat during pretreatment and without use of solvent, are becoming popular in some countries.

Oils Obtained from Tree Crops

Palm oil Recovery of the oil from the fruit of the oil palm (*Elaeis guineensis*) requires early sterilization of the harvested fruit bunches if the formation of free fatty acid in the oil resulting from the action of an enzyme present in the fruitlets is to be avoided. The oil palm plantation should therefore be closely linked, both geographically and organizationally, to a mill capable of processing the harvested fruit with minimum delay. The fruit is steam-sterilized, and the fruitlets are then stripped from the bunch. This operation is followed by pressing to recover the crude palm oil, with the palm kernels left in the press cake. After clarification and drying, the crude palm oil is ready for downstream processing. One hundred tonnes of fresh fruit bunches (FFB) can be expected to produce approximately 23 tonnes of crude palm oil. Crude palm oil from South-east Asia normally contains 4–5% free fatty acids (FFA), but palm oil containing 2–3% FFA can be produced. The crude oil

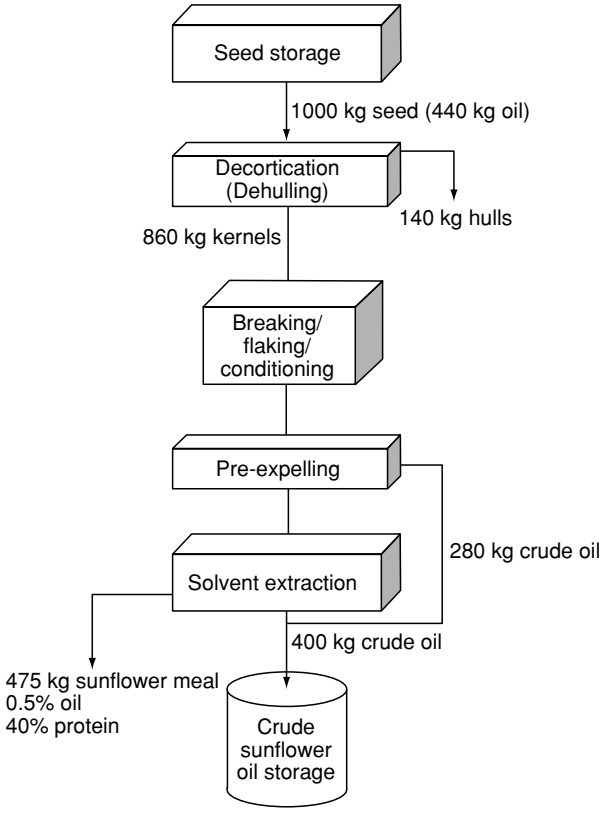

Figure 1 Sunflower oil production: average values for processing of decorticated sunflower seed. Note that oil may be recovered from either whole or partially decorticated seed.

is rich in β-carotene, tocopherols, and tocotrienols, and this feature of the oil is used to enhance its value.

The palm kernels, after separation from fibrous material, are processed for oil recovery using a solvent extraction process, as described above for seed oils. One hundred tonnes of FFB should produce approximately 2.5 tonnes of palm kernel oil, an oil rich in short-chain fatty acids.

Olive oil The olive oil production process involves crushing the fruit followed by a two-stage centrifugal system that serves to separate the oil and then wash it. The fruit contains an average of 20% oil, and the bulk of the oil in the fruit is produced to the standards required to classify it as virgin or extra virgin oil. Olive-pomace oil is the oil obtained by solvent extraction of the residue after the first separation stage. In the European Union, this oil must be refined and then blended with a specified amount of virgin olive oil before it can be sold for human consumption. Internationally agreed specifications for the various grades of olive oil require extra virgin olive oil to have a FFA content not exceeding 1.0% (as oleic acid), whereas virgin olive oil may contain up to a maximum of 2.0% free fatty acid. (*See* **Olive Oil.**)

Rice bran oil Rice bran oil, a byproduct of rice processing, is mainly produced in East Asia. Like the palm fruitlet, rice bran contains an enzyme that rapidly attacks the oil if it is not heat-deactivated immediately after separation from the crop. In practice, unrefined rice bran oil often contains high levels of FFA as a result of delays in processing. The presence of important antioxidants in rice bran oil, ferulic acid esters generally referred to as γ-oryzanols, has recently led to its use as a component of frying oils in order to extend the useful life of industrial frying oils.

Coconut oil Coconut oil is the oil obtained from the dried 'meat,' generally known as 'copra,' of the coconut fruit. The drying of the copra, a necessary step, prior to oil extraction, is sometimes carried out with hot gases containing polycyclic aromatic hydrocarbons (PAHs). These must be removed during refining of the oil. Coconut oil has been the most important of the so-called lauric oils, although palm kernel oil production is now comparable with that of coconut oil. (*See* **Coconut Palm.**)

Oil Movements

The pattern of growth in edible oil production and consumption has meant that large quantities of crude edible oil are moved intercontinentally from producer to user countries. In order to avoid contamination of vegetable oil cargoes with residues of cargoes previously carried by the same vessel, various international bodies such as the London-based Federation of Oils, Seeds and Fats Associations and the US National Institute of Oilseed Products introduced regulations governing the type of cargo that may be carried by a parcel tanker in the three loadings immediately prior to loading an oil to be used for edible purposes.

The possibility of overheating of the oil during the voyage or at the time of discharge is also a cause of oil quality deterioration in the course of long-distance transportation. Table 2 is a schedule of recommended cargo temperatures cited in a recent Codex Alimentarius Commission report.

Ingress of water into the cargo while at sea can also be troublesome, as this can lead to hydrolysis of the triacylglycerols, the major component of the oil, which manifests itself in an increase of the FFA content, the process of hydrolysis being considered to be autocatalytic.

A consequence of these oil movements is the need for substantial storage capacity for vegetable oils at the receiving port, and large tank farms have been

established over the years in the major receiving ports, Rotterdam, Hamburg, and New York. As Malaysia mainly ships refined palm oil fractions (though this may be re-refined by the receiver), there is a growing demand for high-quality storage at the major ports. Nitrogen blanketing of cargoes either during shipment or in port storage has been recommended. In the case of higher-value fats, e.g., confectionery fat components, shipment in special containers has been used.

Oil Processing

Although filtered crude oil is sometimes consumed without any conventional refining, for all practical purposes, vegetable oils must be refined in order to make them suitable for consumption in one of the forms in which oils and fats are used. The refining processes in use nowadays have a number of objectives, the most important of these being

- to produce a clear, stable oil free of undesirable minor components;
- to ensure that the refined oil has a good keepability and retains, as far as possible, those minor components that are considered to be desirable, e.g., tocopherols;
- to avoid structural damage to the oil, e.g., by the formation of *trans*-fatty acids and polar compounds (diacylglycerols, polymers);
- to minimize direct and downgrading losses intrinsic to the process;
- to minimize the cost of processing.

These objectives are achieved by subjecting the oil to either of the two main processes used by the industry. Chemical refining relies on the reaction of the FFA present in the crude oil with an alkaline solution, normally sodium hydroxide, to reduce the FFA content of the oil to an acceptable level. Physical refining, however, makes use of the relative volatility of these FFA to separate them from the major constituents, the triacylglycerols, in a steam-stripping operation carried out at high temperature and low pressure, each process having advantages and disadvantages (Table 3). Both processes (see Figure 2) are considered in more detail in the following sections.

Table 2 Recommended schedule of cargo temperatures (°C)

Oil	*During voyage*		*On discharge*	
	Min	*Max*	*Min*	*Max*
Sunflower	Ambient	Ambient	Ambient	20
Soyabean	Ambient	Ambient	20	25
Groundnut	Ambient	Ambient	20	25
Rapeseed	Ambient	Ambient	Ambient	20
Corn	Ambient	Ambient	Ambient	20
Coconut	27	32	40	45
Palm	32	40	50	55
Palm stearin, RBD	40	45	60	65
Palm olein, RBD	25	30	32	35

Schedule produced by International Seed Crushers' Association.

Degumming

An important characteristic of vegetable oils is their phospholipid content, the seed oils being richest in these (see Table 4). For processing reasons, the phospholipids are often divided into those readily hydratable by water, and the remainder that can only be hydrated (and become insoluble in oil) by the addition of acids. The water-hydratable phospholipids, which are frequently removed immediately after extraction, are useful as a source of lecithin and its derivatives.

The removal of nonhydratable phospholipids is sometimes combined with chemical removal of FFA but can also be carried out as a separate operation, an option that becomes particularly important when the oil is to be refined physically.

Citric acid, though more costly to use than phosphoric acid for the hydration of phospholipids not

Table 3 Advantages and disadvantages of chemical and physical refining of oils

Advantages	*Disadvantages*
Chemical refining	
1. Flexibility (ability to process small batches, dark oils)	1. Refining loss (oil downgrading) can be high
2. Moderate deodorization temperature required	2. Relatively high level of effluent formation
3. Better for recovery of valuable minor components in deodorizer distillate	
Physical refining	
1. Lower refining loss	1. Higher stripping/deodorization temperatures required
2. Reduced effluent formation	2. More sensitive to residual levels of phospholipids and metals
3. Less labor-intensive	3. Increased bleaching earth consumption
	4. Unsuitable for dark oils

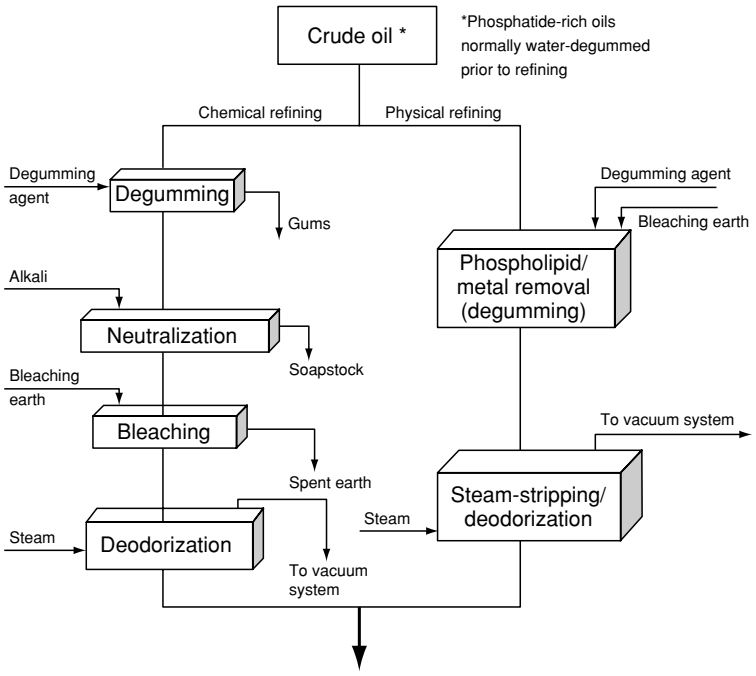

Figure 2 Refining options for edible oil.

Table 4 Phospholipid contents of major oils

Oil	*Phospholipids (%)*	*Phosphorus (p.p.m.)*
Rapeseed	0.5–3.5	200–1400
Soyabean	1.0–3.0	400–1200
Sunflower	0.5–1.3	200–500
Corn	0.7–2.0	250–800
Cottonseed	1.0–2.5	400–1000
Palm	0.03–0.1	15–30

hydrated by water, offers a greater versatility in use and poses no risk of acid residues in the processed oil. Table 5 shows the processes used for degumming oils and their effectiveness in reducing the phospholipid content of oils. Ultrafiltration of the oil (in solvent) to remove the phospholipids is also effective but more costly than other processes.

Removal of the phospholipids present in the oil facilitates the subsequent stages of refining, as it reduces the tendency to form emulsions during neutralization of the FFA and also permits operation at high temperatures (> 200 °C) without any risk of discoloration. When the oil is to be refined physically, i.e., using stripping steam to remove the FFA from the crude oil, the degumming process must also remove traces of metals such as copper and iron, which would discolor the oil at the temperatures used for steam-stripping.

Chemical Refining

Neutralization All crude vegetable oils contain FFA that should be removed in the interests of refined oil quality. The seed oils for the most part contain 0.8–1.5% FFA, but palm oil may contain up to 5.0% FFA as the result of enzymic action in the interval between fruit harvesting and sterilization. Oils such as ricebran oil and several of the more exotic fats, e.g., sheanut butter and illipe fat, may contain more than 5% FFA.

Although, in principle, any alkali may be used to neutralize the FFA present in the crude or degummed oil, sodium hydroxide solution (strength 8–24%, depending on the oil to be refined – a dark oil will normally be neutralized with a stronger alkali than an oil that is light in color) is used almost universally in industrial chemical refining of fatty oils. The neutralization reaction is carried out at 80–90 °C (see Figure 3). In the conditions of intense mixing, as practiced in modern centrifugal refining, it requires contact times between crude oil and alkali of less than 60 s. The soaps formed by neutralization, known as soapstock, are insoluble in the oil and are separated centrifugally, the disk centrifuge being the preferred type nowadays. As the soaps have a low but finite solubility in the oil at the processing temperature (soap content of oil = 150–200 p.p.m.), it is necessary to follow separation of the soapstock by a washing

Table 5 Degumming processes

Process	*Suitable for*	*Residual phospholipids (p.p.m. phosphorus)*
Water degumming	Crude oil	100–250
Acid degumming	Crude oil	20–50
Acid refining	Crude oil	< 20
Superdegumming (Unilever)	Crude or wdg oil	5–15
TOP degumming (Vandenmoortele)	Crude or wdg oil	5–10
Enzymatic degumming	wdg oil	5–10
SOFT degumming (Tirtiaux)	Crude oil	< 5
IMPAC degumming	Crude oil	3–10

wdg, water-degummed.

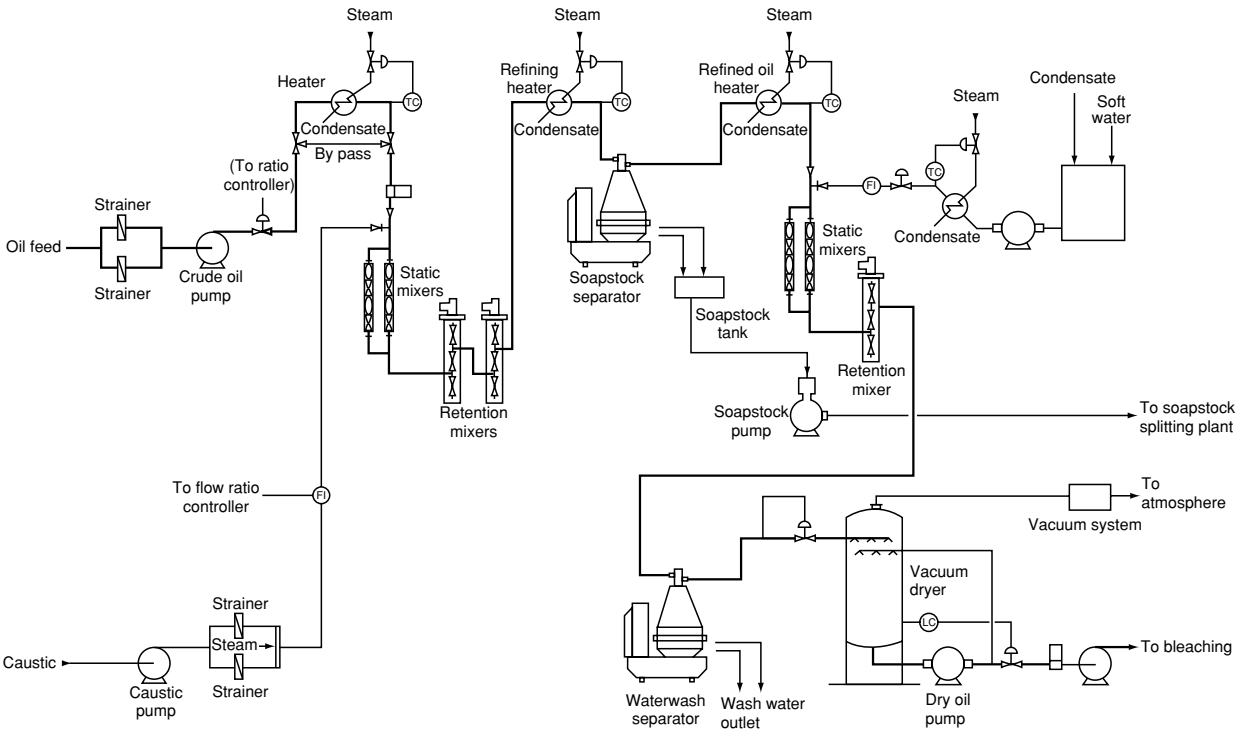

Figure 3 Vegetable oil neutralization plant arrangement. Based on a flowsheet produced by Europa Crown Ltd., Hessle, North Humberside, UK.

stage, in which soft water is used to reduce the soap content of the oil to below 30 p.p.m. The FFA content of the neutralized oil should be below 0.1%.

The soapstock contains a certain amount of neutral oil owing to spontaneous emulsification taking place at the oil–alkali interface in the course of the neutralization reaction. This is reflected in the refining loss experienced by refiners, defined as the ratio of oil loss at the neutralization stage to the FFA present in the crude oil and known as the refining factor. For most oils, the refining loss in chemical refining can be expected to be in the range of 1.5–2.0, although lower refining factors have been obtained when using nonturbulent alkali–oil contactors. The downgrading of sound oil to the lower-valued soapstock or acid oil produced by acidulation of soapstock results in a significant cost charge against the process.

The effluent resulting from chemical neutralization has been recognized as an environmental problem, and this has resulted in a change to physical refining, particularly in Europe. When chemical refining is used to remove only low levels (< 0.15%) of FFA, as in the case of an oil already refined in the country of origin and then transported intercontinentally, it is possible to use a hydrogel silica or a modified magnesium aluminum silicate to adsorb the soaps formed,

instead of washing them out of the oil. Similarly, these materials can be used to replace the wash centrifuge in normal refining in order to remove the soap content remaining in the refined oil after soapstock separation, thus again avoiding the formation of additional liquid effluent.

Bleaching The neutralized oil to be bleached is deaerated and dried at a pressure of approximately 50 mbar to a moisture content of not more than 0.1%. Use of acid-activated bleaching earth (bleaching clay in the USA) will remove traces of metals such as copper and iron from the oil in addition to pigments and residual soap, which is important from the point of view of oil keepability. Cocurrent or countercurrent bleaching may be used to reduce the bleaching earth requirements, with savings of 10–15% relative to batch bleaching being reported for the cocurrent option and up to 40% for countercurrent processing. Oil leaving the bleacher is filtered to separate spent bleaching earth, but this retains high levels of oil unless it is processed further, e.g., by blowing steam or nitrogen through the press. Spent earth destined for disposal to a landfill site will generally still contain approximately 30% oil. This therefore becomes a charge against the process, additional to the refining loss explained above. The spent bleaching earth is also a fire hazard.

In recent years, a number of alternatives to acid-activated bleaching earth have been developed for use in edible oil refining. The use of silica as an adsorbent and filter aid, generally in admixture with bleaching earth, has been shown to be feasible, though the cost is higher than that of the more traditional bleaching earths. A number of multistage bleaching processes are available, their purpose being to reduce the bleaching earth requirement per tonne of oil processed.

An oil contaminated with polycyclic aromatic hydrocarbons (PAH) requires the addition of a small quantity of activated carbon to ensure that the PAH content of the fully processed oil is reduced to acceptable levels.

Deodorization The final stage of oil processing, deodorization, has the function of removing volatile components remaining in the oil after the previous stages of refining as well as those volatile compounds formed during deodorization itself. In addition, some heat-bleaching is likely to occur at this stage, owing to the high temperatures used. The removal of volatile components is achieved by the passage of open steam through the oil at temperatures of 200–220 °C at a pressure of 4–10 mbar. Stripping steam consumption in the process can be expected to be in the range of 0.5–3.0%. In order to ensure stability of the processed oil, its peroxide value should be below 1.0, preferably below 0.5 meq of O_2 per kilogram of oil, which requires rapid cooling of the oil after the completion of deodorization.

Deodorization can be carried out batchwise, semicontinuously or fully continuously, and in the case of a fully continuous operation, either a crosscurrent or countercurrent mode of steam–oil contact may be used. The extent of removal of volatile components is directly proportional to the volatility of the component and is therefore a function of the operating temperature. At the same time, it is also inversely proportional to the operating pressure. Since deodorization is a dynamic process, the interfacial area between the stripping medium and the oil to be deodorized is a vital parameter in determining stripping efficiency.

Modern deodorizers, almost all of which operate semicontinuously or fully continuously, are fitted with heat-recovery systems that can recover up to 80% of the energy needed to heat the incoming oil to the operating temperature. The vacuum system, normally consisting of a series of steam jet ejectors, also uses substantial quantities of steam in the form of motive steam, a ratio of 5–7 kg of motive steam per kilogram of stripping steam evacuated being required in temperate regions when operating at a head pressure of 4–6 mbar. The motive steam requirement can be reduced by chilling the water used to condense the mixed vapors from the deodorizer or by using a 'dry' vacuum system that makes use of mechanical vacuum pumps instead of steam jet ejectors.

Apart from high-energy consumption, the environmental impact of deodorization manifests itself in the liquid and gaseous emissions created by the process. Entrainment of oil in the stripping vapors leaving the deodorizer constitutes a process loss, generally 0.2–0.3% of the oil processed, but if not trapped in a scrubber or condenser, this entrainment will appear in the effluent stream leaving the plant. In recent years, steps have also been taken to reduce the odoriferous gaseous effluent from the deodorizer.

In addition to removing the undesirable minor components of the oil reaching the deodorizer, the stripping steam will also remove, at least partially, valuable components such as tocopherols and sterols. The tocopherols constitute a valuable byproduct that can be converted to natural vitamin E concentrates. However, the α-tocopherol content of most vegetable oils accounts for only a small proportion of the total tocopherol content (Table 6), and the deodorizer distillate requires removal of the nontocopherol components but also conversion of the other tocopherol isomers to the α-form if the product is intended as

Table 6 Typical tocopherol contents of major vegetable oils

Oil	*Total tocopherol (p.p.m.)*	*α-Tocopherol (p.p.m.)*	*β-Tocopherol (p.p.m.)*	*γ-Tocopherol (p.p.m.)*	*Reference*
Soyabean	811	179	28	604	1
Rapeseed	553	216	10	326	2
Sunflower	570	564	2	4	1
Cottonseed	788	403	2	383	1
Corn	840	272	2	566	1
Palm[a]	101	91	2	8	1

[a]Also contains 180 p.p.m. of tocotrienols.

1, *Journal of the American Oil Chemists Society* 62(3): 531–538 (1985).

2, *European Journal of Lipid Science and Technology* 102: 618–623 (2000).

a diet supplement. It is also possible to recover sterols, which can be converted to pharmaceutical products, from the deodorizer distillate when refining soyabean oil.

Physical Refining

Physical refining of a vegetable oil requires operation at a temperature significantly higher than the temperature used in the deodorization stage of chemical refining, as the process depends principally on the volatility of the FFA present in the oil, which are removed by reaction in chemical refining. Processing temperatures in the stripper are therefore generally in the range of 220–250 °C, which means that the oil being processed is subjected to greater thermal stress than the oil deodorized following chemical refining.

Degumming As already indicated, degumming of an oil that is to be physically refined must be designed to achieve higher standards of removal of undesirable components than when degumming as part of a chemical refining process. Acid refining, in which phosphoric acid is used to hydrate the phospholipids not hydrated by water and alkali is then added to neutralize the excess acid, can be used to reduce phospholipids to below 10 p.p.m. Bleaching earth is then added to adsorb the precipitated material, and the mixture of oil and earth is filtered conventionally. With the simultaneous removal of metals, the acid-refined oil is suitable for physical refining.

The bleaching earth requirement for the physical refining process is higher than that for chemical refining, and this fact counterbalances to some extent the lower refining losses in physical refining. Table 7 compares the bleaching earth requirements for various oils in chemical and physical refining.

Table 7 Bleaching earth consumption in physical and chemical refining

Oil	*Percentage of bleaching earth*	
	Fully degummed oil (for physical refining)	*Neutralized (chemical refining)*
Sunflower	0.3	0.1–0.2
Rapeseed	0.7	0.5
Corn	0.8	0.5

From Kokken M (1996) *World Conference on Oilseeds and Edible Oils, Istanbul*.

Stripping/deodorizing The requirement for removal of FFA by steam-stripping in physical refining distinguishes this operation from the apparently closely related operation of deodorization in a number of important respects – a significantly higher operating temperature is needed in order to reach fatty acid vapor pressures at which stripping becomes practical; the molar fraction of fatty acids in the vapor stream leaving the stripper contributes significantly to the total vapor flow and therefore to the design of the condenser used to recover material from the vapour stream; and the higher operating temperature leads to a greater loss of other volatile components, e.g., tocopherols, from the oil. A further consequence of the higher process temperature is the greater possibility of *trans*-fatty acid (TFA) formation.

Stripper design is generally similar to the design of deodorizer discussed earlier, but batch operation is not appropriate. In practice, steam-stripping of fatty acids uses slightly higher quantities of stripping steam per unit of oil processed than deodorization. The condensed stripper vapors, being richer in FFA than deodorizer distillate, have a proportionately lower content of the valuable tocopherols than the distillate and therefore constitute a lower-value byproduct.

Oil Modification

The three oil modification processes – hydrogenation, fractionation, and interesterification – have made a major contribution to the greater use of oils and fats since their introduction in the twentieth century. These processes principally serve to modify the relationship between liquid and solid content of an oil or fat as a function of temperature, thus enhancing their

functionality. Hydrogenation is the most versatile of these processes, primarily because at least three process conditions can be used to modify process performance.

Although the modification processes are used individually in many cases, their coupling, e.g., hydrogenation followed by fractionation, or fractionation coupled with interesterification, can lead to enhanced versatility in the use of different oils and fats.

Hydrogenation

Hydrogenation is used to reduce the unsaturation present in oils and thereby to increase their solid fat content. The conditions for hydrogenation must take into account the different reaction rates for the range of molecular species present in the oil. Nickel, in various forms designed to maximize the active catalyst surface area, is generally used as the reaction catalyst, although other catalysts, e.g., platinum, palladium, and copper, are also effective. The catalyst is precipitated on a carrier for the purpose of maximizing the catalyst surface area available for the reaction. Sulfur addition to the nickel catalyst is used to promote steep melting curves in the hydrogenated fat, but this is achieved by increasing the proportion of TFA in the fat.

When processing a well-refined oil, it is now possible to hydrogenate an oil with catalyst concentrations as low as 40–50 p.p.m. (Table 8). The process conditions employed, including hydrogen pressure, temperature, and catalyst concentration, are chosen to produce a hydrogenated product best suited to the required functionality of the fat produced, bearing in mind that the rate of conversion of linoleic acid (C18:3) to linoleic (C18:2) acid is approximately 40 times greater than that of oleic (C18:1) acid to stearic (C18:0) acid. The selectivity required in the hydrogenation reaction, which expresses the above relationships between the relative reaction rates for the different triacylglycerols present, provides guidance on the choice of catalyst. The process is mostly carried out batchwise or, by linking a series of batch reactors, in a semicontinuous manner. The autoclave may be operated with or without hydrogen recirculation. The loop reactor offers an alternative form of the gas–liquid contactor.

Table 8 Typical hydrogenation conditions

Temperature (°C)	100–250
Catalyst concentration (% Ni)	0.01–1.0
Hydrogen pressure (bar)	1–10
Hydrogen consumption (m^3 per tonne of oil per unit iodine value drop)	0.93
Reaction rate (units of iodine value fall per minute)	1–3

The exothermic hydrogenation reaction leads to a rise in temperature of the bulk oil of approximately 1.6 °C per unit decrease in iodine value (the iodine value of an oil is the standard measure of unsaturation of oils and fats). Efficient cooling must therefore be provided during the reaction in order to maintain the oil at a stable temperature. After hydrogenation the oil must be re-refined in order to remove traces of catalyst from the oil.

Hydrogenation is particularly useful in situations where it is important to maximize the use of one oil, as the process conditions can be used to create a range of hardened oils having different melting characteristics and therefore suiting different applications. In the case of soyabean oil, it is used to convert the relatively unstable linolenic (C18:3) acid to more stable fatty acids.

In recent years, the formation of TFA when hydrogenating has given rise to much reconsideration of the process, since these fatty acids are considered to be health risk (see Figure 4). However, *trans*-isomerization is favored where a fat with a high melting point but without excessive reduction of its degree of unsaturation is required. The formation of TFAs can be controlled by limiting the reuse of catalyst, although this obviously increases process costs. A solution to this problem that is increasingly finding favor is to hydrogenate the oil completely and thereafter interesterify the fully hardened oil with a liquid oil.

Fractionation

In the harvesting year 1999/2000, palm oil accounted for 25% of a total global vegetable oil production of 84 million tonnes, whereas in the year 1974/1975, this oil accounted for only 8% of a total production of approximately 30 million tonnes. The eight fold increase in palm oil production reflected in these figures largely accounts for the important role of fractionation in modern edible oil processing, as palm olein has become a major industrial frying oil.

Fractionation, the separation of liquid and solid fractions of vegetable oils, is based on the reduction in solubility of the more unsaturated in the more saturated triacylglycerols as the bulk temperature is reduced. Thus, cooling of the oil leads to progressive crystallization of, in the first instance, trisaturated triacylglycerols and thereafter disaturated triacylglycerols. As crystallizing triacylglycerols exhibit different polymorphic forms, it is essential to control crystallization conditions in such a way as to minimize occlusion of liquid (uncrystallized) material in the crystallized phase. This normally requires slow

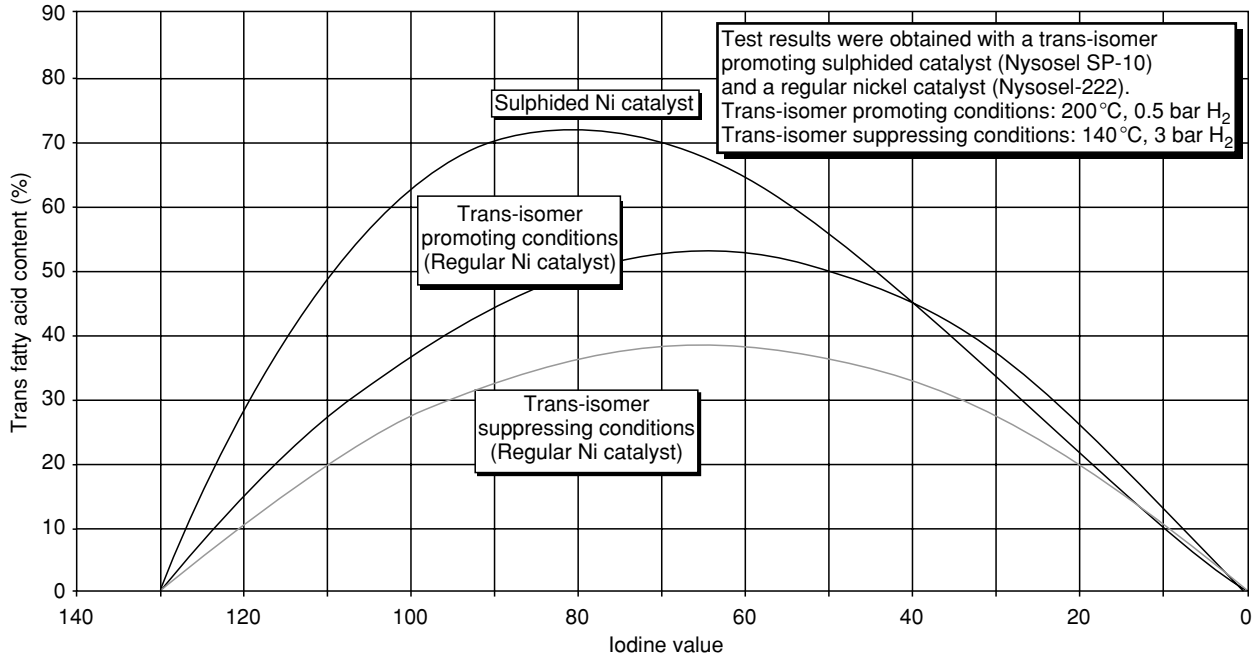

Figure 4 Hydrogenation of soyabean oil using sulfided and nonsulfided nickel catalysts.

cooling of the oil to be fractionated. Separation of the crystallized material from the liquid phase is generally carried out by filtration. The occlusion of liquid oil in the crystallized material means that the yield of filtrate is reduced, which is an important consideration if the filtrate is the primary fraction. Where the filter cake represents the primary fraction, this occlusion reduces its 'purity,' which has an adverse effect on its quality if this is required for specialized applications, e.g., confectionery fats.

The replacement of vacuum filters by the membrane filter press in the 1980s reduced entrainment of liquid oil in the filter cake, but membrane filter press separation at 6 bar still leaves a cake containing 40–60% liquid oil.

When fractionating expensive fats, it becomes commercially feasible to fractionate oils in solvents such as acetone or hexane, as this reduces oil entrainment in the cake to less than 10% and at the same time allows crystallization to be carried out far more quickly.

Fractionation is also used in combination with hydrogenation in the case of soyabean oil. In this case, the conversion of linolenic (C18:3) to linoleic (C18:2) acid is accompanied by the formation of some stearic acid. In order to produce an oil clear at ambient temperature, the hydrogenated oil is then fractionated.

Fractionation has the advantage over other oil modification processes of being a purely physical process, carried out in conditions that do not affect any of the components present beyond crystallizing and melting them. The mild process operating conditions also have the benefit of requiring only moderate processing costs, but the process has the disadvantage of depending on finding rewarding markets for its secondary fractions.

Dewaxing of sunflower oil may be regarded as a fractionation process, though in this case, the purpose of the process is to remove minor quantities of wax esters present in the oil, as these precipitate slowly at low temperatures and give the oil a cloudy appearance. The quantity of wax esters present in the oil depends on whether whole seed or only the kernels have been used for oil production, as more than 80% of the wax present in the whole seed is found in the hull. Oil from undehulled seed may contain more than 1500 p.p.m. wax, whereas oil from dehulled seed is likely to contain less than 500 p.p.m. Dewaxing is normally carried out by chilling the oil, which should have had its phospholipid content reduced to below 10 p.p.m., to approximately 0 °C over a period of 24 h in order to enable the wax to crystallize in discrete form. The subsequent filtration can be accelerated by the addition of a filter aid to the wax–oil slurry. The wax content of the processed oil should not exceed 30 p.p.m. in this case. If a less stringent specification applies, wax separation by centrifugation is feasible.

Interesterification

The fatty acid groups in triacylglycerols can be rearranged within a single oil (intraesterification) or by exchange of fatty acid groups with those of other oils, the latter being known as interesterification. As is the case with the other oil modification processes, the purpose of interesterification is to alter the melting properties of the fat or fat blend in order to improve the functional properties of the product. The reaction requires the presence of a catalyst, and alkaline catalysts are generally used. In particular, sodium methylate, also referred to as sodium methoxide, is extensively used, but metallic sodium has also been used.

In the early stages of the development of this process, both directed interesterification and random interesterification were proposed for use, the former involving interesterification and simultaneous cooling in order to crystallize the saturated triacylglycerols being formed and thus move the equilibrium towards the formation of more saturated triacylglycerols. Random interesterification, however, simply aims to randomize the distribution of the available fatty acid groups on the glycerol molecule and thereby to alter the melting properties of the fat or oil blend.

The aggressive nature of the catalyst makes it essential that the moisture content and the FFA content of the oil to be processed are reduced to the lowest practical values, as hydrolysis of the catalyst results in significant oil losses. It is also of crucial importance that the catalyst is handled with great care by the process operators, as exposure to it is likely to inflict serious injury.

Interesterification (Figure 5) has now become an important tool in the use of fully hardened oils, as this enables an oil processor to minimize the content of TFA in an oil blend. The process finds application in the production of fat blends for use in spreads and of bakery fats, as it facilitates the tailoring of the melting properties of fats to specific requirements (Table 9). It is also used in the production of cocoa-butter replacers.

Immobilized lipases have been used successfully to achieve interesterification on an industrial scale. Lipases have been developed that will retain activity at temperatures up to 80 °C, but increased

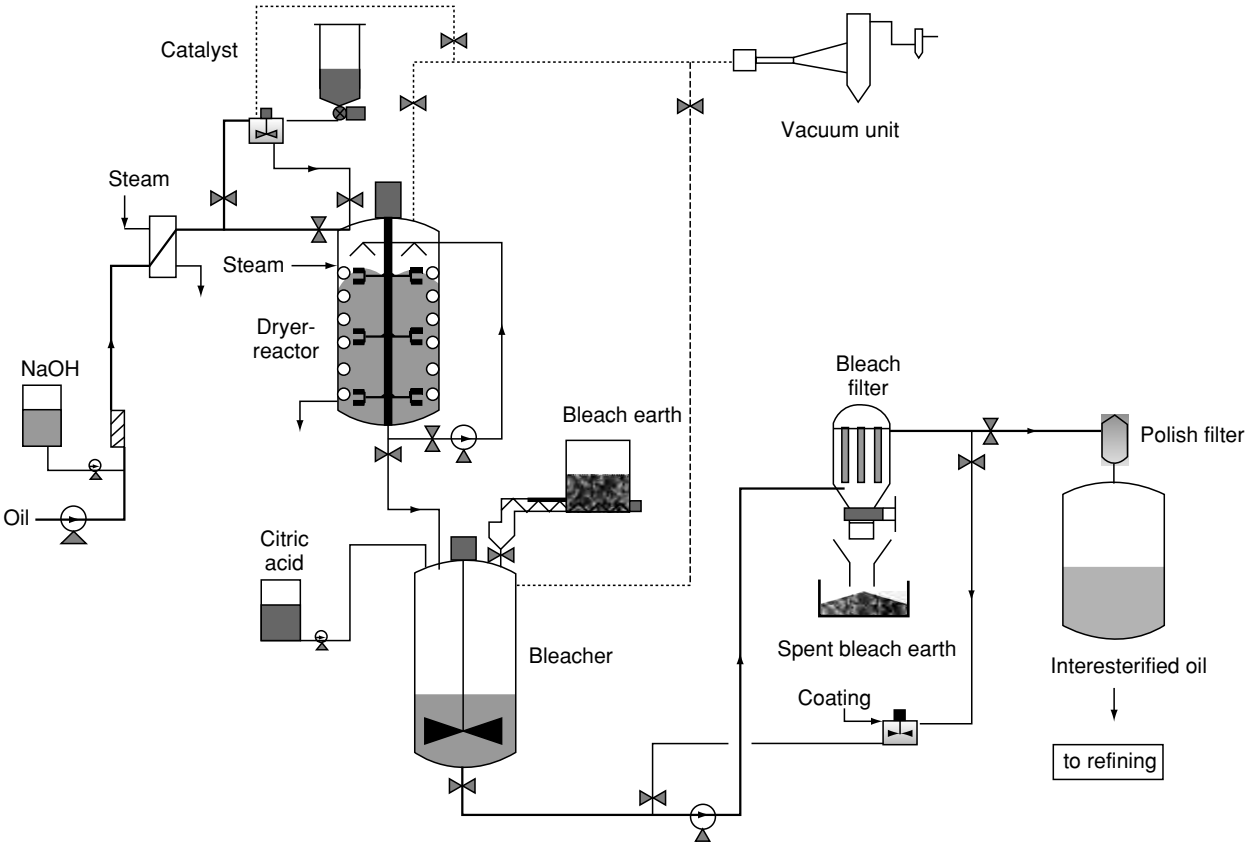

Figure 5 Interesterification plant arrangement.

Table 9 Solid fat content of fat blend (canola/soya stearin 70:30 w/w) before and after interesterification

Catalyst (sodium methoxide) concentration	0.3%				
Temperature	110°C				
Pressure	13 mbar				
	Solid fat content (%)		*Triacylglycerol composition (major components only)*		
Temperature (°C)	*Before interesterification*	*After interesterification*		*Before interesterification*	*After interesterification*
10	42	47	LOO	11.6	5.3
21.1	38	25	OOO	45	17.8
26.7	38	22	OOS	1.6	33.3
33.3	37	15	SSS	24.1	2
40	35	9			

L, linoleic acid; O, oleic acid; S, stearic acid.
From *Journal of the American Oil Chemists' Society* 76: 783 (1999).

temperatures result in higher FFA and diacylglycerol contents, which either require removal by refining or can adversely affect the product properties. Although reuse of the immobilized enzyme has been shown to be feasible, the cost of the immobilized enzyme remains too high for general applicability of enzymic interesterification. Present use of this process appears to be confined to high-value fats.

Applications

The principal use for oils and fats is in the form of liquid oil, either as salad oils or as frying oils. A salad oil must principally display oxidative stability and clarity in appearance, and a wide range of refined oils can meet this requirement. Olive oil is the most obvious example of an oil satisfying these criteria without having been refined, but other oils are sometimes also used in the unrefined state. Frying oils must be resistant to oxidation at high temperature, a high smoke point, and a minimum tendency to foam in use. In the case of oils used for industrial frying, oil cost is an important consideration, and refined, bleached, deodorized palm olein is widely used for this purpose in Europe and various Asian countries. Lightly hydrogenated soyabean oil is the first choice for industrial frying in the USA.

Margarine and spreads have been a major source of fat supply to the Western consumer. In the second half of the twentieth century, the fat component of margarine has increasingly been softened (by skillful use of the various modification processes discussed above) in order to make the product spreadable at refrigerator temperatures. Spreads have a lower fat content than margarine, and are rapidly displacing them. The scraped surface heat exchanger, which replaced the rotary cooler in the production of margarines, has played an important role in the growth of the margarine and spreads industries.

Fats for baked products are widely used both domestically and industrially, with the latter category dominant. In some cases, these fats are blends of individually partially hardened fats, blended to achieve melting properties suitable for a specific application. Emulsifiers may be added to the blend in order to facilitate the incorporation of air in an application.

Confectionery fats constitute a minor, but high-value, part of the total applications field. Their special characteristics of melting over a narrow temperature range has meant that much work has gone into identifying fats suitable for this application. They are characterized by their high content of symmetrical triacylglycerols, obtained by fractionation in the case of cocoa-butter equivalents (CBEs). Cocoa-butter substitutes, however, are based on the oils rich in C14 fatty acids. As these are incompatible with C18-rich fats in crystallized form, care must be taken to exclude the lauric oils from CBEs.

Nutritional aspects of oil and fat consumption can be considered in terms of three characteristics – the type of fatty acid present in the oils and fats consumed, the modification of fatty acids by processing, and the effect of processing on minor components. Clearly, the ongoing debate about healthy eating has favored the consumption of more liquid oils (and thereby a reduction in the consumption of hardened fats) in regions where such choices exist. The choice between polyunsaturated and monounsaturated fatty acids constitutes a second level of consumer choice. The virtual elimination of erucic (C22:1) acid from rapeseed oil grown

in Europe and North America has vastly increased the quantity of this oil suitable for use in the food world. Essential fatty acids are present in only a few vegetable oils, and are often removed by hydrogenation in the interests of oil stability.

The most important modification of fatty acids present in vegetable oils in the course of processing is the formation of TFA. Hydrogenation is the main source of TFA in edible oils, and earlier sections consider some of the changes that have been made in order to reduce the formation of TFA. Although it is generally only a minor contributor to TFA content, high-temperature deodorization (at temperatures of $>250\,^{\circ}C$) does produce measurable quantities of these fatty acids. High-temperature deodorization may also result in an increase in polar lipids, including polymers.

Tocopherols present in crude oils are lost to some extent in the bleaching stage as well as in deodorization, but the latter stage is responsible for the major share of the loss, particularly where very high temperatures are used for this operation. In Europe, deodorization temperatures have therefore been modified in order to reduce this loss. When using chemical refining, the deodorizer distillate is easier to process for tocopherol recovery, and in North America – where chemical refining is preferred to physical refining – addition of tocopherol concentrate to oils is sometimes used to compensate for the loss of tocopherols in the deodorizer.

See also: **Fats**: Production of Animal Fats; Uses in the Food Industry; Digestion, Absorption, and Transport; Requirements; Fat Replacers; Classification; Occurrence; **Phospholipids**: Determination; Physiology; **Vegetable Oils**: Types and Properties; Composition and Analysis; Dietary Importance

Further Reading

Bokisch M (1998) *Fats and Oils Handbook*. Champaign, IL: AOCS Press.

Codex Alimentarius Commission (1999) *ALINORM 99/17 Report of the Sixteenth Session, London, March 1999*.

Erickson DR (ed.) (1995) *Practical Handbook of Soybean Processing and Utilization*. Champaign, IL: AOCS Press.

Gunstone FD and Padley FB (eds) (1997) *Lipid Technologies and Applications*. New York: Marcel Dekker.

Hamm W and Hamilton RJ (eds) (2000) Edible Oil Processing. Sheffield Academic Press.

O'Brien RD (1998) *Fats and Oils: Formulating and Processing for Applications*. Lancaster, PA: Technomic.

Rajah KK (ed.) (in press) *Fats in Food Technology*. Sheffield Academic Press.

Composition and Analysis

E W Hammond, Greenisle-Consulting, Kettering, UK

Analysis

Free Fatty Acids

Free fatty acids (FFA) are not present to any significant level in healthy plant cells. Any fatty acid present is esterified to lipid or associated in some way with protein or other cell constituents. However, FFA at up to 5% (15% in very bad cases) might be found in commercial crude vegetable oils whether pressed or solvent-extracted. The FFA is present as a consequence of cell damage in the vegetable tissue during harvesting, storage, transport, or initial processing. It is a particular problem for the edible oil industry as it must be removed by refining. The cell damage can be caused by bruising of the fruit or seed, freeze–thaw cycling or hot and humid conditions or microbial activity (e.g., fungal growth). The FFA level can also rise during storage of the extracted oil, if the conditions of storage are poor involving humid and warm conditions. The consequence of fruit or seed damage can be easily demonstrated by carefully extracting the lipids from very freshly harvested oil seed. The composition of the lipids is compared with that of an extract of the same batch of seed after it has been frozen at $-20\,^{\circ}C$ for several hours and then left to thaw at room temperature for three or more hours. The freeze–thaw action causes cell damage and the release of hydrolytic enzymes, with resultant lipid hydrolysis and elevated FFA. This can be particularly dramatic in some oil seeds and also can lead to the presence of fatty acids that are not in the native unspoiled oil. Such an example is the seed from *Vernonia anthelmintica* which has (+)*cis*-12,13-epoxyoleic acid (*vernolic acid*) naturally as part of its composition. If the seeds are damaged either by crushing or freeze–thaw action and then left in warm moist conditions, even for a short time, (+)*threo*-12,13-dihydroxyoleic acid is found increasingly to be present at the expense of the vernolic acid. Cell damage causes the release of a lipase and an epoxide hydratase enzyme. While this is an unusual case, it serves to demonstrate the care required when harvesting, storing, and processing vegetable oil crops, particularly so when research is to be undertaken on the lipids of the seeds or fruits. (*See* **Fatty Acids**: Properties.)

The determination of total and individual FFA is covered in detail under the entries on Fatty Acids and will not be repeated here. (*See* **Fatty Acids**: Analysis.)

Fatty Acid Composition

Refined vegetable oils are virtually 100% triacylglyceride (TAG). The analysis of the fatty acid (FA) composition of TAG is covered under Triacylglycerides and Fatty Acids entries. These sections cover the FAs typically found in edible oils including methods of derivatization for chromatography. (*See* **Triglycerides**: Structures and Properties; **Triglycerides**: Characterization and Determination.)

The FAs in most vegetable oils are straight-chain homologous series of saturated and unsaturated carboxylic acids of even-numbered chain lengths from C_8 to C_{24}. The unsaturated series contains one to three methylene-interrupted, *cis* double bonds. Some vegetable oils contain nontypical FAs, but their presence in nutritious edible oils is usually undesirable. For this reason, their analysis is not covered here. A broad review of the types and occurrence of these other FAs can be found in *The Lipid Handbook* (see Further Reading).

Triacylglycerides

The general analysis of TAG is covered under the entry Triglycerides. Some detail on carbon number analysis by nonpolar high-temperature gas chromatography is given there. This technique produces good quantitative information on the overall TAG composition, giving a 'finger print,' which, if linked with fatty acid analysis, enables the determination of fat type. The demanding and detailed analysis of isomeric TAG composition is described here. However, it should be stated that, to obtain the best information, some experience of interpretation and a good knowledge of technique are essential.

High-temperature Polar-phase Gas Chromatography

Polar phases for GC such as the polysiloxanes have a greater selectivity than the nonpolar silicone phases; however, they are also less resistant to high temperatures (>250 °C). This limits their use for lipids of higher molecular weights. TAG have a molecular weight of about 900 and require a temperature of up to 370 °C for efficient chromatography on the nonpolar silicone phases such as OV1, OV101, or SP2100. (*See* **Chromatography**: Gas Chromatography.)

Much excellent work has been published on the application of high-temperature, polar phenylmethyl silicone phases to the analysis of TAG. In this technique, a 25 m × 0.34 mm fused silica column is used, with a bonded phase of 65% phenylmethyl silicone at 0.1 μm film thickness. The separation is in groups of increasing molecular weight, but the groups are further well resolved into both positional isomers and TAG of increasing unsaturation. Quantitation should be expected to be good but is complicated by unequal recoveries due to losses on the column. These losses increase with unsaturation of the TAG and can be severe with highly unsaturated TAG. One has to assume that the phase is responsible since such losses are not seen in the nonpolar phase application. However, reducing the column length to 15 m reduces losses. *Calibration must be used*.

A 15 m × 0.34 mm fused silica column with an 0.1 μm bonded film of 65% phenylmethyl silicone, fitted with a 1 m × 0.53 mm fused silica retention gap is typical for reasonable quantitative analysis. Hydrogen (linear velocity 40 cm s^{-1}) is the preferred carrier gas, although helium (linear velocity 20 cm s^{-1}) will provide good results, but with longer retention times and the possibility of higher losses of the polyunsaturated TAGs. Typically, the initial column temperature is 70 °C with a 2-min hold, then programmed to 320 °C at a 6 °C min^{-1} rise.

Reversed-phase High-performance Liquid Chromatography

This is a nonaqueous type of reversed phase (RP) high-performance liquid chromatography (NARP-HPLC). The resolution achieved is similar to that with the high temperature GC polar-phase technique, but the separation order is much more complex. Because of this, the interpretation of results can be difficult and requires a good table of standardized equivalent chain lengths obtained by analysis of TAG of known structure. The separation achieved is a function of molecular weight, level of unsaturation, and fatty acid position on the TAG. The presence of a double bond is virtually equivalent to a reduction in chain length of two carbon atoms. As an example, triolein (C54) elutes near to tripalmitin (C48). (*See* **Chromatography**: High-performance Liquid Chromatography.)

The separation is achieved using two 25 cm × 4 mm ID Merck Lichrocart Supersphere 100 RP-18 (end capped) columns linked in series. A guard column should always be fitted. The column should be maintained at a constant temperature according to what is being analyzed, but typically 22 °C. Two solvent mixtures are used; A = 10% methanol in acetonitrile and B = 10% tetrahydrofuran in dichloroethane. The start solvent ratio is 52%-A to 48%-B, pumped for 48 min (or an appropriate time to elute all peaks) at 1.0 ml min^{-1}. The system is then reset in between

each sample to a solvent ratio of 95%-A to 5%-B, pumped for 15 min at 1.0 ml min^{-1}. This column equilibration step and time are important for consistent retention times and separations. An HPLC system with an autosampler that will time this period accurately is a great advantage.

RP column packings are not tolerant of high sample loadings and are sensitive to small amounts of other solvents used to dissolve samples. Thus, sample sizes should not exceed 100 µg in 20 µl of solvent. If the sample solvent cannot be that used as the initial mobile phase for HPLC, then its behavior in modifying the column condition must be ascertained. Chloroform should be avoided as it often causes peak splitting, whereas alcohols and hexane create solubility problems. Toluene is a suitable general purpose solvent in this application. Typical chromatography data are shown in Figure 1.

Argentation High-performance Liquid Chromatography

Vegetable fats used for confectionery production, such as for chocolate, require fats with specific physical properties. Such a fat is cocoa butter (CB) and any of the CB equivalents derived by blending of native and fractionated fats of vegetable origin. The specific melting properties and solid/liquid ratios (measured by pulsed nuclear magnetic resonance and plotted as solid fat curves) are created by the presence of a high level (up to 90%) of 'symmetrical' monounsaturated TAG. This TAG has oleic acid (O) at position 2 and a saturated fatty acid, palmitic (P) or stearic (St), at positions 1 and 3, thus producing POP-, POSt-, and StOSt-type TAGs. Perhaps the most rapid method of obtaining quantitative evaluation of such a fat is via argentation HPLC. In this technique, the silver atoms form transient 'complexes' with the π electrons of the unsaturated centers (double bonds). Separation of TAG species depends upon the number of double bonds, their geometry (whether *cis* or *trans*), and the amount of hindrance caused to any 'complex' formation. Thus, monoenes separate from dienes, from trienes, etc. 'Complexes' with *trans* double bonds are less efficient than those with *cis* double bonds and therefore also separate; *trans* being relatively less polar than *cis*. Where there is hindrance to 'complex' formation (due to fatty acid position on the TAG), separation may also occur. For instance, in StOSt, the double bond in oleic is more hindered than in StStO, resulting in StOSt being relatively less polar than StStO causing the two to separate. Figure 2 shows a typical example of the separation of TAG. (*See* **Cocoa**: Production, Products, and Use.)

To achieve the separations in Figure 2, a 20 cm × 4 mm column of LiChrosorb Si-60, 5 µm silica (Merck) coated with 5% silver nitrate, was used. This packing is not available commercially but can be prepared freshly for each column. The silver nitrate, in a foil-covered round flask, is dissolved in acetonitrile and diluted with methanol. The silica is slurried in this and the solvent carefully removed by

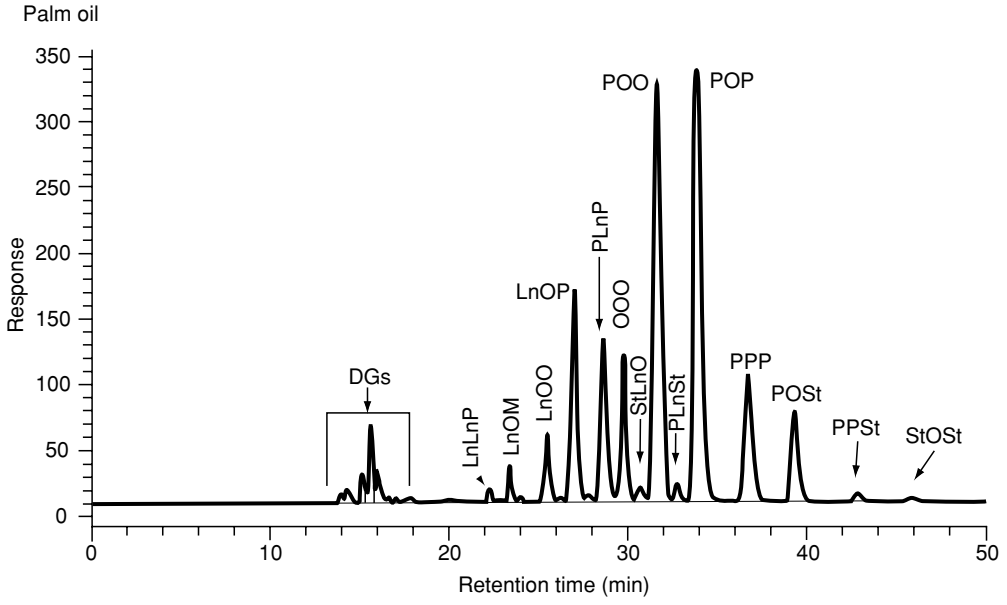

Figure 1 NARP-HPLC of refined, bleached, and deodorized Malaysian palm oil. St, stearic; P, palmitic; O, oleic; Ln, linoleic; M, myristic; DGs, diacylglycerides.

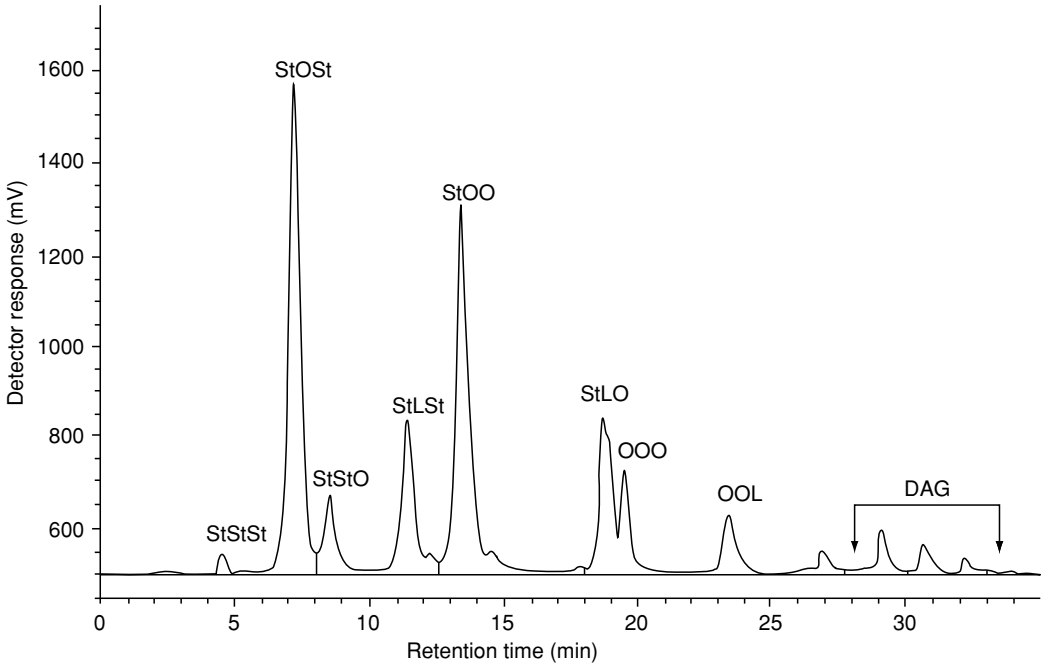

Figure 2 Argentation HPLC of refined palm oil triacylglycerides. The peak labeling corresponds to fatty acids in the order of occurrence on the glyceride: St, saturated (i.e., palmitic, stearic, arachidic); O, oleic; L, linoleic.

rotary evaporation, ensuring constant mixing. When the packing is apparently dry, the last traces of solvent are removed by heating the flask to 70 °C under vacuum for 10 min. This material is slurried in chloroform for column packing and the column packed to 28 MPa (4000 psi) with toluene for 45 min.

The sample size can be up to 600 μg in 30 μl of toluene. The three solvents for HPLC are:

- A = toluene/hexane (1:1, v/v);
- B = toluene/ethyl acetate (3:1, v/v);
- C = toluene with 10 ppm formic acid.

A concave solvent gradient is used from 8%-B to 90%-B in A over 28 min. The regeneration step is very important and uses solvent C for 12 min followed by 8%-B in A for 7 min. Solvent flow is 1.5 ml min^{-1}. Detection of the eluted TAG has to be done by evaporative system detectors such as the Tracor 945 (flame ionization type, not now commercially available), or light-scattering instruments such as the PL-ELS 1000 (Polymer Laboratories). It should be noted that the quantitative behavior of the two instrument types is different. The response curves of the Tracor 945 are essentially linear, while those of light-scattering instruments are curved and sigmoid. Calibration is very important if reliable data are to be obtained, and this should be part of a written procedure. The analysis time is relatively long with a turnaround time of about 40 min per sample, but improvements have been made upon this.

Minor Components

Crude vegetable fats are usually at least 95% TAG. After accounting for any FFA, other minor components can be present, which can include tocopherols, carotenes, gums, waxes, polar lipids (lecithins), unsaponifiable matter, chlorophyl pigments, hydrocarbons, and mineral constituents (present possibly as metal soaps of fatty acids). Refer to individual constituents.

There are standard methods published by the American Oil Chemists' Society (AOCS) and the Association of Official Analytical Chemists (AOAC) for the analysis of the above components. The procedures are fully documented and tested and may be updated at suitable periods, so they will not be reproduced here. However, it is worth mentioning that some of the older published methods have escaped improvement, and more up-to-date procedures can often be found.

For instance, tocopherols have traditionally been estimated as a group, after isolation from the nonsaponifiable fraction of an oil, by reaction with α,α-dipyridyl and colorimetric analysis. However, it is known that the tocopherol composition can aid the identification of a fat and also that different tocopherol compositions lead to different stabilities of a fat to oxidation. Accurate analysis of individual tocopherol compositions can be readily done by HPLC. Essentially, the extracted and dry

fat is dissolved in the mobile phase solvent. A 15 cm × 4 mm column packed with 5-μm Chromegasphere silica (E.S. Industries, Marlton, NJ) is used with a fluorescence detector (e.g., Kontron SFM25 with a flow cell). Excitation is set at 294 nm and emission at 325 nm. The HPLC system is isocratic using a solvent of 3% THF in 2,2,4-trimethyl pentane (iso-octane) as the mobile phase at a flow of 1.5 ml min^{-1}. The sample is applied via a 25-μl loop and Rheodyne valve. Complete separation of tocopherols and tocotrienols is achieved (Figure 3). The column requires flushing with 10% methanol in chloroform every 200 samples to remove any build-up of polar lipid material. Figure 3 shows a typical chromatogram from the analysis of palm oil.

Authentication of Oils

The purity of composition of a natural vegetable oil may be extremely important to the ultimate user. Fats or oils can be among the most expensive ingredients used in the manufacture of foods. Thus, there are situations where, for reasons of unfortunate cross-contamination or adulteration for profit, it becomes important to know the authentic composition of a natural oil. (*See* **Adulteration of Foods**: History and Occurrence.)

The chemical composition of natural vegetable oils is readily obtained by chromatographic analysis. Information on structure can also be obtained by specific enzyme techniques, and minor components can be analyzed either by chemical techniques or again by

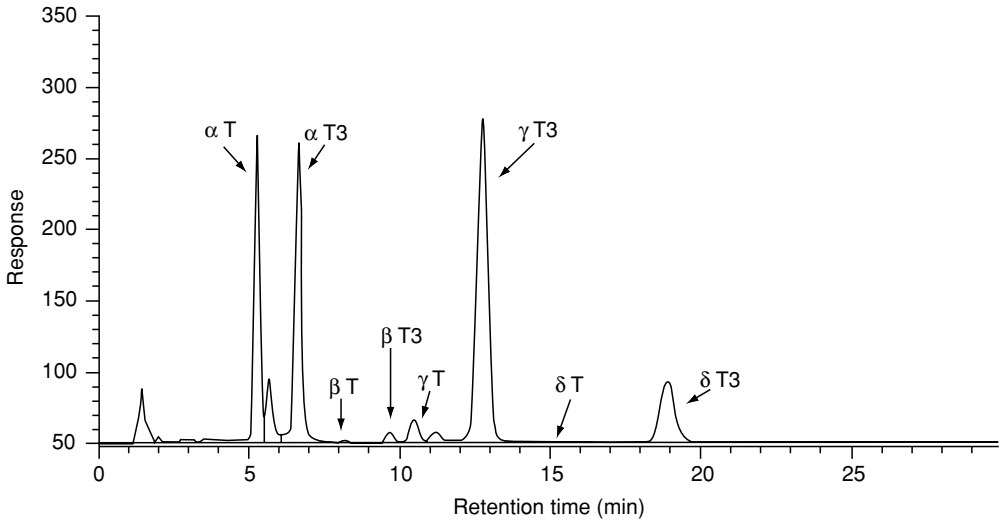

Figure 3 HPLC analysis of palm oil tocotrienols and tocopherols (conditions are described in the text).

Table 1 Fatty acid composition of a range of rapeseed oils

Sample name	*Fatty acid (% composition)*														
	12:0	*14:0*	*16:0*	*16:1*	*18:0*	*18:1*	*18:2*	*18:3*	*20:0*	*20:1*	*20:2*	*22:0*	*22:1*	*24:0*	*24:1*
German, Winter		0.1	4.2	0.3	1.9	59.7	20.2	8.9	0.5	2.0	0.1	0.3	1.5	0.1	0.2
UK, high yield		0.1	5.1	0.4	1.5	59.7	21.0	9.5	0.6	1.4	0.0	0.3	0.4	0.0	0.0
Denmark, Spring		0.0	4.7	0.3	1.4	57.5	21.7	10.9	0.5	1.6	0.1	0.4	1.0	0.0	0.0
France, Tapidor		0.1	4.7	0.3	1.9	59.4	21.0	9.5	0.5	1.5	0.0	0.4	0.7	0.0	0.0
USA, not known		0.0	4.0	0.2	2.1	61.2	20.0	9.5	0.6	1.5	0.1	0.3	0.5	0.0	0.0
China, high erucic		0.1	2.8	0.1	1.3	17.1	13.2	8.5	0.7	9.0	0.2	1.2	44.0	0.8	1.1
China, high erucic		0.1	1.8	0.3	1.7	16.2	12.0	8.3	0.8	9.0	0.2	1.3	46.0	0.7	0.6
Canada, not known		0.1	4.1	0.5	1.7	58.9	21.2	9.7	0.5	1.7	0.0	0.3	1.0	0.2	0.1
Europe, high erucic		0.1	4.0	0.3	1.0	18.5	14.5	11.0	0.8	6.5	0.8	0.5	41.0	1.0	0.0
UK, high yield		0.1	4.0	0.3	1.6	61.5	20.0	10.0	0.5	1.0	0.0	0.3	0.5	0.2	0.0
UK, not known		0.0	4.9	0.3	1.7	59.4	20.8	9.9	0.6	1.4	0.0	0.4	0.3	0.2	0.1
Laurical (GM)[a]	39.9	3.9	2.7	0.1	1.8	30.1	11.0	9.1	0.2	0.8	0.4	0.1	0.0	0.0	0.0
High oleic, UK test crop	0.1	0.2	3.9	0.4	2.4	74.0	9.9	7.0	0.1	1.4	0.1	0.3	0.2	0.1	0.0

[a]Genetically Modified Crop (trade name of Calgene).

chromatography. All this information characterizes the sample being analyzed, but does not necessarily apply exactly to another oil of the same species from a different part of the world. Also, it is important that data be obtained via more than one technique, since actual differences in oil composition may only become apparent with certain techniques.

Within a single species of plant, for instance rapeseed, there can be a number of varieties. It is almost certain that the oil composition of the varieties will be different. This statement is even more important now that genetic modification of oil composition is being done. Also, if one variety is grown in different environments, created either by countries or even by areas of a single country, these situations can lead to oils of variable composition. The differences may be small or large, but they will provide statistical range bands (with statistical means) within which any particular oil composition might lie. This shows that to authenticate a particular oil unambiguously, there must be sufficient world data on the composition as measured by a number of techniques. Table 1 shows typical analytical data that we have found for a range of rapeseed crops. Further data about many vegetable oils are published in *The Lipid Handbook*.

See also: **Adulteration of Foods**: History and Occurrence; **Carotenoids**: Occurrence, Properties, and Determination; **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **Cocoa**: Production, Products, and Use; **Contamination of Food**; **Fats**: Classification; **Fatty Acids**: Properties; Analysis; **Gums**: Properties of Individual Gums; **Phospholipids**: Properties and Occurrence; **Tocopherols**: Properties and Determination; Physiology; **Triglycerides**: Structures and Properties; Characterization and Determination

Further Reading

AOCS (1984) *Recommended Practices*, 3rd edn. Champaign, IL: American Oil Chemists' Society.

AOAC (1984) *Official Methods of Analysis*. Washington, DC: Association of Official Analytical Chemists.

Christie WW (1982) *Lipid Analysis*, 2nd edn. Oxford: Pergamon Press.

Christie WW (1989) *Gas Chromatography and Lipids – A Practical Guide*. Ayr, UK: The Oily Press.

Gunstone FD, Harwood JL and Padley FB (eds) (1986) *The Lipid Handbook*. New York: Chapman & Hall.

Hammond E (1988) Flame ionisation detection for the HPLC of lipids. *Chromatography and Analysis* 1: 7–9.

Hammond EW (1993) *Chromatography for the Analysis of Lipids*. Boca Raton, FL: CRC Press.

Hammond EW (1998) Analysis – Tocopherols and Tocotrienols. *Lipid Technology* 10: 86–88.

Jeffrey BSJ (1991) Silver-complexation liquid chromatography for fast high-resolution separations of triacylglycerols. *Journal of the American Oil Chemists' Society* 68: 289–293.

Dietary Importance

J Dupont, United States Department of Agriculture, Beltsville, MD, USA

Background

Vegetable oils are derived from oil seeds grown mainly for their oil. The major food oils consumed as cooking and salad oils are canola (rapeseed), corn, cottonseed, olive, palm, palm kernel, coconut, peanut, safflower, soya bean and sunflower. They are all triacylglycerols that are liquid at room temperature, but differ greatly in fatty acid composition (**Table 1**). They are classified as short-chain ($C_{6–8}$), medium-chain ($C_{10–12}$), and long-chain ($C_{14–18}$). The plant oils may be saturated (completely hydrogenated) or have one, two, or three double bonds of the *cis* configuration. (*See* **Fatty Acids:** Properties.)

Energy Value

Long-chain acylglycerols provide $38\,kJ\,g^{-1}$, whereas protein and carbohydrate provide $17\,kJ\,g^{-1}$. Because of the difference in energy value of oils as compared with the other energy-yielding nutrients, concentration in the diet is often expressed as a percentage of energy (en%). Vegetable oils are highly digestible (greater than 95%). They are hydrolyzed in the gastrointestinal tract, yielding free fatty acids (FFAs) and *sn*2-monoacylglycerols. By formation of bile micelles, the FFAs and monoacylglycerols come into contact with the mucosal cell membranes and diffuse through the membrane. Within the mucosal cells, they are reesterified to triacylglycerols. They are made water-soluble by incorporation into chylomicrons in the mucosa.

Medium-chain fatty acids occur in triacylglycerols (medium-chain triacylglycerols; MCT) of coconut and palm kernel oils. The MCTs are much more water-soluble than are long-chain triacylglycerols. They are hydrolyzed and enter the mucosal cells but are partially transported as FFAs in the portal circulation.

Chylomicrons and albumin-bound fatty acids are transported to the liver, adipose tissue, and other organs. In the process, chylomicrons are hydrolyzed

Table 1 Percentage fatty acid composition of refined vegetable oils

Oil	*Fatty acid*[a]													
	6:0	*8:0*	*10:0*	*12:0*	*14:0*	*16:0*	*16:1* n-7	*18:0*	*18:1* n-9	*18:2* n-6	*18:3* n-3	*20:1*	*20:4* n-6	*22:1*
Canola[b]						4		2	55	26	10			2
Rapeseed, LEAR[c]						4.8	0.5	1.6	53.8	22.1	11.1	1.0		0.1
Soya bean					0.1	10.3	0.2	3.8	22.8	51.0	6.8	0.2		
Corn						10.9		1.8	24.2	58.0	0.7			
Olive						11.0	0.8	2.2	72.5	7.9	0.6	0.3		
Cottonseed					0.8	22.7	0.8	2.3	17.0	51.5	0.2		0.1	
Safflower						4.8		1.3	75.3	14.2				
Linseed						5.3		4.1	20.2	12.7	53.3			
Palm				0.1	1.0	43.5	0.3	4.3	36.6	9.1	0.2	0.1		
Palm kernel	0.2	3.3	3.7	47.0	16.4	8.1	2.8	11.4	1.6					
Coconut	0.6	7.5	6.0	44.6	16.8	8.2		2.8	5.8	1.8				
Peanut					0.1	9.5	0.1	2.2	44.8	32.0		1.3		

[a]Number of carbons: number of double bonds, location of distal (from *n* or ω) double bond.
[b]From Canola Council of Canada (1987) *Canola Oil–Properties and Performance. Publication No. 60*. Winnipeg: Canola Council of Canada.
All other data from USDA (1979) *Composition of Foods. Agriculture Handbook 8-4*. Washington, DC: US Department of Agriculture.
[c]Low-erucic-acid rapeseed oil.

by the action of lipoprotein lipase, and the fatty acids enter the cells by diffusion and are reesterified to triacylglycerols. During fasting, the triacylglycerol fatty acids are released from adipose tissue by hydrolysis for use as energy. Fatty acids that enter the liver are resecreted as very-low-density lipoproteins, which are hydrolyzed similarly to chylomicrons, yielding low-density lipoproteins (LDLs), which are taken up by tissues via a receptor-mediated process. The LDLs are hydrolyzed to yield fatty acids for energy or reesterification and storage. (*See* **Lipoproteins**.)

Fatty acids are used for energy production in the mitochondria and peroxisomes of cells. Saturated fatty acids (SFAs) are oxidized by β oxidation to yield acetyl coenzyme A (acetyl CoA), which enters the tricarboxylic acid cycle. Unsaturated fatty acids require two more enzymatic steps that are needed for SFAs to change the existing double bonds from *cis* to *trans* and to move them from a γ to a β position. Even so, oxidation of C18:2 *n*-6 is more rapid than that of C18:1 *n*-9. Medium-chain triacylglycerol acids are preferentially oxidized by peroxisomes rather than mitochondria. Some of the acetyl CoA produced is oxidized to hydrogen peroxide, which requires catalase for dismutation to water and oxygen, yielding heat. This reduces the yield of these fatty acids to less than the 37.66 kJ g^{-1} assumed for fats in general. (*See* **Fatty Acids**: Metabolism; Dietary Importance.)

Table 2 Relationship between dietary linoleic acid and biological effects

Dietary concentration of linoleic acid (en%)		*Reported biological effects*
0		Rat: dermatitis, growth retardation, infertility, cellular and humoral immune system defects
		Human: dermatitis
1	Deficient	Rat: increase in triene: tetraene ratio to 0.4, decrease in eicosanoid production
		Mouse: immune system impaired, plaque formation
		Human: increase in triene: tetraene ratio to 0.4, decrease in clotting time
2–3		Rat: maximal amount of C20:4 *n*-6 in cellular lipids, increased eicosanoid production
		Mouse: normal plaque formation
	Saturation	Human: increased plasma cholesterol, decrease in triene: tetraene ratio to 0.1
4		Rat, guinea-pig, mouse: normal mitogenic response
5		DMBA rat: mammary cancer
8		Human: decreased plasma cholesterol
10		Rodents: delayed hypersensitivity normalized
		NMU rat: mammary neoplasia
	Excess	Guinea-pig: decreased mitogenic response
15		Rat: prostatic neoplasia, rat colon cancer

Essential Fatty Acids

Vegetable oils generally have a relatively high concentration of polyunsaturated fatty acids (PUFAs) (Table 1). Exceptions are coconut, palm, and palm kernel oils, which are primarily saturated, and olive and canola oils, which have relatively less PUFAs and more monounsaturated fatty acids. The most abundant PUFA is linoleic acid (C18:2 *n*-6), with canola and soya oils having appreciable concentrations of α-linolenic acid (C18:3 *n*-3). Because these fatty

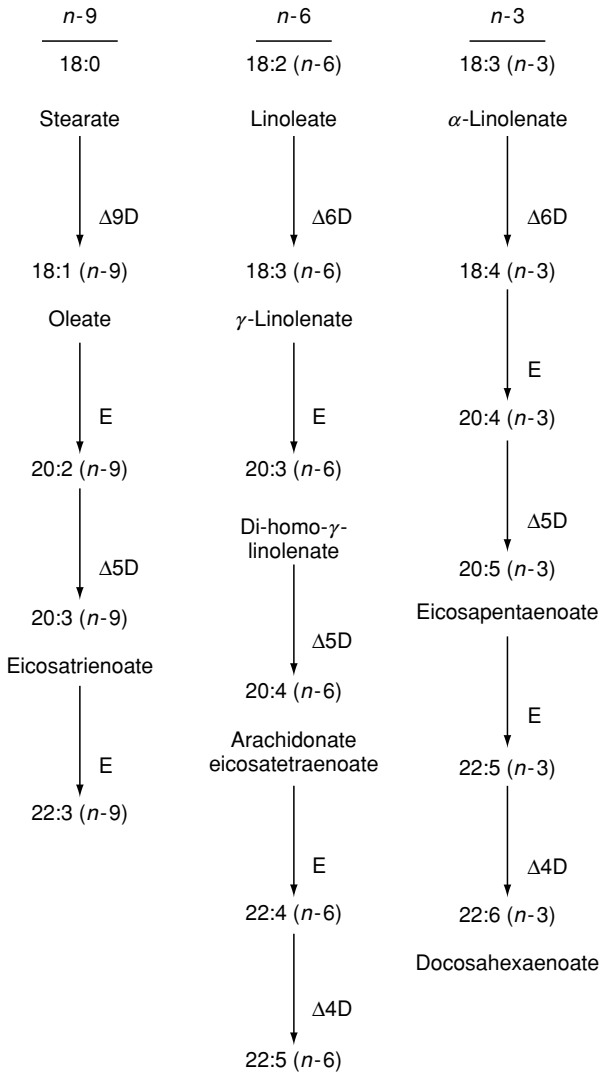

Figure 1 Families of fatty acids formed from C_{18} precursors by desaturation (D) and elongation (E). The *n*-9 pathway is exhibited in mammals only when there is insufficient *n*-6 present in the diet. Reproduced from Vegetable Oils: Dietary Importance, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

acids (1) cannot be synthesized by animal cells and (2) have essential metabolic functions, they are deemed dietary essential fatty acids. Table 2 summarizes current knowledge of the effects of different dietary levels of linoleic acid. (*See* **Essential Fatty Acids.**)

Linoleic acid is a component of ceramides, which are necessary to maintain the transdermal water barrier of the epidermis. It is the precursor by elongation and desaturation (Figure 1) of arachidonic acid (C20:4 *n*-6). Arachidonate is the precursor of the eicosanoids: prostaglandins, thromboxanes, prostacyclin, and leukotrienes (Figure 2). The eicosanoids are autocrine and paracrine factors, which are physiologically active at concentrations of pico- and nanograms. They are required for functions of the cardiovascular, renal, gastrointestinal, immune, and reproductive systems. (*See* **Prostaglandins and Leukotrienes.**)

Many studies indicate that essential fatty acids are necessary for normal immune function. Essential fatty acid deficiency impairs B- and T-cell-mediated responses. These impairments are normalized by inclusion of essential fatty acids in the diet. Most studies have used rodents, and the results cannot be readily extrapolated to humans because the immune systems of different species respond differently to dietary fat alterations. (*See* **Immunology of Food.**)

α-Linolenic is the precursor for eicosapentaenoic acid (C20:5 *n*-3) and docosahexaenoic acid (C22:6 *n*-3) that are required for brain and retinal development and function in animals.

Vitamin E

The vegetable oils that contain high concentrations of PUFA generally contain appreciable quantities of α- and γ-tocopherols (vitamin E) (Table 3). Vitamin E is a membrane-bound, lipophilic antioxidant closely associated with arachidonic acid in the membrane. Vitamin E may protect membranes by trapping lipid peroxyl free radicals produced from PUFA under conditions of oxidative stress. (*See* **Tocopherols:** Physiology.)

Table 3 Vitamin E content of fats and oils (mg per 100 g)[a]

Oil or fat	*α-Tocopherol*	*γ-Tocopherol*	*Total*[b]
Butter	1.76 ± 0.41	0	1.76
Coconut	Trace	Trace	
Corn	11.28 ± 6.25	50.76 ± 19.00	62.04
Cottonseed	32.46 ± 13.10	29.91 ± 13.76	62.37
Lard	1.20	0.70	1.90
Olive	7.94 ± 6.58	Trace	7.94
Palm	22.79 ± 2.86	28.26 ± 1.06	51.05
Peanut	15.12 ± 4.41	23.20 ± 6.90	38.32
Rapeseed	25.82 ± 8.06	30.01 ± 12.73	55.83
Safflower	39.23 ± 3.74	15.76 ± 4.40	54.99
Sesame	17.62 ± 5.69	32.86 ± 5.46	50.48
Soya bean	10.25 ± 5.63	50.48 ± 20.75	60.73
Sunflower	48.70	5.10	53.80

[a]Data are expressed as the mean plus or minus the standard deviation and are from Dupont J, White PJ, Carpenter MP *et al*. (1990) Food uses and health effects of corn oil. *Journal of the American College of Nutrition* 9: 438–470.
[b]Sum of α- and γ-tocopherol.

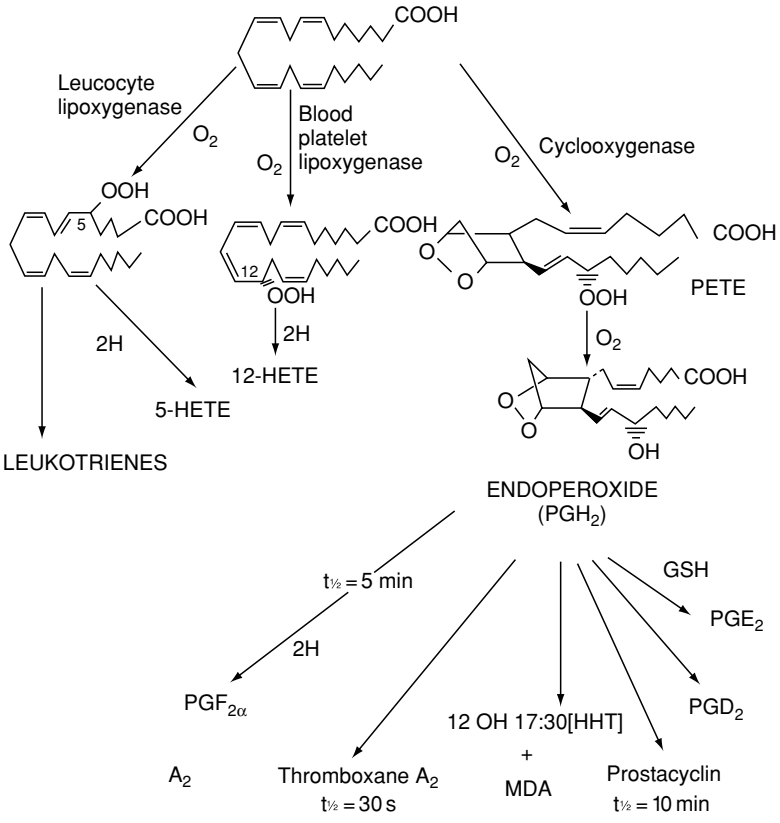

Figure 2 Eicosanoid synthesis from arachidonate. HETE, hydroxyeicosatetraenoic acid; PETE, peroxyeicosatetraenoic acid; PG, prostaglandin; GSH, glutathione; MDA, malondialdehyde; TX, thromboxane; PGI_2, prostacyclin. Reproduced from Vegetable Oils: Dietary Importance, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Cardiovascular Effects

Saturated fatty acid in the diet raises plasma cholesterol, and PUFA lowers it. Many published studies indicate that lowering plasma cholesterol decreases the population risk of heart disease. Levels of LDL-cholesterol are lowered by reducing SFA intake; the effect of concomitantly increasing PUFA is greater than that of merely reducing the SFA. Vegetable oils that increase the percentage of PUFA in the diet contribute to reducing plasma cholesterol. Enrichment of the diet with PUFA reduces blood pressure and has favorably influenced the progression of diabetic angiopathy in adult-onset diabetes mellitus. (*See* **Cholesterol:** Role of Cholesterol in Heart Disease; **Coronary Heart Disease:** Antioxidant Status.)

Oncogenic Aspects

Human surveys and animal experimentation support a role for total fat (excess energy) intake in the development of cancer, particularly in mammary tissue. The impact of linoleic acid intake on mammary and pancreatic tumorigenesis is not documented in human studies, nor is its impact on colon carcinogenesis as in animal studies. A hypothesis arising from human studies is that it is displacement of protective foods such as fruits, vegetables, pulses, and cereals from the diet by fat-rich foods that is detrimental, rather than any direct effects of fat in the diet. (*See* **Cancer:** Diet in Cancer Prevention; **Colon:** Cancer of the Colon; **Margarine:** Dietary Importance.)

See also: **Cancer**: Diet in Cancer Prevention; **Cholesterol**: Role of Cholesterol in Heart Disease; **Colon**: Cancer of the Colon; **Coronary Heart Disease**: Antioxidant Status; **Essential Fatty Acids**; **Fatty Acids**: Properties; Metabolism; Dietary Importance; **Immunology of Food**; **Lipoproteins**; **Margarine**: Dietary Importance; **Prostaglandins and Leukotrienes**; **Tocopherols**: Physiology

Further Reading

Dupont J (1990) Lipids. In: Brown M (ed.) *Present Knowledge of Nutrition*, 6th edn, pp. 56–66. Washington,

DC: International Life Sciences Institute Nutrition Foundation.

Dupont J, White PJ, Carpenter MP *et al*. (1990) Food uses and health effects of corn oil. *Journal of the American College of Nutrition* 9: 438–470.

Dupont J, White PJ and Feldman EB (1991) Saturated and hydrogenated fats in food in relation to health. *Journal of the American College of Nutrition* 10: 577–592.

VEGETABLES OF TEMPERATE CLIMATES

Contents

Commercial and Dietary Importance

S K Roy and A K Chakrabarti, Indian Agricultural Research Institute, New Delhi, India

Introduction

Vegetables are considered as protective foods supplying carbohydrate, protein, minerals, vitamins, and crude fiber, which are essential ingredients of a balanced diet. In addition to their role in nutrition, they increase the attractiveness and palatability of a diet by providing sensory appeal through their variety of color and flavor.

General Nutrient Composition

Vegetables in general, except for a few, are not considered to be the primary source of carbohydrate, protein, and fat. However, some of them with storage roots and tubers are rich in carbohydrate, particularly starch, in amounts comparable to the cereal crops, and the leguminous vegetables supply as much as 14% protein, while dry seeds supply still more. The lipid content in most vegetables is less than 0.1%. Most leafy vegetables and root crops are rich in minerals. Carrot and leafy vegetables are also high in carotene (provitamin A) and vitamin C. Deep-green leafy vegetables are rich in folate (folic acid). Besides, there are some trace elements required by the body like copper, manganese, and zinc, which act as enzyme cofactor. These are found in appreciable quantities in cole crops, roots crops, and some leafy vegeables. (*See* **Carotenoids:** Physiology.)

The nutrient content of a vegetable is dependent on several factors. The environment in which the crop is grown, such as temperature, light, moisture, nutrients, and physical and chemical properties of the soil, plays an important role. The amount of nutrient can vary with variety, cultural practices, stage of maturity, postharvest handling, and storage conditions. Once the vegetable is harvested, its composition goes on changing as a result of physiological and biochemical activities, which are natural processes.

In addition, methods of cooking also influence the nutrient content of vegetables; in particular, leaching losses and oxidative changes are considered as deteriorative processes. The composition of the same vegetable grown in different parts of the world is so different that it becomes difficult to use a general table of composition as a source of precise information of nutritive value of the crop used in the diet. However, such a table gives an estimate of the proximate composition of the nutrients. The nutrient composition of individual vegetables is given in **Tables 1** and **2**, and trace elements in **Table 3**. Refer to individual nutrients.

Table 1 Macronutrient composition of temperate vegetables per 100 g edible portion

Crop	*Energy (kcal)*	*Moisture (g)*	*Protein (g)*	*Fat (g)*	*CHO (g)*
Artichoke	79	77.3	3.6	0.1	16.0
Asparagus	26	91.7	2.5	0.2	5.0
Broad bean	48	85.4	4.5	0.1	7.2
Brussels sprout	45	85.2	4.9	0.4	8.3
Cabbage	24	92.4	1.3	0.2	5.4
Cardoon	10	94.0	0.7	0.1	1.8
Carrot	42	82.2	1.1	0.2	9.7
Cauliflower	27	91.0	2.7	0.2	5.2
Celery	17	94.1	0.9	0.1	3.9
Chicory leaf	13	92.0	1.7	0.3	1.1
Endive	11	95.0	1.3	0.2	1.2
Garden beet	43	87.7	1.7	0.1	8.8
Garlic	30	62.0	6.3	0.1	29.8
Kale	53	82.7	6.0	0.8	9.0
Knol khol	29	90.3	2.0	0.1	6.6
Knol khol green	43	86.7	3.5	0.4	6.4
Leek	77	78.9	1.8	0.1	17.2
Lettuce	14	95.1	1.2	0.2	2.5
Onion	50	86.6	1.2	0.1	11.1
Onion stalk	41	87.6	0.9	0.2	8.9
Parsley	16	90.0	2.2	0.3	1.3
Parsnip	53	81.0	1.2	0.3	11.6
Pea	84	78.0	6.3	0.4	14.4
Potato	97	74.7	1.6	0.1	22.6
Radish	17	94.4	0.7	0.1	3.4
Rutabaga	46	87.0	1.1	0.1	11.0
Spinach	26	90.7	3.2	0.3	4.3
Spinach beet	46	86.4	3.4	0.8	6.5
Sprouting broccoli	32	89.1	3.6	0.3	5.9
Swiss chard	16	92.0	1.8	0.2	1.5
Turnip root	29	91.6	0.5	0.2	6.2
Turnip green	67	81.9	4.0	1.5	9.4

Sources: Gopalan C, Sastri BVR, Balasubramanian SC, Rao BSN, Deosthale YG and Pant KC (1999) *Nutritive Value of Indian Foods*. Hyderabad: National Institute of Nutrition; Watt BK and Merril AL (1963) *Composition of Foods – Raw, Processed, Prepared*. Agricultural Handbook No. 8. Washington, DC: USDA.
CHO, carbohydrate.

Dietary and Commercial Importance

Apart from major nutrients, vitamins, and minerals, there are some phytochemicals present in vegetables which have been indicated by rapidly growing scientific evidence to play a major role in promoting health and in preventing certain diseases. These chemicals have often been referred to as phytonutrients or phytoceuticals. Besides, there are some key micronutrients, flavor compounds, and toxic substances found in vegetables which deserve attention.

Phytoceuticals

Phytoceuticals can be grouped on the basis of their therapeutical activities or their physical and chemical constitution. Some of these groups are bioflavonoids, phenols, phytosterols, thiols, indoles, glucosinolates, and organosulfur compounds.

Bioflavonoids are one of the largest groups of secondary plant metabolites found in several vegetables. They have been reported to increase the effectiveness of vitamin C and are recognized as potent antioxidants. Together with vitamin C they strengthen the capillaries and protect the body from free radicals. They are also antiviral and antiinflammatory. Some of the important bioflavonoids identified from vegetables are quercetin, kaempferol, myricetin, and luteolin. There are reports which suggest that quercetin (found in onion) can inhibit the growth of several types of cancer cells, including breast cancer, ovarian cancer, and leukemia.

The higher content of phenols has been linked to disease resistance in plants. Phenols have also been found to protect our body from oxidative damage caused by free radicals. Some tropical root vegetables like yam, elephant-foot yam, and other vegetables like green banana and eggplant are rich in phenols.

Phytosterols occur in most plants, especially in the seeds of green and yellow vegetables. Phytosterols prevent the absorption of cholesterol and facilitate its excretion from the body. There are some reports which suggest that phytosterols prevent the development of tumors in breast, colon, and prostate glands.

Thiols are important sulfur compounds found in brassicaceous vegetables like broccoli, cabbage, cauliflower, and kale, and in garlic. There is strong evidence to suggest that these compounds can prevent cancers of colon, rectum, and breast. Indoles found in Brassicas bind chemical carcinogens and activate detoxification enzymes, mostly in the gastrointestinal tract.

Glucosinolates, which are secondary plant metabolites, found mostly in cole crops and other Brassicas, are powerful activators of enzymes in the liver which detoxify carcinogens.

There are several organosulfur compounds found in onion and garlic which appear to possess antimicrobial properties and offer protection against cardiovascular diseases.

Key Micronutrients

Vegetables are the major source of vitamins and minerals essential for our body. They are the most important source of vitamin A; deficiency of vitamin A in the diet causes blindness. Vitamin A is essential for growth, bone development, vision, reproduction, and functioning of the immune system. Vitamin A is synthesized in our body from carotene which is present in abundance in green leafy vegetables, carrot, pumpkin, squash, melon, and tomato. Of about 600 carotenoids identified from natural sources, less than

Table 2 Micronutrient composition of temperate vegetables per 100 g edible portion

Crop	Vitamins						Minerals		
	A (IU)	Thiamin (mg)	Riboflavin (mg)	Niacin (mg)	Folic acid (μg)	Ascorbic acid (mg)	Ca (mg)	P (mg)	Fe (mg)
Artichoke	62	0.23	0.01	–	–	0	120	100	2.3
Asparagus	900	0.18	0.20	1.5	109	33	22	62	1.0
Broad bean	15	0.08	–	0.8	–	12	50	64	1.4
Brussels sprout	550	0.10	0.16	0.9	49	102	36	80	1.5
Cabbage	130	0.05	0.05	0.3	32	47	49	29	0.4
Cardoon	120	0.02	0.03	0.3	–	2	70	23	0.7
Carrot	11000	0.06	0.05	0.6	15	8	37	36	0.7
Cauliflower	60	0.11	0.10	0.7	22	78	25	56	1.1
Celery	240	0.03	0.03	0.3	7	9	39	28	0.3
Chicory leaf	4000	0.06	0.10	0.5	–	24	100	47	0.9
Endive	2500	0.07	0.08	0.4	47	8	42	30	2.0
Garden beet	0	0.04	0.09	0.4	–	10	18	55	1.0
Garlic	Trace	0.06	0.23	0.4	–	13	30	310	1.3
Kale	10000	0.16	0.26	2.1	70	186	249	93	2.9
Knol khol	20	0.06	0.04	0.3	–	66	41	51	0.5
Knol khol green	6924	0.25	–	0.3	–	157	740	50	13.3
Leek	30	0.23	–	–	–	11	50	70	2.3
Lettuce	900	0.06	0.06	0.3	25	8	35	26	2.0
Onion	Trace	0.08	0.01	0.4	11	11	47	50	0.7
Onion stalk	993	0.05	0.03	0.3	14	17	50	50	7.4
Parsley	5200	0.08	0.11	0.7	38	90	125	40	2.0
Parsnip	0	0.09	0.05	0.7	–	17	40	69	0.7
Pea	640	0.35	0.14	2.9	25	27	26	116	1.9
Potato	40	0.10	0.01	1.2	7	17	10	40	0.7
Radish	5	0.06	0.02	0.5	7	15	35	22	0.4
Rutabaga	580	0.07	0.07	1.1	5	33	66	39	0.4
Spinach	8100	0.10	0.20	0.6	123	51	93	51	3.1
Spinach beet	5862	0.26	0.56	3.3	60	70	380	30	16.2
Sprouting broccoli	2500	0.10	0.23	0.9	54	113	103	78	1.1
Swiss chard	3300	0.04	0.09	0.4	42	30	51	46	1.8
Tumip green	15691	0.31	0.57	5.4	42	180	710	60	28.4
Turnip root	0	0.04	0.04	0.5	4	43	30	40	0.4

Source: Gopalan C, Sastri BVR, Balasubramanian SC, Rao BSN, Deosthale YG and Pant KC (1999) *Nutritive Value of Indian Foods*. Hyderabad: National Institute of Nutrition; Watt BK and Merril AL (1963) *Composition of Foods – Raw, Processed, Prepared*. Agricultural Handbook No. 8. Washington, DC: USDA.

Table 3 Important trace elements (mg per 100 g) of some temperate vegetables

Crop	Cu	Mn	Zn
Beetroot	0.29	0.19	0.91
Brussels sprouts	0.20	0.18	0.30
Cabbage	0.02	0.18	0.30
Carrot	0.10	0.16	0.36
Cauliflower	0.13	0.10	0.40
Knol khol	0.05	0.11	0.24
Onion	0.18	0.28	0.46
Onion stalk	0.45	0.74	2.29
Pea	0.23		
Potato	0.16	0.13	0.55
Radish	0.40		
Spinach	0.10	0.56	0.30

Source: Gopalan C, Sastri BVR, Balasubramanian SC, Rao BSN, Deosthale YG and Pant KC (1999) *Nutritive Value of Indian Foods*. Hyderabad: National Institute of Nutrition.

10% are precursors of vitamin A. In human beings, β-carotene, α-carotene, and cryptoxanthin are converted to vitamin A, whereas lycopene (present in tomato) and lutein are not. Most important is β-carotene. Some of the carotenoids are powerful antioxidants and have been shown to offer protection against breast, lung, uterine, prostate, and colon cancers.

Vitamin C (ascorbic acid) is essential for wound healing, immune response, and allergic reaction. It is a powerful antioxidant and is found in many vegetables like broccoli, cauliflower, Brussels sprout, tomato, pepper, and potato. It protects our body by scavenging free radicals.

Vitamin E is also an antioxidant which prevents peroxidation of polyunsaturated fatty acids in our body. Increased intake of vitamin E is believed to reduce the incidence of heart attack. The combined action of vitamin E, C, and A has been reported to

delay the onset of cataract. Several vegetables, like broccoli, spinach, parsnip, and watercress, are good sources of vitamin E.

Folic acid, commonly referred to as folate, is an essential vitamin which is especially crucial in fetal development in the early weeks of pregnancy. It is essential for DNA synthesis and its action is interrelated with that of vitamin B_{12}. The Food and Agriculture Organization/World Health Organization (FAO/WHO) expert group has recommended a daily dietary intake of 200 μg of free folate for adults, 400 μg during pregnancy, and 300 μg during lactation. Recent studies have associated folate with a reduced risk of cardiovascular diseases. Deep-green leafy vegetables like spinach, kale, Swiss chard, and parsley, and some beans are rich in folate. Light-green leaves are generally not rich in this vitamin.

Flavor compounds

The flavor compounds contribute to the overall sensory quality of the vegetables. They are a complex mixture of a large number of volatile compounds. Vegetables contain some volatile flavor compounds in the intact tissues and some are produced enzymatically after the tissue is damaged or crushed. These compounds undergo further changes after heating. Flavor compounds of vegetables include (1) nonvolatiles like sugars, amino acids, fatty acids, and organic acids and (2) volatiles like aromatic hydrocarbons, aldehydes, acetals, ketones, alcohols, esters, and sulfur compounds. When used as side-dishes with cereal staple foods, these flavor compounds make the cereal more palatable.

Dietary Fiber

Leafy, root, and a few other vegetables provide dietary fiber, essential for bowel movement and possibly for the prevention of several diseases such as appendicitis, colon cancer, diabetes, diverticulosis, gallstones, and obesity. Dietary fiber consists mainly of cellulose, hemicellulose, pectin, lignin, and mucilage of plants.

Toxic Substances

Some vegetables contain harmful chemical compounds such as trypsin inhibitors in bean seeds, phytate in beans, and oxalate in amaranth and rhubarb, which pose serious nutritional problems. In addition, chemicals such as glucosides induce toxicity by producing hydrogen cyanide (HCN). Other toxicants are nitrates in leafy vegetables, some alkaloids in cucurbits, potato, and asparagus, and glucosinolates in Brassicas. These, when taken in higher amounts, may cause harm.

A balanced diet should contain the amount of essential nutrients needed to supply energy and prevent nutritional deficiencies. Vegetables contain combinations of nutrients and other protective substances. Dietitians recommend an adult daily uptake of 75–125 g of green leafy vegetables, 60–80 g of other vegetables, and 75–100 g of roots and tubers for a balanced diet. (*See* **Dietary Fiber**: Physiological Effects; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Trypsin Inhibitors**.)

With the growing consciousness of health and with the recognition that a diet primarily of animal origin is hazardous to health, the consumption of vegetables has increased considerably, and vegetables have become an important item of commerce. Vegetable growers are able to obtain a better economic return, and this is a great incentive for cultivation of this commodity. Vegetables are items of daily consumption and large quantities are marketed every day; the market in vegetables can, therefore, never stagnate. Increased demand for fresh temperate vegetables in the more affluent countries during winter months has created an opportunity for the tropical and subtropical countries to improve their economy by increasing exports.

Many agro-based industries use vegetables as their raw material. Tomatoes and potatoes are two vegetables which are processed into a wide range of products. Frozen vegetables are also of great commercial importance as they retain most of the qualities of fresh vegetables. In addition, vegetables are utilized by many food-processing units for the production of various vegetable sauces, fermented, and pickled vegetables. (*See* **Freezing**: Structural and Flavor (Flavour) Changes; **Tomatoes**.)

The importance of individual vegetables in commerce, and in different diets with special reference to nutrition, is discussed under the Importance of Specific Temperate Vegetables below.

International Trade

Among all the temperate vegetables discussed in this article, potato occupies the topmost position in world trade. According to an FAO estimate, the total world export of potato in 1997 was 6.87 million tonnes and the European countries ranked first, exporting 5.29 million tonnes. The next important crop in international trade is onion. Out of a total world export of 3.23 million tonnes during 1997, the highest quantity was exported by the European countries with a share of 34.84%, followed by Asian countries with 27.83% share. Besides world trade in the fresh form of potato, both these commodities are exported in various processed products prepared from them.

Potato chips and dehydrated onion occupy a prominent position among them. In tropical countries of Asia, mostly short day varieties of onions are grown. These varieties are more pungent and have higher total soluble solids than the long day types grown in temperate countries. Because of high total soluble solid and pungency, they store well under high temperature conditions with less storage rot and can stand long-distance shipping without refrigeration. Because of these qualities, the demand for these types is very high in world trade, particularly in tropical countries.

Cole crops and root vegetables are much in demand in temperate countries during winter. In tropical and subtropical countries these are grown in the open during winter and thus the cost of production is less compared with those grown under protective cover. Many of these vegetables are exported from Latin American and Asian countries to meet the demand of the European and North American markets. Recently, the demand for minimally processed vegetables has increased considerably in the developed countries, mainly because of convenience. There is a great potential for exporting minimally processed temperate vegetables grown in tropical and subtropical countries during the winter season.

Importance of Specific Temperate Vegetables

Bulb Crops

Bulb crops, such as onion (*Allium cepa*), garlic (*A. sativum*), and leek (*A. porrum*), are cultivated in both the temperate and tropical regions. Leek is particularly adapted to cold weather. (*See* **Onions and Related Crops**.)

Onions are high in energy and rich in calcium and riboflavin (vitamin B_2). The bulbs do not contain starch and most of the carbohydrates are sucrose, glucose, fructose, and fructosan (polymer of fructose). Green onion tops have a high carotene (provitamin A) content. Onions are an indispensable item in every kitchen as a condiment and as a vegetable. Green leaves and immature and mature bulbs are eaten raw in salads or used as flavoring agents in vegetable dishes. Multiplier onion bulbs (shallot) are pickled in vinegar or brine. Dehydrated bulbs and onion powder are in great demand in world trade. The distinctive odor (pungency) of onion is due to a volatile sulfur compound, allyl propyl disulfide. Compounds responsible for alliaceous odors (and flavors, at least in part) are only formed after the onion has been cut or injured. Treatments such as heating or freeze-drying which destroy or suppress enzyme activity, reduce the formation of odor. (*See* **Freeze-drying**: Structural and Flavor (Flavour) Changes; **Sensory Evaluation**: Aroma.)

Garlic is consumed more in Asia, South America, North Africa, and southern Europe. Garlic cloves (bulbs) are rich in protein, carbohydrate, and some minerals. They are used primarily as a flavoring agent in a variety of dishes and vegetable soups, meat, sausage, spaghetti, and vegetable curries. They are used for the preparation of garlic bread and garlic butter. The pungency of garlic is due to a volatile sulfur compound, diallyl disulfide, which is produced by the action of allinase enzyme on the amino acid, alliin, present in the garlic.

The leek is an important ingredient in European dishes and its use in North America is increasing. Leek stalks are used as a base for soups, and they are steamed, boiled, or eaten raw in salads. After boiling or crushing, most of the pungent flavors disappear and this treatment gives a mild flavor to the product.

Onion and garlic have a traditional place in folk medicine in many countries. It has been shown that half a medium-sized raw onion a day can trigger shifts in the ratio of 'good' to 'bad' cholesterol, replacing a significant amount of destructive low-density lipoprotein (LDL) cholesterol with heart-protective high-density lipoprotein (HDL). The HDL boost is greatest from raw white and yellow onions and is reduced with cooking. Onion and garlic are rich sources of quercetin, a bioflavonoid with potent anticarcinogenic activity. They are also known to reduce blood sugar. An antihypoglycemic agent, diphenylamine, has been identified in onion recently. The allicin which is produced in garlic when the tissues are injured by cutting or crushing has been reported to have hypocholesterolemic action. It reduces blood stickiness, thus preventing coronary thrombosis, heart attacks, and strokes. It has been indicated that garlic constitutents like allicin, allistalin, garlicin, diallyl disulfide, diallyl trisulfide, and essential oils are active against certain bacteria and other microorganisms. (*See* **Lipoproteins**.)

Cole crops

Cole crops are one of the largest groups of temperate vegetables, comprising important crops like cauliflower (*Brassica oleracea* var. *botrytis*), cabbage (*B. oleracea* var. *capitata*), sprouting broccoli or broccoli (*B. oleracea* var. *italica*), Brussels sprout (*B. oleracea* var. *gemmifera*), kale (*B. oleracea* var. *acephala*), and knol khol (*B. oleracea* var. *gongylodes*). Cole crops are rich in vitamin C and they are good sources of minerals. Broccoli and kale are rich in

carotene (provitamin A). Several volatile sulfur compounds are responsible for the characteristic flavor of the cole crops. Dimethyl trisulfide has been indicated as the major aroma component in cooked Brassicaceous vegetables. These vegetables contain compounds known as indoles and dithiolthiones which have been linked with the prevention of cancers of the colon, rectum, and breast. (*See* **Cancer**: Diet in Cancer Prevention; **Colon**: Cancer of the Colon.)

Cabbage and broccoli are used raw in salads. Cabbage is used steamed or boiled alone, or in a mixture of other vegetables as main dishes with meat or in casseroles. Sauerkraut is the main processed product from cabbage. Some cabbage is dried after blanching and also pickled. Cauliflower is cooked or steamed. It is used in a mixture with other vegetables for the preparation of curry and fried rice in the Orient. It is also pickled along with other vegetables. Broccoli, Brussels sprout, and cauliflower are frozen. Broccoli is steamed and used in the preparation of many casserole dishes. It may be chopped or prepared whole or as florets. Its versatility of use has made it a staple item in the American diet. Knol khol is a popular vegetable of Germany. It is used raw or uncooked, and the tender leaves are used as greens. Cooked knol khol is scooped out, stuffed with onion and sausage, and then baked.

Leafy vegetables

Spinach (*Spinacea oleracea*), Swiss chard (*Beta vulgaris* var. *cicla*), spinach beet (*B. vulgaris* var. *bengalensis*), and kale (*Brassica oleracea* var. *acephala*) are the major leaf vegetables. The first three crops belong to the family Chenopodiaceae and the last one to Cruciferae (Brassicaceae). Of these, spinach beet is primarily grown in the tropics, usually during winter. Spinach, Swiss chard, and spinach beet leaves are rich sources of carotene and some minerals. Kale is one of the most nutritious vegetables. It is rich in carotene and vitamin C. Besides, there are some vegetables like knol khol, lettuce, spinach, and turnip, the leaves of which contain a good amount of protein. The essential amino acid composition of these leaves is given in Table 4.

Spinach and Swiss chard leaves and tender petioles are cooked, steamed, or boiled for the preparation of dishes like pastries, soufflés, and soups. The stems of chard are cooked and used like celery. Sometimes spinach and chard are cooked combined to give a blend of their pleasant flavors. Some quantity of spinach is frozen and canned. Spinach beet is a popular leaf vegetable of the Indian subcontinent. Leaves and petioles are cooked for preparing stews. Kale is steamed and served with butter or vinegar, or in soups.

Pea and Broad Bean

Pea (*Pisum sativum*) and broad bean (*Vicia faba*) are leguminous vegetables which are characteristically rich in protein. Peas are high in thiamin (vitamin B_1), niacin, and phosphorus. Broad beans are rich in vitamin C. The pea is one of the major vegetables of the temperate and subtropical world. It is cooked with other vegetables or used in soups. It is also used in salads. In India, it is cooked with milk casein to prepare a delicious dish. A large portion of the pea crop is canned, frozen, or dehydrated. The broad bean is cooked with other vegetables or processed like peas. (*See* **Legumes**: Legumes in the Diet; **Peas and Lentils**.)

The essential amino acid composition of pea and broad bean has been given in Table 5. Both of them are deficient in the amino acid tryptophan. Studies have shown that food legumes reduce blood cholesterol concentration and increase bowel movement. Phytic acid (inositol hexaphosphoric acid), which is present in mature seeds of peas and beans, can

Table 4 Essential amino acids content (g) of vegetable leaves per 100 g protein

Amino acids	*Knol khol*	*Lettuce*	*Spinach*	*Turnip*
Arginine		4.8	5.6	4.8
Histidine		1.8	2.2	1.9
Isoleucine	5.3	5.1	4.8	3.8
Leucine	8.0	6.2	8.5	6.6
Lycine	8.6	5.0	6.4	5.1
Methionine	0.8	1.1	1.8	1.3
Phenylalanine	2.6	4.5	5.3	4.2
Threonine	2.1	4.3	4.6	3.8
Tryptophan	5.9	0.8	1.6	1.3
Valine	5.0	5.1	5.6	4.5

Source: Gopalan C, Sastri BVR, Balasubramanian SC, Rao BSN, Deosthale YG and Pant KC (1999) *Nutritive value of Indian Foods*. Hyderabad: National Institute of Nutrition.

Table 5 Essential amino acid content (g) of pea and broad bean seed per 100 g protein

Amino acids	*Broad bean (*Vicia faba*)*	*Pea (*Pisum sativum*)*
Arginine	8.8	9.5
Histidine	2.4	2.3
Isoleucine	4.0	4.2
Leucine	7.0	6.8
Lysine	6.4	7.5
Methionine	0.7	0.9
Phenylalanine	4.3	4.5
Threonine	3.6	4.0
Tryptophan	–	–
Valine	4.3	4.7

Source: Duke JA (1981) *Handbook of Legumes of World Economic Importance*. New York: Plenum Press; Summerfield RJ and Roberts EH (eds) (1985) *Grain Legume Crops*. London: Collins.

combine with calcium and iron from other foods in the diet and render them unavailable. (*See* **Phytic Acid**: Nutritional Impact.)

Perennial Vegetables

Major perennial vegetables are asparagus (*Asparagus officinalis*), artichoke (*Cynara scolymus*), and cardoon (*C. cardunculus*). Asparagus greens are rich in carotene, thiamin, riboflavin, and niacin. Various plant parts of asparagus, such as new shoots or spears, petioles, flower buds, tubers, and roots, are used for the preparation of different dishes. Asparagus is prized for its spears. Fresh asparagus is usually steamed or boiled. It is also frozen and canned. It is especially rich in vitamin E and 60 g of it can provide our daily requirement. It contains an amino acid asparagine which has calming properties. It is diuretic and helps in the elimination of water through urination. It is also helpful in cleaning cholesterol from arteries. However, in excessive amounts, it can irritate the kidneys.

In artichoke and cardoon, the edible portion consists of fleshy bracts, receptacle, and petals. Fleshy petioles of the cardoon are sometimes blanched to make them more succulent. The fresh bud is usually cooked whole and served hot or cold with melted butter, salad dressing, or sauce. The canned or frozen small buds or 'hearts' are popular as an hors d'oeuvre. Artichoke stimulates bile evacuation and acts as a diuretic.

Potato

The potato (*Solanum tuberosum*) ranks first among all the vegetables in terms of area and production in many countries. It occupies a prominent position in the diet as a staple food, as well as a vegetable for the preparation of side-dishes. Potatoes are boiled, baked, mashed, roasted, and fried (French fry). The potato goes well in combination with most vegetables. It is also one of the major vegetables which is processed as chips, dehydrated, and granulated. (*See* **Potatoes and Related Crops**: Processing Potato Tubers.)

Potatoes are rich in starch, minerals, and vitamin C. They are easily digested and assimilated, and thus are considered as a good food for the sick and invalids. Though the tubers contain only about 2% protein on a fresh-weight basis, the content on a dry-weight basis comes to about 10%, which is comparable to wheat flour. This makes the potato an important source of dietary protein in countries where consumption is high.

Potato tubers when exposed to light produce an alkaloid, solanin, which is bitter in taste and has toxic effects when consumed in large amounts. (*See* **Alkaloids**: Toxicology.)

Root crops

Though carrot (*Daucus carota*), radish (*Raphanus sativus*), turnip (*Brassica rapa*), garden beet (*Beta vulgaris*), parsnip (*Pastinaca sativa*), and rutabaga (*Brassica napus* var. *napobrassica*) are major root crops of the temperate region, there are distinct tropical types of the first three crops which are grown extensively in tropical Asia. Of all the vegetables, carrots, particularly yellow and orange flesh cultivars, are the richest source of carotene (provitamin A) which helps prevent night blindness and cancer. The cortical region contains more carotene than the core. Dark red cultivars, mostly of tropical types, contain more anthocyanin pigment and contribute less provitamin A to the diet. Turnip and rutabaga are rich in vitamin C. Turnip and radish greens are rich in provitamin A. Radish is astringent and diuretic and is used to promote bile flow. Garden beet is not nutritionally rich. However, it has a number of useful medicinal properties such as purification of blood, improving circulation, and bowel stimulation. Its red color is due to the pigment betacyanin, the chemical properties of which resemble anthocyanin. It also contains a yellow pigment, betaxanthin. Parsnip is rich in starch. (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments.)

Carrots are primarily consumed in salads and are used in soups, curries, stews, and pies. They are used in the preparation of a popular sweetmeat in the Indian subcontinent. They are frozen or pickled. It has been indicated that there is an inverse relationship between a diet which is rich in carotene and cancer of the lung. Turnip roots are steamed or cooked and the tender leaves, apart from use in salads, are also cooked. In western countries, the radish is primarily used as a salad crop. In the Orient, it is cooked and also pickled. The rutabaga is used almost in the same way as turnip. It is used in salads, soups, and in baked products on account of its pleasant flavor. Parsnip, belonging to the carrot family, gives a pleasant flavor when steamed or cooked, and it is used in the preparation of various dishes. In beet, both roots and tops are steamed or boiled, and a considerable quantity of beetroot is canned, frozen or pickled. Beet juice is used for the coloring of many food products.

Salad crops

Salad crops comprise lettuce (*Lactuca sativa*), celery (*Apium graveolens*), chicory (*Cichorium intybus*), endive (*C. endivia*), and parsley (*Carum petroselinum*). The lettuce is the major salad vegetable in most of the countries of North America, South

America, Europe, Australia, and New Zealand. Its production is increasing in the Middle East, Africa, and Japan. It is rich in provitamin A and minerals. Leaf lettuce is more nutritious than the head types. Celtuce (*L. sativa* var. *augustana*) is very popular in China and Taiwan. Its succulent thick stems and tender leaves are used raw or cooked. Celery is primarily used as a salad crop, and is also used for the preparation of soups, stews, and sauces. Its seeds are used as a spice. It contains a compound, 3-*n*-butyl phthalide, which lowers blood pressure. Chicory (witloof, chicon) and endive are used in salads in Europe. The blanched cluster of leaves (head) has a delicate flavor as compared to the strong flavor of the green, unblanched head. It is considered as a delicacy in France, The Netherlands, and Belgium. The roots of chicory are dried, roasted, and blended with coffee to give it a strong flavor. Parsley leaves are mainly used for garnishing and flavoring. (*See* **Salad Crops**: Dietary Importance; Leaf-types; Root, Bulb, and Tuber Crops.)

See also: **Cancer**: Diet in Cancer Prevention; **Carotenoids**: Physiology; **Colon**: Cancer of the Colon; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Dietary Fiber**: Physiological Effects; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Freeze-drying**: Structural and Flavor (Flavour) Changes; **Freezing**: Structural and Flavor (Flavour) Changes; **Legumes**: Legumes in the Diet; **Onions and Related Crops**; **Peas and Lentils**; **Potatoes and Related Crops**: Processing Potato Tubers; **Salad Crops**: Dietary Importance; **Tomatoes**

Further Reading

Bose TK, Som MG and Kabir J (eds) (1993) *Vegetable Crops*. Calcutta: Naya Prokash.

Choudhury B (1990) *Vegetables*, 8th edn. New Delhi: National Book Trust.

Decoteau DR (2000) *Vegetable Crops*. New Jersey: Prentice Hall.

Duke JA (1981) *Handbook of Legumes of World Economic Importance*. New York: Plenum Press.

Gopalan C, Sastri BVR, Balasubramanian SC, Rao BSN, Deosthale YG and Pant KC (1999) *Nutritive Value of Indian Foods*. Hyderabad: National Institute of Nutrition.

Jeana G (1991) *Pigments in Vegetables*. New York: Van Nostrand Reinhold.

Nonnecke IL (1989) *Vegetable Production*. New York: Van Nostrand Reinhold.

Sharon M (1999) *Nutrients A to Z*. London: Prion Books.

Summerfield RJ and Roberts EH (eds) (1985) *Grain Legume Crops*. London: Collins.

Vaclavik VA (1998) *Essentials of Food Science*. New York: Chapman and Hall.

Watt BK and Merril AL (1963) *Composition of Foods – Raw, Processed, Prepared*. Agricultural Handbook No. 8. Washington, DC: USDA.

Yamaguchi M (1983) *World Vegetables*. New York: Van Nostrand Reinhold.

Cabbage and Related Vegetables

R Fordham, Ludgvan, Cornwall, UK
P Hadley, University of Reading, Whiteknights, Reading, UK

Introduction

Cabbages and closely related vegetables belonging to the same species are collectively referred to as cole crops, which include a number of the world's most commonly cultivated vegetables. They are classified into a number of readily identifiable groups based on morphological differences in their vegetative and reproductive stages. This article deals with those members primarily distinguished by vegetative differences. Members of the species differing in reproductive characters are discussed in a separate article. (*See* **Brassicas**.) Chinese cabbage is a different species and is dealt with separately in the article on Oriental vegetables.

Cabbage: Brassica oleraceae Capitata Group, Family Cruciferae

The cabbage is a biennial plant grown as an annual for its edible leaves, which, at maturity, form a compact head. The species is thought to have originated in western Europe along the coast of the Mediterranean where wild forms still occur. It subsequently spread along the Atlantic coast as far north as Scotland. Earliest selections were probably made to reduce the content of bitter glucosinolates, which are at high levels in the wild types. The crop has a long history of cultivation; the Ancient Greeks grew it primarily for medicinal purposes and its cultivation in Germany took place as early as the twelfth century. Over time a wide range of cultivars has been selected to suit different environments, seasons, and markets. Cabbage production is now worldwide and, although better adapted to cooler climates, it is also grown in tropical regions. Data on world production are presented in **Table 1**.

Table 1 Estimated world production of cabbage

Country	*Production (Mt)*
World total	48 872 193
China	18 502 086
India	4 200 000
Russian Federation	2 850 000
Korea, Republic of	2 755 000
Japan	2 400 000
USA	2 150 000
Poland	1 709 154
Indonesia	1 459 232
Ukraine	1 015 000
Uzbekistan	894 000
Romania	812 000
Germany	791 364
Turkey	732 000
Korea, Democratic People's Republic	600 000
Italy	522 000

Source: FAO (1999) Food and Agriculture Organisation. Website: http://apps.fao.org.

Although there are marked differences between cultivars in terms of their environmental responses, the optimum temperature for growth in cabbage is between 16 and 18 °C. Growth is arrested above 25 °C and below 0 °C, although cold-hardened plants will survive temperatures as low as −10 °C without severe damage.

The dense terminal cabbage head is produced on a short stem. Harvesting is usually carried out at around the time of maximum head density, although immature plants may also be harvested as 'spring greens' in order to extend the season of supply. If harvesting is delayed beyond maturity, further expansion of the inner leaves and continuation of stem growth will result in splitting of the head. Flowering in cabbage, although a necessary prerequisite of seed production, will detract from crop quality. Young plants pass through a juvenile period during which they are unable to flower but following this, exposure to vernalizing temperatures of between 4 and 10 °C may induce premature flowering, referred to as bolting. The stem lengthens to bear a terminal raceme of yellow to white flowers. Following pollination seed development occurs within a fruit called a siliqua. Normal flowering, however, usually occurs in the second year following exposure of the mature plant to vernalizing temperatures during its first winter in the field. The duration of exposure required to initiate flowering depends on the precise pattern of temperatures and on the sensitivity of individual cultivars, but generally varies between 10 and 50 days or longer.

Cabbage cultivars also differ in head shape, varying from pointed to round; leaves may be green or red and smooth or wrinkled. Savoy types, which are particularly tolerant to cold conditions, have deeply wrinkled dark green leaves. For production purposes, cabbages are classified on the basis of their season of maturity. The earliest types for spring harvesting may be cut before they develop hearted heads and are then marketed as spring greens. These are followed by a wide range which mature throughout the summer, fall, and winter. Winter white cultivars, often referred to as Dutch White, have dense-hearted heads of blanched leaves which store well.

Nutritional Value and Chemical Composition

The nutritional value of cabbages depends on their vitamin and mineral contents, which vary significantly with cultivar and method of preparation. Carotene levels, for example, depend on the amount of chlorophyl, and outer green leaves may contain 50 times as much as inner white ones. As a consequence, carotene content in Savoy and kale, for example, are significantly higher than in Dutch White cabbage. Heat treatment will also cause changes in composition and vitamin C levels are markedly reduced by boiling. Estimates of nutritional value and chemical composition of cabbage and related vegetables are summarized in Table 2.

Handling and Storage

Cabbages are a relatively low-value vegetable and are normally grown on a large field scale, either from direct drillings or transplanting. A range of cultivars is available to provide spread in maturity dates but the season can also be extended by growing transplants under protection in glasshouses. Earliness can also be achieved by covering field crops with plastic film covers or nonwoven agritextiles. Under favorable conditions, appropriate cultivars can be harvested throughout the year, using Savoy types and unhearted spring cabbages or greens to cover the winter months.

Continuity of supply can also be maintained by both cold and controlled-atmosphere storage, although the latter is more commonly used for higher-value vegetables. Winter white cabbage, for example, can be stored for at least 6 months in cold stores at an optimum temperature of 0–1 °C, and 95% relative humidity. Storage for longer periods can be achieved under controlled-atmosphere conditions of lower oxygen and higher carbon dioxide relative to ambient conditions, although different vegetables and cultivars vary markedly in their response to specific levels and off-flavors may develop. The optimum concentrations for cabbage lie between 2.5 and 3% oxygen and 2–5% for carbon dioxide which, combined with cold storage, enables white cabbage to be stored for around 9 months.

Table 2 Nutritional value and chemical composition of white and Savoy cabbage, curly kale, and Brussels sprouts. Values are based on 100 g raw flesh

	White cabbage	*Savoy cabbage*	*Curly kale*	*Brussels sprouts*
Edible proportion	0.91	0.77	0.85	0.69
Water (g)	90.7	88.1	88.4	84.3
Total nitrogen (g)	0.23	0.33	0.55	0.56
Protein (g)	1.4	2.1	3.4	3.5
Fat (g)	0.2	0.5	1.6	1.4
Carbohydrate (g)	5.0	3.9	1.4	4.1
Energy value (kJ)	113	114	140	177
Starch (g)	0.1	0.1	0.1	0.8
Total sugars (g)	4.9	3.8	1.3	3.1
Dietary Fiber (g) (Englyst method)	2.1	3.1	3.1	4.1
Sodium (mg)	7	5	43	6
Potassium (mg)	240	320	450	450
Calcium (mg)	49	53	130	26
Magnesium (mg)	6	7	34	8
Phosphorus (mg)	29	44	61	77
Iron (μg)	0.5	1.1	1.7	0.7
Copper (mg)	0.01	0.03	0.03	0.02
Zinc (mg)	0.2	0.3	0.4	0.5
Sulfur (mg)	54	88	N	93
Chloride (mg)	40	48	68	38
Manganese (mg)	0.2	0.2	0.8	0.2
Selenium (μg)	Tr	(2)	(2)	N
Iodine (μg)	2	2	N	1
Carotene (μg)	40	995[c]	3145	215
Vitamin D (μg)	0	0	0	0
Vitamin E (μg)	0.2	0.20[d]	(1.70)	1.00
Thiamin (mg)	0.12	0.15	0.08	0.15
Riboflavin (μg)	0.01	0.03	0.09	0.11
Niacin (μg)	0.3	0.7	1.0	0.2
Vitamin B_6 (μg)	0.18	0.19	0.76	0.37
Vitamin B_{12} (μg)	0	0	0	0
Folate (μg)	34	150	120	135
Pantothenate (mg)	0.21	0.21	0.09	1.00
Biotin (μg)	0.1	0.1	0.5	0.4
Vitamin C (mg)	35	62	110	115

[a]N, no analysis available; Tr, trace.
[b]Values in parentheses are estimated from other related data.
[c]Average figures. The amount of carotene in leafy vegetables depends on the amount of chlorophyl, and the outer green leaves may contain 50 times as much as inner white leaves.
[d]Value for inner leaves.
Reproduced from Holland B, Unwin ID and Buss DH (1991) *Vegetables, Herbs and Spices: Fifth Supplement to McCance and Widdowson's. The Composition of Foods*. London: HMSO, with permission.

Uniformity and head size are important marketing requirements and are achieved through the appropriate choice of cultivar and carefully controlled plant spacing. Spring greens, for example, are grown at densities of up to 40 plants per square meter, in contrast to 9 per square meter for hearted summer harvests and 3–4 for winter and fall crops. Earlier harvests may be by selective hand-cutting, although the majority of commercial crops are now harvested mechanically.

Domestic and Industrial Use

Cabbages may be cooked by boiling or stir-fried, or used raw in salads and in the preparation of coleslaw. Red types are generally sliced and pickled in vinegar. Sauerkraut is manufactured industrially from sliced white cabbage by controlled anaerobic fermentation and the addition of salt. A small proportion of the crop may be dried.

Kale: *Brassica oleraceae* Acephela Group, Family Cruciferae

Kale is a biennial closely related to the cabbage, grown as an annual for its edible shoots and young leaves. It probably originated along with other varieties of the species in western Europe, but is now widely distributed throughout the world. It is one of

the hardiest of all brassicas, capable of withstanding winter temperatures as low as −15 °C. It is also tolerant of high summer temperatures. There are two main groups of kale, the curly-leaved types referred to as Scotch kales or borecole, and the broader, smooth-leaved types. Rape or Siberian kales are similar to the curly-leaved kales but belong to the species *Brassica napus*. Unlike cabbages, kales produce no head, but leaves are borne on a tall stem, the height of which is dependent on cultivar. The crop is mainly used for winter supplies when other green leafy vegetables may be in short supply.

Cultivation of the crop is similar to that for cabbage. Shoots and leaves of kale are cooked by boiling in a similar manner to cabbage. They provide a valuable source of minerals and vitamins. These are summarized for raw curly kale in Table 2.

Brussels Sprout: *Brassica oleracea* Gemmifera Group, Family Cruciferae

The Brussels sprout is a single-stemmed, tall biennial growing to a height of around 1 meter in its first season. The main edible portions of the crop are the small, tightly formed axillary buds that develop in the axils of the expanding leaves. Its loose head of large leaves may also be used as a substitute for cabbage. Flower production in Brussels sprouts is similar to that for cabbage, being triggered by exposure of mature plants to relatively low temperatures. Cultivars vary in their sensitivity to vernalizing temperatures, but Brussels sprouts generally require a longer period of exposure than cabbage, ranging from around 50 to 80 days. The crop has developed from the wild cabbage and is believed to have originated in Belgium where it was cultivated during the nineteenth century. Its cultivation has spread to other European countries: production is carried out on a relatively large scale in the UK. Its importance is more limited outside Europe. In the USA it is mainly cultivated in California.

Brussels sprouts are adapted to relatively cool conditions, and in warmer regions the axillary buds may develop into loose sprouts, in contrast to the preferred tightly formed round buds. Cultivars have been selected to provide a spread in harvests and later-maturing ones are able to withstand temperatures down to −10 °C. Cultivar selection for sprout uniformity has been achieved for use in mechanical harvesting.

Food Value and Chemical Composition

In comparison with other *Brassica* crops, only a relatively small portion of the total plant is consumed. The composition of the edible sprouts is indicated in Table 2. Mineral and vitamin content determines their main value.

Handling and Storage

The cultivation of Brussels sprouts is similar to that of cabbage and other *Brassica* species. Crops may be grown either from direct seeding or from transplanted seedlings raised in open seed beds or, more generally, under protection. Typically, seeds are sown in early spring and harvesting is carried out from fall until the following spring. Continuity of production is achieved primarily through choice of cultivars, but raising transplants under protection enables an advancement of the earlier crops. The sprouts on older cultivars develop from the base of the stem upwards, necessitating selective harvesting by hand. Modern hybrid cultivars, however, mature more uniformly and commercial production is now mainly conducted using mechanical harvesters.

Sprouts are trimmed and marketed loose. Short-term storage in cold stores at 0–1 °C and high humidity can be used to maintain freshness.

Domestic and Industrial Use

Brussels sprouts are cooked by boiling. Processing is primarily by freezing, although a small proportion of the crop is canned.

See also: **Brassicas**; **Vegetables of Temperate Climates**: Oriental Brassicas

Further Reading

Hardy FS and Watson GD (1982) *The Complete Guide to Commercial Vegetable Growing*. London: Frederick Muller.

Holland B, Unwin ID and Buss DH (1991) *Vegetables, Herbs and Spices: Fifth Supplement to McCance and Widdowson's The Composition of Foods*. London: HMSO.

Huxley AJ, Griffiths M and Levy M (eds) (1992) *The New Royal Society Dictionary of Gardening*. London: MacMillan.

Nieuwhof M (1969) *Cole Crops – Botany, Cultivation and Utilization*. London: Leonard Hill.

Nonnecke IL (1989) *Vegetable Production*. New York: AVI.

Rubatzky VE and Yamaguchi MAS (1997) *World Vegetables – Principles and Nutritive Values*, *2nd edn*. Connecticut: AVI.

Salunkhe DK and Kadam SS (eds) (1998) *Handbook of Vegetable Science and Technology: Production, Composition, Storage and Processing*. New York: Marcel Dekker.

Splittstoesser WE (1990) *Vegetable Growing Handbook – Organic and Traditional Methods*. New York: Van Nostrand Reinhold.

Thompson HC and Kelly WC (1957) *Vegetable Crops*. New York: McGraw-Hill.

Weichmann J (ed.) (1987) *Postharvest Physiology of Vegetables*. New York: Marcel Dekker.
Wein HC (1997) *The Physiology of Vegetable Crops*. Wallingford, Oxon, UK: CAB International.
Wein HC and Wurr DCE (1997) Cauliflower, broccoli, cabbage and Brussels sprouts. In: Wien HC (ed.) *The Physiology of Vegetable Crops*. Wallingford, Oxon, UK: CAB International.

Broccoli-type Brassicas *See* **Brassicas**

Leaf Vegetables

R Fordham, Ludgvan, Cornwall, UK
P Hadley, University of Reading, Whiteknights, UK

Introduction

In addition to the diverse selections of the species *Brassica oleraceae* which include nonheading leafy members, there are other species within the genus *Brassica* and also in the family Chenopodiacea which are cultivated as vegetables principally for their edible foliage.

Spinach (*Spinacia oleracea*): Family Chenopodiacea

Spinach is an annual crop grown for its edible leaves. It initially develops as a rosette of leaves on a short stem. Following stem elongation, branching occurs from leaf axils, from which further laterals develop. It is believed to have originated in central Asia and there are records of its cultivation in China in the seventh century and in Europe in the fourteenth century. The Moors introduced it to Spain and the word 'spinach' is derived from the Spanish *hispania*. It is now widely cultivated, particularly in the USA. It has a long history of cultivation. Recent production statistics for the principal growing areas are summarized in **Table 1**.

It is a cool-season crop and grows best at average temperatures in the range 16–18 °C. Young plants can withstand freezing temperatures down to −9 °C. Flowering is induced by long day conditions; the critical day-length ranges from 12.5 to 15 h depending on cultivar and temperature. High temperatures can also cause premature flowering or bolting. A combination of cold exposure followed by high temperatures and long days causes rapid bolting with consequent cessation of leaf growth. Most cultivars produce between 22 and 26 leaves.

The species is mainly dioecious, producing male and female flowers on separate plants, although monoecious plants occur.

Cultivars differ in day-length response and also in their leaf morphology, which may be either smooth or wrinkled (savoyed).

Nutritional Value and Chemical Composition

Spinach is rich in vitamins and is particularly high in carotene, a precursor of vitamin A. It is also rich in iron and calcium, but the calcium is said to be unavailable as it unites with oxalic acid to form calcium oxalate. **Table 2** summarizes the nutritional value and composition of raw spinach leaves. Spinach leaves contain relatively high levels of nitrate (1075–2300 p.p.m.), although these are influenced by a number of factors, including cultivar, fertilizer regime, growing temperatures, and nitrate reductase activity. It has been suggested that frequent consumption of leafy vegetables containing high levels of nitrate may be hazardous as a result of the nitrate being reduced to nitrite by saliva and the upper gastrointestinal tract. It has been shown that nitrate levels are reduced by exposure of light prior to harvesting and that levels increase during storage. Thus, agronomic practices can have a significant influence on the levels of nitrate in the marketed leaves. Processing and culinary treatment will also affect levels of nitrates and other components. For example, canning and freezing at −18 °C can lead to cell rupture and hence ion loss. This has been reported particularly in relation to nitrate

Table 1 Estimated world production of spinach (Mt)

World	8 031 923
China	6 162 000
USA	354 390
Japan	330 000
Turkey	191 000
France	130 000
Korea, Republic of	125 000
Italy	85 367
Pakistan	73 788
Indonesia	65 000
Germany	59 890
Belgium–Luxembourg	58 000
Spain	50 000

Source: Food and Agriculture Organization (2001) http://www.nal.usda.gou/fric/foodcomp.

Table 2 Nutritional value and chemical composition of spinach, New Zealand spinach, and Swiss chard. Values are based on 100 g raw flesh

	Spinach	*New Zealand spinach*	*Swiss chard*
Proximates			
Water (g)	91.58	94.00	92.66
Energy (kJ)	92	59	79
Protein (g)	2.86	1.50	1.80
Total lipid (fat) (g)	0.35	0.20	0.20
Carbohydrate, by difference (g)	3.50	2.50	3.74
Fiber, total dietary (g)	2.7	N	1.6
Ash (g)	1.72	1.80	1.60
Minerals			
Calcium, Ca (mg)	99	58	51
Iron, Fe (mg)	2.71	0.80	1.80
Magnesium, Mg (mg)	79	39	81
Phosphorus, P (mg)	49	28	46
Potassium, K (mg)	558	130	379
Sodium, Na (mg)	79	130	213
Zinc, Zn (mg)	0.53	0.38	0.36
Copper, Cu (mg)	0.130	0.093	0.179
Manganese, Mn (mg)	0.897	0.639	0.366
Selenium, Se (mg)	1.0	0.7	0.9
Vitamins			
Vitamin C total ascorbic acid (mg)	28.1	30.0	30.0
Thiamin (mg)	0.078	0.040	0.040
Riboflavin (mg)	0.189	0.130	0.090
Niacin (mg)	0.724	0.500	0.400
Pantothenic acid (mg)	0.065	0.312	0.172
Vitamin B_6 (mg)	0.195	0.304	0.099
Folate, total (μg)	194	15	14
Vitamin B_{12} (μg)	0.00	0.00	0.00
Vitamin A (μg RE)	672	440	330
Vitamin E (mg ATE)	1.890	N	1.890

N, no analysis available.
RE, retinol equivalents; ATE, alpha-tocopherol equivalents.
Source: US Department of Agriculture (2001) http: //www.nal.usda.gov/fric/foodcomp.

reduction. Sun-drying of spinach before storage is carried out in a number of developing countries but the practice markedly reduces carotene content.

Handling and Storage

Leaf production is at its optimum just below the critical day-length and thus appropriate cultivars are needed to suit latitude and season of production. The crop is established from direct seeding. For overwintering, a hardier form of spinach having prickly seeds is preferable. Spinach leaves can either be harvested individually as required or the entire shoot can be cut off just above ground level, leaving the plant to produce a new growth under favorable conditions. Mechanical harvesting is used commercially, particularly for the processing industries.

Continuous refrigeration just above 0 °C from packaging through to retail is desirable. Hydrocooling or vacuum cooling should be carried out as soon as practical after harvesting in order to remove field heat. Storage at ambient temperatures for as little as 24 h can lead to significant losses in ascorbic acid and vitamin A. Controlled-atmosphere storage can prolong the shelf-life of spinach by about 1 week at 5 °C.

Domestic and Industrial Use

The leaves are prepared by boiling and can also be used raw in salads. The crop is also processed by both freezing and canning.

New Zealand Spinach (*Tetragonia tetragoniodes*): Family Aizoaceae

New Zealand spinach is a minor crop unrelated to true spinach. It is a perennial, trailing vine reaching 2 m in length and cultivated for its edible young shoots and dark green succulent leaves. It is indigenous to New Zealand and the south and western coastal regions of Australia and Tasmania and was introduced to England in the eighteenth century. It has the advantages of being drought-resistant and heat-tolerant but is not as hardy as spinach and is therefore established in the spring after the risk of frost. It can be raised from either direct seeding or from transplants. Seed germination is slow. Harvesting commences about 70 days after sowing by cutting off the terminal 7–8 cm of shoot growth. This may be repeated two or three times during the season following regrowth. The crop is a useful substitute for spinach in hot dry weather. Young shoots are edible either raw or cooked. Information on the nutritional value and chemical composition of New Zealand spinach is included in Table 2.

Swiss Chard *Beta vulgaris* ssp. *cicla*: Family Chenopodiacea

Swiss chard was first recorded in the Mediterranean region and Canary Islands and was popular as long ago as 350 BC. It was introduced to the USA in 1806 by Swiss settlers.

Swiss chard is a biennial cultivated as an annual for its edible leaves. It has a number of synonyms, including silver chard, seakale beet, silver beet, spinach beet, and perpetual spinach.

There are a range of types, including those with dark green spinach-like leaves to examples with large crisp crinkly leaves, pronounced fleshy midribs, and swollen petioles. The latter types are referred to as chards and after removing the leaf lamina the central portions are used as a separate vegetable which may be cooked and served like asparagus. Leaf stalks occur in a range of colors, including white, pink, red and silver.

The crop is less prone to bolting in warm weather than true spinach and forms a useful substitute during the summer when other leafy species are less readily available. Establishment is by seed which is formed in clusters necessitating early thinning. The thinnings can be used as greens and leaves and stems are then harvested 50–60 days later, leaving crowns which form more leaves for later use. The leafy beets are closely related to the beetroot (*B. vulgaris* ssp. *rubra*) and probably developed from an ancestral form native to the Mediterranean region. Cultivation on a small scale is widespread in temperate regions.

The nutritional value and chemical composition of raw Swiss Chard are included in Table 2.

Orache *Atriplex hortensis*: Family Chenopodiacea

Orache or mountain spinach is a tall annual, growing to a height of over 1.5 m. Its edible leaves may be used as a substitute for spinach. They may be eaten raw in salads, stir-fried, and used in soups and various Chinese and Japanese dishes. The crop is established from seed and continuity of supply of tender young leaves may be achieved from successional sowings made between spring and early summer. It is not widely grown as a food crop but colored varieties are more often cultivated as ornamentals.

Texsel Greens *Brassica carinata*: Family Cruciferae

Texsel Greens, also known as Abysinnian cabbage, was introduced into Texas, USA in 1966 as an oil seed crop from which a selection was made with suitable leaf qualities to be used as a vegetable. This annual, originating from the highlands of Ethiopia, initially produces a rosette of leaves from a very short stem. The leaves have slender petioles and are oval or oblong in shape, light green in color, and smooth in texture. They are reported to have a high protein content (3.74%) relative to spinach and they provide a useful source of vitamin C. It has a good flavor, being milder than collard greens and less pungent than mustard greens.

The crop is not frost-hardy but grows rapidly from successional sowings from early spring to produce harvests until autumn. The crop has been trialed in the USA and the UK but is not widely grown.

The young leaves have a short shelf-life. They can be eaten either raw in salads or boiled lightly as a substitute for spinach.

See also: **Brassicas**; Commercial and Dietary Importance; Cabbage and Related Vegetables

Further Reading

Hardy FS and Watson GD (1982) *The Complete Guide to Commercial Vegetable Growing*. London: Frederick Muller.

Huxley AJ, Griffiths M and Levy M (eds) (1992) *The New Royal Society Dictionary of Gardening*. London: MacMillan.

Nonnecke IL (1989) *Vegetable Production*. New York: AVI.

Rubatzky VE and Yamaguchi M (1997) *World Vegetables – Principles and Nutritive Values*, 2nd edn. Connecticut: AVI.

Singhal RS and Kulkarni PR (1998) Leafy vegetables. In: Salunkhe DK and Kadam SS (eds) *Handbook of Vegetable Science and Technology: Production, Composition, Storage and Processing*, pp. 533–587. New York: Marcel Dekker.

Splittstoesser WE (1990) *Vegetable Growing Handbook – Organic and Traditional Methods*. New York: Van Nostrand Reinhold.

Thompson HC and Kelly WC (1957) *Vegetable Crops*. New York: McGraw-Hill.

Weichmann J (ed.) (1987) *Postharvest Physiology of Vegetables*. New York: Marcel Dekker.

Oriental Brassicas

R Fordham, Ludgvan, Cornwall, UK
P Hadley, University of Reading, Whiteknights, Reading, UK

Introduction

Brassica species have been cultivated since antiquity for food and animal fodder. Extensive hybridization and selection over many generations have led to difficulties in classifying crops at the species level. This is particularly true of the brassicas of oriental origin, which occur in many different forms. The present section deals with the main groups of oriental origin within the species *Brassica rapa*, the Alloglabra group of *B. oleraceae*, and finally, the species *Chrysanthemum coronarium* or edible chrysanthemum.

Pak Choi (*Brassica rapa*, Chinensis Group): Family Cruciferae

The Chinensis group includes a number of leafy cultivars referred to as pak choi, plus other flowering types in which the developing flowering shoot is also consumed. The terms Chinese white cabbage, celery mustard, and Chinese mustard are also used to describe this group. Although a biennial, the species is primarily grown as an annual for its edible

leaves. It can, under adverse conditions, flower and produce seed during its first year of growth.

Cultivation of the Chinensis group has been recorded in China since the fifth century. It is now in widespread production throughout the temperate regions of the world and cultivars have also been developed for use in tropical regions. Various sizes, leaf shapes, and green and white leaf stalks are available.

Nutritional Value and Chemical Composition

The oriental brassicas are an excellent source of vitamins and minerals, in particular having high levels of calcium, iron, and vitamin A. Data on nutritional value and composition for pak choi and other oriental brassicas are summarized in Table 1.

Table 1 Nutritional value and chemical composition of pak choi[a], Chinese cabbage[b], and Spinach mustard.[a] Values are based on 100 g raw flesh

	Pak choi	*Chinese cabbage*	*Spinach mustard*
Proximates			
Water (g)	95.32	95.4	92.20
Energy (kcal)	13.00	12	22.00
Energy (kJ)	54.00	49	92.00
Protein (g)	1.50	1.0	2.20
Total lipid (fat) (g)	0.20	0.2	0.30
Carbohydrate, by difference (g)	2.18	1.4	3.90
Fiber, total dietary (g)	1.00	1.2	2.80
Ash (g)	0.80	N[c]	1.40
Minerals			
Calcium, Ca (mg)	105.00	54	210.00
Iron, Fe (mg)	0.80	0.6	1.50
Magnesium, Mg (mg)	19.00	39	11.00
Phosphorus, P (mg)	37.00	28	28.00
Potassium, K (mg)	252.00	130	449.00
Sodium, Na (mg)	65.00	130	21.00
Zinc, Zn (mg)	0.190	0.2	0.170
Copper, Cu (mg)	0.021	0.02	0.075
Manganese, Mn (mg)	0.159	0.3	0.407
Selenium, Se (mg)	0.50	N[d]	0.80
Vitamins			
Vitamin C total ascorbic acid (mg)	45.00	21	130.0
Thiamin (mg)	0.040	0.09	0.068
Riboflavin (mg)	0.070	Tr	0.093
Niacin (mg)	0.500	0.2	0.678
Pantothenic acid (mg)	0.088	0.11	0.178
Vitamin B_6 (mg)	0.194	0.11	0.153
Folate, total (mcg)	65.700	77	158.0
Vitamin B_{12} (mcg)	0.00	0.00	0.00
Vitamin A (IU)[e]	3000	4400	9900
Vitamin A (μg RE)[e]	300	440	990
Vitamin E (mg ATE)	0.120	N[c]	1.704

[a]Source: US Department of Agriculture (2001) http: //www.nal.usda.gov/fric/foodcomp.
[b]Source: Holland B, Unwin ID and Buss DH (1991) *Vegetables, Herbs and Spices: Fifth Supplement to McCance and Widdowson's The Composition of Foods*. London: HMSO.
[c]N, no analysis available.
[d]Tr, trace.
[e]Vitamin A-equivalents.

Handling and Storage

There are many cultivars, ranging from large types over 60 cm tall, to low rosetting forms less than 10 cm. The heads of the larger cultivars can weigh over 2 kg. Plants are established from seed, and spacing is dependent on cultivar. The crop can be harvested in the leaf stage either whole or by picking individual leaves. Flowering shoots are also edible. The leaves are generally consumed fresh but in China, leaves are also dried following blanching in boiling water and stored for winter use in soups.

Chinese Cabbage (*Brassica rapa* Pekinensis Group): Family Cruciferae

Chinese cabbages (also known as Chinese leaves or celery cabbage) are broadly divided into the headed types and the loose-headed types, depending on their propensity to produce tight, compact heads of internal leaves. The crop originated in China and is believed to have been derived from a cross between the Chinensis form from the south and the turnip from the north.

Heading Chinese cabbage is one of the most important vegetables in Eastern Asia. They are the most widely grown vegetable in China and in northern areas of the country account for more than one-quarter of the total annual vegetable consumption. Other major areas of production include Korea and Taiwan. They are also cultivated throughout the world and in recent years have become firmly established as an important vegetable crop.

The Chinese cabbage occurs in both annual and biennial forms but is cultivated as an annual. It forms an upright head of either tightly overlapping leaves or sometimes a looser head of more separated leaves. Shape and size differ markedly between cultivars and head weight may vary between 1.4 and 4.5 kg. Leaf color in the hearted centers is usually a creamy white, but outer leaves vary from dark to light green. The swollen bases of the pronounced mid-ribs often form a solid white 'butt' to the base of the harvested heads.

The headed types vary in shape from tall cylindrical forms which can be 35–45 cm tall and 10–15 cm in diameter to the more squat, barrel-shaped cultivars 20–25 cm in height and 15–23 cm in diameter. There are also loose or semiheading types which are generally easier to grow than the heading types.

Nutritional Value and Chemical Composition

Estimates of nutritional and compositional data are included in Table 1.

Handling and Storage

Chinese cabbage can be either direct-seeded or raised as transplants. A succession of crops can be achieved from a combination of raising early transplants under protection followed by a succession of outdoor sowings. The time from sowing to harvest can vary from around 50 to 100 days depending on cultivar and season. The loose-headed types are more cold-tolerant and can be used to extend the period of supply. The crop is generally consumed fresh but storage at 0–1 °C and high humidity can extend shelf-life for up to 2 months. Trials in Germany have demonstrated the possibility of storage of some cultivars for up to 3 months under controlled atmosphere and low temperatures.

Domestic and Industrial Uses

Chinese cabbage are eaten raw or lightly cooked by, for example, stir-frying. More than 90% of the Chinese cabbage harvested in Korea is used for making Kimchi, a fermented side-dish eaten throughout the year by virtually every family.

Spinach Mustard (*Brassica rapa* Perviridis Group): Family Cruciferae

Spinach mustard or tendergreen is an annual to biennial. It is a type of leafy turnip and has a thickened and tuberous root crown. It is a tall plant, reaching a height of 45 cm and having large leaves, on average 30 cm long and 18 cm wide. It has the main advantage of being cold-tolerant, surviving temperatures down to −14 °C. Although largely unknown in China, it is grown in Japan, Korea, and Taiwan but not grown on a large scale elsewhere. Leaves of spinach mustard are eaten raw in salads to add flavor.

Nutritional Value and Chemical Composition

Estimates of nutritional and compositional data are included in Table 1.

Mizuna and Mibuna Greens (*Brassica rapa* Nipposinica Group): Family Cruciferae

These two vegetables are relatively recent introductions to the west from Japan. Botanically they are closely related to leafy turnips. Mizuna forms a bushy clump of finely dissected, feathery, dark, glossy green leaves. The white leaf stalks are slender and juicy. Plant height is around 23 cm and spread is about 45 cm.

Mibuna has long (30–45 cm), narrow (1–4 cm) leaves in clumps which can be 56 cm in diameter. Mibuna is a cool-weather plant which can survive temperatures below freezing, but, of the two, mizuna is both more heat- and cold-tolerant.

These vegetables are usually eaten raw in salads.

Chinese Broccoli (*Brassica oleraceae* Alboglabra Group): Family Cruciferae

Chinese broccoli or Chinese kale is more closely related to the European cabbages than the Chinese cabbages. It is an annual grown primarily for its flowering stem. It grows to a height of about 45 cm to produce a smooth, succulent flowering stem 1–2 cm in diameter. There are both yellow and white flowering cultivars. It is a vigorous, fast-growing species able to withstand both frost and high temperatures. The flower stems are harvested at an early stage while flowers are still in bud. It is usually cooked by stir-frying.

Chrysathemum Greens (*Chrysanthemum coronarium*): Family Compositae

The edible chrysanthemums are known under a wide range of names, including chop suey greens, Japanese greens, crown daisy and garland chrysanthemum. It originated in the Mediterranean region and spread through Europe into Africa and Asia. In the west, selections have been entirely towards its value as an ornamental species and it is only in China, Japan, and South-East Asia that it has been adopted as a vegetable.

The edible chrysanthemum is an annual, growing to about 30 cm until flowering, when it may reach almost a meter. It is grown mainly for its leaves and young stem which can be used raw in salads, lightly boiled, or in the preparation of soups. The small yellow, daisy-like flowers may also be eaten but have a strong aromatic flavor. Removing flowers as they develop on the growing plant helps to maintain leaves in a tender, mild-flavored condition.

There are three main types distinguished on the basis of leaf form and size. Those with finely divided, narrow leaves are darker in color and generally considered to be more cold-tolerant than those with larger leaves. Other types include those with medium and larger leaf forms.

The crop may be established from seed, either from direct sowings or transplants, and by the use of rooted cuttings taken in the spring from plants overwintered under protection.

See also: **Brassicas**

Further Reading

Herklots GAC (1972) *Vegetables in South-East Asia*. London: Allen and Unwin.

Holland B, Unwin ID and Buss DH (1991) *Vegetables, Herbs and Spices: Fifth Supplement to McCance and Widdowson's The Composition of Foods*. London: HMSO.

Huxley AJ, Griffiths M and Levy M (eds) (1992) *The New Royal Horticultural Society Dictionary of Gardening*, 4 vols. London: MacMillan.

Larkom J (1991) *Oriental Vegetables*. London: Murray.

Opena RT, Kuo CG and Yoon JY (1988) *Breeding and Seed Production of Chinese Cabbage in the Tropics and Subtropics*. Taiwan: Asian Vegetable Research and Development Centre (AVRDC).

Rubatzky VE and Yamaguchi MAS (1997) *World Vegetable–Principles and Nutritive Values*, 2nd edn. Connecticut: AVI.

Table 1 World production of carrots

	Area (10^3 ha)	*Production (10^3 MT)*
World	926	
Continents		
Africa	72	902
North and Central America	80	2639
South America	46	981
Asia	421	7761
European Union	78	3441
Former Soviet Republic	187	2648
Leading countries		
China	289	5120
Russia	92	1350
USA	52	1941
Poland	35	947
Nigeria	27	231
India	24	340
Japan	23	677
Mexico	16	378
UK	13	674
Italy	13	588
France	12	496
Germany	9	432
Canada	8	279

Source FAO (2001), with permission.

Carrot, Parsnip, and Beetroot

P Hadley, University of Reading, Whiteknights, Reading, UK
R Fordham, Ludgvan, Cornwall, UK

Introduction

Carrot, parsnip, and beetroot together represent the most popular root vegetables consumed in temperate areas of the world. Carrot and parsnip are largely grown as a fresh vegetable whereas beetroot is also processed and sold either as boiled roots or diced or sliced and pickled in vinegar.

Carrot – *Daucus carota* var. Sativus, Family Umbeliferae

Carrots are a popular vegetable crop throughout most temperate regions of the world (**Table 1**), with only Brassicas and peas ranking higher in terms of importance. Consumption has increased considerably over the last 40 years. Carrots are generally grown on a large farm scale, using a high degree of mechanization. Large-scale farm operation has become more important for carrot-growing over the last decade, with the increasing need to supply supermarket chains and multiples rather than wholesale markets. Here, success is measured in terms of being able to program carrot production carefully so that supermarket demands can be accurately met throughout the year. Developments in production and storage techniques in recent years now enable carrot growers to achieve this.

Carrot originated from middle Asia (Punjab or Kashmir) but later spread to the Near East and, through further domestication, to Europe and China. It was originally used as a medicinal plant and it was not until the tenth to the twelfth centuries that it became used as a food crop. The taproot is woody in the wild form. However, breeding and selection have enabled the root of the cultivated form (subsp. *sativus*) to become thickened and succulent. Modern European cultivars have resulted from selections made in the seventeenth century, for orange roots, in Holland, and subsequently in France.

Carrot grows vegetatively in the first year, during which it develops an enlarged hypocotyl and a prominent taproot. The root is composed of an internal pith or core composed of xylem and an outer cortex composed largely of phloem. Higher quality is associated with roots composed largely of cortex with a smaller proportion of pith. Color is also important: high quality is associated with a uniform deep orange coloration. The roots have a sweet flavor which is associated with a high concentration of soluble carbohydrates. The roots also have a high carotene content that is associated with the orange pigmentation of the root.

The orange color has been developed through continuous selection and is not a characteristic of its wild forebears. Prior to this, cultivated carrots were purple-rooted owing to the presence of anthocyanin pigments, but white and yellow anthocyanin-free types became preferred and it was from these that modern orange-rooted cultivars have been derived. Some large-rooted white and yellow forms remain in use for stock-feeding.

The leaves form a rosette and are characteristically two- to three-pinnate. During the second year, the stem elongates and flowers are produced initially as terminal umbels and later as secondary umbels. Flowers are pollinated by bees and other winged insects. The seeds are spiny, hooked, and slightly curved.

Cultivars

Carrots fall into different classes according to their shape, size, and time to maturity (Figure 1). Well-known groups of cultivars are as follows:

- Chantenay – Medium-sized, conical, stump-rooted types. Late-maturing. Widely grown for canning whole, for dehydrating, and for the fresh market. Medium and large-sized roots may be sliced or diced and then either canned or frozen.
- Amsterdam forcing – Small to medium-sized, slender, cylindrical, stump-rooted types of good quality. Early maturity. Grown for early-season bunching crops and for prepacking.
- Nantes – Medium, cylindrical, stump-rooted types. Medium maturity. Grown for prepacking and to a lesser extent for the fresh market and for canning.
- Berlicum – Large-sized, cylindrical, stump-rooted types. Late-maturing. Grown for the fresh market and for prepacking.
- Autumn King – Very large-sized, tapering rooted types. Late-maturing and heavy-yielding but flesh color is generally inferior with a rough skin texture. Grown mainly for the fresh market but also to a limited extent for dicing prior to canning and freezing.

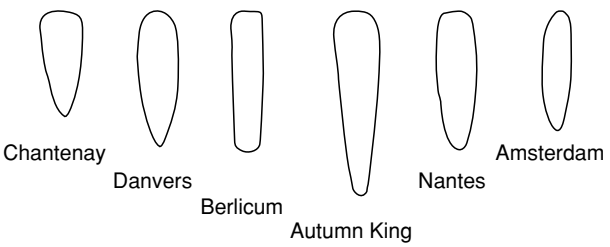

Figure 1 Root shape in carrot types. Reproduced from Vegetables of Temperate Climates: Carrot, Parsnip and Beetroot, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Handling and Storage

Carrots are capable of growing in any area where the growing season is cool; the optimum growth temperature is typically 16–18 °C. Under excessively cool conditions there is a tendency for the plants to bolt (produce flower stalks in the first year); under hot conditions, plants produce excessive leaf growth at the expense of root growth. Soil needs to be free from stones and clods, as this often leads to fanged roots. The crop requires a steady supply of available soil moisture in order to produce high yields and good quality. The crop has a relatively low fertilizer requirement. It is possible for UK carrot growers to supply produce all year round. This has been possible by extending the main growing season using a combination of field and refrigerated storage together with early production under plastic film and nonwoven agrotextile crop covers. In the UK, the earliest sowings for conventional production of carrots (i.e., without the use of plastics) are possible from the last week in February and continue through to the end of April under UK conditions. This provides production of carrots from July until October. Earlier production from the end of May until the beginning of July is possible from sowings made during January and February under perforated plastic-film crop covers. Such early crops (often referred to as second earlies) are typically grown under 200-hole polyethylene (200×1 cm diameter holes m^{-2}) supported by the crop. Certain carrot cultivars can also be overwintered under plastic-film crop covers to get even earlier maturity. This is achieved by sowing in September. Sowing date is critical as, if this is too early, the crop will bolt due to the large plants being sensitive to the vernalizing effect of cold winter temperatures. Alternatively, if the sowing date is too late, the crop will not be large enough to survive the winter satisfactorily. The latest sowings are made in late April and May. These sowings supply crops from October to May through crop storage under field conditions, or until the end of June if carrots are stored in refrigerated cold stores. Seed spacing is varied to produce roots of different sizes. Densities vary from 55–110 plants m^{-2} for large carrots suitable for dicing (greater than 45 mm), 160 plants m^{-2} for medium-sized roots for prepacking (32–43 mm), 380 plants m^{-2} for small roots for canning (18–32 mm), to 430 plants m^{-2} for freezing (20–25 mm).

Harvesting is now fully mechanized. When ready for harvest, roots are typically loosened, lifted by their tops, and the roots and tops conveyed to an

area where the tops are removed. The carrots are then washed and sorted before placing in pallet boxes for storage. Early-season carrots are marketed with the tops attached (bunching carrots). These are still hand-pulled and tied into bunches, washed, and cooled. Baby carrots are produced by growing carrots at high densities. These carrots are harvested using modified radish-harvesting equipment. After removing the tops, the roots are loosened using an undercutting device. The roots are then lifted on a conveyor, and the soil is shaken free to leave the carrots which are then washed.

Carrots are moderately resistant to frosts, and crops can usually be stored in the field during winter until required. Field storage is achieved by earthing up with 150 mm soil or placing 300 mm depth of straw over the crop in November to December to protect the crop from frosts. The stored crops are then lifted as and when required. For longer-term storage, carrots are lifted under good weather conditions and placed in cold stores (normally in bulk bins) at 0–10 °C and 95–98% relative humidity. Carrot roots have high respiration rates and their temperature needs to be closely monitored during storage to insure that carrot temperatures remain low to avoid loss of quality.

Domestic and Industrial Uses

Carrots are marketed washed in prepacks, nets, bags, or in bulk (the latter is common for supermarkets that allow customers to select and weigh out the produce they require). Prepackaging is now common and usually consists of half or 1 kg of produce in perforated plastic bags. Approximately half the UK crop is sold to households for home consumption; the remainder is grown for processing (canning, frozen foods, soup manufacture) or is sold to caterers. Although carrots are mainly marketed as a fresh product, there is a trend, particularly in the USA, for carrots to be processed as a prepared vegetable. Here it is peeled and sliced.

Carrots can be eaten fresh or, with very little preparation, the roots can be boiled, steamed, fried, roasted with meats, sautéed, and pickled. They are processed for canning, freezing, and diced for incorporation into soups. The crop is also processed for production of carrot juice.

Food Value and Chemical Composition

Cultivars grown for human consumption have about 11% dry-matter content, of which approximately 40–50% consists of sugars. The orange color is derived from the high β-carotene content. β-carotene is a precursor of vitamin A and carrots have been recognized as a remedy for night-blindness, which is brought about by vitamin A deficiency. Food value and chemical composition of raw old and young carrots are summarized in Table 2.

Table 2 Nutritional value and chemical composition of carrot, parsnip, and beetroot. Values are based on 100 g raw flesh

	Carrot	*Parsnip*	*Beetroot*
Proximates			
Water	88.8 g	79.3 g	87.1 g
Sugar	5.6 g	5.7 g	7.0 g
Starch	0.1 g	6.2 g	0.6 g
Dietary fiber	2.6 g	4.3 g	2.8 g
Total nitrogen	0.11 g	0.29 g	0.27 g
Energy value (kcal)	30	64	36
Energy value (KJ)	125	271	154
Protein	0.7 g	1.8 g	1.7 g
Fat	0.5 g	1.1 g	0.1 g
Carbohydrate	6.0 g	12.5 g	7.6 g
Minerals			
Na	40 mg	10 mg	66 mg
K	240 mg	450 mg	380 mg
Ca	34 mg	41 mg	20 mg
Mg	9 mg	23 mg	11 mg
P	25 mg	74 mg	51 mg
Fe	0.4 mg	0.6 mg	1.0 mg
Cu	0.02 mg	0.05 mg	0.02 mg
Zn	0.02 mg	0.3 mg	0.4 mg
S	7 mg	17 mg	16 mg
Cl	39 mg	49 mg	59 mg
Vitamins			
Retinol	0 μg	0 μg	0 μg
Carotene	5330 μg	30 μg	20 μg
Vitamin D	0 μg	0 μg	0 μg
Thiamin	0.04 mg	0.23 mg	0.01 mg
Riboflavin	0.04 mg	0.01 mg	0.01 mg
Niacin	0.2 mg	1.0 mg	0.1 mg
Vitamin C	4 mg	17 mg	5 mg
Vitamin E	0.56 mg	1.0 mg	Trace
Vitamin B_6	0.07 mg	0.11 mg	0.03
Vitamin B_{12}	0 μg	0 μg	0 μg
Folate	28 μg	87 μg	150 μg
Pantothenic acid	0.25 mg	0.50 mg	0.12 mg
Biotin	0.6 mg	0.1 mg	Trace

Reproduced from Holland B, Unwin ID and Buss DH (1991) *Vegetables, Herbs and Spices: Fifth Supplement to McCance and Widdowson's The Composition of Foods*. London: HMSO, with permission.

Parsnip–*Pastinacea sativa*, Family Umbelliferae

Parsnip is grown for its edible taproot. It is characteristically very hardy, producing a creamy white root that has been valued as a winter vegetable, although there is increasing demand for parsnip all the year round. It differs from carrot in having simple pinnate leaves and a creamy-white, fleshy root. It also has a longer growing season than carrot. It originates from either the Mediterranean or northern Europe. In the UK, the wild parsnip (subsp. *sylvestris*), which has a

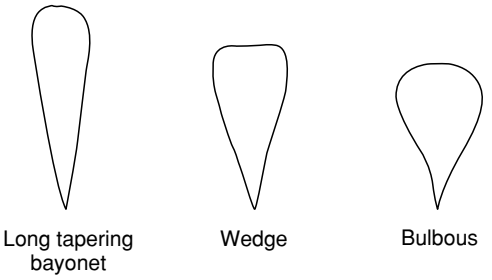

Figure 2 Parsnip root shape. Reproduced from Vegetables of Temperate Climates: Carrot, Parsnip and Beetroot, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

slender woody taproot, inhabits limestone soils. It has had a long period of domestication and was grown as a food crop by the Ancient Greeks.

Parsnip is a biennial producing a fleshy taproot in the first year and developing a flower stalk in the second year: this is terminated by a flower which is a broad, compound umbel. Flowers are pollinated by bees and blowflies. Parsnip types are distinguished by the shape of their roots. Three broad categories can be distinguished: bulbous cultivars, wedge-shaped cultivars, and long, tapering or bayonet cultivars (Figure 2). Long, tapering roots are the more traditional cultivars, although wedge and bulbous cultivars are now favored as they are less prone to breakage during mechanical harvesting. Breeding has largely been toward canker-resistant, shallow-crowned types that are more suitable to mechanical harvesting. Average yields are approximately 20 t ha^{-1}. Generally parsnips are grown in the same areas as those used for carrot production.

Handling and Storage

The crop is slow-growing, preferring a deep, rich, friable sandy loam soil which allows the development of straight roots without fanging. In the UK, the earliest crops are produced in small quantities from July onwards, although the main market for this crop is during the winter months. It is very hardy and resistant to frosts. Frosting the roots during the winter months improves their sweetness.

Parsnips are grown for the fresh market, prepacking, and processing. Small roots (40–65 mm diameter) are required for prepacking, whilst a larger root is required for the fresh market (40–130 mm diameter). Very large roots are suitable for processing in soup manufacture. A range of plant densities are used to achieve these different root sizes. Typically, parsnips for the fresh market are grown at a density of 20 plants m^{-2}, whereas the smaller roots required for prepacking are grown at a density of 40–60 plants m^{-2}. For early crops, sowings in the UK occur in February whilst the main crop is sown from the end of March until the end of April. Seed is slow to germinate and has an unusually short period of viability of between 1 and 2 years. The seed has an unusual flat shape and is difficult to drill. In addition, because of its notoriously poor germination (generally no better than 60%), it is one of the few crops for which there is no statutory minimum germination rate.

Harvesting is similar to that described for carrots. Although the crop can withstand frosts, it is generally lifted and stored in unheated barns for 3 weeks prior to marketing. This short-term storage is to provide stocks if subsequent weather conditions make harvesting difficult.

Domestic and Industrial Uses

Parsnips are generally eaten cooked. They are processed for canning, freezing, and diced for incorporation into soups.

Food Value and Chemical Composition

Food value and chemical composition of raw parsnips are summarized in Table 2.

Beetroot – *Beta vulgaris* subsp. *rubra*, Family Chenopodiacae

Beetroot has been used for human consumption for over 2000 years. The crop was introduced into Europe during the middle of the sixteenth century. Genotypes that are currently cultivated are probably derived from eastern Mediterranean types. Beets (of the same species, *B. vulgaris*) also grown for their swollen roots include sugar beet, mangles, and fodder beet. The wild ancestor grows on the coasts of Europe and North America. Although well-known varieties have deep red flesh with characteristic concentric circles of vascular tissue, yellow and white varieties also exist, although the latter are not commercially important and are grown mainly for their novelty value. The deep red color is caused by anthocyanin pigments within the cell sap and selection has aimed at producing roots of even and very dark color. Beetroot varieties are classified into three groups according to their shape: (1) round beet, which are the most wildly grown, being favored for their tender, sweeter-tasting flesh, e.g., globe Detroit; (2) intermediate-shaped

beet; and (3) long beet, e.g., Cheltenham greentop, which can reach up to 30 cm long and is popular for certain markets. Round, long, and intermediate-rooted cultivars are grown for main crops for winter use; round or flat cultivars are used for earlier production and for forcing. Many cultivars are available; choice is determined by the shape desired, bolting resistance (particularly for early-season production), and internal color.

Domestic and Industrial Uses

Although the root can be marketed fresh, it is often sold ready-cooked. For early and forced beet, the foliage is trimmed and the roots bunched into sixes; later crops are topped and loose-packed. For the supermarket and greengrocery trade, beetroots are now cooked and usually presented in the form of prepacks. Processed beet is the major outlet for beetroot where it is pickled whole, sliced or, to a lesser extent, diced into glass jars. Processors demand beet to be graded into specific sizes for particular outlets.

Handling and Storage

The crop is sown relatively late as plants sown too early in the season are prone to bolt (run to seed). The seed is actually a fruit enclosing one, two, or three seeds. The fruit tissue often contains germination inhibitors which can delay germination. Since the 'seed' can give rise to clusters of up to three seedlings, this tends to increase variability in root size at harvest and thus a significant fraction of roots that lie outside their market specifications. This problem is reduced in new monogerm varieties which only contain one seed per fruit. Seed is sown from late March for early bunched crops and from April until the end of July for the main crop. Early forced beetroot can be ready as bunched beetroot, when the roots are about 35 mm diameter in late spring and early summer (June to July). On light soils, lifting can be carried out by hand. Roots of early red beet are easily damaged by bruising and so care is needed at all stages of handling. Main crop beetroot is generally ready to lift between the end of September and the beginning of November. Maturity is generally indicated when the leaves fall away from the upright position.

The crop is normally topped in the field and harvested mechanically with a digger elevator. The roots are damaged by frosts and must be lifted prior to the onset of winter. Beetroot can be stored in clamps, as for other root crops or potatoes. There is a continuous demand for red beet throughout the year that is met either from the field or from clamps for 10 months of the year. These supplies fall off from the end of April until the middle of June. This is met either through imports or from improved extended-storage methods that have been introduced over the last decade.

Food Value and Chemical Composition

Cooked beet contain approximately 10% carbohydrate, 1.5–2% protein and a calorific value of 46 cal per 100-g root. Its vitamin content, in contrast, is low, and its vitamin C content is reduced to less than 50% by cooking. Food value and chemical composition of raw beetroot are summarized in Table 2.

Glossary

Bolting The premature running to seed of a biennial crop before harvesting. This results, in the case of root crops, in coarse, unmarketable roots.
Clamp An outside store for root crops, consisting of layers of straw and soil covering a heap of roots to protect them from frost.
Digger elevator A harvestor for lifting root crops.
Monogerm seed A seed cluster which germinates to give one seedling.

See also: **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability; **Vegetables of Temperate Climates**: Carrot, Parsnip, and Beetroot; Swede, Turnip, and Radish; Miscellaneous Root Crops; Stem and Other Vegetables

Further Reading

Anonymous (1970) *Parsnips*. Advisory leaflet 151. London: Ministry of Agriculture, Fisheries and Food.
Anonymous (1972) *Carrots for Processing*. Horticultural Enterprises Booklet. London: Ministry of Agriculture, Fisheries and Food.
Benjanin LR, McGarry A and Gray D (1997) The root vegetables; beet, carrot, parsnip and turnip. In: Wien HC (ed.) *The Physiology of Vegetable Crops*. CAB International.
Hardy FS and Watson GD (1982) *The Complete Guide to Vegetable Growing*. London: Frederick Muller.
Holland B, Unwin ID and Buss DH (1991) *Vegetables, Herbs and Spices: Fifth Supplement to McCance and Widdowson's The Composition of Foods*. London: HMSO.
Kotecha PM, Desai BB and Madavi DL (1998) Carrots. In: Salunkhe DK and Kadam SS (eds) *Handbook of Vegetable Science and Technology: Production, Storage and Processing*. New York: Marcel Decker.
Shoemaker JS (1947) *Vegetable Growing*. New York: Wiley.

Swede, Turnip, and Radish

P Hadley, University of Reading, Whiretnights, Reading, UK
R Fordham, Ludgvan, Cornwall, UK

Introduction

The family Cruciferae includes some of the most important temperate vegetables, of which there are three commercially important root crops.

Swedes – *Brassica napus* L. var. *napobrassica Peterm*, Family Cruciferae

Swede (also known as rutabaga, Swedish turnip, or Lapland turnip) is usually considered to be a root crop. However, the 'root' is mainly the swollen hypercotyl of the plant. The crop is grown as an annual, as a fodder crop, or for human consumption, and is generally grown on a farm scale. It is thought to be a relatively recent vegetable that originated as a natural hybrid of *Brassica rapa* (turnip) and *B. oleracea* (cabbage). This hybrid may have occurred first in Bohemia in the seventeenth century, although some reports suggest that its origin is much earlier. The swede was initially grown in Sweden (from where it takes its name) and later spread to the rest of Europe and North America. Swedes have an elongated globe shape with a swollen neck bearing a number of leaf scars; this neck differentiates swedes from turnips.

Roots may be globe or tankard-shaped, or intermediate between the two. The flesh is either white or yellow/orange and has a milder flavor and a higher sugar content than winter turnips. Yellow-fleshed types are more commonly grown. The root develops anthocyanin and/or chlorophyl in its outer skin where it is exposed to the light, giving rise to a range of skin colors. Four skin colors exist: green, light or dark purple, and orange. Green-topped swedes are the poorest croppers, but the best keepers, whilst the reverse is true for purple-topped swedes.

Handling and Storage

Swedes can be grown in a wide range of soils; however, the crop is best suited to a well-drained, medium loam. The crop has a relatively low nutrient requirement and it rarely needs irrigation except on light, sandy soil. In the UK, seeds are sown from March to the end of June and are ready for harvesting 15–21 weeks later. Average yields are 32 t ha^{-1} and good yields are typically 68 t ha^{-1}. The crop is pulled by hand or is machine-harvested by digger elevators. The root is removed and the neck is trimmed off. The crop is sold for both the prepack and the wholesale markets. For prepacking, small roots are required, normally between 80 and 150 mm. Swedes for the wholesale markets are normally marketed in nets with minimum grading.

Storage normally commences in the autumn in traditional clamped heaps covered with straw or soil and is used to supply the market over the winter period. Flavor is improved after the roots have been exposed to early frosts. A more modern approach to storage is to harvest into bulk bins and then surround the bulk bins with straw bales. The crop is a low-value crop and does not merit more sophisticated storage conditions, although, for the longest storage period, the roots can be stored in an insulated store with forced ventilation at a temperature of 0 °C and a relative humidity of 90–95%. The crop can be left standing in the field and harvested as required.

Domestic and Industrial Uses

Swedes are normally eaten as a cooked vegetable but they can also be shredded raw in salads. They are processed for canning, and included in pickles and mixed and diced vegetables.

Food Value and Chemical Composition

Food value and chemical composition of raw swede are summarized in Table 1.

Turnips – *Brassica rapa*, Family Cruciferae

Turnips are very similar vegetables to swedes and the two crops are often confused. Roots are either globe-shaped or have a more flattened shape, with white or yellow flesh, and have a high water content. The turnip is believed to have originated from a wild form of Eurasian origin and has been known for about 4000 years. The same species is believed to have been responsible for Chinese cabbage and other similar oriental vegetable crops. Although the turnip (like swede) is described as a root, it is actually a swollen hypocotyl and contains very little root tissue. The color of the skin may be white, purple, green, or mottled. The yellow-fleshed types are hardier than white types. Roots are generally small and may be round, flat long, half long, or globe-shaped. They rarely exceed 100 mm in diameter. The number of different types allows turnips to be produced throughout the year, although the crop is mainly winter- or spring-grown.

Handling and Storage

Turnips tend to be an intensively grown crop. A sandy loam soil is preferable for early-spring-sown crops;

Table 1 Nutritional value and chemical composition of swede, turnip, and radish. Values are based on 100 g raw flesh

	Swede	*Turnip*	*Radish*
Proximates			
Water	91.2 g	91.2 g	95.4 g
Sugar	4.9 g	4.5 g	1.9 g
Starch	0.1 g	0.2 g	Trace
Dietary fiber	2.4 g	2.5 g	0.9 g
Total nitrogen	0.11 g	0.14 g	0.11
Energy value (kcal)	24	23	12
Energy value (KJ)	101	98	49
Protein	0.7 g	0.9 g	0.7 g
Fat	0.3 g	0.3 g	0.2 g
Carbohydrate	5.0 g	4.7 g	1.9
Minerals			
Na	15	15 mg	11 mg
K	170	280 mg	240 mg
Ca	53 g	48 mg	19 mg
Mg	9 g	8 mg	5 mg
P	40 mg	41 mg	20 mg
Fe	0.1 mg	0.2 mg	0.6 mg
Cu	0.01 mg	0.01 mg	0.01 mg
Zn	0.3 mg	0.1 mg	0.2 mg
S	39 mg	22 mg	38 mg
Cl	31 mg	39 mg	27 mg
Vitamins			
Retinol	0 μg	0 μg	0 μg
Carotene	350 μg	20 μg	Trace
Vitamin D	0 μg	0 μg	0 μg
Thiamin	0.15 mg	0.05 mg	0.03 mg
Riboflavin	Trace	0.01 mg	Trace
Niacin	1.2 mg	0.4 mg	0.4 mg
Vitamin C	31 mg	17 mg	17 mg
Vitamin E	Trace	Trace	0 mg
Vitamin B_6	0.21 mg	0.08 mg	0.07 mg
Vitamin B_{12}	0 μg	0 μg	0 μg
Folate	31 mg	14 mg	38 mg
Pantothenic acid	0.11 mg	0.2	0.18 mg
Biotin	0.1 mg	0.1 mg	N

N, data not available.
Reproduced from Holland B, Unwin ID and Buss DH (1991) *Vegetables, Herbs and Spices: Fifth Supplement to McCance & Widdowson's The Composition of Foods*. London: HMSO, with permission.

however, the crop can be grown in most soils providing that they are reasonably deep, friable, and free-draining. The time from sowing to harvest is generally short (between 6 and 10 weeks, depending on the season). The crop season usually begins in early April in the UK; these crops are normally picked when the roots are approximately 50 mm in diameter, washed, and sold as bunches of three, six, or eight tied together with about 75 mm of the leaf stalk left attached to the root. (Average yields are approximately 24 t ha^{-1}.) The main crop is harvested through to November. Turnips require similar storage conditions to other root crops and can be stored in clamps (as for swedes). The main crop is washed, graded, and marketed in nets (chiefly to the greengrocery trade) and polyethylene film bags (for the supermarkets).

Domestic and Industrial Uses

Turnips are normally eaten as a cooked vegetable. As with swedes, the crop is processed for use in pickles, and in mixed and diced frozen vegetables. Turnip tops are also eaten as an early spring-green vegetable.

Food Value and Chemical Composition

Food value and chemical composition of raw turnips are summarized in Table 1.

Radish – *Raphanus sativus* L., Family Cruciferae

Radish is a fast-maturing root crop, probably originating from central and western China and India. It has been cultivated for thousands of years and was grown by the Egyptians and ancient Greeks. It was first grown in western Europe in the mid sixteenth century.

The main edible part of the plant is a swollen hypercotyl and primary root. The lack of lignification of the vascular tissue and the presence of considerable thin-walled parenchyma are responsible for the succulence of the root system. The enlarged roots are available in a wide range of sizes, shapes, and colors, depending on cultivar. Early-maturing European salad types can be round, oval-shaped, or cylindrical, red, white, or red with a white tip and with crisp white flesh. Winter and oriental radishes are often white (although red and purple skins are also found) with a large conical shape and white flesh. Early-maturing radishes produce small roots which are harvested when approximately 1.25 cm in diameter. Winter and oriental radish roots are often 25–40 cm in length and can weigh approximately 2–3 kg (although even larger roots are available).

Handling and Storage

European cultivars Early radishes are grown under glass or perforated polyethylene and nonwoven agrotextile crop covers to provide a crop before outdoor crops are mature. Commercially, the crop is produced by specialist growers and market gardeners. The crop is grown sequentially to provide a continuous supply to meet market demand. The crop is normally lifted after 3–6 weeks when the plant has approximately four true leaves. The crop can be harvested by hand, although harvesting machinery is available, particularly for growers supplying the prepack market. For the fresh market, plants are bunched (usually 10–12 roots per bunch) and secured with an elastic band. It is possible to store the crop for 3–4 weeks at 0–1 °C.

Winter and oriental cultivars These types take longer to mature. Radishes grow best in moisture-retentive, rich sandy loam soils whilst heavier soils often cause misshapen roots. They are normally harvested during the autumn until the onset of hard frosts between 8 and 14 weeks after sowing. After this radishes can be stored through the winter (typically for up to 6 months) using the same storage conditions as for carrots.

Domestic and Industrial Uses

Early-maturing radishes are eaten raw in salads, whilst winter and oriental radishes are either eaten raw or as a cooked vegetable, rather like turnips. The very characteristic flavor is popular in Japan, the Philippines, and Hawaii and roots are used to prepare food products such as takuwan and cabaizaku. Leaves and pods of some cultivars can be boiled and eaten as a vegetable. Roots have been used medicinally in the treatment of liver and gallbladder complaints.

Food Value and Chemical Composition

Pungent flavor is due to isothiocyanates. Radishes are a good source of ascorbic acid (15–40 mg $100\,g^{-1}$); a higher concentration is found in pink-skinned cultivars. Catechol has been reported to be present in the red cultivars and flavinols have been detected in minute quantities. Food value and chemical composition of raw red and white radishes are summarized in Table 1.

See also: **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability; **Vegetables of Temperate Climates**: Carrot, Parsnip, and Beetroot; Miscellaneous Root Crops; Stem and Other Vegetables

Further Reading

Anonymous (1980) *Grower Guide No. 14: Swedes and Turnips for the Fresh Market*. London: Grower Books.

Benjanin LR, McGarry A and Gray D (1997) The root vegetables; beet, carrot, parsnip and turnip. In: Wien HC (ed.) *The Physiology of Vegetable Crops*. UK: CAB International.

Hardy FS and Watson GD (1982) *The Complete Guide to Vegetable Growing*. London: Frederick Muller.

Holland B, Unwin ID and Buss DH (1991) *Vegetables, Herbs and Spices: Fifth Supplement to McCance & Widdowson's The Composition of Foods*. London: HMSO.

Kay DE (1973) *Tropical Products Institute Crop and Product Digest No. 2: Root Crops*. Tropical Products Institute.

Masalkar DL and Salunkhe DK (1998) Other roots, tubers and rhizomes. In: Salunkhe DK and Kadam SS (eds) *Handbook of Vegetable Science and Technology: Production, Storage and Processing*. New York: Marcel Decker.

Oldham CH (1950) *Vegetable Growers' Guide*. London: Lockwood.

Shoemaker JS (1947) *Vegetable Growing*. New York: Wiley.

Tindall HD (1968) *Commercial Vegetable Growing*. Oxford: Oxford University Press.

Miscellaneous Root Crops

P Hadley, University of Reading, Whiteknights, Reading, UK
R Fordham, Ludgvan, Cornwall, UK

Introduction

A number of commercially less important temperate root crops exist. These are either grown as domestic crops, or on a limited scale by specialist commercial growers.

Salsify – Tragopogon porrifolius, Family Compositae

Salsify (also known as white oyster, vegetable oyster, or oyster plant because of its oyster-like flavor) is a root crop that is little known by general consumers. It therefore has little importance commercially and is only produced by specialist growers. The root is long and tapering (20–30 cm long and 2.5–5 cm in diameter), with a creamy white skin and a pale yellowish flesh. The leaves are long and grass-like and have a grayish-green color. It is favored as a winter vegetable. Large purple flower heads are produced in the second year. The flowers can be eaten if they are cut before they get too woody.

Handling and Storage

Salsify is usually sown at the end of March in the UK and roots can be harvested from the middle of October onwards. A long growing season is required in order to produce large roots. A sandy loam or peat soil with good water retention provides the best growing conditions. Salsify is winter-hardy and can withstand freezing conditions; indeed, it is suggested that freezing can improve the oyster-like flavor. Salsify is treated in the same way as parsnips and can be stored in clamps or can be left in the ground to be lifted as required. It is possible to extend the storage period using controlled-atmosphere storage

with the temperature maintained at 0 °C, carbon dioxide at 3%, and oxygen at 3%. High humidity is essential to prevent shriveling. In the spring the unlifted roots develop young and succulent shoots known as chards.

Food Value and Chemical Composition

Salsify has a similar nutritive value to parsnips but has a lower calorific value (40 cal 100 g^{-1}) and is valued as a source of fiber. It is low in sodium, its vitamin content is low; it is low in fat and cholesterol-free. Food value and chemical composition of raw salsify are summarized in Table 1.

Celeriac – *Apium graveolens*, Family Unbelliferae

Celeriac is also known as turnip-rooted celery or knob celery. It belongs to the same species as celery; however, in this case the wild ancestor has been bred to produce a 'root' that resembles a turnip. Celeriac is not a true root but is a stem which has developed with a large mass measuring up to 15 cm in diameter. It is produced half in the ground and half above, and from this a rosette of leaves is produced with a fibrous root system produced below. The roots have a cream-colored flesh with a characteristic celery flavor. The foliage resembles that of celery but is coarse and not edible.

Handling and Storage

Cultivation of the crop is very similar to that of celery, except that plants are not earthed up during the growing season. As with celery, it cannot tolerate drought conditions, although it can be grown successfully on a wide range of soil types providing that they are free-draining. It can tolerate heavier soils than celery. Celeriac is usually transplanted as the seeds are too small to direct-drill. The crop is usually sown in greenhouse conditions planted in the spring (typically early to mid April in the UK) and is ready to harvest by the middle of September onwards. Like celery, it is a heavy feeder and responds well to bulky organic manures. Roots are usually lifted early in October and stored in dry, cool conditions until required. Roots can remain in the field if soil is drawn up around them or, alternatively, the crop is covered with straw in mid-October to provide some winter protection. Roots can then be lifted as required.

Domestic and Industrial Uses

Celeriac is peeled and can be boiled and served as a hot vegetable or served cold in salads.

Food Value and Chemical Composition

Food value and chemical composition of raw celariac are summarized in Table 1.

Scorzonera – *Scorzonera hispanica*

Scorzonera (also known as black salsify or Spanish salsify) and salsify are usually classed together because their roots have the same root shape and flesh color. However, taxonomically they are unrelated. The skin color of the root of scorzonera is dark purplish, whilst the leaves are broader and shorter than those of salsify. However, crop treatment is similar,

Table 1 Nutritional value and chemical composition of salsify, celeriac, and Jerusalem artichoke. Values are based on 100 g raw flesh

	Salsify	*Celeriac*	*Jerusalem artichoke*[a]
Proximates			
Water	83.3 g	88.8 g	80.2 g
Sugar	1.5 g	1.8 g	1.3 g
Starch	3.5 g	0.5 g	Trace
Dietary fiber	N	5.1 g	N
Total nitrogen	0.22 g	0.20 g	0.25 g
Energy value (kcal)	27	18	41
Energy value (KJ)	113	73	207
Protein	1.3 g	1.2 g	1.6 g
Fat	0.3 g	0.4 g	0.1 g
Carbohydrate	10.2 g	2.3 g	10.6 g
Minerals			
Na	5 mg	91 mg	3 mg
K	310 mg	460 mg	420 mg
Ca	42 mg	40 mg	30 mg
Mg	20 mg	21 mg	11 mg
P	42 mg	63 mg	33 mg
Fe	0.9 mg	0.8 mg	0.4 mg
Cu	0.0.1 mg	0.04 mg	0.12 mg
Zn	0.2 mg	0.3 mg	0.1 mg
S	22 mg	15 mg	22 mg
Cl	37 mg	22 mg	58 mg
Vitamins			
Retinol	0 µg	0 µg	0 µg
Carotene	20 µg	2 µg	20 µg
Vitamin D	0 µg	0 µg	0 µg
Thiamin	0.06 mg	0.18 mg	0.10 mg
Riboflavin	0.01 µg	0.02 mg	Trace
Niacin	0.2 mg	0.5 mg	0.9 mg
Vitamin C	1 mg	14 mg	2 mg
Vitamin E	N	N	N
Vitamin B_6	0.07 mg	0.08 mg	N
Vitamin B_{12}	0 µg	0 µg	0 µg
Folate	57 µg	51 µg	N
Pantothenic acid	N	N	N
Biotin	N	N	N

[a]Boiled in unsalted water.

N, data not available.

Reproduced from Holland B, Unwin ID and Buss DH (1991) *Vegetables, Herbs and Spices: Fifth Supplement to McCance & Widdowson's The Composition of Foods*. London: HMSO, with permission.

maturing in approximately 120 days, and it is prepared for the table in a similar way. Although it is usually grown as an annual to provide a winter root vegetable, scorzonera can be left in the ground for a second season. This will result in a larger root without any apparent loss in quality. Scorzonera has similar storage requirements to salsify.

Domestic and Industrial Uses

The root can be eaten as a cooked vegetable and the leaves can be eaten in salads.

Food Value and Chemical Composition

These are not available.

Hamburg Parsley – *Petroselinum crispum*, Family Umbelliferae

Hamburg parsley (also known as turnip-rooted parsley) is a type of parsley in which the root, rather than the shoot, has been developed as the harvestable part. The white, parsnip-shaped roots which have a parsley/celery flavor can be eaten raw or boiled. The leaves can also be served like spinach. It is more often grown as a domestic crop than grown commercially. The crop needs a long growing season. Hamburg parsley is hardy and can withstand moderate freezing conditions in the open. Roots may be left in the ground or lifted and stored like carrot and parsnip.

Domestic and Industrial Uses

Hamburg parsley is typically peeled and eaten as a boiled vegetable.

Food Value and Chemical Composition

These are not available.

Jerusalem Artichoke – *Helianthus tuberosus* L., Family Asteracea

Jerusalem artichoke (also known as girasole, topinambur, and sunchoke) has edible tuberous roots which are fleshy and irregular in shape and possess a somewhat nutty flavor. The plant is vigorous, with the shoot growing to 3 m or more. It is a native of North America where it was grown by American Indians. It was introduced into France in the early seventeenth century and later appeared in England, Holland, and Italy. The name 'girasole' was given to this plant in Italy and 'Jerusalem' is a corruption of this word. The name artichoke was given to denote a similarity in taste and flavor of the tuber to the globe artichoke. It is now widely grown in temperate, subtropical, and tropical areas, although it has never been very popular and the area under cultivation is not extensive.

Handling and Storage

The crop is grown from sets which are small tubers or pieces of tuber normally weighing approximately 60 g with at least two to three eye buds. A typical spacing would be 30–60 cm apart in rows 90–129 cm apart. Jerusalem artichoke is frost-sensitive and requires at least 125 frost-free days to be grown successfully. The crop normally reaches maturity in 18–24 weeks, although early-maturing types can be harvested after 10–12 weeks. The crop can be harvested as soon as the leaves begin to die back. The tubers can be lifted by hand or by machine using a suitable elevator after first cutting back the tops. Tubers are perishable but can be stored at 0 °C and 90–95% relative humidity for 4–5 months. Yields average approximately 25–30 MT ha^{-1}.

Domestic and Industrial Uses

Tubers are boiled, baked, fried, or steamed and eaten in much the same way as potatoes, but can also be eaten raw in salads. Tubers are also grown for stock feeding, particularly to pigs. The green tops can also be used for stock feeding. Tubers can be utilized for flour, as a source of fructose, and for the preparation of 5-hydroxy-methyl-furfurol. They can also be used as a source of industrial alcohol and for a beer-like beverage.

Food Value and Chemical Composition

The principal storage carbohydrate present is inulin rather than starch. Food value and chemical composition of boiled Jerusalem artichokes are summarised in Table 1.

Chinese Artichoke – *Stachys tuberifa*, Family Labiatae

Chinese artichoke is also known as crosnes and kon loh. There are two other similar and closely related species within the same genus (*S. affinis* and *S. sieboldii*). The plant is a perennial grown for its small edible tubers. It has rough, nettle-like leaves and grows to a height of 30–45 cm. The tubers produced from its creeping root system are 5–8 cm long and 1.5–2 cm in diameter, constricted at the internodes into rows of bead-like segments. It is a native of the Far East and is cultivated in China and Japan.

Handling and Storage

Plants are propagated vegetatively from tubers. They can be overwintered in cool temperate areas providing frosts are not too severe. Discoloration of tubers following harvesting is reduced by storage in the dark at low temperature and high humidity.

Domestic and Industrial Uses

Tubers have a taste similar to those of the larger Jerusalem artichoke and may be roasted, boiled, or fried.

Food Value and Chemical Composition

The major storage carbohydrate in the tubers is tetrasaccharide stachyose. Other nutritional data for this crop are not available.

See also: **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability; **Vegetables of Temperate Climates**: Carrot, Parsnip, and Beetroot; Swede, Turnip, and Radish; Stem and Other Vegetables

Further Reading

Holland B, Unwin ID and Buss DH (1991) *Vegetables, Herbs and Spices: Fifth Supplement to McCance & Widdowson's The Composition of Foods*. London: HMSO.

Huxley AJ, Griffiths M and Levy M (eds) (1992) *The New Royal Society Dictionary of Gardening*, 4 vols. London: MacMillan.

Kay DE (1973) *Tropical Products Institute Crop and Product Digest No. 2: Root Crops*. Tropical Products Institute.

Masalkar DL and Salunkhe DK (1998) Other roots, tubers and rhizomes. In: Salunkhe DK and Kadam SS (eds) *Handbook of Vegetable Science and Technology: Production, Storage and Processing*. New York: Marcel Decker.

Oldham CH (1948) *Brassica Crops and Allied Cruciferous Crops*. London: Lockwood.

Oldham CH (1950) *Vegetable Growers' Guide*. London: Lockwood.

Shoemaker JS (1947) *Vegetable Growing*. New York: Wiley.

Tindall HD (1968) *Commercial Vegetable Growing*. Oxford: Oxford University Press.

Stem and Other Vegetables

R Fordham, Ludgvan, Penzance, Cornwall, UK
P Hadley, University of Reading, Whiteknights, Reading, UK

Introduction

In addition to the relatively small number of vegetable groups grown on a large scale in the temperate regions of the world, there are many lesser-known species whose production is labor-intensive and/or whose shelf-life is short, thus inhibiting large-scale commercial exploitation. Many of these crops are confined to home gardens. However, with increasing public taste for novelty and growers prepared to innovate in order to gain the potentially high market prices, such crops are taking on an increased economic significance, although global statistics are rarely available to quantify their status. This article discusses a number of such vegetables which are representative of a diverse range of plant structures: stems (asparagus and bamboo), leaf petioles (cardoon and seakale), flower buds (globe artichokes), and fruits (okra).

Globe Artichoke (*Cynara scolymus*), Family Compositae

The globe artichoke is a herbaceous perennial about 1.5 m tall. Its large erect leaves are gray-green and it bears large thistle-like flowers about 10 cm in diameter. The floral bracts have a sharp terminal spine, although thornless cultivars are being selected for commercial production. The globe artichoke is unknown in the wild but probably derives from the cardoon (*C. cardunclus)*. It is grown primarily for its immature flower buds, which are partly edible, but young leaves may also be eaten as chards. Globe artichokes were eaten fresh and preserved for year-round use by the Romans centuries before Christ. It was later rediscovered in Italy during the reign of the Medicis, from where it was introduced into France and, via Spain, to America. It is now grown widely in southern Europe and Italy is a major producer. See **Table 1** for an estimation of world production and leading producers.

Nutritional Value and Chemical Composition

The nutritional value and chemical composition of the edible portion of the globe artichoke are indicated in **Table 2**.

Handling and Storage

Propagation is either from seed or vegetatively from suckers of selected cultivars. The crop is a short-term

Table 1 Estimated world production of globe artichoke (Mt)

World	1 367 893
Italy	514 955
Spain	321 600
Egypt	87 923
Argentina	85 000
France	65 000
USA	49 580
Morocco	40 000
Algeria	39 000
China	30 000
Turkey	28 500
Chile	24 650
Greece	23 000

Source: FAO (2001) http://www.apps.fao.org.

Table 2 Nutritional value and chemical composition of globe artichoke, asparagus, bamboo, cardoon, and okra

Nutrient	*Units*	*Artichoke*	*Asparagus*	*Bamboo*	*Cardoon*	*Okra*
Proximates						
Water	g	84.940	92.400	91.000	94.000	89.580
Energy	kJ	197.000	96.000	113.000	84.000	138.000
Protein	g	3.270	2.280	2.600	0.700	2.000
Fat	g	0.150	0.200	0.300	0.100	0.100
Carbohydrate	g	10.510	4.540	5.200	4.890	7.630
Fiber	g	5.400	2.100	2.200	1.600	3.200
Minerals						
Calcium	mg	44.000	21.000	13.000	70.000	81.000
Iron	mg	1.280	0.870	0.500	0.700	0.800
Magnesium	mg	60.000	18.000	3.000	42.000	57.000
Phosphorus	mg	90.000	56.000	59.000	23.000	63.000
Potassium	mg	370.000	273.000	533.000	400.000	303.000
Sodium	mg	94.000	2.000	4.000	170.000	8.000
Vitamins						
Vitamin C	mg	11.700	13.200	4.000	2.000	21.100
Thiamin	mg	0.072	0.140	0.150	0.020	0.200
Riboflavin	mg	0.066	0.128	0.070	0.030	0.060
Niacin	mg	1.046	1.170	0.600	0.300	1.000
Pantothenic acid	mg	0.338	0.174	0.161	0.099	0.245
Vitamin B_6	mg	0.116	0.131	0.240	0.043	0.215
Folate	μg	68.000	128.000	7.100	28.300	87.800
Vitamin A	IU	185.000	583.000	20.000	120.000	660.000

Values per 100 g edible portion.
Source: USDA (2000) http://www.nal.usda.gov/fric/foodcomp.

perennial and it is usual to replant after approximately 3 years depending on soil and environmental conditions. The crop grows well in temperate climates but cannot survive severe frosts. The flower buds are harvested throughout the summer and early crops can be produced under protection in glasshouses or plastic tunnels.

Globe artichokes do not store well but can be kept for a few days in cold stores at 0 °C and 95% relative humidity.

Domestic and Industrial Use

Harvested flower buds are trimmed and boiled before serving entire hot or cold, usually as an hors d'oeuvre. The edible parts are the fleshy basal portions of the bud scales or bracts and the heart or basal portion of the flower. The 'choke,' consisting of immature florets, is inedible and should be discarded. The young shoots of the plant may also be forced and blanched in the dark and used as a substitute for celery in salads.

Asparagus (*Asparagus officinalis*), Family Liliaceae

Asparagus is a dioecious perennial reaching a height of about 1.5 m, grown for its edible young shoots. It originated in the eastern Mediterranean region and occurs wild in maritime regions throughout Europe.

Table 3 Estimated world production of asparagus (Mt)

World	4 204 338
China	3 508 500
Peru	180 600
USA	103 060
Mexico	55 000
Spain	51 300
Germany	50 794
Morocco	36 000
Greece	34 000
Japan	28 700
Italy	27 143
France	25 000
Chile	18 800
Australia	17 000
Netherlands	17 000
Bulgaria	12 000

Source: FAO (2001) http://www.apps.fao.org.

It has been used for over 2000 years, first as a medicine and later as a vegetable. It is now widely grown throughout the world and, although more suited to the cooler temperate regions, it is also a valued export crop for a number of warmer countries. The main producing countries are listed in Table 3, along with estimates of recent production figures.

Nutritional Value and Chemical Composition

Data on the nutritional value and chemical composition of the edible portion of asparagus shoots are included in Table 2.

Handling and Storage

Plants are mainly propagated from seed but direct seeding in the field is not recommended as germination is slow and emergence poor. For this reason plants are usually established from 1-year-old crowns raised by specialist propagators. As asparagus is dioecious, seedlings will normally give rise to a mixture of male and female plants, of which male plants are more productive. For this reason plant breeders have selected for all male cultivars, although the alternative approach of propagating selected male plants by tissue culture is also practiced commercially. By maintaining separate male and female lines through tissue culture it is also possible to produce F1 hybrid seed and a number of improved cultivars have been introduced in this way.

Field production is usually carried out from 1-year-old crowns established in the spring which produce a canopy of fern-like leaves throughout the growing season. This replenishes and builds up the store of carbohydrates in the root system which are necessary for regrowth of new shoots the following year. In temperate regions the leaves die back and the plants become dormant prior to regrowth of shoots the following spring. The earliest shoots are harvested for consumption while later-produced ones are allowed to develop to reinstate a photosynthetic canopy to provide carbohydrates which are stored in the roots for the following year's crop. During the first year of production harvesting is light in order for the plants to build up in size. Once a plantation is established it can remain productive for 20–30 years, although commercially it is usual to replant more frequently with improved stock. The management of the crop requires a careful balance to be made between harvesting and retaining sufficient fern growth in the growing season. In warmer regions of the world, fern growth may be continuous and plants fail to achieve dormancy. It is then necessary to regulate cropping by defoliating a portion of an asparagus plantation which then initiates a crop of new harvestable shoots, a number of which must be left to reestablish new leaf growth. In this way it is possible to harvest shoots throughout the year, in contrast to temperate countries where most of the crop is produced in the spring and early summer.

Crowns may be established either in ridges or flat beds: the choice depends on a number of factors, including soil drainage and the method of harvesting, Shoots may be harvested either green or white. Shoots are cut below ground level when about 25 cm long either when only the tips are showing and the stem is blanched or alternatively when a portion of the shoot has developed above the soil surface and become green. White asparagus production predominates in Europe, South Africa, and Taiwan, whereas green asparagus is more common in North America and Australia. It is possible to advance cropping in cooler regions by using plastic film covers which are removed prior to harvesting.

Due to its high respiration rate, fresh asparagus has a short storage life but this can be extended to around 3 weeks in cool stores set at approximately 2 °C and high humidity, although there is danger of changes in flavor and texture.

Domestic and Industrial Use

Young shoots are cooked by boiling and either served entire or made into soup. The crop may also be processed by freezing, canning, or dehydrating.

Bamboo Shoot – A Large Number of Genera within the Family Graminae

The bamboos constitute a large number of species (as many as 1000 within 90 genera) of the family Graminae or grasses. These multipurpose perennials include many of the world's fastest-growing woody plants. The mature shoots are used as timber for construction purposes but the young blanched shoots of many species are also harvested as vegetables. The most important edible species are mainly natives of China and include mousou-chiku (*Phyllostachys edulis*), ku-chiku or mandake (*P. bambusoides*), and *P. dulcis*. Other important genera include *Dendrocalamus, Bambusa*, and *Sasa*. Production statistics are not readily available but the major producers include China, Japan, Taiwan, and Thailand. New plantations of *D. asper* and *D. latiflorus* have recently been introduced in Australia.

Handling and Storage

Bamboo cultivation is in many ways similar to that of asparagus in that the mature shoots provide photosynthates for the developing shoots which are harvested immature for consumption. The new shoots are blanched by a covering of compost to exclude the light and mature shoots are harvested at intervals, providing a timber crop and stimulating further shoot growth. Exposure of the young emerging shoots to light can result in the photoactivated synthesis of bitter and potentially toxic cyanogenic glycosides. Plantings, which may be dedicated bamboo plantations or intercropped with other crops, are established vegetatively by division of the older crowns. Yields may be as high as $10\,000\,\mathrm{kg\,ha^{-1}\,year^{-1}}$ of edible shoots.

Nutritional Value and Chemical Composition

Bamboo shoots form an important component of Chinese cooking in which they contribute to the mineral and vitamin components of the diet. Estimates of nutritional value and chemical composition for *Phyllostachys* spp. are provided in Table 2.

Domestic and Industrial Use

Fresh bamboo shoots are cooked by boiling for 30 min to tenderize and remove bitter flavors caused by cyanogenic glycosides. Precooked shoots are also conserved by canning and drying.

Cardoon (*Cynara cardunculus*), Family Compositae

Cardoon is a perennial crop, closely related to the globe artichoke, grown for its fleshy petioles which are sometimes blanched to render them more succulent. It is a large plant, reaching a height of around 2 m, and is thistle-like in appearance. It was introduced into the UK during the nineteenth century but has never been grown on a large scale, although it is more popular on the European continent.

Nutritional Value and Chemical Composition

Estimates of nutritional value and chemical composition of cardoon are included in Table 2. Shelf-life is short but may be extended under cold-store conditions of around 0 °C and high relative humidity.

Domestic and Industrial Use

The crop is harvested in the autumn when the petioles and thinly sliced hearts can be eaten raw in salads or boiled in salt water until tender. Shoots may be blanched prior to harvest by excluding light from the lower portion of the stem with, for example, black plastic.

Okra (*Abelmoschus esculentus*), Family Malvaceae

Okra was formerly classified as *Hibiscus esculentus*. Its center of origin was most likely West Africa but transfer to parts of India and South-east Asia occurred early on in the crop's development. It is a traditional vegetable of many tropical and subtropical countries but following its increasing popularity its production has extended to more temperate areas such as the southern USA. It is commonly known as lady's finger. Estimates of world production and leading producers are shown in Table 4.

Table 4 Estimated world production of okra (Mt)

World	3 912 384
India	2 500 000
Nigeria	719 000
Ghana	150 000
Pakistan	109 000
Iraq	85 000
Egypt	76 000
Saudi Arabia	58 500
Benin	39 928
Mexico	36 000
Burkina Faso	26 000
Turkey	26 000
Yemen	19 100

Source: FAO (2001) http://www.apps.fao.org.

The species is very variable; commercial cultivars are erect annuals which become woody at maturity. The plant is primarily grown for it long, finger like fruits which are harvested immature. The mature seeds are relatively large and heavy (60 g per 100 seed) and show potential in various food supplements.

Nutritional Value and Chemical Composition

Tender immature fruits provide dietary fiber, protein, and vitamin C in human nutrition. As fruits mature the increase in fiber content renders them unpalatable. The species is characterized by a high mucilage content made up of acidic polysaccharides and having viscous colloidal dispersion properties in water, valued in soup and stew preparation. Okra seeds contain 12–17% oil, mainly monounsaturated fatty acids, and have potential in cereal-based diets due to their high lysine level. They also show promise for use as meat analogs as their protein concentrates and seed flour are more soluble than commercial soya products. Nutritional value and chemical composition are summarized in Table 2.

Handling and Storage

Okra requires a long growing period of relatively high temperatures. It is sensitive to frosts and commercial production is confined to summer months in subtropical and warmer parts of the temperate zone. Most cultivars require short day lengths for floral initiation and day-neutral selections are therefore required for cultivation in temperate countries. Commercial selections are also available based on visual characters

such as absence of spines and fruit color. The crop is established from seed sown *in situ* as soon as spring temperatures are favorable. Fruit production is not synchronized and selective harvesting must be carried out manually throughout the summer. Exudates from the fruit surface and other plant parts can cause skin inflammation.

Fresh fruits have a short storage life and suffer chilling damage at 6 °C and below. Controlled-atmosphere storage at 5% oxygen and 10% carbon dioxide at 11 °C and 90–93% relative humidity extend shelf-life to 2 weeks.

Domestic and Industrial Use

The fresh immature fruits may be boiled or fried or used as a thickening agent in soups. Processing may be by canning and freezing and cultivars are selected for these outlets. Sun drying of fruit is conducted in developing countries for off-season use.

Seakale (*Crambe maritime*), Family Cruciferae

Seakale is a perennial cultivated for its edible leaves and young shoots. It has a delicate flavor when forced and can be used as a substitute for asparagus. The species is adapted to cool maritime climates and has been collected from the wild since early times. Cultivation has been recorded since the early nineteenth century, although on a relatively small scale. Traditionally, seakale roots were lifted in the autumn and forced in cold frames and covered with horse manure topped with straw. Modern techniques involve raising crowns under field conditions prior to forcing in the dark under controlled environments of 15 °C and high humidity. The young succulent shoots are harvested about 4 weeks later when they will have reached a length of approximately 20 cm. They may be eaten raw in salads or lightly boiled, as for asparagus.

Handling and Storage

The crop is propagated by seed or vegetatively from root cuttings, the latter resulting in earlier and more uniform establishment. Once plants have built up sufficient carbohydrate storage within its root system, they are ready for forcing. This is achieved by covering the dormant crowns in the autumn to produce a crop of young blanched shoots the following spring. Following harvest, plants may remain *in situ* and the process is repeated for the following year. Earlier harvesting can be achieved by lifting crowns and forcing under protection at warmer temperatures. The young freshly harvested shoots do not store well.

Domestic and Industrial Use

Seakale has a unique nutty flavor and crisp texture. It may be used raw in salads or the young shoots may be boiled and used in a similar manner to asparagus.

See also: **Storage Stability**: Mechanisms of Degradation; Parameters Affecting Storage Stability; **Vegetables of Temperate Climates**: Carrot, Parsnip, and Beetroot; Swede, Turnip, and Radish; Miscellaneous Root Crops

Further Reading

Düzyaman E (1997) Okra: botany and horticulture. In: Janick J (ed.) *Horticulture Reviews*, vol. 21, pp. 41–72. New York: John Wiley.

Hardy FS and Watson GD (1982) *The Complete Guide to Commercial Vegetable Growing*. London: Frederick Muller.

Huxley AJ, Griffiths M and Levy M (eds) (1992) *The New Royal Society Dictionary of Gardening*, vols. 4 London: MacMillan.

Nonnecke IL (1989) *Vegetable Production*. New York: AVI.

Rubatzky VE and Yamaguchi MAS (1997) *World Vegetables – Principles and Nutritive Values*, 2nd edn. Connecticut: AVI.

Salunkhe DK and Kadam SS (eds) (1998) *Handbook of Vegetable Science and Technology: Production, Composition, Storage and Processing*. New York: Marcel Dekker.

Splittstoesser WE (1990) *Vegetable Growing Handbook – Organic and Traditional Methods*. New York: Van Nostrand Reinhold.

Thompson HC and Kelly WC (1957) *Vegetable Crops*. New York: McGraw-Hill.

Weichmann J (ed.) (1987) *Postharvest Physiology of Vegetables*. New York: Marcel Dekker.

VEGETABLES OF TROPICAL CLIMATES

Contents

Commercial and Dietary Importance

S K Roy and A K Chakrabarti, Indian Agricultural Research Institute, New Delhi, India

Background

Tropical vegetables are many and varied. They contain a large number of leafy and legume vegetables, some of which are not only rich in proteins but also rich in several vitamins, minerals, and fiber. Tropical root vegetables are rich sources of carbohydrate, and they serve as a staple food in many tropical countries. They have a long shelf-life and have a good export potentiality. Some tropical fruits, e.g., banana, jack fruit, bread fruit, papaya, are consumed as vegetables at the tender stage. These fruit vegetables are rich in carbohydrate, phenolic compounds, and dietary fiber as well as minerals and vitamins. Some of the tropical vegetables are known to have curative properties for certain ailments and are popular in traditional medicine in several countries.

General Nutrient Composition

Tropical vegetables include many roots and tubers like sweet potato, cassava, taro, and yam, which are rich in starch and serve as staple foods in many countries. Most leafy vegetables are rich in carotene (provitamin A), riboflavin (vitamin B_2), and minerals. It has been estimated that 100 g of tropical leafy vegetables can provide 60–140 mg of ascorbic acid (vitamin C), 100 μg of folic acid, 4–7 mg of iron and 200–400 mg of calcium. Consumption of 100 g of leafy vegetables per day can supply 15% or more of the total protein intake. Only 30 g of the leaves will be sufficient to meet the requirements of vitamin A and C. The B vitamins (thiamin, riboflavin, and niacin) will increase by 15–30%, and the entire requirement of folic acid will be met. (*See* **Carotenoids**: Physiology.)

The nutrient composition of the major tropical vegetables is listed in Tables 1 and 2 and is discussed in detail in the section Importance of Specific Tropical Vegetables. The essential amino acid profile of selected, protein rich tropical leafy vegetables is given in Table 3. Refer to individual nutrients.

Dietary and Commercial Importance

The general importance of vegetables in commerce and diet, as discussed for temperate vegetables, holds for the tropical vegetables also. Their commercial value and importance in different diets are discussed in the section Importance of Specific Tropical Vegetables. (*See* **Vegetables of Temperate Climates**: Commercial and Dietary Importance.)

International Trade

Several tropical vegetables, such as sweet potato, taro, yam, dried chilli, squash, and pumpkin, have a good keeping quality. These vegetables are cultivated in the tropical countries of Asia, Africa, and South America, and are new to many people of Europe and North America. These vegetables have added attractions in those countries, and their export is growing steadily.

Some vegetables like the bitter gourd have a therapeutic value and have been in use in traditional medicine in India and China for centuries. Scientific evidence that bitter gourd contains some phytoceutical chemicals that reduce blood sugar rapidly has stimulated interest in Western countries and thus increased its export potential.

Some vegetables like tomato, capsicum, cucumber, and French bean are in great demand in Western countries during the winter. The tropical and subtropical countries that can grow these vegetables in the open during winter have been able to meet the demand to a great extent by exporting these commodities. Those vegetables that are grown in the open fetch a higher price than those grown under protective cover because of their better quality attributes. With the improvement in handling and storage of vegetables by the introduction of modern

Table 1 Nutrient composition of tropical vegetables per 100-g edible portion

Crop	Macronutrients					Vitamins				Minerals		
	Energy (kcal)	Moisture (g)	Protein (g)	Fat (g)	Carbohydrate (g)	Thiamin (mg)	Riboflavin (mg)	Niacin (mg)	Ascorbic acid (mg)	Ca (mg)	P (mg)	Fe (mg)
Amaranth	45	85.7	4.0	0.5	6.1	0.03	0.30	1.2	99	397	83	25.5
Basella	32	90.8	2.8	0.4	4.2	0.03	0.16	0.5	87	200	35	10.0
Bitter gourd	25	92.4	1.6	0.2	4.2	0.07	0.09	0.5	88	20	70	1.8
Bottle gourd	12	96.1	0.2	0.1	2.5	0.03	0.01	0.2	0	20	10	0.7
Bottle gourd leaves	39	87.9	2.3	0.7	6.1							
Bread fruit	71	79.5	1.5	0.2	15.8	0.04	0.07		21	40	30	0.5
Brinjal	24	92.7	1.4	0.3	4.0	0.04	0.11	0.9	12	18	47	0.9
Cassava	157	59.4	0.7	0.2	38.1	0.05	0.10	0.3	25	50	40	0.9
Capsicum	22	93.4	1.2	0.2	4.0	0.06	0.06	0.5	128	9	22	0.7
Chinese cabbage	14	95.0	1.2	0.1	3.0	0.05	0.04	0.6	25	43	40	0.6
Cho cho	27	92.5	0.7	0.1	5.7	0	0.04	0.4	4	140	30	0.6
Cluster bean	16	81.0	3.2	1.4	10.8	0.09	0.03	0.6	49	130	57	4.5
Cowpea	48	85.3	3.5	0.2	8.1	0.07	0.09	0.9	14	72	59	2.5
Cowpea leaves	38	89.0	3.4	0.7	4.1	0.05	0.18	0.6	4	290	58	20.1
Cucumber	13	96.3	0.4	0.1	2.5	0.03	0	0.2	7	10	25	1.5
Drumstick	26	86.9	2.5	0.1	3.7	0.50	0.70	0.2	120	30	110	5.3
Drumstick leaves	92	75.9	6.7	1.7	12.5	0.06	0.05	0.8	220	440	70	7.0
Elephant-foot yam	79	78.7	1.2	0.1	18.4	0.06	0.07	0.7	0	50	34	0.6
Fenugreek leaves	49	86.1	4.4	0.9	6.0	0.04	0.31	0.8	52	395	51	16.5
French bean	32	90.1	1.9	0.2	7.1	0.08	0.11	0.5	19	56	44	0.8
Hyacinth bean	48	86.1	3.8	0.7	6.7	0.10	0.06	0.7	9	210	68	1.7
Indian squash	21	93.5	1.4	0.2	3.4	0.04	0.08	0.3	18	25	24	0.9
Ivy gourd	18	93.5	1.2	0.1	3.1	0.07	0.08	0.7	15	40	30	1.4
Jack fruit tender	51	84.0	2.6	0.3	9.4	0.05	0.04	0.2	14	30	40	1.7
Jute leaves	63	81.4	5.1	1.1	8.1					241	93	
Kakrol	52	84.1	3.1	1.0	7.7	0.05	0.18	0.6		33	42	4.6
Lima bean	123	67.5	8.4	0.5	22.1	0.24	0.12	1.4	29	52	142	2.8
Lotus root	53	85.9	1.7	0.1	11.3	0.10			22	21	74	0.4
Muskmelon	17	95.2	0.3	0.2	3.5	0.11	0.08	0.3	26	32	14	1.4
Mustard leaves	34	89.8	4.0	0.6	3.2	0.03	0.22	0.8	33	155	26	16.3
Okra	35	89.6	1.9	0.2	6.4	0.07	0.10	0.6	13	66	56	1.5
Papaya green	27	92.0	0.7	0.2	5.7	0.01	0.01	0.1	12	28	40	0.9
Plantain	64	83.2	1.4	0.2	14.0	0.05	0.02	0.3	24	10	29	0.6
Pointed gourd	20	92.0	2.0	0.3	2.2	0.05	0.06	0.5	29	30	40	1.7
Pumpkin	25	92.6	1.4	0.1	4.6	0.06	0.04	0.5	2	10	30	0.7
Pumpkin leaves	57	81.9	4.4	0.8	7.9				10	392	112	2.1
Purslane	27	90.5	2.4	0.6	2.9	0.10	0.22	0.7	29	111	45	14.8
Ridge gourd	17	95.2	0.5	0.1	3.4		0.01	0.2	5	18	26	0.5
Snake gourd	18	94.6	0.5	0.3	3.3	0.04	0.06	0.3	0	26	20	0.3
Sponge gourd	18	93.2	1.2	0.2	2.9	0.02	0.06	0.4	0	36	19	1.1
Summer squash	17	94.8	0.5	0.1	3.5	0.02	Trace	0.4	18	10	30	0.6
Sweet potato	114	70.6	1.7	0.4	26.3	0.10	0.06	0.6	21	32	47	0.7
Sweet potato leaves	63	80.7	4.2	0.8	9.7	0.07	0.24	1.7	27	360	60	10.0
Taro	97	73.1	3.0	0.1	21.1	0.09	0.03	0.4	0	40	140	1.7
Taro leaf stalks	24	93.0	0.5	0.2	6.0	0.02	0.04	0.4	13	49	25	0.9
Tomato	22	93.5	1.1	0.2	4.7	0.06	0.04	0.7	23	13	27	0.5
Watermelon	26	92.6	0.5	0.2	6.4	0.03	0.03	0.2	7	7	10	0.5
Water spinach	28	90.3	2.9	0.4	3.1	0.05	0.13	0.6	37	110	46	3.9
Wax gourd	10	96.5	0.4	0.1	1.9	0.06	0.01	0.4	1	30	20	0.8
Winged bean	25	92.0	2.1	0.3	4.0							
Yam	102	74.0	1.5	0.2	24.0	0.1	0.01	0.8	15	12	35	0.8

Source: FAO (1968) *Food Composition Table for Use in Africa*. Bethesda, MD: FAO and US Department of Education and Welfare; FAO (1972) *Food Composition Table for Use in East Asia*. Rome: FAO; Gopalan C, Sastri BVR, Balasubramanian SC *et al*. (1999) *Nutritive Value of Indian Foods*. Hyderabad: National Institute of Nutrition; Watt BK and Merrill AL (1963) *Composition of Foods – Raw, Processed, Prepared. Agricultural Handbook No. 8*. Washington, DC: USDA.

techniques like cold chain, refrigeration, and controlled-atmosphere storage, and with the development of rapid transport by land, sea, and air, trade in such highly perishable commodities is growing steadily. (*See* **Controlled-atmosphere Storage**: Effects on Fruit and Vegetables.)

The USA imports tropical vegetables from Mexico and other Latin American countries, the European countries import these vegetables from the Middle East and some countries of North Africa, and the Persian Gulf countries import vegetables mostly from the Indian subcontinent and other South Asian countries. Apart from fresh vegetables, processed products enjoy a good international trade. Taiwan is the major country exporting processed vegetables to several countries.

Information on international trade in some of the vegetables dealt with is meager, but from the point of view of their nutritive value, these vegetables deserve attention for future exploitation in the world market.

Table 2 Important trace elements of some tropical vegetables (mg per 100 g)

	Cu	*Mn*	*Zn*
Amaranth	0.08	0.36	0.18
Bitter gourd	0.10	0.08	0.46
Bottle gourd	0.03	0.06	0.22
Brinjal	0.12	0.13	0.22
Cluster bean	0.08	0.10	0.36
Cucumber	0.09	0.14	0.23
Drumstick	0.01	0.05	0.16
Drumstick leaves	0.07	0.37	0.16
Fenugreek leaves	0.01	0.23	0.36
French bean	0.06	0.12	0.42
Okra	0.11	0.15	0.42
Ridge gourd	0.08	0.07	0.38
Snake gourd	0.27	0.14	0.31
Sweet potato	0.02	0.22	0.11
Yam	0.12	0.12	0.45

Source: Gopalan C, Sastri BVR, Balasubramanian SC *et al*. (1999) *Nutritive Value of Indian Foods*. Hyderabad: National Institute of Nutrition.

Importance of Specific Tropical Vegetables

The importance of individual vegetables in nutrition, diet, commerce, and world trade is discussed under separate groups.

Beans

The major crops are the cluster bean (*Cyamopsis tetragonoloba*), cowpea (*Vigna unguiculata*), French bean or kidney bean (*Phaseolus vulgaris*), hyacinth bean (*Lablab purpureus*), lima bean (*Phaseolus lunatus*), and winged bean (*Psophocarpus tetragonolobus*). Beans (also called legumes) are second only to cereals as a source of human food. Nutritionally, they are rich in proteins, soluble carbohydrates, and vitamins. The essential amino acid composition of tropical beans is given in Table 4. Most beans are deficient in sulfur-containing amino acids, and a few are deficient in tryptophan. Sprouted seeds contain more vitamin, particularly vitamin C. Consumption of food legumes has been reported to decrease the concentration of cholesterol. (*See* **Beans**; **Legumes**: Legumes in the Diet.)

Legumes offer a variety of edible products. Mature seeds, known as pulses, serve as the major item of diet in the Indian subcontinent and China. Immature pods along with seeds of all the beans listed, except lima bean, are used as a cooked vegetable before the fibers lignify and harden. Lima bean seeds (green or dried) are boiled or baked and used as a side-dish. Only lima beans and French beans are canned or frozen. Canned beans in tomato sauce are an important item of commerce.

Cucurbits

Cucurbits constitute the largest group of tropical vegetables, consisting of the bitter gourd (*Momordica charantia*), bottle gourd (*Lagenaria siceraria*), cho cho or chayote (*Sechium edule*), cucumber (*Cucumis sativus*), Indian squash (*Praecitrullus fistulosus*), ivy

Table 3 Essential amino acid content (g) of vegetable leaves per 100 g of protein

Amino acids	*Amaranth*	*Colocasia*	*Drumstick*	*Fenugreek*	*Pumpkin*
Arginine	3.8	4.0	6.1	5.6	6.9
Histidine	2.1	2.1	2.2	1.8	1.6
Isoleucine	4.6	5.0	4.5	5.3	5.0
Leucine	5.9	5.8	7.4	6.2	10.1
Lycine	4.0	4.2	5.1	4.8	5.4
Methionine	1.1	1.0	1.8	1.4	1.8
Phenylalanine	2.9	2.9	4.6	4.8	5.4
Threonine	2.2	2.4	4.0	3.2	5.0
Tryptophan	1.1	0.6	1.6	1.3	1.3
Valine	4.5	4.5	5.6	5.1	5.8

Source: Gopalan C, Sastri BVR, Balasubramanian SC *et al*. (1999) *Nutritive Value of Indian Foods*. Hyderabad: National Institute of Nutrition.

Table 4 Essential amino acid content (g) of bean seed per 100 g of protein

Amino acids	*Cowpea* (Vigna unguiculata)	*Cluster bean* (Cyamopsis tetragonoloba)	*French bean* (Phaseolus vulgaris)	*Hyacinth bean* (Lablab purpureus)	*Lima bean* (Phaseolus lunatus)	*Winged bean* (Psophocarpus tetragonolobus)
Arginine	7.3	12.5	5.1	6.6	6.3	4.7
Histidine	3.2	2.5	2.6	3.2	3.2	3.2
Isoleucine	4.0	3.2	4.2	4.4	5.3	4.4
Leucine	7.2	5.9	8.1	8.5	8.9	7.0
Lysine	6.4	4.0	6.7	6.8	7.5	7.7
Methionine	1.2	1.4	0.9	0.9	1.2	1.3
Phenylalanine	6.0	3.7	5.3	4.9	6.4	5.0
Threonine	3.6	2.8	4.2	4.2	5.1	4.8
Tryptophan		1.9	1.5			3.5
Valine	4.7	4.2	5.1	5.2	5.8	6.1

Source: Duke JA (1981) *Handbook of Legumes of World Economic Importance*. New York: Plenum Press; Summerfield RJ and Roberts EM (eds) (1985) *Grain Legume Crops*. London: Collins.

gourd (*Coccinia indica*), kakrol (*Momordica dioica* and *M. cochinchinensis*), muskmelon or cantaloupe (*Cucumis melo*), pumpkin (*Cucurbita moschata*), summer squash (*Cucurbita pepo*), winter squash (*Cucurbita maxima*), ridge gourd (*Luffa acutangula*), sponge gourd (*Luffa cylindrica*), snake gourd (*Trichosanthes cucumerina*), pointed gourd (*Trichosanthes dioica*), watermelon (*Citrullus lanatus*), and wax or ash gourd (*Benincasa hispida*). (*See* **Melons, Squashes, and Gourds**.)

Cucurbits, in general, are not rich in calories, minerals, or vitamins. There are a few exceptions like the bitter gourd, which is rich in vitamin C, the squash, pointed gourd, and muskmelon, which contain large amounts of carotene (provitamin A), the kakrol, which is high in protein, and the pointed gourd, which is rich in calcium. Some of the species of *Cucumis, Luffa, Coccinia*, and *Momordica* exhibit varying degrees of bitterness in the fruits, leaves, and twigs, caused by terpenes called momordicins. Owing to its bitter principle, the bitter gourd is used in the Ayurvedic system of medicine in India. Scientists have isolated a hypoglycemic phytoceutical, cheratin, in the fruit, which causes a lowering of blood sugars. (*See* **Hypoglycemia (Hypoglycaemia)**.)

The cucumber is mostly used as a salad crop. Immature fruits, also called gherkins, are pickled in various ways. Muskmelon and watermelon are used fresh as dessert fruits, alone or in combination with other fruits in salads. Watermelon rind is pickled. In Russia, beer is produced from watermelon juice. Pumpkin and squash fruits are cooked with other vegetables or used in pie-making. Luffa, bottle gourd, ash gourd, bitter gourd, pointed gourd, and Indian squash are cooked, and some are fried for the preparation of various dishes. Mature ash gourd fruits are used in making sweet candy in India. Leaves, twigs, and flowers of squash, pumpkin, and ash gourd are consumed in the tropics.

Most cucurbits are used fresh. Some of the cucurbits, such as bitter gourd, pointed gourd, coccinia, squash, and pumpkin, are canned and some frozen. Squash, pumpkin, melons, and cucumber are shipped to distant markets. During the winter, cucumbers from tropical countries are exported to Europe and North America. A variety of muksmelon called Sarda, which is grown in Afghanistan, is exported to several countries.

Leafy vegetables

In the tropics, apart from cultivated leafy vegetables, a large number of leaves from different sources such as perennial trees, aquatic plants, and annuals other than vegetable crops are consumed. The most important of these are amaranth (*Amaranthus* sp.), Basella (*Basella alba*), bitter gourd (*Momordica charantia*), bottle gourd (*Lagenaria siceraria*), cassava (*Manihot esculenta*), Chinese cabbage (*Brassica chinensis*), cowpea (*Vigna unguiculata*), drumstick (*Moringa oleifera*), fenugreek (*Trigonella* sp.), jute (*Corchorus olitorius*), mustard and rape (*Brassica* sp.), pumpkin (*Cucurbita moschata*), purslane (*Portulaca oleracea*), sweet potato (*Ipomoea batatas*), taro (*Colocasia esculenta*), water spinach (*Ipomoea aquatica*), and winged bean (*Psophocarpus tetragonolobus*). Most leafy vegetables are rich in carotenes, fibre, and minerals. Several studies have demonstrated an inverse relationship between a diet rich in carotenoids and lung cancer. Dietary fibre may help prevent certain cancers, such as cancer of the colon and rectum. (*See* **Amaranth**; **Mustard and Condiment Products**.)

Some of the leafy vegetables contain toxic ingredients. Cassava leaves contain glucosides that may liberate hydrocyanic acid by the action of enzymes. It is highly volatile, and therefore, any preparation that involves heating will render the leaves harmless. Amaranth, *Portulaca, Celosia*, and *Basella* contain

appreciable quantities of oxalic acid and oxalates. Oxalic acid converts calcium from food consumed to insoluble calcium oxalate, rendering it unavailable for absorption. In amaranth, under certain conditions, the nitrate content of leaves exceeds tolerable limits. Both of these problems can be eliminated by boiling and discarding the cooking water. Calcium oxalate, present in the form of fine crystals in the leaves of taro, causes itching of the fingers and a pricking sensation on the tongue and throat. This can be removed by adding a little tamarind or citric acid to the cooking water and then discarding the water. (*See* **Cassava**: The Nature of the Tuber; **Plant Antinutritional Factors**: Detoxification.)

Leafy vegetables are mostly used fresh as they are highly perishable. They are fried or boiled singly or in combination. Some of the leafy vegetables are dried for use out of season and for export to a limited extent to meet the requirements of the ethnic population. Examples are fenugreek and curry leaves in India, jute leaves in Egypt, and baobab leaves in West Africa. Mustard leaves are considered an item of delicacy in the Indian subcontinent and are exported, after canning, for the ethnic population abroad. Some of the dried and powdered leafy vegetables, which are rich in carotene and minerals, can be blended with cereal flours to improve the nutritional status of people of the developing world who often suffer from nutritional deficiency diseases.

Okra

Okra (*Abelmoschus esculentus*) fruits are not a rich source of nutrients, but they do supply some minerals and vitamins. A considerable amount of okra is processed by drying, freezing, canning, and pickling. Okra pods have a mucilaginous substance that interferes with processing and remains in the canned product also. To overcome this, slices can be canned in tomato sauce.

A considerable quantity of okra is air-freighted from India to distant markets in the Persian Gulf countries.

Solanaceous Vegetables

Major crops are tomato (*Lycopersicon esculentum*), brinjal or egg plant (*Solanum melongena*), sweet pepper or *Capsicum* and hot pepper or chilli (*Capsicum annum*). (*See* **Potatoes and Related Crops**: Fruits of the Solanaceae; **Tomatoes**.)

In most countries, the tomato ranks second after the potato in production and consumption. It is a low-calorie vegetable and is a good source of vitamins A and C. The content of these vitamins is less in fruits ripened off the vine than in those ripened on the plants. The red color of tomato is due to a carotenoid pigment, lycopene. Egg plants contain substantial amounts of some vitamins and minerals. Several chemicals that help lower cholesterol have been detected in brinjal in a recent study. Brinjal is also rich in bioflavonoids, which can provide protection from stomach cancer. Peppers are rich in vitamins A and C, and the pungency of hot peppers is due to capsaicin.

Tomatoes serve as the major ingredient of salads. Tomatoes are commercially the most important vegetable used for processing to make sauce, ketchup, purée, etc. and are also preserved by canning. In the present fast-food era, the tomato is an indispensable ingredient of hamburgers, pizzas, and pastas. Sweet peppers are boiled, stuffed, and baked to produce various dishes. Hot peppers are dried and used as spices, mostly in the tropics. They are also pickled in oil. The mild, nonpungent varieties of chilli, commonly known as paprika, are valued chiefly for their brilliant red color and mild flavor and are used for coloring dishes. Chilli extracts are used in the preparation of ginger beer and are also used in the preparation of chilli sauces. The eggplant is an important vegetable of Asia, and the fruits are fried, roasted, baked, or cooked to prepare various dishes. Boiled and hollowed eggplants are stuffed with potato, fish, or meat and are also canned.

The tomato occupies a prominent position in world trade, mostly in the form of processed finished products or as an intermediate, tomato paste. Some quantities of fresh fruits are also air-freighted. According to an FAO estimate, the total world export of tomato and all types of pepper (*Capsicum*, chilli, paprika, etc.) during 1997 was 2.80 million tonnes and 0.21 million tonnes, respectively. The European countries rank first in tomato export, exporting 1.99 million tonnes with a market share of 71.07% and for pepper, Asian countries are the largest exporters, with a 61.90% market share, exporting 0.13 million tonnes. India is the largest exporter of dry chillies and oleoresin. There is considerable demand for paprika powder in the Western countries.

Starchy Roots, Tubers, and Corms

The major root crops are sweet potato (*Ipomoea batatas*) and cassava or manioc (*Manihot esculenta*), a tuberous crop is yam (*Dioscorea* sp.), and the corm crops are arum or taro (*Colocasia esculenta*), tannia (*Xanthosoma* sp.), and elephant-foot yam (*Amorphophallus campanulatus*). These crops are important sources of energy in many tropical countries. These vegetables are high in starch. The yellow-fleshed sweet potato is rich in carotene (provitamin A). Cassava and sweet potato contain considerable amounts of vitamin C. Cassava roots contain a toxic

hydrocyanic glucoside, which, on enzymatic action, liberates HCN; the harmful effect can be destroyed by peeling, washing, and boiling. Some species of *Dioscorea* contain a toxic alkaloid, dioscorine, which can be destroyed by roasting or boiling. (*See* **Alkaloids**: Toxicology.)

Sweet potatoes are baked, boiled, fried, or prepared mixed with other vegetables. They are used for the production of alcohol. Taro, tannia, cassava, and yam are fried, roasted, baked, or boiled for the preparation of various dishes. Taro is recommended for gastric patients, and its flour is a good baby food. In Hawaii, a fermented product prepared from taro is very popular. In Africa, a paste prepared from taro, after boiling, is consumed. The leaves and young shoots are used as greens. Cassava roots are processed to produce starch and tapioca granules. Yam is used for the preparation of chips, flakes, and flours. A sapogenin compound called diosgenin, present in some species of yam, is used for the production of cortisone and contraceptive drugs.

Cassava is an important staple food of Africa. Being perishable in nature, it is mostly consumed locally where it is grown. Some fresh cassava roots are air-freighted to the USA from Venezuela. Yams are also a staple food in some African countries and South-east Asia. Mostly, they are consumed locally where they are produced, but some quantities are shipped to distant markets. Taro and tannia can be shipped to distant markets because of their good keeping qualities. Taro is more popular in Asia and tannia in Africa.

Other Vegetables

There are several other vegetables, like plantain (*Musa paradisiaca*), bread fruit (*Artocarpus altilis*), jack fruit (*A. heterophyllus*), papaya (*Carica papaya*), drumstick (*Moringa oleifera*), and lotus (*Nelumbo nucifera*), which are popular in the tropics. The plantain is high in carbohydrate and provitamin A. The fruits, pseudostem, and inflorescence are cooked, as they contain a considerable amount of tannin, which renders them unpalatable even when ripe because of astringency. The fruits are boiled or cooked to prepare various dishes and are also used for making alcohol. The inner sheaths of the pseudostem of banana are cooked in India, and the immature, male flower buds, after removal of the boat-like fibrous bracts, are cooked in many South Asian countries. (*See* **Papayas**.)

Immature bread fruits and jack fruits are used as vegetables in many countries of Asia; the fruits, when tender, contain a high level of starch. After removing the skin and outer layer of the fruit, the flesh is sliced or cubed and then boiled, cooked, or fried. In India, seeds of mature jack fruit are cooked, and drum stick is a popular vegetable. Lotus rhizomes, commonly called roots, are cooked in many Asian countries.

Maize (also called sweetcorn) is popular in Western countries and is rich in starch. The tender cobs are consumed raw, boiled, or roasted and can be canned or frozen.

These vegetables have local demand and are mostly consumed where they are produced.

See also: **Alkaloids**: Toxicology; **Amaranth**; **Beans**; **Cassava**: The Nature of the Tuber; **Controlled-atmosphere Storage**: Effects on Fruit and Vegetables; **Hypoglycemia (Hypoglycaemia)**; **Legumes**: Legumes in the Diet; **Melons, Squashes, and Gourds**; **Mustard and Condiment Products**; **Papayas**; **Plant Antinutritional Factors**: Detoxification; **Potatoes and Related Crops**: Fruits of the Solanaceae; **Tomatoes**; **Vegetables of Temperate Climates**: Commercial and Dietary Importance

Further Reading

Bose TK, Som MG and Kabir J (eds) (1993) *Vegetable Crops*. Calcutta: Naya Prokash.

Choudhury B (1990) *Vegetables*, 8th edn. New Delhi: National Book Trust.

Decoteau DR (2000) *Vegetable Crops*. Englewood Cliffs, NJ: Prentice Hall.

Duke JA (1981) *Handbook of Legumes of World Economic Importance*. New York: Plenum Press.

FAO (1968) *Food Composition Table for Use in Africa*. Bethesda, MD: FAO and US Department of Education and Welfare.

FAO (1972) *Food Composition Table for Use in East Asia*. Rome: FAO.

Gopalan C, Sastri BVR, Balasubramanian SC *et al*. (1999) *Nutritive Value of Indian Foods*. Hyderabad: National Institute of Nutrition.

Jeana G (1991) *Pigments in Vegetables*. New York: Van Nostrand Reinhold.

Oomen HAPC and Grubben GJH (1977) *Tropical Leaf Vegetables in Human Nutrition. Communication 69*. Amsterdam: Department of Agricultural Research, Royal Tropical Institute.

Sharon M (1999) *Nutrients A to Z*. London: Prion Books.

Summerfield RJ and Roberts EH (eds) (1985) *Grain Legume Crops*. London: Collins.

Tindall HD (1983) *Vegetables in the Tropics*. London: Macmillan.

Vaclavik VA (1998) *Essentials of Food Science*. New York: Chapman & Hall.

Watt BK and Merrill AL (1963) *Composition of Foods – Raw, Processed, Prepared. Agricultural Handbook No. 8*. Washington, DC: USDA.

Yamaguchi M (1983) *World Vegetables*. New York: Van Nostrand Reinhold.

Root Crops of Uplands

S K O'Hair, University of Florida, IFAS, Tropical Research and Education Center, Homestead, FL, USA
D N Maynard, University of Florida, IFAS, Gulf Coast Research and Education Center, Bradenton, FL, USA

Introduction

Highland and midaltitude tropical root crops include several little-known crops that are regionally valued as stable foods in the South American Andes. Nearly all are ancient foods, domesticated possibly by pre-Incan people. All are either solely or mainly propagated through vegetative cuttings. It has been speculated that somatic mutations are a likely source of many of the varieties, since flowering and seed set are rare in many. Several are considered to be underexploited and deserving of considerably more scientific and commercial attention. Since these crops are mostly propagated vegetatively, appropriate methods to control the spread of diseases should be observed.

Ahipa

Ahipa (*Pachyrhizus ahipa* (Weddell) Parode, Family Fabaceae), also known as ajipa or Andean yam bean, is one of the few leguminous plants cultivated for its edible roots. When *Rhizobium* is present, the plant can fix nitrogen. Its crisp, white roots are very similar to those of its close relative, jicama (*P. erosus*).

Production is mainly limited to Peru and Bolivia at elevations of 1500–3000 m. Although there is little commercial importance for this crop, the potential exists for it to be grown as a vegetable in temperate regions for use as a fresh or salad vegetable. There are no recorded varieties of this crop. However, landraces are known to exist.

Each plant produces one enlarged root weighing 0.5–0.8 kg. A relatively thin, pale yellow skin covers the white flesh. Root production is significantly increased by manual removal of entire inflorescences.

Most plant parts are toxic, except for the enlarged edible storage root. Roots can be eaten raw or cooked. Specific nutritional data are not known. However, they are likely to be similar to its close relative, jicama.

Little is known about either handling or storage after harvest. Roots are normally harvested as needed since they lose moisture quickly after harvest.

There are no current industrial uses for the crop.

Arracacha

The arracacha or Peruvian carrot (*Arracacia xanthorrhiza* Bancroft, Family Apraceae) is grown for its edible roots, which have been described by the noted naturalist, David Fairchild, as one of the tastiest of all vegetables.

Of the crops covered in this article, arracacha production has spread to other food production areas the most. Current production areas include much of South America, Puerto Rico, Mexico, and parts of Africa. In the Andean region, production is found between 1800 and 2600 m. At higher latitudes, production is found at lower elevations. Commercial trade is limited to local markets. Plants are normally propagated vegetatively as clones of three general types: white, yellow, and purplish. Viable seeds are produced on older plants, resulting in possibilities for numerous varieties.

Storage organs arise as enlargements of a main cylindrical rootstock with enlarged edible lateral, parsnip-like roots near the base of the plant. The young blanched stems can also be eaten.

Peeled roots have nutritional qualities similar to other root and tuber crops (Table 1). Refer to individual nutrients.

Harvested roots have a short shelf-life of 1 week or less. Thus, roots are harvested as needed. However, as the roots age, they become fibrous and, eventually, unpalatable. Little research has been carried out regarding postharvest handling of the roots. Refrigerated storage at 3 °C extends the storage period. They can be frozen, like cassava, for shipment to foreign markets. (*See* **Cassava**: The Nature of the Tuber.)

Table 1 Average composition (per 100 g of raw product) of tropical root crops

Constituent	*Arracacha*	*Yacon*	*Canna*	*Oca*	*Mashua*	*Ulluco*
Moisture (g)	72	79	75	82	85	86
Carbohydrate (g)	25		24	16	11	11
Starch (g)		15	20			
Sugars (g)		12				
Protein (g)	1	1.5	1	1	1.6	1.8
Fat (g)	0.1		0.1	0.1	0.1	0.1
Fiber (g)	0.7	1	0.6	0.8	1.2	0.6
Ash (g)	1.6	1	1.4	0.5	0.5	0.7
Calcium (mg)	24					
Phosphorus (mg)	65					
Iron (mg)	0.7					
Vitamin A (IU)	40					
Ascorbic acid (mg)	21.5					

Data from Kay DE (1973) *Root Crops*. London: Tropical Products Institute; King SR (1987) Four endemic Andean tuber crops: promising food resources for agricultural diversification. *Mountain Research and Development* 7: 43–52; National Research Council (1989) *Lost Crops of the Incas*. Washington, DC: National Academy Press, with permission.

There is no commercial processing of arracacha. However, it is a popular source of carbohydrate among children. Starch can be extracted from both the main rootstock and lateral roots.

Edible Canna

Achira (*Canna edulis* Ker-Gawler Family Cannaceae), sagu, or Queensland arrowroot, is similar in appearance to its close relative, the ornamental canna. A related species grown in South-east Asia is *C. discolor* Lindl.

Production is found throughout the low- to mid-altitude tropics, with most concentrated in South America, especially in Peru, where it is a valued crop. Its commercial importance is limited to starch production in a few regions and production as a staple food in parts of Peru, Ecuador and South-east Asia. Starch extracted from rhizomes is used to produce noodles in Asia. Two main types are described, one with white corms and green foliage, and a second with violet-skinned corms. There are many variations in plant characteristics.

The storage organ forms at the lower part of the stem, at or below the soil surface, forming a corm. Occasionally, plants form a rhizome, which may be a factor of clonal variation or environmental conditions.

Nutritional composition is similar to most of the root and tuber crops (Table 1). The rhizomes are high in food value, containing 70% carbohydrate, 12% protein, and 22% fiber on a dry-weight basis. Edible canna is one of the few root and tuber crops that can be eaten raw. Thus, nutritional value is maximized if the corms are eaten uncooked. The starch granules are among the largest known and considered to be easily digested by infants and invalids.

Little is known about the handling and storage of this crop. Usually, the plant is harvested as needed. However, as the corms age beyond 8–10 months, the fiber content increases.

Mechanization exists for most stages of production, and the crop is widely adaptable. Thus, there is potential for greater exploitation of this crop as a source of speciality starch.

Maca

The turniplike maca (*Lepidium meyenni* Walp., Family Brassicaceae) is among the least-known of the Andean root and tuber crops. There are at least 29 cultivated varieties in the Andes.

Nearly all of the production is between 3500 and 4000 m in the Peruvian and Bolivian Andes. Because of their altitude constraints, the potential for commercialization is limited. Elite plants are selected and cultured with added care to procure botanical seed for the next crop. Yellow-fleshed types are preferred.

Storage organs arise as enlargements of the root region of the plant.

Roots have high concentrations of sugars and starch. Low quantities of alkaloids, fatty acids, tannins and saponins in the roots have been reported. (*See* **Alkaloids: Properties and Determination**; **Saponins**; **Tannins and Polyphenols**.)

Roots are aromatic and occasionally used as a flavoring agent. They are dried and stored for 1 or more years in protected, cool locations. However, the flavor deteriorates over time. In Peru maca is processed and marketed for its medicinal and therapeutic properties.

Mashua

Mashua or añu (*Tropaeolum tuberosum Ruiz* and Pav., Family Tropaeolaceae) is little known outside of its native production areas in the high Andes of South America. However, it resembles its close relative, the ornamental nasturtium.

Most of the production is limited to altitudes above 3000 m in Bolivia and Peru, with little commercialization. The area of production in Peru is estimated to be 4000 ha. Plants are vegetatively propagated as clones, with over 100 being recognized by their variation in tuber coloration.

Edible tubers are conical or ellipsoidal, 5–15 cm in length and 3–6 cm in width. Skin colors can be very attractive, ranging from white or yellow to red or purple. The flowers are also edible.

In comparison with other root and tuber crops, some mashua tubers have a relatively low moisture content of 55%. Most are comparable with the other root and tuber crops (Table 1), yet genetic variation in protein content and quality has been documented. Tubers provide an adequate balance of amino acids, with leucine being low (Table 2). Some clones are reported to have medicinal properties through the production of a mustard oil, which is used for treatment of kidney and liver diseases. Others reportedly have low degrees of antiaphrodisiac properties. Nutritional information for the flowers is not available. (*See* **Protein**: Quality.)

Little is known about handling and storage, since there is little commercial trade in this crop. Typically, one would expect the handling and storage requirements to be similar to those of potato. In the high Andes, tubers are frozen like chuño, a freeze-dried whole tuber that can be reconstituted in water.

Currently, there is little or no production of mashua for industrial purposes. Some cultivars have pesticidal properties.

Table 2 Average amino acid content in tropical root crops

Amino acid	*Amino acid content (mg per gram of protein)*		
	Oca	*Mashua*	*Ulluco*
Lysine	58	38	48
Threonine	46	23	26
Valine	37	36	35
Isoleucine	38	31	41
Leucine	56	39	49
Phenylalanine plus tyrosine	62	26	60
Tryptophan	7	5	9
Methionine plus cystine	30	14	30

Data from King SR (1987) Four endemic Andean tuber crops: promising food resources for agricultural diversification. *Mountain Research and Development* 7: 43–52, with permission.

Oca

Of the Andean root and tuber crops, oca (*Oxalis tuberosa* Molina, Family Oxalidaceae) is possibly the oldest and is second in importance in Andean regions only to potato. However, it is considered to have the potential of great consumer appeal, because of the brilliant color and pleasant flavor of the tubers.

Until recently, production of oca has been limited to altitudes of 2500–4000 m the Andes and parts of Mexico. In Peru, there are more than 16 000 ha in production. It has recently been commercialized in New Zealand. Reportedly, the potential exists for it to become a temperate region crop. In the Andes, there are approximately 50 cultivated clones. Andean germplasm collections have over 400 accessions.

Storage organs are formed on underground stolons, resulting in the production of elongated or rhizomic tubers 3–20 cm long.

Large amounts of variation in nutritional qualities are present among oca clones. In general, they are considered to have a food value comparable with potato (Table 1). The protein is of a high quality, being low in valine and tryptophan (Table 2). The carbohydrate component is largely sugar. Some tubers have levels of oxalic acid that are detectable when tubers are eaten raw, giving them a slightly acid taste. Oxalic acid concentrations can be as high as 500 p.p.m., but are generally only 1% of those found in spinach.

Little information on handling and storage is available. In the Andes, tubers are placed in the sun for a few days to increase the sugar content. However, owing to the high moisture content, handling should include means to control moisture loss. Alternatively, the tubers can be dried, producing a high-carbohydrate food. In the high Andes, tubers are 'freeze-dried' and stored for several years. (*See* **Drying**: Drying Using Natural Radiation; **Freeze-drying**: The Basic Process.)

Currently, there are no industrial uses for the crop. The starch is considered to be a quality product.

Ulluco

Ulluco or ullucu(s) (*Ullucus tuberosus* Caldas, Family Basellaceae) is one of the more popular members of the Andean root and tuber crops, being considered a delicacy by many.

Production is mostly limited to the Andean region. Unlike most of the other crops discussed in this article, there is commercial trade of this crop in larger South American cities. There are more than 15 000 ha under cultivation in Peru. Plants are propagated vegetatively as clones, with most being separated according to their flesh color, which almost covers the color spectrum. It is estimated that there are at least 50–70 distinct clones in the Andean region.

Tubers, resembling small potatoes, are produced on underground stolons. Most are elongated or slightly curved. The tuber skin is thin and soft, with inconspicuous buds. The flesh is low in fiber content.

The nutritional value of ulluco is similar to the other crops discussed in this article (Table 1). Undocumented reports suggest a much higher protein content. The amino acid balance is adequate; valine and tryptophan content are low (Table 2). Detectable levels of saponin are present in plant parts.

Tubers are usually harvested as needed. If stored, they must be kept in dark locations. In the light, tuber skins turn green, like potatoes stored in areas exposed to light. In the high Andes, tubers are 'freeze-dried.'

Tubers are dried and milled into flour. They can also be used to make chuño.

Yacon

Yacon or jiquima (*Polymnia sonchifolia* Poepp & Endl., Family Asteraceae), a native of North America, is a perennial herb being somewhat similar to Jerusalem artichoke.

Most of the yacon production is in the tropical and subtropical Andes of Peru, Bolivia, Colombia, Ecuador, and Argentina at altitudes of 900–2800 m. There is currently little commercialization of the crop. However, it has been proposed as a potential source of high-fructose syrup. The crop is propagated vegetatively as clones, which are separated according to shape and chemical composition.

Tuberous, fusiform roots 20 cm in length and 3 cm in diameter are produced at the base of the plant. The raw flesh is crisp. The main stem of the plant is used as a cooked vegetable.

Tuberous roots contain about 46 % inulin on a dry weight basis. They are also higher in protein than

most other tropical root and tuber crops (Table 1); and they are a good source of potassium. Dried leaves contain 11–17% protein and could be used for livestock feed.

Undamaged roots have a long shelf-life. Roots can also be dried for extended storage.

An alternative use for yacon is as a fodder crop. The inulin has commercial value as a sweetening agent for diabetics and in the production of fructose. (*See* **Fructose**.)

See also: **Alkaloids**: Properties and Determination; **Cassava**: The Nature of the Tuber; **Drying**: Drying Using Natural Radiation; **Freeze-drying**: The Basic Process; **Fructose**; **Protein**: Quality; **Saponins**; **Tannins and Polyphenols**

Further Reading

Hermann M and Heller J (1997) *Andean roots and tubers: ahipa, arracacha, maca and yacon. Promoting the conservation and use of underutilized and neglected crops*, 21 pp. Rome: International Plant Genetic Resources Institute.

Kay DE (1973) *Root Crops*. London: Tropical Products Institute.

King SR (1986) Ancient buried treasure of the Andes. *Garden* 10: 6–11.

King SR (1987) Four endemic Andean tuber crops: promising food resources for agricultural diversification. *Mountain Research and Development* 7: 43–52.

King SR and Bastien HHC (1991) *Oxalis tuberosa* (Oxalidaceae) in Mexico: an Andean tuber crop in Meso-America. *Advances in Economic Botany* 8: 77–91.

National Research Council (1989) *Lost Crops of the Incas*. Washington, DC: National Academy Press.

Sperling CR and King SR (1990) Andean tuber crops: worldwide potential. In: Janick J and Simon JE (eds) *Advances in New Crops*, pp. 428–435. Portland, OR: Timber Press.

Root Crops of Lowlands

D N Maynard, University of Florida, Bradenton, FL, USA
S K O'Hair, University of Florida, Homestead, FL, USA

Crops

Low-land root and tuber crops, which include sweet potato, yam, and, to a lesser extent, yam bean, provide a low-cost and readily available carbohydrate source in the diets of large populations in the tropics. Their value, both economic and nutritional, is based on production of starch-filled storage organs. Furthermore, their usefulness is extended somewhat through in-ground or common storage to provide high-caloric food during drought or other times of shortage.

Sweet Potato

The sweet potato (*Ipomoea batatas* (L.) Lam., Family Convolvulaceae) can be grown throughout the subtropics and tropics for use as a staple food source. Its growth habit is conducive to interplanting with other food crops, and vine tips are often used as a green vegetable in many cultures.

Global Distribution, Commercial Importance, and Varieties

Although widely distributed, sweet potato production historically has been concentrated in Asia; in 2001, production there accounted for about 90% of world production. Africa, the second most important area, contributed only about 7% of world production (**Table 1**). The area devoted to sweet potato production has decreased considerably in the last 20 years; however, yield increases of more than 100% in Asia have more than compensated for the smaller area. Asian yield increases can be attributed in part to improved varieties and production technology developed at the Asian Vegetable Research and Development Center, Tainan, Taiwan. In other parts of the world, sweet potato production area, yields, and total production have remained about the same or have declined.

There is little international trade in sweet potato because of its relatively low value as a world crop, the relatively short postharvest life of uncured roots, limited use in industrial products, and legal restraints associated with the introduction of plant diseases and insects with the roots. In the tropics, sweet potatoes are usually left in the ground until needed and sold in local markets within 1–2 weeks of harvest.

Distinct sweet potato clones that differ in growth habit and pigmentation of vines and roots have been selected by small growers in the tropics. The clones are named according to the source of planting stock or according to some characteristic that distinguishes it from other clones. Thus, a single clone may be known under several different names.

However, cultivars developed by breeding in the USA, the International Institute for Tropical Agriculture in Nigeria, and the Asian Vegetable Research and Development Center in Taiwan are known by specific names and/or number. The principal cultivars grown in the USA are Beauregard in the southeast and Garnet and Beauregard in California.

Table 1 World sweet potato production

Region	Area harvested ($\times 10^3$)			Yield (kg ha^{-1})			Production ($\times 10^3$ t)		
	1989–1991	1995–1996	2001	1989–1991	1995–1996	2001	1989–1991	1995–1996	2001
Africa	1 345	1 544	2 165	4 738	4 602	1 714	6 375	7 098	10 203
North and Central America	176	172	150	6 402	6 453	7 742	1 127	1 130	1 158
South America	127	109	101	10 124	12 000	12 320	1 288	1 307	1 250
Asia	7 347	7 197	6 860	15 608	17 999	17 879	114 672	129 562	122 645
Europe	6	5	6	13 091	12 437	10 498	77	60	60
Oceania	115	114	111	4 885	4 867	5 426	564	552	602
World	9 117	9 139	9 392	13 613	15 285	19 472	12 4104	139 708	135 919

Data from FAO (2002) www.FAO.org.

Morphology and Anatomy of the Root

An enlarged tuberous root, possessing typical root structures and the unique capability of initiating adventitious shoots, is the principal edible portion of the plant. Shortly after transplanting, some fibrous roots begin enlarging, probably in response to endogenous growth regulators. The length of the enlarged root is largely achieved in the first 8 weeks after planting, whereas increases in diameter continue as long as photosynthate is produced by the aerial portions of the plant. Storage parenchyma accounts for a major portion of the storage root, but cortex, secondary phloem, and secondary phloem parenchyma also account for a significant portion of the storage root.

Root shape varies from elongated to nearly globular; the skin color may be white, tan, yellow–orange, salmon, red, or copper, and the flesh color may be white, yellow–orange, orange, salmon, purple, or red. Normally, each plant produces five to 10 roots that usually vary in weight from 100 to 300 g each.

Nutritional Composition

In common with other root crops, sweet potato roots are rich in carbohydrate, which provides high energy value (Table 2). The yellow and orange-fleshed sweet potatoes are exceptional sources of pro Vitamin A, but white-fleshed types do not provide this advantage. Sweet potatoes also provide a good source of vitamin C. A 100 g serving may provide 25–40% of the recommended dietary allowance (RDA) depending on cultivar, root size, harvest date, and storage conditions. Sweet potato is not a superior source of amino acids, but generally exceeds yam and yam bean in amino acid composition (Table 3). The sweet potato protein and amino acid pattern is affected by cultivar, environment, cultural management, and growth duration. Nonetheless, it is generally high in lysine and provides a supplemental protein to cereals. Genotypic variability in provitamin A, ascorbic acid, protein content, and protein quality offer the possibility of enhanced nutritional quality through plant breeding. The potential for increased carbohydrate production from sweet potato can be best achieved by total yield increases derived from improved cultivars and production practices. In order to achieve acceptable levels of digestibility, it is necessary to cook sweet potato roots by boiling or roasting, because of the presence of a trypsin inhibitor in raw roots.

The nutritional composition of sweet potato leaves complements that of the roots (Table 2). Leaves are much lower in carbohydrates and food energy but are substantially higher in the essential elements calcium, iron, magnesium, phosphorus, and potassium. Leaves have a higher thiamin, riboflavin, niacin, and vitamin B_{12} content than roots but are inferior to roots in ascorbic acid and pro Vitamin A content.

Handling and Storage

Harvest of sweet potato may be on an as-needed basis by hand-digging of the roots to be eaten or sold on local markets; this is the pattern in much of the tropics. In the USA, and elsewhere where sweet potatoes are grown on an extensive commercial basis, they are harvested when the maximum number of roots is appropriate to be marketed; depending on cultivar and environmental conditions, this may vary from 90 to 130 days after transplanting. In this case, the entire crop is harvested mechanically on a once-over basis.

Sweet potato roots are subject to a number of postharvest rots that gain entry through wounds inflicted during harvest. Accordingly, it is critical to use harvesting methods that minimize wounds. Fortunately, harvest by hand, as in small-scale tropical agriculture, results in little root damage, so roots can be sold within the usual 1- or 2-week marketing period. With mechanical harvesting and the need for long-term storage to provide orderly marketing as in the USA, it is necessary to insure healing of wounded areas of the roots. This is accomplished by curing, i.e.,

Table 2 Composition (per 100 g of raw product) of tropical root crops

Constituent	*Sweet potato*		*Yam*	*Yam bean*
	Roots	*Leaves*	*Tuber*	*Root*
Water (g)	72.84	87.96	69.60	90.07
Food energy (kJ)	439	146	118	159
Protein (N × 6.25) (g)	1.65	4.00	1.53	0.72
Total lipid (fat) (g)	0.30	0.30	0.17	0.09
Total carbohydrate (g)	24.28	6.38	27.89	8.82
Fiber (g)	3.0	2.0	4.1	4.9
Ash (g)	0.95	1.36	0.82	0.30
Calcium (mg)	22	37	17	12
Iron (mg)	0.59	1.01	0.54	0.60
Magnesium (mg)	10	61	21	12
Phosphorus (mg)	28	94	55	18
Potassium (mg)	204	518	816	150
Sodium (mg)	13	9	9	4
Zinc (mg)	0.28	0.29	0.24	0.16
Copper (mg)	0.169	0.037	0.178	0.048
Manganese (mg)	0.355	0.256	0.397	0.060
Ascorbic acid (mg)	22.7	11.0	17.1	20.2
Thiamin (mg)	0.066	0.156	0.112	0.020
Riboflavin (mg)	0.147	0.345	0.032	0.029
Niacin (mg)	0.674	1.130	0.552	0.200
Pantothenic acid (mg)	0.591	0.225	0.314	0.135
Vitamin B_6 (mg)	0.257	0.190	0.293	0.042
Folacin (μg)	14?	80?	23.0	12?
Vitamin B_{12} (μg)	0	0	0	0
Vitamin A (IU)	20 063	1 028	0	21

Data from USDA (2002) http://www.nal.usda.gov/fnic/foodcomp/.

Table 3 Amino acid composition of tropical root crops

Amino acid	*Amino acid content (mg per 100 g)*		
	Sweet potato	*Yam*	*Yam bean*
Tryptophan	20	12	
Threonine	82	54	18
Isolucine	82	52	16
Leucine	121	96	25
Lysine	81	59	26
Methionine	41	21	7
Cysteine	13	19	6
Phenylalanine	99	71	17
Tyrosine	68	40	12
Valine	108	62	22
Arginine	77	127	37
Histidine	31	34	19
Alanine	90	63	20
Aspartic acid	282	155	200
Glutamic acid	161	181	43
Glycine	74	53	16
Proline	72	54	25
Serine	85	81	25

Data from USDA (2002) http://www.nal.usda.gov/fnic/foodcomp/.

holding the roots at 29 °C and 90–95% relative humidity (RH) in a well-ventilated storage for 4–7 days. Subsequent storage at 13–16 °C and 85–90% RH will maintain sweet potatoes in good condition for several months.

Industrial Uses

Although there is potential for numerous industrial products, only canning of whole, sliced, or puréed sweet potatoes for long-term preservation has achieved any measure of commercial importance. Dehydration of fresh or cooked chips by sun-drying is practiced in some tropical countries for preservation or later conversion to flour. Commercial starch production has been tried in several countries, including the USA, but has not proved to be economically feasible. The same has been true of glucose, pectin, and carotene production on a commercial basis. Recently, there has been considerable interest in utilizing sweet potato roots as a raw material for alcohol production. Where the environment permits continuous sweet potato production, yields surpass those of cassava, and ease of mechanization favors sweet potato production. However, it is likely that cassava or other crop products will be more suitable raw materials under seasonal production conditions. However, cull sweet potatoes have consistently provided a high-energy feed for livestock, especially hogs in the USA and elsewhere.

Yam

The yam (*Dioscorea* spp., family Dioscoreaceae) is a carbohydrate-rich food crop grown in some tropical areas, especially Africa, where its importance exceeds that of sweet potato. In the USA, the term yam is incorrectly applied to orange, moist-fleshed sweet potatoes in the market place. Accordingly, there is some confusion in popular terminology. True yams are not grown commercially in the continental USA, but are occasionally seen in ethnic markets.

Global Distribution, Commercial Importance, and Varieties

About 96% of world yam production occurs in Africa, principally West Africa (Table 4). Other production centers include South America, the Caribbean islands, and Oceania. Nigeria alone produces over two-thirds of the world yam crop. World production has increased in recent years, due primarily to increased area devoted to its production.

International trade in yams is limited to small markets in North America and Europe. For the most part, yams are used within the family or are sold locally. Where yams and other carbohydrate-rich root crops such as cassava and sweet potato are available in the same market, yams usually command a higher price. Furthermore, yams are favored over other root crops for special occasions and to serve to guests.

Five species of *Dioscorea* are grown commonly for food (Table 5) and an additional five of about 60 known species are occasionally used for food. *Dioscorea alata*, the greater yam, is widely distributed throughout the tropics, including South-east Asia, Africa, and the Caribbean islands. The yellow yam (*D. cayenenis*) is grown primarily in West Africa and the Caribbean. *D. esculenta*, the lesser yam, is distributed widely in the tropics, but production is mostly in South-east Asia, the south Pacific islands, and the Caribbean islands. The white yam (*D. rotundata*) is the most important yam in West Africa and is grown extensively in the Caribbean, and to some extent in East Africa. *Dioscorea trifida* is the only yam of New World origin, and its cultivation is mostly in the Caribbean. Distinct varieties of some yam species are available, but different names are used for the same yam in different parts of the world so that variatal designations are not particularly useful.

Morphology and Anatomy of the Edible Portion

The edible part of the yam plant is the tuber that varies in shape according to species (Table 5) and environment in which it was grown. Individual tubers

Table 4 World yam production

Region	Area harvested ($\times 10^3$)			Yield (kg ha^{-1})			Production ($\times 10^3$ t)		
	1989–1991	1995–1996	2001	1989–1991	1995–1996	2001	1989–1991	1995–1996	2001
Africa	2153	3014	3908	9312	9773	9499	20 073	31 113	37 124
North and Central America	57	60	55	6915	8068	7825	396	490	431
South America	41	44	56	7897	9435	9863	317	414	549
Asia	15	17	14	14 314	14 406	16 017	208	237	224
Europe				7953	9955	16 154	1	1	2
Oceania	18	18	18	15 735	15 702	16 181	283	286	283
World	2283	3243	4051	9307	10 320	9532	21 278	32 541	38 614

Data from FAO (2002) www.FAO.org.

Table 5 Descriptions of the most common *Dioscorea* species grown for food

Species	Common name	Origin	Tuber number and shape	Flesh color	Leaf shape and phyllotaxy	Stem appearance
D. alata L.	Water or greater yam	South-east Asia	Single, cylindrical	White to purple	Ovate, opposite	Winged
D. cayenensis Lam.	Yellow or yellow guinea yam	Western Africa	Single, variable	Yellow	Pointed, opposite; alternate	Spiny
D. esculenta Burk.	Lesser yam Potato yam	China	4–20 ovoid	White	Simple, alternate	Spiny
D. rotundata Poir.	White or white guinea yam	Western Africa	Single, cylindrical, spherical	White to yellow	Simple, opposite	Circular, smooth-spiny
D. trifida L.	Cush-cush	Tropical America	Numerous, elongate	White to purple	Lobed, alternate; opposite	Winged

Janick J (1990) Tropical root and tuber crops. *Horticultural Reviews* 12: 157–196. Reprinted by permission of Timber Press, Portland, OR, USA.

may range from a few grams to over 50 kg in weight, and tubers 2–3 m long have been produced under special cultural conditions.

Single or multiple tubers develop from a corm-like structure at the base of the vine. The tubers are covered by a thick layer of cork that often bears many cracks caused by tuber expansion. Roots may arise from the tuber, resulting in a somewhat hairy appearance before the dried roots slough off. The tuber has three general morphological sections – the head, middle, and tail. Buds are initiated from the head section after dormancy has been broken.

The bulk of the tuber is composed of starch-filled, thick-walled parenchyma cells in the central core. Vascular bundles having both xylem and phloem are scattered throughout the central core. A layer of meristematic cells that give rise to the buds surround the central core, and in turn is surrounded by a layer of thin-walled cortical cells. Finally, the entire tuber is surrounded by several layers of cork arranged in radial rows that arise from cork cambia.

Nutritional Composition

The principal nutritional component of yam is carbohydrate (Table 2). Like sweet potato, yam is a good source of ascorbic acid; however, it does not have the advantage of the high pro Vitamin A found in sweet potato. Yam is generally inferior to sweet potato in terms of amino acid content (Table 3). Exceptions are cystein, arginine, histidine, and glutamic acid, the levels of which are slightly higher in yam. However, other literature sources site yam as having a superior-quality protein compared with sweet potato, and being similar to potato in protein quality.

Handling and Storage

Dormant yams of some species may be stored in the ground for as long as 4 months. Harvested yams are stored in 'yam barns' in West Africa; the tubers are tied to vertical poles under a shelter to give the appearance of a barn with walls. Yams will remain in good condition during the dry period with only loss of weight from shrinkage and respiration, but loss from rots occurs quickly with the return of the rainy season. Postharvest applications of gibberellic acid prolong harvest life, and fungicide treatments retard rots, but these treatments are rarely practiced. Care during harvest to prevent cuts and bruises is required for long-term tuber storage. Yams can be cured like sweet potatoes to assist healing of wounds and prolong storage life, but this practice is uncommon. Yams can be stored under refrigeration, if available, at 16 °C and 70% RH, but are subject to chilling injury at temperatures lower than 12 °C.

Industrial Uses

Currently, there is little processing of yams, but the potential is great because the demand for yam as a food is high since fresh yams are not available until the next harvest season once tuber dormancy is broken and sprouting occurs. Dehydrated products such as chips or flour serve to fill the demand for yams to a minor extent, but consumer acceptance of these products is not universal.

Yam Bean

The yam bean (*Pachyrrhizus erosus* L., Family Fabaceae) is a minor tropical root crop known by many common names throughout the world. Jicama or some variation thereof is used most frequently.

Global Distribution, Commercial Importance, and Varieties

Yam bean is native to Mexico and northern Central America and is still cultivated there. In addition, yam bean is produced in South-east Asia, China, the Philippines, South America, the Caribbean islands, and East Africa.

For the most part, yam bean is used locally and does not enter into international trade. Production data are not available, but Mexico is likely to be the leading producer with an estimated 4000 ha under cultivation. Experimental yields of 80–90 t ha^{-1} have been reported from several countries. Traditionally, creole types of yam bean have been grown in Mexico, but they are being replaced by two new varieties – Agreadulce and Cristalina.

Morphology of the Edible Portion

The brownish roots are mostly turnip-shaped, but long and slender types are known. The preferred market size is 10–15 cm in diameter, which is attained in approximately 90 days from planting of seed. However, the root will continue to grow and reach a diameter of 30–45 cm and weight of 2–3 kg if not harvested at market maturity. The thick, tough skin is easily peeled to reveal the white flesh, which is crisp and succulent. Thinly sliced roots are eaten raw with a garnish or used in salads, and occasionally, it is cooked or pickled. In the USA, yam bean is often used as a substitute for water chestnut in oriental cuisine.

Nutritional Composition

Yam bean root is more succulent than that of sweet potato and yam and considerably less carbohydrate-rich (Table 2). Of the vitamins, only ascorbic acid is provided in significant amounts in yam bean. Except

for histidine and aspartic acid, yam bean is inferior to sweet potato and yam in amino acid content (Table 3). Typically, yam bean does not play a major role in the diet, in contrast to sweet potato and yam.

Handling and Storage

Yam bean roots are usually harvested by hand, washed, and stored in baskets or crates until sold. Refrigerated storage at 13–18 °C and 65–70% RH provides conditions for storage of roots for 1–2 months. Chilling injury occurs at temperatures less than 12 °C, and sprouting may occur at temperatures above 18 °C.

Industrial Uses

All yam bean roots are consumed in the fresh state. The mature seeds have 0.12–0.43% rotenone, an insecticide, but this has not been exploited for commercial use.

See also: **Amino Acids**: Properties and Occurrence; **Cassava**: Uses as a Raw Material; **Controlled-atmosphere Storage**: Effects on Fruit and Vegetables; **Dietary Requirements of Adults**; **Drying**: Drying Using Natural Radiation; **Fungicides**; **Protein**: Quality

Further Reading

Coursey DG (1967) *Yams*. London: Longmans.

FAO (1990) *Roots, Tubers, Plantains and Bananas in Human Nutrition*. FAO Food and Nutrition Ser. No. 24, Rome: Food and Agriculture Organization.

Horton D, Prain G and Gregory P (1989) High-level investment returns for global sweet potato research and development. *CIP (International Potato Center) Circular* 17(3): 1–11.

Martin FW (1972) *Yam Production Methods*. US Department of Agriculture Product Research Report 147. Washington, DC: US Government Printing Office.

O'Hair SK (1990) Tropical root and tuber crops. *Horticultural Reviews* 12: 157–196.

Onwueme IC (1978) *The Tropical Tuber Crops*. Chichester, UK: John Wiley.

Ravi V and Indira P (1999) Crop physiology of sweetpotato. *Horticultural Reviews* 23: 277–338.

Rubatzky VE and Yamaguchi M (1997) *World Vegetables. Principles, Production, and Nutritive Values*, 2nd edn. New York: Chapman & Hall.

Steinbauer CE and Kushman LJ (1971) *Sweetpotato Culture and Diseases. US Department of Agriculture Handbook 388*. Washington, DC: US Government Printing Office.

Tindall HD (1983) *Vegetables in the Tropics*. London: MacMillan.

Villareal RL and Griggs TD (eds) (1982) *Sweet Potato*. Tainan, Taiwan: Asian Vegetable Research and Development Center.

Edible Aroids

S K O'Hair, University of Florida, Homestead, FL, USA
D N Maynard, University of Florida, Bradenton, FL, USA

Background

There are plants from five genera within the Araceae family that are cultured for their starch-filled storage organs or herbaceous leaves. Most thrive in moist to flooded conditions and are among the most shade-tolerant of terrestrial food crops. Many are grown in moist tropical environments where other starchy food crops fail to flourish.

The edible aroids are known by many common names, often being confused with one another by the untrained eye. Part of the confusion arises from the low flowering frequency of most species. Some are cultured as foods of famine, and others are valued as staple foods. In the Caribbean, they are often the only food crops remaining after tropical storms.

All plant parts contain varying concentrations of an acrid factor, which appears to be carried on the surface of the needle-like calcium oxalate raphide crystals. It is believed to be either a cysteine proteinase or a glucoside that causes considerable irritation or swelling of affected tissues. Concentrations vary considerably among genera, species, and varieties, also being affected by plant age and cultural conditions. Cooking or other processing procedures destroy the acrid factors in the more popular clones. Naturalized or 'wild' clones need extensive processing to render them palatable. Levels of acridity and fiber content are factors that determine consumer acceptance of varieties. Production figures are not available for most of the crops, since they are grown as subsistence crops in small, often mixed plantings. Taro (*Colocasia*) and tannia (*Xanthosoma*) figures are combined with an estimated 5×10^6 t being produced world-wide.

As with all tropical root and tuber crops, yields and the size of individual edible portions vary considerably with planting density and length of time that the crop was allowed to remain in production. As the density diminishes, the size of individual edible parts increases. However, after a point, the marketable yield or crop value for the area planted declines. Thus, sizes of individual parts can be misleading if planting densities are not included. Early plant growth is characterized by the development of a root system and foliage production. Once total plant leaf area has reached a peak, well into the production period, the rate of corm and cormel production increases. This phase lasts two or more months.

Once the corm or cormel production phase is through, the plants of some species go dormant, and others must be harvested to prevent excessive amounts of sprouting and a decline in starch content and product quality.

Since all are propagated from vegetative cuttings, the development of certified insect and disease-free planting material is essential for optimal crop productivity. Dasheen mosaic virus (DMV) is commonly found throughout all genera, even in symptomless plants. Yield reductions due to DMV infection have been reported in both food and ornamental species.

With the exception of taro and tannia, postharvest and processing technology is very limited, since most are considered minor crops, and mechanization is not well developed. Like potato, the starches and carbohydrates have food and industrial uses.

Elephant (Foot) Yam

Plants in the genus *Amorphophallus* are unique among food crops as the plant produces only one petiole with a single umbrella-like compound leaf at the top. A flattened edible corm is produced at the base of the single compound leaf, enlarging from one year to the next if left unharvested. By the end of the first year, corms can weigh as much as 7 kg. Sections, up to 0.5 kg, of the corm are used as planting material for the next crop.

Amorphophallus is most commonly found in certain Pacific islands and much of Asia. In China and Japan, a popular species is known as konjac or konjac with production in China dating back 2000–3000 years. Its commercial value is greatest in Japan, where it is a valued food crop. Most *Amorphophallus* varieties are separated only by species names. Of 80–90 species, only a few are edible, with elephant foot yam (*Amorphophallus campanulatus* Bl. ex Decaisne) and konjac (konnyaku) (*A. rivieri* du Rieu (syn. *A. konjac* K. Koch)) being the most popular. Others are valued as novel ornamentals or as medicinal plants. Since flowering and seed set are commonn a few improved varieties and interspecific hybrids, many with reduced levels of acridity, exist.

The main carbohydrate, mannose (glucomannan), is a soluble dietary fiber that is not absorbed by humans. Otherwise, corms are similar in nutritional value with the other aroids (Table 1), with the exception of vitamin A levels, which are higher. The corms of some are considered to have medicinal value. In Japan, consumption is believed to reduce blood cholesterol and triglyceride levels in humans. Refer to individual nutrients. (*See* **Carbohydrates:** Classification and Properties; **Cholesterol:** Determining Blood Cholesterol Levels.)

At the end of the warm moist growing period, plants pass through a dormant period, at which time the corms are harvested as needed. Little information is available on postharvest storage. In general, corms are stored under ambient conditions.

The corms are processed to extract mannan, to make a wine, and as konjac curd for food. Glucomannan extracts have value as a fruit and vegetable preservative and as a sweetening agent.

Giant Taro

Alocasia or giant taro (*Alocasia macrorrhiza* (L.) Schott) is probably better known as an ornamental tropical plant than as a food crop. It is easily recognized by the spear-like appearance of the leaf laminae and petioles. Other species, including *A. indica* (Roxb.) Schott, *A. cucullata* Schott and *A. fornicata* Schott have been utilized as food crops.

Alocasia is among the oldest of food crops in Oceania, where, at one time, it was a staple food. Its spread as a food crop has been limited to Oceania and the Asian tropics. Its commercial importance is currently greater as an ornamental, than as a food crop.

Table 1 Average composition (per 100 g of raw product) of edible aroids

Constituent	*Elephant yam*	*Giant taro*	*Swamp taro*	*Tannia*	*Taro*
Moisture (g)	79	84	68	64	75
Carbohydrate (g)	10	15	29	32	21
Protein (g)	1.2	0.6	0.8	2.2	2.2
Fat (g)	0.2	0.3	0.2	0.2	0.4
Fiber (g)	0.8			1.0	0.8
Ash (g)	0.8	0.3		1.2	1.0
Calcium (mg)	43	30	578	16	34
Phosphorus (mg)	22	50	28	47	62
Iron (mg)	0.6	1	1.3	0.9	1.2
Vitamin A (IU)	270				
Ascorbic acid (mg)		5		8	8

Data from Leung WTW, Burtrum RRB and Chang FM (1972) *Food Composition Table for Use in East Asia*. Rome: Food and Agriculture Organization; Murai MF, Pen F and Miller CD (1958) Some Tropical South Pack Island Foods. Honolulu, HI: University of Hawaii Press, with permission.

There are only a few cultivated clones, most of which are found in the South Pacific. Nevertheless, one Indonesian germplasm collection reportedly contains 70 accessions. Flowering is common. However, there is little research in breeding for improved varieties.

The edible portion is the large semicompressed stem containing many starch-filled parenchymatous cells. During the first year of growth, the stems can reach 12 m in height with a diameter of 6 cm and a weight of approximately 18 kg. The plants can be left unharvested until needed.

There is little nutritional information for the stems. However, the carbohydrate content is among the highest of the edible aroids (Table 1). Owing to the nature of erect stems, the structural carbohydrate content is high. Of the edible aroids, alocasia is generally considered to be less palatable and more acrid.

There is no handling or storage information on this crop. It is harvested as needed for local consumption.

There is currently no industrial use for giant taro.

Tannia

Tannia (*Xanthosoma sagittifolium* (L.) Schott), also known as coco(yam), malanga, tanier, or yautia, originated in the wet valleys of northeastern South America. It was domesticated and carried throughout the Caribbean basin by migrating Indian tribes. There is much confusion within this genus, with the same plant being given more than one species name. In addition, tannia is often confused with taro. In parts of Africa, both are known as cocoyam, and in a few other places, both are listed as taro.

Tannia is a valued food crop in most of the Caribbean islands, Florida (USA), western Africa, parts of Egypt, India, and Oceania. In most regions, it is grown for sale in local markets. However, in Florida, the crop is valued at $\$8.6 \times 10^6$ annually. Costa Rica and the Dominican Republic are exporters of the crop. There are only a few varieties of tannia, and most have white flesh. Pink and yellow-fleshed types are less common and occasionally classified as *X. violaceum* Schott and *X. atrovirens* C. Koch, respectively. The same clone may have more than one common name. Common varieties in the English speaking Caribbean are South Dade White, Bruce and Rabess, which are white-fleshed. Vinola and Jamaique are purple- to pink-fleshed. Flowering is rare, and hand pollinations are required for seed production, thus accounting for the few varieties. Formal breeding programs in West Africa have focused on developing root rot-resistant genotypes.

The main stem of the white- and pink-fleshed types is an inedible corm, which is occasionally boiled for animal feed. The edible portions of these two types are club-shaped, underground offshoots of the main corm that develop into cormels weighing as much as 1 kg each. The main corm of the yellow-fleshed types is the part that is consumed. Young leaves are occasionally cooked as potherbs. In the western hemisphere, tropics plants of *X. brasiliense* (Desf.) are occasionally grown solely for their edible leaves.

Table 2 Average composition (per 100 g of raw product) of edible aroid leaves

Constituent	*Tannia*	*Taro*
Moisture (g)	90	90
Carbohydrate (g)	5.3	5.7
Protein (g)	2.5	2.4
Fat (g)	1.0	0.6
Fiber (g)	2.1	1.5
Ash (g)	1.3	1.3
Calcium (mg)	95	98
Phosphorus (mg)	388	49
Iron (mg)	2.0	2.0
Vitamin A (IU)	3300	180
Ascorbic acid (mg)	37	11

Data from Leung WTW, Burtrum RRB and Chang FH (1972) *Food Composition Table for Use in East Asia*. Rome: Food and Agriculture Organization, with permission.

The cormels are mostly starch, possessing the highest starch concentrations among the edible aroids (Table 1). The nutritional value of the leaves is considered to be comparable to spinach, complementing the high energy level of the cormels (Table 2). (*See*: **Starch:** Sources and Processing.)

Cormels of the favored clones have a natural dormancy period of 1–3 months and are planted year-round in moist production areas. They are planted and harvested as needed. Cormels can be removed individually or the entire plant can be uprooted. In commercial settings, cormels are washed and packaged in waxed boxes prior to shipment. The storage period of the cormels can be extended by several months by holding them at 7°C and 80% relative humidity (RH). Although not common, cormels can be dried and stored for extended periods. (*See* **Controlled-atmosphere Storage:** Effects on Fruit and Vegetables.)

Processing of tannia is limited to freezing of peeled cormels and bagging them in consumer packs in the form of a ready to cook frozen food. There are many potential uses similar to those described for taro.

Taro

Taro (*Colocasia esculenta* (L.) Schott), also known as eddoe or dasheen, originated in the Bay of Bengal region of South-east Asia. It was carried by early Polynesians throughout Oceania, where it became a staple food. It is an ancient crop in Asia, being

introduced into Japan more than 2500 years ago. Compared with tannia, it is better adapted to excessively moist areas. Many clones are grown like rice in flooded conditions.

Taro has lost some popularity in the traditional areas of production in the South Pacific and Asia owing to significant crop losses attributed to taro leaf blight disease. Nevertheless, it is still widely grown in this region, being the most important food crop. Promising levels of genetic resistance and other control methods for this malady should help to return this crop to traditional levels of production. Other areas where taro is grown include most islands in the Caribbean, Hawaii, parts of western and northern Africa, the Philippines, Japan, and South-east Asia. There are more varieties of taro than all the other edible aroids combined, with estimates exceeding 1000. Some clones flower and set seed under natural conditions, providing for genetic variability. Formal breeding programs in the Pacific basin and Asia have resulted in the development of improved genotypes. Bun long and Lehua maoli are popular varieties in Hawaii. Varieties are most commonly separated by coloration of the leaves and corms as well as by habit of cormel formation. One of two basic types, the dasheen, produces one large, edible corm up to 3–5 kg in weight, with few offshoots. The other type, the eddoe or Chinese taro, produces a much smaller main corm with several slightly smaller edible offshoot corms.

The corms are the primary edible portion of the plant. Young leaves and blanched leaves are also eaten as potherbs.

The nutritional composition of taro corms (Table 1) and leaves (Table 2) are very similar to those of tannia.

At harvest, approximately 2–3 cm of the petioles should be left attached to the corm in order to minimize problems associated with invasion by rot-causing organisms. Losses can be reduced by allowing time for wound periderm formation before placing corms in bulk storage. Since there is no distinct dormancy period, the shelf-life of taro is shorter than that of tannia. Temperatures as low as 5 °C can be used to extend the storage life.

In the USA, corms are sliced thinly and cooked in hot oil, resulting in a product similar to potato chips or crisps. Experimental procedures indicate that taro products can include flour, cereals, bread, noodles, cake, infant and invalid food, and beverage powder, as well as flaked, canned, or frozen corms.

Swamp Taro

Swamp taro (*Cyrtosperma chamissonis* (Schott) Merr) is pre-Polynesian and considered to be among the oldest of the South Pacific food crops.

Swamp taro flourishes in warm, humid climates. Production is mostly limited to the low islands of Micronesia, where it is primarily grown as a traditional food crop. Varieties are differentiated by their acrid factors, leaf shape, color, spininess of the petiole, and flowering habit. The time to maturity (measured in years) and tolerance to saltwater intrusion vary slightly.

The main corm is the edible portion of the plant, increasing in size from one year to the next; 10-year-old corms can weigh as much as 300 kg. Side-shoots are used as propagules.

The carbohydrate content of swamp taro is similar to that of taro and tannia (Table 1).

Plants are harvested as needed, except when storms raise the salinity levels of the coastal bogs. At such times, plants are harvested immediately and stored until planting conditions are favorable.

There are no industrial uses for swamp taro.

See also: **Controlled-atmosphere Storage**: Effects on Fruit and Vegetables

Further Reading

Coursey DG (1968) The edible aroids. *World Crops* 20: 25–30.

FAO (1990) *Roots, Tubers, Plantains and Bananas in Human Nutrition*. FAO Food and Nutrition Series No. 24. Rome: Food and Agriculture Organization.

Kay DE (1973) *Root Crops*. London: Tropical Products Institute.

Leung WTW, Burtrum RRB and Chang FH (1972) *Food Composition Table for Use in East Asia*. Rome: Food and Agriculture Organization.

Murai MF, Pen F and Miller CD (1958) *Some Tropical South Pack Island Foods*. Honolulu, HI: University of Hawaii Press.

O'Hair SK and Asokan MP (1986) Edible aroids: botany and horticulture. *Horticultural Reviews* 8: 43–99.

O'Hair SK, Snyder G and Morton JF (1982) Wetland taro: a neglected crop for food, feed, and fuel. *Proceedings of the Florida State Horticultural Society* 95: 367–374.

Suzuki M, Kano M, Mitani A, Mochida F and Ariki M (1975) On the irritant substance, 3,4-diglycosilicobenzaldehyde in taro *Colocasia antiquorum*. *Journal of the Japanese Society of Food and Nutrition* 28: 55–59.

Terasawa F, Tsuji K, Tsuji E *et al.* (1979) The effects of konjac flour on the blood lipids in the elderly subjects. *Japanese Journal of Nutrition* 37: 23–28.

Thompson S (1982) *Cyrtosperma chamissonis* (Araceae): ecology, distribution, and economic importance in the South Pacific. *Journal d'Agriculture Traditionelle et de Botanique Appliquee Travaux Detthnobotanique et Dethnozoologie* 29: 195–203.

Wang JK (1983) *Taro: a Review of* Colocasia esculenta *and its Potential*. Honolulu, HI: University of Hawaii Press.

VEGETARIAN DIETS

J Dwyer, Jean Mayer US Department of Agriculture, Human Nutrition Research Center on Aging at Tufts University, Boston, MA, USA

Introduction

Individuals mean different things when they refer to themselves as vegetarians. Those who provide dietary counseling must probe further to determine what foods are actually avoided and also the degree of adherence to the dietary pattern so that dietary recommendations will be appropriate.

Prevalence and Trends in Vegetarian Eating

The prevalence of vegetarian eating patterns varies, depending on the definition employed. Currently approximately 7% of Americans consider themselves to be vegetarian, but fewer than 1% are lactoovo vegetarians who abstain from all meat, fish, and poultry products, and even a smaller percentage – perhaps 0.1% – are vegans, who eat no animal foods at all. Many of those who define themselves as vegetarians are probably actually semivegetarians who limit their consumption of some or most animal products in favor of a plant-based diet, although they have not adopted a dietary pattern consisting exclusively of plants. The growing popularity of meatless meals and meat alternatives is due in large part to increased consumption of these products by individuals who are not necessarily vegetarian, but who want to eat a meatless diet.

Reasons for Adoption of Vegetarian Patterns

Individuals choose vegetarian diets for many reasons. In western countries, the major motivation for adopting vegetarian diets is for health reasons. Some believe that avoidance of animal products yields health benefits because it decreases the risks of bacterial or other foodborne diseases (e.g., bovine spongiform encephalopathy) that are particularly common in unhygienically prepared animal products, or that are associated with excessive intakes of saturated-fat cholesterol and sodium. Others focus on the putative health benefits of plant foods, including not only nutrients but also phytochemicals such as fiber, flavonoids, and other compounds that may have as yet unrecognized health benefits. Individuals who describe themselves as 'ethical vegetarians' often do so because of personal convictions such as opposition to the slaughter of animals for food or to what they view as the inhumane treatment of animals raised for food. Still others adhere to philosophies that encourage vegetarianism among their followers, (e.g., transcendental meditators, anthroposophists), or have religious beliefs that include vegetarian eating patterns (e.g., Seventh-Day Adventists, Buddhists, some Hindus, and other religious groups). Other people are motivated by environmental concerns, such as the conviction that raising food animals consumes natural resources that could otherwise be diverted to raising plant foods. Finally, there are some individuals who follow vegetarian diets simply out of habit. In developing countries vegetarian eating patterns may result from one or more of these motivations or simply because animal foods are expensive or unavailable to some segments of the population.

The degree of adherence to the chosen vegetarian eating pattern depends on motivation, habits, and circumstances. Until recently, the degree of animal food avoidance often predicted the extent of other food avoidance and divergence in other lifestyles from omnivores. Thus, vegans tended often to have belief systems that extended not only to extensive prohibitions against food ingredients made from or by animals (such as honey, casein, whey, rennet, and gelatin) but also to nonuse of animal products such as leather, wool, or silk and nonviolence toward animals. Today, however, the links between diet and belief are not so strong. Some vegans focus solely on the eating pattern for what they view as its health advantages but they do not necessarily subscribe to all the other aspects of the traditional vegan belief system and lifestyle. They are more willing to use meat analogs, nutrient-fortified foods, and vitamin mineral supplements, making it easier to attain nutritional adequacy.

Classification of Vegetarian Eating Patterns

The five major vegetarian dietary patterns are described in Table 1. However, myriad vegetarian dietary patterns exist and they are not easily described by focusing on the single dimension of animal food intake. Many vegetarians' food intakes differ from those of nonvegetarians in other ways as well. For example, some vegetarians eschew the use of vitamin–mineral supplements, and foods that they regard

Table 1 Common types of vegetarian dietary patterns

Pattern	*Comments*
Meat avoiders	Limit or avoid red meat and other flesh foods; may also restrict poultry, fish, and seafood
Lactoovo vegetarians	Avoid all meat, fish, poultry, and often fish but consume milk products and eggs
Lactovegetarians	Avoid all meat, fish, poultry, fish, and eggs
Macrobiotics	Currently popular diets are less restrictive than those popular in the 1970s but still involve numerous restrictions, generally including avoidance of all meat, poultry, milk, and eggs, but may consume fish in small amounts. Also avoid sugar and other refined sweeteners, members of the nightshade family (peppers, eggplant, tomatoes, potatoes) and tropical fruits
Vegans	Avoidance of all animal products including meat, fish, poultry, eggs, and dairy products
Other	Raw food eaters and 'living food' eaters avoid animal foods and eat raw plant foods, including fruits, vegetables, and cereals, with special health foods such as wheatgrass or carrot juice. Fruitarians consume diets mostly of fruits, nuts, honey, and olive oil. Rastafarians eat vegan-like diets but avoid alcohol, salt-preserved foods, and additives

as processed or nonorganic. Others include special foods such as organically grown foods, soy products such as miso, or sea vegetables that they regard as healthful.

A vegetarians eating patterns therefore encompass a continuum of animal and other food use and varies in the extent, intensity, and degree of alterations in both animal and other types of food consumption, and in its involvement with belief systems and other lifestyles. Some vegetarians limit only animal foods. Others may also proscribe processed foods, 'nonorganic' foods, or foods produced through biotechnology such as cheeses made with recombinant enzymes and genetically modified plant foods. At the same time, vegetarians' dietary diversity may be considerable owing to the wide variety of fruits, vegetables, nuts, legumes, meat analogs and fungi they consume.

Dietary intakes

Energy

Vegans often have lower energy intakes and lower weights than omnivores; differences are smaller and less consistent among those who consume other vegetarian patterns. Vegan and, to a lesser extent, vegetarian diets tend to be relatively low in caloric density, somewhat lower in fat, and occasionally lower in protein than those of omnivores. While lower energy intakes pose little problem for most adults, for vegan infants, weanlings, and small children, unless they are fed frequently, energy intakes may be low. Nuts, nut butters, oils, and fruit juices are concentrated sources of calories that are usually acceptable.

Fat and Saturated Fat

Most vegetarian diets eaten in western countries, and especially vegan diets, are lower in fat, saturated fat, and cholesterol, and higher in polyunsaturated and monounsaturated fats than are nonvegetarian patterns. Their low saturated fat and cholesterol and high dietary fibre and polyunsaturated fatty acid content decrease risk factors for coronary artery disease.

Carbohydrate

Intakes of carbohydrates, especially complex carbohydrates and dietary fiber, tend to be higher among vegetarians than among omnivores, and more in line with current dietary recommendations.

Protein

Vegetarians usually consume less total protein and less animal protein than omnivores but their intakes are usually satisfactory and sustain good nutritional status if energy intakes are adequate. Energy intake, protein quantity, quality, and digestibility must all be monitored on extremely restrictive vegetarian diets, however. Sufficient energy intake is important so that protein is available for its unique functions rather than being utilized for energy. Protein quality is also of potential concern, since some plant sources of protein are limited in one or more amino acids. For example, lysine is low in most cereal grains, sulfur-containing amino acids (methionine and cysteine) are lower in legumes and fruits than in other plant sources and animal foods, threonine is lower in cereals, and tryptophan is lower in fruits than in most animal foods. Thus, the biological value of vegan diets based on single cereal staples may be poor, especially at times of rapid growth.

However, the protein quality of appropriate mixtures of plant foods is entirely satisfactory and promotes good nutritional status. When dietary patterns include several plant protein sources (e.g., legumes, cereals, nuts and seeds, fruits, and other vegetables), protein quality is satisfactory since the deficits in one plant food source can be complemented by other plant or animal products that are higher in the

amino acid. Soy-based infant formulas are usually supplemented with small amounts of methionine, and these sustain good growth in infants. If animal foods are also included, protein quality is further increased. Digestibility of protein sources is a concern on some vegan diets that are fed to infants and young children because uncooked cereals and legumes are not well absorbed. Cooking or processing can improve their absorption and inactivate amylase inhibitors, lectins, tannins, and other substances in foods that decrease bioavailability.

Iron

The iron intakes of vegetarians are usually similar to those of omnivores. However, iron bioavailability may be decreased because nonheme iron absorption from plant foods is influenced by the presence of inhibitors, such as polyphenols, phytates, fiber, and tannins. Unleavened and unrefined cereals often contain many inhibitors. However, enhancers of nonheme iron absorption such as citric, ascorbic, and other organic acids that are often high in vegetarian diets may increase absorption. Also small amounts of animal foods, especially flesh foods, enhance absorption. Nevertheless, in infancy, pregnancy, the adolescent growth spurt and during other times of high iron requirements, without supplementation iron needs may not be met. Cereals highly fortified with iron or iron supplements may be acceptable iron sources to supplement other iron-rich foods.

Zinc

Vegetarian diets tend to be lower in zinc than nonvegetarian diets because approximately half of the zinc intake of western populations is from meat, fish, and poultry. The bioavailability of zinc is lower in cereal and other plant sources because phytate, fiber, and other inhibitors of zinc absorption are present, especially in unrefined foods. The presence of inhibitors is highest on vegan diets. Vegetarian diets should be planned to include adequate amounts of this nutrient, but supplements are not recommended since evidence is lacking that they are needed.

Calcium

Vegetarians who consume milk products have intakes of calcium as high as those of omnivores. Vegans have lower calcium intakes, and their intakes of vitamin D are also low. The major source of calcium in most western diets is dairy products. If they are not consumed or consumed in only small amounts, care must be taken to choose plant and other foods that are high in calcium. These include dark leafy green vegetables (chard, broccoli, kale, dandelion, mustard greens) and almonds that are high in calcium, fish such as sardines with bones, calcium-fortified soy milks and formulas, and calcium-set tofu. In recent years calcium-fortified cereals, orange juice, soy milks, and many other products have become widely available and are acceptable options for those who eschew milk. Supplemental calcium is another option, Calcium nutriture is compromised when vitamin D nutritional status is marginal since the vitamin is needed for absorption of the mineral.

Cobalamin (Vitamin B_{12})

Vegan diets are low in vitamin B_{12}, and other vegetarian diets are also somewhat lower than nonvegetarian diets. The reason for concern is that vitamin B_{12} deficiency causes not only a macrocytic megaloblastic anemia but, if long continued, serious, irreversible, neurologic and neuropsychiatric abnormalities. These are difficult to diagnose since the characteristic megaloblastic anemia of vitamin B_{12} deficiency is masked by high intakes of folic acid. Vitamin B_{12} is only found in products of animal origin, foods that contain microorganisms capable of incorporating the vitamin (nutritional yeast grown on vitamin B_{12}-enriched media), and vitamin B_{12}-fortified foods (B_{12}-fortified soy milk, nut milk, and cereals). Individuals such as some vegans who often do not consume any of these foods are at special risk.

Reliable and hygienic sources of vitamin B_{12} are important for those who consume no animal foods. Spirulina, kombu, tempeh, and miso contain inactive forms of the vitamins that cannot be utilized by human beings, and the plankton in seaweeds such as nori are unreliable and highly variable sources of the vitamin. Food and water are sometimes contaminated with vitamin B_{12}-producing bacteria but these are due to poor hygienic practices and are not reliable sources of the vitamin. Among individuals who are likely to have low intakes of vitamin B_{12}, high intakes of folic acid may mask the development of the megaloblastic anemia of vitamin B_{12} deficiency – another reason for insuring adequate intakes of vitamin B_{12} in all persons. For those over 50, regardless of their diets, vitamin B_{12} supplementation is recommended from fortified foods or supplements because the bioavailabiliy of food sources of B_{12} may be limited owing to the presence of unrecognized atrophic gastritis.

Folic Acid

Folic acid intakes are often higher among vegetarians than among omnivores because of high intakes of fruits and vegetables. However, there is the risk of possible inadequacy arising in persons who consume diets consisting of vegetables which are braised or

fried at high temperatures, few fruit juices, and cereals fortified with the nutrient. For women who are at risk of becoming pregnant, regardless of their dietary pattern, supplementation with 400 μg folic acid is currently recommended to decrease risks of neural tube defects.

Vitamin D

Vitamin D is only found in animal foods such as vitamin D-fortified dairy products and highly fortified cereals, and liver. Cod liver oil and menhaden oil contain high but variable amounts of vitamin D. More acceptable sources are water- or fat-soluble vitamin D supplements, commercial infant formula (both soy- and cows' milk-based), and fortified cereals. Sunlight exposure may be insufficient in young infants at weaning and thereafter, and among the elderly, especially those in northern climes with inclement weather.

Riboflavin

The major source of riboflavin in most western diets is milk, and those who avoid milk need to select other foods high in this nutrient.

Fiber

Intakes of whole-grain cereals and nonrefined foods and thus of dietary fiber are higher on most vegetarian diets. On some vegan diets fed to small children, fiber intakes are several times expert recommendations of $0.5\,g\,kg^{-1}\,day^{-1}$, and the sheer bulk of the diet may pose a problem. In adults, vegetarians' fiber intakes tend to be closer to expert recommendations than are those of nonvegetarians.

Other Nutrients

Certain other nutrients such as carnitine and taurine may be low on vegan and some restrictive vegetarian diets, but at present the health significance of these differences is not established. Intakes of sodium vary on vegetarian diets; there are some reports that sodium intakes are lower among vegans than among omnivores. Lower sodium intakes are generally recommended.

Usually vegetarian diets are higher in thiamin, vitamin C, β-carotene and other carotenoids, and vitamin E than are nonvegetarian diets. Currently there is much interest in the health effects of phytochemicals other than nutrients that are present in plant foods, such as the flavonoids, plant stanol esters, and other compounds, which are likely to be higher in the diets of vegetarians than in those of omnivores. The beneficial health effects of these dietary constituents are still under study.

Benefits

Vegetarians who have chosen such an eating pattern generally enjoy their diets and are satisfied with their lifestyles. Those who consume vegetarian diets because of economic necessity may add animal foods when their means permit. Differences in all-cause mortality between vegetarians and nonvegetarians are equivocal, although differences in specific causes of mortality do seem to differ. The health effects of vegetarian diets are more difficult to interpret since vegetarians often not only have divergent dietary patterns but lifestyles that differ in ways that also affect their risks for disease. These include increased physical activity, decreased use of tobacco and alcohol, and moderation in food intake, all of which can be expected also to decrease chronic disease risks regardless of what diet is consumed. Nevertheless, even after taking these factors into account, decreased risks and later onsets for several chronic degenerative diseases are apparent that appear to be due to dietary pattern. In adults, vegetarian diets are positively associated with reduced risks of obesity, coronary artery disease, hypertension, non-insulin-dependent diabetes mellitus, and possibly some types of hormone-dependent cancers. Osteoporosis does not seem to be different from that of the general omnivorous population among vegetarians. It may be higher in vegans, especially if they are very low in weight, and have low intakes of calcium and vitamin D. However, other factors that affect risks, such as weight-bearing exercise and nonsmoking, may help to decrease risks. Differences in infectious diseases between vegetarians and nonvegetarians are not as apparent as are those with the chronic diseases mentioned above. Foodborne illnesses that are more frequently associated with animal foods are probably less common among vegetarians, although there is no evidence that other foodborne diseases are lower among them. Although some vegetarians are more likely to self-treat and less likely than their nonvegetarian counterparts to seek medical care for mild illnesses, for severe illnesses the vast majority of vegetarians seek medical care. Therapeutic diet manuals are available for those with vegetarian eating patterns. However, the claims of macrobiotics, living foods advocates, and fruitarians that their regimens are especially or uniquely helpful for treating cancers or other maladies are not supported.

Risks

Like their omnivorous counterparts, vegetarian diets can be well or poorly planned. The risks to dietary inadequacy increase with the number and degree of

restrictions on the food groups and foods that are consumed, the willingness to use fortified foods or nutrient supplements, and to accept medical advice, and the physiological state of the individual. Vegetarians who consume unplanned regimens or feed them to their children are no more immune to nutritional problems of excess, inadequacy, and imbalance than are omnivores. Particular risks arise when many food groups and foods are eliminated without due care being given to providing alternative sources of nutrients either from other conventional foods, fortified foods, nutrient supplements, or a combination of these. This is a particular problem for those who by virtue of their increased nutritional needs or their health status are especially nutritionally vulnerable. A vegetarian diet, which might suffice for good health in other groups, may not meet their nutritional needs and may be inappropriate for them. Such groups include the very young (weanlings and young children), those who are growing rapidly (adolescents), those likely to become pregnant, pregnant and lactating women, the very old, and the sick. For example, infants who are weaned to vegan diets may develop rickets, iron-deficiency anemia, protein-calorie malnutrition, and other dietary deficiency diseases. There is also a small minority of vegetarians whose convictions and practices place them at increased health risk. They refuse to use prescription drugs, reject immunizations using animal antiserum, medications using recombinant products such as epoietin-α, human recombinant insulin, and blood transfusions, or avoid or delay other conventional medical care. Such practices may increase disease risks or severity. They may self-treat using unproven remedies including herbal and botanical preparations without due caution to make sure that they are safe. Therefore not only dietary patterns but also associated health behaviors need to be assessed.

The most common dietary deficiencies observed on unplanned vegetarian diets are iron-deficiency anemia (particularly in weanlings, young children, and pregnant women), rickets in weanlings and young children, megaloblastic anemia due to vitamin B_{12} deficiency (in pregnant and lactating women, infants and young children, and the aged), and osteomalacia and low bone density in vegan adults and the elderly.

Dietary Planning

Table 2 presents some key points for dietary planning for each of the common vegetarian dietary patterns. Well-planned vegetarian diets that meet the dietary standards for nutrients as determined by scientific expert groups are in line with current dietary recommendations. Balance, variety, and moderation are essential and can be achieved by focusing on alternative sources for nutrients high in foods that are excluded on the vegetarian pattern. The use of a wide variety of plant foods, the advent of nutrient-fortified plant foods, and the judicious use of vitamin and mineral supplements at times of particularly high needs may be helpful. Sound vegetarian dietary guides have been published that assure adequacy of nutrient intakes, and these facilitate dietary planning, especially for vegans. Since periods of rapid growth or physiological stress are times when risks of nutritional deficiencies are greatest, special attention is warranted to planning dietary intakes during these nutritionally vulnerable times during the life cycle, especially for those who consume vegan diets. Because diets that permit good health in adults may not do so during these times of particular stress, special guidance from a dietitian is often helpful as

Table 2 Considerations in dietary planning for vegetarian patterns

Dietary pattern	*Key issues and comments*
Meatless, partial or semivegetarian	These diets are similar in most respects to nonvegetarian diets and fit well with dietary recommendations to decrease risks of chronic degenerative diseases
Lactoovo vegetarian	Liberal amounts of vitamin D-fortified dairy products and some eggs in the diet allow for adequate intakes of nutrients, with the possible exception of iron, which can be obtained from fortified cereals. Low-fat dairy products are preferred to keep intakes of saturated fat and total fat moderate
Lactovegetarian	Same as above
Macrobiotic	Current variations of this diet are less restrictive than versions used two decades ago, but deficiencies of energy, iron, calcium, vitamin B_{12}, vitamin D, and other nutrients may still arise in weanlings, pregnant women, and young children, especially if diets are unplanned
Vegans or 'pure vegetarian'	Without careful planning, energy and vitamins B_{12}, D, and bioavailable sources of iron may be low. Concentrated sources of energy can increase caloric intakes. Vitamin B_{12}, D, and calcium can be supplied through use of soy milk, cereals, and/or supplements of these nutrients. A variety of protein food sources assures adequate complementary proteins. Iron intakes may be enhanced by iron-fortified cereals, green leafy vegetables, and iron supplements if these are acceptable. A vegan food guide can be helpful

well. If there is doubt as to the nutritional adequacy and healthfulness of vegetarian diets, the assistance of a dietitian and dietary assessment of food records of typical intake is warranted. Many computerized dietary assessment programs are available that can assist in this task.

Conclusion

Plant-only (vegan, macrobiotic, and fruitarian), plant-based (lactovegetarian, lactoovo vegetarian, meatless), and plant-rich omnivorous diets that are planned to insure dietary adequacy and minimize chronic disease risk are compatible with good nutritional status and health.

Acknowledgment

This review was supported in part by contract number 53/3 K06-10 from the US Department of Agriculture to Tufts University.

See also: **Anemia (Anaemia)**: Iron-deficiency Anemia; Megaloblastic Anemias; **Calcium**: Properties and Determination; **Cholecalciferol**: Properties and Determination; **Cobalamins**: Properties and Determination; **Dietary Fiber**: Properties and Sources; **Folic Acid**: Properties and Determination; **Functional Foods**; **Iron**: Properties and Determination; **Macrobiotic Diets**; **Protein**: Food Sources; **Riboflavin**: Properties and Determination; **Rickets and Osteomalacia**; **Zinc**: Properties and Determination

Further Reading

American Dietetic Association (1993) Position of the American Dietetic Association on vegetarian diets. *Journal of the American Dietetic Association* 11: 1317–1319.

Dagnelie P, Van Staveren W, Vergate F, Dingan D, Van den Berg H and Hautvast J (1989) Increased risk of vitamin B_{12} and iron deficiency in infants on macrobiotic diets. *American Journal of Clinical Nutrition* 50: 818–824.

Duyff RL (1997) The vegetarian way. In: Duyff RL (ed.) *The American Dietetic Association's Complete Food and Nutrition Guide*, pp. 553–572, 595. Minneapolis: Chronimed Publishing.

Dwyer JT (1996) Vegetarianism for women. In: Krummel D and Kris-Etherton PM (eds) *Nutrition in Women's Health*, pp. 232–263. Gaithersburg, MD: Aspen.

Dwyer JT and Loew FM (1994) Nutritional risks of vegan diets to women and children: are they preventable? *Journal of Agricultural and Environmental Ethics* 7: 87–109.

Haddad EH (1994) Development of a vegetarian food guide. *American Journal of Clinical Nutrition* 59 (suppl.): 1248–1254.

Haddad EH, Sabate J and Whitten CG (1999) Vegetarian food guide pyramid: a conceptual framework. *American Journal of Clinical Nutrition* 70 (suppl.): 615S–619S.

Harper AE (ed.) (2000) Physiologically active food components. *American Journal of Clinical Nutrition* 71: 1651S–1743S.

Herbert V (1997) Vitamin B_{12}. In: Ziegler EE and Filer LJ (eds) *Present Knowledge in Nutrition*, 7th edn, pp. 191–206. Washington, DC: ILSI Press.

Hodgkin G and Mahoney S (eds) (1995) *Seventh-Day Adventist Diet Manual Including a Vegetarian Plan*, 7th edn, pp. 1–559. Loma Linda, CA: Seventh-Day Adventist Dietetic Association.

Johnston PK (ed.) (1994) Vegetarian nutrition: proceedings of a symposium held in Arlington VA. *American Journal of Clinical Nutrition* 59 (suppl.): 1099S–2362S.

Johnston PA (1999) Nutritional implications of vegetarian diets. In: Shils ME, Olson JA, Shike M and Ross AC (eds) *Modern Nutrition in Health and Disease*, 9th edn, pp. 1755–1768. Baltimore: Williams and Wilkins.

Johnston PK and Sabate J (eds) (1999) Third International Congress on Vegetarian Nutrition. *American Journal of Clinical Nutrition* 70: 429S–634S.

Key TF, Fraser GE, Thorogood M *et al*. (1998) Mortality in vegetarians and nonvegetarians: a collaborative analysis of 8300 deaths among 76 000 men and women in five prospective studies. *Public Health Nutrition* 1: 33–41.

Mangels R (1995) Nutrition section. In: Wasserman D (ed.) *Simply Vegan*, 2nd edn. Baltimore: Vegetarian Resource Group.

Messina M and Messina V (1996) *The Dietitian's Guide to Vegetarian Diets*. Gaithersburg, MD: Aspen.

Mutch PB and Johnston PK (eds) (1988) Proceedings of the First International Congress on Vegetarian Nutrition. *American Journal of Clinical Nutrition* 48 (suppl.): 707–927.

Sanders TAB (1999) Essential fatty acid requirements of vegetarians in pregnancy, lactation and infancy. *American Journal of Clinical Nutrition* 70: 555S–559S.

Sanders TAB and Reddy S (1994) Vegetarian diets and children. *American Journal of Clinical Nutrition* 59 (suppl.): 1176–1181.

Smith J and Dwyer JT (1998) Vegetarian diets for children. In: Dershewitz RA (ed.) *Ambulatory Pediatric Care*, 3rd edn, pp. 106–112. Philadelphia: Lippincott Raven.

Wasserman D and Mangels R (1995) *Simply Vegan*, 2nd edn. Baltimore: Vegetarian Resource Group.

Young VR and Pellett PL (1994) Plant proteins in relation to human protein and amino acid nutrition. *American Journal of Clinical Nutrition* 59 (suppl.): 1203–1212.

VERMOUTH

P Liddle and L Boero, Martini & Rossi, Pessione, Torino, Italy

History

The origins of vermouth date back to ancient Mediterranean history. The maceration of herbs and spices in wine was common practice in antiquity, and the invention of aromatized wine, the ancestor of vermouth, has been attributed to Hippocrates. Greece abounds in aromatic plants and Crete in particular provided the plants dittany and wormwood, both of which possess tonic and digestive properties. It is reported that Hippocrates macerated the flowers of these plants in strong, sweet Greek wine, thereby obtaining a satisfying and digestive beverage which throughout antiquity and the Middle Ages was called 'Hippocratic wine' or *vinum absinthianum*. Later, the Romans elaborated on the production of such wines by introducing other herbs such as thyme, rosemary, myrtle, and celery.

In the Middle Ages the Venetians, who had the monopoly of the spice trade, introduced aromatic plants into Italy which until then were unknown in that country, and they were used in the preparation of Hippocratic wine. These plants (cardamom, cinnamon, myrrh, cloves, rhubarb, ginger, and sandalwood) came from East Africa, India, and Indonesia.

Turin, along with Florence and Venice, was one of the major Italian centers of production of Hippocratic wines and liqueurs from the late eighteenth century. Piedmont in particular possessed two assets: aromatic plants abounded in the Piedmont Alps and their foot-hills, and the bouquet of Piedmont dry and sweet white wines combined very well with the fragrance of herbs. These two factors doubtless explain why this area became the major center of the vermouth industry in the nineteenth century, when the leading names in that sector emerged.

The term 'vermouth' derives etymologically from Wermut, the German name for wormwood. While the latter English term refers to the use of the plant as a vermifuge, the German name is supposedly derived from Wer (man) and Mut (courage, spirit). When it was introduced into Bavaria in the first half of the seventeenth century by the Piedmont producer Alessio, *vinum absinthianum* was probably translated literally as Wermutwein, which when it reached France became vermouth.

Today, most of the world's vermouth is produced in Europe (essentially in Italy, France, and Spain) and to a lesser extent in South America, but local varieties are always to be found in all wine-producing countries.

Definition

In basic terms, vermouth is a combination of wine, aromatic plants (hereafter referred to as 'botanicals'), sugar, sometimes grape must in limited quantities, and alcohol. Caramel is the only coloring substance authorized, and is used for red vermouths.

The difference between the various types of vermouth lies essentially in the presence or absence of certain botanicals in their formulas. Broadly speaking, there are two main families – sweet and dry – which differ in their sugar content. Sweet vermouths contain around 150 grams of sugar per liter, while the dry varieties contain less than 50 g l^{-1}. Dry vermouths are all white, and serve mainly as a base for well-known cocktails, although they may also be drunk straight.

In the European Union (EU), vermouth is defined within the European Council Regulation 1601/91 and subsequent amendments. This Regulation applies to 'the definition, description and presentation of aromatized wines, aromatized wine-based drinks, and aromatized wine-product cocktails.' Article 2 (1) (a) of the Regulation states:

> aromatized wine shall mean a drink:
> – obtained from wines defined in points 12 to 18 of Annex I to Regulation (EEC) No. 822/87 (11), as last amended by Regulation (EEC) No. 1325/90 (12), with the exception of retsina table wine, and possibly with added grape must, grape must in fermentation and/or fresh grape must with fermentation arrested by the addition of alcohol, as defined by Community legislation,
> – to which alcohol has been added as defined in Article 3 (d), and
> – which has been flavoured with the aid of:
> – natural flavouring substances and/or natural flavouring preparations as defined in Article 1 (2) (b) (i) and (c) of Directive 88/388/EEC . . .
> – aromatic herbs and/or spices and/or flavouring foodstuffs,
> – which has generally been sweetened and, subject to the exceptions provided for in paragraph 2, has possibly been coloured with caramel,
> – which has a minimum actual alcoholic strength by volume of 14.5% vol or more and a maximum actual alcoholic strength by volume of less than 22 % vol . . .

The wine used in the preparation of an aromatized wine must conform to the relevant EU wine legislation (in particular as far as the minimum alcohol content is concerned), and be present in the finished product in a proportion of not less than 75%.

Vermouth itself is further defined in Article 2 (2) (a) as an:

> aromatized wine . . . the characteristic taste of which is obtained by the use of appropriate derived substances, in particular of the *Artemisia* species, which must always be used; this drink may be sweetened only by means of caramelized sugar, sucrose, grape must, rectified concentrated grape must and concentrated grape must.

Denominations relating to the sugar contents, e.g., sweet, extra-dry, are defined in Article 2 (5). In addition, EU Regulation 122/94 allows for the use of the pure flavoring substance vanillin, identical to that found in natural sources.

Definitions in other parts of the world are broadly similar, with slight variations on the minimum wine content (70% in some cases), and minimum/maximum alcohol contents. In all cases, the base wine and botanicals used must comply with the relevant local legislation concerning these ingredients.

Finally, contrary to some popular beliefs, vermouth is not a 'vin cuit' (cooked wine), and, to put to rest another long-standing rumor, it is not a particularly acidic product – in fact, the acidity is often less than that of the average table wine.

Ingredients

Wine

Quantitatively, wine is the most important ingredient of vermouth, since in the EU it accounts for at least 75% of its volume. The quality of the final product therefore depends on the quality of the wine employed. Generally speaking, only neutral white wines that do not oxidize are employed; they must be low in tannin if they are not to maderize with age, and turn a darker color.

Botanicals

The botanicals (leaves, flowers, fruits, seeds, roots, and barks) are the natural sources of flavor which characterize vermouth. There are so many of these that it would be impossible to list them exhaustively and, of course, the composition of the mixture is a closely guarded secret of each producer, but some examples are given in Table 1, together with their respective Latin botanical names.

As with other foods and beverages, use of traditional botanicals may introduce trace levels of compounds which in their pure (or, at least, more concentrated) form may be suspected of negative effects on the health of consumers, based on toxicological studies. These compounds, usually present at or below the mg l^{-1} level, are hence subject to quantitative limitation by the legislation of different countries. In Europe, these limits are set by the Directive 88/388/EEC, and limits of similar magnitude are found in other national or international legislation.

In the particular case of vermouth, the use of *Artemisia* species forms part of the EU definition of the product, and virtually all of the *Artemisia* species contain thujone, a compound often associated with the negative aspects of the drink absinthe. However, the most common source of thujone in foodstuffs is sage, a botanical whose name and history are synonymous with good health, and which is widely and traditionally used as a culinary herb in many countries.

The EU Directive, for example, currently fixes a maximum limit for thujone of 5 mg kg^{-1} in alcoholic beverages of less than 25% alcohol. Sweet vermouths are well within these limits, generally containing less

Table 1 English and Latin botanical names of examples of natural sources of flavor used in the production of vermouth

English name	*Latin botanical name*
Angelica	*Angelica archangelica*
Artemisia species	Various
Bay	*Pimenta racemosa*
Camomile	*Anthemis nobilis*
Cardamom	*Elettaria cardamomum*
Cinchona	*Cinchona species*
Cinnamon	*Cinnamomum zeylanicum*
Citrus species	Various (orange, lemon, etc.)
Clary sage	*Salvia sclarea*
Clove	*Eugenia caryophyllus*
Coriander	*Coriandrum sativum*
Dittany of Crete	*Origanum dictamnus*
Elder	*Sambucus nigra*
Gentian	*Gentiana lutea*
Ginger	*Zingiber officinale*
Hops	*Humulus lupulus*
Juniper	*Juniperus communis*
Lemon balm	*Melissa officinalis*
Licorice	*Glycyrrhiza glabra*
Marjoram	*Origanum vulgare*
Nutmeg and mace	*Myristica fragrans*
Orris	*Iris florentina*
Quassia	*Quassia amara*
Raspberry	*Rubus idaeus*
Rhubarb	*Rheum officinale*
Rose	*Rosa damascena*
Saffron	*Crocus sativus*
Sage	*Salvia officinalis*
Savory	*Satureia hortensis*
St John's wort	*Hypericum perforatum*
Star anise	*Illicium verum*
Thyme	*Thymus vulgaris*
Vanilla	*Vanilla planifolia*

than 1 mg kg^{-1}, a level considerably lower than those in past or current versions of absinthe; most dry vermouths contain even less.

Other such compounds may be introduced into traditional flavored alcoholic beverages by botanicals that also often have long-standing histories of culinary use. In all cases, careful selection of raw materials, and if necessary the mode of preparation of the botanical extracts, results in products which respect current limits. In addition, it should be remembered that the very presence of alcohol in flavored wines and spirits constitutes an 'automatic' limiting factor on the ingestion of such compounds during moderate consumption of these products.

Alcohol

Ethyl alcohol also enters into the composition of vermouth, both to fortify the wine and as a means of extracting the flavoring substances from the botanicals. It must be of agricultural origin, very pure, and conform to standards laid down by legislation (e.g., concerning the maximum level of methanol and other natural fermentation byproducts). In Europe, such specifications are given in Annex I to Regulation 1601/91.

Water

Although not a major ingredient, the water used (mainly together with alcohol for the botanical extracts) must usually conform to the standards for drinking water. This is the case in Europe, as embodied in the current and future requirements of EU Directives concerning potable water (80/778/EEC and 98/83/EC).

Sugars

In order to give vermouth its required sugar content, mistelle (muted grape must) or a good dessert wine may be added to the base wine, as well as the necessary amount of good-quality white sugar. The sugars slightly attenuate the bitterness of certain substances that would otherwise be too strong on the palate. They give the vermouth body, firmness, and smoothness, and thus play a very important role in the preparation. As mentioned above, sweet vermouths contain about 150 g l^{-1} of sugar, and dry vermouths less than 50 g l^{-1}.

Caramel

Red vermouth generally owes its amber hue to caramel, which in Europe is the only coloring matter authorized under Regulation 1601/91. Apart from imparting color, it also confers a special flavor specific to red vermouth, contributing to the body and smoothness.

Production

Prior to the marriage of the ingredients of vermouth, the base-wine blend and the botanical extracts must be prepared.

Base-wine blend

The preparation of the base-wine blend is performed using traditional processing aids common to all wines and wine-based products, e.g., gelatin, bentonite, charcoal, etc. Colloid formation and coagulation reduce the levels of high-molecular-weight compounds and of traces of iron and other metals, all of which could compromise the clarity and stability of the finished product. As with all wine processing, small amounts of potassium ferrocyanide may be necessary to reduce the trace metals. Due to losses from the base wines during the transfer and blending, small amounts of sulfur dioxide may be added to maintain the antiseptic and antioxidant effects of the latter. After a suitable contact time (up to several days), the blend is decanted and filtered.

Although alcohol is generally added at a later stage of the process, the base wines of some vermouths (particularly in France) may be fortified by the addition of this ingredient at an earlier stage. They are then subjected to a physical aging in wood (sometimes up to a year), as well as a form of biological aging similar to the flor sherry process used in Spain, to enrich the organoleptic qualities of the wines.

Botanical Extracts

The botanicals are usually incorporated into the vermouth in the form of an extract (strictly speaking, a tincture) produced by macerating them in aqueous alcohol or a distillate obtained by distilling them in the presence of aqueous alcohol (an alcoholate).

Maceration is the most commonly used process. The botanicals are placed in a tank, covered with a mixture of aqueous alcohol, and are agitated periodically so that they remain constantly covered by the liquid. The operation is frequently performed in rotating tanks and may last for several weeks. At the end of this time, the extract is drawn off and the botanicals are pressed.

Distillation is used when it is desired to make particular use of the volatile substances contained in the botanicals (e.g., fruits, St John's wort), and to exclude higher-molecular-weight compounds which could have negative physical or organoleptic effects. The botanicals are distilled in the presence of a mixture of alcohol and water, with or without an initial period of maceration. The actual strength of the aqueous alcohol used, and the complete or partial recovery

of the distillate, together form part of the traditional recipe that characterizes a particular vermouth.

In some cases, in particular in the production of a well-known French dry vermouth, the botanicals are macerated directly in the wine, which has already been fortified by alcohol.

Blending the Ingredients

The sugar required in the product is generally incorporated into the base-wine blend, which is then mixed with alcohol, water, and the botanical extracts, together with caramel if required, in blending tanks, which often have a capacity of up to 200 000 l. After careful homogenization of the liquid, the vermouth is then allowed to mature for up to several weeks, in order to achieve a proper harmony and balance of its ingredients.

Stabilization

At the end of this maturation time, the vermouth undergoes cold stabilization treatment. It is refrigerated, and is held for several days at around $-8\,^{\circ}C$, close to its freezing point. This precipitates substances (mainly potassium bitartrate) which may form a natural deposit later if the vermouth is subjected to low temperatures during storage and transport, or to contact with ice when served. The precipitate is removed by a low-temperature filtration, guaranteeing the physical stability of the product under all conditions.

Finally, to insure perfect clarity and brilliance, and also biological stability in the case of products with an alcohol strength of less than 16% vol, the vermouth is generally subjected to a very fine sterilizing filtration immediately prior to bottling. The latter is carried out using the modern techniques common to other sectors of the beverage industry.

An example of an overall production process is shown in Figure 1.

Analysis

Most of the routine analyses of vermouth are of necessity those used for the analysis of wines, since wines form the basis of all vermouths. In addition, modern techniques of gas and liquid chromatography are used to analyze for the volatile and nonvolatile compounds which are derived from the botanicals used, and which contribute to the specific aroma and taste of vermouth. In both cases, these chromatographic techniques often need to be coupled to mass

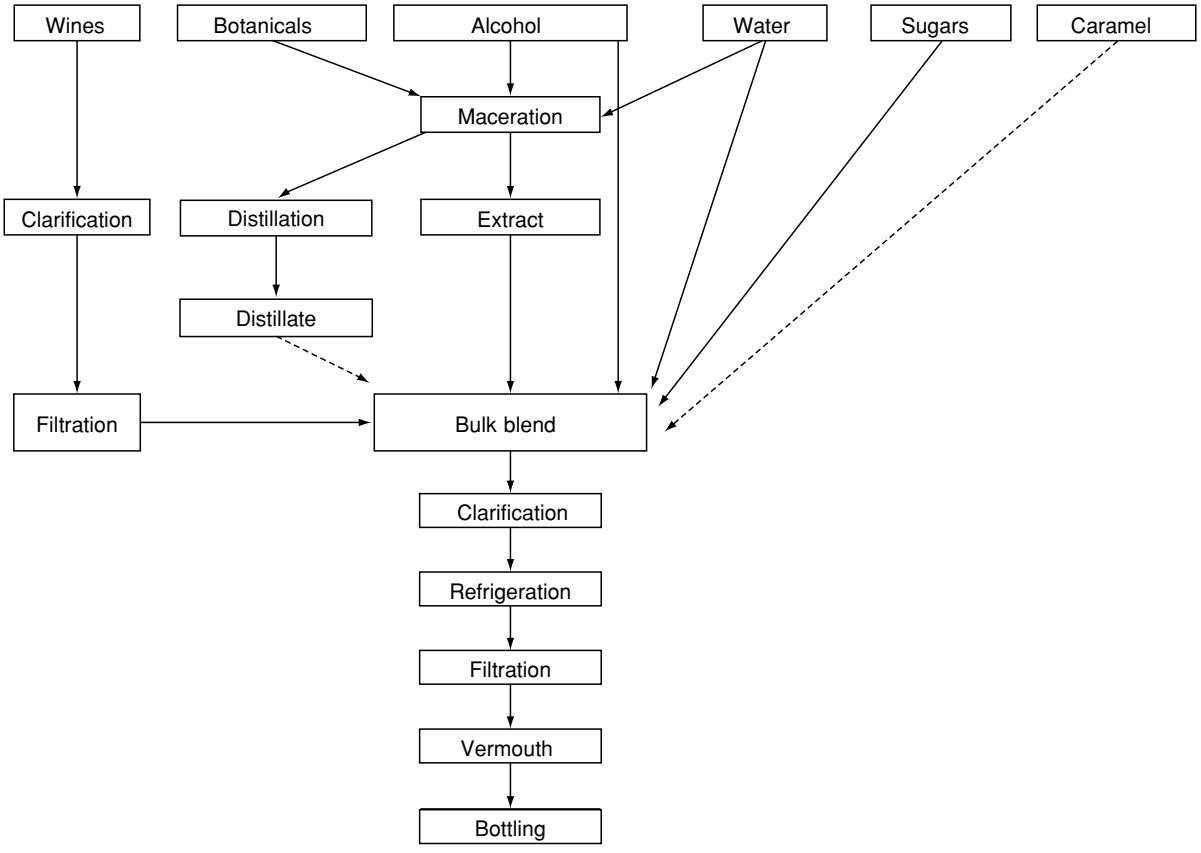

Figure 1 Typical overall production process for vermouth.

spectrometry in order to identify and monitor the low levels of these compounds (typically a total of the order of less than a few tens of $mg\,l^{-1}$).

In common with the rest of the food industry, great care is also taken today to monitor and limit any low levels of contaminants, using these same techniques, which may potentially be introduced by contact with processing or transport equipment, or packaging materials.

See also: **Alcohol**: Properties and Determination; **Caramel**: Properties and Analysis; **Grapes**; **Herbs**: Herbs Used in Alcoholic Drinks; **Spices and Flavoring (Flavouring) Crops**: Use of Spices in the Food Industry; **Wines**: Types of Table Wine

Further Reading

Amerine MA (1973) A multi-language dictionary of vermouth ingredients. *Rivista Italiana EPPOS* 15: 504–516.

Anonymous (1974) From the wine of Hippocrates to vermouth. *Revue Vinicole Internationale* 1–2: 33–36.

Council Directive 88/388/EEC of 22 June 1988 on the approximation of the laws of the Member States relating to flavourings for use in foodstuffs and to source materials for their production. *Official Journal of the Europeon Communities* L 184, 15/07/1988, pp. 61–66.

Council Regulation (EEC) No 1601/91 of 10 June 1991 laying down general rules on the definition, description and presentation of aromatized wines, aromatized wine-based drinks and aromatized wine-product cocktails. *Official Journal of the Europeon Communities* L 149, 14/06/1991, pp. 1–9.

Commission Regulation (EC) N. 122/94 of 25 January 1994 laying down certain detailed rules for the application of Council Regulation (EEC) N. 1601/91 on the definition, description and presentation of aromatized wines, aromatized wine-based drinks, and aromatized wine-product cocktails. *Official Journal of the European Communities* L 21, 26/01/1994, p. 7.

VIBRIOS

Contents

Vibrio cholerae

J G Morris, Jr, University of Maryland School of Medicine, Baltimore, MD, USA

Introduction

Vibrio cholerae is a curved, Gram-negative bacilli. The organism is widely distributed in aquatic environments, where it is a natural, free-living organism. A subset of *V. cholerae* strains, carrying the genes for cholera toxin (CT) and the *Vibrio* pathogenicity island (VPI), cause cholera, a disease characterized by severe, dehydrating diarrhea; virtually all of these latter 'epidemic strains' are within either O group 1 or O group 139. Cholera remains a persistent cause of morbidity and mortality in Asia and Africa. In 1998, the World Health Organization (WHO) received reports of 293 121 cholera cases and 10 586 deaths from 74 countries, with the actual number of cases (including unreported cases) probably greatly exceeding this figure. Cholera can also move explosively through populations that have been free of the disease for generations, as demonstrated by the occurrence of over 100 000 cases in Peru in 1991 and the subsequent rapid spread of cholera to other South American countries.

V. cholerae strains which lack the virulence factors necessary to cause epidemic cholera ('nonepidemic strains') have in the past been referred to as non-O1 *V. cholerae*, nonagglutinable (NAG) *Vibrio*, or noncholera *Vibrio*. While the majority of these environmental, nonepidemic strains appear to be nonpathogenic for humans, some have been linked with human disease.

Cholera

Cholera has been endemic in the delta of the Ganges and Brahmaputra rivers since the beginning of recorded history. Seven pandemics of cholera have occurred since 1817. The seventh pandemic began in Indonesia in 1961, with subsequent spread through Asia, Africa,

parts of Europe, and, beginning in 1991, South America. Transmission during an epidemic is thought to occur primarily through contamination of potable water. However, foodborne transmission is well recognized, and may play an important role in the introduction of the organism to a new area or transmission in areas of low endemicity, such as the USA.

Illness in cholera is due primarily to the action of CT, a protein enterotoxin that stimulates profuse, watery diarrhea, leading to dehydration and death. *V. cholerae* strains may be classified by O group, and prior to 1992, it was assumed that epidemic cholera could only be caused by strains in O group 1 (*V. cholerae* O1). In 1992, cholera caused by a new serogroup, subsequently designated as O group 139 (*V. cholerae* O139), appeared in epidemic form in India and Bangladesh. Genetic studies have demonstrated that O139 strains are virtually identical to O1 strains, except for the replacement of a 22-kb DNA fragment responsible for O1 antigen biosynthesis with a 35-kb capsular polysaccharide biosynthesis region. With this change in surface antigens, *V. cholerae* O139 strains were able to cause cholera in persons with prior immunity to *V. cholerae* O1. Key cholera virulence factors can be transferred among *V. cholerae* strains by phage, and it is likely that other serotypes (such as O37), when carrying an appropriate complement of virulence genes, can also cause cholera.

Occurrence in the Environment

As previously noted, *V. cholerae* is a free-living organism in the aquatic environment. Growth is influenced by salinity (isolation generally occurs in water with a salinity of 2–20 parts per thousand) and water temperature (optimally > 17 °C). *V. cholerae* produces a chitinase, and may bind to and colonize plankton and other aquatic organisms. The bacterium can shift to a 'rugose' morphologic form, in which it produces an amorphous exopolysaccharide that contributes to the formation of biofilms. Strains in the environment may also assume a 'viable but nonculturable' form. Studies in Peru demonstrate that seasonal cholera epidemics are often preceded by increases of *V. cholerae* in the environment, with subsequent 'spill-over' of the bacterium into human populations. As humans become infected, *V. cholerae* present in their feces can contaminate food and water sources and further increase environmental *V. cholerae* levels, resulting in amplification of the organism and initiation of the epidemic cycle.

Entry to Food Chain

Entry of *V. cholerae* into the food chain is most clearly described in association with shellfish. In areas of low endemicity, such as the USA, the occurrence of illness has been linked with the consumption of raw or undercooked shellfish (crabs, oysters) harvested from areas in which the organism is known to be present. Outbreaks of illness have also been reported in association with consumption of raw or undercooked mussels, cockles, and fish. Environmental and/or fecally derived organisms may contaminate other foods as a result of irrigation or when water is used for cleaning or washing of fruits and vegetables or other foods eaten without cooking. Direct fecal contamination of cooked foods within a household can also serve as a route of transmission, particularly if foods are left at ambient temperature for a period of hours after contamination.

Fate on Processing/Storage

There is an extensive body of literature on the survival of *V. cholerae* in food, beginning with the observation of Babes in 1885 that cholera *Vibrio* remained alive up to 48 h on fresh nonacid vegetables, potatoes, and cheese, but not longer than 24 h on sour fruit and vegetables. In general, cooked foods appear to be a more favorable milieu than raw foods as cooking destroys many competing bacteria. Growth is inhibited by low pH and enhanced by neutral or alkaline pH. Recent studies suggest that cooked rice is a particularly good growth medium.

Survival on shellfish has been reported for 15–45 days, depending on the holding temperature. *V. cholerae* are readily killed by high temperatures (5 min at 65 °C, 'seconds' at 80–100 °C). However, studies conducted by the US Centers for Disease Control during the Louisiana cholera outbreaks in the late 1970s demonstrated that, after experimental inoculation of the organism, *V. cholerae* O1 could be cultured from crabs that were boiled up to 8 min or steamed for up to 30 min; crabs boiled for 8 min or steamed for 25 min were red, and the meat was firm and appeared well-cooked. In recent studies with cockles, mild heat treatment for up to 25 s was inadequate to eliminate *V. cholerae* O139; the persistence of *V. cholerae* O1 in frozen shrimp, with shell, has been reported for over a month.

With the rapid means of transportation now available, any raw or undercooked food (particularly raw or undercooked seafood) imported from an area with epidemic or endemic disease has the potential for carrying the organism. This was dramatically illustrated with the 1991 South American epidemics, where transmission of the disease across national borders (and across continents) was associated with importation of food items such as ceviche (raw, marinated fish) and crab meat. The risk of such transmission is dependent on the level of sanitation in the

country of origin, the extent of the disease within the country (i.e., products from a country with low levels of endemic disease constitute a much smaller risk than products from a country in the midst of a major epidemic), and the handling of the product during shipment. In this context, it should be noted that WHO feels that the risk of cholera transmission associated with trade in food products is 'very small,' and considers restrictions on food and beverage importations to be 'inappropriate cholera control measures.'

Detection in Food, Water, and Physiological Samples

Isolation of the organism from environmental samples generally requires an enrichment procedure. Enrichment broths rely on the ability of *Vibrio* species to grow at alkaline pH in the presence of bile salts, or upon the ability of *Vibrio* to use gelatin as the only available carbon and nitrogen source: commonly used enrichment broths include alkaline peptone water and gelatin phosphate saline broth. The enrichment broth should be streaked within 6–8 h after inoculation on to a selective media such as thiosulfate citrate bile salts sucrose (TCBS) agar. Isolation of the organism from stool samples also requires the use of a selective medium such as TCBS, and may be facilitated by an enrichment step in alkaline peptone water.

Identification of typical colonies from TCBS plates is based on standard biochemical characteristics. Epidemic strains can be tentatively identified by serotyping with O1 and O139 specific antisera. *V. cholerae* of O group 1 can be further divided into two serovars, Ogawa and Inaba, and a rare third type, Hikojima, as well as two biotypes, classical and El Tor. Confirmation that an isolate is capable of causing epidemic disease requires demonstration that the strain has the ability to produce cholera toxin. DNA probes are now available for identification of strains carrying the genes for cholera toxin, VPI, and other critical virulence factors. These probes have proven to be extremely useful in epidemiologic studies and in the screening of environmental isolates.

Symptoms and Characteristics

Despite the dread inspired by the term 'cholera', the majority of persons infected with epidemic *V. cholerae* strains do not have severe illness: 75% of persons infected with strains of the classical biotype have inapparent or mild disease, while 93% of infections with biotype El Tor strains (which are responsible for the most recent pandemic) have illnesses that are inapparent or mild. The infectious dose is dependent to a large degree on gastric acidity and the associated buffering capacity of the vehicle carrying the organism. When normal, healthy, adult North American volunteers ingested 10^6 *V. cholerae* O1 organisms in water, neither illness nor infection resulted. However, volunteers consistently became infected after ingesting 10^6 organisms with 2 g of sodium bicarbonate or with food; 10^3 organisms caused illness in four of six volunteers when administered with sodium bicarbonate.

The incubation period for cholera varies from several hours to 7 days. Although there may be premonitory symptoms such as anorexia and abdominal pain, in most cases onset is sudden, with profuse diarrhea. Initially the stool is brown with fecal matter, but soon the evacuations assume a pale gray color; mucus in the stool imparts the characteristic rice-water appearance. Tenesmus is absent; instead, there is a feeling of relief as enormous amounts of fluid are passed effortlessly. Vomiting may be present, occurring within a few hours of the onset of diarrhea. Temperature is usually normal or subnormal. Peak stool losses occur within 24 h of the onset of diarrhea and in cholera gravis (the most severe form of cholera) may exceed 1 l of fluid per hour, resulting in profound dehydration and circulatory collapse.

In epidemic settings there is no other disease that can consistently cause severe, life-threatening diarrhea and dehydration in adults. In individual cases, diarrhea approaching the severity of cholera can be caused by other diarrheal pathogens, including enterotoxigenic *Escherichia coli* and strains of non-O1 *V. cholerae*. Among children, rotavirus can also be a cause of severe, dehydrating diarrhea.

Treatment, Prevention

Treatment of cholera is based on replacement of fluids lost through diarrhea: this can almost always be accomplished by oral rehydration, although in the most severe cases intravenous rehydration may be necessary. Antibiotics such as tetracycline can shorten the duration of diarrhea and reduce the period of carriage. However, antibiotics should always be regarded as ancillary therapy to vigorous rehydration.

Prevention is based on the maintenance of good sanitation, provision of safe potable water, and avoidance of raw or undercooked foods in or from highly endemic or epidemic areas. Currently available parenteral vaccines are of low efficacy, and their use is not recommended. Several oral vaccines are under active development; their role in disease prevention remains to be determined.

'Nonepidemic' *Vibrio cholerae*

Strains of *V. cholerae* that do not carry virulence factors necessary to cause cholera have been implicated as a cause of diarrheal disease and, in

susceptible hosts, may cause septicemia. Epidemiologic studies, and studies with volunteers, have linked the occurrence of diarrheal illness with the production of a heat-stable enterotoxin (NAG-ST) similar to that produced by enterotoxigenic *E. coli*; diarrheal illness has also been linked with the production of cholera toxin or a cholera toxin-like toxin, and with the ability of a strain to colonize the intestine.

Occurrence in the Environment

As previously noted, nonepidemic strains of *V. cholerae* are ubiquitous in the estuarine environment: these organisms have been isolated from surface water at multiple sites in North America, Europe, Asia, and Australia, and it is likely that they are present in coastal and estuarine areas throughout the world. Nonepidemic strains have also been isolated from a variety of wild and domestic animals, including 6% of seagulls sampled in England; 14% of dogs in Calcutta, India; cows, goats, dogs, and chickens in India; ducks in Denmark; and horses, lambs, and bison in Colorado.

Entry to Food Chain

Nonepidemic *V. cholerae* can be isolated from a high percentage of oysters and other filter-feeding shellfish, particularly during warm summer months when counts of the organism are the highest in estuarine waters. Isolation is not associated with the presence of fecal coliforms. Results of a US Food and Drug Administration study of random oyster lots are shown in Table 1: overall, nonepidemic (non-O1) *V. cholerae* strains were isolated from 14% of 790 oyster lots tested, ranging from a high of 38.5% in July to a low of 0% in December and March. Not unexpectedly, epidemiologic studies in the USA have shown a significant association between non-O1/non-O139 *V. cholerae* diarrhea and eating raw oysters.

Table 1 Frequency distribution of oyster samples positive for *Vibrio cholerae* by month of harvest

Month of harvest	*No. of samples*	*No of samples with* V. cholerae[a]	
		O1	*Non-O1*
January	73	0 (0)	1 (1.4)
February	68	0 (0)	3 (4.4)
March	68	0 (0)	0 (0)
April	64	0 (0)	5 (7.8)
May	53	1 (1.9)	7 (13.2)
June	62	2 (3.2)	22 (35.5)
July	104	3 (3.2)	40 (38.5)
August	85	1 (1.2)	20 (23.5)
September	93	0 (0)	11 (11.8)
October	53	0 (0)	1 (1.9)
November	40	0 (0)	1 (2.8)
December	27	0 (0)	0 (0)

[a]Percentages are given within parentheses.

From Twedt RM, Madden JM, Hunt JM *et al*. (1981) Characterization of *Vibrio cholerae* isolated from oysters. *Applied Environmental Microbiology* 41: 1475–1478.

Outbreaks of non-O1/O139 *V. cholerae*-associated diarrhea have also been associated with other foods, including chopped egg (on an airplane flight to Australia), and preprepared gelatin (Mexico). In a study in the Ban Vinai refugee camp in Thailand, nonepidemic strains were isolated from a number of food items, including 4% of vegetables and 39% of meat samples. Routes by which the organism entered the food chain in these instances are unclear.

Fate on Processing

Characteristics of nonepidemic *V. cholerae* strains during processing have not been extensively investigated; preliminary studies suggest that they are similar to those described for epidemic strains.

Detection in Food, Water, and Physiological Samples

Isolation and identification techniques are identical for epidemic and nonepidemic strains. Strains shown by DNA probes or other genetic techniques to carry genes for CT or NAG-ST would appear to have the highest risk of causing illness. However, given our current state of knowledge, it is not possible to differentiate clearly nonepidemic *V. cholerae* which are potentially pathogenic from those which are nonpathogenic.

As noted below, nonepidemic *V. cholerae* can also cause wound infections and septicemia. Isolation is readily made from blood agar or other standard media used in culturing wounds; the organism grows well on standard blood culture media. Isolates from patients with septicemia tend to be more heavily encapsulated than general environmental isolates.

Symptoms and Characteristics

The most common manifestation of nonepidemic *V. cholerae* infection is diarrhea. Based on outbreak reports and volunteer studies, the incubation period appears to be short (< 24 h). Abdominal cramps may be prominent; bloody diarrhea is occasionally reported. Illness is usually mild and self-limited, although a diarrheal stool volume of 5.3 l was seen in one volunteer receiving 10^6 CFU of one nonepidemic strain.

Nonepidemic *V. cholerae* strains have also been isolated from persons with septicemia. Mortality rates in these cases exceed 50%. Illness appears to be confined to persons with underlying liver disease

or persons who are immunocompromised in some way. The route of entry of the organism in these cases is not well defined, although foods containing the organism would be a likely source.

Treatment/Prevention

As with epidemic *V. cholerae*, treatment of diarrhea is dependent on adequate rehydration. Septicemia requires aggressive antibiotic therapy and supportive care. Prevention in the developed world is based on the avoidance of raw or undercooked shellfish. In the developing world prevention is similar to that for other enteric pathogens, including avoidance of raw or undercooked foods of all types.

See also: **Contamination of Food**; **Diarrheal (Diarrhoeal) Diseases**; **Shellfish**: Contamination and Spoilage of Molluscs and Crustaceans; **Vibrios**: *Vibrio parahaemolyticus*; *Vibrio vulnificus*

Further Reading

Albert MJ, Neira M and Motarjemi Y (1997) The role of food in the epidemiology of cholera. *World Health Statistics Quarterly* 50: 111–118.

Blake PA, Allegra DT, Snyder JD *et al.* (1980) Cholera – a possible endemic focus in the United States. *New England Journal Medicine* 302: 305–309.

Franco AA, Fix AD, Prada A *et al.* (1997) Cholera in Lima, Peru, correlates with prior isolation of *Vibrio cholerae* from the environment. *American Journal of Epidemiology* 146: 1067–1075.

Karaolis DKR, Somara S, Maneval DR Jr, Johnson JA and Kaper JB (1999) A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. *Nature* 27: 375–379.

Morris JG, Jr (1990) Non-O group 1 *Vibrio cholerae*: a look at the epidemiology of an occasional pathogen. *Epidemiology Review* 12: 179–191.

Morris JG Jr, Takeda T, Tall BD, *et al.* (1990) Experimental non-O group 1 *Vibrio cholerae* gastroenteritis in humans. *Journal of Clinical Investigation* 85: 697–705.

Mujica OJ, Quick RE, Palacios AM *et al.* (1994) Epidemic cholera in the Amazon: the role of produce in disease risk and prevention. *Journal of Infectious Diseases* 169: 1381–1384.

Nair GB, Ramamurthy T, Bhattacharya SK *et al.* (1994) Spread of *Vibrio cholerae* O139 Bengal in India. *Journal of Infectious Diseases* 169: 1029–1034.

Pollitzer R (1959) *Cholera*. Geneva: World Health Organization.

St Louis ME, Porter JD, Helal A *et al.* (1990) Epidemic cholera in West Africa: the role of food handling and high-risk foods. *American Journal of Epidemiology* 131: 719–728.

Twedt RM, Madden JM, Hunt JM *et al.* (1981) Characterization of *Vibrio cholerae* isolated from oysters. *Applied Environmental Microbiology* 41: 1475–1478.

Wachsmuth IK, Blake PA and Olsvik O (eds) (1994) Vibrio cholerae *and Cholera*. Washington, DC: ASM Press.

Waldor MK and Mekalanos JJ (1996) Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 272: 1910–1914.

Vibrio parahaemolyticus

R Sakazaki, Japan Institute of Biological Sciences, Chiyodaku, Tokyo, Japan

Introduction

Vibrio parahaemolyticus, the causative organism of gastroenteritis associated with the consumption of raw or undercooked seafish, shellfish, and other contaminated foods, is one of the best described of pathogenic vibrios. Although several synonyms have been suggested for it, the name *V. parahaemolyticus* has been formally accepted.

Characteristics

V. parahaemolyticus is a short, slightly curved or straight, Gram-negative, facultatively anaerobic, motile rod. Sometimes it reveals polymorphism with coccoid cells, especially in older strains. In broth cultures, cells each exhibit a single polar, sheathed flagellum, but on the surface of solid media young cultures may also show unsheathed peritrichous flagella.

V. parahaemolyticus is a halophilic organism. It can grow on/in ordinary media containing 1–8% NaCl, but the most abundant growth occurs in the presence of 2–4% salt. It fails to grow on/in media lacking salt. It grows between 10 °C and 44 °C, but the optimal temperature for growth is between 30 °C and 37 °C. It fails to grow at 4 °C and lower temperatures not only arrest multiplication of the organism but also cause a rapid decrease in viable cell numbers. The vibrio can grow at a pH range from 5.6 to 9.6 but grows best at pH 7.6–8.6. Under optimal conditions, the generation time of the vibrio in the exponential phase of growth is estimated at 9–13 min.

Cultures of *V. parahaemolyticus* produce moist, circular colonies, opaque in appearance, attaining a size of 2–3 mm on ordinary agar media containing 2–3% salt after incubation at 37 °C for 24 h. Sometimes mucoid and viscid colonies may be seen. After several subcultures, however, dissociated translucent or rugose colony variants may also occur. Cultures of this vibrio, especially those isolated

from environmental sources, often show swarming on the surface of nutrient agar. In broth containing 2–4% salt, most strains produce heavy homologous turbidity, but rugose variants form a pellicle on the surface of the broth.

With few exceptions, *V. parahaemolyticus* produces both lysine and ornithine decarboxylases as well as indole. Arginine dihydrolase is not produced. It gives a negative reaction in the Voges–Proskauer test, and a positive reaction in the oxidase test. It ferments glucose, maltose, mannose, trehalose, and mannitol without gas production but fails to attack lactose, sucrose, xylose, and sorbitol. The majority of strains ferment arabinose, but occasional strains fail to do so. Urea is not usually decomposed, but some strains are able to hydrolyze it. It has been suggested that the ability to hydrolyze urea may be a marker for those Kanagawa-negative strains which are associated with illness but which nevertheless produce a thermolabile TDH-related hemolysin (TRH) related to, though different from, the direct thermostable direct hemolysin (TDH) of Kanagawa-positive strains.

The vibrio is susceptible to most antibiotics except penicillin, ampicillin, and polymyxin B, but resistant mutants do occur. It is also susceptible to the vibriostatic agent O/129 (2,4-diamino-6,7-diisopropyl pteridine) at a concentration of 150 μg ml^{-1} but resistant at 10 μg ml^{-1}.

Strains of *V. parahaemolyticus* may be typed with a combination of both somatic O and capsular K antigens, based on a serological scheme in which 13 O and 75 K antigens were established. The number of K antigens has since been extended to 84. Most strains of the vibrio are inagglutinable with O antisera in the living state, but after heating suspensions for 30 min at 100 °C they become O-agglutinable. H antigens are not included in the antigenic scheme, since in all strains they are regarded as serologically identical, although antigenically the polar flagellum and the peritrichous flagella are different.

Incidence in the Aquatic Environment

V. parahaemolyticus occurs in estuarine environments throughout the world. It also occurs occasionally in fresh-water sites, especially in brackish water during the summer, but only in small numbers. The halophilic nature of the vibrios is related to their ability to survive in the aquatic environment. They survive in estuarine water for more than 1 week at 15–20 °C, whereas they rapidly become extinct in fresh water under the same conditions as those of estuarine waters. The fate of the vibrios in fresh water is also partly concerned with osmotic pressure which causes autolysis of the cells. In brackish water, however, they may survive as long as in estuarine water in the summer. They are not found in the estuarine water in winter, although the vibrio can usually be isolated from sediment even when the water temperature is less than 10 °C. Indeed, it is likely that it survives in sediment in estuarine areas during the winter. The vibrio produces chitinase and the action of this enzyme enables it to adhere to the chitin of marine life. It thus colonizes zoo-plankton and the surface of shellfish. It therefore occurs in the digestive tracts of shellfish as a result of their ingestion of zooplankton. *V. parahaemolyticus* probably survives in a viable but nonculturable state in a cold environment, as reported for *V. cholerae* and *V. vulnificus*.

Reservoirs

Gastroenteritis due to *V. parahaemolyticus* is usually associated with seafoods. Coastal fish and shellfish are usually contaminated with *V. parahaemolyticus* in the summer. On the surface of market marine fish, the vibrios do not proliferate if kept below 10 °C, but at 20 °C their numbers increase rapidly. About 10–100 cells of *V. parahaemolyticus* have been found on the surface of coastal fish just after landing. If the fish were kept at atmospheric temperature, especially in summer, the number of vibrios increases to more than 10^6 cells within a few hours. It has not been found on market fish in winter.

Seafoods responsible for illness vary with local eating habits in different countries. In Japan, the high incidence of infection is undoubtedly due to the national custom of eating the meat of raw fish, shellfish, and other fish products. Any kind of seafish served as sushi and sashimi can act as the vehicle for the transmission of gastroenteritis. Although the vibrio may be killed within several minutes in 0.5% acetic acid, raw fish or shellfish with vinegar, which are widely eaten in Japan, nevertheless often transmit this infection. Also, infection is sometimes associated with consumption of cured vegetables cross-contaminated with the vibrio in the kitchen. The incidence of *V. parahaemolyticus* infection may be less in other countries where fish are not usually consumed uncooked. Nevertheless, cases of gastroenteritis caused by this organism have been reported in East European countries, the UK, Africa, and the USA. In the USA, for example, 40 outbreaks of infection were reported to the Centers for Disease Control and Prevention between 1973 and 1998. In countries other than Japan, although fish are usually cooked shortly before consumption, crab meat and shrimp, which are the seafoods most often implicated with this infection, are usually extracted either by hand or

mechanically after cooking, becoming contaminated from other sources. Oysters and clams which are consumed raw are also significantly associated with gastroenteritis caused by this vibrio. It is possible that *V. parahaemolyticus* may cause diarrheal diseases in developing countries more often than is currently recognized. In such countries, waterborne infection with this vibrio should also be considered.

Clinical Infection with *V. parahaemolyticus*

Although *V. parahaemolyticus* is associated with occasional skin and soft-tissue infections as well as with otitis media in fish handlers and bathers, gastroenteritis is the essential clinical manifestation of foodborne infection. Symptoms usually occur about 12 h after consumption of contaminated food. The outstanding features are severe abdominal cramps and diarrhea, together with nausea, vomiting, and fever. These symptoms range from mild with a few loose stools to very severe. In addition, dysentery-like stools with blood and mucus may occur in some cases. Recovery of patients and disappearance of the vibrio from stools are usually complete within a few days. The mortality rate is very low. Person-to-person transmission has not been recognized.

Although the size of the infectious dose necessary to produce clinical symptoms may vary with the strain and individual, according to the data from human volunteer experiments it, probably lies between about 10^5 and 10^8 viable cells. However, based on the low numbers of Kanagawa-positive cells usually found in samples of the food implicated or in waterborne infections encountered in developing countries, it appears that the numbers of viable vibrios needed to cause illness may be much smaller.

Virulence Factors

It has been found that strains of *V. parahaemolyticus* isolated from patients with gastroenteritis were hemolytic, on a modified brain–heart infusion agar containing human blood, whereas those from sea fish and the marine environment were nonhemolytic. This hemolysis is referred to as the Kanagawa reaction. It was found that 96% of 2720 strains from patients with gastroenteritis were Kanagawa-positive compared with only 1% of 650 environmental strains. Feeding tests with Kanagawa-negative strains in 15 adult volunteers failed to induce any clinical signs of illness. The hemolysin responsible for the Kanagawa reaction is known as TDH. Although *V. parahaemolyrticus* is known to produce at least four hemolytic substances – TDH, a thermolabile hemolysin, phospholipase A, and lysophospholipase – it is probable that only TDH plays a significant role in the pathogenesis of gastroenteritis.

Occasional outbreaks of gastroenteritis may be associated with Kanagawa-negative vibrios. These strains produce a thermolabile TRH but not the thermostable TDH.

TDH is a 21-kDa protein not affected by heating at 100 °C for 15 min at pH 6.0. It is a pore-forming toxin which expresses hemolytic activity, cytotoxicity, and cardiotoxicity. In hemolytic tests, TDH is strongly active against erythrocytes of dogs, rats, mice, and humans, weakly active against those of rabbits and sheep, and inactive against horse erythrocytes. Although it is clear that TDH is associated with the pathogenesis of infection, it is not clear how this toxin causes diarrhea. TDH, when inoculated at doses of 100 μg into ligated ileal loops in rabbits, failed to produce fluid accumulation but caused erosive lesions and necrosis of the intestinal mucosa. Such histological changes do not occur when whole bacteria are tested in ileal loops.

It has been reported that a TDH-positive parent vibrio strain caused fluid accumulation in ligated ileal loops in rabbits, whereas a TDH-negative mutant did not cause fluid accumulation. Similar results were obtained with culture supernatants of TDH-positive strains tested on rabbit ileal tissues mounted in an Ussing chamber, which represents a more sensitive measure of secretory activity. In these assays, the ability of TDH to alter iron transport in the intestinal tract was demonstrated at nanogram levels, with no histological changes. TDH induces intestinal chloride ion secretion and trisialoganglioside G_{T1b} appears to be the cellular receptor. However, recent work suggests that other unknown receptor(s) may be present. TDH uses Ca^{2+} as an intracellular second messenger, and is thus the first bacterial enterotoxin for which a linkage between changes in intracellular calcium and secretory activity has been established.

TDH is encoded by two copies, *tdh1* and *tdh2*, of the *tdh* gene in *V. parahaemolyticus*. The two gene copies are not identical and the predicted protein products differ in seven amino acid residues, although the proteins themselves are immunologically indistinguishable. The level of TDH production may be under the control of a regulator similar to that of ToxR in *V. cholerae*.

Although TRH and TDH are similar in their biological, immunological, and physicochemical properties, TRH is thermolabile and differs from TDH in its activity on erythrocytes. TRH is encoded by the *trh* gene which shares 69% identitiy with the *tdh2* gene. TRH is linked epidemiologically to gastroenteritis, but the secretory mechanism of this toxin is still uncertain.

Several adhesive factors, including fimbriae, peritrichous flagella, outer membrane proteins, and a mannose-resistant, cell-associated hemagglutinin, have also been proposed, but no substantial studies have yet implicated any of the candidate adhesins in the pathogenicity of the organism.

Isolation and Identification

Several selective agar media have been devised, but thiosulfate citrate bile salts sucrose (TCBS) agar is recommended for the isolation of *V. parahaemolyticus*. In clinical microbiology laboratories, MacConkey agar containing 0.5% additional NaCl is also convenient for routine culture of diarrheal stools.

Enrichment culture is used for the detection of the vibrio from food and marine samples. Polymyxin salt broth, containing 2% NaCl and 50 $\mu g\,ml^{-1}$ of polymyxin B (pH 7.6) may be used for the selective growth of *V. parahaemolyticus*. It should be noted that some factor(s) in shellfish may inhibit the growth of vibrios. It is therefore recommended that shellfish are cut into small pieces, but not homogenized. After shaking the enrichment broth vigorously, the pieces of shellfish are removed with forceps. However, enrichment culture of seafoods incriminated in outbreaks of infection may be unrewarding, because most yield Kanagawa-negative isolates, in contrast to those from patients.

The colonial appearance of *V. parahaemolyticus* on TCBS agar is so typical that provisional identification of the isolates from stool specimens may be made directly from the plates. However, isolates from marine sources must be further examined in order to differentiate them from related organisms. The addition of 1% NaCl to medium for biochemical tests is essential to obtain valid reactions.

The Kanagawa reaction is a reliable test for the recognition of virulent strains. For determination, isolates should if possible be tested using Wagatsuma agar, which contains 0.5% yeast extract, 1% peptone, 0.5% mannitol, 0.05% K_2HPO_4, 7% NaCl, 0.0001% crystal violet, 1.5% agar, and washed human red blood cells. However, as adequate sources of human blood, or of blood from other suitable animals, are not always readily available, other suitable media may be considered. Thus, an enzyme-linked immunosorbent assay has been described for the detection of TDH-producing vibrios.

Several molecular approaches for the detection of Kanagawa-positive vibrios have also been developed. DNA and oligonucleotide probes specific for the genes *tdh* and *trh* have been described. However, the probes also hybridize with *tdh* genes in some strains of non-O1 *V. cholerae*, *V. hollisae*, and *V. mimicus*. A polymerase chain reaction technique has been reported using oligonucleotide primers derived from the nucleotide sequence of the *tdh* gene.

Serotyping of isolates of *V. parahaemolyticus* may be performed by slide agglutination tests using O and K antisera. In outbreaks of *V. parahaemolyticus* infection, however, the same serovar as that identified in patients is seldom detected in incriminated seafoods. Serotyping of isolates from seafoods and marine sources is thus not usually significant unless they are Kanagawa-positive.

See also: **Fish**: Spoilage of Seafood; **Shellfish**: Contamination and Spoilage of Molluscs and Crustaceans; **Vibrios**: *Vibrio cholerae*

Further Reading

Asakawa Y, Akahane S and Noguchi M (1974) Quantitative studies on pollution with *Vibrio parahaemolyticus* during distribution of fish. In: Fujino T *et al.* (eds) *International Symposium on Vibrio parahaemolyticus* pp. 97–103. Tokyo: Saikon Publishing Co.

Daniels NA, MacKinnon L, Bishop R *et al.* (2000) *Vibrio parahaemolyticus* infections in the United tates, 1973–1998. *J Infect Dis* 181: 1661–1666.

Honda T, Ni Y, Yoh M *et al.* (1989) Production of monoclonal antibodies against thermostable direct hemolysin of *Vibrio parahaemolyticus* and application of the antibodies for enzyme-linked immunosorbent assay. *Med Microbiol Immunol* 178: 245–253.

Kato T, Obara Y, Ichinoe H *et al.* (1965) Grouping of *Vibrio parahaemolyticus* (biotype 1) by hemolytic reaction. *Shokuhin Eisei Kenkyu* 15: 83–86 (in Japanese).

Lee C and Pan S-F (1993) Rapid and specific detection of the thermostable direct hemolysin gene in *Vibrio parahaemolyticus* by polymerase chain reaction. *J Gen Microbiol* 139: 3225–3231.

Nishibuchi M, Ishibashi M, Takeda Y *et al.* (1985) Detection of the thermostable direct hemolysin gene and related DNA sequences in DNA colony hybrid-ization test. Infect. *Immun* 49: 481–486.

Nishibuchi M, Hill WE, Zon G *et al.* (1986) Synthetic oligodeoxyribonucleotide probes to detect Kanagawa phenomenon-positive *Vibrio parahaemo-lyticus*. *J Clin Microbiol* 23: 1091–1093.

Nishibuchi M, Taniguchi T, Misawa T *et al.* (1989) Cloning and nucleotide sequence of the gene (*trh*) encoding the hemolysin related to the thermostable direct hemolysin of *Vibrio parahaemolyticus*. Infect. *Immun*. 57: 2691–2697.

Osawa R, Okitsu T, Morozumi H *et al.* (1996) Occurrence of urease-positive *Vibrio parahaemo-lyticus* in Kanagawa, Japan, with special reference to presence of thermostable direct hemolysin (TDH) and the TDH-related hemolysin genes. *Appl Environ Microbiol* 62: 725–727.

Sakazaki R, Iwanami S and Fukumi H (1963) Studies on the enteropathogenic, facultatively halophilic bacteria,

Vibrio parahaemolyticus I. Morphological, cultural and biochemical properties and its taxonomical position. *Jpn J Med Sci Biol* 16: 161–188.

Sakazaki R, Iwanami S and Tamura K (1968a) Studies on the enteropathogenic, facultatively halophilic bacteria, *Vibrio parahaemolyticus*. II. Serological characteristics. *Jpn J Med Sci Biol* 21: 313–324.

Shinoda S, Miwatani T, Honda T *et al*. (1974) Antigenicity of flagella of *Vibrio parahaemolyticus*. In: Fujino T *et al*. (ed.) *International Symposium on Vibrio parahaemolyticus* pp. 193–197. Tokyo: Saikon Publishing Co., Tokeo.

Sanyal SC and Sen PC (1974) Human volunteer study on the pathogenicity of *Vibrio parahaemolyticus* In: Fujino T *et al*. (eds) *International Symposium of Vibrio parahaemo-lyticus* pp. 227–235. Tokyo: Saikon Publishing Co.

Takikawa I (1958) Studies on pathogenic halophilic bacteria. *Yokohama Med Bull* 2: 313–322.

Yamamoto K, Honda T and Miwatani T (1992) Enzyme-labeled oligonucleotide probes for detection of the genes for thermostable direct hemolysin (TDH) and TDH-related hemolysin (TRH) of *Vibrio parahaemolyticus*. *Canad J Microbiol* 38: 410–416.

Vibrio vulnificus

A C Wright, University of Florida, Gainesville, FL, USA
J G Morris, Jr., University of Maryland School of Medicine, Baltimore, MD, USA

Background

Vibrio vulnificus, first recognized as a distinct species in the late 1970s, is a Gram-negative bacterium that is indigenous to the estuarine environment. Virtually all oysters harvested in the USA during warmer, summer months harbor this organism, and it has been reported in association with shellfish and fish in a number of locations in South America, Asia, and Europe. For persons in 'high risk' groups (i.e., persons with liver disease, iron overload states, or immunosuppression), ingestion of *V. vulnificus* in raw or undercooked seafood can result in a syndrome of 'primary septicemia,' characterized by intractable shock and a mortality rate of greater than 50%. These infections are the leading cause of death associated with consumption of shellfish in the USA. The microorganism can also cause serious wound infections and may be a cause of gastroenteritis. In population-based studies in US coastal areas, the incidence of *V. vulnificus* infections is approximately 0.5/100 000 persons per year. Case series from Korea and Taiwan suggest that *V. vulnificus* infections are even more common in these latter areas.

Virulence of *V. vulnificus* is most closely linked with the presence of a polysaccharide capsule and is greatly influenced by availability of iron in the host during infection. *V. vulnificus* is also known to produce a variety of extracellular factors, including a cytolysin, siderophores, pili, lipopolysaccharide, proteases, and other exoenzymes; however, the role of these factors in virulence remains unclear. Animal studies of the lethality of this organism have indicated that mutants deficient in cytolysin and protease do not exhibit reduced virulence. Mutants deficient in siderophores or pili expression are somewhat less virulent than wild-type strains, but these results are complicated by the pleiotropic nature of these mutations. However, excess iron in the host and expression of capsular polysaccharide (CPS) by *V. vulnificus* clearly increase the lethality of this disease. For mice with normal iron status, the LD50 ranges from 10^5 to 10^7, depending on the strain and the ability to produce CPS. The LD50 for unencapsulated mutants in iron-overloaded mice is also about 10^5; however, encapsulated strains in iron-overloaded mice may be lethal at a dose of one bacterium.

There is tremendous diversity in capsular types among strains, and no one type (or group of types) predominates among clinical or environmental isolates. Three biotypes (or two biotypes and one distinct serovar) of *V. vulnificus* are currently recognized. The majority of clinical and environmental *V. vulnificus* isolates reported to date are in biotype 1. Strains initially classified as biotype 2 are responsible for sepsis in eels; they do not cause human disease. More recently, it has been recognized that the biotyping characteristics described for these strains lacked specificity; however, the eel pathogens are homogeneous in their lipopolysaccharide-based serogroup, leading to their reclassification as serovar E. Biotype 3 strains have been described in association with wound infections related to handling of live fish (tilapia) from fish farms in Israel.

Occurrence in the Environment

As noted above, *V. vulnificus* is commonly isolated from water and organisms, particularly shellfish, collected from estuarine environments. The highest numbers are found in areas with intermediate salinities (5–25 p.p.t.) and warmer temperatures (optimally, $>20\,^{\circ}C$). Isolation has been reported from the US Atlantic, Gulf, and Pacific coasts, as well as the Atlantic coast of Europe, the Mediterranean, the Indian Ocean, Malaysia, and the Pacific Coast of Asia. It is likely that the organism is present in

virtually all estuarine areas with appropriate salinity and temperature ranges. Using a DNA probe, *V. vulnificus* has been identified in 80% of Chesapeake Bay water samples collected during months in which water temperatures exceeded 8°C: concentrations ranged from 3.0×10^1 to 2.1×10^2 CFU^{-1} ml^{-1}, representing approximately 8% of the total culturable heterotrophic bacteria. However, none of the samples collected during February and March, when temperatures were less than 8°C, were positive for this organism.

V. vulnificus is generally not recovered from the environment during colder months, and several lines of evidence suggest that it may enter a viable but nonculturable state, where recovery on typical solid media is not possible. Other studies have suggested that the majority of *V. vulnificus* cells die off with the stress of cold temperature or starvation, and only a small resistant subpopulation remains viable. However, both experimental and epidemiological data strongly indicate that nonrecoverable forms probably are not virulent, as little or no disease is observed in colder months. Thus, the relevance of this issue to the seafood industry may be negligible. However, further study may shed light on how the species is maintained in the environment and help elucidate survival mechanisms that could be targets for control measures.

Entry to the Food Chain

Epidemiologically, there is a very strong association between *V. vulnificus* infection and eating raw oysters, particularly those harvested from warmer waters. It should be noted that most, if not all, serious disease is associated with the raw, freshly shucked product; proper cooking eliminates the disease risk. In Chesapeake Bay studies, *V. vulnificus* was found in virtually all oysters collected during warmer months (water temperature 7.6°C or above), with the numbers of this species approximately two orders of magnitude above those found in the surrounding water. In studies conducted by the US Food and Drug Administration, *V. vulnificus* was detected in virtually all oysters from the US Gulf Coast. The numbers varied based on temperature and salinity; representative data from this study are shown in Figure 1. The prevalence of this organism in the environment correlates well with disease incidence, as approximately 90% of the infections occur between April and October.

Unlike other pathogens associated with seafood consumption, several studies have shown no correlation between the presence of *V. vulnificus* and fecal coliforms. *V. vulnificus* is clearly an indigenous estuarine species with an environmental reservoir;

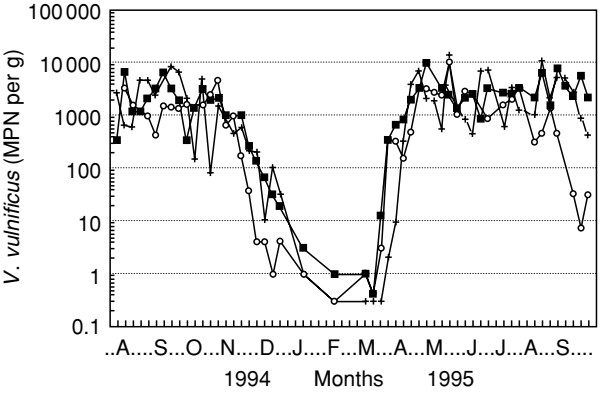

Figure 1 Weekly densities of *V. vulnificus* in oysters from Gulf Coast sites in Alabama (+), Florida (○), and Louisiana (■) during a 14-month period. Each point represents the geometric mean ($n = 2$) of the MPN of bacteria per gram of oyster meat. From Motes ML, DePaola A, Cook DW *et al*. (1998) Influence of water temperature and salinity on *Vibrio vulnificus* in northern Gulf and Atlantic Coast oysters. *Applied and Environmental Microbiology* 64: 1459–1465, with permission.

however, many aspects of this reservoir are still unclear. For example, does this organism exist and replicate as a free-living species or in association with algae, shellfish, or fish? While it may be assumed that concentrations of *V. vulnificus* in shellfish, in part, reflect filtering of the organism from surrounding water, there are data suggesting that *V. vulnificus* can assume a symbiotic relationship with oysters. At a practical level, this relationship implies specific attachment or internalization, which may account for the difficulties in clearing the organism from oysters by using traditional depuration techniques.

The condition and overall health of the oyster may also influence the numbers of *V. vulnificus* in oyster tissue. Numerous variables such as heavy metal concentrations, water turbidity, availability and types of food sources, spawning status, etc. may influence the ability of oysters to clear or control endogenous bacterial populations. Bacterial growth also may be enhanced by coinfection with parasites such as *Perkinsus marinus*, the presence of which may depress innate defense mechanisms.

Fate on Processing/Storage and Control Measures

V. vulnificus can persist in oyster shellstock for extended periods of time. In oysters held at 18°C for 30 h, or at ambient temperatures for 12 h, there are significant increases in *V. vulnificus* counts, highlighting the importance of refrigeration after harvest and during storage and transport. Currently in the USA, Hazard Analysis Critical Control Point

recommendations are being implemented that provide more rapid postharvest refrigeration of shellfish to limit increases in the numbers of *V. vulnificus* resulting from temperature abuse.

A variety of approaches have been proposed to reduce or eliminate *V. vulnificus* from oyster shellstock. Traditionally, depuration or the incubation of oysters in recirculating, sanitized fresh water has been used to reduce bacterial contamination of fecal origin in shellstock; however, several studies found this treatment to be ineffective for *V. vulnificus* control. In one study, *V. vulnificus* was cleared from the digestive organs but persisted in other tissues following depuration treatment. Recently, alternative treatments such as freezing, cold or hot water immersion, relaying to high salinity waters, high pressure, and irradiation have shown promising results for the reduction and possible elimination of this species from oysters. Unfortunately, many of these treatments are lethal to oysters as well, and their application to the 'oyster on the half shell' market may be limited. Some suppliers are currently using these methods to produce a product that greatly resembles a freshly shucked oyster but will not be sold as live product.

Detection in Food, Water, and Physiological Samples

V. vulnificus has traditionally been evaluated in food and water samples using most probable number (MPN) calculations and standard microbiologic techniques. In this procedure, oysters are homogenized in buffer and serially diluted to MPN enrichment tubes, which are subsequently screened for vibrios by growth on selective/differential media. Unfortunately, owing to the diversity and abundance of vibrios in the environment, this method lacks reliability, and further confirmatory methods such as species-specific monoclonal antibody are generally required for identification.

More recently, there has been increasing use of molecular approaches, including DNA probes and PCR techniques. Several of the probes and PCR primers are based upon the DNA sequence for the *V. vulnificus* cytolysin gene, *vvh*. This gene appears to be common to most, if not all, strains and is highly species-specific. Applications include a colony hybridization protocol that can be combined with total viable counts on nonselective medium to provide more direct enumeration. Multiplex PCR analyses have also been described that simultaneously detect *V. vulnificus* and other *Vibrio* spp. Unfortunately, these probes, as well as the standard MPN methods described, do not discriminate between potentially virulent and avirulent strains.

For patient samples, *V. vulnificus* will grow without difficulty in standard blood culture media or on nonselective media (such as blood agar) routinely used for wound cultures; identification and speciation of the organism are possible through any of the standard, commercially available microbiology identification systems. As with all *Vibrio* species, isolation of the organism from stool generally requires the use of a specific selective culture media (thiosulfate citrate bile-salts sucrose).

Symptoms and Characteristics

V. vulnificus has been linked with three different clinical syndromes: (1) 'primary septicemia' (presence of the organism in the blood, without an obvious source such as a wound infection), (2) wound infections, and, possibly, (3) gastroenteritis. Serious *V. vulnificus* infections occur almost exclusively in persons who have underlying liver disease (including alcoholic liver disease), are immunosuppressed, or have chronic diseases such as diabetes or renal failure. Risk of infection is also substantively increased in persons with increased iron stores, as found in patients with hemochromatosis.

Primary Septicemia

Approximately one-third of patients with primary *V. vulnificus* septicemia present in shock or become hypotensive within 12 h of hospital admission. Three-fourths of patients have distinctive bullous skin lesions. Thrombocytopenia is common, and there is often evidence of disseminated intravascular coagulation. Complications such as gastrointestinal bleeding are not infrequent. The mortality rate for persons with primary *V. vulnificus* septicemia is more than 50% overall and more than 90% in persons who are hypotensive. Although the actual infection may be cleared rapidly, patients often require prolonged hospitalization in intensive care units because of the associated shock syndrome and resultant multiorgan system failure.

Wound Infections

V. vulnificus may contaminate wounds exposed to estuarine waters or shellfish. In persons in the risk groups noted above, the infection may spread rapidly, producing severe myositis and fasciitis reminiscent of gas gangrene and requiring amputation in extreme cases.

Gastroenteritis

V. vulnificus has also been associated with gastroenteritis. However, an etiologic role is difficult to

establish, as isolation from stool samples is of uncertain significance, given the ubiquitous presence of the organism in both water and shellfish.

Treatment and Prevention

In *V. vulnificus* septicemia, the sooner antibiotic therapy is initiated, the greater the chance that the patient will survive. Based upon clinical observations and *in vitro* susceptibilities, tetracycline and quinolone antibiotics have been recommended for management of *V. vulnificus* infections. Recent *in vitro* and animal studies from Taiwan indicate that there is synergism between minocycline and cefotaxime in treatment of serious *V. vulnificus* infections, leading to the recommendation that these latter two drugs be used as 'first line' therapy. Patients with serious infections need to be aggressively managed in an intensive care unit setting to minimize the possible consequences of hypotension, septic shock, and multiorgan system failure.

Given the high mortality associated with *V. vulnificus* infections, persons in the high-risk groups noted above should avoid eating raw or undercooked shellfish, particularly oysters. By law, restaurants in many states are now required to post warnings about this risk. Persons at high risk for infection should also avoid situations in which estuarine-associated wounds are likely to occur.

See also: **Fish**: Spoilage of Seafood; **Shellfish**: Contamination and Spoilage of Molluscs and Crustaceans

Further Reading

Bisharat N, Agmon V, Finkelstein R *et al*. (1999) Clinical, epidemiological, and microbiological features of *Vibrio vulnificus* biogroup 3 causing outbreaks of wound infection and bacteremia in Israel. *Lancet* 354: 1421–1424.

Blake PA, Merson MH, Weaver RE, Hollis DG and Heublein PC (1979) Disease caused by a marine Vibrio: clinical characteristics and epidemiology. *New England Journal of Medicine* 300: 1–5.

Brasher CS, Depaola A, Jones DD and Bej AK (1998) Detection of microbial pathogens in shellfish with multiplex PCR. *Current Microbiology* 37: 101–107.

Bush CA, Patel P, Gunawardena S *et al*. (1997) Classification of *Vibrio vulnificus* strains by the carbohydrate composition of their capsular polysaccharides. *Analytical Biochemistry* 250: 186–195.

Chuang YC, Ko WC, Wang ST *et al*. (1998) Minocycline and cefotaxime in the treatment of experimental murine *Vibrio vulnificus* infection. *Antimicrobial Agents and Chemotherapy* 42: 1319–1322.

Cook DW (1997) Refrigeration of oyster shellstock: conditions which minimize the outgrowth of *Vibrio vulnificus*. *Journal of Food Protection* 60: 349–352.

Dixon DW and Rodrick GE (1998) Effect of gamma radiation on shellstock oysters: extension of shelf-life and reduction in bacterial numbers, with particular reference to *Vibrio vulnificus*. In: *Combination Processes for Food Irradiation*, pp. 97–110. Vienna: IAEA.

Elliot EL, Kaysner CA, Jackson L and Tamplin ML (1998) *Vibrio cholerae, V. parahaemolyticus, V. vulnificus* and other *Vibrio* spp. In: *U.S. Food and Drug Administration Bacteriological Analytical Manual*, 8th edn. Gaithersburg, MD: AOAC International.

ISSC (2000) Gulf Oyster Industry Council, FL. Dept of Agriculture, and FDA Office of Seafood Processing controls for vibrios in raw oysters. Online abstracts of the 25th Annual Meeting of Seafood Science and Technology Society, Longboat Key, FL http://www.seafood.ucdavis.edu/organize/SSTSA.htm

Motes ML and DePaola A (1996) Offshore suspension relaying to reduce levels of *Vibrio vulnificus* in oysters (*Crassostrea virginica*). *Applied and Environmental Microbiology* 62: 3875–3877.

Motes ML, DePaola A, Cook DW *et al*. (1998) Influence of water temperature and salinity on *Vibrio vulnificus* in northern Gulf and Atlantic Coast oysters. *Applied and Environmental Microbiology* 64: 1459–1465.

Park SD, Shon HS and Joh NJ (1991) *Vibrio vulnificus* septicemia in Korea: Clincal and epidemiologic findings in seventy patients. *Journal of the American Academy of Dermatology* 24: 397–403.

Strom MS and Paranjpye RN (2000) Epidemiology and pathogenesis of *Vibrio vulnificus*. *Microbes and Infection* 2: 177–188.

Tamplin ML and Capers GM (1991) Persistence of *Vibrio vulnificus* in tissues of Gulf coast oysters (*Crassostrea virginica*) exposed to seawater disinfected with UV light. *Applied and Environmental Microbiology* 58: 1506–1510.

Wright AC, Hill RT, Johnson JA, Roghman M-C, Colwell RR and Morris JG Jr. (1996) Distribution of *Vibrio vulnificus* in the Chesapeake Bay. *Applied and Environmental Microbiology* 62: 717–724.

VINEGAR

M Plessi, Dipartimento di Scienze Farmaceutiche, Università degli Studi di Modena, Modena, Italy

Summary and Background

Vinegar is traditionally the product of the acetous fermentation of dilute alcoholic solutions. At the present time it is produced microbiologically from natural alcoholic solutions or by dilution of acetic acid. Vinegar and its characterization are presented in this article.

Vinegar has been known since ancient times, for more than 10 000 years, and has been used for its typically acid flavor. It uses were many and various for centuries, not only for food but also for medicinal and ritual uses. It is traditionally the product of the acetic fermentation of dilute alcoholic liquors and it was wine that was the first alcoholic liquid used, thus the derivation of its name (from the French *vin aigre* = sour wine).

History

During Babylonian times vinegar was used as a condiment and for food preservation. Hippocrates recognized its medicinal properties and it is also mentioned in the Bible as a remedy on account of its sedative and curative properties. Pliny reported that vinegar, diluted with water (known as posca), was given to the Roman legionnaires during their long marches.

The medicinal use of vinegar was widespread during the Middle Ages and the Renaissance, for both internal and topical use. It was in fact used as a digestive, a prophylactic against liver disorders, an anthelmintic, for sore throats, and to rub on the wrists against fever, but also against hair loss and tinea.

Vinegar production as an industry began during the period of the Communes; those who swore not to reveal the secret of its fabrication were admitted to the corporation of manufacturers. Although widely used, little was known about the real nature of vinegar and its fermentation until the eighteenth century. In 1723, Stahl obtained acetic acid from vinegar, but nothing was known of the causes which determined the formation of acetic acid. The first breakthrough came with Berzelius, who, at the beginning of the nineteenth century, explained how acetic acid was formed from alcohol by a process of oxidation. Persoon (1822) and Kùtzing (1837) were the first to suggest the presence of microorganisms in the process. Finally, in 1864, Pasteur brilliantly demonstrated the biological origin of the acetic fermentation, indicating *Mycoderma aceti* as the fermentation agent. Later studies clarified the nature of the microorganisms, which turned out to belong to several species.

Vinegar is now used above all to give foodstuffs a pleasant acidic flavor; it is also used as a preserver of foodstuffs and is recognized to have certain pharmacological qualities.

Production

At the present time, vinegar is still produced microbiologically from a mostly natural alcoholic solution, but can also be produced by dilution of acetic acid.

Microbiological Production

Vinegar is the product of the acetic fermentation of slightly alcoholic liquids (less than 10–12% by volume of ethyl alcohol); transformation of alcoholic liquids into vinegar is not really fermentation, but oxidation.

Raw materials may be wine, cider, beer, and other liquors deriving from the alcoholic fermentation of cereals, fruit and potatoes, sugar solutions, such as molasses, honey, and whey, and also diluted pure ethanol with the addition of nutrients. These are the so-called brewed vinegars, derived from the oxidation activity of aerobic microorganisms. The microorganisms belong to the genus *Acetobacter* (initially called *Mycoderma*), and the most widespread species are *A. aceti*, *A. pasteurianus*, and *A. hansenii*. These are the ones that oxidize ethyl alcohol into acetic acid. The oxygen used for the bioxidation is that of air.

These microorganisms, in the shape of asporigenous rods or cocci, form membranes of different consistencies (vinegar mother). The optimum temperature for their multiplication is between 18 and 34 °C, depending on the species.

Ethanol is dehydrogenated to acetic acid and the reduced cosubstrates are oxidized via the respiratory chain (**Figure 1**). Similarly, but less abundantly and in an anaerobic atmosphere, acetic acid is formed by dismutation of 2 molecules of acetaldehyde (derived in their turn from alcohol by oxidation). In theory, 1 g of alcohol yields 1.3 g of acetic acid; in practice, however, the yield is 15–20%, lower mainly because alcohol, acetaldehyde, and acetic acid tend to volatilize. The theoretical amount of air required for 1 l of vinegar containing 6% of acetic acid is about 120 l,

whereas, in practice, given the slow rate of liquid–gas exchange, the amount required is much greater.

Acetous fermentation is accompanied by secondary fermentation which combines to produce the flavor and typical aroma; small quantities of volatile substances are formed (e.g., ethane, acetaldehyde, ethyl formate, ethyl acetate, isopentyl acetate, butanol, methylbutanol, 3-hydroxi-2 butanone or acetylmethylcarbinol), which vary from vinegar to vinegar depending on the starting material and which, because of their individual characteristics, produce vinegars with a variety of odor, taste, color, and other properties. The fermentation is usually stopped at a minimum residual ethanol level to avoid overoxidation, as oxidation of acetic acid to water and CO_2. (*See* **Fermented Foods: Origins and Applications.**)

Methods of manufacture The manufacture of vinegar is an ancient craft, known and practiced all over the world from time immemorial. A certain amount of vinegar is still manufactured following the centuries-old empirical methods of the small producer, but since the last century a flourishing vinegar-manufacturing industry has developed.

Industrial vinegar-manufacturing processes fall into three main categories. The slow processes, which are the oldest commercial procedure and resemble home-brewing techniques, are no longer used. The Orleans process is still in use for the production of high-quality vinegar; the procedure has remained unchanged through the years but is very slow and requires a great deal of space (Figure 2). The starting liquor is placed in a large cask, containing wood shavings or grape stalks, where the acetification process gets underway. After about 8 days, the liquid is withdrawn and transferred into barrels, which are left only one-half or two-thirds full and where the liquid remains until acidity reaches its peak (about 3 months). From now on, two-thirds or three-quarters of the vinegar is withdrawn from the bottom of the barrel every week and an equal volume of liquid from the generator cask is added from the top. (*See* **Barrels: Wines, Spirits, and Other Beverages.**)

The processes of the second and third categories aim at a closer contact, with maximum possible

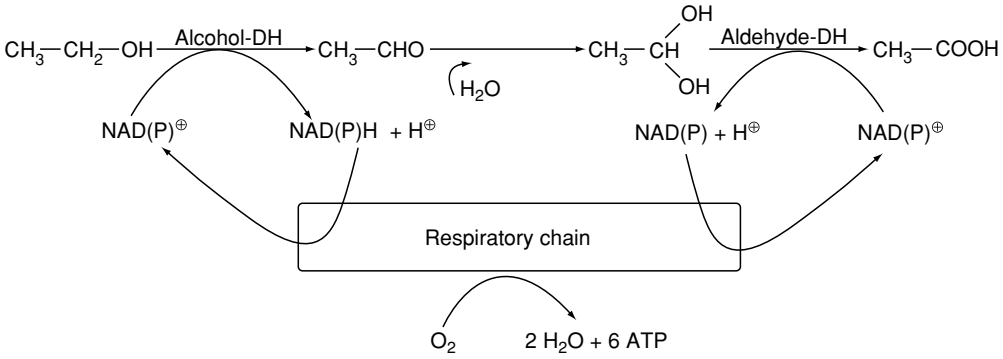

Figure 1 Ethanol oxidation to acetic acid by *Acetobacter* spp.

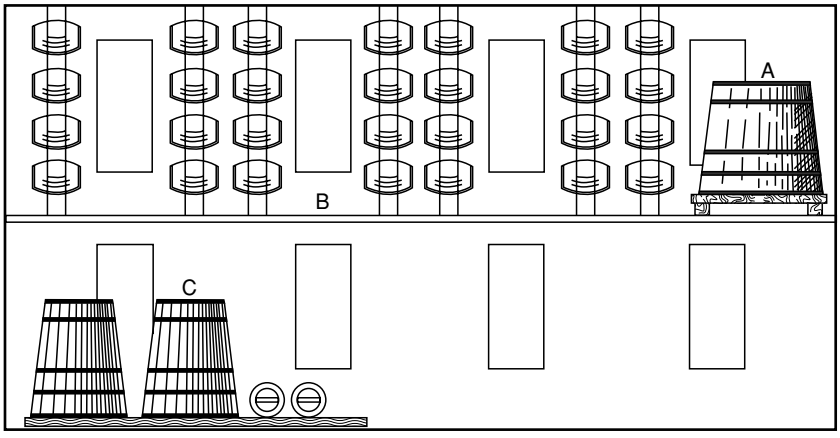

Figure 2 Vinegar making Orleans method: A, starting vat; B, acetifier casks; C, vats of clarification.

surface exposure, between the alcoholic liquid undergoing acetification and the ambient air. The passage of oxygen from the gaseous to the liquid phase is thereby accelerated and intensified and the acetification time, which with traditional methods is rather long, is reduced.

Generator processes, first introduced in Germany, have been in general use for more than 100 years. The generator tanks are generally made of wood or sometimes steel, are equipped with cooling coils, and are vented to allow the air to circulate; they have a perforated false bottom which serves to support the wood shavings (preferably beech) or grape stalks with which the tanks are filled. A spray mechanism distributes the alcoholic liquid over the surface of the packing (Figure 3). The liquid trickles over the wood shavings covered with *Acetobacters* and then is pumped back to the sprinkler header-tank and the cycle is repeated. The process continues until acetification is complete, which generally takes about a week providing the temperature is kept in the optimal range of 27–30 °C. A measured volume of vinegar is then withdrawn from the bottom of the tank and is replaced with an equal volume of fresh liquid. The generator process is a continuous process which gives a fairly clear vinegar with good sensory characteristics. However, it is slow and about 20% of yield is lost through evaporation. Also, the wood shavings have to be replaced at least once a year.

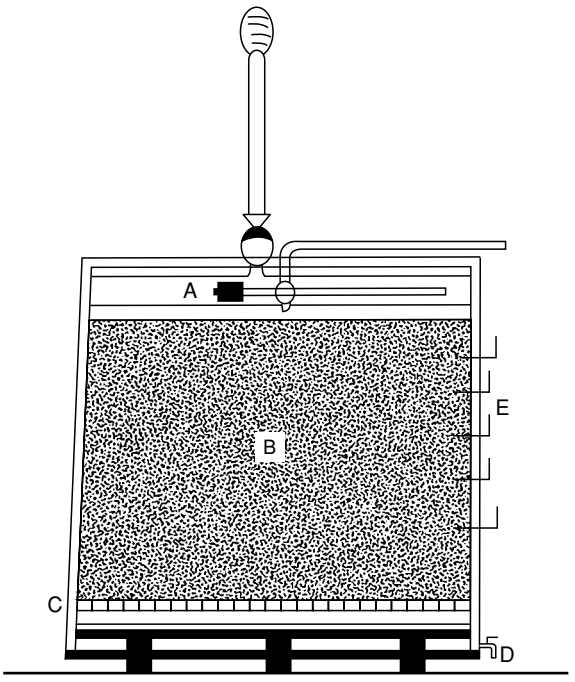

Figure 3 Circulating generator: A, sprinkler for dispersal of the liquid over the wood shavings; B, beechwood shaving; C, perforated false bottom; D, tap for withdrawal of the liquid; E, thermometers.

The third category involves the submerged processes. These are the most up-to-date, benefitting as they do from industrial experience with submerged culture techniques employed in the field of antibiotics. These processes do not use wood shavings, but air forced into the liquid, and the mother of vinegar is often replaced with selected cultures of *Acetobacters*. The generator tank is made of stainless steel, or polypropylene reinforced with fiberglass; it is equipped with devices to insure an adequate and continuous flow of air, with cooling coils which maintain a steady temperature of about 30 °C in the fermenting body of liquor, with thermometers for monitoring temperature and, sometimes, with an automatic gage for measuring the alcohol content of the fermenting liquid. Oxidation of the alcohol is started slowly by means of aeration, and fermentation is apparent after about 24 h. Air is then introduced at regular hourly intervals; it permeates the body of liquor in a uniform manner, causing rapid and lively fermentation. Acetification is deemed complete when the residual alcohol is around 0.2–1.5% w/v (depending on the type of vinegar required). The process is extremely rapid. About half of the resultant vinegar is then withdrawn; the remaining vinegar acts as mother for the next batch.

This method was first developed in the USA and Germany. It spread rapidly throughout the world and displaced the older, slower processes. The yield is very high (90–95%), since there is little loss through evaporation. However, the product is rather cloudy and less aromatic than vinegars obtained by slower fermentation processes. This is due to the fact that, given the brief period of contact, the esterases do not have time to perform their function adequately and the characteristic volatile substance content is low as a result.

The product is therefore filtered and, in some cases, put into wooden casks to age; the result is a vinegar of superior quality which is rich in biological principles, generated by bacterial-cell autolysis, and perfectly clear.

Malt vinegar Malt vinegar, which is the most common variety in UK, is produced without intermediate distillation from the double fermentation (alcoholic and acetous) of malted barley with or without the addition of other cereals, the starch of which has been converted into fermentable sugars from α-amylase of malted barley.

Malt vinegar manufacture involves mashing, fermentation, and acetification. During mashing, the malted barley, sometimes mixed with other cereals such as maize and rice, is milled and mixed with hot water in mash tuns, where the starch is converted by α-amylase into maltose, dextrose, and dextrins. The sweet liquor drains off the mash through the

perforated false bottom of the tun and is collected in vessels where it is fermented by the addition of yeasts, which convert the fermentable sugars to ethanol and carbon dioxide. When fermentation is practically complete, the alcoholic liquor is separated from the yeasts and acetified by inoculation with *Acetobacter* cultures. The resultant alcohol is thus oxidized to acetic acid in the presence of atmospheric oxygen. The process lends itself to the different systems in current use. The vinegars will be more or less aromatic, the superior varieties will be those produced by the old and slow Orleans process.

Malt vinegar is straw-colored and must in any case contain 4% w/v of acetic acid. Distilled malt vinegar is prepared by distilling malt vinegar. This product contains only the volatile constituents of the original vinegar and is colorless; it is generally used for the preparation of pickled silverskin onions. (*See* **Malt**: Chemistry of Malting.)

Wine vinegar Wine vinegar, which is obtained from acetous fermentation of wine, is largely produced in continental Europe.

The wines used for acetification are those with too low an alcohol content (7–9% v/v) or those in which volatile acidity is too high. Spoilt wines cannot be used. If wines with a high alcohol content are to be used, they need to be appropriately diluted with water, since a high concentration of alcohol will inhibit the development of the *Acetobacters*. For the same reason, the wine must be free of sulfur dioxide and dregs.

Both white and red or rosé wines can be used to produce white or red vinegar, respectively. In the domestic or small-scale production of wine vinegar, the wine is poured into small wooden barrels, together with the vinegar mother, which consists of colonies of *Acetobacters* taken from barrels in which vinegar has already been produced. The barrel must contain air, so for this reason it is not filled completely. Acetification is slow and stops spontaneously when acidity reaches 7–8%. The slow transformation of wine into vinegar leads to the formation of many substances which impart excellent organoleptic qualities to the end product. The most important of these are acetaldehyde and ethyl acetate – which in top-grade vinegars are present in quantities far greater than those of wine – volatile alcohols, particularly 3-methyl-1-butanol, and 3-hydroxy-2 butanone (or acetylmethylcarbinol).

The vinegar is then partly withdrawn for use and replaced with fresh wine to be acetified. Vinegar produced in this way is bound to vary in composition and character: it may be more or less cloudy and its acidity may vary according to the degree of alcohol in the wine and the nature of the fermentation.

However, most wine vinegar is produced by industrial processes and, as such, is a standardized industrial product; according to EC regulations, its total acidity must not be less than 6 g of acetic acid per 100 ml and residual ethanol may be present in quantities not exceeding 1.5% v/v. The color varies from pale yellow to red, depending on the wine used; the clear vinegars made from white wine are in great demand for pickling. (*See* **Wines**: Types of Table Wine.)

Cider vinegar Cider vinegar is prepared from apple wine that has undergone acetous fermentation and is widely used as a table vinegar, above all in the USA and UK. It is yellowish in color and may be darkened with caramel. Its acidity is not very high and its acidic, astringent flavor recalls the fruit of origin. (*See* **Cider (Cyder; Hard Cider)**: The Product and its Manufacture; Chemistry and Microbiology of Cidermaking.)

Fruit vinegars Fermented juices from other fruits such as peaches and berries are also used to produce vinegar. Since these alcoholic liquids are not distilled, they maintain the subtle flavors and aromas of the raw ingredients; this confers on the vinegars much-appreciated characteristics.

Rice vinegar In the Far East, where rice is the staple cereal, vinegar is prepared from rice as such, from sake (its fermentation product), or from the by-products of sake manufacture. Traditional methods, similar to the Orleans process, are still in use but have been largely superseded by modern submersion techniques. This vinegar has a fairly low acidity and a high amino acid content. It is light in color and has a clean, delicate flavor. It is therefore highly prized in oriental cooking since it does not significantly alter the taste of the food; it is excellent for flavoring with herbs, spices, and fruits because of its mild flavor.

Molasses vinegar This vinegar is prepared from sugar syrup or molasses by double fermentation, alcoholic fermentation of sugars, and the subsequent acetic bioxidation. It serves to make use of the by-products of the sugar industry, but is not widely used.

Honey vinegar This vinegar is obtained from honey which is added to the right quantity of water, subjected to alcoholic fermentation so as to obtain a drink containing ethanol (mead), which, at the right temperature and in the presence of oxygen is subject to the action of acetic bacteria so as to obtain acetic acid. The vinegar is then clarified by bland filtration or by decanting, so that all the qualities of honey remain unaltered.

Spirit vinegar Spirit vinegar, sometimes referred to as white, distilled, or alcohol vinegar, is prepared by acetous fermentation of an alcoholic distillate obtained from the products of alcoholic fermentation of natural sugar solutions. In countries where it is permitted by law, wide use is made of synthetic ethanol, diluted to 10–14% v/v. Spirit vinegar is colorless and is often colored with caramel; it is strongly acid but not aromatic; the distillation of the alcoholic liquids increases the concentration of ethanol, but reduces flavor. This vinegar is less expensive and is the most widespread in the world. When diluted to 4–5% acidity it is used for pickling.

Balsamic vinegar A particular type of highly prized vinegar has been produced for centuries in the provinces of Modena and Reggio Emilia in Northern Italy.

The raw material is grape must, preferably Trebbiano. When alcoholic fermentation has just started, about 24 h after pressing, the must is boiled gently until it has reduced to about one-half or one-third of its starting volume. The result is a liquor with a high sugar concentration (about 30%) in which alcoholic and acetous fermentation take place together extremely slowly; yeasts (*Saccharomyces* and *Zygosaccharomyces*) and acetic bacteria (*Acetobacter* and *Gluconobacter*) are needed for the formation of this vinegar.

Traditional balsamic vinegar takes many years to produce. During the process of fermentation, maturing, and aging, the product becomes highly concentrated and the sugars, alcohols, aldehydes, and organic acids undergo extremely gradual chemical transformation. The vinegar battery consists of a variable number of barrels (between five and 12 or even more) of different woods and decreasing in size. They are set up in well-ventilated areas which are hot and dry in summer and cold in winter. According to the traditional method, part of the contents of the smallest barrel is withdrawn annually for consumption, and is replaced with an equal volume from the next largest barrel in the battery. This is in turn replenished from its neighbor, and so on up the line. Finally, the largest barrel is topped up with the season's boiled must (**Figure 4**).

The process takes at least 12 years, though it is not uncommon to find vinegars that are 50 or more years old, and the yield is very low (no more than 1 l of vinegar from 100 kg of fresh must). The end product, however, is of exceptionally high quality, dark brown in color, of a syrupy consistency, sweet and sour to the taste, and with a characteristically pleasant aromatic smell. In traditional balsamic vinegar the total solids are very high (20–70%), acidity varies between 6 and 18% w/v acetic acid and there are large amounts of sugars, essentially glucose and fructose, as well as numerous aromatic substances that have gradually been formed over the years.

Chemical Synthesis

Acetic acid is usually synthesized by catalytic oxidation from acetaldehyde. Acetaldehyde is in turn obtained by catalytic hydration of acetylene or by catalytic dehydrogenation of ethanol (**Figure 5**). Formic acid and formaldehyde, byproducts of acetic acid synthesis, are removed by distillation.

The acetic acid is subsequently purified and diluted with water to 60–80% by volume to obtain the

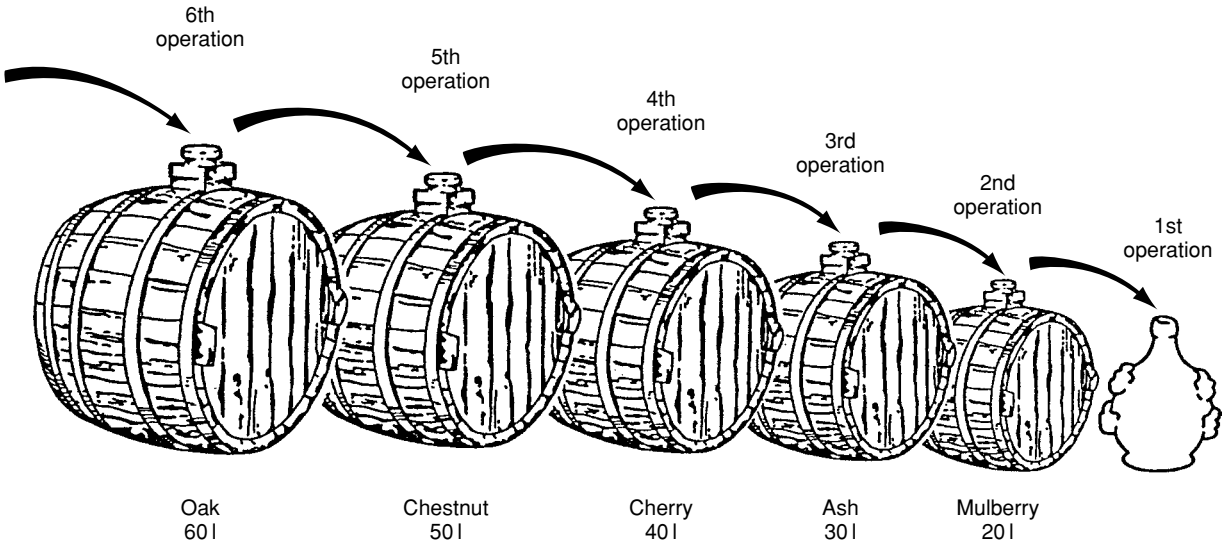

Figure 4 Battery of barrels for manufacture of the 'traditional balsamic vinegar of Modern'.

'vinegar essence.' In order to produce a food-grade vinegar, this 'vinegar essence,' which is a concentrated solution of acetic acid and highly corrosive, is further diluted until it contains 4–5% w/v of acetic acid. This liquid is used in the preparation of synthetic vinegar.

Synthetic Vinegar A number of countries allow a nonfermented vinegar to be used for alimentary purposes which may be produced using synthetic acetic acid that is diluted, aromatized, and colored. The synthetic vinegar thus obtained needs to be colored artificially, and for this purpose caramel is commonly used. It is then aromatized with the addition of sugars, chemical seasoning, and salt or with the addition of natural vinegar. The end product must contain at least 4% w/v of acetic acid, as in the case of fermented vinegars. However, the name 'vinegar' is not always accepted for this product; in the UK for example, it must be labeled 'nonbrewed condiment.'

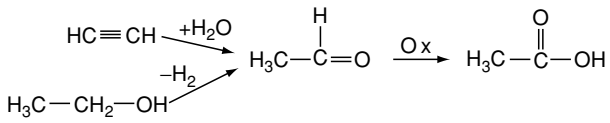

Figure 5 Chemical syntheses of acetic acid.

Characterization and Analysis of Vinegars

The relative density of vinegars ranges from 1.007 for synthetic vinegars to 1.360 for balsamic vinegars. Total solids and ash vary according to type. They are very high in the 'traditional balsamic vinegar of Modena.' They are more abundant in malt, wine, and cider vinegars than they are in spirit vinegars and synthetic vinegars.

Brewed vinegars, being a natural product, contain a large number of components that may derive from the raw materials, be the result of primary or secondary fermentation, or derive from the interaction of components from these two sources (Table 1).

The most abundant volatile constituent is acetic acid, which is largely responsible for the acidity of the vinegar. It ranges from 4–5% for rice vinegars to 10–12.5% for distilled vinegars.

The presence of acetaldehyde, a byproduct of acetous fermentation, is typical. Numerous esters, major contributors to aroma, are present, notably ethyl acetate, the major portion of which was formed during storage, and also methyl and isopentyl acetate; various isomeric butanols and pentanols, present in the raw material, occur in vinegars.

The formation of 3-hydroxy-2-butanone and 2,3-butanedione are typical, deriving from 2,3-butandiol,

Table 1 A comparison of some typical figures for the chemical composition of various vinegars

	Malt vinegar[a]	*Wine vinegar*[b]	*Cider vinegar*[a]	*Spirit vinegar (conc.)*[a]	*Rice vinegar*[c]	*Balsamic vinegar*[d]	*Synthetic vinegar*[a]
Relative density (20°)	1.013–1.022	1.013–1.020	1.013–1.024	1.015–1.020		1.042–1.361	1.007–1.022
Total solids ($g\,l^{-1}$)	3.00–28.40	8.71–24.88	19.00–35.00	1.50–6.00		336.70–873.90	1.00–4.50
Total ash ($g\,l^{-1}$)	0.60–7.60	1.47–3.10	2.00–4.50	0.20–0.50	0.50–7.60	4.00–18.80	0.20–0.50
Alkalinity of ash ($mmol\,l^{-1}$)		9.60–31.60				26–152	
Total acidity as acetic acid (%)	4.30–5.90	5.94–9.20	3.90–9.00	11.50–12.20	4.00–5.24	6.25–14.88	4,10–5.30
Volatile acidity as acetic acid (%)		5.55–7.95			3.79–5.16	3.90–13.60	
Fixed acidity as acetic acid (%)	0.20–0.40	0.02–0.55	0.10–0.20		0.05–0.66	1.58–2.27	Negligible
Alcohol (% v/v)		0–0.14	0–0.81	0.03	0.15	0–0.68	0.04–0.08
Total nitrogen ($g\,l^{-1}$)	0.40–1.40	1.15–1.86		0.03–0.30	0.80–1.29	1.02–2.16	0–2.00
Sugars ($g\,l^{-1}$)		0–6.20	1.50–7.00		0–91.00	351.00–689.70	
Glycerol (%)[e]		0.37	0.28			0.94–1.7	
Gluconic acid (%)[e]		0.02	0.02			0.68–1.14	
Tartaric acid (%)[e]		0.11				0.46–0.52	
Acetic acid (%)[e]		7.2	5.5			6.38–7.64	
Citric acid ($g\,l^{-1}$)		0.26–0.39				1.66–3.47	
Malic acid ($g\,l^{-1}$)		0.47–0.80	0.70–1.60			8.00–37.40	
Acetaldehyde ($mg\,l^{-1}$)		18.9–114.4				84.9–374.7	
2,3 Butanediol ($mg\,l^{-1}$)						470–2890	
2,3 Butanedione ($mg\,l^{-1}$)						17–37	
3 Hydroxy-2 butanone ($mg\,l^{-1}$)		122.0–1257.0				249.1–1597.8	
Ethylacetate ($mg\,l^{-1}$)		206.0–590.0					

[a]Data from Egan H, Kirk RS and Sawyer R (1987) *Pearson's Chemical Analysis of Foods*, 8th edn. Harlow: Longman Scientific and Technical.
[b]Data from Stella C (1976) La composizione degli aceti *Vini d'Italia* 18: 269.
[c]Data from Mizumoto H, Minamisone H, Mori K and Azumo K (1975) Chemical composition of commerical rice vinegars. *Kagoshima-ken Kogjo Shikenjo Nempo* 22: 67.
[d]Data from Plessi M and Coppini D (1984) L'aceto balsamico tradizionale di Modena. *Atti della Società dei Naturalisti e Matematici* 115: 39–46.
[e]Data from Giudici P *et al*. (1994) Origine ed evoluzione degli acidi organici durante l'invecchiamento dell'aceto balsamico. *Industria delle bevande* 23: 569–574.

Table 2 A comparison of major volatile components of various vinegars

Malt vinegar	*Wine vinegar*	*Cider vinegar*	*Distilled vinegar*	*Balsamic vinegar*
Ethyl acetate	Acetaldehyde	Acetaldehyde	Acetaldehyde	Ethyl acetate
Ethyl formate	Ethyl acetate	Ethyl acetate	Ethyl acetate	Ethanol
3-Methyl-1-butanol	Ethyl formate	Ethyl formate	Ethyl formate	3-Hydroxy-2-Butanone
2-Methyl-1-propanol	2-Methyl-1-propanol	3-Methyl-1-butanol	Ethanol	2,3-Butanediol
2-Methyl-1-butanol	3-Hydroxy-2-butanone	2-Pentanol	2-Pentanone	2,3-Butanedione
Acetaldehyde	2,3-Butanediol	2-Methyl-1-butanol	3-Pentanol	Methyl acetate
Ethanol	3-Methyl-1-butanol	2-Methyl-1-propanol	2-Pentanol	Acetaldehyde
Isobutyl acetate	Isopentyl acetate	Ethanol	Methyl formate	Furfural
Isopentyl formate	2-Methyl-1-butanol	2-Pentanone	Propionaldehyde	2,4,5-Trimethyl-1,3-dioxolane
Isopentyl acetate	Ethanol	3-Pentanol	Methanol	Acetone
Benzaldehyde	Isovaleraldehyde	Benzaldehyde	Acetal	3-Methyl-1-butanol
Acetone	Isobutyl acetate	Propionaldehyde	2-Methyl-1-butanol	Ethyl formate
Isobutyl formate	γ-Butyrolactone	Ethyl lactate	3-Methyl-1-butanol	Isobutanal
2-Pentanone	Acetone	Phenylacetaldehyde		2-Methyl butanal
2-Methyl-3-buten-2-ol	Isobutyl formate	Diethyl succinate		2-Butanone
	2-Pentanone			Furan

a product of the condensation of acetaldehyde by microorganisms, which are always, and only, found in brewed vinegars, albeit in variable quantities (Table 2).

Organic acids, other than acetic acid, are also present: tartaric and citric are largely found in wine vinegars, while formic and malic acid are those most commonly found in cider vinegars. During the fermentation process small quantities of other acids, such as succinic acid, lactic acid, and fumaric acid, are formed.

Of the many analyses that vinegars lend themselves to, the most useful from the point of view of general characterization are the following:

- *Sensory examination*: The sensory characteristics to be taken into account are color, odor, and taste, which must conform to those attributed to the type of vinegar being analyzed. If the vinegar is cloudy, it must be examined under strong magnification for vinegar eels (*Anguillula aceti*), a nematode that is a typical parasite of vinegar.
- *Relative density*: This is determined at 20 °C using a Westphal balance or pyknometer.
- *Total acidity*: This is determined by titrating the diluted product with a standard alkaline solution. It is expressed as grams of acetic acid per 100 ml of vinegar.
- *Fixed acidity*: A 10-ml aliquot of vinegar is evaporated to dryness and reconstituted with water, the process being repeated in order to eliminate acetic acid. The final residue is dissolved in water and titrated with an alkaline solution. It is expressed as grams of tartaric acid per 100 ml of vinegar.
- *Volatile acidity*: This is calculated from the difference between total and fixed acidity, both expressed as grams of acetic acid per 100 ml of vinegar.
- *Total solids*: A 50-ml aliquot of vinegar is evaporated in a platinum capsule with water in order to eliminate the acetic acid. It is then dried to constant weight in an oven at 100 °C.
- *Ash*: The total solids are ignited at 500–550 °C. If a high ash figure is obtained, salt should also be determined.
- *Ash alkalinity*: This is chiefly due to potassium carbonate formed during calcination and is determined by dissolving the ash in an exact quantity of standard sulfuric acid, of which the excess is titrated with alkali. It is expressed in mmol of alkali per liter.
- *Ethanol*: A 100-ml aliquot of vinegar is neutralized with sodium hydroxide and double-distilled. The density of the solution is measured and, consequently, the quantity of ethanol is calculated from tables.
- *Total nitrogen*: This is determined by the Kjeldahl method on 25 ml of vinegar.

Methods of Differentiating Between Various Types of Vinegar

Apart from color and odor, the nitrogen, phosphates, ash, and their alkalinity also serve to differentiate between the various types of vinegar. In a malt vinegar, nitrogen and phosphate figures are low, while in a rice vinegar the nitrogen figure is higher and distilled vinegar has a very low ash content.

However, useful methods of routine differentiation are based on those values that depend on the volatile substances present, namely:

- *Oxidation value*: The number of ml of 0.002 mol l^{-1} $KMnO_4$ discolored by 100 ml of sample in 30 min under standard conditions. Alcohol and

acetylmethylcarbinol are mostly responsible for the oxidation value.

- *Iodine value*: The number of ml of 0.01 mol l^{-1} iodine absorbed by 100 ml of sample under standard conditions. This value is mostly influenced by acetylmethylcarbinol and diacetyl.
- *Ester value*: The number of ml of 0.01 mol l^{-1} KOH required to saponify the esters contained in 100 ml of sample under standard conditions.

Brewed vinegars give comparatively high values, but those for artificial products are low as the latter are almost devoid of volatile reducing substances.

Vinegars can be more specifically identified by means of gas chromatography. This technique can distinguish brewed vinegar from synthetic vinegar and also brewed vinegars from different origins by analyzing the accompanying compounds (metabolic byproducts of *Acetobacter* strains and substances derived from the raw material).

In addition, the blending or adulteration of brewed vinegar with synthetic acid can be detected by mass spectrometric determination of the $^{13}C/^{12}C$-isotope ratio; brewed vinegar has 5% more ^{13}C isotope than acetic acid synthesized petrochemically.

Food Uses

Vinegar has long been used worldwide as a basic seasoning in the preparation and cooking of certain foods, because its sharp taste makes it so useful and versatile.

A considerable quantity of vinegar is marketed as such for domestic use. In the UK and USA the table vinegar most widely used is cider vinegar, while in Ireland it is malt vinegar and, in grape-growing countries, such as Italy, France, and Spain, wine vinegar. In the Far East, in addition to the traditional rice vinegar, synthetic vinegar is very common.

Vinegar adds flavor to vegetable and meat products. It is one of the ingredients of salad dressings, sauces, such as tabasco, and tomato products, such as ketchups, mustard, and aspics. Mixed with oil and salt it makes the classic vinaigrette, and it can be used as a condiment for salad and as a sauce for cold, cooked vegetables, meat, and fish.

For condiment uses, vinegar can be aromatized with herbs and spices like tarragon, basil, garlic, shallot, and elder. They are steeped in the vinegar when acetification is complete; in certain types of vinegar sugars are added as well. These vinegars add special and unusual tastes to food.

The aroma and sweet and sour flavor which balsamic vinegar confers on foods to which it has been added is decidedly particular; it is suitable for all types of food, such as sauces, green salads, and meat and sprinkled on strawberries, peaches, or melon adds a pleasant flavor and aroma.

Finally, because of its acetic acid content and low pH, vinegar is used as a preservative for both domestic use and in the food industry. It is in fact used for the preservation, or pickling, of a wide variety of foods such as vegetables, meat, fish products, and spiced fruits. For this purpose, the food-processing industry uses mainly distilled vinegar and, where the law permits, synthetic vinegar.

See also: **Acids**: Properties and Determination; Natural Acids and Acidulants; **Preservatives**: Classifications and Properties; **Spices and Flavoring (Flavouring) Crops**: Use of Spices in the Food Industry

Further Reading

Adams MR (1985) Vinegar. In: Wood BJB (ed.) *Microbiology of Fermented Foods*, pp. 1–47, vol. 1. Essex: Elsevier Applied Science.

AA VV (1996) *L'aceto Balsamico di Modena*. Modena: Consorzio Tutela Aceto Balsamico di Modena.

Carnacini A and Gerbi V (1992) L'aceto di vino, un prodotto da valorizzare e tutelare. *Industrie delle Bevande*, 19: 465–483.

Conner HA and Allgeier RJ (1976) Vinegar: its history and development. *Advances in Applied Microbiology* 20: 81–133.

Coppini D, Plessi M and Monzani A (1978) Ulteriore contributo allo studio dell'aceto Balsamico. *Rivista di Viticoltura ed Enologia* 31: 351–357.

Egan H, Kirk RS and Sawyer R (1987) *Pearson's Chemical Analysis of Foods*, 8th edn. Harlow: Longman Scientific and Technical.

García-Parrilla MC, Gonzáles GA, Heredia FJ and Trocoso AM (1997) Differentiation of wine vinegars based on phenolic composition. *Journal of Agriculture and Food Chemistry* 45: 3487–3492.

Garoglio PG (1973) *Enciclopedia Vitivinicola Mondiale*. Milano: Edizioni Scientifiche UIV.

Masai H (1980) Recent technical developments on vinegar manufacture in Japan. *Proceedings of the Oriental Fermented Foods* 192–198.

Mizumoto H, Minamisone H, Mori K and Azumo K (1975) Chemical composition of commercial rice vinegars. *Kagoshima-ken Kogyo Shikenjo Nempo* 22: 67–68.

Nieto J, Gonzá-Viñas MA, Barba P *et al.* (1993) Recent progress in wine vinegar R & D and some indicator for the future. In: Charalambous G (ed.) *Food Flavours, Ingredients and Composition*, pp. 469–500. Amsterdam: Elsevier Science.

Plessi M and Coppini D (1984) L'aceto balsamico tradizionale di Modena. *Atti della Società dei Naturalisti e Matematici* 115: 39–46.

Plessi M, Monzani A and Coppini D (1987) Determination of the monosaccharide and alcohol content of balsamic

and other vinegars by enzymatic methods. *Agricultural and Biological Chemistry* 52: 25–30.
Plessi M, Monzani A and Coppini D (1989) Quantitative determination of acids and derivatives in balsamic and other vinegars. *Sciences des Aliments* 9: 179–183.
Saccani F and Ferrari Amorotti V (1999) *Il Balsamico della Tradizione Secolare*. Modena: Consorteria dell'Aceto balsamico tradizionale di Modena.
Sacchetti M (1970) *Sull'aceto Balsamico Modenese*. Bologna: Edagricole.
Stella C (1976). La composizione degli aceti. *Vini d'Italia* 18: 269–274.
Tesfuge W, Morales ML, Garcia-Parrilla MC and Troncoso AM (2002). Wine vinegar authenticity and quality evaluation. *Trends in Food Science and Technology* 13: 2–33.

VIRUSES

H Appleton, PHLS Central Public Health Laboratory, London, UK

Background

Viruses are usually transmitted directly from person to person, but it is clearly recognized that, on occasions, some viruses may also be transmitted by foods and water. Unlike bacteria, viruses do not multiply or produce toxins in food: food items merely act as vehicles for their passive transfer. Any virus that is infectious via the gastrointestinal tract could be considered a potential foodborne pathogen, but in practice, the most commonly reported foodborne viral infections are gastroenteritis and, less frequently, hepatitis. Although the clinical characteristics of these two illnesses differ, their food-related mode of transmission is essentially the same. Foodborne spread of both these infections has been clearly defined, but the extent of the problem is unknown. For various reasons, the incidence of both is almost certainly under-reported.

Viral Gastroenteritis

Gastroenteritis is the most common foodborne viral infection. Discovery of the viruses responsible for causing gastroenteritis began with the identification of the Norwalk virus in 1972 and then the human rotavirus in 1973. These viruses are usually transmitted directly from person to person via the fecal–oral route. In 1977, however, there was a report that a virus had been responsible for a community-wide outbreak in England associated with shellfish. At least 800 people were affected, and it is now clearly established that food may also play an important role in transmission.

Prior to the discovery of gastroenteritis viruses and recognition of their foodborne mode of transmission, there were many reports of 'food poisoning outbreaks of unknown etiology,' which were assumed to be bacterial but in retrospect have the characteristics of viral gastroenteritis. Such reports still occur when appropriate viral investigations are not undertaken. Viral gastroenteritis tends to mimic bacterial food poisoning, and invariably, outbreaks are initially investigated for traditional food poisoning bacteria. However, there are characteristic features that should alert investigators to the possibility of a viral cause. The incubation period is somewhat longer and more variable than for most bacteria, with a peak of around 30–36 h (range 12–60 h, which may be dose-dependent). Symptoms are variable and may include both vomiting and diarrhea as well as the usual range of associated complaints such as abdominal pain, general malaise, and nausea.

Secondary person-to-person spread is a characteristic feature. Illness is usually regarded as mild and self-limiting with recovery within 24–48 h. However, people commonly feel debilitated for 2–3 weeks, and this can have considerable impact socially and economically in impaired performance. Outbreaks may be very disruptive and have led to closure of hospitals and schools. Some outbreaks have been very large, involving several hundred people. Treatment is not normally required unless people are severely affected and become dangerously dehydrated. This is a more common problem in children suffering diarrheal diseases in underdeveloped countries. Oral rehydration therapy with simple balanced salt solutions is an effective treatment.

The overall incidence viral gastroenteritis is unknown. A community-based study of infectious intestinal disease, from all causes in England from 1993 to 1996, confirmed that the vast majority of cases go unreported. The study found that viruses were responsible for a far greater proportion of gastrointestinal illness than national surveillance figures had previously indicated. The study also concluded that infections resulting from viruses are less likely to be

reported than bacterial infections. It was estimated that only one in 35 infections with rotavirus and one in 1562 infections caused by Norwalk-like viruses occurring in the community are reported to national surveillance, compared with one in three cases of salmonellosis.

The proportion of these viral gastrointestinal infections that is foodborne is similarly unknown. Viruses account for 6% of foodborne outbreaks and 5% of waterborne outbreaks occurring in England and Wales and reported to national surveillance, but this is likely to be a gross underestimate of the true incidence. The mild nature of the illness means that persons do not tend to consult a medical practitioner. However, the short incubation period means that when an illness is investigated, possible association with a food source is more likely to be recognized than for hepatitis A, for which there is a long incubation period. The highly infectious nature of the viruses and the frequent occurrence of secondary person-to-person spread also add to the difficulties of identifying a food- or waterborne outbreak.

The Gastroenteritis Viruses

Several different viruses cause gastroenteritis (Table 1), but it is the viruses of the Norwalk group that have been particularly linked to foodborne transmission. They form a complex group, and the nomenclature is confusing. The viruses have been known variously as Norwalk-like viruses (NLV), and as small round structured viruses (SRSV) because of their appearance when viewed in the electron microscope (Figure 1). The virus particle measures 30–35 nm in diameter and has an amorphous surface and a ragged outline. They are classified as members of the Caliciviridae family and hence are sometimes referred to as human caliciviruses.

Norwalk-like Viruses The Norwalk virus, which originated from a school outbreak of nonbacterial gastroenteritis at Norwalk, Ohio, in the USA, was the first virus of the group to be discovered. Viruses of this group are frequently the cause of both community and institutional outbreaks of nonbacterial gastroenteritis, whether resulting from person-to-person spread or a food or water source. Outbreaks in hospitals, residential homes, schools, hotels, restaurants, and cruise ships have been reported frequently. The rapid spread of infection can cause major disruption, leading to closure of schools and hospital wards. There are also considerable economic costs in terms of working days lost. They are probably the most important cause of sporadic gastroenteritis in the community, although because of the relatively minor nature of illness, such incidents are not usually investigated. In the UK, NLVs have been implicated in more than 90% of foodborne outbreaks where a virus has been detected, and it is clear that this virus group is the main cause of foodborne viral gastroenteritis in developed countries.

Several different serotypes have been identified. Immunity to NLVs is poorly understood and may be short-lived. Volunteer studies have shown that some

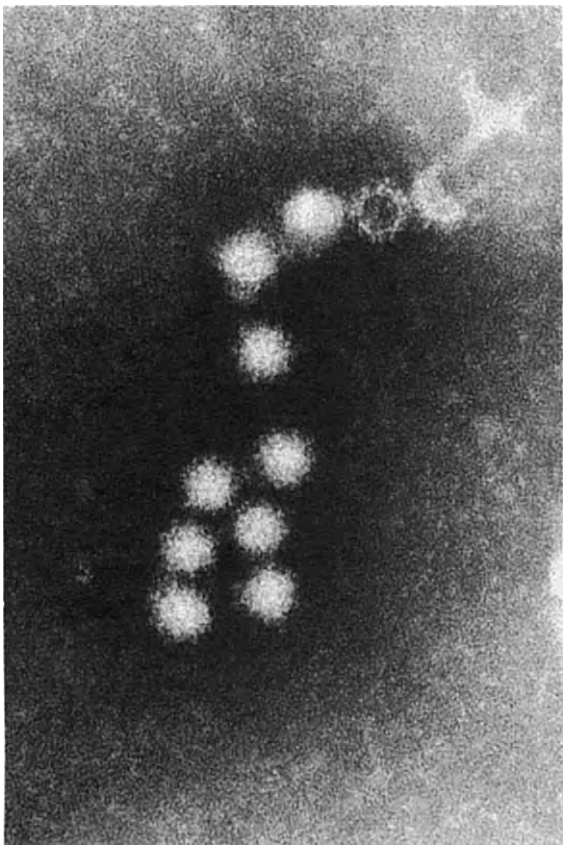

Figure 1 Electron micrograph of a group of Norwalk-like viruses. The virus particles have an amorphous surface and ragged outline. Most particles are complete, but an empty stain-penetrated particle can be seen at the top of the group. Virus particles negatively stained with phosphotungstic acid. Magnification: ×200 000.

Table 1 Viruses commonly causing gastroenteritis

Virus	*Person-to-person spread*	*Food- or waterborne spread*
Norwalk-like virus[a]	Yes	Yes
Rotavirus	Yes	Infrequent[b]
Astrovirus	Yes	Infrequent
Adenovirus (types 40 and 41)	Yes	Not reported

[a]Also known as small round structured viruses.
[b]Food- and waterborne transmission is probably more significant in underdeveloped countries.

people may be infected on more than one occasion by the same virus. Thus, people who experience several incidents of illness may not only be infected with different antigenic types of NLV, but may also develop illness on more than one occasion from the same virus.

The development of polymerase chain-reaction (PCR) assays has allowed the molecular characterization of virus strains from many outbreaks. The human gastroenteritis viruses in the Caliciviridae family form three broad genogroups. Two groups have the morphology of NLVs (Figure 1) and broadly similar genomic features. Members of the third group are known as Sapporo-like viruses (SLVs). The genomic arrangement of this third group is distinct from the other two groups. In addition, SLVs have the classical morphology more usually associated with the calicivirus group, with hollow cup-like depressions on the surface, which are not usually seen in NLVs. (The name calicivirus derives from the Latin *calix* meaning cup.) The different serotypes of the human caliciviruses broadly correspond with the genotypic groups. The majority of reported infections caused by viruses with classical calicivirus morphology have occurred in young children. Outbreaks in nursery groups involving person-to-person transmission have been recorded. The virus has also observed in older persons, and some reported foodborne outbreaks have been attributed to calicivirus.

Rotavirus Rotaviruses mainly infect young children and are the most frequently reported of the gastroenteritis viruses. It is estimated that they may cause a million deaths a year in children under 5 years of age, mostly in developing countries. In developed countries, deaths are relatively rare, but rotavirus gastroenteritis is the most frequent reason for admission of young children to hospital. Foodborne and particularly waterborne spread are probably significant modes of transmission in developing countries, but in developed countries, reports are rare. Although rotavirus has been detected in shellfish, so far, there have been no reports of illness from this source.

Astrovirus Astroviruses form a morphologically distinct group of viruses, and are named from the five- or six-point star seen by electron microscopy on the surface of some particles. Astroviruses are normally associated with gastroenteritis in young children, often under 1 year of age. There have been occasional reports of astrovirus infection in some adults following consumption of shellfish or contaminated water, but these incidents appear comparatively rare.

Adenovirus Adenovirus serotypes 40 and 41 have been specifically linked to gastroenteritis, and infections in young children are frequently reported. Reports in older patients are rare. There have been no reports of food- or water-associated adenovirus infection.

Parvovirus Parvoviruses are small DNA-containing viruses. Their role as causative agents of gastroenteritis in humans is uncertain, since, unlike other gastroenteritis viruses, there is a prolonged carrier state in some persons. However, parvoviruses are an important cause of gastroenteritis in some animal species. Parvoviruses have been observed in school outbreaks of winter vomiting disease and in a number of shellfish associated outbreaks of gastroenteritis. Parvovirus-like particles have been detected occasionally in shellfish that have been implicated in illness and where a similar virus has been found in patients.

Hepatitis

Hepatitis is an inflammation of the liver and is usually recognized by the occurrence of jaundice. Descriptions of hepatitis can be traced back to the time of Hippocrates, but it was not until the eighth century that the infectious nature of the disease was first mentioned in letters from Pope Zacharias to Saint Boniface, Archbishop of Mainz. The viral etiology was first postulated by MacDonald in 1918 and was later confirmed by volunteer studies in the 1940s. It is now known that several different viruses may cause hepatitis, but it is the enterically transmitted viruses, hepatitis A and hepatitis E, that are of importance in foodborne and waterborne transmission.

Types of Hepatitis

Hepatitis A Hepatitis A is transmitted by the fecal–oral route and is prevalent in conditions of overcrowding and poor hygiene. Although it has long been recognized that hepatitis A may be transmitted by food and water, the incubation period of 3–6 weeks makes it difficult to link infection to a food source. Hence, reports of foodborne hepatitis are infrequent and are unlikely to indicate the true extent of the problem. Symptoms in addition to, or instead of, jaundice are nausea and general malaise. Patients may feel unwell for several weeks, but recovery is complete; fatalities are rare. There is no specific treatment, but a vaccine is available. Illness tends to be less severe in children, and in some, infection is symptomless. Once infected, a person is immune from further infection. As hygienic conditions improve, infection becomes less common in children, and a larger proportion of the adult population becomes susceptible.

Serological surveys in the UK indicate that only 10–15% of young adults have antibody to hepatitis A virus (HAV).

HAV is a small, compact icosohedral virus measuring 25–30 nm in diameter. It is classified as a picornavirus (from the Latin *pico* (small) + rna as it has an RNA genome). Within the Picornaviridae family, it is assigned to the hepatovirus genus. There is only one serotype.

Hepatitis E More recently, another enterically transmitted form of hepatitis (designated hepatitis E) has been described. This virus has been associated with large waterborne outbreaks in some developing countries, notably in Asia, Africa, and Central America. Illness appears more severe than hepatitis A, particularly in pregnant women for whom a mortality rate of 17–34% has been reported. Secondary person-to-person transmission is estimated at only 0.7–8%. The primary source of infection appears to be contaminated water rather than person-to-person spread. Cases in the UK are reported infrequently and are mainly imported from endemic areas.

Hepatitis E is classified as a calicivirus but is serologically unrelated to the viruses that cause gastroenteritis. Like all caliciviruses, the virus particles have a diameter of 30–35 nm. When viewed in the electron microscope, the characteristic cup-like depressions of members of the calicivirus group can be seen on the surface of the virus particles.

Other Hepatitis Viruses There are other hepatitis viruses that are not transmitted by the fecal–oral route. Hence, they are not considered to be associated with foodborne transmission. The commonest, hepatitis B and hepatitis C, are usually transmitted parenterally, e.g., by blood transfusions, from injections of contaminated blood products such as factor VIII, or from contaminated needles used by drug addicts. They may also be transmitted by sexual contact.

Other Viruses

Enteroviruses

Enteroviruses form a distinct genus within the Picornaviridae family and include the polioviruses, echoviruses, and Coxsackie viruses. Most infections are asymptomatic, but enteroviruses can cause a variety of symptoms, including paralysis, meningitis, cardiomyopathy, and less severe conditions such as colds and fever. Most enteroviruses grow readily in cell cultures of human and primate origin. An outbreak of poliomyelitis associated with raw milk in 1914 is believed to be the first recorded foodborne outbreak of viral illness. Raw milk was implicated in a further 10 outbreaks in the UK and USA by 1949. Milkborne outbreaks of poliomyelitis then appeared to cease in developed countries, even before the introduction of polio vaccine. This could largely be attributed to the widespread use of pasteurization. Various enteroviruses have been isolated from foods such as meat and shellfish, but there is no clear epidemiological evidence to suggest that foodborne infection is occurring.

Tick-borne Encephalitis

A further mode of foodborne transmission of viruses is by zoonotic infection. Cases of tick-borne encephalitis arising from consumption of raw milk from infected goats, sheep, and cattle or products made from unpasteurized milk have occurred in Eastern Europe. The incidence is fortunately rare, owing to the limited distribution of the appropriate species of ticks.

Entry of Viruses to the Food Chain

Foods may become contaminated with viruses in two main ways. Food may be contaminated at source, usually by coming into contact with sewage-polluted water: this is known as primary contamination. Additionally, food may be contaminated during preparation and serving by infected food handlers, in just the same way as some types of foodborne bacterial contamination can arise. This is known as secondary contamination.

Contamination at Source

Shellfish The most clearly implicated foods in the transmission of viruses are bivalve molluscs (oysters, clams, cockles, and mussels). Indeed, most illness associated with this type of shellfish is viral. These shellfish are harvested from shallow estuarine and coastal waters, which are frequently polluted with sewage. They feed by filtering particulate matter from the large volumes of water passing over their gills, and this can include potentially pathogenic microorganisms. Although human viruses do not replicate in shellfish, they can be concentrated within molluscs to higher concentrations than occur within the surrounding water up to 100-fold. Bivalve molluscs are not usually thoroughly cooked before being eaten, and oysters are frequently consumed raw.

Fruit and Vegetables There is the potential for fruits, vegetables, and salad items to become contaminated during irrigation with polluted water and by fertilization with untreated or inadequately treated sewage sludge. Many vegetables are cooked before

consumption, and there is no evidence that if such food is primarily contaminated, illness ensues. However uncooked soft fruits such as strawberries and raspberries have been implicated in a number of outbreaks when contamination is assumed to have occurred at the growing site. Salad items are commonly implicated in outbreaks, although in many of these incidents, contamination is believed to have originated from infected food handlers at the time of preparation.

Secondary Contamination: Food Handlers

Viruses causing gastroenteritis and hepatitis A are infectious in very low doses, and thus are spread very readily from infected persons. It is now recognized that outbreaks arising from infected food handlers are common occurrences. The foods that pose the greatest risk to the consumer are cold items, such as sandwiches and salads, that require much handling during preparation. Without meticulous attention to personal hygiene and thorough and frequent hand washing, fecally contaminated fingers can contaminate foods and work surfaces. Viral gastroenteritis can be very sudden in onset and commence with projectile vomiting. The virus can be disseminated over a wide area in aerosol droplets so that, in food preparation areas, exposed foods, working surfaces, and door handles are likely to be contaminated.

Virus Survival

Gastroenteritis viruses and hepatitis viruses are robust viruses and survive extremely well in the environment. It is the NLVs and HAV that are the most resistant to inactivation. There is little precise information on the stability of NLVs, since they cannot be cultured in the laboratory. Most information on survival and inactivation has come from epidemiological observations in outbreaks and limited studies of infectivity in volunteers. Some strains of HAV and rotavirus can be cultured, and there have been a small number of experimental studies on their stability. Gastroenteritis and hepatitis viruses have been detected in food, water, and other environmental samples by PCR assays, but such assays only detect the viral genome and do not necessarily correlate with infectivity.

Viruses that infect via the gastrointestinal tract are acid-stable. Both NLVs and HAV retain infectivity after exposure to acidity levels below pH 3. Hence, they are often able to survive the food-processing and preservation conditions used to inhibit bacterial and fungal spoilage of foods, such as pickling in vinegar and fermentation processes to produce foods such as yogurt. Alcohol and high sugar concentrations are unlikely to have any adverse effect on survival. Refrigeration and freezing are routinely used to preserve viruses and have no effect on viability. Frozen foods that are not subsequently cooked, such as soft fruits, have been implicated in a number of incidents of both viral gastroenteritis and hepatitis A.

The viruses are destroyed by normal cooking processes. For foods that only undergo partial cooking in special heat processes, such as the treatment of shellfish and other pasteurization procedures, care must be taken to ensure that inactivation of virus is achieved. Both NLVs and HAV retain infectivity after heating to 60 °C for 30 min.

Gastroenteritis viruses and hepatitis viruses appear to survive in the environment for long periods. They survive on inanimate surfaces, on hands and in dried fecal suspensions. Lingering outbreaks have occurred in hospitals, residential homes and on cruise ships, probably as a result of environmental contamination. NLVs have been detected by PCR in swabs from hospital lockers and hotel carpets supposedly cleaned after incidents of vomiting.

There is conflicting evidence on the efficacy of disinfectants. Chlorine-based disinfectants are usually considered the most effective against enteric viruses and the most suitable in food preparation areas. However, NLVs and HAV are resistant to a level of 0.5–1 mg per liter of free residual chlorine, which is consistent with that present in drinking water. NLVs are inactivated by 10 mg per liter of chlorine. In the USA, a level of 5 mg per liter of chlorine with a contact time of 1 min is recommended for inactivation of HAV. Infectivity of HAV is reduced by sodium hypochlorite, 2% glutaraldehyde, and quarternary ammonium compounds, but there are no comparative data for NLVs.

Virus Detection

In Food

It is not feasible to examine food samples routinely for the presence of viruses. Although it has been possible to culture HAV since 1980, primary isolation is a lengthy and unreliable procedure, and there have been very few isolations of virus from food or water samples. Virus was isolated from a drinking water supply responsible for a large outbreak in the USA, but this involved culture of highly concentrated water samples for up to 21 weeks. There have been several studies involving recovery of virus from artificially contaminated foods, particularly shellfish, but these have used highly adapted laboratory strains of virus. Viruses stick avidly to shellfish meat, and complex extraction techniques are required. Even with the best

of these methods, virus recovery rates are poor. NLVs cannot be cultured in the laboratory. Culture of rotaviruses and astroviruses is complex and unreliable, and the problems of avidity similarly apply. Rapid solid-phase immunoassays are available for the detection of hepatitis and gastroenteritis viruses. These are not sufficiently sensitive, however, to detect the very small numbers of virus particles likely to be present in contaminated foods. Similarly, electron microscopy is insufficiently sensitive for the detection of virus in food samples.

There is considerable interest in the application of PCR assays to the detection of viruses in food and water samples. PCR is an extremely sensitive technique. However, extraction of viral RNA from samples and removal of PCR inhibitors is a fairly complex and time-consuming procedure. Hence, such assays cannot be used for routinely screening food samples. Furthermore, unlike culture, a positive PCR result only indicates the presence of viral nucleic acid and does not necessarily confirm the presence of viable infectious virus. These tests do have an important role in the epidemiological investigation of outbreaks and assessing the efficacy of treatment procedures.

In Patients

HAV is excreted in feces, but it is not usually appropriate to examine fecal specimens for virus as the main excretion period precedes illness. The usual diagnostic test is detection of anti-HAV specific immunoglobulin M (IgM) antibody in serum, for which there are a number of commercially available solid-phase immunoassays. The presence of IgM antibody is indicative of recent infection. If the fecal specimens are collected at the appropriate time, it is quite feasible to detect viral antigen by various techniques, the most rapid and straightforward being the solid-phase immunoassays and PCR tests.

All the gastroenteritis viruses were discovered by electron microscopy, and this has continued to play an important role in diagnosis. Viruses from different groups look different, and it is this characteristic morphology that is the basis for their identification. However, electron microscopy is a time-consuming technique that demands considerable skill and experience on the part of the operator, and it is not a technique that lends itself to examining large numbers of specimens. PCR assays are being used increasingly for the detection of NLVs, but their use is mainly confined to specialist virology laboratories. Although of far greater sensitivity than electron microscopy, current primers will not detect all strains of NLVs. Recombinant capsid antigens have been made from a small number of NLV strains and have been used to develop ELISA tests. So far, only a very few NLV types can be detected, and the tests are not yet widely available. Commercial ELISA and latex agglutination test kits are readily available for the detection of rotaviruses. Astroviruses are usually detected by electron microscopy. Some laboratories use in-house PCR assays or culture combined with immunofluorescence.

Prevention and Control

Elimination of viral contamination at source is an ideal that is unlikely to be attained in the near future. In practice, it would be necessary to prevent the discharge of sewage into rivers and coastal waters, which not only pollutes water supplies that may be used for irrigation but also results in the contamination of shellfish. Sewage sludge provides beneficial nutrients and is applied to agricultural land as fertilizer. In developed countries, increasingly stringent regulations for the use of untreated or partially treated sewage are being introduced. Since the end of 1999, application of all untreated sewage sludge on agricultural land has been prohibited in the UK. Where treated sewage is applied to land for growing salad and vegetable crops, permitted time intervals, between applying the sewage and harvesting the crops, are specified. Hepatitis A is endemic in underdeveloped areas of the world and viral gastroenteritis more prevalent. Viruses commonly occur in polluted coastal waters in these areas, and contamination of shellfish is likely. Standards for the use of sewage in agriculture are also less satisfactory. Imports of fruits, salads, vegetables and shellfish could potentially pose a risk to consumers.

It is usual practice to treat shellfish in some way to remove microbial contamination before they are sold to the public. In the European Union, conditions for bivalve molluscs are laid down in the European Council Directive on Shellfish Hygiene (91/492/ECC), and there are similar regulations in the USA. Shellfish, such as oysters and mussels, may be cleansed, either by relaying in cleaner coastal or estuarine waters or by transfer to land-based depuration facilities. Harmful microorganisms should be washed out during the natural feeding process. These procedures are remarkably successful in removing bacterial contamination but do not completely eliminate viruses. These may remain for several weeks in molluscs held in depuration tanks. Even though molluscs may appear entirely satisfactory in bacteriological tests, incidents of viral illness may follow their consumption.

Up to the middle of the 1980s, many outbreaks of viral illness were recorded, related to the

consumption of inadequately cooked shellfish such as cockles. Prolonged cooking results in a tough unpalatable product, and thus the aim of treatment must be to apply the minimum heat necessary to render shellfish safe. Studies on the inactivation of hepatitis A virus in shellfish led to recommendations that the internal temperature of shellfish meat should be raised to 90 °C and maintained for 1.5 min. These recommendations were adopted by the UK shellfish industry and subsequently accepted as a valid method of treatment by the European Commission. Later studies, using feline calicivirus as a model, indicated that these conditions were also adequate for inactivation of the NLVs. Epidemiological evidence strongly supports the view that this treatment is effective for the inactivation of gastroenteritis viruses. Since early 1988, there have been no further reports in England and Wales of any viral illness, either hepatitis A or gastroenteritis, from shellfish heat-treated according to the recommendations.

Foods may be contaminated during preparation and serving by infected food handlers. Both the gastroenteritis viruses and hepatitis A virus are extremely infectious in low doses and thus are spread easily from infected persons. Persons with symptoms should be excluded from handling food. However, food handlers with very minimal symptoms have been implicated in the transmission of NLVs. There is a little circumstantial evidence that asymptomatic excretion of virus occurs, but in the absence of more definitive data, it is generally recommended that people should be allowed to resume work 48 h after symptoms have ceased. That recommendation was based on the rapid decline in virus shedding observed by electron microscopy and in practice appears to work satisfactorily. However, NLVs often can be detected by PCR for a longer period than by electron microscopy and, in some instances, for up to a week after the onset of symptoms. It is not clear if persons shedding virus detectable by PCR are infectious after symptoms have ceased. Recommendations on how long to exclude people from work need to be kept under review.

Excretion of hepatitis A virus precedes symptoms, and hence, early exclusion of infectious food handlers is not usually possible. An effective vaccine is available. Currently, in the UK, hepatitis A vaccine is used selectively for persons at high risk, such as travelers. It is not generally used for food handlers except in outbreak situations. If exposure is known to have occurred, the use of normal human immunoglobulin might be considered in persons at risk of developing illness.

In the kitchen, prevention of transmission of viruses through foods largely depends on rigorous application of normal good hygiene practices, including frequent hand washing and thorough washing of fruit and vegetables. Shellfish must be regarded as a potential source of infection, and uncooked shellfish should be kept separate from other food items that are not to be cooked. If vomiting occurs, virus may be spread over a wide area. Uncovered food should be discarded. Even food that is to be cooked is a potential source of cross-contamination. Work surfaces, door handles, and toilet areas should be cleaned thoroughly. Chlorine-based disinfectants are considered the most effective. Handling of food should be kept to a minimum. Wearing gloves may prevent fecally contaminated fingers coming into contact with food but will not prevent transfer of organisms from contaminated work surfaces.

See also: **Contamination of Food**; **Food Poisoning**: Statistics; **Shellfish**: Contamination and Spoilage of Molluscs and Crustaceans; **Zoonoses**

Further Reading

Advisory Committee on the Microbiological Safety of Food (1998) *Report on foodborne viral infections*. London: The Stationery Office.

Appleton H (2000) Control of foodborne infections. *British Medical Bulletin* 56: 172–183.

Cliver DO (1997) Foodborne viruses. In: Doyle MP, Beuchat LR and Montville TJ (eds) *Food Microbiology: Fundamentals and Frontiers*, pp. 437–446. Washington, DC: ASM Press.

Desslberger U (2000) Viruses associated with acute diarrhoeal disease. In: Zuckerman AJ, Banatvala JE and Pattison JR (eds) *Principles and Practice in Clinical Virology*, 4th edn., pp. 235–252. Chichester, UK: John Wiley.

Evans HS, Madden P, Douglas C *et al.* (1998) General outbreaks of infectious intestinal disease in England and Wales: 1995 and 1996. *Communicable Disease and Public Health* 1: 165–171.

Food Standards Agency (2000) A report of the study of infectious intestinal disease in England. London: The Stationery Office.

Harrison TJ, Dusheiko GM and Zuckerman AJ (2000) Hepatitis viruses. In: Zuckerman AJ, Banatvala JE and Pattison JR (eds) *Principles and Practice in Clinical Virology*, 4th edn., pp. 187–233. Chichester, UK: John Wiley.

Hui YH, Sattar SA, Murrell KD, Nip W-K and Stanfield PS (eds) (2000) *Foodborne Disease Handbook, Volume 2: Viruses, Parasites, Pathogens, and HACCP*, 2nd edn. Chapters 2–9. New York: Marcel Dekker.

Lees D (2000) Viruses and bivalve shellfish. *International Journal of Food Microbiology* 56: 81–116.

Millard J, Appleton H and Parry JV (1987) Studies on heat inactivation of hepatitis A virus with special reference to shellfish. *Epidemiology and Infection* 98: 397–414.

Public Health Laboratory Service Viral Gastroenteritis Sub-Committee (1993) Outbreaks of gastroenteritis associated with SRSVs. *PHLS Microbiology Digest* 10: 2–8.

Seymour IJ and Appleton H (2001) A review: Foodborne viruses and fresh produce. *Journal of Applied Microbiology* 91: 759–773.

Slomka MJ and Appleton H (1998) Feline calicivirus as a model system for heat inactivation studies of small round structured viruses in shellfish. *Epidemiology and Infection* 121: 401–407.

Wheeler JG, Sethi D, Cowden JM *et al*. (1999) Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. *British Medical Journal* 318: 1046–1050.

Visible Spectroscopy *See* **Spectroscopy**: Overview; Infrared and Raman; Near-infrared; Fluorescence; Atomic Emission and Absorption; Nuclear Magnetic Resonance; Visible Spectroscopy and Colorimetry

Vitamin A *See* **Retinol**: Properties and Determination; Physiology

Vitamin B_1 *See* **Thiamin**: Properties and Determination; Physiology

Vitamin B_{12} *See* **Cobalamins**: Properties and Determination; Physiology

Vitamin B_2 *See* **Riboflavin**: Properties and Determination; Physiology

VITAMIN B_6

Contents
Properties and Determination
Physiology

Properties and Determination

J A Driskell, University of Nebraska, Lincoln, NE, USA

Vitamin B_6: Properties and Analysis

Vitamin B_6 is a B-complex vitamin. Vitamin B_6 is the generic descriptor for all 3-hydroxy-2-methylpyridine derivatives possessing the biological activity of pyridoxine.

Gyürgy, in 1934, identified vitamin B_6 as a curative factor for a characteristic dermatatis in the laboratory rat. The first naturally occurring form of the vitamin was isolated in 1938; its structure was confirmed by chemical synthesis as 3-hydroxy-4,5-bis(hydroxymethyl)-2-methylpyridine in 1939. 'Pyridoxine' was the trivial name of this compound, and 'pyridoxine' and 'vitamin B_6' were synonyms. Other compounds having vitamin B_6 activity were detected and isolated. In the 1960 IUPAC Definitive Rules for the Nomenclature of Vitamins, 'pyridoxine' was recommended as the generic descriptor of the B_6 vitamers and 'pyridoxol' as the trivial name of the alcohol form of the vitamin. In the 1966 IUPAC-IUB Tentative Rules, the suggestion was made that 'pyridoxol' should be designated as 'pyridoxine.' In 1968, the IUPAC-IUB Commission on Biochemical Nomenclature decided to publish a special document regarding vitamin B_6 nomenclature, which clearly stated that 'pyridoxine should not be used as a generic name synonymous with vitamin B_6'; this document, published in 1973, also designated the composition of the six forms of the vitamin and oxidized metabolites, 4-pyridoxic acid and 4-pyridoxolactone. However, confusion still exists in the literature regarding usage of the term 'pyridoxine.'

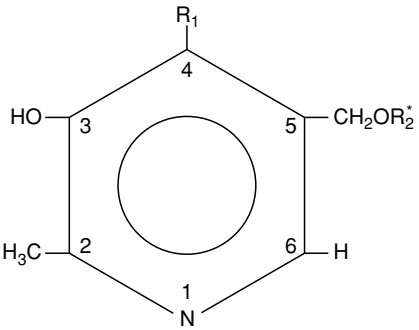

	R_1	R_2
PN	CH_2OH	H
PL	CHO	H
PM	CH_2NH_2	H
4-PA	COOH	H
PNP		PO_3^-

*When $-CH_2O{\cdot}PO_3^=$, then is PNP, PLP, or PMP.

Figure 1 Stuctural formulae of vitamin B_6 compounds.

The chemical structures of the vitamin B_6 compounds and the accepted ring numbering system are shown in Figure 1. The vitamin B_6 congeners are pyridoxine (PN), pyridoxal (PL), and pyridoxamine (PM) – alcohol, aldehyde, or amine groups are at the #4 position of the pyridine ring. Each of these three forms also has a corresponding 5′-phosphate (P), i.e., pyridoxine 5′-phosphate (PNP), pyridoxal 5′-phosphate (PLP), and pyridoxamine 5′-phosphate (PMP). The various forms of the vitamin are biochemically interconvertible, as shown in Figure 2. They can also be chemically interconverted. The major excretory form of the vitamin is 4-pyridoxic acid (4-PA), which is eliminated in the urine.

Physical and Chemical Properties

In general, the various vitamin B_6 compounds are white crystals that are very soluble in water, slightly soluble in alcohol, and often insoluble in ether. They melt, usually with decomposition. Aqueous solutions of the various vitamin B_6 compounds are relatively sensitive to light and often are destroyed by heating. The light sensitivity of the vitamers is affected by pH. The nonphosphorylated vitamers are relatively stable to heat at an acidic pH but not at an alkaline pH. According to Ang, stability values of PN·HCl, PL·HCl, and PM·2HCl at pH 4.5 to white light for 8 h are 97, 97, and 81%, respectively, which decreases with increasing pH. For example PN·HCl stability is 88% pH 7; at PL·HCl, 81% at pH 6; and PM·2HCl, 74% at pH 8. Overall, PN is the most stable, followed by PL and then PM. Greater

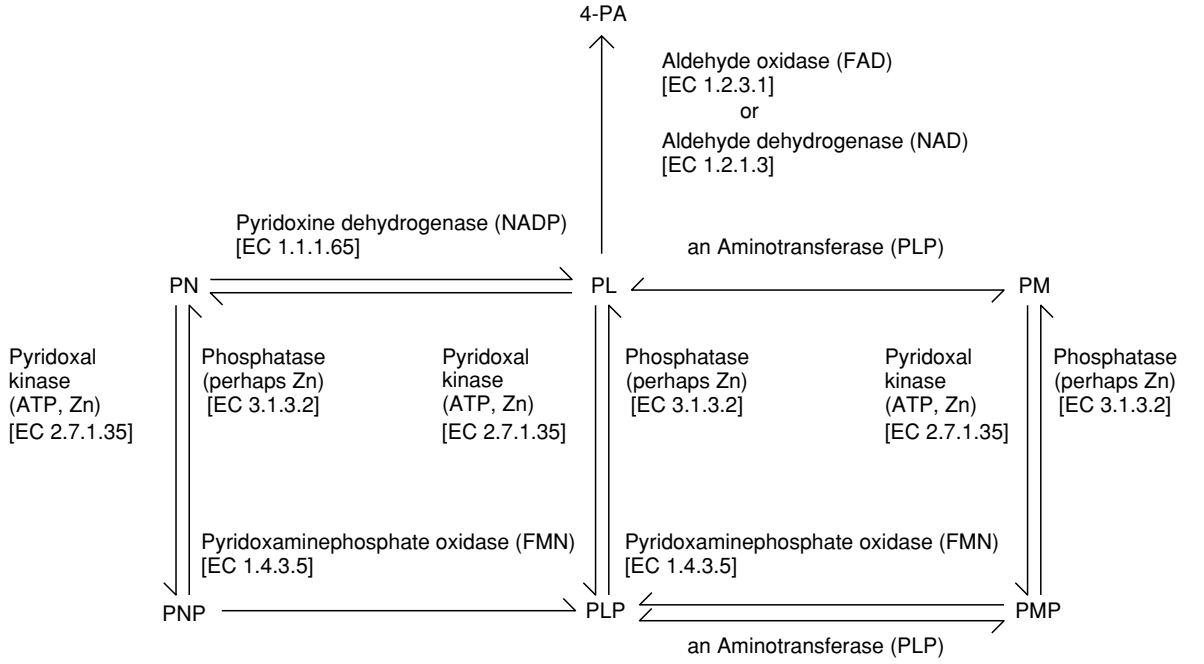

Figure 2 Interconversions of vitamin B_6 compounds.

Table 1 Physical properties of vitamin B_6 compounds[a]

		Ultraviolet absorption spectra							
		0.1 N HCl		*pH 7.0*		*Fluorescence maxima, pH 7*		*pK*	
Congener	*Molecular weight (g)*	λ_{max} *(nm)*	ε_{max}	λ_{max} *(nm)*	ε_{max}	*Activation (nm)*	*Emission (nm)*	pK_1	pK_2
PN	169.1	291	8700	254	3760	325	400	5.0	8.9
PL	167.2	288	9100	390	8800	320	385	4.2	8.7
PM	168.1	293	8500	253	4600	325	405	3.4	8.1
				325	7700				
PNP	249.2	290	8700	253	3700	322	394	–	–
				325	7400				
PLP	247.2	293	7200	388	5500	330	375	< 2.5	4.1
		334	1300						
PMP	248.2	293	9000	253	4700	330	400	< 2.5	3.5
				325	8300				
4-PA	183.2	–	–	–	–	325[b]	425[b]	–	–
						355[c]	445[c]		

[a]Data from Storvick CA, Benson EM, Edwards MA and Woodring MJ (1964) Chemical and microbiological determination of vitamin B_6. *Methods of Biochemical Analysis* XII: 183–276; Harris SA (1968) Pyridoxine. *Encyclopedia of Chemical Toxicology* 16: 806–824; Brin M (1978) Vitamin B_6: chemistry, absorption, metabolism, catabolism, and toxicity. In: *Human Vitamin B_6 Requirements*, pp. 1–20. Washington, DC: National Academy of Sciences; Ubbink JB (1992) Vitamin B_6. In: DeLeenheer AP, Lambert WE, and Nelis HJ (eds) *Modern Chromatographic Analysis of the Vitamins*, 2nd edn., Chapter 10. New York: Marcel Dekker.
[b]pH 3.4, 0.01 N acetic acid.
[c]pH 10.5, 0.1 N NH_4OH, lactone of 4-PA.

decomposition of the vitamin B_6 compounds is observed with ultraviolet and near-ultraviolet radiation. Researchers working with this nutrient generally perform their analyses in darkened laboratories (red or yellow lights may be used); in addition, media containing the vitamin are kept at temperatures of 5 °C or less.

The physical properties of the vitamin B_6 compounds are summarized in Table 1. The ultraviolet absorption spectra of these compounds are dependent upon the pH because of changes to the electronic structure. The pyridine ring of the vitamin B_6 compounds is zwitterionic at neutral pH with ionized phenolate and pyridinium groups. The physical properties of the B_6 compounds are useful with regard to detection and quantitation of the vitamin.

As mentioned previously, the various forms of vitamin B_6 are interconvertible. The various B_6 vitamers

can change from one form to another during food processing and food preparation, and in the body.

Occurrence and Forms in Foods

Information regarding the vitamin B_6 content of foods has been obtained primarily via microbiological assay, although some values are available which were obtained by high-performance liquid chromatography (HPLC) methodologies, which can differentiate between the vitamers. In the microbiological assay techniques, acid (generally hydrochloric acid) is used in extracting the vitamin from the food; the acid hydrolyzes phosphates. Hence, vitamer values obtained using microbiological methods represent the sum of the phosphorylated and nonphosphorylated forms of the vitamin. Frequently, total vitamin B_6 concentration values are obtained rather than values for the individual vitamers.

Food composition values are available for the three forms of the vitamin (PL + PLP, PN + PNP, and PM + PMP) in various foods. These can be found in the USDA's publication *Pantothenic Acid. Vitamin B_6 and Vitamin B_{12} in Foods*. In some reports, the individual phosphate and free-base forms are also available.

A high proportion of the B_6 vitamers in plant-derived foods is glycosylated PN, primarily as 5′-β-D-glucoside. PL, primarily as the phosphorylated form, is generally the predominant B_6 vitamer in animal-derived foods. Different foods vary somewhat as to which form of the vitamin is predominant.

Vitamin B_6 found in unprocessed foods is mainly bound to proteins. Table 2 gives the vitamin B_6 content of selected foods. This information can be obtained by accessing the USDA food composition website or from a variety of printed formats such as McCance and Widdowson. Meats, whole-grain products, and nuts are good sources of vitamin B_6. Cultured cheeses and other fermented products are also good sources of the vitamin. Fruits and vegetables are generally rather poor sources. PN, PM, and PL have been reported to have equal potencies in the rat. In humans, dietary PM and PL are about 10% less effective than PN, as indicated by effects on plasma PLP and urinary 4-PA levels according to a previous report. In the human body, the nonphosphorylated forms of vitamin B_6 can be converted to their respective phosphorylated forms and vice versa by pyridoxal kinase and pyridoxamine (pyridoxine) phosphate oxidase (see Figure 2). Recently, these kinase and oxidase enzymes have been identified as potential targets for pharmacologic agents.

Glycosylated vitamin B_6 is found in all plant-type foods. The glycosylated vitamin B_6 contents of selected foods can be found in various publications. Gregory indicated that a mixed diet contains about 15% of glycosylated vitamin B_6, which is about 50% as available as the other B_6 vitamers. This is often excluded from food composition data. Bioavailability is of importance in that it reflects the quantity of a consumed nutrient that becomes available to the cells of the body.

Effects of Food Processing and Preparation Methods upon Vitamin B_6 Content of Foods

Vitamin B_6 is one of the more labile nutrients. The vitamin is stable to dry heat; its stability in moist heat is largely dependent upon the pH of the media, being relatively stable to heat in an acidic media but not in alkaline media. The vitamin is unstable to light, both visible and ultraviolet. The vitamin undergoes rapid destruction or loss during cooking procedures.

Food processing and subsequent storage may also have an effect upon the vitamin B_6 content of various foods. The retention of the vitamin is quite high following the processing of dry foods. Thirty to 50% of the vitamin B_6 present in whole milk and infant formulas is retained following liquid processing including sterilization treatments, whereas the spray drying of these products results in minor loss. Storage of canned dairy products shows no degradation over 2–3 years. The same is true with regard to commercial canning and irradiation of boned chicken, although the retention is better for the irradiated than the canned product. Heat processing by retorting results in vitamin B_6 losses in several products. The stability of the vitamin in dehydrated foods is affected by moisture content, water activity, and storage conditions. Food-processing techniques generally result in losses of 0–50% of the vitamin.

Vitamin B_6 aldehydes can bind to food proteins as ε-pyridoxyllysine complexes during thermal processing and low-moisture storage. Gregory has shown that the protein-bound ε-pyridoxyllysine has antivitamin B_6 effects in rats. This pyridoxyllysine complex may also affect vitamin B_6 availability to humans, particularly with regard to the vitamin in unfortified infant formulas.

Much of the vitamin loss is due to cooking and holding practices of consumers and food service personnel. The retention of the vitamin in oven-braised beef rounds prepared in food service establishments has been reported to range from 45 to 62%; values for green beans are slightly higher, whereas those for baking potatoes prepared in the skins are around 80%. The vitamin B_6 content of meats and vegetables has been shown to decrease following various

Table 2 Vitamin B_6 content of selected foods

Food	Edible portion (mg per 100 g)	Food	Edible portion (mg per 100 g)
Meats			
Bacon	0.270	Ham, cured, roasted	0.310
Beef, ground, broiled, medium	0.270	Herring, Atlantic, cooked, dry heat	0.348
Beef, components, separable lean and fat, trimmed, choice, cooked	0.320	Lamb, composite, separable lean, choice, cooked	0.160
Bologna, beef	0.180	Liver, chicken, simmered	0.580
Bologna, pork	0.270	Mackerel, Atlantic, cooked, dry heat	0.460
Chicken		Oysters, Eastern, raw	0.050
Dark with skin, roasted	0.310	Pork, composite, separable lean, roasted	0.404
Light with skin, roasted	0.520	Pork sausage, fresh, cooked	0.330
Frankfurters, beef	0.110	Salmon, Sockeye, cooked, drained	0.219
Haddock, cooked, dry heat	0.346	Shrimp, breaded, fried	0.098
Halibut, cooked, dry heat	0.397	Tuna, canned in water, drained	0.350
Dairy products			
Butter	0.003	Eggs	0.139
Buttermilk	0.034	Milk	
Cheese		Whole	0.042
Cottage, creamed	0.067	2%	0.043
Cheddar	0.074	Skim	0.040
Swiss	0.083	Yogurt	0.032
Cereal and grain products			
Bread		Rice	
White	0.064	White, short-grain, cooked	0.059
Whole wheat	0.097	Brown, medium-grain, cooked	0.149
Pumpernickel	0.126	Wheat, germ, toasted, plain	0.978
Vegetables			
Asparagus, solids, drained	0.122	Collards, boiled, drained	0.128
Beans		Corn, sweet, boiled, drained	0.060
Lima, boiled, drained	0.193	Lettuce, looseleaf, raw	0.055
Snap, green, boiled, drained	0.056	Mushrooms, raw	0.102
Beets, boiled, drained	0.067	Onions, raw	0.116
Broccoli, boiled, drained	0.143	Peas, green, boiled, drained	0.216
Brussels sprouts, boiled, drained	0.178	Potatoes, boiled, without skin	0.269
Cabbage, boiled, drained	0.113	Potatoes, flesh	0.301
Carrots, raw	0.147	Potatoes, French-fried, restaurant-prepared	0.243
Cauliflower, boiled, drained	0.173	Spinach, boiled, drained	0.242
Celery, raw	0.087	Tomatoes, red, ripe, raw	0.080
Fruits			
Apples, raw	0.048	Orange, raw	0.060
Avocados, raw	0.280	Orange juice, fresh	0.040
Blueberries, raw	0.036	Peaches, raw	0.018
Cherries, sour, red raw	0.044	Pears, raw	0.018
Grapefruit, raw	0.042	Raisins, seedless	0.249
Grapes, American type, raw	0.110	Strawberries, raw	0.059
Nuts and seeds			
Almonds, dry roasted, unblanched	0.126	Peanuts, dry roasted	0.296
Cashew nuts, dry roasted	0.256	Pecans, dried	0.187
Coconut meat, dried, unsweetened	0.300	Walnuts, dried	0.537
Mixed nuts, dry roasted, with peanuts	0.296		
Alcoholic beverages			
Beer	0.050	Wine, table	0.024

Data from US Department of Agriculture, Agricultural Research Service (1999) *USDA Nutrient Database for Standard Reference, Release 13*.

household cooking procedures, even when foods were cooked according to standard recipes or manufacturer's instructions. The retention of vitamin B_6 in selected vegetables has been found to be highest for those steamed in plastic bags, followed by those boiled in plastic bags, those steamed, and lastly, for those boiled. Boiling results in degradation and leaching but the retention of the vitamin in vegetables following microwave steaming or regular microwave cooking techniques is around 90%. It has been found that more vitamin B_6 is retained when beef and pork strips are cooked by stir-frying ($\sim$67%) than

by microwaving (~60%) or by broiling (~58%). More vitamin B_6 is retained when pork roasts are cooked to an internal temperature of 71 °C (62%) than 82 °C (54%) and this retention is greatest when pork is cooked in a bag than roasted or braised. Vitamin B_6 retention has been found to be similar (68%) in bison patties cooked by grilling and by broiling. In summary, loss of some vitamin B_6 is inevitable during all household/food service preparation, to a greater or lesser extent.

Fortification of Foods with Vitamin B_6

PN·HCl is listed on the generally recognized as safe (GRAS) list of the US Food and Drug Administration and is used in the fortification of foods, particularly various cereal and dairy products. This form of the vitamin is also used in dietary supplements and pharmaceutical products. PN·HCl is resistant to destruction by heat as well as by other factors but should be protected from light and moisture. Pharmaceutical products containing this vitamin have been reported to have less than 10% loss in vitamin activity after storage at room temperature for 1 year.

Methods of Analysis

There are two basic stages in procedural methods dealing with quantitation of vitamin B_6 components – preparative or extraction techniques and determinative methods. Several excellent reviews have been published on methods of vitamin B_6 analysis including chromatographic techniques. Many methods have been utilized in quantitating the vitamin B_6 components in foods. These include animal and microbiological growth, enzymatic, fluorometric, immunologic, gas chromatographic (GC), and HPLC assays.

Extraction Techniques

The major problem encountered in quantitating vitamin B_6 in plant or animal products is extraction. Vitamin B_6 compounds must be extracted from food products before being analyzed by microbiological, chromatographic, and other methods. Vitamin B_6 frequently occurs in biological materials in association with proteins. The most common bond for PLP is the Schiff base linkage with the ε-amino group of lysine. The Schiff base linkage is easily hydrolyzed by deproteinizating agents.

Great care must be taken during sample handling and extraction and with relation to the stability of the different vitamers as well as their interconversions. The samples must be homogenous and representative of the bulk product.

Conventional methods for extracting vitamin B_6 compounds from foods entail the use of high temperature and acidic media to denature the protein and disrupt the sample matrix. The phosphate esters of PNP, PLP, and PMP are hydrolyzed by autoclaving in acidic media. The microorganisms used in microbiological assays lack response to the phosphates. Efficient extraction of vitamin B_6 compounds is obtained using a sample size containing 2–4 μg of the vitamin and hydrolyzing in 180 ml of 0.055 N H_2SO_4 for 1 h at 128 °C (20 psi, 138 kPa); 0.44 N sulfuric acid under the same conditions is essential for all plant products. Thermal extractions of pH 1.7–1.8 have been found to be optimally effective for most samples; this pH range corresponds closely to the 0.055 N sulfuric acid, which was recommended earlier. Hydrochloric acid is also used in thermal extractions of the vitamin. Complete hydrolysis of PLP is achieved after autoclaving the sample in 0.055 N hydrochloric acid at 121 °C (15 psi, 103 kPa) for 1 h; however, 5 h are required for extraction of the vitamin from dried yeast and liver powder. Complete hydrolysis of PMP requires 3 h at 125 °C in 0.055 N hydrochloric acid, whereas PLP is hydrolyzed completely in 30 min. The B_6 vitamers in rice bran are best hydrolyzed using 2 N hydrochloric acid. The AOAC method of thermal hydrolysis for B_6 vitamers is as follows: use 1–2 g of dry product; add 200 ml of 0.44 N hydrochloric acid for plant products followed by autoclaving at 121 °C for 2 h and for animal products, add 200 ml of 0.055 N hydrochloric acid, and autoclave at 121 °C for 5 h; after the solution has cooled to room temperature, adjust the pH to 4.5 with 6 N or saturated potassium hydroxide; then dilute to 250 ml with water and filter, using Whatman No. 40 paper. Complete release of both protein-bound and phosphorylated forms of the vitamin is essential for quantitative assay using microorganisms.

Perchloric acid, trichloroacetic acid, and sulfosalicylic acid may be used as deproteinating agents. These acids are particularly useful when one desires to quantitate phosphorylated, nonphosphorylated, and glycosylated forms of the vitamin individually, as use of these agents does not result in dephosphorylation or deconjugation. If so desired, trichloroacetic acid may be removed from the extract by extraction with diethyl ether or freon amine. Perchloric acid can also be removed from the extract by precipitation with potassium hydroxide; and sulfosalicylic acid can be removed via ion-exchange chromatographic techniques. Metaphosphoric acid has been used in extraction of the vitamin from brain tissue. This acid does not need removal from the extract prior to many of the HPLC techniques for vitamin B_6 assay.

Extraction procedures must solubilize the vitamers for subsequent quantitation. The effectiveness of the extraction procedures may be determined by evaluation of recoveries of B_6 vitamers added during sample homogenization or, even better, by incorporation of radiolabeled forms of the vitamin into the sample, called intrinsic labeling.

Animal Growth Assays

Bioassays are based on the growth response of animals, usually rats or chicks, to purified diets containing graded levels of a standard vitamin to the sample being tested. An advantage of the method is that the vitamin in the sample does not have to be extracted. Animal assays are not used very frequently because of their lack of specificity as well as their expense and time requirement. The intestinal microflora of these animals may affect the data; efforts to prevent coprophagy or the use of antibiotics are not completely effective.

Microbiological Assays

The vitamin B_6 composition of plant and animal products is generally determined using microbiological assays. These assays are tedious, and none are without problems. The most widely used microbiological assay is a growth assay using the yeast *Saccharomyces uvarum* (ATCC 9080; formerly *S. carlsbergenesis*). Growth is measured via turbidity using a spectrophotometer. Unfortunately, the response of many cultures of the yeast to PM is frequently 60–80% of that to PN and PL. It is advisable to evaluate the magnitude of the response differences under the conditions of each assay. PL, PN, and PM can be separated by subjecting the acid extractant to cation-exchange chromatography on Dowex AG 50W-X8 resin using KOAc solutions of three different pHs; each form of the vitamin is then analyzed, utilizing its respective calibration curve. Care must be taken to assure proper separation of the vitamers and avoid cross-contamination of the three fractions.

The yeast *Kloeckera brevis* (ATCC 9774; formerly *K. apiculata*) may be used in quantitating total vitamin B_6. Some researchers indicate that *K. brevis* has an equal growth response to PL, PN, and PM, whereas others indicate that it has an even lower relative response to PM than *S. uvarum*. The *K. brevis* assay has been adapted to radiometric quantitation.

The total, as well as the three nonphosphorylated vitamers, may also be determined by a differential microbiological technique. *S. uvarum* responds to PL, PN, and PM, whereas *Streptococcus faecalis* R (ATCC 8043) responds to PL and PM, and *Lactobacillus casei* (ATCC 7469) responds to only PL. *Tetrahymena pyriformis*, a protozoan, has also been used as an assay organism.

Saccharomyces uvarum is the microorganism used in the AOAC method. The concentrations of the three nonphosphorylated B_6 vitamers are quantitated in this method.

Assays for glycosylated vitamin B_6 compounds in foods have been developed utilizing *S. uvarum* and differential hydrolytic procedures. This method does have a tendency to underestimate the quantity of the glycosylated vitamin in raw plant foods having *p*-glycosidase activity.

Values given in food composition tables have been derived primarily using *S. uvarum* as the test organism. These tables generally list total vitamin B_6 concentrations but may list PL, PN, and PM concentrations. The Center for Nutrient Analysis of the US Food and Drug Administration uses *S. uvarum* for quantitating vitamin B_6 in foods. Data obtained microbiologically using *S. uvarum* do not always agree with those obtained by animal growth assays.

Enzymatic Assays

Several enzymatic methods are available that quantitate PLP, the main circulating form of vitamin B_6. These methods are frequently utilized as parameters for the assessment of vitamin B_6 status of animals, including humans but are not utilized in the analysis of food samples. The enzymatic methods that have been more commonly used include the following: the cleavage of tryptophan using apotryptophanase (EC 4.1.99.1), the measurement of aspartate aminotransferase (EC 2.6.1.1; formerly referred to as glutamic-oxaloacetic transaminase), and alanine aminotransferase (EC 2.6.1.2; formerly referred to as glutamic-pyruvic transaminase) activities, and the radiomonitored decarboxylation of [^{14}C]tyrosine using the apoenzyme of tyrosine decarboxylase (EC 4.1.1.25). Currently, the most acceptable enzymatic method of quantitating PLP is the tyrosine apodecarboxylase technique.

Fluorometric Assays

Direct fluorometric analysis of vitamin B_6 in extracts of plant and animal products is generally unsatisfactory. This is due to the presence of many potentially interfering compounds that fluoresce, as well as the varying spectral characteristics of the B_6 vitamers. As a result, the total vitamin B_6 content of selected foods is higher when measured by fluorometry following open-column chromatography than published values based on microbiological assay.

Immunologic Assays

Antibodies specific for (phospho)pyridoxyl groups have been utilized in immunoassays for certain B_6 vitamers. However, there is a marked variation in the affinity of the antibodies for the B_6 vitamers, so these methods are not widely used.

Gas Chromatographic Assays

Unless B_6 vitamers are nonphosphorylated and their polar functional groups derivatized, these compounds are not well suited for GC analysis. Nonvolatile components and water are not easily removed from sample extracts. Methods involving the formation of acetyl, isopropylidene, trimethylsilyl, trifluoroacetyl, and heptafluorobutyryl derivatives of the nonphosphorylated B_6 vitamers have been developed, but other chromatographic methods are preferred.

HPLC Assays

The physiochemical properties of the B_6 vitamers facilitate their assay by HPLC assay. HPLC has the potential for high resolution and high sensitivity with regard to the B_6 vitamers found in biological materials, without the necessity of complex extraction and derivatization techniques. If desired, the phosphate esters of the phosphorylated B_6 vitamers can be hydrolyzed using phosphatase yielding nonphosphorylated vitamers. Otherwise, the phosphorylated vitamers can be preserved during extraction and quantitated as such. All six B_6 vitamers and 4-PA can be quantitated at nanogram levels using HPLC techniques. Alternatively, all forms of the vitamer can also be converted to PN, followed by HPLC quantitation of total vitamin B_6 as PN techniques. This method is rapidly becoming the preferred way of determining total vitamin B_6. Ion-exchange (anion or cation) and reversed-phase (with or without paired-ion) HPLC methods for assay of B_6 vitamers have been published. Most of the methods utilize fluorescence detectors, although some methods, particularly the earlier methods, do utilize ultraviolet (UV) detectors; the fluorescence detector is far more sensitive and specific. The use of an internal standard, most frequently 4-deoxypyridoxine, increases the accuracy and precision of HPLC analysis of B_6 vitamers, providing that it is not naturally present in the samples. External standards (spiking) are also utilized in proving identities of the peaks. The identities of the peaks should be confirmed by other means such as mass spectrometry and fluorescence spectra.

Ion-exchange chromatography is accomplished using packing material that possesses charge-bearing functional groups. Many of the HPLC separations reported in the 1970s were cation- or anion-exchange methodologies, which generally involved determinations of nonphosphorylated B_6 vitamers in acid hydrolysates. An HPLC method has been developed for the separation of all of the B_6 vitamers using Aminex A-25 anion-exchange resin, isocratic elution, and a pH 10 glycine buffer. This anionic method was successfully applied to the determination of vitamin B_6 components in milk, animal tissues, meats, and fortified cereals. Unfortunately, the manufacturer altered the composition of the resin in such a manner that the method no longer provided satisfactory separation of the vitamers. Also, a cation-exchange HPLC method, fluorometric detection, has been developed, which successfully quantitated all B_6 vitamers in plasma and other animal tissues.

Reversed-phase chromatography is the term given to chromatographic conditions in which a nonpolar stationary phase is used in conjunction with a polar mobile phase. Gregory and Kirk were the first to separate nonphosphorylated B_6 vitamers using reversed-phase HPLC methods, with fluorescence detection, which successfully measured PN in fortified cereals.

Paired-ion (also called ion-pair) chromatography was developed to deal with separating compounds that are either very polar, multiply ionized, or strongly basic. Several paired-ion, reversed-phase HPLC methods, both gradient and isocratic, have been developed for determining vitamin B_6 components in biological materials and foods. These methods have also been utilized successfully in quantitating the six B_6 vitamers and 4-PA in plasma. This type of methodology may also be utilized in quantitating glycosylated and other forms of the vitamin.

Several researchers have reported comparable results when specific foods were analyzed for total vitamin B_6 vitamer content using microbiological (*S. uvarum*) and HPLC techniques. Most researchers in the vitamin B_6 area believe that HPLC is the preferred method for determining the vitamin B_6 content of food because individual vitamers can be determined. However, as indicated earlier, vitamin B_6 content data for food composition tables are obtained primarily using microbiological (*S. uvarum*) methods. As discussed earlier, the potencies of the B_6 vitamers in the human body appear to be rather similar, so one questions the importance of routinely determining each of the six B_6 vitamers in foods. However it does have academic interest and is of clinical significance where drugs or diseases/conditions influence the metabolism of the B_6 vitamers. Information as to the form of the vitamin in food may be of importance therapeutically.

Summary

The quantitation of vitamin B_6 components in biological materials is difficult. The vitamin must be extracted, generally using acidic media, from plant and animal products before being analyzed by one of the many chemical, biochemical or biological methods. The official AOAC method for determination of the vitamin B_6 content of foods involves the use of the microorganism *Saccharomyces uvarum*. Although researchers have developed HPLC methods that quantitate all six forms of the vitamin, these methods are rarely utilized in determining the vitamin B_6 content of food. Little difference appears to exist in bioavailability of the different B_6 vitamers in humans so total vitamin B_6 composition is determined by conversion to pyridoxine. Researchers are continuing to seek improved methodologies for the analysis of vitamin B_6 involving use of the latest technology such as mass spectrometry. Additional research is also needed to determine the importance of specific forms of the vitamin in humans.

See also: **Bioavailability of Nutrients**; **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **Food Fortification**

Further Reading

Ang CY (1979) Stability of three forms of vitamin B_6 to laboratory light conditions. *Journal of the Association of Official Analytical Chemists* 62: 1170–1173.

Association of Official Analytical Chemists (2000) *Official Methods of Analysis*, 19th edn. Arlington, VA: AOAC.

Batenhorst JH, Sun J, Giraud DW *et al*. (1995) Retention of selected nutrients in pork roasts prepared using different cookery methods. *Journal of Muscle Foods* 6: 359–368.

Brin M (1978) Vitamin B_6: chemistry, absorption, metabolism, catabolism, and toxicity. In: *Human Vitamin B_6 Requirements*, pp. 1–20. Washington, DC: National Academy of Sciences.

Chrisley BMc, Thye FW, McNair HM and Driskell JA (1988) Plasma B-6 vitamer and 4-pyridoxic acid concentrations of men fed controlled diets. *Journal of Chromatography* 428: 35–42.

DeRitter E (1976) Stability characteristics of vitamins in processed foods. *Food Technology* 30: 48–54.

Driskell JA (1994) Vitamin B-6 requirements of humans. *Nutrition Research* 14: 293–324.

Driskell JA (1995) Sensory and nutritive qualities of food prepared by microwaving as compared to other cooking methods. *Microwave World* 16(2): 6–9.

Eitenmiller RR and Landen WO, Jr (1999) *Vitamin Analysis for the Health and Food Sciences*, Chapter 10. Boca Raton, FL: CRC Press.

Gregory JF (1980) Effects of ε-pyridoxyllysine bound to dietary protein on the vitamin B-6 status of rats. *Journal of Nutrition* 110: 995–1005.

Gregory JF (1988) Methods for determination of vitamin B_6 in foods and other biological materials: a critical review. *Journal of Food Composition and Analysis* 1: 105–123.

Gregory JF (1990) The bioavailability of vitamin B_6: recent findings. *Annals of the New York Academy of Sciences*, pp. 86–95. New York: New York Academy of Sciences.

Gregory JF (1997) Bioavailability of vitamin B-6. *European Journal of Clinical Nutrition* 51: S43–S48.

Harris SA (1968) Pyridoxine. *Encyclopedia of Chemical Technology* 16: 806–824.

International Union of Pure and Applied Chemistry – International Union of Biochemistry, Commission on Biochemical Nomenclature (1973) Nomenclature for vitamin B_6 and related compounds: recommendations 1973. *European Journal of Biochemistry* 40: 325–327.

Johnson JA and Driskell J (1987) Vitamin retention and sensory evaluation of selected foods prepared by boiling, boiling-in-bag, and steaming methods. *Home Economics Research Journal* 16: 136–142.

Leklem JE and Reynolds RD (eds) (1981) *Methods in Vitamin B-6 Nutrition: Analysis and Status Assessment*. New York: Plenum Press.

Leklem JE (1991) Vitamin B_6. In: Machlin LJ (ed.) *Handbook of Vitamins*, 2nd edn., Chapter 9. New York: Marcel Dekker.

McCance RA, Widdowson EM and Holland B (1993) *The Composition of Foods*, 5th edn. London: Springer Verlag.

McCormick DB and Chen H (1999) Update on interconversions of vitamin B-6 with its coenzyme. *Journal of Nutrition* 129: 325–327.

Orr ML (1969) *Pantothenic Acid, Vitamin B_6 and Vitamin B_{12} in Foods. US Department of Agriculture, Home Economics Research Report No. 36*. Washington, DC: US Government Printing Office.

Reiter LA and Driskell JA (1985) Vitamin B-6 content of selected foods served in dining halls. *Journal of The American Dietetic Association* 85: 1625–1627.

Sauberlich HE (1999) *Laboratory Tests for the Assessment of Nutritional Status*. Boca Raton, FL: CRC Press.

Storvick CA, Benson EM, Edwards MA and Woodring MJ (1964) Chemical and microbiological determination of vitamin B_6. *Methods of Biochemical Analysis* XII: 183–276.

Ubbink JB (1992) Vitamin B_6. In: DeLeenheer AP, Lambert WE and Nelis HJ (eds) *Modern Chromatographic Analysis of the Vitamins*, 2nd edn., Chapter 10. New York: Marcel Dekker.

US Department of Agriculture, Agricultural Research Service (1999) *USDA Nutrient Database for Standard Reference, Release 13*. Nutrient Data Laboratory Home Page, http://www.nal.usda.gov/fnic/foodcomp.

Yang J, Sulaeman A, Setiawan B *et al*. (1994) Sensory qualities and nutrient retention of beef strips prepared by different household cooking techniques. *Journal of The American Dietetic Association* 94: 199–201.

Yang J, Sulaeman A, Setiawan B *et al*. (1994) Sensory and nutritive qualities of pork strips prepared by three household cooking techniques. *Journal of Food Quality* 17: 33–40.

Yuan X, Marchello MJ and Driskell JA (1999) Selected vitamin contents and retentions in bison patties as related to cooking method. *Journal of Food Science* 64: 462–464.

Physiology

D A Bender, University College London, London, UK

Introduction

Vitamin B_6 has a central role in amino acid metabolism, as the coenzyme for a variety of reactions, including transamination and decarboxylation. It is also the coenzyme of glycogen phosphorylase, and acts to modulate the activity of steroid and other hormones (including retinoids and vitamin D) which act by regulation of gene expression.

Dietary Forms, Biological Availability, and Metabolism

The main form of vitamin B_6 in foods is pyridoxal phosphate, bound to enzymes. There is also a small amount of pyridoxamine phosphate. In plant foods a significant amount of the vitamin is present as pyridoxine.

A number of plants contain relatively large amounts of pyridoxine glycosides, which are not biologically available, since they are not substrates for mammalian glycosidases. They are absorbed (passively) from the intestinal lumen and are excreted more or less quantitatively in the urine. Between 5% and 50% of the total vitamin B_6 in some foods may be present as pyridoxine glycosides.

A proportion of the vitamin B_6 in foods may be biologically unavailable, especially after heating, as a result of the formation of (phospho)pyridoxyllysine by reduction of the aldimine (Schiff base), by which pyridoxal phosphate is bound to the ε-amino groups of lysine residues in proteins. While some of this pyridoxyllysine may be useable, since it is a substrate for pyridoxine phosphate oxidase, it is also a vitamin B_6 antimetabolite, and even at relatively low concentrations can accelerate the development of deficiency in experimental animals maintained on deficient diets. In the 1950s there was an outbreak of vitamin B_6 deficiency among infants fed on formula that had been overheated in manufacture, resulting in the formation of relatively large amounts of pyridoxyllysine.

Digestion and Absorption

Pyridoxal phosphate bound as a Schiff base to lysine in dietary proteins is released on digestion of the protein. The phosphorylated vitamers are dephosphorylated by membrane-bound alkaline phosphatase in the intestinal mucosa; pyridoxal, pyridoxamine, and pyridoxine are all absorbed rapidly by passive diffusion. Intestinal mucosal cells have pyridoxine kinase, and pyridoxine phosphate oxidase so that there is net accumulation by metabolic trapping. Much of the ingested pyridoxine is released into the portal circulation as pyridoxal, after dephosphorylation at the serosal surface.

Metabolism and Transport

Most of the absorbed vitamin is taken up by the liver, although other tissues can also take up the unphosphorylated vitamers from the circulation. Uptake is by carrier-mediated diffusion, followed by metabolic trapping as phosphate esters. Pyridoxine and pyridoxamine phosphates are oxidized to pyridoxal phosphate. All tissues have pyridoxine kinase activity, but pyridoxine phosphate oxidase is only found in liver, kidney, and brain (**Figure 1**).

Pyridoxine phosphate oxidase is a flavoprotein, and its activity falls markedly in riboflavin deficiency. Despite this central role of riboflavin in vitamin B_6 metabolism, blood and tissue concentrations of pyridoxal phosphate are not affected by riboflavin deficiency, and riboflavin nutrition appears to have no effect on vitamin B_6 nutritional status.

Pyridoxine phosphate oxidase is inhibited by its product, pyridoxal phosphate. This is not simple product inhibition, but involves binding at a specific inhibitor site on the enzyme. The normal intracellular concentration of free pyridoxal phosphate gives significant inhibition, which indicates that this is a physiologically important mechanism in the control of tissue pyridoxal phosphate.

Pyridoxine is phosphorylated rapidly in liver and other tissues. Pyridoxal phosphate does not cross cell membranes, and uptake and efflux of the vitamin in most tissues are as pyridoxal. Pyridoxal phosphate is exported from the liver bound to albumin. Much of the free pyridoxal phosphate in the liver is hydrolyzed to pyridoxal, which is also exported, and circulates bound both to albumin and to hemoglobin in erythrocytes. Free pyridoxal remaining in the liver is rapidly oxidized to 4-pyridoxic acid, which is the main excretory product of the vitamin.

Extrahepatic tissues take up pyridoxal from the plasma. Pyridoxal phosphate is hydrolyzed to pyridoxal, which can cross cell membranes, by

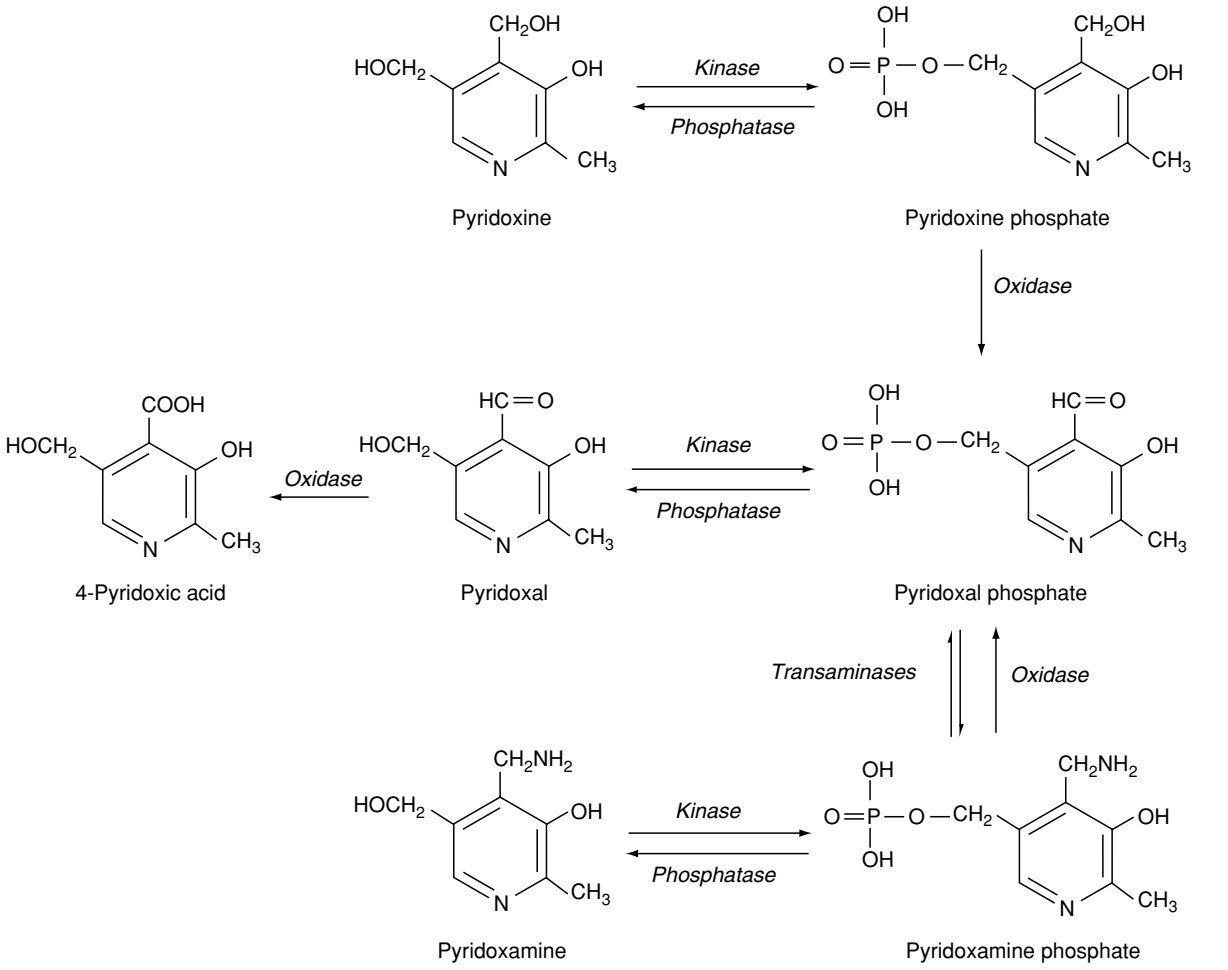

Figure 1 Metabolism of vitamin B_6.

extracellular alkaline phosphatase, then trapped intracellularly by phosphorylation.

Tissue concentrations of pyridoxal phosphate are controlled by the balance between phosphorylation and dephosphorylation. The activity of phosphatases acting on pyridoxal phosphate is greater than that of the kinase in most tissues. This means that pyridoxal phosphate that is not bound to enzymes will be dephosphorylated and hence leave the cell by diffusion. Thus there is little accumulation of pyridoxal phosphate in tissues, other than that which is bound to enzymes and other proteins (e.g., hormone receptors).

Free pyridoxal either leaves the cell or is oxidized to 4-pyridoxic acid by aldehyde dehydrogenase, which is present in all tissues, and also by hepatic and renal aldehyde oxidase. 4-Pyridoxic acid is the main excretory product of vitamin B_6, and its excretion reflects recent intake more than the state of underlying tissue reserves of the vitamin. Small amounts of pyridoxal and pyridoxamine are also excreted in the urine, although much of the active vitamin B_6 which is filtered at the glomerulus is resorbed in the kidney tubules.

Storage and Body Reserves

There is no specific storage of vitamin B_6 in the body; as discussed above, pyridoxal phosphate that is not bound to enzymes is rapidly dephosphorylated, oxidized to 4-pyridoxic acid, and excreted.

The total body pool of vitamin B_6 is of the order of 60 μmol (250 mg); 15 μmol (3.7 mg) per kg body weight. About 80% of this is in muscle, associated with glycogen phosphorylase. This does not seem to function as a true reserve of the vitamin and is not released from muscle in times of deficiency.

Muscle pyridoxal phosphate is released into the circulation (as pyridoxal) in starvation, as muscle glycogen reserves are exhausted, and there is less requirement for glycogen phosphorylase activity. Under these conditions it is available for redistribution to other tissues, especially liver and kidney, to

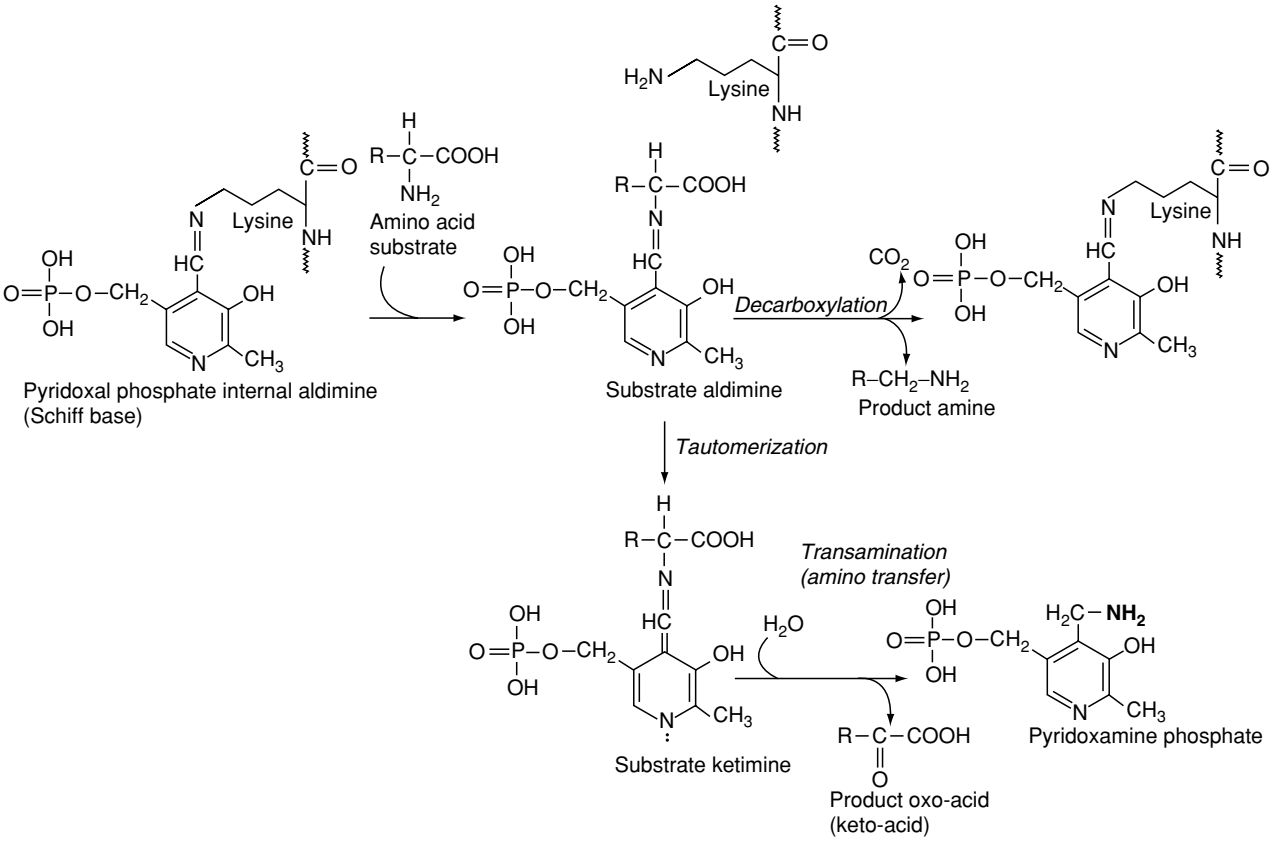

Figure 2 Roles of vitamin B_6 in amino acid metabolism.

meet the increased requirement of transamination of amino acids for gluconeogenesis.

Metabolic Functions of Vitamin B_6

The metabolically active vitamer is pyridoxal phosphate, which is involved in many reactions of amino acid metabolism, where the carbonyl group is the reactive moiety, in glycogen phosphorylase, where it is the phosphate group that is important in catalysis, and in the release of hormone receptors from tight nuclear binding, where again it is the carbonyl group that is important.

The Role of Pyridoxal Phosphate in Amino Acid Metabolism

The various reactions of pyridoxal phosphate in amino acid metabolism (Figure 2) all depend on the same chemical principle – the ability to stabilize amino acid carbanions, and hence to weaken bonds about the α-carbon of the substrate. This is achieved by reaction of the α-amino group with the carbonyl group of the coenzyme to form a Schiff base (aldimine).

Pyridoxal phosphate is bound to enzymes, in the absence of the substrate, by the formation of an internal Schiff base to the ε-amino group of a lysine residue at the active site. Thus the first reaction between the substrate and the coenzyme is transfer of the aldimine linkage from this ε-amino group to the α-amino group of the substrate.

The ring nitrogen of pyridoxal phosphate exerts a strong electron-withdrawing effect on the aldimine, and this leads to weakening of all three bonds about the α-carbon of the substrate. In nonenzymic model systems, all the possible pyridoxal-catalyzed reactions are observed: α-decarboxylation, aminotransfer, racemization, and relevant side-chain elimination and replacement reactions. By contrast, enzymes show specificity for the reaction pathway followed; which bond is cleaved will depend on the orientation of the Schiff base relative to reactive groups of the catalytic site. However, a number of decarboxylases and enzymes that catalyze side-chain elimination reactions of amino acids undergo gradual inactivation as a result of catalyzing the half-reaction of transamination, leaving (catalytically inactive) pyridoxamine phosphate at the catalytic site.

α-Decarboxylation If the electron-withdrawing effect of the ring nitrogen is primarily centered on

the α-carbon-carboxyl bond, the result is decarboxylation of the amino acid with the release of carbon dioxide. The resultant carbanion is then protonated, and the primary amine corresponding to the amino acid is displaced by the lysine residue at the active site, with reformation of the internal Schiff base.

A number of the products of the decarboxylation of amino acids are important as neurotransmitters and hormones: 5-hydroxytryptamine, the catecholamines dopamine, norepinephrine (noradrenaline), and epinephrine (adrenaline), histamine and γ-aminobutyrate (GABA), and as the diamines and polyamines involved in the regulation of DNA metabolism. The decarboxylation of phosphatidylserine to phosphatidylethanolamine is important in phospholipid metabolism.

Racemization of Amino Acids Deprotonation of the α-carbon of the amino acid leads to tautomerization of the Schiff base to yield a quinonoid ketimine. The simplest reaction that the ketimine can undergo is reprotonation at the now symmetrical α-carbon. Displacement of the substrate by the reactive lysine residue results in the racemic mixture of D- and L-amino acid.

Amino acid racemases are important in bacterial metabolism, since several D-amino acids are required for the synthesis of cell wall mucopolysaccharides. There is no evidence that there are any mammalian amino acid racemases; such utilization of D-amino acids as occurs is probably due to the action of renal D-amino acids oxidase to form the symmetrical 2-oxo-acid which is then a substrate for transamination.

Transamination Hydrolysis of the α-carbon-amino bond of the ketimine formed by deprotonation of the α-carbon of the amino acid results in the release of the 2-oxo-acid corresponding to the amino acid substrate, and leaves pyridoxamine phosphate at the catalytic site of the enzyme. In this case there is no reformation of the internal Schiff base to the reactive lysine residue. This is the half-reaction of transamination. The process is completed by reaction of pyridoxamine phosphate with a second oxo-acid substrate, forming an intermediate ketimine, followed by the reverse of the reaction sequence shown in Figure 2, releasing the amino acid corresponding to this second substrate after displacement from the aldimine by the reactive lysine residue to reform the internal Schiff base.

Transamination (Figure 3) is of central importance in amino acid metabolism, providing pathways for the catabolism of all of the amino acids except lysine, which does not undergo transamination. Many of these reactions are linked to the amination of 2-oxoglutarate to glutamate or glyoxylate to glycine, which are substrates for oxidative deamination, reforming the oxo-acids. Transamination reactions also provide a pathway for the synthesis of those amino acids for which there is an alternative source of the oxo-acid (the nonessential amino acids). Indeed, the nonessential amino acids can be defined as those whose oxo-acids can be formed other than from the amino acid itself.

Side-chain elimination and replacement reactions The third bond in the Schiff base aldimine that can be labilized by the electron-withdrawing

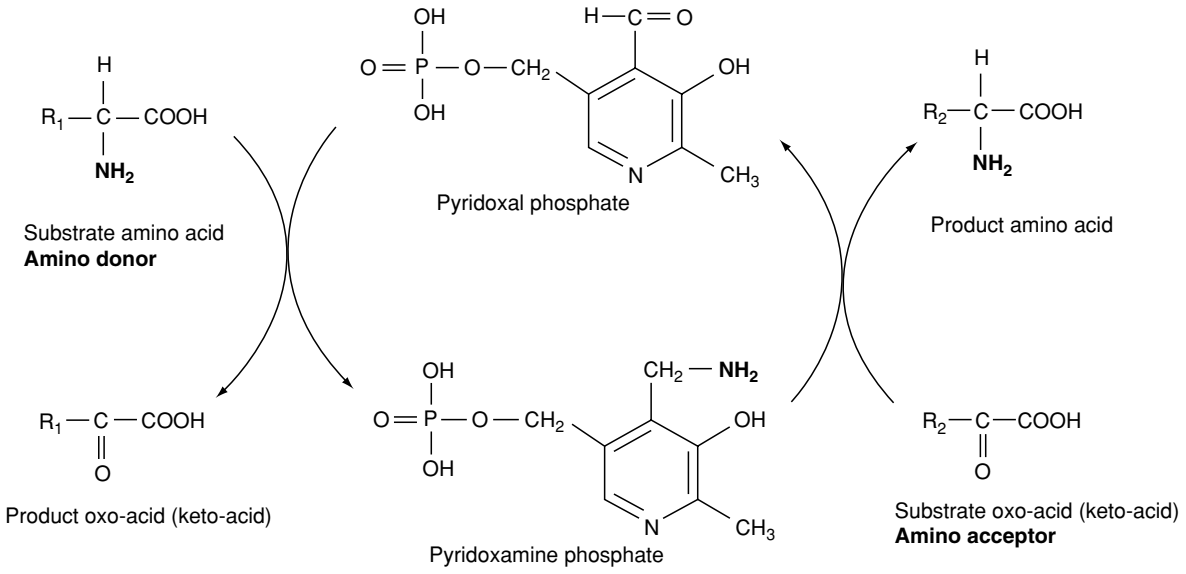

Figure 3 The role of vitamin B_6 in transamination reactions.

effect of the ring nitrogen of pyridoxal phosphate is that between the α-carbon and the side-chain of the amino acids, resulting in a variety of α–β elimination and β–γ replacement reactions.

The Role of Pyridoxal Phosphate in Glycogen Phosphorylase

Glycogen phosphorylase catalyzes the sequential phosphorolysis of glycogen to release glucose 1-phosphate; it is thus the key enzyme in the utilization of muscle and liver reserves of glycogen.

Unlike other pyridoxal phosphate-dependent enzymes, in which the carbonyl group is essential for catalysis, the internal Schiff base between pyridoxal phosphate and lysine in glycogen phosphorylase is not broken during the reaction. The catalytic region of the coenzyme is the 5′-phosphate group. The initial stage in the phosphorolysis of glycogen is protonation of the glycosidic oxygen of the polysaccharide by inorganic phosphate. The resultant oxycarbonium ion is stabilized by the inorganic phosphate. The role of pyridoxal phosphate is as a proton shuttle or buffer to stabilize the oxycarbonium-phosphate ion pair, permitting covalent binding of the phosphate to the oxycarbonium ion, to form glucose 1-phosphate.

The Role of Pyridoxal Phosphate in Steroid Hormone Action

Pyridoxal phosphate has a role in modulating the action of those hormones which act by binding to a nuclear receptor protein, inducing transcription of DNA, and hence regulating gene expression, leading to new protein synthesis. Such hormones include androgens, estrogens, progesterone, glucocorticoids, calcitriol (the active metabolite of vitamin D), retinoic acid and other retinoids, and thyroid hormone. Target-tissue specificity of hormone action is insured by the presence of characteristic receptor proteins with a zinc finger motif, which are responsible for both nuclear uptake of the hormone and the interaction with DNA and nucleoproteins to initiate gene expression.

Pyridoxal phosphate reacts with a lysine residue in the receptor protein, and displaces the hormone–receptor complex from tight nuclear binding. *In vitro*, reaction with pyridoxal phosphate also inhibits the binding of receptor protein to isolated DNA and chromatin. The effect is specific for the phosphorylated vitamer, suggesting that there may be a specific pyridoxal-phosphate binding site on the receptor proteins, and it occurs at low concentrations of pyridoxal phosphate, of the same order of magnitude as occur in tissues under normal conditions.

This suggests that pyridoxal phosphate acts as a cofactor in the release of hormone–receptor complexes from tight nuclear binding, resulting in release of the receptor from the nucleus, termination of hormone action, and recycling receptor protein for further uptake of hormone.

In experimental animals, vitamin B_6 deficiency results in increased and prolonged nuclear uptake and retention of steroid hormones in target tissues, and there is enhanced sensitivity to hormone action. Deficient animals show greater induction of uterine peroxidase, and considerably greater suppression of the hypothalamic secretion of luteinizing hormone, by estrogens than do vitamin B_6-supplemented controls. In vitamin B_6-deficient male animals there is an increased mitotic response of the prostate to low doses of testosterone. Deficient male animals have a higher activity of ornithine decarboxylase (an androgen-induced enzyme) in the liver, and deficient females have higher renal ornithine transaminase (an estrogen-induced enzyme). The induction of hepatic tyrosine transaminase and tryptophan dioxygenase by glucocorticoids is also enhanced in vitamin B_6-deficient animals.

In a variety of cells in culture that have been transfected with a glucocorticoid, estrogen, or progesterone response element linked to a reporter gene, acute vitamin B_6 depletion (by incubation with 4-deoxypyridoxine) leads to a twofold increase in expression of the reporter gene in response to hormone action. Conversely, incubation of these cells with high concentrations of pyridoxal, leading to a high intracellular concentration of pyridoxal phosphate, results in a halving of the expression of the reporter gene in response to hormone stimulation.

Criteria of Adequacy and Assessment of Nutritional Status

Plasma Concentrations of the Vitamin

The fasting plasma concentration of either total vitamin B_6 or, more specifically, pyridoxal phosphate, is widely used as an index of vitamin B_6 nutritional status. The generally accepted criteria of adequacy are shown in Table 1.

Conditions that involve increased plasma activity of alkaline phosphatase may result in reduced plasma pyridoxal phosphate, without affecting tissue concentrations of pyridoxal phosphate or vitamin B_6 nutritional status, as assessed by other criteria. There is a compensatory increase in the circulating concentration of pyridoxal, which is the main form for tissue uptake of vitamin B_6. Despite the fall in plasma pyridoxal phosphate in pregnancy, which has been

Table 1 Indices of vitamin B_6 nutritional status

	Adequate status
Plasma total vitamin B_6	> 40 nmol (10 μg) l^{-1}
Plasma pyridoxal phosphate	> 30 nmol (7.5 μg) l^{-1}
Erythrocyte alanine aminotransferase activation coefficient	< 1.25
Erythrocyte asparate aminotransferase activation coefficient	< 1.80
Erythrocyte asparate aminotransferase	> 0.13 units (8.4 μkat) l^{-1}
Urine 4-pyridoxic acid	> 3.0 μmol $24\,h^{-1}$ > 1.3 mmol mol^{-1} creatinine
Urine total vitamin B_6	> 0.5 μmol $24\,h^{-1}$ > 0.2 mmol mol^{-1} creatinine
Urine xanthurenic acid after 2 g tryptophan load	< 65 μmol $24\,h^{-1}$ increase
Urine cystathionine after 3 g methionine load	< 350 μmol $24\,h^{-1}$ increase

widely interpreted as indicating vitamin B_6 depletion, the plasma concentration of (pyridoxal phosphate plus pyridoxal) is unchanged. This suggests that determination of plasma pyridoxal phosphate alone may not be a reliable index of vitamin B_6 nutritional status.

Urinary Excretion of 4-Pyridoxic Acid

About half of the normal dietary intake of vitamin B_6 is excreted as 4-pyridoxic acid. Urinary excretion of 4-pyridoxic acid will largely reflect recent intake of the vitamin rather than underlying nutritional status; the criteria for assessment of 4-pyridoxic acid excretion are shown in Table 1.

Coenzyme Saturation of Transaminases

Various pyridoxal phosphate-dependent enzymes compete with each other for the available pool of coenzyme. Thus the extent to which an enzyme is saturated with its coenzyme provides a means of assessing the adequacy of the body pool of coenzyme. This can be determined by measuring the activity of the enzyme before and after the activation of any apoenzyme present in the sample by incubation with pyridoxal phosphate added *in vitro*. Erythrocyte aspartate and alanine transaminases are both commonly used; the results are expressed as either the percentage stimulation of activity by added pyridoxal phosphate, or the activation coefficient – the ratio of activity with added coenzyme to that without added coenzyme.

It seems to be normal for a proportion of pyridoxal phosphate-dependent enzymes to be present as inactive apoenzyme, without coenzyme. This may be a mechanism for metabolic regulation. It is possible that increasing the intake of vitamin B_6, so as to insure complete saturation of pyridoxal phosphate-dependent enzymes, may not be desirable.

Metabolic Loading Tests

A direct test of the adequacy of an individual's intake to meet his or her idiosyncratic metabolic requirement is the ability to metabolize a test dose of a substrate whose metabolism is dependent on the vitamin. For vitamin B_6, two metabolic loading tests can be used, although neither can be considered to be reliable for population studies of vitamin B_6 status.

The Tryptophan Load Test The oxidative pathway of tryptophan metabolism is shown in Figure 4. Kynureninase is a pyridoxal phosphate-dependent enzyme, and in vitamin B_6 deficiency its activity is lower than that of tryptophan dioxygenase. This means that there is a considerable accumulation of both hydroxykynurenine and kynurenine, sufficient to permit greater than usual metabolic flux through kynurenine transaminase, resulting in increased formation of kynurenic and xanthurenic acids. Although kynurenine transaminase is also pyridoxal phosphate-dependent, it is relatively unaffected in vitamin B_6 deficiency. Kynureninase is exquisitely sensitive to vitamin B_6 deficiency because it undergoes a slow inactivation as a result of catalyzing the half-reaction of transamination in addition to its normal reaction. The resultant enzyme with pyridoxamine phosphate at the catalytic site is catalytically inactive, and can only be reactivated if there is an adequate concentration of pyridoxal phosphate to displace the pyridoxamine phosphate.

Xanthurenic and kynurenic acids are easy to measure in urine, so that the ability to metabolize a test dose of 2 or 5 g of tryptophan has been widely adopted as a convenient and sensitive index of vitamin B_6 nutritional status. However, induction of tryptophan dioxygenase by glucocorticoid hormones will result in a greater rate of formation of kynurenine and hydroxykynurenine than the capacity of kynureninase, and will thus lead to increased formation of kynurenic and xanthurenic acids – an effect similar to that seen in vitamin B_6 deficiency. Such results may be erroneously interpreted as indicating vitamin B_6 deficiency in a variety of subjects whose problem is increased glucocorticoid secretion as a result of stress or illness, not vitamin B_6 deficiency.

Inhibition of kynureninase, e.g., by estrogen metabolites, also results in accumulation of kynurenine and hydroxykynurenine, and hence increased formation of kynurenic and xanthurenic acids, again giving results which falsely suggest vitamin B_6 deficiency.

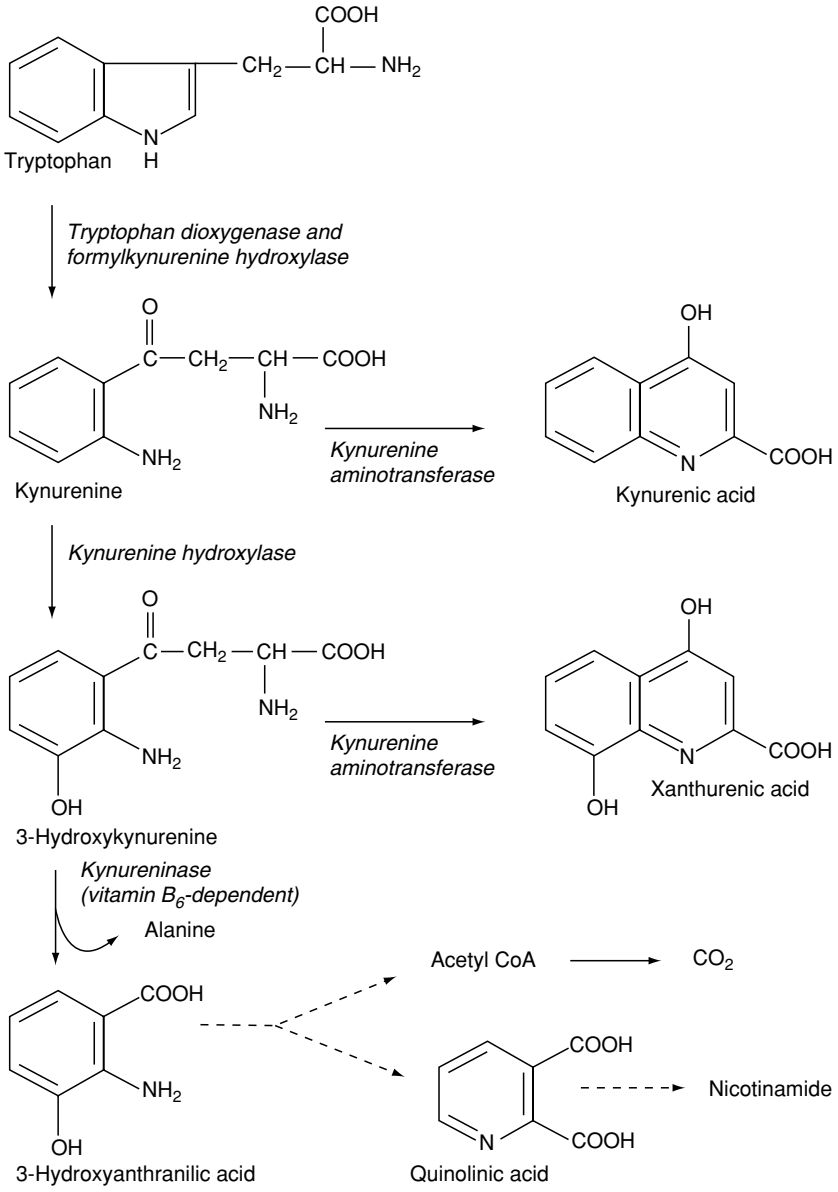

Figure 4 The oxidative pathway of tryptophan metabolism – the basis of the tryptophan load test for vitamin B_6 status.

This has been widely, but incorrectly, interpreted as estrogen-induced vitamin B_6 deficiency – it is in fact simple competitive inhibition by estrogen metabolites of the enzyme that is the basis of the tryptophan load test.

There is normally a considerable excess of either apokynureninase or kynureninase that has undergone transamination, and has pyridoxamine phosphate at the catalytic site, in the liver. This can be activated by (relatively high concentrations of) pyridoxal phosphate. The abnormalities of tryptophan metabolism associated with increased activity of tryptophan dioxygenase, or inhibition of kynureninase by estrogen metabolites, are thus corrected by the administration of high doses of vitamin B_6, although they are not in fact due to deficiency.

This means that, while the tryptophan load test may be an appropriate index of status in controlled depletion/repletion studies to determine vitamin B_6 requirements, it is not an appropriate index of status in population studies.

The methionine loading test The metabolism of methionine, shown in Figure 5, includes two pyridoxal phosphate-dependent steps, catalyzed by cystathionine synthetase and cystathionase. In vitamin B_6 deficiency there is an increase in the plasma concentration of homocysteine, and increased urinary

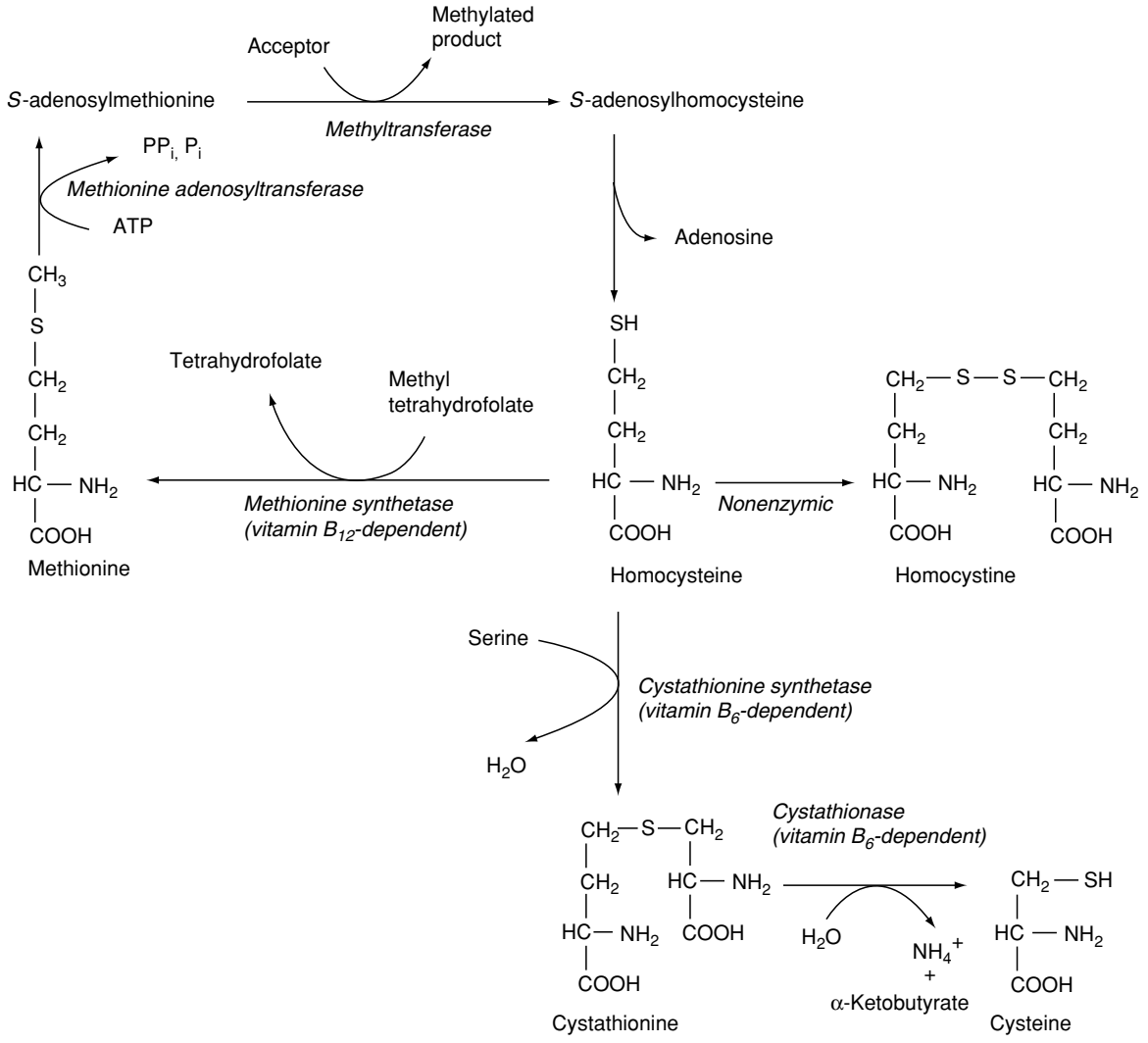

Figure 5 The pathway of methionine metabolism – the basis of the methionine load test for vitamin B_6 status.

excretion of cystathionine and homocysteine, both after a loading dose of methionine and under basal conditions. The ability to metabolize a test dose of methionine therefore provides an index of vitamin B_6 nutritional status. Because measurement of homocysteine and cystathionine is technically less easy than measurement of xanthurenic and kynurenic acids, the methionine load test has been less widely used than the tryptophan load test.

Some 10–25% of the population have a genetic predisposition to hyperhomocysteinemia, which is a risk factor for atherosclerosis and coronary heart disease, as a result of polymorphism in the gene for methylenetetrahydrofolate reductase. As discussed below, there is no evidence that supplements of vitamin B_6 reduce fasting plasma homocysteine in these subjects and, like the tryptophan load test, the methionine load test may be an appropriate index of status in controlled depletion/repletion studies to determine vitamin B_6 requirements, but not in population studies.

Requirements and Recommendations

The total body pool of vitamin B_6 is of the order of 15 μmol (3.7 mg) per kg of body weight. Isotope tracer studies suggest that there is a turnover of about 0.13% per day, and hence a minimum requirement for replacement of 0.02 μmol (5 μg) per kg of body weight – some 350 μg day^{-1} for a 70-kg adult.

Most studies of vitamin B_6 requirements have followed the development of abnormalities of tryptophan (and sometimes also methionine) metabolism during depletion, and normalization during repletion with graded intakes of the vitamin. While the tryptophan and methionine loading tests are unreliable as

indices of vitamin B_6 nutritional status in field studies, under the controlled conditions of depletion/repletion studies they do indeed provide a useful indication of the state of vitamin B_6 nutrition.

Although some 80% of the total body pool of vitamin B_6 is associated with muscle glycogen phosphorylase, this pools turns over relatively slowly. The major metabolic role of the remaining 20% of total body vitamin B_6, which turns over considerably more rapidly, is in amino acid metabolism. Therefore, *a priori*, it seems likely that protein intake will affect vitamin B_6 requirements.

People maintained on (experimental) vitamin B_6-deficient diets develop abnormalities of tryptophan and methionine metabolism faster, and their blood vitamin B_6 falls more rapidly, when their protein intake is high. Similarly, during repletion of deficient subjects, tryptophan and methionine metabolism and blood vitamin B_6 are normalized faster at low than at high levels of protein intake. However, the relevance of these studies to normal nutrition is unclear – they were conducted using extremes of protein intake: 40 g day^{-1} for the low protein intake, which is barely adequate to maintain nitrogen balance, and 150 g day^{-1} for the high intake, which is considerably higher than average western intakes of protein.

These studies suggest a mean requirement of 13 μg of vitamin B_6 per gram of dietary protein; recommended dietary allowances (RDA) are based on 15–16 μg per g of protein. At average intakes of about 100 g of protein per day, this gives an RDA of 1.4–1.6 mg of vitamin B_6.

More recent studies of women's requirements have suggested an RDA of 20 μg g^{-1} protein intake. It is not clear whether this reflects a gender difference (most of the earlier studies were conducted using men) or the use of more sensitive criteria of adequacy than were used previously.

Possible Benefits of Higher Levels of Intake

The identification of hyperhomocysteinemia as an independent risk factor in atherosclerosis and coronary heart disease has led to suggestions that higher intakes of vitamin B_6 than are currently considered adequate to meet requirements may be desirable. As shown in Figure 5, homocysteine is an intermediate in methionine metabolism, and may undergo one of two metabolic fates: remethylation to methionine (a reaction that is dependent on vitamin B_{12} and folic acid), or onward metabolism leading to the synthesis of cysteine (the vitamin B_6-dependent transsulfuration pathway).

Among elderly survivors of the Framingham study cohort (aged 67–96), hyperhomocysteinemia was most significantly correlated with low folate status, but there was also a significant association with low vitamin B_6 status. However, a number of studies have shown that, while folate supplements lower fasting homocysteine in moderately hyperhomocysteinemic subjects, 10 mg day^{-1} vitamin B_6 has no effect.

Vitamin B_6 supplements do, however, reduce the peak plasma concentration of homocysteine following a test dose of methionine. This can probably be explained by the kinetics of the enzymes involved. The K_m of cystathionine synthetase is 10-fold higher than that of methionine synthetase. Under basal conditions, little homocysteine is metabolized by way of the transsulfuration pathway. It is only after a loading dose of methionine, when homocysteine rises to relatively high levels, that the activity of cystathionine synthetase, rather than the availability of its substrate, is limiting or transsulfuration.

It is thus unlikely that intakes of vitamin B_6 above amounts that are adequate to prevent metabolic signs of deficiency will be beneficial in lowering plasma concentrations of homocysteine.

Vitamin B_6 Requirements of Infants

Estimation of the RDA for vitamin B_6 of infants presents a problem, and there is a clear need for further research to achieve a realistic estimate of infants' requirements. Human milk, which must be assumed to be adequate for infant nutrition, provides only some 40–100 μg l^{-1}, or 3–8 μg of vitamin B_6 per gram of protein – very much lower than the apparent requirement for adults. There is no reason why infants should have a lower requirement than adults, and indeed since they must increase their total body pool of the vitamin as they grow, they might be expected to have a proportionally higher requirement than adults.

A first approximation of the vitamin B_6 needs of infants came from studies of those who convulsed as a result of gross deficiency caused by overheated infant milk formula in the 1950s. At intakes of 60 μg day^{-1} there was an incidence of convulsions of 0.3%. Provision of 260 μg day^{-1} prevented or cured convulsions, but 300 μg day^{-1} was required to normalize tryptophan metabolism. This is almost certainly a considerable overestimate of requirements, since pyridoxyllysine, formed by heating the vitamin with proteins, has antivitamin activity, and would therefore result in a higher apparent requirement.

Based on the body content of 15 μmol (3.7 mg) of vitamin B_6 per kg of body weight, and the rate of weight gain, the minimum requirement for infants over the first 6 months of life would appear to be 100 μg (417 nmol) day^{-1} to establish tissue reserves.

Pharmacological Uses and Toxicity of Vitamin B_6 Supplements

Supplements of vitamin B_6 ranging from 25 to 500 mg day^{-1}, and sometimes higher, have been recommended for the treatment of a variety of conditions in which there is an underlying physiological or biochemical mechanism to justify the use of supplements, although in most cases there is little evidence of efficacy. Such conditions include postnatal depression, depression and other side-effects associated with oral contraceptives, hyperemesis of pregnancy, and the premenstrual syndrome.

Supplements have also been used empirically, with little or no rational basis, and little or no evidence of efficacy, in the treatment of a variety of conditions, including acute alcohol intoxication, atopic dermatitis, autism, dental caries, diabetic neuropathy, Down's syndrome, Huntington's chorea, schizophrenia, and steroid-dependent asthma.

Doses of 100 mg day^{-1} have been reported to be beneficial in the treatment of the carpal tunnel syndrome, or what has been called tenosynovitis. However, most of the reports originate from one center, and there appears to be little or no independent confirmation of the usefulness of the vitamin in this condition.

Vitamin B_6 has been reported to be effective in suppression of lactation, although other reports have shown no difference from placebo. Because the vitamin suppresses the increase in prolactin induced by treatment with the dopamine receptor antagonist pimozide, and because lactation is also suppressed by the dopamine agonist bromocriptine, it has been suggested that it acts to stimulate dopaminergic activity in the hypothalamus. However, it is more likely that its action is by reduction in target tissue responsiveness to the steroid hormones that stimulate prolactin secretion.

Doses of 50–200 mg day^{-1} have an antiemetic effect, and the vitamin is widely used, alone or in conjunction with other antiemetics, to minimize the nausea associated with radiotherapy and to treat pregnancy sickness. There is no evidence that vitamin B_6 has any beneficial effect in pregnancy sickness, nor that women who suffer from morning sickness have lower vitamin B_6 nutritional status than other pregnant women. There have been reports of a teratogenic effect of vitamin B_6 used to treat morning sickness, but all of these involved use of the vitamin together with the sedative Debendox, and there is no evidence that vitamin B_6 itself is teratogenic. However, since it downregulates responsiveness to steroid hormones and retinoids, it is possible that high levels of vitamin B_6 intake may affect embryonic or fetal development.

Vitamin B_6 and the Side-Effects of Oral Contraceptives

Although estrogens do not cause vitamin B_6 deficiency, the administration of vitamin B_6 supplements has beneficial effects on some of the side-effects of both administered and endogenous estrogens. The supplements act in two main areas: in normalizing glucose tolerance and as an antidepressant.

Impairment of glucose tolerance is common in pregnancy, and may indeed be severe enough to be classified as gestational diabetes mellitus, which generally resolves at parturition, although in some subjects it may persist, pregnancy having been the trigger for the development of maturity-onset diabetes. High-estrogen oral contraceptives may also cause impaired glucose tolerance. This seems to be the result of increased tissue and blood concentrations of xanthurenic acid, because of the inhibition of kynureninase by estrogen metabolites. Xanthurenic acid forms a complex with insulin which has little or no hormonal activity. Vitamin B_6 supplements may have a beneficial effect on glucose tolerance by activating apokynureninase or kynureninase that has been inactivated by undergoing transamination.

One of the relatively common side-effects of estrogenic oral contraceptives is depression, affecting about 6% of women in some studies. This frequently responds well to the administration of relatively large amounts of vitamin B_6 (generally in excess of 40 mg day^{-1}). Postnatal depression also responds to similar supplements in some studies.

Again, this does not seem to be due to correction of vitamin B_6 deficiency, but rather to a direct effect of pyridoxal phosphate on the metabolism of tryptophan. High concentrations of pyridoxal phosphate attenuate the response to glucocorticoid hormones; tryptophan dioxygenase is a glucocorticoid-induced enzyme, and thus its synthesis and activity will be reduced by high intakes of vitamin B_6. This reduces the oxidative metabolism of tryptophan, and increases the amount available for synthesis of 5-hydroxytryptamine in the brain. Increased brain 5-hydroxytryptamine synthesis has a mood-elevating effect.

Vitamin B_6 in the Premenstrual Syndrome

The studies showing a beneficial action of vitamin B_6 in overcoming depression associated with oral contraceptives have led to the use of the vitamin in depression and other pathology associated with endogenous estrogens, in the premenstrual syndrome. There is no evidence of poorer vitamin B_6 nutritional status in women who suffer from the premenstrual syndrome.

There are few well-controlled studies of the effects of vitamin B_6 in premenstrual syndrome. In general, those that have been properly controlled report little benefit from doses between 50 and 200 mg day^{-1} compared with placebo, although some studies do claim a beneficial effect. Interestingly, meta-analysis of controlled cross-over trials shows that whichever treatment is used second, active vitamin or placebo is (marginally) more effective. There is no obvious explanation for this observation.

Despite the lack of evidence of efficacy, vitamin B_6 is widely prescribed (and self-prescribed) for the treatment of the premenstrual syndrome.

Toxicity of Vitamin B_6

Animal studies have demonstrated the development of signs of peripheral neuropathy, with ataxia, muscle weakness and loss of balance, in dogs given 200 mg pyridoxine HCl per kg of body weight for 40–75 days, and the development of a swaying gait and ataxia within 9 days at a dose of 300 mg per kg of body weight. At a dose of 50 mg per kg of body weight, there are no clinical signs of toxicity, but histologically there is a loss of myelin in dorsal nerve roots. At higher doses there is more widespread neuronal damage, with loss of myelin and degeneration of sensory fibers in peripheral nerves, the dorsal columns of the spinal cord, and the descending spinal tract of the trigeminal nerve. The clinical signs of vitamin B_6 toxicity in animals regress after withdrawal of these massive doses, but sensory nerve conduction velocity, which decreases during the development of the neuropathy, does not recover fully. The mechanism of the neurotoxic action of vitamin B_6 is unknown.

The development of sensory neuropathy has been reported in patients taking 2–7 g of pyridoxine HCl day^{-1}. Although there was residual damage in some patients, withdrawal of these extremely high doses resulted in a considerable recovery of sensory nerve function.

Other reports have suggested that intakes as low as 50 mg day^{-1} are associated with neurological damage, although these have been based on patients reporting symptoms rather than on detailed neurological examination. Nevertheless, this led to a proposal in the UK in 1997 to regulate the sale of vitamin B_6 supplements, with 50 mg day^{-1} being available only on prescription, and between 10 and 50 mg day^{-1} only available from qualified pharmacists. The proposals were put in abeyance in 1998, pending further studies and reexamination of the evidence of toxicity at intakes below 100–200 mg day^{-1}.

Vitamin B_6 Deficiency

Gross clinical deficiency of vitamin B_6 is more or less unknown. The vitamin is widely distributed in foods, and intestinal flora synthesize relatively large amounts, at least some of which is believed to be absorbed and hence available.

In vitamin B_6-deficient experimental animals there are more or less specific skin lesions (e.g., acrodynia in the rat) and fissures or ulceration at the corners of the mouth and over the tongue, as well as a number of endocrine abnormalities, defects in the metabolism of tryptophan, methionine and other amino acids, hypochromic microcytic anemia (the first step of heme biosynthesis is a pyridoxal phosphate-dependent reaction), changes in leukocyte count and activity, a tendency to epileptiform convulsions, and peripheral nervous system damage resulting in ataxia and sensory neuropathy.

Much of our knowledge of human vitamin B_6 deficiency is derived from an outbreak in the early 1950s, which resulted from an infant milk preparation which had undergone severe heating in manufacture. The probable result of this was the formation of pyridoxyllysine by reaction between pyridoxal phosphate and the ε-amino groups of lysine in proteins. In addition to a number of metabolic abnormalities, many of the affected infants convulsed. They responded to the administration of vitamin B_6 supplements.

Investigation of the neurochemical basis of the convulsions in vitamin B_6 deficiency helped to elucidate the role of GABA as a neurotransmitter; GABA is synthesized by the decarboxylation of glutamate. More recent studies have suggested that the accumulation of hydroxykynurenine in the brain may be the critical factor precipitating convulsions in deficiency; GABA is depleted in the brains of deficient adult and neonate animals, while hydroxykynurenine accumulation is considerably more marked in neonates than adults – only neonates convulse in vitamin B_6 deficiency. GABA depletion may be a necessary but not sufficient condition for convulsions in vitamin B_6 deficiency.

Vitamin B_6 Dependency Syndromes

A small number of cases have been reported of patients with genetic defects which result in an abnormally high requirement for vitamin B_6 in order to maintain the activity of the affected enzyme (Table 2). Such vitamin B_6 dependency syndromes have been reported in cases of xanthurenic aciduria, homocystinuria, hypochromic sideroblastic anemia, ornithinemia, and infantile convulsions. The molecular basis of the defects appears to be a severely impaired affinity of the affected enzyme for its

Table 2 Vitamin B_6-responsive inborn errors of metabolism

	Enzyme affected
Convulsions of the newborn	Enzyme defect not known
Cystathioninuria	Cystathionase (**Figure 5**)
Gyrate atrophy with ornithinuria	Ornithine-δ-aminotransferase
Homocystinuria	Cystathionine synthase (**Figure 5**)
Primary hyperoxaluria, type 1	Peroxisomal alanine-glyoxylate transaminase
Sideroblastic anemia	δ-Aminolevulinate synthase (↓ heme synthesis)
Xanthurenic aciduria	Kynureninase (**Figure 4**)

cofactor, and patients respond well to doses of 500–1000 mg of vitamin B_6 per day. Apart from the affected enzyme, other biochemical indices of vitamin B_6 nutritional status are normal in these patients. Interestingly, there are few reports of peripheral neuropathy among such patients treated with high doses of vitamin B_6 for many years.

Groups at Risk of Deficiency

A number of studies have shown that between 10 and 20% of the apparently healthy population have low plasma concentrations of pyridoxal phosphate or abnormal erythrocyte transaminase activation coefficient, suggesting vitamin B_6 inadequacy or deficiency. In most studies, only one of these indices of vitamin B_6 nutritional status has been assessed. Where both have been assessed, while each shows some 10% of the population apparently inadequately provided with vitamin B_6, few of the subjects show inadequacy by both criteria.

There is a decrease in the plasma concentration of vitamin B_6 with increasing age, and some studies have shown a high prevalence of abnormal transaminase activation coefficient in elderly subjects, suggesting that the elderly may be at risk of vitamin B_6 deficiency. It is not known whether this reflects an inadequate intake, a greater requirement, or changes in the tissue distribution and metabolism of the vitamin with increasing age.

Drug-induced vitamin B_6 deficiency A number of drugs which react with carbonyl compounds are capable of causing vitamin B_6 deficiency on prolonged use. These include the antituberculosis drug isoniazid (iso-nicotinic acid hydrazide), penicillamine, and the antiparkinsonian drugs Benserazide and Carbidopa. In general, the main effect is impairment of tryptophan metabolism by inhibition of kynureninase, and hence the development of the niacin deficiency disease pellagra. The condition therefore responds to the administration of either vitamin B_6 or niacin. Isoniazid also causes peripheral neuropathy, which responds to vitamin B_6 supplements, but not to niacin.

Estrogens and vitamin B_6 nutritional status There have been many reports of abnormal tryptophan metabolism in women taking estrogens as oral contraceptives and menopausal hormone replacement therapy. These have been widely interpreted as evidence of estrogen-induced vitamin B_6 deficiency. However, as discussed above, this is the result of inhibition of kynureninase by estrogen metabolites, not estrogen-induced deficiency of the vitamin. Where other indices of vitamin B_6 status have been reported, they have been generally unaffected by contraceptive use, again suggesting an effect on tryptophan metabolism *per se*, rather than on vitamin B_6 nutritional status.

In many cases the metabolism of tryptophan has been normalized by the administration of vitamin B_6 supplements of the order of 20–50 mg day^{-1}, compared with an RDA of 1.4–1.6 mg day^{-1}. It was noted above that there is an apparent excess of apokynureninase in the liver and therefore the administration of vitamin B_6 supplements will increase kynurenine metabolism, even when there is no preexisting deficiency.

See also: **Amino Acids**: Metabolism; **Contraceptives: Nutritional Aspects**; **Glycogen**; **Hormones**: Steroid Hormones; **Infants**: Nutritional Requirements; **Premenstrual Syndrome: Nutritional Aspects**; **Vitamin B_6**: Properties and Determination

Further Reading

Bender DA (1987) Oestrogens and vitamin B_6 – actions and interactions. *World Review of Nutrition and Dietetics* 51: 140–188.

Bender DA (1992) *Nutritional Biochemistry of the Vitamins*. Cambridge: Cambridge University Press.

Bender DA (1999) Non-nutritional uses of vitamin B_6. *British Journal of Nutrition* 81: 7–20.

Bender DA (2003) *Nutritional Biochemistry of the Vitamins*, 2nd edn. New York, Cambridge University Press.

Coburn SP (1996) Modelling vitamin B_6 metabolism. *Advances in Food and Nutrition Research* 40: 107–132.

Fasella PM (1967) Pyridoxal phosphate. *Annual Review of Biochemistry* 36: 185–210.

Hayashi H, Wada H, Yoshimura T, Esaki N and Soda K (1990) Recent topics in pyridoxal 5′-phosphate enzyme studies. *Annual Review of Biochemistry* 59: 87–110.

Ink SL and Henderson LM (1984) Vitamin B_6 metabolism. *Annual Review of Nutrition* 4: 455–470.

Lumeng L, Lui A and Li T-K (1980) Plasma content of B_6 vitamers and its relationship to hepatic vitamin B_6 metabolism. *Journal of Clinical Investigation* 66: 688–695.

Martell AE (1982) Reaction pathways and mechanisms of pyridoxal catalysis. *Advances in Enzymology* 53: 163–169.

Merrill AH and Henderson LM (1987) Diseases associated with defects in vitamin B_6 metabolism of utilization. *Annual Review of Nutrition* 7: 137–156.

Wiss O and Weber F (1964) Biochemical pathology of vitamin B_6 deficiency. *Vitamins and Hormones* 22: 495–501.

Vitamin C *See* **Ascorbic Acid**: Properties and Determination; Physiology

Vitamin D *See* **Cholecalciferol**: Properties and Determination; Physiology

Vitamin E *See* **Tocopherols**: Properties and Determination; Physiology

VITAMIN K

Contents

Properties and Determination

H E Indyk, Anchor Products, NZMP, Waitoa, New Zealand
M J Shearer, St Thomas' Hospital, London, UK
D C Woollard, Agriquality NZ, Auckland 1, New Zealand

Background

Vitamin K is a fat-soluble vitamin widely distributed in nature and comprising several molecular forms. It was discovered as an essential antihemorrhagic factor in the 1930s by the Danish scientist Henrik Dam, who named it *Koagulationsvitamin*. This article reviews the properties and analysis of vitamin K, with special reference to its measurement in foods. Such measurements have proved difficult, and reliable methods have only become available since the 1980s with the development of physicochemical assays based on high-performance liquid chromatography (HPLC). A subsequent article deals with the physiology of vitamin K, where the introduction of new assay techniques for vitamin K and its dependent proteins has led to a greater understanding of its nutritional role. According to recent evidence, this includes the probability that, in addition to its well-defined role in blood coagulation, vitamin K is needed for a variety of physiological processes related to a number of extrahepatic vitamin K-dependent proteins. (*See* **Vitamins**: Overview.)

Chemical Structures and Nomenclature

Naturally occurring compounds with vitamin K activity possess a common 2-methyl-1,4-naphthoquinone

nucleus, but differ in the structures of the side-chain at the 3 position (Figure 1).

In traditional nomenclature (Table 1), the vitamin K found in plants (of which there is only one major chemical form) is known as vitamin K_1 and the multiple forms synthesized by bacteria as vitamin K_2, where the number of carbon atoms (n) in the side chain is indicated as vitamin $K_{2(n)}$. This nomenclature was revised by the International Union of Pure and Applied Chemistry – International Union of Biochemistry (IUPAC–IUB) Subcommittee on Nomenclature of Quinones, who recommended that vitamins K_1 and K_2 be called phylloquinone and menaquinones, respectively. Recommended current abbreviations are K_1 for phylloquinone and MK-n for menaquinones. Table 1 also shows a nomenclature that the International Union of Nutritional Sciences (IUNS) recommended should replace the IUPAC system. This new system, however, has not found favor among most workers in this field. Consequently, the IUPAC-IUB nomenclature is still the one most generally accepted at the present time and will be used throughout this article.

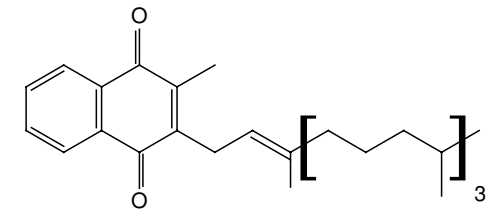

(a) Phylloquinone

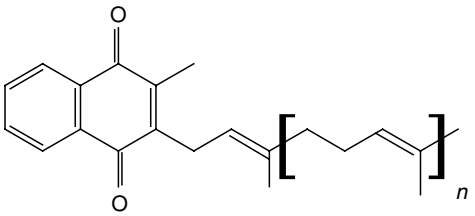

(b) Menaquinone

Figure 1 Chemical structures of naturally occurring K vitamins: (a) phylloquinone (vitamin K_1) and (b) menaquinones (vitamin K_2).

The parent compound of the vitamin K group of vitamers is 2-methyl-1,4-naphthoquinone. This structure is not found in nature but, because of its historical use as a synthetic derivative, was originally designated vitamin K_3. Under the IUPAC system, this artificial vitamin K is called menadione.

Phylloquinone, which has the same phytyl side-chain as chlorophyl, is found throughout the plant kingdom and in cyanobacteria (blue–green algae). In higher plants, the synthesis of phylloquinone is associated with the chloroplast, while the unusual finding of phylloquinone in cyanobacteria has been cited as supportive evidence for the possible bacterial origin of the chloroplast in evolution.

The menaquinones represent a highly diverse family of compounds all synthesized by bacteria. The side-chain of menaquinones is based on a number of repeating unsaturated five-carbon (prenyl) units. The major forms are designated menaquinone-n (MK-n), with n denoting the number of prenyl units. Some bacteria also synthesize menaquinones in which one or more of the multiprenyl side-chain double bonds are saturated. In such cases, the additional hydrogen atoms are indicated by the prefixes dihydro, tetrahydro, hexahydro, etc., and these may be abbreviated to MK-n(H_2), MK-n(H_4), MK-n(H_6), etc. The division between the plant-form phylloquinone and the bacterial menaquinones is somewhat artificial, since phylloquinone may be regarded as a partially saturated form of menaquinone-4, namely MK-4(H_6), but the distinction serves to emphasize their different origins.

Carbon atoms in the side-chain are designated C1′, etc., and the presence of double bonds at C2′ in phylloquinone and at C2′, 6′, 10′, etc. in menaquinones indicates that *cis–trans* isomerism is possible. Natural forms of phylloquinone and menaquinones exist as *trans* isomers, but synthetic forms contain both *trans* and *cis* isomers.

Biological Activities

The biological activity of different vitamin K compounds has been assessed in whole animal systems (*in vivo*), or as a cofactor to the vitamin K-dependent γ-glutamyl carboxylase in isolated enzyme systems

Table 1 Nomenclature of K vitamins

Chemical name	*Old name*	*IUPAC–IUB (abbreviation)*	*IUNS (abbreviation)*
2-Methyl-1,4-naphthoquinone	Vitamin K_3	Menadione	Menaquinone
2-Methyl-3-phytyl-1,4-naphthoquinone	Vitamin K_1	Phylloquinone (K)	Phytylmenaquinone (PMQ)
2-Methyl-3-multiprenyl-1,4-naphthoquinone (class)	Vitamin K_2	Menaquinone-n (MK-n)	Prenylmenaquinone-n (MQ-n)
2-Methyl-3-farnesylfarnesyl-1,4-naphthoquinone	Vitamin $K_{2(30)}$	Menaquinone-6 (MK-6)	Prenylmenaquinone-6 (MQ-7)

(*in vitro*). γ-Glutamyl carboxylase is an integral membrane glycoprotein that requires vitamin K to activate posttranslational modification of glutamate (Glu) to γ-carboxyl glutamate (Gla) in specific vitamin K-dependent proteins. The *in vivo* systems are normally based on the correction of blood clotting time in chicks induced by vitamin K deficiency, or through measurement of undercarboxylated vitamin K-dependent proteins, and their interpretation needs to take account of the route of administration and other metabolic considerations, such as absorption and interconversion to other chemical forms. (*See* **Coenzymes**.)

It has been established that besides the 1,4-naphthoquinone nucleus, a methyl group at the C2 position is essential for both *in vivo* and *in vitro* activity. The activity of menadione *in vivo* is due to the presence of tissue alkylating enzymes that can transform this compound to MK-4. Recent evidence is accumulating that suggests an alternative *in vivo* route to MK-4 via tissue-mediated conversion of phylloquinone. The essential requirement for *in vitro* activity is for a 2-methyl-3-phytyl (or polyisoprenyl)-1,4-naphthoquinone structure, which can be reduced to the corresponding hydroquinone. The biological activity of K_1 is reduced by saturation of the phytyl double bond (2′,3′-dihydrophylloquinone), whereas the desmethyl or *cis* isomers have no appreciable activity *in vitro*, although some conversion of the *cis* form to the active *trans* form may occur *in vivo*.

Physicochemical Properties

Appearance, Solubility, and Stability

Phylloquinone and menaquinones are golden yellow oils at ambient temperature, though the higher menaquinones may also be obtained as fine, yellow crystals.

All K vitamins are insoluble in water but are soluble in normal lipid solvents such as hexane, chloroform, diethyl ether, and acetone, but are less soluble in ethanol. They are only sparingly soluble in methanol, and the degree of solubility depends on their hydrophobicity (i.e., menaquinones with long side-chains are less soluble than those with short side-chains).

Vitamin K compounds are reasonably stable towards heat, oxygen, and mild acidic conditions, but are degraded by UV light, alkalis, and strong acids.

Spectroscopy

The spectroscopic properties of K vitamins are important for their detection, identification, and quantification in assays. All vitamin K compounds possess the same characteristic ultraviolet spectrum, which is a summation of the benzenoid and 1,4-quinone components. Naturally occurring phylloquinone and menaquinones exhibit typical benzenoid bands with maxima at 242, 248, and 238 nm (shoulder), whereas the quinone contributions appear as distinct bands with absorption maxima at 260 and 269 nm and a higher-wavelength maximum at 326 nm. Ultraviolet spectroscopy is particularly useful for the quantification of standard solutions; the molar absorption coefficient (ε) is the same for both phylloquinone and menaquinones and, at 248 nm (the wavelength of highest absorption), has a value of 18 900. (*See* **Spectroscopy**: Overview.)

Vitamin K compounds, in their stable quinone forms, do not possess native fluorescence, but they can be readily reduced to the corresponding quinols, whose fluorescent properties have been exploited for the assay of K vitamins in foods and other biological tissues. (*See* **Spectroscopy**: Fluorescence.)

The K vitamers also exhibit infrared and nuclear magnetic resonance absorption spectra distinctive of the naphthoquinone nucleus, and may be further characterized by mass spectroscopy. (*See* **Mass Spectrometry**: Principles and Instrumentation; Applications.)

Occurrence and Forms in Foods

Vitamin K is largely found in the membranous fractions of cells. Thus, in plants, phylloquinone is concentrated in the chloroplast lamellae, and in bacteria, menaquinones are located in the plasma membrane. In animals, the vitamin is found in various cellular membranes, especially those of the microsomes and mitochondria.

Recent progress in analytical techniques for this vitamin, as well as increased emphasis on its nutritional significance, is responsible for the current development of a comprehensive database of the vitamin K content of foods. Whereas, before 1980, most published values for the vitamin K content of foods originated from bioassays, contemporary data derive from the highly specific and accurate HPLC methods. These techniques have focused predominantly on phylloquinone ubiquitously distributed in foods, although more recently, increasing amounts of data are becoming available for the menaquinones.

Table 2 lists some values of the phylloquinone content of various foods arranged into four concentration ranges to illustrate the wide distribution of the vitamin in the diet.

In general, the relative values confirm the known association of phylloquinone with photosynthetic tissues, with the highest values (400–600 μg per 100 g) being found in green, leafy vegetables. Nutritionally significant amounts of phylloquinone are, however, present in many nonleafy vegetables,

Table 2 Vitamin K_1 content of common foods determined by HPLC (μg per 100 g)[a]

0.1–1.0	*1–10*	*10–100*	*100–1000*
Bananas	Apples	Avocado	Broccoli tops
Beef, steak	Aubergines	Beans, runner	Brussels sprouts
Bread, white	Baked beans	Beans, French	Cabbage, green
Chicken, thigh	Barley	Beans, broad	Kale
Coconut oil	Beef, minced (1–30)	Cabbage, red	Lettuce
Cod, fillet (0.1–2)	Bilberries	Cauliflower	Margarine
Cornflakes	Bran, wheat	Chick peas	Parsley
Flour, white	Bread, wholemeal	Cucumber	Rapeseed oil
Grapefruit	Butter	Greengages	Soybean oil
Ham, tinned	Carrots	Kiwi fruit	Spinach
Icecream	Celery	Margarine	Water cress
Maize	Cheeses (1–80)	Mustard cress	
Mangoes	Chocolate, plain	Natto (900–1200)	
Melon, yellow	Corn oil	Olive oil, ex. Virgin	
Melon, water	Courgettes	Peas	
Milk, cows	Cranberries		
Mushrooms	Cream, double		
Onion	Dates, fresh		
Oranges	Eggs (10–25)		
Parsnips	Figs, fresh		
Peanuts, roast	Grapes, black		
Pilchards, brine (0.1–2)	Grapes, green		
Pineapple	Infant Formula		
Pork chop, lean	Leeks		
Potatoes	Liver, lamb		
Rice white	Liver, ox (10–20)		
Rice brown	Nectarines		
Salmon, brine	Oats		
Sausages, pork/beef	Palm oil		
Spaghetti	Peaches, fresh		
Trout, Rainbow (0.1–3)	Pears		
Tuna, brine	Peppers, green		
Turnips	Peppers, red		
Yogurt (0.1–0.4)	Plums, red		
	Raisins		
	Rhubarb		

[a]MK-*n* ranges are shown in parentheses.

as well as in fruits, oils, dairy products, eggs, and some meat products. With modern HPLC techniques, naturally occurring phylloquinone can be detected in most foods. Vegetable oils show a high degree of variability, ranging from < 5 μg per 100 g in coconut, corn, groundnut, and safflower oils, to 100–200 μg per 100 g in soya bean and rapeseed oils. This leads to a similar variation in the phylloquinone content of margarines according to the oil base. The phylloquinone content of lean meat and fish is generally low (< 1 μg per 100 g), but higher amounts are found in animal livers; this organ acts as a storage site of the vitamin. Please refer to individual foods.

There is less information on the menaquinone content of foods. It is known that animal liver contains substantial concentrations of menaquinones, which normally greatly exceed the phylloquinone content. The livers of ruminants contain particularly large concentrations of long-chain menaquinones (MK-6–13), which derive from bacterial fermentation in the rumen. The only other known sources of significant amounts of menaquinones in the Western diet are fermented foods, such as cheese and yogurt. All cheeses contain relatively high concentrations (≤ 20 μg per 100 g) of MK-8 and MK-9, with the latter predominating. Natto, the Japanese fermented soya bean food, contains substantial amounts of MK-7. Whereas menaquinones are detectable in cow's milk at much lower concentrations than phylloquinone, MK-4 has been confirmed to be present at comparable, or higher, levels than phylloquinone in the milk of several mammalian species.

Foods containing hydrogenated vitamin K-rich oils will also include a contribution from 2′,3′-dihydrophylloquinone. The biological activity of this K-vitamer is currently under investigation, and hence its abundance in the diet is of uncertain nutritional significance.

The processing of foods can have a variable effect on retention of vitamin K, with conventional cooking, canning, γ-irradiation, and freezing practices having a relatively minor influence, whereas significant and rapid loss occurs on exposure to UV light.

Use in Food Fortification

Since the requirements for vitamin K are low, fortification of foods with vitamin K is not common, with the exception of infant formula feeds to which synthetic preparations of phylloquinone are invariably added. Depending on the method of chemical synthesis, the ratio of the natural *trans* to the inactive *cis* form will vary, but is typically *c.* 7:1. (*See* **Food Fortification**; **Infant Foods**: Milk Formulas.)

Although national guidelines for fortification of infant formulas differ, their common aim is to protect against rare, but life-threatening, bleeding due to deficiency of vitamin K in early life. It is now known that the typical fortification levels of *c.* $50\,\mu g\,l^{-1}$ (as fed) are several times higher than human breast milk concentrations of phylloquinone, which average *c.* $2\,\mu g\,l^{-1}$. Although the reasons for adopting such high levels of fortification are somewhat empirical and based on the erroneously high estimates of vitamin K in breast milk by early bioassays, there has been little impetus for change because the low concentrations in breast milk are an established risk factor for neonatal vitamin K deficiency.

The synthetic vitamin K derivative menadione is no longer used to fortify infant formulae because of its association with neonatal hemolytic anemia and liver damage. In fact, the vitamin K activity of menadione is dependent on its *in vivo* conversion to biologically active MK-4. Nevertheless, because menadione is much cheaper to manufacture than phylloquinone, menadione derivatives are widely used in animal husbandry, particularly in poultry feeds, since poultry are very prone to bleeding due to dietary deficiency of vitamin K. In the poultry industry, the birds' dietary requirements may be further increased by the use of antibiotics, the effect of which may be either to inhibit the growth of intestinal microorganisms that synthesize menaquinones or to interfere directly with vitamin K metabolism. Consequently, stabilized water-soluble derivatives of menadione, such as crystalline menadione sodium bisulfite complex, are used universally as poultry feed supplements.

Measurement of Vitamin K in Food

In the 1930s, shortly after the discovery of vitamin K, Dam and colleagues developed a curative bioassay that involved the feeding of substances to vitamin K-deficient chicks and measuring the response on blood clotting. Although values that have been derived from this early work are often given in textbooks and handbooks of nutrition, a closer evaluation of the methodology shows that attempts to use these results as the basis for absolute food values are unjustified. Although such bioassays were greatly improved in the 1960s, little further work has emerged on their routine use and validation for food analysis.

The introduction of modern HPLC techniques has revolutionized the measurement of vitamin K in tissues. This powerful technique has met the criteria of high selectivity and sensitivity, which are also necessary for measurement of the low concentrations in most foods. Although several diverse protocols have been developed for phylloquinone, only one procedure has been validated by collaborative study. A range of sample extract purification strategies may be implemented prior to the final chromatographic analysis, which is usually accomplished by analytical reversed-phase HPLC. Due to the instability of vitamin K at high pH, saponification techniques commonly utilized for other fat-soluble vitamins are inappropriate for vitamin K analysis. The need for preliminary enhancement techniques is dependent principally on food type, analyte level, and final detection mode and may incorporate one or more techniques including (1) lipid extraction, (2) enzymatic lipid digestion, (3) solid-phase extraction (SPE) and (4) semipreparative normal-phase HPLC. (*See* **Chromatography**: High-performance Liquid Chromatography.)

Extraction and Clean-up

No universal extraction and purification procedure exists for vitamin K analysis in all foods. Nonpolar lipid solvents including hexane, isopropanol, dichloromethane, acetone, diethyl ether, and methanol, used either singly or in combination, have been used for initial total lipid extraction depending on the food type. For certain high-fat foods (fats, oils and milk), bulk lipid may be effectively removed with enzymatic digestion prior to hexane extraction of triglyceride-depleted lipid. For low-fat foods (vegetables, fruit, cereals, meats, and fish), acetone is an effective solvent, whereby, after the addition of water and hexane, the K vitamins partition entirely in the upper hexane phase, leaving polar lipids in the acetone–water phase. Alternatively, hexane:isopropanol or chloroform:methanol binary solvents are commonly used.

For protocols relying on a preliminary total lipid extraction, further purification is required in view of the overwhelming proportion of triglycerides and their incompatibility with reversed-phase HPLC.

Various combinations of lipase digestion, SPE and semipreparative normal-phase HPLC techniques are commonly recommended.

The extraction stage may also incorporate the addition of an internal standard, which, in general terms, compensates for procedural losses of the target analyte. This technique is essential when multidimensional protocols for vitamin K are implemented. The internal standard is chosen on the basis of several criteria that include its similar physical, chemical, and chromatographic properties to the analyte. Depending on the sample matrix, several vitamin K compounds and analogues (including radiolabelled compounds) have been found to be suitable internal standards for vitamin K assays.

A general scheme outlining commonly applied extraction and clean-up options is illustrated in Figure 2. Purification techniques usually comprise two successive stages based on the principles of adsorption chromatography, which facilitate the separation of K vitamers from most other potentially interfering lipid classes (e.g., hydrocarbons, sterols, fatty acids). If required, semipreparative normal-phase HPLC can also be used to isolate both K_1 and MK fractions and to separate the *cis* and *trans* isomers of phylloquinone.

The fraction(s) collected at the semipreparative HPLC stage may then be analyzed by reversed-phase HPLC using a variety of detection methods. For certain foods, a lipase digestion and hexane extraction scheme has been found to be sufficient to facilitate a direct analytical HPLC quantitation, without the need for additional purification steps.

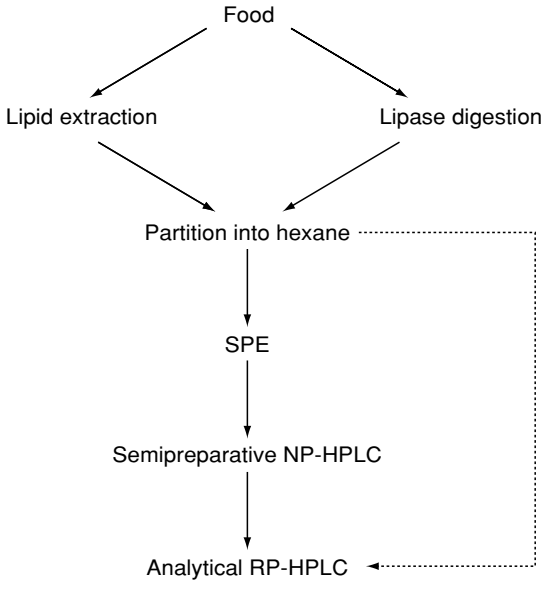

Figure 2 Flow chart summarizing extraction and purification of vitamin K from foods.

Detection and Quantification

The final analytical stage of HPLC is almost exclusively based on the principles of partition chromatography in the reversed-phase mode. This robust mode of liquid chromatography resolves the individual K vitamins from each other, while supporting the mobile phase additives necessary for the most sensitive detection techniques currently utilized. Whereas isocratic systems are generally used for the analysis of vitamin K_1, gradient elution is preferable when profiling the menaquinones, in view of their very wide hydrophobic range. The C18 functionality has most commonly been employed, although the recently developed C30 phase has been shown to allow resolution of the *cis* and *trans* isomers of phylloquinone, previously only accomplished under normal phase conditions.

The properties of K vitamins offer several avenues of approach to their detection by HPLC, and the most common options are summarized in Figure 3. The parent quinone form, in which the vitamin is isolated, may be detected with reasonable sensitivity (*c.* 500 pg) by ultraviolet absorbance detection, but this lacks selectivity and may be subject to interferences from other lipids. This detection method does, however, give reliable results for green vegetables, but is less satisfactory for animal tissues.

Alternatively, the more recently developed strategies exploit the reversible reduction of the quinone to the quinol, which may then be detected by either electrochemical or fluorescent techniques. For electrochemical detection, the greatest sensitivity and selectivity is obtained by employing a redox technique in which the detector has a dual-electrode cell with the electrodes arranged in series. Vitamin K is first

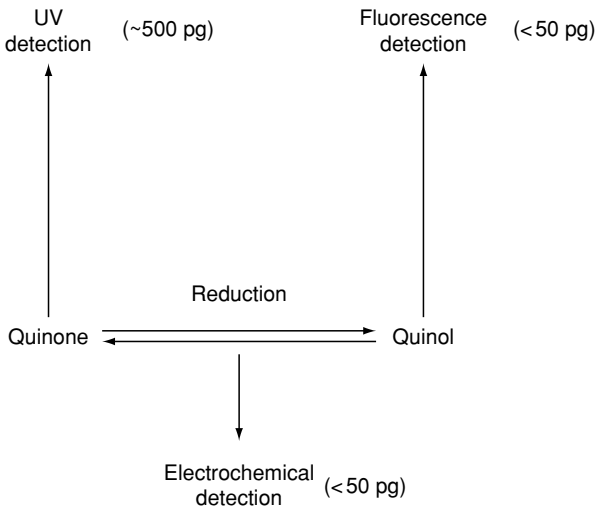

Figure 3 Detection options for vitamin K.

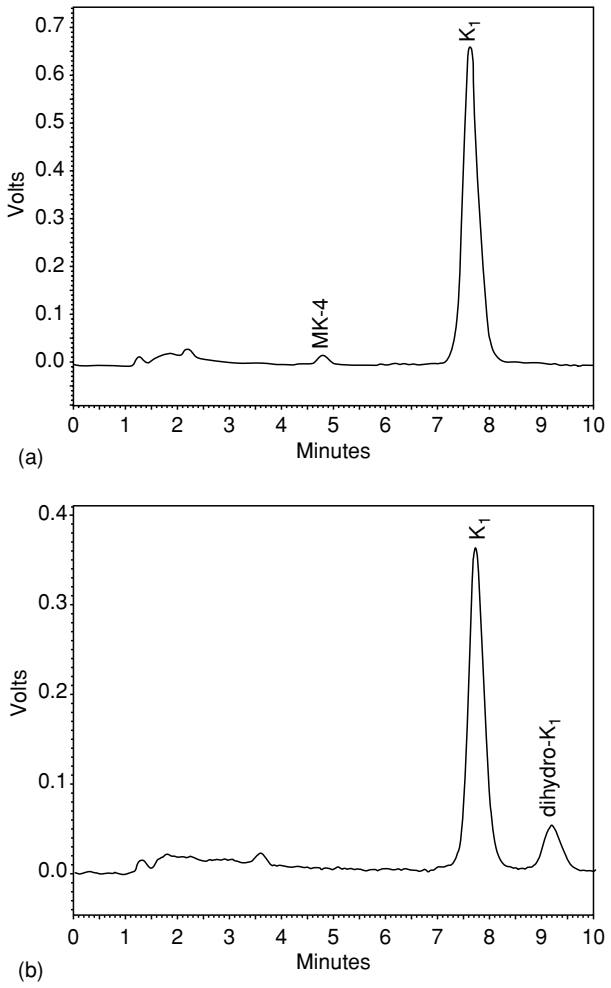

Figure 4 Liquid chromatography obtained for (a) infant formula and (b) margarine. C18 column, isocratic elution, postcolumn reduction, and fluorescence detection.

reduced at the upstream (or generator) electrode to the quinol, which is then reoxidized at the downstream (or detector) electrode: the current generated in the second reoxidation step is measured by the detector and is proportional to the amount of vitamin K passing through the electrochemical cell. This method is capable of detecting 50 pg or less of phylloquinone. Fluorescence detection again requires the postcolumn generation of the quinols after the chromatographic separation of their respective quinones. Reduction can be achieved by chemical, photochemical, or electrochemical techniques, with fluorescence detection providing similar sensitivity to redox electrochemical detection (*c.* 25–50 pg). Currently, the most commonly employed reduction technique involves a short dry-packed zinc or platinum oxide column, configured between the separation column and detector. Examples of chromatography achieved for a representative infant formula and margarine are illustrated in Figure 4, utilizing a C18 column and fluorescence detection.

Further theoretical and practical details are provided in the methodological references listed in the Further Reading section.

See also: **Chromatography**: High-performance Liquid Chromatography; **Coenzymes**; **Food Fortification**; **Infant Foods**: Milk Formulas; **Mass Spectrometry**: Principles and Instrumentation; Applications; **Spectroscopy**: Overview; Fluorescence; **Vitamins**: Overview

Further Reading

Ball GFM (1988) *Fat-soluble Vitamin Assays in Food Analysis: A Comprehensive review*. London: Elsevier.

Bolton-Smith C, Price RJG, Fenton ST, Harrington DJ and Shearer MJ (2000) Compilation of a provisional database for the phylloquinone (vitamin K_1) content of foods. *British Journal of Nutrition* 83: 389–399.

Booth SL, McKeown NM, Lichtenstein AH *et al.* (2000) A hydrogenated form of vitamin K: its relative bioavailability and presence in the food supply. *Journal of Food Composition and Analysis* 13: 311–317.

Booth SL and Sadowski JA (1997) Determination of phylloquinone in foods by high-performance liquid chromatography. *Methods in Enzymology* 282: 446–456.

Davidson RT, Foley AL, Engelke JA and Suttie JW (1998) Conversion of dietary phylloquinone to tissue menaquinone-4 in rats is not dependent on gut bacteria. *Journal of Nutrition* 128: 220–223.

Eitenmiller RR and Landon WO (1999) Vitamin K. In: Eitenmiller RR and Landon WO (eds) *Vitamin Analysis for the Health and Food Sciences*, pp. 149–184. Boca Raton: CRC Press.

Indyk HE and Woollard DC (2000) Determination of vitamin K in milk and infant formulas by liquid chromatography: Collaborative study. *Journal of the Association of Analytical Chemists International* 83: 121–130.

Piironen V and Koivu T (2000) Quality of vitamin K analysis and food composition data in Finland. *Food Chemistry* 68: 223–226.

Schurgers LJ, Geleijnse JM, Grobbee DE *et al.* (1999) Nutritional intake of vitamins K_1 (phylloquinone) and K_2 (menaquinone) in The Netherlands. *Journal of Nutrition and Environmental Medicine* 9: 115–122.

Shearer MJ (1992) Vitamin K metabolism and Nutriture. *Blood Reviews* 6: 92–104.

Shearer MJ, Bach A and Kohlmeier M (1996) Chemistry, nutritional sources, tissue distribution and metabolism of vitamin K with special reference to bone health. *Journal of Nutrition* 126: 1181S–1186S.

Suttie JW (1991) Vitamin K. In: Machlin LJ (ed.) *Handbook of Vitamins*, 2nd edn. pp. 145–194. New York: Marcel Dekker.

Physiology

M J Shearer, St Thomas's Hospital, London, UK

Background

From the 1930s until the 1970s, the only known role of vitamin K was as an essential antihemorrhagic factor, which initially had been shown to promote the synthesis of active prothrombin (factor II), but later also of factors VII, IX, and X. Collectively, these four factors, all synthesized in the liver and then released into the circulation, became known as the classical vitamin K-dependent coagulation factors. In 1974, it was shown that vitamin K was required as a cofactor for the enzymic conversion of specific glutamate residues in the prothrombin molecule to γ-carboxyglutamate (Gla) residues. The elucidation of this biochemical pathway leading to the synthesis of a new amino acid was the impetus that led to the discovery of many more vitamin K-dependent proteins. These proteins, also known as Gla proteins, are now known to have a widespread tissue distribution. They include further coagulation proteins (proteins C, S, and Z) and others with putative roles in the regulation of bone turnover (osteocalcin), and bone calcification (matrix Gla protein (MGP)). New evidence suggests that MGP is also a powerful inhibitor of vascular calcification. Other suspected functions of Gla proteins are in vascular repair processes (Protein S), cell-cycle regulation, cell–cell adhesion (Gas6), and signal transduction (Gas6 and proline-rich Gla proteins). For the most part, the precise physiological function of these new Gla proteins is poorly understood. However, nutritional and epidemiological studies in humans suggest that optimal intakes of vitamin K are necessary to maintain skeletal and cardiovascular health. Despite the possibility of these new health roles, the only clearly defined deficiency syndrome is bleeding due to a lack of functional molecules of the vitamin K-dependent coagulation factors II, VII, IX, and X. Bleeding due to vitamin K deficiency is rare in adults but is more common in infants during the first 6 months of life. The potential risk of serious life-threatening bleeding in infancy means that most countries recommend that all newborns should receive vitamin K prophylaxis.

Intestinal Absorption and Bioavailability

Dietary vitamin K, mainly as phylloquinone, is absorbed chemically unchanged from the proximal intestine after solubilization into mixed micelles composed of bile salts and the products of pancreatic lipolysis. In healthy adults, the efficiency of absorption of phylloquinone in its free form is about 80%. The bioavailability of vitamin K from foods varies according to food type and the associated matrix. Thus, the efficiency of absorption of phylloquinone may be as low as 10% from green leafy vegetables (e.g., spinach) in which the vitamin is tightly associated with chloroplast membranes. Less is known about the absorption of dietary menaquinones, but the efficiency of absorption decreases with increasing length of the side-chain.

Menaquinones (MKs) are also synthesized by the intestinal microflora, predominately in the colon. Major forms are MK-10 and MK-11 synthesized by *Bacteroides* species, MK-8 by *Enterobacteria*, MK-7 by *Veillonella*, and MK-6 by *Eubacterium lentum*. It is commonly held that microfloral synthesis provides an important nutritional source of vitamin K for humans, but direct evidence is lacking. A microfloral contribution of vitamin K is supported by the findings that very-long-chain MKs 10–13 are rare in foods but make up the majority of hepatic stores. However, all experimental evidence suggests that the absorption of menaquinones from the large intestine is an extremely inefficient process.

Transport in Plasma and Tissue Distribution

After intestinal absorption, vitamin K is incorporated into chylomicrons, secreted into the lymph, and enters the blood via the lacteals. Once in the circulation, phylloquinone is rapidly cleared at a rate consistent with its continuing association with chylomicrons and the chylomicron remnants that are produced by lipoprotein lipase hydrolysis at the surface of capillary endothelial cells. After an overnight fast, more than half of the circulating phylloquinone is still associated with triglyceride-rich lipoproteins, with the remainder being equally distributed between low-density and high-density lipoproteins. Phylloquinone is the major circulating form of vitamin K, but menaquinones MK-7 and MK-8 have been detected at lower concentrations.

Dietary vitamin K is delivered to the liver and possibly other tissues, including bone marrow, in the form of chylomicron remnants. The cellular uptake of vitamin K is mediated by apolipoprotein E in the chylomicron remnants and heparin sulfate proteoglycans on the cell surface. The liver, the site of synthesis of the vitamin K-dependent coagulation proteins, has a limited capacity for storing phylloquinone. Thus, the hepatic reserves of phylloquinone are several thousand-fold lower than those of vitamin A, despite the fact that dietary intakes of phylloquinone are

only about 10-fold lower. The distribution of molecular forms of vitamin K in the liver is quite different from that in plasma in that the major transport form, phylloquinone, represents only about 10% of total stores; the remainder comprise bacterial menaquinones, mainly MKs 7–13. Long-chain menaquinones are also present in other nonhepatic tissues but at lower concentrations than the liver. Some tissues such as the brain and pancreas seem to have the capacity to selectively convert phylloquinone to MK-4, but little is known about the biochemistry or physiological significance of this enrichment.

Catabolism and Turnover

The metabolic transformations leading to the elimination of vitamin K from the body occur exclusively in the liver. Phylloquinone is rapidly degraded with about 20% of a single dose being excreted in the urine within 3 days and about 40–50% excreted in the feces via the bile. The hepatic reserves of individuals on a low-phylloquinone diet decline rapidly. The overall picture is that phylloquinone has a fast turnover time, suggesting that the body stores are being constantly replenished.

The route of hepatic catabolism leading to the urinary excretion of vitamin K proceeds by oxidative degradation of the phytyl side-chain, probably involving the same enzymes used for the ω-methyl and β-oxidation of fatty acids, steroids, and prostaglandins. The end products of this pathway are shown in Figure 1. Two major aglycone metabolites have been identified; they are carboxylic acids with 5 and 7 carbon atom side-chains that are excreted in the urine as glucuronide conjugates. The biliary metabolites have not been clearly identified but are initially excreted as water-soluble conjugates and become lipid-soluble during their passage through the gut, probably through deconjugation by the intestinal microflora. There is no evidence that body stores of vitamin K can be conserved by an enterohepatic circulation, since vitamin K itself is not excreted in bile, and the side-chain shortened carboxylic acid metabolites are not biologically active.

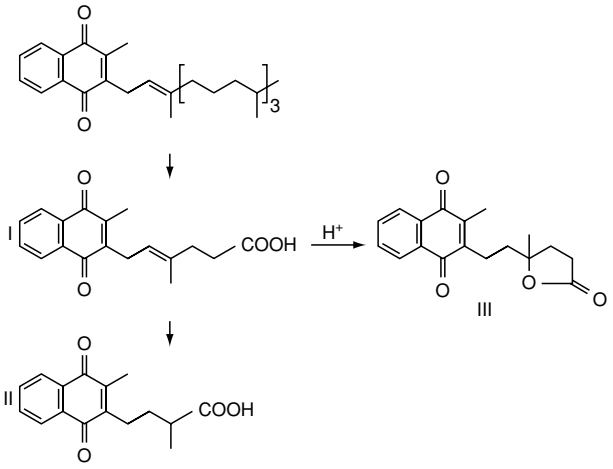

Figure 1 Simplified pathway and chemical structures of the two major urinary metabolites of phylloquinone in their aglycone forms (structures I and II). In acidic conditions, metabolite I may cyclize to produce phylloquinone γ-lactone (structure III). (I) 2-Methyl-3-(5′-carboxy-3′-methyl-2′-pentenyl)-1,4-naphthoquinone; (II) 2-methyl-3-(3′-carboxy-3′-methylpropyl)-1,4-naphthoquinone;(III) 2-methyl-3-(5′-carboxy-3′-hydroxy-3′-methylpentyl)-1,4-naphthoquinone lactone.

Cofactor Role of Vitamin K

Vitamin K-dependent Carboxylation of Glutamate Residues

The only unequivocal biochemical role for vitamin K is as a cofactor for a unique posttranslational chemical modification confined to a small group of unusual calcium-binding proteins. The vitamin K-dependent step is a carboxylation reaction that transforms selective glutamic (Glu) residues in the protein precursor into Gla. The reaction shown in Figure 2 is catalyzed by a microsomal enzyme called vitamin K-dependent (γ-glutamyl) carboxylase and requires molecular oxygen and carbon dioxide. The form of vitamin K required as the cofactor is not the stable quinone structure found in the diet but the reduced quinol (or hydroquinone) structure (KH_2). The subsequent oxidation of vitamin KH_2 to vitamin K 2,3 epoxide is coupled to the abstraction of a proton from the γ-carbon of the Glu residue to generate a carbanion that reacts with CO_2 to yield the final Gla product. The biological relevance of the Gla structure is that it forms a cage structure to which divalent metal ions such as calcium may be bound. In the case of the vitamin K-dependent procoagulants, this induces the conformational change and

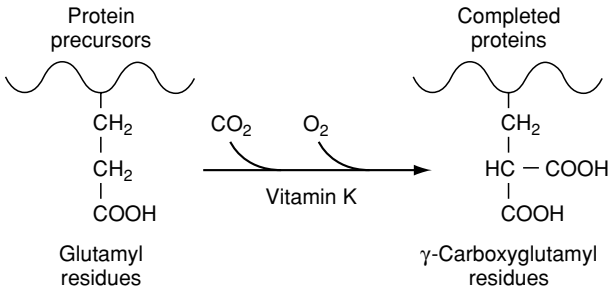

Figure 2 Posttranslational vitamin K-dependent carboxylation reaction showing the conversion of protein-bound glutamyl residues (Glu) to γ-carboxyglutamate (Gla) residues.

membrane-binding properties that are essential for their biological activity.

Vitamin K Epoxide Cycle

In all tissues and cells that synthesize Gla proteins, vitamin K-dependent carboxylation is intimately linked to a metabolic sequence known as the vitamin K-epoxide cycle. This cycle and the associated enzyme activities are shown in Figure 3. During γ-glutamyl carboxylation, vitamin KH_2 is transformed to vitamin K epoxide (enzyme activity E_1). The epoxide product is recycled in two steps; first by a vitamin K epoxide reductase activity (E_2) to produce vitamin K quinone, and second by a quinone reductase activity (E_3) to produce the coenzyme vitamin KH_2. Both these activities are thiol-dependent and are probably effected by the same enzyme.

An important property of the dithiol-dependent epoxide and quinone reductase activities is their sensitivity to certain antagonists, especially those based on 4-hydroxycoumarin (e.g., warfarin) or indandione structures. Drugs based on these structures have long been used as oral anticoagulants, and it is now clear that their anticoagulant action is based on their ability to inhibit the epoxide reductase activity (E_2) and block the recycling of the vitamin. The dithiol-dependent quinone reductase (E_3) is also sensitive to warfarin, but there is a second quinone reductase activity catalyzed by an NAD(P)H-dependent enzyme activity (E_4) that is less sensitive to warfarin inhibition, and this provides an alternative pathway for the reduction of vitamin K to KH_2 in the presence of warfarin or other oral anticoagulant drugs. Therapeutic anticoagulation is dependent on a balance being achieved between the inhibition of the recycling enzymes and the amount of dietary vitamin K that can enter the cycle to support carboxylation at a reduced efficiency.

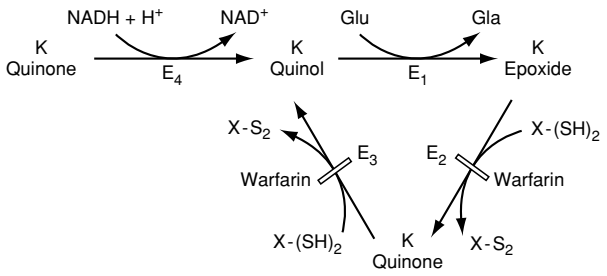

Figure 3 Scheme showing the reactions and enzyme activities (denoted by E_1, E_2, E_3, and E_4) of the vitamin K-epoxide cycle. The active form of vitamin K needed for the conversion of glutamate residues (Glu) to γ-carboxyglutamate (Gla) residues is the reduced form, vitamin K quinol (K quinol). The carboxylation reaction is driven by a vitamin K-dependent carboxylase activity (E_1), which simultaneously converts K quinol to vitamin K 2,3-epoxide (K epoxide). K epoxide is reduced back to the quinone (K quinone) by vitamin K epoxide reductase (E_2). The cycle is completed by the reduction of recycled K quinone by a vitamin K reductase activity (E_3). Both E_2 and E_3 reductase activities are dithiol-dependent (X-(SH_2) and X-S_2 denote reduced and oxidized dithiols) and are inhibited by coumarin anticoagulants such as warfarin. Dietary vitamin K (K quinone) may enter the cycle via an NAD(P)H-dependent vitamin K reductase activity (E_4), which is not inhibited by warfarin.

Vitamin K-dependent (Gla) Proteins

Distribution and Structure

γ-Glutamyl carboxylase activity is present in a wide variety of extrahepatic tissues such as bone, kidney, placenta, pancreas, skin, spleen, lung, testis, and the vessel wall. New Gla proteins continue to be discovered; the present list together with their tissue distribution is shown in Table 1. Some of these Gla proteins are found in only one major tissue (e.g., procoagulant factors in liver and osteocalcin in bone), while others (e.g., matrix Gla protein and protein S) are more widely distributed.

All Gla-containing proteins are secretory proteins and are characterized by an intracellular precursor form possessing a leader sequence that facilitates specific posttranslational events and is then normally cleaved before secretion. Within the leader sequence, a hydrophobic presequence facilitates the translocation of the precursor Gla proteins across the endoplasmic reticulum, while a second prosequence serves as a recognition signal for the vitamin K-dependent carboxylase. With the exception of matrix-Gla protein, the prosequence of currently characterized Gla proteins is located immediately before the N-terminal of the mature protein. In matrix-Gla protein, the carboxylation recognition site is internal and is retained in the mature protein. The number of Gla residues varies: the four classical vitamin K-dependent clotting factors contain 10–12 Gla residues, whereas the much smaller MGP and osteocalcin contain five and three Gla residues, respectively.

Role of Gla proteins

Hemostasis (procoagulants and anticoagulants) The vitamin K-dependent coagulation proteins comprise four plasma procoagulants (factors II, VII, IX, and X) and two feedback anticoagulants (proteins C and S). The physiological function of a seventh plasma protein (protein Z) is less established but seems to play a role in dampening coagulation by acting as a cofactor for the inhibition of factor Xa. The four vitamin K-dependent procoagulants have long been known to play a central role in the blood clotting cascade leading to the formation of a fibrin clot. Circulating

Table 1 Distribution and roles of vitamin K-dependent (Gla) proteins

Gla protein	*Tissue*	*Role*
Prothrombin (factor II), factors VII, IX, and X	Liver (then plasma)	Blood clotting (procoagulants)
Protein C	Liver (then plasma)	Blood clotting (anticoagulant)
Protein S	Liver (then plasma), endothelium, bone	Anticoagulant role (cofactor for protein C); role in bone unknown
Osteocalcin (or bone Gla protein)	Bone	Implicated in bone remodeling
Matrix Gla protein	Most tissues	Inhibits calcification of arteries and cartilage
Gas 6	Most tissues	Implicated in cell-cycle regulation, cell survival, cell proliferation, cell adhesion, platelet signaling
Proline-rich Gla proteins	Most tissues	Unknown, but protein structures suggest roles as cell-surface receptors

as inactive forms (or zymogens) of serine proteases, their biological activity is dependent on their ability to bind to anionic phospholipid surfaces where cleavage of the zymogens yields the active protease clotting factors. The Gla residues of these clotting factors provide an efficient chelating site for calcium ions that enables ion bridges to be made between the factor and the surface phospholipids of platelets and endothelial cells where, together with other cofactors, they form membrane-bound enzyme complexes. The intricate control of coagulation is illustrated by the fact that another Gla protein (protein C), once activated, performs an anticoagulant role by specifically degrading phospholipid-bound activated factors V and VIII in the presence of calcium. This anticoagulant activity of activated protein C is dependent on yet another Gla protein (protein S) that acts as a synergistic cofactor by enhancing the binding of activated protein C to negatively charged phospholipids.

Bone metabolism The discovery in 1975 that bone tissue contains a Gla protein, which is also one of the most abundant proteins in the body, has revolutionized earlier thinking that vitamin K is only needed for coagulation. This protein, called osteocalcin or bone Gla protein, accounts for about 15% of the noncollagenous proteins in the bone of most vertebrate species, although this proportion is less in humans. The three Gla residues of osteocalcin project away from the Gla helix, and their periodic spacing complements the interatomic spacing of the calcium ions of the hydroxyapatite lattice; this is thought to account for the strong mineral-binding characteristics of the fully carboxylated form of osteocalcin. The adsorption affinity of osteocalcin for hydroxyapatite and other features such as its high degree of conservation throughout evolution, its regulation by vitamin D, and its chemotractant properties suggest an important role for osteocalcin in bone metabolism, although the nature of this role still remains unknown. Osteocalcin is specifically synthesized by osteoblasts, and since its concentration in blood reflects osteoblastic activity, measurements of total osteocalcin have become clinically important in the diagnosis of bone disease.

Besides osteocalcin, bone tissue contains lower concentrations of other Gla proteins; matrix Gla-protein (MGP), protein S, and Gas6 (Table 1). MGP is a 10-kDa secretory protein, with five residues of Gla. It was first isolated from demineralized extracts of bone by Price in 1983 but is expressed by a wide variety of tissues and cell types. Its function in bone is unclear but is probably related to its capacity to inhibit localized calcification as recently shown for arteries. Transgenic MGP-deficient mice show inappropriate calcification of the cartilage, including the growth plate, and a phenotype characterized by short stature, osteopenia, and fractures. This function is consistent with the effects of the vitamin K antagonist warfarin, which, when given *in utero* or to young animals, results in excessive mineralization and premature closure of the epiphyseal growth plate. An important role for Gla proteins during human fetal bone development is suggested by the bone defects often seen in infants born to mothers who had received vitamin K antagonists such as warfarin during the first trimester of pregnancy. Evidence that warfarin embryopathy is mediated by a dysfunctional MGP comes from the demonstration that mutations encoding MGP are responsible for a rare inherited disorder called Keutal syndrome, which is characterized by a similar abnormal calcification of cartilage. The functions of protein S and Gas6 in bone are presently unknown.

Vascular calcification A series of remarkable studies since 1997 have firmly established MGP as a potent inhibitor of calcification of arteries and cartilage. Thus, targeted deletion of the *MGP* gene in mice resulted in rapid focal calcification of the elastic

lamellae of the arterial media but not of the arterioles, capillaries, or veins. This process begins at birth and by 3–6 weeks is sufficiently severe to cause the arteries to become rigid tubes that fracture, causing death from bleeding. Most of the phenotypic features of arterial calcification seen in MGP-deficient mice can be mimicked, albeit to a somewhat muted degree, by the vitamin K antagonist warfarin. This effect of warfarin points to the need for Gla residues, and hence vitamin K, to the function of MGP. Overall, the dramatic features of both genetic and biochemical models of MGP depletion strongly suggest that MGP is central to the process by which calcification of arteries is normally inhibited *in vivo* and that there is no effective alternative inhibitor at aortal sites.

Assessment of Vitamin K Status

The traditional criterion for vitamin K sufficiency is the maintenance of normal plasma concentrations of fully γ-carboxylated vitamin K-dependent procoagulants II, VII, IX, and X. The presence of a clinically significant vitamin K-deficient state is usually investigated by global coagulation assays such as the one-stage prothrombin time. Such assays are insensitive to small changes in γ-carboxylation of the vitamin K-dependent coagulants and, while appropriate for detecting overt vitamin K deficiency, are not useful for picking up suboptimal states of vitamin K deficiency.

Modern assessments of vitamin K status need to take into account both coagulation and noncoagulation roles of Gla proteins and their different sites of synthesis. When vitamin K is in short supply, or its action is blocked by an oral anticoagulant such as warfarin, undercarboxylated (des-γ-carboxy) species of Gla proteins are produced at the tissue site of synthesis and, in some species, including humans, are secreted into the circulation. Collectively, these undercarboxylated species are known as **p**roteins **i**nduced by **v**itamin **K** **a**bsence or antagonism (PIVKA). In the case of the coagulation protein prothrombin, specific and sensitive immunoassays for undercarboxylated prothrombin (PIVKA-II) provide a powerful marker of the coagulation status of vitamin K because they can detect small decreases in the Gla content of prothrombin well before any changes occur in conventional coagulation tests. Similarly, the assay of circulating undercarboxylated species of osteocalcin provides a specific functional marker to assess the sufficiency of vitamin K for the carboxylation of bone Gla proteins. Since prothrombin and osteocalcin are synthesized in the liver and bone, respectively, the detection of their respective undercarboxylated species allows the vitamin K status of these tissues to be assessed independently. Such studies have shown that the carboxylation of osteocalcin in bone is more susceptible to reduced dietary intakes of vitamin K than the carboxylation of the classical coagulation proteins. Measurements of undercarboxylated osteocalcin are being increasingly used to assess the importance of optimal vitamin K intakes with respect to bone health. The excretion of free Gla in urine represents the final catabolic product of all Gla proteins, and its measurement has also been shown to reflect dietary intakes and status. Another method for assessing vitamin K status is from direct serum assays of phylloquinone, the predominant transport form of vitamin K. A low serum phylloquinone is a good indicator of a poor overall vitamin K status.

Vitamin K and Health

Role in Hemostasis

Overt vitamin K deficiency in adults, resulting in clinical bleeding, is almost unknown except as a consequence of underlying disease, most commonly resulting from hepatointestinal disorders that disturb the absorption or utilization of the vitamin. However, oral coumarin anticoagulant drugs, such as warfarin, are widely used in the prevention and management of thromboembolic disease. These vitamin K antagonists are used to induce an artifical but controlled state of vitamin K deficiency, the aim being to reduce the circulating levels of the vitamin K-dependent clotting factors to within a predetermined target range depending on the clinical condition.

In contrast to adults, bleeding due to vitamin K deficiency may occur spontaneously in the first few months of life. The resulting bleeding syndrome is traditionally known as hemorrhagic disease of the newborn, in recognition of the Boston physician Townsend, who is credited with the first description of the syndrome in 1894. The modern name is vitamin K deficiency bleeding (VKDB) of early infancy. Until the late 1960s, VKDB was considered to be solely a problem of the first week of life (classical VKDB), but two other forms (early and late VKDB) have now been recognized (Table 2). Late VKDB, peaking between the third and sixth weeks of life, is the most common syndrome and is still a significant worldwide cause of infant morbidity and mortality. In different parts of the world, some two to 100 cases per 100 000 births of late VKDB have been reported, with a particularly high incidence in the Far East. Unlike the classical form, late VKDB has a high incidence of intracranial hemorrhage, resulting in death or severe and permanent brain damage in

Table 2 Classification of vitamin K deficiency in the newborn (VKDB)

VKDB syndrome	*Time of presentation*	*Common bleeding sites*	*Etiological factors*
Early	0–24 h	Cephalohematoma, intracranial, intrathoracic, intraabdominal	Maternal drugs (e.g., warfarin, anticonvulsants)
Classical	1–7 days	Gastrointestinal, skin, nasal, circumcision	Mainly idiopathic, breast-feeding (may be related to low milk intakes)
Late	2–26 weeks	Intracranial, skin, gastrointestinal	Mainly idiopathic, breast feeding, some degree of cholestasis often present; May reflect underlying disease (e.g., biliary atresia, α-1-antitrypsin deficiency, cystic fibrosis)

around 50% of cases. One major risk factor, exclusive breast-feeding, is common to both classical and late syndromes and is probably related to the relatively low concentrations of vitamin K in breast milk ($1–2\,\mu g\,l^{-1}$) compared to about $50\,\mu g\,l^{-1}$ in formula feeds. Another probable etiological factor is the precarious hepatic stores of vitamin K in newborns because of poor placental transfer of phylloquinone and the lack of neonatal stores of bacterial menaquinones, which only build up slowly after bacterial colonization of the intestine. Some infants who develop late VKDB have abnormalities of liver function, leading to cholestasis and an impaired absorption efficiency of vitamin K.

By the early 1980s, reports in the literature were suggesting a worldwide rise in the incidence of late VKDB. This resurgence may have been related to both a decline in routine vitamin K prophylaxis and the promotion of exclusive breast-feeding. In recent years, many countries have reemphasized the need to give all babies extra vitamin K as a public-health measure. Often, this was given as a single intramuscular dose at birth, but since 1992, concerns about an epidemiological association between the intramuscular route and later childhood cancer has seen a shift in some countries towards oral prophylaxis. However, there is evidence that even multidose oral regimens fail to give full protection against late VKDB and that this is because there is an underlying reservoir of infants who have a reduced efficiency of absorption due to underlying and unrecognized cholestasis. Several epidemiological studies conducted since 1992 have not substantiated the link between vitamin K prophylaxis and cancer.

Role in Bone Health

Despite mechanistic gaps in our knowledge, a picture is emerging that a suboptimal vitamin K status may play a role in the pathogenesis of osteoporosis. Much of this evidence is still circumstantial, being based on associations of fracture risk and bone mineral density with indices of poor vitamin K status and low dietary intakes. Of particular note is that undercarboxylated osteocalcin is an independent risk factor of both hip-fracture risk and bone mineral density. However, osteoporosis is a complex disease, and without a fuller understanding of the biological function of osteocalcin, the interpretation of these associations with vitamin K status remains uncertain.

In the near future, the results from a number of long-term intervention studies on the effects of vitamin K supplementation on bone health will become available. These may help to answer the question of whether extra intakes of vitamin K can influence bone mineral density and bone turnover. Most of these studies have used phylloquinone because this is the major dietary form, but in Japan, the effects of very large doses of MK-4 (menatetranone) on bone health are being studied. Both phylloquinone and MK-4 are known to promote the γ-carboxylation of osteocalcin, but these forms differ in other important respects. Thus, there is evidence that tissues can convert phylloquinone into MK-4 *in vivo* (but not vice versa) and that MK-4 may have a direct inhibitory effect on bone resorption. In an intervention study in which Japanese women with osteoporosis were treated with a high-dose MK-4 regimen, the incidence of fractures was significantly reduced, and the rate of bone loss was decreased.

Dietary Intakes and Requirements

The major dietary source of vitamin K is phylloquinone. In Western societies, the majority of phylloquinone intakes (up to 50%) comes from green leafy vegetables followed by certain vegetable oils such as soybean, canola (rapeseed), and olive oils. Mixed dishes and meals contribute about 15% of total phylloquinone intake, largely because of their oil content. The wide range of the vitamin K content of different food items means that individual daily intakes may vary widely, but average adult intakes mostly range from 60 to 120 μg per day. Intakes of this order are more than sufficient to maintain normal hemostasis, and even intakes as low as 10 μg per day for a few weeks have no effect on global coagulation assays.

However, more sensitive measures of vitamin K status (e.g., undercarboxylation of Gla proteins, urinary Gla excretion, and serum phylloquinone) respond readily to moderate changes in dietary intakes. One conclusion from nutritional depletion/repletion studies is that the daily intakes of vitamin K needed to sustain the posttranslational γ-carboxylation of extrahepatic Gla proteins such as osteocalcin are higher than those needed for the coagulation Gla proteins synthesized in the liver. Therefore, previous recommendations of a daily intake of 1 μg per kilogram of body weight, which have been based on the coagulation role of vitamin K, may need to be revised upwards if evidence continues to harden on other potential health benefits such as the amelioration of osteoporosis and arterial calcification. Studies in the UK and USA already suggest that a substantial proportion of their populations have intakes below 1 μg per kilogram of body weight. What this optimal intake should be must await further research on the physiological function of Gla proteins, the further delineation of functional markers of vitamin K status, and studies of the health consequences of different intakes.

See also: **Bone**; **Calcium**: Physiology; **Dietary Reference Values**; **Dietary Requirements of Adults**; **Dietary Surveys**: Surveys of National Food Intake; Surveys of Food Intakes in Groups and Individuals; **Infants**: Nutritional Requirements; **Lipoproteins**; **Microflora of the Intestine**: Role and Effects; **Osteoporosis**; **Vitamins**: Overview

Further Reading

Angelillo-Scherrer A, Garcia de Frutos P, Aparicio C *et al.* (2001) Deficiency or inhibition of Gas6 causes platelet dysfunction and protects mice against thrombosis. *Nature Medicine* 7: 215–221.

Benzakour O and Kanthou C (2000) The anticoagulant factor, protein S, is produced by cultured human vascular smooth muscle cells and its expression is up-regulated by thrombin. *Blood* 95: 2008–2014.

Booth SL and Suttie JW (1998) Dietary intake and adequacy of vitamin K. *Journal of Nutrition* 128: 785–788.

Broze GJ, Jr. (2001) Protein Z-dependent regulation of coagulation. *Thrombosis and Haemostasis* 86: 8–13.

Furie B, Bouchard BA and Furie BC (1999) Vitamin K-dependent biosynthesis of γ-carboxyglutamic acid. *Blood* 93: 1798–1808.

Institute of Medicine (2001) *Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc*, pp. 162–196. Washington, DC: National Academy Press.

Kulman JD, Harris JE, Xie L and Davie EW (2001) Identification of two novel transmembrane γ-carboxyglutamic acid proteins expressed broadly in fetal and adult tissues. *Proceedings of the National Academy of Sciences* 98: 1370–1375.

Luo G, Ducy P, McKee MD *et al.* (1997) Spontaneous calcification of arteries and cartilage in mice lacking matrix GLA protein. *Nature* 386: 78–81.

Nelsestuen GL, Shah AM and Harvey SB (2000) Vitamin K-dependent proteins. *Vitamins and Hormones* 58: 355–389.

Newman P and Shearer MJ (1998) Vitamin K metabolism. In: Quinn PJ and Kagan VE, (eds) *Subcellular Biochemistry, Volume 30, Fat-soluble Vitamins*, pp. 455–488. New York: Plenum Press.

Olson RE (1999) Vitamin K. In: Shils ME, Olson JA, Shike M and Ross AC (eds) *Modern Nutrition in Health and Disease* 9th edn, pp. 363–380. Baltimore, MD: Lippincott, Williams & Wilkins.

Shanahan CM, Proudfoot D, Farznaneh-Far A and Weissberg PL (1998) The role of Gla proteins in vascular calcification. *Critical Reviews in Eukaryotic Gene Expression* 8: 357–375.

Shearer MJ (1997) The roles of vitamins D and K in bone health and osteoporosis prevention. *Proceedings of the Nutrition Society* 56: 915–937.

Shearer MJ (2000) Role of vitamin K and Gla proteins in the pathophysiology of osteoporosis and vascular calcification. *Current Opinion in Clinical Nutrition and Metabolic Care* 3: 433–438.

Tsaioun KI (1999) Vitamin K-dependent proteins in the developing and aging nervous system. *Nutrition Reviews* 57: 231–240.

VITAMINS

Contents
Overview
Determination

Overview

P M Finglas, Institute of Food Research, Norwich Research Park, Norwich, UK

Overview

This article covers the definition, a brief historical account of the discovery, classification and nutritional role, forms found in food, and effects of processing on vitamin stability.

Vitamins are defined as a group of complex organic micronutrients present in small amounts in food or nutritional supplements that the body requires for its normal metabolism and function. They cannot be synthesized by humans in sufficient quantity for normal metabolic requirements. Certain substances that are considered vitamins can be synthesized in small amounts by intestinal-tract bacteria (e.g., B-groups). Each vitamin consists of a mixed group of compounds, which may not be related to each other chemically and are therefore classified by function. (*See* **Microflora of the Intestine:** Role and Effects.)

History

It was not until the early 1900s that the existence of the 'accessory food factors' was recognized. Indeed, the term 'vitamine' was first used by a Polish biochemist, named Funk, to describe an 'accessory food factor.' However, deficiency diseases were well known many centuries earlier, the earliest documented disease being beriberi, which was recorded in China around 2600 BC. Yet, it was not until the end of the eighteenth century that the nutritional etiology of the disease was established. In 1885, a Japanese physician named Takaki succeeded in reducing the effect of beriberi among Japanese sailors by feeding them a diet containing more meat and vegetables, and less rice. Then, in 1890, a Dutch doctor named Eijkman reproduced symptoms of the disease in chickens fed exclusively with shelled rice. His collaborator, Grijns, showed that these symptoms disappeared by adding rice husks to the chicken diet. Finally, in 1912, Funk isolated the antiberriberi factor from the cuticle of the rice and named it 'vitamine,' the 'life-essential amine.' This term was adopted to name all the nutritional factors responsible for deficiency diseases.

Similarly, scurvy and night blindness were described in ancient Egypt around 1500 BC. During the period AD 1400–1600, the explorers Vasco de Gamma and Magellan described in detail the scurvy and night blindness that had devastated their ships' crew. A century earlier, Jacques Cartier had succeeded in saving a good many sailors thanks to an Indian medicine based on pine needles. Finally, in 1747, James Lind found the preventive effect of citrus fruit using controlled experiments at sea, and this was shortly followed by the demonstration, during the voyages of Captain Cook in the period 1768–1771, that prolonged sea voyages were possible without the ravages of scurvy. (*See* **Hypovitaminosis A; Scurvy.**)

Thirteen vitamins were isolated, identified, and synthesized between 1929 and 1975. The last vitamin to be discovered was vitamin B_{12} in 1948. Many researchers who contributed to the fundamental discoveries have been honored by the Nobel Prizes Committee, among them: Szent-Gyorgyi and King for the simultaneous discovery of ascorbic acid in 1928; Karrer for his work on vitamin A and β-carotene between 1928 and 1933; Dam and Doisy for the discovery and chemical nature of vitamin K in 1943; and Hodgkin for the structural determination of vitamin B_{12} in 1964. (*See* **Carotenoids**: Physiology.)

Nowadays, vitamin research is still very active. The understanding of the precise, complex role of vitamins has made remarkable progress owing to the study of vitamin-dependent metabolite diseases of genetic origin. More recently, because of their antioxidant properties, some fat-soluble vitamins, notably vitamins A, E, and C, and provitamin A compounds (carotenoids; including β-carotene, lutein, and lycopene), are also considered to play an important role in the prevention against chronic degenerative diseases and cancers. Also, it has been shown that folic acid can reduce significantly the incidence of neural tube defects (such as spina bifida), and folates, in general, may protect against vascular diseases and some cancers, notably colonic cancer.

(*See* **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; Synthetic Antioxidants, Characterization and Analysis; Role of Antioxidant Nutrients in Defence Systems; **Cancer**: Epidemiology; Carcinogens in the Food Chain; **Coronary Heart Disease**: Etiology and Risk Factor; **Pregnancy**: Maternal Diet, Vitamins, and Neural Tube Defects; **Vitamin B_6**: Properties, and Determination.)

Classification and properties

A number of factors can make the identification and classification of vitamins difficult: (1) a vitamin can exist in a number of different structures with similar, or different, biological activities; (2) the biological activity is organism-dependent; (3) the physiologically or metabolically active form of a vitamin *in vivo* may be different from that found in food (e.g., B vitamins are active in coenzyme forms; 1,25-dihydroxycholecalciferol is the active form of vitamin D_3; (4) several forms of the same vitamin can interconvert during extraction (e.g., pyridoxine, pyridoxamine, pyridoxal, and their phosphorylated derivatives; and (5) synthetic forms of a vitamin are often available commercially in a more stable form, e.g., α-tocopherol acetate and retinol acetate are more stable compared with the free vitamins and are converted to the letter in the body.

Vitamins are generally classified by their solubility characteristics into fat-soluble and water-soluble groups. A full list of vitamins, forms, functions, and health benefits is given in Table 1. The fat-soluble vitamins (A, E, D_3, and K, and carotenoids) are derived from the isoprenoid pathway, whereas the synthesis of the water-soluble vitamins (B_1, B_2, B_6, B_{12}, C, biotin, niacin, and pantothenic acid) is much less connected, and several of them have multiple biosynthetic pathways. Other 'vitamin-like' compounds are sometimes associated with the classical vitamin compounds. These include vitamins F (essential fatty acids), L (*o*-aminobenzoic acid), P (bioflavonoids), U (L-methioninylmethylsulfonium chloride), and B_{13} (orotic acid), choline, α-lipoic acid, myoinositol, plastoquinone, pyrroloquinoline quinone, and ubiquinones. *Refer to* individual vitamins.

During the early years of vitamin discovery, their chemical composition was largely unknown and were each assigned a letter of the alphabet for convenience. This was further complicated when it was found that the activity attributed to a single vitamin was, in some instances, the result of a combination of several vitamins, e.g., B-complex vitamins. Once a vitamin had been isolated from a food and its chemical structure determined, variations in the structure were found within compounds having the same vitamin activity but in different species. A system of numerical subscripts was therefore adopted to overcome this problem, e.g., D_2 and D_3.

In order to simplify the system, letters were sometimes replaced with names, e.g., thiamin (vitamin B_1), riboflavin (B_2), pantothenic acid (B_5) and biotin (vitamin H), based on the chemical structure of a specific function or source. 'Thiamin' contains the prefix 'thi,' which is derived from the Greek word for sulfur, and refers to its sulfur content. 'Riboflavin' indicates the chemical structure and is derived from the chemical names 'ribose' (a pentose sugar) and 'flavin' (a heterocyclic ketone). 'Pantothenic acid' is derived from the Greek word *pantos*, meaning 'found everywhere.' 'Biotin' originates from the German word *haut*, meaning skin, as this vitamin protects the skin.

Occurrence in Food

One of the main differences between fat- and water-soluble vitamins is their distribution in biological tissues. The water-soluble vitamins are found in all living tissues, whereas the fat-soluble vitamins are completely absent from some tissues. Vitamins A and D occur widely in plant tissues in the form of a provitamin. Similarly, the amino acid tryptophan can be converted into nicotinic acid in the body. Although it is generally assumed that sufficient amounts of vitamins can be absorbed from a balanced diet for normal requirements, the use of vitamin supplements, either as single compounds or in multivitamin products, is widely consumed. The major vitamin forms found in foods, some commercial forms, and some examples of rich food sources, are given in Table 2.

Nutritional Aspects

Vitamins are required in trace amounts in the diet for the maintenance of optimal health, growth, and reproduction. Deficiency symptoms will occur if a single vitamin is omitted from a diet of a species that requires it. Many of the vitamins act as coenzymes, whereas others have no single role but perform certain essential functions. A list of known functions for each vitamin is given in Table 1. Vitamin requirements based on metabolic needs are similar for animals and humans, but dietary needs widely among species. (*See* **Coenzymes**.)

The recommended intakes of nutrients, including vitamins, are intended as daily intake guidelines for preventing deficiencies and maintaining body reserves among healthy individuals. The values are derived from studies on volunteers on nutrient balance studies, measures of tissue saturation, normal vitamin intakes in the general population, and extrapolation

Table 1 Fat- and water-soluble vitamins: forms, function, and potential health benefits

Vitamin	*Vitamers and related forms*	*Function*	*Potential health benefits*
Fat-soluble vitamins			
A	Retinal, retinoic acid, retinol *cis*-isomers, β-carotene, other carotenoids	Needed for normal vision, reproduction, and various tissues; immune function	Protection against various cancers
D	Cholecalciferol (D_3), ergocalciferol (D_2), 1-α-25-dihydroxycholecalciferol	Enhances calcium absorption; regulates calcium/phosphate metabolism; promotes bone mineralization	Reduces risk of osteoporosis
E	All-rac-α-tocopherol, β-, δ-, γ-tocopherols, various tocotrienols	Prevents lipid oxidation of membranes and PUFAs from autoxidation; normal immune function and healthy tissues	Reduces risk of cardiovascular disease, precancerous lesions, and cancer
K	Phylloquinone (K_1), menaquinone (K_2), menadione (K_3)	Essential for normal blood clotting, formation/maintenance of healthy bones	Reduces risk of blood clotting and osteoporosis
Water-soluble vitamins			
B_1	Thiamin, thiamin monophosphate (TMP), thiamin pyrophosphate (cocarboxylase, TTP) complexed to proteins	Helps convert carbohydrates into energy metabolism; required by brain, nervous system and heart	Important role in energy metabolism
B_2	Flavin mononucleotide (FMN), flavin adenine dinucleotide (FAD)	Involved in carbohydrate, protein and fat metabolism; helps release energy into cells	Important role in energy metabolism; healthy skin
Niacin	Nicotinic acid, niacinamide (nicotinamide), nicotinamide adenine dinucleodite (NAD) and nicotinamide adenine dinucleotide diphosphate (NADP)	Involved in carbohydrate, protein and fat metabolism	Reduces risk of arteriosclerosis
B_6	Pyridoxine, pyridoxal, pyridoxamine and related phosphorylated forms	Essential for proper protein utilization; involved in homocysteine metabolism	Reduces risk of cardiovascular disease (with folate and B_{12}) and osteoporosis
B_{12}	Cyanocobalamin, methyl- and hydroxycobalamins, adenosylcobalamin (coenzyme B_{12})	Needed for red blood cell formation, DNA and RNA syntheses, involved in homocysteine metabolism	Prevents certain anemias and reduces risk of cardiovascular disease
Folic acid (folacin)	Pteroylmonoglutamic acid, 5-methyl- and 10-formyltetrahydrofolic acid, tetrahydrofolic acid as both mono- and polyglutamates	Involved in amino acid interconversions and methylation reactions, needed for red blood cell formation, DNA and RNA syntheses, involved in homocysteine metabolism and protects against neural tube defects	Prevents certain anemias and birth defects, reduces risk of cardiovascular disease (with B_6 and B_{12}) and certain cancers
Pantothenic acid	Coenzyme A, pantetheine and acyl carrier protein (ACP)	Active in protein, carbohydrate and fat metabolism, formation for certain nerve-regulators	General health and well-being
Biotin	Biotin, desthiobiotin	Active in protein, carbohydrate, and fat metabolism	Helps develop healthy nails
C	Ascorbic and dehydroascorbic acids, isoascorbic acid	Helps form/maintain collagen; important for healthy tissues and wound healing; enhances iron absorption; protects against free-radical damage and infection	Protects against cancer, cardiovascular and eye diseases; reduces symptoms of colds and flu, and risks of osteoporosis

Adapted from Francis FJ (ed.) (2000) Vitamins: survey. In: *Encyclopedia of Food Science and Technology*, 2nd edn., vol. 4, pp. 2440–2449. New York: Wiley.

Table 2 Vitamins: commercial forms, food forms and rich dietary sources

Vitamin	*Major commerical forms*	*Main food forms*	*Rich dietary sources*
Fat-soluble vitamins			
A	Retinol acetate & palmitate, β-carotene	Retinly esters, carotenes (provitamin A compounds)	Fatty fish, butter, margarines and liver
D	Cholecalciferol, ergocalciferol	Cholecalciferol	Fatty fish, eggs, margarines
E	(*RRR*)-α-Tocopherol, (*RRR*)-α-tocopherol acetate, all-*rac*-α-tocopherol, all-*rac*-α -tocopherol succinate	Tocopherols (α-, β- & γ-forms) and tocotrienols (α- & β-)	Vegetable oils
K	2-(*R,S*-)-Vitamin K_1 (*cis/trans* mixture), menadione sodium bisulfite	Phylloquinone (K_1), menaquinones (K_2)	Green leafy vegetables, beef, liver
Water-soluble vitamins			
B_1	Thiamin chloride HCl, thiamin mononitrate, thiamin diphosphate	Thiamin; thiamin mono- and pyrophosphates complexed to proteins	Cereals, potatoes, liver, yeast extract
B_2	Riboflavin, riboflavin sodium phosphate	Riboflavin; riboflavin mononucleotide, flavin adenine dinucleotide complexed to proteins	Liver, milk, eggs, green vegetables
Niacin	Niacin, niacinamide	Nicotinic acid; nicotinamide adenine dinucleotide (NAD) and its phosphate form (NADPH) complexed to proteins	Meat and meat products, cereals
B_6	Pyridoxine hydrochloride, pyridoxal-5′-phosphate	Pyridoxine, pyridoxal and pyridoxamine and their phosphorylated forms bound to proteins	Meat and meat products, eggs, cereals
B_{12}	Cyanocobalamin, hydroxycobalamin	Methyl-, adenosyl-, hydroxy-, and sulfitocobalamins	Kidneys, eggs, milk, and cheese
Folic acid (folacin)	Folic acid	Folic acid; 5-methyltetrahydrofolate, 10-formyltetrahydrofolate, tetrahydrofolate as mono- and polyglutamates	Leafy green vegetables, cereals, yeast, and vegetable extracts
Pantothenic acid	Calcium pantothenate, D-panthenol	Pantothenic acid, coenzyme A and acyl carrier protein (ACP)	Liver, kidney, yeast, wheat germ, and egg yolk
Biotin	D-Biotin	Biotin and biotin bound to lysine and proteins	Widely distributed especially liver, kidney, yeast and eggs
C	Ascorbic acid, ascorbyl palmitate, sodium ascorbate	Ascorbic acid, dehydroascorbic acid and isoascorbic acid	Citrus fruits and vegetables

Information from Francis FJ (ed.) (2000) Vitamins: survey. In: *Encyclopedia of Food Science and Technology*, 2nd edn. vol. 4, pp. 2440–2449. New York: John Wiley also Food Standards Agency (2002) McCance & Widdowson's *The Composition of Foods*, sixth summary edition. Cambridge: Royal Society of Chemistry.

Table 3 Recommended dietary allowances (RDA) and reference nutrient intakes (RNI) for vitamins in adult males and females[a]

Category	*Age (years)*	*A (μg of RE per day)*[b]	*D (μg of D_3 per day)*	*E (mg of α-TE per day)*[c]	*K (μg per day)*	*B_1 (mg per day)*	*B_2 (mg per day)*	*Niacin (mg per day)*[d]	*B_6 (mg per day)*	*B_{12} (μg per day)*	*Folate (μg per day)*[e]	*C (mg per day)*
RDA												
Males	25–50	1000	5	10	80	1.5	1.7	19	2.0	2.0	200 (400)	60
Females	25–50	800	5	8	65	1.1	1.3	15	1.6	2.0	180 (400)	60
RNI												
Males	19–50	700				1.0	1.3	17	1.4	1.5	200	40
Females	19–50	600				0.8	1.1	13	1.2	1.5	200	40

[a]Taken from Food and Nutrition Board, National Research Council (1989) *Recommended Dietary Allowances* (10th edition), Washington, DC: National Academy Press; Report of the Panel on Dietary Reference Values of the Committee on Medical Aspects of Food Policy (1991) *Dietary Reference Values for Food Energy and Nutrients for the UK*. London: HMSO.

[b]RE, retinol equivalents.

[c]TE, α-tocopherol equivalents.

[d]Nicotinic acid equivalents.

[e]Dietary folate equivalent (DFE) for adults in the USA, rising to 600 μg per day during pregnancy and 500 μg per day during lactation (*See* **Folic Acid**: Physiology). Taken from Food and Nutrition Board, Institute of Medicine. National Academy of Sciences, Subcommittee on Folate, other B-Vitamins and Choline (1998) *Dietary Reference Intakes: Thiamin, Riboflavin, Niacin, Vitamin B-6, Folate, Vitamin B-12, Pantothenic Acid, Biotin and Choline*. Washington, DC: National Academy Press.

from animal studies, and are generally revised every 5–10 years by panels of expert scientists. The recommended intakes can vary from country to country owing to a number of factors such as different interpretations of investigations into the determination of requirements, the level of the safety margin needed, and the amounts needed to allow for individual variation in the general, healthy population. The recommended dietary allowances (RDA) and recommended nutrient intakes (RNI) used in the USA and the UK, respectively, are given in Table 3 for adult males and females. (*See* **Coronary Heart Disease**: Intervention Studies; **Dietary Reference Values**; **Dietary Requirements of Adults**; **Dietary Surveys**: Surveys of National Food Intake.)

It is also becoming evident that some vitamins have additional health benefits when given, via the diet or by supplementation, at levels above those required for eliminating classical deficiency states. Although, for most vitamins, the optimal levels are not yet established, for some vitamins such as folic acid, the health benefits have been clearly demonstrated for taking higher intakes to protect against spina bifida or other neural tube defects. (*See* **Cobalamins**: Properties and Determination; Physiology **Folic Acid**: Properties and Determination; Physiology **Food Fortification**.)

Availability

The way in which vitamins are absorbed and utilized in the body can be affected by a number of factors that can be broadly divided into dietary-related, and physiological factors related to the host. The vitamin may not be in a form that can be readily absorbed by the body. For example, nicotinic acid occurs in cereals in a form that is not absorbed from the gut. Other components in the diet may also interact with the vitamin, causing an enhancement, or depression, of absorption. Thus, the nutritive value of a diet for a given vitamin may be somewhat different from the amount actually analyzed chemically. Thiamin requirements are higher when the diet contains large amounts of carbohydrate or alcohol. Similarly, when the intake of polyunsaturated fat is high, vitamin E requirements are also raised. (*See* **Bioavailability of Nutrients**.)

In general, for food tables, allowances are normally made for reduced biological activities of different forms of three vitamins: 13-*cis*-retinol and retinaldehyde (vitamin A), carotenes other than β-carotene (provitamin A compounds), and tocopherols and tocotrienols other than α-tocopherol (vitamin E).

Adverse Effects

Along with increasing evidence of health benefits from consumption of diets containing vitamins at levels much higher than the RDAs, there is concern over potential adverse effects and even toxicity. Vitamins can be generally divided into two groups: first, those vitamins with a safety level of 100 times or more the RDA for that vitamin for which there is no clear indication of any adverse effects, including most of the water-soluble vitamins and the tocopherols; Second, those vitamins with a safety level of about 10 times the RDA, above which serious adverse reactions can occur. These vitamins include retinol, calciferols, phylloquinone, and pyridoxine.

In particular, chronic intakes of vitamins A and D are to be avoided. For example, an excessive intake of vitamin D can lead to vitamin D intoxication, which is characterized by an increase in serum 25-OH-D and is associated with hypercalciuria and hypercaemia. The safe upper limit for vitamin D is 25–50 μg per day (1000–2000 IU per day). Pregnant women should avoid large doses of vitamin A, either by avoiding foods, such as liver, that contain high concentrations of vitamin A, or by avoiding multivitamin supplements that contain doses of vitamin A of > 5000 IU.

As for the water-soluble vitamins, there has been some concern about excessive intakes for nicotinic acid and pyridoxine. Vitamin B_6 intakes as low as 50 mg per day can be associated with possible neurological damage in patients. Similarly, the potential benefits of lipid-lowering effects of high nicotinic acid intakes (2–6 g per day) have to be considered alongside the possible toxic effects, particularly for the liver. One of the main concerns of folic acid fortification is its potential to mask the diagnosis of pernicious anemia, which is caused by vitamin B_{12} malabsorption, at intakes of ≥ 1 mg per day, causing progressive and irreversible nerve damage. (*See* **Bioavailability of Nutrients**; **Functional Foods**.) Refer to individual vitamins.

Effects of Processing on Vitamin Stability

Vitamin stability varies greatly and depends on a number of factors such as temperature, oxygen, acid or alkali strength, and so on. Table 4 provides general information on stability (S) or instability (U) under several conditions. Several processes have been developed to produce stabilized forms of the vitamin, including spray drying in a suitable matrix (e.g., gelatin) and encapsulation.

Consumer interest over recent years in the nutritional labeling of foods, and the increased awareness in health benefits of vitamins, has focused attention on the effects of food processing on vitamin stability. It should be remembered also that one of the major benefits of processing is that it has a marked effect on preserving vitamin concentrations by inactivating

food enzymes despite any immediate loss. (*See* **Food Composition Tables**.)

The content of vitamins in processed foods is determined initially by the content of the raw materials used. There is a wide variability in vitamin concentrations of commercially grown foods. For example, carrots may vary in their carotene concentration 100-fold, and vitamin C can vary by as much as 35-fold. This can be due to a number of factors, including variety differences, climate and growing conditions, maturity at harvest, and postharvest storage conditions. Thus, the vitamin content of raw materials for processing can vary enormously, in many cases exceeding that found in normal processed food.

The different chemical and physical properties of the vitamins mean that they vary widely in their degree of stability. In the pure crystalline form, they retain their activity for long periods of time. However, when foods are processed, vitamins are subjected to a range of conditions that can be detrimental to their stability such as light, moisture, oxygen, heat and pH, these being typical environmental factors to which the food will be subjected in the food chain (see Table 4). The most labile vitamins are retinol, ascorbate, folate, and riboflavin. Niacin and biotin are the most stable, but information on some of the vitamins, especially vitamin D, is incomplete.

Domestic Cooking

Typical percentage losses of vitamins on cooking for selected food groups are given in Table 5.

Table 4 Vitamin stability under different conditions[a]

Vitamin	*Oxygen*	*Light*	*Heat*	*Acid (pH < 7)*[b]	*Alkali (pH > 7)*[b]	*Metals*[c]	*Reducing agents*
A	U	U	U	U	S		
D	U	U	U	U	U		
E	U	U	S	S	S		
K	S	U	S	S	U		
B_1	U	S	U	S	U	U	U
B_2	S	U	U	S	U	U	U
Niacin	S	S	S	S	S	U	S
B_6	S	U	U	S	S	S	U
B_{12}	U	U	U	S	S		
Folic acid	U	U	U	U	S	S	S
Pantothenic acid	S	S	U	U	U		
Biotin	S	S	U	S	S	S	
C	U	U	U	S	U	U	S

[a]Adapted from Chichester CO (1973). Nutrition in food processing. In: Rechcigi M (ed.) *Food Nutrition and Health. World Review of Nutrition and Dietetics*, vol. 16, pp. 319–326. Washington, DC: Karger-Basel.
[b]In the absence of oxygen.
[c]Information from van der Meer MA (1972) *Voeding* 33: 277. S, stable; U, unstable.

Table 5 Typical percentage losses of vitamins on cooking for selected food groups[a]

Vitamin	*Cereals*		*Milk*		*Meats*	*Fish*	*Vegetables*	*Fruit*
	Boiling	*Baking*	*Boiling*[b]	*Baked dishes*	*Meat, grilled or fried*	*Baking*	*Boiling*	*Stewing*
A					0	0	0	0
D						0		
E			20		20	0	0	
B_1	40	25[c]	10	25	20	30	35	25
B_2	40	15	10	15	20	20	20	25
Niacin	40	5	0	5	20	20	30	25
B_6	40	25	10	25	20	10	40	20
B_{12}			5	5	20	10		
Folate	50	50	20	50		20	40	80
Pantothenate	40	25	10	25	20	20		25
Biotin	40	0			10	10		25
C			50	50			45	25

[a]Adapted from Food Standards Agency (2002) McCance & Widdowson's *The Composition of Foods*, sixth summary edn. Cambridge: Royal Society of Chemistry.
[b]In milk-based drinks and custards.
[c]15% in bread-making and toasting.

Vitamin C is one of the most labile of all the vitamins and is readily leached into water on boiling, particularly when the surface of the food has been cut or damaged. Vegetables are an important source of vitamins in the diet, and typical losses on boiling vegetables are quite variable. The use of microwaving can reduce vitamin losses for cooked vegetables (see below). (*See* **Cooking**: Domestic Techniques.)

Meat is also an important source of several vitamins. A considerable amount of B-group vitamins, especially thiamin, can be lost when meat is grilled or fried (up to 40%). Losses are less when pot roasting and braising are used. Some vitamins that leach into the meat juices during cooking will not be lost if the sauce or the gravy is eaten as part of the actual dish. This means that average losses in meat dishes are no higher than for grilled or fried meat, even though the cooking times are longer. (*See* **Meat**: Dietary Importance.) Similarly, when bread is baked, 30% of the thiamin content can be destroyed, and this can increase to as much as 80% if alkaline baking powders are employed.

Microwave cooking can result in retention of vitamins equal to, or greater than, conventionally cooked vegetables. However, a number of studies have shown that cooking time, water volume used, etc. affect considerably vitamin retention with microwave cooking, as with other cooking procedures. For meat and meat products, the effect of microwave versus conventional cooking on vitamin retention seems to be less pronounced. A higher degree of riboflavin retention and a lower degree of thiamin retention have been found, but no difference in niacin concentrations. (*See* **Cooking**: Domestic Use of Microwave Ovens.)

Other Processing Methods

There has been only limited work reported on the effects of extrusion cooking on vitamin retention. In one of the few studies, the retention of B-group vitamins was investigated during crispbread production. The destruction of the more labile vitamins (thiamin, folate, B_6, and B_{12}) was directly proportional to the energy input, although other factors, such as moisture content, mass temperature, screw geometry, and rate of throughput, are also thought to influence retention. It appears that retention can be improved by increasing both the rate of throughout and moisture content. However, carotenoids in corn starch are fairly stable to extrusion, although some of the forms can be oxidized further. Losses of vitamin C during extrusion of 25–40% have been found. (*See* **Extrusion Cooking**: Chemical and Nutritional Changes.)

Radiation-induced changes in vitamin concentrations in foods during irradiation have been reviewed extensively in the literature. At low irradiation doses (< 1 kGy) used for disinfectant treatment, losses are negligible compared with the larger variations in the vitamin content of the raw materials. The most thermolabile vitamins are thiamin, tocopherols, and ascorbate, but, with the medium-dose range (1–10 kGy) used to control microbial contamination, losses can be minimized. (*See* **Irradiation of Foods**: Applications.)

It is well known that fermentation can increase the B-vitamin content (e.g., folates) of some foods, such as cheese, milks, yogurts, and other dairy foods, and this is a potential area for production of foods with enhanced health benefits. There does appear to be some potential for producing significant amounts of vitamins with various fermentation conditions and organisms.

See also: **Antioxidants**: Natural Antioxidants; Synthetic Antioxidants; **Bioavailability of Nutrients**; **Cobalamins**: Properties and Determination; Physiology; **Coenzymes**; **Dietary Reference Values**; **Dietary Requirements of Adults**; **Dietary Surveys**: Surveys of National Food Intake; **Folic Acid**: Properties and Determination; Physiology; **Food Fortification**; **Hypovitaminosis A**; **Pregnancy**: Maternal Diet, Vitamins, and Neural Tube Defects; **Vitamin B_6**: Properties and Determination

Further Reading

Ball GFM (1998) *Bioavailability and Analysis of Vitamins in Foods*. London: Chapman & Hall.

Bates CJ and Heseker H (1994) Human bioavailability of vitamins. *Nutrition Research Reviews* 7: 93–127.

Chichester CO (1973) Nutrition in food processing. In: Rechcigi M (ed.) *Food Nutrition and Health. World Review of Nutrition and Dietetics*, vol. 16, pp. 319–326. Washington, DC: Karger-Basel.

Francis FJ (ed.) (2000) Vitamins: survey. In: *Encyclopedia of Food Science and Technology*, 2nd edn. vol. 4, pp. 2440–2449. New York: John Wiley.

McCance RA and Widdowson EM (2001) *The Composition of Foods*, 6th summary edn. London: HMSO.

Taylor TG and Jenkins NK (eds) (1986) Food processing. In: *Proceedings of the XII International Congress of Nutrition*, pp. 786–803. London: John Libbey.

Van den Berg H, Heseker H, Lamand M, Sandström B and Thurnham D (1993) Flair Concerted Action No 10 Vitamin Status Papers. *International Journal for Vitamin and Nutrition Research* 63: 247–316.

Walter P, Stähelin H and Brubacher G (eds) (1989) *Elevated Dosages of Vitamins*. Toronto: Hans Huber.

Ziegler EE and Filer LJ (eds) (1996) *Present Knowledge in Nutrition*, 7th edn. Washington, DC: ILSI Press.

Determination

R Eitenmiller and L Ye, University of Georgia, Athens, GA, USA

Vitamins and Food Composition

From a metabolic and human health aspect, understanding of the roles that vitamins play is rapidly expanding. Concurrently, new and more precise information on food composition has been required by the scientific community. One of the more challenging areas of food composition research is the development and application of methods suitable to accurately assay water- and fat-soluble vitamins in foods. We have seen this area of bioanalytical chemistry expand rapidly over the past three decades to the point that, for most of the vitamins, we can confidently provide accurate measurements from most matrices. In the following discussion, we present a brief review of useful approaches for the analysis of vitamins in food. Also included are tables giving pertinent details of methods useful for the analysis of the specific vitamins. These approaches from the current literature show the state-of-the-art methods developed by leading vitamin researchers. Because of the significance of microbiological and chromatographic techniques to the area, our discussion is limited to these analytical approaches.

Analytical Methods

Methods available for vitamin assay can be classified as bioassays using animals or humans, microbiological assays that are based upon specific vitamin requirements of bacteria, yeast, molds, or protozoans, and physicochemical methods that are based on instrumental methods of analysis, including, but not limited to, spectrophotometric, fluorometric radiometric, immunological, and chromatographic techniques. Most recent advances in vitamin analysis have come through development and application of high-performance liquid chromatographic (LC) techniques specifically designed for food analysis. Factors influencing the suitability of a method for analysis of a vitamin(s) in food include the specificity of the assay, sensitivity of the method and levels of the vitamin in the food matrix to be assayed, the method's proven value as documented by validation studies including precision and accuracy, the applicability of the method to the food matrix, numbers of samples to be assayed in relation to manpower and sample throughput needs, availability of required instrumentation, and cost requirements. While excellent, sophisticated methods are available for most vitamins, these techniques are often not available for limited analytical efforts because of initial set-up economic considerations. In such instances, lower-cost, older methods of analysis must be utilized. This should not be considered as a limiting factor of the quality of the data because, if properly conducted, older methods based on wet chemistry or microbiological assay can provide data of excellent quality.

Because of the instability of many vitamins to oxidation, light inactivation, pH, heat, and various other chemical and environmental factors, precautions must be taken throughout the analysis to ensure stability of the analyte. For this reason, it is important that the analyst be thoroughly informed about the chemical properties of the specific vitamin being assayed (*See* **Vitamins**: Overview). Additionally, because the stability of the vitamin must be considered throughout the sample-handling period (collection through storage through analysis), poor-quality data will be obtained if degradative processes occur through improper maintenance of conditions necessary to ensure vitamin stability. This outcome negates the worth of even the most advanced analysis technique.

Extraction Methods

Except for bioassays that require the feeding of intact food samples, the vitamin(s) must be extracted from the food matrix in a medium designed to stabilize the vitamin prior to the determinative analytical step. Because of the diverse stability characteristics of the vitamins, specific extraction procedures have been developed over the years that accomplish efficient extraction of the various vitamins. Extraction media must be compatible with the determinative method and at the same time stabilize the vitamin against degradation. Table 1 outlines some common approaches used to extract the various vitamins along with the primary routes that might lead to degradation. Most extraction procedures use a combination of physical and chemical effects to ensure analyte stability. The overall extraction includes physical treatment combined with use of the extraction solvent to remove the vitamin with a high rate of recovery from the sample matrix, provide stability to the extracted vitamin, and provide the chemical environment required for compatibility with the determinative procedure. Physical disruption to reduce the particle size of the food sample is often accomplished through grinding, sonication, or other means. Following the initial sample treatment, solubilization of the vitamin into the extraction solution is completed by heat, acid, alkali, or enzyme digestion, often applied in combination. Extraction procedures

Table 1 Extraction procedures applicable to vitamin analysis of food

	Method	*Example*	*Primary routes of degradation*
Fat-soluble vitamins			
Vitamin A, vitamin A esters, provitamin A, carotenoids, vitamin E, vitamin D	Direct solvent extraction	Hexane or ether containing an antioxidant. Maintain darkness.	Oxidation Light Heat Acid
Vitamin A, vitamin A esters, provitamin A, carotenoids, vitamin E, vitamin D	Saponification	Reflux at 70 °C in ethanolic KOH containing antioxidant. Extract digest with hexane or combination of organic solvents containing an antioxidant.	Oxidation Light Heat Acid
Vitamin K	Hydrolysis and solvent extraction	Lipase hydrolysis. Pentane extraction. Maintain pH control. Solid-phase clean-up.	Unstable at alkaline pH
Water-soluble vitamins			
Ascorbic acid	Metaphosphoric acid extraction	Solvent usually contains reducing agent and metal chelating agent	Oxidation Temperature Light Ascorbic acid Oxidase
Thiamin	Acid hydrolysis followed by enzyme hydrolysis of phosphate esters	0.1 N HCl, autoclave, phosphatase digestion	Heat Unstable at alkaline pH
Riboflavin	Acid hydrolysis	0.1 N HCl, autoclave protect from light	Light Heat
Niacin	Acid or alkaline hydrolysis	HCl or H_2SO_4 digestion cereals – alkaline digestion	Few stability problems
Vitamin B_6	Acid extraction or enzymatic hydrolysis of phosphate esters following extraction in buffered solvent	Plant products – 0.44 N HCl, autoclave. Animal products – 0.055 N HCl, autoclave.	pH, stable in acid solutions, light
Folate	Multienzyme digestion in the presence of a reducing agent	Trienzyme digestion with Pronase® α-amylase and conjugase	Oxidation Light Heat Avoid pH below 5.0
Vitamin B_{12}	Phosphate buffer containing reducing agent	Autoclave in phosphate buffer containing citric acid and sodium metabisulfate	Light Optimum stability between pH 4.0 and 7.0 Oxidation
Biotin	Acid hydrolysis	Autoclave in 2 N H_2SO_4 (plant products) or 6 N H_2SO_4 (animal products)	Oxidation Stable between pH 4.0 and 9.0
Pantothenic acid	Multienzyme digestion	Digestion with alkaline phosphatase and pantotheinase	Stable between pH 5.0 and 7.0

are available for single vitamins and for multianalyte assays in which more than one vitamin is assayed from the same extract.

Water-soluble Vitamins

Microbiological Analysis

The microbiological assay of water-soluble vitamins represents one of the earliest approaches for accurate analysis of water-soluble vitamins in foods. Except for vitamin C, microbiological assays are available for each of the water-soluble vitamins. These methods were originally developed in the 1940s concurrently with, or shortly after, the development of bioassays for the different vitamins. The assays have withstood the test of time, often being applicable to the analysis of low-concentration samples that are problematic for other methods. Microbiological assays are highly specific and sensitive, although less technologically advanced, compared with many physicochemical procedures.

All microbiological vitamin assays are based upon the development of a dose–response curve resulting

from the growth of a microorganism in a medium containing the vitamin that is essential for the specific microorganism. The microorganism must have an absolute nutritional requirement for the vitamin that is not modified by other components in the food extract. If this requirement is met by the microorganism, a basal medium can be formulated that provides all of the growth requirements for the microorganism except for the vitamin being assayed. When a standard or an extract of the material to be assayed is added to the clear growth medium, the microorganism grows in proportion to the amount of added vitamin. Over a defined vitamin concentration range, the growth response will be directly proportional to the amount of vitamin in the growth medium. Within this range, sample and standard solutions can be accurately compared. The dose–response curve is developed by adding known amounts of standard to the growth medium. Unknown levels in the sample are calculated from the dose–response curve using the growth response resulting from the vitamin added to the growth medium through a known amount of sample extract. Growth is measured by turbidity, by production of a metabolite such as lactic acid, by respiration, or gravimetrically.

Characteristics that microorganisms must possess to be useful for water-soluble vitamin assay include: (1) a specific requirement for vitamin forms that are biologically active in higher animals; (2) inability to synthesize the vitamin; (3) inability to utilize metabolites that are not biologically active for the human; (4) inability to utilize compounds that are structurally related to the vitamin that are not biologically active; (5) a rapid growth cycle; (6) a growth response that is not easily modified by other substances present in the food extract (stimulators or inhibitors); (7) genetic stability and (8) nonpathogenicity. A list of accepted microorganisms available from the American Type Culture Collection (ATCC), Rockville, MD is provided in Table 2. Table 2 also lists the Association of Official Analytical Chemists (AOAC) International methods for each water-soluble vitamin assay.

Microbiological vitamin analysis remains an important tool for laboratories involved in food-composition studies. Advantages of the analytical approach include: (1) high sensitivity; (2) capacity to assay naturally occurring levels from complex food samples without concentration of the extract; and (3) low cost due to minimal laboratory requirements for space, materials, and instrumentation. Disadvantages include (1) labor intensiveness; (2) poor precision; (3) difficulty to automate, which leads to low sample throughput; (4) varying growth responses of the microorganism to different vitamin forms, including bound forms and multichemical forms (e.g., folates); and (5) susceptibility of the assay microorganism to stimulators or inhibitors in the food extract that produce drift. In relation to human nutrition, the assays do not provide information on the bioavailability of the vitamin from the food sample.

Table 2 Microbiological assay of water-soluble vitamins

Analyte	*Assay organism*	*AOAC International method*[a]
Thiamin	*Lactobacillus viridescens* ATCC 12706	
Riboflavin	*Lactobacillus rhamnosus (casei)* ATCC 7469	960.46 Chapter 45.2.06
Niacin	*Lactobacillus plantarum* ATCC 8014	944.13 Chapter 45.2.04 985.34 Chapter 50.1.19
Vitamin B_6	*Saccharomyces uvarum* ATCC 9080	961.15 Chapter 45.2.08 985.32 Chapter 50.1.18
Folate	*Enterococcus hirae* ATCC 8043 *Lactobacillus rhamnosus (casei)* ATCC 7469	944.12 Chapter 45.2.03 992.05 Chapter 50.1.21
Vitamin B_{12}	*Lactobacillus delbrueckii* ATCC 7830	952.20 Chapter 45.2.02 986.23 Chapter 50.1.20
Biotin	*Lactobacillus plantarum* ATCC 8014	
Pantothenic acid	*Lactobacillus plantarum* ATCC 8014	945.74 Chapter 45.2.05 992.07 Chapter 50.1.22

[a]From AOAC International (2000) *Official Methods of Analysis* of AOAC International, 17th edn. Arlington, VA: Association of Official Analytical Chemists.

Chromatography for the Assay of Water-soluble Vitamins

High-performance LC methods are available for analysis of water-soluble vitamins in foods. These methods offer speed, selectivity, accuracy, precision, and, depending upon the chemical properties of the specific vitamin, a high degree of sensitivity. Reversed-phase chromatography systems using C-18 or other bonded supports with varying adsorption properties are widely applied to specific analytical methods for the analysis of water-soluble vitamins. Specific applications depending on the chemical properties of the analytes include ion exchange for vitamin C, folates, and niacin, PLRP-S (polymeric styrene reversed-phase macroporous supports) for riboflavin and vitamin C, and C-18 for most water-soluble vitamins. Useful mobile phases for water-soluble vitamin analysis on reversed-phase systems are usually based upon

methanol or acetonitrile, together with water, acetic acid or buffers. Ion-pair chromatography, which combines the principles of bonded-phase and ion-exchange chromatography is often used to improve analyte resolution, retention time, and peak shape.

Often, the only limitation to application of LC to water-soluble vitamins in foods is the low natural levels that must be quantified. Detection systems useful for LC analysis of the water-soluble vitamins include UV, fluorescence, and electrochemical; however, the choice of the detection mode depends entirely on the properties of the vitamin being assayed. Low natural levels in foods are often below the detection capabilities of the LC system. In this case, sample extract concentration using solid-phase extraction (SPE), affinity chromatography, or other means such as freeze-drying might be applied to bring the analyte into the required concentration range. Often, with low-concentration samples, microbiological assay offers a workable solution to the analytical problem.

Pre- and postcolumn derivatization can be used for vitamins such as thiamin, folate, and others to induce or increase the fluorescence of the analyte to increase the sensitivity of the analysis. Fluorescence, because of its sensitivity and selectivity, is a preferred detection mode for LC procedures and should be incorporated into an analysis protocol if chemically feasible. Multianalyte procedures useful for food analysis are mostly limited to fortified foods or supplements in which the analytes are present in high concentrations, as opposed to low concentration in natural foods. Some successful simultaneous assays are available for thiamin and riboflavin, which possess excellent spectral properties that make sensitivity less problematic.

Because of the breadth and complexity of the applications of LC to water-soluble vitamin analysis, examples of literature methods that have been proven useful for food composition research are characterized in Table 3. The reader is urged to access the Further Reading list for detailed information.

Fat-soluble Vitamins

Chromatography presents versatile and highly useful methods for the analysis of fat-soluble vitamins from foods and other biological samples. The analytical field has progressed from applications of paper chromatography, thin-layer chromatography, and gas chromatography to the almost universal use of LC for fat-soluble vitamin analysis. Both normal-phase and reversed-phase systems are applicable to fat-soluble vitamin analysis.

Efficient extraction of the fat-soluble vitamins from the sample matrix is as important as the determinative chromatography. With the fat-soluble vitamins, many LC-based procedures require that the vitamin be isolated from the lipid material prior to injection of the vitamin extract on to the LC column. Applicable extraction methods rely on alkaline hydrolysis or saponification, enzymatic hydrolysis, solvent extraction, and supercritical fluid extraction. Saponification is the most accepted method for extraction of vitamin A, carotenoids, vitamin D, and vitamin E from foods. Saponification involves refluxing the sample with ethanolic KOH solution in the presence of an antioxidant (pyrogallol or vitamin C) at 70 °C. The unsaponifiable fraction in the aqueous digest containing the fat-soluble vitamins, sterols, carotenoids, and other material is extracted with hexane, ethyl ether, or a mixture of hexane and ethyl acetate, leaving the fatty acid salts and other water-soluble components in the aqueous phase. Saponification has the ability to efficiently destroy the sample matrix, which facilitates vitamin extraction, frees cellular bound forms of the vitamins, and destroys substances such as chlorophyls that might interfere with the chromatography. Because of the instability of vitamin K under alkaline conditions, saponification cannot be used for extraction.

If the LC method is a normal-phase system, limited amounts of lipid can be injected directly on to the column. In this case, solvent extraction of the sample can be used with direct injection of the extract containing the fat-soluble vitamins on to the normal-phase column. Direct solvent extraction has been used to develop multianalyte procedures capable of assaying several analytes including ester forms of vitamin A (retinyl palmitate and acetate) and vitamin E (α-tocopheryl acetate).

Vitamin A and Carotenoids

Most recently published methods for retinoids and carotenoids use reversed-phase LC on C-18. The advantages of reversed-phase compared with normal-phase LC include: (1) less sensitivity to changes in retention time due to the presence of water; (2) more easily cleared of contaminants; (3) more stable to small changes in mobile phase composition; (4) more quickly equilibrated to mobile phase composition changes, permitting use of gradients; and (5) capable of resolving compounds with a wide range of polarities. Both isocratic and gradient mobile phases are useful (see Table 4). For vitamin A (retinol) assay, reversed-phase chromatography with simple, isocratic mobile phases consisting of methanol:water, acetonitrile:water, or gradients based on these solvents can be used. Resolution of retinol *cis*- and *trans*-isomers is best accomplished with normal-phase LC.

The conjugated double bond system of vitamin A and its closely related retinoids gives specific and

Table 3 Selected HPLC methods for the analysis of water-soluble vitamins in various foods

Analyte	*Sample type*	*Sample preparation*	*HPLC parameters*				*References*
			Column	*Mobile phase*	*Detection*	*Quality-assurance parameters*	
Ascorbic acid							
AA	Food	Dilute samples to an estimated AA concentration of 100–800 ng ml^{-1} with mobile phase. Filter.	Inersil ODS-3 5 μm 150 × 4.6 mm	100 mM KH_2PO_4 (pH 3) containing 1 mM EDTA-2Na:$2H_2O$. 0.6 ml min^{-1}	EC 400 mV Ag/AgCl	DL (ng) On-column 0.5 % Recovery > 90	*Journal of Chromatography A* 806 (1998) 361–364
AA, DHAA, IAA, DHIAA	Various foods	Extract sample with 1% MPA, 0.5% oxalic acid (pH 2). For samples high in starch, extract with 2% MPA, 1% oxalic acid–EtOH (50:50). Centrifuge Filter. Inject.	Jupiter C_{18} 5 μm 250 × 4.6 mm	2.3 mM DTMAC:2.5 mM Na_2EDTA in a 66 mM phosphate–20 mM acetate butter pH 4.5, 1.2 ml min^{-1}	Postcolumn derivatization of dehydroforms with OPD fluorescence λ_{ex} 350 mm λ_{em} 430 nm	RSDr (%) 1.1–1.6 RSD_R (%) 1.3–3.5	*Journal of Chromatography B* 730 (1999) 101–111
AA	Various foods	Dilute samples to an estimated AA concentration of 1–10 μg ml^{-1} with mobile phase. Filter.	Inersil ODS-3 5 μm 150 × 4.6 mm	20 mM MSG (pH 2.1) or GMP (pH 2.1) 0.8 ml min^{-1}	EC 400 mV Ag/AgCl	% Recovery > 90	*Journal of Chromatography A* 881 (2000) 317–326, 327–330
Thiamin							
Thiamin, TMP, TPP, TTP	Various foods	Add amprolium (IS), homogenize with 5% sulfosalicyclic acid. Centrifuge. Filter water layer.	Perkin Elmer C_{18} 3 μm 300 × 3 mm	Gradient 0.1 M Na_3PO_4, pH 5.5, 6 min 0.1 M Na_3PO_4, pH 2.6, 19 min	Thiochrome postcolumn Fluorescence λ_{ex} 339 nm λ_{em} 432 nm		*Journal of Micronutrient Analysis* 2 (1986) 189–199
Thiamin	Various foods	Add HCl, digest, 100 °C, 30 min. Adjust pH to 4–4.5. Digest with takadiastase, 47 °C, 3 h.	Lichrosphere 100 RP− 18 5 μm 125 × 4 mm	Isocratic H_3PO_4/KH_2PO_4, 10^{-2} M, pH 2.8 : MeOH (85:15) with 5 μM Hex sulfonic acid and 0.1% triethylamine	254 nm		*Journal of Liquid Chromatography and Related Technology* 19 (1996) 2155–2164
Thiamin	Cooked sausages	Homogenize 10 g ground sample with 60 ml 0.1 N HCl. Autoclave, 20 min. Cool. Adjust pH to with 2.5 M NaOAC. Add 5 ml 6% claradiastase, incubate, 3 h. Add 2 ml 50% TCA, 90 °C, 15 min. Filter.	Spherisorb C_{18} 5 μm 250 × 4 mm	Phosphate buffer (5 mM, pH 7.0):MeCN (70:30) 35 °C 0.6 ml min^{-1}	Thiochrome precolumn derivatization Fluorescence λ_{ex} 360 nm λ_{em} 430 nm	% Recovery > 90	*Journal of Agricultural and Food Chemistry* 47 (1999) 170–173
Riboflavin							
Riboflavin FMN, FAD	Wine, beverages	Filter sample through Millex-GV (0.22 μm). Dilute the sample up to four times with water. Inject.	Hypersil ODS C_{18} 5 μm 200 × 2.1 mm	Gradient (A) 0.05 M NaH_2PO_4 (pH 3.0) (B) MeCN	Fluorescence λ_{ex} 265 nm λ_{em} 525 nm	QL (μg l^{-1}) Riboflavin–1.7 FAD – 6.6 FMN – 2.8	*Journal of Chromatography A* 823 (1998) 355–363

Riboflavin	Cooked sausages	Homogenize 10 g ground sample with 60 ml 0.1 N HCl. Autoclave, 20 min Cool. Adjust pH 4–4.5 with 2.5 M NaOAC. Add 6% claradiastase, incubate, 50 °C 3 h. Add 2 ml 50% TCA, 90 °C, 15 min. Filter	Spherisorb ODS-2 5 μm 250 × 4 mm	5 mM HEP sulfonic acid (pH 2.7):MeCN (75:25) 35 °C 0.6 ml min^{-1}	Fluorescence λ_{ex} 277 nm λ_{em} 520 nm	DL (mg/100g) 0.015 % Recovery > 95	*Journal of Agricultural and Food Chemistry* 47 (1999) 1067–1070
Riboflavin FMN, FAD	Various food	Add 19 ml MeOH–CH_2Cl_2 (9:10) to 12 g sample and shake, 1 min. Add 9 ml 0.1 M NH_4OAC (pH 6.0) and shake, 1 min. Centrifuge at 4 °C, 20 min. Filter.	Alphabond C_{18} 10 μm 300 × 4.6 mm	MeOH: 0.05 M NH_4OAC (pH 6.0) (30:70) gradient 1 ml min^{-1}	Fluorescence λ_{ex} 450 nm λ_{em} 530 nm	DL (ng ml^{-1}) Riboflavin, FMN – 1 FAD – 6 % Recovery > 95	*Journal of Chromatography A* 881 (2000) 285–297
Niacin							
Niacin	Infant formula wheat flour	Weigh sample containing 100–200 μg niacin. Add 8 mL DMW, 2 mL (1+1) H_2SO_4, mix. Autoclave, 45 min. Cool. Adjust pH to 6.5 with 7.5 N NaOH. Adjust pH ≤1 with H_2SO_4. Filter.	PRP-X100 Anion-exchange 250 × 4.1 mm	0.1 M NaOAC–HAC (pH 4–4.2) 1.5 ml min^{-1}	260 nm	DL (μg ml^{-1}) 0.2 % Recovery 95–107	*Journal of AOAC International* 82 (1999) 128–133
Niacin	Various foods	Add 30 ml 0.1 M HCl to 1–5 g ground sample, heat at 100 °C, 1 h. Cool. Dilute. Filter. Autoclave, 1 h. Cool. Adjust pH to 4.5, dilute. Filter.	Lichrospher 100 RP-18 5 μm 250 × 5 mm	0.07 M KH_2PO_4, 0.075 M hydrogen peroxide, 5.10 × 10^{-6} M copper (II) sulfate solution 1 ml min^{-1}	Postcolumn derivatization Fluorescence λ_{ex} 322 nm λ_{em} 380 nm	% Recovery 90–107 % CV < 4	*Food Chemistry* 65 (1999) 129–133
Nicotinic acid, nicotinamide	Cooked sausages	Add 30 ml water, mix and homogenize, 2 min. Dilute with water and centrifuge. Filter upper solution. Add 1 ml saturated zinc sulfate solution and 1 ml 1N NaOH. Dilute with water, let stand 30 min. Filter.	Spherisorb ODS-2 5 μm 250 × 4 mm	HEP sulfonic acid (5 mM, pH 3.3):MeCN (75:25) 35 °C 0.65 ml min^{-1}	261 nm	% CV 2.3–4.5 % Recovery 93–95	*Journal of Agricultural and Food Chemistry* 48 (2000) 3392–3395
Vitamin B_6							
PL, PM, PN, PLP, PMP, PNP, 4-PA	Egg, milk, baker's yeast extract, yeast cell-free culture media	Baker's yeast extract – 50 mg sample, add 50 μl 1 mM *iso*-PL, 3 ml of 1 M $ClHO_4$. Egg yolk – 2 g sample, add 100 μl 1 mM *iso*-PL, 6 ml 1 M $ClHO_4$; Milk – 2 g sample, add 10 μl 1 mM iso-PL, 1 ml 1 M $ClHO_4$; mix. Centrifuge. Decant supernatant, adjust pH to 3–4 with 10 M KOH. Refrigerate. Centrifuge. Filter	Phenospher ODS-2 5 μm 250 × 4.6 mm	0.15 M NaH_2PO_4 Adjust pH with 70% $ClHO_4$ to 2.5 1 ml min^{-1}	Postcolumn derivatization sodium bisulfite Fluorescence λ_{ex} 290 nm λ_{em} 389 nm	% Recovery 89–102	*Journal of Chromatography A* 790 (1997) 83–91
PMP, PM, PL, PLP PN	Cooked sausages	Extract with 5% MPA. Centrifuge Filter.	Hypersil BDS C_{18} 5 μm	Isocratic 0.05 M KH_2PO_4 (pH 3.2): MeCN (99:1) 0.8 ml min^{-1} 35 °C	Fluorescence λ_{ex} 290 nm λ_{em} 395 nm	% Recovery 92–100	*Journal of Agricultural and Food Chemistry* 49 (2001) 38–41

Continued

Table 3 Continued

Analyte	*Sample type*	*Sample preparation*	*HPLC parameters*				*References*
			Column	*Mobile phase*	*Detection*	*Quality-assurance parameters*	
Folate							
Folic acid, 5-CH_3-H_4 folate, 5-CHO-H_4 folate, 10-CHO-H_2 folate, 10-CHO-folate	Cereal foods	Add 10 volumes Hepes/Ches buffer, pH 7.85 (50 mM Hepes, 50 mM Ches) containing 2% AA and 10 mM 2-MCE, vortex, boiling water bath, 10 min. Cool. Homogenize, Digest rat plasma conjugase and α-amylase (37 °C, 4 h) followed by protease (37 °C, 1 h), boiling water bath, 5 min. Cool. Centrifuge. Resuspend residue in extraction buffer. Centrifuge. Filter (Whatman #1 flush N_2).	Ultramex C_{18} 5 μm 250 × 4.6 mm	Gradient (A) MeCN (B) 0.033 M H_3PO_4, pH 2.3 0–8 min Isocratic 5%A 8–33 min Linear gradient to 17.5% A 1 ml min^{-1}	PDA UV Monitor at 280 nm	% Recovery 87–107	*Journal of Agricultural and Food Chemistry* 45 (1997) 407–413
Biotin							
d-Biotin, *d*-biocytin	Various foods	Ground sample (5–10 g) for high-starch sample, add 100 mg takadiastase (60.5 U mg^{-1}), add reduced glutathione (300 μl), EDTA (300 μl), citrate buffer (30 mL) and papain (3 ml). Shake and incubate, 37 °C, overnight. Cool. Filter.	Lichrospher 100 RP 18 5 μm 250 × 5 mm	0.1 M phosphate buffer (pH 6):MeOH (81:19) 0.4 ml min^{-1}	Post column derivatization with avidin-fluorescein isothiocyanate Fluorescence λ_{ex} 490 nm λ_{em} 520 nm	% Recovery 90–104 for *d*-biocytin 92–106 for *d*-biotin	*Food Chemistry* 65 (1999) 253–258
Pantothenic acid							
Pantothenic acid	Milk, infant formulas	Add 5 ml warm water (< 40 °C) to 0.5 g sample, vortex. Stand, 20 min. Add 0.5 ml HAC (3%), vortex. Stand, 20 min. Centrifuge. Filter.	Nomura Develosil ODS-MG-5 5 μm 250 × 4.6 mm	0.1 M KH_2PO_4 (pH 2.25): MeCN (95:5) 1.0 ml min^{-1}	UV 200 nm 205 nm 240 nm	DL (mg per 100 g) 0.3 % Recovery 98–100	*Food Chemistry* 69 (2000) 201–208

Table 4 Selected HPLC methods for fat-soluble vitamins in foods

Analyte	*Sample type*	*Sample preparation and clean-up*	*HPLC parameters*				*References*
			Column	*Mobile phase*	*Detection*	*Quality-assurance parameters*	
Vitamin A							
Retinol	Milk powder, flour	Mix 1.25 g sample with 5 ml aqueous NaOH (50%), heat 30 °C, 3 min. Add 25 ml EtOH and 0.5 ml hydroquinone (20% in EtOH), heat, 80 °C, 30 min. Cool. Add water and extract with Et_2O-PE (1:1). Reextract. Filter. Evaporate and redissolve in MeOH.	LiChrosorb RP-18 5 μm 125 × 4.5 nm 30 °C	Isocratic MeCN:water (80:20) 0.8 ml min^{-1}	325 nm	%Recoveries Flour: 89–108 Milk powder: 91–101 QL < 32 μg per 100 g	*Chromatographia* 47 (1998) 716–720
Retinol	Human milk	Add 1 ml EtOH (containing 2% BHT) and 500 μl KOH (12.5 M) to 500 μl sample. Saponify, 80 °C, 25 min. Extract with 1 ml Hex:toluene (1:1). Centrifuge. Inject 200 μl of the organic phase	Grom-Sil-CN-2PR 5 μm 250 × 4.6 mm	Hex:IPA (98:2) 2 ml min^{-1}	Fluorescence λ_{ex} 325 nm λ_{em} 480 nm	RSD (%) Intra – 1.9 Inter – 2.3 % Recovery 102	*Journal of Chromatography A* 898 (2000) 179–183
Carotenoids							
α-, β-carotene; 13-*cis*-β-carotene; 9-*cis*-β-carotene	Green vegetables	Blend with water containing 0.5% AA. Extract with A:PE (3:2) containing 0.5% BHT. saponify extract, ambient, 15 min. Wash 3 × with 10% NaCl solution. Dry over Na_2SO_4. Evaporate. Dissolve in MeOH:CH_2Cl_2 (9:1).	Vydac TP-201 5 μm 250 × 4.6 mm	Isocratic prepared isomers – MeOH:CH_2Cl_2: water (79:15:6) sample–MeOH: CH_2Cl_2: water (80:12.5:4.8) 1 ml min^{-1}	450 nm	% Recovery 96.4–103.5	*Food Chemistry* 55 (1996) 63–72
All-*trans*-lutein, zeaxanthin, cryptoxanthin, lycopene, α-, β-carotene	Human milk	Add 1 ml EtOH and 1 ml Hex to 1 ml sample, vortex, centrifuge. Remove upper organic layer. Repeat one more time with 1 ml Hex. Extract Hex layer twice with 1 ml EiOH:water (9:1). Dry the remaining Hex. Combine the residue with the aqueous phase from first extraction step. Saponify, 37 °C, 2 h. Extract with 1 ml Hex 4×. Combine the ethanol and Hex extract, dry, redissolve in 0.2 ml IPA. Inject.	C_{30} column 5 μm 250 × 4.6 mm	Gradient Solvent A MeOH:water (90:10) with 0.04%NH_4 OAC Solvent B MeOH:MTBE:water (8:90:2) with 0.01% NH_4OAC 1 ml min^{-1}	450 nm	% Recovery > 90	*International Journal for Vitamin and Nutrition Research* 70 (2000) 79–83

Continued

Table 4 Continued

Analyte	*Sample type*	*Sample preparation and clean-up*	*HPLC parameters*				*References*
			Column	*Mobile phase*	*Detection*	*Quality-assurance parameters*	
15-*cis*-β-carotene; 13-*cis*-β-carotene; all-*trans*-α-, β-carotene; 9-*cis*-β-carotene	Carrot juices and vitamin supplemented drinks	Extract sample with A:Hex (1:1). Add NaCl solution (10%) to remove the emulsion. Wash Hex layer with water. Add BHT as an antioxidant (0.1%). Dry with Na_2SO_4. Evaporate. Redissolve in IPA. Inject	C_{30} RP column 5 μm 250 × 4.6 mm	Linear gradient from 100% A to 56% B within 50 min at flow rate 1 ml min^{-1} Eluent A MeOH:MTBE:water (81:15:4) Eluent B MTBE:MeOH: water (90:6:4)	9-*cis*-β-carotene, α-carotene – 445 nm β-carotene – 452 nm 13-*cis*-β-carotene – 471 nm β-apo-8′-carotenal (IS) – 462 nm	% Recovery > 90%	*Food Chemistry* 70 (2000) 403–408
Vitamin D							
D_2,D_3	Baby food	Add D_2 or D_3 (IS). Saponification, 70 °C, 30 min. Extract with Et_2O:PE (1:9). Evaporate. Redissolve in iOCT Clean-up. Semi-Prep LC Polygosil 60–5 μm, 250 × 8 mm.	Two Shandon ODS-5 columns in series 5 μm 250 × 4.6 mm	Isocratic MeOH (100%) 1 ml min^{-1}	265 nm	QL 350–420 IU kg^{-1}% Recovery 84 DL (ng) On-column = 2	*Netherlands Milk and Diary Journal* 48 (1995) 31–39
D_2,D_3	Infant formula, enteral nutritionals	Add D_2 (IS). Saponification, 60 °C, 30 min. Extract with Hex. Evaporate. Clean-up SPE B & J Silica 9054. Evaporate. Redissolve in MeCN.	Vydac 201 TP54 C_{18} 5 μm 250 × 4.6 mm (not endcapped)	Gradient MeCN:MeOH:EtOAC 0.7–2.5 ml min^{-1}	265 nm	QL 8 IU/quart % Recovery > 99 RSD_R = 13.48%	*Journal of AOAC International* 75 (1992) 566–571; 79 (1996) 73–80
Vitamin E							
α-, γ-, δ-T	Margarine, reduced fat products	Dissolve sample (5 g) with Hex containing 0.1% BHT, add $MgSO_4$ to remove water. Stand in dark ≥2 h. Filter.	LiChrosorb Si60 5 μm 250 × 4.6 mm	Isocratic Hex:IPA (99.1:0.9) 1.0 ml min^{-1}	Fluorescence λ_{ex} 290 nm λ_{em}330 nm	QL (μg per 100 g) α-T–39.8 γ-T – 5.0 δ-T – 3.0 % Recovery > 97	*Journal of Liquid Chromatography and Related Technology* 21 (1998) 1227–1238

α-T acetate retinyl palmitate	Soy-, milk-based infant formula, medical foods	Matrix solid-phase extraction, C_{18}	LiChrosorb Si60 5 μm 250 × 4.6 mm	Isocratic IPA:Hex (0.5:99.5) for α-T acetate, (0.125:99.895) for retinyl palmitate 1.0 ml min^{-1}	Fluorescence α-T acetate λ_{ex} 285 nm λ_{em} 310 nm Retinyl palmitate λ_{ex} 325 nm λ_{em} 470 nm	Recoveries > 91%	*Journal of AOAC International* 81 (1998) 582–586; 82 (1999) 107–111; *Journal of Liquid Chromatography and Related Technology* 21 (1998) 2853–2861
Vitamin K							
K_1	Vegetables	To 1–5 g sample, sonicate with 10 ml MeOH, 15 min. Centrifuge. 2 ml of the mix solution extract with 4 ml volume carbonate. Heat at 80 °C, 1 h. Partition with 4 ml Hex (3×). Evaporate Hex layer. Redissolve in MeOH.	LiChrosorb RP-8 10 μm 250 × 4.6 mm	Isocratic 100% MeOH 0.6 ml min^{-1}	247 nm and particle beam mass spectrometry	DL (ng) On-column UV – 0.1 % Recovery 82–90	*Fresenius Journal of Analytical Chemistry* 355 (1996) 48–56
K_1, MK-4, $K_1(H_2)$	Margarine, butter, vegetable oils	Add 20 ml of Hex to samples (0.2 – 0.5 g or 5.0 g for fat-free products). Homogenize. Filter. Evaporate. Redissolve in THF:MeOH (1:1).	YMC C_{30} 3 μm 250 × 4.6 mm followed by a zinc column (20 × 4 mm)	Isocratic 100 ml of CH_2Cl_2 + 900 ml of MeOH + 5 ml of methanolic solution containing 1.37 mg $ZnCl_2$, 0.41 mg NaOAC and 0.3 mg glacial HAC 0.8 ml min^{-1}	Fluorescence λ_{ex} 243 nm λ_{em} 430 nm	% Recovery 93–99	*Food Chemistry* 69 (1999) 79–88
K_1	Milk- and soy-based infant formula	Matrix solid-phase extraction, C_{18}	Alltech C_8 3 μm 150 × 4.6 mm followed by a zinc reduction column (25 × 3.2 mm)	MeOH:reductive ionic solution (RIS) (950:5) containing 10% Hex 1 ml min^{-1} RIS – 2 M $ZnCl_2$ + 1 M NaOAC + 1 M HAC per liter of methanol	Fluorescence λ_{ex} 248 nm λ_{em} 418 nm	DL (pg) K_1–12 QL (pg) K_1 – 38% Recovery 84–102	*Journal of AOAC International* 82 (1999) 1140–1145; *Journal of Liquid Chromatography and Related Technology* 23 (2000) 423–432
Multianalyte methods							
All-*trans* retinol, α-, β-, γ-, δ-T, β-carotene, 13-*cis*-retinol	Italian cheese	Saponification, 70 °C, 30 min. Extract with Hex:EtOAC (90:10). Evaporate. Redissolve in mobile phase.	Ultrasphere Si, 5 μm 250 × 4.6 mm	Isocratic – Hex: IPA (99:1) 1.5 ml min^{-1} gradient – multilinear pump A Hex: IPA (99:1), pump B Hex (100%), 1.5 ml min^{-1}	Tocopherols fluorescence λ_{ex} 280 nm λ_{em} 325 nm Retinol fluorescence λ_{ex} 325 nm λ_{em} 475 nm β-carotene 450 nm	DL (ng) On-column α-T – 0.9, β-T – 0.73 γ-T – 0.55 δ-T – 0.56 13-*cis*-retinol – 0.09 β-carotene – 0.16 all-*trans* retinol – 0.32	*Analyst* 119 (1994) 1161–1165

Continued

Table 4 Continued

Analyte	*Sample type*	*Sample preparation and clean-up*	*HPLC parameters*				*References*
			Column	*Mobile phase*	*Detection*	*Quality-assurance parameters*	
K_1, α-T, all-*trans*-retinol	Milk, milk powder	(a) α-T, retinol saponification, overnight, ambient. Extract with Hex. Evaporate. Redissolve in MeOH (b) K_1 lipase hydrolysis, add alcoholic sodium hydroxide and immediately extract with Hex. Clean-up Sep-Pak silica.	Brownlee OD-224 RP 18 5 μm 220 × 4.6 mm	Isocratic MeOH:water (99:1) containing 2.5 mM HAC:NaOAC, 1.25 ml min^{-1}	EC dual-EC glassy carbon -1100 mV +700 mV vs Ag/AgCl	DL (ng) On-column retinol – 0.06 K_1 – 3.1 α-T – 0.19	*Journal of Chromatography* 623 (1992) 69–74; *Analyst* 120 (1995) 2489–2492
α-T, retinol	Infant formula	Saponification, overnight, ambient. Extract (5×) with Hex. Wash extract with water and dry over Na_2SO_4. Filter. Evaporate, 40 °C. Redissolve in MeOH. Filter.	Spherisorb ODS-2 C_{18} 5 μm 250 × 4.6 mm	Isocratic water:MeCN:MeOH (4:1:95)	α-T – 292 nm retinol – 323 nm	% Recovery α-T – 86 retinol – 98	*Journal of Chromatography A* 778 (1997) 243–246
All-*trans* retinol, 13-*cis* retinol, α-, β-T, β-carotene, retinyl palmitate	Milk	Saponification, 70 °C, 7 min. Extract with Hex:DIPE (75:25) containing 0.04% BHT. Add 30 ml Milli-Q water, gently inverted the tube, centrifuge. Evaporate supernatant. Redissolve in 2 mL Hex containing 0.4% BHT β-T (IS).	Alltech Econosphere silica 3 μm 150 × 4.6 mm	Isocratic 0.01% acetic acid, 0.5% IPA, 0.02 mg l^{-1} α-T in Hex 1.5 ml min^{-1}	Vitamin A fluorescence λ_{ex} 330 nm λ_{em} 470 nm Vitamin E fluorescence λ_{ex} 295 nm λ_{em} 330 nm β-carotene 450 nm	% Recovery 98–100	*Milchwissenschaft* 53 (1998) 11–15
α-, γ-T, α-, β-carotene, lutein, zeaxanthin, β-cryptoxantnin	Corn	Add 6 ml EtOH with 0.1% BHT to 600 mg freeze-dried corn powder. Heat at 85 °C in water bath, 5 min. Add 120 μl 80% KOH, vortex and saponify at 85 °C, 10 min. Add 3 ml cold deionized water, cool. Extract with 3 ml Hex (3×). Wash Hex layer with 3 ml deionized water. Evaporate Hex layer. Redissolve in 200 μL CH_3CN:MeOH:CH_2Cl_2 (45:20:35).	Vydac 201TP54 C_{18} 5 μm 150 × 4.6 mm connected with Nova-Pak C_{18} 4 μm 150 × 3.9 mm	Isocratic CH_3CN:MeOH:CH_2Cl_2 (75:20:5) containing 0.05% TEA and 0.1% BHT 1.8 ml min^{-1}	Carotenoids – 450 nm Tocopherols – 290 nm	% Recovery 85–122	*Journal of Liquid Chromatography and Related Technology* 22 (1999) 2925–2934

α-, γ-T, α-, γ-, δ-T3, carotenoids	Red palm oil	Add 5 ml MeOH, 0.3 ml 60% KOH, 1 ml Methanolic pyrogallol (0.02%) to 400 mg sample. Heat at 100 °C in water bath, 30 min. Cool. Add 10 ml water. Extract with 10 ml ET_2O (3×). Wash ether layer with 10 Ml water (4×). Add anhydrous Na_2SO_4 and filter. Evaporate. Redissolve in HPLC mobile phase.	YMC C_{30} 5 μm 250 × 4.6 mm	(A) MeOH:MTBE:water (81:15:4) (B) MTBE:MeOH (91:9) 0–45 min – 100% A to 50% A 45–56 min – 100% B 56–60 min – 100% A 1 ml min^{-1}	Vitamin E – 295 nm Carotenoids – 450 nm	% CV 2–25 for carotenoids 0.6–20 for tocopherols	*Journal of Liquid Chromatography and Related Technology* 23 (2000) 1873–1885
Vitamin A, E, D_2, D_3, K_1, retinyl acetate, retinyl palmitate, α-T acetate, ergosterol, 7-dehydrocholesterol	Milk	Add 5 ml EtOH with 0.025% BHT to 5 ml sample, sonicate. Extract with 15 ml Hex (2×). Wash Hex layer with MeOH-water (9:1). Filter and evaporate Hex layer. Redissolve in 100 μl MeOH. Inject 60 nl.	Hypersil C_{18} BDS 3 μm 150 × 0.3 mm	(A) MeOH:water (99:1) (B) MeOH: THF (70:30) 0–4 min – 0% B 4–10 min – 100% B 10–15 min – 100% B 15–17 min – 0% B 6 μl min^{-1}	Wavelength (nm) 325–vitamin A 264–vitamin D 280–vitamin K_1, E, provitamin D	%Recovery 89–107 %CV 2–8	*Journal of Chromatography A* 891 (2000) 109–114
α-, β-, γ-, δ-T α-T acetate, β-carotene, retinyl palmitate	Fortified foods	To 3 g sample, add 2 ml 80 °C water, sonicate. Add IPA, $MgSO_4$ and Hex containing 0.003% BHT. Homogenize, 1 min. Filter. Evaporate. Redissolve in Hex. Inject.	LiChrosorb Si60 5 μm 250 × 4.6 mm	Isocratic 0.27% IPA in Hex Gradient flow 0.9–1.5 ml min^{-1} over 5.3 min	Fluorescence (programmable) α-T acetate λ_{ex} 285 nm λ_{em} 315 nm Tocopherol λ_{ex}290 nm λ_{em} 330 nm Retinyl palmitate λ_{ex} 325 nm λ_{em} 470 nm β-carotene PDA 450 nm	%Recovery 99–101	*Journal of Agricultural and Food Chemistry* 48 (2000) 4003–4008
Vitamin A, E, D_2, D_3, K_1, retinyl acetate, retinyl palmitate, α-T acetate, ergosterol, 7-dehydrocholesterol	Milk, butter	Milk – add 500 μl EtOH containing 0.025% BHT to 1 ml fresh or 2.5 ml fortified milk, sonicate, 2 min. Dilute with water. Clean-up. Mega Bond Elut C_{18}. Butter – add 5 ml EtOH containing 0.025% BHT to 1 g sample. Extract with 15 ml Hex (2×). Wash Hex layer with MeOH–water (9:1). Evaporate Hex layer. Redissolve in 1 ml MeOH. Inject.	Extrasil ODS2 3 μm 150 × 2.1 mm	(A) MeOH:water (99:1) (B) MeOH:THF (70:30) 0–20 min –0% B 20–22 min – 85% B 22–29 min – 85% B 29–33 min – 0% B Flow rate gradient 0–5.5 min – 0.15 ml min^{-1} 6.3 min – 0.2 ml min^{-1}	Wavelength (nm) 325 – vitamin A 264 – vitamin D 280 – vitamin K_1, E, provitamin D 0.7 min – 325 nm 7–9.5 min – 264 nm 9.5–24 min – 280 nm 24 min – 325 nm	DL (ng) 0.14–6.9 % CV 1.6–4.5 % Recovery 84–108	*Analyst* 125 (2000) 427–431

A, acetone; AA, ascorbic acid; BHT, butylated hydroxytoluene; CH_2Cl_2, methylene chloride; Ches, 2-[*N*-cyclohexylamino]-ethanesulfonic acid; $ClHO_4$, perchloric acid; (Vitamin) D_3, cholecalciferol; (Vitamin) D_2 ergocalciferol; DHAA, dehydroascorbic acid; DIPE, diisopropyl ether; DMW, demineralized water; EC, electrochemical; EtOAC, ethyl acetate; EtOH, ethanol; Et_2O, diethyl ether; FAD, flavin adenine dinucleotide; FMN, flavin mononucleotide; GMP, disodium guanosine-5′-monophosphate; HAC, acetic acid; HCl, hydrochloric acid; HEP, heptane; Hepes, *N*-[2-hydroxyethyl]pipera zine-*N*′-ethanesulfonic acid; Hex, hexane; HPLC, high-

strong ultraviolet (UV) absorbance properties. Most LC procedures use UV aborbance between 320 and 380 nm for detection. Fluorescence (excitation: 320 nm, emission 470 nm) is an excellent detection mode for retinol and retinyl esters. Carotenoids absorb strongly in the visible region due to the long conjugated double-bond system. For β-carotene, 450 nm is universally used for detection. Variable wavelength, programmable UV/visible and photodiode array detectors have greatly aided the chromatographic study of complex, natural carotenoid mixtures characteristic of fruit and vegetable samples.

Vitamin D

Vitamin D analysis of nonfortified foods is difficult due to the low concentration and the relatively poor spectral properties of the vitamin. Both normal- and reversed-phase LC efficiently resolve vitamin D_2, D_3, previtamins, and hydroxylated metabolites if properly prepared extracts are available. Excellent methods that use competitive protein binding assays (CPBA) or radioreceptor assays (RRA) using (^{3}H) metabolites are available for analysis of biological matrices. These methods use extensive prepurification of extracts by LC or solid-phase chromatography. CPBA and RRA are specific and sensitive enough to measure circulating levels of hydroxylated vitamin D metabolites; however, the methods have not been applied to food analysis. Radioimmune assay with ^{125}I-labeled metabolites is used extensively in clinical chemistry.

Provided the extraction protocol provides an extract that meets the detection limits of the assay, reversed-phase LC with UV detection at 265 nm is useful as the determinative assay. More recent methods applicable to food analysis incorporate preparative LC for sample clean-up and concentration prior to LC resolution on C-18.

Vitamin E

Vitamin E exists in nature as eight closely related compounds (α-, β-, γ-, δ-tocopherols and corresponding tocotrienols). Normal-phase LC on silica resolves the eight vitamin forms. Reversed-phase LC cannot resolve the β- and γ-positional isomers; therefore, most methods for food analysis use normal-phase LC. Detection does not pose a problem because of the strong fluorescence exhibited by the chromanol family. α-Tocopheryl acetate is normally used for food fortification. This compound is much more stable than α-tocopherol and possesses much less native fluorescence. However, highly sensitive fluorescence detectors that are now readily available can quantify α-tocopheryl acetate at the levels that are commonly used for food fortification.

Vitamin K

Methods for vitamin K analysis of foods use a combination of adsorption chromatography for sample clean-up, and reversed-phase LC on C-18 supports for the quantification. Mobile phases are simple, isocratic systems of methanol modified with dichloromethane. The mobile-phase composition depends on the detection mode. If electrochemical detection is used, an electrolyte must be added to the mobile phase to provide for conductivity. More commonly, vitamin K compounds are converted from the quinone form to the highly fluorescent hydroquinone form by postcolumn zinc metal reduction. This conversion allows the use of fluorescence for detection. The instability of vitamin K at alkaline pH dictates that saponification cannot be used for extraction. Because lipids must be removed from vitamin K extracts before reversed-phase LC, novel lipid removal procedures have been developed for vitamin K extraction. Recent methods (Table 4) use lipase hydrolysis followed by solid-phase extraction to provide a suitable concentrated extract for LC resolution.

Multianalyte Procedures for Fat-soluble Vitamin Analysis

Historically, fat-soluble vitamin analysis of foods has been difficult due to methodology deficiencies of poor detector sensitivity, poor precision due to interferences, and labor intensiveness of the methods. Technological advances in detectors and LC instrumentation together with research on extraction methods for the fat-soluble vitamins from foods have led to the availability of methods with improved precision, specificity, and sensitivity. These improvements have led to the development of multianalyte approaches for fat-soluble vitamin analysis that are especially useful for the analysis of foods fortified with combinations of the fat-soluble vitamins. Such methods, when incorporated into the laboratory analysis program, can save a great deal of time and present significant cost savings.

A recent multianalyte approach developed in our laboratory at the University of Georgia is presented to show the analytical power of multianalyte analysis. The procedure is capable of simultaneously assaying, retinol, retinyl palmitate, retinyl acetate, β-carotene, natural vitamin E, and α-tocopheryl acetate. In this method, the vitamins are extracted in 2-propanol and hexane without saponification. Quantification is by normal-phase LC with fluorescence detection of retinol, the retinyl esters, α-tocopheryl acetate and the vitamin E alcohols. β-Carotene is quantified at 450 nm. Fluorescence and photodiode array detectors

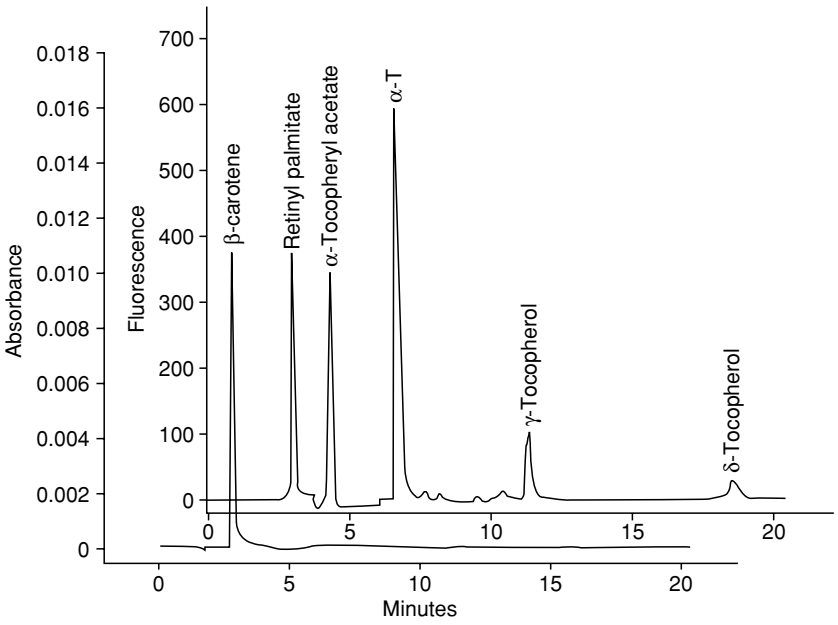

Figure 1 Chromatogram of fat-soluble vitamins in margarine. Chromatography parameters: mobile phase, 0.27% IPA in Hex, gradient flow, 0.9–1.5 ml min^{-1} over 5.3 min. Absorbance β-carotere, 450 nm; fluorescence retinyl palmitate, λ_{ex} 325 nm, λ_{em} 470 nm, α-tocopheryl acetate, λ_{ex} 285 nm, λ^{em} 310 nm; tocopherols, λ_{ex} 290 nm, λ_{em} 330 nm.

in series allow quantification of all analytes from a single injection. Figure 1 shows the elution profile obtained from a margarine sample fortified with retinyl palmitate, β-carotene, and α-tocopheryl acetate.

See also: **Ascorbic Acid**: Properties and Determination; **Carotenoids**: Occurrence, Properties, and Determination; **Cholecalciferol**: Properties and Determination; **Cobalamins**: Properties and Determination; **Retinol**: Properties and Determination; **Riboflavin**: Properties and Determination; Physiology; **Thiamin**: Properties and Determination; Physiology; **Tocopherols**: Properties and Determination; Physiology; **Vitamin K**: Properties and Determination; **Vitamins**: Overview; **Vitamin B_6**: Properties and Determination; Physiology

Further Reading

AOAC International (2000) *Official Methods of Analysis* of AOAC International, 17th edn. Arlington, VA: Association of Official Analytical Chemists.

Ball GFM (1988) *Fat-soluble Vitamin Assays in Food Analysis; A Comprehensive Review*. Oxford: Elsevier Science.

Ball GFM (1994) *Water-soluble Vitamin Assays in Human Nutrition*. New York: Chapman & Hall.

Ball GFM (1998) *Bioavailability and Analysis of Vitamins in Foods*. New York: Chapman & Hall.

De Leenheer AP, Lambert WE and Nelis HJ (1992) *Modern Chromatographic Analysis of Vitamins*. New York: Marcel Dekker.

Eitenmiller RR and DeSouza S (1985) Niacin. In: Augustin J, Klein BP, Becker DA and Venugopal PB (eds) *Methods of Vitamin Assay*, 4th edn. New York: John Wiley.

Eitenmiller RR and Landen WO, Jr. (1995) Vitamins. In: Jeon IJ and Ikins WG (eds), *Analyzing Food for Nutrition Labeling and Hazardous Contaminants*. New York: Marcel Dekker.

Eitenmiller RR and Landen WO, Jr. (1999) *Vitamin Analysis for the Health and Food Sciences*. Boca Raton, FL: CRC Press.

Eitenmiller RR, Landen WO, Jr. and Augustin J (1998) Vitamin analysis. In: Nielsen SS (ed.) *Food Analysis*, 2nd edn. Gaithersburg, MD: Aspen.

Ye L and Eitenmiller RR (2000) Vitamin analysis in food. In: Meyers RA (ed.) *Encyclopedia of Analytical Chemistry*, pp. 4278–4318.Chichester, UK: John Wiley.

VODKA

C F Hayman, Hayman Limited, Witham, Essex, UK

Introduction

The exact origin of vodka is unknown, although both Russia and Poland claim to be the first producers. Vodka was distilled in Russia in the early twelfth century and, as with other spirit drinks, began its life more for medical purposes rather than for enjoyment.

Origin

The word vodka or wodka means 'little water'. It is derived from the Russian word voda. Today the Russians refer to it as vodoshka. Although the origins of vodka may date back some eight centuries, it is generally accepted that vodka as we know it today first appeared during the sixteenth century.

Characteristics of Production

Vodka is a fairly tasteless and neutral spirit product which is mainly derived from the fermentation and distillation of grain. However, some producers will use other raw materials such as potatoes, sugarbeet, grapes, or cassava instead of grain, depending on the local availability and cost. Once the grain or similar has been converted into neutral alcohol by fermentation and distillation, the vodka producer will start the process. Different producers have developed their own methods of production.

Some distillers will further distil or rectify the spirit, retaining only the middle part of the distillation for additional processing. The foreshot and feints are disregarded. The spirit will be passed through charcoal or carbon filters which remove the remaining flavors and odors to give vodka its clean and smooth character.

The various producers have developed their own, often secret, ways of filtering but are very concerned about the contact time of the spirit with the charcoal. Some distillers have a single filter (Figure 1) while others require the spirit to pass through a series of filters. The choice of charcoal varies as it affects the final product. One producer uses charcoal specially made from Sussex oak trees; another uses carbon derived from peat found in the Dutch polders. After filtration, some distilleries will regenerate the charcoal using steam whilst others discard the spent charcoal.

The character of the water added to the finished vodka to reduce it to the required bottling strength is very important. The increased technology of water demineralization plants allows a greater consistency in quality than ever before. Water quality is of a vital significance in the final taste of the product.

In the European Community, the minimum level of alcohol acceptable for vodka is 37.5% (v/v), but often higher-strength vodkas are available, particularly coming from eastern Europe, e.g., 45–55%.

Many vodkas are flavored and this again originated from both Poland and Russia. One Polish vodka is flavoured by steeping buffalo grass (*Hierochloe odorata*) in the product to give it an additional aromatic flavor, and it also takes on some of the green color from the grass. An old Russian vodka is prepared with an infusion of leaves from apple and pear trees, to which is added brandy and port-type wine. This vodka is mellow in taste and pleasing to drink. Some high-strength vodkas are aged in wood, often giving them a brown color derived from the cask. In more recent times, Pepper vodka has been produced,

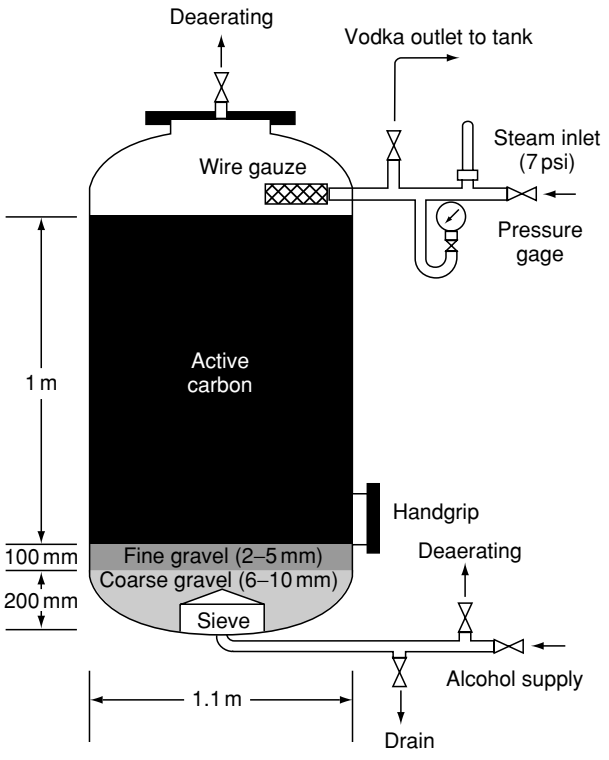

Figure 1 A typical vodka filter. Reproduced from Vodka, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

the pepper enhancing its aroma and giving it a slight burning sensation on consumption. Other vodkas flavored with the addition of lemon, peach, and other fruit flavors are now appearing. (*See* **Barrels**: Wines, Spirits, and other Beverages; **Flavor (Flavour) Compounds**: Structures and Characteristics.)

See also: **Barrels**: Wines, Spirits, and Other Beverages; **Flavor (Flavour) Compounds**: Structures and Characteristics

Further Reading

Begg O (1998) *The Vodka Companion*. London: Quinted Publishing Ltd.

Clutton D (1979) *Brewers Guardian* 108(10): 25.

Pridemore WA (2002) Vodka and violence: alcohol consumption and homicide rates in Russia. *American Journal of Public Health* 92(12): 1921–1930.

Rose AM (1977) *Alcoholic Beverages*. Economic Microbiology Series, vol. 1. London: Academic Press.

W

Wafers *See* **Biscuits, Cookies, and Crackers**: Nature of the Products; Methods of Manufacture; Chemistry of Biscuit Making; Wafers

WALNUTS AND PECANS

R B N Prasad, Indian Institute of Chemical Technology, Hyderabad, India

Introduction

Many of the species of trees in the family Juglandaceae produce edible, oil-rich nuts, and these include walnuts and pecans. The valuable part of the commercial crop is the whole nut. This article reviews the varieties, global distribution, commercial importance, morphology, and anatomy of the fruits and seeds, chemical and nutritional composition, handling, and storage of walnuts and pecans.

Walnuts

Varieties and Global Distribution

Walnuts are common, large, forest deciduous trees primarily of temperate areas, but also of subtropical regions, mostly of eastern North America, Central America, western South America, and Eastern Asia. Walnuts belong to the genus *Juglans*, which consists of about 15 species.

The most important species, producing the common, Persian, or English walnut, is *J. regia* L. It was probably moved by migrating populations from ancient Persia to Greece and later distributed throughout the Roman empire. The early colonists from the UK brought seeds to North America, and the settlers called the resulting trees 'English walnuts' to distinguish them from the native American black walnuts. The English walnut is cultivated commercially in the USA, France, Italy, former Yugoslavia, Poland, Germany, Czech Republic, Bulgaria, Turkey, Chile, Northern India, China, and Australia.

The second principal variety is the black walnut (*J. nigra* L.), which is native to the USA and is found distributed along the coast, west to Michigan and throughout most of the midwestern states.

A third walnut variety, less important from a commercial standpoint, is the butternut (*J. cinerea* L.). This tree is native to North America and is found in the eastern regions.

Other popular species of walnut trees include the following: Arizona (or Nogal) walnut, *J. major* Helier (distributed in New Mexico and Arizona); the California walnut, *J. californica* Walt. (southern California); the Chinese walnut, *J. cathayensis* Dode (central and western China); the Hinds walnut, *J. hindsii* Sarg. (California); the Texas walnut, *J. rupestris* Engelm. (south-western USA to Mexico); the Manchurian walnut, *J. mandshurica* Maxim. (eastern Commonwealth of Independent States and northern China); and the Japanese walnut, *J. sieboldiana* Maxim. (Japan). Some of these species are favored for garden and park plantings.

Commercial Importance

Although the most important and best known commercial species is undoubtedly the English walnut, many commercial varieties of different species are grown around the world. China is the leading producer of walnuts, at 300 000 mt in the year 2000, out of 1 183 898 mt of the world production. The other important walnut-growing countries are the USA (222 260 mt), Iran (142 906 mt), Turkey (122 000), India (31 000 mt), France (27 780 mt) and Romania (25 000 mt). The nuts of all species are edible, and the young green fruits, before the nuts harden, are also eaten pickled in vinegar. Mature nuts can be eaten raw or salted, and are excellent for dessert, in icecreams and candy, and also in baking and

confectionery. Walnuts are a part of holiday celebrations around the world and especially in places with German tradition. Families sit around the table and crack walnuts, eating the fresh meats. Although the black walnut has a richer flavor than the English walnut, it is not as popular because its shell is much thicker and harder to remove without breaking the kernels. The kernel forms 42–60% of the nut. About 300–600 t of walnut oil is produced from waste material from shelling and processing plants, e.g., the inedible kernels rejected during shelling and fragments of kernels recovered from shells. Walnut oil makes a good salad oil, and can also be used in the preparation of soap, and as a drying oil in paints. Defatted walnut cake is useful as an animal feed.

The walnut shells, reduced to flour of various mesh sizes, have found applications as fillers in synthetic resin adhesives, plastics, and industrial tiles, and as an insecticide diluent. About 50% of the annual production of walnut shell is used in the plastics industry alone. The shell powder is also used for drilling mud in oil fields, and as an abrasive in polishing metal castings. Dyes prepared from walnut shells and leaves are still used in Turkey, and the high tannin content of green walnut shells is useful in leather manufacture. A pharmaceutical grade of activated charcoal can be prepared from the inedible parts of the fruit. Various parts of the walnut tree, such as bark, leaves, shells, fruits, and kernels, have been used in folk medicine throughout the world. The timber of walnuts possesses excellent mechanical strength and very good shock resistance, and hence it is much valued for veneers, furniture, boat-building, and gunstocks.

Morphology and Anatomy of the Fruits and Seeds

Walnuts are usually tall, broad-headed, deciduous trees that grow 10–50 m in height. They have rough bark, and compound pinnate leaves that exude an aromatic fragrance when crushed or bruised, a property which varies in degree from one species to another.

The English walnut is a large, deciduous tree attaining over 30 m in height and with a trunk of 6 m in girth; the bark is gray, smooth on young trees, and fissuring with age. The staminate (male) flowers of this species are borne on twigs of the previous season's growth, and the pistillate (female) flowers, usually few, are borne on the twigs of the current season's shoots. Pollination occurs naturally via the wind, but numerous cultivated varieties have recently been developed by other means. The green fruit is 3.7–5.0 cm in length with a subglobose, gland-dotted, and glabrous structure. The fruit is a drupe, and consists of an outer layer and an inner nut, inside which is the single seed containing abundant oil. When the seed is ripe, the husk opens and the ovoid-shaped stone falls to the ground. The surface of the stone is wrinkled, showing the suture of the carpels in the median vertical plane. The walnut has an outer green husk which is a perfect seal from insects. Inside there is a hard shell. Nature has created a perfect package. The seed consists mainly of two large cotyledons storing fatty substances.

The macroscopic and microscopic structures of the walnut (*J. regia*) seed are given in Figure 1a–c. Most of the endocarp consists of stone cells with brown walls. The outermost of these have very thick walls with slit-like pits, but towards the interior the walls are thinner, with relatively large pits. Inside the endocarp there is a zone of flattened parenchyma cells with brown walls. The outer epidermis of the testa consists of rather radially elongated cells up to 90 μm in length. In some places, this layer is biseriate; pigment is found in some of these cells. It has been reported that raised stomata are present in the outer epidermis. The rest of the testa consists of three or four rows of flattened parenchyma cells with brown walls, and also some vascular tissue. The endosperm is normally represented by a single layer of cells, up to 16 μm in radial diameter, with small aleurone grains and oil drops. The parenchyma cells of the cotyledon possess intercellular spaces and oil drops, and contain aleurone grains, up to 10 μm in diameter (Figure 1a–c).

The black walnut tree grows up to 50 m in height and has brownish bark with fissuring. The drupelike fruits are solitary or in pairs, 2.5–5.0 cm in length, globose, oblong and pointed at the apex, or slightly pyriform with a thick husk. The nut is oval or oblong, slightly flattened, 3.0–3.8 cm in diameter, deeply divided on the outer surface into thin or thick, often interrupted, irregular ridges.

Chemical and Nutritional Composition

Chemical characteristics of oil Walnut oil is generally produced from the inedible meat rejected during shelling, plus meat fragments recovered from the shell. The oils from the different species of walnut bear a close resemblance to each other in terms of both physical and chemical characteristics. They are light yellow in color with a greenish tinge and a faint, nutty odor. The range of chemical characteristics of walnut oil is given in Table 1.

Fat content Walnut kernels generally contain about 60% oil, but vary from 52% to 70% depending on the variety. The major constituents of the oil are triacylglycerols. Free fatty acids, diacylglycerol, monoacylglycerol, sterols, sterol esters, phospholipids, and vitamins are present in minor quantities. The triacylglycerol moiety of the oil is a mixture of

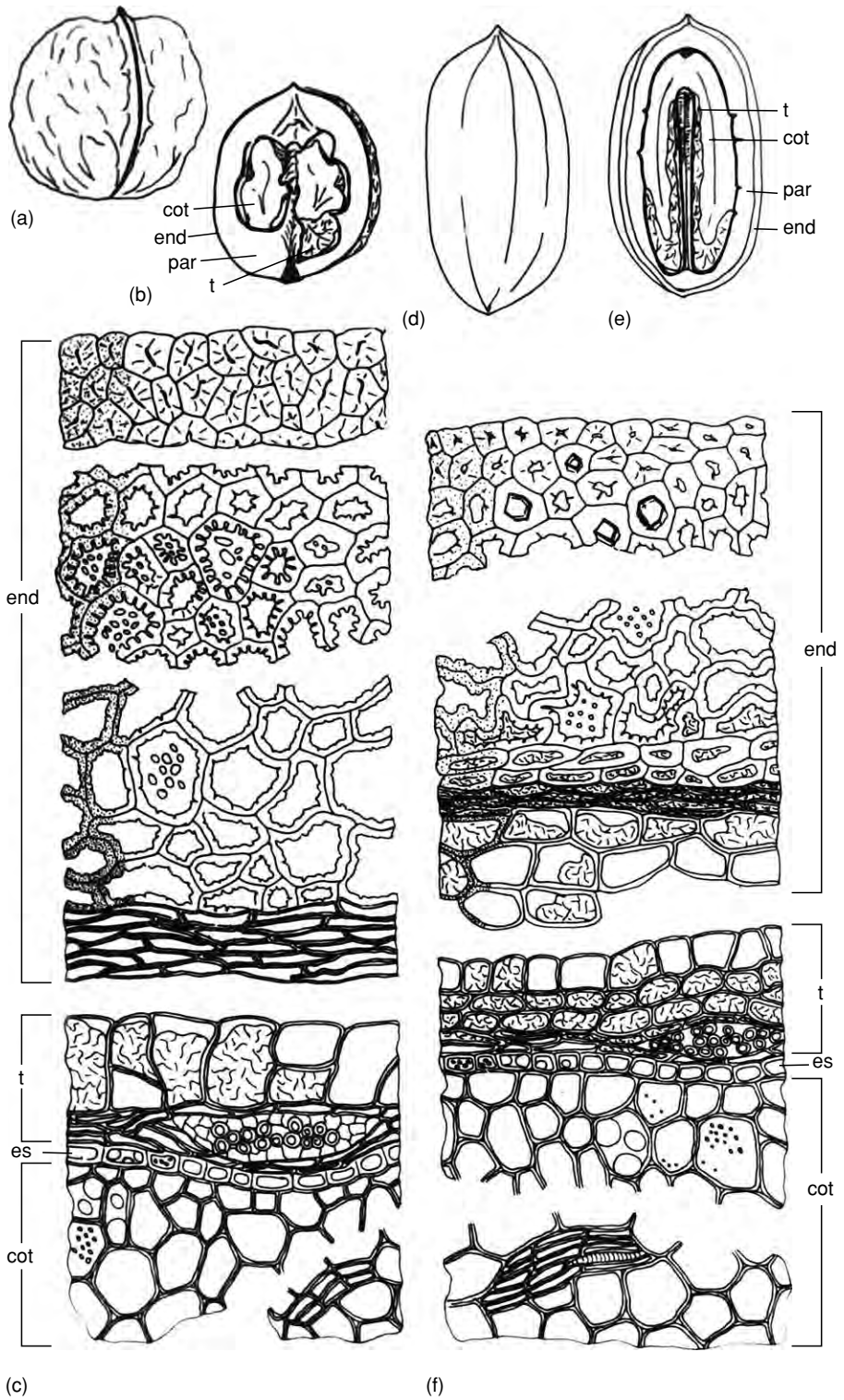

Figure 1 (a–c) Walnut *(Juglans regia)*. (a) Nut. (b) Longitudinal section of nut showing: end, endocarp; par, partition; cot, cotyledone, t, testa. (c) Transverse section: abbreviations as in (b); es, endosperm; × 250. (d–f) Pecan (c*arya illinoensis*). (d) Nut. (e) Longitudinal section of nut; (f) Transverse section of endocarp, testa, endosperm, and cotyledon; × 270. Reproduced from Vaughan JG (1970) Juglandaceae. In: *The Structure and Utilization of Oilseeds*, pp. 113–116. London: Chapman and Hall, with permission.

Table 1 Chemical characteristics of walnut and pecan oils

Property	*Walnuts*	*Pecans*
Specific gravity (at 15 °C)	0.919–0.929	0.911–0.914 (25 °C)
Refractive index (n_D 25 °C)	1.469–1.478	1.469–1.471
Acid value	0.5–10.0	1.1–10.5
Saponification value	189–197	189–198
Iodine value	132–162	77–106
Polenske value	0.2–0.3	0.3
Unsaponifiable matter (g 100 g^{-1})	0.2–1.0	0.35–0.67
Solidification point (°C)	−12 to −28	−18

Data compiled from: Krischenbauer (1960) In: *Fats and Oils – An Outline of their Chemistry and Technology*, pp. 198–203. New York: Reinhold; Eckey EW (1954) In: *Vegetable Fats and Oils*, pp. 378–382. New York: Reinhold; Jamieson GS (1943) In: *Vegetable Fats and Oils*, pp. 158, 333–334. New York: Reinhold; Mehran M (1974) Oil characteristics of Iranian walnuts. *Journal of the American Oil Chemistry Society* 51: 477–478; Council of Scientific and Industrial Research (1959) *The Wealth of India – Raw Materials*, vol. V, pp. 298–305. New Delhi: Council of Scientific and Industrial Research.

triunsaturated and nonsymmetrical diunsaturated glycerides that form up to 83–95% of the total fraction. The 2-position of the triacylglycerol contains primarily linoleic acid. The fatty acids of the walnut oil are predominantly unsaturated and consist mainly of linoleic and oleic acids (Table 2). The linoleic acid content in English walnut is higher than that in black walnut. (*See* **Fatty Acids**: Properties; **Triglycerides**: Structures and Properties.)

Walnut oil contains about 0.9–1.3% phospholipids; the composition of individual components of total phospholipids is: phosphatidylcholine, 18–20.4%; phosphatidylinositol, 28.2–45.0%; phosphatidylethanolamine, 15–30.5%; cardiolipin, 0–20.4%; phosphatidylserine, 0–5%; phosphatidic acid, 0–7%; lysophosphatidylcholine, traces to 2.0%; lysophosphatidylethanolamine, traces to 2.0%.

Walnut oil contains at least 29 volatile components, such as terpenes, alcohols, and carbonyls, and the characteristic odor of the oil is attributable to the collective effect of a number of constituents. (*See* **Sensory Evaluation**: Aroma.)

Protein content Walnut kernels contain about 14.5–24% of protein, and this is enriched up to 61–66% in dry defatted cake. Full details of the nature of the protein are not available, but the molecular weight of the three main proteins was estimated to be about 43 000, 39 000, and 25 000. The content of the sulfhydryl and disulfide bonds in the walnut protein was reported to be higher than that of soybean and yeast protein. Walnut cake contains more arginine and less lysine than does casein and soybean cake. The walnut kernel protein contains 18 amino acids, and glumatic acid (2.809 g per 100 g of kernel) is the main amino acid, followed by arginine (2.103 g per 100 g of kernel). The essential amino acid composition is given in Table 3. (*See* **Amino Acids**: Properties and Occurrence.)

In addition to the normal amino acids, walnut kernels are found to contain a sulfur amino acid, taurine (2-aminoethanesulfonic acid) in concentrations of about 15–46 nmol g^{-1}, depending on the variety. In humans, taurine deficiency may lead to a

Table 2 Fatty acid composition of walnut and pecan oils (g per 100 g total fatty acids)

Fatty acid	*Walnuts*	*Pecans*
Myristic	0–2	
Palmitic	2.9–8.2	3.3–7.1
Palmitoleic	0–0.1	
Stearic	0.6–2.2	0.9–5.5
Oleic	9.0–39.2	40.6–88
Linoleic	42.5–76	13.5–38
Linolenic	2.0–18.0	0.8–2.0
Gadoleic	0–0.4	0.21–0.25

Data compiled from: Mehram M (1974) Oil characteristics of Iranian walnuts. *Journal of the American Oil and Chemistry Society* 51: 477–478; Jamieson GS and Mekinney RS (1936) Fatty acid composition of black walnut oil. *Oil and Soap* 13: 202; Rockland LB and Benedict CD (1970) Occurrence of fatty acid methyl esters in walnut kernel and other oils. *Journal of Agriculture and Food Chemistry* 18: 223–228; Savage GP, Dutta PC and Mc Neil DL (1999) Fatty acid and tocopherol contents and oxidative stability of walnut oils. *Journal of the American Oil and Chemistry Society* 76: 1059–1063; Hilditch TP and Williams PN (1964) In: *the Chemical Constitution of Natural Fats*, 4th edn; London: Chapman & Hall. Young CT (1996) Nuts. In: Kroschwitz J and Howe-Grant M (eds) *Kirk-Othmer Encyclopaedia of Chemical Technology*, 4th edn, vol. 17, pp. 544–579. New York: John Wiley; French RB (1962) Analyses of pecan, peanut and other oils by gas chromatography and ultra violet spectrophotometry. *Journal of the American Oil and Chemistry Society* 39: 176–181; Jamieson GS (1943) *Vegetable Fats and Oils*, p. 158. New York: Reinhold; USDA Database for Standard References *Release 11* September 1996 and *Release 13* November 1999.

Table 3 Essential amino acid component of walnut and pecan kernels

Amino acid	*Amino acid contents (g 100 g^{-1})*	
	Walnuts	*Pecans*
Ariginine	2.103–3.661	0.848–1.185
Histidine	0.359–0.680	0.198–0.273
Isoleucine	0.566–0.978	0.322–0.533
Leucine	0.992–1.704	0.520–0.773
Lysine	0.285–0.721	0.287–0.435
Methionine	0.247–0.473	0.112–0.186
Phenylalamine	0.628–1.107	0.392–0.564
Threonine	0.230–0.589	0.253–0.389
Tryptophan	0.175–0.321	0.093–0.199
Valine	0.723–1.286	0.345–0.525

Data compiled from: USDA Database for Standard Reference *Release 11* September 1996 and *Release 13* November 1999; Considine PEDM and Consodine GD (eds) (1982) Pecan and walnut. In: *Foods and Food Production Encyclopaedia*, pp. 1345–1350. New York: Van Nostrand Reinhold; Young CT (1996) Nuts. In: Kroschwitz J and Howe-Grant M (eds) *Kirk-Othmer Encyclopaedia of Chemical Technology*, 4th edn, vol. 17, pp. 544–579 New York: John Wiley.

decreased electroretinogram (ERG) and to pigmentary degeneration of the retina. Hence walnut kernels have the potential to replace meat as a source of dietary taurine. (*See* **Protein**: Requirements.)

Carbohydrate The total carbohydrate content in different varieties of walnut varies from 12.1 to 18.3 g $100\,g^{-1}$ kernels. English walnut is rich in carbohydrate compared to the black variety. The fiber content in walnut varieties is about 1.7–6.5 g $100\,g^{-1}$ kernel. Some varieties of walnut kernels contain large amounts of sugar(s). (*See* **Carbohydrates**: Classification and Properties.)

Vitamins Walnut kernels are a good source of vitamins. Walnuts contain about 30–300 IU of vitamin A, 0.22–0.45 mg thiamin, 0.10–0.16 mg riboflavin, and 0.7–1.105 mg niacin $100\,g^{-1}$ of kernel. The richest sources of vitamin C are the immature fruits or their green hulls, and the leaves of walnuts. Unripened walnuts are reported to have a very high vitamin C content (2400–3700 mg $100\,g^{-1}$), displaying an activity 40–50 times as high as that of oranges and lemons. However, the dry walnut kernels are reported to contain only 0.4–8.5 mg $100\,g^{-1}$ after ripening. Reports of large amounts of vitamin C in green walnuts have yet to be fully confirmed.

Walnuts contain about 7.3–28.7 μg α-tocopherol, 1.0–8.2 μg β-tocopherol, 205–375.8 μg γ-tocopherol and 28.0–62.1 μg δ-tocopherol g^{-1} of its oil. Walnut kernels are also found to contain vitamin K (0.9–1.1 mg $100\,g^{-1}$). Refer to individual vitamins.

Phenolic acids In food products, the presence of phenolic acids has been associated with astringency, discoloration, inhibition of enzyme activity, and antioxidant properties. Defatted kernels are found to contain some phenolic acids, namely phenylacetic, protocatechuic, syringic, vanillic, gallic, caffeic, and ferulic acids, in very small quantities ranging from 0.02 to 0.20 μg g^{-1} of kernels. (*See* **Phenolic Compounds**.)

Pectic substances Walnuts contain about 2% (dry matter in the edible portion) pectic substances (expressed as anhydrogalacturonic acid). Pectic substances, which make up a component of dietary fiber, are basically polymers of galacturonic acid and may have important physiological and nutritional consequences, such as hypocholesterolemic effects, increased excretion of fecal sterols, and the capacity to bind bile salts. Pectic substances may also slow down the absorption of soluble carbohydrates, causing a smaller increase in the postprandial level of blood sugar; hence walnut kernels could play a role in the design of pectic-rich diets if required for medical reasons.

Juglone Juglone (5-hydroxy-1,4,-naphthoquinone), a yellow-brown compound, was shown to be a host factor that may be associated with resistance of walnut to scab caused by fungus. Juglone is present in considerable amounts in all green and growing parts of the trees and unripe hulls of the nuts. Juglone was also reported in the surface waxes of English walnut leaves (1800 μg g^{-1} of fresh leaf) and fruits (97 μg g^{-1} of fresh fruit) along with hydrocarbons, wax esters, aldehydes, fatty alcohols, and fatty acids. However, juglone levels in kernels was either very low or absent.

Minerals Walnuts are considered to be a good source of dietary minerals. English and black walnuts contain almost similar quantities of minerals, except calcium, which is present only in black walnuts. Potassium, phosphorus, magnesium, and iron are found in significant quantities in these nuts. The mineral composition of walnuts is given in Table 4. Refer to individual minerals.

Nutritional Quality

Walnut kernels exhibit high nutritive value compared to cereals and legumes and therefore find a great

Table 4 Minerals of walnut and pecan kernels

Minerals	*Mineral content (mg 100 g^{-1})*	
	Walnuts	*Pecans*
Calcium	Traces–108	21–139
Copper	0.31–3.2	0.8–1.4
Iron	2.1–7.6	1.9–4.8
Magnesium	131–169	120–170
Manganese	2.4–2.9	1.7–5.3
Phosphorus	309–570	271–610
Potassium	328–687	330–660
Sodium	1–16.5	0–1
Sulfur	104–140	0–113.0
Zinc	2.3–3.6	5.3–10.4

Data compiled from: Senter SD (1976) Mineral composition of pecan nut meats. *Journal of Food Science* 41: 963–964; Hammar HE and Hunte JH (1946) Chemical composition of pecan nut kernels. *Plant Physiology* 21: 476–478; Young CT (1996) Nuts. In: Kroschwitz J and Howe-Grant M (eds), *Kirk-Othmer Encyclopaedia of Chemical Technology*, 4th edn, vol. 17, pp. 544–579. New York: John Wiley; Considine PEDM and Consodine GD (eds) (1982) Pecan and walnut. In: *Foods and Food Production Encyclopaedia*, pp. 1345–1350. New York: Van Nostrand Reinhold; Council of Scientific and Industrial Research (1959) *The Wealth of India – Raw Materials*, vol. V, pp. 298–305. New Delhi: Council of Scientific and Industrial Research; *Composition of Foods, Agriculture Hand Book* no. 8 (1963) Agriculture Research Service, US Department of Agriculture; Cristina VP and Maria ETI (1981) Mineral elements in walnuts. *Anales Bromatologia* 33: 69–76; USDA Database for Standard References *Release 11* September 1996 and *Release 13* November 1999.

importance in food industry. A high level of fat, containing unsaturated fatty acids with considerable amounts of natural antioxidants like tocopherols and phenolic acids, enhances its importance as a good edible oil source. Defatted walnut kernels are useful as a dietary supplement as they contain a very high content of digestible protein and all the essential amino acids, taurine, and significant amounts of vitamins and minerals. Walnuts are claimed to have a hypocholesterolemic effect on the basis of richness in magnesium and fiber, and a low lysine and arginine ratio. The positive nutritional advantages of walnuts in lowering blood cholesterol come from the high levels of mono- and polyunsaturated fatty acids and possibly the tocopherol content. A diet that includes moderate quantities of walnuts without an overall increase in total dietary fat and calories decreases serum cholesterol levels and favorably modifies the lipoprotein profile in normal men. Walnut kernels are a good source of energy and provide about 2686 kJ $100\,g^{-1}$. (*See* **Energy**: Measurement of Energy Expenditure.)

Industrial Uses of Oil

Walnut oil is similar to linseed oil in composition, and its drying properties have been extensively used in paints and artists' colors from early times. In comparison with black walnut oil, English walnut oil contains more linoleic acid and has more active drying properties. Walnut oil is also used in the manufacture of soaps, varnishes, alkyd resins, and styrenated oils. In addition, since the Second World War, walnut oil has been consumed as a specialty food oil.

Handling and Storage

Rancidity is one of the first signs of deterioration in walnuts because they are rich in oil containing unsaturated fatty acids. High temperature and humidity, as well as sunlight, favor the development of rancidity in kernels, and the best storage conditions are therefore low temperature, low humidity, and little or no light. Aflatoxin presents another problem. In general, aflatoxin production occurs during the harvest period after the nuts have begun to dry; the nuts must therefore pass quickly through this critical moisture zone. Storage of nuts under proper temperature and humidity will prevent further contamination, and there are many well-established methods of destroying aflatoxins in the kernels and oils. (*See* **Mycotoxins**: Occurrence and Determination; **Oxidation of Food Components**.)

The optimum moisture level for the storage of walnuts is 3.2–3.8%, and stability of the shelled nuts is significantly reduced at higher or lower moisture levels. Very elaborate machinery is used for cleaning, grading, bleaching, cracking, and packaging of edible tree nuts, including walnuts. Skin color of the nuts is very important, and methods such as the photometric estimation of absorbance of a methanol extract of the whole nuts are useful for determining the rate of darkening of walnuts at regular intervals during storage. Walnuts are bleached for market by dipping them for 5–10 s in an aqueous solution of bleaching powder and sodium carbonate, with a small amount of sulfuric acid to clear the liquid. In another method, walnuts are exposed to ethylene (one part ethylene to 1000 parts air) to preserve the natural color of the kernel. This treatment prevents the hulls from sticking and staining the shells. After hulling and washing, the walnuts are placed in bins and dried with warm air (not over 43 °C) to reduce the moisture content. The best storage conditions for walnuts are 0–2 °C, accompanied by ventilation with air of relative humidity in the range 65–75%. The storage methods adopted for pecan nuts are also useful for walnuts. (*See* **Preservation of Food**; **Storage Stability**: Parameters Affecting Storage Stability.)

Pecans

Varieties and Global Distribution

The species *Carya illinoensis* Koch, or *C. pecan* Marsh, are known internationally as pecans. Botanically pecan is classified as a hickery (carya) and is a member of the walnut family (Juglandaceae). Native to North America, the pecan is grown mainly in southern and central USA and northern Mexico, with limited cultivation in Australia, South Africa, Brazil, and Israel. Commercially, well over 300 varieties of pecan are offered. Pecan is one of the major commercial edible nuts, and is produced in quantities exceeding 153 500 t $year^{-1}$ in 2000–2001 in the USA.

Commercial Importance

The nuts of pecan have been of value for human consumption in North America for the last three centuries. Pecan is next in importance to the English walnut for its valuable nuts; hence more and more orchards are being planted, not only in southern USA, but also in other countries. The coarse, brittle wood is not as strong as walnut, but it is used occasionally to make tool handles and agricultural implements, and as fuel.

The nuts are eaten raw or salted, and are excellent for dessert, icecreams, cakes, bread, candies and other confectionery, and in vegetarian croquettes and sandwiches. Pecans have a mild, sweet, walnut-like flavor.

About 300 t of pecan oil is produced annually from the pecan-shelling industry. The oil is refined and used for edible purpose or for the production of soap cosmetics and some drugs; the cake is used in animal feed.

Waste materials from the pecan-shelling industry have shown commercial possibilities for the recovery and production of oil, tannin, shell flour, and activated charcoal. Tannic acid produced from pecan shells is used by the tanning industry and by the oil industry to control the viscosity of drilling muds. The shell flour of pecan, in common with that of walnut, has many industrial applications.

Morphology and Anatomy of the Fruits and Seeds

The pecan is a large, deciduous tree which grows up to 9–12 m high – some specimens may reach a height of 50 m – and the trunk can be up to 2 m in diameter, with gray, furrowed bark.

The male flowers are borne in three-branched pendulous catkins, the female in (two to 10) spikes. The fruit is a drupe (3–7 cm long) which, at maturity, splits into four valves to reveal the smooth, brown kernel. The nut is ovoid to ellipsoidal, nearly cylindrical or slightly four-angled towards the pointed apex, rounded and usually apiculate at the base. The brown shelled nut is about 3.5 cm in length and is wrapped in a husk which usually splits and falls away. If the husk does not break away, the nuts are usually defective. The nuts contain 39–55% meat, most of which can be separated in a stable form, but a small portion (about 3%) remains with the shell.

The macroscopic and microscopic structures of the pecan nut are shown in Figure 1d–f. The endocarp of the pecan nut has essentially the same structure as the walnut endocarp, except that some of the outermost stone cells have diamond-shaped crystals in their cavities. Some of the cells of the outer epidermis, in a transverse section of the testa, are square-shaped and others tangentially elongated, with a radial diameter of up to 18 μm and a tangential diameter of 11–32 μm. Some of these cells contain a brown pigment, which may also be found in the guard cells of the stomata. The rest of the testa consists of vascular tissue and parenchyma, some cells of which contain pigment. The endosperm of the pecan nut is the same as that of the walnut. The embryo is also similar, but the aleurone grains are smaller, about 5 μm in diameter, and there may be minute starch grains present (Figure 1d–f).

Chemical and Nutritional Composition

Chemical characteristics of oil On account of the value of the nuts for edible purposes, pecan oil is not prepared commercially. However, the unusable fine nut fragments which accumulate at shelling plants can be utilized for the production of oil. The chemical characteristics of pecan oil are given in Table 1.

Fat content Pecan kernels contain 60–75% oil depending on the cultivars. Triacylglycerol forms more than 95% of total lipids, while monoacylglycerols, diacylglycerols, complex lipids, free fatty acids, and sterols are present at levels of less than 1%. The fatty acids of pecan oil are mainly unsaturated, with oleic acid as the predominant fatty acid (unlike walnut oil, in which linoleic is the major acid; Table 2).

Protein content The protein content of pecan kernels varies from 7.8 to 17% depending on the cultivar. The pecan kernel protein also contains 18 amino acids and glutamic acid is the main amino acid (1.829 g 100 g^{-1} kernels), followed by arginine (1.177 g 100 g^{-1} kernels) in pecan kernels. The essential amino acid content of the kernels is given in Table 3.

Carbohydrate content The total carbohydrate and fiber contents of pecan kernels are about 13.9–18.3 and 1.6–9.6 g 100 g^{-1}, respectively. The total reducing sugar content is lower in pecan than most other nuts; the amount of sucrose is 2.0 g 100 g^{-1} kernel.

Juglone Juglone is also present in all green and growing parts of the trees and unripened hulls of pecan nuts. However, the kernels contain minute amounts of juglone.

Vitamins Pecan nuts contain about 77–130 IU vitamin A, 0.66–0.86 mg thiamin, 0.11–0.13 mg riboflavin, 0.89–1.167 mg niacin, and 1.1–2.0 mg ascorbic acid 100 g^{-1} kernel. The nuts contain 170 μg γ-tocopherol and 15 μg α-tocopherol g^{-1}. The δ-isomer is present in minute quantities.

Minerals Although some minerals, such as copper, iron, cobalt, and chromium, decrease the stability of vegetable oils, many minerals are essential for normal metabolic functions and are required components in a balanced diet. Pecan kernels are rich in dietary minerals as they contain phosphorus, potassium, magnesium, and sulfur in fairly large quantities (Table 4), and others, such as iron, copper, zinc, and molybdenum, in microquantities.

Tannins The integument of the pecan contains up to 25% tannin, and hence the crop is a very good source

of commercial vegetable tannin. Tannic acid produced from pecan shells is used by the tanning industry. It is also used by the oil industry to control the viscosity of drilling muds. (*See* **Tannins and Polyphenols**.)

Phenolic acids Pecan kernels also contain phenolic acids, namely gallic, gentisic, vanillic, protocatechuic, *p*-hydroxybenzoic, and *p*-hydroxyphenylacetic acids. Coumaric acid and syringic acids are present in trace amounts. The phenolic compounds were located only in the testa of the kernel. The total phenolic acid content in the defatted kernels was found to be about 171.4 $\mu g\,g^{-1}$; gallic acid was predominant and constituted about 78% of those present. The content of all the major phenolic acids except *p*-hydroxy phenylacetic acid progressively decreases during storage, and thus they appear to function as antioxidants to maintain the chemical integrity of the kernel. (*See* **Antioxidants**: Natural Antioxidants.)

Nutritional Quality

Pecan nuts are mainly used as appetizers and additions to salads, cakes, candies, and cookies or as emergency rations for persons expending great amounts of energy, rather than a basic diet. These kernels are high in fat content with a calorie count close to butter. Pecan kernels are nutritious and a good source of energy and vitamins. Defatted kernel powder is useful as a dietary supplement as it contains all the essential amino acids and significant amounts of minerals and natural antioxidants like phenolic acids and tocopherols. The oil is used in cooking and salads and has good keeping qualities. Pecan kernels are a good source of energy and provide about 2891 $kJ\,100\,g^{-1}$. (*See* **Energy**: Measurement of Food Energy.)

Industrial Uses of Oil

Like walnut oil, pecan nut oil is used in food, soap, enamel paints, oil colors, and printing inks. The refined oil is a clear, neutral, edible oil but is sold mostly for special uses to the pharmaceutical, essential oils, and cosmetics industries. In the foreseeable future, pecan oil will remain a byproduct oil, since the main part of the crop is too valuable in the form of nuts or nut meats.

Handling and Storage

More than 50% of pecan kernels offered for sale may, in any one year, fail to meet minimum US Department of Agriculture (USDA) requirements because of discoloration, rancidity, and staleness. Hence it is very important to harvest at the appropriate time, and store the crop under proper conditions to maintain the quality of the nuts.

Early harvest has distinct advantages as a means of avoiding unfavorable weather conditions, and the color and flavor scores of early advanced pecans that are dried and refrigerated promptly change at a much slower rate during storage than do late-harvested crops. A moisture content of 4–5% is recommended to maintain the color.

Very elaborate machinery is used for cleaning, grading, bleaching, cracking, and packing the pecan nuts. In commercial shelling plants, in-shell pecans are conditioned prior to cracking and shelling to improve the yield of halves. In general, conditioning can be achieved by soaking pecan nuts in large vats of water containing chlorine at 1000 p.p.m. for 1–2 h, and then holding them for 12–24 h before cracking. Another method used by processors is soaking the pecans in water (85 °C) for 3–5 min, followed by holding for 12–24 h before cracking. The most effective process for improving the yield of halves from shells, and for maintaining the quality of the kernels, is the 3-min atmospheric steam process. Immediately after cracking, the moisture content is reduced to less than 4.5% for pecan halves and less than 3.5% for pieces in order to maintain quality.

Acidification of pecan kernels with dilute acid solutions significantly lightens the color of the kernels, and phosphoric acid has proved to be the most effective. Dilute solutions of citric acid can also enhance the color but, unlike phosphoric acid, they adversely affect the taste. Treatment with dilute hydrochloric acid, ammonia, and sulfur dioxide can improve the color but responses are variable. The color of the pecan kernels is considered a primary quality parameter of the shelled nuts and, in the USA, they are graded into four color classes by the USDA: light, light amber, amber, and dark amber.

Rancidity is the major problem for pecan kernels during storage, as the oil is rich in unsaturated fatty acids. The development of rancidity also influences the flavor of pecans, and they can be kept for about 3–4 months only, at ordinary temperatures, without developing rancidity. However, they may be stored, soon after harvest, at 3 °C for 1 year, without developing off-flavors at − 4.5 °C for 3 years and at −17.5 °C for more than 5 years. It is believed that kernels do not become rancid more rapidly after being removed from storage than fresh pecans exposed to the same conditions.

In pecans, molds which are capable of producing aflatoxins are of great concern to processors and marketers because fungi like *Penicillium*, *Aspergillus* and *Fusarium* species are capable of growing on pecans at 0 °C and, to a lesser extent, at −6–5 °C.

However, pecan kernels can be preserved against fungal deterioration and discoloration by reducing the moisture content rapidly, after harvest, with moderately warm (< 40 °C), dry air circulated constantly around the nuts, followed by storing in ventilated chambers containing air of relative humidity in the range 60–68%. Chlorine or propylene oxide treatments are also employed by the pecan industry to disinfect in-shell pecans, halves, and pieces.

See also: **Amino Acids**: Properties and Occurrence; **Antioxidants**: Natural Antioxidants; **Carbohydrates**: Classification and Properties; **Energy**: Measurement of Food Energy; Measurement of Energy Expenditure; **Fatty Acids**: Properties; **Mycotoxins**: Occurrence and Determination; **Oxidation of Food Components**; **Phenolic Compounds**; **Preservation of Food**; **Protein**: Requirements; **Sensory Evaluation**: Aroma; **Storage Stability**: Parameters Affecting Storage Stability; **Tannins and Polyphenols**; **Triglycerides**: Structures and Properties

Further Reading

Brouk B (1975) *Plants Consumed by Man*. London: Academic Press.

Considine PEDM and Considine GD (eds) (1982) Pecan and walnut. In: *Foods and Food Production Encyclopaedia*, pp. 1443–1450, 2088–2098. New York: Van Nostrand Reinhold.

Dhar DN (1968) *The Chemistry and Biochemistry of Walnut Trees*. Kanpur: Indian Institute of Technology.

Eckey EW (1954) *Vegetable Fats and Oils*. New York: Reinhold.

Vaughan JG (ed.) (1970) Juglandaceae. In: *The Structure and Utilization of Oilseeds*, pp. 113–116. London: Chapman and Hall.

Young CT (1981) Nuts. In: Grayson M (ed.) *Kirk-Othmer Encyclopaedia of Chemical Technology* 3rd edn, vol. 16, pp. 248–276. New York: John Wiley.

Young CT (1996) Nuts. In: Kroschwitz J and Howe-Grant M (eds) *Kirk-Othmer Encyclopaedia of Chemical Technology*, 4th edn, vol. 17, pp. 544–579. New York: John Wiley.

WASTAGE OF FOOD

C Hunt, University of Huddersfield, Queensgate, Huddersfield, UK

Background

Some food wastage is inevitable during storage, distribution, processing, and cooking. However, avoidable wastage of food at any point in the food chain, from crop losses in harvesting and storage to food left on the plate, could be considered undesirable in moral, economic, and nutritional terms. There have been relatively few studies on this topic, but interest intensifies if commodities become scarcer and/or more expensive. Pollution problems relating to disposal of organic waste, including food and food by-products also need to be addressed.

Discrepancies between total food supplies and actual food consumption of populations have long been recognized but not fully understood. For example, recent estimates in the UK suggest that there is sufficient food and drink available to provide approximately 13.4 MJ (3200 kcal) per head per day.

These estimates relate to edible food available from farms and from imports after allowance had been made for exports, nonfood usage and any changes in year-end stocks. Yet the actual average energy intake in 1998 (National Food Survey) was calculated at 8.8 MJ (2110 kcal) per person per day, about 12% of which derived from food and drink consumed outside the home. The survey included a sample of 5973 fully participating households, and the home food consumption was based on a 7-day food diary of all food entering the household. The above energy intake figure does not include food wastage in the home. However, in further calculations relating energy intakes to dietary reference values (DRVs), the report assumes that 10% of edible food is wasted via plate wastage, spoilage in the kitchen or on the plate, or fed to domestic pets or livestock. This assumed figure is slightly higher than that measured in the study mentioned below.

Thus, on a national scale, there appears to be a 'crude energy gap' of approximately 30–35% comparing food energy available with energy consumed. Possible contributors to this loss in the food chain could be distribution and storage losses, and waste in markets and shops, food factories, hotel and catering outlets, and domestic households.

Definition

Food waste can be defined as 'any potential source of food that has knowingly been discarded or destroyed,' e.g., food discarded in factories, food not

sold in shops or restaurants, kitchen scraps, and plate waste. *Food loss* could be used to refer to food that has inadvertently been destroyed or spoiled, e.g., stored grain eaten by rats, meat spoiled by bacterial growth and nutrients destroyed or going into solution in vegetable canning.

Food waste and loss can be described in terms of *weight, cost*, or *nutritional value* but does not refer to inedible material such as egg shells, potato peelings, and bones.

Household Waste

In a major study in the UK conducted by the Ministry of Agriculture, Fisheries and Foods (1980) food waste was measured in 672 representative households (338 in summer and 334 in winter). The householders collected all potentially edible food wasted in their homes during 1 week and also kept a record of the food other than commercial pet food, which they gave to pets and wild birds. Each food sample received was weighed and its energy content determined calorimetrically.

Significantly more waste food was collected in summer than in winter, equivalent to 9.3 MJ (2220 kcal) and 7.1 MJ (1700 kcal) per household per week, respectively. In terms of energy, cereals, fat and meat wastage predominated, whereas in terms of weight, milk was more important, and fat less so. Significant quantities of otherwise edible food were also given to pets and birds, accounting on average for a further 2.4 MJ (570 kcal) and 3.0 MJ (710 kcal) per household per week in summer and winter, respectively.

The energy content of all food wasted in the home therefore averaged 11.7 MJ (2790 kcal) per household per week in summer and 10.1 MJ (2410 kcal) in winter, equivalent to an average 0.55 MJ (140 kcal) per person per day. This was therefore less than one-quarter of the gap between food supplies and the amount of food thought to be eaten in the UK. When assessed against the expected usage of food in the home, wastage accounted for 6.5% of the energy intake in summer and 5.4% in winter. Wastage was not related to income or geographical area but was greater per person in smaller compared with larger households.

In another study conducted in Cambridge, UK (1985) on 82 families, wastage of edible energy averaged 3.8% (with a further 3.0% of purchases being consumed by visitors).

Domestic food wastage in the USA is probably higher than in the UK. Few data exist, but one study in Arizona, based on measurement of waste food in garbage cans, estimated approximately 10% of the weight of food that entered the home being discarded. However this did not include food given to pets or washed down sinks or through garbage disposal units.

Catering and Hotel Waste

As mentioned previously, this is another area where food wastage occurs. Reliance by consumers on food provided by the caterer varies widely within and between different groups of people and between countries. In the UK, on average, three to four meals per week per person are consumed outside the home. (*See* **Catering:** Catering Systems.)

Catering outlets vary greatly in size ranging from small cafés, snack bars, and restaurants to large industrial canteens and hospitals. Wastage can occur at various stages of the catering operation. Firstly, poor storage conditions for incoming goods can lead to food being discarded, e.g., from temperatures rising too high in refrigeration, cold storage, etc. Meat, meat products, fruits, and vegetables are particularly vulnerable in this respect. Also, excessive humidity can lead to spoilage of dry goods particularly if unpackaged. (*See* **Catering:** Nutritional Implications.)

Secondly, losses can occur in preparation such as in the trimming of meat, vegetables, and fruit. Spillages can occur, and also significant drip-thaw loss of water-soluble protein, vitamins, and minerals can occur in the defrosting of frozen meat.

Cooking can cause further loss, e.g., loss of weight due to evaporative loss, as in roasting of meat, also a reduction in fat and hence energy content in the grilling of meat. Destruction of heat-labile vitamins (B complex and C) inevitably will occur as well in cooking, together with leaching of water-soluble vitamins. These losses can be minimized by good catering practice. Unserved food may also be discarded, and finally, plate wastage may occur, particularly if the meal is unappetizing, e.g., not hot enough or poorly presented. There have been a few studies of wastage in catering. Platt *et al.*, in a study of UK hospitals, drew attention to alarmingly high levels of food waste, with 25–35% of food (by weight) sent from the kitchen being left over in serving dishes and plate waste averaging a further 10%. In a later study of a variety of UK catering establishments, 30% of total edible food energy input was found to be wasted throughout the catering operation in the hospital sector, but was less in other sectors (cafés and snack bars 4.8%; schools 6.5%; place of work 10.2%; restaurants, hotels and pubs 15.5%; and other welfare 17%), giving an overall average of 11.4%. Higher overall wastage occurred with generally poor control procedures, and high plate wastage was also

noticeable with high fat and high energy meals (served particularly in hotels and restaurants).

A recent study on long-stay elderly patients in hospital maintained concerns about the nutritional impact of waste. The data suggested that, although adequate energy and nutrients per patient were provided from the kitchens, service loss and plate waste resulted in actual intakes being inadequate, on average.

In this respect, arguments remain as to the relative merits of plated meals or bulk trolley service, but there have been recent suggestions that bulk trolleys, when operated on an appropriately 'personalized service' basis, can significantly reduce plate waste and improve patient intakes.

Factory Waste

The extent of wastage, however defined, in food-processing factories varies greatly with country and with the particular food products concerned.

Stocks of raw food materials and of food products after processing are held in factories for variable periods and are potentially open to attack by microbes, insects, rodents, etc. The extent of losses will range from considerable to minimal, depending on how tightly storage conditions are controlled and monitoring procedures enacted.

In developing countries, the food-processing industries are generally less well developed and regulated, and the potential for waste is considerable. In developed countries, processing has been increasingly concentrated in very large factory units with enormous throughputs of food. Control and quality-assurance procedures in the latter are relatively sophisticated and waste can be minimized. Inevitably, though, there is still some loss either directly of the food raw materials and food products or of potentially utilizable byproducts. For example, it has been estimated that 1–4% (by weight) of the milk input to a modern dairy plant is wasted and that 2–5% of total carcass protein is lost in the effluent from abattoirs and poultry plants. However, interest in recovery of waste food products and byproducts has been increasing for both economic and ecological reasons, since very large quantities of biodegradable materials are involved, e.g., aqueous 'washings' from the processing of fruit, vegetables and cereals containing starch, sugars and protein, blood from slaughterhouses, and whey from the dairy industry. These wastes have been sprayed directly on to fields, where they can act as fertilizer but can be degraded by soil bacteria in the process, or discharged into the sewage system, or directly into rivers. Excessive use of the last option has been increasingly controlled since the relatively high biochemical oxygen demand of these biodegradable substances can overwhelm the natural capacity of the river to deal with them.

The technology exists for recovery of these materials by chemical and microbiological means either directly for human food or by conversion to animal feedstuffs, but the extent to which this is done is likely to depend largely on economics and environmental legislation.

See also: **Catering**: Catering Systems; Nutritional Implications

Further Reading

Birch GG, Parker KJ and Worgan JT (1976) *Food from Waste*. London: Applied Sciences.

Collison R and Banks GH (1981) Food waste in catering. *Proceedings of the Institute of Food Science and Technology* 14: 181–189.

Collison R and Colwill JS (1987) Food waste in public houses and restaurants and customer attitudes. *International Journal of Hospitality Management* 6(3): 163–167.

Fenton J, Eves A, Kipps M and O'Donnell CC (1995) The nutritional implications of food wastage in continuing care wards for elderly patients with mental-health problems. *Journal of Human Nutrition & Dietetics* 8(4): 239–248.

Jelen P (1983) Reprocessing of whey and other dairy wastes for use as food ingredients. *Food Technology* 37(2): 81–84.

Kelly L (1999) Audit of food wastage: differences between a plated and bulk system of meal provision. *Journal of Human Nutrition & Dietetics* 12(5): 415–424.

Knorr D (1983) Recovery of functional proteins from food processing wastes. *Food Technology* 37(2): 71–76.

MAFF (1999) *National Food Survey, 1998: Annual Report on Food Expenditure, Consumption and Nutrient Intakes*. London: The Stationery Office.

Nelson M, Dyson PA and Paul AA (1985) Family food purchases and home food consumption: comparison of nutrient contents. *British Journal of Nutrition* 54: 373–387.

Wenlock RW, Buss DH and Derry BJ (1980) Household food wastage in Britain. *British Journal of Nutrition* 43: 53–70.

Waste Water *See* **Effluents from Food Processing**: On-Site Processing of Waste; Microbiology of Treatment Processes; Disposal of Waste Water; Composition and Analysis

WATER

Contents

Structures, Properties, and Determination

M M Pierce, University of Nebraska Lincoln, Lincoln, NE, USA

Introduction

In most food systems, water is a major component and influences food structure, appearance, flavor and susceptibility to spoilage.

Shape of the Molecule

The water molecule is made up of two hydrogen atoms with an oxygen central atom. The water molecule is a polar, angular structure (Figure 1). There is a 105° angle between the covalently bonded hydrogen atoms. The hydrogen bonds are covalent σ bonds with a dissociation energy of 4.614×10^2 kJ mol^{-1}.

Pure water can contain as many as 33 chemical variants of the molecule HOH. These include the common isotopes ^{16}O and ^{1}H, as well as ^{17}O, ^{18}O, ^{2}H (deuterium) and ^{3}H (tritium). Water also contains ionic particles including hydronium ions (H_3O^+), hydroxyl ions (OH^-) and their isotopic counterparts. (*See* **pH – Principles and Measurement.**)

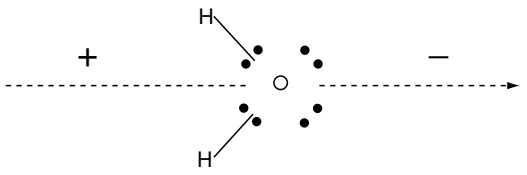

Figure 1 The water molecule. Reproduced from Water: Structure, Properties and Determination, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Hydrogen Bonding

As mentioned, water has a bond angle of 105°, which allows for an approximate tetrahedral structure. The hydrogen atoms are found along two of the axes of the tetrahedron, and the lone pairs of the oxygen atom are located approximately where the other two tetrahedron axes would be. This balance of two hydrogen bond donor and receptor sites allows for three-dimensional hydrogen bonding with a maximum of four other water molecules.

Structure of Ice

It has been shown that the ice crystal structure has a hexagonal symmetry (Figure 2). Three of the hydrogen bonds of the tetrahedron are in the same plane, and the fourth is directly above or below. In the ice crystal, these basal planes are parallel (Figure 3). Therefore, the structure of ice is a series of parallel planes. When ice melts, the structure collapses as hydrogen bonds break. The arrangement becomes more closely packed and, thus, liquid water occupies less space than ice.

Physical Properties

When compared with molecules of a similar molecular weight and composition, water has unusually high melting and boiling points. Water also has large values for surface tension, dielectric constant, heat capacity, and heats of phase transition and a low value for density. Also, the thermal conductivity is large compared with other liquids and nonmetallic solids. These physical properties are influenced by the highly polar nature of the water molecule, which

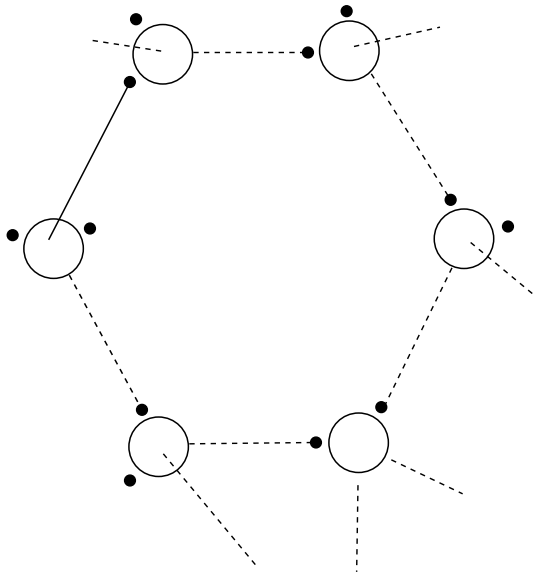

Figure 2 Diagram showing the hexagonal structure of an ice crystal. Reproduced from Water: Structure, Properties and Determination, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), (1993), Academic Press.

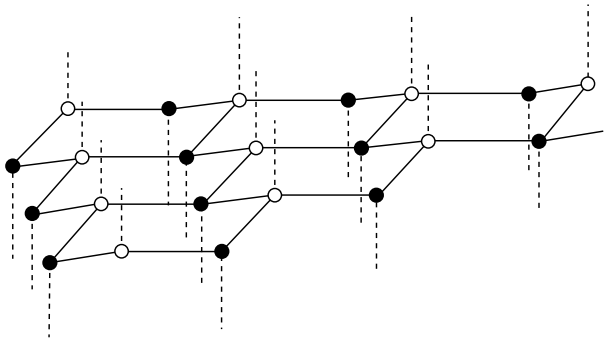

Figure 3 Diagram showing the basal plane of an ice crystal. Reproduced from Water: Structure, Properties and Determination, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), (1993), Academic Press.

allows such phenomena as hydrogen bonding and ionization to occur. Table 1 lists some of the physical properties of water.

Nature of Water in Foods

When associated with food, water can be placed into three categories, free, absorbed, and bound. Free water is the water found within the pores or intergranular spaces of the food. This water retains the usual physical properties of water and often acts as a solvent or dispersing agent for crystalline or colloidal substances. Water absorbed on the surface of macromolecular colloids is termed absorbed water. The macromolecular colloids to which water may be absorbed include starches, pectins, cellulose, and proteins. This absorption is due to Van der Waals' forces and hydrogen bonding. Bound water is the proportion of the water that is in combination with constituents of the food, e.g., water of hydration. This classification system is arbitrary, and there are numerous definitions and variations for water types.

Table 1 Physical properties of water

Molecular weight	18.015	
Phase transitions		
Melting point at 1 atm (°C)	0.00	
Boiling point at 1 atm (°C)	100.00	
Heat of fusion at 0 °C (kJ)	6.012	
Heat of vaporization at 100 °C (kJ)	40.63	
Heat of sublimation at 0 °C (kJ)	50.91	
	0 °C (liquid)	*0 °C (ice)*
Density (kg l^{-1})	0.9998	0.9168
Viscosity (Pa s)	1.787×10^{-3}	—
Surface tension (N m^{-1})	75.6×10^{-3}	—
Heat capacity (J kg^{-1} K^{-1})	4.2177	2.1009
Thermal conductivity (J m^{-1} S^{-1} K^{-1})	5.644×10^{2}	22.40×10^{2}
Thermal diffusivity (m^2 s^{-1})	1.3×10^{-5}	$\sim 1.1 \times 10^{-4}$

Determination in Food

The amount of water in a food is often referred to as the 'moisture content.' This moisture content is important for many reasons. The most important of these is the effect of moisture on the stability and quality of the food substance. This refers to microbiological spoilage as well as undesirable chemical reactions. Moisture is of great economic importance because it is inversely related to the amount of dry material. Other important moisture content issues are processing problems, such as material balance and optimizing processing techniques affected by moisture content. Moisture content is also necessary for calculation of nutritional information, as a means of expressing analytical data on a uniform basis (dry weight basis) and meeting compositional standards and laws. (*See* **Preservation of Food; Spoilage**: Bacterial Spoilage; **Water Activity**: Principles and Measurement; Effect on Food Stability.)

Heating Methods

One of the oldest methods of determining moisture content is heating, and several variations of heating methods exist. They are all common in that the water is driven from the food by the use of a heat source. The loss in weight is measured and assumed to be the moisture content of the food.

The accuracy of these methods is greatly affected by drying temperature, relative humidity and movement

of the air in the chamber, vapor pressure of the chamber (vacuum), sample particle size and quantity, oven construction, position and number of samples in the chamber, the surface of the sample, and the rate of diffusion of water vapor. The drying temperature must be at least the boiling point of water at the given vapor pressure of the chamber. When the chamber is under vacuum, a lower drying temperature must be used. The humidity of the air must be low, and movement of the air is important. Continuous sweeping of the chamber with dry air aids drying and decreases drying time. Using a sample of small particle size increases the surface area and aids drying. A thin layer of sample in a shallow aluminum pan also increases the rate of drying. Allowing liquids to evaporate on a water bath before oven drying is the preferred method. To allow for uniform drying of several samples, pans need to be evenly spaced in the center of the oven 2–5 cm from the walls and 5–8 cm from the door.

There are two common types of ovens, air ovens and vacuum ovens. Air ovens include both convection and forced-air types. Vacuum ovens allow for complete drying with less change to the organic composition of the product.

A less traditional heating method is infrared drying. This type of drying reduces the drying time to 12–33% of the conventional drying time. Infrared drying consists of a 250–500 W lamp, which heats to 2000–2500 K. The lamp is placed above the weighed sample at a distance of about 10 cm. The sample pan may rest on a torsion balance with an indicator scale reading in grams and/or percent moisture. This normally has a drying time of 10–40 min. Microwave ovens can also be used for drying and allow very short drying times.

Distillation Methods

Distillation is almost as old a method as drying. There are two types of distillation apparatus. The first uses a high-boiling-point water-immiscible liquid such as a mineral oil. The sample is suspended in this oil and heated to a temperature high enough to allow the water to be vaporized. As the water is distilled off, it is condensed and collected in a vessel and measured. The second type of distillation involves the use of a solvent that is immiscible with water. Again, the sample is suspended, and the solvent and water are codistilled and collected in a measuring device or moisture trap. The two liquids separate, and the water layer is measured. Often, a reflux system is used to allow recycling of the solvent. **Figure 4** shows a typical distillation system, and **Figure 5** shows various types of moisture traps.

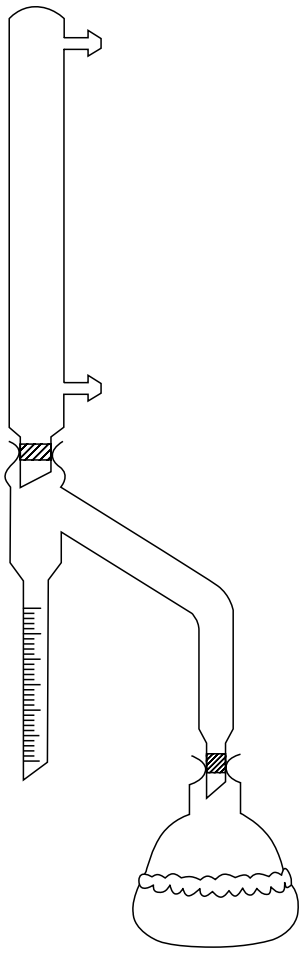

Figure 4 A typical distillation system for determining the water content in food. Reproduced from Water: Structure, Properties and Determination, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Chemical Methods

The Karl–Fischer titration is a good method for samples that are low in moisture, high in sugar, and high in reducing sugar and protein, or samples that give erratic results. The Karl–Fischer method is based upon the reduction of iodine by sulfur dioxide in the presence of water. Methanol and pyridine are used to dissolve the iodine and sulfur dioxide. The reaction occurs in two steps (eqn (1)).

$$C_5H_5N \cdot I_2 + C_5H_5N \cdot SO_2 + C_5H_5N + H_2O \rightarrow 2C_5H_5N \cdot HI + C_5H_5N \cdot SO_3 \quad (1)$$

$$C_5H_5N \cdot OSO_3 + CH_3OH \rightarrow C_5H_5N(OH)SO_4CH_3. \quad (1b)$$

Normally an excess of methanol, pyridine, and sulfur dioxide is used, the strength being dependent on the

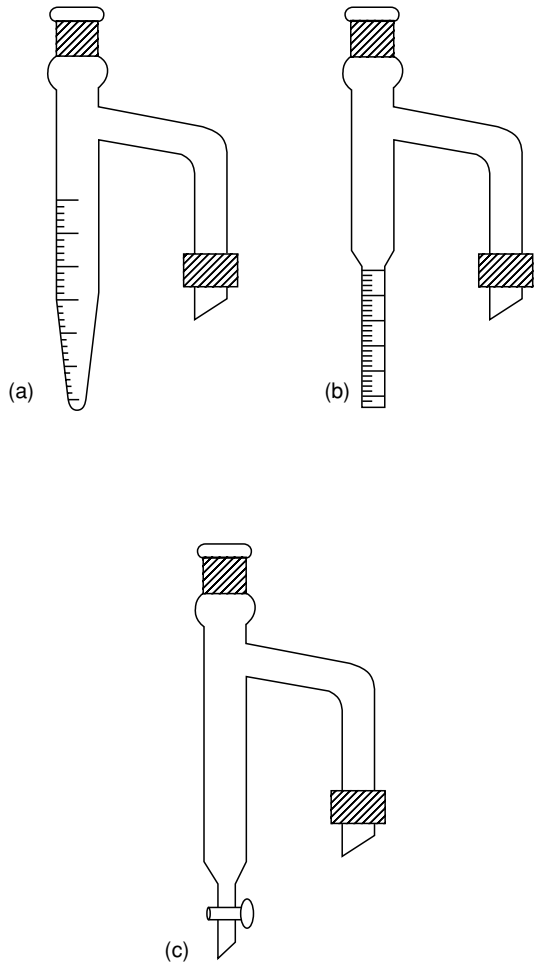

Figure 5 Moisture traps for distillation methods of water determination: (a) Stark and Dean, (b) Bidwell and Sterling, and (c) Barrett. Reproduced from Water: Structure, Properties and Determination, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), (1993), Academic Press.

concentration of iodine. Karl–Fischer reagents can be obtained commercially as two solutions: (1) iodine in methanol and (2) sulfur dioxide in pyridine.

The titration can be carried out in one of two ways. The first requires that the reagents be mixed just before use and the sample dispersed in a solvent (e.g., methanol). The dispersed sample is then titrated with the four-component reagent until a yellow–mahogany brown end point is reached. In the second method, the sample is dispersed in a mixture of sulfur dioxide, pyridine, and methanol. This is then titrated with iodine in methanol. In order to enhance the end point, a few drops of methylene blue may be added, giving a green end point. Alternatively, a photometric or electrochemical method can be used instead of a visual method.

Instrumental Methods

Infrared determination Infrared spectrophotometry is based on absorption of wavelengths characteristic of molecular vibrations. In water, the fundamental vibrational modes are 3000 and 6100 nm with the first overtone of the OH stretch at 1400 nm. (*See* **Spectroscopy**: Infrared and Raman.)

Near-infrared determination This type of spectrophotometry is similar in nature to infrared spectrophotometry. The wavelengths of interest are 1940 nm, as the primary OH bend and stretch, and 1450 nm, from the first overtone of the OH stretch. (*See* **Spectroscopy**: Near-infrared.)

Gas chromatographic determination Moisture can be determined by gas chromatography using a thermal conductivity detector, provided that the water can be extracted using a solvent such as methanol/butanol and that separation of water from other volatiles is possible. (*See* **Chromatography**: Gas Chromatography.)

Nuclear magnetic resonance (NMR) determination NMR determination uses a magnetic field of constant frequency that is applied to the sample, and the strength of the field is varied. The peak-to-peak amplitude of the signals is measured, and moisture is determined by using a standard curve. This method, however, is hydrogen-specific, not water-specific, and may be useful for materials only with a very low solids content. (*See* **Spectroscopy**: Nuclear Magnetic Resonance.)

Electrical determinations Several instruments exist that use conductivity as a method to measure moisture. Most of these instruments use a modified Wheatstone bridge. In these instruments, an electrical current is applied to a food, and the resistance of the food is measured and related to moisture content.

Another electrical method is capacitance. This type of instrument relies on the fact that the dielectric constant of water is higher than that of most other liquids. The sample is placed between two plates, with alternating opposite charge at a fixed frequency, resulting in an alternating current. Polar substances with permanent dipoles orient themselves in this alternating field, increasing the charge of the plates.

Problems of Determination

All of the methods discussed have drawbacks, and while some are excellent methods for one product, they may yield poor results for others. In drying, the most common source of error is the formation of a crust or case hardening. Lowering the oven temperature and increasing the vacuum may help. In the case

of some liquids that are high in sugar, mixing with sand may help. A two-stage drying procedure can also be used, where the temperature starts low and is increased very slowly.

Distillation methods also have many difficulties associated with them. These include adherence of moisture to glass and improper drainage of water into the measuring device, difficulties in reading the meniscus, and low precision of the measuring device. Also, water-soluble components may be distilled, causing erroneous results.

The main difficulty of the Karl–Fischer titration is incomplete extraction of water from the sample. Also, ascorbic acid, a common additive, is an interferent and has an additive effect. Aldehydes and ketones release water when mixed with methanol, causing a fading endpoint.

Thus, although chemical and instrumental methods are, in general, very fast, they have a narrow range of products with which they can be used.

See also: **Chromatography**: Gas Chromatography; **Preservation of Food**; **Spectroscopy**: Infrared and Raman; Near-infrared; Nuclear Magnetic Resonance; **Spoilage**: Bacterial Spoilage; **Water Activity**: Principles and Measurement; Effect on Food Stability

Further Reading

AOAC (1990) *Official Methods of Analysis: Association of Official Analytical Chemists*, 15th edn. Arlington, VA: Association of Official Analytical Chemists.

Aurand WA, Woods AE and Wells MR (1987) *Food Composition and Analysis*. New York: AVI.

Fennema OR (1985) *Food Chemistry*. New York: Marcel Dekker.

Osborne BG and Fearn T (1986) *Near Infrared Spectroscopy in Food Analysis*. Harlow: Longman.

Pomeranz Y and Meloan CE (1987) *Food Analysis: Theory and Practice*, 2nd edn. New York: AVI.

Sisler HH, VanderWerf CA and Davidson AW (1967) *College Chemistry*, 3rd edn. New York: The Mallcom Company.

Physiology

J R Robinson, Dunedin, New Zealand

Introduction

Body water cannot be poured out to measure its volume; it must be determined indirectly. A good indirect method is to introduce a known quantity of a harmless solute and measure its concentration in the blood plasma after mixing is complete. The results depend upon the solute chosen. Very diffusible substances such as urea, ethanol, or isotopic forms of water, give large values that correspond to the total amount of water in the body, 35–45 l, or 50–70% of body weight. Proteins or other colloids that cannot leave the blood give the volume of plasma, around 3 l, or 4–5% of body weight. Substances of intermediate molecular size that cannot penetrate into cells give estimates ranging between 12 and 20 l, or 16–30% of body weight. (*See* **Body Composition**.)

Provided that these substances can penetrate into dense connective tissues and tendons, this intermediate estimate should give the sum of the extracellular fluid outside the blood vessels and the plasma inside them. The results are consistent with the obvious anatomical fact that the water of the body is not one homogeneous phase. The most important subdivision is into two major fluids: one is intracellular, discontinuous, in 10^{14} or so separate tiny portions, each surrounded by a cell membrane; the other is extracellular, the continuous medium in which the cells live, and which provides the medium for their communication with each other, and, through the circulating blood, with the external environment across interfaces at the lungs and the alimentary tract.

These two major fluids are as distinct in composition as in anatomical location (**Table 1**). The solutes in the intracellular fluids (ICFs) are potassium salts, and the anions are mostly organic, chiefly phosphoric esters and proteins, with only a little chloride. The extracellular fluid (ECF) is more like a dilute sea water; its chief solutes are sodium and chloride, with small but important concentrations of potassium, calcium and bicarbonate, together with organic nutrients such as glucose and amino acids, and waste products such as urea.

Table 1 Composition (mmol per kg of water) of typical intracellular and extracellular fluids

	Muscle ICF	*Plasma ECF*
Potassium (K^+)	150	5
Sodium (Na^+)	10	150
Magnesium (Mg^{2+})	10	1
Calcium (Ca^{2+})	10	2
Organic phosphate (Org P^-)	130	5
Bicarbonate (HCO_3^-)	10	27
Chloride (Cl^-)	5	110
Phosphate ($H_2PO_4^-$)		2
Protein ($Prot^{17-}$)	2	1
pH	7.1	7.4
Osmolality (mosmol per kg of water)	285	285

Taken from Table 1.1 in Bray JJ, Cragg PA, Macknight ADC, Mills RG and Taylor DW (1989) *Lecture Notes on Human Physiology*, 2nd edn. Oxford: Blackwell Scientific with permission.

The membranes surrounding the cells are of remarkably low and highly selective permeability, but they are not totally impermeable to sodium, potassium, and other ions. Consequently, the characteristic differences in composition have to be maintained actively, and a major proportion of 'resting' metabolism is required to provide energy to pump out sodium that continually diffuses into cells, and to recover potassium that escapes or is lost through activities such as muscular contraction or conduction of nerve impulses. The membranes are permeable to water, so that the fluids on either side are in osmotic equilibrium. (*See* **Potassium**: Physiology; **Sodium**: Physiology.)

Role as Transport Vehicle for Nutrients and Metabolites

Within the body, nutrients and metabolites are in solution, and the transport system is based on water carriage. That part of the ECF which is the blood plasma is contained in the arteries, capillaries, and veins, and circulated rapidly; even at rest, the heart pumps about 5 $l\,min^{-1}$ from the aorta. The lymph, which is also in vessels, circulates much more slowly in returning escaped plasma proteins from tissue spaces to the blood. The flowing plasma carries nutrients and metabolites close to cells, but cannot quite reach them. Transport of consumables to cells, and products away, depends on diffusion through the capillary walls and across thin layers of interstitial fluid between the capillaries and cells in the tissues. Diffusion in water is slow over distances greater than 0.1 mm, but at distances of the order of cell diameters, it is remarkably rapid, 90% equilibration taking less than 0.05 s at 10 μm, even at room temperature. Between adjacent cells 1 μm apart, it would take less than 5×10^{-4} s, across a synapse, 0.1 μm wide, $< 5 \times 10^{-6}$ s. Thus, the circulating blood provides long-distance transport of oxygen from the lungs, metabolites from the liver, and nutrients from the gut to the tissues; it also carries carbon dioxide to the lungs and nonvolatile waste products to the kidneys for disposal. Diffusion provides rapid transport in the ICFs between organelles within the cells, and rapid exchanges with the ECF in which they live. Both these mechanisms of transport depend on the water in which the transported materials are dissolved. The blood also carries 'messengers' such as carbon dioxide from active tissues to the respiratory center, and hormones from endocrine glands to cells all over the body, which makes transport by water extremely important for the coordinated function of a vast population of separate cells as one organism.

Protection and Lubrication

There are some smaller collections of fluid that are not inside cells and yet cannot exchange freely with the main mass of ECF by diffusion across highly permeable capillary membranes. Examples are the vitreous and aqueous humors in the chambers of the eye, the cerebrospinal fluid that surrounds the brain and the spinal cord, and the synovial fluid in the capsules of joints. These are by no means ICFs, but they are secreted by layers of specialized cells (e.g., in the ciliary body, choroid plexus, or synovial membranes) which surround them and separate them from the main body of ECF. They may be called transcellular fluids, because the only approach to them by diffusion from the continuous ECF is across cellular membranes with their specialized and restricted permeability. These specialized fluids serve to maintain the size and shape of the eyeball, to protect the brain from trauma due to sudden movements by reducing its effective weight by flotation, and to lubricate the joints.

The cells are tiny aquatic organisms, and the ECF is the pond in which the live. It has been aptly called a middleman fluid because it accommodates exchanges of matter and energy between cells and other cells in remote parts of the body as well as between cells inside the body and the external environment from which they eventually derive their nutrients and the oxygen they need to utilize these. The famous French physiologist, Claude Bernard, died in 1878 correcting the proofs of the book in which he characterized the ECF as an internal environment within which the cells live a sheltered life, protected from the vicissitudes of an external world that is variable and often hostile. Bernard enunciated the principle that the constant properties of this internal environment are a necessary condition for the free, independent life of higher organisms. Indeed, a substantial part of the subject matter of physiology is concerned with homeostatic mechanisms that operate to stabilize the chemical and physical properties of the internal environment.

Role in Body Temperature Regulation

The rates of the chemical reactions catalyzed by enzymes have substantial temperature coefficients, so that proper coordination of the activities of the body's many enzymes depends on maintaining a fairly constant temperature despite the variable output of heat by active tissues. The large thermal capacity of water minimizes local increases in temperature at sites where heat is evolved; flowing blood carries heat away from active sites far more rapidly than is possible by conduction, so that the extra heat is

shared with a larger volume of blood and other tissues. Heat produced at active sites within the body must ultimately be conveyed to an external surface before it can be lost; the high heat capacity of water allows this to be achieved with only small changes in temperature of the blood. When the surroundings of the body are warmer than the skin, heat cannot be lost at all by conduction and convection; but the large latent heat of vaporization of water enables large amounts of heat to be lost to hotter surroundings without undue losses of body fluids. The skin is cooled by the evaporation of sweat, a special dilute solution secreted by sweat glands that are activated when the temperature of the blood bathing receptors in the brain rises. Maximal sweating may exceed $1\,l\,h^{-1}$, and people working or exercising in hot environments may lose 15 l per day; but if the latent heat of vaporization were to be 10 times smaller (as were those of some liquids Henderson (1913) took for comparison), 10 times as much would have to be lost, up to three or four times the total amount in the body in the course of a single day! Conversely, in circumstances of extreme cold, a modest flow of blood suffices to bring heat from within the body to combat chilling of the extremities, and the large latent heat of fusion of water protects these extremities from quickly freezing.

Other Inevitable Losses of Body Water

The secretion of sweat is almost entirely dictated by the need to prevent overheating of the brain; sweating cannot be turned off to save water. There is an equally unavoidable loss of about 0.5 l every day in the form of vapor from moist skin between the ducts of the sweat glands ('insensible' because, unlike liquid sweat, it is not obvious). The skin must remain moist to retain its flexibility and avoid cracking (with the risk of bacterial infection) with bending and stretching. The alveolar lining of the lungs must also be moist to permit deformation and to allow oxygen and carbon dioxide to exchange in solution between alveolar gas and blood. Since the expired air is saturated with water vapor at the temperature of the respiratory tract, while the inspired air is drier, an inevitable loss of about 0.5 l of water by evaporation accompanies breathing at sea level. The loss is proportional to the volume of expired gas; mountaineers breathing large volumes of thin air at great heights may lose as much as 2 l in a day. This inevitable loss of water vapor takes with it the latent heat of evaporation, so that they risk severe impairment of their physical condition from chilling as well as dehydration, especially if their exertions lead to additional losses in sweat.

Losses in feces are unimportant unless there is diarrhea. However, one very important function of body water that can only be carried out by expending it is the removal of soluble waste products in the urine. About 600 mmol must be excreted every day, and this requires 0.5 l of maximally concentrated urine, or a minimum of 2 l if, because of impaired function, the kidneys cannot make the urine more concentrated than the blood.

Hence, everyday living involves unavoidable daily losses of 1.5–2 l of water; since water is an essential part of the living machinery of the cells, and effective transport around the body requires adequate volumes of blood to circulate and of ECF to bathe the cells, these losses must be replaced. A small amount of water, about 300 ml, is actually generated each day in the cells by metabolism; solid food (which contains as much, or, if vegetable, more water than our own tissues) provides about a liter, and the rest must be taken in as beverages or soups.

Maintaining the Water Balance of the Body

When water is lost without solute and not replaced, the osmotic pressure of all the body fluids increases, and their combined volume shrinks. Osmoreceptors in the supraoptic nucleus of the hypothalamus at the base of the brain respond by evoking the sensation of thirst and by releasing into the blood an antidiuretic hormone (ADH), which causes the kidneys to conserve water by making the urine more concentrated (if it is not already maximally concentrated). The subject is encouraged to drink more watery fluid, and the kidneys retain water in the body. If the subject drinks more water than is required, the body fluids become diluted, thirst fades, the release of ADH ceases, and the kidneys rapidly excrete excess water in a copious flow of dilute urine. (*See* **Renal Function and Disorders**: Kidney: Structure and Function.)

Not only do the kidneys act with thirst to preserve the total amount of water in the body, but they also regulate the distribution between cells and ECFs. As stated above, the ICFs are primarily solutions of potassium salts, and the ECFs of sodium salts. Neglecting the other ions, human ECF can be likened to a $0.154\,mol\,l^{-1}$ (0.9%) solution of sodium chloride. Hence, for every 9 g of sodium chloride, there is a liter of ECF. The kidneys are controlled by a number of factors, including the adrenal hormone, aldosterone, so that they retain more salt in the body when the volume of blood is diminished. Metabolic activity keeps the extra sodium out of the cells, in the ECF, so that its volume, and that of the blood in the vessels, is restored. (*See* **Hormones**: Adrenal Hormones.)

The volume of ICF is fixed by the prevailing osmotic pressure and the amount of solute that the cells keep within them by metabolic activity. The kidney, controlling the excretion of water under the influence of ADH, fixes the osmotic pressure in the body, thus setting the stage for the cells to maintain their appropriate volume of ICF. Fixing the osmotic pressure throughout the body also determines the amount of water to go with each unit of sodium in the ECF. Hence, with osmotic pressure and the volume of ICF stabilized by controlling the excretion of water, the kidneys can also regulate the volume of ECF by excreting or retaining salt.

See also: **Body Composition**; **Hormones**: Adrenal Hormones; **Potassium**: Physiology; **Renal Function and Disorders**: Kidney: Structure and Function; **Sodium**: Physiology

Further Reading

Bray JJ, Cragg PA, Macknight ADC, Mills RG and Taylor DW (1989) *Lecture Notes on Human Physiology*, 2nd edn. Oxford: Blackwell Scientific.

Henderson LJ (1913) *The Fitness of the Environment*. New York: Macmillan.

Robinson JR (1970) Water and life. *World Reviews of Nutrition and Dietetics* 12: 172–207.

Robinson JR (1978) Water in the animal body. In: McIntyre AK (ed.) *Water: Plants, Planets and People*, pp. 60–70.Canberra: Australian Academy of Science.

Robinson JR (1988) *Reflections on Renal Function*, 2nd edn. Oxford: Blackwell Scientific.

WATER ACTIVITY

Contents
Principles and Measurement
Effect on Food Stability

Principles and Measurement

Y H Roos, University College Cork, Cork, Ireland

Introduction

Water is the main component of most food materials. The control of water content is often used to increase food stability, e.g., by dehydration. It is also known that high concentrations of sugars and salts may increase food stability by affecting the availability of water to microbial growth, although the water content may remain high. It is obvious that the water contents of different foods vary substantially, and water content as such cannot be used as a measure of water availability or food stability. The water availability for microbial growth is a function of the vapor pressure of water in a food, which has led to the development of the *water activity* concept. It has been shown that food storage stability and microbial safety, i.e., rates of chemical and enzymatic reactions and microbial growth, can be controlled by water activity in addition to other factors, such as pH and temperature. Therefore, the development of methods to control water activity of foods and water activity measurement has received much attention in the food industry. Furthermore, reference values for water activity are used in legislation and regulations to insure food safety. This article discusses the basic concepts of water activity, its definition, temperature dependence, measurement, relationship with water content, and mathematical approaches to predict water activity in food systems.

Water Activity

The water activity, a_w, in a solution or a food is derived from its chemical potential, μ_w. The chemical potential, μ, of a pure substance is given by its molar Gibbs free energy, G_m. The Gibbs free energy is pressure-dependent, and it can be shown that the Gibbs free energy of a perfect gas at a pressure, p, can be related to its value, G^*, at a standard pressure, p^*. Then, the Gibbs free energy will be given by eqn (1).

$$G = G^* + nRT \ln \frac{p}{p^*}. \qquad (1)$$

Assuming perfect behavior and taking into account that for water vapor $\mu_{w0} = G_w/n$, eqn (2), where μ_w^* is

the chemical potential at the standard pressure, p^*_w, is obtained.

$$\mu_{w0} = \mu^*_w + RT \ln \frac{p_w}{p^*_w}. \quad (2)$$

The vapor pressure of pure water can be taken as $p_{w0} = p_w/p^*_w$. Thus, the chemical potential is expressed by eqn (3). If water exists in a solution, its chemical potential will be μ_w and vapor pressure $p_{w(liquid)}$. At equilibrium, the chemical potential in the liquid and vapor states is equal, and, therefore, the chemical potential of water in a solution, μ_w, is defined by eqn (4). Eqns (3) and (4) can be combined to give eqn (5) and the chemical potential of liquid water in the solution. The activity, in the case of water, the *water activity*, a_w, is defined by eqn (6).

$$\mu_{w0} = \mu^* + RT \ln p_{w0} \quad (3)$$

$$\mu_w = \mu^* + RT \ln p_w \quad (4)$$

$$\mu_w = \mu_{w0} + RT \ln \frac{p_w}{p_{w0}} \quad (5)$$

$$a_w = \frac{p_w}{p_{w0}}. \quad (6)$$

As defined by the preceding equations, the chemical potential, vapor pressures, and water activity are all temperature-dependent. Therefore, the water activity is a measure of water availability at equilibrium at a constant temperature, e.g., room temperature.

The relative humidity, RH, of an atmosphere is defined as the ratio of the quantity of water present to the saturation quantity at the temperature of the atmosphere, and it is also the ratio of the vapor pressure of water present to that of saturated water vapor at the same temperature, i.e., the relative vapor pressure (RVP), as given by eqn (7).

$$RH = \frac{p_w}{p_{w0}} \times 100\%. \quad (7)$$

If a solution or a food is placed in an atmosphere, it will lose or gain water until equilibrium is reached and the chemical potential and, therefore, the water activity in the atmosphere and the solution or food have become equal. This has a great importance in a_w measurements, because a_w is often derived from the equilibrium relative humidity (ERH) of the headspace of a food system.

Raoult's Law

The vapor pressure of ideal solutions is related to the mole fraction of water, as defined by Raoult's law. Raoult's law is based on the observation that the partial vapor pressure of each component in a solution is a function of its mole fraction and the vapor pressure of a pure liquid at the same temperature. In the case of water, assuming that the solution behaves as an ideal solution, Raoult's law is given by eqn (8). However, Raoult's law assumes linearity, it applies for ideal solutions, and often the vapor pressure deviates from ideal behavior with increasing solute concentration. For real solutions, an activity coefficient is defined, and the water activity is related to the mole fraction of water, x_w, by eqn (9), where γ_w is the activity coefficient of water. Therefore, water activity may also be considered as the 'effective' mole fraction of water.

$$p = x_w p_{w0} \quad (8)$$

$$a_w = \frac{p_w}{p_{w0}} = \gamma_w x_w. \quad (9)$$

Raoult's law is important in understanding the boiling temperature elevation and freezing temperature depression of solutions. The freezing temperature depression may also be measured to derive water activity.

Water Activity Measurement

The water activity of simple, ideal solutions can be derived from Raoult's law by applying eqn (8) and real solutions from eqn (9). Unfortunately, food systems are very complex, the water activity is difficult to estimate, and it must be measured for most foods. A simple, but rough, water activity estimation is based on placing samples in several environments (chambers) with various relative humidities and estimating the water activity from a plot showing weight gain or loss against RH.

The water activity of food materials can be measured using various systems to observe the vapor pressure of water in a food sample or by measuring the relative humidity of air in equilibrium with a food sample. The measurement of the vapor pressure of water can be carried out using a manometer and other relatively simple laboratory equipment. Simple, commercial instruments may apply hair hygrometry. Such devices are not very accurate and require that appropriate time is allowed for the sample material in a container to equilibrate with the measuring unit. Various other commercial hygrometers have been applied in water activity measurements. For these measurements, an appropriate amount of sample material often needs to be placed in a closed container or jar at a constant temperature. The sensor of the instrument is placed in the headspace, and the relative humidity at equilibrium is recorded and expressed as water activity.

Several commercial water activity measurement devices have been developed for rapid and accurate water activity determinations. These devices use either a chilled-mirror dew-point technology or measure relative humidity with sensors that detect electric resistance or capacitance. The dew-point instruments provide relatively rapid and accurate water activity measurements. The main principle of such instruments is based on cooling air at equilibrium water vapor pressure with a food sample without changing the water content and determination of the temperature at which water in the air saturates at a chilled-mirror surface. The water activity of the sample is given by the ratio of the saturation vapor pressure at the dew-point temperature and the saturation vapor pressure at the sample temperature. Instruments using relative humidity sensors require equilibration with the sample at the same temperature in a closed container or jar, as described earlier for hygrometer measurements.

Instrument Calibration

Water activity measurements require adequate and accurate instrument calibration. This is usually carried out with the help of saturated salt solutions or other substances with known water vapor pressures at a given temperature. Different instruments may require different specific calibration procedures, but most instruments should be checked for correct water activity readings. Saturated salt solutions are probably the most common means of producing atmospheres with known relative humidities or water vapor pressures. Saturated salt solutions can be prepared to give a wide range of water activities in a number of temperatures. Although the solutions are relatively easy to prepare, accurate, and safe, appropriate procedures and safety precautions should be followed. The water activities of common saturated salt solutions are listed in Table 1. Other materials, which in various concentrations in solutions with water can be used to adjust water activity, include glycerol and sulfuric acid (H_2SO_4).

Water Sorption

It is often beneficial to establish water sorption isotherms, which show the relationship between water content and water activity for particular food systems. Sorption isotherms can be obtained experimentally by using dehydrated, for example freeze-dried, samples of food solids. The samples should be initially stored over a 'zero' relative humidity in a hermetic container with $CaSO_4$ or P_2O_5. These samples can be used to obtain the 'anhydrous' weight, and the samples are then placed in hermetic containers, preferably evacuated desiccators at a constant temperature over saturated salt solutions with known water activities. The samples in each container are weighed at intervals until subsequent weighings level off, and the weight gain can be considered as equilibrium water content corresponding

Table 1 Water activities of selected saturated salt solutions at various temperatures

Salt	*Water activity at temperature (°C)*									
	5	*10*	*15*	*20*	*25*	*30*	*35*	*40*	*45*	*50*
CsF	0.055	0.049	0.043	0.038	0.034	0.030	0.027	0.024	0.022	0.021
LiBr	0.074	0.071	0.069	0.066	0.064	0.062	0.060	0.058	0.057	0.055
ZnBr	0.087	0.085	0.082	0.079	0.078	0.076	0.076	0.075	0.076	0.077
LiCl	0.113	0.113	0.113	0.113	0.113	0.113	0.113	0.112	0.112	0.111
CaBr		0.216	0.202	0.185	0.165					
LiI	0.217	0.206	0.196	0.186	0.176	0.166	0.156	0.146	0.135	0.124
CH_3COOK		0.234	0.234	0.231	0.225	0.216				
$MgCl_2$	0.336	0.335	0.333	0.331	0.328	0.324	0.321	0.316	0.311	0.305
NaI	0.424	0.418	0.409	0.397	0.382	0.362	0.347	0.329	0.310	0.292
K_2CO_3	0.431	0.431	0.432	0.432	0.432	0.432				
$MgNO_3$	0.589	0.574	0.559	0.544	0.529	0.514	0.499	0.482	0.469	0.454
NaBr	0.635	0.622	0.607	0.591	0.576	0.560	0.546	0.532	0.520	0.509
$CoCl_2$					0.649	0.618	0.586	0.555	0.526	0.500
KI	0.733	0.721	0.710	0.699	0.689	0.679	0.670	0.661	0.653	0.645
$NaNO_2$	0.786	0.775	0.765	0.754	0.743	0.731	0.721	0.710	0.700	0.690
NaCl	0.757	0.757	0.756	0.755	0.753	0.751	0.749	0.747	0.745	0.744
KBr	0.851	0.838	0.826	0.817	0.809	0.803	0.798	0.794	0.792	0.790
KCl	0.877	0.868	0.859	0.851	0.843	0.836	0.830	0.823	0.817	0.812
KNO_2	0.963	0.960	0.954	0.946	0.936	0.923	0.908	0.890	0.870	0.848
K_2SO_4	0.985	0.982	0.979	0.976	0.973	0.970	0.967	0.964	0.961	0.958

to the water activity of the particular salt solution. The equilibration times may vary depending on the container design, water activity, and sample material among other factors. However, equilibration in evacuated desiccators is rapid and often occurs in a few hours and should not take more than a week. Equilibration in nonevacuated containers may take substantially longer and is more subject to errors caused by nonequilibrium conditions and water transfer between the container and external atmosphere.

Food materials exhibit two basic types of sorption isotherms reflecting their either amorphous or crystalline state. Most foods are noncrystalline or partially crystalline systems and have a sigmoid shaped sorption isotherm. Crystalline materials, e.g., crystalline sugars, show a low water sorption until solubilization, and other changes occurring at high water activities result in a substantial increase in sorbed water (Figure 1). Food materials also have a hysteresis between desorption and adsorption isotherms, i.e., the water content in desorption at the same water activity is higher than in adsorption (Figure 2). The hysteresis results most likely from retention of water in capillaries and structural differences. Sorption isotherms are extremely useful in practical applications, as only water content needs to be known to derive water activity. Furthermore, changes in storage relative humidity of low moisture foods can be related to corresponding changes in water content.

Water activity and water sorption isotherms are temperature-dependent. The temperature dependence of water activity follows the well-known Clausius–Clapeyron relationship (eqn (10)), where a_{w1} and a_{w2} refer to water activity at absolute temperatures T_1 and T_2, respectively, Q_s is the heat of sorption, and R is the gas constant ($8.3144\,\mathrm{J\,K^{-1}\,mol^{-1}}$). A schematic representation of the temperature dependence of water activity is shown in Figure 3.

$$\ln\frac{a_{w2}}{a_{w1}} = \frac{Q_s}{R}\left(\frac{1}{T_1} - \frac{1}{T_2}\right). \qquad (10)$$

In general, the water activity of a food with a constant water content increases with increasing temperature, while the water sorption decreases with increasing temperature at the same water activity. This is observed from sorption isotherms determined for the same food at different temperatures (Figure 4). Understanding the temperature dependence of water sorption is extremely important, for example, in estimating the effects of temperature fluctuations on water activity changes inside a food package. The

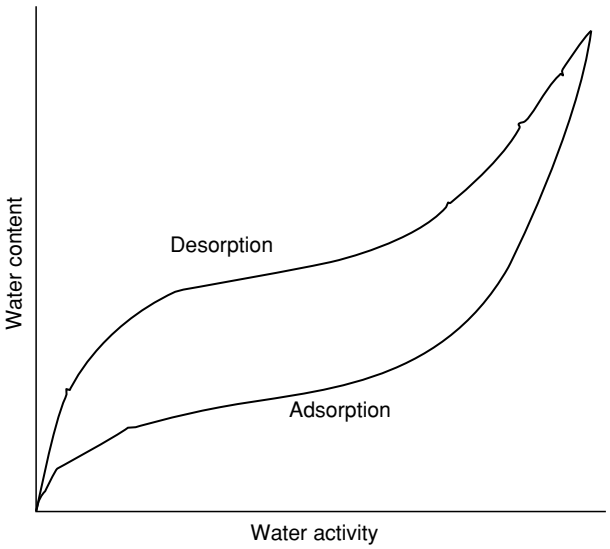

Figure 2 Hysteresis between water desorption and adsorption typical of foods.

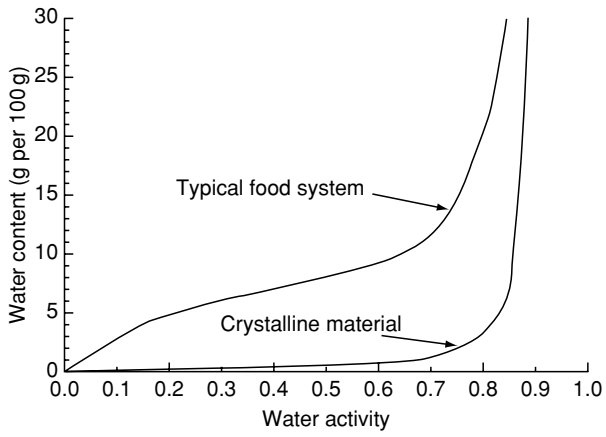

Figure 1 Hypothetical sorption isotherms typical of food systems and crystalline materials, such as sugar crystals.

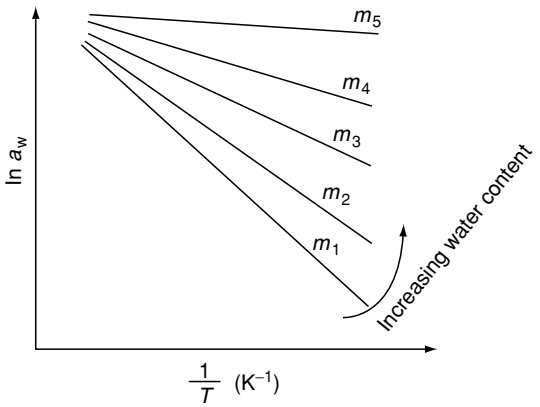

Figure 3 Temperature and water content dependence of water activity. The temperature dependence of water activity increases with decreasing water content.

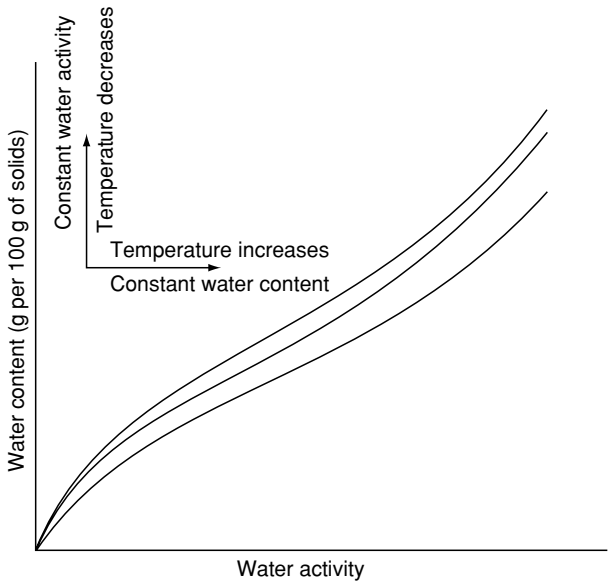

Figure 4 Schematic representation of water sorption isotherms determined at different temperatures.

temperature change under normal storage conditions may result in a water activity shift exceeding 0.1 a_w.

Water Sorption Models

A number of mathematical models have been used to fit experimental water sorption data of food materials. The most common models are the Brunauer–Emmett–Teller (BET) model (eqn (11)) and the extended Guggenheim–Anderson–De Boer (GAB) model (eqn (12)). These models originate from studies of sorption of gases on solid surfaces. The BET model relates the water content (m) to water activity using two parameters, m_0 and C, defined as the monolayer value and a constant related to the excess enthalpy of sorption, respectively. The monolayer value gives the amount of water sorbed initially to first available sorption sites, and it is often found to correspond to the water content of low moisture food systems with improved stability. The BET sorption model can be written to the form of eqn (12), which gives a linear relationship between $a_w/(1-a_w m)$ and a_w. Thus, experimental water activity and water content values obtained from sorption studies can be used to derive the constants m_0 and C. The BET equation, however, fits experimental data only between 0 and 0.5a_w.

$$\frac{m}{m_0} = \frac{Ca_w}{(1-a_w)[1+(C-1)a_w]} \tag{11}$$

$$\frac{a_w}{(1-a_w)m} = \frac{1}{m_0 C} + \frac{C-1}{m_0 C}a_w \tag{12}$$

$$\frac{m}{m_0} = \frac{KC'a_w}{(1-Ka_w)[1+(C'-1)Ka_w]} \tag{13}$$

$$\frac{a_w}{m} = \frac{K}{m_0}\left(\frac{1}{C'}-1\right)a_w^2 + \frac{1}{m_0}\left(1-\frac{2}{C'}\right)a_w + \frac{1}{m_0 KC'}. \tag{14}$$

The GAB model (eqn (13)) has been shown to fit experimental sorption data for most food materials over the entire a_w range. Experimental data may also be used to derive the three parameters, the constants C' and K, and the monolayer value, m_0, of the GAB model. This can be carried out, for example, by using nonlinear regression or rearranging the equation to a form of a second-order polynomial and plotting experimental values of a_w/m against a_w, as suggested by eqn (14). However, the BET and GAB monolayer values are not equal, and neither of these values can be considered as a water content giving maximum stability in a low-moisture state.

Water Activity in Food System

Water is the main component of most fresh foods, and their water activities are high and close to 1. The water activity has increasing importance in concentrated foods, dehydrated foods or systems that contain salts, sugars, and other compounds reducing water activity to provide stability, for example highly salted meats and fish or confectionery and candies. Therefore, it is often of great interest in food product development to estimate effects of food composition and recipe components on end-product water activity. It is also important in packaging of mixtures of different foods, such as dry ingredients, to understand how water activity in a sealed package with several components may change owing to differences in component water activities. Furthermore, water activity in different parts of foods may differ and result in time-dependent water migration until equilibrium is attained. This is often the case in bakery products, which may have a high water activity filling in a lower water activity matrix. However, it is important to note that lipids and fat are not water-soluble, and water activity is mainly affected by water-soluble and water-miscible food components only. Therefore, high-fat products with relatively small amounts of water, such as butter and margarine, may have a relatively high water activity.

The water activity of dilute sugar and salt solutions can be calculated using Raoult's law (eqn (8)) and for more concentrated solutions when the activity coefficient is known using eqn (9), as discussed above. Many food systems, however, deviate from ideal

behavior, and Raoult's law cannot be used. Then, either sorption isotherms or other water activity calculation methods must be followed to estimate water activity in food product development. Several equations are available for predicting water activity of a food product based on the product composition. These include the Norrish equation, which extends the use of mole fractions of food components with additional coefficients to calculate water activity, and the Grover equation, which uses different empirical constants for the various components in a polynomial. The Ross equation (eqn (15)) gives probably the most accurate water activity predictions for complex solutions. The equation is relatively simple and does not require the use of empirical constants.

$$a_w = a_{w1} \times a_{w2} \ldots a_{wi}. \qquad (15)$$

Mixtures of two or more ingredients or food components with different water activities in a sealed package will exchange water until the water activity of the components becomes the same. The water content of the components initially and after equilibration may differ significantly, owing to differences in sorption properties. There are several approaches, which can be applied to estimate the final water activity of a food mixture in a sealed package. These methods are based on knowledge of the sorption isotherms and their use to predict the equilibrium water activity. A simple approach is to establish a sorption isotherm for the mixture and predict water activity using the total water content of the ingredients. Another approach assumes that the sorption isotherms of the components over the applied water activity range follow a straight line. This leads to eqn (16), which can be used to predict the equilibrium water activity, a_w. The method is also useful in estimating an appropriate ratio of food components to keep the final, equilibrium water activity at an acceptable level.

$$a_w = \frac{a_{w1}b_1w_w + a_{w2}b_2w_2 \ldots + a_{wi}b_iw_i}{b_1w_1 + b_2w_2 \ldots + b_iw_i}, \qquad (16)$$

where a_{w1}, $a_{w2} \ldots a_{wi}$, b_1, $b_2 \ldots b_i$, and w_1, $w_2 \ldots w_i$ refer to the initial component 1, 2 . . . i water activity, slope of the sorption isotherm, and weight of dry solids, respectively.

The role of water activity as a food stability parameter has been well recognized. Water activities can be measured with commercially available equipment, and sorption isotherms are available for most food components. Furthermore, computer software for water activity calculations and establishment of sorption isotherms has been developed and used widely in the food industry.

See also: **Water**: Structures, Properties, and Determination; Physiology; **Water Activity**: Effect on Food Stability

Further Reading

Barbosa-Cánovas GV and Welti-Chanes J (eds) (1995) *Food Preservation by Moisture Control: Fundamentals and Applications*. Lancaster, PA: Technomic.

Bell LN and Labuza TP (2000) *Moisture Sorption. Practical Aspects of Isotherm Measurement and Use*. St. Paul, MN: American Association of Cereal Chemists.

Fontana AJ (2000) Understanding the importance of water activity in food. *Cereal Foods World* 45: 7–10.

Iglesias HA and Chirife J (1982) *Handbook of Food Isotherms*. New York: Academic Press.

Jouppila K and Roos YH (1997) Water sorption isotherms of dehydrated milk products: Applicability of linear and nonlinear regression analysis in modeling. *International Journal of Food Science and Technology* 32: 459–471.

Labuza TP (1968) Sorption phenomena in foods. *Food Technology* 22: 263–265, 268, 270, 272.

Labuza TP (1980) The effect of water activity on reaction kinetics of food deterioration. *Food Technology* 34(4): 36–41, 59.

Rockland LB and Stewart GF (eds) (1981) *Water Activity: Influences on Food Quality*. New York: Academic Press.

Roos YH (1995) *Phase Transitions in Foods*. San Diego, CA: Academic Press.

Roos YH (2000) Water activity and plasticization. In: Eskin NAM and Robinson DS (eds) *Food Shelf Life Stability*. Boca Raton, FL: CRC Press.

Roos YH, Leslie RB and Lillford PJ (eds) (1999) *Water Management in the Design and Distribution of Quality Foods*. Lancaster, PA: Technomic.

Seow CC, Teng TT and Quah CH (eds) (1988) *Food Preservation by Moisture Control*. London: Elsevier.

Simatos D and Multon JL (eds) (1985) *Properties of Water in Foods*. Dordrecht: Martinus Nijhoff.

Slade L and Levine H (1991) Beyond water activity: Recent advances based on an alternative approach to the assessment of food quality and safety. *Critical Reviews in Food Science and Nutrition* 30: 115.

Effect on Food Stability

Y H Roos, University College Cork, Cork, Ireland

Introduction

Water activity, together with pH and temperature, is often an important factor in controlling food stability. These three parameters all, but not alone, influence microbial activity, rates of chemical and enzymatic reactions, and changes in food texture. Water activity,

however, has a significant role in controlling texture formation and textural changes in intermediate-moisture foods (IMF) and low-moisture foods. In such foods, removal or sometimes sorption of water results in concentration changes of food solids and thereby has an effect on reaction rates. Enhanced reaction rates resulting from water removal may be desired, as in baking to provide color and flavor, or they may be detrimental as in production of many food powders, such as dairy powders. That makes the time–temperature–water or water activity control in food production and storage extremely important.

In some cases, the maximum food stability at low water contents is achieved at a water activity corresponding to the Brunauer–Emmett–Teller (BET) monolayer water content. Furthermore, water affects the physical state and water plasticization, i.e., softening of water miscible food solids. As a plasticizer, water controls the glass transition of food components, such as carbohydrates and proteins. The glass transition is a well-known property of all inorganic and organic amorphous materials. The glass transition occurs over a temperature range, which is often referred to with the glass transition temperature, T_g. The physico-chemical properties of amorphous materials change considerably over the glass transition, as the transformation includes a change between the liquid-like and solid-like material states. There are also established and possible relationships between the physical state, water plasticization, and rates of diffusion-controlled reactions, such as nonenzymatic browning, enzymatic reactions, and oxidative changes.

In general, water activity provides valuable information of the effects of water content on water availability and the physical state of food solids. Such information can be described using sorption isotherms and state diagrams. They provide *critical* values for water content, water activity, and temperature, which are important in the characterization of food behavior in processing and in establishing criteria for packaging requirements and appropriate storage conditions. These areas and the role of water activity in controlling food stability are discussed in the present article.

Water Activity and Food Stability

Most fresh foods are high-moisture materials, and their shelf-life is reduced by enzymatic changes, the growth of microorganisms, and mechanical damage. High-moisture foods have an a_w of 0.90–0.999, and they often contain more than 50% (w/w) water (Table 1). These foods include fresh meats and seafood, various dairy products, fruits, vegetables, and beverages. Most bacteria, molds, and yeasts are likely to grow in high-moisture foods. However, the types of spoilage microorganisms and the growth of various species are highly dependent on pH, temperature, and water activity.

Table 1 Examples of water activity ranges of various foods

Food systems	*Water activity range*
Fresh meat, fish, vegetables, < 40% (w/w) sucrose, < 7% (w/w) salt	> 0.95
Bread, cooked sausages, medium aged cheese	0.90–0.95
Salami, old cheese, > 65% sucrose, > 15% salt	0.87–0.90
Dried beef, sweet condensed milk, cereals with 15% (w/w) water	< 0.86
Jam, marmalade, old salami, > 26% (w/w) salt	0.80–0.87
Flour, cereals, nuts	0.75–0.80
Caramels, honey, toffee	0.60–0.75
Breakfast cereals, snack foods, food powders	0.20–0.60

IMF have a water activity within the range of 0.60–0.90a_w, and their water contents normally vary between 10 and 50% (w/w) (Table 1). These foods include many traditional low-moisture foods, such as grains, nuts, and dehydrated fruits, but also a number of processed foods or foods designed and manufactured to have a known composition to provide stability. Such foods may have particular applications and requirements for stability, e.g., when used as fillings in bakery products or confectionery. Although microbial spoilage is prevented below 0.60a_w, and many microbes do not grow in IMF, their stability and shelf-life are reduced by deleterious changes, such as structural transformations, enzymatic changes, browning reactions, and oxidation, depending on a_w, pH, and temperature. The rates of such changes are often at least to some extent affected by the physical state of the materials and the extent of water plasticization of water-miscible solids.

Low-moisture foods obviously have the lowest water contents, often below 10% (w/w), and their water activity is lower than 0.6a_w (Table 1). Such foods are not subject to microbial spoilage. Their shelf-life, however, is often limited by chemical and textural changes, particularly browning and other changes in color and flavor as well as oxidation. These foods may be exceptionally hygroscopic and exhibit water sorption from their surroundings. Water sorption is reduced by the use of protective packaging, but sorption occurs during storage as a result of permeation or damage of protective packaging. Many low-moisture foods have a solid appearance, or they have a crispy, solid texture. Water sorption in such foods may lead to stickiness, structural changes, such as loss of crispness and sogginess,

and coincident increases in the rates of browning and enzymatic changes. A number of products contain amorphous components, for example, lactose in dairy powders and sucrose in many bakery products and confectionery. Such amorphous components may crystallize during water sorption, as water activity increases. It is important to notice that low-moisture foods are often glassy materials with T_g values above their normal storage temperature. Water sorption and associated plasticization may decrease the T_g, and the glass transition may occur over the range of the concurrent increase in water activity.

Stability Control

Microbial Stability

Microbial growth requires a minimum a_w, in addition to pH, temperature, and other appropriate conditions that are important for the growth of bacteria, molds, and yeasts. The water activity of high-moisture foods, especially processed foods, can be manipulated to some extent by the addition of salts and sugars or other ingredients, which are known to reduce water activity. Such compositional alterations are highly advantageous in product development and food safety control. For many products, the effects of compositional changes on water activity can be predicted on the basis of composition and confirmed by measuring water activity of the final product. This allows an understanding of storage requirements and estimation of the product shelf-life in various storage conditions. In high-moisture foods, the main role of water activity control is to govern and reduce the risk of the growth of pathogenic and spoilage bacteria. Examples of water activity limits for the growth of selected microorganisms are given in Table 2.

The lowest water activity limit for microbial growth of $0.60a_w$ allows the growth of xerophilic yeasts. Above this limiting water activity, IMFs have an increasing possibility for the growth of various microorganisms with increasing water activity. However, the water activities of IMFs are such that pathogenic bacteria are unable to grow, but there is a possibility for the growth of molds and yeast. The growth of these microorganisms must be controlled by careful adjustment of product water activity, use of protective packaging to avoid contamination, and selection of appropriate humectants, pH control, and use of antimicrobial agents.

Microbial stability is an obvious, and often the most important, criterion in food preservation. The a_w limits for growth of various microorganisms, as shown in Figure 1, are well established and successfully used in food product development and manufacturing as well as control of product safety. Furthermore, in high-moisture foods and several IMF products, water activity is relatively constant and dependent on composition, especially solids content, and the type of water-soluble components.

Table 2 Water activity (a_w) limits for the growth of selected pathogenic and spoilage microorganisms

Microorganism	*Minimum* a_w
Bacteria	
Bacillus cereus	0.930
Bacillus subtilis	0.900
Campylobacter jejuni	0.990
Clostridium botulinum	0.940
Clostridium perfringens	0.945
Escherichia coli	0.935
Halobacterium halobium	0.750
Lactobacillus plantarum	0.940
Listeria monocytogenes	0.920
Salmonella spp.	0.940
Shigella spp.	0.960
Staphylococcus aureus	0.860
Vibrio parahaemolyticus	0.936
Yersinia enterocolitica	0.960
Molds	
Aspergillus candidus	0.750
Aspergillus flavus	0.780
Aspergillus niger	0.770
Erotum echinulatum	0.620
Penicillium citrinum	0.800
Penicillium expansum	0.830
Penicillium patulum	0.810
Rhizopus nigricans	0.930
Xeromyces bisporus	0.610
Yeasts	
Saccharomyces bailii	0.800
Saccharomyces cerevisiae	0.900
Saccharomyces rouxii	0.620

Chemical and Enzymatic Stability

The chemical stability of high-moisture foods is not significantly affected by water activity, as the microbial quality is the main determinant of product shelf-life. The role of water activity in determining the chemical and enzymatic stability of IMF and low moisture foods is more important. As the water activity and water content decrease, the concentration of reactants in the water phase of foods becomes obviously increased. Therefore, rates of several reactions may increase with decreasing water activity. The relative rate of deteriorative changes in intermediate and low moisture foods is traditionally related to water content and a_w, as shown by the 'Food stability map' in Figure 2. Early studies applying nuclear magnetic resonance spectroscopy found 'mobilization points' (water activity allowing reactant mobility) for solutes in low-moisture food matrices. The mobilization

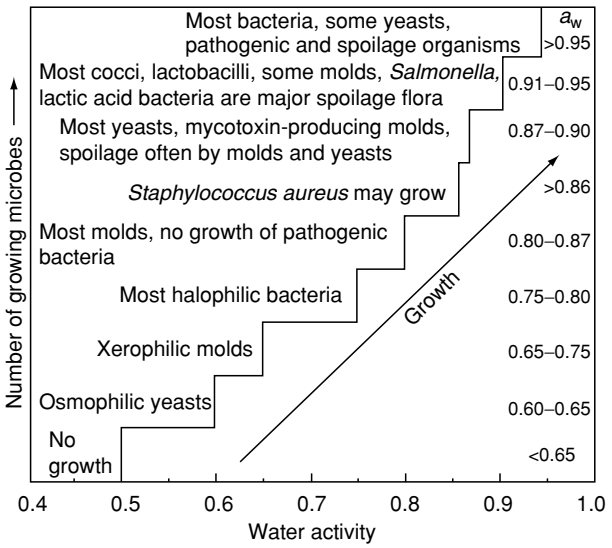

Figure 1 Growth of various microorganisms at different water activity conditions.

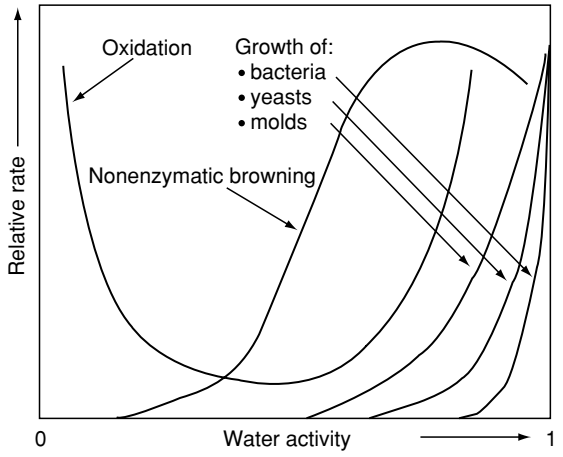

Figure 2 Food stability map describing the relationship between water activity and food stability.

point was peculiar to the system, and the level of hydration needed to achieve mobility was solute-dependent. No solute mobilization occurred below the BET monolayer value, and, for example, experimental data of nonenzymatic browning reaction rates suggested that browning initiated at the mobilization point. An increase in the reaction rate was apparent with increasing a_w, and the rate maximum occurred at the water activity corresponding to a hydration level allowing complete mobilization. However, several other factors, including the glass transition, are important in controlling changes occurring in intermediate and low moisture foods.

Enzyme activity in low-moisture foods is often related to hydration. At low water activities enzymatic activity is generally not observed, as water cannot enhance diffusion of substrates to enzyme molecules. Such water activity dependence applies to hydrolases and oxidases, unless the substrates are nonaqueous liquids, allowing changes to occur fairly independently of water activity. It seems that not only water activity but also the ability of water to give a known mobility to enzymes and substrates is important in controlling enzyme activity. The amount of water needed increases with increasing molecular size owing to impaired diffusion, in particular for enzymes, which are active in the water phase of foods. Therefore, for example, lipase activity is not necessarily related to the mobility provided by water.

Lipids exist often in a separate phase apart from the water phase. At low water activities and water contents, lipids become more accessible to the atmospheric oxygen, and oxidation rates may increase (Figure 2). The rate of oxidation, however, decreases rapidly with increasing water activity, as the lipids may become protected from atmospheric oxygen owing to the formation of protective aqueous layers in the food microstructure. In several low-moisture foods, lipids, flavors, and other oxygen sensitive compounds are entrapped within the amorphous structure. Such encapsulated compounds also exhibit reduced rates of oxidation as compared with 'free' lipids.

Traditional shelf-life predictions of low-moisture foods were based on the information of rates of deteriorative changes and loss of nutrients at various temperatures and water contents. An increase in water content above the BET monolayer value at a constant storage temperature often results in rapid deterioration as reaction rates increase at intermediate water contents. The main factors affecting reaction rates in low-moisture food materials, however, can be controlled by a number of factors, which include food composition and the type of the reaction, temperature, pressure, water content, and pH. In concentrated food systems, changes in viscosity and relaxation times in the vicinity of the glass transition may also affect diffusion and thereby contribute to reaction rates and product shelf-life.

Water Plasticization

Intermediate and low moisture foods often exist in an amorphous (noncrystalline, no defined structure of molecular arrangement), elastic, 'rubbery,' or 'leathery' state or in a solid-like, glassy state. The amorphous state is typical of liquids and remains in

many biological materials owing to complicated chemical composition, rapid removal of solvent water, or quench cooling of a liquid melt in processes not allowing component crystallization. For example, many sugars typical of foods do not crystallize in dehydration, but remain in a supercooled, liquid state after removal of the solvent water. Glassy, solid-like foods may suffer the glass transition as a result of water sorption, which may be observed from rapid changes in structure and appearance, and enhanced rates of chemical and enzymatic reactions.

The glass transitions of food components have a wide range of temperatures depending on the component itself and also on its molecular weight. In general, high-molecular-weight food components, such as carbohydrate polymers and proteins, have glass transition temperatures at very high temperatures well above 100 °C. Low-molecular-weight components may have very low glass transition temperatures, for example, the glass transition of amorphous water occurs at around −135 °C, whereas many anhydrous sugars have a T_g between 0 and 100 °C. The T_g is often measured using differential scanning calorimetry, and it can be observed with a number of other thermal and spectroscopic techniques. These include dielectric analysis, dynamic mechanical analysis, electron spin resonance spectroscopy, infra-red and Raman spectroscopy, and nuclear magnetic resonance spectroscopy. Examples of glass transition temperatures of food components are given in Table 3.

Table 3 Examples of melting and glass transition temperatures of food components

Food component	*Melting temperature of crystalline solids (DSC onset,° C)*	*Glass transition temperature (DSC onset,° C)*
Carbohydrates		
Fructose	108	5
Galactose	163	30
Glucose	143	31
Lactose	223	101
Maltose	160	87
Sorbitol	85	−9
Starch[a]	na	250
Sucrose	173	62
Trehalose	215	100
Xylitol	89	−29
Proteins[a]		
Gelatin	na	207
Gliadin	na	179
Glutenin	na	189
Myoglobin	na	149
Ovalbumin	na	157

[a]Estimated values.

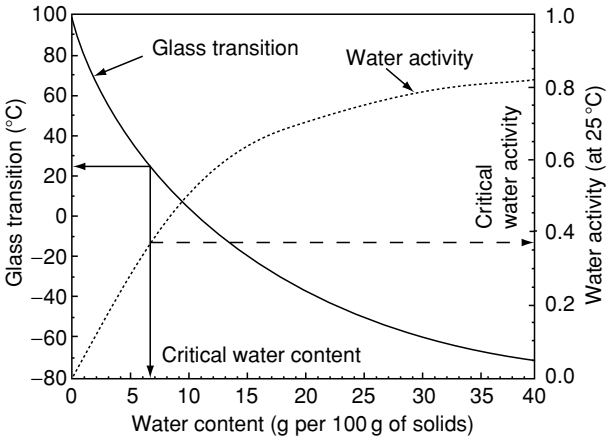

Figure 3 Relationships between glass transition temperature, water activity, and water content for amorphous lactose with critical values for water content and water activity at room temperature.

Carbohydrates and proteins in foods are water-miscible components, and they are plasticized by water. Water plasticization of the amorphous structure results in a decrease in T_g. The decrease in T_g is substantial even at low water contents, and at high water contents, the T_g approaches that of amorphous water. Therefore, at a constant temperature, e.g., room temperature or a typical storage temperature of intermediate- and low-moisture foods, a water activity or water content range can be established, which corresponds to the glass transition of the material. The relationships between glass transition, water activity, and water content are described in Figure 3. The water activity and water content, which correspond to the T_g depression to the storage temperature can be defined as 'critical water activity' and 'critical water content,' respectively.

$$T_g = \frac{w_1 T_{g1} + k w_2 T_{g2}}{w_1 + k w_2}, \qquad (1)$$

where w_1 and w_2 and T_{g1} and T_{g2} are the weight fraction and glass transition temperature of solids (1) and water (2), respectively.

The relationship between the glass transition temperature depression and water content is often predicted with the Gordon–Taylor relationship (eqn (1)). The model can be fitted to experimental T_g and water content data and used together with water sorption data to establish relationships between the T_g and water sorption properties, as shown for amorphous lactose in Figure 3. Such predictions are useful in food product development, because they allow estimation of food stability in terms of the physical state and water activity.

Water Activity and Material Properties

Several properties, such as dielectric properties, mechanical properties, and viscosity, of amorphous food materials are greatly affected by the glass transition. These changes are related to relaxation times and molecular mobility, which may substantially change and affect food stability over the glass transition. Such changes may contribute to observed changes in textural characteristics, e.g., loss of crispness and sogginess, and reaction rates, e.g., nonenzymatic browning and oxidation. Furthermore, the solid, glassy state supports the amorphous state of food components, but many crystallizing components, e.g., sugars, may crystallize in the supercooled liquid state. Both textural changes and crystallization of amorphous components contribute to a loss of flavor and aroma compounds as well as release and oxidation of encapsulated lipids in numerous dehydrated foods and food powders.

In general, the relaxation times of molecular rearrangements in the glassy state are sufficient to provide stability for low-moisture food systems. The changes occurring over the glass transition and in the supercooled liquid state reduce the relaxation times, and the materials lose their long-term stability. For example, amorphous, glassy lactose in food systems is stable, but an increase in water content resulting in glass transition leads to a rapid decrease in viscosity and flow of the material (transformation from solid state to a syrup) accompanied by a loss of structure (stickiness and collapse) and lactose crystallization. Crystallization of amorphous sugars is often observed from water sorption isotherms. When crystallization occurs, sorbed water is not maintained in the crystals, and a decrease in water content is recorded. Such crystallization is time-dependent and detrimental to the quality of amorphous powders.

Stiffness

A single model based on the Fermi's distribution model, as given in eqn (2), allows modelling of the decrease in relaxation times of mechanical properties (stiffness) in the vicinity of the glass transition. According to the model, *stiffness* can be expressed in terms of a defined and measurable stiffness parameter, Y, which may refer to viscosity, flow, sensory crispness, etc. The stiffness parameter values can be modelled as a function of a_w, T, or m relative to its value at a reference state, Y_s. A constant, a_X, defines the broadness of the transition, e.g., the water activity range over which changes in stiffness occur. The reference value, X_s, obtained from $b = -X_s/a_X$, indicates the value for a_w, T, or m that decreases Y to 50% below Y_s.

$$\ln\left(\frac{Y_s}{Y} - 1\right) = b + \frac{1}{a_X}X, \qquad (2)$$

where Y is the stiffness parameter, Y_s is the stiffness parameter at a reference state, X is a_w, T, or water content, m, and b and a_X are constants.

The term *collapse* relates to stiffness and covers various time-dependent structural transformations that may occur in amorphous food and other biological materials at temperatures above the T_g. These changes reflect the effect of changes in relaxation times of mechanical properties on the rates of structural transformations and flow that occur over and above the glass transition temperature range. The rates of these changes increase rapidly above the critical water content or a_w during food storage. Stickiness and caking of food powders and collapse of structure are collapse phenomena that are related to water activity and glass transition. The main cause of stickiness is plasticization of particle surfaces by water, which allows interparticle binding and formation of clusters. Collapse of the structure occurs when the decrease in viscosity results in flow, and the material cannot support its own weight. Structural collapse often leads to a reduction in quality of dehydrated, especially freeze-dried, foods, whereas stickiness and caking are detrimental to the quality of spray-dried foods, such as dairy powders.

Crystallization of Food Components

Crystallization of amorphous sugars is known to result in serious quality losses in food powders. For example, crystallization of amorphous lactose in dehydrated milk products has been observed to result in acceleration of the nonenzymatic browning reaction as well as other deteriorative changes and caking. Crystallization of lactose coincides with an increase in free fat, which presumably facilitates lipid oxidation in milk powders. Crystallization in low-moisture carbohydrate matrices, which contain encapsulated volatiles or lipids, results in a complete loss of flavor and release of lipids from the matrix. Water plasticization and depression of T_g to below the ambient temperature are responsible for crystallization of amorphous sugars in foods as a result of increased free volume and molecular mobility, decreased viscosity, and enhanced diffusion. Crystallization seems to initiate at T_g or corresponding a_w and proceed with a rate determined by the temperature difference $T - T_g$ to a maximum extent also defined by the $T - T_g$. The kinetics of crystallization of sugars and other amorphous food components at a constant temperature above T_g can be related to water content and a_w, which define the rate controlling $T - T_g$.

Water Plasticization and Reaction Rates

It has been well established that water as a plasticizer has a significant effect on molecular mobility and probably rates of quality changes above a critical, temperature-dependent a_w or water content. A chemical reaction requires sufficient mobility of reactants and reaction products in addition to the driving force, e.g., temperature or concentration, of the reaction or change in quality. Information on factors affecting rates of various kinetic processes can be used to manipulate and control rates of changes occurring in food processing and to develop food products that are less sensitive to detrimental changes during storage.

Nonenzymatic Browning

The nonenzymatic browning reaction is one of the most studied deteriorative reactions affected by water activity. Other reactions include destruction of vitamins, enzymatic changes, and oxidation. The rates of most reactions are affected by water activity and glass transition. However, there are no single determinants of rates of chemical reactions and food stability, although an understanding of both the water activity and glass transition of a food material is important in predicting and controlling food stability.

Nonenzymatic browning is often considered as a series of condensations that can be assumed bimolecular. The reaction occurs between a reducing sugar and an amino acid or amino group and produces typical flavors of several foods in processing. When uncontrolled, the reaction produces unpleasant flavors and colors, and decreases dehydrated food quality during storage. The nonenzymatic browning reaction in several foods exhibits increasing rates above the critical a_w values or water contents corresponding to the glass transition. However, reactions in amorphous foods are complex and may be controlled by several other factors, including concentration changes, pH, and temperature. Established relationships between T_g and the rate of the nonenzymatic browning reaction have, however, suggested that the reaction may become diffusion-controlled, and the rate may be affected by the T_g. The size of the reactants may also be important in diffusion-controlled reactions. It may be assumed that the rate of diffusion decreases with increasing size of the diffusant.

The effect of glass transition on the reaction rate has been related to observed discontinuities in Arrhenius plots suggesting a high activation energy in the vicinity of the transition. Studies of effects of glass transition on reaction kinetics and the nonenzymatic browning reaction in particular are complicated, because of several factors affecting plasticization and reaction rates. Several studies, however, have suggested that the reaction rate increases considerably above the glass transition or the critical water activity.

Oxidation

Oxidation is a common reaction in low-moisture foods. Oxidation of lipids directly exposed to atmospheric oxygen at food surfaces or surfaces of porous dehydrated materials may occur freely. It is well known that 'free fat' in dairy powders, for example, is highly susceptible to oxidation causing quality defects. The stability of dairy powders is related to the glassy state of lactose, and the lipids are at least partially encapsulated within the lactose–protein matrix in the spray-drying process.

Enzymatic Changes

Studies of glass transition effects on enzymatic changes have included, for example, sucrose inverson by invertase. Sucrose inversion by invertase in low-moisture systems is dependent on water activity. The reaction rate increases at water activities above 0.6, but does not seem to be affected by the glass transition of the main food matrix. It is possible that the enzyme mobility is not sufficient to allow the reaction to occur until the reacting molecules are plasticized by water to an appropriate extent above $0.6a_w$.

Modern Stability Maps

Deteriorative changes affected by water activity include such changes as enzymatic changes, nonenzymatic browning, lipid oxidation, microbial growth, and overall stability. The rate of these changes seems to be related to water activity. However, water activity limits for chemical, enzymatic, and mechanical changes are material-specific, similarly as the 'mobilization points,' and they may also be related to the physical state, molecular mobility, water plasticization and glass transition of amorphous food solids. Structural transformations, as well as diffusion-controlled deteriorative reactions and changes affected by crystallization phenomena, occur at increasing rates with increasing water activity above the critical a_w of the material. It is likely that water contents lower than the critical water content are needed for maximum stability. Therefore, modern stability maps provide information on both the water activity and material physical state on the relative rates of deteriorative changes (**Figure 4**). In addition, state diagrams and sorption isotherms are useful as stability maps. Knowledge of material properties is extremely useful in the production of encapsulated flavors, extruded products, confectionery, development of new food products, and avoiding quality

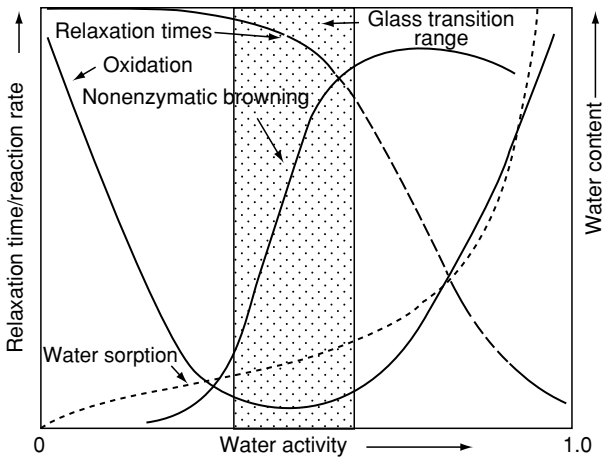

Figure 4 Food stability map taking into account water activity and physical state effects on relaxation times and reaction rates.

changes that may result from mechanical changes, e.g., loss of crispness. The temperature-, water content-, and time-dependent changes that are often problems in manufacturing and storage of food powders and other low-moisture foods can be reduced by avoiding exceeding their critical values of water activity based on the T_g determination or by compositional adjustments that provide sufficiently high values for critical a_w and T_g. The most important applications of producing high-quality dehydrated foods include reduced collapse and improved flavor retention in dehydration processes. The kinetics of enzyme activity is important to food quality, and applying the knowledge in food industry may allow the design of improved products with extended shelf-lives or even improved retention of activity, e.g., in enzyme preparations. Water activity has been successfully used in setting limits for microbial stability in foods and qualitative characterization of relative rates of deteriorative reactions in low-moisture foods. The combined use of a_w and glass transition has a high practical applicability as it provides criteria for critical a_w and, together with sorption isotherm, can be applied to adjust water content at a constant temperature to achieve the maximum food stability.

See also: **Browning**: Nonenzymatic; Toxicology of Nonenzymatic Browning; **Crystallization**: Basic Principles; **Mycotoxins**: Classifications; **Oxidation of Food Components**; **Spoilage**: Chemical and Enzymatic Spoilage; Bacterial Spoilage; Molds in Spoilage; Yeasts in Spoilage; **Water**: Structures, Properties, and Determination; Physiology; **Water Activity**: Principles and Measurement; **Yeasts**

Further Reading

Barbosa-Cánovas GV and Welti-Chanes J (eds) (1995) *Food Preservation by Moisture Control: Fundamentals and Applications*. Lancaster, PA: Technomic.

Bell LN (1996) Kinetics of non-enzymatic browning in amorphous solid systems: distinguishing the effects of water activity and the glass transition. *Food Research International* 28: 591–597.

Buera MP and Karel M (1995) Effect of physical changes on the rates of nonenzymatic browning and related reactions. *Food Chemistry* 52: 167–173

Labuza TP (1980) The effect of water activity on reaction kinetics of food deterioration. *Food Technology* 34: 36–41, 59.

Lievonen SM, Laaksonen TJ and Roos YH (1998) Glass transition and reaction rates: nonenzymatic browning in glassy and liquid systems. *Journal of Agricultural and Food Chemistry* 46: 2778–2784.

Matveev YuI, Grinberg VYa and Tolstoguzov VB (2000) The plasticizing effect of water on proteins, polysaccharides and their mixtures. Glassy state of biopolymers, food and seeds. *Food Hydrocolloids* 14: 425–437.

Nelson KA and Labuza TP (1994) Water activity and food polymer science: implications of state on Arrhenius and WLF models in predicting shelf life. *Journal of Food Engineering* 22: 271–289.

Peleg M (1993) Mapping the stiffness–temperature–moisture relationship of solid biomaterials at and around their glass transition. *Rheologica Acta* 32: 575–580.

Roos YH (1993) Melting and glass transitions of low molecular weight carbohydrates. *Carbohydrate Research* 238: 39–48.

Roos YH (1995) *Phase Transitions in Foods*. San Diego, CA: Academic Press.

Roos YH, Karel M and Kokini JL (1996) Glass transitions in low moisture and frozen foods: effects on shelf life and quality. *Food Technology* 50(11): 95–108.

Roos YH (2000) Water activity and plasticization. In: Eskin NAM and Robinson DS (eds) *Food Shelf Life Stability*. Boca Raton, FL: CRC Press.

Roos YH, Leslie RB and Lillford PJ (eds) (1999) *Water Management in the Design and Distribution of Quality Foods*. Lancaster, PA: Technomic.

Shimada Y, Roos Y and Karel M (1991) Oxidation of methyl linoleate encapsulated in amorphous lactose-based food model. *Journal of Agricultural and Food Chemistry* 39: 637–641.

Slade L and Levine H (1991) Beyond water activity: Recent advances based on an alternative approach to the assessment of food quality and safety. *Critical Reviews in Food Science and Nutrition* 30: 115.

WATER PLANTS

T M Koyama, Ninon University, Fujisawa City, Japan

Lotus (*Nelumbo nucifera* Gaertner) (Nymphaeacea: Waterlily Family)

Lotus is a well-known water plant in Asia, and is widespread from Iran and the Caspian Sea eastwards to China and Japan, and southeastwards to India and Indo-China, then to northern Australia. In India, Thailand, China, and Japan, it is the sacred flower of Buddhism, and is the national flower of Thailand.

It is a large, emergent herb, growing in marshes and shallow ponds with rhizomes creeping in mud and sending out leaves and flowers over the water. The slender rhizomes spread extensively, branching and rooting at nodes, which bear a single leaf and a flower at the axil of some leaves; the round leaf blades are deeply glaucous-green, 70 cm or more across, slightly bowl-shaped and have some 20 nerves radially arranged. They are peltate on a petiole up to 2 m in length, usually sparsely bearing brownish spines. The large flowers are 10–20 cm across, appear in summer in temperate regions, are subtended by four large bracts, and terminate in a scape (about 2 m long) which is sparsely spiny. The petals are many, originally deep pink but sometimes white. In the center of the flowers, there are many yellow stamens and an obconical receptacle with many holes, each containing an ovary. The receptacle develops into a large, obconical, spongy structure, flat on top, and maturing some 20–30 nuts. The nuts are ellipsoidal or obovoidal, deep brown in color with hard shells, and about 2 cm long.

The history of the lotus can be traced back to the diluvian epoch, the formation of which, in central Japan, contains the fossil rhizomes and fruits of lotus. Records show that it was grown in ancient Egypt and that a white-flowered variety was grown in India as early as 500 BC. It reached Japan during the seventh century, at which time it was used mostly as an ornamental. Now, in Japan, there are over 100 cultivars of lotus, including some 10 varieties as food crops, showing varying colors (white, pink-tinged, pink-striped, pink-variegated, yellow, orange–yellow, etc.) and shapes (double-flowered and polycephalous) of flower in the ornamental forms. In Japan, it was also hybridized with the North American yellow-flowered lotus, *N. lutea* Persoon, to produce another ornamental lotus.

The edible parts of the lotus are principally its rhizomes and seeds, though in China, immature receptacles are occasionally eaten boiled or fried as a delicacy, and whole plants are used as a Chinese medicine against high blood pressure.

In China and Japan, lotus cultivated in paddies begins producing tuberous apices of rhizomes in late autumn and early winter. These fat, tuberous apices of rhizomes are much thicker than the leaf-bearing parts of the rhizome, with one to three internodes up to 30 cm long by 10 cm wide, bearing some 10 pipes as air spaces; these are called the 'lotus roots' and are widely marketed. Recently, water-boiled lotus roots have been exported from China as a canned vegetable. In Japan, lotus roots are eaten boiled and seasoned with soy sauce and sugar, or fried as tempura (Japanese fritters), or sometimes pickled, while in China, occasionally, a kind of arrowroot starch is produced, which is used with spice and pork or with pork spareribs for making a popular soup.

After removal of the dark brown, hard shell, the obovoidal-to-round seed is covered with thin brown testa and is 13–17 mm long. To prepare the edible parts, which are the thick cotyledons, rich in raffinose, the testa and hypocotyle, which is bitter, must be removed. Slightly immature seeds have a very sweet taste and can be eaten raw. The cotyledons of mature seeds are hard and must be boiled prior to eating. Water-boiled lotus seeds are canned and exported. They are also candied (i.e., preserved in sugar) and eaten as a dessert sweet.

The yellow-flowered lotus, *N. lutea* Persoon, from the eastern USA, is the American counterpart species of lotus. Its rhizomes are not as thick as those of the Asian lotus, but it has been eaten by American Indians in the past. *Euryale ferox* Roxb., of the same family is an annual water plant occurring in India, China, and southern Japan. It has large, orbicular leaf blades, 30–100 cm across, floating on the water surface, and smaller flowers of some 4 cm in diameter. A rather sweet starch extracted from the seeds is used for making a kind of rice-cake in Japan. In Central China, the caudices and leaf petioles (with the skin removed) of this species are sometimes used as vegetables.

Water Chestnut, also called Water Caltrops or Trapa Nut (*Trapa natans* L.) (Onagraceae: Willow Herb Family)

Water chestnut is an aquatic annual herb that grows in ponds or marshes with its roots in the muddy soil at

the bottom of the water and with its upper leaves floating on the water surface. Its slender stems are of varying length, depending upon the depth of the water, and bear feathery submerged leaves at the nodes. The leaves of a normal kind, with a rhombic blade shallowly serrate on the upper margins, are gathered in a rosette at the tip of the stem; they are deep green and shiny on the upper surface, slightly hairy and often colored with purplish brown beneath. The leaf petioles are swollen at the midway, enclosing spongy aerial tissue, which helps the blade to float on the water surface. Flowers are borne at the leaf axils of floating leaves, and are pedicelled. They have a four-part calyx, four white petals, four stamens and a pistil. Fruits are nut-like, 4–5 cm wide and purple–brown with two 1-cm-long, horn-like spines, one on each side.

The nuts are the edible part of water chestnut, and are eaten boiled, or occasionally roasted; alternatively, the dried nuts are ground into a starchy flour to prepare a kind of gruel or for fermentation into a local liqueur. Young nuts are sometimes eaten raw, with the greenish skin removed. In China, the fruits are listed as a Chinese medicinal plant, to be used as an antipyretic and as an astringent.

The use of the water chestnut as a food can be traced back to Neolithic times, and the fruits are found abundantly in Tertiary sediments. People of Central Europe and Asia eat the typical water chestnut, collecting the fruits from wild plants, but in central and southern China and in India, the variety *T. bispinosa* (Roxb.) Kakino is cultivated in paddy fields. Related species, which are occasionally cultivated in China and Korea, are *T. bicornis* Osbeck and *T. quadrispinosa* Roxb. While the former species bears larger fruits of 7–8 cm in width, the latter bears four-horned fruits. In northeast China, there occurs another wild species, *T. maximowiczii* Korsh., of which the smaller nuts (*c.* 1–1.5 cm in width) are also eaten locally, and sometimes used for fermentation into a local spirit.

Chinese Water Chestnut (*Eleocharis dulcis* Trinius cv. *tuberosa* (Roxb.) T. Koyama) (Cyperaceae: Sedge Family)

The Chinese water chestnut, also known by the Chinese names '*Ma-ti*' or '*Pitzi*', is a stem vegetable crop derived form its wild stock, *E. dulcis* Trinius, which is sporadically, but widely, distributed from Madagascar through the Indian subcontinent and southern China to northern Australia and the Pacific Islands, and also to southern Japan. It is rather widely cultivated in both mainland China and Taiwan, and some parts of Thailand, not only for local consumption but also as a source of preserved (canned) vegetables for export.

Grown in paddy fields, often as a substitute crop of rice (where rice is overproduced), plants of the Chinese water chestnuts form a tuft of many slenderly tubular stems, which are 60–90 cm high, deeply green, hollow inside and transversely septate. Leaves of a 'normal' kind disappear, being reduced to pale or purple–brown sheaths surrounding the base of the stems. Small flowers are borne in cylindrical spikes at the apices of some stems; the spikes bearing yellow–brown scales imbricatedly. The base of the plants send out slender, elongated, underground runners that form a tuber at the apex. Tubers are depressed-globose, 2–5 cm in diameter and 1.5–3 cm high, and are clothed with thin, purple–brown to black–brown scales. Being a high polyploid with approximately 150 chromosomes, the cultivated Chinese water chestnut does not produce seeds and must be propagated by tubers.

The edible part of the Chinese water chestnut is the tuber (see Table 1). It is a considerably popular 'root' vegetable (strictly speaking, a stem vegetable) among the Chinese, and is also used in Thailand, the Philippines and Indonesia, but not in the traditional Japanese kitchen. There, they eat the tubers boiled and seasoned in various vegetable dishes, and extract fine-quality starch grains. In addition, minutely chopped tubers are mixed with ground meat when making Chinese meatballs. The tubers, rich in starch, are crispy and slightly sweet, making a good ingredient in vegetable salads; hence the water-boiled tubers are exported to various parts of the world as a canned vegetable.

Table 1 Chemical components of Chinese water chestnut tubers (with skins removed)

	America data[a]	*Chinese data*[b]
Major components (%)		
Water	78.3	86.0
Protein	1.4	1.2
Fat	0.2	0.1
Carbohydrate	19.0	11.5
Fiber	0.8	0.1
Ash	1.1	1.1
Minerals (mg per 100 g)		
Calcium	4.0	1.0
Phosphorus	65.0	64.0
Iron	0.6	0.9
Vitamin C (mg per 100 g)	4.0	7.0

[a]Watt BK and Merrill AL (1963) Composition of foods. *USDA Agricultural Handbook No. 8*, pp. 1–189. Washington, DC: US Department of Agriculture.
[b]Chang WH and Chow LT (1980) Studies on the chemical and nutritional compositions of 12 unexploited economic plants. *National Science Council Monthly* 8: 827–832.

In 1947, importation of canned Chinese water chestnut into the USA reached 1154 tonnes per year. The yield of this variety in China is 18–37 tonnes per hectare. The wild stock of Chinese water chestnut differs from the cultivated in having much smaller tubers (only as large as 10 mm across) and due to the fertile nature (seed-bearing) of the flowers. Rumphius in 1750 had noted that the natives of Java were eating the tubers of this wild Chinese water chestnut, which were seen in morning markets between August and December. The cultivated Chinese water chestnut, apparently developed from the wild stock in northern Vietnam or adjoining southern China, has two obvious forms: the red-skin type (comparatively large, soft tubers with reddish–purplish skin), and the black-skin type (comparatively small tubers with a harder flesh and covered with black–brown skin). The former is for vegetable use, whereas the latter is used for extracting high-quality starch grains called the Ma-ti powder; the latter is principally used both for cooking and as a stomach powder in Chinese medicine.

Arrowhead (*Sagittaria trifolia* L. cv. *sinensis* (Makino) T. Koyama) (Alismataceae: Arrowhead Family)

A marsh plant of Chinese origin, arrowhead is cultivated in the paddy fields of temperate China and Japan for the tubers. The cultivated arrowhead, cv. *sinensis*, is much larger than its wild counterpart, with its leaves attaining a height of 100 cm. Several leaves grow from a short rhizome; they have a long petiole, emergent from the water. The leaf blades are sagittate (arrow-shaped) with two rather broad and elongated basal lobes, and are 10–15 cm long. Cultivated arrowhead seldom bears flowers. The scape, if produced, is normally lower than the leaves, and bears three branches at several nodes. Flowers are monoecious and unisexual, and the staminate flowers are borne on the upper part of the raceme, the pistillate ones on the basal part of the raceme. Flowers are about 2.5 cm across, and have three green sepals and three white, roundish petals. Staminate flowers have many yellow stamens. Pistillate flowers have a pistil of many carpels, each maturing into an achene.

The short rhizome sends out several slender underground runners, and each develops a tuber at its tip from late summer to autumn. Tubers are globular to somewhat depressed-globular, 3–5 cm across, and their white or bluish surfaces are zonate, bearing deciduous, membranous scales. In winter, the tubers bear a beak-like terminal bud at the apex.

The edible parts of the arrowhead are the tubers, which are considerably rich in starch (24.8%); the other components are water (68.2%), protein (4.4%), fat and oil (0.23%), fiber (0.4%), and inorganic salts (1.37%). Arrowhead tubers are eaten boiled and seasoned in both Chinese and Japanese cooking.

Water Shield (*Brasenia schreberi* Gmelin) (Nymphaeceae: Water Lily Family)

This is a cosmopolitan water plant occurring in Asia, North America, Australia, and Africa, but is eaten only in Japan. The slender rhizome is located in the mud of ponds or lakes, and stretches out its stems into the water. The leaves are alternately arranged on the stems, with the blades floating on the water surface. The blades are broadly elliptic, 5–10 cm long, 'freshly' green and shiny above, and tinged with purple beneath. Their petioles are peltately attached to the center of the blade. Flowers appear during the summer. They are about 2.5 cm across and are a purplish brown color with six petals, some 16 stamens and six to 18 carpels. Sac-like fruits are indehiscent, and usually mature in the water.

The stems, petioles, and lower surface of the leaf blades are covered with a semitranslucent, agar-like substance that is secreted from glands on the plant surfaces. The young shoot apices, in particular, are copiously covered with this gelatinous substance, and this portion is eaten as a delicacy, but rather commonly, in Japan. They are used in Japanese-style clear soup, salads, etc., or they are stored in vinegar and/or in salt. Recently, water-boiled shoots of this plant have been preserved in bottles and cans, and widely sold.

See also: **Ethnic Foods**; **Vegetables of Temperate Climates**: Oriental Brassicas

Further Reading

Chang WH and Chow LT (1980) Studies on the chemical and nutritional compositions of 12 unexploited economic plants. *National Science Council Monthly* 8: 827–832.

Rumphius GE (1750) *Herbarium Amboinense*, vol. 6, p. 7.

Watt BK and Merrill AL (1963) Composition of foods. *USDA Agricultural Handbook No. 8*, pp. 1–189. Washington, DC: US Department of Agriculture.

WATER SUPPLIES

Contents

Water Treatment

J Manning, Wessex Water, Bath, UK

Introduction – Why Treat the Water?

The aim of all water providers throughout the world must be to collect, treat, and distribute a safe supply of drinking water to their consumers that complies with all legislative requirements in a cost-effective way. The natural water cycle provides the raw material but treatment processes are needed to remove chemical and bacteriological contaminants that would otherwise be damaging to health or affect the esthetic appearance of the product at the customer's tap. Technical advances in treatment processes and an understanding of the effects of raw-water contamination on consumers' health will continue to make cost-effective solutions to treatment more accessible. This should mean that compliance with legislation becomes an easier task, but is also likely to mean that quality standards will continue to become more stringent in the UK and Europe.

Water Sources

As the aim is to provide safe clean water, it makes sense to take the raw material from the point in the water cycle where water is least contaminated, in sufficient quantity and well placed geographically, for the supply to reach the customers. Sources can be derived from deep aquifers using boreholes, shallow groundwater such as springs, surface water collected in reservoirs from lowland areas, or less polluted upland areas, or direct abstractions from rivers. Contaminants can come from agricultural sources, surface water drainage or runoff, and sewage treatment. Industrial pollution can also affect water supplies. Harmful bacteria and viruses will be found if there is fecal contamination of the source.

Generally, raw groundwater from boreholes is of good quality as the ground above the aquifer provides a filtration medium to remove particles such as bacteria. However, the time for water to travel through strata can be sufficiently long to dissolve other soluble contaminants such as naturally occurring calcium and magnesium (hardness) but also pesticides and nitrate from agricultural activities. Wherever water is stored for a long period of time, be it as naturally occurring aquifer or an artificial impounding reservoir, reduced oxygen concentrations will dissolve iron and manganese and produce hydrogen sulfide (characterized by a rotten-egg smell). This will result in a dirty and foul-smelling water source, although this can still be successfully treated. The long travel time from surface to aquifer, with naturally occurring treatment on the way, means that groundwater sources are very consistent in their quality and hence desirable for ease of treatment. A large proportion of the UK's water supplies is abstracted from boreholes in both chalk and greensand aquifers.

Springs collect water from the ground into horizontally laid pipework, and as this is very much closer to the ground surface than a borehole, they are more at risk from variable quality and surface water intrusion.

Reservoir, lake, and river water sources, by their nature, are generally contaminated with fecal material, colored yellow or brown by the organic debris of rotting vegetation and turbid from mud and silt, and algae which will grow in suspension. The quality of a surface water source can vary within hours, depending on weather conditions. This is especially true in river intakes and seasonal changes must be expected in reservoirs and lakes. Wet winter conditions will cause runoff of agricultural contaminants such as pesticides and nutrients that can cause excessive algal growth (eutrophication). This can be controlled to some extent by mixing of the reservoir using compressed air lines or wind-powered mixing.

Storage reservoirs provide protection against shortages of rainfall but also blending of source waters to a more consistent quality suitable for treatment. A storage or retention time in excess of 7 days can provide some settlement and a significant reduction in bacterial numbers.

With improvements in technology, we are now seeing a greater variety of source waters becoming available for treatment. For example, reverse osmosis is widely used for desalination of sea water where other sources of fresh water are not available. Many manufacturing industries will use complex treatment processes to recover or reuse effluent produced in manufacturing processes.

Treatment Processes

A summary of some commonly used treatment processes is shown in Table 1. The treatment objectives of a water treatment plant will be site-specific and depend on the raw water quality, and the quantity and standard of the treated water quality that it is required to produce. The main objective will always be to provide water that is microbiologically safe for the consumers to drink. This is achieved using appropriate disinfection processes either to kill or remove bacteria, viruses, and cysts. The control of the disinfection process is therefore considered critical.

Disinfection

Chlorination is the most commonly used disinfection process in the UK. Chlorine is effective because it is toxic, highly oxidizing, soluble in water, and its germicidal effectiveness lasts for days. It is also widely available and therefore cost-effective.

Chlorine will react in water to form hypochlorous acid, which can then dissociate into hydrogen and hypochlorite ions, according to Eqn (1).

$$Cl_2 + H_2O = HOCl + H^+ + OCI^- \qquad (1)$$

This reaction is very important, as the disinfecting power of HOCl, hypochlorous acid, is about 40–80

Table 1 Summary of treatment processes

Example	*Uses*
Disinfection	
Chlorination, ozonation, chlorine dioxide, ultraviolet irradiation	Oxidants used to kill bacteria and viruses
Coagulation	
Aluminum sulfate, ferric chloride, or polyaluminium chloride	Combines small, suspended, and colloidal particles into larger, flocculated particles by chemical attraction. Algae, iron, manganese, color, and turbidity can be entrained in floc particles which are then removed by solid/liquid separation processes
Solid/liquid separation	
Microstraining	Large (20–100 μm) particles such as algae and leaves removed by straining
Dissolved air flotation	Removal of coagulant sludges by floating on a curtain of small air bubbles
Sedimentation	Removal of coagulant sludges by settlement. Suited to sludge with high inorganic content
Slow sand filtration	Removal of colloidal, organic, and suspended material by straining and biological activity on the surface of sand filter beds
Rapid gravity filtration	Removal of suspended material by straining and entrapment within the depth of sand filter beds
Membrane micro- and ultrafiltration (size exclusion)	Removal of submicron-sized particles using membranes manufactured from polymeric material
Membrane nanofiltration and reverse osmosis (diffusion control)	Removal of submicron particles and dissolved inorganic species of high molecular weight
Membrane ion exchange or electrodialysis (charge-controlled)	Removal of salts, i.e., desalination of sea water or nitrate removal
Adsorption	
Granular-activated carbon filtration (fixed bed) or powdered activated carbon (requires filtration process)	Adsorption of pesticides and volatile organic compounds, removal of taste and odors
Aeration	
Packed towers or cascades	Removal of H_2S, volatile organic compounds, or carbon dioxide for corrosivity control. Oxidation of iron and manganese
Chemical addition	
Acid and alkali for pH adjustment	Mineral acid and/or sodium hydroxide, sodium carbonate, or lime for optimum coagulation
Phosphoric acid	Provides phosphate coating on lead pipes to control lead concentrations in water
Softening	
Lime addition	Removal of calcium carbonate by precipitation at high pH. Reduces scale formation in downstream structures

times that of OCl^-, hypochlorite. Chlorination is more effective at pH values below 6 as there is a high proportion of HOCl present and very little OCl^-. The quantity of HOCl and OCl^- that is present in the water is known as the free chlorine. The effectiveness of chlorination will depend on factors listed in Table 2.

Chlorine reacts within minutes with a variety of substances, including nitrite and ferrous ions. The demand for chlorine from these reactions must be satisfied before the slower reactions with the bacteria cell walls can take place and hence provide disinfection.

To insure the disinfection process has taken place satisfactorily, continuous monitoring of the residual chlorine concentration will normally be used along with regular analysis for fecal and heterotrophic bacteria.

For surface water, where there may be higher concentrations of organic contaminants or ammonia, superchlorination, followed by dechlorination, may be used. The chlorine concentration is held for some time at a value much higher than that expected to be found at the customer's taps – around 2–5 mg Cl_2 l^{-1}. This will quickly satisfy the chlorine demand of the water, and the concentration is then reduced to an acceptable level by adding sulfur dioxide. Treated water will leave the plant with a residual of around 0.3 mg Cl_2 l^{-1} to control any subsequent regrowth of bacteria in the distribution pipework to the customers.

Surface waters often contain humic and fulvic acids from the decomposition of vegetation. This reacts with chlorine to form trihalomethanes (THMs) such as chloroform, which are suspected carcinogens. The formation of THMs can be controlled by insuring chlorination takes place after removal of the precursor compounds. Concern in forming these byproducts has led to an increase in alternative disinfectants such as ozone, ultraviolet radiation, and chlorine dioxide.

Research over the last decade has increased awareness and concern over organisms that are harmful to health and not affected by normal disinfection techniques. *Cryptosporidium* and *Giardia* are both difficult to identify analytically, resistant to chlorine, and form very small (2–5 μm) oocysts outside the host. Fecal in origin, they are commonly found in rivers, lakes, and surface water sources, but because of their persistence in the environment and very small size they can also be found in groundwater sources. Alternative treatment strategies based on removing the organisms, using barrier treatment, rather than killing them are required. Accepted practice is to use well-controlled coagulation and rapid gravity filtration, preferably with an additional solids removal stage such as DAF (dissolved air flotation) or sedimentation, or using membrane micro- or ultrafiltration.

Coagulation and Flocculation

Suspended, colloidal, and dissolved material is commonly found in surface water, such as fine clay particles, algae, and color from humic substances. These dissolved substances and submicron-sized particles can react with coagulant to form larger, agglomerated solid particles that can more easily be removed by solid–liquid separation processes. The surfaces of very small particles are charged and the repulsive forces between them help to keep them in suspension. The coagulation mechanism is essentially twofold. First, the coagulant forms highly charged hydrolyzed ions in water which destabilize the charge on the surface of the particles. Second, once the charge has been destabilized, attractive forces, which cause the particles to agglomerate, replace repulsive forces and larger particles called floc are formed. This agglomeration is referred to as flocculation. There are many complex chemical and physical reactions taking place during the coagulation/flocculation process, making exact conditions difficult to predict. The factors affecting efficiency are summarized in Table 3.

Predicting coagulation conditions is very difficult and specific to each water source. Conditions can change quickly with changing raw-water quality.

Table 2 Factors affecting the efficiency of chlorination

Parameter	*Effect on chlorination*	*Typical value for good disinfection*
Turbidity	Absorbs chlorine and protects entrapped bacteria	0.4 NTU or less
pH	Alters the proportions of hypochlorous acid and hypochlorite ion present	pH 6–7.5
Contact time	Allows the oxidizing reactions to take place	30–60 min
Chlorine dose rate	High concentrations will increase reaction rates	0.2–2 mg Cl_2 l^{-1}
Organic contaminants	Reacts to form less effective chloroorganic compounds	Less than 5 mg C l^{-1} as total organic carbon
Temperature	Increases the decay rate of free chlorine to chloride	Up to 20 °C
Ammonia	Reacts to form chloramines	Trace

NTU, nephelometer turbidity units.

Table 3 Factors affecting the efficiency of coagulation/flocculation

Parameter	*Effect on coagulation/flocculation*	*Typical value for coagulation/flocculation*
Coagulant type	Ability to form large polymeric hydrolyzed ions	Aluminum sulfate or iron (III) chloride, polyaluminum chloride
pH	Affects charge destabilization reactions; coagulants become soluble at very high and low pH	pH 6.0–7.5, depends on coagulant used
Coagulant dose rate	Insufficient will cause small floc and ineffective removal, excess may cause dissolution of aluminum/ferric or excess solids	1–10 mg l^{-1} (as aluminum or iron) depending on raw-water conditions
Mixing speed	Coagulation requires immediate contact, i.e., rapid mix; flocculation requires slow mixing to allow particles to coalesce without breaking up	Velocity gradients G, 1000 s^{-1} for coagulation, 20–200 s^{-1} for flocculation
Contact time	Sufficient time is needed to allow reactions to take place	Seconds for coagulation, 5–20 min for flocculation
Organic contaminants	Typically, large molecules with charged surfaces and therefore more difficult to remove	High concentrations will require high coagulant doses and lower pH
Alkalinity	Alkalinity (bicarbonate/carbonate) is used in the hydrolysis reaction. Low-alkalinity water will form weak floc	Greater than 30 mg $CaCO_3$ l^{-1} alkalinity
Temperature	Affects reaction rates	5–25 °C

Solid–Liquid Separation

After coagulation, the flocculated particles must be removed. Removal of gross solids can be achieved with sedimentation or dissolved air flotation and residual smaller particles can be removed using rapid gravity or other filtration techniques.

Sedimentation Flocculated particles are denser than water and will settle readily by gravity. Generally, plants are designed to operate continuously as upflow clarifiers. Water enters at the bottom of the tank at a low-enough velocity (1–3 m h^{-1}) for the particles to settle to the bottom, where the sludge can be removed. The tanks combine the flocculation and removal mechanisms in one plant as flocculation occurs when water entering the bottom of the tank has to travel through the sludge 'blanket' settled at the tank base.

Dissolved-air flotation Algae often have low specific gravity and algae-laden waters are therefore suited to dissolved-air flotation because of their tendency to float. A recirculated stream of clarified water (8–10% of the flow) is saturated with dissolved air under pressure using compressed air introduced into packed pressure vessels. This is returned to the inlet of the flotation tank with the flocculated water where, on releasing the pressure, a curtain of very small bubbles forms as the air comes out of solution. These rise to the surface, attaching floc particles on the way, to form a blanket of sludge that floats on the surface. Periodically, the sludge is removed by scraping the surface. Both sedimentation and dissolved-air flotation will remove coagulant sludge to around 0.2–1.0 mg l^{-1} of aluminum or iron. Sand filtration will be required to remove residual coagulant.

Rapid depth filtration Depth filtration is achieved by passing water that contains particles through a bed of filtration media, usually sand, at relatively high velocity or filtration rate. At a filtration rate of 3–10 m h^{-1}, the particles are drawn into the medium rather than remaining on the surface of the bed. This has two advantages; a smaller structure is required because of the high hydraulic capacity and the pressure drop or head loss is reduced by avoiding the build-up of a blinding layer of particles on the surface. Filters can be constructed as either operating under gravity or enclosed and under pressure. A schematic diagram of a rapid gravity filter is shown in Figure 1.

As particles block the filter, head loss will increase until either flow from the filter reduces or a breakthrough of particles is seen in the filter outlet water as turbidity. Before either of these happens, the filter should be washed. The bed is partly drained and air

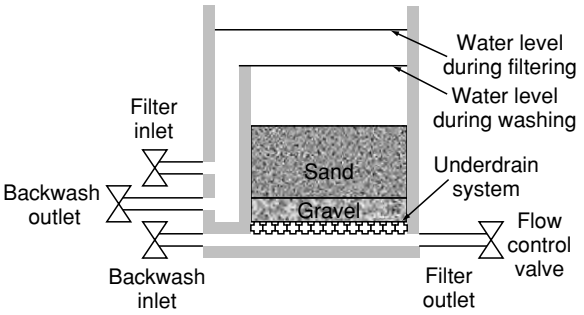

Figure 1 Schematic diagram of a rapid gravity filter.

scoured using compressed air introduced through the underdrain to break up compacted solids. Previously filtered water is then passed upwards, at a velocity of around 20 m h^{-1} for 10 min, to remove entrained solids. Rapid gravity filtration is an important barrier treatment for *Cryptosporidium* and *Giardia*, so close attention is paid to conditions that may cause breakthrough of solids into the filtered water. Improvements in laser techniques to measure particle numbers and size distribution have led to a greater understanding of the correct coagulation conditions, minimization of flow changes through filters, and the ripening phase immediately after a backwash as important factors in controlling the numbers of oocysts in the treated water. The dirty backwash water will contain a higher concentration of oocysts than the raw water and must be disposed of carefully. Often this water is returned to the raw inlet, at a very low dilution rate, following settlement to remove the coagulant sludge.

Slow sand filtration Slow sand filters are used without the addition of coagulant. The lower velocities used mean more particles are retained which in turn encourages biological activity. Within days the layer of organic material on the surface, known as the schmutzdecke, becomes capable of removing color and turbidity and large numbers of bacteria. Principal mechanisms are straining and entrapment but digestion of dissolved organics also takes place by a variety of organisms, such as protozoa, algae, and bacteria living in the schmutzdecke. A comparison of rapid and slow and filtration is shown in Table 4. Slow sand filters are cleaned by removing the surface layer of sand and schmutzdecke.

Membrane filtration Membrane filtration has been in use in the food industry for some decades. Very fine filtration generates a relatively high pressure but produces a very-high-quality product. More recently, improvements in the membrane-manufacturing process and a growing market have made the technology more cost-effective. Increasing public concern over *Cryptosporidium*, a chlorine-resistant, spore-forming organism that can cause severe illness, has resulted in membranes being installed as one of the few effective treatments.

Iron and manganese removal Many groundwater sources can be found that are of good microbiological quality but esthetically unsuitable because of high concentrations of iron and manganese salts resulting in black or brown discolored water. These waters can be successfully treated using filtration techniques. However, the metal salts must be in an insoluble, particulate form to be removed effectively. Soluble metal ions can be converted to particulate forms by oxidation using aeration, chlorination, or ozone. Due to their relative oxidizing power, oxidation can be achieved in minutes with ozone but may take hours with air. Removal of manganese requires a longer reaction time or a stronger oxidant than iron; potassium permanganate or ozone is often used. The oxidation reaction kinetics are dependent on pH and best performed at around pH 8.0.

Organics Removal

The UK and Europe have treatment- and health-related standards for various organic compounds. Contamination can occur from both agricultural and industrial sources. Soluble organic compounds such as herbicides, particularly triazines, chlorophenoxy

Table 4 Comparison of rapid and slow sand filtration

Parameter	*Rapid gravity/pressure*	*Slow sand*
Filter media	Sand, granular-activated carbon or anthracite	Sand
Filtration rate (m h^{-1})	3–10	0.2
Filter run length	24 h	60 days
Cleaning	Air scour and upwash	Surface skim
Mechanism	Physical entrapment in media depth	Physical entrapment on media surface, biological digestion
Pretreatment	Almost always used with coagulation	Microstraining or roughing filtration

acids, and urons, are often found. Industrial solvents, carbon tetrachloride, and trichloroethene (TCE) can also be found to contaminate groundwater. Coagulation and filtration techniques are ineffective but ozonation, granular-activated carbon, and air stripping can be used to remove these contaminants.

Ozone will degrade many pesticides and other organics. Concern for the health effects of the degradation products and the potential for bacterial regrowth mean it is often used with granular-activated carbon to absorb the degraded organics.

Granular-activated carbon (fixed bed) will absorb a number of organic compounds and is commonly used for pesticide removal and for the removal of taste- and odor-causing compounds. It can also be used as a rapid gravity filtration medium, thus combining two processes in one. The carbon has a finite lifetime and, depending on the organic loading and type, will require regeneration. This involves returning the medium to the manufacturer to undergo a high-temperature regeneration process to oxidize the adsorbed organic compounds.

Powdered activated carbon can have similar uses but is dosed into a contact tank or similar structure and is subsequently removed by a downstream separation process such as filtration.

Removing Dissolved Salts

Nitrate Removal

Increasing levels of nitrate in groundwater sources are generating a need for nitrate removal plant in the UK and Europe. Ideally, the problem is best approached by reducing the levels at source, by minimizing the levels of nitrate applied to the land within the water catchment area. There are, in the UK, government-aided grants to identify and assist farmers in controlling the application of nitrate fertilizers.

Where treatment is required, ion exchange and electrodialysis have been used in the past. The principal of ion exchange is to pass the water through a bed of synthetic resin beads, which exchange nitrate ions for chloride. Once the resin is saturated with nitrate, it is regenerated with brine and a waste product containing high concentrations of nitrate and sulfate is produced. Electrodialysis, which is also commonly used for desalination of brackish and sea water, involves applying an electrical current across a membrane that is capable of selectively passing nitrate ions but not water. More recently, with improvements in membrane manufacture, reverse osmosis is becoming a cost-effective solution for nitrate removal.

Softening

Softening is not extensively used in the UK water industry but is used to protect industrial boilers from scaling. Precipitation using lime can be used to precipitate calcium and magnesium. This process produces large quantities of sludge. Alternatives are ion exchange and, more recently, reverse osmosis. Ion exchange resins will replace calcium and magnesium ions for chloride ions and, once saturated, they can be regenerated using brine.

Distribution Systems

Water Quality in Distribution Systems

Water that has been treated at the treatment plant can take days to reach the customer through the underground distribution network of pipes and service reservoirs. To insure that the quality and especially the microbiological safety of the water do not deteriorate on the way, residual disinfectant, pH stabilization, and other corrosivity control processes are used at the treatment works.

A free chlorine residual concentration of around $0.3\,\mathrm{mg\,Cl_2\,l^{-1}}$ will insure bacterial regrowth will not occur for several days. This will however depend on the distance and time of travel, the water temperature, as chlorine will decay more quickly at higher temperatures, the condition and age of the distribution network, and the quality of the water. If the water contains THM precursors, the additional contact time with chlorine in the distribution network can lead to higher concentrations of THMs reaching the customer's tap than found at the treatment works.

Corrosion Control

The chemical balance between carbonates and bicarbonates is especially important in corrosion control. A protective layer of calcium carbonate will form on the inside of distribution pipes if the water is adequately stabilized. This requires the balance of free carbon dioxide, carbonate, and bicarbonate concentrations to be controlled, usually by adjusting the pH to control Eqn (2).

$$CO_2 + CaCO_3 + H_2O = Ca(HCO_3)_2 \qquad (2)$$

Excess carbon dioxide will dissolve calcium carbonate, and cause corrosion to distribution pipework. Carbon dioxide is an acidic gas. If it is removed, for example in an aeration process, it is likely that the pH of the water will rise and calcium carbonate will be precipitated. The corrosivity of water can be estimated using corrosion indices; the most commonly used index is the Langelier saturation

index. The saturation pH is calculated from ionic strength, calcium, alkalinity, and temperature and compared to the measured pH of the water. A positive index indicates the water will precipitate calcium carbonate and a negative index shows that it has corrosive potential. The index can therefore be controlled by adjusting the pH, commonly with lime, as this has the advantage of also adding calcium which reduces the saturation pH. Mineral acid or caustic soda is also used.

Lead plumbing was extensively used in the UK at the turn of the 19th century, but concern over very low levels of lead contamination have prompted the need to replace pipework wherever possible and control lead solubility. This is achieved by dosing phosphoric acid to around 0.2–2.0 mg $P l^{-1}$ as the water enters the distribution network from the treatment works. A protective coating of lead phosphate is formed which reduces the concentration of lead in the water.

Plant Design

In summary, the aim of the design of a water treatment plant will essentially address the following criteria:

1. Establish the difference in quality of the raw source compared to the required treated water quality.
2. Specify process plant, and its operating limitations, capable of producing sufficient quantity and quality to meet customer demand and legislative requirements.
3. Place the plant geographically to insure minimal cost of transport and potential for quality deterioration.
4. Provide a treatment solution at an acceptable capital and operating cost.

See also: **Filtration of Liquids**; **Iron**: Properties and Determination; **Manganese**; **Membrane Techniques**: Principles of Ultrafiltration; Applications of Ultrafiltration; **Nitrates and Nitrites**; **Pesticides and Herbicides**: Residue Determination; Toxicology; **Water Supplies**: Chemical Analysis; Microbiological Analysis

Further Reading

Badenoch J (1990) *Cryptosporidium in Water Supplies*, p. 230. UK: DETR.

Barrett JM, Bryck J, Robin Collins M, Janonis BA and Logsdon GS (1991) *Manual of Design for Slow Sand Filtration*, pp. 5–12, AWWA Research Foundation.

Bouchier I (1998) *Cryptosporidium in Water Supplies*. UK: DETR.

Carrol T and Booker N (2000) *Membrane Technology in Water and Wastewater Treatment*, pp. 93–99. Royal Society of Chemistry.

Drinking Water Inspectorate website. Available online at www.dwi.gov.uk.

Environment Agency web site. Available online at environment–agency.gov.uk.

Hall T and Hyde RA (eds) (1992) *Water Treatment Processes and Practices*. Swindon: WRC.

Hammer MJ (1986) *Water and Wastewater Technology*, 2nd edn. New York: John Wiley.

Kirmeyer GJ *et al*. (1999) *Maintaining Water Quality in Finished Water Storage Facilities*, pp. 137–168. AWWA Research Foundation.

Langlais B, Reckhow D and Brink D (1991) *Ozone in Water Treatment Application and Engineering*, pp. 31–54. Lewis.

Montgomery J (1985) *Water Treatment Principles*. New York: John Wiley.

National Rivers Authority (1992) *The Influence of Agriculture on the Quality of Natural Waters in England and Wales*, 1st edn, pp. 19–23. Bristol: National Rivers Authority.

Pontius FW (1990) *Water Quality and Treatment*, pp. 269–361. Mc-Graw-Hill.

Twort AC, Law FM, Crowley FW and Ratnayaka DD (1994) *Water Supply*, 4th edn, pp. 258–277. New York: John Wiley.

World Health Organization (1993) *Guidelines for Drinking Water Quality*, 2nd edn, pp. 93–106. France: World Health Organization.

Chemical Analysis

P J Walker, Thames Water Utilities Ltd, UK

Introduction

Most of the water supplied for human consumption has to comply with national or international standards with regard to its chemical composition and the concentration of material present. The limits placed on the permitted concentrations of these compounds are not always health-based; some may be set for esthetic reasons or they may be limited by the practicalities of the analytical methods, limit of detection at the time of setting. Chemical analysis of water will provide information about compliance with these standards. It will also assist treatment plant operators to insure that the production processes are working efficiently and that the purity of the water is maintained within the distribution system.

The World Health Organization (WHO) regularly updates and issues drinking-water quality guidelines

with the aim of protecting public health and forming the basis for national water quality criteria. National policy may vary between countries and as a result there are often differences between the limits expressed in the WHO guidelines and those in the national legislation.

In addition to concentration limits, it is becoming more common for national legislation to specify that a quality system must be in place to insure that the results of chemical analysis are valid. Validity in this sense usually means:

- The performance of the analytical method is known and documented for the sample type
- This performance complies with regulatory criteria for limit of detection, systematic, and random error for both analytical standards and samples
- There is documented evidence that there is a system of monitoring instrumental performance and analyst training, and use of formalized methods
- Analytical quality control (AQC) standards are routinely used as additional samples
- There is regular participation in recognized analytical proficiency schemes.

A system where quality control is specified will insure that results from different laboratories can be compared or used in national statistical exercises.

Laboratory standard methods of analysis are published by national agencies e.g., US Environmental Protection Agency (EPA), and the Environment Agency in the UK. They are also produced by the International Standards Organization (ISO), the CEN standards body in Europe, national standards bodies, e.g., BSI and DIN, and also professional bodies such as the American Public Health Association.

Sampling

It is essential that sampling is performed correctly. Failure to sample properly will affect the result and invalidate it. A sample must be representative of the bulk material, it must be taken into the correct container, with an appropriate preservative if required, and the sample container must be clearly and unambiguously labeled.

The sampling procedure is dependent on the analysis required. In most circumstances the sample must be representative of the drinking water delivered to the sample point. To achieve this the water is allowed to flow to waste for a suitable 'flushing period' to clear any nonrepresentative water in the pipes from the supply main to the sampling point. The exception however is when lead analysis is to be carried out. It is essential that the sample collected is representative of the water that has been in prolonged contact with the pipework and any plumbing metals associated with it. In these circumstances no flushing occurs as the sample is drawn directly from the sampling point into the sample container.

The sample container must not affect the sample by adsorption, absorption, or contribution, e.g., leaching or carry-over from previous samples. Glass was the preferred material for sample containers, but the newer plastic materials, e.g., polyethylene teraphthalate (PET), are replacing glass in many circumstances. Metals, inorganics, and some pesticide determinands have all been shown to be stable within plastic bottles. Compared to glass the advantages of plastic bottles are that they are light-weight, more robust, inexpensive, and disposable. Quality control becomes easier in insuring bottle cleanliness as a small representative sample from each delivered batch can be checked before use. As the bottles are used once only and then discarded, there is no problem with carry-over from the previous sample.

Special care has to be taken if the analyte can undergo degradation or reaction between the time of sampling and start of analysis. Most samples are preserved by refrigeration, but dissolved oxygen, sulfide, and polynuclear aromatic hydrocarbons (PAH) all require 'fixing' on-site at the time of sampling. Mercury can be lost prior to analysis and is usually preserved with nitric acid and/or potassium dichromate. Recent work has shown that the addition of gold solution up to $5\,\mathrm{mg\,l^{-1}}$ Au has excellent preservation properties for both glass and plastic (PET) containers. This works well with inductively coupled plasma – mass spectrometry (ICP-MS), but significantly depresses the signal from a cold vapor fluorescence instrument system.

Basic Characterization

By using a suite of simple tests it is possible to characterize water. These tests are pH, electrical conductivity, color, turbidity, alkalinity, and hardness. Additionally, where drinking water has been disinfected with chlorine, the determination of free and combined chlorine is also determined. The organoleptic properties of taste and odor also provide useful monitoring data.

pH

This determinand indicates the acidity or alkalinity of the water. It is measured using a pH meter connected to an electrode combination that is responsive to pH. Calibration utilizes standard pH buffer solutions over the required range. It is usual to supply drinking

water in the range pH 6.0–9.5. Water supplied outside this range is increasingly corrosive.

Electrical Conductivity

The electrical conductivity of drinking water is measured using a pair of electrodes set a specific distance apart. The conductivity of the water is dependent upon the number and type of ions present and will also indicate the amount of dissolved material present in the sample. The meter is calibrated using a standard solution of potassium chloride.

Color

The unit of color is defined by the platinum cobalt method, being the color produced by $1\,mg\,l^{-1}$ platinum as chloroplatinate ion. There is a direct relationship with the Hazen scale.

Color is measured using visual comparison against liquid standards, calibrated colored glass disks, or spectrometrically. With spectrometry the wavelength used determines the result.

The Hazen scale is primarily used in nonlaboratory situations. It comprises a range of colored glass disks, calibrated in color and intensity to match the chloroplatinate colorimetric test. The sample is poured into a viewing tube and matched to the Hazen standard against a white background. A variety of manufacturers produce these calibrated glass disks; ranges of $0–70\,mg\ Pt\,l^{-1}$ are common. The chloroplatinate method involves the preparation of the standard colors from potassium chloroplatinate standard solution.

Color in water is mainly due to free sediment, iron or manganese material, organic matter, and industrial pollutants, e.g., dyes.

Turbidity

This determinand is a measure of the clarity of the water. It is usual to measure turbidity in formazin turbidity units (FTU), following calibration. A variety of techniques exist to measure turbidity but the most popular uses nephelometry, where the intensity of scattered light is measured. It is usual to have the detection at 90° to the light source.

Alkalinity

This general term covers the combined concentrations of hydroxide, carbonate, and bicarbonate present in the sample. Actual concentrations are dependent upon pH, temperature, and total dissolved solids content. A series of nomograms are used to calculate actual concentration.

Alkalinity is determined either colorimetrically or by titration with acid.

Hardness

Calcium and magnesium salts present determine the hardness of the water. Permanent hardness (noncarbonate hardness) is due to the presence of chlorides, sulfates, and nitrates. Carbonate hardness is due to carbonate and bicarbonate salts of calcium and magnesium.

Metals Analysis

A variety of instrumental techniques and colorimetric methods can be used to detect metallic elements. Electrochemical systems, i.e., anodic striping voltametry, atomic absorption – flame or furnace – inductively coupled plasma-optical (OES) or mass spectrometry (MS) are all practical instrumental systems. The correct choice of system is dependent upon:

1. limit of detection required
2. number of samples to be analyzed
3. number and type of metal analyses per sample
4. degree of automation required
5. skill required of analyst

The lowest limits of detection (LoDs) are most likely to be achieved by the use of an ICP-MS, which, when used with an autosampler system, has the potential to analyze large numbers of samples for most metals/metalloids unattended.

However, with ICP-MS, problems arise with sodium, potassium, calcium, and magnesium due to their high concentration in drinking water, and also with a range of interferences developed within the plasma. Use of gold with matrix-matched standards (Na, K, Ca, Hg) allows the determination of mercury (Hg), arsenic (As), selenium (Se), and antimony (Sb). The use of an ultrasonic nebulizer (USN) or specialized nebulizers (e.g., Minehardt) can significantly improve the performance of an ICP-OES instrument, resulting in lower LoDs than the conventional instrument alone. Many ICP-OES instruments have true simultaneous detection for many metals by the incorporation of charge-coupled arrays or similar devices within the detector unit.

Electrochemical methods such as anode striping voltametry can produce results with good precision and low LoDs but they are disadvantaged by being slow compared to ICP use. Improvements have been made by automating sample analysis and the use of multiplace sample holders, but they are not in common use in high-throughput situations.

Mercury, arsenic, selenium, and antimony can be detected using an ICP-MS, but most other flame or plasma systems cannot produce acceptable LODs or precision for these elements. Furnace systems can get low LoDs by multiple additions of analyte to the

pyrolizing component, but this is a slow technique. Mercury is frequently analyzed using the cold vapor fluorescence technique. In this a reducing agent, i.e., stannous chloride or sodium borohydride solutions, reduces the inorganic mercury to its elemental form in solution. Bubbling a gas, usually argon, through this solution strips out the elemental mercury vapor. Detection is usually by fluorescence, although absorption and other systems have been used. Arsenic, antimony, and selenium can be converted to corresponding hydrides by reaction with sodium borohydride solution. These volatile forms can then be directed to flame or optical systems for detection. It is vital to insure that the element of interest is in the appropriate oxidation state, As^{III} Sb, Se^{IV}, prior to hydride formation. Use of an inappropriate oxidation state will lead to a low result or no result being obtained.

In addition to occurring naturally, aluminum salts can be used within the water treatment process to assist coagulation and subsequent settlement. With a WHO limit of 200 $\mu g\,l^{-1}$, aluminum can be analyzed colorimetrically or instrumentally. However aluminum in drinking water can have a detrimental effect on kidney dialysis at concentrations as low as 20 $\mu g\,l^{-1}$ Al. Any method chosen for aluminum should be capable of detecting this concentration where kidney dialysis patients are treated in the supply area.

The metals present in high concentrations – sodium, potassium, calcium, and magnesium – can be determined using flame photometry in addition to flame absorption or ICP-OES.

WHO guidelines for metals are given in Table 1.

Anion Analysis

Ion-selective electrodes, ion chromatography, and many forms of automated and manual colorimetric techniques are used to analyze anionic compounds. Automated methods include air-segmented and flow injection analyzers or discrete systems derived from clinical analyzers. The choice of technique is dependent upon daily sample numbers and range of analysis required.

Table 1 World Health Organization guidelines for metals present in water

Aluminum	200 $\mu g\,l^{-1}$ Al
Antimony	5 $\mu g\,l^{-1}$ Sb
Arsenic	10 $\mu g\,l^{-1}$ As
Barium	700 $\mu g\,l^{-1}$ Ba
Boron	300 $\mu g\,l^{-1}$ B
Cadmium	3 $\mu g\,l^{-1}$ Cd
Chromium	50 $\mu g\,l^{-1}$ Cr
Copper	2000 $\mu g\,l^{-1}$ Cu
Iron	2000 $\mu g\,l^{-1}$ Fe
Lead	10 $\mu g\,l^{-1}$ Pb
Manganese	500 $\mu g\,l^{-1}$ Mn
Mercury	1 $\mu g\,l^{-1}$ Hg
Molybdenum	70 $\mu g\,l^{-1}$ Mo
Nickel	20 $\mu g\,l^{-1}$ Ni
Selenium	10 $\mu g\,l^{-1}$ Se

National limits exist for beryllium, calcium, magnesium, silver, sodium, tin, and zinc.

Ammonia, Nitrite, and Nitrate

These nitrogen-based ions can be determined using all of the above techniques. The calculation of nitrate is usually obtained by subtracting the nitrite result from that of total oxidized nitrogen (TON). Total oxidized nitrogen is determined by the selective reduction of nitrate to nitrite and its determination as nitrite, along with any nitrite in the sample.

Chloride

Chloride can be analyzed by all the above methods plus titration with silver or mercuric nitrate solution. Colorimetric analysis is based on the reaction that liberates thiocyanate ions from mercuric thiocyanate solution. In the presence of iron III ions, the highly colored ferric thiocyanate is formed and this is measured at 480 nm.

Fluoride

Fluoride may occur naturally or may have been added to reduce the occurrence of dental caries. It can be analyzed by all the above techniques.

Sulfate

This ion is usually determined using gravimetry and turbidity following reaction with barium chloride solution to form insoluble barium sulfate. Ion chromatography is also used, as is the colorimetric methyl thymol blue method.

Phosphates

Phosphates exist in a variety of forms in water. The simplest form is orthophosphate, with the increasingly more complex pyro-, meta-, poly-, and organophosphates occurring less frequently and at lower concentrations. Apart from the determination of total phosphorus by ICP, separation into type is dependent upon conversion to soluble orthophosphate by increasingly more vigorous acid/hydrolysis/digestion. Subsequent colorimetric determination of the orthophosphate using the vanodomolybdophosphoric acid method, stannous chloride method, or ascorbic acid method is dependent upon the expected orthophosphate concentration.

Process Byproduct Analysis

The most familiar process byproducts are the formation of the trihalomethane group of organic compounds following the reaction of chlorine with organic material in the water supply. These compounds (chloroform, bromodichloromethane, dibromochloromethane, and bromoform) are normally analyzed by head space gas chromatography using an electronic capture detection (GC-ECD). Other systems, e.g., purge-and-trap (P&T) and GC-MS, have been used.

The more frequent use of ozone as a pretreatment prior to activated charcoal for the removal of complex pesticides has established a need to analyze for bromates, chlorates, and chlorites. These compounds can be formed from the reaction of the ozone with bromide and chloride ions in the supply water. Ion chromatography, usually using a concentration column, is the preferred method of detecting these compounds. A new procedure, which appears to offer very good LoDs, is currently being adopted. It uses postcolumn derivatization with ultraviolet detection with the advantage that all compounds of interest can be analyzed from the same sample injection. Existing methods for bromate require a multistage preparation to remove interfering chloride and sulfate ions. Samples prepared in this way are analyzed using differing chromatographic conditions. For these reasons it is necessary to analyze bromate separately.

Organics Analysis

A large number and range of compounds fall within this area of analysis. They vary from simple molecules to large complex compounds with differing stabilities and solubilities. It is unrealistic to expect one instrument to analyze the whole range of these compounds. The current techniques used are specifically targeted to obtain the best detection performance. Preferred techniques readily utilize technical advances in instrumental techniques, as these become commercially available.

When analyzing for organic compounds present at very low concentrations in water, the first stage of analysis is usually the concentration of the analytes to be determined. Concentration is achieved either by extracting the compound(s) of interest into another solvent – liquid/liquid extraction (LLE) – or on to a specially prepared inert or polymeric material – solid-phase extraction (SPE). In LLE the choice of solvent is critical. It must not interfere with compound(s) being concentrated, nor must it affect the chromatography by obscuring the peaks of interest with the solvent peak. Once extracted, the solvent phase is usually reduced in volume by either evaporation under a gentle flow of inert gas, e.g., nitrogen, or using specialized equipment, such as the Kuderna Danish evaporator.

Where a complex mixture of compounds is being extracted, the relative solubilities and absorption capabilities of each component can differ significantly. These fundamental differences will lead to unequal extraction ratios occurring which, unless corrected, will lead to errors later in the analysis. It is normal in these circumstances to evaluate the analysis of each compound independently so that the percentage extraction ratio is understood for each matrix type. Once the information is available then the constancy of extraction can be monitored by the addition of surrogate standards to each sample prior to extraction. Surrogate standards are usually compounds very similar to the target analyte, differing either by the inclusion of a different substituted chemical group within the molecule, or by the use of deuterated compounds where mass spectrometry is used in the final analysis. By appropriate choice of surrogate standard, the range of extraction efficiencies can be monitored during the extraction and corrected for in the final result.

Internal standards are standards that are added to the extracted material or concentrate and are used to monitor the constancy of the analytical instrument system used. The addition of compound(s) to the extracted material enables the analyst to correct for both random and systematic variations, e.g., partially blocked/worn injection systems, temperature variation, loss of column efficiency, and detector deterioration.

The instrument systems used rely on chromatographic separation using either GC or high-performance liquid chromatography (HPLC). Different detectors are used according to the sensitivity required, but the reduced size and cost of mass spectrometer detectors has led to their increased use, particularly when deuterated compounds are used for surrogate or internal standards. Large libraries of 'most significant masses' for organic compounds have been compiled to assist the analyst to identify compounds when using single-ion mode(s) (SIMs) of MS-detector operation. Two or more 'significant masses' can be used specifically to identify and quantitate a compound, particularly when it has one mass fragment value that is identical to those produced by other different compounds. The use of total ion count (TIC) mode on an MS-detector will produce a more conventional chromatogram of the separation, showing peak height against time.

Typical organic compounds determined are described below.

Phenols

These compounds have detectable tastes at low concentrations and may react with chlorine to produce chlorophenols which have even lower taste thresholds. They are analyzed by LLE using methylene chloride and detected by GC using FID, ECD, or MS detectors as appropriate. Derivitization using pentafluorobenzylbromide produces lower detection limits.

Trihalomethanes (THMs)

These are normally produced when chlorine reacts with naturally occurring humic or fulvic material during disinfection. Industrial contamination can add THM and other chlorinated hydrocarbons to the raw water. THMs are analyzed by head space or P&T techniques, followed by GC-ECD or GC-MS.

Pesticide Residues

Organochlorine (OCP) or organophosphorus (OPPs) can be analyzed together with polychlorinated biphenyls (PCBs) using methylene chloride LLE and GC-ECD or GC-MS instruments.

Phenoxy Acid Herbicides

MCPB and 2.4.D are analyzed for by extraction into diethylether and derivitized using diazomethane to produce their methyl ethers. Analysis of the derivitized compounds is by GC-MS.

Triazine and Uron Herbicides, Including Atrazine, Diuron, and Prometryn

These compounds tend to be thermally unstable and are best suited to analysis using liquid chromatography-MS detection following either methylene chloride LLE or solid-phase extraction using cartridges preprepared for this specific extraction.

Polynuclear Aromatic Hydrocarbons

This class includes the most frequently found compound – fluoranthene – and the most toxic – benzo(a)pyrene. The presence of these compounds is mainly due to the coal tar pitch lining applied to water distribution pipes. The limits applied to these compounds are health-related. Analysis is by HPLC.

Organic Carbon

This empirical determinand is frequently used as a coarse monitor of quality in both supply and treated water. There are many different techniques for analysis, including combustion furnace or chemical oxidation, with or without ultraviolet light to give carbon dioxide, and reducing conditions to give methane. The detection system is appropriate to the preparative stage, e.g., infrared detectors for CO_2 and FID for methane. The term 'total organic carbon (TOC)' is the summation of the various forms of organic carbon capable of being distinguished by differing preparative stages.

1. Dissolved organic carbon (DOC) – not retained on a 0.45-μm filter
2. Particulate (POC) – the fraction retained as 0.45-μm filter
3. Purgeable or volatile (VOC) – the fraction removed by gas stripping
4. Nonpurgeable (NPOC) – the residue after gas stripping

Radioactivity

The determination of radioactivity is sometimes required due to geological or artificial circumstances. It is rare for there to be a requirement to investigate specific radiochemical isotopes, although the monitoring of gross α- and gross β-emitters is used for establishing and monitoring trends. WHO recommends a limit of 1 Bq l^{-1} for α- and 100 Bq l^{-1} for β-emitters. (A becqueral (Bq) is one disintegration per second.)

Samples can be analyzed directly or in most cases concentrated by boiling. Boiling removes any radon present so this must be determined separately. Final sample preparation is dependent upon the limit of detection required and the instrument/detector used to count the sample. Counting times vary between 3 and 10 h.

Quality Control

Quality control is very important if the results obtained are to be credible and useful for comparison. Analytical quality control is only a part of the total system which should include information on the actual performance of the instrument system used at the time of analysis, the status of the analyst, i.e., trained or under training, the performance of the technique on various sample types or matrices, and the conditions under which the analysis was made.

The methods used to establish and maintain analytical quality control include the use of standards as samples, analysis of duplicates, measures of drift, and others, as appropriate. These values are plotted on charts which show when the analysis is moving away from the target value in a regular way or when the results are outside the statistically established method performance. Shewhart and cusum charts are both used to determine if a procedure is within or outside control.

See also: **Aluminum (Aluminium)**: Properties and Determination; **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **Fluoride**; **Legislation**: International Standards; **Nitrates and Nitrites**; **Pesticides and Herbicides**: Residue Determination; **Phenolic Compounds**; **Phosphorus**: Properties and Determination; **Polycyclic Aromatic Hydrocarbons**; **Quality Assurance and Quality Control**; **Water Supplies**: Water Treatment; **World Health Organization**

Further Reading

American Public Health Association (1998) *Standard Methods for the Examination of Water and Wastewater*, 20th edn. American Public Health Association.

Fatemian, Allibone and Walker (1999) *Analyst* 124: 1233–1236.

Fatemian, Allibone and Walker (1999) *Journal of Analysis and Atomic Spectrometry* 14: 235–239.

HMSO (1976–1992) *Index of Methods for the Examination of Waters and Associated Materials*. London: HMSO.

World Health Organization. www.who.int

Microbiological Analysis

C Fricker, CRF Consulting, Reading, UK

Introduction

The dramatic reduction in disease associated with the consumption of drinking water over the last century has been brought about largely by increased awareness and by the considerable advances made in treatment, disinfection, and distribution of potable supplies. The microbiological and chemical examination of water plays an important part in the control of water treatment processes. Microbiological examination of potable water is performed to insure the safety of supply. In general, the tests used are based on indicator organisms, the presence or absence of which provides a measure of the microbiological quality of water. This article deals with aspects of microbiological examination of water from sampling through to detection of some organisms that are important pathogens for human beings.

Sampling and Handling of Samples

Sampling issues have tended to receive less attention than water quality issues in the various standards and regulations that have been laid down. The recommended frequency of sampling varies widely between different regulatory authorities. In designing a microbiological monitoring program, the major consideration is of course the health of the consumer. The program should be designed to take account of the size of the distribution network and should insure that samples are taken from all areas of the network. The frequency with which a particular supply needs to be sampled depends primarily on the number of consumers it supplies; the higher the number of consumers, the more frequently sampling should take place.

An ideal sampling regime should be completely randomized. However, owing to the complexity of many distribution systems, this is seldom possible. For this reason, stratified random sampling is probably the best approach. The entire distribution should be divided into discrete areas such that it is composed of regions where contamination could occur without being detected by sampling other areas. However, whilst this approach may be ideal in theory, it is seldom, if ever, adopted. In general, the sampling of potable water supplies begins with the water entering the distribution system. Samples from large water treatment plants are usually taken on a daily basis, whilst at smaller works regular monitoring of residual chlorine levels, together with less frequent microbiological checks, is often used. In addition to samples of water entering the supply, other sites such as service reservoirs and break-pressure tanks are regularly sampled. Samples from distribution are randomized as much as possible and this also involves sampling from consumers' premises.

The type of sample taken depends on the reason for sampling. Generally bacteriological sampling (for detection of *Escherichia coli* and coliforms) involves the collection of a relatively small volume of water (300–500 ml), whereas samples for detection of specific pathogens such as enteroviruses, *Cryptosporidium*, *Giardia* and *Salmonella* involve larger volumes (10–1000 l). Samples for bacteriological and virological analysis require sterile bottles and the addition of sodium thiosulfate. The sodium thiosulfate is used to neutralize the effects of free chlorine in the water sample. Samples for parasites such as *Cryptosporidium* and *Giardia* may be taken by passing large volumes of water through cartridge filters and the filters are then taken to the laboratory for examination. Alternatively, smaller volumes of water (10–100 l) may be collected into containers and transported to the laboratory for subsequent concentration. Microbiological samples should be examined as soon as possible after collection. This is particularly important for samples for bacteriological and virological examination, where ideally samples should be analysed within 6 h.

Colony Counts

Water contains a variety of bacteria that may have different temperature optima. Bacteria whose natural habitat is cold water will, in general, have a lower optimal growth temperature than those that enter the water distribution system as contaminants. The temperatures most often employed for providing colony counts on potable water are 22 °C and 37 °C. Generally, plate counts are performed as pour plates where a small amount of water (usually 1 ml) is transferred by pipette into a Petri dish and molten agar (at about 47 °C) is added. After mixing, the agar is allowed to solidify and the plates are incubated at the appropriate temperatures for 1–7 days before the colonies are counted. The agar medium used varies between different laboratories, but a general-purpose medium such as yeast extract agar is most often employed, although the less nutritious R_2A is becoming more widely used, as it usually allows detection of a higher number of bacteria. It is sometimes necessary to prepare serial 10-fold dilutions of samples to insure that an accurate plate count can be achieved. It is not appropriate to record plate counts that exceed 300 colonies on a plate, as the results are likely to be highly inaccurate. Plates with this number of colonies are likely to give a count that is lower than the true count owing to two or more bacteria forming a single colony and because colonies will tend to lie beneath one another. Furthermore, some bacteria produce bacteriocins which will inhibit the growth of other organisms. The resultant plate counts are by no means a measure of the total count of bacteria in a water sample but merely indicate the number of organisms that are able to form colonies in the particular medium used under specified growth conditions.

Colony counts are not essential for assessing the safety of potable water, but may be useful operationally and may give an early indication of future problems. The value of colony counts lies in the comparison of a number of counts from an individual source taken over a period of time. When subjected to trend analysis, any significant changes may indicate problems. A sudden increase in the plate count at 37 °C may indicate a pollution incident. Plate counts at 22 °C are subject to environmental and seasonal influences. Plate counts are also particularly useful in determining the efficiency of water treatment processes such as filtration, coagulation, and chemical disinfection.

Use of Automated Methods

The development of automated procedures for the microbiological testing of water samples has been slow. To date, very little has been implemented. Some automation of colony counts has begun and it is not uncommon for laboratories to produce pour plates using commercially available plate pourer/stackers. Water is added to the Petri dish, a defined volume of agar at the appropriate temperature is added and, after mixing, the plates are cooled to allow the agar to solidify. It is also possible to utilize automatic colony counters to perform plate counts. Colony counters composed of a solid-state video camera coupled to sophisticated electronics and software are often used. Such instrumentation can lead to considerable time savings in a busy water microbiology laboratory. New systems are being developed which allow the enumeration of bacteria in a sample without the need to grow them. Such systems employ the use of viability markers such as fluorescein diacetate which is converted to a fluorescent product by viable bacteria. This fluorescent product which is held inside the bacterial cell wall allows sophisticated instrumentation to enumerate the bacteria present on a membrane surface.

Special Problems Associated with Viruses

All warm-blooded animals may harbor enteric viruses and therefore have the potential for contaminating water through defecation. In addition, other viruses from other sites such as the respiratory tract may be found in sewage and therefore may pass into source waters. Fortunately, most viruses are extremely host-specific and will infect only a small number of species. (*See* **Viruses**.)

Human enteric viruses, which are now known to occur frequently in raw water, were first isolated from the environment over 50 years ago. This group includes viruses such as the polio virus and hepatitis A virus. Whilst a direct association between the presence of virus in drinking water and human disease has not been proved for all enteroviruses, their presence in drinking water should be seen as a public health hazard.

Viruses are small in size and may travel long distances through soil. Consequently they may be found in sites where bacteria have been filtered out. For this reason, the absence of indicator bacteria from water does not mean that viruses are absent. It is therefore appropriate to examine water for viruses in some instances. This creates several problems. Because viruses are intracellular parasites incapable of replication outside a host cell, they cannot be cultured on artificial culture media as can most bacteria. Viruses must be grown in tissue cultures, which are cells, usually derived from animal or human tumor cells, which can be grown in the laboratory. Because of

their host-specificity and particular tissue tropism, an individual class of virus will only grow in certain types of tissue culture and some viruses grow very poorly or not at all *in vitro*. In practice, very few methods exist for the detection of viruses in water. Examination for enteroviruses (specifically the polio and Coxsackieviruses) is undertaken most frequently as the methods for these viruses are simple and well established. These viruses are used as indicators of the potential presence of other viruses in much the same way as *Escherichia coli* is used as an indicator of the potential presence of pathogenic bacteria.

The procedure used for the detection of enteroviruses in water involves concentration of the viruses from a large volume of water (usually a minimum of 10 l) on to an electropositive membrane, followed by elution using a small volume of a proteinaceous solution such as beef extract. The concentrated material is used to inoculate cultured cells, which are then incubated at 37 °C for several days. Viruses are detected as holes or plaques in the cell sheet which are created by the virus killing the cells in a small area. Some viruses are not able to cause this so-called cytopathic effect, although they may be able to replicate to some extent within the cell. These viruses are generally detected using enzyme-linked or fluorescently labeled antibodies. More recently, molecular methods based upon the polymerase chain reaction (PCR) have been introduced and these are usually applied to tissue cultures a few days after addition of the sample, to insure that only infectious viruses are detected. (*See* **Immunoassays**: Radioimmunoassay and Enzyme Immunoassay.)

Parasite Problems

In recent years *Cryptosporidium parvum* has become recognized as a major pathogen of human beings and some domestic animals. It causes a self-limiting diarrhea in immunocompetent individuals, but may cause mortality in immunocompromised hosts. One of the modes of transmission is through consumption of contaminated water and several waterborne outbreaks have been reported. The organism survives in water as an environmentally robust oocyst containing four sporozoites. The organism has been reported in drinking water and it survives chlorination during normal water treatment. (*See* **Parasites**: Occurrence and Detection.)

Examination for Specific Groups

Much of the microbiological work within the water industry involves the use of indicator organisms as opposed to specific groups of bacteria or viruses. The organisms most frequently sought in this connection are coliform and *E. coli*. The term coliform has been given several definitions, but in general, a working definition is that it produces P-galactosidase and grows at 37 °C on laboratory media containing bile salts. The definition of *E. coli* is of course much clearer, it being a species in its own right.

Both *E. coli* and coliforms are cultured in the same way. Water is passed through a membrane filter and placed on the surface of a pad soaked in medium containing proteinaceous nutrients, lactose, bile salts (or similar compounds), and a pH indicator. The combination of lactose and an indicator allows the identification of lactose fermenters (a trait common to coliforms and hence to *E. coli*), whilst the bile salts inhibit the growth of nonintestinal bacteria. Nowadays bile salts as such are not generally used, and they have been replaced by purified detergents such as Teepol or sodium lauryl sulfate. The membranes are incubated at 37 °C for coliforms or 44 °C for *E. coli* and incubation is continued for at least 18 h. The incubation often includes a 4-h period at 30 °C to allow organisms that may be slightly damaged to adapt to their new environment. After incubation, the membranes are examined and lactose-fermenting colonies are counted on membranes incubated at 37 °C and 44 °C. These counts are termed 'presumptive coliforms' and 'presumptive *E. coli*,' respectively. Confirmation of the presumptive counts was based on production of acid and gas from lactose, and indole from tryptophan at their respective temperatures (37 °C and 44 °C). However, it has been known for some time that these confirmation tests are not 100% effective because of differences between strains within the groups (e.g., anaerogenic strains). It is becoming a more common practice to identify the isolates by means of more detailed physiological tests, often using one of the commercially available batteries of tests for identification of the Enterobacteriaceae. This permits a much more accurate 'confirmed' result.

The use of simpler tests for the detection of coliforms and *E. coli* is becoming more common and in particular defined substrate technology is extremely popular. These tests rely on the detection of two enzymes, β-D-galactosidase (for coliforms) and β-D-glucuronidase (for *E. coli*) in a medium which contains little carbon and energy source other than the substrate for the two enzymes. Enzyme activity is detected by a color change (for coliforms) and the production of fluorescence (for *E. coli*).

Other groups of indicator organisms which are sometimes used include the enterococci and *Clostridium perfringens*. These organisms are generally more resistant to chlorine than the coliform group

and also tend to survive longer in the water. They may be used to identify waters which have been subjected to less recent fecal contamination. Isolation of these groups of organisms generally involves membrane filtration, followed by incubation on an appropriate selective medium. Specific enteric pathogens such as *Salmonella* spp. and *Campylobacter* are rarely looked for in potable water. However, when they are sought, techniques similar to those employed in food microbiology are used. (*See* ***Campylobacter***: Detection; ***Clostridium***: Detection of *Clostridium perfringens*.)

The two major enteric pathogens that are looked for specifically in water are the protozoan parasites *Giardia* and *Cryptosporidium*. Examination for these organisms involves a very different procedure to those used for bacteria and viruses. Large volumes of water (typically 100–1000 l) are filtered through cartridge filters. The filter matrix is then 'washed' in a dilute detergent solution in an effort to release as much of the retained material as possible. The washings from the filter are then concentrated by centrifugation, purified by immunomagnetic separation, dried on to microscope slides, and stained by immunofluorescently-labeled antibodies. Examination of water samples from 'at-risk' water treatment plants is now a regulatory requirement in the UK and a specific method has been published which must be used for examination of these regulated samples. The process involves concentration of large volumes of water, separation of *Cryptosporidium* oocysts from background debris by immunomagnetic separation, and epifluorescence microscopy. A similar method (USEPA (United States Environmental Protection Agency) method 1623) has been recommended for use in the USA.

See also: ***Campylobacter***: Detection; ***Clostridium***: Detection of *Clostridium perfringens*; **Immunoassays**: Radioimmunoassay and Enzyme Immunoassay; **Parasites**: Occurrence and Detection; **Viruses**

Further Reading

Backer H (1997) Water disinfection strategies for. *Cryptosporidium*. *Wilderness and Environmental Medicine* 8: 96–100.

Cho HB, Lee SH, Cho JC and Kim SJ (2000) Detection of adenoviruses in tap water and river water by reverse transcription multiplex PCR. *Canadian Journal of Microbiology* 46: 417–424.

Craum GF and Frost FJ (2002) Possible information bias in a waterborne outbreak investigation. *International Journal of Environmental Health Research* 12: 5–15.

Gofti L, Zmirou D, Seigle Murandi F, Hartemann P and Potelon JL (1999) Waterborne microbiological risk assessment: a state of the art and perspectives. *Vérité Epidémiologie et Santé Publique* 47: 61–73.

Karanis P and Kimura A (2002) Evaluation of three flocculation methods for the purification of Cryptosporidium parvum oocystis from water samples. *Letters in Applied Microbiology* 34: 444–449.

Weaning *See* **Infant Foods**: Milk Formulas; Weaning Foods; **Infants**: Nutritional Requirements; Breast- and Bottle-feeding; Weaning; Feeding Problems

WEIGHING

R Schulthess, Mettler-Toledo AG, Greifensee, Switzerland

Introduction

Well-organized and effective quality assurance in the food industry is a must for all companies, whatever their size. A quality assurance system has to meet demands from both the legislative and the industrial economic standpoint. Scales and weighing systems represent important tools to enable such quality assurance systems to fulfill these requirements.

The Scale as a Control and Management Tool

The degree of automation of all types of production processes must be constantly increased to withstand the growing pressure from the competition. At the

same time, the demands on quality are increasing, tolerance limits are becoming narrower, and integrated control processes are assuming more and more importance. This makes the use of 'intelligent' (computerized) scales and weighing systems as tools in quality assurance an important factor in production. (*See* **Quality Assurance and Quality Control.**)

Sometimes, only a few grams or even fractions of a gram decide the quality and the profit and loss in mass production. Even minor weight deviations, e.g., losses incurred through the overfilling of solid or liquid foods, can cost a great deal of money. Difficulties must also be anticipated with underfilling. In addition to customers, consumer and other protection associations, which react extremely sensitively to underfillings, each country has its own statutory regulations which, particularly in the food-processing industry, require a weight control.

The Scale: Electronics as Standard

In the last two decades, balance and scale construction has undergone a fundamental changeover from mechanics to electronics. As a result, electronic balances and scales are now standard equipment. They offer the user incomparably more convenience than their mechanical counterparts, and thus make possible extensive rationalization of operations. Electronic scales, however, are also intelligent scales, which offer a wide variety of possibilities to process the weight values further.

Thus weight determinations in different units are currently performed on one and the same scale; percentage weighings and formula weighing (net total determination) are part and parcel of the daily work routine. Furthermore, electronic scales are system-compatible, i.e., they can be incorporated in automated operational routines in a wide variety of ways.

The Measurement Technology: Decisive Factor for the Performance Class of a Scale

Electromagnetic Force Compensation: Synonym for Extremely High Accuracy

At the start of the 1970s, the balance and scale manufacturer Mettler (Switzerland) used the principle of electromagnetic force compensation commercially for the first time. The measurement principle is shown in Figure 1. With this measurement technology, the force action on the weighing platform is compensated by means of a coil in a magnetic field. The current needed for this is proportional to the weight of the loaded sample. It is transformed into a digital signal that can be processed by a powerful microprocessor. This principle allows an extremely high resolution greater than 300 000 certified points (e) and the maximum possible precision, even in surroundings not conducive to highly precise work.

The most conspicuous feature of these scales is the drastic reduction in the number of controls. A single key is all that is needed to initiate all functions, such as on–off switching of the scale, setting the display to

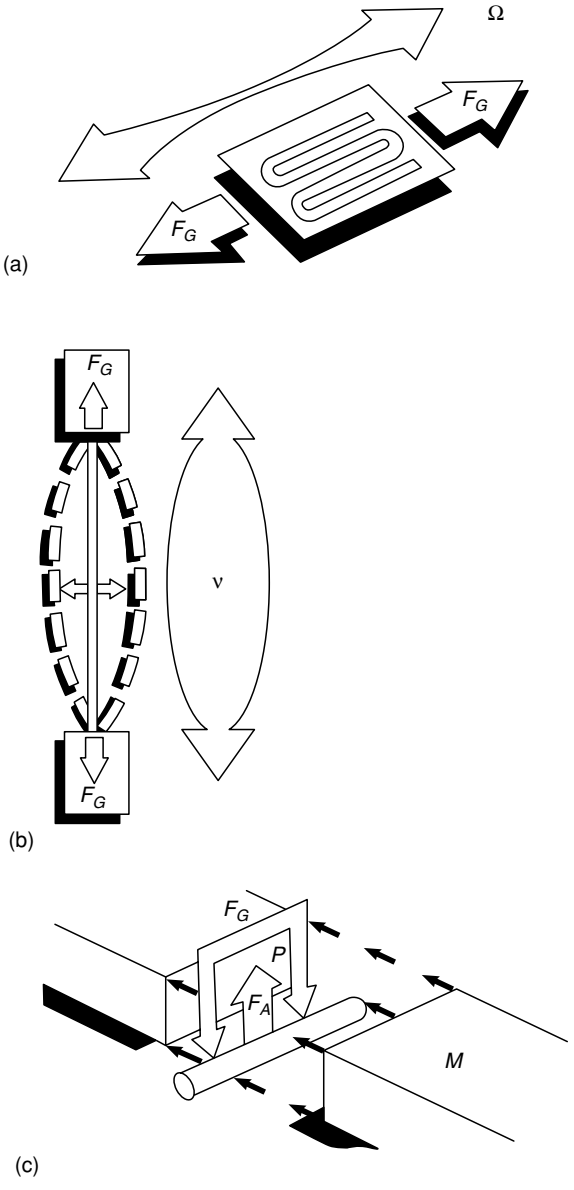

Figure 1 (a) The principle of frequency-modulated force measurement. (b) The principle of strain-gage technology. (c) The principle of electromagnetic force compensation. Reproduced from Weighing, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

zero, taring a vessel on the scale, and selection of additional functions from the menu (Figure 2).

Strain-Gage Technology for the Harsh Industrial Environment

Weighing stations in a harsh industrial environment must be characterized by ruggedness and endurance. With scales employing strain-gage technology, the strain gauge, an electrical resistive film, is mounted on an elastically deformable body and thus allows leverless measurement of loads weighing up to several tonnes. This technology, with its various designs, e.g., extremely flat platforms for wheeled access, and particularly flat ring-load cells, offers weighing results of unvarying accuracy in the lower resolution range at 3000 e (Figure 1).

Frequency-Modulated Force Measurement – Precise and Economically Attractive

The combination of extremely simple mechanics (monobloc) with state-of-the-art sensor technology (vibrating string) allows a very good multirange resolution. A value of 3×3000 e assures flexible use and offers advantages in a wide range of applications, especially in situations where the industrial ruggedness of the sensor (IP67-type protection) is essential. The frequency-modulated force measurement is distinguished by a very high accuracy and, at the same time, by its attractive price: performance ratio (Figure 1).

All three basic weighing technologies described above are used in commercial instruments. The user thus has available a balance or scale product line up to 6 t that is geared to practically all conceivable applications.

Automatic Calibration

To ensure that 100 g is actually measured as 100 g the world over, reference weights (so-called calibrated mass weights) are used with a mass based on the standard stored in Paris. With traditional, mechanical two-pan scales, each weighing was of necessity a comparison with the loaded reference weights. With electronic scales, however, a comparative measurement is performed only when the scale is adjusted on-site with a known weight. This process is called calibration.

Figure 2 Modern electronic balance with a single control bar. Reproduced from Weighing, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Accurate Results – Even Under Unfavorable Ambient Conditions

Anyone who performs weighings wants more than merely accurate results. They should also appear rapidly on the display and be stable. However, rapidity and stability are properties which are mutually inconsistent. For example, an analytical balance with 0.1 mg read ability reacts sensitively to loaded weights, but vibrations of the support and draughts can also influence the displayed weight values. The effects of building oscillations, open windows, and air-conditioning units with their powerful air currents must therefore be effectively suppressed. The measured values were summed over an adjustable time, the so-called integration time, and shown as a mean value. However, this averaging method is associated with a relatively long stabilization time.

Modern technological developments allow better solutions. The built-in processor system performs a trend analysis so that the actual weight value can be displayed rapidly despite a less than ideal location.

Display Behavior Meeting Application Demands

To match the requirements of the application, the behavior of the display can be adjusted with the so-called weighing process adapter: the last decimal place in the display is switched off during rapid weight changes until the weight has gradually stabilized. This prevents the operator being distracted and quickly becoming fatigued by rapidly changing figures that are completely superfluous at the time.

In fine addition, on the other hand, the display should follow the weight increase as rapidly as possible and show the weight continuously with its complete resolution.

Yet another, completely different behavior is demanded from the scale in check weighings. The weight result should be displayed as quickly as possible, and this is allowed for by filters. Intermediate values are of no interest, and hence the display shows all decimal places only when the actual weight has been determined.

Digital or Analog?

Digital numbers are still indispensable for the representation of extremely accurate measured values, but

digital numbers are not geared to the human senses. Our feeling for changing quantities is virtually nonexistent.

Even today digital gages, especially speedometers in cars, have gained little acceptance as it takes too long to read and acquire the measured value. For this reason dynamic analog displays have been developed. Similar to a clock with hands, this angular information supplements the digital values of the display (Figure 3).

Single Range, Dual Range, DeltaRange, MultiRange

The, at first sight, confusing multiplicity of terms hides a number of different advantages, depending on the scale type, which prove their worth in the appropriate application.

In principle, everything could be weighed with the single-range scale. However, owing to the often very high resolution, such scales take a relatively long time to achieve stability and are expensive to manufacture. Selection of the correct scale leads to the best and most favorably priced solution. The concepts can be characterized as follows.

Figure 3 DeltaTrac analog and digital display. Reproduced from Weighing, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Single-range scale provide the same readability over the entire weighing range, e.g., 0–200 g to an accuracy of 0.1 mg.

Dual-range scales have a range with a finer readability, e.g., 0–50 g to an accuracy of 0.01 mg for the weighing of light samples, and for the weighing of heavy samples a range of, for example, 0–200 g with a coarser readability of 0.1 mg. Switching between the ranges can be performed manually or automatically as a function of the loaded sample.

DeltaRange scales have a more accurate fine range which covers part of the entire weighing range and can be shifted at will. If this fine range is exceeded, the scale automatically switches over to the coarser range. By pressing the tare key, however, the start of the fine range can be reset to any position over the entire weighing range. This device is extremely useful primarily when several components have to be weighed very accurately into a relatively heavy tare vessel. After every taring, the fine range is again available.

Finally, multirange scales change the readability in a variable manner. Thus the numerical increments of the display increase from one to two to five increments with increasing load. For example, the multirange KA32s weighing platform has a range of 0–6 kg to an accuracy of 1 g, followed by 6–12 kg to 2 g, and 12–32 kg to 5 g.

Electronic Data-Processing (EDP) Attachment

The time-tested laboratory notebook can still be found next to the majority of scales. However, anyone who has discovered how simple it is to print out weight values or transfer them to a computer will need a lot of persuading to relinquish such conveniences.

Two interfaces with different standards and functions are steadily gaining acceptance as worldwide standards. RS232C is the most widespread standard for attachment to personal computers and printers. It allows data communication in two directions: to the computer and from the computer to the scale. Current loop (CL) is a standard interface that allows distances of up to 1000 m to be covered without any problems, which is not possible with an RS232C connection. This enables several scales and peripheral units, even those at different sites, to be linked and controlled and monitored by a central computer.

Matched to Harsh Surroundings

In the food sector, scales and weighing systems are often used in harsh ambient surroundings: wet conditions and dust, acids and salt, and other corrosive

substances can influence their functional dependability. Despite the harsh conditions, they must offer absolute reliability.

The use of electrolytically polished chrome–nickel steel and other high-grade materials ensures operational reliability even in adverse surroundings.

A high degree of protection (e.g., IP67) assures protection of the sensitive instruments and particularly that of the electronics against the ingress of water, moisture, and dust.

The Hygiene Factor

In the food-processing industry, hygiene is one of the central problems facing a plant. All production equipment must be appropriately designed to ensure strict compliance with the hygiene regulations. Thus the construction of the scales and weighing systems described above also takes into account the stringent requirements with regard to cleanliness. The instruments can be thoroughly cleaned once or several times a day in accordance with the regulations without any problems and without being damaged, whether this involves hosing down, the use of powerful cleaning agents, or other treatments. (*See* **Cleaning Procedures in the Factory**: Overall Approach.)

Increase in Efficiency Linked with Cost Savings

Modern material flow concepts can no longer be imagined without electronic scales and weighing systems. As integratable components, they fulfill a wide variety of functions over the entire goods flow from the incoming goods via production, storage, quality assurance, research and development, up to the outgoing goods. Thanks to its system capability, the scale can be integrated without any problems in fully automated and EDP-controlled production and flow processes.

The high accuracy, the extreme user-friendliness, and the highly diverse application capabilities make modern scales and weighing systems an essential factor to increase efficiency and optimize quality and costs.

The Selection of the Right Material Flow Scale

What are the essential points to consider in the use of scales in material flow? Basically, a distinction must be made between features specific to the scale and those specific to the application.

Bar Codes in Weighing Technology

The use of the bar code technique is now making headlong progress in general industry, and additional bar codes, superficially known as industry codes, have been developed to cover these needs. The main use lies in conveying and handling, as well as warehousing; here the material flow is tracked. In all phases of the transport process, attempts are being made to identify automatically the goods to be transported or stored i.e., by electrooptical reading of the bar code symbols on the packaging, transport containers, accompanying documents, etc., and to transfer the data to process or normal computers and perform automatic crediting and debiting by way of a permanent inventory.

Consideration of the scale as a sensor in material flow assumes increasing significance if the following simple application is pictured: read in the identification data of the weighing sample or sample to be counted directly at the weighing terminal from the bar-coded medium and, supplemented with the weighing result, transmit to an external data receiver via a data interface; alternatively, print out weighing results in bar code form for the rapid and reliable acquisition by bar code reading at the destination.

The features specific to the scale include properties such as industrial compatibility, weighing range and accuracy, user-friendliness, and system capability. The application-related characteristics depend on the location and the actual application. A number of companies offer a whole series of software packages with standard applications, such as weight controls (plus or minus checks), automatic weighing in the material flow, totalization, formula weighing, dispensing, and filling process control. The requirements of the specific departments must also be taken into consideration.

The entire inventory management of the complete chain stands or falls by the precise acquisition of weights and batches in the incoming goods department. Check weighings and data acquisition are of prime importance, whether they involve stationary or mobile, line-independent scales.

In production, the broad spectrum of applications proves its full worth. Here, the production performance has to be supported by efficient control systems. Time-tested standard applications provide assurance in checking and classifying, and in check weighing in the material flow. The use of EDP allows the weighing data to be processed further, and the function of the scale as a sensor to maintain quality and comply with quantity can be utilized to the full. In addition to high quality, the goal is always an increase in efficiency, whenever possible linked with a reduction in costs.

It is impossible to imagine modern warehouses and warehouse management without electronic scales and weighing systems. The precise acquisition of weight

data, piece counts, etc., provides a reliable basis for stock planning and logistics.

In the outgoing goods department – the last link in the production chain – scales again make a contribution by ensuring that with precise check weighings the preceding efforts in the entire material flow are not for nothing. Particularly in the food industry, rapid flows in this department are decisive to assure quality and freshness of the products, but without endangering the economic efficiency of the production.

From the Production to the Front

What is right for production should be proper for commerce. Even at the front, in the retail trade, the picture has changed drastically in the past few years. Scales with price-marking systems are today the order of the day, as is the linkage to form complete network systems in supermarkets. Similar to the situation in production, the scale network distributed over the selling space is centrally monitored and controlled. In the field of price marking and label design, all conditions required by law and economic efficiency can be met: bar code, layout, article data, up to the design of advertisements for the business in question.

Electronics also on the Road

In the era of the EC and the increase in the exchange of goods by road, it is essential to acquire complete delivery and dispatch data. Several vehicle scales employ the principle of strain-gage technology and have the same acquisition and processing possibilities of the weighing data as all other scales.

Filling Process Control in the Food Industry

One of the most important application areas of weighing systems in the food industry involves filling process controls. For here, as in all types of filling processes, the tolerance limits must be defined so that neither unnecessary losses owing to safety overfillings nor any violations of the legal regulations owing to underfilling occur.

The following example will illustrate this point. In the Hamburg factory of Tchibo, a major coffee-roasting company, every year 32 000 t of freshly roasted coffee beans are weighed to the nearest bean and immediately packed. Approximately 580 Tchibo branches and 15 000 fresh product depot partners assure the customer of the proverbial Tchibo freshness, in other words top-class quality and an unvarying fill quantity.

The filling machines run in a three-shift operation with a capacity of 100–120 packages per min and a net weight of 500 g. This high filling rate notwithstanding, the weight of every package must be exactly right. There are two contrasting reasons for this. First, the legislative authorities or the weights and measures office checks that the prepackaged goods regulations are complied with and the average weight is not less than the nominal fill quantity. Second, constant overfilling 'on the side of safety' would be hugely uneconomical with the amount that is processed. In order to handle these two problems, Tchibo has installed a filling process control system which operates on the basis of random sampling. At the 13 packaging lines, per day 3900 random samples are taken and processed by the system using statistical methods.

Use of this filling process control system ensures that not only are the legally specified tolerances complied with, but also approximately 40 kg of coffee is saved every day. Scaled up to the annual production, this means a saving of around 10 t. Assuming the price of green or raw coffee is \$1000 per t and taking into account the tax on coffee in Germany, at present around \$2 per kg, this results in a sum of approximately \$30 000 per year. This figure shows that a well-functioning control system pays for itself many times over after a short time.

The Filling Machine – A Complex System

As a rule, a filling machine is a complex entity. Various influences, which can be controlled only after a great deal of effort, lead to random fluctuations in the weight or volume of the packages. The magnitude of these fluctuations depends on the filling principle, the condition of the filling machine, on the packaged product, and numerous other factors.

Independent of these fluctuations, in the guise of the nominal fill quantity, a lower tolerance limit is more or less rigidly fixed by appropriate statutory regulations. Simple compliance with this lower limit makes a control of the filling process necessary; even with apparently large safety everfillings, a check must be made to ensure that they are actually large enough.

On the other hand, from the industrial economic viewpoint, the interest of the filling plant is centered on the upper limit. The efforts are thus always aimed at keeping the shaded area shown in Figure 4 as small as possible.

The distance between the upper and lower limit is to a large extent a question of the control outlay and is therefore decisive for the selection of a suitable

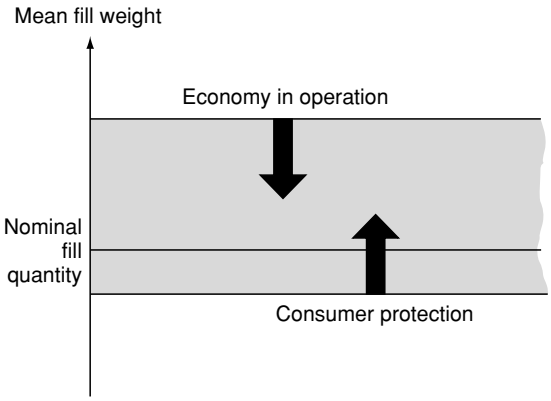

Figure 4 In the interest of the filling plant, the magnitude of the overfilling (shaded area) must be kept as small as possible. Reproduced from Weighing, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

control system. Decisive criteria in this case are the type and speed of the controls as well as the desired informative nature of the control results, rather than the number of packages to be checked.

The Control System – An Adjustment System

A control system should not only be used to check weights or volumes of packaged products, it must also be possible to influence the filling process actively with the aim of ensuring compliance with the statutory limit values in filling at all times with minimal material costs. This results in a control loop (Figure 5), which in practice is by no means easy to achieve.

As every filling process is subject to random influences, it goes without saying that the control and adjustment tasks must be solved by statistical methods. However, an important fact should always be noted even with the statistical weight check: nobody – and this applies however extensive the technical effort – can predict the weight of the next package.

In contrast to a widely held view, 100% inspection does not result in any improvement in this respect. With a check weigher integrated in the filling line, although every single package is checked, statistical methods must still be employed for control of the filling machine.

Selection of the Suitable Organizational Form

In the selection of the suitable organizational form, a fundamental choice has to be made right at the start: the options are statistical filling process control, or

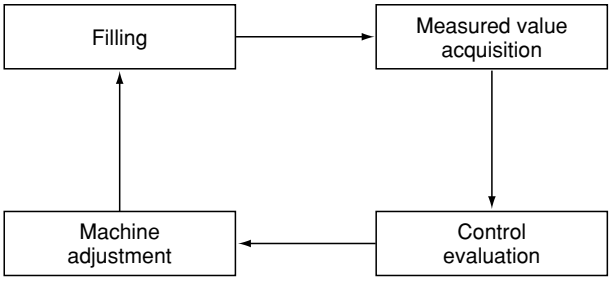

Figure 5 Control loop for active influencing of the filling process. Reproduced from Weighing, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

100% inspection, i.e., a check on every package with a check weigher.

The following description of the two types of control shows that each has certain advantages. It will be apparent that under certain conditions it can also be eminently practical to select a mixed organizational form employing both random sampling and 100% inspection.

100% Inspection

Basically, it is wrong to assume that only 100% inspection can provide the necessary assurance of strict compliance with the statutory regulations. In particular, it is the presence of a lowest tolerance limit which contributes to this erroneous belief; the regulations state quite clearly that to ensure the lowest tolerance limit 100% inspection is not obligatory and a random sample check is admissible.

However, there are a number of cases in which 100% inspection has advantages compared with random sampling:

- in a filling situation where frequently only half-filled or even empty packages are produced and there is no subsequent visual inspection
- if the produced packages have to be classified and/or sorted
- if a complete operational data acquisition or evaluation (production efficiency, throughput, etc.) is required
- if the product or the filling situation experiences extreme density fluctuations

The automatic feed of the packages to a check weigher allows a measured value acquisition without additional personnel costs. However, with a relatively large number of filling machines, the capital investment required for 100% inspection is considerable, as each filling line must be equipped with a check weigher (Figure 6).

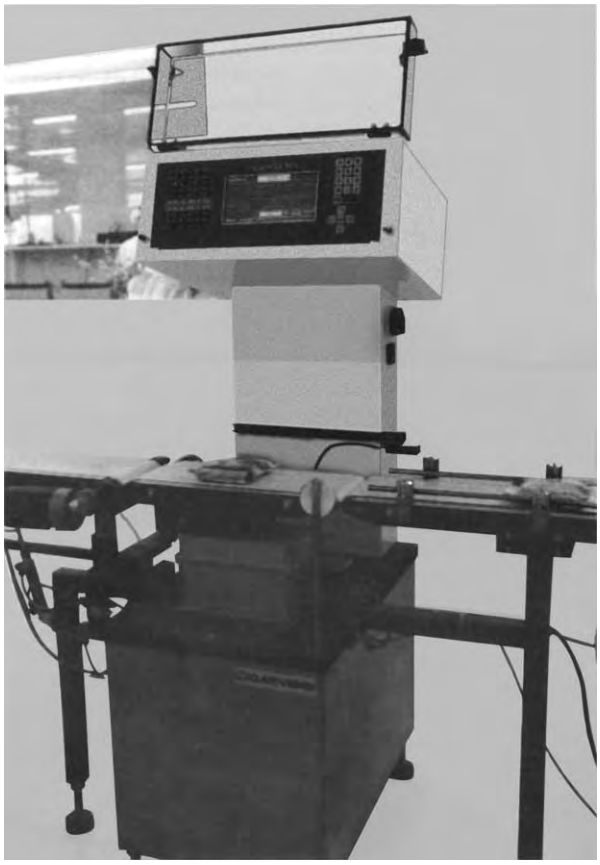

Figure 6 Check weigher (Leisi, Switzerland). Reproduced from Weighing, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Figure 7 Sampling operation (weighing/Leisi). Reproduced from Weighing, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Statistical Filling Process Control

In contrast to 100% inspection, the statistical filling process control has the following advantages:

1. The most important advantage of the statistical control is without doubt the appreciably lower capital investment required in comparison with 100% inspection. The disadvantage of the higher outlay, owing to the need for additional control personnel, is offset to some extent by selection of a suitable organizational form. Hence, in many plants, it is the task of the machine operators to undertake a check of their own machines by periodic sampling through a line inspection.
2. If a measuring station is assigned several machines, the inspection can be made with a so-called continuous inspection. Special control personnel make rounds through the production, collect samples, and then check the packages at the measuring station. This form of organization is selected primarily in plants in which the filling process control is performed by a department independent of the production – as a rule by the quality assurance.
3. Basically, the management of even large measurement series in a statistical filling process control poses no problems since the operating convenience and efficiency of modern electronic scales allow several thousand weighings in the course of a working day (Figure 7).

In summary, the statistical filling process control proves advantageous in those filling situations where a large number of filling stations must be monitored with low capital investment. In contrast, 100% inspection is advantageous in all cases where large batches of a single product have to be subjected to an automated filling control at a few filling stations and hence the ability of a modern check weigher to check several hundred packages per minute is of prime importance.

Selection of the Correct Control System

What criteria are used by the filling plant to select the control system that is optimum for its specific requirements? The cost-conscious plant will decide on the control system that brings it the greatest profit, i.e., the smallest possible overfilling that ensures no violations of the laws. With such a minimum configuration, the filling plant may well comply with the statutory regulations, but it sometimes gives away material worth a considerable amount through overfilling.

From the industrial economics viewpoint, however, a filling control is also indispensable even if there

were no regulations, for the correct selection and use of a control system can be profitable. In the final analysis the success of the filling process control depends to a large extent on the control personnel and machine operators. Only when the degree of automation has been increased will this dependence be reduced.

Filling Process Control in Practice: The SQC52 Quality Control System

The SQC52 quality control system is discussed here in some detail as a representative of the broad spectrum of systems on offer.

This system uses random sampling to supply all information decisive for the optimum adjustment of the filling machine and for the legally required records and evaluations.

The operating philosophy of the SQC52 system has been taken over from the window technique used successfully in EDP. The small number of operating keys and the easily surveyed graphics display contribute decisively to the simple and convenient operation (Figure 8).

All information is displayed in the form of windows. The status field provides general information on the operating mode; the dialogue window is used to support the operating sequences. It indicates when additional details are required or information or error messages have to be displayed for the operator. Each operating mode contains a number of working screens, which can be selected using the function keys.

Dependability – Thanks to Proven Menu Technique

For the control operation, all necessary data are available as a working window. In the working screen product definition, the user defines in a simple manner all necessary product parameters for the products to be checked. Other windows handle the parameter selection (selection of the possible weighing modes), the interface configuration, the memory allocation and instrument settings. The individual windows are selected using cursor keys. A help key is used to call up help texts with information on the respective entry or selection fields.

Graphical Evaluation, Informative Records

The sampling operations can be shown on the display as a mean-value trace together with the deviations with time in the filling process. The last 10 samples in each case are shown in the correct order as so-called box plots with mean value, standard deviation, greatest and smallest mean value.

The histogram reproduces a set of batch statistics with the distribution of all measured values. This allows a qualitative report on the filling plant and the production. On the basis of the statistical values calculated by the system, the machine can be adjusted manually (adjustment message) or automatically (Figure 9).

The SQC52 system allows comprehensive documentation of the samples in four detailed sets of statistics. All parameters, such as date, time, mean value, standard deviation and tolerance violations, as well as all requisite adjustments, are recorded and printed out. The system takes into account most national legal regulations (tolerance system) including pharmacopoeia (an official list with regulations for preparation, quality, and application of drugs).

The SQC52 System in Operation in Leisi

Leisi is a large Swiss bakery and producer of different types of ready-made pastry. It uses an SQC52 system

Figure 8 SQC52 display. Reproduced from Weighing, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

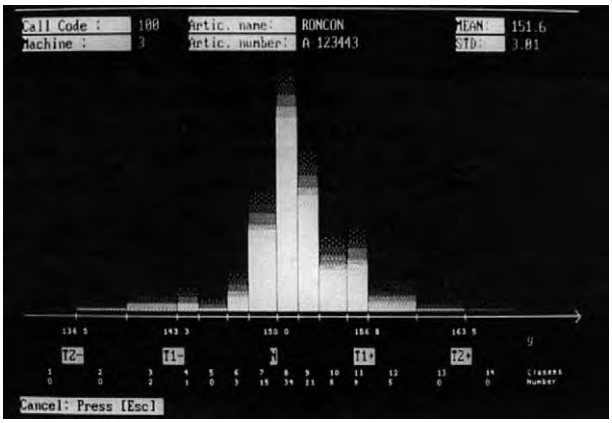

Figure 9 SQC61 evaluations. Reproduced from Weighing,*Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

with five weighing stations. At four of these, random sample checks are made with the machine operators performing the weighing. The fifth line has a Garvens check weigher attached. The reasons for such a distribution are as follows.

At four of the lines with random sample weighings, rolled-out, ready-made pastry is produced. On the basis of the adjustment message of the system (in grams), the machine operators set the pastry thickness on the machines, or they correct it if necessary. Although this information is received in grams and not in correction values in millimeters of pastry thickness, the machine control functions completely satisfactorily thanks to the experience of the operators.

In contrast to earlier generations of control methods, the system has the great advantage that it supplies adjustment messages in absolute values and decisions no longer have to be made on the basis of a mean value alone.

At the fifth line, always the same product, unrolled pastry, is produced and packaged in large quantities. The criteria listed earlier are therefore applicable here and a check weigher is used.

Leisi produces one or more batches daily for a major customer. The system thus supplies the summaries in the form of batch statistics. If necessary, these can be passed on to the customer.

In addition to the statutory requirements which the SQC52 system in operation with Leisi fulfills, it also offers attractive prospects from the economic efficiency standpoint. Quite apart from the savings in material owing to optimum fill quantities, it greatly rationalizes control operations. The data of approximately 50 articles are permanently stored. After every product change at a filling station, the control system is thus immediately ready for operation. If the packaging of a product is changed, only the new tare value must be entered. This is performed simply by overwriting the old value.

See also: **Cleaning Procedures in the Factory**: Overall Approach; **Quality Assurance and Quality Control**

Weight Maintenance *See* **Slimming**: Slimming Diets; Metabolic Conseqences of Slimming Diets and Weight Maintenance

Whales *See* **Marine Foods**: Production and Uses of Marine Algae; Edible Animals Found in the Sea; Marine Mammals as Meat Sources

WHEAT

Contents

The Crop

S T Grundas, Dobrzanski Institute of Agrophysics, Polish Academy of Sciences, Lubin, Poland

Introduction

For 10 000 years, wheat (*Triticum*) has been an object of particular interest, owing to its specific taste and nutritional value, as well as its ease of cultivation and storage. Archeological evidence indicates that primitive unleavened bread, baked on heated stones, was first made in the Neolithic era. The early Egyptians were developers of bread and the use of the oven and developed baking into one of the first large-scale food production industries. Nowadays, wheat constitutes about one-third of the global production of cereals, and its yearly production in the mid-1990s reached 600×10^6 tonnes. Wheat has a predominant role in the grain trade and is utilized as food (67%), feed (20%), seed (7%), and industrial products (6%). Wheat meal used for human consumption provides 20% of all the energy consumed by the human population. It is consumed mainly as bread.

This article provides information on the following: the global distribution of wheat; its varieties and market classes; the morphology and anatomy of the spikelet and grain; its chemical and nutritional composition; grading, handling, storage; wheat processing and end uses; the role of wheat gluten proteins.

Distribution, Varieties, and Market Classes of Wheat

Distribution

There are over 20 species of wheat, differing in their basic number of chromosomes (diplo-, tetra-, and hexaploid), and several thousand varieties. The most economically significant species are the ordinary wheat *Triticum aestivum* sp. *vulgare* (hexaploid) and the hard wheat *T. durum* (tetraploid). The global area of wheat cultivation covers a total of about 240×10^6 ha, and approximately 90% of that is occupied by *T. aestivum*, which has the greatest number of crop-yielding varieties of starchy grains. There are numerous varieties of *T. durum* cultivated in continental climate zones and covering approximately 10% of the total area.

The geography of wheat cultivation is illustrated in Figure 1. The areas of wheat cultivation are mainly found in the northern hemisphere. Wheat is grown both as a winter and as a spring cereal in the moderate climate zones, and in the tropical regions in low lands and uplands alike. The distribution of wheat over the various parts of the world is related to the numerous species and varieties, and to their adaptability to particular environments.

Wheat harvesting is a continuous process around the world and every month is a period of harvest in one of the wheat-growing regions.

Wheat Varieties and Marked Classes

Five marked classes of wheat are distinguished: Hard and Soft Red Winter (HRW and SRW), Hard Red Spring (HRS), durum (Durum), and White, which is classified in the following subclasses: Hard and Soft Winter (HWW and SWW), and Hard and Soft Spring (HWS and SWS).

Although variety is an important factor affecting quality, wheats are seldom classified on the basis of a single variety. In practice, we adopt the classification of wheat according to its class. Each class consists

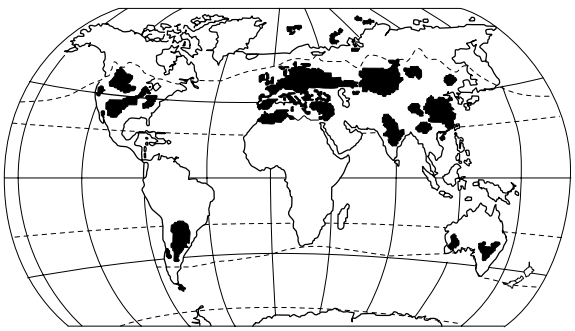

Figure 1 Main wheat-growing regions in the world; the northern and southern distribution boundaries are marked with a broken line. From Grzesiuk S and Kulka K (1988) *Biology of Cereal Kernels*. Warsaw: Polish Scientific Ed. (in Polish) with permission.

of a group of varieties of similar characteristics for specific purposes.

As an example, in Canada a model HRS wheat is the Marquis variety, for years now replaced in production by varieties genetically similar, but more valuable economically, such as Neepawa, Manitou, and Thatcher. Out of Amber classes (CWAD1 and CWAD2) *T. durum* wheats, representative varieties are Kyle and Plenty. Wheat varieties grown in USA are genetically similar to those grown in Canada. In Argentina, approximately 97% of the area of wheat cultivation is taken up by winter wheat similar in its value to the HRS class. In Australia, the primary crop is winter wheat, class ASW (see Grading below), of hard and soft white grain, genetically similar to the Argentinean wheat. In Europe, and especially in France and Germany, the high level of agriculture resulted in these two countries having virtually the highest grain yield per unit of area of cultivation. Recently, 'clubs' have been set up by wheat producers who achieve yields of 10–15 t ha^{-1}. In this region the dominant type are wheats of the SRW class. The EU countries have dozens of high-yield varieties of very good grain quality. In Mediterranean countries such as Italy and Spain some *T. durum* is grown. East European countries grow mainly red winter wheats, some of which have very good baking qualities. In the countries of the Commonwealth of Independent States (CIS) both spring and winter, red and white wheats are grown, as well as some durum wheats. The CIS has the largest area of wheat cultivation in the world. The leading varieties in terms of yield and grain quality are winter wheats Bezostaya 1 and Mironovskaya 808, and the spring wheat Saratovskaya 29.

Morphology and Anatomy of Spikelet and Grain

Spikelet

The central part of the ear or spike is a waveform rachis, composed of a number of short nodes and internodes (Figure 2a). Spikelets grow at the nodes, constituting modified shoots (Figure 2b). Two glumes, located at the base of each spikelet, are shorter than the rest of lemmas in florets. In floret the grain is enclosed between two pales (the palea, from the inside, and the lemma, from the outside). Ordinary wheats which have fusiform spikes are awned or awnless, and easily threshed.

Grain

The dorsal side (the same side as the germ) of the wheat grain can be oval, ovate, or elliptical, with a cluster of long or short brush hairs at the apex, and an oval or circular embryo at the bottom. Wheat kernels have a longitudinal crease or a deep groove the length of the ventral side (opposite the germ). The shape of the groove is a characteristic feature of some species and varieties.

The grain structure is presented in Figure 2c. The grain coat is made up of the fruit and the seed coat, adhering directly to the aleurone layer. The pericarp (fruit coat) consists of several layers of translucent cells, strongly lignified, elongated in shape and encrusted with mineral substances, and of a layer of perpendicular and tubular cells. The seed coat consists of a compact cell layer and pigment strand, providing the grain with its characteristic coloring, and a hyaline layer.

The aleurone layer, made up of large thick-walled cells filled with functional proteins and nutritional components, encloses the endosperm, and disappears around the embryo.

The endosperm consists of large, thin-walled cells, filled mainly with starch and protein. In the subaleurone area, especially on the dorsal side of the grain, the dominant cells are elongated in the direction of the endosperm center. Cells within the endosperm are less regular in shape. Starch grains are enclosed in the thin layer of adherent protein and located within a protein matrix which fills the individual cells of the endosperm to varying degrees. The highest protein content is observed in the cells of the subaleurone layer of the endosperm. The closer to the center of the grain, the lower the protein content.

Chemical and Nutritional Composition of Wheat Grain

Chemical Composition

The chemical composition of wheat grain and of its anatomical parts (Table 1) is determined by genetic and ecological (climate, soil, tillage) factors, and by the physical and chemical effects acting on the grain during its storage and processing.

The primary quantitative component of wheat grain is starch. Cereal grains store energy in the form of starch. The starchy endosperm varies in composition from its outer portion, just beneath the aleurone layer, to its center. The amount of starch contained in wheat grain varies but is generally between 60 and 70% of the weight of the grain. Starch is basically polymers of glucose. Chemically, at least two types of polymers are distinguishable: amylose, an essentially linear polymer, and amylopectin, which is highly branched. The ratio of amylose to amylopectin is relatively constant, at about $23 \pm 3\%$

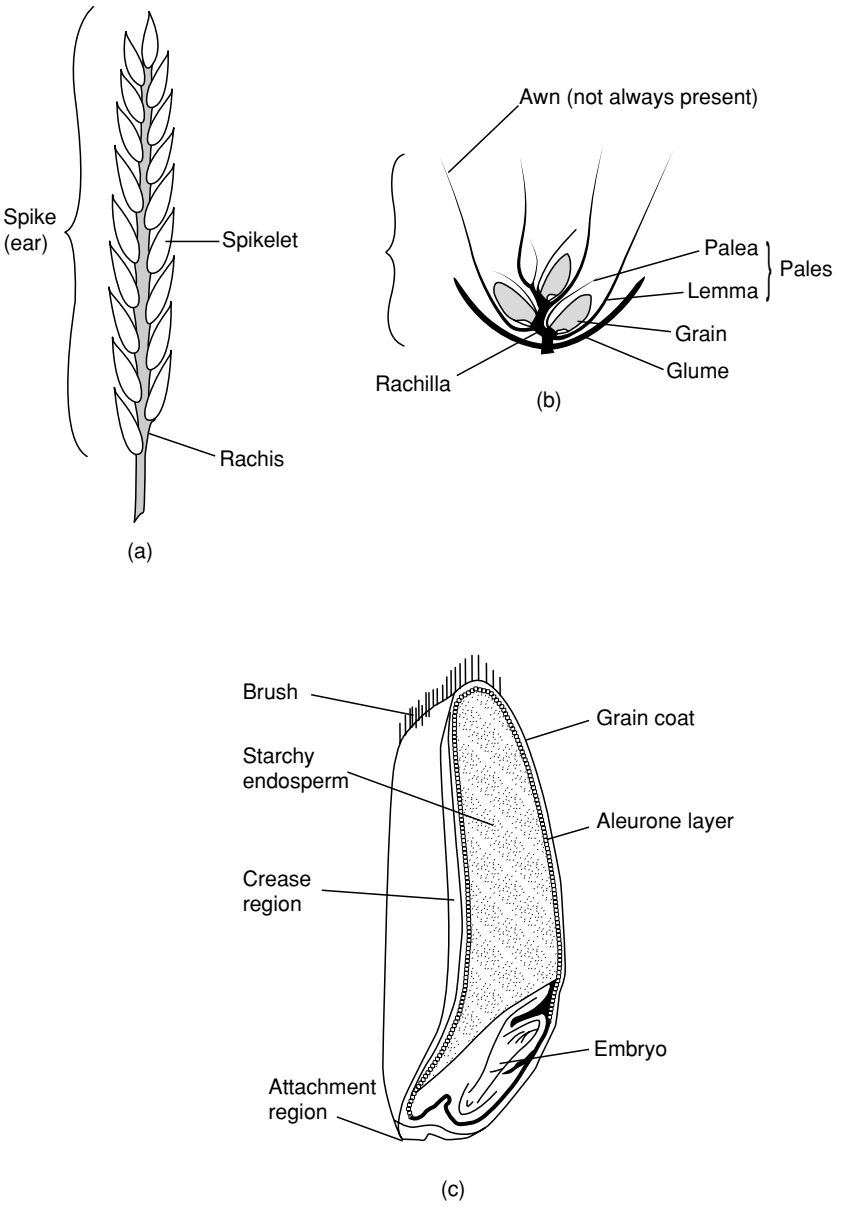

Figure 2 Structure of the spike (ear), spikelet, and grain: (a) spikelets arranged alternately on the main axis or rachis; (b) a mature spikelet; (c) grain structure.

amylose. Wheat endosperm has two types and sizes of starch granules: the large (25–40 μm) lenticular and small (5–10 μm) spherical granules. Apart from starch, the grain and especially the grain coat, the aleurone and the embryo, contain other carbohydrates, such as cellulose, hemicelluloses (pentosans), and sugars. Pentosans, although their content in endosperm is low (2–3%), form the basic structure of endosperm cell walls. Of particular interest is the concept that pentosans could, depending on their water-absorbing capacity (which is 10 times their mass), represent 'moisture pumps' that to a certain extent regulate moisture levels in different flour components after their hydration in the dough-making process.

Proteins are among the basic nutritional components of grain, and their quantity and quality determine the applicability of the grain and its products for particular purposes. The total protein content is determined on the basis of the nitrogen multiplied by factors of 5.7.

The protein substances of wheat are divided, in relation to their solubility, into the following subgroups: (1) albumins, in water; (2) globulins, in

10% sodium chloride; (3) prolamins, in 70% ethanol; (4) glutelins, in dilute alkaline or acid solutions. These substances form groups of particles of similar physicochemical and structural properties but, occurring in various grain tissues, they differ from one another in amino acid composition (Table 2). The proteins of the starchy endosperm contain small quantities of lysine, a considerable amount of glutamic acid, and some prolines. The proteins of the embryo and the aleurone layer of the endosperm are characterized by a relatively high content of leucine, lysine, and aspartic acid.

Fats are the third important component of wheat grain. Their highest contents are observed in the embryo and the aleurone layer (Table 1). A high content of fats in cereal products can result in their rapid spoilage. Carotenoids, found mostly in an embryo and in the aleurone layer, belong in a group of fats which are of practical importance. They have a considerable nutritional value, as they constitute the basis for the formation of vitamin A in humans and animals. They have a yellow or pale orange coloring, which determines the typical creamy look of wheat flour. Whitening flour by chemical processing results in a decrease in its nutritional value through the oxidation of the carotenoids.

The ash formed in the course of burning grain or flour originates from inorganic compounds and organic substances (Table 1). The highest content of mineral substances is found in the aleurone layer and in the grain coat, and the lowest in the endosperm.

Nutritional Composition of Grain and its Products

The nutritional value of wheat is determined by the technological properties of the grain (good milling and baking characteristics) and the nutritional qualities of its products, in combination with taste and digestibility.

The nutritional components of the grain provide energy and 'building materials,' and play a regulatory role. Primary components providing energy are sugars and fats and, less frequently, proteins. Components providing building materials are mainly proteins and minerals, while those performing a regulatory function include vitamins and minerals. The energy values of the major chemical components of grain and its products are presented in Table 3.

The nutritional value is also determined by the content of endogenous amino acid (lysine, methionine, tryptophan), the levels of which are relatively low in wheat. For this reason, an admixture of milk or soya flour to wheat flour improves the quality of bread or other bakery products. A typical day's healthy diet would include 55% or more of daily calories from complex carbohydrates, such as cereal grains. Eating six or more servings of bread, pasta,

Table 2 Amino acid content in proteins from various parts of the wheat grain

	Amino acid content (g per 100 g)		
Amino acid	*Endosperm*	*Embryo*	*Aleurone layer*
Lysine	1.7	8.4	4.6
Histidine	2.3	4.0	4.0
Arginine	3.5		
Aspartic acid	3.3	6.4	8.0
Threonine	2.7	2.6	3.1
Glutamic acid	39.3	8.4	16.4
Proline	14.7	3.5	3.9
Glycine		7.3	
Alanine	3.5	6.4	5.1
Cystine		1.0	1.8
Valine	4.3	4.0	5.4
Methionine	1.9	1.5	1.5
Isoleucine	4.1	2.5	3.1
Phenylalanine	6.0	2.0	4.0
Tyrosine	3.9	1.5	3.0
Leucine	7.6	4.5	6.0
Tryptophan	2.0	1.7	4.0

Data from Grzesiuk S and Kulka K (1988) *Biology of Cereal Kernels*. Warsaw: Polish Scientific Ed. (in Polish).

Table 1 Main chemical components in particular parts of the wheat grain

	Content of chemical components (% db)			
Chemical components	*Whole grain*	*Grain coat with aleurone layer*	*Embryo*	*Endosperm*
Proteins	10–17	23–33	36–42	9–14
Carbohydrates				
Starch	60–70	0	0	78–84
Sugars	3.0–6.0	3.0–5.0	22–28	3.0–4.0
Pentosans	6.0–9.5	30–40	9–11	2.5–3.0
Cellulose	2.5–3.3	12–20	3–5	0.13–0.18
Fats	2.0–2.5	7.0–8.5	12–16	0.5–0.7
Minerals	1.4–2.3	9–11	5–6	0.3–0.5

Data from various sources, selection by author. db, dry-weight bases.

Table 3 Energy value of some grades of flour and wheat bakery products

Product	*Energy value (kJ per 100 g)*	
	Theoretical	*Actual*
Wholemeal flour (97% type 1888)	1575	1400
Bread-making flour (80% type 850)	1533	1365
Patent flour (type 500)	1512	1386
Semolina	1533	1407
Graham bread	1071	945
White bread	1050	1010
Rolls (popular types)	1050	1030

Data from various sources, selected by the author.

Table 4 The B-group vitamins located in endosperm of wheat grain and recommended daily intake for adults

B-group vitamins	*Content in endosperm (%)*	*Daily intake for adults (mg)*
Thiamin	3	2.0–3.0
Riboflavin	32	2.0–2.5
Pyridoxine	6	1.5–2.0
Niacin	12	15.0–20.0
Pantothenic acid	43	10.0–11.0

Data from Pomeranz Y (1988) *Wheat: Chemistry and Technology*, vol. II. St Paul, Minnesota, USA: American Association of Cereal Chemists, and other sources, selected by the author.

and cereal grains each day is the right choice for an overall healthy lifestyle.

Wheat grain is relatively low in vitamins; only the B-group vitamins are present in any notable quantities. The B-group vitamins as a component of wheat grain are presented in Table 4. Likewise, fat, the content of which is low in wheat flour, plays no significant role in the nutritional value of wheat flour and its products.

Grain Grading, Handling, and Storage

Grading

The grading of wheat grain involves the isolation and valuation of grain portions based on the physical and chemical properties of the grain. The scope of wheat grain grading in use is related to the state and development of agricultural knowledge, as well as to the level of organization of the local cereal market (pricing polices, grain-purchasing system).

The systems of wheat grain grading include the following groups of grain properties:

1. Stable properties, related to genetic properties, such as grain size and coloring, protein content, and sedimentation factors.
2. Variable properties, subject to change as a result of drying, cleaning, sorting, and transport (moisture content, contamination, damage to grain), related to climatic conditions (bulk density).
3. Permanent faults and defects, such as stale smell, fermented smell, foreign smell, or faults that can be rectified, e.g., washing grain contaminated with soil or foreign matter.

The stability properties are primary properties, and it is on their basis that the market classes are distinguished. This is the most important group of properties characterizing the market grades of wheats in international trading.

The variable properties are secondary properties, although also important from the viewpoint of grain quality. The group of variable properties, systematized into several ranges in terms of values, constitutes grain grades, which are identical for all the market classes and varieties. The quality grading of wheat grain sometimes includes an assessment of mechanical damage, which occurs wherever grain is subjected to the destructive action of internal or external forces.

The results of research carried out using X-ray methods showed significant differences in grain endosperm cracks between common wheat varieties. It was stated that natural wetting of dry grain (below 15% of moisture content) during rainfall in field conditions is one of the main reasons for its cracking. Considering the results of endosperm cracks which can appear before harvest in some climatic regions of wheat production, special attention should be paid not only to proper combined adjustment but also to the state of natural continuity of endosperm tissue and to genetic features of the variety. The susceptibility of wheat to mechanical damage is determined by genetic factors (hard and soft grains), environmental effects (climatic condition during the preharvest period), and by the conditions of grain storage (excessive humidity). The combination and selection of these properties determine the unification of grain material in terms of quality.

The permanent faults and defects are excluded in grain grading, as their character disqualifies grain from food usage. Defined faults which can be rectified constitute special grades of defective quality.

The standards of wheat valuation have been established by Canada and the USA. These countries have a system of wheat grading with classes based on the varietal properties as expressed by the vegetation period (spring and winter wheats), grain coloring (red, white) and hardness (hard, soft). In recent years, an extra class, Utility, has been added to the grading system, and it is made up of certain genetically

similar varieties of winter and spring wheat. Wheat grading in the USA additionally provides, in the subclasses, for the proportion of vitreous grains, and it is simplified, compared to the Canadian system, by using identical grades for all classes and subclasses. The system also includes an additional grade, called Sample Grade, samples of which do not conform to established requirements of the basic-quality grades.

Argentina uses a system of bread wheat division into subclasses in relation to the grain type (hard, semihard, and soft), and into grades. Australia has developed a grading system for white-grain wheats. The grading is based on the health requirements, cleanliness and bulk density established for the basic class called Australian Standard Wheat (ASW). Grain of lower quality than that of ASW is classified as General Purpose, or as Feed grain.

Wheat grading in the CIS is based on principles similar to those followed in the USA, with the distinction that classes are replaced here by types, and subclasses by subtypes, primarily designed to take into consideration the proportions of vitreous grains. Some European countries (e.g., France, Germany) use a grading system based on the protein content, bulk density, and sedimentation factor. With only slight differences in the appearance of many varieties, and with the possibility of grain quality reduction as a result of environmental effects, grades have been developed to include, in addition, the contamination of the grain and the falling number.

Other countries, in which wheat is grown or only imported, also use specific standards for the evaluation of the quality of wheat grain and its products. Most frequently such systems are modifications of the systems presented above.

In relation to grading, the following definitions are applied:

1. Hard wheat is wheat which, as a result of variety or breeding in combination with environmental factors during growth, has a vitreous endosperm generally considered an advantage for the production of bread-making flours.
2. Soft wheat is wheat which, as a result of variety or breeding in combination with environmental factors during growth, has a white opaque endosperm generally considered more suitable for the production of cake and biscuit flours.

Handling

Grain handling includes harvesting, segregation, cleaning, and sometimes, drying. Preliminary segregation takes place during combine harvesting of the grain which, depending on its moisture content and level of contamination, is then additionally cleaned and dried.

With the use of various means of grain transport, both internal (in the grain-cleaning machinery and for grain loading and off-loading) as well as external (road, rail, or sea transport), loads of wheat grain of initially determined grades are sometimes subject to mechanical damage, which considerably reduces their technological value.

Grain stockpiling is a concentric process in which the grain is gradually moved from the farm to the world market. That movement passes the following stages: farm silos, local, regional, and terminal storage elevators (the latter frequently located at harbors). At every stage, grain is subjected to classification – from highly detailed to more general grading, improvement (drying, cleaning, sorting, and deinfestation, if required), and grouping of supplies into larger batches of constant quality. Isolated fractions of undersize grain and lower-grade wheat are sold on the local market. With depressed demand, the concentric movement may end at the local market. In case of oversupply it proceeds into the regional or the national markets. With grain exports, the concentric movement may flow out of the national borders and continue worldwide.

Storage

The principal factors affecting the intensity of the process of grain respiration, and therefore also the extent of quantitative losses to the mass of the grain in storage, are the following: grain moisture content, temperature, and grain condition relative to its history. Grain respiration is more intensive in soft grain than in hard grain. If grain moisture content is relatively low (below 14% weight basis, wb), and the bulk of grain of that moisture is homogenous, such grain can be stored for several years without significant losses. However, if proper air temperature and humidity conditions are not observed in the storage area, the increasing humidity resulting from the process of grain respiration may lead to the phenomenon of spontaneous heating of the grain, which excludes its use for human consumption.

In some wheat-growing regions, e.g., in the Scandinavian countries, wheat grain is relatively moist during harvesting and has to be dried to maintain its level of quality. Drying must be performed at specific levels of temperature and humidity of the drying air. Exceeding the required temperature may lead to the loss in the baking value of the flour. To avoid damaging grain meant for baking purposes, its temperature should not exceed 35 °C at moisture contents above 20% wb. There is a general view that the so-called 'critical moisture content' (14% wb), at

which changes related to fungal growth are at a minimum and the grain remains in an abiotic condition, can be maintained by air with a relative humidity of less than 70%.

Wheat Processing and End Uses

Processing

Grain for processing must be of a satisfactory technological quality, and is subjected to a process of preparation for milling. The process of preparation consists of sorting, decontamination, and grain surface cleaning. Most frequent is the process of dry cleaning, using the following equipment: winnowers, aspirators, dry stone separators, pearling mills, brushers, or wet brushing machines. An important measure in the process of grain preparation is the conditioning which brings the grain to the most favorable condition from the viewpoint of the grain milling quality and the flour baking value. This process strengthens the grain coat and facilitates the diminution of the endosperm, but usually eliminates the aleurone layer because it is tightly bound to the grain coat. The aleurone layer contains most of cereals' active anticancer compounds. Therefore, whole grain products are recommended. The embryo, on the other hand, owing to its high fat content, easily flakes off after passing between smooth-bore cylinders.

The milling of the wheat grain, understood as a process of grain diminution down to the level of flour, is realized currently in rather complex and highly mechanized processing plants, equipped with special machines for grain cleaning, grinding, and sorting.

One-time diminution of grain into a product of a specific particle size, so-called simple grinding, is used when grinding wheat grain for wholemeal flour. In graded flour production, the grain and the products of its diminution go through a process of repeat grinding, and such grinding is known as complex grinding.

Extraction is the ratio of the mass of flour obtained as a result of milling to the initial grain mass. The level of flour extraction is related to its degree of 'whiteness.' Flours obtained from mealy wheats, at the same level of extraction, is always brighter than flour from vitreous wheats. Fine flour, of fine granulation, is lighter than flour of the same extraction but of coarser granulation. Flours of fine and uniform granulation improve the process of dough formation. With increasing level of extraction, the baking value of the flour from soft wheat grain decreases. Flours of higher extraction have higher protein contents, and also contain more cellulose, vitamin B, and mineral substances such as phosphorus, iron, calcium, magnesium, zinc, and copper, but have a lower content of starch. More starchy flours, i.e., those of lower protein and cellulose content, are made of pure grits and fine-ground, while the highest content of cellulose and ash characterize flours made from the final milling passages. Sometimes, during the process of milling, wheat germs of high nutritional value but short storage life are isolated.

End Uses

Wheat is used in baking, primarily to make white bread and numerous other bakery products. Although wheat is consumed in many forms, including noodles, gruels, cooked cereals, and ready-to-eat cereals, bread has established itself worldwide as a major convenient and delicious food, especially when made according to regional preferences for flat breads, French bread (baguette), Italian breads, Vienna bread, white and variety pan breads, Polish bread specialties, and many other ethnic varieties. A specific form of bread, known in the countries of the Far East, especially in China, is steamed bread.

Soft wheat flour (of generally low extraction and protein content) is used mostly to make cakes, wafers, and dough that is used for cooked products (noodles and some types of pasta) and for special cakes and cookies made with butter and sugar.

Pasta and noodles are wheat-based products that are formed from a dough but are not leavened. It is generally believed that the ideal raw material for pasta is semolina from durum wheat. Durums are usually amber in color; they are high in carotenoid pigments. When durum wheat is expensive, common hard wheat farina (purified middlings from hard wheat) is often blended with durum or used by itself to produce pasta. In general, the hard-wheat farina produces good pasta except that it does not have the yellow color and is not as resistant to overcooking as that produced from semolina.

Generally, we can say that the baker converts the flour produced by the miller from wheat grown by the producer into delicious basic foods favored by the consumer and capable of being eaten with every meal, by every ethnic and age group every day of the year, regardless of economic level, from the cradle to the grave.

Role of Wheat Gluten Proteins

Gluten is a highly molecular complex of proteins – gliadins and glutenins – together with other components of the endosperm cells. Gladins and glutenins are typical storage proteins of wheat. Glutenins, as a

group of wheat proteins, are heterogeneous, and can be separated, using electrophoresis, into over a dozen fractions of different molecular weights. An appropriate percentage of high-molecular fractions in wheat flour is responsible for the baking value of the flour.

Gluten, as a specific and unique substance formed in the dough obtained from ordinary wheat flour, is characterized by physical properties such as cohesion, elasticity, and considerable stretchability. In bakery applications, the basic characteristics of gluten are its high water absorption (it can even triple its mass), and its ability to form viscoelastic films which play a direct part in the proving of the dough. The properties of film-forming are also used in the production of other foodstuffs, such as the glazing of sausage and meat pâté.

Among the different environmental factors affecting grain quality, temperature fluctuation during grain filling is seen to be the most important factor. Most of the studies have examined the effect of extreme temperatures on dough properties. Daily mean temperatures up to 30 °C during grain filling generally increased dough strength, while temperatures above 30 °C produced weaker dough.

In terms of nutritional value, gluten is at the bottom of the scale, owing to its low content of lysine, methionine, and tryptophan, but in combination with other proteins, e.g., with soya, it displays a higher nutritional value than each of the proteins by itself. One of the reasons for the increasingly frequent addition of gluten to wheat flour is the fact that new wheat varieties, grown under intensive cultivation conditions and providing high crop yield, produce protein of lower value.

Gluten cannot be considered as a natural substance occurring in wheat. It is a form given to protein components as a result of modifying measures applied during the extraction of flour for dough-making with the result that certain proteins can expand in a water medium or in a weak solution of sodium chloride, in combination with mechanical activity. No correlation has been ascertained between the amount of gluten washed out, and its quantitative characteristics. Proteins are considered to be a particularly labile (reactive) system of chemical compounds in the grain, and hence the yield of gluten – adopted as a basis for the qualitative evaluation of grain material – depends on the character and rate of interaction taking place within the material under analysis. Studies confirm the technological effects of protein–water, protein–mineral substance, protein–carbohydrate, or protein–lipid interactions. Each of these types of interactions may have a positive or a negative effect under specific conditions. Contemporary baking technology uses a complex system of dough kneading whereby, apart from the addition of water, various substances which react easily with gluten proteins are introduced. For these reasons, countries experiencing a need for standardization of flour and dough quality have developed numerous instrumental methods to support or supplant the determination of gluten yield.

See also: **Bread**: Dietary Importance; **Cereals**: Contribution to the Diet; Bulk Storage of Grain; Handling of Grain for Storage; **Wheat**: Grain Structure of Wheat and Wheat-based Products

Further Reading

Faridi H and Faubion JM (ed.) (1995) *Wheat End Uses Around the World*. St Paul, Minnesota, USA: American Association of Cereal Chemists.

Grundas TS (1997) Advantage of physical properties of single kernels of wheat to evaluation of its utility values. *Europe News (Europe Section of the American Association of Cereal Chemists)* July: 6–7.

Hamer RJ and Hoseney RC (ed.) (1998) *Interactions: The Keys to Cereal Quality*. St Paul, Minnesota, USA: American Association of Cereal Chemists.

Hoseney RC (1994) *Principles of Cereal Science and Technology*, 2nd edn. St Paul, Minnesota, USA: American Association of Cereal Chemists.

Pomeranz Y (ed.) (1988) *Wheat: Chemistry and Technology*, 3rd edn, vols I and II. St Paul, Minnesota, USA: American Association of Cereal Chemists.

Pomeranz Y (ed.) (1989) *Wheat is Unique*. St Paul, Minnesota, USA: American Association of Cereal Chemists.

Sauer DB (ed.) (1992) *Storage of Cereal Grains and Their Products*, 4th edn. St Paul, Minnesota, USA: American Association of Cereal Chemists.

Saurer W (1996) The single kernel classification system for the characterization of wheat. *Proceedings of the 10th International Cereal and Bread Congress*. Greece: Chalkidiki.

Grain Structure of Wheat and Wheat-based Products

S T Grundas, Dobrzanski Institute of Agrophysics, Polish Academy of Sciences, Lublin, Poland

Introduction

The structure of wheat grain is a product of genetic and environmental effects acting on wheat during development. The structure of wheat grain is, for all practical purposes, the effect of a system of genetic

information that has been transferred during development. We can view wheat grain as the system waiting to be exploited to meet our needs. For 10 000 years, evolutionary changes have resulted in the wheat plant that is best suited to its environment. Human intervention in this process has given us the main wheat product – bread that has specific attributes which are not found elsewhere in the plant kingdom. We are only now beginning to understand better the genetics of a few storage proteins and how they relate to quality, let alone which genes are responsible for grain shape, size, and amount of storage components. Many factors that govern the quality properties of wheat grain are determined by the structure of the caryopsis. Indeed, every structural component of wheat kernel has some effect on grain quality as a basic material for human nutrition.

This article provides information on the following: wheat grain structure, techniques of wheat grain structure study, and wheat-based products.

Structure of Wheat Grain

The wheat grain is botanically a single-seeded fruit called a caryopsis or a kernel. The wheat kernel consists of an embryo called germ by millers and an endosperm enclosed by aleurone and bran layers. Location of the individual constituents of these components in the mature kernel is illustrated in both longitudinal and cross-section in Figure 1. In this figure a segment of bran layers with aleurone and endosperm cells is presented in detail.

In practice three dimensions determine the shape of the kernel: thickness, width, and length. The wheat kernel averages about 2.5–3.0 mm thickness, 3.0–3.5 mm width, and 6.0–7.0 mm length. Wheat kernels are rounded on the dorsal side (the same side as the germ) and on the ventral site (opposite the germ) they have a longitudinal cavity (crease). At the distal end of the kernel are a number of hairs (trichomes). The crease, which extends to its center, is masked by two flanks (cheeks). The crease occupies 0.7–1.9% of the total grain volume. The presence of wide and deep creases causes low test weights. Wheat kernels average about 30–35 mg in weight. All anatomical parts of the wheat kernel are presented in Figure 2.

Bran

The pericarp (fruit coat) surrounds the entire seed and consists of two portions, the outer pericarp and inner pericarp. The outer pericarp has the following layers: the epidermis (epicarp), the hypodermis, and the innermost layer, called the remnants of thin-walled cells. These thin-walled cells which have a discontinuous cellular structure form a natural plane of cleavage. Removal of the outer pericarp, which millers call beeswing, also aids movement of water into the kernel. The inner pericarp adjacent to the remnants is composed of intermediate cells – a single layer of cross cells and tube cells. The cross cells are long and cylindrical (about $125 \times 20\ \mu m$), and have a long axis perpendicular to the long axis of grain. They are tightly packed, with little or no intercellular space. The tube cells are of similar size and shape as the cross cells but have their long axis parallel to the long axis of grain. The tube cells are not packed tightly and do not form a continuous layer and thus have many intercellular spaces. They are only recognizable in the mid dorsal region of mature grains. The next layer inwards is the seed coat (testa or integument) which is firmly joined to the tube cells on the outside and the nucellar epidermis on the inside. The seed coat of red wheat consists of a thick outer cuticle, strongly pigmented layer, and thin inner cuticle. The seed coat in white wheat has cell layers containing little or no pigments. The color, usually red or white (although purple is also known), is related to pigment in the testa. Tightly bound to the internal surface of the seed coat is the nucellar epidermis (hyaline layer, perisperm). The thickness of the seed coat varies from 5 to 8 μm. The nucellar epidermis is about 7 μm thick and closely united to both the seed coat and the aleurone layer. The total pericarp has been reported to comprise about 5% of kernel volume.

The aleurone layer, which is generally one cell thick, completely surrounds the kernel, covering both the starchy endosperm and the germ except for that adjacent to the scutellum. Although the aleurone layer is anatomically a part of the endosperm, the miller regards the aleurone as the innermost layer of the bran. The majority of the mineral matter located in bran is found in the aleurone layer, which also contains one-third of the grain's thiamin content. The cytoplasm of the cells contains many small (3–4 μm), round aleurone granules surrounded by lipid droplets. The aleurone granules contain two types of inclusions: type I contains phytin and type II contains protein, carbohydrate, and bound nicotinic acid, which is largely unavailable for human nutrition. The phytin granules are the main source of mineral matter and hence the degree of aleurone (or bran) contamination of flour is frequently evaluated by an ash analysis. In addition, thiamin and riboflavin are higher in the aleurone than in the other parts of the bran, and enzyme activity is high. Over the embryo, the aleurone cells are modified, becoming thin-walled cells that may not contain aleurone granules. The thickness of the aleurone layer over the embryo averages about 13 μm, or less

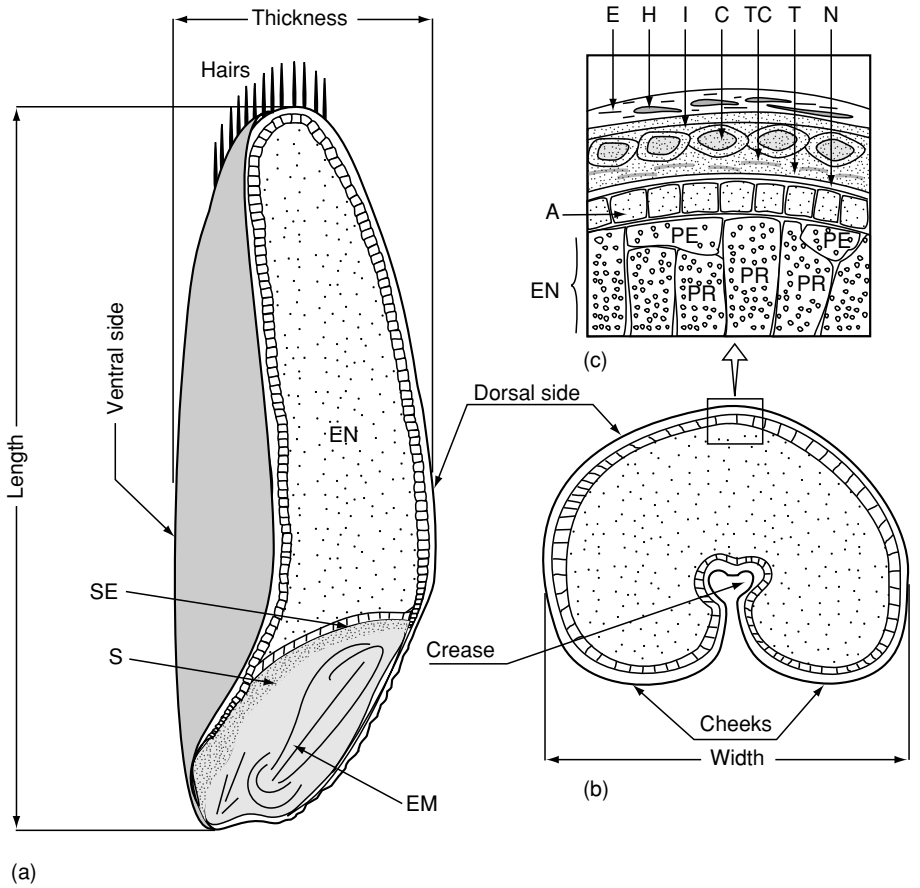

Figure 1 Main morphological components of a wheat kernel. (a) Longitudinal section; (b) cross-section; (c) segment of bran layer with aleurone and endosperm cells. E, epidermis; H, hypodermis; I, inner pericarp; C, cross cells; TC, tube cells; T, testa; N, nucellar layer; A, aleurone layer; EN, endosperm; PE, peripheral cells of endosperm; PR, prismatic cells of endosperm; SE, scutellar epithelium; S, scutellum; EM, embryo.

than one-third the thickness found elsewhere. The aleurone cells are heavy-walled, essentially cube-shaped, and free starch. The aleurone cells can vary in thickness from 30 to 70 μm within a single kernel and have thick (6–8 μm), double-layered cellulosic walls.

Endosperm

The starchy endosperm, excluding the aleurone layer, is composed of three types of cells: peripheral, prismatic, and central. The peripheral cells are the last to be initiated during grain filling and tend to be smaller than the other endosperm cells (60 μm in diameter and 20–60 μm radially) and have thicker cell walls (8 μm). Several rows of elongated prismatic cells are found inside the peripheral cells. They extend inward to about the center of the cheeks and are about 150–200 μm. The central cells are more irregular in size and shape than are the other cells, and are located inside the central endosperm. The central endosperm cells are the first to be formed and have thinner walls (2 μm). Cell wall thickness also appears to vary among cultivars and between hard and soft wheat types. The differences between hard and soft wheat may be the result of selection: hard wheat (bread wheat) has been selected for high water absorption. The endosperm cell walls are composed of pentosans, other hemicelluloses, and beta-glucans, but not cellulose. The pentosans in them absorb large amounts of water. The endosperm cells are packed with starch granules embedded in a protein matrix. Starch is the major component of wheat endosperm, comprising approximately 64–75% of milled endosperm. Generally, the starch granules in wheat are classified into two size groups: large, lenticular (lens-shaped) A granules of up to 40 μm across the flattened side, and small, spherical B granules (2–8 μm in diameter), which are formed later in the grain filling process.

In common bread wheat (*Triticum aestivum*) the endosperm texture of grains varies both in texture

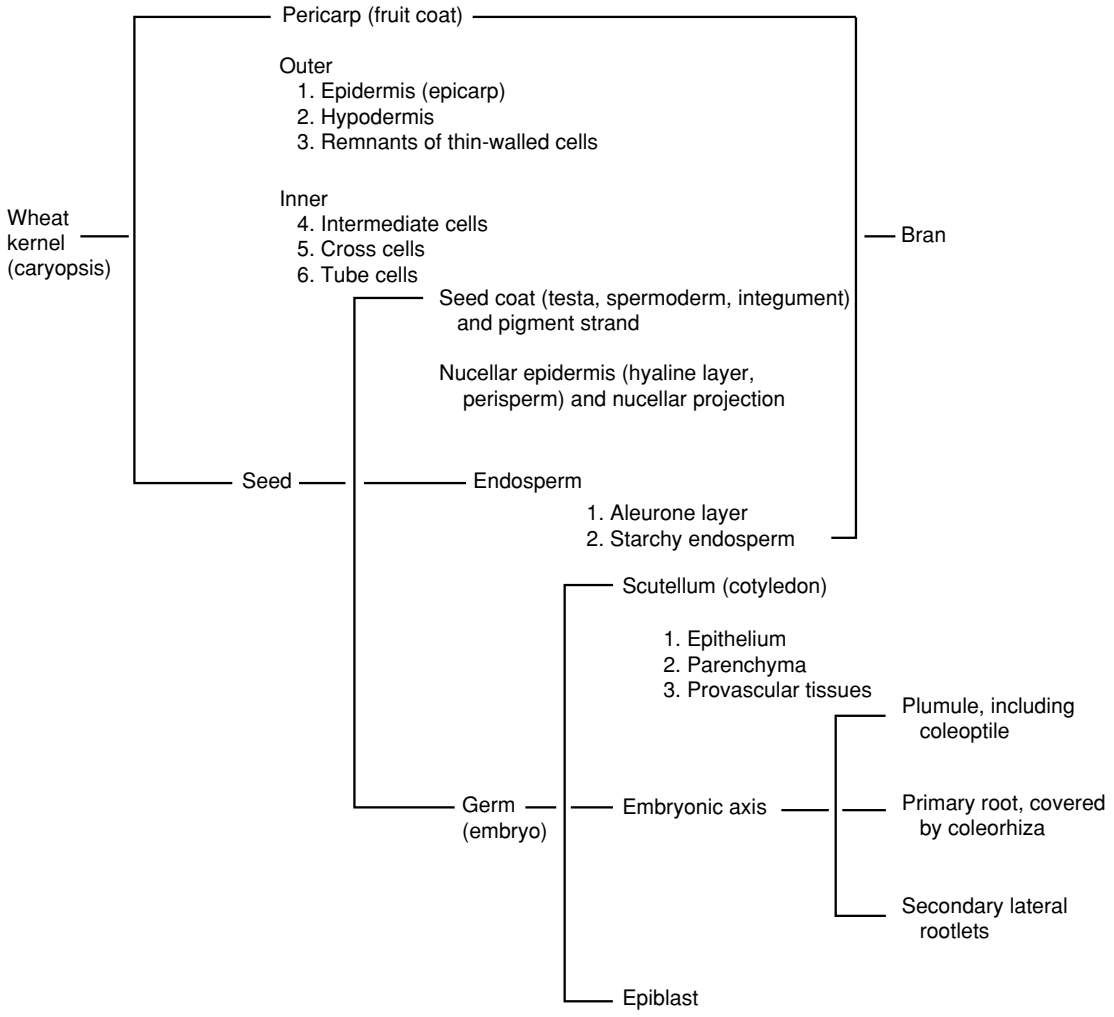

Figure 2 Parts of the wheat kernel. Reproduced from Hoseney RC (1994) Structure of Cereals. In: *Principles of Cereal Science and Technology*, 2nd edn. American Association of Cereal Chemists with permission.

(hardness) and appearance (vitreousness). In general, high-protein hard grains tend to be vitreous and low-protein soft grain tend to be opaque. Some wheat grains are vitreous or translucent in appearance, while others are opaque, mealy, or floury. In vitreous kernels, with no air spaces, light is diffracted at the air–grain interface but then travels through the grain without being diffracted again and again. With vitreous kernels, the protein shrinks but remains intact, giving a denser kernel. As expected, the presence of air spaces within the endosperm makes the opaque grain less dense. The air spaces are apparently formed during drying of the grain. If grain is harvested immature and dried by freeze-drying, it becomes entirely opaque. This shows that the vitreous character results during intensive drying in the field. It is also well known that vitreous grain wetted and dried in the field, or for that matter in the laboratory, will lose its vitreousness. In durum wheat (*T. durum*), which is much harder than the common hard bread wheats, a much larger number of broken starch granules occur when the grain is fractured.

Germ

The wheat germ is composed of two major parts, the embryonic axis (rudimentary root and shoot) and scutellum, which functions as a storage organ. The scutellum is adjacent to the endosperm and contains the remaining two-thirds of the grain's thiamin content. The germ is quite high in vitamin E (total tocopherol) and B-vitamins, and contains many enzymes. It is relatively high in protein (25%), sugar (18%), oil (16% of the embryonic axis and 32% of the scetullum are oil) and ash (5%). The sugars are mainly sucrose and raffinose. The wheat germ comprises 2.5–3.5% of the kernel. Recovery of the germs

during the milling process is an important step because of its value in the food and pharmaceutical industries.

Techniques of Wheat Grain Structure Study

Many cereal scientists and technologists have successfully employed the three main branches of microscopy – light (LM), scanning electron (SEM), and transmission electron microscopy (TEM) – to study the structure and composition of wheat and wheat-based products. The objective of the studies varies from gaining fundamental information on the accumulation of cellular constituents in the developing wheat grain to providing information which can improve our understanding of differences in processing ability or overall quality of wheat and wheat-based products. Microscopic observation should always be linked to other technological, chemical, or physical data. This information will assist the microscopist to select the most appropriate preparation techniques and evaluate the nature and magnitude of any artifacts that may be created. In many cases a number of microscopic techniques are used in order to gain the maximum amount of information on the original structure and composition of the product.

LM and TEM have the advantage that many staining techniques have been developed to assist recognition of constituents and provide data on chemical composition. However, these techniques generally require the use of solvents during sample preparation and these can give rise to artifacts (swelling, shrinking, or leaching of soluble material). Few specific stains are available for use in SEM, and constituents are generally identified by their shape or location. The advantage of SEM is that, in some cases, sample preparation artifacts can be minimized. Recently, X-ray techniques have been used to detect kernel cracks (Figure 3).

Light Microscopy

The main stages in the preparation of samples for examination by LM are fixation, embedding, sectioning, and staining. The aims of fixation are to preserve samples from attack by enzymes or microorganisms, render some constituents insoluble, and strengthen the sample to improve its structural integrity during sectioning. The most commonly used fixation is aqueous, buffered glutaraldehyde but specialized fixation has been developed for specific applications (e.g., fixation of lipid-rich samples). Baked samples, which have been heat-fixed, may not require chemical fixation. Embedding is used to provide additional support during sectioning, and commonly used embedding media are water and aqueous gums for cryostat microtomy, synthetic resins, or special waxes. Resins are generally used where thinner sections are required and are cut using glass knives. Cryostat microtomy and sectioning of wax-embedded samples are carried out using steel knives. Sections, supported on glass slides, are then usually stained prior to mounting and examination.

The stains commonly used for examination with transmitted bright-field illumination are listed in Table 1, together with their substrates. Fluorescence microscopy has been widely used and may rely on autofluorescence or the application of fluorescent dyes. The use of fluorescent dyes coupled to specific antibodies or lectins is rapidly expanding and several are listed in Table 2. Other coupled antibody techniques have been developed whereby colored reaction products are produced. Polarized light can be used to study external starch gelatinization or to provide detail of cell wall structure.

Scanning Electron Microscopy

SEM has a greater depth of focus than LM and therefore it is not necessary to section samples prior to examination. Samples are normally coated with a thin layer of an electrically conducting material, usually gold, platinum, or carbon, prior to examination. Samples of dry wheat (< 12% moisture) or similar products only require air-drying over a desiccant prior to coating. Samples which contain more moisture, may be rapidly frozen prior to dehydration using freeze-drying or critical-point drying. The

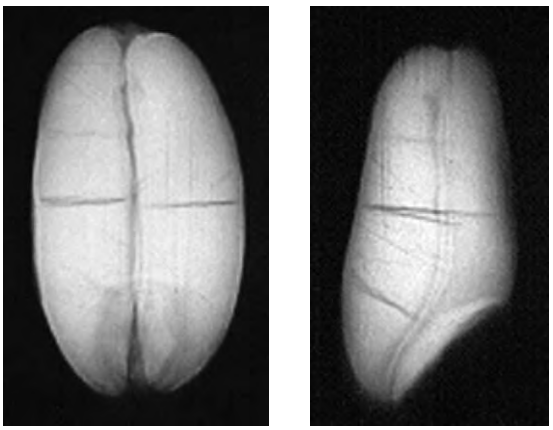

Figure 3 X-ray images of the same wheat kernel with typical radial cracks and transverse to the crease. (a) Front view; (b) side view.

Table 1 List of stains commonly used for bright-field microscopy of wheat and wheat products

Substrate	*Stain*	*Comments*
Starch	Periodic acid–Schiff (PAS) reaction	Covalent, irreversible; may require aldehyde blockade if aldehyde-fixed
	Iodine or potassium iodide	Temporary stain only. Amylose stained blue-black; amylopectin stained red-brown
Protein	Fast green, Ponceau 2R, and other anionic dyes	Ionic, reversible. Intensity influenced by pH and differentiation
	Coupled antibody or lectin stains	Potentially more specific for selected protein
Lipid	Sudan IV or oil red O in 70% ethanol	Temporary stain only; intensity influenced by differentiation
Cell walls	Toluidine blue O	Metachromatic stain. Lignified walls stained green; testa stained purple; other cellulosic walls stained blue
	Zinc-chlor-iodide	Temporary stain only. Lignified walls unstained. Testa stained orange; other cellulosic walls stained blue
Yeast	Methylene blue, pH 4·6	Yeast stained blue; gluten unstained owing to low pH

Table 2 List of stains and conditions commonly used for fluorescent microscopy of wheat and wheat products

Substrate	*Stain*	*Comments*
Starch	Periodic acid–Schiff (PAS) reaction	More sensitive than bright-field PAS for thin sections
Protein	Acid fuchsin	More sensitive than bright-field stain for thin sections
	Coupled antibody or lectin	Potentially more specific
Lipid	Aqueous phosphine 3R	As aqueous solution, lipid distribution less likely to be altered compared to Sudan dyes
Cell walls	Autofluorescence	Walls containing lignin or phenolic acids autofluoresce
	Coupled lectin	Specific for selected glycoprotein

development of cold stages in SEM has obviated the need for drying and this technique, together with freeze-etching, has been successfully used to examine those cereal foods in which moisture is high and forms an integral part of the structure.

Transmission Electron Microscopy

Samples for TEM require fixation and thin sectioning. Initial fixation is usually with glutaraldehyde followed by fixation with osmium tetroxide. The fixed tissue blocks are rinsed, dehydrated, and embedded with a resin prior to sectioning. Heavy metal salts are used to stain the sections and enhance contrast; more recently, antibody staining has been developed whereby the antigenic sites are located at TEM level by the presence of colloidal gold particles, which are coupled to the antibodies.

Image Analysis and Other Quantitative Techniques

During the 1990s the value of all the above work was enhanced by high-tech methods such as magnetic nuclear resonance, laser, and autoradiography methods. More recently, the application of stereology techniques and automatic image analysis of grain have been used. Also, out-of-danger X-ray techniques have been proven to be highly suitable for visualization of inner cracks in kernels. Some internal cracks of wheat endosperm can occur in the field, even before harvesting, due to internal stresses which are mainly induced by high moisture gradients. The same process can be observed during intensive wetting of dry wheat kernel in laboratory conditions. In Figure 3 X-ray images of the same wheat kernel are presented with typical radial cracks of endosperm. X-ray detection makes it possible to identify the position of cracks and also to quantify cracks inside the kernel, and thus to evaluate the physical condition of the grain endosperm.

Wheat-Based Products

The major processing stages to be considered are milling, dough formation, and cooking. The structure of the wheat grain has important implications in the utilization of wheat-based products. This is because millers could grind the grain more finely than is now possible without damaging the starch – a common milling problem. Most breeding programs mainly emphasize grain yield, disease resistance, protein content, and quality while many other features that affect grain quality are considered second or not at all.

Milling

The aim of the milling processes is to remove the endosperm of grain efficiently from the germ and

surrounding bran so that the maximum yield of white flour is achieved with the minimum contamination of nonendosperm material. Fluted rollers are used to break open the grain and scrape the endosperm from the bran; smooth rollers are used to reduce the endosperm particles into flour. The moisture content of cleaned wheat is adjusted to between 15% and 17% prior to being fed to the fluter. This process is called conditioning and its aim is to facilitate separation of bran and endosperm, and to toughen the bran, thereby reducing the amount of ash in the flour.

Conditioning can be completed in approximately 6 h for soft wheat, but hard wheat may require 24 h or longer. The X-ray method has shown that added water moves rapidly through the bran layers, but may remain at the aleurone–endosperm interface for several hours. The rate of water penetration through the endosperm is dependent on protein content, initial moisture, and grain hardness. The air spaces that are present in the endosperm of the soft wheat allow water to move more rapidly and hence conditioning time is less than that required for hard wheat.

Of morphological factors which can influence the yield of clean, white flour, the shape of grain and the amount of endosperm within the grain are particularly important. Grain size has been shown to be significant; larger grains have a higher potential flour yield. Image analysis has been used to measure a large number of morphological parameters, and extraction has been shown to correlate with grain length parameters. However, the manner in which these parameters influence extraction is not clear and further work is required to determine their potential as a screening system at wheat intake.

In addition to the amount of endosperm contained in the grain, the efficiency with which bran and endosperm can be separated ('bran clean-up') is also a major factor influencing flour yield. Because of intrinsic differences in the structure of hard and soft wheat, they must be milled differently. The cell structure of soft wheat is very weak and readily broken. In addition, the endosperm of soft wheat appears to adhere more strongly to the bran. Durum wheat is designed to produce semolina, a granular material analogous to the farina or flour middling of the hard-wheat-milling process. Granular products such as semolina and farina are used for pasta, baby foods, and other specialty foods.

Endosperm hardness influences the manner in which grains fragment, and this can also influence flour yield. Grain hardness is, in turn, influenced by moisture and this is another objective of the conditioning process. The efficiency with which floury endosperm is removed from overlying bran and the degree to which bran is powdered or otherwise damaged are both influenced by the manner in which grain fractures during milling. If fracturing occurs at the boundary between the endosperm cell wall and cell contents (Figure 4), the endosperm is efficiently

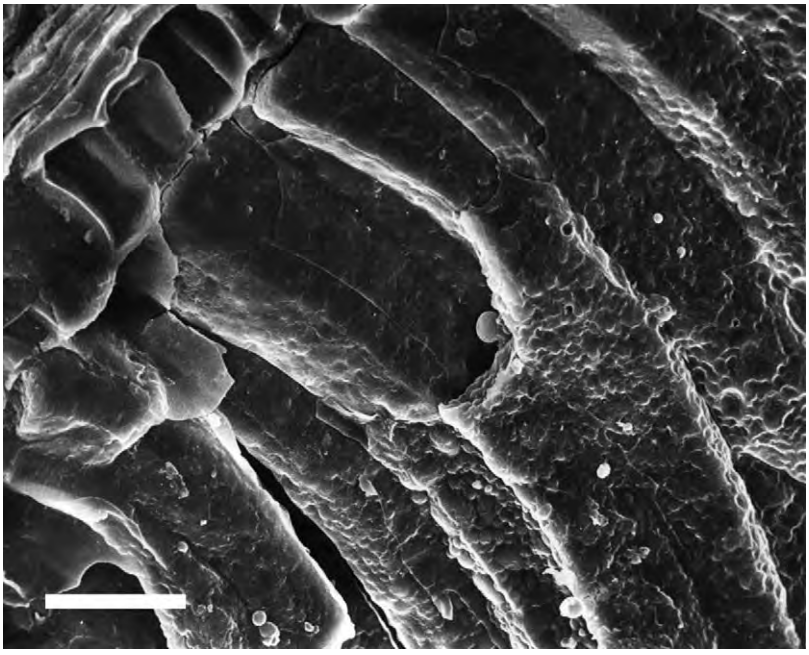

Figure 4 Scanning electron micrograph of subaleurone region of a mealy kernel of spring wheat var. Jara. Note that fracture has taken place at the interface between cell contents and endosperm cell wall. Bar, 50 μm.

removed from the bran, but the bran is fractured into small pieces.

When the endosperm is fractured intracellularly (Figure 5), bran clean-up is poor, but large pieces of bran are produced. The manner in which the grain is fractured is determined by both the inherent hardness of wheat grain and its moisture content. It is thus possible to optimize the relationship between grain hardness, bran clean-up, and powdering by adjustment of moisture level and conditioning time.

Another difference between hard and soft wheat is the point of the endosperm fragmentation. In hard wheat it takes place mainly at the boundary between adjacent cells as the contents of the cells are more firmly bound together by a continuous matrix protein. Thus, hard wheat endosperm can be removed more efficiently from bran as the shear forces imparted by the fluted rollers are directed along the boundary between adjacent endosperm cells towards the bran (Figure 6).

When they reach the bran, some of the forces are deflected along the aleurone–endosperm interface, facilitating separation of endosperm from bran. Of course, as the kernel is reduced to flour size, the hard wheat cell contents are also fractured. In soft wheat, the discontinuities in the protein matrix allow endosperm contents to break apart more easily and cleavage takes place intercellularly. Thus the shear forces are dissipated within the endosperm and are not redirected towards the bran.

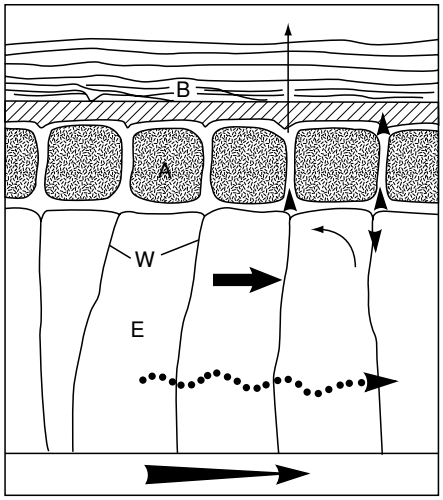

Figure 6 Diagrammatic representation of the forces acting on one endosperm cell of a wheat grain. Large arrow at the base of the figure indicates the direction of the shear force caused by the differential speed of the break rollers. In hard wheat the contents of the cell behave as one unit and hence the cell is cleanly torn from the overlying aleurone layer (indicated by the dashed line). In soft wheat the shear force passes through the content of the cell, as indicated by the dotted arrow. B, bran layer; A, aleurone layer; W, endosperm cell wall; E, endosperm. Reproduced from Wheat: Grain Structure of Wheat and Wheat-Based Products, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

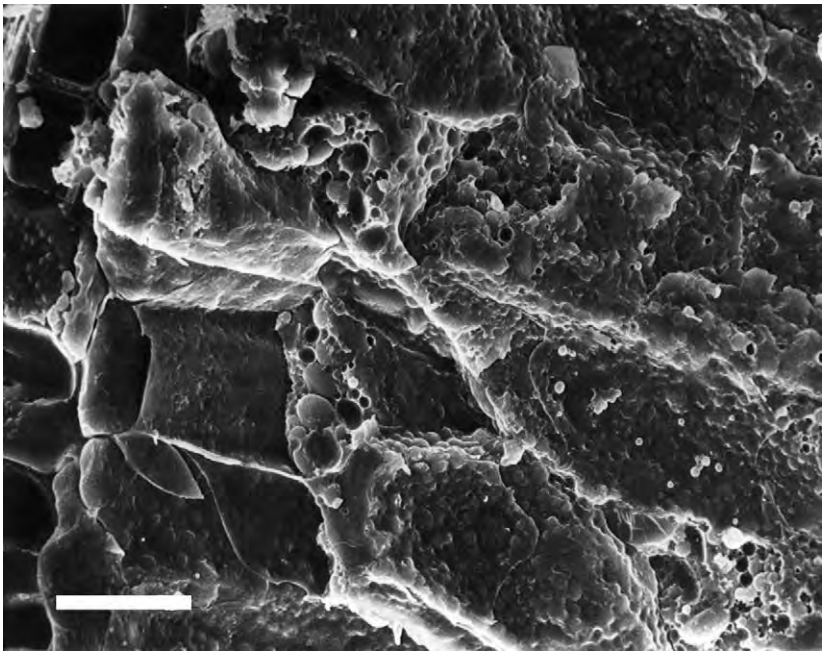

Figure 5 Scanning electron micrograph of the subaleurone region of a mealy kernel of spring wheat var. Jara. Note that the fracture is intercellular. Bar, 50 μm.

Microscopy can also be used to study the mode of action of different items of mill equipment and the effect of processing variables on flour quality. Starch damage is regulated by the amount of pressure applied by the smooth reduction rollers. The physically damaged granules absorb more water and are more susceptible to enzyme attack. A controlled level of starch damage is required for bread flours. The higher roller pressures produce thin flakes of endosperm and microscopic examination has indicated that the appearance of the protein is altered. The protein has been spread over the surface of the starch granules and its functional properties are adversely affected so that cohesive elastic gluten cannot be formed from the protein in the flakes. The appearance of the starch granules in flaked endosperm is usually similar to that seen prior to flaking and the damage is only apparent on hydration, when the granules absorb water and rapidly swell. The amount and nature of bran and wheat germ in stock or flour steams can be determined using microscopy. Bran contamination is important as it adversely influences flour color and ash. Typical wheat has an ash content of about 1.5%. However, the ash is not distributed evenly in the grain. The inner endosperm may only contain 0.3%, whereas the bran may contain 6%. Wheat germ has less effect on flour color, but reduces gluten strength and influences improved requirement in breadmaking.

Dough

When water is added to flour and energy is imparted to the dough by a mixer, the gliadin and glutenin proteins interact to form elastic, cohesive gluten. The mechanism and rate of gluten development depend on the level of water addition and rate of work input. In bread dough the level of water addition (55–62%) is sufficient to cause the mixer initially to pull the gluten away from the starch granules and form coarse, poorly connected masses (Figure 7). As mixing proceeds, these masses are gradually stretched out and become more interconnected and eventually form a uniform, continuous, extensible network, which surrounds the majority of the starch granules in the dough (Figure 8).

Such a gluten network gives dough a smooth external appearance and prevents loss of carbon dioxide produced by yeast during fermentation. This ultimately results in a fine porous structure that is typical of good-quality bread. If mixing is prolonged the protein may become overdeveloped and flow over all starch granules in a thin, veiling film. Carbon dioxide can diffuse through the thin film and escape from the dough, which ultimately results in a small, dense loaf of bread. Overdevelopment only takes place if rate of work input from the mixer is above the critical level for that flour and dough system. If the rate of work input is below the critical level, overdevelopment does not occur and the protein matrix gradually becomes more coarse and discontinuous. This phenomenon has been termed 'unmixing'. Much emphasis has been placed on the mixing process, but

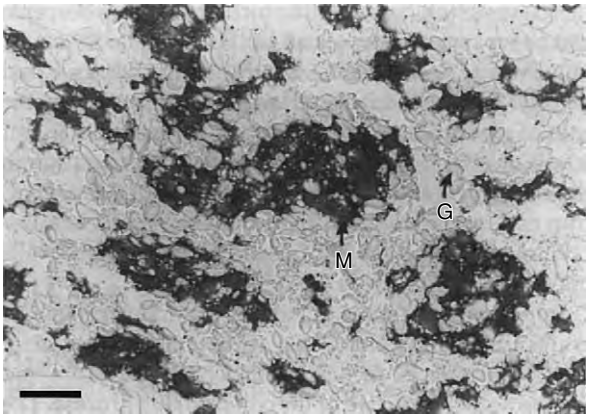

Figure 7 Light micrograph of a cryostat section of an underdeveloped bread dough. The stain was Ponceau 2R. Note the coarse gluten masses (M) that do not surround the majority of the starch granules (G). Bar, 60 μm. Reproduced from Wheat: Grain Structure of Wheat and Wheat-Based Products, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

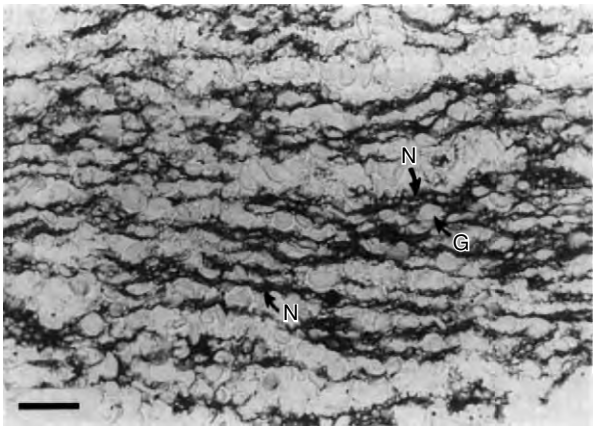

Figure 8 Light micrograph of a cryostat section of an optimally developed bread dough. The stain was Ponceau 2R. Note the coarse gluten masses have become more interconnected and form a continuous network (N) that surrounds the majority of the starch granules (G). Bar, 60 μm. Reproduced from Wheat: Grain Structure of Wheat and Wheat-Based Products, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

the dough structure formed by mixing can be modified by the nature and timing of subsequent dough-processing stages. Dough pieces are usually molded into shape by passage through sheeting rollers. The dough molder can further develop the gluten network or unmix it. Which phenomenon occurs depends on dough strength and the time between molding and the previous processing stages. Generally, flour from the hard wheat grains gives a more adequate structure of dough for breads.

In other types of dough the pattern of gluten development is modified owing to lower levels of water addition or the presence of other ingredients, particularly high levels of fat. In flaky pastries or short, sweet biscuits, a continuous gluten matrix is undesirable, as it would give rise to a tough or hard end product.

The level of water addition in noodle or pasta dough is considerably less (30–33% addition) than that in bread, and no protein pullback occurs during mixing. The function of mixing in these processes is to insure uniform distribution and hydration of ingredients. The continuous gluten matrix is formed by sheeting rollers in the case of noodles and by high-pressure extrusion in the case of pasta.

Cooked Products

One of the major changes that can take place during the cooking of baked products is gelatinization of the starch granules, and the extent to which this has occurred is easily followed using polarized light microscopy. Starch gelatinization can also result in the formation of more significant sample preparation artifacts than occur when samples containing ungelatinized starch are prepared using similar techniques. In some products, the gelatinized starch granules absorb a large amount of water, and freezing or dehydration can give rise to artifacts. However, these can be minimized and appropriate allowance made when interpreting data. In expanded, extruded products the gelatinized starch granules have a high capacity for absorbing water and this gives rise to a high degree of swelling if aqueous reagents are used during processing for LM.

For gelatinization to occur there must be an appropriate combination of heat and moisture and in many baked products the latter can be insufficient. In bread, only a narrow zone of starch granules in the crust still shows some birefringence. However, in biscuits and pastries that contain a high level of fat the majority of the starch granules are frequently still birefringence (i.e., not gelatinized). In some cooked cereal foods the continuous phase, responsible for sample integrity, is gelatinized starch (e.g., products made from batters or from high-pressure cooker-extruders), whereas in others (e.g., all breads, noodles, and pasta) it is gluten. The nature of the continuous phase influences raw material selection and the organoleptic properties of the end product.

Generally, flour from soft wheat grains gives a structure of dough which is more adequate for cooked products. Recent investigations have shown that it is possible to identify the two molecular proteins called puroindolines – pin A and pin B – which correlate perfectly with wheat texture. This discovery should help wheat breeders in the future to develop a supersoft variety for making new kinds of cakes and cookies.

See also: **Microscopy**: Light Microscopy and Histochemical Methods; Scanning Electron Microscopy; Transmission Electron Microscopy; **Milling**: Principles of Milling; Types of Mill and Their Uses; **Wheat**: The Crop

Further Reading

American Association of Cereal Chemists (1990) *Advances in Cereal Science and Technology*. St Paul, MN: American Association of Cereal Chemists.

Angold RE (1979) Cereals and bakery products. In: Vaughan JG (ed.) *Food Microscopy*, pp. 75–136.London: Academic Press.

Betchel DB (1990) Preparation of cereals and grain products for transmission electron microscopy. *Food Structure* 9: 241–251.

Evers AD (1979) Cereal starches and proteins. In: Vaughan JG (ed.) *Food Microscopy*, pp. 139–186. London: Academic Press.

Fornal J, Jeliński T, Sadowska J and Quatrucci E (1999) Comparison of endosperm microstructure of wheat and durum wheat using digital image analysis. *International Agrophysics* 13: 215–220.

Fulcher RG and Wong SI (1980) Inside cereals – a fluorescence microchemical view. In: Inglett GE and Munck L (eds) *Cereals for Food and Beverages*, pp. 1–26. New York: Academic Press.

Hoseney RC (ed.) (1990) Structure of cereals (I). In: *Principles of Cereal Sciences and Technology*. St. Paul, Minnesota, USA: American Association of Cereal Chemists.

Niewczas J, Kudra T, Strumillo P and Wozniak W (2000) Stress cracking as a measure of grain quality after thermal drying. In: Mujumdar AS and Suvachittanont S (eds) *Development in Drying*, vol. II, pp. 1–31. Bangkok, Thailand: Kasertsart University Press.

O'Brien TP and McCully ME (1981) *The Study of Plant Structure. Principles and Selected Methods*, pp. 2.2–4.56. Melbourne: Termarcarphi.

Wozniak W, Grundas ST and Kocon J (1991) Qualitative effect of moisture treatment of wheat grain by means X-ray and SEM techniques. In: Kratochvil J (ed.) *Proceedings of the ICC Symposium*, Prague, vol. 2, pp. 494–499.

WHEY AND WHEY POWDERS

Contents

Production and Uses

J G Zadow, Mount Martha, Victoria, Australia

Origins and Characteristics of Liquid Whey

For many years, whey, the byproduct of casein and cheese manufacture, was treated as a waste product. Most whey was disposed of by feeding to animals, or running to waste in streams or on to the land. The comparatively small amount that was used was mainly directed at the animal-feed market. In the past few decades, however, environmental pressures, coupled with a recognition of the inherent value of whey solids, have resulted in the development of processes for the conversion of liquid whey into a range of valuable food ingredients. Today, whey and whey products are in considerable demand, both as replacement ingredients for more expensive existing products and, more often, as functional and nutritional ingredients in their own right. The use of whey products in the animal feed market is still very important, but in many major markets, such as the USA, domestic utilization of whey powder for human foods is greater than that for animal feed. In Europe, however, animal feed still remains the dominant use for whey powder. This section reviews the source of wheys, compositional factors affecting utilization, alternatives for processing of whey properties of whey powders, and their uses in the food industry.

Little information is available on worldwide production of whey. In 1987, it was reported that the annual world production was about 110 million tonnes, of which about 47 million tonnes were produced in Western Europe, 27 million tonnes in North America, 30 million tonnes in Eastern Europe, and 8 million tonnes elsewhere; Australia produces about 1.8 million tonnes. Given the considerable increase in cheese production worldwide since 1987, it is probable that the current world production of whey is up to 50% higher than the figure quoted above.

Therefore, two pressures exist on whey processors to increase whey utilization. First, there is the rapid increase in the total volume of whey produced, owing to increasing cheese manufacture, and second, there is much greater pressure from environmental regulatory bodies to reduce disposal, and increase utilization of whey. A detailed survey of whey utilization and disposal practices was carried out in Australia in the early 1990s, and many of the conclusions are probably applicable to other major whey-producing countries. In Australia, about 50% of the total whey produced is used (an increase from 46% utilization in the mid-1980s), with the remainder being disposed mostly on to land. However, in spite of the considerable increase in whey production over the past decade, the percentage of whey used is believed to have remained approximately constant. In the 1980s in Australia, membrane processing was a popular means of whey utilization, representing about 25% of the whey used. By the early 1990s, this level had decreased to about 10% and is currently believed to have fallen even lower. By far the greatest means of whey utilization now is via whey powder manufacture (of interest is the fact that, although membrane processing represented only a comparatively small amount of the whey processed, all of the permeate produced was used).

It should be noted that the greatest disposal problems were encountered by factories that produced only moderate amounts of whey. The major options for disposal of whey, either through drying or through membrane processing and permeate utilization, were often too capital-intensive and scale-sensitive to be an option for such companies. The development of economic whey utilization techniques for such manufacturers would be of considerable benefit to the industry.

Source and Composition of Wheys

Whey is the watery substance remaining after coagulation of the casein in milk, either through the

addition of acid (as in casein manufacture) or through the action of a protease such as chymosin (as in cheese manufacture). Clearly, the composition of whey varies considerably, depending on the source of the milk and the manufacturing process involved. However, on average, whey contains about 65 g of solids per kilogram, comprising about 50 g of lactose, 6 g of protein, 6 g of ash, 2 g of nonprotein nitrogen, and 0.5 g of fat. Casein whey generally has a significantly higher level of ash than cheese wheys.

It is convenient to classify whey into three major types:

1. Sweet wheys: titratable acidity 0.10–0.2%, pH typically 5.8–6.6. This category would include wheys produced from chymosin-coagulated cheeses with a low level of acidity.
2. Medium acid wheys: titratable acidity 0.20–0.40%; pH, typically 5.0–5.8. This class could include whey from the manufacture of fresh acid cheese such as ricotta or cottage cheeses.
3. Acid wheys: titratable acidity greater than 0.40%, pH less than 5.0. This class would include casein wheys made by addition of mineral acids, and some fresh acid cheese wheys.

A detailed composition of dried sweet and acid wheys is shown in Table 1.

The following points are of particular relevance in assessing options available for the processing of whey streams:

1. Whey has a total solids of about 6.5%, i.e., it is a fairly dilute product. Thus, to produce 1 kg of whey powder requires the removal of about twice as much water as does the production of 1 kg of milk powder. Removal of water is a costly unit operation, and this factor alone mitigates against many options for whey processing.
2. Of the total solids in whey, more than 75% is lactose. The effective utilization of whey is therefore inextricably linked with the effective utilization of lactose. Unfortunately, lactose is not a commercially valuable sugar, as it is not particularly soluble, and it is not particularly sweet (Table 2). These factors limit the commercial applications of whey solids considerably.
3. The proteins present in whey comprise about 50% β-lactoglobulin, 25% α-lactalbumin, and 25% other proteins. Whey proteins have a very high nutritional profile, are rich in essential amino acids, and can have excellent functional properties. Clearly, the whey proteins are the most valuable components of whey, and most whey-processing operations (e.g., ultrafiltration, manufacture of lactalbumin) therefore aim to increase

Table 1 Typical composition (%) of sweet and acid whey powders, and whey protein concentrates

	Moisture	*Crude protein*	*True protein*	*Lactose*	*Fat*	*Ash*
Sweet whey[a]	3.2	12.9		74.4	1.1	8.4
Acid whey[a]	3.5	11.7		73.4	0.5	10.8
35% WPC[b]	4.6	36.2	29.7	46.5	2.1	7.8
50% WPC[b]	4.3	52.1	40.9	30.9	3.7	6.4
65% WPC[b]	4.2	63.0	59.4	21.1	5.6	3.9
80% WPC[b]	4.0	81.0	75.0	3.5	7.2	3.1

[a]Data from Posati LP and Orr ML (1976) *Agriculture Handbook, No. 8–1*. Washington, DC: US Department of Agriculture.
[b]Data from Glover FA (1985) *Ultrafiltration and Reverse Osmosis for the Dairy Industry. Technical Bulletin, No. 5*. Reading, UK: National Institute for Research In Dairying.

Table 2 Relative sweetness and solubility of lactose, sucrose, and some monosaccharides

Sugar	*Relative sweetness*[a]	*Solubility (g per 100 g of solution)*[b]		
		10 °C	*30 °C*	*50 °C*
Sucrose	100	66	69	73
Lactose	16	13	20	30
D-Galactose	32	28	36	47
D-Glucose	74	40	54	70
D-Fructose	173		82	87

[a]Data from Pazur JH (1970) Oligosaccharides. In: Pigman W, Horton D and Herp A (eds.) *The Carbohydrates: Chemistry and Biochemistry*, p. 69. New York: Academic Press.
[b]Data from Shah NO and Nickerson TA (1978) Functional properties of hydrolyzed lactose: solubility, viscosity and humectant properties. *Journal of Food Science* 43: 1081.

the proportion of whey proteins in the end product.
4. The mineral content and low pH of casein wheys severely limit their commercial exploitation. The vast majority of whey-based products are manufactured commercially from low- or medium-acid wheys.
5. Whey has a very high biochemical oxygen demand (BOD), which poses major difficulties for its disposal. It should be noted also that a number of options for whey processing, particularly those that increase the proportion of protein in the product, also result in the production of a waste product, which contains most of the lactose originally present. This stream, in turn, requires further processing. Thus, the problems posed by the biochemical oxygen demand of the original whey are often only slightly affected by many of the whey-processing options.

Processing Options

The processing options for whey fall into four main areas:

1. Those concerned with simple removal of water (spray or roller drying to yield whey powder).
2. Those concerned with increasing the ratio of protein in the end product (ultrafiltration for the manufacture of whey protein concentrates, fractionation processes for the manufacture of protein isolates, heat treatment for the production of lactalbumin).
3. Those concerned with utilization of the lactose in whey (treatment with lactase or heat/acid for lactose-hydrolyzed products, fermentation to a range of products such as lactic acid, citric acid, and single-cell protein).
4. Those designed to alter the mineral composition of the product (electrodialysis and ion exchange for the manufacture of demineralized products).

Each of these is considered in turn below.

Drying of Whey

The production of whey powder from whey is a mature technology. Perhaps the major difficulty involves the holding period required prior to drying in the manufacture of nonhygroscopic whey powder. Such a holding period (sometimes also involving seeding) is required to allow for crystallization of the lactose into a nonamorphous, nonhygroscopic form prior to drying. The development of technology to reduce this holding period could result in significant savings to the industry.

The spray drying of whey is a fairly straightforward operation, with conditions employed similar to those used for the spray drying of milk. Thus, the whey is concentrated to 40–70% total solids and spray (or roller) dried to a moisture content of less than 5%. As indicated above, the drying of whey is complicated by its high lactose content. Lactose exists in two isomeric forms, α-lactose and β-lactose. α-Lactose crystallizes as a hydrate, whereas solid β-lactose contains no water of crystallization. However, when solutions of whey are dried rapidly, there may be insufficient time for the α-lactose to form as the monohydrate, and it forms as amorphous α-lactose. The dry lactose in the whey product is then essentially in the same form as in the liquid. Neither α-lactose hydrate nor β-lactose is hygroscopic. However, amorphous α-lactose is highly hygroscopic and will absorb moisture from the air, resulting in a hydrate that occupies more space than the amorphous form. This effect causes the commonly observed lumping and caking in many whey powders.

Both hygroscopic and nonhygroscopic whey powders are manufactured. The former are produced by simple drying of the whey concentrate. The manufacture of nonhygroscopic whey powders relies on the conversion of much of the lactose in the liquid concentrate to a crystalline form prior to drying. This is achieved by holding the concentrate under appropriate conditions to allow for extensive formation of α-hydrate crystals. The traditional process requires holding the concentrate for about 16–24 h after addition of seeding crystals. This process requires holding tanks of considerable size and is energy-intensive (good agitation is required), and the viscosity of the concentrate must be as low as possible (heat treatment of the concentrate should be minimized). Recently, a new process has been described, based on the fact that most of the lactose is crystallized in 3–4 h. In this process, the whey is concentrated to 53–55% TS, flash-cooled to about 30 °C, and seeded. The product is pumped to a conventional nonagitated tank, representing 2–3 h of production. The semicrystallized concentrate is pumped to a second flash cooler and cooled to 15 °C, and then to a second crystallization tank, where it is held ready for spray drying. Alternatively, a process similar to instantizing may be used, in which the surface of the partly dried whey powder particles is partially humidified prior to completion of the drying operation. This stage permits additional crystallization of the α-lactose during drying.

The market for permeate powders is not as yet well developed, probably because of difficulties in obtaining a consistent product, and difficulties encountered during drying caused by the high-lactose/low-protein content of the product. Worldwide, the

difficulties encountered in spray-drying permeate have led some companies to cease production of permeate powder, and to adopt perhaps less profitable alternative technologies. Research suggests that the drying properties of permeate may be considerably influenced by the pH and of course the crystallizing conditions. The development of successful means for the drying of permeate would be of particular value to the industry, as the production of permeate powder is a simpler option for many companies than, for example, the production of crystalline lactose. In response to this need, equipment supply companies are continuing to develop systems specifically for permeate drying.

Means for Increasing the Protein Content of the End Product

Whey solids contain about 11% protein. Many of the most popular methods of whey treatment aim to increase this level, with end products containing between 35% and virtually 100% protein. It should be noted that each of these methodologies results in a waste stream high in lactose, which will pose separate utilization or disposal problems.

Ultrafiltration Ultrafiltration is the most common method used by the dairy industry to produce a range of whey products with an increased protein content, known as whey protein concentrates (WPCs). Ultrafiltration relies on the passage of whey near a membrane with a pore size such that low-molecular-weight materials such as lactose and salts pass through the membrane, whereas higher-molecular-weight components, such as proteins, are retained. On ultrafiltration of whey, therefore, the solids content of the product retained by the ultrafiltration membrane (the retentate) is higher in protein and lower in lactose than the original whey, and the solids content of the product that passes through the membrane (the permeate) is high in lactose and ash, and has minimal protein content. WPCs are produced from a wide range of wheys, generally to protein contents of 35%, 50% and 75% (Table 1).

WPC containing 35% protein is often used as a skim milk powder replacer in applications where the specific functionality of skim milk powder is not important (a WPC of 35% protein content is generally significantly less expensive than skim milk powder). WPC with a protein content of 50% is not widely manufactured and generally is used for specific applications only. WPCs containing 75% protein can have very desirable functional properties, and these can be readily manipulated by modifying the manufacturing process. Such products often have excellent water-binding, gelation, and emulsifying properties, making them sought after by the food industry as functional ingredients.

Production of lactalbumin Whey proteins are heat-sensitive and can be precipitated by heat treatment under appropriate conditions of pH and ionic strength. This property is exploited in the manufacture of lactalbumin. (Note that lactalbumin, the product of heat precipitation of the proteins from whey, contains a mixture of denatured α-lactalbumin, β-lactoglobulin and other whey proteins. 'Lactalbumin' should not be confused with α-lactalbumin.) In the manufacture of lactalbumin, whey is heated to denature coagulate and precipitate the whey proteins; the sediment is recovered by settling and decantation (or centrifugation), washed to remove excess salts and lactose, and the product recovered by centrifugation or filtration prior to drying, grinding, and bagging. The heat treatment used in the manufacture of lactalbumin results in extensive denaturation of the whey proteins, resulting in a product of poor functionality. Therefore, lactalbumin finds its best applications in products where protein fortification is necessary, but it is not required to provide any functional properties.

Isolation and fractionation of protein In contrast to the precipitation of whey proteins by heat treatment in the manufacture of lactalbumin, protein isolation and fractionation methodologies aim to separate the proteins from whey in such a form that they remain, as far as possible, fully undenatured and thus retain their functionally. These products (protein concentrates and isolates) have a high protein content and can have exceptional functional properties of considerable value to the food industry.

Protein concentrates contain the whey proteins in about the same proportions as that in whey. (Note that, in this article, the term 'protein concentrates' is used for high-protein products containing the individual whey proteins in about the same ratio as that present in whey, the term 'whey protein concentrate' is used for such products manufactured by ultrafiltration, and the terms 'protein isolates' and 'protein fractions' are used to refer to high-protein products with a higher ratio of a particular protein than that present in whey.) Such protein concentrates are generally manufactured by the use of a nonspecific absorbent to bind the proteins in whey, followed by elution of the proteins by treatment of the absorbent with a specific eluent. Absorbents that have been commercially used include carboxy-methylcellulose and a range of mineral oxides. Although these absorbents are comparatively nonspecific, they can show a preference for binding particular proteins

under set conditions of pH, temperature, and ionic strength. Thus, these processes can be used to produce protein isolates, for example with a higher ratio of β-lactoglobulin to α-lactalbumin than that present in whey.

Protein fractionation technology is also developing, which relies on separation of α-lactalbumin from β-lactoglobulin on the basis of their different solubilities under specified conditions of pH, temperature, and ionic strength. It is therefore possible, for example, to separate by sedimentation most of the α-lactalbumin from whey by careful manipulation of processing conditions. Both the α-lactalbumin that has sedimented and the residual soluble protein (mostly β-lactoglobulin) are comparatively undenatured and thus retain their high functionality. The conditions employed in these processes are very mild and do not result in any denaturation of the whey proteins. With such processes, it is therefore possible to produce isolates rich in β-lactoglobulin (with an extremely high gel strength) and α-lactalbumin (a product that may have considerable potential in non-allergenic infant foods).

Lactose-processing Options

The options for treatment of whey involving lactose may be divided into three groups – those involving a fermentation step, those involving separation of the lactose and its utilization, and those involving enzymatic hydrolysis of the lactose to produce galactose and glucose.

Fermentation Many options for the fermentation of whey are described in the literature, including the production of biogas, biomass, ethanol, lactic acid, and citric acid. However, the dairy industry worldwide has not taken up such opportunities to any great extent.

Separation of lactose In many respects, this is a most attractive option, as it can be used also as a process for the treatment of waste streams from other whey treatment operations such as ultrafiltration. The manufacture of lactose (normally α-lactose hydrate) generally involves the removal of protein, concentration, refiltration, further concentration, induction of crystallization, and separation of crystals with a basket centrifuge.

The production of lactose in the major dairy countries is increasing rapidly – in the USA, production has tripled in the past 14 years, with exports doubling in the past 6 years, and in Australia, production has at least doubled. Similar increases in Europe have benn reported. Overall, the lactose market is holding steady through increased applications for the product, particularly in the confectionery field.

Hydrolysis of lactose The hydrolysis of lactose yields the sweet soluble sugars, glucose and galactose, thus increasing the applications of the product. Such hydrolysis can be carried out by treatment of whey with lactase (β-galactosidase) or by treatment of deproteinized whey at an elevated temperature and low pH. It should be noted that it is difficult to dry hydrolyzed wheys, because of the tendency of the monosaccharides formed by the hydrolysis to produce glasses on the surface of the drier.

It is often suggested that lactose-hydrolyzed products are attractive commercial options, first, because hydrolyzed products will be in demand by lactose-intolerant people, and second, the increased sweetness of the products makes them attractive to the general public. In fact, the demand by lactose-intolerant members of the community has not been significant, and the benefits of increased sweetness are of only marginal commercial value.

Hydrolysis of lactose may be carried out by a number of processes, including heat/acid treatment (for permeate only) and enzymic hydrolysis (for whey, permeate, and WPC). A number of hydrolyzed whey products are manufactured, using both immobilized and free enzyme technology. In Sweden, a beverage ('Nature's Wonder') is based on hydrolyzed whey; in Finland, hydrolyzed whey is used by bakeries and in a flavored whey beverage. In the UK, hydrolyzed whey has been used in confectionery products, generally to replace sweetened condensed milk. However, in spite of the apparent potential of these products, market development appears to be proceeding only slowly.

Changes in Mineral Composition

The mineral composition of whey, particularly casein whey, is such that it deleteriously influences the taste and applications of the product. Whole or partial demineralization of whey is therefore a popular option with manufacturers. In the past, this has been accomplished most often by treatment of whey by ion exchange (no preferential removal of ions) or electrodialysis (preferential removal of monovalent ions). However, these methods appear to be falling into disuse both because of their high cost and because of large volumes of intractable effluents that they produce.

More recently, an option for demineralization using 'open' reverse-osmosis membranes has been developed. This technique, known as ultra-osmosis, uses membranes that allow the passage of ions and

water, whilst retaining all other whey components, including lactose. Ultra-osmosis is rapidly gaining in application, in spite of the fact that its ability to reduce mineral content to much below 50% of original levels is limited. However, it does allow for preferential removal of monovalent ions, a considerable advantage.

Worldwide, demand for demineralized whey products is increasing rapidly, with emphasis on products containing reduced monovalents.

Applications

Whey solids may be used an ingredient in products such as calf milk replacers, infant formulae, whey cheese, beverages, baked goods, icecream and other dairy products, comminuted meat products, and imitation milk products. In most cases, the whey solids contribute little to the functionality of the product, offering only a comparatively low-cost source of protein, carbohydrate, and calcium.

Similarly, lactalbumin is used in foods where protein fortification is required, but additional functionality is not essential. Lactalbumin is commonly used in products such as baked goods and comminuted meat products.

WPCs are used where both protein fortification and functionality are required, although WPC of 35% protein content is generally used directly as a cost-effective skim milk powder replacer. WPC of 75% protein content, with its excellent gelation properties, is often used as a cost-effective egg white replacer.

The main market for demineralized wheys is in the formulation of cows' milk-based infant formulae with a composition closer to that of human milk. Demineralized wheys also may be used effectively as beverage ingredients, where the saltiness of the undemineralized product might normally preclude its use.

Lactose is used in sauces, instant drinks, and meat products, where its low sweetness and ability to enhance flavor are advantages. Lactose is also used extensively in the confectionery and baked goods industries. Highly purified lactose is also used in the pharmaceutical industry for tablet manufacture and as a substrate for the manufacture of penicillin and other fermented products.

Applications of hydrolyzed wheys include ingredients in foodstuffs, such as beverages, and other products, such as most animal foods, where it can be used as a humectant to replace the more expensive glucose commonly used.

See also: **Casein and Caseinates**: Methods of Manufacture; Uses in the Food Industry; **Cheeses**: Types of Cheese; **Drying**: Spray Drying; **Lactose**; **Membrane Techniques**: Principles of Ultrafiltration; Applications of Ultrafiltration; **Powdered Milk**: Milk Powders in the Marketplace; Characteristics of Milk Powders

Further Reading

Fox PF (ed.) (1985) *Developments in Dairy Chemistry*, vol. 3. London: Elsevier.

IDF (1987) *Trends in Whey Utilization. Bulletin No. 212*. Brussels: International Dairy Federation.

IDF (1988) *Trends in the Utilization of Whey and Whey Derivitives. Bulletin No. 223*. Brussels: International Dairy Federation.

Marshall KR (ed.) (1979) Proceedings of Whey Workshop II 2–6 April, 1979. *New Zealand Journal of Dairy Science and Technology* 14: 73.

Morr CC and Melachouris N (eds) (1984) Symposium: Production and Utilization of Whey and Whey Components (1983). *Journal of Dairy Science* 67: 11.

Robinson RK (ed.) (1986) Modern dairy technology. In: *Advances in Milk Processing*, vol. 1. London: Elsevier.

Sienkiewicz T and Riedel C-L (1990) *Whey and Whey Utilization*. Gelsenkirchen-Buer, Germany: Verlag Th. Mann.

Zadow JG (ed.) (1992) *Whey and Lactose Processing*. London: Elsevier.

Protein Concentrates and Fractions

J G Zadow, Mount Martha, Victoria, Australia

Introduction

Over the past 20 years, whey processing has been revolutionized by the development of methodologies for the manufacture of highly functional food products containing a higher proportion of protein than that found in whey. Functionality has been defined as 'any property of a food or food ingredient, except its nutritional ones, that affects its utilization.' The main functional properties of whey protein products used for food applications include solubility, whipping, flavor, water binding, thermal extrusion, viscosity, gelation, foam formation, and emulsification. The technology involved in the manufacture of functional whey-based ingredients generally involves either ultrafiltration or the selective separation of whey proteins (either by adsorption technology or by utilization of the different solubility characteristics of whey proteins under specified conditions of pH and temperature). This article will discuss the properties of the major proteins of whey, whey protein products

and methods for their manufacture, and the application of such products in the food industry.

Characteristics of Whey Proteins

Whey proteins are those which remain in solution after removal of the caseins from milk, either by treatment with chymosin or by acidification. The distribution of proteins in milk is shown in Table 1.

β-Lactoglobulin is the principal protein in whey, comprising about one-half of the total protein present. It occurs as a number of variants, and has a monomer molecular weight of about 18 000. It should be noted that β-lactoglobulin exists as a monomer only outside the pH range 3.5–7.5. Inside this range, it generally exists as a dimer although, under certain circumstances, some variants may form an octomer. β-Lactoglobulin is comparatively heat-sensitive and may be denatured by heat treatment much above 60 °C. α-Lactalbumin has a monomer molecular weight of about 14 000, and is somewhat more heat-resistant than is β-lactoglobulin. There are many other serum proteins (including immunoglobulins) in whey, most of which are readily heat-denaturable. For example, heat treatment of skim milk at 70 °C for 30 min denatures only 6% of the α-lactalbumin, but 32% of the β-lactoglobulin, and 89% of the immunoglobulins; cumulatively, such a heat treatment results in denaturation of about one-third of the total serum proteins. In whey, heat treatment is often carried at a pH quite removed from that of milk. It should be noted that the pH of heat treatment of whey has a considerable influence on the magnitude of the degree of denaturation of each of the whey proteins.

As globular proteins, both α-lactalbumin and β-lactoglobulin have the potential to be highly functional food ingredients. Unfortunately, many of the processes used for whey protein recovery result in their partial denaturation, which generally reduces the functionality of the product.

Table 1 Typical protein composition of milk

	g per 100 g milk
Colloidal casein[a]	2.36
Serum casein[a]	0.26
β-Lactoglobulin[b]	0.32
α-Lactalbumin[b]	0.12
Bovine serum albumin[a]	0.03
Total immunoglobulins[b]	0.075
Other proteins[a]	0.06
Nonprotein nitrogen components[b]	0.08

[a]Data from Farrell HM Jr and Thompson MP (1974) *Physical equilibria in milk proteins*. In: Webb BH, Johnson AH and Alford JA (eds) *Fundamentals of Dairy Chemistry*, p. 465. Westport, CT: AVI Publishing.
[b]Data from Walstra P and Jenness R (eds) *Dairy Physics and Chemistry*, p. 71. New York: John Wiley.

Major Products and Applications

Whey Protein Concentrates

Whey protein concentrates are manufactured by the ultrafiltration of whey. In this process, whey is passed against a semipermeable membrane, which selectively allows passage of low-molecular-weight materials such as water, ions, and lactose, whilst retaining higher-molecular-weight materials such as protein in the retentate. The retentate is then further concentrated by evaporation and spray-dried to yield whey protein concentrates (WPCs). WPCs are generally available commercially with a protein content of about 35, 50, or 75%. For the higher protein products a process known as diafiltration is employed, in which water is added to the retentate during manufacture to wash out more of the low-molecular-weight materials from the retentate.

The development of markets for WPCs has been slowed by the inconsistent quality and composition of the product, particularly during the early years of development. WPC 35% is often considered to be a skim milk powder analog, and for many skim milk powder applications, it may be utilized as a direct replacement. However, it is only in recent years, with improved product consistency and quality, that markets for this product have firmed. It should also be noted that this is not a particularly lucrative market for whey producers, as 35% WPC is generally priced at a significant discount to skim milk powder.

A WPC with a higher protein content, particularly the 75% product, is generally required by the end user to have defined and reproducible functionality. Again, however, inconsistent quality, due perhaps to a lack of understanding of processing conditions on functionality development, has slowed market acceptance. In some cases, however, supplies of 75% WPC have outstripped demand and this highly nutritional product has been sold for animal feed on the basis of its protein content.

A further problem with WPC development is the need to develop markets for the products of permeate processing. Whilst lactose markets are strong and continuing to increase, significant concerns about the ability of the industry to continue to cope with the increasing levels of permeate production worldwide remain.

There appears to be a slowing in the rate of installation of ultrafiltration membrane processing for the production of WPC by the industry. This is probably the result of a number of factors, including:

1. Such technology is only applicable to major whey producers, and many have already installed this technology (although some of these are reverting

to the production of whey powder for whey utilization).

2. The market response to the products has to date been disappointing. The early promise of extremely high returns for high-protein (>75% protein) WPC has often not been achieved, mostly because of inconsistent functionality of such products. However, there is the potential for good returns provided that quality and functionality can be maintained consistently. Products containing 35% protein have in general not proven particularly commercially attractive.
3. Adopting this technology does not solve any disposal problems – the permeate still requires processing or disposal. Further, given the fact that it is virtually free of protein, permeate is generally a less commercially attractive product to utilize than whey.

The inclusion of WPC into foodstuffs may confer a wide range of potential functionality. Important functional properties include solubility, foaming, gelation, emulsification, elasticity, viscosity, and organoleptic characteristics. It is possible, by manipulating compositional factors and processing conditions, to manufacture WPCs with a wide range of functionality, combining many of the above attributes. In general, WPCs of lower protein content have more limited functionality than those of higher protein content. In many cases, WPCs serve more than one functional purpose in foods. For example, as whey proteins remain soluble over a wide pH range, in particular near pH 4.5, they may be used in acidic drinks as protein fortifiers. WPCs can be used as water binders in products such as baked goods and processed meats. In such cases, processing temperature must be sufficiently high to denature the whey proteins, but not so high as to disrupt their water-binding properties. WPCs can also have excellent gelation characteristics, and can assist in the formation of heat-induced gels in meat products and baked goods. WPCs with good emulsification properties would find applications in products such as sausages, salad dressing, coffee whitener, and soup. WPCs may also have useful foaming properties, producing very stable foams on whipping. Applications for such products again vary depending on the functionality required in the end product, but would include desserts, cakes, and whipped products.

A major challenge to whey protein researchers and producers is to determine precisely the compositional and processing properties which influence the final functionality of WPC. It is well known that factors such as the source of whey, its protein content, the heat treatment applied to the whey or ultrafiltration retentate during manufacture, and the lipid content and mineral content of the product influence functionality, but detailed knowledge of mechanisms is far from complete. The manufacture of WPCs with such widely varying and yet specific functional properties certainly requires the careful control of manufacturing conditions. Currently, much ultrafiltration processing of whey is carried out on an *ad hoc* basis, with little understanding of the effects of individual processing parameters on the conformation or structure of the individual proteins, or on the functionality of the product. It should also be noted that the functionality of the whey products can be permanently impaired by the use of excessive temperatures or extremes of pH during processing.

Mineral composition is also an important factor determining whey protein functionality. Many of the functional properties of WPCs are considerably influenced by either demineralization or by the addition of calcium salts. Clearly, the manufacturing process used in the whey production process will influence the mineral content of the product, and thus its functionality.

Whey Protein Isolates

There have been some commercial developments recently in the manufacture of whey protein isolates. Generally these products are manufactured by the use of specific complexing agents which bind with the proteins, allowing their removal from the whey. The processes use absorbents such as carboxy-methylcellulose (such as in the Bi-Pro process), or inorganic oxides (such as with Spherosil reagents) or complex formation (with reagents such as polyphosphates). The end products may then be sold in the form of these complexes, or the proteins eluted and dried for sale as protein isolates. The products of these operations are highly functional, high-value-added whey protein fractions which have high specific functionality and generally sell for a high premium. Those products sold as complexes have less general applicability, and attract reduced profit margins. However, in spite of their availability, the market for these products has proved to be somewhat limited, and demand has not been as high as expected. It is believed that some difficulties have been experienced in insuring that the products of such processes have consistent functionality when utilized in foods.

Whey Protein Fractions

Whey proteins can be recovered as mixtures of the proteins in whey by use of the technologies outlined above. Commercially, ultrafiltration technology is dominant for such processes. However, WPCs

manufactured by ultrafiltration have not, in general, lived up to their promise of offering reliable highly functional cost-effective ingredients to the food industry. In particular, the functionality of WPCs is often poor and/or highly variable. Further, the functional properties of WPCs are often well below those expected from their protein composition. This may be due to mechanical or heat damage to the proteins during manufacture, or to the presence of other compounds in whey which inhibit the development of full functionality.

To overcome such difficulties, a number of alternative options for whey protein fractionation have been developed. These aim at the manufacture of protein fractions containing a higher proportion of a particular protein than that present in whey solids.

Ion exchange Ion exchange or adsorption was discussed briefly above as a means for the separation of proteins from whey. However, the adsorbents used can also selectively adsorb proteins from whey under specified conditions of pH and temperature. For example, under appropriate conditions, certain Spherosil adsorbents can selectively remove a significant proportion of the β-lactoglobulin from whey, leaving a fraction rich in α-lactalbumin as the effluent from the adsorption process.

Such processes generally use batch operation, involving adsorption, elution, and regeneration.

Ion depletion This technology relies on the fact that β-lactoglobulin is insoluble in solutions of low ionic strength, particularly in the vicinity of its isoelectric point. This principle has been used in a number of studies which have led to pilot-scale separation procedures. The products in general were substantially β-lactoglobulin, although other proteins were present. However, the yield of the process was poor, with only about 30% of the protein present in whey recovered as the precipitate.

Thermal separation A series of studies have recently shown that the solubility of α-lactalbumin decreases very markedly under certain conditions of pH, temperature (below that of denaturation), and ionic strength. These conditions have no effect on the solubility of β-lactoglobulin. Clearly, this difference in solubility characteristics may be exploited as a means of preparing whey protein fractions. Two processes based on this principle have reached the stage of near commercialization – one in France, the other in Australia. The major difference between the two is the starting material – the French process uses untreated whey while the Australian process uses whey concentrated to 12% solids by ultrafiltration. In the Australian process, the pH of the ultrafiltration retentate is adjusted to 4.2, and the process of aggregation initiated by heating of the mixture to 64 °C for 5 min. During this process, the α-lactalbumin aggregates into small particles. The product is then diluted with water to assist in the formation of larger aggregates, and the sediment (which is mostly α-lactalbumin) separated, for example, by centrifugation or microfiltration. The separated sludge is evaporated and dried, to yield a fraction rich in α-lactalbumin (α fraction). The supernatant is subjected to ultrafiltration and diafiltration (to assist in the removal of ash and lactose), and dried to yield a fraction rich in β-lactoglobulin (β fraction). The α fraction contains about 50% protein (mostly α-lactalbumin) and 40% lactose, while the β fraction contains about 75% protein (mostly β-lactoglobulin) and 15% lactose.

It is probable that the products from the French and Australian processes have similar functionality – in each case, the β fraction is low in lipid content. The β fraction has been shown to have excellent gelation characteristics (much greater than those shown by the best 75% WPC). Further, the gel strength exhibited by the β fraction can readily be manipulated by minor modifications to processing conditions. Clearly, this product has considerable potential as a highly functional food ingredient, with applications similar to those of egg white.

The α fraction contains most of the lipid and phospholipid in whey, and should be expected to show excellent emulsifying properties. A further application for the α fraction is in (humanized) infant foods. Although whey-based products are common components of infant foods, whey contains a significant amount of β-lactoglobulin, a protein which has no analog in human milk. On the other hand, human milk does contain an analog to bovine α-lactalbumin. Clearly, therefore, infant foods based on (β-lactoglobulin-free) α fraction may offer considerable advantages in reducing allergenic response.

An alternative thermal procedure is used in the manufacture of lactalbumin. In the manufacture of this product, whey is heated to denature, coagulate, and precipitate the whey proteins present, which are recovered by settling and decanting, or by centrifugation. The product (which should not be confused with α-lactalbumin) is generally virtually fully denatured, and comprises a mix of the heat-labile proteins present in whey. Given the fact the functionality is generally associated with proteins being present in a globular native state, it is not surprising that lactalbumin shows little functionality. However it finds specific applications in the food industry, particularly in biscuits and baked goods. Modified techniques for the preparation of lactalbumin, involving somewhat

less severe heat processing, have been reported, with indications that such products show markedly improved solubility and functionality.

Ferric chloride fractionation Techniques have been described in the literature for the treatment of partially demineralized whey with ferric chloride to precipitate β-lactoglobulin selectively as a ferric complex at near neutral pH. By contrast, at an acidic pH, almost all proteins except β-lactoglobulin are preferentially precipitated. In this case, the separated complex can be solubilized by a change in pH to near neutral, and the ferric ion separated by, for example, ultrafiltration. Such processes do not appear to be close to commercialization.

Removal of Lipid from Whey and Whey Protein Fractions

WPCs inevitably contain levels of residual lipid, in spite of attempts by producers to remove as much as possible As would be expected, levels of residual lipid increase with increasing protein concentration. Reports have suggested levels of about 2.1% fat in 35% WPC, 3.7% fat in 50% WPC, and 7.2% fat in 80% WPC. As can be seen in the higher protein content products, the level of lipid is quite significant. Removal of the lipid from whey results not only in an improvement in ultrafiltration flux, but also markedly improves product functionality. For example, higher levels of lipid have been shown to be detrimental to the foaming and flavor characteristics of WPC. Removal of lipids increases overrun whipping substantially, as well as increasing foam stability. Higher levels of lipid also inhibit the heat gelation properties of WPC.

The composition of the lipid fraction in WPC is not similar to that of bulk milk, being higher in phospholipid and milk fat globule membrane material.

Overall, therefore, the lipid fraction in whey is believed to inhibit much of the potential functionality of WPCs, and whey protein fractions. The lipid fraction in whey is also partly responsible for the fouling of the membranes on ultrafiltration processing of whey. Removal of the lipid fraction can thus improve processing efficiency (if ultrafiltration is employed) as well as product functionality.

Removal of such lipids from whey prior to ultrafiltration or fractionation can be achieved by microfiltration using membranes of an appropriate pore size to remove the comparatively large lipid-containing material. However, fouling of microfiltration membranes (presumably also by the lipid-containing fraction) in such processes has, as yet, mitigated against extensive commercial adoption of this approach. It is likely, however, that improvements in microfiltration membranes coupled with appropriate adjustment of processing conditions will result in increased commercialization of such processes in the near future.

The addition of calcium to aggregate lipoproteins in whey has also been proposed as an alternative to microfiltration for lipid removal. Whilst this process is technically effective, it may pose particular difficulties on scaling up to commercial operation.

A number of the whey fractionation processes previously outlined result in the preferential transfer of any lipid-containing portion of the whey into one particular fraction. For example, the lipid fraction in whey is preferentially found in the α fraction produced by the Australian process utilizing thermal aggregation. Such lipid material may be removed from whey fractions by, for example, microfiltration. This would result in a protein fraction with increased functionality, and a lipid fraction with excellent emulsification characteristics.

Summary of Whey Fractionation Methodologies

Of the various whey fractionation processes outlined, only the procedures based on adsorption/ion exchange involving the use of carboxymethylcellulose or Spherosil are in commercial operation, and these only in comparatively small-scale operations. The thermal procedures for whey fractionation are nearing commercialization. Of the remainder, the use of microfiltration to pretreat whey to remove lipid-containing material is likely to become further developed over the next few years. The remainder of the processes outlined seem unlikely to be of immediate commercial interest.

Functionality of Concentrates and Individual Whey Fractions

It is likely that whey protein fractions will have much greater and more reliable functionality than will WPCs, even WPCs containing 75% protein. Although the existing production of whey protein fractions is limited in the main to those from the Bi-Pro and Spherosil processes, and is of comparatively small tonnage, the ongoing development of the thermal aggregation processes will likely see the production of whey fractions increase sharply in the next few years. Already whey protein fractions attract much higher returns than WPCs and, with increased production, it is probable that increased applications will be identified, resulting in increased demand. Some products in which WPCs and whey fractions have been commercially utilized are listed in Table 2.

The individual proteins in whey can offer certain specific functional advantages. α-Lactalbumin is in demand for formulae which mimic human milk.

Table 2 Some uses of whey proteins and whey protein concentrates in foods

Baked custard	Coffee whitener	Meat analogs
Beverages	Cream desserts	Meat extenders
Acid-clear	Cream filling	Meat loaf
Acid-turbid	Cream icings	Meringues
Neutral	Cream sauces	Noodles
Biscuits	Cultured beverages	Pasta
Breads	Doughnuts	Potato flakes
Cake fillings	Egg-yolk replacer	Sauces
Cakes	Egg-white replacer	Puddings
Confectionery	Gravies	Sausage
Caramels	Hot dogs	Sherbet
Milk chocolate	Icecream	Snack foods
Canned beans	Imitation cheese	Tortillas
Cereals	Imitation milk	Whipped toppings
Chocolate drink	Macaroni	Yogurt

Data from Marshall KR and Harper WJ (1987) *Whey Protein Concentrates*. Bulletin No. 388B, p. 21. Brussels: International Dairy Federation.

However, currently the product is very expensive, and some of the technologies available for production have not yet come on stream. β-Lactoglobulin has been shown to possess valuable functional properties, and this product has been developed in Australia as the basis of a solid fat-substitute, with initial development as an ingredient in reduced- or low-fat luncheon meats and salami. Lactoferrin is also a significant component in whey, and various technologies have been described for its isolation. However current medical/nutritional claims for the product still have to be fully verified, and existing production capacity probably exceeds market demand. This situation could, however, change dramatically if the benefits of the product are verified in clinical trials. Extracts containing growth factor from whey are also currently being commercially developed. An Australian consortium of private, university, and medical organizations is planning to undertake clinical trials to assess the benefits of whey growth factor in wound healing and repair, including application to ulcers, burns, and incision wounds.

Whey fractions and WPCs offer many valuable functional properties as food ingredients. They can modify some or all of the organoleptic, visual, hydration, surfactant, structural, textural, and rheological properties of the food, resulting in improved consumer acceptance of the product. It is probable that the further development of specialized whey fractions with reliable and well-defined functional properties will see a marked increase in the application of these products over the next few years. While the technology for the production of such concentrates and fractions is now well known, the commercial success of this technology will turn to a large extent on developing means for the utilization of the lactose-rich byproduct streams emanating from such processes.

See also: **Whey and Whey Powders**: Production and Uses; Fermentation of Whey

Further Reading

du Boisbaudry P (ed.) (1981) Valorisation du lactoserum. *La Technique Laiterie* 3: 952.

Fox PF (ed.) (1989) *Developments in Dairy Chemistry. Functional Milk Proteins*, vol. 4. London: Elsevier Applied Science.

Hudson BJF (1982) *Developments in Food Proteins*, vol. 1. London: Applied Science Publishers.

Mulvihill DM and Fox PF (1983) *Assessment of the Functional Properties of Milk Protein Products*. Bulletin no. F-Doc 93. Brussels: International Dairy Federation.

Rich B (ed.) (1990) *Membrane Technology Today Seminar Proceedings, 20/21 February 1990*. Victoria: CSIRO Division of Food Processing, the Victorian College of Agriculture and Horticulture, and the Food Research Institute. Werribee, CSIRO Australia.

Sienkiewicz T and Riedel C-L (1990) *Whey and Whey Utilization*. Gelsenkirchen-Buer, Germany: Verlag. Th. Mann.

Zadow JG (1983) Whey utilization. *Food Research Quarterly* 1: 12.

Zadow JG (1986) Utilization of milk components whey. In: Robinson RK (ed.) *Modern Dairy Technology – Advances in Milk Processing*, vol. 1. London: Elsevier Applied Science.

Zadow JG (ed.) (1992) *Whey and Lactose Processing*. London: Elsevier Applied Science.

Fermentation of Whey

J Mawson, Massey University, Palmerston North, New Zealand

Process Options – Economic Factors

An extensive range of products can be obtained from whey fermentation. However, although processes for the large-scale utilization of whey were first investigated in the 1930s and 1940s, the commercial realization of these technologies has been slow. The most important factors that impact upon the development of economically feasible whey fermentation processes are:

1. Volume of whey available: 6–9 and 25–30 l of whey are produced per kilogram of cheese or casein manufactured, respectively, but the volume available at individual processing sites can be quite variable. In countries where production efficiencies have driven centralization of dairy

processing, in excess of 2×10^6 l day^{-1} of whey may be produced at a given site; in other countries, factory production volumes may be less than 10^5 l day^{-1}.

2. Polluting capacity: the polluting load of whey is very high. The BOD_5 (biological oxygen demand) value of whole whey is typically 40–50 kg m^{-3} (35 kg m^{-3} for deproteinated whey), or about 100 times the strength of domestic sewage. Combined with the large volumes that can be produced and increasingly stringent legislation limiting discharges to the environment, effective treatment of whey poses a significant challenge. Aerobic treatment is expensive and alternative technologies are favored.
3. Microbial utilization: lactose is the principal carbon and energy source available for microbial growth but is utilized by a relatively narrow range of organisms, and in some cases not by those that would normally be selected for the commercial production of a given fermentation product. Furthermore, the natural concentration of lactose in whey of 4–5 wt% is low in comparison with traditional fermentation substrates such as molasses or sugar syrups. This presents two challenges: it may limit the product concentration that can be achieved in some fermentations, leading to high product recovery costs, and also precludes transportation of whey beyond the production site to centralized processing facilities unless the whey is first concentrated by evaporation or reverse osmosis.
4. Milk production: in some countries, milk production is seasonal and whey fermentation plants will be idle for part of the year unless other substrates can be processed during these periods.
5. Markets for whey fermentation products: the market for whey fermentation products is influenced by several factors. Some products may not have a significant local market and export costs may be high. Where markets exist, of most importance are competition from similar products produced from other sugar sources (e.g., corn syrups) and the demand for alternative whey products (e.g., whey powders, lactose).

As a consequence of these various constraints, the emphasis in whey fermentation has been largely on the production of commodity chemicals or biomass as a means of reducing the effluent load of large volumes of whey. Such products command low prices in competitive markets; recovery costs dominate the economics of the process and high fermentation productivity must be achieved. The economic return is necessarily limited, although this is offset by the low cost of the substrate. In some cases, a negative cost may be assigned to the whey in recognition of the expense of alternative disposal methods.

Those fermentation processes that have proved commercially successful have developed largely to service particular niches in local markets. These include the production of potable and industrial ethanol, yeast biomass products, methane for use as a fuel, some organic acids and their derivatives, and various whey beverages. These processes are summarized in Figure 1 and will be the major focus of this article. Briefer consideration will be given to several other fermentation products that have been extensively researched and offer potential for commercial production.

Ethanol

Whey distilleries are in operation in Ireland, New Zealand, and the US. A typical industrial process flowsheet is shown in Figure 2. Deproteinated cheese or casein whey serum is used as substrate and may be either concentrated by reverse osmosis or supplemented with other lactose-rich streams to increase the lactose concentration to 10–13 wt%. The fermentation employs strains of the yeast *Kluyveromyces marxianus* var. *marxianus* (synonyms: *K. fragilis, Saccharomyces fragilis*). The process may be operated under aseptic conditions using pasteurized whey, although this is not necessarily required provided that good serum handling methods are practiced. Fermentation temperature is commonly in the range of 24–34 °C. The pH is not controlled and falls to 4.0 or less during the fermentation, which may take 12–36 h, depending on the initial lactose and yeast concentrations. At least one distillery operates with continuous fermentation at an overall dilution rate of 0.07 h^{-1}. Typically, the fermenter vessels are of 120–250 m^3 total volume and some mixing is required to prevent supersaturation of CO_2 in the vessel. This may be achieved mechanically or by recovering and reinjecting CO_2. Almost complete utilization of lactose can be achieved (and is a necessary requirement for effective waste abatement) and the yield of ethanol is in the range of 75–85% of the theoretical value of 0.538 kg ethanol per kg lactose metabolized. This corresponds to an ethanol concentration of about 2 wt% for fermentation of natural whey or 5–6 wt% for concentrated whey streams.

The ethanol is recovered by distillation using conventional techniques. The process shown in Figure 2 incorporates two columns (stripping and rectifying 'beer' or boiling columns) for initial concentration and partial purification, an extractive distillation

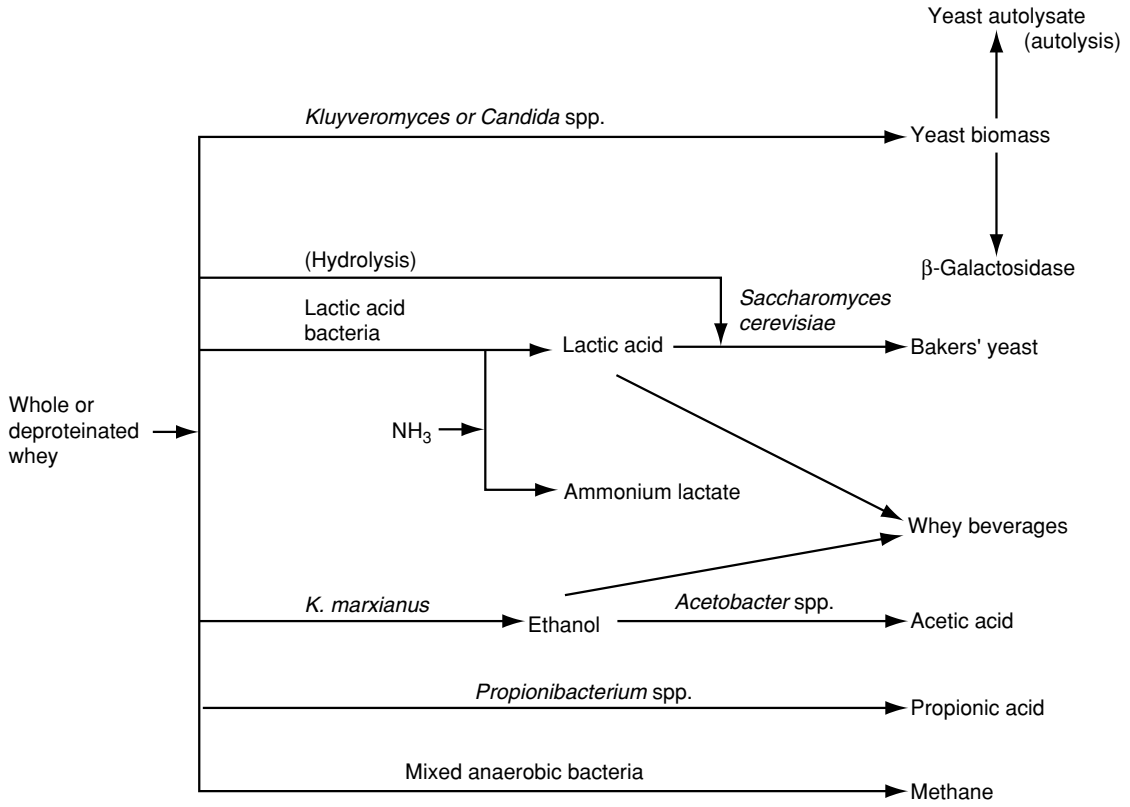

Figure 1 Major commercial products obtained by whey fermentation. Reproduced from Whey and Whey Powders: Fermentation of Whey, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

column using recycled hot water for removal of higher alcohols (fusel oils), and a rectifying column for final purification and concentration. For crude industrial or fuel alcohol, only the boiling columns and possibly a smaller side-draws purification column are required. For anhydrous ethanol production, additional azeotropic or extractive distillation columns, a zeolite molecular sieve, or a pervaporation membrane plant must be used.

Yeast Biomass

Single-cell Protein

Yeast biomass has been produced commercially by whey fermentation from the 1940s until the present. Strains of *K. marxianus* var. *lactis* or var. *marxianus* are those most commonly used, although other strains may be produced successfully. In some operations, ethanol is also obtained as a co-product by lowering the aeration rate during the fermentation.

The requirements of a suitable strain for biomass production are summarized as follows.

1. High specific growth rate and biomass yield to insure a high productivity
2. The strain must not be affected by whey proteins if these are present
3. The strain must be suited to continuous culture
4. The strain must be acid-resistant. (To control contamination, it is necessary to operate the process at low pH or to wash the yeast at frequent intervals with acid to remove contaminants)
5. Large cell size and uniform morphology to aid cell separation and concentration
6. Adequate protein content and acceptability in feeding trials

The Bel process (Figure 3), developed in France in the late 1950s, is a frequently cited example of this fermentation. Sweet cheese whey is first deproteinated and diluted to a lactose concentration of 20–25 kg m^{-3}. Whey is limiting in nitrogen sources for yeast growth, so ammonium salts are added to maintain a high nitrogen content and growth rate, and trace metals (Fe, Cu, Mn, Zn) may also be added. The culture has been shown to comprise three cooperating yeast species. The continuous fermentation is

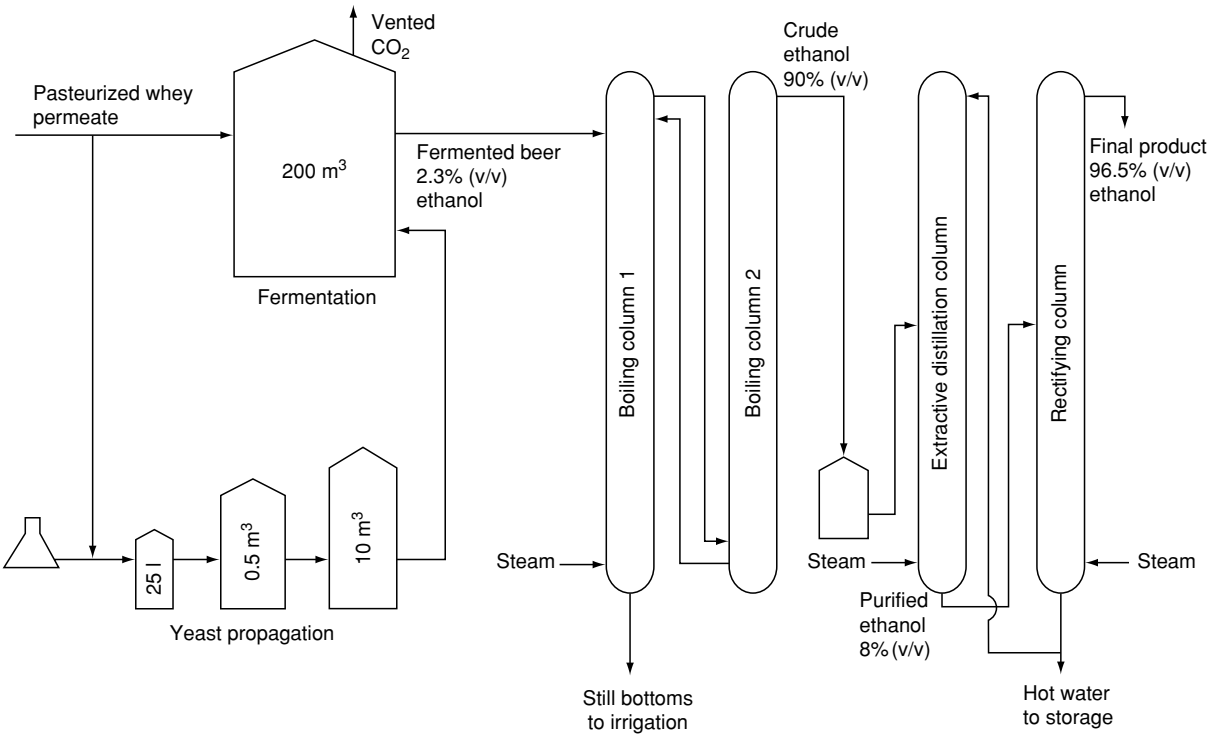

Figure 2 Typical process for production of ethanol from natural-strength whey permeate. Reproduced from Whey and Whey Powders: Fermentation of Whey, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

operated at a dilution rate of $0.33\,h^{-1}$ and a temperature of 38 °C. For a serum flow rate of 5–6 $m^3\,h^{-1}$, air is added at about $1800\,m^3\,h^{-1}$ and provides both the mixing and the aeration of the culture in an air-lift fermenter. The residual sugar level is about $1\,kg\,m^{-3}$, the biomass yield is 0.55–0.6 kg yeast (dry basis) per kg lactose metabolized, and the biomass productivity is approximately $4.5\,kg\,m^{-3}\,h^{-1}$. The yeast is separated and concentrated in a two-stage washing and centrifugation process, plasmolyzed to render the yeast protein more accessible, and dried.

The composition and uses of biomass produced from whey are similar to those of other food yeasts. The crude protein content is about 50 wt% on a dry basis and the only significant limitation is of the sulfur-containing amino acids. Overall, the protein value is similar to that of casein. Fermenting whole whey and recovering a product containing both the yeast biomass and the whey proteins may increase the food value of the yeast.

Bakers' Yeast

Two processes have been developed for the production of bakers' yeast to overcome the limitation of *Saccharomyces cerevisiae* not being able to utilize lactose. In the first, the lactose is hydrolyzed using β-galactosidase, and the glucose and galactose are consumed simultaneously by the yeast in fed-batch or continuous culture. The second process utilizes a two-stage fermentation system. In the initial stage, lactic acid bacteria convert lactose to lactate and this is consumed in the subsequent stage by the yeast. Although the bakers' yeast so produced appears comparable in quality to that from conventional processes, only very limited quantities are manufactured by these methods.

Products from Yeast Biomass

Yeast cells can be a useful source of several food additives or processing aids. Following cell growth and recovery, the biomass can be broken down mechanically or biologically to release cellular components which may be purified or transformed to yield high-value products. The major products that can be derived from biomass are yeast autolysate, flavour nucleotides, enzymes, and cell wall components such as glucan and mannan.

Autolysate and Flavor Nucleotides

Autolysis is a process of self-digestion mediated by degradative enzymes present within the cell. Elevated

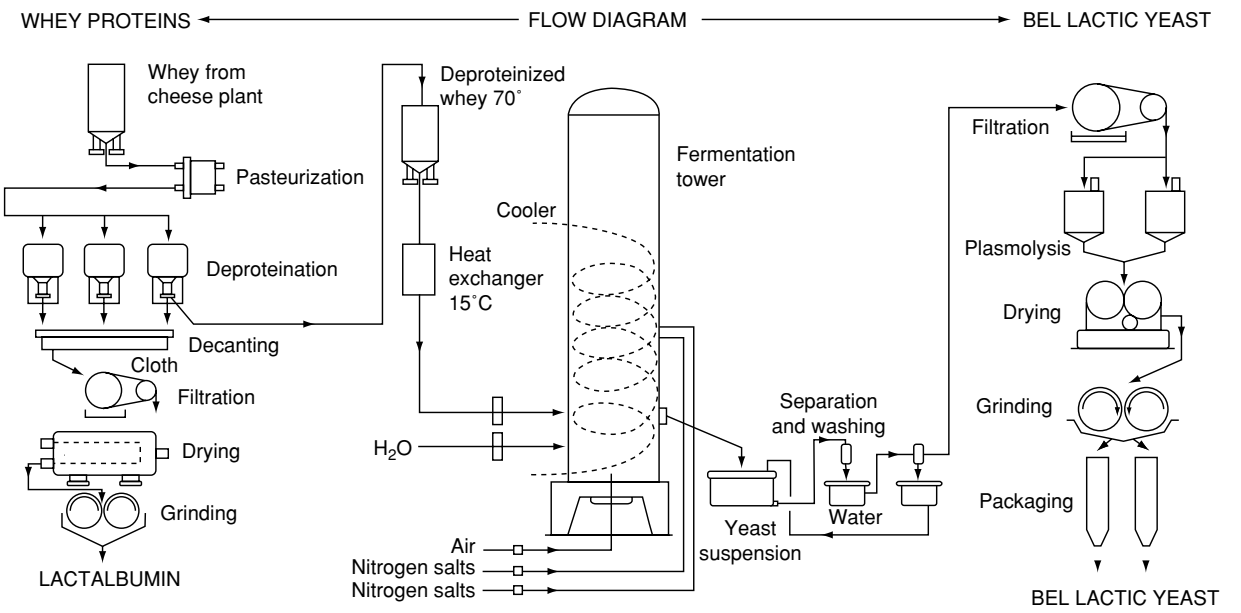

Figure 3 Fromageries Bel process for yeast production from deproteinated whey. Reproduced from Whey and Whey Powders: Fermentation of Whey, *Encyclopaedia of Food Science, Food Technology and Nutrition*, Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

temperatures and the presence of NaCl enhance the rate of lysis during which protein is solubilized and cell wall polymers are broken down. Optimal conditions for autolysis of *Klyveromyces* strains are 50–55 °C, pH 5.1–6.5, and 8–12 h; addition of salt may not be necessary. Yeast nutrients are more readily available in autolysates than in intact cells and the autolysates find uses in foods, animal feeds, and microbial cultures, e.g., for enhancing the growth of dairy starter cultures.

RNA can be isolated from lysed cells and used to produce 5′-ribonucleotides which are important flavor enhancers in food systems.

β-Galactosidase

The enzyme β-galactosidase (EC 3.2.1.23), or lactase, can be produced from selected strains of *Kluyveromyces* spp. following growth on diluted whey. The enzyme is intracellular and is induced by lactose and galactose. One study showed that expressed levels may vary up to 60-fold between strains, but commercial yield data are not available. The optimum conditions for enzyme activity are pH 6–7 and 35–45 °C, and Mn^{2+} and Mg^{2+} are strong activators. β-Galactosidase may be used to hydrolyze lactose to overcome the problem of lactose intolerance and to generate syrups for food use which are sweeter than lactose and which do not crystallize as readily (hence reducing 'sandy' defects). Oligosaccharides, which have useful nutraceutical properties, are alternative and interesting byproducts of the hydrolysis process.

Organic Acid Fermentations

Lactic Acid and Ammonium Lactate

Whey has been a traditional feedstock for lactic acid production by fermentation, but the process struggles to be economic compared to the utilization of corn syrups. Lactic fermentations have typically been conducted in batch mode using homofermentative strains of *Lactobacillus* such as *L. delbrueckii* sbp. *bulgaricus*. Complex nutritional sources such as corn steep liquor, malt sprouts, and malt or yeast extracts may be supplemented. The pH is controlled in the range 5.5–6.5 by addition of $Ca(OH)_2$ or $CaCO_3$ and the optimum temperature is about 43 °C. The medium is usually pasteurized, but contamination is not usually a serious problem. For natural whey the fermentation will be complete in less than 24 h with a yield of 90–95%. The productivity of the fermentation is typically 1–3 kg m^3 h^{-1}.

Recovery and purification of lactate represents a major cost in the process and are complicated by the highly corrosive nature of the product. This requires all equipment to be 316 stainless steel or better and removal of metal ions introduced by corrosion may be required for some grades of product. In traditional processes, the fermentation liquor is initially heated

to 80–100 °C and the pH is raised above 10 to solubilize calcium lactate, kill all microorganisms, and precipitate calcium phosphate and proteins. The liquor is then decanted, filtered, and treated with carbon to increase the purity. Following evaporation under vacuum, a technical-grade lactic acid can be recovered by acidification with sulfuric acid, filtration of the precipitated $CaSO_4$, and further carbon and purification treatments. To obtain higher-grade products, calcium lactate crystals may be recovered following cooling of the liquor; these are then washed, redissolved, and carbon-treated in several cycles. Several alternative recovery processes are available, e.g., both solvent extraction and distillation of lactate esters are practiced commercially.

Lactic acid has attracted much recent interest as a base for the manufacture of biodegradable polylactate plastics. A large-scale pilot plant was commissioned in Wisconsin, US, but full-scale production did not proceed.

Ammonium lactate is produced by the addition of liquid ammonia to the fermentation broth. After completion of the fermentation, the broth is evaporated, neutralized by addition of further ammonia, and processed to a range of liquid or solid products. These are used as animal feeds and are especially suitable for ruminants. Ammonium lactate is superior to urea and similar to soybean meal in its nutritive value and digestibility.

Acetic Acid

Following the ethanol fermentation of whey, the alcohol can be further metabolized to acetic acid by *Acetobacter* spp. This process has been commercialized, most notably by Kraft, and the resulting whey vinegar may be used in salad dressings and other foods. The manufacturing process is assumed to be similar to that for conventional vinegar production. The production of acetic acid from whey as a base for synthesis of other chemicals has also been considered. Recent attention has focussed on production of acetate deicers as a possible market.

Propionic Acid

Propionic acid is added as a fungistatic agent to bread and bakery products and can be produced by whey fermentation using strains of *Propionibacterium freudenreichii* ssp. *shermanii*. Whole whey is typically used and the fermentation must proceed under sterile conditions at around 30 °C and pH 6.5–7.5. Nutritional supplements, such as yeast extract, considerably enhance propionate production and a typical yield of about 40% of the lactose fermented is achieved after 60–70 h. The culture fluid is spray-dried to obtain a powder containing both the acid and whey proteins.

Methane Fermentation

The microbial production of methane from whey offers the major benefits of an efficient waste treatment process coupled with the production of a z-convenient energy source that can be utilized on-site. The methane fermentation, or anaerobic digestion, requires a limited input of energy and nutrients and produces very little sludge for ultimate disposal in comparison to conventional aerobic treatment processes. A number of bioreactor (digester) configurations are available for the treatment of high-strength effluents such as whey. All of these rely on retaining the microorganisms mediating the fermentation in the digester, so that the hydraulic residence time of the waste stream can be very much less than the biological solids (or biomass) residence time. This separation is necessary as the cells grow only slowly, yet large volumes of waste must be processed daily.

Several large-scale whey or whey permeate digesters are in operation and many laboratory or pilot-scale trials have been reported for a variety of retained-cell systems. Mesophilic operation at 30–37 °C is favored. Overall COD (chemical oxygen demand) reductions of about 90% are possible for loading rates of up to $20\,kg\,m^{-3}\ day^{-1}$, although loading rates in large-scale operations are typically lower, at $1.5–10\,kg\,m^{-3}\ day^{-1}$. Gas production is about $35\,l^{-1}$ of whey processed and has a methane content of 55–65%. These figures are consistent with the theoretical yield of 350 litres CH_4 at standard temperature and pressure per kg COD removed.

Success has also been achieved with treatment of slops from whey distilleries by anaerobic digestion. The reduced strength of the slops from ethanol fermentation can promote more stable operation of the digester, although the total methane yield is less. However, regardless of the initial substrate, some aerobic polishing of the effluent is required before digester effluent can be discharged to natural waterways.

Fermented Whey Beverages

Whole or deproteinated whey can be fermented to produce a range of beverage products. The major advantages offered by whey as a substrate for beverage production are that whey drinks have a greater nutritive value and are more thirst-quenching than most soft drinks, and are less acidic than fruit juices. The marketing of these products generally emphasizes the health and nutritional benefits of the

Table 1 Production of a fermented-fruit whey drink from whey

Recipe	
Stabilizer	0·30%
Fruit concentrate	2–7%
Sour whey	85·7%
Sugar	6–10%
Flavor and color	As required, to 100%
Possible technical pH range	3·6–4·2
Method	
1. Coarse filtration of sour whey	
2. Disperse stabilizer in the sour whey, avoiding lumps	
3. Maintain gentle agitation for 30 min to allow the stabilizer to swell	
4. Heat to 85 °C and cool to incubation temperature (*c.* 42 °C)	
5. Inoculate with *Lactobacillus helveticus*	
6. Incubate to pH 4·0	
7. Cool	
8. Add fruit and sugar	
9. Check pH and adjust, if necessary, with 50% citric acid solution to pH 3·6–4·2	
10. Thermal treatment	
11. Place in chill	

beverage, especially if they still contain the whey proteins. A variety of whey drinks are available in many countries, although they are most popular in Europe.

Lactic or alcoholic fermentations can provide desirable properties in the beverage. Lactic fermentations use conventional starter organisms or probiotic strains and *Kluyveromyces* yeast strains are commonly used for the alcoholic fermentation. A combined lactic–ethanol fermentation may be conducted with a kefir-type culture, but whey health drinks usually have a low (< 1%) or negligible alcohol content. A typical recipe for preparation of a fermented-fruit whey beverage is outlined in Table 1. Whey 'wines' can be manufactured by fermentation of deproteinated wheys and can serve as a base for wine coolers.

Other Fermentation Products

Many alternative fermentation processes have been investigated for whey utilization. Possible products include mixed solvents (acetone, butanol, and ethanol), yeast oils and specialized lipids, glycerol, extracellular polysaccharides, other organic acids (citric, acrylic), vitamins, poly(β-hydroxybutyrate), amino acids, flavoring agents, pigments, and lantibiotics such as nisin. Of these, polysaccharides (xanthan gum) and flavoring agents (based on diacetyl) have been commercially produced in limited quantities, and the production of lantibiotics using lactic acid bacteria also appears feasible.

See also: **Casein and Caseinates**: Methods of Manufacture; **Cheeses**: Chemistry of Gel Formation; **Effluents from Food Processing**: On-Site Processing of Waste; Composition and Analysis; **Enzymes**: Uses in Food Processing; **Fermented Milks**: Other Relevant Products; **Lactic Acid Bacteria**; **Lactose**; **Vinegar**

Further Reading

Belem MAF and Lee BH (1998) Production of bioingredients from *Kluyveromyces marxianus* grown on whey: an alternative. *Critical Reviews in Food Science and Nutrition* 38: 565–598.

Driessen FM and van den Berg MG (1990) New developments in whey drinks. *Bulletin of the International Dairy Federation* 250: 11–19.

Kilara A and Patel MT (1992) Whey and lactose fermentation. In: Zadow JG (ed.) *Whey and Lactose Processing*, pp. 409–448. London: Elsevier Applied Science.

Litchfield JH (1996) Microbiological production of lactic acid. *Advances in Applied Microbiology* 42: 45–95.

Mawson AJ (1994) Bioconversions for whey utilisation and waste abatement. *Bioresource Technology* 47: 195–203.

Meyrath J and Bayer K (1979) Biomass from whey. In: Rose AH (ed.) *Microbial Biomass. Economic Microbiology*, vol. 4, pp. 207–269. London: Academic Press.

Sienkiewicz T and Riedel C-L (1990) *Whey and Whey Utilization*, 2nd edn. Gelsenkirchen-Buer: Verlag Th. Mann.

Yang ST and Silva EM (1995) Novel products and new technologies for use of a familiar carbohydrate, milk lactose. *Journal of Dairy Science* 78: 2541–2562.

Principles of Dialysis

J J Roberts, University of Pretoria, Pretoria, South Africa
J F Mostert, ARC-Animal Nutrition and Animal Products Institute, Irene, South Africa

Background

Dialysis is the oldest, most natural, and flexible technique for separating components from a solution. Electrodialysis (ED), however, has been known and studied since the beginning of the twentieth century, but it was not until around the end of the 1940s that stable membranes of low electrical resistance were developed, allowing the commercial application of ED using membrane stacks with several membranes in series. The modern process of ED has been made possible by the development of highly selective membranes for anions and cations. Both processes may be used in conjunction with other

membrane-separation processes (osmosis, reverse osmosis, ultrafiltration, and microfiltration) in experimental and large-scale industrial applications. A major disadvantage is that an equilibrium is quickly reached. However, specific molecules may be completely separated, provided a high concentration gradient is maintained, and the permeate is constantly removed. (*See* **Membrane Techniques**: Principles of Reverse Osmosis; Applications of Reverse Osmosis; Principles of Ultrafiltration.)

Theory of Dialysis and Electrodialysis

Dialysis

Dialysis is a separation process whereby smaller molecules are segregated from larger molecules (macromolecules) in a solution by virtue of their different rates of diffusion through a specific membrane (or by nondiffusion, which is the criterion of the colloidal state and forms the basis of separation of colloidal materials from crystalloids). Alternatively, dialysis may be regarded as a membrane transport process in which solute molecules are exchanged between two liquids separated by a membrane. Hence, the primary driving force is the difference in concentration of the permeable species between the solution in the dialysis membrane system and that on the outside. The membrane might be natural, such as a pig bladder, or artificial (i.e., consisting of materials such as cellulose derivatives or collodion).

To illustrate the principles of dialysis, a test solution can be placed in a sack or bag that may consist of a membrane with, for example, a 24-nm pore diameter. The bag is closed/knotted and soaked in distilled water at a given temperature and for a given time; the bags are usually agitated. The small molecules diffuse out into the water, while the larger molecules remain in the solution in the bag. This method is commonly used to separate proteins from salt solutions. In addition, it is an indispensable technique in the recovery and purification of materials in food, chemical, biological, and pharmaceutical preparations.

Electrodialysis

ED is a unit operation that uses semipermeable membranes for the separation or concentration of electrically charged particles (ions) from nonionic particles or species in a solution. The key to the process is the use of ion-selective membranes. These membranes are ion-exchange resins cast in sheet form, which allow the passage of positively charged cations (e.g., sodium or potassium) or anions (e.g., chloride or phosphate). To achieve separation by means of ED, cation and anion membranes are alternated with plastic spacers in a stack configuration with the anode at the one end and the cathode at the other. Spacers are usually made of low-density polyethylene and arranged in the membrane stack so that all the mineralized streams are manifolded together, and all the concentrated streams are manifolded together. The spaces between the membranes represent the flow paths of the demineralized and concentrate streams. Hence, a repeating section, called a cell pair, consists of a cation transfer membrane, demineralized water flow spacer, anion transfer membrane, and concentrate water flow spacer. A typical membrane stack may have from 300 to 500 cell pairs.

A direct electric current (d.c.) (or an electrostatic potential or gradient) applied across the electrodes creates a driving force. This induces anions to migrate in the direction of the anode (+) and cations to move in the direction of the cathode (−). The ion-selective membranes form barriers to ions of opposite charge. The net result is that the anions attempting to migrate to the anode pass through the anion barrier or membrane but are stopped by the cation membrane. Likewise, cations trying to migrate to the cathode pass through the cation membrane but are stopped by the anion membrane. The overall effect is that the barriers form alternate compartments of ion-diluting cells and ion-concentrating cells. Therefore, ED relies primarily on voltage or electromotive force and the use of ion-selective membranes to effect separation between charged particles. Hence, on circulation of a specific liquid product or test solution through the diluting cells and a brine solution through the concentrating cells, free mineral ions leave the test solution and collect in the brine stream. The level of demineralization achieved depends upon the initial ash content, current density, and the duration of time the test solution is within the membrane cells.

Figure 1 shows a schematic diagram of a typical ED process for the demineralization of whey in an ED membrane stack. One stream enters the membrane stack and flows in parallel only through the demineralizing compartments, while a brine stream enters the membrane stack and flows in parallel only through the concentrating compartments. Note that the solutions flow across, not through, the membranes. When a d.c. voltage is applied across the electrodes, the electrical potential gradient created causes anions to move in the direction of the anode and cations to move in the direction of the cathode. The ion-selective membranes form barriers to ions of opposite charge. The result is that anions attempting to migrate to the anode pass through the anion membrane but are stopped by the cation membrane; cations trying to migrate to the cathode pass through the cation membrane but are stopped by the

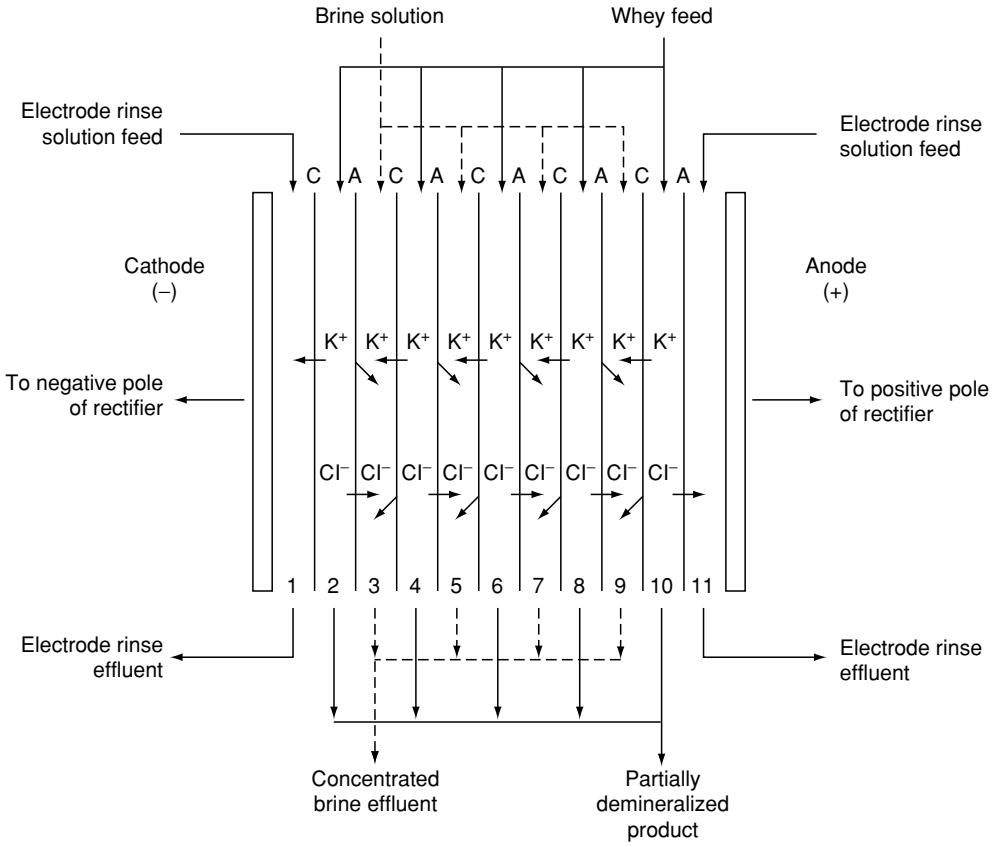

Figure 1 Schematic diagram of the electrodialysis process showing the demineralization of whey. From Batchelder BT (1987) Electrodialysis applications in whey processing. In: *Trends in Whey Processing, Bulletin of the International Dairy Federation No. 212*, pp. 84–90. Brussels: International Dairy Federation, with permission.

anion membrane. Hence, the membranes form alternate compartments of ion-diluting cells and ion-concentrating cells. By circulating whey through the diluting cells and a brine solution through the concentrating cells, free mineral ions leave the whey and collect in the brine stream. The level of demineralization achieved depends on the initial ash content, current density, and duration of time the whey solution is within the membrane cells.

For a better comprehension of the process of ED dialysis (especially the efficiency of the mass-transfer reaction), the functioning of an electrolytic cell will be discussed briefly. A pair of electrodes as well as an electrolytic solution (i.e., a solution containing ions) are placed in such a cell. When a voltage is applied to the electrodes, a migration of cations and anions takes place. According to Faraday's law, the number of mass equivalents transported by this mechanism is directly proportional to the amount of electricity which passes through the cell. In mathematical terms, this can be expressed as follows:

$$\text{Current} \times \text{time} = F \times E\,(\text{equivalent mass transported}), \quad (1)$$

where E = number of moles × valency, and Faraday's constant $F = 96\,490$ A s per equivalent or 26.8 A h per equivalent.

Faraday's law states that if, for example, a current of 26.8 A is applied in an electrolytic cell for 1 h, it will transport one equivalent of matter from the solution to the electrodes. Hence,

$$i = FE(J_j Z_j) \quad (2)$$

where i = current density (A m^{-2}), J_j = molar flux of ion j (mol s^{-1} m^{-2}), and Z_j = valency of ion j.

Eqn (2), therefore, states that the total current in an electrolytic cell is equivalent to the sum of the current conveyed by each species in the cell.

Similarly, the amount of electricity necessary to produce separation by means of electrodialysis can be calculated by means of eqn (1). First, we must consider the cell efficiency coefficient (N_c), which accounts for the nonideality of the cell stack. This is due mainly to the presence of the selective membranes, but also to the loss of current through pipes and manifolds as well as the diffusion of electrolyte through the membranes. Thus, for a stack with n cell pairs:

$$n \times \text{current} \times \text{time} = FE\,(\text{equivalents removed})/N_c, \quad (3)$$

N_c can be expressed as the product of three efficiencies.

$$N_c = N_m \times N_l \times N_w \quad (4)$$

N_m accounts for the imperfection of the permselective membrane and is also called the Coulomb efficiency; N_l corresponds to the inefficiency due to current leakage through pipes and manifolds; N_w accounts for water transport, which always follows the migration of ions. The latter is also referred to as electroosmosis. For commercial equipment N_c is about 50%, and for laboratory equipment 90%. By manipulating eqn (3):

$$N_c = F(Q_{in}C_{in} - Q_{out}C_{out})/(In). \quad (5)$$

where Q_{in} = input flow ($m^3\,s^{-1}$); Q_{out} = output flow ($m^3\,s^{-1}$); C_{in} = input concentration (equivalents m^{-3}); C_{out} = output concentration (equivalents m^{-3}).

Owing to electroosmosis, Q_{in} and Q_{out} are generally different.

The energy consumed by an electrodialysis system may be computed from

$$E = RI^2t. \quad (6)$$

where E = electric energy (W s = J), R = total resistance of the system (ohm), I = electric current (A), and t = time (s). From eqn (3), the number of equivalents removed = ItN_c n/F. Hence from eqn (6) the energy per unit of equivalent removed is

$$e = RIF/N_c(\text{W s per equivalent}).$$

Types of Electrodialysis Membranes and Fouling Problems

The following types of membranes are used in ED:

- anion-permeable membranes (—NH_3^+) (A), which contain fixed cationic groups and repel cations;
- cation-permeable membranes (C), which contain fixed anionic groups (—SO_3^-) and repel anions; and
- nonselective membranes (N), which are permeable to both anions and cations.

Commercially available ED membranes are manufactured as heterogeneous and homogeneous membranes. The first type is made by mixing ion-exchange resins with a solution of polymeric materials, which is used as a thin coating material on a fine cloth or mesh as a mechanical reinforcement. Anionic or cationic membranes are obtained by using the corresponding ion-exchange resins. In homogenous membranes, the active cationic or anionic groups are induced or created in the polymeric material. Nonselective membranes are usually manufactured from cellulose. Pore sizes of the membranes are in the range 1–2 nm.

The desired characteristics of a good-quality membrane for electrodialysis include:

- good electrical conductivity;
- good mechanical strength;
- high ionic permselectivity;
- chemical stability;
- resistance to fouling by organic molecules;
- insolubility in aqueous solutions;
- resistance to chemical change in pH from 1 to 10;
- long life expectancy;
- impermeability to water under pressure;
- operation at temperatures in excess of 46 °C.

A major problem with ED is that of membrane polarization, also known as 'limiting current density.' Limiting current density is usually expressed as $(CD/N_d)_{lim}$, where CD = current density (the amount of current carried by a unit area of membrane surface), and N_d = normality of the demineralized outlet stream. This limit is a function of the fluid velocity in the flow path, the stream temperature, and the types of ions present. This condition may arise when too large a driving potential is applied to a system. The latter results in a rapid depletion of both anions and cations from the region immediately in contact with the membranes of the feed cell. Hence, ions are removed from the boundary layers through the membranes more rapidly than they can be replaced by diffusion from the bulk feed fluid, so that nothing remains to conduct the current. This polarization potential is a combination of applied voltage, diffusion characteristics of the ions, flow characteristics of the feed, and the ionic concentration of the feed. Note that polarization does not usually become significant at both membranes at the same time. Other typical problems with ED systems include current leakage, back diffusion, and membrane poisoning and fouling. Membrane poisoning is the result of ions passing into the membrane, attaching irreversibly, thereby obstructing the passage of following ions. In contrast, fouling is a surface phenomenon and is the result of the deposition on the surface of a fouling layer that forms a barrier to diffusion. These effects are largely eliminated in an electrodialysis reversal system.

Electrodialysis Reversal

The electrodialysis reversal (EDR) system is the first commercially available membrane process that is symmetrically reversible. This means that the direction of ion movement through the membranes and the

identity of the concentrating and demineralizing compartments are reversed at periodic intervals (e.g., every 20 min). To bring this about, the d.c. polarity is reversed. The mechanism of EDR differs from that of unidirectional ED in a very simple but profound way, and the advantages of EDR system are many. The process is more tolerant than other commercially available membrane processes of a wide range of troublesome organic, inorganic, colloidal, and biological contaminants. The single most important advantage of EDR is that any scale or film build-up, common in unidirectional processes such as electrodialysis and reverse osmosis, is limited. There is a tendency for constituents that are capable of forming slimes and precipitates to be deposited at the membrane/solution interface, and over a period of time, these deposits affect the membrane desalting/filtration performance. The self-cleaning characteristic of EDR systems is analogous to that of reversing heat exchangers used in the cryogenic industry for removal of carbon dioxide and water scale from low-temperature surfaces.

Transport Depletion

The fouling of anion exchange membranes by negatively charged organic collodial matter is a serious problem in conventional ED systems. This problem can be avoided by replacing anion membranes with nonselective (neutral) membranes, with the order of the membranes being –C–N–C–N–C–, so that concentration polarization is avoided.

Ion Substitution

Ion substitution is another variant of ED in which one species of ion is removed and substituted by another. Instead of the simple, alternating –anionic–cationic–anionic–(–A–C–A–)– membrane arrangement, the sequence –A–C–C–A– solution to be demineralized is fed into the C–C compartment.

Ion Replacement

Another modification of classical electrodialysis is ion replacement, where the membranes used are anionic or cationic only. In this application, the solutions are neither enriched nor depleted in ions; only an exchange in specific types of ions is accomplished.

Advantages of Electrodialysis

The many advantages of electrodialysis, e.g., in protein separations, include fast and controlled removal of salts, no product dilution, low membrane area requirements, negligible product adsorption, easy salt recovery in the same unit, and the use of lower concentrations of salting-out agents than are required by conventional direct addition processes.

See also: **Membrane Techniques**: Principles of Reverse Osmosis; Applications of Reverse Osmosis; Principles of Ultrafiltration

Further Reading

Batchelder BT (1987) Electrodialysis applications in whey processing. In: *Trends in Whey Processing, Bulletin of the International Dairy Federation No. 212*, pp. 84–90. Brussels: International Dairy Federation.

Brocklebank MP (1987) Large scale separation and isolation of proteins. In: King RD and Cheetham PSJ (eds) *Food Biotechnology – 1*, pp. 139–192. New York: Elsevier Applied Science.

Glover FA (1986) Modifications to the composition of milk. In: Robinson RK (ed.) *Modern Dairy Technology*, vol. 1, pp. 235–271. New York: Elsevier Applied Science.

López-Leiva MH (1988) The use of electrodialysis in food processing. Part 1: Some theoretical concepts. *Lebensmittel Wissenschaft und Technologie* 21: 119–125.

Applications of Dialysis

J F Mostert and L Frylinck, ARC-Animal Nutrition and Animal Products Institute, Irene, South Africa

Introduction

The concentration, removal, or separation of components in aqueous mixtures by membrane techniques, e.g., dialysis and electrodialysis, have many applications in the processing of liquid foods and areas closely related to food.

Dialysis is a technique employing the difference in concentration as a driving force to separate large particles from small ones in a solution, for example, proteins from salts. The solution to be treated is placed on one side of a membrane, and a solvent (water) on the other side. The membrane has pores of a diameter that allows the small salt molecules to pass through, but is too small for the protein molecules to pass (**Figure 1**). A disadvantage of dialysis is that the rate of diffusion varies with the difference in concentration. Dialysis can, however, be speeded up if the solvent on the other side of the membrane is changed frequently.

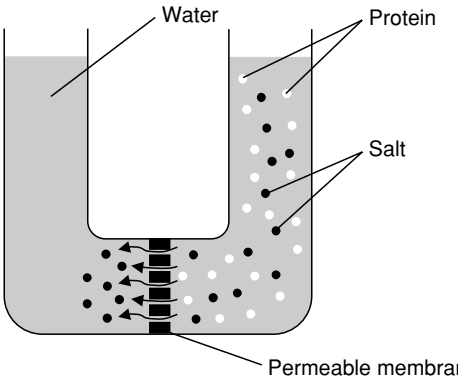

Figure 1 The separation of protein molecules from a salt solution, using a dialysis membrane. Reproduced from Bylund G (1995) The chemistry of milk. In: *Dairy Processing Handbook*, pp. 17. Lund, Sweden; Tetra Pak (Processing System Division), with permission.

Electrodialysis (ED), on the other hand, is defined as the transport of ions through nonselective semipermeable membranes under the driving force of a direct current and an applied potential. The membranes used have both anion and cation exchange functions, thus making the electrodialysis process capable of reducing the mineral content of a process liquid, e.g., whey or sea water. The most important use of ED is in the desalination of brackish water for the production of potable water as well as in the production of table salt from sea water. ED is a classic whey demineralization process with a well-proven technology and with a multitude of systems operating worldwide. Other important applications are found in the dairy, wine, and beverage industries. The integration of dialysis and ED in food-processing systems, together with new advances in fermentation systems, has a profound effect on new biotechnological applications for the manufacture of specialized and highly differentiated food products with enhanced functional properties.

Dialysis in the Dairy Industry

In the dairy industry ED is primarily used for the demineralization (desalting) of whey. Large volumes of whey are produced as a byproduct from cheesemaking and, to a lesser extent, from casein manufacture. For each tonne of cheese produced, about 8 tonnes of whey are left over, resulting in huge volumes. In the past whey has been returned to farmers for animal feeding, irrigation of lands and pastures, or discharged directly into rivers, waterways, or sewage systems. Considerable quantities of whey are still being discharged into the environment worldwide.

Whey is the liquid residue of cheese and casein production and is, to a certain extent, one of the biggest reservoirs of high-quality food protein not fully utilized in human consumption levels. The composition of whey will vary substantially depending on several factors, including the process involved in the production of whey. On average, whey contains about 65 g kg^{-1} of solids, comprising about 50 g lactose, 6 g protein, 6 g ash, 2 g nonprotein nitrogen, and 0.5 g fat. The whey protein fraction of bovine milk contains four main proteins: β-lactoglobulin (β-Lg, 50%), α-lactalbumin (α-La, 20%), blood serum albumin (10%), and immunoglobulins (lg, 10%; mainly lgG_1, with lesser amounts of lgG_2, lgA, and IgM). Wheys are broadly classified as 'sweet' wheys produced from rennet-coagulated cheeses and acid wheys, which have a similar composition but with a higher acidity and ash content.

The unique technological, nutritional, physiological, and even nutraceutical properties of whey proteins, together with the high disposal cost of dairy effluent, stimulated research in whey utilization within the food industry. Industrial whey processing began with fluid whey being concentrated and dried. New membrane techniques and processes revolutionized whey processing by providing breakthrough methods for the separation of dissolved or suspended solids, for the removal of water and, especially, for the combination of these. Nowadays demineralized whey powders, whey protein concentrates, whey protein isolates, individual proteins and derivates, as well as permeate products/lactose derivates are manufactured and are available on the market.

It is not easy, from an economic point of view, to process whey into products of higher commercial value. As whey has a fairly high salt content – about 8–12% calculated on dry weight – its usefulness as an ingredient in human foods is limited. By having the whey demineralized, various fields of application can, however, be found for whey, which is partially (25–30%) or highly (90–95%) demineralized. Partially demineralized whey concentrate can, for example, be used in the manufacture of icecream and bakery products or even in quarg, whereas highly demineralized whey concentrate or powder can be utilized in formulas for infants and also in a wide range of other products.

The two main methods used commercially for high-degree demineralization of whey, namely ion exchange and ED, result in products of somewhat different composition. The ion exchange process is relatively nonselective and removes both monovalent and polyvalent ions, whereas ED is more dependent on ionic mobility and tends preferentially to remove monovalent ions. Economically, only about 90% demineralization is possible with ED; 50–75% is regarded as being more viable from a practical and

economical point of view. The value of 90% is considered the practical limit of salt removal due to decreasing electrical conductivity with demineralization. ED is likely to have advantages for plants with high hourly utilization rates, for the manufacture of products that do not require high levels of demineralization, and where low-cost electricity is available. Based strictly on economics, the recent trend is towards combined ED ion exchange plants. The majority of ED plants use preconcentrated sweet whey as well as acid whey, skimmed milk, delactosed whey, milk, whey protein concentrate, and whey ultrafiltration (UF) permeates.

Whey produced by ED has a reduced mineral content and is rich in lactose and whey proteins, and contains appropriate low levels of essential minerals, which makes it an ideal ingredient for infant formulas. Raw whey is concentrated to 23–27% solids, and clarified before ED. Liquid or dried reduced-minerals whey could be used for blending infant formulas, but a high level of demineralization (90%) is required for sweet whey to be used in an infant formula. The use of reduced-minerals whey as an ingredient in infant formulas must be ranked as one of the major successes in terms of the valuable and profitable use of a byproduct. Reduced-minerals whey could be used in animal feeds, protein and citrus drinks, dry mixes, confectionery coatings, icecream, and bakery goods.

Although the trend has been towards the use of whey protein-based infant formulas, there is also a considerable percentage of infant formulas produced in which nonfat milk is used as the sole source or protein. These products tend to have higher ash contents, including higher levels of phosphorus, calcium, sodium, potassium, and chloride, than whey protein-based formulas. Other ED applications include the partial demineralization of whey protein concentrate, which is ideal for products that require a higher protein-to-lactose ratio; acid whey for mixing with sweet whey in the manufacture of whole whey powder; and of cottage cheese whey in order to reduce disposal problems.

ED is also used in various integrated membrane systems for specific purposes, e.g.:

- Nanofiltration/ED systems: nanofiltration is used to concentrate and partially demineralize whey prior to ED, which in turn completes demineralization to the desired level.
- UF/ED systems: UF allows molecules, such as sugars and salts, to pass through the membrane while protein molecules are retained. UF is used to produce whey protein concentrate (WPC) while ED demineralizes the WPC to produce a low-ash, high-protein ingredient for infant formulas.

Microbial Fermentation Systems

Membrane techniques are well established in fermentation processes and are increasingly used in integrated bioprocesses. Figures 2 and 3 give an indication of the possible implementation of an integrated bioprocess in the dairy industry using cheese whey/waste water as substrate for the production of lactic acid. Regarding the fermentation of concentrated substrates, one application for the ED technique during upstream processing may be a partial reduction of the salt content and to obtain osmotic pressure to their optimum values. A final microfiltration step can be incorporated to obtain a continuous

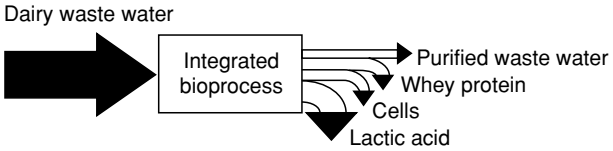

Figure 2 The recovery of lactic acid and other products of value from whey/dairy waste water, using an integrated bioprocess system. Courtesy of Fraunhofer Institute for Interfacial Engineering and Biotechnology, Stuttgart, Germany.

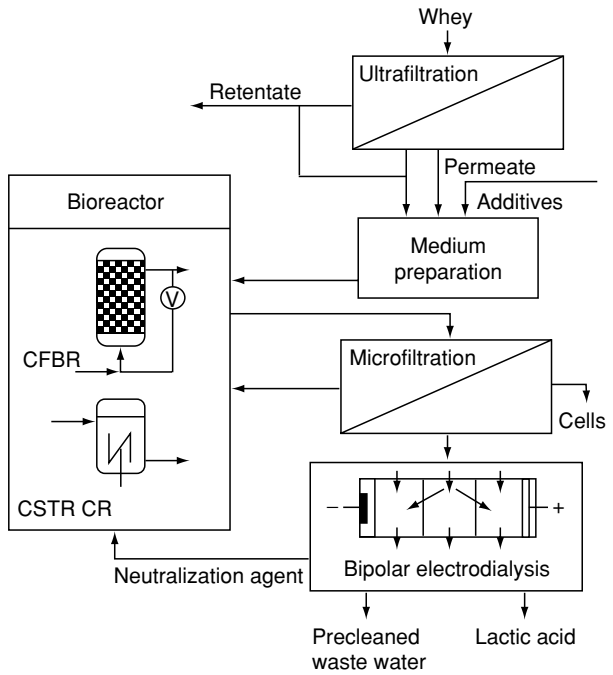

Figure 3 An integrated bioprocess for the production of lactic acid from whey using an integration of membrane systems in combination with a fermentation process. CSTR CR, continuous stirred tank reactor with cell recycling; CFBR, continuous fluidized-bed reactor. Courtesy of Fraunhofer Institute for Interfacial Engineering and Biotechnology, Stuttgart, Germany.

delivery of sterile substrate to the bioreactor. In continuous fermentation systems, membranes can be used for cell separation, retention of metabolic compounds, and recirculation. The reaction compounds are discharged from the membrane reactor with the permeate, which is then replaced by a new substrate solution. Owing to cell retention, cell density and, hence, substrate conversion velocity can be significantly enhanced.

The application of ED in downstream processing offers many economically competitive methods for the recovery of whey-based lactic acid:

- Continuous cell recycling with UF or microfiltration membranes followed by the removal of lactic acid by ED units.
- Acid recovery by means of ED with bipolar membranes (Figures 3 and 4). Bipolar membranes consist of cation and anion exchange layers arranged in parallel between two electrodes which enable the lactate salt to be split and electrically neutralized by H^+ ions and the corresponding alkali. The function of the bipolar membrane in the electrodialytic production of acids and bases from the corresponding salt solution is based on its capability for enhanced water dissociation under the influence of an electric potential gradient.

The combination of technologies and the demand for specialized products will in future lead to the next generation of highly differentiated whey protein-based products, with specific functionalities.

Other Applications of Dialysis

Dialysis and ED techniques are also applied in the demineralization of foods such as soy sauce, fruit juice, vinegar, hydrolyzed vegetable protein, sugars, and wine; deacidification of foods; purification of fermented products, hydrolysates, lipopolysaccharides, and protein fractions; as well as for the isolation of flavor substances and the production of enzymes. Other applications include the recovery of salts, acids, or alkali from industrial rinse waters as well as the purification of drinking water supplies, e.g., drinking water nitrate reduction.

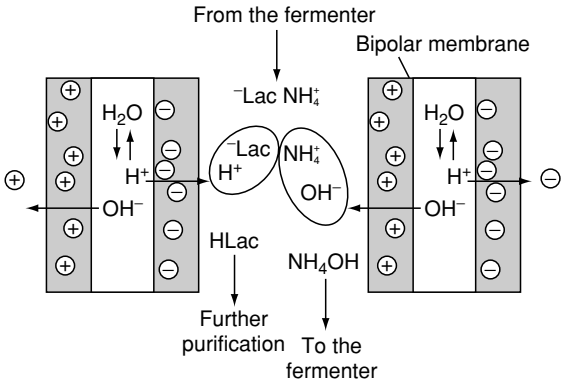

Figure 4 Schematic diagram of the function of bipolar electrodialysis membrane. Reproduced from Kulozik U (1992) membranes in microbial fermentations. In: *New Applications of Membrane Processes*. Special issue 9201, pp. 141–160. Brussels: International Dairy Federation, with permission.

See also: **Cheeses**: Chemistry of Gel Formation; Manufacture of Extra-hard Cheeses; Manufacture of Hard and Semi-hard Varieties of Cheese; **Effluents from Food Processing**: Microbiology of Treatment Processes; Disposal of Waste Water; **Infant Foods**: Milk Formulas; Weaning Foods; **Membrane Techniques**: Principles of Reverse Osmosis; Applications of Reverse Osmosis; Principles of Ultrafiltration; Applications of Ultrafiltration; **Whey and Whey Powders**: Production and Uses; Protein Concentrates and Fractions; Fermentation of Whey; Principles of Dialysis

Further Reading

Batchelder BT (1987) Electrodialysis applications in whey processing. In: *Trends in Whey Utilization*. IDF Bulletin no. 212, pp. 84–90. Brussels: International Dairy Federation.

Bylund G (1995) The chemistry of milk. In: *Dairy Processing Handbook*, p. 17. Lund, Sweden: Tetra Pak (Processing System Division).

Bylund G (1995) Whey processing. In: *Dairy Processing Handbook*, pp. 331–351. Lund, Sweden: Tetra Pak (Processing System Division).

Fraunhofer Institut Grenzflächen- und Bioverfahrenstechnik (2001) Integrated bioprocesses – recovery of value products from waste materials. Internet: http://www.igb.fhg.de/UWbio/en/Lactic_acid.en.html. Accessed 6 November 2001.

Horton B (1998) The whey processing industry – into the 21st century. In: *Whey*, special issue 9804, pp. 12–25.Brussels: International Dairy Federation.

Ionics (2001) Membrane systems for food and dairy processing. Internet: http://www.ionics.com/products/WaterFoodChemical/FoodProcessing/fooddairy.htm. Accessed 12 January 2001.

Kulozik U (1992) Membranes in microbial fermentations. In: *New Applications of Membrane Processes*. Special issue 9201, pp. 141–160. Brussels: International Dairy Federation.

López-Leiva MH (1988) The use of electrodialysis in food processing. Part 2: Review of practical applications. *Lebens-mittel-Wissenschaft und Technologie* 21: 177–182.

Mostert JF and Roberts JJ (1993) Applications of dialysis. In: Macrae R, Robinson RK and Sadler MJ (eds) *Encyclopaedia of Food Science, Food Technology and Nutrition*, vol. 8, pp. 4905–4908. London: Academic Press.

WHISKY, WHISKEY, AND BOURBON

Contents

Products and Manufacture

J R Piggott, University of Strathclyde, Glasgow, UK
J M Conner, The Scotch Whisky Research Institute, Riccarton, Edinburgh, UK

Synopsis

Whiskies are distilled beverages which have been produced by the distillation of fermented cereals and have been matured in oak casks. They can be classified according to starting material, blending and country of origin. The major grains used in whisky production are maize, barley, wheat and rye. Amylase activity required for starch breakdown is obtained from malted barley, and some regulations permit the use of other enzymes. The yeast used for fermentation of the cereal extract is normally a specially developed strain of *Saccharomyces cerevisiae*. Distillation is in small pot stills or continuous column stills, followed by maturation normally for at least 3 years in oak casks.

Types of Product

Whiskies are distilled beverages that have been produced by the distillation of fermented cereals and have been matured in oak casks. They can be classified on the basis of the nature of the cereal, the manner in which the product is blended, if at all, and the country of origin. Although there is some evidence for a distillation-type process in ancient Egypt, the first reports of whisky production arise in Ireland in the twelfth century, and the first description of a whisky production process is found in Scottish records dating from 1494. The term 'whisky' is derived from the Gaelic '*uisge beathe*,' which means 'water of life.' Major countries of production are Scotland, the USA, Ireland, Canada, and Japan, while similar products are produced in many other countries including Brazil, Korea, and Spain.

Regulations in the European Union (EU) define whisky as a spirit produced by the distillation of a mash of cereals and matured for at least 3 years in wooden casks. There are limitations on the enzymes that can be used to hydrolyze starch to fermentable sugars and on the distillation strength, and yeast must be used for the fermentation. Within this definition, Scotch and Irish whiskies are further restricted. In practice, this provides for three distinct products: malt whisky, produced from 100% malted barley; grain whisky, produced from unmalted cereal grains; and the bulk of the whisky produced in Scotland, which is blended from 60–70% grain whisky and 30–40% malt whiskies. This blended whisky usually contains up to 40 individual malts that are blended to produce a consistent brand flavor, and every component of the blend must be matured for the minimum period, or the age specified on the bottle. Malt whiskies may contain either a blend of malt whiskies of a given minimum age or malt whisky from a single distillery. Scottish malts have a different character depending upon whether they are from the west coast, such as Campbeltown and the Islay region, which produce highly peated malts, or from the Highlands or the Lowlands. Irish whiskies have a similar flavor to Scotch but do not have a peaty character.

Whisky produced in the USA has evolved in a different legislative framework, allowing for a greater variety of distinctive products. The basic definition of US whiskey is that it is distilled from a fermented grain mash at less than 95% alcohol (190° US proof). Within this broad definition, there are two special categories, which are bottled as such or provide flavoring whiskies for blending. In the first group, the cereal content of the mash is restricted to not less than 51% of the named cereal (e.g., rye for rye whiskey, maize for Bourbon), the distillation must be at less than 80% alcohol, and maturation must be in new charred oak. Bourbon takes its name from Bourbon County in Kentucky. Corn whiskey must be from a mash of 80% maize, and matured in new uncharred or used oak. If matured for 2 or more years, these are 'straight whiskies.' They are to the blender the equivalent of the Scotch malts, and provide the majority of the flavor in a blend with lighter flavored grain spirits. US blenders are also permitted to use 'blending materials,' such as sherry or blending wine, at up to 2.5%. (*See* **Barrels**: Wines, Spirits, and

Other Beverages; **Sherry**: The Product and its Manufacture; **Wines**: Production of Table Wines.)

Canadian whisky production follows the same general practice as in Scotland and the USA. Canadian blends are light column still whiskies with normally 10–15% of heavier-bodied flavoring whiskies. Up to 9.09% (on the basis of absolute alcohol content) of 'flavorings' is permitted, which may include wines and sherries, rum, brandy, bourbon, or malt whisky. Maturation must be for not less than 3 years, and oak casks of 182 l (40 imperial gallons) are commonly used. The range in flavor required by the blender cannot be achieved by using the products of many different distilleries, but techniques such as variation of the composition of the mash (corn, wheat, rye or barley), cooking procedures, yeast strain, distillation procedure, barrel type, and maturation period can be used. Traditionally, pot stills were used but these have largely been phased out. (*See* **Brandy and Cognac**: Armagnac, Brandy, and Cognac and their Manufacture; **Rum**.)

Similar products are produced in many other countries and are subject to local regulations, which, while generally similar in principle to the EU definition, allow different extents of variation. In Brazil, regulations permit the addition of a proportion of agricultural alcohol to blended whisky, and in Japan, the locally produced spirit is frequently blended with imported Scottish malts. Spain is a substantial producer of whisky, which must meet the EU requirements. EU and USA regulations also include provision for grain spirits or grain brandies, which are essentially similar to whiskies but not matured. In other parts of Europe (e.g., Poland), 'whiskies' are produced that are essentially vodkas, matured briefly and flavored with wood extracts or other materials, and which may or may not meet regulations commonly accepted for whiskies.

Raw Materials

The major grains used in whisky production are maize (*Zea mays*), barley (*Hordeum polysticum*), wheat (*Triticum vulgare*), and rye (*Secale montanum*). These are the grains traditionally used for whisky, on the basis of a high fermentable starch content providing a good yield of alcohol. A typical composition of these grains is shown in Table 1. In Scotland and Ireland, all the amylase activity required for starch breakdown is obtained from malted barley, in the form of either a kilned malt, which is stored at a moisture content of 5%, or a green malt, which is prepared specifically on site as a source of enzyme activity for production of grain whisky. Some regulations also permit the use of other (normally microbial) enzymes during wort production. Cultivars are normally selected on the basis of extract and fermentability, corresponding to spirit yield, or enzyme activity in the case of malt for grain distilling. Enzyme activity should be considered under process conditions, as this will not be the same as activity measured under optimum laboratory conditions. Ethyl carbamate is an undesirable trace component in distilled beverages, and its content in malt whisky is largely determined by the cyanogenic precursor, epi-heterodendrin, in malt. Barley cultivars that do not contain the precursor are being developed. (*See* **Barley**; **Rye**; **Wheat**: Grain Structure of Wheat and Wheat-based Products.)

Water is required for mashing, cooling during distillation, dilution of the spirit before maturation, and blending. The quality of the product may be influenced by the nature of the water supplied, and many distilleries take their water supply from individual water courses or special wells. Soft water is normally used for malt distilleries, but some degree of hardness is preferred in grain distilleries since this has a stabilizing effect on the amylase enzymes. The water should be monitored for chemical and microbiological contamination and be free of decaying vegetable matter. The water used for dilution of the spirit must have a low calcium and iron content, and is normally demineralized, to minimize the risk of discoloration or precipitation occurring in the final product.

The yeast used for fermentation is normally a specially grown distilling yeast that has good physiological and microbiological properties for producing alcohol. However, some recycled brewer's yeast is commonly used in Scotland, and yeasts from previous distillery fermentations may be used in some Bourbon

Table 1 Composition of cereals used in the production of whisky worts

	Corn	*Rye*	*Barley*	*Wheat*
Percentage of total				
Endosperm	82	87	84	85
Germ	12	3	3	3
Bran	6	10	13	12
Chemical composition (%, dry-weight basis)				
Nitrogen-free extract	69.2	70.9	66.6	69.9
Starch	72	68	63–65	69
Sugars	2.6	0	2–3	0
Protein	8	12.6	12	13.2
Soluble nitrogen (% of total)	4.7	0	11	0
Crude fiber	2	2.4	5.4	2.6
Fat	3.9	1.7	1.9	1.9
Ash	1.2	1.1	2	1.9

From Bronsky AJ and Schumann RW (1989) Cereals for whisky production. In: Piggott JR, Sharpe R and Duncan REB (eds) *The Science and Technology of Whiskies*, pp. 1–18. Harlow, UK: Longman Scientific and Technical, with permission.

processes, which may influence the range of flavor products in the wash. The main criterion for yeast strain selection is normally efficiency in converting fermentable sugars to ethanol. To a limited extent, strains may be selected to provide the required composition and flavor characteristics in the distillate, particularly the ester content, and yeast cultivation and storage conditions also affect its performance. Yeasts expressing glucoamylase activity have been developed.

Barrels used in whisky production are produced from either American white oak, *Quercus alba* and related species, or the Spanish oak, *Quercus robur*. Scotch whisky is traditionally matured in barrels which have previously been used for Bourbon maturation or sherry production. Spanish sherry production is carried out in barrels constructed from either Spanish or American white oak. Except where the regulations provide otherwise, barrels are commonly reused until they fail to provide a satisfactory maturation. Such barrels may be regenerated by scraping out the inner surface and charring the inside of the barrel, but this does not produce performance equivalent to a new barrel. During periods when sherry casks have been in short supply, barrels have also been wine treated or wine seasoned (filled with sherry for up to 6 months), and Madeira or port casks are also occasionally used. (*See* **Barrels:** Beer Making; **Sherry**: The Product and its Manufacture.)

Method of Manufacture

Malting and Mashing

The traditional method of production involves the production of wort from cereal grains using amylases from barley. These α- and β-amylases are produced by the barley during the malting process. Malting of barley is carried out by steeping the grain in water for 2–3 days and permitting the grain to germinate until the root is approximately one-third the length of the grain. When this stage is reached, the grain is dried down to a moisture level of 5% by kilning. The grain is then stored until immediately prior to use, when it is milled before being transferred to a mash tun. In Scotch whisky production, peated malt is prepared by drying the malt in the presence of smoke from a peat fire and is an important source of flavor compounds, normally assessed on the basis of the total phenols content of the malt (up to 50 p.p.m.), although other components may contribute to flavor. The nitrosamine contamination of malt has recently been of concern in Scotch whisky, arising from the use of natural gas in directly fired kilns. This can be controlled by the use of burners designed to reduce nitrogen oxides, by burning sulfur to add sulfur dioxide to the air stream, or by indirect firing. (*See* **Flavor (Flavour) Compounds**: Structures and Characteristics; **Phenolic Compounds**.)

Most of the breakdown of the starch occurs during the process of mashing, which is carried out in a mash tun. Barley grist can be mashed directly since it has a high amylase activity, whereas worts produced from maize and other cereals require the addition of malted barley as a source of enzymes. The process of mashing may be a simple batch process similar to the infusion mashing techniques used in brewing beer, in which the grist is mixed with water at 63–68 °C for 0.5–1.5 h before the wort is filtered off. However, in the production of grain whisky, an initial cooking stage, to gelatinize the starch of the nonbarley cereals, is required. The cooking may be carried out as a batch process at 120 °C for 1.5 h or, as is more usual in North America, as a continuous process. In either case, the mash is cooled to 60–65 °C before the addition of the malted barley. Careful control is required at this stage to prevent the solidification of the gelatinized starch. When a cooking stage is used, amylase activity is normally provided by adding 10–15% freshly malted barley (green malt) after the cooking stage (Figure 1). (*See* **Starch**: Structure, Properties, and Determination.)

During the malting stage, β-glucanase enzymes are also produced; these facilitate the breakdown of the endosperm cell walls and the release of starch and proteases. The latter are involved in the breakdown of some of the barley grain proteins to amino acids and peptides. These amino acids are an essential nitrogen source for the growth of the yeast. (*See* **Amino Acids**: Properties and Occurrence; **Peptides**.)

The mash is usually filtered prior to the fermentation to give a clear wort, but in some processes, a higher yield is obtained by fermenting the whole mash directly. The composition of a typical Scotch grain whisky wort is shown in Table 2. Refer to individual ingredients.

Fermentation

The wort, which is produced when the contents of the mash tun has been filtered, is then transferred to a fermentation vessel and fermented using one or more strains of the yeast *Saccharomyces cerevisiae* to give a wash containing approximately 7% ethanol and a large number of flavor compounds. When the wort is not boiled (as is the case in malt whisky production), it contains microorganisms and active amylases. It is cooled to 20–25 °C before inoculation with a strain of distilling yeast and, in some processes, secondary yeasts. Yeast growth occurs rapidly and is completed in the first 8–12 h. Ethanol production follows

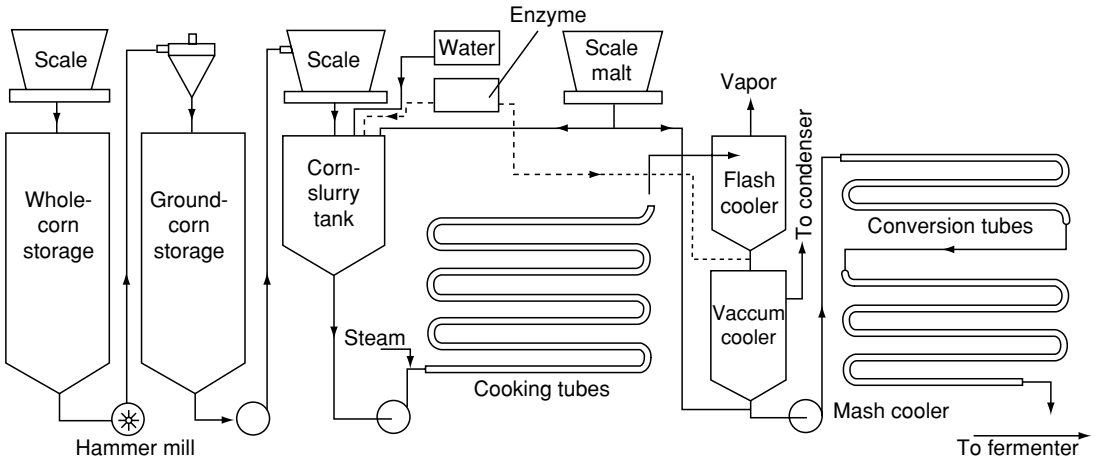

Figure 1 Maize process for whisky. From Simpson AC (1985) Advances in the spirits industry. In: Birch GG and Lindley MG (eds) *Alcoholic Beverages*. London: Elsevier Applied Science, pp. 51–67, with permission.

Table 2 Chemical composition of a typical Scotch grain wort

	Content
Total soluble carbohydrate (as glucose) (%)	9.00
Insoluble solids (%)	2.20
Sugars (%)	
Fructose	0.13
Glucose	0.29
Sucrose	0.28
Maltose	4.65
Maltotriose	0.96
Maltotetraose	0.15
Dextrin	2.54
Amino nitrogen (as leucine) (%)	0.09
Ash (%)	0.27
Containing P_2O_5	0.09
Containing K_2O	0.09
Containing MgO	0.02
Vitamins ($\mu g\ ml^{-1}$)	
Thiamin	0.46
Pyridoxine	0.61
Biotin	0.01
Inositol	236
Nicotinic acid	11.1
Pantothenate	0.71

From Pyke M (1965) Manufacture of Scotch grain whisky. *Journal of the Institute of Brewing* 71: 209, with permission.

growth initially, then continues in a linear manner until the metabolizable sugars have been consumed. During the growth period, wort sugars are rapidly metabolized, and the temperature may rise to 33 °C if it is not controlled. The pH drops from an initial value of 5.0–5.5 to 4.2–4.5 as a result of the production of organic acids, such as acetate, succinate, and pyruvate. A wide range of flavor compounds, such as higher alcohols, fatty acids, esters, and carbonyl compounds, are also produced throughout the fermentation; they pass through into the distillate and constitute an important part of the product. In contrast, glycerol, which is also produced, is nonvolatile and does not pass into the distillate; this represents an important loss of yield. At the end of the fermentation, usually between 48 and 72 h, death and autolysis of the yeast may occur. The fermentation is not a sterile process, so bacterial contamination can occur. Most of the bacteria that are introduced from the raw materials die during the early stages of the fermentation. However, lactobacilli can survive and grow with a concurrent formation of lactic acid during the later stages of the fermentation. There is some suggestion that limited growth of lactic acid bacteria is beneficial in flavor terms, but to prevent excessive growth and resulting yield losses, fermentation is normally completed in 40–48 h, rather than the longer times traditionally used. Heavy contamination by lactic acid bacteria can also result in a loss of yield and the production of acrid off-flavors. (*See* **Lactic Acid Bacteria**.)

Distillation

The fermented wash is then distilled either in a series of two or three pot stills (Figure 2), or in a series of column stills. The distillation process varies with the type of product and the country of origin. Historically, the first whisky processes used copper pot stills, and this type of process is still used for the production of malt whisky in Scotland, where two stills – the wash still and the spirit still – are used. In Ireland, a three-still process was typical. Pot stills have also been used in the USA and Canada, but grain spirit is normally produced by continuous distillation in column stills. Again, the first continuous process is associated with Scotch whisky production and the development of the Coffey or Patent still by Aeneas Coffey in 1830. The ethanol concentration of the

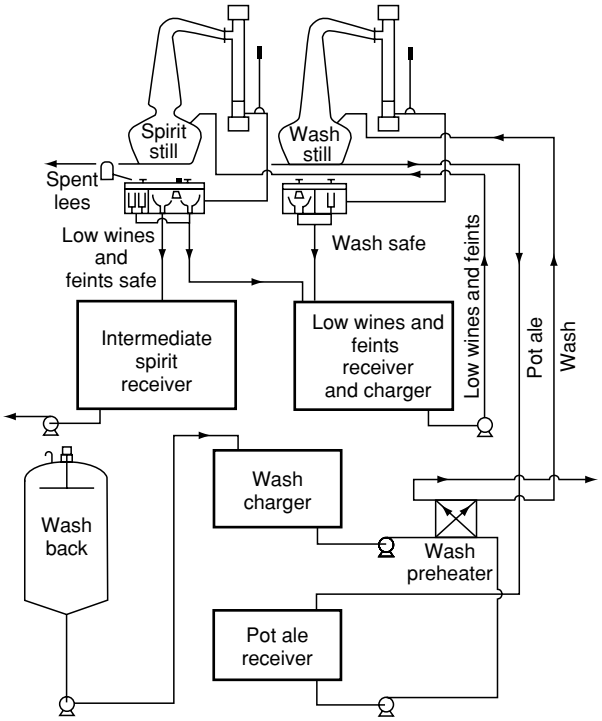

Figure 2 Typical malt distillery flow diagram. From Nicol D (1989) Batch distillation. In: Piggott JR, Sharpe R and Duncan REB (eds) *The Science and Technology of Whiskies*, pp. 118–149. Harlow, UK: Longman Scientific and Technical, with permission.

distillate determines the extent of impurities (congeners) present, a higher strength distillate being purer aqueous ethanol, and thus its strength of flavor.

In batch distillation processes, all the contents of the fermenter are transferred to the large wash still and boiled for 5–6 h to produce a distillate, low wines, which contains 20–25% (v/v) ethanol. This is transferred to the smaller low wines or spirit still where it is distilled to give a spirit containing around 70% (v/v) ethanol (Figure 3). The Coffey still consists of two columns, referred to as the 'analyzer' or 'beer column', and the 'rectifier column.' The wash is fed towards the top of the first column, and alcohol-free stillage is withdrawn at the bottom of the column; vapors are fed from the top to the base of the rectifying column (Figure 4). This produces a 190° US Proof spirit and is still used for the production of grain spirit in Scotland. In the USA, whisky production employs more complex distillation processes involving the use of up to five columns. The heat for most distillation processes is supplied as steam at the present time, although pot stills directly fired by solid fuel or gas are also in use. In addition to producing spirit, distillation processes produce a large quantity of stillage, which is processed into various by-products, such as dark grains and distiller's solubles.

Maturation

All whiskies are matured, and most legal definitions of whisky incorporate a minimum maturation period. The legal minimum varies from 1 year for some American grain whiskies to 3 years for Scotch, Irish, and Canadian whiskies. In effect, these are the minimum requirements, and most spirits are matured for longer periods. Branded products are generally matured for between 4 and 8 years, whereas single malts and premium straight whiskies are matured for 8 to more than 18 years. The process of maturation involves storing the raw spirit from the still, diluted to 50–80% ethanol, in barrels made from a limited range of oak woods. The type of cask used greatly affects the properties of the mature spirit. American straight whiskies (e.g., Bourbon) are matured in new charred American oak barrels that by law cannot be reused. Maturation in new charred casks is associated with an increase in color, solids, acids, esters, aldehydes, and tannins, together with increases in sugars, such as arabinose, glucose, and xylose. (*See* **Carbohydrates:** Classification and Properties; **Tannins and Polyphenols.**) New charred casks give a strong sweet, wood-derived flavor to the final product. The casks used for straight whisky maturation are American standard barrels (180 l) made from American white oak. 'White oak' is a cooperage trade classification used for at least 10 oak species, the principal species being *Quercus alba*.

Most other types of whisky are matured in barrels previously used either for maturation of Bourbon or in sherry production. Bourbon barrels may be used 'as is' or reassembled as hogsheads, with additional staves and new ends to increase capacity (250 l). Reuse of casks decreases the levels of wood components extracted during maturation and gives a much less intense wood-derived flavor. Sherry casks used for the maturation of some whiskies are generally butts or puncheons (480–520 l), which have been used for the fermentation and/or aging of various styles of sherry. Sherry casks can be either American or Spanish oak casks (predominantly *Quercus petraea* and *Quercus robur*). American oak is generally used for fino and amontillado sherries, whereas Spanish oak is used for oloroso sherry. For sherry casks, distillers will specify the species of oak, whether the cask is used for fermentation and the contact time with maturing sherry, as all are known to affect the final properties of the mature spirit. In all but Bourbon production, casks are reused until they are exhausted, i.e., fail to produce a satisfactory change in the spirit over an economical maturation period. The life of a cask can be extended by removal of the old char layer and refiring the cask. This regeneration

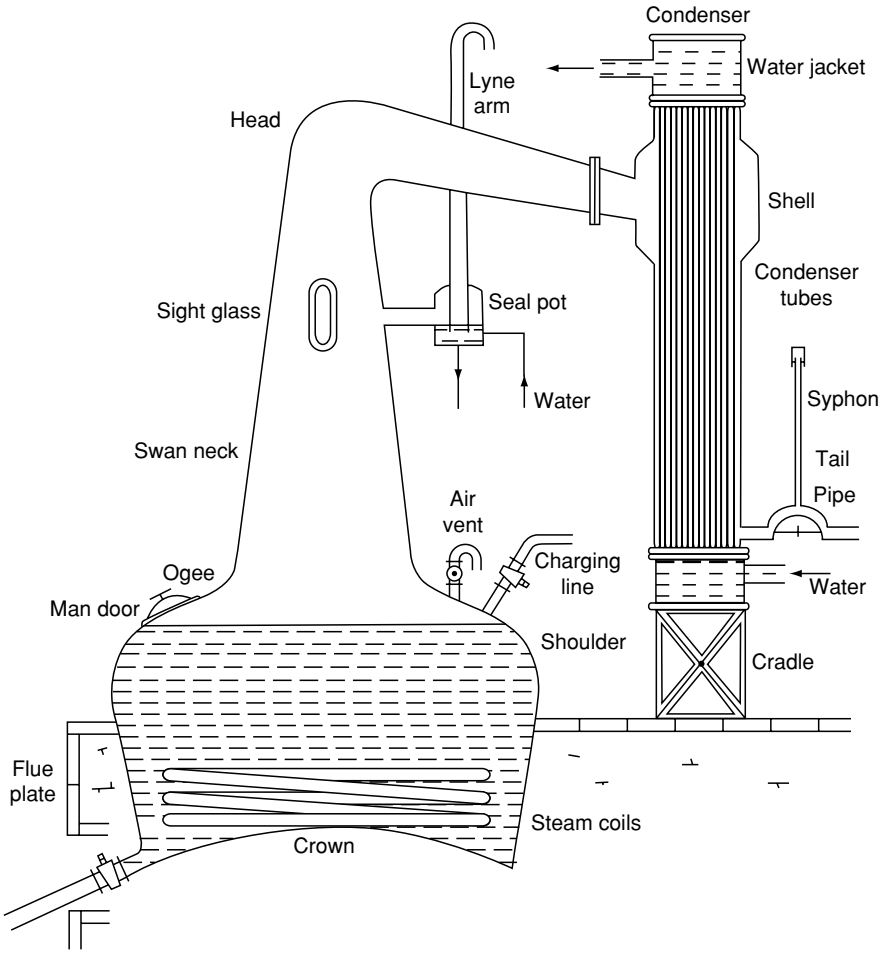

Figure 3 Plain wash still. From Nicol D (1989) Batch distillation. In: Piggott JR, Sharpe R and Duncan REB (eds) *The Science and Technology of Whiskies*, pp. 118–149. Harlow, UK: Longman Scientific and Technical, with permission.

increases the levels of wood polymer breakdown products extracted during subsequent maturation, but does not regenerate all wood constituents, and such casks give different sensory properties when compared with a new cask.

Maturing whisky was traditionally stored in stone-built warehouses located beside the distillery. Casks were stored in racks or in 'stows,' two or three high on the earth floor. Expansion in production resulted in much larger warehouses with steel racking, which allowed casks to be stored up to 12 high. The temperature and humidity within warehouses vary between different countries, and this can affect the way in which a spirit matures. In cool damp climates, such as Scotland and Ireland, more ethanol than water evaporates from the cask and consequently strength decreases during the maturation period. In warm dry climates, such as the USA, more water than ethanol evaporates and strength increases. During maturation, many of the chemical changes that occur are influenced by the storage conditions, particularly temperature. However though higher maturation temperatures have been shown to increase the rate of extraction of wood components, this does not generally correlate with improvements in sensory quality.

Blending to Bottling

All whiskies are to some extent blended. Even in the most tightly controlled environments, batch-to-batch variations can occur at any stage in the production process. Blending is used to produce a consistent product that has the distinctive flavors of that particular brand. For Bourbon and single malt production, variations are evened out to create the recognized brand from different batches of whisky. True blended whiskies consist of light bodied spirits mixed with a number of heavier bodied spirits in a wide range of proportions. 'Light-bodied spirits' are those distilled to high ethanol concentrations using continuous column stills and include Scotch grain and American light whiskies. 'Heavier bodied' whiskies are either

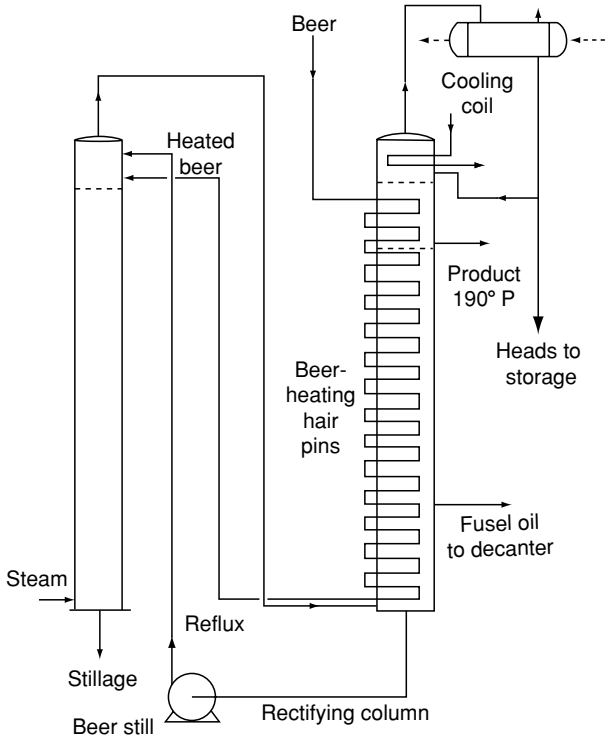

Figure 4 Coffey still. From Panek RJ and Boucher AR (1989) Continuous distillation. In: Piggott JR, Sharpe R and Duncan REB (eds) *The Science and Technology of Whiskies*, pp. 150–181. Harlow, UK: Longman Scientific and Technical, with permission.

batch pot still products or column still products distilled to lower ethanol concentrations. The nature and components of blends are determined by the traditions and regulations of the country of origin. In Scotland, there are a large number of distilleries, and blend recipes are complex with as many as 50 malt whiskies used. Purchasing or exchanging new whiskies by producers maintains the complexity of these blends. In North America, where there are fewer distilleries, and trading between competitors is uncommon, the variety of whiskies for blending can be increased by variations in mash bills, fermentation conditions, distillation parameters, maturation periods, and cask types.

The physical process of blending is, however, very similar. Approved whiskies are drained from casks into passivated steel troughs, which convey the whisky to a blending vat. Mixing uses mechanical agitators and compressed air. When the blend is correct, deproofing water is added to reduce the alcohol content to bottle strength. In Scotland, blending is sometimes followed by a further maturation or marrying period to allow integration of the individual components into the final product. An additional maturation or finishing period is also used for some single malts. This frequently employs a distinctive wood type, such as port casks, and is used to add complexity to a whisky or to increase the number of styles available from a single distillery.

Most whiskies are filtered prior to bottling to reduce the risk of haze formation during storage. In heavier bodied older whiskies and whiskies matured at high strengths, high-molecular-weight lipids and wood components slowly flocculate when the whisky is reduced to bottling strength. This is controlled by chill filtration, where whisky is cooled to between -10 and $10\,^{\circ}C$ for a specified period of time, and the problem compounds are removed by physical separation and adsorption by a filter. The conventional filter is a plate and frame filter with preformed pads of cellulose, or cellulose impregnated or precoated with diatomaceous earth. Operational parameters depend on the batch size, the nature of the product, and the filtration rate required.

See also: **Amino Acids**: Properties and Occurrence; **Barley**; **Barrels**: Wines, Spirits, and Other Beverages; **Brandy and Cognac**: Armagnac, Brandy, and Cognac and their Manufacture; **Carbohydrates**: Classification and Properties; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Lactic Acid Bacteria**; **Phenolic Compounds**; **Rum**; **Rye**; **Sherry**: The Product and its Manufacture; **Starch**: Structure, Properties, and Determination; **Tannins and Polyphenols**; **Wheat**: Grain Structure of Wheat and Wheat-based Products; **Wines**: Production of Table Wines

Further Reading

Berry DR (1984) Physiology and microbiology of Scotch whisky production. In: Bushell ME (ed.) *Progress in Industrial Microbiology*, pp. 199–243. Amsterdam: Elsevier Science.

Campbell I (ed.) (1990) *Proceedings of the Third Aviemore Conference on Malting, Brewing and Distilling*. London: Institute of Brewing.

Campbell I (ed.) (1999) *Proceedings of the Fifth Aviemore Conference on Malting, Brewing and Distilling*. London: Institute of Brewing.

Campbell I and Priest FG (eds) (1986) *Proceedings of the Second Aviemore Conference on Malting, Brewing and Distilling*. London: Institute of Brewing.

Campbell I and Priest FG (eds) (1995) *Proceedings of the Fourth Aviemore Conference on Malting, Brewing and Distilling*. London: Institute of Brewing.

Lyons TP and Rose AH (1977) Whisky. In: Rose AH (ed.) *Economic Microbiology. 1. Alcoholic Beverages*, pp. 635–692. London: Academic Press.

Piggott JR, Sharp R and Duncan REB (eds) (1989) *The Science and Technology of Whiskies*. Harlow, UK: Longman.

Priest FG and Campbell I (eds) (1983) *Current Developments in Malting, Brewing and Distilling*. London: Institute of Brewing.

Composition and Analysis of Whisky

J R Piggott, University of Strathclyde, Glasgow, UK
J M Conner, The Scotch Whisky Research Institute, Riccarton, Edinburgh, UK

Introduction

Whisky has traditionally been assessed by expert blenders with many years of experience, by comparison of samples with experience or a reference. Vocabularies of descriptive terms have been used, which have often been presented in the form of a 'flavor wheel.' With the advancement in chromatographic techniques and their coupling with mass spectrometry, the number of compounds identified in whisky has increased dramatically. They include alcohols, carboxylic acids, esters, aliphatic carbonyl compounds, heterocyclic sulfur compounds and aliphatic sulfides, heterocyclic nitrogen compounds, a wide range of simple phenolic compounds, phenolic aldehydes, and heterocyclic oxygen compounds.

Sensory Assessment

Whisky has traditionally been assessed by expert blenders with many years of experience and training within the industry. An experienced blender knows what flavors of distillate a still can produce, which are desirable, and how a whisky is likely to develop during maturation. The blender's task is then to identify faults and deviations from the expected path of maturation, and to select a specific maturation point at which a whisky can contribute to the blend. Blenders use a system of flavor description to assist their work, but their method of working is largely by comparison of samples with experience or a reference. Each sample is evaluated in terms of its similarity to an expected or acceptable product. (*See* **Sensory Evaluation**: Taste.)

Scotch grain whiskies are light-bodied continuous still spirits, largely regarded as providing a background for a blend. Malt whiskies provide the majority of the character of a particular blend, and, though the distilleries all produce slightly different whiskies, for convenience, they are divided into groups sharing common characteristics. The Highland malts, the most numerous, are further divided into first, second and third class. These are not intended to be quality evaluations, but are simply descriptive classes. The Lowland malts are not regarded as having a distinctive character, and thus can be used in relatively large quantities in blends. There are only two remaining Campbeltown distilleries, and this group has, as a result, declined in importance. The Islay whiskies tend to be full-flavored and are usually heavily peated, and thus have a significant impact on a blend even when used in small proportions. The Highland group contains most of the great whiskies, which have established reputations as single malts and for their value in blending. A wide range of flavors is available, and a skilled blender can make use of it to create a distinctive blend.

Formal methods of laboratory sensory analysis have also been used. Many companies use difference tests for quality-control monitoring, and the whisky industry seems to have been one of the first to adopt the duo–trio and triangle tests for such purposes. Over the past 25 years or so, there has also been an increasing interest in the assessment of whiskies by profiling or descriptive methods, largely for research into the origins and development of flavor in whiskies. Several broadly similar vocabularies of descriptive terms have been used. These vary mainly in the number of terms they include, and therefore in the detail with which they attempt to describe flavor. They have often been presented in the form of a 'flavor wheel,' with varying numbers of concentric tiers, which, while having no particular theoretical significance, provide a framework for communication and training (Figure 1).

The Chemistry of Flavor Compounds

With the advancement in chromatographic techniques and, in particular, their coupling with mass spectrometry, the number of compounds identified from whisky has increased dramatically. Particular use is made of gas chromatography with a number of detectors for the measurement of volatile organic compounds. Nonvolatile constituents may be measured as a whole spectrophotometrically, or individually by high-performance liquid chromatography. Problems arise, as compounds of sensory importance may be present at very low concentrations (nanograms per gram). Typical analyses of a Scotch malt whisky new distillate are shown in Table 1, and the major categories of compounds thought to be important in the development of flavor are discussed below. (*See* **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Mass Spectrometry**: Principles and Instrumentation.)

Alcohols, Carboxylic Acids, and Esters

Aliphatic alcohols, acids, and esters are produced by the yeast during fermentation, though ester formation

may also occur during distillation and maturation. Saturated and unsaturated chains are found, ranging from one to 16 carbon atoms. Short-chain alcohols such as propanols, butanols, and pentanols are the most abundant. Nonaliphatic alcohols include 2-phenylethanol and 2,3-butanediol. The principal organic acid is acetic acid, accounting for between 50 and 90% of the total content of volatile acids. Octanoic, decanoic, and dodecanoic acids are quantitatively important, though lower-molecular-weight acids such as butanoic and pentanoic acids may have more sensory impact. The most abundant esters are the ethyl esters of the acids named, their concentrations reflecting the relative abundance of the acid. Any other combination of the acids and alcohols present may occur. After ethyl esters, the most abundant are acetate esters such as isoamyl acetate and phenylethyl acetate. The esters of hydroxy acids, such as lactic acid, and of dicarboxylic acids, such as succinic acid, can be distilled along with alcohol and steam and so occur in whisky.

Aliphatic Carbonyl Compounds

The majority of carbonyl compounds are produced by the metabolic actions of the yeast. However, their formation by oxidation of unsaturated fatty acids or Strecker degradations has been shown to occur at almost any stage of production. Acetaldehyde is the most common and is often used to determine total aldehydic content. Many other aldehydes, ketones, and acetals have also been identified. For aldehydes and ketones, aliphatic chain lengths, both saturated and unsaturated, range from two to 14 carbon atoms. Due to the high alcoholic strengths of the distillate,

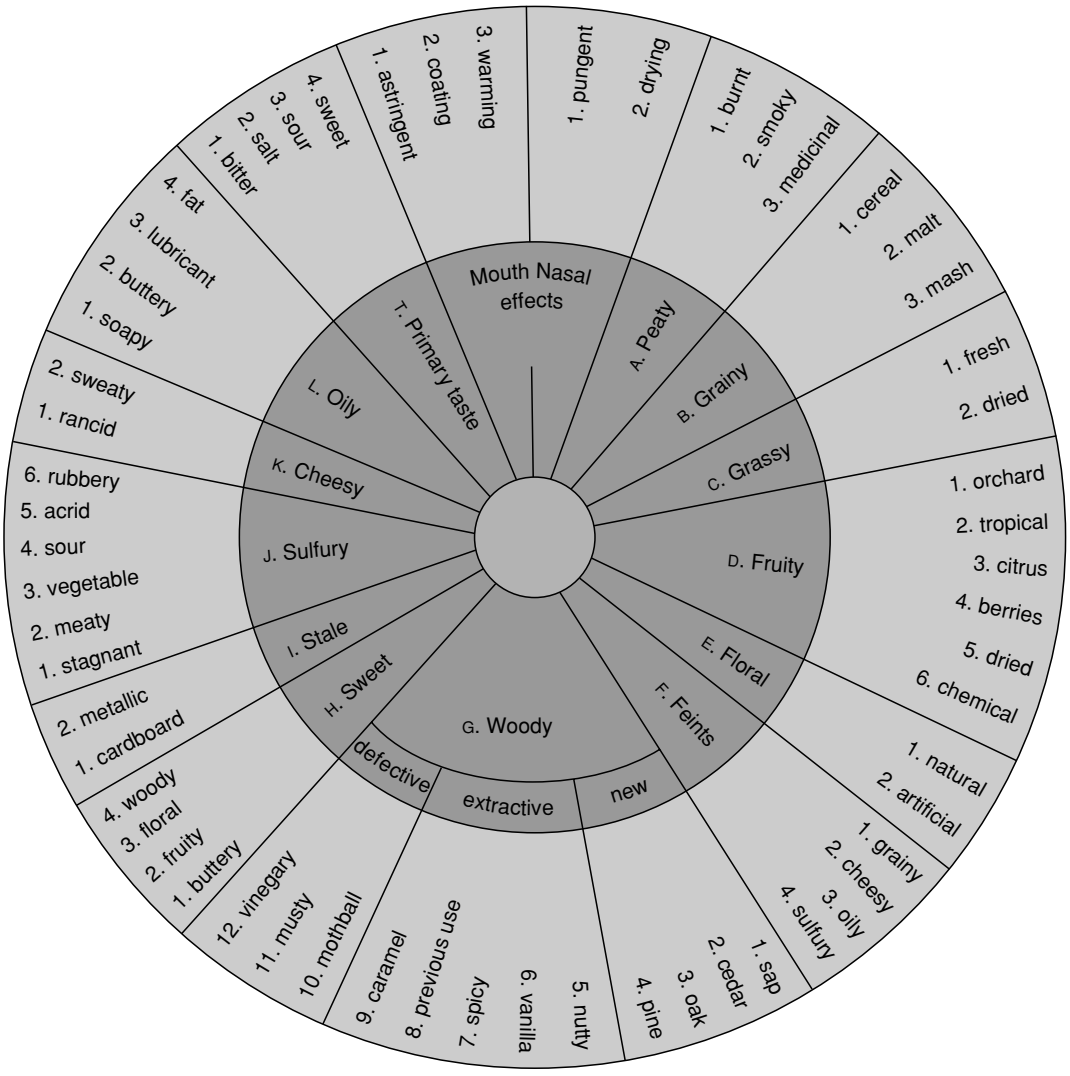

Figure 1 Whisky flavor wheel. From Lee KYM, Paterson A, Piggott JR and Richardson GD (2001) Origins of flavour in whiskies and a revised flavour wheel: a review. *Journal of the Institute of Brewing* 107: 287, with permission.

Table 1 Chromatographic analysis of three samples of a Scotch malt whisky new distillate[a]

Period (accounting period)	4	4	4
Charge (batch number)	12	14	16
Strength (% v/v)	63.5	71.5	69.4
Acetaldehyde	3.2	3.8	6.8
Ethyl acetate	23.7	25.5	27.0
Diethyl acetal	1.7	1.2	2.2
Methanol	5.1	4.6	5.3
Propanol	40.8	42.7	41.9
Isobutanol	79.8	80.8	80.5
Amyl alcohol[b]	47.7	44.7	49.5
Isoamyl alcohol	142.5	145.5	142.5
Total higher alcohols	331.1	313.7	314.4
Ethyl lactate	4.7	2.5	4.1
Ethyl octanoate	1.6	1.9	1.7
Furfural	3.3	3.9	4.2
Ethyl decanoate	5.7	5.6	4.5
β-Phenethyl acetate	5.7	7.5	5.9
Ethyl laurate	2.1	2.6	2.1
β-Phenylethanol	3.8	0.6	0.6
Ethyl myristate	0.6	1.1	0.6
Ethyl palmitate	2.7	3.3	2.6
Ethyl palmitoleate	1.5	1.9	1.4

[a]Concentrations are grams per 100 liters of alcohol.
[b]Optically active.
From Nicol D (1989) Batch distillation. In: Piggott JR, Sharpe R and Duncan REB (eds) *The Science and Technology of Whiskies*, pp. 118–149. Harlow, UK: Longman Scientific and Technical, with permission.

aldehydes exist in equilibrium with the corresponding acetal and hemiacetal. The most frequently reported acetal is 1,1-diethoxyethane, though a range of aldehydes up to six carbon atoms have been shown to form acetals.

Sulfur and Nitrogen Compounds

Yeasts have the ability to utilize inorganic sulfur, sulfur-containing amino acids, and vitamins, and from these sources, volatile sulfur-containing compounds are produced. The range of compounds detected includes heterocycles, such as thiazoles and thiophenes, and aliphatic mono-, di-, and trisulfides. Formation of sulfur compounds has also been shown to occur during distillation, and concentrations are highest in new spirit. Wood extract and traces of elemental copper are important in the degradation of such compounds during maturation (Table 2). Heterocyclic nitrogen compounds such as pyrazines, pyrroles, and pyridines are formed via the Maillard reaction during cooking, mashing, distillation, and barrel charring. As they have relatively low odor thresholds, they may well make a contribution to the aroma of whisky. (*See* **Browning**: Nonenzymatic.)

Phenols

A wide range of phenolic compounds are found in whisky. Simple phenols such as phenol, the isomeric cresols, xylenols, ethyl phenols, and guaiacols arise through the thermal degradation of benzoic acid derivatives from malt and from peat smoke. Phenolic aldehydes such as vanillin, syringaldehyde, coniferaldehyde, and sinapaldehyde are formed from the breakdown of wood lignin during cask charring and maturation. Their corresponding acids and ethyl esters have also been detected. Polyphenols extracted from the wood include ethanol-soluble lignin and tannins derived from both gallic and ellagic acids. (*See* **Phenolic Compounds**; **Tannins and Polyphenols**.)

Table 2 Concentrations (micrograms per liter at 100% alcohol) of sulfur compounds in malt whiskies of various ages

Compounds	*Age (years)*				
	0	*1*	*2*	*3–5*	*6–10*
Dimethyl disulfide	359	345	272	149	91
Dimethyl trisulfide	12	16	15	19	14
3-(Methylthio)propanal	7	t	0	0	0
Dihydro-2-methyl-3(2H)-thiophenone	77	53	3	0	0
Ethyl 3-(methylthio)propanoate	21	11	t	0	0
3-(Methylthio)propyl acetate	274	139	3	0	0
2-Thiophenecarboxaldehyde	66	83	49	60	54
3-(Methylthio)propanal	8	0	0	0	0
5-Methyl-2-thiophenecarboxaldehyde	15	15	10	16	16
Benzothiophene	39	27	27	31	38
Benzothiazole	18	16	14	15	32

t, trace.
From Masuda M and Nishimura K (1981) Changes in volatile sulfur compounds of whisky during aging. *Journal of Food Science* 47: 104, with permission.

Heterocyclic Oxygen Compounds

The main types of these compounds found in whisky are furaldehydes and lactones. Furfural (2-furaldehyde) is formed from pentoses and 5-hydroxymethyl-2-furaldehyde from hexoses via the Maillard reaction during mashing, distillation, and cask charring. Lactones are formed by the dehydration of aliphatic hydroxy acids, and most are derivatives of 2(3H)-furanones. The most abundant of these are the *cis* and *trans* isomers of 5-butyl-4-methyldihydro-2(3H)-furanone, the oak or whisky lactones, which are formed and extracted from oak casks during maturation.

Variations with Processing

Though there are only four basic processes in whisky production; local, regional, and national variations in these processes are responsible for the large number of whiskies available. Each stage of production, from the production of a fermentable extract, through to fermentation, distillation, and maturation, can affect the character of the final product. However, each stage is interlinked, and identifying one, as the most important for a particular type of whisky, is often impossible.

Production of a Fermentable Extract

The cereal grains in common use are wheat, corn (maize), rye, and barley, though the choice of cereal is often legally restricted. Flavor differences between cereals arise from two quite separate sources. The choice of cereal can also indirectly affect the flavor of the final product. The composition of the fermentable extract varies with cereal used and can influence the metabolism of the yeast. Factors such as pH, amino acid concentration, and levels of insoluble material are known to affect the levels of fusel alcohols, acids, and esters produced, which may influence the flavor of the final product. Between cereals, differences also exist in the levels of precursors for a number of classes of flavor compounds. Examples are compounds formed by the chemical and enzymatic oxidation of fatty acids, the reaction of free amino acids and reducing sugars (Maillard reaction), and the thermal degradation of biosynthesized precursors like *S*-methylmethionine and benzoic acid derivatives. However, the breakdown of these compounds frequently occurs during fermentation and distillation and is affected by variations in these processes. Consequently, no flavors in the final product have been unequivocally linked to components of the original cereal.

Fermentation

Fermentation is responsible for the primary production of a large number of flavor compounds in whiskies. Factors that are known to alter the production of these compounds are the strain or mix of yeasts used, design and level of aeration in the fermenter, the duration of the fermentation, and the presence of contaminating bacteria. These factors alter according to distillery design and the raw materials employed and so may be important in producing regional and national differences in whiskies.

Distillation

In batch distillation using pot stills, the selection of the cut points in the still run has a marked influence on the flavor of the product. Generally, aldehyde and short-chain ester concentrations are determined by the primary cut from foreshots to spirit, while the concentration of fusel alcohols and acids is determined by the cut from spirit to feints. Other factors, which affect the physical process of separation during distillation, are still design parameters such as the still height, the angle and length of the lyne arm, and the type of condenser used. In addition, a number of chemical reactions can occur when the wash or low wines are heated for distillation. This is particularly true for batch distillation in pot stills where the heat impact is greater. Examples include the formation of sulfur compounds from sulfur-containing amino acids, the breakdown of unsaturated fatty acids to carbonyl compounds, and the dehydration of β-hydroxypropionaldehyde to acrolein. Copper catalyzes many of these reactions.

In continuous stills, congener concentrations are determined by the design, operation, and efficiency of the still. The range of reactions described for pot stills can still occur but becomes progressively less important as the distillate strength increases. Collection of spirit at high ethanol concentrations (> 90%) results in low levels of congeners in the spirit.

Maturation

During the maturation period, the new distillate becomes highly modified as a result of its contact with the cask (Figure 2). The composition of the cask wood has a major effect on the extent of this modification and varies with the species of oak used, the pretreatment applied to the wood, and the previous use of the cask. In general, American oak (*Quercus alba*) yields higher concentrations of vanillin and *cis* oak lactone, whereas Spanish oak (*Quercus robur*) gives higher concentrations of tannins. Charring increases the levels of lignin breakdown products and lactones extracted during maturation, whereas previous use of a cask greatly reduces the amount of extractable lignin breakdown products and lactones. The use of sherry casks results in raised levels of tannins and sugars. The cask is the major determinant of mature quality, but other factors such as filling proof, climate, and warehouse conditions also affect the course of maturation. Higher temperatures produce greater evaporation rates and also increase extract levels and some reaction rates, but do not necessarily produce a more mature product.

Correlation of Chemical Composition with Sensory Data

The flavor of whisky is dependent upon a complex mixture of components that originate from, and are modified by, different stages of production. Although improved analytical methods have revealed much about the chemistry of whisky, this has not resulted in a better understanding of the chemical basis of whisky flavor and quality. Frequently, chemical differences can be related to variation in production processes, but identifying the actual compounds responsible for the flavor differences has been much more elusive.

The extent to which volatile compounds from cereals survive fermentation, distillation, and maturation is the hardest to establish. Whisky prepared

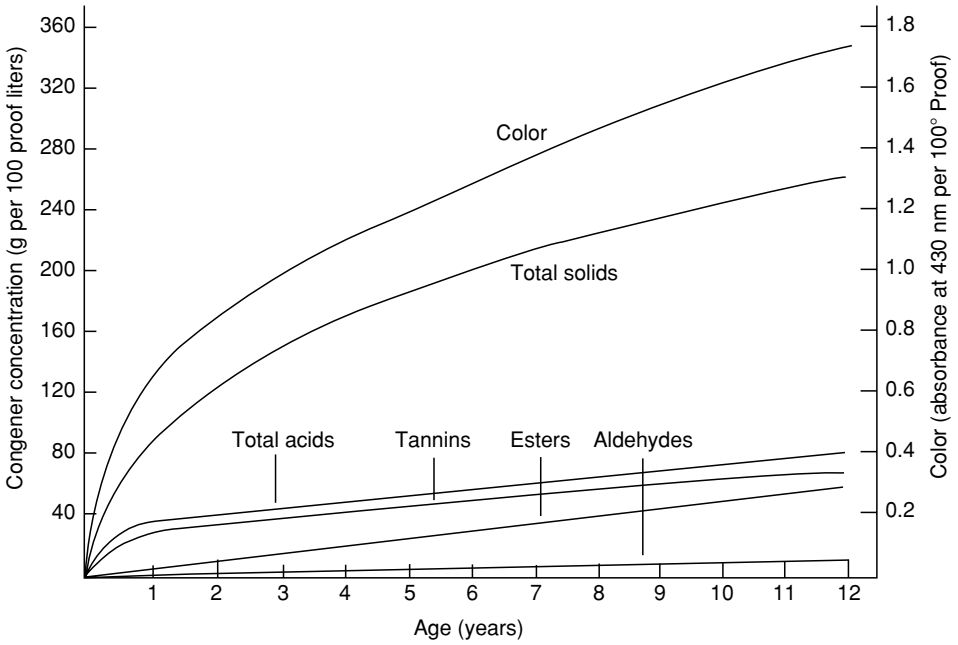

Figure 2 Congener changes during maturation of whisky. From Reazin GH (1981) Chemical mechanisms of whisky maturation. *American Journal of Enology and Viticulture* 32: 283, with permission.

from heavily peated malt retains a characteristic flavor that is strongly correlated with the total phenols content of the malt and of the whisky. However, the range and concentration of phenols present are greatly altered by mashing, fermentation, and distillation, and whether the phenols fully explain the 'peated' flavor is still uncertain. Likewise, some whiskies (especially Canadian) have a characteristic 'spicy' or 'minty' flavor, which is possibly related to the use of rye in the mash, but no compound has been identified as responsible.

Various flavors in new distillates have been correlated with sulfur compounds. Polysulfides are present in whisky at levels around and above their sensory thresholds (3–6 $ng\,g^{-1}$ for dimethyl trisulfide), and the data available suggest that these compounds are correlated with 'sulfury' and 'heavy' characters in fresh distillates. Other sulfur-containing compounds are thought to be important contributors to flavor, and their formation is dependent on the levels of precursors present in different cereals/cultivars. However, such compounds are frequently formed during fermentation and distillation and then degraded during maturation. Each stage has its own impact on the concentration in the final product, making it harder to establish their role in whisky flavor.

The products of fermentation and distillation make a substantial contribution to whisky flavor. However, these compounds, except when present in excess, appear to form what has been described as the 'integrated whisky complex,' and their individual contributions have not been determined. It does appear that relatively constant proportions of some of these compounds must be present for a recognizable 'whisky' character, though substantial changes can occur with only minor changes in flavor.

Compounds considered to be important in maturation are vanillin, eugenol, and *cis* oak lactone. These have been correlated to 'vanilla' and 'clove-like' flavors in some whiskies matured in new casks. However, cask reuse causes a general decrease in wood extractives, and their importance in whisky matured in refill casks is still uncertain.

Attempts to create a model of whisky flavor have been largely unsuccessful. Attempts were based on compositional analyses of the liquid. Research has since shown that there are complex relationships governing the release of flavor compounds from whiskies. Major factors identified include alcoholic strength and the nonvolatile content of the whisky matrix. Dilution of spirits for sensory assessment, to reduce the pungency of ethanol, can greatly alter the release of aroma compounds more soluble in ethanol than in water. Dilution also increases the capacity for matrix effects that suppress the release of ethyl esters and other aroma compounds. Matrix effects may occur through competition for release at the liquid surface or changes in solubility of flavor components. Changes in solubility can be related to the physical structuring of the ethanol–water

mixture. Differential scanning calorimetry, light scattering, and mass spectrometry have all demonstrated that physical changes occur during maturation, adding another layer of complexity to whisky flavor and reinforcing the uniqueness of each individual product.

See also: **Browning**: Nonenzymatic; **Chromatography**: High-performance Liquid Chromatography; Gas Chromatography; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Mass Spectrometry**: Principles and Instrumentation; **Phenolic Compounds**; **Sensory Evaluation**: Taste; **Tannins and Polyphenols**

Further Reading

Campbell I (ed.) (1990) *Proceedings of the Third Aviemore Conference on Malting, Brewing and Distilling*. London: Institute of Brewing.

Campbell I (ed.) (1999) *Proceedings of the Fifth Aviemore Conference on Malting, Brewing and Distilling*. London: Institute of Brewing.

Campbell I and Priest FG (eds) (1986) *Proceedings of the Second Aviemore Conference on Malting, Brewing and Distilling*. London: Institute of Brewing.

Campbell I and Priest FG (eds) (1995) *Proceedings of the Fourth Aviemore Conference on Malting, Brewing and Distilling*. London: Institute of Brewing.

Nykänen L and Lehtonen P (eds) (1984) *Flavour Research of Alcoholic Beverages*. Helsinki: Foundation for Biotechnical and Industrial Fermentation Research.

Nykänen L and Suomalainen H (1983) *Aroma of Beer, Wine and Distilled Alcoholic Beverages*. Dordrecht: Reidel.

Piggott JR (ed.) (1983) *Flavour of Distilled Beverages: Origin and Development*. Chichester, UK: Ellis Horwood.

Piggott JR and Paterson A (eds) (1989) *Distilled Beverages Flavour: Recent Developments*. Chichester, UK: Ellis Horwood.

Piggott JR, Sharp R and Duncan REB (eds) (1989) *The Science and Technology of Whiskies*. Harlow, UK: Longman.

Priest FG and Campbell I (eds) (1983) *Current Developments in Malting, Brewing and Distilling*. London: Institute of Brewing.

WILD RICE

O W Archibold, University of Saskatchewan, Saskatoon, Canada

Background

Wild rice is an aquatic grass that grows naturally in shallow lakes and slow-moving rivers (**Figure 1**) in the Great Lakes region of North America. It is North America's only native cereal, and until very recently had not been improved by breeding and selection. Lake-grown wild rice is unique amongst commercial crops in that it is grown and harvested on water and, once established, will reseed itself indefinitely. Wild rice was an important staple of the indigenous peoples of the Great Lakes region and remains part of many legends and ceremonies. The range of the species was initially extended by the early European voyagers who took wild rice with them to establish food supplies along the waterways that they traveled. Recent commercial exploitation has greatly increased production and extended the crop well beyond its native range. Most of the wild rice is now produced in artificial paddies and combines have replaced the traditional hand-harvesting methods. Vigorous marketing and product development have created global markets for this highly nutritious crop.

Taxonomy and Tradition

The contemporary common name 'wild rice' applies to annual species of *Zizania*, emergent aquatic grasses, which are native to eastern and central North America. Four forms are generally recognized on the basis

Figure 1 Shallow bays supplied with nutrients from slow-moving streams provide the ideal wild rice habitat.

of morphology, range, and habitat. The exclusively riparian species *Z. aquatica* L. includes the variety *aquatica* (Hitch.) Fassett (southern wild rice) and the variety *brevis* (Fassett) Dore (estuarine wild rice). A second species, *Z. palustris* L. is comprised of the variety *palustris* Dore (northern wild rice) and the variety *interior* (Fassett) Dore (interior wild rice). *Zizania palustris* is the most widespread species; it is also the species of commercial value, owing to its large seed size. Northern and interior wild rice are found in the upper Mississippi–Great Lakes region, extending into the boreal forest. These two varieties are sympatric over much of their ranges, although a latitudinal gradation in dominance is acknowledged.

Wild rice grows naturally in shallow lakes and streams, and is a traditional crop of the native peoples of the upper midwestern USA and central Canada. The grain was a staple diet for these seminomadic peoples, and was a cherished resource that was vigorously guarded for food and trade. French explorers noted the resemblance of wild rice to tares, a troublesome weed of wheat, and during this period it was known as 'folle avoine.' Indeed, the scientific name chosen by Linnaeus comes from the Greek *zizanion* – a weed of Mediterranean grainfields. By the early nineteenth century, English-speaking explorers called the grain 'wild rice' or 'Indian rice' because its aquatic habitat reminded them of Asiatic white rice (*Oryza* spp.). Wild rice, and the French Canadian equivalent 'riz sauvage,' remain the common names for the species. By the end of the nineteenth century, non-Native Americans started to become interested in the potential of wild rice as a commodity, first as brokers assuming control of wild rice processing and sale, then as planters and farmers, ultimately gaining control of the industry. Records of treaty negotiations show repeatedly the attempts by native Americans to explain the importance of their natural resources. Various federal and state laws have been passed which recognize the traditional value of the crop, and specify the type of equipment that can be used and the amount of wild rice that can be collected. Consequently, most of the lake-grown wild rice in Minnesota and Wisconsin is under control of Native Americans. Legislation in Canada similarly provides for aboriginal involvement in wild rice production in the pristine northern lakes. However, increased commercial demand and high commodity prices in the late 1970s resulted in the rapid establishment of artificial paddies which now account for most of the production.

Life Cycle of Wild Rice

Wild rice is an annual plant that grows each year from seed and requires approximately 120 days to mature. Under natural conditions ripe seeds shatter into the lakes. The seeds do not germinate until they have been subjected to an afterripening period of several months at low temperatures. Dormancy is broken naturally during the long boreal winters while the seeds lie embedded in the mud beneath the frozen lake. Freezing does not harm the seeds and germination begins in late spring, when water temperatures rise to about 5 °C. For the first 3 or 4 weeks of growth the young plants are under water, then long, thin leaves reach the surface as the plant enters the floating-leaf stage. By the end of June the stem and leaves begin to emerge from the water and the plant enters the aerial-leaf stage. Additional stems (tillers) may arise from the lower part of the stem and increase the density of the stands.

Flowering commences in mid-July, with separate male and female florets forming a branching panicle up to 50 cm in length, although considerable variation in panicle size occurs within and between natural stands. The flowers mature from the top down. The female (grain-producing) flowers nearest the top are the first to bloom, and it can take several weeks for all of the flowers to mature. Wild rice is wind-pollinated. The female flowers remain receptive to pollen for only a few days and so receive pollen from the male flowers on other plants; this insures cross-pollination and high genetic diversity. The grains take 4–6 weeks to mature following pollination. During this time the maturing grain passes through the soft-textured milk and dough stages until it becomes firm and greenish-black in color within the encasing hull. The first seeds usually ripen by the end of August, and must be harvested quickly because they shatter readily. However, seeds will continue to ripen over a period of several weeks until the plant is killed by frost. Seed loss, even in harvested stands, is normally sufficient to maintain stands from year to year.

Environmental Conditions in Natural Stands

Wild rice growing in unmanaged habitats is sensitive to the vagaries of the weather. In the early stages of growth, water depth and wind are critical because the young plants can be drowned or uprooted by wave action. The ideal water depth is 75–100 cm. Mature grains can be lost under windy conditions, and in some areas the harvest period is often cut short because of frosts. Wild rice production is also influenced by numerous site-specific environmental factors. Wild rice does not grow well in acidic water that is low in essential nutrients; neither will it grow well in alkaline and saline water. Total alkalinity ranging from 40 to 80 mg l^{-1} and pH values ranging from

6.9 to 7.4 are considered optimal, although the relationship between sediment chemistry and wild rice production is tenuous. Wild rice will grow on a variety of sediments, but the most productive stands are typically associated with soft organic materials that are easily permeated by the coarse fibrous roots. The state of oxidation of the sediment is an important factor affecting production, and many lakes seeded to wild rice seem incapable of nurturing the crop when E_h (redox) values are below −200 mV.

Wild rice growing in lakes does not compete well with taller, emergent perennials such as spike-rush (*Eleocharis palustris*) and cattails (*Typha latifolia*), which renew growth early in the spring. Similarly, floating-leaved aquatics such as water lilies (*Nymphaea* spp. and *Nuphar* spp.) and bur reed (*Sparganium* spp.) can also be problematic. In paddies common water-plantain (*Alisma trivale*) and various aquatic grassy weeds are of concern as no herbicides have been cleared for their control. Intraspecific competition for light and nutrients can also occur in very dense stands of wild rice and some paddies are thinned to increase in production.

Lake-grown wild rice is relished as a food source by waterfowl, muskrats, beavers, and moose, while red wing blackbirds (*Agelaius phoeniceus*) and crayfish (*Orconectes virilis*) are pests in the paddies. Insect depradation and losses from fungal, bacterial, and viral diseases have also been reported. Most damage is attributed to the riceworm (*Apamea apamiformis*) which feeds on the developing grain and the rice stalk borer (*Chilo plejadelus*) which feeds on the lower stems, making them susceptible to breakage or causing incomplete development of the grain. Reduced yields are also attributed to leaf spot, smut, and similar microbial pathogens. Ergot has long been associated with wild rice.

Agronomic Development

The natural range of wild rice has been extended through deliberate introduction, but early expansion of the industry was hampered by unreliable yields from stands in natural lake environments. The first attempts to domesticate the crop began in the 1950s when wild rice was grown in Minnesota in dyked paddies where a water depth of about 30 cm could be maintained during the growing season. A further significant development occurred in 1968 when the nonshattering Johnson cultivar was introduced. In Minnesota this has largely been replaced by earlier-maturing cultivars such as K2, Voyager, and Franklin. Shattering resistance requires the crop to be harvested by specially modified combines and grain losses have been significantly reduced. The paddies are drained prior to harvest and must be reseeded to wild rice in much the same way as other grain crops.

Agrochemicals are also used to maximize paddy production. Ammonium phosphate or urea is commonly incorporated into the upper few centimeters of the soil during seedbed preparation together with phosphate and potash; nitrogen topdress applications may be desirable as the stem begins to elongate. Fungicides may also be needed. Weeds are usually controlled by water management, while pests, such as blackbirds and crayfish, may also need to be controlled. Such agronomic developments have reduced the problem of unpredictable yields, thereby facilitating expansion of commercial markets for wild rice.

By 1978 more than 4000 ha of paddies had been established in Minnesota, and production approached 1.0 million kg of processed wild rice compared to 0.46 million kg from lakes. This year is significant because in 1978 paddy wild rice cultivation commenced in California. By 1986 production in California exceeded 4 million kg (processed) compared to 2.3 million kg in Minnesota. Paddy acreage in Minnesota is now about 6500 ha compared to about 3200 ha in California, with combined production currently about 5.0 million kg; this accounts for more than 90% of the world's supply. Approximately 0.4 million kg is produced annually in Canada, mainly in Saskatchewan, Manitoba, and Ontario. The remainder comes mostly from Idaho, Wisconsin, and Oregon. Although trials have been conducted in Finland, at present wild rice is not grown commercially outside North America.

Harvest

In traditional harvests wild rice is collected by knocking the grain into canoes using two sticks. It is then dried in partial shade to avoid mildew, picked over to remove pieces of stem and leaves, then parched by smoke drying or scorching in kettles. Each kernel is enclosed by a hull comprised of a scabrous lemma and palea which is difficult to remove until it has been heated. Once detached from the seed by treading or threshing, the chaff is removed by tossing in the air with large winnowing trays. The kernels left in the tray are then ready for cooking or storage. The traditional stages of processing have been adopted by the commercial plants that handle the massive increase in yield harvested mechanically from the paddies. Dyked and ditched fields allow water levels to be closely regulated at all stages of production, and fertilizers and pesticides are applied as necessary to optimize yields. In August the paddies are drained to allow access by modified combine harvesters. After a period of drying to help

eliminate weeds, the soils are prepared for seeding the next year's crop. Paddies in Minnesota are planted in the fall and reflooded the following spring. In California the seed must be kept in cold storage to break dormancy prior to planting in the spring.

Lake-grown wild rice differs from other crops because it is grown and harvested on water. Unlike the nonshattering paddy cultivars, lake-grown plants must be harvested regularly to minimize grain losses. Wild rice kernels mature gradually, beginning in August and continuing for 15–30 days, depending on weather conditions. About 3–6% of the potential yield matures each day. Maximum production can therefore be achieved if harvesting is repeated regularly during the ripening stage. A wild rice plant can develop many stems, particularly in shallow-water sites. These tillers are capable of producing the same-quality grain as the main stem. However, tillers generally come into flower later than the main stems. Maximum yields can only be achieved if the area is harvested several times. Normally, harvesting is repeated every 4–7 days. Most of the lake-grown wild rice now comes from Canada and is harvested by propeller-driven airboats. The most widely used design consists of a blunt-bowed, flat-bottomed aluminum hull about 3.6 m long and 1.5 m wide fitted with a collecting tray, popularly known as a speedhead (Figure 2). The sides and rear of the speedhead frame are covered with window-screening material. The speedhead has no moving parts; the kernels are dislodged through the impact of the wild rice plants against the leading edge and they fall into the tray.

Processing and Grading

Freshly harvested wild rice for processing must be delivered to the plant as soon as possible. The grade and appearance of the finished product depend a great deal on how the wild rice is handled. Careful storage is important in that it helps begin the curing process. Not all of the wild rice is at the same stage of maturity, and it is necessary to pile it in windrows to allow it to cure. The windrows are watered regularly to prevent self-heating and drying. They must also be turned daily to prevent decay. It is during this 4–10-day period that the wild rice acquires its familiar black color and flavor, and the hulls also begin to loosen. When the grain is properly cured it passes into the parching ovens. Parching removes moisture from the kernels and toughens them so they will not break as easily. Smaller batch parchers will hold up to 275 kg of wild rice, and the drum is rotated to prevent the grain from burning. Larger processing plants use continuous-flow parching. It takes about 2 h for the grain to pass through the parcher at temperatures of 145 °C in these automated systems. Wild rice can be loaded into continuous-flow parchers at a rate of up to 400 kg h^{-1}.

Figure 2 Propeller-driven airboat used to harvest lake-grown wild rice.

The primary purpose of parching is to reduce the moisture content of the kernel to about 7% without fracturing. In addition, the hull is loosened, pigmentation of the pericarp is completed, and the flavor of the grain is brought out. Steam retention in closed parchers helps to give a glassy, translucent appearance to the grain by gelatinization of the starchy endosperm. After parching, leaf and stem fragments are removed by screening, and ferrous metals are trapped by magnets on the conveyors. The hot kernels immediately pass into the hulling machines where the dry hulls are removed. The kernels are dehulled by passing through a roll huller which consists of two closely spaced, rubber-faced rollers which rotate towards each other at different speeds. This strips the hulls off and makes the kernels look black and shiny. Some processing plants use a barrel huller in which rubber cogs rotate at high speed. The hulls and the kernels fall into an aspirator that sucks off the lighter chaff. Once the hulls are removed, the wild rice is cleaned by passing over vibrating screens to separate unhulled grain, stones, and other heavy materials. Once broken kernels have been sorted, the intact kernels are graded by size on a gravity table prior to inspection and packaging. Most plants use 45-kg sacks to bag the final product for storage.

Various grades of wild rice have been proposed by the Processors Committee of the International Wild Rice Association, mainly on the basis of size. The longest kernels command the highest market prices, as they are preferred by the gourmet culinary trade. For long-grain wild rice (grades 1 and 2), kernel length is greater than 1.03 cm, medium-grain wild rice (grade 3) is 0.79 cm, short-grain wild rice (grades

4 and 5) is 0.40 cm, with smaller kernels classed as small and broken wild rice (grades 6 and 7). These grades are further defined according to the percentages of kernels of each size. Long-grain wild rice requires a minimum of 85% long grains, and a maximum of 10% medium, 7% short, and 3% broken grains. Medium-grain wild rice contains a minimum of 80% medium grains and a maximum of 10% long grains, 10% short grains, and 7% broken; short-grain wild rice contains a minimum of 80% short grains, a maximum of 10% broken, 5% long, and 10% medium kernels. The average kernel weighs approximately 50 mg.

Four quality grades of wild rice are recognized – premium, choice, good, and standard – on the basis of foreign materials such as dust and chaff, unhulled grain, or weed seeds, damaged kernels, and percent moisture. Color, taste, and odor also affect product quality. Blackish-brown kernels are favored and quality is reduced if the thin bran peels off and exposes the white endosperm. The kernels also lose some of their color when some of the bran is removed through scarification. However, this is necessary to hasten cooking time and allows wild rice to be used in blended-rice products. Improperly processed rice which contains green kernels that have not been fully fermented or which lacks the preferred toasted flavor or has a musty odor is not acceptable. Moisture content is reduced to less than 7% during processing. This facilitates long-term storage, but stress cracks which develop due to excessive heat during parching make it susceptible to damage during subsequent handling.

Wild rice is considered a specialty by retailers and is often sold in regional airports and tourist resorts in attractive transparent packages or ornamental containers. Appearance is an important aspect of quality and customer appeal and broken kernels render the product less desirable. Lake-grown wild rice is normally larger than paddy-grown cultivars and so commands higher prices, but they are indistinguishable in terms of nutritional value.

Composition and Food Quality

In the traditional native diet, wild rice was more nutritious than other naturally available fruit, grain, or animal food sources. The wild rice kernel is similar in structure to other cereals. It consists of an embryo, a starchy endosperm, and an aleurone layer encased in a tough, impermeable pericarp. Wild rice has good nutritional value; it is rich in carbohydrates and high in protein and minerals, but low in fats and oils compared to other cereals. Starch accounts for approximately 75% of the dry weight of the kernels, protein averages about 13%, dietary fiber 7%, sugars 2%, minerals 2%, and oils and fats 1% (Table 1). Many amino acids are present at concentrations that are adequate to meet daily human requirements. Lysine, threonine, and methionine, important indicators of nutritional quality of cereals, are typically present at higher concentration in wild rice than in other grains (Table 2). Very little amino acid is lost in processing, whereas essential minerals (Table 3) are mainly present in the pericarp, and a considerable reduction occurs when wild rice is scarified or polished. Vitamin A is absent from processed wild rice, but it is a good source of dietary B-vitamins, especially niacin and thiamin (Table 4). Linoleic and linolenic acids are comparatively abundant and together make up about 70% of the total fatty acid content in wild rice. Although this is desirable from a nutritional standpoint, it can lead to problems of rancid odor if oxidation occurs because of improper storage. However, properly cured and stored wild rice remains wholesome for several years.

Wild rice has also attracted the interest of the food-processing industry because it contains the antioxidant phytate (myoinositol hexaphosphate). Additions of 15% by weight of cooked wild rice to beef patties has been shown to reduce rancidity by reducing the production of thiobarbituric reactive substances by as much as 50%. This has a significant potential in microwavable precooked meat products, especially as studies have shown consumer taste preferences for meat patties blended with wild rice. In addition, nutritive properties are enhanced, cholesterol and fat percentages are reduced, and cooking yields are augmented. Similar antioxidant properties were found with extracts from wild rice hulls; these are presently discarded at the processing plants. Wild rice starch has been shown to have a low degree of retrogradation and this may be useful in the paper, textile, and adhesive industries where changes in prepared batches of pastes are undesirable.

Table 1 Typical composition of wild rice and other cereals (g 100 g^{-1} of dry product)

	Wild rice	*Brown rice*	*Yellow corn*	*Hard wheat*	*Oat groat*
Starch	74.0	78.0	71.5	66.5	62.0
Protein (N × 5.7)	13.5	8.7	9.0	14.5	15.5
Dietary fiber	6.8	5.3	9.5	11.5	11.0
Sugars	1.7	1.3	2.3	1.7	1.4
Oils and fats	0.8	2.6	4.7	1.8	6.5
Ash	1.8	1.5	1.5	2.0	2.0

After Anderson RA (1976) Wild rice: nutritional review. *Cereal Chemistry* 53: 949–955, with permission.

Table 2 Composition of essential amino acids for human nutrition in wild rice and other cereals (g 100 g^{-1} of protein)

	Wild rice	*Brown rice*	*Yellow corn*	*Hard wheat*	*Oat groat*	*Human requirements*
Lysine	4.5	3.9	2.5	2.6	4.1	5.8
Methionine	2.8	2.1	1.8	1.5	2.2	2.5
Tryptophan	1.6	1.5	0.9	1.2	1.6	1.1
Threonine	3.4	3.7	3.8	2.8	3.4	3.4
Phenylalanine	5.1	5.0	5.0	4.7	5.5	6.3
Histidine	2.8	2.6	2.4	2.2	2.3	1.9
Isoleucine	4.4	4.0	3.9	3.7	4.0	2.8
Leucine	7.4	8.3	11.4	6.7	7.7	6.6
Valine	5.9	5.9	4.7	4.5	5.6	3.5

After Anderson RA (1976) Wild rice: nutritional review. *Cereal Chemistry* 53: 949–955, with permission.

Table 3 Minerals in wild rice and other cereals (mg 100 g^{-1})

	Wild rice	*Brown rice*	*Polished white rice*	*Oats*	*Winter Wheat*	*Corn*
Calcium	17–22	32	24	53	46	22
Iron	4.2	1.6	0.8	4.5	3.4	2.1
Magnesium	80–161		28	144	160	147
Potassium	55–344	214	92	352	370	284
Phosphorus	298–400	221	94	405	354	268
Zinc	3.3–6.5		1.3	3.4	3.4	2.1

After Anderson RA (1976) Wild rice: nutritional review. *Cereal Chemistry* 53: 949–955, with permission.

Table 4 Vitamin content of wild rice and other cereals (mg 100 g^{-1})

	Wild rice	*Brown rice*	*Polished white rice*	*Oats*	*Winter wheat*	*Corn*
Thiamin	0.45	0.34	0.07	0.60	0.52	0.37
Riboflavin	0.63	0.05	0.03	0.14	0.12	0.12
Niacin	6.20	4.70	1.60	1.00	4.30	2.20

After Anderson RA (1976) Wild rice: nutritional review. *Cereal Chemistry* 53: 949–955, with permission.

Contamination of the wild rice habitat has prompted concern about elevated levels of toxic metals in the kernels which might pose a health hazard. Copper concentrations as high as 14.4 μg g^{-1} and lead and cadmium levels as high as 6.2 and 6.7 μg g^{-1} respectively have been reported in some samples of parched and dehulled wild rice. Locally elevated metal concentrations have been attributed to atmospheric deposition from smelters, and it has also been reported that lead derived from shotgun shells can be taken up by wild rice. Concentrations of lead (0.5–11.5 g 100 g^{-1} dry weight), cadmium (1.0–10.2 g 100 g^{-1} dry weight), and arsenic (0.6–14.2 g 100 g^{-1} dry weight) have been found in wild rice sold in the USA.

Concluding Remarks

Interest in wild rice from companies like Uncle Ben's and General Foods has spurred the development of the industry. Blended products containing 12–18% wild rice mixed with long-grain white or brown rice have increased on the supermarket shelves. The smaller grain sizes developed by breeders have eliminated the problems of lengthy cooking times and, coincidentally, increased the number of grains per package, thereby making it more attractive to consumers. With increasing acceptance of wild rice, a variety of products began to emerge, including stuffings, frozen casseroles, bread and pancake mixes, and puffed rice snacks, as well as a myriad of cookbooks and recipes. However, natural lake-grown wild rice continues to be the preferred choice of the gourmet and health-food markets. In an effort to maintain some control of the market, native growers were successful in getting legislation requiring that paddy-grown rice be so labeled, thus distinguishing it from the traditional product. Attempts by Native American organizations to compete or cooperate with the larger companies have not been overly successful. In effect, the white-monopolized industry has caused a reversion of wild rice to its traditional role as a subsistence item in the Native American economy.

See also: **Rice**

Further Reading

Aiken SG, Lee PF, Punter D and Stewart JM (1988) *Wild Rice in Canada*. Toronto: NC Press.

Anderson RA (1976) Wild rice: nutritional review. *Cereal Chemistry* 53: 949–955.

Fyles F (1915) A preliminary study of ergot of wild rice. *Phytopathology* 5: 186–192.

Hoover R, Sailaja Y and Sosulski FW (1996) Characterization of starches from wild rice and long grain brown rice. *Food Research International* 29: 99–107.

Minerich PL, Addis PB, Eppley RJ and Bingham C (1991) Properties of wild rice/ground beef mixtures. *Journal of Food Science* 5: 1154–1157.

Nriagu JO and Lin T-S (1995) Trace metals in wild rice sold in the United States. *The Science of the Total Environment* 172: 223–228.

Oelke EA, Grava J, Noetzel D, *et al.* (1982) *Wild Rice Production in Minnesota*. Extension Bulletin 464. Minnesota: University of Minnesota Agricultural Extension Service.

Oelke EA, Porter RA, Grombacher AW and Addis PB (1997) Wild rice – new interest in an old crop. *Cereal Foods World* 42: 234–247.

Painchaud DL and Archibold OW (1990) The effect of sediment chemistry on the successful establishment of wild rice (*Zizania palustris* L.) in northern Saskatchewan waterbodies. *Plant and Soil* 129: 109–116.

Peterson AG, Noetzel DM, Sargent JE, Hanson PE, Johnson CB and Soemawinata AT (1981) *Insects of Wild Rice in Minnesota*. Misc. Rep. 157. Minnesota: University of Minnesota Agriculture Experimental Station.

Vennum T (1988) *Wild Rice and the Ojibway People*. St Paul: Minnesota Historical Society Press.

Wu K, Zhang W, Addis PB, Epley RJ, Salih AM and Lehrfeld J (1994) Antioxidant properties of wild rice. *Journal of Agriculture and Food Chemistry* 42: 34–37.

WINES

Contents

Types of Table Wine

R S Jackson, Bic, QC, Canada

Background

Classification is a common human trait. Without it, precise language would be nonexistent. Unfortunately, the natural desire to group wines into logical categories is fraught with difficulty. Wine is an eclectic collection of grape-based beverages that reflects local conditions and grape cultivars, as well as adaptions to changes in cultural attitudes and technological advances over several millennia. There is no evolutionary progression to connect these diverse products. In the absence of an inherently rational basis for organizing wines, they have generally been grouped according to those features that tend to be relatively stable and easily detected (e.g., color, sparkle, alcohol content), or origin (e.g., geographic or varietal).

Classification based on these features has potential value because it supplies a crude idea of the sensory attributes of the wine. For example, red wines are typically more flavorful, drier, and astringent than white wines, which generally are milder tasting, and may come in sweet and semisweet styles (red wines seldom do). The obvious presence of an effervescence has become culturally associated with celebration. The mention of a particular grape cultivar or region can suggest, to those possessing sufficient experience, the likely presence of a particular set of flavor and taste characteristics.

The problem with empirical systems, such as those noted above, is that they depend heavily on experience most consumers lack. Although people readily distinguish color categories and may recognize geographic names, they seldom have a clear concept of their sensory significance. Despite the need for a clear, rational classification of table wines, the diversity of wine is too great to permit presentation of a readily understandable, coherent categorization. What is feasible is an organization based on those features that generate the diversity found in table wines.

The most critical element influencing the sensory features of a wine is the winemaker. Starting with red grapes, a winemaker can make an almost infinite spectrum of wines – from almost colorless, through pink, to light red, to dark red–purple, with equivalent variations in flavor intensity. In addition, the winemaker may direct production toward sparkling or sweet versions of any of the former styles, or choose to produce sherry- or port-like wines. Because the effect of the winemaker can be so marked, and unpredictable, it is typically neglected or submerged under regional or stylistic categories. This is possible because tradition, if not legal constraints, often limit winemaker freedom in the style(s) produced.

In Europe, most regions have tended to specialize in the production of one or very few wine styles, made from one or a few, typically indigenous, grape cultivars. This feature has permitted the association of

particular flavor traits with wines from particular regions. However, cultivars generally produce similar flavor attributes wherever they are grown. For example, 'Cabernet Sauvignon' and 'Chardonnay' often produce wines undistinguishable from those produced in their homelands of Bordeaux and Burgundy, respectively. Thus, the predictive value of varietal origin on sensory attributes depends more on the skill of the winemaker, and the maturity and health of the grapes, than geographic origin. Although winemaker skill is difficult to quantify, grape maturity and health are often quantified for consumers in vintage ranking. The central roles of grape variety and production style are outlined in Figure 1 and discussed below.

Varietal Origin

In a fundamental sense, wine taxonomy should be based on the cultivar or cultivars used in their production. Varietal origin sets the outer limits of the wine's sensory attributes. Unfortunately, most grape cultivars do not, or are not known to, possess a readily distinguishable varietal aroma. Thus, varietal origin is of significance only for those cultivars that possess unique aromatic potential. Several of the better known of these are briefly noted below.

Red Cultivars

'Cabernet Sauvignon' is the most well-known member of the Carmenet family of grape cultivars. Its renown comes from its involvement in most Bordeaux wines (and equivalents). Other members of the family include 'Merlot' and 'Malbec.' Their inclusion in Bordeaux blends moderates the tannin content donated primarily by 'Cabernet Sauvignon.' The tendency of 'Merlot' to mature more quickly has made it a popular substitute for 'Cabernet Sauvignon.' Under optimal conditions, 'Cabernet Sauvignon' yields a fragrant wine possessing a blackcurrant aroma. Under less favorable conditions, it generates a bell-pepper aroma. 'Cabernet Sauvignon' probably is the result of an accidental crossing between grapes related to, if not identical with, 'Cabernet Franc' and 'Sauvignon blanc.'

'Gamay noir à jus blanc' is the primary, white-juiced Gamay cultivar. Its reputation has risen in association with the increased popularity of Beaujolais wines. Crushed and fermented by standard procedures, 'Gamay' produces a light red wine with few distinctive characteristics. However, when processed by carbonic maceration, it yields a distinctively fruity wine. Most of these features come from the grape-berry fermentation phase of carbonic maceration.

'Nebbiolo' is generally acknowledged as producing the most highly regarded red wine in north-western Italy. With traditional vinification, it produces a wine high in acid and tannin content that requires years to mellow. The color has a tendency to oxidize rapidly. Common aroma descriptors include tar, violets, and truffles.

'Pinot noir' is the most famous Noirien grape variety. It is particularly environmentally sensitive, producing its 'typical' fragrance (beets, peppermint, or cherries) only occasionally. The cultivar exists as a varied collection of distinctive clones. Usually, the more prostrate, lower-yielding clones produce more flavorful wines. The upright, higher-yielding clones are more suited to the production of rosé and sparkling wines. The South African cultivar, 'Pinotage,' is a cross between 'Pinot noir' and 'Cinsaut.'

'Sangiovese' is an ancient cultivar consisting of many distinctive clones, grown extensively throughout central Italy. It is most well known for the light- to full-bodied wines from Chianti, but also produces many fine red wines elsewhere in Italy. Under optimal conditions, it yields a wine possessing an aroma reminiscent of cherries, violets, and licorice. 'Sangiovese' is also grown under local synonyms, such as 'Brunello' and 'Prugnolo,' used in producing Brunello di Montalcino and Vino Nobile di Montepulciano wines, respectively.

'Shiraz' ('Syrah' in France) has become famous for yielding a deep red tannic wine with long aging potential in Australia. This has helped 'Syrah' regain the prominence it once held in the Rhône Valley of France. Its wines are peppery with aspects reminiscent of violets, raspberries, and currants.

'Tempranillo' is probably the finest Spanish red-grape variety. Under favorable conditions, it yields a delicate, subtle wine that ages well. It is the most important red cultivar in Rioja. Outside Spain, it is primarily grown in Argentina. In California, it usually goes under the name 'Valdepeñas.' 'Tempranillo' generates an aroma distinguished by a complex, berry-jam fragrance, with nuances of citrus and incense.

'Zinfandel' is extensively grown in California. Its European counterpart is probably the Italian variety 'Primitivo,' or its Hungarian equivalent, 'Plavac.' 'Zinfandel' is used to produce a wide range of wines, from ports to light blush wines. In rosé versions, it shows a raspberry fragrance, whereas full-bodied red wines possess rich berry flavors.

White Cultivars

'Chardonnay' is the most famous white member of the Noirien family of cultivars. It not only produces

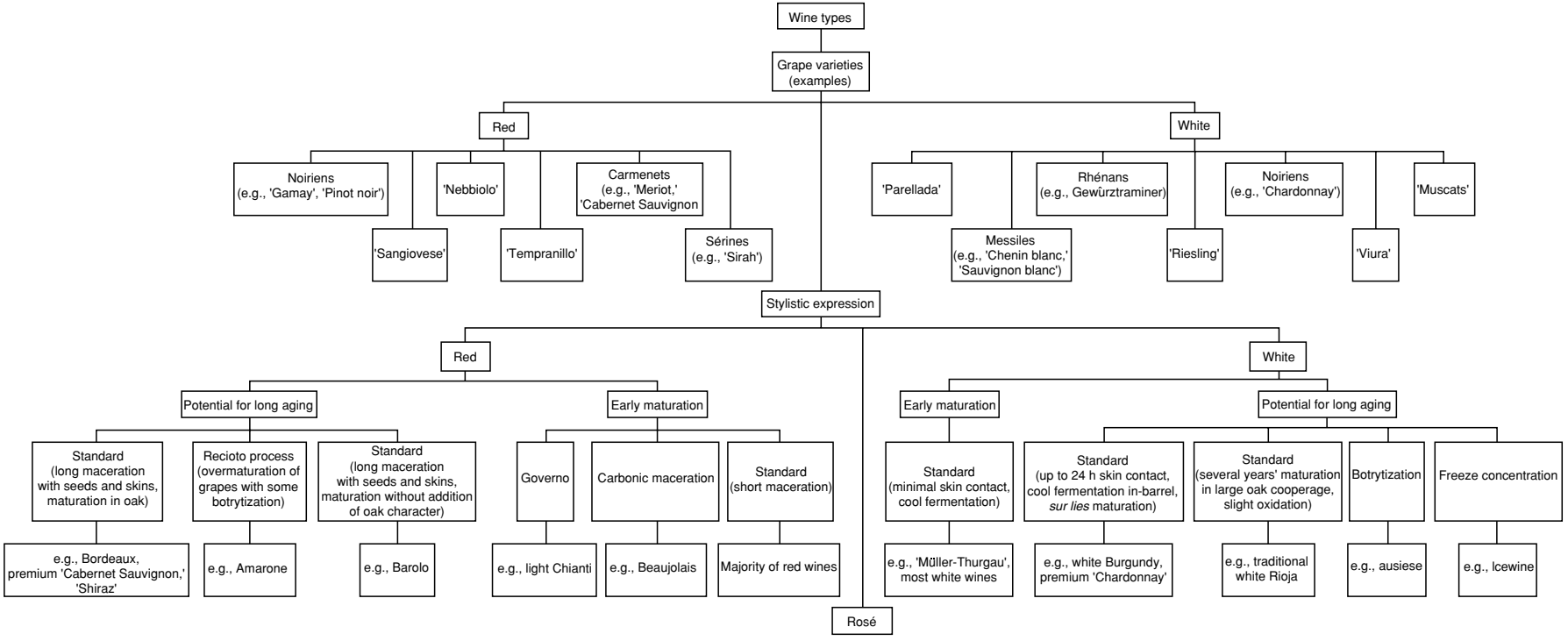

Figure 1 Major sources of the different types of table wine.

wines with an appealing fruit fragrance, but also tends to do well world-wide. In addition to producing fine table wines, it yields one of the finest sparkling wines (champagne). Under optimal conditions, the wine develops aspects reminiscent of various fruits, including apple, peach, and melon.

'Müller-Thurgau' is possibly the most well-known modern *V. vinifera* cultivar, constituting about 30% of German hectarage. It is most likely a crossing between 'Riesling' and 'Chasselas de Courtillier.' Its mild acidity, subtle fruity fragrance, and early maturity are ideal for producing light wines in cooler climates.

'Muscat blanc' is one of many members of the Muscat family of cultivars. Their aroma is so distinctive that it is described in terms of the cultivar name – muscaty. Because of the intense flavor, slight bitterness, and tendency to oxidize, Muscat grapes have been used most commonly in producing dessert wines. The reduced bitterness and lower oxidation susceptibility of the new Muscat cultivar, 'Symphony,' permit the production of dryer wines with a better aging potential. 'Moscato bianco' is the primary variety used in the flourishing sparkling wine industry in Asti, Italy.

'Parellada' is a variety distinctive to the Catalonian region of Spain. It produces an aroma that is apple- to citrus-like in character, occasionally showing hints of licorice or cinnamon.

'Pinot gris' and 'Pinot blanc' are color mutants of 'Pinot noir.' They are both cultivated throughout the cooler climatic regions of Europe for the production of dry, botrytized, and sparkling wines. 'Pinot gris' typically yields subtly fragrant wines with aspects of passion fruit, whereas 'Pinot blanc' is more fruity, with suggestions of hard cheese.

'Riesling' is without doubt Germany's most highly respected grape variety. It can produce fresh, aromatic, well-aged wines, which can vary from dry to sweet. Its complex floral aroma, commonly reminiscent of roses and pine, has made it acclaimed throughout central Europe and much of the world. Outside Germany, its largest plantings are in California and Australia.

'Sauvignon blanc' is one of the main white Carmenet varieties in Bordeaux, and the principal white cultivar in the upper Loire Valley. It has become popular in California and New Zealand in recent years. Often, its aroma shows elements of green peppers, as well as a herbaceous aspect, especially in cooler climates. Better clones possess a subtly floral character.

'Traminer' is a distinctively aromatic cultivar grown throughout the cooler regions of Europe and much of the world. Although possessing a rose blush in the skin, it produces a white wine. 'Traminer' is fermented to produce both dry and sweet styles, depending on regional preferences. Intensely fragrant ('Gewürztraminer') clones generally possess an aroma resembling lichi fruit.

'Viognier' has become popular in the USA and Australia, following its languishing in the Rhône Valley since the phylloxera epidemic of the late 1800s. The wine matures quickly and is characterized by the development a fragrant peach to apricot aroma.

'Viura' is the main white variety cultivated in Rioja. In cooler areas, it produces a fresh wine possessing a subtle floral aroma with aspects of citron. After extended aging in large wood cooperage, it develops a golden color and rich butterscotch to banana fragrance that characterizes the traditional white wines of Rioja.

Stylistic Origin

Although varietal attributes set the outer limits of a wine's characteristics, their expression may be influenced by growing conditions. Equally, the way in which varietal attributes exhibit themselves depends on grape health and maturity, how the grapes are transformed into wine, and how the wine is subsequently stored. Thus, after varietal origin, vintage conditions and production style are the most important features establishing the features that characterize the wine.

Because wine styles have arisen independently, there is no obvious or evolutionary logic by which wine can be classified. For simplicity, they can be organized temporally, relative to whether they occur before, during, or after grape fermentation.

Type Reflecting Procedures Before Fermentation

One of the oldest processes that may be involved in wine production is carbonic maceration. This occurs whenever whole grapes collect in large piles (in the absence of oxygen). Before the mid-1800s, carbonic maceration was involved in the production of essentially all wines. Inefficient manual crushing resulted in many whole grapes remaining after crushing. Thus, when yeast fermentation consumed the oxygen dissolved in the freed juice, whole grapes began a grape-cell fermentation (carbonic maceration). Carbonic maceration generates a unique set of aromatic compounds that give the wine a distinct, raspberry–kirsch fragrance. When the majority of the grapes undergo carbonic maceration, as in the production of Beaujolais wines, the fruity aromas generated often mask the presence of varietal aromas. Only as the wine ages, and grape-fermentation odors fade, may a varietal character express itself. Most Beaujolais wines do

not age well because the cultivar from which they are made ('Gamay noir') possesses little varietal aroma to replace the fading carbonic maceration fragrance.

Another, technique that can fundamentally modify a wine's sensory attributes is over-maturation. Over-maturation generally occurs by leaving the grapes on the vine past maturity (late-harvest). For white wines, it is used to produce a more flavorful, usually sweet wine. Its most classic expression is in the production of botrytized wines, such as the ausleses of Germany, the Tokaji of Hungary, and the Sauternes of France. During over-maturation, infection by the fungus *Botrytis cinerea* concentrates the juice, reducing the varietal aroma, and generating its own unique apricot, honey-like fragrance. Late-harvest wines not involving the intervention of *Botrytis cinerea* generally have only an enhanced over-matured varietal aroma. Ice wines (*eisweins*) are a distinctive category of late-harvest wines. They undergo a particularly long (up to several months) over-maturation on the vine, until frigid (−7 to −8 °C) conditions freeze the grapes. The grapes are harvested and crushed while frozen to extract only the highly concentrated juice (most of the water in the grapes remaining in the crusher as ice).

Although more commonly used in the production of white wines, over-maturation is occasionally employed in the making of particular red wines. For example, in the production of Chianti, a portion of the harvest may be set aside to slowly dry and over-mature. Most of the crop is crushed and fermented into wine following standard procedures. A month or so later, the over-matured portion of the crop is crushed, begins fermentation, and is added to the majority of the harvest that usually has just completed fermentation. The effect of the second fermentation is to produce a wine that is fruitier and matures earlier than otherwise. This process, called the *governo* process, is less popular than in the past because it does not promote aging potential – the touchstone of quality for most connoisseurs.

In Veneto and, to a lesser extent, in Lombardy, a technique called the *recioto* (*appassimento*) process involves the exclusive use of over-matured grapes. Especially selected mature clusters are laid on trays in thin layers, and the trays stacked in well-ventilated, unheated, storage sheds. The grapes are left to dry and over-mature slowly for several months. Frequently, a portion of the grapes undergo a slow transformation under the action of *Botrytis cinerea*. This results in a marked increase in the concentration of glycerol and sugars, as well as the oxidation of grape phenolics. The latter appears to generate several volatile phenols that give the wine its distinctive tulip–daffodil pungency. After several months, the grapes are crushed and fermented to produce dry, semisweet, or sparkling wines. The most commonly exported version is the dry Recioto della Valpolicella Amarone.

Type Reflecting Conditions Immediately Following Crushing

One of the most fundamental decisions in winemaking involves the conditions immediately following crushing. For white wines, the juice is separated from the seeds and skins shortly after crushing and before fermentation begins. In contrast, the juice is left in contact with the seeds and skins throughout most of the fermentation period for the production of red wine.

Separation of the juice from the seeds and skins, directly after crushing, limits the uptake of most grape flavorants (aroma), and yields an almost colorless wine. In contrast, leaving the juice in contact with the seeds and skins progressively permits the extraction of flavorants and colorants, notably from the skins. For white wines, skin contact (maceration) is short (often only a few hours) to minimize the uptake of bitter-tasting phenolics.

Type Reflecting Conditions During Fermentation

The duration of joint fermentation and maceration significantly affects a red wine's characteristics. Gently pressing to remove the seeds and skins immediately after crushing may result in the production of a white wine. Maceration with the skins for the first 24–48 h usually generates a rosé wine. Standard red wines typically undergo maceration for 4–5 days, during which fermentation becomes fully active. Depending on the efficiency of the juice–skin contact, moderate to long aging, deeply pigmented wines are produced. Traditionally, wines designed for long aging have been left on the seeds and skins for upwards of 2–3 weeks. This may generate flavorful wines, but prolonged maturation is required before their high tannin content 'softens.' Newer techniques optimize flavor and pigment uptake without the extraction of excessive amounts of bitter, astringent tannins. Most varietal flavors are dissolved from the skins within about 3–5 days, whereas tannins are removed more slowly.

Another condition that fundamentally affects the characteristics of a wine is fermentation temperature. White wines are typically fermented at cooler temperatures than red wines. Cool temperatures (about 15 °C) enhance the production and retention of 'fruit' esters synthesized by yeasts, giving the wine a fruity fragrance. This is particularly important for white wines produced from cultivars with little distinctive aroma. Wines from aromatically unique varieties may be fermented at somewhat higher temperatures to

accentuate the aroma versus general fruitiness. Red wines often are fermented at warm (about 22–25 °C) temperatures to facilitate yeast growth in the presence of grape phenolics, and to promote pigment and flavor extraction from the skins.

Several other fermentation conditions influence the stylistic expression of a wine. However, most are so extensively, but variably, used across varietal and regional styles as to negate their value in characterizing wine types. Examples are the use of different yeast strains, the employment of malolactic fermentation, and the use of in-barrel vs. tank fermentation.

Type Reflecting Conditions During Maturation

For the majority of wines designed for early consumption, maturation prior to bottling typically occurs in stainless steel tanks, or other inert cooperage. This avoids the uptake of additional tannins from wood cooperage that can require extensive aging to mellow. It also avoids the extraction of oak flavors that might mask the wine's mild fragrance.

Wines benefiting from extended aging, or possessing a distinct varietal aroma, are often aged in oak cooperage. For white wines, this is seldom more than for 6–12 months. This may or may not be combined with *sur lies* maturation (3–6 months' contact with the dead and dying yeast cells following the end of fermentation). The procedure is often thought to enhance the wine's flavor complexity. For red wines, maturation often occurs in new or used oak barrels (approximately 250-l capacity), for a period of up to 2 years. An older procedure, less commonly used today, is maturation in large (1000–10 000 l) oak tanks. The latter supplies less oak character to the wine, while providing slower but increased moderate oxidation. In-barrel maturation accentuates the varietal character of several major grape cultivars, such as 'Cabernet Sauvignon,' whereas tank maturation tends to moderate varietal distinctiveness.

Geographic Expression

Although it is often implied that geographic wine expressions are clearly recognizable, this is rarely the case. Even experienced wine tasters often fail to identify the varietal origin of wines, let alone their regional variations. If one knows that the wines come from a single region, are made from the same cultivars, and have been produced using similar techniques, it may be possible, under ideal conditions, to differentiate subregional expressions. However, if the varietal wines come from several different countries, it is unlikely that the individual geographic origins of the wines can be consistently recognizable. Even the more distinctive varietal feature of many New World wines is not dependable. Yearly variations in weather conditions can produce greater fluctuations in wine attributes than climatic differences between regions.

The prominence given to geographic origin in most wine classifications comes from the familiarity that consumers have, or can obtain, with regional names. As already noted, in several European regions, legislation limits the use of grape cultivars and wine production techniques. Although this helps maintain stylistic consistency, it limits innovation.

Appellation Control (AC) Laws

The frequent grouping of wine by geographic origin, and the economic value of clear distinction (if only by name), have combined to promote the establishment of appellation control laws. Although these laws vary considerably from country to country, and even from region to region, they always act as a certification of geographic origin. They may also stipulate varietal composition, conditions of grape culture, as well as fermentation and wine-processing procedures. Although important, what is of greater general interest is how these laws reflect cultural views on the origins of wine quality.

In most established European regions, it is often assumed that historically accepted practices and geographic features define wine quality. However, they have limited improvement, leading to less regulated regions rapidly reaching and surpassing the quality of established regions. Furthermore, weather conditions can be so variable as to nullify the law's attempt to 'guarantee' wine quality. Finally, in the absence of uniform yearly assessments of wine quality, AC laws effectively assure only geographic authenticity.

In Europe, appellation control laws fall into two philosophically different groups. One ascribes greatest importance to geographic features (e.g., France, Italy, and Spain), whereas the other accentuates the importance of grape maturity (e.g., Germany and Austria).

In the French version, regions are ranked relative to geography (incorporating traditional winemaking and grape-growing procedures). Regions are classed according to historic perceptions of their winemaking quality – premium (AC), superior (VDQS, vins délimités de qualité supérieure), good (vins de pays), and basic (vins de table). AC regions are often subdivided, with smaller regions viewed as producing more distinctive (and better) wines than those from larger regions. This subdivision on geographic origin may be taken down to single vineyards. In most cases, though, it goes down only to single parishes or townships. As the area becomes smaller, production regulations often become more precise and restrictive.

In addition to the classification of regions, individual vineyards (or wineries) may be hierarchically ordered. The most well-known example is the 1855 grouping of some Bordeaux vineyards. This arrangement is called the cru classé system. As in AC designations, the cru classé system is based on the historic potential of the vineyards to produce quality wine. Thus, yearly sensory assessments of wine quality are not considered necessary for use of the designations.

In Italy, most wines are classified heirarchically into DOCG (denominazione di origine controllata e garantita), DOC (denominazione di origine controllata), IGP (indicazione geografica tipica), and Vino da Tavola categories. One of the unique features of the Italian AC system is the potential pyramidal interconnection. Wine production exceeding the quantities stipulated in the regulations progressively demotes the wine to lower and lower categories. Use of the highest (DOCG) designation requires that the wine pass yearly governmental sensory evaluation tests.

In the Germanic concept of quality, greatest importance is placed on grape maturity. Thus, it has a prominent place, along with geographic origin, in the naming of wines. Correspondingly, annual achievement of a minimum score on a governmental sensory evaluation test is required before permission is granted to use any of the QbA (qualitätswein bestimmter anbaugebietes) or QmP (qualitätswein mit prädikat) quality designations. Historical reputation has no influence on the quality or geographic designations in the German system.

Most New World countries are developing AC regulations. In most instances, they do not have the same significance as their European equivalents. This results from the diversity of wine styles produced within the designated regions. Thus, New World appellation laws generally certify only geographic origin. This, however, achieves the primary advantage of appellation designations – nominal differentiation of regional wines in the marketplace.

See also: **Grapes**; **Wines**: Production of Table Wines; Dietary Importance; **Yeasts**

Further Reading

Amerine MA and Singleton VL (1977) *Wine – An Introduction*, 2nd edn. Berkeley, CA: University of California Press.

Anonymous (1987) *Symposium 'Les Appellations d'Origine Historiques.'* Paris: Office International de la Vigne et du Vin.

Clarke O (1995) *Oz Clarke's Wine Atlas*. Boston, MA: Little Brown.

Jackson RS (2000) *Wine Science: Principles, Practice, Perception*, 2nd edn. San Diego, CA: Academic Press.

Johnson H and Halliday J (1992) *The Vintner's Art*. New York: Simon & Schuster.

Johnson H and Robinson J (2001) *The World Atlas of Wine,* 5th edn. London: Mitchell Beazley.

Production of Table Wines

K Wucherpfennig, Baumann-Gonser-Foundation, Wiesbaden, Germany

Introduction

Wine is legally defined as the product that is recovered exclusively by complete or partial alcoholic fermentation of the fresh, treated grapes or the grape must. If alcohol is added to the wine, e.g., in the form of distillate of wine, dessert wine is obtained with an alcohol content between 15 and 22% by volume (sherry, port, etc.). If a second fermentation is carried out without the escape of the formed carbon dioxide, champagne or sekt is produced. If carbon dioxide is artificially introduced, pearl wine or simple sparkling wine is obtained. (*See* **Port**: The Product and its Manufacture; **Sherry**: The Product and its Manufacture; **Wines**: Production of Sparkling Wines.)

Wines, dessert wines, pearl wines, and sparkling wines can be further processed to vinous beverages. These are products in which the proportion of wine is in excess of 50%. The best known products of this group are the aromatized wines such as vermouth. (*See* **Vermouth**.)

If fruits or other sugar- or starch-containing raw materials are processed to alcoholic beverages, the end product is not wine within the meaning of the pertinent wine legislation, but a beverage similar to wine. These include the fruit wines, the main representative being cider.

Table, land, and quality wines made from grapes make up by far the largest category among the different sorts of wine. For this reason, the main focus will be on their method of production. The production of, for example, fruit wines is very similar, however.

Viticulture

The vine belongs to the genus *Vitis*, of which 50–60 species are known. One of these is the European vine, *Vitis vinifera*, from which a large number of varieties stem. Over the centuries, the variety best suited to the growing area has been selected, i.e., the variety that offers the highest quality together with adequate potential crop and resistance to diseases of the vine.

Vine Varieties

As the *Vitis vinifera* varieties used for the production of high-quality wines are susceptible to attack by the phylloxera, they can only be grown grafted on to the stock of American vine. World-wide, a group of 15 vine varieties has established itself for the production of red wines, and a group of 20 vine varieties has proved suitable for the production of white wines. Scientific breeding of vines has been practiced since 1900 with the aim of improving the performance of the vine with respect to yield and quality.

Methods of Cultivation

In the cultivation of vines, certain planting systems and forms of training are applied with a view to attaining regular and high-quality crops. It must be insured that the sprouting flower buds with two bunches per shoot on average are exposed to as much sun as possible, and that every shoot has about 10–15 leaves.

Maturity and Vintage

In the course of several growth stages, the setting gives rise to the grape consisting of the skin (exocarp), the pulp (mesocarp), and two to four seeds. The largest part of the berry is the succulent pulp. The grape increases most in weight during the final growth stage; sugar (glucose and fructose) is stored, and water is absorbed, while the concentration of total acid and malic acid decreases substantially. The grapes can be considered physiologically ripe when a sugar content of 75–80° Oechsle (18.20–20.47° Brix) is reached, corresponding to an alcohol content of 9% by volume. From this point onwards, the stems develop into mature cane, and nutritious matter from the soil and the sugar formed in the leaves can no longer penetrate into the grapes. If the grapes have reached the state of full ripeness and are not harvested immediately, their constituents are concentrated owing to evaporation of water. The grapes become overripe and soft, as a result of which the sugar content increases. Further concentration can take place if the grapes are attacked by the fungus, *Botrytis cinerea*. If this happens at an early stage, it causes 'grey rot,' which is responsible for a deterioration in quality; however, if the disease strikes at a late stage, the so-called 'noble rot' develops, which is desirable in special areas, e.g., for the production of Sauterne. (*See* **Ripening of Fruit**.)

Choosing the right time for harvesting is of particular significance, as the current consensus of opinion is that the volume of alcohol to be expected from the sugar content in the grapes determines the quality The more sugar the grapes contain when harvested, the higher the quality rating of the wine produced.

The vintage has been carried out manually for centuries. However, as an expense of work of 250 h ha^{-1} is required, there is an increasing tendency to use machines that straddle the rows of vines and shake off the bunches of grapes. The yield varies between 60 and 150 hl ha^{-1}, depending on the vine variety, climate, soil, and care.

Processing of White Wine

The flow chart of white wine processing is shown in Figure 1. After harvesting, the grapes should preferably be transported to the press room in an uncrushed state to prevent undesirable excessive extraction of tannins from the stems and seeds (1). There, they are weighed and crushed. The must weight is also determined at this point, which is an important criterion for the quality of the subsequent wine.

Pressing

The grapes are often stemmed (2) to produce wines with a low tannin content, although this is not absolutely essential. The crushed grapes are then transported direct into the press or into storage tanks or straining bins (3). (*See* **Tannins and Polyphenols**.)

The untreated or treated crushed grapes can be pressed in discontinuously operating horizontal presses. These comprise a perforated press cage in which one or two mechanically or hydraulically actuated pressure plates gradually reduce the space available to the crushed grapes, causing the must to be pressed through the fine slots (4).

In the recently introduced pneumatic presses, the crushed grapes are pressed by air pressure acting on a plastic membrane against a perforated metal plate and are, therefore, pressed extremely gently. Good must and wine qualities can be attained using continuous, large-diameter screw presses that rotate at a low speed.

Freshly pressed musts still contain insoluble parts of the cellular tissue, protein–tannin compounds, microorganisms, crystal compounds, e.g. tartar, dirt particles, and substances of an unknown nature that can adversely affect the quality of the wine. For this reason, they are preclarified by sedimentation, defecation, or centrifugation (6).

During the sedimentation, which takes about 10 h, pectolytic enzymes natural to the fruit degrade pectic substances and other cell wall components that would later stabilize the lees.

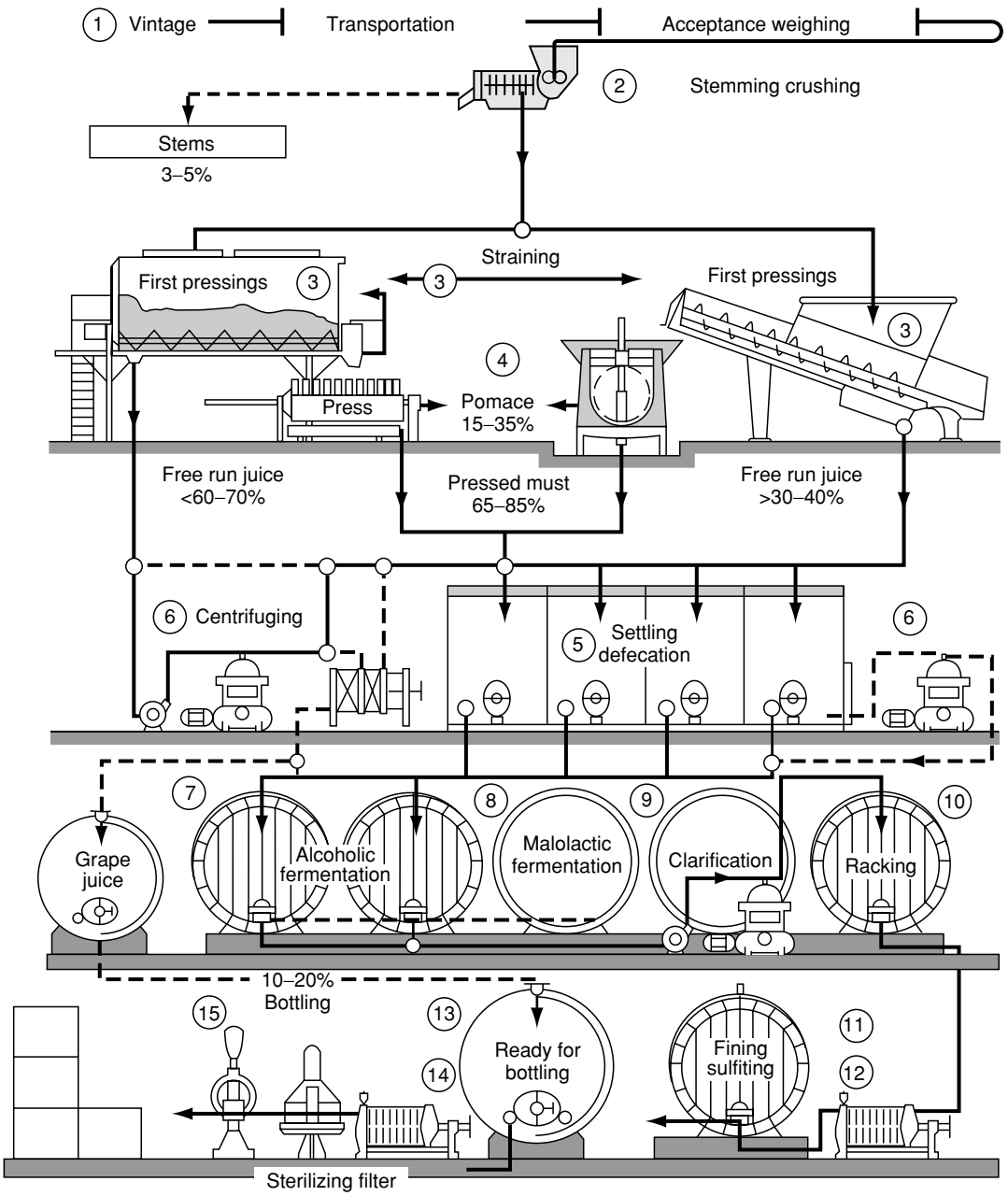

Figure 1 Flow sheet of processing for white wines according to Troost G (1988) Technologie des Weines, 6th edn. Stuttgart: Eugen Ulmer. (1) Acceptance of grapes; (2) stemming, crushing; (3) straining; (4) pressing; (5) settling, defecation; (6) centrifugation; (7) alcoholic fermentation; (8) malolactic fermentation; (9) clarification; (10) racking; (11) fining; (12) sulfiting; (13) reading for filling; (14) sterilizing filter; (15) bottling.

Amelioration of Alcohol Content and Regulation of Acidity

Freshly pressed musts frequently exhibit a disharmonious sugar–acid ratio. Musts from northern growing areas often have too much acid, and musts from more southern areas too little. Thus, during climatically unfavorable vintages, neither is likely to produce the required alcohol content. For this reason, the acid content must be regulated and the sugar content increased. To compensate for adding acid (tartaric or citric) to the wine, a deacidification process is normally carried out chemically using calcium carbonate, but treatment of the must is the preferred option. (*See* **Acids**: Natural Acids and Acidulants.)

To compensate for insuffiency of natural alcohol, it is permissible in most wine-growing countries to

add sugar (saccharose) to the crushed grapes, must, or wine (chaptalization). The sugar is intended exclusively for alcoholic fermentation and not as a sweetener.

Alcoholic Fermentation

The most important stage in wine processing is the alcoholic fermentation, i.e., the anaerobic degradation of sugar typical of yeasts, such as *Saccharomyces cerevisiae*.

Fermentation of grape must or crushed grapes is possible with the mixture of yeast strains that occur on the skins, or within damaged grapes. This is referred to as 'spontaneous fermentation.' In order to avoid the risks associated with this type of fermentation, yeast strains selected especially for their suitability for use in wine processing are used. The fermentable monosaccharides – glucose and fructose – which are present in a ratio of approximately 1:1, predominate. Saccharose (sucrose) added to low-sugar musts is likewise fermented owing to the high content of β-fructofuranosidase (invertase) in yeasts.

While the theoretical yield through fermentation of these hexoses amounts to 51.1% ethanol (by weight) and 48.9% CO_2 (by weight), the actual alcohol yield is only around 47%. Normally, 'spontaneous fermentation' can scarcely produce more than 17.5% by volume in a wine.

The large difference between the theoretical and practical yields of ethanol is attributable to loss of ethanol during fermentation, and the formation of fermentation byproducts. One of the most important fermentation byproducts is glycerol, which is a major contributory factor in forming the body of the wine. Although smaller in volume, the byproducts that react with sulfurous acid are equally important.

Heat is generated during fermentation of the must, and in the case of large must volumes, the must has to be cooled. Heating of the fermentation product accelerates the fermentation process up to a certain point. The fermentation activity of most yeast strains decreases at must temperatures in excess of 30 °C. Once fermentation has finished, the yeast settles on the tank bottom.

Malolactic Fermentation

After alcoholic fermentation and a storage time of several weeks, low-acidity wines with a pH above 3.3 often experience a decrease in the acid content. This is due not only to precipitation of the acid potassium salt of tartaric acid, but primarily to the action of lactic acid bacteria of the genera *Lactobacillus*, *Leuconostoc* or *Pediococcus*. This causes fermentation of the L-malic acid to L-lactic acid and carbon dioxide (8). (*See* **Lactic Acid Bacteria**.)

Since the malolactic fermentation usually takes place after the alcoholic fermentation, vigorous carbon dioxide production can stimulate a secondary fermentation. In France and Switzerland, this is consciously initiated and forced, while opinion is divided in other countries, especially in the case of white wine. By contrast, this 'second fermentation' is practically obligatory for red wines.

Processing of Rosé and Red Wine

The flow chart for rosé and red wine processing is shown in Figure 2. If red grapes are pressed in the same way as described for white wines, pale-pink musts are obtained from which rosé wines are produced (1, 2, 4). Although they are produced from red grapes, they are actually classified as white wines as regards their components and the technology applied.

As the pigments of red grapes are in the skin, the cells of the skin must be made permeable, which is achieved by fermenting or heating the crushed grapes (3, 6). (*See* **Colorants (Colourants)**: Properties and Determination of Natural Pigments.)

Fermentation on Skins

The classical process for the production of red wines is so-called 'fermentation on the skins' (6). The initial processing steps utilize the same methods and machines as in the production of white wine and comprise vintage, transport, grape receiving, stemming, and crushing (1, 2).

In fermentation on the skins, the red grapes are not pressed to must before the alcoholic fermentation. The crushed grapes are partially or completely fermented, in which process, the alcohol produced extracts the pigments (6).

The crushed grapes are usually left to ferment spontaneously with the natural mixture of yeast strains in a special fermentation tank. In the case of fermentation on the skins, the carbon dioxide bubbles convey the pulp and skin to the surface of the fermentation product, producing a so-called 'head' or 'cap,' which must be kept below the surface of the fermentation liquid. A number of methods can be employed to bring the skins in contact with the color-extracting liquid, such as the use of submerged caps, pumps, agitators, mechanical removal of the floating head by screw conveyors or rakes, or sudden decompression of the compressed carbon dioxide in the tank (6, 7).

Heating of Crushed Grapes

In the increasingly widespread new process for red wine production of heating the crushed grapes, the stemmed and crushed grapes are conveyed to the

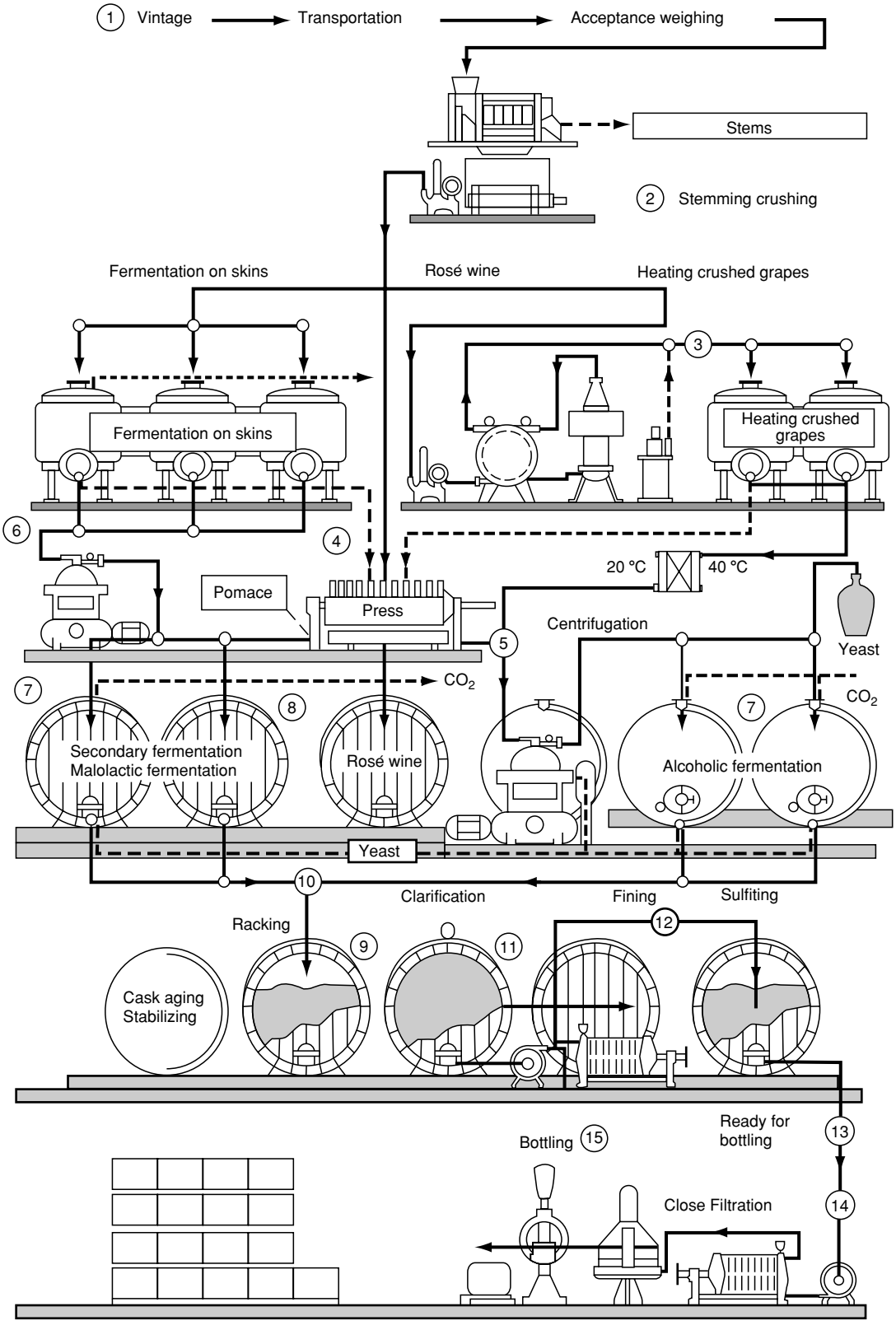

Figure 2 Flow sheet of processing rosé and red wines according to Troost G (1988) Technologie des Weines, 6th edn. Stuttgart: Eugen Ulmer. (1) Acceptance of grapes; (2) stemming, crushing; (3) heating crushed grapes; (4) pressing; (5) centrifugation; (6) fermentation on skins; (7) alcoholic fermentation; (8) malolactic fermentation; (9) clarification; (10) racking; (11) fining; (12) sulfiting: (13) ready for bottling; (14) sterilizing filter; (15) bottling.

heater by means of a smooth-action pump (3). In this process, which is also termed plasmolysis, the semipermeable, cytoplasmic membrane becomes permeable to the dissolved substances.

Heaters for crushed grapes consist of a system of tubes in a steam-heated room. It has been established that heating the crushed grapes to 45–55 °C is adequate to obtain dark-colored musts. Nowadays, it is common practice to heat the crushed grapes above the plasmolysis temperature to about 87 °C for a brief period in order to inactivate the enzymes inherent in the fruit, and to kill off the microorganisms. Subsequently, the crushed grapes are cooled to a moderate temperature.

The pretreated crushed grapes can now be conveyed directly into a press (4) or into a straining bin. After heating, processing is identical to that described for white wine. The use of selected yeasts for initiation of the fermentation process is strongly recommended.

After the alcoholic fermentation, the red wines are developed in the same way as white wines, but with the difference that the malic acid is degraded to lactic acid almost completely (8). This 'second fermentation' is brought about by the activity of lactic acid bacteria, but this frequently happens a number of months after the alcoholic fermentation.

Development

After fermentation of the must, which may have been ameliorated or whose acidity may have been adjusted, numerous processes take place in the young wine. It develops to a peak, termed maturity. This aging and finishing process takes place in a tank or in the bottle. If the maturity peak is exceeded, the wine becomes 'firne,' which can still be considered a positive quality characteristic. However, if this development continues, the wine will develop a rancid taste and become 'maderized' and oxidized – a sign of rapidly decreasing quality.

The finishing and aging processes can be influenced to a certain extent and for this reason one often talks of 'developing' the wine. This involves treating the wine with the aim of supporting the natural maturation and clarification processes, and to protect the wine from undesirable changes.

Racking

The first step in the development of wine in the cellar is racking (10, 10). By this is meant the separation of lees from the wine by decantation.

Despite racking, visible fine lees can still be suspended in the wine; in any case, the wine contains invisible colloidal dissolved substances. If the beverages are to be limpid and clear when bottled and sold, they must be clarified. This can be done by centrifugation or filtration. Before filtration, beverages are usually treated with fining agents, i.e., a portion of the substances to be separated out are flocculated, which improves the filtering efficiency. This treatment is denoted preclarification and fining of the wines (11, 11). (*See* **Filtration of Liquids**.)

Fining

The invisible particles present in the wine consist largely of polymeric carbohydrates, proteins accounting for 10–20%, and oxidized and condensed polyphenols making up the rest. All particles become electrically charged in solution; in wine, they normally carry a negative charge. If the negative charge of a wine is altered by adding particles with the opposite charge, particle repulsion is neutralized. Particles combine to form larger aggregates that sink to the bottom or float to the top, i.e., the particles are flocculated or coagulated. (*See* **Flocculation**.)

A number of organic and inorganic substances are available for treating the wine. With the exception of silica sol, the inorganic fining agents serve primarily to prevent turbidity and to eliminate nonbacterial disorders. In other words, their main function is to stabilize the wine. Today, fining is usually carried out with gelatine and silica sol, and bentonite is used for stabilization. Special effects can also be achieved with casein, isinglass, albumen, and tannin.

Sulfiting

Prior to filtration, the wines must be adjusted to a certain sulfurous acid concentration. Sulfur dioxide (SO_2) dissolves in wine to form an equilibrium mixture [1]:

$$SO_2 \rightleftharpoons HSO_3^- \rightleftharpoons SO_3^{2-} \qquad (1)$$

and it has been added to wines since time immemorial to improve the storage characteristics. It inhibits the growth of, or even kills, certain bacteria, and also prevents browning and oxidation reactions; certain substances are reduced, and taste is improved. This versatility makes sulfite indispensable in the cellar. (*See* **Preservation of Food**.)

One of the essential tasks of the 'free' SO_2 is to bond the small quantities of organoleptically perceivable acetaldehyde, thereby neutralizing its sensory characteristics. If free acetaldehyde is present in the wine, the wine creates a sensory impression that is

labelled rancid, 'maderized' and oxidized or 'sherry taste,' and avoidance of this defect is vital in white wines.

Sulfurous acid has no significance as a preservative in wine as the permissible maximum contents are too low.

Stabilization

Although fining and filtration are normally sufficient to obtain a limpid and clear wine, they can become turbid again after this treatment or can exhibit deposits. Colloidal proteins can alter and become insoluble. Wines containing thermolabile proteins are most frequently stabilized by treatment with minerals containing aluminum silicate. Bentonites are most suitable for this purpose. It is also possible that heavy-metal ions will cause discoloration or turbidity because they may either form insoluble colored compounds with the wine components or become insoluble and precipitate out. These phenomena are primarily caused by iron and copper ions.

In 1922, success was achieved in bonding the iron and precipitating it to a large extent by adding a precisely determined amount of potassium ferrocyanide (potassium hexacyanoferrate (II)). Subsequently, it was found that a series of other metal ions, such as those of copper, zinc, manganese, nickel, and silver, and the extremely toxic metal ions of lead and cadmium, could be precipitated out as ferrocyanide complexes. (*See* **Heavy Metal Toxicology**.)

After fining of the wines to protect them against changes, a further stabilization is normally required to inhibit crystalline precipitation. This is aimed chiefly at separating out potassium bitartrate together with calcium tartrate, i.e., the so-called cream of tartar. To prevent tartrate precipitation, it is usual to cool the wine down to a temperature of between +4 and 0 °C for a period of 4–10 days. After fining and stabilization, the wines must be completely clarified prior to bottling.

Whereas centrifugation is only suitable for separating out relatively coarse trub particles, filters can remove even particles down to the molecular range from the wine.

Three different types of filter are applied to wine filtration: the so-called precoat filter, in which the porous mass constantly reforms during filtration; the sheet filter, in which the porous mass is prepared beforehand and placed in the filter; and the membrane filter, which has a defined pore size. Purely mechanical clarification has a negligible effect on the chemical composition of white wines. However, the sensory properties are improved substantially.

Bottling

After development of the wine in the cellar, the wine must be 'bottle ripe', i.e., it must be physically, chemically, and biologically stable. White wines in particular should be optically clear as well as tasting fresh and elegant to meet the requirements of today's consumer. In the case of red wines, long aging in wooden casks is desirable (Bordeaux wines). (*See* **Barrels**: Wines, Spirits, and other Beverages.)

A bottling line consists of machines and equipment for cleaning the bottles, bottling, corking, capsuling, labelling, packing (in plastic cases or boxes) and palleting (15, 15). The most important feature of bottling of wine is the microbiological stabilization of the wine to ensure that the bottled wines do not become turbid due to the activity of microorganisms. In practice, this means cold-sterile or hot-sterile bottling of the wine.

Central to cold-sterile bottling is bacterial clarification of the wine by filtration with fine-pore filter sheets (sterilizing filters or germ-proof filters) or membranes (Figure 1 (14), Figure 2 (15)). The difficulties associated with cold-sterile bottling can be overcome by heating the wine. For an alcohol content of *c*. 10% by volume, a bottling temperature of 55 °C is adequate.

Classification

The composition and nature of a wine depend on the grape variety, the site and soil of the vineyard, the climate and growth conditions of the vine, and the state of the grapes. They are also greatly influenced by the must treatment, the yeast, and the treatment of the wine in the development stage. The quality of the wine generally depends on the amount of volatile compounds, such as ethanol and aromatic substances, and the content of nonvolatile compounds, which are classified under the heading 'extract.'

Wines can be classified, to a certain extent, by the analytical determination of individual components. In particular, it can be established whether the legally defined maximum and minimum limits have been adhered to. Adulteration by addition of nonpermissible substances such as acids can be detected. (*See* **Adulteration of Foods**: History and Occurrence.)

However, it is difficult to make an objective assessment of the quality of the wine with the aid of chemical analysis. Table 1 gives an overview of a number of important wine components.

Classification of a wine, therefore, necessitates a sensory evaluation as well as the results of chemical analysis. The current consensus is that analytical and sensory tests are equally important and complement

Table 1 Important components (g l^{-1}) of wines[a]

Alcohols	
Ethanol	
Table wines	50.0–110.0
Dessert wines	104.0–176.0
Methanol	
White wines	0.02–0.1
Red wines	0.09–0.75
Glycerol	3.5–25.0
Butylene glycol	*c.* 0.6
Isobutylene glycol	*c.* 0.12
Sorbitol	*c.* 0.02
Acids	
Total	
White wines	4.0–9.0
Red wines	4.0–6.0
Malic	0.0–6.0
Tartaric	0.5–4.0
Malic	0.8–3.3
Succinic	0.5–1.3
Citric	0.0–0.3
Volatile acids	
Acetic	0.15–1.2
Carbonic	0.0–1.5
Sulfurous acid free	0.002–0.05
Sulfurous acid bound	0.08–0.40
Sugars	
Fermentable	0.0–150.0
Nonfermentable	1.0–2.5
Arabinose	0.5–0.8
Tannins	
White wines	0.05–0.4
Red wines	1.0–2.5
Nitrogen compounds	
Total	0.1–0.9
Protein nitrogen	0.03–0.04
Proline	0.14
Glutamic acid	0.21
Histamine	0.0–0.22
Mineral matter (ash)	1.5–3.0
Potassium	0.5–2.5
Magnesium	0.07–0.24
Calcium	0.1–0.2
Sulfate	0.15–1.0
Nitrate	0.015

[a]From Gollmick F, Bocker Hand Grünzel H (eds) (1991) *Das Weinbuch*. Leipzig: Fachbuchverlag, with permission.

one another. In sensory evaluation of the quality of wine, color, odor, and flavor are evaluated.

Physiology and Toxicology

The value of a wine lies in the combination of sensory impressions such as odor and taste and the specific effect of the alcohol, which together make drinking in moderation a pleasurable experience. The predominant component is the ethanol, which enters the bloodstream relatively rapidly as it is easily absorbed by the mucous membranes in the mouth and stomach, but mainly in the small intestine.(*See* **Sensory Evaluation:** Sensory Characteristics of Human Foods.)

As is the case with all other aliphatic alcohols, ethanol affects the central nervous system, causing anesthesia and intoxication. It is degraded chiefly in the liver, where, in the first stage, acetaldehyde is formed by alcohol dehydrogenase, this is converted to acetic acid in the second stage by aldehyde hydrogenase and is metabolized to carbon dioxide and water in the citrate cycle. (*See* **Alcohol:** Metabolism, Beneficial Effects, and Toxicology.)

Sulfur dioxide is the only component with any toxicological significance that is used in the development of wine. Approximately 95% of the sulfite in the wine is absorbed by the body. It is oxidized to sulfate by the sulfitoxydase present to some degree in all body tissue. The latter is essential for the human body and is needed for the degradation of sulfur-bearing amino acids. The oxidation of sulfite to sulfate takes place with amazing speed. To date, it has not been possible to detect sulfurous acid in body fluids. The sulfate is discharged almost completely via the urine within 24 h, as has been demonstrated in tests using radioactively labeled sulfur. For this reason, there is, at present, no evidence that sulfurous acid poses a threat to health if it is added to wines in the legally permitted concentrations.

See also: **Acids**: Natural Acids and Acidulants; **Colorants (Colourants)**: Properties and Determination of Natural Pigments; **Filtration of Liquids**; **Flocculation**; **Heavy Metal Toxicology**; **Port**: The Product and its Manufacture; **Ripening of Fruit**; **Sensory Evaluation**: Aroma; Taste; **Sherry**: The Product and its Manufacture; **Tannins and Polyphenols**; **Vermouth**; **Wines**: Production of Sparkling Wines; **Yeasts**

Further Reading

Amerine MA, Berg HW and Kunkee RE (eds) (1982) *The Technology of Wine Making*. Westport, CT: AVI.

Bergner KG and Lemperle E (2001) *Weinkompendium*, 3rd edn. Stuttgart, Leipzig: Verlag Hirzel.

Farkas J (1988) *Technology and Biochemistry of Wine*. Montreux: Gordon and Breach.

Gollmick F, Bocker H and Grünzel H (eds) (1991) *Das Weinbuch*. Leipzig: Fachbuchverlag.

Gussek KD and Bocker H (1990) *Der Wein (Lexikon)*. Leipzig: Bibliographisches Institut.

Jackson RS (1994) *Wine Science: Principles and Applications*. London: Academic Press.

Jakob L (1995) *Lexikon der Önologie*, 3rd edn. Neustadt: Verlag und Druckerei Meininger GmbH.

Jakob L (1997) *Der Wein*, 10th edn. Stuttgart: Verlag E. Ulmer.

Lea AGH and Piggott JR (eds) (1995) *Fermented Beverage Production*. Glasgow: Blackie.

Ribéreau-Gayon J, Peynaud E and Sudraud P (eds) (1972–1977) *Traité d'Oenologie, Science et Techniques du Vin. I, Analyse et Contrôle des Vins* (1972); *II, Composition, Transformation et Traitements des Vins* (1966); *III, Vinifications, Transformation du Vin* (1976); *IV, Clarification et Stabilisation. Materiels et Installations* (1977). Paris: Dunod.

Troost G (1988) *Technologie des Weines*, 6th edn. Stuttgart: Eugen Ulmer.

Würdig G and Woller R (eds) (1989) *Chemie des Weines*. Stuttgart: Eugen Ulmer.

Production of Sparkling Wines

S Buxaderas and E López-Tamames, University of Barcelona, Barcelona, Spain

Background

Carbonic acid gas and the capacity to form bubbles are the characteristics that differentiate sparkling wines from other types of wines on the market. These wines are consumed for the most part in cocktails, aperitifs, and special occasions (weddings, baptisms, birthdays, etc.), probably because they are considered elegant wines. They are required to have a clear sparkle, a perfect tonality (especially white sparkling wines, the most popular kind), a harmonious bouquet, and a refined and attractive presence. These qualities can be achieved only with raw materials of the best quality and a meticulous elaboration, which explains their tendency to be on the expensive side.

Although sparkling wines are mostly consumed in social events, in the producing countries, different types of sparkling wines are prepared for other occasions. The *brut* type (dry or almost dry) may be drunk as a table wine, as an aperitif, or with fish, and also along with cheese. Others are dessert wines, such as the sparkling *Asti*, or appetizer wines, such as *Prosecco*.

Sparkling wines are gassy beverages in which the carbon dioxide is found in a state of oversaturation (generally, 4–6 bar at 20 °C). When the wine is poured into a glass, CO_2 is rapidly released as a result of the difference in pressure between the hermetically sealed bottle and atmospheric pressure. The effervescent CO_2 bubbles reach the wine surface forming, during the first seconds, a great volume of foam, which seconds later, is reduced to a foam collar. The endogenous or exogenous origin of the CO_2 allows these wines to be classified into two large groups: natural sparkling wines (those produced by the Champenoise, Charmat method (or similar methods), or the 'pearl' wines, which have a natural 'sparkle') (**Figure 1**), and those carbonated artificially (aerated sparkling wines).

In the first case, a dry wine to which sugar (saccharose or concentrated must) has been added is refermented or subjected to a second alcoholic fermentation either in a sealed bottled (as in the Champenoise method and other similar methods) or in pressurized tanks (as in the Charmat method). Other sparkling wines are processed by refermentation of a must partially fermented using their own sugar. In these products, the pressure of carbon dioxide forms effervescence, but it does not reach the values necessary to form foam. These wines can also be included within this group, since the CO_2 origin is endogenous. In the European Union (EEC regulation of the 1493/99), the denomination of sparkling wine is only applied to those wines that, when uncorked and poured into the glass, are able to form foam. This would explain why pearl wines, which do not reach the 3.5 bar of CO_2 pressure (measured at 20 °C), are not considered sparkling wines, but semisparkling wines. In the second case, gassy carbon dioxide of exogenous origin is added directly to a wine (aerated wines), thus creating the more modest types of sparkling wine.

Champenoise Method

The first historical information concerning the preparation of a sparkling wine dates back to the seventeenth century in France. At the end of that century, the abbot Dom Perignon (1638–1715) tried to rationalize the natural tendency of Champagne wines to sparkle lightly in the bottle. The Champenoise method, probably because of its geographical proximity, expanded to the north of Spain (Catalonia) around 1842 (Montferrant Cellar) and 1878 (Codorniu Cellar).

At present, within the European Union (EU), the Champagne region in France (located in the Epernay and Rheims area) is the only region that can use the denominations of the Champenoise method and Champagne for the production of sparkling wines. Other natural sparkling wines produced in other regions of France are known as *vins mousseux*. In Italy and Spain, the Champenoise method is designated classic (*metodo classico, Talento* wines, and *Spumanti Metodo Classico*) and traditional (*método tradicional, Cava* wines), respectively. These sparkling wines are considered to be of premium category

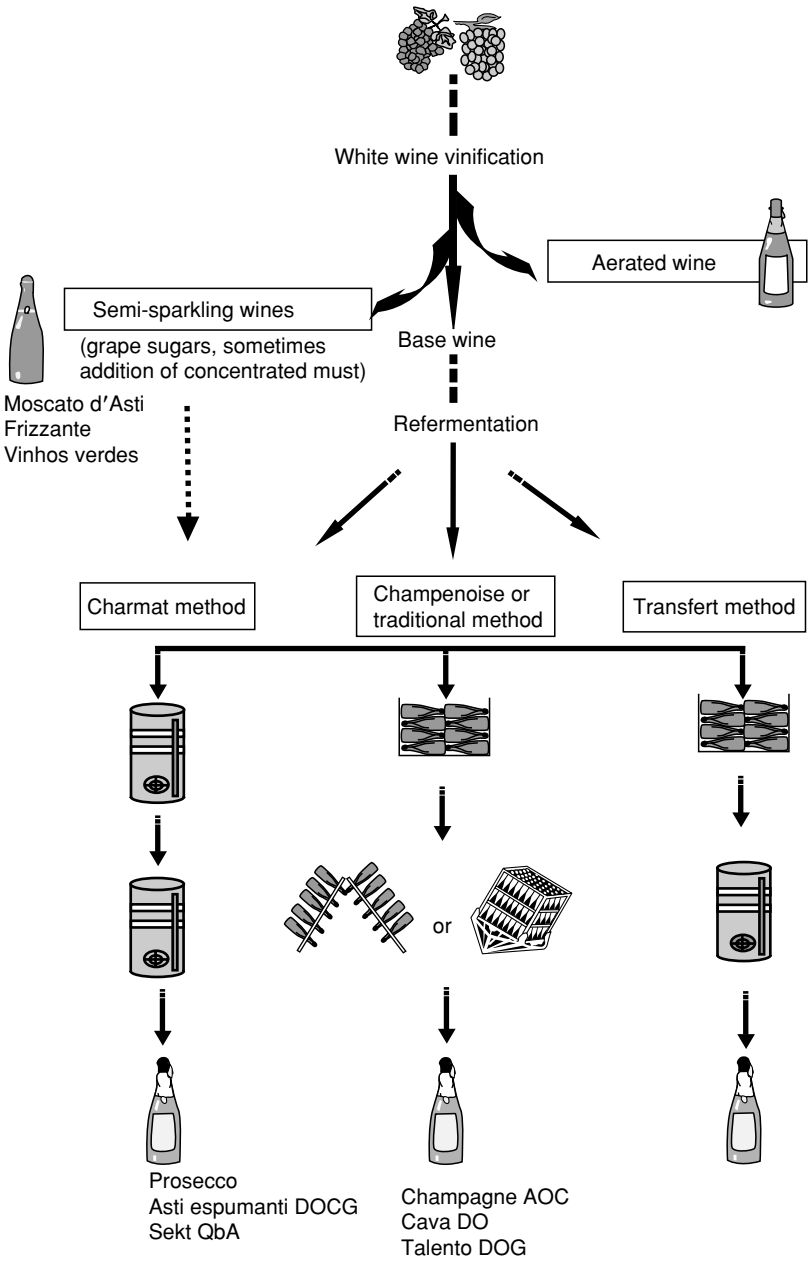

Figure 1 Process for making sparkling wines.

when their quality is guaranteed by a Certified Brand of Origin (CBO) (quality wines called *p.s.r.*, 'produced in specific regions,' according to the EU's terminology) (EEC regulations 823/87 and 2043/89). The regulations that entitle natural sparklers to receive a CBO are strict and establish the authorized grape varieties and their place of cultivation, the maximum yield per vineyard (kilograms of grapes per hectare), the viticulture, and the winemaking processes, sometimes to the point of specifying details such as the type of material to be used in ageing wine and the period of this ageing. In the USA, the denomination of Champagne for sparkling wines produced according to the method champenoise (or called 'fermented in the bottle') can appear on the label on the condition that the geographic place of its processing is mentioned (and thus labeled as Californian Champagne, New York Champagne, and so on).

In the Champenoise method, as well as in the classic and traditional methods, there are two production stages: preparation of the base wine (pressing of the grapes, clarification and fermentation of the must, and clarification and stabilizing of the wine)

and the sparkle stage (bottling and 'tirage' process, second fermentation, ageing of the wine on its lees, elimination of yeast or riddling, disgorging and addition of the 'expedition liqueur'). The differences detected among the Champagne sparkling wines, the Spumanti Metodo Classico, and the Cava are due to the different grape varieties used, the viticultural ecosystem of the geographic area, and the enological practices applied (type of pressing, blending (*coupage*) of the wines that will make up the base wine, yeast strains used for the first and second fermentation, period of contact with the yeast during maturation, etc.).

The French grape varieties Pinot noir, Pinot meunier, and Chardonnay are the main varieties used all over the world to produce natural sparkling wines (in Italy, England, USA, Australia, New Zealand, etc.). Sometimes, these wines are blended with wines from other varieties that may or may not be native to the area. In Italy, Pinot blanc is also used. In England, these varieties are blended with hybrid varieties (e.g., Bacchus, a cross between a Silvaner and Riesling hybrid with a Müller-Thurgau grape, and Kerner, a cross between Riesling and Schiave grossa grapes) and other English varieties (e.g., Schonburger). Spanish Cava is probably one of the few exceptions. The main grapes used are native varieties (Macabeo, Xarello, and Parellada) as a way of fostering Cava's distinctive sensory characteristics. Only the foreign variety of Chardonnay is allowed for white sparkling wines and the Pinot noir variety for Cava *rosé*.

Base Wine Preparation

This preparation involves off skin fermentation, but with certain peculiarities. Grapes are picked in 20–25-kg lugs (in Italy and Spain) or 35-kg lugs (in France) so that the grapes reach the winery uncrushed. It is better to avoid the release of phenolic compounds from the broken skins and the reactions of enzymatic oxidation. That is also why grapes are pressed gently with mechanical or pneumatic horizontal or lateral presses, which have replaced the old vertical basket presses. The free-running juice produced during the loading of the press and the grape juice, which is extracted with very gentle pressure, are used to produce the highest-quality natural sparkling wines, whereas the must obtained by increasing the pressure is used for lower-quality sparkling wines or other wines such as table wines The regulation of the Champagne Certified Brand of Origin (A.O.C., 'Appellation d'Origine Contrôlée)' and that of Cava (D.O., Denominación de Origen) requires a 66.6% must yield, which is slightly higher (max. 70%) for the Talento (D.O.C., 'Denominazione di origine controllata'). Approximately 2.7 l are obtained from 4 kg of grapes, constituting the free-running juice and the juice of the first press fraction.

The must is treated with sulfur dioxide to block polyphenol oxidase and prevent uncontrolled fermentation. Before alcoholic fermentation occurs, the must is racked by physical methods (sedimentation for 12–24 h, then filtration or centrifuging) and its acid and sugar levels corrected. The cold climate of the Champagne region and Germany is responsible for the high acid but low sugar levels, thus necessitating chaptalization (addition of sugars). Musts obtained in the Mediterranean climates of Italy, Greece, and Spain have sufficient quantities of sugar to ensure a minimum alcohol content of 9–9.5% of base wines, but it is often necessary to add tartaric and/or citric acid. Alcoholic fermentation is carried out at 16–20 °C so as to minimize the loss of aroma. Temperature control is facilitated when fermentation occurs in stainless steel tanks, which incorporate a refrigerating system (double jacket). Since fermentation may take several days to finalize, wine is devatted before fermentation is completed to prevent the appearance of mercaptan compounds resulting from prolonged contact with the lees. Dry wine, with a residual sugar content usually no higher than $1.5\,g\,l^{-1}$, is left to settle for several months (cloudy wines). During this time, it is clarified by sedimentation, and if this is not adequate, fining agents are employed (bentonite, gelatine, caseinates, etc.). Malolactic fermentation is favored only when the acidity is excessive (15 g of malic acid per liter). Malolactic fermentation should not occur in the bottle, since this would increase the CO_2 pressure, and colonies of lactic bacteria would form viscous lees attached to the inside of the bottle. Tartaric stabilizing involves precipitation of tartrate salts. Tartaric acid and its salts become less soluble in hydroalcoholic solution when sparkling wine is at a low temperature. To prevent tartaric crystals forming the tank temperature is reduced below 0 °C (−4 °C). Finally, a sterilizing filtration is carried out (base wine). Generally, the base wine is not a monovarietal wine. The blending, or *coupage*, among wines of different characteristics can be done in the tank for tartaric stabilization or before filtration to yield a base wine with more rounded sensory properties. It is permissible for some natural sparkling wine labels to indicate the proportion of monovarietal wines used in the preparation of the base wine (as in France, Italy, Germany, USA, and South Africa), but in other countries, such as Spain, Greece, and Argentina, the wine varieties may be mentioned, but not the proportions of the blend.

Fermentation and Aging in Bottle

The settling of the sparkle, or *prise de mousse* phase, starts with the 'tirage' process, when the wine is bottled, and the sugar and yeast solution are added. The 'tirage liqueur' is added to the base wine (55 ml of wine-liqueur mix per liter). The composition of this liqueur is: 1–2 million cells of yeast per milliliter, ≈ 500 g of saccharose per liter and 0.1–0.2 g of bentonite per liter to agglutinate the lees during riddling, and/or ammonium phosphate (50 ml hl^{-1}) and others compounds to stimulate yeast growth. The bottle glass is usually thick to withstand the CO_2 pressure generated by the second fermentation (5–6 bar at 20 °C before opening). The bottles are hermetically sealed with a cup-shaped plastic insert, known as a bidule, and a metal crown cap, and are stacked horizontally in bins or along cellar walls, at a temperature of 12–14 °C. The settling of the sparkle phase varies in duration; a minimum of 9 months for Cava and 15 months for Champagne and Spumanti Metodo Classico. This length of time is sufficient for the second alcoholic fermentation to take place and for the exchange of compounds between the yeasts and the wine. However, the ageing period for natural sparkling wines tends to be much longer. Yeast autolysis is believed to occur 9 months after the tirage. This enzymatic process implies the release of compounds from the cell walls followed by a progressive dissolution of the cytoplasmic compounds in wine. Between the 12th and 18th month, marked changes are detected in the wine composition (the content of polysaccharides, peptides, and free amino acids rises, and that of proteins falls). Undoubtedly, the special ageing of natural sparkling wines depends on this contact between the wine and the less of the second fermentation, but the reducing atmosphere created by the CO_2 also plays a part. The complex reactions that take place during the settling of the sparkling wine are still under study, but they are known to give rise to each wine's special and characteristic bouquet (matured aromas and flavors that supply the fruity, nutty, yeasty, and toasty notes). The most frequent ageing period is 2–3 years, although in some cases, it may be up to 6 years, but sparkling wines tend to lose their freshness if left to age excessively. When the wine has reached the desired ageing, the dead yeast cells have to be eliminated through the riddling (*remuage*) process. The sediment that has accumulated in the bottle on the side on which it has been lying is directed to the mouth of the bottle and is gathered in the bidule of the inverted bottle. In manual riddling, the bottles are placed in classical racks, called *pupitre*, at an angle of 25–30 ° from the ground. Bottles are progressively given an 1/8 twist, and their inclination is also increased, so that the sediments slide to the mouth of the bottle. At the end of the riddling operation, the bottle is neck down and the sediment is in the bidule, ready for disgorging. Classical hand riddling requires many specialized workers and lasts for 1, 2, or more months. This is why most wineries have replaced the racks with 500-bottle bins, which rest on an inverted octagonal pyramid. Each change of the octagonal face on which the bin is resting represents the 1/8 spin of the rack. At the same time, bascule guides increase the inclination of the bin until it is vertical. The most sophisticated riddling system has replaced the mechanical control of the bin movements on the octagonal base with automatic riddling machines. These modern systems take only 2–4 weeks to do all the riddling. There are also automatic racks (Pupi-matic) that continually shake and tilt 240 bottles. The process only lasts 10 days but requires more workhands to load and unload the rack than the bin system, whose unloading into the disgorging machine is mechanized. The bottles sitting in the neck-down position facilitate the disgorging process. The mouth of the bottles is plunged into a freezing solution of ethylene glycol (45%). When the crown cap is removed, the ice plug with the frozen sediment shoots out of the bottle due to the CO_2 pressure. Pressure loss is approximately 1 bar, and wine loss is 10–15 ml. The volume lost is restored by adding the filling solution or 'expedition liqueur'. This is made up by wine recovered from the disgorging, or an aged wine, to which sugar and sometimes liquor (e.g., Armagnac, brandy, or rum) or a fortified wine is generally added. Sparkling natural wines in the EU receive different appellations, depending on the amount of sugar added to the expedition liqueur: Brut nature < 3 g l^{-1} (only residual sugars), extra brut 0–6 g l^{-1}, brut < 15 g l^{-1}, extra dry 12–20 g l^{-1}, dry 17–35 g l^{-1}, medium dry 33–50 g l^{-1}, and sweet > 50 g l^{-1}. Sweetness tends to disguise some defects in the wine. Therefore, natural sparkling wines of premium quality are the brut nature and extra brut wines, which contain only the residual sugar of the second fermentation or have been sweetened only very slightly. For these natural sparkling wines, the raw material and the production process must be impeccable. The bottle used in the settling of the sparkle is the same bottle that reaches the consumer, corked, foiled, and duly labeled.

The Charmat or Granvas Method (or Bulk Method)

The basic difference in this method compared with the Champenoise method (fermentation in the bottle)

is that the second fermentation is carried out in a Charmat tank (a sealed tank that traps the carbon dioxide gas produced during the second fermentation). This is the main method used in the USA, Australia, and New Zealand) (Table 1). The period of time during which the wine is in contact with the lees of the fermentation distinguishes the short Charmat from the long Charmat. Young and fruity wines preferably should not lose these sensorial characteristics by ageing in contact with the yeasts. The organoleptic result of ageing on the lees reduces their value. This is the case with the Italian Asti and Prosecco. The production method most suitable for these is the short Charmat.

Table 1 1998 world production[a] of sparkling wine

Origin	*Type of sparkling wine*	*Quantity of sparkling wine bottled produced (× 1000)*	*Percentage of the sparkling wine type*
Argentinean	Others	23.148	100
Australian	Method traditional	17.700	34.3
	Others	33.900	65.7
Austrian	Sekt	18.024	86.6
	Frizzante	2.832	13.5
	Others	84	0.1
Brazilian	Method traditional	1.302	6.4
	Others	19.050	93.6
Canadian	Method traditional	1.061	13.4
	Others	6.859	86.6
Chilean	Others	10.104	100
Chinese	Others	15.342	100
French	Champagne	272.160	53.2
	Method traditional	22.232	4.3
	Others	217.056	42.4
German	Sekt	257.904	59.1
	Others	178.320	40.9
Greek	Others	888	100
Italian	Method traditional	16.925	9
	Asti	49.968	26.9
	Prosecco	26.472	14.3
	Others	92.196	49.7
Japanese	Others	1.650	100
Mexican	Method traditional	324	45.0
	Others	396	55.0
New Zealand	Method traditional	4.464	41.4
	Others	6.312	58.6
Portuguese	Others	12.924	100
Russian	Others	2652	100
South African	Method traditional	663	6.2
	Others	10.116	93.8
Spain	Cava	190.508	81.4
	Others	43.632	18.6
Swiss	Others	264	100
Turkish	Others	204	100
Ukrainian	Sekt	90	100
Uruguayan	Others	408	100
USA	Method traditional	40.440	33.3
	Others	80.856	66.7
Venezuelan	Others	666	100

[a]Data provided by Freixenet, S.A. cellar.

Short Charmat Method

The base wine (monovarietal or *coupage*) ferments in a Charmat tank to which the 'tirage liqueur' (sugar, nutrients, and yeast) has been added. This liqueur has a similar saccharose concentration to that used in the Champenoise method, but less volume is added (44 ml instead of 55 ml of wine-liqueur mix per liter). Then, 10% of yeast-enriched wine is added to the sweetened base wine. The ideal temperature for fermentation is 12–13 °C. When the desired CO_2 pressure has been reached (approximately 4 bar), the temperature of the Charmat tank is reduced to 8 °C to arrest the fermentation (with 10% of residual sugars). The low temperature causes the suspended yeast to sink to the bottom. The next step is to transfer under pressure the partially decanted wine, already bubbly, to the Charmat tank where it will be cold stabilized. When the wine is transferred from one Charmat tank to another, it is filtered or centrifuged to separate it from the lees, thus ensuring minimal contact between the wine and the lees to prevent the yeast from absorbing bouquet notes. Filtration by counter-pressure uses a gas (usually carbon dioxide) to compensate the empty volume of the Charmat tank while it is being emptied of the wine. The gas that is introduced will be at least 1 bar higher than the pressure of the sparkling wine to ensure that it remains invariable. This is also why the gas introduced into the second Charmat tank, which will receive the sparkling wine through the filter or hermetic centrifuge, remains at the same pressure as that of the sparkling wine. The gas used most often is nitrogen, rather than compressed air or carbon dioxide. Compressed air is not favored because of the negative influence of oxygen on wine. In some countries, such as France, Italy, or Spain, the use of industrial carbon dioxide is forbidden in cellars that produce natural sparkling wines, and can be employed only by companies that produce artificially gasified sparkling wines. However, it is possible to use the carbon dioxide that is recovered from the alcoholic fermentation as a compensation gas. If necessary, tartaric cold-stabilizing (−4 °C) is carried out when it has not previously been applied to the base wine. The wine is then refrigerated at 0 °C to prevent liquid loss, especially between the bottling and corking process.

There are several sparkling wines of excellent quality that deserve special attention, such as the Moscato d'Asti (a semisparkling wine, according to the EU's denomination) and the Asti spumate (Asti DOGC: denominazione di origine controllata e garantita)

(Table 2). The Moscato d'Asti is thought to be originally from Piamonte in Italy, and as ancient as Champagne. There are dessert wines produced from muscat grapes (Moscato bianco) and with a low alcohol content (4.5–5.5%), which differ from the sparkling wines previously described in that the base wine is a must wine whose second fermentation is carried out using sugars from the grapes of origin. In the past, the artisan production of the Moscato d'Asti meant that alcoholic fermentation had to be arrested when the desired alcohol content, CO_2 pressure, and quantity of residual sugars had been reached. Successive filtrations and the elimination of nitrogenated compounds gradually diminished the fermentation process until the critical point was reached. This point was determined by the percentage of nitrogen present and the quantity of live cells that made the wine incapable of fermenting when exposed to low-temperature, alcoholic, and reducing (CO_2) conditions. At present, the procedure for producing Moscato d'Asti is essentially the same, but modern techniques are available to facilitate the process (temperature of $-3\,^{\circ}C$, sterilizing microfiltration, and the use of selected yeasts). The filtered must is left to lie at $0\,^{\circ}C$ for 12–15 months. During this time, as soon as incipient fermentation is detected, the must is filtered and refrigerated again to eliminate and deactivate any yeasts, and it reaches a maximum alcohol content of 1%. Once the ageing period has passed, selected yeast strains are added, and the must is left to ferment at a controlled temperature of 18–20 °C. When an alcohol content of 4.5–5% is reached, and approximately 80 g of residual sugars per liter are left over, fermentation is stopped by refrigeration at $-3\,^{\circ}C$, followed by microfiltration and bottling. In some cellars, the product may be pasteurized to ensure that a second fermentation does not reoccur. The difference between the Moscato d'Asti and the Asti spumante is the alcohol content and CO_2 pressure in the bottled product. Asti spumante ferments aged must wine with 1% alcohol in a Charmat tank until an alcohol content of 7.5–9% is reached. As for the pressure, once fermentation has started, CO_2 is released, but when the product approaches the 5% alcohol content, the Charmat tank valve is closed. The pressure in the tank will reach 6–7 bar at $20\,^{\circ}C$ so as to ensure a pressure of 3.5 bar once the product is bottled.

Long Charmat Method

This method is used to produce sparkling wines with organoleptic characteristics that are closer to those achieved by the Champenoise method, especially wines of the Pinot and Chardonnay varieties. This is because the period during which the wine remains in the Charmat tank of fermentation is prolonged for several months (minimum 6 months). Wine ageing in contact with the lees gives rise to the development of aromatic compounds typical of this special ageing process. Another difference with respect to the short Charmat is that the sugar added for the second fermentation should be the necessary one to obtain a dry sparkling wine with a pressure of around 6 bar. Once the 6-bar pressure (measured at $20\,^{\circ}C$) is reached, the wine is kept in contact with its lees for 9 or more months at room temperature and with intermittent riddling. The riddling is delayed 2 or 3 weeks before the end of the ageing period, and the wine is then decanted. The decanted wine is centrifuged, and at the exit, a dosage pump injects expedition liqueur suitable for the type of sparkling wine desired. Then, the centrifuged wine is filtered into the Charmat refrigeration tank. The refrigeration temperature

Table 2 Enological characteristics of some natural sparkling wines

Parameters	*Cava*	*Champagne*	*Moscato d'Asti*[a]	*Asti Spumante*	*Talento*
Etanol (% v/v)	10.8–12.8	> 11	4.5–6.5	7–9.5	> 11.5
Total acidity ($g\,l^{-1}$)	> 5.5	> 5.5	> 5	> 5	> 5.5
Residual sugar ($g\,l^{-1}$)	< 3	< 3	> 50	> 50	< 3
Extracto seco ($g\,l^{-1}$)	13–22	> 16	> 15	> 17	> 16
Pressure CO_2 (bar)	> 3.5	> 3.5	< 1.7	> 3	> 3.5
Aroma/notes	Butter Honeysuckle Honey Breadcrumbs Nuts	Fruity Floral Woody Spicy Bread	Ripe fruits Muscatel Terpenes: linalool, nerol, geraniol	Muscatel Cooked apple Furanic oxydes α-terpineol Ho-trienol	Fruity
Color	Pale yellow Straw-colored Green tint	Golden yellow	Yellow, made of straw	Pale golden, made of straw	Yellow with golden shade

[a]A semisparkling wine.

depends, as for the short Charmat, on whether tartaric stabilization is needed. After bottling under pressure at a low temperature, the sparkling wine is left to age for a long period (7–14 months) in bins so as to acquire the premium organoleptic qualities, especially with wines of Pinot origin (such as the German dry sparkling wines). A typical example of this type of production is the German and Austrian natural sparkling wine known as *Sekt*. Even though this sparkling wine can be produced following the traditional method, Charmat method, or Transfer method, for the most part, the wine is produced using the long Charmat method (85%), only 7% by the transfer method, and 8% by the traditional method.

In some cellars, once fermentation is complete, yeast autolysis is accelerated by increasing the temperature in the Charmat tank to 42 °C for 72 h. The disadvantage is that the CO_2 pressure is doubled from 5 bar at 20 °C to 10 bar. Therefore, Charmat tanks resistant to 13 bar are used, even though the pressure in the interior is not allowed to exceed 8–9 bar because decompression is provoked. The temperature stimulates the activity of the proteolytic enzymes and the lysis phenomenon to improve similar organoleptic characteristics to those of sparkling wines fermented in the bottle. This shortens the long Charmat method, even though an aging period of several months is still recommended after bottling.

Transfer Method

This system, of Italian origin, obviates the need for the riddling and disgorging steps characteristic of the Champenoise, traditional, or classical methods. The term 'transfer' refers to transferring the wine, which has undergone a second fermentation in a bottle to a Charmat tank (**Figure 1**). Therefore, it can be considered an intermediate system between the Champenoise method and the Charmat method. The wine is left in the bottle with only a crown cap during a minimum of 9 months, sometimes more. It is then transferred under pressure with the lees to the Charmat tank. From this moment on, the process is exactly like that of the Charmat method. The drawback of this method is that it is as laborious as (if not more so) the Champenoise method (from bottle to tank to bottle), and that filtration is carried out on a product that contains a Champenoise bouquet, significantly reducing the sensorial qualities of the sparkling wine. This method is declining and has already been abandoned in Italy, but is still employed on a limited scale in some German, Austrian, Greek, Hungarian, American, and Argentinean companies. In the USA when the sparkling wine has been transferred to another bottle different to that used for fermentation, the label may mention that it has been 'fermented in bottle,' but not 'fermented in *this* bottle,' as in the case of sparkling wines produced by the method champenoise.

See also: **Grapes**; Types of Table Wine

Further Reading

De Rosa T (1987) *Tecnologia dei vini spumanti*. Brescia: Edizioni AEB.

Flanzy C (1999) *Oenologie: Fondements scientifiques et technologiques*. Paris: Technique et Documentation.

Naudin C and Flavigny L (eds) (1999) *Larousse des Vins. Tous les Vins du Monde (1998)* Paris: Larousse-Bordas.

Rankine B (1995) *Making Good Wine. A Manual of Winemaking Practice for Australia and New Zealand*. Sydney: Pan Macmillan Australia.

Ribérau-Gayon P, Glories Y, Maujean A and Dubourdieu D (1998) *Traité d'Oenologie*. Paris: Dunod.

Stevenson T (1998) *Christie's World Encyclopedia of Champagne & Sparkling Wine*. Bath, UK: Absolute Press.

Wiesenthal M and Torres MA (eds) (1988) *Enciclopedia del vino: Enología, viticultura y cata*. Barcelona: Ediciones Orbis.

http://www.astidocg.it
http://www.champagne.fr
http://www.crcava.es
http://www. englishbubbly.com/html
http://www. europa.eu.int/eur-lex
http://www.talento.it

Dietary Importance

S S Percival and S L Talcott, University of Florida, Gainesville, FL, USA

Background

The health benefits ascribed to wine are associated with both moderate alcohol intake and the content of polyphenols that have antioxidant capability. This review will cover wine composition, epidemiological data, and experimental data related to beneficial effects of wine consumption. It is important to note that alcohol consumption and mortality follow a J-shaped curve (**Figure 1**); that is, if nondrinkers are assigned a relative risk of 1.0, then the odds ratio for relative risk of adverse health effects is lower for those

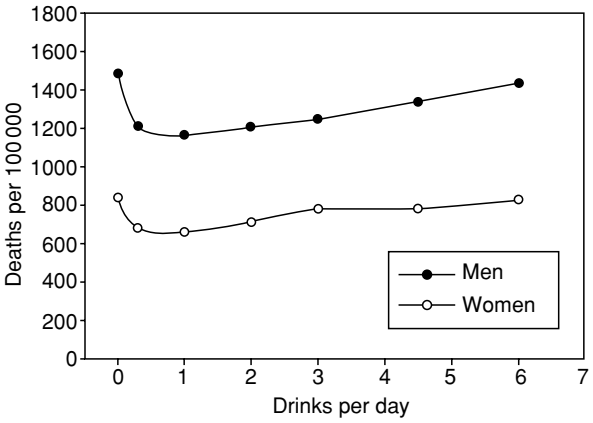

Figure 1 Rates of death from all causes.

consuming one to three drinks per day. When consumption is greater than three drinks per day, the odds ratio is greater than for nondrinkers. Thus, teetotalers and heavy drinkers have a greater risk of adverse effects and death than moderate consumers. Health benefits, accordingly, are manifested only in relation to a moderate consumption of wine, moderate in terms of quantity related to body size and therefore defined as two glasses (150 ml per glass or 5 oz per glass).

The grape species used in wine production is usually *Vitis vinifera*. When information is available, we will also include data related to the species *Vitis rotundifolia*, also known as the muscadine grape. This grape has a higher antioxidant profile and contains unique components and amounts compared with *Vitis vinifera*. Very few studies have been done with the *Vitis labrusca*, which is also known as the Concord grape.

Wine Composition

The phytochemical content of grapes of all species is linked to environmental conditions, but levels in wines are also influenced by viticultural practices, variety, and processing procedures (Table 1). In addition to ethanol, wine (especially red wine) contains a range of phytochemical compounds that confer health benefits for a variety of reasons that will be discussed. These include phenolic acids (*p*-coumaric, cinnamic, caffeic, gentisic, ferulic, and vanillic acids), trihydroxy stilbenes (resveratrol and polydatin), and flavonoids (anthocyanins, catechin, epicatechin, myricetin, kaempferol, and quercetin). An estimate of the different classes of compounds in wine follows (Table 2). Nonflavonoid compounds such as the hydroxy benzoic acids, hydroxycinnamic acids and stilbenes make up about 40% of the total

Table 1 Variability in total phenolics analyzed by different laboratories

	Investigator 1	*Investigator 2*	*Investigator 3*
Cabernet Sauvignon	1800	2600	821
Petite Syrah	3200	2266	
Muscadine			1269

phenolic acids and polyphenols. Flavonoids make up the remaining 60%. Hydroxybenzoic acids make up about 60% of the nonflavonoid fraction with gallic acid, about 120 mg l^{-1}, found in the highest quantity. Anthocyanins make up the largest portion of the flavonoids with the distribution as 50% anthocyanins, 30% flavanols such as catechin and epicatechin and 20% flavonols such as quercetin, myricetin, and kaempferol. Procyanidins are large complexes of epicatechin, catechin, and gallic acid. They comprise the largest fraction of the flavanols. A high level of flavonoids is found in the skin and seeds of the grape. Therefore, the longer the fermentation time with skin and seeds, the more flavonoids are extracted. Consequently, white wine has relatively low levels of flavonoids compared with red (Table 2).

V. rotundifolia, the muscadine, has approximately twice the polyphenolic content of the more commonly consumed *V. vinifera* and significantly more gallic acid and epicatechin (Table 3). *V. rotundifolia* has less catechin and procyanidin B3 than *V. vinifera*. Also, *V. rotundifolia* is found to have approximately 10 times more resveratrol than *V. vinifera*.

Muscadine grapes are similar to *V. vinifera* in that both contain high concentrations of neutral and acidic polyphenolic compounds. Neutral phenolics contain characteristic structural features that have been associated with radical scavenging abilities, including the number and location of hydroxyl groups, and the presence of a 2, 3 double bond. Additionally, these compounds have the ability to act synergistically with other antioxidant compounds or chelate metals that participate in free radical generation.

Epidemiological Evidence for Health Benefits of Red Wine

Epidemiological evidence that led to the suggestion that red wine confers benefits was first revealed in the Mediterranean population. Despite an intake of saturated fat about three times greater than Americans, the French population has only about one-third of the incidence of coronary disease. Consumption of red wine has been hypothesized to be the contributing factor in protection against cardiovascular disease

Table 2 Relative distribution of phenols and polyphenols in red and white wines

	Red wine	*White wine*
Total polyphenols and phenolic acids	1200 mg l^{-1}	200 mg l^{-1}
Nonflavonoids	240–500 mg l^{-1} 60%	160–260 mg l^{-1} 85%
OH-benzoic acids (gallic acid, protocatechuric acid, *p*-hydroxybenzoic acid)	15%	45%
OH-cinnamic acids (coutaric, caffeic, caftaric, coumaric, ferulic)	80%	55%
Stilbenes (resveratrol, polydatin)	5%	trace
Flavonoids	750–1060 mg l^{-1} 40%	25–30 mg l^{-1} 15%
Flavonols (quercetin, myricetin, kaempferol, rutin)	10%	0
Flavanols (catechin, epicatechin, procyanidins)	30%	100%
Anthocyanins (delphinidin, cyanidin, petunidin, peonidin, malvidin)	60%	0

Table 3 Comparison of polyphenolics in Muscadine and Cabernet Sauvignon

Compounds	*Cabernet Sauvignon 13-day skin fermentation*	*Muscadine, cv. Noble 14-day skin fermentation*
Total phenolics	821 mg per liter of GAE[1]	1269 mg per liter of GAE
Gallic acid	15.9	200
Caftaric	4.6	0.9
Coutaric	0	0.3
Catechin	32.6	2.6
Ellagic	0	4.1
Epicatechin	40.3	205
Procyanidin B3	30.2	0.9
Procyanidin B4	5.2	
Total added	128.8 (15.7% of GAE)	413.8 (32.6%)

GAE, gallic acid equivalents.

and is popularly known as the 'French Paradox.' The positive attributes of the Mediterranean diet, besides a moderate consumption of alcohol, mainly as wine, include: high monounsaturated to saturated fat ratio; high consumption of vegetables, fruits, legumes, and grains; moderate consumption of dairy, mostly as cheese; and low consumption of meat. The French population consumes less milk, and more garlic, cheese, fruits, vegetables, and wine. Therefore, this postulate, that red wine is the protective factor in these diets, cannot be wholly ascribed to the wine. Other dietary factors may play a considerable role.

Epidemiological data showing a relationship between chronic disease and red wine consumption are limited, but the data suggest a significant correlation between red wine and reducing the risk of heart disease. The fact that red wine is high in flavonoids and polyphenols and therefore antioxidants is often the reason why the consumption of red wine is thought to reduce the risk of disease.

Although benefits due to red wine consumption and heart disease are largely thought to be due to polyphenols, some studies suggest a benefit due to alcohol regardless of the form in which it was consumed. One study was not successful in showing a wine and heart disease correlation, but the consumption of wine by this particular population was very low.

Epidemiological evidence supporting a reduction in risk for certain cancers is strong for consumption of fruits and vegetables or for the consumption of green tea, yet data do not exist for red wine consumption. Some evidence exists from animal studies and will be discussed.

Relation to Diet

Bioavailability

Research on the absorption of polyphenolics from foods is limited, and yet it is the critical element in understanding health benefits conferred by red wine as well as by fruits, vegetables, and teas. Limited data are available on the bioavailability of bioactive compounds from red wine; however, anthocyanins have been detected in human plasma as well as in the urine of volunteers who drank red wine. Furthermore, anthocyanins have been detected in human and rat plasma as intact glycosides with no deglycosylation or further methylation or sulfation, although the level was less than 1% of the oral dose. A methylated form of a particular anthocyanin was detected in the liver of the animals. Catechin has been detected in

mouse plasma after consumption of a diet containing wine solids.

Colonic bacteria can metabolize polymeric procyanidins. An *in vitro* study showed that they were able to be completely degraded to low-molecular-weight aromatic compounds. These monomers may be easily absorbed in the large intestine and may therefore result in further beneficial effects.

Numerous studies indirectly assess bioavailability by measuring plasma antioxidant levels as an index of absorption and transport to the blood. Each day, 356 ml (12 oz) of red Muscadine wine (cv. Noble) were consumed for 4 days. The oxygen radical absorptive capacity (ORAC) in human plasma was increased by 10–15% (Figure 2). When the serum sample was deproteinized by acetone, the plasma ORAC values were lower but increased 21–33% over the 4 days. Mice consumed muscadine wine or Cabernet Sauvignon over 9 days as their sole source of fluid. Based on preconsumption levels, plasma ORAC levels rose by two-thirds in the case of the cabernet and doubled in the case of the muscadine (Figure 3). The results showed that the antioxidant capacity in plasma increased two- to three-fold with the consumption of the wines. The antioxidant capacity in this study was measured by DPPH (1, 1-diphenyl-2-picrylhydrazyl). This has been shown by others and confirms that the compounds that confer antioxidant activity are absorbed and transported to the plasma.

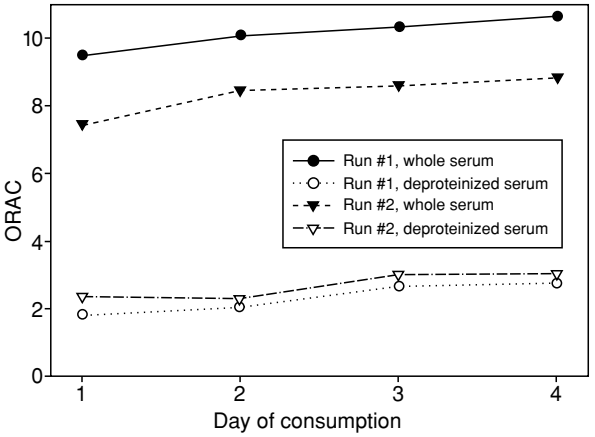

Figure 2 Human serum of oxygen radical absorptive capacity (ORAC) values after 9 days of consumption.

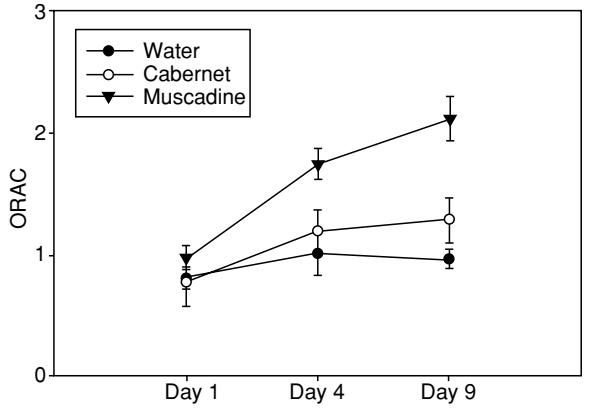

Figure 3 Mouse serum oxygen radical absorptive capacity (ORAC) values after 9 days of consumption.

Intake

The estimated intake of total polyphenols in the diet per day is 1 g. Although this value was derived over 25 years ago, it appears still to hold true. Wine composition is complex and variable. Because of the variability among grapes and wine, it would be difficult to estimate the consumption of polyphenols, even if consumption of the wine volumes were known. Teissedre, however, calculates the flavonoid consumption by the French population to be about 400 mg per day from red wine. In the USA, the consumption would probably be less.

Individual Components of Red Wine having Potential Benefits

The phenolic compounds in wine all have the potential to act as antioxidants. Because of structural differences, the phenolics vary in their relative activity.

Ellagic acid has been well studied as an isolated compound. Ellagic acid is an important phenolic compound found in many small fruits including cranberry, blackberry, blackcurrant, mayhaw, strawberry, and raspberry. Various reports have shown a positive effect for ellagic acid as a functional antimutagen, anticarcinogen, and inhibitor of various chemically induced cancers. Muscadine grapes have appreciable amounts of ellagic acid (10–25 mg l^{-1}) thought to be derived either directly from hydrolysis of ellagitannins or from dimerization of gallic acid.

Resveratrol is a phytoalexin that has been associated with anticancer, antiinflammatory, and antimutagenic activity. Resveratrol is found in the highest concentration in the skin and seeds, especially when the plant has been stressed. The content of resveratrol in the red wines from *V. vinifera* contains very little resveratrol, whereas it is found in higher concentrations in the muscadine wines. During processing, the longer the skins and seeds are fermented with the crushed grapes, the more resveratrol will be found in the wine. Resveratrol was first discovered in the *V. vinifera* grape species. It is synthesized by the leaf and skin in response to fungal infection. Thus, wine

derived from healthy plants has little to no resveratrol. Muscadine, however, has higher levels of resveratrol. The biological activities of resveratrol are numerous. It has been shown that resveratrol has antiinflammatory activity from its ability to inhibit eicosanoid production, antioxidant activity by scavenging radicals and chelating metals, and anticancer activity including antimutagenic, anticarcinogenic, and chemopreventative activity. Heart health benefits may be derived from its ability to inhibit platelet aggregation and induce vasorelaxation. Resveratrol has also been shown to have estrogenic activity.

Catechin and epicatechin are major components in red wine and are responsible for all the flavonoid content of white wine. These polyphenols are high-molecular-weight molecules with the ability to complex with other phenolic acids making a near endless array of procyanidins. They are also found in green tea and chocolate.

Health Benefits

Alcohol, Flavonoids, and Immunity

Ethanol is known to modify immune cell function detrimentally. In animals that drank 5% ethanol daily, there was a marked reduction in total splenocytes and total thymocytes, and an increase in T-cells relative to B-cells in the periphery. An increase in natural killer (NK) cytotoxic activity in both the spleen and the periphery was also found, but there was little change in NK cell number. We also found a modest, but significant, change in peripheral NK cell numbers.

Alcohol consumption for 2, 4, and 8 weeks was associated with a decrease in the number of NK cells in the spleen and periphery over time. The cytotoxic activity of peripheral and splenic NK cells was decreased modestly, in contrast to the aforementioned study.

The effect of alcohol on NK activity is one of the most controversial because of the variable results among laboratories. A reduction in the number of NK cells or their activity would have serious consequences in terms of host defense to viruses and tumors. Further research is needed to understand the role of both alcohol and wine polyphenolics on NK cell activity.

The detrimental effect of alcohol on immune function is thought to be due to free radical generation related to acetaldehyde. By extension, then, will the consumption of red wine reduce the detrimental effects of alcohol on immunity by virtue of its antioxidant polyphenolics? Consider, first, the data that show an effect of polyphenols on the modification of immune cell function. Although the studies showing modification of immune cell function have been largely done with isolated compounds, the research suggests that flavonoids can alter immunity.

Flavonoids have been studied in relation to NK cell function. A dose-dependent antitumor activity against solid tumors was exhibited using a type of flavone in mice. This was attributed to a stimulation of cytokine gene expression caused by the flavone. Specifically, mRNA for interferon α was upregulated. This was an *ex vivo* experiment, so it is not known if naturally occurring flavones show the same effect or if the food itself is potent enough to alter cytokine expression.

Flavonoids commonly found in red wine have been shown to be antiallergenic. Certain flavonoids were found to inhibit the stimulated release of mast cell, eosinophil, and basophil granular components, with quercetin being particularly active. Antiviral activity was demonstrated by the inhibition of viral infection of cells and inhibition of the replication of virus, once the cell was infected. Different flavonoids have different capabilities. Antiinflammatory activity was demonstrated by the inhibition of neutrophil activation via inhibition of the NADPH oxidase activity. Inhibition of phospholipase A_2 was also demonstrated, resulting in an inhibition of the release of arachidonic acid and reduced production of prostaglandins and leukotrienes. These activities have not been demonstrated in studies utilizing the wine itself.

A study was conducted in 4-week-old mice in which they consumed one of four beverages put into their drinking bottles as the only fluid available to them. They drank a Muscadine wine, a Cabernet Sauvignon wine, ethanol, or water for 8 weeks. The Muscadine, Cabernet, and ethanol were adjusted to 6% alcohol, and it was shown that the mice drank an equal volume of fluid without any differences in body weight or food consumption among the groups. The amount of wine or ethanol consumed by the mice was equivalent to two 178-ml (6-oz) glasses of wine for a human when based upon caloric consumption. This model utilized a moderate consumption of the food itself rather than isolated compounds or pharmacological doses.

Some aspects of immune function were impaired in the mice that consumed ethanol, but not in those that consumed wine. Ethanol consumption decreased the basal numbers of NK cells and T-lymphocytes. However, the mice that consumed the same amount of ethanol as wine had basal levels of NK cells and T-lymphocytes equal to the water-consuming controls.

Mice were stimulated with a low level of lipopolysaccharide (LPS) in order to induce an immune response. We asked, first, if ethanol would reduce

the response and, second, if the red wine would prevent the reduced response. In control mice, 24 h after injection of LPS, blood lymphocyte numbers were reduced due to emigration and turnover. However, in the mice that consumed ethanol, the T-lymphocyte numbers were the same before and after the LPS. The mice that consumed either of the red wines showed a normal emigration of T-lymphocytes after LPS injection.

NK cell numbers, however, increase after LPS, likely due to the reduced numbers of other lymphocytes from the blood and/or possibly due to the migration of NK cells from the bone marrow. The mice that consumed ethanol had the lowest increase in NK cell numbers, whereas the mice consuming the red wine had an increase equal to that of the animals that drank water. These data suggest an impaired immune response to the LPS when even a moderate amount of ethanol was consumed. The mice that consumed wine as their sole beverage responded identically to the mice that consumed water. Perhaps free radicals generated during alcohol metabolism damaged the immune response, but the polyphenolic compounds in the red wine prevented the damage. It was also hypothesized that another mechanism could be functioning, i.e., the alteration in the phase I and phase II enzymes involved in alcohol degradation.

Free radical damage due to ethanol consumption was not found in any of these groups. Liver and lymphocyte glutathione levels and thio barbituric acid-reacting substances (TBARS) were not significantly different among the groups, stimulated or nonstimulated. The data suggest that immunity can be compromised without any overt signs of free radical damage.

Cardiovascular Disease

The health benefits of red wine related to cardiovascular disease have been documented in epidemiological studies, as well as experimental animals and humans. An excellent review of this was presented by German and Walzem in the Annual Reviews of Nutrition (2000).

Research has shown that moderate alcohol intake, regardless of the type of beverage, lowers cardiovascular mortality risk. The mechanism is related to the elevations in high-density lipoprotein (HDL) cholesterol. The rise in HDL cholesterol accounts for nearly half of the reduction in risk for death due to heart disease. Ethanol also stimulates liver lipogenesis, resulting in increased plasma triglycerides, increased lipoprotein lipase, thereby increasing triglyceride-rich lipoproteins, and increased expression of ApoAI and ApoAII, which together have a net effect of increasing the production of HDL.

The other nonalcoholic components of wine have also been shown to modify physiological cardiovascular health. One of the initial factors related to plaque formation in heart disease is the oxidation of low-density lipoprotein (LDL). Once oxidized, the molecule is rapidly taken up by macrophages that then take on the appearance of 'foam cells.' These foam cells modify the endothelium and contribute to the formation of blockage. Many *ex vivo* studies have measured the oxidation rates of LDL in the presence of wine phenolics and shown it to be reduced. It is not clear how much or to what extent the LDL is protected *in vivo*.

In addition to the protection against LDL oxidation, the polyphenols and phenolic acids in wines affect endothelial cells, platelets, lymphocytes, and macrophages through changes in cell signaling. The phenolic compounds in wine, as well as the ethanol, have been shown to reduce platelet aggregation. In addition, phenolics and ethanol have both been shown to improve endothelial function, as measured by vasodilation, vasorelaxation, or cellular adhesion molecule expression. Again, the extent to which this happens *in vivo* is not known.

Cancer

One of the few studies that used wine itself in the experimental design showed a reduction in cancer frequency. Mice consumed commercial wines or 12% ethanol with or without ethyl carbamate at 10 or 20 mg per kilogram of body weight for 41 weeks. Liver and lung were examined, and the incidence of tumors (percentage of mice with tumors) and the frequency (number of tumors per tumor-bearing mouse) were evaluated. Although the incidence of liver tumors was not consistently affected by the treatments, the frequency of tumors was significantly less for the wine-drinking animals than for the water- or ethanol-drinking animals. Similar results were found for the incidence and frequency of lung adenomas.

Moderate wine consumption in humans was associated with protection against hydrogen peroxide-induced DNA damage, a preliminary event related to cancer. Four volunteers drank 300 ml (9 oz) of red or white wine. Blood samples were collected before drinking and 1, 3, 8, and 24 h after drinking. Hydrogen peroxide-induced micronucleated cell frequency was determined in fresh lymphocytes that were incubated in the donor's homologous serum. Less damage occurred when the lymphocytes were incubated with their respective plasma 1 h after consumption, but not before consumption or at 8 or 24 h, suggesting that the protective effects are transient due to turnover of the bioactive compounds.

HTLV-1 transgenic mice that develop clinical and pathological features resembling human neurofibromatoses were fed dehydrated and dealcoholized wine (wine solids). The wine solids diet contained 0.04% polyphenols. Mice were examined daily for tumors, and the endpoint was age of tumor onset (latency). The age at which the first tumor appeared in the control mice and that in the experimental mice were 71 ± 4 days and 102 ± 10 days, respectively in the first experiment and 63 ± 4 and 86 ± 6 days in the second. They calculated the wine solid concentration at approximately one-quarter of that of a human consumption of 1 g of polyphenols per day. Catechin was the major polyphenol measured in the wine.

Other Diseases

Many other chronic diseases exist that are thought to be due to free radical damage. For example, Parkinson's and Alzheimer's diseases, diabetes, arthritis, inflammatory disorders, radiation injury, reperfusion injury, and macular degeneration as a few examples are thought to result from the accumulated damage caused by free radicals. No data exist to suggest that red wine consumption will prevent or alter the course of these diseases.

Mechanisms

Antioxidation

The mechanisms by which the compounds in red wines are thought to provide benefits related to risk reduction of chronic disease are via the antioxidant capabilities of the polyphenolic compounds. In general, antioxidants work by different mechanisms. Firstly, a free radical is defined as a compound missing an electron. Polyphenols, and antioxidants in general, can donate electrons to the free radical to prevent the free radical from obtaining electrons from the lipids of cell membranes, proteins, or nucleic acids. Secondly, the phenolics in red wine, shown by *in vitro* studies, can chelate metals, particularly iron and copper, which participate in the generation of free radicals. Thirdly, the phenolics in red wine can spare or help in the regeneration of other antioxidants. Flavonoids have been shown to conserve endogenous vitamin E contained in the LDL molecule.

The relative strengths of the flavonoids in their ability to prevent free radical damage by any of the above mechanisms is dependent upon the system used to study it. Numerous systems exist, in both aqueous and hydrophobic phases. The radical generated in the procedure as well as the detection system are different and can result in different effects. Some generalizations of structural requirements can be drawn to signify relative strengths of the polyphenols. Conjugation between the A and B rings in the form of a 2,3 double bond, a 3′,4′ di-hydroxyl configuration of the B ring, a 3-hydroxyl group in the C ring, and a 5-hydroxyl group in the A ring is highly significant for the radical scavenging function. Glycosylation of the polyphenols as well as sulfation and methylation may result in changes to the relative antioxidant strength. As stated earlier, the absorption and relative bioavailability are factors in assessing relative antioxidant strength. Much is yet to be done to understand the structure–function relationship of the thousands of polyphenolic compounds. Although the mechanisms, components, and availability are not yet clear, consumption of red wine is known to increase the antioxidant capacity of the plasma.

Cell Signaling

Flavonoids and therefore, by extrapolation, bioactive red wine components produce other metabolic changes that could be interpreted as beneficial. Components found in red wine are known to inhibit the arachidonic acid cascade via inhibition of phospholipase A2, lipoxygenase, and cyclooxygenase, thus accounting for potential antiinflammatory activity. Growth inhibitory activity via, for example, p53, tyrosine kinase, or polyamine metabolism have been shown to result from bioactive compounds found in red wine. Very little work has been done *in vivo* or with the wine itself.

Microsomal Detoxification Enzymes

Changes in the levels of Phase I and Phase II microsomal detoxification enzymes are associated with both beneficial and harmful effects. Inhibition of the Phase I microsomal detoxification enzymes is associated with drug toxicity due to a reduction in drug metabolism. Activation of the phase I enzymes is associated with drug tolerance or the conversion of innocuous compounds to toxic or carcinogenic compounds. Induction of phase I enzymes has been associated with the promotion stage of carcinogenesis. Potential phase I enzymes that may be involved in carcinogenesis include cytochrome P450 2E1 (CYP2E1), CYP1A1, CYP1A2, and CYP3A4. Alcohol is known to induce CYP2E1.

Induction of phase II enzymes is associated with anti-initiation processes related to cancer. Enzymes such as glutathione *S*-transferase (GST), epoxide hydrolase, *N*-acetyltransferase, and quinone reductase have been linked to beneficial anticancer activities.

Different dietary agents result in distinct alterations in the expression in phase I and II enzymes. For example, butylated hydroxy toluene (BHT), ethoxyquin, indol-3-carbinol, and phenethyl isothiocyanate induced both the phase I and the phase II isozymes and are therefore referred to as bifunctional agents. Caffeic acid, garlic oil, and propyl gallate induced phase II enzymes, but not phase I. 4-Methyl catechol, α-tocopheral, and red wine decreased phase I and induced GST. The type of red wine used in this study was not defined. The ratio of the phase I/phase II is more important in understanding protection or toxicity as opposed to the levels of one or the other. If the relative increase in phase II enzymes is greater than the phase I enzymes, then the assumption is that it is overall protective. If phase I enzymes increase relatively greater than phase II, then the potential for damage, by free radicals or production of toxic compounds, is greater. Alcohol consumption is known to induce phase I CYP2E1, resulting in production of acetaldehyde. Acetaldehyde, whether generated by this mechanism or by alcohol dehydrogenase, is potentially damaging due to its participation in free radical generation.

In one of the few articles that used the wine itself, Chan *et al*. showed that red wine solids, but not white wine solids, inhibited CYP3A4. In the study by Percival and Sims, liver microsomal cytochrome CYP2E1 and GST were altered by wine consumption, and Muscadine wine resulted in different patterns relative to the Cabernet Sauvignon. The Muscadine wine reduced CYP2E1 about 80%, whereas the Cabernet had no effect. Although the GST activity was also reduced in the mice that drank Muscadine, the ratio between the two enzymes was increased threefold, suggesting a beneficial effect of the Muscadine wine. GST activity was no different in the ethanol group or the Cabernet group compared with the water-consuming group, suggesting that there are unique components in the Muscadine wine that resulted in the observed changes. Moreover, the mice that drank the Cabernet Sauvignon did not have an extensive change in the microsomal detoxification enzymes, signifying a unique response of the mice to the Muscadine wine.

Ellagic acid has been shown to increase phase II enzymes GST and quinone reductase in mice. It has also been shown to inhibit metabolic activation of carcinogens. For example, ellagic acid inhibits the conversion of polycylic aromatic hydrocarbons, nitroso compounds, and aflatoxin B1 into compounds that cause genetic damage. Ellagic acid was also shown to inhibit CYP1A1. The anticarcinogenic effectiveness of ellagic acid was due to several different structural aspects of the molecule.

Summary

Red wine consumption results in an increase in antioxidant levels in the blood. Data on benefits of the individual compounds are emerging, but the study of red wine itself is limited. Health benefits can be demonstrated in epidemiological and animal studies regarding heart disease, but information is less solid regarding cancer, and nonexistent for other disease. The data cannot be ascribed solely to the phytochemical content of the wine since moderate alcohol intake is known to have benefits, both physiologically and psychologically.

The mechanisms by which the phenolic/polyphenolic compounds confer benefits are via their antioxidant activity, intracellular signal modification, alterations in gene expression, and alterations in microsomal detoxification enzyme activity. Other functional modifications that occur due to phenolics/polyphenolics, such as antimutagenic, anticarcinogenic, or antiinflammatory activity are derived from one of these major mechanisms.

The contribution of red wine to health can only be understood in the context of the whole diet. In populations that consume limited amounts of fruits and vegetables, as in Denmark, the consumption of wine is likely a significant contributing factor of phenolics and other phytochemicals and therefore more likely to provide benefits. In other populations, such as those that consume a Mediterranean-style diet, red wine is less likely to contribute to the overall health and disease risk reduction due to the consumption of olive oil, fruits, and vegetables. Recommendations to consume red wine for its beneficial effects may be imprudent, due to the health problems associated with overuse. However, if a person drinks alcohol sensibly, red wine has an advantage due to its contents of beneficial compounds.

See also: **Antioxidants**: Role of Antioxidant Nutrients in Defense Systems; **Bioavailability of Nutrients**; **Cancer**: Epidemiology; **Coronary Heart Disease**: Etiology and Risk Factor; **Immunology of Food**; **Phenolic Compounds**; **Tannins and Polyphenols**

Further Reading

Auw J, Blanco V, O'Keefe S and Sims CA (1996) Effect of processing on the phenolics and color of Cabernet Sauvignon, Chambourcin, and Noble wines and juices. *American Journal of Enology and Viticulture* 47: 279–286.

Bravo L (1998) Polyphenols: chemistry, dietary sources, metabolism, and nutritional significance. *Nutrition Reviews* 56: 317–333.

Cao G, Sofic E and Prior RL (1997) Antioxidant and prooxidant behavior of flavonoids: structure–activity relationships. *Free Radical Biology and Medicine* 22: 749–760.

Clifford AJ, Ebeler SE, Ebeler JD *et al*. (1996) Delayed tumor onset in transgenic mice fed an amino acid-based diet supplemented with red wine solids. *American Journal of Clinical Nutrition* 64: 748–756.

Fremont L (2000) Biological effects of resveratrol. *Life Sciences* 66: 663–673.

German JB and Walzem RL (2000) The health benefits of wine. In: McCormick DB, Bier DM and Cousins RJ (eds) *Annual Reviews of Nutrition*, pp. 561–593. Palo Alto, CA: Annual Reviews.

Manthey JA and Buslig BS (1998) *Advances in Experimental Medicine and Biology*. New York: Plenum Press.

Middleton E and Kandaswami C (2000) Plant flavonoid modulation of immune and inflammatory cell functions. In: Klurfeld DM (ed.) *Nutrition and Immunity* , vol. 8, pp. 239–266. New York: Plenum Press.

Percival SS and Sims CA (2000) Wine modifies the effects of alcohol on immune cells of mice. *Journal of Nutrition* 130: 1091–1094.

Rice-Evans CA, Miller N, Bolwell P, Bramley P and Pridhan J (1995) The relative antioxidant activites of plant derived polyphenolic flavonoids. *Free Radical Research* 22: 375–383.

Stoewsand GS, Anderson JL and Munson L (1991) Inhibition by wine of tumorigenesis induced by ethyl carbamate (urethane) in mice. *Food and Chemical Toxicology* 29: 291–295.

Teissedre P and Landrault N (2000) Wine phenolics: contribution to dietary intake and bioavailability. *Food Research International* 33: 461–467.

Wine Tasting

A C Noble, University of California, Davis, CA, USA

Introduction

The phrase 'wine tasting' is most frequently associated with subjective determination of quality of wine by experts at competitions or wine fairs. Increasingly often, consumers are learning about wine through tastings conducted by wine societies; similarly, wine bars offer the opportunity to taste single glasses of wines. Although wines are informally 'tasted' throughout all phases of wine-making, for example during aging in oak barrels, analytical sensory methods must be used to determine if wines differ in flavor or to measure the effects of a specific treatment. In this article, standardized wine terminology and objective sensory methods for evaluating wine flavor will be introduced and contrasted with more subjective quality evaluations. (*See* **Sensory Evaluation**: Sensory Characteristics of Human Foods.)

Quality Evaluation

Assessment of quality is made as a composite response to the sensory properties of wine based on expectations for a given wine type. These expectations are a function of previous experience with wines. Thus, while quality judgments are required for wine shows, for example, they are subjective. Responses vary among individuals because of different expectations and preferences. Therefore, the most experienced, skilled, and sensitive of wine judges will have differences of opinion about wine quality. Considerable experience with wines is required to make meaningful wine-quality evaluations; nearly all quality scorecards or schemes for awarding medals at competitions assume that the judges have a concept of the 'standard' for each wine type being judged. (*See* **Sensory Evaluation**: Practical Considerations; Food Acceptability and Sensory Evaluation.)

One of the most widely used scorecards is the Davis 20-point system (**Table 1**). For both systems, each factor is assessed separately and judged relative to an existing internal standard for the class of wine being evaluated. For example, for a young white Riesling, points are subtracted for a dark gold color which suggests excessive oxidation in this generally very light-colored variety, whereas a dark gold in an older white Riesling may be consistent with a wine of that age, so no points are subtracted.

These types of scorecard serve as a training device by forcing judges to examine the same factors, but experienced judges seldom rate each category independently, using the specific weighting assignments. Instead, the attributes are confounded; dark color in a young white Riesling influences the judge to look for and generally subtract points for an oxidized aroma and flavor by mouth. Most experienced assessors tend to evaluate color, aroma, taste, and flavor by mouth and come up with a final score using their own criteria. Further, differences in preferences result in greatly differing overall quality evaluations. Judges who like varietally distinct Sauvignon blanc or Cabernet Sauvignon wines (e.g., wines with intense asparagus or bell-pepper aromas) give high-quality ratings for these wines, whereas judges who dislike the intensely vegetative aromas rate the wines as being defective. (*See* **Sensory Evaluation**: Appearance; Aroma; Taste.)

Wine Terminology

An alternative and more analytical approach to wine tasting is descriptive analysis. This can be done

informally by recording specific terms to describe a wine. However, when naive wine drinkers first try to describe wine flavor, they usually find it difficult to name specific and familiar flavor notes. Standardized terminology for wine evaluation has been developed which facilitates development of precise descriptions. A standardized terminology using analytical terms for evaluation of table wines is presented in a hierarchical system in the wine aroma wheel (Figure 1). (A similar scheme for evaluation of sparkling wine aroma has also been developed.) The most general terms are located in the inner tier, with increasingly specific terms placed in the second and third tiers. In construction of this word list, subjective, hedonic terms, such as 'rich,' 'well-rounded,' 'harmonious,' 'elegant,' and 'austere,' which may have meaning to an individual or group of people but which cannot be readily defined or standardized, were specifically excluded. Terms are grouped by similarity of the aromas where possible. To define the terms, reference standards can be easily prepared (Table 2).

Descriptive Analysis

In contrast to quality ratings of wine experts, which provide an overall impression of their preferences but do not provide description of the flavor or reasons for the quality assignment, descriptive analysis provides specific information about the wines, which is figuratively a 'photograph of the flavor.' Analytical sensory evaluations, such as descriptive analysis, are conducted under controlled conditions which minimize bias. For example, to exclude the influence of color, tests are conducted under red lights or using black glasses. Tulip-shaped glasses are used for all evaluations of aroma. The optimum shape for increasing intensity has recently been demonstrated to be a tall glass with a wide bowl and narrow mouth. (*See* **Sensory Evaluation**.)

In descriptive analysis, initially analytical descriptive terms or attributes, such as those in the wine aroma wheel, are identified which are necessary to describe and differentiate the wines. For each attribute, reference standards are provided which define the specific aroma or taste characteristics and provide a common point of reference for communication. Even if descriptive analysis is done informally by consumers, profiles of wine flavors can be developed which permit specific comparisons of the differences among wines. More formally, data obtained using trained judges under controlled conditions provide precise information about wine flavor. Judges are trained extensively in the use of these terms in discussion sessions and in formal scoring tests, with reference standards available to define each attribute clearly. The intensity of each term is then rated in each of the wines and the data are analyzed statistically to confirm that the judges are consistent and reproducible and to determine whether differences among wines are statistically significant. (*See* **Sensory Evaluation**: Sensory Rating and Scoring Methods.)

The information provided by this type of analysis can be illustrated briefly, first, by examination of results from a descriptive analysis of Napa Valley Cabernet Sauvignon wines made from vineyards with different soils. Trained judges rated flavor by mouth and aroma attributes using the reference standards described in **Table 3**. How soil type may have influenced wine sensory properties is seen in Figure 2 for two wines which were produced at the

Table 1 Point assignment of the Davis 20-point scorecard[a]

Characteristic	*Weight*
Appearance	2
Color	2
Aroma and bouquet	4
Volatile acidity	2
Total acid	2
Sugar	1
Body	1
Flavor	2
Astringency	2
General quality	2

[a]Explained in detail in Amerine MA and Roessler EB (1983) *Wines, Their Sensory Evaluation*. San Francisco: W.H. Freeman.
17–20, superior wines: must have some outstanding characteristic and no marked defect; 13–16: standard wines with neither an outstanding character nor defect; 9–12, wines of commercial acceptability but with a noticeable defect; 5–8, wines of below commercial acceptability; 1–4, completely spoiled wines.

Table 2 Composition of typical aroma reference standards

Term	*Composition*
Berry	1 blackberry, 15 g raspberries, 10 g strawberry jam, 10 g raspberry jam in 3 ml base wine
Vanilla	0.1 ml vanilla extract in 25 ml base wine
Butterscotch	0.2–0.3 ml butter extract in 25 ml base wine
Vegetative	0.2 ml asparagus brine, 0.4 ml black olive brine, 1.5 ml green-bean brine in 25 ml base wine
Bell pepper	0.1 g fresh bell pepper in 25 ml base wine
Cloves	3 whole cloves in 25 ml base wine
Soy	0.4 ml soya sauce in 2.5 ml base wine
Mint	Piece of eucalyptus leaf and 0.05 ml mint eucalyptus extract in 25 ml base wine

Base wine was a neutral, low-intensity red wine.
Data from Spears TA (1990) Evaluation of the effects of soil and other geographic parameters on the composition and flavour of Cabernet Sauvignon wines from the Napa Valley. Unpublished. MSc Thesis, University of California, Davis.

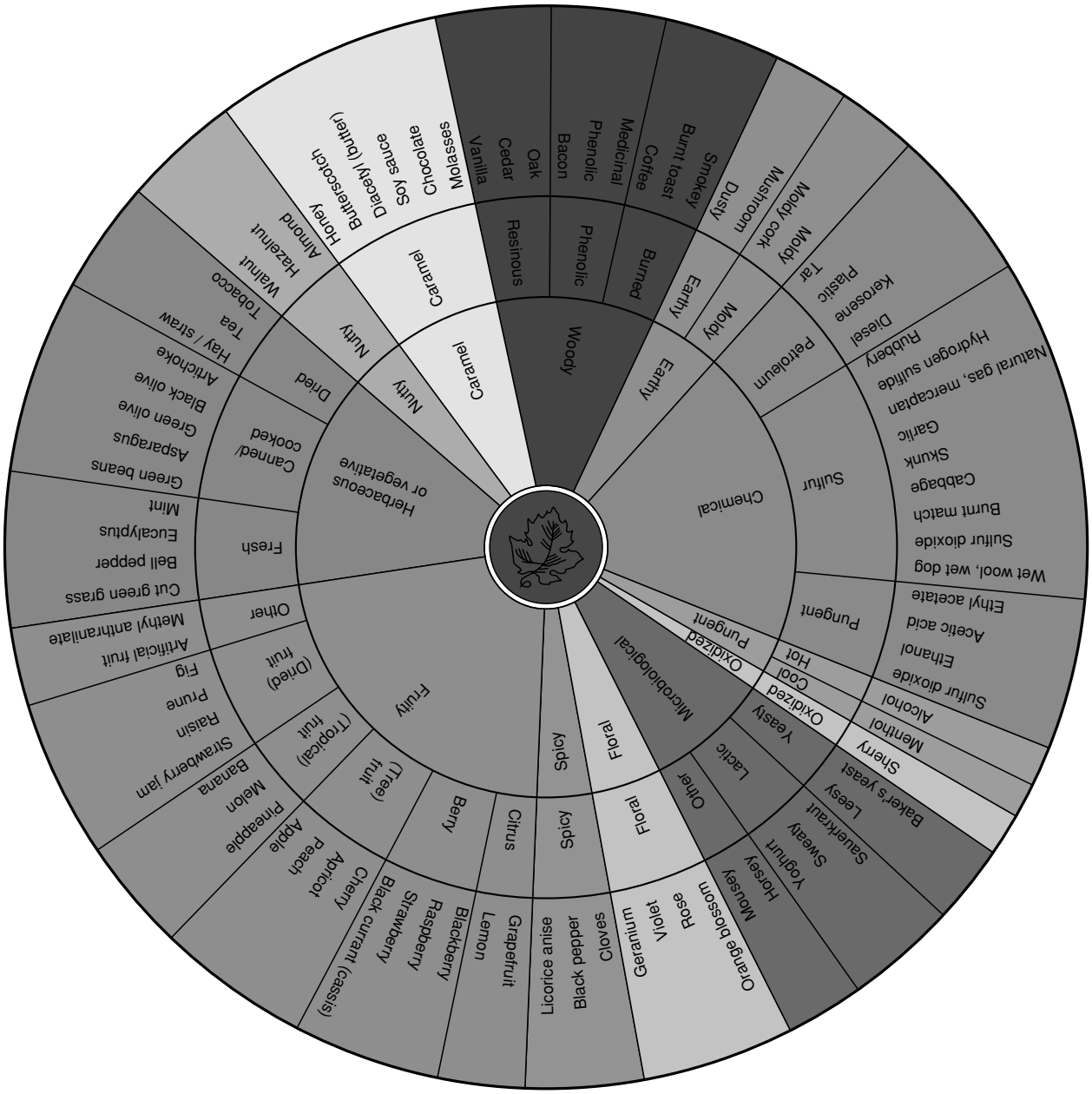

Figure 1 Wine aroma wheel. Standardized terminology showing most general terms (first tier) to most specific ones (inner tier). Copyright 1990, AC Noble, www.winearomawheel.com.

same winery from vines grown in very different soil types. In this polar coordinate graph, the center of the figure represents low intensity with the distance from the center to the intensity rating corresponding to the relative intensity of each wine for the attribute. By connecting the mean ratings, the profiles of wines are shown.

Wine 12 is characterized by a fruity (berry) aroma and flavor by mouth and is high in intensity of vanilla and butterscotch aromas. In contrast, wine 13 is low in these notes, and high in bell pepper, soy, and vegetative aromas and in vegetative flavor. Soil at site 12 is young, shallow, sandy, and nutrient-poor with a low water-holding capacity, whereas at site 13 there is a deep, clay-rich, older soil which is high in nutrients with a high water-holding capacity. Using a multivariate statistical method, the soil data were related to the sensory descriptive analyses for these sites and wines, demonstrating an association between wines with higher intensity of vegetative aroma and flavor by mouth and soils with higher water-holding capacity. Conversely, fruitier wines, high in berry aroma and flavor, were associated with older, gravelly soils with poor water-holding capacity.

As a second illustration of the application of descriptive analysis, 58 Chardonnay wines spanning three vintages were evaluated by descriptive analysis in 1985. For clarity, mean values for only two vintages are plotted in Figure 3. The younger wines from the 1983 vintage are higher in fruity (peach, citrus, and floral) aromas. The older wines from the 1981 vintage are lower in these fruity characteristics but higher in vanilla and bell-pepper aromas. Previously, it was claimed that younger wines were fruitier than older wines, but this study was the first quantitative evaluation to substantiate informal observations. More recently it was demonstrated that storage at high temperatures can decrease the fruitiness of white wines in less than 15 days. The intensity of citrus, floral, and green apple aromas decreased by 50% in Chardonnay wines stored at 40 °C for 15 days. After 30 days, the fruity notes were almost absent, while tea/tobacco, honey, and vanilla were the dominant aroma characteristics (Figure 4).

The 1981, 1982, and 1983 Chardonnays were also scored for quality by expert judges using the Davis 20-point scorecard (Table 1). No significant differences in quality scores were found across the 58 wines, but quality ratings for the vintages were statistically significantly different, although they were too small to be meaningful. (The oldest vintage was scored 14.80, and the youngest vintage the lowest, 14.08.) From examination of the aroma profiles of these vintages in Figure 3, it can be seen that higher-quality ratings were assigned to the older wines which were higher in vanilla (from oak aging) and lower in fruity notes which decreased over time because of chemical changes, such as hydrolysis of fruity esters to odorless or nonfruity compounds. Whether these differences in flavor resulted in the higher-quality ratings for the older wines, or factors such as the

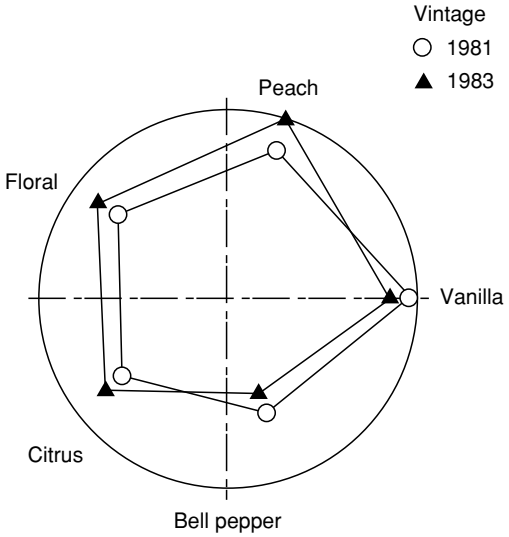

Figure 3 Comparison of mean aroma profiles for two vintages of California Chardonnay wine. Data from Ohkubo T, Noble AC and Ough CS (1987) Evaluation of California Chardonnay wines by sensory and chemical analyses. *Science des Aliments* 7: 573–587.

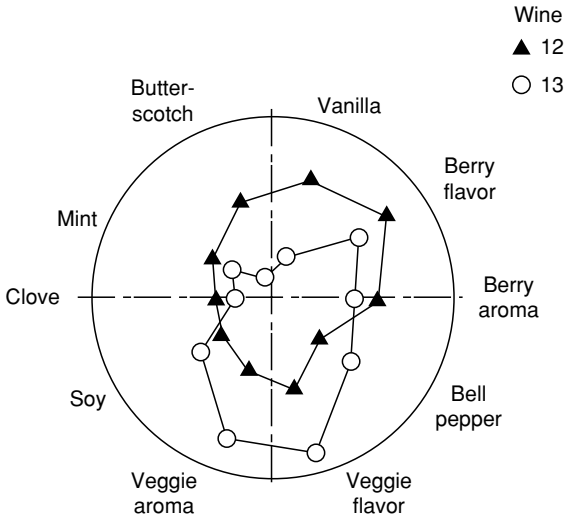

Figure 2 Comparison of flavor profiles of two Napa Valley Cabernet Sauvignon wines. At the origin, intensity = 0; at the perimeter, intensity = 4.0. By connecting the average intensities for each attribute, the profile for a wine is shown. Reproduced from Spears TA (1990) Evaluation of the effects of soil and other geographic parameters on the composition and flavor of Cabernet Sauvignon wines from the Napa Valley. Unpublished. MSc thesis. University of California, Davis.

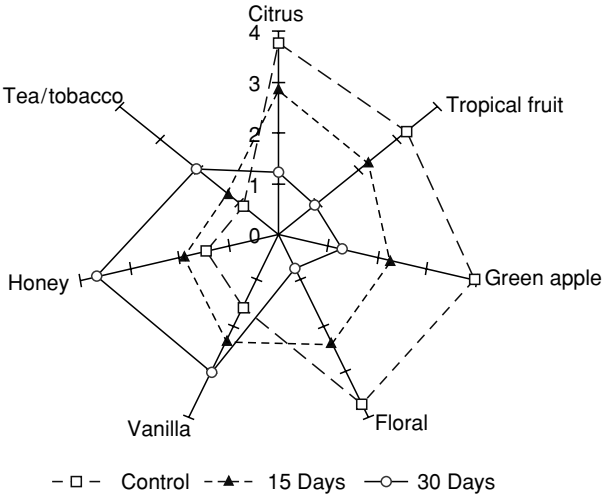

Figure 4 Comparison of aroma profiles of Chardonnay wines heated for 0 (control), 15, and 30 days at 40 °C. Adapted from de la Presa Owens C and Noble AC (1997) Effect of storage at elevated temperatures on aroma of Chardonnay wine. *American Journal of Enology and Viticulture* 48: 310–316, with permission.

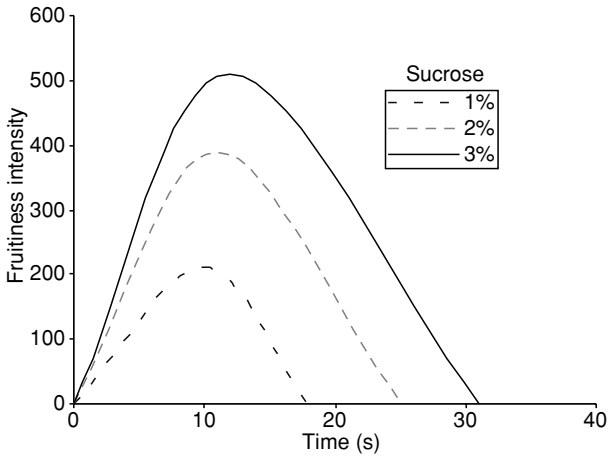

Figure 5 Average curve for intensity of the fruity flavor by mouth in white wine sweetened with 1, 2, and 3% sucrose. From Noble and Richards, unpublished data.

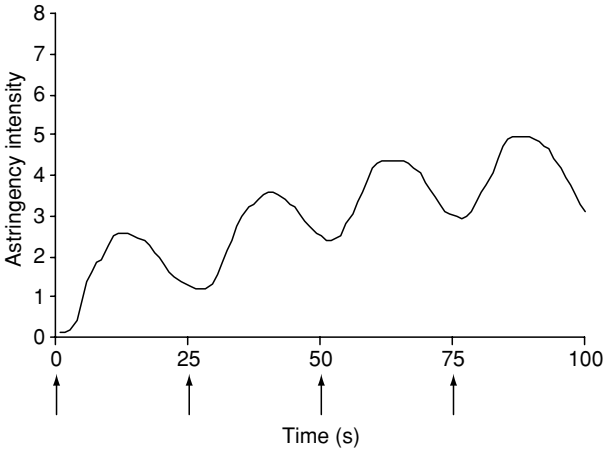

Figure 6 Average curve for intensity of astringency over time. Four sips of the same wine were taken at 25-s intervals. Arrows at time of sips. From Noble and LeDrean, unpublished data.

greater complexity of older wines influenced the experts, can only be speculated upon. However, this example illustrates the greater information provided by wine descriptive analysis than that provided by quality scores. (*See* **Sensory Evaluation:** Descriptive Analysis.)

Time–Intensity Evaluation

When wine is tasted, the flavor by mouth often lingers. To measure aftertaste, the method of time–intensity evaluation in which judges rate intensity continuously is useful. When a wine is sweetened, the fruitiness is perceived to increase. As shown in Figure 5, as the concentration of sugar was increased from 1% to 3%, the fruity flavor by mouth became more intense and lasted longer. Time–intensity analysis is useful for characterizing the temporal profiles of bitterness and astringency. Both are very persistent and increase in intensity upon repeated sips. For example, astringency intensity was rated continuously, while the tasters took four sips of the same wine at 25-s intervals. The average time–intensity curve shown in Figure 6 illustrates the build-up of astringency when red wine is sipped rapidly.

See also: **Sensory Evaluation**: Practical Considerations; Sensory Rating and Scoring Methods; Descriptive Analysis; Appearance; Aroma; Taste; Sensory Characteristics of Human Foods; Food Acceptability and Sensory Evaluation

Further Reading

Amerine MA and Roessler EB (1983) *Wines, Their Sensory Evaluation*. San Francisco: W.H. Freeman.

de la Presa Owens C and Noble AC (1997) Effect of storage at elevated temperatures on aroma of Chardonnay wines. *American Journal of Enology and Viticulture* 48: 310–316.

Fischer U (2001) Practical applications of sensory research: effect of glass shape, terroir and yeast strain on wine flavor. *American Journal of Enology and Viticulture* 51 (in press).

Noble AC (1988) Analysis of wine sensory properties. In: Linskens HF and Jackson JF (eds) *Wine Analysis*, vol. 6. *Modern Methods of Plant Analysis*, pp. 9–28. London: Springer-Verlag.

Noble AC (1994) Application of time–intensity procedures for the evaluation of taste and mouthfeel. *American Journal of Enology and Viticulture* 46: 128–133.

Noble AC (1998) Why do wines taste bitter and feel astringent? In: Waterhouse AL and Ebeler S (eds) *Chemistry of Wine Flavor*, pp. 156–165. ACS Symposium Series no. 714. Washington, DC: American Chemistry Society.

Noble AC, Arnold RA, Buechsenstein J *et al*. (1987) Modification of a standardized system of wine aroma terminology. *American Journal of Enology and Viticulture* 38: 143–146.

Noble AC, Elliott-Fisk DL and Allen MS (1995) Vegetative flavor and methoxypyrazines in Cabernet Sauvignon. In: Rousell RL and Leahy MM (eds) *Fruit Flavors: Biogenesis, Characterization, and Authentication*, pp. 226–234. ACS Symposium Series 596. Washington, DC: American Chemical Society.

Ohkubo T, Noble AC and Ough CS (1987) Evaluation of California Chardonnay wines by sensory and chemical analyses. *Sciences des Aliments* 7: 573–587.

Rankine B (1986) Roseworthy develops new wine scorecard. *Australian Grapegrower and Winemaker* Feb: p. 16.

WORLD HEALTH ORGANIZATION

J Akré, World Health Organization, Geneva, Switzerland

The Birth of the World Health Organization

The World Health Organization (WHO; Figure 1) is the culmination of efforts at international health cooperation that began 150 years ago. Although people have exchanged remedies and diseases for many thousands of years, only recently has thought been given to ways of working together to promote health that could go beyond purely parochial concerns. Early attempts at international cooperation in health were limited to small groups of countries, which discussed a few obviously contagious diseases such as cholera and smallpox, and strategies such as quarantine to keep them at bay. Although 11 international sanitary conferences were held in Europe between 1851 and 1903, it was not until 1907 that a world-wide international institution – the Office International d'Hygiène Publique – was founded to prepare and administer international sanitary conventions and to provide national health administrations an opportunity for regular contact and discussion.

The terrible epidemics that raged in Europe at the end of World War I, coupled with the mass movement of liberated prisoners of war, constituted a menace of such magnitude as to require coordinated international effort. The League of Red Cross Societies, created in 1919, attempted the task but quickly realized that intergovernmental action was essential to cope with a problem of such magnitude. As the charter of the Office International d'Hygiène Publique did not give enough power for action in individual countries, provision for necessary measures had to be made by the League of Nations, then in the process of creation.

The Geneva-based League, during its short and unhappy history between the two world wars, had been the first to invoke international health cooperation to deal with many kinds of health problems. In the same period, the Pan American Sanitary Organization, originally established in 1902, continued to work in its own geographic sphere. (In 1958, it became the Pan American Health Organization, which also serves as WHO's Regional Office for the Americas.)

In 1945, the United Nations (UN) Conference on International Organization, meeting in San Francisco, unanimously approved a proposal by Brazil and China to establish an autonomous international health organization within the UN system. The following year, a conference held in New York set up an interim commission and approved the *Constitution of the World Health Organization*, which came into force on 7 April 1948. Since then, 7 April is celebrated every year as World Health Day, when attention all around the globe is focused on a theme of major international public health importance.

The Task Entrusted to WHO

The agreed policy of WHO is a structured effort by all countries to bring health within the reach of everyone. The Organization's objective is the attainment by all peoples of the highest possible level of health. Health, as defined in WHO's Constitution, is a state of complete physical, mental, and social well-being, and not merely the absence of disease or infirmity. It is seen as a shared responsibility, calling for a high degree of self-reliance from the individual, the family, the community, and the nation as a whole. Because the determinants of health are so broad, the efforts of the health sector must be supported and augmented by those of many other sectors, including agriculture, water and sanitation, finance, communications, planning, and education. In carrying out its activities, the WHO Secretariat focuses on six core functions:

- Articulating consistent, ethical, and evidence-based policy and advocacy positions.

Figure 1 The WHO emblem.

- Managing information by assessing trends and comparing performance; setting the agenda for, and stimulating, research and development.
- Catalyzing change through technical and policy support, in ways that stimulate cooperation and action and help to build sustainable national and intercountry capacity.
- Negotiating and sustaining national and global partnerships.
- Setting, validating, monitoring, and promoting implementation of norms and standards.
- Stimulating development and testing of new technologies, tools, and guidelines for disease control, risk reduction, health care management, and service delivery.

WHO also proposes conventions, agreements, and regulations, and makes recommendations about international nomenclature of diseases, causes of death, and public health practices. It develops, establishes, and promotes international standards concerning foods and biological, pharmaceutical, and similar substances.

Structure and Management

WHO is a UN specialized agency, as provided for in the *Charter of the United Nations*. A goal-oriented organization with policies, programs, and budget defined through well-developed mechanisms, WHO consists of three constituent bodies.

The *World Health Assembly*, which is the highest decision-making body, is held in May of each year, and is attended by delegations from WHO's 191 Member States and two Associate Members. Its main function is to determine the policies of the Organization; it also appoints the Director-General and reviews and approves the budget.

The *Executive Board*, comprising 32 members designated by Member States, meets twice a year. It functions as the executive organ of the Assembly, prepares the agenda for each session, and submits to it a general program of work. The Assembly elects countries entitled to designate members of the Executive Board, and one-third of the membership is renewed annually.

The *Secretariat* is headed by the Director-General who, as the chief technical and administrative officer of the Organization, is appointed for a 5-year term by the World Health Assembly on the nomination of the Executive Board. Dr Gro Harlem Brundtland (Norway) took office on 21 July 1998. Past Directors-General have been Dr Brock Chisholm (Canada, 1948–1953), Dr Marcolino Candau (Brazil, 1953–1973), Dr Halfdan Mahler (Denmark, 1973–1988), and Dr Hiroshi Nakajima (Japan, 1988–1998).

In general, all technical activities that are of universal applicability – such as biological standardization, the overall assessment of the efficacy of methods and materials, and promoting the control of diseases – are led from *headquarters* in Geneva, Switzerland (Figure 2). WHO's decentralized structure enables it to respond directly to the needs of its membership, upon request, through its six regions – Africa, the Americas, South-east Asia, the Eastern Mediterranean,

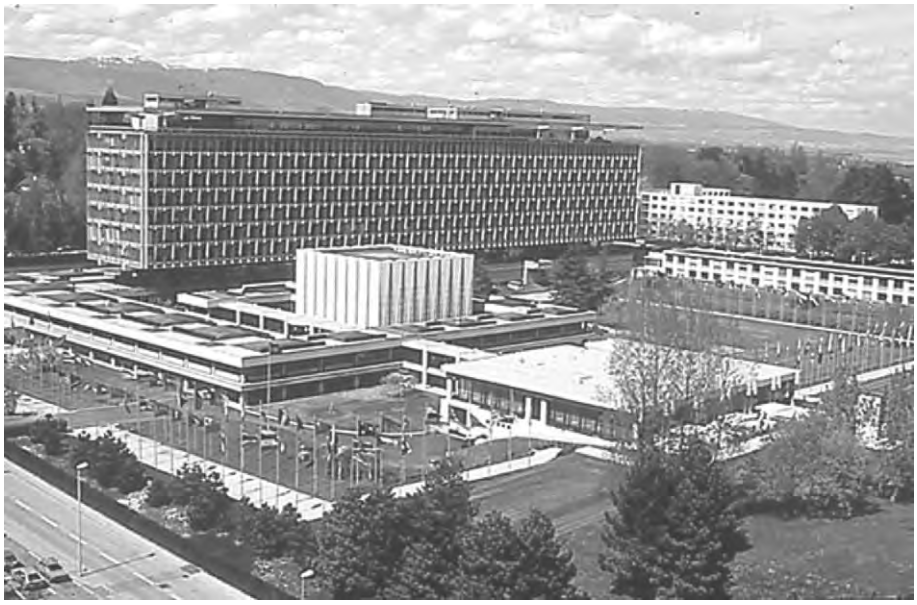

Figure 2 General view of the WHO headquarters in Geneva, Switzerland (photograph by J. Germain, WHO).

Europe, and the Western Pacific – each consisting of a *regional committee* and a *regional office* (Figure 3). The regional offices, with their own directors, are responsible for formulating policies of a regional character and for monitoring regional activities. In many countries, there is a resident WHO representative, who is the main intermediary for WHO support, and who participates, with the government, in planning and managing national health programs.

Some 27% of WHO's approximately 4300 staff members work in countries all over the world, either in field programs or as WHO representatives; 44% are in the regional offices, and 29% at headquarters.

The WHO's standard-setting functions include preparation and updating of the *International Classification of Diseases*, assignment of generic names for pharmaceuticals, and development of the authoritative scientific knowledge base – methodologies, standards, norms and criteria, and guidelines and strategies – for preventing, detecting and managing the major forms of malnutrition. Since 1957, evaluation of the safety for human consumption of selected food additives and contaminants in food and establishment of acceptable daily intakes for these substances has been the focus of the Joint FAO (Food and Agriculture Organization of the United Nations)/WHO Expert Committee on Food Additives. The committee's reports, as well as those of a similar FAO/WHO group responsible for evaluating the safety of pesticide residues, are used in formulating national food legislation intended to protect consumers from hazardous additives or contaminants and by the Codex Alimentarius Commission – another joint FAO/WHO body – in establishing international food standards. Food legislation is one of the many topics covered by WHO's *International Digest of Health Legislation* database.

Apart from its regular budget, WHO receives voluntary contributions from both governmental and nongovernmental sources. The regular and projected extra-budgetary figures for the biennium 2002–2003 are US$855.6 million and US$1.380 billion, respectively. Voluntary contributions include subaccounts for purposes such as developing a new international growth reference based on an international sample

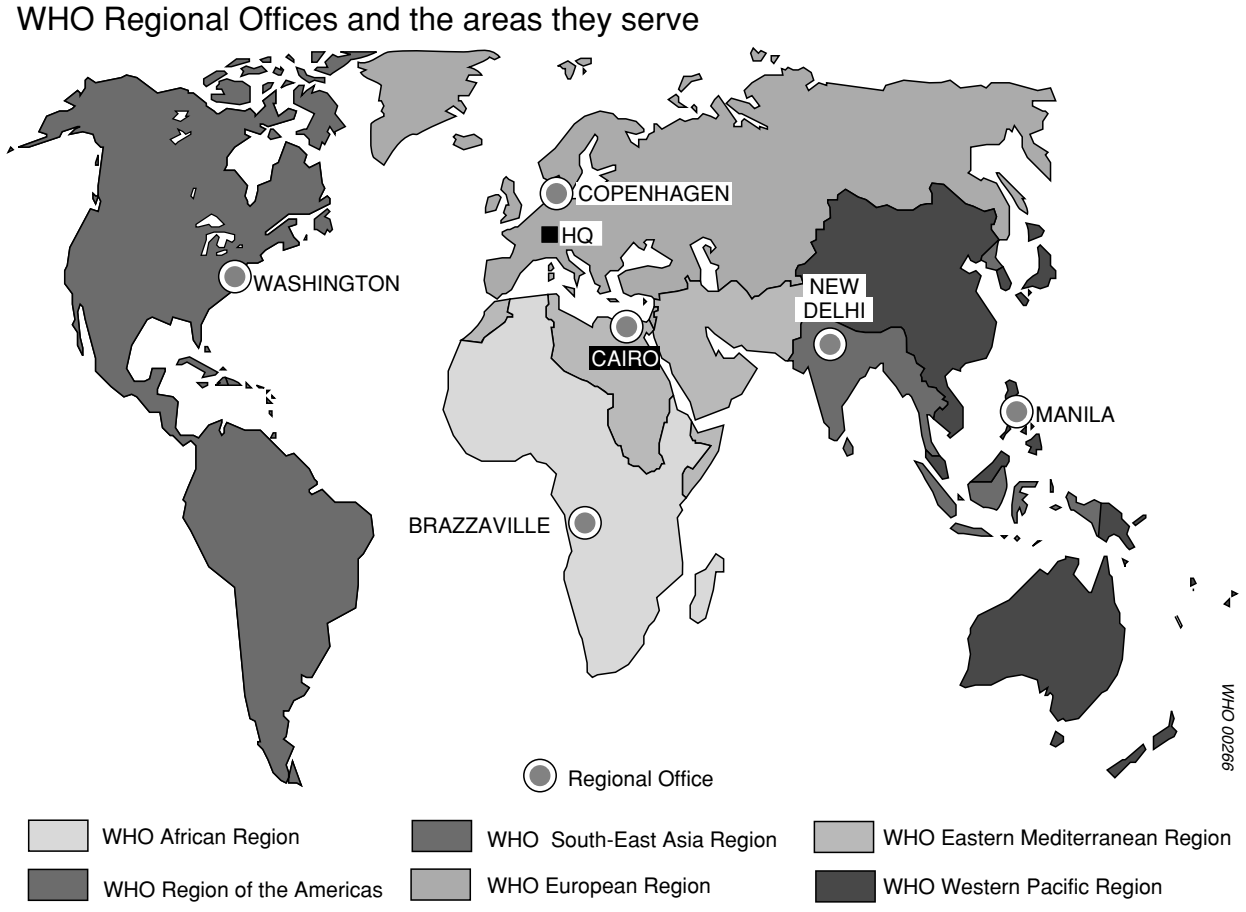

Figure 3 WHO regional offices and the areas they serve.

of breastfed infants from healthy populations with unconstrained growth; fostering health research; improving community water supply; expanding immunization; eradicating polio; preventing and controlling HIV/AIDS, diarrheal disease and yaws; preparing a credible emergency health response to disasters and natural catastrophes; and responding to new and re-emerging diseases, e.g., viral hemorrhaghic fevers like Ebola, and tuberculosis (see below concerning public–private partnerships).

Working Together for Better Health

From its beginning, WHO set out to work not through its small staff alone but with and through others. Many thousands of individual researchers and scientists, including Nobel laureates, have put their talents at the Organization's disposal – and their number continues to grow. The same is true of the nearly 1200 leading health-related institutions around the world that have been officially designated as WHO collaborating centers in a wide range of disciplines.

WHO was also mandated to work with other agencies within the UN family of organizations. Nutrition work, for example, quite naturally came to involve FAO, as did work against animal diseases and, as mentioned above, in the area of food additives, pesticide residues, and contaminants. WHO cosponsored, with FAO, the International Conference on Nutrition (Rome, 1992) when 159 governments reached a new global consensus on the nature of malnutrition and its multiple causes, and on the urgent action required to prevent, reduce, and ultimately eliminate malnutrition. Programs aimed at educating teachers on how to deal with mentally retarded children came within the purview of the United Nations Educational, Scientific and Cultural Organization (UNESCO). Occupational health was a shared activity with the International Labour Organization. Drug dependence and abuse called for collaboration with the United Nations Division of Narcotic Drugs. WHO, together with the United Nations Children's Fund (UNICEF), the United Nations Development Programme, the United Nations Population Fund, UNESCO and the World Bank, was a founding cosponsor of UNAIDS, the global effort to control the spread of HIV/AIDS.

One of WHO's closest partners has been UNICEF, with which, for example, the early yaws and malaria campaigns were carried out, and it prepared the International Code of Marketing of Breast-milk Substitutes (1981). In 1991, the two agencies jointly launched the Baby-friendly Hospital Initiative to promote breastfeeding, and they continue to collaborate in the development of a global strategy for infant and young child feeding.

Collaboration was also initiated with professional, charitable, and other nongovernmental organizations (NGOs) pursuing aims consonant with those of WHO. By the end of the first decade, WHO had established official relations with 40 such bodies, ranging from the International Council of Nurses to the International Commission on Radiation Units and Measurements, and from the World Federation for Mental Health to the International Leprosy Association. Work of vital importance has been made possible through the enthusiasm of these and other valuable organizations – for example, the Bill and Melinda Gates Foundation – which have in turn benefited from the moral support and the technical information provided by WHO.

Collaboration continues unabated; examples include cancer pain relief with the International Association for the Study of Pain, efforts to insure safe blood and plasma supplies undertaken with, among other groups, the International Federation of Red Cross and Red Crescent Societies, and support for polio eradication provided through Rotary International. The success of these joint ventures is illustrated by the ever-lengthening list of NGOs admitted into official relations with WHO, which now numbers 192 with the recent acceptance of Timor-Leste.

To further its mission and policies, WHO also regularly interacts with commercial enterprises, including through participation in alliances and other relationships, exchange of information, product research and development aimed at improving health, generation of cash and in-kind donations to WHO, and advocacy for health. While stressing the need to be fully transparent and to avoid conflict of interest arising in any relationship – work with the private sector to achieve health outcomes is governed by formal guidelines – the Executive Board has encouraged new forms of partnerships and other relationships in order to advance WHO's health agenda. A number of groundbreaking public–private partnerships – WHO, government agencies, NGOs, and the pharmaceutical industry – have been recently developed to tackle several particularly challenging diseases, including leprosy, malaria, and trachoma.

World Health in the Twenty-first Century

Using pragmatic tools in pursuit of its visionary goal, WHO strives to respond to its constitutional mandate through its evolving program of activities. These are based on the needs and priorities determined by Member States themselves through the World Health Assembly and the Executive Board. Changes in program emphasis occur in response to altered political, social, economic, and environmental realities.

For example, in 1977, the World Health Assembly decided that the major social goal of governments and WHO should be the attainment by all people of the world by the year 2000 of a level of health that would permit them to lead a socially and economically productive life. This was the birth of the Health for All movement, which was followed, in 1981, with the adoption of the Global Strategy for Health for All. In 1994, WHO's Member States, while acknowledging that significant global changes had occurred since the adoption of the goal of Health for All, called for a renewal of the strategy to meet challenges, exploit opportunities, and overcome obstacles at the dawn of the twenty-first century.

Respecting the new issues and priorities of the times for WHO means grappling with the multiple consequences for health and nutritional status of accelerating social and demographic changes, including population growth, ageing, international movements of migrants and refugees, and complex emergencies resulting from natural and ecological disasters, wars, civil disturbances, and serious disease outbreaks. It calls for dealing creatively with the impact of indebtedness, incurred over time, that has led many countries to reduce public spending in health services, which are often regarded as being only an expenditure rather than an investment in human potential. There continues to be a massive gap between the resources needed to help poor communities tackle different causes of illness – malnutrition, malaria, tuberculosis, and HIV/AIDS being among the most urgent examples – and the funding and human capacity currently available to them.

WHO must also face up to threats to human health that respect no national boundaries, such as uncontrolled dumping of toxic wastes and pollution of land, water and air, and prodigious consumption and mismanagement of natural resources. The globalization of infectious diseases is not a new phenomenon. However, increased population movements; growth in international trade in food and biological products; social and environmental developments linked with urbanization; deforestation and climate change; and modification of food-processing methods, distribution, and consumer habits have reaffirmed that infectious disease events in one country are potentially a concern for the entire world. The solution of these problems hinges, to a large extent, on the degree to which WHO is able to harness the interdependence of nations and peoples as a positive force for conciliating between competing present and future needs.

Maintaining that health should be seen as central for development and to the quality of human life, WHO seeks to achieve a dynamic yet harmonious balance between health in terms of consumption and health as an investment in human capital with dividends earned in the future. The Organization's history has been one of adaptation to evolving circumstances, giving priority first to one health problem, then to another, and it continues to adjust its priorities to meet the world's ever-changing health requirements. At the same time, WHO is being asked to respond to unprecedented requests for support in addressing the problems brought about by dramatically declining resources, political upheaval, war and civil strife, together with the pandemic of AIDS, epidemics of cholera and bovine spongiform encephalopathy (and its human form, Variant Creutzfeldt–Jakob Disease), environmental degradation, as well as natural disasters such as earthquakes, floods and drought. In 2001, WHO opened an office in Lyon (France) to help with the technical training of specialists from all over the world who work to control epidemics. WHO's Tobacco Free Initiative was created to focus international attention, resources and action on the global tobacco pandemic that kills four million people a year today. WHO is leading the way with the Framework Convention on Tobacco Control – the world's first multilaterally negotiated rules devoted entirely to a major health issue – which will be ready for signature no later than 2003.

In this light, and as a reflection of the decisions taken during the World Health Assembly, the Executive Board, and the regional committees, WHO has designated its priorities for the start of the twenty-first century. These include an approach that focuses particularly on links between health and poverty reduction; a greater role in establishing consensus on health policies, strategies and standards; more effective action to improve health and decrease inequities in health outcomes; and an organizational culture that encourages strategic thinking, global influence, prompt action, creative networking, and innovation. To pursue these ends, the work of WHO has four strategic directions:

- reducing excess mortality, morbidity, and disability, especially in poor and marginalized populations;
- promoting healthy lifestyles and reducing factors of risk to human health that arise from economic, social, and behavioral causes;
- developing health systems that equitably improve health outcomes, respond to people's legitimate demands, and are financially fair;
- developing an enabling policy and institutional environment in the health sector and promoting an effective health dimension to social, economic, environmental, and development policy.

The Foundation of Nutritional Well-being

Nutrition is a cornerstone of health and development; it affects and defines the health of all people, rich and poor. It is a fundamental pillar of human life, health, and development across the entire life span and an essential foundation for human and national development. From the earliest stages of fetal development, at birth, through infancy, childhood, adolescence, and on into adulthood and old age, proper food and good nutrition are essential for survival, physical growth, mental development, performance and productivity, health, and well-being. Conversely, malnutrition spells increased vulnerability to disease and premature death.

Malnutrition means 'badly nourished,' but it is far more than just a reflection of what people eat, or fail to eat. Clinically, malnutrition is the result of inappropriate intake and absorption of protein, energy, and micronutrients, whether too little or too much, and by frequent infections or disease. Nutritional status is the result of the complex interaction between the food we eat, our overall state of health, and the social and physical environment in which we live – in short, food, health and caring, the three 'pillars of well-being.' Good health is as essential to nutritional well-being as good nutrition is crucial for maintaining healthy growth and development. Preventing infection and managing infectious diseases – minimizing their incidence, duration, and severity – are essential for optimizing nutrition.

The fundamental WHO goal of Health for All – that people everywhere have the opportunity to reach and maintain the highest attainable level of health – is impossible in the presence of hunger, starvation, and malnutrition, which remain among the most devastating problems facing the majority of the world's poor and needy and continue to dominate the health of the world's poorest nations. Indeed, poverty, inequality, and underdevelopment will not be truly resolved until both malnutrition and ill-health are effectively addressed.

Malnutrition: Casting Long Shadows

Nearly 30% of the world's population, especially in low-income countries, bears the triple burden of poverty, malnutrition, and ill-health, which stalk one another in a vicious circle. From a human-development perspective, poverty compromises health and nutritional status, and undermines socioeconomic development. In a world that has both the resources and knowledge to end this catastrophe, the situation is a continuing travesty of justice and an abrogation of the fundamental human right to adequate food and nutrition, and to be free from hunger and malnutrition. In addition to suffering from one or more of the multiple forms of malnutrition, 800 million people – many of whom depend on food aid for their survival – are threatened by food insecurity. Malnutrition kills, maims, cripples, and blinds on a massive scale world-wide; and it is both a major cause and effect – as well as a key indicator – of poverty and underdevelopment.

The point here is as simple to state as it is complex to remedy – proper nutrition is a prerequisite both for insuring good health and for reducing poverty. Where hunger and malnutrition prevail, unnecessarily high levels of morbidity and mortality, and unacceptable levels of economic and social deprivation are sure to be found. Thus, for example:

- *Poverty* underlies most of the world's *malnutrition*, with attendant inadequate and insecure food supply, inappropriate feeding practices and care, nutritional emergencies, and widespread infection and infestation that are compounded by lack of health services.
- *Nutritional status*, determined using nutritional indices, is a reliable and easily measurable objective indicator of poverty and progress achieved in national development.
- *Maternal malnutrition* remains a major factor for the 30 million infants born each year (23.8% of all births) with *intrauterine growth retardation* leading to retarded physical, mental, and intellectual growth, and heightened risk of infectious diseases and death.
- *Protein–energy malnutrition* affects more than a quarter of the world's children – 26.7% (150 million) underweight and almost 30% (161 million) stunted – of which 70% are in Asia, over 25% in Africa, and 4% in Latin America and the Caribbean.
- *Malnutrition* contributes to 60% of the 10.9 million *deaths* each year among children under five in developing countries. Well over two-thirds of these deaths, which are often associated with poor feeding practices, occur during the first year of life.
- Some 740 million children and adults in 130 countries are affected by *iodine deficiency disorders*, still the greatest single cause of preventable brain damage to the fetus, infant, and young child.
- *Vitamin A deficiency* affects more than 250 million preschool children in 118 countries, mainly in Africa and South-east Asia, causing blindness and increased risk of infection and death.
- Vast problems of *iron* and *folate deficiency*, resulting in anemia, affect more than 60% of

women of childbearing age in developing countries, and millions of young children.
- *Noncommunicable diseases* contribute to approximately 60% of deaths in the world and 43% of the global burden of disease. About half of these deaths are attributable to *cardiovascular diseases*.
- Major natural and human-induced *emergencies* of all types are increasing in number and intensity, further compromising the care and feeding of vulnerable populations, including refugees and internally displaced persons who currently number more than 40 million (including 5.5 million under-five children).

Responding to the Challenge of the Twenty-first Century – A Two-Pronged Approach

To promote healthy nutrition for all people, it is necessary to take simultaneous action on two different, if overlapping, fronts. On the one hand, many less-favored nations remain handicapped by a formidable array of development constraints, including rapidly increasing population, unproductive agriculture, environmental degradation, limited health service coverage, and war and civil strife. Among the most visible – and tragic – consequences are the many millions of wasted and stunted children who do not have enough protein and energy in their diets, who suffer from cretinism and other permanent brain damage because their diets and those of their parents are deficient in iodine, or who go blind or even die for lack of vitamin A. It is in just such environments that diarrheal disease, frequently compounded by seasonal or chronic food shortages, take their heaviest toll in terms of malnutrition, ill-health, and premature death.

On the other hand, there has been a significant drop in recent years in the prevalence of infectious diseases, while food availability and the quality of diets have improved for population groups all over the world. The result in many countries has been a sharp reduction in infant and child mortality and longer adult life expectancy. These and related factors have paved the way for a dramatic expansion of a different type of nutrition crisis – diet-related chronic disorders are now flourishing where, not so long ago, infectious diseases were seen as the greatest menace to health. Chronic disorders are fast becoming major killers, and their prevalence is especially high in rapidly industrializing low- and middle-income countries among the urban poor who pursue the unhealthy dietary habits and lifestyles that, until recently, were considered the province of more affluent groups.

The rapidly increasing burden of noncommunicable diseases is a key determinant of global public health. On the basis of current estimates, deaths from noncommunicable diseases are expected to account for 73% of deaths and 60% of the disease burden by the year 2020. Already 79% of deaths attributed to noncommunicable diseases occur in developing countries, where most people affected by these diseases are between 45 and 65 years old. In China and India alone, the burden of cardiovascular disease is greater than that in the industrialized countries as a whole. One of the main objectives of WHO's global strategy for the prevention and control of noncommunicable diseases is to reduce the level of exposure to the major risk factors, namely, tobacco use, unhealthy diet, and physical inactivity, which should be tackled in an integrated manner.

WHO is committed to reducing malnutrition's global burden and its associated morbidity and mortality. The overall aim of its work in nutrition is to promote healthy nutrition and appropriate feeding for all population groups based on sound scientific evidence. WHO seeks to prevent, reduce, and eliminate malnutrition world-wide, thereby promoting sustainable health and well-being, accelerating poverty reduction, and enhancing human and national development. Within this context, WHO's seven priority action areas are:

- *protein–energy malnutrition* – assessment, monitoring/management, prevention, and reduction;
- *micronutrient malnutrition:*
  - Elimination of vitamin A deficiency and iodine deficiency disorders;
  - Significant reduction of iron deficiency anemia;
- *obesity* and other *diet-related diseases* – epidemiology, prevention and management;
- *national nutrition policies and programs* – development and implementation;
- *infant and young child feeding* – appropriate breastfeeding and complementary feeding;
- *nutrition in emergencies* – preparedness and response;
- *food aid for development* – guidance for food-assisted development activities.

Against this backdrop, WHO strives to:

- strengthen the technical capabilities of its Member States for assessing and coping with malnutrition;
- develop authoritative standards, norms, criteria, methodologies, guidelines and strategies for preventing, detecting, and managing all major forms of malnutrition;
- promote the health and nutrition benefits, for food-insecure populations, of food-assisted development and emergency activities;

- monitor, evaluate, and report on the world's major forms of malnutrition and progress achieved in combating them.

Given malnutrition's multicausal nature and multisectoral dimensions, WHO collaborates extensively with many multilateral (United Nations) and bilateral (national) development agencies (see examples above), and international scientific advisory bodies and NGOs. A world-wide network of 28 distinguished collaborating centers in nutrition also undertake a wide range of nutrition training and research activities.

In view of nutrition's foundational importance for health and development, WHO focuses on priority issues in basic nutritional science, nutritional care throughout the life span from infancy to old age, and nutrition policies and programs for sustainable development. Today's most urgent priority in tackling these vast nutritional challenges is for WHO to focus its combined normative and collaborative strength, particularly through its technical outreach in regions and countries, to strengthen the ability of all its 191 Member States to effectively prevent, control, reduce, and, ultimately, eliminate malnutrition in all its forms.

See also: **Anemia (Anaemia)**: Iron-deficiency Anemia; **Carotenoids**: Occurrence, Properties, and Determination; **Folic Acid**: Properties and Determination; **Food and Agriculture Organization of the United Nations**; **Infants**: Breast- and Bottle-feeding; **Iodine**: Iodine-deficiency Disorders; **Legislation**: Codex; **Malnutrition**: The Problem of Malnutrition; Malnutrition in Developed Countries

Further Reading

International Digest of Health Legislation database, http://www.who.int/idhl/
Joint FAO/WHO Expert Committee on Food Additives http://www.who.int/pcs/jecfa/jecfa.htm
Joint FAO/WHO Expert Committee on Pesticide Residues http://www.who.int/pcs/jmpr/jmpr.htm
Monitoring Chemical Contaminants in Food http://www.who.int/fsf/fdstxt1.htm#monitoring
Regional Office for Africa, Harare, Zimbabwe http://www.whoafr.org/
Regional Office for the Americas, Washington, DC, USA http://www.paho.org/
Regional Office for the Eastern Mediterranean, Cairo, Egypt http://www.who.sci.eg/
Regional Office for Europe, Copenhagen, Denmark http://www.who.dk/
Regional Office for South-East Asia, New Delhi, India http://www.whosea.org/
Regional Office for the Western Pacific, Manila, Philippines http://www.who.org.ph/
WHO Department of Nutrition for Health and Development, http://www.who.int/nut/
WHO Global Database on Child Growth and Malnutrition http://www.who.int/nutgrowthdb/
WHO Department of Child and Adolescent Health and Development http://www.who.int/child-adolescent-health/
WHO Programme on Cardiovascular Diseases http://www.who.int/ncd/cvd/index.htm
WHO Food Safety Programme http://www.who.int/fsf/
WHO (1988) *Four Decades of Achievement: Highlights of the Work of WHO*. Geneva: WHO.
WHO (1996) *Physical Status: The Use and Interpretation of Anthropometry. Report of a WHO Expert Committee. World Health Organization Technical Report Series 854*. Geneva: WHO.
WHO (1998) *Complementary Feeding of Young Children in Developing Countries: A Review of Current Scientific Knowledge*. Geneva: WHO.
WHO (1998) *Preparation and Use of Food-based Dietary Guidelines. Report of a Joint FAO/WHO Consultation. World Health Organization Technical Report Series 880*. Geneva: WHO.
WHO (1998) *HIV and Infant Feeding. Guidelines for Decision-makers: A Guide for Health Care Managers and Supervisors; A Review of HIV Transmission through Breastfeeding (WHO/FRH/NUT/CHD/98.1)*. Geneva: WHO.
WHO (1999) *Management of Severe Malnutrition: A Manual for Physicians and Other Senior Health Workers*. Geneva: WHO.
WHO (2000) *The Management of Nutrition in Major Emergencies*. Geneva: WHO.
WHO (2000) *Obesity: Preventing and Managing the Global Epidemic. Report of a WHO Consultation. World Health Organization Technical Report Series 894*. Geneva: WHO.
WHO (2000) *Complementary Feeding: Family Foods for Breastfed Children (WHO/NHD/00.1)*. Geneva: WHO.
WHO (2000) *Nutrition for Health and Development: A Global Agenda for Combating Malnutrition. Progress Report (WHO/NHD/00.6)*. Geneva: WHO.
WHO (2000) *Foodborne Disease: A Focus for Health Education*. Geneva: WHO.
WHO (2001) *The Optimal Duration of Exclusive Breastfeeding: A Systematic Review (WHO/NHD/01.08)*. Geneva: WHO.
WHO (2002) *Programming of Chronic Disease by Impaired Fetal Nutrition. Evidence and Implications for Policy and Intervention Strategies (WHO/NHD/02.3, WHO/NPH/02.1)*. Geneva: WHO.
WHO (2002) *Keep Fit for Life: Meeting the Nutritional Needs of Older Persons*. Geneva: WHO.
WHO (2002) *Nutrient Adequacy of Exclusive Breastfeeding for the Term Infant During the First Six Months of Life*. Geneva: WHO.
World Health Organization, Geneva, Switzerland http://www.who.int/

WORLD TRADE ORGANIZATION

S D Kalyankar and S P Kalyankar, Marathwada Agricultural University, Parbhani, Maharashtra, India
C D Khedkar and A S Khojare, College of Dairy Technology, Warud (Pusad), Maharashtra, India

Emergence of WTO

As per the proposal made by the USA, the world economy organized around the International Monetary Fund and International Bank for Reconstruction and Development. In 1947, the General Agreement on Tariffs and Trade (GATT) was created and signed by 23 countries on 30 October at the Palais des Nations in Geneva and came into force in 1948 with its first meeting at Havana, Cuba. Thereafter, GATT completed seven rounds. The eighth round, the largest trade round (1986–1993), was known as the Uruguay Round (UR). The UR proposals had covered agriculture in a comprehensive manner for the first time, since the creation of GATT in 1947. The problems of stagnation in agricultural trade were presented in the UR at Punda del est in 1986 and articulated in the Dunkel proposals in December 1991. The Dunkel proposals were eventually ratified, and these form part of the rules of the World Trade Organization (WTO).

Objectives

WTO is an international body dealing with the rules of the international trade. The basis for the organization are the agreements and negotiations signed by the bulk of the world's trading nations, which are essentially contracts, binding governments to keep their trade policies within agreed limits. It aims to facilitate trade flow as freely as possible, as long as there are no undesirable side-effects, and serves as a forum for trade negotiations and as a dispute settlement body. It encourages countries to enter into free trade gradually through negotiation by promoting fair competitions and trade without discrimination.

Structure

Membership

WTO currently has 132 member countries. Thirty-two countries are currently negotiating for accession. These include China and the Russian Federation, and a number of developing countries.

Decision-making System

WTO has a three-tier system of decision-making. The decisions are made at three levels: ministerial conference, general council and other councils, and heads of delegations.

The top-level decision-making body is the ministerial conference, which meets at least once in a year and can make decisions on all matters under any of the multilateral trade agreements. WTO has organized two ministerial conferences: Singapore (1996) and Geneva (1998).

In the intervening period between two conferences, the general council acts on behalf of the ministerial conference and carries out the day-to-day work of WTO. The general council of WTO, together with other councils for trade in goods, trade in services, and intellectual property, performs WTO functions. These councils are responsible for implementing WTO agreements in their respective areas of specialization.

The third tier of decision-making system is at the heads of delegation level. This is most effective in overcoming inherent impediments for reaching consensus on trade-related issues. This level is usually represented by different sets of people, depending on the need of the nation and their expertise and experience. The usual process is to hold a meeting of about 40 countries, which are most interested in a particular issue. Such meetings are termed 'green room meetings'. An example of the success of such a meeting is the market access negotiation. In the last 5 years of the operations of WTO, its growing importance and membership have proved to be effective in this new millennium, and will grow further with changing requirements. The WTO structure and decision-making process is adequately equipped to face future challenges.

Agreements

These rules are contained in the following legal instruments.

- Annex 1A. Multilateral Agreement on Trade in Goods.
- Annex 1B. General Agreement on Trade in Services and Annexes.
- Annex 1C. Agreement on Trade-related Aspects of Intellectual Property Rights.
- Annex 2. Understanding the Rules and Procedures Governing the Settlement of Dispute
- Annex 3. Trade policy Review Mechanism.
- Annex 4. Plurilateral Trade Agreement.

Member countries are under an obligation to ensure that their national legislation, regulations,

and procedures are in full conformity with the provisions of these agreements. The harmonized rules stipulate that there should be no unnecessary barriers to trade and that a country's exports are not disrupted by the sudden imposition of higher tarrifs or other barriers to trade.

Annex 1A. Multilateral Agreement on Trade in Goods

1. Agreement on Agriculture;
2. Agreement on application of sanitary and phytosanitary measures;
3. Agreement on technical barriers to trade;
4. Agreement on textiles and clothing;
5. Agreement on TRIMS;
6. Agreement on the implementation of Article VI of GATT 1994;
7. Agreement on preshipment inspection;
8. Agreement on rules of origin;
9. Agreement on import licensing procedures;
10. Agreement on subsidies and countervailing measures;
11. Agreement on safeguards.

Agreement on Agriculture (AOA)

AOA is a set of rules governing international agriculture policy development to liberalize agricultural trade and reduce the existing distortions in the sector, if any. The agreement itself has concessions and commitments of members regarding market access – removing various trade restrictions confronting exports, domestic support programs, and tackling the problem of making exports artificially competitive. The reform program requires that countries that have applied quantitative restrictions to import be permitted to add to existing import tariffs on the prices of imported products. This process has come to be known as 'tariffication.' Under the reform program, the countries, which granted subsidies to their farmers will have agreed to reduce by agreed percentages domestic subsidies that distort trade and export subsidies.

General Agreement on Trade in Services (GATS)

Services cover a wide range of economic activities from banking, insurance, and telecommunications to recreation, cultural, and sporting services. One of the main characteristics of services is that they are intangible and invisible. These characteristics influence the modes in which international trade transactions in goods and services take place. While international trade in goods involves the physical movement of goods from one country to another, relatively few service transactions involve cross-border movements.

Trade in services is growing and accounts at present for over 20% of international trade. GATS has created a framework for bringing this trade under international discipline. Its provisions apply to all modes of supply of service in international trade, viz.:

- cross-border trade in services;
- establishment of a commercial presence in foreign country;
- temporary movement of natural persons to another country to provide services there; and
- movement of consumers to the country of importation.

The agreement consists of a framework text, which lays down a set of general principles for measures affecting trade in services. The annexes applicable to specific sectors complement the text.

Agreement on Trade-related Aspects of Intellectual Property Rights (TRIPS)

Intellectual property right is the legally enforceable power to exclude others from using the information created or to set the terms on which it can be used. The objective of the agreement on TRIPS is to give adequate and effective protection to the intellectual property rights to reward creativity and inventiveness. TRIPS sets out minimum protection to be given to each category of intellectual property in the national laws of each WTO member. It lays down procedures and remedies to be provided by each country for intellectual rights enforcement. The agreement on TRIPS lays down minimum standards of protection that countries must provide for intellectual property rights. One of the other important features of the agreement is that it also requires countries to ensure effective enforcement of these rights. Its rules apply to:

- copyright and related rights;
- trademarks;
- patents;
- geographical indications;
- industrial designs;
- layout designs of integrated circuits; and
- undisclosed information.

See also: **Legislation**: History; International Standards

Wort *See* **Beers**: History and Types; Raw Materials; Wort Production; Chemistry of Brewing; Biochemistry of Fermentation; Microbreweries

Y

YEASTS

T Deák, Szent István University, Budapest, Hungary

Introduction

Humans have consumed cereals since prehistoric times. In Mesopotamia, baked pastes of ground grain were consumed, while, in Egypt, leavened bread was prepared. For thousands of years, humans were unaware of the existence of yeasts but were able to use them not only for breadmaking but also for brewing. The fundamental role of yeasts and bacteria in fermentation processes was only discovered in the last century by Pasteur. Soon after this, Hansen established the use of pure cultures for brewing and baking. Until that time, bakers used spent brewer's yeast for leavening, or seeded the fresh dough with a small part of already leavened dough. The commercial production of yeast for baking purposes started after about 1850, and since then, yeast technology has grown into a highly productive and economic industry. 'Yeast' in this article refers solely to the species *Saccharomyces cerevisiae*.

The Organism

Baker's yeast is a biotype of *S. cerevisiae* that can metabolize sugars both aerobically, producing the end products carbon dioxide and water, and anaerobically, producing ethanol and carbon dioxide. Thus, baker's yeast can be propagated in large quantities under aerobic conditions, and the cell mass added to dough can produce carbon dioxide under various conditions to leaven bakery products. These are the fundamental properties of baker's yeast that render it indispensable to the production of bread. Through this century-old application, many specific requirements have been imposed by the producer and the user, and strains have been selected with improved properties to meet these requirements. The strains used today are able to reproduce rapidly under strong aeration and limited nutrient supply. The yeast cells show remarkable tolerance to various storage conditions, and survive drying. Baker's yeast possesses strong fermentative activity and flavor development in various conditions of processing.

Production

There are at least four major steps in the manufacture of baker's yeast: preparation, fermentation, separation, and packaging (Figure 1).

Preparation

The manufacture of baker's yeast starts in two separate areas: in the laboratory with the propagation of a pure yeast culture, and in the factory with the preparation of the fermenters and the nutrient medium.

In the laboratory, where pure stock cultures are maintained, a small flask of sterile fresh culture is prepared after one or more subcultivations. This sample is then inoculated into the first pure culture tank. Two or three tanks of increasing capacity from 50 to 400 l may be applied. The yeast produced in an earlier stage is used to seed the next stage, while transfer is made under sterile conditions. In these early stages of propagation, the main concern is to maintain purity. The pure culture fermenters are fed with sterile molasses medium supplemented with the necessary growth factors, but aeration with filtered air is not at full capacity in the first batch fermentation.

Before the First World War, grain mash was used for the commercial propagation of yeast. During the war, shortages of grain led to its replacement by molasses, a relatively cheap byproduct of cane and beet-sugar production. Since then, molasses has become the traditional source of carbon and energy for yeast growth. It is usually fortified with a source of nitrogen, minerals and growth factors. Nitrogen is added in the form of ammonia, its salts, or urea; phosphorus is supplied in the form of phosphoric acid or ammonium phosphate. Depending on the

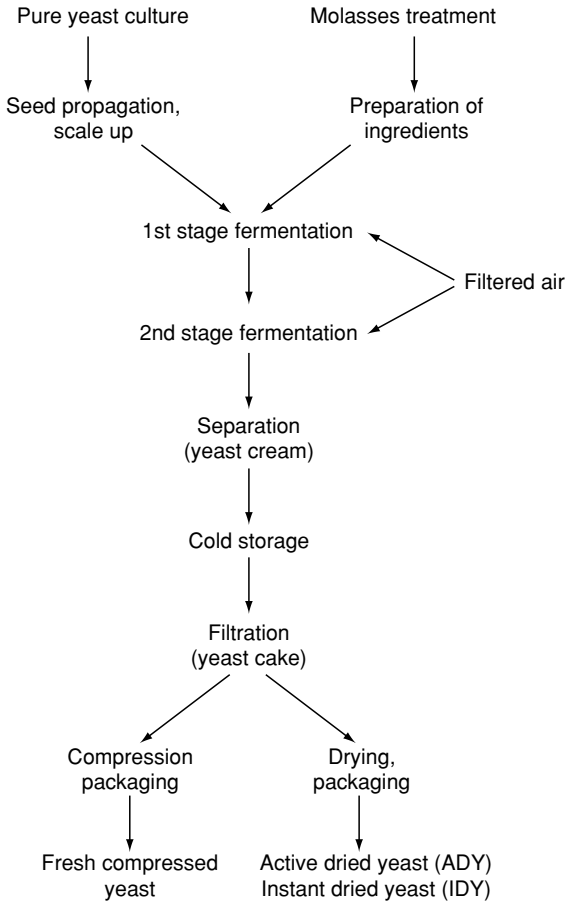

Figure 1 Flow chart for the production of baker's yeast.

composition of the raw molasses, certain growth factors, most frequently biotin, are also added to the wort. The concentrated molasses is diluted, purified, supplemented with other nutrients, and sterilized before use.

Fermentation

Baker's yeast is usually produced in a multiple-stage process. During scale-up, strong aeration and incremental feeding are introduced. Full-scale fermentation is conducted in large (100 m^3 or larger) tanks. There is a great variation in the size and shape of fermenters. However, in their design, an important requirement is to insure maximum aeration, because it is the oxygen transfer that is usually the rate-limiting factor in yeast propagation. Mechanical and sparger aeration systems are generally used. During the aerobic growth of yeast, a considerable amount of heat is liberated, and an efficient cooling system is also an integral part of the fermenter. The strict demands for hygiene determine the overall construction of the fermenter and the materials used therein. Facilities for cleaning in place are always integrated.

After cleaning and disinfection, the fermenter is fed with water, in which the pure seed yeast is suspended, then mixed with wort, and the propagation starts with vigorous aeration. Baker's yeast 'fermentation' is a typical fed-batch process in that, after commencing the propagation, nutrients are fed incrementally, maintaining at all times a very low sugar concentration at full aeration. The protocols for nutrient feed rate, temperature, pH, and aeration are specifically set up and strictly controlled to optimize yield, productivity, and product quality. Special attention is paid to prevent underaeration, which leads to excessive alcohol formation and a decrease in productivity. Instrumental process control and automation are necessary to produce baker's yeast economically. Adequate sensors and computer applications now make it possible to control the most sophisticated fermenter systems. Baker's yeast producers, however, have to consider the baking quality (stability and activity) of the product, which can be attained at the cost of productivity. As a satisfactory compromise, at the final stage of fermentation, nutrient feeding is stopped, and aeration is continued for about an hour. During this ripening period, the properties of baker's yeast are improved. Nitrogen starvation increases stability, but fermentative activity decreases. At the end of a typical fermentation, the yeast solid content may vary between 3 and 8%, which means a yield of about 20 000–30 000 kg of fresh yeast in one batch propagation at 28–30 °C for 12–18 h.

Efforts to introduce continuous fermentations on a commercial scale have remained unsuccessful. Although continuous systems can be maintained at a maximum yield, a good product quality can be achieved only with propagation regimes that do not easily lend themselves to continuous culture. Moreover, the problem of preventing contamination raises the cost and makes the process economically unfeasible.

Separation and Filtration

At the end of each fed-batch propagation period, the yeast cells are recovered from the spent medium by centrifugation. Water wash is applied between two passages through centrifugal separators. A yeast cream is obtained with 18–20% dry weight, which can be stored in agitated tanks at 2–4 °C for a few days without any loss of quality.

The yeast cream is further concentrated by filtration on rotary vacuum filters or filter presses. Filtration yields a yeast cake of about 27–30% dry matter content. (*See* **Filtration of Liquids**.)

Packaging

After filtration, the yeast cake is mixed with oils, emulsifiers, and a small amount of water, then compressed and extruded into blocks, or granulated for bulk distribution. The oil and emulsifiers improve product appearance and aid the formation of blocks (extrusion, cutting). (*See* **Emulsifiers**: Uses in Processed Foods.)

The Product

Compressed Yeast

This is the form of baker's yeast produced by the process outlined above. Compressed yeast is the traditional form of baker's yeast, which is available to wholesale bakers in 0.5–2.0-kg blocks, while smaller (10–50 g) blocks are prepared for households. Compressed yeast wrapped with waxed paper and stored at 4 °C keeps for a few weeks. A granulated pressed cake form of the same product packed in 10–20-kg bags can also be prepared for large-scale bakeries.

Compressed yeast cells are alive. They use their reserve carbohydrates (glycogen, trehalose) for energy to survive. Storage under refrigeration retards metabolism, and wrapping inhibits drying out. Under improper storage conditions, deterioration processes (autofermentation, autolysis) may start, resulting in heat build-up and loss of yeast activity.

Dried Yeast

In a dried form, baker's yeast possesses a longer shelf-life than compressed yeast. Dried yeast retains stability even when stored at room temperature. It offers benefits by reducing the cost of refrigeration, transport and storage, but drying increases the expenses of the manufacturer. The disadvantage for the user arising from the lower activity of heat-stressed cells is compensated by the ready availability of dried yeast. In all, there is an increasing preference and a growing share in the market for dried yeast. There are two forms of dried yeast. The first, introduced about 50 years ago, active dried yeast (ADY) needs to be rehydrated in warm water before use. Developed only in the last decade, instant dried yeast (IDY) does not require rehydration and can be mixed directly with flour in making dough. The early procedures of yeast propagation for drying are basically similar to those of the traditional baker's yeast fermentation. However, specific yeast strains selected to withstand drying stresses are used, and the final stages of propagation are set in order to increase yeast resistance to drying. The feeding schedule and maturation period are controlled to produce yeast with lower protein but higher trehalose and lipid contents.

Preparation of the yeast for drying begins with extrusion of the compressed yeast cake into slender strands (1–3-mm diameter) that are cut into short pieces. These particles are dried in a hot air current; the early procedure of tunnel drying has been mostly replaced by tumble or rotating driers and, increasingly, by fluidized bed driers. Only the latter are suitable for the production of IDY. Airlift driers employ a blast of hot air at a velocity sufficient to suspend the yeast particles in a fluidized bed. Air temperatures of 160 °C can be used for quick drying but, after loss of the free cell water content (at about 35% moisture), temperatures should not exceed 40 °C in order to minimize cell-membrane damage and reduction of enzyme activity. Cells are rapidly killed at temperatures exceeding 50 °C. (*See* **Drying**: Spray Drying.)

The moisture content of ADY ranges from 6 to 8%, whereas that of IDY is only between 4 and 6%. ADY possesses only one-third to one-half of the leavening activity of fresh compressed yeast (Table 1). The instant drying procedure allows the production of yeast with a leavening activity comparable with that of compressed yeast.

ADY can be stored without refrigeration. During storage, ADY loses its activity by 1% per month if packed under vacuum or under nitrogen. On storage in air at ambient temperature, the loss of activity is faster. In order to restore activity, rehydration of ADY should be carried out by adding warm (40 °C) water to the yeast in a 4:1 ratio. During rehydration, 20–30% leakage of intracellular materials occurs, which leads to a loss of fermentation activity. A

Table 1 Main characteristics of commercial baker's yeast products

Product	*Form*	*Method of drying*	*Dry (%)*	*Protein (%)*[a]	*Gas production (%)*[b]
Compressed yeast	Blocks or granules	None	30	52	100
Active dried yeast	Irregular spheres	Drum	94	40	45
	Irregular particles	Belt	94	40	40
Instant dried yeast	Short rods	Fluidized bed	96	48	80

[a]Nitrogen × 6.5.
[b]Related to the activity of compressed yeast.

further disadvantage of this leaching is that it releases reducing substances, such as glutathione, which may cause slackening in the dough.

IDY particles are highly porous and easy to rehydrate. This allows immediate use without prior rehydration. However, air may also access the cells, resulting in rapid oxidation and loss of activity. Hence, IDY must be packed under vacuum or in a nitrogen atmosphere, and must be used within a few days after opening the package.

Addition of a variety of agents to the yeast cake prior to drying can improve activity and stability of dried yeast. Emulsifiers (e.g., 1% sorbitan esters) facilitate rehydration of IDY, and antioxidants (e.g., 0.1% butylated hydroxyanisole) increase the stability of ADY. (*See* **Antioxidants**: Synthetic Antioxidants.)

Applications

The application of baker's yeast is indispensable to the production of leavened baked products, such as breads, rolls, pastries, doughnuts, etc. Recipes and technologies for these products vary world-wide, but the essence of the process is the same, in that after mixing flour, water, yeast, salt, and optional ingredients, the dough undergoes panary fermentation before baking. The primary role of baker's yeast in the baking industry will be illustrated using as an example the predominant product, white bread.

Breadmaking

Conventional breadmaking technology involves sponge dough. This dough comprises about two-thirds of the total flour mixed with water, salt, and yeast, and is left for a fermentation period of 4–5 h. The sponge is then added to the balance of flour, water, and all remaining ingredients and thoroughly mixed mechanically until it is transformed into a smooth dough. The characteristic rheological properties of the dough are due to the structure of gluten, a cross-linked network formed from wheat proteins and lipids. This allows the elasticity of dough to retain gas evolved by yeast and thus to leaven. (*See* **Wheat**.)

The dough undergoes a series of mechanical operations (divided into pieces, rounded, and moulded) while being allowed to rest between these procedures for short periods. During these proofing periods, fermentation proceeds, and leavening continues. After the final proof, loaves are placed into a hot oven for baking. Within the loaf, gas expands, steam and alcohol evaporate to form holes in the coagulated matrix of gluten, and the characteristic structure of the crumb sets. While the temperature in the center of the loaf remains below 100 °C, the surface reaches 140 °C, to form a hard, brown-colored crust. The baked bread is left to cool before the finishing operations (slicing, wrapping) and distribution. (*See* **Bread**: Breadmaking Processes.)

The conventional sponge dough technology requires about 8 h to finish, and several alternative methods have been developed to shorten this period (Table 2). In the straight dough method, all the ingredients are mixed at the start, and one bulk fermentation period of 2–4 h is allowed for leavening. In the short-time dough process, only 15–30 min are allotted for the dough to rest, and intense mechanical working brings about the structure of the dough. Time is also saved by the continuous mix processes, in which a ferment or brew is first prepared from yeast with little or no flour (liquid ferment), and after about 2 h of fermentation, the dough is mechanically developed in a continuous mixer. Bulk fermentation of the dough can be replaced by intense mechanical working and/or the addition of chemical improvers in other process variants. Improvements in equipment design have brought about savings in labor, better control and automation, effective sanitation, and greater processing flexibility of breadmaking technology.

Role of Yeast

Yeast plays three major functions in the dough: leavening, maturing, and flavor development.

Leavening The increase of dough volume is due to the production of carbon dioxide during yeast fermentation of the carbohydrates available in the flour. Dry flour contains approximately 1–8% fermentable sugars (glucose, fructose, sucrose), whereas maltose is produced from starch granules by wheat amylases after wetting the flour. (*See* **Flour**: Dietary Importance.)

Yeast has to adapt to the mostly anaerobic environment in the dough as well as to the fermentation of maltose after depletion of available free sugars. Yeast also has to tolerate a certain degree of osmotic pressure exerted by salt (and sugars, if added). The concentration of solutes is higher at the first stage of dough preparation when only half of the regular water is added. In certain formulae, sucrose or high-fructose syrup is used to sweeten the dough. Increasing the osmotic stress not only reduces the fermentation rate but also induces glycerol production.

Maturation To some extent, both yeast itself and its fermentation activity play important roles in developing the texture of the dough, called maturation. This involves complex changes, including

Table 2 Schematic comparison of breadmaking processes

Time (h)	Sponge dough	Straight dough	Continuous mix	Short-time dough
7	Mixing			
6	↓			
5	Sponge	Mixing		
4	↓	↓	Liquid ferment	
3	Mixing	Dividing Rounding	↓	
	↓ Dough	↓		
	↓			
2	Dividing Rounding 1st proof Molding	1st proof ↓	Mixing Developing Dividing Rounding	High-speed mixing Dividing Rounding 1st proof Molding
1	2nd proof ↓	2nd proof ↓	Mixing Proofing ↓	2nd proof ↓
0	Baking	Baking	Baking	Baking

the mechanical forces of mixing leading to gluten formation. Carbon dioxide developed during yeast fermentation would not produce gas cells without the ability of viscoelastic gluten films to retain the gas. In fact, it is the air bubbles preformed in the dough during mixing into which the carbon dioxide diffuses. The rheological properties of the dough are influenced by the fermentation products (ethanol, pH decrease) in a way not yet clearly understood. Reduced compound (e.g., glutathione) liberated from yeast cells may split the disulfide bonds between gluten molecules, leading to the cleavage of the gluten structure.

Taste and Flavor The characteristic and appealing bread aroma would not develop without yeast. It is difficult to characterize the complex nature of bread aroma and to determine the precise role of yeast fermentation in its development. More than 200 volatile compounds have been identified by gas chromatography, and many of these organic esters and acids, alcohols, and carbonyl compounds are formed as byproducts of yeast fermentation. Other compounds, such as amino acids, originate from the yeast cells. In addition to fermentation, bread aroma is determined by the process of baking, which leads to crust browning. This is a Maillard-type reaction, and its extent is influenced by the fermentation products of yeast and its cell constituents. (*See* **Browning**: Nonenzymatic; **Flavor (Flavour) Compounds**: Structures and Characteristics.)

Sour Dough

Before baker's yeast was available commercially, part of leavened dough was added as an inoculum to the fresh dough. Acidification normally took place in this old dough, hence the name: sour dough. In recent years, consumption of sour-dough breads has greatly increased. The use of sour dough is necessary for the development of characteristic properties of rye breads. Sour doughs contain both heterofermentative lactic acid bacteria and yeasts, and the mixed population usually comprises several different species of both groups. Unlike baker's yeast consisting overwhelmingly of a single yeast species, *S. cerevisiae*, even commercial sour-dough starters are not made of defined pure cultures. *Lactobacillus brevis* and *L. sanfrancisco* are characteristic lactic acid bacteria in sour dough, whereas *Candida milleri* and *S. exiguus* are the predominant yeast species.

Yeast Starter Cultures

While about 1.5 million tons of baker's yeast are produced annually throughout the world, the estimated production of yeast starter cultures is less then 1000 tons, although these are made commercially in many countries. Wine starter cultures are used to initiate the fermentation of must instead of a natural (spontaneous) process. The use of yeast starter cultures in wine-making has increased strongly over the last decade. Freeze-dried or active dry yeast starters are made of selected strains of various useful

technological properties, such as a tolerance of high concentrations of ethanol, sugar, and sulfur dioxide, low temperature, etc. Their use is also advantageous in sparkling-wine production.

The use of dried yeast in breweries is not common, although it may offer advantages such as yeast availability, flexibility, and cost-effectiveness in particular for producing specialty beers in small quantities.

Future Developments

Although baker's yeast is one of the oldest biotechnological products that is used in large quantities and has been the subject of basic biochemical and genetic studies, there have remained a number of unresolved questions about its physiology relevant to production and application.

The carbohydrate metabolism of *S. cerevisiae* is under strict regulation, and it has been long recognized that a high sugar concentration triggers alcoholic fermentation, even under fully aerobic conditions (the so-called 'Pasteur effect'). Hence, baker's yeast is produced with strong aeration and limited sugar addition in a fed-batch process. However, cultures adapted to these conditions are required to perform leavening in drastically different conditions prevailing in the dough, i.e., anaerobically and with sugars in excess. Moreover, when glucose and fructose present initially are exhausted, yeast cells have to adapt the metabolism of maltose originating from flour starch. Hence, baker's yeast with a good leavening capacity should have a high glycolytic activity, a high potential of maltose fermentation, and an ability to synthesize enzymes under anaerobic conditions to adapt rapidly to changing substrates. In attempts to improve these characteristics, overexpression of certain glycolytic enzymes, modulation of the balance between glycolysis and gluconeogenesis, and induction or derepression of maltose utilization enzymes have been targeted using genetic transformants, producing mutants and constructing hybrids. Recombinant DNA technology provides further ways to enhance biomass yield on molasses and alternative carbon sources. Molasses contains raffinose, of which only the fructose part can be utilized by baker's yeast. Brewer's yeast, a close relative, produces melibiase, permitting complete fermentation of the trisaccharide. Baking yeast strains with this ability have been constructed, resulting in an increase in biomass yield up to 8%. A genetically modified baker's yeast, developed at Gist-Brocades (Delft, The Netherlands) in 1990, was the first to be approved for use. Currently, strains constructed by intraspecific hybridization and gene transfer are not regarded as true genetically modified organisms (GMOs); nevertheless, commercial use of genetically manipulated yeast, or any other organisms, raises issues of safety, regulation, and labelling by authorities and receives serious concern from the public. To date, successful strategies of genetic engineering to introduce foreign genes into *S. cerevisiae* enabling the utilization of alternative carbon sources have remained at laboratory scale.

Recombinant DNA techniques have been applied to incorporate and express amylase genes from bacilli, molds, and different yeast species into baker's yeast, enabling starch hydrolysis. *S. cerevisiae* strains have also been created that are able to utilize lactose, pentoses from hemicellulose, and even cellulose. Whey, which is rich in lactose, would be a good alternative carbon source, but baker's yeast lacks the necessary enzymes to utilize lactose. Laboratory strains of *S. cerevisiae* have been constructed by both interspecies protoplast fusion with the yeast *Kluyveromyces lactis* and genetic transformation of genes from the bacterium *Escherichia coli* or the mold *Aspergillus niger.*

The keeping quality of baker's yeast is a complex property that is not well characterized in terms of physiology and biochemistry. Factors influencing cryoresistance, osmotolerance, and tolerance to drying have a great impact on the loss of leavening performance of ADY in frozen doughs and sweet doughs. In this respect, the cryoprotective effect of trehalose, the osmoprotective role of glycerol, and the stress tolerance attributed to membrane composition have received most attention. Molecular genetic studies have allowed a deeper insight into the regulation of trehalose, glycerol, and phospholipid metabolism, the knowledge on which strategies to improve keeping quality of baker's yeast can be based. Conditions of yeast propagation and storage have been optimized to produce yeast cells with higher contents of trehalose, phospholipids, unsaturated fatty acids, and ergosterol. Glycerol accumulation has been achieved in genetically altered strains.

Strain development by recombinant DNA technology may have a further impact on the application of baker's yeast. The supplementation of bread doughs with enzymes has increased greatly and has become a standard practice in breadmaking. The use of amylases, often together with protease, lipase, lipoxi genase or other enzymes, improves the functional properties of dough, decreases the mixing time, and contributes to flavor development. These commercial enzymes are produced with filamentous fungi or bacilli, and may contain impurities adversely affecting bread quality or acting as allergens. Possibilities exist to genetically engineer baker's yeast

secreting one or more enzyme additives during the process of breadmaking. Success achieved with laboratory strains may be impractical, however, at the industrial level. Moreover, the approval for commercial use of genetically manipulated yeast expressing heterologous proteins is still problematic and has met with public resistance. Nevertheless, taking measures to prevent any possibility of harmful consequence of genetic manipulation, recombinant strains will have an important role in future developments in the production and application of baker's yeast.

See also: **Antioxidants**: Synthetic Antioxidants; **Bread**: Breadmaking Processes; **Browning**: Nonenzymatic; **Drying**: Spray Drying; **Emulsifiers**: Uses in Processed Foods; **Filtration of Liquids**; **Flavor (Flavour) Compounds**: Structures and Characteristics; **Flour**: Dietary Importance; **Starter Cultures**; **Wheat**: The Crop; Grain Structure of Wheat and Wheat-based Products; **Wines**: Production of Table Wines; Production of Sparkling Wines; **Yeasts**

Further Reading

Beudeker RF, van Dam HW, van der Plaat JB and Vellenga K (1990) Developments in baker's yeast production. In: Verachtert H and De Mot R (eds) *Biotechnology and Biocatalysis*, pp. 103–146. New York: Marcel Dekker.

Caron C (1995) Commercial production of baker's yeast and wine yeast. In: Rehm H-J and Reed G (eds) *Biotechnology*, 2nd edn, vol. 9, *Enzymes, Biomass, Food and Feed*, pp. 321–351. Weinheim: VCH.

Evans IH (1990) Yeast strains for baking: recent developments. In: Spencer JFT and Spencer DM (eds) *Yeast Technology*, pp. 13–54. Berlin: Springer.

Linko Y-Y, Javanainen P and Linko S (1997) Biotechnology of bread baking. *Trends in Food Science and Technology* 8: 339–344.

Randez-Gil F, Sanz P and Prieto JA (1999) Engineering baker's yeast: room for improvement. *Trends in Biotechnology* 17: 237–244.

Reed G and Nagodawithana TW (1991) *Yeast Technology*, 2nd edn. New York: Van Nostrand Reinhold.

Spicher G and Brümmer J-M (1995) Baked goods. In: Rehm H-J and Reed G (eds) *Biotechnology*, 2nd edn, vol. 9. *Enzymes, Biomass, Food and Feed*, pp. 241–319. Weinheim: VCH.

YERSINIA ENTEROCOLITICA

Contents

Properties and Occurrence

G Zurera Cosano and R M García Gimeno, University of Cordoba, Cordoba, Spain

Background

The genus *Yersinia* comprises 11 species, of which three (*Yersinia enterocolitica*, *Yersinia pseudotuberculosis*, and *Yersinia pestis*) have clearly been shown to cause human disease. *Y. pseudotuberculosis* is associated with mesenteric adenitis, and *Y. pestis* is responsible for the bubonic plague. The remaining eight species (*Y. frederiksenii, Y. intermedia, Y. kristensenii, Y. mollareti, Y. bercovierii, Y. rohdei, Y. ruckeri*, and *Y. aldovae*) are sometimes referred to as *Yersinia enterocolitica*-like and have not yet been clearly shown to cause human disease and are easily distinguished from *Y. enterocolitica* by their biochemical characteristics. *Y. enterocolitica* is classified in the family Enterobacteriaceae. It is a small (1–3.5 μm × 0.5–1.3 μm), ovoid, sometimes rod-shaped, Gram-negative bacterium and facultative anaerobe. It grows within a wide range of temperatures, from −1 to 40 °C, achieving optimum growth at around 29 °C. It is nonmotile at 37 °C, but is motile and produces peritrichous flagella when it is grown at less than 30 °C. It produces colonies of 1.0 mm or less on nutrient agar. It is oxidase-negative and ferments glucose with little or no gas. It lacks phenylalanine deaminase and is urease-positive. *Y. enterocolitica* produce gastroenteritis (yersiniosis), and the symptoms include severe abdominal pain, suggesting an

appendicitis-like attack, diarrhea (lasting several weeks), fever, rash, nausea, headache, and vomiting. The minimum infectious dose is unknown. Illness onset is usually between 24 and 48 h after intake, although the incubation period can be as long as 11 days.

Y. enterocolitica is widely distributed in the environment, in lakes and stream waters, which are sources of the organisms to warm-blooded animals. Animals from which *Y. enterocolitica* has been isolated include cats, birds, dogs, guinea-pigs, rats, horses, chickens, cattle, swine, lambs, and fish. Certain biotypes and serogroups are associated with a specific host. A principal source or reservoir of virulent *Y. enterocolitica* are pigs. Virulent strains of *Y. enterocolitica* often reside in the oral cavity and the gastrointestinal tract of pigs. Serogroups 0:3 and 0:9 are mainly associated with pork tongues and tonsillar area. Despite this, it has been shown that pork meat has accidentally been a vehicle for yersiniosis.

There are four serogroups of *Y. enterocolitica* associated with pathogenicity: 0:3, 0:5, 0:8, and 0:9. The genes encoding for invasion of mammalian cells are located on the chromosome, while a 40–50-MDa plasmid encodes most of the other virulence-associated phenotypes. The 40–50-MDa plasmid is present in almost all pathogenic *Yersinia* species, and the plasmids appear to be homologous. Although heat-stable enterotoxin has been isolated from most clinical isolates, it is uncertain whether this toxin plays a role in the pathogenicity of the organism.

Y. enterocolitica have been detected in a wide variety of processed foods, particularly pasteurized milk, dairy products, vacuum-packaged meats, seafood, vegetables, and many ready-to-use foods. Incidence varies between countries. The most common *Y. enterocolitica* serogroups in human infections are 0:3, 0.5, 27, 0:8, and 0:9. In Europe, Africa, and Japan, serogroups 0:3 and 0:9 are the most common. These serogroups are prevalent in Denmark, Belgium, and Sweden, and are a significant cause of gastrointestinal illnesses in those countries.

Growth Requirements

Temperature

Y. enterocolitica is a psychrotroph and can grow at temperatures as low as 0 °C and as high as 44 °C, the optimum range being 22–29 °C. For biochemical reactions, the optimum temperature is about 28 °C in pure culture, but it is a poor competitor at this temperature. Multiplication can occur between −2 and 0 °C, but growth is very slow. At 3 °C, it takes 200 h to increase 2 logarithmic cycles in culture medium. Growth at 0–2 °C in milk after 20 days has been observed. Generation times at 4 °C appear to be strain-dependent, varying between 16 and 26 h in pasteurized milk, and growth is preceded by an extended lag time of up to 40 h. Increases in levels of up to 5 log increments in 10–14 days have been recorded in a variety of artificially contaminated cooked foods stored at 4 °C. Growth at 0–1 °C on pork and chicken has been observed and on raw beef kept for 10 days at 0–1 °C. Growth is slightly slower in the presence of competitor organisms but at 4 °C competes better against other psychrotrophic bacteria. In milk at 4 °C, *Y. enterocolitica* grow and reach numbers of up to 10^7 CFU ml^{-1} in 7 days, and compete well with the background flora.

pH

The pH range for growth is 4.6–9.0, with an optimum pH of 7.0–8.0. The minimum pH at which growth can occur varies between strains and also depends on storage temperature and the acidulant used. The following values have been found for six strains of *Y. enterocolitica* with the pH adjusted with HCl and incubated for 21 days: 4.42–4.80 at 4 °C; 4.36–4.83 at 7 °C; 4.26–4.50 at 10 °C and 4.18–4.36 at 20 °C. Growth inhibition at low pH is greater at 4 °C than at 20 °C, and the use of organic acids raises the minimum pH for growth. When organic acids are used to adjust pH, the order of their effectiveness is acetic > lactic > citric. Acetic acid is the most inhibitory organic acid commonly used in food, preventing growth below 5.7 at 4 °C and 5.1 at 20 °C. It has also been shown that tolerance to high pH values is dependent on temperature and on the medium used. Growth has been demonstrated at pH 9.5 at 10 °C but not at 6 °C. Under the same pH and temperature conditions, the inhibitory effects of acids are quite different. At a high pH (5.8), the order of inhibition of *Y. enterocolitica* 0:9 at 4 °C is formic acid > acetic acid > propionic acid > lactic acid, whereas at a lower pH (4.2) the order is formic acid > lactic acid > acetic acid > propionic acid. The mechanisms of the interaction of these acids inside bacterial cells differ under aerobic and anaerobic conditions. While the inhibitory effect of lactic acid is enhanced under anaerobic conditions, the acetic and formic acids are less effective.

Heat Tolerance

According to available data, the heat resistance of *Y. enterocolitica* seems to vary widely among different strains. Strains of *Y. enterocolitica* are not particularly heat-resistant. Sublethal injury to

Y. enterocolitica cells has been reported after heat treatment at 47 °C for more than 12 min. However, the magnitude of the effect depends on growth temperature. In the range of 4–20 °C, the growth temperature does not influence heat resistance at 54–66 °C for *Y. enterocolitica*. However, when cells are grown at 37 °C, heat resistance increases fourfold. The heat resistance of *Y. enterocolitica* is also influenced by the pH of the heating medium. The pH of maximum heat resistance in citrate phosphate buffer is pH 7 for cells grown at 37 °C but pH 5 for those grown at 4 °C. In both, the magnitude of the pH effect on heat resistance is constant at all heating temperatures. Different studies have been carried out in milk as pasteurization and thermalization are achieved effectively when the inoculum does not exceed 10^8 CFU ml^{-1}.

Freezing

Refrigeration reduces the multiplication capacity of *Y. enterocolitica* but does not inhibit it completely. Growth at low temperatures depends on other factors, mainly on pH. Good growth has been observed at 4 °C, when the pH is neutral, but very little growth has been detected at a slightly acidic pH (5.4). The ability of *Y. enterocolitica* to withstand freezing has been shown: cultures of *Y. enterocolitica* can survive for several weeks at −20 °C without any reduction in the number of viable cells. Also, in studies with frozen foods, 3–4 logarithmic reductions have been found to occur after 4 weeks of storage at −18 °C. Despite this, *Y. enterocolitica* has been isolated from a wide range of frozen foods.

Irradiation

The ability of *Y. enterocolitica* to survive γ-radiation depends on the food composition/treatment and the presence of preservatives. As *Y. enterocolitica* is more radiosensitive than other common foodborne pathogens such as *Salmonella* and *Listeria*, it can be completely eliminated at a dose range of 3–4 kGy. With low numbers of *Y. enterocolitica* (D_{10} values between 0.25 and 0.30 kGy), a contamination of up to 10^3–10^4 CFU g^{-1} can be positively controlled by a 1–1.5-kGy dose. Thus, the average dose of 1 kGy recommended by the FDA appears to be adequate for eliminating *Yersinia*. While irradiation even at 6 kGy fails to prevent survival in raw pork, the bacterium is completely suppressed in cooked ham and salami treated with lower doses of radiation. Faster recovery of radiation-injured cells in raw meat than in cured meat with spices or in cooked ham demonstrates that *Y. enterocolitica* in irradiated products can be inhibited by using combination processes including irradiation.

Preservatives and Packaging

The behavior of *Y. enterocolitica* in response to various preservatives depends on the temperature and level of the chemical, pH, and amount of sodium chloride present. At low pHs and temperatures, potassium sorbate is very effective against *Y. enterocolitica*. In contrast, other preservatives such as sodium phosphite and sodium hypophosphite are inhibitory at relatively high pH values. Packaging containing high levels of carbon dioxide is very effective at delaying growth and can cause complete inhibition. *Y. enterocolitica* can survive sodium chloride concentrations of 6–7% at specific temperature and pH levels. Although 7% NaCl is inhibitory at 3 °C, growth occurs with 5% NaCl. Other preservatives such as potassium sorbate, sodium hypophosphite, sodium phosphite, and curing salt delay or inhibit the growth of this microorganism.

Predictive Modeling

Predictive models are used to describe the behavior of microorganisms under different conditions such as temperature, pH, water activity, oxidation–reduction potential, and additives. They may be applied in various ways to ensure the microbial safety and quality. The main objective is to describe mathematically the growth or inactivation of microorganisms in food under prescribed food conditions.

Several models charting the growth and survival of *Yersinia enterocolitica* have been developed by studying the effect of different factors, such as: pH (4–8.5); organic acids (acetic, lactic, sulfuric, and citric acid); temperature (0–42 °C); NaCl concentration (0–10.5%); sodium nitrite (0–200 p.p.m.), strain with and without plasmids related to virulence agents, and mild heat treatment (50–60 °C). These models have been validated in different foods such as pâté, packed cold smoked salmon, UHT milk, yogurt, mayonnaise, meat and meat products, eggs, fish, and tofu; in general, predictive models overstimate growth rate values on the fail-safe side, i.e., the prediction always describes faster growth rates than the values observed in real foods.

Epidemiological and Clinical Features

From an epidemiological standpoint, *Y. enterocolitica* is widespread in nature, occurring in the gastrointestinal tract of an array of animal hosts, including mammals, avian species, and cold-blooded species, and is occasionally isolated in humans. Of all sources, swine appear to be the major source of strains

pathogenic for humans. Several outbreaks have been reported where infection has been associated with contact with pigs, and indirect contact with contaminated pork products and other foods (Table 1). Some outbreaks in families were attributed to exposure to cats and dogs. On rare occasions, it can be transmitted as a result of the bacterium passing from stools or soiled fingers of one person to the mouth of another person. This may happen when basic hygiene and handwashing habits are inadequate.

In general, sporadic or food-related outbreaks are more frequent in regions where pigs are the carriers of this microorganism. Seasonality seems to play a role in the incidence of this microorganism in focal geographical areas, with a higher frequency noted during cooler periods of the year.

Children are infected more often than adults, and confirmed cases or *Y. enterocolitica* in humans correspond to 1 in 100 000 persons per year. In the USA, it is estimated to cause 3000–20 000 cases each year. Most recently, it has been shown that humans can be asymptomatic, since clinically normal blood donors have supplied *Yersinia enterocolitica*-infected blood resulting in fatal septicemia in transfused patients. Nevertheless, 90% of cases originate from food.

Human clinical infections with *Y. enterocolitica* occur most frequently after the intake of food or water contaminated with this pathogen. The infectious dose estimated to be sufficient to cause human illness is approximately 10^7 cells, but consumer susceptibility will vary according to individual levels of immunocompetence, age, and general health. After intake, the bacteria pass through the stomach, colonize the epithelial cell surface of the small intestinal mucosa, and cause diarrhea, possibly by producing the enterotoxin YST. This microorganism can also invade epithelial cells, attach to M cells, and penetrate Peyer's patches, where it multiplies with the help of several surface components that enable it to resist phagocytosis and survive complement-mediated killing. The invasion process and virulence factor production are complex processes that are thermoregulated by both plasmid and chromosomally encoded information. Continued proliferation of bacteria results in inflammation leading to the formation of local microabscesses and ulceration of the overlying epithelium. In a relatively short period, ingested *Y. enterocolitica* reach the terminal ileum, spread to the mesenteric lymph nodes, and enter the blood stream.

From a clinical standpoint, *Y. enterocolitica* is a common human pathogen that causes a variety

Table 1 Outbreaks of foodborne *Y. enterocolitica* infection

Vehicle	*Number of cases (confirmed)*	*Biogroup(s)*	*Country*	*Year*
Pasteurized milk	1000 (172)[a]	13a, 13b	USA	1982
	3(3)[b]	6,30	UK	1984
	19(19)[b]	10K	UK	1985
	17(17)[b]	6,30	UK	1985
Chocolate milk	218 (38)[c]	8	USA	1976
	250[d]		USA	1978
Cream deserts	4[e]		Belgium	1975
Powered milk/turkey chow mein	239(159)[b]	8	USA	1981
Bean sprouts	16[f]	8	USA	1982
Tofu	87(56)[g]	8, tacoma(21)	USA	1981
Chitterlings	14[h]		USA	1988
Pork	1[i]		USA	1994
Pork (pork cheese)	5(3)[b]	3	Hungary	1983

[a]From Tacket C, Narain J, Sattin R *et al*. (1984) A multistate outbreak of infections caused by *Yersinia enterocolitica* transmitted by pasteurized milk. *Journal of the American Medical Association* 251: 383–486.

[b]From Greenwood MH (1993) *Yersinia enterocolitica*. Properties and occurrence In: Macrae R, Robinson RK and Sadler MJ (eds) *Encyclopaedia of Food Science, Food Technology and Nutrition*. London: Academic Press.

[c]From Black R, Jackson R, Tsai T *et al*. (1978) Epidemic *Yersinia enterocolitica* infection due to contaminated chocolate milk. *New England Journal of Medicine* 298: 76–79.

[d]From Silliker J (1986) *Yersinia enterocolitica. Food Technology* 8: 22.

[e]From Richard J, Moiner J, Cornu C, Goor M and Dereume G (1975) Léndemie *à Yersinia enterocolitica* dans le Tournaisis. *Lille Medical* 20(1): 36–43.

[f]From Cover T and Aber R (1989) *Yersinia enterocolitica. New England Journal of Medicine* 231: 16–24.

[g]From Aulisio C, Stanfield J, Weagent S and Hill W (1983) Yersiniosis associated with tofu consumption: serological, biochemical and pathogenicity studies of *Yersinia enterocolitica* isolates. *Journal of Food Protection* 46: 226–230.

[h]From Lee L, Gerber A and Lonsway D *et al*. (1990) *Yersinia enterocolitica* 0:3 infections in infants and children, associated with the household preparation of chitterlings. *New England Journal of Medicine* 322: 984–987.

[i]From Stoddard J, Wechsler D, Nataro J and Casella J (1994) *Yersinia enterocolitica* infection in a patient with sickle cell disease after the exposure to chitterlings. *American Journal of Pediatric Hematology and Oncology* 16: 153–155.

of intestinal and extraintestinal syndromes, with varying degrees of severity, ranging from mild gastroenteritis to mesenteric lymphadenitis (often confused with appendicitis), and septicemia. Infections caused by this microorganism may also reveal postinfectious immunological sequelae, including erythema nodosum, arthritis, miocarditis, and glomerulonephritis.

Occurrence in Foods

Foods analyzed for *Y. enterocolitica* show that the distribution of this microorganism is fairly extensive; it has been isolated in beef, lamb, pork, oysters, shrimp, crabs, and water. The vast majority of these isolates are avirulent, and virulent strains seldom occur in any of these foods. Studies have examined the fecal droppings of rats near slaughterhouses and have found that the rats carried virulent *Y. enterocolitica* probably picked up from the pork that was being processed at the slaughterhouses.

The 68% positive samples of porcine tongues taken from freshly slaughtered animals have been isolated, and these represent six serogroups, 0:8 being the most common and 0:6, 30 the second most common, followed by other serogroups including 0:3, 0:13, 7, 0:18, and 0:46. Several studies performed in Brazil have reported positive isolations of *Yersinia enterocolitica* from raw milk (16.8%), pasteurized milk (1.7%), raw beef (80%) and chicken (60%), ground beef, liver, and pork (20%). In one study performed in the USA, after examining 100 milk samples, 12 raw samples and one pasteurized sample yielded *Yersinia enterocolitica*. Another study of this product in France reported that 81% of raw milk samples were positive, the predominant serovar being 0:5 (Table 2).

Outbreaks of yersiniosis have been reported in which chocolate, coleslaw, tofu, bean sprout, pasteurized milk, and pork chitterlings have been singled out as vehicles (Table 1). One of the first and most important outbreaks occurred in the USA, where 220 schoolchildren were taken ill; it was confirmed that *Y. enterocolitica* had caused the problem and that chocolate milk was the vehicle of infection. This milk was prepared at a small dairy by means of vat pasteurization; the chocolate was added after pasteurization and mixed with a paddle; hence, there was plenty of potential postprocessing contamination. In another case, the vehicle of transmission was tofu (soyabean curd) that had been packaged in untreated spring water contaminated with *Y. enterocolitica*.

Both virulent and avirulent strains of *Y. enterocolitica* have been isolated from a variety of dairy products. Three milkborne outbreaks occurred in the USA, two attributable to serogroup 0:8 and one to serogroup 0:13a, 13b. In 1982, there was an outbreak linked to pasteurized milk; 170 cases were confirmed, although epidemiological studies revealed that more than 1000 people were infected. The strain was also isolated from the plastic carrying cases used to carry the milk after bottling. In this case, outdated milk had been brought to a pig farm in carrying cases. The cases were placed in mud and manure, the milk poured into troughs for the pigs, and the carrying cases brought back to the plant and superficially washed. Although it was confirmed that the pigs were the source of infection, it was impossible to determine exactly how the microorganism had reached the milk from the carrying cases.

Prevention of Food Contamination

Food-related toxic infections caused by *Yersinia enterocolitica* seem to be related to foods in cold storage. In this situation, it is difficult to define preventive measures (repeated washing of vegetables, mainly in collective restaurants, strengthening of the cold chain, etc.) that enable the spread of this food-related infection to be restricted. These limitations to prevention highlight the importance of monitoring human infections (networks of expert doctors, control laboratories, and reference laboratories for studying strains). The prevention of foodborne disease depends on careful food production, handling of raw products, and preparation of finished foods. Hazards can be introduced at any point from farm to table. Monitoring and control technologies systematically applied to food production by the HACCP are highly effective for preventing foodborne illness.

Primary control of the microorganism would require changes to current slaughtering practice. The 'bagging technique' – the method of enclosing the anus and rectum of pigs in a plastic bag during slaughter – has been introduced in several slaughterhouses, mainly in Norway, and it has been shown that this effectively reduced carcass contamination by *Yersinia enterocolitica*. The removal of the head, tongue, and pharynx at an early stage of slaughter and improvements in the way in which intestines are removed would also help to prevent carcass contamination. Additional practices, such as the exclusion of tonsillar tissue from sale or the scalding of tongue and head meat prior to the dispatch of products may also be necessary.

Table 2 Occurrence of *Yersinia enterocolitica* on food

Food	*Percentage of samples*	*Serogroup(s)*	*Country*
Raw milk	81[a]	0:5	France
	12[b]		USA
	18.2[c]		Canada
	16.8[d]		Brazil
	30[e]		Morocco
Raw milk, goat	71[f]		Australia
Pasteurized milk	1[b]		USA
	13.7[d]		Brazil
Traditional fermeted milk	6.3[e]		Morocco
Cheese	9.2[c]		Canada
	4[e]		Morocco
Turkish Feta cheese	28.8[g]		Turkey
Salmon and trout	23[h]		UK
Raw beef and chicken	80[i]		Brazil
Ground meat and liver	60[i]		Brazil
Ground meat	48[j]		Germany
Porcine tongues	68[k]	0:8; 1:6, 30; 0:3; 0:13, 7; 0:8; 0:46	USA
	50.8[l]		USA
Chitterling	> 50[m]	0:3	USA
Pork meat	26.7[e]		USA
	20[i]		Brazil
	10.83[n]	0:3, 3B phage type 2; 0:3, 4 phage type 8; biotype 1; 2	Japan
	17[o]	0:3	Norway
Drinking water	5[p]		Germany
	54[q]		USA
Vegetables	1[r]		Italy
	7[s]		France
	50[t]		France

[a]From Vidon D and Delmas C (1981) Incidence of *Yersinia enterocolitica* in raw milk in Eastern France. *Applied and Environmental Microbiology* 41: 353–359.
[b]From Moustafa M, Ahmed A and Marth E (1983) Occurrence of *Yersinia enterocolitica* in raw and pasteurized milk. *Journal of Food Protection* 46: 276–278.
[c]From Schiemann D (1978) Association of *Y. enterocolitica* with the manufacture of cheese and occurrence in pasteurized milk. *Applied and Environmental Microbiology* 36: 274–277.
[d]From Tibana A, Warnken M, Nunes M *et al*. (1987) Occurrence of *Yersinia enterocolitica* in raw and pasteurized milk in Rio de Janeiro, Brazil. *Journal of Food Protection* 50: 580–583.
[e]From Hamama A, El Marrakchi A and El Othmani F (1992) Occurence of *Yersinia enterocolitica* in milk and dairy products in Morocco. *International Journal of Food Microbiology* 16: 69–77.
[f]From Hughes J and Jensen N (1981) *Yersinia enterocolitica* in raw goat's milk. *Applied and Environmental Microbiology* 41: 309–310.
[g]From Simsek O and Arici M (1994) Pathogenic microorganisms in cheese and their foodborne disease. In: *Cheese With All Aspects*, 2nd edn, Tekirdağ: Trakya University.
[h]From Davies A, Capell C, Jehanno D, Nychas G and Kirby R (2001) Incidence of foodborne pathogens on European fish. *Food Control* 12: 67–71.
[i]From Warnken A, Nunes M and Noleto A (1987) Incidence of *Yersinia enterocolitica* in meat samples purchased in Rio de Janeiro, Brazil. *Journal of Food Protection* 50: 578–579, 583.
[j]From Kleinlein N, Unterman F and Beissner H (1989) Occurrence of *Salmonella* and *Yersinia* species as well as *Listeria monocytogenes* in minced meat. *Fleischwirtschaft* 69: 1474–1476.
[k]From Doyle MP, Hugdahl MB and Taylor SL (1981) Isolation of virulent *Yersinia enterocolitica* from porcine tongues. *Applied and Environmental Microbiology* 42: 661–666.
[l]Harmon MC (1984) A study of the incidence and pathogenic potential of *Yersinia enterocolitica* in Indiana pork. *Dissertation Abstracts International, B* 44: 3707.
[m]From Lee LA, Gerber DR, Lonsway JD *et al*. *Yersinia enterocolitica* O:3 infections in infants and children, associated with the household preparation of chitterlings. *New England Journal of Medicine* 322: 984–987.
[n]From Fukushima H, Hoshima K, Nakamura R and Ito Y (1987) Occurrence of *Yersinia* spp. in raw beef, pork and chicken. *Zentralblatt fur Bakteriologie Mikrobiologie und Hygiene* (B) 184: 50–59.
[o]From Johannessen G, Kapperud G and Kruse H (2000) Occurrence of *Yersinia enterocolitica* in Norwegian pork products determined by PCR method and a traditional culturing method. *International Journal of Food Microbiology* 54: 75–80.
[p]From Schindler P (1984) Isolation of *Yersinia enterocolitica* from drinking water in South Bavaria. *Zentralblatt fur Bakteriologie Mikrobiologie und Hygiene* 180: 76–84.
[q]From Langeland G (1983) *Yersinia enterocolitica* and *Yersinia enterocolitica*-like bacteria in drinking water and sewage sludge. *Acta Pathologica, Microbiologica et Immunologica Scandinavica* 91: 179–185.
[r]From Gola S, Previdi M, Mutti P and Belloli S (1990) Microbiological investigation of frozen vegetables, incidence of *Listeria* and other psychrotrophic pathogens. *Industria Conserve* 65: 36–38.
[s]From Catteau M, Krembel C and Wauters G (1985) *Yersinia enterocolitica* in raw vegetables. *Science Aliments* 5: 103–106.
[t]From Darbas H, Riviere M and Obertic J (1985) *Yersinia* in refrigerated vegetables. *Science Aliments* 5: 81–84.

Although normal detergents and disinfectants used in the food industry are effective for eliminating *Yersinia enterocolitica*, their effect is reduced if fatty residues are present. The regular use of steam cleaning instead of cleaning with hot water at 85 °C has been used successfully at a milk pasteurization plant.

Handlers of meal are the last critical control point before foods reach the table. Therefore, several practical food-handling precautions could reduce the risk of contamination of food by this microorganism. Cross-contamination of foods can be avoided by separating cooked and raw food and also by washing food handlers' hands, cutting boards, and contaminated surfaces. Practical recommendations can be followed by consumers, such as avoiding the consumption of raw or undercooked pork, raw milk, or related byproducts, and washing hands carefully before eating, preparing food after contact with animals, and after handling raw meat. After handling raw chitterlings, it is recommended that hands and fingernails be cleaned thoroughly with soap and water before touching children or their toys, bottles, or pacifiers.

See also: **Epidemiology**; **Freezing**: Principles; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; Chemical and Microbiological Changes; Electrical Process Heating; **Irradiation of Foods**: Basic Principles; **pH – Principles and Measurement**; **Preservatives**: Classifications and Properties; ***Yersinia enterocolitica***: Detection and Treatment

Further Reading

Adams MR and Moss MO (1995) *Food Microbiology*. Cambridge: The Royal Society of Chemistry.

Bottone EJ (1999) *Yersinia enterocolitica*: overview and epidemiological correlates. *Microbes and Infection* 1: 323–333.

Doyle MP and Cliver DO (1990) *Yersinia enterocolitica*. In: Cliver DO (ed.) *Foodborne Diseases*, pp. 223–228.San Diego, CA: Academic Press.

Eley AR (1992) *Microbial Food Poisoning*. London: Chapman & Hall.

Forsythe SJ (2000) *The Microbiology of Safe Food*. Oxford: Blackwell Science.

Greewood MH (1993) *Yersinia enterocolitica*. Properties and occurrence. In: Macrae R, Robinson RK and Sadler MJ (eds) *Encyclopaedia of Food Science, Food Technology and Nutrition*. London: Academic Press.

Jay JM (2000) *Modern Food Microbiology*, 6th edn. Gaithersburg, MD: Aspen.

Schiemann DA (1989) *Yersinia enterocolitica* and *Yersinia pseudotuberculosis* In: Doyle MP (ed.) *Foodborne Bacterial Pathogens*, pp. 601–672. New York: Marcel Dekker.

Detection and Treatment

A Ramesh, B P Padmapriya, S Bharathi and M C Varadaraj, Central Food Technological Research Institute, Mysore, India

Background

Yersinia enterocolitica is an invasive enteropathogen prevalent in the environment and often isolated from soil, water, food, and clinical sources. The serotypes of *Y. enterocolitica* associated with disease in humans are O:3, O:9, O:8, and O:5, 27, and all harbor a 70–75-kb virulence plasmid. The bacterium has been recognized as a significant foodborne pathogen owing to its ability to survive in both vacuum-packed and refrigerated food samples. Thus, in the light of food hygiene and public health, it is imperative to develop rapid methods of detecting *Y. enterocolitica*. Various methods consisting of conventional cultural methods, detection of virulence plasmid, immunological methods, and rapid nucleic acid-based techniques have been described for the detection and identification of *Y. enterocolitica*. In an infection, the principal clinical feature in humans is abdominal pain and diarrhea accompanied by fever. Reactive arthritis and erythema nodosum have also been reported as common complications. Several cases of diarrhea resulting from *Y. enterocolitica* infection resolve on their own, thus avoiding the need for an antibiotic treatment. However, in the case of a severe infection, antibiotics are administered. The following section describes the methods of detection and treatment of *Y. enterocolitica*.

Detection of *Yersinia enterocolitica*

The common food samples associated with *Y. enterocolitica* infection are pork, milk and milk products, and raw vegetables. The detection of pathogenic strains of *Y. enterocolitica* from food samples and clinical specimens is mainly based on the isolation of the bacterium. A scheme of the different detection methods followed for *Y. enterocolitica* is shown in **Figure 1**.

Conventional Methods of Detection

Yersinia enterocolitica is a Gram-negative rod or ovoid, facultatively anaerobic, oxidase-negative, catalase-positive nonsporulating motile bacterium. The isolation of this bacterium from food samples is difficult, because it is present in low numbers and is slow growing compared to other coexisting members

of the enterobacteriaceae present in the background microflora. This underscores the need for a selective enrichment technique to facilitate the isolation of the bacterium. A basic isolation scheme involves preenrichment in a highly nutritious, well-buffered medium containing peptone, yeast extract, and mineral salts at temperatures ranging from 4 to 37 °C over 1–9 days. The salient features of the various types of enrichment procedures are indicated in Table 1.

Selective enrichment media like bile oxalate sorbose (BOS) broth, and phosphate buffer saline (PBS) supplemented with sorbitol and bile salts with a postenrichment alkali treatment has been widely used to enrich *Y. enterocolitica* from food samples. This method is simple, sensitive, and rapid, eliminating a large number of contaminating non-*Yersinia* species. *Y. enterocolitica* is also a psychrotroph, capable of growing under refrigerated conditions. Hence, cold enrichment can increase the isolation rate of the bacterium. Irgasan–ticarcillin–chlorate (ITC) enrichment has been shown to be more efficient in isolating pathogenic bioserotype from pork. The enriched culture is plated on to the selective media cefsulodin–irgasan–novabiocin

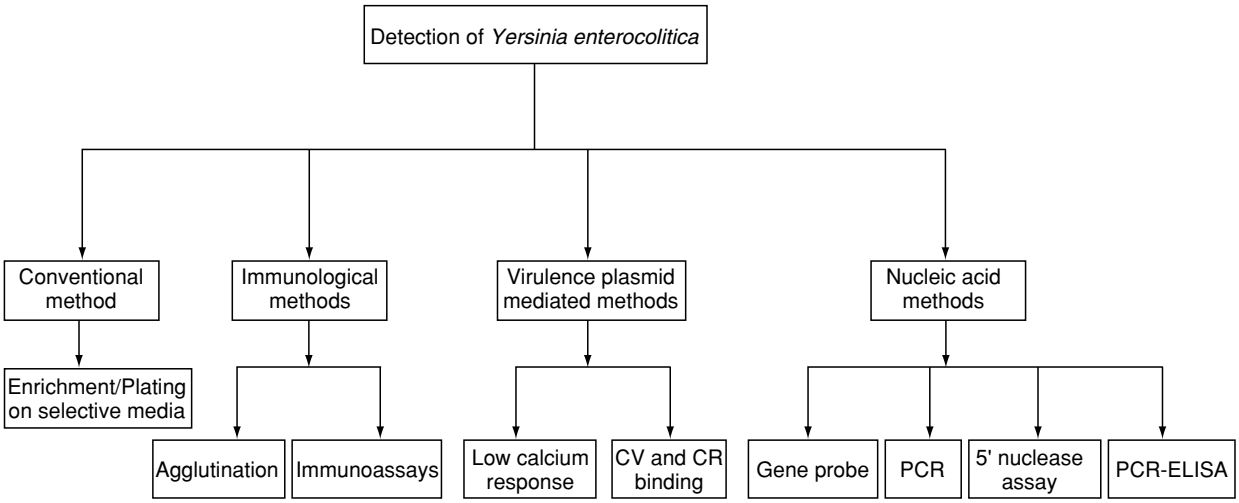

Figure 1 Schematic representation of the various detection methods for *Yersinia enterocolitica*.

Table 1 Enrichment broth and growth conditions used for the isolation of *Yersinia enterocolitica*

Primary enrichment	*Incubation temperature (°C)*	*Incubation period (days)*	*Secondary enrichment*	*Incubation period (days)*	*Incubation temperature (°C)*
MRB	23	5			
PB	4	14			
PB	4	14	MRB	23	5
CMB	23	28	MRB	23	5
SC	37	1			
RMC	22	3			
BHI	8–10	2	SB	22–28	2
ITC	RT	2			
PBS	4	2			
PBSSB	4	21			
BHI	8–10	2	KOH	RT	
BHI	8–10	2	SC	RT	2
HOBS	4	5			
PBS	4	21	MRB	27	2
BOS	22	3–5			

BHI, brain–heart infusion broth; BOS, bile–oxalate–sorbitol broth; CMB, cooked meat broth; HOBS, hemoglobin–oxalate–bile–sorbitol broth;ITC, irgasan–ticarcillin–chlorate; MRB, modified Rappaport broth; PB, Butterfields phosphate buffer; PBS, phosphate-buffered saline; PBSSB, phosphate-buffered saline–sorbitol–bile salts; RMC, restrictive $MgCl_2$ broth; SB, sorbitol bile broth; SC, selenite cystine.

(CIN) agar and incubated for 24–48 h at 32–37 °C to obtain characteristic colonies with a deep red center (like a bull's eye) with an entire edge and a transparent border measuring 0.5–1.0 mm in diameter. Suspected *Yersinia* colonies from CIN plates are selected and suspended individually in trypticase soya broth (TSB) for further characterization through diagnostic biochemical tests, as indicated in Table 2. *Yersinia enterocolitica* has been essentially classified into six biogroups, i.e., 1A, 1B, 2, 3, 4, and 5. Biotype 6 is rare.

Current methods based on selective enrichment merely detect the presence of *Y. enterocolitica*, and these methods fail to recover all pathogenic bioserotypes. Further improvements in the specific detection of the bacteria can be achieved through immunological and nucleic acid-based methods.

Immunological Detection of *Y. enterocolitica*

Immunological methods have been reported since the early detection of *Y. enterocolitica* as an advancement over conventional microbiological tests. The methods that have been widely accepted include agglutination tests and immunoassay.

Agglutination tests Latex slide agglutination and tube agglutination (Widal) tests are commonly used and form the basis for many commercially available kits for the rapid detection of *Y. enterocolitica* O:3 and O:9 antibodies in the serum. Latex of 0.9 μm particle size was coated with specific antibodies raised against serotypes O:3, O:5, O:6, O:8, and O:9 of *Y. enterocolitica*. The latex reagents specifically detected the O-antigen of *Y. enterocolitica*, with no cross-reactivity between the serotypes. Alcohol-treated bacterial cells are used as antigens in some cases, and an agglutination titer of 640 or a fourfold or greater increase in the titer is an indication of recent or actual infection. Commercial kits are available to detect O:3 and O:9 antigens of *Y. enterocolitica* individually or in combination, using the latex agglutination method.

Immunoassays The lack of sensitivity and specificity of the agglutination tests hinders their application as a routine detection method. However, immunoassays like surface adhesion immunofluorescence (SAIF) technique, enzyme immunoassay (EIA), immunoblot, and enzyme-linked immunosorbent assay (ELISA) can alleviate this problem. SAIF has been used to detect *Y. enterocolitica* in broth and enriched meat cultures, with a polycarbonate membrane to trap the cells on the surface of the membrane prior to detection by immunofluorescence.

EIA was developed for the detection of *Yersinia* immunoglobulin (Ig) complexes of known Ig class. Immune complexes are attached to polystyrene microtiter plates by rabbit antihuman immunoglobulins. Detection of *Y. enterocolitica* O:3 antigens was demonstrated using Fab fragments of alkaline phosphatase-conjugate antibody against the same serotype. Methods to differentiate serological responses to *Y. enterocolitica* O:9 from responses to *Brucella* have been very successful. However, immunoblots

Table 2 Differentation of *Yersinia enterocolitica* biogroups

Test	*Biogroup*						
	1A	*1B*	*2*	*3*	*4*	*5*	*6*[a]
Indole production	+	+	−	−	−	−	−
Esculin hydrolysis	+/−	−	−	−	−	−	−
Voges–Prausker test	+	+	+	+	+	+	−
L-Ornithine decarboxylase	+	+	+	+	+	+/−	+
Pyrazinamidase	+	−	−	−	−	−	+
Lipase	+	+	−	−	−	−	−
D-Glucosidase	+	−	−	−	−	−	−
Acid from							
Inositol	+	+	+	+	+	+	+/−
Salicin	+	−	−	−	−	−	−
L-Sorbose	+	+	+	+	+	−	+
Trehalose	+	+	+	+	+	−	+
D-Xylose	+	+	+	+	−	V	+
Sucrose	+	+	+	+	+	+	−

[a]This biotype is relatively rare.
V, variable reacton.
Modified from Bottone EJ (1999) *Yersinia enterocolitica*: overview and epidemiological correlates. *Microbes and Infection* 1: 323–333; Wauters G (1981) Antigens of *Yersinia enterocolitica*. In: Bottone EJ (ed.), pp. 41–53. Boca Raton, FL: CRC Press.

and ELISA targeting the *Yersinia* outer membrane proteins (YOPs) are more sensitive and specific than other serological tests for yersiniosis.

Monoclonal antibodies specific for the lipopolysaccharide (LPS) O antigens of *Y. enterocolitica* serotype O:3 can be used for the detection of *Y. enterocolitica* in colony immunoblots. This method has been shown to be sensitive in detecting a single colony of *Y. enterocolitica* in the presence of 3.1×10^8 heterologous organisms. Besides, the assay was specific with no significant cross-reactivity against any related serotypes. An immunoblot kit called *recom* Blot *Yersinia* IgG and *recom* Blot IgM/IgA has been developed and provides recombinant *Yersinia* antigens in a test strip format to detect IgG, IgM, or IgA antibodies directed against *Y. enterocolitica* and *Y. pseudotuberculosis*.

ELISAs using either purified lipopolysaccharide or formalinized whole cells expressing virulence plasmid-coded surface antigen (pYV^+ cells) can be performed to detect *Y. enterocolitica*. It has been shown that ELISA can specifically detect pYV antibodies in sera, cross-adsorbed to formalinized pYV-cured *Y. enterocolitica* O:3 cells prior to the assay. The serological response against YOPs using AB-ELISA was used to detect the O:9 serotype of *Y. enterocolitica*. In order to avoid interference by *Y. enterocolitica* O:9 LPS with YOPs in serological analysis, an antigenic preparation of YOPs was obtained from *Y. enterocolitica* O:7. There is a high degree of homology between YOPs of different *Yersinia* species and an absence of cross-reactivity between *Y. enterocolitica* O:7 LPS and *Y. enterocolitica* O:9 LPS. The commercially available *Y. enterocolitica* ELISA kit provides a high diagnostic sensitivity capable of differentiating IgG from IgA and IgM.

Although a host of immunological methods are available to detect *Y. enterocolitica* for clinical applications, the use of these assays in a food system is comparatively limited. An immunomagnetic separation technique has been shown as an efficient device to sequester *Y. enterocolitica* in food samples and facilitate its sensitive detection. The major factors limiting the use of immunological techniques for the detection of *Y. enterocolitica* in food systems are the cross-reactivity with the background microflora present in food samples and the inhibitors present in the food matrix itself.

Detection of Virulent Plasmid-bearing (pYV^+) Strains

The pathogenic strains of *Y. enterocolitica* are characterized by the presence of a 70–75-kb plasmid which encodes for a number of virulence factors manifested as conspicuous colony morphology, autoagglutination, dye binding capacity, and low calcium response. Plasmid-mediated phenotypic expression has been used to detect virulent strains and to discern them from avirulent plasmid-less strains. In a calcium-deficient medium like the brain–heart infusion agarose (BHO) calcium response of virulent strain was tested. At 37 °C, plasmid-bearing virulent cells (pYV^+) appear as pin-point colonies 0.36 mm in diameter, whereas avirulent cells (pYV^-) form distinctly larger colonies. The CIN agar medium incorporated with crystal violet can distinguish between pathogenic and nonpathogenic strains. At 32 °C, the colony morphology appears similar for both the strains. However it has been demonstrated that at 37 °C, plasmid-bearing pathogenic strains appear as densely stained microcolonies compared to the nonpathogenic strains. Direct detection of a plasmid-bearing virulent strain of *Y. enterocolitica* has been carried out in artificially contaminated food samples like pork chops, ground pork, cheese, and zucchini. Enriched samples plated on BHO incorporated with Congo red and incubated at 37 °C form red pin-point colonies as a manifestation of plasmid-mediated dye binding and a low calcium response. Thus, the Congo red binding technique is suitable for direct detection and isolation of pathogenic *Y. enterocolitica* using a single medium.

Detection by Nucleic Acid Probes

The physiological traits encoded by the virulent plasmid are only expressed at 37 °C, which also favors plasmid loss accompanied by elimination of virulent traits. Thus, the instability of the plasmid limits its use as a tool for detecting virulent *Y. enterocolitica* in clinical and food samples. The emergence of nucleic acid probes has provided an efficient tool to detect genes in *Y. enterocolitica* that encode specific pathogenic traits. Such gene probes can be used in colony- or dot-blot hybridization to directly detect *Y. enterocolitica* in the presence of high background microflora. Additionally, these probes can be evaluated as a substitute to tissue culture-based assay to ascertain the invasive phenotype of pathogenic strains.

The target regions in a probe-based diagnostic assay can be unique sequences within chromosomal DNA, ribosomal RNA, or plasmid DNA. The ribosomal RNA is widely preferred as a target gene owing to its ubiquity and relatively high copy number, which enhances the assay sensitivity. The GENE-TRAK colorimetric hybridization assay for *Y. enterocolitica* has been developed, employing probes directed against specific signature sequences of 16 s rRNA. This assay has been shown to be a convenient alternative to conventional microbiological procedures, yielding presumptive results in a significantly shorter time.

However, this method lacks the discriminatory power to differentiate virulent from avirulent strains.

Probes targeting virulence plasmid genes can be employed to specifically detect pathogenic *Y. enterocolitica* serogroups and distinguish them from environmental *Yersinia* species, which are frequently encountered in food samples. A synthetic radioactively labeled 19-mer oligonucleotide probe based on the plasmid-encoded *Yersinia* adhesin A (*Yad* A) gene has been used in colony hybridization to detect pathogenic *Y. enterocolitica* in pork samples. The probe demonstrates a superior estimation of virulent strains in the samples than conventional methods. Polymerase chain reaction (PCR)-generated nonradioactive probes are safe, inexpensive, and shelf-stable, and can be used in the detection of the organism. In colony hybridization, the DIG (digoxigenin)-labeled PCR-generated *Vir* F and *Yad* A amplicons are able to detect pathogenic *Y. enterocolitica* in artifically contaminated food samples and indicate equivalence with the PCR assay and phenotypic test for pathogenicity.

Spontaneous loss of plasmids in *Y. enterocolitica* during *in-vitro* replication is frequently encountered either in the environment or in the laboratory. This fosters the need to develop probes targeting the chromosomally encoded virulence factors like the attachment invasion locus (*Ail*) and the heat-stable enterotoxin (*Yst*). The gene coding for the attachment invasion locus (*Ail*) has been found uniquely in pathogenic strains of *Y. enterocolitica*, whereas the invasion (*Inv*) gene in all *Yersinia* spp. DIG-labeled *Inv* and *Ail* probes has been used to differentiate pathogenic from nonpathogenic strains. A PCR-amplified DIG-labeled *Yst* probe has been used in colony hybridization to detect and differentiate between pathogenic and nonpathogenic strains in naturally and artificially contaminated pork and milk samples following enrichment in ITC enrichment broth. The hybridization results are concordant with those from biochemical and serological tests, and a detection limit of 10 colony-forming units (CFU) ml^{-1} in the original sample has been achieved. The use of a multiplex riboprobe can be adopted as a comprehensive strategy to detect multiple targets, thus enhancing assay sensitivity and ensuring detection of virulent *Y. enterocolitica* irrespective of plasmid loss.

Detection by Polymerase Chain Reaction

The use of gene probe-based hybridization assays to detect *Y. enterocolitica* is limited by the lack of sensitivity. PCR allows rapid and selective identification of *Y. enterocolitica* and fulfils the criteria for specificity and sensitivity. The *Yad* A gene located on the virulence plasmid (pYV) encoding for an outer membrane polymeric protein has been used as a target to detect the presence of pathogenic *Y. enterocolitica* in retailed pig tongue and minced meat. However, loss of pYV is frequently encountered under certain conditions of handling, especially in food samples. This leads to the possibility of false-negative results. The chromosomally encoded attachment invasion locus (*Ail*) is characteristically associated with virulence and can be an appropriate target to detect pathogenic *Y. enterocolitica*. *Ail*-specific primers have been used in PCR to identify virulent strains in blood, and the levels of sensitivity achieved indicate the potential of PCR in the diagnosis of transmission-related bacteremia caused by *Y. enterocolitica*. A two-step PCR targeting the *Ail* gene has been employed to detect pathogenic strains in environmental water, and the results have demonstrated a superior specificity and sensitivity over cultural methods.

A PCR method based on the targeting of the *rfb* C gene involved in the biosynthesis of the O side chain of *Y. enterocolitica* O:3 has enabled the detection and differentiation between pathogenic strain of *Y. enterocolitica* of serotype O:3 and *Y. pseudotuberculosis* and other pathogenic *Y. enterocolitica*. PCR enables reliable and sensitive detection when applied to pure cultures, but its efficiency is markedly curtailed when applied to food and clinical samples. This is mainly due to the potential inhibitors or high levels of background microflora present in such samples. Therefore, adequate sample preparation is required to either eliminate such inhibitors or concentrate target cells to yield satisfactory detection levels. Selective enrichment followed by buoyant density gradient centrifugation provides an efficient sample preparation method prior to PCR. This treatment method has facilitated the detection of 10 CFU ml^{-1} of the bacterium in pork using a multiplex PCR simultaneously targeting the plasmid-borne *Yad* A gene and the *Yersinia*-specific 16S rRNA region. Immunomagnetic separation (IMS) of the target cells from meat samples has enabled a nested PCR assay to detect 10–30 CFU g^{-1} meat in the presence of 10^6-fold excess background microflora. Thus, IMS provides an efficient way of excluding the PCR inhibitors from the sample and has the potential to be exploited for automation.

Methods for detecting virulent *Y. enterocolitica* based on PCR involve gel electrophoresis for product analysis and the use of ethidium bromide to stain the gels, which is hazardous and impractical for routine analysis in food monitoring laboratories. The fluorogenic 5′ nuclease assay is an attractive option for rapid detection of virulent *Y. enterocolitica*. The assay exploits the 5′–3′ nuclease activity of *Taq* DNA polymerase and releases a probe with a fluorescent reporter dye and a quencher, which hybridizes to

an internal fragment of the target gene. With increasing cycles of PCR amplification, the fluorescent signal increases proportionately. A 5′ nuclease assay targeting the *Yst* gene has been shown to be highly specific in detecting virulent *Y. enterocolitica*. This assay has allowed a sensitive detection of $\geq 10^2$ CFU ml^{-1} in pure culture and $\geq 10\,CFU\,g^{-1}$ in spiked ground pork. The 5′ nuclease assay developed for detecting virulent strains has also been shown to yield a better estimation of contamination of pork and other processed meat and a better detection level than the 'gold standard' recommended by the US FDA in the Bacteriological Analytical Manual for the detection of pathogenic *Y. enterocolitica*. The assay is rapid and quantitative, can be performed for a number of samples at a time, and is suitable for application as an automated high-throughput detection system.

Conventional PCR-based assays fail to discriminate between viable and nonviable cells. However, a reverse-transcriptase 5′ nuclease assay targeting a specific gene transcript can be potentially exploited for the specific detection of viable *Y. enterocolitica*. PCR–ELISA is another alternative to circumvent electrophoretic analysis of PCR products and reduce the post-PCR analysis time. Essentially, a biotinylated capture probe is employed to immobilize the labeled PCR products on to microtiter plates. This colorimetric assay performed in the microtiter format is amicable to automation, can be applied to a large number of samples simultaneously, and allows the PCR product to be quantified directly. A PCR–ELISA adopted for detecting virulent strains has demonstrated a high specificity and sensitivity comparable to gel electrophoresis in detecting 6 CFU of the bacterium per PCR reaction. For the detection of *Y. enterocolitica*, the BAC *Screen* and BAC *Ident* kits are available commercially for conventional PCR and agarose gel detection and the LightCycler and ABI Prism PCR systems, respectively.

Random-amplified polymorphic DNA (RAPD) analysis is the amplification of random DNA segments with single primers of arbitrary nucleotide sequence. The difference in banding pattern allows strains belonging to the same species to be identified. Arbitrary primers used in RAPD have been shown to differentiate pathogenic and nonpathogenic strains and their serotypes. RAPD can also reveal coinfection with more than one strain of *Y. enterocolitica*, indicating the potential of this technique to detect strain variations in epidemiological studies.

Treatment of *Yersinia enterocolitica*

Yersinia enterocolitica is significantly implicated in foodborne illness leading to yersiniosis, arthritis, or even septicemia in children. In the majority of cases, patients recover without any treatment. Uncomplicated enterocolitis is a self-limiting disease. However, in older children, certain complications like pseudoappendicular syndrome, mesentric adenitis, ileitis, and extramesentric complications can be prevented by administering antibiotics, like cotrimoxazole. Uncomplicated gastroenteritis is often treated with norfloxacin, ciprofloxacin, or ciprofloxacin/doxycycline.

Regardless of the age, any patient with culture-proven septicemia should be treated. Treatment may shorten both the duration of the disease and the shedding of the bacterium. Aminoglycosides, in combination with third-generation cephalosporins, quinolones, or rifampin and trimethoprim-sulfamethoxazole, are effective against septicemia. No controlled clinical trials have demonstrated the efficacy of treatment, although septicemic patients should receive ampicillin and tetracycline because severe infection has a higher mortality (around 50%) rate.

Systemic infections, extraintestinal fecal infections, and diseases in immunocompromised children have been treated with cotrimoxazole (bactrim, septra, sulfatrim), cefotaxime (cloforan), and for a long time gentamycin (garamycin), which was later found as a failed drug. For cotrimoxazole, the recommended dosage is $160\,mg\,day^{-1}$ PO bid in adults weighing $>40\,kg$. The pediatric prescribed dosage is $8–10\,mg\,kg^{-1}\,day^{-1}$ PO divided bid. Dose-related complications include renal or hepatic impairment, crystaluria, stone formation, and possible hemolysis in glucose-6-phosphate dehydrogenase deficiency. Cefotaxime (claforan) is recommended in an adult dosage of $1–2\,g\,day^{-1}$ divided $q^{6–8\,h}$ intravenously/intramuscularly (IV/IM). Children are injected IV/IM with $150–180\,mg\,kg^{-1}\,day^{-1}$ divided $q^{6–8\,h}$. In patients with meningitis, $300\,mg\,kg^{-1}\,day^{-1}$ IV/IM divided $q^{6–8\,h}$ is recommended. Hypersensitivity to cephalosporin antibiotics has been documented. A dosage of $100\,mg\,kg^{-1}\,day^{-1}$ ciprofloxacin started early in the infection is an effective antibiotic treatment recommended in *Yersinia*-triggered reactive arthritis. However, antibiotic treatment has no effect on fully developed arthritis.

The agents traditionally used to treat human infections including cotrimoxazole, doxycycline, and chloramphenicol may remain the drugs of first choice. Very high failure rates of β-lactams like amoxycillin with or without clavulanate, benzylpenicillin, erythromycin, clindamycin, and first-generation cephalosporins have been observed. Ninety-nine per cent of all *Yersinia* strains have been found to be resistant to amoxycillin. Variations in minimum inhibitory concentrations (MICs) of different biovars are shown in Table 3.

This implies a complex regulation of β-lactamases in *Y. enterocolitica*. Some β-lactamases are prevalent and significantly expressed in specific biovars. Biovar 1A, 1B 2, 3, and 4 and serotypes O:3; O:5, 27; O:6, 30; O:7, 8; O:8, and O:9 are all found to be susceptible to most of the cephalosporins, aminoglycosides, quinolones, cotrimoxazole, doxycycline, piperacillin, tazobactam, and imipenem and highly resistant to cefazolins, amoxycillin, coamoxyclov, and macrolides. In clinical trials, cotrimoxazole and doxycycline have been found to be effective to an extent of 70–75%, whereas cefurozine, ceftazidine, cefaparazone, piperacillin, and gentamycin failed in seven of the eight courses.

Hygienic practices avoiding close contact with infected persons can prevent transmission of yersiniosis, thus obviating the need for drugs. Extreme care needs to be taken while handling natural warm-blooded reservoirs like pigs. Infection of *Y. enterocolitica* starts with the spread of the bacterium from the intestinal contents of the host carrier to the carcass and body parts. Virulent *Y. enterocolitica* serogroup O:3 has been isolated from various surfaces in slaughter houses. Stringent measures must be taken to prevent *Y. enterocolitica* contamination during slaughter with potential growth during subsequent storage and refrigeration. To prevent the initial colonization of *Y. enterocolitica* in reservoirs, application of vaccines or even feeding with an avirulent biotype 1A *Y. enterocolitica* capable of producing an *in-vivo* antagonistic substance (bacteriocin) against virulent *Y. enterocolitica* may be employed.

Research has focused on the development of vaccines and has shown a positive application when tested in mice. Intranasal immunization with *Y. enterocolitica* O:8 extracellular extracts have been shown to induce local and systemic immunity and subsequent protection against nasal infection. Efforts have also been made to design a rational live vaccine by mutating virulence-associated genes like *Yersinia* adhesion A (*Yad* A), Mn-cofactored superoxide dismutase. It has been found that the mutants are able to colonize Peyer's patches for at least 12 days, resulting in the induction of significant antibody titers against *Yersinia* outer-membrane proteins and in priming of *Yersinia* antigen specific $CD4^+$ Th1 cell isolates from spleen. A high level of attenuation does not diminish the immunogenic properties of the mutant strains.

In many experimental cases, oral vaccination is less successful as the colonization of *Y. enterocolitica* is inhibited in the gut. Vaccination has been shown to be more efficient by parenteral immunization than by oral route. Research has also paved the way for designing subunit vaccines using microbial heat-shock protein as a candidate and also interleukin-12 as an efficient alternative adjuvant to immunostimulating complexes for the induction of protective $CD4^+$ Th1-cell-dependent immune responses against bacterial pathogens.

Table 3 Antibiogram of *Yersinia enterocolitica*

Type	*Response*	*Antibiotic*	*MIC (mg l^{-1})*
Biovar 1A	R/IS	Ticarcillin	16–64
		Fosfomycin	
		Amoxycillin/clavulanate	4–32
	S	Chloramphenicol	4.0
		Trimethoprim	1.0
		Ciprofloxacin	0.016
Biovar 1B	S	Amoxycillin/clavulanate	0.5–2.0
Biovar 2	S	Fosfomycin	4–16
Biovar 3	R	Amoxycillin/clavulanate	
		Ampicillin	
		Cefoxitin	
	S	Ticarcillin	4–32
		Fosfomycin	
Biovar 4	R/IS	Ticarcillin	
		Amoxycillin/clavulanate	
	S	Fosfomycin	1–4
Biovar 5	IS	Amoxycillin/clavulanate	
	S	Fosfomycin	1–32

IS, intermediate sensitivity; R, resistant; S, sensitive.

Concluding Remarks

Yersinia enterocolitica remains in the limelight as a prototype to study bacterial invasion and associated pathogenesis. Armed with an array of temperature-inducible, plasmid, and chromosomal-encoded virulence factors, and capable of being transmitted through food, water, and blood transfusions, *Y. enterocolitica* has instigated significant public health concerns. Consequently, extensive research is being carried out to detect its presence in food products and blood meant for human transfusion. Elegant detection techniques have been devised, embracing inputs from immunological and nucleic acid-based techniques. Antibiotic therapy has been adopted for effective treatment in patients with active infection and immunocompromised individuals.

See also: ***Yersinia enterocolitica***: Properties and Occurrence

Further Reading

Aulisio CCG, Mehlman IJ and Snaders AC (1980) Alkali method for rapid recovery of *Yersinia enterocolitica* and *Yersinia pseudotuberculosis* from foods. *Applied and Environmental Microbiology* 39: 135–140.

Bhaduri S, Turner-Jones C, Taylor MM and Lachica VR (1990) Simple assay of calcium dependency for virulent plasmid-bearing clones of *Yersinia enterocolitica*. *Applied and Environmental Microbiology* 28: 798–800.

Bottone EJ (1999) *Yersinia enterocolitica*: overview and epidemiological correlates. *Microbes and Infection* 1: 323–333.

Cleary TG (2000) *Yersinia*. In: Behrman RE, Kliegman RM and Jenson HB (eds) *Nelson Textbook of Pediatrics*, 16th edn, pp. 857–859. London: W.B. Saunders, Harcourt Brace Jovanovich.

Cloak OM, Duffy G, Sheridan JJ, Blair IS and McDowell DA (1999) Isolation and detection of *Listeria* spp, *Salmonella* spp and *Yersinia* spp using a simultaneous enrichment step followed by a surface adhesion immunofluorescent technique. *Journal of Microbiological Methods* 39: 33–43.

Feng P (1992) Identification of invasive *Yersinia* species using oligonucleotide probes. *Molecular and Cellular Probes* 6: 291–297.

Gray JT, WaKabongo M, Campos FE *et al*. (2001) Recognition of *Yersinia enterocolitca* multiple strain infection in twin infants using PCR-based DNA fingerprinting. *Journal of Applied Microbiology* 90: 358–364.

Kampfer P (1999) *Yersinia*. Introduction. In: Robinson RK, Batt CA and Patel PD (eds) *Encyclopedia of Food Microbiology*, vol. III, pp. 2342–2350. London: Academic Press.

Kapperaud G, Varadund T, Skjerve E, Hornes E and Michaelsen TE (1993) Detection of pathogenic *Yersinia enterocolitica* in foods and water by immunomagnetic separation, nested PCR and colorimetric detection of amplified DNA. *Applied and Environmental Microbiology* 59: 2938–2944.

Khare SS, Kamat AS, Doctor TR and Nair PM (1996) Incidence of *Yersinia enterocolitica* and related species in some fish, meat and meat products in India. *Journal of Science Food and Agriculture* 72: 187–195.

Manafi M and Holzhammer E (1994) Comparison of the vitek, API 20E and Gene-trak systems for the identification of *Yersinia enterocolitica*. *Letters in Applied Microbiology* 18: 90–92.

Nagaraju NR, Isloor S, Rao MS and Rajasekhar M (2001) Prevalence of yersinia antibodies in serodiagnosis of bovine brucellosis. *Journal of Veterinary Research* 5: 197–204.

Schiemann DA and Toma S (1978) Isolation of *Yersinia enterocolitica* from raw milk. *Applied and Environmental Microbiology* 35: 54–58.

Stock I and Wiedemann B (1999) An *in vitro* study of the antimicrobial susceptibilities of *Yersinia enterocolitica* and the definition of a database. *Journal of Antimicrobial and Chemotherapy* 43: 37–45.

Vishnubhatla A, Fung DYC, Oberst RD *et al*. (2000) Rapid 5′ nuclease (TaqMan) assay for detection of virulent strains of *Yersinia enterocolitica*. *Applied and Environmental Microbiology* 66: 4131–4135.

Wauters G (1981) Antigens of *Yersinia enterocolitica*. In: Bottone EJ (ed.) *Yersinia enterocolitica*, pp. 41–53. Boca Raton, FL: CRC Press.

YOGURT

Contents

The Product and its Manufacture

N Shah, Victoria University of Technology, Victoria, Australia

Introduction

At the beginning of this century, Nobel Laureate, Elie Metchnikoff, at the Pasteur Institute, linked health and longevity to the ingestion of bacteria present in fermented foods such as yogurt. This observation provided a major boost to the manufacture and consumption of yogurt and other fermented milk products.

Although no records are available to trace the origin of yogurt, it is believed that fermentation was the first technique employed by humans to preserve foods. The word 'yogurt' was derived from the Turkish word 'Jugurt.' Yogurt is defined as 'a product resulting from milk by fermentation with a mixed starter culture consisting of *Streptococcus*

thermophilus and *Lactobacillus delbrueckii* ssp. *bulgaricus*.' However, in some countries, including Australia, other suitable lactic acid bacteria are permitted for use as starter cultures. As a result, some yogurt manufacturers use *Lactobacillus helveticus* and *Lactobacillus jugurti* for yogurt manufacture. The first US Federal Standards of Identity for yogurt were published in 1981. The standard does not permit the use of culture organisms other than *Lb. delbrueckii* ssp. *bulgaricus* and *Sc. thermophilus* for making yogurt, and the titratable acidity must be at least 0.9%, expressed as lactic acid.

Yogurts differ according to their chemical composition, method of production, flavour used and the nature of postincubation processing. Based on the fat content, there are three main types of yogurt: full-fat yogurt, reduced-fat yogurt, and low-fat yogurt (Table 1).

On the basis of the method of production and the physical structure of the coagulum, yogurts are classified as either set or stirred. Set yogurt is the product formed when the fermentation of milk is carried out in a retail container, and the yogurt produced is in a continuous semisolid mass. In contrast, stirred yogurt results when the coagulum is produced from milk, and the gel structure is broken before cooling and packaging. Fluid yogurt can be considered as stirred yogurt of low viscosity.

On the basis of flavorings, yogurts are divided into three categories. Plain or natural yogurt is the traditional product, which has a typical sharp 'nutty' flavor. Fruit yogurts are made by addition of fruits, usually in the form of fruit preserves, puree or jam. Flavored yogurts are prepared from plain or natural yogurt by adding sugar and/or other sweetening agents, flavorings and colorings.

The method of yogurt production has changed very little over the years. However, some newer varieties have been added including frozen-, concentrated-, dried-, and pasteurized yogurt. Postincubation processing of yogurt may lead to pasteurized/UHT yogurt, concentrated yogurt, frozen yogurt, and dried yogurt. Pasteurized/UHT yogurt is prepared by heat- treating yogurt after incubation. Heat treatment destroys yogurt starter bacteria and reduces the levels of volatile compounds associated with the flavor of yogurt. In some countries, such as Australia, heat treatment after production to inactivate the starter culture is not permitted. In other countries, such as the USA and several countries in Europe, heat treatment is permitted; however, auxiliary labeling 'heat-treated after culturing' is required if yogurts are heat-treated after culturing. Frozen yogurt resembles ice cream, and the chemical composition and manufacturing details up to the freezing stage are similar to those of yogurt. Frozen yogurt can be either soft or hard. Dried yogurt can be produced by sun-drying, spray-drying or freeze-drying of yogurt. The drying process transforms the junket into powder. Drying also causes loss of some flavor compounds and destruction of the starter culture. Another type of yogurt, which may find favor among diet-conscious consumers is low-calorie yogurt, in which intense sweeteners are used, and the viscosity of the product is improved by the addition of stabilizers and thickening agents such as carrageenan and/or gelatin.

In the Middle East, concentrated yogurt is prepared by adding wheat flour or parboiled wheat, and the yogurt–wheat mixture is shaped into rolls and sun-dried. The product is popularly known as 'kishk.' In some countries, such as India, Nepal, and Bangladesh, yogurt is made in earthenware pots, and because of evaporation of moisture through the pores of the earthenware pots, the product becomes concentrated.

Table 1 Compositional standards for yogurt

	Full-fat yogurt	*Reduced-fat yogurt*	*Low-fat yogurt*
Fat (%)	≥3.0%	0.5–2	≤0.5
Milk solids not fat (%)	≥8.25	≥8.25	≥8.25
Titratable acidity (%)	≥0.9	> 0.9	> 0.9
pH	≤4.5	≤4.5	≤4.5

Processing and Manufacture of Yogurt

Preparation of Mix

A flow diagram for making yogurt is shown in Figure 1. Traditionally, yogurt is manufactured from milk, which has been concentrated by boiling to increase the viscosity of the product. Nowadays, milk is fortified with nonfat dry milk for the same purpose. The level of fortification varies from as little as 1% to as much as 6%. However, the generally recommended level of fortification is around 3–4%. This increases the total solids level in yogurt mix to 15–17%.

Ultrafiltration and reverse osmosis may also be used to achieve a higher solids content in order to improve the viscosity of the product. Milk concentrated by ultrafiltration to 18–20% total solids has been reported to produce good-quality yogurt. The viscosity of yogurt is almost wholly dependent on the protein content of milk. Hence, a high protein concentration is essential for the production of viscous yogurt. Casein is the major contributor of viscosity

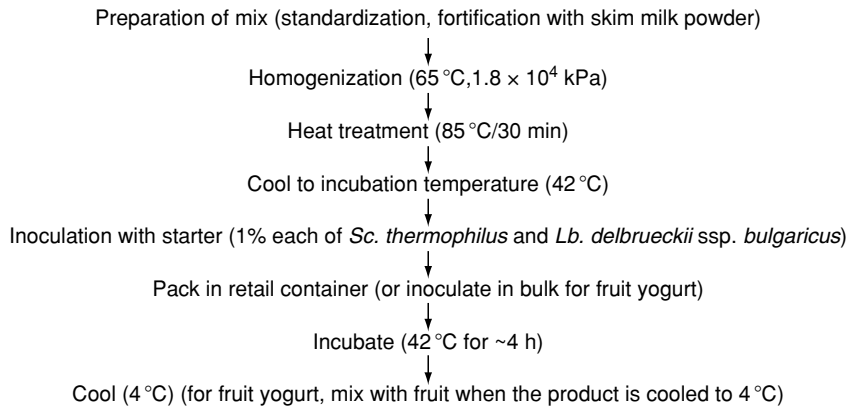

Figure 1 Flow diagram for yogurt manufacture.

followed by fat and whey proteins. Stabilizers can also influence the consistency of yogurt. In practice, gelatin, starch, vegetable gums, carrageenan, and pectin are used most widely as stabilizers for yogurt. The best yogurt texture is achieved by using gelatin at 0.3–0.8%. Good yogurt can be made without the use of added stabilizers, but yogurt without a stabilizer is more vulnerable to a number of stress factors than one that has not been stabilized. When properly chosen and used, stabilizers play an important role in improving the body, texture, mouth feel, and appearance of yogurt. The fat content may vary from 0.5 to 3.5%. Milk fat also tends to 'mask' the acid flavor of yogurt. Obviously, when milk fat is incorporated into the mix formulation, homogenization becomes important for the overall texture quality of yogurt.

Sweeteners are added to plain yogurts, generally with fruits. Addition of sugar is a method of cutting the sharpness of yogurt flavor. An appropriate level of sugar should be used to mask the full degree of acidity, but enough sugar should remain in the product after fermentation for a desirable acid–sugar blend. Up to 10% sucrose is usually used when fruit is added to yogurt. According to Ravula and Shah, a mix containing 4% or more sucrose, may reduce acid production and lower cell counts of both microbial species when incorporated in the mix prior to fermentation. Nonnutritive or intense sweeteners such as aspartame (NutraSweet) may find favor among diet conscious consumers. Since aspartame is approximately 200 times sweeter than sucrose, only a small amount is required for desirable sweetness. Because of this, yogurts containing intense sweeteners require a bulking agent such as polydextrose.

As yogurt mix is prepared, particular attention should be given to blending, homogenization, and heat treatment of the mix. The blending and homogenization steps are important to the uniformity of ingredient distribution. Homogenization is usually carried out before heat treatment, but in some cases, it may take place after the heat treatment.

Homogenization

Homogenization is the typical industrial process used to effect stabilization of the lipid phase against separation by gravity. The specific gravity of milk fat is 0.9, whereas that of skim milk is 1.036. Fat, being lighter, tends to separate from the serum or skim milk if left undisturbed. The diameter of the fat globules in milk ranges from 1 to 15 μm, with an average of 3–4 μm. Homogenization reduces the average diameter of fat globules to 1 or 2 μm. As a result, the fat globules do not cream during the incubation of yogurt. Because of the size reduction, there is usually a four-to sixfold increase in the surface area.

The fat-globule membrane protects the fat globules from lipase. Upon homogenization, the fat-globule membrane is destroyed, and the fat globule is vulnerable to attack by lipase, which is naturally present in milk. Because of this, the milk is pasteurized to inactivate lipase before homogenization. If the milk is to be homogenized before pasteurization, the milk must be pasteurized immediately to prevent lipolysis.

Since the efficiency of homogenization is much better at higher temperatures, the mix is warmed to around 65 °C to liquefy fat and then forced at high pressure (1.8×10^4 kPa) through a small orifice to reduce the size of the fat globules.

Heat Treatment

Yogurt mixes are heated to a high temperature, typically 85 °C for 30 min. Such a high heat treatment has many objectives: (1) to destroy all pathogenic bacteria, (2) to inactivate all the enzymes that may be

present in milk, including lipase, (3) to destroy most of the spoilage-causing bacteria, including thermodurics, and, most importantly, (4) to denature whey proteins, β-lactoglobulin and α-lactalbumin. This heat treatment denatures more than 90% of the β-lactoglobulin compared with 60% of the α-lactoglobulin. A complex is formed between κ-casein and denatured β-lactoglobulin, which increases the hydrophilic properties of the casein, reduces the propensity of the gel to syneresis, and facilitates the formation of a stable coagulum. The hydration effect of the protein is maximal when milk is heated at 85 °C but decreases as the temperature is raised above 85 °C. A higher heat treatment decreases the hydrophilic properties of the β-lactoglobulin–κ-casein complex. As a result, syneresis occurs, and the structure of the yogurt gel becomes weak and fragile.

After heat treatment and complex formation between β-lactoglobulin and κ-casein, the heated milk forms a smooth gel-like structure when the pH drops during fermentation to 4.6, which is the isoelectric point of casein (the pH at which casein proteins carry no electrical charge).

Inoculation and Incubation

After pasteurization, the mix is cooled to 45 °C and inoculated at a level that varies from 0.5 to 5%. The former level (0.5–1%) may be susceptible to growth conditions and a longer incubation time. The maximum amount recommended is 5%. This level will cause very rapid acid production, but leads to defects in aroma, and a large amount of culture must be prepared. The optimum level is 2%, 1% each of *Lb. delbrueckii* ssp. *bulgaricus* and *Sc. thermophilus*.

Lb. delbrueckii ssp. *bulgaricus* hydrolyzes milk proteins, the caseins, thus releasing essential amino acids, including valine, which stimulate the growth of *Sc. thermophilus*. Initially, *Sc. thermophilus* grows rapidly, reducing the pH to around 5.4, which stimulates the growth of *Lb. delbrueckii* ssp. *bulgaricus*, which is acid-tolerant and produces large amounts of lactic acid, which reduces the pH. *Sc. thermophilus* uses oxygen during its growth, which makes oxidation–reduction potential more favorable for *Lb. delbrueckii* ssp. *bulgaricus*; it also produces formic acid, which stimulates the growth of the lactobacillus.

The starter bacterial cells are distributed in the inoculated mix by stirring for 10–15 minutes after the addition of the cultures. This is followed by dispensing the mixes into consumer-sized containers and incubating at 42 °C for approximately 4 h until the pH decreases to 4.5. For the manufacture of stirred yogurt, the inoculated mix is incubated in bulk, and when the pH reaches the desired value, the yogurt is stirred, cooled, mixed with various flavorings, if required, and filled into retail containers.

The incubation temperature of 42 °C is optimal for the associated growth; the optimum temperature of the growth of *Sc. thermophilus* alone is 37 °C, and that for *Lb. delbrueckii* ssp. *bulgaricus* is 45 °C. Incubation temperatures above 42 °C will promote the growth of the *Lactobacillus*, whereas an incubation temperature of less than 42 °C will favor the growth of *Sc. thermophilus*, and lack of flavor results due to the poor growth of *Lb. delbrueckii* ssp. *bulgaricus*. Deviation in either direction will lead to a disturbance in the ratio of lactobacilli to streptococci. The ratio of lactobacilli and streptococci in natural yogurt for optimum flavor should be 1:1. By using monocultures, this ratio of 1:1 can be maintained easily by adjusting the inoculum level of both *Lb. delbrueckii* ssp. *bulgaricus* and *Sc. thermophilus*.

When a 2% inoculum is used at an incubation temperature of 42 °C, the milk will coagulate, and a firm gel will form in 3.5–4 h, and the pH will decrease to 4.4–4.5. When frozen concentrated cultures are used, an incubation period of 5–8 h may be required. Similarly, freeze-dried cultures may require longer times than fresh and active cultures.

A new trend in yogurt making is to incubate the yogurt mix at a significantly lower temperature than normal for a longer period. One of the advantages is that products require less starter cultures and shorter cooling times. An incubation temperature of 30 °C with 0.25% starter culture may take 12–14 h. If the incubation temperature is too low, the growth of yogurt starter may be affected adversely.

The final acidity of yogurt is 0.9–1%. The US standard requires a titratable acidity of 0.9% or higher, whereas the Australian standard requires a pH of 4.5 or lower. According to IDF, the minimum acidity of yogurt should be 0.7%.

The yogurt cultures produce acetaldehyde, which gives natural yogurt its typical flavor that resembles that of green apples. Other volatile compounds include acetic acid and diacetyl.

Cooling

Once the desired acidity has been reached, the product is cooled to <10 °C as quickly as possible. In the case of stirred yogurt, in one-phase cooling, the product is cooled from the incubation temperature to <10 °C prior to the addition of flavoring material and filling. In two-phase cooling, the temperature of the product is reduced to 15–20 °C during the first phase cooling before addition of flavoring materials and filling of containers followed by the second stage

cooling to <10 °C in a cold store. Yogurt quality may be improved by packaging at 24 °C followed by final cooling of the product in the container. To maximize the effect, the second stage of cooling should be carried out as slowly as possible over a 12-h period.

The viscosity of yogurt improves during storage for 1–2 days. During the first 24–48 h of cold storage, an improvement in the physical characteristics of the coagulum is observed, mainly hydration and/or stabilization of casein micelles. Proper hydration is required to avoid syneresis. It is therefore important to delay the sale or distribution of yogurt for 24–48 h.

During cold storage, it is important to minimize rough mechanical handling of the packaged yogurt, and to maintain the temperature of <5 °C. During transport, especially in summer, shaking of yogurt can lead to a reduction in viscosity and syneresis.

Flavored yogurts

Flavored yogurts are prepared by adding flavorings to plain yogurt. Sundae-style yogurt is prepared by layering 15–18% of total weight of yogurt with fruit purée or syrup on the bottom of the containers and then filling the containers with warm inoculated mix, followed by sealing the containers and incubation. The fruit in the product may be mixed with the yogurt gel by consumers before eating.

Swiss or stirred-style yogurts are prepared by blending fruit purée, sucrose, or glucose into bulk prepared fresh plain yogurt. Since the coagulum is broken during blending, plain yogurt is usually prepared with a higher level of stabilizer (0.7%) than normal (0.3%). The product after mixing with fruit is chilled to 4 °C.

Frozen yogurt

Frozen yogurt is a fermented milk or yogurt mix containing a typical yogurt flavor and has been subjected to fermentation by the two yogurt bacteria to a pH of 4.5 or lower or a titratable acidity of 0.9% or higher, expressed as lactic acid. Frozen yogurts, frozen flavored yogurt desserts, yogurt frozen on a stick, or frozen fermented dairy desserts are relatively new products that were developed in the 1960s. Frozen yogurts are prepared in a similar fashion as icecream, and therefore are made most conveniently in icecream factories.

Low-lactose Yogurt

Lactose malabsorbers do not produce sufficient lactase (β-D-galactosidase) and thus cannot hydrolyze the ingested lactose completely. In addition to the gastrointestinal discomfort, such as stomach upset and diarrhea brought about by ingestion of milk by these individuals, a general impairment in the normal digestion process has been observed.

Low-lactose or lactose-free milk for lactose sensitive individuals can be prepared either by the physical removal of lactose by ultrafiltration or by hydrolysis of lactose into the corresponding monosaccharides, glucose, and galactose.

Lactose can be hydrolyzed using strong mineral acid or enzymes. The enzyme β-galactosidase (β-D-galactoside galactohydrolase, EC 3.2.1.23), commonly called lactase, catalyzes the hydrolysis of the β-1 → 4-galactosidic linkage present in lactose. Lactase has been isolated commercially from the fungi, *Aspergillus niger*, *A. oryzae*, and *A. flavus*, or the yeast *Saccharomyces lactis* or from the bacterium *Escherichia coli*. Yogurt bacteria (*Lb. delbrueckii* ssp. *bulgaricus* and *Sc. thermophilus*) possess the highest amount of β-galactosidase activity among the lactic acid bacteria.

β-Galactosidase from *Saccharomyces* (known as neutral lactase) has an optimum activity at pH 6.8–7.0, is stable in the pH range of 6.0–8.5, and works best at 35 °C. It is suitable for treating milk (pH 6.6) or sweet whey (pH 6.2), but the lack of stability below pH 6.0 precludes its use for treating acid whey (pH 4.5). *A. niger* lactase (also known as acid lactase), with a pH optimum of 4.0–4.5, good stability over a wide range of pH (3.0–7.0), and an optimum temperature of 55 °C, is suitable for the modification of acid whey.

Low-lactose yogurt is then produced by the use of low-lactose milk during processing.

Heated or Pasteurized Yogurt

Even though natural yogurt is held at 4 °C, the titratable acidity increases due to residual activity of the starter bacteria. This process is referred in the industry as 'post-acidification,' as a result of which the product becomes too tart. Without heat treatment, the shelf-life of yogurt is 4–5 weeks at 4 °C. Heating destroys most of the starter bacteria and yeast and molds, and if post-processing contamination is avoided, the shelf-life of the product can be extended to 8 weeks. No postacidification occurs in heated yogurt. Heated or pasteurized yogurt can be prepared by heat-treating yogurt in the package at about 55 °C for 30 min, followed by cooling. The main problems associated with pasteurized yogurt are loss of flavor and syneresis. Stabilizers can be used to overcome the latter problem.

Controversies still exist regarding the heat treatment of yogurt after fermentation as heat-treated yogurt contains very few or no viable yogurt bacteria. It also raises questions about the definition of yogurt,

which must contain an abundant viable count of *Sc. thermophilus* and *Lb. delbrueckii* ssp. *bulgaricus*. Yogurt is considered to be an excellent nutritional food, and consumers ingest several million live cells through a typical serving of yogurt.

Prebiotics, Probiotics, and Synbiotics

The term 'probiotic' is derived from two Greek words meaning 'for life.' Probiotic organisms and substances produced by these organisms contribute to the microbial balance in the intestines. The generally accepted definition of probiotics is that 'they are live microbial food or feed supplements that provide a beneficial effect on hosts (human or animal) by improving the microbial balance in the intestine.' The intake of these bacteria is reported to help restore the balance in the intestinal microflora, which may have been lost due to stress, antibiotic use, or illness. The major strains of bacteria used in probiotics, *Lb. acidophilus* and *Bifidobacterium* spp., are dominant organisms of human large intestines. These microorganisms are claimed to inhibit the growth of pathogenic organisms through the production of organic acids and bacteriocins. Other benefits include reduction in lactose malabsorption, suppression of potentially harmful enzymes, increased immune response due to increased production of secretory immunoglobulins A, reduction in serum cholesterol, and antimutagenic effects.

The beneficial effects of the presence of bifidobacteria in the gastrointestinal tract are dependent on their viability and metabolic activity. The growth of bifidobacteria is dependent on the presence of complex carbohydrates such as oligosaccharides and other substrates such as *N*-acetyl glucosamine and lactulose. These carbohydrates that stimulate the growth of bifidobacteria are known as 'bifidogenic factors.' Some oligosaccharides, due to their chemical structure, are resistant to digestive enzymes and therefore pass into the large intestine, where they become available for fermentation by bifidobacteria.

Compounds that are either partially degraded or not degraded by the host and are preferentially utilized by bifidobacteria as carbon and energy sources are defined as 'prebiotics.' Some of the bifidogenic factors that are of commercial significance include fructo-oligosaccharides, lactose derivatives (such as lactulose, galacto-oligosaccharides), isomalto-oligosaccharides, xylo-oligosaccharides, gluco-oligosaccharides and soybean oligosaccharides. Resistant starch and nonstarch polysaccharides are classified as colonic foods but not as prebiotics because they are not metabolized by certain beneficial bacteria. Oligosaccharides have been recognized for their health benefits in Japan, and many products containing oligosaccharides were developed during the 1990s. Products that contain both prebiotics and probiotics are referred to as 'synbiotics.' An example of synbiotic includes SymBalance yogurt, which contains inulin as prebiotic and *Lb. reuteri*, *Lb. acidophilus*, and *Lb. casei* as probiotics.

It is not clear how probiotics work. Acidification of the gut is claimed to be one of the mechanisms. Breast-fed infants have a much higher percentage of *Bifidobacterium bifidum* than formula-fed infants. The intestinal microflora of the latter group is more like that of adults; it is a mixed microflora, including coliform bacteria. Bifidobacteria produce acetic acid, butyric acid, lactic acid, and pyruvic acid. The lactic acid and acetic acid account for more than 90% of organic acids produced. It is widely accepted that because of acid production by *Lb. acidophilus* and bifidobacteria, the enteropathogenic bacteria are unable to grow. The growth of clostridia and *E. coli*, when cocultured with bifidobacteria, has been found to be inhibited, even at a neutral pH, suggesting that acid production may not be solely responsible for inhibition. Metabolites produced by bifidobacteria may be partly responsible for the inhibition of pathogens.

Acidophilus and Bifidus Yogurt (AB Yogurt)

Lb. acidophilus and bifidobacteria are normal inhabitants of the intestine of many animals including man. *Lb. acidophilus* is Gram positive and rod-shaped, while bifidobacteria are Gram-positive rods of variable morphology that show branching and pleomorphism. Bifidobacteria were first isolated by Tissier at the Pasteur Institute, Paris, France, and predominate the gut flora in breast-fed infants.

Yogurt containing *Lb. acidophilus* and bifidobacteria has gained popularity in many countries, including Japan, France, Germany, Canada, Australia, and USA, and more than 70 products containing *Lb. acidophilus* and bifidobacteria, including sour cream, buttermilk, yogurt, milk powder, and frozen desserts, are produced world-wide. It is estimated that about 11% of all yogurt sold in France now contains *Lb. acidophilus* and *Bifidobacterium* spp. In 1978, the Yakult Co. launched a Bifidus fluid yogurt named Milmil™, which contains *Bifidobacterium breve*, *B. bifidum* and *Lb. acidophilus*. More than 54 different types of milk products containing *Lb. acidophilus* and *Bifidobacterium bifidum* are marketed in Japan.

Lb. acidophilus and *Bifidobacterium* spp. are difficult to propagate because of their specific nutritional requirements. Bifidobacteria are not as acid-tolerant as *Lb. acidophilus*, and the growth of

Bifidobacterium species is significantly retarded below pH 4.0. *Lb. acidophilus* and *Bifidobacterium* spp. are slow acid producers; the slow growth rate of these organisms can be compensated by adding a higher level of inoculum, such as 5 or 10%. Yogurt bacteria are usually added to carry out fermentation. If pure cultures of *Lb. acidophilus* and/or *Bifidobacterium* spp. are used, the time required to reduce the pH of milk to 4.5 could be as long as 18–24 h at 37 °C. When yogurt bacteria (*Lb. delbruecckii* ssp. *bulgaricus* and *Sc. thermophilus*) and AB (*Lb. acidophilus* and *Bifidobacterium* spp.) cultures are used, the incubation time is about 4 h.

Recent Advances in Probiotic Yogurt

The most commonly used species in commercial probiotic products are *Lb. acidophilus*, *Lb. casei*, *Lactobacillus* GG (a close relative of *Lb. casei* subgroup *rhamnosus*, ATCC 53103), *B. bifidum*, *B. longum*, *B. breve*, and *B. infantis*. Additional blends are also being investigated, such as *Lb. reuteri*, *Lb. plantarum*, and *Lb. casei*.

In Australia and Europe, yogurt containing *Lb. acidophilus* and *Bifidobacterium* spp. is referred to as AB yogurt. The recent trend is to incorporate *Lb. casei* in addition to *Lb. acidophilus* and *Bifidobacterium* spp, and such products are known as 'ABC yogurt.'

A wide variety of probiotic cultured products are now available world-wide such as Nu-Trish a/B, USA (acidophilus + bifidus), AB-yogurt, Denmark (acidophilus + bifidobacteria + yogurt culture), Miru-Miru, Japan (*Lb. acidophilus* + *Lb. casei* + *B. breve*), Bifighurt (*B. longum* + *Sc. thermophilus*). Products that contain both prebiotics and probiotics include SymBalance yogurt (*Lb. reuteri* + *Lb. acidophilus* + *Lb. casei* + bifidobacteria, inulin) produced in Switzerland and Fysiq (*Lb. acidophilus* + Raftiline brand prebiotic).

Several probiotic products contain both yogurt bacteria and one or more types of probiotic bacteria. Because of sensitivity to acid, *Lb. acidophilus* and *Bifidobacterium* spp. in yogurt begin to die within a few days after manufacture because of acid produced during manufacture and storage. In order to provide health benefits, the suggested level for probiotic bacteria is 10^6 cfu per gram of a product. However, studies have shown a low viability of probiotics in market preparations. Many yogurt manufacturers use a starter culture devoid of *Lb. delbrueckii* ssp. *bulgaricus* but a combination of *Lb. acidophilus*, bifidobacteria and *Sc. thermophilus* (known as 'ABT') as starter cultures to overcome the postacidification problem. However, the use of ABT starter culture increases incubation time significantly as *Sc. thermophilus* is the main organism responsible for fermentation in ABT starter cultures, and this organism is less proteolytic than *Lb. delbrueckii* ssp. *bulgaricus*. Bifidobacteria are anaerobic and are fastidious organisms, requiring specific growth factors and prefering a low oxidation–reduction potential for growth.

Cysteine is usually incorporated in the media used for selective enumeration of bifidobacteria. Cysteine is an essential growth factor for bifidobacteria and also reduces the oxidation–reduction potential for optimum growth of anaerobic bifidobacteria. Dave and Shah have shown a better viability of bifidobacteria, in yogurt made from ABT starter cultures, with an additional source of nitrogen, such as acid casein hydrolysate, than with lowering the oxidation–reduction potential with cysteine. These authors have also shown that incorporation of cysteine (at $> 250\,mg\,l^{-1}$) to lower the redox potential affected the growth of *Sc. thermophilus*, as this organism is considered aerobic. Since *Sc. thermophilus* is the sole fermenting organism in ABT starter cultures, the incubation time increased significantly.

The viability of bifidobacteria in yogurt made with starter cultures containing both yogurt bacteria was not affected as much, as *Lb. delbrueckii* ssp. *bulgaricus* is proteolytic and provides peptides and amino acids for the growth of bifidobacteria.

See also: ***Bifidobacteria* in Foods**; **Casein and Caseinates**: Uses in the Food Industry; **Heat Treatment**: Chemical and Microbiological Changes; **Homogenization**; **Lactic Acid Bacteria**; **Pasteurization**: Principles; **Prebiotics**; **Probiotics**; **Starter Cultures**; **Sweeteners**: Intensive; Others; **Yogurt**: Yogurt-based Products; Dietary Importance

Further Reading

Augustin MA, Cheng LJ and Clarke PT (1999) Effects of preheat treatment of milk powder on the properties of reconstituted set skim milk yogurts. *International Dairy Journal* 9: 415–416.

Dave RI and Shah NP (1997a) Viability of yoghurt and probiotic bacteria in yoghurts made from commercial starter cultures. *International Dairy Journal* 7: 31–41.

Dave RI and Shah NP (1997b) Effect of cysteine on the viability of yoghurt and probiotic bacteria in yoghurt made with commerical starter cultures. *International Dairy Journal* 7: 537–545.

Dave RI and Shah NP (1998) Ingredient supplementation effects on viability of probiotic bacteria in yoghurt. *Journal of Dairy Science* 81: 2804–2816.

Guzman-Gonzalez M, Morais F and Amigo L (2000) Influence of skimmed milk concentrate replacement by dry dairy products in a low-fat set yoghurt model system.

Journal of the Science of Food and Agriculture 80: 433–438.
Hekmat S and McMahon DJ (1997) Manufacture and quality of iron-fortified yogurt. *Journal of Dairy Science* 80: 3114–3122.
IDF (1996) *Oligosaccharides and Probiotic Bacteria. Bulletin 313*. Brussels: International Dairy Federation.
Kosikowski FV and Mistry VV (1997) *Cheese and Fermented Milk Foods*. Westport, CT: FV Kosikowski.
Kurmann JA and Rasic JLJ (1991) Health potential of products containing bifidobacteria. In: Robinson RK (ed.) *Health Potential of Products Containing Bifidobacteria: Therapeutic Properties of Fermented Milks*. London: Elsevier Applied Science.
Mazza G (ed.) (1998) *Functional Foods: Biochemical and Processing Aspects*. Lancaster, PA: Technomic Publication.
Modler WH (1994) Bifidogenic factors – sources, metabolism and applications. *International Dairy Journal* 4: 383–407.
Ordonez GA, Jeon IJ and Roberts HA (2000) Manufacture of frozen yogurt with ultrafiltered milk and probiotic lactic acid bacteria. *Journal of Food Processing and Preservation* 24: 163–176.
Ravula RR and Shah NP (2000) Influence of water activity on fermentation, organic acid production and viability of yoghurt and probiotic bacteria. *Australian Journal of Dairy Technology* 55: 127–131.
Robinson RK (ed.) (1986) *Modern Dairy Technology, vol. II. Advances in Milk Products*. London: Elsevier Applied Science.
Shah NP (2000) Probiotic bacteria: selective enumeration and survival in dairy foods. *Journal of Dairy Science* 83: 894–907.
Shah NP, Lankaputhra WEV, Britz M and Kyle WSA (1995) Survival of *L. acidophilus* and *Bifidobacterium bifidum* in commercial yoghurt during refrigerated storage. *International Dairy Journal* 5: 515–521.
Tamime AY and Deeth HC (1980) Yoghurt: technology and biochemistry. *Journal of Food Protection* 43: 939–977.
Tamime AY and Robinson RK (1999) *Yoghurt: Science and Technology*. Cambridge: Woodhead Publishing Ltd.

Yogurt-based Products

A Y Tamime, Ayr, UK

Introduction

The transformation of milk into yogurt, with a starter culture composed of *Streptococcus thermophilus* and *Lactobacillus delbrueckii* ssp. *bulgaricus*, can easily extend the shelf-life of milk from a few days to about 3 weeks. However, the starter cultures employed in the manufacture of biofermented milks including yogurt-related products have been reviewed recently. (*See* **Yogurt**: The Product and its Manufacture.) The main organisms belong to the following genera: *Lactobacillus*, *Bifidobacterium*, *Enterococcus*, and *Pediococcus*. The application of the different preservation techniques known to humans for thousands of years could extend the shelf-life of yogurt to a few months, or even indefinitely. Examples of such postfermentation processes include heating, concentration, freezing, and/or drying, but it is evident that such treatments will alter the characteristics of the end product. Some of these processes are still carried out using traditional methods, but limited data are available in the literature; however, some of these processes have been mechanized and developed by industrial organizations, and as a consequence, sometechnical data are somewhat limited. (*See* **Starter Cultures**.)

Methods of Manufacture

Pasteurization

Traditionally, natural yogurt in rural areas in the Middle East is heated gently for a few hours over low fires using a special wood, and the product is known as 'laban mudakhan' or 'smoked yoghurt.' The application of heat inactivates the starter culture organisms and their enzymes, as well as other contaminants, e.g., undesirable bacteria, yeasts, and molds. The use of such methods enables the nomads in that region to extend the shelf-life of yogurt for a few weeks, or until they reach a market to sell their product. Alternatively, the hot yogurt is placed in a clean jar and covered with a layer of olive oil or tallow, so that the 'smoked yogurt' is preserved over the winter months.

In mechanized plants, natural/plain, fruit, flavored, or drinking yogurts are subjected to heat treatment after the fermentation stage to prolong the shelf-life of the product. The time–temperature relationships, which are used to achieve this objective, are dependent on: (1) the level of acidity, (2) the method of heating and packaging, and (3) the storage conditions. Set-type yogurt is heated between 60 and 85 °C for up to 50 min, and in some instances, the heating is carried out under a pressure of about 0.2 MPa. The other types of yogurt are pasteurized/ultrahigh temperature (UHT), which is heated at 65 to >100 °C for up to 50 s. In general, the heating and packaging methods can be divided into the categories of pasteurized/UHT yogurts given in **Table 1**. (*See* **Heat Treatment**: Ultrahigh Temperature (UHT) Treatments.)

Postfermentation heating of yogurt causes the separation of the aqueous phase from the suspended casein particles. This is mainly due to the

phenomenon of aggregation/dehydration of the caseins caused by heating at the isoelectric point (pH 4.6). Certain hydrocolloids (e.g., pectins, alginate, carboxymethyl cellulose, or prophylene glycol) are negatively charged, and, when added to yogurt before the heat treatment stage, they interact with the positive charges of the casein below its isoelectric point, and the separation of the two phases in the product is avoided. (*See* **Casein and Caseinates**: Methods of Manufacture.)

Concentration

Concentrated or strained yogurt is a popular product in the Middle East region. The product is known as labneh, labaneh, or lebneh (in most Arab countries), mast or mastou (in Iraq), leben zeer (in Egypt), tan or than (in Armenia), Greek yogurt or Greek-style yogurt (in the UK), stragisto or tzatziki (in Greece), torba or suzme (in Turkey), basa or zimne (in the Balkans), and yogurt cheese (in some parts of the world). Other closely related products, which are produced in different countries and using mesophilic, thermophilic, and/or lactose-fermenting yeast starter cultures, are ymer (in Denmark), chakka and shirkhand (in India) and skyr (in Iceland).

The traditional and rural method of production consists of straining cold and unsweetened natural/plain yogurt using a cloth bag, animal skin, or earthenware vessel. The method, in comparison with large, factory-scale operations, is slow, labor-intensive, unhygienic, and cumbersome, and gives low yields due to the residues left in the bag. Typical compositional standards for labneh (e.g., in Saudi Arabia and Lebanon) range between 22 and 26 g 100 g^{-1} total solids and 7 and 10 g 100 g^{-1} fat, respectively.

The factory-scale production of concentrated or strained yogurt includes the methods described below.

Traditional Process

The cold natural/plain yogurt is stirred and emptied into cloth bags of about 25 kg. The bags are stacked on top of each other in a vertical press that is located in a refrigerated room. Pressure is applied in order to assist in whey drainage for a duation of 12–18 h. On the following day, the concentrated product is emptied into a mixing bowl to obtain a uniform texture prior to packaging. Alternatively, long, horizontal filter cloths can be used; the long sides are supported on poles and gently oscillated up and down, while slight lateral pressure is applied. This method of processing is known as the modified 'Berge' system, and was developed in France in the mid-1960s for the production of fresh curd cheese.

The application of more pressure and a longer dewheying stage will yield a product that contains $\geq$ 30 g 100 g^{-1} total solids, and is known as 'yogurt cheese.' This highly concentrated product can be shaped into balls by hand, placed in a jar, and preserved in oil. In Lebanon, mainly goat's milk is used for the production of yogurt cheese, which is known locally as 'labneh anbaris.' Herbs and spices are added to the curd after concentration, and before it is shaped into balls; the product is referred to as 'shanklish' (in the Lebanon).

Table 1 Categories of pasteurized/UHT yogurts

Temperature	*Packaging*	*Shelf-life/storage*
Pasteurization (low)	Hot	2–3 weeks/cold
Pasteurization (high)	Cold/aseptic	1–2 months/cold
UHT	Hot/aseptic	Several months/ambient

Mechanical Separators

The production of strained yogurt by centrifugation of heated yogurt has been used successfully in experimental trials and commercially in different countries; concentration is achieved using a nozzle or Quarg separator. Skimmed milk should be used for the manufacture of the yogurt, and the fermented milk is stirred vigorously, heated to about 60 °C, cooled to about 40 °C and concentrated to 18 g 100 g^{-1} total solids, cooled to about 12 °C, blended with cream, and finally packaged. If whole milk is used instead, the nozzles of the separator will clog. However, recent developments in the design of the centrifugal separators have made it feasible to use fermented whole milk for the production of strained yogurt.

After acidification (i.e., pH 4.6–4.8), the fermented milk is heated to 60 °C to inactivate the culture and control the level of acidity, and then deaerated for 15–20 min to assist the separation of whey in the separator. A centrifugal pump transports the fermented milk through a switchable double strainer to break up any lumps before it enters the separator. The concentrated product leaving the separator is blended with cream and seasoning (e.g., salt, herbs or fruit flavors – optional), cooled, and packaged. A typical chemical composition for strained yoghurt is 24 g 100 g^{-1} total solids and 9.6 g 100 g^{-1} fat (about 40% fat-in-dry-matter); the composition of the whey is 6.1 g 100 g^{-1} total solids, consisting mainly of lactose and minerals, but also about 0.5 g 100 g^{-1} fat. Capacities of such separators are up to 6.5 tonnes h^{-1}, depending on the composition of the milk used and the acidity of the fermented milk prior to concentration.

Ultrafiltration (UF)

Two different systems of UF have been used for the production of strained yogurt. Firstly, the standardized milk is concentrated by UF to the desired solids content before homogenization, heat treatment, and fermentation, and secondly, the warm yogurt (about 45 °C) is concentrated by UF. (*See* **Filtration of Liquids.**)

In the former system of production, the UF retentate is fermented in the retail container – as with the manufacture of natural set yogurt – and the firmness of the product is much greater than that of products made using the traditional (cloth bag) method or by the UF of warm yogurt. However, when the product is broken and smoothed by passing it through a lactic curd homogenizer, the strained yogurt showed signs of whey syneresis, and an excessive reduction in its firmness is observed. These faults are not evident in the other types of strained yogurt, and the product should be processed differently in order to overcome such faults.

The quality of strained yogurt made by UF of warm yogurt closely resembles the traditional product in terms of elasticity, firmness, and structure. The stages of manufacture consist of: (1) production of yogurt from whole milk, (2) UF of yogurt at 35–45 °C using batch or multistage systems, (3) partial cooling to 20 °C, fruit blending (optional), and packaging, and (4) cooling in the retail container to < 10 °C. No fat losses are observed in the permeates, and the solids content consists almost exclusively of lactose, organic acids, minerals and *very* slight losses of low-molecular-weight nonprotein nitrogen.

A wide range of UF plants are available on the market that can be used for the production of strained yogurt on a large scale. A typical example is illustrated in Figure 1, and according to the supplier, the manufacturing process is as follows. The standardized milk (e.g., 12.4 g 100 g^{-1} total solids and 3.5 g 100 g^{-1} fat) is preheated to 60 °C, homogenized at 14.7 MPa, heated in a plate heat exchanger (PHE) to 95 °C for 5 min in a holding tank before cooling to 40–45 °C in the regeneration section of the PHE. After the fermentation period, the warm yogurt is heated at 58–60 °C for 3 min in the PHE, cooled to 40 °C, concentrated in a two- to four-stage UF plant, cooled in a plate cooler to about 20 °C, and finally packaged. The degree of concentration of yogurt using a four-stage UF plant, for example, could be adjusted to give 14, 16, 19, and 22 g 100 g^{-1} total solids, respectively.

Product Formulation

It is feasible to manufacture strained yogurt from recombined dairy ingredients. The process involves reconstitution of powder(s) (whole, skimmed milk, retentate, and/or caseinate) in water and blending it with anhydrous milk fat/butter oil, stabilizer (e.g., Cremodan Mousse 31 (Danisco Ingredients (UK) Ltd.)), and salt (optional). The recombined milk is

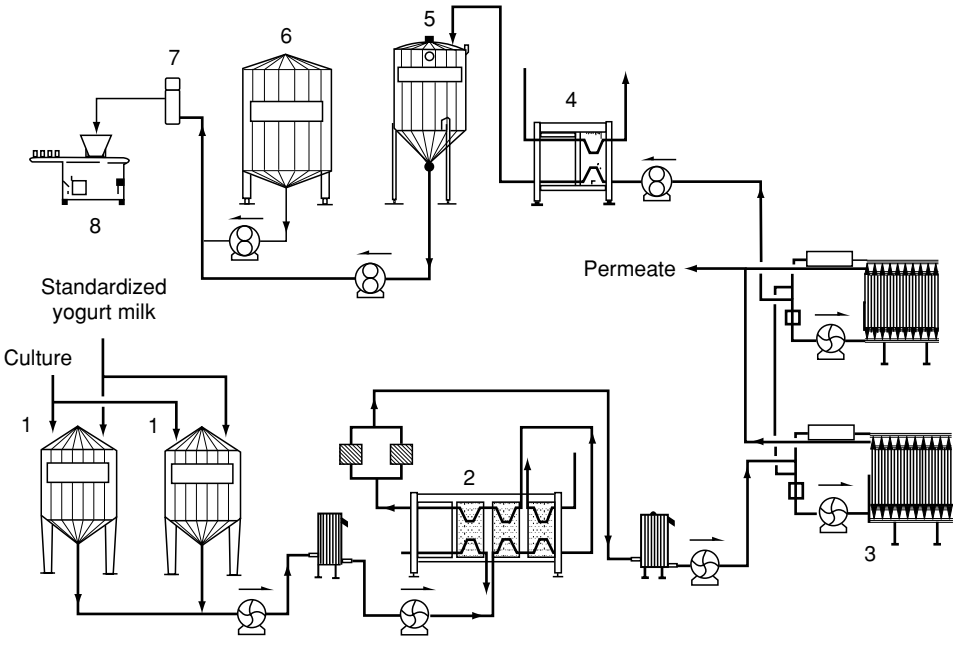

Figure 1 Simplified flow chart for the manufacture of strained yogurt by UF: 1, fermentation tanks; 2, plate heat exchanger; 3, two- to four-stage UF plant; 4, plate cooler; 5, buffer tank; 6, fruit tank (optional); 7, in-line mixer; 8, packaging machine. Reproduced courtesy of Tetra Pak (Processing System Division) A/B, Lund, Sweden.

Table 2 Typical compositions (g 100 g^{-1}) of full- and low-fat strained yogurts

Component	*Full-fat*	*Low-fat*
Fat	10.0	4.2
Solids-not-fat	13.8	17.4
Salt	0.5	0.5
Cremodan Mousse 31	0.8	0.9
Total solids	26.1	23.0

Table 3 Chemical composition (g 100 g^{-1}) of different brands of strained yogurt that are marketed in the UK

	Brands					
Composition	*A*	*B*	*C*	*D*	*E*	*F*
Total solids	21.70	19.25	21.98	21.56	19.78	14.27
Protein	7.31	4.46	4.89	7.74	6.00	9.90
Fat	10.68	9.99	10.11	10.48	9.44	0.21
Carbohydrate[a]	3.14	3.55	6.00	2.76	3.70	3.53
Ash	0.57	0.75	0.98	0.58	0.64	0.63

[a]Calculated by difference.
After Tamime (unpublished data).

handled and processed in a similar way to the production of yogurt. After the fermentation stage, the product is precooled to about 20 °C and packaged, and the final cooling to 5 °C takes place in the cold store. Typical compositions of full- and low-fat strained yogurts are given in Table 2. (*See* **Recombined and Filled Milks.**)

At present, there are no compositional standards for this product in the UK. However, Table 3 illustrates the variation in the chemical composition of leading brands of strained yogurt that are marketed in the UK.

Freezing

Frozen yogurt resembles icecream, in that the fresh stirred yogurt is stabilized, fortified with fruit base (syrup or pieces), whipped, and frozen. In general, the product is classified into three main categories: soft, hard, or mousse. The recommended ratio of yogurt to fruit base for the manufacture of soft and hard frozen yogurts is 80:20 and 65:35, respectively. The outlet freezer temperature of these two products is −6 °C, and the storage temperature for soft and hard frozen yogurts is −6 °C and −25 °C, respectively. Although air is normally used at the whipping/freezing stage, a longer-shelf-life product can be achieved by using nitrogen rather than air.

The chemical composition of frozen yogurt may differ from one manufacturer to another, or between countries. This product is very popular in the USA, and the variation in the chemical composition of a number of commercial samples is shown in Table 4.

Drying

The objective of manufacturing powdered yogurt is to produce a product that is stable during prolonged storage and readily utilizable. The stages of manufacture depend on the method of drying used.

Traditionally, natural/plain yogurt or, in some instances, strained yogurt is mixed with 'burghol or bourghoul' (i.e., parboiled cracked wheat) in a ratio of 4:1. The mixture is kneaded daily for around a week, salted, passed through a meat mincer, shaped into rolls, and placed in the sun to dry. The dried product is called 'kishk', and it is either sold as rolls or ground into flour. Kishk as a dish is prepared by rehydrating the dried product with water and then simmering the mix gently over a fire. The product is rather thick, similar to porridge. Chopped onions, coriander, and/or tomatoes can be added to the reconstituted 'kishk,' and such a dish is normally consumed with bread. (*See* **Fermented Milks**: Types of Fermented Milks.)

Other types of dried, fermented dairy product are known in the Arabian peninsula as 'oggtt' or 'maddeer' or 'buqle'. The desert dwellers allow goat's or camel's milk to sour naturally, churn it to make butter, and then boil the buttermilk until it thickens. When cooled, the concentrated buttermilk is shaped into small cakes and left in the sun to dry.

At present, powdered yogurt, which is sometimes known as dried or instant yogurt, is manufactured in relatively small quantities using normal driers. Some of the manufacturing techniques have been patented, and the dried product is aimed towards the do-it-yourself consumer market, the baby food manufacturers, and the food/baking industries.

Table 4 Range of composition (g 100 g^{-1}) of commercial frozen yogurts sold in the US market

Flavor	*Fat*	*Protein*	*Ash*	*Total solids*	*pH*
Vanilla	1.79–5.94	3.52–3.84	0.70–1.01	28.83–34.21	6.37–7.10
Chocolate	3.15–5.73	2.94–4.19	0.87–1.06	31.08–36.70	6.36–7.10
Strawberry	1.69–5.31	1.61–3.15	0.82–1.05	31.20–37.60	4.37–5.70

Data from Tieszen KM and Baer RJ (1989) Composition and microbiological quality of frozen yoghurts. *Cultured Dairy Products Journal* 24(4): 11–14, with permission.

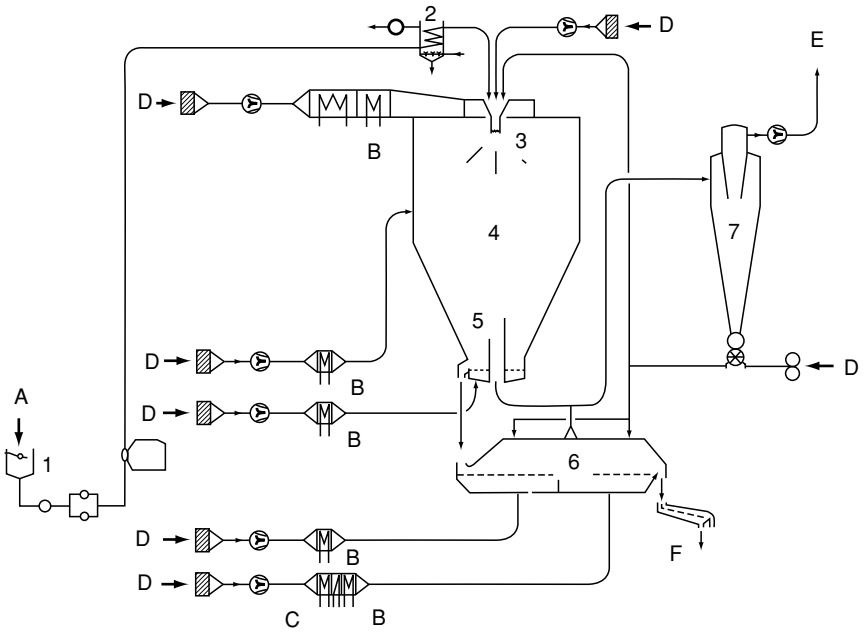

Figure 2 Schematic illustration of a three-stage drying plant for the manufacture of dried yoghurt; A, product inlet; B, steam; C, cooling water; D, air inlet; E, air outlet; F, product outlet; 1, feed tank; 2, preheater; 3, atomizer; 4, spray drying chamber; 5, integrated fluid bed; 6, external fluid bed; 7, cyclone. Reproduced courtesy of APV Norcid A/S, Copenhagen.

Yogurt presents no difficulties when spray dried, but some precautionary measures that should be considered included the following: firstly, the concentration stage, before drying, should be carried out at a low temperature (about 50–60 °C) to minimize scorching on to the surfaces of the evaporator, or discoloration of the final powder, and secondly, the processing conditions must be moderate in order to ensure a high viable cell count of starter culture organisms in the dried product. The acidified product is concentrated using an ordinary evaporator to 25–36 g 100 g^{-1} solids at 57–58 °C. The concentrated buttermilk, which is highly viscous, is pumped to the spray drier at 43 °C and dried using an inlet air drying temperature between 175 and 190 °C. The moisture content in the dried product is around 4 g 100 g^{-1}, and the tapped bulk density is 0.77–0.83 g cm^{-3}. (*See* **Drying**: Spray Drying.)

According to APV Nordic A/S, which is one of the leading drier manufacturers, yogurt can be dried in a three-stage drying plant, and an illustration is shown in Figure 2. The yogurt is concentrated to 35 g 100 g^{-1} total solids, preheated, and nozzle atomized into the drying chamber. The inlet and outlet air

Table 5 Range of composition (g 100 g^{-1}) of different types of powdered yogurt products

Product	*Total solids*	*Protein*	*Fat*	*Total sugars*	*Ash*
Oggt					
Sheep	91.9–95.6	31.7–37.3	14.5–39.3	19.3–32.5	5.3–7.6
Cow	93.7–96.3	26.2–31.3	11.0–25.4	38.7–44.4	6.0–7.2
Goat	92.5–93.1	26.3–30.4	18.9–28.9	34.8–37.3	5.3–6.5
Madeer	91.8–96.1	35.5–36.4	13.4–15.3	34.4–37.4	7.6–7.9
Yogurt/plain kishk	95.6–96.0	33.0–54.4	1.0–7.9	29.5–54.0	3.8–7.0
Kishk					
Commercial	87.0–94.5	8.9–23.5	1.6–16.1	31.0–68.8	2.0–9.1
Laboratory	90.5–92.5	14.5–19.7	nr[a]	nr	4.4–8.7
Different cereals	87.5–94.8	17.6–25.8	6.4–9.7	56.3–67.9	2.2–4.9
Milk and soy	90.1–94.8	16.1–28.2	nr	nr	3.5–5.3
Whey	90.3	13.3	nr	nr	nr

[a]nr: data not reported.

Data compiled from Tamime AY and Robinson RK (1999) *Yoghurt Science and Technology*, 2nd edn. Cambridge: Woodhead, with permission.

drying temperatures are 160 and 65 °C, respectively. The partially dried yogurt particles fall down on to the integrated fluid bed drier in the bottom of the drying chamber. Such particles form a fluidized layer, which is further dried.

Finally, the powder is transferred to an external fluid bed drier for final drying and cooling. The spent drying air from both the drying chamber andthe external fluid bed is drawn through cyclone(s) to separate the very fine powder particles (fines) from the air. The fines are fed back to the external fluid bed drier, where they are mixed with the bulk of the powder. During drying, the maximum temperature of the powder is 55 °C and that of the product outlet, 25 °C. Such powder contains $2\,g\,100\,g^{-1}$ moisture and has a tapped bulk density of $0.5\,g\,cm^{-3}$.

It is evident that dried-yogurt products (traditional or industrial) are available to consumers in rural or industrialized countries, and Table 5 illustrates some typical compositional analyses. Note that some products contain low quantities of fat, which indicate that: (1) skimmed milk has been used rather than full-fat milk, or (2) in some traditional dried products, the fermented milk is first churned to make butter, and then the buttermilk is dried for preservation purposes.

See also: **Casein and Caseinates**: Methods of Manufacture; **Drying**: Spray Drying; **Fermented Milks**: Types of Fermented Milks; **Filtration of Liquids**; **Heat Treatment**: Ultra-high Temperature (UHT) Treatments; **Recombined and Filled Milks**; **Starter Cultures**; **Yogurt**: The Product and its Manufacture

Further Reading

Bylund G (1995) *Dairy Processing Handbook*. Lund: Tetra Pak Processing Systems.

Lehmann HR, Dolle E and Bucker H (1991) *Processing Lines for the Production of Soft Cheese*, 3rd edn. Technical and Scientific Document No. 8. Germany: Westfalia Separator AG.

Marshall VME and Tamime AY (1997) In: Law BA (ed.), *Microbiology and Biochemistry of Cheese and Fermented Milk*, 2nd edn. London: Blackie Academic & Professional.

Masters K (1991) *Spray Drying Handbook*, 5th edn. Harlow: Longman Scientific and Technical.

Pisecky J (1997) *Handbook of Milk Powder Manufacture*. Copenhagen: Niro A/S.

Tamime AY and Marshall VME (1997) In: Law BA (ed.) *Microbiology and Biochemistry of Cheese and Fermented Milk*, 2nd edn. London: Blackie Academic & Professional.

Tamime AY and Robinson RK (1999) *Yoghurt Science and Technology*, 2nd edn. Cambridge: Woodhead.

Tieszen KM and Baer RJ (1989) Composition and microbiological quality of frozen yoghurts. *Cultured Dairy Products Journal* 24(4): 11–14.

Dietary Importance

J Buttriss, British Nutrition Foundation, London, UK

Introduction

Yogurt is a fermented milk product and as such is a means of preserving the nutrients in milk. A wide variety of yogurts are now available around the world, ranging from very-low-fat fruit yogurts to Greek-style yogurt with a fat content around 8 g per 100 g. Yogurt can be made from cows', ewes', goats', or buffalo's milk. This article reviews the nutritional composition of a range of yogurts, provides data on yogurt consumption around the world and discusses the importance of yogurt for different population groups. (*See* **Buffalo**: Milk; **Milk**: Dietary Importance; **Sheep**: Milk.)

Since yogurt is derived from milk, it provides protein, calcium, and other minerals, and a range of vitamins (**Table 1**). Levels of some vitamins, such as vitamin B_1 and pantothenic acid, are reduced as they are utilized by the bacterial culture used to produce the yogurt. However, folic acid levels are typically higher than in milk since folic acid is produced by the bacteria. Refer to individual nutrients.

Some people have speculated that the bioavailability of some minerals in yogurt is enhanced, but research to support this claim is inconclusive. (*See* **Bioavailability of Nutrients**.)

Sugar and Artificial Sweeteners

Reflected in the calorific content, yogurts with a range of sugar contents are now available. In plain yogurt, it is lactose, naturally present in yogurt, that provides the sweetness. But in other forms, sugar is added either as sucrose or as fructose and glucose from fruit. In 'diet' or 'very-low-fat' yogurt, artificial sweeteners are often used to reduce the calorie content. (*See* **Carbohydrates**: Sensory Properties; **Sweeteners**: Intensive.)

Trends in Yogurt Consumption

Yogurt and other cultured milk products have traditionally been popular dietary items in Eastern Europe. Popularity in the UK has increased considerably in the past 25 years: consumption

Table 1 Nutritional value of varieties of yogurt

	Low-fat plain (per 100 g)	*Low-fat fruit (per 100 g)*	*Whole-milk fruit (per 100 g)*	*Low-calorie fruit (per 100 g)*
Energy (kcal)	56	90	105	41
(kJ)	236	382	441	177
Protein (g)	5.1	4.1	5.1	4.3
Carbohydrate	7.3	17.1	15.4	5.8
Sugars (g)	7.3	17.1	15.4	5.8
Fat (g)	0.8	0.7	2.8	0.2
Saturates (g)	0.5	0.4	1.5	0.1
Monosaturates (g)	0.2	0.2	0.8	0.1
Polyunsaturates (g)	Trace	Trace	0.2	Trace
Sodium (mg)	83	64	82	73
Dietary fiber (g)	Nil	0.5[a]	0.5[a]	0.5[a]
Vitamin A (μg)	9	11	42	Trace
Thiamin (mg)	0.05	0.05	0.06	0.04
Riboflavin (mg)	0.25	0.21	0.30	0.29
Nicotinic acid (mg)	0.15	0.14	0.13	0.13
Potential nicotinic acid from tryptophan (mg)	1.20	0.96	1.29	1.00
Vitamin B_6 (mg)	0.09	0.08	0.07	0.07
Folic acid (μg)	17	16	10	8
Vitamin B_{12} (μg)	0.2	0.2	0.1	(0.2)
Pantothenic acid (mg)	0.45	0.33	0.30	N/A
Biotin (μg)	2.9	2.3	2.0	N/A
Vitamin C (mg)	1.0	1.0	1.0	1.0
Vitamin D (μg)	0.01	(0.01)	(0.04)	Trace
Vitamin E (mg)	0.01	(0.01)	(0.05)	0.03
Vitamin K (mg)	na	na	na	na
Calcium (mg)	190	150	160	130
Chlorine (mg)	150	130	150	120
Copper (mg)	Trace	Trace	Trace	Trace
Iodine (μg)	63	48	(48)	na
Iron (mg)	0.1	0.1	Trace	0.1
Magnesium (mg)	19	15	16	13
Phosphorus (mg)	160	120	130	110
Potassium (mg)	250	210	210	180
Selenium (μg)	1	(1)	(1)	(1)
Zinc (mg)	0.6	0.5	0.5	0.4

[a]This figure will vary depending on the type of fruit used.
na, data not available.
Source: Holland B, Unwin ID and Buss DH (1989) *Milk Products and Eggs*. Fourth supplement to McCane and Widdowson's *The Composition of Foods*, 4th edn. London: Royal Society of Chemistry and Ministry of Agriculture, Fisheries and Food, with permission.

increased 10-fold from 0.4 kg per person per year in 1966 to 4.1 kg per person per year in 1989, according to the *International Dairy Federation (IDF) Annual Statistics*. By 1996, the per capita consumption reached 6.18 kg. Much of this increase has been due to the introduction in the UK of flavoured and fruit yogurts.

The UK's National Food Survey for 1998 indicated that yogurt consumption is 109 ml per person per week. Comparable consumption levels for fromage frais and other dairy desserts are 16 and 27 ml, respectively. In this data set, yogurt is responsible for 72% of the market for these three types of products.

Table 2 shows the consumption of yogurt, fromage frais, and, for comparison, dairy desserts, in a selection of countries. Again, yogurt represents about 72% of the market.

Importance of Yogurt for Different Population Groups

Yogurt and Health

Apart from its contribution to nutrient needs, the perception of yogurt as a 'healthy' food has been augmented by claims of health benefits attributed to specific live bacteria present in some yogurts, in particular *Lactobacillus acidophilus* and bifidobacteria. Both of these types of bacteria are to be found in the human gastrointestinal tract, especially in breast-fed infants, and it has been suggested that these microorganisms may be able to colonize the gut when consumed in yogurt, and protect against pathogens. It has also been speculated that they may be of benefit in a number of intestinal disorders, including

Table 2 Yogurt, fromage frais, and dairy desserts in 1998

	Yogurt and dairy desserts market			*Segmentation (percent volume)*		
	Thousand tonnes	*Percentage change (1992–1998)*	*Per capita (kg)*	*Yogurt*	*Fromage frais*	*Dairy desserts*
Belgium	148.1	+16.7	14.52	50.6	26.1	23.4
France	1643.3	+10.6	27.91	56.6	20.7	22.7
Germany	1703.8	+28.2	20.75	74.4	7.6	18.0
Italy	306.2	+23.7	5.33	84.7	3.7	11.7
Netherlands	573.5	−1.1	36.33	56.6	5.5	37.9
Spain	501.8	+13.9	12.73	69.4	4.8.	25.8
UK	500.8	+45.9	8.53	72.5	10.8	16.8
Total	5377.5	+18.1	16.67	66.4	11.7	22.0

Source: ERC Group Plc (1999) *Yogurt and Dairy Desserts – Mainstream Europe*. Newmarket, UK: ERC.

those precipitated by antibiotic treatment or by diseases such as cancer and liver or kidney disease. Claims have been made that specific bacteria used in the production of a certain brand of yogurt (no longer on the market) have the potential to reduce blood levels of low-density lipoproteins (LDL). On the basis of existing research it is not possible to substantiate these various claims, although evidence is increasing, and inconsistencies in the findings may in part be explained by differences in strain and species of bacterial cultures, and differences in experimental design.

There is, however, a substantial body of evidence to indicate that fermented dairy products such as yogurt are well tolerated by individuals who are lactose-intolerant. It has been suggested that this is because of the bacterial enzyme, β-galactosidase (produced by the culture) in 'live' yogurt. This enzyme, which is able to digest lactose to glucose and galactose, is intracellular and hence is thought to survive gastric digestion. However, as lactose maldigesters tolerate yogurts with varying β-galactosidase activities equally well, it would seem that other factors may also be important, including the rate of gastrointestinal transit of yogurt.

Intake of the types of lactic acid-producing organisms found in yogurt has also been postulated to prevent or inhibit intestinal growth of a variety of foodborne, disease-causing organisms. Most of the evidence supporting a role for cultured dairy products, or the bacteria used to make them, in controlling intestinal pathogens comes from experimental animal and *in vitro* studies. Findings are inconsistent, but various mechanisms have been proposed to explain reported protective effects. For example, the ability of lactic acid cultures to lower intestinal pH favors growth of lactic acid bacteria but provides a hostile environment for pathogens. It has also been suggested that lactic acid bacteria may produce bacteriocins, peptides with a direct antibiotic effect.

In summary, while a fairly clear case has been made for tolerance of yogurt by lactose maldigesters unable to tolerate milk, the potential benefit of yogurt (or specific types of yogurt) in protecting against pathogens, in recolonizing the gut after illness, or in lowering LDL cholesterol concentrations in the blood needs further investigation.

See also: **Bioavailability of Nutrients**; **Buffalo**: Milk; **Carbohydrates**: Sensory Properties; **Folic Acid**: Properties and Determination; **Lactic Acid Bacteria**; **Milk**: Dietary Importance; **Pantothenic Acid**: Properties and Determination; **Sheep**: Milk; **Sweeteners**: Intensive; **Thiamin**: Properties and Determination

Further Reading

Buttriss J (1997) Nutritional properties of fermented milk products. *International Journal of Dairy Technology* 50(1): 21–27.

Chandan RC (1989) *Yogurt: Nutritional and Health Properties*. VA: National Yoghurt Association.

ERC Group plc (1999) *Yoghurt and Dairy Desserts: Mainstream Europe*. Newmarket, UK: ERC Group:

Hamilton-Miller JMT and Fuller R (1996) Probiotics – panacea or nostrum? *BNF Nutrition Bulletin* 21: 199–208.

Holland B, Unwin ID and Buss DH (1989) *Milk Products and Eggs*. Fourth supplement to McCance and Widdowson's *The Composition of Foods*, 4th edn. London: Royal Society of Chemistry and Ministry of Agriculture, Fisheries and Food.

Ratcliffe B and McMillan S (1999) Gut microflora. *BNF Nutrition Bulletin* 24: 82–91.

Renner E (1989) *Micronutrients in Milk and Milk-based Food Products*. London: Elsevier Applied Science.

Robinson RK (ed.) (1991) *Therapeutic Properties of Fermented Milks*. London: Elsevier Applied Science.

Tamime AY and Robinson AY (1985) *Yogurt: Science and Technology*. Oxford: Pergamon Press.

The Dairy Council (2000) *Milk Matters: a guide to the UK dairy industry and its products*. London: The Dairy Industry.

Z

ZINC

Contents

Properties and Determination

M Ruz, University of Chile, Santiago, Chile

Importance of Zinc in Human Biology

Zinc is the 23rd most abundant element on the Earth's crust. Among the transition elements, zinc is the second in abundance in eukaryotic cells. In mammals, zinc is found, on average, at a concentration of 0.02 mM in the extracellular space and 1 mM in the intracellular space. The total amount of zinc in the body of an adult ranges from 1.5 g (female) to 2.5 g (male). The distribution of zinc in the organism of a man is detailed in Table 1.

The essentiality of zinc for growth for lower forms of life was discovered in 1869 by Raulin studying the effects of zinc on *Aspergillus niger*. Todd documented the essentiality of zinc for rats in 1934. Today, the requirement of zinc for the growth and replication of all cells is recognized. Zinc is required for virtually all aspects of cell metabolism, i.e., DNA synthesis, DNA transcription, translation of mRNA into proteins, and structure and stabilization of translated proteins. Thus, at a whole organism level, zinc is necessary not only for growth but also for development and highly relevant functions, such as immunity, tissue repair, vitamin A metabolism, protection against oxidative damage, neuropsychological functions, apoptosis, cellular signaling, and hormone action, among others. (*See* **Zinc**: Physiology.)

Prior to 1963, zinc deficiency in humans was unrecognized. Today, as a result of remarkable advances in the area of zinc metabolism, we know that mild zinc deficiency is a widespread condition, probably as prevalent as iron deficiency. This has led to the suggestion that zinc should be added to iron, iodine, and vitamin A, to the group of the micronutrients with the highest risk of suboptimal nutrition in the world.

Severe human zinc deficiency is now a rare condition. It has been reported mainly in patients with acrodermatitis enteropathica and in patients receiving total parenteral nutrition without added zinc. Clinical manifestations include growth retardation, delayed sexual maturation, alopecia, epithelial lesions, immune deficiencies, hypogeusia, hyposmia, delayed wound healing, altered appetite, and night blindness. Mild zinc deficiency is much more difficult to characterize, and much information has been obtained from zinc supplementation studies in populations around the world. Growth retardation has been the most extensively studied feature, and immunological alterations, neuropsychological manifestations, and increased morbidity have also been reported. (*See* **Zinc**: Deficiency.)

In order to understand why zinc is crucial for life and the importance of achieving a normal nutritional status regarding this mineral, it is necessary to analyze the biochemical characteristics of this ion.

General Properties

Zinc is a IIB element with a complete d subshell and two additional s electrons. It is a small ion, with a radius of 0.65 Å and an oxidation state of +2. It has five stable isotopes and several radioactive isotopes with variable half-lives, some of which are shown in Table 2.

Zinc is a ubiquitous element and is mainly found in intracellular space (extracellular/intracellular ratio 1:50). The concentration of free zinc ions varies from

Table 1 Distribution of zinc in the human body

	Total Zn (g)	*Percentage*
Skeletal muscle	1.53	57
Bone	0.77	29
Skin	0.16	6
Liver	0.13	5
Brain	0.04	1.5
Kidneys	0.02	0.7
Heart	0.01	0.4
Hair	< 0.01	0.1 (approx)
Blood plasma	< 0.01	0.1 (approx)

Modified from Jackson MJ (1989) Physiology of zinc: general aspects. In: Mills CF (ed.) *Zinc in Human Biology*. pp. 1–14. London: Springer-Verlag.

Table 2 Stable and radioactive isotopes of zinc

Stable isotopes		*Radioactive isotopes*	
Isotope	*Natural abundance (%)*	*Isotope*	*Half-life*
^{64}Zn	48.89	^{62}Zn	9.22 h
^{66}Zn	27.91	^{63}Zn	38.5 min
^{67}Zn	4.11	^{65}Zn	243.8 days
^{68}Zn	18.57	^{69m}Zn	13.8 h
^{70}Zn	0.62	^{71m}Zn	3.97 h
		^{72}Zn	46.5 h

$\leq 10^{-9}$ M in the cytoplasm of many cells to $\geq 10^{-3}$ M in vesicles. Zinc has a highly concentrated charge and a small size, and a modest binding capacity to anions such as carboxylate and phosphate. Like others ions, zinc has a high affinity for electrons. It is a strong Lewis acid and can bind strongly to donors such as thiolates and amines.

Zinc is a flexible atom that is able to adopt four, five, and six coordinated geometries. It has a distinctive ability to form stable complexes with several molecules. Zn readily complexes amino acids, peptides, proteins, and nucleotides, and has an affinity for thiol and hydroxy groups and for ligands containing electron–rich nitrogen as a donor. For instance, zinc is often found in internal cross-links of intracellular proteins, where it apparently functions in a similar way to the disulfide bridges of extracellular proteins. Disulfide bridges are unstable to reduction in the intracellular environment. Zinc does not show any changes in its oxidation state, and this is an important advantage in reactions in which redox activities may represent a risk for oxidative damage.

Table 3 lists the important properties of zinc in a biological context.

Link Between Properties and Major Roles

The exceptional ability of the zinc atom to participate in strong but readily exchangeable ligand binding, the flexibility of its coordination geometry and its stable oxidation state are extremely useful in biological systems. In other words, zinc is important for its ubiquity and versatility. These characteristics mostly explain the importance of zinc involvement in biological functions.

Table 3 Importance of zinc in biology

1. High availability
2. Strongly retained
3. Fast ligand exchange
4. No redox reaction
5. Flexible coordination geometry
6. Good electron acceptor or Lewis acid
7. Supplies hard base, hydroxide

Modified from Williams RJP (1989) An introduction to the biochemistry of zinc. In: Mills CF (ed.) *Zinc in Human Biology*, pp. 15–31. London: Springer-Verlag.

The biological functions of zinc can be classified into three main categories: catalytic, structural, and regulatory. Although there are good examples for each of these, this should be considered only as a rough classification, because there is often some degree of overlap.

Catalytic

The catalytic role of zinc is related to the requirement of zinc by more than 200 enzymes. Zinc metalloenzymes are found in all six enzyme classes: oxidoreductases, transferases, hydrolases, lyases, isomerases, and ligases. Zinc metalloenzymes are involved in nucleic acid, protein, carbohydrate, lipid, and vitamin metabolism.

A combination of properties of zinc accounts for its utility in zinc metalloenzymes. Zn as a Lewis acid is able to accept a pair of electrons, and it does not participate directly in redox reactions. This is highly valuable in reactions such as proteolysis and the hydration of carbon dioxide. As result of having filled the d-shell orbitals, no specific geometry is more stable than any other, although the binding geometry observed most frequently is tetrahedral, with the metal ion coordinating three or four protein side-chains. This flexibility is widely used by zinc metalloenzymes during catalysis. Another important characteristic is its rapid ligand exchange.

Zinc present in enzymes plays distinct roles, from catalytic or cocatalytic to structural. It is not uncommon to observe mixed functions.

Catalytic Zinc is located in the active site interacting directly with the molecule involved in the reaction. Zinc has a highly localized charge and electron affinity, so it is a very effective attacking element. The

most frequently observed ligand in the active site of zinc metalloenzymes is histidine, followed by glutamic acid, aspartic acid, and cysteine. These ligands account for three or four ligands that can be incorporated with at least one water molecule. The zinc-bound water is a critical component for the catalytic site. It can be ionized to zinc bound hydroxide, polarized by a general base to generate a nucleophile for catalysis, or displaced by the substrate. Thus, a given substrate binds directly to Zn and may displace the metal-bound H_2O molecule. Zinc in this case would act as a Lewis acid polarizing the bound substrate facilitating the nucleophilic attack. This mechanism is observed in aldolases and peptidases. Alternatively, zinc may not bind to the substrate but mediate its function through the metal-bound water molecule. In this case, zinc lowers the pK_a of the bound water molecule. The resultant metal-bound hydroxide ion is the direct agent attacking the substrate. Examples of enzymes in which zinc plays catalytic roles are carbonic anhydrase (EC 4.2.1.1) and carboxypeptidase A (EC 3.4.17.1).

Because of the mechanisms of action and the characteristics of the substrates of zinc metalloenzymes, these enzymes are used in nature preferably for acting on small substrates that cannot be handled by organic groups in a multipoint attachment or when a chemically less specific but very rapid attack is needed.

Cocatalytic This term refers to the participation of more than one metal ion in the catalysis, i.e., the situation in which there are several metal ions bound in proximity to each other, and although all ions play crucial roles in the catalysis, only one is involved directly in the attack. Examples of this group are alkaline phosphatase, which has two zinc atoms and one magnesium atom per mole (EC 3.1.3.1), phospholipase C, which has three zinc atoms per mole (EC 3.1.4.3), and leucine aminopeptidase, which contains two Zn ions per mole (EC 3.4.11.1).

Structural In its structural role, zinc is coordinated with four amino acid chains, usually in a tetrahedral geometry. Cysteine is the most frequently observed ligand, followed by histidine and aspartate. In this situation, zinc atoms contribute to the stabilization of the tertiary or quaternary structure of the protein but do not participate in the catalysis. One such example is the enzyme aspartate carbamyl transferase (EC 2.1.3.2), which has six zinc atoms per mole. These atoms are found in the R subunits tetrahedrally bound by four cysteine thiolates, stabilizing the quaternary structure of the protein.

Structural

The structural role of zinc is related to the participation of zinc in specific cellular or subcellular components. It is an important component in proteins involved in gene regulation and as part of cellular membranes.

Zinc and gene expression Important cellular processes such as transcription, RNA processing, and replication are regulated by protein–DNA and protein–protein interactions. A significant number of such regulatory proteins present domains (or motifs) that are involved in binding to the DNA molecule. A major breakthrough was the identification of such motifs as zinc-containing structures generically termed 'zinc fingers.'

Zinc fingers are usually formed by zinc and cysteine and histidine residues. Tetrahedral coordination of zinc to thio-sulfur of cysteine residues and imidazole–nitrogen of histidine stabilizes the segment in a structure that resembles a finger. In this structure, the zinc atom does not interact directly with the DNA: it acts as a type of scaffolding that confers the required conformation to the overall structure, enabling the DNA-binding domain to bind DNA. Site-specific binding of Zn finger domains to DNA is one of the mechanisms by which proteins modulate gene expression. Zinc finger proteins contain one or more zinc finger domains and include transcription factors, hormone receptors, DNA-repair enzymes, products of oncogenesis, and tumor suppressor genes. Genes encoding zinc finger proteins could amount to as much as 1% of the human genome.

The first transcription factor to be identified as a zinc finger-containing protein was TFIIIA in 1985. TFIIIA is required for the correct initiation of transcription of *Xenopus laevis* 5S RNA genes by RNA polymerase III. The structure of TFIIIA encompasses zinc bound to two cysteine and two histidine residues. This is referred as the classic zinc finger DNA-binding motif.

The steroid (glucocorticoid and estrogen) receptors are members of the multigene family, which also includes receptors for thyroid hormone, retinoic acid, and vitamin D. The steroid hormone receptor has two zinc-finger loops. Steroid hormone receptors are ligand-inducible transcription enhancers that bind specifically to short DNA sequences or hormone response elements and regulate a number of genes. The steroid hormones induce gene expression by binding to receptor proteins in the cytoplasm, causing these proteins to enter the nucleus, bind to DNA, and activate transcription of certain genes.

Other zinc finger-containing proteins are: transcription factor Sp1, NGFI-A/EGR/Krox 24/Zif 268

family, GATA-1, GAL4, the so-called RING-finger proteins (*R*eally *I*nteresting *N*ew *G*ene), and p53. p53 zinc-binding proteins seem to play important roles in the development/suppression of cancer. p53 seems to be involved in the induction of cell-cycle arrest upon identification of increased genomic DNA damage. About 50% of human cancers may involve a p53 dysfunction.

Zinc and membranes Zinc is a structural component of cellular and subcellular membranes. Although its specific roles are only partially known, a number of studies have demonstrated alterations when zinc is reduced in these structures. These include changes in protein structure and lipid concentration, increased red cell fragility and abnormal platelet aggregation, and reduced activities of selected enzymes, such as Ca-ATPase, 5′ nucleotidase and alkaline phosphatase.

Under physiological conditions the major role of zinc is not to be a membrane stabilizer, as previously considered, but a permissive factor for cell signaling. Zn acts by binding to, and maintaining, the tertiary or quaternary structure of several key proteins involved in signal transduction. A crucial feature in zinc deficiency is a defect in calcium channels function, specifically a reduced calcium uptake as result of an abnormal sulfhydril redox state in the membrane channel protein.

Regulatory

Zinc is involved in regulatory functions. One remarkable example is the regulation of the protein metallothionein (MT). MT is a small metal-binding protein characterized by a high cysteine content and a high affinity for heavy metals. There are four isoforms of MT. The functional significance of MTs is not well understood. It is a rather nonspecific metal-buffering ligand to retain or deliver metal ions. Among the proposed roles for MT are: metal donation to target apometallothioneins such as zinc fingers proteins and enzymes, metal detoxification, and protection against oxidative damage. It can also act as a short-term storage reservoir for the metals.

The transcription of MT genes can be activated, among others, by a number of metals, including zinc, copper, and cadmium. Zn is the most important in eukaryotic cells. The regulation of MT gene expression by zinc and other metals occurs via the interaction of the metals with multiple *cis*-acting DNA elements called metal response elements located in the MT gene promoters.

Another regulatory function of zinc is that related to the modulation of the gene-directed cell death process, or apoptosis. A number of studies have demonstrated the importance of maintaining an intracellular pool of zinc atoms to regulate this process. *In vitro* studies have shown that apoptosis is induced when cells are cultured in a zinc-free medium or when zinc is depleted by membrane-permeant zinc chelators. The exact mechanism by which Zn ions intervenes in preventing or reducing apoptosis is not completely understood. Zinc may interact with one or more participants of the apoptosis process. Thus, initial studies have proposed that zinc inhibits the activity of the Ca–Mg-dependent endonuclease, whereas others have proposed that zinc might block the mechanism by which the inactive procaspase-3 is activated. Zinc is a selective inhibitor of caspase 6, which promotes activation of caspase-3. It has also been proposed that zinc may participate through its antioxidant roles.

A series of recent reports have documented the participation of zinc in a variety of signal-related functions. Thus, β-2 adrenergic receptors are allosterically modulated by zinc. Zinc has been shown to increase the excitability of dopaminergic neurons in rat substantia nigra, to modulate signals from red- and short wavelength-sensitive cones to horizontal cells in carp retina, and to induce the stimulation of the c-Jun N-terminal kinase signaling pathway through phosphoinositide 3-kinase. Some observations in cultured cells have suggested the existence of a zinc-sensing receptor that triggers the release of intracellular Ca^{2+} and regulates ion transport.

Zinc Determination

Compared with other essential or toxic trace elements, such as selenium or arsenic, zinc does not present any major problems for its determination in a wide variety of matrices, and consequently, a large number of methods are used for its quantification. A detailed discussion of the basis, advantages, and limitations of each methodology is beyond the scope of this chapter, but there is abundant literature elsewhere on specific issues relating to zinc.

Principal Methods to Determine Zinc in Biological and Environmental Samples

The methods used to determine zinc are based on several physical, chemical, and nuclear characteristics of the atoms, including zinc. Some of the methods used to determine zinc are listed in Table 4.

Atomic absorption and atomic emission spectrometry are based on the property of the atoms to absorb or emit light energy during energy transition from ground states to excited states or during decay. By far the most commonly used method to determine zinc is atomic absorption spectrometry (AAS). For zinc, the analysis is usually conducted using a

Table 4 Methods to determine zinc in biological and environmental samples

Flame atomic absorption spectrometry
Graphite furnace atomic absorption spectrometry
Atomic emission spectrometry
Inductively coupled plasma-atomic emission spectrometry
Differential pulse anodic stripping voltammetry
Energy-dispersive X-ray fluorescence
Total reflection X-ray fluorescence
Proton-induced X-ray emission
Photon activation analysis
Instrumental neutron activation analysis
Inductively coupled plasma mass spectrometry

wavelength of 213.9 nm. This method provides accurate and precise results, and the cost of the equipment and maintenance are low compared with most of the other techniques. The detection limit in its flame mode is about 100 $\mu g\,l^{-1}$ (or 100 p.p.b.). This detection limit is greatly improved, more than 200 times, when used with the graphite furnace. The significant improvement in sensitivity is also accompanied by a higher degree of interferences, mainly matrix effects. Similarly, the sensitivity of atomic emission spectrometry is improved when inductively coupled plasma is used. Unlike flame atomic absorption spectrometry (FAAS) or graphite furnace atomic absorption spectrometry, inductively coupled plasma-atomic emission spectrometry provides the capacity for performing multielemental analyses. Potentiometric methods, such as differential pulse anodic stripping voltammetry have shown analytical performances comparable with FAAS in simple matrices. The main advantage of potentiometric methods is the simplicity of the equipment and lower costs. It has been suggested as a candidate for routine clinical chemistry.

The group of nuclear analytical methods suitable for zinc determination includes a number of methods, such as: energy-dispersive X-ray fluorescence, total reflection X-ray fluorescence, proton-induced X-ray emission (PIXE), photon activation analysis, and instrumental neutron activation analysis (INAA), among others. These methods have detection limits in the range of p.p.b. to p.p.t. Some of them are based on the excitation of characteristic X-rays in a sample by electromagnetic radiation from an X-ray tube or radioactive source (X-ray fluorescence methods) or by accelerated charged particles (PIXE). Photon-activated analysis is based on photon emission after irradiation. INAA is a multielemental method that has been widely used for zinc and other trace element analysis. It is based on the measurement of the γ radiation produced after irradiation. For zinc, the reaction is ^{64}Zn (n, γ) ^{65}Zn. Gamma radiation from the long-lived nuclide ^{65}Zn is usually measured 15 days after irradiation.

Inductively coupled plasma mass spectrometry (ICP-MS) has shown remarkable advances in sensitivity and reliability for mineral analysis. The regular equipment is an ICP-quadrupole mass spectrometer, which has a detection limit in the range of p.p.t. with concomitant high precision and accuracy. This is a versatile technique that can be coupled to chromatography systems for other applications such as high-performance liquid chromatography–ICP-MS. New versions of high-performance ICP-MS include ICP magnetic sector-MS and ICP ion trap-MS. These are intended to reach detection limits in the low p.p.t.–p.p.q. range. ICP-MS can determine not only the total element concentration but also its individual stable isotopes.

The use of stable isotopes has proven to be a powerful tool in the field of zinc nutrition and zinc metabolism. Because they do not emit radiation, they can be safely used as tracers in any age/physiological condition group. This requires techniques with the capacity for performing highly accurate and precise determination of isotope ratios. Among the methods suitable for stable isotope determinations are: neutron activation analysis, electron ionization-mass spectrometry, gas chromatography-mass spectrometry, fast atom bombardment-mass spectrometry, thermal ionization mass spectrometry (TIMS), and ICP-MS. Of these, TIMS and ICP-MS are recognized as the techniques with the best results in terms of accuracy, precision, and sensitivity.

Regardless of the method used to determine zinc, an appropriate quality assessment of the procedures should be conducted, considering both precision and bias evaluation. The former is conducted through repetitive measurements and internal test samples, and the latter is accomplished through a series of actions such as the use of different operators or equipments, use of reference materials, participation in intercomparison rounds, external audits, etc.

The determination of zinc in biological samples can serve several purposes. One of the most commonly used applications has been the assessment of zinc status. Thus, zinc has been determined in: plasma, serum, hair, skin, semen, sweat, saliva, urine, neutrophils, lymphocytes, platelets, erythrocytes, and erythrocyte membranes. Although the determinations of zinc in these fluids or tissues have been conducted appropriately from an analytical point of view, they have shown limitations as indicators of zinc nutriture. These limitations also extend to other methods such as the activity of several zinc enzymes, including alkaline phosphatase (EC 3.1.3.1), α-D-mannosidase (EC 3.2.1.24),

5′ nucleotidase (EC 3.1.3.5), angiotensin-1-converting enzyme (EC 3.4.15.1), and several functional tests. (*See* **Nutritional Assessment:** Biochemical Tests for Vitamins and Minerals.)

See also: **Nutritional Assessment**: Biochemical Tests for Vitamins and Minerals; **Zinc**: Physiology; Deficiency

Further Reading

Berthon G (ed.) (1995) *Handbook of Metal–Ligand Interactions in Biological Fluids*, vol. 1. New York: Marcel Dekker.

Davis SR and Cousins RJ (2000) Metallothionein expression in animals: a physiological perspective on function. *Journal of Nutrition* 130: 1085–1088.

de Goeij JJM (1994) Nuclear analytical methods in the life sciences. *Biological Trace Element Research* 43–45: 9–17.

Green A, Parker M, Conte D and Sarkar B (1998) Zinc finger proteins: a bridge between transition metals and gene regulation. *Journal of Trace Elements in Experimental Medicine* 11: 103–118.

Locatelli C and Torsi G (2001) Heavy metal determination in aquatic species for food purposes. *Annali di Chimica* 91: 65–72.

McCall KA, Huang Ch and Fierke CA (2000) Function and mechanism of zinc metalloenzymes. *Journal of Nutrition* 130: 1437S–1446S.

Mellon FA and Sandstrom B (1996) *Stable Isotopes in Human Nutrition*. London: Academic Press.

Mills CF (ed.) (1989) *Zinc in Human Biology*. London: Springer-Verlag.

O'Dell BL (2000) Role of zinc in plasma membrane function. *Journal of Nutrition* 130: 1432S–1436S.

Yergey AL (1996) Analytical instruments for stable isotopic tracers in mineral metabolism. *Journal of Nutrition* 126: 355S–361S.

Physiology

N W Solomons, CeSSIAM, Guatemala City, Guatemala

Background

It was first established in 1869 by Raulin that zinc was essential for the growth of bread mold. In the twentieth century, it was discovered that zinc was essential for the growth of rats and to prevent the parakeratitis dermatological condition of swine. The essentiality of zinc, and its requirements, has been established in species from the protozoa, *Euglena gracilis*, to the steelhead trout. Most recently, the critical role that nutrition plays in the nutrition of the retrovirus of AIDS has been revealed. This article considers three topics: the normal metabolism of zinc by the body; aberrations of its metabolism; the physiological and biochemical functions of zinc for the organism. Emphasis will be placed on zinc in human nutrition.

Normal Metabolism of Zinc

Absorption of Zinc

Mechanisms of normal absorption Zinc is not absorbed in the esophagus or stomach but can be taken up along all parts of the small intestine (duodenum, jejunum, and ileum) with varying efficiencies and, to some extent, in the large bowel. The form of zinc in the diet, soluble in beverages or bound in food matrices, may determine the part of the alimentary tract in which the zinc is absorbed. In healthy humans, the fractional absorption of isotopically labeled oral zinc ranges from 40 to 70% from aqueous solutions and 2 to 40% in the presence of various foods and meals.

The cellular mechanisms and kinetics of zinc absorption have been studied extensively. It is likely that zinc, freed from food, is complexed to some small molecules as it approaches the brush border. Uptake across the mucosal membrane can be by a simple diffusion (nonsaturable) mechanism of low permeability and dependent upon the concentration gradient between lumen and blood, taking a paracellular pathway after membrane transport or by a carrier-mediated and energy-dependent (saturable) mechanism, which is most important at low intraluminal concentrations.

Once within the cell, an element of zinc has a multitude of possible designations. It can become part of the zinc nutrition of the enterocyte, itself; it can remain in the cell, eventually being bound to metallothionein (MT), a 6000–7000-Da peptide, with 61 amino acid residues, and then lost back into the lumen of the intestine when the cell is desquamated; or it can pass on to the serosal surface of the cell, possibly bound to intracellular ligands or binding proteins to be released into the body. The level of preexisting MT in the cell has been thought to regulate the amount of zinc available for exit into the bloodstream, but microvesicles may participate in the transport of the zinc that is eventually taken up by the circulation. The final exit step of zinc into the body may be energy-dependent, coupled to basal membrane adenosine triphosphate.

Zinc absorption appears to be homeostatically regulated, in so far as uptake is reduced in sufficiency

or excess, whereas in deficiency, the uptake is enhanced. In zinc-depleted subjects, more than 90% of a highly available dose can be absorbed. The amount that passes through the enterocyte by diffusion is related to the mucosal–serosal gradient and is unrelated to the zinc status of the host. It is the carrier-mediated active absorption that is responsive to underlying zinc nutrition. In animal experiments and human experience, zinc absorption from the diet has been shown to increase during both pregnancy and lactation.

The more zinc there is available to the intestine in a meal, the lower is its fractional absorption. However, the amount of zinc that reenters the intestinal lumen is now recognized to be a dynamic function of the quantity of the mineral absorbed from the diet. (*See* **Lactation**: Physiology; Human Milk: Composition and Nutritional Value.)

Bioavailability of zinc from the diet Bioavailability strictly refers to both the uptake and metabolic utilization of a nutrient. Since different valence states or chemical species are not an issue with zinc, the biological availability of zinc relates exclusively to the efficiency of its absorption. (*See* **Bioavailability of Nutrients**.)

In theory, zinc must be either in soluble form or bound to soluble, absorbable species to be absorbed by the intestine. The efficiency of mastication and gastric and pancreatic protease digestion of food determines the extraction of zinc from its food matrices. Once liberated and dissolved, the zinc ion needs to remain in solution and proceed in its transit through or around the enterocytes.

Substances that influence intestinal permeability to zinc are also mediators of bioavailability. A series of dietary substances influence the bioavailability of zinc, either by chelating or binding zinc or by competitively inhibiting its absorption. Two common components of whole-grain cereals – phytic acid and dietary fibre – are the most important substances that interfere with zinc absorption. Polyphenolic substances (tannins) in tea and coffee and oxalic acid of green leafy herbs are also inhibitors. Calcium and dietary nonheme iron are two inorganic elements that interfere with zinc uptake by direct competition. (*See* **Dietary Fiber**: Effects of Fiber on Absorption; **Phytic Acid**: Nutritional Impact.)

Alternatively, a series of compounds have been suggested as *enhancers* or *promoters* of zinc absorption, including amino acids, sugars, picolinic acid, citric acid, and prostaglandins. Only the first two classes of compound are likely to be of physiological significance. Red wine has been shown to enhance zinc balance under metabolic conditions. Ascorbic acid and lactose have a *neutral* influence on zinc bioavailability.

It should be remembered that bioavailability refers to a fractional efficiency of uptake, and the *total amount* of zinc ingested is a codeterminant of how much net nutrient will be absorbed. If the zinc intake is high in a diet of poor zinc bioavailability, the amount absorbed may still be sufficient to support true nutritional requirements.

Enteropancreatic circulation With each meal, a large volume of bile, pancreatic juice, and salivary, gastric, and intestinal secretions are secreted into the intestinal lumen, with pancreatic juice being the richest in the metal. Within the lumen, the endogenous zinc forms a common pool with that from oral intake. Some of the pancreatic zinc is not recovered and passes into the stools, but a substantial part is reabsorbed and becomes part of internal metabolism once more. This constitutes an enteropancreatic circulation for endogenous zinc. By virtue of this 'double jeopardy' of zinc to dietary inhibitors, the body can be more rapidly depleted of this nutrient than of most others.

Net daily zinc metabolic balance In growing individuals, more zinc should be taken in than excreted *on average* on any given day. Adults with a normal total body content of zinc should be in essentially zero balance, with the amount absorbed equal to the amount lost by all routes (see below). In pregnancy and lactation, the products of conception and the daily secretion of milk represent additional depots for newly absorbed zinc. In 1973, an expert committee of the WHO first used a factorial approach to estimate the amount of zinc that must be absorbed daily to account for deposition for growth and losses from the body in infants, children, and adolescents, and to account for losses in adults. This amount of zinc is often referred to as the 'true zinc requirement,' to distinguish it from the daily dietary *intake* requirement, and the daily dietary *allowance*. Estimations of the amount of *dietary* zinc to be absorbed daily have been refined by subsequent consultancies, the most recent being from the World Health Organization and the International Atomic Energy Agency in 1996 and from the US Food and Nutrition Board in their *Dietary Reference Intakes* process in 2001. For infants aged 6–12 months, it is 0.8 mg; for children aged 6–10 years, 1.1 mg; for men over 18 years, 1.4 mg; and for women over 18 years, 1.0 mg. During pregnancy, absorption daily from zinc in the diet should be 1.5 mg, rising to 1.8 mg during the first 6 months of lactation, and declining to 1.4 mg during the duration of the period of breast-feeding. Of

course, the amount of zinc reabsorbed from the endogenous sources of the mineral that enters intestinal lumen as part of pancreatic, biliary, and intestinal secretions with meals is at least equivalent to the dietary requirement. So, a rough calculation of the daily net absorptive demand across the intestine would involve a doubling of the aforementioned age- and physiology-specific values. (*See* **Dietary Requirements of Adults**.)

Distribution and Transport of Zinc

Distribution of zinc The total body content of zinc in normal individuals has been measured by isotopic dilution studies and ranges from 1.5 to 2.5 g. It is higher in men than in women and decreases with increasing age in adults. At the tissue level, the content of zinc is heterogeneous, ranging from 20 to 200 $\mu g\,g^{-1}$ in most tissues, and up to 600–800 $\mu g\,dl^{-1}$ in pancreas, gonads, and retina. Within the cell, zinc is ubiquitous but compartmentalized, being found in nuclear, mitochondrial, and cytosolic fractions.

The two major depots for zinc in the body are skeletal muscle and the skeleton, which contains, respectively, 57 and 29% of the zinc in an individual. Skin and liver share an additional 10%. The percentage of the total body content of zinc in the 7-l volume of peripheral circulation is less than 0.1%.

The uptake of newly absorbed zinc is greatest in the liver. Structural tissues, such as bone and skin, and functional tissues, such as kidney, thymus, pancreas, and gonads, have high rates of uptake, which implies high rates of intracellular turnover.

Transport of zinc Zinc that passes through and around the enteral cell, from the intestinal lumen, enters the capillary circulation of the gut and returns by the mesenteric and portal circulation to the liver. Some of the zinc is taken up by hepatic cells, and the remainder passes into the systemic circulation.

Since zinc would be filtered and lost by the kidney were it not bound to large, nonfilterable molecules or to reabsorbed species, it circulates in bound states in the systemic circulation. Forty per cent of zinc is a covalently linked component of α_2-macroglobulin, and most of the remainder is loosely bound to serum albumin. A small fraction is chelated to serum free amino acids or small peptides. About 3–5 mg of newly absorbed zinc and additional quantities moving from one site to another pass through the circulation daily; the flux of the transport pool is rapid. Hormones (glucocorticoids, glucagon catecholamines) and monokines (interleukin-1, tumor necrosis factor) modify zinc uptake and turnover in peripheral tissues. Whether or not a peripheral cell zinc receptor exists for the capture of the nutrient from the systemic bloodstream has not been determined.

In reproductive biology, two additional aspects of zinc transport – transplacental and mammary gland transport – are important in normal metabolism. Zinc can move bidirectionally from the maternal circulation to the fetus, and from the fetus to the mother, but the net flux is toward the fetus. The flux is rapid, and the size of the transfer is proportional to the mass of the fetus, increasing with advancing gestation. Some 75% of the zinc in amniotic fluid is bound to albumin. The exact mechanism for transfer of zinc from plasma to mammary gland, and within the gland into the breast milk, is poorly understood. The concentration of zinc decreases progressively with duration of lactation. A multinational study of seven countries, conducted by the International Atomic Energy Agency, showed that zinc concentrations in breast milk vary from nation to nation, and within countries from rural to urban dwellers. However, in experimental studies, high oral doses of zinc did not modify the zinc content of healthy mothers.

Retention and Excretion of Zinc

Zinc storage and retention Despite the relatively large amount of zinc in the body, and its ubiquitous distribution, most is either in active functions in metabolism or in nonretrievable deposits in bone, muscle, and integument. There is a consensus that if any metabolic storage pool of zinc exists to act as a reserve against nutritional compromise, it is small in relation to the total body content of the mineral. It is now postulated that the mobile exchangeable zinc pool constitutes about 10% of the total body zinc, i.e., is 150–250 mg. Thus, a cumulative deficit of 100–200 mg, i.e., the obligatory losses of 20–40 days, could be critical.

In tissues with active uptake and turnover of zinc, such as liver, kidney, pancreas, and gut, the metal-binding protein MT is prominent, and its synthesis is regulated and responsive to zinc availability. MT expresses its high affinity for divalent cations and is related to the 20 sulfur-containing cysteine residues. It is believed to constitute a cytosolic storage pool or buffer to regulate concentrations of free ions within the cytoplasm. It has a rapid turnover of 20 h and is subject to regulation or influence by a number of hormones and pathological conditions.

Zinc excretion A number of routes are available for zinc excretion from the body. In the zinc-adequate state, the excretion is proportional to the total body burden of the nutrient. Some pathways for zinc excretion are not regulated. Among these are those

via the integumentary tissues, i.e., the growing out of hair, the desquamation of skin, and sweating.

The major route for excretion of the zinc of the active metabolic pool is into the fecal stream from pancreatic and intestinal secretions. Low zinc status produces conservation of zinc, whereas zinc excess leads to enhanced fecal excretion. Adults chronically consuming 10–15 mg of dietary zinc excrete about 500 μg of zinc in the urine daily. With severe zinc restriction, the renal excretion in 24 h is reduced to 200 μg. When zinc is perfused parenterally or consumed orally in pharmacological amounts, a manifold increase in renal excretion is observed.

Pathophysiological Metabolism of Zinc

Several conditions can produce a pathophysiological metabolism of zinc. Its malabsorption is produced in short-gut or pancreatic insufficiency. Diarrheal disease, protein-losing (exudative) enteropathy, and enterocutaneous fistulas produce a wasting of endogenous zinc into the stools. Excessive renal losses are seen in primary renal disease, in certain endocrinopathies such as diabetes mellitus, with infection and stress, and with severe exercise.

Infection and stress induce a cascade of hormonal responses including the secretion of glucocorticoid and catecholamine hormones and the peptides from circulating monocytes and fixed macrophages (cytokines), which produce an internal redistribution of zinc with redirection of the mineral to the liver and the bone marrow. A primary effect of the hormones and cytokines may be to induce the synthesis of MT, which captures and holds zinc in the cells. A teleological explanation of this filtering of zinc from blood to liver and bone marrow would be to facilitate the hepatic synthesis of acute-phase protein and the proliferation of new lymphocytes or neutrophils.

Physiological, Metabolic, and Biochemical Functions of Zinc

Physiology at the System Level

Below is a comprehensive listing of the physiological and metabolic functions of mammalian organisms, including humans, that are dependent on zinc. The inference of their dependence is based primarily on zinc-restriction experiments in laboratory animals and livestock, along with some zinc-deprivation experiments in humans, in which the performance is altered with decreasing zinc status.

- growth of cells;
- replication of cells;
- antioxidant (antifree radical) protection;
- regulation of neurotransmission;
- insulin and leptin metabolism;
- sexual maturation;
- spermatogenesis;
- oogenesis;
- cellular apoptosis;
- dark adaptation or scotopic vision;
- cellular immunity;
- humoral immunity;
- appetite regulation;
- taste acuity;
- olfactory acuity;
- cognitive function.

Biochemistry at the Cellular and Molecular Level

Zinc is a unique element among the traditional metals in so far as it exists exclusively in the divalent oxidation state and has no redox potential. Its coordination chemistry is alternatively tetrahedral (four ligands) or trigonal bipyramidal (five ligands) geometry, coordinated by the negative charges of a sulfur of cysteine, a nitrogen of histidine, or an oxygen from aspartate and glutamate in various combinations. Within the organized complexes, the protein at the center exerts an intense positive charge. Zinc in proteins can participate directly in catalytic reactions or maintain stability and structure conformation of proteins.

Zinc metalloenzymes A zinc metalloenzyme is an enzyme that has a specific requirement for zinc as a cofactor or as an integral, coordinated firmly complexed moiety within the structure of the protein. In the latter circumstances, the zinc ion(s) can participate at the active site of the enzyme in its catalytic activity in either the 4- or 5-coordination configurations in association with a molecule of water. The transformation of water by the dense charge and acid–base equilibrium provides the energy for the catalytic properties. Alternatively, complexed zinc can serve a role in stabilizing the conformation of the protein. A selected number of the estimated 300 enzymes in nature that have been confirmed as zinc metalloenzymes are listed below:

- aldolase (fructose 1, 6-biphosphatase);
- alcohol dehydrogenase;
- alkaline phosphatase;
- carbonic anhydrase;
- carboxypeptidase A;
- carboxypeptidase B;
- collagenase;
- δ-aminolaevulinate dehydratase;
- deoxynucleotidyl transferase;
- DNA polymerase;
- enkephalinase;

- glutamic acid dehydrogenase;
- lactic dehydrogenase;
- malic acid dehydrogenase;
- pyruvate carboxylase;
- nucleoside phosphorylase;
- reverse transcriptase;
- DNA polymerase I;
- RNA polymerases I, II, and III;
- thermolysin;
- zinc–copper superoxide dismutase.

Zinc metalloenzymes are not restricted to mammalian enzymes: some examples are found in plants, bacteria, and viruses. Some of the pathogenesis of zinc deficiency can be attributed to defects in the reactions catalyzed by zinc metalloenzymes. For example, the impaired dark adaptation and night blindness of zinc deficiency have been attributed to a defective functioning of the alcohol dehydrogenase (retinol dehydrogenase), found in the retina, that is responsible for the conversion of retinol to retinaldehyde for the regeneration of the visual pigment, rhodopsin, in the rods of the retina. (*See* **Coenzymes**.)

Other putative cellular and molecular roles of zinc The other generic manner in which zinc, possibly in a freely exchangeable, ionic form within the cell, is in the stabilization of membranes, of membrane–protein associations, and of macromolecular structures. The association of protein kinase C with cell membranes, for example, depends on zinc.

Zinc has long been known to stabilize polyribosomes, and this is one of the putative roles of the nutrient in protein synthesis. More recently, a role for zinc, tightly complexed to amino acids in chains of peptides within the chromatin proteins of the nucleus of the cell, has been shown for expression of genetic information. A series of 'zinc finger' proteins have been characterized. The name derives from the rigid projection of a loop of protein (a finger), stabilized by the zinc in tetrahedral coordination with four amino acids; the most common motif is one coordinated with two cysteine and two histidine residues. Over 2000 different domains of these zinc finger proteins for regulating the expression of specific genes along the chromosomes, allowing a determined amount of the genetic code to be translated to its complementary messenger RNA, have been identified in the last decade. Functions of enzymes that do not contain zinc themselves may be impaired in zinc deficiency because of a defect in a zinc finger protein that governs its translation in the nuclear reserves of genetic information.

A specific form of transcriptional regulation is known as the metal regulating element, which is involved in the regulation of MT synthesis presumably in response to the cellular levels of zinc that exist. Hence, the amount of zinc entering the cell influences an MT transcription factor associated with this MRE in the chromosome coding for MT and accounts for the up- or downregulation.

Two levels of zinc's participation in genetic expression, i.e., in zinc finger transcription regulators and in polyribosome, may also help to support the theory, proposed by Mills, that increased protein degradation is a generic consequence of zinc deficiency, one that explains the defects in growth and the rapidity of the onset of manifestations after restriction of dietary zinc.

Tissue cells are constantly dividing, but the net cell mass of an organ remains constant. Apoptosis is the process of genetically programmed death of cells by which certain cells within the tissue 'commit suicide' to maintain the balance of mass and structure. Each cell has this potential in a series of 'procapase' protease proenzymes that can become activated to produce a cascade effect leading to nuclear involution, cytosolic dissolution, and loss of attachment. Zinc is among the dietary constituents involved in this process. A decrease in intracellular zinc concentration precedes the physiological and anatomical cellular changes of apoptosis.

Finally, the molecular basis of zinc's role in immune protection has not been elucidated. Its roles in cell proliferation and thymic function are basic but combined with regulation of cytokine production, modulation of apoptosis, and a function of induced metallothionein as a scavenger of harmful free radicals in the cell cytoplasm.

See also: **Bioavailability of Nutrients**; **Coenzymes**; **Dietary Fiber**: Effects of Fiber on Absorption; **Dietary Requirements of Adults**; **Immunology of Food**; **Lactation**: Human Milk: Composition and Nutritional Value; Physiology; **Phytic Acid**: Nutritional Impact

Further Reading

Davis JS and Cousins RJ (2000) Metallothionein expression in animals: a physiological perspective on function. *Journal of Nutrition* 130: 1085–1088.

Dibley MJ (2001) Zinc. In: Bowman BA and Russell RM (eds) *Present Knowledge in Nutrition*, 8th edn, pp. 329–343. Washington, DC: ILSI Press.

Institute of Medicine, Food and Nutrition Board (2001) *Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc*. Washington, DC: National Academy Press.

Solomons NW (2001) Dietary sources of zinc and factors affecting its bioavailability. *Food and Nutrition Bulletin* 22: 138–154.

Solomons NW and Ruz M (1998) Trace element requirements in humans: An update. *Journal of Trace Elements in Experimental Medicine* 11(supplement): 177–195.

World Health Organization/International Atomic Energy Agency (1996) *Trace Elements in Human Nutrition and Health*. Geneva: WHO.

Deficiency

N W Solomons, CeSSIAM, Guatemala City, Guatemala

Background

Zinc is an essential nutrient for all living species. Human zinc deficiency was first described in patients as a secondary consequence of alcoholic cirrhosis in 1956. In 1961, human zinc deficiency was first described on a *dietary* basis in rural, Iranian children. Since then, clinical zinc deficiency has been described in all age groups from breast-fed infants to the very elderly. For both the clinical nutrition of hospitalized patients and the public health at the community level, human zinc deficiency has emerged as a relevant concern over the past three decades.

Mechanisms for the Pathogenesis of Zinc Deficiency

Only a very small 'exchangeable zinc pool,' sufficient for a few days' nutrition, exists in human body, and the major deposits of the metal in bone or muscle cannot be readily mobilized for nutritional purposes. Zinc in excess of short-term metabolic needs is either excluded from absorption or excreted by the kidneys or via the alimentary tract. The human organism lives with a perpetually marginal zinc nutriture by evolutionary design and is dependent upon regular dietary replenishment.

Certain extrinsic conditions and intrinsic situations can place an individual at increased risk of zinc depletion. Victor Herbert developed a conceptual framework for understanding the genesis of deficiencies of any nutrient based on six mechanisms that operate alone, or in combination, to contribute to the development of deficiency; these include: decreased intake, decreased absorption, increased destruction, increased excretion, decreased utilization, and increased requirements. With the exception of destruction, which is not relevant to inorganic nutrients, this scheme provides a focus for understanding the ways in which human zinc deficiency can occur.

With respect to zinc, primary deficiencies arise when the supply to the body, either by an oral/enteral (dietary) route or by parenteral nutrition, is insufficient to match losses and/or meet demands for growth. Selection of a zinc-poor diet, be it for cultural, socioeconomic or eating-disorder motives, can produce deficiency. When a patient is dependent on intravenous nutrition, improper formulation of the nutrient fluids can precipitate a primary deficiency. Impaired absorption of zinc results from digestive difficulties that impede its release from foods, interference with its mobility or solubility that inhibit its absorption, or intestinal membrane pathologies that impair its uptake. Increased zinc losses from the body can occur as a result of a hemorrhagia, superficial losses, excessive diuresis, or gastrointestinal tract wastage into stool. Decreased utilization can occur in inflammation, infection, and metabolic disorders. Thermal burns, catch-up growth, multiple pregnancies (twinning, triplets), lactation, and rampant malignancies are conditions that place additional demands on the supply of zinc. These factors can operate alone or in various combinations. In the case of an inflammatory bowel disease such as severe Crohn's enteritis, one might see the combination of: (1) reduced intake owing to anorexia; (2) impaired absorption owing to mucosal damage; (3) increased losses through bleeding and protein-losing enteropathy; and (4) impaired utilization owing to activation of the acute-phase response. Increased requirements for zinc might even occur in this disease if an adolescent patient develops a remission leading to rapid compensatory growth and sexual development after a period of growth arrest. (*See* **Bioavailability of Nutrients**.)

Once the body pool of zinc has contracted beyond a critical level, however, the exact molecular basis of the pathogenesis of zinc deficiency is poorly understood. Michael Golden has classified the deficiency of zinc as a type II nutritional deficiency, similar to deficiencies of the intracellular nutrients nitrogen, phosphorus, magnesium, and potassium, in which there is conservation of tissue concentrations at the expense of growth and development. This is especially true of zinc deficiency in experimental animals, in which restriction of the nutrient produces marked anorexia, immediate cessation of growth and reproductive capacity, and even weight loss; combined with regulated conservation of nutrient, it allows the ratio of body zinc content to body mass to remain relatively constant. Animals will not readily eat a zinc-deficient ration that is producing a deficiency state. Only with force-feeding will a diet of low zinc density produce tissue desaturation. In humans, however, social and cultural factors intercede to overcome whatever

anorexia might result from consumption of a zinc-poor diet or one that has a poor zinc bioavailability; thus, people – as opposed to animals – will consume generous amounts of energy and protein while on zinc-deficient diets. An opposing combination of conservation and desaturation of tissue zinc can therefore occur in humans. The fact that a certain degree of conservation of tissue zinc is maintained under conditions of zinc-poor diets does not eliminate the possibility of the intracellular and molecular roles of the nutrient being disrupted and altered by the responses to zinc deprivation.

How the deficiency manifestations, described below, are related to zinc deficiency at the tissue and cellular level is still poorly understood. Between 60 (conservative estimate) and 300 (generous estimate) enzymes, are believed to be zinc metalloenzymes or enzymes dependent on zinc as a cofactor in their catalytic activity. Specifically, the disruption of cell proliferation that leads to growth failure and to impairment of function of cells in the gut and bone marrow that are rapidly turning over can be attributed to the defects in enzymes responsible for DNA regeneration and RNA production. Furthermore, over 2000 nuclear transcription factors that employ zinc to regulate their protein configuration during gene activation (zinc finger proteins) have been identified in the past decade. Recently, *in vitro* experiments have confirmed that decreasing cellular zinc restricts the expression of a zinc-finger transcription factor for regulation of interleukin-2 in lymphocytes. Zinc deficiency promotes protein degradation and enhances protein turnover; this would explain both the rapid response to zinc deficiency by growing animals and the limitation of total growth. (*See* **Coenzymes**.)

The loss of stabilization of intracellular structures by zinc in a freely exchangeable intracellular form also accounts for some of the molecular defects in zinc deficiency. Zinc is thought to stabilize membranes, to stabilize ribosomes, and to contribute to associations between membranes and functional proteins. Soluble zinc plays a role in signaling among cells and between compartments within cells.

Manifestations of Zinc Deficiency

Clinical Signs and Symptoms

Deficiency in laboratory animals and livestock We have understood human zinc deficiency better using the basis in its observation and study in laboratory rodents, as well as poultry, swine, ruminants, felines, canines, and subhuman primates. The dietary regimens have various formats, including *ad libitum* feeding of a zinc-poor diet with pair-feeding of control animals as well as force-feeding of an energy-adequate, zinc-depleted ration. Clinical manifestations documented in zinc restricted animals have included dermatitis and hair loss, depressed immune functions, altered reproductive performance, teratogenesis, and skeletal abnormalities.

Deficiency in human adults Although the anorectic response does not dominate in humans with either primary, experimental or secondary zinc deficiency, substantial homology exists between the clinical manifestations in animals and those in humans. The manifestations of zinc deficiency in humans have been catalogued from observations of iatrogenic deficiency, genetic predisposition (acrodermatitis enteropathica), patients with secondary deficiency, endemic deficiency situations, and experimental depletion in volunteers. Recently, there has been a trend to use the response of populations suspected of endemic zinc deficiency to define functional manifestations of deficiency. As a rule, as the severity of zinc deprivation increases, there are a wider array and a greater intensity of clinical manifestations. In chronic deficiency, the zinc deficiency manifestations have a gradual onset, whereas in acute deficiency, signs and symptoms appear rapidly.

The manifestations of human zinc deficiency include the following: behavioral alterations and central nervous system change (anorexia, depression, psychosis, hypogeusia, hyposmia, night blindness, seizure disorder, impaired cognitive performance); integumentary changes (bullous, pustular lesions; keratotic lesions; rough skin; loss of scalp, facial and body hair; impaired wound healing); gastrointestinal manifestations (diarrhea, impaired nutrient digestion and absorption); impaired growth and development (linear growth retardation; weight loss); altered reproductive biology (oligospermia; hypogonadism and reduced potency; fetal teratogenesis); and impaired immunity (recurrent infections; reduced delayed cutaneous hypersensitivity). Situations such as acute deficiency induced during total parenteral nutrition, or the congenital zinc malabsorption syndrome, *acrodermatitis enteropathica*, can bring out the most severe extremes, with manifestations from all systems. In a mild deficiency, anorexia, dry rough skin and oligospermia may be the only manifestations.

Deficiency in children and adolescents All of the signs and symptoms discussed above for adults, with the exception of those related to reproductive function, can be manifested in children and adolescents. However, the imperative of growth and development add additional dimensions to clinical manifestations.

Slow growth leading to growth retardation or even periods of growth arrest are seen in pediatric zinc deficiency. The juvenile deficiency also delays development of puberty and secondary sexual characteristics in both boys and girls. A neonatal seizure disorder has been ascribed to zinc deficiency in preterm infants, a manifestation apparently unique to this specific age-group. (*See* **Adolescents; Children:** Nutritional Problems; Nutritional Requirements; **Infants:** Nutritional Requirements.)

Physiological and Metabolic Consequences

Certain physiological and metabolic consequences of human zinc deficiency do not manifest with clinical signs and symptoms, but are detectable in the laboratory. The physiological alterations of zinc deficiency (information gathered from animal experiments or experimental restrictions in humans), including structural changes in microstructures, hematological abnormalities, metabolic and endocrinological alterations, immune defense defects, nervous system problems, changes in the homeostatic regulation of zinc supply, and assorted miscellaneous changes, are summarized below:

- impaired cell replication;
- impaired collagen synthesis and cross-linking;
- disorganized tubulin synthesis;
- dysplasia of buccal and esophageal mucosa;
- dense inclusion bodies in Paneth cells of the intestine;
- changes in pancreatic exocrine secretion;
- impaired tocopherol absorption;
- impaired platelet aggregation and prolonged bleeding time;
- increased amino acid oxidation;
- reduced leptin gene expression and leptin secretion;
- elevated plasma ammonium levels;
- impaired rhodopsin regeneration in retina;
- synthesis of retinol-binding protein and transport of retinol;
- impaired insulin secretion;
- impaired glucose tolerance;
- elevated cholesterol;
- deranged essential fatty acid and prostaglandin metabolism;
- decreased gonadal hormone production;
- increased lipid peroxidation;
- impaired cellular immunity;
- decreased CD4+: CD8+ T-cell ratio;
- decreased CD73+/CD11b T-cell ratio in the CD8+ subset;
- thymic hypoplasia and atrophy;
- decreased production of thymulin;
- structural changes in the brain;
- decreased seizure thresholds;
- changes in neurotransmitter storage and metabolism;
- decreased peripheral nerve conduction time;
- alteration of taste bud morphology;
- decreased metallothionein synthesis in liver and intestine;
- enhanced absorption of dietary zinc;
- increased retention of zinc; decreased urinary and fecal loss;
- decreased growth of tumors;
- impaired thermal regulation.

Assessment of Human Zinc Nutriture

The assessment of human zinc nutriture is complicated and confounded by issues related both to zinc biology and to analytical chemistry. In theory, the approach to define the status of an individual for any nutrient is to measure the total body pool and/or its availability to the sites of activity. In Golden's concept of a type II nutritional deficiency, the overall amount of zinc in the body or its concentrations in the tissues will not be severely altered; rather, adaptations such as growth arrest and retarded development will occur at an early stage of sensing zinc restriction, and tissue concentrations will be relatively conserved.

Tests that Assess the Total Body Pool and Various Component Pools

The only conceivable direct approach to estimating total-body zinc content is in a postmortem chemical examination or by *in vivo* neutron activation analysis. Neither is practical in the context of patients or populations. One can estimate total-body zinc by isotopic techniques. Radiozinc (^{65}Zn) and several stable isotopes serve in this connection. However, such is the nature of zinc biology that it makes it inappropriate to speak of standing zinc 'stores' or 'reserves.' For this reason, biopsies of various tissues such as bone and muscle would provide a less than useful index of zinc that is available for metabolic purposes.

Rather than ask how much total zinc is in the body, we must search for indicators of the metabolically active or nutritionally relevant zinc. Using multiple isotopes and complex mathematics, the size of various component pools, including the 'exchangeable zinc pool' can be estimated to provide a reliable estimation of individual 'nutriture.' More commonly, less expensive indirect measures are used, with the circulating concentration of zinc (in plasma or serum) being the most common, followed by whole-blood

zinc. In zinc deficiency, the concentration in the circulation declines. Unfortunately, it also declines for a myriad of reasons from normal meal patterns to stress and infection to prolonged fasting to use of contraceptive hormones, etc. The zinc in the circulation at any given moment, however, represents less than 0.01% of the total body supply. For the individual, a normal zinc level is a reasonable indication of adequate status, but an isolated low value is difficult to interpret. For populations, however, the prevalence of low zinc concentrations provides the best epidemiological diagnostic assessment.

The recent development of assays for circulating metallothionein I suggests the possibility of a substance that is low when the body demand is high and high when the demand is low. Newer tests (functional, see below) that are based on the state of regulation and expression of the metallothionein gene may be more sensitive with fewer confounding factors than any assay for the metallothionein protein itself, and more reflective of an active zinc pool. (*See* **Nutritional Assessment**: Importance of Measuring Nutritional Status; Biochemical Tests for Vitamins and Minerals.)

Tests that Assess Tissue Concentrations

A series of accessible tissues can be biopsied and assessed as indicators of zinc availability at peripheral sites, including: platelet zinc, lymphocyte zinc, granulocyte zinc, red cell zinc, buccal mucosa zinc, hair zinc, and nail zinc. The former three circulating cellular elements, in theory, represent tissue that is more rapidly turning over, and hence would represent more current peripheral availability. More than 50% of all circulating red cells are over 60 days old. Studies in experimental mild zinc deprivation in volunteers have shown that levels of zinc decline much earlier in platelets, lymphocytes, and granulocytes than in red cells or plasma. Buccal mucosa is easy to obtain, but it can be contaminated by other sources of zinc. The final three options represent integumentary tissue in which the zinc may be deposited in proportion to the zinc that was mobilized to the underlying, active dermal, or mucosal tissue. For hair and nails, however, the rate of outward growth operates independently of the deposition of the nutrient into the roots, confounding the interpretation of zinc content and concentration.

Tests that Assess the Intactness of Zinc-dependent Functions

The functional assessment of nutritional status involves evaluating the intactness of zinc-dependent physiological functions. For zinc, performance-related tests include growth attainment, gustatory acuity, immune functions, dark adaptation, serum thymulin activity and reconstitution, and zinc metalloenzyme activities, to mention some of the more readily available options. The ratios of specific CD markers on T-lymphocytes represent a newer version of an immune-function measure. When evaluated at a single point in time, the values for these or other functional test must be classified in terms of normative curves of standard populations. As healthy individuals vary widely along these parameters, this weakens their reliability for clinical assessment. A substantial overlap in the distributions of performance results exists for truly deficient and truly sufficient individuals. Moreover, deficits in most of these functions are not specific to zinc deficiency and may be caused by other nutritional or pathological factors. (*See* **Nutritional Assessment**: Functional Tests.)

Tests that Assess the Homeostatic Regulation of Zinc

More specific functional tests for zinc are found in those that examine its regulation. When the metabolism of a nutrient is homeostatically regulated to achieve fairly constant total body content, as in the case with zinc, functional assessment of nutritional status can also involve evaluation of the operation or regulation. Isotopic turnover, again with ^{65}Zn, is an index. Urinary zinc output after a zinc load has been used to determine the threshold when body needs have been met, and the body is no longer in a strict conservation mode. Zinc absorption tests have also been used to gauge regulation, as high rates of uptake will occur in the zinc-deficient state. The regulation of zinc by assessment of the function of metal regulatory elements in gene expression or of rapidly turning over products of the homeostatic mechanism, such as the messenger RNA for metallothionein I, is the most modern and sophisticated candidate in this line.

Responses to Zinc Supplementation

It has been stated correctly that, given the limitations and caveats in the interpretations of all of the classes of zinc status tests, the surest approach to diagnosis of zinc status is retrospective, after a therapeutic trial with zinc. Static indices such as plasma zinc levels or the concentrations in accessible tissues are of limited interpretation in a zinc supplementation trial, as they would be expected to rise passively with increased availability. However, a positive response in *functional* tests either of tests of zinc-dependent functions or of tests of homeostatic regulation would be a significant indication that a zinc-responsive deficit in function or an upregulation state had existed at the time of the original inquiry. Zinc doses should be in the dietary range, as high-dose supplements raise the

specter of a pharmacological effect of zinc, unrelated to underlying nutritional status.

Practical Recommendations for Zinc Status Evaluation

Despite the theoretical considerations that the major response to zinc deficiency involves conservative and adaptive mechanism of a nonspecific nature, total-body zinc and zinc in various tissues and compartments are influenced in most forms and etiologies of human zinc deficiency. Given the pitfalls in the interpretation of any given tests, a battery of tests, combining circulating zinc level assay with several of the other tissue assays and functional tests, is recommended. When a short-term clinical supplementation trial is possible, it can give the most definitive diagnosis short of isotopic studies.

Clinical Conditions Associated with Zinc Deficiency

It follows from the discussion of pathogenesis of zinc deficiency that any disease or disorder that produces decreased intake, decreased absorption, excess losses, impaired utilization, or increased demand could predispose to zinc deficiency. The clinical conditions that have been associated with human zinc deficiency are as follows:

- protein–energy malnutrition;
- vegetarianism;
- zinc-poor diet;
- experimental depletion;
- psychogenic eating disorders (anorexia nervosa; bulimia);
- high-fiber, high phytate diets;
- high-dose iron supplements;
- cancer cachexia;
- synthetic diets;
- zinc-poor total parenteral nutrition;
- acarodermatitis enteropathica;
- Crohn's disease;
- jejuno-ileal bypass;
- chronic renal failure;
- ulcerative colitis;
- celiac disease;
- intestinal helminthiasis (ascariasis, trichuriasis);
- intestinal giardiasis;
- alcoholism;
- alcoholic hepatitis;
- viral hepatitis;
- alcoholic cirrhosis;
- acute alcoholic pancreatitis;
- primary biliary cirrhosis;
- infantile biliary atresia;
- cystic fibrosis;
- Schwachman's syndrome;
- phenylketonuria;
- sickle cell disease;
- thalassemia;
- porphyria;
- muscular dystrophy;
- acquired immunodeficiency syndrome;
- hyperparathyroidism;
- thermal burns;
- short-gut syndrome;
- exocrine pancreatic insufficiency;
- thiazide diuretic use;
- parenteral ethylene diamine tetraacetate chelation therapy;
- oral ethylene diamine pentaacetate chelation therapy;
- oral D(−) penicillamine therapy;
- diabetes mellitus;
- persistent diarrhea;
- ileostomy fluid loss;
- exfoliative dermatitis;
- psoriasis;
- prolonged lactation;
- twin pregnancies;
- very low birth weight;
- rapidly proliferating malignancies;
- synthetic growth hormone therapy.

Prevalence of Zinc Deficiency

When the parameter is zinc deficiency in a medical and clinical context, the prevalence and incidence of the diseases and conditions discussed in the preceding section – and their management – would determine the global prevalence of zinc deficiency. Viewed as a public health problem on a population and epidemiological basis, however, marginal and deficient zinc status is emerging as a subject of great interest. In the classical regions of endemic zinc deficiency, the rural oases of Iran and Egypt, its occurrence has been attributed to interference with zinc bioavailability by the high-extraction, unrefined cereal diets. Characteristics of the diet and the occurrence of adverse health conditions would determine endemic zinc deficiency. Brown and Wuehler have made a crude estimate of risk of endemic zinc deficiency through an assessment of zinc in the global food supply. This is based on national food balance sheets and estimates of the content of zinc and of major inhibitors of its biological availability, primarily phytic acid. They conclude that the daily rations provide 7–9 mg of absorbable zinc per capita in South and South-east Asia, the Middle East, and throughout Africa, with 9–11 mg available in Latin America and China, and 11–12 ng in the affluent countries of North America, Western Europe, and Oceania, placing the former

two geographic categories at risk of endemic zinc deficiency. This has potentially important implications, as a recent population intervention trial with oral zinc has suggested that the incidence of malarial attacks, diarrheal episodes, and acute respiratory infections may be reduced by daily zinc supplements.

Within societies, social and economic deprivation is a common denominator for the expression of zinc deficiency or marginal zinc status in both industrialized and nonindustrialized countries.

As with other nutrients, it is believed that young children, pregnant women, and the elderly would be the most at risk for zinc deficiency. Infants who are exclusively breast-fed are generally protected against zinc deficiency, but formula-fed infants, especially those with iron-fortified diets, could also share some risk for zinc deficiency. Adolescence is associated with a rapid growth spurt, and this could attenuate the adequacy of zinc from a diet that was sufficient when less of the nutrient was being deposited or utilized for cell growth and replication. Zinc loss resulting from recurrent infections, parasites, and climatic conditions (heat, humidity) would also be factors in the etiology of marginal zinc deficiency. (*See* **Elderly**: Nutritional Status.)

Despite these theoretical considerations, bolstered by some experience in attempting to document the zinc status of the various populations and subgroups discussed above, the pitfalls and limitations of diagnostic indices for zinc status have limited both the number of population-based surveys of human zinc status that have been performed and the reliability of the estimates of prevalence of deficiency. The most reliable approach to assessing the prevalence of zinc deficiency would alter that prevalence at the same time, namely a therapeutic trial of zinc supplementation. The individuals who respond with a *functional* improvement, i.e., accelerated growth, better appetite, and improved physiological or cognitive performance, could be classified retrospectively as having been zinc-deficient.

Treatment of Zinc Deficiency

The treatment of clinical or subclinical zinc deficiency requires the restoration of adequate body pools and tissue concentrations of zinc. It invariably involves the administration of amounts of zinc in excess of the daily requirements as zinc-rich foods or in the form of zinc salts or chelates. Salts such as zinc sulfate and chelates such as zinc gluconate are available commercially as oral tablets; no convincing evidence of improved bioavailability of chelates has been forthcoming. The most efficient uptake of zinc from supplements is achieved on an empty stomach, but gastric irritation symptoms often move patients to combining their zinc with meals, to the detriment of its absorption rate.

When subjects cannot tolerate food or oral medications, the intravenous (parenteral) route can be used. For parenteral administration, pharmacological preparations of zinc chloride or zinc sulfate are available in single formulation or in combination with trace elements. The daily dosage original guidelines they put forth by expert panels of various North American physicians groups have remained valid for decades: they put forth the following guidelines for daily parenteral administration of zinc: stable adults, 2–5-4.0 mg; adults with disease-related, enteral losses, 12–17 mg; term infants, $\geq 150\,\mu g\,kg^{-1}$; premature, low-birthweight infants, $> 300\,\mu g\,kg^{-1}$. To avoid the toxicity reported with excessively rapid dosing, parenteral zinc should never be given as a bolus. Zinc chloride and zinc sulfate are compatible in solution with most other intravenous forms of essential micronutrients.

Prevention of Zinc Deficiency

Prevention of zinc deficiency can be considered at the level of the healthy individual adult or child, at the level of the zinc deficiency or a predisposing medical condition, or at the level of public health of nutritionally vulnerable populations or subgroups.

Prevention in Healthy, Normal Individuals

Spontaneous clinical zinc deficiency is unlikely to occur in healthy individuals on a dietary basis, but they are at risk of marginal deficits. The US Institute of Medicine's Food and Nutrition Board in its *Dietary Reference Intakes* has revised the recommendations for dietary intakes that would keep most healthy individuals replete, as has an expert panel for the UN agencies of the World Health Organization and International Atomic Energy Agencies. In a departure from the approach of the past, in which the safe and adequate level for the individual (recommended dietary allowance, safe dietary intake) was the focus of recommendations, the new approach is focused more on the normative distribution of zinc intakes within a population that would be considered to be at low risk for endemic zinc deficiency. For the US population, this is based on the estimated average requirement derived from actual normative zinc intake data. For the world at large, the UN still used a factorial approach to determine the requirement for uptake of zinc, and then added assumptions for different levels of bioavailability in different regional cuisines (high (45–55%), medium (30–35%), and low (10–15%)), and finally adjusted for a standard population variance. For instance, for an infant, 6–12 months of age,

the WHO/IAEA recommended population daily zinc intake for a diet of medium bioavailability is 5.6 mg. For children aged 6–10 years, this figure is 7.5 mg, rising to 13.1 mg in late adolescence; for men above 18 years, 9.4 mg; for women above 18 years, 6.5 mg; for pregnant women, 10 mg; and for lactating women, 12.2 in the first 6 months of breast-feeding. To calculate the intake that would cover 97.5% of the population, a 50% overage is added to these values for the respective individual protective intakes. (*See* **Dietary Requirements of Adults**.)

Sources of highly available zinc include oysters, herring, lean beef, and lean pork. Increasingly, ready-to-eat breakfast cereals and breakfast snack bars are zinc-fortified. Almost all multivitamin-mineral supplements currently contain zinc. Although breast milk has relatively low zinc concentrations, its zinc is highly absorbable. Exclusive breast-feeding should be protective for the infant, although an occasional case of zinc deficiency from an idiopathic abnormal milk supply has led to infantile zinc deficiency.

Prevention in Persons Predisposed to Deficiency

If an individual has a history of zinc deficiency, or has one of the conditions listed above (see Section Physiological and Metabolic Consequences), more attention to primary prevention, or to avoiding recurrence, is required. Eliminating or ameliorating the primary medical or psychological condition should reduce the tendency to underconsumption and/or malutilization of the nutrient. If the underlying disorder cannot be addressed, chronic use of some multiple – two or three times – of the usual recommended level may be necessary.

Prevention in Vulnerable Populations

If epidemiology suggests that a population is prone to zinc deficiency or marginal zinc status, some public health measures are warranted. The classical populations in which zinc deficiency on a dietary basis was described were juveniles in rural areas of Iran and Egypt with a common factor of total zinc intakes in the acceptable range but a consumption of inhibitors of bioavailability (dietary fiber, phytic acid) that reduced its absorption. For such a population, changes in dietary practices to introduce sources of highly absorbable zinc or to adopt alternative processing of traditional cereal grains (germination, fermentation, phytase treatment) might be a solution. Possibilities in plant genetics and plant breeding to reduce inhibitors such as phytic acid or to increase the affinity of the leaves and other tissue to hold divalent cations are under exploration. The final approach is fortification of a staple food with zinc. Hence, the strategies for ensuring a generous intake of zinc in underprivileged populations include popular nutrition education for better food choices, and fortification of some common food-stuffs with additional zinc. As growing children are the usual victims of endemic marginal zinc status, a food vehicle targeted to this age group can provide more effective coverage. Confectionary, breakfast cereals and soft drinks are among the options for reaching the vulnerable population with fortification of zinc. (*See* **Food Fortification**.)

In underprivileged groups in both developed and developing countries, subsections of the population of children who respond to zinc supplementation with improved growth have been identified. Whether or not the additional burden of zinc loss owing to recurrent infections combines with poor appetite and restricted variety of foods to generate a subclass of zinc marginal individuals is worth considering. Once again, increased loss of zinc can be overcome by compensatory increased intakes of the mineral. Chronic supplementation with zinc is likely to be unsustainable and poses a potential risk to copper nutriture and immune function in at least a subsegment of a long-term supplemented population.

See also: **Adolescents**; **Bioavailability of Nutrients**; **Children**: Nutritional Requirements; Nutritional Problems; **Coenzymes**; **Dietary Requirements of Adults**; **Elderly**: Nutritional Status; **Food Fortification**; **Infants**: Nutritional Requirements; **Nutritional Assessment**: Importance of Measuring Nutritional Status; Biochemical Tests for Vitamins and Minerals; Functional Tests

Further Reading

Black RE (1998) Zinc deficiency and child development. *American Journal of Clinical Nutrition* 68(supplement): 464S–469S.

Brown KH and Wuehler S (2000) *Zinc and Human Health: Results of Recent Trials and Implication for Program Interventions and Research*. Ottawa: Micronutrient Initiative.

Gibson RS (1994) Zinc nutrition in developing countries. *Nutrition Research Reviews* 7: 151–173.

Hambidge KM (2000) Zinc and human health: current status and future directions. *Journal of Nutrition* 130: 1341S–1343S.

Prasad AS (1990) Discovery of human zinc deficiency and marginal deficiency of zinc. In: Tomita H (ed.) *Trace Elements in Clinical Medicine*, pp. 3–14. Tokyo: Springer-Verlag.

Sandström BM (2001) Diagnosis of zinc deficiency and excess in individuals and populations. *Food and Nutrition Bulletin* 22: 133–137.

World Health Organization/International Atomic Energy Agency (1996) *Trace Elements in Human Nutrition and Health*. Geneva: WHO.

ZOONOSES

R Cichon, A Platt-Samoraj and J Uradziński,
Warmia and Masuria University in Olsztyn, Poland

Introduction

Zoonoses are defined as 'those diseases and infections which are naturally transmitted between vertebrate animals and man.' In an obligate zoonotic disease, such as anthrax, transmission occurs only from animal to human, whereas in facultative zoonoses, infections are mostly transmitted among humans. Animal-based food may harbor the etiological agent which has infected its living host.

None of these zoonotic diseases is transmitted to humans exclusively by means of food. Food is only occasionally the vehicle of transmission. Since the oral route represents only one of several possible ways of infection, zoonoses are not then foodborne diseases in the general meaning.

It is impossible to give here a complete survey of all zoonoses potentially transmitted to humans via food. This article highlights three of the most important: brucellosis, Q fever, and leptospirosis.

Brucellosis

Brucellosis is an acute febrile disease, or a chronic disease with a wide variety of symptoms. It is a true (an obligate) zoonosis transmitted to humans generally by domestic animals, especially cattle, pigs, and goats. Brucellosis is variously known as undulant fever, Malta fever, Mediterranean fever, or mimic fever.

Occurrence and Prevalence of Disease

This disease is caused by pleomorphic organisms of the genus *Brucella*, which also includes the organisms that are responsible for infectious abortion (Bang's disease) in cattle and swine. Brucellae are highly infectious, even in very small numbers. The organisms can appear as a coccus, a bacillus, or with characteristics of both (a coccobacillus). They may appear singly and, less frequently, in pairs, short chains, or small groups. The brucellae are nonmotile and nonsporing, and they possess neither capsules nor flagellae. Resting stages are not known.

The genus *Brucella* contains six species of bacteria. Important causes of human disease are the three main species: *Brucella melitensis* of goats and sheep, biotypes 1–3; *B. abortus* of cattle, biotypes 1–9; and *B. suis* of pigs, European hares, and reindeer, biotypes 1–4, with the possible exception of biovar 2. Of the others, *B. ovis* causes disease only in sheep, mainly rams; *B. canis* affects dogs, but is occasionally transmitted to humans, causing a mild type of brucellosis; and *B. neatomae* has been isolated from the desert wood rat. Brucellae have also been isolated from many other animal species, both wild and domestic, e.g., poultry, rabbits, rodents, foxes, hares, elk, deer, moose, polecats, hedgehogs, and insects (flies, mosquitoes, ticks, and bedbugs). It seems, therefore, that there are many types of animal reservoir of brucellae. The extent of the role of insects and wild animals in the transmission of brucellosis to humans is not known. The fact that *B. ovis* and *B. canis* were not discovered until 1953 and 1967, respectively, has led to speculation that the genus *Brucella* may be in a phase of rapid evolution and that new species may be emerging.

Brucellosis is distributed everywhere throughout the world. The following groups of people whose occupation brings them into contact with animals are especially endangered: farmers, veterinarians, workers in slaughterhouses and meat-packing and -processing plants, as well as laboratory personnel. It is particularly common in areas where goats, sheep, cattle, and swine are raised. *B. melitensis* infection is prevalent in the Mediterranean area, central Asia, and some parts of Latin America, especially Mexico and Peru; it does not occur in northern Europe, the USA, Canada, South-east Asia, or Australia. Bovine brucellosis (*B. abortus* infection) is moderately prevalent in most parts of the world, with the exception of many developed countries, although it persists in some southern areas of the USA. Porcine brucellosis (*B. suis* infection) is still fairly widespread in Latin America, South-east Asia, and the Pacific region. Canine brucellosis due to *B. canis* occurs in many countries, but poses only a minimal threat to human health.

Mechanism of Entry into Food and Transfer to Human

In the more susceptible species of domestic animal (cattle, goats, sheep, and pigs), enormous multiplication of brucellae occurs in the uterus in the latter part of pregnancy and, to a lesser extent, in the mammary glands during lactation. Abortion, and sometimes parturition at the normal time, leads to

massive contamination of the environment. On the other hand, excretion of brucellae in the milk, usually in much smaller numbers, may continue for years. Thus, milk and dairy products are frequently the source of foodborne brucellosis. Moreover, brucellae are intracellular parasites of the reticuloendothelial system and penetrate the whole body of the infected host.

B. melitensis and *B. suis* transmit more easily to humans then *B. abortus*. The usual portal of entry in humans is by the mouth, either directly by consuming infected dairy and meat products, or indirectly through contact with hands contaminated during work. Less commonly infection may occur via conjunctiva or by inhalation. An outbreak of brucellosis from consumption of wild greens contaminated by sheep and goat urine and feces has also been reported.

Detection in Foods

The detection of *Brucella* species involves cultivation, various serological methods, animal tests and, occasionally, microscopic examination.

There is a simple and sensitive method of detecting *Brucella* contamination in milk, known as the milk ring test. In this test, a *Brucella* antigen stained with hematoxylin is added to milk. Any agglutinins in the milk, which are carried by the cream as it rises to the surface, react with the antigen. Clumping follows, and a blue ring, indicative of the presence of brucellae, is formed. This test can be used to detect contaminated cows' milk but a similar ring test is available for goats' milk. Milk ring tests are unreadable with sheep milk. Recently, a rapid polymerase chain reaction (PCR) method has been developed for detection of *Brucella* species in milk and in milk products. There is also an immunoenzymatic assay (milk enzyme-linked immunosorbent assay (ELISA)), which is more sensitive then the milk ring test and it is a better screening test then PCR.

Fate of Bacteria During Processing of Storage

Brucellae in milk and meat are readily destroyed by conventional pasteurization methods properly used in dairy or meat processing. They are very sensitive to sunlight, being killed in a few hours, and also quite sensitive to acid conditions (pH 4 or below). However, highly relevant to the transmission of brucellae by food products is the extent of survival of these organisms in food. The organisms survive for only a few hours in raw milk at 37 °C, and for 48 h at 8 °C. At ordinary room temperature the organism will survive, according to another report, for several days until souring of the milk kills them. It is possible that other organisms may influence the survival of brucellae since, in sterile milk, they may survive for 17 days. Moreover, they can survive in tapwater for 10 days at 25 °C or 57 days at 8 °C, but in the presence of organic matter, such as cattle urine and feces, soil, and even lake water at 25 °C, they can survive for more than 2 years.

It appears that the brucellae are a group of sturdy organisms, which can survive prolonged periods in milk and dairy products as well as in meat products.

Methods of processing milk products, e.g., cheesemaking from raw milk, generally do not destroy them. Survival of *Brucella* in unpasteurized soft cheese is discussed. If present in cheese, the organism will survive from 38 days to more than 180 days, but aging of cheese tends to destroy the bacteria. In Cheddar cheese, no brucellae survived after 1 year. However, the European Union veterinary advisory committee in 1995 concluded that a ripening time or cheese age of 60 days is sufficient to eliminate a potential brucellosis risk to the consumer. Brucellae in naturally contaminated milk have been shown to survive during the processing of the milk to butter, buttermilk, and icecream. The organism has also been shown to survive in icecream stored below freezing for more than 1 month. Brucellae may survive in ham during pickling in brine for 21 days, and in salted and otherwise cured meats for 150 days, but not after the smoking process. *B. suis* has also been shown to survive for 21 days in pork under refrigeration and *B. abortus* in sausages for 175 days.

Symptoms of Illness and Treatment

Brucellosis in humans provides serious diagnostic problems, partly because the clinical symptoms are variable and often ill-defined and the complications can affect many systems. The incubation period ranges from 10 to 14 days. The onset is influenza-like with recurrent or undulating fever reaching temperatures up to 40–41 °C in the evening, falling to normal each morning. Limb and back pains are unusually severe. Sweating, headache, muscle ache, malaise, and fatigue, as well as anorexia and weight loss are the other symptoms. If untreated, symptoms may persist for months and complications may enter an ill-defined chronic syndrome, involving several body systems and organs.

For reasons which are not known, children, who are the main milk drinkers, seem to be less susceptible to brucellosis than adults. Evidence suggests that in humans, natural infection and complete recovery produce some resistance to reinfection.

Tetracycline alone or in several drug combinations may determine the effective treatment of acute brucellosis in humans if the treatment is prolonged for

6 weeks. Patients with chronic brucellosis frequently need symptomatic treatment in addition to antibiotics, and their response to antibiotic treatment is often disappointing.

Prevention of Brucellosis

Vaccination of animals, particularly when young, and testing animals and slaughtering reactors are probably the best means of controlling brucellosis. Occupationally exposed workers can be protected by wearing impermeable clothing, rubber boots, gloves and face masks, and by practicing good personal hygiene.

Since food and water are not the only modes of transmission, control in these aspects only will not minimize the occurrence of the disease. Nevertheless, removal of these sources of infection will do much to minimize human brucellosis. Thus, meat of all kinds, especially from unknown sources or from contaminated areas, must be cooked thoroughly. Raw milk should never be drunk without pasteurization, no matter what the source. Milk products, such as cheese, butter, and icecream, should be prepared from pasteurized milk. It must be emphasized that, in countries in which brucellosis has been eradicated, individuals can acquire the disease by consuming imported dairy products or by visiting countries in which the disease is endemic.

Q Fever

The Q (for 'query') fever infection was observed for the first time in Queensland, Australia, in 1935, during an outbreak of the febrile illness among abattoir workers. The disease also occurs in the USA, Canada, and Europe. Although a number of animals are susceptible, it is not certain that infection results in overt disease, except in the guinea-pig and in humans.

Occurrence and Prevalence of the Disease

The disease is transmitted generally by tick bites from wild animals (the natural reservoir) to domestic animals, especially ruminants. The infection is frequently symptomless in animals; the mammary gland, parenchymatous organs, muscles, lymph nodes, and uterus of infected slaughter cattle may harbor the organism. The microorganisms grow particularly well in the placenta of these animals, where they may reach a titer value of 10^9 viable organisms per gram of tissue. Transmission of disease from ticks to humans is uncommon. Sporadic cases or outbreaks of the human disease, Q fever, occur primarily among individuals involved in the transport of infected domestic animals and in the meat and dairy industries.

Mechanism of Entry into Food and Transfer to Humans

The main source of infection is highly infectious dust from tick feces deposited on the hides of animals or from dried placenta following parturition. The disease is usually transmitted to humans in three ways: by airborne dissemination in aerosols or dust contaminated with organisms; by processing of animal products, usually wool and hides; or by ingestion of infected food, especially raw milk. It seems to be well established that the infection may be transmitted by drinking infected milk. Milk may contain up to 10^3 infective doses per gram, even when serologically negative.

Detection in Food

Coxiella burnetii, the infectious agent of Q fever, differs sufficiently from other organisms of the genus *Rickettsia* to be classified in a separate genus, *Coxiella*, in the family Rickettsiaceae. Attempts to isolate *C. burnetii* in the laboratory are hazardous and should be avoided. The diagnosis of Q fever can be confirmed serologically by significant increases in complement fixation, agglutination, or direct fluorescent antibody levels to phase I or II antigens. The complement fixation test (CFT) is widely used and regarded as reliable. However, some cows excrete coxiellae without positive CFT titers. The agglutination test is the method of choice for epidemiological survey work. Additionally, the capillary agglutination test (CAT) requires very small amounts of reagents; it is performed similarly to the *Brucella* ring test by overlaying serum, milk, or whey with a partially purified phase I *C. burnetii* antigen. Positive results are indicated by a colored ring (cream layer) and/or agglutination throughout the serum or milk column. Also, ELISA, PCR, and PCR-ELISA techniques are available. PCR technique is more sensitive than standard culture techniques for retrospective diagnosis with frozen samples.

Fate of Bacteria During Processing or Storage

C. burnetii has a high degree of resistance to physical and chemical agents, comparable to that of sporogenic bacteria. This resistance might be attributable to the endospore-like forms. At 4 °C, viability is retained for over 1 year in dried form such as tick feces or wool, as well as in sterile skim milk or unchlorinated water. Meats remain infected for at least 1 month. Whether common dairy pasteurization will kill the organism depends on the procedure applied. Complete inactivation is not always accomplished by exposure to 63 °C for 30 min or 85–90 °C for a few seconds. However, with the

high-temperature, short-time method, no viable coxiellae have been demonstrated as yet in naturally infected milk.

Symptoms of Illness and Treatment

Q fever usually becomes apparent 2–4 weeks after exposure to the causative agent. In most cases Q fever may resemble a mild type of influenza. However, the severity of illness ranges from a subclinical infection to a febrile flu-like illness accompanied by respiratory symptoms, severe headache, arthralgia, and myalgia. Interstitial pneumonia, resembling a viral or primary atypical pneumonia, occurs in about 50% of patients. The characteristic rash of other rickettsial diseases does not occur. Usually, Q fever is a self-limited disease which lasts for 2–3 weeks, but fulminating pneumonia and chronic infections in the form of lipogranulomatous liver disease and endocarditis may complicate its course. To prevent complications, all patients with acute Q fever should be treated with tetracycline and chloramphenicol. Prolonged therapy with tetracycline, usually in combination with a rickettsiacidal drug, is required. *C. burnetii* is also inhibited by rifampicin, and only to a very limited extent by penicillin, streptomycin, or erythromycin.

Prevention of Disease

Persons at risk, such as abattoir workers, dairy and farm workers, veterinary personnel, woolsorters, tanners, and some laboratory research workers, should be monitored serologically at regular intervals and during respiratory or influenza-like illnesses. Also an effective vaccination as a preventive measure should be considered for such groups of population. If Q fever is diagnosed, treatment should be instituted promptly to prevent the serious complications that can develop after the illness.

Leptospirosis

Leptospirosis is a zoonosis that is worldwide in distribution and encompasses a broad spectrum of animal hosts. The primary reservoir hosts are wild animals. Domestic animals, especially dogs, cattle, and swine, may also be an important source of human infection.

Occurrence and Prevalence of the Disease

Leptospires, the causative agents of leptospirosis, have been isolated from approximately 160 mammalian species in the temperate zone. The disease is more widespread in tropical countries where the infectious agent may be one of some 150 pathogenic serotypes of *Leptospira interrogans*, carried by a large variety of hosts. The most important reservoirs or carriers are various rodents, particularly the common Norway rat (*Rattus norvegicus*), which is the principal carrier of the often very virulent serotype *L. interrogans* serovar *icterohaemorrhagiae*. In animals, particularly rodents and dogs, pathogens may colonize the kidneys following infection, which may or may not be symptomatic. Pathogenic leptospires, multiplying in the convoluted tubules, are shed in the urine of infected animals, and susceptible hosts become infected through contact with urine or urine-contaminated environs such as moist soil and surface water. Humans are accidental hosts. In general, infections in humans occur through broken skin and mucous membranes. Thus, persons who work in foul, watery places such as sewers, rice fields, and poorly built mines are exposed to the causative agent.

Mechanisms of Entry into Food and Transfer to Humans

There are two sources of leptospires which may be involved in contamination of food: (1) urine of infected animals, especially rats; and (2) infected soil, dust, and water. However, the possibility of leptospirosis transmitted by contaminated food or water is rare. This is understandable in view of the extreme sensitivity of the organism to acidity. Thus, the stomach serves as a protective barrier. Nevertheless, one case of foodborne leptospirosis has been reported: 8 days previously, the patient had consumed ham which appeared to have been gnawed by rats, and presumably have been urinated on by the rodents. While leptospires may be shed into the milk from infected animals, there are no proven cases of milk-borne leptospirosis in humans. However, milk has a leptospiricidal property which is not destroyed by pasteurization.

Fate of Bacteria During Processing or Storage

Leptospires are very sensitive to heat. They are killed within 10 s at 60 °C and 2 h at 32 °C. Therefore, any pasteurization method used should be effective in killing the organism during food processing.

Leptospires are quite sensitive to many chemical and environmental factors. They are killed by long-chain fatty acids as well as by detergents. These substances, however, have no effect in the presence of albumin or other serum proteins which combine with these chemicals. The organisms are also killed by sea water, chlorine in the water supply, and ultraviolet light. They are killed at varying rates (a few minutes or hours) in liquids at pH values below 5 and above 8.5.

Moisture may play an important role in their survival. Thus, leptospires have been shown to survive for 279 days during the winter in soils with 70% moisture and only 3 days with 10% moisture. Leptospires have also shown a remarkable ability to survive at extremely low temperatures. Thus, when quick-frozen to −70 °C, they survive for many years with no loss in virulence.

In milk, stored at 40 °C for 2 months, pasteurized or heated at 80 °C for 5 min, leptospires are killed within 3 h. This leptospiricidal property of raw or pasteurized milk is destroyed by boiling. Dilution of the milk also decreases the leptospiricidal property. Thus, it has been shown that 1:20 to 1:80 dilutions allow the organisms to survive in milk for 60 days. Leptospires have been shown to survive for 33 h at 5 °C in refrigerated bovine meats. Survival of the organism from 3 to 12 days in muscle, liver, kidney, and spleen at 3–5 °C has also been reported.

Symptoms of Illness and Treatment

Leptospirosis may appear in either a mild or severe form, and both types have similar incubation periods – 7–14 days. Most leptospiral infections are mild and virtually all pathogenic serotypes cause febrile illness in humans. The most prominent symptom is continuous fluctuating fever with temperatures of 38.9–41.9 °C for 2–12 days or more, severe headache, muscle pain, and nausea that persists for approximately 7 days. Jaundice occurs during this phase in more severe infections. The different forms of leptospirosis are known by other names which are suggestive of the causative serotype, duration of symptoms, or the mode or place of transmission or original discovery. Thus, the disease is also known as Fort Bragg fever, canicola fever, 7 days' fever, mud or swamp fever, canecutter's disease, pea picker's disease, and swineherd disease. In the severe form of leptospirosis, also known as Weil's disease, these symptoms are more intense, particularly gastrointestinal bleeding, renal failure, and central nervous system symptoms such as anorexia, nausea, and vomiting.

Clinical manifestations of leptospirosis, however, are not sufficiently characteristic for a specific diagnosis of leptospirosis. As a result, the disease is often initially misdiagnosed as meningitis or hepatitis. Thus, microscopic demonstration of the organisms, serological tests, or both, is necessary for diagnosis. The microscopic agglutination test is more frequently used for serodiagnosis of this disease. Highly sensitive immunoenzymatic assays (ELISA) have recently been used for the detection of leptospiral antibodies in serum. Although leptospires are sensitive to penicillin and tetracycline *in vitro*, use of these drugs in the treatment of leptospirosis is controversial. If treatment is initiated within the first week of disease onset, the general belief is that it can be effective. Later in the disease, immunological damage may already have begun, rendering antimicrobial therapy ineffective.

Prevention of the Disease

Human leptospirosis may be controlled by reducing its prevalence in wild and domestic animals. Although little can be done about controlling the disease in wild animals, leptospirosis in domestic animals can be reduced through vaccination. Doxycycline is a valued preventive drug. Administration of 200 mg weekly may be effective prevention against leptospirosis in endangered people. This group of persons should wear impermeable clothes and boots. An awareness of the possible contamination of, for example, soil, mud, stagnant water, ponds, and streams, and avoiding polluted food or water, along with control of rat populations, is essential to the prevention of the infection. Even though transmission of the disease by ingestion is very rare, there remains the hazard of foodborne transmission. Heating and cooking food and avoiding unchlorinated water would still be a safe rule to follow, particularly if the source of drinking water may be contaminated by animal discharges.

See also: **Contamination of Food**; ***Escherichia coli***: Food Poisoning by Species other than *Escherichia coli*; **Infection, Fever, and Nutrition**; **Milk**: Processing of Liquid Milk

Further Reading

Anusz Z (ed.) (1995) *Q fever in humans and animals*. Poland: Olsztyn University of Agriculture and Technology Press.

Baron S (ed.) (1986) *Medical Microbiology*, 2nd edn. Menlo Park: Addison-Wesley.

Concon JM (1988) *Food Toxicology. Principles and Concepts*, parts A and B. New York: Marcel Dekker.

Elliot RP (ed.) (1978) *Microorganisms in Foods 1. Their Significance and Methods of Enumeration*, 2nd edn. Toronto: University of Toronto Press.

Fournier P, Narrie TJ and Raoult D *et al.* (1998) Diagnosis of Q fever. *Journal of Clinical Microbiology* 36: 1823–1834.

Holt JG (ed.) (1986) *Bergey's Manual of Systematic Bacteriology*, vols. 1 and 2. Baltimore: Williams and Wilkins.

Hui YH, Gorham JR, Murrel KD and Cliver DO (1994) *Foodborne Disease Handbook. Diseases Caused by Bacteria*, vol. 1. New York, USA: Marcel Dekker.

Joint Food and Agricultural Organization/World Health Organization Expert Committee on Zoonoses (1967) *Third Report. World Health Organization, Technical Report Series* no. 378. Geneva: World Health Organization.

Quinn PJ, Carter MEB and Carter GR (1994) *Clinical Veterinary Microbiology*. London: Wolfe Publishing.

Roberts TA and Skinner FA (eds) (1983) *Food Microbiology Advances and Prospects*. Ontario: Academic Press.

Plate 1 Bananas and Plantains A wide range of banana varieties are available in South-east Asian markets. Reproduced from Bananas and Plantains. *Encyclopaedia of Food Science, Food Technology and Nutrition,* Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Plate 2 Bananas and Plantains Bananas for export are usually transported to the packing shed by cableways to minimize skin blemishes due to mechanical damage. Reproduced from Bananas and Plantains. *Encyclopaedia of Food Science, Food Technology and Nutrition,* Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Plate 3 Barrels Stave wood drying in the open air.

Plate 4 Barrels A barrel placed over a brazier for bending.

Plate 5 Barrels Preparation for bending with a brazier.

Plate 6 Barrels Heating barrels.

Plate 7 Barrels Aging spirits in oak barrels. Reproduced from Barrels *Encyclopaedia of Food Science, Food Technology and Nutrition,* Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Plate 8 Beers View of the brewhouse in a microbrewery.

Plate 9 Biscuits, Cookies, and Crackers Biscuits come in a wide range of shapes, sizes, textures, flavors, and colors.

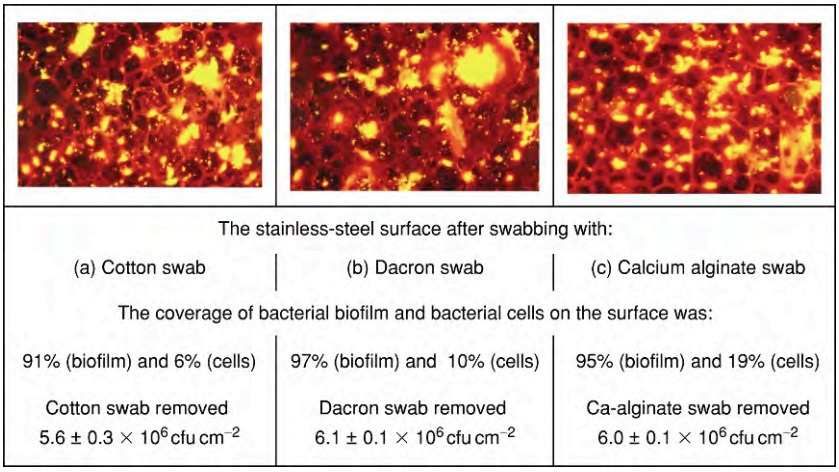

Plate 10 Biofilms Epifluorescence images of 5-day-old *Pseudomonas fragi* biofilm on stainless-steel surfaces (AISI 304,2B) after sampling with cotton, Dacron, and calcium alginate swabs.

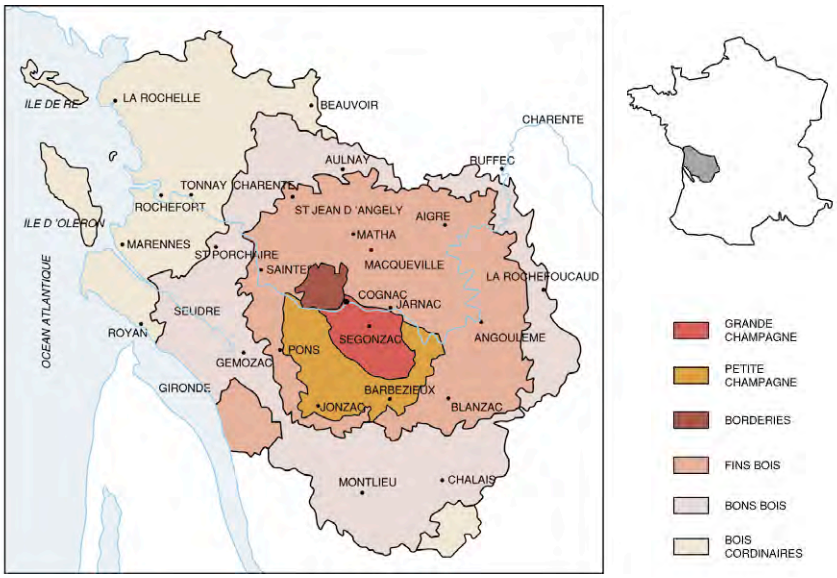

Plate 11 Brandy and Cognac Cognac in France and the different appellation areas. See text for further details.

Plate 12 Butter Stainless steel batch butter churn (twentieth century). From APV Unit Systems, Denmark, with permission.

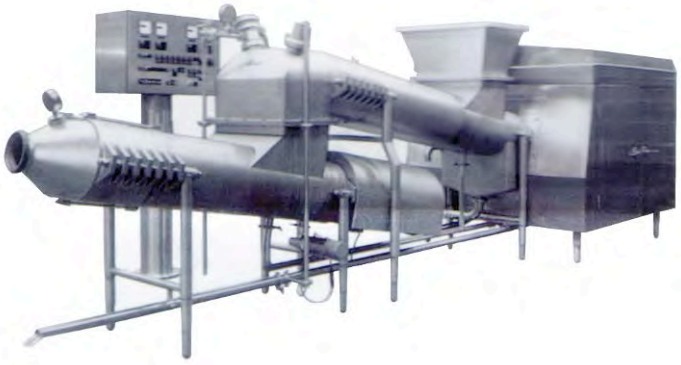

Plate 13 Butter Butter reworking system. From APV Unit Systems, Denmark, with permission.

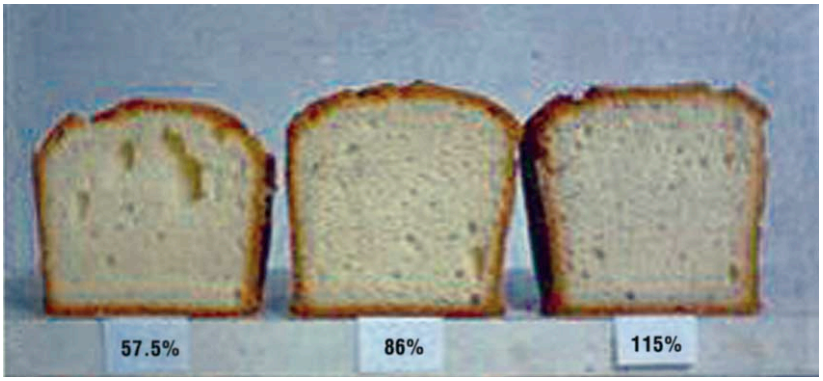

Plate 14 Cakes Cakes with (left) 57.5, (middle) 86 and (right) 115% sucrose based on flour weight.

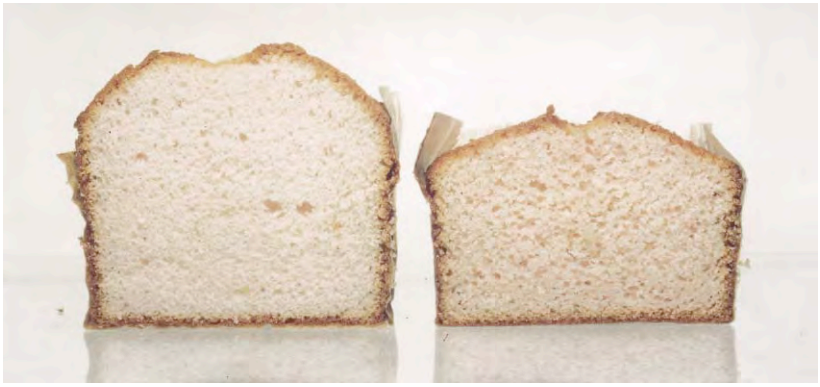

Plate 15 Cakes Cake (left) with and (right) without chemical aeration.

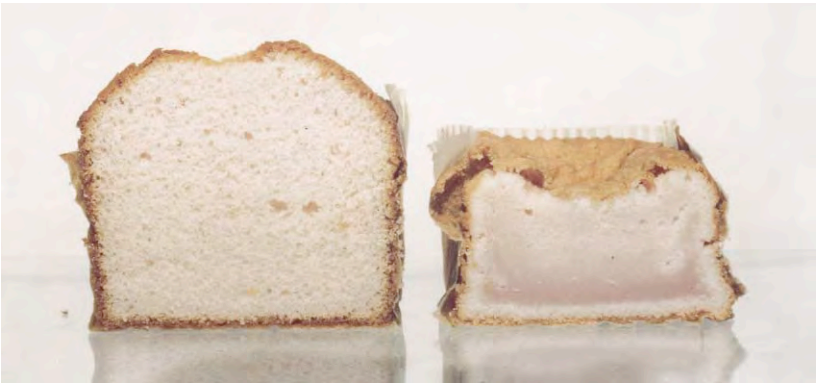

Plate 16 Cakes Cake (left) with and (right) without mechanical aeration.

Plate 17 Casein and Caseinates Biscotti, representing baked products that can be made using high-calcium milk protein. Courtesy of NZMP (North America) Inc.

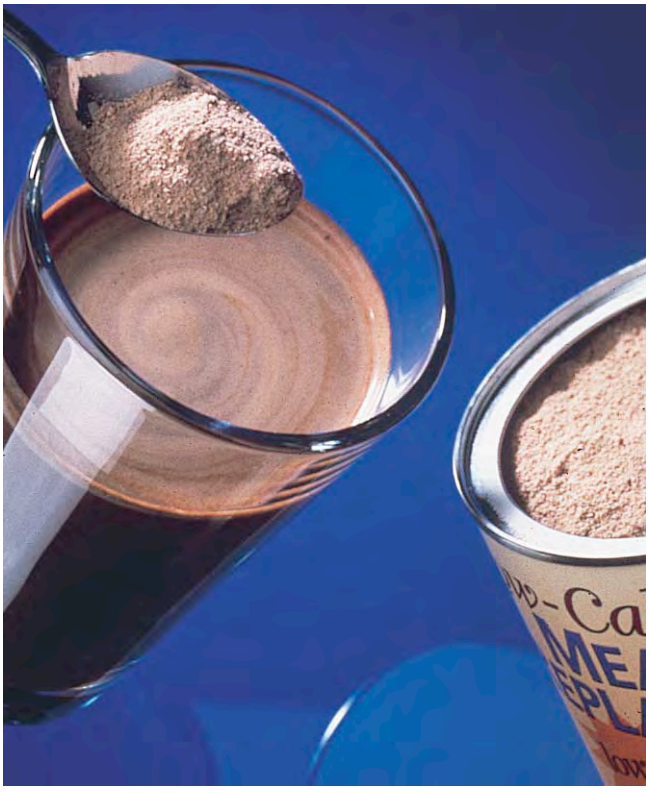

Plate 18 Casein and Caseinates Chocolate drink, representing products for health and nutrition that can be fortified using milk protein isolates and calcium, potassium, or magnesium caseinates. Courtesy of NZMP (North America) Inc.

Plate 19 Casein and Caseinates Pizza, representing the use of imitation Mozzarella cheese that can be made using rennet casein. Courtesy of NZMP (Wellington) Ltd.

Plate 20 Casein and Caseinates Frozen dairy dessert, representing dairy products and analogs that can be made using milk protein isolates. Courtesy of NZMP (North America) Inc.

Plate 21 Casein and Caseinates Candy bars, representing confectionery products that can be made using modified milk proteins and milk protein isolates. Courtesy of NZMP (North America) Inc.

(a)

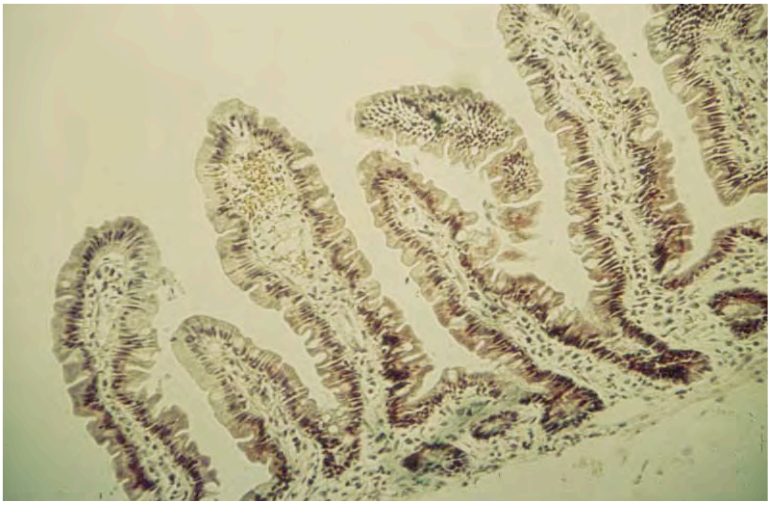

(b)

Plate 22 Celiac (Coeliac) Disease
(a) Dissecting-microscopic appearance of normal mucosal biopsy showing finger-like villi; (b) light-microscopic appearance of normal mucosal biopsy.

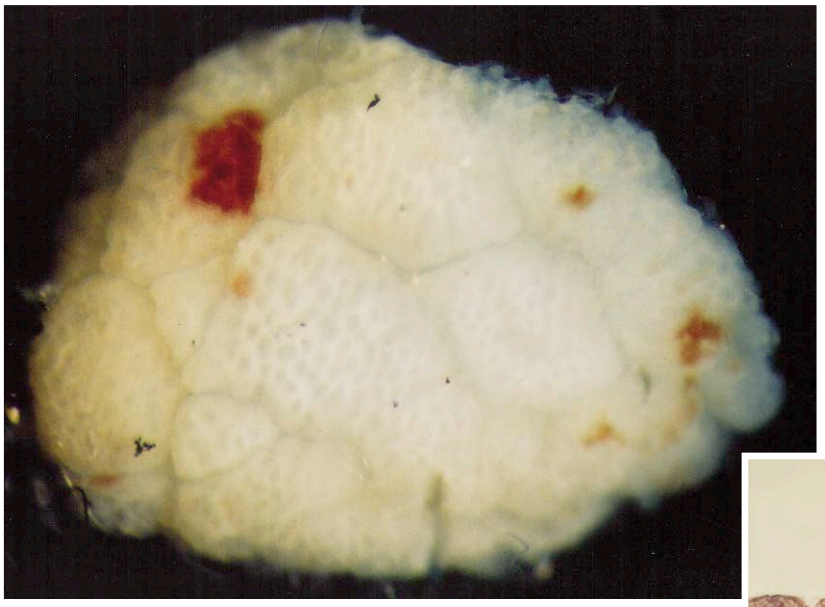

(a)

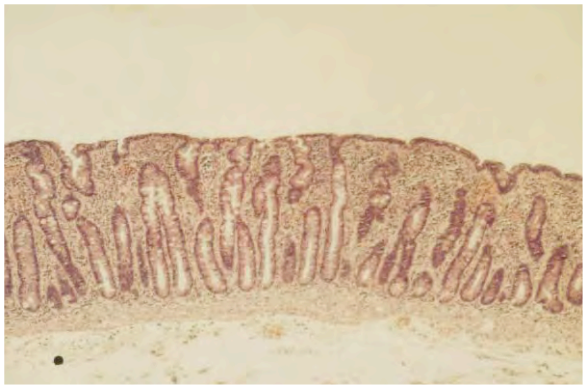

(b)

Plate 23 Celiac (Coeliac) Disease (a) Dissecting-microscopic appearance of biopsy from untreated celiac patient with total absence of villi and a 'flat'-looking mucosal surface; (b) light-microscopic appearance of untreated celiac mucosa with absence of villi, deep crypts, and an increase in inflammatory cells.

Plate 24 Cereals Harbor elevator in Rouen, France, with a total storage capacity of 300 000 tonnes. This is able to load grains into vessels at a rate of 5000 tonnes h^{-1}. Photo courtesy of J. Pfeiffer.

Plate 25 Citrus Fruits Navels are the most important sweet, juicy, and seedless orange marketed for fresh fruit consumption.

Plate 26 Citrus Fruits Limes are normally harvested for fresh marketing when mature but wth green or light green to silver-colored rinds.

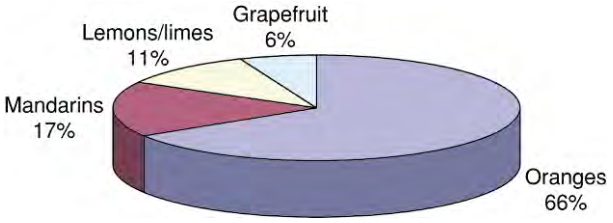

Plate 27 Citrus Fruits Average production percentages for the main citrus types grown in the world between 1995 and 1996 and between 1998 and 1999.

Plate 28 Citrus Fruits A range of internal pigments are now available in grapefruit, with consumer interest in the newer red fleshed cultivars (left to right: Marsh, Thompson, Rio Red, Flame, and Star Ruby).

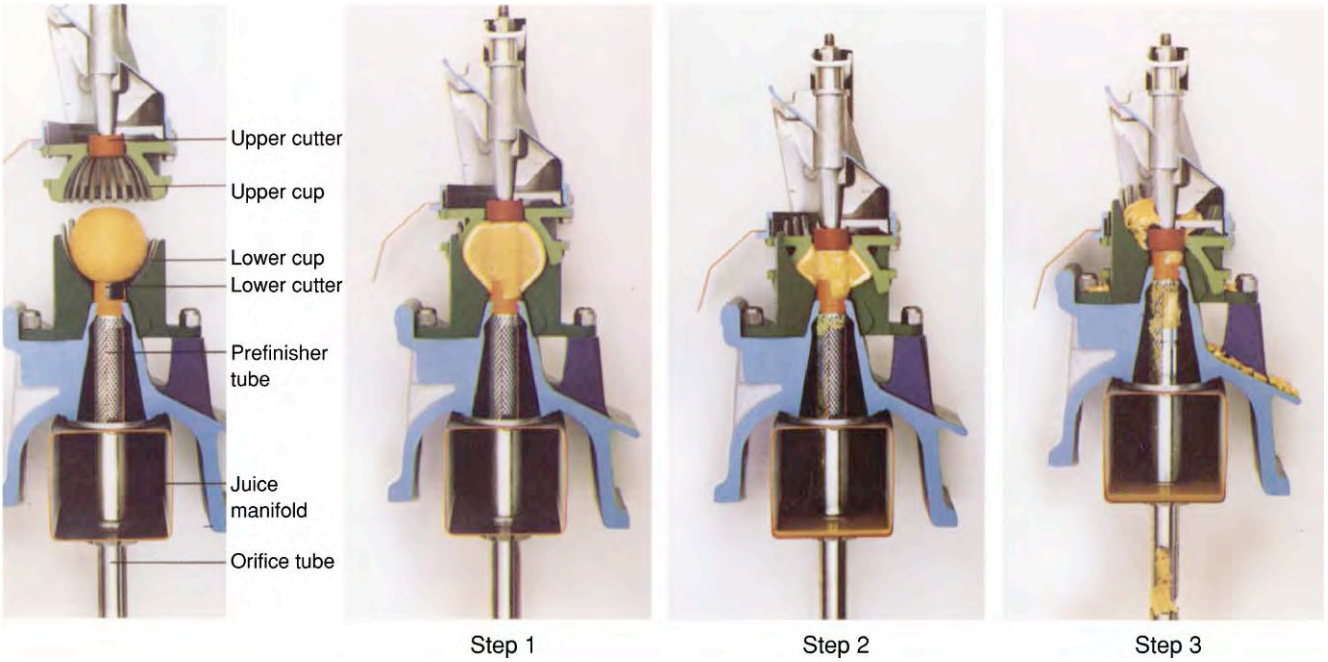

Plate 29 Citrus Fruits FMC juice extractor principle. Courtesy of FMC Food Tech Citrus System, Catania (Sicily) Italy.

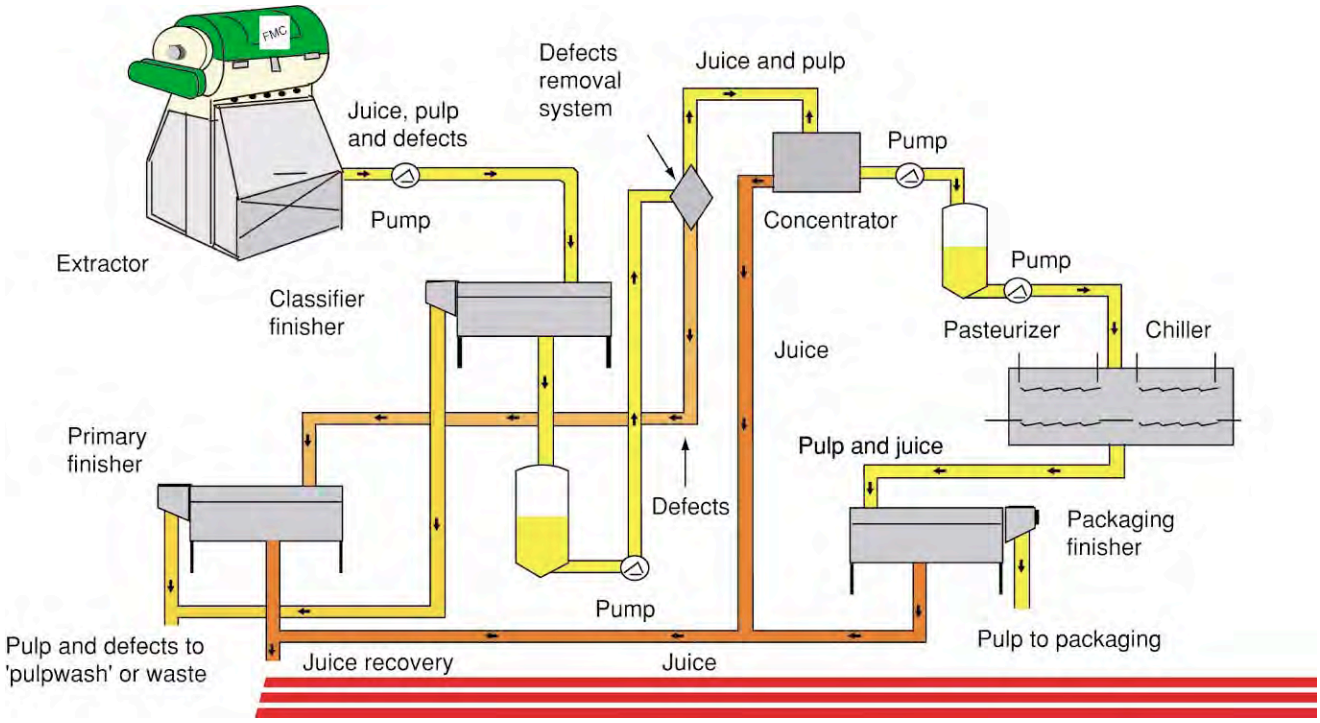

Plate 30 Citrus Fruits Premium pulp recovery process. Courtesy of FMC Food Tech Citrus System, Catania (Sicily) Italy.

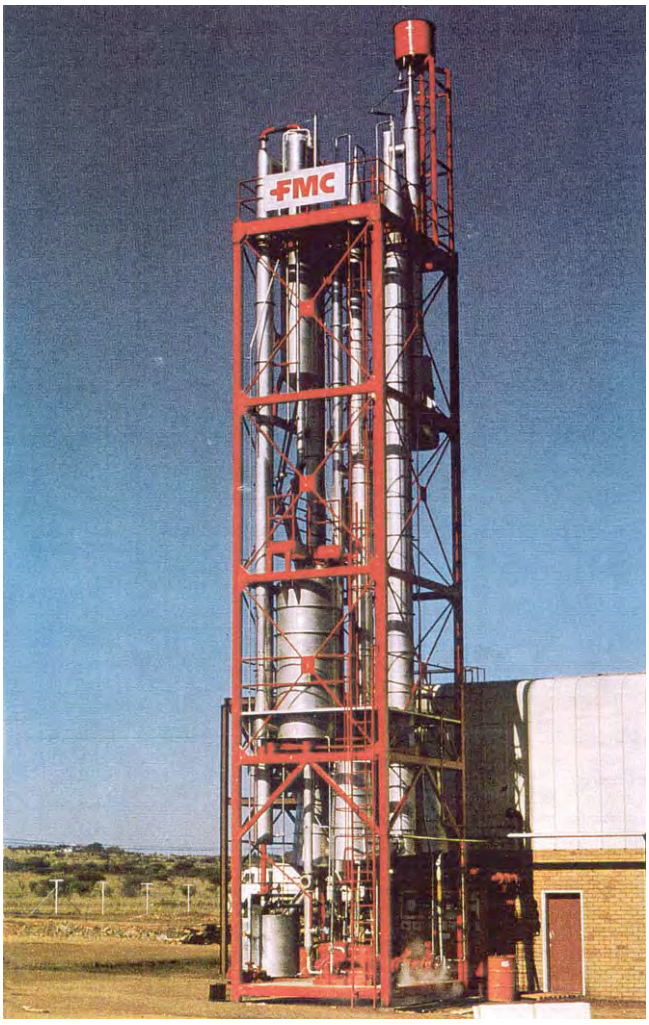

Plate 31 Citrus Fruits Model thermally accelerated short-time evaporator. Courtesy of FMC Food Tech Citrus System, Catania (Sicily) Italy.

Plate 32 Citrus Fruits Harvesting Fortune mandarins for export in Spain.

Plate 33 Citrus Fruits
Moroccan clementine exports
in open 10-kg packages.

Plate 34 Citrus Fruits
Grapefruit are often borne in clusters, like a small bunch of grapes and this has been suggested as the origin of the fruits name.

Plate 35 Citrus Fruits Oroblanco is a new low acid white-fleshed hybrid grapefruit.

Plate 36 Citrus Fruits Star Ruby is a newer grapefruit cultivar with a deep red pigmented flesh.

Plate 37 Coconut Palm Coconut palm.

Plate 38 Coconut Palm Inflorescence.

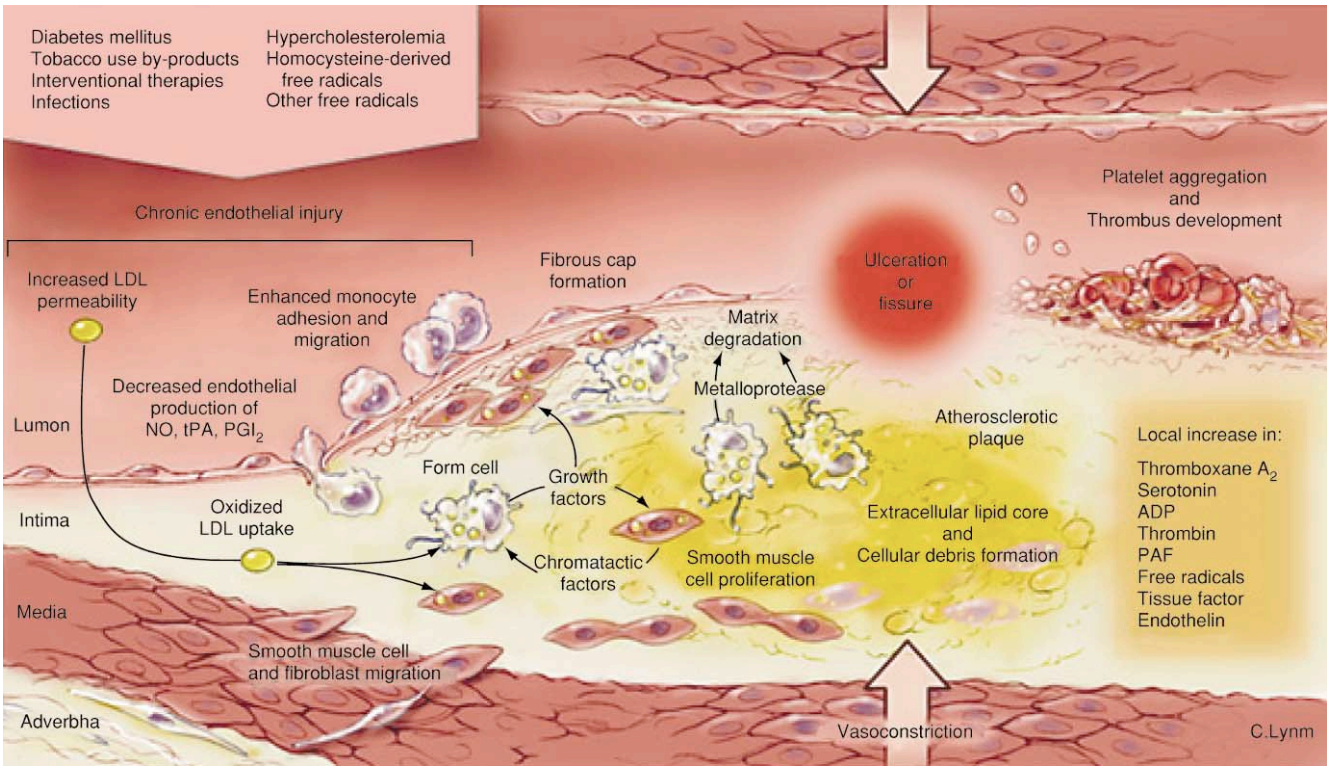

Plate 39 Coronary Heart Disease Mechanisms in the initiation and progression of atherosclerosis. Endothelial injury results from a variety of factors such as smoking, high blood cholesterol and homocysteine, and ulceration of atherosclerotic plaques, etc. (see text). At the site of endothelial injury, there is a downregulation of endothelial-derived nitric oxide (NO), tissue plasminogen activator (tPA), and prostacyclin (PGI_2) with increased adhesion of platelets and leukocytes, increased permeability to lipoproteins, and vasoconstriction. ADP, adenosine diphosphate; LDL, low-density lipoprotein; PAF, platelet activating factor. From Lefkowitz RJ and Willerson JT (2001) Prospects for cardiovascular research. *Journal of the American Medical Association* 285: 581-587 with permission.

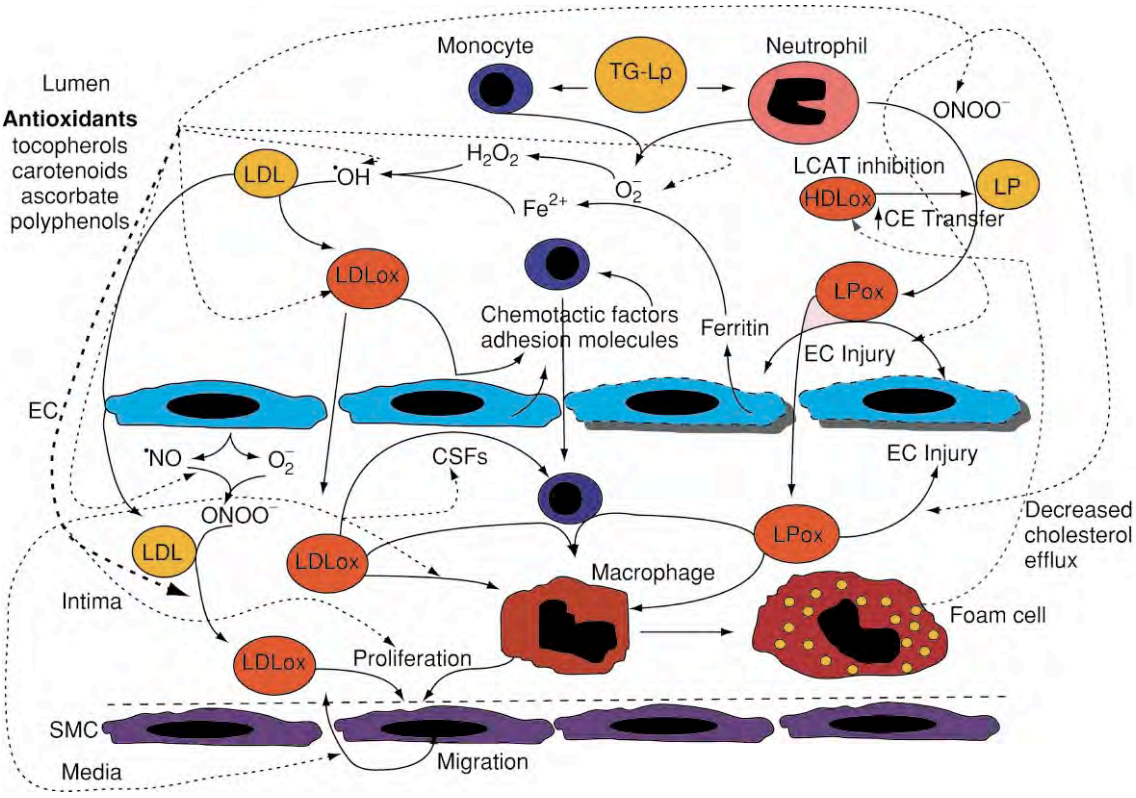

Plate 40 Coronary Heart Diseases Possible effects of antioxidants in the atherogenic process. CE, cholesteryl ester; CSFs, colony-stimulating factors; HDLox, oxidized HDL; LCAT, lecithin: cholesterol acyl transferase; LDLox, oxidized LDL; Lpox, oxidized lipoproteins; TG-Lp, triacylglicerol-rich lipoproteins; EC, endothelial cells; SMC, smooth muscle cells.

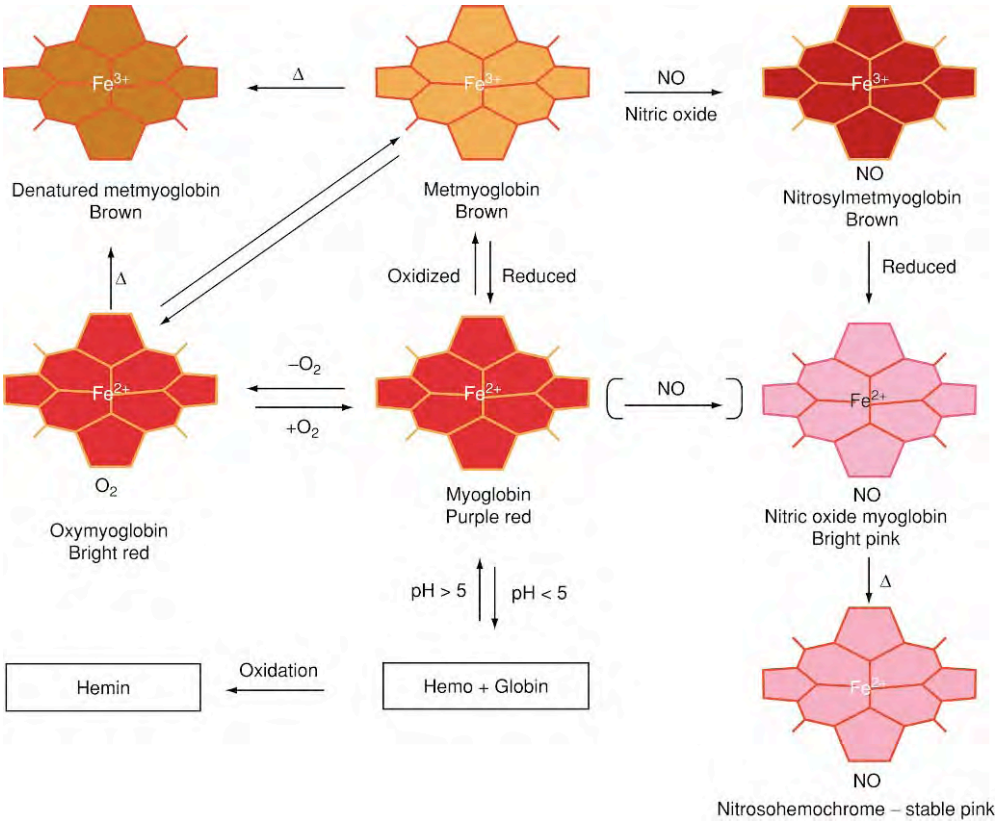

Plate 41 Curing
Changes in the pigment heme during curing and respective color and iron state.

Plate 42 Date Palms (a) Young actively bearing date palm with offshoots or suckers. (b) Bunches of fruit from the same tree photographed at the end of the 'khalaal' stage.

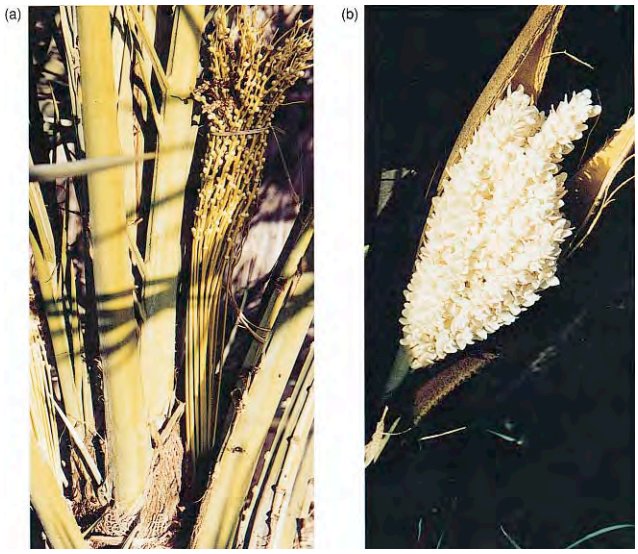

Plate 43 Date Palms (a) Female flower with branches of male inflorescence inserted and tied to the female. (b) Opened male inflorescence with abundant pollen.

Plate 44 Date Palms Eighteen of the most popular date cultivars sold in Riyadh, Saudi Arabia. Their names are as follows: 1=Assela, 2=Labana, 3=Sabbaka, 4=Rothana, 5=Manifi, 6=Naboot-Ali, 7=Sefri, 8=Mabroom, 9=Roshidia, 10=Sokary, 11=Khodri, 12=Safawi, 13=Halaw/Helwa, 14=Hanini, 15=Barhi, 16=Juffair, 17=Qattar, 18=Seeleg.

Plate 45 Date Palms Basket, woven from date fiber and filled with 110 different confectionery products made from dates and date syrup (and dates from 14 dry date cultivars). The background shows semisoft brown date, date leaves, and bunches of mature dates.

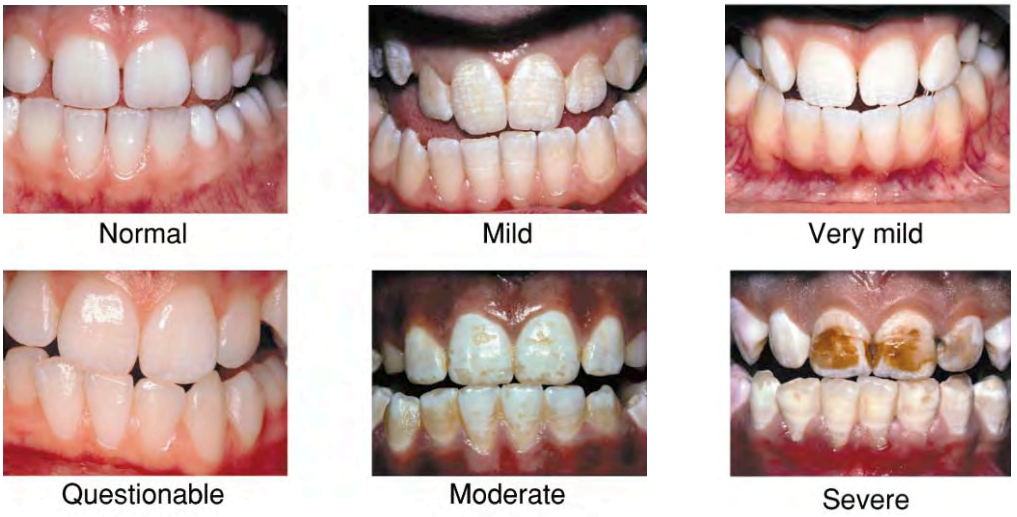

Plate 46 Dental Disease
Categories of dental fluorosis according to Dean's index.

Plate 47 Effluents from Food Processing Round-ended rods. False trichome branching. Note the appearance of the trichome branches being 'stuck' together. False trichome branching does not have contiguous cytoplasm between the branches. Filamentous bacteria, *Sphaerotilus natans.*

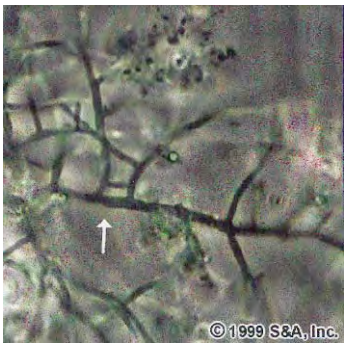

Plate 48 Effluents from Food Processing True trichome branching. True branching refers to contiguous cytoplasm between the branches. Filamentous bacteria, nocardia form. Subject to copyright and patent protection of Stover & Associates, Inc., as part of the FIL-IDENTPro™ Filamentous Bacteria Typing Program.

Plate 49 Effluents from Food Processing
Filamentous bacteria type 021N. This slide was photographed at 1000 x magnification under phase contrast. Subject to copyright and patent protection of Stover & Associates, Inc., as part of the FIL-IDENTPro™ Filamentous Bacteria Typing Program.

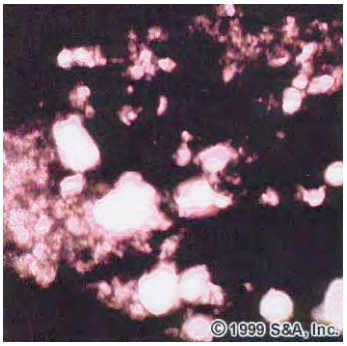

Plate 50 Effluents from Food Processing India ink staining for extracellular polysaccharide. India ink particles penetrate the flocs almost completely, at most leaving a small clear center. In activated sludge containing large amounts of extracellular material (as shown in picture), large clear areas indicate areas of low cell density. This slide was photographed at 10 x magnification. Subject to copyright and patent protection of Stover & Associates, Inc., as part of the FIL-IDENTPro™ Filamentous Bacteria Typing Program.

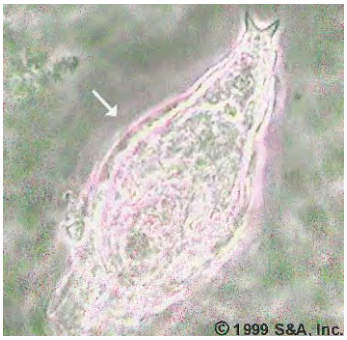

Plate 51 Effluents from Food Processing
Unidentified species of the phylum Rotatoria, commonly called rotifers. This slide was photographed at 100 x magnification. Subject to copyright and patent protection of Stover & Associates, Inc., as part of the FIL-IDENTPro™ Filamentous Bacteria Typing Program.

Plate 52 Fermented Foods Tempe. Growth of a fungus, *Rhizopus oligosporus*, binds cooked, dehulled soybeans into a solid cake. Tempe is a traditional Indonesian meat substitue that also has a small part of the vegetarian market in Japan, the USA and Europe.

Plate 53 Fermented Foods Oncom (ontjom) in Bandung market, Indonesia. Oncom is similar to tempe but uses peanut press cake and the orange-spored fungus, *Neurospora sitophila.* The material in the background is fried tempe.

Plate 54 Fermented Foods Natto. A Japanese *Bacillus* proteolytic fermentation of whole soybeans producing a characteristically slimy, strongly flavoured product.

Plate 55 Fermented Foods Thu-nuo, dried disks. A traditional Thai *Bacillus* proteolytic fermentation of soybeans.

Plate 56 Fermented Foods Tapé in Bandung market, Indonesia. Cooked cassava roots fermented with fungi to produce a slightly soured, slightly alcoholic, and sweetened cassava.

Plate 57 Flavor (Flavour) Compounds Flavor compounds: highly efficient distillation columns. Reproduced from Flavour Compounds Production Methods, *Encyclopaedia of Food Science, Food Technology and Nutrition,* Macrae R, Robinson RK and Sadler M (eds), 1993, Academic Press.

Plate 58 Flavor (Flavour) Compounds Yellowfin tuna. Reproduced from Flavour Compounds, Production Methods, *Encyclopaedia of Food Science, Food Technology and Nutrition,* Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Plate 59 Flavor (Flavour) Compounds Effluents from food processing. Reproduced from Flavour Compounds, Production Methods, *Encyclopaedia of Food Science, Food Technology an Nutrition,* Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

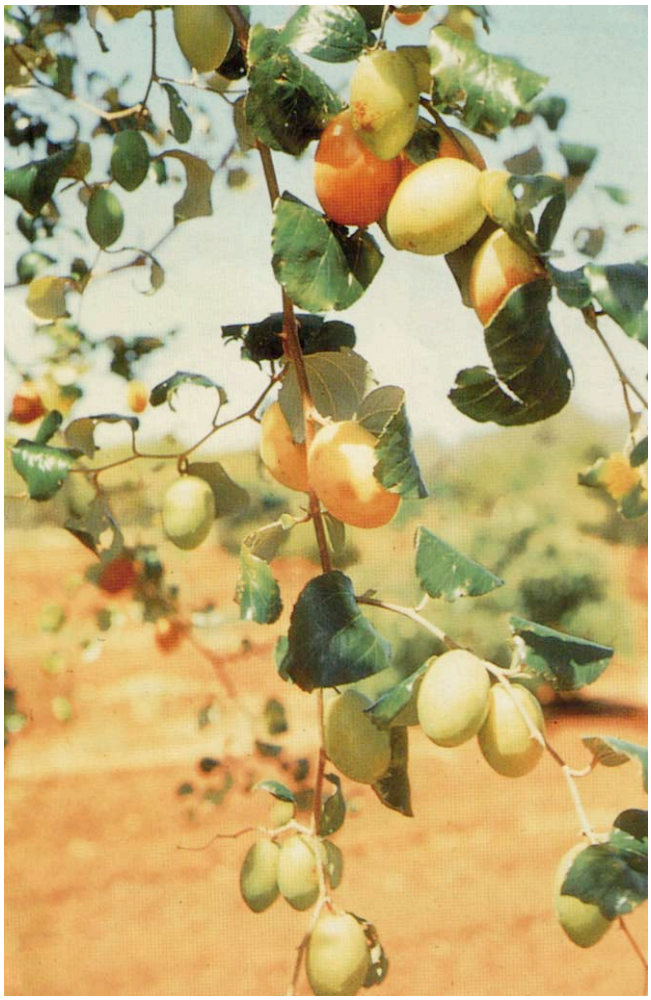

Plate 60 Flavor (Flavour) Compounds Fruits of tropical climates: *Zizyphus mauritiana* Lam. (Indian jujube). Reproduced from Flavour Compounds, Production Methods, *Encyclopaedia of Food Science, Food Technology and Nutrition,* Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Plate 61 Flavor (Flavour) Compounds Road-side markets in Bangkok displaying citrus, banana, rambutan, mango, and lychee. Reproduced from Flavour Compounds, Production Methods, *Encyclopaedia of Food Science, Food Technology and Nutrition,* Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Plate 62 Flavor (Flavour) Compounds A collection of the more exotic tropical fruit of South-east Asia. Courtesy of TA Cooke, Plant Pathology, QDPI, Australia. Reproduced from Flavour Compounds, Production Methods, *Encyclopaedia of Food Science, Food Technology and Nutrition,* Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Plate 63 Flavor (Flavour) Compounds Fruits of tropical climates: *Prunus serotina* var. *salicifolia* (capulin cherry). Reproduced from Flavour Compounds, Production Methods, *Encyclopaedia of Food Science, Food Technology and Nutrition,* Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Plate 64 Flavor (Flavour) Compounds Filtration line for 450hl output per hour; consisting of FILTROstar candle-type precoat filte and PVPP (polyvinyl pyrrolidone, a filter aid dispersed into a cellulose-based matrix) stabilizing plant Filter-o-mat S. Reproduce from Flavour Compounds, Production Methods, *Encyclopaedia of ood Science, Food Technology and Nutrition,* Macrae R, Robinso RK and Sadler MJ (eds), 1993, Academic Press.

Plate 65 Flavor (Flavour) Compounds Kieselgur filtration plant with 35m^2 filtering area and slurry feeder DOSIMAT 500. Special execution with automatic filtration monitoring. Reproduced from Flavour Compounds, Production Methods, *Encyclopaedia of Food Science, Food Technology and Nutrition,* Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

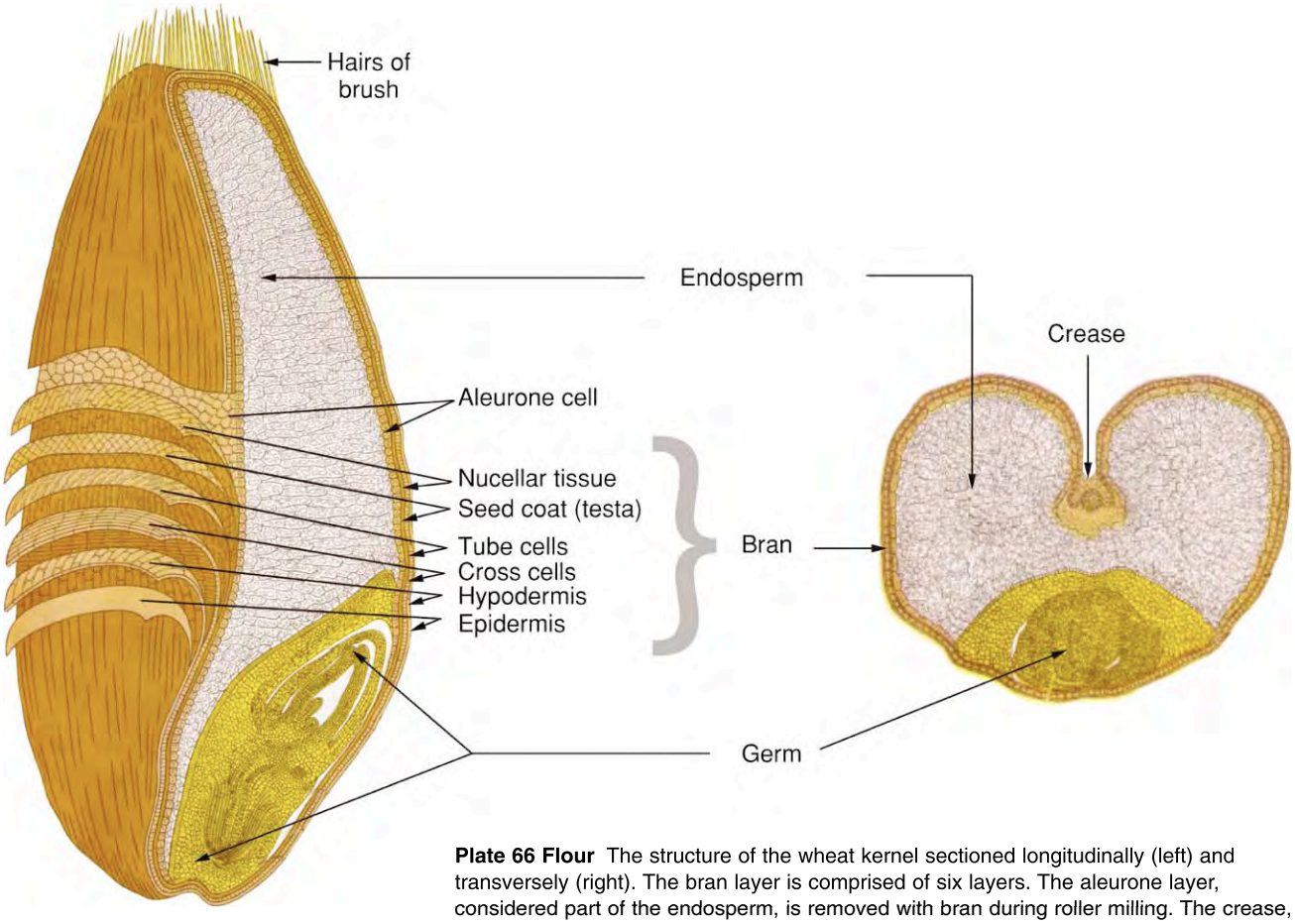

Plate 66 Flour The structure of the wheat kernel sectioned longitudinally (left) and transversely (right). The bran layer is comprised of six layers. The aleurone layer, considered part of the endosperm, is removed with bran during roller milling. The crease, which runs the full length of the kernel, prevents fully efficient separation of bran from endosperm.

Plate 67 Fruits of Temperate Climates Interspecific hybridisation in *Rubus* breeding (R. *phoenicolasius* is used as the donor of resistance to raspberry beetle (*Butyrus tomentosus)),* showing the wild accession (left), a commercial red raspberry (right), and the resultant F_1 hybrid.

Plate 68 Fruits of Tropical Climates Flowers and fruit of *Carissa macrocarpa* (num-num).

Plate 69 Fruits of Tropical Climates Fruit of *Dovyalis caffra* (kei-apple).

Plate 70 Fruits of Tropical Climates Fruit of *Mimusops zeyheri* (moepel).

Plate 71 Fruits of Tropical Climates
Fruit of *Garcinia livingstonei* (African mangosteen).

Plate 72 Fruits of Tropical Climates
Leaves and mature fruit of *Parinari curatellifolia* (mobola plum).

Plate 73 Fruits of Tropical Climates
Fruit of *Vangueria infausta* (wild maedlar).

Plate 74 Fruits of Tropical Climates Fruits of *Baccaurea motleyana.*

Plate 75 Fruits of Tropical Climates Fruits of *Baccaurea polyneura.*

Plate 76 Fruits of Tropical Climates Fruits of *Flacourtia rukam.*

Plate 77 Fruits of Tropical Climates Fruits of *Garcinia atroviridis.*

Plate 78 Gin Juniper berries, the essential botanical ingredient in gin, together with commonly used coriander seeds.

(a)

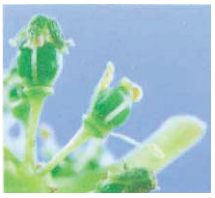

(b)

(c)

(d)

(e)

(f)

Plate 79 Grapes Stages in the flowering and fruit development of grapes: A, flowers with the fused petals (calyptra) about to be shed; B, calyptra in the last stage of separation and anthers about to shed pollen; C, very young grape still showing filaments of anthers; D, young cluster of green grapes; E, maturing grapes beginning to change color *(vèraison)*, the bloom on the berry surface is evident; F, fully mature grapes of 'Cabernet Sauvignon.' From Flaherty DL, Christensen LP, Lanini WT *et al. (1992) Grape Pest Management, Publication No. 3343,* 2nd edn. Oakland, CA: University of California, with permission.

Plate 80 Gums Typical desserts incorporating gums. The additives contained in each dessert include the following: chocolate delight: modified starch, carrageenan, sodium alginate; caramel dessert: modified starch, carrageenan, pectin; raspberry royale: amidated pectin, sodium alginate; pineapple cheesecake: modified starch, gelatin, pectin, guar gum; strawberry fruit fool: gelatin, sodium alginate.

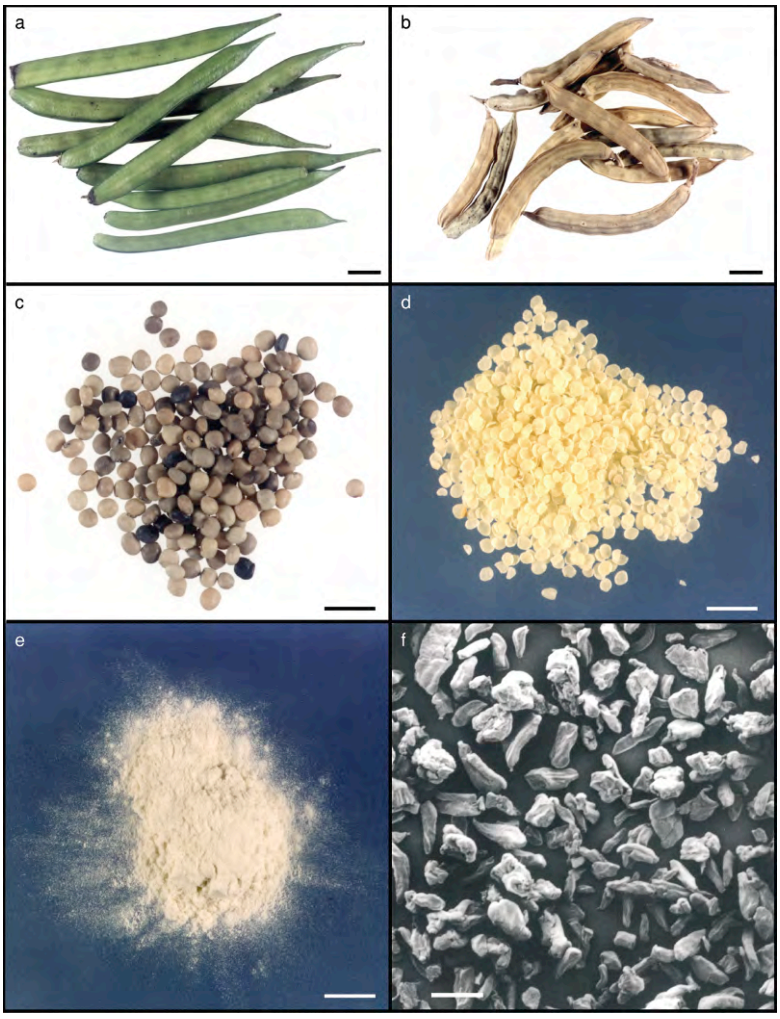

Plate 81 Gums Pictures of pods, seeds, and endosperm extract of the guar plant (*Cyamopsis tetragonoloba* (L.) Taub.), a member of the Leguminosae family. (a) Green pods, scale bar = 1.3 cm. (b) Dried pods, scale bar = 1.6 cm. (c) Seeds, scale bar = 1 cm. (d) Endosperm halves (splits), scale bar = 1 cm. (e) Guar gum flour, scale bar = 1 cm. (f) Scanning electron micrograph of guar gum flour, scale bar = 100 μm.

Plate 82 Herbs Origanum vulgare.

Plate 83 Herbs Coridothymus capitatus.

Plate 84 Herbs Salvia officinalis.

Plate 85 Herbs Herbs of the Umbelliferae: anise. Reproduced from Herbs/Herbs of the Umbeliferae, *Encyclopaedia of Food Science, Food Technology and Nutrition,* Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Plate 86 Herbs Herbs of the Umbelliferae: chervil. Reproduced from Herbs/Herbs of the Umbelliferae, *Encyclopaedia of Food Science, Food Technology and Nutrition,* Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Plate 87 Herbs Herbs of the Umbelliferae: coriander. Reproduced from Herbs/Herbs of the Umbelliferae, *Encyclopaedia of Food Science, Food Technology and Nutrition,* Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Plate 88 Herbs Herbs of the Umbelliferae: fennel. Reproduced from Herbs/Herbs of the Umbelliferae, *Encyclopaedia of Food Science, Food Technology and Nutrition,* Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

Plate 89 Herbs Chamomile field in Norfolk.

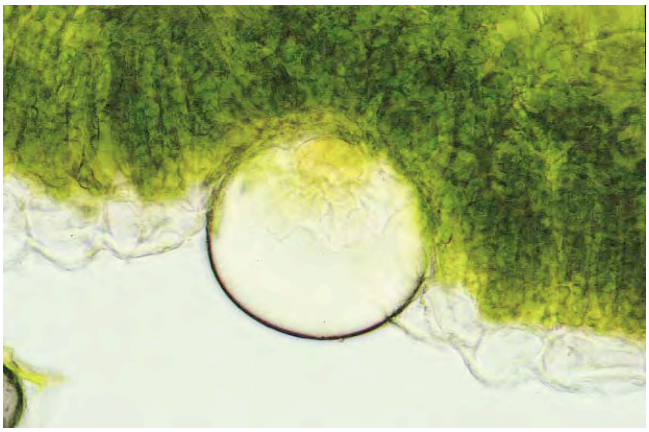

Plate 90 Herbs Volatile oil gland on the surface of a Greek oregano *(Origanum heracleoticum)* leaf.

Plate 91 Herbs *Agastache* species (Korean mint) at the Auchincruive herb garden.

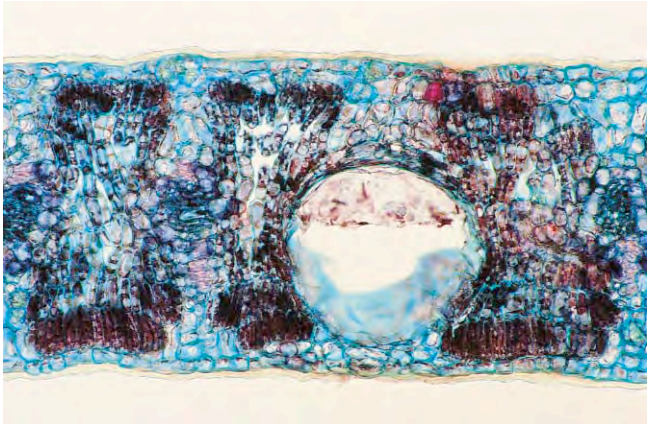

Plate 92 Herbs Cross-section through a eucalyptus leaf, showing the oil cavity.

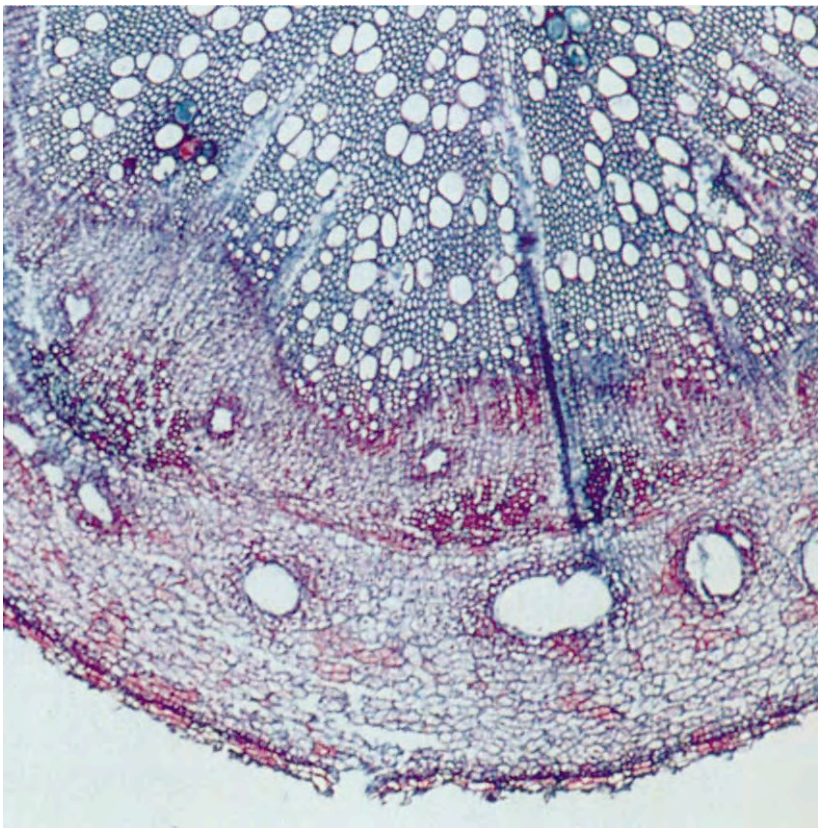

Plate 93 Herbs Volatile oil canals in root of *Artemisia absinthium* (wormwood). Bright field x 40. Reproduced from Herbs: Herbs of the Compositae, *Encyclopaedia of Food Science, Food Technology and Nutrition,* Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

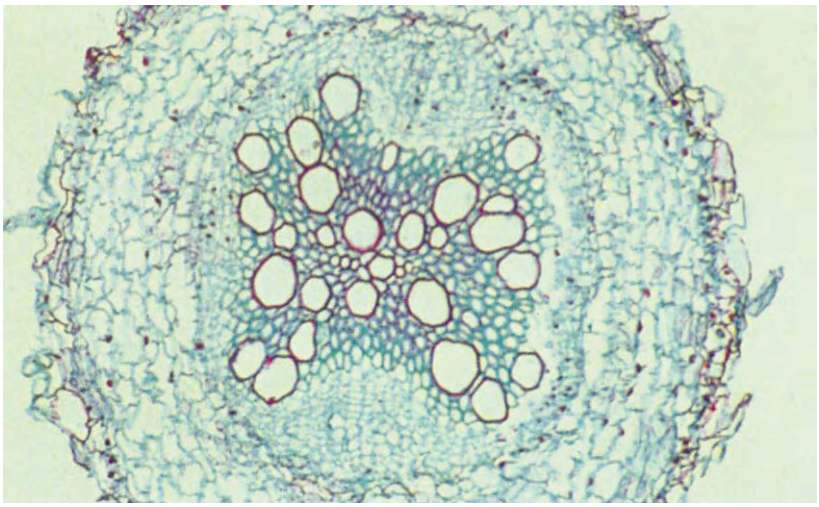

Plate 94 Herbs Transverse section of *Artemisia absinthium* (wormwood) root showing vascular bundle. x 200. Reproduced from Herbs: Herbs of the Compositae, *Encyclopaedia of Food Science, Food Technology and Nutrition,* Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

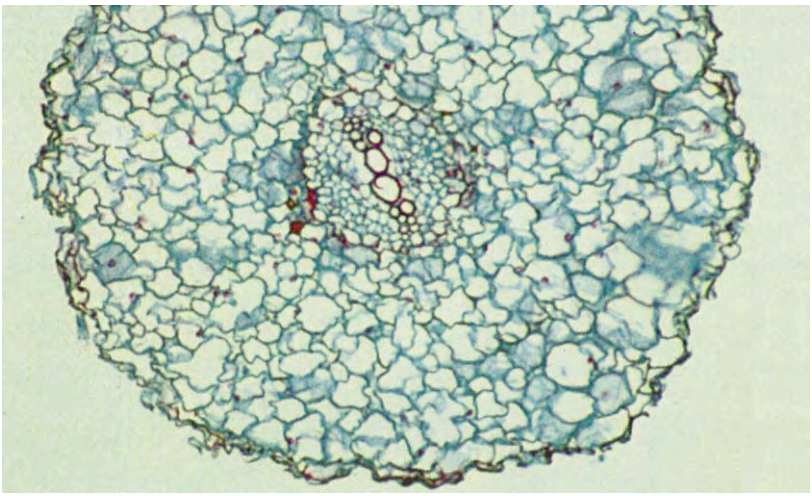

Plate 95 Herbs Transverse section of *Artemisia absinthium* (wormwood) root showing vascular bundle in transformed root. x 200. Reproduced from Herbs: Herbs of the Compositae, *Encyclopaedia of Food Science, Food Technology and Nutrition,* Macrae R, Robinson RK and Sadler MJ (eds), 1993, Academic Press.

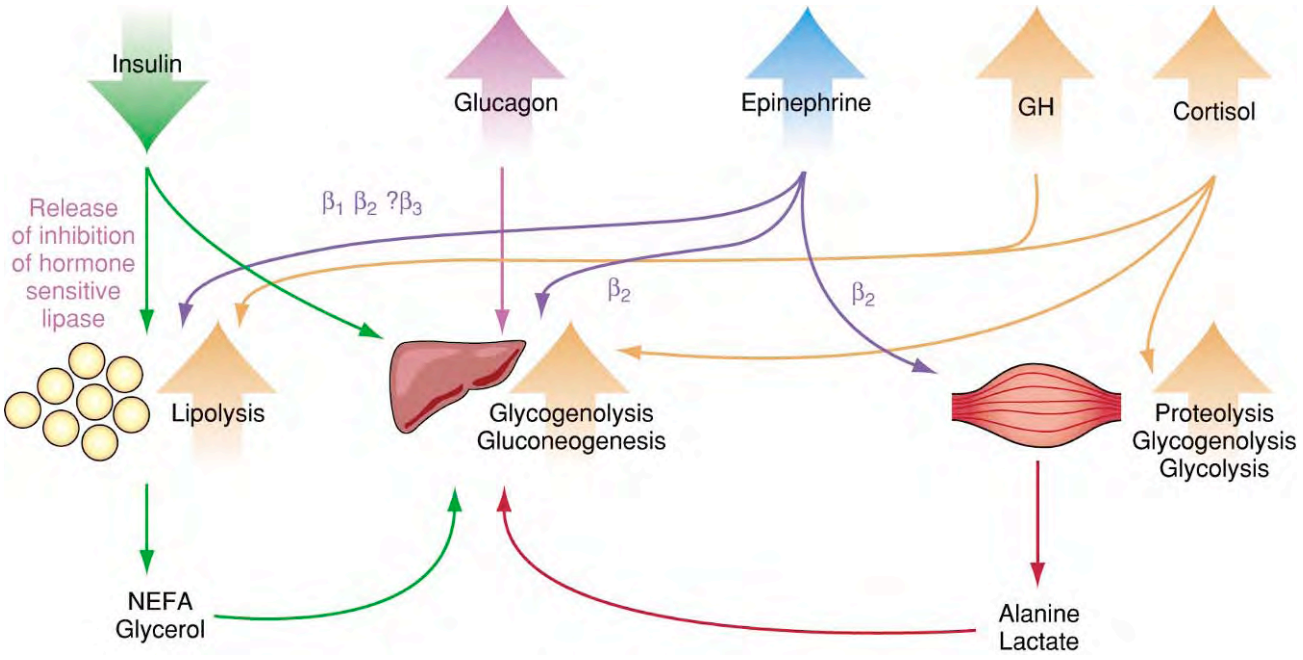

Plate 96 Hypoglycemia (Hypoglycaemia) Hormonal effects on glucose metabolism during fasting. ß$_1$, ß$_2$ and ß$_3$, receptor-mediated effects of epinephrine; NEFA, nonesterified fatty acids.

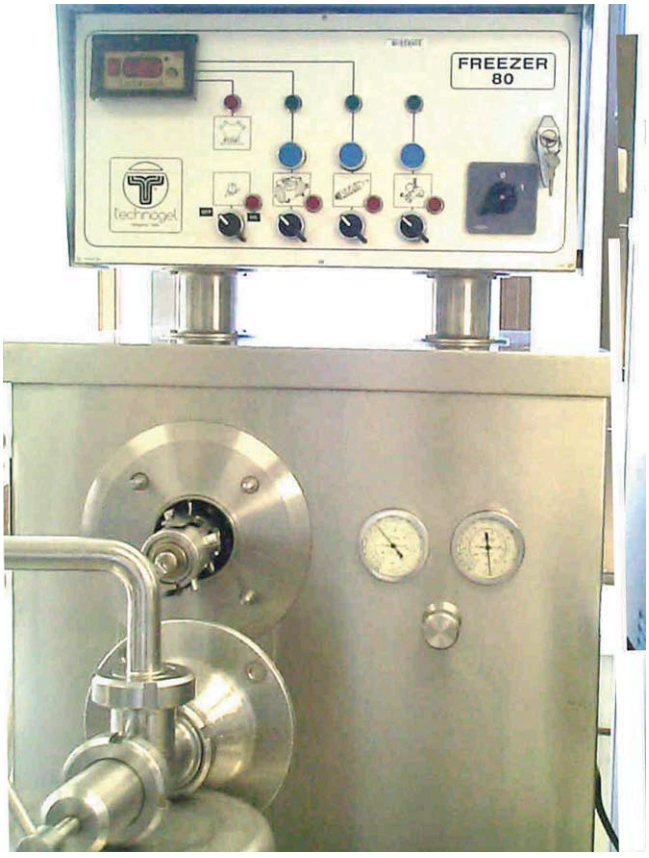

Plate 97 Ice cream Continuous-type freezer with front removed revealing the freezing cylinder and dasher. To regulate the flow of mix and the extent of freezing, pressure is exerted on the freezing chamber by turning the handle located at the outlet for the frozen product in the cylinder cover.

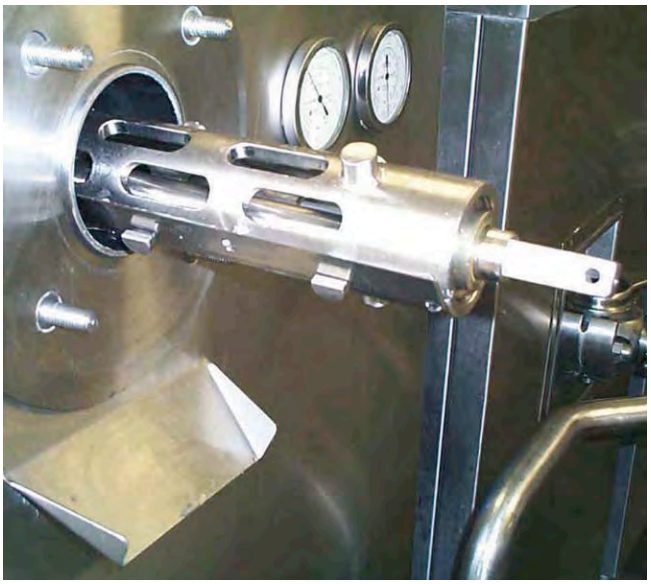

Plate 98 Ice cream Open-type dasher with sharp scraper blades. The dasher is partially inserted into the freezing cylinder of the continuous freezer.

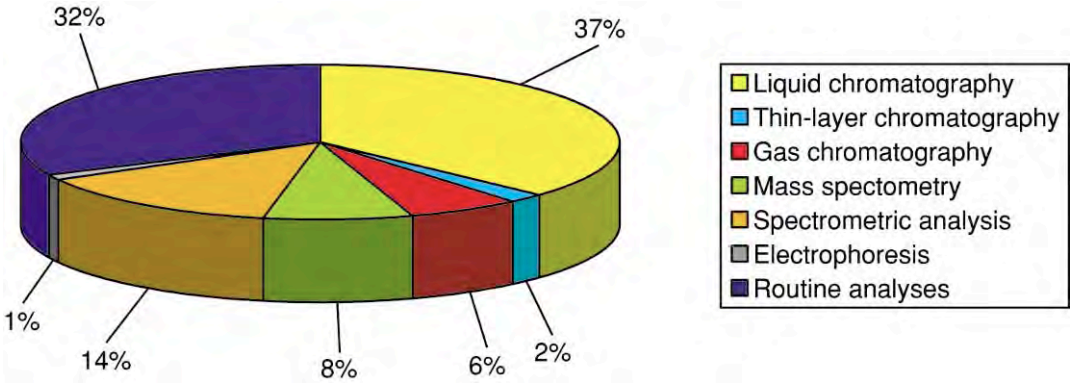

Plate 99 Inborn Errors of Metabolism Laboratory analyses performed in body fluids of 2150 patients investigated over a period of 5 years in the laboratory for metabolic diseases of the Vrije Universiteit Medical Center in Amsterdam. Investigations were performed in whole blood, plasma, urine, amniotic fluid, and cerebrospinal fluid.

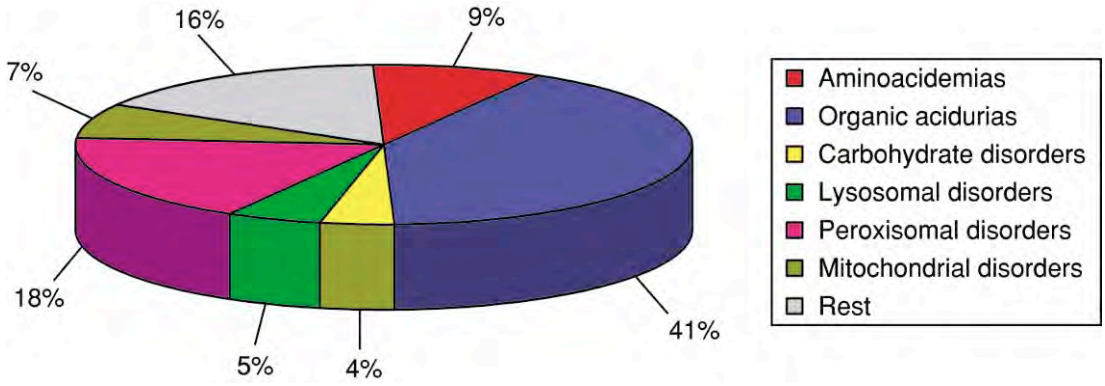

Plate 100 Inborn Errors of Metabolism Frequency distribution according to diagnostic group in 263 patients newly identified with an inborn error of metabolism over a period of 5 years in the Laboratory for Metabolic Diseases of the Vrije Universiteit Medical Center in Amsterdam. Fatty acid oxidation defects included in organic acidurias.

Plate 101 Lupin The diversity of lupin seeds. Different species show remarkable differences in size, color and shape.

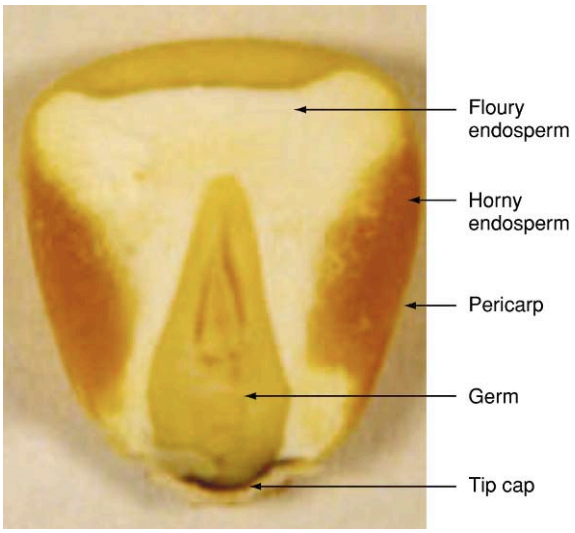

Plate 102 Maize Maize Kernel.

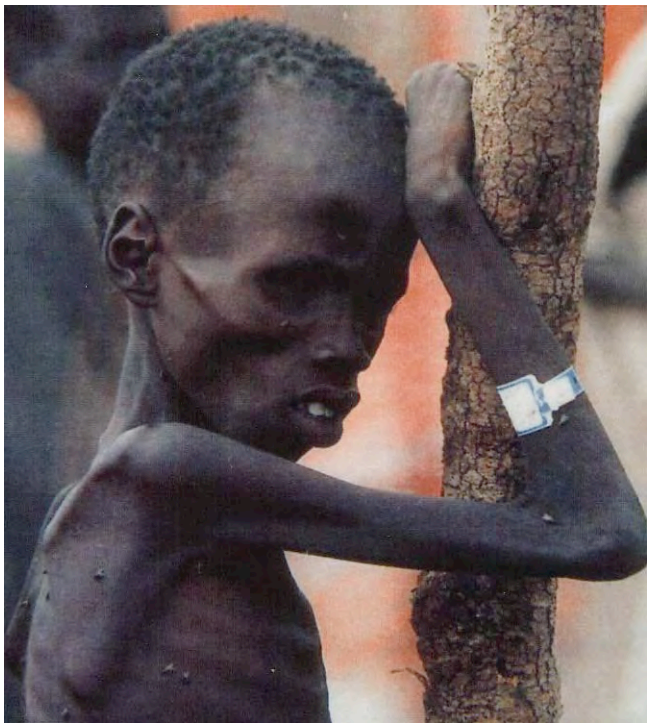

Plate 103 Marasmus Expression of sadness in a marasmic person. Courtesy of Cruz Roja España.

Plate 104 Melons, Squashes and Gourds Melons.

Plate 105 Melons, Squashes and Gourds Squash and pumpkins.

Plate 106 Metals Used in the Food Industry Stainless steel fermentation vessels and associated pipework in a modern brewery. (Courtesy of Courage Ltd and British Steel plc.)

Plate 107 Metals Used in the Food Industry Stainless steel's hygienic and corrosion resistance properties are vital in the brewing industry. Today, all new beer barrels are made with type 304 stainless steel. (Courtesy of British Steel plc.)

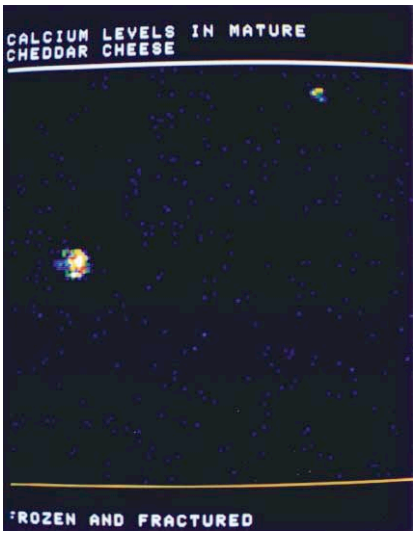

Plate 108
Microscopy Digital X-ray map showing calcium distribution (of calcium phosphate) in frozen hydrated cheddar cheese.

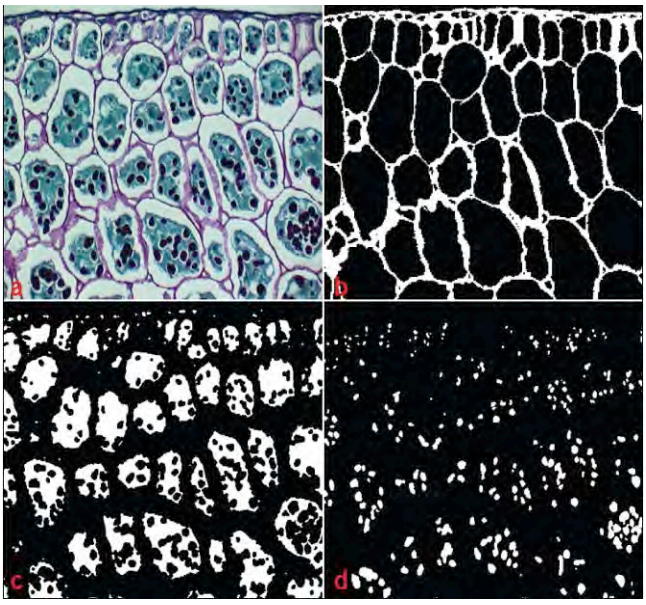

Plate 109 Microscopy Color segmentation of a section from a pea cooked from frozen. (a) Original image; (b) cell walls; (c) cytoplasm; (d) gelatinized starch granules.

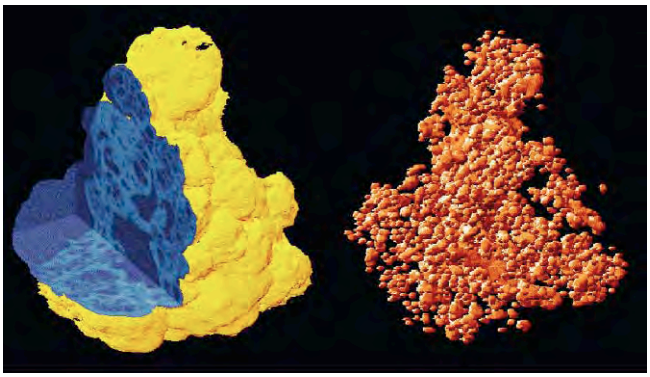

Plate 110 Microscopy Reconstruction of puffed wheat from X-ray tomographic sections. Left, reconstruction with cut-out to show the internal structure; right, reconstruction of the voids inside the puffed wheat. Courtesy of Dr Paul Jenneson, Surrey University, UK.

Plate 111 Oats Oat panicle of cultivated *Avena sativa.*

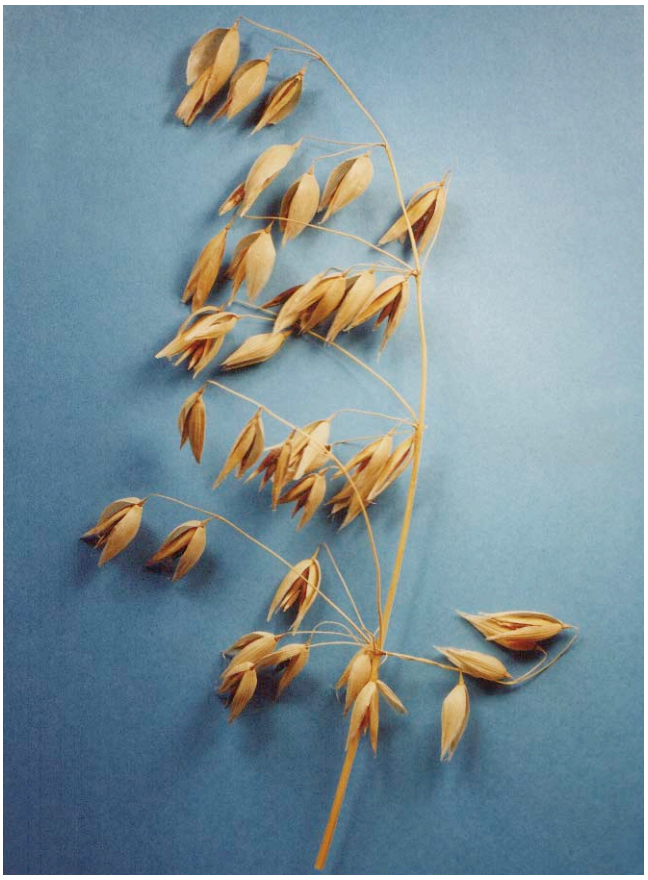

Plate 112 Oats Oat panicle of *Avena* spp. From a genetic cross between wild and hull-less oats. The oat derivative can be used for decorative purposes.

Plate 113 Oats Oat panicle of naked (hull-less) *Avena nuda*.

Plate 114 Oats Kernels of cultivated white oat, *Avena sativa* (left), hull-less oat, *Avena nuda* (center), and a derivative of a wild *Avena* sp. (right). Whole oat kernels are pictured on the top row and the corresponding groats on the bottom row.

Plate 115 Olive Oil Picture of a fossilized olive leaf. Reproduced from Psillakis N and Kastanas E (1999) *The civilization of olive: olive oil,* 2nd edn. Greek Academy of Taste, Karmanor, Iraklion, Crete (Greece), with permission.

Plate 116 Olive Oil Olive tree in Rethimnon, Crete (Greece), courtesy of V. Zambounis, Athens.

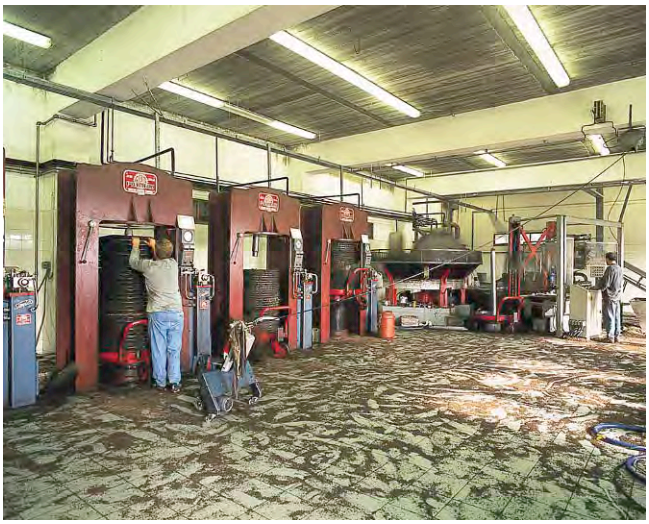

Plate 117 Olive Oil Hydraulic pressure unit (courtesy of ELAIS SA, Piraeus, Greece).

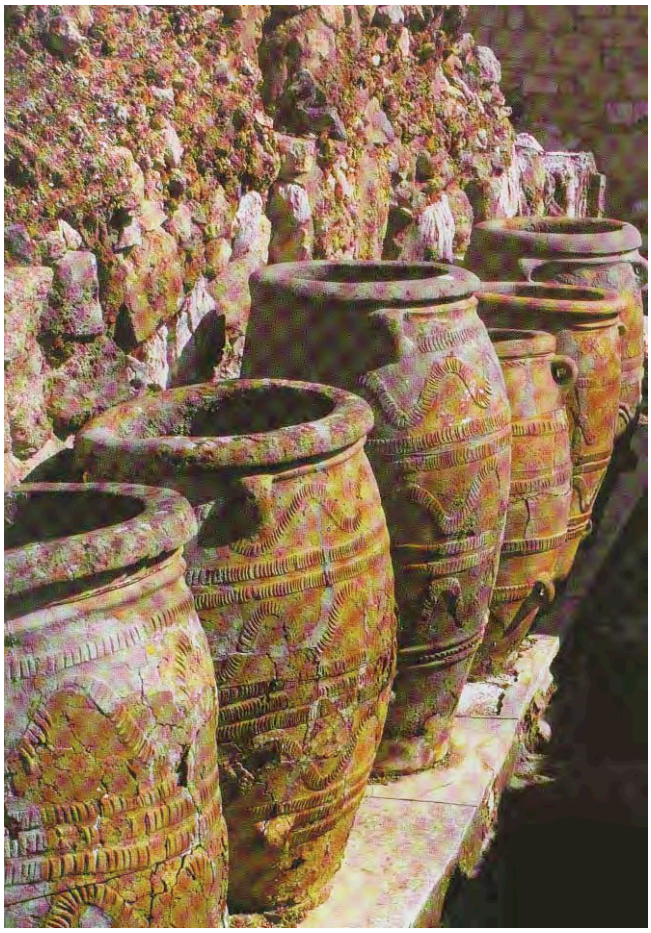

Plate 118 Olive Oil Pottery from Knossos Palace. Reproduced from Psillakis N and Kastanas E (1999) *The civilization of olive: olive oil,* 2nd edn. Greek Academy of Taste, Karmanor, Iraklion, Crete (Greece), with permission.

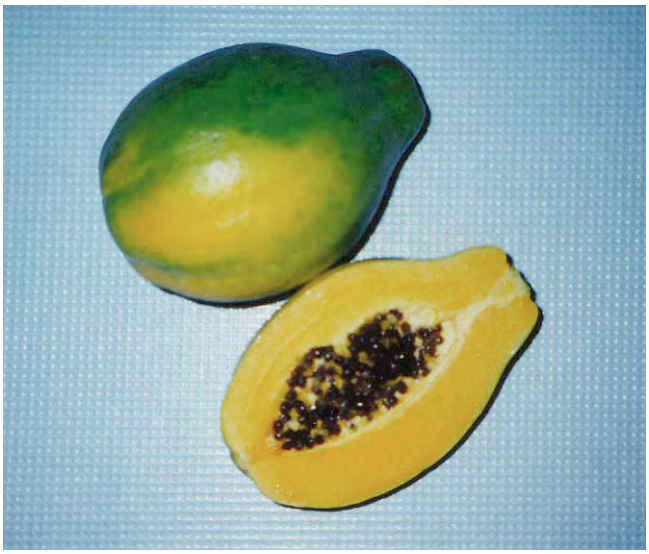

Plate 119 Papayas Rainbow cultivar growing in Hawaii, weighing 0.66kg.

Plate 120 Papayas Thailand cultivar growing in Thailand, weighing 2.50kg.

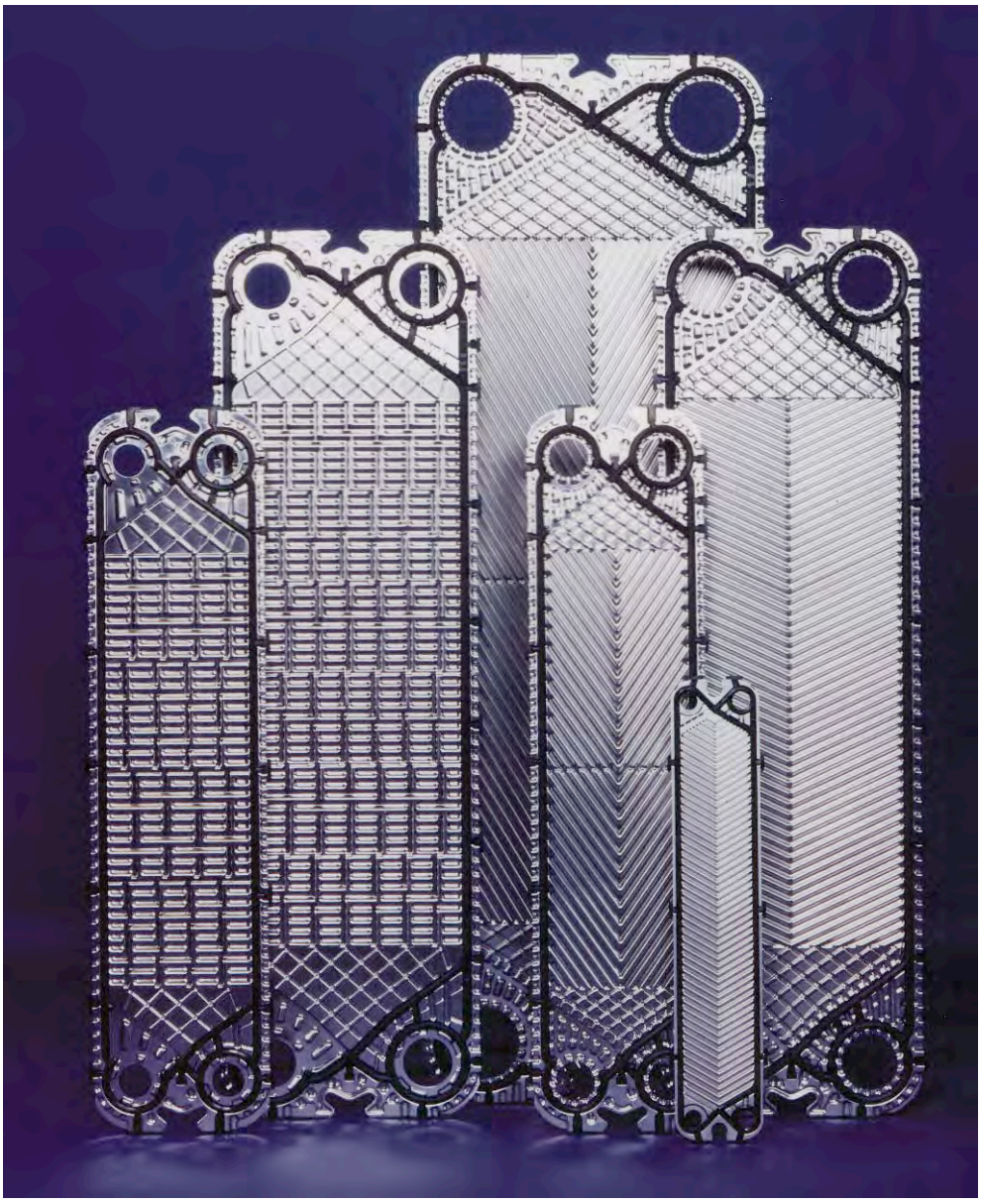

Plate 121 Pasteurization
Heat exchanger plates.
(Photo by courtesy of Tetra Pak Processing UK.)

Plate 122 Port Vila Nova de Gaia; bonded aread in Portugal.

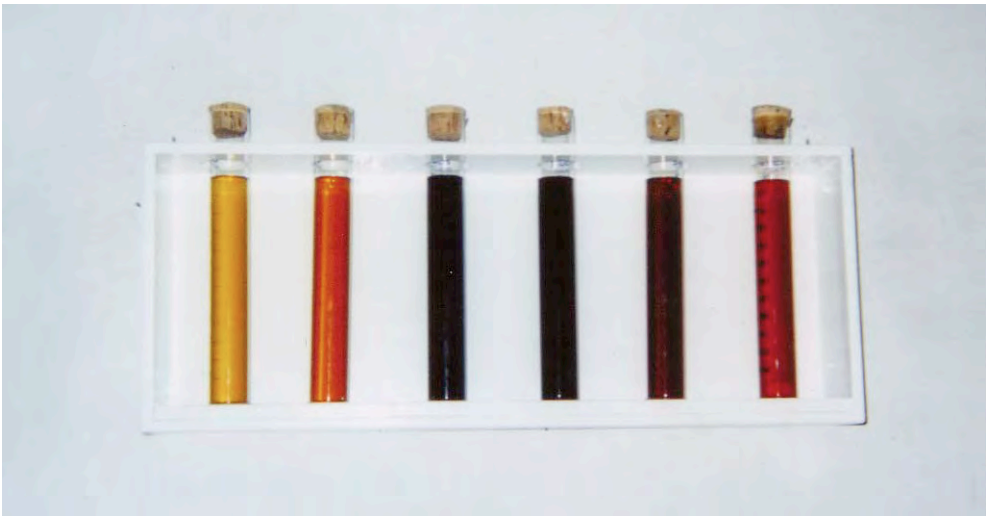

Plate 123 Port Samples of different types of port, showing the range of colors.

Plate 124 Quinoa Quinoa plant. Courtesy of Ing. Carlos Nieto C. Pronaleg-Iniap.

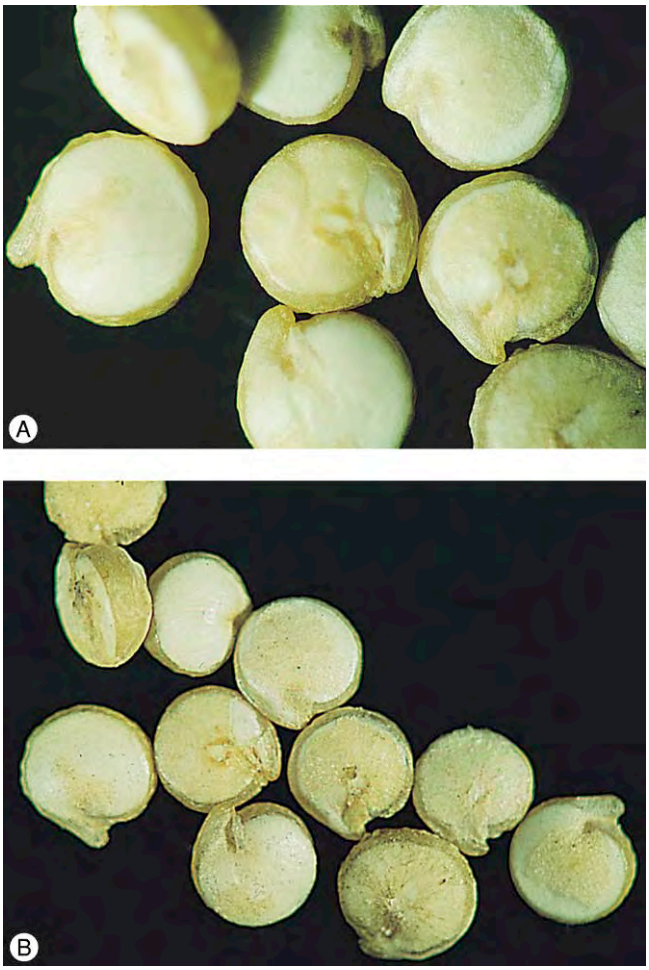

Plate 125 Quinoa Quinoa seeds. Courtesy of Silvia Valencia Ch.

Plate 126 Single-cell Protein Two commercial products elaborated with 100% dried *Spirulina maxima* biomass, sold as nutraceuticals in health food stores in Mexico City.

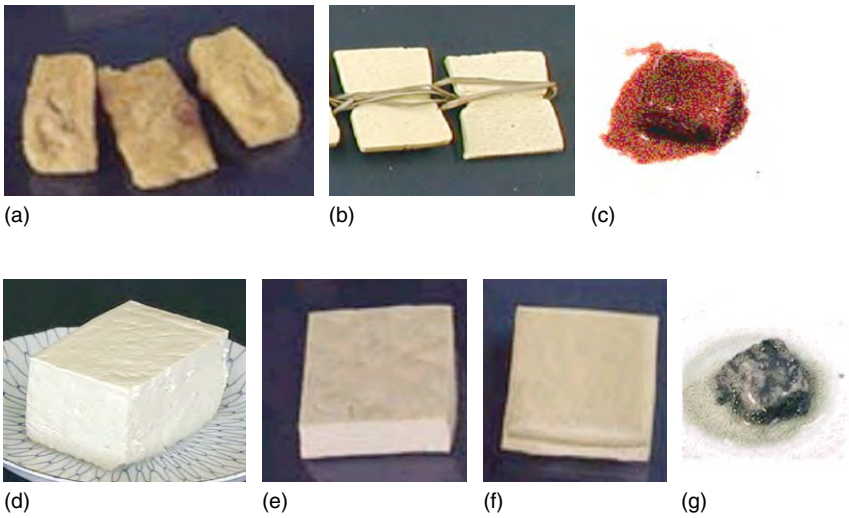

Plate 127 Soy (Soya) Cheese Photographs of various tofu and its derivatives. a, fried tofu; b, dried frozen tofu; c, furu; d, momen tofu; e, firm tofu; f, dried tofu; g, choutofu.

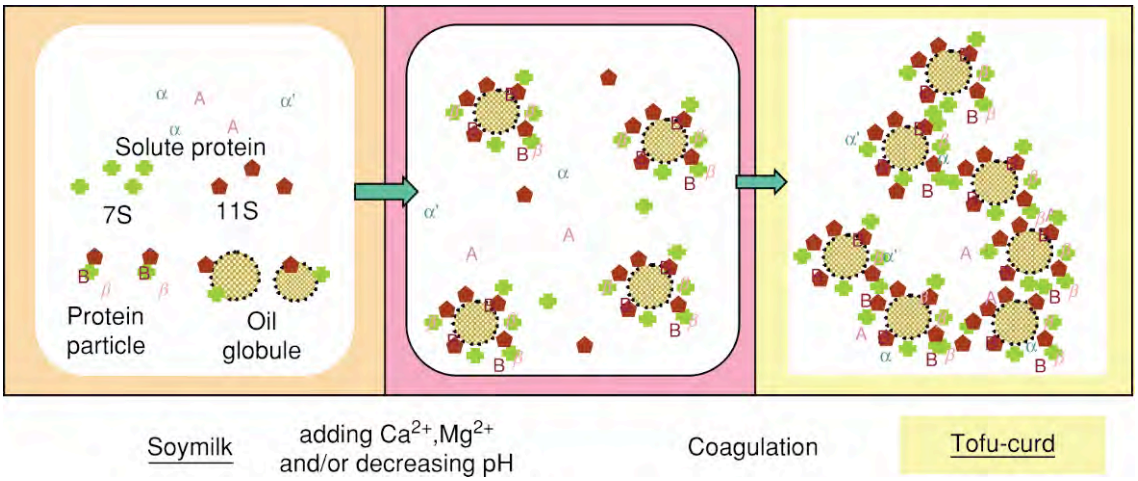

Plate 128 Soy (Soya) Cheese Formation of tofu-curd from soymilk (speculation). Oil globule is packed with triple layers of the proteins; oil-body's (oleosin), particulate and solute proteins.

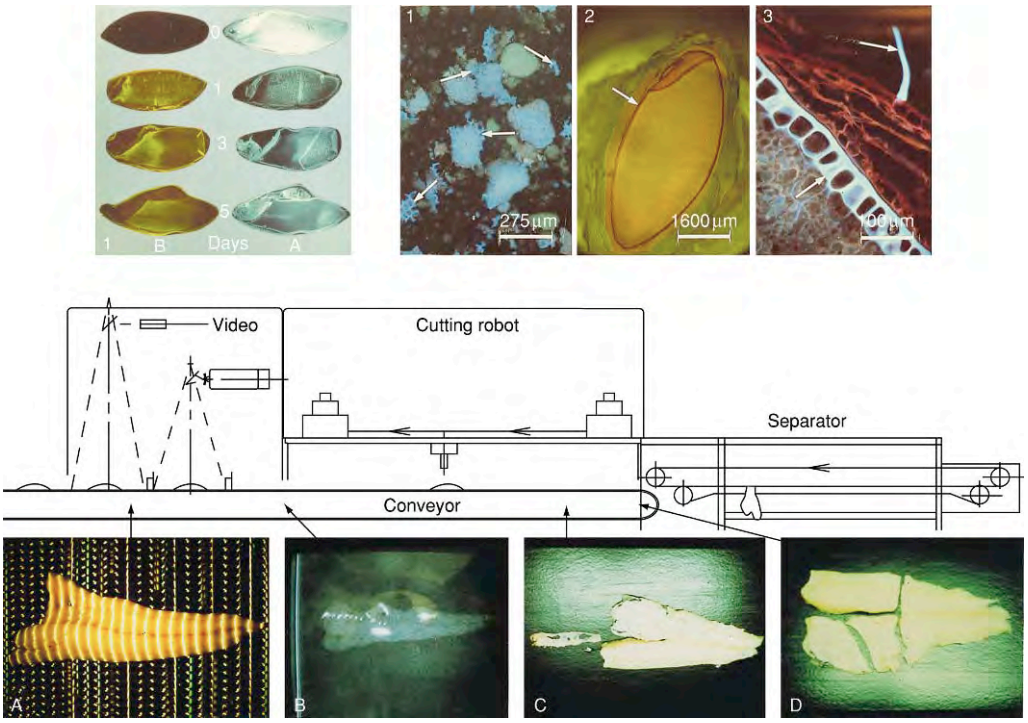

Plate 129 Spectroscopy
Examples of fluorescence microscopy and imaging reproduced from articles referred to in Munck L. (ed.) (1989) Fluorescence Analysis in Foods, Harlow: Longman. pp. 289. Munck L. Practical experiences in the development of fluorescence analyses in an applied food research laboratory, ibid. p.1-32. Fulcher RG, Irwing DW and de Francisco A. Fluorescence microscopy: Applications in food analysis, ibid. 59-109. Jensen SAa, Reenberg S and Munck L. Fluorescence analysis in fish and meat technology, ibid. 171-180.

Plate 130 Spices and Flavoring (Flavouring) Crops Spices and flavouring crops: Rhizome of green ginger (*Zingiber officinale)*.

Plate 131 Spices and Flavoring (Flavouring) Crops Spices and flavouring crops: Rhizome of turmeric (*Curcuma domestica)*.

Plate 132 Spices and Flavoring (Flavouring) Crops Spices and flavouring crops: Rhizome of mango ginger (*Curcuma amada).*

Plate 133 Tea Tea-tasting session (courtesy of the Tea Board of Kenya, Nairobi).

Plate 134 Tomatoes Harvest aid for fresh market tomatoes. Fruit are sorted and accumulated on a hopper on the machine. The fruit are periodically transferred to half-tonne bins for transport to the packing house.

Plate 135 Tomatoes Mechanical harvesting of processing tomatoes. Plants are cut at ground level, and elevated on to a shaker which separates the fruit from the plants.

Plate 136 Tomatoes Tray packs of glasshouse tomatoes, harvested once ripe.